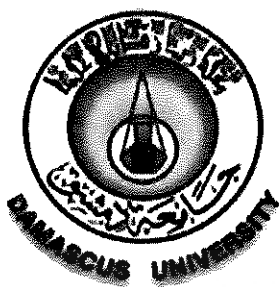


The logo of Damascus University is a circular emblem. It features a central yellow oil lamp with a flame, set against a background of radiating yellow lines. The lamp is encircled by a purple ring. The entire emblem is surrounded by a larger circle containing Arabic text at the top and English text at the bottom. The Arabic text reads 'وقل رب زدني علما' (O Lord, increase my knowledge). The English text reads 'Damascus University'.

برامج إحصائية (1) Statisitcal Software



برامج إحصائية (1)
السنة الثالثة إحصاء
قسم الإحصاء الرياضي



منشورات جامعة دمشق

كلية العلوم

برامج إحصائية (1)

الدكتور

محمد جاسم المحمد

مدرس في قسم الرياضيات

الدكتور

حسان عاقل

أستاذ مساعد في قسم الإحصاء

1434 - 1435 هـ

2013 - 2014 م

جامعة دمشق



٢٠١٠

9 PREFACE مقدمة 9

13PSEUDO-RANDOM NUMBERS الفصل الأول: الأعداد شبه العشوائية

13 INTRODUCTION مقدمة 1.1

16THE RANDOM NUMBER CYCLE دورة الرقم العشوائي 2.1

3.1 طرق الحصول على أعداد عشوائية

17RANDOM NUMBER GENERATORS

17 RANDOM NUMBER TABLES جداول الأرقام العشوائية 1.3.1

19مولدات الأعداد العشوائية 2.3.1

21متتاليات الأعداد شبه العشوائية 3.3.1

30 نمذجة المتغيرات العشوائية حاسوبياً 4.1

39 نمذجة المتغيرات العشوائية المنفصلة 5.1

42 EXERCISES تمارين 6.1

الفصل الثاني: المحاكاة باستخدام الخوارزميات العشوائية

45 SIMULATIONS USING RANDOMIZED ALGORITHMS

45 INTRODUCTION مقدمة 1.2

46 RANDOMIZED ALGORITHMS الخوارزميات العشوائية 2.2

47 أمثلة على الخوارزميات العشوائية 3.2

51 MONTE CARLO SIMULATION المحاكاة بطريقة مونتي كارلو 4.2

51 COMPUTER SIMULATION المحاكاة الحاسوبية 1.4.2

52 MONTE CARLO METHODS طرق مونتي كارلو 2.4.2

57 تطبيقات على طرق مونتي كارلو 3.4.2

84 EXERCISES تمارين الفصل 5.2

87 DATA ANALYSIS الفصل الثالث: تحليل البيانات

87 مقدمة 1.3

89 أهمية تحليل البيانات 2.3

92 مجال تحليل البيانات 3.3

98 خوارزمية تحليل البيانات 4.3

101 ملخص الفصل 5.3

الفصل الرابع: عرض خصائص البيانات الإحصائية ودراساتها

103 .. DISPLAYING AND EXPLORING STATISTICAL DATA

103 INTRODUCTION مقدمة 1.4

106	2.4 العرض البياني GRAPHICAL DISPLAY
106	1.2.4 وصف البيانات في متغير واحد
110	2.2.4 المدرج التكراري HISTOGRAM
113	2.2.4 مخطط مقارنة العينات مع توزيع نظري
115	3.2.4 العرض الصندوقي BOX PLOTS
119	4.2.4 مقارنة البيانات باستخدام المخطط الصندوقي BOX PLOTS
121	5.2.4 مخطط الأخطاء ERROR BARS
125	6.2.4 مخطط الساق والورقة STEM AND LEAF PLOT
128	7.2.4 المخطط التبعثري SCATTER PLOT
131	3.4 التحليل الاستكشافي للبيانات EXPLORATORY DATA ANALYSIS
144	4.4 تمرينات EXERCISES

الفصل الخامس: الفرضيات و الاختبارات الإحصائية

147	STATISTICAL TESTS AND HYPOTHESIS
147	1.5 الفرضية الاحصائية THE STATISTICAL HYPOTHESIS
148	1.1.5 الفرضية البديلة ALTERNATIVE HYPOTHESIS
148	2.1.5 فرضية العدم NULL HYPOTHESIS
148	3.1.5 اختبار الفرضيات HYPOTHESIS TEST
149	2.5 خوارزمية اختبار الفرضيات
150	3.5 أنواع الخطأ في اختبار الفرضيات
151	1.3.5 الخطأ من النوع الأول TYPE I ERROR
151	2.3.5 الخطأ من النوع الثاني TYPE II ERROR
153	4.5 الاختبارات الإحصائية STATISTICAL TESTS
153	1.4.5 اختبارات وسيطية (OR TESTS) PARAMETRIC METHODS
154	2.4.5 اختبارات غير وسيطية (OR TESTS) NONPARAMETRIC METHODS
155	5.5 الاختبارات الوسيطية PARAMETRIC TESTS
156	1.5.5 اختبارات ستودنت STUDENT TEST
156	1.1.5.5 اختبار ستودنت حول مجتمع واحد
160	2.1.5.5 اختبار ستودنت حول مجتمعين Two-Sample T Test
	3.1.5.5 اختبار ستودنت حول مجتمعين مستقلين
163	Two-Sample Independent T Test
	4.1.5.5 اختبار ستودنت لمقارنة متوسطي عينتين مرتبطتين

169	Two-Sample Paired T Test
173	2.5.5 ANOVA اختبارات التباين
175	1.2.5.5 One-way ANOVA تحليل التباين بمتغير واحد مستقل
182	2.2.5.5 Two-way ANOVA تحليل التباين بمتغيرين مستقلين
	3.2.5.5 اختبار تحليل التباين للعينات المرتبطة
192	Repeated Measures ANOVA
209	6.5 تمرينات EXERCISES
223	الفصل السادس: الاختبارات غير الوسيطة
223	1.6 مقدمة INTRODUCTION
224	2.6 اختبارات كاي مربع CHI-SQUARE TESTS
226	1.2.6 اختبار الاستقلالية CHI-SQUARE INDEPENDENCE TEST
232	2.2.6 اختبار جودة التوافق CHI-SQUARE GOODNESS-OF-FIT TEST
237	4.5 اختبار توافق المراقبين OBSERVER AGREEMENT
242	4.6 اختبار MANN-WHITNEY TEST
245	5.6 اختبار الإشارة WILCOXON SIGNED RANKED TEST
249	6.6 اختبار كروسكال واليس لأجل K عينة مستقلة
251	7.6 اختبار FRIEDMAN ANOVA
258	8.6 تمرينات EXERCISES
265	الفصل السابع: تطبيقات عملية
303	الملحق A: مدخل إلى SPSS SPSS TUTORIAL
	الملحق B: مهارات أساسية في الـ SPSS
325	SPSS BASIC SKILLS TUTORIAL: BASIC ANALYSES
345	الملحق C
391	الملحق D
417	المصطلحات العلمية
425	المراجع REFERENCES



بسم الله الرحمن الرحيم

مقدمة Preface

يفرز العالم الواقعي مسائل معقدة يتطلب حلّها التغلب على الكثير من التحديات. من جهة أن هذه المسائل تتسم بالتعقيد لكونها تتأثر بمؤثرات كثيرة يتطلب تحديدها جهداً ووقتاً كبيراً. و من جهة ثانية ، طريقة تأثير هذه المؤثرات ليست واضحة دائماً بل في كثير من الأحيان يكتنفها الغموض- أي من الصعب التنبؤ بطريقة و حجم تأثيرها.

أصبح لدينا كمّ هائل من البيانات حول مختلف الظواهر المحيطة بنا بسبب التطور المعرفي الهائل. لهذه البيانات أهمية خاصة كونها تساعد في فهم هذه الظواهر بشكل أفضل. ينتج تقريباً عن كل فعاليات الحياة بيانات. بيانات عن الرأي العام حول قضية معينة ، بيانات عن تجارب بيولوجية ، بيانات عن البيئة المحيطة بنا ، بيانات عن الأدوية الجديدة ، و غيرها كثير. تتسم البيانات بأنها تخفي معلومات كثيرة بداخلها. أي أن البيانات هي مواد أولية لا تظهر أي معلومات بشكل مباشر ، و يجب تحليلها للحصول منها على معلومات مفيدة حول الظواهر التي تصفها.

الهدف الرئيسي لهذا الكتاب هو تزويد القارئ بمجموعة من التقنيات و الأدوات التي تمكن من التعامل مع مسائل واقعية. تعد طرق مونتي كارلو من الطرق المهمة التي تستخدم لحل مسائل واقعية و ذلك عن طريق بناء نماذج و محاكاة هذه النماذج ، و من ثم استنباط حلول للمسائل من خلال نتائج المحاكاة. طرق مونتي كارلو هي طرق عامة لحل المسائل و تعد وسيلة فعالة للتعامل مع المسائل التي يكتنفها الغموض. لذلك تمت مناقشة طرق مونتي كارلو في هذه الكتاب، بما أن طرق مونتي كارلو ترتبط بمبدأ العشوائية فقد أفرد فصل كامل للحديث عن الأعداد العشوائية و بعض خصائصها.

يناقش هذا الكتاب من جهة ثانية ، موضوع التحليل الإحصائي للبيانات و استنباط نتائج منها. حيث جرى مناقشة العديد من تقنيات إختبار الفرضيات الإحصائية و كيفية تحليل البيانات المتوفرة للحصول على استنتاجات حول هذه الفرضيات.

تم عرض المادة العلمية لهذا الكتاب على عدة فصول و ملاحق. يعطي الفصل الأول مقدمة الأعداد العشوائية و استخداماتها في مختلف العلوم. كما حددنا في هذا الفصل مواصفات خوارزميات الأعداد العشوائية و بيّنا معايير جودة مولدات الأعداد العشوائية. يعرّض الفصل الثاني الخوارزميات العشوائية و استخداماتها في مختلف المسائل العملية. بشكل خاص ، تمت مناقشة طرق مونتي كارلو و المحاكاة الحاسوبية ، و تم التركيز على توصيف المسائل التي تنطبق عليها طرق مونتي كارلو. يعرّض الفصل الرابع تقنيات عرض البيانات و تنقيتها من الشذوذ. إذ يتعلم القارئ في هذا الفصل العديد من طرق عرض البيانات بيانياً و مجالات استخدام كل وسيلة عرض. كذلك يعرض الفصل إجراء متكاملأ لاستكشاف البيانات و معرفة طبيعة توزيعها و تنقيتها من الشذوذ. يعرض الفصل الخامس موضوعات مهمة للغاية. يدخل القارئ إلى مفهوم الفرضية الاحصائية ، مفهوم الأخطاء و أنواعها و أهمية كل نوع ، و يعرض كذلك العديد من الاختبارات الاحصائية الوسيطة. أعطيت الأمثلة المناسبة في كل الأماكن التي يحتاج فيها القارئ إلى التوضيح. يعرض الفصل السادس الاختبارات غير الوسيطة. حيث تم عرض العديد من الاختبارات غير الوسيطة مع الأمثلة المناسبة.

لقد اتبعنا أسلوباً جديداً في الكتاب و ذلك بعرض الإجراءات الإحصائية الوسيطة و غير الوسيطة مع إعطاء النموذج العام للمسائل التي ينطبق عليها هذا الإجراء. الهدف هو تسهيل عملية ربط هذه الإجراءات بالمسائل الواقعية بحيث يستطيع الطالب بسهولة (أو بمحاكمة بسيطة) معرفة أي إجراء مناسب للمسألة التي يتعامل معها. وأخيراً اهتم الكتاب كثيراً بتنمية مهارة كتابة نتائج تحليل البيانات بشكل علمي ، مستعدين عن المصطلحات التقنية و يعرض النتائج بشكل مختصر و لكن كافٍ. يعرض الفصل السابع للتطبيقات العملية ، فتم طرح مجموعة من البيانات المأخوذة من تجارب عمالية في مجالات مختلفة مع التركيز أكثر على البيانات البيولوجية. الهدف من هذا الفصل هو إغناء تجربة القارئ في التعامل مع معطيات كبيرة. كذلك، تعليم القارئ كيفية التحليل الدقيق للبيانات و استنباط أسئلة منها و من

ثم محاولة الإجابة عن هذه الأسئلة من خلال إجراء تحليلات مختلفة للبيانات. تركت أيضاً بعض المسائل و البيانات للقارئ للتمرين مع إعطاء بعض الإرشادات التي تساعد على التعامل مع البيانات.

أفردت الملاحق A, B, C لعرض واحد من أهم الحزم البرمجية الاحصائية SPSS. حيث عُرِضت مجموعة محدودة و لكن كافية من الإجراءات الاحصائية التي تضع القارئ على أرضية صلبة للانطلاق و التعمق في استخدام البرنامج الاحصائي. أيضاً خصصنا مساحة لا بأس بها للتعريف بكيفية تفريغ البيانات بشكل صحيح في شاشات الإدخال. خُصص الملحق D لإعطاء مجموعتين كبيرتين من البيانات المأخوذة من تجارب عالمية كبيرة. يمكن للقارئ تفريغ هذه البيانات و تحليلها.

طُرِحت في كل فصل مجموعة من التمرينات التي تختبر مدى استيعاب القارئ للمادة العلمية المطروحة و تساعد على تنمية قدراته على تحليل المسائل الواقعية و اختيار أفضل إجراء لمعالجتها. بعض هذه التمرينات مأخوذة عن دراسات علمية، لذلك أنصح كثيراً القراء بمحاولة حلها كاملة.

إرشادات لقراءة الكتاب

تتمحور مفردات الكتاب حول مفهومين واسعين: (1) الطرق العشوائية لحل المسائل (الفصلين الأول و الثاني) ، (2) مفهوم التحليل الإحصائي باستخدام معالج البيانات الإحصائية SPSS (الفصول الثالث إلى السابع). لذلك ، فإن إحدى النصائح لقراءة الكتاب تتم بالشكل التالي:

1. قراءة الفصلين الأول و الثاني بهذا الترتيب. على القارئ أن يضع في حسبانته أن هذين الفصلين يشكلان وحدة متكاملة تدخل القارئ بشكل معمق نوعاً ما إلى خوارزميات الأعداد العشوائية و طرق مونتّي كارلو.
2. تُقرأ الفصول من الثالث إلى السادس بهذا الترتيب. من المفيد للطالب أن ينتقل إلى الملاحق عند الضرورة للإستزادة حول كيفية إدخال البيانات و تفعيل الاختبارات الإحصائية التي أعطت الخرج المعروض في هذه الفصول.

أنصح بعد الإنتهاء من كل فصل بمحاولة حل التمارين لأنها مصممة للمساعدة في تعميق فهم القارئ للمفاهيم المطروحة في هذه الفصول.

3. يُقرأ بعددِ الفصل السابع الذي يعطي القارئ المزيد من المرونة في تحليل المعطيات. هناك الكثير من التطبيقات المُعالجة و لكن تركت بعض الثغرات قصداً لتعطي القارئ فرصة للتفكير.

أخيراً أمل أن يحقق الكتاب الهدف المرجو منه و أن يكون عوناً للطلاب على التعرف إلى تحليل البيانات و استخدامها في مجال عملهم.

و الله الموفق

دمشق في 2013/4/24

المؤلفون

الأعداد شبه العشوائية Pseudo-Random Numbers

Anyone who considers arithmetical methods of producing random digits is, of course, in a state of sin.

John von Neumann (1951)

إن أي شخص يولد أعداداً عشوائية من صيغ رياضية هو بالطبع غير صائب (مذنب).

جون فون بايمان 1951

1.1 مقدمة Introduction

تلعب الأعداد (شبه) العشوائية دوراً هاماً في كثير من التطبيقات العملية. يمكن لنا الحديث عن العديد من التطبيقات العملية للأعداد العشوائية. أولاً، تستخدم هذه الأعداد في محاكاة النظم الواقعية، كما سوف نرى في الفصل الثاني. إن طرق مونتي كارلو و طرق لاس فيغاس في المحاكاة تعتمد بشكل جوهري على هذه الأعداد. ثانياً، تستخدم الأعداد العشوائية أيضاً في تطبيقات عملية أخرى غير المحاكاة: التشفير، لعب القمار على الأنترنت، و الألعاب المنزلية هي أمثلة قليلة على استخداماتها. في التشفير تستخدم

الأعداد العشوائية لتوليد مفاتيح من الصعب التنبؤ بها (تستخدم لفتح رسائل مُشفرة). تستخدم الأعداد العشوائية في صناعة القمار على الأنترنت لتقديم نتائج عشوائية لمجموعة من الألعاب مثل كازينو لعبة البوكر ، لعبة الورق ، و لعب أخرى. وإلى جانب الاستخدام المهني ، غالباً ما تستخدم الأرقام العشوائية بشكل كبير في الحياة اليومية. حيث تستخدم في المقام الأول بهدف استبعاد الإنحياز البشري والغش. على سبيل المثال، إذا أردنا إختيار شخص ما من مجموعة من الأشخاص و ذلك لإعطائه جائزة معينة ، فإن الأعداد العشوائية تلعب دوراً كبيراً في عملية الإختيار لأنها تسمح بإنتقاء عادل و غير منحاز .

لايوجد في الواقع العملي شئ اسمه عدد عشوائي. على سبيل المثال هل العدد 3 هو عدد عشوائي؟ و بالتالي فإن الحديث عن أعداد عشوائية بالمعنى المطلق للكلمة غير دقيق من الناحية العلمية. نستطيع الحديث عن متتالية من الأعداد المستقلة عن بعضها البعض التي تخضع لتوزيع معين. بسبب استقلالية هذه الأعداد عن بعضها البعض فإننا نقول أن امكانية ظهور أي عدد في المتتالية لايعتمد على ظهور عدد معين آخر في المتتالية أو عدمه. أي أن جميع الأعداد لها الفرصة نفسها في الظهور في المتتالية (بمعني آخر عملية إنتقاء الأعداد تكون عشوائية). بالإضافة إلى ذلك ، أي عدد له قيمة احتمالية محددة ليتم إنتقاؤه من مجال عددي معطى. على سبيل المثال ، إذا إقترعنا عينة من الأعداد العشوائية من مجموعة أعداد أكبر تخضع لتوزيع منتظم، فإن لكل عدد من المجموعة الكبيرة الفرصة نفسها للظهور في العينة.

تعتمد كل مولدات الأعداد العشوائية على خوارزميات رياضية لتوليد هذه الأعداد. لذلك، فإن الأعداد المولدة من هذه الخوارزميات ليست عشوائية بالمعنى الدقيق لكلمة عشوائية (بسبب إعتمادها على قانون محدد). بالرغم من أن توليد أعداد عشوائية ليس ممكناً ، فإننا يجب أن نحاول الحصول على أعداد قريبة للعشوائية (أي شبه عشوائية). يعتمد إتسام مجموعة متتالية من الأعداد بصفة العشوائية على مجموعة من الخصائص. بمعنى آخر جودة متتالية أعداد عشوائية

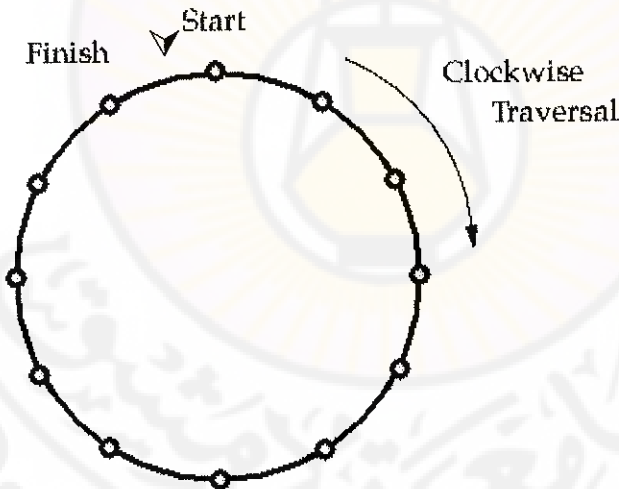
نقاس بقربها من مفهوم العشوائية وهذا القرب يمكن أن يقاس بمجموعة من الخصائص. هناك العديد من الخصائص التي يجب أن تتمتع بها خوارزميات توليد الأعداد العشوائية (أو بشكل أكثر دقة: شبه العشوائية) حتى تولد متتاليات أعداد عشوائية بمواصفات جيدة. من أهم هذه الخصائص نذكر:

- يجب أن تكون متتاليات الأعداد شبه العشوائية المولدة مستقلة (Uncorrelated Sequences). هذا يعني أن أي متتالية من أعداد عشوائية مولدة بواسطة خوارزمية معينة تكون مستقلة عن أي متتالية أخرى مولدة باستخدام الخوارزمية نفسها.
- يجب أن تملك دوراً طويلاً Long Period. يجب على أي خوارزمية توليد أعداد شبه عشوائية أن تملك دوراً طويلاً. بمعنى آخر ، إن أي تكرار لنفس المتتالية يجب أن يحدث بعد توليد متتالية طويلة من الأعداد العشوائية. طول الدور مهم للغاية لأنه إذا كان دور الخوارزمية قصيراً فإن متتالية الأرقام سوف تتكرر بعد عدد معين من الأعداد العشوائية المولدة فعلياً. و بالتالي، تفقد هذه الأعداد قيمتها خصوصاً في التطبيقات التي تحتاج إلى توليد عدد كبير (جداً) من الأعداد شبه العشوائية.
- يجب أن تكون منتظمة التوزيع Uniformity. الخوارزمية يجب أن تولد أعداد منتظمة التوزيع ضمن مجال محدد. على سبيل المثال ، إذا أردنا أن نولد أعداداً عشوائية ضمن المجال $[0, 1]$ ، عندئذٍ إذا حدث أن أكثر من نصف هذه الأعداد المولدة تقع ضمن المجال ، مثلاً $[0, 0.1]$ ، فإن هذه المتتالية هي متتالية سيئة. هناك ثمة علاقة بين الإنتظام في التوزيع و الإستقلالية بين متتاليات الأعداد العشوائية. عندما تكون خاصة الإنتظام في التوزيع غير محققة فإنه غالباً ماتكون بعض المتتاليات المولدة مترابطة (أي ليست مستقلة).
- يجب أن تملك الكفاءة أو الفعالية Efficiency. فعالية الخوارزميات المولدة للأرقام العشوائية مهمة للغاية. كما سوف نرى في الفصل الثاني ،

فإن أي عملية محاكاة لنظام معين ، تحتاج إلى ملايين أو حتى مليارات الأعداد العشوائية. و بالتالي ، إذا لم تكن الخوارزمية سريعة فإن عملية المحاكاة سوف تكون بطيئة و غير عملية.

2.1 دورة الرقم العشوائي The Random Number Cycle

تنتج جميع المولدات العشوائية تقريباً متتالية أعداد عشوائية صحيحة. للحصول على أعداد عشوائية ، مثلاً حقيقية ، فإنه يجب معالجة هذه الأعداد رياضياً للحصول منها على أعداد حقيقية. يمكن عرض دورة الرقم العشوائي بدلالة الأعداد الصحيحة أو الأعداد الحقيقية. للسهولة ، سوف نعرض دورة الأعداد العشوائية الصحيحة. يمثل الشكل التالي دورة الرقم العشوائي الصحيح.



يوضح الشكل أعلاه دورة عدد عشوائي ممثل على شكل متتالية من 12 عدد صحيح. كل نقطة على محيط الدائرة هي عبارة عن عدد صحيح مختلف. لتوليد أعداد عشوائية نبدأ من الدائرة المشار إليها بالسهم (Start) ثم نتحرك باتجاه عقارب الساعة و

ننتهي عند العدد الصحيح الذي يسبق مباشرة العدد الذي بدأنا منه (الدائرة المشار إليها بـ Finish). توضح طبيعة دورة الرقم العشوائي القضايا التالية:

1. المتتالية المولدة لها عدد منتهى من الأعداد.
2. يتم توليد المتتالية بترتيب معين.
3. المتتالية المولدة تتكرر إذا تجاوزنا طول دور المولد.

3.1 طرق الحصول على أعداد عشوائية

Random Number Generators

سوف نعرض في هذا الجزء بعض المولدات العشوائية. هدفنا ليس تغطية شاملة لجميع المولدات العشوائية المعروفة و لا الإسهاب في دراسة خصائص هذه المولدات لأن هذا المقرر ليس مخصصاً لدراسة نظرية الأعداد العشوائية. إنَّ ، الهدف هو تسليط بعض الضوء على بعض هذه المولدات و على بعض خصائصها و ذلك لإعطاء الطالب فكرة عن كيفية توليد الأعداد العشوائية.

توجد ثلاث طرائق للحصول على الأرقام العشوائية: جداول الأرقام العشوائية، مولدات الأعداد العشوائية، و متتاليات الأعداد شبه العشوائية.

1.3.1 جداول الأرقام العشوائية Random Number Tables

يحتوي الجدول (2) على مجموعة من الأرقام العشوائية. يتم تشكيل جداول الأرقام العشوائية مثل الذي يظهر بالجدول (2) من خلال ترقيم مجموعة من الأوراق أو كرات (أو ماشابه ذلك) من 0 إلى 9 ثم توضع هذه الأوراق أو الكرات في وعاء وتخلط جيداً. نسحب بشكل عشوائي ورقة ثم نسجل الرقم الظاهر عليها ونعيدها مرة أخرى إلى الوعاء، ثم نعيد التجربة حتى نحصل على العدد الذي نريد من الأرقام العشوائية.

73735	45963	78134	63873
02965	58303	90708	20025
98859	23851	27965	62394
33666	62570	64775	78428
81666	26440	20422	05720
15838	47174	76866	14330
89793	34378	08730	56522
78155	22466	81978	57323
16381	66207	11698	99314
75002	80827	53867	37797
99982	27601	62686	44711
84543	87442	50033	14021
77757	54043	46176	42391
80871	32792	87989	72248
30500	28220	12444	71840

الجدول (2): جدول أرقام عشوائية.

هناك العديد من الطرق الأخرى للحصول على الأرقام العشوائية في الجدول (2). من هذه الطرق نذكر دولا ب الحظ. حيث يقسم هذا الدولا ب إلى عشرة أقسام متساوية وتكتب الأرقام العشرية من 0 إلى 9 بشكل متتالٍ على أقسام الدولا ب. يتم تدوير الدولا ب ويُقرأ الرقم الظاهر من خلال فتحة الدولا ب ويسجل. تعاد العملية ذاتها حتى نحصل على عدد الأرقام العشوائية التي نريد.

أسئلة للتأمل:

1. لماذا تعد الأرقام التي ينتجها دولا ب الحظ عشوائية؟
2. أذكر طريقة لتوليد أرقام عشوائية من 1 إلى 6. أذكر طريقة أخرى لتوليد أرقام عشوائية ثنائية (0 أو 1).

لاحظ أن كل رقم في الجدول (2) هو في الواقع قيمة للمتغير العشوائي Z المنتظم التوزيع على المجموعة $\{0, 1, \dots, 9\}$ ، الذي يمكن التعبير عنه بالشكل التالي:

$$f_z = p(Z = z) = 0.1 \quad ; \quad z = 0, 1, \dots, 9 \quad (1)$$

إذا أردنا الحصول على 7 أرقام للمتغير العشوائي Z ، فإننا نحصل على الأرقام العشوائية 0, 9, 5, 1, 6, 8، إذا أخذنا 7 أرقام من السطر الأول بشكل متتالي من بداية الجدول (2). ونحصل على متتالية الأرقام العشوائية 7, 5, 8, 4, 6, 8، إذا أخذنا 7 أرقام من بداية العمود الأول.

أما إذا أردنا الحصول على قيم لمتغير عشوائي منتظم التوزيع على المجال المفتوح $[0, 1]$ ، نسحب من الجدول (2) مجموعة من الأرقام و نضعها على يمين الفاصلة العشرية. على سبيل المثال، تمثل الأعداد:

$$(2) \quad 0.865, 0.159, 0.079$$

قيماً لمتغير عشوائي منتظم التوزيع على المجال $[0, 1]$ ، حيث تتألف كل قيمة من ثلاثة أرقام عشرية مأخوذة من السطر الأول. بينما تمثل القيم:

$$(3) \quad 0.8651, 0.5907, 0.9566, \dots$$

قيماً لمتغير عشوائي منتظم التوزيع على المجال $[0, 1]$ ، حيث تتألف كل قيمة من أربعة أرقام عشرية مأخوذة من السطر الأول. و أخيراً تمثل القيم التالية

$$(4) \quad 0.86, 0.48, \dots, 0.75, \dots$$

قيماً لمتغير عشوائي منتظم التوزيع على المجال $[0, 1]$ ، حيث تتألف كل قيمة من رقمين عشريين مأخوذة من العمود الأول.

إن متتالية القيم في (2) و (3) و (4) هي قيم مستقلة لمتغير عشوائي مُقَرَّب و منتظم على المجال المفتوح $[0, 1]$.

2.3.1 مولدات الأعداد العشوائية

ندعو مولداً عشوائية كل واسطة تقنية تحضر أرقام عشوائية. سوف نعرض في هذه الفقرة فكرة عن المولدات الفيزيائية

المولدات الفيزيائية

عادةً ما تستخدم أجهزة التشويش الإلكترونية كالمتممات و الترانزستورات لبناء المولدات الإلكترونية. لن ندخل هنا في التفاصيل التقنية لهذه الأجهزة وسندرس فقط واحدة من الطرائق الممكنة لبناء مولد أرقام عشوائية ثنائية.

لنفترض أن لدينا عدداً V يحسب عدد نهايات بيان تيار التشويش (الضجيج) لدارة إلكترونية التي تزيد عن مستوى محدد E_0 في زمن محدد Δt . يمكن تبسيط العمل أكثر باعتبار أن المولد يُنتج العدد الثنائي $v \bmod 2$ (v هي القيمة التي يعطيها العداد V)، أي يعطي الرقم 0 عندما يكون v زوجياً و الرقم 1 عندما يكون v فردياً. يمكن تمييز حالتين:

أولاً، إذا كان احتمال ظهور الصفر يساوي احتمال ظهور الواحد، عندها يمكن القول إن المولد ينتج متتالية أرقام ثنائية عشوائية.

ثانياً، إذا كان احتمال ظهور الصفر مختلفاً عن احتمال ظهور الواحد، عندها يمكن إضافة دارة جديدة تنظم الاحتمال أو وضع خوارزمية تنظم الاحتمال. على سبيل المثال، يمكن تجميع الأرقام الناتجة على شكل أزواج و إعطاء الرقم 0 عند الحصول على الزوج 01 و الرقم 1 عند الحصول على الزوج 10، و إهمال الزوجين 00 و 11. بما أن $P(01) = P(10) = P(1-p)$ ، فإن هذا المولد سينتج الرقمين 0 و 1 باحتمالين متساويين.

يحتوي مولد الأعداد العشوائية عادةً على m مولداً من الشكل السابق وتعمل المولدات بشكل مستقل عن بعضها البعض، و تمكنا من الحصول على أعداد عشوائية تقريبية من الشكل: $\gamma = 0.\alpha_0\alpha_1\dots\alpha_m$. على الرغم من أن هذه المولدات أفضل من طرق الحصول على الأعداد العشوائية من خلال الجداول كونها لا تحتاج لنفس المساحة التخزينية اللازمة لتخزين الجداول إلا أنها تعاني من مشاكل كثيرة.

1. لا يمكن استرجاع الأعداد العشوائية المحضرة بهذه المولدات مما يجعل عملية التحكم بالحسابات صعبة ولا يمكن إعادة الحسابات مرة أخرى باستخدام نفس المتتالية.

2. من الصعب إجراء الاختبارات الإحصائية للأعداد المولدة للتأكد من عشوائيتها لأن هذه الأرقام تولد وتستخدم في الآن ذاته. مما يستدعي إجراء اختبارات دورية للمولدات أو وضع دالة إضافية لاختبارها من خلال اختبارات إحصائية خاصة و كشف أي خلل إلكتروني قد يؤدي إلى توقف أحد المولدات مما يجعل التوزيع منحازاً.

3.3.1 متتاليات الأعداد شبه العشوائية

تولد متتاليات الأعداد شبه العشوائية من خلال معادلات رياضية. بسبب طبيعة توليدها، فإن هذه الأعداد تسمى بشبه العشوائية و يبرهن أن هذه المتتاليات تحقق جميع اختبارات العشوائية.

الشكل العام للعلاقات الرياضية التي تولد أعداد شبه عشوائية هو كما يلي:

$$\gamma_{n+1} = f(\gamma_n) ; n=0,1,...,i$$

حيث γ هو عدد معلوم يسمى القيمة البدائية أو Seed. يوجد العديد من العلاقات الرياضية لتوليد أعداد عشوائية لكل محاسنها و مساوئها.

طريقة منتصف التربيعات Middle-Square Method

تعتبر طريقة منتصف التربيعات من أقدم طرق توليد الأعداد العشوائية التي اقترحها نايمان في عام 1946. تعتمد هذه الطريقة على البدء من عدد معين مكون من عدد زوجي من الأرقام، ثم نربع هذا العدد، ثم نأخذ مجموعة الأرقام الواقعة منتصف مربع العدد، يستخدم العدد الناتج لتوليد عدد عشوائي جديد باستخدام الأسلوب نفسه.

يمكن تلخيص طريقة منتصف التربيعات خوارزمية بالخطوات التالية:

1. نقوم باختيار عدد X_0 مولفاً من n رقم (n زوجي). يسمى هذا العدد نواة أو Seed.

2. نقوم بتربيع هذا العدد لنحصل على عدد جديد R مكون من $2n$ رقم. إذا كان كان عدد أرقام R أقل من $2n$ ، نضيف أصفاراً إلى يسار R حتى تصبح عدد أرقامه مساوياً $2n$.

3. نقوم بإقتطاع الـ n رقم الواقعة بالمنتصف لنحصل على العدد العشوائي التالي الذي سوف يصبح أيضاً نواة لاقتراع عدد عشوائي جديد. كرر الخطوات الثلاث لنحصل على المتتالية المطلوبة من الأعداد العشوائية.

سوف نوضح عمل هذه الطريقة بالمثال التالي. لنفرض أننا بدأنا بالعدد $x_0=2041$ كنواة. بتربيع هذا العدد نحصل على العدد 4165681. بما أن عدد أرقام العدد الناتج أقل من 8 أرقام (2×4) ، نكمل الأرقام إلى ثمانية و ذلك بإضافة أصفار إلى يسار العدد ليصبح 04165681. نقوم بإقتطاع الأرقام الأربعة الواقعة منتصف العدد لنحصل على العدد العشوائي 1656. بتكرار هذه العملية نحصل على الأعداد العشوائية المبينة في الجدول التالي.

k	0	1	2	3	4	5	6	7	8
x_k	2041	1656	7423	1009	0180	0324	1049	1004	80

على الرغم من بساطة هذه الطريقة إلا أنها تعاني من مشكلة كبيرة. تتمثل هذه المشكلة بنزعة متتالية الأعداد العشوائية المولدة للإقتراب بسرعة من الصفر. نتيجة لإقتراب المتتالية من الصفر، فإننا لانستطيع الحصول على أعداد عشوائية جديدة و بالتالي فإن دور هذه الخوارزمية غير طويل بشكل عام. على سبيل المثال، إذا استمرينا بتوليد المزيد من الأعداد العشوائية باستخدام المثال أعلاه، فإننا نتوقف عند العدد العشوائي الرابع عشر لأنه يصبح صفراً. (تحقق من ذلك بنفسك). إضافةً لذلك، فإن دور المتتالية يعتمد بشكل كبير على الاختيار الجيد للنواة (Seed). بعض الأعداد التي تُستخدم كنواة تجعل دور المتتالية المولدة يساوي 1. على سبيل المثال، إذا بدأنا بالرقم

50 كنواة فإننا نحصل على متتالية دورها 1 فقط و ذلك لأن مربع العدد هو 2500 و بالتالي فالعدد العشوائي الجديد هو 50 (العدد السابق نفسه).

يمكن تلخيص مميزات طريقة نايمان (منتصف التربيعات) بالنقاط التالية:

1. سهولة التوليد لأنها تعتمد على عملية تربيع عدد ما (النواة) و من ثم اقتطاع الأرقام الواقعة في منتصف العدد الناتج.
2. دور هذه المتتالية يعتمد على الاختيار الجيد للنواة. بعض الاختيارات كما رأينا تولد متتاليات قصيرة الدور.
3. الأرقام الناتجة ليست جيدة العشوائية و بالتالي ليست مناسبة لكل التطبيقات العملية.

المولدات العشوائية الخطية Linear Congruential Generators

المولدات العشوائية الخطية من أكثر المولدات شيوعاً حيث تشكل تقريباً الأساس لجميع المولدات العشوائية المستخدمة حالياً. تعطي هذه المولدات أعداداً عشوائية منتظمة التوزيع على المجال الكلي. تستخدم المولدات الخطية العلاقة التالية التي طورها Lehmer في عام 1948.

$$X_{n+1} = (aX_n + c) \bmod m \quad (1)$$

حيث: تمثل m طول الدور ، a عبارة عن مضروب ما يجب أن يحدد ، c مقدار الزيادة، و X_0 تمثل قيمة بدائية أو Seed. يجب أن تحقق هذه المقادير الشروط التالية:

1. $m > 0$
2. $0 \leq a \leq m$
3. $0 \leq c \leq m$
4. $0 \leq X_0 \leq m$

في حالة $c=0$ فإن المولد العشوائي يدعى مولداً ضربياً أما إذا كانت c أكبر من الصفر فإن المولد يدعى مولداً مختلطاً (جمعي و ضربى).

تبيين العلاقة أعلاه أن أي عدد عشوائي لاحق يعتمد فقط على العدد العشوائي السابق له مباشرة. هذه خاصية مهمة للمولدات العشوائية الخطية فهي تقلل متطلبات التخزين (أي توليدها يحتاج إلى ذاكرة حاسوبية أقل). إضافة لذلك ، العلاقة (1) تضع قيداً على دور المولدات الخطية. توضيحاً لذلك فإن أطول دور ممكن هو m . بمعنى آخر ، نحصل فقط على متتالية من " $m - 1$ " عدد عشوائي و كل المتتاليات الأخرى ستكون تكراراً لهذه المتتالية.

الأعداد المولدة بوساطة العلاقة (1) أعداد صحيحة. للحصول على أعداد حقيقية لابد من إجراء تحويل رياضي بسيط. يتمثل هذا التحويل بقسمة العدد الصحيح على طول الدور أو على طول الدور ناقص واحد كما موضح أدناه.

عدد حقيقي منتظم التوزيع في المجال $[0, 1]$	$R_n = X_n / m$
عدد حقيقي منتظم التوزيع في المجال $[0, 1]$	$R_n = X_n / (m-1)$

تعتمد جودة المولد بشكل كبير على القيم الأربعة (a, c, m, X_0) الواردة في الصيغة (1) وتتأثر بشكل كبيرها. أي اختيار سيئ لهذه القيم قد يؤثر على طول دور المولد و قد يؤثر أيضاً على جودة الأعداد العشوائية المولدة. على سبيل المثال، قيمة المضروب a تؤثر على كل من طول الدور و عشوائية متتالية الأعداد المولدة. الجدول التالي يوضح تأثير قيمة a في طول الدور و جودة الأعداد العشوائية المولدة.

قيمة a	المتتالية المولدة	طول الدور Period
0	0, 0, ...	1
1	1, 1, ...	1
2	1, 2, 4, 8, 3, 6, 12, 11, 9, 5, 10, 7, 1, ...	12
3	1, 3, 9, 1, ...	3

6	1, 4, 3, 12, 9, 10, 1, ...	4
4	1, 5, 12, 8, 1, ...	5
12	1, 6, 10, 8, 9, 2, 12, 7, 3, 5, 4, 11, 1,	6
12	1, 7, 10, 5, 9, 11, 12, 6, 3, 8, 4, 2, 1, ...	7
4	1, 8, 12, 5, 1, ...	8
3	1, 9, 3, 1, ...	9
6	1, 10, 9, 12, 3, 4, 1, ...	10
12	1, 11, 4, 5, 3, 7, 12, 2, 9, 8, 10, 6, 1, ...	11
2	1, 12, 1, ...	12

لاحظ تأثير قيمة a على طول الدور. بعض قيم a تجعل طول الدور واحداً فقط و بعضها الآخر تجعل طول الدور 12.

تمرين: تذكر أن المولد الضربي هو مولد يستخدم العلاقة (1) حيث $C=0$ أما المولد المختلط فهو يستخدم كامل العلاقة (1) أي C مختلفة عن الصفر. السؤال أي المولدين (الضربي أم المختلط) أسرع من الناحية الحاسوبية؟ علل إجابتك.

متوسط الأعداد العشوائية و تباينها

Average and Variance of Random Numbers

المتوسط المثالي لمجموعة من الأعداد العشوائية المنتظمة التوزيع في المجال نصف المفتوح $[0, 1]$ هو $1/2$ و التباين المثالي هو $1/12$. هاتان القيمتان للمتوسط و الإنحراف المعياري هما شرطان ضروريان لجودة المولد العشوائي و لكنهما غير كافيتين لجودة المولد العشوائي. هذا يعني أنه في حال عدم تحقق هذين الشرطين على المتوسط و التباين، فإن المولد ليس جيداً. تحقق هذين الشرطين لا يؤدي بالضرورة إلى مولد عشوائي جيد. المثال التالي يوضح ذلك. إذا فرضنا أن قيمة $a=1, c=1, m=4096$, $X_0=1$ في المعادلة (1) أعلاه، فإن متوسط أي متتالية أعداد عشوائية مولدة من هذا المولد و تباينها ستكون على الترتيب 0.4999 و 0.0833 و هما قيمتان قريبتان من

القيمتين المثاليتين $1/2$ و $1/12$. و لكن المثالية العشوائية المولدة باستخدام هذه القيم ليست جيدة. هذه المتتالية: $0, 1, 2, 3, \dots, 4093, 4094, 4095, 0, \dots$

تطبيق برمجي

سوف نكتب برنامجاً بلغة الجافا يقوم بتوليد أعداد عشوائية باستخدام العلاقة (1) ثم يحسب المتوسط و الإنحراف المعياري للمتتالية الناتجة. تذكر أن المتوسط و الإنحراف المعياري يعطيان بالعلاقات التالية:

$$\bar{X} = \frac{1}{n}(x_1 + x_2 + \dots + x_n) = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

البرنامج بلغة الجافا JAVA

```
public static void GRandom(int a, int c,
    long m, long x)
{
    long A[] = new long[500];

    int p;
    double mean, sum, var;
    for (int i=1; i<=m+5; i++){
        x=(a*x+c)% m;  // إقتراع العدد العشوائي
        A[i] =x;
    }
    int i =1;
    int j =2;
    while (A[i] != A[j]){
        if (j > m){
            i++;
            j=i;
        }
        // حساب طول الدور
    }
}
```

```

        j++;
    }
    p = j - i;
    sum = 0;
    for (int k = i; k <= i + p - 1; k++)
        sum += A[k] / (m + 0.0);
    mean = sum / p; // حساب المتوسط
    var = 0.0;
    for (int k = i; k <= i + p - 1; k++) // حساب التباين
        var += (A[k] / (m + 0.0) -
                mean) * (A[k] / (m + 0.0) - mean);
    var = var / p;
    System.out.println("Period: " + p + " mean =
                        "+ mean + "Variance =" + var);
}

public static void main(String[] args)
{
    GRandom(6, 3, 13, 1);
}
}

```

الخرج Output

العدد العشوائي الصحيح	العدد العشوائي المنتظم التوزيع في المجال [0, 1]	نحصل على العدد العشوائي المنتظم التوزيع في المجال [0, 1] بقسمة العدد الصحيح على الطول الأعظمي للدور (m). في مثالنا m=13.
9	0.6923076923076923	
5	0.38461538461538464	
7	0.5384615384615384	
6	0.46153846153846156	
0	0.0	
3	0.23076923076923078	

8	0.6153846153846154	
12	0.9230769230769231	
10	0.7692307692307693	
11	0.8461538461538461	
4	0.3076923076923077	لاحظ أن المتتالية بدأت تكرر نفسها بعد هذا العدد..... مما يدل على أن دور المتتالية هو 12
1	0.07692307692307693	
9	0.6923076923076923	
5	0.38461538461538464	
7	0.5384615384615384	
6	0.46153846153846156	
0	0.0	
3	0.23076923076923078	
طول الدور: 12 المتوسط: 0.487179 التباين: 0.08119		

يبين الخرج مجموعة من الأعداد العشوائية التي اقترعت باستخدام العلاقة (1) و ذلك من أجل قيم $a=6, c=3, m=13, X_0=1$. يمثل العمود الأول من اليسار القيم العشوائية الصحيحة. بينما يمثل العمود الثاني من اليسار القيم العشوائية المقابلة لها في المجال $[0, 1]$ التي تم الحصول عليها كما أشرنا سابقاً بنحويلات رياضية (قسمة كل عدد على طول الدور +1). إن دور المتتالية السابقة هو 12 لأن الأرقام العشوائية المولدة بعد العدد العشوائي الثاني عشر (مشار إليه بالسهم «») بدأت تكرر نفسها. يبين الجدول أعلاه أيضاً متوسط المتتالية المولدة وهو قريب من المتوسط المثالي $1/2$ و كذلك التباين و هو قريب من الانحراف المعياري المثالي $1/12$.

Advantages and Disadvantages of Linear Generators

سوف نعرض في هذه الفقرة بعض مساوئ و إيجابيات المولدات الخطية. هدفنا هنا تسليط الضوء على بعض المحاسن و المساوئ و ليس مناقشتها جميعاً. المراجع بنهاية الفصل توفر معلومات أكثر لمن يرغب الإستزادة حول الموضوع.

إيجابيات المولدات الخطية

- المولدات العشوائية الخطية سهلة البرمجة و سريعة لأن العمليات الحسابية المطلوبة لاقتراح أعداد عشوائية بسيطة.
- المولدات العشوائية الخطية تحتاج إلى تخزين أقل. لاقتراح عدد عشوائي فإننا نحتاج فقط لتخزين العدد السابق له مباشرة. هاتان الإيجابيتان مهمتان للغاية لأن أي كلفة عالية من الناحية الزمنية أو التخزينية تجعل استخدام الطريقة غير مقبول حاسوبياً.

مساوئ المولدات الخطية

- جودة الأعداد العشوائية المولدة بهذه الطريقة ليست عالية لأن هناك ارتباطاً بين كل عددين عشوائيين متتاليين. لذلك ينصح بعدم استخدامها في التطبيقات التي تحتاج إلى أعداد عشوائية بجودة عالية. مثالاً على ذلك، لا ينصح باستخدام الأعداد العشوائية المولدة بهذه الطريقة مع طرق مونتي كارلو (الفصل القادم). إضافة إلى ذلك، يجب عدم استخدام الأعداد العشوائية المولدة بهذه الطريقة في تطبيقات أمن المعلومات أو التشفير Cryptographic Applications. حيث يمكن تفكيك الشيفرة المشفرة باستخدام المولدات الخطية بسهولة.
- يعتمد طول دور المتتالية المولدة بطريقة المولدات الخطية على الاختيار الجيد لوسطاء العلاقة (1). من المعلوم أن دور المتتالية العشوائية المولدة باستخدام

- العلاقة (1) هو على الأكثر m . لكن أي اختيارات غير موفقة للوسيط a تجعل طول الدور أقصر من m . في الواقع هناك شروط لازمة و كافية لضمان توليد متتالية أعداد عشوائية بدور كامل (مساوياً m). هذه الشروط:
- الوسطاء a و m أوليان فيما بينهما.
 - المقدار $a-1$ قابل للقسمة على كل القواسم الأولية للوسيط m .
 - المقدار $a-1$ هو من مضاعفات 4 إذا كانت m من مضاعفات 4.
 - تعتمد جودة الأعداد العشوائية على اختيارات الوسطاء. أي أن هناك إختيارات للوسطاء يمكن أن ينتج عنها أعداد عشوائية ذات جودة غير مقبولة.

4.1 نمذجة المتغيرات العشوائية حاسوبياً

لاقتراح قيم عشوائية لمتغير ما X حاسوبياً فإننا ننتقل من دالة توليد الأرقام العشوائية المنتظمة التوزيع في المجال $[0, 1[$ والموجودة في كل الحواسيب تقريباً. تعد كل قيمة منتظمة على المجال $[0, 1[$ قيمة قياسية يتم بوسطاتها الحصول على قيمة لأي متغير عشوائي حاسوبياً. فإذا كان X متغيراً عشوائياً منتظماً على المجال $[0, 1[$ ، يكون لـ X الكثافة الاحتمالية التالية:

$$f(x) = \begin{cases} 1 & ; 0 < x < 1 \\ 0 & ; \text{خلاف ذلك} \end{cases}$$

$$F_x(x) = \begin{cases} 0 & ; x \leq 0 \\ x & ; 0 < x < 1 \\ 1 & ; x \geq 1 \end{cases}$$

و التوقع الرياضي:

$$\mu_x = E(x) = \frac{1}{2} \quad , \quad V(x) = E(x^2) - \mu_x^2 = \frac{1}{12}$$

إذا رمزنا للقيمة العشوائية للمتغير العشوائي X المنتظم التوزيع على المجال المفتوح $[0, 1[$ والمولدة (المُقتَرعة) حاسوبياً بـ RND، عندئذ فإن المتتالية التالية المولدة $x_1 =$

التوزيع على المجال $[0, 1]$. $x_2=RND, \dots, x_n = RND$ تمثل n قيمة عشوائية للمتغير X المنتظم

إن قيم هذه المتتالية ليست مستقلة عشوائياً من الناحية النظرية و ذلك لأسباب تعود لطبيعة الخوارزميات التي تنتجها، وهذا ما دعى لتسميتها أعداداً شبه عشوائية، ما دامت تحقق اختبارات إحصائية تدعم اعتبارنا لهذه القيم بأنها مستقلة عشوائياً، ومن مجتمع منتظم على المجال $[0, 1]$.

نمذجة المتغيرات العشوائية المستمرة

ليكن Y متغيراً عشوائياً مستمراً كثافته الاحتمالية $f_Y(y)$ وتابع توزيعه $F_Y(y)$ ، عندئذٍ إذا فرضنا أن $X = F_Y(y)$ ، يكون لـ X التوزيع المنتظم على المجال $[0, 1]$ ، أي أن لـ X تابع التوزيع:

$$F_X(x) = \begin{cases} 0 & ; \quad x \leq 0 \\ x & ; \quad 0 < x < 1 \\ 1 & ; \quad x \geq 1 \end{cases}$$

و تابع الكثافة:

$$f_X(x) = \begin{cases} 1 & ; \quad 0 < x < 1 \\ 0 & ; \quad \text{خلاف ذلك} \end{cases}$$

وهنا نميز حالين:

الحالة الأولى: يمكن إيجاد الدالة العكسية لدالة التوزيع $x = F_Y(y)$ تحليلياً.

في هذه الحالة تكون القيمة $F_Y^{-1}(RND)$ قيمة عشوائية للمتغير العشوائي Y . نقدم فيما يلي علاقات الاقتراع لبعض المتغيرات العشوائية.

مثال (1): متغير منتظم التوزيع على مجال معين

إذا كان المتغير X منتظم التوزيع على المجال $[a, b]$ ، فإن الكثافة:

$$f(x) = \begin{cases} \frac{1}{b-a} & ; \quad a < x < b \\ 0 & \end{cases}$$

وتابع التوزيع :

$$F(x) = \begin{cases} 0 & ; x \leq a \\ \frac{x-a}{b-a} & ; a < x < b \\ 1 & ; x \geq b \end{cases}$$

يمكن إنشاء علاقة اقتراع للمتغير X من خلال المساواة التالية:

$$RND = F(x) = \frac{x-a}{b-a}$$

أي أن:

$$x = a + (b - a) * RND$$

هي عبارة عن قيمة عشوائية للمتغير العشوائي X .

على سبيل المثال، إذا أردنا اقتراع عشر قيم للمتغير العشوائي X ، فإننا نقوم بتوليد عشر

قيم عشوائية RND و في كل مرة نستخدم علاقة الاقتراع $x = a + (b - a) * RND$

لتوليد قيمة عشوائية لـ X .

مثال (2): متغير أسّي

إذا كان المتغير العشوائي X أسّي التوزيع بوسيط λ و معرف بالكثافة التالية:

$$f(x) = \begin{cases} 0 & ; x < 0 \\ \lambda e^{-\lambda x} & ; x \geq 0 \end{cases}$$

و تابع التوزيع:

$$F(x) = 1 - e^{-\lambda x} ; x \geq 0 ; \lambda > 0$$

يمكن إنشاء علاقة اقتراع للمتغير العشوائي X الأسّي التوزيع من خلال المساواة

التالية:

$$RND = 1 - e^{-\lambda x}$$

التي تنتج بإصلاح بسيط للعلاقة التالية:

$$x = \frac{-1}{\lambda} \ln(1 - RND)$$

التي تكافئ العلاقة التالية:

$$x = \frac{-1}{\lambda} \ln(RND)$$

و هذا ينتج من حقيقة أنه إذا كان المتغير العشوائي Y المنتظم التوزيع على المجال $[0, 1[$ ، فإنه يكون للمتغير العشوائي $Z = 1 - Y$ التوزيع المنتظم على المجال $[0, 1[$ (يترك برهان ذلك للطالب).

يمكن تطبيق الطريقة السابقة (طريقة الدالة العكسية) حتى لو كانت دالة التوزيع $F(y)$ تحتوي على نقاط انقطاع. المثال التالي يوضح ذلك.

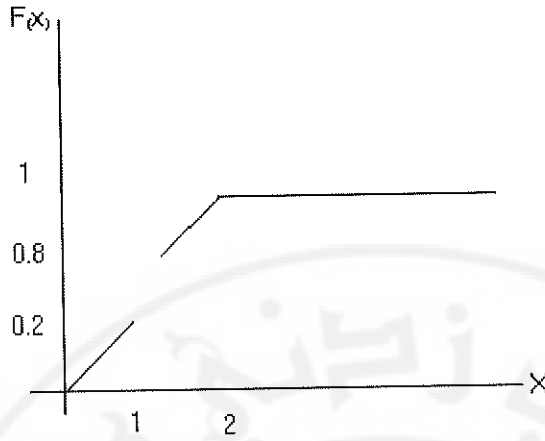
مثال (3)

ليكن المتغير X متغيراً عشوائياً معرّفاً بتابع التوزيع التالي:

$$F(x) = \begin{cases} 0.2x & ; 0 < x < 1 \\ 0.2x + 0.6 & ; 1 \leq x \leq 2 \\ 1 & ; x > 2 \end{cases}$$

المطلوب:

1. أرسم دالة التوزيع $y = F(x)$.
 2. أوجد الدالة العكسية لدالة التوزيع $F(x)$.
 3. أوجد علاقة اقتراع لقيم X .
- يوضح الشكل التالي مخططاً بيانياً لدالة التوزيع.



الدالة العكسية لتابع التوزيع $F(x)$ هي:

$$x = \begin{cases} 5y & ; 0 < y < 0.2 \\ 1 & ; 0.2 \leq y \leq 0.8 \\ 5y - 3 & ; 0.8 < y < 1 \end{cases}$$

و بالتالي تكون علاقة الاقتراع بالشكل التالي:

$$x = \begin{cases} 5RND & ; 0 < RND < 0.2 \\ 1 & ; 0.2 \leq RND \leq 0.8 \\ 5 * RND - 3 & ; 0.8 < RND < 1 \end{cases}$$

الحالة الثانية: لا يمكن إيجاد الدالة العكسية لتابع التوزيع تحليلياً.

في هذه الحالة يمكن تطبيق بعض الطرائق مستفيدين من بعض المبرهنات في نظرية الاحتمالات التي نقبلها بدون برهان.

طريقة نايمان لاقتراع قيم عشوائية لمتغير مستمر

ليكن X متغيراً عشوائياً مستمراً على المجال $[a, b[$ و كثافته $f(x)$ محدودة، أي

يوجد عدد حقيقي موجب m بحيث يكون:

$f(x) \leq m$ من أجل جميع قيم x في المجال $[a, b]$. ضمن هذه الشروط فإننا يمكن أن نلخص طريقة نايمان لتوليد أعداد عشوائية بالخطوات التالية:

1. توليد عددين عشوائيين $x_1 = \text{RND}$ و $x_2 = \text{RND}$.

2. ننشئ النقطة A التي إحداثياتها:

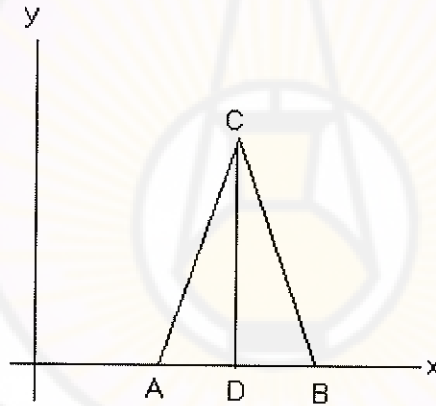
$$x = a + (b-a) * x_1 \text{ و } y = m * x_2$$

3. إذا وقعت النقطة A تحت منحنى الكثافة $f(x)$ نقبل القيمة x_1 كقيمة مقترعة

للمتغير X و إلا نرفضها و في هذا الحالة نعود إلى الخطوة 1.

مثال (4)

بفرض أن X متغير عشوائي مستمر على المجال $[a, b]$ و له كثافة احتمالية مثلثية مبينة بالشكل التالي:



علماً أن إحداثيات رؤوس المثلث A, B, C معرفة بالشكل:

$$A(a, 0), B(b, 0), C\left(\frac{a+b}{2}, \frac{2}{b-a}\right), D\left(\frac{a+b}{2}, 0\right)$$

المطلوب اكتب خوارزمية اقتراح n قيمة للمتغير العشوائي X بطريقة نايمان.

الحل:

1. نوجد معادلة المستقيم AC التي يمكن أن تصاغ بالشكل التالي:

$$y = 4(x - a)/(b - a)^2$$

2. نوجد معادلة المستقيم BC التي يمكن أن تصاغ بالشكل التالي:

$$y = 4(b - x)/(b - a)^2$$

3. أنشئ النقطة $M(\alpha, \beta)$ ، حيث النقاط α و β :

$$\alpha = a + (b - a) * \text{RND}$$

$$\beta = 2 * \text{RND} / (b - a)$$

4. نختبر وقوع النقطة $M(\alpha, \beta)$ تحت منحنى كثافة X (أي داخل المثلث (ABC) .

إذا تحقق الشرطان:

$$\beta < 4(x - a)/(b - a)^2$$

$$\beta < 4(b - x)/(b - a)$$

نقبل α باعتبارها قيمة عشوائية للمتغير العشوائي X ، وإلا نعود إلى الخطوة (3).

5. نكرر الخطوتين 3 و 4 حتى نحصل على n قيمة.

اقتراع قيم لمتغير عشوائي طبيعي

هناك العديد من الطرق لتوليد قيم عشوائية لمتغير طبيعي، سوف نعرض بعضها في هذه الفقرة.

الطريقة الأولى: استخدام مبرهنة النهاية المركزية

ليكن X متغيراً عشوائياً له التوزيع الطبيعي بالمتوسط μ و التباين σ^2 أي أن $X \sim (\mu, \sigma^2)$ ، و نريد اقتراع قيم لهذا المتغير.

من أجل ذلك نقوم بالخطوات التالية:

1. نقترح n قيمة عشوائية منتظمة التوزيع على المجال $[0, 1]$ و لتكن هذه القيم هي: $x_1=RND, x_2=RND, \dots, x_n = RND$.

2. لنفترض أن $Y = x_1 + x_2 + \dots + x_n = \sum_{i=1}^n x_i$ عندئذ فإن توزيع المتغير العشوائي Y يتقارب من التوزيع الطبيعي $N(\frac{n}{2}, \frac{n}{12})$ و بالتالي فإن المتغير

$$Z = \frac{(Y - \frac{n}{2})}{\sqrt{\frac{n}{12}}}$$

يتقارب من التوزيع الطبيعي المعياري $N(0, 1)$.

3. بأخذ التركيب التالي $X = \sigma Z + \mu$ نجد أن X يتقارب من التوزيع الطبيعي $N(\mu, \sigma^2)$.

ملاحظة هامة

يكفي أن نأخذ $n = 12$ ليتقارب المتغير $Y = \sum_{i=1}^{12} x_i$ من التوزيع الطبيعي $N(6, 1)$

1) وبالتالي يتقارب المتغير العشوائي Z من $N(0, 1)$. اعتمادا على ذلك تكون خوارزمية اقتراح قيمة لمتغير عشوائي طبيعي X متوسطه μ و تباينه σ^2 على الشكل التالي:

1. نقترح 12 قيمة منتظمة التوزيع في المجال $[0, 1]$ ولتكن هذه القيم هي: $x_1, x_2, \dots, x_n$.

2. نشكل المقدار $Z = \sum_{i=1}^{12} x_i - 6$.

3. نحسب المقدار $\gamma = \sigma Z + \mu$ فتكون هذه القيمة هي قيمة عشوائية مقترعة للمتغير العشوائي X الذي له التوزيع الطبيعي $N(\mu, \sigma^2)$.

الطريقة الثانية

نعلم أنه إذا كان $X \sim (\mu, \sigma^2)$ فإن كثافة X تعطى بالعلاقة التالية:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}; -\infty < x < +\infty, -\infty < \mu < +\infty, \sigma > 0$$

يبرهن على أن $P(\mu - 3\sigma < X < \mu + 3\sigma) = 0.99$ و بالتالي يمكننا في حالة $X \sim (\mu, \sigma^2)$ أن نعتبر X معرفاً و مستمراً على المجال $[a, b[$ حيث $a = \mu - 3\sigma$ و $b = \mu + 3\sigma$. بناءً على ذلك، يمكن تطبيق طريقة نايمان لاقتراح قيمة للمتغير العشوائي الطبيعي X .

الطريقة الثالثة

إذا كان للمتغير العشوائي X التوزيع الطبيعي المعياري، فإنه يمكن تعريف علاقة اقتراح لـ X بالعلاقة التالية:

$$x = \sqrt{-2\ln(RND)} \cos(2\pi RND)$$

اقتراح قيم عشوائية للمتغير X الذي كثافته

$$f(x) = \begin{cases} nx^{n-1} & ; 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

الطريقة الأولى: باستخدام الدالة العكسية:

من الواضح أن :

$$F(x) = \begin{cases} 0 & ; x \leq 0 \\ x^n & ; 0 < x < 1 \\ 1 & ; x \geq 1 \end{cases}$$

يمكن الحصول على علاقة الاقتراح من تابع التوزيع بالشكل التالي:

$$RND = F(x) \Rightarrow RND = x^n \Rightarrow x = (RND)^{\frac{1}{n}}$$

أي أن، علاقة الاقتراع للمتغير العشوائي X هي:

$$x = \sqrt[n]{RND}$$

الطريقة الثانية: يمكن اقتراع قيم X طبقاً للعلاقة:

$$x = \max(\gamma_1, \gamma_2, \dots, \gamma_n)$$

حيث γ_i من أجل $i=1, 2, \dots, n$ هي متغيرات عشوائية منتظمة التوزيع على المجال $[0, 1]$.

5.1 نمذجة المتغيرات العشوائية المنفصلة

حاول الباحثون بناء دولاب خاص بكل متغير عشوائي منفصل على حدة لاقتراع قيم لذلك المتغير. على سبيل المثال، للمتغير العشوائي X المُمثل بجدول التوزيع الاحتمالي التالي:

X	1	2	3	4
$P(X)$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{1}{8}$

يمكننا بناء دولاب بحيث تعكس تقسيماته احتمال ظهور كل قيمة من القيم الظاهرة في الجدول أعلاه. بمعرفة أن الدولاب هو عبارة عن دائرة وتساوي 360 درجة، فإنه و بحسابات رياضية بسيطة نجد أن حصة الأرقام 1، 2، 3، 4 من الدولاب هي على الترتيب: 45، 180، 45، 90 درجة.

يمكن تعميم هذه الفكرة بحيث نتحدث عن متغير عشوائي منفصل له n قيمة و احتمال p_i ($i=1, 2, \dots, n$).

على الرغم من بساطة هذه الفكرة إلا أنها تقتصر إلى إمكانية التطبيق العملي.

لذلك تم الاستغناء عنها و الاعتماد على المولد العشوائي القياسي RND المنتظم التوزيع.

على المجال $[0, 1]$ لتوليد قيم عشوائية لمتغيرات منفصلة. يمكن توليد قيم عشوائية لأي متغير منفصل كما يلي:

نفرض أن لدينا المتغير العشوائي المنفصل X والمعرف بالجدول التالي:

X	x_1	x_2	...	x_n
$f(x)$	p_1	p_2		p_n

لاقتراح قيم لهذا المتغير، فإننا نقوم بتقسيم المجال $[0, 1]$ إلى n قسم بحيث تكون أطوال هذه الأقسام هي $p_1, p_2, \dots, p_n$. لنفترض أن نقاط التقسيم هي $s_1, s_2, \dots, s_n$ فتكون فواصل نقاط التقسيم هي على الترتيب $p_1, p_1 + p_2, \dots, 1$. يمكن بسهولة ملاحظة أن المجالات الجزئية المحددة بنقاط التقسيم تعطى بالمجال التالي:

$$\Delta_i = [\sum_{j=1}^{i-1} p_j, \sum_{j=1}^i p_j]$$

و بالتالي يمكن وضع معادلة الاقتراح التالي لقيم المتغير المنفصل X كما يلي

$$RND \in \Delta_i \Leftrightarrow X = x_i, i = 1, 2, \dots, n \quad (*)$$

أي إن القيمة المقترحة لـ X هي x_i إذا كانت قيمة RND تنتمي إلى المجال

$$\Delta_i = [\sum_{j=1}^{i-1} p_j, \sum_{j=1}^i p_j]$$

يمكن تلخيص ذلك خوارزمية بالخطوات التالية:

```

input n
i = 1
while i < n do
    input  $x_i, p_i$ 
end while

```

```

 $q_0 = 0, i = 1$ 
while  $i \leq n$  do

```

$q_i = q_i + p_i$
end while

$g = \text{rnd}$
 $i = 1$
while $g > q_i$ do
 $i = i + 1$
end while

print x_i
end.

يمكن استخدام الحالة العامة السابقة لإنشاء صيغة لاقتراح قيم عشوائية لمتغير X معرف الجدول التالي:

X	x_1	x_2	...	x_n
$f(X)$	$\frac{1}{n}$	$\frac{1}{n}$	...	$\frac{1}{n}$

يمكننا باستخدام العلاقة السابقة (*) أن نوجد علاقة الاقتراح لهذا المتغير

بالشكل:

$$x = x_i \Leftrightarrow i = 1 + \text{int}(n * RND)$$

حيث $\text{int}(a)$ تعني الجزء الصحيح من a .

و ذلك لأن:

$$x = x_i \Leftrightarrow \frac{i-1}{n} \leq RND \leq \frac{i}{n}$$

يمكن كتابة خوارزمية اقتراح للمتغير العشوائي المعرف في الجدول أعلاه كما

يلي:

Input n
 $I=1$
While $i \leq n$ do
 Input x_i

End While

$i = 1 + \text{int}(n * \text{RND})$

Print x_i

End

6.1 تمارين Exercises

1. لماذا يجب أن نهتم بتوليد الأعداد العشوائية؟
2. ما مواصفات خوارزمية توليد الأعداد العشوائية؟
3. لماذا نهتم بكون المولد العشوائي فعالاً؟
4. ما الذي نعنيه بدور المولد العشوائي؟ هل للدور أهمية في التطبيقات العملية؟
5. لماذا تحتاج المولدات العشوائية الخطية تخزيناً أقل؟
6. لماذا الأعداد العشوائية المولدة بواسطة المولدات الخطية ليست ذات جودة عالية؟
7. يقال إن الأعداد العشوائية الناتجة عن المولدات الخطية أنها مرتبطة Correlated؟ هل تتفق مع هذه المقولة؟ أعط تفسيراً كافياً لجوابك.
8. إذا كان لدينا مولد عشوائي يولد أعداداً عشوائية في المجال $[0, 1]$ و كان متوسط المتتاليات المولدة 0.39 وانحرافها المعياري 0.12. ناقش جودة هذا المولد.
9. ما الذي نعنيه بالعبرة: "متتاليات الأعداد العشوائية يجب أن تكون مستقلة"؟ ما أهمية استقلالية هذه المتتاليات في التطبيقات العملية؟ أعط تفسيراً مقنعاً لجوابك.
10. لماذا تكون المولدات الخطية سريعة؟ لماذا السرعة مهمة؟
11. استخدم طريقة منتصف التربيعة لاقتراح
a. 10 أعداد عشوائية باستخدام القيمة البدائية $x_0=1009$
b. 20 عدد عشوائي باستخدام القيمة البدائية $x_0=653217$

- c. 15 عدد عشوائي باستخدام القيمة البدائية $x_0 = 3043$
- d. ماذا تلاحظ من المتتاليات العشوائية السابقة؟ هل تتقارب كل متتالية ناتجة من الصفر بسرعة؟

12. استخدم طريقة المولدات الخطية (العلاقة (1)) لإقتراح

- a. 10 أعداد عشوائية باستخدام $a = 5, b = 1, c = 8$
- b. 15 عدد عشوائي باستخدام $a = 1, b = 7, c = 10$
- c. 20 عدد عشوائي باستخدام $a = 5, b = 3, c = 16$
- d. ماذا تلاحظ من النتائج السابقة؟ هل هناك حلقات؟ إذا وجدت حلقات ، متى تحدث هذه الحلقات؟

13. ليكن لدينا المتحول العشوائي X المُعرف بالكثافة الاحتمالية التالية:

X	x_1	x_2	x_3	x_4	x_5	x_6
$P(X)$	$\frac{1}{16}$	$\frac{8}{16}$	$\frac{2}{16}$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{2}{16}$

- a. أوجد علاقة لاقتراح قيم عشوائية لهذه المتحول.
- b. اكتب خوارزمية لاقتراح قيم عشوائية مستخدماً علاقة الاقتراح التي أوجدتها في (a).

14. ليكن لدينا المتحول العشوائي X المُعرف من خلال الكثافة التالية:

$$f(X) = \begin{cases} \frac{1}{6} & \text{if } 2 \leq X \leq 8 \\ 0 & \text{Otherwise} \end{cases}$$

- a. أوجد علاقة اقتراح قيم عشوائية للمتحول X .
- b. اكتب خوارزمية و برنامجاً بأي لغة نشاء لاقتراح قيم عشوائية لهذا المتحول.

15. أوجد علاقة اقتراح قيم عشوائية لمتحول طبيعي $N(6, 1.3)$ بطريقتين مختلفتين.

16. أوجد علاقة اقتراح لمتحول عشوائي برنولي. اكتب خوارزمية لتوليد n عدداً عشوائياً لهذا المتحول.

★ ★ ★ ★ ★ ★ ★ ★ ★ ★ ★ ★ ★ ★ ★ ★



المحاكاة باستخدام الخوارزميات العشوائية Simulations using Randomized Algorithms

1.2 مقدمة Introduction

هناك العديد من التطبيقات التي أصبحت ممكنة بفضل وجود الحواسيب السريعة. لقد أصبح ممكناً بناء نماذج تحتاج إلى قدرة حسابية عالية تحاكي الواقع و تعطي مؤشرات تساعد في التنبؤ بالنتائج الممكن وقوعها في المستقبل. إن المجالات الهندسية، الدراسات الإحصائية، التطبيقات الإدارية، التطبيقات المالية و الإستثمارية ما هي إلا بعض التطبيقات الممكن بناء نماذج محاكاة لها و المساعدة في التنبؤ بحالتها المستقبلية.

أحد أهم إنجازات المحاكاة للنظم الواقعية هي القدرة على التنبؤ بالحالات المستقبلية التي يمكن أن تمر بها هذه النظم و ذلك بوجود الغموض الذي يحيط بالمتغيرات التي تؤثر في هذه النظم خاصة في متغيرات الدخل. يتم ذلك من خلال إجراء مئات الآلاف من التجارب على نماذج المحاكاة التي تحاكي السيناريوهات المختلفة التي يمكن أن يمر بها النظام نتيجة التغيرات في البيئة المحيطة. من أهم نتائج محاكاة النظام الواقعي هي المساعدة في إتخاذ قرارات أقرب ماتكون للصواب إزاء النظام.

سوف نقوم في هذا الفصل بتعريف العديد من المفاهيم الهامة في النمذجة العشوائية. كذلك سوف نطرح العديد من الأمثلة المناسبة التي تساعد في تنمية إدراك

القارئ بأهمية هذا الفرع من النمذجة. يجب أن ننوه أن الهدف ليس تغطية شاملة لهذا الموضوع لأنه بحاجة إلى أكثر من مقرر واحد لتغطيته. لذلك الهدف هنا هو تعريف القارئ بهذا الموضوع و من يريد الإستزادة يمكنه الإطلاع على المراجع في نهاية الفصل.

2.2 الخوارزميات العشوائية Randomized Algorithms

الخوارزمية العشوائية هي خوارزمية تستخدم ، بالإضافة لدخلها، مجموعة من الأعداد العشوائية. تستخدم الخوارزمية الأعداد العشوائية للقيام بإجراء إختيارات عشوائية. هذا يعني أن أداء الخوارزمية على الدخل نفسه يمكن أن يختلف من تنفيذ لآخر و ذلك إعتتماداً على الأعداد العشوائية المستخدمة في التنفيذ. على سبيل المثال ، عند دخل معين، فإن الزمن اللازم لتنفيذ الخوارزمية على هذا الدخل هو متغير عشوائي لايمكن تحديد قيمه مسبقاً.

ظهرت الخوارزميات العشوائية منذ نشوء الحاسبات و لكن لم يتم الإنتباه إلى أهميتها إلا منذ سبعينات القرن الماضي. من الخوارزميات المشهورة جداً نذكر خوارزمية إختبار فيما إذا كان عدداً ما أولياً أم لا.

من المؤكد أن العشوائية تلعب دوراً مهماً في العديد من التطبيقات العملية و تقدم أدوات غاية في الأهمية لإنشاء الخوارزميات. هناك فائدتان هامتان جداً للخوارزميات العشوائية.

1. غالباً ما يكون التعقيد الزمني و احتياجات التخزين في الخوارزميات العشوائية أقل بكثير من أفضل الخوارزميات غير العشوائية التي تحل نفس المسألة. تعتبر هذه الخاصة غاية في الأهمية لأن التعقيد الزمني العالي للخوارزمية أو احتياجات التخزين الكبيرة تجعل تنفيذها على الكمبيوتر أمراً متعذراً.
2. الخوارزميات العشوائية سهلة الفهم و سهلة البرمجة.

3.2 أمثلة على الخوارزميات العشوائية

هناك العديد من الخوارزميات العشوائية الشهيرة التي تستعمل في العديد من التطبيقات. سوف نطرح هنا فقط إثنين منها: خوارزمية إيجاد الوسط Median لعينة معينة و خوارزمية مقارنة شريطين (Two Strings).

الخوارزمية العشوائية لحساب الوسط

وسط العينة (Median) ، بالتعريف ، هو القيمة التي تقسم عناصر العينة إلى قسمين متساويين بعد ترتيب عناصرها تصاعدياً أو تنازلياً. و بالتالي يحتاج إيجاد وسط عينة إلى ترتيب العينة. يحتاج ترتيب العينة في أفضل خوارزميات الترتيب الحاسوبية إلى عدد من العمليات يقدر بـ $n \log n$ حيث n هي عدد عناصر العينة. في الواقع باستخدام الخوارزميات العشوائية نستطيع حساب الوسط بعدد عمليات أقل من $n \log n$. في الواقع تستطيع الخوارزمية التي سوف نطرحها في هذه الفقرة إيجاد الوسط بعدد عمليات من رتبة n . الخوارزمية العشوائية التي سوف نطرحها هنا أكثر عمومية حيث إنها تحسب العنصر الأصغر ذا الترتيب k . و بالتالي فإن حساب الوسط سوف يكون حالة خاصة. الفكرة العامة للخوارزمية العشوائية لإيجاد العنصر الأصغر ذي الترتيب k هي أن نقوم بتجزئة العينة إلى متجهتين و ذلك باستخدام قيمة مقترعة عشوائياً تسمى نقطة الفصل (Pivot). إحدى المتجهتين تحوي الأعداد التي هي أصغر من قيمة الفصل و المتجهة الأخرى تحوي القيم التي تكون أكبر (أو تساوي) قيمة الفصل. (للتوضيح: نقطة الفصل هي قيمة مختارة عشوائياً من قيم العينة.) ثم نقوم بالبحث عن أصغر عنصر ذي الترتيب k في المصفوفة المحددة. يمكننا تحديد المصفوفة التي تحوي العنصر الأصغر باستخدام فقط حجم الجزئين. أي أننا في كل عملية نحتاج للبحث عن العنصر في مجموعة أقل.

بشكل أكثر تحديداً لنفرض أننا قسمنا العينة إلى متجهتين المنجهة LESS و تحوي جميع عناصر العينة التي تقل عن نقطة الفصل والمنجهة GREATER و تحوي

جميع عناصر العينة التي هي أكبر أو تساوي نقطة الفصل. فإذا كنا نبحث مثلاً عن العنصر الأصغر ذي الترتيب 87 ، و لنفرض أن عدد عناصر LESS هي 200 عندئذٍ فإن هذا العنصر موجود في المتجهة LESS و ليس في المتجهة GREATER. أما إذا كان عدد عناصر LESS 40 مثلاً ، فإن العنصر الذي يتم البحث عنه موجود في المتجهة GREATER. في الحالة الأخيرة نحتاج إلى إيجاد العنصر الأصغر ذي الترتيب 46 المتجهة GREATER (لماذا؟). أما إذا كان حجم LESS يساوي 86 فإن العنصر الأصغر ذا الترتيب 87 هو عنصر الفصل نفسه (لماذا؟).

يمكن طرح الخوارزمية العشوائية لحساب الوسط بالخطوات التالية. لنفرض أن لدينا عينة ممثلة بالمتجهة غير المرتبة A و التي حجمها n و عدد k أصغر أو يساوي n و يمثل دليل عنصر الفصل.

الخوارزمية

FindMedian (A, n)

1. اختر عشوائياً أحد عناصر المتجهة A و ليكن هذا العنصر p .
2. نقوم بتجزئة المتجهة A إلى متجهتين جزئيتين LESS و GREATER حيث LESS تحوي جميع العناصر التي تقل عن p و GREATER تحوي بقية العناصر. في أثناء التجزئة، نقوم بعد العناصر في المصفوفة LESS و لنفرض أن هذا العدد هو L .
3. لدينا ثلاثة حالات
 - I. إذا كان $L = k - 1$ فإن العنصر المطلوب هو p .
 - II. إذا كان $L > k - 1$ عندئذٍ نبحث عن العنصر في المصفوفة LESS. أي *FindMedian (LESS, k)*
 - III. إذا كان $L < k - 1$ عندئذٍ نبحث عن العنصر المطلوب في المصفوفة GREATER. أي *FindMedian (GREATER, k-L-1)*

خوارزمية مقارنة شريطين String Comparison

لنفرض أن إحدى الشركات لديها فرعان الأول في دمشق و الثاني في حلب. لدى الشركة قاعدة بيانات في دمشق تحتوي على جميع فعاليات الشركة. و لدى الشركة قاعدة البيانات نفسها في حلب (أي تحتوي على ذات البيانات). يجب أن تكون قاعدتا البيانات دائماً متطابقتين. لضمان تطابق المعلومات في قاعدتي البيانات ، هناك برمجيات تقوم بتحديث قاعدة البيانات في حلب عبر الإنترنت و ذلك كلما تغيرت قاعدة البيانات في دمشق. إذا كانت الشركة تريد التأكد من هذا التطابق، فإن أحد الحلول هي أن تُرسل قاعدة البيانات في حلب إلى دمشق عبر الإنترنت للمطابقة. لكن هذا مكلف للغاية لأن نقل كمية كبيرة من المعلومات على الإنترنت مكلفة مادياً و غير آمن.

تقنياً يمكن تمثيل كل قاعدة بيانات على شكل شريط من البتات (البت هو رقم ثنائي: 0 أو 1). يمكننا الآن تحويل كل شريط من البتات إلى عدد صحيح. لنفرض أن قاعدة البيانات في دمشق مُمثلة على شكل عدد صحيح x وقاعدة البيانات في حلب مُمثلة على شكل شريط y . نختار الآن عدداً أولياً p و نحسب بصمة (Fingerprint) y بالشكل: $y_p = y \bmod p$. بفرض أن p ليس عدداً كبيراً جداً ، يمكن إرسال البصمة عبر الإنترنت إلى دمشق للمطابقة مع بصمة مولدة من قاعدة البيانات في دمشق و ذلك باستخدام ذات العدد الأولي P . تعتمد الخوارزمية العشوائية على أخذ بصمة من شريط قاعدة البيانات في حلب ، ترسل هذه البصمة عبر الإنترنت إلى دمشق ، و من ثم تقارن مع بصمة شريط قاعدة البيانات في دمشق x_p . إذا كان $x_p \neq y_p$ فإن قاعدتي البيانات غير متطابقتين. أما في حال التساوي فإن هناك ثقة بأن القاعدتين متطابقتان.

في الواقع إن إختيار عدد أولي p بشكل محدد غير مفيد تطبيقياً. هناك أعداد صحيحة مختلفة x, y بحيث $x - y$ مضاعف لـ p و أن $x \equiv y \bmod p$ ولكن الشريطين غير متساويين. للتخلص من هذه المشكلة ، يتم إختيار p بشكل عشوائي في كل مرة تحسب فيها بصمة الشريط لكي تقارن مع بصمة الشريط الثاني. فيما يلي نعرض خوارزمية المقارنة العشوائية.

1. اختر عدداً عشوائياً أولياً p حيث $p < M$ و M عدد موجب مختار.
2. احسب بصمة الشريط y بالشكل: $y_p = y \bmod p$. ثم أرسل البصمة y_p و العدد الأولي p إلى دمشق.
3. في دمشق تحسب بصمة الشريط x بالشكل: $x_p = x \bmod p$.
4. تقارن x_p مع y_p . إذا كان $x_p \neq y_p$ فإن الشريطين غير متساويين و إلا يكونان متساويين.

يُبرهن على أن إحتمال فشل هذه الخوارزمية في الوصول إلى قرار صحيح حول تطابق شريطين يعطى بالعلاقة التالية:

$$p(\text{fail}) \leq \frac{\pi(n)}{\pi(M)}$$

حيث $\pi(n)$ هي عدد الأعداد الأولية الأصغر من n (عدد البتات في الشريط) و $\pi(M)$ هي عدد الأعداد الأولية التي هي أصغر من M . على سبيل المثال، إذا كان عدد البتات $n=10^{12}$ و العدد $M=2^{64}$ فإن احتمال فشل الخوارزمية:

$$P(\text{fail}) \leq 8.71 \times 10^{-8} \text{ و هو احتمال صغير.}$$

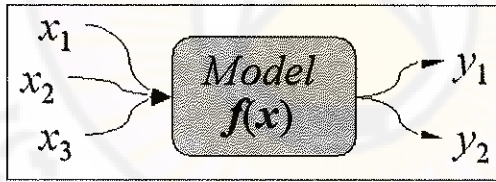
سوف نخصص بقية الفصل لدراسة نوعين من الخوارزميات العشوائية. النوع الأول يسمى خوارزمية مونت كارلو Monte Carlo Algorithm وتضم مجموعة من الخوارزميات التي تعطي دائماً نتيجة. في حالات نادرة قد تكون النتيجة غير صحيحة. النوع الثاني يسمى خوارزمية لاس فيغاس Las Vegas Algorithm و تضم مجموعة من الخوارزميات التي تعطي دائماً نتيجة صحيحة أو لا تعطي أي نتيجة (أي أنها لا يمكن أن تعطي نتيجة خاطئة).

4.2 المحاكاة بطريقة مونت كارلو Monte Carlo Simulation

يهدف هذا القسم إلى تعريف المحاكاة الحاسوبية و النماذج المقررة مركزين على محدوية النماذج المقررة و خصوصاً في ظل وجود الغموضية. بعد ذلك سوف ندخل مفهوم النمذجة و المحاكاة بطريقة مونت كارلو. يتضمن هذا القسم أيضاً العديد من الأمثلة في طيف واسع من التطبيقات العملية.

1.4.2 المحاكاة الحاسوبية Computer Simulation

تتعامل المحاكاة الحاسوبية عادة مع نماذج لنظم واقعية و ذلك بهدف دراسة خصائص تلك النظم أو القيام بتنبؤات حولها. تتألف هذه النماذج من مجموعة من معادلات رياضية تصف النظام الواقعي التي تأخذ دخلاً معيناً و تنتج خرجاً معيناً. هذا النوع من النماذج يسمى نموذج مقرر (Deterministic Model) بمعنى أن هذا النموذج يعطي نفس النتائج بغض النظر عن عدد المرات التي تعيد فيها الحسابات. يبين الشكل التالي نموذجاً مقررأ.



لاحظ أن النموذج المقرر و الممثل بالدالة $f(x)$ يستقبل دخلاً عبارة عن مجموعة من القيم ممثلة بالمتغيرات x_1, x_2, x_3 و يعطي خرجاً ممثلاً بالمتغيرات y_1, y_2 . على الرغم من أن هذا النوع من المحاكاة يساعد في فهم بعض آليات عمل النظام إلا أنه غير مناسب للقيام بتنبؤات حول النظم. إنَّ العيب الواضح في هذه النماذج هو أنها تعتمد على فرضية أن قيم متغيرات الدخل يمكن تحديدها مسبقاً بدقة و بالتالي يتجاهل تماماً أن هذه المتغيرات تخضع لظروف البيئة المحيطة بالنظام. بكلمات أخرى، فإن قيم بعض متغيرات الدخل أو كلها هي قيم متغيرة وتخضع للتغيرات في البيئة المحيطة بالنظام و

بالتالي فإن افتراض أن قيم هذه المتغيرات ثابتة و يمكن تحديدها مسبقاً هو تبسيط كبير للواقع قد ينتج عنه معلومات غير دقيقة و لا تفيد في إتخاذ قرار حول النظام.

2.4.2 طرق مونتني كارلو Monte Carlo Methods

طرق مونتني كارلو هي طرق عشوائية لحل المسائل بمختلف أشكالها. يمكن لهذه الخوارزمية أن تُخطأ في الوصول إلى حل دقيق للمسألة ، ولكنها تُوجد حلاً صحيحاً بإحتمال كبير مهما تكن المسألة. أي إن إحتمال الخطأ في الوصول إلى حل صغير جداً. لوضع هذا الوصف في شكل أكثر دقة لنفرض أن $0 < p < 1$ ، عندئذ نقول إن طريقة مونتني كارلو صحيحة بمقدار p إذا كانت تعيد حلاً صحيحاً للمسألة بإحتمال لا يقل عن p . من الخصائص الهامة لطرق مونتني كارلو أنه يمكن دائماً تصغير الخطأ كما نريد و لكن على حساب سرعة الوصول إلى الحل.

توفر خوارزمية مقارنة شريطين المطروحة أعلاه مثلاً لطرق مونتني كارلو. اعتماداً على مناقشتنا السابقة ، تعطي خوارزمية مقارنة شريطين حلاً صحيحاً بإحتمال أكبر أو يساوي $1 - \frac{\pi(n)}{\pi(M)}$. أي أن هذه الخوارزمية صحيحة بمقدار $1 - \frac{\pi(n)}{\pi(M)}$. وهذا ما يوضح أن هناك امكانية الحصول على جواب خاطئ لأن إحتمال الحصول على جواب صحيح أقل من 1.

المحاكاة بطرق مونتني كارلو

المحاكاة بطريقة مونتني كارلو هي أحد أشكال المحاكاة الحاسوبية. تمتاز هذه الطريقة بأنها طريقة تكرارية لتقييم نموذج مقرر باستخدام مجموعة من أعداد عشوائية على أنها دخل لهذا النموذج. باستخدام الأعداد العشوائية نكون قد حولنا النموذج من نموذج مقرر إلى نموذج عشوائي (Stochastic). تعطي نتائج دراسة النموذج العشوائي صورة أكثر واقعية عن الغموضية المرافقة لعدد من المتغيرات المؤثرة في المسألة.

المبدأ الأساسي في طرق مونتني كارلو هو القدرة على بناء نماذج لنظم واقعية و من ثم استخدام هذه النماذج لدراسة سلوك هذه النظم. يمكن لهذه النماذج أن تكون مقررّة أو

نماذج عشوائية. تتم عملية محاكاة النظم الواقعية من خلال تنفيذ النموذج على دخل من أعداد عشوائية.

الميزة الهامة لطرق مونتى كارلو هي قدرتها على تحديد (أو وصف) إنتقال الغموضية (Uncertainty) الموجودة في نموذج عناصر الدخل (متغيرات الدخل) إلى غموضية في عناصر الخرج (النتائج). و بالتالي ، تهتم طرق مونتى كارلو بتمثيل الغموضية بشكل واضح. تستخدم طرق مونتى كارلو التوزيعات الإحتمالية لتمثيل الغموضية في متغيرات الدخل. إذا كانت متغيرات الدخل غامضة فإن أية تنبؤات بكفاءة النظام المستقبلية الذي يستخدم تلك المتغيرات الغامضة ستكون أيضاً غامضة. بكلمات أخرى، نتائج التحليل التي تعتمد على متغيرات دخل ممثلة بتوزيعات إحتمالية هي بالضرورة توزيعات إحتمالية.

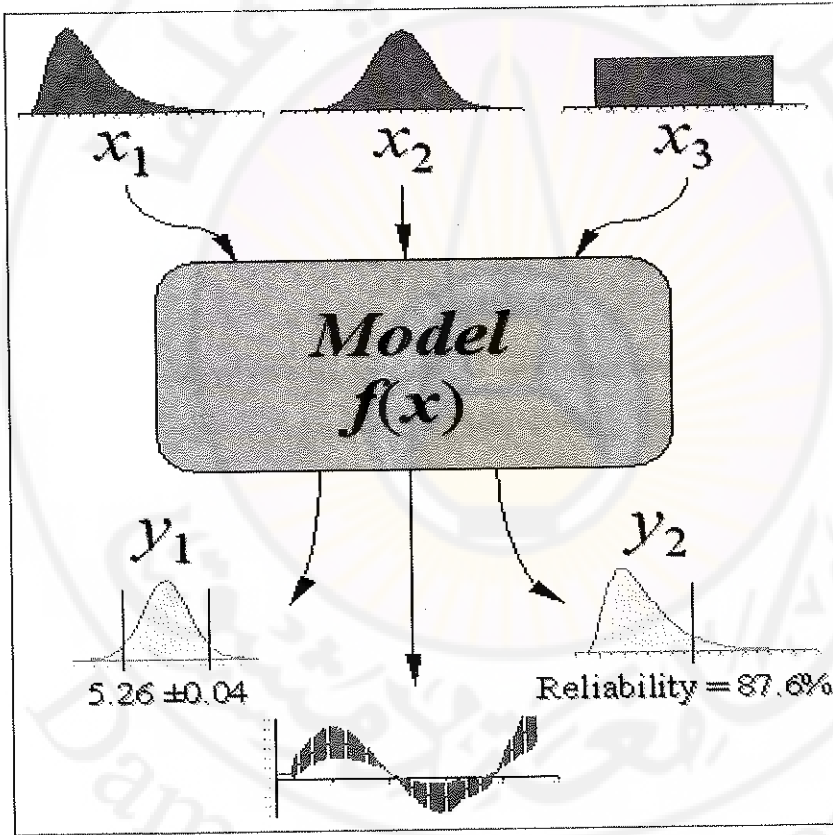
يوضح الشكل (1) بيانياً كيف أن نتائج محاكاة نموذج عناصر دخله غامضة (و بالتالي ممثلة بتوزيعات إحتمالية) تكون أيضاً غامضة و بالتالي ممثلة بتوزيعات إحتمالية مختلفة.

لكي نعين التوزيع الإحتمالي للتنبؤ بكفاءة نظام معين يجب أن نحدد بدقة كيف تنتقل الغموضية المرافقة لمتغيرات الدخل إلى متغيرات الخرج (النتيجة). للقيام بذلك، تتم محاكاة النظام عدداً كبيراً من المرات (1000 مرة مثلاً). في كل عملية محاكاة تقترع قيمة لكل متغير غامض. (أي قيمة عشوائية مقترعة من التوزيع الإحتمالي المحدد لكل متغير). تحسب كفاءة النظام في كل محاكاة. و بالنتيجة نحصل على عدد كبير من نتائج المحاكاة وكلّ منها (النتائج) تمثل حالة ممكنة للنظام. نتائج المحاكاة تستخدم لبناء توزيعات احتمالية للنتائج الممكنة.

خطوات المحاكاة باستخدام طرق مونتى كارلو

1. بناء نموذج محاكاة للظاهرة المدروسة. يشتمل هذا النموذج على متغير الظاهرة و المتغيرات المستقلة التي تؤثر في الظاهرة. غالباً ما تكون نماذج المحاكاة عبارة عن معادلات رياضية.

2. تقدير قيم المتغيرات المستقلة. و غالباً ما تتم من خلال تقدير مجال القيم أو تعريف هذه القيم على شكل توزيعات احتمالية.
3. اقتراح قيمة عشوائية لكل متغير من خلال مجال قيمه أو من خلال التوزيع الإحتمالي الذي يحدد قيم هذا المتغير.
4. تقدير قيمة النموذج عند هذه القيم و من ثم تسجل هذه القيم. تعاد الخطوات (3) و (4) عدداً من المرات.
5. دراسة النتائج لتبيان احتمال الحصول على نتائج معينة للنموذج.



الشكل 1: شكل توضيحي لانتقال الغموضية من عناصر الدخل إلى عناصر الخرج (النتائج).

نموذجياً ، تتم عملية محاكاة النموذج مئات أو حتى آلاف المرات. كلما زاد عدد مرات المحاكاة كلما أصبحت الدقة بوصف الظاهرة الحقيقية أعلى. غير أننا يجب أن نكون حذرين هنا لأن هناك عوامل أخرى تلعب دوراً في دقة النتائج.

- جودة القيم المولدة و خصوصاً إذا كانت هذه القيم عشوائية.
- دقة تقدير مجالات قيم المتغيرات الغامضة أو التوزيعات الإحتمالية المعرفة لقيم المتغيرات.

متى نستخدم المحاكاة باستخدام مونتني كارلو
تستخدم طريقة مونتني كارلو عند الحاجة لتقدير قيم متغيرات، التنبؤ عن ظاهرة معينة، أو إتخاذ قرار مهم حول قضية معينة حيث إن هناك غموضية في المتغيرات التي تحكم المسألة. معظم الفعاليات في الأعمال و التجارة، العمليات ، و التخطيط تكون معقدة للغاية ليتم حلها تحليلياً. إحدى الطرق للتغلب على الغموضية هنا هو أن نبني نموذجاً ثم نعطي هذا النموذج قيم دخل مختلفة و نحلل النتائج. هذه الطريقة السهلة ظاهرياً تنجح عندما يكون عدد المتغيرات واحداً أو اثنين و لكنها تفشل حتماً مع زيادة عدد المتغيرات. في الواقع لو أخذنا مثلاً بسيطاً من الأعمال نجد صعوبة تطبيق طريقة التغيرير اليدوي للدخل بسبب وجود العديد من عوامل الغموضية. على سبيل المثال، المتطلبات المتغيرة للسوق، خطط المنافسين التي لا نعلمها، الغموضية التي تتجلى بعدم القدرة على تحديد النفقات ، و هناك الكثير. و بالتالي يمكن القول بأنه إذا كانت المسألة المدروسة تشتمل على عناصر غموضية (لا يمكن التكهّن بها) ، فإن طرق مونتني كارلو تصبح خياراً مفضلاً أو ربما وحيداً.

قبل الإنتقال إلى بعض الأمثلة التطبيقية لطرق مونتني كارلو ، ندخل مفهوم طرق لاس فيغاس Las Vegas Algorithms.

خوارزمية لاس فيغاس Las Vegas Algorithm

خوارزمية لاس فيغاس هي أحد أشكال الخوارزميات العشوائية التي تقوم باختيارات عشوائية في أثناء البحث عن حل للوصول بسرعة إلى ذلك الحل. بشكل مغاير

لخوارزمية مونتي كارلو ، خوارزمية لاس فيغاس إما أن تعيد حلاً صحيحاً أو لا تعيد حلاً أبداً. أي أنها لا تعيد حلاً صحيحاً باحتمال معين كما هو الحال بالنسبة لطرق مونتي كارلو.

سوف نحاول في هذه الفقرة إعطاء وصف مختصر لبعض الخصائص الهامة لخوارزمية لاس فيغاس ، و لكننا لن نغوص كثيراً في التفاصيل لسببين: (1) يحتاج فهم هذه الخوارزمية معرفة جيدة بالخوارزميات العامة لحل المسائل و لا أعتقد أن طلاب الإحصاء لديهم الخلفية العلمية الكافية لهذا الموضوع ، (2) تطبيقاتها مماثلة لتطبيقات طرق مونتي كارلو و بالتالي نفضل عدم إغراق القارئ في التفاصيل. هناك نوعان مهمان لخوارزميات لاس فيغاس.

النوع الأول The First Category

يهتم هذا النوع في تحسين أداء بعض الخوارزميات التي لا تعتمد على العشوائية. هناك بعض الخوارزميات التي تؤدي بشكل جيد في بعض الحالات و بشكل سيء في حالات أخرى. من الأمثلة الشهيرة نذكر خوارزمية الترتيب المسماة Quick sort. هذه الخوارزمية تكون سريعة في حالات تكون فيها عناصر المتجه المراد ترتيبها موزعة بشكل عشوائي. لكن إذا كانت عناصر هذه المصفوفة مرتبة أو شبه مرتبة فإن هذه الخوارزمية تصبح بطيئة. نستخدم عادةً خوارزمية لاس فيغاس لإصلاح هذه المشكلة و جعل أداء الخوارزمية متماثلاً بشكل مستقل عن وضع العناصر في المتجه.

النوع الثاني The Second Category

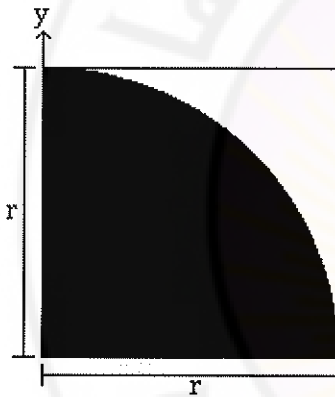
يهتم هذا النوع من خوارزميات لاس فيغاس ، بسحب المسألة إلى حالة عدم وجود حل عندما تكون المسألة لا تملك حلاً دقيقاً ضمن المعطيات الحالية للمسألة. في هذه الحالة يتم تغيير بعض المعطيات بغية البحث عن امكانية وجود حل جديد. هذا في الواقع ما يجعل خوارزمية لاس فيغاس لا تعيد حلاً غير دقيق. من الأمثلة الهامة للمسائل التي ينطبق عليها هذا النوع من خوارزميات لاس فيغاس نذكر: البحث عن المسارات الهاملتونية Finding Hamiltonian Paths.

يُعرف المسار الهاملتوني في بيان ما على أنه المسار الذي يمر من كل رأس من رؤوس البيان مرة واحدة فقط.

3.4.2 تطبيقات على طرق مونت كارلو

تهدف هذه الفقرة إلى استخدام طرق مونت كارلو في أنواع مختلفة من التطبيقات التي تختلف بصعوبتها. من المهم أن يلاحظ القارئ كيف نقوم بنمذجة كل تطبيق و كيف نحلل نتائج النموذج.

حساب قيمة π



لحساب قيمة π باستخدام طرق مونت كارلو، نحتاج إلى بناء نموذج محاكاة و من ثم نقوم باستخدام هذا النموذج لحساب π . سوف نستخدم مفهوم الدائرة لبناء نموذج المحاكاة. يبين الشكل الجانبي مربعاً طول ضلعه r رسمت بداخله ربع دائرة. لنفرض أننا نرمي نقاطاً بشكل عشوائي في المربع في الشكل المجاور، عندئذٍ ليس من الصعب الاستنتاج بأن عدد النقاط الواقعة ضمن ربع الدائرة يتناسب مع مساحتها. يعبر عن ذلك رياضياً بـ:

عدد النقاط الواقعة في ربع الدائرة	=	مساحة ربع الدائرة
عدد النقاط في المربع		مساحة المربع

من معلوماتنا الهندسية ، يمكن أن نجد بسهولة أن:

عدد النقاط الواقعة في ربع الدائرة	=	$\frac{1}{4} \pi r^2$
عدد النقاط في المربع		r^2

عدد النقاط في ربع الدائرة	$\pi = 4$
عدد النقاط الواقعة في المربع	

لحساب العدد π من خلال النموذج أعلاه ، نحتاج إلى آلية لتحديد فيما إذا كانت نقطة مولدة ضمن المربع تقع ضمن ربع الدائرة أم لا. بفرض أن نصف قطر الدائرة يساوي 1. عندئذٍ لأجل كل تجربة نولد عددين عشوائيين ، مثلاً x و y ، و من ثم نحسب بعد النقطتين من نقطة المركز $(0, 0)$ باستخدام معادلة الدائرة. إذا كانت المسافة من نقطة الأصل أصغر أو تساوي الواحد ، فإن هذه النقطة ستكون ضمن المنطقة المظللة و إلا فهي خارج هذه المنطقة.

بتوليد آلاف (أو ملايين) النقاط فإننا نحصل على تقريب لقيمة العدد π . جودة التقريب تعتمد على عاملين إثنين:

1. عدد النقاط العشوائية المولدة لأن كثرة هذه الأعداد يزيد من توزيعها بشكل منتظم في مربع الوحدة.
2. جودة مولد الأعداد العشوائية لأن عدم العشوائية قد يجعل بعض مواضع المربع تحصل على عدد أعلى من النقاط.

بشكل مختصر، يمكن أن نضع الخوارزمية التالية التي تنفذ التجربة أعلاه و تحسب قيمة العدد π . حيث نولد مجموعة من النقاط العشوائية x و y . يمثل المتغير total عدد النقاط العشوائية المولدة. ثم تحسب المسافة بين النقاط المولدة و نقطة الأصل. فإذا كانت المسافة أقل أو تساوي الواحد ، فإن النقطة تنتمي إلى ربع الدائرة و بالتالي تخزن في المتغير hits. قيمة العدد π هي نسبة hits إلى total.

كرر $N \leftarrow 1$

1. اقترح عددين عشوائيين x, y

2. ليكن total يشير إلى عدد النقاط المولدة. 3. $total \rightarrow total + 1$ 4. ليكن distance يمثل بعد النقطة العشوائية (x, y) عن الصفر. أحسب $distance \rightarrow \sqrt{x^2 + y^2}$ 5. إذا كانت $distance \leq 1$، عندئذٍ: $hits \rightarrow hits + 1$	أحسب:
$\pi = 4 * hits / total$	

تقدير الزمن اللازم لإنهاء مشروع

لنفرض أن لدينا مشروعاً معيناً و هذا المشروع يشتمل على ثلاثة أجزاء يجب أن تنجز واحداً تلو الآخر. يبين الجدول التالي أجزاء المشروع و الأزمنة المقدرة لإنهاء كل جزء.

المشروع	الزمن اللازم لإنهاء كل جزء (تقديراً)
الجزء الأول	5 شهور
الجزء الثاني	4 شهور
الجزء الثالث	5 شهور
المجموع	14 شهراً

بطبيعة الحال من المفيد تقدير الأزمنة اللازمة لإنجاز كل جزء من المشروع كما فعلنا في الجدول أعلاه. بحسب التقديرات في الجدول ، فإن المشروع يحتاج إلى 14 شهراً لإنجازه. لكن هذا التقدير يشتمل نوعاً من المجازفة لأنه في كثير من الأحيان لا يمكن التنبؤ و بالتالي التحكم بالمتغيرات التي قد تؤثر على سير المشروع. بمعنى آخر فإن التقدير المطروح في الجدول ربما فيه تبسيط كبير للمسألة. بمعنى آخر إن هذا

النموذج المعتمد على تقدير الزمن بقيمة واحدة محددة ليس فعالاً لأنه لا يفيدنا في تحديد المخاطر (أو الغموضية).

لإنشاء نموذج أكثر فعالية يأخذ بعين الاعتبار عوامل الغموضية و يعطينا احتمالية إنجاز المشروع خلال زمن معين ، نستخدم طرق مونتى كارلو للمحاكاة. على سبيل المثال ، يمكننا استخدام تجاربنا و خبراتنا لتقدير زمن إنهاء كل جزء من أجزاء المشروع الثلاثة. لأجل كل جزء نقدر أصغر زمن ، أكبر زمن ، و الزمن المحتمل. يبين الجدول التالي تقديراً لهذه الأزمنة و الذي سوف نستخدمه في المحاكاة.

المشروع	الزمن اللازم لإنهاء كل جزء (تقديراً)		
	أصغر زمن	الزمن المحتمل	أعلى زمن
الجزء الأول	4 شهور	5 شهور	7 شهور
الجزء الثاني	3 شهور	4 شهور	6 شهور
الجزء الثالث	4 شهور	5 شهور	6 شهور
المجموع	11 شهراً	14 شهراً	19 شهراً

نستطيع القول من خلال المعلومات في هذا الجدول إن المشروع لن ينجز بأقل من 11 شهر و لن يتجاوز زمن إنجازه 19 شهراً. للقيام بالمحاكاة بطريقة مونتى كارلو نقوم بإقتراح زمن عشوائي ، بين أصغر و أكبر زمن ، لكل جزء من الأجزاء الثلاثة و من ثم نقدر زمن إنتهاء المشروع. (نجمع الأزمنة اللازمة لإنجاز كل جزء).

لتحديد إمكانية نتيجة ما (زمن إنهاء المشروع) ، نقوم بتعداد عدد المرات التي أعاد النموذج هذه النتيجة. في حالتنا نريد أن نحدد احتمال أن يتم إنجاز المشروع ضمن فترة محددة. يبين الجدول التالي نتائج المحاكاة (تمت المحاكاة 500 مرة). يمثل العمود الثاني عدد المرات التي كان فيها زمن إنجاز المشروع أقل أو يساوي الزمن المقابل و يمثل العمود الثالث النسبة المئوية.

النسبة المئوية	عدد المرات التي كان فيها زمن تنفيذ المشروع أقل أو يساوي الزمن	الزمن
0%	1	12 شهراً
6%	31	13 شهراً
34%	171	14 شهراً
79%	394	15 شهراً
96%	482	16 شهراً
100%	499	17 شهراً
100%	500	18 شهراً

كما رأينا الزمن المتوقع هو 14 شهر. و لكن بالنظر إلى نتيجة 500 مرة محاكاة فإن احتمال أن يحتاج المشروع إلى 14 شهر أو أقل هو 34%. بكلمات أبسط هناك فرصة مقدارها فقط 34% (أو 1 من 3) أن المشروع سوف ينجز في زمن 14 شهر أو أقل. من ناحية أخرى ، فإن هناك فرصة مقدارها 79% أن المشروع سوف ينجز في 15 شهر أو أقل. علاوة على ذلك ، فإن احتمال أن ينجز المشروع بفترة مقدارها 12 شهر شبه مستحيلة و كذلك فإن احتمال أن يحتاج المشروع إلى 18 شهر شبه مستحيل. و لكن ما هو مقدار الثقة التي نعطيها للنتائج أعلاه. ببساطة إن دقة النتائج تعتمد على جودة تقديرنا لقيم المتغيرات. كذلك فإننا يجب أن ندرك بأن نتيجة المحاكاة هو مؤشرات احتمالية و ليست مؤشرات صحيحة بالمطلق. و مع ذلك ، فإن طرق مونتني كارلو للمحاكاة هي وسائل فعالة تمكننا من القيام بتنبؤات حول مسألة معينة لانملك عنها معلومات كافية.

تقدير أداء شركة تجارية

لنفرض أن شركة تسويق معدات إلكترونية تريد أن تحدد كمية الربح الناتج من تسويق منتج إلكتروني جديد. هناك العديد من العوامل الغامضة Uncertainty التي

يمكن أن تؤثر في الربح. نذكر من هذه العوامل حجم السوق Market Size ، نفقات الإنتاج Expenses ، و العوائد Revenue.

سوف نستخدم طريقة مونتي كارلو لمساعدة هذه الشركة لتحديد الربح مع افتراض وجود الغموض. بالتحديد ، سوف نستخدم مونتي كارلو لتحديد الربح و تقييم المخاطر المرافقة لعملية التسويق.

كما نعلم أن استخدام طريقة مونتي يستوجب بناء نموذج محاكاة عشوائي. لبناء هذا النموذج سوف نبدأ بالنموذج البسيط التالي الذي يربط بين الربح و الدخل و النفقات. يمكن التعبير عن هذا النموذج بسهولة بالشكل:

$$Profit = Income - Expenses$$

أي أن الربح $Profit$ يساوي العائدات من المبيعات $Income$ مطروحاً منه نفقات الإنتاج $Expenses$ للسلع المباعة. كل من العائدات $Income$ و النفقات $Expenses$ هي متغيرات غامضة لا يمكن تقديرها بشكل مباشر. يجب علينا التحقق من استقلالية جميع عناصر الدخل (العائدات و النفقات). على سبيل المثال إذا كان هناك علاقة بين النفقات و العائدات (مثلاً العائدات تعتمد على النفقات) فإن النموذج يجب أن يأخذ ذلك بعين الاعتبار.

سوف نفترض أن الدخل ($Income$) هو جداء عدد القطع المباعة و نرمز له بـ S بسعر مبيع القطعة الواحدة و نرمز له بـ P . لذلك فإن الدخل $Income = S * P$. يمكن أن نطور النموذج بحيث نفترض أن عدد المبيعات S هو عبارة عن عدد العقود L التي تبيعها الشركة شهرياً مضروباً بنسبة العقود التي ينتج عنها مبيعات فعلية R . وبالتالي فإن دخل الشركة الشهري يعطى بالعلاقة التالية:

$$Income = L * R * P$$

إضافةً إلى ذلك فإننا سوف نفترض أن النفقات $Expenses$ هي عبارة عن تكلفة ثابتة (H) مضافاً إليها التكلفة الإجمالية للعقود. لنفترض أن تكلفة عقد واحد هي C و هذه التكلفة تتراوح بين \$0.20 و \$0.80. اعتماداً على دراسة السوق فإن عدد العقود

بالشهر الواحد يتراوح بين 1200 و 1800. النموذج النهائي لربح الشركة يعطى بالصيغة التالية:

$$Profit = L * R * P - (H + L * C)$$

سوف نفترض في هذا النموذج أن H ثابت (أي لا يحتوي أية غموضية). بينما المتغيرات R, L, C, P هي متغيرات غامضة (لا يمكن التكهّن بقيمها بشكل دقيق). يوضح هذا المثال قضية مهمة في إنتقال الغموضية (Uncertainty Propagation)؛ بالتحديد الارتباط Correlation الممكن بين متغيرات الدخل. كما لاحظنا أنه وبعد تحليل كل من الدخل و النفقات إلى مكوناتهم الأساسية وجدنا أن عدد العقود (L) يؤثر على الدخل و يؤثر على النفقات. أي أن الدخل و النفقات غير مستقلين.

محاكاة النموذج

تتم محاكاة النموذج السابق بتوليد قيم عشوائية للمتغيرات الغامضة. سوف نفترض بغية السهولة أن توزيع متغيرات الدخل الأربعة الغامضة هو توزيع منتظم. (بطبيعة الحال ، يمكن القيام بدراسة أعمق على متغيرات الدخل لتحديد التوزيع الذي يتبع له كل متغير. في الواقع في التطبيقات العملية لا يتم فرض توزيع معين لقيم الدخل ، بل تُجرى دراسات دقيقة لتحديد أفضل توزيع احتمالي يمثل هذه القيم.) بشكل خاص ، لنفترض أن قيم متغيرات الدخل محدد بالشكل ادناه.

Input Values (input)

	Nominal	Min	Max
Leads per Month (L)	1500	1200	1800
Cost Per Lead (C)	\$0.50	\$0.20	\$0.80
Conversion Rate (R)	3.0%	1.0%	5.0%
Profit per Sale (P)	\$50.00	\$47.00	\$53.00
Overhead per Month (H)	\$800.00		

القيم في العمود Nominal هي القيم المعطاة للمتغيرات المقابلة. القيم تحت العمودين Min و Max تشير إلى الغموضية في قيم المتغيرات المقابلة. على سبيل المثال، قيم المتغير Leads Per month (L) يمكن أن تكون ضمن المجال [1200, 1800]. لاحظ أن المتغير Overhead Per Month (H) غير غامض بحسب فرضياتنا لذلك له قيمة محددة \$800.00. للحصول على قيمة عشوائية للمتغيرات الغامضة ضمن مجالاتها نستخدم الصيغة التالية:

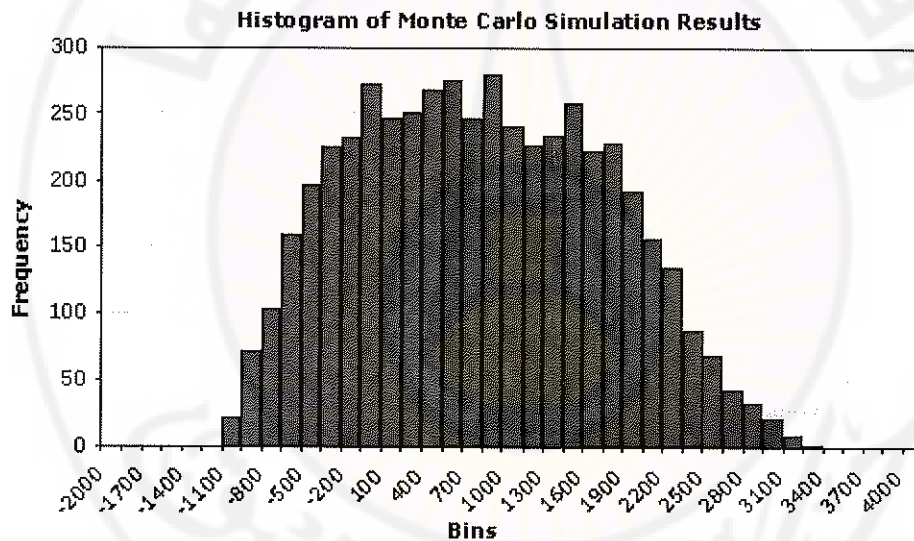
$$\text{min} + \text{RAND}() * (\text{max} - \text{min})$$

حيث إن $\text{Rand}()$ هي دالة تعيد رقماً عشوائياً منتظم التوزيع ضمن المجال $[0, 1]$. لنفترض أننا أجرينا 5000 تجربة على النموذج. في كل تجربة نولد قيمة عشوائية لكل من المتغيرات الأربعة الغامضة و من ثم نحسب ربح الشركة عند هذه القيم. الجدول التالي يبين الـ 11 تجربة الأولى. القيم في العمود الأخير من اليمين تمثل الربح. القيم الموجبة تمثل الربح و القيم السالبة تمثل الخسارة. لنأخذ السطر الأول على سبيل المثال، القيمة العشوائية المولدة لعدد العقود ($L(\text{rand})$) هي 1255.3 و القيمة العشوائية المولدة لتكلفة كل عقد ($C(\text{rand})$) هي \$0.40 و القيمة العشوائية لنسب العقود التي ينجم عنها بيع فعلي ($R(\text{rand})$) هي 1.495% و سعر مبيع السلعة الواحدة ($P(\text{rand})$) \$50.27 و التكلفة الإضافية ($H(\text{rand})$) \$800.00. ينتج عن هذه القيم ربح سالب (أي خسارة) مقدارها -\$352.90.

1255.3	\$0.40	1.495%	\$50.27	\$800.00	-\$352.90
1686.749	\$0.77	1.840%	\$47.31	\$800.00	-\$634.11
1401.01	\$0.76	4.678%	\$52.23	\$800.00	\$1,562.7
1447.243	\$0.72	1.499%	\$48.06	\$800.00	-\$797.22
1595.739	\$0.43	4.866%	\$49.79	\$800.00	\$2,376.83
1628.912	\$0.47	1.694%	\$50.47	\$800.00	-\$166.81

1641.045	\$0.54	4.182%	\$47.44	\$800.00	\$1,563.93
1328.759	\$0.73	1.616%	\$51.48	\$800.00	-\$662.22
1517.658	\$0.21	4.185%	\$47.67	\$800.00	\$1,905.83
1587.563	\$0.78	3.069%	\$48.65	\$800.00	\$331.43
1700.266	\$0.48	3.429%	\$49.90	\$800.00	\$1,295.40

بهذا الشكل يكون لدينا 5000 قيمة مختلفة لمتغير الربح Profit. لنقوم الآن بتحليل نتائج المحاكاة للنموذج باستخدام طرق مونت كارلو. لنبدأ برسم صورة بيانية للنتيجة باستخدام مخطط الأعمدة Histogram. الشكل التالي يمثل مخطط الـ Histogram لبيانات المتغير Profit.



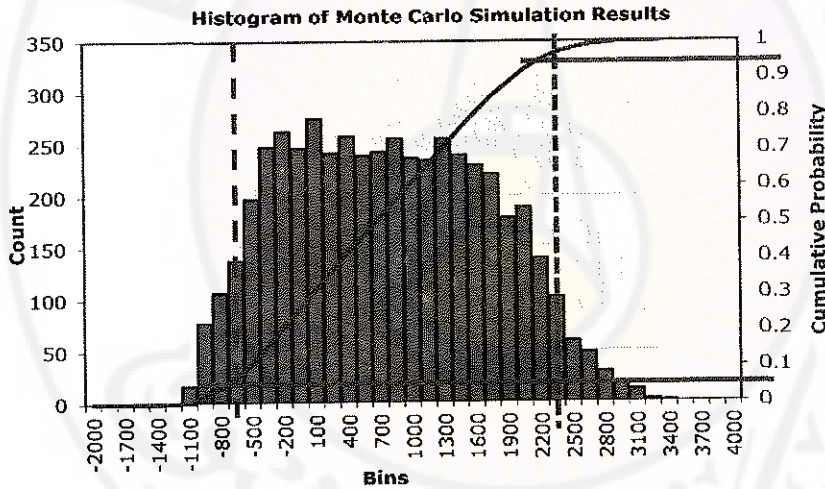
حيث يمثل المحور الأفقي Bins مجالات قيم متغير الربح و المحور الشاقولي يمثل عدد التكرارات Frequency في كل مجال. يمكن من خلال هذه الشكل ملاحظة المعلومات التالية حول ربح الشركة:

- الربح موجب في أكثر الحالات. و هذا يلاحظ من خلال عدد و ارتفاع الأعمدة الواقعة إلى يسار القيمة 100.

- الغموضية كبيرة إلى حد ما حيث إن قيم ربح الشركة تتراوح ما بين 1000- (خسارة) و 3400 (ربح).
- التوزيع لا يظهر أنه طبيعي. لاحظ أن الشكل ليس قريباً من الشكل الجرسى.

على الرغم من أهمية المعلومات التي يعطيها الشكل أعلاه، فإن معلومات أكثر أهمية يمكن الحصول عليها. حساب احتمال أن تكون قيمة الربح أعلى أو أصغر من قيمة معينة أو بين قيمتين معينتين هي معلومات غاية في الأهمية و تساعد في القيام بتنبؤات عن ربح الشركة.

يظهر الشكل التالي نفس مخطط الـ Histogram أعلاه و لكن بعد رسم دالة الكثافة التراكمية التي تظهر على شكل منحنى في الشكل أدناه. لاحظ أن المحور الشاقول الأيمن يمثل القيمة التراكمية للإحتمال من 0 إلى 1.



يمكننا من خلال هذا الشكل الإجابة عن عدد من الأسئلة المهمة التي تتعلق بإعطاء تنبؤات عن أرباح الشركة أو خسائرها. على سبيل المثال، يمكننا حساب احتمال أن تكون خسارة الشركة أكبر من \$700.00 (الخط الشاقول المتقطع على يسار الشكل). إن قيمة هذا الاحتمال لا تتجاوز 5% كما هو واضح في الشكل. بشكل مشابه، فإن احتمال أن تكون أرباح أقل من \$2300.00 (يسار الخط الشاقول اليميني المتقطع) هو تقريباً 95%.

الشكل التالي يعطي ثلاثة احتمالات: احتمال أن يكون ربح الشركة موجباً، احتمال أن تكون خسارة الشركة أكثر من -300، احتمال أن يكون ربح الشركة أكبر من 2000، وأخيراً أن يكون ربح الشركة بين -300 و 2000.

Probabilities			
$Pr(y > 0):$		73%	
$Pr(y < -300)$	=		17%
$Pr(y > 2000)$	=		9%
$Pr(-300 < y < 2000)$		=	74%
Alpha (α):			0.261

نستنتج من التحليل السابق أن طريقة مونتلي كارلو تفيد في إعطاء مؤشرات غاية في الأهمية يمكن أن تبني عليها قرارات صائبة بنسب معينة. تبين نتائج المحاكاة بأن إمكانية الوصول إلى ربح أكثر من 2000 ضئيل (9%) و أن إمكانية عدم الخسارة (الربح موجب) هي 73% مما يعني أن امكانية الخسارة تصل إلى 27%.

خدمة السيارات في محطة وقود

يصل إلى محطة الوقود سيارات بفواصل زمنية عشوائية مختلفة نرمز لها ب X . لنفترض أن محطة الوقود مزودة بمنبع واحد لتزويد السيارات. إن زمن خدمة أي سيارة هو عشوائي ونرمز له بالمتغير العشوائي Y . عندما تكون المحطة مشغولة، فإن السيارات القادمة تنتظر مشكلة طابور (أو رتل) من السيارات.

بفرض أن X متغير عشوائي منتظم التوزيع على المجال $[0, 800]$ وأن Y منتظم التوزيع على المجال $[100, 500]$. إذا فرضنا أن وحدة الزمن هي الثانية، فإن الفترة الزمنية بين وصوليين متعاقبين لسيارتين إلى المحطة تكون بين 0 و 800 ثانية و أن زمن الخدمة هو على الأقل 100 ثانية و على الأكثر 500 ثانية. لاحظ أن هناك أزمنة انتظار للسيارات القادمة و أن هناك أزمنة انتظار لمحطة الوقود حتى تصل إليها سيارة للخدمة. نريد الآن الإجابة عن السؤالين التاليين:

السؤال الأول: ما الزمن الوسطي الذي تحتاجه سيارة من لحظة وصولها حتى لحظة المغادرة (الزمن الوسطي لـ 100 سيارة)؟

السؤال الثاني: ما النسبة المئوية من الزمن الكلي لتخديم 100 سيارة الذي تكون فيه محطة الوقود مشغولة (دون توقف)؟

يمكننا للإجابة عن هذين السؤالين ببساطة إذا كانت محطة الوقود موجودة و بهذه المواصفات، و ذلك إذا قمنا بمراقبة عمل هذه المحطة للحصول على البيانات الإحصائية المطلوبة.

لكن الحصول على هذه المعلومات غير ممكن في الحالات التالية:

- إذا كانت محطة الوقود غير موجودة بالخدمة بعد و هي قيد الإنشاء.
- إذا انحرفت الفترات الزمنية (خلال المراقبة و جمع المعلومات) بين الوصلات المتعاقبة للسيارات القادمة عن مجال قيمها المفروضة.
- إذا انحرفت أزمنة الخدمة عن المجال المفروض.

على الرغم من التحديات أعلاه، يمكننا الحصول على إجابة عن السؤالين السابقين، وذلك من خلال إنشاء نموذج و دراسة ذلك النموذج للحصول على استنتاجات متعلقة بالظاهرة المُنمّجة.

توجد في الواقع نماذج مختلفة من أجل مسألتنا (تقديم السيارات القادمة إلى المحطة)، لكن ربما نماذج المحاكاة العشوائية هي الأكثر مناسبة لوجود العديد من المتغيرات الغامضة (لا يمكن التنبؤ بها سلفاً).

نعلم أنه إذا كان Z متغيراً عشوائياً منتظماً على المجال $[0, 1]$ فإن المتغير العشوائي $L = a + (b - a) * Z$ يكون منتظم التوزيع على المجال $[a, b]$.

بالعودة إلى مثالنا يمكننا أن نكتب:

$$X = 0 + (800 - 0) * Z = 800 * Z \quad (1)$$

$$Y = 100 + (500 - 100) * Z \quad (2)$$

من أجل متتالية القيم العشوائية و المأخوذة من الفصل الأول:

0.865 , 0.159 , 0.079

نحصل من خلال العلاقة (1) على القيم التالية للمتغير العشوائي X:

(3) 692.0 , 127.2 , 63.2

ونحصل من خلال العلاقة (2) على القيم التالية للمتغير Y:

(4) 446.0 , 163.6 , 131.6

كما هو واضح فإن حدود المتتالية في (3) تقع في المجال $[0, 800]$ و حدود

المتتالية في (4) تقع في المجال $[100, 500]$.

إنشاء نموذج المحاكاة والمحاكاة

يمكن تطوير نموذج محاكاة بالاستفادة من علاقات رياضية حسابية بسيطة

تمثل (تُتمذج) زمن الوصول و زمن الخدمة لكل سيارة على حدة. يمثل الجدول (1) نموذجاً حسابياً يوضح نموذج المحاكاة. يوضح هذا الجدول البيانات المتعلقة بكل سيارة تصل إلى محطة الوقود للخدمة.

نموذج المحاكاة لعمل محطة الوقود موجود أيضاً في الجدول (1). بغية تبسيط الحسابات اكتفينا برقم واحد بعد الفاصلة عند اقتراح قيم المتغير العشوائي Z (المنتظم التوزيع على المجال $[0, 1]$) من جدول الأرقام العشوائية (2) (انظر الفصل الأول). سوف نعتبر أن بداية زمن المحاكاة موافق لـ $t = 0$.

عدد عشوائي في المجال $[0, 1]$	Z	0.8	0.5	0.5	0.0	0.9
الفاصل الزمني بين وصولين متعاقبين	$X=800 * Z$	640	400	400	0	720
لحظة الوصول	$T1 = T1 + X$	640	1040	1440	1440	2160
رقم العربة +	$NR = NR + 1$	1	2	3	4	5
انتهت خدمة العربة السابقة؟	$T1 < T3 ?$	n	N	n	J	J
لحظة بداية الخدمة	$T2 = \begin{cases} T1 & T1 < T3 \\ T3 & T1 > T3 \end{cases}$	640	1040	1440	1900	2280
لحظة انتظار العربة +	$W = T3 - T1$	-	-	-	460	120

مجموع أزمان الانتظار +	$Sw = SW + W$	-	-	-	460	580
عدد عشوائي في المجال $[0, 1]$	Z	0.6	0.1	0.9	0.7	0.5
زمن الخدمة	$Y = 100 + 400 * Z$	340	140	460	380	300
مجموع أزمان الخدمة +	$SB = SB + y$	340	480	940	1320	1620
لحظة انتهاء الخدمة	$T3 = T2 + y$	980	1180	1900	2280	2580
زمن توقف العربة +	$D = T3 - T1$	340	140	460	840	420
مجموع أزمان التوقف +	$SD = SD + D$	340	480	940	1780	2200

الجدول (1): نموذج المحاكاة لمحطة تخديم السيارات.

يوضح لنا الجدول (1) أن أول قيمة لـ Z هي 0.8 وبالتالي تكون قيمة X طبقاً للعلاقة (1) هي:

$$X = 800 * 0.8 = 640$$

أي أن القيمة 640 تمثل الزمن المنقضي قبل وصول أول سيارة إلى محطة الوقود، وبالتالي تكون لحظة خدمة هذه العربة هي اللحظة المقابلة لـ $t = 640$. من أجل القيمة $Z = 0.6$ نجد وباستخدام العلاقة (2) أن زمن خدمة العربة هو:

$$Y = 100 + 400 * 0.6 = 340$$

وبالتالي الزمن من بداية المحاكاة (اللحظة $t = 0$) إلى زمن انتهاء خدمة العربة الأولى هو: $T3 = 640 + 340 = 980$.

نعني بالرمز + الموجود في الجدول (1) أن الأعمدة الموافقة لها يمكن ملؤها بعد إنجاز جميع العمليات الحسابية و تدوين النتائج في الجدول. إن الأسطر المشار إليها بـ + تحوي بيانات إحصائية لا تؤثر على مسار المحاكاة.

لقد تم تدوين جميع البيانات اللازمة لإنجاز الحسابات المتعلقة بالعربة الأولى و بحيث يُمكننا العمود الأول من الانتقال إلى حساب القيم الموجودة في العمود التالي. يبدأ العمود الثاني بإيجاد قيمة عشوائية من الجدول (2)، ثم حساب قيمة x التي تساوي: $X = 800 * 0.5 = 400$. و نحسب أيضاً قيمة لحظة وصول العربة الثانية: $640 + 400 = 1040$.

إن حساب قيم العمود الثاني و الثالث يتم بطريقة مشابهة لطريقة الحساب في العمود الأول.

في العمود الرابع من الجدول (1) نلاحظ أن محطة الوقود تكون مشغولة عند وصول العربة إليها، وبالتالي لابد لهذه العربة من الانتظار، حيث يتم تخديمها بعد مغادرة العربة الثالثة المحطة. يجب ملاحظة أن إملاء عمود ما بالقيم تقتضي بالضرورة أن يكون العمود السابق له مليئاً بالقيم.

يمكن استخدام الطريقة المطبقة في هذا المثال (محاكاة عناصر متتالية) في محاكاة العمليات التي تكون عندها الحسابات المتعلقة بعناصر متعلقة فقط بالعناصر السابقة لها وليس بالعناصر اللاحقة لها.

تبين الخوارزمية التالية طريقة حساب القيم الموجودة في الجدول (1).

Input N

Step 1: set $T1=0$, $T3=0$, $SB=0$, $SD=0$, $SW=0$, $X=0$, $NR=0$

Step2: while $i \leq N$

Begin

$Z = \text{RND}$

$X = 800 * Z$; $T1 = T1 + X$

$NR = NR + 1$

If $T1 < T3$ then

Begin

$T3 = T2$

$W = T3 - T1$

$SW = SW + W$

End

Else

$T2 = T1$

$Y = 100 + 400 * Z$, $SB = SB + B$

$T3 = T2 + B$

$D = T3 - T1$, $SD = SD + D$

End

STEP3: Output (...)

نسمي الزمن اللازم حتى نهاية خدمة آخر سيارة بالزمن الكلي للنمذجة. و كما هو واضح فإن هذا الزمن هو قيمة لمتغير عشوائي. في مثالنا: الزمن الكلي للنمذجة من أجل السيارات الخمسة هو 2580 ثانية. إذا كان عدد السيارات المراد نمذجتها ثابتاً $N = \text{constant}$ ، فإن الزمن الكلي للنمذجة يكون متغيراً عشوائياً. و بالعكس إذا كان زمن النمذجة الكلي ثابتاً، فإن عدد السيارات المُخدمة يكون متغيراً عشوائياً مرتبطاً باختيارنا لقيم المتغيرين العشوائيين X و Y .

من الواضح أنه إذا كررنا النمذجة (كما في الجدول (1)) عدداً من المرات و ذلك باختيار قيم مختلفة للأعداد العشوائية (قيم Z)، و قيمة ثابتة لـ N (عدد السيارات)، فإن الزمن الكلي للنمذجة يأخذ قيماً مختلفة. لذلك لابد من التساؤل عن التوقع الرياضي و التشتت و تابع التوزيع للزمن الكلي للنمذجة (زمن تخدم N سيارة) و هذه مسألة إحصائية رياضية.

الآن سوف نُجيب عن السؤالين المطروحين في بداية الفقرة، آخذين بعين الاعتبار قيمة $N=5$ بدلاً من $N = 100$ بغية إيضاح الإجابة عددياً.

- جواب السؤال الأول:

الزمن الوسطي الذي تحتاجه السيارة لحظة وصولها حتى لحظة مغادرتها:

$$\frac{SD}{N} = \frac{2200}{5} = 440 s$$

حيث SD هو مجموع أزمنة التوقف و N هو عدد السيارات.

-جواب السؤال الثاني:

تحسب نسبة انشغال المحطة في أثناء تخدم السيارات الخمسة على الشكل التالي:

$$\frac{100 * 1620}{2580} = 62.8\%$$

إن القيمتين اللتين حصلنا عليهما تمثلان قيمتين لمتغيرين عشوائيين. هذان المتغيران العشوائيان يأخذان قيماً مختلفة من أجل كل نمذجة جديدة بأعداد عشوائية مختلفة. إذا كررنا نمذجة عمل المحطة مرات عديدة، يمكننا تقدير التوقع الرياضي و الانحراف المعياري، كما يمكننا اختبار فرضيات حول تابع توزيع المتغيرين العشوائيين: "الزمن الوسطي الذي تحتاجه عربة من لحظة وصولها حتى لحظة مغادرتها المحطة." "النسبة المئوية لانشغال المحطة من الزمن الكلي."

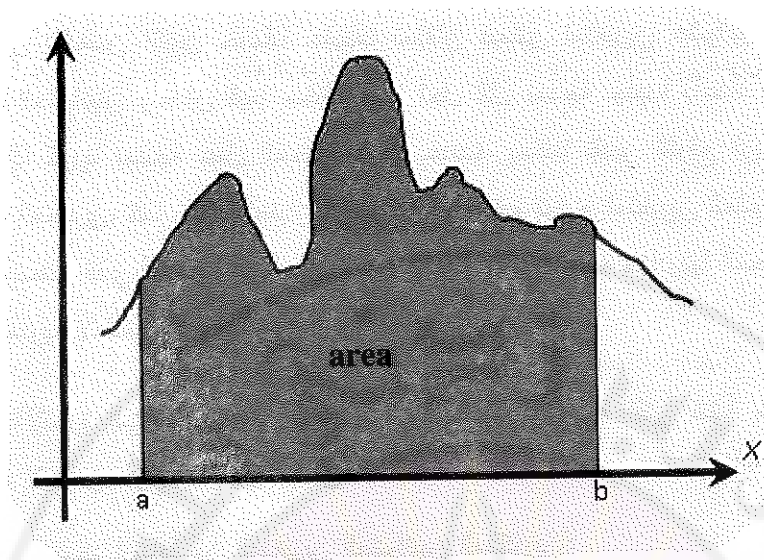
قبل الإشارة إلى كثير من المسائل الإحصائية الرياضية التي تنشأ من تقييم المحاكاة، لابد من طرح السؤالين التاليين:

1. هل يوجد ارتباط بين المتغيرين العشوائيين المذكورين:
"الزمن الوسطي الذي تحتاجه عربة من لحظة وصولها حتى لحظة مغادرتها المحطة."
"النسبة المئوية لانشغال المحطة من الزمن الكلي."
و كيف يمكن إيجاد هذه العلاقة؟

2. هل لعملية خدمة السيارات مرحلة يكون فيها لـ "الزمن الوسطي الذي تحتاجه عربة من لحظة وصولها حتى لحظة مغادرتها المحطة." متوسطاً (توقعاً رياضياً) مختلفاً عن سواء في مرحلة أخرى.

التكامل باستخدام مونت كارلو

لتوضيح فكرة التكامل باستخدام طريقة مونت كارلو من المفيد أن نفهم كيف يعمل التكامل العددي. لنأخذ أبسط حالة و هي التكامل في بعد واحد. يمكن تعريف المسألة على أنها محاولة حساب المساحة area الواقعة تحت منحنى معين و في مجال محدود $[a, b]$.



يمكننا حساب هذه المساحة في أبسط حالة على أنها المجموع المباشر لـ n نقطة تقع في مجال منتظم dx :

$$Area = \sum_{i=1}^n f(x_i) dx \quad (1)$$

حيث $x_i = a + (i-0.5)dx$ و $dx = \frac{b-a}{n}$ وبالتالي تكون المساحة:

$area = \frac{b-a}{n} \sum_{i=1}^n f(x_i)$ من الواضح أن المساحة تساوي متوسط قيم الدالة الكاملة f عند النقاط x_i مضروباً بطول مجال الكاملة.

يمكننا حساب هذه المساحة في أبسط حالة على أنها المجموع المباشر لـ n نقطة تقع في مجال منتظم dx :

$$Area = \sum_{i=1}^n f(x_i) dx \quad (2)$$

حيث $x_i = a + (i - 0.5)dx$ و $dx = \frac{b-a}{n}$ و بالتالي تكون المساحة:

$$area = \frac{b-a}{n} \sum_{i=1}^n f(x_i)$$

المكاملة f عند النقاط x_i مضروباً بطول مجال المكاملة.

بالطبع يمكن زيادة دقة التكامل باستخدام طرق معروفة مثل طريقة Simpson. و مع ذلك لن نوسع المناقشة أكثر من ذلك لأن الهدف هو إعطاء تلميح حول كيفية إجراء المكاملة حسب طرق مونتني كارلو.

يمكن تعميم الصيغة البسيطة السابقة على دالة في الفضاء ذي البعد M و ذلك إجراء الحسابات على M مجال من الشكل:

$$([a_1, b_1], [a_2, b_2], \dots, [a_M, b_M])$$

و التالي فإن قيمة التكامل للدالة ذات M بعد تعطى بالصيغة التالية:

$$V^{(M+1)} = \frac{(b_1 - a_1)(b_2 - a_2) \dots (b_M - a_M)}{N_1 N_2 \dots N_M} \sum_{i_1=1}^{N_1} \sum_{i_2=1}^{N_2} \dots \sum_{i_M=1}^{N_M} f(x_i)$$

حيث $x_i = (x_{i_1}, x_{i_2}, \dots, x_{i_M})$. للحصول على صيغة تفيدنا في نمذجة هذا التكامل وفي طرق مونتني كارلو، فإننا نعيد كتابة الصيغة السابقة بالشكل التالي:

$$V^{(M+1)} = \frac{V^{(M)}}{N} \sum_{i_1=1}^{N_1} \sum_{i_2=1}^{N_2} \dots \sum_{i_M=1}^{N_M} f(x_i) = V^{(M)} \frac{\sum_{i_1=1}^{N_1} \sum_{i_2=1}^{N_2} \dots \sum_{i_M=1}^{N_M} f(x_i)}{N} \quad (3)$$

وضوحاً أن الصيغة الأخيرة تشير إلى أن قيمة التكامل لدالة من البعد M يمكن أن تحسب بحساب متوسط الدالة f على مجال تعريف الدالة و ضرب النتيجة بطول المجال. و بالتالي:

$$V^{(M+1)} = V^M \langle f \rangle$$

$$\langle f \rangle = \frac{\sum_{i=1}^N f(x_i)}{N} \quad \text{حيث}$$

طريقة المعاينة Sampling Method

طريقة المعاينة لتكامل مونتني كارلو هي طريقة مشابهة لقانون التكامل في (1) و (3). الاختلاف هي أن انتخاب النقاط ضمن المجالات Δx يتم بطريقة عشوائية بدلاً من إختيار هذه النقاط بشكل منتظم و من ثم نحسب المتوسط على قيم الدالة عند النقاط العشوائية. على سبيل المثال، لأجل دوال وحيدة البعد (بمتغير واحد)، نختار N نقطة عشوائية في المجال $[a, b]$ و من ثم نحسب التكامل بالشكل:

$$A = \frac{b-a}{N} \sum_{i=1}^N f(x_i)$$

و هي صيغة مماثلة للصيغة (1).

بشكل أكثر عمومياً، من أجل الدوال ذات الـ M بعد فإننا نولد النقاط $x = (x_1, x_2, \dots, x_M)$ عشوائياً في المجالات $([a_1, b_1], [a_2, b_2], \dots, [a_M, b_M])$ حيث $x_i \in [a_i, b_i]$ و من ثم نحسب التكامل:

$$V^{(M+1)} \approx V^M \frac{\sum_{i=1}^N f(x_i)}{N} = V^M \langle f \rangle$$

مثال: حساب مساحة الدائرة Circle Area

لنفرض أن لدينا دائرة في المستوى xy نصف قطرها r . نعلم أنه حتى تكامل وفق مونتني كارلو، فإنه يجب أن نوجد الدالة المكاملة. نعلم أن معادلة الدائرة هي من الشكل:

$$r^2 = x^2 + y^2$$

و بالتالي فإن:

$$f(x, y) = \sqrt{r^2 - (x^2 + y^2)}$$

مكاملة هذه المعادلة تعطي بالطبع نصف مساحة الكرة.

- أبسط طريقة لإنجاز ذلك تكون بإختيار نقاط عشوائية مولدة ضمن المجال $([-r, r], [-r, r])$. ثم نحدد النقاط التي تقع ضمن الدائرة و ذلك من خلال معادلة الدائرة و نحسب قيم الدالة عند النقاط الداخلية. أخيراً نحسب مجموع قيم الدالة.
- يمكن لنا أن نستخدم إجراء معيناً يولد نقاطاً مباشرة ضمن الدائرة. لكن هذا الإجراء يصبح أكثر تعقيداً عند زيادة بعد الدالة.

سوف نعتمد على الطريقة الأولى لكتابة برنامج بلغة C++ لحساب مساحة الكرة بنصف قطر r .

```
#include <math.h>
#include <stdio.h>
main()
{
    int seed=45629;
    float pi=3.141592653589793238;
    int npoints=100000;
    int n, npointsinside;
    float x,y,r,sq,f,sum,fmean,l;
    float ran2(); /* Random number generator provided */
    r=1.0;
    fsum=0.0;
    npointsinside=0;
    for(n=0;n<npoints;n++) {
        x=r*(2.0*ran2(&seed)-1.0);
        y=r*(2.0*ran2(&seed)-1.0);
        /*
        Evaluate function i.e. calculate sqrt(r^2-(x1^2+x2^2))
        but only for points inside the 2D circle
        */
        sq=x*x+y*y;
        if (sq < r*r) {
            f=sqrt(r*r-sq);
```



```

fsum+=f;
npointsinside++
}
}
if (npointsinside==0) {
printf("No points inside. Increase npoints\n");
exit(0);
}
/* MC estimate of <f> */
fmean=fsum/npointsinside;
/* Actual integral: 2 V <f> ; volume is now pi r^2 */
I=2*pi*r*fmean;
printf("Sphere volume is %.6f hits %d\n",I,npointsinside);
}

```

Output

Sphere volume is 4.182319 hits 78575
so the answer is quite close to the correct one, $4\pi/3 = 4.18879020$.

متى تكون طريقة مونتى كارلو أفضل من الطرق العددية للتكامل

دعنا نقارن بين صيغة حساب التكامل عددياً (الصيغة 6) مع صيغة حساب التكامل بوساطة مونتى كارلو (الصيغة 7).

$$V^{(M+1)} = \frac{V^{(M)}}{N} \sum_{i_1=1}^{N_1} \sum_{i_2=1}^{N_2} \dots \sum_{i_M=1}^{N_M} f(x_i) \quad (6)$$

$$V^{(M+1)} \approx \frac{V^M}{N} \sum_{i=1}^N f(x_i) \quad (7)$$

نلاحظ أن هناك إختلافاً كبيراً بين الصيغتين. في صيغة حساب التكامل عددياً، نحتاج إلى إيجاد M مجموع مختلف. أما في صيغة حساب التكامل باستخدام مونتى كارلو، فإننا نحتاج إلى مجموع واحد. و بالتالي، يمكن إن نقول أن طرق مونتى كارلو لحساب تكامل لدوال بعدة متغيرات أفضل من الناحية الحسابية.

من أجل دوال بمتغير واحد، فإنه لا يوجد خلاف جوهري بين الطريقتين. و في الواقع في حالة دوال بمتغير واحد، فإن طرق التكامل العددي التقليدية مثل طرق سيمبسون

تكون دقيقة وربما أكثر فعالية من طرق مونتّي كارلو. ولكن مع ازدياد عدد المتغيرات M للدالة الكاملة، فإن حساب M مجموع يصبح أكثر صعوبةً و بالنتيجة طرق مونتّي كارلو التي تحتاج مجموع واحد أكثر سهولة.

لتوضيح صعوبة التكامل العددي لدوال متعددة المتغيرات، فإننا نحتاج على الأقل مجموعة من النقاط N_i لأجل كل متغير للحصول على جواب قريب و أن عدد المجاميع يزداد بالشكل N_i^M . و بالتالي فإن:

✱ لأجل $N_i = 5$ على سبيل المثال، لأجل كل قيم i (قيمة صغيرة جداً)، عندئذٍ من أجل دالة بعشرة متغيرات نحتاج إلى 10 Million $\approx 5^{10}$ نقطة للحصول على جواب جيد.

✱ و لكن يمكننا استخدام عدد النقاط نفسه (10,000,000) للحصول على دقة جيدة لأجل جالة بعدد إختياري من الأبعاد.

لتوضيح هذه النتائج، لنقم بحساب مساحة كرة في فضاء ذي بعد M و ذلك بالطريقة العددية (باستخدام طريقة النقطة المتوسطة) و طريقة مونتّي كارلو. عدد المجالات الجزئية التي نختار منها نقاط الإرتكاز هو 20 و ذلك لأجل بعد. بينما عدد النقاط المولدة لأجل طريقة مونتّي كارلو هي ثابتة و تساوي 10^5 . بهذه الخيارات لكلٍ من الطريقة العددية و طريقة مونتكارلو نحصل على النتائج الملخصة هذ الجدول.

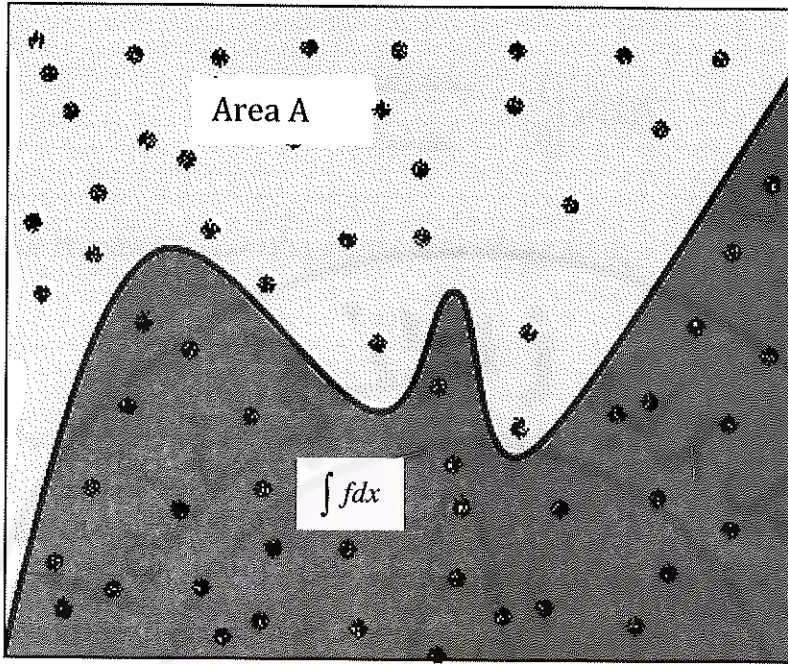
البعد M	الزمن باستخدام الطرق العددية	قيمة التكامل	الزمن باستخدام طرق مونتّي كارلو	قيمة التكامل	القيمة الحقيقية للتكامل
2	0.00	3.1524	0.01	3.1435	3.1415
3	0.00	4.1737	0.07	4.1896	4.1887

4	0.00	4.9023	0.08	4.9330	4.9348
5	0.02	5.2381	0.10	5.2787	5.2637
6	0.30	5.1451	0.13	5.1748	5.1677
7	5.02	4.6704	0.15	4.7098	4.7247
8	89.9	3.9595	0.17	4.0479	4.0587
9	1320	3.3998	0.20	3.3191	3.2985

العمود الأول من اليسار يمثل عدد أبعاد الدالة المكاملة. العمود الثاني و الثالث يحتوي على زمن التنفيذ و قيمة التكامل بالطريقة العددية. العمود الرابع و الخامس يحتوي زمن التنفيذ و قيمة التكامل بطريقة مونتني كارلو. العمود الأخير يمثل القيمة الحقيقية للتكامل. تبين النتائج في الجدول أنه لأجل دوال بعدد متغيرات أقل من 6 فإن الطريقة العددية أسرع. من أجل دوال بأبعاد أعلى فإن الطرق العددية تصبح أبطأ بكثير من طرق مونتني كارلو. النتيجة المهمة التي يمكن استخلاصها من النتائج هي أن الزمن اللازم لطرق مونتني كارلو لحساب التكامل يزداد ببطء شديد على الرغم من أن النتيجة بقيت بنفس الدقة. أي أن طرق مونتني كارلو مهمة في حالة تكامل دوال بمتغيرات متعددة.

طريقة الإصابة و الخطأ Hit and Miss Method

هناك طريقة أخرى لمونتني كارلو و هي أسهل بكثير من الطريقة السابقة. تسمى هذه الطريقة طريقة الإصابة و الخطأ Hit and Miss. الفكرة الرئيسية لهذه الطريقة هي حصر السطح (أو الحجم) الذي تشكله الدالة المكاملة ضمن سطح (أو حجم) ، وليكن A، ممكناً حساب مساحته و من ثم نولد نقاط عشوائية ضمن A و نحسب عدد النقاط التي تقع ضمن السطح المكامل.



لنفرض أن حجم المنطقة الخارجية V_e و نسبة النقاط الواقعة ضمن المنطقة المكاملة (تحت منحنى الدالة في الشكل أعلاه) هو f_h . عندئذٍ يمكن تقريب المساحة تحت منحنى الدالة بالعلاقة:

$$V = V_e f_h$$

يمكن للعلاقة السابقة أن تكتب بشكل مكافئ لحساب قيمة التكامل بطريقة المعاينة. لنفرض أن لدينا دالة M متغير. لنفرض أن عدد النقاط في التجربة N نقطة و أن عدد النقاط الواقعة ضمن السطح المكامل هي N_h . عندئذٍ، يكمن كتابة الصيغة السابقة الشكل:

$$V = V_e f_h = V_e \frac{N_h}{N} = V_e \frac{\sum_{i=1}^{N_h} 1}{N}$$

لنعرف الدالة $f(x)$ بعدد متغيرات M بالشكل التالي:

$$f(x) = \begin{cases} 1 & \text{إذا كانت } x \text{ داخل الحجم} \\ 0 & \text{خلاف ذلك} \end{cases}$$

يمكننا الآن كتابة الصيغة الأخيرة بدلالة الدالة الأخيرة كما يلي:

$$V = V_e \frac{\sum_{i=1}^N f(x_i)}{N}$$

و هي مطابقة لصيغة حساب التكامل بطريقة المعاينة.

تحليل الخطأ لطرق مونت كارلو في المكاملة

Error Analysis of MC Integration

حساب الخطأ في تقدير التكامل بطرق مونت كارلو أمر ضروري. سوف نناقش حساب الخطأ في طريقة المعاينة لحساب المكاملة و طريقة الإصابة و الخطأ.

طريقة المعاينة Sampling Method

يمكن حساب الخطأ في طريقة المعاينة كمايلي. نعلم أن حساب التكامل في هذه الطريقة يعتمد على حساب متوسط الدالة المكاملة. لذلك من المنطقي أن نستخدم الطرق الإحصائية المعروفة لحساب الخطأ في المتوسط. المعادلة العامة لحساب الخطأ في المتوسط و سوف نرمز له بالرمز $\sigma_{\bar{x}}$ يعطى العلاقة:

$$\sigma_{\bar{x}} \approx \frac{\sigma}{\sqrt{N}}$$

حيث يمكن حساب التباين من عينة قيم الدالة عند النقاط العشوائية المولدة في مجال التكامل كمايلي:

$$\sigma^2 = \frac{1}{N-1} \left[\left(\sum_{i=1}^N x_i^2 \right) - N(\bar{x})^2 \right]$$

بفرض أن $N \gg 1$ و بالتعويض العلاقة الأخيرة في علاقة حساب الخطأ في المتوسط حصل على:

$$\sigma_{\bar{x}} = \frac{1}{\sqrt{N}} \sqrt{\frac{\sum_{i=1}^N x_i^2}{N} - (\bar{x})^2}$$

من الواضح أن النقاط x_i هي قيم الدالة f . باستخدام نفس الرموز التي استخدمناها لمتوسط الدالة نجد:

$$\langle f \rangle = \frac{\sum_{i=1}^N f(x_i)}{N}$$

$$\langle f^2 \rangle = \frac{\sum_{i=1}^N f^2(x_i)}{N} \text{ و}$$

و بالتالي فإن الخطأ بحساب التكامل بطريقة المعاينة تعطى بالشكل:

$$\sigma_{V\langle f \rangle} \approx V \sqrt{\frac{\langle f^2 \rangle - \langle f \rangle^2}{N}}$$

بالنتيجة فإن التكامل لدالة f يعطى كمايلي:

$$\int f dv \approx V \langle f \rangle \pm V \sqrt{\frac{\langle f^2 \rangle - \langle f \rangle^2}{N}}$$

تعطى العلاقة الأخيرة أحد أهم خصائص التكامل بطرق مونت كارلو. تزداد دقة التكامل مع زيادة عدد النقاط المولدة N و لاعلاقة لهذه الدقة بعدد متغيرات الدالة. اي نفس عدد النقاط الذي يعطي دقة معينة لتكامل دالة يمكن أن يعطي نفس الدقة و ذلك مهما تغير بعد (عدد متغيرات) الدالة. توضيحاً لذلك فإن هذه النتيجة تعاكس التكامل

العددي المباشر فكلما زادت ابعاد الدالة كلما تطلب ذلك زيادة عدد النقاط في كل بعد K
 N_1^d . حيث N_1 عدد النقاط في بعد معين و d عدد الأبعاد.

5.2 تمارين الفصل Exercises

ملاحظة: عندما نضع العلامة "*" على رقم السؤال فهذا يدل على أن السؤال متوسط الصعوبة. تزداد صعوبة الأسئلة بازدياد عدد العلامات "*".

1. ماهي الخوارميات العشوائية؟ ما مبدأ عملها؟ اذكر بعض المسائل التي يمكن حلها بواسطة هذه الخوارزميات.
2. ما المقصود بالمحاكاة الحاسوبية؟ هل تعتقد أن هناك مصداقية للمحاكاة الحاسوبية؟
3. عرف طرق مونت كارلو. ما هي طبيعة المسائل التي يمكن حلها بواسطة هذه الطرق؟
4. كيف تستطيع طرق مونت كارلو التغلب على الغموض في قيم متغيرات الدخل؟
5. لنفترض أن بعض متغيرات الدخل أو كلها لمسألة معينة هي متغيرات غامضة. ما الذي يمكن قوله بالنسبة لمتغيرات الخرج لهذا المسألة؟
6. لنفترض أن لدينا عينة عشوائية و نريد حساب مئينيات هذه العينة. صمم خوارزمية عشوائية لإيجاد مئيني معين. كيف تستخدم الخوارزمية العشوائية التي صممتهما لحساب الوسط؟
7. لنفترض أن لدينا شريطين من المحارف طول كل منهما K محرف. كيف يمكن استخدام طرق مونت كارلو للتحقق من أن الشريطين متطابقين أو مختلفين بدون الحاجة لمقارنة جميع محارف الشريطين؟
- 8*. لنفترض أن لدينا عينة عشوائية حجمها M و يُظن أن هذه العينة لا تملك منوالاً. كيف تستخدم طريقة مونت كارلو لتأكيد هذا الظن أو نفيه؟ هل الجواب الذي توصلت إليه صحيح 100% لماذا؟

9*. لنفترض أن لدينا عينة عشوائية حجمها M . كيف تستخدم طرق مونتي كارلو لحساب منوال هذه العينة؟ هل النتيجة التي توصلت إليها صحيحة 100% و لماذا؟

10*. لنفرض أن لدينا عينة حجمها N و لنفترض أن لانعرف على وجه التحديد منوال هذه العينة و لكن هناك إدعاء بأن العنصر x هو منوال. كيف تستخدم طريقة مونتي كارلو للتحقق من صحة الإدعاء؟

11**. لنفترض أن لدينا عينة حجمها N و أن هناك هناك ثلاثة عناصر هي الأكثر تكراراً في العينة و لكن أحدهما فقط هو المنوال (العنصر الأكثر تكراراً). كيف تستخدم طرق مونتي كارلو لإيجاد المنوال الصحيح؟ ما مقدار ثقتك بالجواب؟

12. لنفترض أن أحد المستثمرين يريد الاستثمار في بطاقات اليانصيب. بالطبع هناك العديد من المتغيرات الغامضة (لا يمكن تخمين قيمها على وجه الدقة). على سبيل المثال ، عدد البطاقات التي يمكن أن تباع قبل كل سحب و قيمة البطاقة. المطلوب:

A*. ناقش امكانية تطبيق طرق مونتي كارلو على هذه المسألة.

B*. حدد نموذجاً لهذه المسألة.

C**. هل تستطيع أن تحدد ، باحتمال معين ، قيمة البطاقة الواحدة حتى يزداد احتمال ربح المستثمر؟

13**. لنفرض أن لدينا ثلاث مصفوفات A, B, C و لنفرض أن أحدهم يدعي صحة العلاقة التالية: $C = AB$ ، أي أن C هي جداء المصفوفتين A و B . كيف تستطيع أن تتأكد من صحة الإدعاء باستخدام طرق مونتي كارلو و دون القيام بضرب المصفوفتين A و B بشكل مباشر. ما مقدار ثقتك بالجواب؟

14*. لنفرض أن لدينا عينة عشوائية عدد عناصرها N و لنفرض أن هناك شكاً بأن العنصر x هو أكبر عنصر في العينة و أنه وحيد. كيف تستخدم طرق مونتي كارلو للتأكد من أن x هو أكبر عنصر و أن هذا العنصر وحيد؟ ما ثقتك بالجواب.

15. لنفرض أن لدينا عدداً صحيحاً موجباً x و لنفرض أننا نريد معرفة فيما إذا كان هذا العدد أولياً. كيف تستطيع استخدام طرق مونتّي كارلو للتأكد من كون العدد أولياً أم لا؟

16*. لنفرض أنه طُلب منك دراسة جدوى فتح شركة جديدة للمنتجات الغذائية برأس مال معين. يهتم المساهمون بدراسة جدوى فتح مثل هكذا شركة و لا سيما أن هناك العديد من عوامل السوق الغامضة التي لا يمكن معرفتها على وجه التحديد. بشكل خاص يهتم المساهمون بمعرفة الأرباح التي يمكن الحصول عليها و الخسائر المحتملة تحت مختلف السيناريوهات. كيف تستخدم طرق مونتّي كارلو لمساعدة المساهمين. ما مقدار الثقة التي تعطيتها للمستثمرين بعد إنتهاء الدراسة.

17**. لنفرض أن لدينا متجهتين $A1$ و $A2$ لهما الطول نفسه و نفس النوع. نريد أن نتأكد من أن المتجهتين متساويتان دون مقارنة جميع العناصر. أوجد مقدار الثقة في الطريقة التي قمت في تطويرها.

18. أكتب خوارزمية عشوائية تقوم باختبار فيما إذا كانت متجهة ما مرتبة دون القيام بفحص جميع العناصر. ما مقدار الثقة في الجواب الذي تعطيه الخوارزمية؟ هل يمكن أن نعتبر أن الجواب صحيح 100% ؟ لماذا؟

★★★★★★★★★★

تحليل البيانات Data Analysis

1.3 مقدمة

تحيط بنا البيانات (المعطيات) من كل جانب إلى حد أن حياتنا اليومية أصبحت كلها بيانات. كل فعاليات الحياة تقريباً تولد بيانات يجب أن ننتبه إليها و نستفيد منها، لا أن نغض الطرف عنها ببساطة. بعض هذه البيانات ينطوي على أهمية كبيرة بالنسبة إلينا و بعضها الآخر لانهمم بها بشكل مباشر. هذه البيانات تأتي من مصادر متنوعة. لعلّ الانترنت و استخدام الحاسبات قد وفر وسائل لتجميع هذه البيانات و حفظها بكميات كبيرة جداً. يحصل رجل الأعمال على بيانات تتعلق بالسوق ، أسعار المواد الخام، كمية المبيعات في أشهر سابقة و هلم جرا من خلال نظم المعلومات المتوفرة. كما يحصل الطالب على بيانات كثيرة من خلال تجارب يجريها على جراثيم معينة ، من خلال دراسة سلوك مجموعة من الحشرات الضارة للتنبؤ بطريقة تكاثرها ، إلخ. هذه البيانات هي في معظمها مواد خام لا تفضي بشكل مباشر عما تشتمل عليه من معلومات. لكي نحصل على معلومات مفيدة من هذه البيانات فإنه يجب معالجتها.

تسمى عملية معالجة البيانات للحصول على معلومات منها "بتحليل البيانات". كما تحدثنا ، البيانات هي المواد الخام (الأولية) للمعرفة. إلا أن التحدي الأكبر الذي يواجهنا هو كيف نستطيع التعامل مع هذه البيانات و معالجتها بحيث نحصل منها على معلومات و إستنتاجات مفيدة. تحليل البيانات هو العلم الذي يدرس تحويل البيانات من

مجرد أرقام و أشياء لا تفصح عن أي معنى إلى معلومات تفيد في شتى المجالات بدءاً من تأكيد نتائج الأبحاث العملية إلى اتخاذ قرارات إستراتيجية قد تكون ذات حساسية كبيرة.

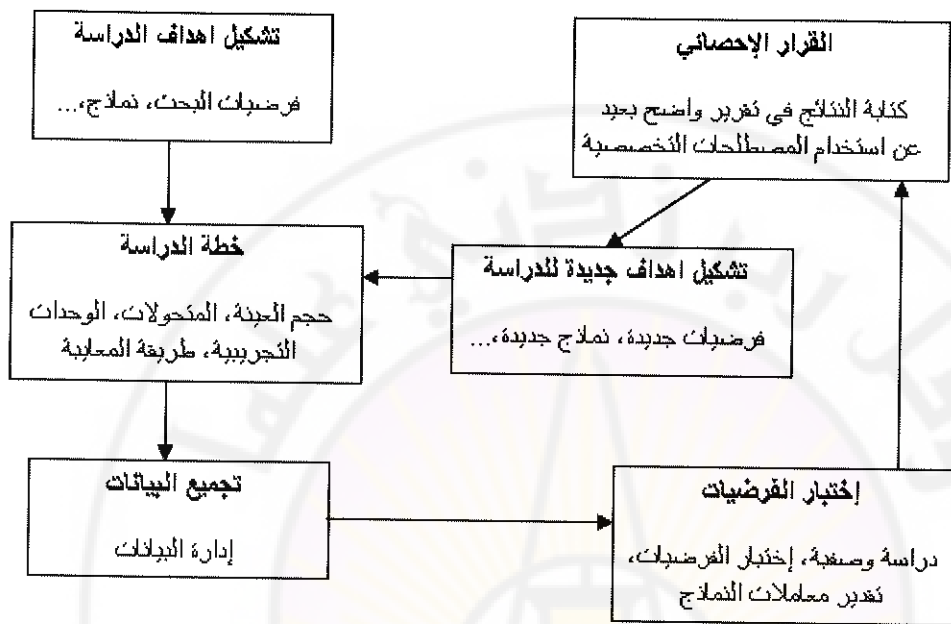
يمكن تعريف تحليل البيانات بشكل عام على أنها عملية تشتمل على الخطوات التالية:

1. تعريف المسألة المطلوب دراستها و فهمها بشكل عميق.
2. تجميع بيانات حول المسألة المدروسة.
3. تلخيص البيانات و تنويعها، حيث يمكن استخدام طرق استكشاف البيانات Exploratory Data Analysis التي سوف ندرسها لاحقاً في هذا الكتاب.
4. تحليل البيانات التي تم تجميعها و تفسير النتائج ثم كتابة تقرير بالخصائص التي تم الوصول إليها.

هذه الخطوات الأربع لتحليل البيانات تتسجم مع الطرق العلمية التي هي مجموعة من المبادئ و الإجراءات التي تستخدم من قبل الباحثين للوصول إلى نتائج (معرفة) حول ظاهرة ما. بشكل خاص تشتمل الطرق العلمية للوصول إلى نتائج و ذلك اعتماداً على تحليل بيانات على الخطوات التالية (الشكل 1) يوضح الترابط ما بين هذه الخطوات بيانياً):

1. تشكيل أهداف الدراسة أو البحث.
2. تصميم مشاهدات أو تجارب للحصول على بيانات حول البحث.
3. تجميع البيانات الناتجة عن التجارب.
4. نمذجة البيانات و تحليل هذه البيانات وصفيًا.
5. وضع الفرضيات المناسبة و التأكد من أن البيانات كاملة و تكفي لغرض اختبار الفرضيات الموضوعية.

6. اختبار الفرضيات و الوصول إلى استنتاجات. لاحظ أنه يمكن للنتائج المستخلصة أن تولد فرضيات جديدة تتطلب بدورها إعادة دورة تحليل البيانات.



الشكل (1): خطوات الطريقة العلمية لحل المسائل الواقعية.

2.3 أهمية تحليل البيانات Importance of Data Analysis

هناك العديد من الأسباب التي تدعونا إلى دراسة تحليل البيانات. السبب الأول هو الحاجة إلى معرفة كيفية تعميم بعض الحقائق الرقمية المنشورة في المجالات و غيرها. نتعامل بشكل يومي مع ادعاءات تأتي إلينا من اتجاهات مختلفة. ادعاءات حول منتجات دوائية قادرة على شفاء مرض معين، نتائج دراسات علوم الاجتماع ، أو نتائج الانتخابات في بلد معين هي جزء صغير جداً من هذه الادعاءات. أيضاً نتعامل يومياً مع نتائج أبحاث علمية منشورة و العديد من هذه النتائج مستقاة من عينات. بعض هذه الاستنتاجات صحيح و بعضها الآخر غير صحيح تماماً. البعض يرى أن تحليل البيانات يمكن أن يستخدم لدعم نتائج الأبحاث العلمية ، و البعض الآخر يرى تحليل

البيانات مجرد وسيلة للكذب. كل هذه العبارات صحيحة لذلك يجب علينا أن نفهم مبادئ تحليل البيانات كي نتمكن من كشف أي نتائج خاطئة. بدون هذه المعرفة فإنه من السهل أن نقع ضحية "التضليل" الإحصائي.

السبب الثاني لدراسة تحليل البيانات هو أنه من الممكن أن يقوم مدير الشركة أو مركز الأبحاث التي تعمل لديها بتكليفك بتفسير بعض نتائج المعاينة أو القيام باستخدام الطرق الإحصائية لتحليل بعض البيانات للحصول على استنتاجات حول واقع التسويق في الشركة. على سبيل المثال ، إذا عملت في حقل العلوم الطبية فإنك سوف تحصل على العديد من الإعلانات التي تصف فاعلية دواء جديد. تعرض هذه الإعلانات بشكل متكرر تجارب و نتائج رقمية تعبر عن فروق حقيقية بين فعالية الدواء الجديد و الدواء القديم. و السؤال الجوهرى هنا: هل هذه النتائج تعبر عن فروق حقيقية بين فعالية الدوائين أو أنها ناتجة عن المصادفة و الخطأ في المعاينة؟ إذا لم تكن واعياً من الناحية الإحصائية بمداول تحليل البيانات ، فإن هذه الأرقام قد تخدعك بسهولة و تقوم بتبني ادعاء منتجي الدواء الجديد ويكون قرارك غير صحيح.

الاتجاه الحالي في المحاكمات مثلاً تسير إلى استخدام متزايد لتحليل البيانات لتقويم قوة الأدلة. استخدام تحليل البيانات في العلوم الاجتماعية ، العلوم البيولوجية ، العلوم الفيزيائية جوهرى لأن هذه العلوم تستخدم الملاحظات حول الظواهر الطبيعية (من خلال المراقبة أو التجارب) لتطوير نظريات حول هذه العلوم أو التحقق منها. يستخدم تحليل البيانات أيضاً في مجال الأعمال و بالتحديد للتنبؤ عن المبيعات و الأرباح. بالإضافة إلى ذلك ، فإن تحليل البيانات يستخدم أيضاً في الهندسة و الصناعة لمراقبة جودة المنتجات.

بناءً على ماسبق ، نجد أن تحليل البيانات له دور محوري في العلوم ، مجال الأعمال ، و الصناعة ، و بالتالي العاملين في هذه المجالات يجب أن يكون لديهم وعي جيد و معرفة كافية كي يميزوا بين الاستنتاجات الصحيحة ، والاستنتاجات غير الصحيحة.

بعض الأمثلة على التفسيرات غير الصحيحة:

1. السببية في الارتباط بين متغيرين. أحد التفسيرات الخاطئة و الناتجة من التحليلات الإحصائية هي عندما تكون هناك علاقة معنوية إحصائياً بين متغيرين. بناءً على وجود هذه العلاقة يتم استنتاج أن التغير في المتغير المستقل يسبب تغيراً في المتغير التابع. هذا الاستنتاج يمكن أن يكون صحيحاً ضمن قيود شديدة على التجربة. على سبيل المثال ، هناك مقالة تناقش العلاقة بين قوة الإيمان الديني و الشفاء من مرض معين. أي كلما ازداد الإيمان الديني ازدادت فرصة الشفاء من المرض. لكن هذا الاستنتاج غير صحيح بالمطلق، و ذلك لأن هناك العديد من العوامل التي تلعب دوراً في الشفاء. و بذلك القول إن درجة الإيمان الديني تؤدي إلى الشفاء غير صحيح بشكل عام.
2. الفرق بين النتائج المعنوية إحصائياً و المعنوية عملياً. تظهر هذه المشكلة في حالة القيام باستنتاجات اعتماداً على عينات كبيرة جداً. نوضح هذا الخطأ بالمثال التالي:

هناك دراسة أجريت لمعرفة فيما إذا كانت هناك اختلافات بين أطوال (ارتفاع) نباتات البازلاء التي تنبت في الربيع و تلك التي تنبت في الخريف. لقد وجدت الدراسة التي أجريت على حوالي 500,000 نبتة أن هناك فرقاً معنوياً إحصائياً بين متوسط طول النباتات التي تنبت في الربيع و متوسط طول النباتات التي تنبت في الخريف مقداره 3 مم (الفرق لصالح النباتات التي تنبت في الربيع). و لكن بسبب كبر العينة فإن هذا الفرق الضئيل لا يملك أي معنوية تطبيقية (حتى لو كان معنوياً من الناحية الإحصائية).
3. حجم العينة. يمكن أن يكون حجم العينة عاملاً مسبباً للاستنتاج الخاطئ. حيث يمكن أن يكون حجم العينة صغيراً جداً لكشف الفروق بين المجتمعات المدروسة. لذلك يجب أن نحدد حجم العينة اللازم للوصول إلى استنتاجات صحيحة. بطبيعة الحال سيكون هناك خطأ و نسعى لكي يكون هذا الخطأ أصغر ما يكون. بالتالي من الأفضل أن نختار عينة بحجم معين بحيث نعرف مسبقاً (تقديرياً) مقدار الخطأ الذي يمكن أن نقع فيه.

3.3 مجال تحليل البيانات Data Analysis Scope

يدخل تحليل البيانات في كل مجالات العلوم سواء أكانت علوماً نظرية أم تطبيقية. متى وجدت بيانات يراد الحصول منها على معلومات فسوف تكون هناك حاجة لتطبيق تحليل البيانات.

سوف نستعرض في هذا الجزء من الفصل بعض التطبيقات الحالية لتحليل البيانات و في مجالات تطبيقية مختلفة. الهدف هو توسيع مدارك الطالب و إلمامه بمجالات تطبيق تحليل البيانات.

تعريف المشكلة: تقليل خطر الأمطار الحامضية على البيئة

Definition of the Problem: Reducing the Threat of Acid Rain on Environment

سبب الأمطار الحامضية هو أحماض النتريك و الكبريت. تنبعث هذه الحوامض من الوقود الجاف (مثل الفحم الحجري) عند الاحتراق و تتطلق إلى الغلاف الجوي. للعلم فقط نذكر ببعض مخاطر الأمطار الحامضية.

1. عندما تذوب الثلوج التي تحوي أحماض النتريك و الكبريت في الربيع و تغذي بعض مواقع تربية الاسماك ، فإن هذه المياه الملوثة تؤدي إلى تقليل نمو الاسماك، و ربما تصبح غير قابلة للأكل.

2. تسبب الأمطار الحامضية إضعاف أشجار الغابات و جعلها عرضة لهجوم الحشرات و الأمراض. كذلك تؤدي الأمطار الحامضية إلى تقليل خصوبة التربة و ذلك بتفكيك الكثير من المواد الغذائية المخزنة في التربة التي تستخدمها النباتات للنمو.

3. تؤدي الأمطار الحامضية الأبنية و البنى التحتية الأخرى كالجسور و الطرقات. في الواقع يعزو الخبراء في الولايات المتحدة الخسائر الكبيرة في البنية التحتية التي تصل إلى مليارات الدولارات إلى الأمطار الحامضية.

الحلول للمشاكل المرتبطة بالأمطار الحامضية ليست سهلة. خصوصاً إذا عرفنا أن هذه المواد تنطلق من الفحم الحجري ، و أن هذا النوع من الوقود يستخدم في الكثير من الدول مصدراً رئيسياً للطاقة. الانتقال إلى فحم حجري أقل بثاً لحمض النتريك و الكبريت ليس سهلاً دائماً. من الناحية العملية فإن إيجاد وسائل احتراق أفضل لإنقاص كمية حمض النتريك و الكبريت هو الحل الأفضل. كذلك فإن محركات الإحتراق الداخلي المستخدمة في السيارات تساهم بشكل كبير في هذه الظاهرة. و بالتالي يجب أيضاً تحسين بنية هذه المحركات لتقليل كمية الانبعاثات الخطرة.

يمكن لتحليل البيانات أن تلعب دوراً كبيراً في مراقبة ظروف الغلاف الجوي و التحقق من فعالية وسائط تقليل انبعاث الأحماض المسببة للأمطار الحامضية و تطوير تكنولوجيا تحكم أفضل وتطوير وسائل بديلة للطاقة.

تعريف المسألة: تحديد فعالية دواء جديد

Defining the Problem: Determining the Effectiveness of New Drug Product

تطوير و اختبار لقاح للحماية من مرض شلل الأطفال Poliomyelitis تعطي مثلاً رائعاً عن كيفية استخدام تحليل البيانات في حل مسائل واقعية. معظم الآباء و الأطفال الذين ولدوا قبل عام 1954 يتذكرون الهلع الذي عانوا منه بسبب الانتشار الواسع لهذا المرض في أشهر الصيف. على الرغم من أن حالات الوفاة بسبب هذا المرض كانت قليلة نسبياً إلا أن الانتشار الواسع و غير المتوقع لهذا المرض جعله موضع اهتمام بسبب احتمال الشلل التام للطفل أو الموت.

اختبار فعالية دواء تتم عن طريق التجربة على مرضى (أحياناً يدعى ذلك بالتجارب السريرية clinical trials). يتم في بعض التجارب السريرية على الدواء تقسيم المرضى عشوائياً قسمين حيث يعالج القسم الأول بالدواء القديم و القسم الثاني بالدواء الجديد. أحد المشاكل المهمة التي يجب حلها في هذه الحالة تتعلق بحجم العينة (أي عدد المرضى المطلوب في التجارب السريرية على الدواء). هذه المشكلة كانت ذات أهمية خاصة في اختبار لقاح شلل الأطفال لأن البيانات من سنوات سابقة تظهر أن حالات الإصابة هي

أقل من 50 حالة لكل 100,000 طفل. لذلك ، نحتاج إلى عدد كبير من المشاركين في الاختبارات السريرية لإظهار الفرق مابين الدواء الجديد و الدواء المستخدم سابقاً. لقد حدد العاملون في تحليل البيانات أن حجم العينة المطلوب هو 400,000 مشارك في التجارب السريرية التي بدأت في عام 1954. تم إعطاء اللقاح الجديد لنصف هؤلاء المرضى و النصف الآخر أعطي اللقاح البديل. (تم تقسيم المشاركين عشوائياً عينتين). و بمراقبة المشاركين على طول أشهر الصيف تبين أن هناك 200 حالة شلل أطفال بين المشاركين و أن معظم هذه الحالات كانت في المجموعة التي تلقت اللقاح البديل. هذه النتائج مدعومة ببعض الحسابات الإحصائية مكنت محلي البيانات من الاستنتاج بأن العقار الجديد أكثر فعالية من الدواء المستخدم سابقاً.

في الواقع كل شركات صناعة الأدوية يجب أن تقوم بإجراء دراسة تقييم لمنتجاتهم الدوائية. مثلاً إدارة الدواء و الغذاء المعروفة بـ (Food and Drug Administration) في الولايات المتحدة تطلب ، و قبل إجازة استخدام أي دواء، القيام بتجارب سريرية يقوم بها العاملون في تحليل البيانات. كذلك هناك بعض الشركات السورية تشتمل على قسم كامل لتحليل البيانات و التأكد من سلامة المنتجات الدوائية و مطابقتها لمعايير الجودة قبل طرحها على المستهلك.

تعريف المشكلة: استخدام البيانات العلمية في المحاكم و تفسيرها

Defining the Problem: Use and Interpretation of Scientific Data in our Courts

القضايا المتعلقة بالمواد الأساسية للمواطنين تشغل بال كل مواطن. حيث ارتفاع الأسعار المتعلق بجشع بعض التجار يمس كل شخص. يمكن لتحليل البيانات أن تلعب دوراً مهماً في هذا المجال. حيث تجمع البيانات حول منتجات معينة و من خلال طرق تحليل البيانات نستطيع الإثبات للمحكمة وجود مخالفات (زيادة غير مسوغة للسعر) في بيع هذه المنتجات. دعونا هنا نتطرق إلى بعض الأمثلة المحددة.

من الأمثلة الطبية المهمة هنا هذه الحالة. يمكن للطبيب المختص بالأورام استخدام تحليل البيانات لتحديد فيما إذا كان هناك ارتباط إحصائي مابين التسرب في طعوم

السلوك المزرعة في صدر المرأة و مرض السرطان ، و ذلك من خلال بيانات مأخوذة لهذا الغرض. إذا استطاع الطبيب أن يجد علاقة ما بين التسرب في هذه الطعوم و مرض السرطان فإن على الطبيب أن يتحقق من كون هذا الارتباط هو بالمصادفة ، أو هو ارتباط حقيقي. المناقشة حول هذه القضية في قاعة المحكمة حول تفسير هذه النتيجة و صلاحيتها تشتمل على تحليلات للبيانات باستخدام المبادئ الإحصائية إضافة إلى التفسيرات السريرية لهذه البيانات.

من الأمثلة الأخرى، نذكر القضايا التي تتعلق بالتمييز في الراتب. في كثير من الأحيان يرفع الموظفون قضايا أمام المحاكم بدعوى التمييز بالراتب على أساس العمر، العرق، الجنس. لأجل ذلك تم بناء بعض النماذج الإحصائية لشرح الاختلافات بين الرواتب اعتماداً على العديد من العوامل مثل الخبرة في العمل، سنوات التعليم ، الأداء في العمل. بعد تعديل الرواتب اعتماداً على العوامل المذكورة أعلاه ، تمت مقارنة الرواتب بين الفئات العمرية المختلفة و بين أبناء الأعراق المختلفة لتحديد فيما إذا كانت هناك فروق حقيقية بين الرواتب ناتجة عن الاختلاف في العمر و العرق. و بالتالي نرى كيف تستطيع تحليل البيانات المساعدة في حل إشكالات كهذه ، و تقديم أدلة أمام المحاكم حول صحة بعض الادعاءات أو بطلانها.

تعريف المشكلة: تقدير عدد الحيتان

Defining the Problem: Estimating Bowhead Whale Population Size

تقدير عدد الحيتان الزرقاء و معدل زيادة هذه الحيتان هي من المواضيع المهمة. أهمية هذه الدراسة تنبع من كون هذه الحيتان هي من أولى الأجناس التي تم منع صيدها لأغراض تجارية. لذلك ، فإن زيادة عددها يثبت نجاعة الإجراءات المتخذة لمنع صيدها. للحصول على البيانات الضرورية للدراسة ، قام الباحثون بإجراء تعداد لهذه الحيتان. بعد ذلك طبق الباحثون الطرق الإحصائية و طرق التقدير على البيانات لتحديد فيما إذا كان عدد الحيتان من النوع المدروس قد إزداد أو نقص منذ منع الصيد التجاري لها. التقديرات الإحصائية بينت أن عدد هذه الحيتان في ازدياد مما يشير إلى أن إجراءات

منع صيد الحيتان تجدي نفعاً و تسهم في زيادة أعدادها. هل تستطيع أن تحدد فائدة تحليل البيانات في هذه الدراسة؟

تعريف المشكلة: التعرض للأوزون و الكثافة السكانية

Defining the Problem: Ozone Exposure and Population Density

مشكلة التلوث من جراء ثقب الأوزون في المدن تشكل واحدة من أكبر التحديات للبيئة. حيث إن ضعف طبقة الأوزون يمكن أن تؤدي إلى انتشار سرطانات الجلد. بينما الكثافة العالية لطبقة الأوزون تسبب مشاكل للجهاز التنفسي و المحاصيل الزراعية و الأشجار.

من الدراسات التي أجريت هي تقدير تركيز الأوزون في منطقة هيوستن في تكساس. بنتيجة الدراسة تبين على أن تركيز الأوزون في تكساس هو أعلى من المستوى المعياري. للتعامل مع هذه المشكلة قام العاملون في تحليل البيانات بتجميع بيانات على مدار الساعة و للفترة من 1980-1993 باستخدام 12 محطة مراقبة. تتعلق هذه البيانات بمستوى الأوزون، درجة الحرارة ، سرعة الرياح ، و اتجاه الرياح. الهدف من تجميع هذه البيانات كان:

1. الحصول على معلومات حول كمية البيانات المفقودة و نمطها إضافة إلى نوعية الأوزون و العوامل المناخية الأخرى (الحرارة ، سرعة و اتجاه الرياح).
2. بناء نموذج إحصائي لكثافة الأوزون يمكن من التنبؤ بتركيز الأوزون في أي موقع (في منطقة الدراسة).
3. تطبيق النموذج الإحصائي لتقدير مؤشرات التعرض التي تأخذ بعين الاعتبار إما التعرض على المدى الطويل أو التعرض على المدى القصير. و ربط هذه المعلومات مع معلومات أخرى عن مؤشرات تعرض مختلفة للحصول على مؤشرات تعرض للأوزون تخص كامل المجتمع.

النموذج الذي بناه الباحثون أعطى تقديرات وضحت أن أعلى مستويات الأوزون تحدث في المواقع التي تكون فيها نسبة الشباب قليلة. أيضاً قدر النموذج أن تعرض الشباب للأوزون تقلص بنسبة 20% من 1980 إلى عام 1993. تفحص توزيع تعرض

السكان للأوزون له العديد من النتائج. بشكل خاص ، تم التحقق من أن الوضع الحالي لمحطات مراقبة الأوزون ليس مثالياً إذا كان الهدف يقتصر على تقويم تعرض السكان للأوزون. يوضح هذا المثال الكثير من القضايا.

1. التعلم من البيانات: بعد تحليل البيانات تكتب النتائج على شكل تقارير تمثل النتائج التي تم التوصل إليها من الدراسة.
2. التخطيط للتجربة و التأكد من صحة هذه التجربة ، و ملائمتها للمسألة المدروسة.
3. كيف يجب أن تجمع البيانات. على سبيل المثال، البيانات أخذت على فترات زمنية و من مواقع مختلفة.
4. ما المتغيرات التي يجب أن تؤخذ بعين الاعتبار. على سبيل المثال ، تم أخذ معلومات عن مستوى الأوزون و متغيرات ثلاثة أخرى: الحرارة ، سرعة الرياح ، و اتجاه الرياح.

تعريف المشكلة: تقويم الرأي العام

Defining the Problem: Assessing Public Opinion

تقويم الرأي العام هو وسيلة لمعرفة رأي الجمهور في قضية تتعلق بالمجتمع. على سبيل المثال ، استفتاء رأي المستهلكين حول سلعة معينة أو استطلاعات آراء جمهور الناخبين حول مرشح ما لإعطاء تقديرات لفرص هذا المرشح في الفوز قبيل الانتخابات. نحن ، وبشكل يومي تقريباً ، على تماس مع نتائج استطلاعات آراء الناس حول موضوعات شتى بدءاً من التعليم ، المنتجات الغذائية ، الرواتب ، مرشحي الانتخابات و غيرها كثير. بشكل خاص المواضيع التالية تظهر بشكل متكرر في الصحف و غيرها من وسائل الاتصال مع الناس.

- ثقة الناس باقتصاد البلد و توقعاتهم المستقبلية حول هذا الاقتصاد.
- المرشحون المفضلون لدى الناس في انتخابات قادمة.

- مواقف الناس من التهرب الضريبي الذي يمارسه العديد من التجار و آثار ذلك على رفاهية المواطن.
- استطلاعات رأي الناس حول رأيهم بالمنتجات الوطنية بالمقارنة مع المنتجات المستوردة.
- ردود أفعال الناس حول المشروعية الدينية للتدخين.
- آراء الناس حول زيادة الضريبة في البلد.

العديد من الأسئلة يمكن أن تطرح حول استطلاعات الرأي. لنفرض أن استطلاعاً للرأي قد أجري حول زيادة مقترحة للضريبة في سورية. ما المجتمع الذي يهم مستطلع الرأي؟ هل المجتمع هو جميع سكان سورية أم هو العاملون الذين يدفعون هذه الضريبة فقط؟ هل سحبت العينة من مجتمع دافعي الضرائب فقط أم من كامل المجتمع السوري؟ إذا كان المجتمع الذي يهم مستطلع الرأي هو فقط دافعو الضرائب ، هل تحقق مستطلع الرأي من أن أفراد العينة المسحوبة كلهم يدفعون ضرائب حالياً؟ ما الأسئلة التي تم طرحها على العينة؟ وكيف صيغت هذه الأسئلة؟ هل كل فرد من أفراد العينة سئل نفس السؤال؟ هل صيغت الأسئلة بحيث تولد انحيازاً إلى إجابة ما؟ هل نستطيع أن نثق بنتائج استطلاعات الرأي؟ هل تمثل نتائج استطلاع الرأي رأي كل المجتمع حول القضايا التي يطرحها الإستطلاع؟

من خلال هذا المثال ، نرى أن بعض التطبيقات لتحليل البيانات تحتاج إلى دقة و دراية كبيرة من قبل المختصين بتحليل البيانات و إلا سوف تكون النتائج المستنبطة من هذه الدراسات موضع شك.

4.3 خوارزمية تحليل البيانات

لا توجد طريقة محددة و عامة يمكن أن ترتب على شكل خطوات يجب اتباعها عند تحليل البيانات. و السبب في ذلك هو أن هناك تنوعاً كبيراً في طبيعة البيانات، السياق المأخوذة منه هذه البيانات، و طبيعة النتائج أو المعلومات المطلوب استخلاصها من البيانات. تقريباً كل توليفة من البيانات و طبيعة المسألة تملك خصوصية تجعلها إلى حد

ما مختلفة عن المسائل الأخرى. بناءً على ذلك فإن أفضل ما يمكن أن يعطى هنا هو مجموعة من الإرشادات العامة ، التي أثبتت جدوى في دراسة مسائل تحليل البيانات.

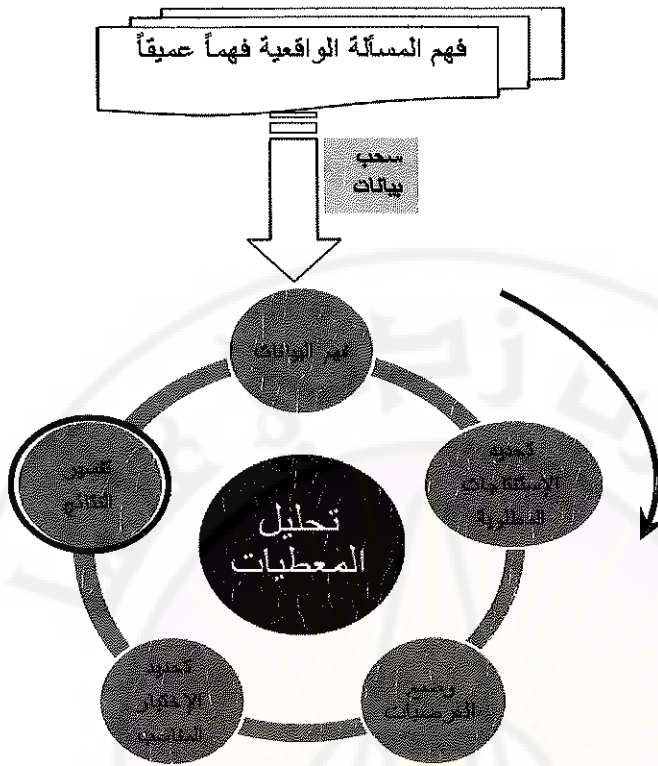
1. يجب على محلل البيانات فهم السياق الذي تنتمي إليه البيانات فهماً عميقاً.
2. أن يحدد محلل البيانات نوعية المعلومات المراد الحصول عليها أو الاستنتاجات التي يراد استخلاصها من البيانات.

3. أن يضع محلل البيانات الفرضيات المطلوب اختبارها. هذه الفرضيات يجب أن تتسجم مع الأسئلة المطروحة.

4. من خلال نوعية البيانات و الفرضيات المطلوب اختبارها ، فإنه يجب على محلل البيانات أن يحدد الاختبارات الإحصائية الأفضل. يجب الانتباه هنا بأنه قد يوجد أكثر من اختبار واحد ممكن و بالتالي فإن اختيار أفضل اختبار هو عمل ليس بالسهل و يحتاج إلى خبرة كبيرة.

5. تفسير نتائج الاختبارات الإحصائية المستخدمة. تشتمل هذه الخطوة على الإجابة على كل فرضية موضوعة و تفسير النتائج بشكل يعكس حقائق هذه البيانات على أرض الواقع. يوضح الشكل (3) الخطوات الخمسة أعلاه بيانياً.

علينا الانتباه على أن اختيار الاختبار الإحصائي المناسب لتحليل مجموعة من البيانات عملية معقدة. تعود هذه الصعوبة إلى أن البيانات الناتجة من مسائل واقعية (Real-World Data) غالباً ما تكون عبارة عن خليط من معطيات متعددة الأنواع. عدم تجانس البيانات يجعل من عملية اختيار الاختبار الإحصائي الصحيح صعبة. وغالباً ما تعود عملية الاختيار إلى تقدير الباحث بدلاً من الاعتماد على طريقة معيارية لتحديد الاختبار الإحصائي الصحيح. هذا يعني أنه من الممكن لخبريين إحصائيين أن يستخدموا طريقتين مختلفتين لاختبار نفس البيانات ، و ذلك اعتماداً على الفرضيات التي يريدون أخذها بعين الاعتبار و طريقة تفسير النتائج. لفهم هذه المشكلة بشكل أعمق ، دعونا نفرض أن لدينا مجموعة من البيانات و فيها:



الشكل (3): خوارزمية تحليل البيانات. كما يوضح الشكل فإن الخطوة الأولى و المهمة هي فهم المسألة المطلوب دراستها. من خلال هذا الفهم يتم انتقاء البيانات المناسبة لفرضيات المسألة. هذه البيانات يجب أن تفهم و تتقن من أي قيم شاذة قد تجعل الاستنتاجات التالية خاطئة.

1. متغير تابع كمي (أي يقبل أي قيمة في مجال تعريفه).
2. خمسة متغيرات مستقلة.
 - a. ثلاثة من هذه المتغيرات المستقلة كمي.
 - b. واحد اسمي كيفي (ليس مجالياً) — اسمي بمعنى أن قيمه لا تقبل الترتيب.
 - c. و المتغير الخامس ترتيبي كيفي — ترتيبي بمعنى أن قيمه تقبل الترتيب.

في بيانات من هذه الأنواع من الممكن أن تستخدم طرق مختلفة لمعالجتها. يمكن لبعض الإحصائيين أن يستخدموا طريقة تحليل الانحدار المتعدد Multiple Regression Analysis حيث يتم التعامل مع المتغير الترتيبي على أنه متغير كمي و يستخدمون متغيرات مساعدة للمتغير الاسمي. يمكن لمجموعة أخرى من الإحصائيين أن يستخدموا طرق تحليل التباين Analysis of Variance لتحليل البيانات.

بشكل عام يمكن الاعتماد على العوامل التالية لاختيار التحليل المناسب للبيانات.

1. اعتماداً على الأسئلة المطروحة و التي يراد الإجابة عنها من البيانات.
2. اعتماداً على خصائص البيانات نفسها و التي تفرض شكلاً معيناً من الاختبارات الإحصائية.
3. اعتماداً على خصائص المعاينة (تصميم المعاينة).

5.3 ملخص الفصل

يتعامل مجال تحليل البيانات مع مبدأ التعلم من البيانات. لذلك فإن الباحثين في المجالات الطبية، الاجتماعية، المحاسبة و الاقتصاد، المستهلكين و قادة الحكومات كلهم يملكون دوراً في عملية تحليل البيانات. و بالتحديد لهم دور في جمع البيانات، تلخيص البيانات، تحليل البيانات، والتوصيل الفعال لنتائج تحليل البيانات.

6.3 تمارين الفصل Exercises

1. ما تحليل البيانات؟ هل تعتبر أن تحليل البيانات مهم؟ عدد بعض المجالات التي يمكن أن نطبق فيها تحليل البيانات.
2. عندما تريد أن تتخذ قراراً حول ظاهرة معينة، هل يمكنك أن تستفيد من تحليل البيانات؟ إذا كان جوابك بالايجاب، بيّن كيف يمكنك الاستفادة من تحليل البيانات.
3. إذا كان جوابك بالنفي، بيّن كيف يمكنك أن تتخذ قراراً حول الظاهرة؟ كيف تصف العلاقة بين تحليل البيانات و الإحصاء؟ هل يمكنك إجراء أي دراسة للبيانات دون استخدام القوانين الاحصائية.



عرض خصائص البيانات الإحصائية و دراستها Displaying and Exploring Statistical Data

1.4 مقدمة Introduction

يمكن تقسيم علم الاحصاء إلى قسمين رئيسيين: احصاء وصفي Descriptive Statistics و احصاء استنتاجي Inferential Statistics. يهتم الاحصاء الاستنتاجي بالحصول على معلومات جديدة من خلال بيانات متوفرة. هذه المعلومات ربما تكون حول كامل مجتمع البيانات أو جزء من هذه المجتمع. سيكون الاحصاء الاستنتاجي موضوع الفصلين القادمين. يهتم الاحصاء الوصفي بوصف البيانات دون محاولة تعميم النتائج على كامل مجتمع البيانات لأن تعميم النتائج يندرج تحت موضوع الاحصاء الاستنتاجي.

سواءً كنا نصف بيانات كامل المجتمع أو عينة من هذا المجتمع و ذلك للحصول على استنتاجات حول المجتمع فإن وصف البيانات أولاً هو خطوة مهمة للغاية. وصف البيانات يعطي معلومات كثيرة نحتاجها للقيام باستنتاجات من العينات أو نحتاجها لتطبيق الاجراءات المناسبة لاستنتاج حقائق حول المجتمع.

هناك طريقتان أساسيتان لوصف مجموعة من البيانات هما: (1) طرق بيانية تشتمل على تمثيل البيانات بمخططات مناسبة و (2) طريقة الوصف العددي التي تصف البيانات بدلالة تكرارات ، متوسطات ، وغيرها من القيم الوصفية.

هناك العديد من طرق عرض البيانات التي يمكن أن تعطي صورة واضحة عن هذه البيانات. و لكن و قبل البدء في مناقشة تلك الطرق ، نبين أهمية تنظيم البيانات و بشكل خاص سوف نتحدث عن تنظيم البيانات في جداول. تنظيم البيانات في جدول هي طريقة لفهم البيانات التي يتم فيها تبويب البيانات حسب خاصية أو مجموعة خواص. على سبيل المثال، يمكن أن تحسب تكرارات البيانات و من ثم ترتب وفق قيم التكرار. لتكن لدينا البيانات التالية:

4 workers, 3 workers, \$2000, %15, %70, \$40, %12, \$30, \$3600, \$26.5

من الصعب القيام بأي استنتاجات حول هذه البيانات. لنأخذ على سبيل المثال البيان \$30. لا يمكن من خلال النظر إلى البيانات معرفة ماذا يمثل بالضبط. هل هو سعر المبيع لعبة حلوى؟ هل هو الأجر الساعي أو اليومي لأحد العاملين في شركة؟ أو هل هو كمية الربح نتيجة بيع كمية من سلعة معينة؟ لكن دعونا ننظر الآن إلى الجدول (1) الذي يظهر البيانات السابقة نفسها و لكن بعد عرضها على شكل جدول.

رقم	مقدار الإنفاق		الربح الكلي خلال	عدد العمال المكرمين	سعر السكر
	استهلاك مواد	هدر عام			
الأول	%70	15%	\$2000	3	\$30
الثاني	%60	%12	-	0	\$26.5
الثالث			\$3600	4	\$40
الرابع	-	-	-	-	-

الجدول (1): عرض البيانات أعلاه بشكل جدولي. تعطي عناوين الأعمدة

معلومات إضافية (مُعرفات) توضح معنى كل بيان.

بسبب تبويب هذه البيانات على شكل جدول و إضافة بعض المعلومات المُعرفة أصبحت هذه البيانات أكثر وضوحاً. على سبيل المثال، لقد أصبح واضحاً أن البيان

\$30 هو سعر وحدة السكر و كذلك أصبح واضحاً الآن أن "3 workers" يعني عدد العمال الذين تم تكريمهم في الشهر الأول من السنة و ليس مثلاً عدد العمال الذين تم تعيينهم حديثاً. كذلك فإن النسبة 15% هي نسبة الهدر العام للموارد و أن 75% هي نسبة المواد المستهلكة. يوضح هذا المثال أهمية عرض البيانات لأن ذلك يزيد من وضوح معنى البيانات.

عرض البيانات وحده قد لا يكفي لفهمها (البيانات) و خاصة إذا كنا نريد تحليلها لإستخلاص نتائج منها. لذلك لابد لمحلل البيانات أن يتعلم تقنيات إضافية لدراسة خصائص البيانات بشكل أكثر عمقاً و ذلك للمساعدة في اختيار الوسيلة الأكثر فعالية لمعالجتها. (سوف نتحدث بشكل مفصل عن معالجة البيانات في الفصول القادمة). تشمل دراسة خصائص البيانات على العديد من القضايا نذكر منها القضايا التالية:

1. القيم الشاذة. هناك مصدران اثنان على الأقل للقيم الشاذة. المصدر الأول يعود إلى عدم تجانس مجتمع العينة. على سبيل المثال، لدى قياس ضغط الدم لدى مجموعة من الأصحاء بينهم مريض مصاب بضغط الدم العالي. من المحتمل أن يكون القياس الذي يعود للشخص المريض بضغط الدم قيمةً شاذةً. المصدر الثاني هو الخطأ في جدولة البيانات حيث يمكن أن نكون قد سجلنا سهواً رقماً بشكل خاطئ. القيم الشاذة تؤثر بشكل كبير في النتائج، و بالتالي يجب اكتشافها والتخلص منها مبكراً قبل البدء بتحليل البيانات.

2. شكل توزيع مجتمع العينة. معرفة شكل توزيع العينة مهم جداً لأن غالبية الاختبارات الإحصائية تفترض شكلاً معيناً للتوزيع. على سبيل المثال، الاختبارات الإحصائية الوسيطة التي سندرسها في الفصول القادمة تفترض أن تكون العينة مسحوبة من مجتمع يتوزع بشكل طبيعي أو على الأقل قريب من الطبيعي.

عرض البيانات و دراسة خصائصها ليس بالضرورة عمليتين منفصلتين. بعض وسائل عرض البيانات التي سندرسها في هذا الفصل تستطيع الكشف عن طبيعة البيانات. على سبيل المثال، من خلال الرسم البياني يمكن أن نكتشف القيم الشاذة. كذلك من خلال الرسم البياني يمكن أن نتعرف إلى شكل توزيع مجتمع العينة. و بالعكس،

توفر طرق دراسة خصائص البيانات أيضاً، و سائل عرض للبيانات على شكل رسوم بيانية.

سوف ندرس في هذا الفصل مجموعة مختارة بعناية من وسائل عرض البيانات مثل العرض الصندوقي و العرض التبعثري. و كذلك سوف ندرس إجراءً لدراسة خصائص البيانات يسمى Exploratory Data Analysis (EDA) (أي التحليل الاستكشافي للبيانات).

2.4 العرض البياني Graphical Display

بعد جمع البيانات فإنه من المفضل تنظيمها على شكل جدول مثلاً ، و تعرض ، ويتم التدقيق فيها باستخدام عدة طرق عرض بياني. العرض البياني له أهمية خاصة لكونه يعطي صورة بصرية عن البيانات و الصور البصرية سهلة الفهم و التدقيق كما يقال "صورة تعبر عن ألف كلمة".

لذلك تم تخصيص هذا الفصل للتعريف بعدد من تقنيات العرض البياني التي يستخدم بعضها لعرض بيانات متغير واحد و بعضها الآخر لعرض بيانات متغيرين أو أكثر. سوف تطرح الأمثلة المناسبة من مخرجات البرنامج الاحصائي SPSS.

1.2.4 وصف البيانات في متغير واحد

Describing Data on a Single Variable

بشكل عام و قبل البدء بعرض البيانات لابد من ترتيبها في فئات بحيث يكون كل قياس ينتمي إلى فئة واحدة فقط. هذا الاجراء يزيل أي غموض قد يظهر في أثناء عرض البيانات. لتوضيح ذلك دعونا ندرس بدقة طريقتي التنظيم المبينة في الجدولين التاليين. لاحظ أن تقسيم البيانات في الجدول الأيمن تضع كل مقدار راتب في فئة واحدة فقط. على العكس من ذلك ، فإن تنظيم البيانات في الجدول الأيسر يكتنفه بعض الغموض. على سبيل المثال، إلى أي فئة ينتمي شخص راتبه 20,000؟ حسب الجدول الأيسر فإنه ينتمي إلى فئة الدخل الأولى و الثانية بأن واحد.

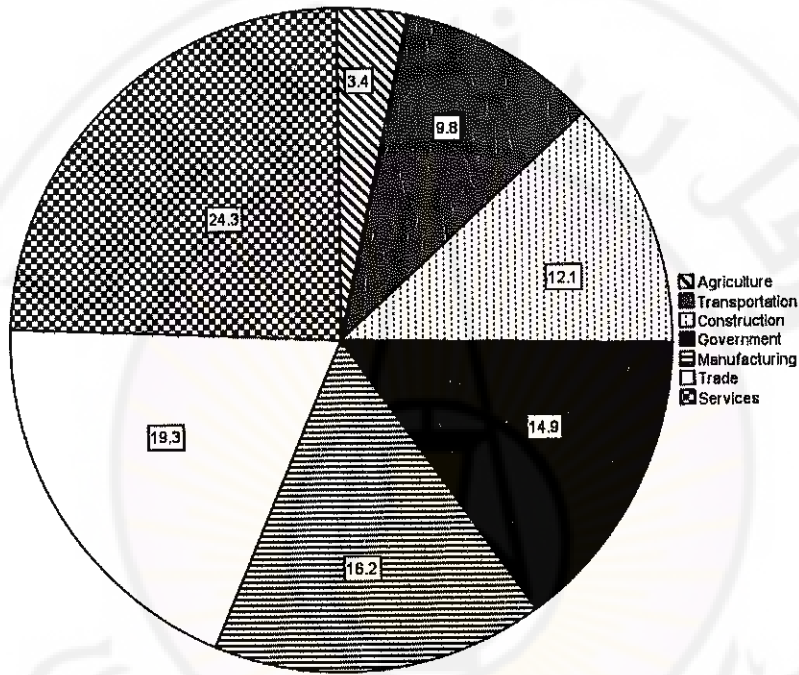
فئة الدخل	الراتب	فئة الدخل	الراتب
1	أقل من 20,000	1	أقل من 20,000
2	20,000 إلى 30,000	2	20,000 إلى 29,999
3	30,000 إلى 40,000	3	30,000 إلى 39,999
4	40,000 إلى أكثر	4	40,000 إلى أكثر

بعد ترتيب البيانات على شكل فئات و بطريقة تخلو من الغموضيّة فإن هناك العديد من الطرق لعرض البيانات بشكل مخططات. الطريقة الأولى و ربما الأسهل هي طريقة المخطط الدائري Pie chart. يستخدم المخطط الدائري لعرض النسب المئوية لعدد القياسات التي تقع في فئة معينة حيث تظهر كل فئة على شكل قطاع من الدائرة مساحته تتناسب مع النسبة المئوية للقياسات في الفئة. على سبيل المثال ، لندرس البيانات في الجدول التالي الذي يبين عدد العمال الذين أصيبوا بعجز نتيجة إصابة عمل في مختلف الوظائف في الجمهورية العربية السورية.

قطاع العمل	عدد حالات العجز	النسبة المئوية
الزراعي	130	3.4
الإنشاءات	470	12.1
الصناعة	630	16.2
النقل	300	9.8
التجارة	380	19.3
الخدمات	750	24.3
الحكومة	580	14.9

على الرغم من أنه يمكن قراءة المعلومات في الجدول و التعرف على الأعمال الأكثر خطورةً (بدلالة عدد الإعاقات) إلا أن المخطط الدائري يظهر البيانات بشكل أكثر

وضوحاً. من المخطط البياني لبيانات الجدول يمكن بسهولة أن نقوم ببعض الاستنتاجات. على سبيل المثال ، يمكن بسهولة ملاحظة أن نسبة الاصابات في قطاع الخدمات Services هي الأكبر (تشكل تقريباً ربع الدائرة) بينما نسبة الاصابات في قطاع الزراعة Agriculture هي الأقل.



لكن يجب الحذر بشكل عام عند القيام باستنتاجات من خلال بيانات الجداول أو المخططات البيانية. النسب المئوية المبينة في الجدول و المخطط لا تأخذ بعين الاعتبار عدد العاملين في القطاعات المشار إليها. لكي نحدد بشكل أكثر واقعية مقدار الخطورة في كل قطاع من القطاعات لابد من الأخذ بعين الاعتبار عدد العاملين في كل قطاع. على سبيل على الرغم من أن عدد إصابات العمل التي تقود إلى عجز هي الأكبر في قطاع الخدمات كما تشير البيانات إلا أن قطاع الخدمات هو القطاع الأكبر من حيث عدد العمال. إذا أخذنا ذلك بعين الاعتبار ، فإن نسبة العاجزين نتيجة الاصابات في

قطاع الخدمات يكون هو الأقل. المغزى هنا أنه من الضروري أن نتفحص النتائج المعروضة بشكل دقيق قبل محاولة الوصول إلى استنتاجات حولها.

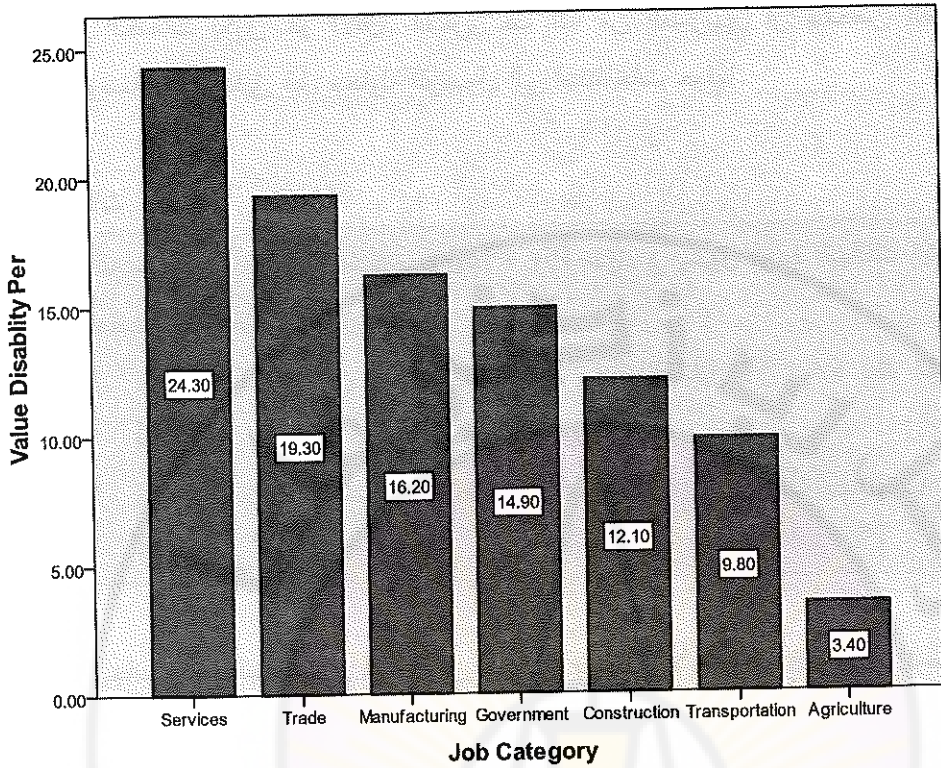
إرشادات

على الرغم من أن المخطط الدائري Pie Chart سهل و يسهل قراءة البيانات إلا أنه من المفيد الأخذ بعين الاعتبار التالي.

1. اختر عدداً قليلاً من الفئات (خمس أو ستة) للمتغير لأن زيادة عدد الفئات أكثر يجعل المخطط الدائري صعب التفسير.
2. متى كان ممكناً ، أشئ المخطط الدائري بحيث تظهرالنسب بالترتيب التصاعدي أو التنازلي.

بجانب المخطط الدائري فإنه يمكن استخدام مخطط الأعمدة Bar Chart لوصف بيانات لمتغير واحد. يمثل المخطط التالي مخطط الأعمدة لنفس البيانات في الجدول السابق. يمكن قراءة المعلومات في هذا المخطط بنفس الطريقة التي قرأنا فيها المخطط الدائري. ارتفاع العمود الخاص بفئة معينة يتناسب مع النسبة المئوية لتلك الفئة. على سبيل المثال ، إن العمود الذي يمثل قطاع الخدمات Services هو الأطول مما يدل على أن قطاع الخدمات يمتلك النسبة الأعلى لإصابات العمل التي تقود إلى عجز. بالمثل ، قطاع الزراعة يمتلك النسبة الأصغر لعدد إصابات العمل التي تقود إلى عجز لأن العمود الممثل لها هو الأقصر.

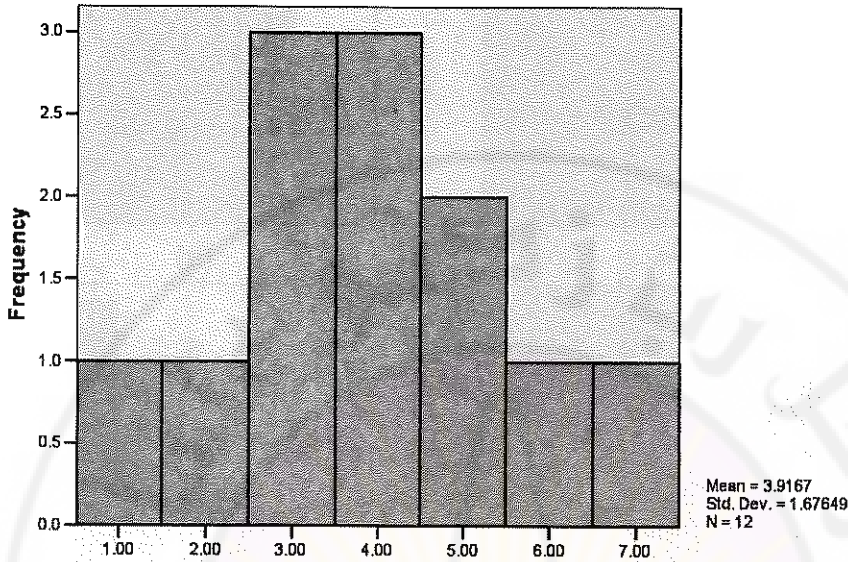
كما أشرنا سابقاً ، علينا أن نكون حذرين من القيام باستنتاجات غير صحيحة بناءً على المعلومات المبينة في مخطط الأعمدة. ربما من غير الصواب الاستنتاج من خلال مخطط الأعمدة أن قطاع الخدمات هو القطاع الأخطر من ناحية العمل و ذلك لأن عدد الإصابات في هذا القطاع هي الأكبر.



2.2.4 المدرج التكراري Histogram

المدرج التكراري هو شكل من أشكال عرض البيانات. حيث يتم تقسيم البيانات إلى مجالات متساوية العرض ، و من ثم تُحسب عدد قياسات العينة الواقعة في كل مجال (أي التكرارات ضمن المجال). هذه التكرارات تترجم بيانياً إلى أعمدة ارتفاعها يتناسب مع عدد التكرارات في كل مجال. يُظهر الشكل (1) مثالاً لمدرج تكراري. المحور الأفقي يمثل مجالات البيانات و المحور الشاقولي يمثل تكرارات القياسات في هذه المجالات. لاحظ أن البيانات مقسمة إلى مجالات عرض الواحد منها 1.5 ، و أن ارتفاع كل عمود يتناسب مع عدد التكرارات في ذلك المجال. على سبيل المثال ، ارتفاع العمود الأول المقابل للمجال [0.0, 1.5] يساوي 1 لأن هناك قياساً واحداً فقط في العينة ينتمي إلى هذا المجال.

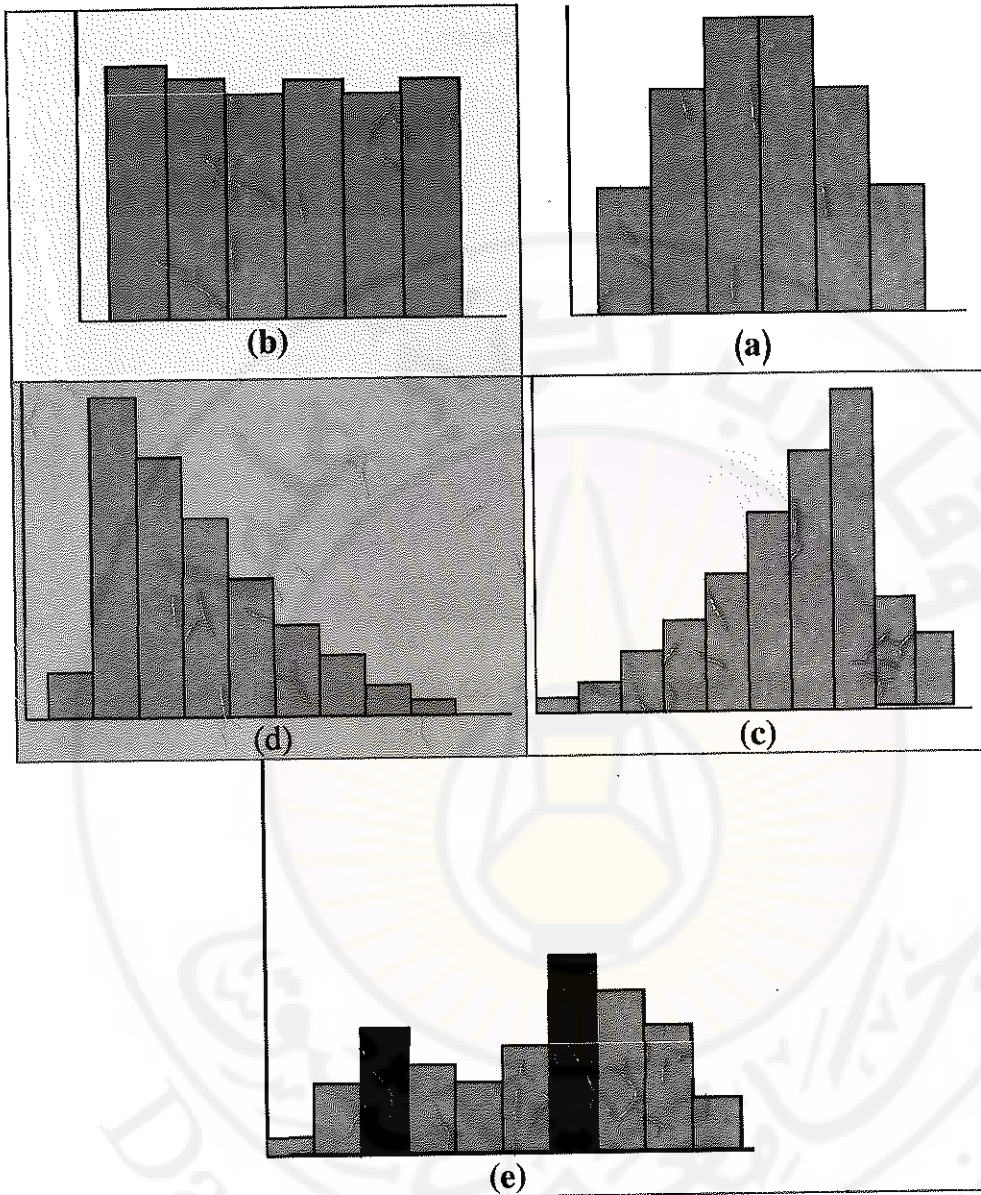
Histogram



الشكل (1): مخطط المدرج التكراري Histogram.

يفيد المخطط التكراري في دراسة شكل توزيع العينة. حيث إنه من خلال المقارنة المباشرة لشكل المدرج التكراري مع شكل توزيع نظري يمكن أن نأخذ فكرة عن شكل توزيع مجتمع العينة. على سبيل المثال ، المدرج التكراري في الشكل (1) قريب من شكل التوزيع الطبيعي (الشكل الجرسى) ، و بالتالي يمكن القول إن هذه العينة مسحوبة من مجتمع ربما يتوزع توزيعاً طبيعياً. (لاحظ أننا نقول إن مجتمع العينة ربما يتوزع توزيعاً طبيعياً و هذا أدق من قولنا إن العينة تتوزع وفقاً للتوزيع الطبيعي.)

يبين الشكل (2) مجموعة من المدرجات التكرارية. عندما يكون للمدرج التكراري ذروة واحدة عندئذٍ يسمى نموذجاً أحادياً Unimodal (انظر الأشكال (a-d)). و عندما يكون له ذروتان يسمى عندئذٍ نموذج ثنائي Bimodal (انظر الشكل (e)). في كثير من الأحيان وعندما يكون المدرج التكراري ثنائي الذروة فإن ذلك يشير إلى البيانات تنتمي إلى مجتمعين مختلفين.



الشكل (2): أشكال مختلفة لتوزيع العينات كما تظهر في المدرج التكراري.
 يكون المدرج التكراري متناظراً إذا كان الطرف الأيمن و الأيسر لهما نفس الشكل.
 الأشكال (a, b) تمثل مدرجاً تكرارياً متناظراً و بالتالي توزيع العينة متناظر. عندما يشمل
 الطرف الأيمن على عدد أكبر من القياسات يمتد المدرج التكراري بشكل أطول بالاتجاه

الأيمن و يدعى عندئذٍ المدرج التكراري بملتوي نحو اليمين **Right-Skewed**. الشكل (d) يمثل مدرجاً تكرارياً ملتوياً نحو اليمين. يكون المدرج ملتوياً نحو اليسار **Left-Skewed** عندما يمتد أكثر باتجاه اليسار. الشكل (c) يمثل مدرجاً تكرارياً ملتوياً نحو اليسار.

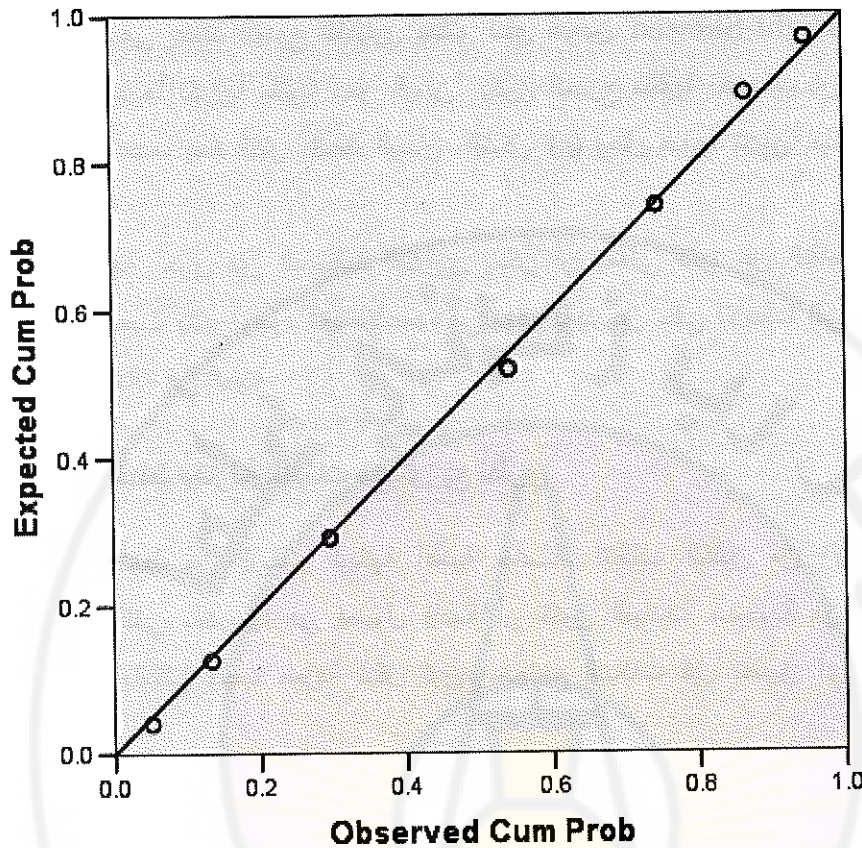
من جهة أخرى فإن المخطط التكراري يمكن أن يعطي فكرة عن النقاط الشاذة **outliers**. حيث يشير وجود أحد الأعمدة في موقع بعيد عن باقي الأعمدة إلى احتمال وجود نقاط شاذة في العينة.

2.2.4 مخطط مقارنة العينات مع توزيع نظري

لهذا النمط من الرسوم البيانية أهمية خاصة. تتطلب الكثير من الاختبارات الإحصائية التي سندرسها في فصول قادمة أن تكون البيانات مسحوبة من مجتمع طبيعي. و بالتالي يجب معرفة شكل توزيع مجتمع البيانات قبل تطبيق أي اختبار إحصائي.

هناك مخططان بيانيان يمكن أن يُستخدما لمعرفة فيما إذا كانت البيانات (أي العينة) تتوافق مع توزيع نظري معين. يُدعى المخطط الأول **Normal Probability Plots** أو اختصاراً يرمز بـ **P-P plot** و يُدعى المخطط الآخر **Quantile-Quantile plot** أو اختصاراً يرمز بـ **Q-Q plots**. على الرغم من أن هذين المخططين يستخدمان أسلوبين مختلفين لإظهار فيما إذا كانت العينة تتفق مع توزيع نظري معين ، فإن المخططات الناتجة عنهما متشابهة و تُفسر بالطريقة نفسها.

سوف نعرض في هذه الفقرة طريقة رسم المخطط **P-P plot** و كيفية قراءة المعلومات المبينة فيه. المخطط الآخر **Q-Q plot** سوف نشرحه في مثال لاحق. يبين الشكل (2) المخطط النموذجي لـ **P-P plot** الذي يرسم البيانات في محورين أفقي و شاقولي. يمثل المحور الأفقي الاحتمالات المشاهدة من خلال بيانات العينة بينما يمثل المحور الشاقولي الاحتمالات المتوقعة من خلال توزيع نظري.



الشكل (2): المخطط P-P الذي يبين التوافق بين التوزيع الطبيعي النظري و توزيع العينة.

الخط المستقيم يمثل الرسم البياني لنقاط مولدة عشوائياً من توزيع نظري (في الشكل (2) التوزيع النظري هو التوزيع الطبيعي). النقاط (الدوائر الصغيرة) تمثل النقاط الفعلية (أي نقاط العينة). يفسر هذا المخطط على النحو التالي:

إذا كانت العينة تتوزع وفق التوزيع النظري فإن نقاط العينة يجب أن تنطبق على المستقيم أو تتجمع (تتراص) حوله. وكلما ابتعدت النقاط عن المستقيم كلما ضعف التوافق ما بين التوزيع النظري و توزيع مجتمع العينة.

من الواضح أنه في الشكل (2) هناك توافق شبه تام بين التوزيع الطبيعي و توزيع مجتمع العينة لأن النقاط تكاد تنطبق على المستقيم (في الواقع ، بعض هذه النقاط ينطبق فعلياً على المستقيم). و بالتالي نستطيع الاستنتاج بأن العينة مسحوبة من مجتمع طبيعي.

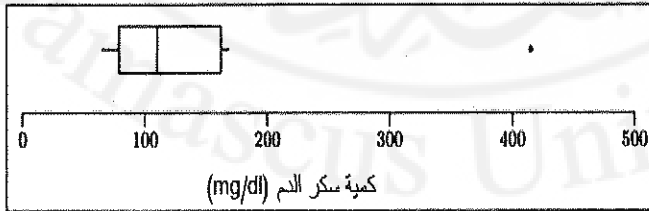
3.2.4 العرض الصندوقي Box Plots

عندما تكون البيانات تتعلق بمتغير عشوائي واحد فإن أول ما يخطر للذهن هو استخدام ال Histogram أو مخطط الأعمدة. أحد المشاكل الواضحة لهذا المخطط هو أنه يُمكن أن يشتمل على الكثير من المعلومات التي لا يمكن استنتاجها بسهولة. المشكلة الثانية ، و ربما الأهم ، هي أن مخطط الأعمدة غير مناسب لمقارنة متغيرين. يتميز ال Box Plots أو العرض الصندوقي بأنه يستطيع إظهار العديد من الإحصائيات الخاصة بمتغير بشكل واضح و قابل للتفسير بسهولة.

بشكل عام يُظهر المخطط الصندوقي الوسط Median ، الربيعيات Quartiles ، المدى الذي تغطيه البيانات ، و القيم الشاذة Outliers إذا كانت بالطبع موجودة. من الميزات المهمة لهذا الشكل من العرض هو أنه يُعطي صورة واضحة عن الإحصائيات التي ذكرناها. كما أنه يعطي صورة واضحة جداً عن تناظر البيانات أو عدمه. لنكن لدينا البيانات التالية التي تعبر عن كمية سكر الدم (mg/dl) عند عينة من المرضى.

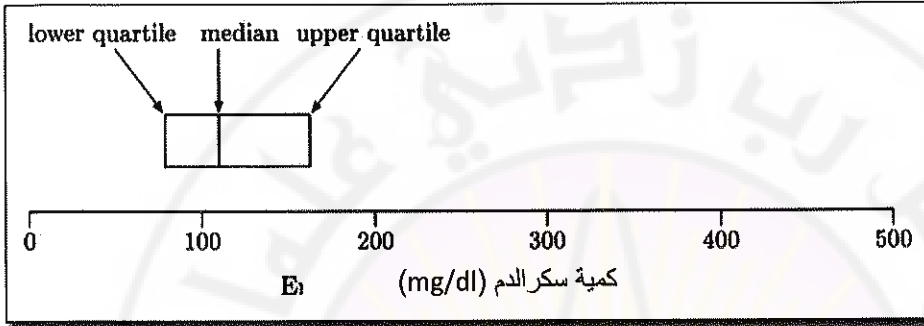
66 72 79 84 102 110 123 144 162 169 414

المخطط الصندوقي لهذه البيانات مُعطى بالشكل (3).



الشكل (3): المخطط الصندوقي Boxplot للبيانات أعلاه.

أفضل طريقة لفهم هذا المخطط الصندوقي هو أن نقوم بإنشائه خطوة بخطوة. لإنشاء الصندوق (المستطيل) نحتاج إلى حساب كل من الوسط و الربيعي الأول و الربيعي الثالث. إنَّ وسط البيانات هو 110 والربيعي الأول هو 79 و الربيعي الثالث هو 162.



الشكل (4): جزء من المخطط الصندوقي للبيانات أعلاه.

هذه المعلومات كافية لرسم المستطيل كما هو مبين بالشكل (4) حيث تقع حدود المستطيل على الربيعي الأول lower quartile و الربيعي الثالث upper quartile أما الوسط median فهو ممثل بالخط الشاقولي داخل المستطيل. ليس من الصعب أن تستنتج أن المستطيل يغطي الـ 50% من القيم الواقعة في الوسط لأن مقدار البيانات بين الربيعي الأول و الثالث هو نصف البيانات.

تغطي المستقيمات الممتدة من طرفي المستطيل، التي تسمى *Whiskers*، البيانات إلى أبعد بيان غير شاذ في كل طرف. لتحديد البيانات الطرفية غير الشاذة ، فإننا نقوم بحساب المدى الربيعي. كما نعلم فإنَّ المدى الربيعي يساوي ناتج طرح قيمة الربيعي الأول من قيمة الربيعي الثالث. في مثالنا الحالي، فإن المدى الربيعي يساوي 162 - 79 = 83. تُرسم المستقيمات من كل طرف من أطراف المستطيل لتصل إلى أبعد قيمة تدعى القيمة المجاورة (غير الشاذة) أو *Adjacent value*. أكبر قيمة مجاورة - غير شاذة - (*Upper adjacent value*) هي التي تبعد بمقدار 1.5 من المدى الربيعي الثالث

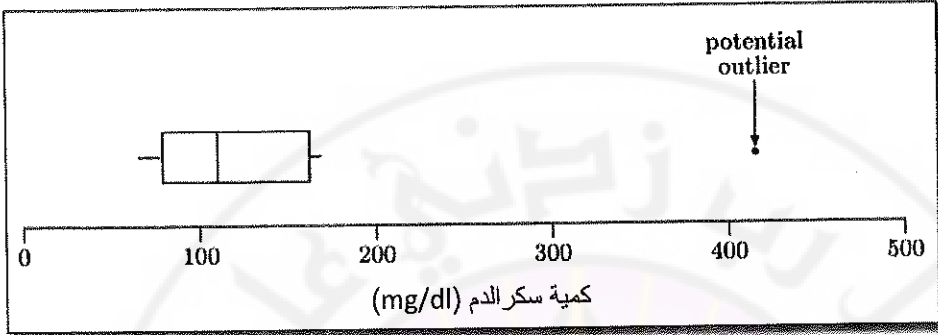
(أي من النهاية اليمينية للمستطيل). أصغر قيمة مجاورة -غير شاذة- (Lower adjacent value) هي التي تبعد بمقدار 1.5 من المدى الربيعي الأول (أي من النهاية اليسارية للمستطيل).

في المثال الحالي أكبر قيمة مجاورة -غير شاذة- يجب أن لا تتجاوز القيمة المحددة بالتعبير التالي $286.5 = 83 * 1.5 + 162$ ، و بالتالي فإن 169 هي أكبر قيمة مجاورة -غير شاذة- في البيانات (لأن القيمة 169 هي أكبر قيمة أصغر من 286.5). على هذا فإن المستقيم اليميني يمتد من الربيعي الثالث إلى أكبر قيمة غير شاذة (169). وبالمثل ، فإن أصغر قيمة مجاورة -غير شاذة- يجب أن لا تقل عن $79 - 1.5 * 83 = -45.5$ ، و بالتالي فإن أصغر قيمة مجاورة -غير شاذة- هي 66. على هذا فإن المستقيم اليساري يمتد من الربيعي الأول إلى أصغر قيمة غير شاذة (66). بما أن القيمة 414 هي أكبر من أكبر نقطة مجاورة -غير شاذة- ، فإن القيمة 414 تُصنف على أنها نقطة شاذة (Outlier). الشكل (5) يمثل المخطط الصندوقي للبيانات أعلاه.

كما هو واضح من الشكل (5) فإن المخطط الصندوقي يُعطي معلومات واضحة عن البيانات. طول المستطيل يمثل المدى الربيعي و طول المستقيمين الممتدين على يمين المستطيل و يساره يمثلان انتشار البيانات. على هذا فإن المخطط الصندوقي يعطي فكرة واضحة عن تباين البيانات. كلما كانت المستقيمتان تمتد بعيداً عن حدود المستطيل كلما كان تباين العينة أكبر. يظهر المخطط الصندوقي كذلك البيانات الشاذة. على سبيل المثال، البيان 414 هو شاذ، لذلك فإن المخطط الصندوقي يعلمه على أنه Potential outlier.

يعطي المخطط الصندوقي فكرة أيضاً عن تناظر البيانات. إذا كانت البيانات متناظرة فإن المخطط الصندوقي سيكون متناظراً بالنسبة إلى الوسط. نلاحظ في المثال أعلاه أن البيانات غير متناظرة بسبب كون الوسط يميل أكثر باتجاه يسار المستطيل وبالتالي فإن الجزء الأيمن من المستطيل أكبر من الجزء الأيسر ، مما يعني أن البيانات ملتوية نحو اليمين (Right-skewed data). جدير بالذكر أن عدم تناظر المستقيمتان على جانبي المستطيل (أي أحدهما أطول من الآخر) يشير أيضاً إلى عدم تناظر

البيانات. على سبيل المثال، إذا كان المستقيم الأيمن أطول من المستقيم الأيسر، فإنه من المحتمل أن تكون البيانات ملتوية نحو اليمين.



الشكل (5): المخطط الصندوقي للبيانات في المثال أعلاه.

أحياناً يحصل لدينا تناقض ما بين جهة الالتواء التي يشير إليها موضع الوسط بالنسبة إلى المستطيل و جهة الالتواء التي تشير إليها المستقيمتان. على سبيل المثال ، الوسط يميل أكثر باتجاه النهاية اليسرى للمستطيل ، بينما المستقيم الأيمن يمتد أكثر من الأيسر. والسؤال: أين هي جهة الالتواء؟ ليس من السهل أحياناً أن نحل هذا التناقض اعتماداً على المخطط الصندوقي فقط بل نحتاج إلى معلومات إضافية سوف نناقشها لاحقاً في هذا الفصل.

اختصاراً، من خلال معاينة المخطط الصندوقي في الشكل (5) نجد الخصائص التالية للبيانات:

1. البيانات تحوي نقطة شاذة و هي 414 كما هو موضح بالشكل.
2. البيانات غير متناظرة حول الوسط لأن الوسط يميل باتجاه يسار المستطيل مما يدل على البيانات ملتوية نحو اليمين.
3. البيانات على يسار الربيعي الأول تنتشر بشكل أكبر من البيانات على يمين الربيعي الثالث.

4.2.4 مقارنة البيانات باستخدام المخطط الصندوقي Box plots

يستخدم المخطط الصندوقي أيضاً من أجل مقارنة العينات. نستطيع من خلال المقارنة أن نستعرض خصائص العينتين. إعطاء مثال يمكن أن يوضح الفكرة بشكل أفضل. تمثل البيانات في الجدول التالي وزن الأطفال المصابين بمتلازمة التضيق التنفسي SIRS عند الولادة. البيانات المعلمة بـ "*" تعني أن الطفل قد توفي نتيجة المرض. السؤال الذي نريد أن نجيب عليه من خلال المخطط الصندوقي هو: هل نستطيع ربط إمكانية النجاة من المرض مع وزن الطفل عند الولادة؟ للإجابة عن هذا السؤال نقوم برسم المخطط الصندوقي للبيانات. يظهر الشكل (6) المخطط الصندوقي لأوزان الأطفال الناجين و الضحايا عند الولادة.

1.050*	2.500*	1.890*	1.760	2.830
1.175*	1.030*	1.940*	1.930	1.410
1.230*	1.100*	2.200*	2.015	1.715
1.310*	1.185*	2.270*	2.090	1.720
1.500*	1.225*	2.440*	2.600	2.040
1.600*	1.262*	2.560*	2.700	2.200
1.720*	1.295*	2.730*	2.950	2.400
1.750*	1.300*	1.130	3.160	2.550
1.770*	1.550*	1.575	3.400	2.570
2.275*	1.820*	1.680	3.640	3.005

الجدول (2): أوزان (بالكيلو غرام) الأطفال المصابين بمتلازمة التضيق التنفسي SIRS.

يصف المخطط الصندوقي في الشكل (6) البيانات في الجدول (2) بدلالة مركز البيانات (Median)، التشتت (Dispersion)، الالتواء (Skewness)، كما يعطي فكرة عن القيم الشاذة (Outliers) إن وجدت. بمقارنة وسط العينتين نجد أن وسط أوزان الولادة للأطفال الذين ماتوا (died) أصغر من وسط أوزان الولادة للأطفال الذين نجوا

(survived). الأمر الذي يوحي بأنه كلما زاد وزن الطفل المريض ازدادت فرصته بالنجاة.

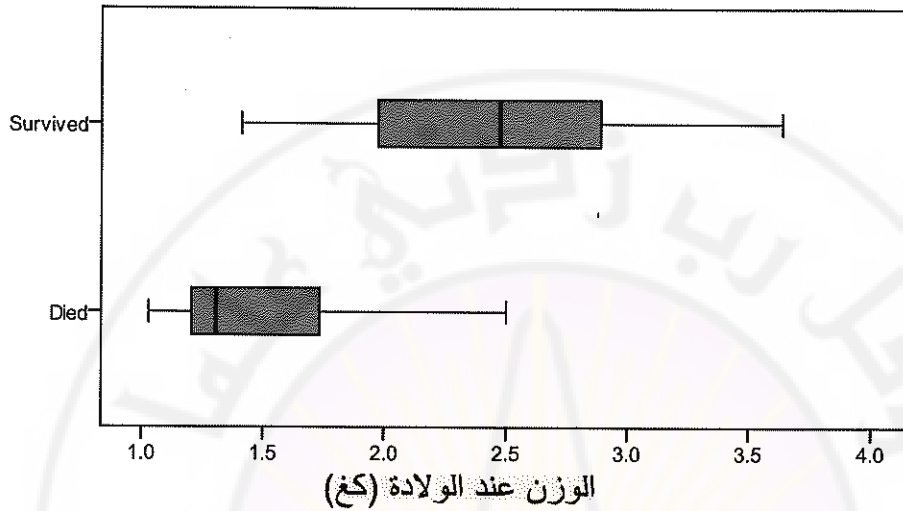
يبين الشكل (6) أيضاً انتشار البيانات لكلا العينتين. من خلال مقارنة طولي الصندوقين (المدى الربيعي) نجد أن انتشار البيانات في كلا العينتين متساو تقريباً لأن طولي الصندوقين متقارب. و لكن مقياس المدى يبين انتشار عينة أوزان الأطفال الذين نجوا أكبر من انتشار عينة أوزان الأطفال الذي ماتوا و يتضح ذلك من خلال المسافة بين طرفي المستقيمين. علاوةً على ذلك فإن مجموعتي البيانات تظهران التواء نحو اليمين على الرغم من أن هذا الالتواء يبدو أكثر وضوحاً في البيانات الخاصة بأوزان الأطفال الذين ماتوا لأن خط الوسط يبتعد أكثر عن وسط الصندوق. هذا الالتواء يبدو كبيراً لأن خط الوسط في كلتا الحالتين يبتعد عن وسط كل صندوق. و بالتالي يمكن القول إن بيانات الأطفال الذين ماتوا تفقر إلى التناظر أكثر من بيانات الأطفال الناجين. لا يُظهر الشكل (6) أية نقاط شاذة في البيانات.

يمكن أن نستنتج من خلال هذه المناقشة أن البيانات الخاصة بأوزان الأطفال سواء الناجين أو غير الناجين تتوزع بنفس الطريقة تقريباً. كذلك يمكن أن نرى بسهولة من خلال الشكل (6) أن وسط أوزان الأطفال الناجين أعلى من وسط الأطفال غير الناجين. و بالتالي يمكننا القول بثقة إن النجاة تتعلق بوزن الطفل المريض.

عند مقارنة مخططات صندوقية Box plots فإن الإرشادات التالية مفيدة.

1. قارن بين وسط العينات لتحديد مراكز هذه العينات.
2. قارن تشتت البيانات من خلال المدى الربيعي (طول الصندوق)، و من خلال المدى (البعد بين الخططين الواصلة بين أكبر و أصغر قيمة).
3. ابحث عن وجود التواء في البيانات، و يتم ذلك من خلال موقع الوسط بالنسبة للمستطيل، و أيضاً من خلال طول المستقيمات على يمين الصندوق و يساره.
4. ابحث عن وجود قيم شاذة.

كتابة أي تقرير حول نتائج المقارنة بين مخططين صندوقين أو أكثر يجب أن يتضمن على الأقل النقاط الأربعة الواردة في الإرشادات أعلاه.



الشكل (6): المخطط الصندوقي للبيانات في الجدول (2).

5.2.4 مخطط الأخطاء Error Bars

يستخدم مخطط الأخطاء لإظهار بيانياً المتوسطات و مجال ثقة 95% أو أي ثقة أخرى حول المتوسطات. لذلك يعد هذه المخطط مفيداً جداً عندما نريد:

1. مقارنة متوسطات عدة عينات بيانياً.
2. إظهار مجالات الثقة حول المتوسطات بيانياً.
3. إظهار مدى تباين القياسات حول المتوسط و يظهر ذلك بوضوح من خلال عرض مجالات الثقة.

سوف نأخذ البيانات التالية لتوضيح فكرة Error Bars. يبين الجدول التالي ، عدد الإصابات في أربع قطاعات عمل خلال الخمسة أعوام الأخيرة.

قطاع الخدمات	قطاع الزراعة	قطاع الصناعة	قطاع الحرف اليدوية
Service	Agriculture	Industry	Handcraft
48	12	74	52
66	30	82	58
74	25	72	82
86	18	80	35
78	26	79	84

ادرس هذه البيانات باستخدام مخطط Error Bars.

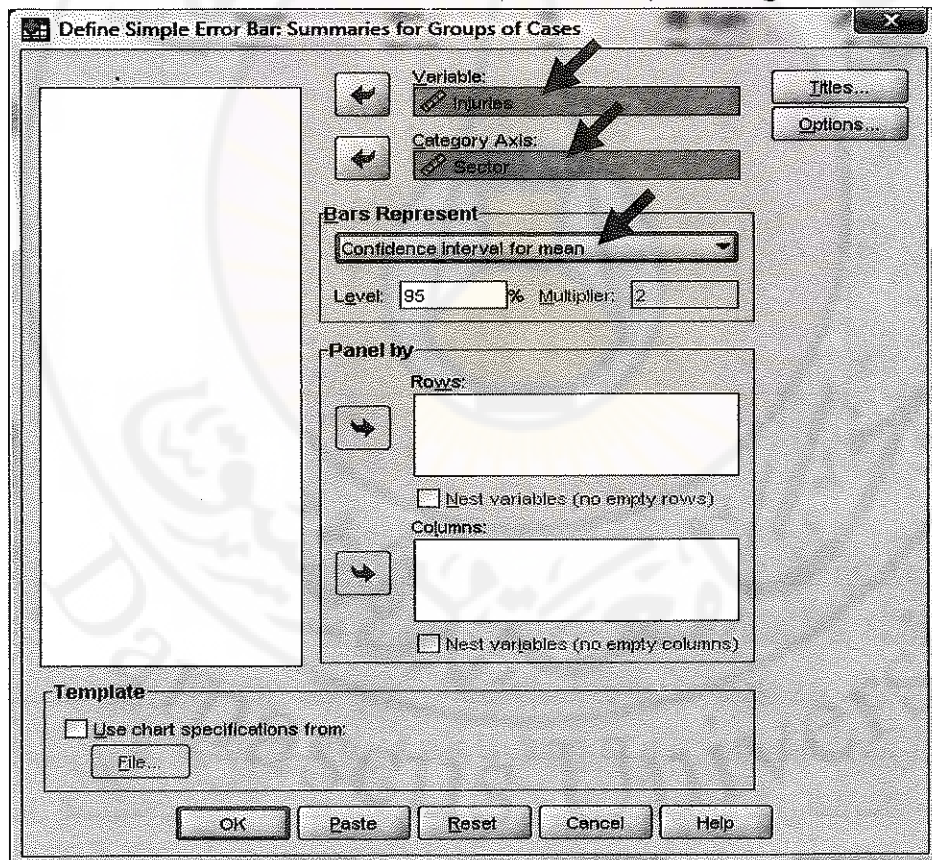
إدخال البيانات

توضع البيانات في عمود منفصل و يُعرف عمود مرجعي يبين مرجعية البيانات في عمود البيانات. يبين الشكل التالي شاشة SPSS بعد إدخال البيانات بالشكل الصحيح. العمود Injuries يحوي عدد الإصابات السنوية للقطاعات المشمولة بالدراسة بينما يحوي العمود Sector على القيم المرجعية 1 إلى 4 حيث 1 يرمز لقطاع الخدمات Service و 2 يمثل الحرف اليدوية Handcraft ، و هكذا.

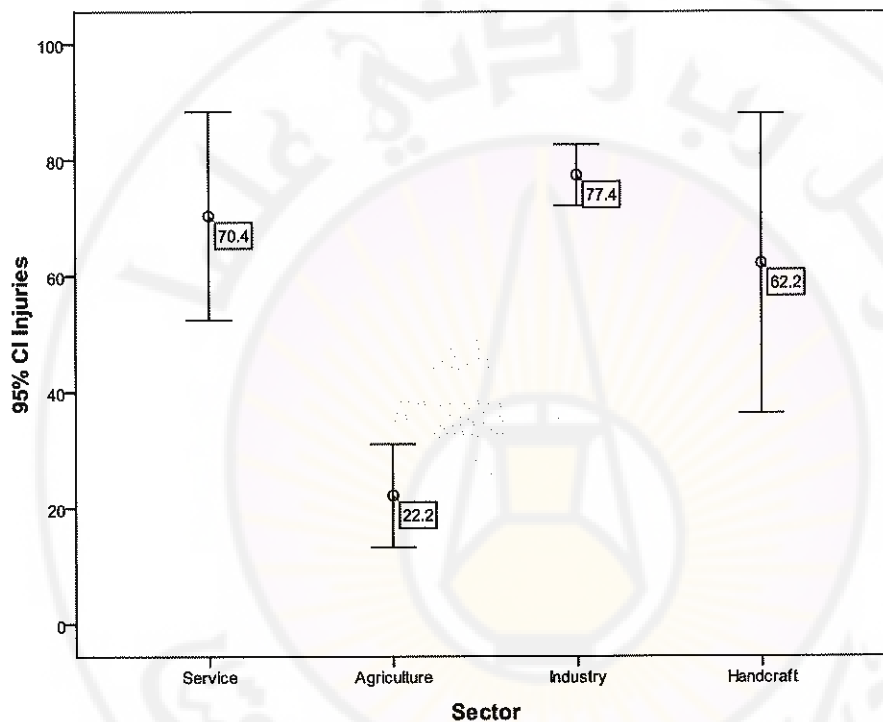
	Injuries	Sector
1	52	1
2	58	1
3	82	1
4	35	1
5	84	4
6	48	1

يمكن تفعيل أمر Error Bars ، بالنقر على القائمة Graphs ، اختيار Legacy Dialog ، من ثم إختيار الأمر الفرعي Error Bar... يظهر مربع حوار تمهيدي ، حيث نختار منه نوع المخطط "Simple" و نختار أيضاً الخيار "Summaries for Groups of Cases". ننقر على المفتاح Define فيظهر مربع الحوار التالي الخاص بهذا الأمر. الخيارات التي نحتاجها في هذا المربع مٌشار إليها بالأسهم.

ننقل العمود الذي يحوي البيانات Injuries إلى الخانة المعنونة "Variable" و العمود المرجعي Sector إلى الخانة المعنونة "Category Axis". نُبقي الوضع الافتراضي "Confidence Interval for mean" في الخانة Bars Represent. ننقر على المفتاح Ok ليقوم SPSS برسم المخطط.



يبين مخطط الخطأ مقدار تباين القياسات عن متوسطها. كلما زاد عرض مجال الثقة زاد تباين البيانات عن المتوسط وقلت ثقتنا بالمتوسط كمقدر. تشير الدائرة في وسط المستقيمات إلى المتوسط. تمثل المستقيمات على جانبي الدائرة عرض مجال الثقة. الأرقام التي تظهر بجانب دائرة المتوسط هي قيمة المتوسط.

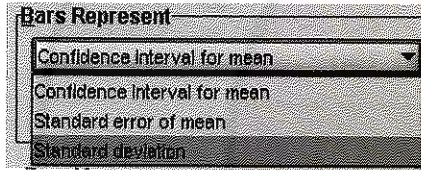


كما يبدو من الشكل فإن أكثر القياسات تشتتاً عن المتوسط هي عدد الإصابات الناتجة عن الحرف اليدوية Handcraft ، تليها عدد الإصابات الناتجة عن قطاع الخدمات Service. يشير ذلك إلى أن أعداد الإصابات تختلف بشكل كبير من سنة لأخرى. إن أقل القياسات تشتتاً هي الناتجة عن قطاع الصناعة Industry. لاحظ أن طول المستقيمات حول المتوسط صغيرة مما يشير إلى تجمع القياسات حول المتوسط.

بالنظر إلى مخطط Error Bar ، نلاحظ أن متوسطات الإصابات في جميع القطاعات ، ماعدا قطاع الزراعة Agriculture ، غير مختلفة معنوياً بثقة 95% لأن مجالات الثقة حول المتوسطات متداخلة. بيد أن مجال الثقة حول متوسط عدد الإصابات في قطاع الزراعة هو الأقل من حيث القيمة و لا يتداخل مع أي مجال آخر و بالتالي هذا المتوسط يختلف معنوياً عن باقي المتوسطات. هذا يشير إلى أن قطاع الزراعة هو الأقل خطورة لأن متوسط عدد الإصابات هو الأقل بين القطاعات الأخرى.

هذا الاستنتاج يكون صحيحاً فقط إذا كان أعداد العاملين في هذه القطاعات متساوياً تقريباً. أما إذا كان عدد العاملين في هذه القطاعات مختلفاً بشكل كبير فإن هذا الاستنتاج قد لا يكون دقيقاً.

ملاحظة: يمكن استخدام مربع الحوار الخاص بمخطط Error Bars لرسم مخططات



للخطأ المعياري في المتوسط Standard error of mean و للانحرافات المعيارية Standard deviation. يتم ذلك

من خلال النقر على سهم القائمة المنسدلة المعنونة Bars Represent لتفتح الخيارات الممكنة و المبينة في الشكل الجانبي.

6.2.4 مخطط الساق و الورقة Stem and Leaf Plot

يُعد مخطط الساق (Stem) و الورقة (Leaf) أحد أشكال عرض البيانات الفعالة التي تعطي فكرة واضحة عن شكل توزيع العينة و عن كل قياس من قياسات العينة. لعرض البيانات بطريقة الساق و الورقة نقوم بترتيب البيانات بحسب مرتبة الرقم في كل قياس. الرقم ذو أعلى ترتيب يسمى الساق و الأرقام ذوات الترتيب الأقل تسمى الورقة (أو الأوراق).

الساق و الورقة هي وسيلة عرض مناسبة جداً عندما تكون العينة تحوي كمية كبيرة من الأرقام. مجموعة كبيرة من نتائج مباريات كرة قدم ، مجموعة كبيرة من درجات

الحرارة أو كميات الأمطار المسجلة على فترة طويلة من الزمن ، كمية كبيرة من نتائج الطلاب في الجامعة هي بضعة أمثلة على البيانات التي يمكن أن تحوي كمية كبيرة من الأرقام.

يمثل الشكل (7) مخطط الساق و الورقة لعلامات مجموعة من الطلاب في مقرر تحليل البيانات. كما يتضح من الشكل أن الساق (Stem) تمثل العشرات و الورقة (Leaf) تمثل الواحدات. بالنظر إلى الشكل نستطيع ببساطة الحصول على كمية كبيرة من المعلومات حول علامات الطلاب. نستطيع أن نفهم أن هناك أربعة طلاب حصلوا على علامة أكثر من 90 ، إثنان منهم حصل على 92 ، وأن طالباً واحداً فقط حصل على علامة 98 ، و هي أعلى علامة. علاوةً على ذلك و من خلال العودة إلى الشكل (7) يمكن أن نلاحظ أنه لا يوجد أي طالب حصل على علامة أقل من 50.

Test Scores Out of 100	
Stem	Leaf
9	2 2 6 8
8	3 5
7	2 4 6 8 8 8
6	1 4 4 7 8
5	0 0 2 8 8

الشكل (7): الساق و الورقة لعلامات الطلاب

يمكن من خلال الساق و الورقة أن نحدد عدد أفراد العينة و ذلك بالقيام بعد الأوراق. كذلك يمكننا بسهولة إيجاد المنوال حيث يقابل المنوال القيمة الأكثر تكراراً في الورقة. على سبيل المثال نجد أن الـ 8 في الورقة المقابلة للساق 7 هي القيمة الأكثر تكراراً (تكررت ثلاثة مرات). و بالتالي فإن المنوال لهذه القياسات هو 78.

يمكننا أيضاً من خلال الساق و الورقة أن نتعرف على شكل التوزيع (طبيعي أو غير طبيعي) ، و ذلك بمقارنة مخطط الساق و الورقة مع التوزيع النظري. على سبيل المثال، بيانات علامات الطلاب في المثال لا تتوزع بشكل طبيعي ، لأن الشكل (7) ليس قريباً من الشكل الجرسى للتوزيع الطبيعي.

يمكننا أيضاً استخدام مخطط الساق و الورقة لمقارنة عينتين. على سبيل المثال الشكل (8) يقارن بين علامات الطلاب و علامات الطالبات في مادة تحليل البيانات.

Scores		
Leaf	Stem	Leaf
طالبات		طلاب
0 3 7 9	9	2 2
2 8	8	3 5 5
1 3 9 7	7	4 6 8 8 9
8	6	6 2

الشكل (8): الساق و الورقة لمقارنة عينتين.

الساق (العشرات) تقع في الوسط و عمودي الواحدات (الورقة) يقعان على يمين الساق و يساره. بالرجوع إلى الشكل (8) نستطيع أن نرى مباشرة أن عدد الطالبات اللاتي حصلن على علامة 90 أو أكثر يفوق عدد الطلاب ، (طالبان مقابل أربع طالبات). كذلك نرى أن إحدى الطالبات حصلت على أعلى علامة (99) ، بينما أعلى علامة حصل عليها الطلاب هي (92). كذلك فإن أقل علامة حصلت عليها الطالبات هي (68) ، بينما أقل علامة حصل عليها الطلاب هي (62). أيضاً عدد الطلاب الذين حصلوا على علامة تقع بين (80) و (90) هو ثلاثة مقابل طالبتين فقط حصلن على علامات في نفس المجال. هناك معلومات أخرى تتعلق بشكل توزيع البيانات ، المنوال، إلخ. هل تستطيع استنتاج كل هذه المعلومات؟

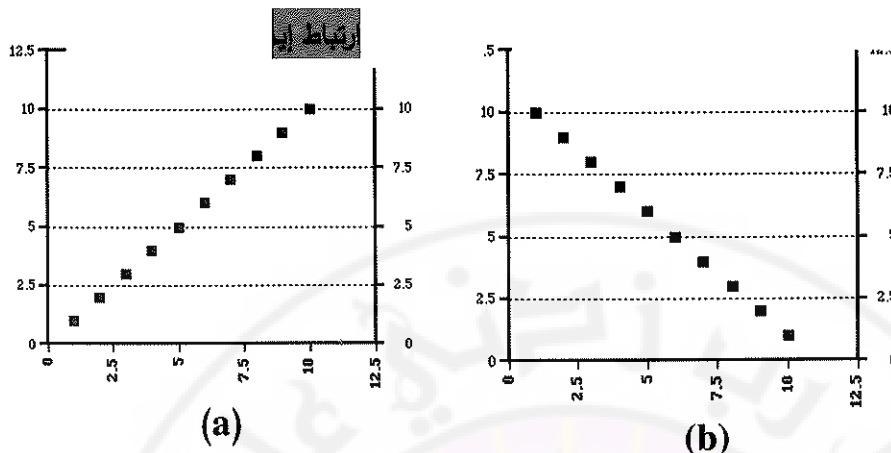
من الخصائص المهمة لمخطط الساق و الورقة نذكر:

1. يعد مخطط الساق و الورقة مفيداً جداً عندما يكون المطلوب عرض بيانات تشتمل على عدد كبير من القياسات الرقمية.
2. يُعطي مخطط الساق و الورقة فكرة واضحة عن توزيع البيانات و بالتحديد فيما إذا كانت البيانات تتوزع طبيعياً.
3. يمكن من خلال مخطط الساق و الورقة أن نوجد المنوال. من السهل أيضاً استخدام مخطط الساق و الورقة لإيجاد وسط القياسات. هل تستطيع رؤية كيف يتم ذلك؟

7.2.4 المخطط التبعثري Scatter plot

يساعدنا المخطط التبعثري على دراسة تأثير متغير عشوائي على متغير عشوائي آخر ، وذلك برسم البيانات في المستوي المحدد بالمحور x و المحور y . بكلمات أخرى، يستخدم المخطط لإظهار شكل الارتباط بين متغيرين عشوائيين. تتحدد شكل العلاقة ودرجة قوتها من خلال شكل انتشار النقاط في المستوي. كلما كان شكل انتشار البيانات أقرب إلى خط مستقيم كانت العلاقة (الارتباط) بين المتغيرين أكبر. إذا كانت البيانات تشكل خطاً مستقيماً عندئذ يكون الارتباط تاماً. و كلما ابتعدت النقاط عن تشكيل خط مستقيم ضعف الارتباط. أما إذا كان توزيع النقاط بشكل عشوائي في المستوي ، فنقول إنه لا يوجد ارتباط.

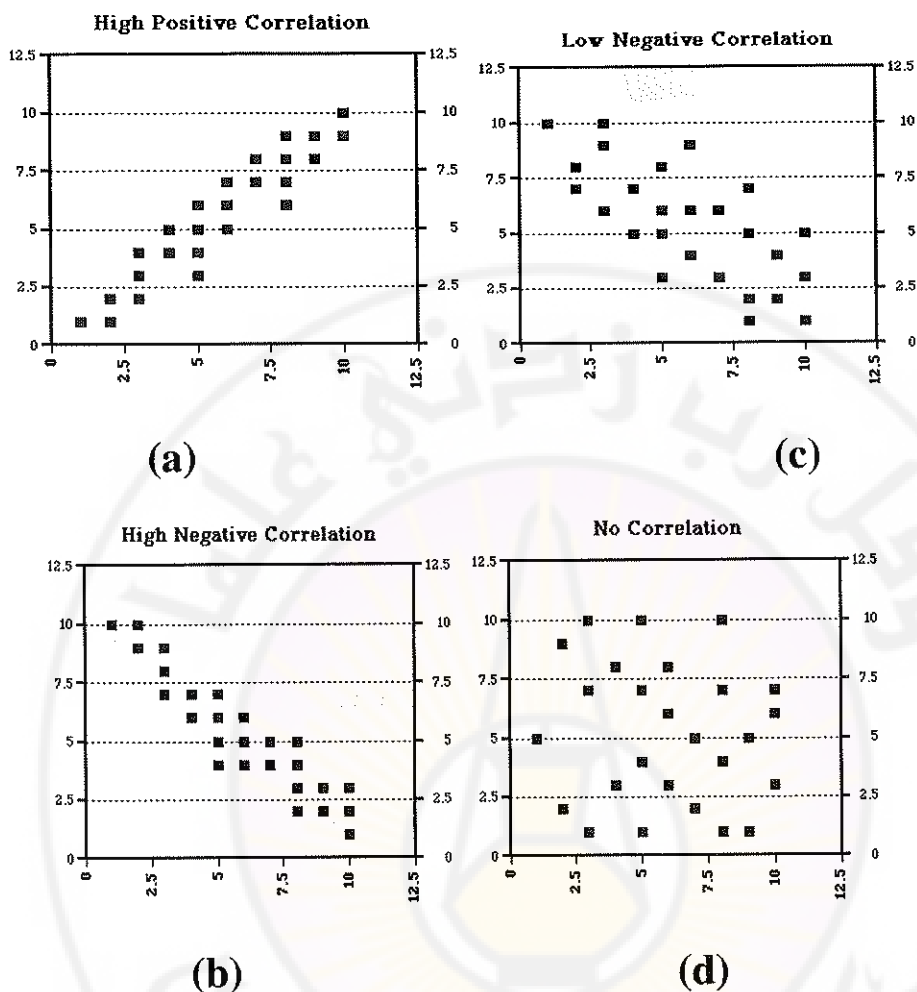
يحدد اتجاه الخط المستقيم اتجاه الارتباط. إذا كان انتشار البيانات يتجه بحيث كلما ازدادت قيم أحد المتغيرين ازدادت قيم المتغير الآخر ، فإن العلاقة تسمى علاقة موجبة. و في الحالة المعاكسة ، أي كلما ازدادت قيم أحد المتغيرين نقصت قيم المتغير الآخر ، فإننا نقول عن العلاقة أنها سالبة.



الشكل (9): المخطط التبعثري

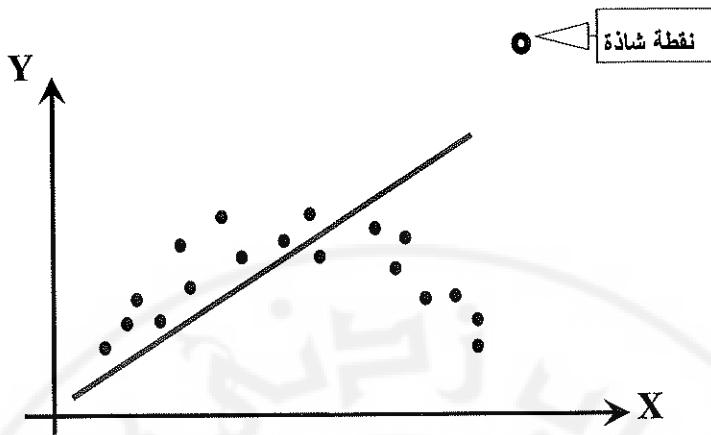
الشكل (9) يمثل ارتباطاً تاماً بين متغيرين عشوائيين لأن القيم تشكل خطاً مستقيماً. الشكل (9.a) يمثل ارتباطاً إيجابياً تاماً لأنه كلما ازدادت قيمة أحد المتغيرين ازدادت قيمة المتغير الآخر و بنفس النسبة. أما الشكل (9.b) فيمثل ارتباطاً سلبياً تاماً لأنه كلما ازدادت قيمة أحد المتغيرين نقصت قيمة المتغير الآخر و بنفس النسبة. و مع ذلك في التطبيقات العملية فإنه نادراً ما يكون لدينا ارتباط تام.

يعطي الشكل (10) صورة أكثر واقعية عن ارتباطات المتغيرات. الشكل (10.a) يمثل ارتباطاً إيجابياً لأن زيادة أحد المتغيرات تؤدي إلى زيادة في المتغير الآخر ، و هذا الارتباط قوي و لكنه غير تام ، لأن زيادة أحد المتغيرات تقود إلى زيادة في المتغير الآخر ، و لكن ليس بنفس النسبة. الشكل (10.b) يمثل ارتباطاً سلبياً لأن زيادة أحد المتغيرات تؤدي إلى نقصان في المتغير الآخر ، و هذا الارتباط قوي و لكنه غير تام ، لأن زيادة أحد المتغيرات تقود إلى نقصان في المتغير الآخر و لكن ليس بنفس النسبة. يمثل الشكل (10.c) ارتباطاً ضعيفاً لأن البيانات تتبعثر بشكل أكبر حول المستقيم، و هذا الارتباط سلبي. يبين الشكل (10.d) حالة عدم وجود ارتباط بين المتغيرين ، لأن انتشار النقاط عشوائي في المستوى.



الشكل (10): أشكال ترابط متغيرين عشوائيين.

هناك أشكال أخرى للارتباط بجانب أن يكون الارتباط الخطي. يبين الشكل (11) ارتباطاً تربيعياً بين المتغيرين العشوائيين X و Y . من الواضح أن البيانات غير مرتبطة خطياً، لأنها (البيانات) منتشرة بشكل متباعد حول الخط المستقيم. نلاحظ أيضاً أن المخطط التبعثري في الشكل (11) يلفت النظر إلى النقاط الشاذة. لاحظ أن النقطة المشار إليها بـ "نقطة شاذة" في الشكل أعلاه هي نقطة شاذة لأنها بعيدة عن بقية النقاط.



الشكل (11): ارتباط تربيعي بين المتغيرين.

على هذا يمكن القول: استخدم المخطط التبصري إذا أردت:

1. أن تدرس العلاقة بين متغيرين أو أكثر. و بالتحديد:

a. قوة الارتباط بين المتغيرات.

b. شكل العلاقة: خطية، تربيعية، لغارتمية، إلخ.

c. اتجاه العلاقة: موجبة أو سالبة.

2. أن تكشف عن وجود النقاط الشاذة التي يمكن أن تؤثر على دراسة العينة.

3.4 التحليل الاستكشافي للبيانات Exploratory Data Analysis

يؤدي التحليل الاستكشافي (يرمز له اختصاراً بـ EDA) دوراً مهماً للغاية عند دراسة و تحليل البيانات. في الواقع يتفق تقريباً جميع محلي البيانات على أن التحليل الإستكشافي هو الخطوة الأولى التي يجب عدم تجاوزها عند دراسة البيانات و تحليلها، كما يوضح ذلك الاقتباس التالي من أحد مراجع الكتاب.

“Never, ever, run any statistical test without performing EDA first!”

"أي لا تتم مطلقاً بإجراء أي اختبار إحصائي على بيانات بدون القيام أولاً بإجراء التحليل الاستكشافي لهذه البيانات"

لماذا نجري التحليل الاستكشافي؟

نقوم بإجراء التحليل الاستكشافي للأهداف التالية:

1. للكشف عن الأخطاء في البيانات (مثلاً النقاط الشاذة outliers) و خصوصاً تلك الناتجة عن خطأ في تفرغ البيانات أو الناتجة عن أخطاء تجريبية.
2. للكشف عن خصائص البيانات و خصوصاً: تناظر البيانات، التواء البيانات و جهة هذا الالتواء، و تبعثر البيانات.
3. اختبار فيما إذا كانت البيانات تتوزع وفق توزيع نظري معين (مثلاً هل تتوزع وفقاً للتوزيع الطبيعي).
4. لانتقاء الاختبار الإحصائي المناسب للبيانات: اختبارات و سيطوية أو غير وسيطوية. (سوف ندرس الاختبارات الوسيطية و الاختبارات غير الوسيطية في فصول قادمة).

تتجلى أهمية التحليل الاستكشافي من خلال الأهداف الآتية الذكر. فوجود القيم الشاذة ربما يجعل نتائج الاختبارات الإحصائية غير صحيحة. تناظر البيانات هو شرط ضروري لتطبيق الاختبارات الإحصائية، و مقياس الالتواء يحدد كمية الابتعاد عن التناظر. معرفة شكل توزيع البيانات هو غاية في الأهمية لأنه يحدد الاختبار الإحصائي الذي يجب استخدامه. كما سوف نرى في الفصول القادمة بأنه لا يمكن تطبيق عدد كبير من الاختبارات الإحصائية المهمة على بيانات مالم يكن توزيع مجتمع تلك البيانات يتبع التوزيع الطبيعي أو قريباً من الطبيعي.

ما العمليات التي يتضمنها التحليل الاستكشافي للبيانات؟

يتضمن هذا الاختبار العديد من الاختبارات الإحصائية و الرسوم البيانية التي تعطي فكرة واضحة عن البيانات المدروسة. نذكر هنا أهم هذه الاختبارات و الرسوم البيانية التي يتضمنها التحليل الاستكشافي.

الإحصاء الوصفي **Descriptive Statistics**: يتضمن العديد من الإحصائيات التي تعطي مقاييس النزعة المركزية إضافة إلى مقاييس التشتت. من الإحصائيات المهمة نذكر:

- المتوسط، الوسط، عدد القياسات في العينة، القيمة العظمى، القيمة الصغرى، إلخ.
- مقاييس التشتت مثل: التشتت، الانحراف المعياري، المدى، المدى الربيعي، إلخ.

اختبارات نوع توزيع البيانات (التوزيع الطبيعي مثلاً) و يتم فيه مقارنة توزيع العينة مع توزيع نظري معين (مثلاً التوزيع الطبيعي أو كاي مربع).

الرسوم البيانية **Graphical methods**: تستخدم الطرق البيانية لمعرفة توزيع العينة والقيم الشاذة. عادةً ما تكون الطرق البيانية أكثر دقةً في تحديد نوع توزيع العينة و ربما تكون هذه الطرق هي الوسيلة الفعالة لاكتشاف القيم الشاذة.

- المخططات التكرارية Histogram Plots
- مخطط الساق و الورقة Stem and Leaf
- المخطط التبعثري Scatter Plots
- المخطط الصندوقي Box Plot
- مخططات المقارنة مع توزيع نظري:
Normal probability plots (P-P plots) و Quantile-Quantile plot(Q-Q plots)

ملاحظة: كل من P-P plots و Q-Q plots تعطي فكرة عن ابتعاد العينة من التوزيع الطبيعي. الـ Q-Q plot يتحسس كثيراً انحراف العينة عن التوزيع الطبيعي في المناطق

الطرفية للتوزيع (distribution tails). الـ P-P plots يتحسس كثيراً عن انحراف العينة عن التوزيع الطبيعي قرب المتوسط.

سوف نستخدم مثلاً لتوضيح فكرة التحليل الاستكشافي وسوف نستخدم حزمة البرامج الإحصائية SPSS لهذه المهمة كونه البرنامج الإحصائي الواسع الانتشار و نظراً إلى فعاليته العالية في تحليل البيانات. في الواقع يوفر الـ SPSS طرقاً سهلة تمكن من إنجاز تحليل البيانات الاستكشافي بسلاسة و من خلال أمر واحد يدعى Explore.

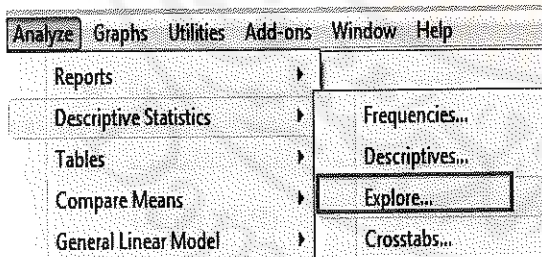
مثال توضيحي Illustrative Example

لتكن لدينا البيانات التالية التي تمثل كمية سكر الدم في عينة من 12 مريض.

50	98	100	101
95	89	103	110
98	91	98	100

المطلوب: اجري التحليل الاستكشافي للبيانات في الجدول أعلاه.

يتم إجراء التحليل الاستكشافي باستخدام الـ SPSS من خلال الأمر Explore.



ينتمي هذا الأمر إلى القائمة Analyze التي تحتوي على جميع الاختبارات الإحصائية. يمكن للقارئ أن يذهب إلى الملاحق للاستزادة أكثر حول استخدام

الـ SPSS. يبين الشكل الجانبي كيف يتم تفعيل أمر التحليل الاستكشافي ... Explore. ينتج عن تفعيل الأمر Explore مربع الحوار في الشكل (12.a).

في مربع الحوار في الشكل (12.a) ننقل المتغير SugerLevel من القائمة اليمنى إلى القائمة اليسرى المعنونة Dependent List، و ذلك بتعليم اسم المتغير و النقر على السهم ► ليتم نقله. (للمزيد من المعلومات حول استخدام الـ SPSS إذهب إلى الملاحق A, B, C). ننقر بعد ذلك على المفتاح Statistics لكي يظهر مربع الحوار في الشكل (12.b).

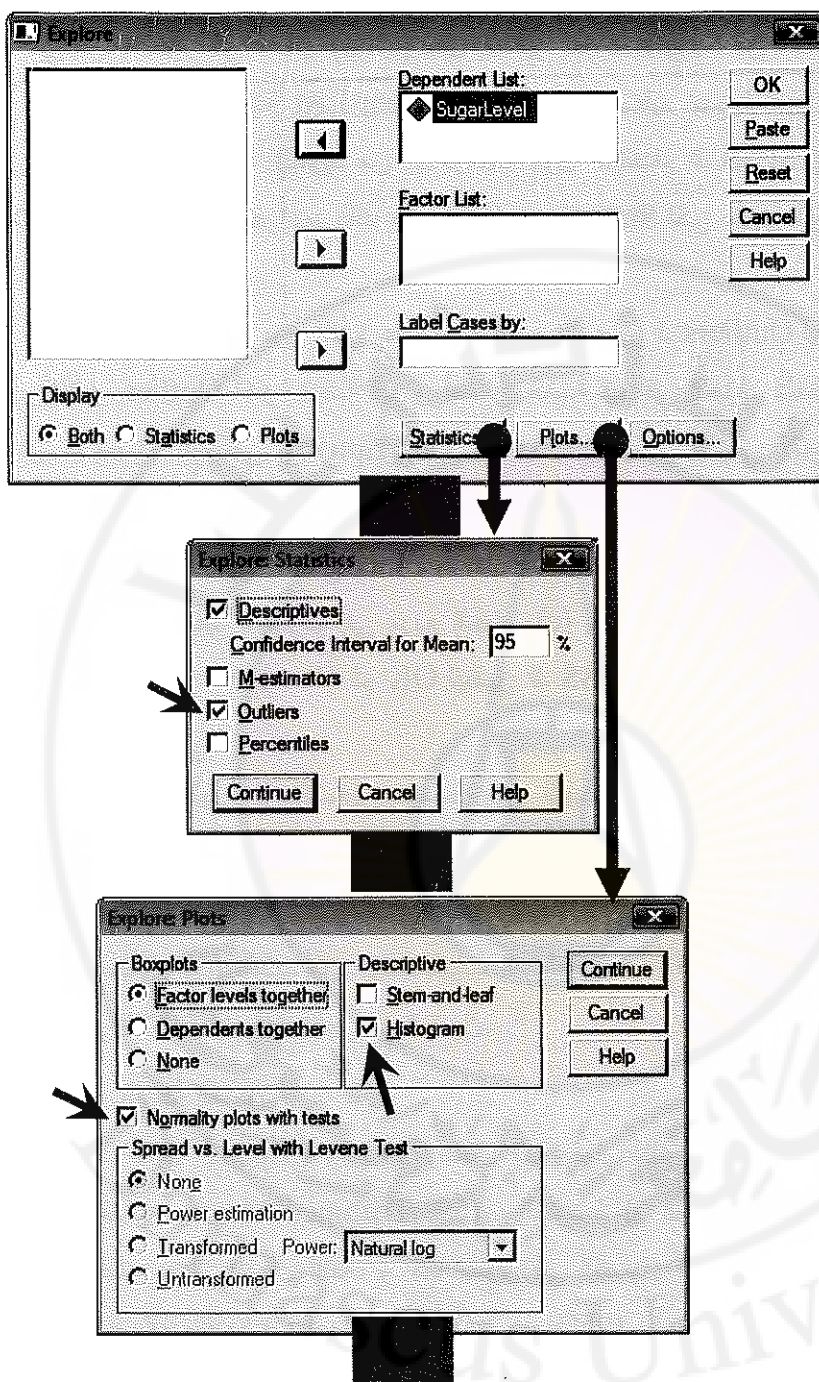
في مربع الحوار انقر بزر الفأرة على مربع النص المجاور لـ Outliers ليتم تفعيله. يظهر الأمر Outlier أكبر خمسة قيم و أصغر خمسة قيم في العينة. لاحظ أن أمر عرض الاحصاء الوصفي Descriptives هو فعال تلقائياً و بالتالي لا توجد حاجة لإعادة تفعيله. انقر بزر الفأرة على المفتاح Continue لكي تعود إلى مربع الحوار (a). انقر بزر الفأرة على المفتاح Plots و فعل الخيارات المشار عليها في السهم في مربع الحوار (c). تقيد الخيارات في مربع الحوار (c) باعطاء صورة بيانية عن شكل توزيع البيانات. انقر Continue ثم OK ليظهر الـ SPSS نتائج أمر تحليل البيانات الاستكشافي.

ينتج عن التحليل الاستكشافي مجموعة من الجداول والمخططات البيانية التي تعطي صورة واضحة عن طبيعة البيانات. فيما يلي سوف نقدم شرحاً مفصلاً عن نتائج التحليل الاستكشافي للعينة قيد الدراسة.

يبين الجدول (3) العديد من الإحصائيات المتعلقة بالعينة قيد الدراسة التي سوف نصنفها على ثلاث فئات كما يظهر في الجدول.

أولاً مقاييس النزعة المركزية

متوسط العينة (Mean) و 95% مجال ثقة حول ذلك المتوسط و المتوسط المشذب بمقدار (5% Trimmed Mean). المتوسط (و المتوسط المشذب) يحدد مركز البيانات. يعطي مجال الثقة، و كما سنرى في الفصول القادمة، تقديراً لمتوسط مجتمع العينة بثقة محددة (مقدار الثقة في مثالنا هو 95%). وأخيراً الوسط (Median) الذي يمثل القيمة التي تحصر نصف البيانات على اليمين و النصف الآخر على اليسار.



الشكل (12): مربعات الحوار الخاصة بأمر استكشاف البيانات .Explore...

Descriptives

SugarLevel			Statistic	Std. Error
مقاييس التزعة مقاييس التشتت	Mean		94.42	4.330
	95% Confidence Interval for Mean	Lower Bound	84.89	
		Upper Bound	103.95	
	5% Trimmed Mean		96.02	
	Median		98.00	
	Variance		224.992	
	Std. Deviation		15.000	
	Minimum		50	
	Maximum		110	
	Range		60	
	Interquartile Range		9	
	Skewness		-2.671	.637
	Kurtosis		8.299	1.232

الجدول (3): مجموعة من الإحصائيات الوافية للعينة.

ثانياً مقاييس التشتت

تحدد مقاييس التشتت تبعثر البيانات حول مركزها. مقياس التشتت (Variance)، والانحراف المعياري (Std. Deviation) هي من أهم المقاييس التي تحدد بعد (أو تبعثر) البيانات عن مركزها (المتوسط). يحدد التشتت (أو الانحراف المعياري) مدى فعالية المتوسط في تقدير متوسط المجتمع. كلما كان التشتت كبيراً فقد المتوسط فعاليتها باعتباره مقدراً لمتوسط المجتمع. من مقاييس التشتت الأخرى نذكر المدى (Range)، و يمثل طول حزمة البيانات (الفرق مابين القيمة العظمى Maximum و القيمة الصغرى Minimum). يفيد المدى الربيعي (Inter-quartile Range) في تحديد مقدار التشتت في الـ 50% من البيانات الوسطى.

ثالثاً معامل الالتواء

يفيد معامل الالتواء (Skewness) في معرفة مدى تناظر البيانات حول مركز هذه البيانات (غالباً المتوسط). إذا كانت البيانات متناظرة حول مركزها فإن معامل الالتواء

يساوي الصفر. إشارة معامل الالتواء تحدد جهة الالتواء. القيم السالبة تشير إلى التواء نحو اليسار، و القيم الموجبة تشير إلى التواء نحو اليمين. على سبيل المثال، البيانات قيد الدراسة ملتوية نحو اليسار بسبب كون قيمة معامل الالتواء سالبة (-2.671). بطبيعة الحال كلما ابتعدت قيمة معامل الالتواء عن الصفر سواء بالاتجاه الموجب أو الاتجاه السالب كانت البيانات غير متناظرة.

يعطي الجدول (4) نظرة عن البيانات من زاوية مختلفة. حيث يظهر أكبر خمسة قياسات في العينة، و أصغر خمسة قياسات مرتبة من الأكبر إلى الأصغر. هناك أهمية واضحة لهذا الجدول. فهو يبين البعد بين القيم الخمسة الكبرى و القيم الخمسة الصغرى في العينة المدروسة. و هذا بدوره يساعد في تحديد تجمع البيانات (أي فيما إذا كانت البيانات مبعثرة أو متجمعة). يساعد هذا الجدول أيضاً في تحديد الحاجة لاستخدام المتوسط المشذب. حيث إنه إذا كانت الخمسة قيم الكبرى بعيدة بشكل كبير عن القيم الصغرى، يمكن حذف هذه النقاط ، إذا كان حجم العينة يسمح بذلك ، و حساب المتوسط المشذب للبقية.

Extreme Values

			Case Number	Value
SugarLevel Highest	1	8	110	
	2	7	103	
	3	4	101	
	4	3	100	
	5	12	100	
Lowest	1	1	50	
	2	6	89	
	3	10	91	
	4	5	95	
	5	11	98	

الجدول (4): أكبر و أصغر خمسة قياسات في العينة مرتبة من الأكبر إلى الأصغر..

يظهر الجدول (5) نتائج اختبار التوزيع الطبيعي لمجتمع العينة (Test of Normality). يمكن تحديد توزيع مجتمع العينة من خلال إحصائية Kolmogorov-Simirnov أو إحصائية Shapiro-Wilk. بشكل عام، نستخدم إحصائية Shapiro-Wilk ومعنوية الاختبار المقابلة لها Sig معياراً لاتخاذ قرار حول توزيع مجتمع العينة و ذلك عندما يكون حجم العينة أقل من 2000. أما إذا كان حجم العينة أكثر من 2000 فإنه يفضل استخدام إحصائية Kolmogorov-Smirnov و قيمة المعنوية Sig المقابلة لها.

Tests of Normality

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	Df	Sig.
SugerLevel	.276	12	.012	.673	12	.000

جدول(5): نتائج اختبار طبيعية توزيع العينة.

الفرضيات الإحصائية التي نريد اختبارها هي (سوف نناقش الفرضيات الإحصائية بشكل مفصل في فصول قادمة):

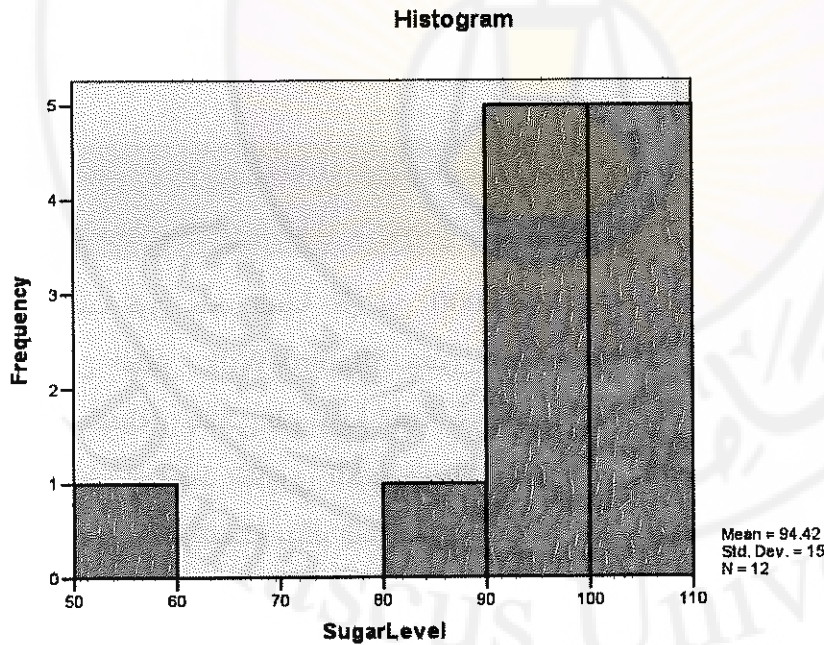
فرضية العدم: إن توزيع العينة هو توزيع طبيعي.
الفرضية البديلة: إن توزيع العينة لا يخضع للتوزيع الطبيعي.

بسبب كون حجم العينة أقل من 2000 فإننا نستخدم قيمة Shapiro-Wilk. بسبب أن اختبار Shapiro-Wilk معنوي (قيمة Sig تساوي الصفر) فإننا نقبل الفرضية البديلة و بالتالي فإن توزيع العينة غير طبيعي.¹
يظهر المخطط التكراري للعينة في الشكل (13). يمثل ارتفاع كل عمود عدد التكرارات في الفئة الممثلة بذلك العمود. كما يبدو من المخطط التكراري فإن العينة غير متناظرة حيث القيم تمتد بعيداً عن مركز البيانات (المتوسط 94.42)، و باتجاه اليسار.

¹ سوف نقدم في فصول قادمة الفرضيات الإحصائية و معايير قبول الفرضيات أو رفضها. لذلك في هذه المرحلة نشير فقط إلى أن الاختبار معنوي أو لا.

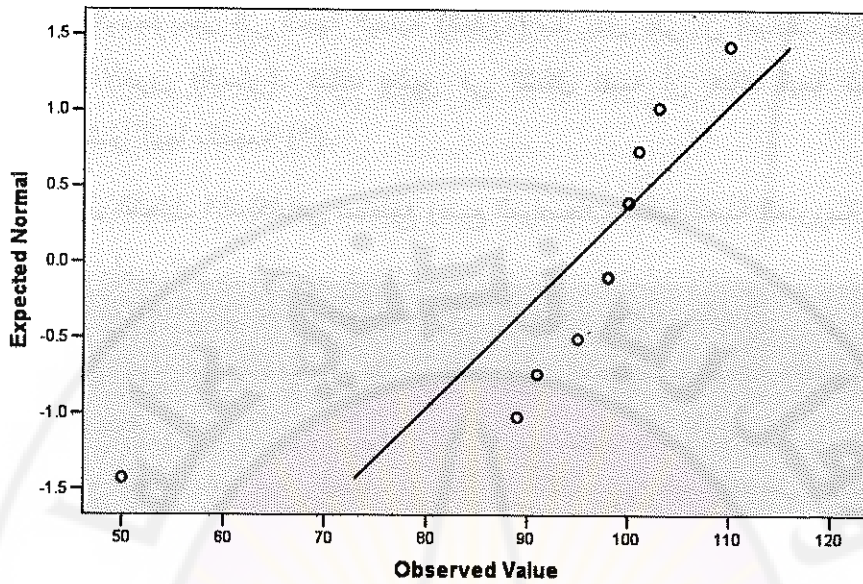
مما يشير إلى أن توزيع العينة غير متناظر وملتبس نحو اليسار و هي نفس النتيجة التي توصلنا إليها سابقاً. كذلك فإن العينة لا تخضع للتوزيع الطبيعي حيث إن المخطط التكراري لا يشبه شكل دالة التوزيع الطبيعي (تذكر أن التوزيع الطبيعي يأخذ شكلاً جرسياً).

يظهر مخطط الـ Q-Q في الشكل (14). يستخدم هذا المخطط لمعرفة فيما إذا كان هناك توافق مابين التوزيع الطبيعي و توزيع العينة. الخط المستقيم يمثل النقاط المتوقعة عندما يكون توزيع مجتمع البيانات (العينة) طبيعياً. النقاط (الدوائر الصغيرة) تمثل نقاط العينة. كما ذكرنا سابقاً حتى يكون توزيع العينة طبيعياً فإنه يجب أن تتطبق هذه الدوائر الصغيرة على المستقيم، أو تتجمع حوله بشكل قريب كفاية. بالعودة إلى الشكل (14) نجد أن النقاط تتبعثر حول المستقيم (لا تتطبق عليه و لا تتجمع حوله) ، و بالتالي فإن هذا دليل آخر على أن مجتمع العينة لا يخضع للتوزيع الطبيعي.

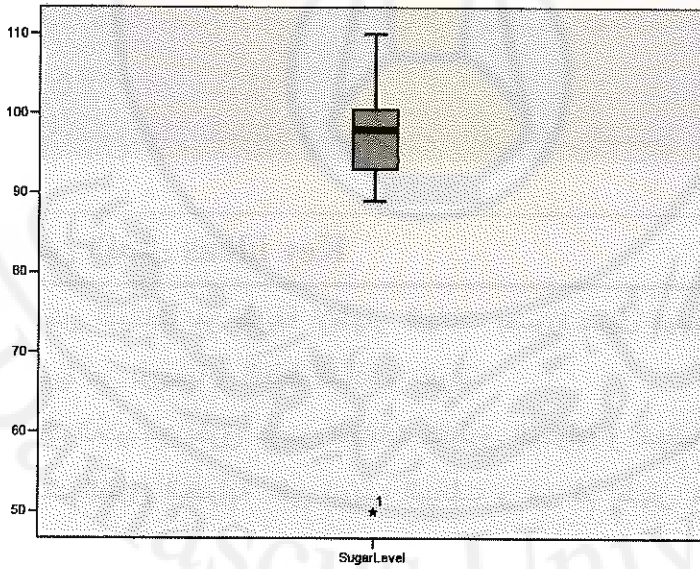


الشكل (13): المخطط التكراري للعينة المدروسة.

Normal Q-Q Plot of SugarLevel



الشكل (14): مخطط Q-Q لاختبار فيما إذا كانت العينة تتوزع وفق التوزيع الطبيعي.



الشكل (15): المخطط الصندوقي SugarLevel

يظهر المخطط الصندوقي للعينه في الشكل (15). من الواضح أن العينه غير متناظرة لأن الوسط يميل إلى الجهه اليمنيه (العلويه) مما يدل على أن البيانات ملتويه نحو الجهه الأخرى (اليسار). كذلك يظهر في الشكل نفسه أن هناك قيمه شاذة (outlier) مشاراً إليها بالنجمه (*).

من خلال التحليل الاستكشافي نستطيع وضع المواصفات التاليه للعينه المدروسه.

- (1) متوسط العينه ليس ممثلاً جيداً للعينه لأن الانحراف المعياري كبير.
- (2) توزيع العينه ليس متناظراً و هذا ما أشار عليه المخططات البيانيه (التكراري، والصندوقي).
- (3) توزيع العينه ملتوٍ نحو اليسار، و هذا ما أشار عليه قيمه معامل الالتواء (Skewness) في الجدول (3). و كذلك فإن الالتواء نحو اليسار كما هو واضح في المخططات البيانيه (التكراري، والصندوقي).
- (4) من خلال اختبار Shapiro-Wilk نستنتج أن مجتمع العينه لا يتوزع توزيعاً طبيعياً. كذلك وضع مخطط Q-Q Plot بيانياً أن هذه العينه لا تتوزع وفق التوزيع الطبيعي.
- (5) العينه تحتوي على نقاط شاذة كما هو واضح من خلال المخطط الصندوقي.

ملخص الفصل Key ideas

- (1) عرض البيانات هو طريقه لإظهار البيانات بصورة تسهل فهمها و إعطاء صورة عن توزيعها و النقاط الشاذة و غيرها.
- (2) هناك العديد من طرق عرض البيانات كالطرق الجدوليه و البيانيه.
- (3) تعتمد الطرق الجدوليه على تبويب البيانات و من ثم إظهار هذه البيانات بشكل جدولي.
- (4) الطريقه البيانيه: تعتمد هذه الطريقه على إظهار البيانات على شكل رسوم بيانيه مختلفه توضح كل واحده منها مجموعه من خصائص القياسات. هناك

العديد من الرسوم البيانية التي تم اختيارها بعناية فائقة لتوفير مجموعة صغيرة و لكنها كافية و تلبي جميع حاجات محلل البيانات.

● المدرج التكراري Histogram و يظهر البيانات على شكل أعمدة حيث ارتفاع كل عمود متناسب مع تكرارات عناصر العينة في كل مجال. يفيد المدرج التكراري بشكل خاص في توضيح شكل توزيع العينة.

● مخطط مقارنة توزيع مجتمع العينة مع توزيع نظري مثل التوزيع الطبيعي. يوفر هذا المخطط وسيلة فعالة لمقارنة توزيع العينة مع توزيع نظري. من أهم مواصفات هذا المخطط أنه ذو فاعلية كبيرة في كشف توزيع العينة وسهولة تفسير نتائجه.

● المخطط الصندوقي Box Plots. يمكن هذا المخطط البياني من إظهار المعلومات التالية لكل عينة.

- تناظر العينة و شكل الالتواء و اتجاهه إن وجد.
- القيم الشاذة outliers.
- الوسط و المدى الربيعي.
- المقارنة الفعالة لعينتين أو أكثر

● مخطط الساق و الورقة Stem and Leaf. و يفيد هذا المخطط بإظهار عدد كبير من بيانات رقمية بشكل ملخص و واضح. كما يساعد أيضاً في تبيان فيما إذا كانت العينة تتوزع وفق التوزيع الطبيعي.

● المخطط التبعثري Scatter Plot. يساعد في دراسة العلاقة ما بين متغيرين أو أكثر. من المعلومات المهمة التي يظهرها هذا المخطط: شكل العلاقة ما بين المتغيرات، قوة علاقة الارتباط، اتجاه الارتباط، القيم الشاذة.

(5) تحليل البيانات الاستكشافي Exploratory Data Analysis. يستخدم مجموعة من الإحصائيات و التقنيات البيانية لدراسة خصائص العينة. يعد هذا التحليل غاية في الأهمية و يجب أن نبدأ به قبل إجراء أي دراسات إحصائية للعينة. يكشف هذا التحليل عن مجموعة من المعلومات المهمة:

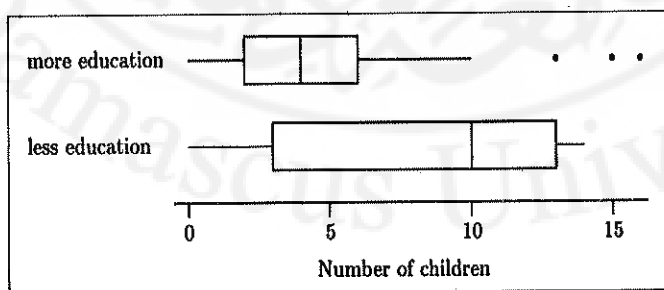
- مقاييس النزعة المركزية و مقاييس التشتت.
- شكل توزيع العينة (مثلاً التوزيع الطبيعي) و نظامية هذا التوزيع. إذا كان التوزيع غير نظامي فإننا نستطيع من خلال هذا التحليل أن نحدد مقدار الإلتواء و جهته.
- القيم الشاذة outliers.

4.4 تمرينات Exercises

أولاً: ادرس البيانات التالية و المخطط الصندوقي المرافق. تدرس البيانات في الجدول العلاقة بين مستوى التعليم و عدد أفراد الأسرة. الأرقام في الجدول تمثل عدد أفراد الأسرة للأمهات المتعلّقات فقط للصف السادس أو أقل و الأمهات المتعلّقات للصف السابع أو أكثر.

المطلوب قارن بين العيّنتين. ما الاستنتاجات التي توصلت إليها حول علاقة مستوى التعليم مع حجم الأسرة (عدد أفراد الأسرة).

Mother educated for six years or less
أمهات متعلّقات حتى الصف السادس أو أقل
14 , 13 , 4 , 14 , 10 , 2 , 13 , 5 , 0 , 0 , 13 , 3 , 9 , 2 , 10 , 11 , 13 , 5 , 14
Mother educated for seven years or more
أمهات صف سابع أو أكثر
0 , 4 , 0 , 2 , 3 , 3 , 0 , 4 , 7 , 1 , 9 , 4 , 3 , 2 , 3 , 2 , 16 , 6 , 0 , 13 , 6 , 1 , 6 , 5 , 9 , 10 , 5 , 4 , 3 , 3 , 5 , 2 , 3 , 5 , 15 , 5

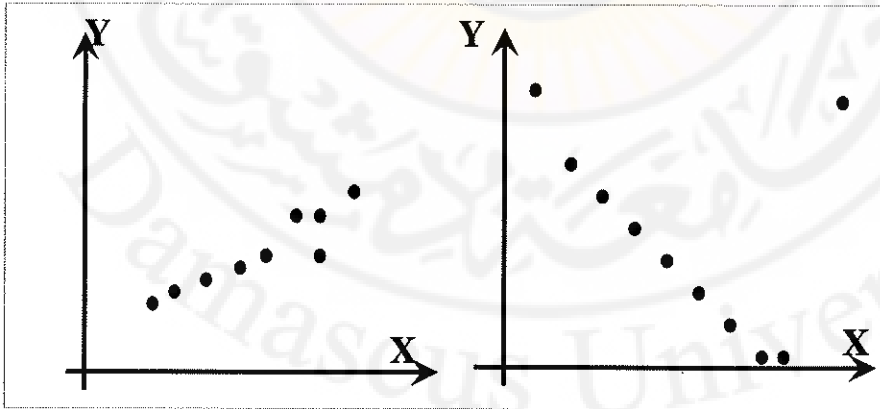


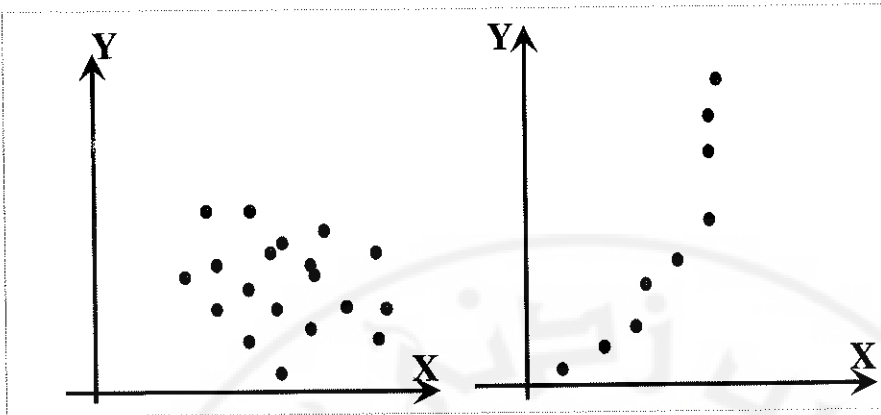
ثانياً: ادرس مخطط الساق و الورقة التالي:

Temperatures	
Tens	Ones
5	0 7 9
6	1 1 2 2 4 5 5 5 7 8 9
7	0 0 1 3 6 7 7 9 9
8	0 0 0 2 2 3 7

- مالذي يمكن أن تستنتجه عن البيانات؟
- هل المخطط أعلاه يوحي بأن البيانات تتوزع وفقاً للتوزيع الطبيعي؟
- استخدم المخطط أعلاه لإيجاد المنوال للعينة.
- استخدم المخطط أعلاه لإيجاد الوسط للعينة.

ثالثاً: بين نوع العلاقة بين المتغيرات الممثلة بالمخططات التبعثرية التالية و بين فيما إذا وجدت نقاط شاذة:





★ ★ ★ ★ ★ ★ ★ ★ ★ ★

الفرضيات و الاختبارات الإحصائية Statistical Tests and Hypothesis

1.5 الفرضية الاحصائية The Statistical Hypothesis

الفرضية الإحصائية هي مقولة حول ظاهرة معينة لا نعرف مسبقاً صحتها أو خطأها. يتم إثبات المقولة أو دحضها من خلال جمع بيانات عن الظاهرة و تحليلها. نتيجة تحليل البيانات إما أن نؤيد المقولة أو أن ندحضها. إحصائياً ، تصاغ الفرضية بشكل أو بآخر باستخدام وسطاء المجتمع (كالمتوسط Mean، الوسط Median، ...). الأمثلة التالية هي فرضيات قد تكون صحيحة و قد تكون خاطئة.

- متوسط كمية السكر في دم شخص غير مصاب بمرض سكر الدم يكون بين 60 – 95.
- تغيرات درجة الحرارة لا تؤثر في فعالية مخصب كيميائي ما.
- متوسط كمية الامطار الهاطلة في مدينة ما لا يتجاوز 200 مم.
- هناك علاقة بين ارتفاع ضغط الدم و كل من كمية الشحوم الثلاثية و العمر.
- تتعلق فعالية مسكنات الألم Paracetamole و Ibuprofen بجنس المريض.

من الواضح أن كلاً من هذه المقولات لا نعرف مسبقاً صحتها و بالتالي نسميها بالفرضيات. يُعرف الإحصائيون نوعين للفرضيات.

1.1.5 الفرضية البديلة Alternative Hypothesis

تشكل الفرضية البديلة هدف الدراسة ، أو فرضية البحث Research hypothesis ، و يقترحها الباحث الذي يقوم بالدراسة. على سبيل المثال ، يمكن لباحث يقوم باختبار فعالية مبيد حشري أن يفترض أن المبيد الحشري يقضي على أكثر من 99% من الحشرات أو رمزياً $\mu > 99\%$. يرمز الإحصائيون اصطلاحاً للفرضية البديلة بـ H_1 .

2.1.5 فرضية العدم Null Hypothesis

بعد وضع فرضية البحث ، فإننا يجب أن نضع فرضية نفي أو نقيض لفرضية البحث و يسمى هذا النقيض بفرضية العدم — أي تعدم أو تنفي صحة فرضية البحث. الهدف من فرضية العدم هو تحديد ما يجب قبوله في حال كون فرضية البحث غير صحيحة. على سبيل المثال ، بالعودة إلى مثال المبيد الحشري في الفقرة السابقة ، فإن فرضية العدم ستكون: المبيد الحشري يستطيع إبادة 99% أو أقل من الحشرات أو رمزياً $\mu \leq 99\%$. يرمز الإحصائيون اصطلاحاً لفرضية العدم بـ H_0 .

تمرين: ادرس المقولة التالية بعناية:
"إن معدل كمية الكولسترول Cholesterol لدى الأشخاص الذين تقع أعمارهم ضمن المجال 30 – 50 يساوي 190 mg/dL". أكتب الفرضيات المناسبة لهذه المقولة.

3.1.5 اختبار الفرضيات Hypothesis Test

تسمى عملية التحقق من صحة فرضية باختبار الفرضية. يشتمل اختبار أي فرضية على مجموعة من الإجراءات تبدأ (1) بتجميع البيانات حول الظاهرة التي وضعنا الفرضية حولها ، (2) اختيار المعالجة المناسبة لهذه البيانات (اختبار الفرضية) ، (3)

ومن ثم تفسير نتائج المعالجة. إذا كانت نتيجة المعالجة تؤيد الفرضية فإننا نقر بصحة هذه الفرضية و إلا نرفضها.

بشكل عام عند اختبار أي فرضية إحصائية ، فإن هناك نتيجتين محتملتين فقط:

1. إما قبول فرضية العدم و بالتالي رفض الفرضية البديلة (لأن الفرضية البديلة معاكسة تماماً لفرضية العدم، و بالتالي لا يمكن منطقياً أن يكون الشيء و ضده صحيحين بأن واحد).
2. أو رفض فرضية العدم، و بالتالي قبول الفرضية البديلة للأسباب نفسها المشار إليها في (1).

تجدر الإشارة هنا أن قبول فرضية إحصائية لا يعني أنها صحيحة بالمطلق. إنما يعني فقط عدم وجود دليل إحصائي كافٍ يثبت خطأ هذه الفرضية. و بالمثل ، إن رفض فرضية معينة لا يدل بالضرورة على أن هذه الفرضية خاطئة بالمطلق. بل يعني فقط أنه لا يوجد دليل إحصائي كافٍ يثبت صحة هذه الفرضية.

2.5 خوارزمية اختبار الفرضيات

بعد و ضع الفرضيات الإحصائية ، فإنه يمكن بشكل عام اتباع الخطوات التالية لاختبار هذه الفرضيات:

1. جمع العينات المناسبة للفرضيات المطلوب اختبارها. هذه العينات يجب أن تخدم حصراً عملية اختبار الفرضيات المطلوبة.
2. تحديد مستوى الأهمية و يرمز له اصطلاحاً ب α . يحدد مستوى أهمية الاختبار مقدار الخطأ الذي نقبل به برفض فرضية العدم H_0 ، علماً أن هذه الفرضية صحيحة. فمثلاً عند تحديد قيمة $\alpha = 5\%$ فإن ذلك يعني أننا نقبل بخطأ مقداره الأعظمي 5% بأن نرفض H_0 و هي في الواقع صحيحة.

3. تحديد نوع الاختبار أو الاختبارات الإحصائية المناسبة لهذه العينات. عملياً ، كل من نوع البيانات (كمي أو كيفي) و الفرضيات الإحصائية تحدد نوعية الاختبارات التي يجب أن تطبق على البيانات.

4. قاعدة القرار الإحصائي تعتمد على كل من مستوى الأهمية و إحصائية أخرى تدعى معنوية الاختبار Significance of the Test. يرمز عادةً لمعنوية الاختبار بـ Sig. أو P-value. بما أن الأداة الإحصائية التي نعتمد عليها في هذا الكتاب هي حزمة البرامج الإحصائية SPSS ، فإننا سوف نستخدم الرمز Sig. لمعنوية الاختبار انسجماً مع ترميز الـ SPSS. اعتماداً على ذلك فإن رفض H_0 أو قبولها يعتمد على:

❖ إذا كانت $Sig \geq \alpha$ ، عندئذٍ نقبل H_0 . السبب في قبول فرضية العدم في هذه الحالة هو أن الخطأ الفعلي Sig لرفض هذه الفرضية أكبر من الخطأ النظري α الذي نقبل به ، و بالتالي لا نملك إلا أن نقبل فرضية العدم.

❖ إذا كانت $Sig < \alpha$ ، عندئذٍ نرفض H_0 و نقبل H_1 . مرةً أخرى فإن الأساس المنطقي لهذا القرار هو أن الخطأ الفعلي Sig لرفض فرضية العدم أصغر من الخطأ النظري الذي نقبل به α و بالتالي من المنطقي أن نرفض فرضية العدم.

قبل البدء بالحديث عن الاختبارات الإحصائية، لابد من التعريف بمفهومين مهمين هما الخطأ من النوع الأول و الخطأ من النوع الثاني.

3.5 أنواع الخطأ في اختبار الفرضيات

عند اختبار فرضية ما ، فإن هناك احتمالاً بدرجة معينة لاتخاذ قرار خاطئ حول هذه الفرضية. هناك نوعان من الأخطاء لهما أهمية خاصة و يجدر بنا الحديث عنهما بشيء من التفصيل: الخطأ من النوع الأول و الخطأ من النوع الثاني.

1.3.5 الخطأ من النوع الأول Type I Error

يُعرف الخطأ من النوع الأول على أنه احتمال (قيمة بين 0 و 1) رفض فرضية العدم H_0 بينما هي صحيحة. و يرمز عادةً لهذا الخطأ بـ α . وضوحاً أن الخطأ من النوع الأول يقابل مستوى الأهمية كما تم تعريفه في الفقرة 2.5 أعلاه.

2.3.5 الخطأ من النوع الثاني Type II Error

يُعرف الخطأ من النوع الثاني على أنه احتمال قبول فرضية العدم H_0 بينما هي خاطئة. و يرمز عادةً لهذا الخطأ بـ β . في الواقع ، من الصعب إعطاء تفسير واضح للخطأ من النوع الثاني. لذلك عادةً ما نقول إن الخطأ β هو خطأ فقط لأننا فقدنا إمكانية رفض H_0 علماً أنها خاطئة و بالتالي لم نتمكن من استخلاص نتيجة كان من الممكن استخلاصها لولا وقوعنا في هذا الخطأ. يعطي الجدول (1) فكرة أوضح عن هذين الخطئين.

القرار الاحصائي	الفرضية H_0 صحيحة	الفرضية H_0 خاطئة
رفض H_0	خطأ من النوع الأول	قرار صحيح
قبول H_0	قرار صحيح	خطأ من النوع الثاني

جدول (1): الخطئين من النوع الأول I و النوع الثاني II.

إن أحد الأسئلة المهمة هنا هو: أيّ من الخطئين أكثر أهمية؟ بكلمات أبسط ، أيّ من الخطئين يجب أن نتجنب الوقوع فيه. لا يوجد جواب قطعي لهذا السؤال لأن أهمية أيّ من الخطئين تعتمد على طبيعة المسألة المدروسة و شكل الفرضيات. لتوضيح هذه الفكرة نأخذ المثالين التاليين.

مثال (1) Example (1)

لنفرض أن متهماً ما قد مثّل أمام القاضي بتهمة معينة. في هذه الحالة أمام القاضي فرضيتان هما:

الفرضيات الاحتمالية
H_0 : المتهم بريء (أي لا توجد أدلة كافية لاتهامه)
H_1 : المتهم مذنب (أي توجد أدلة كافية لاتهامه)

الخطأ من النوع الأول يعني أن القاضي رفض H_0 و هي صحيحة و حكم على المتهم علماً أن هذا المتهم في الواقع بريء. نتيجةً لهذا الخطأ نكون ليس فقط أرسلنا بريئاً إلى السجن بل (و ربما أخطر) ، تركنا المذنب الفعلي حراً. أما الخطأ من النوع الثاني سيكون قبول H_0 و هي خاطئة (أي تبرئة المتهم لعدم كفاية الأدلة). نتيجةً لذلك ليس فقط أننا لم نظلم بريئاً بل و لدينا فرصة لجمع أدلة (بيانات) إضافية ربما تدين هذا المتهم — طبعاً إذا كان فعلاً مذنباً. نستنتج بمقارنة الخطئين أن الخطأ من النوع الأول أسوأ و كان يجب تجنب الوقوع فيه.

مثال (2) Example (2)

لنفرض أن قسم ضبط الجودة (Quality Control) في إحدى الشركات الدوائية يريد فحص مجموعة عشوائية من منتج دوائي بهدف التأكد من جودته. للقيام بذلك وضع المختصون في تحليل البيانات الفرضيات التالية:

الفرضيات الاحتمالية
H_0 : المنتج الدوائي جيد و مطابق لمعايير الجودة
H_1 : المنتج الدوائي سيئ و غير مطابق لمعايير الجودة

لنفرض أنه لدى اختبار عينة من هذه الأدوية تم رفض فرضية العدم (H_0) على الرغم من أن الأدوية جيدة (خطأ من النوع الأول). في هذه الحالة يتوجب على الشركة

التخلص من هذه الأدوية و بالتالي تخسر ثمن الأدوية. و لكن دعونا الآن نفترض سيناريو آخر و هو أن المنتج في الواقع سيئ و مع ذلك قبلنا فرضية العدم. بهذا نكون قد ارتكبنا الخطأ من النوع الثاني. هناك تكلفتان واضحتان لهذا الخطأ. أولاً سوف تخسر الشركة ثمن هذه الأدوية لأن المستهلكين سوف يرجعون هذه الأدوية إلى الشركة. ثانياً (و ربما الأكثر خطورة على الشركة) ضياع سمعة الشركة ناهيك عن القضايا التي قد يرفعها المستهلكون على الشركة إذا أصيب أحد بأذى نتيجة استهلاك هذه الأدوية — ربما تقوم الشركة بدفع مبالغ مادية تعويضاً عن كل إصابة. بمقارنة الخطئين، فإنه يبدو واضحاً أن الخطأ من النوع الثاني أسوأ من الخطأ من النوع الأول و كان يجب تجنب الوقوع فيه.

يتضح من المثالين أعلاه أن أهمية أيّ من الخطئين ترتبط بطبيعة المسألة المدروسة. لذلك ينبغي علينا أن نكون حذرين و مراعاة خصوصية كل مسألة عند اتخاذ قرار حول رفض الفرضيات الخاصة بهذه المسألة أو قبولها.

4.5 الاختبارات الإحصائية Statistical Tests

الاختبارات الإحصائية هي اجراءات تُنفذ على البيانات الإحصائية بهدف الحصول على نتائج. تستخدم نتائج الاختبارات لتأييد فرضية العدم أو رفضها. بشكل عام يمكن تقسيم الاختبارات الإحصائية إلى نوعين.

1.4.5 اختبارات وسيطية (Parametric Methods (or Tests)

تتعلق هذه الاختبارات عادةً بوسطاء المجتمع (مثلاً المتوسط، الانحراف المعياري، و غيرها). تشترط هذه الاختبارات أن يكون المجتمع الذي سُحبت منه البيانات مجتمعاً طبيعياً (يخضع للتوزيع الطبيعي). شرط كون مجتمع البيانات طبيعياً هو شرط ضروري لا يمكن الاستغناء عنه. من الأمثلة على هذه الاختبارات نذكر:

1. اختبار ستودنت أو ما يعرف بـ T-Student Test
2. اختبارات تحليل التباين أو ما يعرف بـ ANOVA Test

هذه الاختبارات يمكن أن تطبق فقط إذا كانت البيانات عديدة بالإضافة إلى تحقق شرط طبيعية التوزيع. لا يمكن بالطبع تطبيق هذه الاختبارات الوسيطة على بيانات محرفية (ليست عددية) لأنها تعتمد على قيم المتوسط و الانحراف المعياري و حساب هاتين القيمتين يقتضي أن تكون البيانات عديدة.

2.4.5 اختبارات غير وسيطة (Nonparametric Methods (or Tests)

نلجأ عادةً لهذا النوع من الاختبارات عندما تكون شروط الاختبارات الوسيطة غير محققة. تتطلب هذه الاختبارات شروطاً أقل من تلك التي تتطلبها الاختبارات الوسيطة. على سبيل المثال فإن هذه الاختبارات لا تشترط أن يكون مجتمع البيانات يتوزع وفق التوزيع الطبيعي. يمكن تقسيم هذه الطرق إلى:

1. طرق تستخدم مع البيانات الكمية Quantitative Data
2. طرق تستخدم مع البيانات الكيفية Qualitative Data

بناءً على ذلك ، فإنه يجب استخدام الاختبارات غير الوسيطة فقط إذا كانت شروط تطبيق الاختبارات الوسيطة غير محققة. (إذا كانت شروط تطبيق الاختبارات الوسيطة غير محققة فإن النتائج الحاصلة من تطبيق هذه الاختبارات تكون محل شك.)

من الاختبارات غير الوسيطة نذكر:

- ❖ اختبار كاي - مربع Chi-squared Test
- ❖ اختبار الرتب لسبيرمان Spearman Rank Correlation Coefficient
- ❖ اختبار الإشارة Wilcoxon Signed-Rank Test
- ❖ اختبار مان وتني Mann-Whitney-Welcoxon Test

تصنف هذه الاختبارات على أنها مستقلة عن شكل التوزيع. بكلمات أخرى ، الاختبارات غير الوسيطة لا تشترط أن تكون المجتمعات المسحوبة منها العينات تتوزع وفق التوزيع الطبيعي. نتيجة لذلك ، فإنه يمكن استخدام الاختبارات غير الوسيطة بغض النظر عن توزيع مجتمع العينات المدروسة على عكس الاختبارات الوسيطة التي تشترط على أن يكون توزيع العينات طبيعياً لكي تصبح قابلة للتطبيق على هذه العينات.

تتويبه هام

الاختبارات غير الوسيطة أقل مرونة من الاختبارات الوسيطة من الناحية العملية و ربما أقل دقة لأنها تعتمد ، كما سوف نرى ، على كمية أقل من المعلومات حول العينة (رتب البيانات) بعكس الاختبارات الوسيطة التي تستخدم معلومات أكثر عن العينة. لذلك في الحالات التي تكون فيها البيانات مناسبة للاختبارات الوسيطة و شروط تطبيق هذه الاختبارات محققة فإنه يُفضل (بعض الإحصائيين يقولون يجب) استخدام الاختبارات الوسيطة.

في هذا الفصل سوف نناقش فقط الاختبارات الوسيطة أما الاختبارات غير الوسيطة فسوف تكون موضوع الفصل التالي.

5.5 الاختبارات الوسيطة Parametric Tests

كما رأينا سابقاً إن الاختبارات الوسيطة تعتمد على وسطاء المجتمع ، و لها شروط يجب أن نتحقق من أجل ضمان صحة نتائج الاختبار. سنناقش هنا مجموعة من الاختبارات الوسيطة و نحدد شروط تطبيق كل اختبار. سوف نطور المناقشة أكثر بحيث نُعرّف النموذج العام للمسائل الواقعية التي ينطبق عليها كل اختبار. يعد تعريف النموذج العام لهذه المسائل خاصة فريدة لهذا الكتاب لأنها تساعد الدارس على الحصول على مقياس يساعد في اختيار الاختبار المناسب لكل مسألة واقعية تصادفه. سنستخدم

الأمثلة المناسبة بهدف تسهيل المناقشة و تعزيز فهم الاختبار و نستخدم مخرجات البرنامج الإحصائي SPSS في أثناء المناقشة.

1.5.5 اختبارات ستودنت Student Test

اختبار ستودنت (يُعرف بـ T-Test أو Student-Test) هو أحد الاختبارات الوسيطة التي تستخدم لاختبار فرضيات حول متوسط مجتمع واحد أو فرضيات حول الفرق بين متوسطي مجتمعين.

هناك عدة شروط لتطبيق هذا الاختبار:

1. أن تكون البيانات عددية وليست محرفية أو سلاسل محرفية.
2. أن تكون العينة مسحوبة من مجتمع يخضع للتوزيع الطبيعي أو قريب من الطبيعي. هذا شرط ضروري لا يمكن إهماله.
3. أن تكون الفرضيات المطلوب اختبارها تتعلق بوسطاء المجتمع أو المجتمعات.
4. أن يكون حجم العينة مناسباً و لا يقل عن خمسة قياسات.

1.1.5.5 اختبار ستودنت حول مجتمع واحد

يفيد هذا الاختبار بالتحقق من فرضيات حول متوسط مجتمع واحد اعتماداً على عينة مسحوبة من هذا المجتمع. تنص فرضية العدم على أن المتوسط يساوي قيمة ثابتة و تنص الفرضية البديلة على المعاكس المنطقي لذلك. على هذا يمكن صياغة الفرضيات بالشكل:

الفرضيات الإحصائية	
H_0	نفي صحة المقولة في الفرضية H_1
H_1	يفترض الباحث علاقة بين متوسط المجتمع و قيمة ثابتة

هناك العديد من التطبيقات العملية لهذا الاختبار نذكر منها.

❖ إن متوسط عدد الجراثيم التي يبيدها المضاد الحيوي A هو أكبر من 95%.
يمكن صياغة ذلك بدلالة الفرضية البديلة و العدم بالشكل: $H_1: \mu > 95\%$ و $H_0: \mu \leq 95\%$.

❖ متوسط كمية سكر الدم في شخص بالغ لا يعاني من مرض السكر (diabetes) هو أقل من 100 mg/dl. يمكن صياغة ذلك بدلالة الفرضية البديلة و العدم بالشكل: $H_1: \mu < 100 \text{ mg/dl}$ و $H_0: \mu \geq 100 \text{ mg/dl}$.

❖ متوسط معدل التخرج لطالبات جامعة دمشق يساوي 70. يمكن أن يصاغ ذلك بدلالة الفرضية البديلة و العدم بالشكل: $H_1: \mu = 70$ و $H_0: \mu \neq 70$.

مثال تطبيقي

لنفرض أنه طلب منك التأكد من صحة الفرضية التالية:
متوسط عدد الأشقاء الأكبر لدى طلاب قسم الإحصاء الذين يحضرون مادة برامج احصائية (1) لا يساوي "1". الجدول التالي يبين عدد الأشقاء الأكبر لدى عينة طلاب قسم الإحصاء.

0, 0, 0, 3, 1, 2, 0, 0, 2, ...

تحليل المسألة

يتم تحليل بيانات ما من خلال الإجابة عن السؤالين التاليين:

— ما الأسئلة التي يمكن للبيانات المتاحة الإجابة عنها؟
— ماهي المسائل التي نهتم بها من بين كل المسائل التي يمكن للبيانات الإجابة عنها؟

لعلّ المسألة الوحيدة المهمة التي يمكن لهذه البيانات الإجابة عنها هي: هل متوسط المجتمع يساوي قيمة معينة؟ في الواقع يمكن الإجابة عن هذا السؤال لأنه يمكننا حساب

متوسط العينة ، و بالتالي تقدير متوسط مجتمع العينة. و من ثم يجري مقارنة المتوسط المقدر مع القيمة المفروضة.

الاختبار المناسب لهذه المسألة هو اختبار ستودنت لعينة واحدة لأن هذه المسألة تحقق تعريف هذا الاختبار (مقارنة متوسط مجتمع مع قيمة مفروضة). بمعرفة الاختبار المناسب نستطيع وضع الفرضيات المناسبة.

الفرضيات المطلوبة اختبارها:

فرضية العدم H_0 : متوسط عدد الأشقاء الأكبر يساوي 1

الفرضية البديلة H_1 : متوسط عدد الأشقاء الأكبر لا يساوي 1

النتائج SPSS Output

يبين الشكل (1) نتائج تحليل البيانات باستخدام برنامج الـ SPSS. يمكن للقارئ التحول إلى الملحق C لمعرفة كيفية إدخال البيانات و تفعيل هذا الاختبار. تظهر نتائج الاختبار بدلالة جدولين. يبين الجدول العلوي (A) وصفاً للعينة بدلالة:

1. عدد أفراد العينة N . و يفيد في التحقق من صحة إدخال البيانات، و كذلك في دلالة العينة.

2. المتوسط $Mean$. تشير قيمة المتوسط 1.26 إلى أن كل فرد من أفراد العينة له بالمتوسط شقيق واحد

3. الانحراف المعياري $Std. Deviation$. يعبر عن تبعثر القيم حول المتوسط.

4. الخطأ المعياري في المتوسط $Std. Error Mean$. يعبر عن الخطأ في اعتبار المتوسط ممثلاً لمتوسط المجتمع.

يبين الجدول السفلي (B) إحصائيات اختبار ستودنت التي تمكننا من التحقق من الفرضيات السابقة. بفرض أن مستوى الأهمية هو $\alpha=0.05$. بمقارنة مستوى الأهمية

المفروض مع معنوية الاختبار الصادرة عن معالجة البيانات Sig = 0.165 نجد أن اختبار ستودنت غير معنوي. هذا يعني أن البيانات لا توفر دليلاً إحصائياً كافياً لرفض فرضية العدم H_0 ، و بالتالي نستنتج أن المتوسط لا يختلف عن 1. أيضاً 95% مجال الثقة حول $\mu - 1$ يؤيد هذا الاستنتاج لأن هذا المجال $[-0.11, 0.63]$ يحوي الصفر، و بالتالي فإن الفرق بين القيمة المختبرة "1" و المتوسط الحقيقي للمجتمع يمكن أن يكون صفراً.

One-Sample Statistics				
	N	Mean	Std. Deviation	Std. Error Mean
Number of Older Siblings	46	1.26	1.255	.185

مقارنة المتوسط "1.26" إلى القيمة الافتراضية "1"

One-Sample Test						
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Number of Older Siblings	1.410	45	.165	.261	-.11	.63

نقبل فرضية العدم و نعزو الفرق الظاهري بين المتوسط و القيمة الافتراضية لخطأ المعاينة لأن الاختبار غير معنوي (Sig > 0.05)

الشكل(1): وصف العينة بدلالة مقاييس النزعة المركزية و جدول الاختبار

كتابة النتائج Reporting Results

البيانات في العينة لا تؤيد فرضية البحث التي تنص على أن متوسط عدد الأشقاء الأكبر لطلاب قسم الاحصاء الذين يحضرون مقرر برامج احصائية (1) لايساوي "1" و ذلك

بخطأ أعظمي مقداره "0.05". و بالتالي يمكن القول إن البيانات المتوفرة تؤيد أن متوسط عدد الأشقاء الأكبر لطلاب قسم الاحصاء الذين يحضرون مقرر برامج احصائية (1) غير مختلف معنوياً عن "1".

النموذج العام لمسائل اختبار ستودنت لعينة واحدة

هناك ظاهرة ما بوسيط (المتوسط مثلاً) غير معروف و نريد اختبار فرضيات حول هذا الوسيط.

مثال: متوسط معدلات النجاح لطالبات قسم الاحصاء أكبر أو يساوي 70%. وضوحاً أننا لانعلم متوسط معدل التخرج لكل طالبات قسم الاحصاء لذلك وضعنا فرضية حوله. يقوم اختبار ستودنت لمجتمع واحد بالتحقق من صحة (أي اختبار) هذه الفرضية.

2.1.5.5 اختبار ستودنت حول مجتمعين Two-Sample T Test

ناقشنا في الفقرة السابقة كيفية استخلاص نتائج حول متوسط مجتمع واحد. يطلب بشكل متكرر في التطبيقات العملية مقارنة متوسطي مجتمعين مختلفين و ذلك بغية استخلاص نتيجة حول هذين المجتمعين. على سبيل المثال ، يمكن أن يُطلب مقارنة متوسط انتاجية نوعين مختلفين من الذرة أو مقارنة متوسط الدخل السنوي للعاملين في القطاع الخاص و القطاع العام.

هناك الكثير من التطبيقات العملية لهذا الاختبار. فيما يلي بعض الأمثلة:

❖ تم تطبيق طريقة جديدة لزيادة نمو النباتات و يطلب تحديد فعالية هذه الطريقة مقارنة بالطريقة الأخرى قيد الاستخدام. إن هذه المسألة هي تطبيق مباشر على هذا الاختبار حيث يمكن أن نعرف فعالية الطريقة الجديدة من خلال، مثلاً، متوسط ارتفاع النباتات التي طبقت عليها هذه الطريقة و مقارنة هذا المتوسط مع متوسط أطوال النباتات التي طبقت عليها الطرق السابقة.

❖ مقارنة معدلات الطلاب الخريجين و الطالبات الخريجات من جامعة دمشق بهدف معرفة الفروق بين الأداء العلمي للطلاب و الطالبات. هذه أيضاً مسألة مقارنة متوسطين حيث نأخذ عينة من مجتمع الطلاب و عينة من مجتمع الطالبات ، و يحسب المتوسط في كل حالة ثم نقارن هذين المتوسطين.

❖ هناك دواء لخفض ضغط الدم الشرياني يسمى Amlodipine، و لهذا الدواء عياران معروفان 30 mg و 60 mg. يراد معرفة أي من هذين العيارين أكثر فعالية في ضبط (خفض) ضغط الدم. المسألة هنا إذا كان متوسط الضغط الناتج بعد المعالجة بهذين العيارين للدواء غير مختلفة إحصائياً فإن إعطاء الدواء بعيار أقل (30 mg) يكون أفضل (ربما من الناحية المادية و الصحية). من الواضح مرةً أخرى أن هذه المسألة هي مسألة مقارنة متوسطين لمؤشر ضغط الدم الشرياني بعد تطبيق هذين العيارين للدواء على مجموعة من المرضى، و من ثم اختبار تساوي هذين المتوسطين.

يمكن بشكل عام صياغة الفرضيات لاختبار ستودنت لمجتمعين كما يلي:

الفرضيات الاحصائية	
$H_0: 1. \mu_1 \leq \mu_2$	لاحظ أننا عرفنا كل الأشكال التي يمكن أن تأخذها فرضية البحث H_1 . عرفنا في الفرضية النقيضة (العدم) H_0 نقيض كل شكل من الأشكال الأربعة التي يمكن أن تأخذها فرضية البحث. على سبيل المثال ، فرضية البحث (1) $\mu_1 > \mu_2$ يقابلها فرضية العدم (1) $\mu_1 \leq \mu_2$ و الأخيرة هي النقيض.
2. $\mu_1 \geq \mu_2$	
3. $\mu_1 = \mu_2$	
4. $\mu_1 \neq \mu_2$	
$H_1: 1. \mu_1 > \mu_2$	
2. $\mu_1 < \mu_2$	
3. $\mu_1 \neq \mu_2$	
4. $\mu_1 = \mu_2$	

أي إن الفرضيات التي يتم فحصها باستخدام هذا الاختبار هي اختبارات حول المتوسطات (لذلك يسمى هذا الاختبار باختبار المتوسطات). فرضية البحث H_1 تضع فرضيات مختلفة حول علاقة المتوسطين. فرضية العدم تنفي ما تفترضه فرضية البحث.

لكي نقرر إحصائياً قبول أي من هاتين الفرضيتين أو رفضها نتبع الأسلوب التالي:

1. نحدد مستوى الأهمية α . يتم تحديد مستوى الأهمية α من قبل الباحث و تعتمد قيمته على تقدير الباحث و أهمية المسألة. في كثير من الأبحاث فإن قيمة α تحدد بـ 1% أو 5%، و لكن مرةً أخرى يعود تقدير قيمة α للباحث نفسه.

2. يتم قبول فرضية العدم أو رفضها من خلال أحد الأسلوبين التاليين

1.2 نقارن α مع معنوية الاختبار، و يرمز لها Sig. كما في الـ SPSS. كمايلي:

❖ إذا كانت $\alpha > sig$ ، عندئذٍ نقبل H_0 (و حتماً نرفض H_1)

❖ إذا كانت $\alpha \leq sig$ ، عندئذٍ نرفض H_0 (و حتماً نقبل H_1)

2.2 من خلال مجالات الثقة المبنية حول الفرق ما بين المتوسطين كمايلي:

❖ إذا كان $\mu_1 - \mu_2 \in [-v_1, +v_2]$ فإننا نقبل فرضية العدم H_0 و ذلك لكون هذا المجال يحوي الصفر (طرفا المجال من إشارتين مختلفتين سالب و موجب).

❖ إذا كان $\mu_1 - \mu_2 \in [+v_1, +v_2]$ فإننا نرفض فرضية العدم H_0 و ذلك لكون هذا المجال لا يحوي الصفر (طرفا المجال من إشارة واحدة موجب). من الواضح في هذه الحالة أن $\mu_1 > \mu_2$.

❖ إذا كان $\mu_1 - \mu_2 \in [-v_1, -v_2]$ فإننا نرفض فرضية العدم H_0 و ذلك لكون هذا المجال لا يحوي الصفر (طرفا المجال من إشارة واحدة سالب). من الواضح في هذه الحالة أيضاً أن $\mu_1 < \mu_2$.

3.1.5.5 اختبار ستودنت حول مجتمعين مستقلين

Two-Sample Independent T Test

و يستخدم هذا الاختبار في حالة مقارنة متوسطي عينتين عشوائيتين مسحوبتين بشكل مستقل من مجتمعين مختلفين. على سبيل المثال ، تقدير فعالية مضاد حيوي معين على نوعين مختلفين من الجراثيم. فالظاهرة المدروسة هي مقدار تأثير المضاد الحيوي بينما الجراثيم هي متغير مستقل له مستويان هما نوعا الجراثيم. بالطبع هذا يكافئ قولنا بأننا سحبنا عينتين مستقلتين من مجتمعين مختلفين (نوعي الجراثيم) و نريد مقارنة متوسطي مقدار تأثير المضاد الحيوي في هذين النوعين. يضع هذا الاختبار بعض الشروط على العينتين لكي نستطيع استخدامه.

1. العينتان مسحوبتان من مجتمعين طبيعيين. هذا الشرط ضروري لا يمكن الاستغناء عنه.
2. العينتان مستقلتان. بمعنى أنه لا يوجد رابط بين العينتين المسحوبتين و إلا لا يجوز استخدام هذا الاختبار.
3. تساوي تباينات المجتمعين. تحقق هذا الشرط غير ضروري و يمكن تجاهله.

بشكل عام يمكن وضع النموذج العام التالي للمسائل التي ينطبق عليها اختبار ستودنت لعينتين مستقلتين:

النموذج العام

هناك ظاهرة يهتم بها الباحث و يؤثر في هذه الظاهرة متغير بمستويين اثنين من التأثير، و يطلب تقويم الفرق (إن وجد) بين تأثير هذين المستويين في الظاهرة. مثال توضيحي: نريد دراسة تأثير عيارين مختلفين لدواء خافض للضغط على مرض الضغط. الظاهرة التي نهتم بها هي مرض الضغط أما المتغير المؤثر هو الدواء ومستوياته هي العياران المختلفان.

مثال تطبيقي : اختبار ستودينت لعينتين مستقلتين

أراد أحد علماء النفس معرفة فيما إذا كان الضغط النفسي من الأقران في النوادي النسائية الجامعية يفاقم اضطرابات الأكل عند النساء، وبالتالي يسبب زيادة في الوزن لدى هؤلاء الذين يعانون من مثل هذه الضغوطات. لأجل ذلك تم أخذ عينة من النساء داخل النوادي النسائية الجامعية، و عينة من نساء من خارج النوادي النسائية الجامعية، و تم أخذ أوزان أفراد العينتين. الجدول التالي يبين أوزان العينتين.

125	120	90	115	105	أوزان النساء اللواتي يرتدن النوادي النسائية الجامعية
110	130	120	135	140	أوزان النساء اللواتي لا يرتدن النوادي الجامعية

هل يوجد فرق معنوي إحصائياً بين أوزان العينتين؟

تحليل المسألة

يمكن لهذه البيانات أن تجيب عن الأسئلة التالية:

- ما متوسط أوزان مجتمع النساء اللواتي يرتدن النوادي النسائية الجامعية؟
- ما متوسط أوزان مجتمع النساء اللواتي لا يرتدن النوادي النسائية الجامعية؟
- هل متوسط أوزان النساء اللواتي يرتدن النوادي النسائية الجامعية مختلف عن متوسط النساء اللواتي لا يرتدن النوادي النسائية الجامعية؟

على الرغم من أن البيانات تستطيع الإجابة على كل الأسئلة السابقة و لكن السؤال المركزي بالنسبة للباحث هو السؤال الأخير: هل هناك فرق بين متوسط أوزان النساء اللواتي يرتدن أو لا يرتدن النوادي الجامعية. هناك تطابق مباشر بين هذه المسألة و اختبار ستودنت لعينتين مستقلتين فالظاهرة محور الاهتمام هي أوزان النساء و المتغير المستقل المؤثر هو النوادي الجامعية و له مستويان: إما النساء يرتدن أو لا يرتدن.

بنتيجة هذين المستويين لدينا عينتان مستقلتان: أوزان النساء المقابلة لكل مستوى. وبالتالي فإن الاختبار المناسب لهذه البيانات هو اختبار ستودنت لعينتين مستقلتين.

الفرضيات المطلوب اختبارها:

أي لا يوجد فرق معنوي إحصائياً بين أوزان النساء في العينتين	$H_0 : \mu_s = \mu_{ns}$
أي يوجد فرق معنوي إحصائياً بين أوزان النساء في العينتين	$H_1 : \mu_s \neq \mu_{ns}$

نتائج الاختبار باستخدام البرنامج SPSS

أولاً يقوم الـ SPSS بعرض ملخص عن العينات قيد التحليل من خلال جدول الإحصاء الوصفي Descriptive Statistics المبين أدناه.

يتضمن الملخص الإحصائي المعلومات المهمة التالية عن العينة:

1. العينتان المدروستان **sorority** (نساء في النوادي النسائية الجامعية) و **other** (نساء لسن في النوادي النسائية الجامعية).
2. عدد عناصر كل عينة N . كما هو ملاحظ من الجدول أعلاه أن حجم العينتين متساوٍ و يساوي 5.
3. متوسط العينة **Mean**. كما يبدو من الجدول أن متوسط العينة **sorority** هو 111 و أن متوسط العينة **others** هو 127.
4. الانحراف المعياري **Std. Deviation**. حيث يظهر من الجدول أن الانحراف المعياري للعينة **sorority** يساوي 13.874 و الانحراف المعياري للعينة **others** يساوي 12.042.
5. الخطأ المعياري في المتوسط **Std. Error Mean**. يمثل الخطأ المعياري في المتوسط مقدار الخطأ الذي نرتكبه عند اعتبار متوسط العينة ممثلاً (مقدراً) لمتوسط المجتمع. على سبيل المثال، إذا عدنا متوسط العينة **sorority** ممثلاً لمتوسط مجتمع هذه العينة فإننا نرتكب خطأً

أعظمية مقداره "6.205". و بالمثل، إذا عددنا متوسط العينة others ممثلاً لمتوسط مجتمع هذه العينة فإننا نرتكب خطأ أعظمية مقداره "5.385".

		متوسط العينة		الانحراف المعياري	
Group Statistics					
Group		N	Mean	Std. Deviation	Std. Error Mean
Weight	sorority	5	111	13.874	6.205
	Others	5	127	12.042	5.385

الخطأ المعياري في المتوسط

ثانياً جدول إحصائية ستودنت. يقدم هذا الجدول معلومات كافية (واقعياً أكثر من كافية) لاتخاذ قرار حول الفرضيات الإحصائية. هناك مجموعتان من الفرضيات نريد اختبارهم. بالتحديد فرضيات حول تساوي (أو عدم تساوي) تباينات المجتمعين (مجتمع العينة sorority و مجتمع العينة others)، و فرضيات تساوي (أو عدم تساوي) متوسطي المجتمعين.

فيما يلي شرح مفصل حول نتائج اختبار الفرضيات.

اختبار فرضيات التباين: الفرضيات حول تباين المجتمعين هي:

(أي تباين المجتمعين متساويان)	$H_0 : \sigma_{sorority} = \sigma_{others}$
(أي تباين المجتمعين غير متساويين)	$H_1 : \sigma_{sorority} \neq \sigma_{others}$

لاختبار هاتين الفرضيتين نحدد مستوى الأهمية α . سوف نعد من الآن فصاعداً أن $\alpha = 5\%$ مالم نحدد غير ذلك صراحةً. المعلومات التي نحتاجها من الجدول أعلاه هي الإحصائية $F=0.094$ و معنوية الاختبار الموافقة لها $\text{Sig.} = 0.767$. حسب القاعدة التي اعتمدناها سابقاً نقبل فرضية العدم H_0 لأن معنوية الاختبار أكبر من مستوى الأهمية α — $\text{Sig.} > \alpha$ — أي أن اختبار تساوي التباينات غير معنوي. و بالتالي نستنتج أنه لا يوجد فرق معنوي بين تباينات (Variances) المجتمعين.

اختبار فرضيات حول الفرق بين متوسط المجتمع: ربما الفرضيات الأهم التي يعالجها هذا الاختبار هي فرضيات حول الفرق بين متوسطي مجتمعين.

(أي متوسطا المجتمعين متساويان)	$H_0 : \mu_{\text{sorority}} = \mu_{\text{others}}$
(أي متوسطا المجتمعين غير متساويين)	$H_1 : \mu_{\text{sorority}} \neq \mu_{\text{others}}$

Independent Samples Test

تستخدم لاختبار تساوي التباين

		t-test for Equality of Means								
		Test for Equal Var.		T	Df	Sig.	Mean	Std.	95% Conf Int	
		F	Sig.			2-tail	Diff.	Error		
								Diff.	Lower	Upper
Weight	Eq.	0.09	0.76	-1.94	8	0.087	-16.0	8.216	-34.95	2.946
	Uneq			-1.94	7.8	0.088	-16.0	8.216	-35.01	3.011

تستخدم لاختبار فرضيات حول المتوسطات

باختيار مستوى الأهمية $\alpha = 5\%$ و بالنظر إلى الجدول أعلاه نجد اختبار ستودنت غير معنوي لأن معنوية الاختبار $\text{Sig.} = 0.087$ و هي أكبر من مستوى الأهمية. (لاحظ أننا استخدمنا معنوية الاختبار في السطر Eq. لأن التباينات متساوية.) بناءً على ذلك فإننا نقبل فرضية العدم H_0 . أي أنه لا يوجد فرق معنوي بين متوسطي أوزان أفراد المجتمعين.

يمكن الوصول إلى الاستنتاج ذاته إذا نظرنا إلى مجالات الثقة المبنية حول الفرق بين المتوسطات. لاحظ ، كما يوضح الجدول ، أنه بثقة مقدارها 95% فإن الفرق بين متوسطي أوزان النساء في النوادي النسائية الجامعية و النساء اللاتي لسن في النوادي النسائية الجامعية هو ضمن المجال $[-34.95, 2.946]$. بما أن هذا المجال يحوي الصفر (لأن طرفي المجال من إشارتين مختلفتين) فإنه يمكن القول و بثقة مقداره 95% أن الفرق يمكن أن يكون مساوياً للصفر ، و بالتالي المتوسطين غير مختلفين بشكل معنوي.

هناك بعض الإحصائيات التي تظهر في الجدول أعلاه التي سوف نلقي بعض الضوء عليها:

❖ **Mean Diff.**: الفرق الملاحظ بين متوسطي العينتين (Observed

$$\mu_{\text{sorority}} - \mu_{\text{others}} = 111 - 127 = -16.0 \text{ (Difference)}$$

❖ **Std. Error Diff.**: الفرق المتوقع بين متوسطي العينتين (Expected

(Difference)

ملاحظة Remark

يفضل بعض الإحصائيين استخدام السطر المقابل لـ Eq. في الجدول أعلاه لقراءة قيم T و Sig. حتى لو كانت تباينات مجتمعي العينتين غير متساوٍ.

عموماً إن الإحصائيات في جدول اختبار ستودنت التي نعتمد عليها بشكل مباشر لاتخاذ قرار (رفض أو قبول) حول الفرضيات الإحصائية هي:

1. الإحصائية F و مستوى معنوية الاختبار المقابلة لها $Sig.$ و تستخدم هاتان الإحصائيتان لاختبار فرضيات حول تساوي التباينات.
2. الإحصائية T و مستوى معنوية الاختبار المقابلة لها $Sig.$ و تستخدم هاتان الإحصائيتان لاختبار فرضيات حول تساوي متوسطات المجتمعين. و يمكن أيضاً النظر إلى مجال الثقة لاستقراء نتائج حول المتوسطات كما ناقشنا سابقاً.

كتابة النتائج Result Reporting

يظهر التحليل الإحصائي للبيانات المتوفرة عدم وجود دليل على أن الضغط من الأقران في النوادي النسائية الجامعية يؤدي إلى زيادة وزن النسوة. حيث تبيّن بنتيجة استخدام الاختبار الإحصائي المناسب (اختبار ستودنت لعينات مستقلة) أنه لا يوجد فرق ذو دلالة احصائية ، و بخطأ حده الأعظمي 5% ، بين أوزان النساء اللواتي يرتدن النوادي النسائية الجامعية و أولئك اللواتي لا يرتدن النوادي الجامعية. لكن وجود الفرق الظاهر في متوسط العينات (وهو كبير إلى حد ما) ربما يشير إلى أن البيانات المأخوذة غير كافية و أنه يتوجب استقصاء المسألة بشكل أعمق.

4.1.5.5 اختبار ستودنت لمقارنة متوسطي عينتين مرتبطتين

Two-Sample Paired T Test

و يستخدم هذا الاختبار لمقارنة عينتين من القياسات و ذلك عندما تكون هذه القياسات متعلقة ببعضها البعض بطريقة معينة. بكلمات أبسط ، مقارنة نتائج مأخوذة حول وحدات تجريبية بعد إخضاع هذه الوحدات إلى تجربتين مختلفتين. على سبيل المثال ، لمعرفة فيما إذا كانت القهوة تزيد تركيز النيكوتين في دم الانسان تم أخذ مجموعة من الوحدات التجريبية (أشخاص مدخنين) و تم قياس تركيز النيكوتين لهذه المجموعة قبل إعطاء كل منهم كوباً من القهوة ، و من ثم تم قياس تركيز النيكوتين في دمهم بعد إعطائهم كوباً من القهوة.

مجموعة من الوحدات التجريبية تجرى عليهم تجربتين مختلفتين و يراد دراسة تأثير هاتين التجربتين على الوحدات التجريبية.

مثال توضيحي: نريد دراسة فعالية طريقة حمية جديدة على خفض الوزن لدى مجموعة من الشباب ذوي الوزن الزائد. الظاهرة هنا هي وزن الشباب و المتغير المطلوب تقييم تأثيره هو طريقة الحمية. الآن بمقارنة الوزن قبل الحمية مع الوزن بعد الحمية، نستطيع الوصول إلى استنتاجات حول فعالية الطريقة الجديدة للحمية.

مثال تطبيقي: اختبار ستودنت لعينتين مرتبطتين

Paired Sample T Test

لمعرفة فعالية دواء مهدئ على خفض ضغط الدم لدى المرضى الذين يعانون من ضغط دم مرتفع ، تم أخذ عينة من عشرة مرضى و تم قياس ضغط الدم لديهم قبل إعطائهم الدواء المهدئ ، ثم أعيد قياس ضغط الدم بعد ساعتين من إعطاء الدواء المهدئ. الجدول التالي يبين قياس ضغط الدم قبل المعالجة (إعطاء الدواء المهدئ) وبعدها للعينة المدروسة.

201	180	196	187	170	208	205	192	170	180	قبل المعالجة
200	178	198	185	162	200	201	190	170	182	بعد المعالجة

هل تظهر هذه البيانات أي فاعلية للدواء؟ (تقاس الفاعلية بمدى قدرة الدواء على تخفيض ضغط الدم).

تحليل المسألة

هناك العديد من الأسئلة التي يمكن لهذه البيانات الإجابة عنها. لكنني سوف أهتم هنا فقط بأهم هذه الأسئلة التي تطابق السؤال المطلوب. أي هل تظهر هذه البيانات فعالية للدواء المهدئ؟

تستطيع هذه البيانات الإجابة عن هذا السؤال لأنه يمكن حساب متوسط كل من العينتين، ومقارنة متوسطي مجتمع العينتين باستخدام الاختبارات الإحصائية المناسبة. إذا كان ضغط الدم بعد أخذ الدواء أقل بشكل معنوي من متوسط ضغط الدم قبل المعالجة فإن الدواء فعال ، و إلا فالدواء ليس فعالاً.

هناك اختباران محتملان للإجابة عن السؤال المطروح. اختبار ستودنت للعينات المستقلة و اختبار ستودنت للعينات المرتبطة. بما أن أفراد العينة أنفسهم قبل التجربة (إعطاء الدواء) و بعدها ، فإن الاختبار الأفضل هو اختبار ستودنت للعينات المرتبطة كون هذه البيانات تحقق تعريف هذا الاختبار.

الفرضيات المطلوب اختبارها

$H_0 : \mu_{before} = \mu_{after}$	متوسط ضغط الدم قبل المعالجة يساوي المتوسط بعد المعالجة.
$H_1 : \mu_{before} \neq \mu_{after}$	متوسط ضغط الدم قبل المعالجة لا يساوي المتوسط بعد المعالجة .

ملاحظة: يمكن للقارئ الانتقال إلى الملحق C للمزيد من المعلومات حول كيفية إدخال البيانات في شاشة الـ SPSS و تفعيل اختبار ستودنت المتعلق بعينتين مرتبطتين.

نتائج الاختبار باستخدام الـ SPSS

أولاً يقوم الـ SPSS بعرض ملخص إحصائي عن البيانات كما هو موضح في الجدول التالي.

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	BeforeTreatment	188.90	10	13.772	4.355
	AfterTreatment	186.60	10	13.689	4.329

حيث يظهر في الجدول المتوسط و الانحراف المعياري للعينة قبل المعالجة (Mean = 188.90 و Std. Deviation = 13.772) و كذلك

المتوسط والانحراف المعياري للعينة بعد المعالجة AfterTreatment) Mean = 186.60 و Std Deviation = 13.689 .

ثانياً يبين الجدول التالي الارتباط بين العينتين. كما يتضح من الجدول فإن قيمة هذا الارتباط Correlation تساوي 0.967 و هو ارتباط قوي جداً بين العينتين. لاختبار فيما إذا كانت قيمة هذا الارتباط معنوية إحصائياً (بمعنى أنها لا تساوي الصفر) نستخدم معنوية الاختبار Sig في الجدول. بما أن قيمة معنوية الاختبار "Sig = 0.000" فهذا يدل على أن الاختبار معنوي. بكلمات أخرى، إن الارتباط بين المتغيرين BeforeTreatment و AfterTreatment هو معنوي — مختلف عن الصفر.

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	BeforeTreatment & AfterTreatment	10	.967	.000

ثالثاً تمكن مخرجات الـ SPSS في الجدول التالي من اختبار الفرضيات أعلاه. نلاحظ أن اختبار ستودنت غير معنوي لأن معنوية الاختبار "0.069" وهي أكبر من مستوى الأهمية المفروض "0.05". بسبب عدم معنوية الاختبار فإننا لا نملك دليلاً كافياً يسمح لنا برفض فرضية العدم H_0 . وبالتالي فإن المتوسطين غير مختلفين معنوياً. يؤيد هذا الاستنتاج مجال الثقة 95% المحسوب حول الفرق بين المتوسطين. بما أن هذا المجال يحوي الصفر، فإنه لا يوجد فرق معنوي إحصائياً بين المتوسطين.

Paired Sample Test

Paired Sample Test								
	Paired Differences					T	Df	Sig
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
Pair 1 B.Treatment – A.Treatment	2.300	3.529	1.116	-0.225	4.825	2.061	9	

كتابة النتائج Result Reporting

لقياس فعالية دواء مهدئ في خفض ضغط الدم ، تمت دراسة عينتين من قياسات ضغط الدم لمجموعة من المرضى قبل أخذ الدواء المهدئ و بعد أخذ هذا الدواء. بإجراء الاختبار الإحصائي المناسب لهذه البيانات (اختبار ستودنت للعينات المرتبطة) تبين أن لا يوجد فرق معنوي إحصائياً بين متوسط ضغط الدم لدى العينة المدروسة قبل أخذ الدواء و متوسط ضغط الدم لدى نفس العينة بعد أخذ الدواء ، وذلك بنسبة خطأ مقداره "0.05". على هذا الأساس يمكن القول إن البيانات المتوفرة لا تكشف أي فعالية للدواء المهدئ في خفض ضغط الدم.

2.5.5 اختبارات التباين ANOVA

اختبار التباين هو اختبار وسيطي يستخدم لاختبار فرضيات حول الفرق بين متوسطات مجتمعين إثنيين أو أكثر. في حالة مجتمعين إثنيين فقط يكون اختبار التباين و اختبار ستودنت متكافئين و غالباً ما يتم استخدام اختبار ستودنت في هذه الحالة. أما في حالة أكثر من مجتمعين فإن اختبار التباين يصبح الخيار الأفضل.

السؤال المهم هنا إذا كان لدينا أكثر من مجتمعين لماذا لا نستخدم اختبار ستودنت حيث نقوم بتقسيم المجتمعات إلى ثنائيات و نطبق اختبار ستودنت على كل ثنائية؟ في الواقع إن السبب الرئيسي في عدم استخدام اختبار ستودنت في حالة مقارنة متوسطات أكثر من مجتمعين هو أن قيمة الخطأ في اتخاذ قرار إحصائي حول متوسطات المجتمعات مُجمعة تزداد طردياً مع زيادة عدد المجتمعات.

بشكل خاص لنفرض أن لدينا ثلاثة مجتمعات A و B و أخيراً C. إذا أردنا مقارنة متوسطات هذه المجتمعات باستخدام اختبار ستودنت ، فإننا نقوم بتقسيمهم إلى ثلاث ثنائيات هي (A, B) ، (A, C) ، و أخيراً (B, C). لو فرضنا أن مستوى الأهمية المفروض $\alpha = 5\%$ فإن احتمال أن يكون قرارنا صحيحاً هو 0.95 من أجل كل ثنائية. الآن احتمال أن يكون قرارنا صائباً على الثنائيات الثلاثة مجتمعة هو $(0.95)^3$

0.8574. و بالتالي يكون احتمال اتخاذنا قراراً خاطئاً حول تساوي متوسطات المجتمعات الثلاثة هو $1 - (0.95)^3 = 14.26\%$. الجدول (2) يبين زيادة مقدار هذه الأخطاء كلما ازداد عدد المجتمعات.

عدد المجتمعات	عدد الثنائيات الممكنة	إحتمال اتخاذ قرار خاطئ
2	1	$1 - (0.95)^1 = 5\%$
3	3	$1 - (0.95)^3 = 14.26\%$
4	6	$1 - (0.95)^6 = 26\%$
5	10	$1 - (0.95)^{10} = 40.1\%$
6	15	$1 - (0.95)^{15} = 53.7\%$
7	21	$1 - (0.95)^{21} = 65.9\%$

جدول (2): احتمال الخطأ المرتكب عند استخدام اختبار ستودنت لاختبار فرضيات حول فرق مجتمعين أو أكثر.

لاحظ أنه من أجل مجتمعين اثنين فإن احتمال اتخاذ قرار خاطئ حول الفرضيات (بالتحديد رفض فرضية العدم و هي صحيحة) يساوي 5%. و لكن احتمال اتخاذ قرار خاطئ حول الفرضيات يزداد بسرعة مع زيادة عدد الفرضيات. لاحظ أنه من أجل 3 مجتمعات فإن احتمال الوقوع في خطأ حول الفرضيات يزداد بسرعة ليبلغ 14% تقريباً. من أجل 5 مجتمعات ، فإن احتمال اتخاذ قرار خاطئ حول فرضيات مقارنة متوسطات المجتمعات يساوي تقريباً 40% و هو احتمال كبير و لا يمكن القبول به! اعتماداً على ماسبق ، نستنتج أن استخدام اختبار ستودنت غير مقبول و يبقى اختبار التباين هو الخيار المناسب.

يتطلب استخدام اختبار تحليل التباين تحقق الشروط التالية.

1. العينات عبارة عن معطيات عددية مسحوية من مجتمعات طبيعية أو قريبة من الطبيعية.

2. يجب أن تكون تباينات العينات متجانسة. كما سوف نرى لاحقاً أنه يمكن تجاوز هذا الشرط، و ذلك باستخدام أوامر فرعية من تحليل التباين.
3. حجوم العينات متقاربة.
4. العينات مستقلة.

1.2.5.5 تحليل التباين بمتغير واحد مستقل One-way ANOVA

في هذا النوع من الاختبارات يكون لدينا متغير تابع تخضع قياساته لتأثير متغير آخر يسمى المتغير المستقل. يمكن للمتغير التابع أن يمثل ظاهرة معينة يراد دراستها والمتغير المستقل هو مؤثر معين يعتقد الباحث أن له دوراً مهماً في التأثير في خصائص الظاهرة. على سبيل المثال، يتأثر عدد الجراثيم التي تنمو في مستعمرة جرثومية بمتغير درجات الحرارة داخل هذه المستعمرة. إن عدد الجراثيم هو الظاهرة التي نهتم بها (أو المتغير التابع) و أن درجة الحرارة هي المتغير المستقل الذي نريد تقويم تأثير تغيراته في متغير عدد الجراثيم.

النموذج العام

ظاهرة تخضع لتأثير متغير له مستويان من التأثير على الأقل. المطلوب هو دراسة كيفية تأثير مستويات هذا المتغير المستقل في الظاهرة. أي هل المستويات لها التأثير نفسه أم تختلف هذه المستويات في تأثيرها.

مثال توضيحي: أجريت دراسة في إحدى مزارع تربية الفروج في ريف دمشق لمعرفة العلاقة بين نمو الفروج (يقاس النمو بزيادة وزن الفروج) و الفترة التي تتم فيها تربية الفروج. لأجل ذلك تم تقسيم فترات السنة إلى ست فترات طول كل فترة شهرين. أخذت عينات من الفروج في كل فترة و تم أخذ أوزان هذه الفرايج. المطلوب هل الفترات المختلفة لها دور في زيادة نمو الفروج؟ إن هذه المسألة ينطبق عليها اختبار تحليل التباين بمتغير مستقل واحد حيث الظاهرة المدروسة هي نمو الفروج و المتغير هو الفترة الزمنية و له ستة مستويات تأثير.

مثال تطبيقي Illustrative Example

تشير البيانات في الجدول التالي إلى شدة الألم المتبقي عند المرضى ممثلة بالمتغير Pain Score بعد أخذ مجموعة من أربعة مسكنات ألم مختلفة ممثلة بالمتغير Drug. المتغير التابع هو درجة الألم Pain Score و المتغير المستقل هو الدواء المسكن Drug. لاحظ أن المتغير المستقل يملك أربعة مستويات تأثير ، و هذه المستويات تقسم المتغير التابع إلى أربع مجموعات groups تقابل تأثيرات المتغير المستقل على الظاهرة. بالتحديد هذه المستويات هي: المجموعة المقابلة للدواء المسكن Diclofenac ، المجموعة المقابلة للدواء المسكن Ibuprophen ، المجموعة المقابلة للدواء المسكن Paracetamol ، و المجموعة المقابلة للدواء المسكن Aspirin. و السؤال المطلوب الإجابة عليه هو: هل تؤثر مستويات المتغير Drug بنفس أم بدرجات مختلفة على المتغير التابع؟ بكلمات أبسط هل تملك الأدوية درجات متساوية أم مختلفة من التأثير على تهدئة الألم؟

Pain Scores for Analgesics		توضيح بما أن القيم في الجدول تمثل شدة الألم المتبقي بعد إعطاء المسكن ، فإن المتوسط الأصغر هو الأفضل.
Drug	Pain Score	
Diclofenac	0, 35, 31, 29, 20, 7, 43, 16	
Ibuprophen	30, 40, 27, 25, 39, 15, 30, 45	
Paracetamol	16, 33, 25, 32, 21, 54, 57, 19	
Aspirin	55, 58, 56, 57, 56, 53, 59, 55	

مستويات المتغير المستقل أو المؤثر

يمكن و باستخدام نفس الأسلوب الذي اعتمدناه في تحليل المسائل السابقة أن نقول إن أفضل اختبار متاح هو تحليل التباين بمتغير واحد مستقل. لماذا؟ أجر هذه المحاكمة بنفسك.

لكي نجيب على هذا السؤال ، يجب أولاً أن نضع الفرضيات المناسبة.

الفرضيات التي نريد اختبارها هي:
 فرضية العدم H_0 : متوسطات درجة الألم متساوية بالنسبة إلى المسكنات الأربعة
 الفرضية البديلة H_1 : على الأقل أحد المتوسطات مختلف.

نتائج الاختبار باستخدام الـ SPSS

أولاً: من المهم جداً أن ننظر إلى نتائج اختبار تساوي تباينات المجموعات المختلفة المشكلة نتيجة مستويات تأثير المتغير المستقل. لأنه كما ناقشنا سابقاً إن أحد شروط تطبيق هذا الاختبار هو أن يكون التباين بين المجموعات متجانساً Homogeneity of Variance. الفرضيات التي نريد اختبارها هي:

H_0	التباينات بين المجموعات متجانسة
H_1	التباينات بين المجموعات غير متجانسة

يظهر الشكل (2) جميع الإحصائيات اللازمة لاختبار الفرضيات حول تباين المجموعات الأربعة المقابلة لمسكنات الألم المختلفة. اختبار التباين معنوي لأن قيمة

Test of Homogeneity of Variances: Pain اختبار تجانس التباين			
Levene Statistic	df1	df2	Sig.
4.837	3	28	.008

معنوية الاختبار "0.008" و هي أصغر من مستوى الأهمية المفروض 5% و بالتالي ، فإننا نرفض فرضية العدم و نقبل أن التباينات غير متجانسة

(الفرضية البديلة). هذا يعني أن أحد شروط تطبيق ANOVA غير محقق. بكلمات أخرى يجب النظر بحذر إلى نتائج تحليل التباين. و لكن من أجل الاستمرار في المناقشة سنعرض نتائج تحليل التباين حتى لو كانت مشكوكاً بصحتها.

ثانياً: يظهر الشكل (3) نتائج تحليل التباين. يعطي هذا الشكل معلومات كافية لاختبار الفرضيات حول المتوسط التي تم صياغتها أعلاه. نلاحظ أن اختبار التباين معنوي لأن

قيمة معنوية الاختبار "0.000" و هي أصغر من مستوى الأهمية 0.05. بالنتيجة نقبل الفرضية البديلة "يوجد على الأقل متوسط واحد مختلف". (تذكر أن شرط تحليل التباين غير محقق و بالتالي يجب النظر إلى النتائج بحذر.)

ANOVA: Pain

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	4956.375	3	1652.125	11.967	0.000
Within Groups	3865.500	28	138.054		
Total	8821.875	31			

الشكل (3): إحصائيات ANOVA لاختبار الفرضيات حول المتوسطات.

اختبارات لا تتأثر بعدم تساوي التباين

Robust Tests of Equality of Means

عندما يكون شرط تساوي التباينات غير محقق، فإن هناك إحصائيات بديلة يمكننا من اختبار فرضيات التباين. في الواقع هناك مجموعة من الإحصائيات التي يمكن استخدامها لاختبار فرضيات تحليل نذكر منها: إحصائية Welch و إحصائية Brown-Forsythe. هاتان الإحصائيتان لا تتأثران بعدم تساوي التباينات ، و بالتالي يمكن استخدامهما في المسألة التي نقوم بدراستها.

يظهر الشكل (4) إحصائيات Welch و Brown-Forsyth للبيانات في مسألتنا الحالية. تختبر إحصائيات Welch و Brown-Forsythe نفس الفرضيات التي يختبرها تحليل التباين. نلاحظ أن كلا من الإحصائيتين Welch و Brown-Forsythe معنوية لأن قيمة معنوية الاختبار Sig المقابلة لكلا منهما تساوي الصفر "0.000" و هي أصغر من مستوى الأهمية المفروض 5%. بكلمات أخرى نرفض فرضية العدم و

نقبل الفرضية البديلة "هناك متوسط واحد على الأقل مختلف"؛ نفس الاستنتاج الذي توصلنا إليه باستخدام اختبار تحليل التباين.

Robust Tests of Equality of Means: Pain				
	Statistic	df1	df2	Sig.
Welch	32.064	3	12.171	.000
Brown-Forsythe	11.967	3	18.889	.000
الشكل (4): نتائج اختباري ولش و براون-فورسيث.				

المقارنات المتعددة Multiple Comparisons

بوفر الـ SPSS مجموعة من الاختبارات التي تمكنا من تحديد المتوسطات المختلفة. لعل اختبار **Tukey** و اختبار **Games-Howell** هما أكثر هذه الاختبارات شيوعاً (استخداماً). يفيد هذان الاختباران بمقارنة المتوسطات متنى متنى و تحديد المتوسطات المختلفة.

ملاحظة: لمعرفة كيف يتم تفعيل هذين الخيارين في الـ SPSS الرجاء الانتقال إلى الملحق **C**.

يظهر الشكل (5) مخرجات الـ SPSS الخاصة بهذين الاختبارين. اختبار **Tukey** يعتمد في الواقع على تساوي التباينات. و بما أن هذا الشرط غير محقق في حالتنا، فإننا نتجاهل نتائج هذا الاختبار. اختبار **Games-Howell** لا يفترض تساوي التباينات ، و بالتالي يمكن الوثوق به أكثر في حالتنا هذه.

يظهر الشكل (5) الفروق بين متوسطي جميع الثنائيات المختلفة مرمزة بالشكل I – (J). بما أن شروط اختبار **Tukey** غير متوفرة كون تباينات المجموعات غير متساوية ، فإننا سنركز على اختبار **Games-Howell** كونه لا يشترط تساوي التباينات.

بالنظر إلى النصف السفلي من الشكل (5) نجد ، على سبيل المثال ، أن الفرق بين متوسط "4" (Aspirin) و متوسط "1" (Diclofenac) هو 33.5000 (موجب) و 95% مجال الثقة حول هذا الفرق هو [16.45, 50.55] و بما أن طرفي مجال الثقة موجب فإن متوسط الـ Aspirin أعلى من متوسط Diclofenac. عموماً يقوم الـ SPSS بتعليم جميع الفروق المعنوية عند مستوى أهمية 5% بـ "**".

كما يبدو واضحاً من الشكل (5) أن مسكن الألم 4 (Aspirin) يملك المتوسط الأكبر بين باقي المتوسطات مما يدل على أن الأسبرين هو الدواء الأقل فاعلية في تسكين الألم. كذلك ، فإنه لا يوجد أي فرق معنوي إحصائياً بين متوسطات بقية المسكنات الأخرى. (لاحظ أنه لا توجد العلامة "*" على أي من الفروق بين متوسطات المسكنات الأخرى غير الـ Aspirin). وبناءً على ذلك ، نجد أن المتوسط المختلف بين هذه المتوسطات هو فقط متوسط المسكن Aspirin. أما باقي المتوسطات فهي غير مختلفة إحصائياً.

Multiple Comparisons Dependent Variable: Pain

	(I) Drug	(J) Drug	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
لايستخدم لأن البيانات غير متجانسة Tukey HSD	1	2	-8.750	5.875	.457	-24.79	7.29
		3	-9.500	5.875	.386	-25.54	6.54
		4	-33.500(*)	5.875	.000	-49.54	-17.46
	2	1	8.750	5.875	.457	-7.29	24.79
		3	-.750	5.875	.999	-16.79	15.29
		4	-24.750(*)	5.875	.001	-40.79	-8.71
	3	1	9.500	5.875	.386	-6.54	25.54
		2	.750	5.875	.999	-15.29	16.79
		4	-24.000(*)	5.875	.002	-40.04	-7.96
	4	1	33.500(*)	5.875	.000	17.46	49.54
		2	24.750(*)	5.875	.001	8.71	40.79

		3	24.000(*)	5.875	.002	7.96	40.04
Games-Howell	1	2	-8.750	6.176	.513	-27.05	9.55
		3	-9.500	7.548	.602	-31.45	12.45
		4	-33.500(*)	5.194	.001	-50.55	-16.45
	2	1	8.750	6.176	.513	-9.55	27.05
		3	-.750	6.485	.999	-20.09	18.59
		4	-24.750(*)	3.471	.001	-36.03	-13.47
	3	1	9.500	7.548	.602	-12.45	31.45
		2	.750	6.485	.999	-18.59	20.09
		4	-24.000(*)	5.558	.014	-42.26	-5.74
	4	1	33.500(*)	5.194	.001	16.45	50.55
		2	24.750(*)	3.471	.001	13.47	36.03
		3	24.000(*)	5.558	.014	5.74	42.26

* The mean difference is significant at the .05 level.

الشكل (5): نتائج تطبيق اختباري Games-Howell و Tukey.

كتابة النتائج Reporting Results

نتيجة تحليل البيانات المتوفرة باستخدام تحليل التباين تبين أن متوسطات كمية الألم المتبقية بعد المعالجة بكل نوع من المسكنات مختلفة و ذلك بخطأ حده الأعظمي 5%. مما يدل على أن كفاءة هذه الأدوية في تسكين الألم مختلفة. و بإستقصاء المسألة أكثر لمعرفة أي المتوسطات هو المختلف و بالتالي تحديد الدواء الأكثر فعالية و جدنا أن متوسط الأسبرين هو الأعلى مما يدل على أن فعاليته هي الأقل بين المسكنات الأربعة. تجدر الإشارة إلى أنه لم تظهر أية فروقات معنوية إحصائياً بين بقية المسكنات. و مع ذلك فإن التفحص الدقيق للنتائج يشير إلى أن متوسط الألم المتبقي بعد أخذ المسكن Diclofinac هو الأصغر ، مما قد يشير إلى أن فعالية هذا المسكن ربما تكون أفضل. لذلك ننصح بتقصي المسألة أكثر و ذلك إما بزيادة حجم العينات أو التأكد من صحة و دقة الظروف التجريبية.

ملاحظات أساسية Teaching Notes

هناك بعض النقاط نريد أن نركز عليها حول اختبار One-Way ANOVA

1. قبل البدء بتفسير جدول تحليل التباين ، يجب الانتباه إلى نتيجة تحليل التباين. إذا كان التباين متجانساً بين كل طبقات المتغير المستقل ، عندئذٍ نستخدم جدول تحليل التباين لتحديد قبول فرضية العدم أو رفضها. أما إذا كان هذا التباين غير متجانس فإننا نقوم بإجراء اختبائي Welch, Brown-Forsythe.

2. إذا تم قبول فرضية العدم H_0 تنتهي المسألة. أما إذا تم رفض فرضية العدم عندئذٍ نستخدم اختبائي Tukey و Games-Howell لتحديد المتوسطات المختلفة.

3. تفسير النتائج و كتابة تقرير يجيب عن كل أسئلة الدراسة مستشهدين بالإحصائيات المناسبة.

2.2.5.5 تحليل التباين بمتغيرين مستقلين Two-way ANOVA

هناك تشابه بين تحليل التباين بمتغيرين مستقلين و تحليل التباين بمتغير مستقل واحد من حيث شروط التطبيق. و لكن الاختلاف المهم هو أن تحليل التباين بمتغيرين مستقلين يدرس تأثير عاملين اثنين (Two Factors) في المتغير التابع بدلاً من عامل واحد. وجود متغيرين مستقلين يؤثران على الظاهرة يعطي بعداً مهماً للمسألة. بجانب دراسة تأثير كل متغير على حدة في الظاهرة ، ندرس أيضاً إمكانية التأثير المتبادل؛ أو مايسمى بالتفاعل بين المتغيرين. التأثير المتبادل بين المتغيرين المستقلين ، إن وجد، يعني أن شكل تأثير أحد المتغيرين يتأثر بتغيرات المتغير الثاني.

من الأمثلة التي يعالجها هذا الاختبار نذكر:

- ❖ دراسة تأثير كل من مستوى التعليم و سنوات الخبرة في رواتب الموظفين في الدولة. لاحظ أن المتغير التابع هو مقدار الراتب ، و أن المتغيرات المستقلة التي نريد دراسة تأثيرها في الراتب هي مستوى التعليم ، و عدد سنوات الخبرة.
- ❖ دراسة تأثير كل من كمية السماد الكيميائي و طريقة الري (تنقيط ، رش ، غمر ، ...) في إنتاجية نبات القطن. مرةً أخرى المتغير التابع هو إنتاجية القطن و المتغيرات المستقلة هي السماد الكيميائي و طريقة الري.
- ❖ دراسة تأثير كل من الترسيب Sedimentation و مراحل التنقية Filtration في كمية العكر في مياه البحيرات السورية. هل تستطيع تحديد المتغير التابع (الظاهرة المدروسة) و المتغيرات المستقلة المؤثرة على المتغير التابع؟

النموذج العام

متغير تابع يمثل ظاهرة معينة و نريد دراسة تأثير متغيرين مستقلين في هذه الظاهرة. يملك كل متغير مستقل مستويين اثنين على الأقل.

مثال توضيحي: أجريت دراسة في إحدى مزارع تربية الفروج في ريف دمشق لمعرفة العلاقة بين نمو الفروج (يقاس النمو بزيادة وزن الفروج) و كل من الفترة التي تتم فيها تربية الفروج و جنس الفروج. متغير الفترة الزمنية له ست فترات طول كل فترة شهرين، و متغير الجنس له مستويان (ذكر، أنثى). أخذت عينات من الفروج في كل فترة و تم أخذ أوزان هذه الفرائج لكل من الذكور و الإناث. المطلوب هل الفترات المختلفة لها دور في زيادة نمو الفروج؟ و هل هذا التأثير يتعلق بجنس الفروج؟ للتوضيح: إن هذه المسألة ينطبق عليها اختبار تحليل التباين بمتغيرين مستقلين حيث الظاهرة المدروسة هي نمو الفروج و المتغير الأول هو الفترة الزمنية، و المتغير الثاني هو الجنس.

الفرضيات التي يختبرها Two-way ANOVA

هناك ثلاث مجموعات من الفرضيات يتم اختبارها وفق هذا الاختبار.

1. هل هناك فرق معنوي بين متوسطات المجموعات الناتجة عن مستويات المتغير المستقل الأول بصرف النظر عن المتغير الثاني؟
2. هل هناك فرق معنوي بين متوسطات المجموعات الناتجة عن مستويات المتغير المستقل الثاني بصرف النظر عن المتغير الأول؟
3. هل يوجد تفاعل أو تأثير متبادل Interaction بين المتغيرين؟

مثال تطبيقي Illustrative Example

أجرت شركة دوائية دراسة لمعرفة تأثير الجنس على مجموعة من الأدوية التي تنتجها و المخصصة لمعالجة مرض السرطان. لأجل ذلك قام خبراء تحليل البيانات بتجريب ثلاثة أنواع من الأدوية (Drug) وهي: cisplatin و vinblastine و 5-fluorouracil على عينات من الذكور Male و الإناث Female و قاموا بقياس حجم الورم (Tumor) المتبقي بعد المعالجة. يظهر الجدول (3) هذه القياسات.

Drug:	Cisplatin		Vinblastine		5-fluorouracil	
Gender:	Female	Male	Female	Male	Female	Male
Tumor Size:	65	50	70	45	55	35
	70	55	65	60	65	40
	60	80	60	85	70	35
	60	65	70	65	55	55
	60	70	65	70	55	35
	55	75	60	70	60	40
	60	75	60	80	50	45
	50	65	50	60	50	40

جدول (3): حجم الورم بعد المعالجة و لكل من الذكور و الإناث

قبل البدء بمعالجة هذه المسألة يجب ملاحظة أن هذه المسألة تطابق النموذج العام لمسائل اختبار تحليل التباين بمتغيرين.

- المتغير التابع (أو الظاهرة المدروسة) هو قياس حجم الورم **Tumor Size**.
- المتغير المستقل الأول هو نوع الدواء **Drug** و له ثلاثة مستويات تأثير تمثل الأدوية المختلفة.
- المتغير المستقل الثاني هو جنس المريض **Gender** و له مستويان اثنان من التأثير يمثلان جنس المريض.

سوف نقدم شرحاً مفصلاً عن كيفية تطبيق هذا الاختبار على هذه المسألة النموذجية و كيف يتم تفسير النتائج. يجب الانتباه إلى أن جزءاً من الأرقام التي تظهر في جداول SPSS غير مهمة من ناحية تفسير النتائج و بالتالي سوف نتجاهلها.

نتائج الاختبار باستخدام الـ SPSS

يظهر الشكل (7) الإحصائيات اللازمة لاختبار فرضية تساوي التباينات.

Levene's Test of Equality of Error Variances^(a)

Dependent Variable: Diameter

F	df1	df2	Sig.
1.462	5	42	0.223
Tests the null hypothesis that the error variance of the dependent variable is equal across groups.			
a Design: Intercept + Gender + Drug + Gender * Drug			

الشكل (7): جدول اختبار تساوي التباينات.

كما رأينا في الفقرة السابقة أن فرضية العدم تنص على تساوي التباينات بين مختلف المجموعات المشكلة نتيجة للعاملين **Gender** و **Drug**. تشير المخرجات في الشكل (7) أن الاختبار غير معنوي لأن معنوية الاختبار "0.223" و هي أكبر من مستوى

الأهمية المفروض "0.05". لذلك نقبل فرضية العدم و بالتالي تباينات المجموعات متساوية. لاحظ أنه في هذه الحالة فإن شرط تساوي التباينات محقق و بالتالي لا حاجة لإجراء اختبارات بديلة كما فعلنا في الفقرة السابقة.

هناك نقطة مهمة يجب التنويه إليها و هي أن مصادر التباين في هذا النوع من التجارب كما يبدو في الشكل (7) مبينة في النموذج:

a Design: Intercept + Gender + Drug + Gender * Drug

أي أن مصادر التباين هي الثابت Intercept ، جنس المريض Gender ، نوع الدواء Drug ، و التأثير المتبادل بين الجنس و نوع الدواء Gender * Drug.

يظهر الشكل (8) نتائج تحليل التباين. هناك العديد من الأرقام التي تظهر في هذا الجدول ، و لكن سنركز فقط على الأرقام التي سنستخدمها لاختبار فرضيات المسألة.

المجموعة الأولى من الفرضيات تتعلق بتأثير متغير الجنس.
فرضية العدم H_0 : متوسط قطر الأورام بعد المعالجة لدى الذكور مساوٍ لمتوسط قطر الورم بعد المعالجة لدى الإناث.
فرضية البديلة H_1 : متوسط قطر الأورام بعد المعالجة لدى الذكور لا يساوي متوسط قطر الورم بعد المعالجة لدى الإناث.

لاختبار هذه الفرضيات نستخدم معنوية الاختبار في الشكل (8) المقابلة ل Gender. نلاحظ أن الاختبار ليس معنوياً لأن قيمة "0.451" أكبر من مستوى الأهمية المفروض "0.05". و بالتالي نقبل فرضية العدم H_0 و نستنتج أن متوسط أقطار الأورام بعد المعالجة متساوٍ لدى الذكور و الإناث. و يُفسّر هذا على أن الدواء يؤثر بشكل متساوٍ لدى الجنسين.

المجموعة الثانية من الفرضيات تتعلق بتأثير متغير أنواع الأدوية.
فرضية العدم H_0 : متوسط قطر الأورام بعد المعالجة متساوية بالنسبة إلى كل أنواع الأدوية.
الفرضية البديلة H_1 : على الأقل متوسط قطر الأورام المعالجة بأحد أنواع الأدوية مختلف.

بالعودة إلى الشكل (8) نجد أن اختبار الفرضيات المتعلقة بتأثير أنواع الأدوية معنوي. لاحظ أن قيمة معنوية الاختبار Sig المقابلة لمتغير الأدوية Drug تساوي الصفر. هذا يعني أن متوسطات أقطار الأورام بعد معالجتها بأدوية مختلفة ليست متساوية.

المجموعة الثالثة من الفرضيات تتعلق بالتأثير المتبادل بين متغير نوع الأدوية Drug و متغير الجنس Gender.

فرضية العدم H_0 : لا يوجد تأثير متبادل (تفاعل) بين المتغيرين.

الفرضية البديلة H_1 : يوجد تأثير متبادل بين المتغيرين.

بالنظر إلى قيمة معنوية الاختبار المقابلة للتفاعل $\text{Gender} * \text{Drug}$ نجد أن هذه القيمة معنوية "0.000". هذا يعني أن هناك تفاعلاً بين المتغيرات المدروسة؛ قبول الفرضية البديلة H_1 . أي أن أدوية مختلفة تملك تأثيرات مختلفة على الإناث و الذكور من حيث تقليصها لقطر الورم.

بسبب وجود التفاعل بين متغير الجنس و متغير نوع الدواء فإنه يجب أن ننظر إلى نتائج اختبارات الفرضيات حول تأثير نوع الدواء و الجنس بحدٍ. و السبب هو أن فعالية الدواء ربما تختلف باختلاف الجنس. بكلمات أخرى ربما تكون هناك أدوية تؤثر بشكل أفضل في الإناث ، و هناك أدوية تؤثر بشكل أفضل في الذكور. بناءً على ذلك، فإنه يبدو قولنا إن الجنس لا يملك تأثيراً (كما توصلنا سابقاً) محل شك.

يوضح الشكل (9) بيانياً التفاعل بين الجنس و نوع الدواء. لاحظ أنه في حالة الذكور فإن تأثيرات الدواء تختلف. حيث الدواءان Cisplatin و Vinplastine لهما تقريباً نفس التأثير في الذكور و الإناث من حيث تقليص حجم الورم. لكن تأثير الدواء 5-Flouroucil يختلف بشكل كبير بين الذكور و الإناث. حيث إن هذه الدواء يقلص متوسط حجم الورم بشكل أكبر بكثير عند الذكور منه عند الإناث. لاحظ أيضاً أنه لا يوجد اختلاف يذكر بين تأثيرات الأدوية الثلاثة في الإناث. كما هو واضح من الشكل

فإن متوسطات أقطار الورم تكاد تكون متساوية بين الأدوية الثلاثة مما يعني أنه لا يوجد فرق بين تأثير هذه الأدوية في الإناث.

بما أن متوسطات حجوم الأورام مختلفة معنوياً عبر مستويات الدواء فإن الخطوة التالية هي معرفة أي من المتوسطات يكون مختلفاً بشكل معنوي عن باقي المتوسطات. لمعرفة ذلك نستخدم اختباري Tukey و Games-Howell كما ناقشنا في الفقرة السابقة.

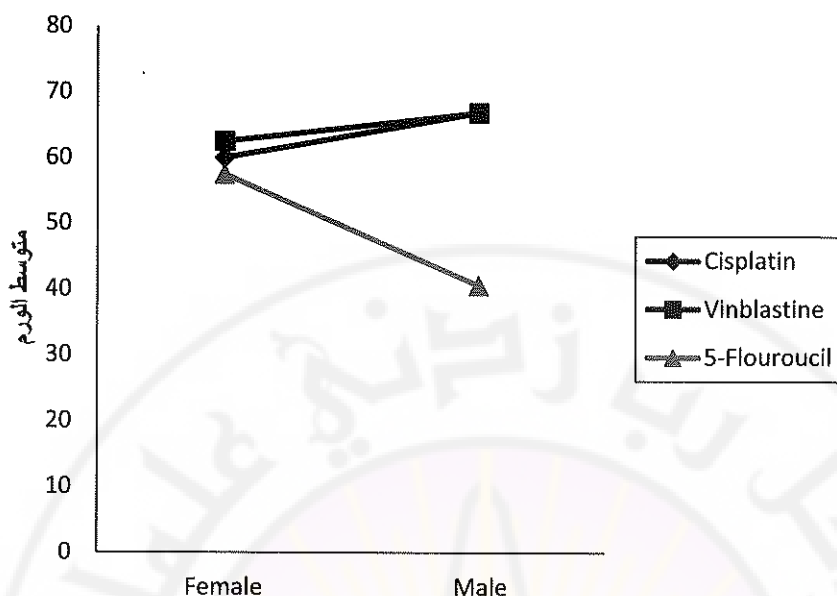
Tests of Between-Subjects Effects Dependent Variable: Diameter

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	3817.188(a)	5	763.438	10.459	.000
Intercept	167442.188	1	167442.188	2294.9	.000
Gender	42.188	1	42.188	.578	.451
Drug	2412.500	2	1206.250	16.526	.000
Gender * Drug	1362.500	2	681.250	9.333	.000
Error	3065.625	42	72.991		
Total	174325.000	48			
Corrected Total	6882.813	47			

a R Squared = .555 (Adjusted R Squared = .502)

التفاعل بين الجنس و نوعية

الشكل (8): جدول تحليل التباين.



الشكل (9): التمثيل البياني للتفاعل بين الجنس (ذكور، إناث) و أنواع الدواء الثلاثة (Cisplatin, Vinblastine, 5-Flourouracil). لاحظ أن تأثيرات الأدوية الثلاثة متقاربة بالنسبة للإناث Female من حيث تقليلها لحجم الورم. لاحظ أيضاً أن الأدوية الثلاثة تملك تأثيرات مختلفة على الذكور Male. من الواضح أن الدواء 5-Flourouracil يعطي أفضل نتيجة من حيث تقليل حجم الورم عند الذكور.

يظهر الشكل (10) نتائج اختباري Tukey و Games-Howell. لاحظ تطابق النتائج بين هذين الاختبارين و ذلك لأن جميع شروط اختبار Tukey محققة. (بالنسبة لاختبار Games-Howell فإنه لا يتأثر بتساوي التباينات أو عدم تساويها.) نلاحظ أن المتوسط المختلف هو متوسط قطر الأورام التي تمت معالجتها بوساطة الدواء 5-flourouracil حيث إن متوسط حجم الورم المتبقي بعد المعالجة بهذا الدواء هو أصغر بشكل معنوي من متوسطات حجوم الأورام المتبقية بعد المعالجة بـ vinblastine و cisplatin. كذلك لا يلاحظ وجود فرق معنوي إحصائياً بين متوسطات حجوم الأورام بعد المعالجة بالدوائين vinblastine و cisplatin.

Multiple Comparisons
Dependent Variable: Diameter

	(I) Drug	(J) Drug	Mean Difference (I – J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Tukey HSD	Cisplatin	Vinblastine	-1.25	3.021	.910	-8.59	6.09
		5-flourouracil	14.38(*)	3.021	.000	7.04	21.71
	Vinblastine	Cisplatin	1.25	3.021	.910	-6.09	8.59
		5-flourouracil	15.63(*)	3.021	.000	8.29	22.96
	5-flourouracil	Cisplatin	-14.38(*)	3.021	.000	-21.71	-7.04
		Vinblastine	-15.63(*)	3.021	.000	-22.96	-8.29
Games-Howell	Cisplatin	Vinblastine	-1.25	3.329	.925	-9.46	6.96
		5-flourouracil	14.38(*)	3.534	.001	5.64	23.11
	Vinblastine	Cisplatin	1.25	3.329	.925	-6.96	9.46
		5-flourouracil	15.63(*)	3.699	.001	6.50	24.75
	5-flourouracil	Cisplatin	-14.38(*)	3.534	.001	-23.11	-5.64
		Vinblastine	-15.63(*)	3.699	.001	-24.75	-6.50
Based on observed means.							
* The mean difference is significant at the .05 level.							

الشكل(10): نتائج اختباري Tukey و Games-Howell حول المقارنات المتعددة بين متوسطات حجوم الأورام بعد المعالجة بأنواع الأدوية المختلفة Drug.

كتابة التقرير Result Reporting

بناءً على العينة المتوفرة ، إن متغير الجنس ليس له دور في فعالية الدواء: متوسط حجوم الأورام بعد المعالجة بالأدوية المختلفة غير مختلف بشكل معنوي عند

الذكور و الإناث و بخطأ مقداره 5%. من جهة أخرى ، نلاحظ أن هناك فرقاً معنوياً بين متوسطات حجوم الأورام بالنسبة إلى الأدوية المختلفة (cisplatin, Drug) و (vinblastine, and 5-flourourcile) و بخطأ مقداره 5%. نلاحظ أيضاً أن هناك تأثيراً متبادلاً بين المتغيرين المستقلين Drug و Gender ، و هذا يدل على أن متوسطات حجوم الأورام بعد المعالجة بأدوية مختلفة غير متساوية فيما يتعلق بمتغير الجنس Gender. و بالنتيجة، يمكن القول إن الأدوية المختلفة Drug تؤثر بشكل مختلف في الجنسين. كما لاحظنا من نتائج التحليل أن الأدوية الثلاثة تؤثر في الإناث بالدرجة نفسها تقريباً بينما الدواء 5-flouroucile يؤثر في الذكور بشكل أفضل من حيث تقليص حجم الورم.

ملاحظات توجيهية Teaching Notes

هناك نقاط يجب التركيز عليها عند إجراء اختبار Two-Way ANOVA.

1. يجب أن لا يفهم خطأ أن مستويات المتغيرين المستقلين هي مقدرات لقيم التابع بل هي مؤثرات فقط تؤثر في قيم (سلوك) المتغير التابع.
2. قبل البدء بتفسير جدول تحليل التباين، يجب الانتباه إلى نتيجة تحليل تجانس التباين. إذا كان التباين متجانساً بين كل الطبقات، عندئذٍ نستخدم جدول تحليل التباين لأخذ قرار حول الفرضيات المدروسة و إلا نجري اختباري Welch, Brown-Forsythe لأنهما لا يتأثران بتجانس التباين بين الطبقات.
3. يجب الانتباه إلى وجود التفاعل. إذا كان التفاعل موجوداً بين المتغيرات المستقلة في تأثيرها في المتغير التابع ، فإن يجب إعادة التفكير في نتائج اختبار الفرضيات المتعلقة بكل متغير على حدة. قد يحجب التفاعل تأثير أحد المتغيرات كما شاهدنا في حالة متغير الجنس في المثال أعلاه. حيث لاحظنا أن الذكور هم أكثر استجابة للدواء 5-flouroucile.

4. إذا تم قبول فرضية العدم H_0 تنتهي المسألة. أما إذا تم رفض فرضية العدم عندئذٍ نستخدم اختباري Tukey و Games-Howell لاكتشاف المتوسط (أو المتوسطات) المختلف.

3.2.5.5 اختبار تحليل التباين للعينات المرتبطة

Repeated Measures ANOVA

يستخدم مصطلح القياسات المتكررة للإشارة إلى أن أفراد العينة أنفسهم يشتركون في كل ظروف تجربة معينة. بكلمات أبسط ، تُجرى عمليات مختلفة ضمن التجربة على نفس أفراد العينة. على سبيل المثال ، يمكن تقييم العيارات المختلفة لدواء على مجموعة من المرضى. من المهم في هذا النوع من التجارب ضبط الفروق الفردية في العيارات المختلفة. بعض الناس يستجيبون لتراكيز دواء منخفضة بينما البعض الآخر يحتاج إلى عيار أعلى للإستجابة. لضبط الفروق الفردية يمكن أن نختبر نفس أفراد العينة في ظروف التجربة المختلفة (عيارات الدواء المختلفة). بكلمات أبسط، فإننا نختبر استجابة كل فرد من أفراد العينة بعد إعطائهم كل عيار من عيارات الدواء.

توضيحاً لذلك فإن، نفس أفراد العينة الذين خضعوا للإجراءات المختلفة وبالتالي لن تكون القياسات الناتجة بعد كل إجراء مستقلة عن بعضها البعض. هذا يخالف شرط الاستقلالية بين العينات الذي تتطلبه طرق تحليل التباين لعينات مستقلة. بسبب كون القياسات الناتجة عن الإجراءات المختلفة مرتبطة Related فإن معالجة هذه البيانات بشكل صحيح يقتضي بعض التعديلات على اختبارات التباين للعينات المستقلة. بالتحديد يجب أن نضع شرطاً إضافياً إلى تلك الخاصة بتحليل التباين و ينص هذا الشرط على أن العلاقة بين أزواج القياسات متشابهة (أي قيمة الارتباط متساوية تقريباً). نسمي فرضية تساوي الارتباطات بين أزواج القياسات بـ **Sphericity**.

يمكن ربط فرضية تساوي الارتباطات أو Sphericity إلى فرضية تجانس التباين Homogeneity of Variance على الشكل التالي: إذا أخذنا كل زوج من المشاهدات

و حسبنا الفروق بين القياسات المتقابلة ، عندئذٍ من الضروري أن يكون تباينات الفروقات متساوية احصائياً.

بالعودة إلى شرط تساوي الترابطات أو Sphericity ، إن عدم تحقق هذا الشرط يؤدي إلى اضعاف حساسية الاختبار مما يزيد من احتمال الوقوع في الخطأ من النوع الثاني (قبول خاطئ لفرضية العدم H_0). بالنتيجة تصبح قيمة F المحسوبة غير قابلة للمقارنة مع F النظرية أو بحسب مصطلحاتنا في هذا الكتاب تصبح معنوية الاختبار غير دقيقة مما يجعل المقارنة مع مستوى الأهمية α غير صحيح.

هناك اختبار شهير للتأكد من تحقق شرط تساوي الترابطات يسمى باختبار Mauchly's test. الفرضيات المُختبرة هي:

فرضيات Mauchly's test

شرط تساوي الترابطات (Sphericity) محقق: H_0

شرط تساوي الترابطات غير محقق: H_1

إذا كان اختبار Mauchly معنوياً عند مستوى أهمية α (أي معنوية الاختبار Sig أصغر من α) فإن شرط تساوي التباينات غير محقق. أما إذا كان الاختبار غير معنوي فإن شرط تساوي الارتباطات محقق.

إذا لم يكن شرط تساوي الارتباطات محققاً فلا بد من تصحيح قيمة F المحسوبة لكي نتمكن من استخدام معنوية الاختبار المقابلة لها. يتم تصحيح قيمة F من خلال تقليص درجات الحرية و يجب أن يتناسب هذا التقليص مع درجة مخالفة شرط تساوي الارتباطات. هناك ثلاثة تقديرات مختلفة لشرط تساوي الارتباطات Sphericity التي تستخدم لتصحيح درجات الحرية.

1. تصحيح: Greenhouse and Geisser

2. تصحيح: Huynh and Feldt

3. تصحيح: The Lower Bound estimate

نعمد لإستخدام أي من هذه التصحيحات الثلاثة على معيار اسمه ϵ الذي يظهر في جدول نتائج اختبار Mauchly's test.

« إذا كانت $\epsilon > 0.75$ ، استخدم تصحيح Huynh and Feldt

« إذا كان $\epsilon < 0.75$ ، استخدم تصحيح Greenhouse and Geisser.

بعد هذه الفكرة النظرية العامة حول تحليل التباين للعينات المرتبطة ، سوف نعطي أمثلة لتحليل التباين بمتغير مستقل واحد و بمتغيرين مستقلين.

تحليل التباين بمتغير واحد للعينات المرتبطة

One-way Repeated Measures ANOVA

النموذج العام للمسائل التي تتبع هذا النوع من الإختبارات مماثلة للنموذج العام للمسائل التي تتبع إختبارات التباين بمتغير واحد مستقل و للعينات المستقلة. الفرق الجوهرى هو أن الوحدات التجريبية تبقى نفسها في كل شروط التجربة.

النموذج العام

لدينا مجموعة من الوحدات التجريبية التي تخضع إلى تجربة مؤلفة من عدة مراحل و نريد معرفة هل المراحل المختلفة تؤثر بشكل مختلف على الوحدات التجريبية. مثال توضيحي: نريد دراسة تأثير مادة منشطة على أداء مجموعة من الرياضيين. تتألف المادة المنشطة من ثلاثة عيارات مختلفة. تم إعطاء المادة المنشطة لنفس الرياضيين على مدى ثلاثة أيام ، يوم لكل عيار. المتغير المستقل هو المادة المنشطة أما المتغير التابع هو أداء الرياضيين.

لنفرض أننا نريد معرفة هل يتعلق عدد الساعات التي يقضيها القراء يومياً بنوع المادة المقروءة. من أجل ذلك تم اختيار عينة من ثمانية قراء و سُجلت عدد الساعات

التي يقضونها بشكل يومي من أجل أربعة أنواع لمواد القراءة: خيال علمي Science fiction ، روايات Novels ، مواد تخصصية Major ، و أخيراً أخبار رياضية Sports. الجدول التالي يحتوي على عدد الساعات التي قضاها كل قارئ من القراء الثمانية بشكل متواصل.

القراء	Sports	Major	Novels	Science fiction
1	6	1	7	8
2	5	2	5	9
3	8	3	2	6
4	9	1	3	5
5	8	5	4	8
6	7	6	5	7
7	2	7	2	10
8	1	8	6	12

ملاحظة: يتم إدخال البيانات في شاشة الـ SPSS بتعريف متغير (عمود) لكل نوع قراءة.

يوفر الجدول (4) وصفاً احصائياً للمستويات المختلفة لمتغير نوع المادة المقروءة. نستطيع القول اعتماداً على قيم الجدول أن متوسط Mean عدد الساعات التي يقضيها القراء الثمانية في قراءة الكتب التخصصية هي الأقل (4.13) و أن متوسط عدد الساعات التي يقضيها القراء الثمانية في قراءة قصص الخيال العلمي هو الأكبر (8.13).

Descriptive Statistics

	Mean	Std. Deviation	N
Science Fiction	8.13	2.232	8
Novels	4.25	1.832	8
Major	4.13	2.748	8
Sports	5.75	2.915	8

الجدول (4): وصف احصائي للبيانات.

تقييم شرط تساوي الارتباطات Sphericity

يظهر الجدول (5) نتائج اختبار Mauchly's test حول تساوي الارتباطات. القيم في العمود Sig و مجموعة القيم المعنونة بـ Epsilon لها أهمية استثنائية كونها تحدد هل شرط تساوي الارتباطات محقق (القيم في العمود Sig) و مقدار مخالفة شرط تساوي الارتباطات إذا كان الشرط غير محقق (القيم المعنونة بـ Epsilon).

Mauchly's Test of Sphericity^b

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi- Square	df	Sig.	Epsilon		
					Greenhouse- Geisser	Huynh- Feldt	Lower- bound
Topic	.136	11.406	5	.047	.533	.666	.333

الجدول (5): مخرجات الـ SPSS لنتائج اختبار Mauchly.

كما يبدو من الجدول فإن معنوية الإختبار $\text{Sig} = 0.047$. يشير هذا إلى أن الإختبار معنوي كون معنوية بمستوى أهمية "0.05". و بالنتيجة نقبل الفرضية البديلة التي تنص على عدم تحقق شرط تساوي الارتباطات (أي إن تباينات الفروق بين مستويات المتغير المستقل مختلفة معنوياً).

يظهر الجدول (6) نتائج تحليل التباين. تم تقسيم الجدول إلى مقاطع لتسهيل عملية الشرح و المتابعة من قبل القارئ.

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	
Topic	Sphericity Assumed	83.125	3	27.708	3.794	.026	الشرط محقق
	Greenhouse-Geisser	83.125	1.599	52.001	3.794	.063	شرط الارتباط غير محقق
	Huynh-Feldt	83.125	1.997	41.619	3.794	.048	
	Lower-bound	83.125	1.000	83.125	3.794	.092	
Error (Topic)	Sphericity Assumed	153.375	21	7.304			
	Greenhouse-Geisser	153.375	11.190	13.707			
	Huynh-Feldt	153.375	13.98	10.970			
	Lower-bound	153.375	7.000	21.911			

الجدول (6): مخرجات الـ SPSS لنتائج تحليل التباين.

المقطع الأول هو نتائج تحليل التباين تحت فرضية أن شرط تساوي الارتباطات محقق "Sphericity Assumed". يمكن قراءة المعلومات في الجدول بسهولة. مجموع المربعات Sum of Squares (83.125) المقابل للمتغير Topic (المتغير المستقل) هو كمية التباين التي يمكن تفسيرها بوساطة مستويات هذا المتغير. أما مجموع مربعات المقابلة للخطأ Error (153.375) فهي عبارة عن كمية التباين التي لا يمكن تفسيرها بوساطة مستويات المتغير المستقل. متوسط المربعات Mean Squares يساوي مجموع المربعات مقسوماً على عدد درجات الحرية. قيمة F هي متوسط التباينات المفسرة مقسوماً على متوسط التباينات غير المفسرة $\frac{27.798}{7.304} = 3.794$. معنوية الاختبار Sig = 0.026. كما يتضح من الجدول فإن معنوية الاختبار تحت فرضية تساوي الارتباطات هي Sig = 0.026 و التي تعني أن

الاختبار معنوي بمستوى أهمية "0.05". و مع ذلك فإن نتائج هذا المقطع لا يعول عليها كون شرط Sphericity غير محقق.

لحسن الحظ يمكن تصحيح قيمة F لكي تصبح قابلة للاستخدام عندما يكون شرط Sphericity غير محقق و ذلك عن طريق تصحيح درجات الحرية. كما نلاحظ من المقطع الثاني في الجدول (الجزء المظلل) ، فإن قيمة F لم تتغير و لكن تم تصحيح درجات الحرية اعتماداً على الإجراءات الثلاثة المبينة في المقطع الثاني. بسبب التغير في درجات الحرية فإن القيمة المعنوية المقابلة ستتغير أيضاً. بسبب وجود ثلاثة تصحيحات فإننا يجب أن نأخذ واحداً فقط من هذه التصحيحات و يتم ذلك اعتماداً على المعيار ϵ الذي انشأناه سابقاً.² بما أن جميع قيم ϵ أصغر من 0.75 فإننا ، و بحسب المعيار أعلاه ، نستخدم قيمة المعنوية المقابلة لتصحيح Greenhouse-Geisser. بما أن قيمة المعنوية المقابلة لهذا التصحيح هي "0.063" فإن الاختبار ليس معنوياً بمستوى أهمية "0.05". و بالتالي لا يوجد فرق معنوي احصائياً بين متوسط الساعات التي يقضيها القارئ يومياً في قراءة الأنواع الأربعة.

تحليل التباين بمتغيرين مستقلين للعينات المرتبطة

Two-Way Repeated Measures ANOVA

يدرس هذا الشكل من تحليل التباين تأثير مستويات متغيرين مستقلين على وحدات تجريبية و التأثير المتبادل (التفاعل) لهذين المتغيرين في الوحدات التجريبية. الوحدات التجريبية ذاتها تخضع لكل ظروف التجربة الممثلة بتقاطع المستويات المختلفة. على سبيل التوضيح يمكن للمتغيرين المستقلين أن يكونا نوع الدواء (نباتي، كيميائي) و الحماية (شديدة أو متوسطة أو خفيفة) و الدراسة تتمحور حول دراسة تأثير نوع الدواء ودرجات الحماية على تعافي المرضى و البحث عن أي تفاعل بينهما. نفس المرضى يخضعون لنوعي الدواء و درجات الحماية الثلاث.

² إن ϵ هو تقدير الارتباط Sphericity.

الفرضيات الأساسية

المجموعة الأولى: هل تؤثر مستويات المتغير الأول بشكل مختلف في الوحدات التجريبية؟

حيث الفرضيات:

H_0 : تؤثر مستويات العامل الأول بشكل متساوٍ

H_1 : هناك اختلاف بتأثير على الأقل مستوى واحد

المجموعة الثانية: هل تؤثر مستويات العامل الثاني بشكل مختلف في الوحدات التجريبية؟ و الفرضيات هنا مشابهة لتلك الخاصة بالمتغير الأول.

المجموعة الثالثة: هل هناك تأثير متبادل (تفاعل) بين مستويات المتغيرين في تأثيرهما على الوحدات التجريبية؟
حيث الفرضيات:

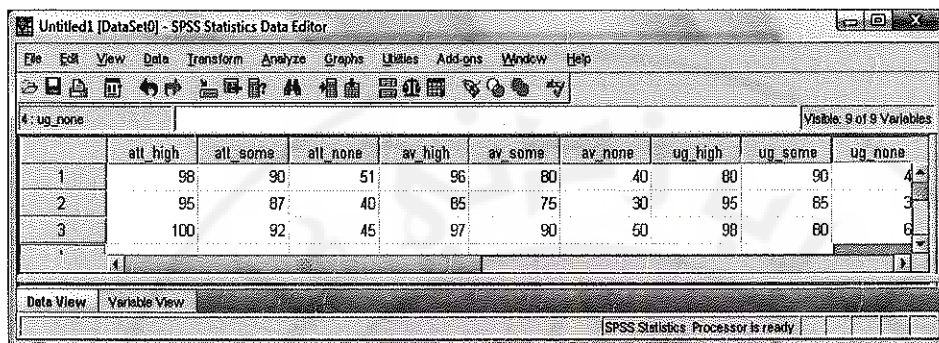
H_0 : لا يوجد تأثير متبادل بين مستويات المتغيرين

H_1 : يوجد تأثير متبادل بين مستويات المتغيرين

مثال

في دراسة أجريت لمعرفة أي نوع من الأفلام يثير اهتمام الشباب أكثر. لأجل ذلك تم أخذ تسع أفلام قُسمت إلى ثلاثة مجموعات كل مجموعة تحوي ثلاث أفلام. المجموعة الأولى تحوي ثلاث أفلام ترفيهية Entertainment تختلف بطول العرض: الأول طويل Lengthy ، الثاني متوسط Average ، أما الثالث فكان قصير Short. المجموعة الثانية تحوي ثلاث أفلام وثائقية Documentary تختلف أيضاً بطول العرض. المجموعة الأخيرة تحوي ثلاث أفلام علمية Sceintific تختلف بطول العرض (من طويل إلى قصير). تم إختيار عشرة شباب من الجنسين و طلب من كل شاب

حضور الأفلام في المجموعات الثلاثة لمدة خمسة دقائق لكل فلم من كل مجموعة، ثم طلب منهم إعطاء نسبة مئوية تحدد مدى اهتمام الشباب بهذه الأفلام. تبين شاشة الـ SPSS التالية جزء من البيانات.



SPSS Statistics Data Editor - Untitled1 [DataSet0]

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

4: ug_none Visible: 9 of 9 Variables

	alt_high	all_some	all_none	av_high	av some	av none	ug_high	ug some	ug none
1	98	90	51	96	80	40	80	90	4
2	95	87	40	85	75	30	95	65	3
3	100	92	45	97	90	50	98	80	6

Data View Variable View

SPSS Statistics Processor is ready

نستطيع أن نرى من المسألة أن هناك متغيرين مستقلين الأول هو نوع الفيلم -File Type (ترفيهي Entertaining ، وثائقي Documentary ، وأخيراً علمي Scientific) و المتغير الثاني طول العرض Duration (طويل lengthy، متوسط Average، قصير Short). الوحدات التجريبية هي الشباب العشرة الذين شاركوا في كل ظرف من ظروف التجربة.

نتائج الاختبار

يوفر الجدول (7) وصفاً إحصائياً للبيانات. يتم الوصف من خلال المتوسط Mean، الانحراف المعياري Std. deviation، و عدد أفراد العينة N. يوحي وصف العينة أن لدى الشباب ميلاً أكبر للأفلام الترفيهية Entertaining و الطويلة lengthy لأن متوسط النسبة المئوية لدرجة اهتمام الشباب بهذا النوع من الأفلام هي الأعلى 89.60 %. إضافةً إلى ذلك يُبين وصف البيانات أن الشباب يفضلون طول العرض على نوع الفيلم حيث أن النسب المئوية عالية عند الأفلام الطويلة بغض النظر عن نوع الفيلم و تنخفض هذه النسبة بشكل كبير عندما يكون المشاهد هو فيلم قصير. لاحظ على سبيل المثال أن النسبة المئوية المقابلة للفيلم الترفيهي و لكن القصير هي 51.8% بينما النسبة المئوية المقابلة للفيلم العلمي و لكن الطويل هي 86.7%.

Descriptive Statistics

	Mean	Std. deviation	N
Entertaining and Lengthy	89.60	6.637	10
Entertaining and Average	87.10	6.806	10
Entertaining and Short	51.80	3.458	10
Documentary and Lengthy	88.40	8.329	10
Documentary and Average	68.90	5.953	10
Documentary and Short	47.00	3.742	10
Scientific and Lengthy	86.70	5.438	10
Scientific and Average	51.20	5.453	10
Scientific and Short	46.10	3.071	10

الجدول (7): وصف احصائي للعينات.

يُظهر الجدول (8) مخرجات الـ SPSS لإختبار Mauchly حول المؤثرات الثلاثة: نوع الفيلم Film، طول الفيلم Duration، والتفاعل بينهما Film* Duration.

Mauchly's Test of Sphericity

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Film	.904	.810	2	.667	.912	1.000	.500
Duration	.851	1.292	2	.524	.870	1.000	.500
Film * Duration	.046	22.761	9	.008	.579	.791	.250

الجدول (8): نتائج اختبار Mauchly.

لاحظ من الجدول (العمود Sig) أن إختبار Mauchly غير معنوي بمستوى أهمية 0.05 بالنسبة للمتغير Film و Duration و بالتالي فإن شرط تساوي الارتباطات محقق — لا حاجة لتصحيح قيمة F. لاحظ أيضاً من اختبار Mauchly معنوي بالنسبة للتفاعل بين المتغيرين و بالتالي فإن شرط تساوي الارتباطات غير محقق — يجب تصحيح قيمة F.

يوفر الجدول (9) نتائج اختبار تحليل التباين للعينات المترابطة مع القيم المصححة ل F. ينقسم الخرج إلى ثلاثة أقسام تقابل كل متغير من متغيرات النموذج (, Film Duration) و التفاعل بين هذين المتغيرين (Film * Duration). لاحظ من الجدول أن تأثير كل من المتغيرين و التفاعل بينهما معنوي بمستوى الأهمية المفروض "0.05". لندرس تأثير كل مؤثر على حدة و بشيء من التفصيل.

تأثير متغير نوعية الفيلم Film

كما يبدو من الجدول (9) أن هناك تأثيراً معنوياً للمتغير Film لأن معنوية الاختبار تساوي "0.000" و هي أصغر من مستوى الأهمية 0.05. يمكن تفسير هذه المعنوية للمتغير Film على أنه ، و باهمال المتغير الآخر Duration ، فإن متوسطات الاستجابات لمستويات المتغير مختلفة معنوياً. بكلمات أبسط ، الشباب يقيمون مستويات المتغير Film بدرجات مختلفة.

Tests of Within-Subjects Effects

Source		Type III Sum of Squares	Df	Mean Square	F	Sig
Films	Sphericity Assumed	3308.87	2	1664.4	6.4	.000
	Greenhouse-Geisser	3308.87	1.82	1813.7	6.4	.000
	Huynh-Feldt	3308.87	2.00	1654.4	6.4	.000
	Lower-bound	3308.87	1.00	3308.8	6.4	.000
Error (Films)	Sphericity Assumed	448.24	8	24.90		
	Greenhouse-	448.24	6.42	27.30		

	Geisser					
	Huynh-Feldt	448.24	8.00	24.90		
	Lower-bound	448.24	9.00	49.80		
Duration	Sphericity Assumed	23932.87		11963.4	74.9	.000
	Greenhouse-Geisser	23932.87	.74	13751.5	74.9	.000
	Huynh-Feldt	23932.87	.00	11966.4	74.9	.000
	Lower-bound	23932.87	.00	23932.8	74.9	.000
Error (Duration)	Sphericity Assumed	783.58	8.00	43.53		
	Greenhouse-Geisser	783.58	5.66	50.03		
	Huynh-Feldt	783.58	8.00	43.53		
	Lower-bound	783.58	9.00	87.06		
Film* Duration	Sphericity Assumed	3365.87	4	841.47	4.9	.000
	Greenhouse-Geisser	3365.87	2.32	1453.6	4.9	.000
	Huynh-Feldt	3365.87	.16	1036.6	4.9	.000
	Lower-bound	3365.87	.00	3365.8	4.9	.000
Error	Sphericity Assumed	867.69	6	24.10		
	Greenhouse-Geisser	867.69	0.84	31.64		
	Huynh-Feldt	867.69	8.48	30.46		
	Lower-bound	867.69	.00	96.41		

الجدول (9): نتائج تحليل التباين للعينات المترابطة.

سؤال: أي قيمة لـ Sig نأخذها بالنسبة للمتغير Film؟ هل هي القيمة المقابلة لـ Sphericity Assumed؟ أم القيمة المقابلة لأحد التصحيحات الثلاثة الأخرى؟ أي تصحيح يجب أن يُؤخذ؟

يوفر الجدول (10) معلومات عن المتوسط ، الخطأ المعياري في الوسط ، و

95% مجال ثقة حول المتوسط. مستويات المتغير Film عُرضت بدلالة الأرقام 1 و 2

و 3 التي تقابل على الترتيب ترفيهي Entertaining، وثائقي Documentary، و أخيراً علمي Scientific. من المهم جداً ملاحظة كيف ينقص المتوسط بالانتقال من فيلم ترفيهي إلى فيلم وثائقي إلى فيلم علمي: المتوسط المقابل للمستوى 1 (ترفيهي) هو الأعلى يليه المتوسط المقابل للمستوى 2 ثم المتوسط المقابل للمستوى 3. يمكن من خلال مجالات الثقة أيضاً ملاحظة أن هذه المتوسطات غير متساوية. تعكس النتائج أن الشباب يفضلون الأفلام الترفيهية أكثر من غيرها.

Estimates

Measure: MEASURE_1

Film	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	76.167	1.013	73.876	78.457
2	68.100	1.218	65.344	70.856
3	61.333	1.018	59.030	63.637

الجدول (10): بعض الاحصائيات حول المتغير Film

تأثير متغير طول الفيلم Duration

نستنتج من خلال الجدول (9) ذات النتيجة التي حصلنا عليها من أجل متغير نوعية الفيلم Film. بسبب كون الاختبار معنوياً فإننا نرفض فرضية تساوي متوسطات الاستجابات للمستويات الثلاثة (فرضية العدم) و بالتالي نقبل فرضية وجود متوسط واحد على الأقل مختلف. بكلمات أبسط، الشباب يقيمون مستويات طول الفيلم بدرجات مختلفة.

يوفر الجدول (11) معلومات عن المتوسط ، الخطأ المعياري في الوسط ، و 95% مجال ثقة حول المتوسط. مستويات المتغير Duration عُرضت بدلالة الأرقام 1 و 2 و 3 التي تقابل على الترتيب فيلم طويل Lengthy، فيلم متوسط Average، و أخيراً فيلم قصير Short. لاحظ أن هناك نزوعاً لهبوط المتوسطات مع التغير من

المستوى Lengthy إلى Short. يمكن من خلال مجالات الثقة أيضاً ملاحظة أن هذه المتوسطات غير متساوية. تعكس النتائج أن الشباب يفضلون الفيلم الطويل أكثر من الفيلم المتوسط أو القصير.

Estimates

Measure: MEASURE_1

Duration	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	88.233	1.598	84.619	91.848
2	69.067	1.293	66.141	71.991
3	48.300	0.751	46.601	49.999

الجدول (11): احصائيات حول المتغير Duration

التفاعل بين نوعية الفيلم و طول العرض

تبين النتائج في الجدول (9) أن هناك تفاعلاً بين طول الفيلم Duration و نوعية الفيلم Film لأن معنوية الاختبار المقابلة للتفاعل (المتغير Film * Duration) تساوي "0.000" و هي تقل عن مستوى الأهمية 5%. تذكر أن شرط Sphericity غير محقق بالنسبة للتفاعل كما رأينا في جدول Mauchly's test. و بسبب كون $\epsilon > 0.75$ فإن معنوية الاختبار المستخدمة هي التي تقابل السطر Huynh Feldt. يُفسر وجود التفاعل بين المتغيرين على أن القيم التي أعطيت اعتماداً على مستويات طول الفيلم تختلف باختلاف مستويات متغير نوع الفيلم. يظهر وجود التفاعل بين المتغيرين بوضوح في الجدول (12) و بشكل أكثر وضوحاً في الشكل (11).

إذا نظرنا بامعان إلى المخطط في الشكل (11) نجد مايلي. أولاً ، لننظر إلى المستوى Duration. من الملاحظ أن قيم هذا المستوى لا تتغير بشكل جوهري مع تغيرات مستويات المتغير Film. بكلمات أخرى فإن درجة استحسان الشباب (ممثلة بمتوسط النسب) للفيلم الطويل لا تتأثر بتغيرات مستويات نوع الفيلم (ترفيهي أو وثائقي

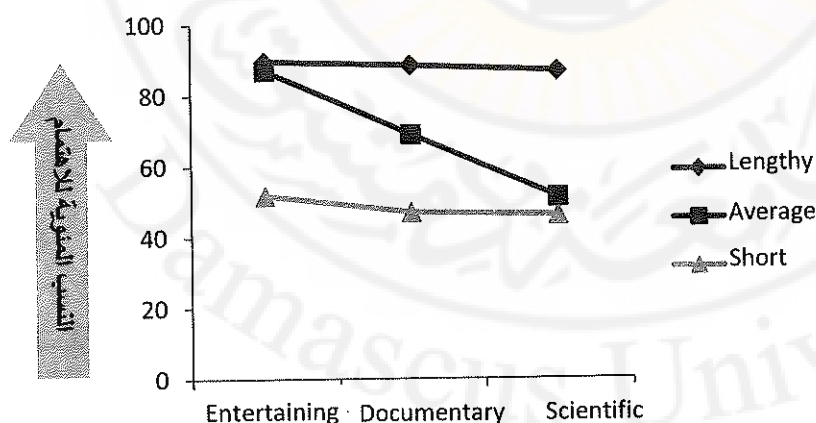
أو علمي). أي أن الشباب يُحبذون الفيلم الطويل بغض النظر عن نوعية هذا الفيلم.
(سؤال: هل يدل هذا على وجود تفاعل بين هذا المستوى و المتغير Film؟)

Look * Charisma

Measure: MEASURE_1

Film	Duration	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
1	1	89.600	2.099	84.852	94.348
	2	87.100	2.151	82.231	91.963
	3	51.800	1.093	49.327	54.273
2	1	88.400	2.634	82.442	94.358
	2	68.900	1.882	54.642	73.158
	3	47.000	1.183	44.323	49.677
3	1	86.700	1.719	82.810	90.690
	2	51.200	1.724	47.299	55.101
	3	46.100	.971	43.903	48.297

الجدول (12): احصائيات حول التفاعل بين المتغيرين Film و Duration.



الشكل (11): مخطط التأثير المتبادل بين المتغيرين Film و Duration.

لننتقل ثانياً إلى المستوى Short (أي فيلم قصير). نلاحظ أن درجة استحسان الشباب للأفلام القصيرة لا تتغير مع تغيرات مستويات نوعية الفيلم (Film) و أن درجة الاستحسان منخفضة. بكلمات أخرى الشباب لا يُحبذون الأفلام القصيرة مهما كانت نوعية الفيلم.

لننظر أخيراً إلى المستوى الأخير Average (أي الأفلام القصيرة) حيث نلاحظ صورة مختلفة تماماً. نلاحظ أن تغيرات مستويات المتغير Film لها دور بارز في تغيرات قيمة المستوى Average. لقد تغيرت درجة استحسان الشباب للأفلام متوسطة طول العرض بتغير نوعية الفيلم. لاحظ لقد حصلت الأفلام الترفيهية على أعلى درجة استحسان و أن الأفلام العلمية متوسطة طول العرض حصلت على أقل استحسان. بكلمات أبسط هناك تفاعل (تأثير متبادل) واضح بين المستوى Average و مستويات المتغير Film.

Key Notes الأفكار الرئيسية للفصل

1. الفرضية هي مقولة يمكن أن تكون صحيحة أو خاطئة. مستوى الأهمية هو قيمة الخطأ الذي نقبل به عند رفض H_0 و هي صحيحة. أما معنوية الاختبار فتتمثل الخطأ الفعلي الذي نرتكبه برفض H_0 و هي صحيحة.
2. هناك خطأ أن يجب الانتباه إليهما بحذر هما الخطأ من النوع الأول (رفض خاطئ لـ H_0) و الخطأ من النوع الثاني (قبول خاطئ لـ H_0). كما ناقشنا سابقاً أن أهمية أي من الخطئين تعتمد بشكل كبير على طبيعة المسألة المدروسة. في بعض الأحيان يكون الخطأ من النوع الأول هو الأكثر أهمية و لذلك نحاول تجنب الوقوع فيه قدر الإمكان و في أحيان أخرى نحاول تجنب الوقوع في الخطأ من النوع الثاني لأنه أكثر أهمية.
3. نستخدم اختبارات ستودنت لاختبار فرضيات حول متوسط مجتمع واحد أو فرضيات حول الفرق بين متوسطي مجتمعين. يجب أن ننتبه إلى تحقق شروط هذا الاختبار قبل تطبيقه.

4. في حالة مقارنة متوسطات أكثر من مجتمعين فلا يمكن استخدام اختبار ستودنت نظراً إلى أن الخطأ يكبر بسرعة مع زيادة عدد المجتمعات المقارنة. في هذه الحالة يجب استخدام اختبارات ANOVA. هذه الاختبارات هي اختبارات وسيطية تستخدم لمقارنة متوسطات مجتمعين أو أكثر. عدة اختبارات متوفرة:

① اختبارات التباين بمتغير مستقل واحد One-Way ANOVA حيث لدينا متغير تابع يراد دراسة تأثير متغير آخر عليه يدعى المتغير المستقل. وتتم دراسة تأثير المتغير المستقل في التابع من خلال مقارنة المتوسطات. إذا تم رفض H_0 فإننا نحتاج إلى إجراء اختبارات Tukey و Games-Howell لمعرفة المتوسطات المختلفة.

② اختبارات التباين بمتغيرين مستقلين Two-Way ANOVA حيث لدينا متغير تابع يراد دراسة تأثير متغيرين آخرين (مستقلين) على هذا المتغير التابع. هناك ثلاث فرضيات يجري اختبارها: فرضية حول كل من المتغيرين المستقلين و فرضية حول التفاعل (Interaction) بين المتغيرين المستقلين.

③ اختبار تحليل التباين للعينات المرتبطة Repeated Measures ANOVA. يستخدم هذا التحليل عندما تكون العينات غير مستقلة وبالتالي لا يمكن استخدام One-Way ANOVA و لا Two-Way ANOVA. الفرضيات التي يجري اختبارها هنا هي فرضية العدم التي تنص على أن المتوسطات متساوية مقابل الفرضية البديلة و التي تنص على وجود متوسط واحد مختلف على الأقل. عند قبول الفرضية البديلة فيجب تحديد أي المتوسطات مختلفة باستخدام الاختبارات الإحصائية المناسبة.

6.5 تمارينات Exercises

نقدم في هذه الفقرة مجموعة من الأسئلة المتعلقة بهذا الفصل و تساعد الطالب في تقويم مدى استيعابه للموضوعات التي تم طرحها في هذا الفصل. لقد صممت هذه الأسئلة بعناية كبيرة لكي تغطي معظم المبادئ الأساسية. لذلك نحث الطالب بقوة على محاولة الإجابة عن جميع هذه الأسئلة لتحقيق الفائدة المرجوة.

توجيه: اقرأ الأسئلة بشكل جيد قبل البدء في الإجابة عنها

أولاً: نعم لا مع التعليل

1. يستخدم اختبار ستودنت T-Student Test من أجل فقط اختبار فرضيات حول متوسط مجتمع واحد؟ نعم أو لا مع التعليل.
2. في حالة استخدام اختبار ستودنت T-Student Test من أجل عينة واحدة فإنه يجب أولاً أن نختبر تجانس تباين هذه العينة قبل محاولة تفسير نتائج الاختبار حول متوسط هذه العينة؟ نعم أو لا مع التعليل.
3. يمكن استخدام اختبار ستودنت T-Student Test مع البيانات الوصفية الاسمية (Nominal Data)؟ نعم أو لا مع التعليل.
4. يمكن استخدام اختبار ستودنت T-Student Test مع البيانات الوصفية الترتيبية (Ordinal Data)؟ نعم أو لا مع التعليل.
5. اختبار التباين و اختبار ستودنت متكافئين و يمكن استخدام أي منهما في اختبار عينتين أو أكثر؟ نعم أو لا مع التعليل.

ثانياً: أسئلة حول الاختبارات الإحصائية Applicability Conditions for

Statistical Tests

6. ما شروط تطبيق اختبار ستودنت؟ حدد جميع الشروط.
7. ما شروط تطبيق اختبار تحليل التباين؟ حدد جميع الشروط.

8. لنفترض أن لدينا ثلاث عينات A و B و أخيراً C. و لنفرض أننا نريد إجراء اختبار حول تساوي متوسطات مجتمعات هذه العينات، و ذلك باستخدام اختبار ستودنت T-Student Test. و المطلوب:

- حدد عدد الثنائيات المطلوب مقارنتها.
 - أحسب الخطأ المرتكب في القرار الإحصائي حول فرضيات تساوي (أو عدم تساوي) هذه المتوسطات.
9. أعد السؤال 8 من أجل أربع عينات و خمس عينات. ماذا تستنتج؟ هل ستستخدم اختبار ستودنت من أجل أكثر من مجتمعين؟

ثالثاً: أسئلة متعددة الخيارات Multi-choice questions

10. يمكن استخدام اختبار ستودنت و اختبار تحليل التباين من أجل مجتمعين و لكن

- A. اختبار ستودنت أكثر دقة ووثوقية.
- B. اختبار تحليل التباين أكثر دقة ووثوقية.
- C. كلاهما يعطي نفس النتائج.
- D. غالباً ما يعطي هذان الاختباران نتائج مختلفة تماماً حتى و لو كانت شروط تطبيقهم محققة.
- E. في حالة مجتمعين فقط فإنه يفضل استخدام اختبار ستودنت لأنه يحتاج إلى زمن أقل.

ملاحظة: في أسئلة الخيارات المتعددة يمكن أن يكون أكثر من خيار واحد بشكل إجابة صحيحة للسؤال. لذلك حتى تكون إجابتك كاملة يجب أن تحدد كل الإجابات الصحيحة. مثلاً، يمكن أن تقول k و m هي إجابات صحيحة. لانتس كذلك تليل إجابتك.

11. يستخدم اختبار ستودنت للعينات المترابطة Two-Sample Paired Test من أجل:

- A. اختبار الفرق بين متوسطي عينتين مختلفتين.
B. اختبار الفرق بين متوسطي عينة واحدة تم حساب متوسطها قبل تطبيق إجراء ما عليها و بعده.
C. اختبار متوسط عينة واحدة.
D. اختبار الفرق بين متوسطات أكثر من عینتين.
E. و لا واحدة من الإجابات أعلاه صحيحة.

12. يستخدم اختبار ستودنت للعينات المستقلة Two-Sample Independent Test من أجل:

- A. اختبار الفرق بين متوسطي عينتين مختلفتين.
B. اختبار الفرق بين متوسطي عينة واحدة تم حساب متوسطها قبل تطبيق إجراء ما عليها و بعده.
C. اختبار متوسط عينة واحدة.
D. اختبار الفرق بين متوسطات أكثر من عینتين.
E. و لا واحدة من الإجابات أعلاه صحيحة.

13. عندما نريد اختبار متوسطات أكثر من مجتمعين فإننا

- A. لانستخدم اختبار ستودنت لأنه مكلف من الناحية الزمنية لأننا نحتاج إلى إجراء الاختبار من أجل كل الثنائيات المختلفة للعينات.
B. لانستخدم اختبار ستودنت لأن الخطأ في القرار الإحصائي سوف يكون كبيراً و يزداد بزيادة عدد العينات.
C. في الواقع يمكن استخدام اختبار ستودنت للعينات المرتبطة فقط.
D. لايمكن استخدام اختبار ستودنت للعينات المرتبطة، و بالتالي يبقى اختبار استخدام ستودنت للعينات المستقلة هو الخيار الأفضل.
E. و لا واحدة من الخيارات صحيحة.

رابعاً: صياغة الفرضيات Hypothesis Formulation

1. لنفرض أنه لدينا ستة أنواع مختلفة من التمارين الرياضية A, B, C, D, E, F و لنفرض أن مصممي هذه التمارين قد طلبوا تقويم هذه التمارين من حيث قدرتها على تخفيض الوزن و سرعة هذا التخفيض. كيف نستطيع تقويم هذه التمارين المختلفة من حيث القدرة على تخفيض الوزن و سرعة هذا التخفيض؟ أي ما الفرضيات الإحصائية التي يجب اختبارها لتقويم فعالية هذه التمارين؟
2. لنفرض أن إحدى شركات تصنيع الأدوية تريد اختبار فعالية مضاد حيوي معين. ما الفرضيات الإحصائية التي يجب اختبارها؟
3. يدعي بعض الباحثين أن متوسط كمية السكر في الدم لدى الأشخاص الأصحاء (أي الذين لا يعانون من مرض السكر) يبلغ قيمة ثابتة و تساوي 77 mg/dl (أي 77 ميليغرام على الديسلتر). ما الفرضيات الإحصائية التي يجب اختبارها للتحقق من صحة الادعاء؟

خامساً: كيفية معالجة المسائل و تحديد نوع الاختبار الإحصائي

1. أراد أحد الباحثين التأكد من أن إنتاجية أشجار الزيتون لا تتأثر بالاسمدة الكيميائية (أي إعطاء أشجار الزيتون كمية من الاسمدة الكيميائية أو عدم إعطائها لا يؤثر في إنتاجية شجرة الزيتون). كيف نتحقق من هذا الادعاء؟
 - a. ما البيانات المطلوبة للتحقق من صحة هذا الادعاء؟
 - b. أكتب الفرضيات الإحصائية المطلوب التحقق منها.
 - c. ما الاختبار الإحصائي المناسب؟

2. أجريت تجربة لتقويم فعالية معالجة الدودة الشريطية في معدة الخراف. لأجل ذلك تم اختيار 24 خروفاً مصاباً بهذه النوع من الديدان. (هذه الخراف من نفس العمر تقريباً و لها نفس الحالة الصحية.) تم تقسيم هذه الخراف إلى مجموعتين حيث عولجت المجموعة الأولى بأدوية مضادة للديدان و تركت المجموعة الثانية بدون معالجة. بعد ستة أشهر تم ذبح هذه الخراف و تم تعداد

الديدان المتبقية في معدتها. الجدول التالي يعطي عدد الديدان الشريطية المتبقية في معدة كل خروف.

عدد الديدان الشريطية داخل معدة كل خروف												
خراف	18	43	28	50	16	32	13	35	38	33	6	7
معالجة												
خراف غير	40	54	26	63	21	37	39	23	48	58	28	39
معالجة												

- a. ما الفرضيات الإحصائية التي يجب اختبارها؟
 - b. أي اختبار من اختبارات الـ SPSS يجب استخدامه؟
 - c. اكتب تقريراً علمياً يلخص النتائج.
3. تدعي إحدى شركات الأدوية بأنها أنتجت مسكن ألم قوي (Pain Killer) يستطيع إزالة معظم الآلام الشديدة (Sever Pain). و أكثر من ذلك فإن الشركة تدعي أن فعالية الدواء هي أكبر من فعالية دوائين آخرين هما Aspirin و Ibuprophen. كيف تتحقق من هذا الادعاء؟
- a. ما البيانات المطلوبة للتحقق من صحة هذا الادعاء؟
 - b. اكتب الفرضيات الإحصائية المطلوب التحقق منها.
 - c. ما الاختبار الإحصائي المناسب؟
4. يعتقد الخبراء في شركة أدوية أن فعالية الدواء drug potency تقل مع الزمن. لأجل ذلك قام الخبراء بأخذ عشر علب دوائية مباشرة من خط الإنتاج و عشر علب من الدواء نفسه و المنتج قبل سنة واحدة ثم فحصوا فعالية هذه الأدوية. يبين الجدول التالي نتيجة فحص الفعالية.

دواء حديث الصنع		دواء مصنوع منذ سنة	
10.2	10.6	9.8	9.7
10.5	10.7	9.6	9.5
10.3	10.2	10.1	9.6
10.8	10.0	10.2	9.8
9.8	10.1	10.1	9.9

هل تؤيد البيانات في الجدول اعتقاد الباحثين؟

5. أجريت تجربة لتقييم تأثير أربعة أنواع من المبيدات الحشرية على انتاجية ثلاثة أنواع من أشجار الحمضيات. يبين الجدول التالي الانتاجية للشجرة الواحدة بحسب نوع الشجرة و نوع المبيد الحشري.

نوع الشجرة (Variety)	مبيد حشري (Pesticide)			
	1	2	3	4
1	49	50	43	53
	39	55	38	48
2	55	67	53	85
	41	58	42	73
3	66	85	69	85
	68	92	62	99

المطلوب:

- اكتب الفرضيات المناسبة.
- اختبر هذه الفرضيات باستخدام الاختبارات الاحصائية المناسبة.
- اكتب تقريراً علمياً يلخص النتائج.

6. في محاولة لربط البيئة الباردة مع ارتفاع التوتر الشرياني (ارتفاع ضغط الدم أو ارتفاع دقات القلب)، تم إجراء تجربة لمعرفة تأثير البيئة الباردة على ارتفاع الضغط الشرياني عند الفئران. تم وضع عينتين من 6 فئران في بيئتين بدرجات حرارة مختلفة. العينة الأولى تم وضعها في بيئة درجة حرارتها 26 درجة مئوية

و العينة الثانية في بيئة درجة حرارتها 5 درجات مئوية. بعد فترة من الزمن تم قياس ضغط الدم لدى هذه الفئران. الجدول التالي يبين قياسات ضغط الدم للفئران في العينتين.

البيئة الأولى: 5 درجات مئوية		البيئة الثانية: 26 درجة مئوية	
الفأر	ضغط الدم	الفأر	ضغط الدم
1	384	7	152
2	369	8	157
3	354	9	179
4	375	10	182
5	366	11	176
6	423	12	149

المطلوب:

- ما الفرضيات الإحصائية التي تريد اختبارها؟
- ما الاختبار الإحصائي المناسب لاختبار الفرضيات (a)؟
- هل البيانات في الجدول أعلاه تقدم دليلاً كافياً بأن الفئران في البيئة الباردة (5 درجات مئوية)، لديها ضغط دم أعلى من الفئران في البيئة ذات درجة الحرارة الأعلى.
- ما القضايا التي يجب أن تؤخذ بعين الاعتبار حتى تكون النتائج أعلاه (c) دقيقة؟

7. يريد أحد الأطباء النفسيين دراسة تأثير المعالجة بالأدوية المضادة للكآبة Antidepressant و المعالجة السلوكية Cognitive Behavior Therapy (CBT) على التفكير بالانتحار. لأجل هذه الغاية تم أخذ أربعة أشخاص ممن

يعرفون بميلهم الكبير للانتحار لإجراء التجربة عليهم. في الشهر الأول للتجربة تم إعطاء كل شخص حبوياً وهمية (لا تحوي أي مادة فعالة Placebo) و لم يخضعوا لأيّة معالجة أخرى. في الشهر الثاني تم إعطاؤهم حبوياً وهمية و لكن خضعوا أيضاً لمعالجة سلوكية. في الشهر الثالث تم إعطاء الأشخاص الأربعة أدوية مضادة للكآبة و لكن بدون أي معالجة سلوكية. و أخيراً في الشهر الرابع تم إعطاؤهم أدوية مضادة للكآبة مع معالجة سلوكية. في الأسبوع الأخير من كل شهر تم الطلب من كل مشارك تسجيل عدد المرات التي فكر فيها بالانتحار. الجدول التالي يلخص البيانات المسجلة.

الدواء (Drug)	وهمي (Placebo)		مضاد إكتئاب (Antidepressant)	
المعالجة (Therapy)	None	CBT	None	CBT
James	70	60	81	52
Andy	66	52	70	40
Mike	56	41	60	31
Erin	68	59	77	49
Mean	65	53	72	43

المطلوب: اجري التحليل الاحصائي المناسب لمعرفة فيما إذا كان عدد مرات التفكير بالانتحار قد تأثر بشكل معنوي بنوع الدواء و المعالجة السلوكية التي حصل عليها كل مشارك و هل هناك تفاعل بين الدواء و المعالجة.

← ما هي المتغيرات المستقلة و كم عدد مستوياتها؟

← ما هو المتغير التابع؟

← هل شرط تساوي الارتباطات Sphericity محقق؟

← اكتب تقريراً يلخص النتائج التي توصلت إليها من خلال التحليل الاحصائي.

8. توجد شكوى دائمة من الطلبة بأن هناك اختلافاً بين المحاضرين عند تصحيح الأوراق الامتحانية. يدعي الطلاب أن بعض المحاضرين يعطي الطالب علامة أكثر من العلامة المستحقة و البعض الآخر يعطي أقل من العلامة المستحق. نادراً ما يدعي الطلاب بأن المحاضرين يعطون العلامة المستحقة. للتحقق من صحة إدعاءات الطلاب تم اختيار أربعة محاضرين عشوائياً و تم اعطاء نفس الورقة الامتحانية إلى المحاضرين الأربعة لتصحيحها. تم إعادة التجربة لثمانية مرات في كل مرة تعطى ورقة معينة للمحاضرين الأربعة ليضعوا علاماتهم المستحقة لها. يلخص الجدول التالي النتائج.

المحاضر الرابع	المحاضر الثالث	المحاضر الثاني	المحاضر الأول	الورقة الامتحانية:
64	63	58	62	1
65	68	60	63	2
65	72	61	65	3
61	58	64	68	4
59	54	65	69	5
50	65	67	71	6
50	67	66	78	7
45	75	73	75	8

← اجري الاختبار الاحصائي المناسب للبيانات لمعرفة هل هناك فرق بين المحاضرين في تصحيحهم للورقة.

← هل تشير النتائج إلى وجود اختلاف بين المحاضرين في تصحيحهم للأوراق الامتحانية؟

9. يعتقد أحد الباحثين في مجال مراقبة التلوث البيئي أن القاطنين على إحدى ضفتي النهر يقوم بتصريف مياه الصرف الصحي نصف المعالجة إلى النهر. نتيجة لذلك فإن مستويات انحلال الأوكسجين في مياه النهر تتغير. للتحقق من هذه الفرضية، قام هذا الباحث باختيار 15 عينة من مياه النهر قبل المدينة و 15 عينة من مياه النهر بعد المدينة. نسبة الأوكسجين المنحل في هذه المياه معطاة في الجدول التالي.

كمية الأوكسجين المنحل في عينات المياه المأخوذة بعد المدينة			كمية الأوكسجين المنحل في عينات المياه المأخوذة قبل المدينة		
5.5	4.8	4.2	4.7	4.8	5.2
4.7	4.9	4.4	5.1	5.0	4.8
4.9	4.6	4.7	5.0	4.7	5.1
4.8	5.1	4.9	4.9	4.7	5.0
4.7	4.3	4.6	4.9	5.0	4.9

المطلوب:

- ما الفرضيات الإحصائية التي تريد اختبارها؟
- ما الاختبار الإحصائي المناسب لاختبار الفرضيات (a)؟
- هل البيانات في الجدول أعلاه تقدم دليلاً كافياً على أن تصريف المياه نصف المعالجة في مياه النهر تسبب ضعف انحلال الأوكسجين.
- ما القضايا التي يجب أن تؤخذ بعين الاعتبار حتى تكون النتائج أعلاه (c) دقيقة؟

10. هناك ثلاث طرق مقترحة لعلاج مرض الخرف الشيخي (Alzheimer's Disease): طريقة A، طريقة B، و أخيراً طريقة C. تقاس درجة هذا المرض على مقياس من 0 إلى 10، حيث صفر تعني أن الشخص قد شفي تماماً من

المرض، و10 تعني أنه لا يوجد تحسن البتة. على هذا الأساس تقاس فعالية طريقة ما بدرجة المرض بعد إخضاع المريض للمعالجة بهذه الطريقة. و أكثر من ذلك يعتقد أيضاً أن نسبة الشفاء تتعلق بجنس المريض (ذكر أو أنثى). كيف تتحقق من هذا الإدعاء؟

a. ما البيانات المطلوبة للتحقق من صحة هذا الإدعاء؟

b. اكتب الفرضيات الإحصائية المطلوب التحقق منها.

c. ما الاختبار الإحصائي المناسب؟

تلميح: عند الإجابة عن هذا الأسئلة فإنه يجب التركيز على مايلي: يجب تحديد البيانات المطلوبة، صياغة الفرضيات المناسبة، و من ثم يجب اختيار الاختبار المناسب مع تعليل كافٍ يبين لماذا استخدمت هذا الاختبار.

سادساً: تفسير النتائج و كتابة تقرير Reporting Results

1. لنفرض أن لديك الجدولين التاليين اللذين يحتويان على نتيجة إجراء تحليل التباين One-Way ANOVA للمتغيرين نمو النبات PlantGrowth و المتغير المستقل مخصب كيميائي Fertilizer.

Test of Homogeneity of Variances

PlantGrowth

Levene Statistic	df1	df2	Sig.
4.494	2	27	.021

ANOVA

PlantGrowth

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	5943.467	2	2971.733	44.616	.000
Within Groups	1798.400	27	66.607		
Total	7741.867	29			

- i. هل من الممكن استخدام الجدول السفلي (جدول اختبار ANOVA) لاختبار فرضية العدم و الفرضية البديلة بمستوى أهمية 5%؟
- ii. إذا كان الجواب نعم، علل لماذا؟ اكتب تقريراً بالنتائج.
- iii. إذا كان لا، علل لماذا؟ بين مالذي يجب فعله لاختبار الفرضيات (فرضية العدم و الفرضية البديلة). ثم اكتب تقريراً بالنتائج.
2. بالعودة إلى (1)

- i. بين لماذا قمنا بإجراء اختباري Welch و Brown-Forsythe و الذي تظهر نتائجه في الجدول التالي:

Robust Tests of Equality of Means				
PlantGrowth	Statistic ^a	df1	df2	Sig.
Welch	66.342	2	16.747	.000
Brown-Forsythe	44.616	2	17.266	.000

a. Asymptotically F distributed.

- ii. يظهر الجدول التالي نتائج اختبار Games-Howell الذي يقارن متوسطات المجموعات الثلاث المشكلة من قياسات المتغير التابع PlantGrowth و المقابلة للمستويات الثلاثة للمتغير المستقل Fertilizer. حدد أياً من هذه المتوسطات مختلفة؟

Multiple Comparisons						
Dependent Variable: PlantGrowth						
Games-Howell						
(I) Fertilizer	(J) Fertilizer	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	-7.200	4.182	.233	-18.18	3.78
	3	-32.800*	4.023	.000	-43.51	-22.09
2	1	7.200	4.182	.233	-3.78	18.18
	3	-25.600*	2.508	.000	-32.02	-19.18
3	1	32.800*	4.023	.000	22.09	43.51
	2	25.600*	2.508	.000	19.18	32.02

*. The mean difference is significant at the .05 level.

سابعاً: تمرينات متنوعة

أجريت دراسة لتقويم فعالية أحد الأدوية المخفضة للضغط Antihypertensive. اختيرت ثلاث مجموعات بحجم 20 من الفئران المصابة بضغط الدم. تم إعطاء المجموعة الأولى جرعة منخفضة من الدواء (Low Dose)، والمجموعة الثانية جرعة أعلى من الدواء (High Dose)، بينما أعطيت المجموعة الثالثة مخفض ضغط عادياً (Control). يبين الجدول التالي نتائج قياسات ضغط الدم بعد فترة من إعطاء الدواء. الرقم السالب يمثل الزيادة في ضغط الدم عن القيمة المقروءة قبل إعطاء الدواء.

Low Dose	High Dose	Control	Low Dose	High Dose	Control
-54.1	54.2	18.2	-65.4	92.6	-33.1
-59.8	89.1	17.2	-17.5	55.3	55.1
58.1	89.6	34.8	22.1	103.2	84.6
-23.7	98.8	3.2	-15.4	45.4	40.3
64.9	107.3	42.9	96.5	70.9	30.5
12.1	65.1	-27.2	-27.7	29.7	18.5
10.5	75.6	42.6	-16.7	40.3	29.3
42.5	52.0	10.0	39.5	73.3	-19.7
48.5	50.2	102.3	-4.2	21.0	37.2
-1.7	80.9	61.0	-41.3	73.2	48.8

1. حدد الفرضيات التي تريد اختبارها.
2. استخدم الـ SPSS لاختبار هذه الفرضيات.

أراد أحد علماء الحياة البرية أن يقيم تأثير البيئة التي يربى فيها الغزلان على الزيادة في حجم هذه الغزلان. لأجل ذلك قام العالم بأخذ ثلاث عينات عشوائية من الغزلان من ثلاث بيئات مختلفة: بيئة برية Wild، بيئة الأدغال Ranches، و بيئة حديقة الحيوان Zoo. ثم قام بوزن هذه الغزلان. يبين الجدول التالي الأوزان بالباوند (1 باوند = 454 غرام).

Environment	Weight (in pounds) of deer							
Wild	114.7	128.9	111.5	116.4	134.5	126.7	120.6	129.6
Ranch	120.4	91.0	119.6	119.6	150.0	169.7	100.9	76.1
Zoo	103.1	90.7	129.5	129.5	182.5	76.8	87.3	77.3

1. حدد الفرضيات التي تريد اختبارها.
2. استخدم الـ SPSS لاختبار الفرضيات.

★★★★★★

الاختبارات غير الوسيطة Nonparametric Statistical Tests

1.6 مقدمة Introduction

الاختبارات غير الوسيطة Nonparametric tests ، أو أحياناً تسمى الاختبارات المستقلة عن التوزيع ، هي اختبارات لا تعتمد على وسطاء المجتمع كالانحراف المعياري و المتوسط — لذلك سميت غير وسيطة. كذلك فإن هذه الاختبارات مستقلة عن توزيع المجتمع فهي لا تشترط أن يكون مجتمع العينة يتبع التوزيع الطبيعي. بالنتيجة ، فإن لها شروط تطبيق أقل من شروط تطبيق الاختبارات الوسيطة.

تستخدم الاختبارات غير الوسيطة عندما لا نعرف هل مجتمع العينة يتبع التوزيع الطبيعي وبالتالي لا يمكن استخدام الاختبارات الوسيطة. على سبيل المثال ، إذا كان توزيع مجتمع العينة ملتوياً Skewed. و لكن ينبغي أن نعلم أنه على الرغم من أن الطرق غير الوسيطة قابلة للتطبيق بغض النظر عن توزيع مجتمع العينة ، إلا أنها أقل كفاءة من الطرق الوسيطة في كشف الفروق بين المجتمعات. على النقيض تستخدم الطرق الوسيطة معلومات أكثر عن العينة كالمتوسط ، الانحراف المعياري ، شكل التوزيع ، إلخ و بالتالي سوف تكون أكثر حساسية لاكتشاف الفروق بين العينات من الاختبارات غير الوسيطة التي تأخذ فقط ، كما سوف نرى ، رتب البيانات.

سوف ندرس في هذا الفصل مجموعة من الاختبارات غير الوسيطة التي تشمل على مايلي.

- اختبارات الاستقلالية Variable-Independence Tests
- اختبار جودة التطابق Goodness-of-fit Test
- اختبار اتفاق المراقبين Observer Agreement
- اختبارات مقارنة مجتمعين Two-Sample tests
- اختبارات أكثر من مجتمعين Population Comparison Tests

2.6 اختبارات كاي مربع Chi-square tests

اختبار كاي مربع هو أحد الاختبارات غير الوسيطة التي تستخدم مع البيانات الكيفية Qualitative Data. يستخدم هذا الاختبار في العديد من المسائل التطبيقية بدءاً من فحص استقلالية المتغيرات إلى مطابقة نتائج مشاهدة من خلال تجربة معينة مع قانون علمي يعتقد أن الظاهرة تخضع له.

في اختبارات كاي مربع ، عادةً ما تكون المشاهدات مرتبة بشكل جدولي حيث تمثل الأسطر مستويات المتغير الأول و الأعمدة مستويات المتغير الثاني. لذلك اختبار كاي مربع هو أحد اختبارات الجداول الثنائية المتقاطعة Crosstabs. الشكل التالي يمثل كيف ترتب البيانات في اختبار كاي مربع.

		المتغير الثاني			
المتغير الأول		L_1	L_2	...	L_n
	D1	F_{11}	F_{12}		F_{1n}
	D2	F_{21}	F_{22}	...	F_{2n}
	:				
	Dm	F_{m1}	F_{m2}	...	F_{mn}

كما هو واضح بالشكل لدينا متغير أول و له m مستوى $D_1, D_2, \dots, D_m$ و متغير ثانٍ و له n مستوى $L_1, L_2, \dots, L_n$. تقاطع مستوى من العمود و مستوى من السطر يشكل خلية. كل خلية تحوي مجموعة من التكرارات F_{ij} — أي عدد المشاهدات في السطر i و العمود j . على سبيل المثال، F_{13} تمثل عدد المشاهدات (التكرارات) الخاصة بالمستوى الأول للمتغير الأول و المستوى الثالث للمتغير الثاني.

هناك نوعان لاختبار كاي مربع:

- ① اختبار استقلالية متغيرين Two Variables Independent Test. يستخدم لدراسة العلاقة بين متغيرين في سياق ظاهرة ما.
- ② اختبار جودة التطابق Chi-square goodness-of-fit Test. يستخدم بشكل خاص لمعرفة فيما إذا كانت مشاهدات ناتجة من تجربة معينة تتطابق مع توزيع احتمالي نظري يضبط هذه المشاهدات.

قبل أن نعرض شرحاً مفصلاً لهذين الاختبارين ، نتوقف قليلاً عند شروط تطبيق اختبار كاي مربع Chi-square test بشكل عام. هناك العديد من الشروط لعل أهمها الشروط الأربعة التالية:

1. أن تكون البيانات (المشاهدات) مسحوبة بشكل عشوائي.
2. أن يكون حجم العينة كبيراً بشكل كافٍ. في الحقيقة لا يوجد مقياس محدد و متفق عليه يبين ماذا نعني بـ "كبير بشكل كافٍ". البعض يقول يجب أن لا يقل عدد المشاهدات عن 50 ، و البعض الآخر يقول إنه يجب أن لا يقل عدد المشاهدات عن 20. في جميع الأحوال هناك حقيقة ثابتة عن هذا الاختبار مفادها أنه كلما كان عدد المشاهدات أقل كبر مقدار الخطأ من النوع الثاني.
3. يجب أن يكون عدد المشاهدات في كل خلية كبيراً كفاية. هنا أيضاً نتعرض إلى معنى "كبير كفاية". البعض يرى يجب أن تحتوي كل خلية على 5 مشاهدات أو أكثر ، البعض الآخر يقول إن عدد المشاهدات في الخلية الواحدة يجب أن

يكون أكثر من 5 أو حتى أكثر من 10. القاعدة الشائعة هي أنه في جدول من عمودين و سطرين يجب أن يكون عدد المشاهدات 5 أو أكثر. أما في جداول أكبر من ذلك فيجب أن يكون على الأقل 80% من الخلايا تحوي 5 مشاهدات أو أكثر لكل خلية. كذلك يجب أن لا تكون أي خلية واحدة تحوي صفراً مشاهدة.

4. يجب أن تكون المشاهدات مستقلة و لا تتكرر نفس المشاهدة في أكثر من خلية واحدة.

ملاحظة عامة

سوف نستخدم معالج البيانات الإحصائية SPSS و ذلك لأجل كل الاختبارات الإحصائية التي سوف ندرسها في هذا الفصل. في هذا الفصل نكتفي فقط بعرض نتائج التحليل باستخدام الـ SPSS و تفسير هذه النتائج. يمكن للقارئ الذهاب إلى الملحق C لمعرفة تفاصيل كافية عن كيفية إدخال البيانات في شاشة الـ SPSS ، كيفية تفعيل الاختبارات ، و قضايا أخرى. ننصح كثيراً أن يقوم القارئ أولاً بفهم تفسير نتائج الاختبارات ، و من ثم الذهاب إلى الملحق C لمعرفة كيف يتم استخدام الـ SPSS لإجراء الاختبار.

1.2.6 اختبار الاستقلالية Chi-square Independence Test

يستخدم اختبار الاستقلالية لتقييم العلاقة بين متغيرين يؤثران في ظاهرة ما. حيث يمكن لهذين المتغيرين أن يكونا مستقلين عن بعضهما البعض أو مرتبطين. إذا كان المتغيران مستقلين ، فإن طريقة تأثير أحدهما في الظاهرة المدروسة لا تؤثر في طريقة تأثير المتغير الآخر في الظاهرة نفسها. وجود ارتباط بين المتغيرين يعني أن طريقة تأثير أحد المتغيرين تؤثر في طريقة تأثير المتغير الآخر على الظاهرة.

يجدر بنا أن نشير إلى أن اختبار الاستقلالية يجيب فقط عن السؤال: "المتغيران مستقلان أم لا". لكنه لا يستطيع تحديد فيما إذا كانت العلاقة بينهما قوية أو ذات فائدة

تطبيقية. تفسير الارتباط و قوته يقع على عاتق الاحصائي (أو محلل البيانات) باستخدام تقنية تحليل النسب كما سنرى. عادة ما يتم حساب بعض معاملات الارتباط مثل Cramer's V و Phi على الرغم من أنني لا أحب ذلك.

الشكل العام للفرضيات التي يختبرها هذا الاختبار:
فرضية العدم H_0 : المتغيران مستقلان في تأثيرهما على الظاهرة.
فرضية البحث H_1 : المتغيران مرتبطان "طريقة تأثير أحدهما في الظاهرة تعتمد على طريقة تأثير الآخر".

هناك العديد من المسائل الواقعية التي ينطبق عليها اختبار استقلالية المتغيرات. سوف نذكر عدداً من هذه المسائل التي تساعدنا على إعطاء تصور عن النموذج العام لهذه المسائل.

1. دراسة العلاقة بين كميات التلوث في البيئة و تلوث البحيرات.
2. دراسة العلاقة بين زواج الأقارب و مرض الصرع epilepsy.
3. دراسة العلاقة بين الجنس و العنف في المدارس الثانوية.
4. دراسة العلاقة بين المستوى التعليمي و الطلاق.

النموذج العام

متغيران كيفيان (أي قيمهما تكرارية) يراد دراسة العلاقة بينهما حول ظاهرة معينة.
مثال توضيحي: من الواضح أن الأمثلة المطروحة أعلاه تنضوي تحت هذا النموذج. على سبيل المثال، دراسة ظاهرة الشغب و ارتباطها بالجنس. حيث يمكن مراقبة مجموعة من، مثلاً، طلاب المدارس الثانوية و تصنيف كل طالب على أنه مثير للشغب أم لا، و كذلك معرفة جنس الطالب المشاهد. هل نستطيع تحديد لماذا باقي الأمثلة تنضوي تحت هذا النموذج؟

مثال (1)

لنكن لدينا البيانات في الجدول التالي التي تربط بين الشغب Trouble في المدارس الاعدادية و جنس الطلاب Gender. متغير الشغب له مستويان مشاغب " Trouble Maker" و غير مشاغب "Peaceful" و متغير الجنس له مستويان ذكر "Male" و أنثى "Female". هل تشير هذه البيانات إلى وجود علاقة بين الشغب و جنس الطالب؟

Gender	Trouble		Total
	Trouble Maker	Peaceful	
Male	40	160	200
Female	80	320	400
Total	120	480	600

الفرضيات الإحصائية

فرضية العدم H_0 : متغير الشغب مستقل على الجنس. أي كلا الجنسين يمكن أن يكون مشاغب بالدرجة نفسها.

الفرضية البديلة H_1 : متغير الشغب مرتبط بمتغير الجنس. أي أن أحد الجنسين يمكن أن يكون مشاغباً من الثاني.

نتائج اختبار الاستقلالية

يبين الجدول (1) مخرجات الـ SPSS لاختبار الاستقلالية. تحوي الخلايا المشكلة بتقاطع مستويات متغير السطر Gender و مستويات متغير العمود على التكرارات الواقعة ضمن الخلايا والنسب المئوية المقابل للسطر و العمود. على سبيل المثال ، الخلية المشكلة بالمستوى Peaceful و Male تحوي 40 تكراراً و نسبة السطر (رمزة %within Gender) المقابلة 20% و نسبة العمود (رمزة %within Trouble) المقابلة 33.3%. (نسب الأسطر تحسب بقسمة العدد في الخلية على

المجموع السطري. نسبة العمود تحسب بقسمة العدد في الخلية على المجموع العمودي.)

لاحظ أن 20% من المشاغبين هم ذكور و 20% من المشاغبين هم إناث. أي أن كون الشخص مشاغباً لم يتأثر بكونه أنثى أو ذكر. علاوةً على ذلك ، لاحظ أن 33.5% من الذكور مشاغبون و النسبة نفسها غير مشاغبين. مما يشير إلى إمكانية أن يكون المتغيران Gender و Trouble مستقلين.

Gender * Trouble Crosstabulation

			Trouble		Total
			Trouble Maker	Peaceful	
Gender	Male	Count	40	160	200
		% within Gender	20.0%	80.0%	100.0%
		% within Trouble	33.3%	33.3%	33.3%
	Female	Count	80	320	400
		% within Gender	20.0%	80.0%	100.0%
		% within Trouble	66.7%	66.7%	66.7%
Total	Count	120	480	600	
	% within Gender	20.0%	80.0%	100.0%	
	% within Trouble	100.0%	100.0%	100.0%	

الجدول (1): وصف البيانات بدلالة التكرارات و النسب.

يبين الجدول (2) مخرجات الـ SPSS لمعنوية اختبار Chi-Square و بعض الاحصائيات الأخرى. ما يهمنا هنا هو السطر Pearson Chi-Square الذي يظهر مظللاً في الجدول. نلاحظ أن الاختبار غير معنوي لأن معنويته تساوي "1" و هي أكبر من أي مستوى أهمية يمكن أن نختاره. بالنتيجة نقبل فرضية العدم التي تنص على أن المتغيرين مستقلان.

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	.000^a	1	1.000
Continuity Correction ^b	.000	1	1.000
Likelihood Ratio	.000	1	1.000
Fisher's Exact Test			
Linear-by-Linear Association	.000	1	1.000
N of Valid Cases	600		

الجدول (2): معنوية Chi-Square

مثال (2)

ادرس البيانات التالية التي تمثل مسحاً لأحد المنتجات الرائدة في السوق. المتغيرات التي نريد دراسة العلاقة بينها هي متغير درجة استخدام هذا المنتج Use و له مستويات ثلاثة: نادر Rare و بعض الأحيان Occasional و متكرر Frequent. و المتغير الآخر كيف يقيم الزبائن هذا المنتج Rating و له ثلاثة مستويات: مقبول Fair و جيد Good و ممتاز Excellent.

	Rating			Total
	Use	Fair	Good	
Rare		64	123	137
Occasional		131	256	129
Frequent		209	171	45
Total		404	550	311

الفرضيات الإحصائية

فرضية العدم H_0 : متغير درجة الاستخدام Use مستقل عن تقييم المنتج Rating.
 الفرضية البديلة H_1 : درجة الاستخدام Use مرتبط بدرجة تقييم المنتج Rating.

نتائج اختبار الاستقلالية

يبين الجدول (3) وصفاً للبيانات بدلالة التكرارات و النسب المئوية للأسطر. تشير النسب المئوية الواردة في الجدول (3) إلى وجود علاقة بين المتغيرين Rating و Use. لاحظ النسب في العمود Excellent. نسبة الذين استخدموا المنتج بشكل قليل Rare الذين يعتبرون المنتج ممتازاً Excellent هي 42.3%. بينما نسب الذين استخدموا المنتج بالمناسبات أو بشكل متكرر و قيموا المنتج على أنه ممتاز هي على الترتيب 25% و 10.6%. و بالتالي كلما ارتفع الاستخدام كلما انخفضت نسبة الذين يعتبرون المنتج ممتازاً. العكس صحيح لأجل نسب الذين يعتبرون المنتج مقبولاً Fair. تشير النسب في العمود الثاني Good إلى عدم وجود علاقة مع الاستخدام Use. لا يوجد اتجاه لتغير النسب.

Use * Rating Crosstabulation

			Rating			Total
			Fair	Good	Excellent	
Use	Rare	Count	64	123	137	324
		% within Use	19.8%	38.0%	42.3%	100.0%
	Occasional	Count	131	256	129	516
		% within Use	25.4%	49.6%	25.0%	100.0%
	Frequent	Count	209	171	45	425
		% within Use	49.2%	40.2%	10.6%	100.0%
Total	Count		404	550	311	1265
	% within Use		31.9%	43.5%	24.6%	100.0%

الجدول (3): وصف البيانات بدلالة التكرارات و النسب.

تشير القيم في الجدول (4) إلى أن اختبار الاستقلالية معنوي لأن معنوية الاختبار المقابلة للسطر Pearson Chi-Square هي "0.000" و هي أصغر من مستوى

الأهمية المفروض "0.05". إذا نرفض فرضية العدم و نقبل الفرضية البديلة التي تنص على وجود علاقة بين المتغيرين Rating و Use.

Chi-Square Tests

	Value	Df	Asymp. Sig. (2-sided)
Pearson Chi-Square	143.587 ^a	4	.000
Likelihood Ratio	143.587	4	.000
Linear-by-Linear Association	126.964	1	.000
N of Valid Cases	1265		

الجدول (4): معنوية اختبار الاستقلالية Chi-Square

إختبار Chi-Square يبين فقط وجود علاقة ارتباط بين المتغيرات. تحليل النسب يظهر قوة العلاقة بين المتغيرين. و بالتالي اختبار Chi-Square و تحليل النسب متكاملان.

2.2.6 اختبار جودة التطابق Chi-square Goodness-of-fit Test

يدرس هذا الاختبار تطابق نتائج مشاهدة من تجربة معينة أجريت على ظاهرة ما مع نتائج متوقعة من نموذج نظري (مثلاً نموذج احتمالي). على سبيل المثال ، من خلال تجربة إلقاء حجر نرد عدة مرات نحصل على عدد من التكرارات لكل وجه من الوجوه الستة. هل تطابق هذه النتائج مع التكرار المتوقع نظرياً: احتمال ظهور أي وجه من الوجوه الستة متساوي. بكلمات أخرى ، هل عدد مرات ظهور كل وجه متساوية تقريباً؟

هناك الكثير من الأمثلة التي ينطبق عليها هذا الاختبار المهم:

1. من الناحية العلمية فإن احتمال ولادة الذكر مساوية لاحتمال ولادة الأنثى. بفرض أن السجلات في مشفى لآخر 20 و لادة بينت أنه من بين المواليد الـ 20 هناك 8 ذكور و 12 أنثى. هل تتفق هذه النتائج مع التوقع العلمي؟

2. لدى إجراء التصلب بين نباتات بازلاء ملساء البذور مع نباتات بازلاء مجعدة البذور حصلنا نتيجة لهذا التصلب على 680 نبتة. من بين هذه النباتات 420 نبتة مجعدة البذور و 260 نبتة ملساء البذور. هل تتفق هذه النتائج مع قانون Mendel's Law الذي ينص على أن نسبة النباتات ببذور مجعدة إلى نسبة النباتات ببذور ملساء هي 3:1.

3. إن تكاثر حشرة عثة الغجر Gypsy Moth تكون في الفصول (الصيف ، الخريف ، الشتاء ، و الربيع) بالشكل 5:2:3:4 (أي كل حشرة تنتج أربع حشرات في الصيف ، و ثلاث حشرات في الخريف ، و حشرتين في الشتاء ، و خمس حشرات في الربيع). لدى مراقبة تكاثر 100 حشرة ، تبين أن 360 حشرة في الصيف، و 310 حشرة في الخريف، و 180 حشرة في الشتاء ، و 600 حشرة في الربيع. هل تتفق هذه المشاهدات مع القانون؟

الشكل العام للفرضيات التي يتم اختبارها وفق هذا الإجراء
فرضية العدم H_0 : النتائج المشاهدة تطابق النتائج المتوقعة حسب قانون نظري معين.
الفرضية البديلة (فرضية البحث) H_1 : النتائج المشاهدة لا تطابق النتائج المتوقعة حسب قانون نظري معين.

مثال (1)

لدى التصلب بين فصيلتين من نبات البازلاء واحدة ببذور خضراء و الأخرى ببذور صفراء حصلنا على 880 نبتة بازلاء نتيجة لهذا التصلب. هناك 639 نبتة ببذور خضراء ، و 241 نبتة ببذور صفراء. المتوقع وفق قانون ماندل هو أن نسبة النباتات ببذور خضراء إلى النباتات ببذور صفراء تعطى بـ 3:1 (أي ثلاثة أرباع النباتات ببذور خضراء و الربع الباقي نباتات ببذور صفراء). هل تتطابق هذه النتائج المشاهدة مع قانون ماندل؟

تحديد الاختبار المناسب

من الواضح أن المطلوب من خلال التجربة هو التأكد من قانون ماندل الذي يعطي احتمال الحصول على بذور خضراء و بذور صفراء من خلال التصلب بين النوعين من نباتات البازلاء. و بالتالي المسألة هي قياس مدى التوافق بين قانون علمي و توزيع مشاهد من خلال التجربة. أي اختبار جودة تطابق.

الفرضيات التي يراد اختبارها في هذه المثال هي:

فرضية العدم H_0 : النتائج المشاهدة تتفق مع النتائج المتوقعة و فق قانون ماندل.

الفرضية البديلة H_1 : النتائج المشاهدة لا تتفق مع النتائج المتوقعة وفق قانون ماندل.

نتائج اختبار جودة التطابق

يظهر الجدول (5) عدد التكرارات المشاهدة "Observed N" و التكرارات المتوقعة "Expected N" للنباتات ذات البذور الخضراء Green و للنباتات ذات البذور الصفراء Yellow. كما يبدو من الجدول أن عدد النباتات المشاهدة "Observed N" ببذور خضراء هو 639 بينما العدد المتوقع "Expected N" من خلال قانون ماندل هو 660. كذلك ، عدد النباتات المشاهدة ببذور صفراء هو 241 بينما العدد المتوقع خلال قانون ماندل هو 220. فهل الفروقات بين هذه الأعداد معنوية إحصائياً أو تعود للصدفة فقط؟ الجدول (6) يعطي الجواب لهذه السؤال.

Observed			
	Observed N	Expected N	Residual
نباتات بذور خضراء	639	660.0	-21.0
نباتات بذور صفراء	241	220.0	21.0
Total	880		

الجدول (5): عدد النباتات ببذور خضراء Green و بذور صفراء Yellow.

يبين الجدول (6) الإحصائيات اللازمة لاختبار الفرضيات أعلاه حول التطابق ما بين التوزيع المشاهد و المتوقع وفق قانون ماندل للتصالب بين نباتات البازلاء. نلاحظ أن قيمة إحصائية كاي مربع "2.673" و معنوية الاختبار (القيمة داخل الدائرة في الجدول) "0.102". نلاحظ أن اختبار جودة التطابق غير معنوي لأن معنوية الاختبار أكبر من مستوى الأهمية "0.05". نقبل فرضية العدم H_0 و نستنتج أنه لا يوجد دليل إحصائي كافٍ للاعتقاد بأن النتائج المشاهدة تختلف معنوياً عن النتائج المتوقعة وفق قانون ماندل.

Test Statistics	
	Observed
Chi-Square ^a	2.673
df	1
Asymp. Sig.	.102

الجدول (6): إحصائيات اختبار جودة التطابق.

مثال (2)

في تجربة تزاوج بين 30 طائراً أسود مع 30 طائراً أبيض حصلنا نتيجة لهذا التزاوج على 42 طائراً أسود و 60 طائراً أبيض. حسب قوانين الوراثة عند تزاوج طائر أسود مع طائر أبيض فإن احتمال ظهور طائر بأحد اللونين (أسود أو أبيض) متساو. هل تتطابق النتائج المشاهدة مع ما هو متوقع وراثياً؟

من الواضح أن هذه المسألة هي من نماذج مسألة تحديد جودة التطابق ما بين توزيع مشاهد (102 طائراً ناتجة عن التجربة) و توزيع احتمالي. لذلك سوف نستخدم لمعالجتها اختبار Chi-square goodness-of-fit. الفرضيات التي نريد اختبارها:

فرضية العدم H_0 : النتائج المشاهدة تتطابق مع المتوقع وفق القانون الاحتمالي.

الفرضية البديلة H_1 : النتائج المشاهدة لا تطابق النتائج المتوقعة وفق القانون الاحتمالي.

يظهر الجدول (7) ملخصاً للتكرار المشاهد "Observed N"، و التكرار المتوقع "Expected N" لكل من الطيور البيضاء White و السوداء Black. لاحظ أن التكرار المتوقع بالنسبة إلى الطيور البيضاء و السوداء متساوٍ و يساوي 51.0 لأنه لدينا 102 طائراً أبيض و أسود و احتمال ظهور كل لون يساوي $\frac{1}{2}$ ، و بالتالي تكرار كل لون $102 * \frac{1}{2} = 51.0$. بينما التكرارات المشاهدة لا تتطابق ظاهرياً مع التكرار المتوقع. فهل يدل الاختلاف بين التكرار الملاحظ و التكرار المتوقع على عدم تطابق بين المشاهد من خلال التجربة و المتوقع من خلال القانون؟

	Observed N	Expected N	Residual
White	42	51.0	-9.0
Black	60	51.0	9.0
Total	102		

الجدول (7): عدد التكرارات الملاحظة Observed و المتوقعة Expected.

يحتوي الجدول (8) على كل الإحصائيات اللازمة لاختبار الفرضيات أعلاه. نلاحظ من خلال الجدول أن معنوية الاختبار $\text{Asymp. Sig} = 0.075$. و هذه القيمة أكبر من مستوى الأهمية المحدد 0.05. و في هذه الحالة لانستطيع رفض فرضية العدم H_0 ، و بالتالي نستنتج أنه لا يوجد دليل إحصائي كافٍ للاعتقاد بأن التوزيع المشاهد مختلف عن التوزيع المتوقع بالقانون الوراثي.

Test Statistics	
	Type
Chi-Square ^a	3.176
df	1
Asymp. Sig.	.075

الجدول (8): إحصائيات اختبار جودة التطابق.

1. اختبار Chi-Square لجودة التطابق يقارن تكرار مشاهد مع تكرار محدد بواسطة قانون نظري. اختبار Chi-Square للاستقلالية من جهة ثانية يقارن تكرارات في سطر (أو عمود) مع تكرارات سطر آخر. يمكن الاستفادة من القاعدة التالية للتمييز بين الاختبارين: إذا كان المطلوب المقارنة مع تكرارات محددة بقانون نظري فإن الاختبار المناسب هو اختبار جودة التطابق و إلا اختبار الاستقلالية.
2. الهدف الرئيسي لاختبار Chi-Square هو تحديد فيما إذا كانت الفروق الظاهرة في القياسات هي فروق نتيجة المصادفة أم هي فروق حقيقية. لكنه لا يقيس أبداً كبر هذه الفروقات.
3. بشكل خاص ، إذا كان اختبار الاستقلالية معنوياً فهو لا يعني وجود علاقة قوية. كذلك فإن عدم معنوية اختبار جودة التطابق لا يعني أبداً أن القياسات النظرية قريبة جداً من القياسات المحددة بالقانون النظري.
4. يجب النظر بتمعن إلى النسب في اختبار الاستقلالية لرؤية هل تظهر النتائج أي أهمية عملية.

4.5 اختبار توافق المراقبين Observer Agreement

اختبار اتفاق المراقبين (و يسمى أيضاً (Inter-rater Reliability (Kappa) هو أحد اختبارات الجداول المتقاطعة Crosstabs. يقيس هذا الاختبار درجة الاتفاق بين مجموعتين من المراقبين (raters/observers) يقومون بتصنيف ظاهرة ما. يجب أن يكون مقياس تصنيف الظاهرة ثابتاً و متفقاً عليه بين مجموعتي المراقبين قبل البدء فعلياً بالتصنيف. على سبيل المثال ، إذا أردنا أن تصنيف مجموعة من النباتات وفق مقياس التصنيف A, B, C ، فإن كل مجموعة من مجموعتي المراقبين تصنف نبتة معينة على أنها تعود إلى الصنف A أو B أو C فقط. بالنتيجة تظهر التصنيفات على شكل جدول كما هو موضح بالجدول التالي.

	Observer 2				
		T_1	T_2	...	T_n
Observer 1	T_1	F_{11}	F_{12}	...	...
	T_2	...	...	...	...
	..	...	...	...	...
	T_n	...	...	...	...

مقاييس التصنيف $T_1, T_2, \dots, T_n$ تكون موحدة و متفقاً عليها بين مجموعتي المراقبين Observer 1 و Observer 2. القيمة F_{11} تمثل عدد العناصر من الظاهرة التي اتفق المراقبون Observer 1 و المراقبون Observer 2 على أنها تعود للصنف T_1 . القيمة F_{12} تمثل عدد العناصر من الظاهرة التي اختلف فيها المراقبون Observer 1 و المراقبون Observer 2. حيث صنفها المراقبون Observer 1 على أنها تعود للصنف T_1 بينما صنفها المراقبون Observer 2 على أنها تعود للصنف T_2 . لاحظ أن القطر الرئيسي للجدول أعلاه F_{ii} يمثل الاتفاق بين المراقبين ، أما باقي خلايا الجدول فهي تمثل اختلافاً في التصنيف بين المراقبين.

يمكن للمراقبين أن يكونوا خبراء بشريين أو أجهزة قياس أو نظم برمجية. على سبيل المثال، لنفترض أن لدينا مجموعة من سلاسل الـ DNA (DNA Texa) التي تعود للأنواع K_1, K_2, K_2 . للقيام بهذا التصنيف فإن إدارة المخبر تتطلب أن يتم هذا التصنيف بوساطة نظام برمجي و مجموعة من خبراء تصنيف الـ DNA. لاحظ أن المراقبين هنا هما النظام البرمجي و مجموعة الخبراء.

المقياس الأساسي لتقويم درجة الاتفاق بين المراقبين هو إحصائية تسمى Kappa و تسمى أيضاً Cohen's Kappa. تتراوح قيمة Kappa بين -1 و 1 ، حيث القيم الموجبة القريبة من الواحد تدل على اتفاق كبير بين المراقبين بينما يدل الصفر على انعدام التوافق (أو ما يسمى توافقاً بالمصادفة و ليس حقيقياً). أما القيم السالبة فيصعب تفسيرها.

من الأمثلة التي تنضوي تحت هذا الاختبار:

1. مجموعتان من علماء النبات يصنفون مجموعة من النباتات على أنها من الفصيلة A أو الفصيلة B أو الفصيلة C ، و يطلب هل هناك توافق بين مجموعتي العلماء على هذا التصنيف.
 2. جهاز معين لقياس جودة مخصب كيميائي حيث إن هذا الجهاز يصنف مخصباً ما على أنه سيئ الجودة، متوسط الجودة ، أو عالي الجودة. و يراد معرفة هل تصنيف هذا الجهاز صحيح ، و ذلك بقياس درجة الاتفاق بين نتائج الجهاز و نتائج تصنيف نفس المخصبات من قبل خبراء.
- يلاحظ من خلال الأمثلة السابقة أن المسائل التي ينطبق عليها هذا الاختبار تأخذ الشكل العام التالي.

الشكل العام

تصنف مجموعتان من المراقبين ظاهرة ما وفق مقياس معين و متفق عليه و يراد قياس درجة الاتفاق ، إن وجدت ، بين هاتين المجموعتين من المراقبين. على الرغم من بساطته، فإن هذا الشكل يضم الكثير من المسائل التطبيقية في علم الحياة و جميع العلوم الأخرى (هل تستطيع أن تفكر في مسائل أخرى غير الأمثلة السابقة؟).

يعتمد القرار الإحصائي حول وجود الاتفاق بين المراقبين على قيمة الإحصائية Kappa. اعتماداً على قاعدة Landis (Landis and Koch, 1977) فإن قيم Kappa تُفسر وفقاً للسلم التالي:

تفسير قيمة Kappa	قيمة Kappa
درجة الاتفاق بين المراقبين ضعيفة	$Kappa < 0.40$
هناك اتفاق متوسط القوة بين المراقبين	$0.40 \leq Kappa < 0.60$
هناك اتفاق جيد بين المراقبين	$0.60 \leq Kappa < 0.80$
هناك اتفاق قوي بين المراقبين	$Kappa \geq 0.80$

مثال (1)

لنفرض أنه تم فحص 100 مريض اضطراب سلوك بوساطة طبييين. و لنفرض أن أي مريض يُصنّف على أنه مضطرب السلوك بسبب: نفسي Psychological، أو عصبي، Neurological، أو عضوي Organic. يبين الجدول التالي تصنيفات حالات المرضى اعتماداً على المراقب Rater A و المراقب الآخر Rater B.

السؤال هل هناك اتفاق مابين المراقبين A و B؟

RATER A

RATER B		Psychological	Neurological	Organic
	Psychological	75	1	4
	Neurological	5	4	1
	Organic	0	0	10

لماذا تدرج هذه المسألة تحت مجموعة مسائل اتفاق المراقبين؟ من خلال نص المسألة هناك ظاهرة (المرضى) و يراد معرفة منشأ الاضطراب السلوكي لديهم (أي هناك تصنيف لحالتهم المرضية) من خلال مجموعتين من المراقبين (الأطباء). و بالتالي فإن هذه المسألة هي مسألة اتفاق مراقبين.³

يحتوي الجدول (9) قيمة Kappa التي نحتاجها لتحديد درجة الاتفاق بين المراقبين. قيمة Kappa (القيمة ضمن الدائرة) تساوي "0.676". تفسر هذه القيمة، بحسب جدول قيم Kappa، على أنها تمثل اتفاقاً جيداً بين المراقبين. القيمة الأخرى المهمة في

³ إن هذه المسألة ليست اختبار استقلالية لأنه لا يوجد متغيران يؤثران في الظاهرة لكي ندرس استقلالية (أو ارتباط) تأثيرهما في الظاهرة. كما أن هذه المسألة ليست اختبار جودة تطابق لأنه لا يوجد قانون محدد (علمي أو غير ذلك) يبين احتمال أن يكون منشأ المرض لدى مريض معين هو نفسي أو عصبي أو غير ذلك.

الجدول (12) هي قيمة معنوية الاختبار أو "0.000". تفيد هذه القيمة باختبار فيما إذا كانت القيمة المحسوبة لـ Kappa تختلف معنوياً عن الصفر. و حيث إن قيمة معنوية الاختبار تساوي "0.000" و هي أصغر من مستوى الأهمية المحدد "0.05" فإننا نرفض فرضية العدم القائلة إن $Kappa = 0$ و نقبل الفرضية البديلة التي تقول إن $Kappa$ مختلفة معنوياً عن الصفر. أي الاتفاق بين مجموعتي المراقبين حقيقية و ليس بالمصادفة.

Symmetric Measures

	Value	Asymp. Std. Error	Approx. T	Approx. Sig
Measure of Agreement Kappa	0.676	0.088	8.879	0.000
N of Valid Cases	100			

الجدول (9): نتائج تحليل اختبار اتفاق المراقبين Observer Agreement

الاستنتاج الإحصائي

نتيجة اجراء الاختبار الاحصائي المناسب (اختبار اتفاق المراقبين) على البيانات المتوفرة تبين أن هناك اتفاقاً جيداً بين المراقبين Rater A و Rater B على تصنيف الاضطراب النفسي لدى كل مريض ($Kappa = 0.676$). هذا الاتفاق هو اتفاق حقيقي و لا يمكن رده إلى الصدفة بسبب كون قيمة Kappa مختلفة إحصائياً عن الصفر (0.000).

لقد ناقشنا، في الفقرات السابقة ، مجموعة من الاختبارات الإحصائية نلخصها بالنقاط التالية:

- تحديد فيما إذا كان متغيران يؤثران في ظاهرة معينة مرتبطين أو مستقلين.

لذلك تحدثنا عن اختبار الاستقلالية Chi-square Independence Test

• اختبار جودة التطابق Chi-square goodness-of-fit Test الذي يبحث عن وجود علاقة ما بين نتائج مشاهدة مع نتائج ناتجة عن قانون ما (مثل قوانين ماندل في الوراثة).

• اختبار تطابق المراقبين Inter-Observer agreement Test و يفيد هذا الاختبار في معرفة التوافق بين مجموعتين من المراقبين على تصنيف ظاهرة ما وفق مقياس متفق عليه.

سوف نناقش في الفقرات التالية مجموعة من الاختبارات الإحصائية غير الوسيطة التي تقابل الاختبارات الوسيطة التي تحدثنا عنها في فصل سابق. بالتحديد ، نعرض الاختبارات الخاصة باختبار فرضيات حول مجتمعين (مماثل اختبار ستودنت) واختبار فرضيات حول أكثر من مجتمعين (مماثل لاختبارات تحليل التباين ANOVA).

4.6 اختبار Mann-Whitney Test

هو اختبار غير وسيطي يستخدم لمقارنة عينتين عشوائيتين مستقلتين. لا يفترض هذا الاختبار أي شروط حول توزيع المجتمع لذلك يستخدم بدلاً لاختبار ستودنت لعينتين مستقلتين عندما تكون شروط تطبيق اختبار ستودنت غير محققة.

يعتمد هذا الاختبار على رتب البيانات Data Ranks بدلاً من المتوسطات و الانحرافات المعيارية كما جرت العادة في الاختبارات الوسيطة. (تحسب الرتب بحسب التوضيح في الصفحة التالية.) و لكن يجب الانتباه إلى هذا الاختبار لا يشترط شكل التوزيع ، و لا يأخذ بعين الاعتبار كل المعلومات المتوفرة في العينة كما في الاختبارات الوسيطة و بالتحديد اختبار ستودنت. لذلك من الطبيعي أن يكون هذا الاختبار أقل حساسية من اختبار ستودنت في اكتشاف الفروق بين المجتمعين.

الفرضيات التي يجري اختبارها هي:

فرضية العدم H_0 : العينتان مسحوبتان من مجتمعين متماثلين.

الفرضية البديلة H_1 : العينتان مسحوبتان من مجتمعين مختلفين.

ملاحظة حول حساب الرتب: يتم حساب الرتب وفق الخطوات التالية:

1. رتب البيانات في العينتين تصاعدياً من الأصغر إلى الأكبر.
2. أعط رتباً للبيانات من 1 إلى N حيث 1 هي الرتبة المعطاة لأصغر عنصر
3. إذا وجدت بيانات متساوية فإنها تعطى متوسط الرتب.



لاحظ أن هاتين الفرضيتين لا تضعان أية شروط على شكل التوزيع (لم يذكر فيهما لا متوسط أو انحراف معياري أو وسط أو ...). يمكن إعادة صياغة هاتين الفرضيتين بحسب طبيعة البيانات ، و بحيث نحدد بشكل دقيق ماذا نعني بتمائل أو اختلاف مجتمعي العينتين. على سبيل المثال، يمكن إعادة صياغة الفرضيات أعلاه بالشكل:

H_0	متوسطا المجتمعين المسحوبة منهما العينتين متساويان (أي $\mu_1 = \mu_2$)
H_1	متوسطا المجتمعين المسحوبة منهما العينتين غير متساويين (أي $\mu_1 \neq \mu_2$)

مثال (1)

أجريت تجربة لتقويم فعالية نوعين من السماد الكيميائي A و B على زيادة نمو النباتات. حيث تقاس فعالية السماد الكيميائي بارتفاع النباتات — كلما زاد ارتفاع النبتة كانت فعالية السماد أفضل. يبين الجدول التالي ارتفاعات النباتات لأجل نوعي السماد A و B.

ارتفاع النباتات (سم)	نوع السماد الكيميائي
100, 90, 82, 92, 96, 87, 110, 98	A
105, 102, 91, 89, 87, 83, 98, 97, 96, 95	B

السؤال: هل هناك فرق بين فعالية السمادين الكيميائيين المستخدمين؟

لماذا اختبار Mann-Whitney؟ إذا كنا نعرف مسبقاً أن توزيع مجتمعي العينتين في الجدول أعلاه طبيعي ، فإن الاختبار المناسب هو اختبار ستودنت لعينتين مستقلتين. و لكن بسبب أننا لانعلم توزيع المجتمعين، فإنه من الأسلم أن نستخدم اختبار Mann-Whitney غير الوسيطى.

الفرضيات المطلوبة:

فرضية العدم H_0 : لا يوجد فرق معنوي بين متوسطي المجتمعين.

الفرضية البديلة H_1 : متوسطي المجتمعين مختلفين.

الخطوات المتبعة في اختبار Mann-Whitney

نتائج اختبار Mann-Whitney

يظهر الجدولين 10 و 11 مخرجات الـ SPSS لاختبار Mann-Whitney. يبين الجدول (10) ملخصاً إحصائياً عن العينتين. حيث يظهر في الجدول (10) عدد قياسات N لكل من العينتين A و B ، متوسط الرتب Mean Rank ، و مجموع الرتب Sum of Ranks.

تحسب الرتب للبيانات بعد ترتيبها تصاعدياً. يعطى البيان الأصغر الرتبة 1 و البيان الذي يليه الرتبة 2 و هكذا. إذا تساوى قياسان أو أكثر فإننا نأخذ متوسط الرتب المقابلة لهذه القياسات ، و يوضع المتوسط على أنه رتبة لهذه القياسات. متوسط الرتب هو مجموع الرتب على عدد القياسات. على سبيل المثال ، متوسط رتب العينة A هو 9.44 (يساوي مجموع الرتب 75.50 مقسوماً على عدد قياسات العينة A "8").

يوفر الجدول (11) الإحصائيات اللازمة لاختبار الفرضيات المطروحة. القيمة الأهم في هذا الجدول هي قيمة معنوية الاختبار (القيمة المظللة). إن قيمة معنوية

الاختبار تساوي "0.965" و هي تفوق مستوى الأهمية المفروض "0.05" ، لذلك فإن اختبار Mann-Whitney غير معنوي. و بالتالي لا نستطيع رفض فرضية العدم H_0 . و نستنتج أنه لا يوجد فرق معنوي إحصائياً بين متوسط طول النباتات المُخصَّبة بالمخصب A و متوسط النباتات المُخصَّبة بالمخصب B.

Ranks			
FerlizerType	N	Mean Rank	Sum of Ranks
Heights A	8	9.44	75.50
B	10	9.55	95.50
Total	18		

الجدول (10): ملخص احصائي للعينتين بدلالة الرتب Ranks

Test Statistics	
	Heights
Mann-Whitney U	39.500
Wilcoxon W	75.500
Z	-.044
Asymp. Sig. (2-tailed)	.965
Exact Sig. [2*(1-tailed Sig.)]	.965 ^a

الجدول (11): نتائج الاختبار Mann-Whitney U Test

5.6 اختبار الإشارة Wilcoxon Signed Ranked Test

يستخدم لاختبار فرضيات حول عينتين مرتبطتين Two Related Samples. لذلك هو إجراء بديل لاختبار ستودنت لعينتين مرتبطتين و يستخدم في حالة كون

مجتمعي العينتين لا يتوزعان بشكل طبيعي. نمط المسائل التي ينطبق عليها هذا الإجراء مماثلة لتلك التي ينطبق عليها اختبار ستودنت لعينتين مرتبطتين.

يستخدم هذا الإجراء رتب Ranks البيانات لحساب الإحصائيات اللازمة لاختبار الفرضيات. بشكل مبسط ، يقوم الاختبار بحساب الفروق بين كل قياسين متقابلين من العينتين ثم تعطى رتب للفروق السالبة ، و رتب للفروق الموجبة بعد ترتيب هذه الفروق تصاعدياً بغض النظر عن الإشارة (أي حسب القيم المطلقة). من خلال هذه الفروق تحسب قيمة إحصائية Z التي تستخدم لاختبار الفرضيات. تجدر الإشارة إلى أنه من خلال استخدام حزم البرامج الإحصائية (مثل الـ SPSS) فإننا لانهتم بكل هذه التفاصيل حيث تُنفذ حزم البرامج الإحصائية كل الحسابات اللازمة.

مثال (1)

أجريت تجربة لمعرفة فيما إذا كانت القردة تفضل تنبيه الجزء A أو الجزء B من أدمغتها. في هذه التجربة تم تدريب 14 قرداً على الضغط على مفتاحين Key 1 و Key 2. عند الضغط على المفتاح Key 1 فإن تياراً كهربائياً ينبه الجزء A من دماغ القرد، و عند الضغط على المفتاح Key 2 فإن تياراً كهربائياً ينبه الجزء B. بعد تدريب هذه القردة على كيفية الضغط على المفتاحين، ترك كل قرد لمدة 15 دقيقة للضغط على المفاتيح و سجلت عدد مرات الضغط على كل مفتاح.

يظهر الجدول التالي عدد المرات التي قام كل قرد من القردة الأربعة عشر بالضغط على المفتاح Key 1 و المفتاح Key 2. على سبيل المثال، ضغط القرد 1 على المفتاح Key 1 عشرين مرة، بينما ضغط على المفتاح Key 2 أربعين مرة.

القرد	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Key 1	20	18	24	14	5	26	15	29	15	9	25	31	35	12
Key 2	40	25	38	27	31	21	32	38	25	18	32	28	33	29

السؤال: هل تظهر القياسات في الجدول أن القردة تفضل تنبيه الجزء A أو الجزء B؟

لماذا تخضع هذه المسألة لإجراء Wilcoxon؟ من الواضح أن البيانات هي بيانات تكرارية لأنها تمثل عدد مرات ضغط كل مفتاح. أيضاً توزيع مجتمع العينات غير معروف، و بالتالي لا يمكن أن نعد -ببساطة- توزيع المجتمع طبيعياً. بناءً عليه، لا يمكن تطبيق الاختبارات الوسيطية عليها. كذلك، العینتان مرتبطتان لأنهما يتعلقان بنفس الوحدة التجريبية (القرد نفسه يضغط على المفتاح الأول Key 1، و يضغط أيضاً على المفتاح Key 2)، و بالتالي الاختبار المناسب هو اختبار Wilcoxon.

الفرضيات التي نريد اختبارها:

فرضية العدم H_0 : القردة تفضل تنبيه الجزء A و الجزء B بنفس الدرجة
الفرضية البديلة H_1 : القردة تفضل تنبيه أحد الجزئين على الآخر.

يبين الجدول (12) مخرجات الـ SPSS لنتائج اختبار Wilcoxon على البيانات في الجدول أعلاه.

يبين الجدول (12.a) عدد رتب الفروق السالبة "2"، و عدد رتب الفروق الموجبة "11"، و عدد رتب الفروق الصفرية "0". كذلك يبين الجدول نفسه متوسط رتب الفروق السالبة ("2")، و متوسط رتب الفروق الموجبة ("9"). أيضاً مجموع رتب الفروق السالبة ("6")، و مجموع رتب الفروق الموجبة ("99") موضح في الجدول نفسه. إذا كانت فرضية العدم صحيحة فإن مجموع رتب الفروق الموجبة سيكون قريباً من مجموع رتب الفروق السالبة. أما إذا كان الفرق بين مجموعي الرتب كبيراً نسبياً فهذا يشير إلى أن البيانات لا تقدم دليلاً كافياً على صحة فرضية العدم. في المثال قيد الدراسة، الفرق بين مجاميع الرتب كبير (قارن 99 إلى 6). و بالتالي فإن هذا يعطي مؤشراً لإمكانية رفض فرضية العدم.

يظهر الجدول (12.b) الإحصائيات اللازمة لاختبار الفرضيات أعلاه. نلاحظ من الجدول أن اختبار الإشارة معنوي لأن معنوية الاختبار "0.003" و هي أصغر من متسوى الأهمية "0.05". نرفض فرضية العدم و نستنتج أن القردة تفضل تنبيه أحد

الأجزاء من دماغها على الجزء الآخر. يمكن الاستنتاج من الجدول (12.b) أن القردة تفضل تنبيه المنطقة A من الدماغ لأن مجموع رتب الفرق Key 2 – Key 1 الموجبة أكبر من رتب الفرق السالبة.

Ranks				
		N	Mean Rank	Sum of Ranks
Key2 - Key1	Negative Ranks	3 ^a	2.00	6.00
	Positive Ranks	11 ^b	9.00	99.00
	Ties	0 ^c		
	Total	14		

a. Key2 < Key1
b. Key2 > Key1
c. Key2 = Key1

(a)

Test Statistics^b

	Key2 - Key1
Z	-2.921 ^a
Asymp. Sig. (2-tailed)	.003

a. Based on negative ranks.
b. Wilcoxon Signed Ranks Test

(b)

جدول (12): نتائج تنفيذ الاختبار Wilcoxon على البيانات في الجدول أعلاه.

6.6 اختبار كروسكال واليس لأجل k عينة مستقلة

Kruskal-Wallis H-test (for k independent samples)

يستخدم هذا الإجراء بدلاً عن اختبار One-Way ANOVA عندما تكون مجتمعات العينات لا تتوزع توزيعاً طبيعياً أو قريباً من الطبيعي. نستخدم هذا الإجراء عندما تكون لدينا K عينة مستقلة حيث K أكبر من 2. (إذا كانت $k = 2$ فإننا نفضل استخدام Mann-Whitney test). إضافة لذلك ، العينات في هذا الاختبار مستقلة ، و بالتالي يمكن لعدد القياسات في هذه العينات أن تختلف.

الفرضيات التي يجري اختبارها

فرضية العدم H_0 : العينات مسحوبة من مجتمعات غير مختلفة.

الفرضية البديلة H_1 : على الأقل واحد من المجتمعات مختلف.

مثال Example

لنفرض أننا قمنا بتجربة لإيجاد أعداء حيويين لقتل حشرة عثة الغجر Gypsy Moth التي تتغذى على أشجار الغابات. تم تدريب أربعة أنواع مختلفة من الحشرات على مهاجمة هذه الحشرة و قتلها. بعد تدريب استمر لمدة شهر واحد تم اختيار عينة من كل نوع ، و تم إطلاقها في مكان وجود حشرة عثة الغجر لفترة زمنية ثابتة مدتها ساعتان.

يبين الجدول التالي أداء كل حشرة و المعطى بدلالة عدد المرات التي استطاعت بها كل حشرة القيام بهجوم ناجح (أي قتل عثة الغجر).

Group 1	Group 2	Group 3	Group 4
63	78	94	54
58	69	83	51

51	58	73	32
41	57	67	74
63	53	66	65
		61	80

السؤال: هل تدل الأرقام في الجدول على وجود فرق حقيقي بين أداء الحشرات المختلفة فيما يتعلق بقدرتها على مهاجمة عثة الغجر و قتلها ؟

الفرضيات المطلوب اختبارها هي:

فرضية العدم H_0 : لا يوجد فرق بين أداء الحشرات في قدرتها على مهاجمة عثة الغجر و القضاء عليها.

الفرضية البديلة H_1 : على الأقل نوع واحد من الحشرات له فعالية مختلفة.

يبين الجدول (13) نتائج اختبار Kruskal-Wallis Test على العينات. يبين الجدول (13.a) متوسط رتب عناصر كل عينة من العينات الأربعة المختلفة (القيم ضمن الإطار). على سبيل المثال متوسط رتب العينة Group 1 يساوي "8.00". من الواضح أن متوسطات رتب العينات الأربعة متقاربة إلى حد ما. و ربما يكون ذلك مؤشراً على أنه لا توجد فروق معنوية بين قدرات هذه الأنواع المختلفة من الحشرات. حتى نتأكد من هذا الاستنتاج المبدئي ننتقل إلى النتائج في الجدول (13.b).

إن اختبار kruskal-Wallis غير معنوي لأن قيمة معنوية الاختبار "0.182" أصغر من مستوى الأهمية المفروض "0.05". و بالتالي فإننا نقبل فرضية العدم التي تنص على عدم وجود فرق معنوي بين أداء الحشرات المختلفة فيما يتعلق بكفاءتها في قتل حشرة عثة الغجر.

تمرين: اكتب تقريراً علمياً يلخص النتائج السابقة. تذكر أن التقرير يجب أن لا يحتوي على مصطلحات تقنية.

Ranks

	Index	N	Mean Rank
SuccessfulHit	Group 1	5	8.00
	Group 2	5	11.10
	Group 3	6	16.17
	Group	6	10.08
	Total	22	

(a)

Test Statistics^{a,b}

	SuccessfulHit
Chi-Square	4.861
df	3
Asymp. Sig.	.182

a. Kruskal Wallis Test

b. Grouping Variable: Index

(b)

جدول (13): نتائج تطبيق اختبار Kruskal-Wallis Test على العينات في الجدول أعلاه.

7.6 اختبار Friedman ANOVA

هو أحد الاختبارات الإحصائية البديلة لاختبار تحليل التباين لعينات مترابطة Repeated Measured Test. لذلك يسمى هذا الاختبار أيضاً *Friedman two-way analysis of variance by ranks*. يعد هذا الاختبار مناسباً جداً بشكل خاص عندما تكون العينات صغيرة ، و في حالة كون البيانات اسمية أو ترتيبية (أي ليست

كمية). الهدف الأساسي لهذا الاختبار هو التحقق من فرضيات حول K مجموعة من البيانات و هذه البيانات متكررة (متراصة).

يمكن صياغة الفرضيات التي يتم اختبارها بالشكل:
 فرضية العدم H_0 : العينات المدروسة تعود إلى المجتمع نفسه (أي لا يوجد فرق معنوي بين هذه العينات).
 الفرضية البديلة H_1 : على الأقل واحدة من العينات تعود إلى مجتمع مختلف (أي على الأقل واحدة من العينات مختلفة عن بقية العينات).

مثال Example

بالعودة إلى المثال في الفقرة السابقة. لنفرض أنه تم اختيار مجموعة واحدة من مجموعات الحشرات و تم تدريبها على مهاجمة عثة الغجر لمدة شهر واحد ثم أختبرت كفاءتها على مهاجمة العثة. و بعد ذلك تم تدريب نفس مجموعة الحشرات لمدة شهر آخر ثم أختبرت قدرتها مرة ثانية. يبين الجدول التالي عدد المرات التي قامت بها كل حشرة بهجوم ناجح على عثة الغجر، و بعد ثلاث فترات تدريبية Phase 1 و Phase 2 و أخيراً Phase 3.

Phase 1	Phase 2	Phase 3
63	76	120
58	69	83
51	80	96
41	70	95
63	66	83

السؤال: هل توجد فروق معنوية بين كفاءة هذه الحشرات بعد فترات التدريب الثلاث؟

تظهر نتائج هذا الاختبار في الجدول (14). يبين الجدول (14.a) رتب المتوسط لكل عمود من الأعمدة الثلاثة في جدول البيانات أعلاه. على سبيل المثال، رتبة متوسط البيانات في العمود Phase 1 هو واحد.

يظهر الجدول (14.b) الإحصائيات اللازمة لاختبار الفرضيات الآتية الذكر. بالتحديد يظهر هذا الجدول الإحصائية Chi-square و قيمتها "10" و معنوية الاختبار و قيمتها "0.007" (القيم ضمن الدائرة في الجدول (b)). لاحظ أن الاختبار معنوي بمستوى أهمية 5% لذلك نرفض فرضية العدم و بالتالي نقبل الفرضية البديلة.

اعتماداً على ذلك يوجد فرق معنوي إحصائياً بين كفاءة الحشرات بعد فترات التدريب الثلاث المتتالية. علاوةً على ذلك فإنه يبدو من خلال رتب المتوسطات أن كفاءة الحشرات تتطور بعد كل فترة تدريب. لاحظ أن رتب المتوسطات ازدادت من 1 في المرحلة الأولى (Phase 1) إلى 3 المرحلة الثالثة (Phase 3).

Ranks	
	Mean Rank
Phase1	1.00
Phase2	2.00
Phase3	3.00

(a)

Test Statistics ^a	
N	5
Chi-Square	10.000
df	2
Asymp. Sig.	.007

a. Friedman Test

(b)

جدول (14): نتائج تطبيق اختبار Friedman على البيانات في الجدول أعلاه.

تمرين: اكتب تقريراً علمياً يلخص النتائج السابقة. تذكر أن التقرير يجب أن لا يحتوي على مصطلحات تقنية و يجب أن يجيب عن كل التساؤلات المطروحة.

الأفكار الرئيسية للفصل Key Notes

1. الاختبارات غير الوسيطة تستخدم بديلاً للاختبارات الوسيطة عندما تكون شروط الأخيرة غير محققة.
2. الاختبارات غير الوسيطة أقل مرونة من الاختبارات الوسيطة و لذلك إذا كانت شروط تطبيق الاختبارات الوسيطة محققة فيجب استخدامها.
3. البيانات المناسبة لهذه الاختبارات هي البيانات الوصفية النوعية Qualitative Data سواء أكانت بيانات ترتيبية Ordinal Data أم بيانات اسمية Nominal Data. و يمكن أيضاً استخدام الاختبارات غير الوسيطة مع البيانات الكمية مباشرة. لكن كما عرضنا سابقاً إذا كانت البيانات كمية و تخضع للتوزيع الطبيعي فإننا نفضل استخدام الاختبارات الوسيطة.
4. اختبارات كاي مربع Chi-square tests هي اختبارات غير وسيطة لاتتعلق بطبيعة توزيع المجتمع المدروس (لاتتأثر بكون المجتمع يتوزع بشكل طبيعي أم لا). و قد ركزنا في هذا الفصل على الاستخدامات التالية لهذا الاختبار:
 - ① اختبار الاستقلالية بين متغيرين Chi-square Independence Test: يستخدم بشكل خاص لمعرفة هل المتغيران اللذان يؤثران في ظاهرة ما مستقلان أو مرتبطان.
 - ② اختبار جودة التطابق Chi-square goodness-of-fit Test: يبحث عن وجود علاقة بين نتائج مشاهدة و نتائج متوقعة من قانون ما.
 - ③ اختبار تطابق المراقبين Observer Agreement Test: يستخدم

للبحث عن توافق بين مجموعتين من المراقبين حول تصنيف ظاهرة ما. يجب ملاحظة أن هذا الاختبار لا يختبر فرضيات بشكل مباشر بل يحاول تحديد قوة الإتفاق بين المراقبين من خلال حساب الإحصائية Cohen's Kappa. قوة الإتفاق بين المراقبين تعتمد على قيمة Kappa و تفسر حسب الجدول المعطى في هذا الفصل.

5. الاختبارات غير الوسيطة البديلة للاختبارات الوسيطة. عندما تكون العينات مسحوبة من مجتمعات لا تتوزع طبيعياً أو بيانات هذه العينات اسمية أو ترتيبية، فإنه لا يمكن استخدام الاختبارات الوسيطة. و بالتالي فإن استخدام الاختبارات غير الوسيطة البديلة يصبح الملجأ الوحيد. الاختبارات غير الوسيطة التي تمت دراستها في هذا الفصل مايلي:

① اختبار Mann Whitney Test الذي يستخدم في حالة عينتين عشوائيتين مستقلتين. لا يشترط هذا الاختبار أن يكون التوزيع طبيعياً. علاوة على ذلك يمكن للبيانات أن تكون اسمية أو ترتيبية. و بالتالي فإن هذا الاختبار هو اختبار بديل عن اختبار ستودنت لعينتين مستقلتين عندما تكون شروط اختبار ستودنت غير محققة.

② اختبار Wilcoxon Signed Ranked Test الذي يستخدم في حالة عينتين عشوائيتين مرتبطتين. شروط تطبيقه هي نفس شروط تطبيق اختبار Mann Whitney. و هو الاختبار غير الوسيطى البديل لاختبار ستودنت لعينتين مرتبطتين.

③ اختبار Kruskal-Wallis H-test (for k independent samples) الذي يستخدم في حالة أكثر من عينتين عشوائيتين مستقلتين. شروط

تطبيقه مماثلة لشروط تطبيق الاختبار غير الوسيطية السابقة. و هو الاختبار غير الوسيطى البديل لاختبار تحليل التباين بمتغير مستقل واحد. ④ اختبار Friedman ANOVA الذي يستخدم في حالة أكثر من عينتين عشوائيتين مرتبطتين. شروط تطبيقه مماثلة لشروط تطبيق الاختبار غير الوسيطية السابقة. و هو الاختبار غير الوسيطى البديل لاختبار تحليل التباين للعينات المترابطة.

قبل أن ننهي هذا الفصل نورد بعض فوائد الاختبارات غير الوسيطية و مساوئها حتى نفهم خصائصها أكثر ، و بالتالي استخدامها بشكل علمي صحيح.

فوائد الاختبارات غير الوسيطية Advantages of Nonparametric Test:

- إذا كان حجم العينة صغيراً فإن الاختبارات غير الوسيطية تصبح الخيار الوحيد مالم يكن توزيع المجتمع طبيعياً.
- تضع الاختبارات غير الوسيطية شروطاً أقل على البيانات، و بالتالي يمكن أن تكون مناسبة أكثر من غيرها لبعض الحالات الخاصة. إضافة إلى أن شكل الفرضيات التي يجري اختبارها بواسطة هذه الاختبارات يمكن أن تكون مناسبة للبحث المدروس أفضل من الفرضيات التي تختبرها الاختبارات الوسيطية.
- تستخدم الاختبارات غير الوسيطية في حالة البيانات التصنيفية أو الفئوية (أي مقاسة وفق مقياس ترتيبي). لا يوجد أي اختبار وسيطي يستطيع معالجة هذه البيانات.

توجد اختبارات غير وسيطية لمعالجة عينات مبنية من مشاهدات مسحوبة من مجتمعات مختلفة.

الاختبارات غير الوسيطية أسهل من ناحية التعلم و التطبيق. إضافة إلى ذلك فإن تفسير نتائج هذه الاختبارات أسهل نسبياً من تفسير نتائج الاختبارات الوسيطية.

تستخدم الاختبارات غير الوسيطية للبيانات الرتبية أو التي قيمها العددية لها قوة الرتب. بكلمات أبسط، قد يستطيع الباحث فقط أن يقول أن وحدة تجريبية معينة تملك من الصفة المدروسة أكثر من وحدة تجريبية أخرى و لكن لا يستطيع الباحث أن يكم مقدار هذه الزيادة. على سبيل المثال، إذا كنا ندرس متغير القلق فإننا يمكن أن نقول أن A أكثر قلقاً من B و لكن قد لا نستطيع تحديد كمية زيادة القلق عند A على B.

مساوئ الطرق غير الوسيطية :Shortcoming of Nonparametric Tests

لعل أهم عيب لهذه الاختبارات هي أنها أقل كفاءة من الاختبارات الوسيطية. و السبب أن الاختبارات الوسيطية تستخدم كمية أكبر من المعلومات المتوفرة في العينة (مثل المتوسط، الانحراف المعياري) بينما تستخدم الاختبارات غير الوسيطية فقط رتب البيانات.

الاختبارات الوسيطية أكثر قوة من الاختبارات غير الوسيطية. مثلاً يمكننا من خلال اختبار Two-Way ANOVA اختبار التفاعل بين المتغيرات و لكن هذه الميزة غير متوفرة بنفس القوة في الاختبارات غير الوسيطية.

8.6 تمرينات Exercises

أولاً: أسئلة عامة Comprehension Questions

1. ما شروط استخدام اختبار كاي مربع (Chi-square Test)؟ هل يمكن تطبيق اختبار كاي مربع إذا كانت البيانات مسحوبة من مجتمع لا يتوزع طبيعياً و لماذا؟
2. من خلال دراستك لاختبار كاي مربع. هل تعتقد أنه يمكن التبديل بين الفرضية البديلة و فرضية العدم بحيث تصبح فرضية العدم هي أن المتغيرات مرتبطة و أن الفرضية البديلة هي أن المتغيرات مستقلة؟
3. اذكر الشكل العام للمسائل التي ينطبق عليها اختبار الاستقلالية Independence Test. اعط مثلاً لهذا الشكل العام.
4. لنفرض أن قيمة Kappa موجبة، لنقل 0.40، وأن معنوية الاختبار الموافقة لها $\text{Approx. Sig} = 8\%$. ما قوة الاتفاق بين المراقبين؟ مالذي يمكن قوله عن الإحصائية Kappa بمستوى أهمية محدد بـ 5%؟ لماذا؟
5. أي اختبار هو أكثر قوة من حيث كشف الفروق بين المجتمعات: اختبار Wilcoxon signed ranks أو اختبار ستودنت لعينتين مترابطتين؟ علل إجابتك.
6. أعط سببين اثنين يبينان أن الاختبارات الوسيطة أفضل من الاختبارات غير الوسيطة.

ثانياً: أسئلة متعددة الخيارات Multi-choice Questions

1. أي من الخيارات التالية تعبر بشكل أدق عن العلاقة مابين اختبار الاستقلالية و اختبار توافق المراقبين؟
- A. اختبار الاستقلالية هو حالة خاصة من اختبار توافق المراقبين.

- B. اختبار توافق المراقبين هو حالة خاصة من اختبار الاستقلالية.
- C. لا توجد أية علاقة بين هذين الاختبارين كون أحدهما يبحث عن العلاقة مابين متغيرين، و الآخر يبحث عن اتفاق بين مراقبين.
- D. الإجابات A و B كلها صحيحة لأنه يمكن النظر إلى كل الاختبارات الإحصائية على أنها حالات خاصة من بعضها البعض.
- E. لا تعبر واحدة من الإجابات السابقة بشكل دقيق عن العلاقة بين هذين الاختبارين.

2. أي من الخيارات التالية تعبر بشكل دقيق عن العلاقة مابين الاختبارات الوسيطة Parametric Tests و الاختبارات غير الوسيطة Nonparametric Tests؟

- A. الاختبارات الوسيطة هي أعم من الاختبارات غير الوسيطة.
- B. الاختبارات غير الوسيطة هي أعم من الاختبارات الوسيطة.
- C. الاختبارات الوسيطة يمكن أن تستخدم دائماً مكان الاختبارات غير الوسيطة.
- D. الخيارات B و C هي أفضل تعبير للعلاقة بين هذين الاختبارين.
- E. لا تعبر واحدة من الخيارات أعلاه بشكل دقيق عن العلاقة بين الاختبارين.

3. يستخدم اختبار جودة التطابق Chi-square goodness-of-fit Test لآياً من المسائل التالية

- A. اختبار التوافق مابين نتائج مشاهدة و أخرى متوقعة وفق قانون ما.
- B. اختبار الاتفاق بين مجموعتين من المراقبين الذين يصنفون ظاهرة معينة.
- C. اختبار استقلالية متغيرين.
- D. اختبار فيما إذا كان حجر نرد متوازناً أم لا.
- E. لا يوجد اختبار اسمه اختبار جودة التطابق.

4. إن الإحصائية Cohen's Kappa هي

- A. مقياس لمدى الارتباط بين المتغيرات في اختبار الاستقلالية.
- B. مقياس لدرجة الاتفاق بين مراقبين يقومون بتصنيف ظاهرة ما.
- C. مقياس الارتباط بين متغيرين كميين.
- D. مقياس لدرجة الذكاء بين الطلاب و الطالبات.
- E. لشيء، أي أنه مجرد إحصائية لاتقيس أي شيء.

5. إذا كان المجتمع المدروس طبيعياً فإن

- A. لا يمكن استخدام الاختبارات غير الوسيطة لأنها تستخدم أصلاً عندما تكون المجتمعات غير طبيعية التوزيع.
- B. لا يمكن استخدام الاختبارات غير الوسيطة لأن التوزيع الطبيعي يؤثر سلباً على معنوية الاختبار.
- C. الاختبارات غير الوسيطة تبقى قابلة للتطبيق لأنها لاتتأثر بشكل التوزيع.
- D. كل من الخيارات A و B صحيحة لأن الاختبارات غير الوسيطة حساسة و فعالة فقط في حالات خاصة.
- E. لا واحدة من الخيارات أعلاه صحيحة.

ثالثاً: تطبيقات عملية Real-world Applications

6. يعتقد أحد علماء الأحياء أن إحدى الحشرات التي تتغذى على نبات القطن (تسمى هذه الحشرة Aphid) تبلغ ذروة تكاثرها في شهر تموز، و تتناقص قدرتها على التكاثر تدريجياً في الأشهر الثلاثة اللاحقة. فإذا علمت أن التكاثر المتوقع لهذه الحشرة يعطى بـ 1:2:3:4 (أي كل حشرة تعطي 4 حشرات في شهر تموز، 3 حشرات في شهر آب، 2 حشرة في شهر أيلول، و حشرة واحدة في تشرين الأول). و المطلوب:
- (1). مالذي يجب عليك فعله للتأكد من صحة (أو خطأ) اعتقاد عالم الأحياء؟

(2). ما الاختبار الذي يجب عليك استخدامه — تلميح: الاختبار المناسب

هو أحد اختبارات كاي مربع Chi-square Test.

7. تدعي إحدى شركات تطوير البرمجيات أنها طورت نظاماً برمجياً يستطيع

تصنيف سلاسل الحمض النووي DNA. حيث بإعطاء سلسلة DNA، يقوم

النظام البرمجي بتحديد فيما إذا كانت هذه السلسلة تنتمي للنوع الحيوي A أو B

أو C. لنفرض أن الجامعة قد طلبت منك التأكد من أن النظام البرمجي يقوم

فعلاً بالتصنيف الصحيح لسلاسل الـ DNA. و المطلوب:

(1) كيف تستطيع القيام بهذه المهمة؟

(2) ما الاختبار الإحصائي الذي يجب استخدامه لاتخاذ قرار حول

كفاءة النظام البرمجي؟

8. يعتقد أحد علماء البيولوجية أن كفاءة الأنسولين Insulin Drug في ضبط

السكر تتأثر بحالة المريض النفسية. لأجل ذلك قام هذا العالم بفحص كمية

سكر الدم عند أشخاص لا يعانون من أي توتر نفسي، و كذلك قام بفحص

كمية سكر الدم عند أشخاص يعانون من توتر نفسي. الجدول التالي يعطي

قياسات السكر عند الأشخاص الذين لا يعانون من ضغط نفسي Relaxed و

عند الأشخاص الذين يعانون من ضغط نفسي Stressed.

الوضع النفسي	كمية السكر (mg/dl)
Relaxed	90, 88, 90, 88, 98, 87, 91, 92, 89, 86
Stressed	91, 87, 100, 94, 90, 89, 91, 84, 100, 96

المطلوب:

(1) أجز اختباراً واحداً من الاختبارات التي تعلمتها في الفصول السابقة

لتحديد فيما إذا كانت العينتان في الجدول أعلاه تتوزعان و فق التوزيع

الطبيعي؟

(2) أجز اختبار Mann-Whitney U على هذه البيانات باستخدام SPSS.

(3) فسر النتائج الصادرة عن SPSS.

(4) هل الضغط النفسي له أي دور في كفاءة الأسولين؟

(5) أجز اختبار ستودنت لعينتين مستقلتين Two-Independent T

Test، هل تفسير النتائج يطابق تفسير نتائج تحليل Mann-Whitney U؟

9. لنفرض أن لدينا مجموعة من الأسماك التي تم تدريبها لمعرفة كيف تعود إلى المكان الصحيح (الذي انطلقت منه). طبقت أربع طرق تعليم مختلفة على هذه المجموعة و بفواصل زمنية مقداره 3 أشهر بين كل طريقة و بحيث لا تؤثر نتيجة التعلم السابقة في نتائج الطريقة الحالية. بعد تعليم الأسماك وفق طريقة معينة تم إطلاق الأسماك من المكان L وتم اختبار فعالية هذه الطريقة بدلالة عدد المرات التي أخطأت بها كل سمكة بالعودة إلى المكان L. الجدول التالي يبين النتائج.

Method 1	Method 2	Method 3	Method 4
5	5	4	3
7	6	6	4
8	8	7	5
2	4	3	4
0	1	0	2

المطلوب:

(1) بين إمكانية استخدام الاختبارات الوسيطة على هذه البيانات. أعط تعليلاً كافياً لإجابتك؟

(2) حدد النموذج المناسب لهذه المسألة. لماذا تعتقد أن النموذج الذي اخترته مناسب؟

- (3) حدد الاختبار الإحصائي المناسب.
- (4) استخدم الـ SPSS لإجراء الاختبار الذي اخترته في (3) ثم فسر النتائج، و اكتب تقريراً علمياً يلخص النتائج.
10. لنفرض أن المطلوب دراسة فعالية ثلاثة أنواع للأدوية A, B, C على مساعدة المدخنين على التخفيف من عدد السجائر المدخنة اليومية. بعد إعطاء كل دواء إلى عينة المدخنين تم تسجيل عدد السجائر التي يدخنها كل فرد من أفراد العينة في اليوم. الجدول التالي يبين عدد السجائر المدخنة يومياً.
- و المطلوب:

- (1) بين إمكانية استخدام الاختبارات الوسيطة على هذه البيانات. أعطِ تعليلاً كافياً لإجابتك؟
- (2) حدد النموذج المناسب لهذه المسألة. لماذا تعتقد أن النموذج الذي اخترته مناسب؟
- (3) حدد الاختبار الإحصائي المناسب.
- (4) استخدم الـ SPSS لإجراء الاختبار الذي اخترته في (3) ثم فسر النتائج، و اكتب تقريراً علمياً يلخص النتائج.

	Drug A	Drug B	Drug C
عدد السجائر المدخنة يومياً	12	16	10
	8	24	8
	16	16	9
	30	18	12
	18	9	6
	22	19	15

★★★★★★



تطبيقات عملية

Real-World Applications

يهدف هذه الفصل إلى عرض عددٍ من التطبيقات العملية التي تعطي القارئ الخبرة الكافية للتعرف على كيفية التعامل مع بيانات مختلفة. سوف نبدأ بمجموعة من التدريبات الصغيرة ، و من ثم ننتقل إلى التعامل مع بيانات كبيرة.

1.7 التطبيق الأول The First Application

لنفترض أنه تم إدخال إجراء معين على عينة عشوائية من 10 عاملين في مؤسسة معينة. المجموعة الأولى مؤلفة من خمسة عمال لديهم خبرة محددة بالمتغير التالي (الخبرة بالشهر):

$$X = 13, 22, 30, 35, 40$$

أي العامل الأول لديه 13 شهراً خبرة ، العامل الثاني لديه 22 شهراً خبرة ، العامل الثالث لديه 30 شهر خبرة ، العامل الرابع لديه 35 شهراً خبرة ، وأخيراً العامل الخامس لديه 40 شهراً خبرة.

المجموعة الثانية مؤلفة من خمسة عمال لديهم خبرة (بعدد الأشهر) محددة بالمتغير التالي:

$$X = 4, 9, 12, 14, 23$$

ليكن المتغير Y يدل على أن العمال قد تعلموا كيفية استخدام الإجراء ($Y = 1$) أو لم يتعلموا كيفية استخدام الإجراء ($Y = 0$). و لنفرض أن المجموعة الأولى تعلمت كيفية استخدام الإجراء لذلك Y المقابلة لهم تساوي 1 ($Y=1$ لأجل هذه المجموعة). و لنفرض أن المجموعة الثانية من العمال لم يتعلموا استخدام الإجراء لذلك $Y=0$ لأجل هذه المجموعة.

السؤال: إذا كانت لديك هذه البيانات ، أي اختبار من الاختبارات الإحصائية سوف تستخدم؟

في الواقع الجواب على هذا السؤال ليس سهلاً و يعتمد بشكل مباشر على الإستنتاجات التي نريد الوصول إليها من خلال هذه البيانات. لكي نوضح ذلك ، سوف نطرح بعض الأسئلة.

1. هل متوسط الخبرة لدى العاملين الذين تعلموا الإجراء مساوٍ لمتوسط سنوات الخبرة لدى العاملين الذين لم يتعلموا الإجراء؟ لاحظ أن البيانات قادرة على الإجابة عن هذا السؤال لأن لدينا: سنوات الخبرة للعاملين الذين تعلموا و سنوات الخبرة للذين لم يتعلموا الإجراء.

2. هل يمكن التنبؤ باحتمال نجاح عامل ما في استخدام الإجراء اعتماداً على سنوات الخبرة؟ لاحظ أن لدينا البيانات اللازمة التي تربط احتمال نجاح شخص و سنوات الخبرة في نموذج. بيد أننا نحتاج إلى تقنية تسمى تحليل الانحدار اللوجستي لإيجاد مثل هذا النموذج. سوف ندرس مواضيع الانحدار اللوجستي في مقرر برامج إحصائية (2).

الفرضية الإحصائية

سوف نكتب فقط الفرضيات الخاصة بالسؤال الأول لأن السؤال الثاني خارج موضوع هذا الكتاب.

فرضية العدم H_0 : لا يوجد فرق معنوي إحصائياً بين متوسط الخبرة لدى العاملين الذين تعلموا الإجراء و متوسط الخبرة لدى الذين لم يتعلموا الإجراء.
الفرضية البديلة H_1 : يوجد فرق معنوي إحصائياً بين متوسط الخبرة لدى العاملين الذين تعلموا الإجراء و متوسط الخبرة لدى الذين لم يتعلموا الإجراء.

الاختبارات الإحصائية

سوف نفترض أن البيانات تتوزع وفق التوزيع الطبيعي. لدينا ظاهرة هي سنوات الخبرة و لدينا متغير مستقل يؤثر في هذه الظاهرة هو سنوات الخبرة و له مستويان. نعلم أن هذه المسألة تطابق نموذج المسائل التي ينطبق عليها إختبار ستودنت لعينتين. يمكن بسهولة استبعاد إختبار ستودنا لعينتين مرتبطتين لأن أفراد العينتين مختلفين. لذلك الإختبار المناسب هو اختبار ستودنت لعينتين مستقلتين.

إجراء اختبار ستودنت لعينتين مستقلتين

قبل البدء باستخدام اختبار ستودنت لعينتين مستقلتين ، يجب تفريغ البيانات في شاشة الإدخال. يتم إدخال البيانات في شاشة الـ SPSS كما يظهر في الشكل (1).

يتم تفعيل اختبار ستودنت كما يلي:

اختار القائمة Analyze ، الأمر Compare Means و من ثم اختار الأمر Independent-Samples T test لكي يفتح مربع الحوار الخاص بهذا الأمر. لمعرفة كيفية التعامل مع مربعات حوار الاختبارات الإحصائية ، اذهب إلى الملاحق A, B, C.

بعد وضع المتغيرات Months و Success في الأماكن المخصصة لهم و النقر على المفتاح Ok نحصل على النتائج.

	Months	Success
1	13	1
2	22	1
3	30	1
4	35	1
5	40	1
6	4	0
7	9	0
8	12	0
9	14	0
10	23	0

الشكل (1): تفريغ البيانات في شاشة الـ SPSS

مناقشة خرج الـ SPSS

يظهر تحليل ستودنت من خلال الجدولين التاليين. أولاً ، جدول الاحصاء الوصفي الذي يصف العينتين من خلال المتوسط ، الانحراف المعياري ، و غيرها.

Group Statistics

Success		N	Mean	Std. Deviation	Std. Error Mean
Months	Failed	5	12.40	7.021	3.140
	Succeeded	5	28.00	10.700	4.785

لاحظ أن متوسط سنوات الخبرة لدى الذين فشلوا في تعلم الإجراء (failed) يساوي 12.40 و هو أقل من متوسط سنوات الخبرة لدى الذين نجحوا في تعلم الإجراء (succeeded) الذي يساوي 28.00. فهل يدل هذا الفرق الظاهري بين المتوسطين على فرق حقيقي بين المتوسطين؟ تتم الإجابة عن هذا السؤال من خلال الجدول التالي.

ثانياً ، إحصائيات اختبار ستودنت الذي نعرضه في الجدول التالي.

Independent Samples Test

		اختبار تساوي التباينات		t-test for Equality of Means						
		Levene's Test for Equality of Variances		الاحصائيات اللازمة لاختبار تساوي المتوسطات					95% Confidence Interval of the Difference	
		F	Sig.	T	Df	Sig.	Mean Difference	Std. Error Difference	Lower	Upper
Months	Equal variances assumed	1.35	.27	-2.7	8	.026	-15.6	5.72	-28.7	-2.4
	Equal variances not assumed			-2.7	6.9	.030	-15.6	5.72	-29.1	-2.0

نلاحظ من الجدول أن اختبار تساوي التباينات Levene's Test غير معنوي مما يدل على أنه لا فرق معنوي بين تباينات المجتمعين (العيتان متجانستان من حيث التباين). بسبب تساوي التباينات فإننا نختار الإحصائيات في السطر المشار إليه بالسهم لاختبار فرضيات تساوي المتوسطات. نلاحظ من هذه الإحصائيات أن اختبار تساوي المتوسطات معنوي لأن معنوية الاختبار Sig تساوي 0.026 و هي أصغر من مستوى الأهمية المفروض 0.05. النتيجة: رفض فرضية العدم و قبول فرضية عدم تساوي متوسطات سنوات الخبرة لدى أفراد العيتين.

2.7 التطبيق الثاني The Second Application

بيانات هذا التطبيق مأخوذة من إحدى الدراسات الشاملة حول داء السكري. هناك عدد كبير من المتغيرات المدروسة التي يبينها الجدول الجزئي التالي. من هذه

المتغيرات نذكر الجنس، نوع الداء السكري، العمر، ارتفاع الضغط الشرياني، القلق، عامل الوراثة من الأب أو الأم، إلخ.

Glucose	Mother	Father	Stress	Hypertension	Smoking	Age	Type2	Type1	Sex
154	Yes	Yes	No	No	No	55	Yes	0	Female
437	No	No	No	No	Yes	52	Yes	0	Male
264	No	No	No	Yes	No	64	Yes	0	Female
185	No	No	No	Yes	No	63	Yes	0	Female
316	No	No	No	No	No	16	No	1	Female
141	No	No	No	Yes	No	67	Yes	0	Female
231	No	No	No	Yes	No	45	Yes	0	Female
260	Yes	No	No	No	Yes	54	Yes	0	Male
236	No	No	No	Yes	No	53	Yes	0	Female
184	No	No	No	Yes	No	45	Yes	0	Female
314	No	No	No	No	No	56	Yes	0	Male
296	No	No	Yes	No	No	53	Yes	0	Female
203	No	No	No	No	No	54	Yes	0	Female
346	Yes	No	No	No	No	49	Yes	0	Female
167	Yes	No	No	Yes	No	36	Yes	0	Male

Male	0	Yes	45	Yes	No	No	No	No	519
Female	0	Yes	42	No	Yes	No	No	Yes	200
Female	0	Yes	82	No	No	No	No	No	193
Female	0	Yes	51	No	Yes	No	No	No	236
Male	0	yes	62	No	Yes	No	No	No	171
Male	1	No	28	Yes	No	No	No	Yes	207
Male	0	Yes	50	Yes	No	No	No	No	160
Male	0	Yes	48	No	No	No	Yes	No	264
Female	0	Yes	42	No	No	No	Yes	No	218
Female	1	No	8	No	No	No	No	No	330

البيانات الكاملة حول هذه الدراسة عن داء السكري و المتغيرات المستقلة المدروسة معروضة في الملحق D. من الأفضل ، كتمرين للطلاب ، تفرغ هذه البيانات في شاشة الـ SPSS و من ثم إجراء بعض التحليلات عليها.

سوف نقوم هنا بإجراء بعض التحليلات الإحصائية على هذه البيانات. الهدف هو إعطاء القارئ المهارة و طريقة التفكير المنطقي في أثناء دراسة بيانات إحصائية. لابد من التنويه هنا بأن التحليلات و الاستنتاجات المعروضة في هذه الفقرة هي جزء من التحليلات الممكنة على هذه البيانات و جزء أيضاً من النتائج التي يمكن الحصول عليها من البيانات. يمكن للقارئ الحصول على استنتاجات أخرى من البيانات.

من الأسئلة المهمة التي يمكن طرحها هي السؤال التالي: هل هناك علاقة ما بين الجنس Sex و نوع الداء السكري (Type 1 أو Type 2)؟

قبل الإجابة عن هذا السؤال يجب أن نحدد بدقة ما يلي:

- ما البيانات المطلوبة للإجابة عن هذا السؤال؟
- هل البيانات المطلوبة متوفرة في العينة المدروسة؟
- ما التحليل الإحصائي المناسب على البيانات بحيث نجيب بشكل دقيق عن السؤال؟

بطبيعة الحال حتى نجيب عن السؤال المطروح يجب أن تكون لدينا بيانات عن جنس كل مريض في العينة و معلومات أخرى عن نوع الداء السكري لدى هذه المريض. بالنظر بدقة إلى البيانات و بالتحديد إلى الجدول أعلاه نجد أن هناك بيانات عن جنس كل مريض (العمود Sex) و هناك معلومات عن نوع الداء السكري (العمود Type). و بالتالي فإن جميع المعلومات الضرورية للإجابة عن هذا السؤال متوفرة. يبقى أن نحدد الاختبار الإحصائي المناسب. هنا يجب أن نكون حذرين جداً لأن أي اختبار خاطئ للاختبار سوف يترتب عليه نتائج ربما تكون خاطئة تماماً. نوع البيانات و السؤال المطروح أساسية جداً لاختيار الاختبار الإحصائي المناسب. لاحظ أن السؤال يبحث عن علاقة بين متغيرين هما الجنس و نوع الداء السكري. على هذا ربما لدينا خياران: (1) اختبار الارتباط لبيرسون أو (2) اختبار الاستقلالية لـ Chi-square. و لكن بالتفحص الدقيق للبيانات نجد أن هذين المتغيرين كفيان (ليسا كميين) و كل متغير له مستويان اثنان — الجنس (ذكر، أنثى) و نوع الداء السكري (Type 1 و Type 2). كذلك السؤال يبحث عن علاقة من الشكل هل تغيرات مستويات أحد المتغيرين ترتبط (أو مستقلة) عن تغيرات المتغير الثاني. من خلال هذا التحليل المنطقي و البسيط يمكن أن نستنتج تحليل الارتباط. و بالتالي فإن الاختبار المناسب هو اختبار الاستقلالية لـ Chi-square.

نتائج الاختبار

تظهر النتائج من خلال الجداول الثلاثة التالية. استخدامات هذه الجداول كيفية قراءتها يجب أن تكون مألوفة بالنسبة إلى القارئ. لذلك سوف أكتفي بتعليق بسيط على النتائج.

الجدول الأول يعطي وصفاً للبيانات حيث يظهر عدد الحالات الموجودة Valid و النسبة المئوية لها Percent . كذلك يظهر فيما إذا وجدت أية قيم غير مدخلة Missing. على الرغم من بساطة المعلومات التي يعطيها هذا الجدول إلا أنها مهمة للغاية لأنها تساعدنا على اكتشاف أية إدخلات خاطئة للبيانات. على سبيل المثال ، إذا وجدت قياسات غير مدخلة فهذا يمكن أن يؤشر إلى أننا قد نسينا إدخال بعض البيانات. بالنظر إلى الجدول نجد أن لدينا 200 بيان مدخل. كذلك لا توجد بيانات مفقودة (غير مدخلة) ، حيث Missing تساوي الصفر.

Case Processing Summary						
	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Gender * Type 2	200	100.0%	0	.0%	200	100.0%

يبين الجدول التالي تعداد النسب المئوية للذكور و الإناث الذين لديهم داء سكري من النوع الثاني yes و الذين لديهم داء سكري من النوع الأول no. على سبيل المثال، لدينا 5 إناث مصابات بالنوع الأول من الداء السكري.

نلفت نظر القارئ إلى النسب المئوية المظللة. لاحظ أن نسبة الإناث اللواتي يعانين من الداء السكري من النوع الثاني هي 95.4% و أن نسبة الذكور الذين يعانون من الداء السكري من النوع الثاني هي 91.2% (نسب متقاربة بلا شك). كذلك يمكن الملاحظة من الجدول نفسه أن نسب الذكور و الإناث الذين يعانون من الداء السكري من النوع الأول متقاربة (على الترتيب 4.6% و 8.8%). يشير هذا التقارب في النسب إلى عدم وجود علاقة بين متغير الجنس و متغير نوع الداء السكري. أي أن الذكور و الإناث يصابون بنفس الدرجة بنوعي الداء السكري.

Gender * Type 2 Crosstabulation

			Type 2		Total
			No	Yes	
Gender	Female	Count	5	104	109
		Expected Count	7.1	101.9	109.0
		% within Gender	4.6%	95.4%	100.0%
	Male	Count	8	83	91
		Expected Count	5.9	85.1	91.0
		% within Gender	8.8%	91.2%	100.0%
Total	Count		13	187	200
	Expected Count		13.0	187.0	200.0
	% within Gender		6.5%	93.5%	100.0%

يبين الجدول التالي نتائج اختبار الاستقلالية لـ Chi-square. كما نلاحظ من الجدول أن الاختبار ليس معنوياً لأن قيمة Sig هي 0.230 و هي أكبر من مستوى الأهمية المفروض 5%. و بناء عليه فإننا لا نستطيع رفض فرضية الاستقلالية (فرضية العدم). و بالتالي نستنتج أن متغير الجنس مستقل عن متغير نوع الداء السكري. بكلمات أبسط، كلا الجنسين يصاب بنوعي الداء السكري بنفس الدرجة.

Chi-Square Tests

	Value	Df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	1.442 ^(b)	1	.230		
Continuity Correction ^(a)	.833	1	.361		
Likelihood Ratio	1.440	1	.230		
Fisher's Exact Test				.261	.181
N of Valid Cases	200				

a Computed only for a 2x2 table

b 0 cells (.0%) have expected count less than 5. The minimum expected count is 5.92.

سؤال: هل هناك علاقة بين متغير نوع الداء السكري و درجة القلق؟

لاحظ أن الإجابة عن هذا السؤال ممكنة أيضاً لأنه لدينا بيانات عن نوع الداء السكري و درجة القلق (Stress). كذلك فإن اختبار الاستقلالية χ^2 هو الاختبار المناسب لهذه البيانات. لماذا؟

نتائج الاختبار

يُترك تفسير هذه النتائج للقارئ كتمرين.

إرشادات: عند تفسير النتائج تأكد من الإجابة عن مجموعة من الأسئلة المهمة كما فعلنا في المثال السابق. من بين الأسئلة المهمة نذكر ما يلي :

- هل جميع القيم مدخلة؟ و هل توجد أية قيم مفقودة؟ تلميح: استخدم النتائج في الجدول الأول.
- هل النسب المئوية عبر مستويات المتغيرين تشير إلى وجود علاقة بين متغير نوع الداء السكري و متغير القلق؟ تلميح: استخدم النتائج في الجدول الثاني.
- هل الاختبار معنوي؟ تلميح: استخدم النتائج في الجدول الثالث.
- ما الإستنتاج الذي يمكن التوصل إليه؟
- هل المعلومات في الجداول تساعد في تقييم قوة الارتباط بين المتغيرين؟

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Stress * Type 2	200	100.0%	0	.0%	200	100.0%

Stress * Type 2 Crosstabulation

	Type 2	Total
	Yes	

			No		
Stress no	Count		11	149	160
	Expected Count		10.4	149.6	160.0
	% within Stress		6.9%	93.1%	100.0%
yes	Count		2	38	40
	Expected Count		2.6	37.4	40.0
	% within Stress		5.0%	95.0%	100.0%
Total	Count		13	187	200
	Expected Count		13.0	187.0	200.0
	% within Stress		6.5%	93.5%	100.0%

Chi-Square Tests

	Value	Df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.185 ^(b)	1	.667		
Continuity Correction ^(a)	.005	1	.943		
Likelihood Ratio	.196	1	.658		
Fisher's Exact Test				1.000	.498
N of Valid Cases	200				

a. Computed only for a 2x2 table

b. 1 cells (25.0%) have expected count less than 5. The minimum expected count is 2.60.

سؤال: هل متوسط كمية السكر في الدم متساوية في نوعي الداء السكري؟

يمكننا القيام بالإجابة عن هذا السؤال لأن هناك بيانات عن كمية سكر الدم لدى المرضى. فلدينا قياسات كمية سكر الدم لدى مرضى النوع الأول من الداء

السكري Type 1 ، و قياسات كمية سكر الدم لدى مرضى النوع الثاني من الداء السكري Type 2. أيضاً يمكن القول إن الاختبار المناسب للإجابة عن هذا السؤال هو اختبار ستودنت للعينات المستقلة Two – Independent Sample T Test. لماذا؟

نتائج الاختبار

تظهر النتائج من خلال الجداول التالية. استخدامات هذه الجداول و كيفية قراءتها يجب أن تكون مألوفة بالنسبة للقارئ. لذلك سوف أكتفي بتعليق بسيط على النتائج.

يبين الجدول الأول وصفاً إحصائياً للعينات. كما هو ملاحظ في الجدول فإن متوسط كمية السكر في المصابين بالنوع الثاني من داء السكري (السطر المقابل لـ No) هو 220.29 و بانحراف معياري 86.783. أما متوسط كمية السكر في دم المصابين بالنوع الأول من داء السكري (السطر المقابل لـ Yes) هو 286.92 و بانحراف معياري 95.123. كما يتضح من الجدول أن متوسط سكر الدم لدى المرضى بالنوع الأول أكبر من متوسط سكر الدم لدى المرضى بالنوع الثاني.

Group Statistics

		N	Mean	Std. Deviation	Std. Error Mean
Glucose	No	187	220.29	86.783	6.346
	Yes	13	286.92	95.123	26.382

يبين الجدول التالي الإحصائيات اللازمة لاختبار الفرضيات.

Independent Samples Test

	Levene's Test for Equality of Variances		t-test for Equality Means							
	F	Sig.	T	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Differences		
								Lower	Upper	
Glucose	Equal variances assumed	.088	.767	-2.66	198	.008	-66.62	-17.24	-116.01	-17.24
	Equal variances not assumed			-2.45	13.4	.028	-66.62	-8.19	-125.06	-8.19

هناك نوعان للفرضيات. الفرضيات الخاصة بتجانس التباين و يتم اختبارها من خلال قيمة F و قيمة معنوية الاختبار المقابلة لها (الجزء المظلل من الجدول). لاحظ من الجدول أن اختبار فرضيات التباين ليس معنوياً لأن قيمة معنوية الاختبار $Sig = 0.767$ هي أكبر من مستوى الأهمية المفروض 5%. بناءً عليه لا نستطيع رفض فرضية تجانس التباينات (فرضية العدم). و بالتالي فإن التباينات متجانسة.

الفرضيات الخاصة بتساوي المتوسطات. يتم ذلك من خلال قيم معنويات الاختبار في العمود المسمى (Sig (2-tailed). نلاحظ أن الاختبار معنوي لأن قيمة معنوية الاختبار $Sig = 0.008$ و هي أصغر من مستوى الأهمية المفروض 5%. (لاحظ أننا استخدمنا السطر المقابل لـ Equal Variances assumed لأن تباينات العينتين متجانس). و بالتالي نرفض فرضية العدم و نقبل الفرضية البديلة التي تنص على أن المتوسطين مختلفان. يظهر أيضاً في الجدول أعلاه 95% مجال ثقة للفرق $Mean_{No} - Mean_{Yes}$ ، حيث يدل $Mean_{No}$ على متوسط

كمية سكر الدم لدى مرضى الداء السكري من النوع الثاني و يدل $Mean_{Yes}$ على متوسط كمية السكر لدى مرضى الداء السكري من النوع الأول. بما أن طرفي المجال سالبين $([-116.01, -17.243])$ فإننا نستنتج و بثقة مقدارها 95% أن متوسط كمية سكر الدم لدى المرضى الذين لديهم داء سكري من النوع الأول ($Mean_{Yes}$) أكبر من متوسط كمية سكر الدم لدى المرضى الذين لديهم داء سكري من النوع الثاني ($Mean_{No}$).

تمرين: هناك العديد من الأسئلة الأخرى المهمة التي يمكن للبيانات في الجدول أعلاه الإجابة عنها.

- اكتب هذه الأسئلة و من ثمّ برر لماذا البيانات قادرة على الإجابة عن هذه الأسئلة.
- اكتب الفرضيات المناسبة لهذه الأسئلة.
- حدد الاختبارات المناسبة لاختبار الفرضيات.
- اكتب تقرير علمي يلخص هذه النتائج.

3.7 التطبيق الثالث The Third Application

يبين الجدول التالي عدد القطع المعطوبة من الإنتاج لأربعة آلات لصنع المصابيح الكهربائية ضمن فترة المراقبة التي امتدت لعدة شهور. الآلات مستوردة من أربعة دول مختلفة الصين China ، اليابان Japan ، ألمانيا Germany ، و بريطانية Britain. حلل هذه البيانات و اكتب تقرير يلخص النتائج حول أداء الآلات الأربع.

Bretain	Germany	Japan	China
0	30	16	55
35	40	33	57

31	27	25	55
29	25	32	57
20	40	20	56
7	15	54	53
43	30	57	59
16	46	19	55

يتم تفريغ هذه البيانات في شاشة SPSS و ذلك بوضع القياسات في الجدول في عمود واحد و نعطيه اسماً و ليكن Defects و عمود آخر يحدد مستويات المتغير التابع Defects و نعطيه الاسم Origin. للمزيد حول كيفية تفريغ البيانات، اذهب إلى الملاحق.

الأسئلة التي يمكن للبيانات الإجابة عنها: لعل السؤال الوحيد المهم الذي يمكن لهذا البيانات الإجابة عنه هو: هل هناك فرق بين متوسطات عدد القطع المعطوبة عبر مستويات المتغير origin؟

الفرضيات الإحصائية

فرضية العدم H_0 : لا يوجد فرق بين متوسطات العينات الأربعة.

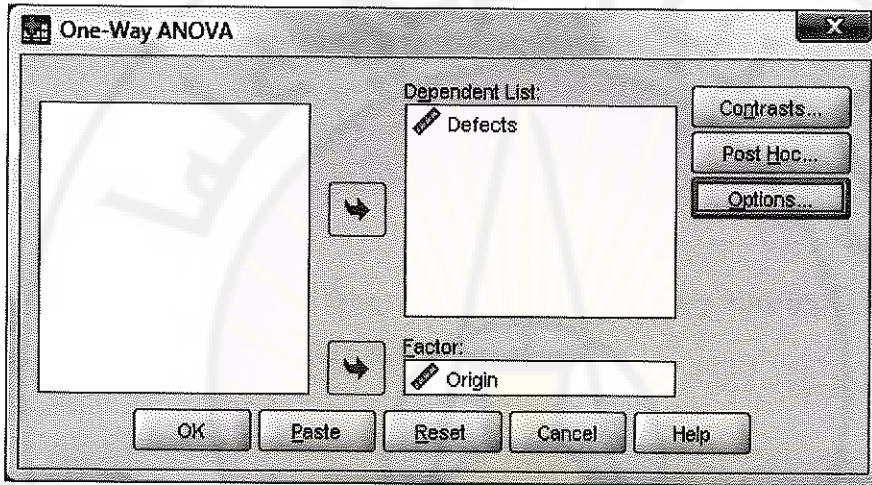
الفرضية البديلة H_1 : على الأقل واحد من المتوسطات مختلف.

الاختبار الإحصائي

بفرض أن البيانات تتوزع وفق التوزيع الطبيعي. لدينا متغير معتمد يمثل ظاهرة المصاييح المعطوبة يخضع لتأثير متغير مستقل هو البلد الذي تم استيراد المصنع منه ، و لهذا المتغير المستقل أربعة مستويات. بمقارنة هذه المسألة مع نماذج الاختبارات الوسيطة ، نجد أن هذه المسألة تطابق نموذج المسائل التي ينطبق عليها اختبار تحليل التباين بمتغير واحد مستقل One-Way ANOVA.

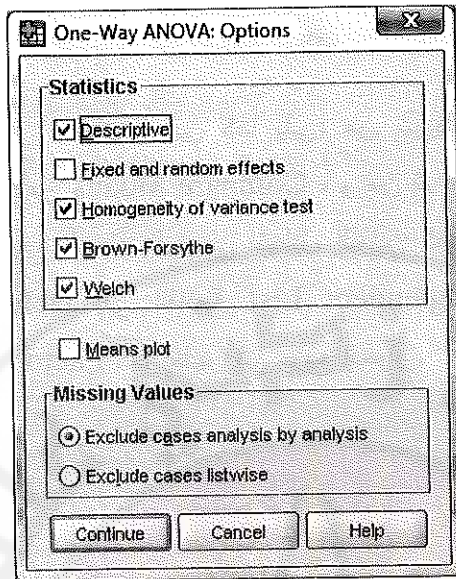
إجراء الاختبار

يتم تفعيل اختبار تحليل التباين كما يلي: من شريط القوائم ، انقر القائمة Analyze ، اختر الأمر Compare Means ، ومن ثم اختر الأمر One-way ANOVA. يفتح مربع الحوار الرئيسي لتحليل التباين كما في الشكل أدناه. ننقل المتغير التابع Defects إلى القائمة Dependent List و المتغير المستقل origin إلى القائمة Factor كما يظهر في مربع الحوار.

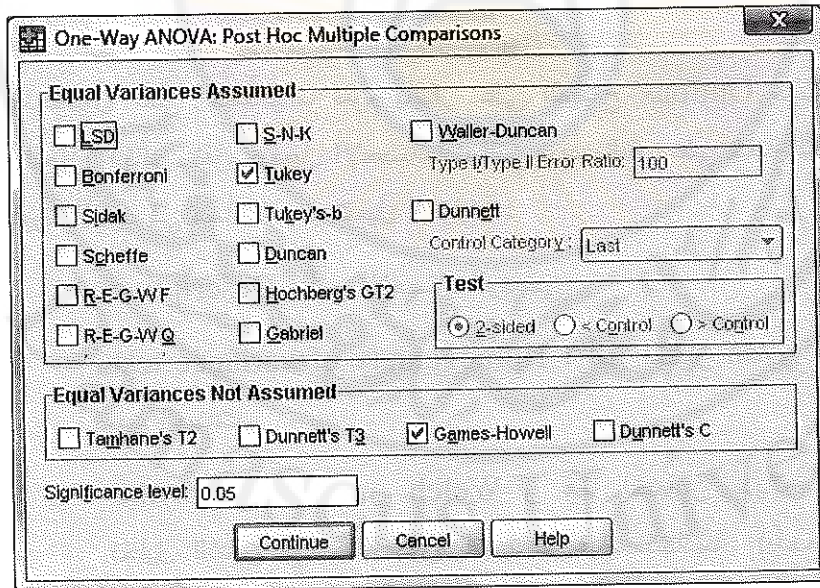


للحصول على إحصائيات نقوم بالنقر على المفتاح Options حيث يظهر مربع الحوار التالي. نقوم بتفعيل الإحصائيات التي نريد بالنقر على خانة الخيار ☒ المجاورة للأمر. في مسألتنا ، تم تفعيل الخيارات:

- Descriptive : أمر الإحصاء الوصفي.
- Homogeneity of variance test : أمر اختبار تساوي التباينات بين العينات.
- Brown-Forsythe و Welch و هما اختباران بديلان لإختبار تحليل التباين في حالة كون التباينات غير متجانسة. (راجع الفصل الخامس للمزيد من المعلومات).



بعد الإنتهاء من إنتقاء الأوامر المطلوبة ننقر على المفتاح Continue للعودة إلى مربع الحوار الرئيسي لتحليل التباين.



ننقر أيضاً على المفتاح Post Hoc فس مربع حوار تحليل التباين ليظهر مربع الحوار التالي. حيث نفعل إختبارات المقارنات الثنائية Tukey و Games-Howell. ننقر على المفتاح Continue لكي نعود إلى مربع الحوار الرئيسي لتحليل التباين.

النتائج

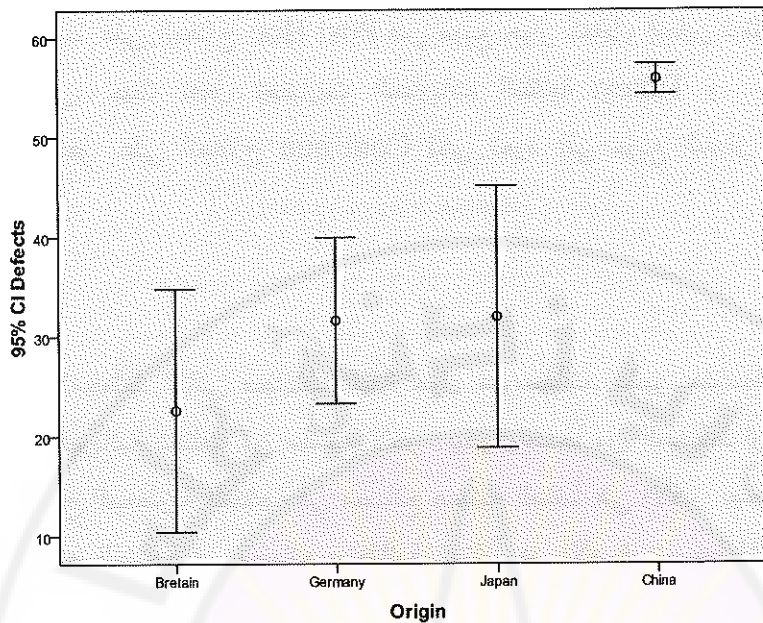
يظهر المخطط التالي مجالات الثقة حول متوسطات. تشير متوسطات عدد المصابيح الكهربائية تختلف من مصنع لآخر. يبدو واضحاً من المخطط أن متوسط عدد المصابيح المعطوبة المنتجة في مصانع صينية المنشأ هي الأكبر بينما متوسط عدد المصابيح المنتجة في مصانع بريطانية المنشأ هو الأصغر. مخططات الخطأ هي 95% مجال ثقة حول المتوسطات الأربعة.

تفسر النتائج كما يلي:

إذا كان هناك تداخل بين الرسوم البيانية لمجالات الثقة فإن الفرق بين هذه المتوسطات غير معنوي إحصائياً. إذا لم يكن هناك تداخل ، فإن هناك فرقاً معنوياً بين المتوسطات.

اعتماداً على ذلك ، يبدو من مخطط الأخطاء أنه لا يوجد فرق معنوي إحصائياً بين متوسطات عدد المصابيح المعطوبة المنتجة في مصانع كل الدول ماعدا الصين لأن الخطوط البيانية لـ 95% مجالات ثقة متداخلة. مجال الثقة الوحيد غير المتداخل هو حول متوسط عدد المصابيح المعطوبة المنتجة في مصانع صينية المنشأ. لاحظ أن هذه المجال لا يتداخل مع أي من المجالات الأخرى مما يشير إلى أنه مختلف عن هذه المجالات.

سؤال: هل متوسط عدد المصابيح المنتجة في مصانع صينية المنشأ أكبر من متوسط عدد المصابيح المنتجة في مصانع الدول الأخرى؟



يعزز هذه الاستنتاج الإحصاء الوصفي التالي. لاحظ أن مجال الثقة الوحيد غير المتداخل هو مجال الثقة حول متوسط عدد المصابيح المعطوبة المنتجة في مصانع صينية المنشأ.

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean	
					Lower Bound	Upper Bound
Bretain	8	22.63	14.569	5.151	10.44	34.81
Germany	8	31.63	9.956	3.520	23.30	39.95
Japan	8	32.00	15.712	5.555	18.86	45.14
China	8	55.88	1.808	.639	54.36	57.39
Total	32	35.53	16.845	2.978	29.46	41.60

أحد الجداول المهمة هو جدول اختبار Levene الذي يختبر تجانس التباين. نلاحظ من الجدول أن اختبار Levene معنوي لأن معنوية الاختبار Sig تساوي

0.008 و هي أصغر من مستوى الأهمية المفروض 0.05. و بالتالي فإن التباينات غير متجانسة. هذا يعني ، كما نعلم ، أن أحد شروط تحليل التباين غير محقق.

Test of Homogeneity of Variances

Defects

Levene Statistic	df1	df2	Sig.
4.845	3	28	.008

يظهر الجدول التالي ، نتائج اختبار تحليل التباين.

ANOVA

Defects

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	4865.344	3	1621.781	11.553	.000
Within Groups	3930.625	28	140.379		
Total	8795.969	31			

كما نرى من الجدول فإن التباين ينتج عن تباين نتيجة المجموعات المختلفة Between Groups وقيمه 4865.344 تباين ناتج عن الاختلافات ضمن المجموعات و الناتجة عى الفروق بين قياسات العينة الواحدة و قيمته 3930.625. لاحظ أن التباين بين المجموعات أكبر من التباين ضمن المجموعات مما يزيد من احتمال وجود فروق معنوية بين متوسطات عدد المصابيح المعطوبة المنتجة بواسطة هذه الآلات. متوسط التباينات Mean Square هو التباين المقابل مقسوم على درجات الحرية.

نلاحظ من الجدول أن اختبار التباين ANOVA معنوي ، مما يدل على أن هناك فروقاً معنوية بين متوسطات المصابيح المعطوبة المنتجة في المصانع الأربعة ؛ الاستنتاج ذاته الذي توصلنا إليه من خلال مخطط الخطأ. و لكن بما أن أحد شروط

تحليل التباين غير محقق (تجانس التباين) ، يمكن أن تكون نتائج تحليل التباين غير دقيقة و نحتاج إلى إجراء إضافي لا يتأثر بعدم تجانس التباين. لأجل ذلك تم تفعيل إختبار Welch. يظهر الجدول التالي نتائج إختبار Welch.

Robust Tests of Equality of Means

Defects

	Statistic	df1	df2	Sig.
Welch	30.477	3	12.116	.000

نلاحظ أيضاً أن إختبار Welch معنوي لأن معنوية الاختبار 0.000 هي أصغر من مستوى الأهمية 0.05. و بالتالي ، مرة أخرى تظهر هناك فروق بين المتوسطات. لتبيان أي متوسط هو المختلف ، نستخدم اختبارات المقارنات الثنائية في SPSS. لقد فعلنا سابقاً إختباري Tukey و Games-Howell و هما من اختبارات المقارنات الثنائية. الجدول التالي يبين نتائج إختبار المقارنات الثنائية. بسبب كون إختبار Tukey يشترط تساوي التباينات و أن شرط تساوي التباينات غير محقق ، فقد تم حذف نتائج إختبار Tukey من الجدول و احتفظنا فقط بنتائج إختبار Games-Howell.

تبين نتائج إختبار Games-Howell للمقارنات الثنائية و المعروضة في الجدول أدناه ، أن متوسطات عدد المصابيح المعطوبة غير مختلفة معنوياً بالنسبة للمصانع ذات المنشأ البريطاني ، الياباني ، و الألماني. لاحظ أن مستويات الأهمية المقابلة لفروق المتوسطات الخاصة بمصانع هذه الدول أكبر من مستوى الأهمية المفروض. كما و تبين ذلك مجالات الثقة حول الفروق (لاحظ مجالات الثقة حول الفروق تحوي الصفر). الفروق ذات الدلالة الإحصائية مشار إليها بالخط العريض في الجدول. كما يبين الجدول أن المتوسط المختلف هو متوسط عدد المصابيح المعطوبة المنتجة من المصنع الصيني المنشأ و أن هذه الفروق لصالح متوسط عدد المصابيح المعطوبة المصنعة في هذا المصنع. (هل تستطيع رؤية ذلك من الجدول؟)

Multiple Comparisons

Dependent Variable: Defects

(I) Origin (J) Origin		Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Games-Howell	Bretain Germany	-9.000	6.239	.498	-27.44	9.44
	Japan	-9.375	7.576	.615	-31.41	12.66
	China	-33.250*	5.191	.001	-50.30	-16.20
	German Bretain	9.000	6.239	.498	-9.44	27.44
	Japan	-.375	6.576	1.000	-19.94	19.19
	China	-24.250*	3.578	.001	-35.90	-12.60
	Japan Bretain	9.375	7.576	.615	-12.66	31.41
	Germany	.375	6.576	1.000	-19.19	19.94
	China	-23.875*	5.592	.014	-42.26	-5.49
	China Bretain	33.250*	5.191	.001	16.20	50.30
	Germany	24.250*	3.578	.001	12.60	35.90
	Japan	23.875*	5.592	.014	5.49	42.26

*. The mean difference is significant at the 0.05 level.

4.7 التطبيق الرابع The Fourth Application

أراد أحد الباحثين معرفة وجود علاقة بين الاهتمامات السياسية للشخص و درجة تعليم الشخص و جنسه. لأجل ذلك ، اختار الباحث مجموعة من الأشخاص و طلب منهم تقييم اهتمامهم السياسي على مقياس من 0 إلى 100 ، حيث الصفر تعني عدم وجود إهتمام نهائياً و 100 تعني أنه منغمس تماماً باهتماماته السياسية. تُظهر شاشة SPSS التالية جزءاً من البيانات.

	Gender	Edu_Level	Int_Politics	var	
1	Male	School	34.00		
2	Male	School	35.00		
3	Male	School	32.00		
4	Male	School	35.00		
5	Male	School	40.00		
6	Male	School	40.00		
7	Male	School	37.00		
8	Male	School	33.00		
9	Male	School	34.00		
10	Male	School	28.00		
11	Male	College	49.00		
12	Male	College	50.00		
13	Male	College	47.00		

حيث العمود Gende يرمز معلومات الجنس (ذكر Male و أنثى Female)،
العمود Edu_Level المستوى التعليمي (مدرسة School ، شهادة جامعية
College ، و شهادة فوق جامعية University) ، و أخيراً العمود Int_Politics
هو درجات الإهتمام السياسي لدى الشخص.

هناك عدة أسئلة يمكنني تهاديكم بها:

- هل هناك تأثير للجنس في الإهتمام السياسي؟
- هل هناك تأثير للمستوى التعليمي في الإهتمام السياسي؟
- هل هناك تأثير متبادل للجنس و المستوى التعليمي على الإهتمام السياسي (التفاعل)؟

الأسئلة المطروحة للإجابة عليها:

السؤال الأول: لا يوجد تأثير للجنس في الإهتمام السياسي (فرضية العدم) ، يوجد
تأثير للجنس في الإهتمام السياسي (الفرضية البديلة).
السؤال الثاني: لا يوجد تأثير للمستوى التعليمي في الإهتمام السياسي (فرضية
العدم)، يوجد تأثير للمستوى التعليمي في الإهتمام السياسي (الفرضية البديلة).

السؤال الثالث: لايوجد تفاعل بين الجنس و المستوى التعليمي في تأثيرهما في الاهتمام السياسي (فرضية العدم)، يوجد تفاعل بين الجنس و المستوى التعليمي في تأثيرهما في الاهتمام السياسي (الفرضية البديلة).

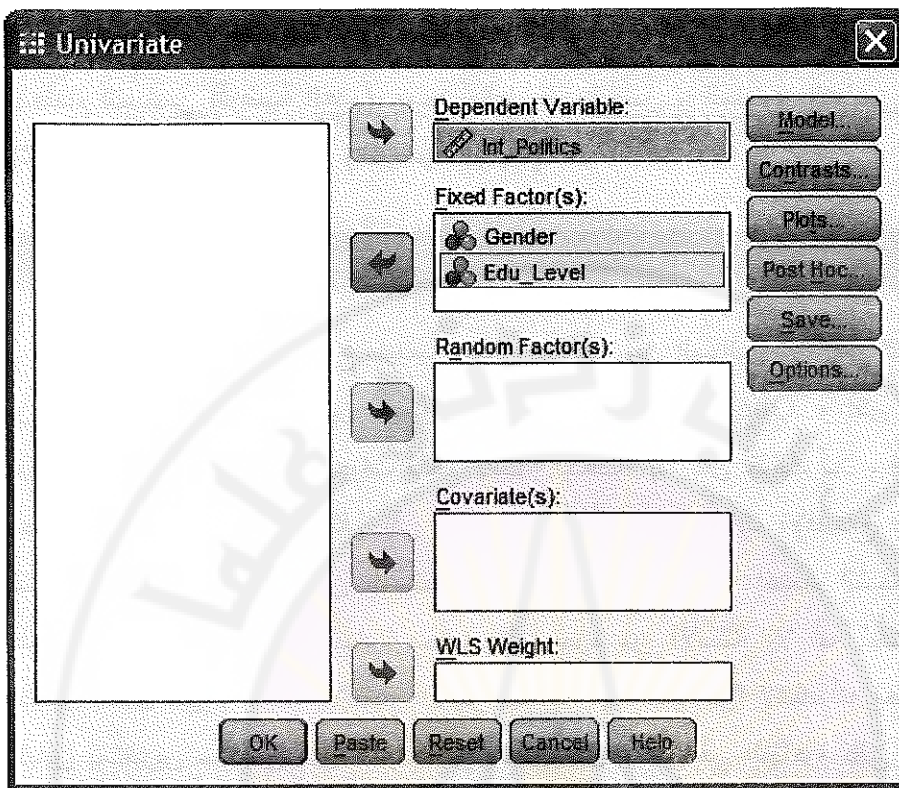
الاختبار الإحصائي المناسب

لنفترض أن البيانات تتوزع طبيعياً. يمكن الإجابة عن السؤالين الأول و الثاني من خلال تطبيق اختبار التباين لمتغير واحد One-Way ANOVA. و لكن فكر قليلاً: هل هذا يكفي لإعطاء جواب ناجع لأسئلة البحث؟ إذا كان جوابك بلا ، فأنت مُحقّق تماماً! فهذا الجواب يفنقر إلى بُعد مهم يمكن لهذا البيانات الإفصاح عنه: التأثير المتبادل بين المتغيرين المستقلين في تأثيرهما في درجة إهتمام الشخص. على الرغم من أن تحليل التباين يجيب على جزء من الأسئلة إلا أنه يُضعف التحليل بشكل كبير. إذاً ما الاختبار الأكثر مناسبة لهذه البيانات؟ لاشك أنه اختبار التباين بمتغيرين مستقلين لأن لا يأخذ فقط تأثير كل متغير في الظاهرة على حدة بل أيضاً يبحث عن أي تأثير متبادل بين المتغيرين في تأثيرهما على الظاهرة.

إجراء اختبار Two-Way ANOVA

تفعيل اختبار **Two-Way ANOVA**: انقر على القائمة Analyze ، اختر الأمر General Linear Model ، ثم الأمر Univariate... يظهر نتيجة لذلك مربع الحوار الرئيسي الخاص بهذا الاختبار في الشكل (1.6).

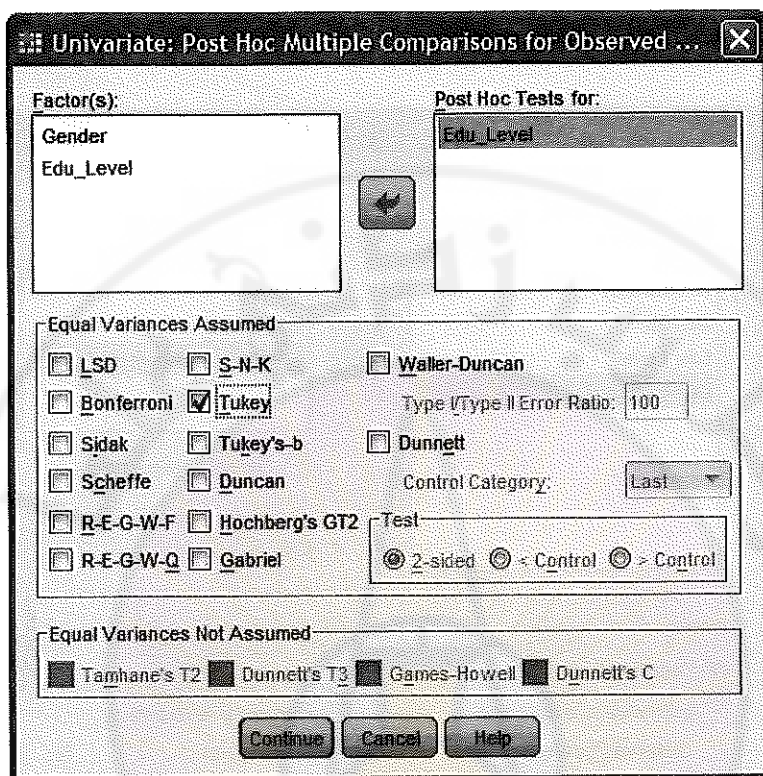
في هذا المربع تم نقل المتغير التابع Int_Politics إلى الخانة Dependent Variable و المتغيرين المستقلين Gender و Edu_Level إلى الخانة Fixed Factor(s). (لمعرفة كيف يتم نقل المتغيرات في مربعات حوار SPSS إذهب إلى الملاحق).



الشكل (1.6): مربع حوار تحليل التباين بمتغيرين Two-Way ANOVA.

إذا كنا نرغب بإجراء مقارنات ثنائية ، ننقر على المفتاح Post Hoc ليظهر مربع الحوار الخاص بهذا الأمر كما في الشكل (2.6).

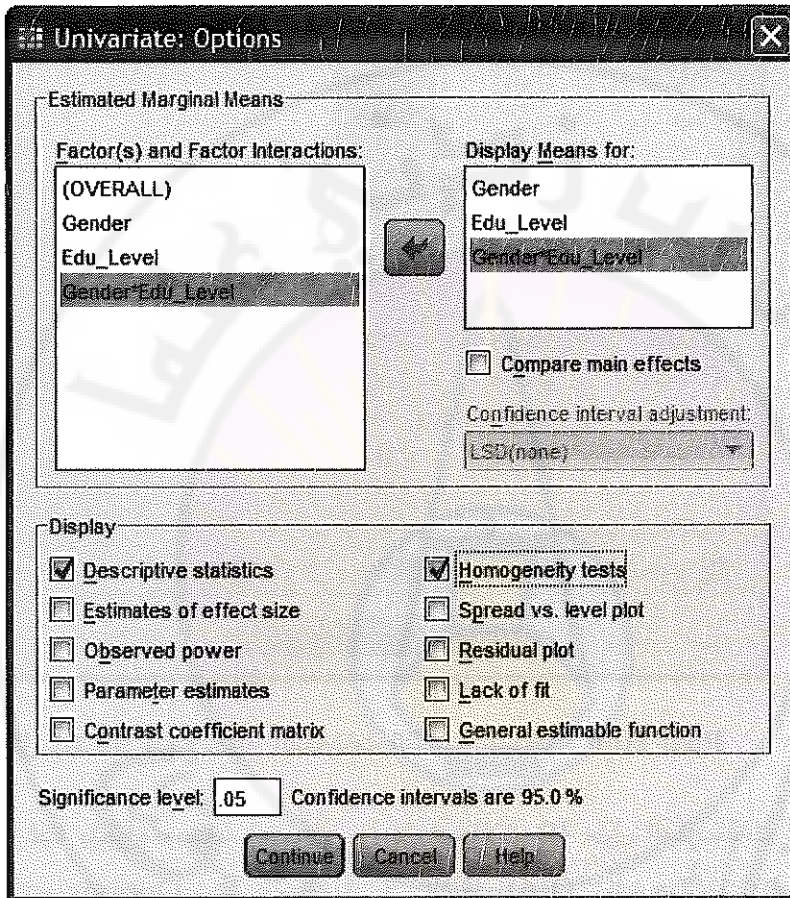
ننقل المتغير الذي نريد أن نجري مقارنات ثنائية اعتماداً على مستوياته إلى الخانة **Post Hoc Tests for**. كما يبدو من الشكل (2.6)، فقد قمنا بنقل المتغير Edu_Level إلى هذه الخانة (**Post Hoc Tests for**). لاحظ أنه لا حاجة لنقل المتغير Gender إلى هذه الخانة (لماذا؟). نقوم بتفعيل اختبار Tukey للمقارنات الثنائية كما يتضح بالشكل (2.6). ننقر بعدئذٍ على المفتاح Continue لنعود إلى مربع الحوار الأساسي (الشكل (1.6)).



الشكل (2.6): مربع حوار المقارنات الثنائية المتعددة Post Hoc

إذا رغبتنا بإجراء وصف إحصائي و بناء مجالات ثقة حول المتوسطات ، ننقر على المفتاح Options من مربع الحوار في الشكل (1.6). يظهر نتيجة لذلك مربع الحوار في الشكل (3.6). ننقل المتغيرات التي نريد عرض وصف احصائي لها إلى الخانة المعنونة **Display Means for**. لاحظ من المربع أننا نقلنا المتغيرين Edu_Level و Gender و كذلك المتغير الدال على التفاعل بين المتغيرين Gender * Edu_Level. نفعل خيار الوصف الإحصائي Descriptive Statistics و خيار اختبار تجانس التباين Homogeneity Test (لاحظ علامات التفعيل بجانب هذين الخيارين).

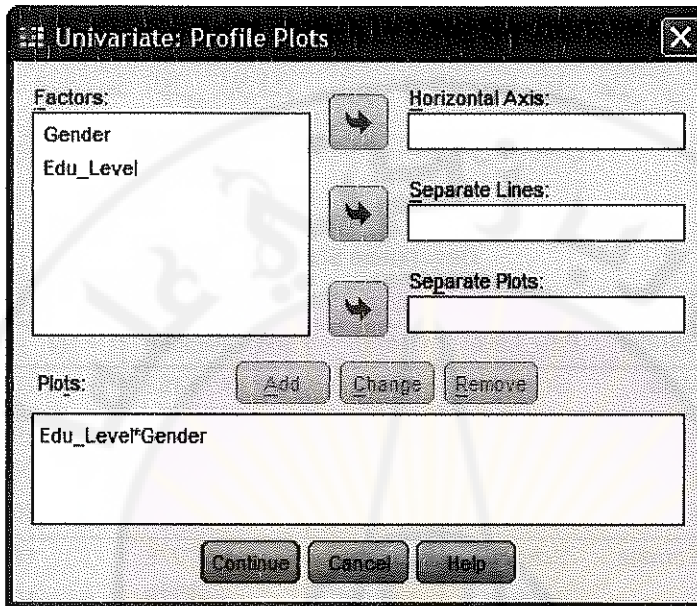
ننقر على المفتاح Continue لكي نعود إلى مربع الحوار الأساسي في الشكل (1.6).



الشكل (3.6): مربع حوار يسمح بتحديد بعض الخيارات.

أخيراً ، قد نرغب أيضاً بعرض بعض الرسوم البيانية. للقيام بذلك ننقر على المفتاح Plots في مربع الحوار في الشكل (1.6). يفتح مربع الحوار في الشكل (4.6). ننقل المتغير Edu_Level إلى الخانة المعنونة Horizontal Axis و المتغير Gender إلى خانة الحوار Separate Lines ، ثم نضغط المفتاح Add

ليظهر المتغير Gender * Edu_Level في خانة الحوار المعنونة Plots. ينتج عن ذلك رسم مخطط بياني لكل من المتغيرين.



الشكل (4.6): مربع اختيار رسوم بيانية لتمثيل البيانات.

ننقر على المفتاح Continue ثم على المفتاح OK في مربع الحوار الرئيسي ليقوم SPSS بعرض نتائج تحليل التباين بمتغيرين مستقلين.

نتائج الاختبار

تظهر نتائج اختبار تحليل التباين لمتغيرين مستقلين من خلال العديد من الجداول التي يعطي كل منها بعض الحقائق عن البيانات.

الوصف الإحصائي

يبين الجدول التالي وصفاً إحصائياً للبيانات. يبين الوصف الإحصائي المتوسط و الانحراف المعياري لدرجة الإهتمام السياسي لكل من الذكور و الإناث مقسماً بحسب

المستوى التعليمي. على سبيل المثال ، متوسط الاهتمام السياسي للجامعيين الذكور يساوي 44.1 و الانحراف المعياري 4.26745 بينما المتوسط و الانحراف المعياري لدرجة الاهتمام عند الإناث هي على الترتيب: 44.6 و 3.27278. لاحظ أيضاً أن متوسط الاهتمام بالسياسة لدى الذكور بغض النظر عن المستوى التعليمي يساوي 48.8 بينما المتوسط الكلي للإناث يساوي 47.4 مما يدل على أن المتوسطين متقاربان و بالتالي يمكن أن لا يتأثر الاهتمام السياسي بالجنس.

Descriptive Statistics

Dependent Variable: Int_Politics

Gender	Edu_Level	Mean	Std. Deviation	N
Male	School	38.2000	4.18463	10
	College	44.1000	4.26745	10
	University	64.1000	3.07137	10
	Total	48.8000	11.87841	30
Female	School	39.6000	3.27278	10
	College	44.6000	3.27278	10
	University	58.0000	6.46357	10
	Total	47.4000	9.05767	30
Total	School	38.9000	3.72615	20
	College	44.3500	3.71023	20
	University	61.0500	5.83524	20
	Total	48.1000	10.49649	60

اختبار تجانس التباين

يبين الجدول التالي تحليل تجانس التباين. لاحظ أن اختبار Levene غير معنوي بمستوى أهمية 0.05 (لماذا؟) و بالتالي فإن التباينات بين مختلف المجموعات غير مختلفة معنوياً. (تذكر أن الفرضيات الخاصة بتجانس التباين هي: التباينات بين المجموعات متساوية <<فرضية العدم>> ، التباينات مختلفة معنوياً

<<الفرضية البديلة>>). و بالتالي فإن شرط تجانس التباين لاختبار تحليل التباين محقق.

Levene's Test of Equality of Error Variances^a

Dependent Variable: Int_Politics

F	df1	df2	Sig.
1.106	5	54	.368

إختبار التباين لتساوي التباينات

يبين الجدول التالي نتائج اختبار تحليل التباين.

Tests of Between-Subjects Effects

Dependent Variable: Int_Politics

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	5525.200 ^a	5	1105.040	61.190	.000
Intercept	138816.600	1	138816.600	7686.727	.000
Gender	29.400	1	29.400	1.628	.207
Edu_Level	5328.100	2	2664.050	147.517	.000
Gender * Edu_Level	167.700	2	83.850	4.643	.014
Error	975.200	54	18.059		
Total	145317.000	60			
Corrected Total	6500.400	59			

a. R Squared = .850 (Adjusted R Squared = .836)

يوفر هذا الجدول الإحصائيات اللازمة لاختبار الفرضيات الثلاث أعلاه. الإحصائيات المهمة هي الموجودة ضمن الإطار. يُلاحظ من الجدول مايلي:

1. لا يوجد تأثير للجنس في الإهتمام السياسي (الذكور والإناث لديهم درجة الإهتمام نفسها) لأن معنوية الاختبار المقابلة لمتغير الجنس Gender أكبر من مستوى الأهمية المفروض 0.05.

2. يؤثر المستوى التعليمي على درجة الإهتمام السياسي (أي أن الأشخاص بمستويات تعليمية مختلفة ليس لديهم نفس درجات الإهتمام السياسي). لاحظ أن الإختبار معنوي بمستوى أهمية 0.05.

3. النتيجة الأهم بالنسبة لهذا الاختبار هي تقييم التفاعل بين المتغيرين المستقلين. لاحظ أن اختبار التفاعل بين المتغيرين معنوي لأن معنوية الاختبار Sig المقابلة لـ Gender * Edu_Level تساوي 0.014 و هي أصغر من مستوى الأهمية ، و بالتالي فإن هناك تفاعلاً بين المتغيرين: الجنس و المستوى التعليمي.

المقارنات الثنائية

يبين الجدول التالي اختبار Tukey للمقارنات الثنائية بين متوسطات المستويات التعليمية.

Multiple Comparisons

Int_Politics
Tukey HSD

(I) Edu_Level	(J) Edu_Level	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
School	College	-5.4500*	1.34385	.000	-8.6887	-2.2113
	University	-22.1500*	1.34385	.000	-25.3887	-18.9113
College	School	5.4500*	1.34385	.000	2.2113	8.6887
	University	-16.7000*	1.34385	.000	-19.9387	-13.4613
University	School	22.1500*	1.34385	.000	18.9113	25.3887
	College	16.7000*	1.34385	.000	13.4613	19.9387

Based on observed means.

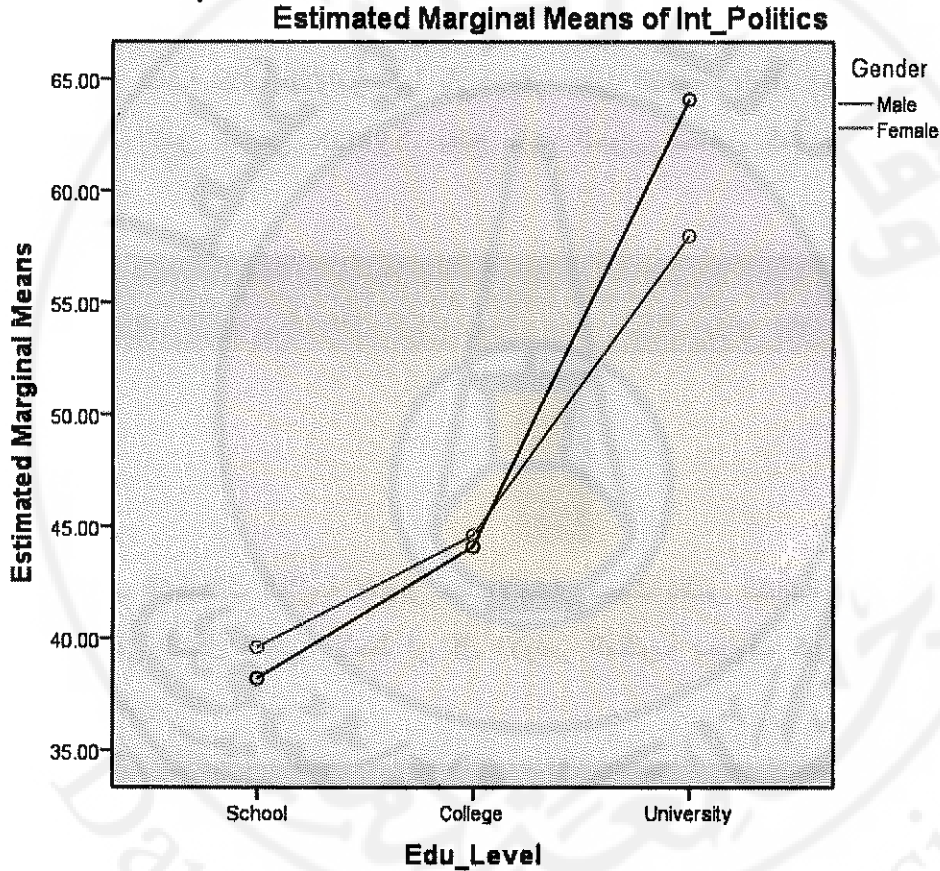
The error term is Mean Square(Error) = 18.059.

*. The mean difference is significant at the .05 level.

نستطيع أن نرى من الجدول أن هناك فروقاً بين متوسطات درجة الإهتمام السياسي بين مستويات التعليم المختلفة. كل المتوسطات مختلفة بمستوى أهمية 0.05 ، حيث المتوسط الأكبر لمستوى التعليم فوق الجامعي يليه متوسط الإهتمام عند الجامعيين ، وأخيراً متوسط الإهتمام عند الذين شهادة مدرسية.

الرسم البياني للمتوسطات

يبين المخطط التالي رسماً بيانياً لمتوسطات درجة الاهتمام لدى الذكور و الإناث مقسماً حسب المستوى التعليمي. يفيد هذا المخطط بإظهار وجود التفاعل و ذلك من خلال علاقة المستقيمات ببعضها: لا يوجد تفاعل إذا كانت المستقيمات متوازية و إلا يوجد تفاعل.

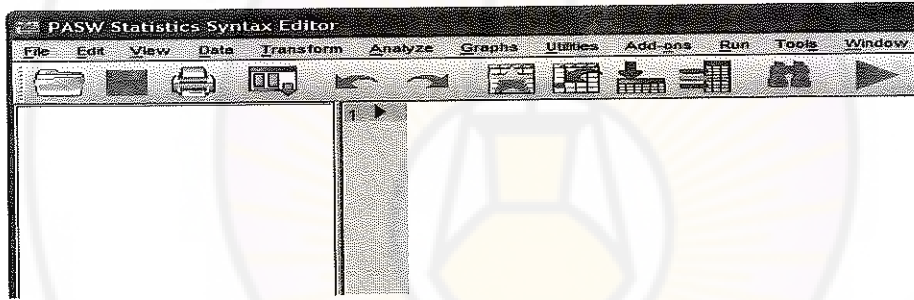


نلاحظ من المخطط أعلاه أن المستقيمات غير متوازية مما يشير إلى وجود تفاعل بين متغير الجنس و المستوى التعليمي.

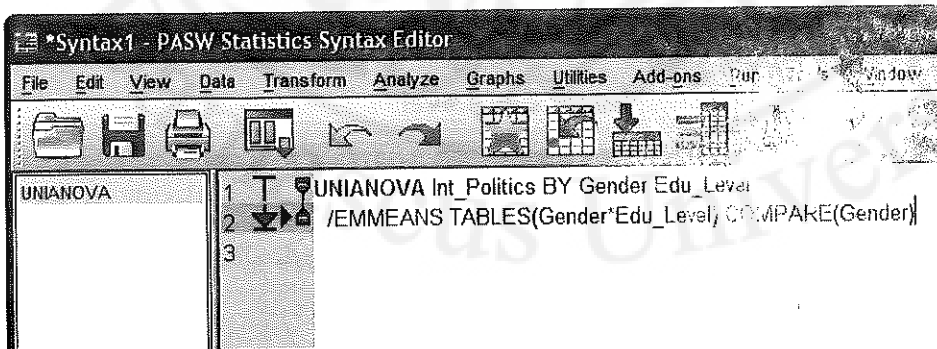
تحديد المستويات المتفاعلة

كما لاحظنا في الفقرات السابقة أن هناك تفاعلاً بين الجنس و المستوى التعليمي ، لكن لم نحدد أيّاً من مستويات المتغيرين تتفاعل. يمكن القيام بذلك اختبار من خلال الاختبار simple main effects الذي يبحث عن اختلاف متوسطات الذكور و الإناث عند كل مستوى تعليمي. للأسف لا يملك SPSS إجراءً وحيداً للقيام بذلك و لابد للمستخدم أن يكتب بعض التعليمات البسيطة لإنجاز الاختبار. سوف نقوم بكتابة مجموعة التعليمات اللازمة كما يلي:

نفتح شاشة كتابة الأوامر كما يلي: انقر على القائمة File، اختر الأمر New، ثم الأمر الفرعي Syntax. يفتح مربع الحوار الخاص كما في الشكل التالي.



نكتب التعليمات اللازمة لإنجاز هذا الاختبار بحيث تظهر الشاشة أعلاه كما هو مبين في الشكل التالي. لاحظ أن كل الأوامر المكتوبة بحروف كبيرة مطلوبة تكتب كما هي.



مجموعة الأوامر في الشكل أعلاه بسيطة و هي عامة و يمكن استخدامها مع أي بيانات. فيما يلي شرح لهذه الأوامر لكي تصبح مألوفة للطالب.

UNIANOVA	تجعل SPSS يستخدم أمر تحليل التباين لمتغيرين مستقلين.
Int_Politics BY Gender Edu_Level	المتغير التابع Int_Politics مدروسٌ بحسب المتغيرين المستقلين Gender , Edu_Level (يجب ترك فراغ بين الأوامر و المتغيرات)
/EMMEANS	تجعل SPSS بحسب المتوسطات الهامشية المقدرة.
TABLES(Gender*Edu_Level)	تجعل SPSS يولد إحصائيات التفاعل بين مختلف المستويات. يجب وضع المتغيرين المستقلين بعد الأمر TABLES و هذان المتغيران مفصولين بـ "*"
COMPARE(Gender)	تجعل SPSS يقارن حدود التفاعل بين مستويات متغير الجنس Gender

بعد الإنتهاء من كتابة الأوامر كما في الشكل أعلاه ، ننقر على المفتاح المشار إليه بالسهم ليتم تنفيذ الأمر و إظهار النتائج.

يبين الجدول التالي نتائج هذا الاختبار. تفيد الإحصائيات في هذا الجدول بتحديد فيما إذا وجدت اختلافات بين متوسطات اهتمام الجنسين في السياسة عند كل مستوى من مستويات التعليم. كما يلاحظ من الجدول ، لا يوجد فرق معنوي إحصائياً بين متوسطات درجة الإهتمام السياسي عند مستوى التعليم مدرسة School و مستوى التعليم جامعي College لأن الاختبار في كلتا الحالتين غير معنوي (معنوية الاختبار في كلتا الحالتين أكبر من مستوى الأهمية المفروض 0.05). بيد أن هناك فرقاً معنوياً بين متوسطات درجة الاهتمام السياسي بين الجنسين عند المستوى التعليمي فوق الجامعي. لاحظ أن الاختبار في هذه الحالة معنوي لأن معنوية الاختبار 0.002 أصغر من مستوى الأهمية المفروض 0.05.

Univariate Tests

Dependent Variable: Int_Politics

Edu_Level		Sum of Squares	df	Mean Square	F	Sig.
School	Contrast	9.800	1	9.800	.543	.465
	Error	975.200	54	18.059		
College	Contrast	1.250	1	1.250	.069	.793
	Error	975.200	54	18.059		
University	Contrast	186.050	1	186.050	10.302	.002
	Error	975.200	54	18.059		

5.7 التطبيق الخامس The Fifth Application

عند هذه النقطة من الكتاب يجب أن يكون القارئ قد امتلك القدرة الكافية للتعامل مع البيانات. بشكل خاص ، يجب أن يكون القارئ قادراً على القيام بالمهام التالية:

- إدخال البيانات بشكل صحيح إلى شاشة البيانات لبرنامج الـ SPSS.
 - والتعامل مع المتغيرات مثل التسمية و تحديد خصائص هذه المتغيرات.
 - القيام بالدراسة الوصفية للبيانات. بشكل خاص، وصف البيانات (1) على شكل مخططات بيانية و (2) من خلال إجراء الدراسة الاستكشافية للتعرف إلى خصائص البيانات و شكل التوزيع الذي تتبع له.
 - وضع الأسئلة المناسبة التي يمكن أن تجيب عنها البيانات و من ثم وضع الفرضيات المناسبة.
 - اختيار الإجراءات الإحصائية المناسبة لتحليل البيانات و الوصول إلى إحصائيات تمكن من القيام باستنتاجات حول البيانات.
 - كتابة تقارير علمية تلخص بشكل كافٍ النتائج التي تم التوصل إليها.
- (تذكر: يجب أن تكون التقارير العلمية موجزة، كافية، دقيقة، و بعيدة عن المصطلحات التخصصية.)

بناءً على ذلك، تم ترك هذا التطبيق تمريناً للقارئ. بيانات هذا التطبيق موجودة في الملحق D — المجموعة الثانية. تم أخذ هذه البيانات من دراسة علمية شاملة عن علاقة مرض القلب التاجي بمجموعة من المتغيرات.⁴ من المتغيرات المدروسة نذكر:

- عمر المريض Age.
- كمية التدخين Tobacco و تقاس بكمية النيكوتين المتراكمة في كل كيلو غرام من وزن الجسم.
- السمنة Adiposity.
- القصة المرضية. Famhist (أي هل لدى عائلة المريض أي حالة إصابة بمرض القلب التاجي).
- شحوم الدم منخفضة الكثافة Low Density Lipoprotein Cholesterol.

المطلوب:

- إدخال هذه البيانات إلى شاشة الـ SPSS، و دراستها بشكل وصفي لتحديد خصائصها.
- وضع كل الأسئلة التي يمكن الإجابة عنها من خلال البيانات.
- وضع الفرضيات المناسبة للأسئلة في الفقرة أعلاه.
- إجراء الاختبارات الإحصائية المناسبة.
- كتابة تقارير تلخص النتائج.

★★★★★★★★★★

⁴ البيانات مأخوذة من دراسة طبية في مجلة العلوم الطبية في جنوب إفريقيا. يمكن الوصول إلى البيانات من خلال العنوان التالي:

<http://www-stat.stanford.edu/~tibs/ElemStatLearn/datasets/SAheart.data>



SPSS Tutorial

الجزء الأول

تجهيز البيانات

Setting up the data file

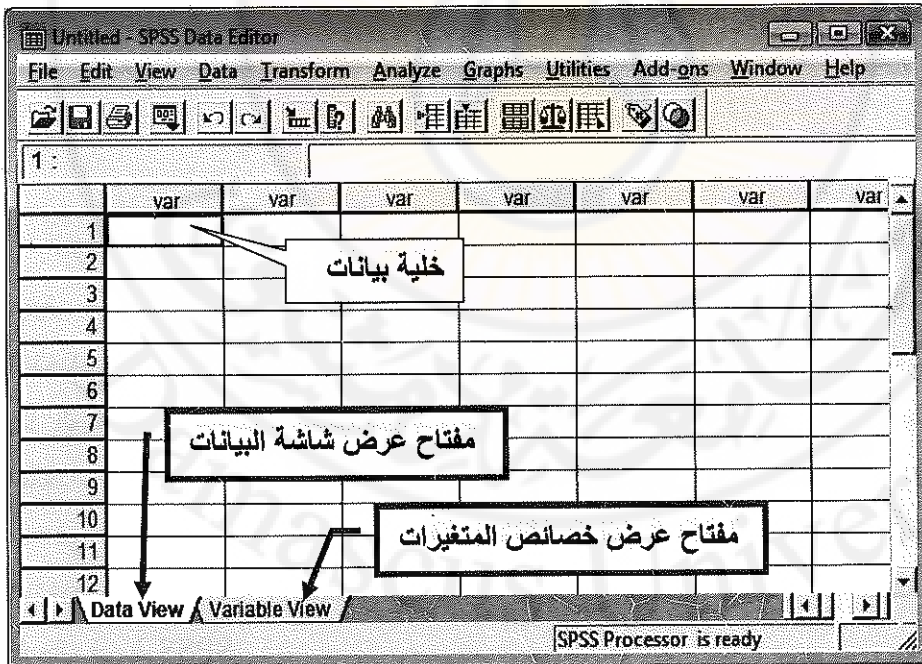
سوف نعرض في هذا الملحق معلومات مختصرة و لكن كافية لكي يستطيع القارئ التعرف إلى كيفية استخدام الـ SPSS. لا يهدف هذا الملحق أبداً إعطاء دليل متكامل لاستخدام الـ SPSS ، و لا إلى تغطية جميع تفاصيل هذا البرنامج. إنّ الهدف الرئيسي لهذا الملحق و الملاحق التالية هو تزويد الطالب بالمعلومات الضرورية لكي يبدأ باستخدام الـ SPSS باعتبارها أداة مساعدة لإجراء مختلف الإجراءات الإحصائية. في البداية و بفرض أن الطالب قام بتشغيل برنامج الـ SPSS. الخطوة الأولى لاستخدام هذا البرنامج الإحصائي هي كيف يستطيع الطالب التعامل مع شاشات الإدخال في الـ SPSS.

قبل البدء بإدخال هذه البيانات يجب أن نشغل برنامج الـ SPSS الذي يظهر كما في الشكل (1). كما يبدو من الشكل (1) فإن شاشة البيانات فارغة و هي جاهزة لاستقبال البيانات.

الشاشة الرئيسية في الـ SPSS تحتوي على المكونات التالية:

1. شريط القوائم مثل File, Edit, View, Data, Transform. تحتوي كل قائمة على مجموعة من الأوامر التي يمكن استخدامها لتنفيذ عمليات متعددة على البيانات. على سبيل المثال ، يمكن استخدام القائمة File لتخزين البيانات ، فتح ملف يحوي بيانات ، طباعة ملف ، إغلاق شاشة البيانات الفعالة حالياً ، إلخ.

2. شاشة إدخال البيانات و تتألف من أعمدة مسماة بأسماء افتراضية Var وأسطر مرقمة تصاعدياً من 1 إلى أعلى رقم يمكن لذاكرة الحاسوب استيعابه.

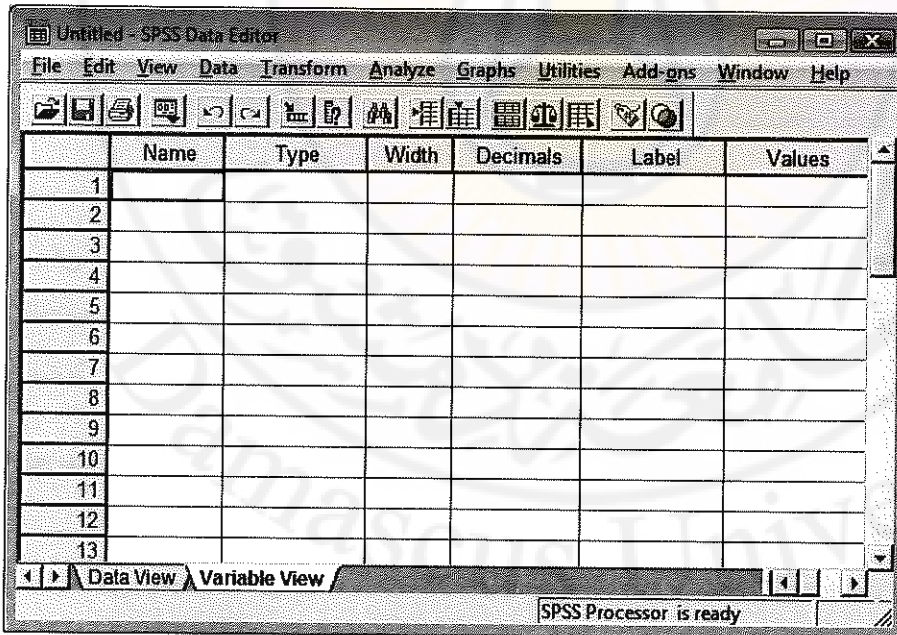


الشكل (1): شاشة إدخال البيانات في الـ SPSS

خلية البيانات يسمى تقاطع سطر مع عمود بخلية بيانات. خلية البيانات هي المكان الذي ندخل فيه البيانات.

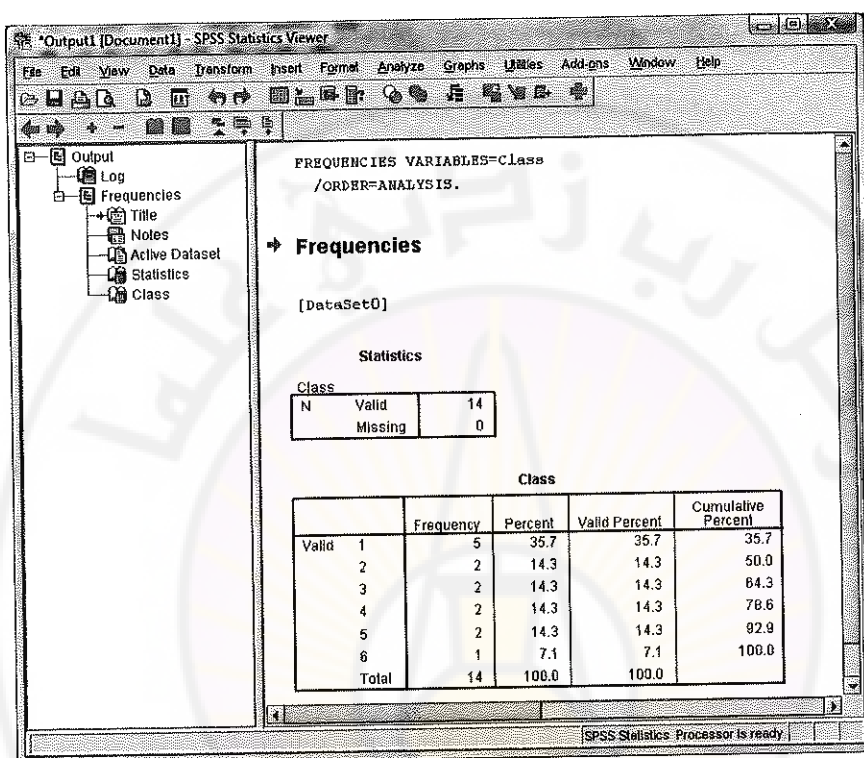
مفتاح عرض شاشة البيانات **Data View** يؤدي النقر على هذا المفتاح إلى فتح شاشة البيانات وهي الشاشة الرئيسية التي تظهر لحظة تشغيل البرنامج.

مفتاح عرض خصائص المتغيرات **Variable View**. يؤدي النقر على هذا المفتاح إلى إظهار شاشة خصائص المتغيرات. كما يبدو بالشكل (2). تحتوي هذه الشاشة على خيارات تحديد خصائص متغير. من هذه الخصائص نذكر: إعطاء اسم للمتغير Name ، تحديد نوع بيانات المتغير Type ، تحديد عرض المتغير (عدد المحارف) Width ، تحديد عدد المحارف التي تظهر بعد الفاصلة العشرية (خاصة بالبيانات الرقمية) Decimal ، كتابة نص توضيحي عن المتغير Label ، إعطاء ترميز لقيم متغير Values ، إلخ. كل سطر مخصص لتعريف متغير واحد. على سبيل المثال ، المتغير الأول يعرف في السطر الأول و الثاني يعرف في السطر الثاني و هكذا.



الشكل (2): تحديد نوع البيانات و بعض التوضيحات مثل نصوص التوضيح Labels.

شاشة الخرج Output Screen و فيها تظهر نتائج المعالجة من جداول و رسومات بيانية و غيرها. يظهر الشكل (3) شاشة الخرج.



الشكل (3): شاشة الخرج.

يمكن الانتقال بين الشاشات المختلفة عن طريق القائمة Window. عندما نفتح هذه القائمة تظهر لدينا كل الشاشات الفعالة و يمكن الانتقال إلى أي واحدة بالنقر على اسمها.

تعريف المتغيرات

لنفترض أننا قمنا بدراسة جدوى برنامج ما لتخفيض الوزن ، و ذلك بتجريبه على 12 متبرع. حيث تم في أثناء الدراسة إعطاء كل متبرع رقماً. بعد الانتهاء من برنامج تخفيض الوزن ، تم أخذ البيانات التالية عن كل مشارك في برنامج تخفيض الوزن.

1. الجنس sex. متغير اسمه Sex ، نوع البيانات عددي Numeric ، و له قيمتان ذكر 0 و أنثى 1 ، وسوف نعطيه اسم معبر Label: "Gender of the participant".

2. الطول height.

3. الوزن قبل الالتحاق ببرنامج تخفيض الوزن before.

4. الوزن بعد الالتحاق ببرنامج تخفيض الوزن after.

5. مكان الإقامة residency.

سوف نعرف خصائص متغير الجنس بشكل تفصيلي و نترك تعريف بقية المتغيرات على أنها تمرين. لنفرض أننا نريد تعريف متغير الجنس بالخصائص التالية:

اسم المتغير Sex ، نوع البيانات Numeric ، له قيمتان 0 يعني ذكر (Male) و 1 تعني أنثى (Female) و سوف نعطيه نصاً معبراً "Gender of Participant".

أولاً نفتح شاشة تعريف المتغيرات المبينة بالشكل أعلاه. سوف نستخدم السطر الأول لتعريف هذا المتغير.

1. ضع الاسم Sex في السطر الأول في العمود المعنون Name.

2. لتغيير نوع البيانات التي يمكن تخزينها في هذا العمود ، انقر بزر الفأرة الأيسر

على الأولى في العمود المعنون Type. يظهر نتيجةً لذلك مربع الحوار في

الشكل (3). فعل الخيار Numeric و ذلك بالنقر بزر الفأرة الأيسر على خانة

الاختيار المجاورة. بما أننا سنرمز الجنس بـ "0" و "1" ، لا نحتاج لأرقام بعد

الفاصلة العشرية لذلك ضع الرقم "0" في الخانة Decimal Places.

3. في الخلية الأولى المعنونة Label ضع نصاً يصف المتغير (اختياري).

سوف نضع النص Gender of the participant

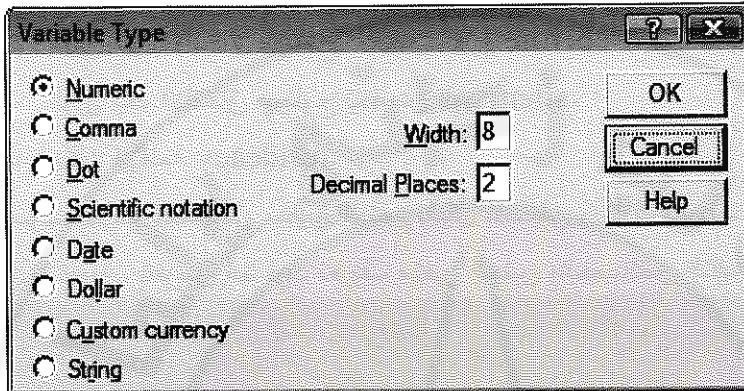
4. عادةً ما نعيد ترميز بعض القيم و نعطيها اسم معبر. مثلاً بدلاً من ظهور 0

في النتائج لتعبر عن الذكر ، نعيد ترميزها إلى Male. للقيام بذلك ننقر

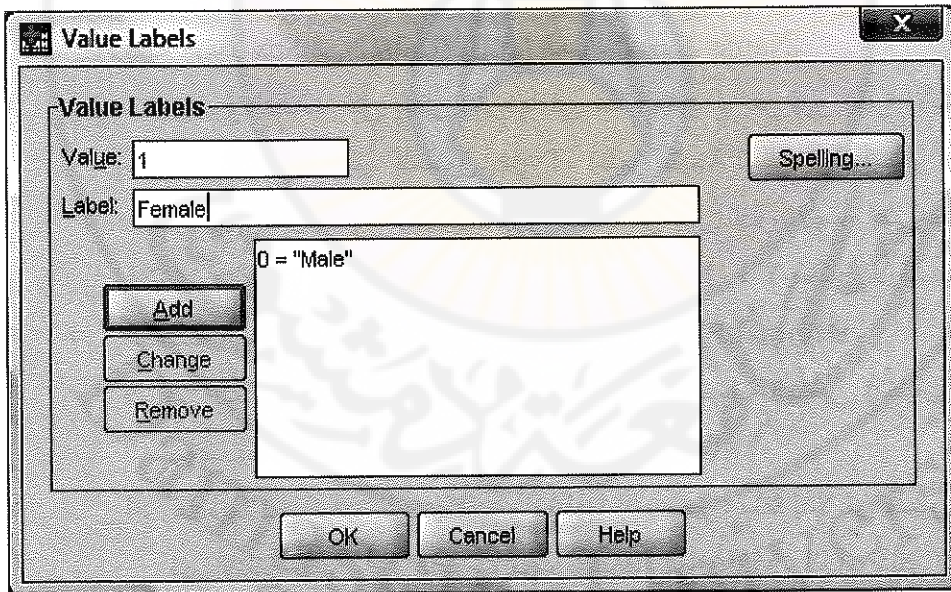
على الخلية الأولى في العمود Value فيفتح مربع حوار ، الشكل (4) ،

يسمح لنا بإعادة ترميز القيم 0 و 1 بتراميز أكثر وضوحاً مثل Male و

Female. في مربع الحوار ضع القيمة 0 في خانة النص المجاورة و
ترميزها Male في الخانة Value Label ثم اضغط على المفتاح Add.
كرر من أجل القيمة 1.

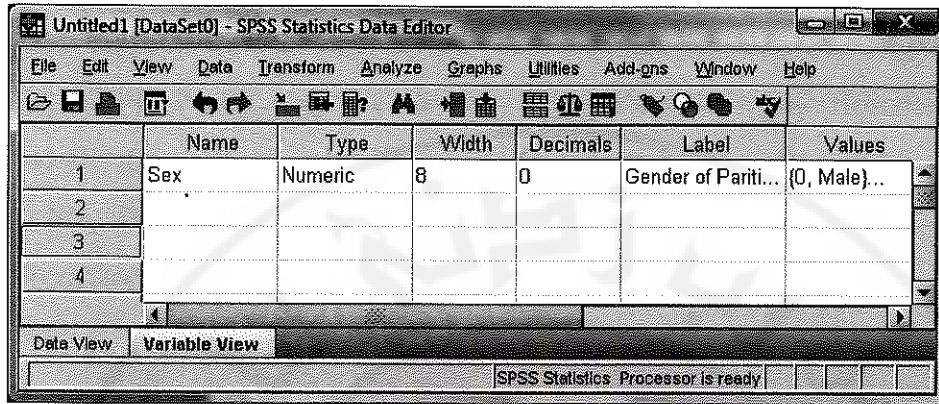


الشكل (3): أنواع البيانات المختلفة التي يمكن تحديدها بواسطة مربع الخيار.

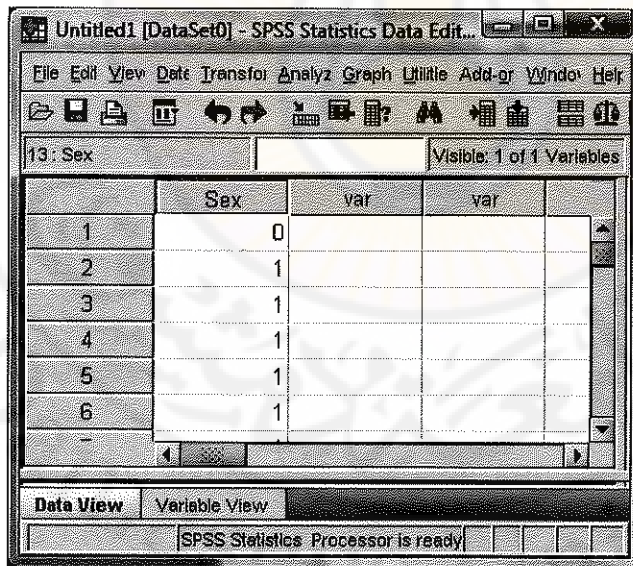


الشكل (4): تحديد النصوص التوضيحية.

بعد هذه الخطوات يجب أن يظهر الشكل (2) كما يلي:



الآن يمكننا إدخال جنس المشاركين في برنامج إنفاص الوزن مستخدمين الأرقام رموزاً للذكور و الإناث. سوف نستخدم القيم التالية: 0,1,1,1,1,1,0,0,0,1,0. يجب أن يظهر الشكل (1) بعد إدخال البيانات في **sex** كما في الشكل التالي:



بعد هذا الشرح حول كيفية تهيئة عمود معين و وضع قيم فيه ، نترك بقية الأعمدة للقارئ تمريناً.

حفظ ملفات الـ SPSS

يجب حفظ البيانات بعد إدخالها ، إذا أردنا استخدامها في المستقبل. كذلك يمكن حفظ المخرجات في شاشة الخرج لكي نوفر نسخة نعود إليها عند الحاجة. حفظ البيانات في الـ SPSS لا يختلف عن حفظ ملفات في البرامج الأخرى مثل برامج Microsoft Office (محرر النصوص Word ، برامج الإكسل Excel ، و غيرها). لحفظ ملف ما ، نفتح القائمة ملف (أو File) و من ثم نختار الأمر Save. يظهر مربع حوار مشابهاً لمربع الحوار في الشكل التالي:

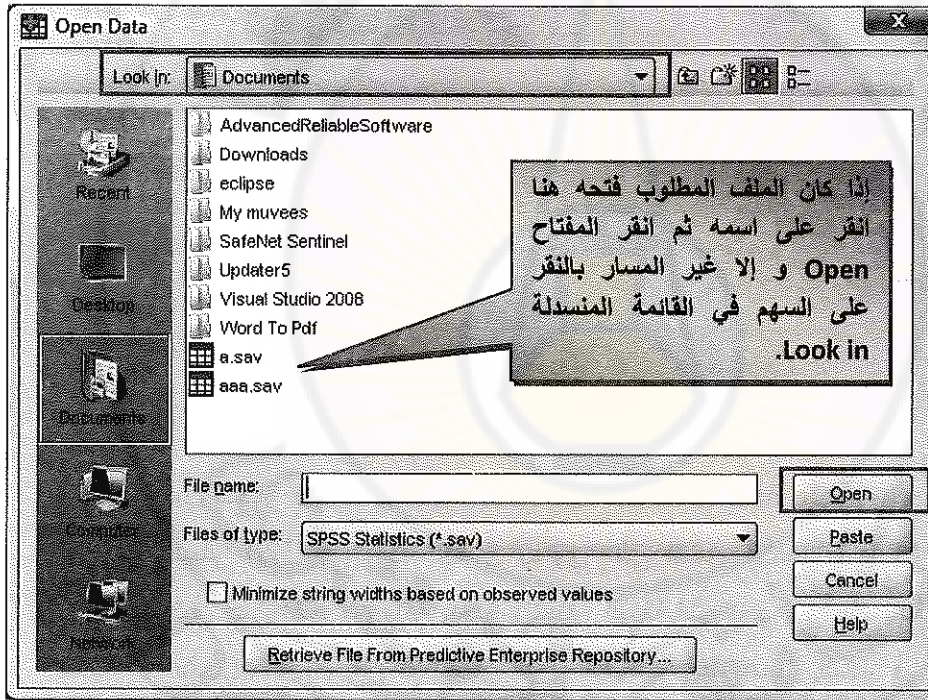


يمكننا تغيير مكان التخزين إلى محرك أقراص معين أو أي مجلد عن طريق الضغط على الخانة Look in و المشار إليها بالسهم (تغيير مكان الحفظ). بعد اختيار مكان التخزين ، إذا كان ذلك ضرورياً ، نكتب اسماً للملف في الخانة File name و المشار إليها بالسهم (اكتب اسم الملف هنا). يمكن بالطبع تحديد نوع ملف التخزين من خلال الخانة Save as type. الوضع الافتراضي هو تخزين الملف كشاشة بيانات

SPSS ، ولكن يمكن تغيير هذا الوضع للتخزين على نوع ملف آخر مثل الإكسل (Excel). بعد ذلك نضغط المفتاح Save ليتم تخزين الملف في المكان المحدد و بالنوع المحدد و تحت الاسم المحدد.

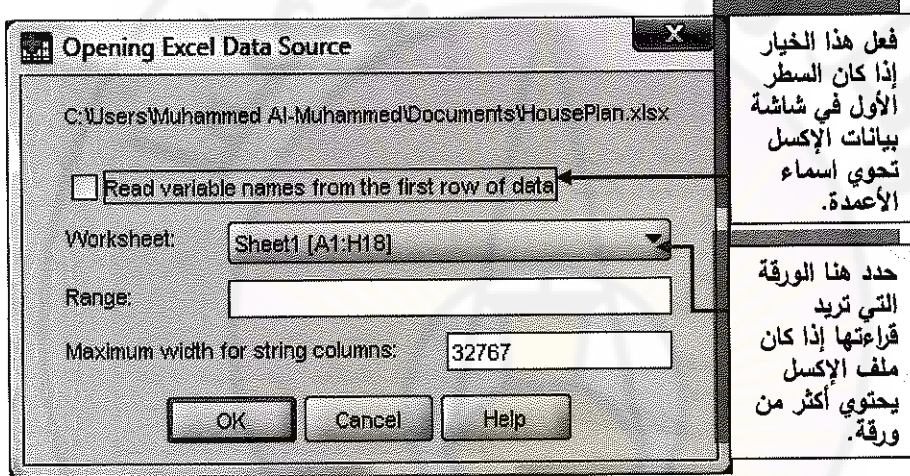
فتح ملف Open a File

يمكن فتح أي ملف SPSS مخزن سابقاً ، وذلك باستخدام القائمة ملف أو File كمايلي: نفتح القائمة ملف (أو File) ، نختار الأمر Open و من ثم الأمر Data من القائمة الفرعية لكي يفتح مربع حوار مشابهاً للمربع التالي:



إذا كان اسم الملف الذي نريد فتحه ظاهراً في المسار الحالي ، فإننا ننقر على اسم الملف و ننقر المفتاح Open. أما إذا كان الملف مخزناً في مكان آخر ، فيجب الانتقال إلى المسار الذي يحوي الملف عن طريق الخانة Look in ، و من ثم نفتح الملف كما سبق.

يمكن فتح ملفات مخزنة بواسطة برامج أخرى مثل Excel. طريقة فتح الملف مماثلة لفتح ملف SPSS مع إجراء إضافي يتمثل بتحديد عدد الأعمدة المراد قراءتها ، و كذلك تحديد فيما إذا كانت أسماء الأعمدة مخزنة في الملف المصدر. مثلاً لفتح ملف Excel ، نتبع الإجراءات نفسها التي قمنا بها أعلاه لفتح ملف SPSS. لكن قبل أن يفتح الـ SPSS الملف فإن البرنامج يظهر مربع الحوار التالي الذي يمكننا من تحديد فيما إذا كانت أسماء الأعمدة موجودة في السطر الأول من الورقة و تحديد الورقة Sheet المطلوبة إذا احتوى الملف أكثر من ورقة واحدة.



نفعال الخيار Read variable names from the first row of data

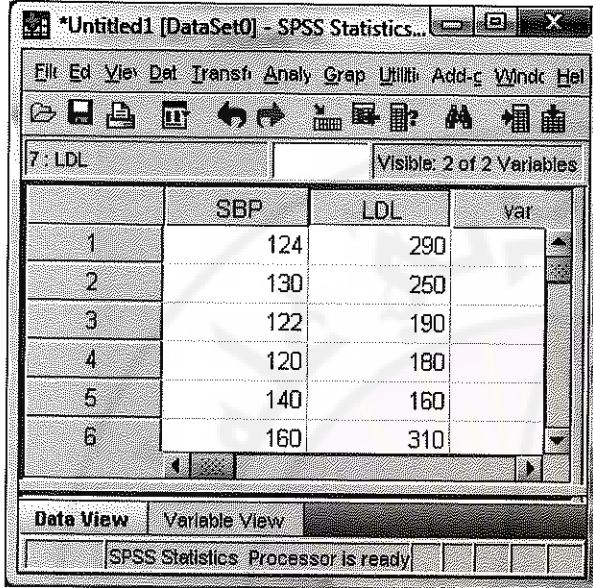
إذا كان السطر الأول في ورقة الإكسل يحتوي على أسماء الأعمدة. إذا كان ملف الإكسل يحتوي على أكثر من ورقة واحدة ، نختار الورقة التي نريد ، و ذلك بالنقر على القائمة Worksheet ، و من ثم اختيار الورقة المناسبة. على سبيل المثال في المربع أعلاه ، تم اختيار الورقة Sheet 1 من ملف الإكسل.

تنفيذ بعض الإجراءات الإحصائية البسيطة

في هذه الفقرة سوف ننفذ بعض الإجراءات الإحصائية البسيطة التي سوف نقيّد باعتبارها مدخلاً إلى البرنامج الإحصائي SPSS.

حساب الارتباط بين متغيرين

سوف نقوم في هذه الفقرة بإجراء اختبار الارتباط الخطي البسيط على البيانات في



The screenshot shows the SPSS Statistics interface. The title bar reads '*Untitled1 [DataSet0] - SPSS Statistics...'. The menu bar includes File, Edit, View, Data, Transform, Analyze, Graph, Utilities, Add-ons, Window, and Help. The toolbar contains various icons for file operations and data manipulation. The main window displays a data set with two variables, SBP and LDL, and a third variable 'var'. The data is organized into a table with 6 rows and 4 columns. The first column is labeled 'var' and contains values 1 through 6. The second column is labeled 'SBP' and contains values 124, 130, 122, 120, 140, and 160. The third column is labeled 'LDL' and contains values 290, 250, 190, 180, 160, and 310. The status bar at the bottom indicates 'SPSS Statistics Processor is ready'.

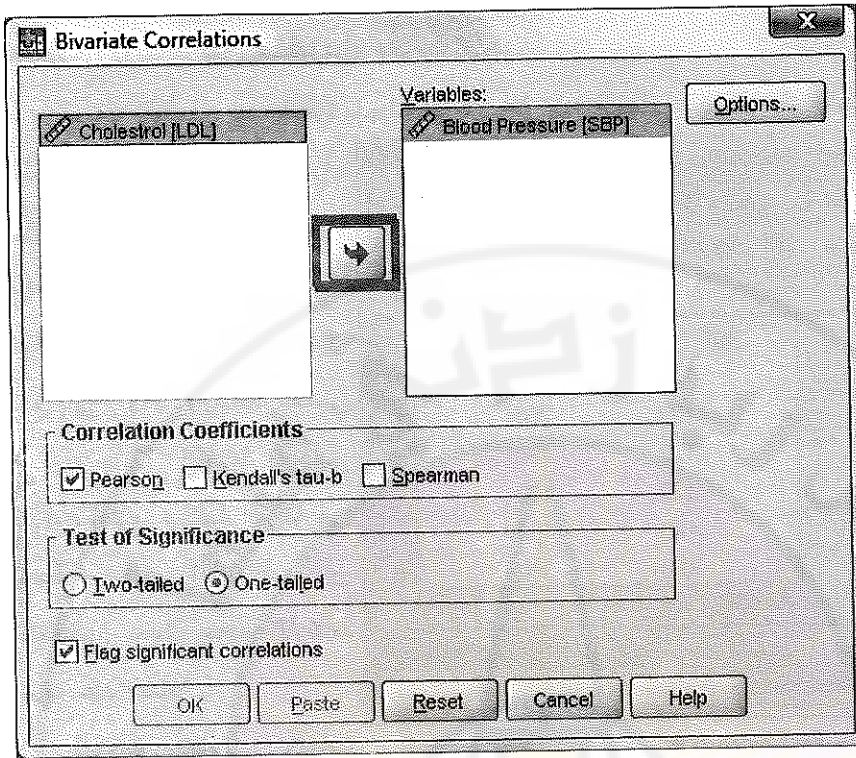
	SBP	LDL	var
1	124	290	
2	130	250	
3	122	190	
4	120	180	
5	140	160	
6	160	310	

الشكل الجانبي. يقوم إجراء بيرسون Pearson بحساب معامل الارتباط الخطي بين متغيرين. قيمة معامل بيرسون تقع بين "1-" (ارتباط عكسي تام) و "1" (ارتباط إيجابي تام). بالطبع ، كلما اقتربنا من الصفر قل الارتباط بين المتغيرين بحيث ينعقد الارتباط تماماً عند الصفر.

لنفترض أننا نريد حساب

قيمة معامل بيرسون للارتباط الخطي بين متغير ضغط الدم SBP و كمية الكوليسترول LDL. لاحظ أن بيانات هذين المتغيرين هي بيانات كمية. لذلك فإن حساب ارتباط بيرسون بينهما له معنى. لحساب معامل ارتباط بيرسون سوف نقوم بالخطوات التالية.

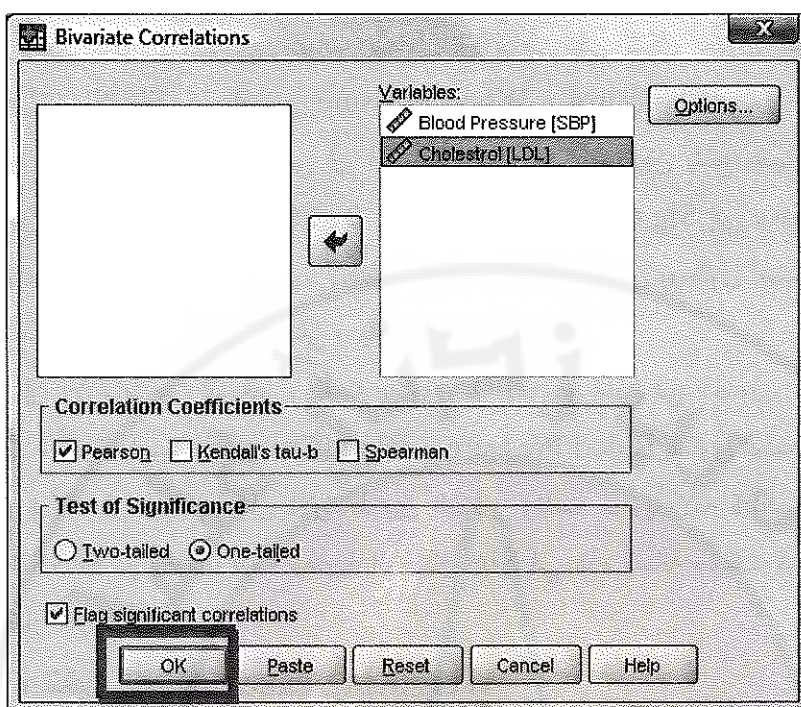
1. افتح القائمة Analyze ، ثم الأمر Correlate ، و من ثم الأمر الفرعي Bivariate. يفتح مربع الحوار في الشكل (6).
2. حدد المتغير "SBS" و ذلك بالنقر عليه الفأرة الأيسر.
3. انقل المتغير المحدد إلى المستطيل المجاور المعنون بـ "Variables" و ذلك بالنقر على السهم بزر الفأرة الأيسر. كرر العملية نفسها لنقل المتغير الآخر "LDL" إلى المستطيل المعنون بـ "Variables". لا توجد حاجة لتفعيل إجراء ارتباط بيرسون لأنه مفعّل افتراضياً في الشكل (6) (لاحظ علامة التفعيل بجانب الأمر).



الشكل (6): مربع حوار إجراء الارتباط

ملاحظة: يمكن نقل متغيرين (أو أكثر) دفعة واحدة إلى المستطيل "Variables" وذلك بتحديد العمودين المراد نقلهما دفعة واحدة ، و النقر على سهم النقل. يتم تحديد عمودين متجاورين أو أكثر من خلال النقر على العمود الأول و من ثم الضغط بشكل مستمر على المفتاح Shift (من لوحة المفاتيح) ثم النقر على آخر عمود. أما إذا كانت الأعمدة غير متجاورة فإن التحديد يتم بالنقر على العمود الأول و من ثم نضغط بشكل مستمر على المفتاح Ctrl (من لوحة المفاتيح) و ننقر على بقية الأعمدة واحداً تلو الآخر حتى يتم تحديد جميع الأعمدة المطلوب نقلها.

بعد ذلك يجب أن يظهر مربع الحوار في الشكل (6) كما يلي:



اضغط المفتاح OK لتنفيذ الأمر.

تظهر نتائج حساب معامل الارتباط في الشكل (7). بشكل خاص نلفت نظر القارئ إلى القيم المظللة في الشكل (7).

Correlations		Blood Pressure	Cholestrol
Blood Pressure	Pearson Correlation	1	.456
	Sig. (1-tailed)		.182
	N	6	6
Cholestrol	Pearson Correlation	.456	1
	Sig. (1-tailed)	.182	
	N	6	6

الشكل (7): نتائج اختبار الارتباط

ينظم SPSS نتائج حساب معامل الارتباط على شكل جدول تتقاطع فيه المتغيرات. الارتباط بين المتغير Blood Pressure و Cholestrol يساوي 0.456 و هو ارتباط موجب. هذا يعني أن أي زيادة في أحد المتغيرين ينتج عنها زيادة في المتغير الآخر. يشير الرقم 0.182 في السطر Sig (1-tailed) إلى معنوية معامل الارتباط بين المتغيرين المقابلين. لاحظ أن الارتباط غير معنوي بمستوى أهمية 0.05 و بالتالي فإن الارتباط غير حقيقي و يمكن أن يساوي الصفر.

إن الارتباط بين المتغير و نفسه يساوي "1" كما هو مبين في الجدول. على سبيل المثال ، الارتباط بين Blood Pressure و Blood Pressure يساوي الواحد. لا توجد ضرورة لاختبار فيما إذا كان الارتباط بين المتغير ونفسه معنوياً لأنه كذلك. و بالتالي فإن SPSS لم يحسب قيمة معنوية الاختبار Sig في هذه الحالة وتركت قيمة معنوية الاختبار فارغة.

إجراء تثقيف البيانات Weighting cases

هذا الإجراء مفيد للغاية و خصوصاً عندما تكون لدينا معطيات تكرارية. على سبيل المثال ، قبل إجراء اختبارات الاستقلالية أو اتفاق المراقبين فإنه يجب أن تدخل البيانات إلى شاشة البيانات على شكل تكرار لعناصر كل ظاهرة. على سبيل المثال ، إذا كانت لدينا القيمة "1" مكرر 1000 مرة فإن هذه القيمة تخزن في 1000 خلية متتالية في العمود المحدد. و لكن هذه العملية مكلفة من الناحية الزمنية و تزيد فرص الإدخالات الخاطئة أيضاً.

من الأفضل أن نضع ، على سبيل المثال ، القيمة "1" في خلية معينة و من ثم نربط هذه القيمة برقم يدل على عدد تكراراتها. هذا بالضبط ما يقوم به إجراء تثقيف البيانات. على سبيل المثال ، لنفرض أن لدينا البيانات التكرارية التالية:

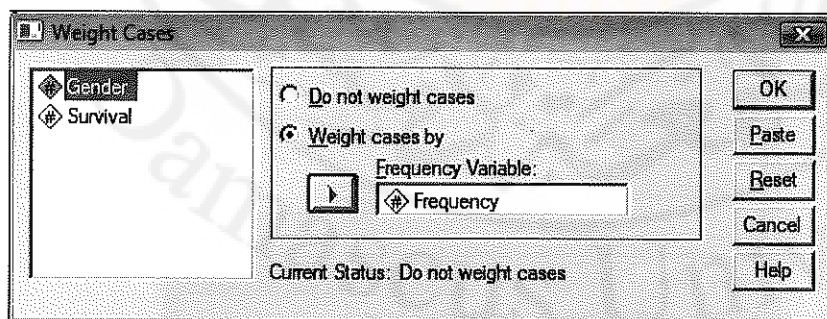
Survival		
Gender	1	2
	680	400
2	1000	120

Untitled - SPSS Data Editor			
File Edit View Data Transform Analyze			
6 :			
	Gender	Survival	Frequency
1	1	1	680
2	1	2	400
3	2	1	1000
4	2	2	120
5			
6			

الشكل (8): تفريغ البيانات الجدول أعلاه إلى شاشة الـ SPSS

يظهر الشكل (8) البيانات في الجدول أعلاه بعد تفريغها في شاشة إدخال البيانات. العمودان الأول والثاني من اليسار (Survival و Gender) يحتويان على القيم المكررة. العمود Frequency يشتمل على تكرارات كل زوج من قيم العمودين Gender و Survival. على سبيل المثال ، الزوج (2,1) في السطر الثالث مكرر 1000 مرة. حتى يعتبر الـ SPSS أن قيم المتغير Frequency هي تكرارات أزواج القيم المقابلة في العمودين Survival و Gender ، فإننا نستخدم الإجراء Weight Cases.

1. تفعيل إجراء تثقيب البيانات من خلال القائمة Data ، و من ثم اختيار الأمر الرئيسي Weight Cases... يظهر نتيجة لهذا التفعيل مربع الحوار في الشكل (9).



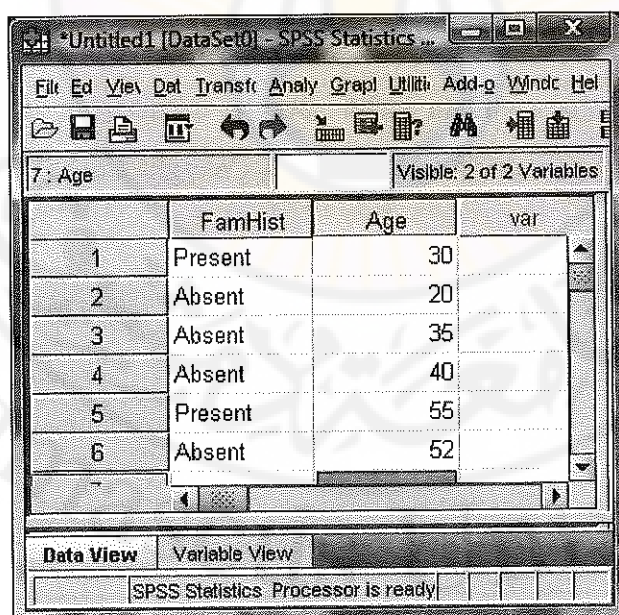
الشكل (9): مربع الحوار الخاص بإجراء تثقيب البيانات.

2. نلغي تفعيل الوضع الافتراضي "Do not weight cases" و نفعل بدلاً من ذلك الخيار "Weight cases by". نختار العمود Frequency و ننقله إلى الخانة المعنونة "Frequency Variable" كما يظهر في الشكل (9).

3. نضغط على المفتاح OK فيقوم الـ SPSS بتنقيح البيانات بالأوزان المحددة في العمود Frequency.

إجراء تشفير (إعادة ترميز) البيانات Data Recoding

يفيد هذا الإجراء في إعادة ترميز مجموعة من البيانات ، و ذلك لتحويلها إلى شكل مناسب لاختبار معين. على سبيل المثال ، في كثير من المسائل التطبيقية تكون الإجابات على شكل شريطي ، مثلاً Yes و No. الاختبارات الإحصائية تفترض أن تكون البيانات رقمية و بالتالي لابد من إعادة التشفير. لنفرض أن لدينا البيانات التالية المعروضة في شاشة الـ SPSS.



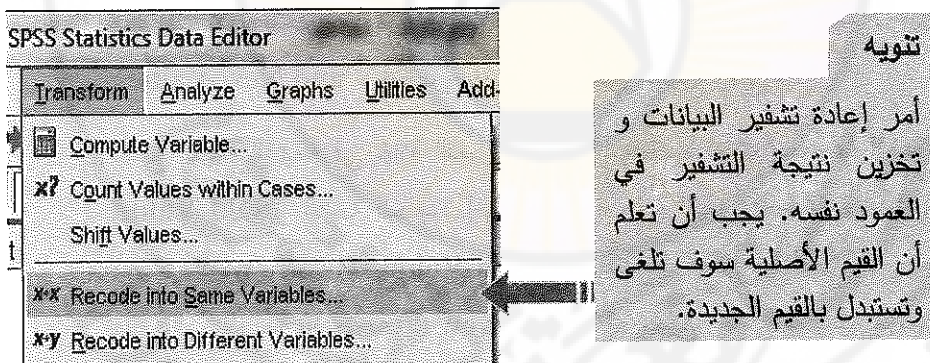
	FamHist	Age	var
1	Present	30	
2	Absent	20	
3	Absent	35	
4	Absent	40	
5	Present	55	
6	Absent	52	

و لنفترض أن البيانات في العمود Age هي كمية و يراد إعادة ترميزها بحيث تصبح فئوية (مجالات العمر بدلاً من القيم الفعلية). و لنفترض أيضاً أن القيم في العمود FamHist هي بيانات شريطية و نريد إعادة ترميزها إلى بيانات رقمية.

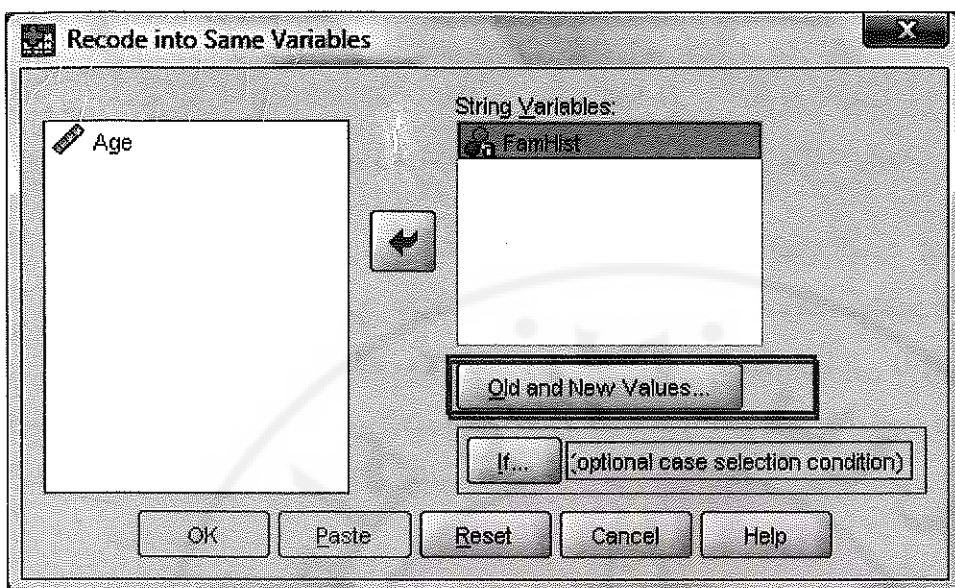
يوفر الـ SPSS إمكانية تشفير البيانات ووضعها في نفس العمود (تحل محل العمود الأصلي) أو في عمود مختلف يعرفه المستخدم. سوف نستخدم الطريقتين لتوضيح هذا الإجراء. لذلك سوف نعيد ترميز قيم المتغير FamHist إلى 0 (Absent) و 1 (Present) ونخزن النتيجة في العمود نفسه ، و نعيد ترميز المتغير Age إلى فئتين 0 (من 1 إلى 30 سنة) و 1 (من 31 إلى 60 سنة) و نخزن العمود الناتج في عمود جديد.

أولاً إعادة ترميز العمود FamHist

يتم تفعيل أمر إعادة الترميز بحيث تخزن النتيجة في العمود نفسه كما هو موضح بالشكل التالي:



بالنتيجة يظهر مربع الحوار الخاص بإعادة تشفير القيم و تخزينها بنفس العمود (الشكل (10)).



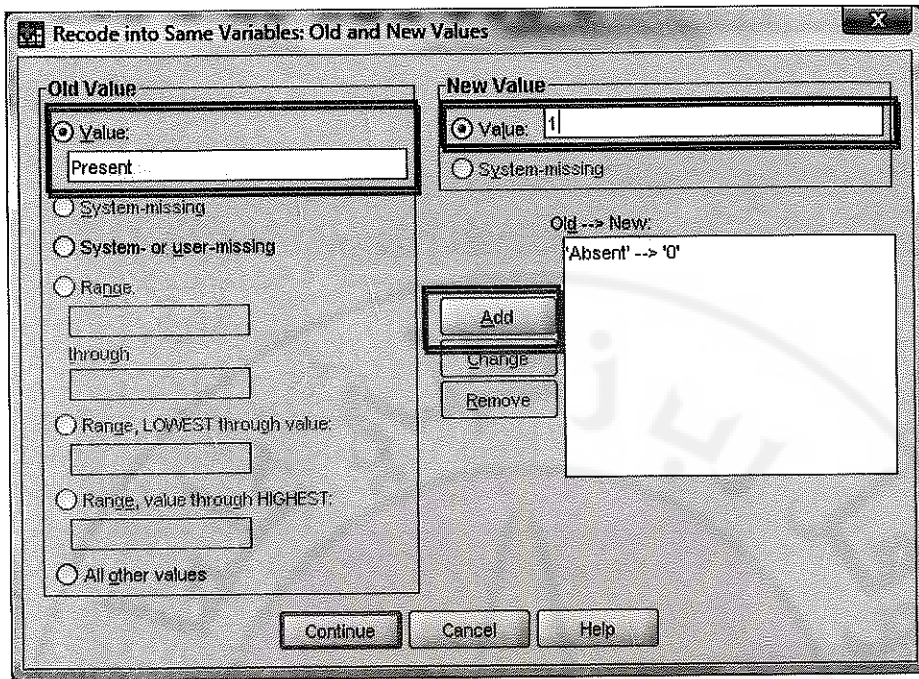
الشكل (10): تشفير البيانات و التخزين في عمود نفسه.

✖ نقوم بنقل العمود الذي نريد تشفيره من القائمة اليسارية إلى القائمة اليمينية. في الشكل (10) ، تم نقل المتغير FamHist إلى المكان المحدد.

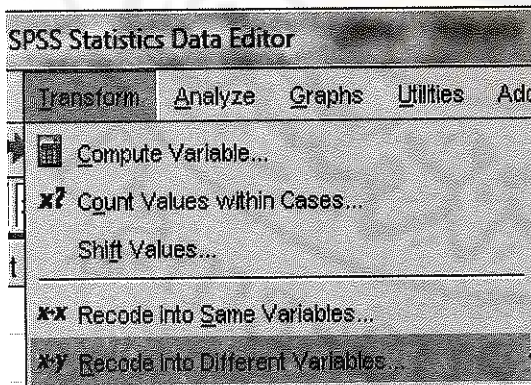
✖ بعد ذلك نقوم بتحديد الترميز للقيم القديمة و ذلك بالضغط على المفتاح " Old and New Values " ، فيظهر مربع الحوار في الشكل (11).

✖ نضع القيمة القديمة في خانة النص المجاورة لـ Value تحت مجموعة الخيارات Old Value و القيمة الجديدة (الترميز) في خانة النص المجاورة لـ Value تحت مجموعة الخيارات New Value. ثم نضغط المفتاح Add ليتم إضافة ذلك إلى القائمة المعنونة بـ Old → New. جميع هذه العمليات موضحة بالشكل (11).

✖ نضغط الآن Continue في مربع الحوار في الشكل (11) لكي نعود إلى الشكل (10) و فيه نضغط OK ليتم تشفير القيم.

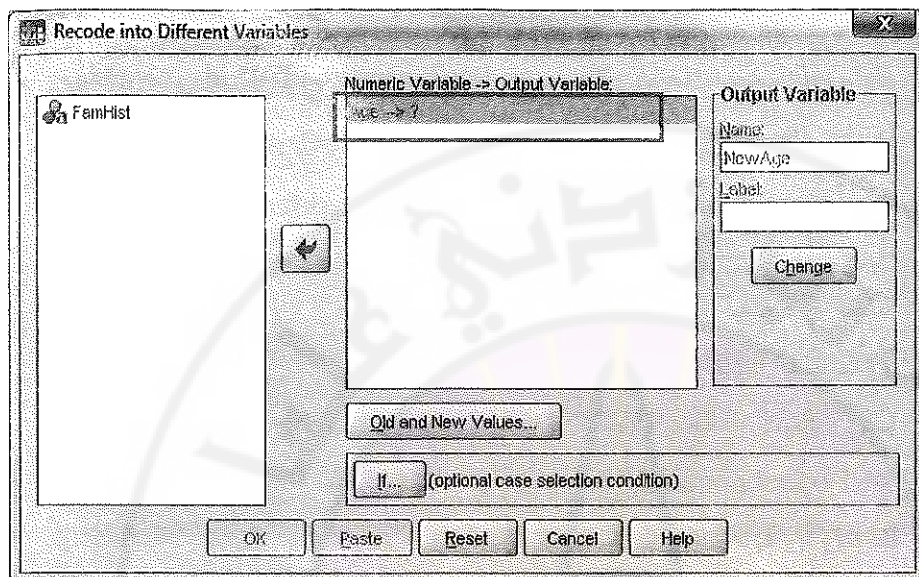


الشكل (11): تحديد القيم القديمة و القيم الجديدة المقابلة.
 ثانياً إعادة ترميز العمود Age و تخزين النتيجة في عمود جديد
 يتم تفعيل أمر إعادة الترميز بحيث تخزن النتيجة في عمود مختلف كما هو موضح
 بالشكل التالي:



تنويه
 أمر إعادة تشفير البيانات و
 تخزين نتيجة التشفير في
 عمود آخر. يجب أن تعلم
 أن القيم الأصلية سوف تبقى
 و تخزن نتيجة التشفير في
 عمود آخر.

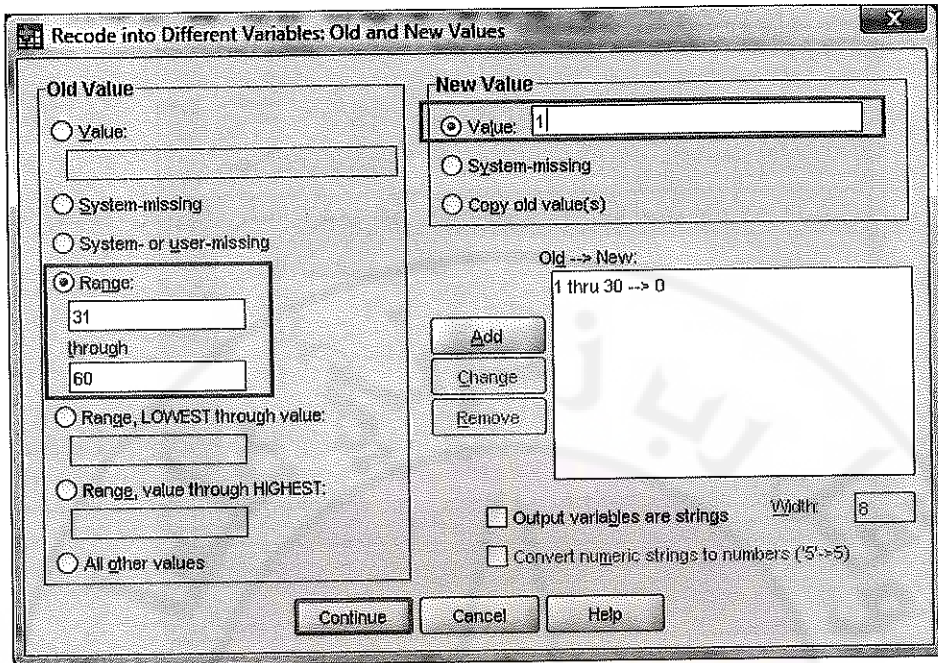
بالنتيجة يظهر مربع الحوار الخاص بإعادة تشفير القيم و تخزينها في عمود جديد (الشكل (12)).



الشكل (12): تشفير البيانات و التخزين في عمود جديد.

✧ بعد نقل العمود الذي نريد إعادة ترميز قيمه إلى القائمة المعنونة بـ Numeric Variable → Output Variable ، فإنه يجب تحديد اسم للعمود الذي ستخزن فيه نتيجة التشفير. يشير الـ SPSS إلى ذلك بـ "?" كما تلاحظ في الشكل. يمكن تعريف اسم للمتغير في الخانة Name تحت المجموعة المعنونة بـ Output Variable كما هو موضح بالشكل (12). فقط اكتب الاسم في خانة النص Name كما فعلنا في الشكل (12) ، حيث كتبنا الاسم "NewAge" ، و من ثم نضغط المفتاح Change ليحل هذا الاسم محل إشارة الاستفهام.

✧ نضغط على المفتاح "Old and New Values" للقيام بتحديد القيم القديمة و الشيفرة المقابلة لها. يظهر مربع الحوار الخاص بهذا الأمر في الشكل (13).



الشكل (13): تحديد القيم القديمة و الجديدة المقابلة لها.

✳ بما أننا نريد إعادة ترميز مجالات بقيم فإننا نستخدم خيار المجال Range حيث نحدد القيمة الدنيا و العليا للمجال كما هو موضح بالشكل. نحدد القيمة الجديدة في خانة النص Value الواقعة تحت New Value. بعدها نضغط المفتاح Add لكي يتم أخذ الخيار. في مثالنا رمزنا المجال من 1 إلى 30 بـ "0" و بالتالي ظهرت القيمة "0 → 1 thru 30" في مربع الحوار. نكرر العملية نفسها لتشفير المجال الثاني.

✳ نضغط المفتاح Continue في الشكل (13) و من ثم نضغط المفتاح OK في مربع الحوار في الشكل (12) ليتم التشفير.

يوضح الشكل (14) شاشة البيانات بعد تشفير قيم العمودين Age و FamHist.

Untitled4 [DataSet3] - SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graph Utilities Add-on Window Help

1: NewAge 0.0 Visible: 3 of 3 Variables

	FamHist	Age	NewAge
1	1	30	0
2	0	20	0
3	0	35	1
4	0	40	1
5	1	55	1
6	0	52	1
7			

Data View Variable View

SPSS Statistics Processor is ready

الشكل(14): شاشة البيانات بعد تشفير قيم العמוד Age و FamHist.

★ ★ ★ ★ ★ ★ ★ ★ ★ ★ ★ ★ ★ ★ ★ ★

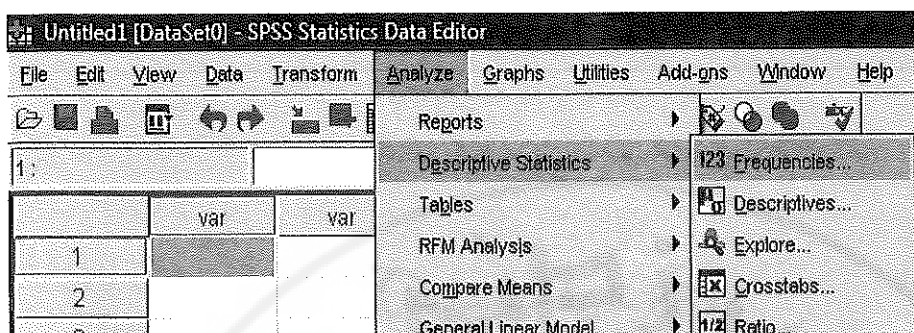
مهارات أساسية في الـ SPSS SPSS Basic Skills Tutorial: Basic Analyses

القائمة Analyze هي القائمة الرئيسية في الـ SPSS التي تستخدم لإجراء جميع التحليلات الإحصائية. كثرة المواضيع تحد من قدرتنا في تغطية كاملة لكل الموضوعات الإحصائية. لذلك سوف أحاول في هذا الملحق تغطية معظم المواضيع الهامة التي تضع القارئ في نقطة مناسبة يمكنه أن ينطلق من خلالها لدراسة أعمق لمواضيع الـ SPSS. سوف نبدأ باستخدام أوامر القائمة الإحصائية Analyze بادئين بالأوامر البسيطة.

التركرارات Frequencies

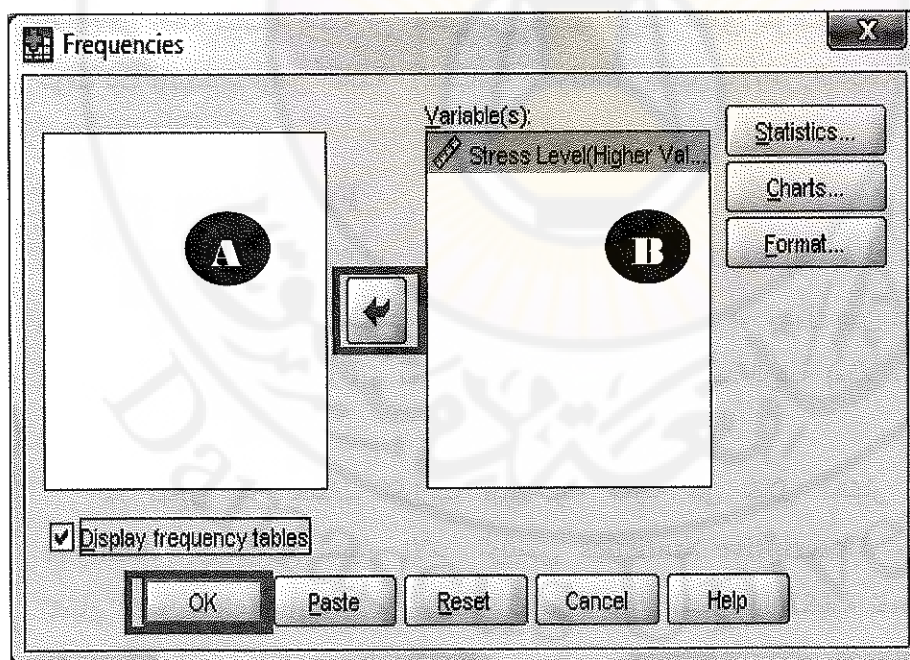
يستخدم أمر حساب التكرارات بشكل خاص مع البيانات الكيفية (اسمية أو ترتيبية) و يقوم بحساب تكرارات كل مشاهدة في العينة. نتيجة هذا الأمر هو جدول يبين كل قيمة في العينة ، و عدد تكراراتها ، و مقاييس أخرى يمكن تفعيلها عند الرغبة (مثلاً ، النسب المئوية).

يتم تفعيل أمر حساب تكرارات قيم عينة معينة من خلال اختيار القائمة Analyze ، و من ثم اختيار الأمر الرئيسي Descriptive Statistics. و أخيراً اختيار الأمر Frequencies من القائمة الفرعية. يبين الشكل (1) كيفية تفعيل هذا الأمر.



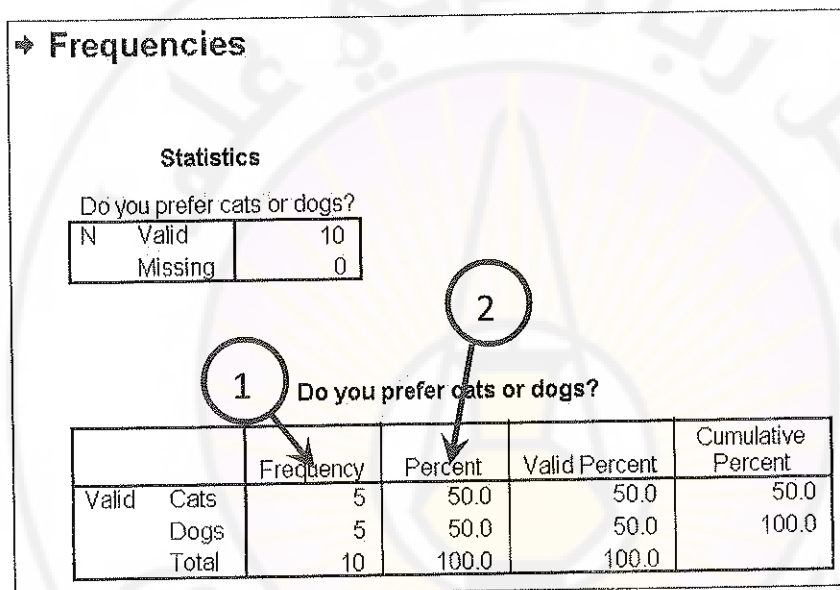
شكل(1): اختبار الأمر Frequencies

يظهر الـ SPSS مربع الحوار الخاص بهذا الأمر الذي يمكننا من تحديد العمود (أو الأعمدة) التي نريد حساب تكرارات كل قيمة فيها (الشكل (2)).



الشكل(2): مربع الحوار الخاص بالأمر Frequencies.

يجب أولاً نقل العمود أو الأعمدة التي نريد حساب تكرارات قيمها من القائمة التيسارية. استشار إليها بـ **A** إلى القائمة اليمينية المشار إليها بـ **B** و ذلك بتحديد العمود أو الأعمدة و من ثم النقر السهم  بالضغط على المفتاح OK يقوم البرنامج الإحصائي بحساب التكرارات و عرض النتيجة في شاشة الخرج (Output Screen).
يُظهر الشكل (3) شاشة الخرج و فيها جدول التكرارات للعمود المحدد. نلفت انتباه القارئ إلى القيم المشار إليها بالأسهم في شاشة الخرج في الشكل (3).



الشكل(3): شاشة الخرج و فيها جدول التكرارات.

① يبين العمود **Frequency** عدد تكرارات كل قيمة من القيم. مثلاً عدد تكرارات Cat هو "5" تكرارات.

② يبين العمود **Percent** النسبة المئوية للتكرار المقابل من التكرارات الكلية لقياسات العمود.

إذا كان المتغير الذي نريد حساب تكرارات قيمه المختلفة يملك قيمةً ترتيبيةً ، فإن SPSS يوفر مجموعة أخرى من الأوامر التي تساعد على إعطاء معلومات أكثر عن

العمود بجانب التكرارات. من المعلومات الإضافية التي يمكن الحصول عليها نذكر مايلي:

✓ مقاييس النزعة المركزية Central Tendency: المتوسط ، الوسط ، المنوال، المجموع.

✓ مقاييس التشتت Dispersion: التشتت ، الانحراف المعياري ، المدى ، القيمة الصغرى ، القيمة العظمى.

✓ الربيعيات: الربيعيات الأول ، الثاني ، و الثالث.

✓ المئينيات percentiles.

للحصول على معلومات إضافية (إذا كان ذلك مناسباً) ، يمكن الضغط على المفتاح Statistics في مربع الحوار المعروض في الشكل (2). نتيجة لذلك يظهر مربع الحوار في الشكل (4).

فيما يلي شرح لبعض الخيارات المهمة في الشكل (4).

✓ الربيعيات Quartiles: تحسب قيم ربيعيات العينة.

✓ لمئينيات Percentiles: تمكن من إيجاد قيمة أي عدد من المئينات يحددها المستخدم في خانة الحوار المقابلة.

✓ المتوسط Mean: يحسب متوسط العينة.

✓ الوسط Median: يحسب وسط العينة.

✓ المنوال Mode: يحسب منوال العينة.

✓ المجموع Sum: يحسب مجموع قياسات العينة.

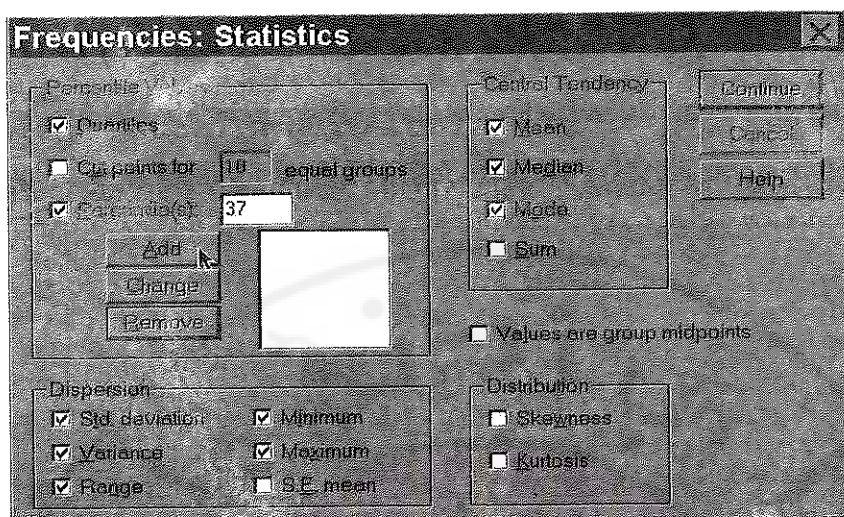
✓ الالتواء Skewness: يبين هل التوزيع منتظم أم ملتوٍ بأحد الاتجاهين.

✓ التفرطح Kurtosis: يبين هل قمة التوزيع مدببة أم مفرطحة.



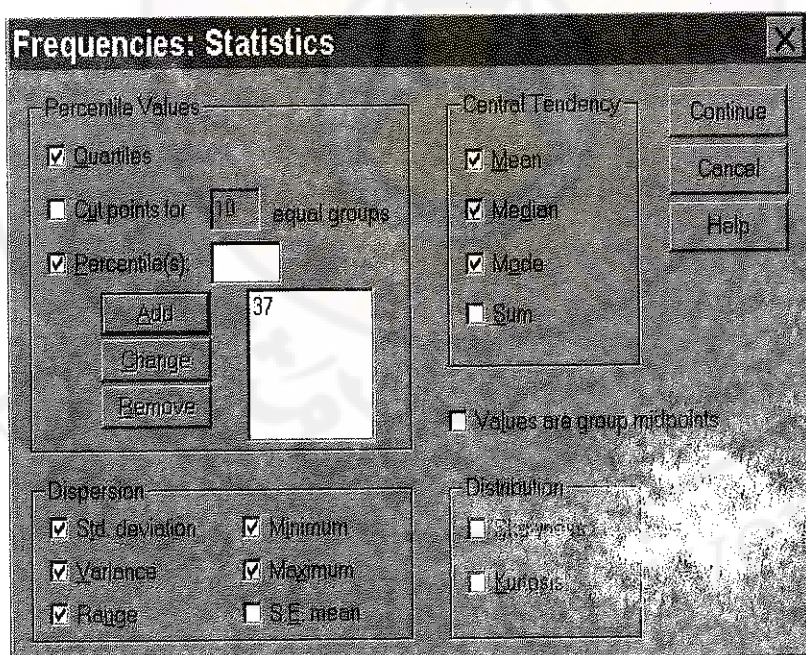
الشكل (4): مربع الحوار الخاص بالأمر الفرعي Statistics

يمكن تفعيل معظم الخيارات في الشكل (4) بالنقر على خانة الخيار Checkbox المجاور لها. إذا أردنا تحديد مئيني معين فإن ذلك يتم أولاً بتفعيل الخيار Percentile (s) و من ثم وضع قيمة المئيني المطلوب في مربع النص المجاور ، وأخيراً الضغط على المفتاح Add لإضافته إلى قائمة الأوامر الفعالة. يوضح الشكلان (5) و (6) عملية تحديد بعض الخيارات في مربع الحوار و تحديد المئيني 37. لاحظ أنه يمكن دائماً إضافة أكثر من مئيني ، حذف مئيني معين Remove ، و تغيير قيمة مئيني معين Change.



الشكل (5): تفعيل بعض الخيارات في مربع الحوار

الخيارات المحددة في الشكل (6) تعطي خرج المبين في الشكل (7).



الشكل (6): إضافة المئيني 37.

عدد الحالات	N	Valid	10
		Missing	0
المتوسط	Mean		37.70
الوسط	Median		36.00
المنوال	Mode		28 ^a
الانحراف المعياري	Std. Deviation		7.86
التباين	Variance		61.79
المدى	Range		23
القيمة الصغرى	Minimum		28
القيمة العظمى	Maximum		51
المئينيات	Percentile	25	30.75
		37	32.21
		50	36.00
		75	45.25

الشكل (7): شاشة الخرج.

و فيما يلي تفسير بعض القيم في الشكل (7).

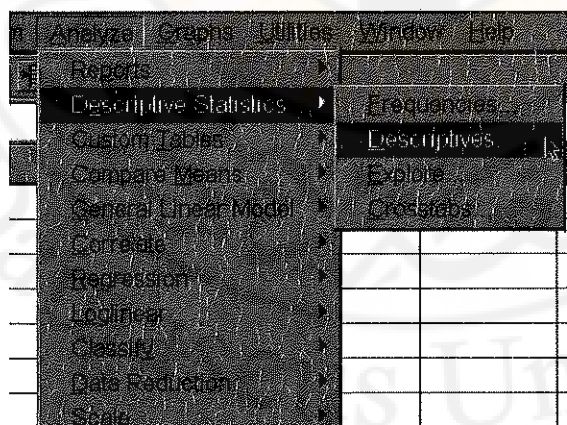
1. المتوسط Mean و قيمته في مثالنا 37.70.
2. الوسط Median و قيمته في مثالنا 36.00.
3. المنوال Mode و قيمته في مثالنا 28.00.

4. التباين Variance و الانحراف المعياري Std Deviation و قيمهما في مثالنا 7961 و 7.86 على الترتيب.
5. المدى Range و قيمته في مثالنا 23.
6. القيمة العظمى Max و القيمة الصغرى Min و قيمهما في مثالنا 51 و 28 على الترتيب.
7. المئينيات Percentiles: المئيني 25 له القيمة 30.75 و المئيني 37 له القيمة 32.21 و هكذا.

الاحصاء الوصفي Descriptive Statistics

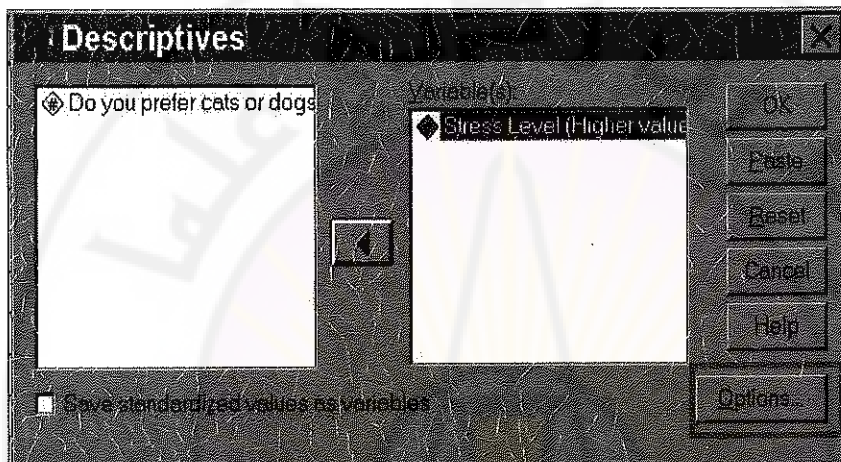
يفيد أمر الاحصاء الوصفي بوصف عينة من المشاهدات من خلال العديد من الإحصائيات نذكر منها: المتوسط ، الوسط ، الانحراف المعياري. يستخدم الاحصاء الوصفي مع البيانات الكمية أو عندما يكون للمتوسط أو الانحراف المعياري معنى. على سبيل المثال ، إذا كان لدينا متغير اسمي مثل الجنس حيث يرمز للذكر بـ 1 و للأنثى بـ 2 فإنه من الواضح أن حساب المتوسط لمتغير الجنس لا معنى له.

يتم تفعيل هذا الأمر من خلال القائمة **Analyze** ، و من ثم اختيار الأمر **Descriptives** كما هو موضح في الشكل (8).



الشكل (8): اختيار أمر الاحصاء الوصفي من القائمة **Analyze**.

ونتيجة لذلك يظهر الـ SPSS نافذة الحوار في الشكل (9) التي تمكن من القيام بعدة مهام من أهمها تحديد المتغير (أو المتغيرات التي نريد وصفها) و الإحصائيات التي نريد إظهارها في شاشة الخرج Output Screen. تحديد المتغيرات التي نريد وصفها بسيط. فقط حدد المتغير (أو المتغيرات) في الجزء الأيسر من مربع الحوار و من ثم انقر على السهم ► لنقلها إلى الجانب الأيمن حيث تصبح جاهزة للوصف.



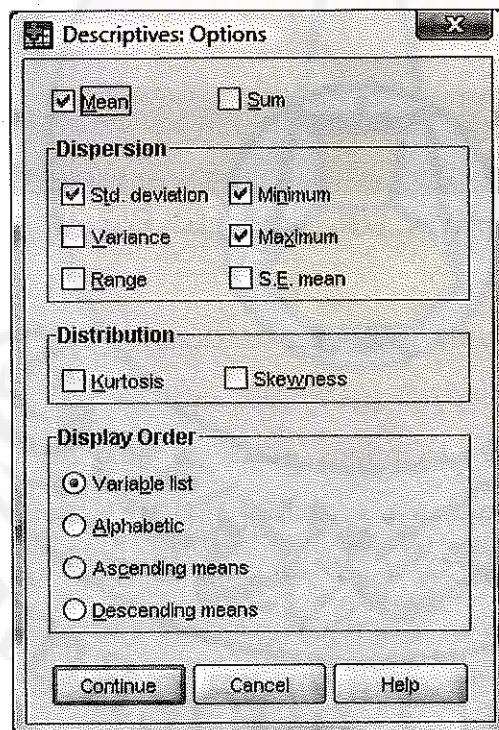
الشكل (9): مربع الحوار الخاص بالأمر Descriptives

يحدد الـ SPSS بعض الإحصائيات بشكل تلقائي ومنها المتوسط ، الانحراف المعياري. لكن هناك العديد من الإحصائيات الأخرى المتوفرة التي يمكن تفعيلها يدوياً من قبل المستخدم و ذلك بالنقر على المفتاح Options في مربع الحوار أعلاه حيث يظهر مربع الحوار في الشكل (10). في هذا المربع سوف تجد بعض الخيارات المفعلة ألياً (يظهر ذلك بوجود الإشارة ☒ بجانب اسم الأمر). يمكن بالطبع تفعيل إحصائيات إضافية و ذلك بالنقر بزر الفأرة الأيسر على مربع الاختيار بجانب الإحصائية المطلوبة. بهذه الخيارات التي تظهر في الشكل (10) ، يعرض الـ SPSS النتائج التالية و التي تمثل الوصف الإحصائي للعينة من خلال مجموعة من الإحصائيات: عدد أفراد العينة N ويساوي "10" ، أصغر قيمة Minimum و تساوي "28" ، أكبر قيمة Maximum

و تساوي "51" ، المتوسط Mean و يساوي "37.70" ، و أخيراً الانحراف المعياري Std. Deviation و يساوي "7.86".

Descriptive Statistics

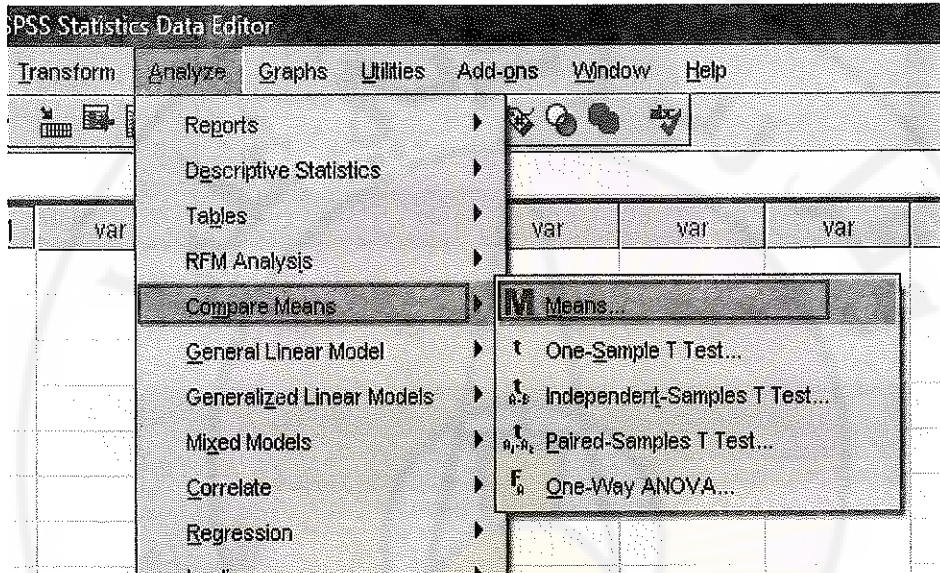
		Minimum	Maximum	Mean	Std Deviation
Stress Level (Higher values= more stressed)	10	28	51	37.70	7.86
Valid N (listwise)	10				



الشكل (10): مربع الحوار الخاص بـ Options

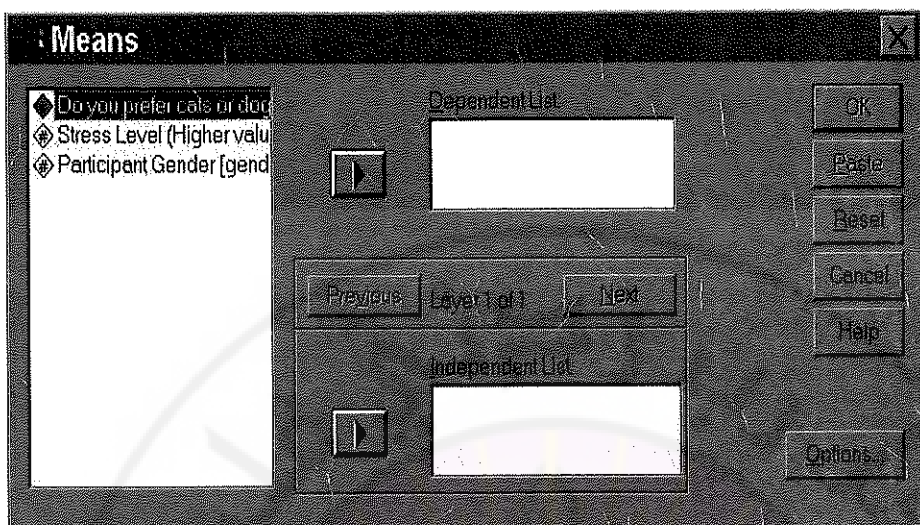
مقارنة المتوسطات Compare Means

مقارنة المتوسطات هي وسيلة لتوليد إحصائيات وصفية لمجموعات متعددة. يمكن تفعيل هذا الأمر من خلال القائمة **Analyze** و من ثم اختيار الأمر **Means** كما يوضح ذلك الشكل (11). بتفعيل هذا الخيار يظهر مربع الحوار في الشكل (12).

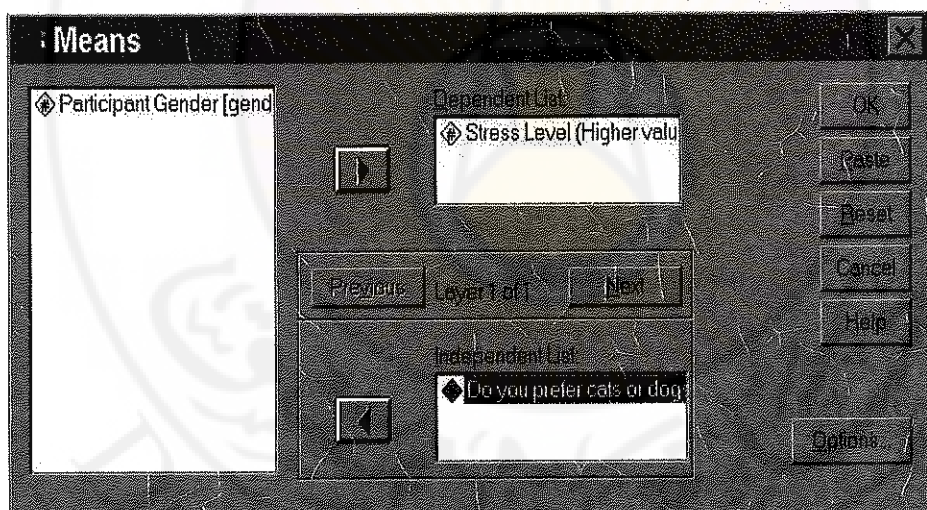


الشكل (11) تفعيل الأمر Compare Means

لتنفيذ هذا الأمر نحتاج إلى تحديد متغيرين في مربع الحوار في الشكل (12). المتغير التابع **Dependent Variable** و يمثل المتغير الذي نريد وصفه ، و المتغير الثاني هو قائمة من المتغيرات المستقلة **Independent List** التي تمثل الأعمدة التي تحوي قيماً تحدد المجموعات المختلفة في المتغير التابع. نضع المتغيرات التي نريد وصفها بالطريقة المعتادة و ذلك بنقلها من الجزء اليساري إلى الأماكن المخصصة لها في الجزء اليميني. يوضح الشكل (13) مربع الحوار بعد نقل المتغيرات إلى مكانها المخصص.



الشكل (12): مربع الحوار الخاص بالأمر Means



الشكل (13): وضع المتغيرات قيد الوصف في أماكنها المخصصة.

يُظهر الشكل (14) الخرج الكلاسيكي لأمر مقارنة المتوسطات. يظهر الشكل (14.a) عدد القياسات في العينة ، ويشتمل ذلك على عدد الحالات المشمولة في

الحسابات (included) ، وعدد الحالات المستثناة من الحسابات (excluded). كما هو واضح في الشكل (14.b) فإن متوسط درجة القلق لدى مالكي القطط هو 37.20 و الانحراف المعياري 8.29 أما متوسط درجة القلق لدي مالكي الكلاب 38.20 بانحراف معياري قدره 8.35. المتوسط الكلي لدرجة القلق عند أفراد العينة بغض النظر عن كونهم يملكون قططاً أو كلاباً هو 37.4 و بانحراف معياري قدره 7.86.

	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
Stress Level (Higher values = more stressed) * Do you prefer cats or dogs?	10	100.0%	0	.0%	10	100.0%

(a)

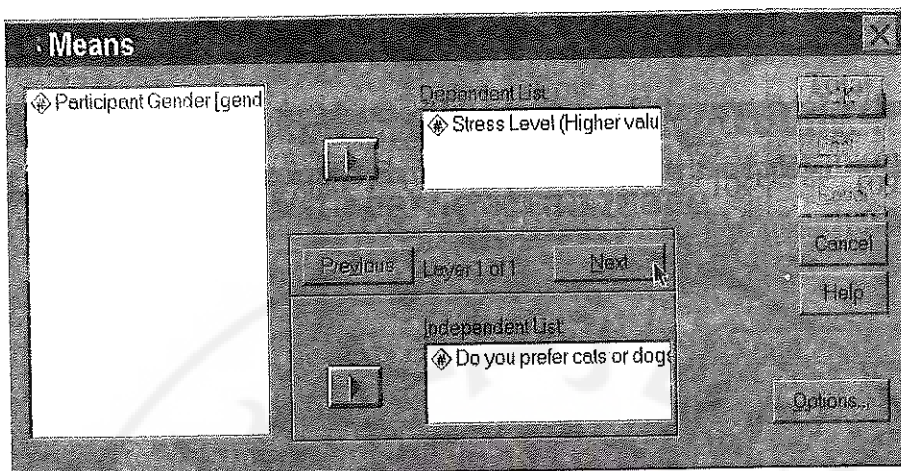
Stress Level (Higher values = more stressed)

Do you prefer cats or dogs?	Mean	N	Std. Deviation
Cats	37.20	5	8.29
Dogs	38.20	5	8.35
Total	37.70	10	7.86

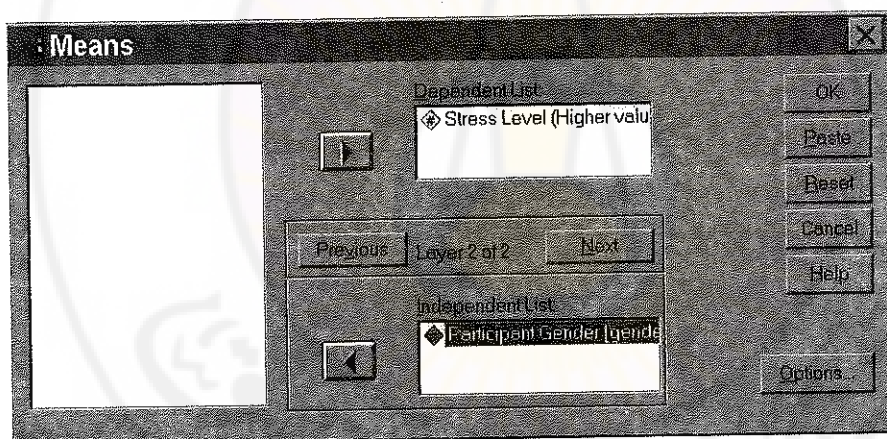
(b)

الشكل (14): نتائج تنفيذ الأمر Means على البيانات.

إذا كنا نريد دراسة أكثر من مجموعة مختلفة ، نستطيع أن نستخدم أمر الطبقات أو Layers في مربع الحوار في الشكل (15). على سبيل المثال ، يمكن أن نحسب متوسط درجة القلق لدى النساء و لدى الرجال ، و ذلك لأجل كل نوع من الحيوانات الأليفة. نبدأ إضافة طبقات بالضغط على الزر Next كما موضح بالشكل (16). نتيجة ذلك يفتح حقل جديد ضمن Independent List تمكنا من تعريف متغير تصنيف جديد كما يبدو في الشكل (16). في هذا الحقل نضيف المتغير Gender أي الجنس.



الشكل (15): فتح طبقة جديدة لإضافة متغير تصنيف جديد.



الشكل (16): إضافة متغير التصنيف الإضافي Gender.

يظهر الشكل (17) نتائج هذا الأمر. لاحظ أنه تم حساب درجة القلق عند الذكور، و ذلك لأجل كل نوع من أنواع الحيوانات الأليفة (هنا كلاب و قطط). كذلك تم حساب درجة القلق عند الإناث و ذلك لأجل كل نوع من أنواع الحيوانات الأليفة.

Stress Level (Higher values = more stressed)

Do you prefer cats or dogs?	Participant Gender	Mean	N	Std. Deviation
Cats	Male	41.50	2	6.36
	Female	34.33	3	9.29
	Total	37.20	5	8.29
Dogs	Male	35.00	3	6.08
	Female	43.00	2	11.31
	Total	38.20	5	8.35
Total	Male	37.60	5	6.43
	Female	37.80	5	9.88
	Total	37.70	10	7.86

الشكل (17): نتائج وصف درجة القلق عند الجنسين و بالنسبة لأنواع الحيوانات الأليفة.

تمرين Exercise

1. أدخل البيانات التالية بشكل صحيح إلى شاشة البيانات في الـ SPSS.

Group	Gender	Hw1	Hw2	Hw3
Expt	Male	92	84	93
Expt	Female	77	84	85
Expt	Male	87	86	81
Expt	Female	89	90	93
Expt	Male	64	73	78
Control	Female	81	84	93
Control	Male	83	90	91
Control	Female	84	88	86
Control	Male	82	80	78
Control	Female	96	91	88

2. أوجد كلاً من المتوسط Mean ، الوسط Median ، المنوال Mode ، الربعيات Quartiles ، التباين Variance ، الانحراف المعياري Standard Deviation ، القيمة الصغرى Minimum ، و القيمة العظمى Maximum وذلك لأجل الوظائف Hw1, Hw2, Hw3.

3. استخدم أمر مقارنة المتوسطات Compare Means و ذلك لإيجاد متوسط:
كل أفراد العينة ، المجموعة التجريبية (expt) Experimental Group ،
المجموعة الضابطة (control) Control Group ، لأجل الإناث ، لأجل
الذكور ، و أخيراً لأجل كل تركيبات الجنس و المجموعات المختلفة.

فحص عشوائية العينات

نفترض العديد من الإجراءات الإحصائية استقلالية قياسات العينة. بكلمات أخرى،
إن الترتيب الذي جمعت به العينات غير مهم. أما إذا كان الترتيب مهماً فإن العينة
ليست مستقلة؛ أو ليست عشوائية. إذا لم تكن العينة عشوائية فإننا لانستطيع من حيث
المبدأ الحصول على استنتاجات صحيحة حول مجتمع العينة. لذلك يعد إجراء اختبار
عشوائية العينة مهماً للغاية.

يمكن استخدام برنامج الـ SPSS لإجراء اختبار العشوائية. يقوم الإجراء بأخذ
نقطة مرجعية أو ما يسمى **cut point** و من ثم تحديد موضع كل قياس في العينة
بالنسبة لهذه النقطة المرجعية ومن ثم التأكد من أن مواضع النقاط لا تملك نمطاً محدداً.
إذا كان موضع النقاط بالنسبة للنقطة المرجعية يملك نمطاً معيناً فإن ذلك يشير إلى أن
العينة غير عشوائية.

يمكن للنقطة المرجعية أن تكون أيّاً من مقاييس النزعة المركزية (المتوسط ، الوسط
، المنوال). كما يمكن للنقطة المرجعية أن تكون أي نقطة اختيارية يحددها الباحث.

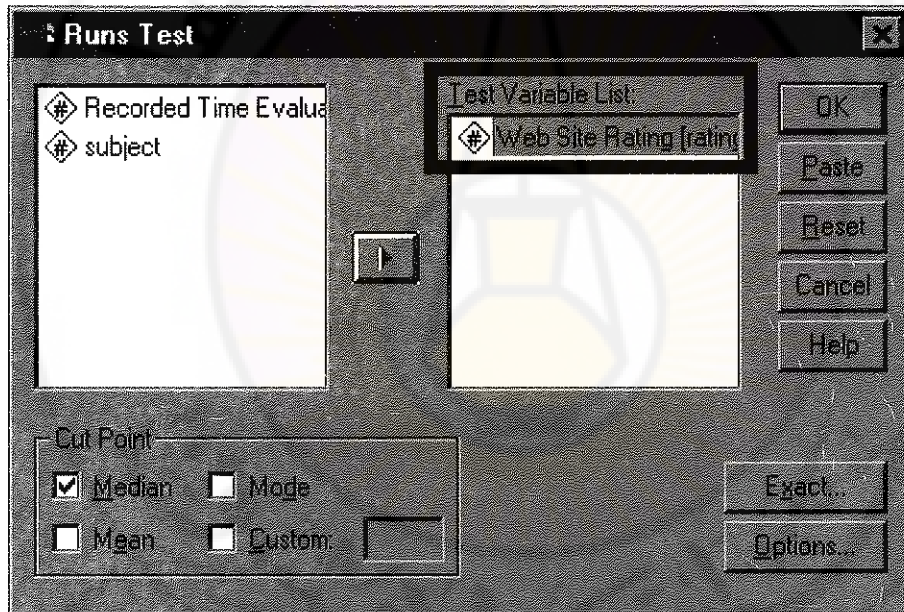
لتكن لدينا مجموعة القياسات التالية:

8	8	12	11
7	9	9	14
8	7	11	9
6	1	12	14
1	7	10	11
8	8	13	12
6	1	13	11
7	1	12	13

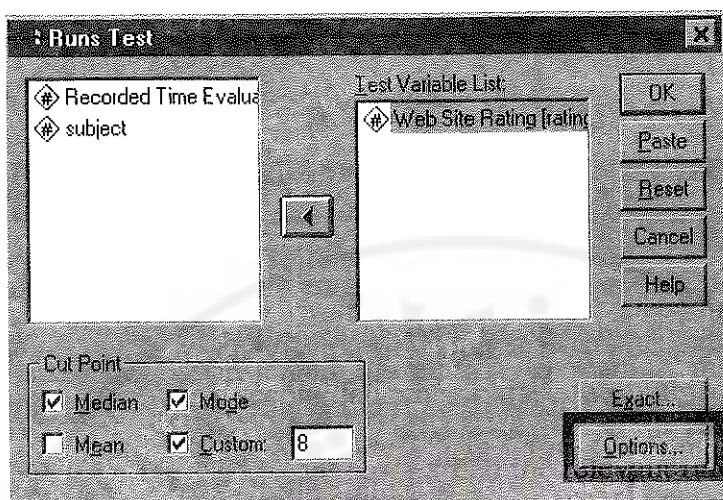
هل العينة في الجدول أعلاه عشوائية؟

بفرض أن هذه البيانات مدخلة إلى شاشة الـ SPSS في عمود اسمه " Web Site Rating". يتم اختبار عشوائية العينة وفق الخطوات التالية:

1. انقر على القائمة Analyze ، اختر الأمر Nonparametric Tests ، ثم الأمر الفرعي Runs... يظهر نتيجة لذلك مربع الحوار في الشكل (18).
2. نقل العمود المطلوب إلى المربع اليميني كما هو موضح بالشكل (18).
3. تحديد النقاط المرجعية لاختبار العشوائية كما هو موضح بالشكل (19).



الشكل (18): تحديد العمود الذي يحوي العينة المطلوب اختبار عشوائيتها.

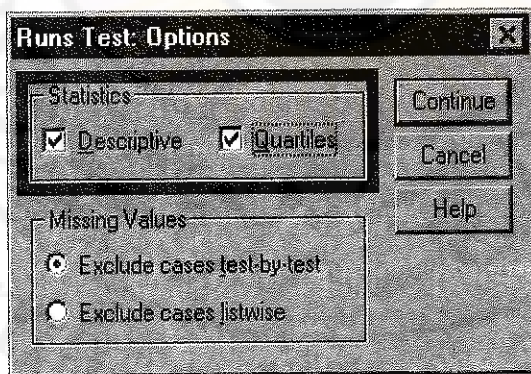


الشكل (19): تحديد النقاط المرجعية.

بشكل اختياري يمكن الحصول على بعض الإحصائيات الوصفية مثل المتوسط و الانحراف المعياري و غيره من خلال الخطوات التالية:

1. النقر على **Options** في الشكل (19) حيث يفتح مربع الحوار في الشكل (20).

2. تفعيل الخيارات **Descriptive** و **Quartiles** كما هو موضح بالشكل (20).



الشكل (20): تفعيل بعض الخيارات الوصفية.

بعد الانتهاء من هذه التحديدات يقوم الـ SPSS بعرض نتائج اختبار العشوائية. يبين الشكل (21) نتائج اختبار العشوائية و ذلك لأجل أربعة نقاط مرجعية: الوسط Median (الشكل 21.A) ، المتوسط Mean (الشكل 21.B) ، المنوال Mode (الشكل 21.C) ، و أخيراً نقطة محددة من قبل المستخدم (الشكل 21.D).

Runs Test	
	Web Site Rating
Test Value ^a	10
Cases < Test Value	14
Cases >= Test Value	18
Total Cases	32
Number of Runs	10
Z	-2.283
Asymp. Sig. (2-tailed)	.022

a. Median

(A)

Runs Test 2	
	Web Site Rating
Test Value ^a	9.94
Cases < Test Value	14
Cases >= Test Value	18
Total Cases	32
Number of Runs	10
Z	-2.283
Asymp. Sig. (2-tailed)	.022

a. Mean

(B)

Runs Test 3	
	Web Site Rating
Test Value ^a	12 ^b
Cases < Test Value	22
Cases >= Test Value	10
Total Cases	32
Number of Runs	16
Z	.315
Asymp. Sig. (2-tailed)	.752

a. Mode

b. There are multiple modes. The mode with the largest data value is used.

(C)

Runs Test 4	
	Web Site Rating
Test Value	8.00
Total Cases	32
Number of Runs	11
Z	.000
Asymp. Sig. (2-tailed)	1.000

a. User-specified.

(D)

الشكل (21): نتائج اختبار العشوائية.

الفرضيات التي يجري اختبارها.
فرضية العدم H_0 : العينة عشوائية.
الفرضية البديلة H_1 : العينة ليست عشوائية.

بالنظر إلى الشكل (23.A) نجد المعلومات التالية:

✧ القيمة المرجعية Test Value. في هذا الشكل هي الوسط Median و تساوي

"10"

✧ عدد القيم التي هي أصغر من الوسط Cases < Test Value و عددها في مثالنا

"14"

✧ عدد القيم التي هي أكبر أو تساوي الوسط Cases >= Test Value و عددها في

مثالنا "18"

✧ عدد الحالات الكلي Total Cases و عددها في مثالنا "32"

✧ عدد الأشواط Number of Runs فوق و تحت القيمة المرجعية و في مثالنا "10"

✧ الإحصائية Z التي تستخدم من أجل حساب قيمة معنوية الاختبار.

✧ معنوية الاختبار Asymp. Sig و تستخدم لاختبار الفرضيات حول العشوائية.

لاحظ أن الاختبار معنوي لأن $Sig < 0.05$ و بالتالي نرفض فرضية العدم. مما

يدل على أن العينة ليست عشوائية إذا كان الوسط Median هو القيمة المرجعية.

لاحظ أيضاً من الشكل (23.B) أن العينة ليست عشوائية إذا أخذنا المتوسط قيمة

مرجعية لأن الاختبار معنوي ($Sig < 0.05$). الأشكال (23.c و 23.D) تظهر أن

اختبار العشوائية ليس معنوياً ، و بالتالي فإن العينة عشوائية إذا أخذنا أكبر منوال في

العينة قيمة مرجعية ، و كذلك إذا أخذنا القيمة المحددة من قبل الباحث "8" قيمة

مرجعية.

★★★★★★★★★★

الاختبارات الوسيطة و غير الوسيطة

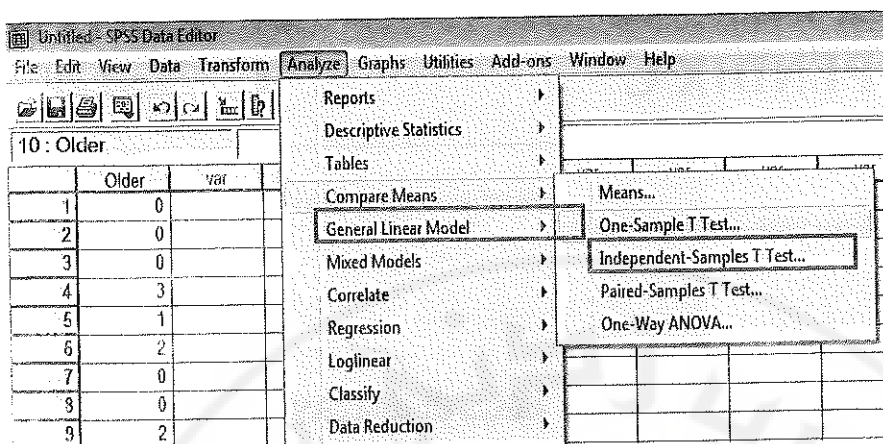
في هذا الملحق نقدم تدريبات حول كيفية إدخال البيانات في الأمثلة المطروحة في الاختبارات الوسيطة و غير الوسيطة (الفصلان الخامس و السادس). بالإضافة إلى كيفية اختيار كل اختبار إحصائي و تحديد وسطائه. كل ذلك بهدف إغناء تجربة القارئ و تزويده بالمهارة الكافية لاستخدام البرنامج الإحصائي SPSS.

إجراء اختبار ستودينت لعينة واحدة باستخدام الـ SPSS

بفرض أن لدينا العينة المطروحة في الفصل الخامس التي نعرضها مرة أخرى هنا للتسهيل.

البيانات: 0, 0, 0, 3, 1, 2, 0, 0, 2, ...

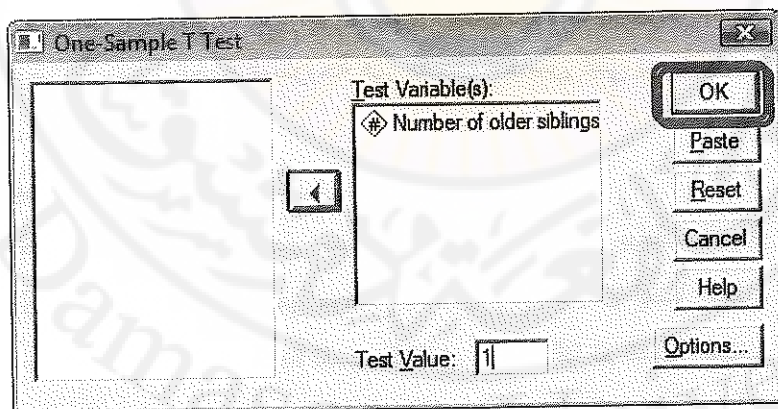
خطوات إجراء هذا الاختبار سهلة للغاية. نقوم بإدخال بيانات العينة إلى شاشة الـ SPSS في عمود واحد كما هو موضح بالشكل (1). يفضل أيضاً إعطاء اسم معبر لذلك العمود و ذلك لسهولة العودة تفسير النتائج. (في مثالنا تم تسمية المتغير اسم معبر ، مثل Older ، كما هو موضح بالشكل (1). بعد إدخال البيانات و التأكد من صحة الإدخال ، نقوم بتفعيل اختبار الإحصائي المناسب و هو في حالتنا اختبار ستودنت لعينة واحدة (One-Sample T Test) كما مبين في الشكل (1).



الشكل (1): يبين شاشة البيانات في الـ SPSS بعد إدخال البيانات و يظهر كذلك

في الشكل الأمر One-Sample T Test

بعد تفعيل الاختبار تظهر الشاشة في الشكل (2) التي يمكننا من تحديد العمود الذي يحوي العينة و تحديد الثابت الذي نقارن معه. ننقل العمود **Number of older siblings** (وهو اسم آخر لـ Old) إلى القائمة اليمينية وذلك بتعليم اسم العمود في القائمة اليسارية و من ثم نقر على السهم  نضع الثابت الذي نقارن مع "1" بكتابته مباشرة في خانة النص **Test Value**.



الشكل (2): نافذة الأمر One-Sample T Test بعد نقل العمود و وضع

الثابت المختبر "1" في خانة النص المخصصة.

بالضغط على المفتاح OK في الشكل (2) نحصل على النتائج التي تم عرضها و مناقشتها في الفصل الخامس.

إجراء اختبار ستودنت لعينتين مستقلتين

يفرض أن لدينا العينة المطروحة في الفصل الخامس التي نعرضها مرة أخرى هنا للتسهيل. تذكر أن هذه البيانات تتعلق بدراسة تأثير الضغط من الأقران في الأندية النسائية الجامعية على عدم الانتظام بالأكل و بالتالي حدوث السمنة لدى النساء في النوادي الجامعية.

125	120	90	115	105	نساء يعانين من الضغط النفسي من الأقران
110	130	120	135	140	نساء لا يعانين من الضغط النفسي من الأقران

إدخال البيانات إلى شاشة بيانات الـ SPSS

يتطلب استخدام اختبار ستودنت لمقارنة متوسطي عينتين مستقلتين إدخال البيانات كما هو موضح في الشكل (3). يحتوي العمود المسمى **Weight** على أوزان أفراد العينة. يمثل العمود الأول من اليسار المجموعات (**group**) ، حيث يمثل الرقم "1" مجموعة النساء في النوادي الجامعية (**sorority**) ، و الرقم "2" مجموعة النساء من خارج النوادي الجامعية (**others**).

ملاحظة: من الأفضل إعادة ترميز الأرقام "1" و "2" في العمود **group** بما تعنيه (**sorority** و **others**) ، و ذلك لتسهيل عرض النتائج و تفسيرها. تتم عملية إعادة ترميز القيم كما تعلمنا في الملحق السابق.

الاختبار Two-Sample Independent T Test

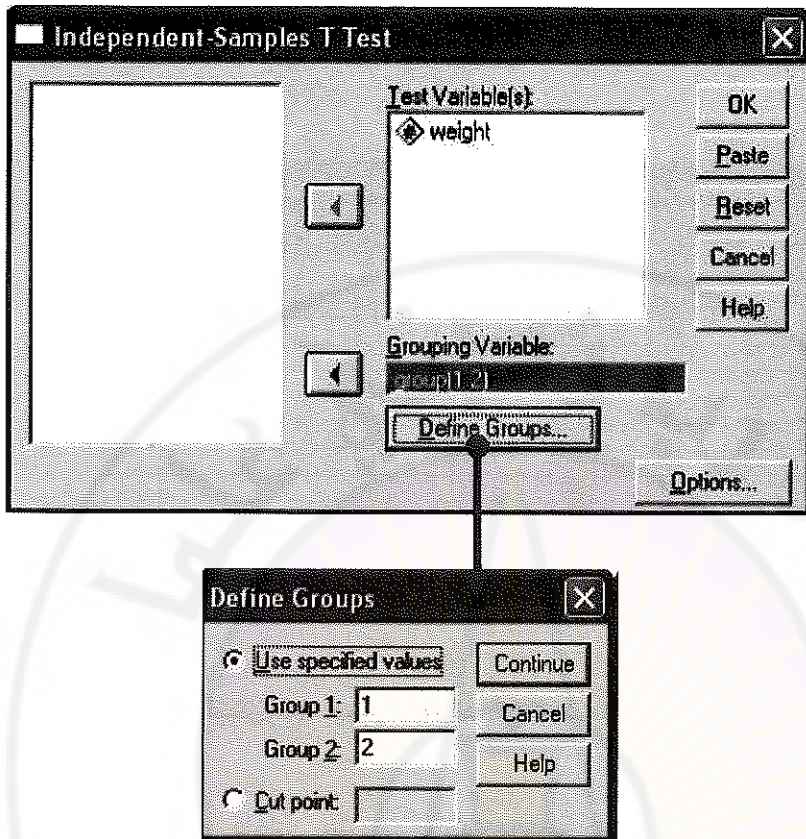
يتم التفعيل و التعامل مع الشاشات المختلفة لاختبار ستودنت للعينات المستقلة كمايلي:

1. افتح القائمة **Analyze** ، اختر الأمر **Compare Means** ، ثم اختر الأمر الفرعي **Two-Sample Independent T Test**. نتيجة لذلك تظهر الشاشة المعنونة بـ **Independent Samples T Test** في الشكل (4).

	Group	Weight	var	var	var	var	var
1	1	120					
2	1	125					
3	1	90					
4	1	115					
5	1	105					
6	2	110					
7	2	130					
8	2	120					
9	2	135					
10	2	140					
11							
12							

الشكل (3): إدخال البيانات في شاشة الـ SPSS

2. نضع المتغير الذي يحوي البيانات في مربع النص المجاور لـ **Test Variable** (في مثالنا **weight**) و نضع المتغير الذي يُعرّف المجموعتين **group** في مربع النص المجاور لـ **Grouping Variable**.
3. ننقر على المفتاح **Define Groups** فتظهر الشاشة المعنونة بـ **Define Groups** كما يظهر في الشكل (4). في هذه الشاشة نعرف المجموعتين اللتين نريد تعريفهما حيث نضع "1" في **Group 1** و "2" في **Group 2**.
- ننقر على **Continue** ومن ثم على **OK** في الشاشة المبينة في الشكل (4) فيقوم الـ SPSS بإجراء الحسابات اللازمة و إصدار النتائج.



الشكل (4): الشاشات الخاصة باختبار ستودنت للعينات المستقلة.

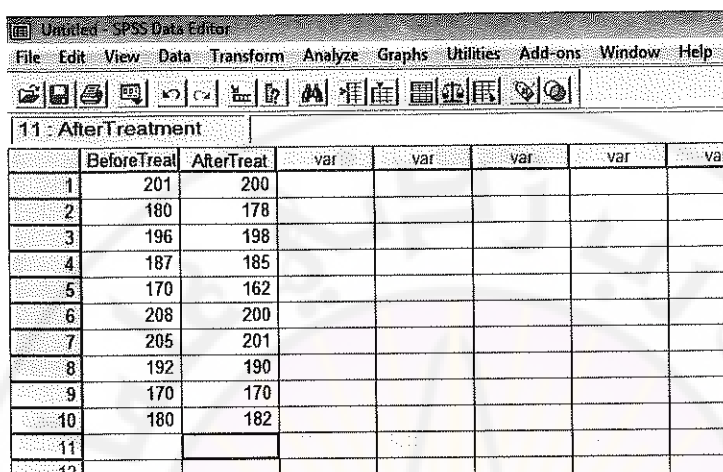
إجراء اختبار ستودنت لعينتين مرتبطتين

بفرض أن لدينا العينة المطروحة في الفصل (5) التي نعرضها مرة أخرى هنا للتسهيل.

101	180	196	187	170	208	105	192	170	180	قبل المعالجة
200	178	198	185	162	200	101	190	170	182	بعد المعالجة

لإجراء هذا الاختبار ، ندخل البيانات في شاشة الـ SPSS و في عمودين مختلفين كما موضح بالشكل (5). حيث يمثل العمود BeforeTreat القياسات قبل المعالجة ،

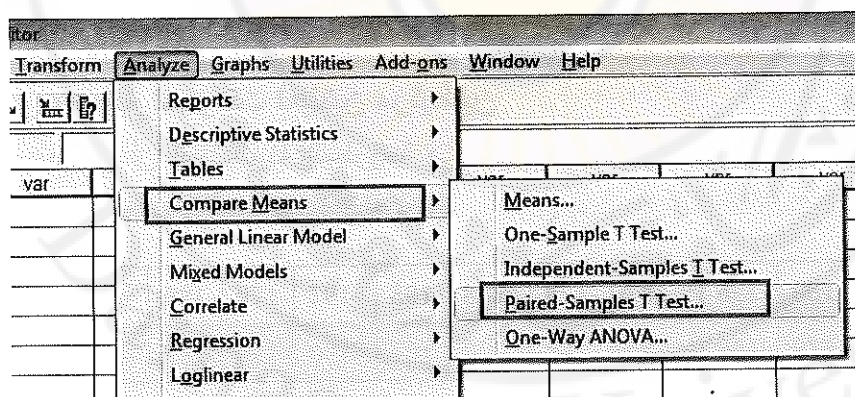
ويمثل العمود AfterTreat القياسات بعد المعالجة. بعد إدخال البيانات يمكن إجراء التحليل طبقاً للخطوات التالية.



	BeforeTreat	AfterTreat	var	var	var	var	var
1	201	200					
2	180	178					
3	196	198					
4	187	185					
5	170	162					
6	208	200					
7	205	201					
8	192	190					
9	170	170					
10	180	182					
11							
12							

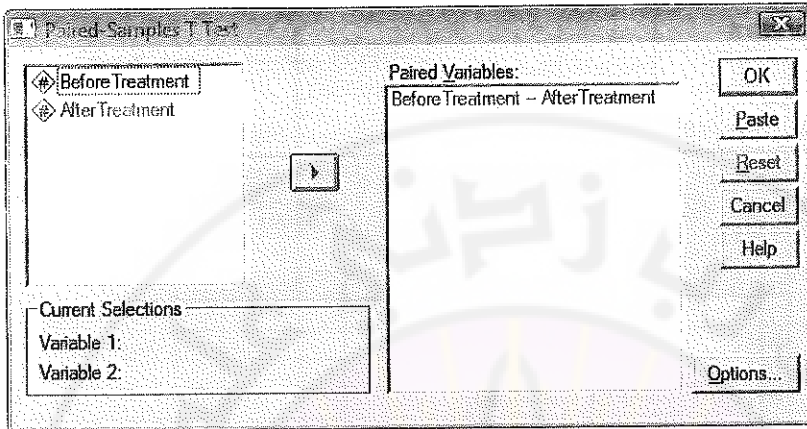
الشكل (5): إدخال البيانات في شاشة الـ SPSS.

نفعل الأمر الخاص باختبار ستودنت لعينات مرتبطة كما يظهر في الشكل التالي:



حيث يظهر المربع الخاص بهذا الأمر في الشكل (6). ننقل العمودين BeforeTreat و AfterTreat كما تعلمنا سابقاً إلى المربع المعنون بـ Paired

Variables حيث يصبحان جاهزين للمعالجة. بعد ذلك نقوم بالضغط على المفتاح OK فيقوم الـ SPSS بإجراء الاختبار و إظهار النتائج.



الشكل (6): مربع الحوار الخاص باختبار ستودنت لعينتين مرتبطتين.

اختبار تحليل التباين بمتغير واحد

بفرض أن لدينا العينة المطروحة في الفصل الخامس التي نعيد عرضها مرة أخرى هنا للسهولة.

<i>Pain Scores for Analgesics</i>	
Drug	Pain Score
Diclofenac	0, 35, 31, 29, 20, 7, 43, 16
Ibuprophen	30, 40, 27, 25, 39, 15, 30, 45
Paracetamol	16, 33, 25, 32, 21, 54, 57, 19
Aspirin	55, 58, 56, 57, 56, 53, 59, 55

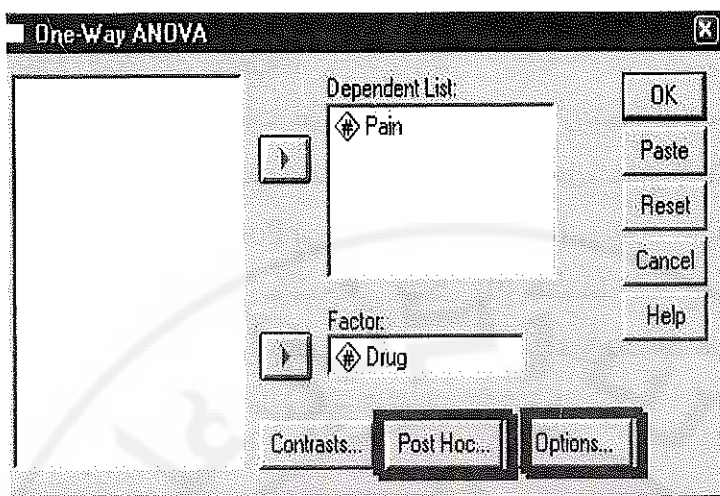
لاختبار فرضيات تحليل التباين باستخدام الـ SPSS نقوم بالخطوات التالية:

1. إدخال البيانات في شاشة البيانات كما مبين في الشكل (7). لاحظ أننا أعدنا ترميز قيم المتغير المستقل Drug باستخدام الأرقام 1, 2, 3, 4 لكي نستطيع استخدام تحليل التباين. بعد إدخال البيانات كما هو موضح بالشكل (7) ، أصبح جاهزين لتشغيل الاختبار.

File Edit View Data Transform Analyze Graphs Utilities Window Help					
10:					
	Drug	Pain	var	var	var
1	1	0			
2	1	35			
3	1	31			
4	1	29			
5	1	20			
6	1	7			
7	1	43			
8	1	16			
9	2	30			
10	2	40			
11	2	27			
12	2	25			
13	2	39			
14	2	15			
15	2	30			
16	2	45			
17	3	16			
18	3	33			
19	3	25			
20	3	32			
21	3	21			

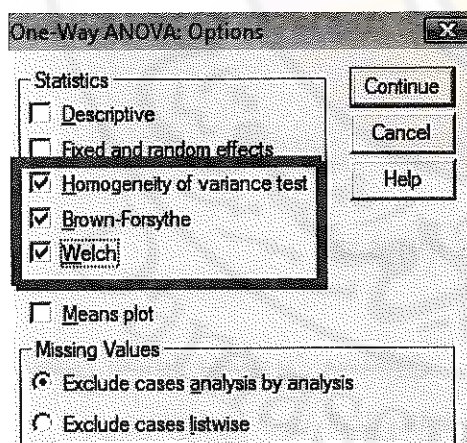
الشكل (7): إدخال البيانات.

2. إفتح القائمة Analyze ، اختر الأمر Compare Means ، ثم اختر الأمر الفرعي One-way ANOVA. نتيجة لذلك يظهر مربع الحوار المبين في الشكل (8).



الشكل (8): شاشة الأمر One-way ANOVA.

3. نضع المتغير التابع Pain في مربع النص المجاور Dependent List و المتغير المستقل Drug في مربع النص المجاور Factor. نضغط أيضاً على الأمر Options لتفعيل خيارات اختبار تجانس التباين Homogeneity of variance test ، اختبار Brown-Forsythe ، و اختبار Welch كما يظهر

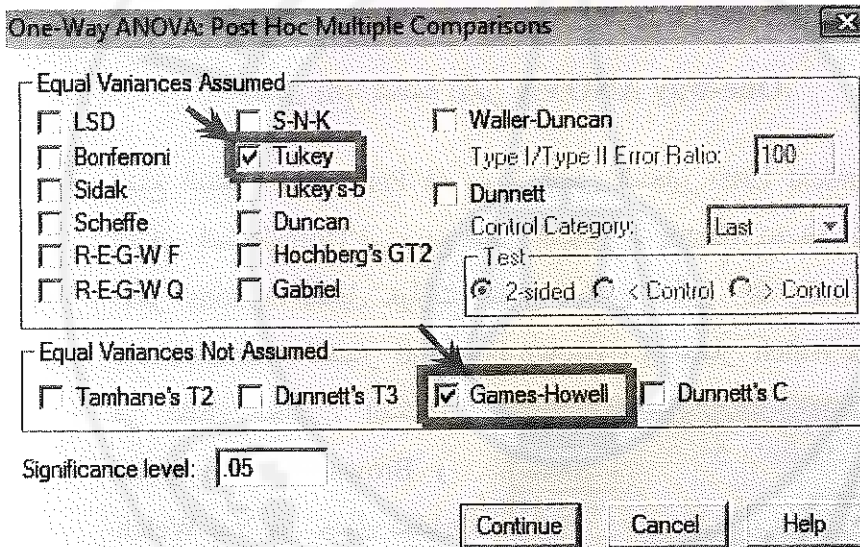


في الشكل الجانبي. (يستخدم اختبارا Brown-Forsythe و Welch بدلاً من جدول تحليل التباين في حال كون التباين غير متجانس بين الكتل المشكلة نتيجة تأثير المتغير المستقل). نضغط Continue في مربع الحوار أدناه لكي نعود إلى مربع الحوار في الشكل (8).

4. بالضغط على OK نحصل على نتائج تحليل التباين.

تفعيل اختبار المقارنات الثنائية

في حال كون المتوسطات غير متساوية فإننا يجب أن نحدد المتوسطات المختلفة. للقيام بذلك نقوم بتفعيل اختبار المقارنات الثنائية. يتم تفعيل اختبار المقارنات الثنائية ، بالنقر على المفتاح Post Hoc في النافذة في الشكل (8) و تفعيل اختبائي Tukey و Games-Howell كما هو موضح في الشكل (9). نضغط المفتاح Continue في الشكل (9) لكي نعود إلى مربع الحوار في الشكل (8). أخيراً نضغط على المفتاح OK في الشكل (8) لكل نحصل على نتائج المقارنات الثنائية.



الشكل(9): نافذة المقارنات الثنائية.

تحليل التباين بمتغيرين مستقلين

بفرض أن لدينا العينة المطروحة في الفصل (5) التي نعيد عرضها مرة أخرى هنا للتسهيل.

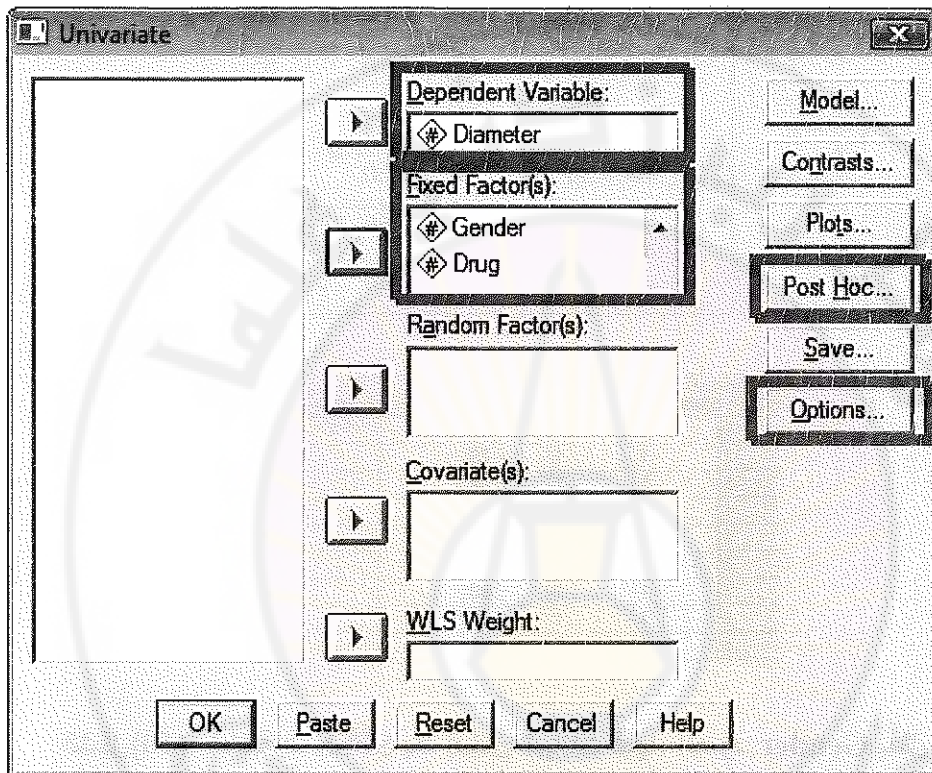
Drug:	Cisplatin		Vinblastine	
Gender:	Female	Male	Female	Male
Tumor Size:	65	50	70	45
	70	55	65	60
	60	80	60	85
	60	65	70	65
	60	70	65	70
	55	75	60	70
	60	75	60	80
	50	65	50	60

	Gender	Drug	Diameter	var	var	var	var
1	1	1	65				
2	1	1	70				
3	1	1	80				
4	1	1	80				
5	1	1	60				
6	1	1	55				
7	1	1	60				
8	1	1	60				
9	1	2	70				

الشكل (10): شاشة الـ SPSS بعد إدخال البيانات فيها. العمود Gender يمثل الجنس ، Drug يمثل أنواع الأدوية و Diameter يمثل قطر الورم.

أولاً لابد من إدخال البيانات في شاشة بيانات قياسات حجم المرض. لأجل ذلك نحتاج إلى ترميز أسماء الأدوية بالرموز "1" و "2" و "3" التي تقابل على الترتيب cisplatin و vinblastine و 5-fluorouracil. كذلك نرسم الجنس بالرموز "1" و "2" التي تقابل على الترتيب Male و Female. يبين الشكل (10) جزءاً من البيانات بعد إدخالها إلى شاشة الـ SPSS. لاحظ أن العمود Diameter يحتوي على البيانات ، العمود Drug يحتوي على ترميز الأدوية الثلاثة ، و Gender يمثل ترميز الجنس.

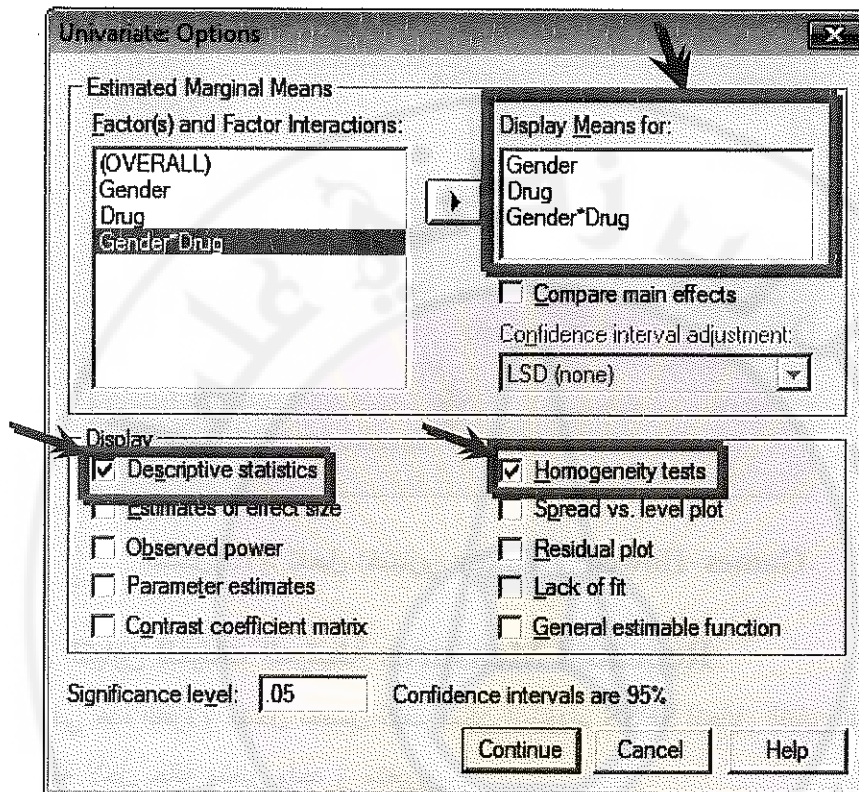
ثانياً يتم تفعيل إجراء تحليل التباين بمتغيرين مستقلين بالنقر على القائمة Analyze ، اختيار الأمر General Linear Model ، و من ثم نختار الأمر الفرعي Univariate. نتيجة لذلك تظهر شاشة الاختبار كما في الشكل (11).



الشكل (11): اختبار تحليل التباين بمتغيرين مستقلين Univariate

في الشكل (11) ، نضع Diameter في مربع النص Dependent Variable و كل من Drug و Gender في مربع النص Fixed Factors. لأجل اختبار فرضية تساوي التباينات نُفعل الأمر Options من الشاشة في الشكل (11) حيث تظهر الشاشة في الشكل (12). نفعل كلاً من تحليل تساوي التباينات Homogeneity test و الاحصاء الوصفي Descriptive statistics. نضع أيضاً

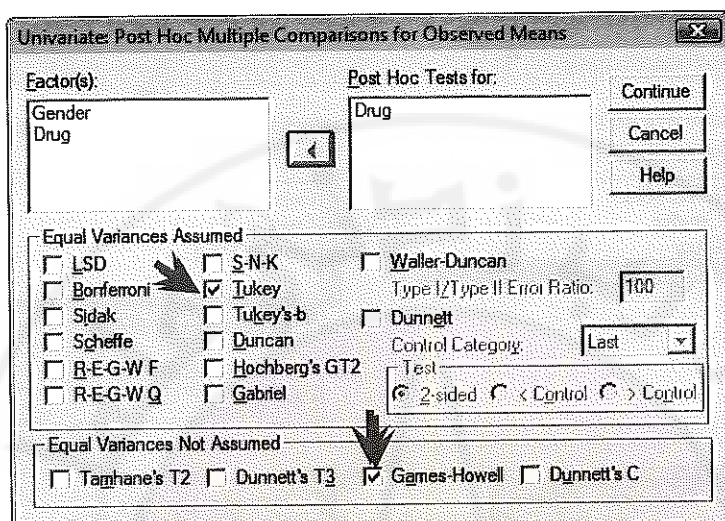
إظهار نتيجة اختبار التأثير المتبادل بين المتغيرات (التفاعل).
 Gender, Drug, Gender*Drug في خانة النص Display Means for و ذلك لأجل



الشكل(12): تفعيل خيارات تحليل التباين

إذا كانت متوسطات حجوم الأورام مختلفة معنوياً عند مستويات الدواء فإن الخطوة التالية هي معرفة أي من المتوسطات يكون مختلفاً بشكل معنوي عن باقي المتوسطات. يستخدم الاختباران Tukey و Games-Howell ، كما نعم ، لإجراء المقارنات الثنائية بين المتوسطات. يتم تفعيل هذين الاختبارين بالضغط على المفتاح Post Hoc في الشاشة (15) حيث تظهر الشاشة في الشكل (13). يتم تفعيل اختبار Tukey و اختبار

Games-Howell بالنقر على مربع الاختيار ☒ المجاور لكلٍ منهما. بعد تفعيل هذين الاختبارين يجب أن تظهر الشاشة كما يلي.



الشكل (13): تفعيل اختباري Tukey و Games-Howell.

تحليل التباين للعينات المرتبطة Repeated Measures

سنبين في هذه الفقرة و بشكل مفصل كيف يمكن استخدام SPSS لتحليل التباين للعينات المترابطة لأجل حالة متغير واحد و متغيرين.

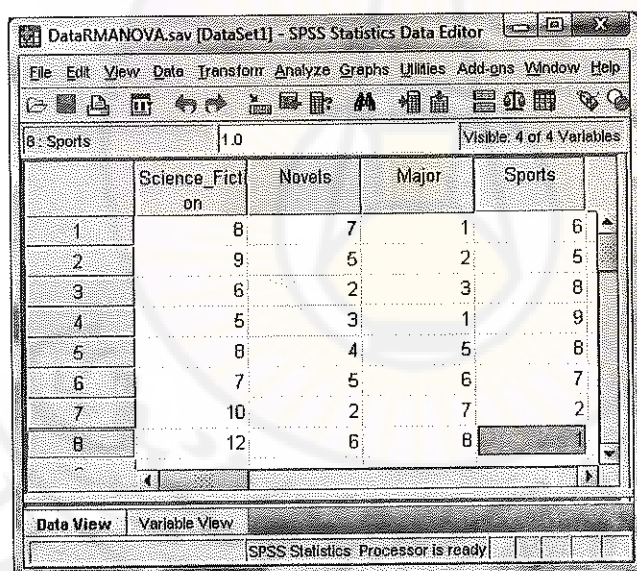
أولاً تحليل التباين بمتغير واحد للعينات المرتبطة

لنفترض أن لدينا البيانات التالية التي تبين عدد الساعات التي يقضيها كل القراء في قراءة أربعة أنواع لمواد القراءة. (للمزيد من المعلومات حول هذه البيانات ، يرجى الانتقال إلى فصل الاختبارات الوسيطة.)

القراء	Sports	Major	Novels	Science fiction
1	6	1	7	8
2	5	2	5	9

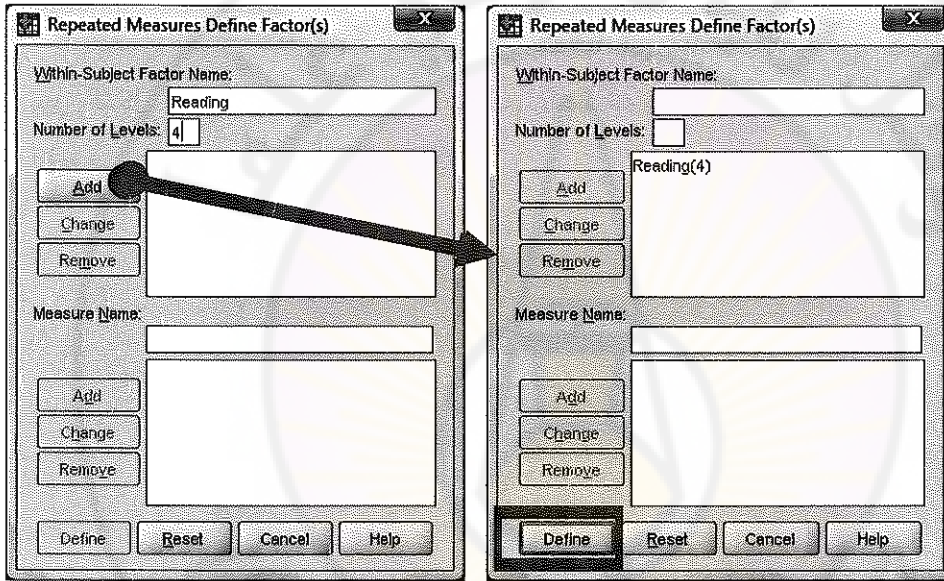
6	2	3	8	3
5	3	1	9	4
8	4	5	8	5
7	5	6	7	6
10	2	7	2	7
12	6	8	1	8

الخطوة الأولى لتحليل هذه البيانات هي إدخالها إلى شاشة البيانات في SPSS. يبين الشكل التالي كيفية إدخال هذه البيانات. لاحظ أنه يجب إدخال قياسات كل مادة قراءة في عمود مستقل في شاشة الـ SPSS.



يتم تفعيل أمر تحليل التباين لعينات مترابطة من خلال اختيار القائمة Analyze، اختيار الأمر General Linear Model، ثم اختيار الأمر الفرعي Repeated Measures. يظهر مربع الحوار الخاص في الشكل (17.A). في هذا المربع نضع

اسم المتغير المرتبط في خانة النص المعنونة "Within-Subject Factor Name".
 في حالتنا المتغير المرتبط هو نوع المادة المقروءة لذلك وضعنا الاسم Reading. بما
 أن متغير نوع المادة المقروءة له أربعة مستويات (أنواع مواد القراءة) فيجب أن نُعرّف
 عدد المستويات في خانة النص "Number of Levels" و لذلك وضعنا 4 في هذه
 الخانة. ننقر على المفتاح Add ليتم تعريف المستويات الأربعة و إحصائيات أخرى.
 يظهر نتيجة لذلك مربع الحوار في الشكل (17.B).

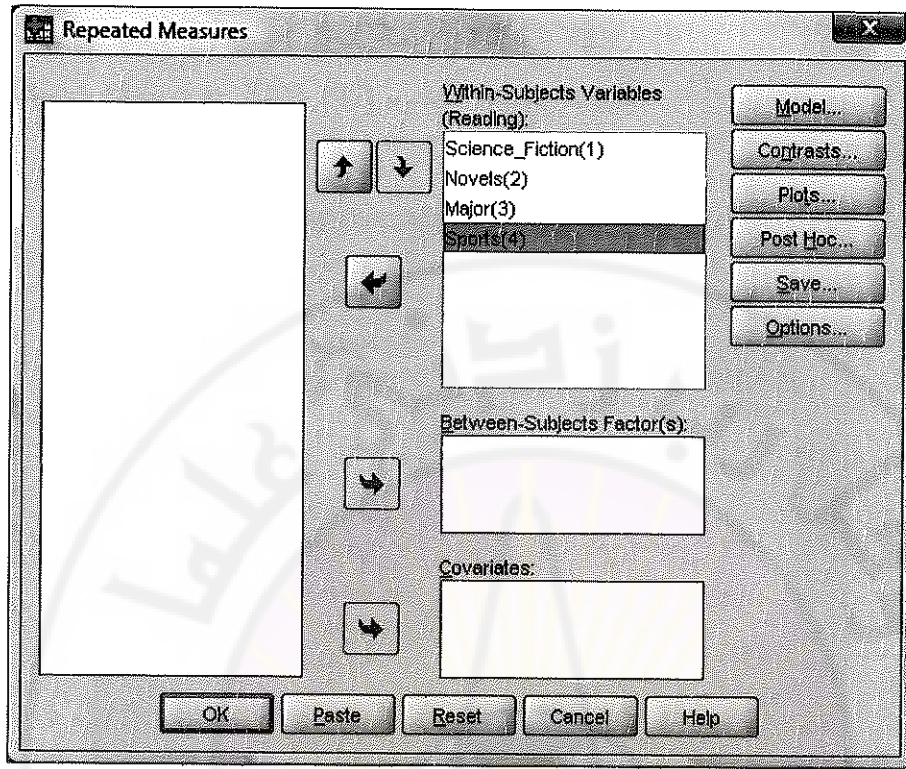


(A)

(B)

الشكل (17): مربعات حوار تحليل التباين لقياسات مترابطة.

ننقر على الأمر Define حيث يفتح مربع الحوار في الشكل (18). نستخدم مربع
 الحوار في الشكل (18) لكي نحدد الأعمدة الأربعة التي يجري عليها تحليل التباين
 للعينات المترابطة. ننقل الأعمدة الأربعة من القائمة اليسارية إلى القائمة اليمينية كي
 تصبح جاهزة للتحليل. لاحظ أننا نقلنا المتغيرات الأربعة: Sceince_Fiction ،
 Sports ، Major ، Novels إلى القائمة اليمينية في الشكل (18).

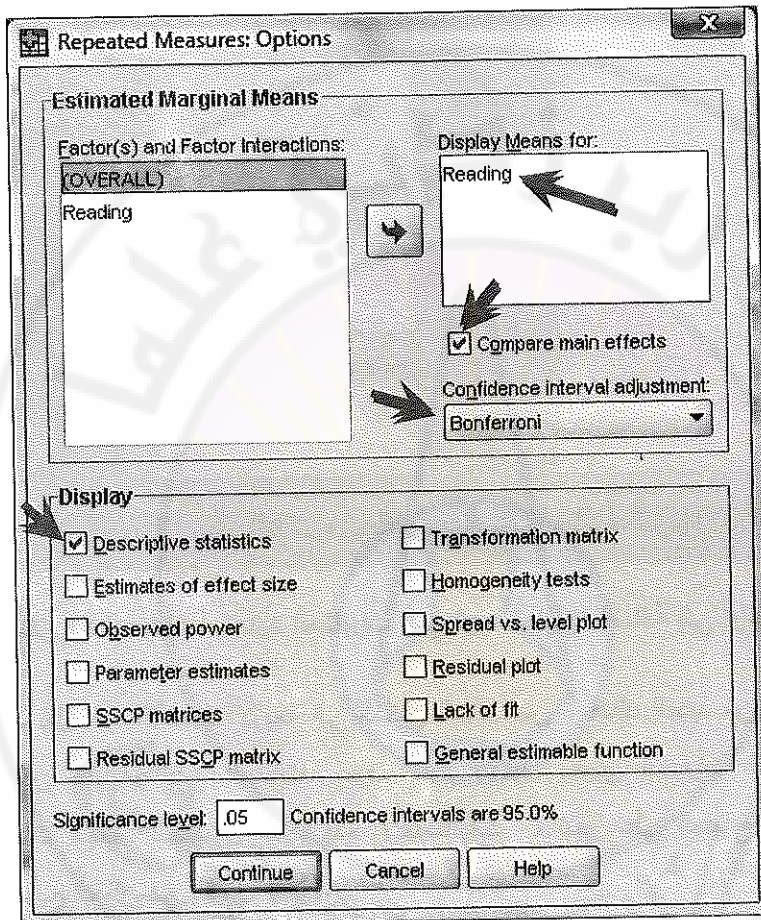


الشكل (18): شاشة تحديد المتغيرات التي نريد تحليلها إضافة إلى أوامر أخرى

بعد نقل هذه الأعمدة الأربعة إلى المكان المخصص ، هناك العديد من الإجراءات التي يمكن تنفيذها على البيانات. سوف نقوم بتفعيل بعض الخيارات التي تزودنا بإحصائيات مهمة. للحصول على هذه الإحصائيات ، نقر على المفتاح Options ليظهر مربع الحوار في الشكل (19).

ننقل اسم المتغير Reading إلى مربع الحوار Display Means for. نفعّل الأمر Compare Main effects و ذلك لإجراء مقارنات ثنائية للمتوسطات. نفتح القائمة المعنونة Confidence interval adjustment ونختار إجراء التعديل Bonferroni. وأخيراً نفعّل الأمر Descriptive statistics للحصول (إذا رغبتنا بذلك) على وصف إحصائي للعينة.

بالنتيجة ، يجب أن يظهر مربع الحوار كما في الشكل (19). تمت الإشارة إلى التحديدات المطلوبة في الشكل (19) بوساطة الأسهم للسهولة.

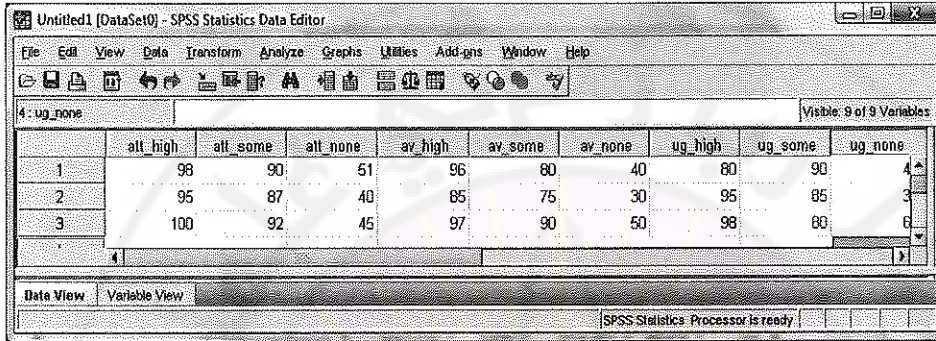


الشكل(19): مربع حوار تحديد المتوسطات المختلفة.

ننقر على المفتاح Continue في الشكل (19) ثم على المفتاح Ok في الشكل (18). يُظهر SPSS النتائج التي ناقشناها في فصل الاختبارات الوسيطة.

ثانياً تحليل التباين بمتغيرين للعينات المرتبطة

لتكن لدينا البيانات الخاصة بدراسة نوع شخصية الرجال يثير إعجاب النساء و التي درست في فصل الاختبارات الوسيطة. جزء من البيانات يظهر بشاشة SPSS التالية.



4: ug_none Visible: 9 of 9 Variables

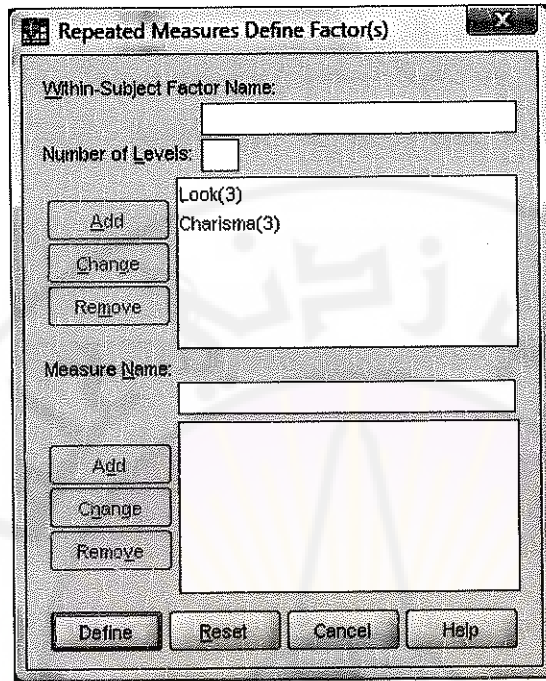
	alt_high	alt_some	alt_none	av_high	av_some	av_none	ug_high	ug_some	ug_none
1	98	90	51	96	80	40	80	90	4
2	95	87	40	85	75	30	95	85	3
3	100	92	45	97	90	50	98	80	6

Data View Variable View SPSS Statistics Processor is ready

خطوات إجراء اختبار تحليل التباين لمتغيرين هي تقريباً نفس خطوات تحليل التباين لمتغير واحد مع فوارق بسيطة. لذلك سوف نستعرض هذه الخطوات بسرعة.

في البداية يجب تعريف متغيري التجربة Look و Charisma و تحديد عدد مستويات كل واحد منهم. (تذكر أن عدد مستويات كل متغير هي ثلاثة كما رأينا في فصل الاختبارات الوسيطة.) نقوم بتفعيل اختبار تحليل التباين لعينات مرتبطة بالشكل:

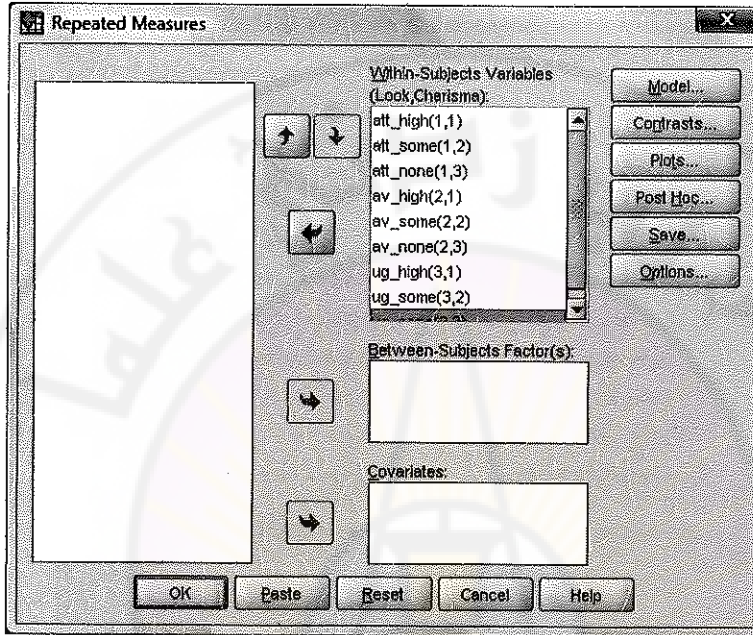
افتح القائمة **ANALYZE** ، اختر الأمر **GENERAL LINEAR MODEL** من هذه القائمة ، ثم اختر الأمر الفرعي **REPEATED MEASURES**. يفتح مربع الحوار الخاص. نحتاج لتعريف المتغيرين (Look, Charisma) و ذلك بكتابة اسم كل واحد على حدة في خانة النص Within-Subject Factor Name و من ثم نحدد عدد مستوياته. فمثلاً ، نضع المتغير Look بكتابة اسم المتغير في خانة النص Within-Subject Factor Name ، نضع عدد المستويات في خانة النص Number of Levels ، ثم نضغط المفتاح Add. نكرر ذات الطريقة من أجل المتغير Charisma. بعد تعريف هذين المتغيرين ، يظهر مربع الحوار الخاص كما في الشكل (20).



الشكل (20): مربع حوار تحليل التباين للعينات المترابطة.

ننقر على المفتاح Define لتعريف متغيرات المسألة. يفتح مربع الحوار الخاص، الشكل (21)، الذي نستطيع من خلاله تعريف متغيرات المسألة. لاحظ أننا نقلنا متغيرات المسألة إلى القائمة اليمينية كما فعلنا في حالة تحليل التباين لمتغير واحد. ننقر على المفتاح Options ليفتح مربع الحوار الخاص كما في الشكل (22). نحدد المتغيرات التي نريد عرض متوسطاتها في مربع النص Display Means for، نفعل الخيار Compare mean effects لإجراء مقارنات ثنائية، و نختار تصحيح Benferroni لبناء مجالات ثقة حول المتوسطات، وأخيراً نحدد الخيار Descriptive Statistics لعرض وصف لمتغيرات المسألة. بعد ذلك يجب أن يظهر مربع الحوار الخاص بـ Options كما في الشكل (22). لاحظ أننا نقلنا المتغيرات Look و

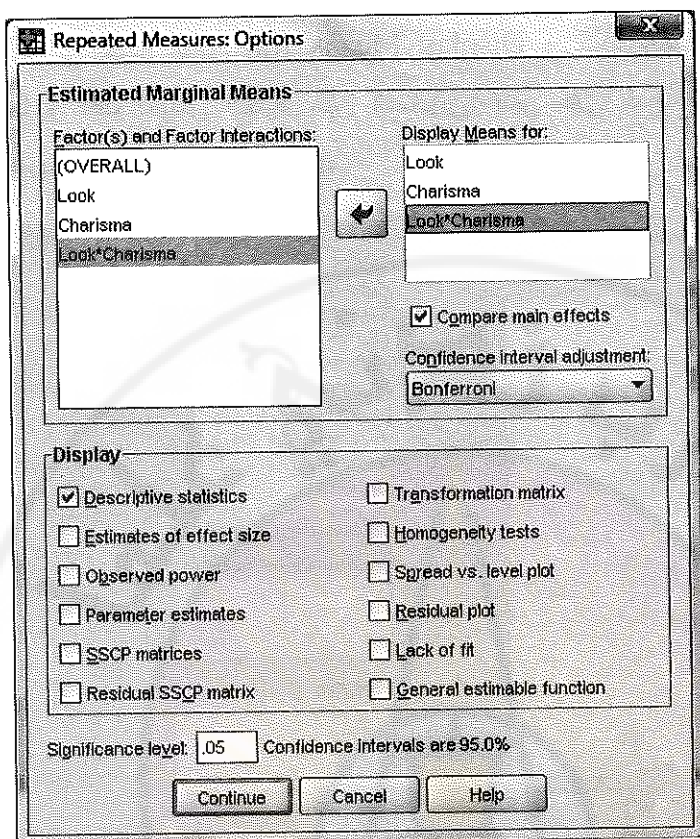
Charisma إلى خانة النص Display Means for Look * و كذلك نقلنا Charisma و ذلك لعرض التفاعل بين المتغيرين. ننقر على المفتاح Continue للعودة إلى مربع الحوار في الشكل (21).



الشكل (21): مربع حوار تعريف المتغيرات.

يمكن لنا أن نطلب من الـ SPSS إظهار مخطط يبين التفاعل بين المتغيرين Look و Charisma. لأجل ذلك ننقر على المفتاح Plots في مربع الحوار (الشكل (21)) ، ليظهر مربع الحوار الخاص في الشكل (23). ننقل المتغير Look إلى مربع النص المعنون "Horizontal Axis" والمتغير Charisma إلى مربع النص Separate Lines ، ثم ننقر المفتاح Add لإضافة المتغيرين إلى القائمة Plots. بهذه التحديدات يجب أن يظهر مربع الحوار كما في الشكل (23).

ننقر الآن على المفتاح Continue ثم ننقر على المفتاح OK في مربع الحوار (الشكل (21)). يقوم الـ SPSS بعرض النتائج. يمكن العودة إلى فصل الاختبارات الوسيطة للإطلاع على النتائج و كيفية تفسيرها.



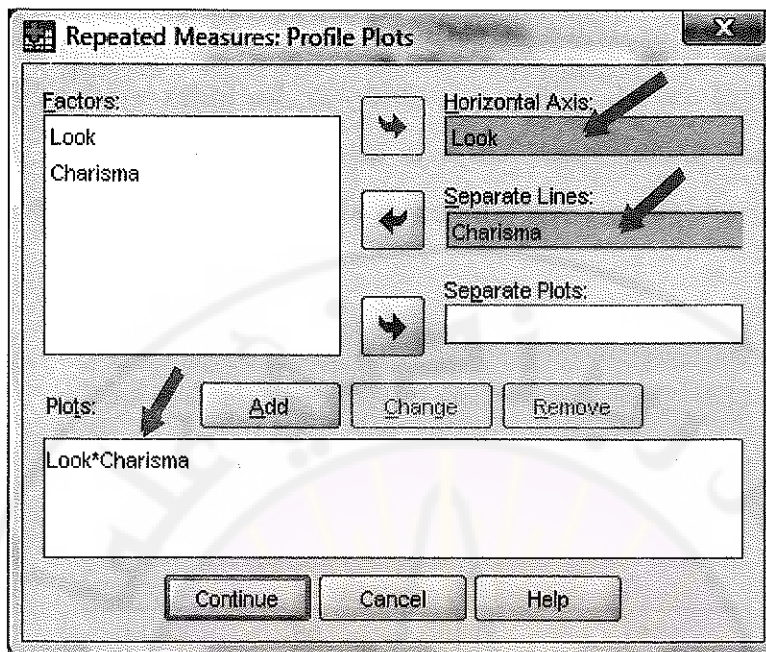
الشكل (22): مربع حوار خيارات Options.

الاختبارات غير الوسيطة Nonparametric Tests

سوف نبين في الفقرات التالية كيفية استخدام SPSS لتحليل المسائل التي طُرحت في فصل الاختبارات غير الوسيطة. لن نناقش النتائج التي تصدر عن التحليل لأن هذه المناقشة موجودة في الفصل المذكور و يمكن العودة إليها.

إجراء اختبار الاستقلالية Chi-Square Independence Test

بفرض أن لدينا البيانات في الجدول (1) التي تصف عدد الناجين و غير الناجين من حادث سفينة Titanic بحسب الجنس.

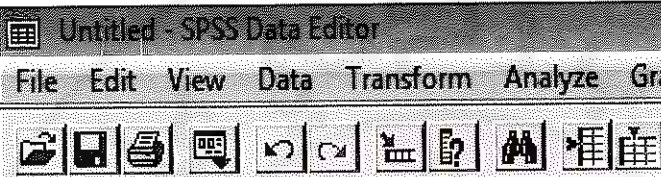


الشكل (23): بعض الاحصائيات المناسبة لتحليل التباين لمتغيرين.

		Survival	
Gender		Died	Survived
	Male	680	168
	Female	126	317

الجدول (1): بيانات عن الناجين نتيجة غرق السفينة Titanic حسب الجنس. لدراسة العلاقة بين المتغيرين Gender و Survival ، فإنه يجب أولاً إدخال هذه البيانات في شاشة بيانات الـ SPSS. لدينا متغيران: المتغير Gender و له قيمتان "1"، "2" و هاتان القيمتان تقابلان على الترتيب Female و Male والمتغير Survival و له أيضاً قيمتان "1"، "2" وهاتان القيمتان تقابلان على الترتيب

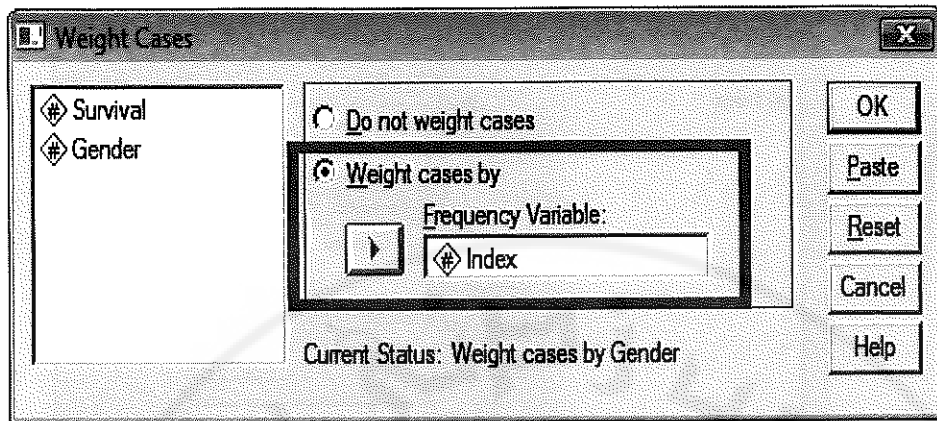
Survived و Died. وبالتالي يوجد أربعة أزواج ممكنة للمتغيرين كما هو موضح بالشكل (24). العمود Index يمثل تكرار كل زوج من القيم. مثلاً عدد تكرارات الزوج (1, 1) أي (Died, Male) هي 680 كما هو موضح بالجدول لذلك تم وضع التكرار 680 لهذا الزوج في الخلية المقابلة من العمود Index.



	Survival	Gender	Index
1	1	1	680
2	2	1	168
3	1	2	126
4	2	2	317

الشكل (24): شاشة البيانات في الـ SPSS

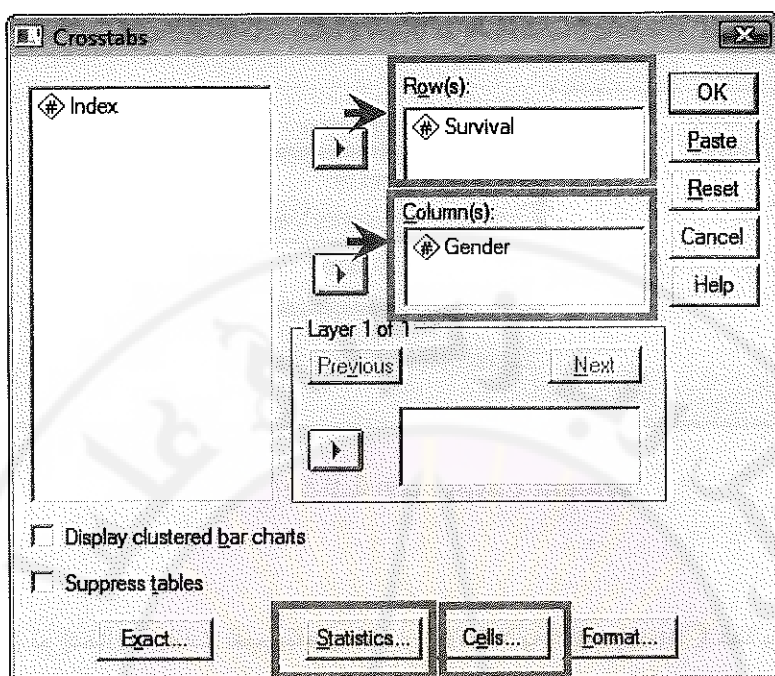
قبل البدء بإجراء تحليل الاستقلالية ، يجب أن نحدد لـ SPSS أن القيم في المتغير Index هي تكرارات الأزواج المقابلة من قيم Survival, Gender. لأجل ذلك نستخدم الأمر Weight cases من القائمة Data و التي تفتح لنا مربع الحوار في الشكل (25). نلغي تفعيل الوضع الافتراضي "Do not weight cases" و ذلك بالنقر على زر الاختيار ☒ المجاور للأمر "Weight cases by". بعد تفعيل هذا الخيار ننقل العمود Index في مربع الحوار المناسب كما يظهر بالشكل (25) أعلاه. بالنقر على OK تصبح تكرارات كل زوج محددة بالقيمة المقابلة لهذا الزوج من العمود Index.



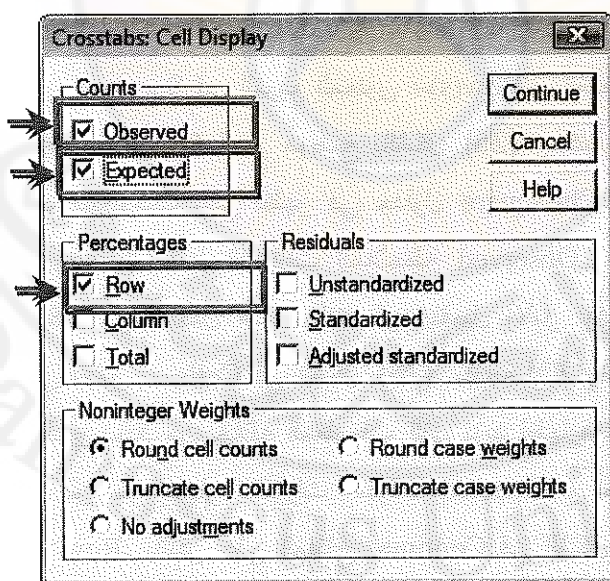
الشكل (25): شاشة الأمر Weight Cases

بعد أن تم إدخال البيانات نقوم بتفعيل اختبار الاستقلالية Chi-square كما يلي:
 أولاً: إختار **ANALYSIS : DESCRIPTIVE STATISTICS : CROSSTABS** . يظهر نتيجة لذلك مربع الحوار الخاص بهذا الأمر بالشكل (26).

ثانياً: نضع المتغير Survival في المربع المعنون بـ Row(s) و المتغير Gender في المربع المعنون بـ Column(s) كما هو موضح بالشكل (26). يمكننا أيضاً من خلال مربع الحوار تحديد المعلومات التي تظهر في جداول Chi-square. بشكل إفتراضي يُظهر اختبار Chi-square فقط التكرارات الملاحظة Observed. إذا أردنا إظهار معلومات أكثر ، كما سوف نفعل هنا ، فإننا نقوم بتفعيل الأمر Cell Display من خلال الضغط على الأمر Cells... في الشكل (26). نتيجة لذلك يظهر مربع الحوار في الشكل (27) الذي نحدد من خلاله المعلومات التي نريد إظهارها. مثلاً في الشكل (27) تم تفعيل كل من أمر إظهار التكرارات المتوقعة Expected و أمر إظهار النسب المئوية لكل سطر Row. بالنقر على المفتاح Continue في الشكل (27) نعود إلى مربع الحوار في الشكل (26).



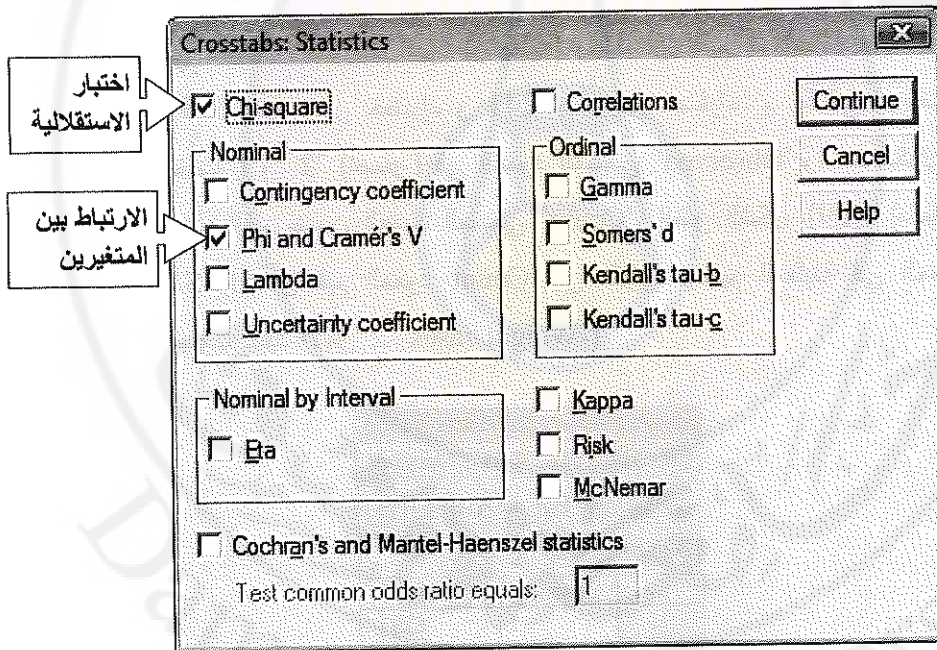
الشكل (26): الشاشة الرئيسية للأمر Crosstabs.



الشكل (27): شاشة الأمر Cells والخيارات المناسبة للمسألة التي نعالجها.

ثالثاً: نقوم الآن بتحديد نوع الاختبار الذي نريد إجراءه من خلال الأمر Statistics... في الشكل (26) حيث يظهر مربع الحوار في الشكل (28). نقوم بتفعيل اختبار الاستقلالية Chi-square الذي يقوم باختبار فيما إذا كان المتغيران مرتبطين أم مستقلين. نفعل كذلك الخيار Phi and Cramer's V الذي يفيد في قياس قوة الارتباط بين المتغيرين إن وجدت. (يتم تفعيل هذين الخيارين من خلال النقر بزر الفأرة الأيسر على خانة الاختبار ☒ المجاورة للخيار).

بالضغط على المفتاح Continue في الشكل (28) نعود إلى مربع الحوار في الشكل (26). نضغط المفتاح OK في مربع الحوار الرئيسي يقوم SPSS بإجراء الاختبار و إصدار النتائج.



الشكل (28): شاشة الأمر Statistics و الخيارات المناسبة للمسألة المُعالجة.

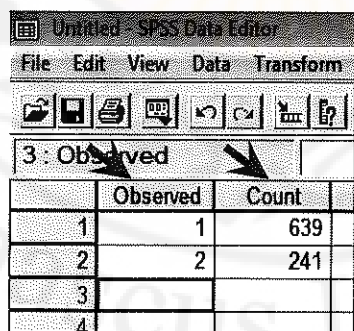
اختبار جودة التطابق Chi-square Goodness-of-fit Test

سوف نبين في هذه الفقرة كيف نستخدم SPSS لإجراء اختبار جودة التطابق. الهدف هنا التعريف بكيفية إدخال البيانات و تطبيق الاختبار. سوف نناقش حالتين أساسيتين. الحالة الأولى عندما تكون القيم المتوقعة وفق قانون معين غير متساوية. في هذه الحالة يجب حساب (يدوياً) القيم المتوقعة اعتماداً على القانون المحدد. الحالة الثانية عندما تكون القيم المتوقعة متساوية. في هذه الحالة يقوم SPSS آلياً بحساب التكرارات المتوقعة.

مثال (1)

أولاً: نقوم بإدخال البيانات في شاشة الـ SPSS.

في تجربة التصالب بين نباتات ببذور صفراء و نباتات ببذور خضراء (عُرضت هذه المسألة في فصل الاختبارات غير الوسيطة) وجدنا أن القيم الملاحظة من تجربة التصالب هي 639 نبتة ببذور خضراء و 241 نبتة ببذور صفراء. و بالتالي تم وضع الرموز "1" (أخضر) و "2" (أصفر) في العمود المسمى "Observed" و عدد تكرارات كل قيمة من هذه القيم ضمن العمود المسمى "Count". انظر الشكل (29). (بالطبع حتى يعتبر الـ SPSS أن قيم العمود Count هي تكرارات للقيم المقابلة في العمود Observed ، يجب تنفيذ الأمر ... Weight cases).

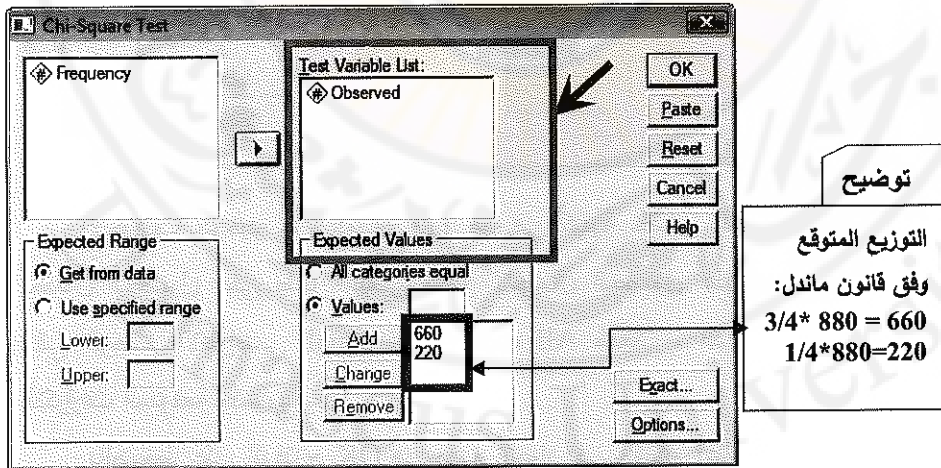


	Observed	Count
1	1	639
2	2	241
3		
4		

الشكل (29): إدخال البيانات في شاشة الـ SPSS.

ثانياً: نقوم بتفعيل اختبار جودة التطابق من خلال النقر على القائمة **ANALYZE** ،
نختار الأمر **NONPARAMETRIC TESTS** ، ثم نختار الأمر الفرعي **CHI-SQUARE** .
حيث يظهر مربع الحوار الخاص بهذا الأمر كما هو مبين بالشكل (30).

نضع في مربع النص المعنون Test Variable List: المتغير الذي يحوي التكرارات المشاهدة (Observed في مثالنا). نلغي تفعيل الوضع الافتراضي " All categories equal " الذي يعني أن جميع الاحتمالات متساوية و نفعّل بدلاً الخيار "Values" و ذلك بالنقر على خانة التفعيل ☒ المجاورة. بعدها نبدأ بحساب القيم المتوقعة يدوياً وفق قانون ماندل ، و من ثم نضع القيم المحسوبة يدوياً في مربع النص المجاور لـ Values ، و نضغط على المفتاح Add عند إضافة كل قيمة. (يبين المربع التوضيحي في الشكل (30) كيف يتم يدوياً حساب القيم المتوقعة وفق قانون ماندل. على سبيل المثال ، العدد المتوقع للنباتات ببذور خضراء 660 ناتجة عن ضرب 880 بـ $\frac{3}{4}$ ؛ نسبة النباتات ببذور خضراء حسب قانون ماندل.) لاحظ أنه من المهم التقيد بترتيب القيم المتوقعة بحيث تطابق ترتيب القيم الملاحظة و إلا نحصل على نتائج غير صحيحة.



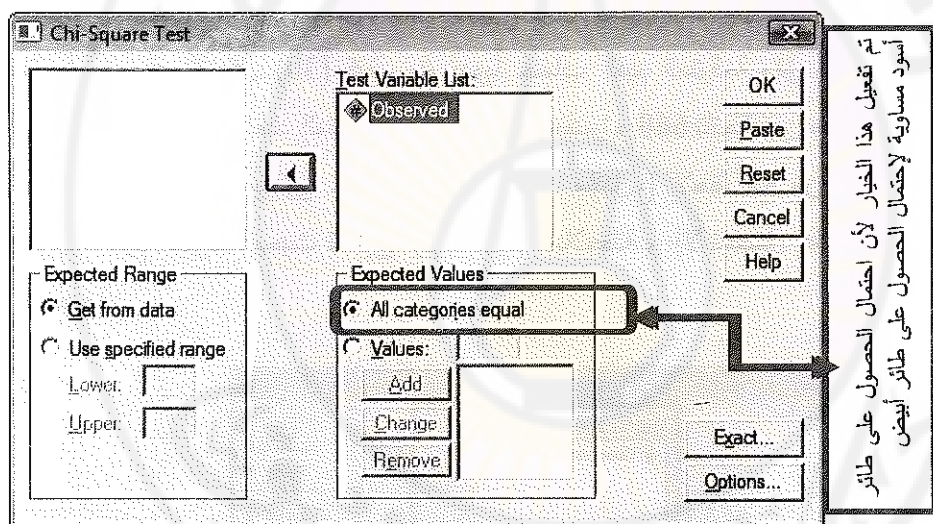
الشكل (30): مربع الحوار الخاص باختبار جودة التطابق.

بالضغط على المفتاح OK في مربع الحوار في الشكل (30) يصدر SPSS نتائج اختبار جودة التطابق.

مثال (2): تجربة التزاوج بين طيور سوداء و أخرى بيضاء.

في تجربة تزاوج بين 30 طائراً أسود مع 30 طائراً أبيض حصلنا نتيجة لهذا التزاوج على 42 طائراً أسود و 60 طائراً أبيض. حسب قوانين الوراثة عند تزاوج طائر أسود مع طائر أبيض ، فإن احتمال ظهور طائر بأحد اللونين (أسود أو أبيض) متساوي. هل تتطابق النتائج المشاهدة مع ما هو متوقع وراثياً؟

لن نتحدث هنا عن كيفية إدخال البيانات في شاشة الـ SPSS ، و لا عن كيفية تفعيل اختبار جودة التطابق لأنها مماثلة لتلك في المثال (1) أعلاه.



الاختلاف الوحيد عن المثال (1) هو في كيفية حساب القيم المتوقعة. بما أن احتمال الحصول على طائر أبيض مساوي لاحتمال الحصول على طائر أسود ، فإننا نبقى الوضع الافتراضي All categories equal مُفعلاً ، الذي يفترض أن جميع التكرارات المتوقعة متساوية.

بالضغط على المفتاح OK في مربع الحوار في الشكل أعلاه نحصل على نتائج اختبار جودة التطابق.

اختبار توافق المراقبين Observer Agreement

بفرض أن لدينا المسألة التالية التي تم عرضها في فصل الاختبارات الوسيطة.

لنفرض أنه تم فحص 100 مريض إضطراب سلوك بوساطة طبيين. و لنفرض أن كل مريض يصنف على أنه مضطرب السلوك بسبب: نفسي Psychological ، أو عصبي Neurological ، أو عضوي Organic. يبين الجدول التالي تصنيفات حالات المرضى اعتماداً على المراقب Rater A و المراقب الآخر Rater B.

هل هناك إتفاق ما بين المراقبين A و B؟

		RATER A		
RATER B		Psychological	Neurological	Organic
	Psychological	75	1	4
	Neurological	5	4	1
	Organic	0	0	10

لإجراء اختبار توافق المراقبين باستخدام SPSS ، يجب أن نضع البيانات في شاشة الـ SPSS كما هو مبين في الشكل (31). لاحظ كيف تم ترتيب قيم المتغير Rater A و قيم المتغير Rater B. المتغير Count يمثل التكرار في كل خلية من خلايا الجدول. على سبيل المثال ، تكرار الخلية (1,1) يساوي 75 ، لذلك تم وضع القيمة 75 في العمود Count ، و في السطر المقابل لهذه الخلية. (حتى يعتبر SPSS أن القيم في العمود Count هي تكرارات لقيم العمودين Rater A و Rater B ، يجب استخدام الأمر Weight cases من القائمة Data كما تعلمنا من الأمثلة السابقة.)

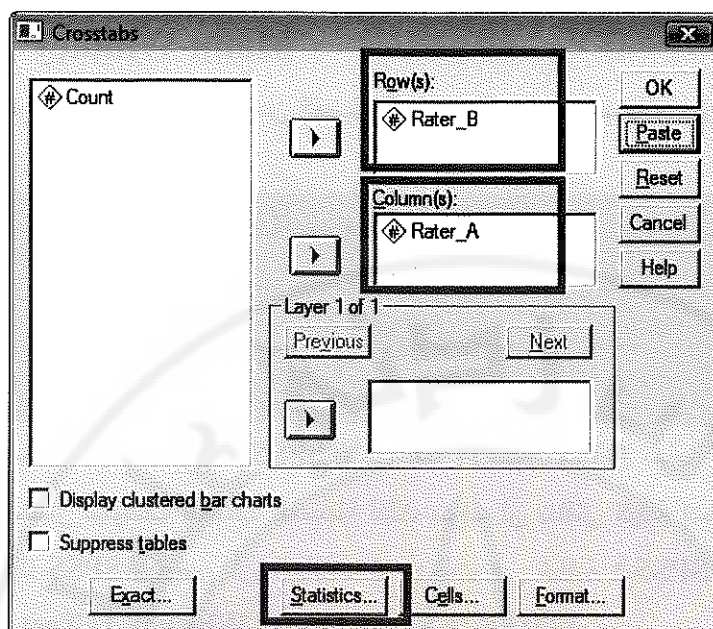
نقوم بتفعيل الاختبار Observer Agreement بالنقر على القائمة **ANALYZE** ، اختيار الأمر **DESCRIPTIVE STATISTICS** ، ثم اختيار الأمر الفرعي **CROSSTABS**. حيث يظهر مربع الحوار في الشكل (32).

في هذا المربع ، نضع المتغير Rater_B في خانة النص Rows و المتغير Rater_A في خانة النص Column(s). جدير بالذكر أن وضع Rater_A في خانة النص Rows أو Columns غير مهم ، و بالتالي لا يؤثر في النتائج. نضغط على المفتاح Statistics لتفعيل اختبار إتفاق المراقبين حيث يظهر مربع الحوار في الشكل (33). يجب تفعيل الخيار Kappa بالنقر على خانة الخيار المقابلة بحيث تصبح مُفعلة (لاحظ الإشارة ☒ بجانب الأمر Kappa كما هو مبين في الشكل). بعد الانتهاء من كل التحديدات ، يجب الضغط على المفتاح Continue للعودة إلى مربع الحوار الرئيسي.

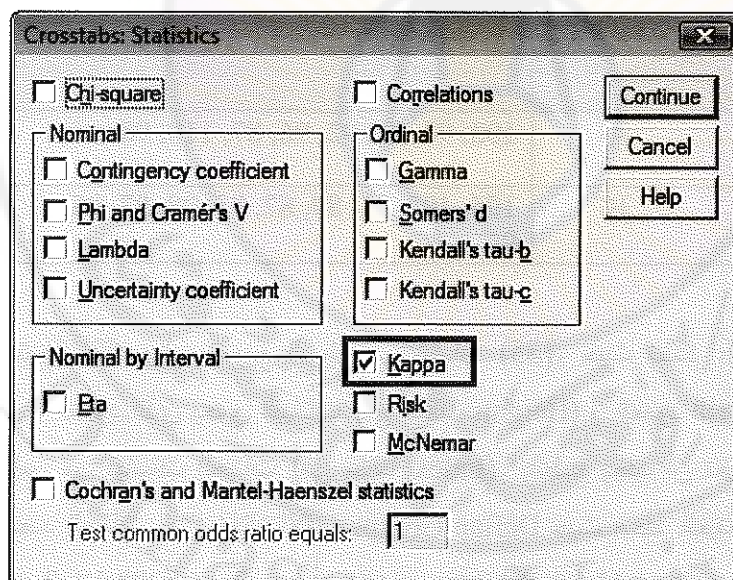
	Rater B	Rater A	Count	var	va
1	1	1	75		
2	1	2	1		
3	1	3	4		
4	2	1	5		
5	2	2	4		
6	2	3	0		
7	3	1	0		
8	3	2	1		
9	3	3	10		
10					

الشكل (31): إدخال البيانات في شاشة الـ SPSS.

بالضغط على المفتاح OK في مربع الحوار في الشكل (32) نحصل على نتائج اختبار تطابق المراقبين.



الشكل (32): مربع الحوار الخاص بالأمر Crosstabs.



الشكل (33): مربع الحوار الخاص بتحديد اختبار تطابق المراقبين.

اختبار Mann-Whitney

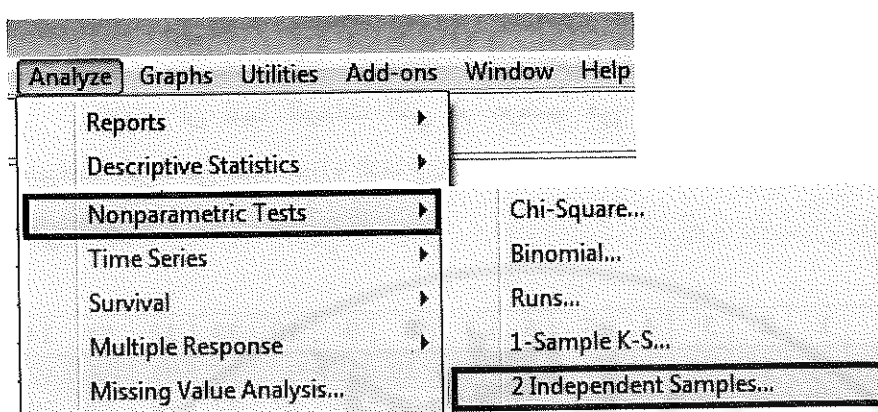
بفرض أن لدينا المسألة من الفصل السادس التي نعيد نصها و نعرض بياناتها هنا للسهولة.

لنفرض أننا نريد إجراء تجربة لتقويم فعالية نوعين من السماد الكيميائي A و B على نمو النباتات. حيث تقاس فعالية السماد الكيميائي بارتفاع النباتات — كلما زاد ارتفاع النبتة كانت فعالية السماد أفضل. يبين الجدول التالي ارتفاعات النباتات لأجل نوعي السماد A و B.

إدخال البيانات في الـ SPSS لأجل هذا الاختبار مطابق تماماً لإدخال البيانات لأجل اختبار ستودنت للعينات المستقلة لذلك سوف لن نناقش كيفية إدخال البيانات في شاشة الـ SPSS. بفرض أن البيانات قد أدخلت في شاشة الـ SPSS ، يمكن تفعيل اختبار Mann-Whitney U Test كما في الشكل (34).

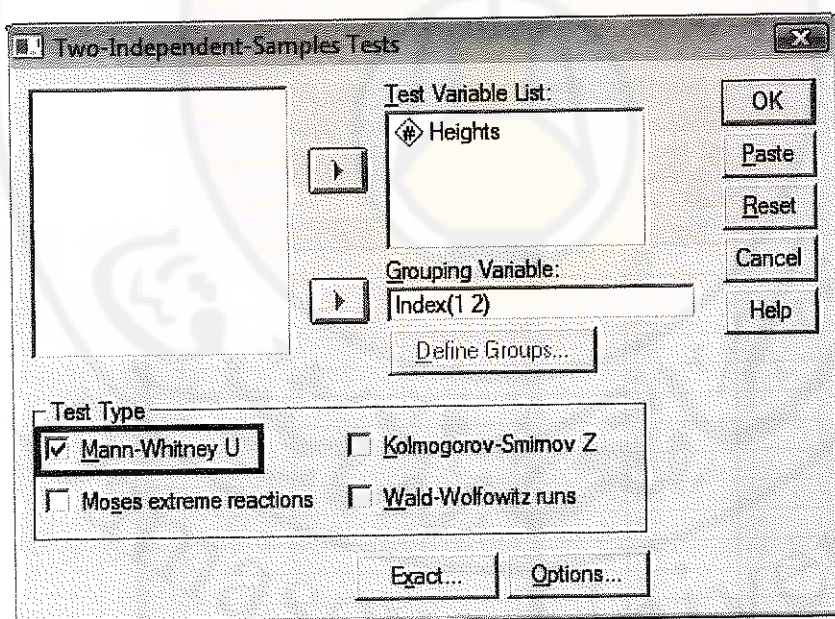
ارتفاع النباتات (سم)	نوع السماد الكيميائي المستخدم في التجربة
100, 90, 82, 92, 96, 87, 110, 98	A
105, 102, 91, 89, 87, 83, 98, 97, 96, 95	B

نتيجة لذلك يظهر مربع الحوار في الشكل (35). حيث نحدد ، كما هو مبين بالشكل ، العمود الذي يحوي البيانات (في مثالنا Heights) و العمود الذي يحدد مرجعية كل قياس من قياسات العينات (في مثالنا Index). بالطبع يجب أن نعرف أيضاً القيم التي تحدد العينات. من خلال الأمر Define Groups حيث مُحدد العينة الأولى هو 1 و مُحدد العينة الثانية هو 2. بما أن اختبار Mann-Whitney هو الوضع الافتراضي ، فإننا نكون قد أنهينا جميع التحديدات اللازمة.



الشكل (34): تفعيل اختبار Mann-Whitney.

بالضغط على المفتاح OK في مربع الحوار في الشكل (35) نحصل على نتائج اختبار تحليل Mann-Whitney.



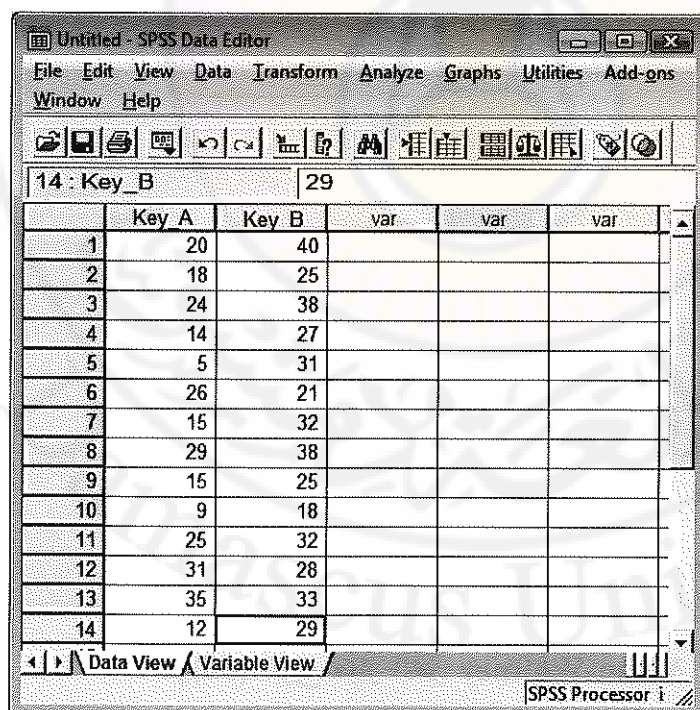
الشكل (35): مربع حوار الأمر Mann-Whitney.

اختبار الإشارة Wilcoxon Signed Ranked Test

بفرض أن لدينا البيانات التي تم عرضها في الفصل السادس التي نقيس تفضيلات القردة في تنبيه أجزاء الدماغ A و B. نعيد عرض هذه البيانات في الجدول التالي للسهولة.

القرد	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Key A	20	18	24	14	5	26	15	29	15	9	25	31	35	12
Key B	40	25	38	27	31	21	32	38	25	18	32	28	33	29

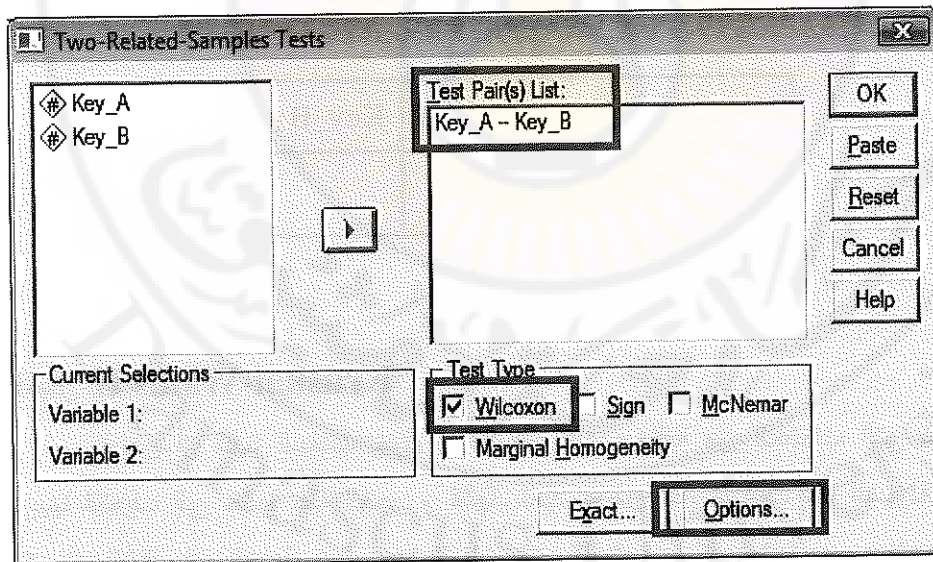
لكي نعالج البيانات في الجدول أعلاه فإنه يجب تفريغها في شاشة البيانات SPSS كما يظهر في الشكل أعلاه. كما تلاحظ تم تفريغ البيانات في عمودين منفصلين كما يتطلب اختبار الإشارة. و قد تم إعطاء أسماء للأعمدة مماثلة للأسماء في الجدول Key_A, Key_B. بهذا الشكل تكون البيانات جاهزة للمعالجة.



يتم تفعيل اختبار الإشارة Wilcoxon من خلال اختيار القائمة Analyze من شريط القوائم و اختيار الأمر الرئيسي Nonparametric Tests... من القائمة الفرعية نختار الأمر "2 Related Samples...". نتيجةً لذلك ، يفتح مربع الحوار الخاص بهذا الأمر و الذي يظهر في الشكل أدناه.

التعامل مع مربع الحوار الخاص باختبار Wilcoxon بسيط للغاية. أولاً ننقل العمودين Key_A و Key_B من القائمة اليسارية إلى القائمة اليمينية كما هو موضح بالشكل. بما أن اختبار Wilcoxon هو الاختبار الافتراضي فإننا لا نحتاج لإعادة تفعيله. (سوف تلاحظ عند فتح مربع حوار Wilcoxon أن علامة التفعيل ☒ موجودة آلياً بجانب اسم الاختبار.)

من المفضل عادةً أن نقوم بوصف العينة. إذا أردنا ذلك (كما سوف نفعل هنا) فما علينا إلا أن ننقر على المفتاح Options و من ثم نفعل الأمر Descriptive ذلك من خلال مربع الحوار الذي سيظهر نتيجة لذلك.



الخرج لهذا الأمر يظهر من خلال الجداول التالية:

وصف العينة Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
Key_A	14	19.86	8.787	5	35
Key_B	14	29.79	6.411	18	40

الرتب Ranks

	N	Mean Rank	Sum of Ranks
Key_B - Key_A Negative Ranks	3 ^(a)	2.00	6.00
Key_B - Key_A Positive Ranks	11 ^(b)	9.00	99.00
Key_B - Key_A Ties	0 ^(c)		
Key_B - Key_A Total	14		

a Key_B < Key_A

b Key_B > Key_A

c Key_B = Key_A

Wilcoxon Test Statistics^(b) إحصائيات اختبار

	Key_B - Key_A
Z	-2.921(a)
Asymp. Sig. (2-tailed)	.003

a Based on negative ranks.

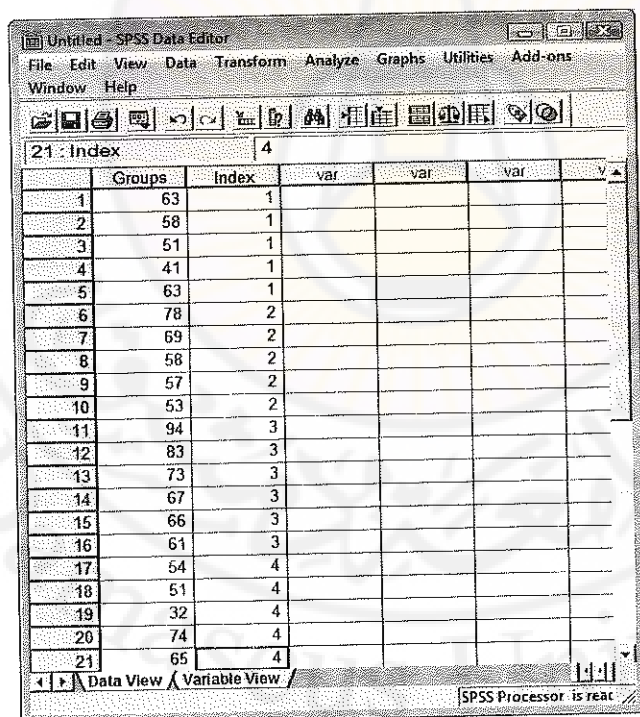
b Wilcoxon Signed Ranks Test

اختبار Kruskal-Wallis H

بفرض أن لدينا البيانات حول أداء مجموعات مختلفة من الحشرات التي تم تدريبها للتعرف إلى حشرة عثة الغجر و قتلها. نعيد عرض هذه البيانات هنا للسهولة. تذكر أن الأرقام في الجدول تمثل عدد المرات التي قامت فيها كل حشرة بهجوم ناجح على عثة الغجر.

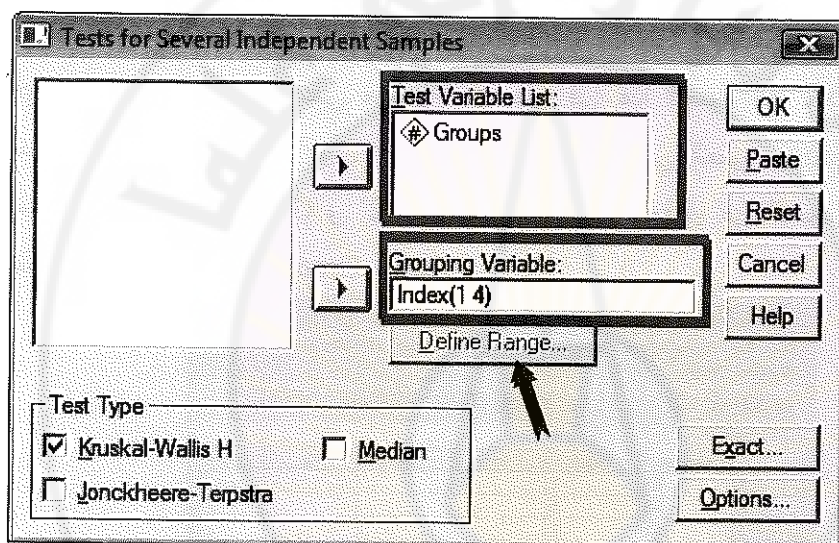
Group 1	Group 2	Group 3	Group 4
63	78	94	54
58	69	83	51
51	58	73	32
41	57	67	74
63	53	66	65
		61	80

يبين الشكل (32) شاشة SPSS بعد تفريغ البيانات فيها. كما تلاحظ لقد تم وضع العينات الأربعة Group 1, Group 2, Group 3, Group 4 في عمود واحد اسمه Groups و تمت الدلالة على مرجعية كل بيان في هذا العمود من خلال العمود المرجعي Index حيث "1" يدل على أن البيان المقابل في العمود Groups ينتمي للمجموعة الأولى Group 1 و "2" يدل على أن البيان المقابل في العمود Groups ينتمي للمجموعة الثانية Group 2 وهكذا.



الشكل (32): شاشة بيانات SPSS بعد تفريغ البيانات فيها.

نقوم بتفعيل اختبار Kruskal-Wallis من خلال القائمة Analyze واختيار الأمر الرئيسي Nonparametric Tests... و من ثم اختيار الأمر الفرعي K Independent Samples. لن نعيد هنا كيف يتم ذلك بشكل مباشر لأنه من المفترض أن يكون القارئ قد أصبح معتاداً على ذلك. نتيجة لذلك يظهر مربع الحوار الخاص بهذا الأمر في الشكل (33).



الشكل (33): مربع الحوار الخاص باختبار عدة عينات مستقلة.

في مربع الحوار في الشكل (33) ، نضع المتغير الذي يحوي القيم الفعلية للعينات Groups في مربع النص Test Variable List ، و المتغير المرجعي Index في مربع النص Grouping Variable. أيضاً من الضروري تحديد مجال القيم المرجعية في العمود Index. بما أن مجال القيم المرجعية من 1 حتى 4 ، لذلك تم تحديد هذا المجال في مربع الحوار من خلال أمر تعريف مجال القيم المرجعية Define Range المشار إليه بالسهم في الشكل (33). لاحظ في مربع الحوار أعلاه أن اختبار Kruskal-

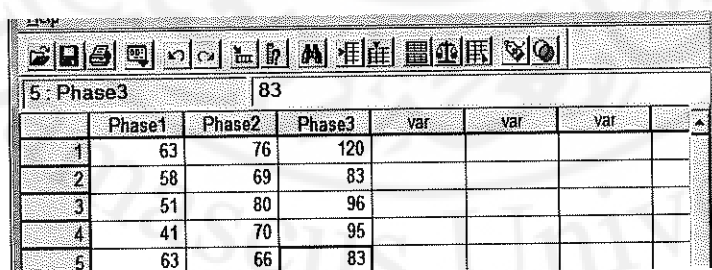
Wallis H مفعّل آلياً لذلك لا حاجة لإعادة تفعيله. بالضغط على المفتاح OK نحصل على خرج الـ SPSS لنتائج الاختبار.

اختبار Friedman

بفرض أن لدينا العينة المطروحة في فصل الاختبارات غير الوسيطية التي نعيد عرضها مرة أخرى في الجدول التالي للسهولة.

Phase 1	Phase 2	Phase 3
63	76	120
58	69	83
51	80	96
41	70	95
63	66	83

لكي نعالج هذه البيانات باستخدام الـ SPSS ، يجب أن ندخل البيانات إلى شاشة البيانات بشكل يناسب اختبار Friedman. إدخال هذه البيانات يتم كما هو موضح في الشكل (34). حيث عرفنا ثلاثة أعمدة Phase1, Phase2, Phase3 وأدخلنا قيم كل عينة من الجدول في عمود خاص. الآن أصبحت البيانات جاهزة للتحليل بواسطة اختبار Friedman.

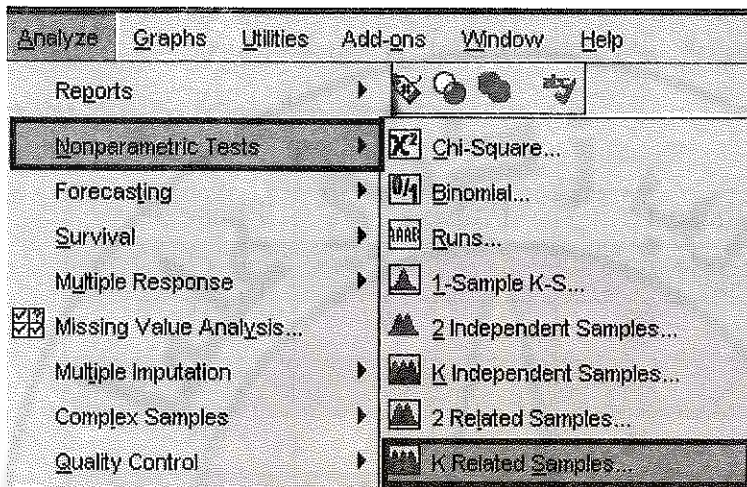


The screenshot shows the SPSS 'Data View' window for a file named '5 : Phase3'. The window has a toolbar at the top with various icons. Below the toolbar, there is a table with 5 rows and 7 columns. The first three columns are labeled 'Phase1', 'Phase2', and 'Phase3'. The next three columns are labeled 'var', 'var', and 'var'. The data entered in the first three columns matches the data from the table in the previous block. The fifth row is highlighted in blue.

	Phase1	Phase2	Phase3	var	var	var
1	63	76	120			
2	58	69	83			
3	51	80	96			
4	41	70	95			
5	63	66	83			

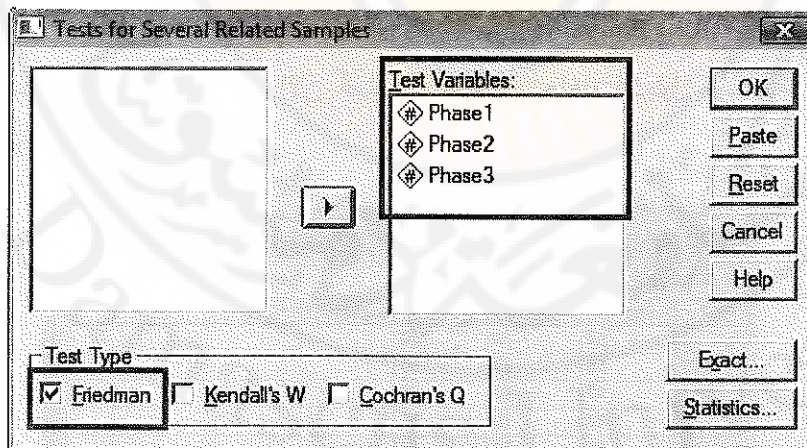
الشكل (34): تفريغ البيانات في شاشة البيانات (SPSS)

يتم تفعيل اختبار Friedman كما هو موضح بالشكل (35). نختار القائمة Analyze من شريط القوائم ، ثم نختار الأمر الرئيسي Nonparametric Tests ... ، ومن ثم من القائمة الفرعية نختار الأمر K Related Samples



الشكل (35): تفعيل اختبار Friedman.

نتيجة لذلك يظهر مربع الحوار الخاص بهذا الأمر كما يبدو في الشكل (36).



الشكل (36): مربع حوار اختبار Friedman

كما تلاحظ فإنه تم نقل المتغيرات الثلاثة Phase1, Phase2, Phase3 إلى القائمة المعنونة بـ Test Variables (بالطرق التي تعلمناها مسبقاً) لكي تصبح جاهزة للمعالجة. بما أن اختبار Friedman مُفعّل آلياً ، فلا توجد حاجة لإعادة تفعيله. بالضغط على المفتاح OK نحصل على النتائج المعروضة في الفصل السادس. (يمكن للقارئ العودة إلى الفصل السادس للتعرف إلى النتائج و كيفية تفسيرها.)

مثال آخر على اختبار Friedman

البيانات التالية تتعلق بكمية المواد الغذائية التي يستهلكها 8 جردان Rats بعد حرمانهم من الغذاء لمدة 0 ساعة ، 24 ساعة ، و 72 ساعة.⁵

كمية الغذاء المستهلك من بعد 72 ساعة حرمان الطعام	كمية الغذاء المستهلك بعد 24 ساعة حرمان من الطعام	كمية الغذاء المستهلك بعد صفر ساعة حرمان من الطعام
13.9	5.9	3.5
12.6	8.1	3.7
8.1	8.1	1.6
6.8	8.6	2.5
14.3	8.1	2.8
4.2	5.9	2.0
14.5	9.5	5.9
7.9	7.9	2.5

⁵ هذه البيانات مأخوذة من بحث علمي بعنوان:

“The Relation Between Differences in Level of Food Deprivation and Dominance in Food Getting in the Rat”

هل تشير هذه البيانات إلى تغير حقيقي في حجم استهلاك الطعام لدى الجرذان؟
 لن نعرض كيف يتم إدخال البيانات شاشة SPSS و لا كيف نعمل اختبار
 Friedman ، لأننا قمنا بالتعرف إلى ذلك في الأمثلة السابقة. سوف نركز هنا فقط على
 عرض النتائج والتعليق عليها بشكل مختصر. يبين الشكل التالي وصفاً للعينات. لاحظ
 من الشكل أن متوسط استهلاك الجرذ من الطعام يزداد كلما ازدادت فترة حرمان الجرذ
 من الطعام. على سبيل المثال ، متوسط استهلاك الجرذ بدون حرمان (0 HOURS)
 يساوي 3.0625 بينما بعد 72 ساعة حرمان فإن المتوسط يرتفع إلى 10.2875.

وصف العينة Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
0 HOURS	8	3.06 25	1.34264	1.60	5.90
24 HOURS	8	7.76 25	1.25463	5.90	9.50
72 HOURS	8	10.2 875	3.99908	4.20	14.50

يبين الشكل التالي متوسطات رتب العينات الثلاثة. نلاحظ أن متوسط رتب كمية
 إستهلاك الطعام تزداد بزيادة ساعات الحرمان. مما قد يشير إلى وجود فرق معنوي في
 متوسط كمية استهلاك الطعام.

الرتب Ranks

	Mean Rank
0 HOURS	1.00
24 HOURS	2.38
72 HOURS	2.63

يبين الجدول التالي نتائج اختبار Friedman حول متوسطات العينات الثلاث. كما هو ملاحظ فإن الاختبار معنوي لأن معنوية الاختبار $\text{Asymp Sig.} = 0.001$ و هي أصغر من مستوى الأهمية المفروض 5%. بناءً على ذلك نرفض فرضية العدم التي تنص على أن المتوسطات متساوية و نقبل الفرضية البديلة التي تنص على أن واحداً على الأقل من المتوسطات مختلف.

إحصائيات اختبار Friedman
Test Statistics^(a)

N	8
Chi-Square	13.067
Df	2
Asymp. Sig.	.001

a. Friedman Test

★ ★ ★ ★ ★ ★ ★ ★ ★ ★ ★ ★ ★ ★ ★ ★



هذا الملحق مخصص لعرض بيانات مأخوذة من دراسات علمية كبيرة. لذلك من الأفضل أن يستوعب القارئ كيفية تنظيم هذه البيانات ، و إجراء دراسات استكشافية عليها (مثل الوصف البياني) و إجراء كل التحليلات الإحصائية المناسبة. و أيضاً من المهم كتابة تقارير علمية تلخص بشكل كافٍ النتائج التي تم التوصل إليها بعيداً عن استخدام المصطلحات التقنية.

مجموعة البيانات الأولى: بيانات حول الداء السكري.

تعريف اسماء المتغيرات الواردة في الجدول

- **Sex:** جنس المريض أنثى (Female) أو ذكر (Male)
- **Type 1, Type 2** السكر من النوع الأول و النوع الثاني.
- **Age:** عمر المريض بالسنوات.
- **Smoking:** هل المريض مدخن (Yes) أو غير مدخن (No).
- **Hypertension:** هل لدى المريض ارتفاع الضغط — (Yes) لديه و (No) ليس لديه ضغط دم.
- **Stress:** هل يعاني المريض من التوتر — (Yes) يعاني و (No) لايعاني.
- **Mother, Father:** هل لدى الأم أو الأب مرض السكر سواء النوع الأول أو الثاني. القيم (Yes) تعني وجود مرض السكر و (No) تعني عدم وجود هذا المرض.
- **Blood Glucose:** مستوى سكر الدم.

Sex	Type 1	Type 2	Age	Smoking	Hypertension	Stress	Father	Mother	Blood Glucose
Female	0	Yes	55	No	No	No	Yes	Yes	154
Male	0	Yes	52	Yes	No	No	No	No	437
Female	0	Yes	64	No	Yes	No	No	No	264
Female	0	Yes	63	No	Yes	No	No	No	185
Female	1	No	16	No	No	No	No	No	316
Female	0	Yes	67	No	Yes	No	No	No	141
Female	0	Yes	45	No	Yes	No	No	No	231
Male	0	Yes	54	Yes	No	No	No	Yes	260
Female	0	Yes	53	No	Yes	No	No	No	236
Female	0	Yes	45	No	Yes	No	No	No	184
Male	0	Yes	56	No	No	No	No	No	314
Female	0	Yes	53	No	No	Yes	No	No	296
Female	0	Yes	54	No	No	No	No	No	203
Female	0	Yes	49	No	No	No	No	Yes	346
Male	0	Yes	36	No	Yes	No	No	Yes	167
Male	0	Yes	45	Yes	No	No	No	No	519

Female	0	Yes	42	No	Yes	No	No	Yes	200
Female	0	Yes	82	No	No	No	No	No	193
Female	0	Yes	51	No	Yes	No	No	No	236
Male	0	yes	62	No	Yes	No	No	No	171
Male	1	No	28	Yes	No	No	No	Yes	207
Male	0	Yes	50	Yes	No	No	No	No	160
Male	0	Yes	48	No	No	No	Yes	No	264
Female	0	Yes	42	No	No	No	Yes	No	218
Female	1	No	8	No	No	No	No	No	330
Male	0	Yes	67	Yes	No	No	No	No	238
Male	1	No	13	No	No	Yes	Yes	No	282
Female	0	Yes	58	Yes	No	No	Yes	No	209
Male	0	Yes	64	No	Yes	No	No	No	297
Male	1	No	13	No	No	No	No	No	333
Female	0	Yes	54	No	Yes	No	No	No	215
Female	0	Yes	45	No	No	No	No	No	225
Female	0	Yes	47	Yes	Yes	No	Yes	No	202
Male	0	Yes	41	No	No	No	No	No	202
Male	0	Yes	50	No	No	Yes	No	No	178

Male	1	No	32	Yes	No	No	No	No	336
Female	0	Yes	51	No	No	No	No	Yes	375
Female	0	Yes	52	No	Yes	No	No	No	133
Female	1	No	35	No	No	No	Yes	No	222
Female	0	Yes	28	No	No	Yes	Yes	No	188
Female	0	Yes	57	No	Yes	No	No	No	253
Male	0	Yes	62	Yes	No	No	No	No	140
Female	0	Yes	53	No	No	No	No	Yes	206
Male	0	Yes	49	Yes	No	No	No	Yes	277
Male	1	No	29	No	No	No	No	No	200
Female	0	Yes	45	No	Yes	No	No	No	249
Female	0	Yes	50	No	Yes	No	No	Yes	295
Male	0	Yes	40	No	Yes	No	No	Yes	289
Male	0	Yes	28	Yes	Yes	No	No	Yes	138
Female	0	Yes	54	No	No	No	No	No	151
Female	0	Yes	49	No	No	No	No	No	232
Male	0	Yes	49	No	Yes	No	No	Yes	32
Female	0	Yes	44	Yes	No	No	No	No	51
Male	0	Yes	45	Yes	Yes	No	No	No	554

Female	0	Yes	58	No	No	No	No	Yes	135
Male	0	Yes	51	No	No	No	No	Yes	408
Male	0	Yes	37	Yes	No	No	Yes	No	227
Female	0	Yes	44	Yes	No	No	Yes	Yes	141
Female	0	Yes	50	Yes	Yes	No	No	No	190
Female	0	Yes	58	No	Yes	No	No	No	284
Male	0	Yes	63	No	No	No	No	No	142
Male	0	Yes	53	No	Yes	No	No	No	183
Female	0	Yes	70	No	Yes	No	No	Yes	121
Female	0	Yes	49	No	No	No	No	Yes	259
Female	0	Yes	33	No	No	No	No	No	241
Female	0	Yes	34	No	No	No	No	No	250
Male	0	Yes	60	No	Yes	No	No	No	303
Female	0	Yes	55	No	Yes	No	Yes	No	170
Male	0	Yes	67	No	Yes	No	No	Yes	382
Male	0	Yes	55	No	Yes	No	No	No	240
Male	0	Yes	60	Yes	No	Yes	No	Yes	270
Female	0	Yes	45	No	Yes	Yes	No	No	153
Female	0	Yes	62	No	Yes	No	No	No	260

Female	0	Yes	41	No	No	No	No	No	252
Male	0	Yes	42	No	No	No	No	No	170
Female	0	Yes	47	No	No	No	No	Yes	164
Male	0	Yes	52	No	No	No	No	No	147
Female	0	Yes	61	No	No	No	No	No	343
Male	0	Yes	70	No	Yes	No	No	No	229
Female	0	Yes	41	No	Yes	No	No	Yes	233
Male	0	Yes	43	No	Yes	No	No	Yes	161
Male	0	Yes	36	Yes	No	No	No	Yes	239
Male	0	Yes	49	No	Yes	No	Yes	No	202
Female	0	Yes	59	No	No	Yes	No	No	267
Male	0	Yes	53	No	No	No	No	Yes	313
Female	0	Yes	73	Yes	Yes	No	No	Yes	173
Male	0	Yes	45	No	No	No	No	No	273
Female	0	Yes	44	No	No	No	No	No	186
Female	1	No	13	No	No	No	No	No	402
Male	0	Yes	57	Yes	Yes	Yes	No	No	166
Male	0	Yes	48	Yes	Yes	No	No	No	432
Female	0	Yes	49	No	No	Yes	No	Yes	331

Male	0	Yes	48	Yes	No	Yes	No	No	298
Female	0	Yes	45	Yes	No	Yes	No	Yes	280
Male	0	Yes	50	No	Yes	Yes	Yes	No	221
Female	0	Yes	60	No	No	Yes	No	Yes	299
Male	0	Yes	43	Yes	Yes	No	Yes	Yes	211
Male	0	Yes	72	No	Yes	Yes	No	No	251
Male	0	Yes	37	No	No	Yes	No	Yes	119
Female	0	Yes	40	Yes	No	Yes	No	No	223
Male	0	Yes	48	Yes	No	Yes	No	No	276
Male	0	Yes	57	No	No	Yes	No	Yes	181
Male	0	Yes	45	No	No	Yes	No	Yes	340
Female	0	Yes	53	No	Yes	Yes	No	No	223
Female	0	Yes	43	No	Yes	Yes	No	Yes	180
Male	0	Yes	59	No	No	No	No	Yes	166
Female	0	Yes	52	No	Yes	Yes	No	Yes	135
Male	0	Yes	61	No	Yes	Yes	No	No	188
Male	0	Yes	53	No	No	Yes	No	No	205
Male	0	Yes	43	Yes	No	Yes	No	Yes	287
Male	0	Yes	54	Yes	No	Yes	No	No	207

Female	0	Yes	52	No	Yes	Yes	No	No	390
Female	0	Yes	52	Yes	Yes	No	No	No	274
Male	0	Yes	63	Yes	No	Yes	No	No	88
Female	0	Yes	39	No	Yes	Yes	Yes	Yes	161
Male	0	Yes	47	Yes	Yes	Yes	Yes	Yes	222
Female	0	Yes	65	No	Yes	Yes	No	No	95
Female	0	Yes	52	No	Yes	Yes	No	No	370
Male	0	Yes	45	No	No	No	Yes	No	123
Male	0	Yes	53	Yes	No	No	No	No	102
Female	0	Yes	49	Yes	No	No	No	No	98
Female	0	Yes	50	No	No	Yes	No	No	147
Female	0	Yes	48	No	Yes	Yes	No	No	336
Male	0	Yes	79	No	No	No	No	No	113
Female	0	Yes	69	No	No	No	Yes	Yes	107
Female	0	Yes	43	No	No	No	No	Yes	104
Female	0	Yes	45	No	No	No	No	Yes	255
Male	0	Yes	62	No	No	No	No	No	280
Female	0	Yes	63	No	No	No	No	Yes	245
Female	0	Yes	43	No	Yes	No	No	No	376

Male	0	Yes	63	Yes	Yes	No	Yes	No	77
Female	0	Yes	64	No	Yes	No	Yes	No	418
Female	0	Yes	64	No	Yes	No	No	No	101
Female	0	Yes	49	No	No	No	No	No	163
Female	0	Yes	56	No	No	No	Yes	No	163
Male	0	Yes	64	Yes	No	No	No	Yes	299
Female	0	Yes	51	No	Yes	No	No	Yes	231
Female	0	Yes	47	Yes	No	No	No	Yes	298
Female	0	Yes	74	Yes	Yes	No	No	No	127
Male	0	Yes	62	No	No	No	No	No	191
Male	0	Yes	61	No	Yes	No	No	Yes	95
Female	0	Yes	63	No	No	No	No	No	299
Female	0	Yes	69	No	No	No	No	No	224
Female	0	Yes	43	No	Yes	No	No	Yes	109
Female	0	Yes	44	No	Yes	No	No	Yes	170
Male	1	No	19	No	No	No	Yes	Yes	280
Female	0	Yes	65	No	Yes	No	No	Yes	231
Female	0	Yes	61	No	No	No	Yes	Yes	75
Female	1		19	Yes	No	Yes	No	Yes	327

Male	0	Yes	62	Yes	Yes	No	No	No	147
Female	0	Yes	47	No	Yes	No	No	Yes	156
Male	0	Yes	38	Yes	No	No	No	Yes	150
Female	0	Yes	42	Yes	Yes	No	Yes	No	129
Male	0	Yes	31	Yes	Yes	No	No	Yes	233
Female	0	Yes	52	No	No	No	Yes	No	161
Female	0	Yes	50	No	No	No	No	No	143
Male	0	Yes	63	No	No	Yes	No	Yes	204
Male	0	Yes	48	Yes	No	No	No	Yes	209
Male	0	Yes	49	No	Yes	No	Yes	Yes	150
Female	0	Yes	53	No	No	No	No	Yes	87
Female	0	Yes	52	Yes	Yes	No	No	Yes	192
Female	0	Yes	40	No	Yes	No	No	No	240
Female	0	Yes	51	Yes	Yes	No	Yes	No	210
Male	0	Yes	71	No	Yes	No	No	No	83
Male	0	Yes	43	No	No	No	No	No	221
Male	0	Yes	33	No	Yes	No	No	No	175
Female	0	Yes	50	Yes	Yes	No	No	No	438
Male	0	Yes	50	Yes	Yes	No	No	Yes	219

Female	0	Yes	41	Yes	No	No	No	Yes	154
Male	0	Yes	52	Yes	No	No	No	No	112
Female	0	Yes	48	No	No	No	Yes	Yes	116
Female	0	Yes	58	No	Yes	No	No	No	265
Male	0	Yes	45	Yes	No	No	No	No	326
Female	0	Yes	43	No	No	No	No	No	172
Female	0	Yes	43	No	No	No	Yes	No	130
Male	0	Yes	63	No	Yes	No	No	No	346
Male	0	Yes	52	No	No	No	No	Yes	198
Male	0	Yes	47	No	No	No	Yes	No	308
Male	1	No	6	Yes	No	No	No	No	428
Female	0	Yes	36	No	No	No	No	No	215
Male	1	No	35	Yes	No	No	Yes	No	67
Male	0	Yes	57	No	No	No	No	No	124
Female	0	Yes	57	No	No	No	No	Yes	112
Male	0	Yes	70	Yes	No	No	No	No	221
Male	0	Yes	41	No	No	No	No	No	232
Male	0	Yes	49	No	No	No	No	No	217
Male	0	Yes	40	Yes	No	No	Yes	No	264

Female	0	Yes	54	No	No	No	No	No	110
Female	0	Yes	51	No	No	No	No	No	123
Female	0	Yes	45	No	Yes	No	No	No	477
Male	0	Yes	41	No	No	No	No	No	257
Female	0	Yes	54	No	Yes	No	No	No	175
Female	0	Yes	49	No	No	No	No	No	225
Female	0	Yes	60	No	No	No	No	No	205
Male	0	Yes	57	No	No	Yes	No	No	210
Female	0	Yes	55	Yes	No	Yes	No	Yes	292
Male	0	Yes	41	No	No	Yes	No	No	297
Female	0	Yes	59	No	Yes	Yes	No	No	119
Male	0	Yes	49	No	No	No	No	No	172
Female	0	Yes	60	No	No	No	No	No	325

مجموعة البيانات الثانية: بيانات عن مرض القلب وعلاقته بالعديد من العوامل المهنية.

تعريف اسماء المتغيرات الواردة في الجدول

- **SBP**: Systolic Blood Pressure (ضغط الدم الانقباضي).
- **Tobacco** (كمية النيكوتين المتراكمة في كل كيلو غرام من الجسم).
- **LDL**: Low Density Lipoprotein Cholesterol (الكولسترول منخفض الكثافة).

• **Adiposity**: السمنة

• **Famhist**: family history of heart disease (Present,

Absent) (هل هناك قصة مرضية عائلية. الإجابة الممكنة: العائلة لديها

تاريخ في مرض القلب Present أو مرض القلب ليس موجوداً في عائلة المريض Absent).

• **Typea**: type-A behavior (السلوك من النوع A).

• **Obesity**: زيادة الوزن.

• **Alcohol**: current alcohol consumption (كمية الكحول التي

يستهلكها المريض حالياً)

• **Age**: (عمر المريض أثناء الفحص الطبي).

• **CHD**: response, coronary heart disease (هل الشخص

مصاب بمرض القلب "1" أو لا "0")

SBP	Tobacco	LDL	Adiposity	Famhist	Typea	Obesity	Alcohol	Age	CHD
160	12	5.73	23.11	Present	49	25.3	97.2	52	1
144	0.01	4.41	28.61	Absent	55	28.87	2.06	63	1
118	0.08	3.48	32.28	Present	52	29.14	3.81	46	0
170	7.5	6.41	38.03	Present	51	31.99	24.26	58	1
134	13.6	3.5	27.78	Present	60	25.99	57.34	49	1
132	6.2	6.47	36.21	Present	62	30.77	14.14	45	0
142	4.05	3.38	16.2	Absent	59	20.81	2.62	38	0
114	4.08	4.59	14.6	Present	62	23.11	6.72	58	1
114	0	3.83	19.4	Present	49	24.86	2.49	29	0
132	0	5.8	30.96	Present	69	30.11	0	53	1
206	6	2.95	32.27	Absent	72	26.81	56.06	60	1
134	14.1	4.44	22.39	Present	65	23.09	0	40	1

118	0	1.88	10.05	Absent	59	21.57	0	17	0
132	0	1.87	17.21	Absent	49	23.63	0.97	15	0
112	9.65	2.29	17.2	Present	54	23.53	0.68	53	0
117	1.53	2.44	28.95	Present	35	25.89	30.03	46	0
120	7.5	15.33	22	Absent	60	25.31	34.49	49	0
146	10.5	8.29	35.36	Present	78	32.73	13.89	53	1
158	2.6	7.46	34.07	Present	61	29.3	53.28	62	1
124	14	6.23	35.96	Present	45	30.09	0	59	1
106	1.61	1.74	12.32	Absent	74	20.92	13.37	20	1
132	7.9	2.85	26.5	Present	51	26.16	25.71	44	0
150	0.3	6.38	33.99	Present	62	24.64	0	50	0
138	0.6	3.81	28.66	Absent	54	28.7	1.46	58	0
142	18.2	4.34	24.38	Absent	61	26.19	0	50	0
124	4	12.42	31.29	Present	54	23.23	2.06	42	1
118	6	9.65	33.91	Absent	60	38.8	0	48	0
145	9.1	5.24	27.55	Absent	59	20.96	21.6	61	1
144	4.09	5.55	31.4	Present	60	29.43	5.55	56	0
146	0	6.62	25.69	Absent	60	28.07	8.23	63	1
136	2.52	3.95	25.63	Absent	51	21.86	0	45	1
158	1.02	6.33	23.88	Absent	66	22.13	24.99	46	1
122	6.6	5.58	35.95	Present	53	28.07	12.55	59	1
126	8.75	6.53	34.02	Absent	49	30.25	0	41	1
148	5.5	7.1	25.31	Absent	56	29.84	3.6	48	0
122	4.26	4.44	13.04	Absent	57	19.49	48.99	28	1
140	3.9	7.32	25.05	Absent	47	27.36	36.77	32	0
110	4.64	4.55	30.46	Absent	48	30.9	15.22	46	0
130	0	2.82	19.63	Present	70	24.86	0	29	0
136	11.2	5.81	31.85	Present	75	27.68	22.94	58	1
118	0.28	5.8	33.7	Present	60	30.98	0	41	1
144	0.04	3.38	23.61	Absent	30	23.75	4.66	30	0
120	0	1.07	16.02	Absent	47	22.15	0	15	0
130	2.61	2.72	22.99	Present	51	26.29	13.37	51	1
114	0	2.99	9.74	Absent	54	46.58	0	17	0
128	4.65	3.31	22.74	Absent	62	22.95	0.51	48	0
162	7.4	8.55	24.65	Present	64	25.71	5.86	58	1

116	1.91	7.56	26.45	Present	52	30.01	3.6	33	1
114	0	1.94	11.02	Absent	54	20.17	38.98	16	0
126	3.8	3.88	31.79	Absent	57	30.53	0	30	0
122	0	5.75	30.9	Present	46	29.01	4.11	42	0
134	2.5	3.66	30.9	Absent	52	27.19	23.66	49	0
152	0.9	9.12	30.23	Absent	56	28.64	0.37	42	1
134	8.08	1.55	17.5	Present	56	22.65	66.65	31	1
156	3	1.82	27.55	Absent	60	23.91	54	53	0
152	5.99	7.99	32.48	Absent	45	26.57	100.32	48	0
118	0	2.99	16.17	Absent	49	23.83	3.22	28	0
126	5.1	2.96	26.5	Absent	55	25.52	12.34	38	1
103	0.03	4.21	18.96	Absent	48	22.94	2.62	18	0
121	0.8	5.29	18.95	Present	47	22.51	0	61	0
142	0.28	1.8	21.03	Absent	57	23.65	2.93	33	0
138	1.15	5.09	27.87	Present	61	25.65	2.34	44	0
152	10.1	4.71	24.65	Present	65	26.21	24.53	57	0
140	0.45	4.3	24.33	Absent	41	27.23	10.08	38	0
130	0	1.82	10.45	Absent	57	22.07	2.06	17	0
136	7.36	2.19	28.11	Present	61	25	61.71	54	0
124	4.82	3.24	21.1	Present	48	28.49	8.42	30	0
112	0.41	1.88	10.29	Absent	39	22.08	20.98	27	0
118	4.46	7.27	29.13	Present	48	29.01	11.11	33	0
122	0	3.37	16.1	Absent	67	21.06	0	32	1
118	0	3.67	12.13	Absent	51	19.15	0.6	15	0
130	1.72	2.66	10.38	Absent	68	17.81	11.1	26	0
130	5.6	3.37	24.8	Absent	58	25.76	43.2	36	0
126	0.09	5.03	13.27	Present	50	17.75	4.63	20	0
128	0.4	6.17	26.35	Absent	64	27.86	11.11	34	0
136	0	4.12	17.42	Absent	52	21.66	12.86	40	0
134	0	5.9	30.84	Absent	49	29.16	0	55	0
140	0.6	5.56	33.39	Present	58	27.19	0	55	1
168	4.5	6.68	28.47	Absent	43	24.25	24.38	56	1
108	0.4	5.91	22.92	Present	57	25.72	72	39	0
114	3	7.04	22.64	Present	55	22.59	0	45	1
140	8.14	4.93	42.49	Absent	53	45.72	6.43	53	1

148	4.8	6.09	36.55	Present	63	25.44	0.88	55	1
148	12.2	3.79	34.15	Absent	57	26.38	14.4	57	1
128	0	2.43	13.15	Present	63	20.75	0	17	0
130	0.56	3.3	30.86	Absent	49	27.52	33.33	45	0
126	10.5	4.49	17.33	Absent	67	19.37	0	49	1
140	0	5.08	27.33	Present	41	27.83	1.25	38	0
126	0.9	5.64	17.78	Present	55	21.94	0	41	0
122	0.72	4.04	32.38	Absent	34	28.34	0	55	0
116	1.03	2.83	10.85	Absent	45	21.59	1.75	21	0
120	3.7	4.02	39.66	Absent	61	30.57	0	64	1
143	0.46	2.4	22.87	Absent	62	29.17	15.43	29	0
118	4	3.95	18.96	Absent	54	25.15	8.33	49	1
194	1.7	6.32	33.67	Absent	47	30.16	0.19	56	0
134	3	4.37	23.07	Absent	56	20.54	9.65	62	0
138	2.16	4.9	24.83	Present	39	26.06	28.29	29	0
136	0	5	27.58	Present	49	27.59	1.47	39	0
122	3.2	11.32	35.36	Present	55	27.07	0	51	1
164	12	3.91	19.59	Absent	51	23.44	19.75	39	0
136	8	7.85	23.81	Present	51	22.69	2.78	50	0
166	0.07	4.03	29.29	Absent	53	28.37	0	27	0
118	0	4.34	30.12	Present	52	32.18	3.91	46	0
128	0.42	4.6	26.68	Absent	41	30.97	10.33	31	0
118	1.5	5.38	25.84	Absent	64	28.63	3.89	29	0
158	3.6	2.97	30.11	Absent	63	26.64	108	64	0
108	1.5	4.33	24.99	Absent	66	22.29	21.6	61	1
170	7.6	5.5	37.83	Present	42	37.41	6.17	54	1
118	1	5.76	22.1	Absent	62	23.48	7.71	42	0
124	0	3.04	17.33	Absent	49	22.04	0	18	0
114	0	8.01	21.64	Absent	66	25.51	2.49	16	0
168	9	8.53	24.48	Present	69	26.18	4.63	54	1
134	2	3.66	14.69	Absent	52	21.03	2.06	37	0
174	0	8.46	35.1	Present	35	25.27	0	61	1
116	31.2	3.17	14.99	Absent	47	19.4	49.06	59	1
128	0	10.58	31.81	Present	46	28.41	14.66	48	0
140	4.5	4.59	18.01	Absent	63	21.91	22.09	32	1

154	0.7	5.91	25	Absent	13	20.6	0	42	0
150	3.5	6.99	25.39	Present	50	23.35	23.48	61	1
130	0	3.92	25.55	Absent	68	28.02	0.68	27	0
128	2	6.13	21.31	Absent	66	22.86	11.83	60	0
120	1.4	6.25	20.47	Absent	60	25.85	8.51	28	0
120	0	5.01	26.13	Absent	64	26.21	12.24	33	0
138	4.5	2.85	30.11	Absent	55	24.78	24.89	56	1
153	7.8	3.96	25.73	Absent	54	25.91	27.03	45	0
123	8.6	11.17	35.28	Present	70	33.14	0	59	1
148	4.04	3.99	20.69	Absent	60	27.78	1.75	28	0
136	3.96	2.76	30.28	Present	50	34.42	18.51	38	0
134	8.8	7.41	26.84	Absent	35	29.44	29.52	60	1
152	12.18	4.04	37.83	Present	63	34.57	4.17	64	0
158	13.5	5.04	30.79	Absent	54	24.79	21.5	62	0
132	2	3.08	35.39	Absent	45	31.44	79.82	58	1
134	1.5	3.73	21.53	Absent	41	24.7	11.11	30	1
142	7.44	5.52	33.97	Absent	47	29.29	24.27	54	0
134	6	3.3	28.45	Absent	65	26.09	58.11	40	0
122	4.18	9.05	29.27	Present	44	24.05	19.34	52	1
116	2.7	3.69	13.52	Absent	55	21.13	18.51	32	0
128	0.5	3.7	12.81	Present	66	21.25	22.73	28	0
120	0	3.68	12.24	Absent	51	20.52	0.51	20	0
124	0	3.95	36.35	Present	59	32.83	9.59	54	0
160	14	5.9	37.12	Absent	58	33.87	3.52	54	1
130	2.78	4.89	9.39	Present	63	19.3	17.47	25	1
128	2.8	5.53	14.29	Absent	64	24.97	0.51	38	0
130	4.5	5.86	37.43	Absent	61	31.21	32.3	58	0
109	1.2	6.14	29.26	Absent	47	24.72	10.46	40	0
144	0	3.84	18.72	Absent	56	22.1	4.8	40	0
118	1.05	3.16	12.98	Present	46	22.09	16.35	31	0
136	3.46	6.38	32.25	Present	43	28.73	3.13	43	1
136	1.5	6.06	26.54	Absent	54	29.38	14.5	33	1
124	15.5	5.05	24.06	Absent	46	23.22	0	61	1
148	6	6.49	26.47	Absent	48	24.7	0	55	0
128	6.6	3.58	20.71	Absent	55	24.15	0	52	0

122	0.28	4.19	19.97	Absent	61	25.63	0	24	0
108	0	2.74	11.17	Absent	53	22.61	0.95	20	0
124	3.04	4.8	19.52	Present	60	21.78	147.19	41	1
138	8.8	3.12	22.41	Present	63	23.33	120.03	55	1
127	0	2.81	15.7	Absent	42	22.03	1.03	17	0
174	9.45	5.13	35.54	Absent	55	30.71	59.79	53	0
122	0	3.05	23.51	Absent	46	25.81	0	38	0
144	6.75	5.45	29.81	Absent	53	25.62	26.23	43	1
126	1.8	6.22	19.71	Absent	65	24.81	0.69	31	0
208	27.4	3.12	26.63	Absent	66	27.45	33.07	62	1
138	0	2.68	17.04	Absent	42	22.16	0	16	0
148	0	3.84	17.26	Absent	70	20	0	21	0
122	0	3.08	16.3	Absent	43	22.13	0	16	0
132	7	3.2	23.26	Absent	77	23.64	23.14	49	0
110	12.16	4.99	28.56	Absent	44	27.14	21.6	55	1
160	1.52	8.12	29.3	Present	54	25.87	12.86	43	1
126	0.54	4.39	21.13	Present	45	25.99	0	25	0
162	5.3	7.95	33.58	Present	58	36.06	8.23	48	0
194	2.55	6.89	33.88	Present	69	29.33	0	41	0
118	0.75	2.58	20.25	Absent	59	24.46	0	32	0
124	0	4.79	34.71	Absent	49	26.09	9.26	47	0
160	0	2.42	34.46	Absent	48	29.83	1.03	61	0
128	0	2.51	29.35	Present	53	22.05	1.37	62	0
122	4	5.24	27.89	Present	45	26.52	0	61	1
132	2	2.7	21.57	Present	50	27.95	9.26	37	0
120	0	2.42	16.66	Absent	46	20.16	0	17	0
128	0.04	8.22	28.17	Absent	65	26.24	11.73	24	0
108	15	4.91	34.65	Absent	41	27.96	14.4	56	0
166	0	4.31	34.27	Absent	45	30.14	13.27	56	0
152	0	6.06	41.05	Present	51	40.34	0	51	0
170	4.2	4.67	35.45	Present	50	27.14	7.92	60	1
156	4	2.05	19.48	Present	50	21.48	27.77	39	1
116	8	6.73	28.81	Present	41	26.74	40.94	48	1
122	4.4	3.18	11.59	Present	59	21.94	0	33	1
150	20	6.4	35.04	Absent	53	28.88	8.33	63	0

129	2.15	5.17	27.57	Absent	52	25.42	2.06	39	0
134	4.8	6.58	29.89	Present	55	24.73	23.66	63	0
126	0	5.98	29.06	Present	56	25.39	11.52	64	1
142	0	3.72	25.68	Absent	48	24.37	5.25	40	1
128	0.7	4.9	37.42	Present	72	35.94	3.09	49	1
102	0.4	3.41	17.22	Present	56	23.59	2.06	39	1
130	0	4.89	25.98	Absent	72	30.42	14.71	23	0
138	0.05	2.79	10.35	Absent	46	21.62	0	18	0
138	0	1.96	11.82	Present	54	22.01	8.13	21	0
128	0	3.09	20.57	Absent	54	25.63	0.51	17	0
162	2.92	3.63	31.33	Absent	62	31.59	18.51	42	0
160	3	9.19	26.47	Present	39	28.25	14.4	54	1
148	0	4.66	24.39	Absent	50	25.26	4.03	27	0
124	0.16	2.44	16.67	Absent	65	24.58	74.91	23	0
136	3.15	4.37	20.22	Present	59	25.12	47.16	31	1
134	2.75	5.51	26.17	Absent	57	29.87	8.33	33	0
128	0.73	3.97	23.52	Absent	54	23.81	19.2	64	0
122	3.2	3.59	22.49	Present	45	24.96	36.17	58	0
152	3	4.64	31.29	Absent	41	29.34	4.53	40	0
162	0	5.09	24.6	Present	64	26.71	3.81	18	0
124	4	6.65	30.84	Present	54	28.4	33.51	60	0
136	5.8	5.9	27.55	Absent	65	25.71	14.4	59	0
136	8.8	4.26	32.03	Present	52	31.44	34.35	60	0
134	0.05	8.03	27.95	Absent	48	26.88	0	60	0
122	1	5.88	34.81	Present	69	31.27	15.94	40	1
116	3	3.05	30.31	Absent	41	23.63	0.86	44	0
132	0	0.98	21.39	Absent	62	26.75	0	53	0
134	0	2.4	21.11	Absent	57	22.45	1.37	18	0
160	7.77	8.07	34.8	Absent	64	31.15	0	62	1
180	0.52	4.23	16.38	Absent	55	22.56	14.77	45	1
124	0.81	6.16	11.61	Absent	35	21.47	10.49	26	0
114	0	4.97	9.69	Absent	26	22.6	0	25	0
208	7.4	7.41	32.03	Absent	50	27.62	7.85	57	0
138	0	3.14	12	Absent	54	20.28	0	16	0
164	0.5	6.95	39.64	Present	47	41.76	3.81	46	1

144	2.4	8.13	35.61	Absent	46	27.38	13.37	60	0
136	7.5	7.39	28.04	Present	50	25.01	0	45	1
132	7.28	3.52	12.33	Absent	60	19.48	2.06	56	0
143	5.04	4.86	23.59	Absent	58	24.69	18.72	42	0
112	4.46	7.18	26.25	Present	69	27.29	0	32	1
134	10	3.79	34.72	Absent	42	28.33	28.8	52	1
138	2	5.11	31.4	Present	49	27.25	2.06	64	1
188	0	5.47	32.44	Present	71	28.99	7.41	50	1
110	2.35	3.36	26.72	Present	54	26.08	109.8	58	1
136	13.2	7.18	35.95	Absent	48	29.19	0	62	0
130	1.75	5.46	34.34	Absent	53	29.42	0	58	1
122	0	3.76	24.59	Absent	56	24.36	0	30	0
138	0	3.24	27.68	Absent	60	25.7	88.66	29	0
130	18	4.13	27.43	Absent	54	27.44	0	51	1
126	5.5	3.78	34.15	Absent	55	28.85	3.18	61	0
176	5.76	4.89	26.1	Present	46	27.3	19.44	57	0
122	0	5.49	19.56	Absent	57	23.12	14.02	27	0
124	0	3.23	9.64	Absent	59	22.7	0	16	0
140	5.2	3.58	29.26	Absent	70	27.29	20.17	45	1
128	6	4.37	22.98	Present	50	26.01	0	47	0
190	4.18	5.05	24.83	Absent	45	26.09	82.85	41	0
144	0.76	10.53	35.66	Absent	63	34.35	0	55	1
126	4.6	7.4	31.99	Present	57	28.67	0.37	60	1
128	0	2.63	23.88	Absent	45	21.59	6.54	57	0
136	0.4	3.91	21.1	Present	63	22.3	0	56	1
158	4	4.18	28.61	Present	42	25.11	0	60	0
160	0.6	6.94	30.53	Absent	36	25.68	1.42	64	0
124	6	5.21	33.02	Present	64	29.37	7.61	58	1
158	6.17	8.12	30.75	Absent	46	27.84	92.62	48	0
128	0	6.34	11.87	Absent	57	23.14	0	17	0
166	3	3.82	26.75	Absent	45	20.86	0	63	1
146	7.5	7.21	25.93	Present	55	22.51	0.51	42	0
161	9	4.65	15.16	Present	58	23.76	43.2	46	0
164	13.02	6.26	29.38	Present	47	22.75	37.03	54	1
146	5.08	7.03	27.41	Present	63	36.46	24.48	37	1

142	4.48	3.57	19.75	Present	51	23.54	3.29	49	0
138	12	5.13	28.34	Absent	59	24.49	32.81	58	1
154	1.8	7.13	34.04	Present	52	35.51	39.36	44	0
118	0	2.39	12.13	Absent	49	18.46	0.26	17	1
124	0.61	2.69	17.15	Present	61	22.76	11.55	20	0
124	1.04	2.84	16.42	Present	46	20.17	0	61	0
136	5	4.19	23.99	Present	68	27.8	25.86	35	0
132	9.9	4.63	27.86	Present	46	23.39	0.51	52	1
118	0.12	1.96	20.31	Absent	37	20.01	2.42	18	0
118	0.12	4.16	9.37	Absent	57	19.61	0	17	0
134	12	4.96	29.79	Absent	53	24.86	8.23	57	0
114	0.1	3.95	15.89	Present	57	20.31	17.14	16	0
136	6.8	7.84	30.74	Present	58	26.2	23.66	45	1
130	0	4.16	39.43	Present	46	30.01	0	55	1
136	2.2	4.16	38.02	Absent	65	37.24	4.11	41	1
136	1.36	3.16	14.97	Present	56	24.98	7.3	24	0
154	4.2	5.59	25.02	Absent	58	25.02	1.54	43	0
108	0.8	2.47	17.53	Absent	47	22.18	0	55	1
136	8.8	4.69	36.07	Present	38	26.56	2.78	63	1
174	2.02	6.57	31.9	Present	50	28.75	11.83	64	1
124	4.25	8.22	30.77	Absent	56	25.8	0	43	0
114	0	2.63	9.69	Absent	45	17.89	0	16	0
118	0.12	3.26	12.26	Absent	55	22.65	0	16	0
106	1.08	4.37	26.08	Absent	67	24.07	17.74	28	1
146	3.6	3.51	22.67	Absent	51	22.29	43.71	42	0
206	0	4.17	33.23	Absent	69	27.36	6.17	50	1
134	3	3.17	17.91	Absent	35	26.37	15.12	27	0
148	15	4.98	36.94	Present	72	31.83	66.27	41	1
126	0.21	3.95	15.11	Absent	61	22.17	2.42	17	0
134	0	3.69	13.92	Absent	43	27.66	0	19	0
134	0.02	2.8	18.84	Absent	45	24.82	0	17	0
123	0.05	4.61	13.69	Absent	51	23.23	2.78	16	0
112	0.6	5.28	25.71	Absent	55	27.02	27.77	38	1
112	0	1.71	15.96	Absent	42	22.03	3.5	16	0
101	0.48	7.26	13	Absent	50	19.82	5.19	16	0

150	0.18	4.14	14.4	Absent	53	23.43	7.71	44	0
170	2.6	7.22	28.69	Present	71	27.87	37.65	56	1
134	0	5.63	29.12	Absent	68	32.33	2.02	34	0
142	0	4.19	18.04	Absent	56	23.65	20.78	42	1
132	0.1	3.28	10.73	Absent	73	20.42	0	17	0
136	0	2.28	18.14	Absent	55	22.59	0	17	0
132	12	4.51	21.93	Absent	61	26.07	64.8	46	1
166	4.1	4	34.3	Present	32	29.51	8.23	53	0
138	0	3.96	24.7	Present	53	23.8	0	45	0
138	2.27	6.41	29.07	Absent	58	30.22	2.93	32	1
170	0	3.12	37.15	Absent	47	35.42	0	53	0
128	0	8.41	28.82	Present	60	26.86	0	59	1
136	1.2	2.78	7.12	Absent	52	22.51	3.41	27	0
128	0	3.22	26.55	Present	39	26.59	16.71	49	0
150	14.4	5.04	26.52	Present	60	28.84	0	45	0
132	8.4	3.57	13.68	Absent	42	18.75	15.43	59	1
142	2.4	2.55	23.89	Absent	54	26.09	59.14	37	0
130	0.05	2.44	28.25	Present	67	30.86	40.32	34	0
174	3.5	5.26	21.97	Present	36	22.04	8.33	59	1
114	9.6	2.51	29.18	Absent	49	25.67	40.63	46	0
162	1.5	2.46	19.39	Present	49	24.32	0	59	1
174	0	3.27	35.4	Absent	58	37.71	24.95	44	0
190	5.15	6.03	36.59	Absent	42	30.31	72	50	0
154	1.4	1.72	18.86	Absent	58	22.67	43.2	59	0
124	0	2.28	24.86	Present	50	22.24	8.26	38	0
114	1.2	3.98	14.9	Absent	49	23.79	25.82	26	0
168	11.4	5.08	26.66	Present	56	27.04	2.61	59	1
142	3.72	4.24	32.57	Absent	52	24.98	7.61	51	0
154	0	4.81	28.11	Present	56	25.67	75.77	59	0
146	4.36	4.31	18.44	Present	47	24.72	10.8	38	0
166	6	3.02	29.3	Absent	35	24.38	38.06	61	0
140	8.6	3.9	32.16	Present	52	28.51	11.11	64	1
136	1.7	3.53	20.13	Absent	56	19.44	14.4	55	0
156	0	3.47	21.1	Absent	73	28.4	0	36	1
132	0	6.63	29.58	Present	37	29.41	2.57	62	0

128	0	2.98	12.59	Absent	65	20.74	2.06	19	0
106	5.6	3.2	12.3	Absent	49	20.29	0	39	0
144	0.4	4.64	30.09	Absent	30	27.39	0.74	55	0
154	0.31	2.33	16.48	Absent	33	24	11.83	17	0
126	3.1	2.01	32.97	Present	56	28.63	26.74	45	0
134	6.4	8.49	37.25	Present	56	28.94	10.49	51	1
152	19.45	4.22	29.81	Absent	28	23.95	0	59	1
146	1.35	6.39	34.21	Absent	51	26.43	0	59	1
162	6.94	4.55	33.36	Present	52	27.09	32.06	43	0
130	7.28	3.56	23.29	Present	20	26.8	51.87	58	1
138	6	7.24	37.05	Absent	38	28.69	0	59	0
148	0	5.32	26.71	Present	52	32.21	32.78	27	0
124	4.2	2.94	27.59	Absent	50	30.31	85.06	30	0
118	1.62	9.01	21.7	Absent	59	25.89	21.19	40	0
116	4.28	7.02	19.99	Present	68	23.31	0	52	1
162	6.3	5.73	22.61	Present	46	20.43	62.54	53	1
138	0.87	1.87	15.89	Absent	44	26.76	42.99	31	0
137	1.2	3.14	23.87	Absent	66	24.13	45	37	0
198	0.52	11.89	27.68	Present	48	28.4	78.99	26	1
154	4.5	4.75	23.52	Present	43	25.76	0	53	1
128	5.4	2.36	12.98	Absent	51	18.36	6.69	61	0
130	0.08	5.59	25.42	Present	50	24.98	6.27	43	1
162	5.6	4.24	22.53	Absent	29	22.91	5.66	60	0
120	10.5	2.7	29.87	Present	54	24.5	16.46	49	0
136	3.99	2.58	16.38	Present	53	22.41	27.67	36	0
176	1.2	8.28	36.16	Present	42	27.81	11.6	58	1
134	11.79	4.01	26.57	Present	38	21.79	38.88	61	1
122	1.7	5.28	32.23	Present	51	24.08	0	54	0
134	0.9	3.18	23.66	Present	52	23.26	27.36	58	1
134	0	2.43	22.24	Absent	52	26.49	41.66	24	0
136	6.6	6.08	32.74	Absent	64	33.28	2.72	49	0
132	4.05	5.15	26.51	Present	31	26.67	16.3	50	0
152	1.68	3.58	25.43	Absent	50	27.03	0	32	0
132	12.3	5.96	32.79	Present	57	30.12	21.5	62	1
124	0.4	3.67	25.76	Absent	43	28.08	20.57	34	0

140	4.2	2.91	28.83	Present	43	24.7	47.52	48	0
166	0.6	2.42	34.03	Present	53	26.96	54	60	0
156	3.02	5.35	25.72	Present	53	25.22	28.11	52	1
132	0.72	4.37	19.54	Absent	48	26.11	49.37	28	0
150	0	4.99	27.73	Absent	57	30.92	8.33	24	0
134	0.12	3.4	21.18	Present	33	26.27	14.21	30	0
126	3.4	4.87	15.16	Present	65	22.01	11.11	38	0
148	0.5	5.97	32.88	Absent	54	29.27	6.43	42	0
148	8.2	7.75	34.46	Present	46	26.53	6.04	64	1
132	6	5.97	25.73	Present	66	24.18	145.29	41	0
128	1.6	5.41	29.3	Absent	68	29.38	23.97	32	0
128	5.16	4.9	31.35	Present	57	26.42	0	64	0
140	0	2.4	27.89	Present	70	30.74	144	29	0
126	0	5.29	27.64	Absent	25	27.62	2.06	45	0
114	3.6	4.16	22.58	Absent	60	24.49	65.31	31	0
118	1.25	4.69	31.58	Present	52	27.16	4.11	53	0
126	0.96	4.99	29.74	Absent	66	33.35	58.32	38	0
154	4.5	4.68	39.97	Absent	61	33.17	1.54	64	1
112	1.44	2.71	22.92	Absent	59	24.81	0	52	0
140	8	4.42	33.15	Present	47	32.77	66.86	44	0
140	1.68	11.41	29.54	Present	74	30.75	2.06	38	1
128	2.6	4.94	21.36	Absent	61	21.3	0	31	0
126	19.6	6.03	34.99	Absent	49	26.99	55.89	44	0
160	4.2	6.76	37.99	Present	61	32.91	3.09	54	1
144	0	4.17	29.63	Present	52	21.83	0	59	0
148	4.5	10.49	33.27	Absent	50	25.92	2.06	53	1
146	0	4.92	18.53	Absent	57	24.2	34.97	26	0
164	5.6	3.17	30.98	Present	44	25.99	43.2	53	1
130	0.54	3.63	22.03	Present	69	24.34	12.86	39	1
154	2.4	5.63	42.17	Present	59	35.07	12.86	50	1
178	0.95	4.75	21.06	Absent	49	23.74	24.69	61	0
180	3.57	3.57	36.1	Absent	36	26.7	19.95	64	0
134	12.5	2.73	39.35	Absent	48	35.58	0	48	0
142	0	3.54	16.64	Absent	58	25.97	8.36	27	0
162	7	7.67	34.34	Present	33	30.77	0	62	0

218	11.2	2.77	30.79	Absent	38	24.86	90.93	48	1
126	8.75	6.06	32.72	Present	33	27	62.43	55	1
126	0	3.57	26.01	Absent	61	26.3	7.97	47	0
134	6.1	4.77	26.08	Absent	47	23.82	1.03	49	0
132	0	4.17	36.57	Absent	57	30.61	18	49	0
178	5.5	3.79	23.92	Present	45	21.26	6.17	62	1
208	5.04	5.19	20.71	Present	52	25.12	24.27	58	1
160	1.15	10.19	39.71	Absent	31	31.65	20.52	57	0
116	2.38	5.67	29.01	Present	54	27.26	15.77	51	0
180	25.01	3.7	38.11	Present	57	30.54	0	61	1
200	19.2	4.43	40.6	Present	55	32.04	36	60	1
112	4.2	3.58	27.14	Absent	52	26.83	2.06	40	0
120	0	3.1	26.97	Absent	41	24.8	0	16	0
178	20	9.78	33.55	Absent	37	27.29	2.88	62	1
166	0.8	5.63	36.21	Absent	50	34.72	28.8	60	0
164	8.2	14.16	36.85	Absent	52	28.5	17.02	55	1
216	0.92	2.66	19.85	Present	49	20.58	0.51	63	1
146	6.4	5.62	33.05	Present	57	31.03	0.74	46	0
134	1.1	3.54	20.41	Present	58	24.54	39.91	39	1
158	16	5.56	29.35	Absent	36	25.92	58.32	60	0
176	0	3.14	31.04	Present	45	30.18	4.63	45	0
132	2.8	4.79	20.47	Present	50	22.15	11.73	48	0
126	0	4.55	29.18	Absent	48	24.94	36	41	0
120	5.5	3.51	23.23	Absent	46	22.4	90.31	43	0
174	0	3.86	21.73	Absent	42	23.37	0	63	0
150	13.8	5.1	29.45	Present	52	27.92	77.76	55	1
176	6	3.98	17.2	Present	52	21.07	4.11	61	1
142	2.2	3.29	22.7	Absent	44	23.66	5.66	42	1
132	0	3.3	21.61	Absent	42	24.92	32.61	33	0
142	1.32	7.63	29.98	Present	57	31.16	72.93	33	0
146	1.16	2.28	34.53	Absent	50	28.71	45	49	0
132	7.2	3.65	17.16	Present	56	23.25	0	34	0
120	0	3.57	23.22	Absent	58	27.2	0	32	0
118	0	3.89	15.96	Absent	65	20.18	0	16	0
108	0	1.43	26.26	Absent	42	19.38	0	16	0

136	0	4	19.06	Absent	40	21.94	2.06	16	0
120	0	2.46	13.39	Absent	47	22.01	0.51	18	0
132	0	3.55	8.66	Present	61	18.5	3.87	16	0
136	0	1.77	20.37	Absent	45	21.51	2.06	16	0
138	0	1.86	18.35	Present	59	25.38	6.51	17	0
138	0.06	4.15	20.66	Absent	49	22.59	2.49	16	0
130	1.22	3.3	13.65	Absent	50	21.4	3.81	31	0
130	4	2.4	17.42	Absent	60	22.05	0	40	0
110	0	7.14	28.28	Absent	57	29	0	32	0
120	0	3.98	13.19	Present	47	21.89	0	16	0
166	6	8.8	37.89	Absent	39	28.7	43.2	52	0
134	0.57	4.75	23.07	Absent	67	26.33	0	37	0
142	3	3.69	25.1	Absent	60	30.08	38.88	27	0
136	2.8	2.53	9.28	Present	61	20.7	4.55	25	0
142	0	4.32	25.22	Absent	47	28.92	6.53	34	1
130	0	1.88	12.51	Present	52	20.28	0	17	0
124	1.8	3.74	16.64	Present	42	22.26	10.49	20	0
144	4	5.03	25.78	Present	57	27.55	90	48	1
136	1.81	3.31	6.74	Absent	63	19.57	24.94	24	0
120	0	2.77	13.35	Absent	67	23.37	1.03	18	0
154	5.53	3.2	28.81	Present	61	26.15	42.79	42	0
124	1.6	7.22	39.68	Present	36	31.5	0	51	1
146	0.64	4.82	28.02	Absent	60	28.11	8.23	39	1
128	2.24	2.83	26.48	Absent	48	23.96	47.42	27	1
170	0.4	4.11	42.06	Present	56	33.1	2.06	57	0
214	0.4	5.98	31.72	Absent	64	28.45	0	58	0
182	4.2	4.41	32.1	Absent	52	28.61	18.72	52	1
108	3	1.59	15.23	Absent	40	20.09	26.64	55	0
118	5.4	11.61	30.79	Absent	64	27.35	23.97	40	0
132	0	4.82	33.41	Present	62	14.7	0	46	1

★★★★★★★★★★

A

Adjacent value	قيمة مجاورة (غير شاذة)
Alternative	بديل
Alternative Hypothesis	الفرضية البديلة
Analysis	احليل
Analysis of Variance	تحليل التباين

B

Brown-Forsythe	اختبار Brown-Forsythe
Binary Data	بيانات ثنائية
Binary Variable	متغير ثنائي القيمة (يأخذ قيمتين فقط)
Box Plot	المخطط الصندوقي

C

Correlation	ارتباط
Categorical Data	بيانات متقطعة تأخذ قيمتين أو أكثر
Cluster	تجمع (عقود)
Covariance	التغاير
Classification	تصنيف
Coefficient	معامل
Coefficient of Variation	معامل التباين
Cluster Sampling	معاينة عنقودية

D

Descriptive Statistics	إحصاء وصفي
------------------------	------------

Data Analysis	تحليل البيانات
Decision	قرار
Dependent Variable	متغير تابع
Dependent	معتمد
Data	معطيات
Decision Theory	نظرية القرار

E

Element	عنصر
Endogenous	مرتبط
Equation	معادلة
Error	خطأ
Estimation	تقدير
Estimation Error	خطأ تقدير
Evaluation	تقويم
Exogenous	مستقل
Experimental unit	وحدة تجريبية (في العينة)
Exploratory	استكشافي
Exploratory Data Analysis	تحليل البيانات الإستكشافي
Explore	يستكشف

F

Factor	مؤثر (عامل تأثير)
Formula	صيغة (رياضية ، منطقية ، إلخ)

G

Games-Howell Test	اختبار Games-Howell
Graphical Methods	طرق بيانية

II

Heterogeneity	عدم تجانس (إختلاف)
Homogeneity	تجانس
Homogeneity of Variance	تجانس التباين
Hypothesis	فرضية
Hypothesis Test	اختبار فرضية

I

Importance Level	مستوى الأهمية
Independent	مستقل
Independent Samples	عينات مستقلة
Independent Variable	متغير مستقل
Interaction	تفاعل (تأثير متبادل)
Inter-Quartile Range	المدى الربيعي

L

Leaf	ورقة
Linear	خطي
Lower Quartile	الربيعي الأول (الأدنى)

M

Mean	متوسط
Measure	قياس
Median	وسط
Mode	منوال
Model	نموذج
Multi	متعدد
Multi-Comparisons	مقارنات متعددة (غالباً ثنائية)

Multiple

متعدد

N

Nominal

اسمي

Nominal Data

بيانات اسمية

Nonparametric Test

اختبار غير وسيطي

Null

عدم (صفرى)

Null Hypothesis

فرضية العدم (الفرضية الصفرية)

O

Odds

احتمال أو إمكانية

One-Sample T Test

اختبار ستودنت لعينة واحدة

One –variable linear correlation

انحدار خطي بمتغير واحد

Ordinal Data

بيانات ترتيبية

Observational Data

بيانات مشاهدة أو وصفية و تأخذ

بالملاحظة المباشرة

One-Way ANOVA

تحليل التباين بمتغير واحد مستقل

Ordinal

ترتيبي

Outliers

نقاط شاذة

P

Paired Samples

عينات مترابطة

Parameter

وسيط

Parametric Test

اختبار وسيطي

Percentiles

المئينيات

Plot

مخطط (رسم بياني)

Point

نقطة

Population	مجتمع
Potential Outlier	نقطة شاذة ممكنة

Q

Qualitative Data	بيانات كيفية
Quantitative Data	بيانات كمية (متصلة)
Quartiles	ربيعيات
Quota	حصة
Quota Sampling	معاينة المحاصصة

R

Random	عشوائي
Random Variable	متغير عشوائي
Range	المدى
Ratio	نسبة أو معدل
Result	نتيجة

S

Sample	عينة
Sampling	معاينة
Sampling Method	طريقة معاينة
Scatter Plot	المخطط التبعثري
Score	قياس
Significance Level	مستوى المعنوية
Significance of Test	معنوية الاختبار
Significant	مهم ، معنوي
Simple	بسيط
Simple Random Sample	عينة عشوائية بسيطة

Skewness	التواء
Software	برمجيات (حزمة برامج)
SPSS (Statistical Processing for Social Statistics)	حزمة برامج الإحصائية SPSS
Square	مربع
Standard	معياري
Standard Deviation	إنحراف معياري
Standard Scores	مقاييس معيارية
Statistical Hypothesis	فرضية إحصائية
Statistical Software	حزمة برامج إحصائية
Statistical Test	اختبار إحصائي
Statistics	علم إحصاء
Stem	الساق
Stem and Leaf Plot	مخطط الساق و الورقة
Stratified	طبقي
Stratified Sampling	معاينة طبقية
Student Test	اختبار ستودنت
Sum	مجموع
Sum of Squares	مجموع المربعات
Systematic	منتظم
Systematic Sampling	معاينة منتظمة
T	
Table	جدول
Tabulation	جدولة
Test	اختبار

Trimmed Mean	متوسط مشذب
Tukey Test	اختبار Tukey
Two-Independent Samples T Test	اختبار ستودنت لعينتين مستقلتين
Two-Paired Samples T Test	اختبار ستودنت لعينتين مرتبطتين
Two-Way ANOVA	تحليل التباين بمتغيرين مستقلتين
Type	نوع
Type I Error	خطأ من النوع الأول
Type II Error	خطأ من النوع الثاني
U	
Unit	وحدة
Univariate	تحليل تباين بمتغيرين مستقلين
Upper Quartile	الرابعي الثالث (الأعلى)
V	
Value	قيمة
Variable	متغير (متغير)
Variance	تباين (تشتت ، تبعثر)
W	
Welch Test	اختبار Welch



References المراجع

- Lyman Ott R. and Longnecker M. (2010). An Introduction to Statistical Methods and Data Analysis. 6th edition, Brooks/Cole, Cengage Learning, Belmont, California, USA, 2010.
- James H. Matis and Muhammed J. Al-Muhammed. (2010). Theory and Application of the Logistic Probability Density Function as a Population Growth Model. Journal of Damascus University, Volume to appear.
- James H. Matis, Muhammed J. Al-Muhammed, and Der Werf, W. V (2010). Using the Logistic pdf Model to Mitigate Autocorrelation in Growth Curve Analysis. Journal of the Indian Society of Agricultural Statistics, Volume 64, No. 2, pages 229 - 236.
- James H. Matis, Muhammed J. Al-Muhammed, and Kaes Mohammed (2010). Mitigating Autocorrelation in Richards Growth Model Analysis Using Incremental Growth Data: A Case Study for Turkey Growth. Journal of the Indian Society of Agricultural Statistics, Volume to appear
- Li, Qi and Maasoumi, Esfandiar and Racine, Jeffrey S (2009). A Nonparametric Test for Equality of Distributions with Mixed Categorical and Continuous Data. *Journal of Econometrics*, volume 148(2), pages 186 – 200. February 2009.
- Field Andy. Discovering Statistics Using SPSS. Third Edition, SAGE Publications Ltd , 2009. Available at: <http://www.uk.sagepub.com/field3e/MCQ.htm>.
- Frank R. Giordano, W. P. (2009). A Frist Course in Mathematical Modeling (4th Edition ed.). Cengage Learning Inc.
- Ann E. Watkins, Scheaffer L. Richard, and Cobb W. George (2008). *Statistics in Action: Understanding a world of Data*. John Wiley & Sons Inc., USA. 2008.
- James H. Matis (2008). Sampling Designs for Experienced Teachers. Texas A & M University, Texas, USA. 2008.

Roxy Peck, Chris Olsen, and Jay Devore (2008). *Introduction to Statistics and Data Analysis*. Thirs Edition, Duxbury, USA, 2008.

Jeffery L. Whitten and Lonnie D. Bentley (2007). *System Analysis and Design for the Global Enterprise*. McGraw-Hill/Irwin, New York, USA. 2007.

Robert V. Hogg, Joseph W. McKean, and Allen T. Craig (2005). *Introduction to Mathematical Statistics*. 6th edition. Upper Saddle River, New Jersey, USA: Pearson Education, INC, 2005.

Theus, M. (2005). *Data Analysis Principles in Interactive Statistical Graphics*. Workshop on modern statistical visualization and related topics, Tokey, Japan. March 11-12, 2005.

Ann E. Watkins, Scheaffer L. Richard, and Cobb W. George (2004). *Statistics in Action: Understanding a world of Data*. Key Curriculum Press (New Edition), 2004.

Norusis, M. (2004). *SPSS 13.0 Guide to Data Analysis*. Upper Saddle-River, N.J.: Prentice Hall, Inc. 2004.

Norusis, M. (2004). *SPSS 13.0 Statistical Procedures Companion*. Upper Saddle-River, N.J.: Prentice Hall, Inc. 2004.

Li, Qi and Racine, Jeff, (2003). *Nonparametric Estimation of Distributions with Categorical and Continuous data*. *Journal of Multivariate Analysis*, Elsevier, volume 86(2), pages 266-292, August, 2003.

Gentle, J. E. (2003). *Random Number Generation and Monte Carlo Methods* (2nd edition ed.). Springer.

V. V. Patel, G. D. (2002). *Some applications of randomized algorithms for control system design* (Vol. 38). Automatica.

Westbrook, L (2000). *Identifying and Analyzing User Needs: a Complete Handbook and Ready-to-Use Assessment Workbook*. Neal-Schuman, 2000, 250pp.

Moore, D.S and McCabe, G.P. (1999). *Introduction to the practice of statistics* (2nd edition). New York: W. H. Freeman & Company, USA. 1999.

Garfield, J. (1995). How students learn statistics. *International Statistical Review*, 63, 25-34. 1995

Raghavan, R. M. (1995). *Randomized Algorithms*. New York: Cambridge University.

Frieze, M. E. (1989). A randomized algorithm for fixed-dimensional linear programming. *Journal of Mathematical Programming* , 44.

Miller, S. P. (1988). Random Number Generators: Good Ones Are Hard To Find. 31 (10) , 1192–1201. *Communications of the ACM*.

Siegel, S., & Castellan, N.J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd Ed.). New York, NY: McGraw-Hill. 1988.

Eckhardt, R. (1987). Stan Ulam, John von Neumann, and the Monte Carlo method. 15 , 131-137. *Los Alamos Science*, Special Issue.

Conover, W. J. (1980). *Practical Nonparametric Statistics*. New York: John Wiley & Sons. 1980.

عدنان عمورة (2010). الاختبارات غير الوسيطة. منشورات جامعة دمشق ، 2010.



التدقيق العلمي Academic Revision

أ.م.د. عماد فتاش كلية الهندسة الميكانيكية والكهربائية قسم العلوم الأساسية	أ.م.د. عدنان عمورة كلية العلوم قسم الإحصاء	أ.م.د. عزات قاسم كلية العلوم قسم الإحصاء
--	--	--

التدقيق اللغوي
د. منى الياس

حقوق الطبع و الترجمة و النشر محفوظة لمديرية الكتب والمطبوعات

DAMASCUS UNIVERSITY
FACULTY OF SCIENCE

STATISTICAL SOFTWARE (1)

FOR STATISTICS STUDENTS
THIRD YEAR

ACADEMIC YEAR 2013 – 2014

DR. MUHAMMED J. AL-MUHAMMED
DR. HASAN AKEL