



د. سبأ محمد علوان
أستاذ مساعد قسم الإحصاء وبحوث العمليات
جامعة الملك سعود

2014



مقدمة

على الرغم من أهمية استخدام الإحصاء في تحليل البيانات للبحوث والرسائل العلمية فإن هناك بعض المشكلات التي قد تنشأ عن هذا الاستخدام والتي تكون غالباً بدون قصد وبسبب عدم التخصص في الإحصاء لدى البعض وبالتالي عدم الإلمام ببعض الجوانب العلمية الإحصائية الدقيقة .

وفي هذه الورشة نحاول معرفة التحليل الإحصائي المناسب من خلال طريقة سهلة وهي الأمثلة المباشرة والتي تحاكي بعض الحالات التي قد يكون فيها الباحث في مساره البحثي، مع ملاحظة أننا نركز على البحوث العلمية وبما هو متاح من الوقت لهذه الورشة .



Saba Mohammed Alwan
Statistical and O.R
Faculty of Science. KSU
salwan@ksu.edu.sa

تحديد حجم العينة المناسب



يتوقف حجم العينة الواجب دراسته على تفاعل عوامل أربعة:

١- مدى التباين في خصائص المجتمع المراد دراسته

(كلما زاد التباين، يزيد حجم العينة المطلوب)،

٢- مدى التفصيل المطلوب في نتائج العينة كتقديرات لخصائص المجتمع

(كلما زادت درجة التفصيل المطلوبة، زاد حجم العينة)

٣- مدى الخطأ الذي يُسمح به في نتائج العينة كتقديرات لخصائص المجتمع

(كلما قل مدى الخطأ الذي يمكن السماح به، زاد حجم العينة)

٤- درجة الثقة التي نود أن نتمتع بها في تحقق السمات السابقة

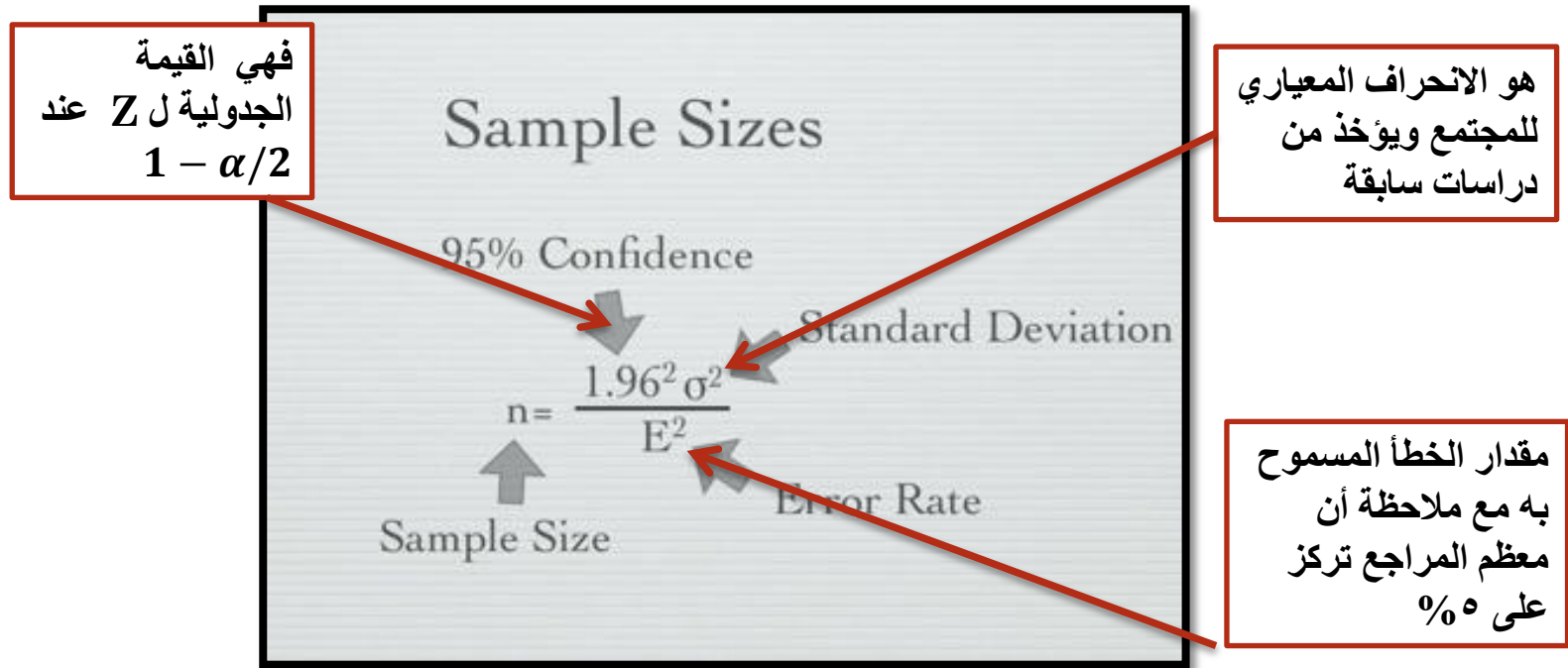
(كلما زادت درجة الثقة المطلوبة، زاد حجم العينة اللازم).

كلما زاد حجم العينة كان ذلك في مصلحة الدقة وصحة النتائج إذا ضمن الباحث السيطرة على المعاملات الداخلة في البحث



وهناك عدد من المعادلات الإحصائية لتحديد حجم العينة المناسب ومنها:

يعتمد تحديد الحد الأدنى لحجم العينة n على أكبر خطأ مسموح به وبالتالي فالعلاقة عكسية بينهما كلما كبر حجم العينة كلما قل الخطأ ويتحدد الحد الأدنى لحجم العينة والذي يتحدد من خلال القانون



ملاحظة:

إن زيادة حجم العينة يجعل من النتائج أكثر وثوقيه ، لكن بالمقابل فإن ذلك يكون على حساب الدقة في جمع البيانات لذلك يفضل تحديد حجم العينة بناء على العوامل التالية:
الميزانية المتاحة، الوقت المتاح ، درجة الدقة المتاحة في ظل العاملين السابقين



مثال : 1

Sample Size

$$n = \frac{1.96^2 \sigma^2}{E^2}$$

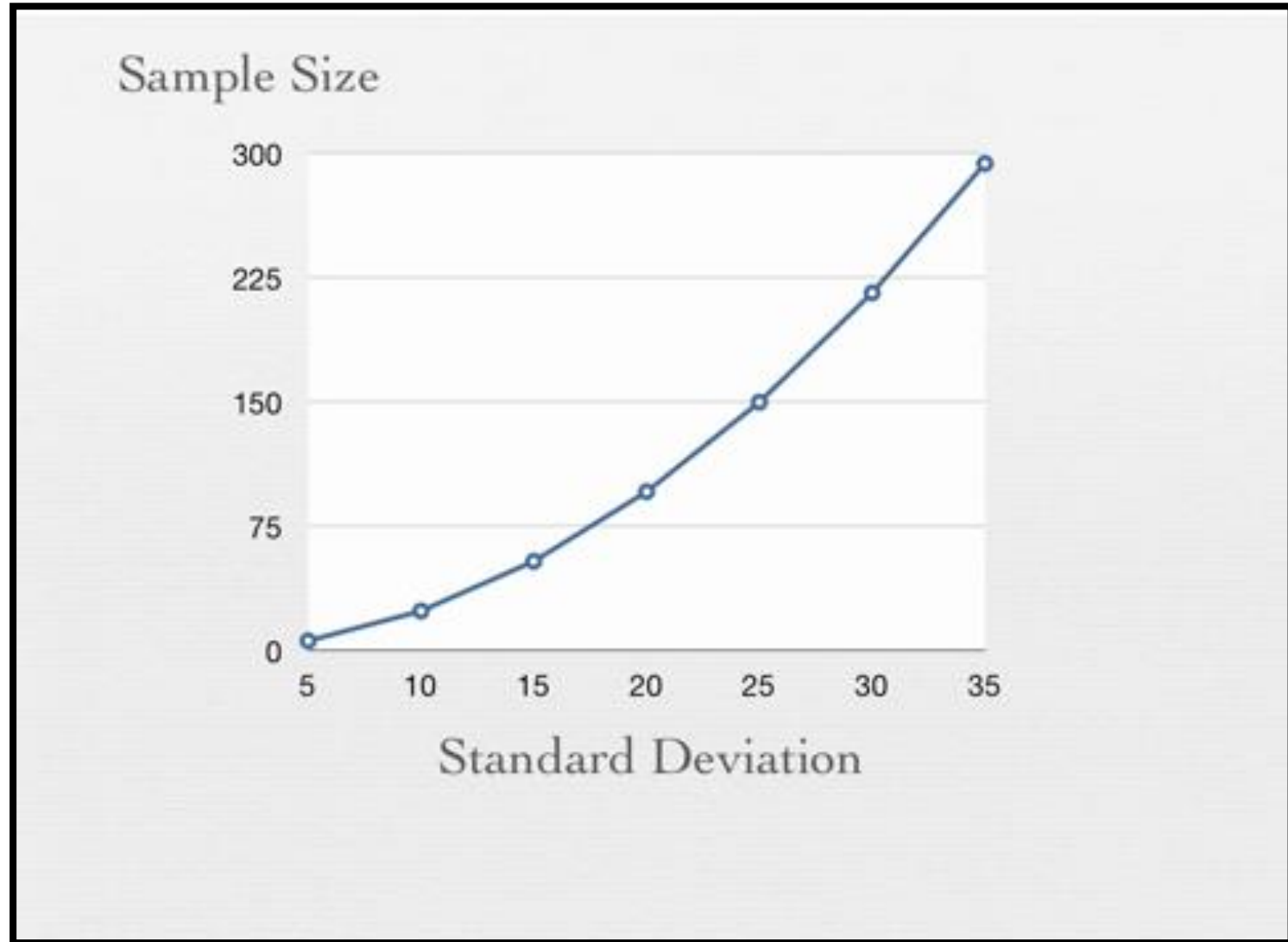
$\sigma = 20 \longleftrightarrow \sigma = 25$
 $E = 4 \qquad E = 4 \text{ units}$

sample size

$$n = \frac{3.84 \times 20^2}{4^2} = \frac{3.84 \times 400}{16} = 3.84 \times 25 = 96$$
$$n = \frac{3.84 \times 25^2}{4^2} = \frac{3.84 \times 625}{16} = 3.84 \times 39 = 150$$

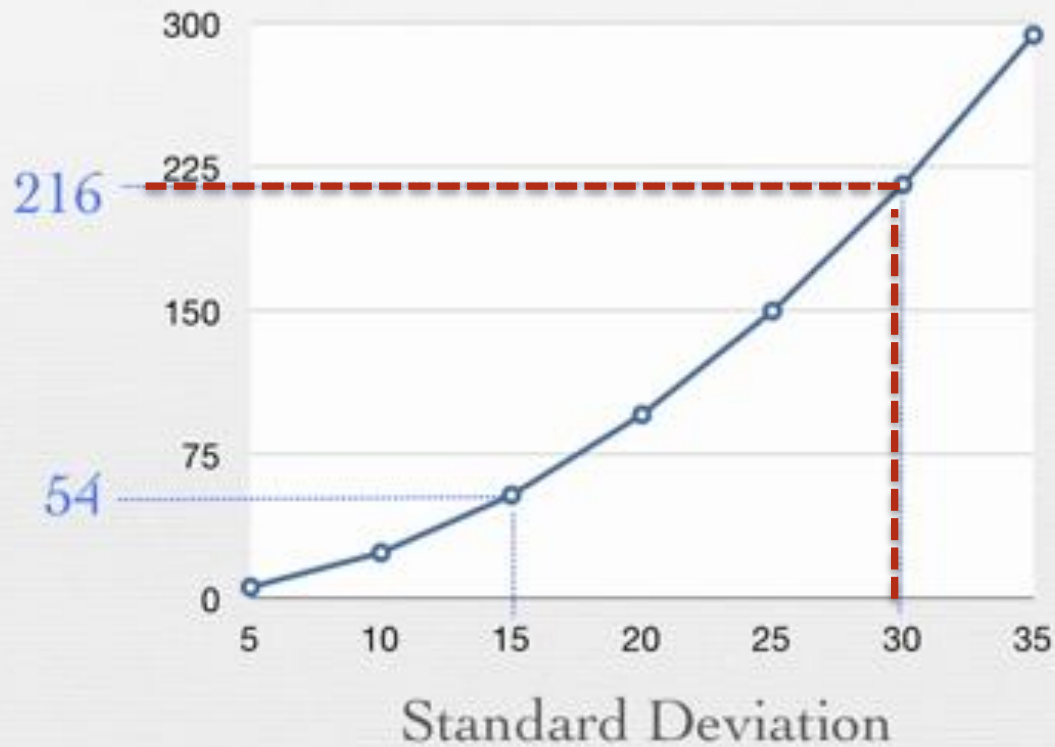


Saba Mohammed Alwan
Statistical and O.R
Faculty of Science. KSU
salwan@ksu.edu.sa





Sample Size





مثال: 2

دراسة تأثير برنامج نشاط بدني على نسبة الشحوم في الجسم لدى عينة من الأشخاص البدناء، علماً بأن الانحراف المعياري لنسبة الشحوم من دراسات سابقة على البدناء يساوي ١٠ %، ومقدار الخطأ الذي نقبل به هو ٥ %.

بتطبيق المعادلة السابقة سيكون الجواب كالتالي:



أومقدار الفرق الذي نريد أن نكتشفه هو ٥ % فأكثر

$$n = \frac{(Z_{\alpha})^2 * (S)^2}{(d)^2}$$

$$n = \frac{(1.96)^2 * (10)^2}{(5)^2}$$

$$n = \frac{(3.84) * (100)}{(25)} = 15.36$$

أي أن العدد
المطلوب هو ١٦
شخص

لكن لاحظ أن التصميم البحث الأمثل في مثل هذه الحالة ينبغي أن يتضمن مجموعة تجريبية ومجموعة ضابطة أو قياس نسبة الشحوم قبل وبعد تطبيق النشاط فالعينتين الآن مرتبطتين لذلك



: لو أخذنا المثال السابق، بجعل مجموعة ضابطة مستقلة عن المجموعة فإننا بتطبيق المعادلة أعلاه

$$n = \frac{(Z_{\alpha})^2 * 2 * (S)^2}{(d)^2}$$

سنحصل على $n = 32$

أي أننا نحتاج في هذا النوع من المعادلات ضعف العدد

بعض المعادلات الإحصائية المستخدمة

في تحديد عينة البحث

المصدر:

الهزاع، هزاع محمد. فسيولوجيا الجهد البدني: الأسس النظرية والإجراءات العملية للقياسات
الفسيولوجية. الفصل الرابع. كتاب مقدم للنشر



معادلة ستيفن ثامبسون (Steven K. Thampson) - لتحديد حجم العينة - تبعاً لحجم المجتمع

$$n = \frac{N * p(1 - p)}{\left[(N - 1) * \left(\frac{d^2}{Z^2} \right) + p(1 - p) \right]}$$

الصيغة :

حيث :

N : حجم المجتمع
Z : الدرجة العيارية لمستوى المعنوية 0.05 و مستوى الثقة 0.95 و تساوي 1.96
d : نسبة الخطأ و تساوي 0.05
p : القيمة الاحتمالية = 0.5

مثال: اذا كان حجم المجتمع 9000 بتطبيق المعادلة نحصل على حجم عينة 368.47 = 369
مثال لعينة طبقية:

اذا كان حجم المجتمع 9000 وكان منهم 5000 ذكور و 4000 اناث فكم عدد الذكور والاناث في العينة؟

حجم العينة الكلي كما سبق 369

- بضرب حجم العينة الكلي بنسبة الذكور في المجتمع ($0.56 = \frac{5000}{9000}$) نحصل على : $0.56 * 369 = 207$
- وبقيّة العينة اناث و عددهن 162 ليصبح مجموع الذكور و الاناث يساوي حجم العينة الكلي .



يريد الباحث تحديد عدد الصفائح الدموية لمجتمع المدمنين في مدينة ما ولم يجد سوى من هم في مصحة علاجية،

س: ماهي الأساليب الاحصائية الممكن استخدامها ؟

ج: لا يستطيع الباحث استخدام الأساليب من تقدير وفرضيات لأن ذلك لا يطبق الا على العينات العشوائية وليس القصدية أو العمدية .



طرق جمع البيانات

الطريقة التجريبية
Experimental method

المسح
Survey method

تبدأ هذه الطريقة بجمع البيانات عن طريق (العد، أو قياس) من العناصر تحت الدراسة بناء على خطة أو تصميم

تبدأ هذه الطريقة بالملاحظات

مثال 6

: يرغب أحد الباحثين أن يحدد العلاقة بين نوع من العلائق والزيادة في أوزان الأبقار (كيف يحصل على بيانات ؟

يقوم الباحث بأخذ عينة عشوائية من الأبقار وأخذ وزنها قبل تطبيق العليقة ثم يقوم بتغذية الأبقار على العليقة وبعد فترة زمنية مناسبة يقوم بقياس أوزان الأبقار ومن هذه القياسات قبل وبعد وبتطبيق التحليل الإحصائي المناسب يمكن أن يحدد طبيعة العلاقة

مثال 5

لاحظ أحد الباحثين أن أغلب المصابين بسرطان الرئة هم من المدخنين، فاعتقد أن هنالك علاقة بين التدخين وسرطان الرئة. ولتدعيم هذا الاعتقاد . كيف يحصل على بيانات ؟

يقوم الباحث بأخذ عينة من المصابين بسرطان الرئة ويصنفهم إلى مدخن وغير مدخن ، ويمكن باستخدام النسبة والتحليل الإحصائي لهذه النسبة أن يصل إلى قرار بهذا الشأن



قبل البدء فى التحليل الإحصائى

شروط التحليل الإحصائى قبل بداية تحليل التباين للبيانات وهى :

- = NORMALITY. الطبيعية
- = HOMOGENITY OF VARIANCES. التجانس
- = INDEPENDENCE OF MEANS AND VARIANCES
الاستقلال.



في حالة عدم توفر شرط او اكثر من شروط تحليل التباين في البيانات

NON PARAMETRIC TEST

استخدام الاختبارات اللامعلمية المقابلة
لكل حالة.
وسنأخذ بعض منها لاحقا

TRANSFORMATION OF DATA

عمل تحويل للبيانات قبل إجراء عملية التحليل
الإحصائي بإتباع طريقة من طرق التحويل ومنها:

أخذ لو غاريتم البيانات 1- LOG TRANSFORMATION

أخذ الجذر التربيعي للبيانات 2- SQUARE ROOT TRANSFORMATION

(Rare events, the data follow a Poisson distribution).

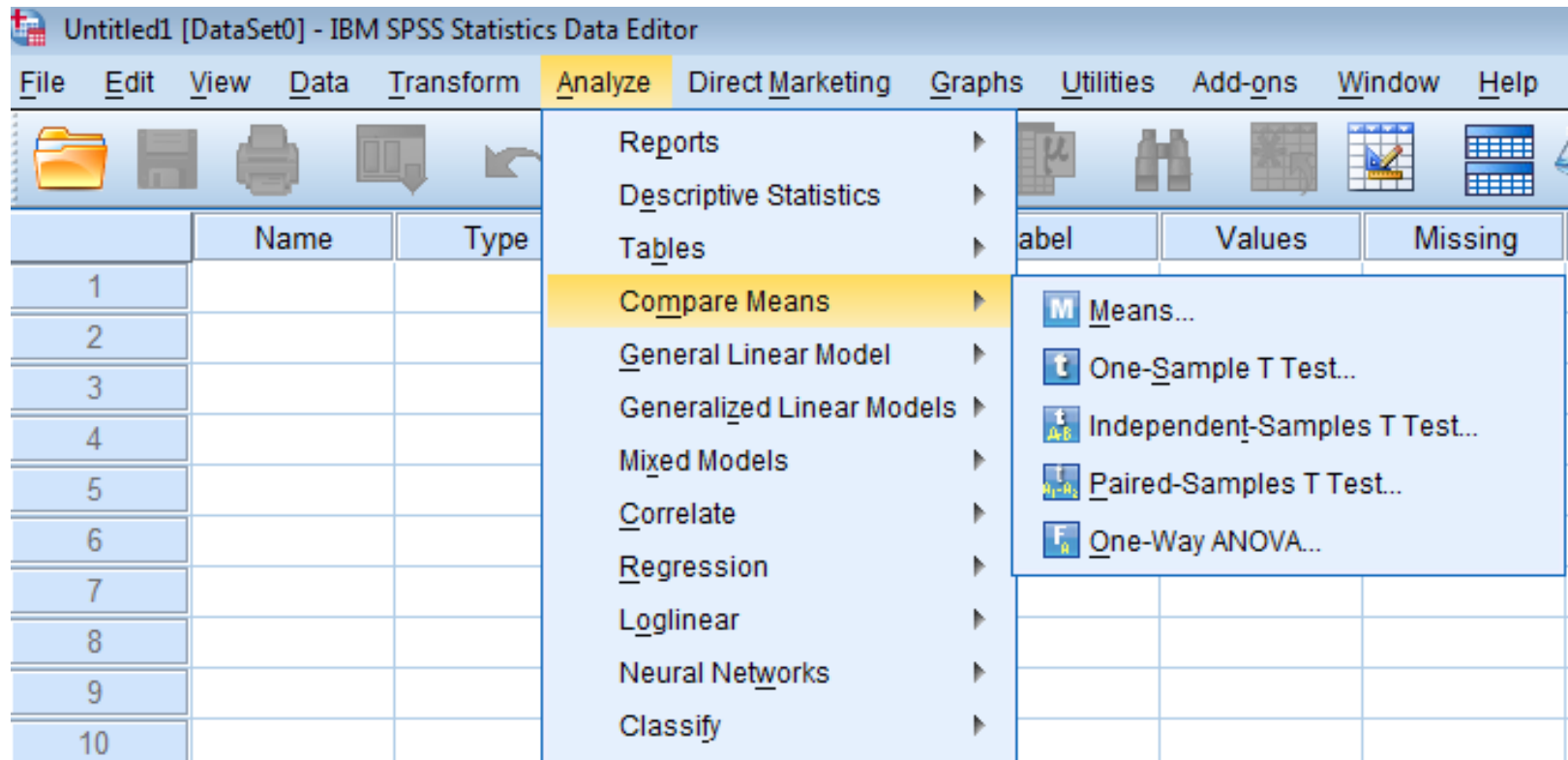
3- ARCSINE TRANSFORMATION.

(Data based on counts expressed as percentages or proportions of the total sample and followed the Binomial Distribution :variances are related to the means).



الاختبارات المعلمية

- أ. اختبار t لعينة واحدة
- ب. اختبار t لعينتين مستقلتين
- ج. اختبار t لعينتين مرتبطتين
- د. اختبار تحليل التباين الأحادي الاتجاه ANOVA
- هـ. اختبار تحليل التباين الثنائي الاتجاه MNOVA





اختبارات T (T-Test)

One Sample T-Test

Independent sample T-Test

Paired Sample T-Test

اختبار t لعينة واحدة

اختبار t لعينتين مستقلتين

اختبار t لعينتين مرتبطتين



اختبر صحة الفرض الذي ينص علي :

لا توجد فروق دالة احصائية عند مستوى دلالة 0.05 بين متوسط أداء طلاب كلية التجارة في مادة الحاسوب (والمتوسط الفرضي لمادة الحاسوب هو ٢٣) .

سيقوم الباحث:

بأخذ عينة عشوائية من طلاب كلية التجارة الدارسين لمادة الرياضيات حجمها أكبر من ٣٠ (لتحقق الطبيعية في البيانات). وتسجيل درجاتهم تحت اسم computer

ماهو التحليل (الاختبار) الإحصائي المناسب لاختبار صحة هذا الفرض

(One Sample T-Test)

اختبار T للعينة الواحدة

يستخدم هذا الاختبار لفحص فرضية تتعلق بالوسط الحسابي للمجتمع. وما إذا كان متوسط المجتمع يساوي قيمة محددة أو أكبر أو أصغر منها، ويجب تحقق الشرطين التاليين:

١. يجب أن يتبع توزيع المتغير التوزيع الطبيعي، ويستعاض عن هذا الشرط بزيادة حجم العينة إلى أكثر من ٣٠ مفردة.

٢. يجب أن تكون العينة عشوائية أي لا تعتمد مفرداتها على بعضها (وهذا محقق عند جمع البيانات)



لم يجد الباحث سوى ١٥ طالب ووجد أن بياناتهم لا تتوزع توزيعاً طبيعياً ، .. فهل يستخدم نفس الاختبار؟

SIGN TEST

اختبار الإشارة

هذا الاختبار اللامعلمي هو المقابل في حال أن البيانات لا تتوزع طبيعياً ونصل إليه من خلال المسار

Analyze -- Nonparametric tests-- Binomial

ادخال المتغيرات

ادخال القيمة المراد اختبار مساواة الوسط لها في مربع Cut point

ثم Ok



من قائمة التحليل Analyze اختر Compare Means ثم One Sample t-test

36 : math								
	math	var	var	var	var	var	var	var
14	36							
15	37							
16	37							
17	37							
18	39							
19	28							
20	29							
21	29							
22	28							
23	27							
24	26							
25	30							
26	30							
27	27							
28								

One-Sample T Test

Test Variable(s):
degrees [math]

Test Value: 23

Options...
Bootstrap...

OK Paste Reset Cancel Help

انقل المتغير
إلى computer
(جهة اليمين) تحت
نافذة Test
Variables

كتب قيمة المتوسط
الفرض لمادة
الحاسوب (والذي =
23) في مربع Test
Value

One-Sample T Test: Options

Confidence Interval Percentage: 95 %

Missing Values

☒ Exclude cases analysis by analysis
☐ Exclude cases listwise

Continue Cancel Help

وعند الضغط علي زر [Option...]
بالفارة سوف تظهر الشاشة التالية :

ثم continue
ثم ok



→ T-Test

[DataSet0]

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
degrees	35	30.46	4.648	.786

One-Sample Test

	Test Value = 23					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
degrees	9.491	34	.000	7.457	5.86	9.05

الفرضية الصفرية: المتوسط = 23
الفرضية البديلة: المتوسط لا يساوي 23



أن أحد الباحثين يريد أن يقف علي فاعلية أحد جرعتين مختلفتين من دواء خفض ضغط الدم (قرصين يوميا ، وثلاثة أقراص يوميا) علي عينة من المرضى المتطوعين قوامها (٣٦) مريضا بارتفاع ضغط الدم .

قام الباحث بتسجيل قيمة الانخفاض في ضغط الدم لعينة المرضى المتطوعين بعد أخذ الجرعة الأولى من الدواء وكذا بعد أخذ الجرعة الثانية من الدواء بفواصل زمني قدره أسبوعين . فحصل على البيانات التالية:

SN	doss1	doss2
1	45	45
2	44	43
3	40	40
4	30	30
5	34	33
6	35	35
7	36	34
8	37	37
9	37	36
10	38	38
11	39	38
12	40	40
13	40	39
14	40	39
15	40	40
16	32	32
17	33	32
18	34	33

وكان لدى الباحث الفرض
الصفرى التالى :

لا توجد فروقا عند
مستوى دلالة 0.01
أو (0.05) بين متوسطات
انخفاض ضغط الدم لدى
أفراد العينة من المرضى
في الحالتين (الجرعة
الأولى ، والجرعة الثانية).

ماهو التحليل
الإحصائى المناسب
لاختبار صحة هذا
الفرض



ماهو التحليل الإحصائي المناسب لاختبار صحة هذا الفرض

(Paired Sample T-Test)

اختبار T للعينات المرتبطة

يستخدم هذا الاختبار في فحص الفرضيات المتعلقة بمساواة متوسط متغيرين لعينتين غير مستقلتين .
وتكتب الفرضية المبدئية والبديلة بالطريقة التالية

شروط استخدام الاختبار:

١. يجب أن يتبع توزيع الفرق بين المتغيرين طبيعياً.
٢. يجب أن تكون العينات عشوائية و مستقلة و حجمهما صغير بتباين متساوي

الفرضية المبدئية: $H_0 : \mu_1 = \mu_2$

لفرضية البديلة: $H_a : \mu_1 \neq \mu_2$

الاختبار اللامعلمي البديل في
حال عدم توفر الشروط او احدها

2 Related Samples

هذا الاختبار اللامعلمي هو المقابل في حال أن البيانات لا تتوزع طبيعياً

Analyze -- Nonparametric tests--2 related samples

ادخال المتغيرات

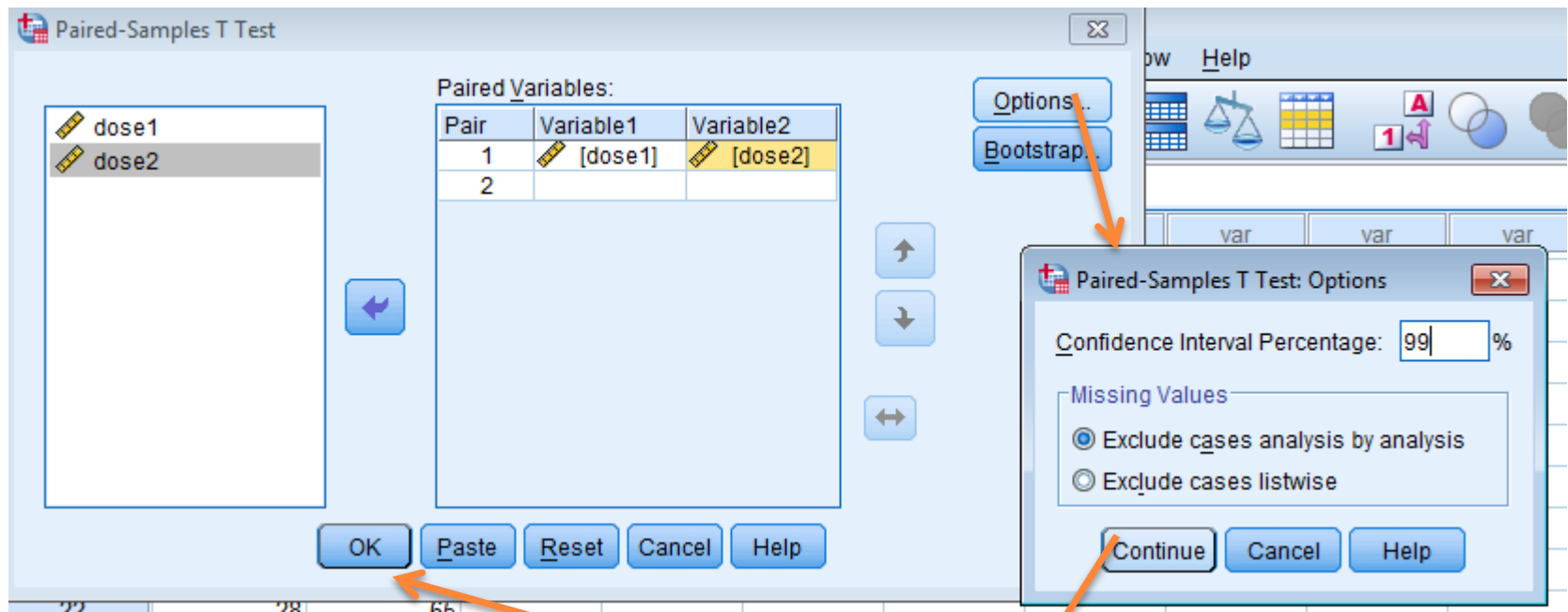
واختبار Sign و Wilcoxon

ok



الخطوات باستخدام SPSS

من قائمة التحليل Analyze اختر Compare Means ثم Paired Sample t-test سوف تفتح النافذة القافزة التالية



٢٢



Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	99% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	dose1 - dose2	-.257-	10.254	1.733	-4.986-	4.472	-.148-	34	.883

نقبل الفرض الصفري
ونقبل بوجود فروق بين
المتوسطات

أكبر من 0.01

القرار الاحصائي لا يوجد فروق دالة احصائية عند مستوى الدلالة 0.01 بين فاعلية الجرعتين علي مرضى ضغط الدم.



بفرض أن أحد الباحثين يريد أن يقف علي فاعلية التعليم التعاوني في تدريسه لمادة الحاسوب للفرقة الثانية الثانوية علي تحصيل الطلاب وبقاء اثر التعلم . تكونت عينة البحث من (٨١ طالبا موزعين عشوائيا علي مجموعتين (تجريبية : وبها ٩ طالبا ، وضابطة: وبها ٩ طالبا) . درست المجموعة التجريبية المحتوى بطريقة التعليم التعاوني بينما درست المجموعة الضابطة بالطريقة التقليدية .

قام الباحث بتطبيق اختبار تحصيلي علي طلاب المجموعتين بعد الانتهاء من تدريس المحتوى لقياس تحصيل الطلاب Ach كما قام الباحث بتطبيق نفس الاختبار بفارق زمن ثلاثة أشهر لقياس بقاء أثر التعلم للطلاب Ret



تكويد
متغير
المعالجات
Treatment
(التجريبية
'
الضابطة)
ب (1, 2)

درجات
الطلاب
في
الاختبار
التحصيلي
ي Ach

اختبار
بقاء أثر
التعلم
Ret

رقم
الطالب

قام الباحث بتكويد
متغير المعالجات

Treatment

(التجريبية ،
الضابطة)

بالأرقام (1, 2) .

ثم قام بإدخال

درجات الطلاب

في الاختبار

التحصيلي Ach

ودرجاتهم في

اختبار بقاء أثر

التعلم Ret

مستخدما محرر

بيانات الحزمة

الإحصائية

SPSS : علي

النحو التالي :

SN	tret	Ach	Ret
1	1	45	45
2	1	44	43
3	1	40	40
4	1	30	30
5	1	34	33
6	1	35	35
7	1	36	34
8	1	37	37
9	1	37	36
10	2	38	38
11	2	39	38
12	2	40	40
13	2	40	39
14	2	40	39
15	2	40	40
16	2	32	32
17	2	33	32
18	2	34	33

وكان لدى الباحث الفرض
الصفري التالي : " لا توجد
فروقا دالة إحصائية بين
متوسطات تحصيل طلاب
المجموعتين (التجريبية
والضابطة) في معدل
الاحتفاظ بالتعلم" .



الفرضية المبدئية: $H_0 : \mu_1 = \mu_2$
لفرضية البديلة: $H_a : \mu_1 \neq \mu_2$

ماهو التحليل الإحصائي
المناسب لاختبار صحة
هذا الفرض

لاحظ أن
العينات مختلفة
ولاتؤثر أحدهما على
الأخرى



اختبار T للعينات المستقلة

(Independent sample T-Test)

• شروط اختبار T للعينات المستقلة

لضمان دقة نتائج اختبار T يجب أن تتوافر الشروط الثلاثة التالية:

1. يجب أن يكون متغير الاختبار طبيعياً في كل فئة من فئات متغير التجميع
2. يجب أن يكون تباين متغير الاختبار متساوياً في كلا فئتي متغير التجميع، وإذا لم يتحقق هذا الشرط فإن نتيجة اختبار T غير دقيقة، وفي هذه الحالة يمكن حساب قيمة تقديرية للإحصائي T لا يشترط لها مساواة التباين للعينتين.
3. يجب أن تكون العينة عشوائية، ويجب أن تكون قيم متغير الاختبار مستقلة عن بعضها.



الاختبار اللامعلمي البديل في
حال عدم توفر الشروط أو احدها

اختبار مان - وتني (Mann-Whitney test (U-Test

هو المقابل في حالة عدم التأكد من أن توزيع

العينتين طبيعياً وكذلك تباين المجتمعين متساويين، أو أن تكون البيانات المأخوذة من العينتين غير دقيقة أو تعتمد على ترتيب عناصر العينتين من حيث القيمة.

Analyze -- Nonparametric tests--2 independent samples

ادخال المتغيرات اضغط على Define Groups وادخل الرقم ١ داخل المستطيل المقابل ل Group1

وادخل الرقم ٢ داخل المستطيل المقابل ل Group2

ثم Continue ثم ok



مستوى الدالة	قيمة	الخطأ المعياري	الانحراف المعياري	متوسط الاحتفاظ	العدد	مجموعات المقارنة
0.214	0.1249	0.009	0.667	0.38	50	التجريبية
		0.157		0.125	72	الضابطة

يتضح من هذا الجدول (حيث مستوى الدلالة $(0.2 > 0.01)$ أي نقبل الفرض الصفري وعليه لا توجد فروقا دالة إحصائية بين متوسطات طلاب المجموعتين (التجريبية والضابطة) في معدل الاحتفاظ بالتعلم . وعلي الباحث قبول الفرض.



بالمسبة للمثال في البيانات باسم : Independent Sample t test

Untitled1 [DataSet0] - IBM SPSS Statistics Data Editor

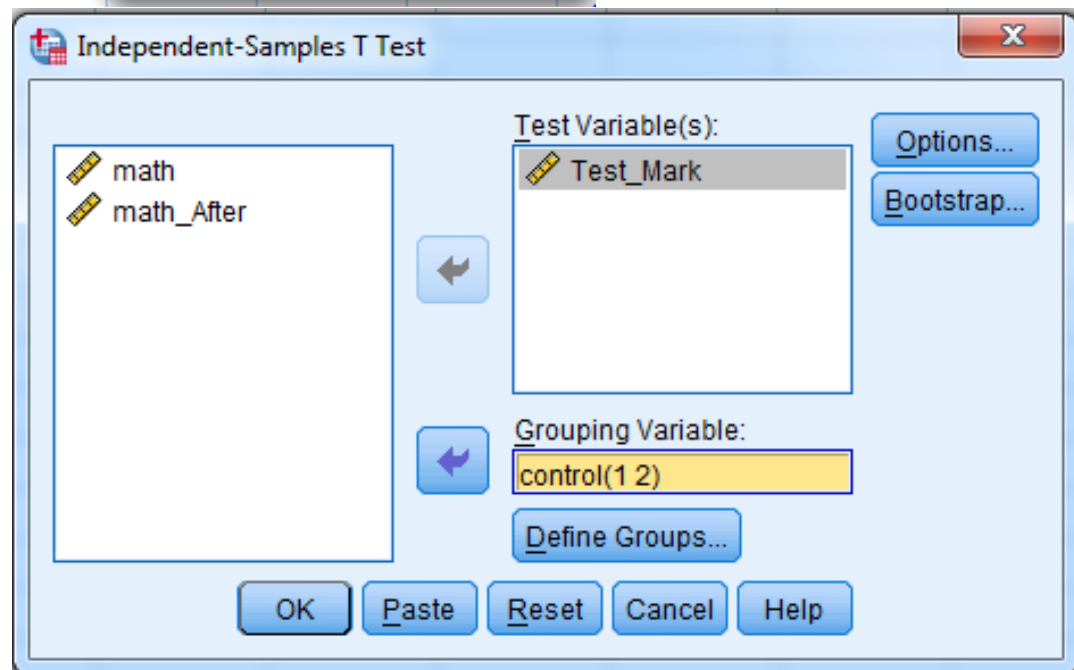
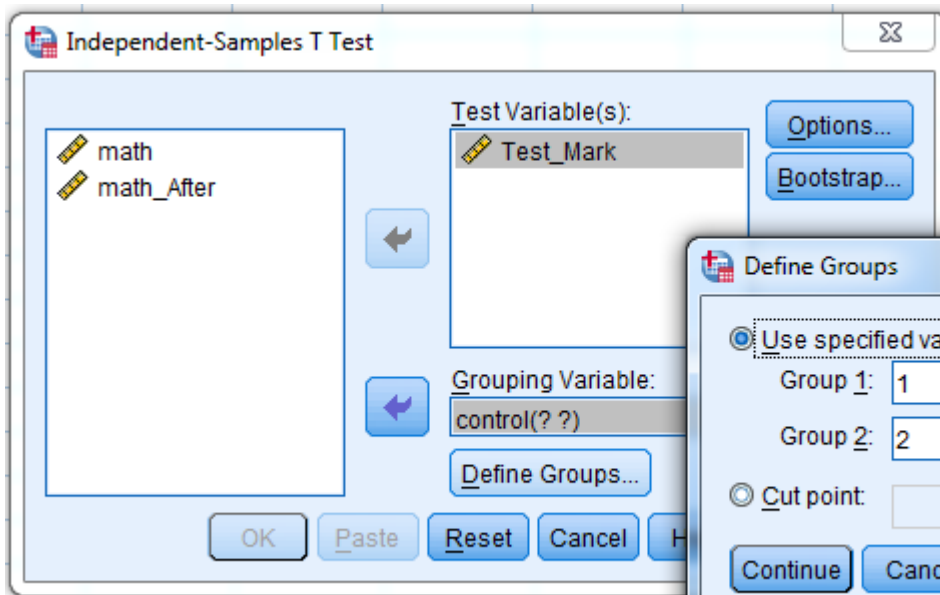
File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

Reports
Descriptive Statistics
Tables
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear
Neural Networks
Classify
Dimension Reduction
Scale
Nonparametric Tests

Label Values Missing

M Means...
t One-Sample T Test...
t Independent-Samples T Test...
t Paired-Samples T Test...
F One-Way ANOVA...

	Name	Type
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		
11		
12		
13		



Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Test_Mark	Equal variances assumed	2.506	.118	-1.034-	73	.305	-8.275-	8.002	-24.223-	7.673
	Equal variances not assumed			-1.106-	39.374	.276	-8.275-	7.483	-23.407-	6.857

يتضح من هذا الجدول أي نقبل الفرض الصفري وعليه لا توجد فروقا دالة إحصائية بين متوسطات طلاب المجموعتين (التجريبية والضابطة) في معدل الاحتفاظ بالتعلم . وعلي الباحث قبول الفرض.



أراد أحد المهتمين ببيانات التعدادات السكانية للدول باختبار ما إذا كان متوسط أعمار الذكور غير المتزوجين يساوي متوسط أعمار الإناث غير المتزوجات في كل مجتمع على حدة .

ما هو الأسلوب الإحصائي الصحيح للإجابة على تساؤل الباحث؟.

هنا الباحث لا يحتاج إلى وضع فرضيات ليستدل بنتيجة اختبارها حول المجتمع بل مقارنة **صريحة** بالمتوسطات فلا داعي في حال الحصر شامل (كما في مثالنا) للأساليب الاستدلالية الإحصائية من تقدير واختبارات للفروض. بل يستخدم الباحث **فقط** الأساليب الإحصائية الوصفية .

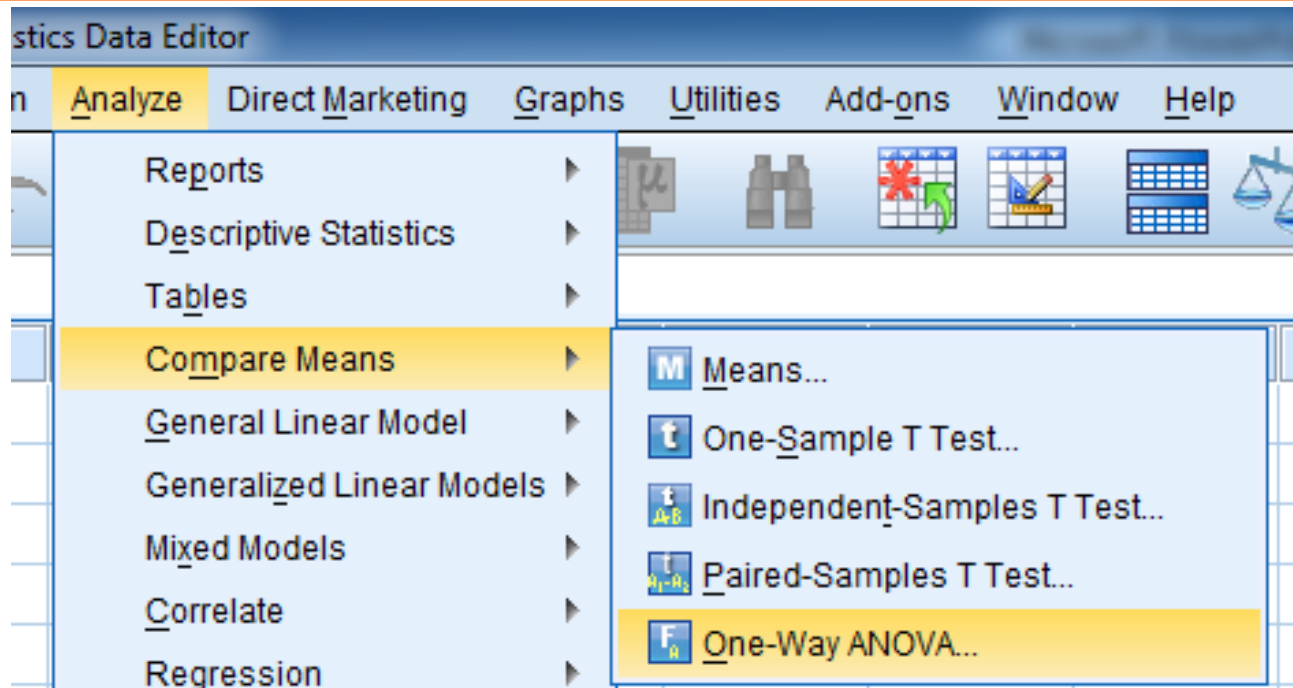


تحليل التباين (ANOVA) Analysis of Variance

إن مقارنة عدة مجتمعات في الوقت ذاته توفر الوقت والجهد والتكاليف فيما لو قارنا بين كل مجتمعين اثنين على حدة المصطلح " النموذج الكامل العشوائية " مرادف لأخذ عينات عشوائية مستقلة من عدة مجتمعات.

ANOVA:

هو أسهل أنواع تحليل التباين وهو التصنيف الأحادي حيث تصنف المشاهدات إلى عدة مجموعات على أساس معامل واحد





الاختبارات اللازمة قبل تحليل التباين :

- اختبار الطبيعية لكل عينة (إذا كل عينه حجمها أقل من 30)
- اختبار التجانس
- العشوائية (مأخوذة في الاعتبار)
- الاستقلال : العينات مستقلة بطبيعة الحال (محقق)
- القياسات كمية: كمية بطبيعة الحال (محقق)

- نستغني عن اختبار الطبيعية (إذا كان كل عينه حجمها أكثر من 30)
- اختبار التجانس يتم فقط في حالة تحليل التباين.
- بقية الشروط لا يتم اختبارها لانه لابد أن تؤخذ في الاعتبار عند تصميم الدراسة وأخذ العينات.



اختبار التباين الغير معلمي

اختبار كروسكال-والس (H-Test) Wallis- Kruskal

يستخدم هذا الاختبار عندما يكون حجم العينات صغيرا أو لا يتبع للتوزيع الطبيعي

الاختبار الالامعلمي البديل في
حال عدم توفر الشروط او احدها



اختبار كروسكال-والس (H-Test) Wallis- Kruskal

يستخدم هذا الاختبار عندما يكون حجم العينات صغيرا أو لا يتبع للتوزيع الطبيعي

Analyze -- Nonparametric tests--K independent samples

ادخال المتغيرات اضغط على Define variable وادخل الرقم ١ داخل المستطيل المقابل ل

minimum وادخل الرقم ٣ داخل المستطيل المقابل ل maximum

ثم Continue ثم Krouskal-Wallis H ثم ok



مثال: 13

استخدم 22 عجلا متماثلا في العمر والوزن في تجربة لدراسة إضافة أربعة مستويات من الفيتامينات A,B,C,D, على الزيادة في الوزن. 5 عجول اختيرت عشوائيا للمعاملة A و 6 اختيرت عشوائيا للمعاملة B، 6 اختيرت عشوائيا للمعاملة C و ال 5 المتبقية للمعاملة D. وكانت النتائج كالتالي:

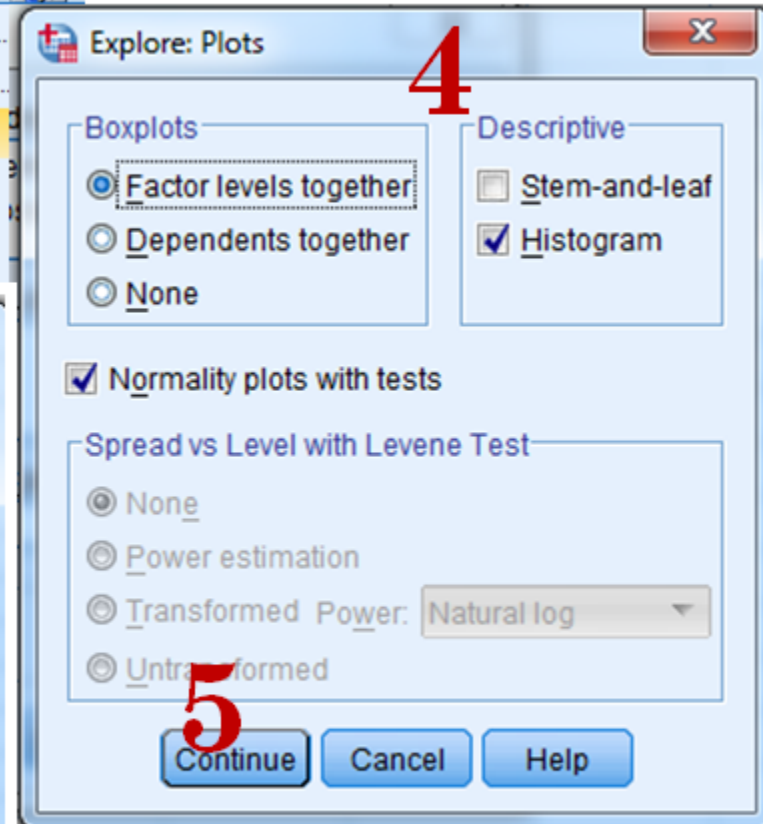
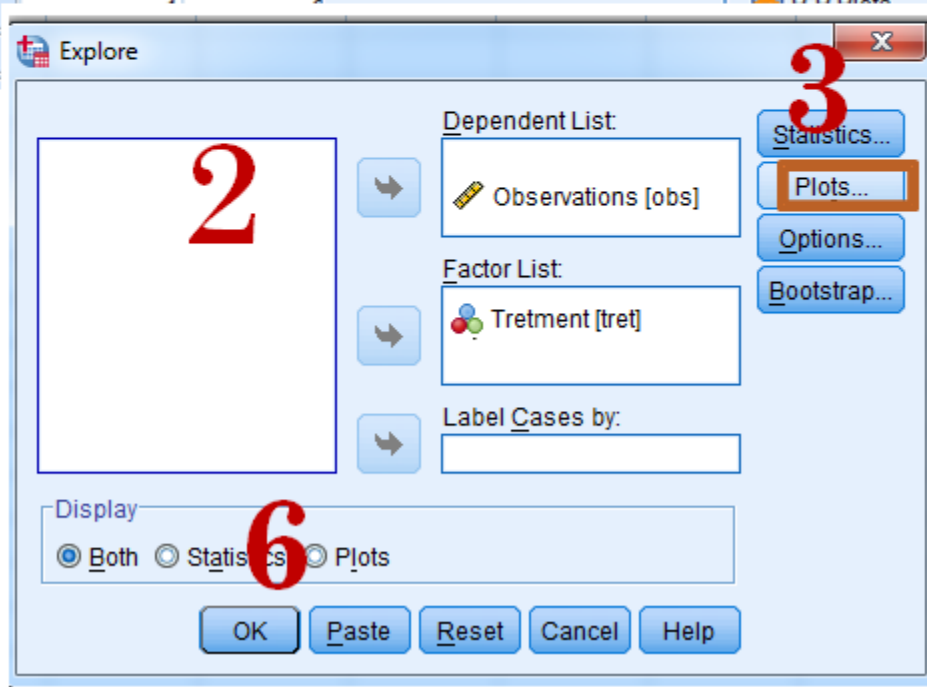
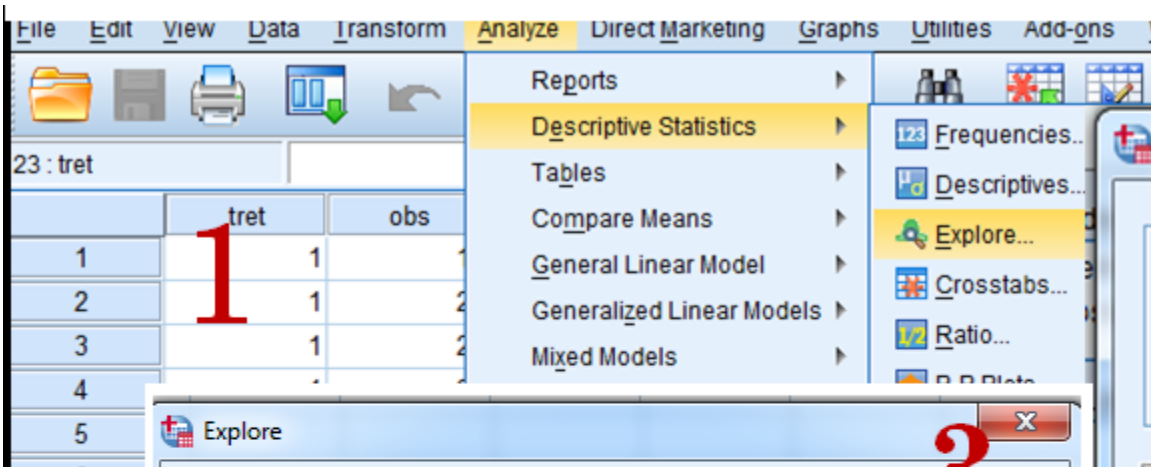
Treatment	الزيادة في الوزن Observations						n
A	18	24	26	21	19		5
B	13	17	13	14	16	14	6
C	18	16	13	21	14	11	6
D	20	26	21	19	24		5
							N=22

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$ (at level of sig. 0.05)

H_1 : على الأقل متوسطين مختلفين

Normality test

اختبار الطبيعية



Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Observations	.140	22	.200 [*]	.949	22	.295

كلاهما
أكبر من
0.05

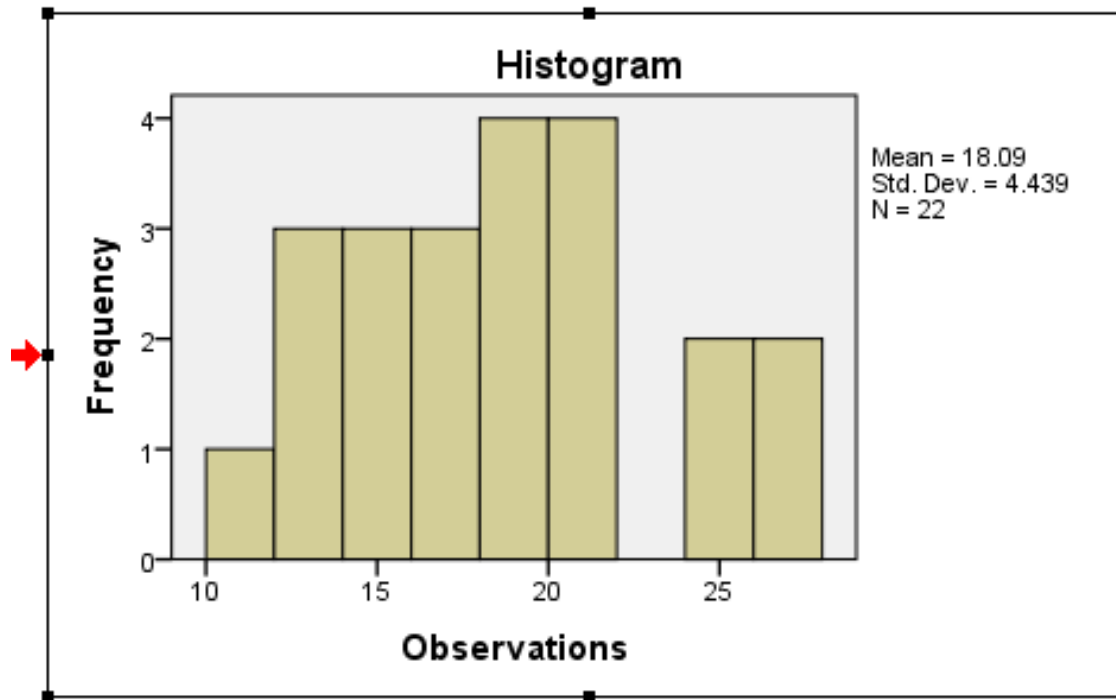
*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

الفرضية الصفرية : البيانات تخضع للتوزيع الطبيعي
الفرضية البديلة : البيانات لا تخضع للتوزيع الطبيعي

Observations

إذن نقبل أن البيانات تتوزع طبيعياً



Shapiro-Wilk test is a specific test for normality, whereas the method used by Kolmogorov-Smirnov test is more general, but less powerful (meaning it correctly rejects the null hypothesis of normality less often).

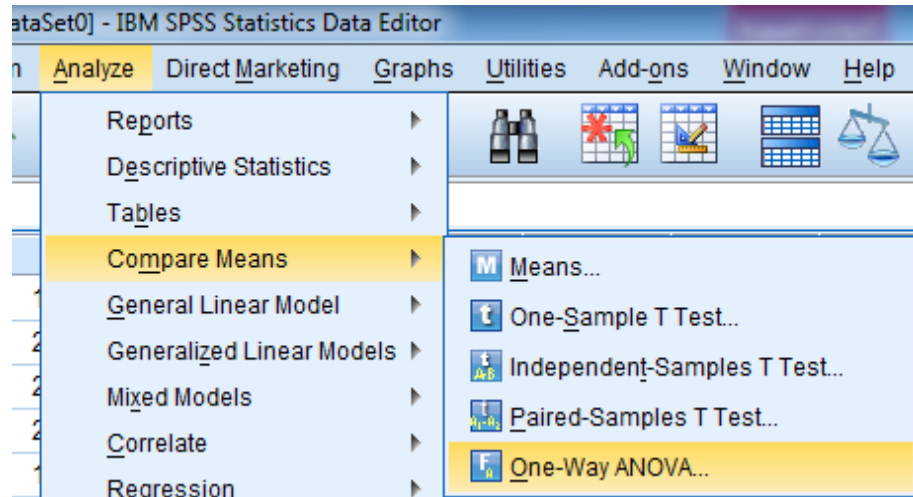
They do so is different from one another in ways that make them more or less sensitive to features of normal distributions.

واختبار شابيرو-ويلك هو اختبار محدد لاختبار الطبيعية، أما كولموغوروف-سميرنوف أكثر عمومية، ولكن أقل قوة (بمعنى أنها ترفض فرضية العدم أقل في كثير من الأحيان بشكل صحيح).

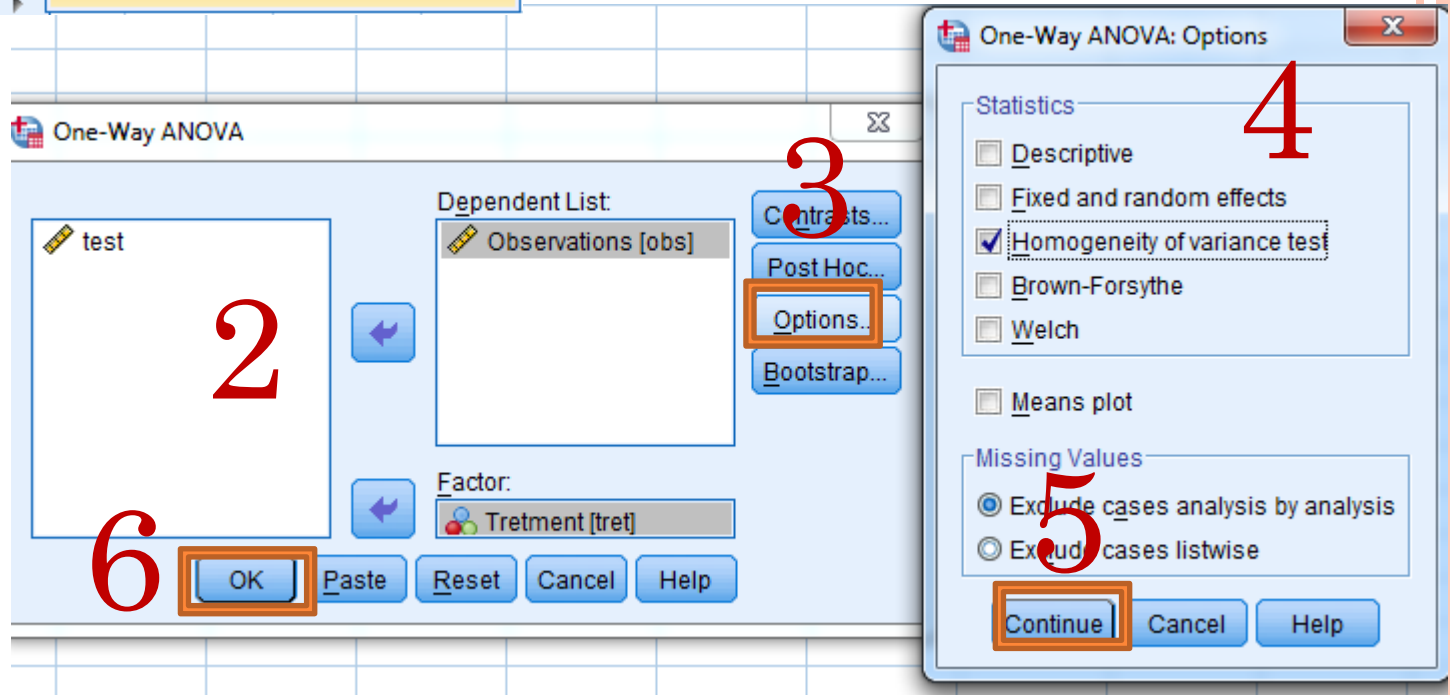
هي مختلفة عن بعضها البعض بطرق تجعلها أكثر أو أقل حساسية لخصائص التوزيع الطبيعي.

Homogeneity test

اختبار التجانس



الفرضية الصفرية : البيانات للأربع عينات متجانسة
الفرضية البديلة : البيانات غير متجانسة





Test of Homogeneity of Variances			
Observations			
Levene Statistic	df1	df2	Sig.
1.473	3	18	.255

أكبر من
0.05

إذن نقبل أن البيانات متجانسة



ANOVA

Observations

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	255.618	3	85.206	9.695	.000
Within Groups	158.200	18	8.789		
Total	413.818	21			

$$P\text{-value} = 0.000 < \alpha = 0.05$$

نرفض الفرضية الصفرية ونقبل الفرضية البديل والتي تقضي بعدم تساوي المتوسطات بمعنى أنه على الأقل متوسطين مختلفين

أي من هذه المتوسطات مختلف ؟؟؟؟



لا نقوم بها إلا بعد معرفة نتيجة تحليل التباين ورفض الفرضية العدم، وقبول وجود اختلاف دال احصائياً

Multiple Comparison

المقارنات المتعددة

One-Way ANOVA: Post Hoc Multiple Comparisons

Equal Variances Assumed

☒ LSD ☐ S-N-K ☐ Waller-Duncan
☐ Bonferroni ☐ Tukey Type I/Type II Error Ratio: 100
☐ Sidak ☐ Tukey's-b ☐ Dunnett
☐ Scheffe ☐ Duncan Control Category: Last
☐ R-E-G-W F ☐ Hochberg's GT2
☐ R-E-G-W Q ☐ Gabriel

Test
☒ 2-sided ☐ < Control ☐ > Control

Equal Variances Not Assumed

☐ Tamhane's T2 ☐ Dunnett's T3 ☐ Games-Howell ☐ Dunnett's C

Significance level: 0.05

Continue Cancel Help

في حال التجانس

في حال عدم التجانس

Multiple Comparisons

Dependent Variable: Over wight

			Mean Difference (I- J)	Std. Error	Sig.	95% Confidence Interval	
(I) Vitamen type	(J) Vitamen type	Lower Bound				Upper Bound	
Tukey HSD	A	B	9.267 [*]	2.521	.009	2.14	16.39
		C	6.100	2.521	.109	-1.03-	13.23
		D	-.400-	2.633	.999	-7.84-	7.04
	B	A	-9.267- [*]	2.521	.009	-16.39-	-2.14-
		C	-3.167-	2.404	.564	-9.96-	3.63
		D	-9.667- [*]	2.521	.006	-16.79-	-2.54-
	C	A	-6.100-	2.521	.109	-13.23-	1.03
		B	3.167	2.404	.564	-3.63-	9.96
		D	-6.500-	2.521	.081	-13.63-	.63
	D	A	.400	2.633	.999	-7.04-	7.84
		B	9.667 [*]	2.521	.006	2.54	16.79
		C	6.500	2.521	.081	-.63-	13.63
LSD	A	B	9.267 [*]	2.521	.002	3.97	14.56
		C	6.100 [*]	2.521	.026	.80	11.40
		D	-.400-	2.633	.881	-5.93-	5.13
	B	A	-9.267- [*]	2.521	.002	-14.56-	-3.97-
		C	-3.167-	2.404	.204	-8.22-	1.88
		D	-9.667- [*]	2.521	.001	-14.96-	-4.37-
	C	A	-6.100- [*]	2.521	.026	-11.40-	-.80-
		B	3.167	2.404	.204	-1.88-	8.22
		D	-6.500- [*]	2.521	.019	-11.80-	-1.20-
	D	A	.400	2.633	.881	-5.13-	5.93
		B	9.667 [*]	2.521	.001	4.37	14.96
		C	6.500 [*]	2.521	.019	1.20	11.80

*. The mean difference is significant at the 0.05 level.



يعتمد اختيار الاختبار البعدي على طبيعة الموقف الذي يكون الباحث فيه وفقا لبياناته وطبيعتها من حيث حجوم العينة، وهل هي متساوية أم لا، هل تتوزع توزيعا طبيعيا أم لا ، تساوي أو اختلاف التباينات :

* عندما تكون المقارنات المتعددة صغيرة === فيمكننا استخدام

Bonferroni

فمثلا

* عندما تكون المقارنات المتعددة كبيرة === فيمكننا استخدام

Tukey

يعتبر أكثر قوة ، كما أن هذه الطريقة أقوى إحصائيا من **Dunn & Scheffe**

* إذا كانت حجوم العينات متساوية، وتباينات المجتمعات متساوية === فيمكننا استخدام

Tukey HSD

أو

R-E-G-W-Q

فكلاهما قويتان إحصائيا كما أنهما يضبطان وعلى نحو كبير الخطأ من النوع الأول ألفا

* إذا كانت حجوم العينات مختلفة بعض الشيء === فيمكن استخدام

فسيكون الاختبار الأقوى إحصائيا في هذه الحالة

Gabriel



* إذا كانت أحجام العينات مختلفة بشكل كبير، === فيمكن استخدام

Hochberg GT2

* إذا كانت تباينات المجتمعات غير متساوية === فيجب استخدام

Games - Howell

ملاحظة: معظم هذه الاختبارات التتبعية تكون جيدة في حال أن البيانات تتوزع توزيعاً طبيعياً، وتباين المجتمعات متساوٍ، وأحجام العينات متساوية....

لذا **نصيحة:** على الباحث محاولة توفير هذه الشروط في بياناته قدر الإمكان حتى يحصل على نتائج دقيقة علماً أن التصميم المسبق للتجربة لجمع البيانات يوفر على الباحث الكثير من الوقت وتعفيه من الدخول في مثل هذه الحالات.

فالعشوائية والاستقلال يجب أن تؤخذ في الاعتبار عند تصميم أي دراسة إحصائية



تحليل التباين في اتجاهين- Two ways Analysis of Variance

تأثير مشترك أو تفاعلي

فهو ناتج عن تفاعل المتغيرين المستقلين ويكون هناك تفاعل بينهم حال اختلاف تأثير أي منهم يختلف باختلاف مستويات الآخر.

أنواع التأثيرات
للمتغير المستقل

تأثير أساسي

يقصد به تأثيره على المتغير التابع بصرف النظر عن المتغير المستقل الآخر

جدول تحليل التباين بمعيارين مع التفاعل الداخلي

Two-Ways Analysis of
Variance with internal
interaction

جدول تحليل التباين بمعيارين مع عدم وجود تفاعل

داخلي
Two-Ways Analysis of
Variance without internal
interaction



باحث ما مهتما في ما إذا كان اهتمام الفرد بالسياسة يتأثر بمستوى التعليم وبالجنس. .. جمعت عينة عشوائية وتم سؤالهم عن اهتمامهم بالسياسة، والتي سجلت على مقياس من 0-100، (أعلى الدرجات يعني اهتماما أكبر) الباحث قام بتقسيم المشاركين حسب نوع الجنس (ذكر / أنثى) وبعد ذلك مرة أخرى حسب المستوى التعليمي (المدرسة / الكلية / الجامعة).

وعليه فإن : المتغير التابع «الاهتمام بالسياسة» وكانت المتغيرات المستقلة: "الجنس" و "التعليم".

		مستوى التعليم Edu_Level		
		مدرسة School	كلية College	جامعي University
Gender	ذكر Male	34	49	60
		35	50	56
		32	47	78
	
	أنثى Female	34	44	57
		33	56	76
	

فما هو الاسلوب الاحصائي المناسب ؟



	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		

Male=1
Female=2

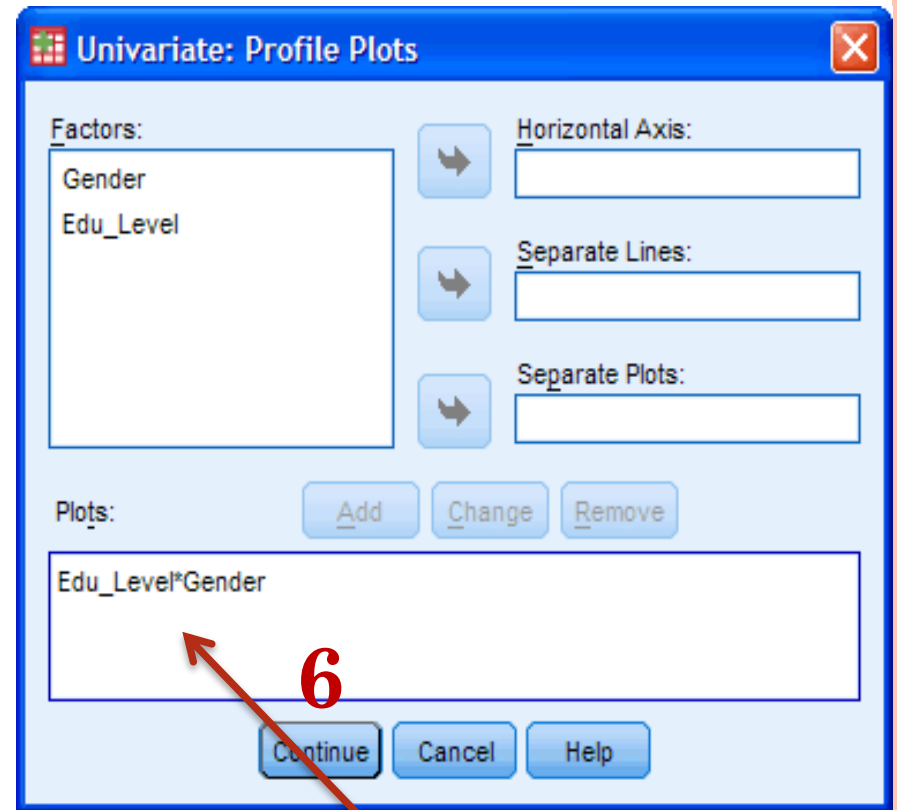
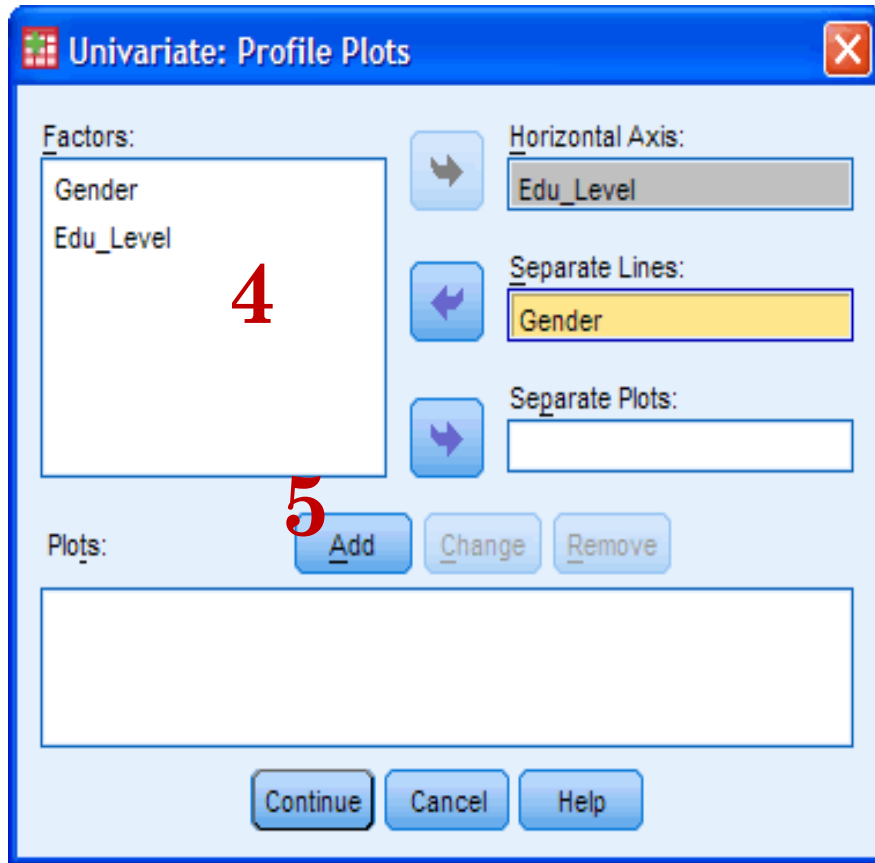
School =1
College =2
University =3

بعد ادخال البيانات والتكويد كما سبق



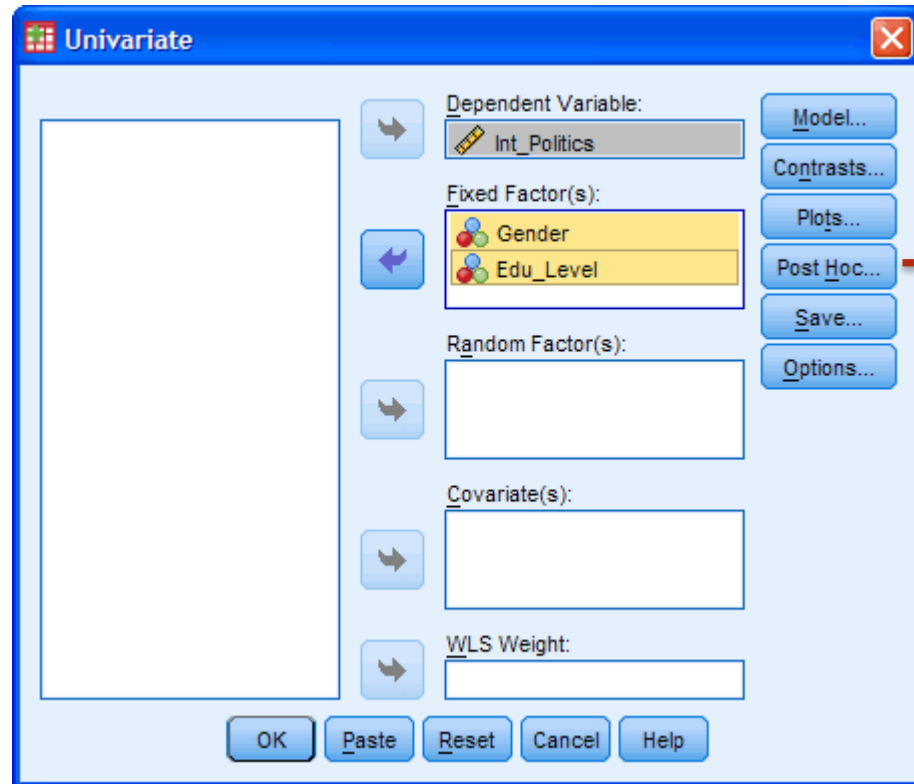
Analyze > General Linear Model > Univariate... on the top menu

The image shows two parts of the SPSS interface. On the left, the 'Analyze' menu is open, and the path 'General Linear Model > Univariate...' is highlighted. A red number '1' is placed over the data grid. On the right, the 'Univariate' dialog box is open. The 'Dependent Variable' is 'Int_Politics'. The 'Fixed Factor(s)' are 'Gender' and 'Edu_Level'. A red number '2' is in the empty box on the left of the dialog. On the right side of the dialog, a vertical list of buttons includes 'Model...', 'Contrasts...', 'Plots...', 'Post Hoc...', 'Save...', and 'Options...'. A red number '3' is next to the 'Plots...' button, with an arrow pointing to it. At the bottom of the dialog are buttons for 'OK', 'Paste', 'Reset', 'Cancel', and 'Help'.



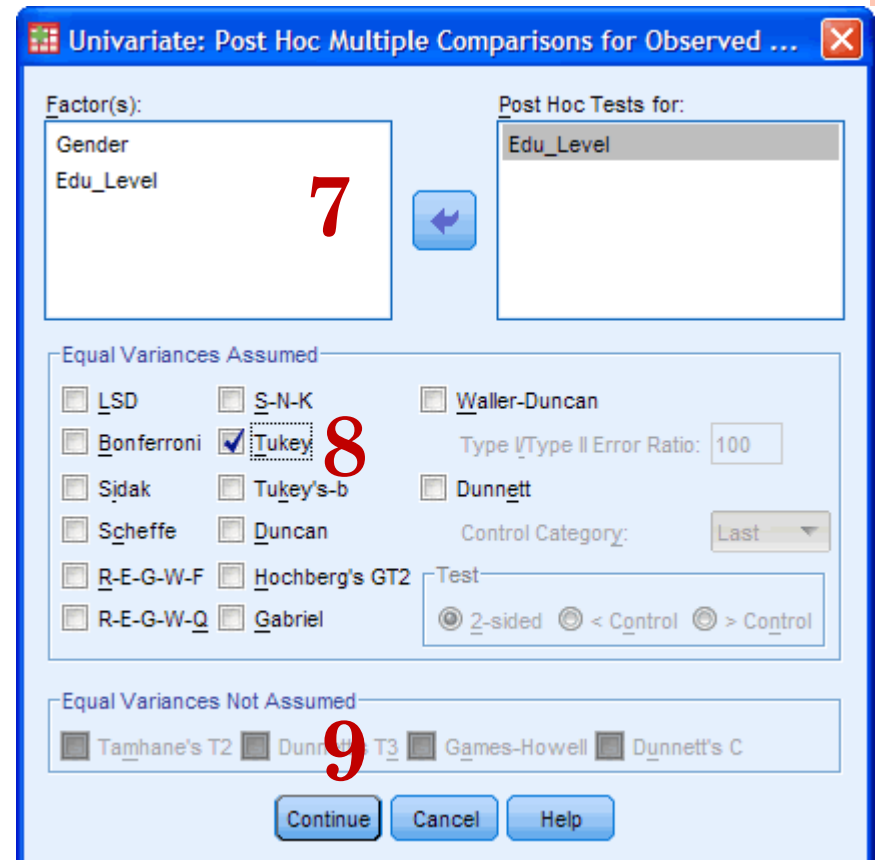
ملاحظة : ضع المتغير المستقل ذي العدد الأكبر
من المستويات في المحور الأفقي

سيظهر المتغير Edu_level , Gender
في صندوق plot



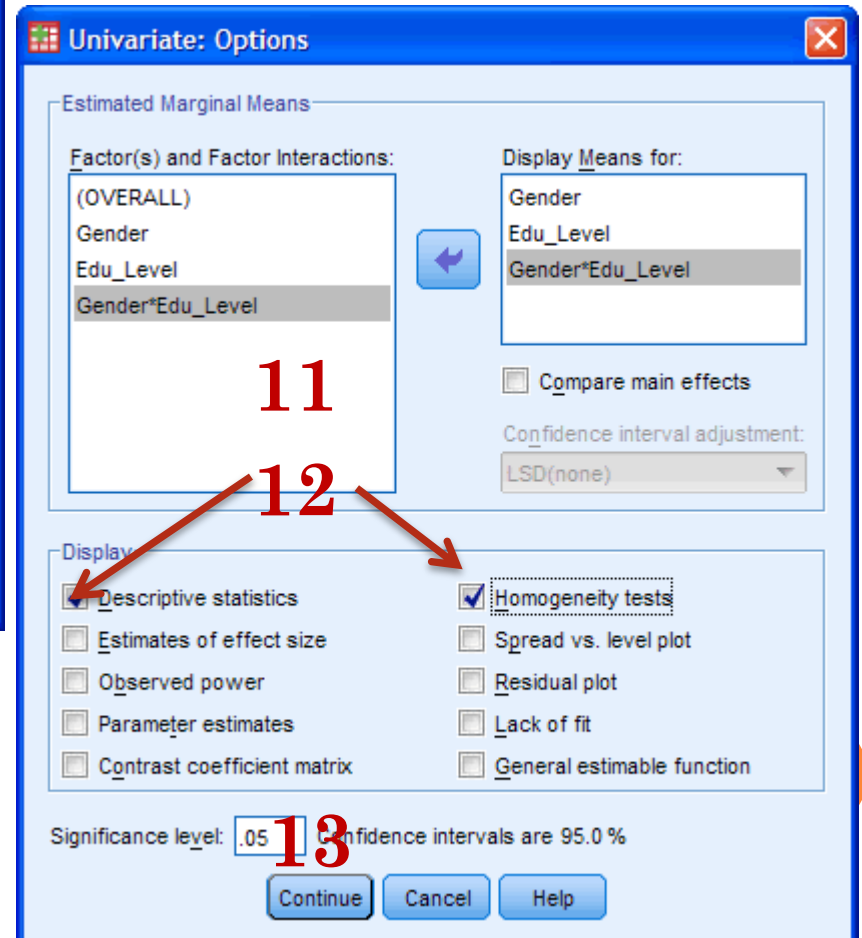
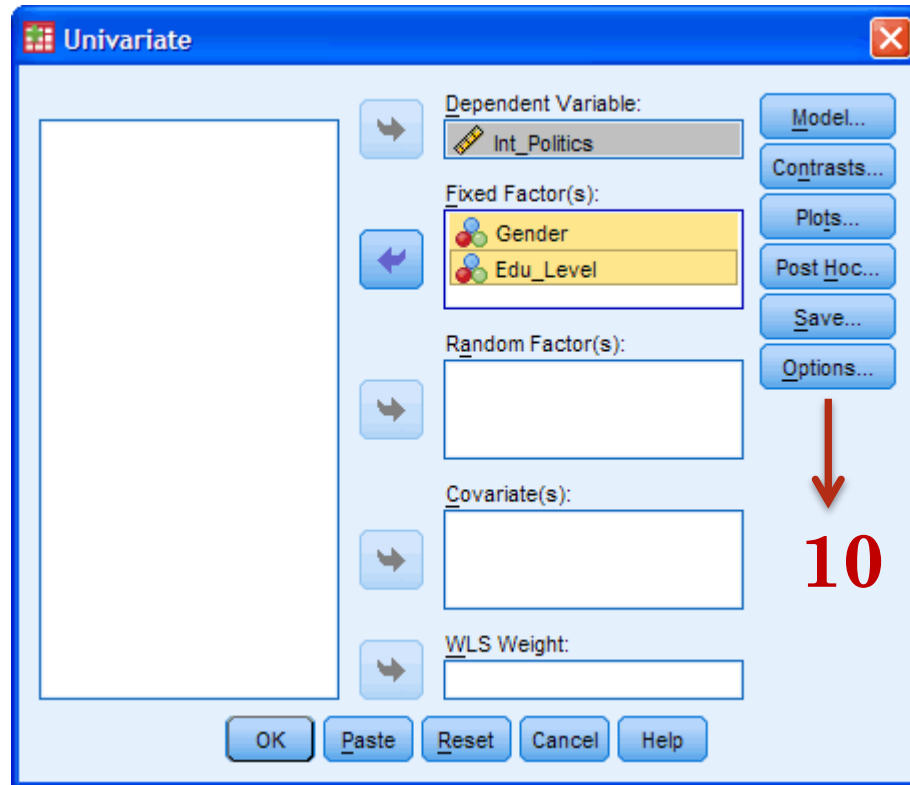
ملاحظة ١

هنا سننقل فقط المتغير الذي يحتوي
على أكثر من مستويين



ملاحظة ٢

إن الاختبارات التتابعية التي أخذناها في
تحليل التباين الأحادي تظل صالحة في
تحليل التباين الثنائي



ثم اضغط ok في نافذة
univariate



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



اختبارات التفاعل بين المتغيرات هذا الجدول يبين النتائج الفعلية للاتجاهين :ANOVA

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)

لا يوجد فرق كبير (معنوي) في الاهتمام بالسياسة بين الجنسين $P=0.207$ ، ولكن كانت هناك فروق ذات دلالة إحصائية بين المستويات التعليمية ($P=0.000$). كما أنه لا يوجد تفاعل مشترك ذو دلالة إحصائية عند 0.05 حيث $P=0.014$



جدول المقارنات المتعدد

هذا الجدول يبين نتائج اختبار توكي للمستويات المختلفة للمتغير Edu_level كونه ذو تأثير معنوي

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

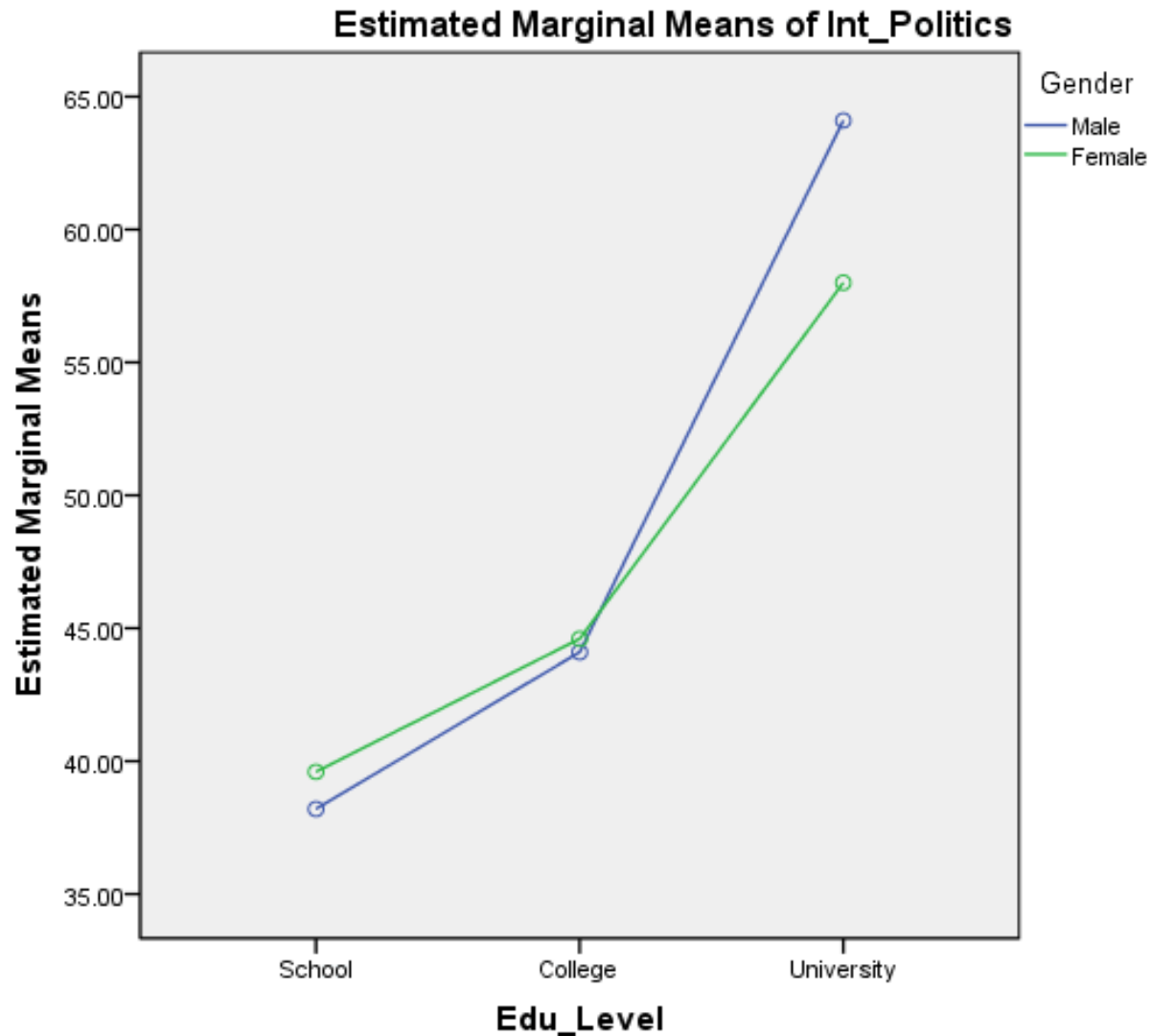
Based on observed means.

The error term is Mean Square(Error) = 18.059.

*. The mean difference is significant at the .05 level.

ونحن مهتمون في الاختلافات بين (١) مدرسة وكلية، (٢) مدرسة وجامعة، و (٣) كلية وجامعة. من نتائج، يمكننا أن نرى أن هناك فرقا معنويا بين الثلاثة مستويات المختلفة من المستوى التعليمي حيث

) $P < 0.005$



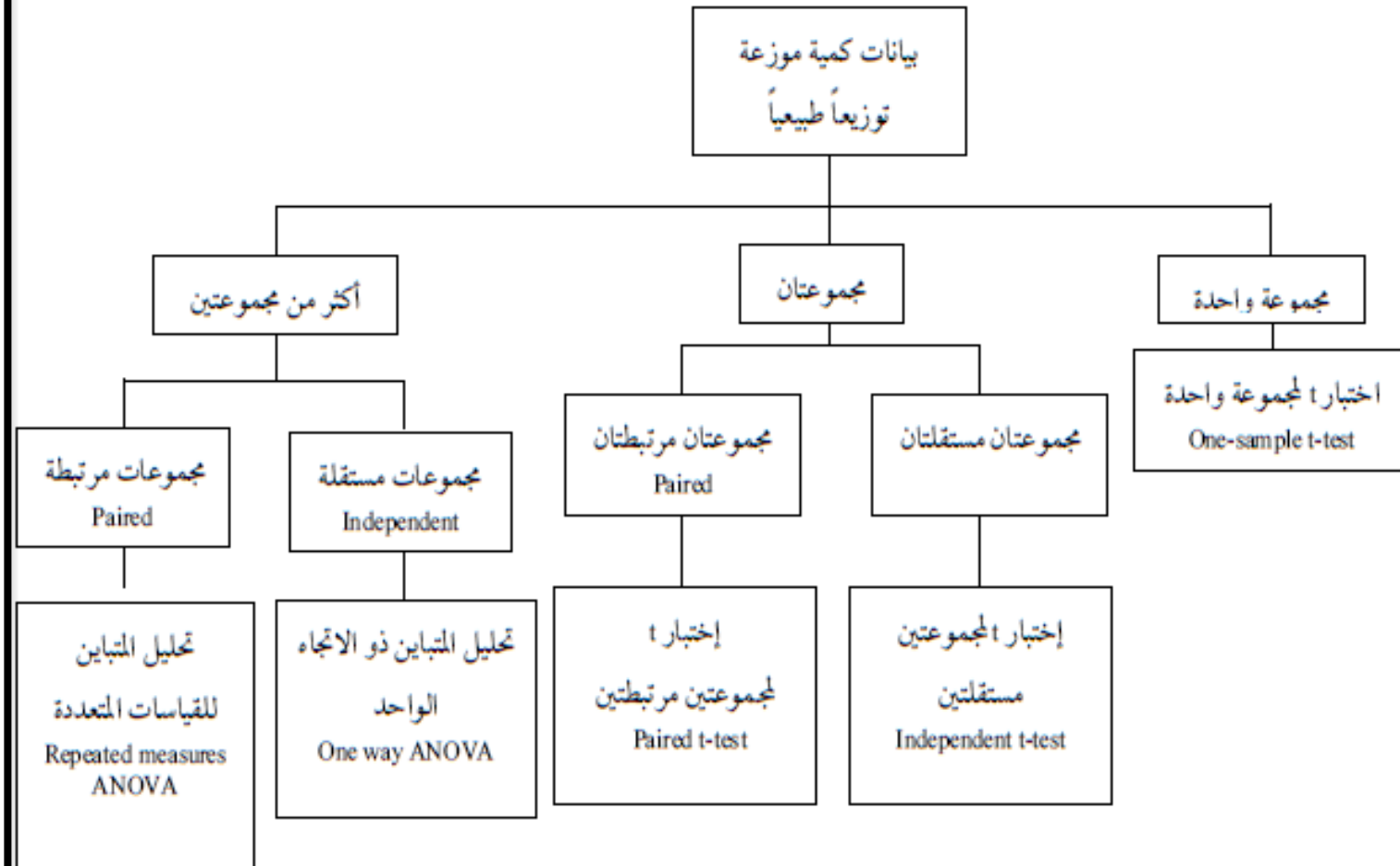
إن وجود تقاطع يعني وجود تفاعل ، بينما الخطوط المتوازية تعني أن لا تفاعل مشترك

تمرين: تأثير مستوى الدخل و فترة الدوام (الدراسة) على نسبة الطالب المؤوية

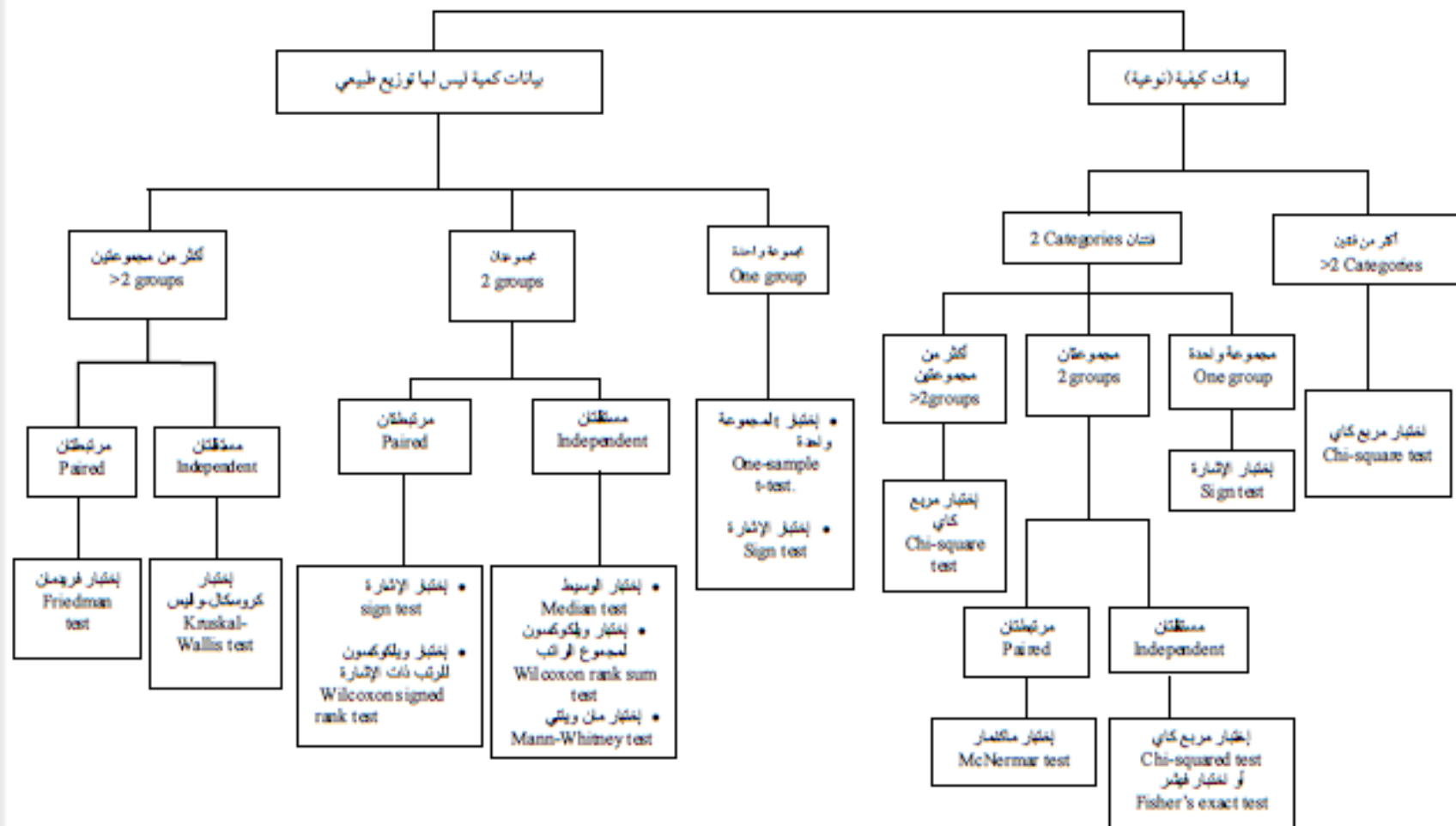
اكتبى فرضيات التحليل و حددي الاسلوب الاحصائي الملائم

		مستويات الدخل			
		L1	L2	L3	L4
		أوقات الدوام اليومي	AM	60	85
75	80			85	85
80	79			67	86
67	66			90	77
PM	90		78	88	85
	90		59	80	66
	88		67	67	67
	77		85	77	84

الشكل 8.2 اختيار الاختبار الإحصائي المناسب للبيانات الكمية ذات التوزيع الطبيعي :



الشكل 8.3 اختيار الاختبار الإحصائي المناسب للبيانات الكيفية (والكمية بدون توزيع طبيعي):



عدد العينات	الفرض	التصميم التجريبي	نوع البيانات	الاختبار الإحصائي
عينة واحدة	التحقق من جودة المطابقة	مجموعة واحدة ذات الاختبار الواحد	اسمية	ذى الحدين - كا ^٢ - سمير نوف
			رتبية	سمير نوف - الإشارة
			فترية	اختبار Z - اختبارات
عينتان مستقلتان	الفروق بين المجموعات	مجموعتان تجريبية - ضابطة	اسمية	كا ^٢ - فشر - سمير نوف
			رتبية	الوسيط - مان ويتنى - التابع
			فترية	اختبارات
عينتان مترابطتان	الفروق بين القياسات	مجموعة واحدة ذات اختبارين قبلي وبعدي	اسمية	ماكنمار
			رتبية	ولكوكسن - الإشارة
			فترية	اختبارات
عدة عينات مستقلة	الفروق بين المجموعات	المجموعات المتعددة	اسمية	كا ^٢
			رتبية	الوسيط - كروسكال وللاس
			فترية	تحليل التباين - تحليل التباين
عدة عينات مترابطة	الفروق بين القياسات	مجموعة واحدة ذات الاختبارات المتعددة	اسمية	كوجران
			رتبية	فريدمان
			فترية	تحليل التباين ذو القياسات المتكررة

يتبع ----->

معامل ارتباط فاي - معامل التوافق - معامل الاقتران الرباعي	اسمية	مجموعة واحدة ذات اختبار قبلي أو بعدى أو عدة اختبارات	الارتباط بين القياسات أو العلاقة بين المتغيرات "دراسات ارتباطية"	عينة واحدة أو عينتان أو عدة عينات
معامل ارتباط سبيرمان - معامل ارتباط كندال	رتبية			
معامل ارتباط بيرسون - الارتباط القانوني - الارتباط المتعدد	فترية			
تحليل الانحدار بأنواعه المختلفة - السلاسل الزمنية	فترية	مجموعة واحدة أو عدة مجموعات مع عدة اختبارات	"دراسات تنبؤية" للمتغيرات أو عضوية الجماعة	عينة واحدة أو عينتان أو عدة عينات
التحليل التمييزي بأنواعه المختلفة				
التحليل العاُملي الاستكشافي - التحليل العاُملي التوكيدي	فترية	مجموعة واحدة أو عدة مجموعة مع عدة اختبارات	"دراسات عاملية" البناء العاُملي	عينة واحدة أو عينتان أو عدة عينات



Saba Mohammed Alwan
Statistical and O.R
Faculty of Science. KSU
salwan@ksu.edu.sa

المراجع

- ١) ابو سريع، رضا. (٢٠٠٤). تحليل البيانات باستخدام برنامج spss ، دار الفكر، عمان.
- ٢) البشير، سعد. (٢٠٠٣)، دليلك الى البرنامج الاحصائي spss ،المعهد العربي للتدريب والبحوث الاحصائية، العراق
- ٣) الاختبارات الاحصائية البارامترية واللابارامترية باستخدام SPSS الاصدار السابع عشر ، د.سوسن ابراهيم أبو العلا شلبي، جامعة الملك سعود
- ٤) الارتباط والانحدار د. كامل أبو ضاهر ، الجامعة الاسلامية - غزة
- ٥) -بعض المعادلات الإحصائية المستخدمة في تحديد عينة البحث، الهزاع، هزاع محمد. فسيولوجيا الجهد البدني: الأسس النظرية والإجراءات العملية للقياسات الفسيولوجية. الفصل الرابع. كتاب مقدم للنشر.

Multiple Regression, Abdelfatah Mustafa, Mansuora University(٦

http://www.nca.umich.edu/sample_size_chart

<http://www.surveysystem.com/sscalc.htm>