

SAS, SPSS, R, Minitab and Excel Data Analysis

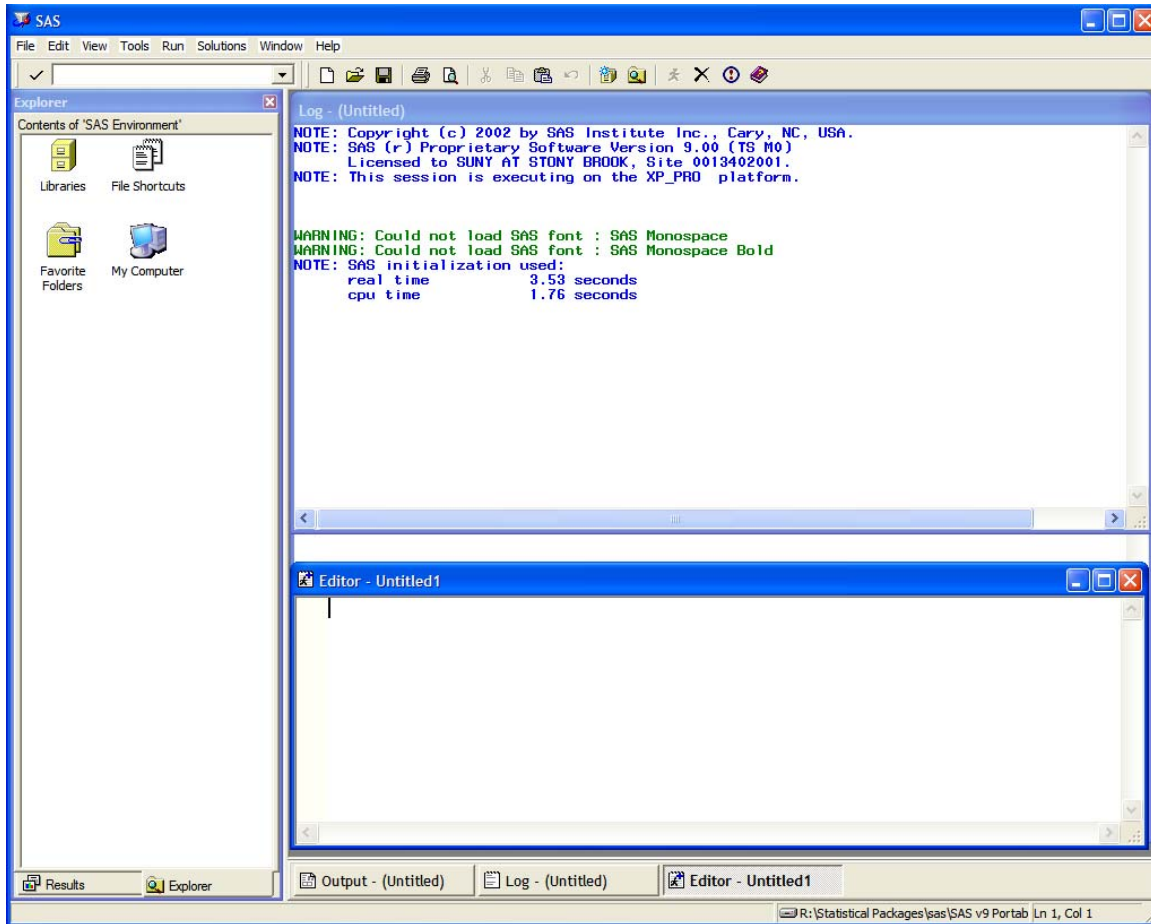
مقدمة:

في هذه المادة سوف أستخدم القواعد اللغوية (او لغة الأوامر) Syntax في إدخال الأوامر وذلك للأسباب التالية:

- 1- القواعد اللغوية او لغة الأوامر أكثر مرونة من خيارات نوافذ الإسقاط.
 - 2- لغة الأوامر لها ميزات أكثر (خيارات أكثر).
 - 3- لغة الأوامر تحفظ للإستخدام لاحقا.
 - 4- لغة الأوامر يمكن تكرار العمل بها.
- نوافذ الإسقاط قد تكون مربكة وغير واضحة المقاصد للمبتدئ وقد تتداخل النوافذ كل نافذة تؤدي لنافذة داخلية اخرى بحيث يضيع المستخدم المبتدئ و قد يختار خيارات غير مناسبة تؤدي لنتائج غير صحيحة.
- إستخدام نوافذ الإسقاط يعطي خيارات إفتراضية قد لاتفي بغرض المحلل كما أن أي تحليل بواسطتها لايمكن تكراره أليا بل يجب إستخدام النوافذ من جديد. كما انها تعطي مخرجات مهما كانت المدخلات بدون التحقق من صحة البيانات او التحليل المجرى عليها . في الحقيقة نوافذ الإسقاط هي واجه يختفي خلفها محرر قواعد اللغة او إختصارا محرر الصيغ والذي يقوم فعلا بالتفاعل مع البرنامج.

مقدمة للحزمة SAS :

عند تفعيل SAS تظهر في النافذة الرئيسية ثلاثة نوافذ:



- 1- محرر قواعد اللغة Syntax Editor او محرر الأوامر أو محرر البرامج Program Editor. والذي يستخدم لإدخال أوامر البرنامج.
- 2- نافذة المتابعة Log. والذي يظهر الأخطاء والتحذيرات أثناء إدخال وتشغيل البرنامج.

3- نافذة الإخراج Output. والتي يظهر فيها نتائج إجراء البرنامج.
 طبعا واحدة من هذه النوافذ يكون مفعلا فقط والتي كان قد تم إختيارها. (لتفعيل أي نافذة
 نضغط على الفارة عندما يكون المؤشر داخل تلك النافذة).

إدخال البيانات:

سوف نستخدم البيانات التالية عن سيارات والمتغيرات هي make نوع السيارة و
 price السعر و mpg عدد الأميال للجالون و rep78 معدل الإصلاح لسنة 78 و
 weight وزن السيارة و length طول السيارة و foreign محلية 0 أو أجنبية 1:

make	price	mpg	rep78	weight	length	foreign
AMC	4099	22	3	2930	186	0
AMC	4749	17	3	3350	173	0
AMC	3799	22	3	2640	168	0
Audi	9690	17	5	2830	189	1
Audi	6295	23	3	2070	174	1
BMW	9735	25	4	2650	177	1
Buick	4816	20	3	3250	196	0
Buick	7827	15	4	4080	222	0
Buick	5788	18	3	3670	218	0
Buick	4453	26	3	2230	170	0
Buick	5189	20	3	3280	200	0
Buick	10372	16	3	3880	207	0
Buick	4082	19	3	3400	200	0
Cad.	11385	14	3	4330	221	0
Cad.	14500	14	2	3900	204	0
Cad.	15906	21	3	4290	204	0
Chev.	3299	29	3	2110	163	0
Chev.	5705	16	4	3690	212	0
Chev.	4504	22	3	3180	193	0
Chev.	5104	22	2	3220	200	0
Chev.	3667	24	2	2750	179	0
Chev.	3955	19	3	3430	197	0
Datsun	6229	23	4	2370	170	1
Datsun	4589	35	5	2020	165	1
Datsun	5079	24	4	2280	170	1
Datsun	8129	21	4	2750	184	1

البرنامج التالي يقرأ البيانات ويخزنها في ملف داخلي يسمى auto ثم يخرجها بـ PROC

PRINT

```
DATA auto ;
  INPUT make $ price mpg rep78 weight length foreign ;
```

```

DATALINES;
AMC      4099 22 3      2930 186 0
AMC      4749 17 3      3350 173 0
AMC      3799 22 3      2640 168 0
Audi     9690 17 5      2830 189 1
Audi     6295 23 3      2070 174 1
BMW      9735 25 4      2650 177 1
Buick    4816 20 3      3250 196 0
Buick    7827 15 4      4080 222 0
Buick    5788 18 3      3670 218 0
Buick    4453 26 3      2230 170 0
Buick    5189 20 3      3280 200 0
Buick    10372 16 3      3880 207 0
Buick    4082 19 3      3400 200 0
Cad.     11385 14 3      4330 221 0
Cad.     14500 14 2      3900 204 0
Cad.     15906 21 3      4290 204 0
Chev.    3299 29 3      2110 163 0
Chev.    5705 16 4      3690 212 0
Chev.    4504 22 3      3180 193 0
Chev.    5104 22 2      3220 200 0
Chev.    3667 24 2      2750 179 0
Chev.    3955 19 3      3430 197 0
Datsun   6229 23 4      2370 170 1
Datsun   4589 35 5      2020 165 1
Datsun   5079 24 4      2280 170 1
Datsun   8129 21 4      2750 184 1
;
RUN;

PROC PRINT DATA=auto(obs=10);
RUN;

```

لاحظ obs =10 والذي يخرج 10 مشاهدات والتالي هو المخرجات من PROC

PRINT

OBS	MAKE	PRICE	MPG	REP78	WEIGHT	LENGTH	FOREIGN
1	AMC	4099	22	3	2930	186	0
2	AMC	4749	17	3	3350	173	0
3	AMC	3799	22	3	2640	168	0
4	Audi	9690	17	5	2830	189	1
5	Audi	6295	23	3	2070	174	1
6	BMW	9735	25	4	2650	177	1
7	Buick	4816	20	3	3250	196	0
8	Buick	7827	15	4	4080	222	0
9	Buick	5788	18	3	3670	218	0
10	Buick	4453	26	3	2230	170	0

إدخال بيانات من الإنترنت في SAS

```
filename eggs url 'http://lib.stat.cmu.edu:80/crab/eggs';
data eggdata;
  infile eggs;
  input year 1-2 numeggs 4-9;
run;
proc print data=eggdata;
run;
```

LOG

NOTE: AUTOEXEC processing completed.

```
1 filename eggs url 'http://lib.stat.cmu.edu:80/crab/eggs';
2 data eggdata;
3   infile eggs;
4   input year 1-2 numeggs 4-9;
5   run;
```

NOTE: The infile EGGS is:
Filename=http://lib.stat.cmu.edu:80/crab/eggs,
Local Host Name=alpha,
Local Host IP addr=192.168.2.4,
Service Hostname Name=FRET.stat.cmu.edu,
Service IP addr=128.2.234.14,Service Name=N/A,
Service Portno=80,Lrecl=256,Recfm=Variable

NOTE: 14 records were read from the infile EGGS.
The minimum record length was 9.
The maximum record length was 9.

NOTE: The data set WORK.EGGDATA has 14 observations and 2 variables.

NOTE: DATA statement used (Total process time):
real time 1.12 seconds
cpu time 0.15 seconds

```
6 proc print data=eggdata; run;
```

NOTE: There were 14 observations read from the data set WORK.EGGDATA.

NOTE: PROCEDURE PRINT used (Total process time):
real time 0.06 seconds

cpu time 0.03 seconds

OUTPUT

Thursday, March 4, 2013 1

The SAS System

18:46

Obs	year	numeggs			
1	73	143822			
2	74	151908			
3	75	178474			
4	76	156195			
5	77	124766			
6	78	101115			
7	79	135421			
8	80	148988			
9	81	149471			
10	82	170201			
11	83	161676			
12	84	170715			
13	85	180938			
			14	86	204797

الإحصاء الوصفي في SAS Descriptive statistics in SAS

لإيجاد الإحصاءات الوصفية نستخدم PROC MEANS كالتالي

```
PROC MEANS DATA=auto;
RUN;
```

فينتج

Variable	N	Mean	Std Dev	Minimum	Maximum
PRICE	26	6651.73	3371.12	3299.00	15906.00
MPG	26	20.9230769	4.7575042	14.0000000	35.0000000
REP78	26	3.2692308	0.7775702	2.0000000	5.0000000
WEIGHT	26	3099.23	695.0794089	2020.00	4330.00
LENGTH	26	190.0769231	18.1701361	163.0000000	222.0000000
FOREIGN	26	0.2692308	0.4523443	0	1.0000000

التالي يعطي إحصاءات وصفية مصنفة حسب المتغير foreign

```
PROC MEANS DATA=auto;
  CLASS foreign;
RUN;
```

المخرج

FOREIGN	N	Obs	Variable	N	Mean	Std Dev	Minimum
0	19		PRICE	19	6484.16	3768.46	3299.00
			MPG	19	19.7894737	4.0356598	14.0000000
			REP78	19	2.9473684	0.5242650	2.0000000
			WEIGHT	19	3347.89	627.1769106	2110.00
			LENGTH	19	195.4210526	17.9639014	163.0000000
1	7		PRICE	7	7106.57	2101.83	4589.00
			MPG	7	24.0000000	5.5075705	17.0000000
			REP78	7	4.1428571	0.6900656	3.0000000
			WEIGHT	7	2424.29	325.1593016	2020.00
			LENGTH	7	175.5714286	8.4628038	165.0000000

FOREIGN	N	Obs	Variable	Maximum
0	19		PRICE	15906.00
			MPG	29.0000000
			REP78	4.0000000
			WEIGHT	4330.00
			LENGTH	222.0000000
1	7		PRICE	9735.00
			MPG	35.0000000
			REP78	5.0000000
			WEIGHT	2830.00

LENGTH 189.0000000

نستطيع الحصول على إحصاء وصفي مفصل للسعر باستخدام PROC UNIVARIATE

```
PROC UNIVARIATE DATA=auto;
  VAR PRICE;
RUN;
```

و المخرجات

Univariate Procedure
Variable=PRICE

Moments			
N	26	Sum Wgts	26
Mean	6651.731	Sum	172945
Std Dev	3371.12	Variance	11364449
Skewness	1.470727	Kurtosis	1.534672
USS	1.4345E9	CSS	2.8411E8
CV	50.68034	Std Mean	661.131
T:Mean=0	10.06114	Pr> T	0.0001
Num ^= 0	26	Num > 0	26
M(Sign)	13	Pr>= M	0.0001
Sgn Rank	175.5	Pr>= S	0.0001

Quantiles(Def=5)

100% Max	15906	99%	15906
75% Q3	8129	95%	14500
50% Med	5146.5	90%	11385
25% Q1	4453	10%	3799
0% Min	3299	5%	3667
		1%	3299
Range	12607		
Q3-Q1	3676		
Mode	3299		

Extremes			
Lowest	Obs	Highest	Obs
3299(	17)	9735(	6)
3667(	21)	10372(	12)
3799(	3)	11385(	14)
3955(	22)	14500(	15)
4082(	13)	15906(	16)

نحصل على توزيع تكراري للمتغير rep78 باستخدام PROC FREQ

```
PROC FREQ DATA=auto;
  TABLES rep78 ;
RUN;
```

و الناتج

REP78	Frequency	Percent	Cumulative Frequency	Cumulative Percent
2	3	11.5	3	11.5
3	15	57.7	18	69.2
4	6	23.1	24	92.3
5	2	7.7	26	100.0

ويمكن إنشاء جدول ثنائي two way table لتكرارات rep78 مع المتغير foreign

```
PROC FREQ DATA=auto ;
  TABLES rep78 * foreign ;
RUN;
```

المخرجات

TABLE OF REP78 BY FOREIGN

REP78	FOREIGN		
	0	1	
Frequency			
Percent			
Row Pct			
Col Pct			
	0	1	Total
2	3	0	3
	11.54	0.00	11.54
	100.00	0.00	
	15.79	0.00	
3	14	1	15
	53.85	3.85	57.69
	93.33	6.67	
	73.68	14.29	
4	2	4	6
	7.69	15.38	23.08
	33.33	66.67	
	10.53	57.14	
5	0	2	2
	0.00	7.69	7.69
	0.00	100.00	
	0.00	28.57	
Total	19	7	26
	73.08	26.92	100.00

الرسومات البسيطة:

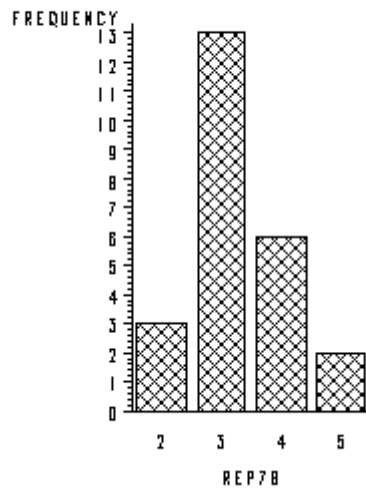
سوف نرسم تكرارات rep78 في رسم عامودي

```
TITLE 'Bar Chart with Discrete Option';
PROC GCHART DATA=auto;
    VBAR rep78/ DISCRETE;

RUN;
```

فينتج

Bar Chart with Discrete Option



الترباط و الإنحدار وتحليل التباين Correlation, regression and analysis of variance

سوف نستخدم PROC CORR لإيجاد الترابط بين price و mpg و weight و length

```
PROC CORR DATA=auto ;
    VAR price mpg weight length ;
RUN;
```

المخرجات

Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
PRICE	26	6652	3371	172945	3299	15906
MPG	26	20.92308	4.75750	544.00000	14.00000	35.00000
WEIGHT	26	3099	695.07941	80580	2020	4330
LENGTH	26	190.07692	18.17014	4942	163.00000	222.00000

Pearson Correlation Coefficients / Prob > |R| under Ho: Rho=0 / N = 26

	PRICE	MPG	WEIGHT	LENGTH
PRICE	1.00000 0.0	-0.43846 0.0251	0.55607 0.0032	0.43604 0.0260
MPG	-0.43846 0.0251	1.00000 0.0	-0.80816 0.0001	-0.76805 0.0001
WEIGHT	0.55607 0.0032	-0.80816 0.0001	1.00000 0.0	0.90654 0.0001
LENGTH	0.43604 0.0260	-0.76805 0.0001	0.90654 0.0001	1.00000 0.0

تستخدم PROC REG لإيجاد إنحدار متعدد للمتغير التابع mpg على المتغيرات weight و length و foreign كالتالي

```
PROC REG DATA=auto;
  MODEL mpg = weight length foreign ;
RUN;
```

المخرجات

Model: MODEL1
Dependent Variable: MPG

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	3	378.69701	126.23234	14.839	0.0001
Error	22	187.14915	8.50678		
C Total	25	565.84615			
Root MSE	2.91664	R-square	0.6693		
Dep Mean	20.92308	Adj R-sq	0.6242		
C.V.	13.93982				

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > T
INTERCEP	1	44.968582	9.32267757	4.824	0.0001
WEIGHT	1	-0.005008	0.00218752	-2.289	0.0320
LENGTH	1	-0.043056	0.07692650	-0.560	0.5813
FOREIGN	1	-1.269211	1.63213395	-0.778	0.4451

تستخدم PROC GLM لتحليل التباين لكي نختبر فيما إذا كان متوسط mpg هو نفسه للسيارات المحلية و الأجنبية

```
PROC GLM DATA=auto;
  CLASS foreign ;
  MODEL mpg = foreign ;
RUN;
```

المخرج

General Linear Models Procedure Class Level Information

Class	Levels	Values
-------	--------	--------

FOREIGN	2	0 1
---------	---	-----

Number of observations in data set = 26

General Linear Models Procedure

Dependent Variable: MPG

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	90.68825911	90.68825911	4.58	0.0427
Error	24	475.15789474	19.79824561		
Corrected Total	25	565.84615385			

R-Square	0.160270	C.V.	21.26610	Root MSE	4.4495220	MPG Mean	20.923077
----------	----------	------	----------	----------	-----------	----------	-----------

Source	DF	Type I SS	Mean Square	F Value	Pr > F
FOREIGN	1	90.68825911	90.68825911	4.58	0.0427

Source	DF	Type III SS	Mean Square	F Value	Pr > F
FOREIGN	1	90.68825911	90.68825911	4.58	0.0427

أجري و أدرس الأمثلة التالية بعناية:

```
PROC FREQ DATA=auto;  
  TABLES make ;  
RUN;
```

```
PROC FREQ DATA=auto;  
  TABLES rep78 ;  
RUN;
```

```
PROC FREQ DATA=auto;  
  TABLES foreign ;  
RUN;
```

```
PROC FREQ DATA=auto;  
  TABLES make rep78 foreign ;  
RUN;
```

```
PROC FREQ DATA=auto;  
  TABLES rep78*foreign ;  
RUN;
```

```
PROC FREQ DATA=auto;  
  TABLES rep78*foreign / NOROW NOCOL NOFREQ ;  
RUN;
```

```
PROC FREQ DATA=auto;  
  TABLES rep78*foreign / NOFREQ NOROW NOCOL ;  
RUN;  
PROC MEANS DATA=auto;  
  VAR mpg;  
RUN;
```

```
PROC UNIVARIATE DATA=auto;  
  VAR mpg;  
RUN;
```

```
PROC UNIVARIATE DATA=auto;  
  CLASS foreign;  
  VAR mpg;  
RUN;
```

:Statistical Tests in SAS الإختبارات الإحصائية

سوف نستخدم البيانات التالية:

```
DATA auto ;
  LENGTH make $ 20 ;
  INPUT make $ 1-17 price mpg rep78 hdroom trunk weight
        length turn displ gratio foreign ;
CARDS;
AMC Concord      4099   . 3 2.5 11 2930 186 40 121 3.58 0
AMC Pacer        4749   . 3 3.0 11 3350 173 40 258 2.53 0
AMC Spirit       3799   . . 3.0 12 2640 168 35 121 3.08 0
Audi 5000        9690 17 5 3.0 15 2830 189 37 131 3.20 1
Audi Fox         6295 23 3 2.5 11 2070 174 36  97 3.70 1
BMW 320i         9735 25 4 2.5 12 2650 177 34 121 3.64 1
Buick Century    4816 20 3 4.5 16 3250 196 40 196 2.93 0
Buick Electra    7827 15 4 4.0 20 4080 222 43 350 2.41 0
Buick LeSabre    5788 18 3 4.0 21 3670 218 43 231 2.73 0
Buick Opel       4453 26 . 3.0 10 2230 170 34 304 2.87 0
Buick Regal      5189 20 3 2.0 16 3280 200 42 196 2.93 0
Buick Riviera   10372 16 3 3.5 17 3880 207 43 231 2.93 0
Buick Skylark    4082 19 3 3.5 13 3400 200 42 231 3.08 0
Cad. Deville    11385 14 3 4.0 20 4330 221 44 425 2.28 0
Cad. Eldorado   14500 14 2 3.5 16 3900 204 43 350 2.19 0
Cad. Seville    15906 21 3 3.0 13 4290 204 45 350 2.24 0
Chev. Chevette   3299 29 3 2.5  9 2110 163 34 231 2.93 0
Chev. Impala     5705 16 4 4.0 20 3690 212 43 250 2.56 0
Chev. Malibu     4504 22 3 3.5 17 3180 193 31 200 2.73 0
Chev. Monte Carlo 5104 22 2 2.0 16 3220 200 41 200 2.73 0
Chev. Monza      3667 24 2 2.0  7 2750 179 40 151 2.73 0
Chev. Nova       3955 19 3 3.5 13 3430 197 43 250 2.56 0
Datsun 200       6229 23 4 1.5  6 2370 170 35 119 3.89 1
Datsun 210       4589 35 5 2.0  8 2020 165 32  85 3.70 1
Datsun 510       5079 24 4 2.5  8 2280 170 34 119 3.54 1
Datsun 810       8129 21 4 2.5  8 2750 184 38 146 3.55 1
Dodge Colt       3984 30 5 2.0  8 2120 163 35  98 3.54 0
Dodge Diplomat   4010 18 2 4.0 17 3600 206 46 318 2.47 0
Dodge Magnum     5886 16 2 4.0 17 3600 206 46 318 2.47 0
Dodge St. Regis  6342 17 2 4.5 21 3740 220 46 225 2.94 0
Fiat Strada      4296 21 3 2.5 16 2130 161 36 105 3.37 1
Ford Fiesta      4389 28 4 1.5  9 1800 147 33  98 3.15 0
Ford Mustang     4187 21 3 2.0 10 2650 179 43 140 3.08 0
Honda Accord     5799 25 5 3.0 10 2240 172 36 107 3.05 1
Honda Civic      4499 28 4 2.5  5 1760 149 34  91 3.30 1
Linc. Continental 11497 12 3 3.5 22 4840 233 51 400 2.47 0
Linc. Mark V     13594 12 3 2.5 18 4720 230 48 400 2.47 0
Linc. Versailles 13466 14 3 3.5 15 3830 201 41 302 2.47 0
Mazda GLC        3995 30 4 3.5 11 1980 154 33  86 3.73 1
Merc. Bobcat     3829 22 4 3.0  9 2580 169 39 140 2.73 0
Merc. Cougar     5379 14 4 3.5 16 4060 221 48 302 2.75 0
Merc. Marquis    6165 15 3 3.5 23 3720 212 44 302 2.26 0
Merc. Monarch    4516 18 3 3.0 15 3370 198 41 250 2.43 0
Merc. XR-7       6303 14 4 3.0 16 4130 217 45 302 2.75 0
Merc. Zephyr     3291 20 3 3.5 17 2830 195 43 140 3.08 0
Olds 98          8814 21 4 4.0 20 4060 220 43 350 2.41 0
Olds Cutl Supr   5172 19 3 2.0 16 3310 198 42 231 2.93 0
```

```

Olds Cutlass      4733 19 3 4.5 16 3300 198 42 231 2.93 0
Olds Delta 88    4890 18 4 4.0 20 3690 218 42 231 2.73 0
Olds Omega       4181 19 3 4.5 14 3370 200 43 231 3.08 0
Olds Starfire    4195 24 1 2.0 10 2730 180 40 151 2.73 0
Olds Toronado    10371 16 3 3.5 17 4030 206 43 350 2.41 0
Peugeot 604      12990 14 . 3.5 14 3420 192 38 163 3.58 1
Plym. Arrow      4647 28 3 2.0 11 3260 170 37 156 3.05 0
Plym. Champ      4425 34 5 2.5 11 1800 157 37 86 2.97 0
Plym. Horizon    4482 25 3 4.0 17 2200 165 36 105 3.37 0
Plym. Sapporo    6486 26 . 1.5 8 2520 182 38 119 3.54 0
Plym. Volare     4060 18 2 5.0 16 3330 201 44 225 3.23 0
Pont. Catalina   5798 18 4 4.0 20 3700 214 42 231 2.73 0
Pont. Firebird   4934 18 1 1.5 7 3470 198 42 231 3.08 0
Pont. Grand Prix 5222 19 3 2.0 16 3210 201 45 231 2.93 0
Pont. Le Mans    4723 19 3 3.5 17 3200 199 40 231 2.93 0
Pont. Phoenix    4424 19 . 3.5 13 3420 203 43 231 3.08 0
Pont. Sunbird    4172 24 2 2.0 7 2690 179 41 151 2.73 0
Renault Le Car   3895 26 3 3.0 10 1830 142 34 79 3.72 1
Subaru           3798 35 5 2.5 11 2050 164 36 97 3.81 1
Toyota Celica    5899 18 5 2.5 14 2410 174 36 134 3.06 1
Toyota Corolla   3748 31 5 3.0 9 2200 165 35 97 3.21 1
Toyota Corona    5719 18 5 2.0 11 2670 175 36 134 3.05 1
Volvo 260        11995 17 5 2.5 14 3170 193 37 163 2.98 1
VW Dasher        7140 23 4 2.5 12 2160 172 36 97 3.74 1
VW Diesel        5397 41 5 3.0 15 2040 155 35 90 3.78 1
VW Rabbit        4697 25 4 3.0 15 1930 155 35 89 3.78 1
VW Scirocco      6850 25 4 2.0 16 1990 156 36 97 3.78 1
;
RUN;

```

إختبارات T : T-tests

نستخدم إختبار t لإختبار مساواة متوسط mpg للسيارات المحلية و الأجنبية

```

PROC TTEST DATA=auto;
  CLASS foreign;
  VAR mpg;
RUN;

```

المخرج:

TTEST PROCEDURE

Variable: MPG

FOREIGN	N	Mean	Std Dev	Std Error	Minimum	Maximum
0	49	19.79591837	4.85188791	0.69312684	12.00000000	34.00000000
1	22	24.77272727	6.61118690	1.40950978	14.00000000	41.00000000

Variances	T	DF	Prob> T
Unequal	-3.1685	31.6	0.0034
Equal	-3.5597	69.0	0.0007

For H0: Variances are equal, F' = 1.86 DF = (21,48) Prob>F' = 0.0776

إختبارات مربع كاي Chi-square tests:

المتغير rep78 يأخذ القيمة 1 كأ سوء قيمة و 5 كأفضل قيمة

```
PROC FREQ DATA=auto;
  TABLES rep78*foreign / CHISQ ;
RUN;
```

المخرجات

TABLE OF REP78 BY FOREIGN

REP78		FOREIGN		
Frequency				
Percent				
Row Pct				
Col Pct	0	1	Total	
1	2	0	2	
	2.90	0.00	2.90	
	100.00	0.00		
	4.17	0.00		
2	8	0	8	
	11.59	0.00	11.59	
	100.00	0.00		
	16.67	0.00		
3	27	3	30	
	39.13	4.35	43.48	
	90.00	10.00		
	56.25	14.29		
4	9	9	18	
	13.04	13.04	26.09	
	50.00	50.00		
	18.75	42.86		
5	2	9	11	
	2.90	13.04	15.94	
	18.18	81.82		
	4.17	42.86		
Total	48	21	69	
	69.57	30.43	100.00	

Frequency Missing = 5

STATISTICS FOR TABLE OF REP78 BY FOREIGN			
Statistic	DF	Value	Prob
Chi-Square	4	27.264	0.001
Likelihood Ratio Chi-Square	4	29.912	0.001
Mantel-Haenszel Chi-Square	1	23.851	0.001
Phi Coefficient		0.629	
Contingency Coefficient		0.532	
Cramer's V		0.629	

Effective Sample Size = 69

Frequency Missing = 5

WARNING: 40% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

```
PROC FREQ DATA=auto;
  TABLES rep78*foreign / CHISQ EXACT ;
RUN;
```

STATISTICS FOR TABLE OF REP78 BY FOREIGN			
Statistic	DF	Value	Prob
Chi-Square	4	27.264	0.001
Likelihood Ratio Chi-Square	4	29.912	0.001
Mantel-Haenszel Chi-Square	1	23.851	0.001
Fisher's Exact Test (2-Tail)			6.27E-06
Phi Coefficient		0.629	
Contingency Coefficient		0.532	
Cramer's V		0.629	

إستخدام where في SAS

Using where with SAS procedures

للبيانات

```
DATA auto ;
  INPUT make $ 1-17 price mpg rep78 hdroom trunk weight length turn
        displ gratio foreign ;
DATALINES;
AMC Concord      4099 22 3 2.5 11 2930 186 40 121 3.58 0
AMC Pacer        4749 17 3 3.0 11 3350 173 40 258 2.53 0
AMC Spirit       3799 22 . 3.0 12 2640 168 35 121 3.08 0
Audi 5000        9690 17 5 3.0 15 2830 189 37 131 3.20 1
```

Audi Fox	6295	23	3	2.5	11	2070	174	36	97	3.70	1
BMW 320i	9735	25	4	2.5	12	2650	177	34	121	3.64	1
Buick Century	4816	20	3	4.5	16	3250	196	40	196	2.93	0
Buick Electra	7827	15	4	4.0	20	4080	222	43	350	2.41	0
Buick LeSabre	5788	18	3	4.0	21	3670	218	43	231	2.73	0
Buick Opel	4453	26	.	3.0	10	2230	170	34	304	2.87	0
Buick Regal	5189	20	3	2.0	16	3280	200	42	196	2.93	0
Buick Riviera	10372	16	3	3.5	17	3880	207	43	231	2.93	0
Buick Skylark	4082	19	3	3.5	13	3400	200	42	231	3.08	0
Cad. Deville	11385	14	3	4.0	20	4330	221	44	425	2.28	0
Cad. Eldorado	14500	14	2	3.5	16	3900	204	43	350	2.19	0
Cad. Seville	15906	21	3	3.0	13	4290	204	45	350	2.24	0
Chev. Chevette	3299	29	3	2.5	9	2110	163	34	231	2.93	0
Chev. Impala	5705	16	4	4.0	20	3690	212	43	250	2.56	0
Chev. Malibu	4504	22	3	3.5	17	3180	193	31	200	2.73	0
Chev. Monte Carlo	5104	22	2	2.0	16	3220	200	41	200	2.73	0
Chev. Monza	3667	24	2	2.0	7	2750	179	40	151	2.73	0
Chev. Nova	3955	19	3	3.5	13	3430	197	43	250	2.56	0
Datsun 200	6229	23	4	1.5	6	2370	170	35	119	3.89	1
Datsun 210	4589	35	5	2.0	8	2020	165	32	85	3.70	1
Datsun 510	5079	24	4	2.5	8	2280	170	34	119	3.54	1
Datsun 810	8129	21	4	2.5	8	2750	184	38	146	3.55	1
Dodge Colt	3984	30	5	2.0	8	2120	163	35	98	3.54	0
Dodge Diplomat	4010	18	2	4.0	17	3600	206	46	318	2.47	0
Dodge Magnum	5886	16	2	4.0	17	3600	206	46	318	2.47	0
Dodge St. Regis	6342	17	2	4.5	21	3740	220	46	225	2.94	0
Fiat Strada	4296	21	3	2.5	16	2130	161	36	105	3.37	1
Ford Fiesta	4389	28	4	1.5	9	1800	147	33	98	3.15	0
Ford Mustang	4187	21	3	2.0	10	2650	179	43	140	3.08	0
Honda Accord	5799	25	5	3.0	10	2240	172	36	107	3.05	1
Honda Civic	4499	28	4	2.5	5	1760	149	34	91	3.30	1
Linc. Continental	11497	12	3	3.5	22	4840	233	51	400	2.47	0
Linc. Mark V	13594	12	3	2.5	18	4720	230	48	400	2.47	0
Linc. Versailles	13466	14	3	3.5	15	3830	201	41	302	2.47	0
Mazda GLC	3995	30	4	3.5	11	1980	154	33	86	3.73	1
Merc. Bobcat	3829	22	4	3.0	9	2580	169	39	140	2.73	0
Merc. Cougar	5379	14	4	3.5	16	4060	221	48	302	2.75	0
Merc. Marquis	6165	15	3	3.5	23	3720	212	44	302	2.26	0
Merc. Monarch	4516	18	3	3.0	15	3370	198	41	250	2.43	0
Merc. XR-7	6303	14	4	3.0	16	4130	217	45	302	2.75	0
Merc. Zephyr	3291	20	3	3.5	17	2830	195	43	140	3.08	0
Olds 98	8814	21	4	4.0	20	4060	220	43	350	2.41	0
Olds Cutl Supr	5172	19	3	2.0	16	3310	198	42	231	2.93	0
Olds Cutlass	4733	19	3	4.5	16	3300	198	42	231	2.93	0
Olds Delta 88	4890	18	4	4.0	20	3690	218	42	231	2.73	0
Olds Omega	4181	19	3	4.5	14	3370	200	43	231	3.08	0
Olds Starfire	4195	24	1	2.0	10	2730	180	40	151	2.73	0
Olds Toronado	10371	16	3	3.5	17	4030	206	43	350	2.41	0
Peugeot 604	12990	14	.	3.5	14	3420	192	38	163	3.58	1
Plym. Arrow	4647	28	3	2.0	11	3260	170	37	156	3.05	0
Plym. Champ	4425	34	5	2.5	11	1800	157	37	86	2.97	0
Plym. Horizon	4482	25	3	4.0	17	2200	165	36	105	3.37	0
Plym. Sapporo	6486	26	.	1.5	8	2520	182	38	119	3.54	0
Plym. Volare	4060	18	2	5.0	16	3330	201	44	225	3.23	0
Pont. Catalina	5798	18	4	4.0	20	3700	214	42	231	2.73	0
Pont. Firebird	4934	18	1	1.5	7	3470	198	42	231	3.08	0
Pont. Grand Prix	5222	19	3	2.0	16	3210	201	45	231	2.93	0

```

Pont. Le Mans      4723 19 3 3.5 17 3200 199 40 231 2.93 0
Pont. Phoenix      4424 19 . 3.5 13 3420 203 43 231 3.08 0
Pont. Sunbird      4172 24 2 2.0 7 2690 179 41 151 2.73 0
Renault Le Car     3895 26 3 3.0 10 1830 142 34 79 3.72 1
Subaru             3798 35 5 2.5 11 2050 164 36 97 3.81 1
Toyota Celica      5899 18 5 2.5 14 2410 174 36 134 3.06 1
Toyota Corolla     3748 31 5 3.0 9 2200 165 35 97 3.21 1
Toyota Corona      5719 18 5 2.0 11 2670 175 36 134 3.05 1
Volvo 260          11995 17 5 2.5 14 3170 193 37 163 2.98 1
VW Dasher          7140 23 4 2.5 12 2160 172 36 97 3.74 1
VW Diesel          5397 41 5 3.0 15 2040 155 35 90 3.78 1
VW Rabbit          4697 25 4 3.0 15 1930 155 35 89 3.78 1
VW Scirocco        6850 25 4 2.0 16 1990 156 36 97 3.78 1
;
RUN;

```

البرنامج التالي يطبع الحالات التي يكون فيها rep78 أكبر من أو يساوي 3

```

PROC PRINT DATA=auto;
  WHERE rep78 >= 3;
  VAR make rep78;
RUN;

```

و الناتج

Obs	make	rep78
1	AMC Concord	3
2	AMC Pacer	3
4	Audi 5000	5
5	Audi Fox	3
6	BMW 320i	4
7	Buick Century	3
8	Buick Electra	4
9	Buick LeSabre	3
11	Buick Regal	3
12	Buick Riviera	3
13	Buick Skylark	3
14	Cad. Deville	3
16	Cad. Seville	3
17	Chev. Chevette	3
18	Chev. Impala	4

Obs	make	rep78
19	Chev. Malibu	3
22	Chev. Nova	3
23	Datsun 200	4
24	Datsun 210	5
25	Datsun 510	4
26	Datsun 810	4
27	Dodge Colt	5
31	Fiat Strada	3
32	Ford Fiesta	4
33	Ford Mustang	3
34	Honda Accord	5
35	Honda Civic	4
36	Linc. Continental	3
37	Linc. Mark V	3
38	Linc. Versailles	3
39	Mazda GLC	4
40	Merc. Bobcat	4
41	Merc. Cougar	4
42	Merc. Marquis	3
43	Merc. Monarch	3
44	Merc. XR-7	4
45	Merc. Zephyr	3
46	Olds 98	4
47	Olds Cutl Supr	3
48	Olds Cutlass	3
49	Olds Delta 88	4
50	Olds Omega	3
52	Olds Toronado	3
54	Plym. Arrow	3

Obs	make	rep78
55	Plym. Champ	5
56	Plym. Horizon	3
59	Pont. Catalina	4
61	Pont. Grand Prix	3
62	Pont. Le Mans	3
65	Renault Le Car	3
66	Subaru	5
67	Toyota Celica	5
68	Toyota Corolla	5
69	Toyota Corona	5
70	Volvo 260	5
71	VW Dasher	4
72	VW Diesel	5
73	VW Rabbit	4
74	VW Scirocco	4

أدرس البرنامجين التاليين

```
PROC FREQ DATA=auto;
  TABLES rep78*foreign ;
RUN;
```

Table of rep78 by foreign

rep78		foreign		
Frequency				
Percent				
Row Pct				
Col Pct		0	1	Total
1	2	0	2	
	2.90	0.00	2.90	
	100.00	0.00		
	4.17	0.00		
2	8	0	8	
	11.59	0.00	11.59	
	100.00	0.00		
	16.67	0.00		
3	27	3	30	

		39.13	4.35	43.48
		90.00	10.00	
		56.25	14.29	
-----+-----+-----+-----+-----				
4	9	9	18	
	13.04	13.04	26.09	
	50.00	50.00		
	18.75	42.86		
-----+-----+-----+-----+-----				
5	2	9	11	
	2.90	13.04	15.94	
	18.18	81.82		
	4.17	42.86		
-----+-----+-----+-----+-----				
Total	48	21	69	
	69.57	30.43	100.00	

```
PROC FREQ DATA=auto;
  WHERE rep78 >= 3;
  TABLES rep78*foreign ;
RUN;
```

Table of rep78 by foreign

rep78 foreign

Frequency				
Percent				
Row Pct				
Col Pct				
		0	1	Total
-----+-----+-----+-----+-----				
3	27	3	30	
	45.76	5.08	50.85	
	90.00	10.00		
	71.05	14.29		
-----+-----+-----+-----+-----				
4	9	9	18	
	15.25	15.25	30.51	
	50.00	50.00		
	23.68	42.86		
-----+-----+-----+-----+-----				
5	2	9	11	
	3.39	15.25	18.64	
	18.18	81.82		
	5.26	42.86		
-----+-----+-----+-----+-----				
Total	38	21	59	
	64.41	35.59	100.00	

```
PROC PRINT DATA=auto;
  WHERE rep78 <= 2;
  VAR make price rep78 ;
RUN;
```

Obs	make	price	rep78
3	AMC Spirit	3799	.
10	Buick Opel	4453	.
15	Cad. Eldorado	14500	2
20	Chev. Monte Carlo	5104	2
21	Chev. Monza	3667	2
28	Dodge Diplomat	4010	2
29	Dodge Magnum	5886	2
30	Dodge St. Regis	6342	2
51	Olds Starfire	4195	1
53	Peugeot 604	12990	.
57	Plym. Sapporo	6486	.
58	Plym. Volare	4060	2
60	Pont. Firebird	4934	1
63	Pont. Phoenix	4424	.
64	Pont. Sunbird	4172	2

القيم المفقودة and the where statement

لكي نستقصي القيم المفقودة من التحليل نستخدم

```
PROC PRINT DATA=auto;
  WHERE rep78 <= 2 and rep78 ^= . ;
  VAR make price rep78 ;
RUN;
```

لاحظ أن ^= تعني لا يساوي

Obs	make	price	rep78
15	Cad. Eldorado	14500	2
20	Chev. Monte Carlo	5104	2

Obs	make	price	rep78
21	Chev. Monza	3667	2
28	Dodge Diplomat	4010	2
29	Dodge Magnum	5886	2
30	Dodge St. Regis	6342	2
51	Olds Starfire	4195	1
58	Plym. Volare	4060	2
60	Pont. Firebird	4934	1
64	Pont. Sunbird	4172	2

أو البرنامج المكافئ

```
PROC PRINT DATA=auto;
  WHERE . < rep78 <= 2;
  VAR make price rep78 ;
RUN;
```

أدرس عمل البرامج التالية

```
PROC MEANS DATA=auto;
  WHERE rep78 = 1 OR rep78 = 2 ;
  VAR price ;
RUN;
```

```
PROC MEANS DATA=auto;
  WHERE rep78 = 3 or rep78 = 4 or rep78 = 5 ;
  VAR price ;
RUN;
```

```
PROC MEANS DATA=auto;
  WHERE 3 <= rep78 <= 5 ;
  VAR price ;
RUN;
```

```
PROC MEANS DATA=auto;
  WHERE rep78 in (3,4,5) ;
  VAR price ;
RUN;
```

Further SAS Examples

```
* SAS Program for Basic Statistical Tests;  
* Filename: Statistics.sas ;
```

```
LIBNAME myLib 'K:\myRfolder';
```

```
DATA temp;  
  SET myLib.mydata100;  
  * pretend q2 and q1 are the same score  
  * measured at two times & subtract;  
  myDiff=q2-q1; run;
```

```
* Basic stats in compact form;  
PROC MEANS; VAR q1--posttest; RUN;
```

```
* Basic stats of every sort;  
PROC UNIVARIATE; VAR q1--posttest; RUN;
```

```
* Frequencies & percents;  
PROC FREQ; TABLES workshop--q4; RUN;
```

```
* Chi-square;  
PROC FREQ;  
  TABLES workshop*gender/CHISQ; RUN;
```

```
* ---Measures of association---;
```

```
* Pearson correlations;  
PROC CORR; VAR q1-q4; RUN;
```

```
* Spearman correlations;  
PROC CORR SPEARMAN; VAR q1-q4; RUN;
```

```
* Linear regression;  
PROC REG;  
  MODEL q4=q1-q3;  
  RUN;
```

```
* ---Group comparisons---;
```

```
* Independent samples t-test;  
PROC TTEST;  
  CLASS gender;  
  VAR q1; RUN;
```

```
* Nonparametric version of above  
  * using Wilcoxon/Mann-Whitney test;  
PROC NPAR1WAY;  
  CLASS gender;  
  VAR q1; RUN;
```

```
* Paired samples t-test;  
PROC TTEST;
```

```

    PAIRED pretest*posttest; RUN;

* Nonparametric version of above using
  both Signed Rank test and Sign test;
PROC UNIVARIATE;
  VAR myDiff;
  RUN;

*Oneway Analysis of Variance (ANOVA);
PROC GLM;
  CLASS workshop;
  MODEL posttest=workshop;
  MEANS workshop / TUKEY; RUN;

*Nonparametric version of above using
  Kruskal-Wallis test;
PROC npar1way;
  CLASS workshop;

  VAR posttest; RUN;

```

The SAS System

11:14

Thursday, March 17, 2013 1

The MEANS Procedure

Variable	Label	N	Mean
Std Dev	Minimum		
<pre> ff ffffffffffffffffffffffffff </pre>			
q1	The instructor was well prepared.	100	3.4500000
1.0952146	1.0000000		
q2	The instructor communicated well.	100	3.0600000
1.2211933	1.0000000		
q3	The course materials were helpful.	99	3.0808081
1.1665120	1.0000000		
q4	Overall, I found this workshop useful.	100	3.4000000
1.1370705	1.0000000		
pretest		100	74.9700000
5.2961869	58.0000000		
posttest		100	82.0600000
6.5901978	59.0000000		
<pre> ff ffffffffffffffffffffffffff </pre>			
	Variable	Label	
Maximum			
<pre> ff </pre>			
	q1	The instructor was well prepared.	
5.0000000			
	q2	The instructor communicated well.	
5.0000000			

5.0000000	q3	The course materials were helpful.
5.0000000	q4	Overall, I found this workshop useful.
86.0000000	pretest	
98.0000000	posttest	
ff		

The UNIVARIATE Procedure
Variable: q1 (The instructor was well prepared.)

Moments

N	100	Sum Weights	
Mean	3.45	Sum Observations	
Std Deviation	1.09521457	Variance	
Skewness	-0.1759436	Kurtosis	-
Uncorrected SS	1309	Corrected SS	
Coeff Variation	31.7453498	Std Error Mean	

Basic Statistical Measures

Location		Variability	
Mean	3.450000	Std Deviation	1.09521
Median	3.000000	Variance	1.19949
Mode	3.000000	Range	4.00000
		Interquartile Range	1.00000

Tests for Location: $\mu_0=0$

Test	-Statistic-	-----p Value-----	
Student's t	t 31.50068	Pr > t	<.0001
Sign	M 50	Pr >= M	<.0001
Signed Rank	S 2525	Pr >= S	<.0001

Quantiles (Definition 5)

Quantile	Estimate
100% Max	5
99%	5
95%	5
90%	5
75% Q3	4
50% Median	3
25% Q1	3
10%	2
5%	2

1%	1
0% Min	1

The UNIVARIATE Procedure
Variable: q1 (The instructor was well prepared.)

Extreme Observations

----Lowest----

----Highest---

Value	Obs	Value	Obs
1	69	5	77
1	62	5	83
1	58	5	84
1	22	5	89
2	100	5	93

The UNIVARIATE Procedure
Variable: q2 (The instructor communicated well.)

Moments

100	N	100	Sum Weights
306	Mean	3.06	Sum Observations
1.49131313	Std Deviation	1.22119332	Variance
1.0594354	Skewness	-0.0146968	Kurtosis -
147.64	Uncorrected SS	1084	Corrected SS
0.12211933	Coeff Variation	39.9082785	Std Error Mean

Basic Statistical Measures

Location		Variability	
Mean	3.060000	Std Deviation	1.22119
Median	3.000000	Variance	1.49131
Mode	2.000000	Range	4.00000
		Interquartile Range	2.00000

NOTE: The mode displayed is the smallest of 2 modes with a count of 28.

Tests for Location: Mu0=0

Test	-Statistic-	-----p Value-----
Student's t	t 25.05746	Pr > t <.0001
Sign	M 50	Pr >= M <.0001
Signed Rank	S 2525	Pr >= S <.0001

Quantiles (Definition 5)

Quantile	Estimate
100% Max	5.0
99%	5.0
95%	5.0
90%	5.0
75% Q3	4.0
50% Median	3.0

25% Q1	2.0
10%	1.5
5%	1.0
1%	1.0
0% Min	1.0

The UNIVARIATE Procedure
Variable: q2 (The instructor communicated well.)

Extreme Observations

----Lowest----

----Highest---

Value	Obs	Value	Obs
1	88	5	77
1	86	5	89
1	75	5	90
1	68	5	93
1	64	5	96

The UNIVARIATE Procedure
Variable: q3 (The course materials were helpful.)

Moments

99	N	99	Sum Weights
305	Mean	3.08080808	Sum Observations
1.36075036	Std Deviation	1.16651205	Variance
0.7105137	Skewness	-0.0418616	Kurtosis -
133.353535	Uncorrected SS	1073	Corrected SS
0.11723887	Coeff Variation	37.8638337	Std Error Mean

Basic Statistical Measures

Location		Variability	
Mean	3.080808	Std Deviation	1.16651
Median	3.000000	Variance	1.36075
Mode	3.000000	Range	4.00000
		Interquartile Range	2.00000

Tests for Location: Mu0=0

Test	-Statistic-	-----p Value-----
Student's t	t 26.27804	Pr > t <.0001
Sign	M 49.5	Pr >= M <.0001
Signed Rank	S 2475	Pr >= S <.0001

Quantiles (Definition 5)

Quantile	Estimate
100% Max	5
99%	5
95%	5
90%	5
75% Q3	4
50% Median	3
25% Q1	2
10%	1
5%	1

1%	1
0% Min	1

The UNIVARIATE Procedure
Variable: q3 (The course materials were helpful.)

Extreme Observations

----Lowest----		----Highest---	
Value	Obs	Value	Obs
1	98	5	49
1	94	5	74
1	86	5	83
1	75	5	89
1	69	5	93

Missing Values

Missing Value	Count	-----Percent Of-----	
		All Obs	Missing Obs
.	1	1.00	100.00

The UNIVARIATE Procedure
Variable: q4 (Overall, I found this workshop useful.)

Moments

N	100	Sum Weights	
Mean	3.4	Sum Observations	
Std Deviation	1.13707049	Variance	
Skewness	-0.252394	Kurtosis	-
Uncorrected SS	1284	Corrected SS	
Coeff Variation	33.4432496	Std Error Mean	

Basic Statistical Measures

Location		Variability	
Mean	3.400000	Std Deviation	1.13707
Median	3.000000	Variance	1.29293
Mode	3.000000	Range	4.00000
		Interquartile Range	1.00000

Tests for Location: Mu0=0

Test	-Statistic-	-----p Value-----
Student's t	t 29.9014	Pr > t <.0001
Sign	M 50	Pr >= M <.0001
Signed Rank	S 2525	Pr >= S <.0001

Quantiles (Definition 5)

Quantile	Estimate
100% Max	5
99%	5
95%	5
90%	5
75% Q3	4
50% Median	3
25% Q1	3
10%	2
5%	1

1%	1
0% Min	1

The UNIVARIATE Procedure
Variable: q4 (Overall, I found this workshop useful.)

Extreme Observations

----Lowest----

----Highest---

Value	Obs	Value	Obs
1	100	5	83
1	68	5	89
1	58	5	91
1	54	5	93
1	22	5	96

The UNIVARIATE Procedure
Variable: pretest

Moments

100	N	100	Sum Weights
7497	Mean	74.97	Sum Observations
28.049596	Std Deviation	5.29618693	Variance
0.4608774	Skewness	-0.3625279	Kurtosis
2776.91	Uncorrected SS	564827	Corrected SS
0.52961869	Coeff Variation	7.06440833	Std Error Mean

Basic Statistical Measures

Location		Variability	
Mean	74.97000	Std Deviation	5.29619
Median	75.00000	Variance	28.04960
Mode	72.00000	Range	28.00000
		Interquartile Range	7.00000

Tests for Location: Mu0=0

Test	-Statistic-	-----p Value-----
Student's t	t 141.5547	Pr > t <.0001
Sign	M 50	Pr >= M <.0001
Signed Rank	S 2525	Pr >= S <.0001

Quantiles (Definition 5)

Quantile	Estimate
100% Max	86.0
99%	85.5
95%	83.0
90%	82.0
75% Q3	79.0
50% Median	75.0
25% Q1	72.0
10%	69.0
5%	66.5

1%	60.0
0% Min	58.0

Thursday, March 17, 2013 11

The SAS System

11:14

The UNIVARIATE Procedure
Variable: pretest

Extreme Observations

----Lowest----

----Highest---

Value	Obs	Value	Obs
58	92	83	91
62	82	84	57
63	51	85	60
63	48	85	62
66	81	86	70

The UNIVARIATE Procedure
Variable: posttest

Moments

100	N	100	Sum Weights
8206	Mean	82.06	Sum Observations
43.4307071	Std Deviation	6.5901978	Variance
0.84229821	Skewness	-0.2897484	Kurtosis
4299.64	Uncorrected SS	677684	Corrected SS
0.65901978	Coeff Variation	8.03095028	Std Error Mean

Basic Statistical Measures

Location		Variability	
Mean	82.06000	Std Deviation	6.59020
Median	82.00000	Variance	43.43071
Mode	81.00000	Range	39.00000
		Interquartile Range	9.00000

Tests for Location: Mu0=0

Test	-Statistic-	-----p Value-----
Student's t	t 124.5183	Pr > t <.0001
Sign	M 50	Pr >= M <.0001
Signed Rank	S 2525	Pr >= S <.0001

Quantiles (Definition 5)

Quantile	Estimate
100% Max	98.0
99%	97.0
95%	92.0
90%	90.0
75% Q3	86.0
50% Median	82.0
25% Q1	77.0
10%	75.0
5%	71.5

1%	63.0
0% Min	59.0

Thursday, March 17, 2013 13

The SAS System

11:14

The UNIVARIATE Procedure
Variable: posttest

Extreme Observations

----Lowest----

----Highest---

Value	Obs	Value	Obs
59	92	92	98
67	48	95	60
68	67	95	62
69	81	96	86
71	82	98	91

The FREQ Procedure

Cumulative Percent	workshop	Frequency	Percent	Cumulative Frequency	
ff					
	NA	1	1.00	1	1.00
	R	31	31.00	32	32.00
	SAS	24	24.00	56	56.00
	SPSS	25	25.00	81	81.00
	Stata	19	19.00	100	100.00

The instructor was well prepared.

q1	Frequency	Percent	Cumulative Frequency	Cumulative Percent
ff				
1	4	4.00	4	4.00
2	14	14.00	18	18.00
3	36	36.00	54	54.00
4	25	25.00	79	79.00
5	21	21.00	100	100.00

The instructor communicated well.

q2	Frequency	Percent	Cumulative Frequency	Cumulative Percent
ff				
1	10	10.00	10	10.00
2	28	28.00	38	38.00
3	21	21.00	59	59.00
4	28	28.00	87	87.00
5	13	13.00	100	100.00

The course materials were helpful.

q3	Frequency	Percent	Cumulative Frequency	Cumulative Percent
ff				
1	10	10.10	10	10.10
2	20	20.20	30	30.30
3	34	34.34	64	64.65
4	22	22.22	86	86.87
5	13	13.13	99	100.00

Frequency Missing = 1

The FREQ Procedure

Overall, I found this workshop useful.

q4	Frequency	Percent	Cumulative Frequency	Cumulative Percent
ff				
1	6	6.00	6	6.00
2	14	14.00	20	20.00
3	34	34.00	54	54.00
4	26	26.00	80	80.00
5	20	20.00	100	100.00

The FREQ Procedure

Table of workshop by gender

workshop	gender			
Frequency,				
Percent				
Row Pct				
Col Pct	Female	Male	NA	Total
NA	0	0	1	1
	0.00	0.00	1.00	1.00
	0.00	0.00	100.00	
	0.00	0.00	100.00	
R	14	17	0	31
	14.00	17.00	0.00	31.00
	45.16	54.84	0.00	
	29.79	32.69	0.00	
SAS	11	13	0	24
	11.00	13.00	0.00	24.00
	45.83	54.17	0.00	
	23.40	25.00	0.00	
SPSS	13	12	0	25
	13.00	12.00	0.00	25.00
	52.00	48.00	0.00	
	27.66	23.08	0.00	
Stata	9	10	0	19
	9.00	10.00	0.00	19.00
	47.37	52.63	0.00	
	19.15	19.23	0.00	
Total	47	52	1	100
	47.00	52.00	1.00	100.00

The FREQ Procedure

Statistics for Table of workshop by gender

Statistic	DF	Value	Prob
ff			
Chi-Square	8	100.3009	<.0001
Likelihood Ratio Chi-Square	8	11.4980	0.1750
Mantel-Haenszel Chi-Square	1	0.7853	0.3755
Phi Coefficient		1.0015	
Contingency Coefficient		0.7076	
Cramer's V		0.7082	

WARNING: 47% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Sample Size = 100

Thursday, March 17, 2013 18

The SAS System

11:14

The CORR Procedure

4 Variables: q1 q2 q3 q4

Simple Statistics

Variable		N	Mean	Std Dev	Sum
Minimum	Maximum				
q1		100	3.45000	1.09521	345.00000
1.00000	5.00000				
q2		100	3.06000	1.22119	306.00000
1.00000	5.00000				
q3		99	3.08081	1.16651	305.00000
1.00000	5.00000				
q4		100	3.40000	1.13707	340.00000
1.00000	5.00000				

Simple Statistics

Variable	Label
q1	The instructor was well prepared.
q2	The instructor communicated well.
q3	The course materials were helpful.
q4	Overall, I found this workshop useful.

Pearson Correlation Coefficients

Prob > |r| under H0: Rho=0

Number of Observations

		q1	q2
q3	q4		
q1		1.00000	0.66687
0.69484	0.57589		
The instructor was well prepared.			<.0001
<.0001	<.0001		
		100	100
99	100		
q2		0.66687	1.00000
0.60407	0.49174		
The instructor communicated well.		<.0001	
<.0001	<.0001		
		100	100
99	100		

q3		0.69484	0.60407
1.00000	0.47307		
The course materials were helpful.		<.0001	<.0001
<.0001		99	99
99	99		
q4		0.57589	0.49174
0.47307	1.00000		
Overall, I found this workshop useful.		<.0001	<.0001
<.0001		100	100
99	100		

The CORR Procedure

4 Variables: q1 q2 q3 q4

Simple Statistics

Variable		N	Mean	Std Dev	Median
Minimum	Maximum				
q1		100	3.45000	1.09521	3.00000
1.00000	5.00000				
q2		100	3.06000	1.22119	3.00000
1.00000	5.00000				
q3		99	3.08081	1.16651	3.00000
1.00000	5.00000				
q4		100	3.40000	1.13707	3.00000
1.00000	5.00000				

Simple Statistics

Variable	Label
q1	The instructor was well prepared.
q2	The instructor communicated well.
q3	The course materials were helpful.
q4	Overall, I found this workshop useful.

Spearman Correlation Coefficients

Prob > |r| under H0: Rho=0

Number of Observations

		q1	q2
q3	q4		
q1		1.00000	0.66833
0.66849	0.55179		
The instructor was well prepared.			<.0001
<.0001	<.0001		
		100	100
99	100		
q2		0.66833	1.00000
0.60961	0.49100		
The instructor communicated well.		<.0001	
<.0001	<.0001		
		100	100
99	100		

q3		0.66849	0.60961
1.00000	0.42701		
The course materials were helpful.		<.0001	<.0001
<.0001			
		99	99
99	99		
q4		0.55179	0.49100
0.42701	1.00000		
Overall, I found this workshop useful.		<.0001	<.0001
<.0001			
		100	100
99	100		

Thursday, March 17, 2013 20

The SAS System

11:14

The REG Procedure

Model: MODEL1

Dependent Variable: q4 Overall, I found this workshop

useful.

Number of Observations Read	100
Number of Observations Used	99
Number of Observations with Missing Values	1

Analysis of Variance

Value	Source Pr > F	DF	Sum of Squares	Mean Square	F
17.51	Model <.0001	3	45.52673	15.17558	
	Error	95	82.31165	0.86644	
	Corrected Total	98	127.83838		

Root MSE	0.93083	R-Square	0.3561
Dependent Mean	3.40404	Adj R-Sq	0.3358
Coeff Var	27.34476		

Parameter Estimates

Variable t Value	Label Pr > t	DF	Parameter Estimate	Standard Error
Intercept	Intercept	1	1.20940	0.31787
3.80	0.0003			
q1	The instructor was well prepared.	1	0.41134	0.13170
3.12	0.0024			
q2	The instructor communicated well.	1	0.15791	0.10690
1.48	0.1429			
q3	The course materials were helpful.	1	0.09372	0.11617
0.81	0.4218			

The NPAR1WAY Procedure

Analysis of Variance for Variable q1
Classified by Variable gender

gender	N	Mean
ff		
Female	47	3.978723
Male	52	2.980769
NA	1	3.000000

	Source	DF	Sum of Squares	Mean Square	F Value
Pr > F					
ff					
	Among	2	24.790507	12.395254	12.7964
<.0001	Within	97	93.959493	0.968655	

Average scores were used for ties.

The NPAR1WAY Procedure

Wilcoxon Scores (Rank Sums) for Variable q1
Classified by Variable gender

		N	Sum of Scores	Expected Under H0	Std Dev Under H0
Mean Score	gender				
////////////////////////////////////					
63.989362	Female	47	3007.50	2373.50	139.316337
38.576923	Male	52	2006.00	2626.00	139.456086
36.500000	NA	1	36.50	50.50	27.773639

Average scores were used for ties.

Kruskal-Wallis Test

Chi-Square	20.7152
DF	2
Pr > Chi-Square	<.0001

The NPAR1WAY Procedure

Median Scores (Number of Points Above Median) for Variable
q1
Classified by Variable gender

				Sum of	Expected	Std Dev
Mean	gender	N		Scores	Under H0	Under H0
Score						
////////////////////////////////////						
	Female	47		33.333333	23.50	2.322879
0.709220						
	Male	52		16.555556	26.00	2.325210
0.318376						
	NA	1		0.111111	0.50	0.463081
0.111111						

Average scores were used for ties.

Median One-Way Analysis

Chi-Square	18.1150
DF	2
Pr > Chi-Square	0.0001

The NPAR1WAY Procedure

Van der Waerden Scores (Normal) for Variable q1
Classified by Variable gender

			Sum of	Expected	Std Dev	
Mean	gender	N	Scores	Under H0	Under H0	
Score						
ff						
0.444116	Female	47	20.873461	0.0	4.597370	
0.394308	Male	52	-20.504001	0.0	4.601982	-
0.369460	NA	1	-0.369460	0.0	0.916516	-

Average scores were used for ties.

Van der Waerden One-Way Analysis

Chi-Square	20.6151
DF	2
Pr > Chi-Square	<.0001

The NPAR1WAY Procedure

Savage Scores (Exponential) for Variable q1
Classified by Variable gender

Mean			Sum of	Expected	Std Dev	
Score	gender	N	Scores	Under H0	Under H0	
////////////////////						
0.404272	Female	47	19.000764	0.0	4.349995	
0.355106	Male	52	-18.465529	0.0	4.354358	-
0.535235	NA	1	-0.535235	0.0	0.867200	-

Average scores were used for ties.

Savage One-Way Analysis

Chi-Square	19.1213
DF	2
Pr > Chi-Square	<.0001

The NPAR1WAY Procedure

Kolmogorov-Smirnov Test for Variable q1
Classified by Variable gender

gender	N	EDF at Maximum	Deviation from Mean at Maximum
ff			
Female	47	0.319149	-1.514079
Male	52	0.730769	1.375656
NA	1	1.000000	0.460000
Total	100	0.540000	

Maximum Deviation Occurred at Observation 53
Value of q1 at Maximum = 3.0

Kolmogorov-Smirnov Statistics (Asymptotic)
KS 0.209677 KSa 2.096775

Cramer-von Mises Test for Variable q1
Classified by Variable gender

gender	N	Summed Deviation from Mean
ff		
Female	47	1.187494
Male	52	0.999975
NA	1	0.091801

Cramer-von Mises Statistics (Asymptotic)
CM 0.022793 CMa 2.279270

Thursday, March 17, 2013 27

The SAS System

11:14

The TTEST Procedure

Statistics

Upper CL		Lower CL		Upper CL		Lower CL	
Difference		N	Mean	Mean	Mean	Std Dev	Std Dev
Std Dev	Std Err						
pretest - posttest		100	-8.063	-7.09	-6.117	4.3051	4.9033
5.696	0.4903						

T-Tests

Difference	DF	t Value	Pr > t
pretest - posttest	99	-14.46	<.0001

The UNIVARIATE Procedure
Variable: myDiff

Moments

100	N	100	Sum Weights	
-39	Mean	-0.39	Sum Observations	
0.9069697	Std Deviation	0.95234957	Variance	
0.9784446	Skewness	0.07143331	Kurtosis	-
89.79	Uncorrected SS	105	Corrected SS	
0.09523496	Coeff Variation	-244.1922	Std Error Mean	

Basic Statistical Measures

Location		Variability	
Mean	-0.39000	Std Deviation	0.95235
Median	-0.50000	Variance	0.90697
Mode	-1.00000	Range	3.00000
		Interquartile Range	1.00000

Tests for Location: Mu0=0

Test	-Statistic-	-----p Value-----	
Student's t	t -4.09513	Pr > t	<.0001
Sign	M -14	Pr >= M	0.0013
Signed Rank	S -632	Pr >= S	<.0001

Quantiles (Definition 5)

Quantile	Estimate
100% Max	1.0
99%	1.0
95%	1.0
90%	1.0
75% Q3	0.0
50% Median	-0.5
25% Q1	-1.0
10%	-2.0
5%	-2.0

1%	-2.0
0% Min	-2.0

Thursday, March 17, 2013 29

The SAS System

11:14

The UNIVARIATE Procedure
Variable: myDiff

Extreme Observations

----Lowest----		----Highest---	
Value	Obs	Value	Obs
-2	86	1	87
-2	84	1	90
-2	82	1	91
-2	81	1	92
-2	64	1	96

Thursday, March 17, 2013 30

The SAS System

11:14

The GLM Procedure

Class Level Information

Class	Levels	Values
workshop	5	NA R SAS SPSS Stata

Number of Observations Read	100
Number of Observations Used	100

The GLM Procedure

Dependent Variable: posttest

Value	Source Pr > F	DF	Sum of Squares	Mean Square	F
6.22	Model 0.0002	4	892.092148	223.023037	
	Error	95	3407.547852	35.868925	
	Corrected Total	99	4299.640000		

R-Square	Coeff Var	Root MSE	posttest Mean
0.207481	7.298400	5.989067	82.06000

Value	Source Pr > F	DF	Type I SS	Mean Square	F
6.22	workshop 0.0002	4	892.0921477	223.0230369	

Value	Source Pr > F	DF	Type III SS	Mean Square	F
6.22	workshop 0.0002	4	892.0921477	223.0230369	

The GLM Procedure

Tukey's Studentized Range (HSD) Test for posttest

NOTE: This test controls the Type I experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	95
Error Mean Square	35.86892
Critical Value of Studentized Range	3.93274

Comparisons significant at the 0.05 level are indicated by

***.

workshop Comparison		Difference Between Means	Simultaneous 95% Confidence Limits		
R	- SPSS	4.538	0.061	9.015	***
R	- SAS	6.633	2.105	11.161	***
R	- Stata	7.311	2.458	12.163	***
R	- NA	8.258	-8.663	25.179	
SPSS	- R	-4.538	-9.015	-0.061	***
SPSS	- SAS	2.095	-2.665	6.855	
SPSS	- Stata	2.773	-2.296	7.842	
SPSS	- NA	3.720	-13.265	20.705	
SAS	- R	-6.633	-11.161	-2.105	***
SAS	- SPSS	-2.095	-6.855	2.665	
SAS	- Stata	0.678	-4.437	5.792	
SAS	- NA	1.625	-15.373	18.623	
Stata	- R	-7.311	-12.163	-2.458	***
Stata	- SPSS	-2.773	-7.842	2.296	
Stata	- SAS	-0.678	-5.792	4.437	
Stata	- NA	0.947	-16.140	18.035	
NA	- R	-8.258	-25.179	8.663	
NA	- SPSS	-3.720	-20.705	13.265	
NA	- SAS	-1.625	-18.623	15.373	
NA	- Stata	-0.947	-18.035	16.140	

The NPAR1WAY Procedure

Analysis of Variance for Variable posttest
Classified by Variable workshop

workshop	N	Mean
ff		
R	31	86.258065
SPSS	25	81.720000
NA	1	78.000000
Stata	19	78.947368
SAS	24	79.625000

Pr > F	Source	DF	Sum of Squares	Mean Square	F Value
ff					
	Among	4	892.092148	223.023037	6.2177
0.0002	Within	95	3407.547852	35.868925	

Average scores were used for ties.

The NPAR1WAY Procedure

Wilcoxon Scores (Rank Sums) for Variable posttest
Classified by Variable workshop

Mean			Sum of	Expected	Std Dev
Score	workshop	N	Scores	Under H0	Under H0
ff					
69.887097	R	31	2166.50	1565.50	133.960239
47.960000	SPSS	25	1199.00	1262.50	125.421260
29.000000	NA	1	29.00	50.50	28.819611
36.473684	Stata	19	693.00	959.50	113.629170
40.104167	SAS	24	962.50	1212.00	123.703771

Average scores were used for ties.

Kruskal-Wallis Test

Chi-Square	22.1786
DF	4
Pr > Chi-Square	0.0002

The NPAR1WAY Procedure

Median Scores (Number of Points Above Median) for Variable
posttest
Classified by Variable workshop

Mean			Sum of	Expected	Std Dev
Score	workshop	N	Scores	Under H0	Under H0
ff					
0.774194	R	31	24.00	15.50	2.267651
0.448000	SPSS	25	11.20	12.50	2.123105
0.000000	NA	1	0.00	0.50	0.487852
0.347368	Stata	19	6.60	9.50	1.923491
0.341667	SAS	24	8.20	12.00	2.094032

Average scores were used for ties.

Median One-Way Analysis

Chi-Square	15.3597
DF	4
Pr > Chi-Square	0.0040

The NPAR1WAY Procedure

Van der Waerden Scores (Normal) for Variable posttest
Classified by Variable workshop

Mean			Sum of	Expected	Std Dev	
Score	workshop	N	Scores	Under H0	Under H0	
ff						
0.628863	R	31	19.494765	0.0	4.457588	
0.063804	SPSS	25	-1.595099	0.0	4.173450	-
0.562268	NA	1	-0.562268	0.0	0.958986	-
0.461561	Stata	19	-8.769668	0.0	3.781063	-
0.356989	SAS	24	-8.567730	0.0	4.116300	-

Average scores were used for ties.

Van der Waerden One-Way Analysis

Chi-Square	21.2971
DF	4
Pr > Chi-Square	0.0003

The NPAR1WAY Procedure

Savage Scores (Exponential) for Variable posttest
Classified by Variable workshop

Mean			Sum of	Expected	Std Dev	
Score	workshop	N	Scores	Under H0	Under H0	
////////////////////						
0.579366	R	31	17.960343	0.0	4.517077	
0.197242	SPSS	25	-4.931054	0.0	4.229146	-
0.659348	NA	1	-0.659348	0.0	0.971784	-
0.223208	Stata	19	-4.240955	0.0	3.831522	-
0.338708	SAS	24	-8.128986	0.0	4.171233	-

Average scores were used for ties.

Savage One-Way Analysis

Chi-Square	16.2626
DF	4
Pr > Chi-Square	0.0027

The NPAR1WAY Procedure

Kolmogorov-Smirnov Test for Variable posttest
Classified by Variable workshop

workshop	N	EDF at Maximum	Deviation from Mean at Maximum
ff			
R	31	0.258065	-1.792461
SPSS	25	0.720000	0.700000
NA	1	1.000000	0.420000
Stata	19	0.736842	0.683659
SAS	24	0.708333	0.628702
Total	100	0.580000	

Maximum Deviation Occurred at Observation 50
Value of posttest at Maximum = 83.0

Kolmogorov-Smirnov Statistics (Asymptotic)
KS 0.217761 KSa 2.177607

Cramer-von Mises Test for Variable posttest
Classified by Variable workshop

workshop	N	Summed Deviation from Mean
fff		
R	31	1.461853
SPSS	25	0.169875
NA	1	0.127763
Stata	19	0.502013
SAS	24	0.355879

Cramer-von Mises Statistics (Asymptotic)

CM 0.026174 CMa 2.617382

LOG FILE

NOTE: AUTOEXEC processing completed.

```
1 * SAS Program for Basic Statistical Tests;
2 * Filename: Statistics.sas ;
3
4 LIBNAME myLib 'K:\myRfolder';
```

NOTE: Libref MYLIB was successfully assigned as follows:

```
Engine:          V9
Physical Name: K:\myRfolder
5  DATA  temp;
6    SET myLib.mydata100;
7    * pretend q2 and q1 are the same score
8      measured at two times & subtract;
9    myDiff=q2-q1; run;
```

NOTE: There were 100 observations read from the data set MYLIB.MYDATA100.

NOTE: The data set WORK.TEMP has 100 observations and 10 variables.

NOTE: DATA statement used (Total process time):

```
real time          0.34 seconds
cpu time           0.07 seconds
```

10

```
11  * Basic stats in compact form;
12  PROC MEANS; VAR q1--posttest; RUN;
```

NOTE: There were 100 observations read from the data set WORK.TEMP.

NOTE: PROCEDURE MEANS used (Total process time):

```
real time          0.15 seconds
cpu time           0.04 seconds
```

13

```
14  * Basic stats of every sort;
15  PROC UNIVARIATE; VAR q1--posttest; RUN;
```

NOTE: PROCEDURE UNIVARIATE used (Total process time):

```
real time          0.10 seconds
cpu time           0.03 seconds
```

16

```
17  * Frequencies & percents;
18  PROC FREQ; TABLES workshop--q4; RUN;
```

NOTE: There were 100 observations read from the data set WORK.TEMP.

NOTE: PROCEDURE FREQ used (Total process time):

```
real time          0.35 seconds
cpu time           0.04 seconds
```

19

```
20  * Chi-square;
21  PROC FREQ;
22    TABLES workshop*gender/CHISQ; RUN;
```

NOTE: There were 100 observations read from the data set WORK.TEMP.

NOTE: PROCEDURE FREQ used (Total process time):

```
real time          0.34 seconds
cpu time           0.03 seconds
```

```

23
24 * ---Measures of association---;
25
26 * Pearson correlations;
27 PROC CORR; VAR q1-q4; RUN;

NOTE: PROCEDURE CORR used (Total process time):
      real time          0.01 seconds
      cpu time           0.00 seconds

28
29 * Spearman correlations;
30 PROC CORR SPEARMAN; VAR q1-q4; RUN;

NOTE: PROCEDURE CORR used (Total process time):
      real time          0.01 seconds
      cpu time           0.01 seconds

31
32 * Linear regression;
33 PROC REG;
34     MODEL q4=q1-q3;
35     RUN;

36
37 * ---Group comparisons---;
38
39 * Independent samples t-test;

NOTE: PROCEDURE REG used (Total process time):
      real time          1.40 seconds
      cpu time           0.03 seconds

40 PROC TTEST;
41     CLASS gender;
42     VAR q1; RUN;

ERROR: The CLASS variable has more than two levels.
NOTE: The SAS System stopped processing this step because of errors.
NOTE: There were 3 observations read from the data set WORK.TEMP.
NOTE: PROCEDURE TTEST used (Total process time):
      real time          0.01 seconds
      cpu time           0.01 seconds

43
44 * Nonparametric version of above
45     using Wilcoxon/Mann-Whitney test;

46 PROC NPAR1WAY;
47     CLASS gender;

```

```
48     VAR q1; RUN;
```

NOTE: There were 100 observations read from the data set WORK.TEMP.

NOTE: PROCEDURE NPAR1WAY used (Total process time):

real time 0.18 seconds

cpu time 0.06 seconds

```
49
```

```
50     * Paired samples t-test;
```

```
51     PROC TTEST;
```

```
52     PAIRED pretest*posttest; RUN;
```

NOTE: There were 100 observations read from the data set WORK.TEMP.

NOTE: PROCEDURE TTEST used (Total process time):

real time 0.00 seconds

cpu time 0.00 seconds

```
53
```

```
54     * Nonparametric version of above using
```

```
55     both Signed Rank test and Sign test;
```

```
56     PROC UNIVARIATE;
```

```
57     VAR myDiff;
```

```
58     RUN;
```

NOTE: PROCEDURE UNIVARIATE used (Total process time):

real time 0.01 seconds

cpu time 0.01 seconds

```
59
```

```
60     *Oneway Analysis of Variance (ANOVA);
```

```
61     PROC GLM;
```

```
62     CLASS workshop;
```

```
63     MODEL posttest=workshop;
```

```
64     MEANS workshop / TUKEY; RUN;
```

```
65
```

```
66     *Nonparametric version of above using
```

```
67     Kruskal-Wallis test;
```

NOTE: PROCEDURE GLM used (Total process time):

real time 0.71 seconds

cpu time 0.12 seconds

```
68     PROC npar1way;
```

```
69     CLASS workshop;
```

```
70     VAR posttest; RUN;
```

NOTE: There were 100 observations read from the data set WORK.TEMP.

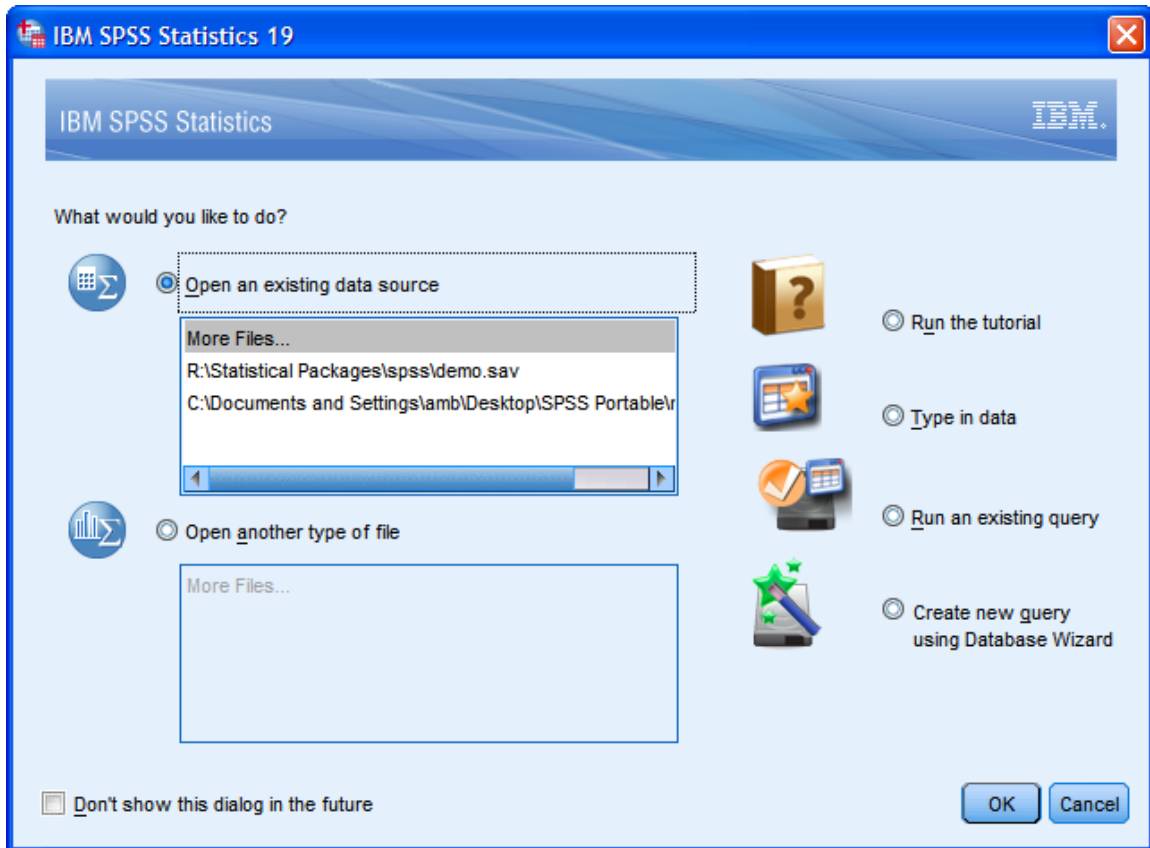
NOTE: PROCEDURE NPAR1WAY used (Total process time):

real time 0.09 seconds

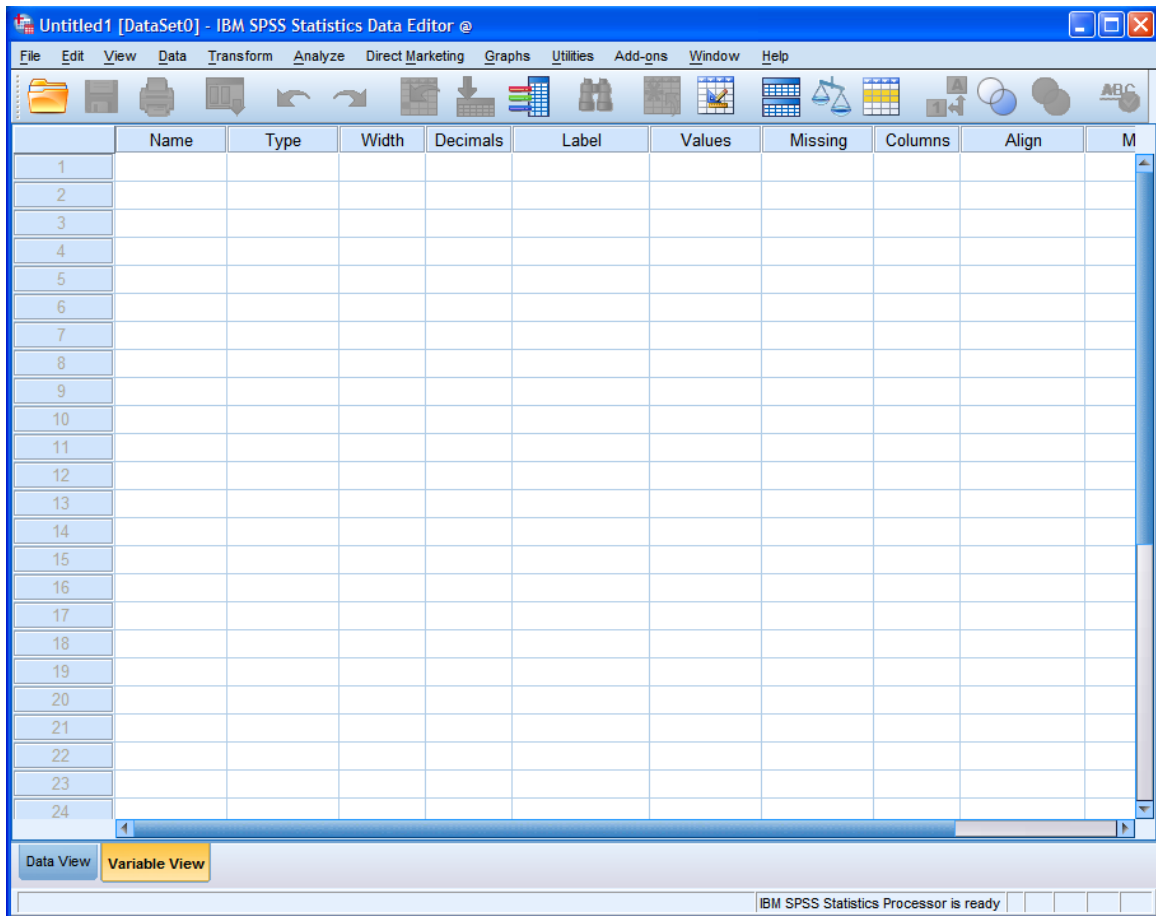
cpu time 0.04 seconds

مقدمة للحزمة SPSS

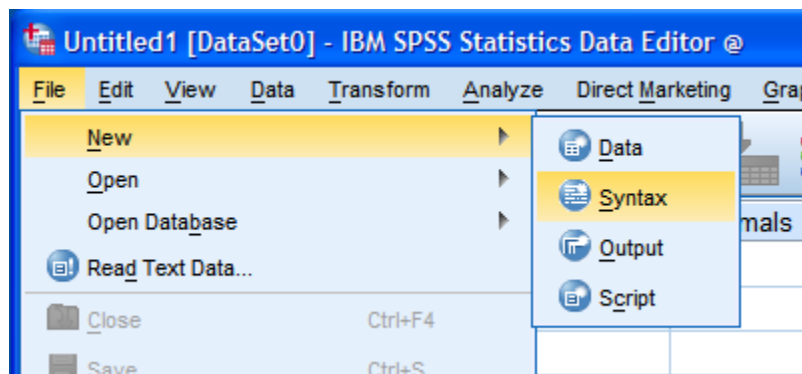
عند إجراء البرنامج تظهر نوافذ الحوار



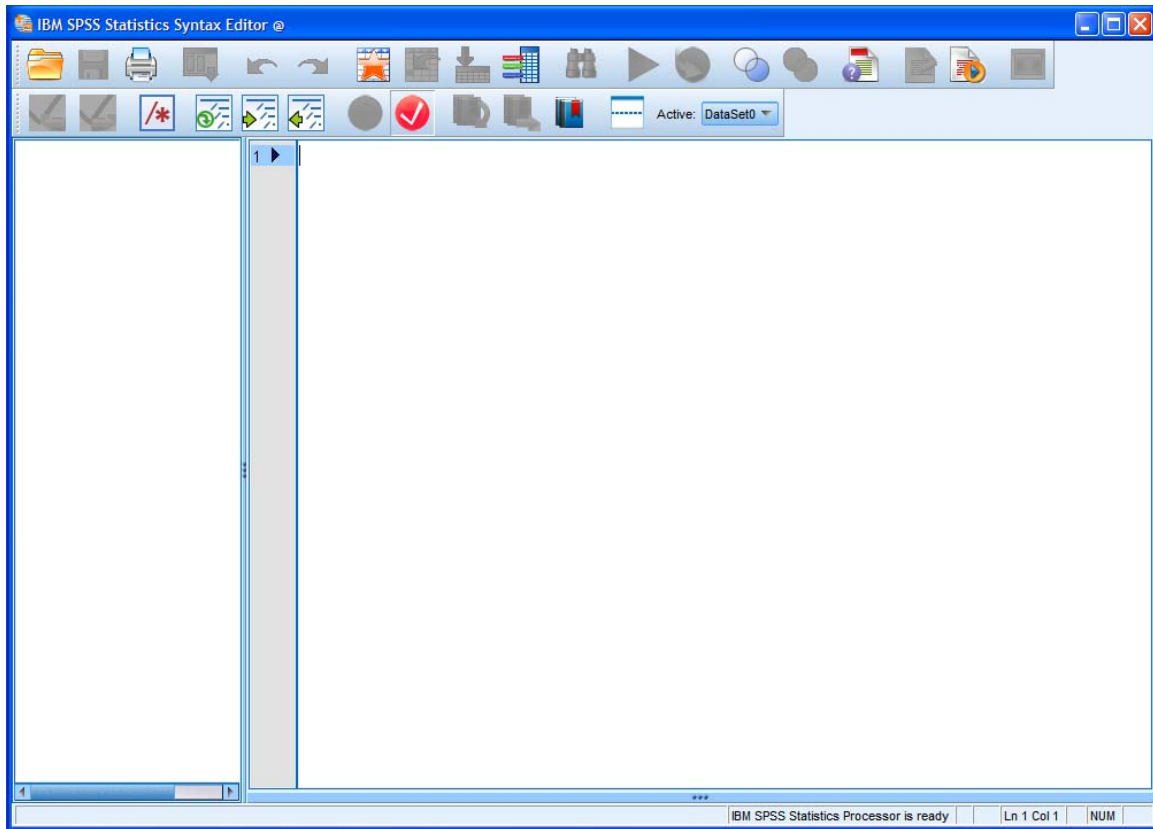
أغلق هذه النافذة لأننا لا نحتاج إليها وتبقى النافذة



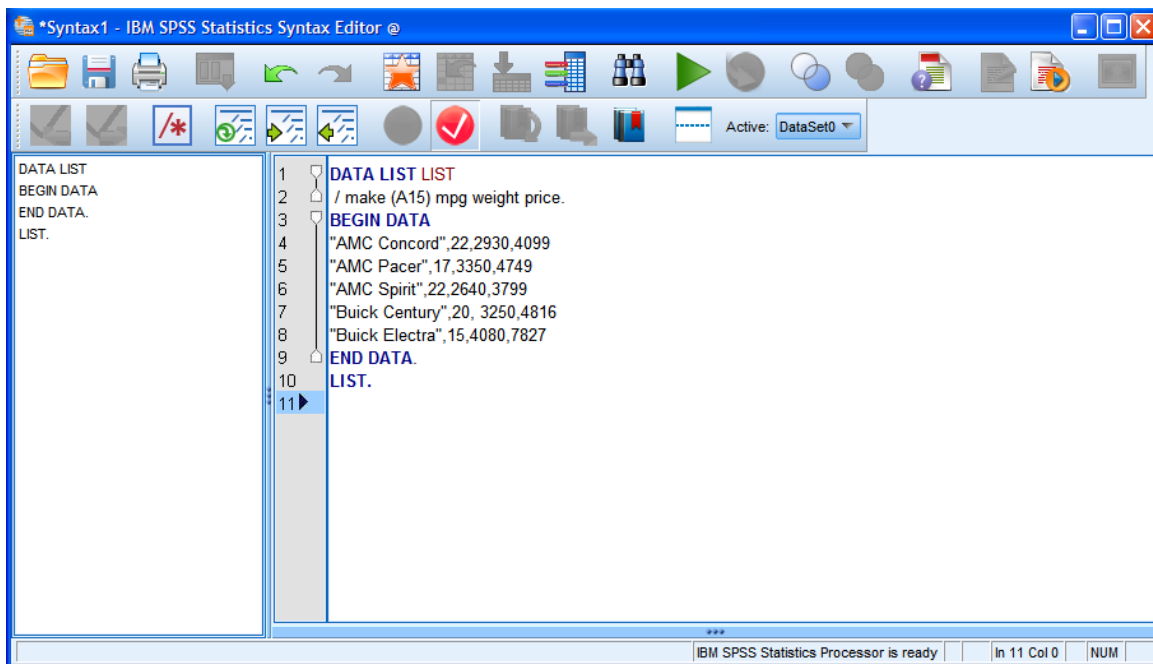
أختار File => New => Syntax



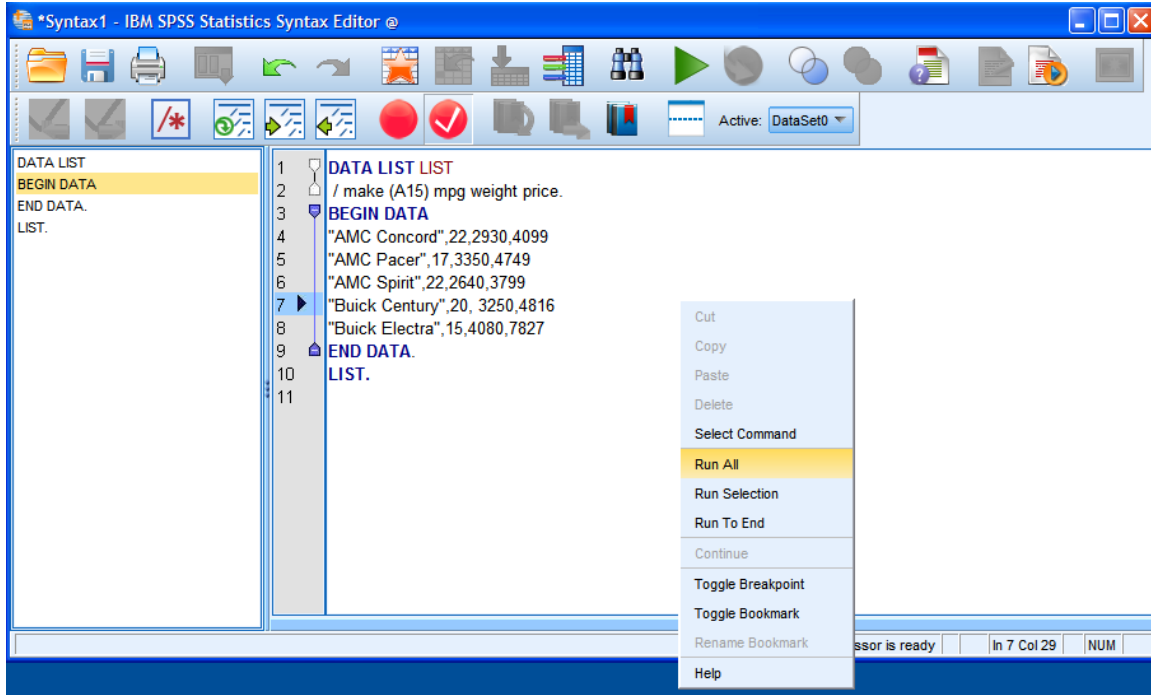
فتظهر النافذة



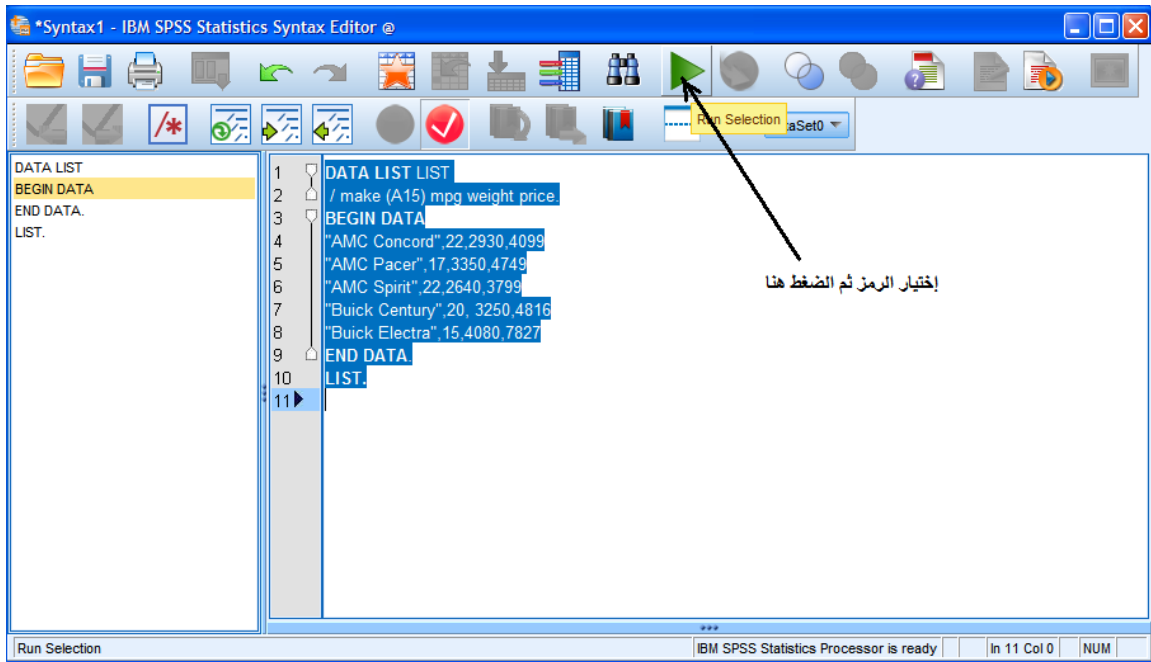
أنسخ و ألصق ترميز البرامج مثلا



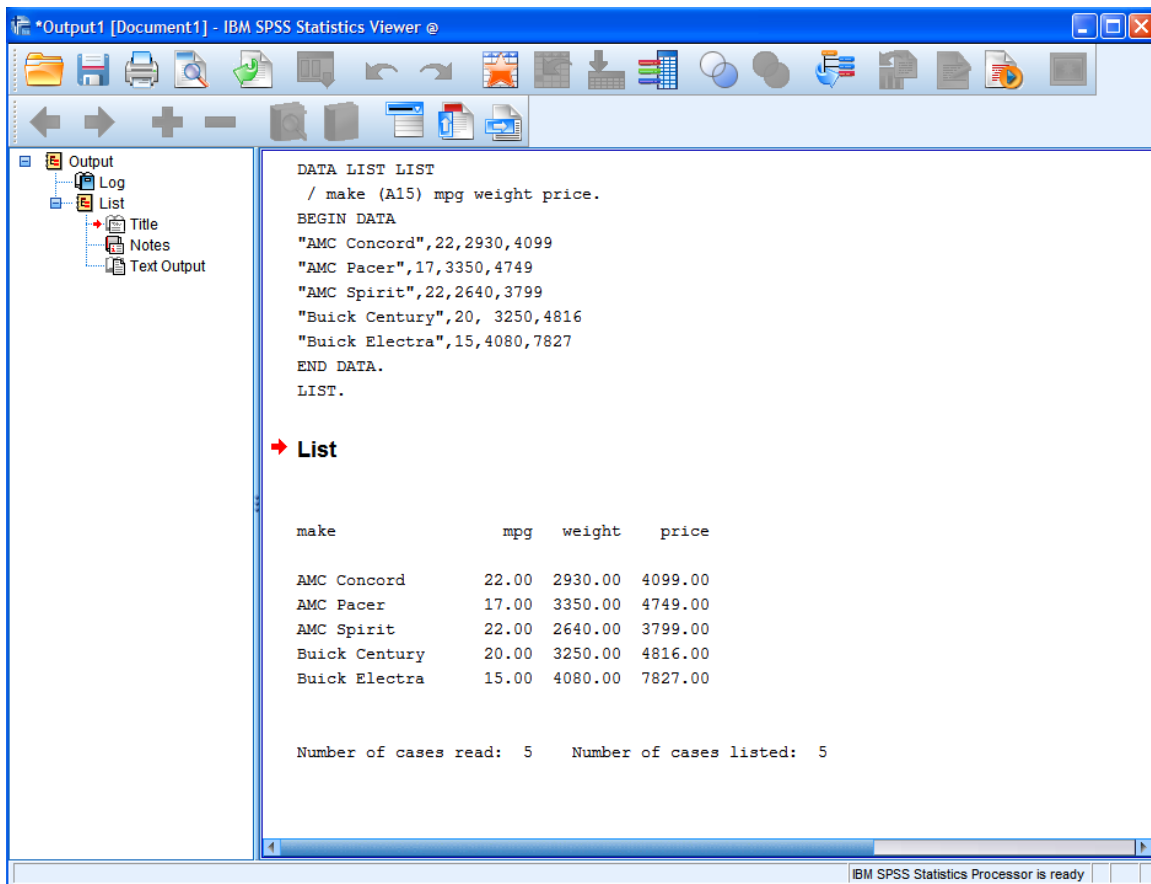
الآن يمكنك إجراء البرنامج بالضغط بالفارة اليمنى فيظهر



أختار Run All .
والطريقة الثانية بإختيار الترميز



وتظهر نافذة الإخراج



Entering Data إدخال البيانات

Reading in an Excel file إكسل من ملف

```
get data
/type=xls
/file = 'c:\spss_data\hs0.xls'
/sheet=name 'hs0'
/readnames=on.
```

Reading in a comma-separated- values text file
القراءة من ملف قيم مفصولة بفواصل

```
get data
/type = txt
/file = 'c:\spss_data\hs0.csv'
/delimiters = ","
/firstcase = 2
/variables =
gender f1.0
id f3.0
race f1.0
ses f1.0
schtype f1.0
prog a10
read f2.0
write f2.0
math f2.0
science f2.0
socst f2.0.
```

```
execute.
```

Reading in an ASCII fixed-format text file نصي من ملف

```
data list fixed file="c:\spss_data\schdat.fix"
/ id 1-2 a1 3-4 t1 5-6 gender 7 a2 8-9 t2 10-11 tgender 12.
list

/cases = from 1 to 10.
```

Inputting data via the syntax editor إدخال البيانات من خلال محرر اللغة

```

data list list
  /id female race ses * schtype (A3) prog read write math science socst.
begin data.
147 1 1 3 pub 1 47 62 53 53 61
108 0 1 2 pub 2 34 33 41 36 36
 18 0 3 2 pub 3 50 33 49 44 36
153 0 1 2 pub 3 39 31 40 39 51
 50 0 2 2 pub 2 50 59 42 53 61
 51 1 2 1 pub 2 42 36 42 31 39
102 0 1 1 pub 1 52 41 51 53 56
 57 1 1 2 pub 1 71 65 72 66 56
160 1 1 2 pub 1 55 65 55 50 61
136 0 1 2 pub 1 65 59 70 63 51

end data.

```

فتح ملف بيانات SPSS Opening an SPSS data file

```
get file 'c:\spss_data\hs0.sav'.
```

بيانات مدخلة مع البرنامج محددة بفاصلة Comma delimited data, inline

```

DATA LIST LIST
  / make (A15) mpg weight price .
BEGIN DATA.
"AMC Concord",22,2930,4099
"AMC Pacer",17,3350,4749
"AMC Spirit",22,2640,3799
"Buick Century",20, 3250,4816
"Buick Electra",15,4080,7827
END DATA.

```

```
LIST.
```

MAKE	MPG	WEIGHT	PRICE
AMC Concord	22.00	2930.00	4099.00
AMC Pacer	17.00	3350.00	4749.00
AMC Spirit	22.00	2640.00	3799.00
Buick Century	20.00	3250.00	4816.00
Buick Electra	15.00	4080.00	7827.00

بيانات مدخلة مع البرنامج محددة بفراغ Space delimited data, inline

```

DATA LIST LIST
  / make (A15) mpg weight price .
BEGIN DATA.
"AMC Concord" 22 2930 4099
"AMC Pacer" 17 3350 4749
"AMC Spirit" 22 2640 3799
"Buick Century" 20 3250 4816
"Buick Electra" 15 4080 7827

```

END DATA.

LIST.

Comma delimited data from an external file بيانات محددة بفاصلة مدخلة من ملف خارجي

```
DATA LIST LIST FILE="auto1.txt"  
/ make (A15) mpg weight price.
```

LIST.

Space delimited data from an external file بيانات محددة بفراغات مدخلة من ملف خارجي

```
DATA LIST LIST FILE="auto2.txt"  
/ make (A15) mpg weight price.
```

LIST.

Fixed format data from an external file بيانات ذات تشكيل ثابت مدخلة من ملف خارجي

```
AMC Concord    22 2930 4099  
AMC Pacer      17 3350 4749  
AMC Spirit     22 2640 3799  
Buick Century  20 3250 4816  
Buick Electra  15 4080 7827
```

```
DATA LIST FIXED FILE="auto3.txt"  
/ make 1-13 (A) mpg 15-16 weight 18-21 price 23-26.
```

LIST.

```
MAKE          MPG WEIGHT PRICE  
AMC Concord   22  2930  4099  
AMC Pacer     17  3350  4749  
AMC Spirit    22  2640  3799  
Buick Century 20  3250  4816  
Buick Electra 15  4080  7827
```

ملفات محددة بمؤشر الجدولة وملفات إكسل files

make	mpg	weight	price
AMC Concord	22	2930	4099
AMC Pacer	17	3350	4749
AMC Spirit	22	2640	3799
Buick Century	20	3250	4816
Buick Electra	15	4080	7827

```
GET TRANSLATE FILE="auto4.txt"  
  /TYPE=TAB  
  /FIELDNAMES.
```

```
LIST.
```

```
GET TRANSLATE FILE="auto5.txt"  
  /TYPE=TAB.  
RENAME VARIABLES (var1 to var4 = make mpg weight price).
```

```
LIST.
```

```
GET TRANSLATE FILE="auto.xls"  
  /TYPE=XLS  
  /FIELDNAMES.
```

```
LIST.
```

القراءة من ملفات SAS data files into SPSS

```
get sas data="d:\hsb2.sas7bdat".  
exe.
```

:Descriptive statistics الإحصاء الوصفي

سوف نستخدم البيانات التالية

MAKE	PRICE	MPG	REP78	FOREIGN
AMC	4099	22	3	0
AMC	4749	17	3	0
AMC	3799	22	3	0
Audi	9690	17	5	1
Audi	6295	23	3	1
BMW	9735	25	4	1
Buick	4816	20	3	0
Buick	7827	15	4	0

```

Buick  5788 18 3 0
Buick  4453 26 3 0
Buick  5189 20 3 0
Buick 10372 16 3 0
Buick  4082 19 3 0
Cad.   11385 14 3 0
Cad.   14500 14 2 0
Cad.   15906 21 3 0
Chev.   3299 29 3 0
Chev.   5705 16 4 0
Chev.   4504 22 3 0
Chev.   5104 22 2 0
Chev.   3667 24 2 0
Chev.   3955 19 3 0
Datsun 6229 23 4 1
Datsun 4589 35 5 1
Datsun 5079 24 4 1
Datsun 8129 21 4 1

```

```

DATA LIST FREE/
  make (A8) price mpg  rep78 foreign .

```

```

BEGIN DATA.

```

```

AMC      4099 22 3 0
AMC      4749 17 3 0
AMC      3799 22 3 0
Audi     9690 17 5 1
Audi     6295 23 3 1
BMW      9735 25 4 1
Buick    4816 20 3 0
Buick    7827 15 4 0
Buick    5788 18 3 0
Buick    4453 26 3 0
Buick    5189 20 3 0
Buick   10372 16 3 0
Buick    4082 19 3 0
Cad.     11385 14 3 0
Cad.     14500 14 2 0
Cad.     15906 21 3 0
Chev.     3299 29 3 0
Chev.     5705 16 4 0
Chev.     4504 22 3 0
Chev.     5104 22 2 0
Chev.     3667 24 2 0
Chev.     3955 19 3 0
Datsun   6229 23 4 1
Datsun   4589 35 5 1
Datsun   5079 24 4 1
Datsun   8129 21 4 1

```

```

END DATA.

```

```

EXECUTE.

```

```

LIST

```

```

  /CASES=10.

```

```

EXECUTE.

```

make	price	mpg	rep78	foreign
AMC	4099.00	22.00	3.00	.00
AMC	4749.00	17.00	3.00	.00
AMC	3799.00	22.00	3.00	.00
Audi	9690.00	17.00	5.00	1.00
Audi	6295.00	23.00	3.00	1.00
BMW	9735.00	25.00	4.00	1.00
Buick	4816.00	20.00	3.00	.00
Buick	7827.00	15.00	4.00	.00
Buick	5788.00	18.00	3.00	.00
Buick	4453.00	26.00	3.00	.00

Number of cases read: 10 Number of cases listed: 10

Using the **FREQ** و الأمر **CROSSTABS** للتعداد
frequencies or crosstabs command for counts

FREQ
/VARIABLES= make.

FREQ
/VARIABLES= rep78.

FREQ
/VAR= foreign.

Statistics

make		
N	Valid	26
	Missing	0

		make			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	AMC	3	11.5	11.5	11.5
	Audi	2	7.7	7.7	19.2
	BMW	1	3.8	3.8	23.1
	Buick	7	26.9	26.9	50.0
	Cad.	3	11.5	11.5	61.5
	Chev.	6	23.1	23.1	84.6
	Datsun	4	15.4	15.4	100.0
	Total	26	100.0	100.0	

Statistics

rep78

N	Valid	26
	Missing	0

rep78

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	2.00	3	11.5	11.5	11.5
	3.00	15	57.7	57.7	69.2
	4.00	6	23.1	23.1	92.3
	5.00	2	7.7	7.7	100.0
Total		26	100.0	100.0	

Statistics

foreign

N	Valid	26
	Missing	0

foreign

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	19	73.1	73.1	73.1
	1.00	7	26.9	26.9	100.0
Total		26	100.0	100.0	

FREQ

/VARIABLES= make rep78 foreign.

CROSSTABS

/TABLES=rep78 BY foreign.

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
rep78 * foreign	26	100.0%	0	.0%	26	100.0%

rep78 * foreign Crosstabulation

Count		foreign		Total
		.00	1.00	
rep78	2.00	3	0	3
	3.00	14	1	15
	4.00	2	4	6
	5.00	0	2	2
Total		19	7	26

CROSSTABS

/TABLES=rep78 BY foreign

/CELLS= COUNT ROW COLUMN TOTAL.

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
rep78 * foreign	26	100.0%	0	.0%	26	100.0%

rep78 * foreign Crosstabulation

			foreign		Total
			.00	1.00	
rep78	2.00	Count	3	0	3
		% within rep78	100.0%	.0%	100.0%
		% within foreign	15.8%	.0%	11.5%
		% of Total	11.5%	.0%	11.5%
	3.00	Count	14	1	15
		% within rep78	93.3%	6.7%	100.0%
		% within foreign	73.7%	14.3%	57.7%
		% of Total	53.8%	3.8%	57.7%
	4.00	Count	2	4	6
		% within rep78	33.3%	66.7%	100.0%
		% within foreign	10.5%	57.1%	23.1%
		% of Total	7.7%	15.4%	23.1%
	5.00	Count	0	2	2
		% within rep78	.0%	100.0%	100.0%
		% within foreign	.0%	28.6%	7.7%
		% of Total	.0%	7.7%	7.7%
	Total	Count	19	7	26
		% within rep78	73.1%	26.9%	100.0%
		% within foreign	100.0%	100.0%	100.0%
		% of Total	73.1%	26.9%	100.0%

CROSSTABS

/TABLES=rep78 BY foreign

/CELLS= TOTAL COUNT ROW COLUMN.

إستخدام الأوامر DESCRIPTIVES و MEANS للتلخيص
descriptives or means command for summary statistics

DESCRIPTIVES

/VARIABLES=mpg.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
mpg	26	14.00	35.00	20.9231	4.75750
Valid N (listwise)	26				

MEANS

/TABLES=mpg BY foreign.

Case Processing Summary

	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
mpg * foreign	26	100.0%	0	.0%	26	100.0%

Report

mpg			
foreign	Mean	N	Std. Deviation
.00	19.7895	19	4.03566
1.00	24.0000	7	5.50757
Total	20.9231	26	4.75750

Using the examine command **EXAMINE** لتلخيص مفصل استخدام الأمر
for detailed summary statistics

EXAMINE
/VARIABLES=mpg.

Below are the results of the **examine** command.

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
mpg	26	100.0%	0	.0%	26	100.0%

Descriptives

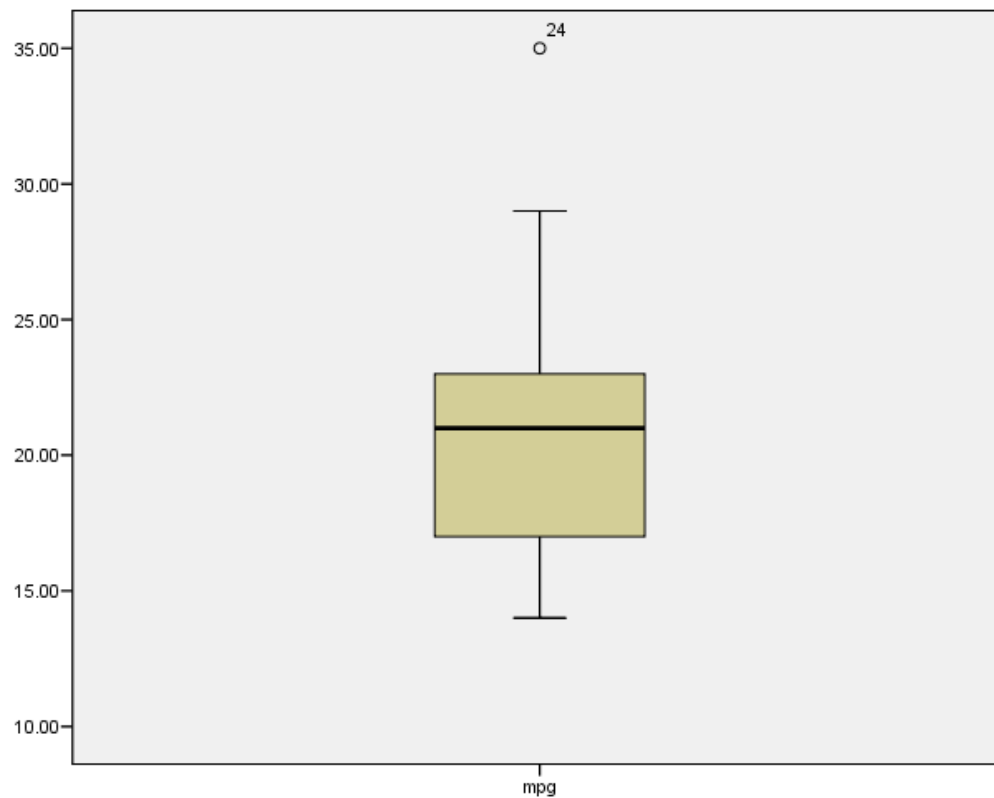
			Statistic	Std. Error
mpg	Mean		20.9231	.93302
	95% Confidence Interval for Mean	Lower Bound	19.0015	
		Upper Bound	22.8447	
	5% Trimmed Mean		20.6026	
	Median		21.0000	
	Variance		22.634	
	Std. Deviation		4.75750	
	Minimum		14.00	
	Maximum		35.00	
	Range		21.00	
	Interquartile Range		6.25	
	Skewness		.935	.456
	Kurtosis		1.793	.887

mpg

mpg Stem-and-Leaf Plot

Frequency	Stem &	Leaf
.00	1 .	
3.00	1 .	445
4.00	1 .	6677
3.00	1 .	899
4.00	2 .	0011
6.00	2 .	222233
3.00	2 .	445
1.00	2 .	6
1.00	2 .	9
1.00	Extremes	(>=35)

Stem width: 10.00
Each leaf: 1 case(s)



نظرة للإختبارات الإحصائية في SPSS An overview of statistical tests in SPSS

سوف نستخدم البيانات

```
DATA LIST FIXED/
  make (A17) price 19-23 mpg 25-26 rep78 28 hdroom 30-32
  trunk 34-35 weight 37-40 length 42-44 turn 46-47
  displ 49-51 gratio 53-56 foreign 58 .
BEGIN DATA.
AMC Concord      4099      3 2.5 11 2930 186 40 121 3.58 0
AMC Pacer        4749      3 3.0 11 3350 173 40 258 2.53 0
AMC Spirit       3799      3.0 12 2640 168 35 121 3.08 0
Audi 5000        9690     17 5 3.0 15 2830 189 37 131 3.20 1
Audi Fox         6295     23 3 2.5 11 2070 174 36  97 3.70 1
BMW 320i         9735     25 4 2.5 12 2650 177 34 121 3.64 1
Buick Century    4816     20 3 4.5 16 3250 196 40 196 2.93 0
Buick Electra    7827     15 4 4.0 20 4080 222 43 350 2.41 0
Buick LeSabre   5788     18 3 4.0 21 3670 218 43 231 2.73 0
Buick Opel       4453     26  3.0 10 2230 170 34 304 2.87 0
Buick Regal      5189     20 3 2.0 16 3280 200 42 196 2.93 0
Buick Riviera   10372     16 3 3.5 17 3880 207 43 231 2.93 0
```

Buick Skylark	4082	19	3	3.5	13	3400	200	42	231	3.08	0
Cad. Deville	11385	14	3	4.0	20	4330	221	44	425	2.28	0
Cad. Eldorado	14500	14	2	3.5	16	3900	204	43	350	2.19	0
Cad. Seville	15906	21	3	3.0	13	4290	204	45	350	2.24	0
Chev. Chevette	3299	29	3	2.5	9	2110	163	34	231	2.93	0
Chev. Impala	5705	16	4	4.0	20	3690	212	43	250	2.56	0
Chev. Malibu	4504	22	3	3.5	17	3180	193	31	200	2.73	0
Chev. Monte Carlo	5104	22	2	2.0	16	3220	200	41	200	2.73	0
Chev. Monza	3667	24	2	2.0	7	2750	179	40	151	2.73	0
Chev. Nova	3955	19	3	3.5	13	3430	197	43	250	2.56	0
Datsun 200	6229	23	4	1.5	6	2370	170	35	119	3.89	1
Datsun 210	4589	35	5	2.0	8	2020	165	32	85	3.70	1
Datsun 510	5079	24	4	2.5	8	2280	170	34	119	3.54	1
Datsun 810	8129	21	4	2.5	8	2750	184	38	146	3.55	1
Dodge Colt	3984	30	5	2.0	8	2120	163	35	98	3.54	0
Dodge Diplomat	4010	18	2	4.0	17	3600	206	46	318	2.47	0
Dodge Magnum	5886	16	2	4.0	17	3600	206	46	318	2.47	0
Dodge St. Regis	6342	17	2	4.5	21	3740	220	46	225	2.94	0
Fiat Strada	4296	21	3	2.5	16	2130	161	36	105	3.37	1
Ford Fiesta	4389	28	4	1.5	9	1800	147	33	98	3.15	0
Ford Mustang	4187	21	3	2.0	10	2650	179	43	140	3.08	0
Honda Accord	5799	25	5	3.0	10	2240	172	36	107	3.05	1
Honda Civic	4499	28	4	2.5	5	1760	149	34	91	3.30	1
Linc. Continental	11497	12	3	3.5	22	4840	233	51	400	2.47	0
Linc. Mark V	13594	12	3	2.5	18	4720	230	48	400	2.47	0
Linc. Versailles	13466	14	3	3.5	15	3830	201	41	302	2.47	0
Mazda GLC	3995	30	4	3.5	11	1980	154	33	86	3.73	1
Merc. Bobcat	3829	22	4	3.0	9	2580	169	39	140	2.73	0
Merc. Cougar	5379	14	4	3.5	16	4060	221	48	302	2.75	0
Merc. Marquis	6165	15	3	3.5	23	3720	212	44	302	2.26	0
Merc. Monarch	4516	18	3	3.0	15	3370	198	41	250	2.43	0
Merc. XR-7	6303	14	4	3.0	16	4130	217	45	302	2.75	0
Merc. Zephyr	3291	20	3	3.5	17	2830	195	43	140	3.08	0
Olds 98	8814	21	4	4.0	20	4060	220	43	350	2.41	0
Olds Cutl Supr	5172	19	3	2.0	16	3310	198	42	231	2.93	0
Olds Cutlass	4733	19	3	4.5	16	3300	198	42	231	2.93	0
Olds Delta 88	4890	18	4	4.0	20	3690	218	42	231	2.73	0
Olds Omega	4181	19	3	4.5	14	3370	200	43	231	3.08	0
Olds Starfire	4195	24	1	2.0	10	2730	180	40	151	2.73	0
Olds Toronado	10371	16	3	3.5	17	4030	206	43	350	2.41	0
Peugeot 604	12990	14		3.5	14	3420	192	38	163	3.58	1
Plym. Arrow	4647	28	3	2.0	11	3260	170	37	156	3.05	0
Plym. Champ	4425	34	5	2.5	11	1800	157	37	86	2.97	0
Plym. Horizon	4482	25	3	4.0	17	2200	165	36	105	3.37	0
Plym. Sapporo	6486	26		1.5	8	2520	182	38	119	3.54	0
Plym. Volare	4060	18	2	5.0	16	3330	201	44	225	3.23	0
Pont. Catalina	5798	18	4	4.0	20	3700	214	42	231	2.73	0
Pont. Firebird	4934	18	1	1.5	7	3470	198	42	231	3.08	0
Pont. Grand Prix	5222	19	3	2.0	16	3210	201	45	231	2.93	0
Pont. Le Mans	4723	19	3	3.5	17	3200	199	40	231	2.93	0
Pont. Phoenix	4424	19		3.5	13	3420	203	43	231	3.08	0
Pont. Sunbird	4172	24	2	2.0	7	2690	179	41	151	2.73	0
Renault Le Car	3895	26	3	3.0	10	1830	142	34	79	3.72	1
Subaru	3798	35	5	2.5	11	2050	164	36	97	3.81	1
Toyota Celica	5899	18	5	2.5	14	2410	174	36	134	3.06	1
Toyota Corolla	3748	31	5	3.0	9	2200	165	35	97	3.21	1
Toyota Corona	5719	18	5	2.0	11	2670	175	36	134	3.05	1

```

Volvo 260      11995 17 5 2.5 14 3170 193 37 163 2.98 1
VW Dasher     7140 23 4 2.5 12 2160 172 36 97 3.74 1
VW Diesel     5397 41 5 3.0 15 2040 155 35 90 3.78 1
VW Rabbit     4697 25 4 3.0 15 1930 155 35 89 3.78 1
VW Scirocco   6850 25 4 2.0 16 1990 156 36 97 3.78 1
END DATA.
FORMATS hdroom (F3.1) gratio (F4.2) .

```

T-tests t إختبار

T-TEST

```

/GROUPS=foreign(0 1)
/VARIABLES=mpg.

```

Group Statistics

foreign	N	Mean	Std. Deviation	Std. Error Mean
mpg 0	49	19.80	4.852	.693
1	22	24.77	6.611	1.410

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
mpg	Equal variances assumed	1.618	.208	-3.560	69	.001	-4.977	1.398	-7.766	-2.188
	Equal variances not assumed			-3.169	31.577	.003	-4.977	1.571	-8.178	-1.776

Chi-square tests إختبار مربع كاي

CROSSTABS

```

/TABLES=rep78 BY foreign
/STATISTICS=CHISQ.

```

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
rep78 * foreign	69	93.2%	5	6.8%	74	100.0%

rep78 * foreign Crosstabulation

Count		foreign		Total
		0	1	
rep78	1	2	0	2
	2	8	0	8
	3	27	3	30
	4	9	9	18
	5	2	9	11
Total		48	21	69

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	27.264 ^a	4	.000
Likelihood Ratio	29.912	4	.000
Linear-by-Linear Association	23.851	1	.000
N of Valid Cases	69		

a. 4 cells (40.0%) have expected count less than 5. The minimum expected count is .61.

Correlation الترابط

CORRELATIONS

/VARIABLES=price mpg weight.

Correlations

		price	mpg	weight
price	Pearson Correlation	1	-.478	.539
	Sig. (2-tailed)		.000	.000
	N	74	71	74
mpg	Pearson Correlation	-.478	1	-.807
	Sig. (2-tailed)	.000		.000
	N	71	71	71
weight	Pearson Correlation	.539	-.807	1
	Sig. (2-tailed)	.000	.000	
	N	74	71	74

CORR

/VARIABLES=price mpg weight
/MISSING=LISTWISE.

Correlations^a

		price	mpg	weight
price	Pearson Correlation	1	-.478	.542
	Sig. (2-tailed)		.000	.000
mpg	Pearson Correlation	-.478	1	-.807
	Sig. (2-tailed)	.000		.000
weight	Pearson Correlation	.542	-.807	1
	Sig. (2-tailed)	.000	.000	

a. Listwise N=71

الإنحدار Regression

REG

/DEPENDENT price

/METHOD=ENTER mpg weight.

Variables Entered/Removed^b

Model	Variables Entered	Variables Removed	Method
1	weight, mpg	.	Enter

a. All requested variables entered.

b. Dependent Variable: price

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.546 ^a	.298	.278	2535.160

a. Predictors: (Constant), weight, mpg

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1.9E+008	2	92835327.81	14.444	.000 ^a
	Residual	4.4E+008	68	6427037.719		
	Total	6.2E+008	70			

a. Predictors: (Constant), weight, mpg

b. Dependent Variable: price

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2394.285	3647.875		.656	.514
	mpg	-58.669	87.294	-.116	-.672	.504
	weight	1.690	.649	.448	2.603	.011

a. Dependent Variable: price

تحليل التباين (والتغاير) Analysis of variance (and analysis of covariance)

ANOVA

/VARIABLES=mpg BY rep78(3,5)

/METHOD=EXPERIM

/STATISTICS MEAN.

Case Processing Summary^a

Cases					
Included		Excluded		Total	
N	Percent	N	Percent	N	Percent
57	77.0%	17	23.0%	74	100.0%

a. mpg by rep78

Cell Means^b

rep78	mpg	
	Mean	N
3	19.43	28
4	21.67	18
5	27.36	11
Total	21.67 ^a	57

a. Grand Mean

b. mpg by rep78

ANOVA^a

			Experimental Method				
			Sum of Squares	df	Mean Square	F	Sig.
mpg	Main Effects	rep78	497.264	2	248.632	8.081	.001
	Model		497.264	2	248.632	8.081	.001
	Residual		1661.403	54	30.767		
	Total		2158.667	56	38.548		

a. mpg by rep78

تحليل التباين باستخدام الأمر glm
Analysis of variance with the glm command

```
COMPUTE filt345=(ANY(rep78 ,3,4,5)).
FILTER BY filt345.
EXECUTE.
```

```
GLM mpg BY rep78
/POSTHOC = rep78 ( TUKEY )
/EMMEANS = TABLES(rep78).
```

```
FILTER OFF.
EXECUTE.
```

Between-Subjects Factors

	N
rep78 3	28
4	18
5	11

Tests of Between-Subjects Effects

Dependent Variable: mpg

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	497.264 ^a	2	248.632	8.081	.001
Intercept	25725.354	1	25725.354	836.142	.000
rep78	497.264	2	248.632	8.081	.001
Error	1661.403	54	30.767		
Total	28917.000	57			
Corrected Total	2158.667	56			

a. R Squared = .230 (Adjusted R Squared = .202)

rep78

Dependent Variable: mpg

rep78	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
3	19.429	1.048	17.327	21.530
4	21.667	1.307	19.046	24.288
5	27.364	1.672	24.011	30.717

Multiple Comparisons

Dependent Variable: mpg

Tukey HSD

(I) rep78	(J) rep78	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
3	4	-2.24	1.676	.382	-6.28	1.80
	5	-7.94*	1.974	.001	-12.69	-3.18
4	3	2.24	1.676	.382	-1.80	6.28
	5	-5.70*	2.123	.026	-10.81	-.58
5	3	7.94*	1.974	.001	3.18	12.69
	4	5.70*	2.123	.026	.58	10.81

Based on observed means.

*. The mean difference is significant at the .050 level.

mpgTukey HSD^{a,b}

rep78	N	Subset	
		1	2
3	28	19.43	27.36
4	18	21.67	
5	11		
Sig.		.483	1.000

Means for groups in homogeneous subsets are displayed.

Based on Type III Sum of Squares

The error term is Mean Square(Error) = 30.767.

a. Uses Harmonic Mean Sample Size = 16.467.

b. Alpha = .050.

Further SPSS Examples

```
* SPSS Program for Basic Statistical Tests.
* Filename: Statistics.sps.

* CD 'C:\myRfolder'.
GET FILE='K:\Statistical Packages\spss\SPSS 19 Portable\mydata100.sav'.
DATASET NAME DataSet2 WINDOW=FRONT.

* Descriptive stats in compact form.
DESCRIPTIVES VARIABLES=q1 to posttest
  /STATISTICS=MEAN STDDEV VARIANCE
  MIN MAX SEMEAN.

* Descriptive stats of every sort.
EXAMINE VARIABLES=q1 to posttest
  /PLOT BOXPLOT STEMLEAF NPLOT
  /COMPARE GROUP
  /STATISTICS DESCRIPTIVES EXTREME
  /MISSING PAIRWISE.

* Frequencies and percents.
FREQUENCIES VARIABLES=workshop TO q4.

* Chi-squared.
CROSSTABS
  /TABLES=workshop BY gender
  /FORMAT= AVALUE TABLES
  /STATISTIC=CHISQ
  /CELLS= COUNT ROW
  /COUNT ROUND CELL .

* ---Measures of association---.

* Person correlations.
CORRELATIONS
  /VARIABLES=q1 TO q4.

* Spearman correlations.
NONPAR CORR
  /VARIABLES=q1 to q4
  /PRINT=SPEARMAN.

* Linear regression.
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT q4
  /METHOD=ENTER q1 q2 q3.
```

```

REGRESSION
  /DEPENDENT q4
  /METHOD=ENTER q1 q2 q3.

* ---Group comparisons---.

* Independent samples t-test.
T-TEST
  GROUPS = gender('m' 'f')
  /VARIABLES = q1.

* Nonparametric version of above using
* Wilcoxon/Mann-Whitney test.
NPTESTS
  /INDEPENDENT TEST (posttest)
  GROUP (gender) MANN_WHITNEY.

* Paired samples t-test.
T-TEST
  PAIRS = pretest WITH posttest (PAIRED).

* Nonparametric version of above using
* Wilcoxon Signed-Rank test and Sign test.
NPTESTS
  /RELATED TEST(pretest posttest) SIGN WILCOXON.

* Oneway analysis of variance (ANOVA).
UNIANOVA posttest BY workshop
  /POSTHOC = workshop ( TUKEY )
  /PRINT = ETASQ HOMOGENEITY
  /DESIGN = workshop.

* Nonparametric version of above using
  Kruskal Wallis test.
NPAR TESTS
  /K-W=posttest BY workshop(1 3).

NPAR TESTS
  /CHISQUARE=happy life
  /EXPECTED=EQUAL
  /MISSING ANALYSIS.

```

Chi-Square Test

Frequencies

General Happiness

	Observed N	Expected N	Residual
Very Happy	467	501.3	-34.3
Pretty Happy	872	501.3	370.7
Not Too Happy	165	501.3	-336.3
Total	1504		

Is Life Exciting or Dull			
	Observed N	Expected N	Residual
Exciting	434	326.7	107.3
Routine	505	326.7	178.3
Dull	41	326.7	-285.7
Total	980		

Test Statistics		
	General Happiness	Is Life Exciting or Dull
Chi-Square	502.047 ^a	382.435 ^b
df	2	2
Asymp. Sig.	.000	.000

a. 0 cells (0.0%) have expected frequencies less than 5. The minimum expected cell frequency is 501.3.

b. 0 cells (0.0%) have expected frequencies less than 5. The minimum expected cell frequency is 326.7.

```
GET DATA
/TYPE=TXT
/FILE='http://www.ats.ucla.edu/stat/data/hsb2.txt'.
```

```
GET FILE='K:\Statistical Packages\spss\SPSS 19 Portable\mydata100.sav'.
DATASET NAME DataSet2 WINDOW=FRONT.
```

```
* Descriptive stats in compact form.
DESCRIPTIVES VARIABLES=q1 to posttest
/STATISTICS=MEAN STDDEV VARIANCE
MIN MAX SEMEAN.
```


Descriptive Statistics

	N		Minimum		Maximum	
	Statistic	Std. Error	Statistic	Std. Error	Statistic	Std. Error
The instructor was well prepared.	100		1		5	
The instructor communicated well.	100		1		5	
The course materials were helpful.	99		1		5	
Overall, I found this workshop useful.	100		1		5	
pretest	100		58		86	
posttest	100		59		98	
Valid N (listwise)	99					

Descriptive Statistics

	Mean		Std. Deviation		Variance	
	Statistic	Std. Error	Statistic	Std. Error	Statistic	Std. Error
The instructor was well prepared.	3.45	.110	1.095		1.199	
The instructor communicated well.	3.06	.122	1.221		1.491	
The course materials were helpful.	3.08	.117	1.167		1.361	
Overall, I found this workshop useful.	3.40	.114	1.137		1.293	
pretest	74.97	.530	5.296		28.050	
posttest	82.06	.659	6.590		43.431	
Valid N (listwise)						

* Descriptive stats of every sort.
 EXAMINE VARIABLES=q1 to posttest
 /PLOT BOXPLOT STEMLEAF NPLOT
 /COMPARE GROUP
 /STATISTICS DESCRIPTIVES EXTREME
 /MISSING PAIRWISE.

Explore

Total Sample

Case Processing Summary

Cases		
Valid	Missing	Total

	N	Percent	N	Percent	N	Percent
The instructor was well prepared.	100	100.0%	0	0.0%	100	100.0%
The instructor communicated well.	100	100.0%	0	0.0%	100	100.0%
The course materials were helpful.	99	99.0%	1	1.0%	100	100.0%
Overall, I found this workshop useful.	100	100.0%	0	0.0%	100	100.0%
pretest	100	100.0%	0	0.0%	100	100.0%
posttest	100	100.0%	0	0.0%	100	100.0%

Descriptives

			Statistic	Std. Error
The instructor was well prepared.	Mean		3.45	.110
	95% Confidence Interval for Mean	Lower Bound	3.23	
		Upper Bound	3.67	
	5% Trimmed Mean		3.49	
	Median		3.00	
	Variance		1.199	
	Std. Deviation		1.095	
	Minimum		1	
	Maximum		5	
	Range		4	
	Interquartile Range		1	
	Skewness		-.176	.241
	Kurtosis		-.624	.478
The instructor communicated well.	Mean		3.06	.122
	95% Confidence Interval for Mean	Lower Bound	2.82	
		Upper Bound	3.30	
	5% Trimmed Mean		3.07	
	Median		3.00	
	Variance		1.491	

The course materials were helpful.	Std. Deviation		1.221	
	Minimum		1	
	Maximum		5	
	Range		4	
	Interquartile Range		2	
	Skewness		-.015	.241
	Kurtosis		-1.059	.478
	Mean		3.08	.117
	95% Confidence Interval for Mean	Lower Bound	2.85	
		Upper Bound	3.31	
	5% Trimmed Mean		3.09	
	Median		3.00	
	Variance		1.361	
	Std. Deviation		1.167	
	Minimum		1	
	Maximum		5	
	Range		4	
	Interquartile Range		2	

Descriptives

			Statistic	Std. Error
The course materials were helpful.	Skewness		-.042	.243
	Kurtosis		-.711	.481
Overall, I found this workshop useful.	Mean		3.40	.114
	95% Confidence Interval for Mean	Lower Bound	3.17	
		Upper Bound	3.63	
	5% Trimmed Mean		3.44	
	Median		3.00	
	Variance		1.293	
	Std. Deviation		1.137	
	Minimum		1	
	Maximum		5	
	Range		4	
	Interquartile Range		1	
	Skewness		-.252	.241
	Kurtosis		-.601	.478

pretest	Mean		74.97	.530
	95% Confidence Interval for Mean	Lower		
		Bound	73.92	
		Upper		
		Bound	76.02	
	5% Trimmed Mean		75.13	
	Median		75.00	
	Variance		28.050	
	Std. Deviation		5.296	
	Minimum		58	
	Maximum		86	
	Range		28	
	Interquartile Range		7	
	Skewness		-.363	.241
	Kurtosis		.461	.478
posttest	Mean		82.06	.659
	95% Confidence Interval for Mean	Lower		
		Bound	80.75	
		Upper		
		Bound	83.37	
	5% Trimmed Mean		82.18	
	Median		82.00	
	Variance		43.431	
	Std. Deviation		6.590	
	Maximum		98	

Descriptives

		Statistic	Std. Error
posttest	Range	39	
	Interquartile Range	9	
	Skewness	-.290	.241
	Kurtosis	.842	.478

Extreme Values

		Case Number	Value
The instructor was well	Highest	1	4 5

prepared.		2	6	5
		3	10	5
		4	14	5
		5	16	5 ^a
	Lowest	1	69	1
		2	62	1
		3	58	1
		4	22	1
		5	100	2 ^b
The instructor communicated well.	Highest	1	10	5
		2	20	5
		3	31	5
		4	40	5
		5	44	5 ^a
	Lowest	1	88	1
		2	86	1
		3	75	1
		4	68	1
		5	64	1 ^c
The course materials were helpful.	Highest	1	4	5
		2	10	5
		3	16	5
		4	18	5
		5	34	5 ^a
	Lowest	1	98	1
		2	94	1
		3	86	1
		4	75	1
		5	69	1 ^c
Overall, I found this workshop useful.	Highest	1	1	5
		2	6	5
		3	8	5
		4	10	5
		5	16	5 ^a
	Lowest	1	100	1
		2	68	1

Extreme Values

	Case Number	Value
--	----------------	-------

Overall, I found this workshop useful.	Lowest	3	58	1
		4	54	1
		5	22	1
	Highest	1	70	86
		2	60	85 ^a
		3	62	85
		4	57	84
		5	8	83
	Lowest	1	92	58
		2	82	62 ^b
		3	51	63
		4	48	63
		5	81	66
pretest	Highest	1	91	98
		2	86	96 ^a
		3	60	95
		4	62	95
		5	8	92
	Lowest	1	92	59
		2	48	67 ^c
		3	67	68
		4	81	69
		5	82	71

a. Only a partial list of cases with the value 5 are shown in the table of upper extremes.

b. Only a partial list of cases with the value 2 are shown in the table of lower extremes.

c. Only a partial list of cases with the value 1 are shown in the table of lower extremes.

d. Only a partial list of cases with the value 83 are shown in the table of upper extremes.

e. Only a partial list of cases with the value 92 are shown in the table of upper extremes.

Tests of Normality

Kolmogorov-Smirnov ^a			Shapiro-Wilk		
Statistic	df	Sig.	Statistic	df	Sig.

The instructor was well prepared.	.199	100	.000	.900	100	.000
The instructor communicated well.	.189	100	.000	.906	100	.000
The course materials were helpful.	.174	99	.000	.916	99	.000
Overall, I found this workshop useful.	.177	100	.000	.904	100	.000
pretest	.087	100	.061	.978	100	.101
posttest	.072	100	.200 [*]	.984	100	.275

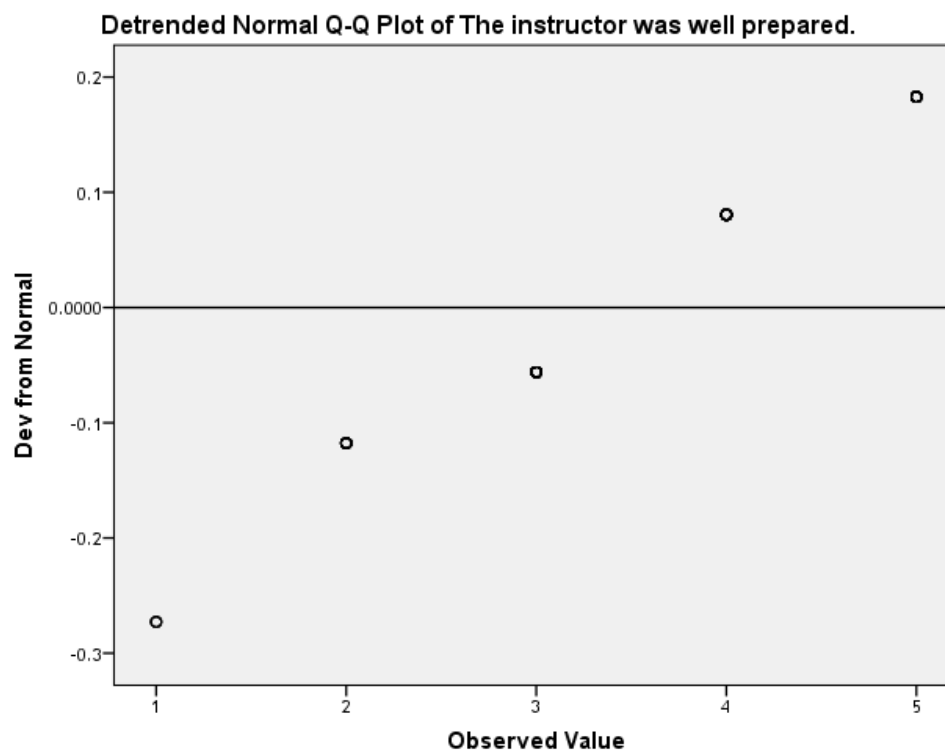
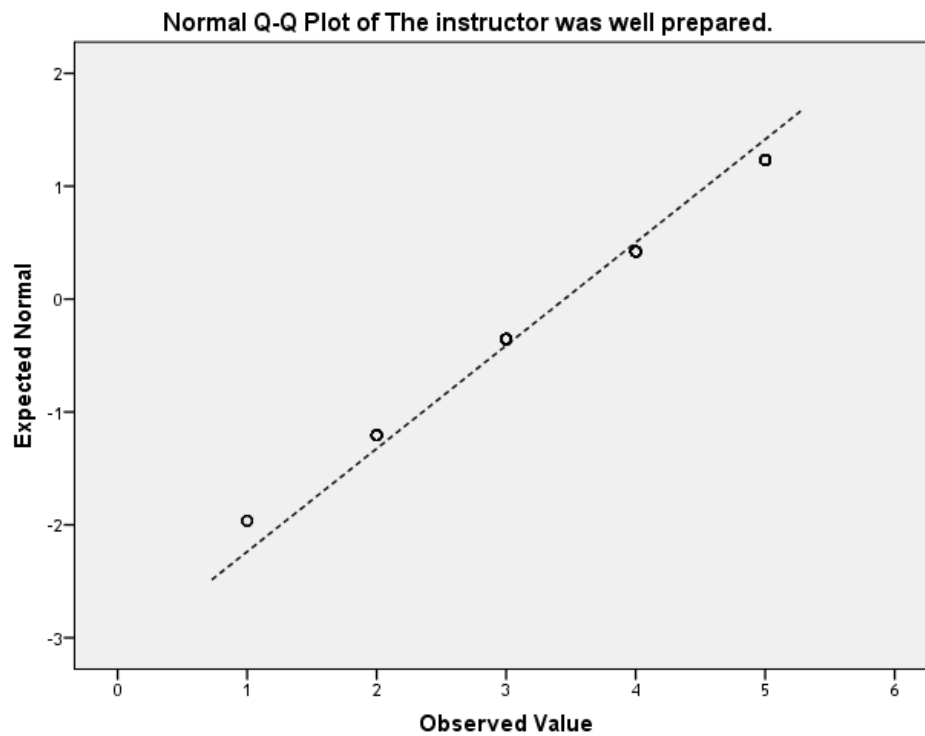
*. This is a lower bound of the true significance.

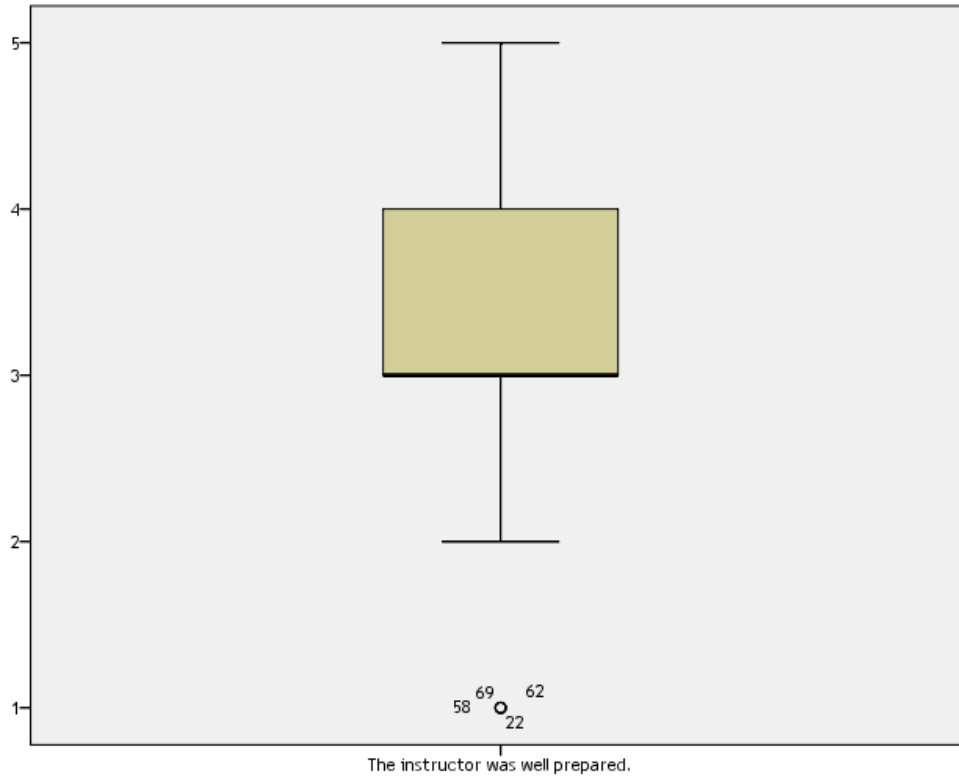
a. Lilliefors Significance Correction

The instructor was well prepared.

The instructor was well prepared. Stem-and-Leaf Plot

[illegible]



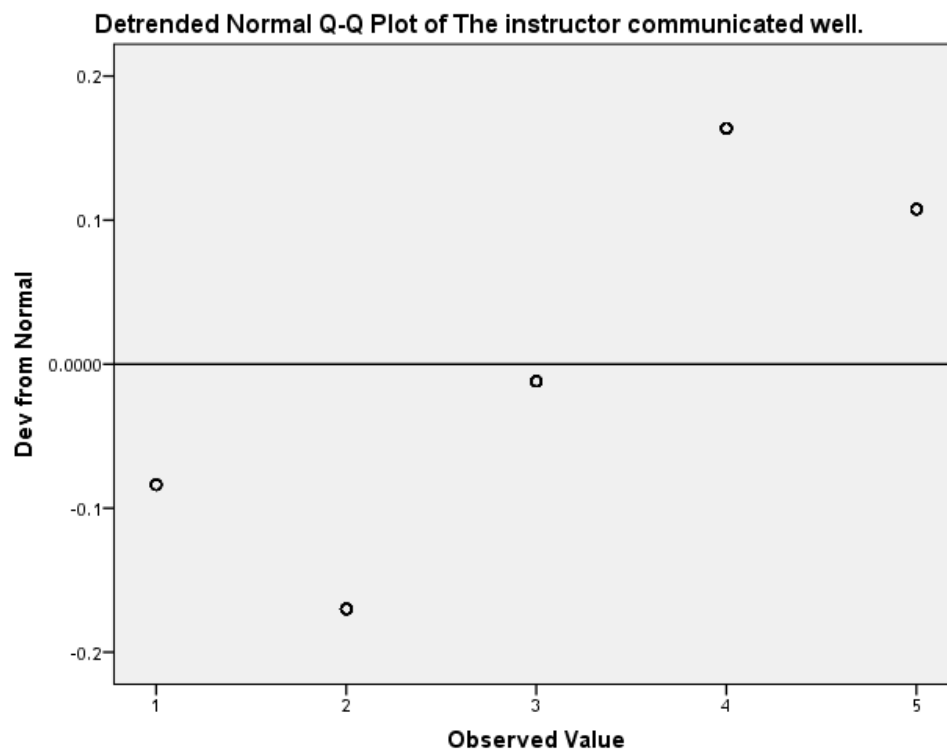
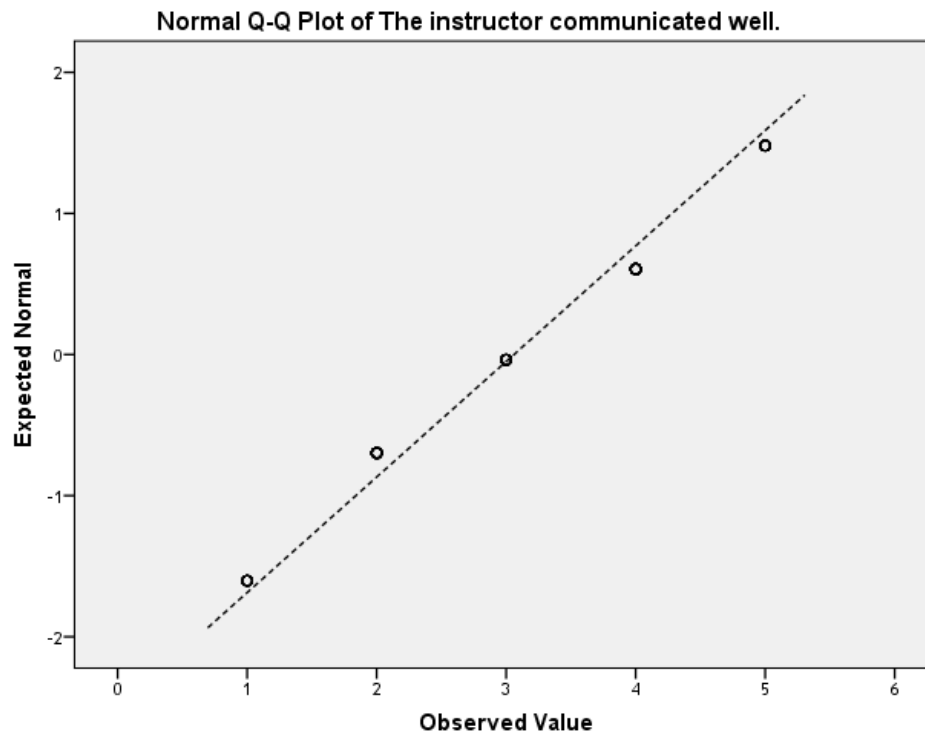


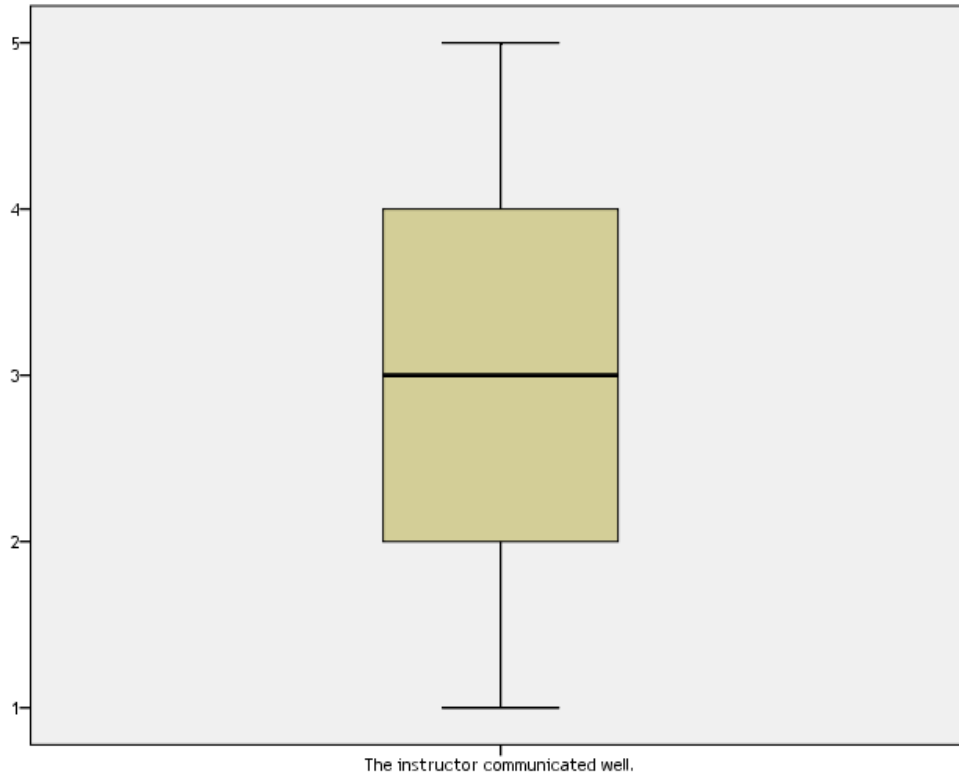
The instructor communicated well.

The instructor communicated well. Stem-and-Leaf Plot

Frequency	Stem &	Leaf
10.00	1 .	0000000000
.00	1 .	
28.00	2 .	000000000000000000000000000000
.00	2 .	
21.00	3 .	000000000000000000000000
.00	3 .	
28.00	4 .	000000000000000000000000000000
.00	4 .	
13.00	5 .	00000000000000

Stem width: 1
Each leaf: 1 case(s)



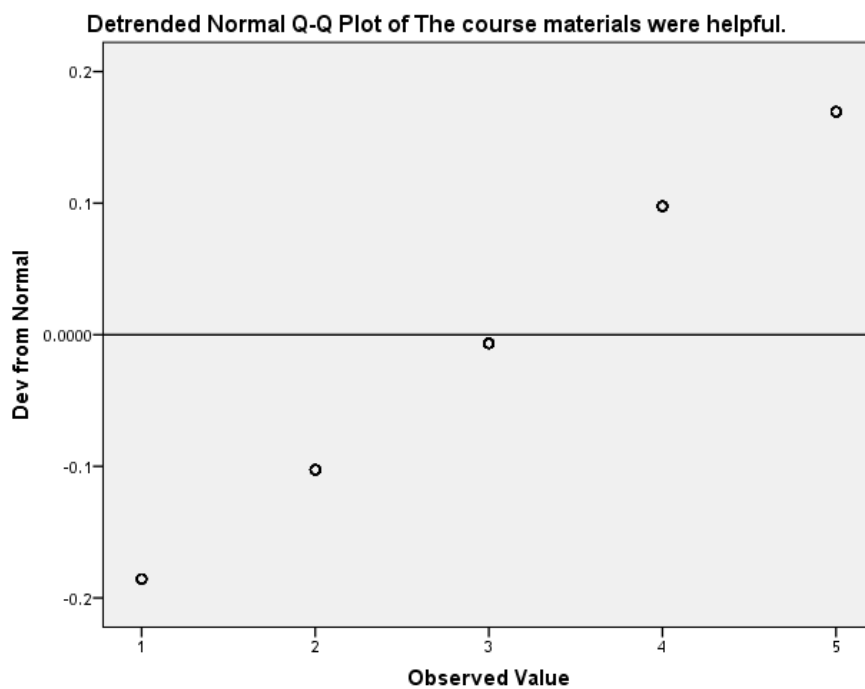
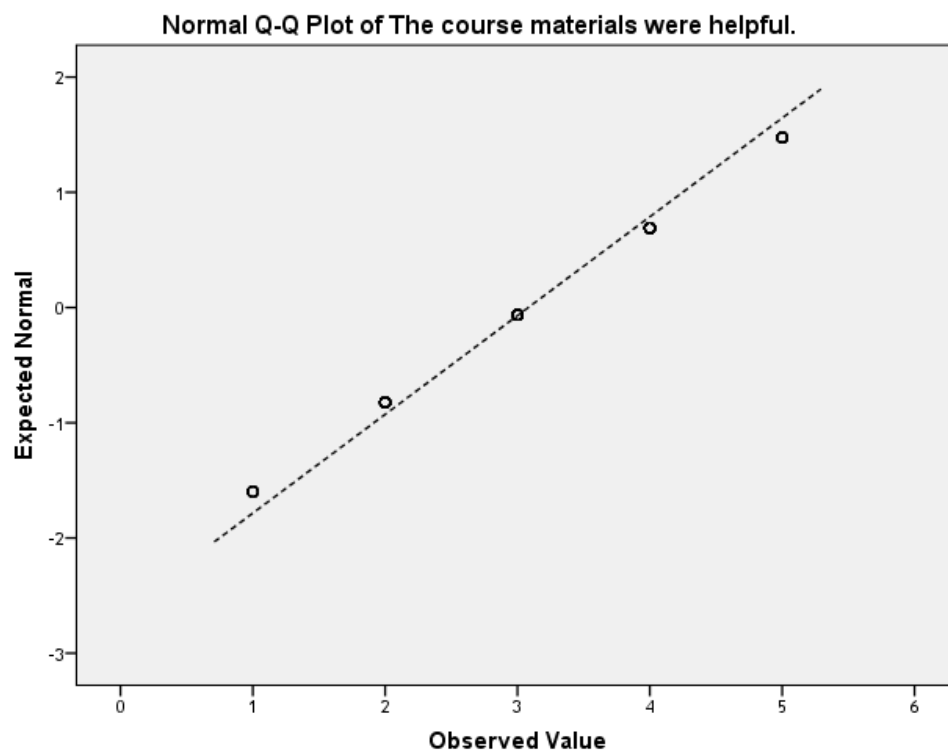


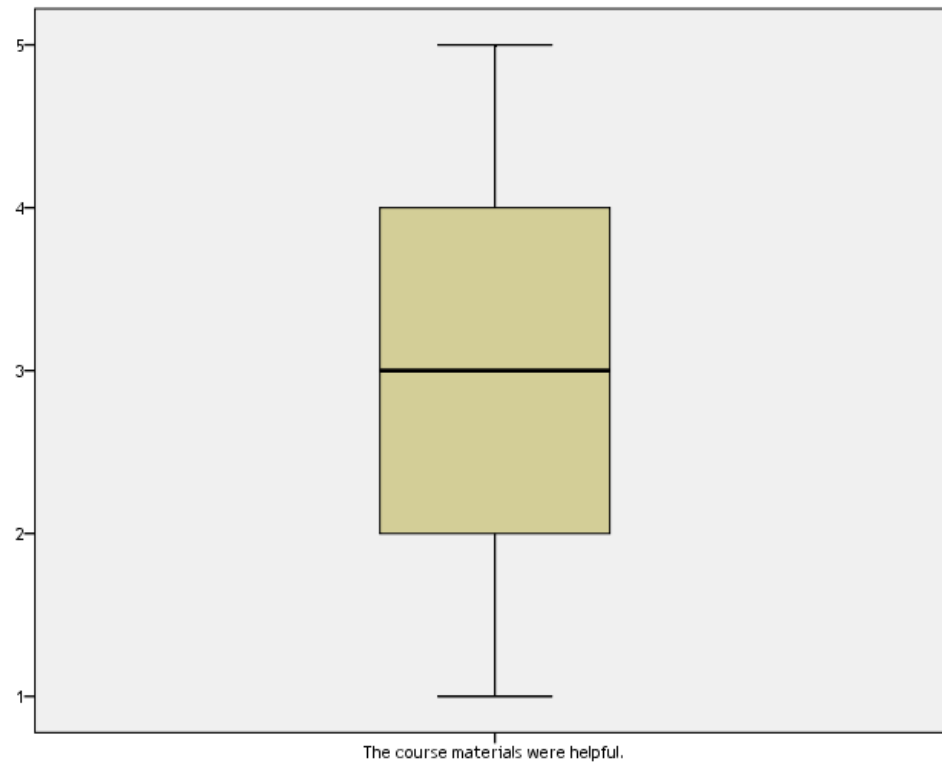
The course materials were helpful.

The course materials were helpful. Stem-and-Leaf Plot

Frequency	Stem &	Leaf
10.00	1 .	0000000000
.00	1 .	
20.00	2 .	00000000000000000000
.00	2 .	
34.00	3 .	00000000000000000000000000000000
.00	3 .	
22.00	4 .	0000000000000000000000
.00	4 .	
13.00	5 .	00000000000000

Stem width: 1
Each leaf: 1 case(s)



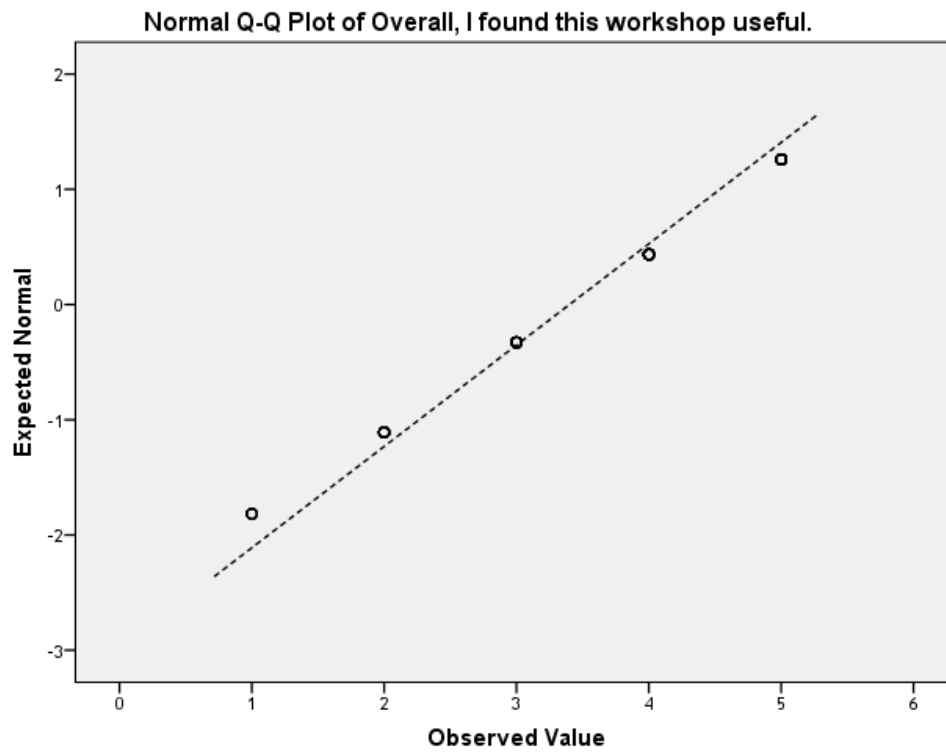


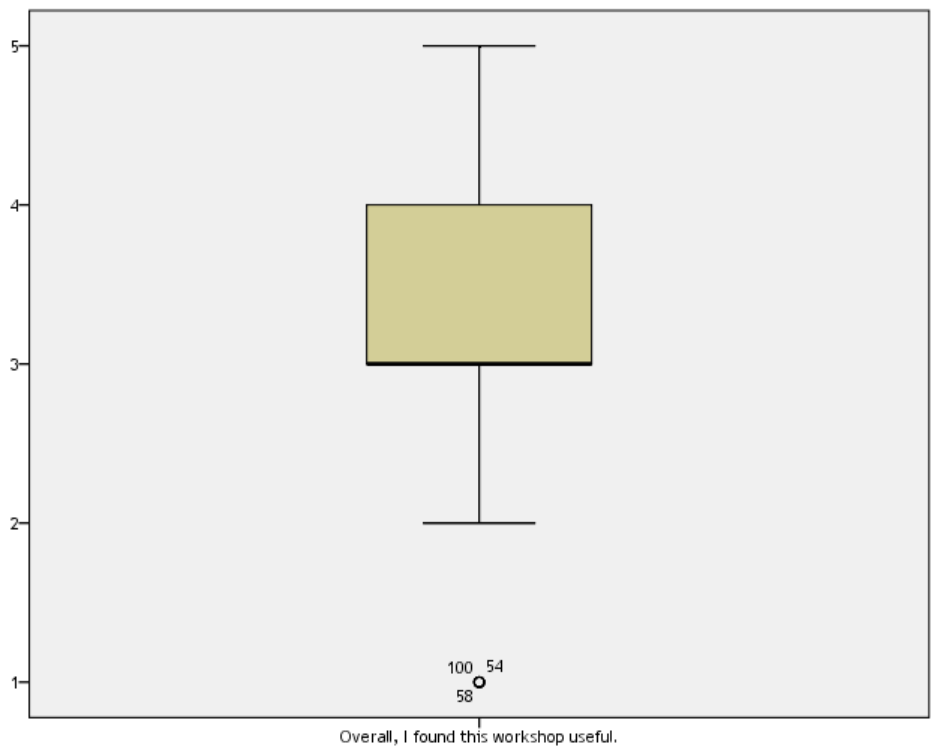
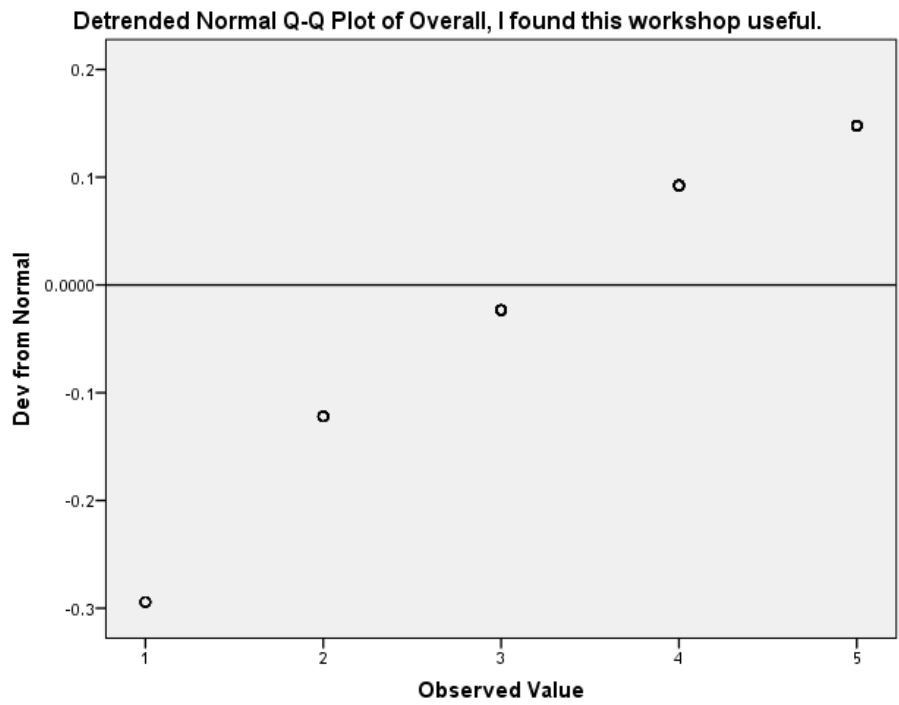
Overall, I found this workshop useful.

Overall, I found this workshop useful. Stem-and-Leaf Plot

[illegible]

Stem width: 1
Each leaf: 1 case(s)



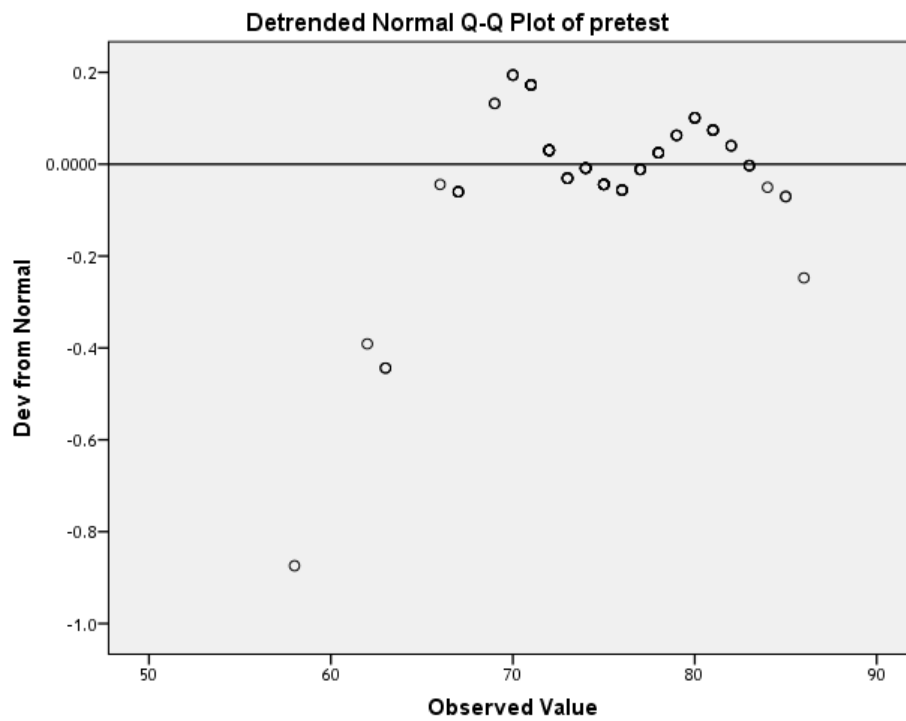
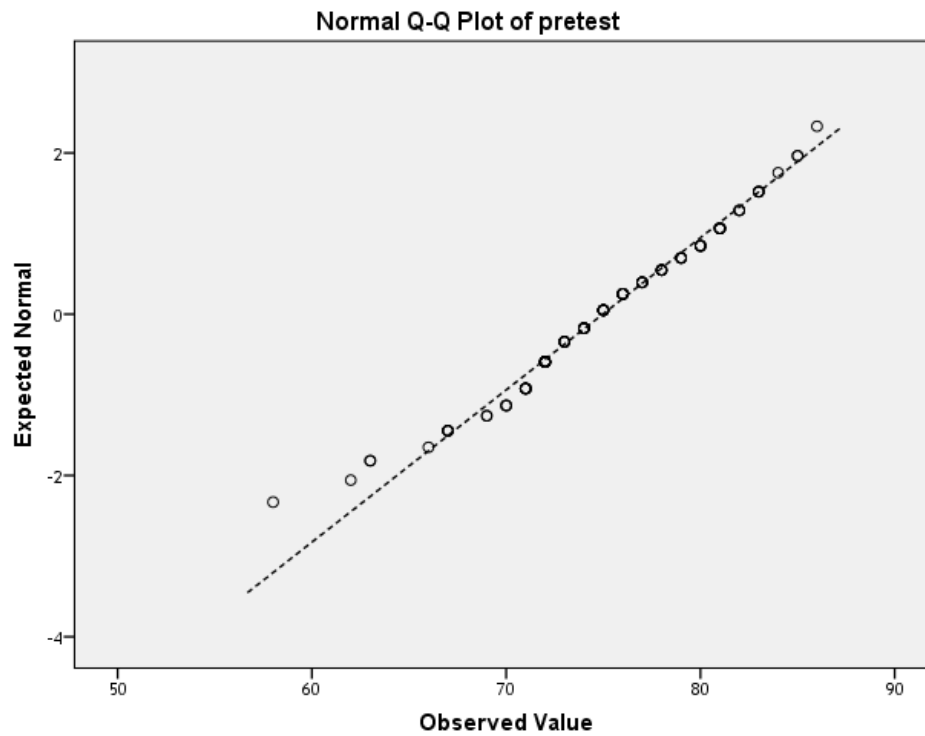


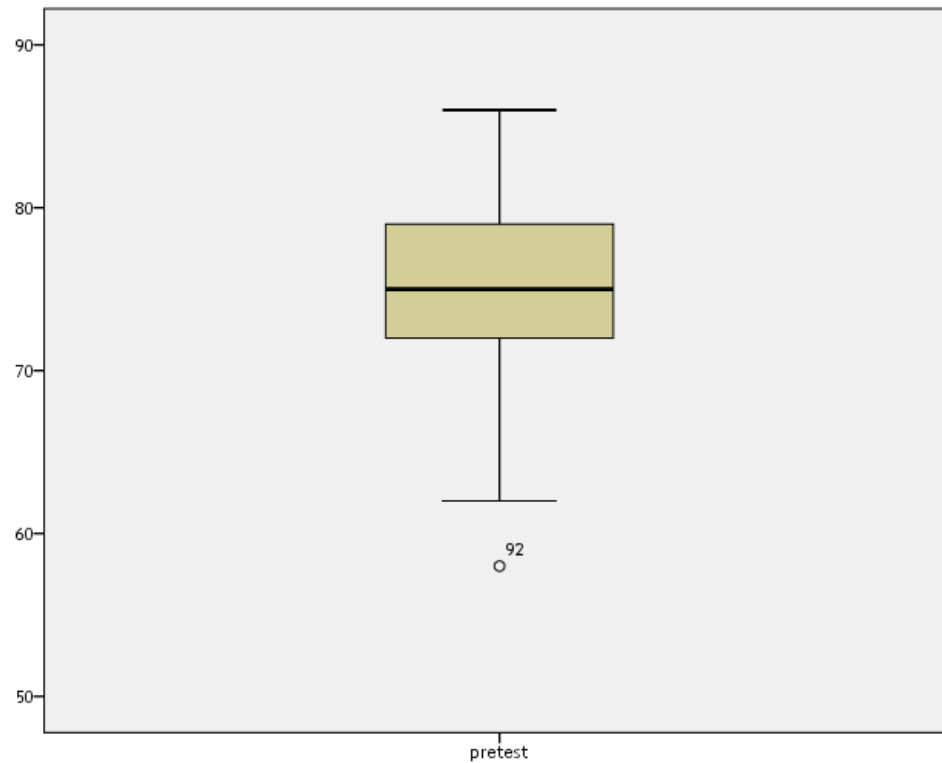
pretest

pretest Stem-and-Leaf Plot

Frequency	Stem &	Leaf
1.00	Extremes	(=<58)
.00	6 .	
3.00	6 .	233
.00	6 .	
5.00	6 .	67777
2.00	6 .	99
10.00	7 .	0001111111
18.00	7 .	22222222222233333
18.00	7 .	4444444445555555555
11.00	7 .	66666677777
10.00	7 .	8888889999
11.00	8 .	00000111111
7.00	8 .	2223333
3.00	8 .	455
1.00	8 .	6

Stem width: 10
Each leaf: 1 case(s)



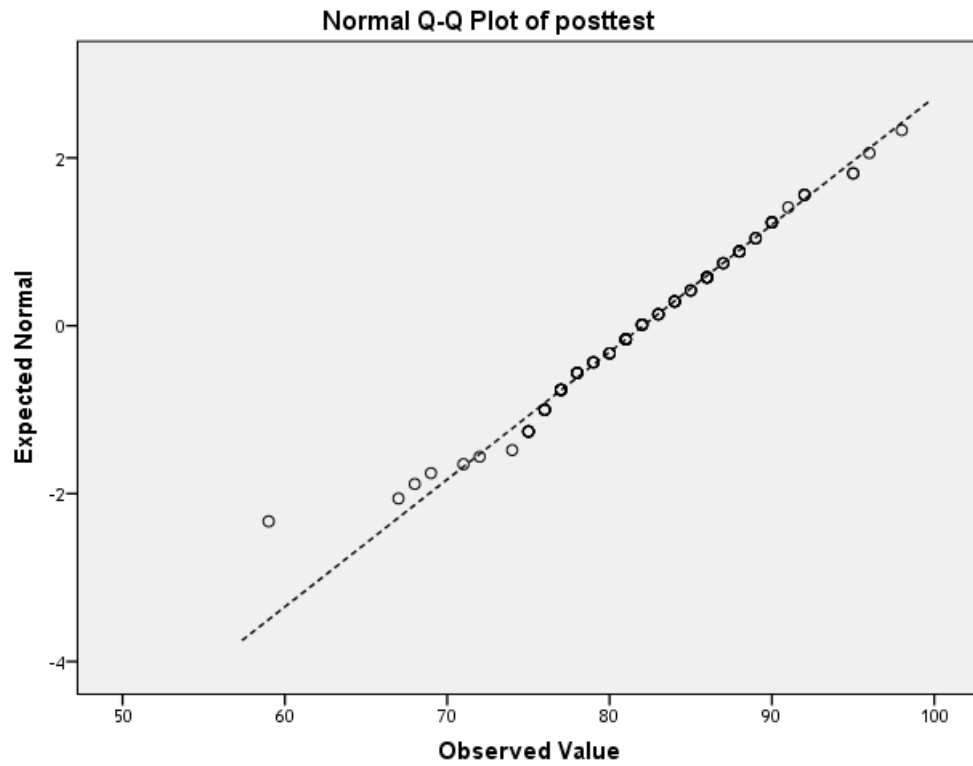


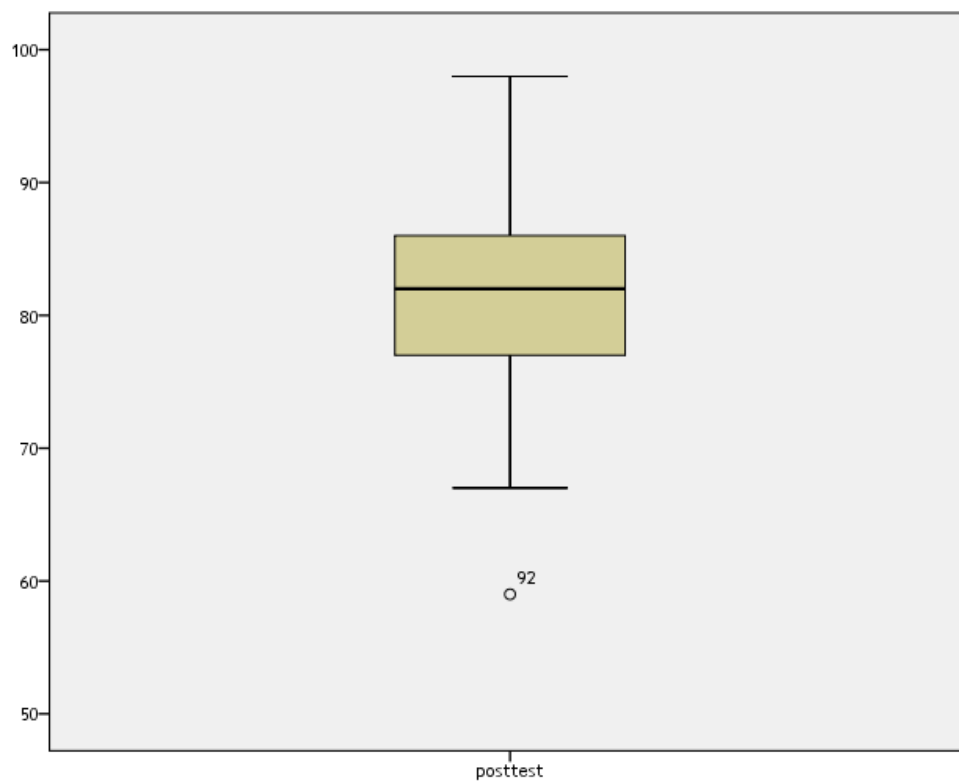
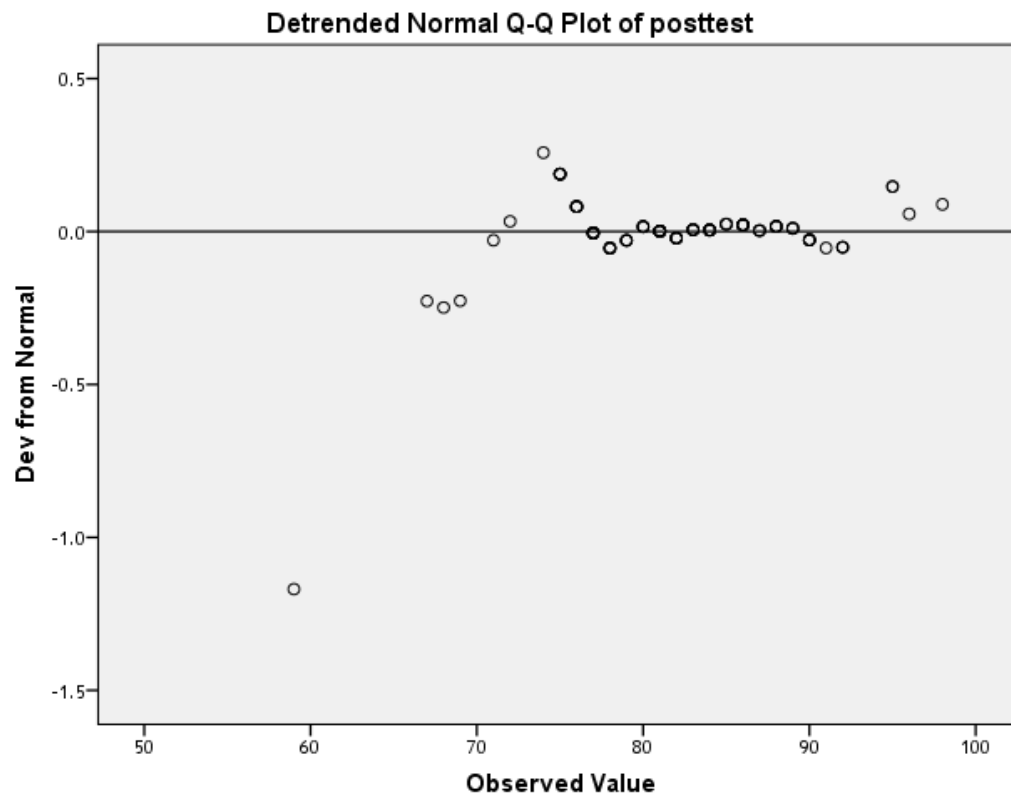
posttest

posttest Stem-and-Leaf Plot

Frequency	Stem &	Leaf
1.00	Extremes	(=<59)
1.00	6 .	7
2.00	6 .	89
1.00	7 .	1
1.00	7 .	2
7.00	7 .	4555555
13.00	7 .	6666677777777
9.00	7 .	888889999
13.00	8 .	0000111111111
10.00	8 .	2222233333
10.00	8 .	4444444555
11.00	8 .	66666666777
8.00	8 .	88888999
6.00	9 .	000001
3.00	9 .	222
2.00	9 .	55
1.00	9 .	6

1.00 9 . 8
Stem width: 10
Each leaf: 1 case(s)





* Frequencies and percents.
 FREQUENCIES VARIABLES=workshop TO q4.

Frequencies

		Statistics				
		workshop	The instructor was well prepared.	The instructor communicated well.	The course materials were helpful.	Overall, I found this workshop useful.
N	Valid	100	100	100	99	100
	Missing	0	0	0	1	0

Frequency Table

		workshop			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	NA	1	1.0	1.0	1.0
	R	31	31.0	31.0	32.0
	SAS	24	24.0	24.0	56.0
	SPSS	25	25.0	25.0	81.0
	Stata	19	19.0	19.0	100.0
	Total	100	100.0	100.0	

		The instructor was well prepared.			
		Frequency	Percent	Valid Percent	Cumulative Percent

Valid	1	4	4.0	4.0	4.0
	2	14	14.0	14.0	18.0
	3	36	36.0	36.0	54.0
	4	25	25.0	25.0	79.0
	5	21	21.0	21.0	100.0
	Total	100	100.0	100.0	

The instructor communicated well.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	10	10.0	10.0	10.0
	2	28	28.0	28.0	38.0
	3	21	21.0	21.0	59.0
	4	28	28.0	28.0	87.0
	5	13	13.0	13.0	100.0
	Total	100	100.0	100.0	

The course materials were helpful.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	10	10.0	10.1	10.1
	2	20	20.0	20.2	30.3
	3	34	34.0	34.3	64.6
	4	22	22.0	22.2	86.9
	5	13	13.0	13.1	100.0
	Total	99	99.0	100.0	
Missing	System	1	1.0		
Total		100	100.0		

Overall, I found this workshop useful.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	6	6.0	6.0	6.0

2	14	14.0	14.0	20.0
3	34	34.0	34.0	54.0
4	26	26.0	26.0	80.0
5	20	20.0	20.0	100.0
Total	100	100.0	100.0	

* Chi-squared.

CROSSTABS

/TABLES=workshop BY gender

/FORMAT= AVALUE TABLES

/STATISTIC=CHISQ

/CELLS= COUNT ROW

/COUNT ROUND CELL .

Crosstabs

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
workshop * gender	100	100.0%	0	0.0%	100	100.0%

workshop * gender Crosstabulation

			gender			Total
			Female	Male	NA	
workshop	NA	Count	0	0	1	1
		% within workshop	0.0%	0.0%	100.0%	100.0%
	R	Count	14	17	0	31
		% within workshop	45.2%	54.8%	0.0%	100.0%
	SAS	Count	11	13	0	24
		% within workshop	45.8%	54.2%	0.0%	100.0%
	SPSS	Count	13	12	0	25

	% within workshop	52.0%	48.0%	0.0%	100.0%
Stata	Count	9	10	0	19
	% within workshop	47.4%	52.6%	0.0%	100.0%
Total	Count	47	52	1	100
	% within workshop	47.0%	52.0%	1.0%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	100.301 ^a	8	.000
Continuity Correction			
Likelihood Ratio	11.498	8	.175
Linear-by-Linear Association			
N of Valid Cases	100		

a. 7 cells (46.7%) have expected count less than 5. The minimum expected count is .01.

* ---Measures of association---.

* Person correlations.

CORRELATIONS

/VARIABLES=q1 TO q4.

Correlations

Correlations

		The instructor was well prepared.	The instructor communicated well.	The course materials were helpful.
The instructor was well prepared.	Pearson Correlation	1	.667	.695
	Sig. (2-tailed)		.000	.000

	N	100	100	99
The instructor communicated well.	Pearson			
	Correlation	.667	1	.604
	Sig. (2-tailed)	.000		.000
	N	100	100	99
The course materials were helpful.	Pearson			
	Correlation	.695	.604	1
	Sig. (2-tailed)	.000	.000	
	N	99	99	99
Overall, I found this workshop useful.	Pearson			
	Correlation	.576	.492	.473
	Sig. (2-tailed)	.000	.000	.000
	N	100	100	99

Correlations

		Overall, I found this workshop useful.
The instructor was well prepared.	Pearson Correlation	.576
	Sig. (2-tailed)	.000
	N	100
The instructor communicated well.	Pearson Correlation	.492
	Sig. (2-tailed)	.000
	N	100
The course materials were helpful.	Pearson Correlation	.473
	Sig. (2-tailed)	.000
	N	99
Overall, I found this workshop useful.	Pearson Correlation	1
	Sig. (2-tailed)	
	N	100

* Spearman correlations.
 NONPAR CORR
 /VARIABLES=q1 to q4
 /PRINT=SPEARMAN.

Nonparametric Correlations

Correlations

			The instructor was well prepared.	The instructor communicated well.
Spearman's rho	The instructor was well prepared.	Correlation		
		Coefficient	1.000	.668
		Sig. (2-tailed)	.	.000
		N	100	100
	The instructor communicated well.	Correlation		
		Coefficient	.668	1.000
		Sig. (2-tailed)	.000	.
		N	100	100
	The course materials were helpful.	Correlation		
		Coefficient	.668	.610
		Sig. (2-tailed)	.000	.000
		N	99	99
	Overall, I found this workshop useful.	Correlation		
		Coefficient	.552	.491
		Sig. (2-tailed)	.000	.000
		N	100	100

Correlations

			The course materials were helpful.	Overall, I found this workshop useful.
Spearman's rho	The instructor was well prepared.	Correlation		
		Coefficient	.668	.552
		Sig. (2-tailed)	.000	.000
		N	99	100
	The instructor communicated well.	Correlation		
		Coefficient	.610	.491
		Sig. (2-tailed)	.000	.000
		N	99	100
	The course materials were helpful.	Correlation		
		Coefficient	1.000	.427
		Sig. (2-tailed)	.	.000
		N	99	99

Overall, I found this workshop useful.	Correlation Coefficient	.427	1.000
	Sig. (2-tailed)	.000	.
	N	99	100

* Linear regression.

```
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT q4
  /METHOD=ENTER q1 q2 q3.
```

Regression

Variables Entered/Removed ^a			
Model	Variables Entered	Variables Removed	Method
1	The course materials were helpful., The instructor communicated well., The instructor was well prepared. ^b		. Enter

a. Dependent Variable: Overall, I found this workshop useful.

b. All requested variables entered.

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.597 ^a	.356	.336	.931

a. Predictors: (Constant), The course materials were helpful., The instructor communicated well., The instructor was well prepared.

ANOVA ^a						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	45.527	3	15.176	17.515	.000^b
	Residual	82.312	95	.866		
	Total	127.838	98			

a. Dependent Variable: Overall, I found this workshop useful.

b. Predictors: (Constant), The course materials were helpful., The instructor communicated well., The instructor was well prepared.

Coefficients ^a						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1.209	.318		3.805	.000
	The instructor was well prepared.	.411	.132	.396	3.123	.002
	The instructor communicated well.	.158	.107	.169	1.477	.143
	The course materials were helpful.	.094	.116	.096	.807	.422

Coefficients ^a			
Model		Fraction Missing Info.	Relative Increase Variance
1	(Constant)		
	The instructor was well prepared.		
	The instructor communicated well.		
	The course materials were helpful.		

a. Dependent Variable: Overall, I found this workshop useful.

REGRESSION
 /DEPENDENT q4
 /METHOD=ENTER q1 q2 q3.

Regression

Variables Entered/Removed^a

Model	Variables Entered	Variables Removed	Method
1	The course materials were helpful., The instructor communicated well., The instructor was well prepared. ^b		. Enter

a. Dependent Variable: Overall, I found this workshop useful.

b. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.597^a	.356	.336	.931

a. Predictors: (Constant), The course materials were helpful., The instructor communicated well., The instructor was well prepared.

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	45.527	3	15.176	17.515	.000^b
	Residual	82.312	95	.866		
	Total	127.838	98			

a. Dependent Variable: Overall, I found this workshop useful.

b. Predictors: (Constant), The course materials were helpful., The instructor communicated well., The instructor was well prepared.

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1.209	.318		3.805	.000
	The instructor was well prepared.	.411	.132	.396	3.123	.002
	The instructor communicated well.	.158	.107	.169	1.477	.143
	The course materials were helpful.	.094	.116	.096	.807	.422

Coefficients^a

Model		Fraction Missing Info.	Relative Increase Variance	Relative Efficiency
1	(Constant)			
	The instructor was well prepared.			
	The instructor communicated well.			
	The course materials were helpful.			

a. Dependent Variable: Overall, I found this workshop useful.

* ---Group comparisons---

* Independent samples t-test.

T-TEST

```
GROUPS = gender('m' 'f')
/VARIABLES = q1.
```

T-Test

Warnings

The Independent Samples table is not produced.

Group Statistics

	gender	N	Mean	Std. Deviation	Std. Error Mean
The instructor was well prepared.	m	0 ^a	.	.	.
	f	0 ^a	.	.	.

a. t cannot be computed because at least one of the groups is empty.

- * Nonparametric version of above using
- * Wilcoxon/Mann-Whitney test.

NPTESTS

/INDEPENDENT TEST (posttest)
GROUP (gender) MANN_WHITNEY.

Nonparametric Tests

Hypothesis Test Summary				
	Null Hypothesis	Test	Sig.	Decision
1	The distribution of posttest is the same across categories of gender.	Independent-Samples Mann-Whitney U Test	.	Unable to compute

Asymptotic significances are displayed. The significance level is .05.

- * Paired samples t-test.

T-TEST

PAIRS = pretest WITH posttest (PAIRED).

T-Test

Paired Samples Statistics					
		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	pretest	74.97	100	5.296	.530
	posttest	82.06	100	6.590	.659

Paired Samples Correlations			
	N	Correlation	Sig.

Pair 1	pretest & posttest	100	.680	.000
--------	--------------------	-----	------	------

Paired Samples Test

Paired Differences					
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference
					Lower Upper
Pair 1	pretest - posttest	-7.090	4.903	.490	-8.063 -6.117

Paired Samples Test

		t	df	Sig. (2-tailed)
Pair 1	pretest - posttest	-14.460	99	.000

- * Nonparametric version of above using
- * Wilcoxon Signed-Rank test and Sign test.

NPTESTS

/RELATED TEST(pretest posttest) SIGN WILCOXON.

Nonparametric Tests

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The median of differences between posttest and pretest equals 0.	Related-Samples Sign Test	.000	Reject the null hypothesis.
2	The median of differences between pretest and posttest equals 0.	Related-Samples Wilcoxon Signed Rank Test	.000	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

- * Oneway analysis of variance (ANOVA).

UNIANOVA posttest BY workshop

/POSTHOC = workshop (TUKEY)


```

/PRINT = ETASQ HOMOGENEITY
/DESIGN = workshop.

```

Univariate Analysis of Variance

Warnings

Post hoc tests are not performed for workshop because at least one group has fewer than two cases.

Between-Subjects Factors

	Value Label	N
workshop	NA	1
	R	31
	SAS	24
	SPSS	25
	Stata	19

Levene's Test of Equality of Error Variances^a

Dependent Variable: posttest

F	df1	df2	Sig.
2.431	4	95	.053

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.^a

a. Design: Intercept + workshop

Tests of Between-Subjects Effects

Dependent Variable: posttest

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	892.092 ^a	4	223.023	6.218	.000	.207

Intercept	140294.173	1	140294.173	3911.301	.000	.976
workshop	892.092	4	223.023	6.218	.000	.207
Error	3407.548	95	35.869			
Total	677684.000	100				
Corrected Total	4299.640	99				

a. R Squared = .207 (Adjusted R Squared = .174)

* Nonparametric version of above using
Kruskal Wallis test.

NPARTESTS

/K-W=posttest BY workshop(1 3).

NPar Tests

Warnings

Text: workshop Command: NPARTESTS

A string variable was used in a variable list where only numeric
variables are allowed.

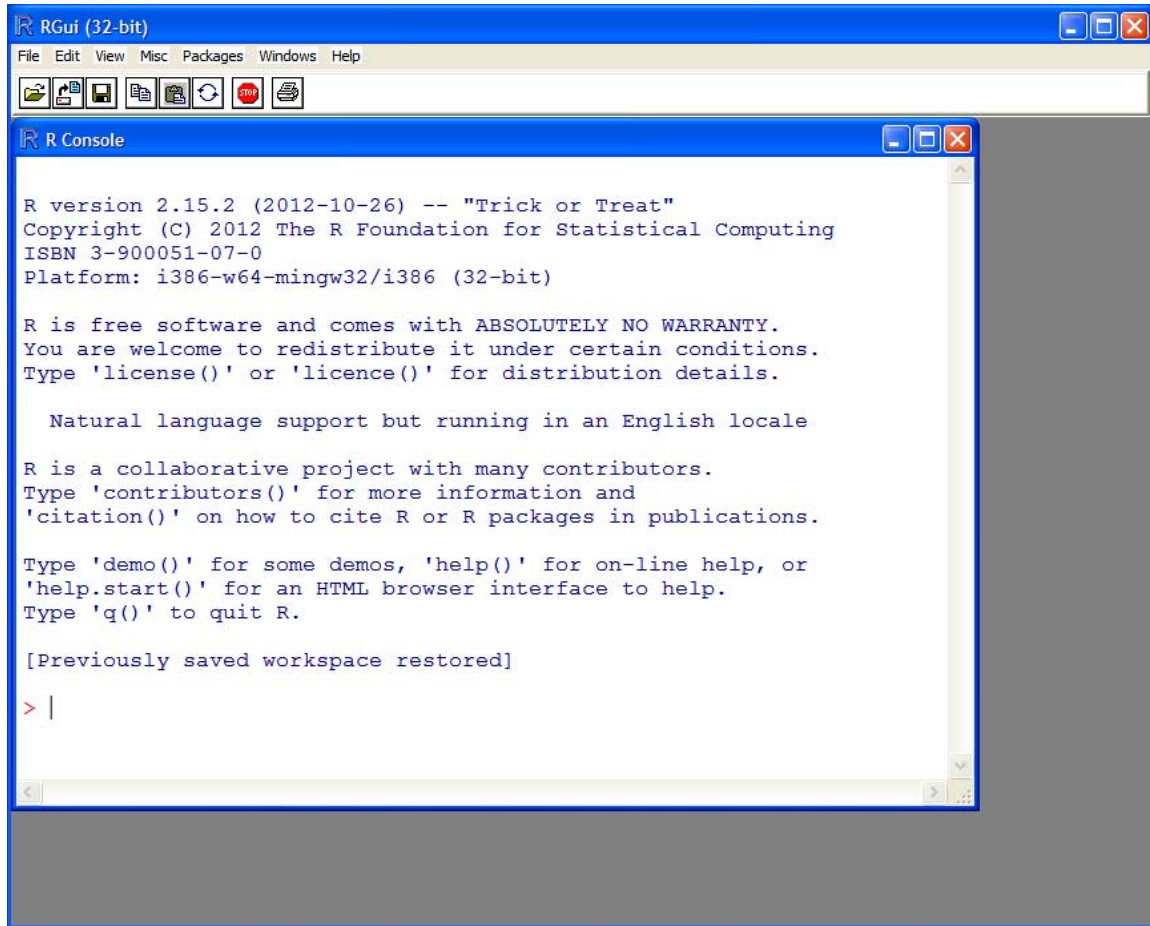
Execution of this command stops.

GET FILE='http://www.ats.ucla.edu/stat/data/hsb2.txt'
/TYPE=TEXT.

Error # 61 in column 10. Text:
http://www.ats.ucla.edu/stat/data/hsb2.txt
The filename is not valid.
Execution of this command stops.

مقدمة للحزمة R

عند تشغيل R تظهر نافذة الأوامر



تدخل الأوامر عند المحفز > .

إدخال بيانات Entering Data

ملفات مجموعة البيانات Dataset files

بيانات مفصولة (أو محددة) بفراغات

```
# tab separated values
dat.tab <- read.table("http://www.ats.ucla.edu/stat/data/hsb2.txt",
  header=TRUE, sep = "\t")
```

بيانات مفصولة (أو محددة) بفواصل

```
# comma separated values
```

```
dat.csv <- read.csv("http://www.ats.ucla.edu/stat/data/hsb2.csv")
```

بيانات من SPSS

```
require(foreign)
# SPSS files
dat.spss <- read.spss("http://www.ats.ucla.edu/stat/data/hsb2.sav",
  to.data.frame=TRUE)
```

بيانات من نوع Excel من الإنترنت

```
# these two steps only needed to read excel files from the internet
f <- tempfile("hsb2", fileext=".xls")

download.file("http://www.ats.ucla.edu/stat/data/hsb2.xls", f,
mode="wb")

dat.xls <- read.xlsx(f, sheetIndex=1)
```

إستعراض البيانات Viewing Data

```
# first few rows
head(dat.csv)
##      id female race ses schtyp prog read write math science socst
## 1  70      0    4  1      1    1  57   52   41      47    57
## 2 121      1    4  2      1    3  68   59   53      63    61
## 3  86      0    4  3      1    1  44   33   54      58    31
## 4 141      0    4  3      1    3  63   44   47      53    56
## 5 172      0    4  2      1    2  47   52   57      53    61
## 6 113      0    4  2      1    2  44   52   51      63    61

# last few rows
tail(dat.csv)
##      id female race ses schtyp prog read write math science socst
## 195 179      1    4  2      2    2  47   65   60      50    56
## 196  31      1    2  2      2    1  55   59   52      42    56
## 197 145      1    4  2      1    3  42   46   38      36    46
## 198 187      1    4  2      2    1  57   41   57      55    52
## 199 118      1    4  2      1    1  55   62   58      58    61
## 200 137      1    4  3      1    2  63   65   65      53    61

# variable names
colnames(dat.csv)
## [1] "id"      "female"  "race"    "ses"      "schtyp"  "prog"    "read"
## [8] "write"   "math"    "science" "socst"
```

```
# pop-up view of entire data set (uncomment to run)
# View(dat.csv)
```

إطرات البيانات Data frames

```
# single cell value
dat.csv[2, 3]
```

```
## [1] 4
# omitting row value implies all rows; here all rows in column 3
dat.csv[, 3]
## [1] 4 4 4 4 4 4 3 1 4 3 4 4 4 4 3 4 4 4 4 4 4 4 3 1 1 3 4 4 4 2 4 4 4 4 4
## [36] 4 4 4 1 4 4 4 4 4 3 4 4 3 4 4 1 2 4 1 4 4 1 4 1 4 1 4 4 4 4 4 4 4 4 4 1
## [71] 4 4 4 4 4 1 4 4 4 1 4 4 4 1 4 4 4 4 4 2 4 4 1 4 4 4 4 1 4 4 4 3 4 4
## [106] 4 4 4 3 4 4 1 4 4 1 4 4 4 4 3 1 4 4 4 3 4 4 2 4 3 4 2 4 4 4 4 4 3 1 3
## [141] 1 4 4 1 4 4 4 4 1 3 3 4 4 1 4 4 4 4 3 4 4 4 4 4 4 4 4 4 4 4 4 1 3 2 3
## [176] 4 4 4 4 4 4 4 4 2 2 4 2 4 3 4 4 4 2 4 2 4 4 4
# omitting column values implies all columns; here all columns in row 2
dat.csv[2, ]
## id female race ses schtyp prog read write math science socst
## 2 121 1 4 2 1 3 68 59 53 63 61
# can also use ranges - rows 2 and 3, columns 2 and 3
dat.csv[2:3, 2:3]
## female race
## 2 1 4
## 3 0 4
```

تأشير المتغيرات More variable indexing

```
# get first 10 rows of variable female using two methods
dat.csv[1:10, "female"]
## [1] 0 1 0 0 0 0 0 0 0 0
dat.csv$female[1:10]
## [1] 0 1 0 0 0 0 0 0 0 0
```

The c function

```
# get column 1 for rows 1, 3 and 5
dat.csv[c(1, 3, 5), 1]
## [1] 70 86 172
# get row 1 values for variables female, prog and socst
dat.csv[1, c("female", "prog", "socst")]
## female prog socst
## 1 0 1 57
```

أسماء المتغيرات Variable Names

```
colnames(dat.csv) <- c("ID", "Sex", "Ethnicity", "SES", "SchoolType",
+"Program", "Reading", "Writing", "Math", "Science", "SocialStudies")
# to change one variable name, just use indexing
colnames(dat.csv)[1] <- "ID2"
```

حفظ البيانات Saving Data

```
write.csv(dat.csv, file = "path/to/save/filename.csv")
write.table(dat.csv, file = "path/to/save/filename.txt", sep = "\t",
na=".")
write.dta(dat.csv, file = "path/to/save/filename.dta")
```

```
write.xlsx(dat.csv, file = "path/to/save/filename.xlsx",  
sheetName="hsb2")  
  
# save to binary R format (can save multiple datasets and R objects)  
save(dat.csv, dat.dta, dat.spss, dat.txt, file =  
"path/to/save/filename.RData")
```

Exploring Data فحص و إستكشاف البيانات

نقرأ البيانات من الإنترنت

```
d <- read.csv("http://www.ats.ucla.edu/stat/data/hsb2.csv")
```

Description of Dataset وصف البيانات

```
dim(d)
```

```
## [1] 200 11
```

```
str(d)
```

```
## 'data.frame':      200 obs. of  11 variables:
## $ id      : int  70 121 86 141 172 113 50 11 84 48 ...
## $ female  : int  0 1 0 0 0 0 0 0 0 0 ...
## $ race    : int  4 4 4 4 4 4 3 1 4 3 ...
## $ ses     : int  1 2 3 3 2 2 2 2 2 2 ...
## $ schtyp  : int  1 1 1 1 1 1 1 1 1 1 ...
## $ prog    : int  1 3 1 3 2 2 1 2 1 2 ...
## $ read    : int  57 68 44 63 47 44 50 34 63 57 ...
## $ write   : int  52 59 33 44 52 52 59 46 57 55 ...
## $ math    : int  41 53 54 47 57 51 42 45 54 52 ...
## $ science : int  47 63 58 53 53 63 53 39 58 50 ...
## $ socst   : int  57 61 31 56 61 61 61 36 51 51 ...
```

Descriptive Stats الإحصاء الوصفي

```
summary(d)
```

```
##           id           female           race           ses
## Min.      : 1.0      Min.    :0.000      Min.    :1.00      Min.    :1.00
## 1st Qu.: 50.8      1st Qu.:0.000      1st Qu.:3.00      1st Qu.:2.00
## Median :100.5      Median :1.000      Median :4.00      Median :2.00
## Mean   :100.5      Mean   :0.545      Mean   :3.43      Mean   :2.06
## 3rd Qu.:150.2      3rd Qu.:1.000      3rd Qu.:4.00      3rd Qu.:3.00
## Max.    :200.0      Max.    :1.000      Max.    :4.00      Max.    :3.00
##          schtyp          prog          read          write
## Min.      :1.00      Min.    :1.00      Min.    :28.0      Min.    :31.0
## 1st Qu.:1.00      1st Qu.:2.00      1st Qu.:44.0      1st Qu.:45.8
## Median :1.00      Median :2.00      Median :50.0      Median :54.0
## Mean   :1.16      Mean   :2.02      Mean   :52.2      Mean   :52.8
## 3rd Qu.:1.00      3rd Qu.:2.25      3rd Qu.:60.0      3rd Qu.:60.0
## Max.    :2.00      Max.    :3.00      Max.    :76.0      Max.    :67.0
##          math          science          socst
## Min.      :33.0      Min.    :26.0      Min.    :26.0
## 1st Qu.:45.0      1st Qu.:44.0      1st Qu.:46.0
## Median :52.0      Median :53.0      Median :52.0
## Mean   :52.6      Mean   :51.9      Mean   :52.4
## 3rd Qu.:59.0      3rd Qu.:58.0      3rd Qu.:61.0
```

```
## Max. :75.0 Max. :74.0 Max. :71.0
```

Conditional Summaries التلخيص الشرطي

```
summary(subset(d, read >= 60))
```

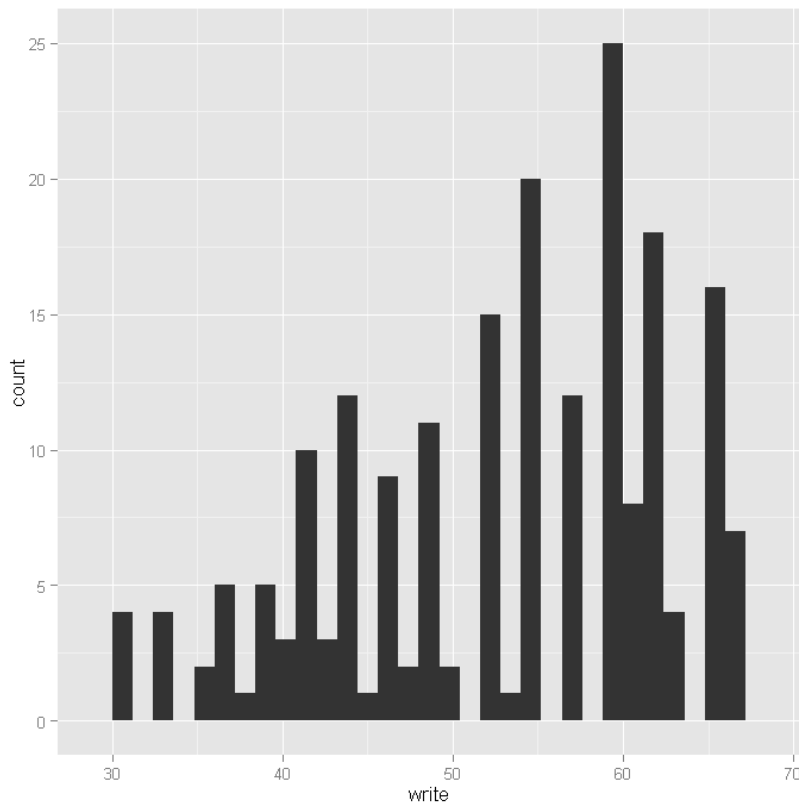
```
##      id      female      race      ses
## Min.   : 3.0   Min.   :0.000   Min.   :1.0   Min.   :1.00
## 1st Qu.: 76.5   1st Qu.:0.000   1st Qu.:4.0   1st Qu.:2.00
## Median :108.5   Median :0.000   Median :4.0   Median :3.00
## Mean   :109.8   Mean   :0.482   Mean   :3.7   Mean   :2.38
## 3rd Qu.:143.2   3rd Qu.:1.000   3rd Qu.:4.0   3rd Qu.:3.00
## Max.   :200.0   Max.   :1.000   Max.   :4.0   Max.   :3.00
##      schtyp      prog      read      write
## Min.   :1.00   Min.   :1.00   Min.   :60.0   Min.   :43.0
## 1st Qu.:1.00   1st Qu.:2.00   1st Qu.:63.0   1st Qu.:57.0
## Median :1.00   Median :2.00   Median :65.0   Median :60.0
## Mean   :1.18   Mean   :1.95   Mean   :65.5   Mean   :59.5
## 3rd Qu.:1.00   3rd Qu.:2.00   3rd Qu.:68.0   3rd Qu.:65.0
## Max.   :2.00   Max.   :3.00   Max.   :76.0   Max.   :67.0
##      math      science      socst
## Min.   :35.0   Min.   :44.0   Min.   :41.0
## 1st Qu.:55.8   1st Qu.:55.0   1st Qu.:56.0
## Median :60.5   Median :61.0   Median :61.0
## Mean   :60.2   Mean   :59.7   Mean   :60.9
## 3rd Qu.:66.2   3rd Qu.:65.2   3rd Qu.:66.0
## Max.   :75.0   Max.   :74.0   Max.   :71.0
```

```
by(d[, 7:11], d$prog, colMeans)
```

```
## d$prog: 1
##      read      write      math science      socst
##      49.76      51.33      50.02      52.44      50.60
## -----
## d$prog: 2
##      read      write      math science      socst
##      56.16      56.26      56.73      53.80      56.70
## -----
## d$prog: 3
##      read      write      math science      socst
##      46.20      46.76      46.42      47.22      45.02
```

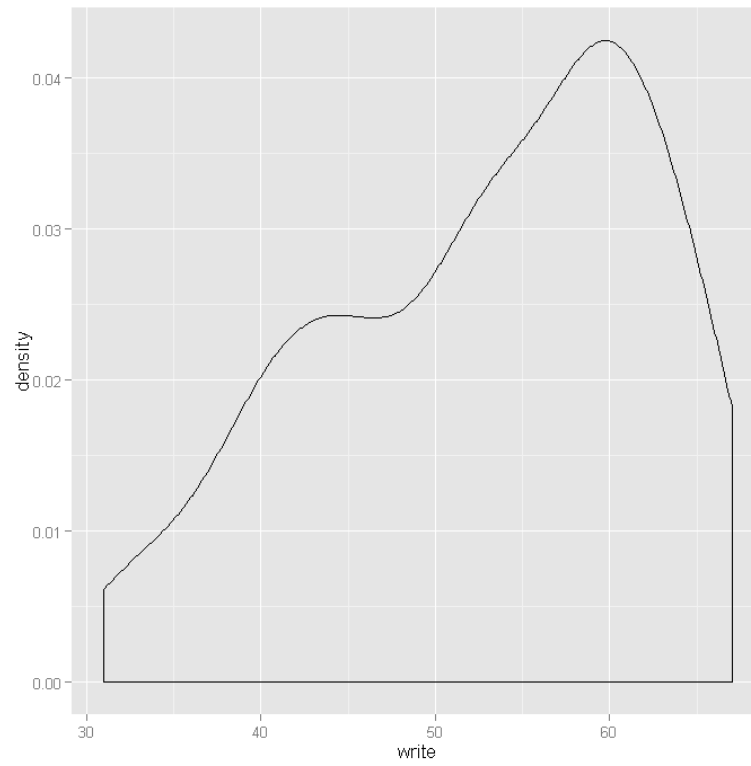

Histograms المدرج التكراري

```
ggplot(d, aes(x = write)) + geom_histogram()
```



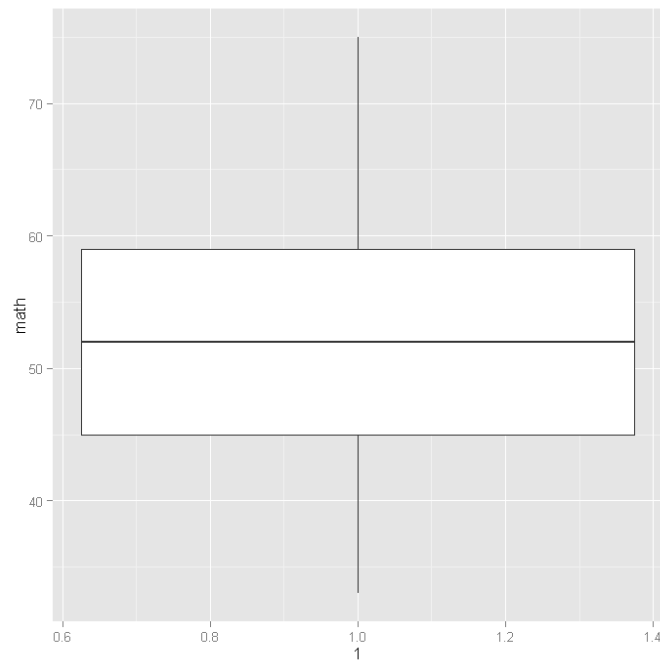
Density Plots رسومات الكثافة

```
ggplot(d, aes(x = write)) + geom_density()
```



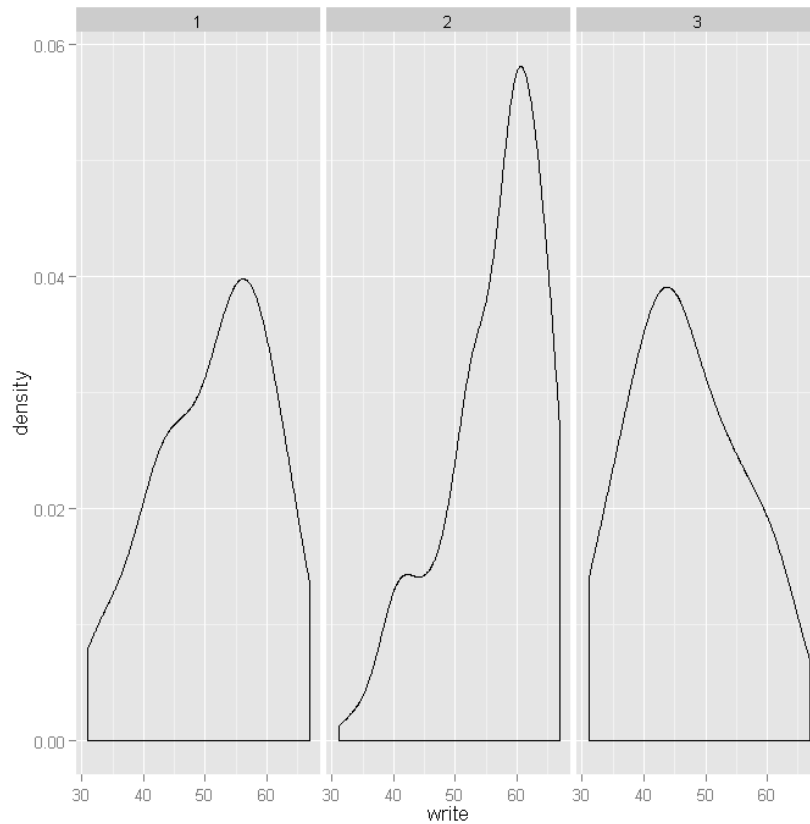
رسومات الصندوق Boxplots

`ggplot(d, aes(x = 1, y = math)) + geom_boxplot()`

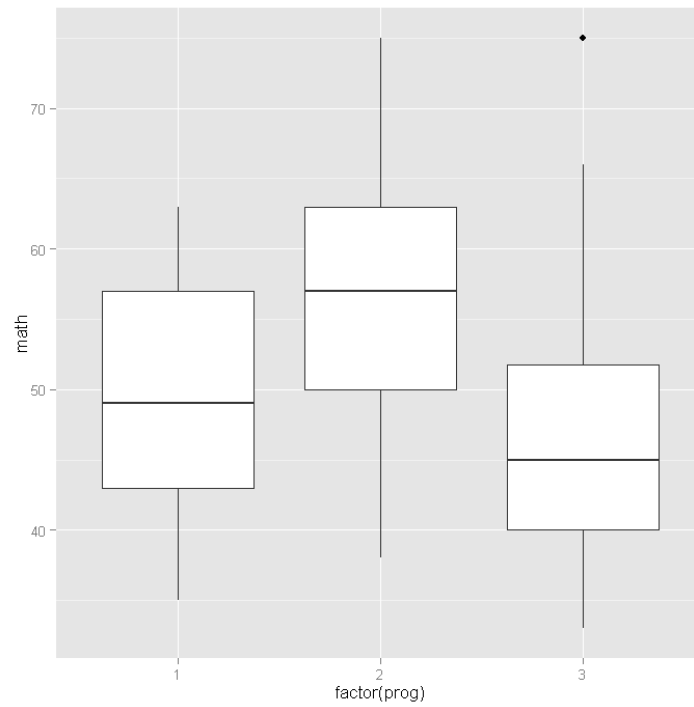


الفحص المشروط Conditional Visualization

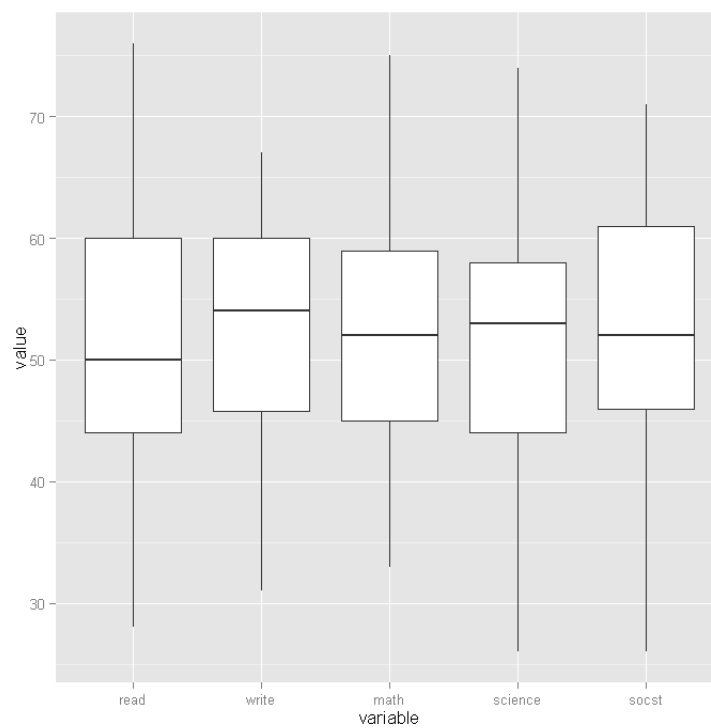
```
# density plots by program type  
ggplot(d, aes(x = write)) + geom_density() + facet_wrap(~prog)
```



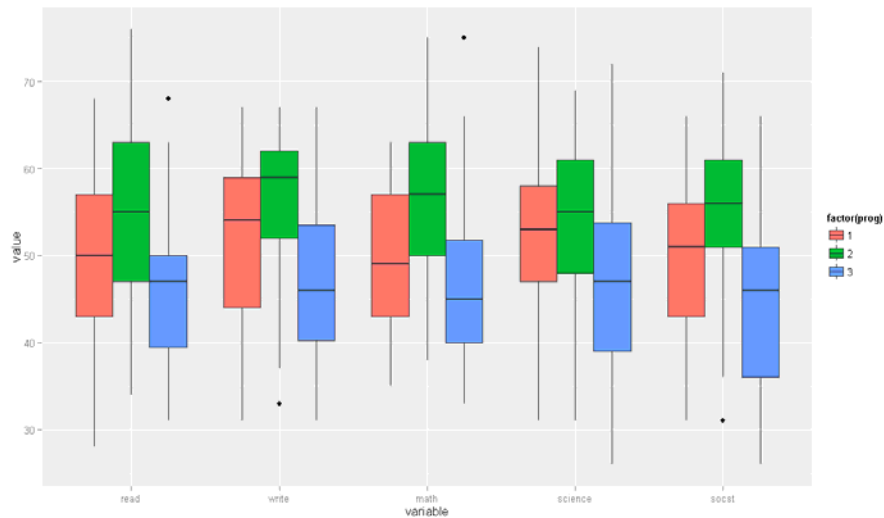
```
ggplot(d, aes(x = factor(prog), y = math)) + geom_boxplot()
```



```
ggplot(melt(d[, 7:11]), aes(x = variable, y = value)) + geom_boxplot()
## Using as id variables
```

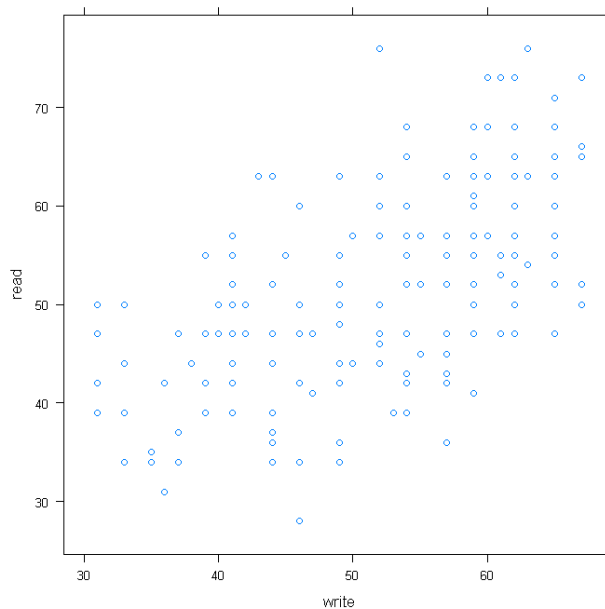


```
ggplot(melt(d[, 6:11], id.vars = "prog"),
       aes(x = variable, y = value, fill = factor(prog))) +
  geom_boxplot()
```

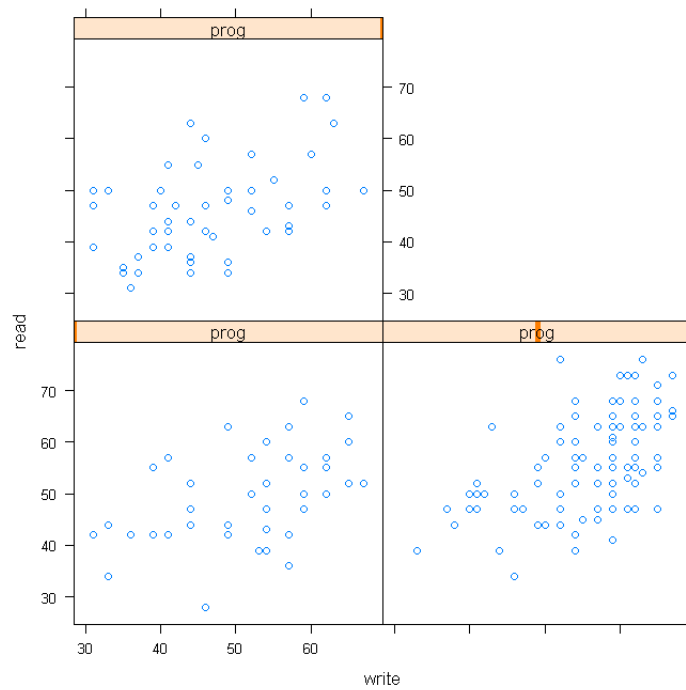


الرسومات في R Graphics in R

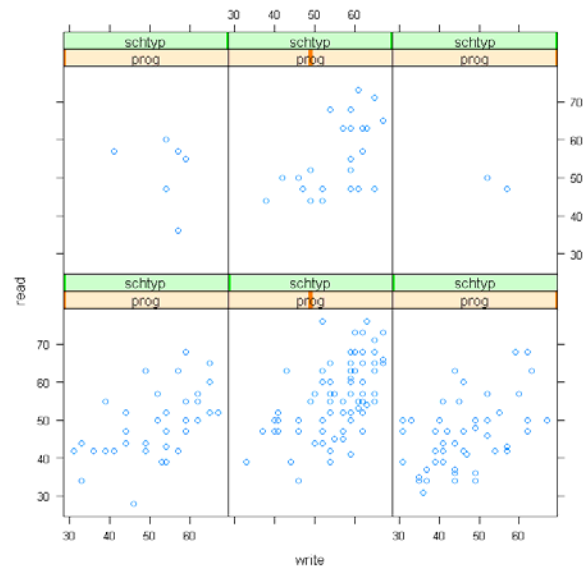
```
# load lattice
require(lattice)
## Loading required package: lattice
# simple scatter plot
xyplot(read ~ write, data = d)
```



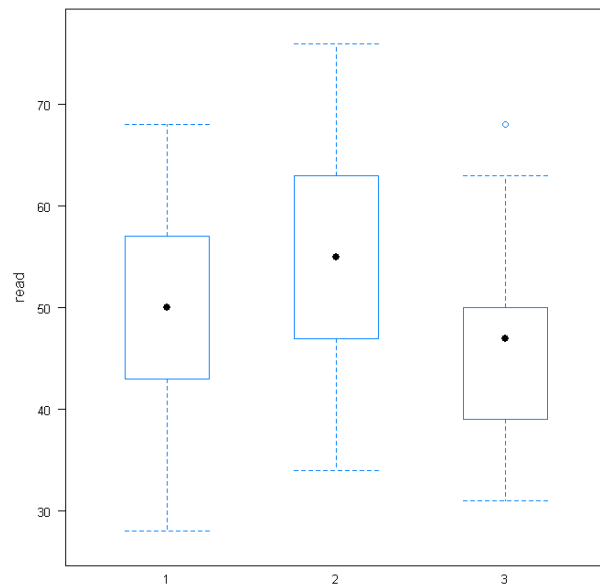
```
# conditioned scatter plot
xyplot(read ~ write | prog, data = d)
```



```
# conditioning on two variables
xyplot(read ~ write | prog * schtyp, data = d)
```



```
# box and whisker plots
bwplot(read ~ factor(prog), data = d)
```



Categorical Data البيانات الوصفية

```
xtabs(~female, data = d)
```

```
## female
##    0    1
##  91 109
```

```
xtabs(~race, data = d)
```

```
## race
##    1    2    3    4
##  24   11  20 145
```

```
xtabs(~prog, data = d)
```

```
## prog
##    1    2    3
##  45 105  50
```

```
xtabs(~ses + schtyp, data = d)
```

```
##      schtyp
## ses  1  2
##    1  45  2
##    2  76 19
##    3  47 11
```

Categorical Data 3

```
(tab3 <- xtabs(~ses + prog + schtyp, data = d))

## , , schtyp = 1
##
##      prog
## ses   1   2   3
##   1 14 19 12
##   2 17 30 29
##   3   8 32   7
##
## , , schtyp = 2
##
##      prog
## ses   1   2   3
##   1   2   0   0
##   2   3 14   2
##   3   1 10   0

(tab2 <- xtabs(~ses + schtyp, data = d))

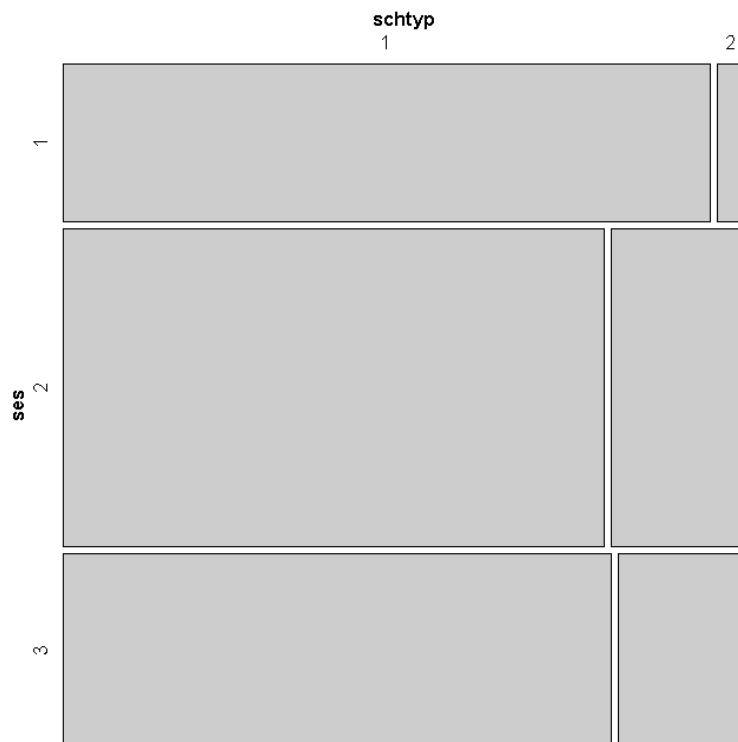
##      schtyp
## ses   1   2
##   1 45   2
##   2 76 19
##   3 47 11

set.seed(10)
(testtab2 <- coinddep_test(tab2, n = 5000))

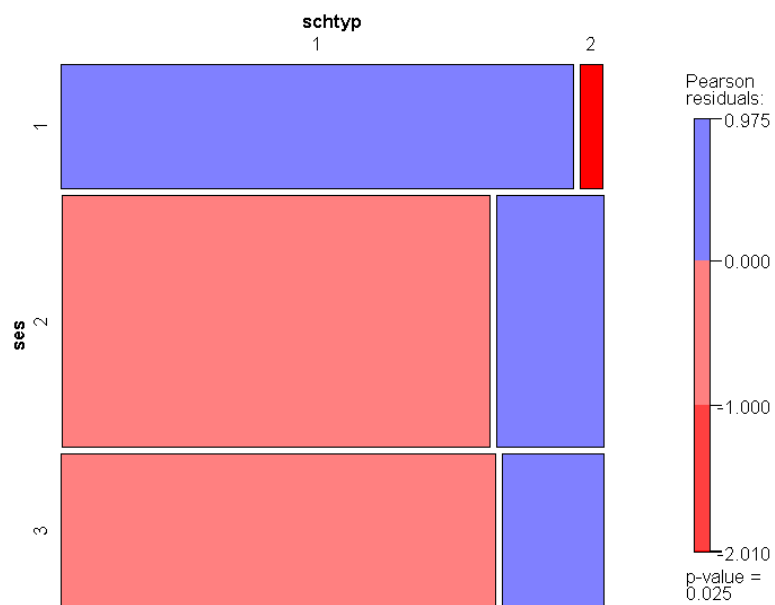
##
##      Permutation test for conditional independence
##
## data:  tab2
## f(x) = 2.013, p-value = 0.0248
```

تصوير أو مشاهدة البيانات الوصفية Visualizing Cat Data

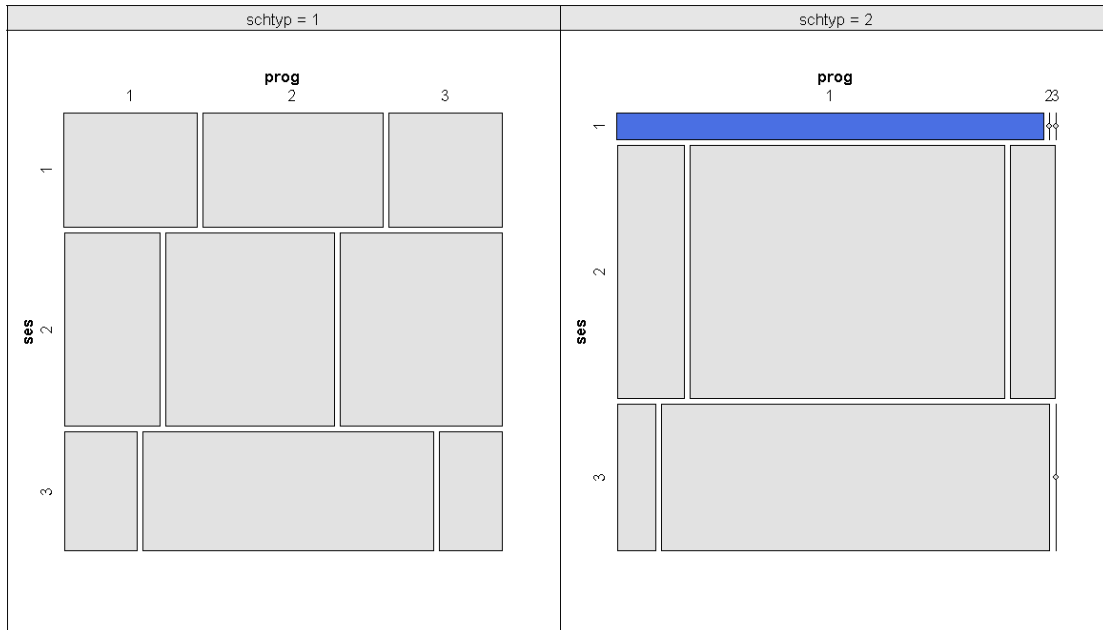
```
# simple mosaic plot
mosaic(tab2)
```

```
mosaic(tab2, gp = shading_hsv, gp_args = list(p.value =
testtab2$p.value, interpolate = -1:2))
```



```
cotabplot(~ses + prog | schtyp, data = d, panel = cotab_coinddep, n = 5000)
```



Correlations الترابط

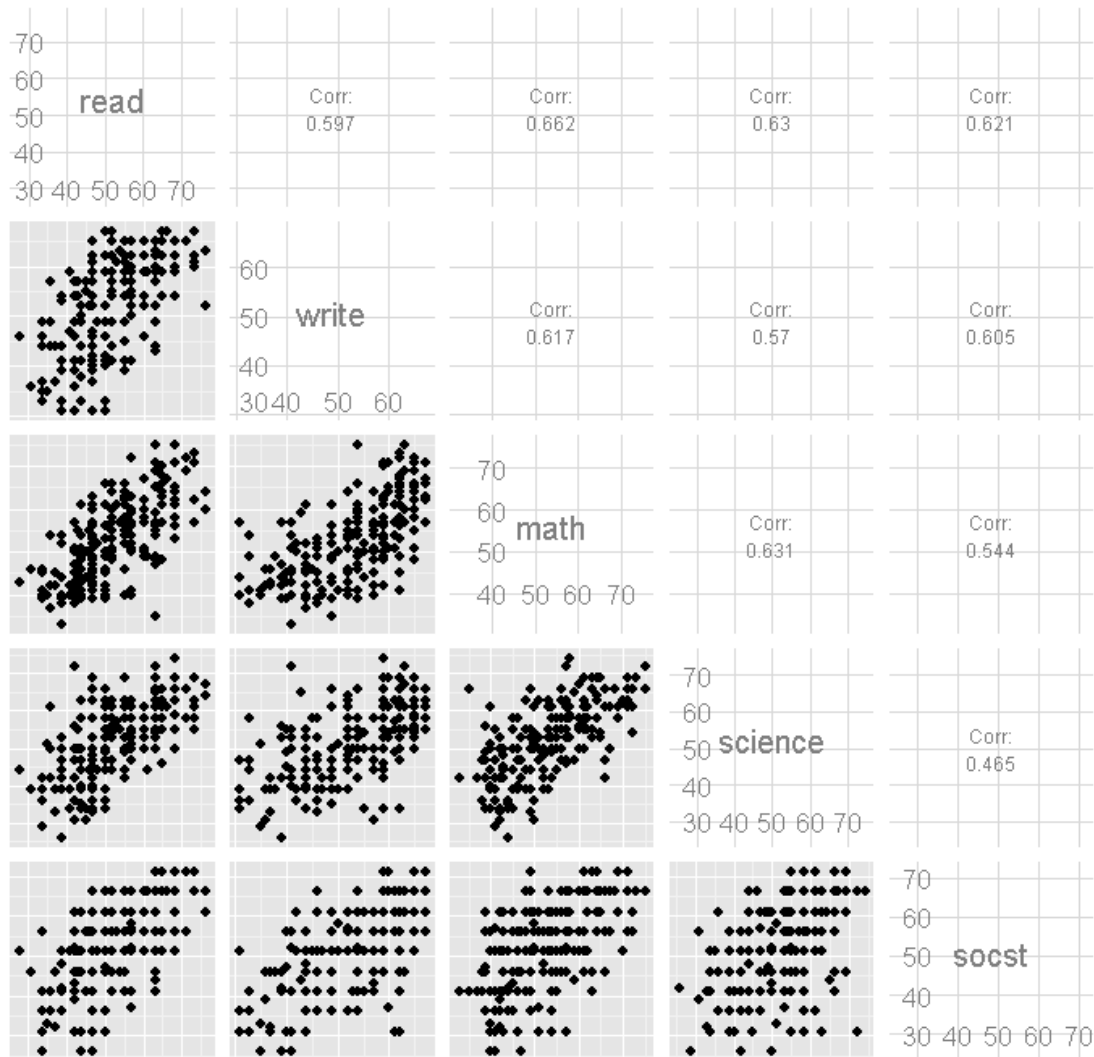
```
cor(d[, 7:11])
##          read  write   math science  socst
## read      1.0000 0.5968 0.6623  0.6302 0.6215
## write     0.5968 1.0000 0.6174  0.5704 0.6048
## math      0.6623 0.6174 1.0000  0.6307 0.5445
## science   0.6302 0.5704 0.6307  1.0000 0.4651
## socst     0.6215 0.6048 0.5445  0.4651 1.0000
```

```
cor(d[, 7:11], use = "complete.obs")
##          read  write   math science  socst
## read      1.0000 0.5968 0.6623  0.6302 0.6215
## write     0.5968 1.0000 0.6174  0.5704 0.6048
## math      0.6623 0.6174 1.0000  0.6307 0.5445
## science   0.6302 0.5704 0.6307  1.0000 0.4651
## socst     0.6215 0.6048 0.5445  0.4651 1.0000
```

```
cor(d[, 7:11], use = "pairwise.complete.obs")
##          read  write   math science  socst
## read      1.0000 0.5968 0.6623  0.6302 0.6215
## write     0.5968 1.0000 0.6174  0.5704 0.6048
## math      0.6623 0.6174 1.0000  0.6307 0.5445
## science   0.6302 0.5704 0.6307  1.0000 0.4651
## socst     0.6215 0.6048 0.5445  0.4651 1.0000
```

Visual Summaries, Continuous تلخيص بصري للمتغيرات المتصلة Variables

```
ggpairs(d[, 7:11])
```



Modifying Data تعديل البيانات سوف نستعرضها على البيانات

```
# read data in and store in an easy to use name to save typing
d <- read.csv("http://www.ats.ucla.edu/stat/data/hsb2.csv")
```

الترتيب Sorting

```
d <- d[order(d$id, d$female), ]
head(d)
##      id female race ses schtyp prog read write math science socst
## 99    1      1    1    1      1    3   34    44   40     39    41
## 139   2      1    1    2      1    3   39    41   33     42    41
## 84    3      0    1    1      1    2   63    65   48     63    56
## 112   4      1    1    1      1    2   44    50   41     39    51
## 76    5      0    1    1      1    2   47    40   43     45    31
## 149   6      1    1    1      1    2   47    41   46     40    41
```

توسيم المتغيرات الوصفية Labeling Categorical Vars

```
str(d)
## 'data.frame':      200 obs. of  11 variables:
## $ id      : int  1 2 3 4 5 6 7 8 9 10 ...
## $ female  : int  1 1 0 1 0 1 0 1 0 1 ...
## $ race    : int  1 1 1 1 1 1 1 1 1 1 ...
## $ ses     : int  1 2 1 1 1 1 2 1 2 2 ...
## $ schtyp  : int  1 1 1 1 1 1 1 1 1 1 ...
## $ prog    : int  3 3 2 2 2 2 2 2 3 1 ...
## $ read    : int  34 39 63 44 47 47 57 39 48 47 ...
## $ write   : int  44 41 65 50 40 41 54 44 49 54 ...
## $ math    : int  40 33 48 41 43 46 59 52 52 49 ...
## $ science : int  39 42 63 39 45 40 47 44 44 53 ...
## $ socst   : int  41 41 56 51 31 41 51 48 51 61 ...
d <- within(d, {
  id <- factor(id)
  female <- factor(female, levels = 0:1, labels = c("male", "female"))
  race <- factor(race, levels = 1:4, labels = c("Hispanic", "Asian",
"African American", "White"))
  schtyp <- factor(schtyp, levels = 1:2, labels = c("public",
"private"))
  prog <- factor(prog, levels = 1:3, labels = c("general", "academic",
"vocational"))
})
```

النتائج

```
str(d)
## 'data.frame':      200 obs. of  11 variables:
## $ id      : Factor w/ 200 levels "1","2","3","4",...: 1 2 3 4 5 6 7 8
9 10 ...
## $ female  : Factor w/ 2 levels "male","female": 2 2 1 2 1 2 1 2 1 2
...
## $ race    : Factor w/ 4 levels "Hispanic","Asian",...: 1 1 1 1 1 1 1
1 1 1 ...
## $ ses     : int  1 2 1 1 1 1 2 1 2 2 ...
## $ schtyp  : Factor w/ 2 levels "public","private": 1 1 1 1 1 1 1 1
1 ...
## $ prog    : Factor w/ 3 levels "general","academic",...: 3 3 2 2 2 2
2 2 3 1 ...
## $ read    : int  34 39 63 44 47 47 57 39 48 47 ...
## $ write   : int  44 41 65 50 40 41 54 44 49 54 ...
```

```
## $ math : int 40 33 48 41 43 46 59 52 52 49 ...
## $ science: int 39 42 63 39 45 40 47 44 44 53 ...
## $ socst : int 41 41 56 51 31 41 51 48 51 61 ...
summary(d)
##      id      female      race      ses
## 1      : 1   male : 91   Hispanic : 24   Min. :1.00
## 2      : 1   female:109   Asian : 11   1st Qu.:2.00
## 3      : 1      African American: 20   Median :2.00
## 4      : 1      White :145   Mean :2.06
## 5      : 1      3rd Qu.:3.00
## 6      : 1      Max. :3.00
## (Other):194
##      schtyp      prog      read      write
## public :168   general : 45   Min. :28.0   Min. :31.0
## private: 32   academic :105   1st Qu.:44.0   1st Qu.:45.8
##          vocational: 50   Median :50.0   Median :54.0
##          Mean :52.2   Mean :52.8
##          3rd Qu.:60.0   3rd Qu.:60.0
##          Max. :76.0   Max. :67.0
##
##      math      science      socst
## Min. :33.0   Min. :26.0   Min. :26.0
## 1st Qu.:45.0   1st Qu.:44.0   1st Qu.:46.0
## Median :52.0   Median :53.0   Median :52.0
## Mean :52.6   Mean :51.9   Mean :52.4
## 3rd Qu.:59.0   3rd Qu.:58.0   3rd Qu.:61.0
## Max. :75.0   Max. :74.0   Max. :71.0
##
```

ترقيم و ترميز Scoring & Recoding

```
d$total <- rowSums(d[, 7:10], na.rm=FALSE)
# recode by breaking into categories
d$grade <- cut(d$total,
  breaks = c(0, 140, 180, 210, 234, 300),
  labels = c("F", "D", "C", "B", "A"))
# view results
summary(d[, c("total", "grade")])
##      total      grade
## Min. :139   F: 1
## 1st Qu.:180   D:51
## Median :210   C:50
## Mean :210   B:49
## 3rd Qu.:234   A:49
## Max. :277
```

معايرة ومتوسط Standardize and Average

```
d <- within(d, {
  zread <- scale(read)
  readmean <- ave(read, ses, FUN = mean)
})

head(d[, c("read", "zread", "readmean")])
##      read      zread readmean
## 99      34 -1.7780      48.28
```

```
## 139    39 -1.2904    51.58
## 84     63  1.0504    48.28
## 112    44 -0.8027    48.28
## 76     47 -0.5101    48.28
## 149    47 -0.5101    48.28
```

ترقيم Scoring

```
d$rowmean <- rowMeans(d[, 7:10], na.rm = TRUE)
```

إدارة البيانات Managing Data

الأدلة Directories in R

```
getwd()
## [1] "D:/www/stat/r/seminars"
list.files()
## [1] "buildRpagesScript.R" "figures"          "intro.htm"
## [4] "intro.R"             "intro.Rhtml"      "R.svg"
## [7] "R.vsd"               "Repeated_Measures" "ui"
# setwd('/path/to/directory')
```

تقسيم أو تجزئ المشاهدات Subsetting Observations

```
dfemale <- subset(d, female == "female")
dmale <- subset(d, female == "male")
```

تقسيم المتغيرات Subsetting Variables

```
# note that select is special, so we do not need to quote the variable
# names
duse <- subset(d, select = c(id, female, read, write))
# note the - preceding c(female... , which means drop these variables
ddropped <- subset(d, select = -c(female, read, write))
```

إضافة مشاهدات Adding Observations (appending)

```
dboth <- rbind(dfemale, dmale)
dim(dfemale)
## [1] 109 16
dim(dmale)
## [1] 91 16
dim(dboth)
## [1] 200 16
```

Merging Data دمج البيانات

```
dall <- merge(duse, ddropped, by = "id", all = TRUE)
dim(duse)
## [1] 200 4
dim(ddropped)
## [1] 200 13
dim(dall)
## [1] 200 16
```

Analyzing Data تحليل البيانات

Analyzing Cat Data تحليل البيانات الوصفية

تحلل العلاقة بين متغيرات وصفية بواسطة إختبارات كاي تربيع. الدالة `chisq.test` يمكن أن تستخدم على بيانات خامة أو يمكن إستخدامها مع جداول تكرارية كآلاتي

```
(tab <- xtabs(~ses + schtyp, data = d))
##      schtyp
## ses public private
## 1      45         2
## 2      76        19
## 3      47        11
chisq.test(tab)
##
##      Pearson's Chi-squared test
##
## data:  tab
## X-squared = 6.334, df = 2, p-value = 0.04213
```

في حالة الخلايا التي تحوي قليل من البيانات نستطيع إستخدام Monte Carlo Simulation كالتالي

```
chisq.test(tab, simulate.p.value = TRUE, B = 90000)
##
##      Pearson's Chi-squared test with simulated p-value (based on
##90000 replicates)
##
## data:  tab
## X-squared = 6.334, df = NA, p-value = 0.04216
```

t-tests إختبارات t

```
t.test(d$write, mu = 50)
##
##      One Sample t-test
##
## data:  d$write
## t = 4.14, df = 199, p-value = 5.121e-05
## alternative hypothesis: true mean is not equal to 50
```

```

## 95 percent confidence interval:
##  51.45 54.10
## sample estimates:
## mean of x
##      52.77

with(d, t.test(write, read, paired = TRUE))

##
##      Paired t-test
##
## data:  write and read
## t = 0.8673, df = 199, p-value = 0.3868
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.6941  1.7841
## sample estimates:
## mean of the differences
##                  0.545

t.test(write ~ female, data = d, var.equal = TRUE)

##
##      Two Sample t-test
##
## data:  write by female
## t = -3.734, df = 198, p-value = 0.0002463
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -7.442 -2.298
## sample estimates:
##  mean in group male mean in group female
##                50.12                54.99

t.test(write ~ female, data = d)

##
##      Welch Two Sample t-test
##
## data:  write by female
## t = -3.656, df = 169.7, p-value = 0.0003409
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -7.499 -2.241
## sample estimates:
##  mean in group male mean in group female
##                50.12                54.99

```

ANOVA and Regression تحليل التباين والانحدار

```

m <- lm(write ~ prog * female, data = d)
anova(m)

```

```
## Analysis of Variance Table
```



```
##
## Response: write
##           Df Sum Sq Mean Sq F value    Pr(>F)
## prog         2   3176     1588    23.25 8.9e-10 ***
## female       1   1129     1129    16.53 7.0e-05 ***
## prog:female   2    326      163     2.39  0.095 .
## Residuals   194  13249         68
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

summary(m)

```
##
## Call:
## lm(formula = write ~ prog * female, data = d)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -21.62  -5.14   1.04   6.12  21.17
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      49.14       1.80   27.25 <2e-16 ***
## progacademic       5.47       2.17    2.52  0.0124 *
## progvocational    -7.32       2.49   -2.93  0.0038 **
## femalefemale       4.11       2.47    1.66  0.0979 .
## progacademic:femalefemale -1.14       2.95   -0.39  0.7005
## progvocational:femalefemale  5.03       3.41    1.48  0.1413
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.26 on 194 degrees of freedom
## Multiple R-squared:  0.259, Adjusted R-squared:  0.24
## F-statistic: 13.6 on 5 and 194 DF, p-value: 2.38e-11
```

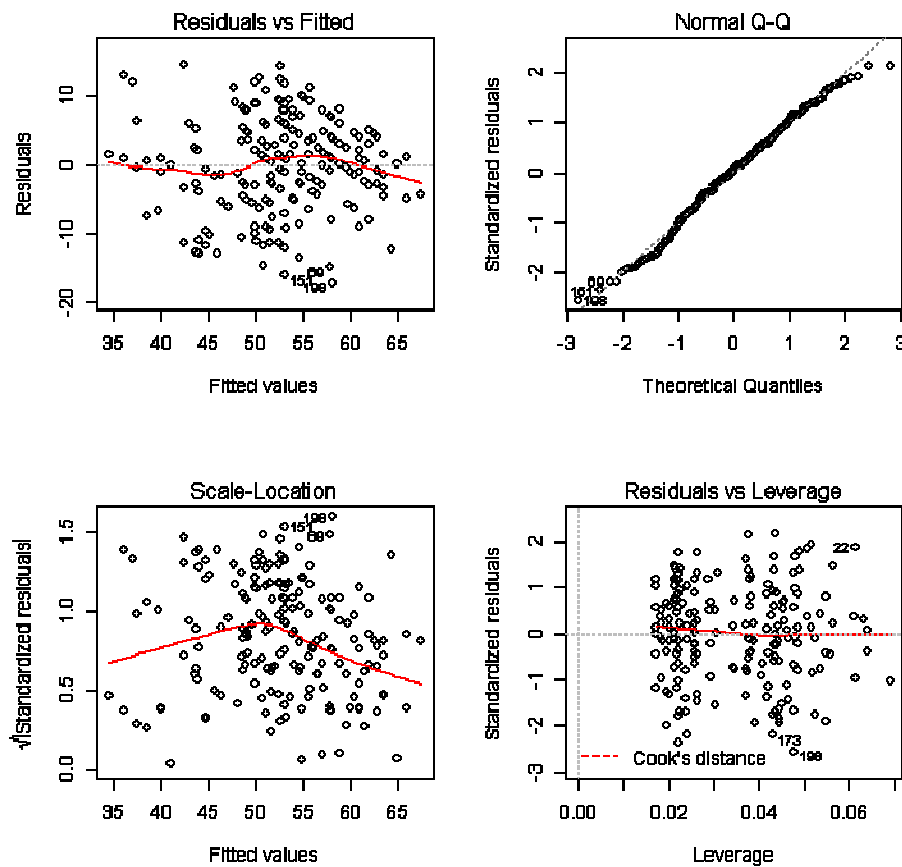
summary(m2 <- update(m, . ~ . + read))

```
##
## Call:
## lm(formula = write ~ prog + female + read + prog:female, data = d)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -17.219  -4.467   0.426   4.608  14.507
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    22.9406     3.1881    7.20 1.3e-11 ***
## progacademic     3.8293     1.8147    2.11 0.03614 *
## progvocational  -3.7044     2.1127   -1.75 0.08113 .
## femalefemale     7.0732     2.0806    3.40 0.00082 ***
## read            0.4948     0.0531    9.32 < 2e-16 ***
## progacademic:femalefemale -4.0012     2.4791   -1.61 0.10816
## progvocational:femalefemale  1.5617     2.8598    0.55 0.58563
## ---
```

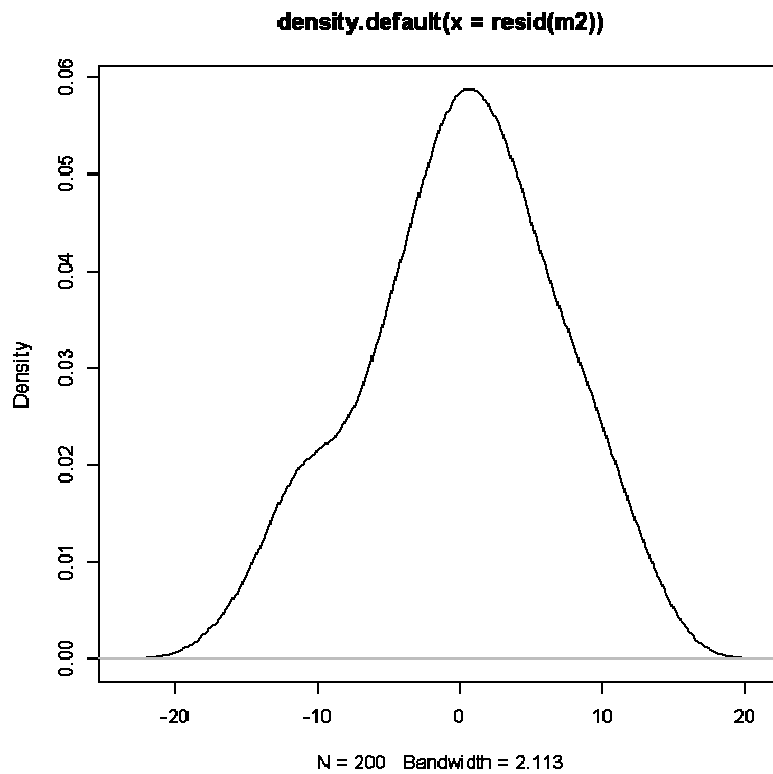
```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.88 on 193 degrees of freedom
## Multiple R-squared:  0.489, Adjusted R-squared:  0.473
## F-statistic: 30.8 on 6 and 193 DF,  p-value: <2e-16
```

Regression Diagnostics تشخيص الإنحدار

```
par(mfrow = c(2, 2))
plot(m2)
```



```
plot(density(resid(m2)))
```



```
summary(m3 <- lm(write ~ prog * read, data = d))
```

```
##
## Call:
## lm(formula = write ~ prog * read, data = d)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -17.595  -4.773   0.557   5.375  18.405
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    27.33720     6.17869   4.42 1.6e-05 ***
## progacademic     2.82708     7.56135   0.37 0.70890
## progvocational  -2.88581     8.36821  -0.34 0.73058
## read             0.48228     0.12214   3.95 0.00011 ***
## progacademic:read -0.01768     0.14413  -0.12 0.90250
## progvocational:read 0.00059     0.17122   0.00 0.99725
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.48 on 194 degrees of freedom
## Multiple R-squared:  0.393, Adjusted R-squared:  0.377
## F-statistic: 25.1 on 5 and 194 DF, p-value: <2e-16
```

```

m3b <- update(m3, . ~ . - prog:read)
anova(m3b, m3)

## Analysis of Variance Table
##
## Model 1: write ~ prog + read
## Model 2: write ~ prog * read
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1      196 10861
## 2      194 10860  2      1.37 0.01  0.99

```

Estimation التقدير

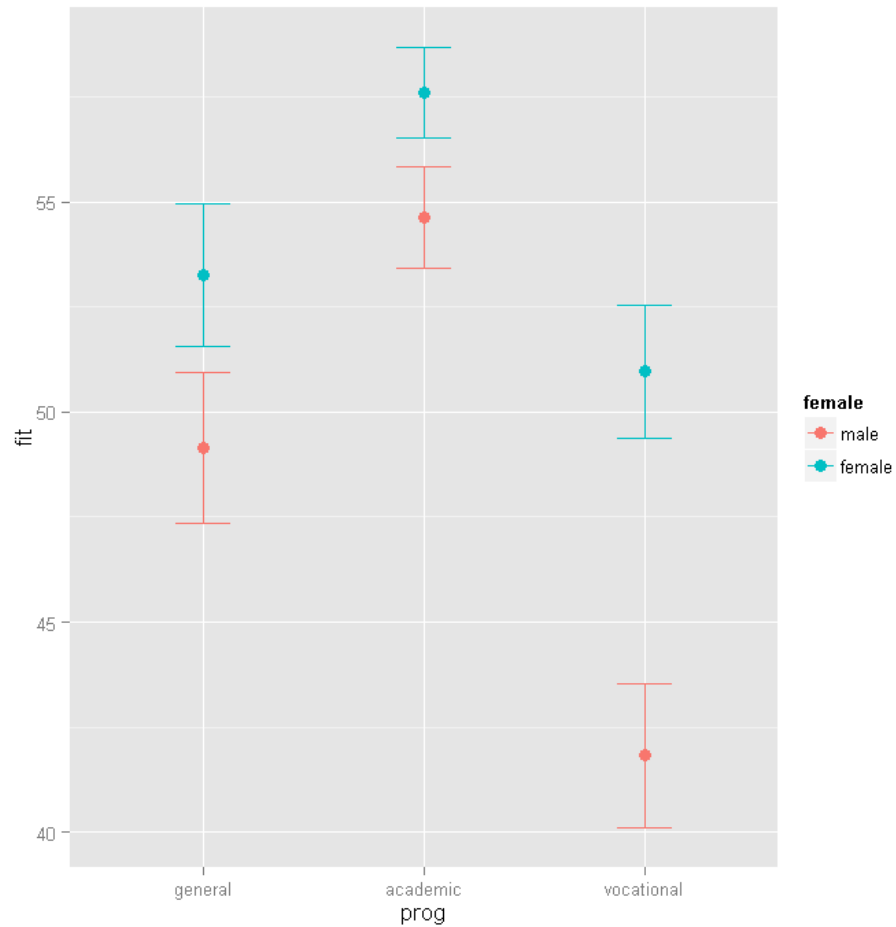
```

newdat <- with(d, expand.grid(prog = levels(prog), female =
levels(female)))
(newdat <- cbind(newdat, predict(m, newdat, se = TRUE)))

##      prog female   fit se.fit  df residual.scale
## 1  general   male 49.14  1.803 194          8.264
## 2  academic   male 54.62  1.205 194          8.264
## 3 vocational   male 41.83  1.723 194          8.264
## 4  general female 53.25  1.687 194          8.264
## 5  academic female 57.59  1.085 194          8.264
## 6 vocational female 50.96  1.590 194          8.264

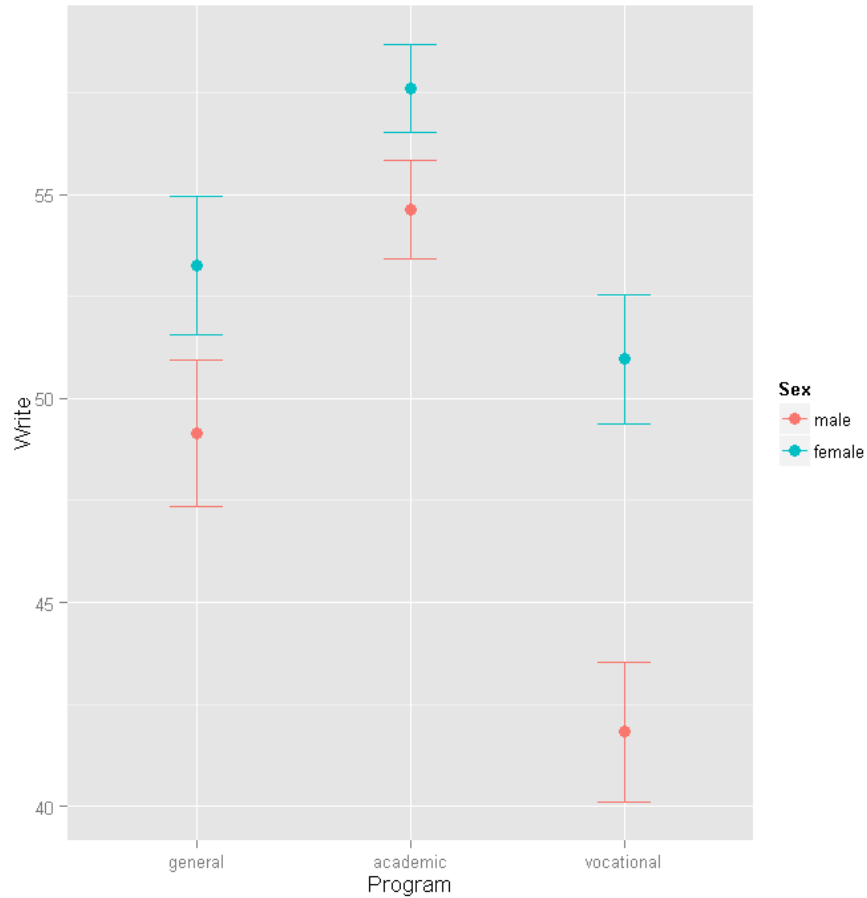
ggplot(newdat, aes(x = prog, y = fit, colour = female)) +
  geom_errorbar(aes(ymin = fit - se.fit, ymax = fit + se.fit),
width=.25) +
  geom_point(size=3)

```



```
colnames(newdat)[1:3] <- c("Program", "Sex", "Write")
```

```
ggplot(newdat, aes(x = Program, y = Write, colour = Sex)) +  
  geom_errorbar(aes(ymin = Write - se.fit, ymax = Write + se.fit),  
width=.25) +  
  geom_point(size=3)
```



R Learning Module

ادرس التالي:

```
# Reading in numeric data
> x <- scan()
1: 3 5 6
4: 3 5 78 29
8: 34 5 1 78
12:
Read 11 items
> x
[1] 3 5 6 3 5 78 29 34 5 1 78
> mode(x)
[1] "numeric"
# Reading in string data
# empty quotes indicates character input
> y <- scan(what=" ")
1: red blue
3: green red
```

```

5: blue yellow
7:
Read 6 items
> y
[1] "red"    "blue"   "green"  "red"    "blue"   "yellow"
> mode(y)
[1] "character"

# inputting a text file and outputting a list
(x <- scan("http://www.ats.ucla.edu/stat/data/scan.txt", what =
list(age = 0,
      name = "")))
## $age
## [1] 12 24 35 20
##
## $name
## [1] "bobby"   "kate"    "david"   "michael"
##

# using the same text file and saving only the names as a vector
x <- scan("http://www.ats.ucla.edu/stat/data/scan.txt", what =
list(NULL,
      name = character()))

(x <- x[sapply(x, length) > 0])
## $name
## [1] "bobby"   "kate"    "david"   "michael"
##
is.vector(x)
## [1] TRUE

# complete data, space delimited, variable names in first row
(test <- read.table("http://www.ats.ucla.edu/stat/data/test.txt",
header = TRUE))

##      prgtype gender   id ses schtyp level
## 1  general      0  70   4      1      1
## 2  vocati      1 121   4      2      1
## 3  general      0  86   4      3      1
## 4  vocati      0 141   4      3      1
## 5 academic     0 172   4      2      1
## 6 academic     0 113   4      2      1
## 7  general      0  50   3      2      1
## 8 academic     0  11   1      2      1

# showing the file with missing values, space delimited
(test_missing.txt data file)
prgtype gender   id ses schtyp level
general      0  70   4      1      1
vocati      1 121   4      1
general      0  86      1
vocati      0 141   4      3      1
academic     0 172   4      2      1
academic     0 113   4      2      1

```

```

    general    0  50    3  2    1
    academic   0  11    1  2    1
test.missing <-
read.table("http://www.ats.ucla.edu/stat/data/test_missing.txt",
  header = TRUE)
## Error: line 2 did not have 6 elements
# showing the file with missing data, comma delimited
(test_missing_comma.txt data file)
prgtype, gender, id, ses, schtyp, level
general,    0,  70,    4,    1,    1
vocati,    1, 121,    4,    ,    1
general,    0,  86,    ,    ,    1
vocati,    0, 141,    4,    3,    1
academic,   0, 172,    4,    2,    1
academic,   0, 113,    4,    2,    1
general,    0,  50,    3,    2,    1
academic,   0,  11,    1,    2,    1
(test.missing <-
read.table("http://www.ats.ucla.edu/stat/data/test_missing_comma.txt",
  header = TRUE, sep = ","))
##      prgtype gender   id ses schtyp level
## 1  general      0   70   4      1      1
## 2  vocati      1 121   4     NA      1
## 3  general      0   86  NA     NA      1
## 4  vocati      0 141   4      3      1
## 5 academic     0 172   4      2      1
## 6 academic     0 113   4      2      1
## 7  general      0   50   3      2      1
## 8 academic     0   11   1      2      1

test.csv <- read.csv("http://www.ats.ucla.edu/stat/data/test.csv",
  header = TRUE)

print(test.csv)
##      prgtype gender   id ses schtyp level
## 1  general      0   70   4      1      1
## 2  vocati      1 121   4      2      1
## 3  general      0   86   4      3      1
## 4  vocati      0 141   4      3      1
## 5 academic     0 172   4      2      1
## 6 academic     0 113   4      2      1
## 7  general      0   50   3      2      1
## 8 academic     0   11   1      2      1
test.csv1 <- read.table("http://www.ats.ucla.edu/stat/data/test.csv",
  header = TRUE, sep = ",")

print(test.csv1)
##      prgtype gender   id ses schtyp level
## 1  general      0   70   4      1      1
## 2  vocati      1 121   4      2      1
## 3  general      0   86   4      3      1
## 4  vocati      0 141   4      3      1
## 5 academic     0 172   4      2      1
## 6 academic     0 113   4      2      1
## 7  general      0   50   3      2      1
## 8 academic     0   11   1      2      1

```



```

test.semi <-
read.table("http://www.ats.ucla.edu/stat/data/testsemicolon.txt",
  header = TRUE, sep = ";")
print(test.semi)
##      prgtype gender  id ses schtyp level
## 1 general      0  70   4     1     1
## 2 vocati      1 121   4     2     1
## 3 general      0  86   4     3     1
## 4 vocati      0 141   4     3     1
## 5 academic     0 172   4     2     1
## 6 academic     0 113   4     2     1
## 7 general      0  50   3     2     1
## 8 academic     0  11   1     2     1
test.z <- read.table("http://www.ats.ucla.edu/stat/data/testz.txt",
  header = TRUE, sep = "z")
print(test.z)
##      prgtype gender  id ses schtyp level
## 1 general      0  70   4     1     1
## 2 vocati      1 121   4     2     1
## 3 general      0  86   4     3     1
## 4 vocati      0 141   4     3     1
## 5 academic     0 172   4     2     1
## 6 academic     0 113   4     2     1
## 7 general      0  50   3     2     1
## 8 academic     0  11   1     2     1

test.fixed <-
read.fwf("http://www.ats.ucla.edu/stat/data/test_fixed.txt",
  width = c(8, 1, 3, 1, 1, 1))

dimnames(test.fixed)[[2]] <- c("prgtyp", "gender", "id", "ses",
"schtyp",
"level")

test.fixed
##      prgtyp gender  id ses schtyp level
## 1 general      0  70   4     1     1
## 2 vocati      1 121   4     2     1
## 3 general      0  86   4     3     1
## 4 vocati      0 141   4     3     1
## 5 academic     0 172   4     2     1
## 6 academic     0 113   4     2     1
## 7 general      0  50   3     2     1
## 8 academic     0  11   1     2     1

(names <- scan("http://www.ats.ucla.edu/stat/data/names.txt", what =
character()))
## [1] "prgtyp" "gender" "id"      "ses"      "schtyp" "level"
(test.fixed <-
read.fwf("http://www.ats.ucla.edu/stat/data/test_fixed.txt",
  col.names = names, width = c(8, 1, 3, 1, 1, 1)))

```

```
##      prgtyp gender  id ses schtyp level
## 1 general      0  70   4      1      1
## 2 vocati      1 121   4      2      1
## 3 general      0  86   4      3      1
## 4 vocati      0 141   4      3      1
## 5 academic     0 172   4      2      1
## 6 academic     0 113   4      2      1
## 7 general      0  50   3      2      1
## 8 academic     0  11   1      2      1
```

```
# using the test.csv data frame to write a text file with no row names
and
# without quotes around the character values (both column names and
string
# variables)
write.table(test.csv, file.path(tempdir(), "test1.txt"), row.names =
FALSE,
           quote = FALSE)
```

```
data(WorldPhones)
is.data.frame(WorldPhones)
## [1] FALSE
WorldPhones
##      N.Amer Europe Asia S.Amer Oceania Africa Mid.Amer
## 1951  45939  21574 2876   1815   1646    89    555
## 1956  60423  29990 4708   2568   2366  1411    733
## 1957  64721  32510 5230   2695   2526  1546    773
## 1958  68484  35218 6662   2845   2691  1663    836
## 1959  71799  37598 6856   3000   2868  1769    911
## 1960  76036  40341 8220   3145   3054  1905   1008
## 1961  79831  43173 9053   3338   3224  2005   1076
(phones_d <- as.data.frame(WorldPhones))
##      N.Amer Europe Asia S.Amer Oceania Africa Mid.Amer
## 1951  45939  21574 2876   1815   1646    89    555
## 1956  60423  29990 4708   2568   2366  1411    733
## 1957  64721  32510 5230   2695   2526  1546    773
## 1958  68484  35218 6662   2845   2691  1663    836
## 1959  71799  37598 6856   3000   2868  1769    911
## 1960  76036  40341 8220   3145   3054  1905   1008
## 1961  79831  43173 9053   3338   3224  2005   1076
is.data.frame(phones_d)
## [1] TRUE
write.dta(phones_d, file = file.path(tempdir(), "phones.dta"))
```

```
data.restore      Read an S3 Binary or data.dump File
lookup.xport      Lookup Information on a SAS XPORT Format
                  Library
read.arff          Read Data from ARFF Files
read.dbf           Read a DBF File
read.dta           Read Stata Binary Files
read.epiinfo       Read Epi Info Data Files
```

<code>read.mtp</code>	Read a Minitab Portable Worksheet
<code>read.octave</code>	Read Octave Text Data Files
<code>read.spss</code>	Read an SPSS Data File
<code>read.ssd</code>	Obtain a Data Frame from a SAS Permanent Dataset, via <code>read.xport</code>
<code>read.systat</code>	Obtain a Data Frame from a Systat File
<code>read.xport</code>	Read a SAS XPORT Format Library
<code>write.arff</code>	Write Data into ARFF Files
<code>write.dbf</code>	Write a DBF File
<code>write.dta</code>	Write Files in Stata Binary Format
<code>write.foreign</code>	Write Text Files and Code to Read Them

```

hsb2.small <-
read.csv("http://www.ats.ucla.edu/stat/data/hsb2_small.csv")

# using the names function to see names of the variables and which
column
# of data to which they correspond
names(hsb2.small)
## [1] "id"      "female"  "race"    "ses"     "schtyp"  "prog"
"read"
## [8] "write"   "math"    "science" "socst"
(hsb3 <- hsb2.small[, c(1, 7, 8)])
##      id read write
## 1    70   57    52
## 2   121   68    59
## 3    86   44    33
## 4   141   63    44
## 5   172   47    52
## 6   113   44    52
## 7    50   50    59
## 8    11   34    46
## 9    84   63    57
## 10   48   57    55
## 11   75   60    46
## 12   60   57    65
## 13   95   73    60
## 14  104   54    63
## 15   38   45    57
## 16  115   42    49
## 17   76   47    52
## 18  195   57    57
## 19  114   68    65
## 20   85   55    39
## 21  167   63    49
## 22  143   63    63
## 23   41   50    40
## 24   20   60    52
## 25   12   37    44

(hsb4 <- hsb2.small[, 1:4])
##      id female race ses
## 1    70      0    4    1
## 2   121      1    4    2
## 3    86      0    4    3
## 4   141      0    4    3
## 5   172      0    4    2
## 6   113      0    4    2
## 7    50      0    3    2
## 8    11      0    1    2
## 9    84      0    4    2
## 10   48      0    3    2
## 11   75      0    4    2
## 12   60      0    4    2

```

```
## 13 95      0      4      3
## 14 104     0      4      3
## 15 38      0      3      1
## 16 115     0      4      1
## 17 76      0      4      3
## 18 195     0      4      2
## 19 114     0      4      3
## 20 85      0      4      2
## 21 167     0      4      2
## 22 143     0      4      2
## 23 41      0      3      2
## 24 20      0      1      3
## 25 12      0      1      2
```

```
(hsb5 <- hsb2.small[1:10, ])
##      id female race ses schtyp prog read write math science socst
## 1   70      0      4      1      1      1  57   52  41      47   57
## 2  121      1      4      2      1      3  68   59  53      63   61
## 3   86      0      4      3      1      1  44   33  54      58   31
## 4  141      0      4      3      1      3  63   44  47      53   56
## 5  172      0      4      2      1      2  47   52  57      53   61
## 6  113      0      4      2      1      2  44   52  51      63   61
## 7   50      0      3      2      1      1  50   59  42      53   61
## 8   11      0      1      2      1      2  34   46  45      39   36
## 9   84      0      4      2      1      1  63   57  54      58   51
## 10  48      0      3      2      1      2  57   55  52      50   51
```

```
(hsb6 <- hsb2.small[hsb2.small$ses == 1, ])
##      id female race ses schtyp prog read write math science socst
## 1   70      0      4      1      1      1  57   52  41      47   57
## 15  38      0      3      1      1      2  45   57  50      31   56
## 16 115      0      4      1      1      1  42   49  43      50   56
```

```
(hsb7 <- hsb2.small[hsb2.small$id %in% c(12, 48, 86, 11, 20, 195),
])
##      id female race ses schtyp prog read write math science socst
## 3   86      0      4      3      1      1  44   33  54      58   31
## 8   11      0      1      2      1      2  34   46  45      39   36
## 10  48      0      3      2      1      2  57   55  52      50   51
## 18 195     0      4      2      2      1  57   57  60      58   56
## 24  20      0      1      3      1      2  60   52  57      61   61
## 25  12      0      1      2      1      3  37   44  45      39   46
```

```
(hsb8 <- hsb2.small[with(hsb2.small, ses == 3 & female == 0), ])
##      id female race ses schtyp prog read write math science socst
## 3   86      0      4      3      1      1  44   33  54      58   31
## 4  141      0      4      3      1      3  63   44  47      53   56
## 13  95      0      4      3      1      2  73   60  71      61   71
## 14 104      0      4      3      1      2  54   63  57      55   46
## 17  76      0      4      3      1      2  47   52  51      50   56
## 19 114      0      4      3      1      2  68   65  62      55   61
## 24  20      0      1      3      1      2  60   52  57      61   61
```

```
(write.50 <- subset(hsb2.small, write > 50))
##      id female race ses schtyp prog read write math science socst
## 1    70      0   4   1     1     1   57   52   41      47    57
## 2   121      1   4   2     1     3   68   59   53      63    61
## 5   172      0   4   2     1     2   47   52   57      53    61
## 6   113      0   4   2     1     2   44   52   51      63    61
## 7    50      0   3   2     1     1   50   59   42      53    61
## 9    84      0   4   2     1     1   63   57   54      58    51
## 10   48      0   3   2     1     2   57   55   52      50    51
## 12   60      0   4   2     1     2   57   65   51      63    61
## 13   95      0   4   3     1     2   73   60   71      61    71
## 14  104      0   4   3     1     2   54   63   57      55    46
## 15   38      0   3   1     1     2   45   57   50      31    56
## 17   76      0   4   3     1     2   47   52   51      50    56
## 18  195      0   4   2     2     1   57   57   60      58    56
## 19  114      0   4   3     1     2   68   65   62      55    61
## 22  143      0   4   2     1     3   63   63   75      72    66
## 24   20      0   1   3     1     2   60   52   57      61    61
```

```
(write.1 <- subset(hsb2.small, write > 50 & read > 60))
##      id female race ses schtyp prog read write math science socst
## 2   121      1   4   2     1     3   68   59   53      63    61
## 9    84      0   4   2     1     1   63   57   54      58    51
## 13   95      0   4   3     1     2   73   60   71      61    71
## 19  114      0   4   3     1     2   68   65   62      55    61
## 22  143      0   4   2     1     3   63   63   75      72    66
```

```
(write.2 <- subset(hsb2.small, write > 50 & read > 60, select =
c(write,
  read)))
##      write read
## 2       59   68
## 9       57   63
## 13      60   73
## 19      65   68
## 22      63   63
```

```
(write.3 <- subset(hsb2.small, science < 55, select = read:science))
##      read write math science
## 1       57   52   41      47
## 4       63   44   47      53
## 5       47   52   57      53
## 7       50   59   42      53
## 8       34   46   45      39
## 10      57   55   52      50
## 11      60   46   51      53
## 15      45   57   50      31
## 16      42   49   43      50
```

```
## 17 47 52 51 50
## 20 55 39 57 53
## 25 37 44 45 39
```

3. Subsetting both variables and observations

```
# using the names function to see names of the variables and which
# column
# of data to which they correspond
names(hsb2.small)
## [1] "id"      "female"  "race"    "ses"     "schtyp"  "prog"
"read"
## [8] "write"   "math"    "science" "socst"
(hsb9 <- hsb2.small[hsb2.small$ses == 3, c(1:4, 7)])
##      id female race ses read
## 3    86      0     4   3   44
## 4   141      0     4   3   63
## 13   95      0     4   3   73
## 14  104      0     4   3   54
## 17   76      0     4   3   47
## 19  114      0     4   3   68
## 24   20      0     1   3   60
```

الإحتمالات والتوزيعات Probabilities and Distributions

توليد عينات عشوائية من التوزيع الطبيعي
Generating random samples from a normal distribution

```
norm <- rnorm(100)
norm[1:10]

[1] -1.38507062 0.03832318 -0.76303016 0.21230614 1.42553797 0.74447982
0.70022940 -0.22935461
[9] 0.19709386 1.20715377

mean(norm)

[1] 0.009620551

sd(norm)

[1] 0.883841

set.seed(124)

norm <- rnorm(100, 2, 5)

norm[1:10]

[1] -4.925353 2.191616 -1.815151 3.061531 9.127690 5.722399 5.501147
0.853227 2.985469 8.035769
```

```
mean(norm)
```

```
[1] 2.048103
```

```
sd(norm)
```

```
[1] 4.419205
```

تولید عینات عشوائیة من توزيعات اخرى
Generating random samples from other distributions

```
#Generating a random sample from a Poisson distribution with lambda=3  
set.seed(124)
```

```
pois <- rpois(100, lambda=3)
```

```
pois[1:10]
```

```
[1] 1 2 3 2 2 2 3 3 6 2
```

```
mean(pois)
```

```
[1] 2.83
```

```
var(pois)
```

```
[1] 2.344545
```

```
#Generating a random sample from a Binomial distribution with size=20  
and prob=.2  
set.seed(124)
```

```
binom <- rbinom(100, 20, .2)
```

```
binom[1:10]
```

```
[1] 2 3 4 3 3 3 4 4 7 3
```

```
mean(binom)
```

```
[1] 3.85
```

```
sd(binom)
```

```
[1] 1.604130
```

احتمالات ودوال توزيع اخرى
Other probability and distribution functions

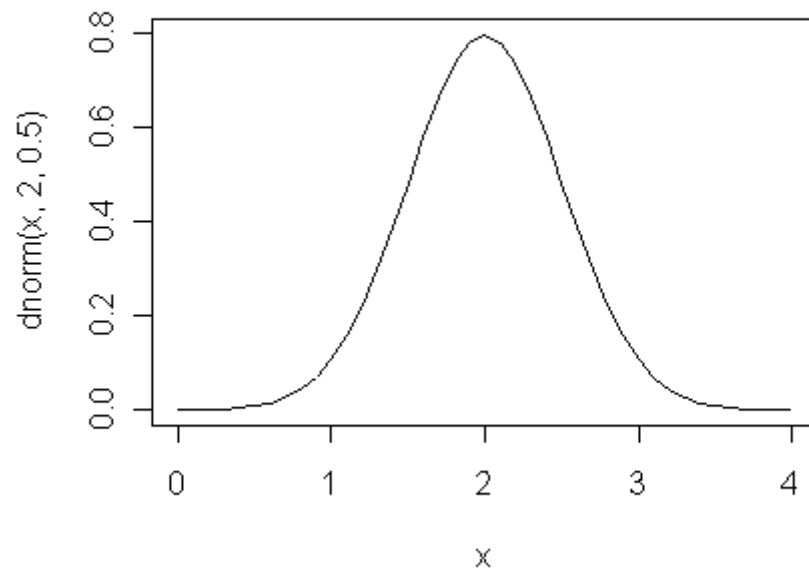
```
#point probability for a specific value of a standard normal dist  
dnorm(-1.96)
```

```
[1] 0.05844094
```



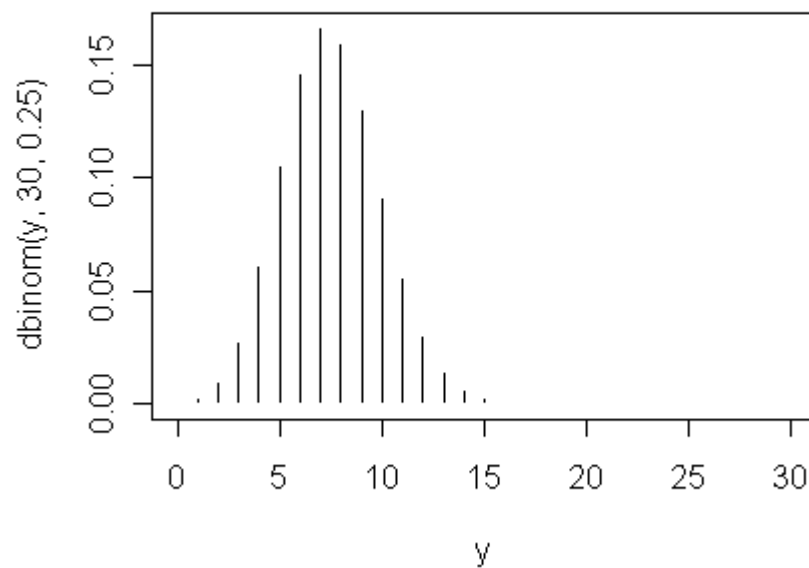
```
#plotting the density function of a normal distribution: N(2, .25)
x <- seq(0, 4, 0.1)

plot(x, dnorm(x, 2, .5), type="l")
```



```
#plotting the density function of a binomial distribution: Binom(30,
.25)
y <- 0:30

plot(y, dbinom(y, 30, .25), type="h")
```



```
#calculating the p-values for the quantiles of a standard normal
1 - pnorm(1.959964)
```

```
[1] 0.025
```

```
1 - pnorm(1.644854)
```

```
[1] 0.04999996
```

```
#calculating the quantiles for the standard normal
qnorm(0.05)
```

```
[1] -1.644854
```

```
qnorm(0.025)
```

```
[1] -1.959964
```

The sample function دالة المعاينة

```
# random sample of size 8 from sequence [5, 15]
set.seed(124)
```

```
sample(seq(5:15), 8)
```

```
[1] 1 5 10 4 2 7 3 6
```

```
# random permutation of sequence [1, 10]
set.seed(124)
```

```
sample(10)
```

```

[1] 8 2 5 1 6 10 7 9 4 3

# random sample of size 10 from sequence [1, 5] with
# unequal probabilities of being chosen
set.seed(124)

sample(5, 10, prob = c(0.3, 0.4, 0.1, 0.1, 0.1), replace = T)

[1] 2 1 1 2 2 2 1 1 4 2

```

إستكشاف البيانات بالرسوم
سوف نستخدم البيانات [hsb2.csv](#)

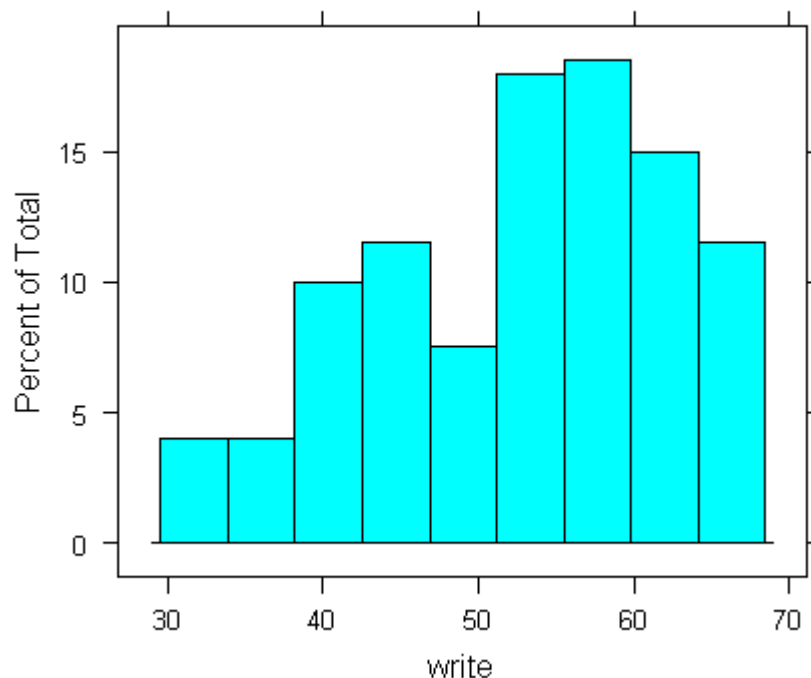
رسومات المتغير الواحد Univariate Plots

```

hsb2 <- read.table('http://www.ats.ucla.edu/stat/r/modules/hsb2.csv',
header=T, sep=",")
attach(hsb2)
library(lattice)
#defining ses.f to be a factor variable
hsb2$ses.f = factor(hsb2$ses, labels=c("low", "middle", "high"))

#histograms
histogram(~write, hsb2)

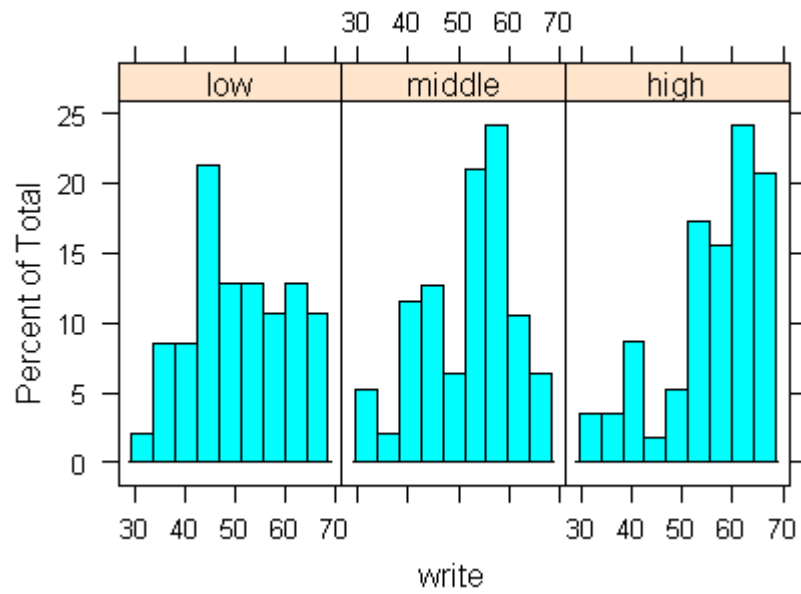
```



```

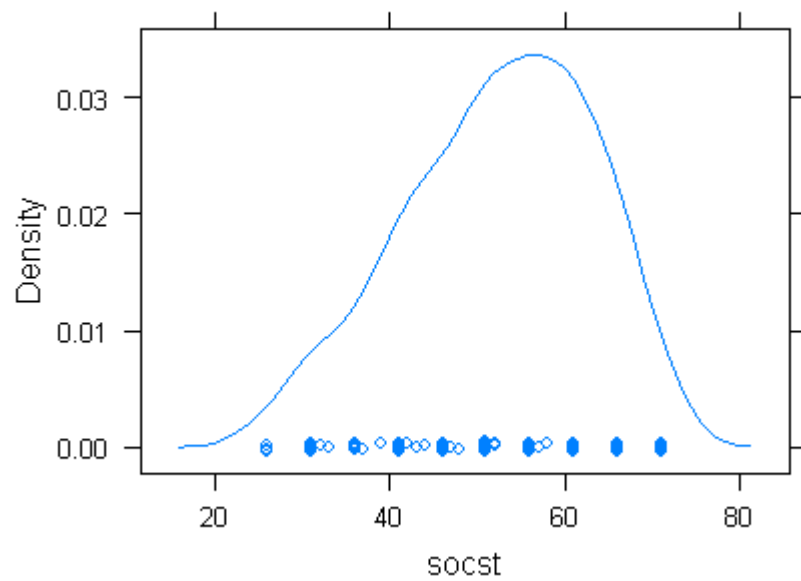
#conditional plot
histogram(~write | ses.f, hsb2)

```

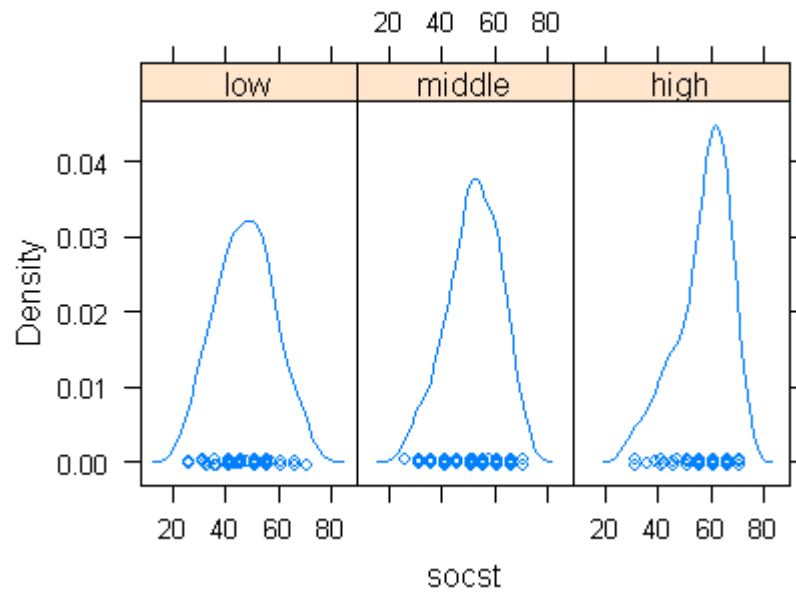


Density plots رسومات الكثافة

`densityplot(~socst, hsb2)`

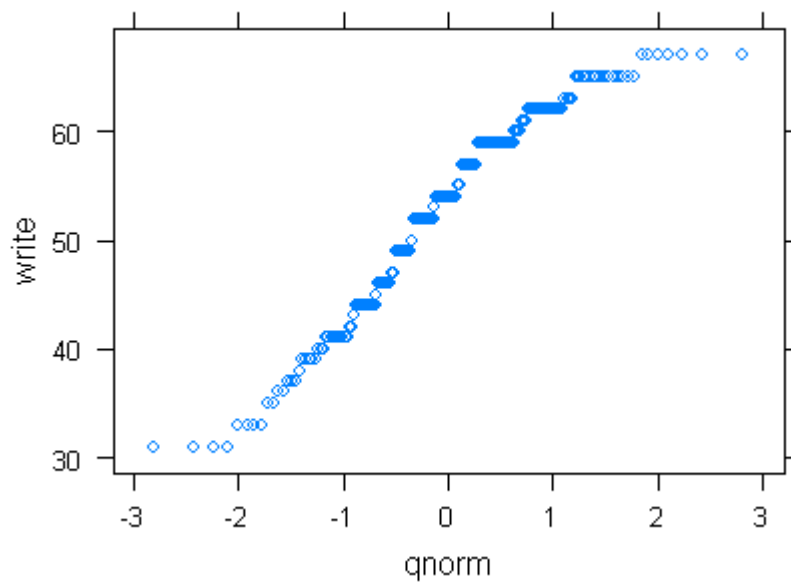


`#conditional plot`
`densityplot(~socst | ses.f, hsb2)`



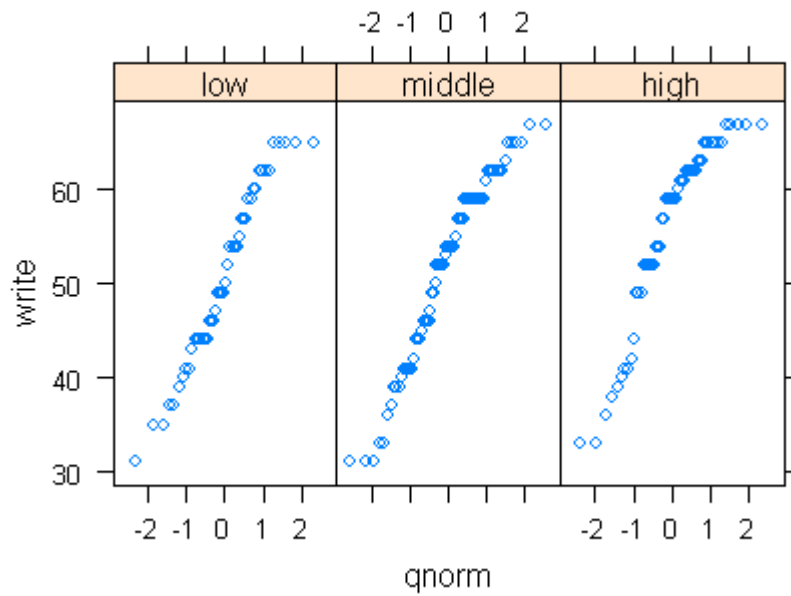
رسومات الربعيات - الربعيات Quantile-quantile plots

```
qqmath(~write, hsb2)
```



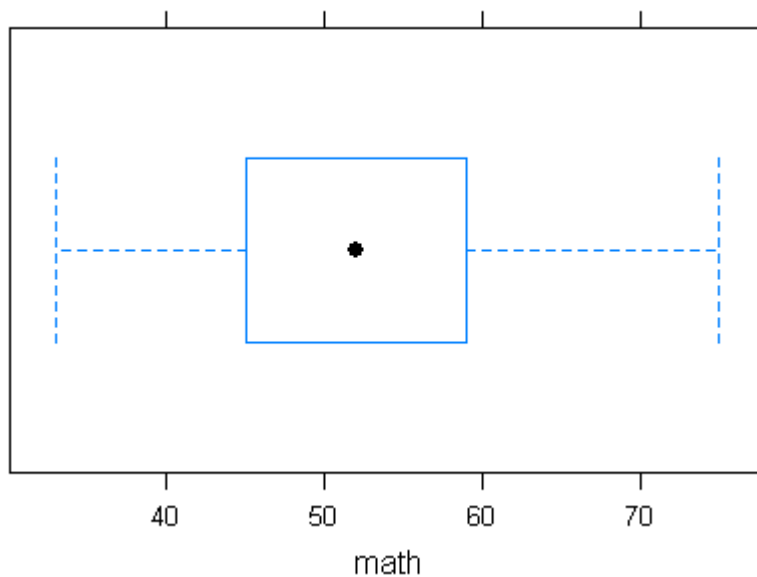
```
#conditional plot
```

```
qqmath(~write | ses.f, hsb2)
```

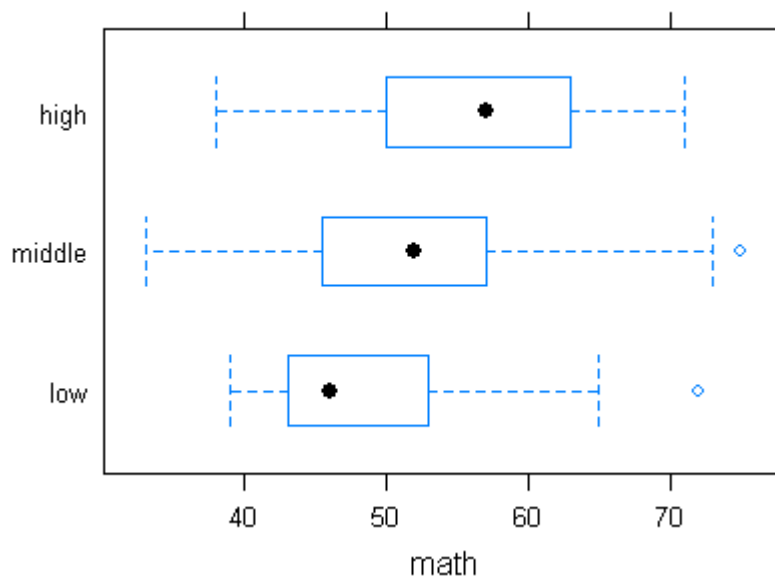


رسومات الصندوق و السبلة Box and whiskers plots

`bwplot(~math, hsb2)`

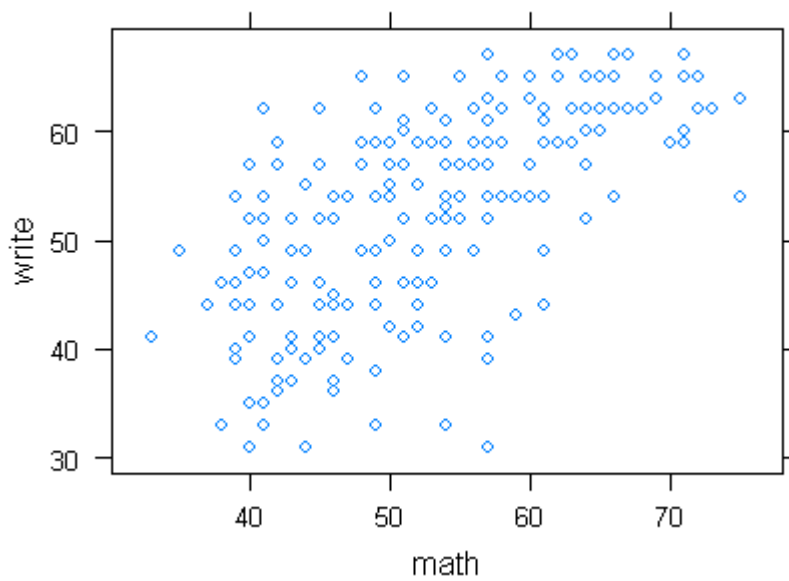


`#conditional plot`
`bwplot(ses.f~math, hsb2)`

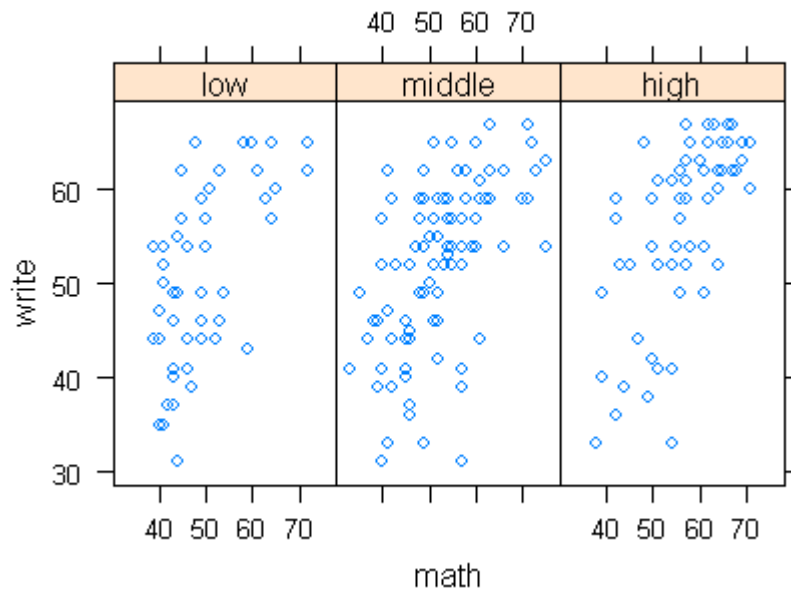


Multivariate Plots رسومات متعددة المتغيرات
Scatter plots رسوم الانتشار

`xyplot(write~math, hsb2)`

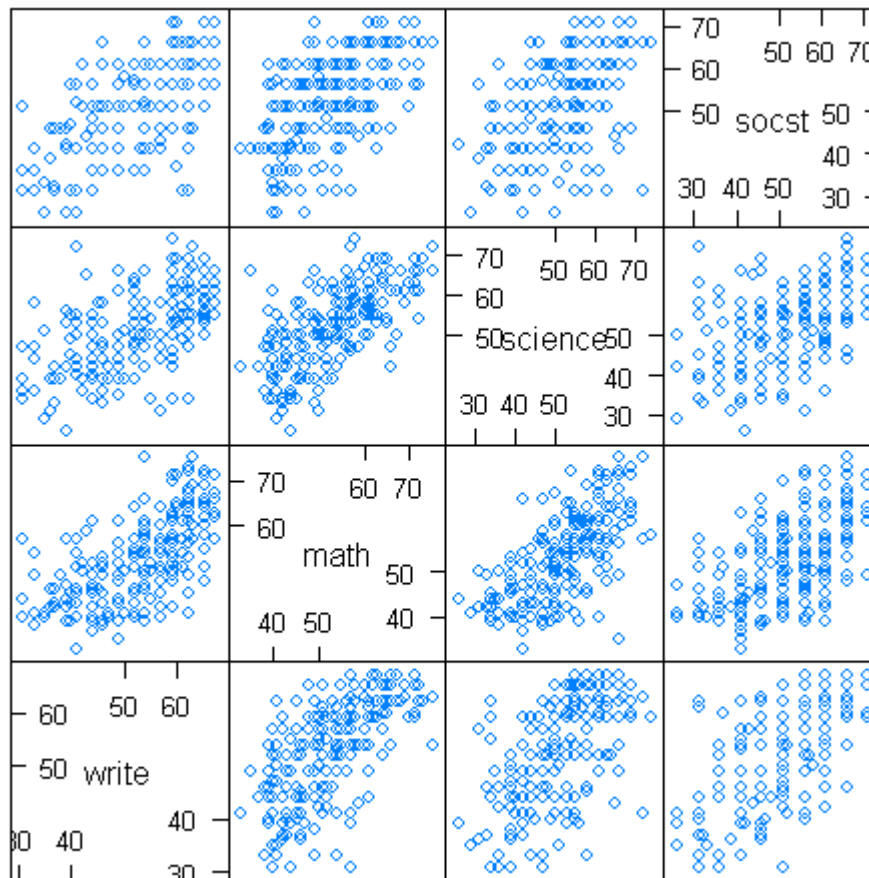


`#conditional plot`
`xyplot(write~math | ses.f, hsb2)`



رسم مصفوفة الانتشار Scatter plot matrices

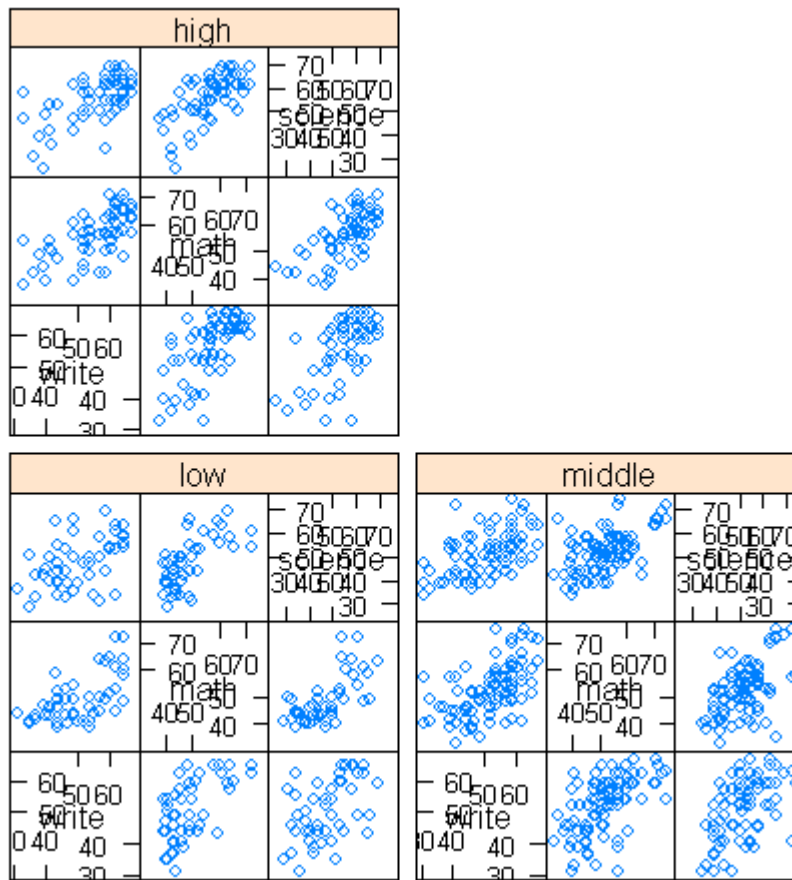
```
subset <- hsb2[ , 8:12]
splom(~subset[ , 1:4])
```



Scatter Plot Matrix

#conditional plot

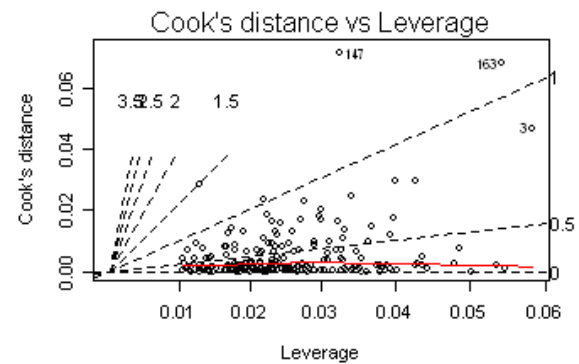
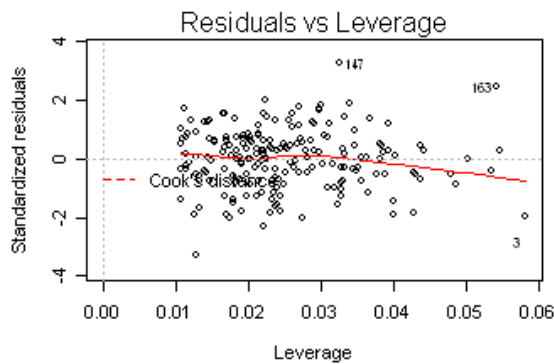
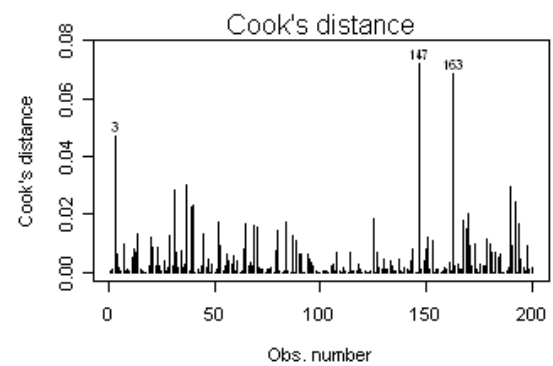
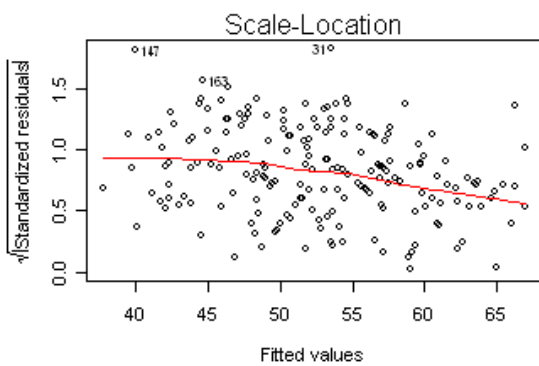
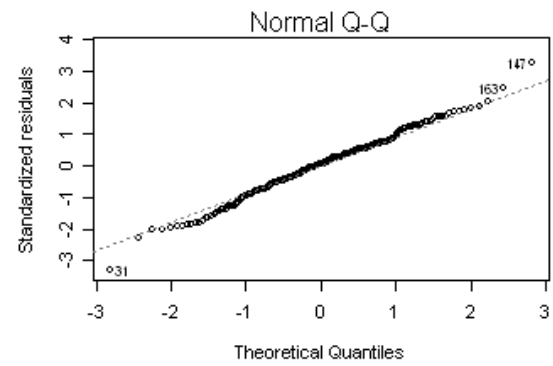
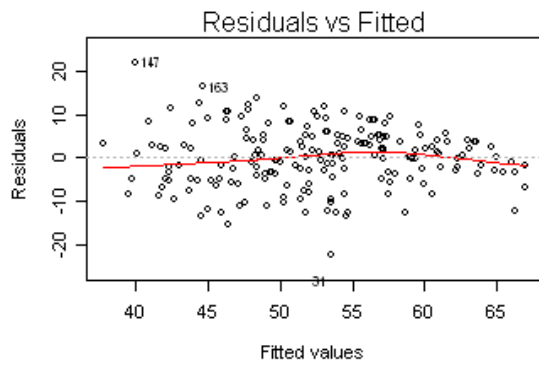

```
splom(~subset[, 1:3] | subset[, 5])
```



Scatter Plot Matrix

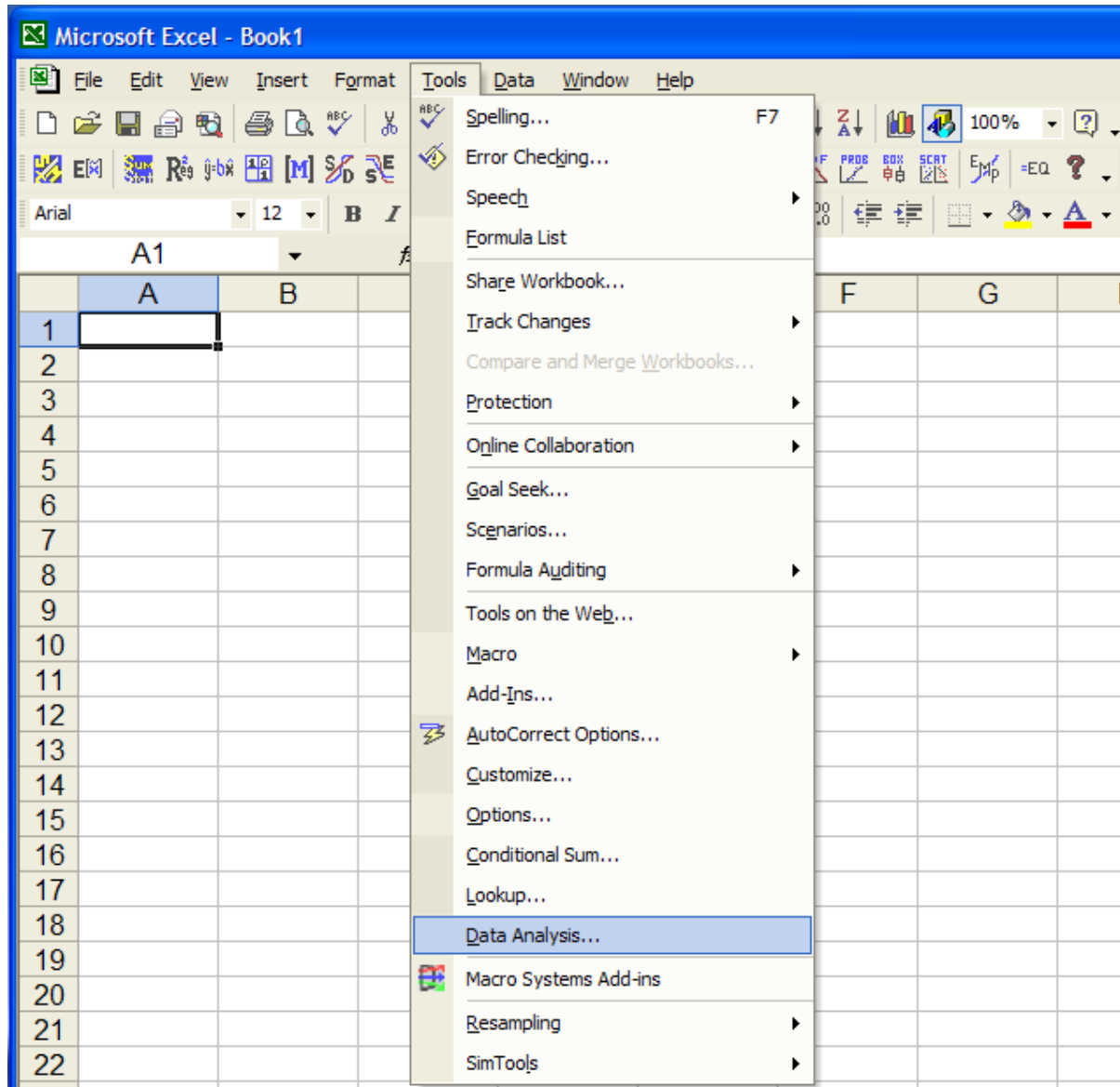
امثلة على الرسومات المستخدمة في التحليل الإحصائي
Examples of Plots Used in Statistical Analysis

```
reg <- lm(write~math+socst+ses.f, hsb2)
par(mfrow=c(3,2))
plot(reg, which=1:2)
plot(reg, which=3:4)
plot(reg, which=5:6)
```

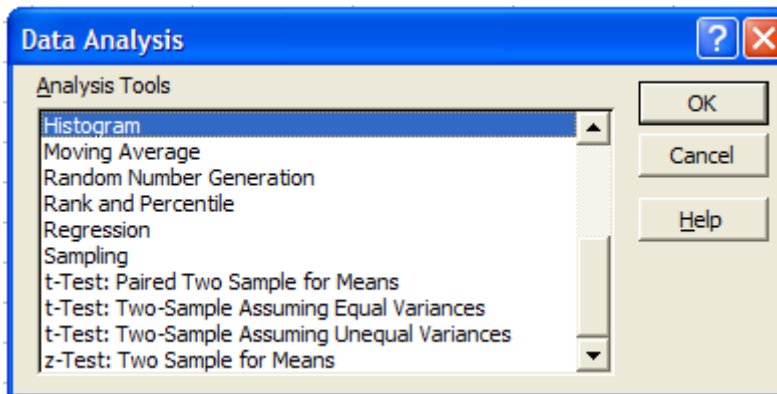
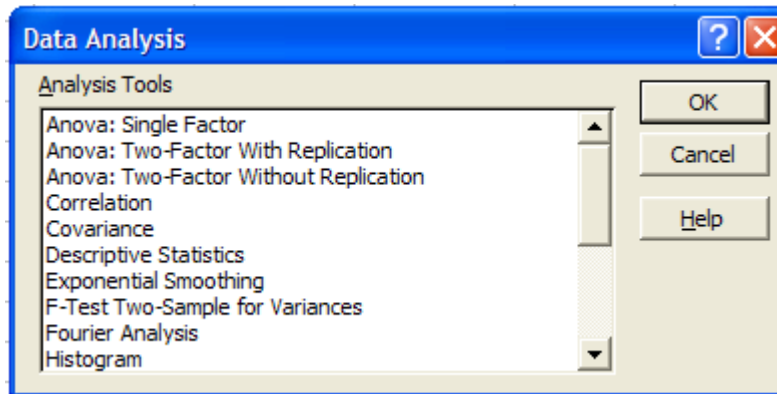


detach(hsb2)

مقدمة للدوال التحليل الإحصائي في Excel:
إستخدام تحليل البيانات المبني في إكسل
يوجد في إكسل إختيار ضمن قائمة الأدوات لتحليل البيانات



والذي يحوي التالي:



- 1- تحليل التباين لعامل واحد
- 2- تحليل التباين لعاملين مع تكرار
- 3- تحليل التباين لعاملين بدون تكرار
- 4- الترابط
- 5- التغيرات
- 6- إحصائيات وصفية
- 7- التمهيد الاسي
- 8- إختبار F للتباين لعينتين
- 9- تحليل فورييه
- 10- المدرج التكراري
- 11- المتوسط المتحرك

- 12- توليد ارقام عشوائية
 - 13- الرتب والمئينات
 - 14- الإنحدار
 - 15- المعاينة
 - 16- إختبار t للمتوسطات لعينتين متقارنة
 - 17- إختبار t لعينتين على إفتراض تساوي التباين
 - 18- إختبار t لعينتين على إفتراض عدم تساوي التباين
 - 19- إختبار z للمتوسطات لعينتين
- وسوف نستعرض بعض هذه الطرق فيما يلي:

تحليل التباين لعامل واحد Anova: Single Factor

اجريت دراسة لمعرفة الفرق بين تأثير ثلاثة طرق لتدريس مبادئ الحساب لطلاب المرحلة الأولى الابتدائية فاختير 27 تلميذا عشوائيا وتم تخصيص 9 تلاميذ بطريقة عشوائية لكل طريقة من الطرق الثلاثة. تم اختبار جميع التلاميذ بعد فترة معينة وكانت نتائج الاختبارات كالتالي:

رقم الطالب	1	2	3	4	5	6	7	8	9	المجموع
الطريقة 1	4	5	4	3	6	10	1	8	5	46
الطريقة 2	12	8	10	5	7	9	14	9	4	78
الطريقة 3	1	3	4	6	8	5	3	2	2	34

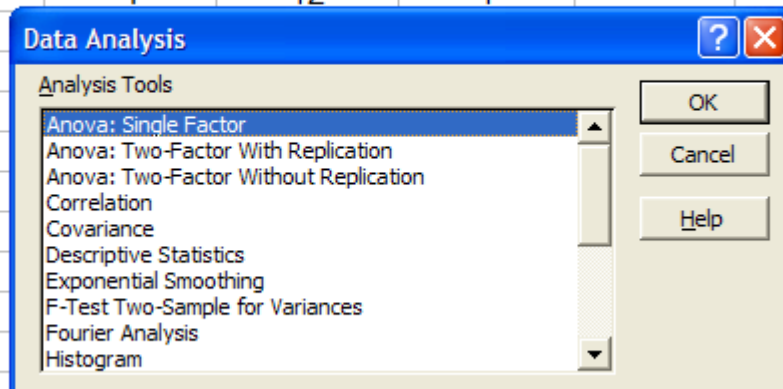
المطلوب معرفة هل هناك فرق معنوي بين طرق التدريس المختلفة. اختبر عند مستوى معنوية 0.05

ندخل البيانات في صفحة من إكسل كالتالي:

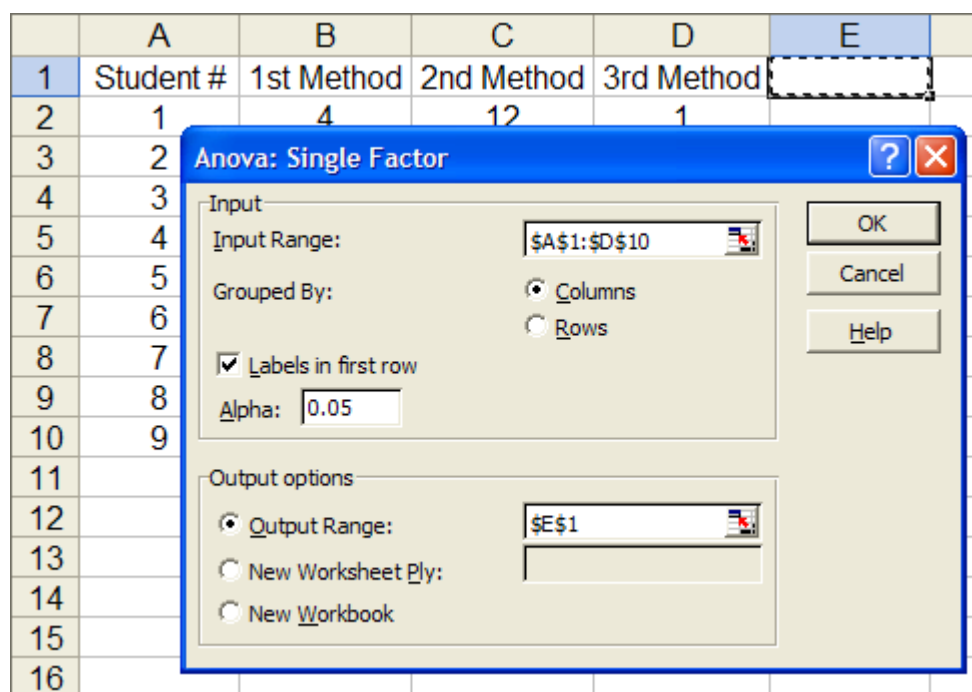
	A	B	C	D
1	Student #	1st Method	2nd Method	3rd Method
2	1	4	12	1
3	2	5	8	3
4	3	4	10	4
5	4	3	5	6
6	5	6	7	8
7	6	10	9	5
8	7	1	14	3
9	8	8	9	2
10	9	5	4	2

من قائمة الأدوات Tools نختار Data Analysis فتظهر النافذة

	A	B	C	D	E	F
1	Student #	1st Method	2nd Method	3rd Method		
2	1	4	12	1		
3	2					
4	3					
5	4					
6	5					
7	6					
8	7					
9	8					
10	9					
11						
12						
13						



نختار تحليل التباين لعامل واحد Anova: Single Factor فتظهر النافذة



ندخل البيانات المطلوبة ثم OK فينتج

E	F	G	H	I	J	K
Anova: Single Factor						
SUMMARY						
<i>Groups</i>	<i>Count</i>	<i>Sum</i>	<i>Average</i>	<i>Variance</i>		
Student #	9	45	5	7.5		
1st Method	9	46	5.111111111	7.111111111		
2nd Method	9	78	8.666666667	10		
3rd Method	9	34	3.777777778	4.944444444		
ANOVA						
<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
Between Groups	119.8611111	3	39.9537037	5.40726817	0.003997338	2.901117568
Within Groups	236.4444444	32	7.388888889			
Total	356.3055556	35				

تحليل التباين لعاملين مع تكرار Anova: Two-Factor With Replication
 قام احد الباحثين بتجربتين على مجموعتين لتحديد درجة الاستيعاب التي تقاس كجزء
 من 100 فتحصل على النتائج التالية

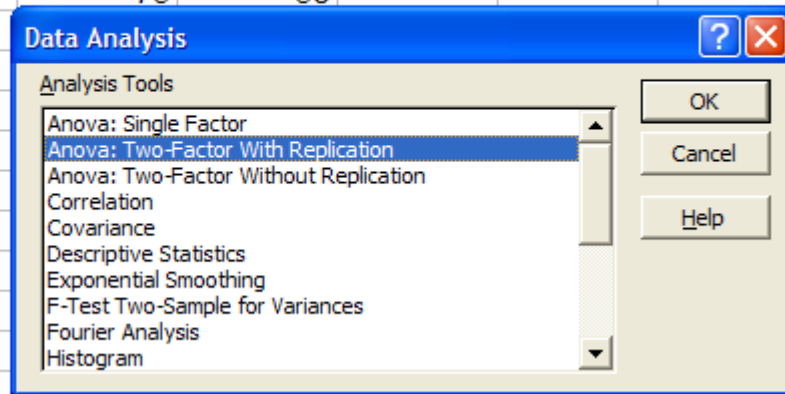
مجموعة 2	مجموعة 1	
58	75	تجربة 1
56	68	
61	71	
60	75	
62	66	تجربة 2
60	70	
59	68	
68	68	

هل هناك فرق بين التجارب وفرق بين المجموعات ؟ اختبر عند مستوى معنوية 0.05
 ندخل البيانات في صفحة من إكسل

	A	B	C
1		Group 1	Group 2
2	Trial 1	75	58
3		68	56
4		71	61
5		75	60
6	Trial 2	66	62
7		70	60
8		68	59
9		68	68

كالسابق من قائمة الأدوات وتحت تحليل البيانات نختار تحليل التباين لعاملين مع تكرار

	A	B	C	D	E	F
1		Group 1	Group 2			
2	Trial 1	75	58			
3						
4						
5						
6	Trial 2					
7						
8						
9						
10						
11						
12						



فتظهر النافذة

Anova: Two-Factor With Replication

Input

Input Range:

Rows per sample:

Alpha:

Output options

☒ Output Range:

☐ New Worksheet Ply:

☐ New Workbook

OK Cancel Help

ندخل البيانات كما هو موضح ثم OK فينتج

D	E	F	G
Anova: Two-Factor With Replication			
SUMMARY	Group 1	Group 2	Total
<i>Trial 1</i>			
Count	4	4	8
Sum	289	235	524
Average	72.25	58.75	65.5
Variance	11.58333333	4.916666667	59.14285714
<i>Trial 2</i>			
Count	4	4	8
Sum	272	249	521
Average	68	62.25	65.125
Variance	2.666666667	16.25	17.55357143
<i>Total</i>			
Count	8	8	
Sum	561	484	
Average	70.125	60.5	
Variance	11.26785714	12.57142857	

ANOVA		
<i>Source of Variation</i>	<i>SS</i>	<i>df</i>
Sample	0.5625	1
Columns	370.5625	1
Interaction	60.0625	1
Within	106.25	12
Total	537.4375	15

<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
0.5625	0.063529412	0.805267228	4.747221283
370.5625	41.85176471	3.07191E-05	4.747221283
60.0625	6.783529412	0.023033141	4.747221283
8.854166667			

Anova: Two-Factor Without Replication

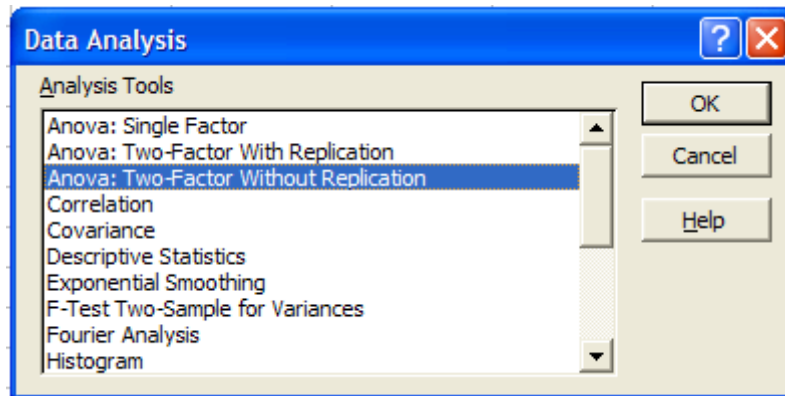
استخدم أحد الباحثين 4 انواع من السماد A,B,C,D لمعالجة 4 قطاعات من الأراضي قطاع 1 وحتى قطاع 4 فتحصل على الإنتاج التالي بالأطنان

Treatment	Sector 1	Sector 2	Sector 3	Sector 4
A	9.3	9.4	9.6	10
B	9.4	9.3	9.8	9.9
C	9.2	9.4	9.5	9.7
D	9.7	9.6	10	10.2

هل هناك فرق بين المعالجات؟ هل هناك فرق بين القطاعات؟ اختبر عند 0.05
تدخل البيانات كالتالي:

	A	B	C	D	E
1	Treatment	Sector 1	Sector 2	Sector 3	Sector 4
2	A	9.3	9.4	9.6	10
3	B	9.4	9.3	9.8	9.9
4	C	9.2	9.4	9.5	9.7
5	D	9.7	9.6	10	10.2
6					

كالسابق من قائمة الأدوات وتحت تحليل البيانات نختار تحليل التباين لعاملين بدون تكرار



ثم ندخل المطلوب كالتالي

	A	B	C	D	E	F
1	Treatment	Sector 1	Sector 2	Sector 3	Sector 4	
2	A	9.3	9.4	9.6	10	
3	B	9.4	9.3	9.8	9.9	
4	C	9.2	9.4	9.5	9.7	
5	D	9.7	9.6	10	10.2	
6						
7						
8						
9						
10						
11						
12						
13						
14						
15						
16						
17						

Anova: Two-Factor Without Replication

Input
 Input Range: 
☒ Labels
 Alpha:

Output options
☒ Output Range: 
☐ New Worksheet Ply:
☐ New Workbook

OK Cancel Help

فينتج

Anova: Two-Factor Without Replication				
SUMMARY	Count	Sum	Average	Variance
A	4	38.3	9.575	0.095833333
B	4	38.4	9.6	0.086666667
C	4	37.8	9.45	0.043333333
D	4	39.5	9.875	0.075833333
Sector 1	4	37.6	9.4	0.046666667
Sector 2	4	37.7	9.425	0.015833333
Sector 3	4	38.9	9.725	0.049166667
Sector 4	4	39.8	9.95	0.043333333

ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Rows	0.385	3	0.128333333	14.4375	0.000871272	3.862538733
Columns	0.825	3	0.275	30.9375	4.52327E-05	3.862538733
Error	0.08	9	0.008888889			
Total	1.29	15				

الترابط Correlation

الجدول التالي يوضح السن X وضغط الدم Y لثمان من الإناث :

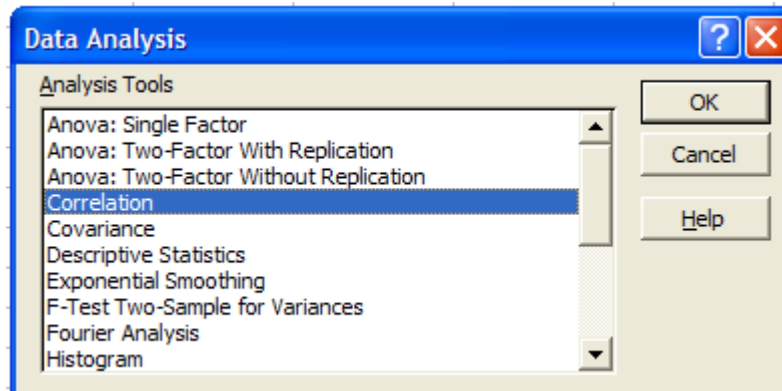
السن X	42	36	63	55	42	60	49	68
ضغط الدم Y	125	118	140	150	140	155	145	152

أوجد معامل الارتباط بين X و Y .

تدخل البيانات في صفحة من إكسل

	A	B
1	X	Y
2	42	125
3	36	118
4	63	140
5	55	150
6	42	140
7	60	155
8	49	145
9	68	152

كالسابق من قائمة الأدوات وتحت تحليل البيانات نختار الترابط



ثم OK فتظهر النافذة

	A	B	C	D	E	F	G
1	X	Y					
2	42	125					
3	36	118					
4	63	140					
5	55	150					
6	42	140					
7	60	155					
8	49	145					
9	68	152					
10							
11							
12							
13							

Correlation

Input

Input Range:

Grouped By: ☒ Columns ☐ Rows

☒ Labels in first row

Output options

☒ Output Range:

☐ New Worksheet Ply:

☐ New Workbook

OK Cancel Help

ندخل المطلوب كما في الشكل فينتج

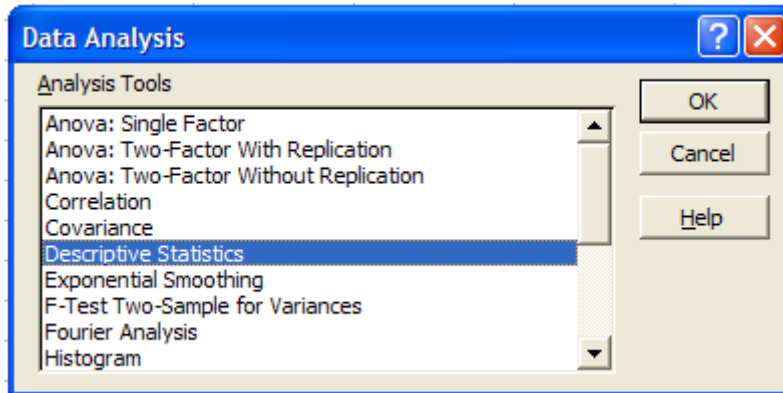
	A	B	C	D	E
1	X	Y		X	Y
2	42	125	X	1	
3	36	118	Y	0.791832	1
4	63	140			
5	55	150			
6	42	140			
7	60	155			
8	49	145			
9	68	152			

الإحصاءات الوصفية Descriptive Statistica

البيانات التالية هي الدخل الشهري بالريال (لأقرب هللة) لعينة من 50 متخرجاً من جامعة الملك سعود (للكتليات غير الطبية)

4932.40, 2625.58, 6691.17, 9172.67, 9053.80, 9659.41, 1918.87,
5140.86, 8878.62, 2936.39, 3809.27, 2172.88, 2065.52, 3145.85,
3600.81, 1940.14, 4137.35, 4613.33, 6339.82, 4730.45, 4849.07,
4715.93, 9264.51, 5621.34, 5294.52, 4292.01, 9800.80, 8414.65,
9928.18, 3901.36 9603.85, 2238.19, 7581.32, 8495.49, 9774.52,
5623.85, 4261.73, 7951.69, 4682.15, 8160.40, 2409.61, 3427.14,
2325.28, 4738.46, 5793.77, 5991.97, 4862.33, 9884.38, 2133.84,
3691.90

بإختيار إحصاءات وصفية من قائمة الإختيارات



تظهر النافذة

	A	B	C	D	E	F
1	Income					
2	4932.40					
3	2625.58					
4	6691.17					
5	9172.67					
6	9053.80					
7	9659.41					
8	1918.87					
9	5140.86					
10	8878.62					
11	2936.39					
12	3809.27					
13	2172.88					
14	2065.52					
15	3145.85					
16	3600.81					
17	1940.14					
18	4137.35					
19	4613.33					
20	6339.82					

Descriptive Statistics

Input

Input Range:

Grouped By: ☒ Columns ☐ Rows

☒ Labels in first row

Output options

☒ Output Range:

☐ New Worksheet Ply:

☐ New Workbook

☒ Summary statistics

☒ Confidence Level for Mean: %

☐ Kth Largest:

☐ Kth Smallest:

OK Cancel Help

هنا اخترنا جميع الإحصائيات الملخصة وكذلك فترة 95% للثقة بالضغط على OK

ينتج

	A	B	C
1	Income	Income	
2	4932.40		
3	2625.58	Mean	5545.59
4	6691.17	Standard Error	369.81
5	9172.67	Median	4855.70
6	9053.80	Mode	#N/A
7	9659.41	Standard Deviation	2614.93
8	1918.87	Sample Variance	6837855.24
9	5140.86	Kurtosis	-1.16
10	8878.62	Skewness	0.37
11	2936.39	Range	8009.31
12	3809.27	Minimum	1918.87
13	2172.88	Maximum	9928.18
14	2065.52	Sum	277279.42
15	3145.85	Count	50.00
16	3600.81	Confidence Level(95.0%)	743.15
17	1940.14		

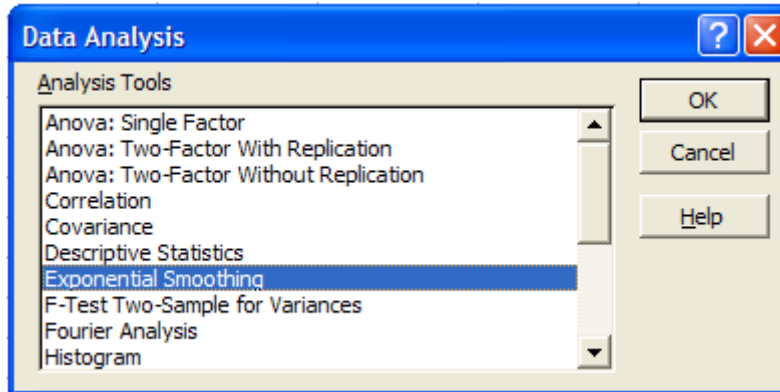
التمهيد الاسي Exponential Smoothing

يستخدم التمهيد الاسي للتنبؤ عن القيمة المستقبلية التالية في سلسلة من المشاهدات لمتغير عشوائي معطى. التمهيد الاسي هو احد الطرق المستخدمة في التنبؤ الإحصائي (انظر كتاب: طرق التنبؤ الإحصائي – الجزء الأول- تأليف: د. عدنان ماجد بري)

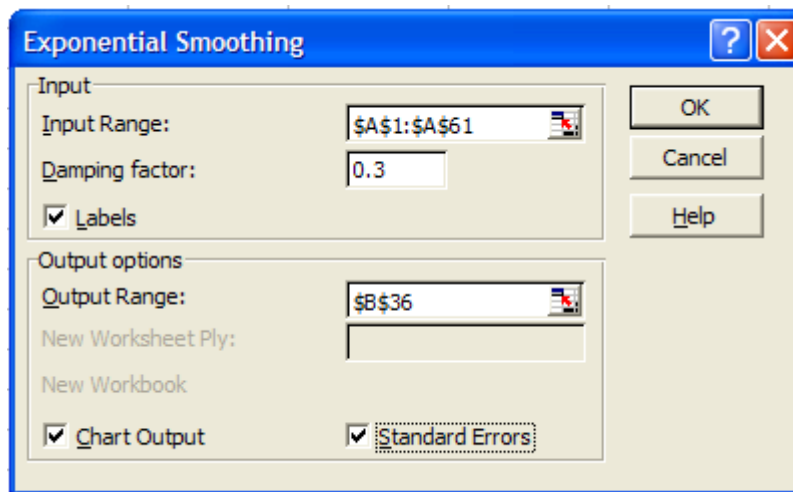
البيانات التالية لظاهرة عشوائية

44.2	44.3	44.4	43.4	42.8	44.3
44.4	44.8	44.4	43.1	42.6	42.4
42.2	41.8	40.1	42.0	42.4	43.1
42.4	43.1	43.2	42.8	43.0	42.8
42.5	42.6	42.3	42.9	43.6	44.7
44.5	45.0	44.8	44.9	45.2	45.2
45.0	45.5	46.2	46.8	47.5	48.3
48.3	49.1	48.9	49.4	50.0	50.0
49.6	49.9	49.6	50.7	50.7	50.9
50.5	51.2	50.7	50.3	49.2	48.1

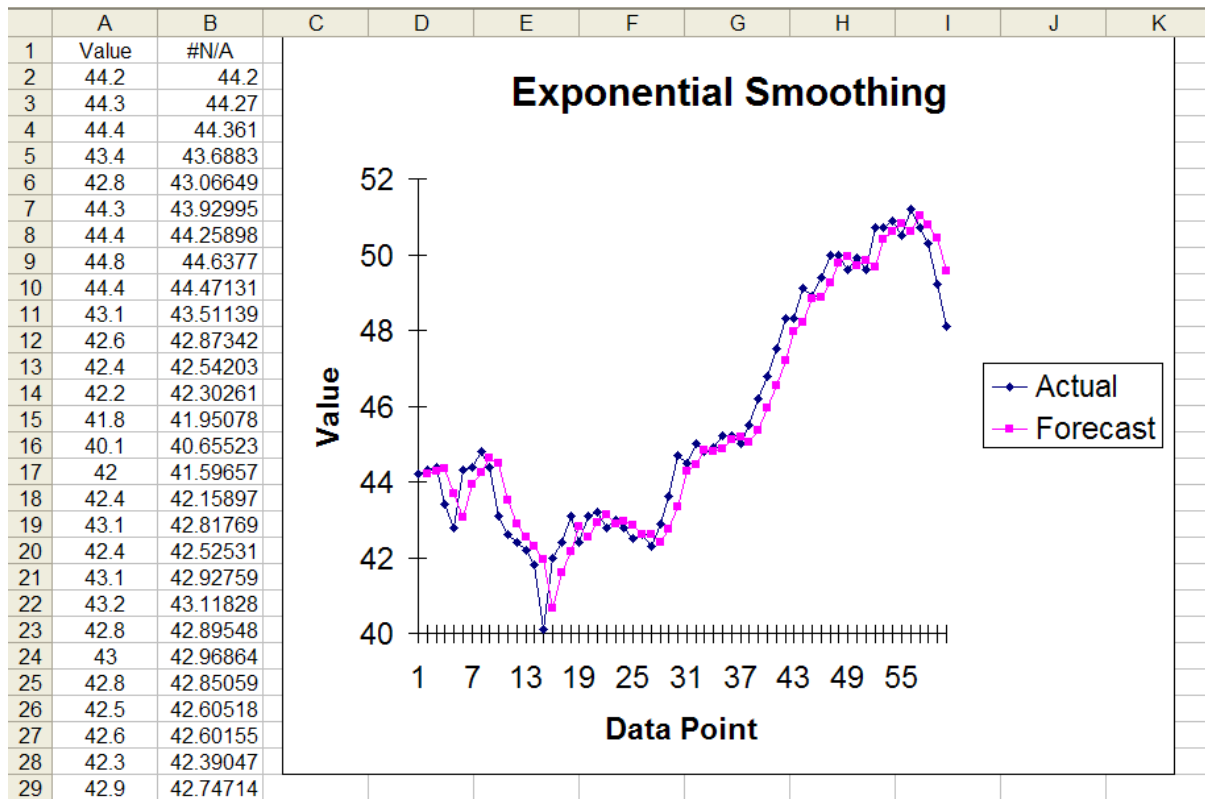
سوف نستخدم التمهيد الاسي (البسيط) علي البيانات



فتظهر نافذة الإدخال



وينتج



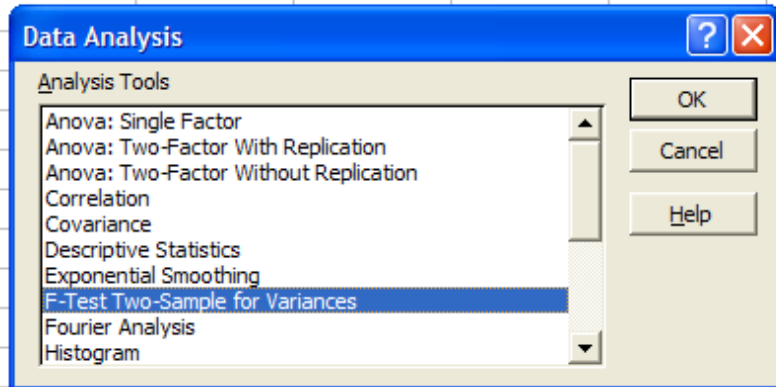
إختبار F للتباين لعينتين:

نريد ان نختبر تساوي تباين العينتين التالية:

عينة 1: 70 59 69 81 66 61 72 90

عينة 2: 84 77 69 91 80 66 78 85 61

	A	B	C	D	E	F	G
1	Sample 1	Sample 2					
2	90	62					
3	72	85					
4	61	78					
5	66	66					
6	81	80					
7	69	91					
8	59	69					
9	70	77					
10		84					



وندخل البيانات كالتالي

F-Test Two-Sample for Variances

Input

Variable 1 Range: \$A\$1:\$A\$9

Variable 2 Range: \$B\$1:\$B\$10

☒ Labels

Alpha: 0.05

Output options

☒ Output Range: \$C\$1

☐ New Worksheet Ply:

☐ New Workbook

OK Cancel Help

فينتج

C	D	E
F-Test Two-Sample for Variances		
	<i>Sample 1</i>	<i>Sample 2</i>
Mean	71	76.88888889
Variance	105.1428571	91.11111111
Observations	8	9
df	7	8
F	1.154006969	
P(F<=f) one-tail	0.418377164	
F Critical one-tail	3.500460366	

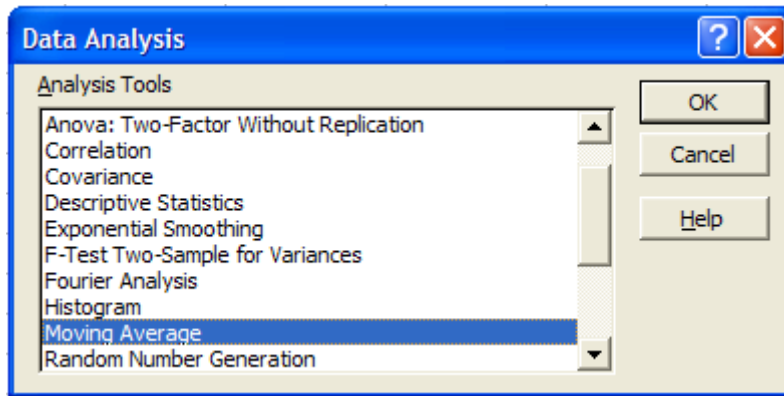
Moving Average المتوسط المتحرك

يستخدم المتوسط المتحرك مثل التمهيد الاسي للتنبؤ عن القيمة المستقبلية التالية في سلسلة من المشاهدات لمتغير عشوائي معطى. المتوسط المتحرك هو احد الطرق المستخدمة في التنبؤ الإحصائي (انظر كتاب: طرق التنبؤ الإحصائي – الجزء الأول - تأليف: د. عدنان ماجد بري)

البيانات التالية لظاهرة عشوائية

44.2	44.3	44.4	43.4	42.8	44.3	44.4	44.8	44.4	43.1
42.6	42.4	42.2	41.8	40.1	42.0	42.4	43.1	42.4	43.1
43.2	42.8	43.0	42.8	42.5	42.6	42.3	42.9	43.6	44.7
44.5	45.0	44.8	44.9	45.2	45.2	45.0	45.5	46.2	46.8
47.5	48.3	48.3	49.1	48.9	49.4	50.0	50.0	49.6	49.9
49.6	50.7	50.7	50.9	50.5	51.2	50.7	50.3	49.2	48.1

سوف نستخدم المتوسط المتحرك عليها



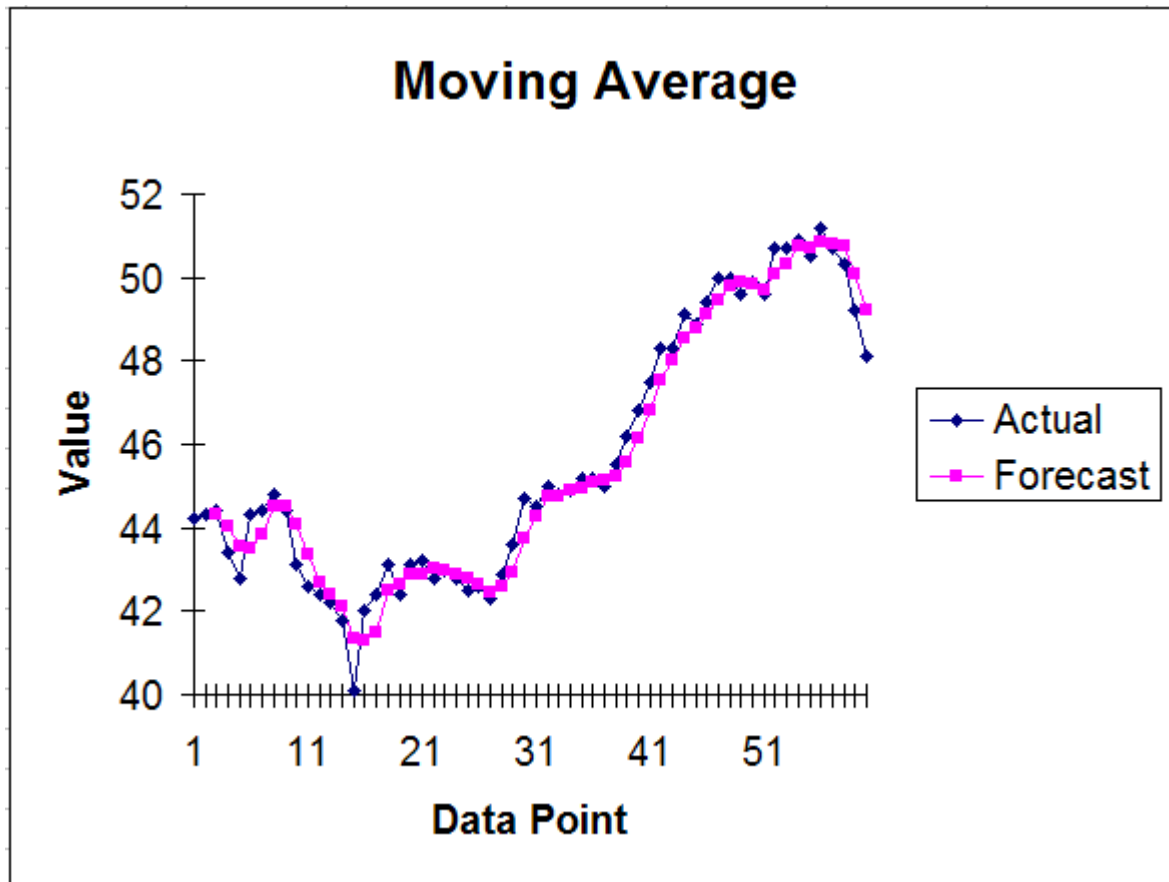
فتظهر نافذة الإدخال

Moving Average

Input
 Input Range: 
☒ Labels in First Row
 Interval:
 OK Cancel Help

Output options
 Output Range: 
 New Worksheet Ply:
 New Workbook
☒ Chart Output ☐ Standard Errors

وينتج

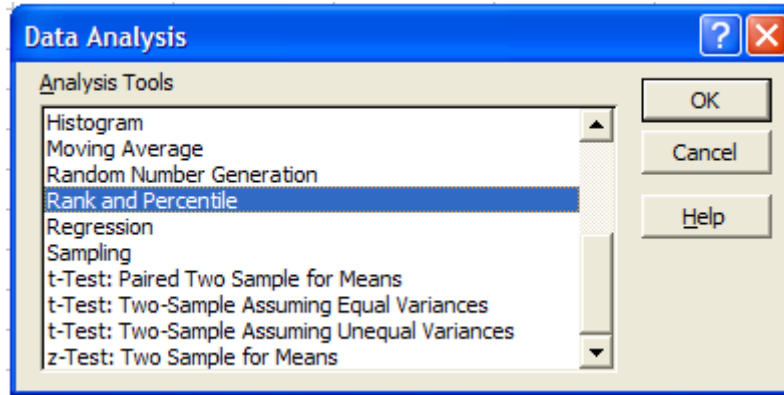


الرتب والمئينات Rank and Percentile

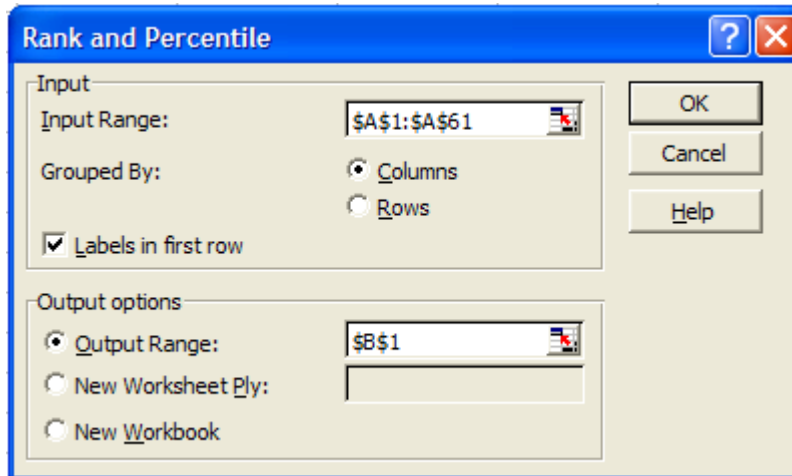
سوف نوجد رتب ومئينات البيانات التالية

44.2	44.3	44.4	43.4	42.8	44.3	44.4	44.8	44.4	43.1	42.6
42.4	42.2	41.8	40.1	42.0	42.4	43.1	42.4	43.1	43.2	42.8
43.0	42.8	42.5	42.6	42.3	42.9	43.6	44.7	44.5	45.0	44.8
44.9	45.2	45.2	45.0	45.5	46.2	46.8	47.5	48.3	48.3	49.1
48.9	49.4	50.0	50.0	49.6	49.9	49.6	50.7	50.7	50.9	50.5
51.2	50.7	50.3	49.2	48.1						

من قائمة الأدوات ومن تحليل البيانات نختار Rank and Percentile كالتالي:



وفي نافذة المدخلات ندخل المطلوب كالتالي:



فينتج

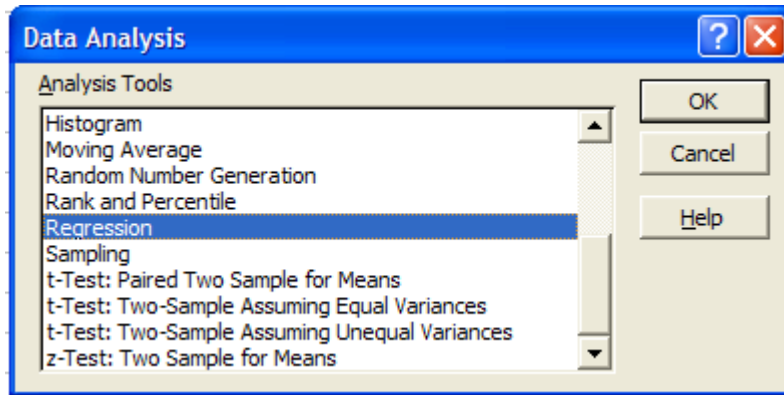
	A	B	C	D	E
1	Value	Point	Value	Rank	Percent
2	44.2	56	51.2	1	100.00%
3	44.3	54	50.9	2	98.30%
4	44.4	52	50.7	3	93.20%
5	43.4	53	50.7	3	93.20%
6	42.8	57	50.7	3	93.20%
7	44.3	55	50.5	6	91.50%
8	44.4	58	50.3	7	89.80%
9	44.8	47	50	8	86.40%
10	44.4	48	50	8	86.40%
11	43.1	50	49.9	10	84.70%
12	42.6	49	49.6	11	81.30%
13	42.4	51	49.6	11	81.30%
14	42.2	46	49.4	13	79.60%
15	41.8	59	49.2	14	77.90%

16	40.1	44	49.1	15	76.20%
17	42	45	48.9	16	74.50%
18	42.4	42	48.3	17	71.10%
19	43.1	43	48.3	17	71.10%
20	42.4	60	48.1	19	69.40%
21	43.1	41	47.5	20	67.70%
22	43.2	40	46.8	21	66.10%
23	42.8	39	46.2	22	64.40%
24	43	38	45.5	23	62.70%
25	42.8	35	45.2	24	59.30%
26	42.5	36	45.2	24	59.30%
27	42.6	32	45	26	55.90%
28	42.3	37	45	26	55.90%
29	42.9	34	44.9	28	54.20%
30	43.6	8	44.8	29	50.80%
31	44.7	33	44.8	29	50.80%
32	44.5	30	44.7	31	49.10%

الإنحدار Regression

سوف نستعرض الإنحدار الخطي على البيانات التالية

X	Y
42	125
36	118
63	140
55	150
42	140
60	155
49	145
68	152



فتظهر نافذة المدخلات

Regression

Input

Input Y Range:

Input X Range:

☒ Labels ☐ Constant is Zero

☒ Confidence Level: %

Output options

☒ Output Range:

☐ New Worksheet Ply:

☐ New Workbook

Residuals

☐ Residuals ☒ Residual Plots

☐ Standardized Residuals ☐ Line Fit Plots

Normal Probability

☒ Normal Probability Plots

OK Cancel Help

والنتائج

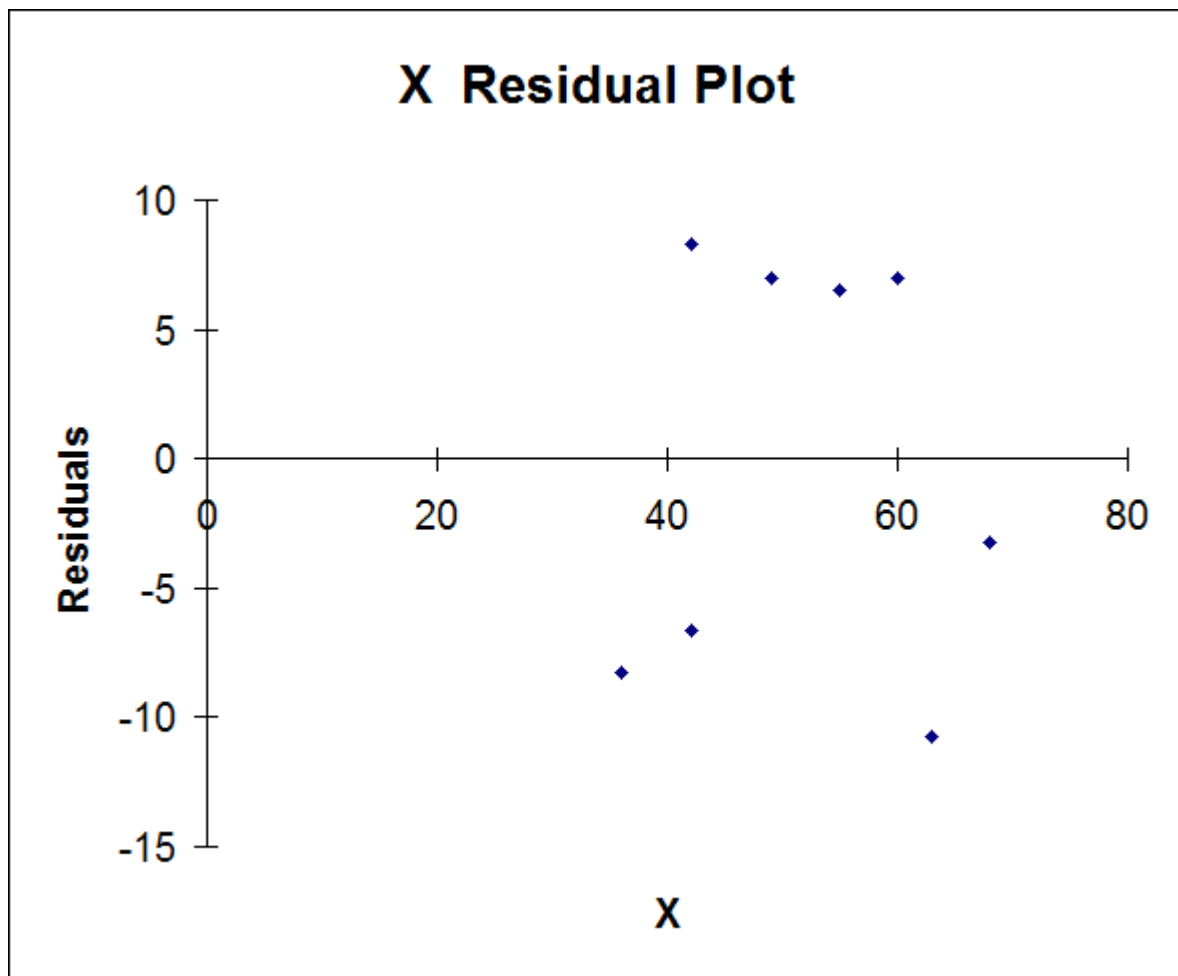
SUMMARY OUTPUT					
Regression Statistics					
Multiple R	0.791831817				
R Square	0.626997626				
Adjusted R Square	0.564830564				
Standard Error	8.636706775				
Observations	8				
ANOVA					
	df	SS	MS	F	Significance F
Regression	1	752.3187765	752.3187765	10.0856885	0.019177545
Residual	6	447.5562235	74.59270392		
Total	7	1199.875			

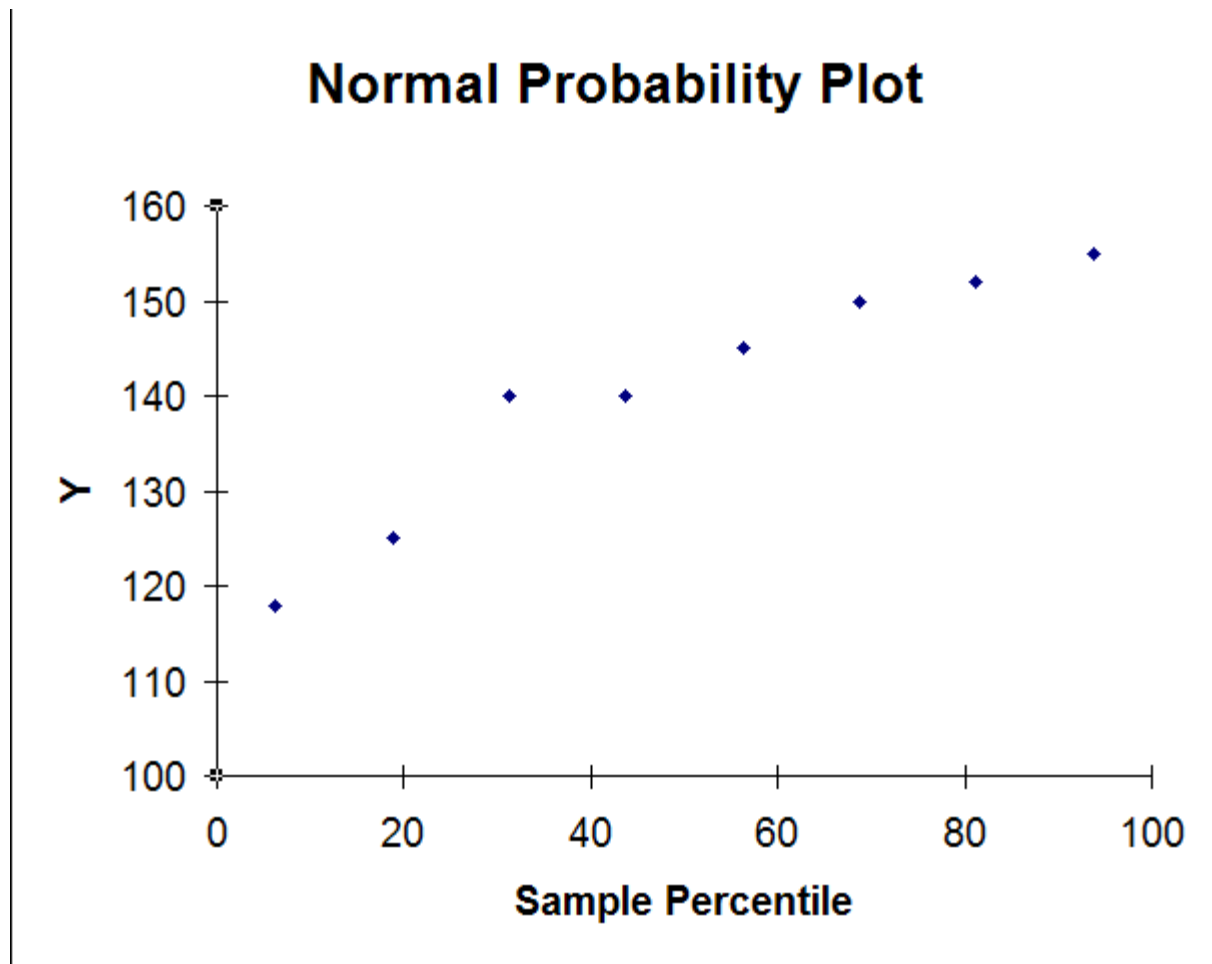
	Coefficients	Standard Error	t Stat	P-value
Intercept	93.58382293	15.12386375	6.187825046	0.000819999
X	0.906817871	0.285540223	3.175797302	0.019177545

Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
56.57703441	130.5906114	56.57703441	130.5906114
0.208125604	1.605510139	0.208125604	1.605510139

RESIDUAL OUTPUT		
<i>Observation</i>	<i>Predicted Y</i>	<i>Residuals</i>
1	131.6701735	-6.670173521
2	126.2292663	-8.229266293
3	150.7133488	-10.71334882
4	143.4588058	6.541194152
5	131.6701735	8.329826479
6	147.9928952	7.007104796
7	138.0178986	6.98210138
8	155.2474382	-3.247438175

PROBABILITY OUTPUT	
<i>Percentile</i>	<i>Y</i>
6.25	118
18.75	125
31.25	140
43.75	140
56.25	145
68.75	150
81.25	152
93.75	155





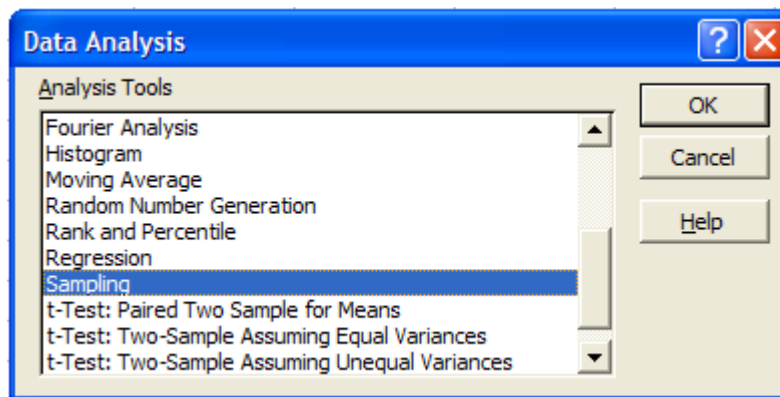
المعاينة Sampling

سوف نقوم بسحب عينة عشوائية من المجتمع $\{0,1\}$ حجمها 60 وحدة كالتالي:

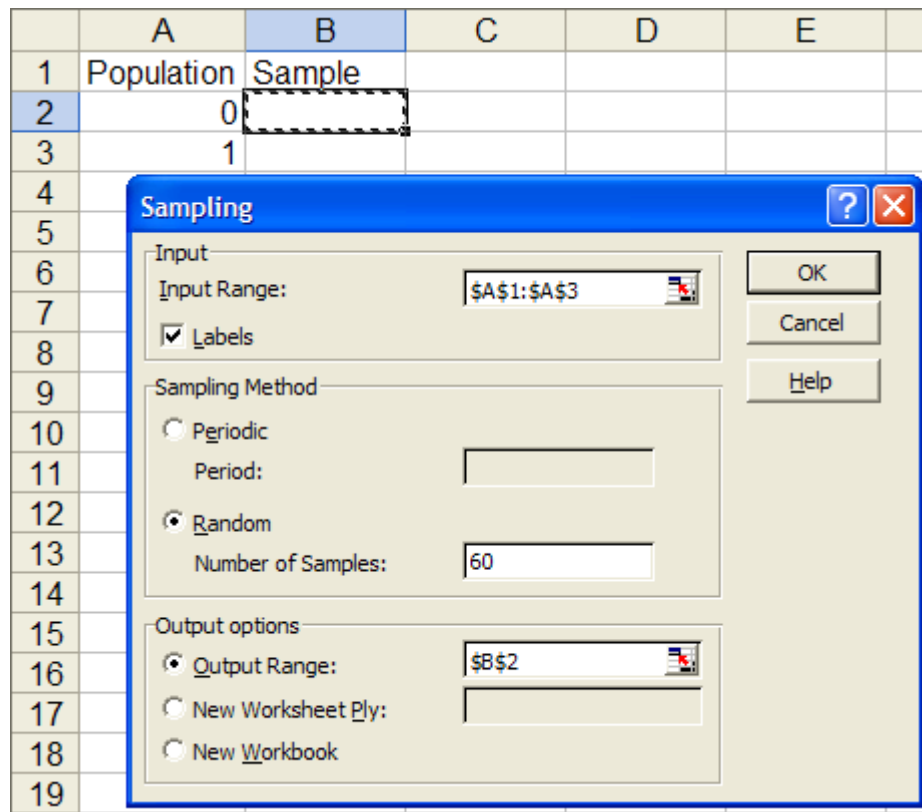
في صفحة من إكسل ادخل التالي

	A	B
1	Population	Sample
2	0	
3	1	

من قائمة الأدوات نختار تحليل البيانات ثم المعاينة



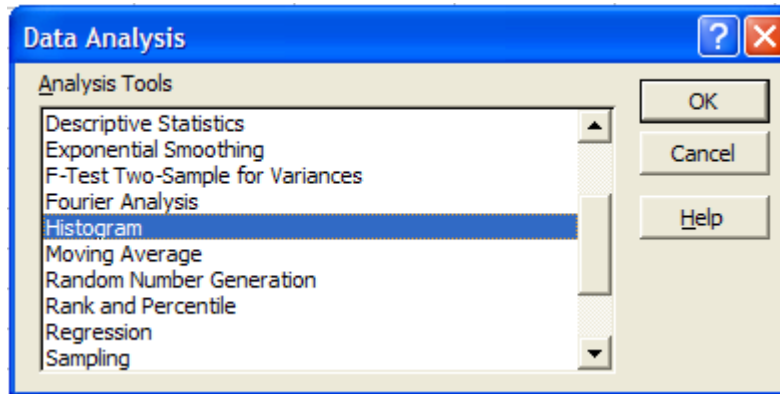
فتظهر نافذة إدخال المعلومات



فينتج (جزء من المخرجات)

	A	B	C
1	Population	Sample	
2	0	1	
3	1	0	
4		0	
5		0	
6		1	
7		0	
8		1	
9		1	
10		0	

المعينة هنا كانت بإحلال، سوف نوجد التوزيع التكراري للعينة باستخدام أداة المدرج التكراري HISTOGRAM الموجودة ضمن تحليل البيانات

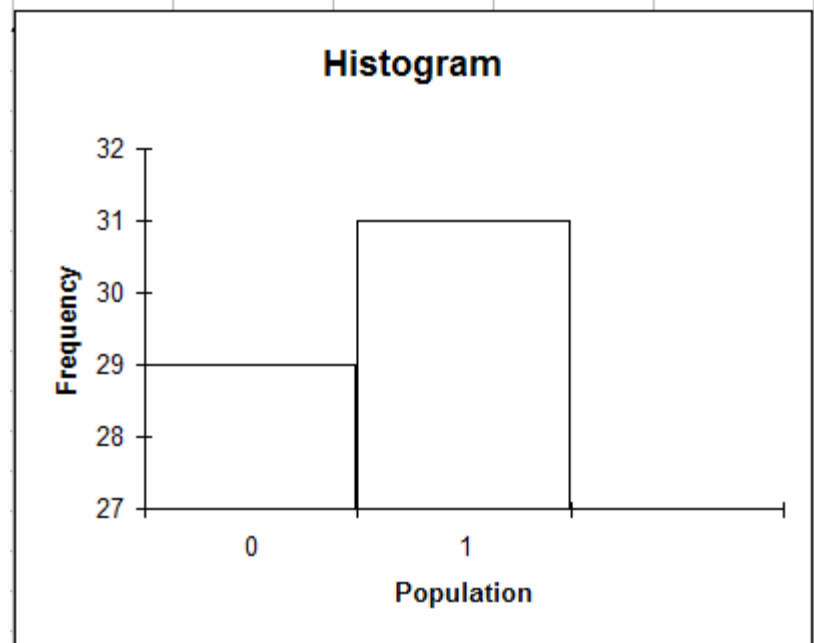


وفي نافذة الإدخال

	A	B	C	D	E	F	G
1	Population	Sample					
2	0	1					
3	1	0					
4		0					
5		0					
6		1					
7		0					
8		1					
9		1					
10		0					
11		0					
12		0					
13		0					
14		1					
15		0					
16		1					
17		1					

فينتج

<i>Population</i>	<i>Frequency</i>			
0	29			
1	31			

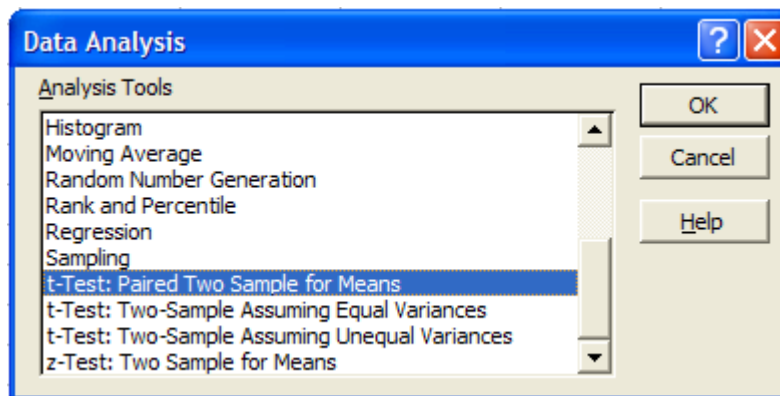


t-Test: Paired Two Sample for Means

البيانات التالية هي درجات 7 طلاب في مادتين مختلفتين

	A	B
1	X	Y
2	62	53
3	82	75
4	77	65
5	57	55
6	62	67
7	90	85
8	82	79

اختبر الفرض القائل انه لا يوجد فرق بين متوسطي درجات المادتين عند مستوى
معنوية 0.05



وندخل البيانات

t-Test: Paired Two Sample for Means

Input

Variable 1 Range:

Variable 2 Range:

Hypothesized Mean Difference:

☒ Labels

Alpha:

Output options

☒ Output Range:

☐ New Worksheet Ply:

☐ New Workbook

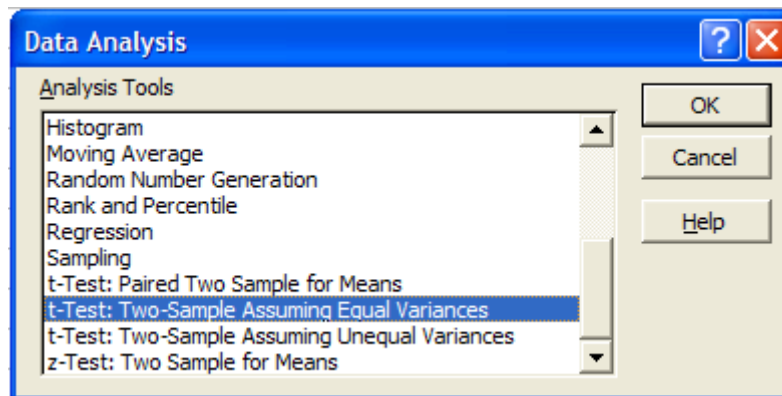
OK Cancel Help

والنتائج

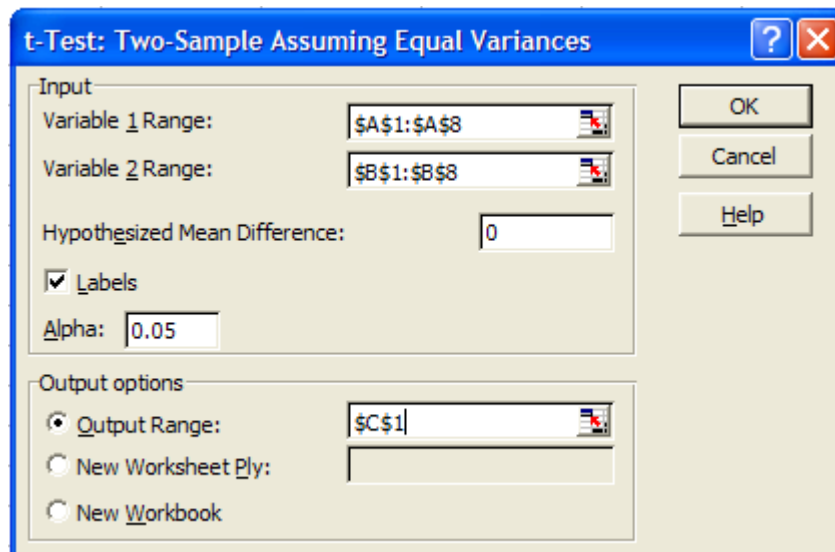
C	D	E
t-Test: Paired Two Sample for Means		
	X	Y
Mean	73.14285714	68.42857143
Variance	160.8095238	143.6190476
Observations	7	7
Pearson Correlation	0.902112013	
Hypothesized Mean Difference	0	
df	6	
t Stat	2.268233209	
P(T<=t) one-tail	0.031911178	
t Critical one-tail	1.943180905	
P(T<=t) two-tail	0.063822356	
t Critical two-tail	2.446913641	

إختبار t لعينتين على إفتراض تساوي التباين Assuming Equal Variances

سوف نقوم بهذا الإختبار للمثال السابق



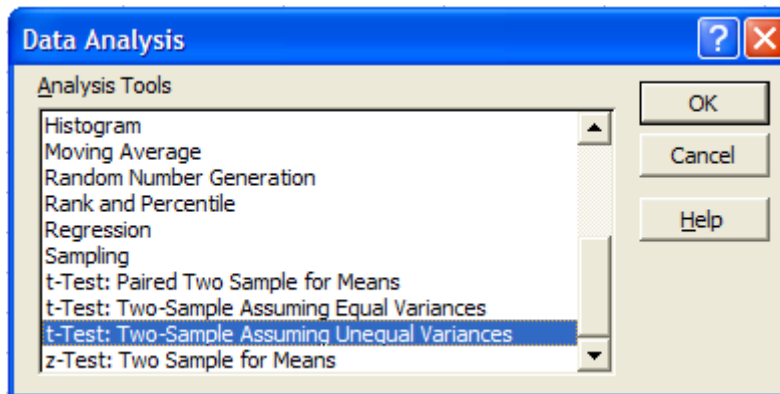
تدخل البيانات المطلوبة



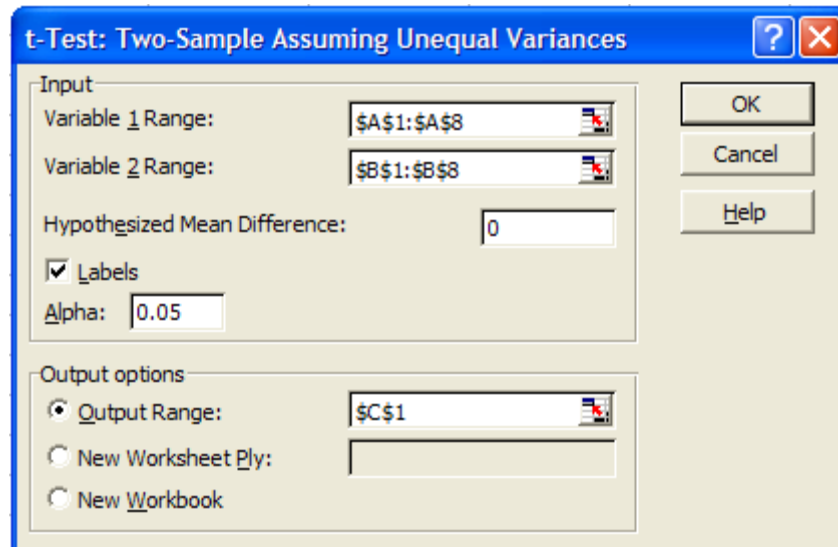
C	D	E
t-Test: Two-Sample Assuming Equal Variances		
	X	Y
Mean	73.14285714	68.42857143
Variance	160.8095238	143.6190476
Observations	7	7
Pooled Variance	152.2142857	
Hypothesized Mean Difference	0	
df	12	
t Stat	0.714862005	
P(T<=t) one-tail	0.244184537	
t Critical one-tail	1.782286745	
P(T<=t) two-tail	0.488369074	
t Critical two-tail	2.178812792	

إختبار t لعينتين على إفتراض عدم تساوي التباين t-Test: Two-Sample Assuming Unequal Variances

سوف نقوم بهذا الإختبار للمثال السابق



تدخل البيانات المطلوبة

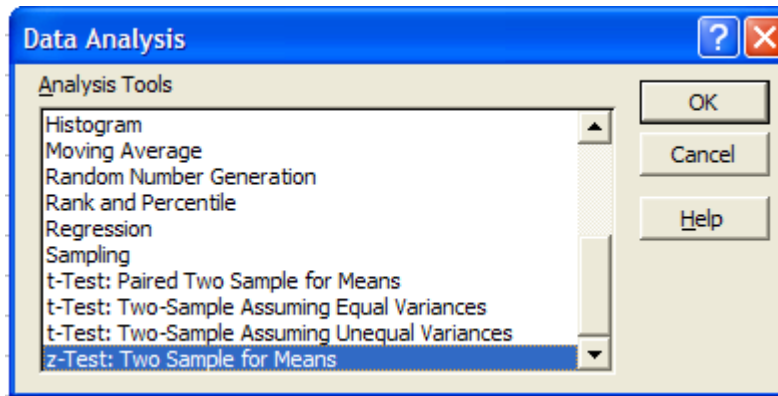


وينتج

C	D	E
t-Test: Two-Sample Assuming Unequal Variances		
	X	Y
Mean	73.14285714	68.42857143
Variance	160.8095238	143.6190476
Observations	7	7
Hypothesized Mean Difference	0	
df	12	
t Stat	0.714862005	
P(T<=t) one-tail	0.244184537	
t Critical one-tail	1.782286745	
P(T<=t) two-tail	0.488369074	
t Critical two-tail	2.178812792	

إختبار z للمتوسطات لعينتين z-Test: Two-Sample for Means

سوف نستعرض هذا الإختبار على المثال السابق



ندخل البيانات المطلوبة

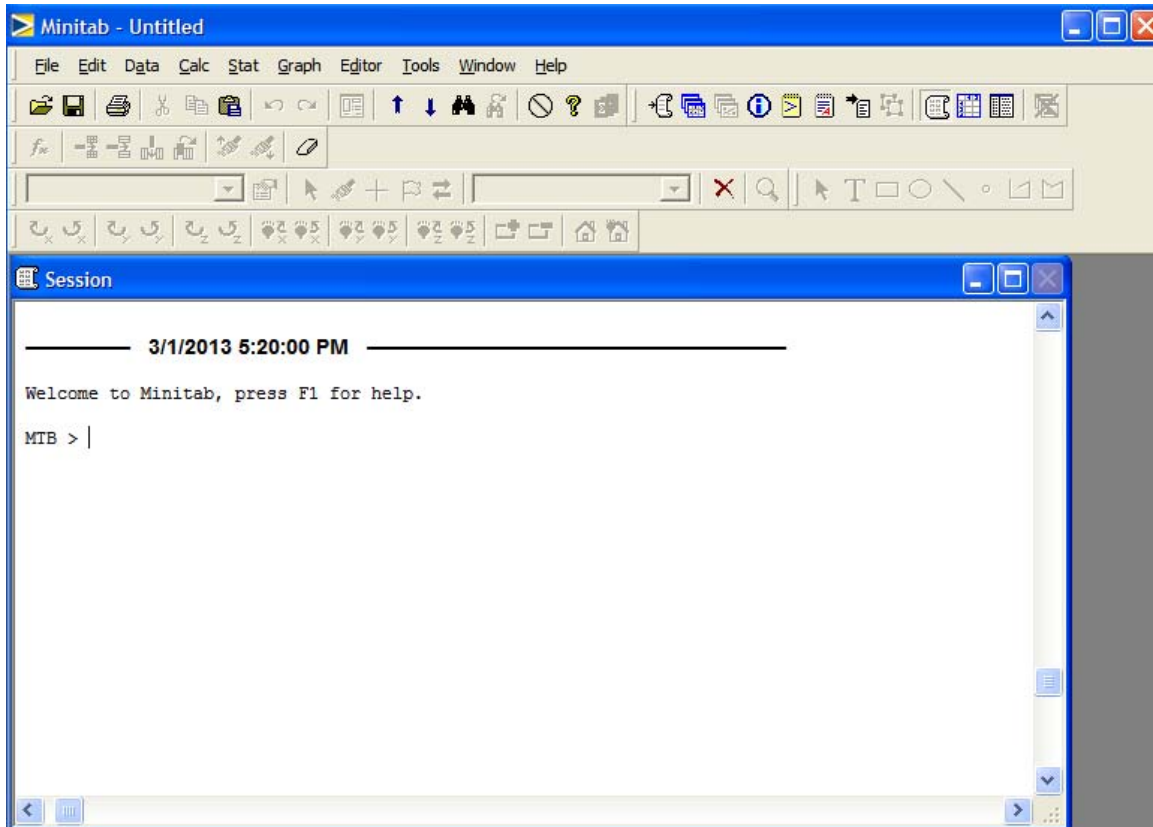
The image shows the 'z-Test: Two Sample for Means' dialog box. The 'Input' section contains the following fields: 'Variable 1 Range' is '\$A\$1:\$A\$8', 'Variable 2 Range' is '\$B\$1:\$B\$8', 'Hypothesized Mean Difference' is '0', 'Variable 1 Variance (known)' is '161', and 'Variable 2 Variance (known)' is '145'. The 'Labels' checkbox is checked, and the 'Alpha' field is set to '0.05'. The 'Output options' section has three radio buttons: 'Output Range' (selected), 'New Worksheet Ply', and 'New Workbook'. The 'Output Range' field is set to '\$C\$1'. On the right side, there are buttons for 'OK', 'Cancel', and 'Help'.

ويكون الناتج

C	D	E
z-Test: Two Sample for Means		
	X	Y
Mean	73.14285714	68.42857143
Known Variance	161	145
Observations	7	7
Hypothesized Mean Difference	0	
z	0.713024096	
P(Z<=z) one-tail	0.237915349	
z Critical one-tail	1.644853476	
P(Z<=z) two-tail	0.475830698	
z Critical two-tail	1.959962787	

ويترك للطالب المقارنة بين الإختبارات السابقة.

الحزمة الإحصائية Minitab

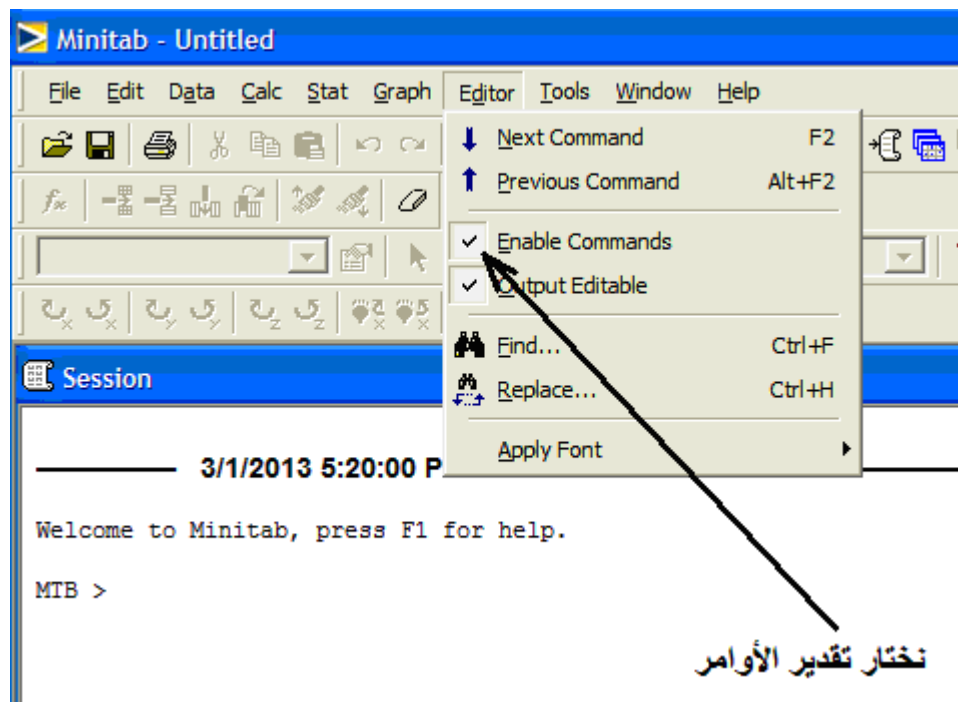


Worksheet 1 ***

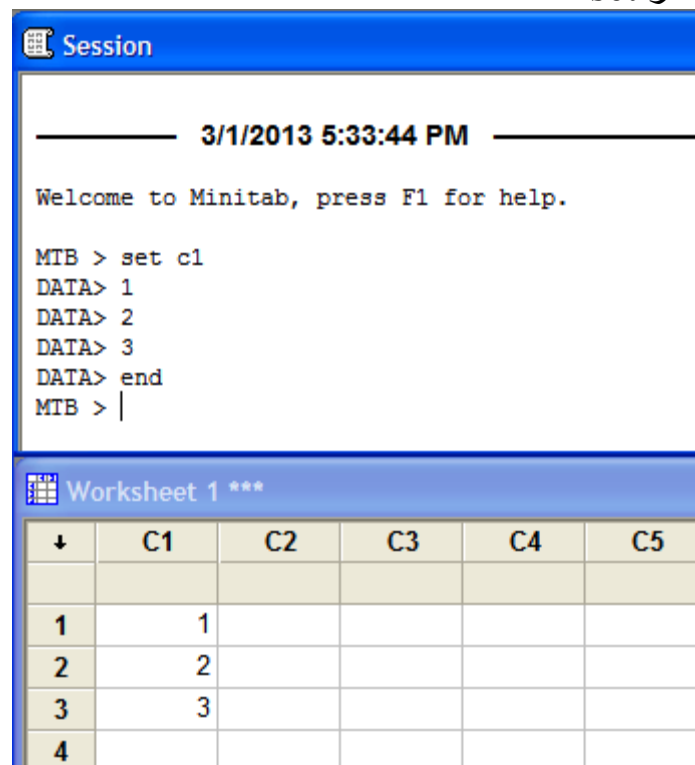
↓	C1-T	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11
	make	price	mpg	rep78	weight	length	foreign				
1	AMC	4099	22	3	2930	186	0				
2	AMC	4749	17	3	3350	173	0				
3	AMC	3799	22	3	2640	168	0				
4	Audi	9690	17	5	2830	189	1				
5	Audi	6295	23	3	2070	174	1				
6	BMW	9735	25	4	2650	177	1				
7	Buick	4816	20	3	3250	196	0				
8	Buick	7827	15	4	4080	222	0				
9	Buick	5788	18	3	3670	218	0				
10	Buick	4453	26	3	2230	170	0				
11	Buick	5189	20	3	3280	200	0				
12	Buick	10372	16	3	3880	207	0				
13	Buick	4082	19	3	3400	200	0				
14	Cad.	11385	14	3	4330	221	0				
15	Cad.	14500	14	2	3900	204	0				

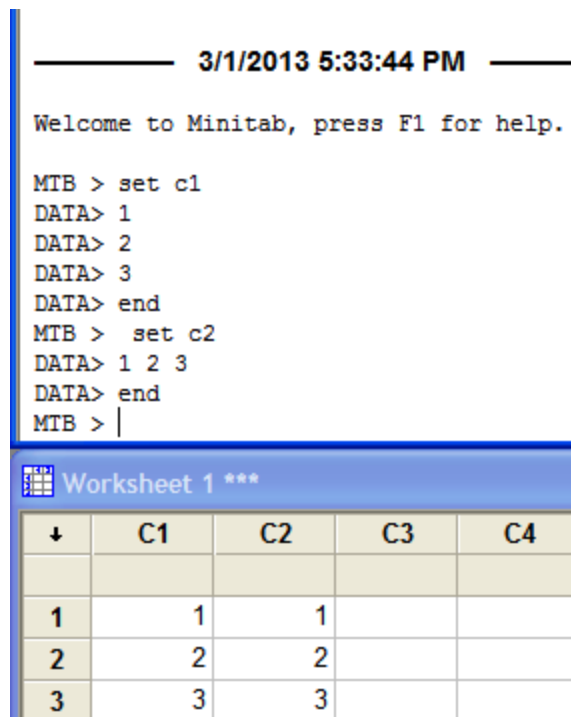
Current Worksheet: Worksheet 1 Editable

إدخال البيانات
من المحرر : لكي نظهر محفز المحرر > MTB نتأكد من إختيارنا لنافذة التحرير ثم



وتدخل البيانات بالأمر set





إدخال بيانات من ملف

MTB > WOpen "K:\Statistical Packages\minitab\cars.MTW".
Retrieving worksheet from file: 'K:\Statistical
Packages\minitab\cars.MTW'
Worksheet was saved on Fri Mar 01 2013

Results for: cars.MTW

MTB > info

Information on the Worksheet

	Column	Count	Name
T	C1	26	make
	C2	26	price
	C3	26	mpg
	C4	26	rep78
	C5	26	weight
	C6	26	length
	C7	26	foreign

طباعة البيانات

MTB > print c1-c7

Data Display

Row	make	price	mpg	rep78	weight	length	foreign
1	AMC	4099	22	3	2930	186	0
2	AMC	4749	17	3	3350	173	0
3	AMC	3799	22	3	2640	168	0

4	Audi	9690	17	5	2830	189	1
5	Audi	6295	23	3	2070	174	1
6	BMW	9735	25	4	2650	177	1
7	Buick	4816	20	3	3250	196	0
8	Buick	7827	15	4	4080	222	0
9	Buick	5788	18	3	3670	218	0
10	Buick	4453	26	3	2230	170	0
11	Buick	5189	20	3	3280	200	0
12	Buick	10372	16	3	3880	207	0
13	Buick	4082	19	3	3400	200	0
14	Cad.	11385	14	3	4330	221	0
15	Cad.	14500	14	2	3900	204	0
16	Cad.	15906	21	3	4290	204	0
17	Chev.	3299	29	3	2110	163	0
18	Chev.	5705	16	4	3690	212	0
19	Chev.	4504	22	3	3180	193	0
20	Chev.	5104	22	2	3220	200	0
21	Chev.	3667	24	2	2750	179	0
22	Chev.	3955	19	3	3430	197	0
23	Datsun	6229	23	4	2370	170	1
24	Datsun	4589	35	5	2020	165	1
25	Datsun	5079	24	4	2280	170	1
26	Datsun	8129	21	4	2750	184	1

MTB >

الإحصاء الوصفي

MTB > Describe 'price' 'mpg' 'rep78' 'weight' 'length' 'foreign';

```

SUBC> Mean;
SUBC> SEMean;
SUBC> StDeviation;
SUBC> Variance;
SUBC> CVariation;
SUBC> QOne;
SUBC> Median;
SUBC> QThree;
SUBC> IQRRange;
SUBC> Mode;
SUBC> TRMean;
SUBC> Sums;
SUBC> Minimum;
SUBC> Maximum;
SUBC> Range;
SUBC> SSQ;
SUBC> Skewness;
SUBC> Kurtosis;
SUBC> MSSD;
SUBC> N;
SUBC> NMissing;
SUBC> Count;
SUBC> CumN;
SUBC> Percent;
SUBC> CumPercent;
SUBC> GHist;
SUBC> GNHhist;
SUBC> GBoxplot.

```

Descriptive Statistics: price, mpg, rep78, weight, length, foreign

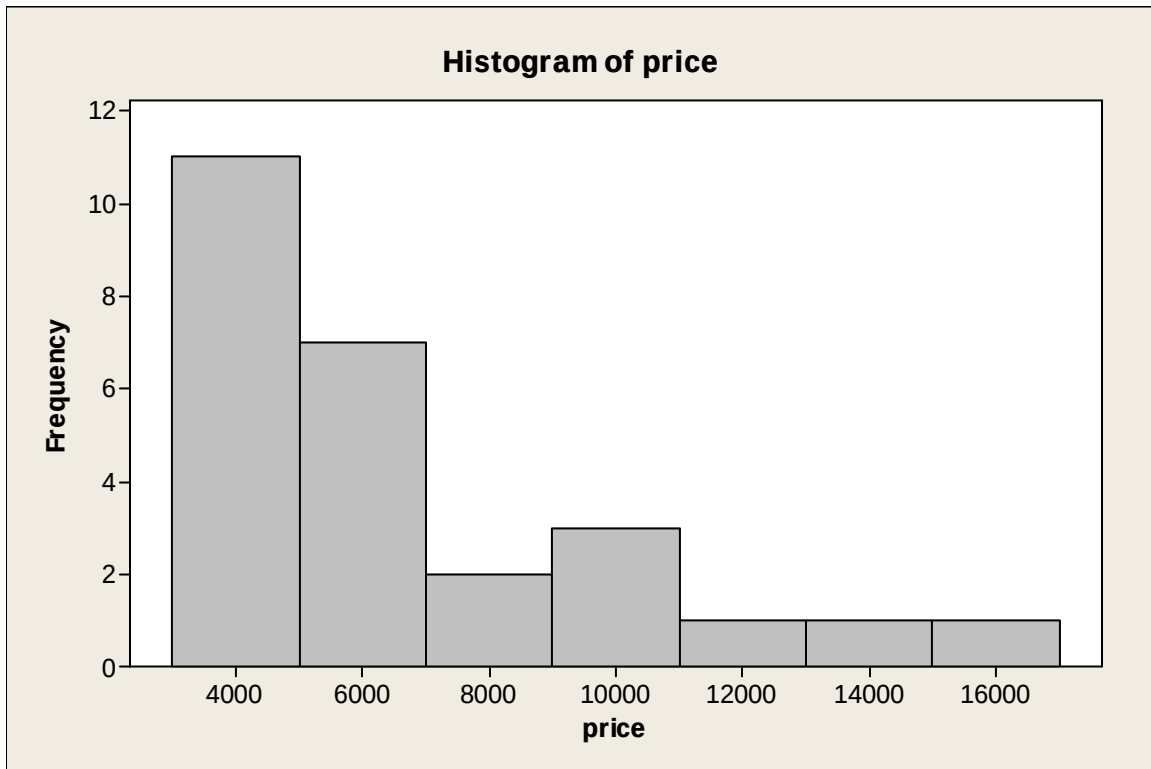
Variable	Total Count	N	N*	CumN	Percent	CumPct	Mean	SE Mean	TrMean	StDev
price	26	26	0	26	100	100	6652	661	6406	3371
mpg	26	26	0	26	100	100	20.923	0.933	20.625	4.758
rep78	26	26	0	26	100	100	3.269	0.152	3.250	0.778
weight	26	26	0	26	100	100	3099	136	3093	695
length	26	26	0	26	100	100	190.08	3.56	189.88	18.17
foreign	26	26	0	26	100	100	0.2692	0.0887	0.2500	0.4523

Variable	Variance	CoefVar	Sum	Sum of Squares	Minimum	Q1	Median
price	11364449	50.68	172945	1434494797	3299	4365	5147
mpg	22.634	22.74	544.000	11948.000	14.000	17.000	21.000
rep78	0.605	23.78	85.000	293.000	2.000	3.000	3.000
weight	483135	22.43	80580	261814400	2020	2573	3200
length	330.15	9.56	4942.00	947614.00	163.00	172.25	191.00
foreign	0.2046	168.01	7.0000	7.0000	0.0000	0.0000	0.0000

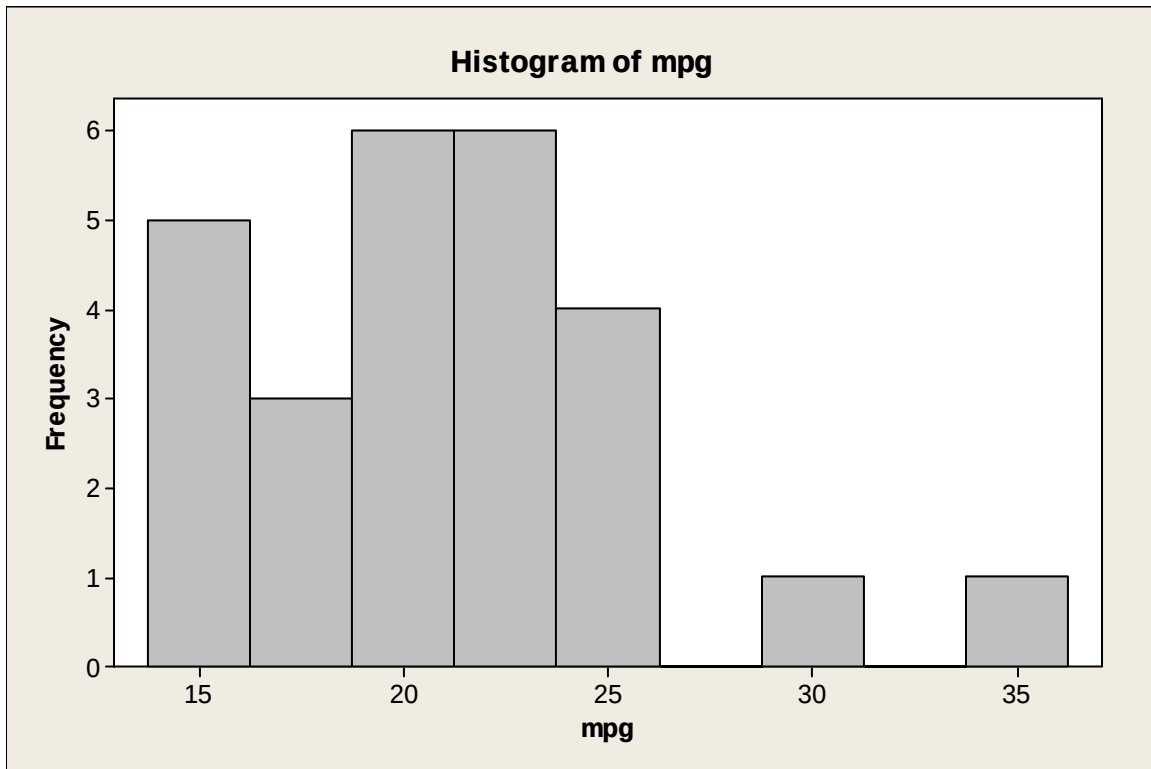
Variable	Q3	Maximum	Range	IQR	Mode	N for Mode	Skewness	Kurtosis
price	8519	15906	12607	4155	*	0	1.47	1.53
mpg	23.250	35.000	21.000	6.250	22	4	0.94	1.79
rep78	4.000	5.000	3.000	1.000	3	15	0.58	0.43
weight	3675	4330	2310	1103	2750	2	0.10	-0.97
length	204.00	222.00	59.00	31.75	170, 200	3	0.18	-1.14
foreign	1.0000	1.0000	1.0000	1.0000	0	19	1.11	-0.85

Variable	MSSD
price	8333059
mpg	19.220
rep78	0.420
weight	350888
length	243.72
foreign	0.0600

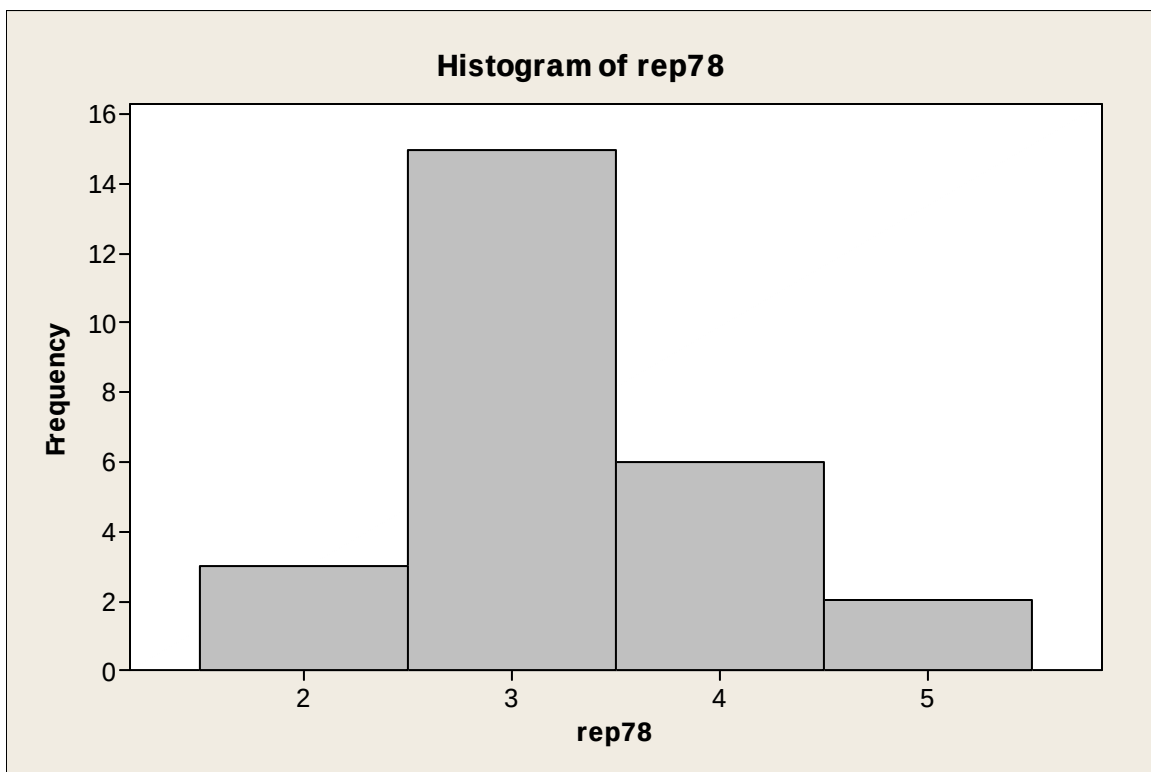
Histogram of price



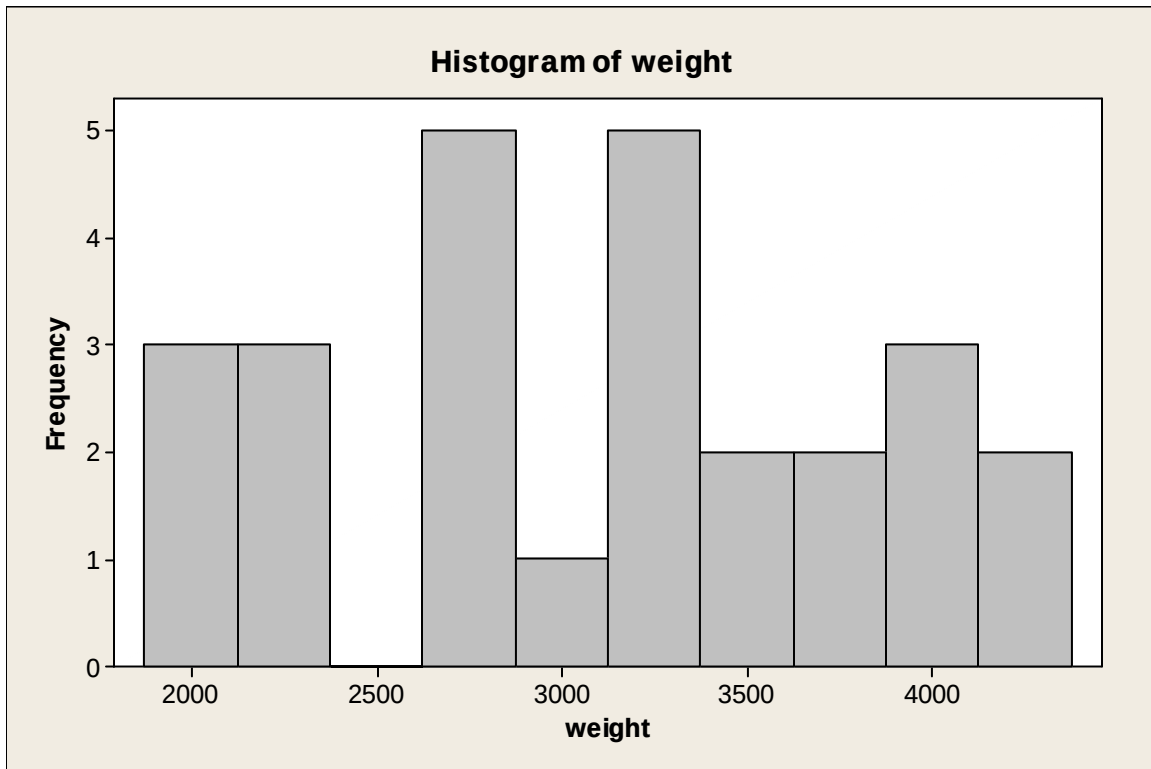
Histogram of mpg



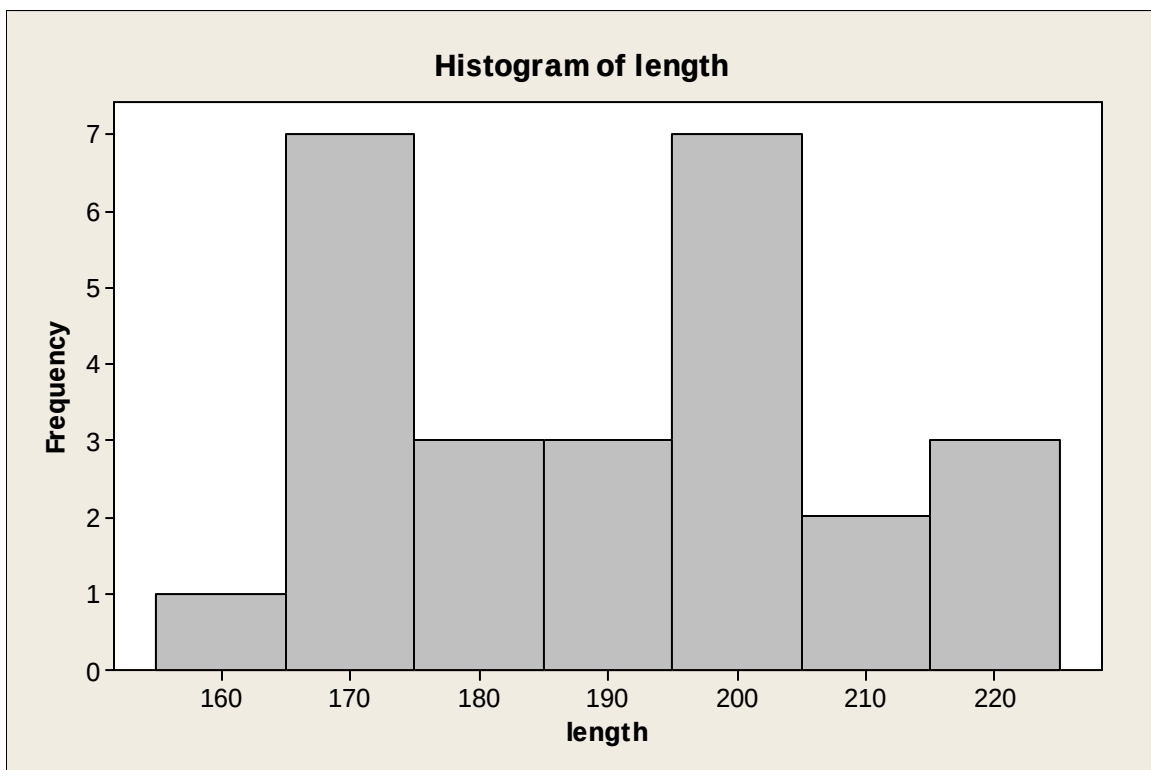
Histogram of rep78



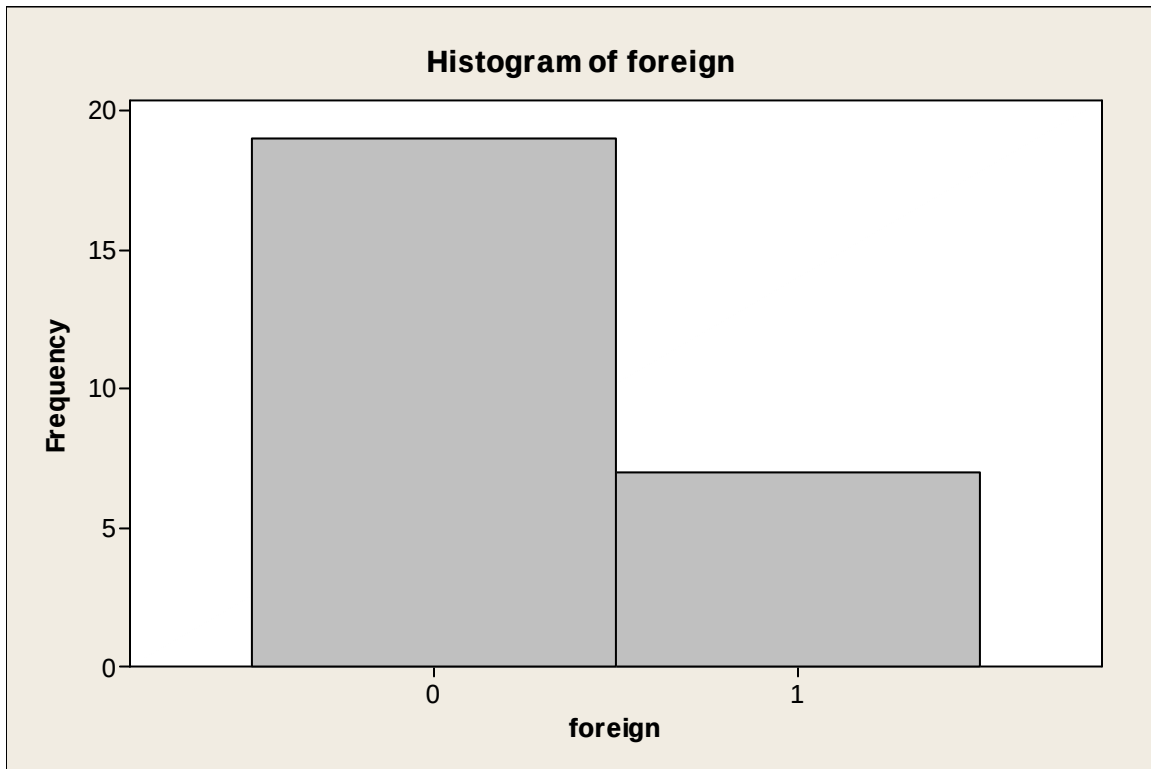
Histogram of weight



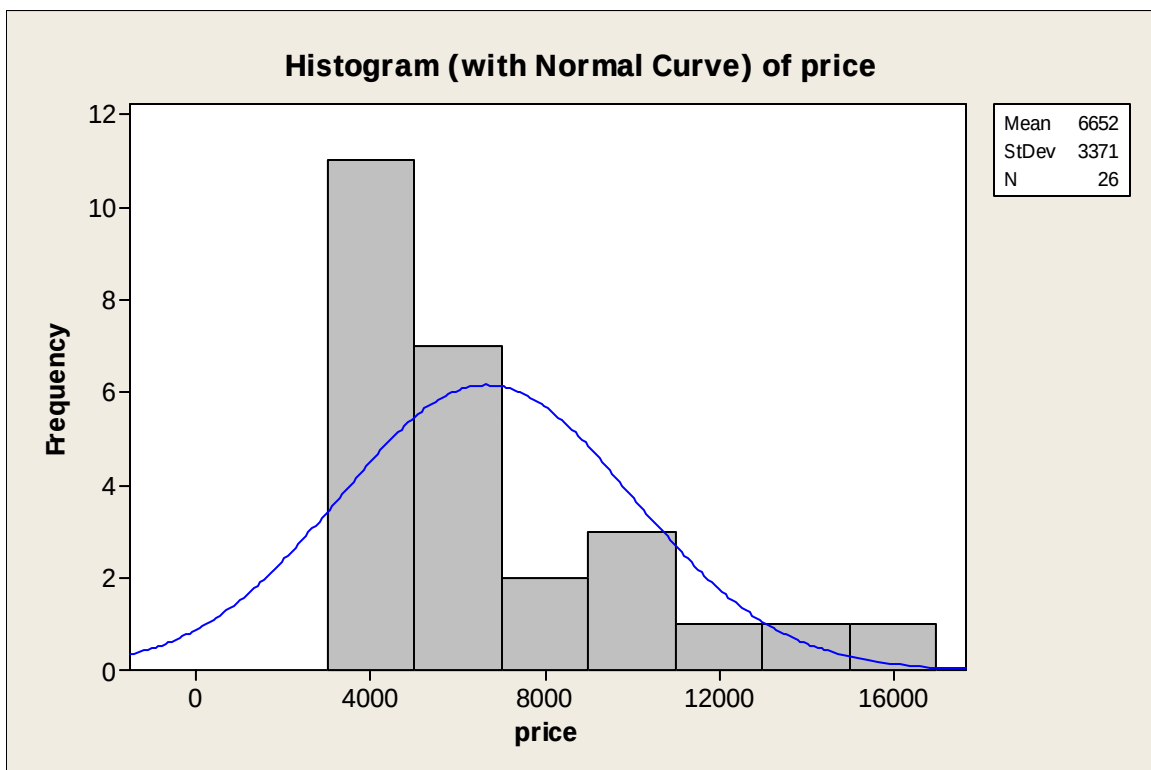
Histogram of length



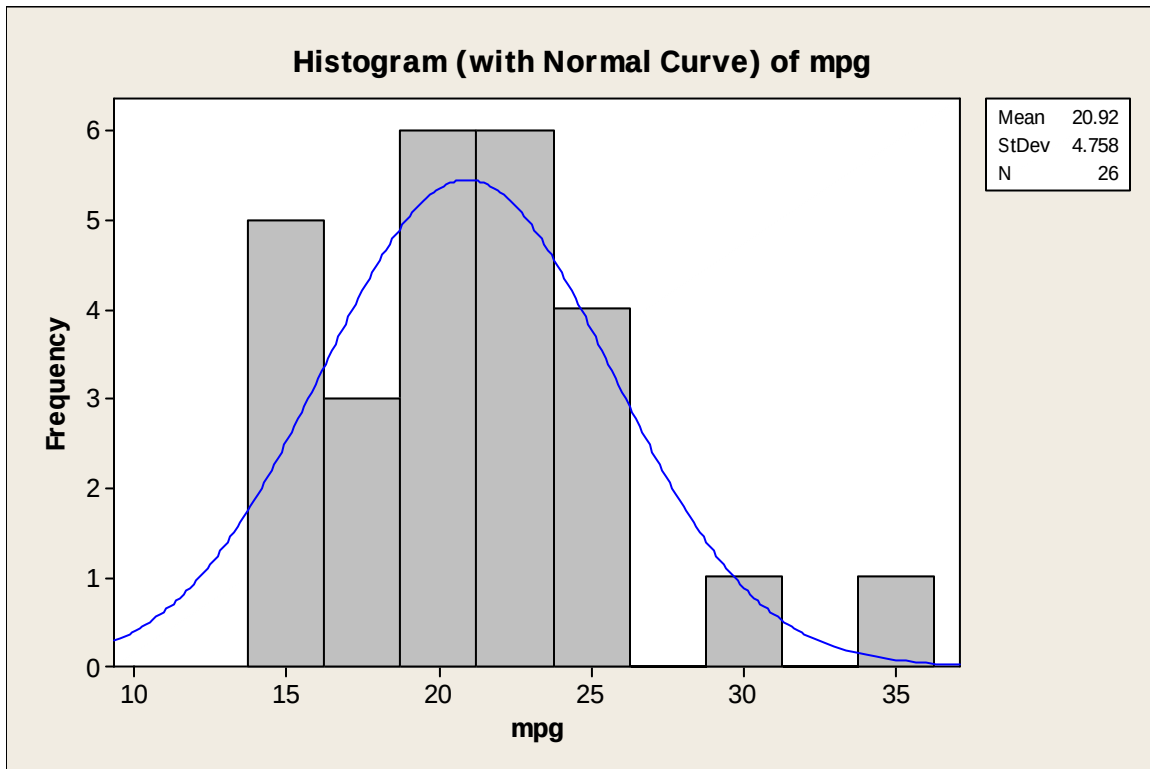
Histogram of foreign



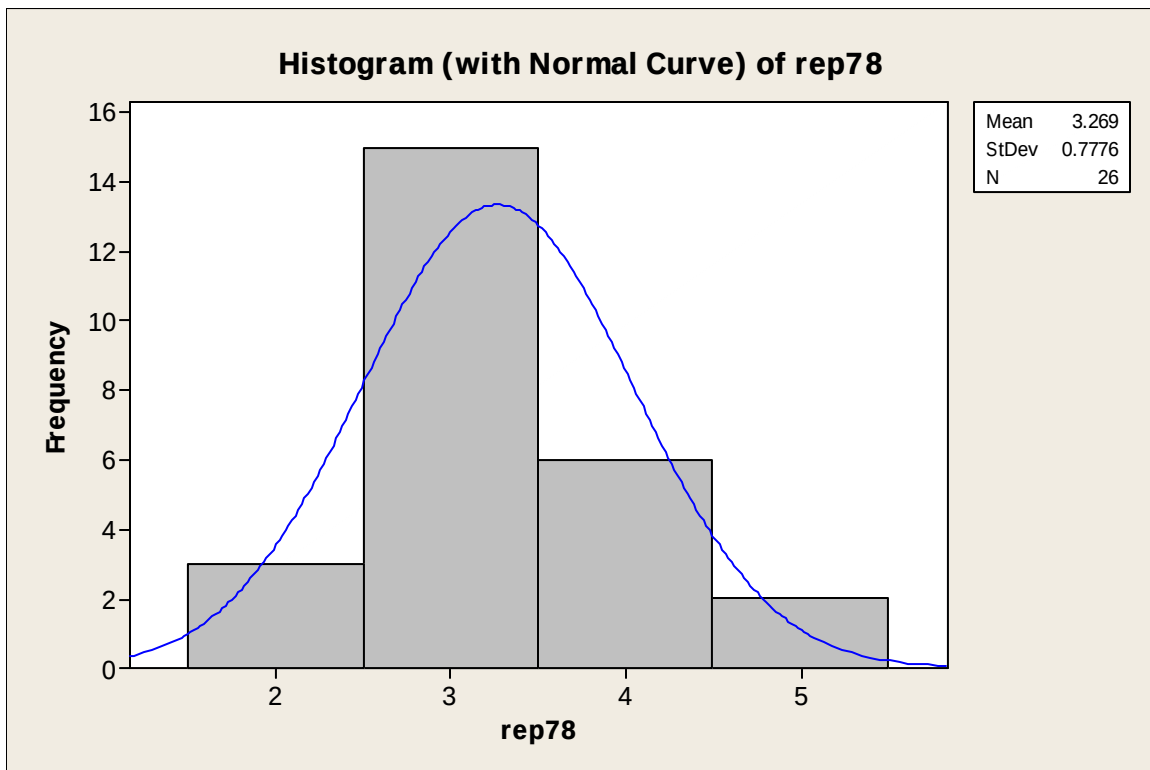
Histogram (with Normal Curve) of price



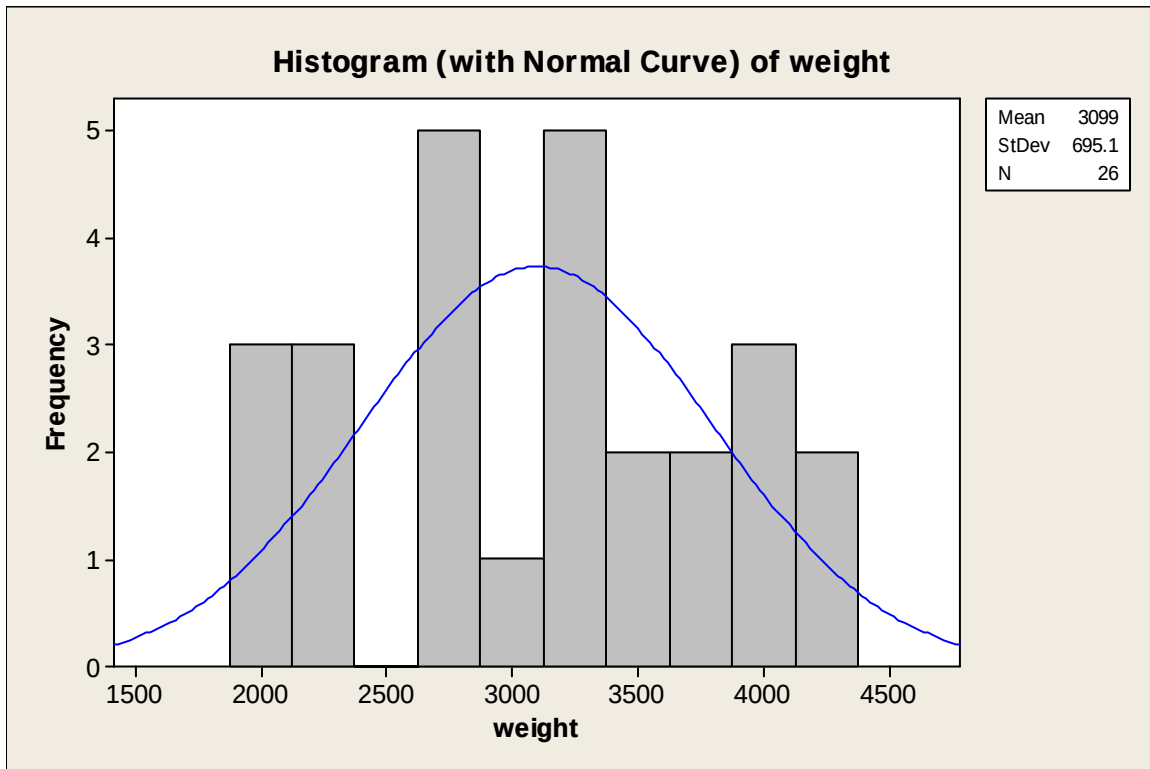
Histogram (with Normal Curve) of mpg



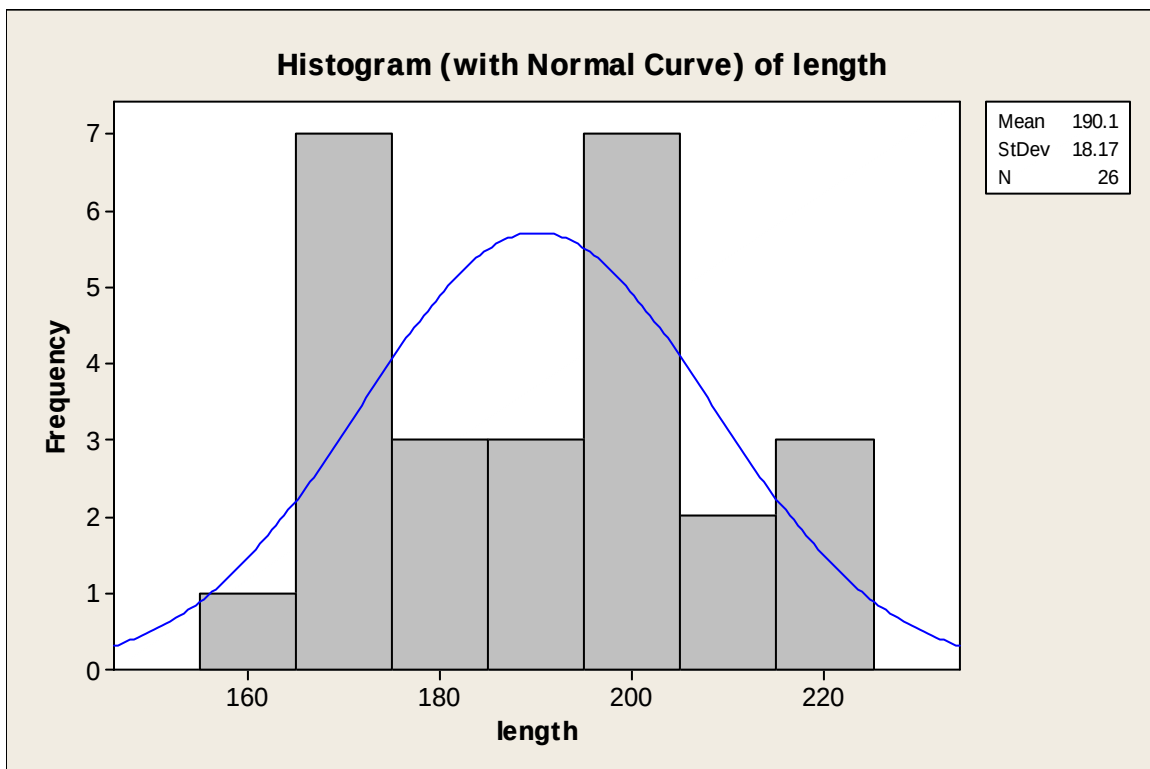
Histogram (with Normal Curve) of rep78



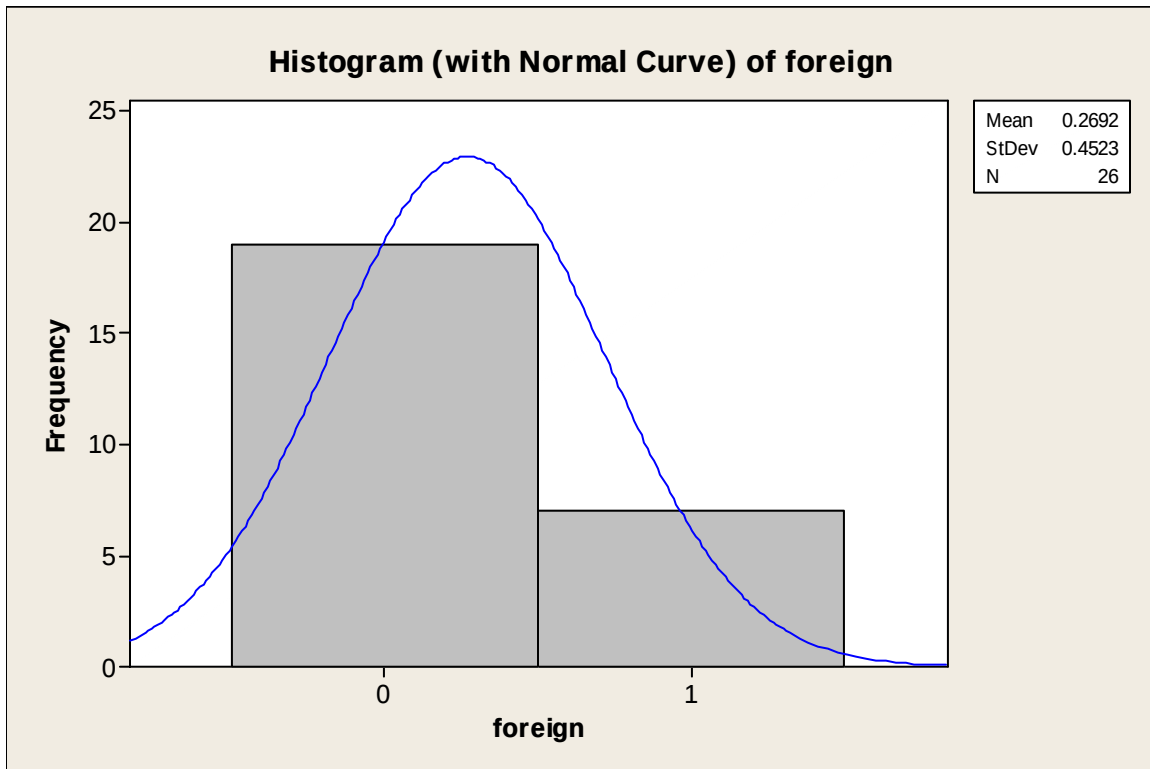
Histogram (with Normal Curve) of weight



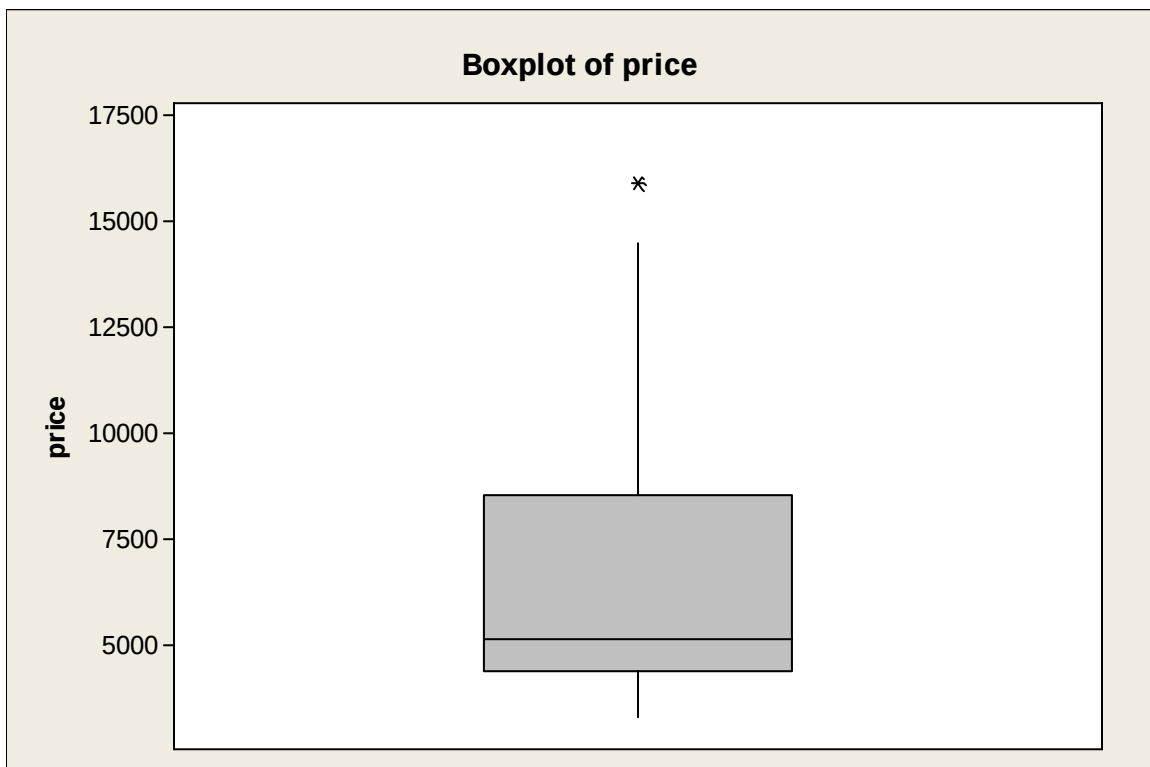
Histogram (with Normal Curve) of length



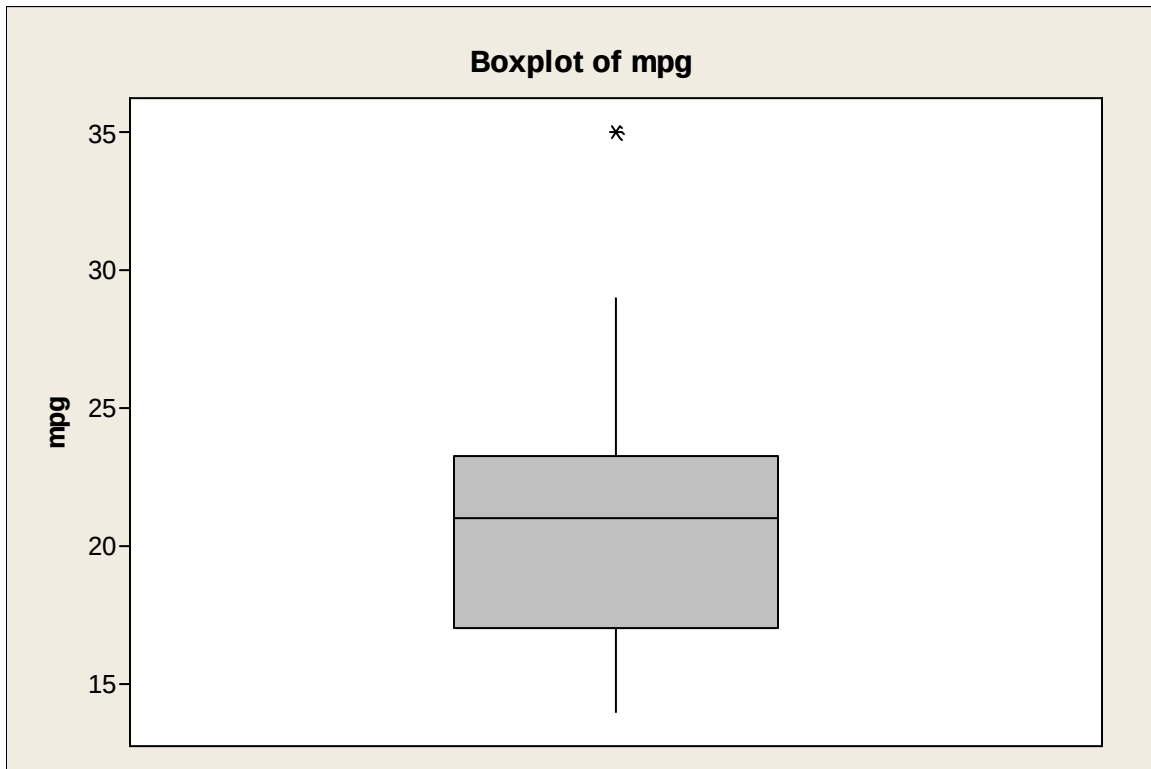
Histogram (with Normal Curve) of foreign



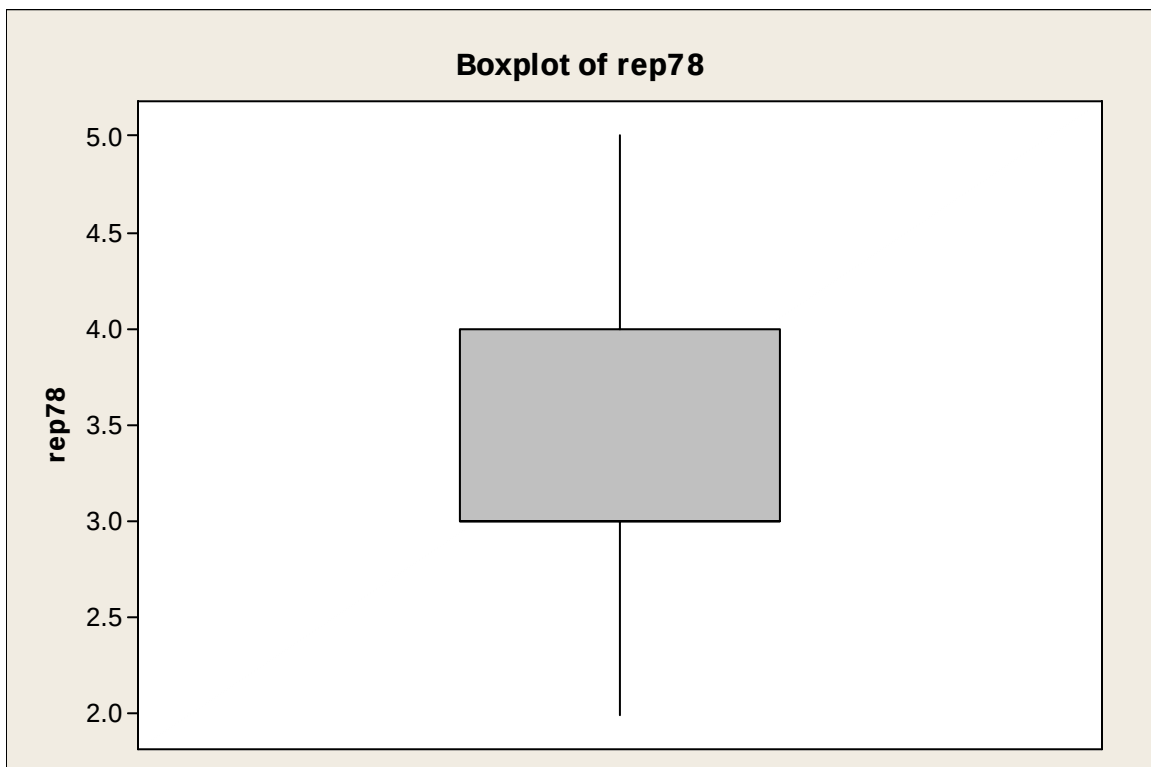
Boxplot of price



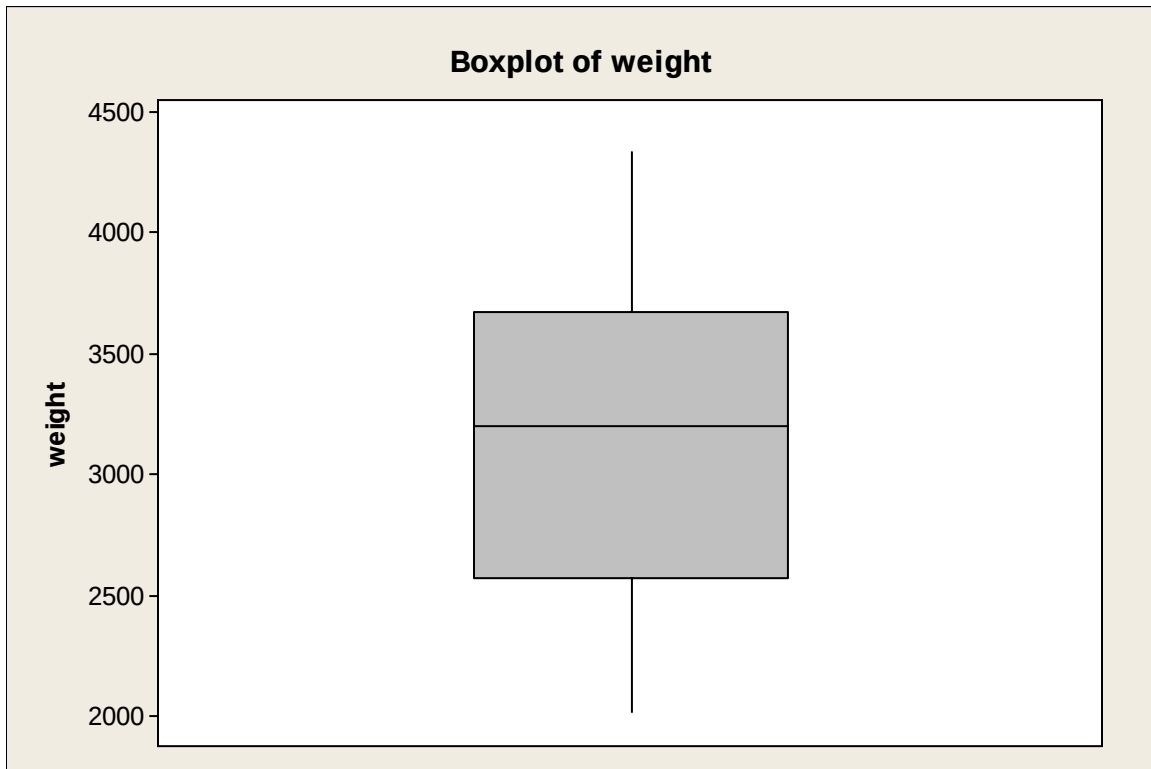
Boxplot of mpg



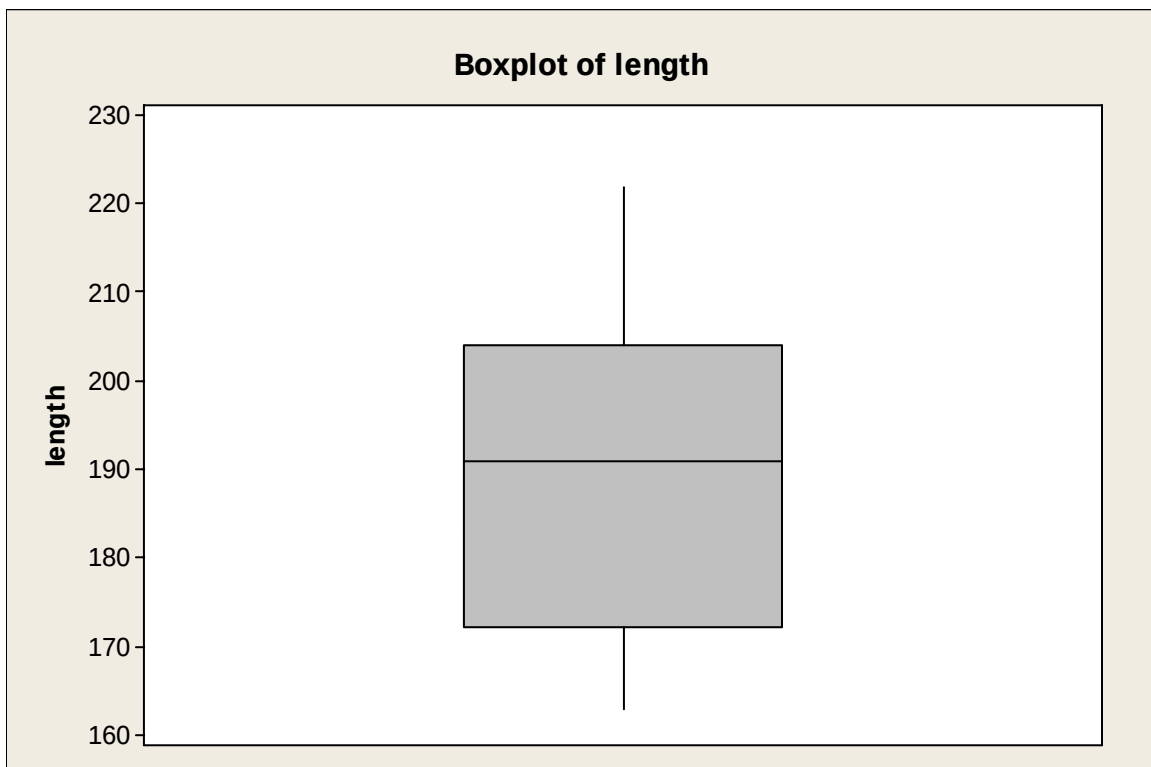
Boxplot of rep78



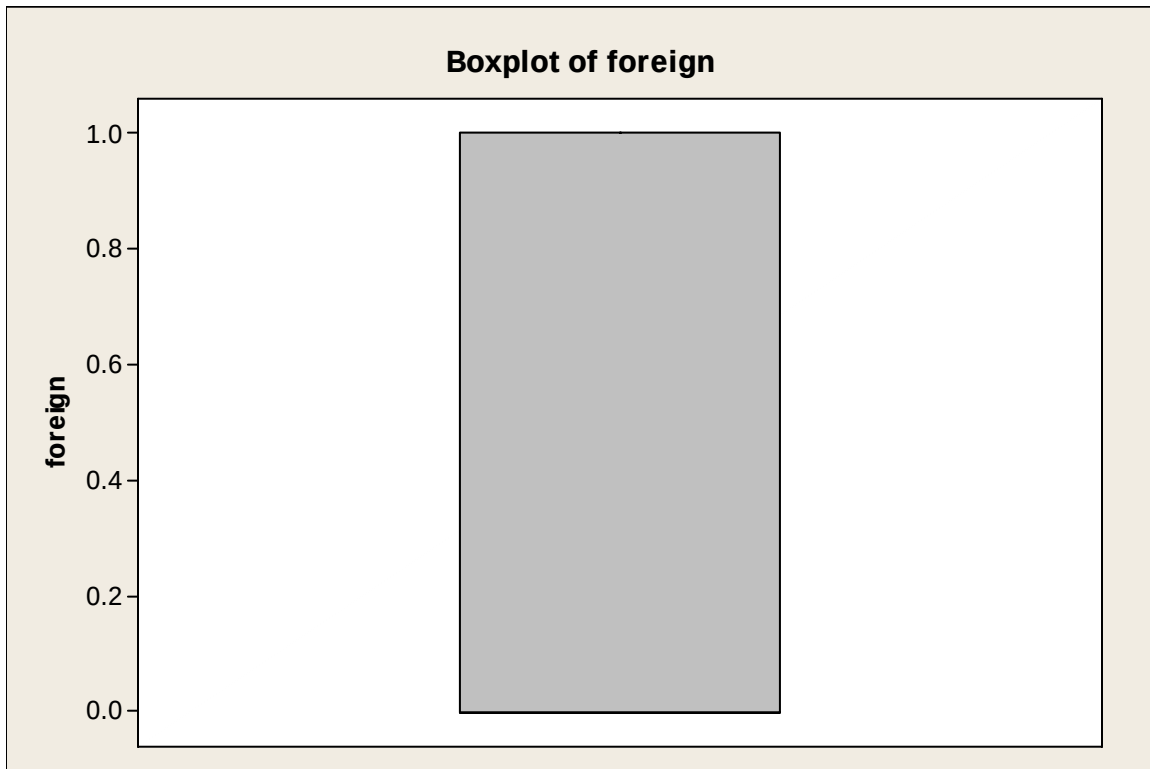
Boxplot of weight



Boxplot of length



Boxplot of foreign



MTB >

الإختبارات الإحصائية

MTB > Onet 'price' 'mpg' 'weight' 'length'.

One-Sample T: price, mpg, weight, length

Variable	N	Mean	StDev	SE Mean	95% CI
price	26	6652	3371	661	(5290, 8013)
mpg	26	20.923	4.758	0.933	(19.001, 22.845)
weight	26	3099	695	136	(2818, 3380)
length	26	190.08	18.17	3.56	(182.74, 197.42)

MTB >

MTB > TwoT 'price' 'foreign'.

Two-Sample T-Test and CI: price, foreign

Two-sample T for price

foreign	N	Mean	StDev	SE Mean
0	19	6484	3768	865
1	7	7107	2102	794

Difference = mu (0) - mu (1)
 Estimate for difference: -622
 95% CI for difference: (-3080, 1835)

T-Test of difference = 0 (vs not =): T-Value = -0.53 P-Value = 0.602
DF = 19

MTB >

تحليل الانحدار

MTB > Regress 'price' 3 'mpg' 'weight' 'length';
SUBC> GFourpack;
SUBC> RType 1;
SUBC> Constant;
SUBC> Brief 2.

Regression Analysis: price versus mpg, weight, length

The regression equation is
price = 6798 - 5 mpg + 4.35 weight - 71.1 length

Predictor	Coef	SE Coef	T	P
Constant	6798	13173	0.52	0.611
mpg	-5.4	211.3	-0.03	0.980
weight	4.353	2.194	1.98	0.060
length	-71.15	77.19	-0.92	0.367

S = 2930.00 R-Sq = 33.5% R-Sq(adj) = 24.5%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	95243488	31747829	3.70	0.027
Residual Error	22	188867731	8584897		
Total	25	284111219			

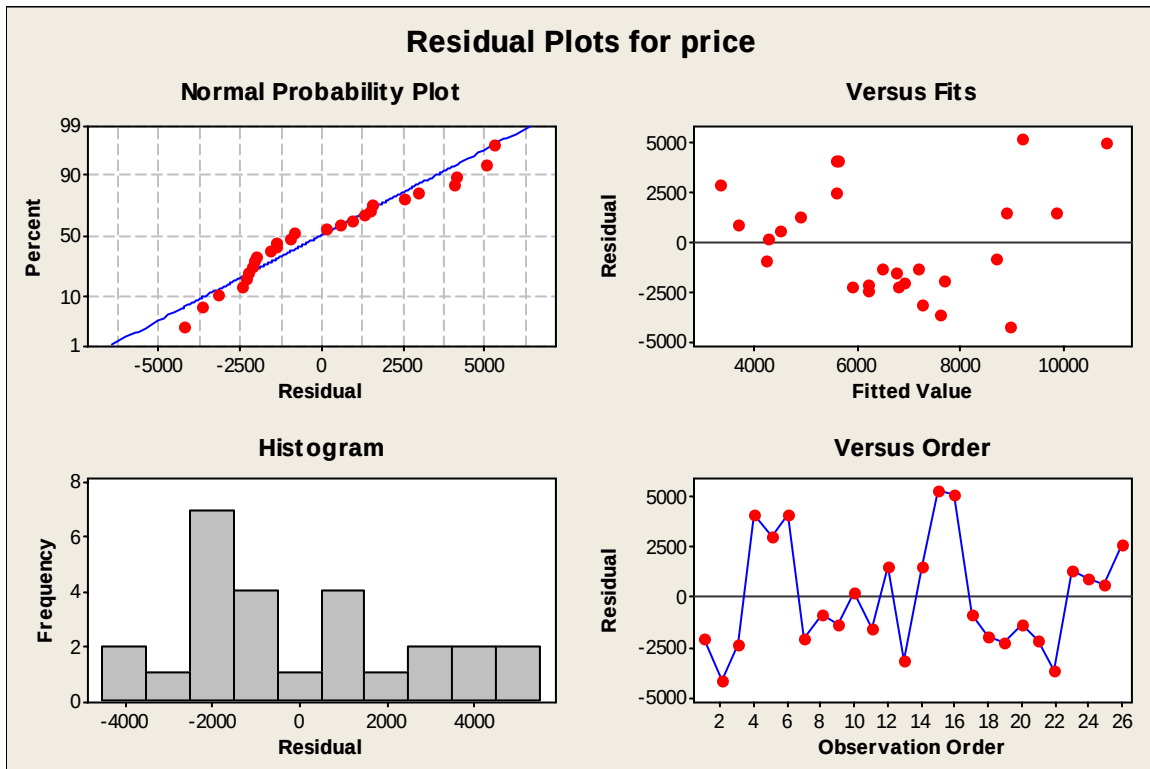
Source	DF	Seq SS
mpg	1	54620028
weight	1	33329492
length	1	7293968

Unusual Observations

Obs	mpg	price	Fit	SE Fit	Residual	St Resid
2	17.0	4749	8980	2022	-4231	-2.00 X
16	21.0	15906	10844	2024	5062	2.39RX
24	35.0	4589	3662	2028	927	0.44 X

R denotes an observation with a large standardized residual.
X denotes an observation whose X value gives it large leverage.

Residual Plots for price



MTB >

تحليل التباين

MTB > Oneway 'price' 'foreign';
SUBC> GFourpack.

One-way ANOVA: price versus foreign

Source	DF	SS	MS	F	P
foreign	1	1981693	1981693	0.17	0.685
Error	24	282129526	11755397		
Total	25	284111219			

S = 3429 R-Sq = 0.70% R-Sq(adj) = 0.00%

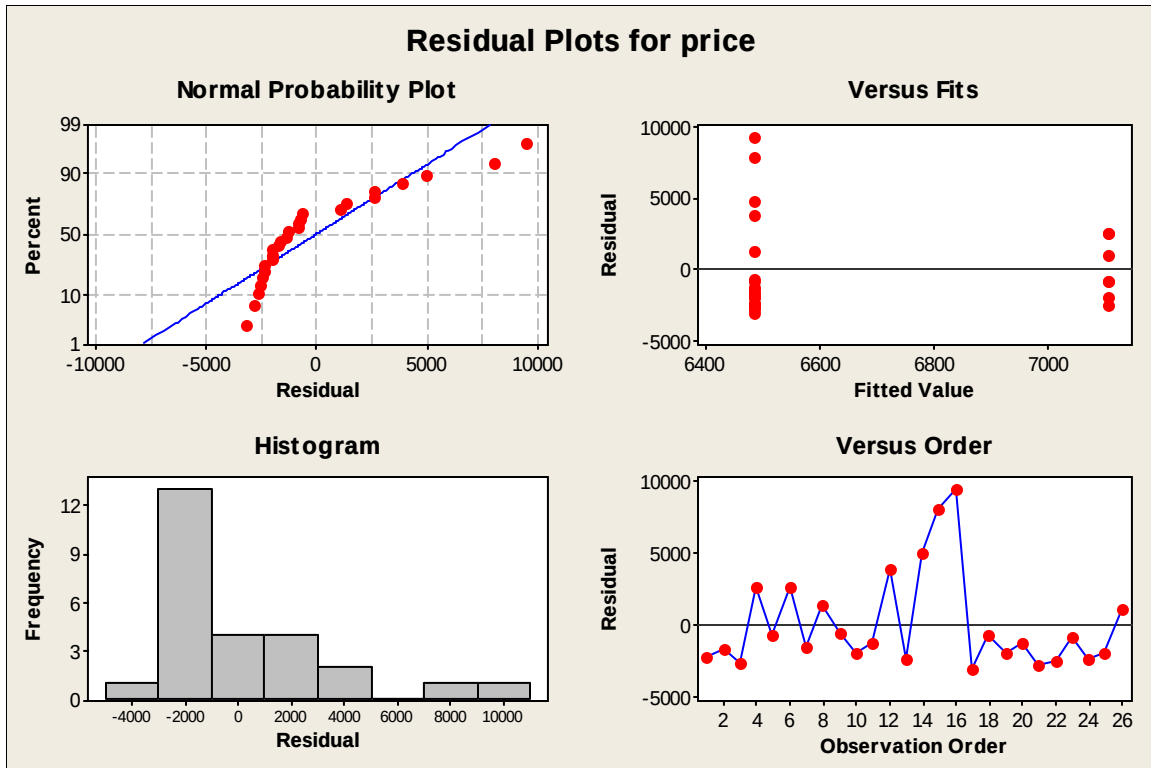
Level	N	Mean	StDev
0	19	6484	3768
1	7	7107	2102

Individual 95% CIs For Mean Based on Pooled StDev

4500 6000 7500 9000

Pooled StDev = 3429

Residual Plots for price



MTB >

الترايط

Correlations: price, mpg, weight, length

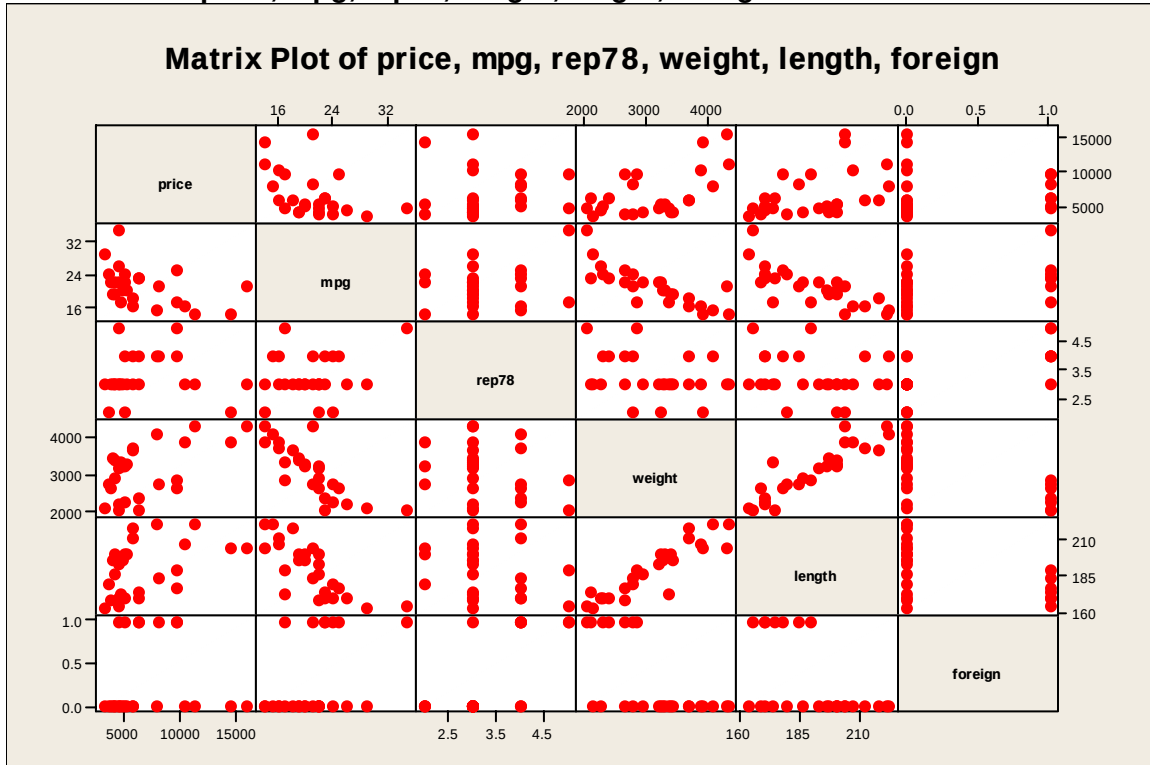
	price	mpg	weight
mpg	-0.438 0.025		
weight	0.556 0.003	-0.808 0.000	
length	0.436 0.026	-0.768 0.000	0.907 0.000

**Cell Contents: Pearson correlation
P-Value**

MTB >

```
MTB > Matrixplot 'price' 'mpg' 'rep78' 'weight' 'length' 'foreign';
SUBC> Symbol.
```

Matrix Plot of price, mpg, rep78, weight, length, foreign



MTB >

ملحق:

تدريج او مستويات القياس Scales of Measurement

لكي نستعرض ونفسر ونحلل البيانات لابد أن نميز بين تدريج أو مستويات قياس المتغيرات.

يوجد أربع تصانيف أساسية لمستويات قياس المتغيرات

1- **التدريج أو التصنيف الإسمي Nominal Scale** ويصف المتغيرات النوعية أو الفئوية Categorical or Qualitative Variables فمثلا المتغير جنس gender يصنف ذكر أو انثى والمتغير جنسية يصنف سعودي أو اماراتي أو عماني أو مصري أو جزائري الخ. في بعض الحزم الإحصائية لابد من إعادة ترميز المتغيرات النوعية حتى يمكن تحليلها فمثلا الذكر يرمز له بالرقم 1 والانثى الرقم 2 فقط لاحظ ان ارقام الترميز هذه لا تعنى اي ترتيب.

2- **التدريج أو التصنيف الترتيبي Ordinal Scale** ويصف ايضا المتغيرات النوعية أو الفئوية وهذا التدريج يوسع المعلومات عن التدريج الإسمي ويعطيه ترتيب فمثلا المتغير المستوى الدراسي يمكن ان يصنف اول أو ثاني أو ثالث وهذا يعطي وصفا وترتيب ايضا المتغير مستوى الرضى عن خدمة يمكن ان يوصف غير راضي أو راضي بعض الشيء أو راضي الخ.

3- **التدريج أو التصنيف الفترتي Interval Scale** ويصنف المتغيرات الكمية Quantitative Variables فبالإضافة الى الترتيب بين قيم المتغير هناك معنى لكمية الفرق بين القياسات. التصنيف الفترتي لا يوجد له نقطة صفر حقيقية فمثلا درجة حرارة 40 مئوية لاتعني ان الحرارة مرتين اعلى من 20 درجة مئوية لأن الصفر في التدريج المؤي اختياري حيث ان درجة الحرارة صفر لاتعنى عدم وجود حرارة.

4- **التدريج أو التصنيف النسبي Ratio Scale** ويصنف المتغيرات الكمية ايضا وهو مثل التدريج الفترتي إلا انه يمتلك صفر حقيقي فمثلا إذا اخذنا الوزن بوحدات الكيلوجرام فإن الفرق في الوزن بين شخصين وزن احدهم 82 كجم والآخر 69 كجم هو نفسه كالفرق في الوزن بين شخصين وزن احدهم 64 كجم و آخر وزنه 51 كجم أي ان الفرق 13 كجم وله نفس المعنى والتفسير ايضا شخص وزنه 100 كجم يزن ضعف شخص وزنه 50 كجم الوزن له صفر حقيقي.

طرق و انواع التحليل الإحصائي على المتغير يعتمد على تدريجه كالتالي:

- 1- المتغير الإسمي لا يوجد معنى لمتوسطه او وسيطه ولكن يوجد له منوال ونسبة.
- 2- المتغير الترتيبي لا يوجد معنى لمتوسطه ولكن يمكن ايجاد وسيطه ومنواله ونسبته.
- 3- المتغيرات الفترتيه والنسبية يمكن ايجاد المتوسط الخ لها.

الربيعات والعشيرات والمئينات Quantiles, Deciles and Percentiles

إذا رتبنا المشاهدات تصاعدياً فإن المشاهدات التي يكون $n\%$ من المشاهدات أقل منها في القيمة تسمى المئين n ويرمز له P_n . المئينات 25 و 50 و 75 تعرف على أنها الربع الأول و الربع الثاني (او الوسيط) والربع الثالث أي

$$Q_1 = P_{25}$$

$$Q_2 = P_{50} = \text{median}$$

$$Q_3 = P_{75}$$

المئينات 10 و 20 و ... و 90 تعرف على أنها العشير الأول والعشير الثاني و ... والعشير التاسع أي

$$D_1 = P_{10}$$

$$D_2 = P_{20}$$

$$D_5 = P_{50} = \text{median}$$

$$D_9 = P_{90}$$

معامل الاختلاف Coefficient of Variation:

أو التشتت النسبي ويعرف كالتالي :

$$C.V. = \frac{s}{\bar{x}}$$

$$C.V. = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

المتغير المعياري والدرجات المعيارية (مقياس التمرکز):

إذا كان لدينا المتغير العشوائي X الذي له القيم الممكنة $x_1, x_2, \dots, x_n$ والتي متوسطها $\bar{x}$ وانحرافها المعياري s فإن المتغير Z والذي له القيم $z_1, z_2, \dots, z_n$ والتي تعطى بالعلاقة التالية :

$$z_i = \frac{x_i - \bar{x}}{s}, \quad i=1, 2, \dots, n$$

حيث z_i تقيس الانحرافات عن المتوسط الحسابي بوحدات من الانحراف المعياري يسمى بالمتغير المعياري ويستخدم للمقارنة بين التوزيعات المختلفة.

مثال

حصل طالب على 82 درجة في مقرر للإحصاء حيث كان متوسط الدرجات هو 75 درجة وانحراف معياري 10 درجات ثم حصل على 89 درجة في مقرر للرياضيات وكان متوسط الدرجات للرياضيات هو 81 درجة وانحراف معياري 16 درجة في أي من المقررين كانت درجة استيعاب هذا الطالب أعلى ؟

الحل

إذا كانت z_1 ترمز للدرجة المعيارية للإحصاء فإن :

$$z_1 = \frac{82 - 75}{10} = 0.7$$

وإذا كانت z_2 ترمز للدرجة المعيارية للرياضيات فإن :

$$z_2 = \frac{89 - 81}{16} = 0.5$$

وهذا يعطي أن إستيعاب الطالب النسبي لمقرر الإحصاء أعلى من الرياضيات .

العزوم Moments :

للمتغير العشوائي X والذي له دالة توزيع $F_X(x)$ معرفة على جميع قيم $-\infty < x < \infty$ يعرف العزوم الـ r حول الصفر

$$E(X^r) = \int_x x^r dF_X(x)$$

والعزم الـ r حول المتوسط

$$E((X - \mu)^r) = \int_x (x - \mu)^r dF_X(x)$$

حيث

$$\mu = E(X) = \int_x x dF_X(x)$$

العزوم من عينة $x_1, x_2, \dots, x_n$

1- العزوم حول المتوسط

يعرف العزم النوني حول المتوسط بالعلاقة

$$m_r = \frac{\sum (x - \bar{x})^r}{n}$$

2- العزوم حول الصفر

$$m'_r = \frac{\sum x^r}{n}$$

مقياس الالتواء (v)

$$v = \frac{m_3}{s^3}$$

$$m_3 = \frac{\sum (x - \bar{x})^3}{n}$$

$$s = \sqrt{\frac{1}{n-1} \sum (x - \bar{x})^2}$$

للتوزيع الطبيعي $v = 0$.

التقلطح k بالعلاقة التالية:

$$k = \frac{m_4}{s^4}$$

حيث :

$$m_4 = \frac{\sum (x - \bar{x})^4}{n}$$

للتوزيع الطبيعي $k=3$.

نظرية النهاية المركزية Central limit Theorem

وتعتبر هذه النظرية أساسية في علم الإحصاء والتي تعطي التوزيع العيني للمتوسط وهي كالتالي :

إذا فرضنا أننا أخذنا عينة حجمها n من مجتمع له متوسط μ وانحراف معياري σ . فإن المتغير العشوائي $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ له الخصائص التالية :

(1) يكون له (تقريباً) التوزيع الطبيعي القياسي إذا كانت $n \geq 30$ مهما كان توزيع المجتمع .

(2) يكون له (تماماً) التوزيع الطبيعي القياسي إذا كان توزيع المجتمع طبيعياً مهما كان حجم العينة .

ملاحظة

أ - إذا كانت $n \geq 30$ فإن $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

ب - إذا كان المجتمع طبيعياً فإن : $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

مهما كانت n .

تقدير معالم المجتمع واختبار الفروض Estimation and Testing Hypothese

التقدير بنقطة

يحتوي اي مجتمع إحصائي على معالم تكون غير معلومة مثل متوسطه μ أو انحرافه المعياري σ أو نسبة معينة R الخ ويمكن إيجاد تقديرات لهذه المعالم من بيانات مأخوذة من عينة عشوائية من هذا المجتمع الإحصائي وذلك بحساب ما يسمى بالإحصاءات (وهي دوال في المشاهدات) فمثلاً متوسط العينة العشوائية $\bar{X}$ يستخدم كمقدر لمتوسط المجتمع μ وكذلك الانحراف المعياري σ يستخدم كمقدر للانحراف المعياري للمجتمع σ وهكذا.... وتسمى هذه التقديرات التقدير بنقطة لأنها قيمة وحيدة محسوبة من العينة.

التقدير بفترة

التقدير بفترة لإحدى معالم المجتمع المجهولة مثل μ أو σ أو R هي عبارة عن إيجاد فترة تحدد بقيمتين تحسب من مشاهدات العينة العشوائية المأخوذة من المجتمع محل الدراسة ، ونتوقع احتواء هذه الفترة على معلمة المجتمع باحتمال معين $(1 - \alpha)$ حيث عادة α تأخذ قيمة صغيرة مثل 0.01, 0.05, 0.1 ويمكن أيضاً إيجاد دقة التقدير بفترة للمعلمة . وكلما كان طول الفترة صغيراً زادت دقة التقدير . لذلك سميت بتقدير فترة الثقة . فإذا كان مثلاً درجة الدقة في الخطأ بين متوسط المجتمع μ ومتوسط العينة العشوائية $\bar{X}$ هو المقدار الموجب ε فإنه يمكن تحديد حد أدنى للاحتمال يكتب كالتالي :

$$P(|\mu - \bar{x}| \leq \varepsilon) \geq 1 - \alpha$$

ويمكن كتابة العلاقة السابقة كالتالي :

$$P(\bar{x} - \varepsilon \leq \mu \leq \bar{x} + \varepsilon) \geq 1 - \alpha$$

- وهذه العلاقة تمثل فترة ثقة $(\bar{x} - \varepsilon, \bar{x} + \varepsilon)$ للمعلمة المجهولة μ باحتمال لا يقل عن $(1 - \alpha)$. والمقدار $(1 - \alpha)$ يسمى درجة الثقة فإذا كانت قيم α هي 0.1, 0.05, 0.01 فإن درجات الثقة المناظرة هي على الترتيب 90 % , 95 % , 99 % .
- وسوف نقوم باستعراض كيفية تقدير فترة الثقة لبعض المعالم المجهولة للمجتمع .
- 1 - تقدير فترة الثقة للمتوسط μ لمجتمع تباينه σ^2 **معروف** أو غير معروف للعينات الكبيرة .
 - 2 - تقدير فترة الثقة للمتوسط μ لمجتمع ذات توزيع طبيعي تباينه σ^2 **معروف** أو غير معروف للعينات الصغيرة .
 - 3 - تقدير فترة الثقة للنسبة R .

وقبل التعرض لدراسة كيفية حساب تقدير فترات الثقة للحالات السابقة . سوف نستعرض فيما يلي ما يسمى بالقيمة العظمى في خطأ التقدير ، والتي تساعدنا في إيجاد حجم العينة عند مستوى دقة معين α ، وكذلك في إيجاد فترات الثقة السابق ذكرها .

القيمة العظمى للخطأ في التقدير

لقد سبق دراسة توزيع المعاينة لمتوسط العينة $\bar{X}$ ووجد أن الخطأ المعياري له $\sigma_{\bar{X}}$ يساوي $\frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$ وذلك عندما يكون حجم العينة n مأخوذاً من مجتمع صغير أو محدود حجمه N . أو يساوي $\frac{\sigma}{\sqrt{n}}$ في المجتمعات الكبيرة جداً أي عندما يكون حجم العينة n يمثل نسبة صغيرة جداً من المجتمع الذي حجمه N . ولقد سبق أيضاً دراسة نظرية النهاية المركزية . والتي تقول إن توزيع المعاينة للمتوسط يقترب من التوزيع الطبيعي وأنه يمكن أن يؤكد احتمال قدره $(1 - \alpha)$ بأن متوسط العينة $\bar{X}$ يختلف عن

متوسط المجتمع μ بمقدار يقل عن $z_{\alpha/2}$ من الخطأ المعياري $\sigma_{\bar{x}}$. ويمكن التعبير عما سبق بالعلاقة الاحتمالية التالية :

$$P\left(\left|\bar{X} - \mu\right| \leq z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

وذلك في حالة المجتمعات ذات الحجم الكبير أو الانهائية. والمقدار $|\bar{X} - \mu|$ يسمى خطأ التقدير ويرمز له بالرمز E وذلك عندما يزيد مقدار الخطأ في التقدير إلى نهايته العظمى. فمن المعادلة السابقة نجد أن خطأ التقدير يأخذ القيم التالية:

$$E = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

أحياناً تسمى القيمة العظمى في خطأ التقدير بالدقة في التقدير.

حجم العينة

عند تحديد مقدار الدقة المطلوب أو القيمة العظمى للخطأ في التقدير عند احتمال معين $(1-\alpha)$ يمكن تحديد حجم العينة n من المعادلة السابقة ويكون كالتالي :

$$n = \left(\frac{z_{\alpha/2} \cdot \sigma}{E}\right)^2$$

تقدير فترة الثقة للمتوسط

في العينات الكبيرة

يمكننا أن نقول إنه باحتمال $(1-\alpha)$ أن :

$$|\bar{x} - \mu| \leq z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

أي أن :

$$-z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \bar{x} - \mu \leq z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

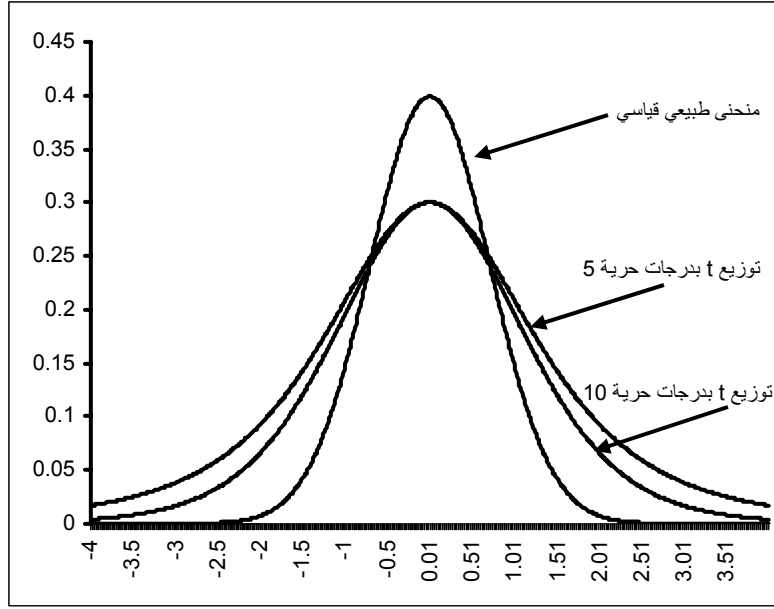
أي أن :

$$\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

وهذا يعني بأن متوسط المجتمع μ واقع داخل الفترة الممتدة من الحد الأعلى $\bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ إلى الحد الأدنى $\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ ويسمى هذا بتقدير فترة الثقة عند مستوى معنوي أو بدرجة ثقة قدرها $(1-\alpha)$ 100 . فإذا كانت $\alpha = 0.05$ فإن درجة الثقة هي 95% . وإذا كانت $\alpha = 0.01$ فإن درجة الثقة تكون 90% وهكذا .

في العينات الصغيرة

توزيع المعاينة للمتوسط يقترب من التوزيع الطبيعي عندما يكون **الانحراف المعياري σ للمجتمع معلوماً أو غير معلوم للعينات الكبيرة أو معلوماً للعينات الصغيرة من مجتمع ذا توزيع طبيعي** ولكن عندما يكون الانحراف المعياري للمجتمع الطبيعي σ مجهولاً ونستعوض عنه بالانحراف المعياري المقدر من العينة s و توزيع المعاينة يكون في هذه الحالة توزيع t و هو توزيع متماثل ويختلف عن التوزيع الطبيعي . وتوزيع t يشبه التوزيع الطبيعي ويقترب من التوزيع الطبيعي القياسي كلما كبر حجم العينة n أو زادت درجات الحرية ν حيث $\nu = n - 1$ وينطبق على التوزيع الطبيعي عندما يصبح $\nu = 30$. ونوضح شكل توزيع t عند درجات حرية 5 , 10 ومنحنى التوزيع الطبيعي القياسي بالشكل التالي :



شكل يمثل توزيع t مقارنة بالتوزيع الطبيعي

ويمكن كتابة فترة الثقة للعينات الصغيرة باحتمال قدره $(1-\alpha)$ لمتوسط المجتمع μ مثل ما تم بالنسبة للعينات الكبيرة كالتالي :

$$\bar{x} - t_{\left(\frac{\alpha}{2}, \nu\right)} \left(\frac{s}{\sqrt{n}} \right) < \mu < \bar{x} + t_{\left(\frac{\alpha}{2}, \nu\right)} \left(\frac{s}{\sqrt{n}} \right)$$

حيث استبدلنا $Z_{\alpha/2}$ بالقيمة $t_{\alpha/2}$. واستبدلنا الانحراف المعياري σ للمجتمع بالانحراف المعياري للعينة s . و $t_{\alpha/2}$ تعطى من الحزم الإحصائية حيث تعطى قيم t لكل درجة من درجات الحرية $\nu = n-1$ وذلك لبعض قيم α مثل 0.1, 0.05, 0.01,

التقدير باستخدام الحزم الإحصائية:

SAS

```
DATA height;
    INPUT x @@;
    DATALINES;
56 70 53 71 73 50 46 54 44
48 61 61 48 53 62 57 61 73
59 49 52 65 55 78 80 72 52
```



```

71 56 70 62 51 67 59 53 55
46 70 54 60 63 65 60 56 64
56 72 66 63 67 34 72 60 62
44 62 56 67 43 47 47 55 73
48 67 72 46 58 68 49 35 71
74 65 45 57 48 71 69 69 44
57 43 68 58 49 57 75 55 66
59 75 74 51 48 62 52 50 63
73

```

```

;
run;
proc print data = height;
run;
proc means data= height n mean stddev clm alpha=0.05;
run;

```

المخرجات:

February 28, 2013 1

The SAS System

10:20 Thursday,

Obs	x
1	56
2	70
3	53
4	71
5	73
6	50
7	46
8	54
9	44
10	48
11	61
12	61
13	48
14	53
15	62
16	57
17	61
18	73
19	59
20	49
21	52
22	65
23	55
24	78
25	80
26	72
27	52
28	71

29	56
30	70
31	62
32	51
33	67
34	59
35	53
36	55
37	46
38	70
39	54
40	60
41	63
42	65
43	60
44	56
45	64
46	56
47	72
48	66
49	63
50	67
51	34
52	72

February 28, 2013 2

The SAS System

10:20 Thursday,

Obs	x
53	60
54	62
55	44
56	62
57	56
58	67
59	43
60	47
61	47
62	55
63	73
64	48
65	67
66	72
67	46
68	58
69	68
70	49
71	35
72	71
73	74
74	65
75	45
76	57
77	48

R

```
> height = read.table("clipboard", sep=" ", header=TRUE)
> height
      height
1         56
2         70
.....
99        63
100       73
> n = length(height)
> a = mean(height)
> b = sd(height)
> err = qt(0.975, df = n-1)*b/sqrt(n)
> lft = a - err
> rt = a + err
> lft
      height
57.21374
> rt
      height
61.22626
> t.test(height)
```

One Sample t-test

data: height

t = 58.5694, df = 99, p-value < 2.2e-16

alternative hypothesis: true mean is not equal to 0

95 percent confidence interval:

57.21374 61.22626

sample estimates:

mean of x

59.22

SPSS

DATA LIST LIST

/ height.

BEGIN DATA

```
56 70 53 71 73 50 46 54 44 48 61 61 48 53 62 57 61
73 59 49 52 65 55 78 80 72 52 71 56 70 62 51 67 59
53 55 46 70 54 60 63 65 60 56 64 56 72 66 63 67 34
72 60 62 44 62 56 67 43 47 47 55 73 48 67 72 46 58
68 49 35 71 74 65 45 57 48 71 69 69 44 57 43 68 58
49 57 75 55 66 59 75 74 51 48 62 52 50 63 73
```

END DATA.

LIST.

DESCRIPTIVES

/VAR=height.

المخرجات:

Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
height	100	34.00	80.00	59.2200	10.11108
Valid N (listwise)	100				

EXAMINE

/VARIABLES=height.

Descriptives				
		Statistic	Std. Error	
height	Mean	59.2200	1.01111	
	95% Confidence Interval for Mean	Lower Bound	57.2137	

	Upper Bound	61.2263	
5% Trimmed Mean		59.3444	
Median		59.0000	
Variance		102.234	
Std. Deviation		10.11108	
Minimum		34.00	
Maximum		80.00	
Range		46.00	
Interquartile Range		16.50	
Skewness		-.126	.241
Kurtosis		-.701	.478

Minitab

```
MTB > set c1
DATA> 56 70 53 71 73 50 46 54 44
DATA> 48 61 61 48 53 62 57 61 73
DATA> 59 49 52 65 55 78 80 72 52
DATA> 71 56 70 62 51 67 59 53 55
DATA> 46 70 54 60 63 65 60 56 64
DATA> 56 72 66 63 67 34 72 60 62
DATA> 44 62 56 67 43 47 47 55 73
DATA> 48 67 72 46 58 68 49 35 71
DATA> 74 65 45 57 48 71 69 69 44
DATA> 57 43 68 58 49 57 75 55 66
DATA> 59 75 74 51 48 62 52 50 63
DATA> 73
DATA> end
MTB > Describe 'height';
SUBC> Mean;
SUBC> SEMean;
SUBC> StDeviation;
SUBC> Variance;
SUBC> CVariation;
SUBC> QOne;
SUBC> Median;
SUBC> QThree;
SUBC> IQRRange;
SUBC> Mode;
SUBC> TRMean;
SUBC> Sums;
SUBC> Minimum;
SUBC> Maximum;
SUBC> Range;
SUBC> SSQ;
SUBC> Skewness;
SUBC> Kurtosis;
SUBC> MSSD;
SUBC> N;
SUBC> NMissing;
SUBC> Count;
SUBC> CumN;
SUBC> Percent;
SUBC> CumPercent.
```


Descriptive Statistics: height

Variable	Total Count	N	N*	CumN	Percent	CumPct	Mean	SE Mean	TrMean	StDev
height	100	100	0	100	100	100	59.22	1.01	59.34	10.11

Variable	Variance	CoefVar	Sum	Sum of Squares	Minimum	Q1	Median	Q3
height	102.23	17.07	5922.00	360822.00	34.00	51.25	59.00	67.75

Variable	Maximum	Range	IQR	Mode	N for Mode	Skewness	Kurtosis	MSSD
height	80.00	46.00	16.50	48, 56, 62	5	-0.13	-0.70	101.66

MTB >

Excel

المخرجات:

<i>height</i>	
Mean	59.22
Standard Error	1.011108003
Median	59
Mode	56
Standard Deviation	10.11108003
Sample Variance	102.2339394
Kurtosis	-0.701168047
Skewness	-0.126404815
Range	46
Minimum	34
Maximum	80
Sum	5922
Count	100
Largest(75)	52
Smallest(25)	51
Confidence Level(95.0%)	2.006257588

اختبارات الفروض الإحصائية Test of Statistical Hypothesis

الفرع الثاني من الاستدلال الإحصائي هو اختبارات الفروض . يحاول الباحث في كثير من الأحيان اتخاذ قرار بشأن خواص توزيع مجتمع ما ، وذلك بناءً على بيانات عينة عشوائية اختيرت من المجتمع نفسه . فمثلاً يريد طبيب أن يختبر فعالية دواء جديد بالنسبة لعلاج مرض معين . أو يريد باحث في التربية أن يختبر كون منهج معين أكثر فائدة من منهج آخر الخ . وهكذا وللوصول إلى هذه القرارات الإحصائية (Statistical Decision) نقوم عادة بوضع فروض عن خواص المجتمع ، متوسطه ، انحرافه المعياري ... الخ . ونختبر هذا الفرض بناءً على عينة عشوائية نختارها من المجتمع ، وهذه الفروض - التي قد تكون صحيحة ، أو غير صحيحة - تسمى بالفروض الإحصائية وتنقسم الفروض إلى قسمين :

i () فروض عن معالم المجتمع (Parametric) .

ii () فروض عن صورة دالة التوزيع (Nonparametric) .

وسوف نكتفي بدراسة الفروض عن بعض معالم المجتمع فقط .

ونبدأ بافتراض إحصائي يسمى **فرض العدم (Null Hypothesis)** (وسمي بهذا لأن الغرض منه هو عدم قبوله أو محوه) ويرمز له بـ H_0 ، ويسمى الافتراض الذي يختلف عن H_0 **بفرض البديل** ويرمز له بـ H_1 (Alternative) .

وكما سبق أنوضحنا بأن اختبار الفروض يعتمد على بيانات العينة وهي تعني بالإجابة على السؤال التالي : هل بيانات العينة متناسقة مع فرض معين ؟ وهل تميل إلى تأكيده أو نفيه ؟

إذا فرضنا قيمة لمعلم من معالم المجتمع (مثلاً المتوسط) وأخذنا عينة عشوائية من المجتمع فسيكون هناك غالباً فرق راجع إلى مجرد الصدفة والخطأ المتوقع نتيجة العينة ؟ أم أنه فرق حقيقي معنوي (Significant) ، وأن هناك أسباباً أخرى ساعدت على كبر

هذا الفرق ؟ فإذا وجدنا أن الفرق معنوي فإننا نميل إلى أن الفرض H_0 قد يكون غير صحيح (أو على الأقل نرفضه بناء على البيانات المتاحة) . وإذا كان الفرق غير معنوي (Nonsignificant) فليس هناك ما يدعونا لرفض فرض العدم H_0 .
 فمثلاً إذا قذفنا قطعة نقود 20 مرة وكانت النتيجة المشاهدة هي 16 للصورة و 4 للكتابة ، فإننا قطعاً نميل إلى رفض H_0 بأن العملة متزنة ، على الرغم من أننا قد نكون مخطئين في اتخاذ هذا القرار الإحصائي . ولكن إن كانت النتيجة مثلاً 11 صورة و 9 كتابة فليس هناك ما يبرر رفض H_0 .

وكمثال آخر لفرض العدم (الفرضية الأولية) . إذا افترض أحد الباحثين بأن أطوال الذكور البالغين في إحدى البلاد هو 165 سم واختار عينة مكونة من 64 ذكراً بالغاً . وحسب متوسط الطول لهم فكان 160 سم ، وكان الانحراف المعياري لهذا المجتمع معروفاً ويساوي 5 سم . فيكون فرض العدم في هذه الحالة $H_0 : \mu = 165 \text{ cm}$ (أو الفرضية الأولية H_0 أن متوسط المجتمع الذي سحبت منه العينة هو 165 سم) . حيث نفترض عدم وجود فروق حقيقية بين متوسط المجتمع والقيمة المفروضة . والفروق المشاهدة إنما تعزى للصدفة ويقابل الفرضية الأولية فرضية أخرى تسمى بالفرضية البديلة ونرمز لها بالرمز H_1 ، في حالة متوسط أطوال الذكور البالغين يمكن أن يأخذ الفرض البديل إحدى الحالات التالية :

$$\mu \neq 165 \text{ أو } \mu < 165 \text{ أو } \mu > 165 \text{ أو } \mu = 170 \text{ الخ}$$

اختبارات المعنوية

ولاختبار صحة الفرضية الأولية H_0 يجب علينا تكوين إحصاءة وهي دالة من مشاهدات العينة العشوائية ومعالم H_0 مثل $z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$ ، وعادة يكون توزيع الإحصاءة معروفاً ويقسم المجال المقابل لهذه الإحصاءة إلى منطقتين ، المنطقة الأولى يمكن تسميتها منطقة عدم الرفض وهي التي يكون فيها احتمال حدوث قيم الإحصاءة والذي

هو $1-\alpha$ كبيراً عندما تكون الفرضية الأولية H_0 صحيحة . والمنطقة الثانية تسمى بمنطقة الرفض وهي التي يكون احتمال حدوث قيم الإحصاءة والذي هو α صغيراً عندما تكون الفرضية الأولية صحيحة . وفي حالة متوسط أطوال الذكور البالغين نأخذ

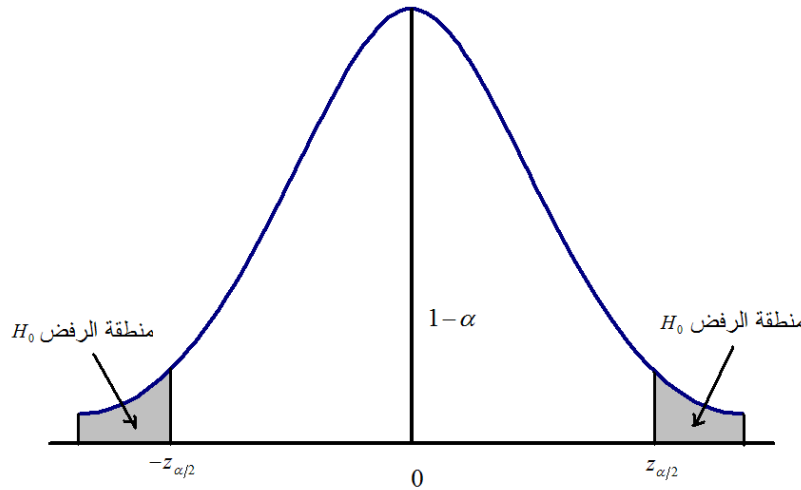
الإحصائية $Z_0 = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}}$ حيث سبق معرفة توزيعها بأنه يقترب من التوزيع

الطبيعي القياسي $N(0, 1)$ ويمكن توضيح منطقة القبول والرفض للفرضية $H_0 : \mu = 160$ على أساس صحة الفرضية H_0 مع الأخذ في الاعتبار الفرضية البديلة H_1 في الحالات التالية :

$$H_0 : \mu = 160$$

$$H_1 : \mu \neq 160$$

في هذه الحالة الفرض البديل H_1 له طرفان كما هو موضح بالمنطقة المظللة في الرسم التالي .



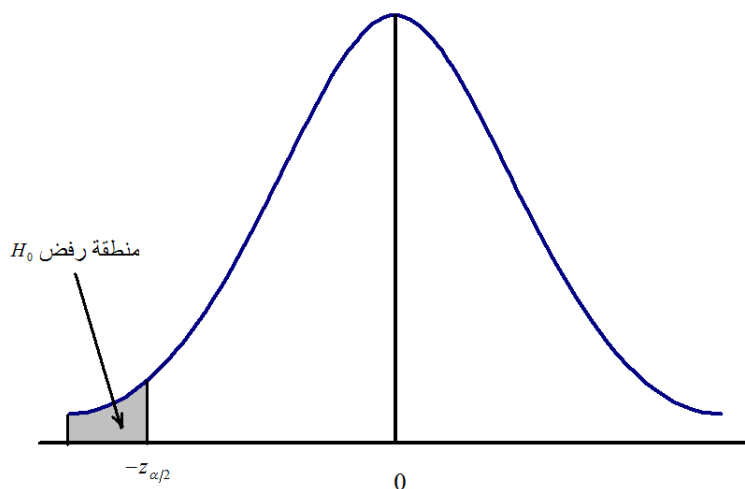
شكل يمثل منطقة الرفض للفرض H_0 من الطرفين

منطقة عدم الرفض H_0 هي $-z_{\alpha/2} \leq z_0 \leq z_{\alpha/2}$ وهي غير مظلمة في شكل السابق .

$$H_0 : \mu = 160$$

$$H_1 : \mu < 160$$

الفرض البديل H_1 له طرف واحد أدنى ونوضح ذلك بالشكل التالي :



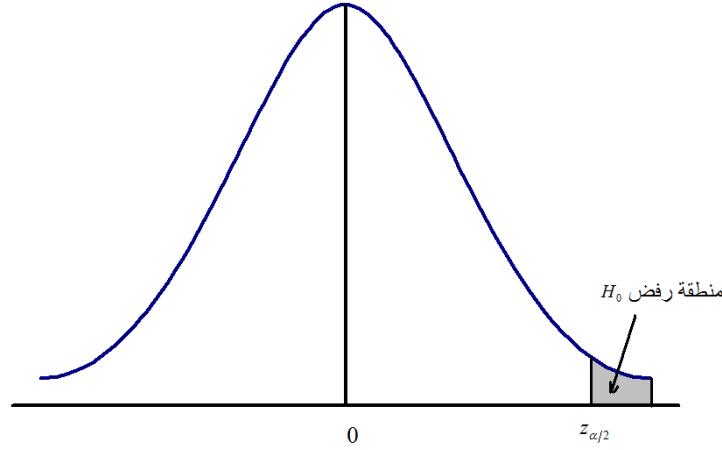
شكل يمثل منطقة الرفض للفرض H_0 من الطرف الأيسر

منطقة عدم رفض H_0 هي $z_0 > z_{\alpha}$.

$$H_0 : \mu = 160$$

$$H_1 : \mu > 160$$

الفرض البديل له طرف واحد من أعلى ونوضح ذلك بالشكل التالي :



شكل يمثل منطقة الرفض للفرض H_0 من الطرف الأيمن

منطقة عدم رفض H_0 هي $z_0 < z_{\alpha}$.

وعموماً إن كانت الفرضية الأولية $H_0: \mu = \mu_0$ فإنه يمكن تلخيص مناطق الرفض وعدم الرفض للفرضية الأولية وذلك حسب الفرض البديل H_1 . وذلك لقيم الإحصاء z_0 المحسوبة من العينة كالتالي :

$\mu > \mu_0$	$\mu < \mu_0$	$\mu \neq \mu_0$	H_1 الفرض البديل
$z_0 < z_{\alpha}$	$z_0 < -z_{\alpha}$	$z_0 < -z_{\alpha/2}$ أو $z_0 < z_{\alpha/2}$	منطقة الرفض لـ H_0
$z_0 < z_{\alpha}$	$z_0 < -z_{\alpha}$	$-z_{\alpha/2} < z_0 < z_{\alpha/2}$	منطقة عدم الرفض لـ H_0

الخطأ من النوع الأول α والخطأ من النوع الثاني β

أي قرار إحصائي يترتب عليه نوعان من الخطأ . خطأ من النوع الأول : هو عندما نرفض الفرضية الأولية H_0 وهي صحيحة ، ويكون احتمال هذا الخطأ هو قيمة α ، وهي عادة ما تكون صغيرة ومثال على ذلك لقيم α , 0.01, 0.05, 0.1 وأحياناً تسمى α مستوى المعنوية . وعندما لانرفض الفرضية الأولية H_0 وهي خاطئة نكون قد ارتكبنا خطأ من النوع الثاني ، ويكون احتمال هذا الخطأ هو القيمة β . ويمكن تلخيص ذلك في الجدول التالي :

القرار	H_0 قبول	H_0 رفض
H_0 الفرضية	صحيحة H_0 الفرضية	خطأ من النوع الأول α
H_0 خطأ الفرضية	خطأ من النوع الثاني β	قرار صحيح

اختبار الفرضيات للمتوسط μ

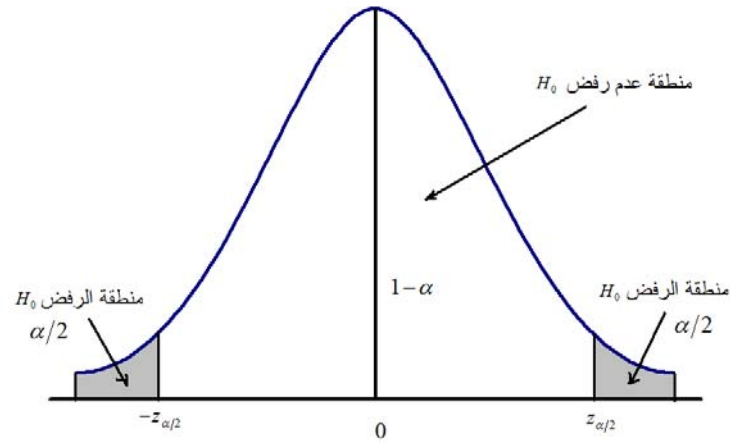
عندما يكون **الانحراف المعياري σ للمجتمع معلوماً** فإن الإحصائية $\frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$ يكون لها

توزيع يقترب من التوزيع الطبيعي القياسي $N(0, 1)$ وبذلك يمكننا من تكوين الفرضية الأولية (فرض عدم) H_0 والفرضية البديلة H_1 كما يلي :

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$

حيث الفرض البديل له طرفان كما هو موضح بالرسم التالي :

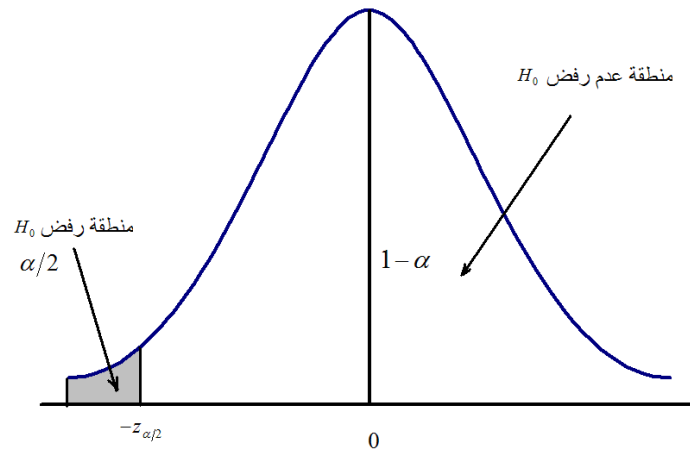


شكل يمثل منطقة الرفض وعدم الرفض للفرض H_0

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu < \mu_0$$

حيث الفرض البديل له طرف واحد كما هو موضح بالرسم التالي:



شكل يمثل منطقة الرفض للفرض H_0 للطرف الأدنى

ثم تكون الإحصائية $\left(z_0 = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \right)$ باستخدام البيانات المشاهدة من العينة

ثم نختار مستوى المعنوية α الذي على أساسه نحدد القيم الحرجة $z_{\alpha/2}, -z_{\alpha/2}$ عندما تكون الفرضية البديلة H_1 من طرفين ونوضح ذلك بالمثال التالي :

درجات الحرية

إن كان لدينا مجتمع ما ويوجد به عدد من المعالم . ونرغب في تقدير هذه المعالم من عينة تحتوي على n من البيانات المستقلة . فإن درجة الحرية التي يرمز لها بالرمز ν تساوي عدد البيانات المستقلة للعينة مطروحاً منها عدد المعالم للمجتمع التي يجب تقديرها من العينة ويعبر عن ذلك رياضياً كما يلي :

$$\nu = n - k$$

حيث k تساوي عدد المعالم التي يجب تقديرها . فإن كان المطلوب تقدير المتوسط μ بـ $\bar{X}$ من العينة فإن $k = 1$ وبالتالي $(\nu = n - 1)$ وإن كان المطلوب تقدير كل من μ و σ فإن $k = 2$ أي أن درجات الحرية في هذه الحالة $(\nu = n - 2)$ وهكذا

إختبارات t :

إختبار t يستخدم لتثمين الفرق بين متوسطين لمجموعتين مستقلتين من البيانات قد يكون أحد المتوسطين مفترض نظرياً . فمثلاً يمكن إختبار نتائج فحص مجموعتين من المرضى اعطي لمجموعة دواء حقيقي و لآخرى دواء مزيف اونتائج فحص مجموعة من المرضى مع نتيجة او معيار سابق. ويستخدم إختبار t على عينة مسحوبة من توزيع طبيعي غير معلوم تباينه وعلى اساس ان العينات مسحوبة من مجتمعين لهم نفس التباين.

إختبار t للعينات المستقلة:

ويستخدم لإختبار الفرضية الصفرية لتساوي متوسطات مجتمعين أي

$$H_0 : \mu_1 = \mu_2$$

عند توفر عينات مستقلة من كل من المجتمعين. كما أن المتغير الذي نقارنه يفترض ان له توزيع طبيعي وله نفس الانحراف المعياري في كلا المجتمعين.
إحصائية الاختبار هي

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

حيث $\bar{x}_1$ و $\bar{x}_2$ متوسطات العينات و n_1 و n_2 احجام العينات و s الانحراف المعياري المجمع والذي يحسب من العلاقة

$$s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

حيث s_1 و s_2 الانحرافات المعيارية المحسوبة من العينات.

تحت الفرضية الصفرية الإحصائية t لها توزيع t بدرجات حرية $n_1 + n_2 - 2$.

فترة الثقة الناتجة من إختبار عند مستوى معنوية α يعطى بالعلاقة

$$(\bar{x}_1 - \bar{x}_2) \pm t_\alpha s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

حيث t_α هي القيمة الحرجة لإختبار بذيولين بدرجات حرية $n_1 + n_2 - 2$.

إختبار t للعينات المتزاوجة:

الإختبار المتزاوج Paired t-test يستخدم لإختبار متوسطات مجتمعين والتي يكون فيها كل فرد من المجتمع الأول متزاوج (مرتبط) بفرد من المجتمع الثاني فمثلا مقارنة أوزان لأطفال اوزانهم فوق العادة مع أخوانهم العاديين أو مقارنة حالة مريض قبل وبعد العلاج.

إذا رمزنا للمتغير الذي نهتم به بالرمز x فيكون x_1 و x_2 رموز للمتغير للقياسات الأولى و الثانية على التوالي ونرمز للزوج i بالرموز x_{1i} و x_{2i} ويكون الفرق $d_i = x_{1i} - x_{2i}$ و الذي نفترض ان له توزيع طبيعي. الفرضية الصفرية هي

$$H_0 : \mu_d = 0$$

و إحصائية الاختبار

$$t = \frac{\bar{d}}{s_d / \sqrt{n}}$$

حيث $\bar{d}$ متوسط الفروق d_i و s_d الانحراف المعياري للفروق d_i و n عدد الأزواج. تحت الفرضية الصفرية الإحصائية t لها توزيع t بدرجات حرية $n - 1$. فترة ثقة $100(1 - \alpha)\%$ تعطى بالعلاقة

$$\bar{d} \pm t_\alpha s_d / \sqrt{n}$$

حيث t_α هي القيمة الحرجة لإختبار بذيلين بدرجات حرية $n - 1$.
مثال:

SAS:

```
DATA bodyfat;
  INPUT gender $ fatpct @@;
  DATALINES;
m 13.3 f 22 m 19 f 26 m 20 f 16 m 8 f 12 m 18 f
21.7 m 22 f 23.2 m 20 f 21 m 31 f 28 m 21 f 30 m 12
f 23 m 16 m 12 m 24
;
RUN;
PROC PRINT DATA = bodyfat;
RUN;
ods select equality;
proc ttest data=bodyfat;
  class gender;
  var fatpct;
```

```

run;
ods select conflimits;
proc ttest data=bodyfat alpha=0.10;
  class gender;
  var fatpct;
run;

```

المخرجات:

February 27, 2013 1

The SAS System

18:25 Thursday,

Obs	gender	fatpct
1	m	13.3
2	f	22.0
3	m	19.0
4	f	26.0
5	m	20.0
6	f	16.0
7	m	8.0
8	f	12.0
9	m	18.0
10	f	21.7
11	m	22.0
12	f	23.2
13	m	20.0
14	f	21.0
15	m	31.0
16	f	28.0
17	m	21.0
18	f	30.0
19	m	12.0
20	f	23.0
21	m	16.0
22	m	12.0
23	m	24.0

February 27, 2013 2

The SAS System

18:25 Thursday,

The TTEST Procedure

Equality of Variances

Variable	Method	Num DF	Den DF	F Value	Pr > F
fatpct	Folded F	12	9	1.29	0.7182

R

```
> male = c(13.3, 19, 20, 8, 18, 22, 20, 31, 21, 12, 16, 12, 24)
> female = c(22, 26, 16, 12, 21.7, 23.2, 21, 28, 30, 23)
> t.test(male, female)
```

Welch Two Sample t-test

```
data: male and female
t = -1.7336, df = 20.539, p-value = 0.09798
alternative hypothesis: true difference in means is not
equal to 0
95 percent confidence interval:
 -9.0538774  0.8277235
sample estimates:
mean of x mean of y
 18.17692  22.29000

>
```

SPSS

DATA LIST LIST

/ gender (A1) fatpct.

BEGIN DATA

"m" 13.3

"f" 22

"m" 19

"f" 26

"m" 20

"f" 16

"m" 8

"f" 12

"m" 18

"f" 21.7

"m" 22

"f" 23.2

"m" 20

"f" 21

"m" 31

"f" 28

"m" 21

"f" 30

"m" 12

"f" 23

"m" 16

"m" 12

"m" 24

END DATA.

LIST.

T-TEST


```

/GROUPS=gender ("m" "f")
/VARIABLES=fatpct.

```

المخرجات:

T-Test

Group Statistics

gender	N	Mean	Std. Deviation	Std. Error Mean
fatpct m	13	18.1769	6.03243	1.67310
f	10	22.2900	5.31966	1.68222

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
fatpct	Equal variances assumed	.293	.594	-1.704	21	.103	-4.11308	2.41345	-9.13212	.90597
	Equal variances not assumed			-1.734					-9.05388	.82772

Minitab

```
MTB > set c1
DATA> 13.3, 19, 20, 8, 18, 22, 20, 31, 21, 12, 16, 12, 24
DATA> end
MTB > set c2
DATA> 22, 26, 16, 12, 21.7, 23.2, 21, 28, 30, 23
DATA> end
MTB > TwoSample 'male' 'female'.
```

المخرجات:

Two-Sample T-Test and CI: male, female

Two-sample T for male vs female

	N	Mean	StDev	SE Mean
male	13	18.18	6.03	1.7
female	10	22.29	5.32	1.7

Difference = μ (male) - μ (female)
Estimate for difference: -4.11
95% CI for difference: (-9.06, 0.84)
T-Test of difference = 0 (vs not =): T-Value = -1.73 P-Value = 0.098 DF = 20

MTB >

Excel

المخرجات:

m		f	t-Test: Two-Sample Assuming Unequal Variances		
13.3	22				
19	26				
20	16				
				<i>m</i>	<i>f</i>
			Mean	18.17692308	22.29
					28.2987777
8	12		Variance	36.39025641	8
18	21.7		Observations	13	10
22	23.2		Hypothesized Mean Difference	0	
20	21		df	21	
				-	
31	28		t Stat	1.733589466	
21	30		P(T<=t) one-tail	0.048825134	
12	23		t Critical one-tail	1.720742871	
16			P(T<=t) two-tail	0.097650268	
12			t Critical two-tail	2.079613837	
24					

إختبارات مربع كاي Chi Square Tests:

إختبارات حسن المطابقة Goodness of Fit Test:

لنفترض اننا اخذنا عينة من 100 طفل تمت ولادتهم في مستشفى الجامعة فوجدنا 66 طفلا من نوع الذكور و 34 طفلا من نوع الإناث و المفترض نظريا ان يكون عدد الذكور وعدد الإناث متساويا فهل هذا الاختلاف عائد للمعاينة و الصدفة أم أن هناك إختلاف بيئي و إجتماعي.

إذا فرضيتنا هي ان نسبة عدد الذكور لعدد الإناث هي 1:1 أي في عينة من 100 طفل نتوقع أن يكون عدد الذكور 50 وعدد الإناث 50 ولكي نختبر هذا نكون الإحصائية

$$\chi_0^2 = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2}$$

حيث O_i العدد المشاهد من النوع i , $i = 1, 2$ (1 ذكر و 2 انثى) و E_i العدد المتوقع.

للمثال السابق

$$\chi_0^2 = \frac{(66-50)^2}{50} + \frac{(34-50)^2}{50} = 5.12 + 5.12$$

$$= 10.24$$

هذه الإحصائية لهذا توزيع كاي تربيع بدرجة حرية 1 تحت الفرضية الصفرية. القيمة الناتجة لها $p\text{-value} = 0.00137$ وهي تدل على رفض الفرضية الصفرية ان نسبة عدد الذكور لعدد الإناث هي 1:1 .

أو في رمي زهرة نرد متزنة 60 مرة ونتج التالي:

Draw	1	2	3	4	5	6
Number Observed	10	10	8	11	12	9
Number Expected	10	10	10	10	10	10

تحت الفرضية الصفرية أن الزهرة متوازنة نحسب الإحصائية

$$\chi_0^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

حيث O_i العدد المشاهد من النوع $i = 1, 2, \dots, n$ و E_i العدد المتوقع.

$$\chi_0^2 = \frac{(10-10)^2}{10} + \frac{(10-10)^2}{10} + \dots + \frac{(9-10)^2}{10} = 1$$

ولها درجات حرية $5 = 6 - 1$ ونجد قيمة $p\text{-value} = 0.962566$ أي لانرفض الفرضية الصفرية.

كمثال لحسن مطابقة بيانات على توزيع معين لنفترض أن أحد الباحثين تحصل على البيانات التالية والتي يعتقد انها موزعة توزيع بواسون بمتوسط 4.

2	2	3	6	4	3	5	6	2	5	6
4	4	3	4	1	7	2	4	3	0	2
2	5	3	5	0	5	7	4	3	3	8

2 4 5 4 5 5 4 5 4 2 5
3 11 7 5 4 7

نكون توزيع تكراري للبيانات

value	0	1	2	3	4	5	6	7	8	9	10	11
obs	2	1	8	8	11	11	3	4	1	0	0	1
exp	0.9	3.66	7.3	9.77	9.77	7.8	5.21	2.98	1.49	0.66	0.26	0.1

تحسب القيم المتوقعة تحت الفرضية الصفرية ان البيانات تأتي من توزيع بواسون

بمتوسط 4 كالتالي:

من العلاقة

$$E_i = n \times P(x) = n \frac{\lambda^x}{x!} e^{-\lambda}, x = 0, 1, 2, \dots$$

حيث n عدد المشاهدات و $P(x) = \frac{\lambda^x}{x!} e^{-\lambda}, x = 0, 1, 2, \dots$ دالة الكتلة لتوزيع بواسون

ونحسب القيم المتوقعة:

$$E_0 = 50 \times P(x=0) = 50 \frac{4^0}{0!} e^{-4} = 0.915782$$

$$E_1 = 50 \times P(x=1) = 50 \frac{4^1}{1!} e^{-4} = 3.663129$$

$$E_2 = 50 \times P(x=2) = 50 \frac{4^2}{2!} e^{-4} = 7.326256$$

الخ

نحسب الإحصائية

$$\chi_0^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

$$\chi_0^2 = \frac{(2-0.9)^2}{0.9} + \frac{(1-3.66)^2}{3.66} + \dots + \frac{(1-0.1)^2}{0.1} = 15.919$$

ولها 11 درجات حرية (عدد الخلايا - 1) القيمة الإحصائية هي 0.144159 وهذه القيمة تدعوا لعدم رفض الفرضية الصفرية.
مثال على إختبار حسن التطابق:

SAS

```
DATA GENE;
INPUT @1 COLOR $13. @15 NUMBER 3.;
DATALINES;
YELLOWSMOOTH 152
YELLOWWRINKLE 39
GREENSMOOTH 53
GREENWRINKLE 6
;
* HYPOTHESIZING A 9:3:3:1 RATIO;
PROC FREQ ORDER=DATA; WEIGHT NUMBER;
TITLE3 'GOODNESS OF FIT ANALYSIS';
TABLES COLOR / NOCUM CHISQ TESTP=(0.5625 0.1875 0.1875
0.0625);
RUN;
```

Thursday, March 14, 2013 3

The SAS System

16:06

GOODNESS OF FIT ANALYSIS

The FREQ Procedure

COLOR	Frequency	Percent	Test Percent
YELLOWSMOOTH	1	6.67	56.25
YELLOWWRINKLE	3	20.00	18.75
GREENSMOOTH	5	33.33	18.75
GREENWRINKLE	6	40.00	6.25

Chi-Square Test
for Specified Proportions
Chi-Square 35.6074

DF	3
Pr > ChiSq	<.0001

WARNING: 75% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Sample Size = 15

SPSS

DATA LIST

/ ID 1-14 (A) ORDER 16 NUMBER 18-20.

BEGIN DATA

YELLOWSMOOTH	1	152
YELLOWWRINKLE	2	39
GREENSMOOTH	3	53
GREENWRINKLE	4	6

END DATA.

LIST.

WEIGHT BY ORDER.

NPAR TESTS

/CHISQUARE=number
 /EXPECTED=0.5625 0.1875 0.1875 0.0625
 /MISSING ANALYSIS.

R

```
> observed <- c(152,39,53,6)
> chisq.test(observed, p=c(9,3,3,1)/16) -> results
> results
```

Chi-squared test for given probabilities

data: observed

X-squared = 8.9724, df = 3, p-value = 0.02966

```
> results$expected
[1] 140.625 46.875 46.875 15.625
> results$residuals
[1] 0.9592242 -1.1502174 0.8946135 -2.4349538
>
```

Excel

	A	B	C	D	E	F
1	observed	ratios	expected	chi	chi-test =	=CHITEST(A2:A5,C2:C5)
2	152	9	=(B2/\$B\$7)*\$A\$7	=(A2-C2)*(A2-C2)/C2		
3	39	3	=(B3/\$B\$7)*\$A\$7	=(A3-C3)*(A3-C3)/C3		
4	53	3	=(B4/\$B\$7)*\$A\$7	=(A4-C4)*(A4-C4)/C4		
5	6	1	=(B5/\$B\$7)*\$A\$7	=(A5-C5)*(A5-C5)/C5		
6	sum			chi-sq	p-value	
7	=SUM(A2:A5)	=SUM(B2:B5)	=SUM(C2:C5)	=SUM(D2:D5)	=CHIDIST(D7,3)	
8						

	A	B	C	D	E	F
1	observed	ratios	expected	chi	chi-test =	0.029659504
2	152	9	140.625	0.920111111		
3	39	3	46.875	1.323		
4	53	3	46.875	0.800333333		
5	6	1	15.625	5.929		
6	sum			chi-sq	p-value	
7	250	16	250	8.972444444	0.029659504	
8						

Minitab

```
MTB > name c1 "observed"
MTB > set c1
DATA> 152 39 53 6
DATA> end
MTB > name c2 "ratios"
MTB > set c2
DATA> 9 3 3 1
DATA> end
MTB > TChiSquare;
SUBC>   Observed 'observed';
SUBC>   Counts 'ratios';
SUBC>   GBar;
SUBC>   GChiSQ;
SUBC>   Pareto;
SUBC>   RTable.
```

Chi-Square Goodness-of-Fit Test for Observed Counts in Variable: observed

Category	Observed	Historical Counts	Test Proportion	Expected	Contribution to Chi-Sq
1	152	9	0.5625	140.625	0.92011
2	39	3	0.1875	46.875	1.32300

3	53	3	0.1875	46.875	0.80033
4	6	1	0.0625	15.625	5.92900

N	DF	Chi-Sq	P-Value
250	3	8.97244	0.030

MTB >

جداول الرجحان و إختبار التوافق و الإقتران (Independence and Association)

ونشرحها بمثال: البيانات التالية اخذت من مجموعة من زوار معرض الكتاب لتفضيلهم للكتب الإلكترونية أو الورقية

	Education Level		
Book Type	High School	Bachelors	Master
eBook	23	35	42
Paper Book	45	30	25

الفرضية الصفرية: لا توجد علاقة بين مستوى التعليم ونوع الكتب المفضلة
لحساب القيم المتوقعة نكون الجدول التالي

	Education Level			Row Totals
Book Type	High School	Bachelors	Master	
eBook	23 (34.0) (3.559)	35 (32.5) (0.192)	42 (33.5) (2.157)	100
Paper Book	45 (34.0)	30 (32.5)	25 (33.5)	100

	(3.559)	(0.192)	(2.157)	
Column Totals	68	65	67	200

القيمة المتوقعة للخلية = (مجموع السطر الذي تقع فيه الخلية X مجموع العمود الذي تقع فيه الخلية) \ المجموع الكلي

القيمة المتوقعة لخلية (مستوى التعليم ثانوي و كتاب إلكتروني) تحسب من

$$100 \times 68 / 200 = 34.0$$

القيمة المتوقعة لخلية (مستوى التعليم جامعي و كتاب إلكتروني) تحسب من

$$100 \times 65 / 200 = 32.5$$

وهكذا. نحسب إحصائية الاختبار من

$$\chi^2_0 = \sum_{i=1}^n \sum_{j=1}^m \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

نحسب قيمة مربع كاي للخلية الواحدة من $(O - E)^2 / E$ فمثلا للخلية (مستوى التعليم

ثانوي و كتاب إلكتروني) نجد $O = 23$ و $E = 34.0$ فيكون

كاي تربيع (مستوى التعليم ثانوي و كتاب إلكتروني) =

$$(23 - 34.0)^2 / 34.0 = 3.5588$$

وهكذا الخ. مربع كاي الكلي ينتج عن جمع مربعات كاي الجزئية $chi-sq = 11.816$

تحسب درجات الحرية $\nu = (n - 1) \times (m - 1)$ حيث m عدد الأعمدة و n عدد

السطور ويكون $\nu = (2 - 1) \times (3 - 1) = 2$ القيمة الاحتمالية لقيمة $chi-sq =$

11.816 بدرجات حرية 2 هي $p-value = 0.003$ أي ترفض الفرضية الصفرية.

مثال :

```

DATA DRNKRS;
INPUT CRM $ DRNKR COUNT;
DATALINES;
Arsn 1      50
Arsn 0      43
Rp    1      88
Rp    0      62
Vlnc 1     155
Vlnc 0     110
Stlng 1    379
Stlng 0    300
Cnng  1     18
Cnng  0     14
Frd   1     63
Frd   0    144
;
PROC FREQ DATA=DRNKRS;
WEIGHT COUNT;
TABLES DRNKR*CRM/CHISQ;
TITLE 'Chi Square Analysis of a Contingency Table';
RUN;

```

Drug	Arsn	Rp	Vlnc	Stlng	Cnng	Frd
1	50	88	155	379	18	63
0	43	62	110	300	14	144

SPSS

DATA LIST/

CRM (A5) ORDER DRG COUNT .

BEGIN DATA.

```

Arsn      1      1      50
Arsn      1      0      43
Rp         2      1      88
Rp         2      0      62
Vlnc       3      1     155

```

Vlnc	3	0	110
Stlng	4	1	379
Stlng	4	0	300
Cnng	5	1	18
Cnng	5	0	14
Frd	6	1	63
Frd	6	0	144

END DATA.

EXECUTE.

LIST.

WEIGHT BY ORDER.

CROSSTABS

/TABLES= DRG BY COUNT

/STATISTICS=CHISQ.

Excel

	A	B	C	D
1	observed			
2	crm	Arsn	Rp	Vlnc
3	1	50	88	155
4	0	43	62	110
5	col totals	=SUM(B3:B4)	=SUM(C3:C4)	=SUM(D3:D4)
6	expected			
7	crm	Arsn	Rp	Vlnc
8	1	=\$H3*B\$5/\$H\$5	=\$H3*C\$5/\$H\$5	=\$H3*D\$5/\$H\$5
9	0	=\$H4*B\$5/\$H\$5	=\$H4*C\$5/\$H\$5	=\$H4*D\$5/\$H\$5
10		=SUM(B8:B9)	=SUM(C8:C9)	=SUM(D8:D9)
11	chi	=(B3-B8)*(B3-B8)/B8	=(C3-C8)*(C3-C8)/C8	=(D3-D8)*(D3-D8)/D8
12		=(B4-B9)*(B4-B9)/B9	=(C4-C9)*(C4-C9)/C9	=(D4-D9)*(D4-D9)/D9
13				
14	chi-sq	=SUM(B11:G12)		
15	p-value	=CHIDIST(B14,5)		

E	F	G	H
Stlng	Cnng	Frd	row total
379	18	63	=SUM(B3:G3)
300	14	144	=SUM(B4:G4)
=SUM(E3:E4)	=SUM(F3:F4)	=SUM(G3:G4)	=SUM(H3:H4)
Stlng	Cnng	Frd	
=\$H3*\$E5/\$H\$5	=\$H3*\$F5/\$H\$5	=\$H3*\$G5/\$H\$5	=SUM(B8:G8)
=\$H4*\$E5/\$H\$5	=\$H4*\$F5/\$H\$5	=\$H4*\$G5/\$H\$5	=SUM(B9:G9)
=SUM(E8:E9)	=SUM(F8:F9)	=SUM(G8:G9)	=SUM(H8:H9)
=(E3-E8)*(E3-E8)/E8	=(F3-F8)*(F3-F8)/F8	=(G3-G8)*(G3-G8)/G8	
=(E4-E9)*(E4-E9)/E9	=(F4-F9)*(F4-F9)/F9	=(G4-G9)*(G4-G9)/G9	

	A	B	C	D	E	F	G	H
1	observed							
2	crm	Arsn	Rp	Vlnc	Stlng	Cnng	Frd	row total
3	1	50	88	155	379	18	63	753
4	0	43	62	110	300	14	144	673
5	col totals	93	150	265	679	32	207	1426
6	expected							
7	crm	Arsn	Rp	Vlnc	Stlng	Cnng	Frd	
8	1	49.10869565	79.20757	139.9334	358.5463	16.89762	109.3065	753
9	0	43.89130435	70.79243	125.0666	320.4537	15.10238	97.69355	673
10		93	150	265	679	32	207	1426
11	chi	0.016176839	0.976002	1.622222	1.166808	0.071918	19.61721	
12		0.018099791	1.09202	1.815057	1.305507	0.080468	21.94912	
13								
14	chi-sq	49.73060762						
15	p-value	1.57332E-09						

Minitab

Data Display

Row	crm	order	count
1	Arsn	1	50
2	Arsn	0	43
3	Rp	1	88
4	Rp	0	62
5	Vlnc	1	155
6	Vlnc	0	110
7	Stlng	1	379
8	Stlng	0	300
9	Cnng	1	18
10	Cnng	0	14
11	Frd	1	63

12 Frd 0 144

MTB >

MTB > ChiSquare 'order' 'count'.

Chi-Square Test: order, count

Expected counts are printed below observed counts
Chi-Square contributions are printed below expected counts

	order	count	Total
1	1	50	51
	0.21	50.79	
	2.893	0.012	
2	0	43	43
	0.18	42.82	
	0.180	0.001	
3	1	88	89
	0.37	88.63	
	1.055	0.004	
4	0	62	62
	0.26	61.74	
	0.260	0.001	
5	1	155	156
	0.65	155.35	
	0.184	0.001	
6	0	110	110
	0.46	109.54	
	0.461	0.002	
7	1	379	380
	1.59	378.41	
	0.220	0.001	
8	0	300	300

	1.26	298.74	
	1.257	0.005	
9	1	18	19
	0.08	18.92	
	10.641	0.045	
10	0	14	14
	0.06	13.94	
	0.059	0.000	
11	1	63	64
	0.27	63.73	
	1.997	0.008	
12	0	144	144
	0.60	143.40	
	0.603	0.003	
Total	6	1426	1432

Chi-Sq = 19.893, DF = 11

WARNING: 10 cells with expected counts less than 1.

Chi-Square approximation

probably invalid.

12 cells with expected counts less than 5.

الارتباط والانحدار Correlation and Regression

سوف نتناول دراسة البيانات التي يكون لأفرادها متغيران يتغيران معاً في وقت واحد ، وذلك لمعرفة نوع العلاقة التي تربط بينهما . والأمثلة كثيرة على هذا النوع من البيانات ، مثل دراسة العلاقة بين أوزان وأطوال مجموعة من الطلاب ، أو أعمار ودرجات مجموعة من الطلاب أو أجور و إنتاج مجموعة من العمال ، أو الدخل والإنفاق لمجموعة من الأسر ، أو العلاقة بين صفة الطول للأب والابن ، أو صفة الذكاء للأب والابن وهكذا . ثم إيجاد مقاييس تقيس درجة هذه العلاقة .

كذلك سوف نتناول أيضاً دراسة العلاقة بين المتغيرين (X, Y) ، فإذا كان هناك علاقة بين المتغير X والمتغير Y ، فكيف يمكن التعبير عنها بمعادلة رياضية ومنها يمكن التنبؤ بقيمة أحد المتغيرين إذا علمت قيمة المتغير الآخر .

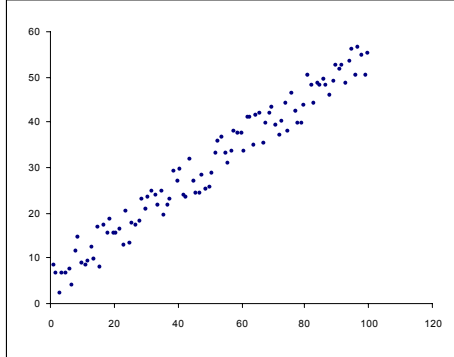
فإذا أردنا دراسة العلاقة بين الطول Y والوزن X لمجموعة عددها n من طلاب جامعة الملك سعود . فإنه يمكن لكل طالب قياس طوله ووزنه ويصبح لدينا أزواج من القراءات التالية :

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$$

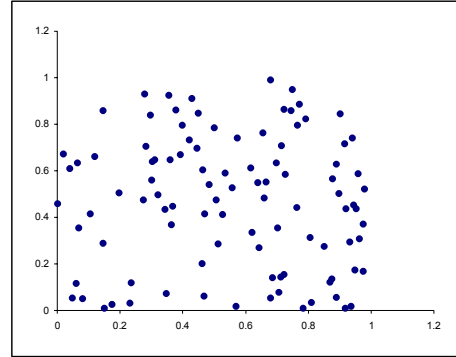
ولتمثيل هذه الأزواج من القراءات بيانياً نرسم محورين متعامدين . المحور الأفقي يمثل الوزن X مثلاً والمحور الرأسي ويمثل الطول Y ونقوم بتمثيل القراءات السابقة بنقاط فنحصل على ما يسمى بشكل الانتشار (Scatter Diagram) .

أشكال الانتشار

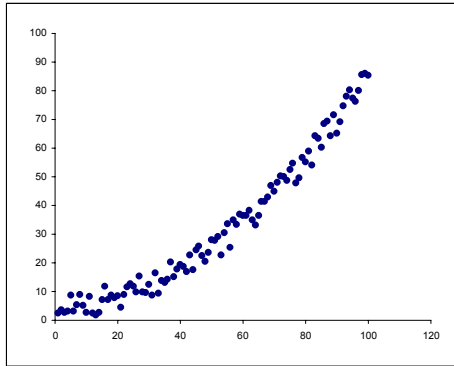
وأشكال الانتشار تأخذ صوراً مختلفة وذلك حسب طبيعة العلاقة بين المتغيرين (X, Y) تحت الدراسة . وفيما يلي نعرض بعض أشكال الانتشار (1) ، (2) ، (3) ، (4) التالية .



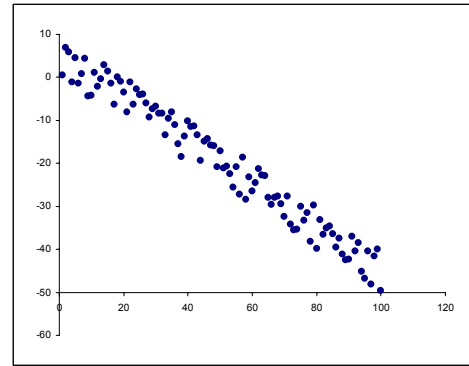
(2) شكل



(1) شكل



(4) شكل



(3) شكل

ونلاحظ من أشكال الانتشار السابقة ما يلي :

شكل الانتشار (1)

تكون فيه النقاط منتشرة بدون ترابط حول اتجاه محدد مما يدل على أنه لا توجد علاقة بين المتغيرين (X, Y) .

شكل الانتشار (2)

تكون فيه النقاط منتشرة حول خط مستقيم تزيد فيه قيم Y مع زيادة قيم X ، ونستنتج منه وجود علاقة خطية طريفة بين المتغيرين (X, Y) .

شكل الانتشار (3)

تكون فيه النقاط منتشرة حول خط مستقيم وفيه تنقص قيم Y مع زيادة قيم X ونستنتج منه وجود علاقة خطية عكسية بين المتغيرين (X, Y) .

شكل الانتشار (4)

تكون فيه النقاط منتشرة حول منحنى نستنتج منه وجود علاقة غير خطية بين المتغيرين (X, Y) .

وسوف نكتفي بإيجاد مقاييس تقيس قوة الارتباط بين المتغيرين (X, Y) في الحالة الخطية فقط وسندرس منها معامل الارتباط الخطي لبيرسون (Person) ، وكذلك دراسة معادلة خط الانحدار للمتغير Y على المتغير X أو العكس للمتغير X على Y .

معامل الارتباط الخطي لبيرسون Coefficient of Linear Correlation

ويستخدم معامل الارتباط الخطي لبيرسون لقياس التغير الذي يطرأ على المتغير Y عندما تتغير قيم X أو العكس . ويستخدم عادة في حالة البيانات الكمية .

معامل ارتباط بيرسون

إذا كان لدينا ازدواج المشاهدات التالية :

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$$

فإن معامل الارتباط r لبيرسون يعطي بمتوسط مجموعة حاصل ضرب القيم المعيارية للمتغيرين X', Y' ويكون كالتالي :

$$r = \frac{1}{n} \sum x'y'$$

حيث :

$$y' = \frac{y - \bar{y}}{\sigma_y}, \quad x' = \frac{x - \bar{x}}{\sigma_x}$$

أي أن العلاقة السابقة يمكن كتابتها كالتالي :

$$r = \frac{1}{n} \frac{\sum (x - \bar{x})(y - \bar{y})}{\sigma_x \sigma_y}$$

ويفضل في حالة البيانات المأخوذة من عينة وضع العلاقة السابقة في الصورة التالية :

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{(n-1)s_x s_y}$$

ويكون لمعامل الارتباط r الخصائص التالية .

خصائص معامل الارتباط الخطي لبيرسون

- (1) قيمته تساوي الصفر عندما تكون الظاهرتان مستقلتين تماماً ولا يوجد ارتباط بينهما ألبتة.
- (2) قيمته مقدار موجب عندما يكون الارتباط بين المتغيرين طردياً . ويكون قوياً عندما يكون المقدار الموجب قريباً من الواحد الصحيح وضعيفاً عندما يكون المقدار الموجب قريباً من الصفر .
- (3) قيمته مقدار سالب عندما يكون الارتباط بين المتغيرين عكسياً ، ويكون قوياً عندما تقترب قيمته من المقدار -1 وضعيفاً عندما تقترب من الصفر .

ملاحظة

يجب ملاحظة عدم ربط الارتباط بين متغيرين بالسببية أي أن التغير في أحد المتغيرين ليس بالضرورة يؤدي إلى التغير في المتغير الثاني (يتسبب فيه) فمثلاً إذا وجدنا ارتباطاً قوياً بين سلسلة من بيانات خاصة بأعداد مرضى السكري في سنوات متتالية وسلسلة بيانات الزيادة في دخل الوظائف العسكرية فهذا لا يدل على أن هناك سبب بين الظاهرتين. أو إذا كان يوجد ارتباط عكسي قوي بين سلسلة مبيعات البطاطين في المملكة العربية السعودية في شهور سنة ما مع درجات الحرارة في إنجلترا في نفس شهور السنة فلا يمكن تفسير ذلك بأن انخفاض درجة الحرارة في إنجلترا يتسبب في زيادة مبيعات البطاطين في السعودية.

خط الانحدار Regression Line

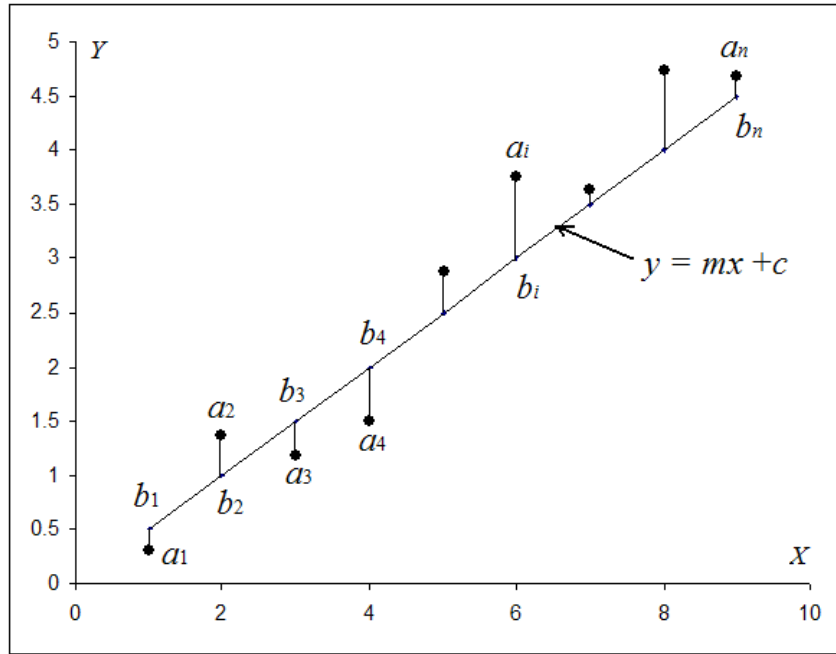
سبق لنا أن بينّا أنه إذا كان لدينا متغيران (X, Y) أحدهما متغير مستقل X مثلاً والآخر متغير تابع وليكن Y مثلاً لظاهرتين محل الدراسة مثل العلاقة بين الطول X والوزن Y . أو العلاقة بين الإنفاق Y والدخل X . فإنه يمكن عمل أشكال الانتشار لهذه المتغيرات . وقد تم دراسة إيجاد معامل الارتباط بعدة طرق لقياس قوة الارتباط بين المتغيرين (X, Y) وذلك في الحالة التي يكون فيها شكل الانتشار في صورة خطية . وفيما يلي نبحث عن إيجاد معادلة رياضية تمثل أحسن توفيق لخط مستقيم يعبر عن البيانات في شكلها الخطي . وهذه العلاقة تسمى بمعادلة خط الانحدار . فإذا كان X متغيراً مستقلاً ، و Y متغيراً تابعاً ، فإن المعادلة التي نحصل عليها تسمى بمعادلة خط انحدار Y على X وهي على الصورة التالية :

$$y = mx + c$$

وإذا كان المتغير Y متغيراً مستقلاً والمتغير X متغيراً تابعاً فإن معادلة الخط المستقيم تسمى بمعادلة خط انحدار X على Y وتعطى بالعلاقة التالية :

$$x = m'y + c'$$

والغرض من إيجاد معادلة خط الانحدار هو التنبؤ بقيمة المتغير التابع لقيمة محددة من قيم المتغير المستقل . ويمكن الحصول على رسم خط الانحدار للبيانات من شكل الانتشار وذلك بالتمهيد باليد ، ولكن هذا التمهيد باليد يختلف من شخص إلى آخر ولذلك دعت الحاجة إلى إيجاد خط الانحدار بطريقة لا تعتمد على الرسم وإنما تعتمد على استخدام البيانات المعطاة للحصول على قيم الثوابت c , m في حالة خط انحدار Y على X وللحصول على c' , m' في حالة خط الانحدار X على Y وفي حالة انحدار Y على X يمكن إيجاد قيمة الثابتين c , m بطريقة المربعات الصغرى كما يلي :



شكل يبين انحرافات النقط التي تمثل المشاهدات عن خط الانحدار

نفرض أن لدينا مجموعة أزواج المشاهدات $(x_1, y_1), \dots, (x_n, y_n)$ ويرسم شكل الانتشار لهذه الأزواج نحصل على النقاط $a_1, \dots, a_n$ نفرض أن خط الانحدار $y = mx + c$ كما هو موضح بالرسم . ويرسم خطوط موازية لمحور Y مارة بالنقاط $a_1, \dots, a_n$ فإنها تقطع خط الانحدار في النقاط $b_1, \dots, b_n$. نرمز للأطوال $a_1b_1, \dots, a_nb_n$ بالقيم $d_1, \dots, d_n$ وهذه الأطوال عبارة عن انحرافات المشاهدات عن خط الانحدار . وقد تكون هذه الانحرافات سالبة أو موجبة حسب وضع النقط $a_1, \dots, a_n$ بالنسبة لخط الانحدار .

لكي نحصل على أجود خط لابد أن يكون مجموع مربعات هذه الانحرافات أقل ما يمكن أي أن :

$$d = \sum_{i=1}^n (mx_i + c - y_i)^2$$

أقل ما يمكن .

بتفاضل d بالنسبة إلى m, c ومساواة الناتج بالصفر نحصل على :

$$\begin{aligned} \sum y &= m \sum x + x + nc \\ \sum xy &= m \sum x^2 + c \sum x \end{aligned}$$

وبحل هاتين المعادلتين نحصل على :

$$m = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2}$$

$$c = \frac{\sum y}{n} - m \frac{\sum x}{n}$$

وبالطريقة نفسها يمكن الحصول على m', c' :

$$m' = \frac{n \sum xy - \sum x \sum y}{n \sum y^2 - (\sum y)^2}$$

$$c' = \frac{\sum x}{n} - m' \frac{\sum y}{n}$$

سوف نطبق نموذج الانحدار

$$Y = \beta X + \varepsilon$$

على البيانات

$$Y = \begin{pmatrix} 136 \\ 144 \\ 145 \\ 169 \\ 176 \\ 195 \\ 211 \\ 224 \\ 231 \\ 256 \\ 281 \\ 312 \\ 347 \\ 377 \\ 423 \\ 477 \\ 553 \\ 613 \end{pmatrix} \quad X = \begin{pmatrix} 1 & 91 & 11 \\ 1 & 105 & 13 \\ 1 & 109 & 17 \\ 1 & 130 & 19 \\ 1 & 146 & 23 \\ 1 & 155 & 29 \\ 1 & 160 & 36 \\ 1 & 180 & 41 \\ 1 & 200 & 59 \\ 1 & 215 & 82 \\ 1 & 240 & 100 \\ 1 & 275 & 110 \\ 1 & 320 & 134 \\ 1 & 360 & 139 \\ 1 & 410 & 138 \\ 1 & 460 & 182 \\ 1 & 510 & 220 \\ 1 & 575 & 271 \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_{18} \end{pmatrix}$$

باستخدام R:

أولاً: بالطريقة المطولة:

تقدر β من العلاقة (أنظر الورقة النظرية)

$$\hat{\beta} = (X'X)^{-1} X'Y$$

أدخل التالي في R

X = matrix(c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,


```
91,105,109,130,146,155,160,180,200,215,240,275,320,360,410,
460,510,575,11,13,17,19,23,29,36,41,59,82,100,110,134,139,
138,182,220,271),byrow=F,18,3)
```

```
Y = matrix(c(136,144,145,169,176,195,211,224,231,256,281,
312,347,377,423,477,553,613),byrow=F,18,1)
```

```
XX = solve(t(X)%*%X)
XY = t(X)%*%Y
beta = XX%*%XY
options(digits=4)
beta
```

```
      [,1]
[1,] 60.4599
[2,]  0.7563
[3,]  0.4135
```

أي

$$\hat{\beta} = \begin{pmatrix} 60.4599 \\ 0.7563 \\ 0.4135 \end{pmatrix}$$

والنموذج المقدر

$$\hat{y}_i = 60.4599 + 0.7563x_{i1} + 0.4135x_{i2}, i = 1, 2, \dots, 18$$

تقدير الخطأ المعياري للبواقي:

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^n \hat{\varepsilon}_i^2}{n-k}}$$

حيث $n = 18$ (حجم العينة) و $k = 3$ (عدد المعالم المقدرة) و

$$\hat{\varepsilon}_i = y_i - \hat{y}_i, i = 1, 2, \dots, 18$$

هي البواقي أو مقدرات للخطأ.

نحسب $\hat{y}_i, i = 1, 2, \dots, 18$ من

```
Ye = beta[1]*X[,1]+beta[2]*X[,2]+beta[3]*X[,3]
erro = Y-Ye
```

مجموع مربعات الأخطاء $\sum_{i=1}^{18} \hat{\varepsilon}_i^2$ يحسب من

```
sum(erro^2)
```

[1] 1084

ويكون الخطأ المعياري للبواقي

```
n = length(Y)
k = 3
sigma2 = sum(erro^2)/(n-k)
sigma = sqrt(sigma2)
sigma
[1] 8.503
```

إذا الخطأ المعياري للبواقي $\hat{\sigma} = 8.503$ بدرجات حرية 15

الخطأ المعياري لمقدرات المعالم:

الاطءاء المعيارية هي عناصر القطر لمصفوفة التباين - التغيرات

$$V(\hat{\beta}) = \hat{\sigma}^2 (\mathbf{X}'\mathbf{X})^{-1}$$

```
var.beta = sigma2*XX
var.beta
      [,1]      [,2]      [,3]
[1,] 67.6512 -0.665858  1.19755
[2,] -0.6659  0.007659 -0.01451
[3,]  1.1976 -0.014508  0.02819
```

قطر المصفوفة

```
diag(var.beta)
[1] 67.651189  0.007659  0.028188
```

والأخطاء المعيارية للمعالم

```
sqrt(diag(var.beta))
[1] 8.22503 0.08752 0.16789
```

ونضع النموذج على الشكل

$$\hat{y}_i = 60.4599 + 0.7563x_{i1} + 0.4135x_{i2}, i = 1, 2, \dots, 18$$

(8.225) (0.088) (0.168)

معامل التحديد:

معامل التحديد R^2 غير المعدل يعطى بالعلاقة

$$R^2 = 1 - \frac{\sum_{i=1}^n \hat{\varepsilon}_i^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

ويحسب

```
R2 <- 1-sum(erro^2)/sum((Y-mean(Y))^2)
R2
[1] 0.9969
```

لحساب معامل التحديد R_a^2 المعدل

$$R_a^2 = 1 - (1 - R^2) \frac{n-1}{n-k}$$

$$R22 <- 1 - (1 - R2) * (18-1) / (18-3)$$

$$R22$$

$$[1] \quad 0.9965$$

إختبار فرضيات:

سوف نختبر هل هناك علاقة بين المتغيرات المستقلة والمتغير التابع كالآتي:

$$H_0 : \beta_1 = 0$$

$$\beta_2 = 0$$

وبوضعها في شكل مصفوفة

$$H_0 : \mathbf{RB} = \mathbf{r}$$

حيث:

$$\mathbf{R} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}, \mathbf{r} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

ونحسب النسبة F

$$F = \frac{(\mathbf{R}\hat{\beta} - \mathbf{r})' [\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}]^{-1} (\mathbf{R}\hat{\beta} - \mathbf{r}) / q}{\frac{\sum_{i=1}^n \hat{\varepsilon}_i^2}{n-k}}$$

حيث q هو عدد القيود على الفرضية الصفرية ويساوي هنا 2

$$\mathbf{R} = \text{matrix}(c(0, 1, 0, 0, 0, 1), 2, 3, \text{byrow}=T)$$

$$\mathbf{r} = \text{matrix}(c(0, 0), 2, \text{byrow}=T)$$

$$q = 2$$

$$(t(\mathbf{R} \% \beta - \mathbf{r}) \% \text{solve}(\mathbf{R} \% \mathbf{X} \mathbf{X} \% t(\mathbf{R})) \% (\mathbf{R} \% \beta - \mathbf{r}) / q) / (\text{sum}(\text{erro}^2) / (n - k))$$

$$[, 1]$$

$$[1,] \quad 2450$$

$$\text{qf}(0.95, 2, 15)$$

$$[1] \quad 3.682$$

وحيث أن $2450 \in [3.682, \infty)$ فإننا نرفض الفرضية الصفرية.

ثانياً بالطريقة المختصرة:

باستخدام lm:

```
summary(lm(Y~X[,2]+X[,3]))
```

Call:

```
lm(formula = Y ~ X[, 2] + X[, 3])
```

Residuals:

Min	1Q	Median	3Q	Max
-13.22	-4.85	-1.60	4.58	15.83

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	60.4599	8.2250	7.35	2.4e-06	***
X[, 2]	0.7563	0.0875	8.64	3.3e-07	***
X[, 3]	0.4135	0.1679	2.46	0.026	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.5 on 15 degrees of freedom

Multiple R-squared: 0.997, Adjusted R-squared: 0.997

F-statistic: 2.45e+03 on 2 and 15 DF, p-value: <2e-16

باستخدام SAS:

```
DATA multreg;  
    INPUT y x1 x2 ;  
    DATALINES;
```

136	91	11
144	105	13
145	109	17
169	130	19
176	146	23
195	155	29
211	160	36
224	180	41
231	200	59
256	215	82
281	240	100
312	275	110
347	320	134
377	360	139

423	410	138
477	460	182
553	510	220
613	575	271

```
;
ODS RTF;
PROC REG DATA = multreg;
MODEL y = x1 x2;
RUN;
ODS RTF CLOSE;
```

المخرجات:

The SAS System 19:22 Thursday, February 25, 2013 1

The REG Procedure
Model: MODEL1
Dependent Variable: y

Number of Observations Read	18
Number of Observations Used	18

Analysis of Variance

Value	Source Pr > F	DF	Sum of Squares	Mean Square	F
2450.17	Model	2	354269	177134	
	<.0001				
	Error	15	1084.41994	72.29466	
	Corrected Total	17	355353		

Root MSE	8.50263	R-Square	0.9969
Dependent Mean	292.77778	Adj R-Sq	0.9965
Coeff Var	2.90412		

Parameter Estimates

> t	Variable	DF	Parameter Estimate	Standard Error	t Value	Pr
	Intercept	1	60.45995	8.22503	7.35	
<.0001	x1	1	0.75635	0.08752	8.64	
<.0001	x2	1	0.41349	0.16789	2.46	
0.0264						

Number of Observations Read	18
Number of Observations Used	18

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	354269	177134	2450.17	<.0001
Error	15	1084.4199 4	72.2946 6		
Corrected Total	17	355353			

Root MSE	8.50263	R-Square	0.9969
Dependent Mean	292.7777 8	Adj R-Sq	0.9965
Coeff Var	2.90412		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	60.45995	8.22503	7.35	<.0001
x1	1	0.75635	0.08752	8.64	<.0001
x2	1	0.41349	0.16789	2.46	0.0264

SPSS

DATA LIST LIST

/ y x1 x2.

BEGIN DATA

```
136  91   11
144 105   13
145 109   17
169 130   19
176 146   23
195 155   29
211 160   36
224 180   41
231 200   59
256 215   82
281 240  100
312 275  110
347 320  134
377 360  139
423 410  138
477 460  182
553 510  220
613 575  271
END DATA.
```

LIST.

REGRESSION

/DEPENDENT y

/METHOD = ENTER x1 x2.

y	x1	x2
---	----	----

136.00	91.00	11.00
144.00	105.00	13.00
145.00	109.00	17.00
169.00	130.00	19.00
176.00	146.00	23.00
195.00	155.00	29.00
211.00	160.00	36.00
224.00	180.00	41.00
231.00	200.00	59.00
256.00	215.00	82.00
281.00	240.00	100.00
312.00	275.00	110.00
347.00	320.00	134.00
377.00	360.00	139.00
423.00	410.00	138.00
477.00	460.00	182.00
553.00	510.00	220.00
613.00	575.00	271.00

Number of cases read: 18 Number of cases listed: 18

REGRESSION

/DEPENDENT y

/METHOD = ENTER x1 x2.

المخرجات:

Variables Entered/Removed^b

Model	Variables Entered	Variables Removed	Method
1	x2, x1 ^a	.	Enter

a. All requested variables entered.

b. Dependent Variable: y

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.998 ^a	.997	.997	8.50263

a. Predictors: (Constant), x2, x1

Variables Entered/Removed^b

Model	Variables Entered	Variables Removed	Method
1	x2, x1 ^a	.	Enter

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	354268.691	2	177134.346	2450.172	.000 ^a
	Residual	1084.420	15	72.295		
	Total	355353.111	17			

a. Predictors: (Constant), x2, x1

b. Dependent Variable: y

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	60.460	8.225		7.351	.000
	x1	.756	.088	.779	8.642	.000
	x2	.413	.168	.222	2.463	.026

a. Dependent Variable: y

Minitab

```
MTB > Regress 'y' 2 'x1' 'x2';  
SUBC> Constant;  
SUBC> VIF;  
SUBC> DW;  
SUBC> Press;  
SUBC> Pure;  
SUBC> XLOF;  
SUBC> Brief 2.
```

المخرجات:

Regression Analysis: y versus x1, x2

The regression equation is
 $y = 60.5 + 0.756 x1 + 0.413 x2$

Predictor	Coef	SE Coef	T	P	VIF
Constant	60.460	8.225	7.35	0.000	
x1	0.75635	0.08752	8.64	0.000	39.910
x2	0.4135	0.1679	2.46	0.026	39.910

S = 8.50263 R-Sq = 99.7% R-Sq(adj) = 99.7%

PRESS = 1578.36 R-Sq(pred) = 99.56%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	354269	177134	2450.17	0.000
Residual Error	15	1084	72		
Total	17	355353			

No replicates.
Cannot do pure error test.

Source	DF	Seq SS
x1	1	353830
x2	1	439

Unusual Observations

Obs	x1	y	Fit	SE Fit	Residual	St Resid
17	510	553.00	537.17	4.03	15.83	2.11R

R denotes an observation with a large standardized residual.

Durbin-Watson statistic = 1.25226

No evidence of lack of fit ($P \geq 0.1$).

MTB >

Excel

المخرجات:

SUMMARY OUTPUT

<i>Regression Statistics</i>	
Multiple R	0.998473
R Square	0.996948331
Adjusted R Square	0.996541442
Standard Error	8.502626793
Observations	18

ANOVA

	<i>df</i>	<i>SS</i>	<i>MS</i>
Regression	2	354268.6912	177134.3456
Residual	15	1084.419936	72.29466238
Total	17	355353.1111	

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>
Intercept	60.45994709	8.225034304	7.350722788
x1	0.75634722	0.087517909	8.642199411
x2	0.413493537	0.167893881	2.462826727

<i>F</i>	<i>Significance F</i>
2450.171835	1.36154E-19

<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>
2.40303E-06	42.92870154	77.99119264	42.92870154	77.99119264
3.28542E-07	0.569807214	0.942887226	0.569807214	0.942887226
0.026368095	0.055636203	0.771350871	0.055636203	0.771350871

أمثلة متنوعة:

```
DATA CHILDREN;
INPUT WEIGHT HEIGHT AGE;
DATALINES;
64 57 8
71 59 10
53 49 6
67 62 11
55 51 8
58 50 8
77 55 10
57 48 9
56 42 10
51 42 6
76 61 12
68 57 9
;
ODS RTF;
proc means;
Title 'Example 1a - PROC MEANS, simplest use';
run;
proc means maxdec=2;
    var WEIGHT HEIGHT;
Title 'Example 1b - PROC MEANS, limit decimals,
specify variables';
run;
proc means maxdec=2 n mean stderr median;
    var WEIGHT HEIGHT;
Title 'Example 1c - PROC MEANS, specify statistics
to report';
run;
ODS RTF CLOSE;
```

```
DATA FERTILIZER;
INPUT FEEDTYPE WEIGHTGAIN;
DATALINES;
1 46.20
1 55.60
1 53.30
```

```

1    44.80
1    55.40
1    56.00
1    48.90
2    51.30
2    52.40
2    54.60
2    52.20
2    64.30
2    55.00
;
ODS    RTF;
PROC SORT DATA=FERTILIZER;
      BY FEEDTYPE;
PROC    MEANS;
      VAR WEIGHTGAIN;
      BY    FEEDTYPE;
TITLE   'Summary    statistics by group';
RUN;
PROC    MEANS;
      VAR WEIGHTGAIN;
      CLASS    FEEDTYPE;
TITLE   'Summary    statistics USING CLASS';
RUN;
ODS RTF    CLOSE;

DATA WEIGHT;
      INPUT TREATMENT LOSS @@;
      DATALINES;
      2 1.0 1 3.0 1 -1.0 1 1.5 1 0.5 1 3.5 1 -99
      2 4.5 3 6.0 2 3.5 2 7.5 2 7.0 2 6.0 2 5.5
      1 1.5 3 -2.5 3 -0.5 3 1.0 3 .5 3 78 1 .6 2 3
2 4 3 9 1 7 2 2
;
ODS RTF;
PROC MEAN;
      VAR LOSS;
TITLE 'Find largest and smallest values';
RUN;
ODS RTF CLOSE;

```

```

DATA WEIGHT;
INPUT WBEFORE WAFTER;
* Calculate WLOSS in the DATA step *;
    WLOSS=WAFTER-WBEFORE;
DATA LINES;
200 190
175 154
188 176
198 193
197 198
310 240
245 204
202 178
;
ODS RTF;
PROC MEANS N MEAN T PRT;
    VAR WLOSS;
TITLE 'Paired t-test example using PROC MEANS';
RUN;
ODS RTF CLOSE;

```

```

DATA WT;
INPUT WEIGHT;
DATA LINES;
64
71
53
67
55
58
77
57
56
51
76
68
;

```



```

PROC MEANS NOPRINT DATA=WT;
    VAR WEIGHT;
    OUTPUT OUT=WTMEANS;
    MEAN=WTMEAN STDDEV=WTSD;
RUN;
DATA WTDIFF;
SET WT;
IF _N_=1 THEN SET WTMEANS;
    DIFF=WEIGHT-WTMEAN;
    Z=DIFF/WTSD; * CREATES STANDARDIZED SCORE (Z-
SCORE);
RUN;
ODS RTF;
PROC PRINT DATA= WTDIFF;
    VAR WEIGHT DIFF Z;
RUN;
ODS RTF CLOSE;

```

```

DATA auto ;
    input MAKE $ PRICE MPG REP78 FOREIGN ;
DATALINES;
AMC      4099 22 3 0
AMC      4749 17 3 0
AMC      3799 22 3 0
Audi     9690 17 5 1
Audi     6295 23 3 1
BMW      9735 25 4 1
Buick    4816 20 3 0
Buick    7827 15 4 0
Buick    5788 18 3 0
Buick    4453 26 3 0
Buick    5189 20 3 0
Buick    10372 16 3 0
Buick    4082 19 3 0
Cad.     11385 14 3 0
Cad.     14500 14 2 0
Cad.     15906 21 3 0
Chev.    3299 29 3 0
Chev.    5705 16 4 0

```

```

Chev.   4504 22 3 0
Chev.   5104 22 2 0
Chev.   3667 24 2 0
Chev.   3955 19 3 0
Datsun  6229 23 4 1
Datsun  4589 35 5 1
Datsun  5079 24 4 1
Datsun  8129 21 4 1
;
RUN;

PROC MEANS ;
RUN;

PROC MEANS DATA=auto;
RUN;

PROC MEANS DATA=auto N MEAN STD ;
RUN;

PROC MEANS DATA=auto;
  VAR price ;
RUN;

PROC MEANS DATA=auto;
  CLASS foreign ;
  VAR price ;
RUN;

PROC REG DATA=auto ;
  MODEL price = mpg ;
RUN;
QUIT;

PROC FREQ DATA=auto;
RUN;

PROC FREQ DATA=auto PAGE;
  TABLES rep78 ;
  TABLES price ;

```

```
RUN;  
  
PROC FREQ DATA=auto PAGE;  
  TABLES rep78 / NOPERCENT ;  
  TABLES price ;  
RUN;
```

البيانات التالية تمثل درجات 50 طالباً في إحدى المواد :

51	95	70	74	73	90	71	74	90	67
91	72	83	89	50	80	72	84	85	69
62	82	87	76	91	76	87	75	78	79
71	96	81	88	64	82	73	57	86	70
80	81	75	85	74	90	83	66	77	91

```
student_scores =
c(67,90,74,71,90,73,74,70,95,51,69,85,84,72,
80,50,89,83,72,91,79,78,75,87,76,91,76,87,82
,62,70,86,57,73,82,64,88,81,96,71,91,77,66,8
3,90,74,85,75,81,80)
student_scores
mean(student_scores)
sd(student_scores)
library(mosaic)
tally(student_scores)
histogram(student_scores)
t.test(student_scores, alternative =
c("two.sided"),mu = 75,conf.level = 0.95)
```

```
> rm(list=ls(all=TRUE))
> student_scores =
c(67,90,74,71,90,73,74,70,95,51,69,85,84,72,80,50,89,83,72,
91,79,78,75,87,76,91,76,87,82,62,70,86,57,73,82,64,88,81,96
,71,91,77,66,83,90,74,85,75,81,80)
> student_scores
[1] 67 90 74 71 90 73 74 70 95 51 69 85 84 72 80 50 89 83
72 91 79 78 75 87 76
[26] 91 76 87 82 62 70 86 57 73 82 64 88 81 96 71 91 77 66
83 90 74 85 75 81 80
>
> mean(student_scores)
[1] 77.86
> sd(student_scores)
[1] 10.51337
> library(mosaic)
```

```
> tally(student_scores)
```

50	51	57	62	64	66	67	69	70	71	72	73	74
1	1	1	1	1	1	1	1	2	2	2	2	3
75	76	77	78	79	80	81	82	83	84	85	86	87
2	2	1	1	1	2	2	2	2	1	2	1	2
88	89	90	91	95	96	Total						
1	1	3	3	1	1	50						

```
> histogram(student_scores)
```

```
> t.test(student_scores, alternative = c("two.sided"), mu = 75, conf.level = 0.95)
```

One Sample t-test

data: student_scores

t = 1.9236, df = 49, p-value = 0.06023

alternative hypothesis: true mean is not equal to 75

95 percent confidence interval:

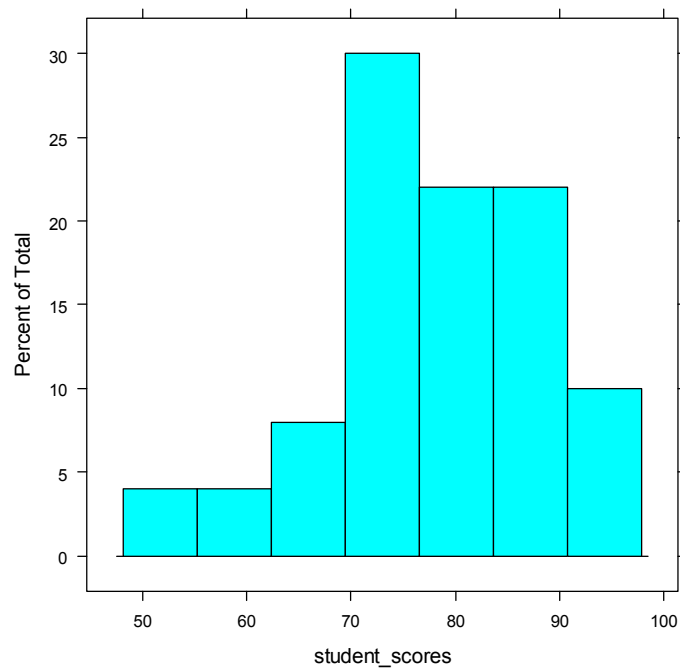
74.87213 80.84787

sample estimates:

mean of x

77.86

```
>
```



Exam Data

Student	Exam 1	Exam 2
1	93	98
2	88	74
3	89	67
4	88	92
5	67	83
6	89	90
7	83	74
8	94	97
9	89	96
10	55	81
11	88	83
12	91	94
13	85	89
14	70	78
15	90	96
16	90	93
17	94	81
18	67	81
19	87	93
20	91	83

```

data STAT328;
input student exam1 exam2 @@;
scorediff = exam2 - exam1 ;
label scorediff='Differences in Exam Scores';
datalines;
1 93 98 2 88 74 3 89 67 4 88 92 5 67 83 6 89 90
7 83 74 8 94 97 9 89 96 10 55 81 11 88 83 12 91 94
13 85 89 14 70 78 15 90 96 16 90 93 17 94 81
18 67 81 19 87 93 20 83 91
;
run;
=====
proc ttest data=chromat;
paired hp*std;
title 'Paired Differences with PROC TTEST';
run;
=====

```

```
PROC UNIVARIATE DATA = AGE;
VAR x;
RUN;
```

```
PROC UNIVARIATE DATA = AGE MU0= 30 ;
VAR x ;
RUN ;
```

```
DATA AGE;
    INPUT x y@@;
    DATALINES;
22 21 40 46 23 22 31 33 24 25
38 42 26 25 51 47 23 21 25 24
31 35 29 27 38 33 25 25 46 37
42 24 32 39 30 27 27 25 24 23
23 21 36 35 36 27 24 24 18 20
55 53 68 52 50 55 62 60 26 27
24 23 19 16 24 21 35 22 25 22
41 40 52 39 27 34 26 27 29 24
26 24 24 22 22 20 24 23 34 39
24 23 22 23 29 28 37 36 26 18
19 19 29 30 36 34 22 20 51 50
42 38 54 44 22 26 24 27 21 20
34 32 35 36 32 32 27 21 23 23
31 36 22 21 51 39 23 22 26 24
45 38 44 44 28 24 31 30 20 22
33 27 33 37 66 53 32 37 25 32
54 47 21 20 20 21 23 21 32 31
20 18 31 23 29 26 41 34 48 43
43 39 21 22 25 20 71 73 54 47
24 23 35 42 54 51 26 33 60 45
;
RUN;
PROC PRINT DATA = AGE;
RUN;
PROC UNIVARIATE DATA = AGE;
VAR x y;
RUN;
PROC UNIVARIATE DATA = AGE MU0= 30 ;
VAR x y;
RUN;
=====
DATA AGE;
    INPUT x y@@;
```

```

        DATALINES;
22 21 40 46 23 22 31 33 24 25
38 42 26 25 51 47 23 21 25 24
31 35 29 27 38 33 25 25 46 37
42 24 32 39 30 27 27 25 24 23
23 21 36 35 36 27 24 24 18 20
55 53 68 52 50 55 62 60 26 27
24 23 19 16 24 21 35 22 25 22
41 40 52 39 27 34 26 27 29 24
26 24 24 22 22 20 24 23 34 39
24 23 22 23 29 28 37 36 26 18
19 19 29 30 36 34 22 20 51 50
42 38 54 44 22 26 24 27 21 20
34 32 35 36 32 32 27 21 23 23
31 36 22 21 51 39 23 22 26 24
45 38 44 44 28 24 31 30 20 22
33 27 33 37 66 53 32 37 25 32
54 47 21 20 20 21 23 21 32 31
20 18 31 23 29 26 41 34 48 43
43 39 21 22 25 20 71 73 54 47
24 23 35 42 54 51 26 33 60 45
;
RUN;
PROC PRINT DATA = AGE;
RUN;
PROC UNIVARIATE DATA = AGE;
VAR x y;
RUN;
PROC UNIVARIATE DATA = AGE MU0= 30 ;
VAR x y;
RUN;
=====
PROC TTEST DATA = AGE;
PAIRED x*y;
TITLE 'Paired Differences with PROC TTEST';
RUN;

```