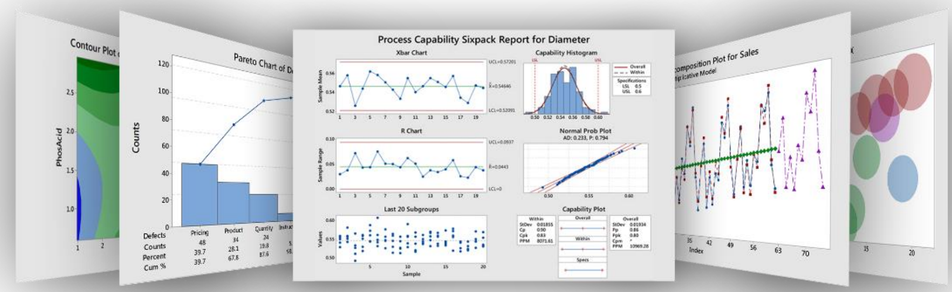
Department of Statistics and Operations Research

College of Science

King Saud University

<u>Exercises</u>

STAT 328 (Statistical Packages)



nashmiah r.alshammari

^_^

Excel

and

Minitab

MATHEMATICAL FUNCTIONS

Write the commands of the following:

		By Excel (using (fx))	By Minitab calc → calculator
Absolute value	$\lvert -4 \rvert$=4	ABS(-4)	
Combinations	$\binom{10}{6}$=10C6=210	COMBIN(10;6)	
The exponential function	$e^{-1.6}$=0.201897	EXP(-1.6)	
Factorial	110! =1.5882E+178	FACT(110)	
Floor function	$[-3.15]$= -4	INT(-3.15)	
Natural logarithm	ln(23)= 3.135494216	LN(23)	
Logarithm with respect to any base	$\log_9(4)$ = 0.630929754	LOG(4;9)	
Logarithm with respect to base 10	log(12) = 1.079181246	LOG10(12)	
Multinomial Coefficient	$\binom{9}{2\quad 2\quad 5}$= 756	MULTINOMIAL(2;2;5)	
Square root	$\sqrt{85}$= 9.219544457	SQRT(85)	
Summation	Summation of: 450,11,20,5 = 486	SUM(450;11;20;5)	
Permutations	10P6=151200	PERMUT(10;6)	
Product	Product of: 450,11,20,5 = 495000	PRODUCT(450;11;20;5)	
Powers	10^{-4}= 0.0001	POWER(10;-4)	

MATRICES

Write the commands of the following:

		By Excel **(**using (fx)**)**	**By Minitab** 1) data → copy → columns in matrix ↘ display data 2) calc → matrices → arithmetic ↘ invers •The name of matrices in <u>columns in matrix</u> keeps their names + × Names of matrix containing.... •The name of new matrices in <u>arithmetic and invers</u> is (M#).
Addition of Matrices	$A=\begin{bmatrix}-5 & 0\\ 4 & 1\end{bmatrix}, B=\begin{bmatrix}6 & -3\\ 2 & 3\end{bmatrix}$ $\Rightarrow A+B=\begin{bmatrix}-5+6 & 0+-3\\ 4+2 & 1+3\end{bmatrix}=\begin{bmatrix}1 & -3\\ 6 & 4\end{bmatrix}$		
Subtract of Matrices	$C=\begin{bmatrix}1 & 2\\ -2 & 0\\ -3 & -1\end{bmatrix}, D=\begin{bmatrix}1 & -1\\ 1 & 3\\ 2 & 3\end{bmatrix}$ $\Rightarrow C-D=\begin{bmatrix}1-1 & 2-(-1)\\ -2-1 & 0-3\\ -3-2 & -1-3\end{bmatrix}=\begin{bmatrix}0 & 3\\ -3 & -3\\ -5 & -4\end{bmatrix}$		
Additive Inverse of Matrix	$A=\begin{bmatrix}1 & 0 & 2\\ 3 & -1 & 5\end{bmatrix}$ $\Rightarrow -A=\begin{bmatrix}-1 & 0 & -2\\ -3 & 1 & -5\end{bmatrix}$		
Scalar Multiplication of Matrices	$D=\begin{bmatrix}-3 & 0\\ 4 & 5\end{bmatrix}$ $\Rightarrow 3D=\begin{bmatrix}-9 & 0\\ 12 & 15\end{bmatrix}$		
Matrix Multiplication	$E=\begin{bmatrix}1 & 4 & 7\\ 2 & 5 & 8\\ 3 & 6 & 9\end{bmatrix}, F=\begin{bmatrix}1 & 4\\ 2 & 5\\ 3 & 6\end{bmatrix}$ $\Rightarrow E\times F=\begin{bmatrix}30 & 66\\ 36 & 81\\ 42 & 96\end{bmatrix}$		
Determinant and Inverse Matrices	$G=\begin{bmatrix}3 & -1\\ -5 & 2\end{bmatrix}$ $\Rightarrow \det(G)=1 \text{ and } G^{-1}=\begin{bmatrix}2 & 1\\ 5 & 3\end{bmatrix}$		

CONDITIONAL FUNCTION (IF) AND COUNT CONDITIONAL FUNCTION

By Excel

(using (fx)**)**

We have grades of 10 students

73	45	32	85	98	78	82	87	60	25	64	72	12	90

1. Print student case being successful (Mark >=60) and being a failure (Mark <60).
2. How many successful students?
3. How many students whose grades are less than or equal to 80?

DESCRIPTIVE STATISTICS

We have students' weights as follows: 44, 40, 42,48,46,44. Find:

	By Excel (using (fx) and (Data Analysis))	**By Minitab** stat → basic statistics → display descriptive statistics + See Appendix -1-
Mean=44	AVERAGE(C2:C7)	
Median=44	MEDIAN(C2:C7)	
Mode=44	MODE.SNGL(C2:C7)	
Sample standard deviation=2.828	STDEV.S(C2:C7)	
Sample variance=8	VAR.S(C2:C7)	
Kurtosis=-0.3	KURT(C2:C7)	
Skewness=4.996E-17	SKEW(C2:C7)	
Minimum=40	MIN(C2:C7)	
Maximum=48	MAX(C2:C7)	
Range=8	MAX(C2:C7)-MIN(C2:C7)	
Count=6	COUNT(C2:C7)	
Coefficient of variation=6.428%	STDEV.S(C2:C7)/AVERAGE(C2:C7)*100	

⋆ Range= Maximum-Minimum

⋆⋆ Coefficient of variation= $\frac{\text{Sample standard deviation}}{\text{Mean}} \times 100\%$

PEARSON CORRELATION COEFFICIENT

We have the table illustrates the age X and blood pressure Y for eight female.

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

Find:

	By Excel **(**using (fx) and (Data Analysis)**)**	**By Minitab** stat → basic statistics → correlation + ✓Display p-value
Correlation=0.791832	CORREL(M3:M10;N3:N10)	

PROBABILITY DISTRIBUTION FUNCTIONS

Discrete Distributions

Notes

If X is discrete random variable, then

1) $P(a < X \le b) = P(X \le b) - P(X \le a)$

and so, if

$P(a \le X < b) = P((a-1) < X \le (b-1)) = P(X \le (b-1)) - P(X \le (a-1))$ or

$P(a \le X \le b) = P((a-1) < X \le b) = P(X \le b) - P(X \le (a-1))$ or

$P(a < X < b) = P(a < X \le (b-1)) = P(X \le (b-1)) - P(X \le a)$.

2) $P(X > a) = 1 - P(X \le a)$,
$P(X \ge a) = 1 - P(X < a) = 1 - P(X \le (a-1))$,
$P(X < a) = P(X \le (a-1))$

1. Binomial Distribution

A biased coin is tossed 6 times. The probability of heads on any toss is 0.3. Let X denote the number of heads that come up. Calculate:

(i) If we call heads a success then this X has a binomial distribution with parameters $n = 6$ and $p = 0.3$.

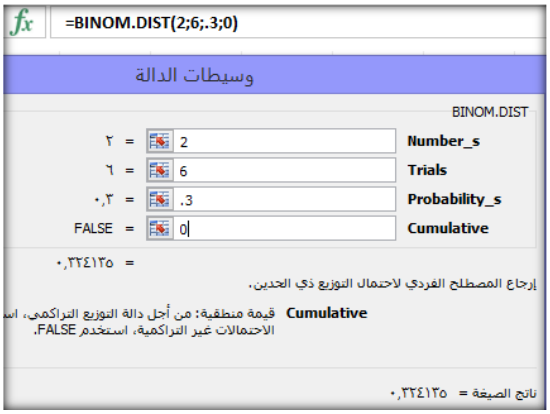
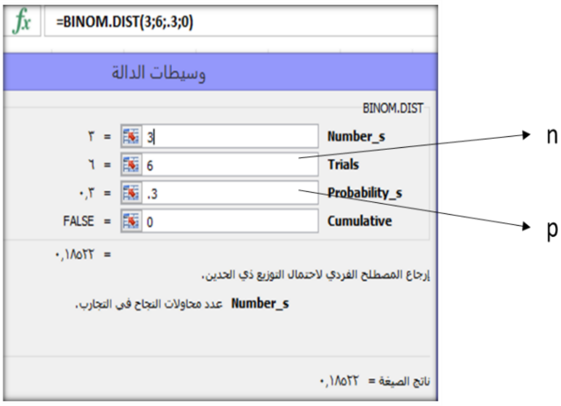
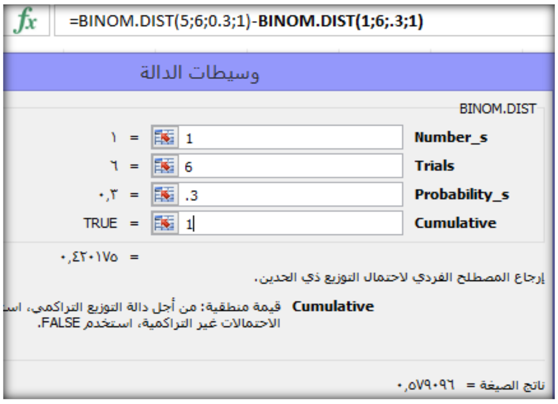
$$P(X = 2) = \binom{6}{2}(0.3)^2(0.7)^4 = 0.324135$$

(ii)

$$P(X = 3) = \binom{6}{3}(0.3)^3(0.7)^3 = 0.18522.$$

(iii) We need $P(1 < X \le 5)$

$$\begin{aligned} & P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5) \\ = \; & 0.324 + 0.185 + 0.059 + 0.01 \\ = \; & 0.578 \end{aligned}$$

<table>
<tr><th></th><th>By Excel
(using (fx))</th><th>By Minitab
calc → probability distribution</th></tr>
<tr><td>i</td><td>
</td><td></td></tr>
<tr><td>ii</td><td>
</td><td></td></tr>
<tr><td>iii</td><td>
</td><td></td></tr>
</table>

2. Poisson Distribution

Births in a hospital occur randomly at an average rate of 1.8 births per hour.

What is the probability of observing 4 births in a given hour at the hospital?

Let X = No. of births in a given hour

(i) Events occur randomly
(ii) Mean rate $\lambda = 1.8$ $\Rightarrow X \sim \text{Po}(1.8)$

We can now use the formula to calculate the probability of observing exactly 4 births in a given hour

$P(X = 4) = e^{-1.8}\frac{1.8^4}{4!} = 0.0723$

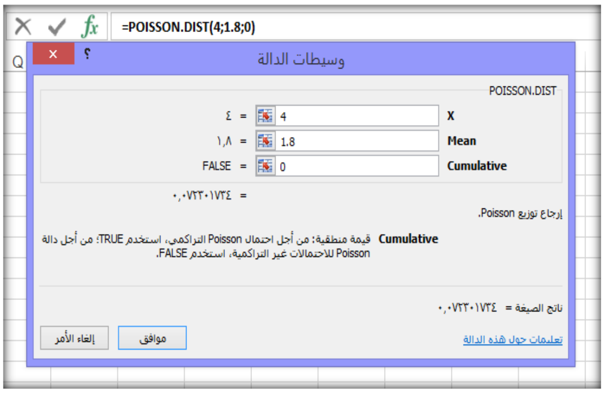
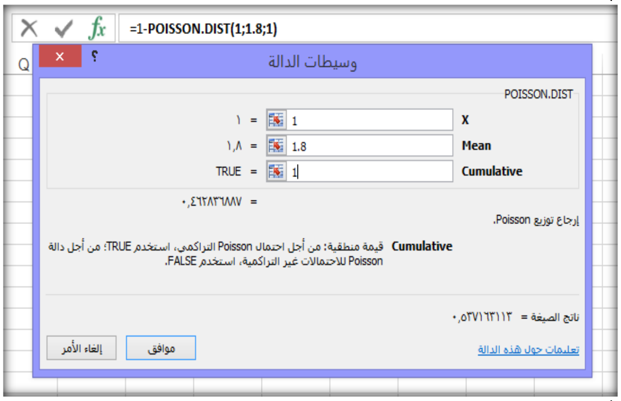
What about the probability of observing more than or equal to 2 births in a given hour at the hospital?

We want $P(X \geq 2) = P(X = 2) + P(X = 3) + \ldots$

i.e. an infinite number of probabilities to calculate

but

$$\begin{aligned} P(X \geq 2) &= P(X = 2) + P(X = 3) + \ldots \\ &= 1 - P(X < 2) \\ &= 1 - (P(X = 0) + P(X = 1)) \\ &= 1 - (e^{-1.8}\frac{1.8^0}{0!} + e^{-1.8}\frac{1.8^1}{1!}) \\ &= 1 - (0.16529 + 0.29753) \\ &= 0.537 \end{aligned}$$

By Excel (using (fx))	**By Minitab** calc → probability distribution
	
	

Continuous Distributions

Notes

If X is continuous symmetric random variable (as Normal distribution and Student's t-distribution), then

1) $P(X \geq x) = 1 - P(X \leq x)$ and $P(X \leq x) = 1 - P(X \geq x)$
2) $P(X \leq x) = 1 - P(X \leq -x)$ and $P(X \geq x) = 1 - P(X \geq -x)$

1. Exponential Distribution

On the average, a certain computer part lasts 10 years. The length of time the computer part lasts is exponentially distributed.

What is the probability that a computer part lasts more than 7 years?

Solution

Let X = the amount of time (in years) a computer part lasts.

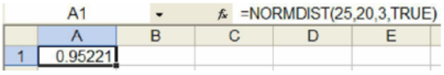
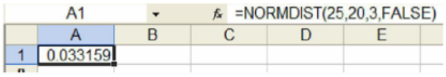
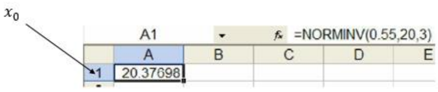
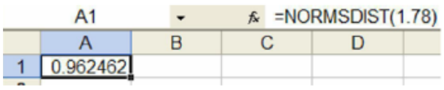
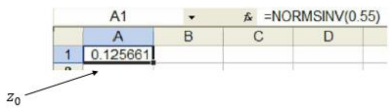
$\mu = 10$ so $m = \frac{1}{\mu} = \frac{1}{10} = 0.1$

$P(X > 7) = 1 - P(X < 7)$.

$P(X > 7) = e^{-0.1 \cdot 7} = 0.4966$. The probability that a computer part lasts more than 7 years is 0.4966.

By Excel (using (fx))	**By Minitab** calc → probability distribution
	

2. Normal Distribution

	By Excel **(using (fx))**	**By Minitab** calc → probability distribution
$P(X \le 25)$ $= P(X < 25)$ at $\mu = 20$ and $\sigma = 3$		
$f_X(25)$ at $\mu = 20$ and $\sigma = 3$		
$P(X \le x_0)$ $= P(X < x_0)$ $= .55$ at $\mu = 20$ and $\sigma = 3$		
$P(Z \le 1.78)$ $= P(Z < 1.78)$ at $\mu = 0$ and $\sigma = 1$		
$P(Z \le z_0)$ $= .55$ at $\mu = 0$ and $\sigma = 1$		

3. Student's t Distribution

Notes in Excel

1) =T.DIST(t, ν, 0) $\leftrightarrow f_{T_\nu}(t)$
2) =T.DIST(t, ν, 1) $\leftrightarrow P(T_\nu \le t)$
3) =T.DIST.RT(t, ν) $\leftrightarrow P(T_\nu \ge t)$
4) =T.DIST.2T(t, ν) $|\leftrightarrow 2P(T_\nu \ge t)$
5) =T.INV(p, ν) $\leftrightarrow P(T_\nu \le t_0) = p$
6) =T.INV.2T(p, ν) $\leftrightarrow 2P(T_\nu \ge t_0) = p$

Find: $(a)\, t_{0.025}$ *when* $\nu = 14$

$(b)\, t_{0.01}$ *when* $\nu = 10$

$(c)\, t_{0.995}$ *when* $\nu = 7$

<table>
<tr><th></th><th>By Excel
(using (fx))</th><th>By Minitab
calc → probability distribution</th></tr>
<tr><td>(a)
$P(T_{14} \le t) = 0.025$</td><td>
</td><td></td></tr>
<tr><td>(b)
$P(T_{10} < t) = 0.01$</td><td>
</td><td></td></tr>
<tr><td>(c)
$P(T_7 < t) = 0.995$</td><td>
</td><td></td></tr>
</table>

Given a random sample of size **24** from a normal distribution, find **k** such that:

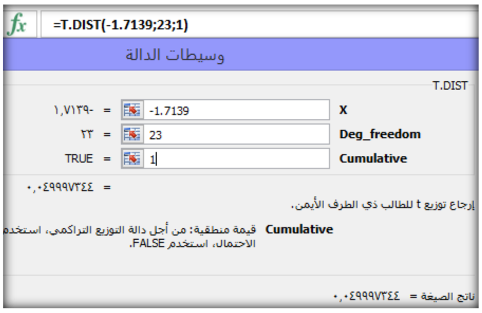
$(a)\ P(-1.7139 < T < k) = 0.90$

$(b)\ P(k < T < 2.807) = 0.95$

$(c)\ P(-k < T < k) = 0.90$

(a)

$P(-1.7139 < T_{23} < k) = 0.9$
$\leftrightarrow P(T_{23} < k) - P(T_{23} < -1.7139) = 0.9$
$\leftrightarrow P(T_{23} < k) = 0.9 + P(T_{23} < -1.7139)$
$\leftrightarrow P(T_{23} < k) = \ 0.949997$

By Excel **(using (fx))**	**By Minitab** calc → probability distribution
  **O** *In excel you might make it in one step too* $P(T_{23} < k) = 0.9 + P(T_{23} < -1.7139)$ *so,* $= T.INV(0.9 + T.DIST(-1.7139, 23, 1), 23) \ = 1.713839369$	

(b)

$$P(k < T_{23} < 2.807) = 0.95$$
$$\leftrightarrow P(T_{23} < 2.807) - P(T_{23} < k) = 0.95$$
$$\leftrightarrow P(T_{23} < k) = (T_{23} < 2.807) - 0.95$$
$$\leftrightarrow P(T_{23} < k) \; = 0.044996$$

<table>
<tr><th>By Excel
(using (fx))</th><th>By Minitab
calc → probability distribution</th></tr>
<tr><td>

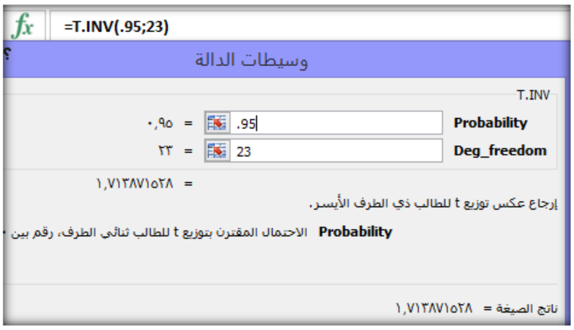
○ *In excel you might make it in one step too*
$P(T_{23} < k) = (T_{23} < 2.807) - 0.95$
$so, \quad = T.INV(-0.95 + T.DIST(2.807,23,1), 23) \; = -1.769952576$

</td><td></td></tr>
</table>

(c)

$$(i)\ P(T_{23} < k) - P(T_{23} < -k) = .9$$
$$\leftrightarrow P(T_{23} < k) - \{1 - P(T_{23} < k)\} = 0.9$$
$$\leftrightarrow 2P(T_{23} < k) - 1 = 0.9$$
$$\leftrightarrow 2P(T_{23} < k) = 1.9$$
$$\leftrightarrow P(T_{23} < k) = 0.95$$
$$so, \quad = T.inv\ (0.95{,}23) = 1.71387$$

$$(ii) P(T_{23} < k) - P(T_{23} < -k) = .9$$
$$\leftrightarrow 1 - P(T_{23} > k) - P(T_{23} < -k) = 0.9$$
$$\leftrightarrow 1 - P(T_{23} > k) - \{1 - P(T_{23} > -k)\} = 0.9$$
$$\leftrightarrow 1 - P(T_{23} > k) - \{1 - [1 - P(T_{23} > k)]\} = 0.9$$
$$\leftrightarrow 1 - P(T_{23} > k) - \{1 - 1 + P(T_{23} > k)\} = 0.9$$
$$\leftrightarrow 1 - P(T_{23} > k) - P(T_{23} > k) = 0.9$$
$$\leftrightarrow 1 - 2P(T_{23} > k) = 0.9$$
$$\leftrightarrow 2P(T23 > k) = 0.1$$
$$so, \quad = T.inv.2t(0.1{,}23) = 1.71387$$

By Excel **(**using (fx)**)**	**By Minitab** calc → probability distribution
 	

4. Chi-Square Distribution

Notes in Excel

1) $= \textbf{CHISQ.DIST}(x, v, 0) \leftrightarrow f_{\chi_v}(x)$
2) $= \textbf{CHISQ.DIST}(x, v, 1) \leftrightarrow P(\chi_v \le x)$
3) $= \textbf{CHISQ.DIST.RT}(x, v, 1) \leftrightarrow P(\chi_v \ge x)$
4) $= \textbf{CHISQ.INV}(p, v) \leftrightarrow P(\chi_v \le x_0) = p$
5) $= \textbf{CHISQ.INV.RT}(p, v) \leftrightarrow P(\chi_v \ge x_0) = p$

By using chi- square distribution ,Find:

$$\chi^2_{0.995} \; when \;\; v = 19$$

<table>
<tr><th></th><th>By Excel
(using (fx))</th><th>By Minitab
calc → probability distribution</th></tr>
<tr><td>$P(\chi_{19} < x)$
=0.995</td><td>
</td><td></td></tr>
</table>

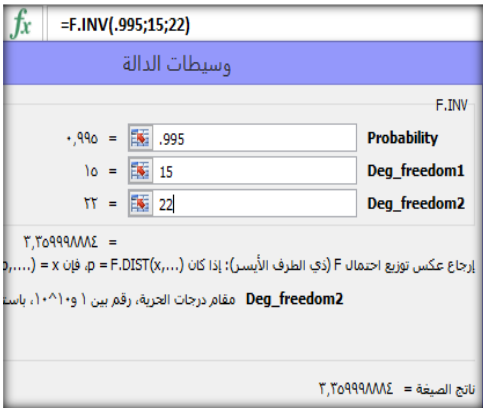
5. F Distribution

Notes in Excel

1) $= \mathbf{F.DIST}(f, v_1, v_2, 0) \quad \leftrightarrow \quad f_{F_{v_1,v_2}}(f)$
2) $= \mathbf{F.DIST}(f, v_1, v_2, 1) \quad \leftrightarrow P(F_{v_1,v_2} \le f)$
3) $= \mathbf{F.DIST.RT}(f, v_1, v_2, 1) \quad \leftrightarrow P(F_{v_1,v_2} \ge f)$
4) $= \mathbf{F.INV}(p, v_1, v_2) \quad \leftrightarrow P(F_{v_1,v_2} \le f_0) = p$
5) $= \mathbf{F.INV.RT}(p, v_1, v_2) \quad \leftrightarrow P(F_{v_1,v_2} \ge f_0) = p$

From the tables of F- distribution ,Find:

$F_{0.995,15,22}$

	By Excel (using (fx))	**By Minitab** calc → probability distribution
$P(F_{15,22} < f) = 0.995$		

HYPOTHESIS TESTING STATISTICS AND CONFIDENCES INTERVAL

	By Excel (using (Data Analysis))	**By Minitab**
Z test one sample for mean with known variance	×	✓ stat → basic statistics
T test one sample for mean with unknown variance	×	✓ stat → basic statistics
T test two samples for means assuming equal variance and unequal variance	✓	✓ stat → basic statistics
T test parried two samples for means	✓	✓ stat → basic statistics
One-way ANOVA (Single Factor ANOVA)	✓	✓ stat → ANOVA → one way → response data are in one column for all factor levels. → response data are in a separate column for each factor level.
Linear regression	✓	✓ Stat → regression → regression → fit regression model
Independent test	×	✓ Stat → tables → chi-square test for association → summarized data in a two-way table; → cross tabulation and chi-square → raw data (categorical variables) + See Appendix -2-

Notes

p-value

(1) $H_1: \theta \neq \theta_0 \rightarrow p-value_{\text{two tail}} = 2P(distribution\ of test\ statistical\ > |test\ statistical|)$

(2) $H_1: \theta > \theta_0 \rightarrow p-value_{\text{one tail }(>)} = P(distribution\ of test\ statistical\ > test\ statistical)$

(3) $H_1: \theta < \theta_0 \rightarrow p-value_{\text{one tail }(<)} = P(distribution\ of test\ statistical\ < test\ statistical)$

In the programs (Excel and Spss for symmetric distribution), how to find p-value for the one tail from p-value for two tail?

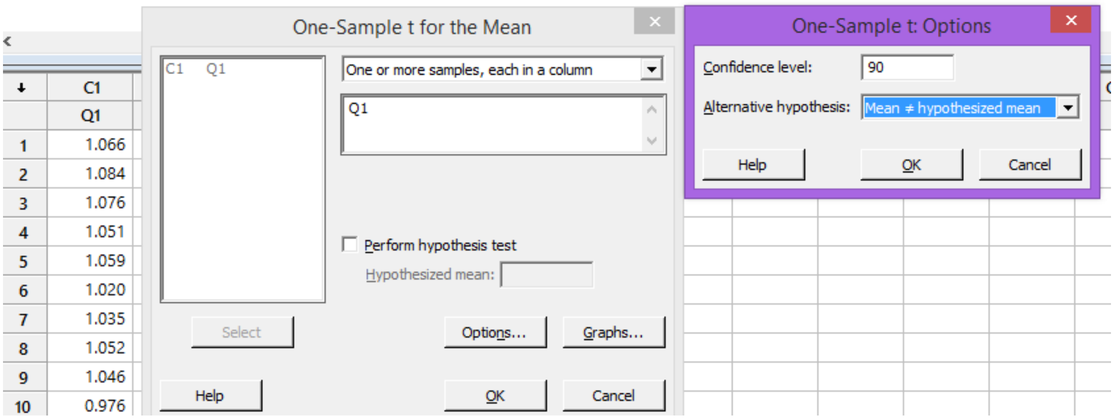
$p-value\ \ one\ tail = \dfrac{p-value_{\text{two tail}}}{2}$	$test\ statistical > 0$	Then we have $p-value_{\text{one tail }(>)}$ and $p-value_{\text{one tail }(<)}$=1-$p-value_{\text{one tail }(>)}$
	$test\ statistical < 0$	Then we have $p-value_{\text{one tail }(<)}$ and $p-value_{\text{one tail }(>)}$=1-$p-value_{\text{one tail }(<)}$

328 stat

1)

For a sample of 10 fruits from thirteen-year-old acidless orange trees, the fruit shape (determined as adiameter divided by height) wae measured [Shaheen and Hamouda (1984b)]:
1.066 1.084 1.076 1.051 1.059 1.020 1.035 1.052 1.046 0.976
Assuming that fruit shapes are approximately normally distributed, find and interpret a 90% confidence interval for the average fruit shape.

(T test one sample for mean with unknown variance By Minitab)



One-Sample T: Q1

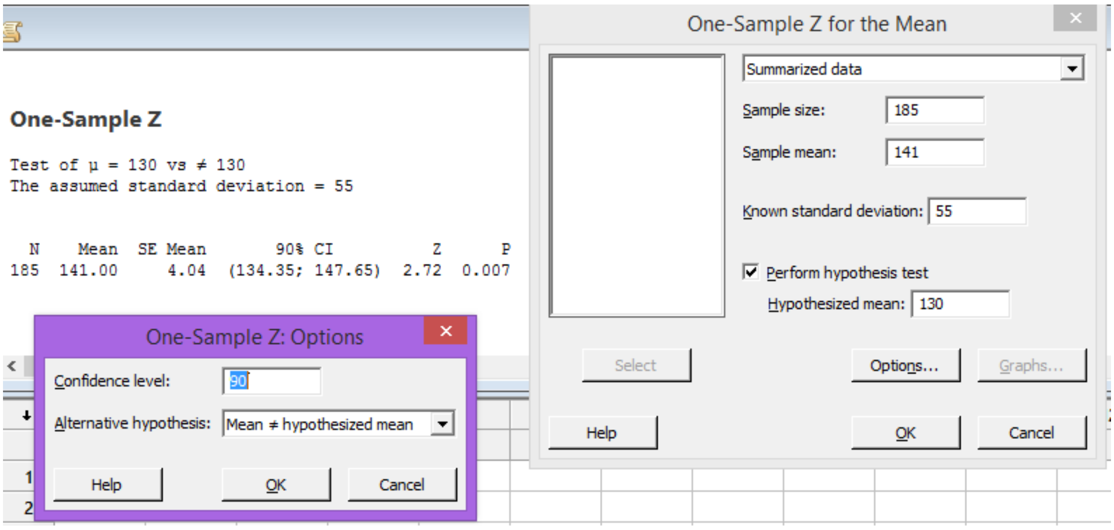
```
Variable   N     Mean    StDev  SE Mean        90% CI
Q1        10  1.04650  0.03103  0.00981  (1.02851; 1.06449)
```

2)

[In a sample of 185 people in the Western Region who had a particular bacterial infection, the mean egg count (per gram of stool) was 141 [Ghandour et. al. (1991)]. Assume that egg counts of such people are normally distributed with a variance of 3025.]
Find and interpret a 90% confidence interval for the average egg count.

** In a sample of 185 people in the Western Region who had a particular bacterial infection, the mean egg count (per gram of stool) was 141. Assume that egg counts of such people are normally distributed with a variance of 3025. Can we conclude that the true mean egg count is different from 130. . Use α=0.10.

(Z test one sample for mean with known variance By Minitab)



328 stat

3)

The phosphorus content was measured for independent samples of skim and whole

Whole: 94.95 95.15 94.85 94.55 94.55 93.40 95.05 94.35 94.70 94.90

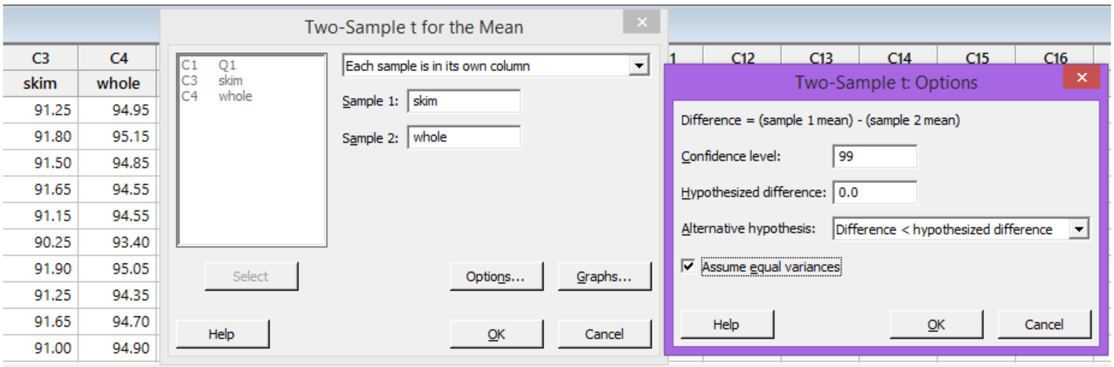
Skim: 91.25 91.80 91.50 91.65 91.15 90.25 91.90 91.25 91.65 91.00

Assuming normal populations with equal variances

a) Test whether the average phosphorus content of skim milk is less than the average phosphorus content of whole milk. Use α=0.01

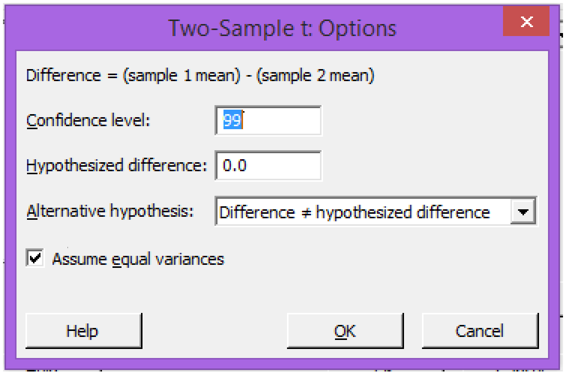
b) Find and interpret a 99% confidence interval for the difference in average phosphorus contents of whole and skim milk

(T test two samples for means assuming equal variance By Minitab)



Two-Sample T-Test and CI: skim; whole

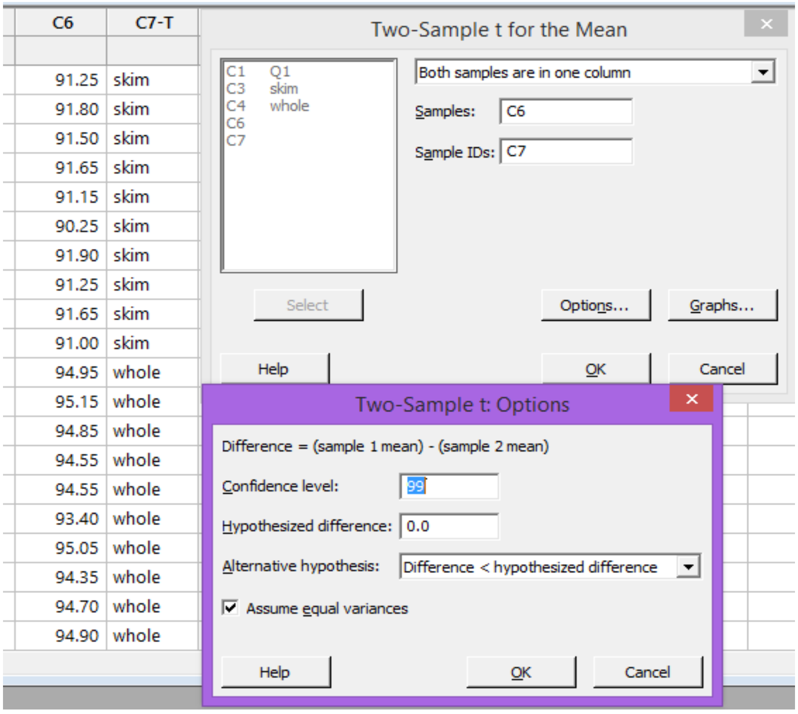
```
Two-sample T for skim vs whole

        N    Mean  StDev  SE Mean
skim   10  91.340  0.483     0.15
whole  10  94.645  0.503     0.16


Difference = μ (skim) - μ (whole)
Estimate for difference:  -3.305
99% upper bound for difference:  -2.742
T-Test of difference = 0 (vs <): T-Value = -14.99  P-Value = 0.000  DF = 18
Both use Pooled StDev = 0.4931
```



Two-Sample T-Test and CI: skim; whole

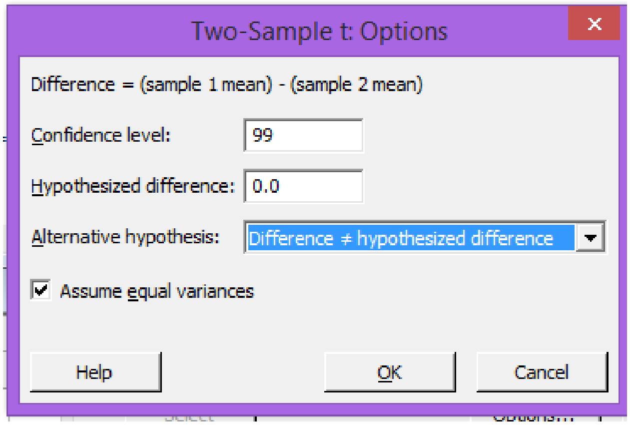
```
Two-sample T for skim vs whole

        N    Mean  StDev  SE Mean
skim   10  91.340  0.483     0.15
whole  10  94.645  0.503     0.16


Difference = μ (skim) - μ (whole)
Estimate for difference:  -3.305
99% CI for difference:  (-3.940; -2.670)
T-Test of difference = 0 (vs ≠): T-Value = -14.99  P-Value = 0.000  DF = 18

Both use Pooled StDev = 0.4931
```

Or



Two-Sample T-Test and CI: C6; C7

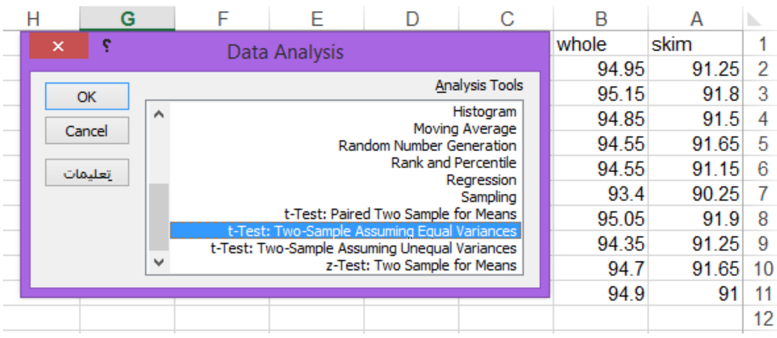
```
Two-sample T for C6

C7      N    Mean  StDev  SE Mean
skim   10  91.340  0.483     0.15
whole  10  94.645  0.503     0.16


Difference = μ (skim) - μ (whole)
Estimate for difference:  -3.305
99% upper bound for difference:  -2.742
T-Test of difference = 0 (vs <): T-Value = -14.99  P-Value = 0.000  DF = 18
Both use Pooled StDev = 0.4931
```



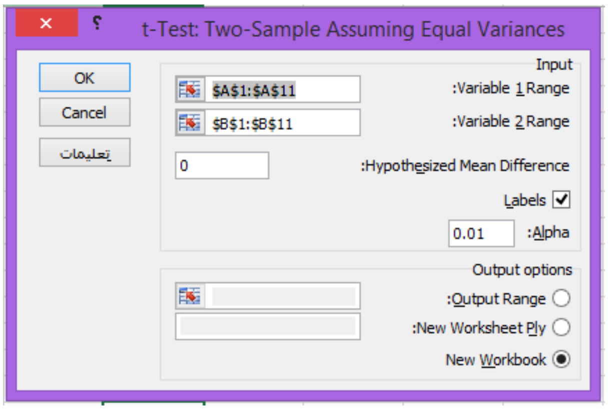
Two-Sample T-Test and CI: C6; C7

```
Two-sample T for C6

C7      N    Mean  StDev  SE Mean
skim   10  91.340  0.483     0.15
whole  10  94.645  0.503     0.16


Difference = μ (skim) - μ (whole)
Estimate for difference:  -3.305
99% CI for difference:  (-3.940; -2.670)
T-Test of difference = 0 (vs ≠): T-Value = -14.99  P-Value = 0.000  DF = 18
Both use Pooled StDev = 0.4931
```

(T test two samples for means assuming equal variance By Excel)



328 stat



t-Test: Two-Sample Assuming Equal Variances

	skim	*whole*
Mean	91.34	94.645
Variance	0.233222	0.253028
Observations	10	10
Pooled Variance	0.243125	
Hypothesized Mean Difference	0	
df	18	
t Stat	-14.9879	
P(T<=t) one-tail	6.53E-12	
t Critical one-tail	2.55238	
P(T<=t) two-tail	1.31E-11	
t Critical two-tail	2.87844	

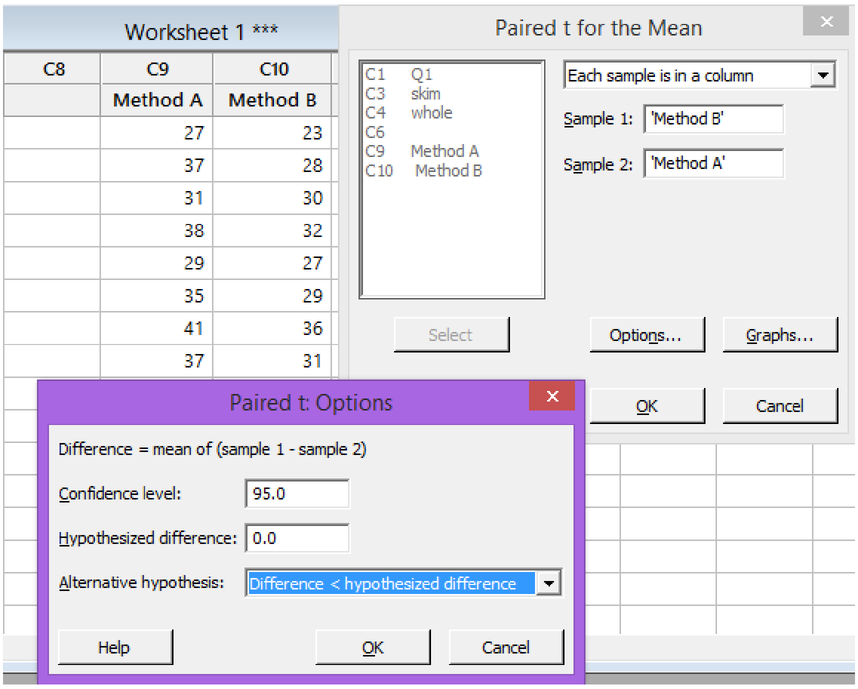
4)

In an experiment comparing 2 feeding methods for caves, eight pairs of twins were used – one twin receiving Method A and other twin receiving Method B. At the end of a given time, the calves were slaughtered and cooked, and the meat was rated for its taste (with a higher number indicating a better taste):

Twin pair	Method A	Method B
1	27	23
2	37	28
3	31	30
4	38	32
5	29	27
6	35	29
7	41	36
8	37	31

Assuming approximate normality, test if the average taste score for calves fed by Method B is less than the average taste foe calves fed by Method A. Use **α**=0.05.

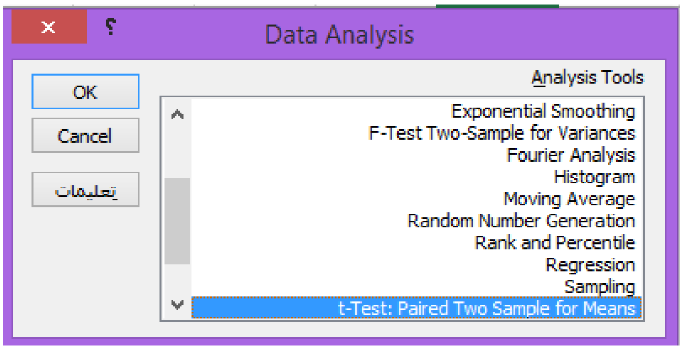
(T test parried two samples for means By Minitab)



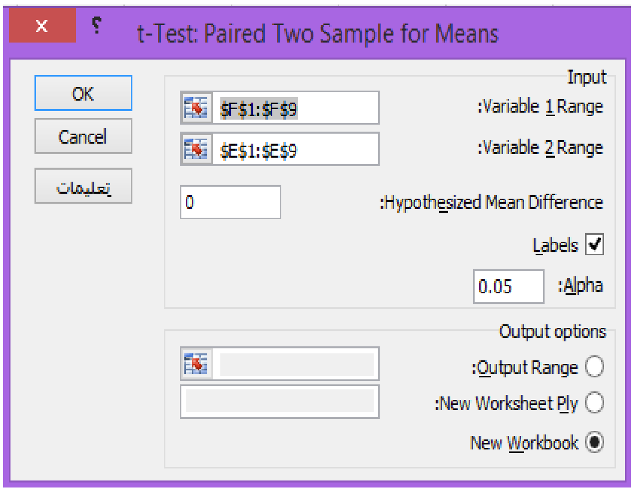
Paired T-Test and CI: Method B; Method A

```
Paired T for Method B - Method A

             N    Mean   StDev  SE Mean
Method B     8   29.50    3.82     1.35
Method A     8   34.38    4.87     1.72
Difference   8  -4.875   2.532    0.895


95% upper bound for mean difference: -3.179
T-Test of mean difference = 0 (vs < 0): T-Value = -5.45  P-Value = 0.000
```

(T test parried two samples for means By Excel)



Method B	Method A
23	27
28	37
30	31
32	38
27	29
29	35
36	41
31	37



t-Test: Paired Two Sample for Means

	Method B	*Method A*
Mean	29.5	34.375
Variance	14.57143	23.69643
Observations	8	8
Pearson Correlation	0.857204	
Hypothesized Mean Difference	0	
df	7	
t Stat	-5.44586	
P(T<=t) one-tail	0.00048	
t Critical one-tail	1.894579	
P(T<=t) two-tail	0.00096	
t Critical two-tail	2.364624	

5)

Two independent samples of dates were taken-one from dates in the Khalal stage and one from dates at the Tamr stage. The calcium (in mg/100g) was measured [Sawaya (1986)]:

Khalal:30,57,29,23,55,50,49,74,101,97,79,158,112,107,93,63,70,90,98,48,75,64,71,72,146,37, 82,19,115,36,34,27,38,42,18,21,75,37,80,72,73,198,107,107,35,56,25,35,26,40,75,109,27,101

Tamr:14,25,21,18,28,14,19,20,44,18,24,47,19,52,31,38,41,39,35,16,47,26,26,30,81,18,42,9,49,2 3,27,14,15,17,10,16,18,14,13,32,42,55,42,27,30,17,24,14,20,17,48,20, 76

Assuming normal populations with unequal variances (α=0.05)

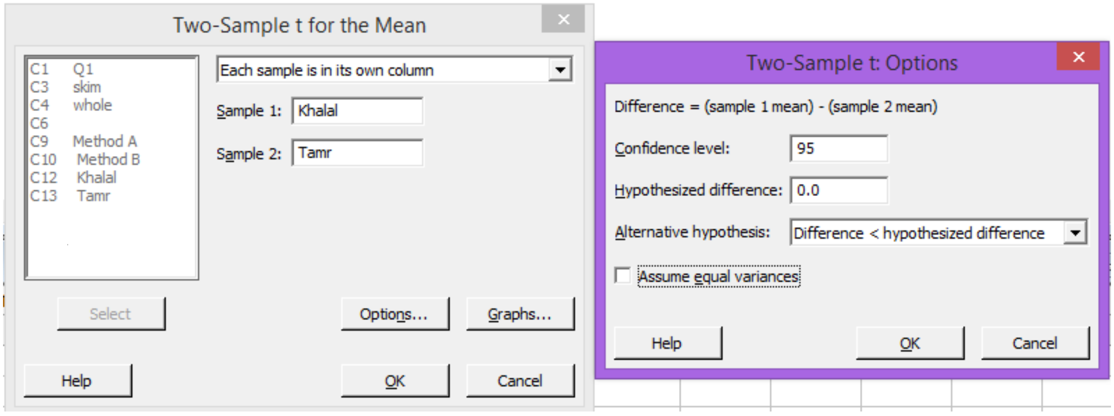
a) Test whether the average calcium of dates at the khalal stage is more than this average for Tamar stage dates

b) Find the confidence interval for the difference in the average calcium of dates at the two stage

328 stat

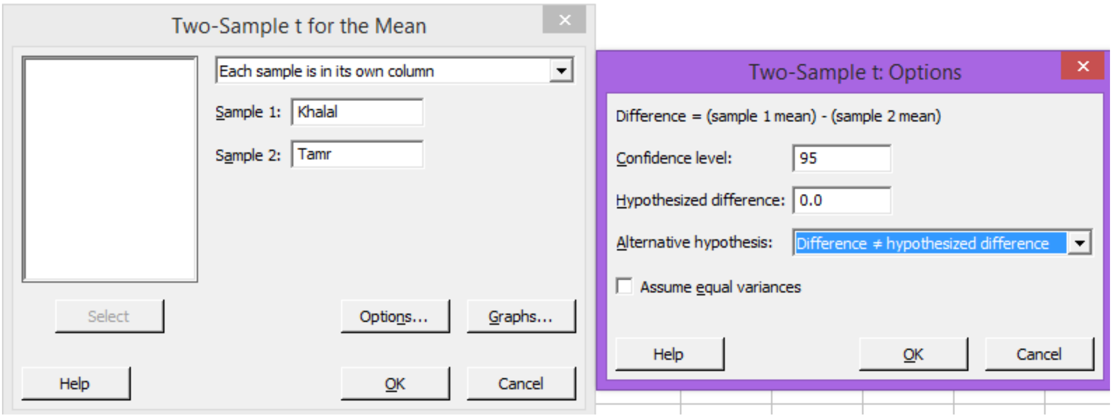
(T test two samples for means assuming unequal variance By Minitab)

	Khalal	Tamr
1	30	14
2	57	25
3	29	21
4	23	18
5	55	28
6	50	14
7	49	19
8	74	20
9	101	44
10	97	18
11	79	24
12	158	47
13	112	19
14	107	52
15	93	31
16	63	38
17	70	41
18	90	39
19	98	35
20	48	16
21	75	47
22	64	26
23	71	26
24	72	30
25	146	81
26	37	18
27	82	42
28	19	9
29	115	49
30	36	23
31	34	27
32	27	14
33	38	15
34	42	17
35	18	10
36	21	16
37	75	18
38	37	14
39	80	13
40	72	32
41	73	42
42	198	55
43	107	42
44	107	27
45	35	30
46	56	17
47	25	24
48	35	14
49	26	20
50	40	17
51	75	48
52	109	20
53	27	76
54	101	



Two-Sample T-Test and CI: Khalal; Tamr

```
Two-sample T for Khalal vs Tamr

         N  Mean  StDev  SE Mean
Khalal  54  67.7   38.0      5.2
Tamr    53  28.7   15.7      2.2


Difference = μ (Khalal) - μ (Tamr)
Estimate for difference:  39.02
95% upper bound for difference:  48.36
T-Test of difference = 0 (vs <): T-Value = 6.97  P-Value = 1.000  DF = 70
```

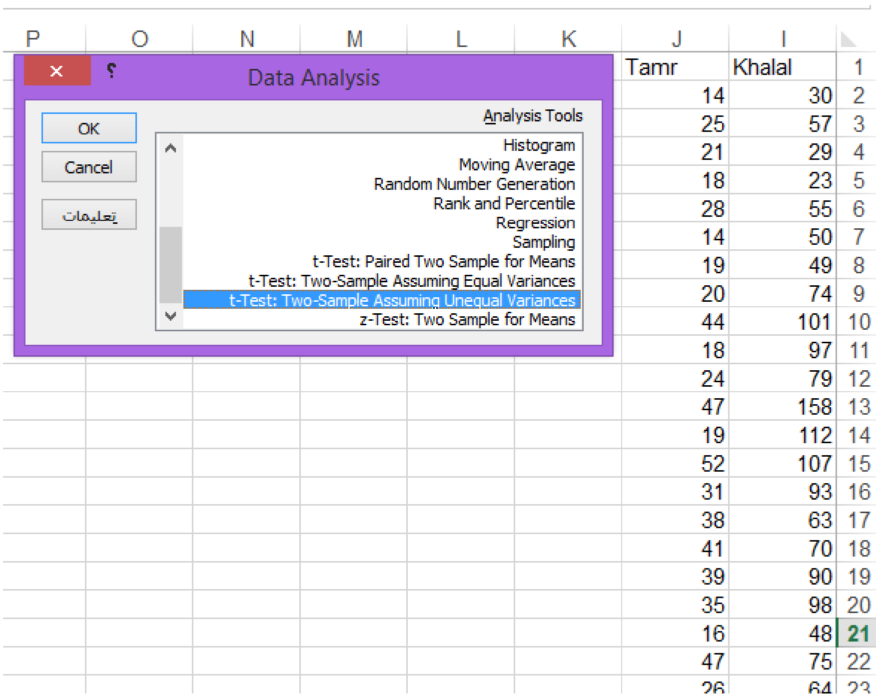


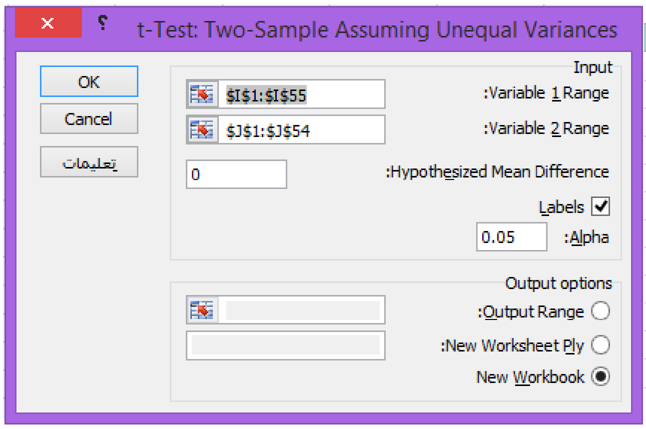
Two-Sample T-Test and CI: Khalal; Tamr

```
Two-sample T for Khalal vs Tamr

         N  Mean  StDev  SE Mean
Khalal  54  67.7   38.0      5.2
Tamr    53  28.7   15.7      2.2


Difference = μ (Khalal) - μ (Tamr)
Estimate for difference:  39.02
95% CI for difference:  (27.85; 50.20)
T-Test of difference = 0 (vs ≠): T-Value = 6.97  P-Value = 0.000  DF = 70
```

328 stat

(T test two samples for means assuming unequal variance By Excel)





t-Test: Two-Sample Assuming Unequal Variances

	Khalal	*Tamr*
Mean	67.74074074	28.71698
Variance	1443.139064	247.2837
Observations	54	53
Hypothesized Mean Difference	0	
df	71	
t Stat	6.965139095	
P(T<=t) one-tail	6.80179E-10	
t Critical one-tail	1.666599658	
P(T<=t) two-tail	1.36036E-09	
t Critical two-tail	1.993943368	

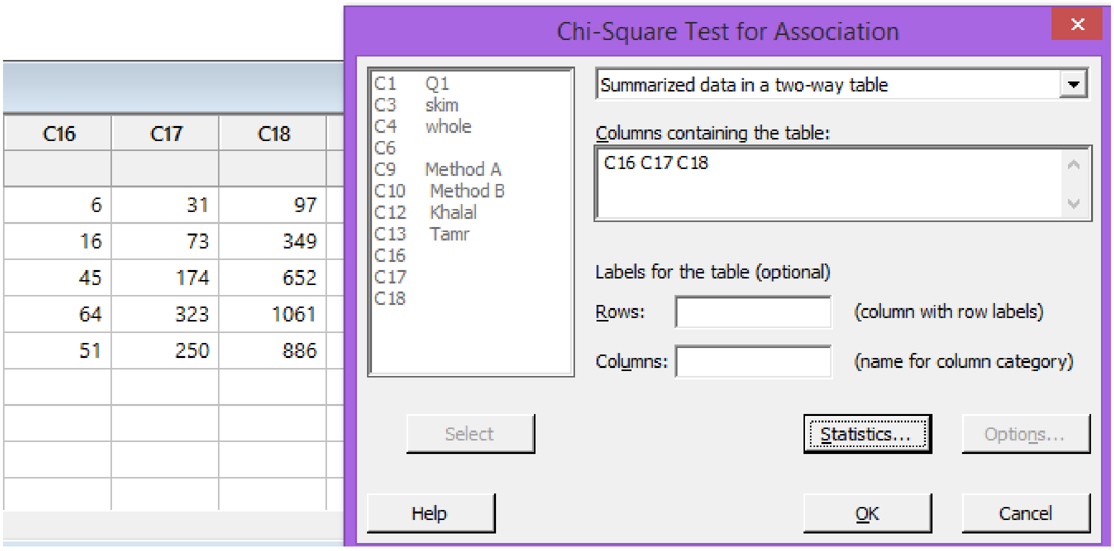
6)

Formation of vitamin D depends on exposure to ultraviolet radiation in sunlight. A sample of Saudis was classified by the type of residence and the level of vitamin D [Sedrani et al. (1992)]:

	Vitamin D Level			
Residence type	Insufficient < 5 ng/ml	Low 5-10 ng/ml	Sufficient > 10 ng/ml	Total
Tent	6	31	97	134
Mud house	16	73	349	438
Flat	45	174	652	871
Villa	64	323	1061	1448
Brick house	51	250	886	1187
Total	182	851	3045	4078

Test whether the Vitamin D level of Saudis is related to the type of residence. Use a level of significance of 0.05.

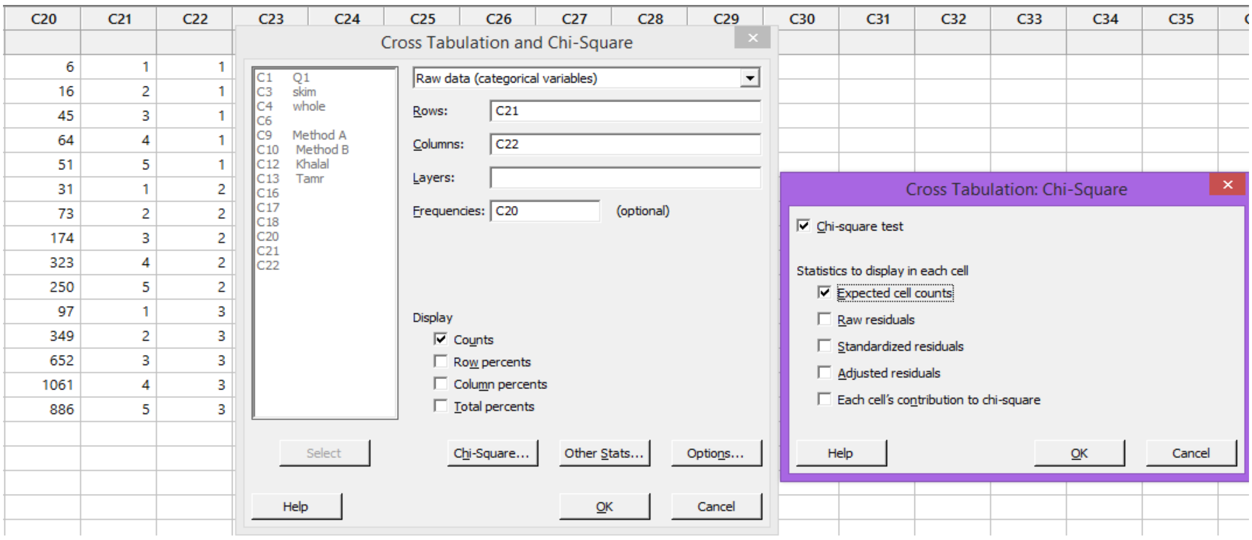
(Independent test By Minitab)



Chi-Square Test for Association: Worksheet rows; Worksheet columns

```
Rows: Worksheet rows   Columns: Worksheet columns

        C16    C17     C18   All

1         6     31      97   134
        6.0   28.0   100.1

2        16     73     349   438
       19.5   91.4   327.1

3        45    174     652   871
       38.9  181.8   650.4

4        64    323    1061  1448
       64.6  302.2  1081.2

5        51    250     886  1187
       53.0  247.7   886.3

All     182    851    3045  4078

Cell Contents:      Count
                    Expected count


Pearson Chi-Square = 9.461; DF = 8; P-Value = 0.305
Likelihood Ratio Chi-Square = 9.668; DF = 8; P-Value = 0.289
```

Or



328 stat

Tabulated Statistics: C21; C22

Using frequencies in C20

Rows: C21   Columns: C22

	1	2	3	All
1	6	31	97	134
	6.0	28.0	100.1	
2	16	73	349	438
	19.5	91.4	327.1	
3	45	174	652	871
	38.9	181.8	650.4	
4	64	323	1061	1448
	64.6	302.2	1081.2	
5	51	250	886	1187
	53.0	247.7	886.3	
All	182	851	3045	4078

Cell Contents: Count
                    Expected count

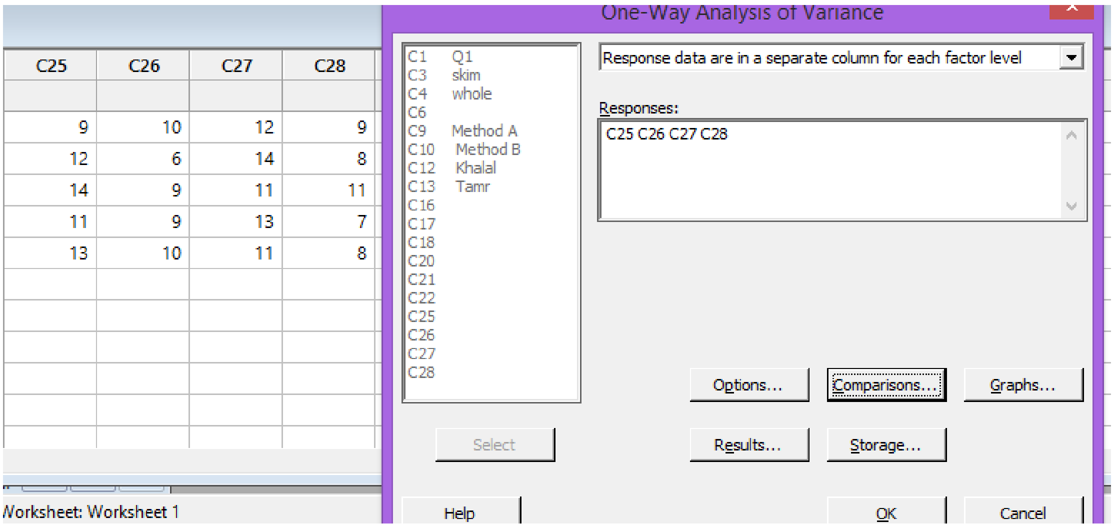
Pearson Chi-Square = 9.461; DF = 8; P-Value = 0.305
Likelihood Ratio Chi-Square = 9.668; DF = 8; P-Value = 0.289

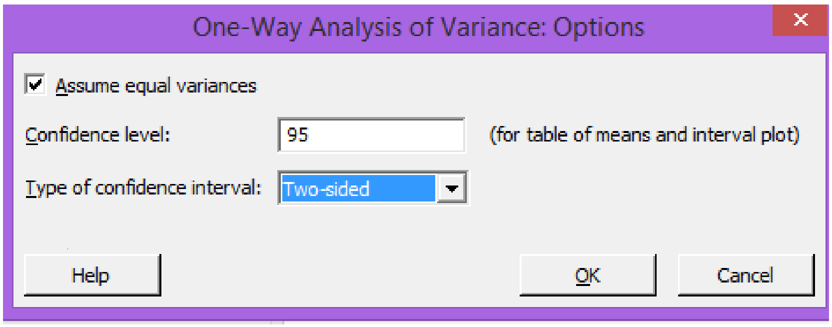
7)

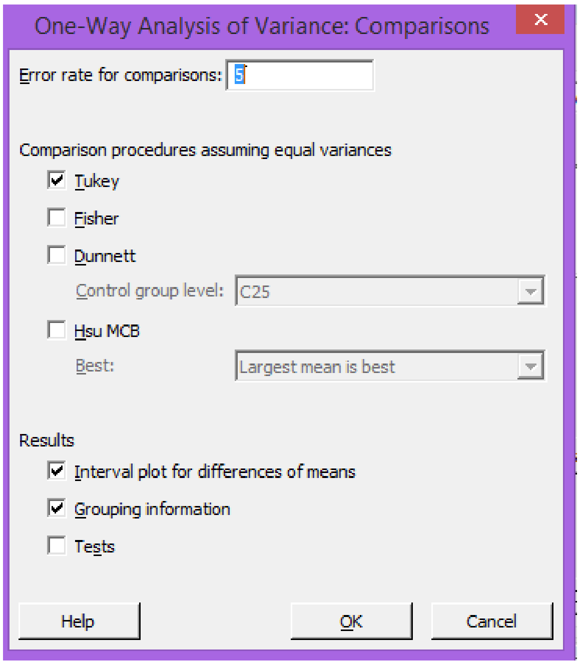
A firm wishes to compare four programs for training workers to perform a certain manual task. Twenty new employees are randomly assigned to the training programs, with 5 in each program. At the end of the training period, a test is conducted to see how quickly trainees can perform the task. The number of times the task is performed per minute is recorded for each trainee, with the following results:

Observation	*Program 1*	*Program 2*	*Program 3*	*Program 4*
1	9	10	12	9
2	12	6	14	8
3	14	9	11	11
4	11	9	13	7
5	13	10	11	8

(One-way ANOVA by Minitab)



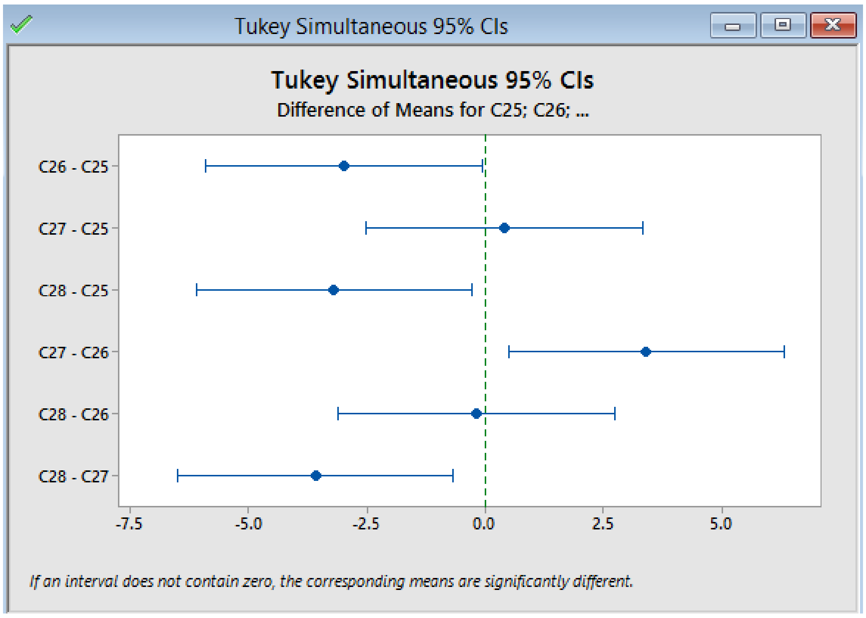




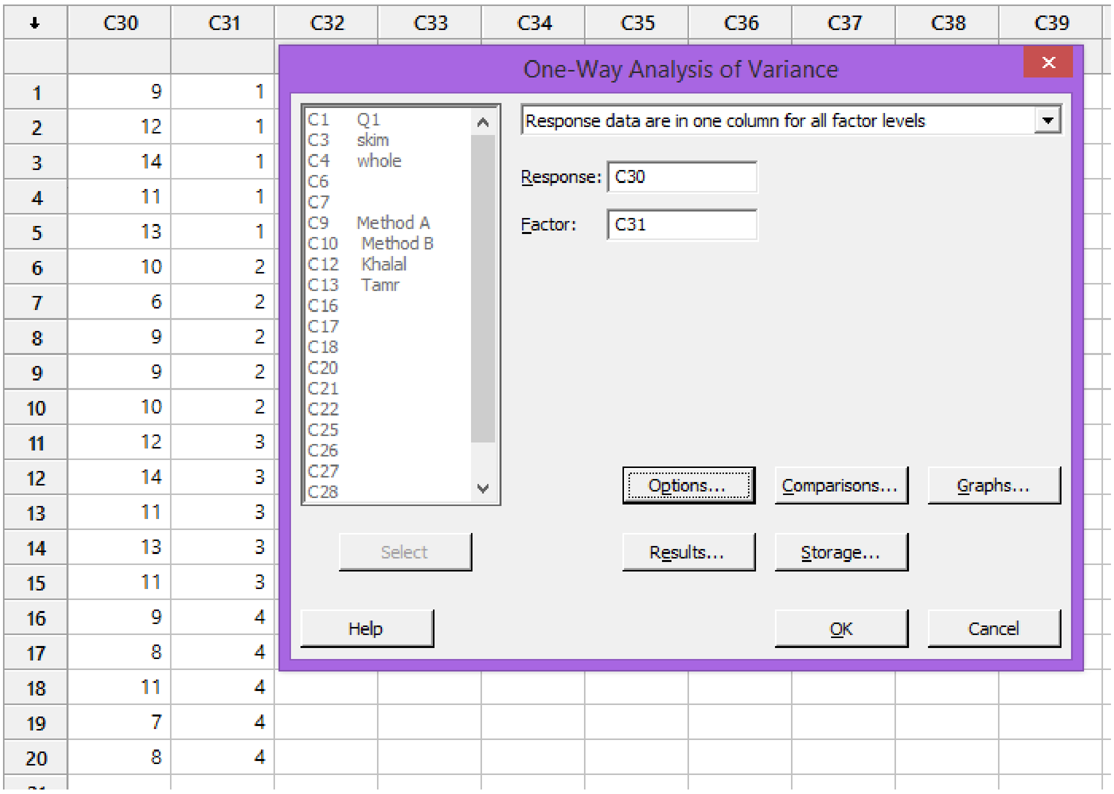
One-way ANOVA: C25; C26; C27; C28

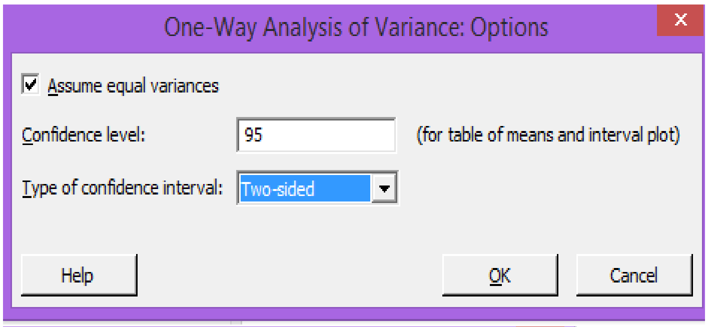
```
Method

Null hypothesis         All means are equal
Alternative hypothesis  At least one mean is different
Significance level      α = 0.05

Equal variances were assumed for the analysis.


Factor Information

Factor  Levels  Values
Factor       4  C25; C26; C27; C28


Analysis of Variance

Source  DF  Adj SS  Adj MS  F-Value  P-Value
Factor   3   54.95  18.317     7.04    0.003
Error   16   41.60   2.600
Total   19   96.55
```

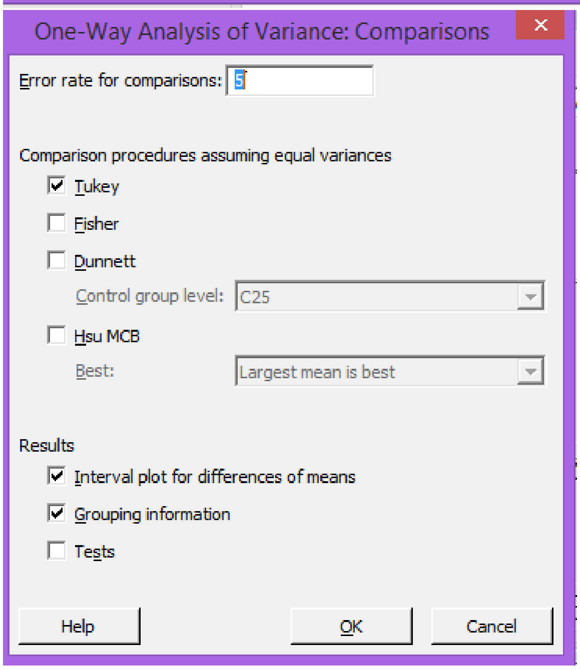
Tukey Simultaneous 95% CIs



Or

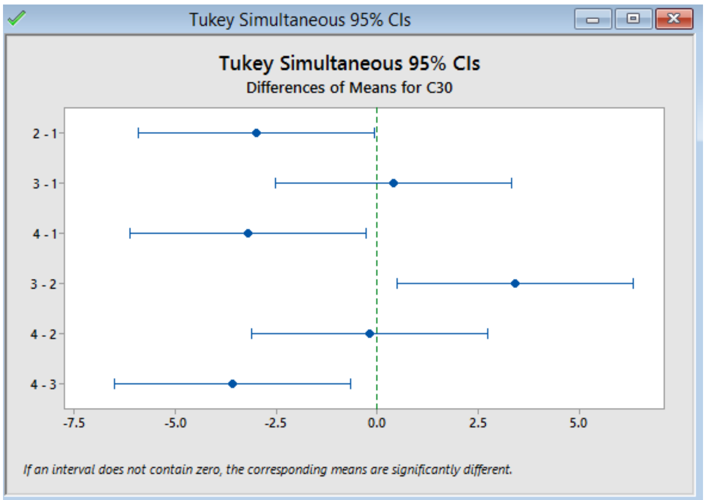

C30
C31
C32
C33
C34
C35
C36
C37
C38
C39
One-Way Analysis of Variance
C1 Q1
C3 skim
C4 whole
C6
C7
C9 Method A
C10 Method B
C12 Khalal
C13 Tamr
C16
C17
C18
C20
C21
C22
C25
C26
C27
C28
Response data are in one column for all factor levels
Response: C30
Factor: C31
Options...
Comparisons...
Graphs...
Select
Results...
Storage...
Help
OK
Cancel
1 9 1
2 12 1
3 14 1
4 11 1
5 13 1
6 10 2
7 6 2
8 9 2
9 9 2
10 10 2
11 12 3
12 14 3
13 11 3
14 13 3
15 11 3
16 9 4
17 8 4
18 11 4
19 7 4
20 8 4


One-Way Analysis of Variance: Options
Assume equal variances
Confidence level: 95
(for table of means and interval plot)
Type of confidence interval: Two-sided
Help
OK
Cancel


One-Way Analysis of Variance: Comparisons
Error rate for comparisons: 5
Comparison procedures assuming equal variances
Tukey
Fisher
Dunnett
Control group level: C25
Hsu MCB
Best: Largest mean is best
Results
Interval plot for differences of means
Grouping information
Tests
Help
OK
Cancel

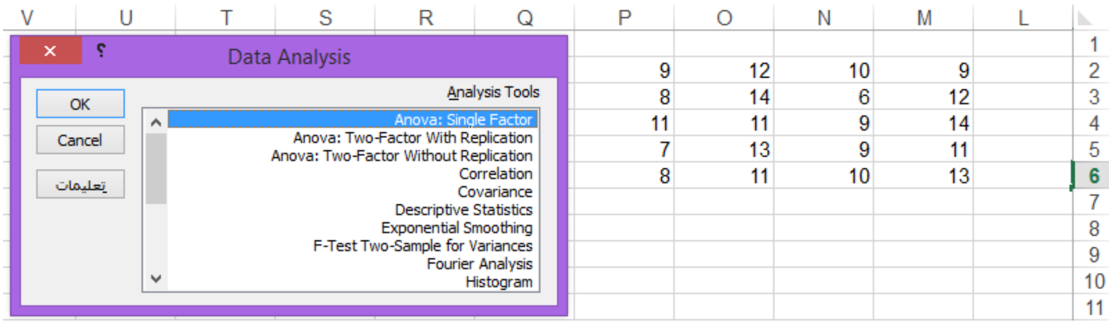
One-way ANOVA: C30 versus C31

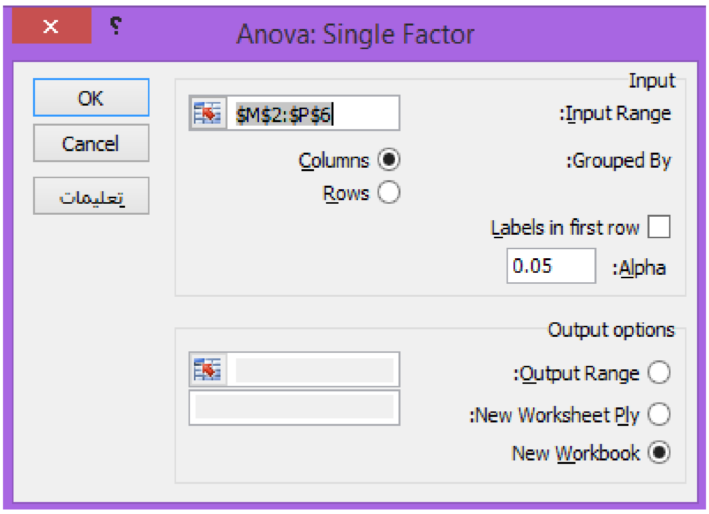
```
Method

Null hypothesis         All means are equal
Alternative hypothesis  At least one mean is different
Significance level      α = 0.05

Equal variances were assumed for the analysis.


Factor Information

Factor  Levels  Values
C31          4  1; 2; 3; 4


Analysis of Variance

Source  DF  Adj SS  Adj MS  F-Value  P-Value
C31      3   54.95  18.317     7.04    0.003
Error   16   41.60   2.600
Total   19   96.55
```



Tukey Simultaneous 95% CIs

(One-way ANOVA by Excel)





328 stat

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Column 1	5	59	11.8	3.7
Column 2	5	44	8.8	2.7
Column 3	5	61	12.2	1.7
Column 4	5	43	8.6	2.3

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	54.95	3	18.31667	7.044872	0.003113	3.238872
Within Groups	41.6	16	2.6			
Total	96.55	19				

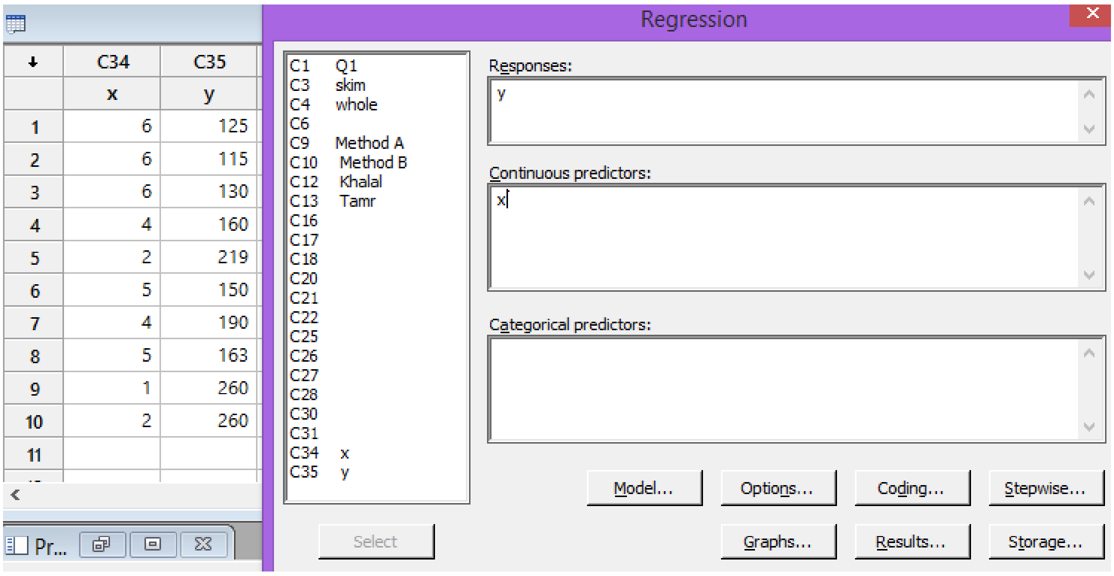
8)

Ten Corvettes between 1 and 6 years old were randomly selected from last year's sales records in Virginia Beach, Virginia. The following data were obtained, where x denotes age, in years, and y denotes sales price, in hundreds of dollars.

x	6	6	6	4	2	5	4	5	1	2
y	125	115	130	160	219	150	190	163	260	260

a) Determine the regression equation for the data.
b) Compute and interpret the coefficient of determination, r^2 .
c) Obtain a point estimate for the mean sales price of all 4-year-old Corvettes.

(Linear regression by Minitab)

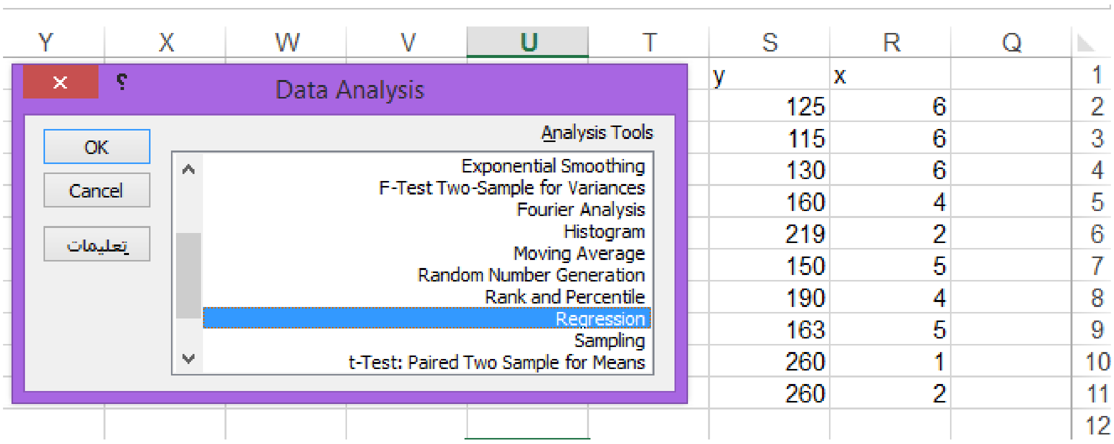


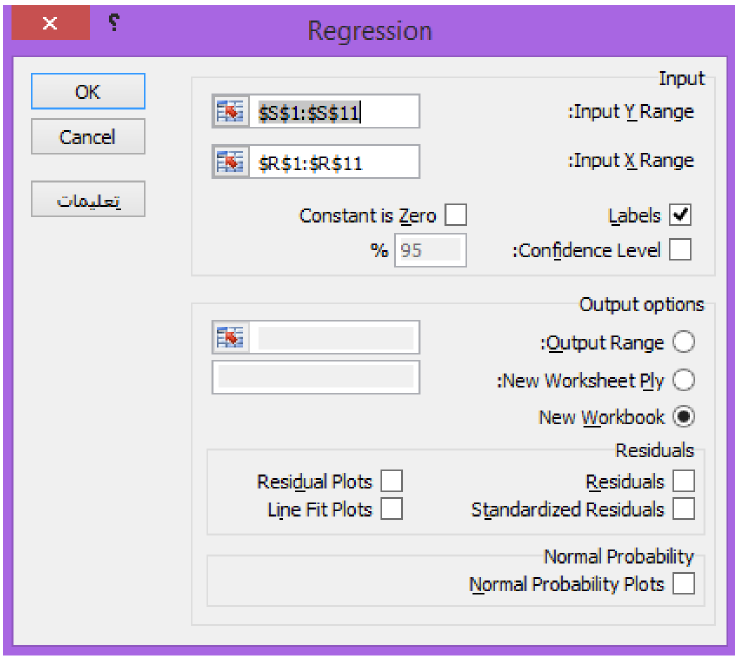
Regression Analysis: y versus x

|

```
Model Summary

      S    R-sq  R-sq(adj)  R-sq(pred)
14.2465  93.68%     92.89%      90.16%


Coefficients

Term         Coef  SE Coef  T-Value  P-Value   VIF
Constant    291.6     11.4    25.51    0.000
x          -27.90     2.56   -10.89    0.000  1.00


Regression Equation

y = 291.6 - 27.90 x
```

(Linear regression by Excel)





328 stat

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.967871585
R Square	0.936775406
Adjusted R Square	0.928872332
Standard Error	14.24652913
Observations	10

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	24057.89126	24057.89	118.533	4.48427E-06
Residual	8	1623.708738	202.9636		
Total	9	25681.6			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	291.6019417	11.43289905	25.50551	5.98E-09	265.2376293	317.9662542	265.2376293	317.9662542
x	-27.90291262	2.562889198	-10.8873	4.48E-06	-33.81294571	-21.99287953	-33.81294571	-21.99287953

Spss

Q1)

For a sample of 10 fruits from thirteen-year-old acidless orange trees, the fruit shape (determined as adiameter divided by height) wae measured [Shaheen and Hamouda (1984b)]:
1.066 1.084 1.076 1.051 1.059 1.020 1.035 1.052 1.046 0.976
Assuming that fruit shapes are approximately normally distributed, find and interpret a 90% confidence interval for the average fruit shape.

Q2)

The phosphorus content was measured for independent samples of skim and whole

Whole: 94.95 95.15 94.85 94.55 94.55 93.40 95.05 94.35 94.70 94.90

Skim: 91.25 91.80 91.50 91.65 91.15 90.25 91.90 91.25 91.65 91.00

Assuming normal populations with equal variances

a) Test whether the average phosphorus content of skim milk is less than the average phosphorus content of whole milk. Use $\alpha=0.01$
b) Find and interpret a 99% confidence interval for the difference in average phosphorus contents of whole and skim milk

Q3)

What is the relationship between the gender of the students and the assignment of a Pass or No Pass test grade? (Pass = score 70 or above).

	Pass	No Pass	Row Totals
Males	12	3	15
Females	13	2	15
Column Totals	25	5	30

Q4)

A firm wishes to compare four programs for training workers to perform a certain manual task. Twenty new employees are randomly assigned to the training programs, with 5 in each program. At the end of the training period, a test is conducted to see how quickly trainees can perform the task. The number of times the task is performed per minute is recorded for each trainee, with the following results:

Observation	*Program 1*	*Program 2*	*Program 3*	*Program 4*
1	9	10	12	9
2	12	6	14	8
3	14	9	11	11
4	11	9	13	7
5	13	10	11	8

Q5)

Ten Corvettes between 1 and 6 years old were randomly selected from last year's sales records in Virginia Beach, Virginia. The following data were obtained, where x denotes age, in years, and y denotes sales price, in hundreds of dollars.

x	6	6	6	4	2	5	4	5	1	2
y	125	115	130	160	219	150	190	163	260	260

a) Compute and interpret the linear correlation coefficient, r.
b) Determine the regression equation for the data.
c) Compute and interpret the coefficient of determination, r^2 .
d) Obtain a point estimate for the mean sales price of all 4-year-old Corvettes.

Q1) to use the T- test, we need to make sure that the population follows a normal distribution [redacted] i.e.

$$H_0: the\ population\ follows \text{ a normal distribution}$$

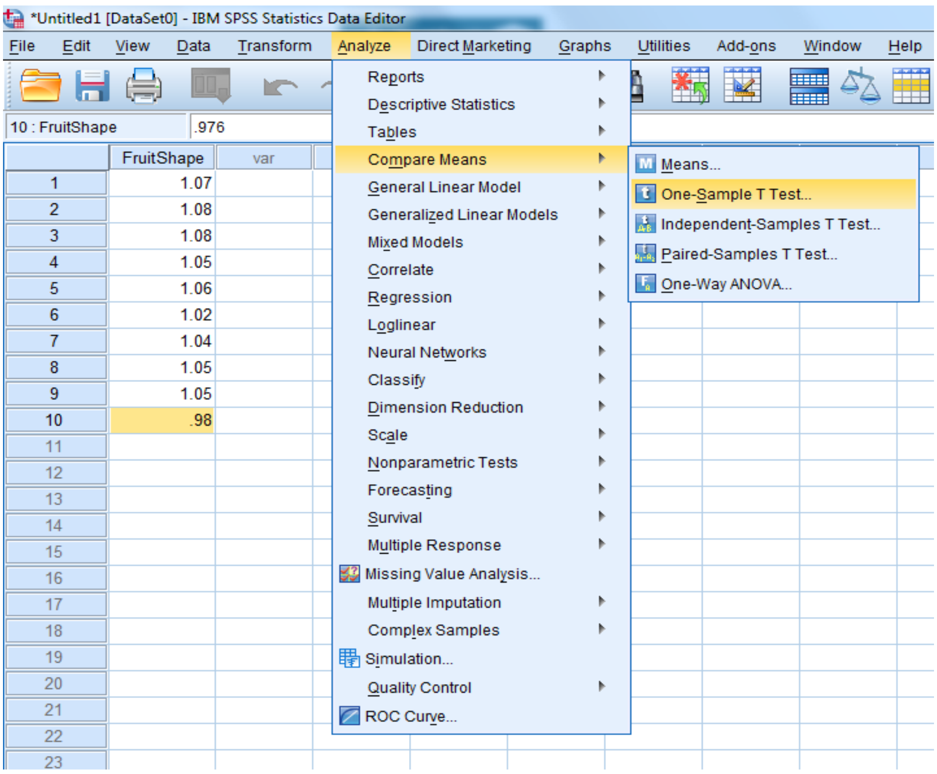
$$Vs$$

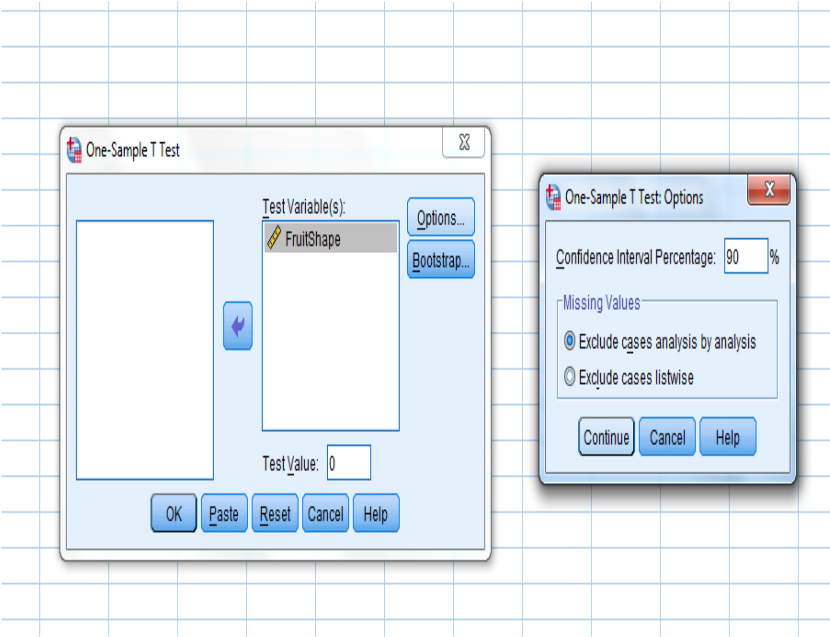
$$H_1: the\ population\ \ does\ not\ follow \text{ a normal distribution}$$

However, we find the question he said that the population follows a normal distribution, so is not necessary to make this test.

Now, 90% Confidence interval of the mean can be found in two ways:

1) The first method:





T-Test

[DataSet0]

One-Sample Statistics

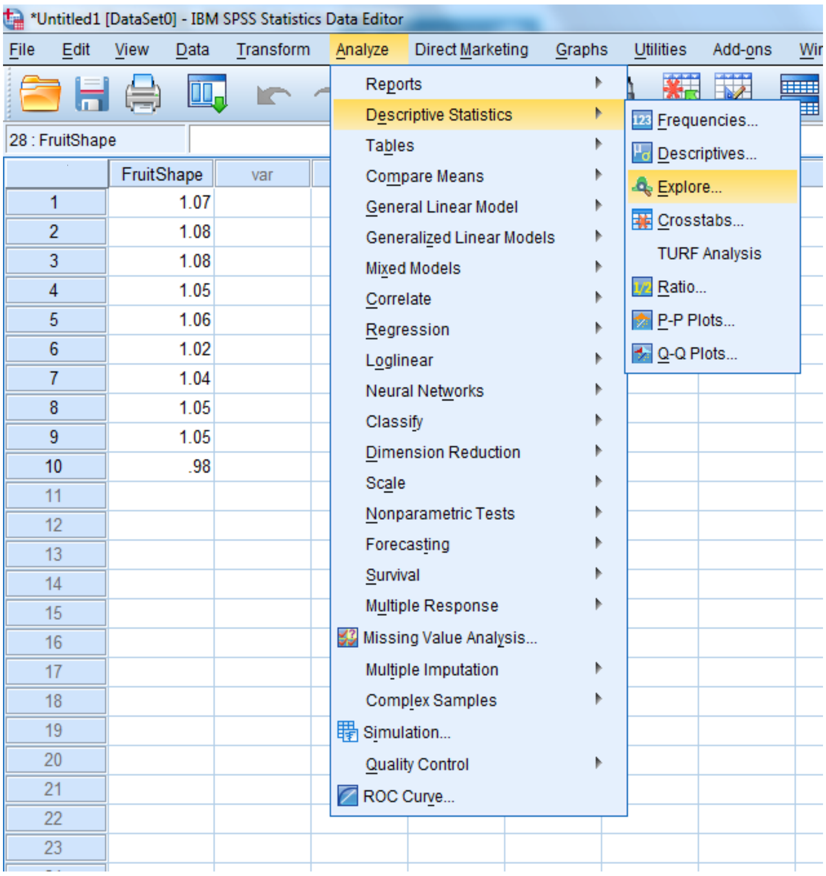
	N	Mean	Std. Deviation	Std. Error Mean
FruitShape	10	1.0465	.03103	.00981

One-Sample Test

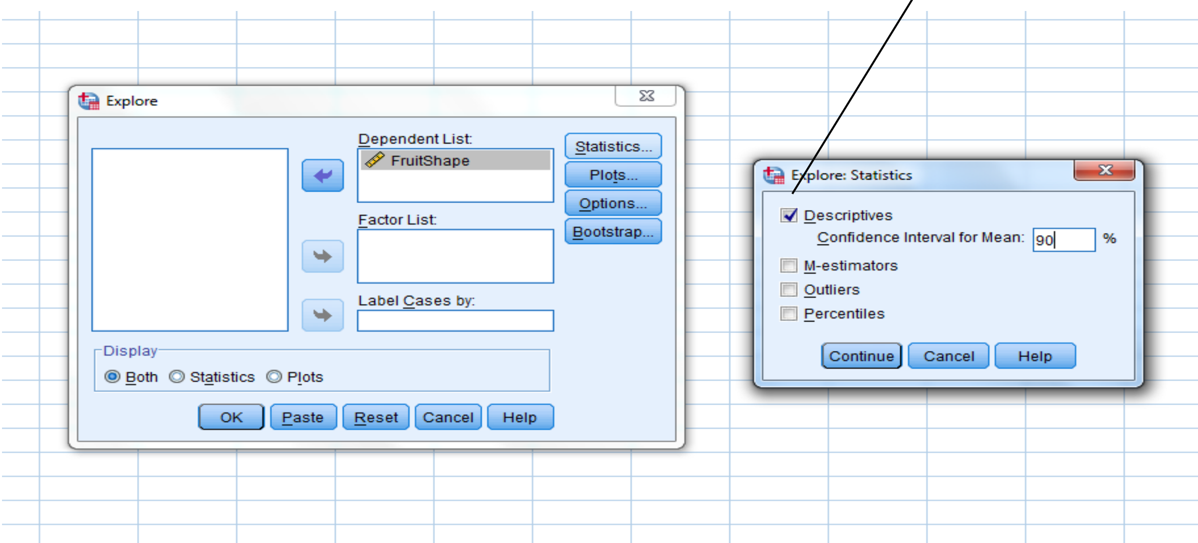
	Test Value = 0					
					90% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
FruitShape	106.632	9	.000	1.04650	1.0285	1.0645

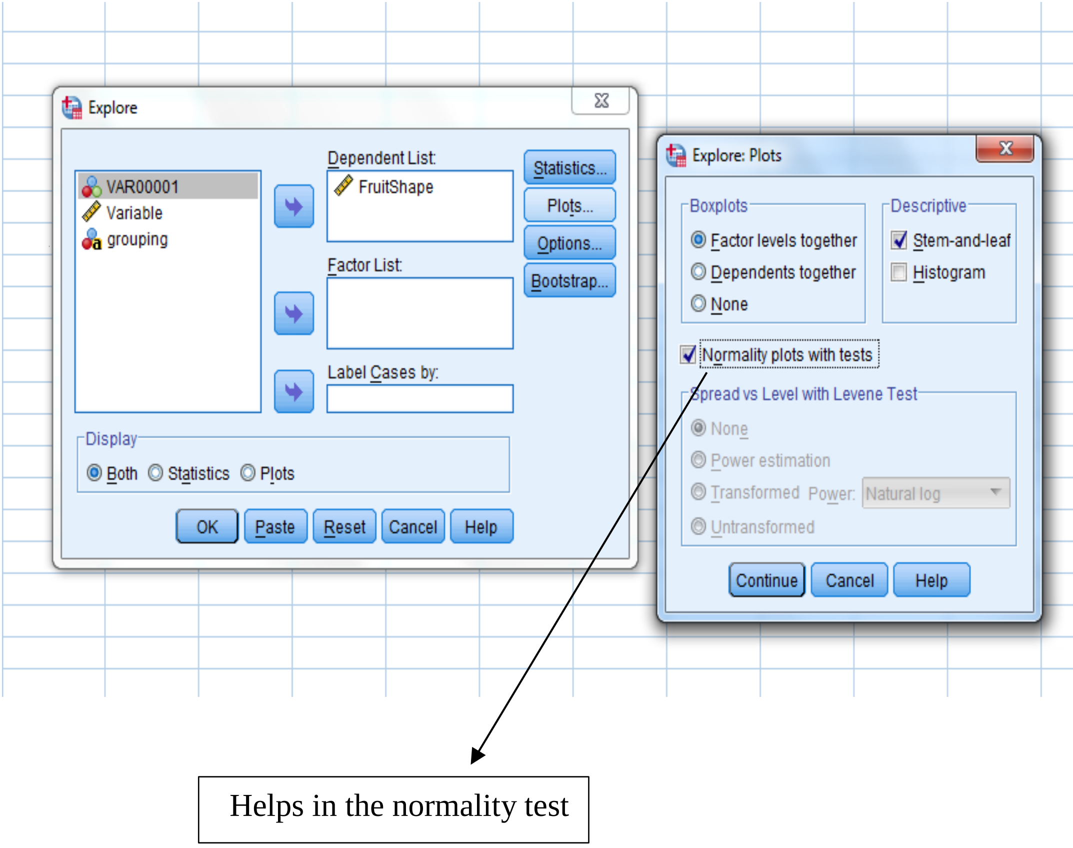
C.I for the mean

2) The second method:



It helps in the calculation of the confidence interval and find the statistical measures




Explore
VAR00001
Variable
grouping
Dependent List:
FruitShape
Factor List:
Label Cases by:
Statistics...
Plots...
Options...
Bootstrap...
Display
Both
Statistics
Plots
OK
Paste
Reset
Cancel
Help
Explore: Plots
Boxplots
Factor levels together
Dependents together
None
Descriptive
Stem-and-leaf
Histogram
Normality plots with tests
Spread vs Level with Levene Test
None
Power estimation
Transformed
Power:
Natural log
Untransformed
Continue
Cancel
Help
Helps in the normality test

Explore

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
FruitShape	10	50.0%	10	50.0%	20	100.0%

Descriptives

			Statistic	Std. Error
FruitShape	Mean		1.0465	.00981
	90% Confidence Interval for Mean	Lower Bound	1.0285	
		Upper Bound	1.0645	
	5% Trimmed Mean		1.0483	
	Median		1.0515	
	Variance		.001	
	Std. Deviation		.03103	
	Minimum		.98	
	Maximum		1.08	
	Range		.11	
	Interquartile Range		.04	
	Skewness		-1.313	.687
	Kurtosis		2.276	1.334

C.I for the mean

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
FruitShape	.194	10	.200*	.907	10	.260

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

a. Lilliefors Significance Correction

As $\text{P} - \text{value} > .1$

So, we except H_0: $the\ population\ follows$ a normal distribution

<u>Q2)</u> to use the T- test for two sample, we need to make sure that

1) The independence of the two samples: It is very clear that there is no correlation between the values of the two samples.

2) The populations follow a normal distribution [redacted] i.e.

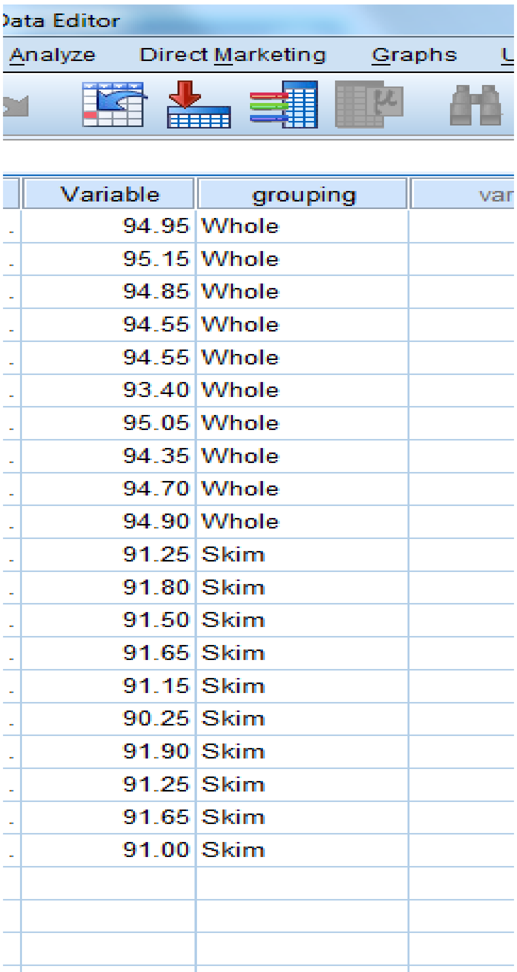
$$H_0: the\ two\ populations\ follow \text{ a normal distribution}$$

$$Vs$$

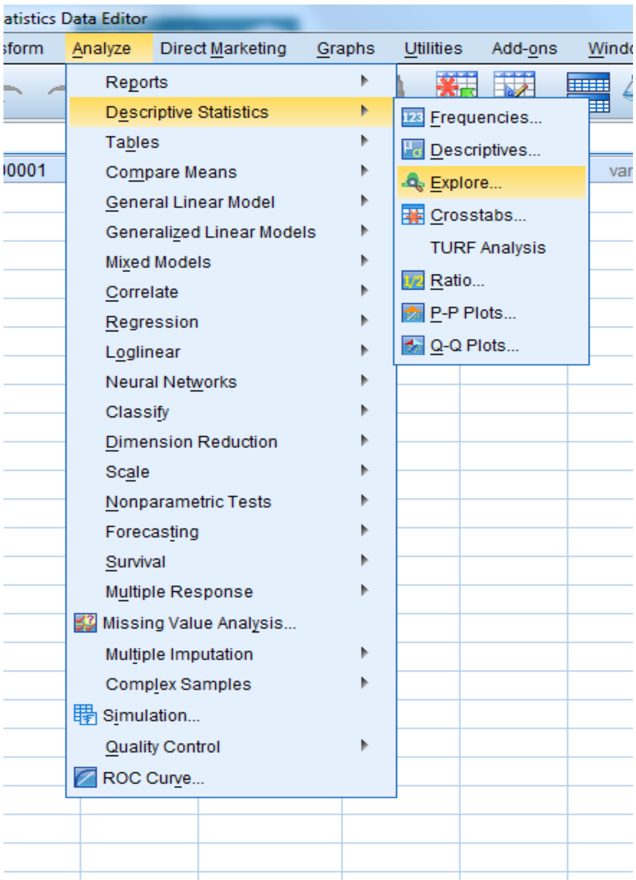
$$H_1: the\ two\ populations\ \ do\ not\ follow \text{ a normal distribution}$$

However, we find the question he said that the populations follows a normal distribution, so is not necessary to make this test.

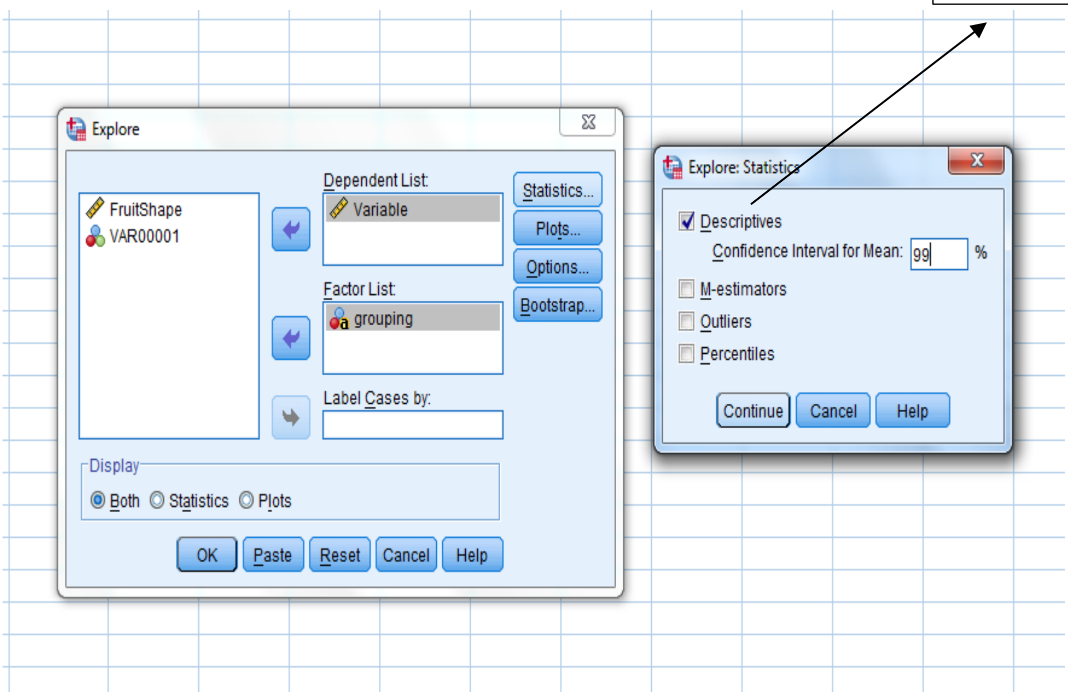
*To make sure no more..............

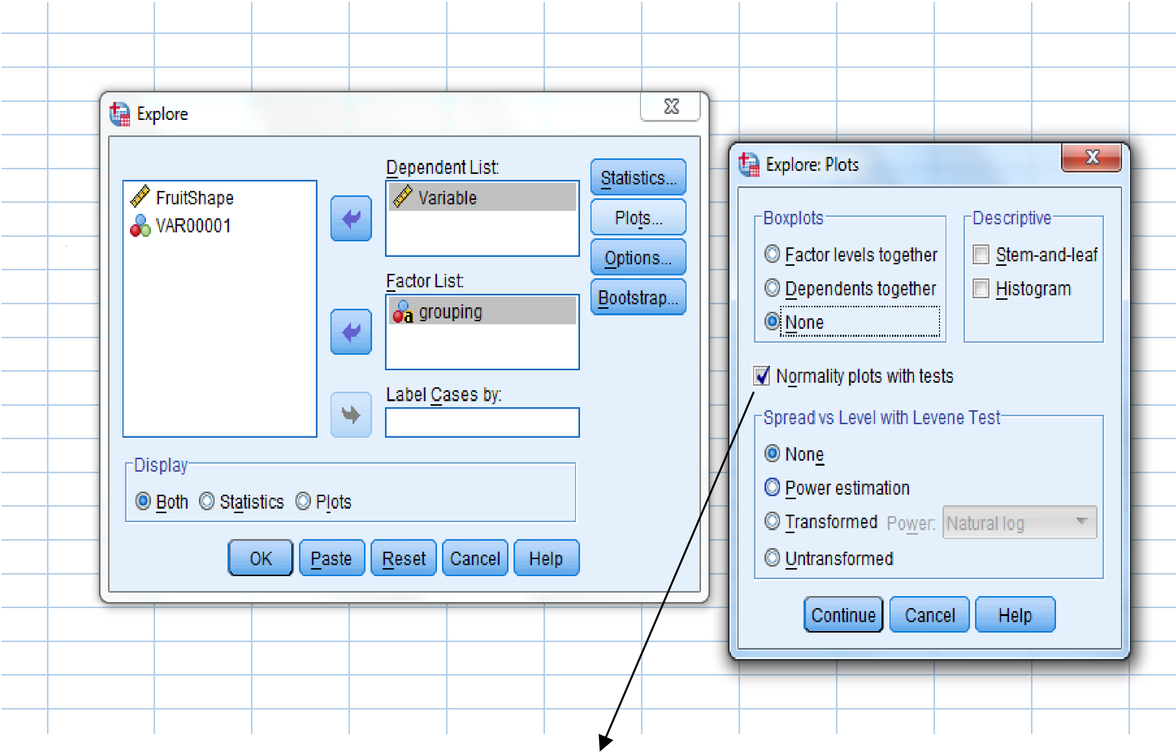


Variable	grouping	var
94.95	Whole	
95.15	Whole	
94.85	Whole	
94.55	Whole	
94.55	Whole	
93.40	Whole	
95.05	Whole	
94.35	Whole	
94.70	Whole	
94.90	Whole	
91.25	Skim	
91.80	Skim	
91.50	Skim	
91.65	Skim	
91.15	Skim	
90.25	Skim	
91.90	Skim	
91.25	Skim	
91.65	Skim	
91.00	Skim	



It helps in the calculation of the confidence interval and find statistical measures for each sample





Helps in the normality test

Explore

grouping

Case Processing Summary

	grouping	Cases					
		Valid		Missing		Total	
		N	Percent	N	Percent	N	Percent
Variable	Skim	10	100.0%	0	0.0%	10	100.0%
	Whole	10	100.0%	0	0.0%	10	100.0%

Descriptives

	grouping			Statistic	Std. Error
Variable	Skim	Mean		91.3400	.15272
		99% Confidence Interval for Mean	Lower Bound	90.8437	
			Upper Bound	91.8363	
		5% Trimmed Mean		91.3694	
		Median		91.3750	
		Variance		.233	
		Std. Deviation		.48293	
		Minimum		90.25	
		Maximum		91.90	
		Range		1.65	
		Interquartile Range		.57	
		Skewness		-1.241	.687
		Kurtosis		2.035	1.334
	Whole	Mean		94.6450	.15907
		99% Confidence Interval for Mean	Lower Bound	94.1281	
			Upper Bound	95.1619	
		5% Trimmed Mean		94.6861	
		Median		94.7750	
		Variance		.253	
		Std. Deviation		.50302	
		Minimum		93.40	
		Maximum		95.15	
		Range		1.75	
		Interquartile Range		.47	
		Skewness		-1.864	.687
		Kurtosis		4.241	1.334

Double-click to activate

C.I for the mean for the skim

C.I for the mean for the whole

Tests of Normality

	grouping	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Variable	Skim	.147	10	.200*	.902	10	.232
	Whole	.225	10	.163	.823	10	.028

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

a. Lilliefors Significance Correction

As $P - \text{value} > .01$ for both populations.

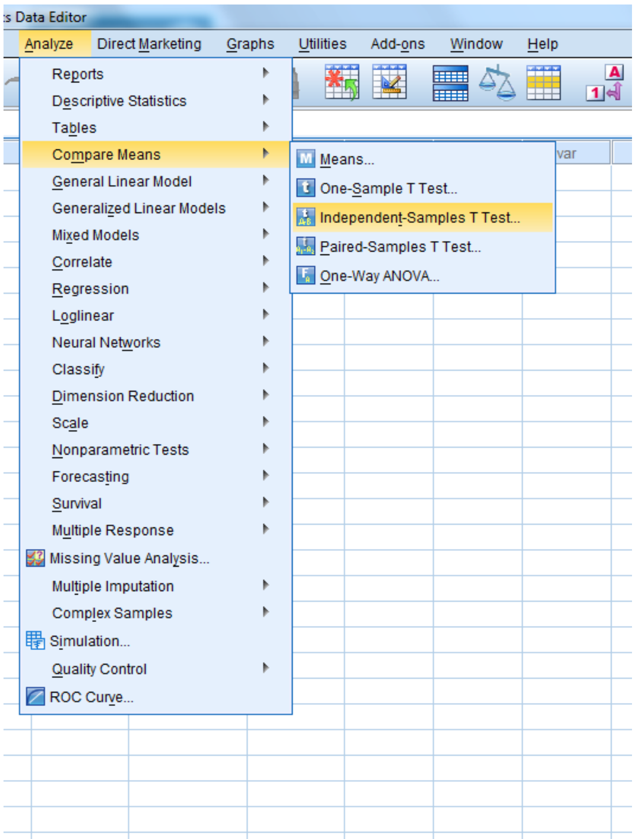
So, we except H_0: *the two populations follow* a normal distribution

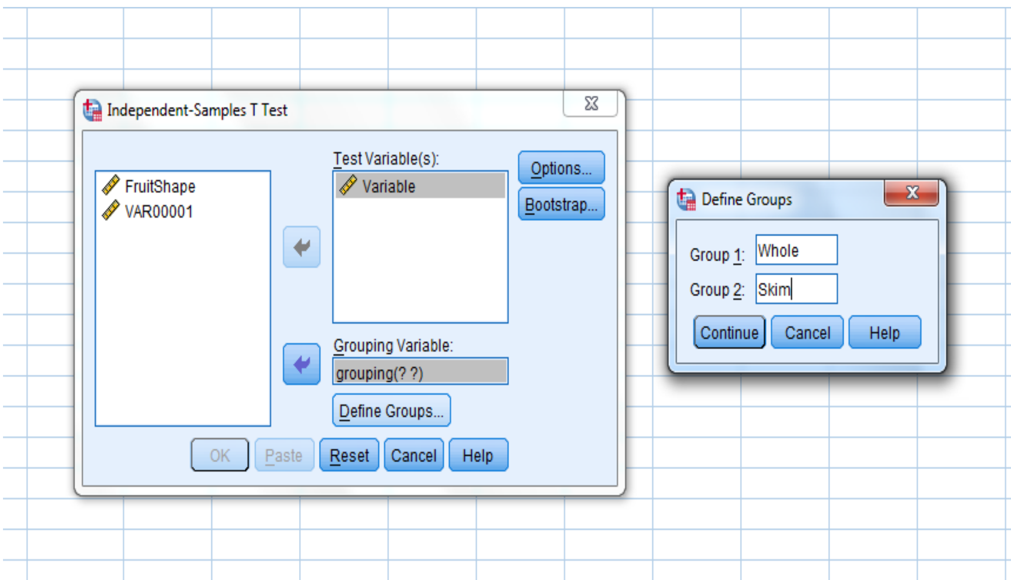
Now, the goal of the question:

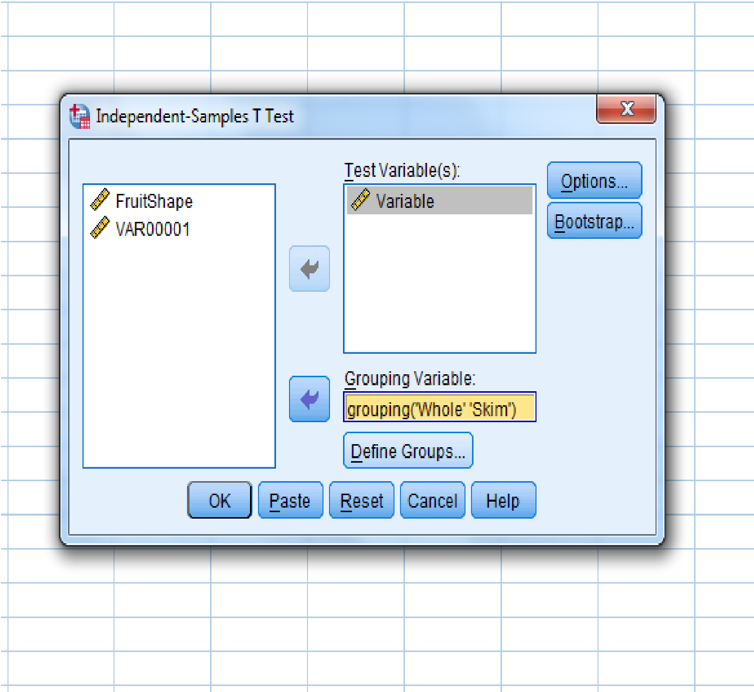
a) $H_0: \mu_{whole} - \mu_{skim} = 0 \;\; Vs \;\; H_1: \mu_{whole} - \mu_{skim} > 0 \;\; at\ \alpha = .01$

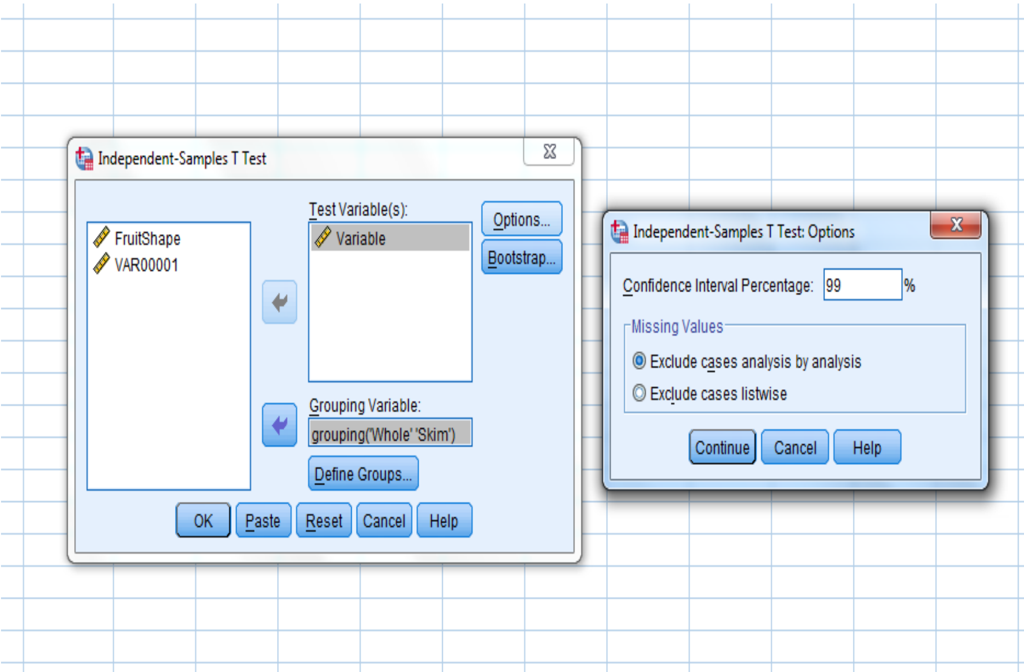
and

b) 90% Confidence interval of $\mu_{whole} - \mu_{skim}$






Independent-Samples T Test
FruitShape
VAR00001
Test Variable(s):
Variable
Options...
Bootstrap...
Grouping Variable:
grouping('Whole' 'Skim')
Define Groups...
OK
Paste
Reset
Cancel
Help


Independent-Samples T Test
FruitShape
VAR00001
Test Variable(s):
Variable
Options...
Bootstrap...
Grouping Variable:
grouping('Whole' 'Skim')
Define Groups...
OK
Paste
Reset
Cancel
Help
Independent-Samples T Test: Options
Confidence Interval Percentage:
99
%
Missing Values
Exclude cases analysis by analysis
Exclude cases listwise
Continue
Cancel
Help

This for test

$$H_0: \sigma^2_{whole} = \sigma^2_{skim} \quad Vs \quad H_1: \sigma^2_{whole} \neq \sigma^2_{skim}$$

As P − value > .01 .So, we except H_0. However, it is given in question.

T-Test

Group Statistics

	grouping	N	Mean	Std. Deviation	Std. Error Mean
Variable	Whole	10	94.6450	.50302	.15907
	Skim	10	91.3400	.48293	.15272

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									99% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Variable	Equal variances assumed	.009	.924	14.988	18	.000	3.30500	.22051	2.67027	3.93973
	Equal variances not assumed			14.988	17.970	.000	3.30500	.22051	2.67015	3.93985

$0/2 = 0$ but as $t = 14.988 > 0$ so $P - value = P(T_{18} > t) = 0$ then we reject $H_0: \mu_{whole} - \mu_{skim} = 0$.

99% C.I for $\mu_{whole} - \mu_{skim}$

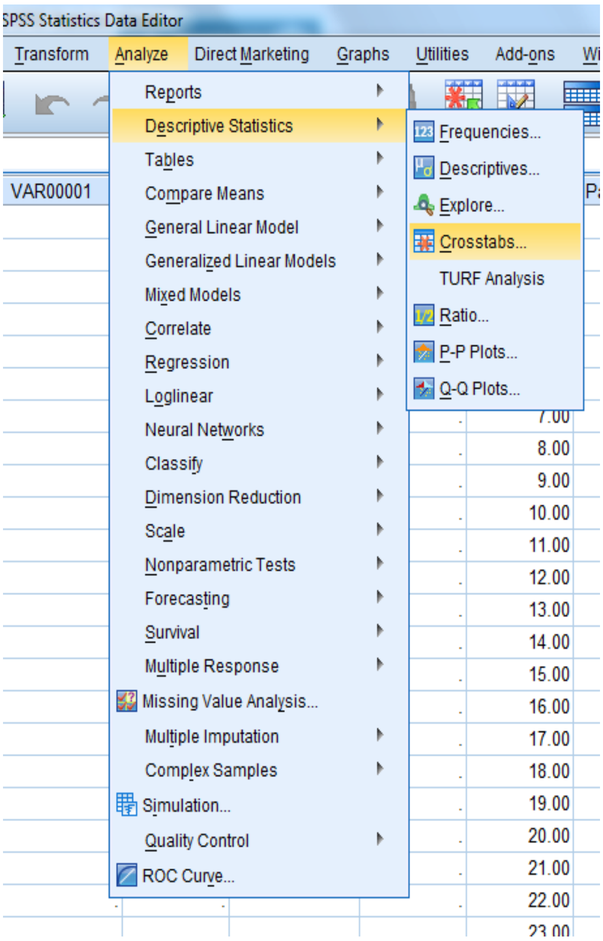
Q3)

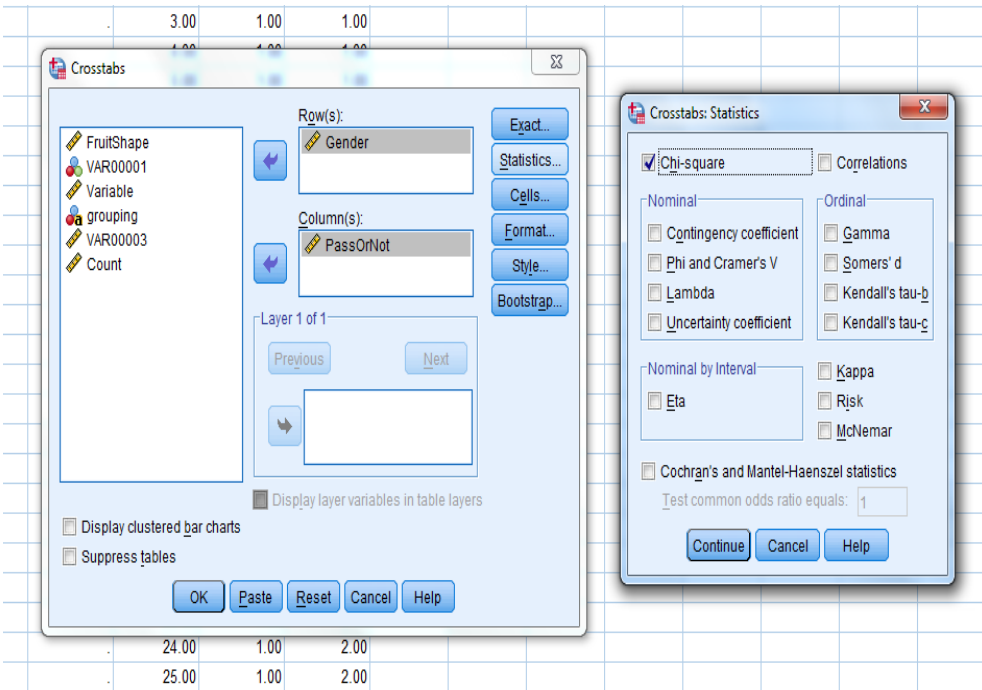
H_0: the gender of the students is indep. of a Pass or No Pass test grade

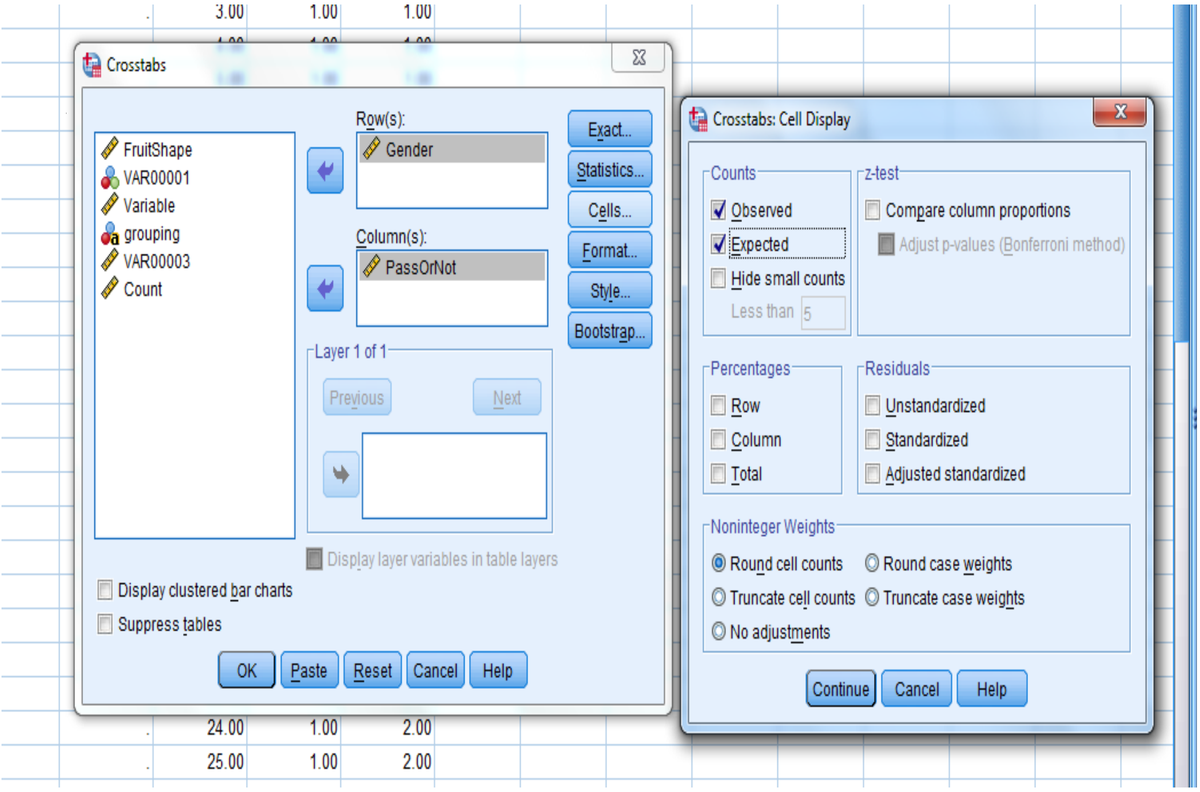
Vs

H_1: the gender of the students is not indep. of a Pass or No Pass test grade

Count	PassOrNot	Gender	var
1.00	1.00	1.00	
2.00	1.00	1.00	
3.00	1.00	1.00	
4.00	1.00	1.00	
5.00	1.00	1.00	
6.00	1.00	1.00	
7.00	1.00	1.00	
8.00	1.00	1.00	
9.00	1.00	1.00	
10.00	1.00	1.00	
11.00	1.00	1.00	
12.00	1.00	1.00	
13.00	2.00	1.00	
14.00	2.00	1.00	
15.00	2.00	1.00	
16.00	1.00	2.00	
17.00	1.00	2.00	
18.00	1.00	2.00	
19.00	1.00	2.00	
20.00	1.00	2.00	
21.00	1.00	2.00	
22.00	1.00	2.00	
23.00	1.00	2.00	
24.00	1.00	2.00	
25.00	1.00	2.00	
26.00	1.00	2.00	
27.00	1.00	2.00	
28.00	1.00	2.00	
29.00	2.00	2.00	
30.00	2.00	2.00	


SPSS Statistics Data Editor
Transform
Analyze
Direct Marketing
Graphs
Utilities
Add-ons
VAR00001
Reports
Descriptive Statistics
Tables
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear
Neural Networks
Classify
Dimension Reduction
Scale
Nonparametric Tests
Forecasting
Survival
Multiple Response
Missing Value Analysis...
Multiple Imputation
Complex Samples
Simulation...
Quality Control
ROC Curve...
Frequencies...
Descriptives...
Explore...
Crosstabs...
TURF Analysis
Ratio...
P-P Plots...
Q-Q Plots...
8.00
9.00
10.00
11.00
12.00
13.00
14.00
15.00
16.00
17.00
18.00
19.00
20.00
21.00
22.00


Crosstabs
FruitShape
VAR00001
Variable
grouping
VAR00003
Count
Row(s):
Gender
Column(s):
PassOrNot
Layer 1 of 1
Previous
Next
Display layer variables in table layers
Display clustered bar charts
Suppress tables
Exact...
Statistics...
Cells...
Format...
Style...
Bootstrap...
OK
Paste
Reset
Cancel
Help
Crosstabs: Statistics
Chi-square
Correlations
Nominal
Contingency coefficient
Phi and Cramer's V
Lambda
Uncertainty coefficient
Ordinal
Gamma
Somers' d
Kendall's tau-b
Kendall's tau-c
Nominal by Interval
Eta
Kappa
Risk
McNemar
Cochran's and Mantel-Haenszel statistics
Test common odds ratio equals:
Continue
Cancel
Help


Crosstabs
FruitShape
VAR00001
Variable
grouping
VAR00003
Count
Row(s):
Gender
Column(s):
PassOrNot
Layer 1 of 1
Previous
Next
Display layer variables in table layers
Display clustered bar charts
Suppress tables
Exact...
Statistics...
Cells...
Format...
Style...
Bootstrap...
OK
Paste
Reset
Cancel
Help
Crosstabs: Cell Display
Counts
Observed
Expected
Hide small counts
Less than 5
z-test
Compare column proportions
Adjust p-values (Bonferroni method)
Percentages
Row
Column
Total
Residuals
Unstandardized
Standardized
Adjusted standardized
Noninteger Weights
Round cell counts
Round case weights
Truncate cell counts
Truncate case weights
No adjustments
Continue
Cancel
Help

Crosstabs

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Gender * PassOrNot	30	100.0%	0	0.0%	30	100.0%

Gender * PassOrNot Crosstabulation

			PassOrNot		Total
			1.00	2.00	
Gender	1.00	Count	12	3	15
		Expected Count	12.5	2.5	15.0
	2.00	Count	13	2	15
		Expected Count	12.5	2.5	15.0
Total		Count	25	5	30
		Expected Count	25.0	5.0	30.0

The Chi-Square statistic

$df = (2 - 1) * (2 - 1)$

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.240[a]	1	.624		
Continuity Correction[b]	.000	1	1.000		
Likelihood Ratio	.241	1	.623		
Fisher's Exact Test				1.000	.500
Linear-by-Linear Association	.232	1	.630		
N of Valid Cases	30				

a. 2 cells (50.0%) have expected count less than 5. The minimum expected count is 2.50.

b. Computed only for a 2x2 table

As we can see that 2 cells have expected count less than 5 because these 2 cells contain less than 5 observations. So the solution is will be Merge cells until we get the expectation greater than 5 but here it is not possible, so take a larger sample.

$P - value > (\alpha = .05)$ so we except H_0

<u>Q4)</u> to use the one way ANOVA- test, we need to make sure that

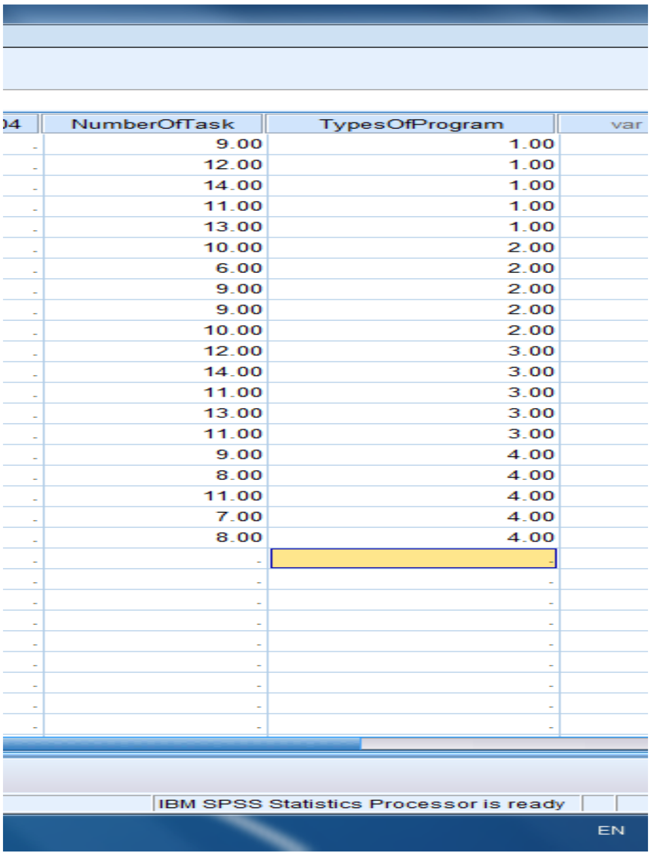
1) The independence of the four samples: It is very clear that there is no correlation between the values of the four samples.

2) The populations follow a normal distribution [redacted] i.e.

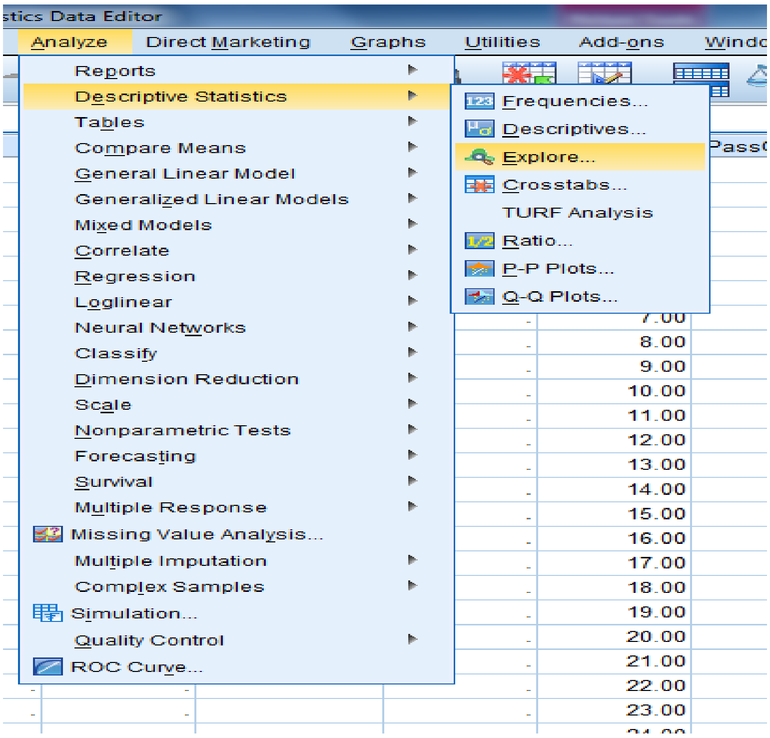
$$H_0: the\ four\ populations\ follow \text{ a normal distribution}$$

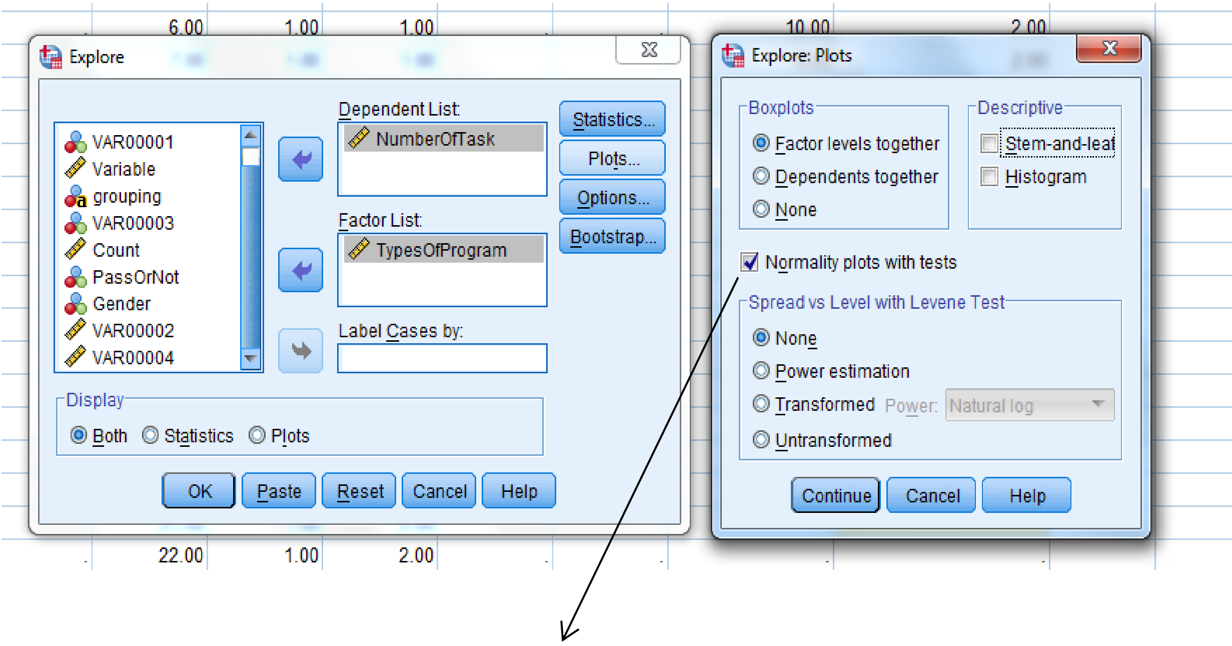
$$Vs$$

$$H_1: the\ four\ populations\ \ do\ not\ follow \text{ a normal distribution}$$



NumberOfTask	TypesOfProgram
9.00	1.00
12.00	1.00
14.00	1.00
11.00	1.00
13.00	1.00
10.00	2.00
6.00	2.00
9.00	2.00
9.00	2.00
10.00	2.00
12.00	3.00
14.00	3.00
11.00	3.00
13.00	3.00
11.00	3.00
9.00	4.00
8.00	4.00
11.00	4.00
7.00	4.00
8.00	4.00





Helps in the normality test

Explore

[DataSet1] E:\328\7 الدرس\Untitled1.sav

TypesOfProgram

Case Processing Summary

		Cases					
		Valid		Missing		Total	
	TypesOfProgram	N	Percent	N	Percent	N	Percent
NumberOfTask	1.00	5	100.0%	0	0.0%	5	100.0%
	2.00	5	100.0%	0	0.0%	5	100.0%
	3.00	5	100.0%	0	0.0%	5	100.0%
	4.00	5	100.0%	0	0.0%	5	100.0%

Descriptives

	TypesOfProgram			Statistic	Std. Error
NumberOfTask	1.00	Mean		11.8000	.86023
		95% Confidence Interval for Mean	Lower Bound	9.4116	
			Upper Bound	14.1884	
		5% Trimmed Mean		11.8333	
		Median		12.0000	
		Variance		3.700	
		Std. Deviation		1.92354	
		Minimum		9.00	
		Maximum		14.00	
		Range		5.00	
		Interquartile Range		3.50	
		Skewness		-.590	.913
		Kurtosis		-.022	2.000
	2.00	Mean		8.8000	.73485

				Statistic	Std. Error
	2.00	Mean		8.8000	.73485
		95% Confidence Interval for Mean	Lower Bound	6.7597	
			Upper Bound	10.8403	
		5% Trimmed Mean		8.8889	
		Median		9.0000	
		Variance		2.700	
		Std. Deviation		1.64317	
		Minimum		6.00	
		Maximum		10.00	
		Range		4.00	
		Interquartile Range		2.50	
		Skewness		-1.736	.913
		Kurtosis		3.251	2.000
	3.00	Mean		12.2000	.58310
		95% Confidence Interval for Mean	Lower Bound	10.5811	
			Upper Bound	13.8189	
		5% Trimmed Mean		12.1667	
		Median		12.0000	
		Variance		1.700	
		Std. Deviation		1.30384	
		Minimum		11.00	
		Maximum		14.00	
		Range		3.00	
		Interquartile Range		2.50	
		Skewness		.541	.913
		Kurtosis		-1.488	2.000
	4.00	Mean		8.6000	.67823
		95% Confidence Interval for Mean	Lower Bound	6.7169	
			Upper Bound	10.4831	
		5% Trimmed Mean		8.5556	
		Median		8.0000	
		Variance		2.300	
		Std. Deviation		1.51658	
		Minimum		7.00	
		Maximum		11.00	
		Range		4.00	
		Interquartile Range		2.50	
		Skewness		1.118	.913
		Kurtosis		1.456	2.000

Tests of Normality

	TypesOfProgram	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
NumberOfTask	1.00	.141	5	.200*	.979	5	.928
	2.00	.348	5	.047	.779	5	.054
	3.00	.221	5	.200*	.902	5	.421
	4.00	.254	5	.200*	.914	5	.492

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

a. Lilliefors Significance Correction

As P − value $>$ .05 for the four populations.

So, we except H_0: $the\ four\ populations\ follow$ a normal distribution

3) Homogeneity of Variance (to get a test of the assumption of homogeneity of variance) i.e.

$$H_0: \sigma^2_{program\ 1} = \sigma^2_{program\ 2} = \sigma^2_{program\ 3} = \sigma^2_{program\ 4}$$

i.e. the variances of each sample are equal

$$Vs$$

$$H_1: The\ variances\ are\ not\ all\ equal$$

This will be clear later.

Now, the **goal** of the question:

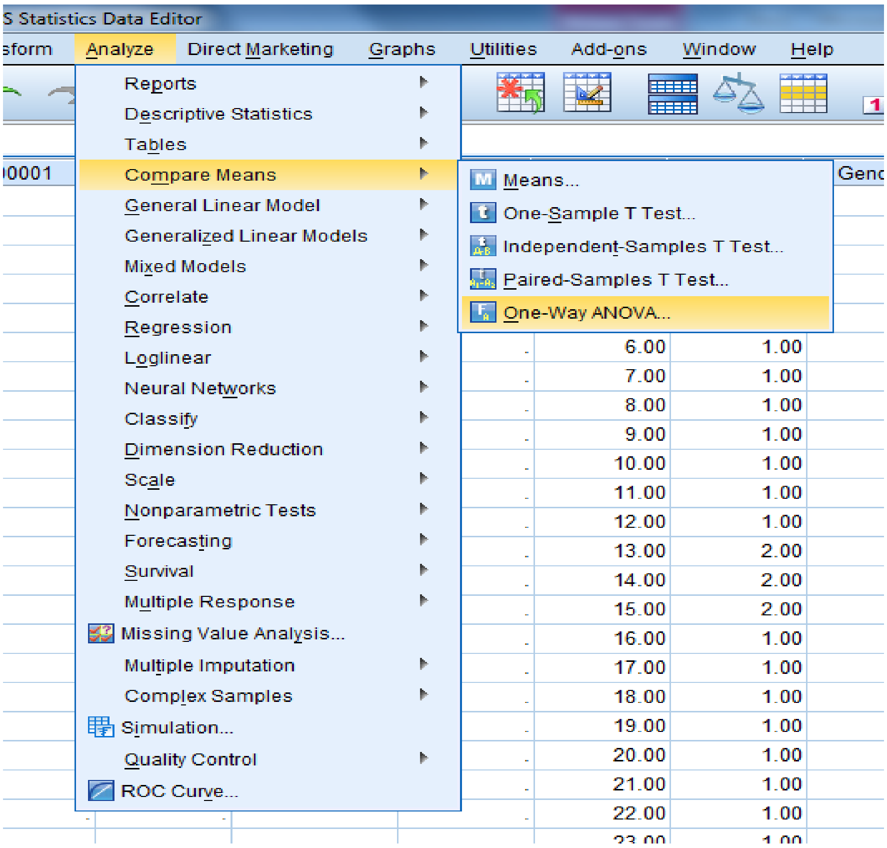
$$H_0: \mu_{program\ 1} = \mu_{program\ 2} = \mu_{program\ 3} = \mu_{program\ 4}$$

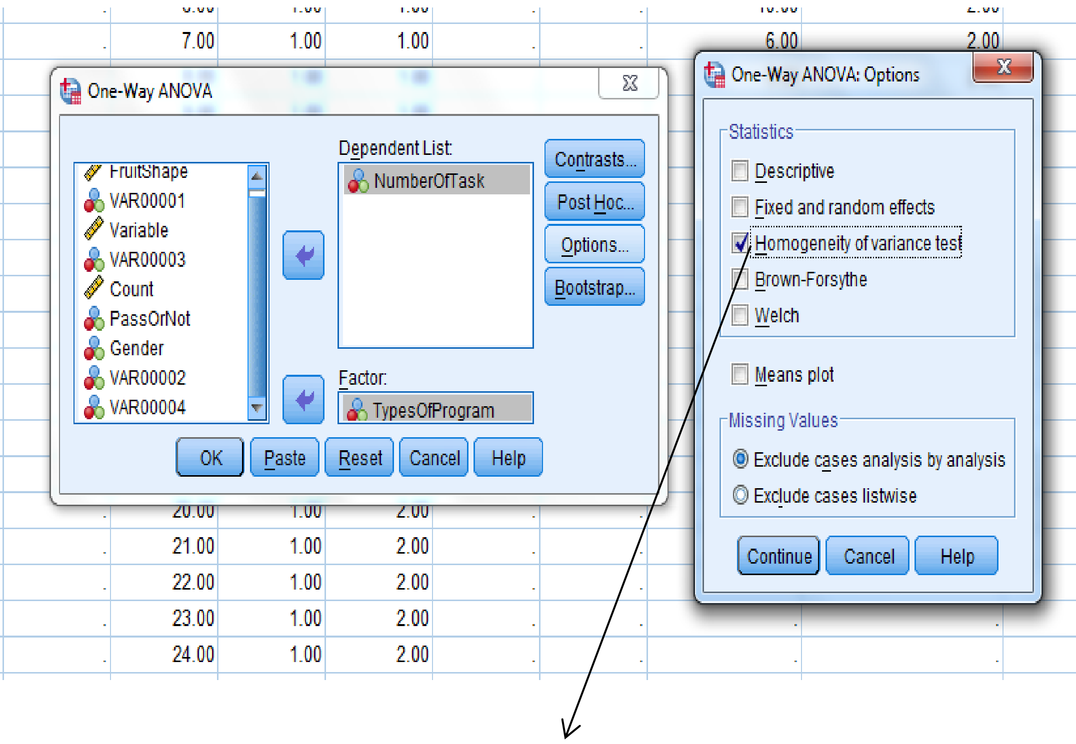
i.e. treatments are equally effective

$$Vs$$

$$H_1: The\ means\ are\ not\ all\ equal$$

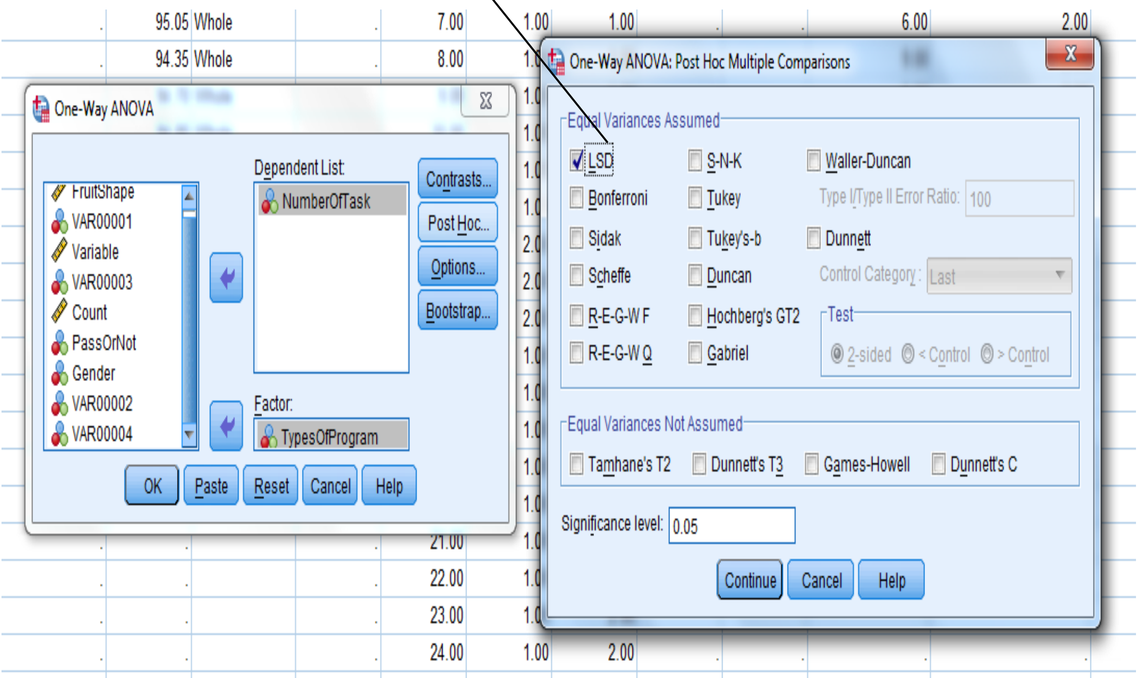
$$at\ \alpha = .05$$





Helps in the homogeneity of variance test

If we reject H_0 in Analysis of Variance (ANOVA one way-test) we need to look at the multiple comparisons output by use the appropriate post hoc procedure (LSD) to determine whether unique pairwise comparisons are significant.



As $P-\text{value} > .05$ .So, we except $H_0: \sigma^2_{program\ 1} = \sigma^2_{program\ 2} = \sigma^2_{program\ 3} = \sigma^2_{program\ 4}$

Oneway

Test of Homogeneity of Variances

NumberOfTask

Levene Statistic	df1	df2	Sig.
.190	3	16	.902

$= 4-1$

$= 20-4$

$= 20-1$

ANOVA

NumberOfTask

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	54.950	3	18.317	7.045	.003
Within Groups	41.600	16	2.600		
Total	96.550	19			

as $P-value < .05$,then we reject $H_0: \mu_{program\ 1} = \mu_{program\ 2} = \mu_{program\ 3} = \mu_{program\ 4}$.

Post Hoc Tests

Multiple Comparisons

Dependent Variable: NumberOfTask
LSD

(I) TypesOfProgram	(J) TypesOfProgram	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1.00	2.00	3.00000*	1.01980	.010	.8381	5.1619
	3.00	-.40000	1.01980	.700	-2.5619	1.7619
	4.00	3.20000*	1.01980	.006	1.0381	5.3619
2.00	1.00	-3.00000*	1.01980	.010	-5.1619	-.8381
	3.00	-3.40000*	1.01980	.004	-5.5619	-1.2381
	4.00	.20000	1.01980	.847	-1.9619	2.3619
3.00	1.00	.40000	1.01980	.700	-1.7619	2.5619
	2.00	3.40000*	1.01980	.004	1.2381	5.5619
	4.00	3.60000*	1.01980	.003	1.4381	5.7619
4.00	1.00	-3.20000*	1.01980	.006	-5.3619	-1.0381
	2.00	-.20000	1.01980	.847	-2.3619	1.9619
	3.00	-3.60000*	1.01980	.003	-5.7619	-1.4381

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

1) $H_0: \mu_{program\ 1} = \mu_{program\ 2}\ vs\ H_1: \mu_{program\ 1} \neq \mu_{program\ 2}\ at\ \alpha = .05$

$as\ \ P - value = .01 < .05$, then we reject H_0.

2) $H_0: \mu_{program\ 1} = \mu_{program\ 3}\ vs\ H_1: \mu_{program\ 1} \neq \mu_{program\ 3}\ at\ \alpha = .05$

$\text{as P} - \text{value} = .7 > .05$, then we except H_0.

3) $H_0: \mu_{program\ 1} = \mu_{program\ 4}\ vs\ H_1: \mu_{program\ 1} \neq \mu_{program\ 4}\ at\ \alpha = .05$

$as\ \ P - value = .006 < .05$, then we reject H_0.

4) $H_0: \mu_{program\ 2} = \mu_{program\ 3}\ vs\ H_1: \mu_{program\ 2} \neq \mu_{program\ 3}\ at\ \alpha = .05$

$as\ \ P - value = .004 < .05$, then we reject H_0.

5) $H_0: \mu_{program\ 2} = \mu_{program\ 4}\ vs\ H_1: \mu_{program\ 2} \neq \mu_{program\ 4}\ at\ \alpha = .05$

$\text{as P} - \text{value} = .847 > .05$, then we except H_0.

6) $H_0: \mu_{program\ 3} = \mu_{program\ 4}\ vs\ H_1: \mu_{program\ 3} \neq \mu_{program\ 4}\ at\ \alpha = .05$

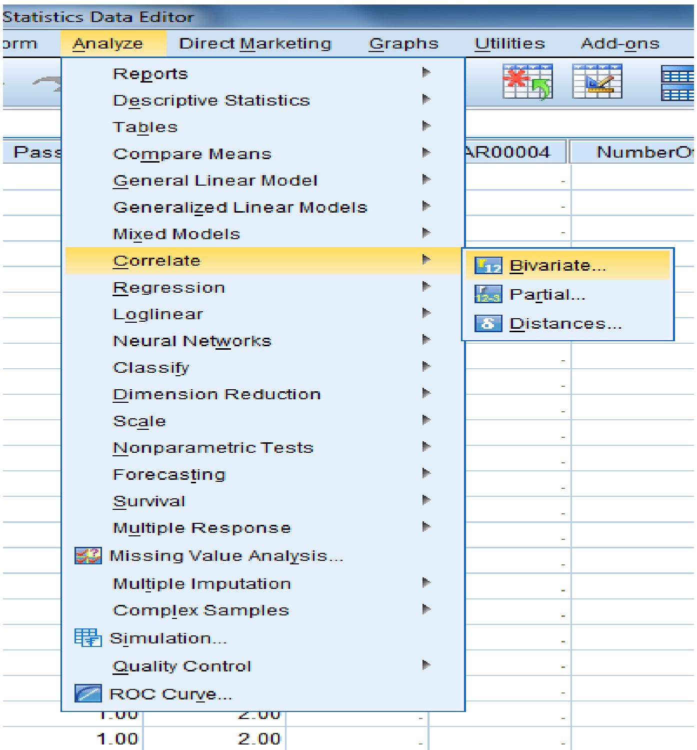
$as\ \ P - value = .003 < .05$, then we reject H_0.

Q5)

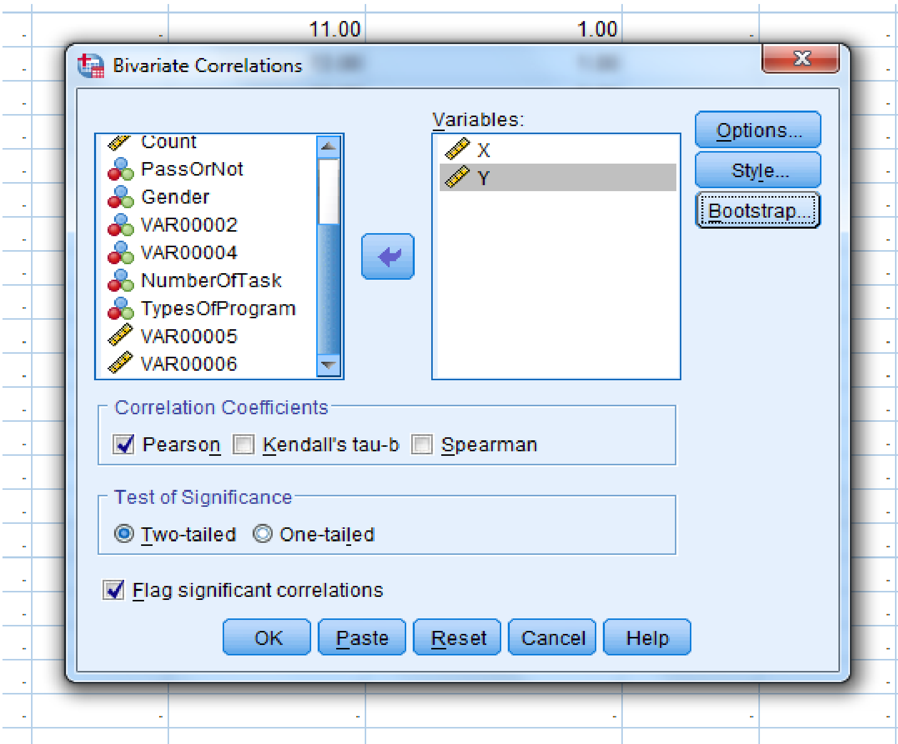
Enter the age values into one variable and the corresponding sales price values into another variable (see figure, below).

X	Y	var
6.00	125.00	
6.00	115.00	
6.00	130.00	
4.00	160.00	
2.00	219.00	
5.00	150.00	
4.00	190.00	
5.00	163.00	
1.00	260.00	
2.00	260.00	
.	.	
.	.	
.	.	
.	.	
.	.	
.	.	
.	.	
.	.	
.	.	
.	.	
.	.	
.	.	
.	.	
.	.	
.	.	
.	.	

a) Select Analyze ◊ Correlate ◊ Bivariate… (see figure, below).



Select "x" and "y" as the variables, select "Pearson" as the correlation coefficient, and click " "OK" (see the left figure, below).



Correlations

Correlations

		X	Y
X	Pearson Correlation	1	-.968**
	Sig. (2-tailed)		.000
	N	10	10
Y	Pearson Correlation	-.968**	1
	Sig. (2-tailed)	.000	
	N	10	10

**. Correlation is significant at the 0.01 level (2-tailed).

The correlation coefficient is –0.9679 which we can see that the relationship between x and y are –ve and strong.

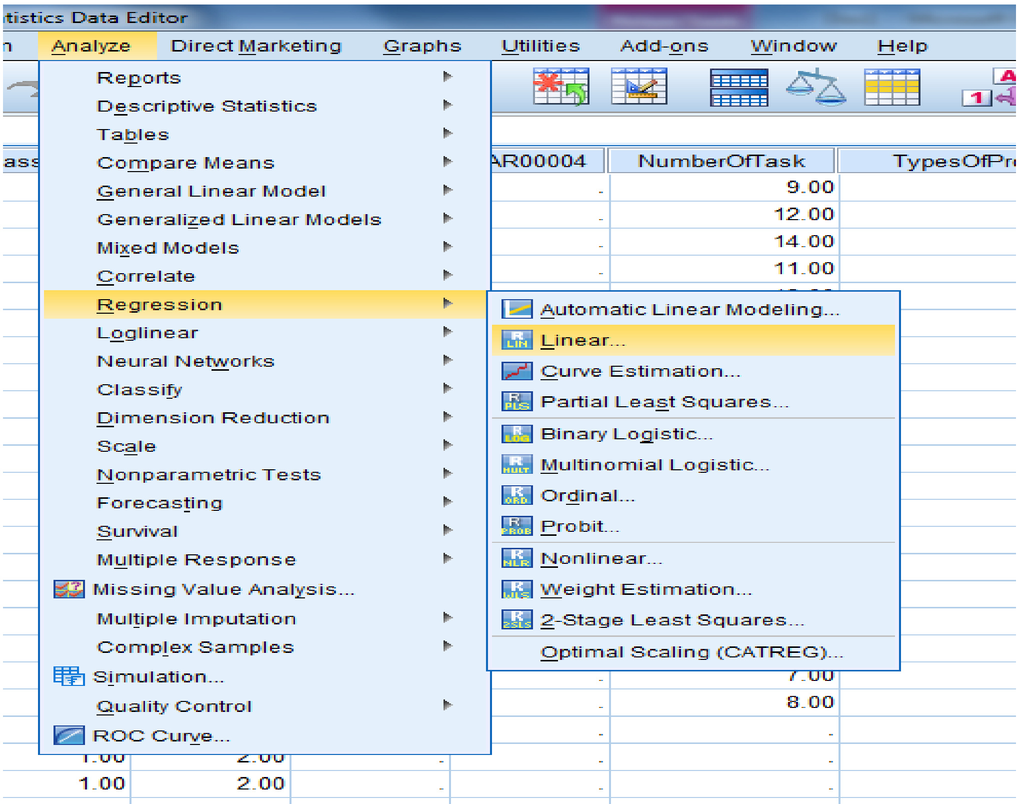
b, c and d)

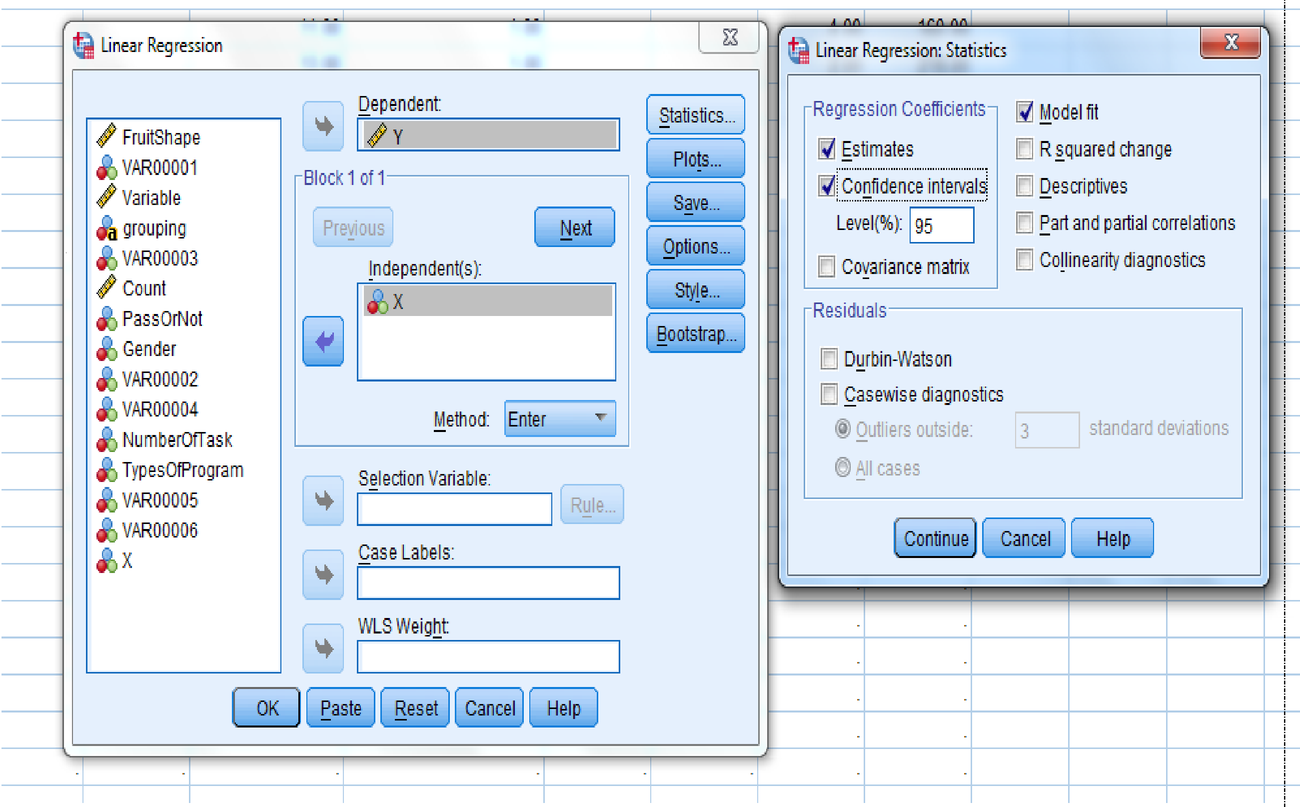
Since we eventually want to predict the price of 4-year-old Corvettes, enter the number "4" in the "x" variable column of the data window after the last row. Enter a "." for the corresponding "y" variable value (this lets SPSS know that we want a prediction for this value and not to include the value in any other computations) (see figure, below).

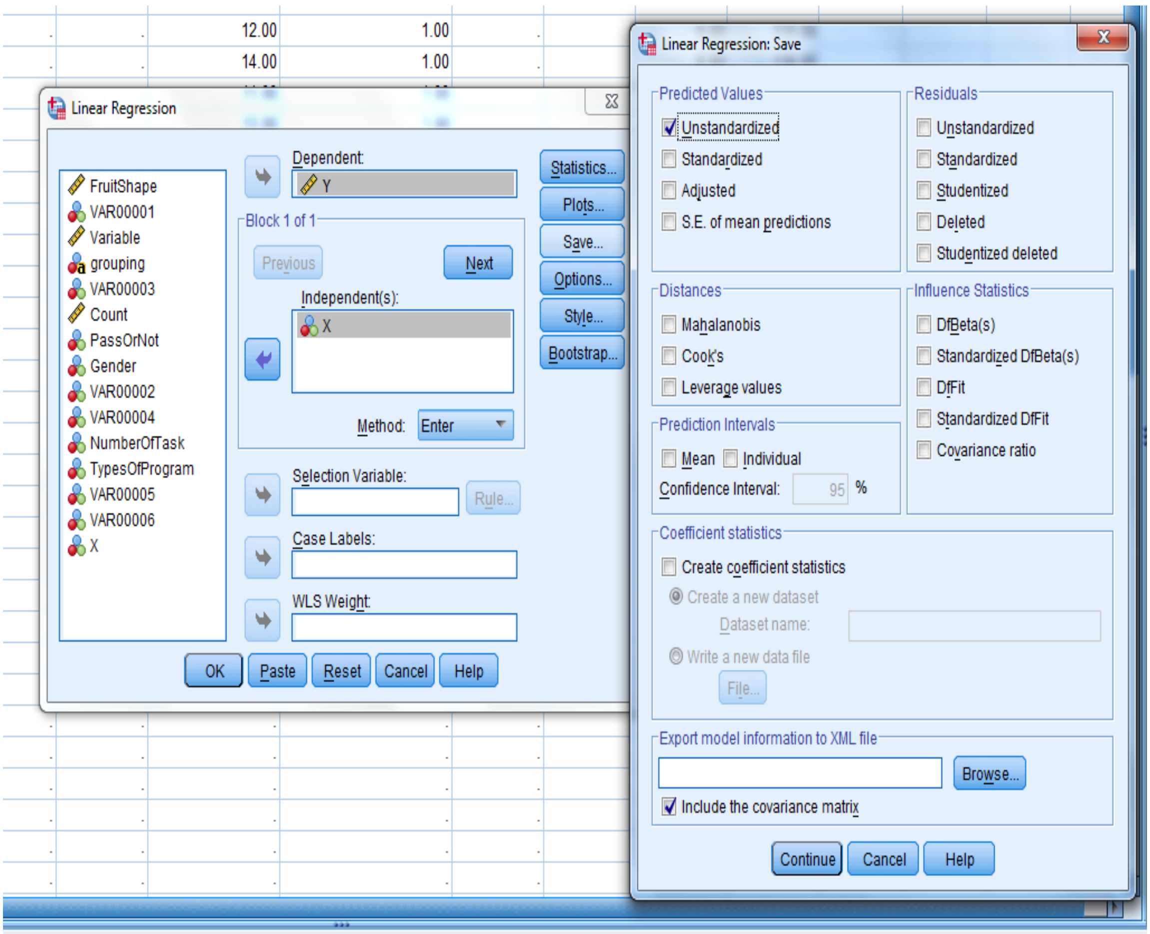
X	Y
6.00	125.00
6.00	115.00
6.00	130.00
4.00	160.00
2.00	219.00
5.00	150.00
4.00	190.00
5.00	163.00
1.00	260.00
2.00	260.00
4.00	.
.	.
.	.
.	.

Select Analyze ◊ Regression ◊ Linear… (see figure).

Select “y” as the dependent variable and “x” as the independent variable. Click “Statistics”, select “Estimates” and “Confidence Intervals” for the regression coefficients, select “Model fit” to obtain r ^{2} , and click “Continue”. Click “Save…”, select “Unstandardized” predicted values and click “Continue”. Click “OK”.




Linear Regression
FruitShape
VAR00001
Variable
grouping
VAR00003
Count
PassOrNot
Gender
VAR00002
VAR00004
NumberOfTask
TypesOfProgram
VAR00005
VAR00006
X
Dependent:
Y
Block 1 of 1
Previous
Next
Independent(s):
X
Method:
Enter
Selection Variable:
Rule...
Case Labels:
WLS Weight:
Statistics...
Plots...
Save...
Options...
Style...
Bootstrap...
OK
Paste
Reset
Cancel
Help
Linear Regression: Statistics
Regression Coefficients
Estimates
Confidence intervals
Level(%):
95
Covariance matrix
Model fit
R squared change
Descriptives
Part and partial correlations
Collinearity diagnostics
Residuals
Durbin-Watson
Casewise diagnostics
Outliers outside:
3
standard deviations
All cases
Continue
Cancel
Help


12.00
1.00
14.00
1.00
Linear Regression
FruitShape
VAR00001
Variable
grouping
VAR00003
Count
PassOrNot
Gender
VAR00002
VAR00004
NumberOfTask
TypesOfProgram
VAR00005
VAR00006
X
Dependent:
Y
Block 1 of 1
Previous
Next
Independent(s):
X
Method:
Enter
Selection Variable:
Rule...
Case Labels:
WLS Weight:
Statistics...
Plots...
Save...
Options...
Style...
Bootstrap...
OK
Paste
Reset
Cancel
Help
Linear Regression: Save
Predicted Values
Unstandardized
Standardized
Adjusted
S.E. of mean predictions
Residuals
Unstandardized
Standardized
Studentized
Deleted
Studentized deleted
Distances
Mahalanobis
Cook's
Leverage values
Influence Statistics
DfBeta(s)
Standardized DfBeta(s)
DfFit
Standardized DfFit
Covariance ratio
Prediction Intervals
Mean
Individual
Confidence Interval:
95
%
Coefficient statistics
Create coefficient statistics
Create a new dataset
Dataset name:
Write a new data file
File...
Export model information to XML file
Browse...
Include the covariance matrix
Continue
Cancel
Help

Regression

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.968[a]	.937	.929	14.24653

a. Predictors: (Constant), X

b. Dependent Variable: Y

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	24057.891	1	24057.891	118.533	.000[b]
	Residual	1623.709	8	202.964		
	Total	25681.600	9			

a. Dependent Variable: Y

b. Predictors: (Constant), X

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	291.602	11.433		25.506	.000	265.238	317.966
	X	-27.903	2.563	-.968	-10.887	.000	-33.813	-21.993

a. Dependent Variable: Y

X	Y	PRE_1
6.00	125.00	124.18447
6.00	115.00	124.18447
6.00	130.00	124.18447
4.00	160.00	179.99029
2.00	219.00	235.79612
5.00	150.00	152.08738
4.00	190.00	179.99029
5.00	163.00	152.08738
1.00	260.00	263.69903
2.00	260.00	235.79612
4.00	.	179.99029
.	.	.
.	.	.

From above, the regression equation is: y = 29160.1942 – (2790.2913)(x).

The coefficient of determination is 0.9368; therefore, about 93.68% of the variation in y data is explained by x.

328 stat

R

328 stat

Q1)

i- $\binom{150}{30}$, $\Gamma(18)$, $\ln(14)$, $\log(17)$

ii- $P(2 < X \le 4)$ *when* $X \sim Poisson(3)$

iii- Write R loop and the results to calculate

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \ z = -3.1, -3.0, \ldots -0.1, 0, 0.1, \ldots 3.0, 3.1.$$

iv- Write R code and the results to calculate:

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt, \ z = -3, -2, -1, 0, 1, 2, 3.$$

v- Write loop structure in R for generating 5 samples each of size 100 from $Binomial(5, 0.7)$. Then calculate the mean, standard deviation and coefficient of variation for each sample.

Q2)

(a) Write the commends and results to calculate the following:

(i) $P(-1.0 < T < 1.5)$, $\nu = 10$,

(ii) Find k such that $P(T < k) = 0.025$, $\nu = 12$,

(iii) $\binom{15}{9}$, $\log_{10}(25)$, $28!$

means $\frac{1}{rate} = \frac{1}{3}$
rate=3

(b) Generate a random sample of size 12 from the exponential (3) distribution and save it to A. Next, write an R command to create the column B such that

$$B_i = \begin{cases} 1, & A_i \le 3 \\ 2, & A_i > 3 \end{cases}, \ i = 1, 2, \ldots, 12.$$

328 stat

Q3)

(a) Find k when $P(X > k) = 0.04, \quad X \sim F(12,10)$

$\Rightarrow 1 - P(X < k) = .04$

$\Rightarrow 1 - .04 = P(X < K)$

(b) $P(3 < X \leq 7) \quad when \quad X \sim Poisson(3)$

beta dis. $= \dfrac{\Gamma 11}{Beta(6,5)}$

(c) $\int_0^1 x^5[1-x]^4 dx \qquad f(x) = \dfrac{\Gamma 6+5}{\Gamma 6 \, \Gamma 5} x^{6-1} (1-x)^{5-1}, \quad 0 < x < 1$

Q4)

$$A = \begin{bmatrix} 1 & 6 & 3 & -1 \\ 5 & 2 & 7 & 4 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 9 & 8 \\ 7 & 4 & 2 \\ 5 & 1 & 5 \\ 1 & 1 & 9 \end{bmatrix}, \quad C = \begin{bmatrix} 3 & 4 & 2 & 7 \\ 4 & 9 & 0 & 6 \\ 3 & 8 & 3 & 2 \\ 3 & 4 & 6 & 2 \end{bmatrix}$$

(a) A*B

(b) Determinant of C

(c) Inverse of C

4) The following data are two independent random samples from two independent populations $A \sim (\mu_A, \sigma^2)$ and $B \sim (\mu_B, \sigma^2)$, respectively.

A:	48	39	42	52	40	48	52	52	54	48
B:	50	48	42	40	43	48	50	46	38	38

Write R command and the results to

i- Test whether $\mu_A > \mu_B$.

ii- Construct 90% confidence interval of the difference $\mu_B - \mu_A$.

328 stat

Q5)

(a) Write the command and result to calculate the following:

log(80) = مكرر

ln(40) = مكرر

40!= مكرر

$\binom{50}{15}$= مكرر

$P(X > 2.22),\ where\ X \sim N(2,5)$

⇒ 1- P(X < 2.22)

(b) Write the commands and results to find the determent of the matrix and its inverse

مكرر

$$\begin{bmatrix} 1 & 0 & 4 & 7 \\ 8 & 3 & 1 & 9 \\ 7 & 4 & 2 & 8 \\ 0 & 9 & 5 & 6 \end{bmatrix}$$

```
###Q1
##i
   choose(150,30)
   gamma(18)
   log(14);log(14,base=exp(1))
   log(17,base=10);log10(17)

##ii
   ppois(4,lambda=3)-ppois(2,lambda=3)

##iii
   z <- seq(-3.1,3.1,by=.1)
   z
   for(i in z) {
   a=dnorm(i, mean = 0, sd = 1)
   cat(i," ",a,"\n")
   }
 #or
   for(i in z) {
   a=dnorm(i, mean = 0, sd = 1)
   print(c(i,a))
   }

##vi
   z <- seq(-3,3,by=1)
   z
   for(i in z) {
   b=pnorm(i, mean = 0, sd = 1)
   cat(i," ",b,"\n")
   }
 #or
   for(i in z) {
   b=pnorm(i, mean = 0, sd = 1)
   print(c(i,b))
   }


##v
   generating <- seq(1,5,by=1)
   generating
   generating <- c(1,2,3,4,5)
   generating
   for(i in generating) {
   c=rbinom(100, size=5, prob=.7)
   d <- mean(c)
   e <- sd(c)
   f <- e/d
   cat("sample:",c," ","mean=",d," ","sd=",e," ","cv=",f,"\n")
   }
 #or
   for(i in generating) {
   c=rbinom(n=100, size=5, prob=.7)
   d <- mean(c)
   e <- sd(c)
   f <- e/d
   print(c(c,d,e,f))
   }
```

```
###Q2
##ai
   pt(1.5,df=10)-pt(-1,df=10)

##aii
   k=qt(.025, df=12)
   k

##aiii
   choose(15,9)
   log(25,base=10);log10(25)
   factorial(28)

##b
   A <- rexp(12, rate=3)
   A
   for(i in A) {
     if(i<=3)  print(1) else  print(2)
   }
 #or
   B <- vector(mode = "numeric")
   j <-0
   for(i in A) {
     j <- j+1
     if(i<=3)  B[j]=1 else  B[j]=2
   }
   B

###Q3
##a
   k=qf(1-.04, df1=12, df2=10)
   k

##c
   f <- function(x) { (x^5)*((1-x)^4) }
   i <- integrate(f,lower=0,upper=1)$value
   i

 #or
   f <- function(x) { dbeta(x, shape1=6, shape2=5) * (beta(6,5)) }
   i <- integrate(f,lower=0,upper=1)$value
   i


###Q4
   a <- c(1,6,3,-1,5,2,7,4)
   A <- matrix(a,nrow = 2, ncol =4,byrow=T)
   A
   b <- c(1,9,8,7,4,2,5,1,5,1,1,9)
   B <- matrix(b,nrow = 4, ncol =3,byrow=T)
   B
   c <- c(3,4,2,7,4,9,0,6,3,8,3,2,3,4,6,2)
   C <- matrix(c,nrow = 4, ncol =4,byrow=T)
   C
##a
   A%*%B

##b
   det(C)

##c
   solve(C)
```

```
##d
   A <- c(48,39,42,52,40,48,52,52,54,48)
   A
   B <- c(50,48,42,40,43,48,50,46,38,38)
   B
 #or
 قبل استدعاء الملفاضع المؤشر في صفحة النتائج ثم اختار ملف ثم اختار الخيار رقم ٩ ثم احدد المجلد الذي فيه ملف البيانات#
   data <- read.csv("data.csv",header=T, sep=";")
   data
   A <- data$A
   A
   B <- data$B
   B
##di
   t.test(A,B,alternative = "greater", paired = FALSE, var.equal = T,conf.level = 0.95)

##dii
   t.test(B,A,alternative = "two.sided", paired = FALSE, var.equal = T,conf.level = 0.90)

###Q5
   1-pnorm(2.22, mean = 2, sd = sqrt(5))
```

+ See Appendix -3-