السؤال الأول: في المثال التالي تم سؤال عشرة موظفين من النساء والرجال في شركة ما عن المؤهل العلمي وعدد سنوات الخبرة والراتب الحالي، نصنف البيانات باستخدام المتغيرات التالية ثم نقوم بإدخالها

الجنس:

Male: 1 female: 2

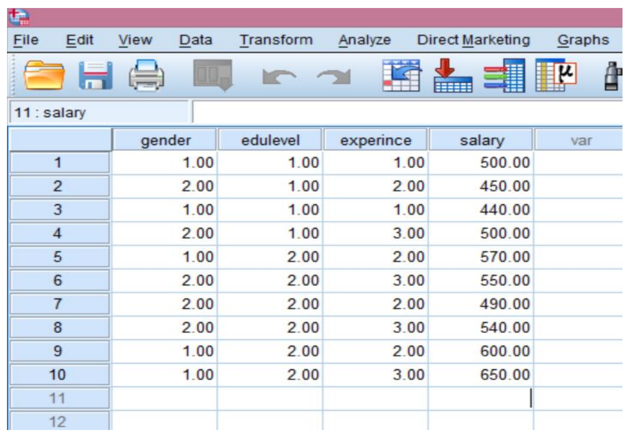
المؤهل الدراسي:

Bachelor’s degree: 1 master’s degree: 2

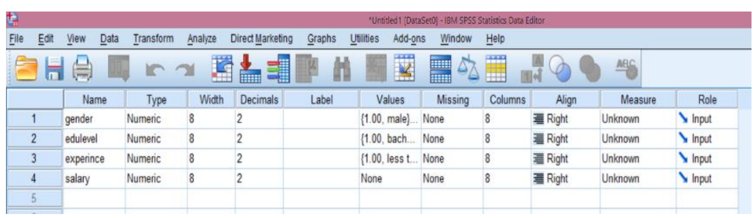
الخبرة:

Less than 5: 1 between 5 and 10:2 greater than 5:3

الراتب:

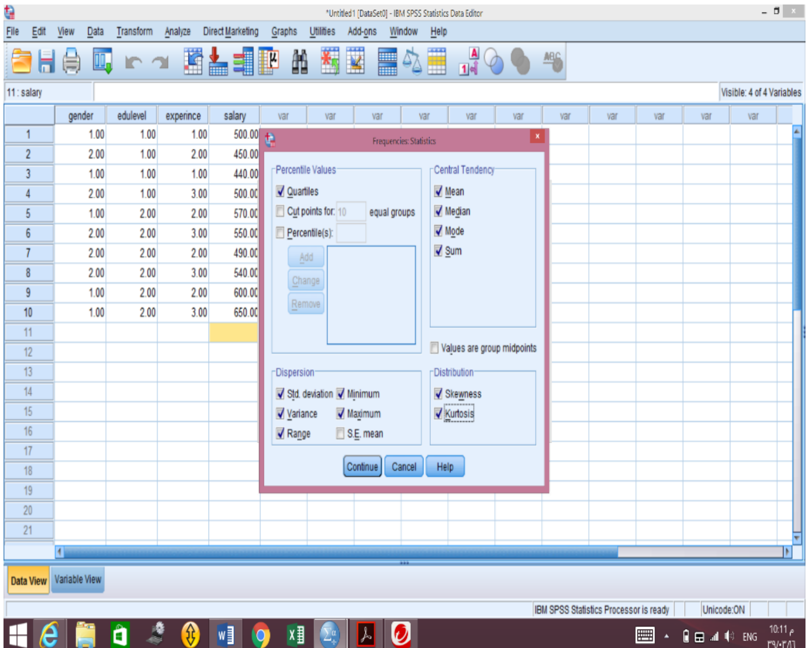


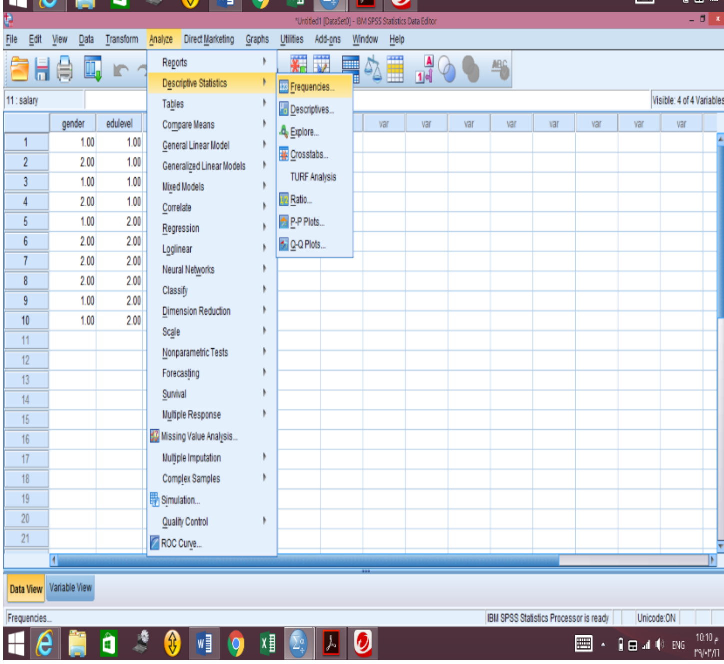
	gender	edulevel	experince	salary	var
1	1.00	1.00	1.00	500.00	
2	2.00	1.00	2.00	450.00	
3	1.00	1.00	1.00	440.00	
4	2.00	1.00	3.00	500.00	
5	1.00	2.00	2.00	570.00	
6	2.00	2.00	3.00	550.00	
7	2.00	2.00	2.00	490.00	
8	2.00	2.00	3.00	540.00	
9	1.00	2.00	2.00	600.00	
10	1.00	2.00	3.00	650.00	
11					
12					

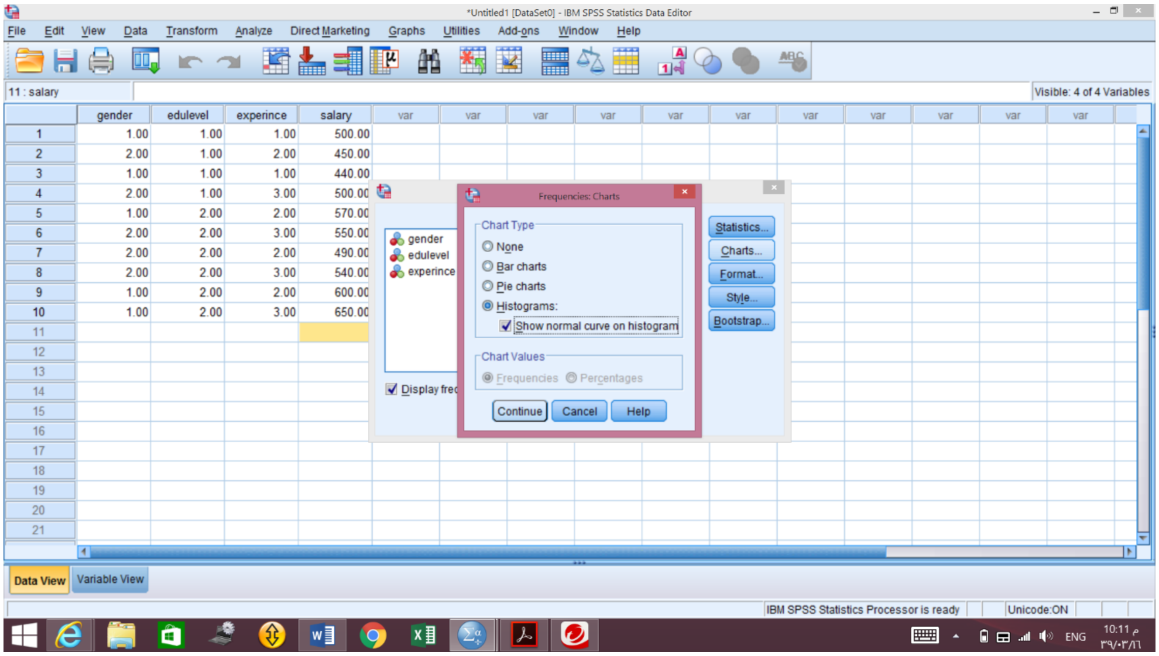


	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	gender	Numeric	8	2		{1.00, male}...	None	8	Right	Unknown	Input
2	edulevel	Numeric	8	2		{1.00, bach...	None	8	Right	Unknown	Input
3	experince	Numeric	8	2		{1.00, less t...	None	8	Right	Unknown	Input
4	salary	Numeric	8	2		None	None	8	Right	Unknown	Input
5											

استخدام الخيارFrequencies لحساب المقاييس الإحصائية والجداول التكرارية





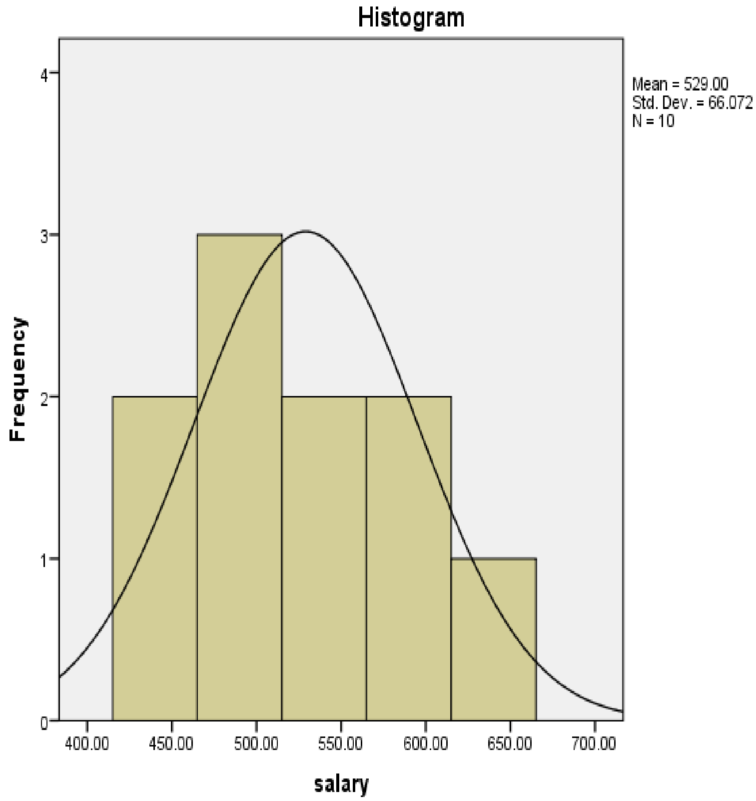


Frequencies

Statistics

salary

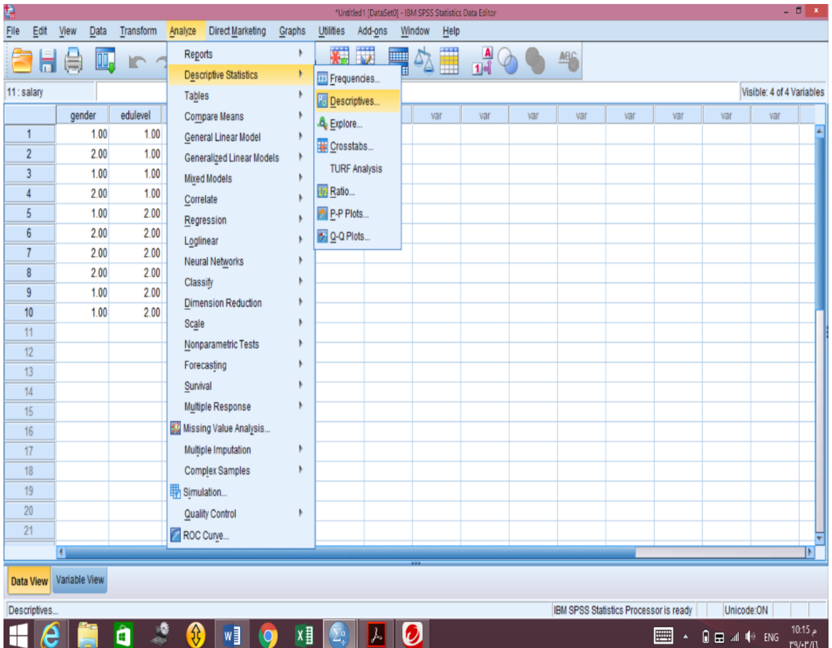
N	Valid	10
	Missing	0
Mean		529.0000
Median		520.0000
Mode		500.00
Std. Deviation		66.07235
Variance		4365.556
Skewness		.435
Std. Error of Skewness		.687
Kurtosis		-.351-
Std. Error of Kurtosis		1.334
Range		210.00
Minimum		440.00
Maximum		650.00
Sum		5290.00
Percentiles	25	480.0000
	50	520.0000
	75	577.5000

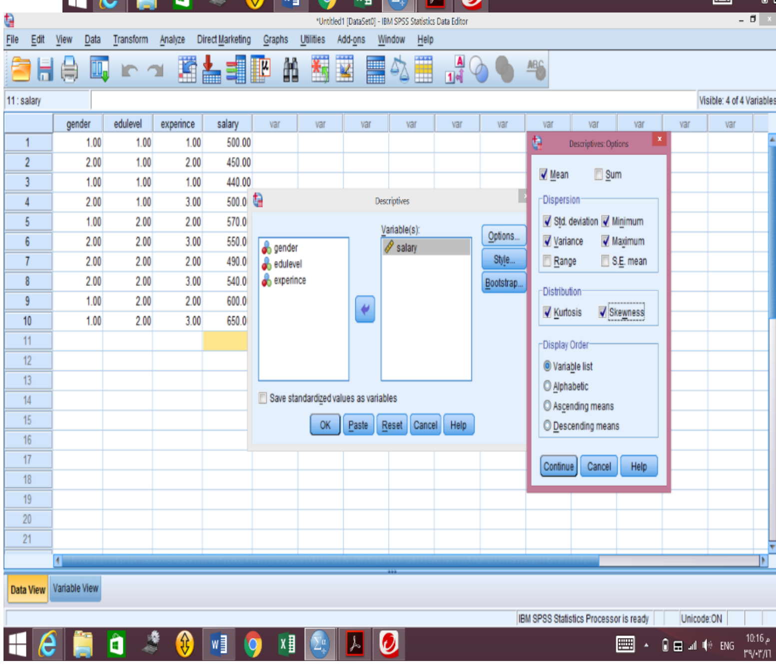


salary

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	440.00	1	10.0	10.0	10.0
	450.00	1	10.0	10.0	20.0
	490.00	1	10.0	10.0	30.0
	500.00	2	20.0	20.0	50.0
	540.00	1	10.0	10.0	60.0
	550.00	1	10.0	10.0	70.0
	570.00	1	10.0	10.0	80.0
	600.00	1	10.0	10.0	90.0
	650.00	1	10.0	10.0	100.0
	Total	10	100.0	100.0	

استخدام الخيار descriptive لحساب المقاييس الإحصائية





تظهر لنا النتائج التالية

Descriptives

Notes

	Minimum	Maximum	Mean	Std. Deviation	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
salary	440.00	650.00	529.0000	66.07235	4365.556	.435	.687	-.351-	1.334
Valid N (listwise)									

Q2)

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.

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

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

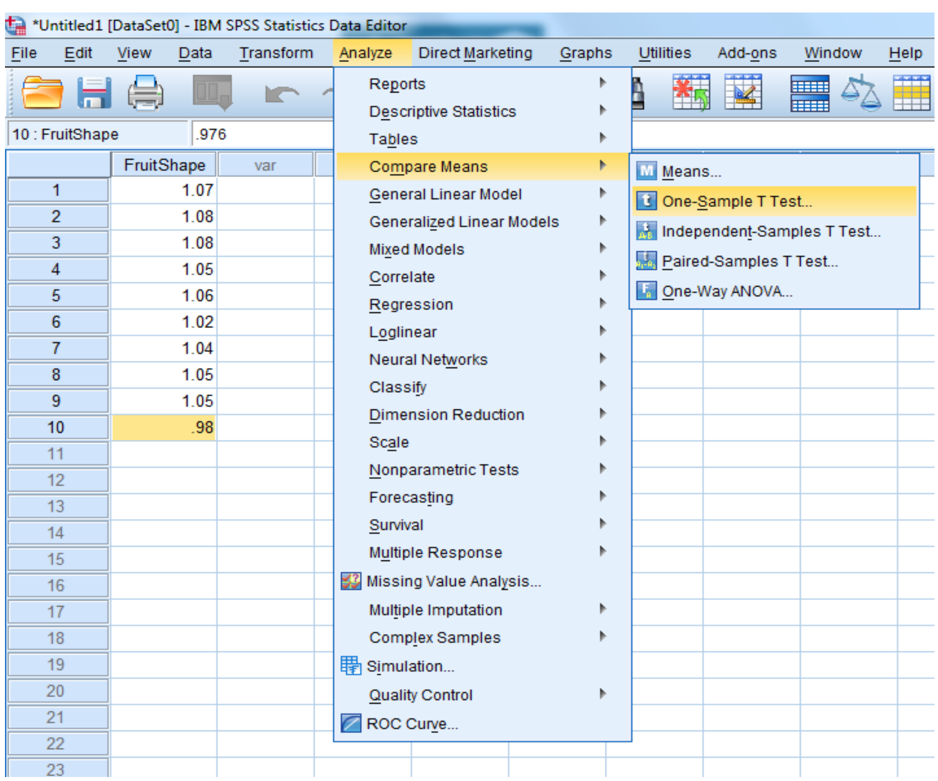
$$Vs$$

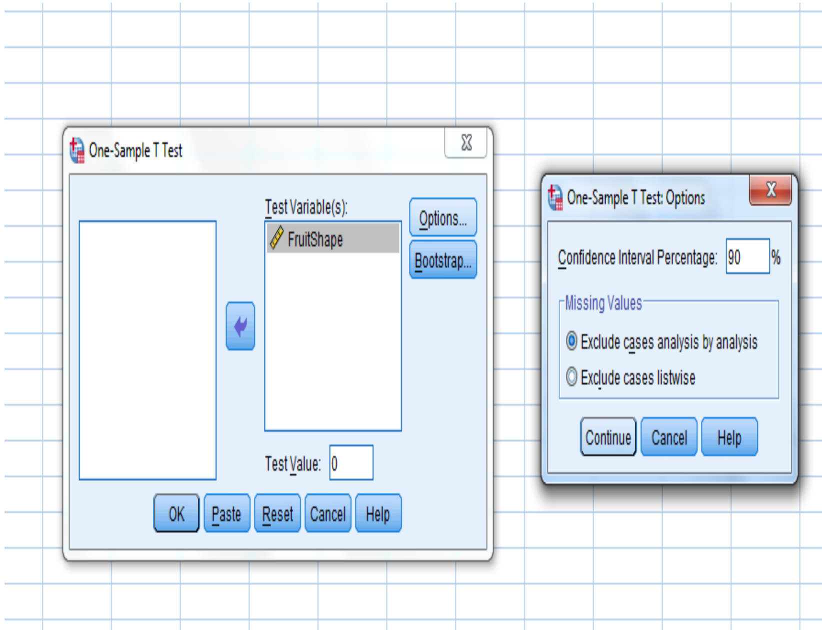
$$H_1: the\ population\ \ does\ not\ follow \text{ a normal distribution}$$

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

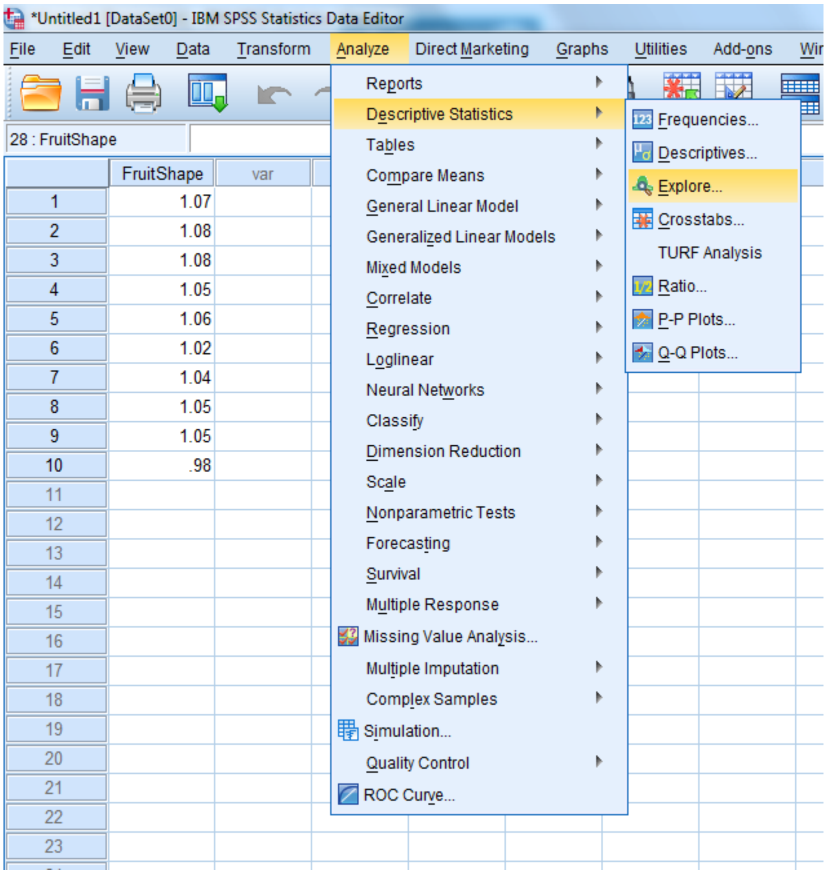
	N	Mean	Std. Deviation	Std. Error Mean
FruitShape	10	1.0465	.03103	.00981

One-Sample Test

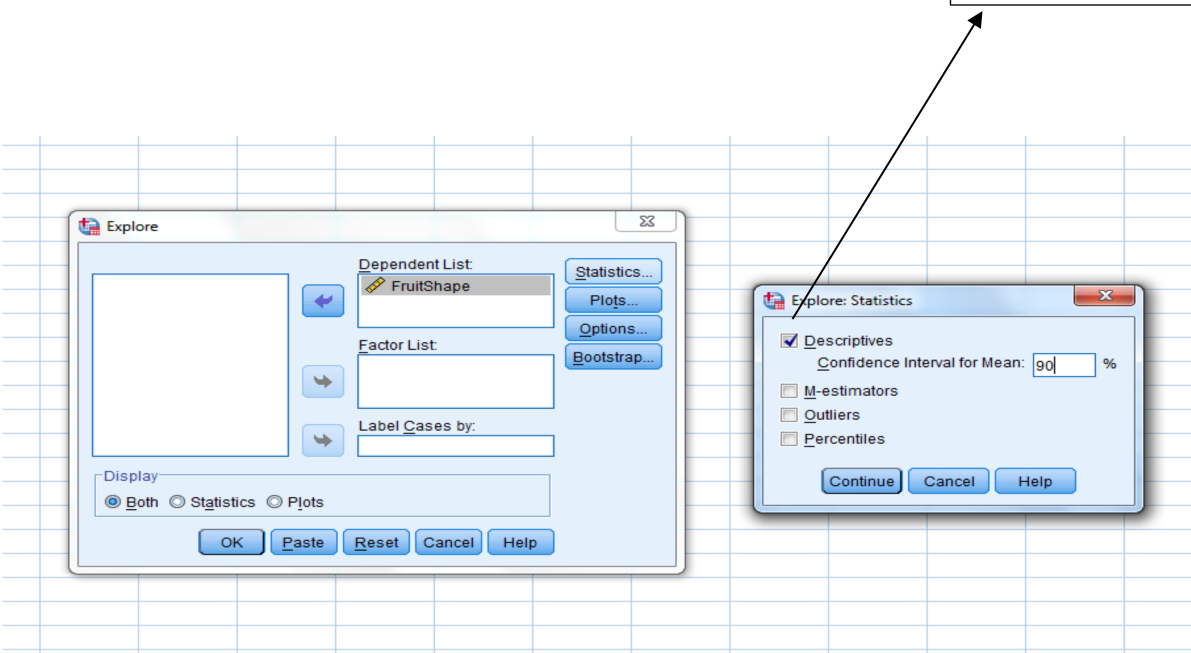
	Test Value = 0					
	t	df	Sig. (2-tailed)	Mean Difference	90% Confidence Interval of the Difference	
					Lower	Upper
FruitShape	106.632	9	.000	1.04650	1.0285	1.0645

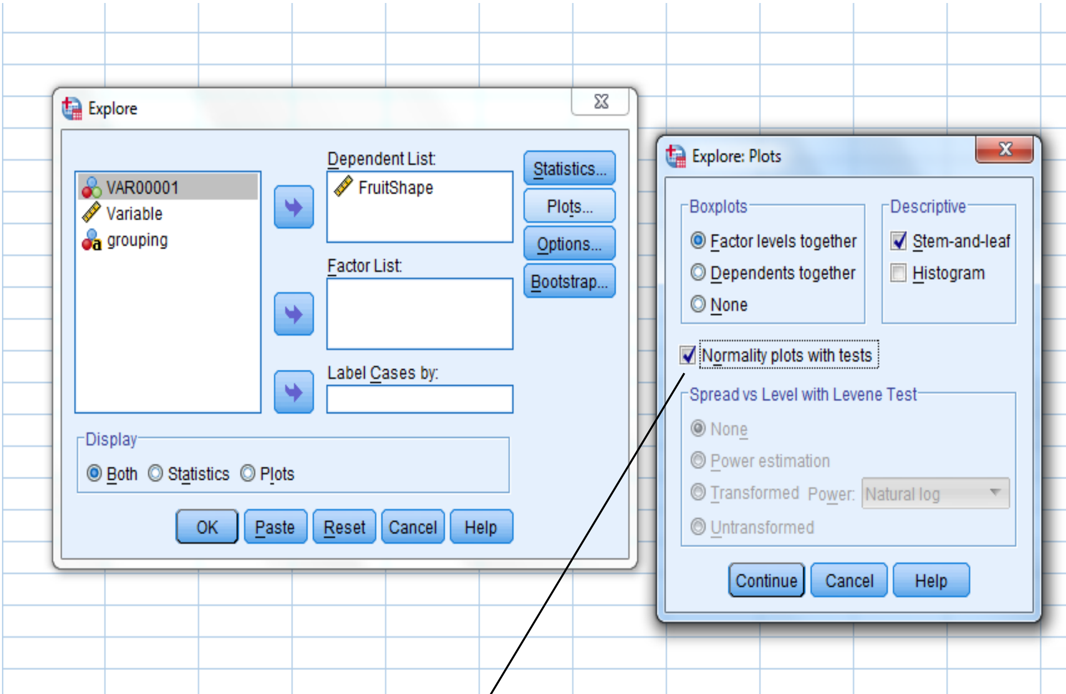
C.I for the mean

2) The second method:



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





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

Q3)

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

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.

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

$$Vs$$

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

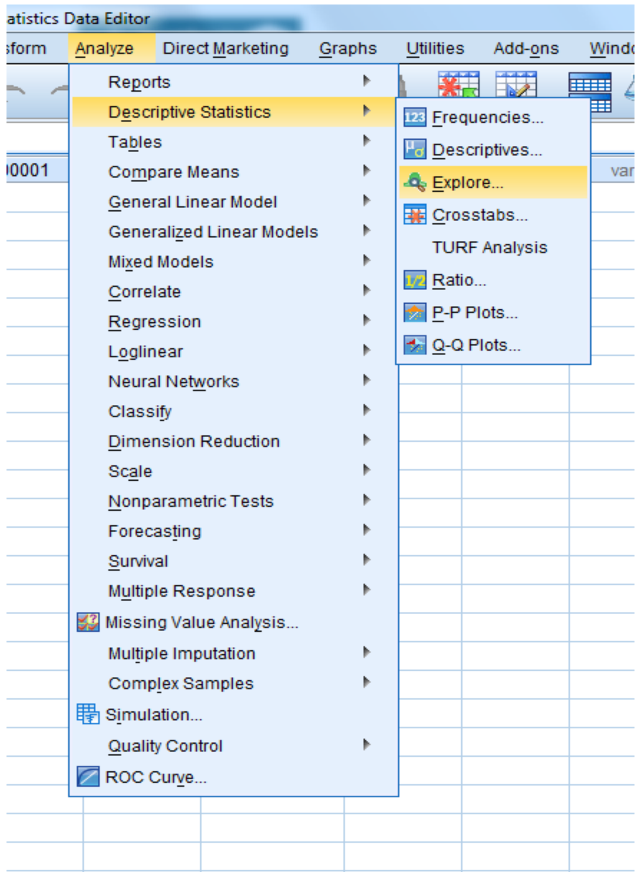
However, we find the question he said that the populations follows a normal distribution, so is not necessary to make this test.

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

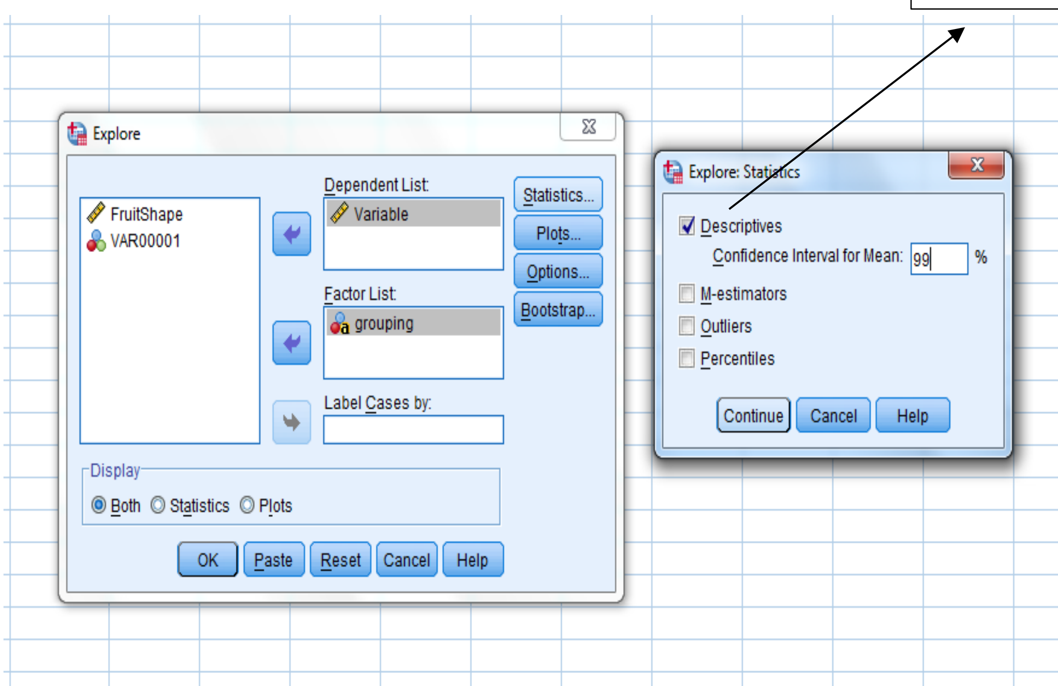
ata Editor

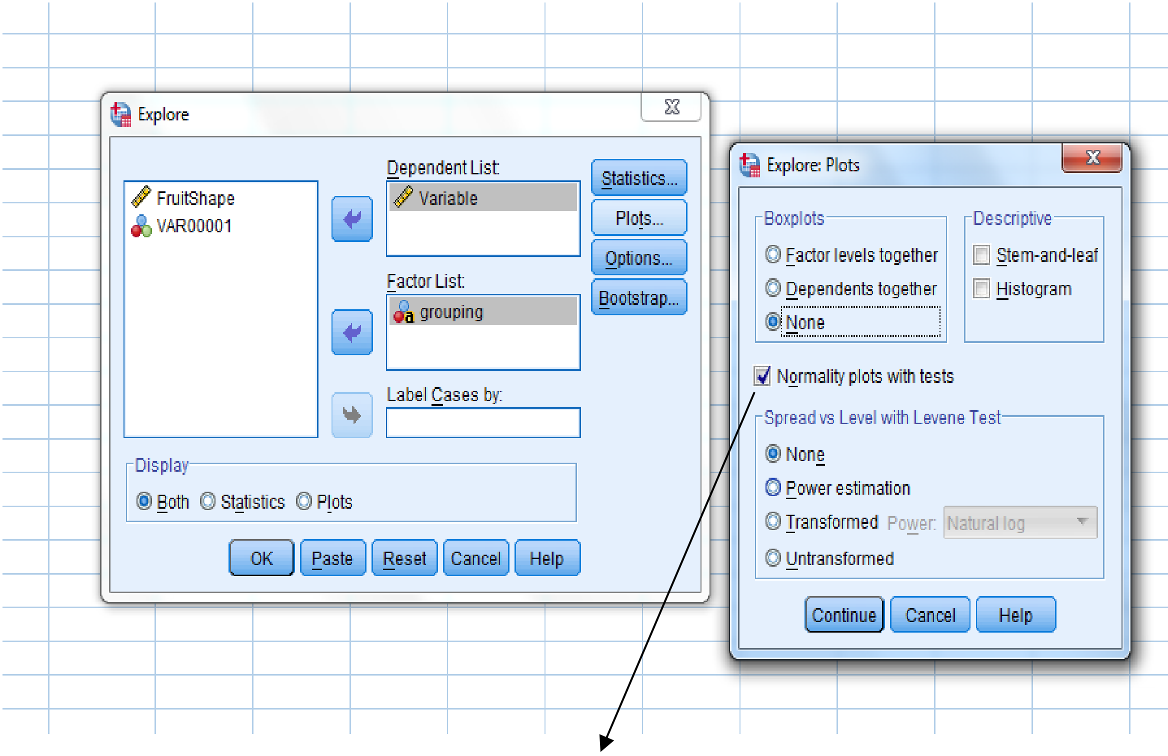
Analyze Direct Marketing Graphs U

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

C.I for the mean for the skim

C.I for the mean for the whole

Double-click to activate

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.

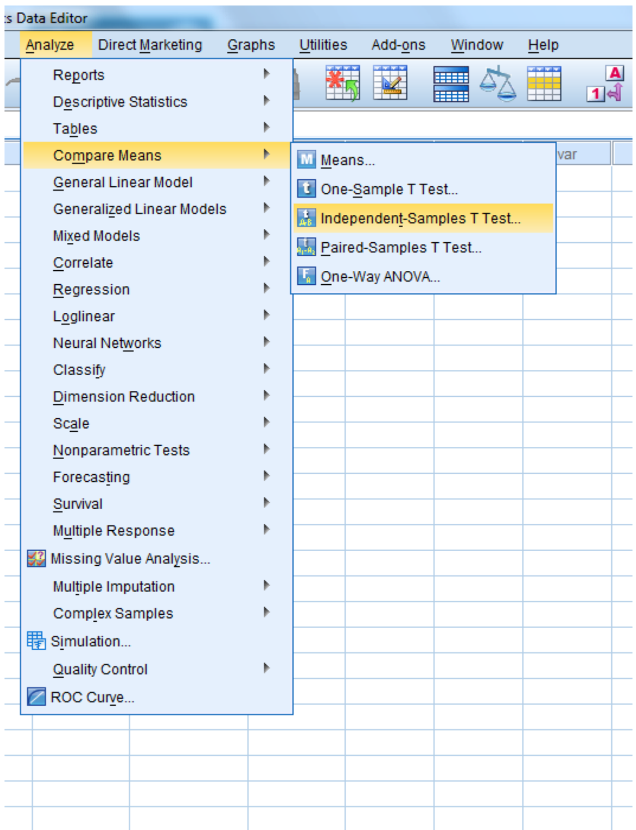
So, we except H_0: *the two populations follow* a normal distribution

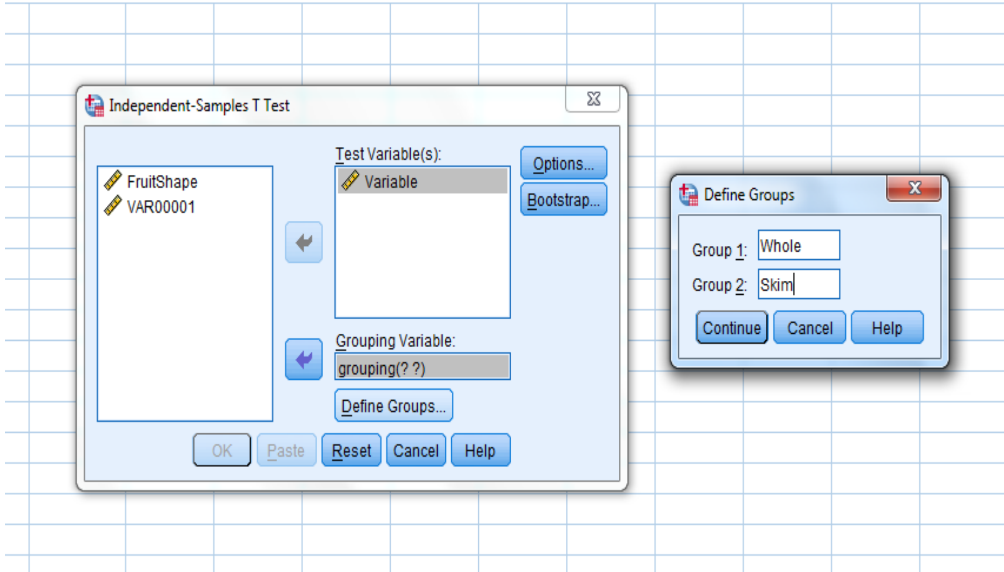
Now, the goal of the question:

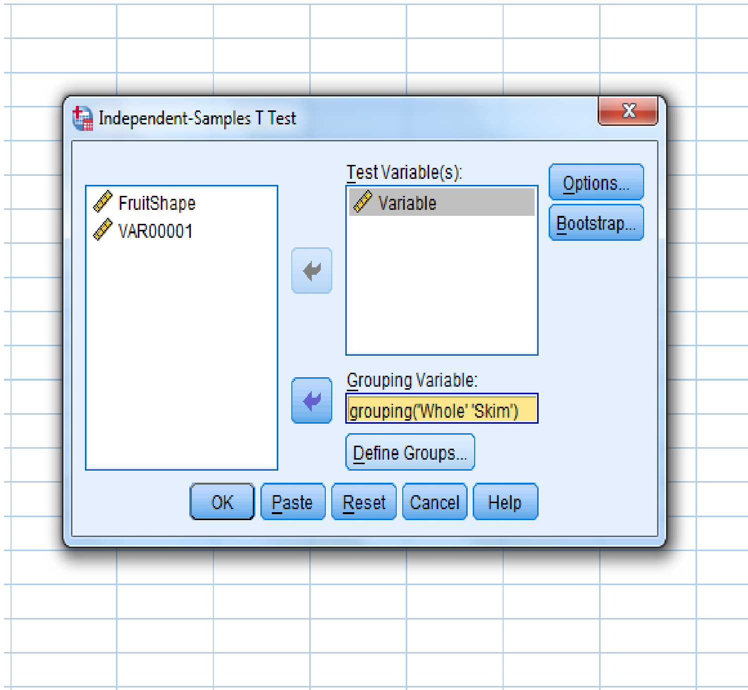
a) $H_0: \mu_{whole} - \mu_{skim} = 0 \quad Vs \quad H_1: \mu_{whole} - \mu_{skim} > 0 \quad at\ \alpha = .01$

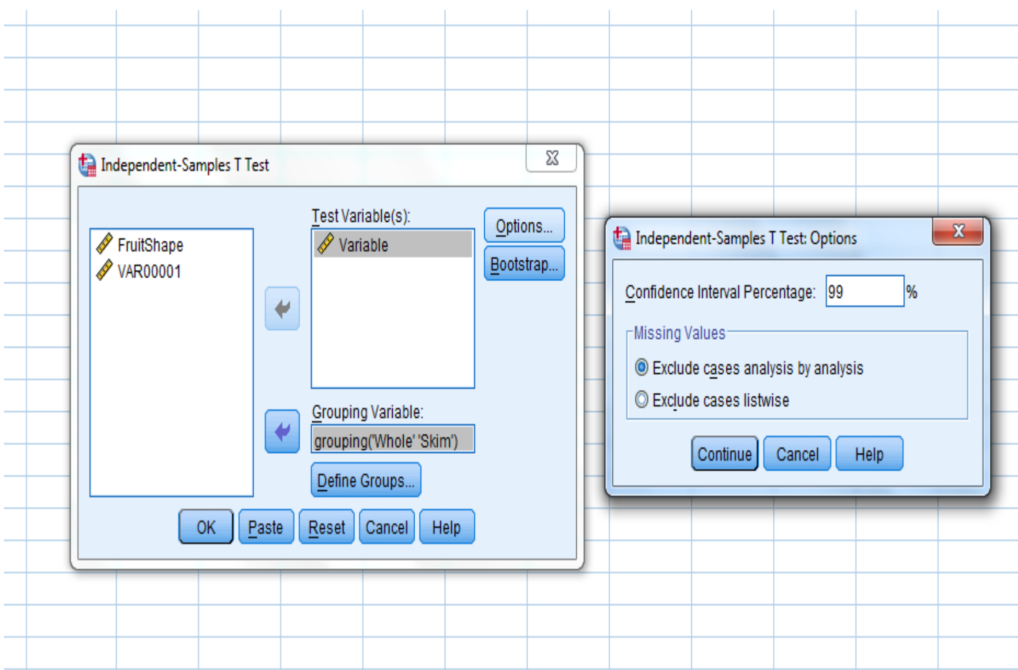
and

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




Independent-Samples T Test
FruitShape
VAR00001
Test Variable(s):
Variable
Options...
Bootstrap...
Grouping Variable:
grouping(? ?)
Define Groups...
OK
Paste
Reset
Cancel
Help
Define Groups
Group 1:
Whole
Group 2:
Skim
Continue
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
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 $\mu_{whole} - \mu_{skim}$

Q4) 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).

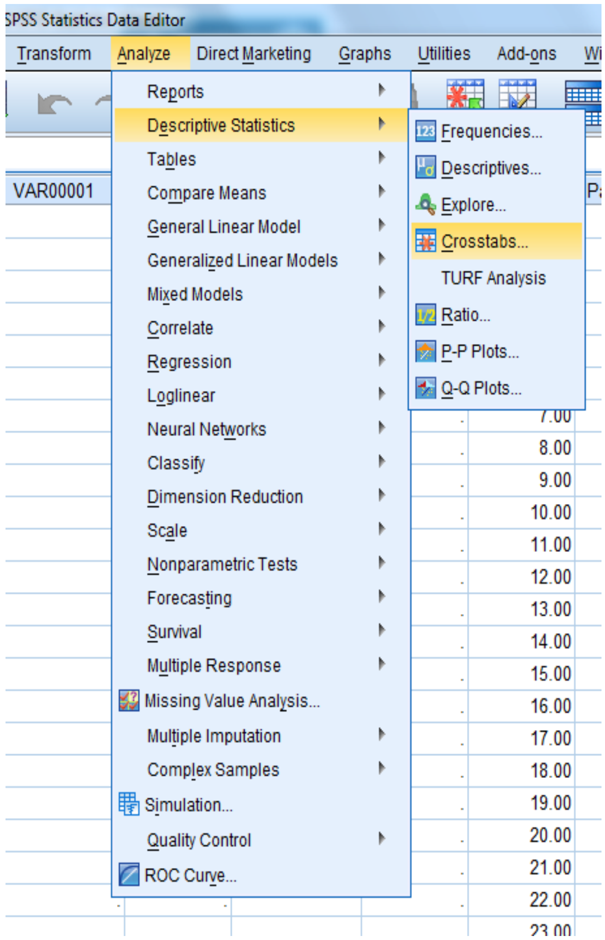
	Pass	No Pass	Row Totals
Males	12	3	15
Females	13	2	15
Column Totals	25	5	30

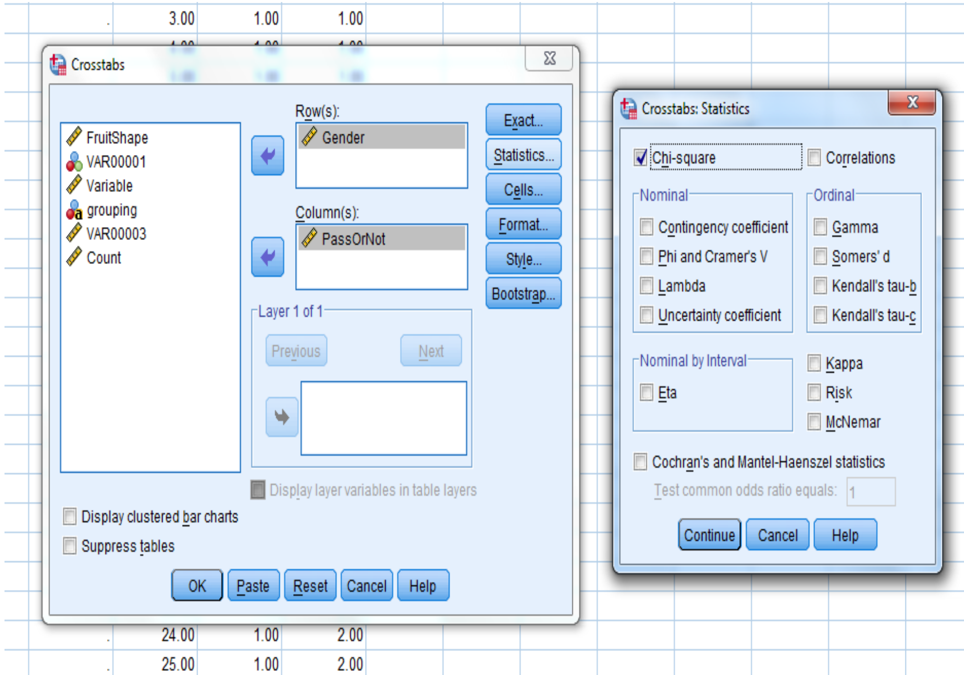
H_0: the gender of the students is indep. of a Pass or No Pass test grade

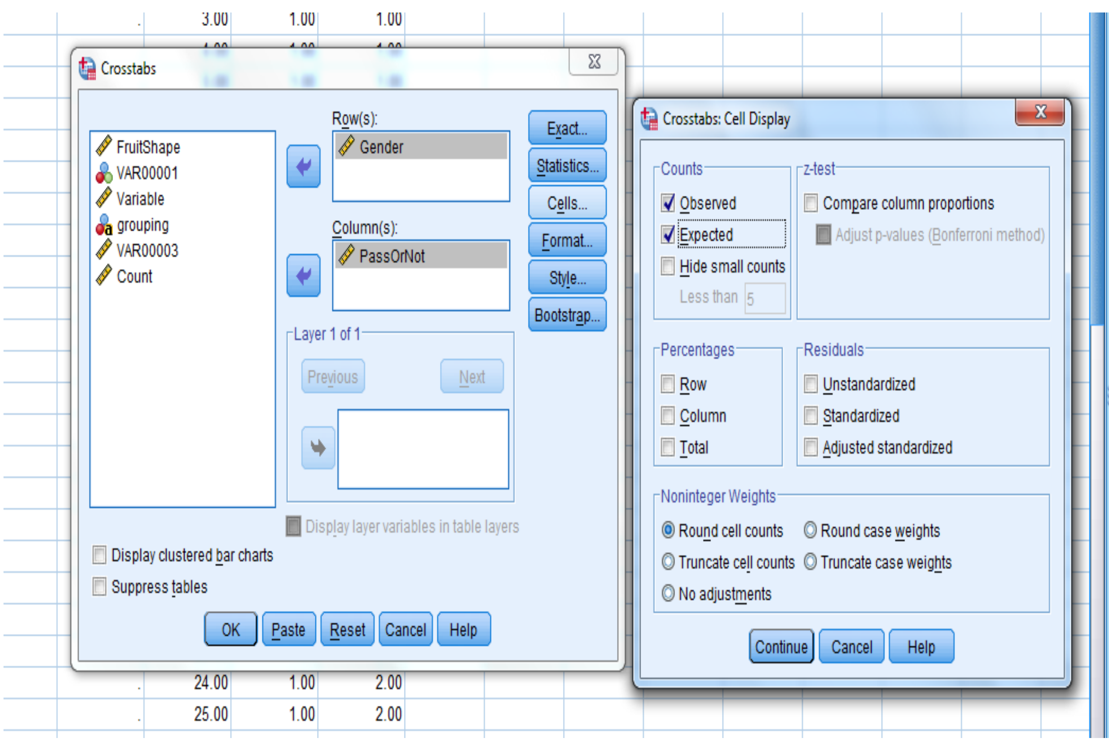
Vs

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
OK
Paste
Reset
Cancel
Help
Exact...
Statistics...
Cells...
Format...
Style...
Bootstrap...
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:
1
Continue
Cancel
Help
3.00
1.00
1.00
24.00
1.00
2.00
25.00
1.00
2.00


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
OK
Paste
Reset
Cancel
Help
Exact...
Statistics...
Cells...
Format...
Style...
Bootstrap...
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
3.00
1.00
1.00
24.00
1.00
2.00
25.00
1.00
2.00

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

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.

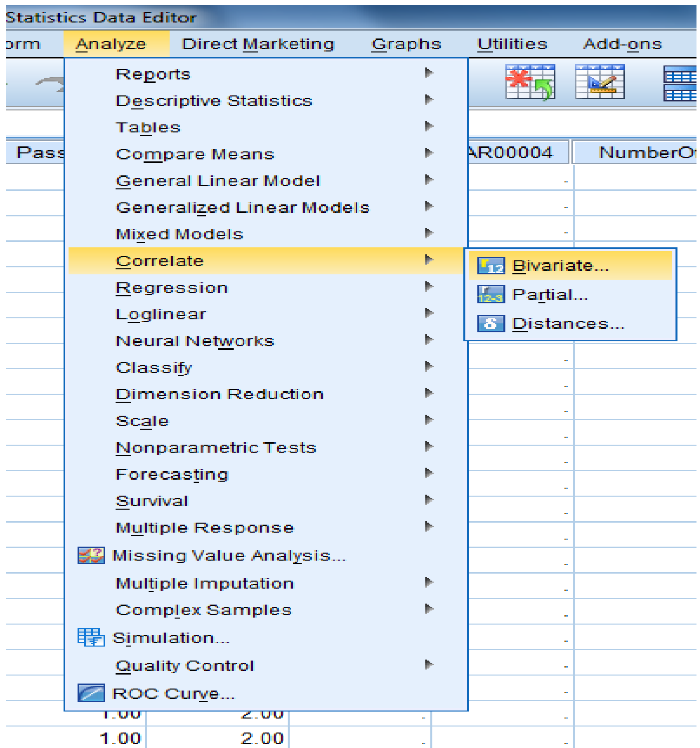
x	6	6	6	4	2	5	4	5	1	2
y	125	115	130	160	219	150	190	163	260	260

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.

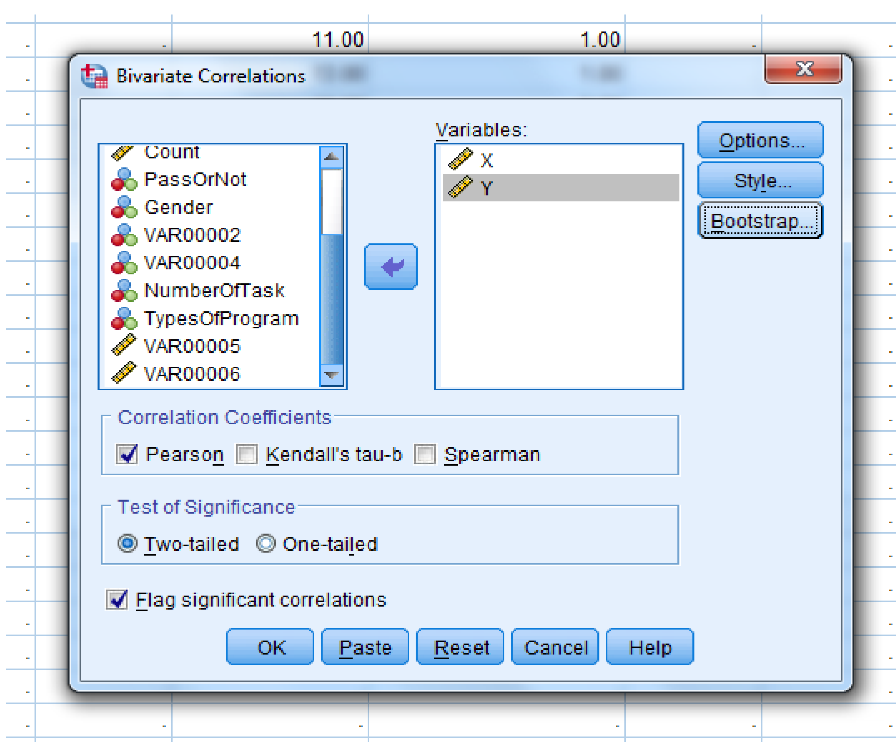
Enter the age values into one variable and the corresponding sales price values into another variable (see figure, below).

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.

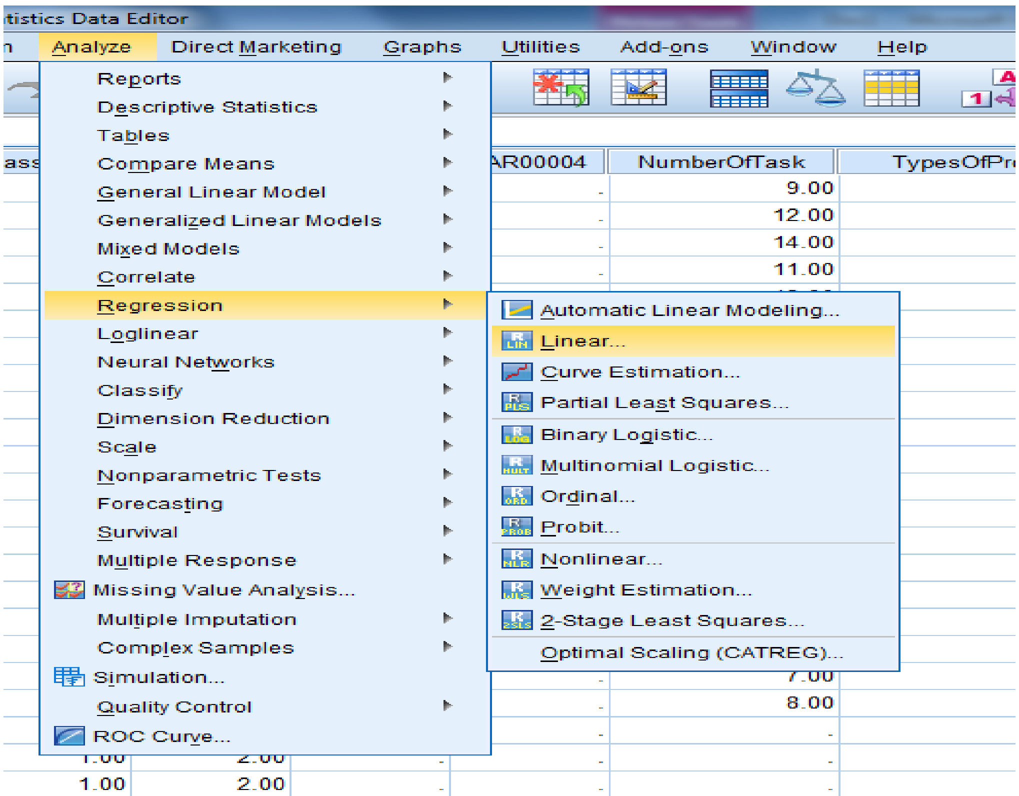
b, c and d)

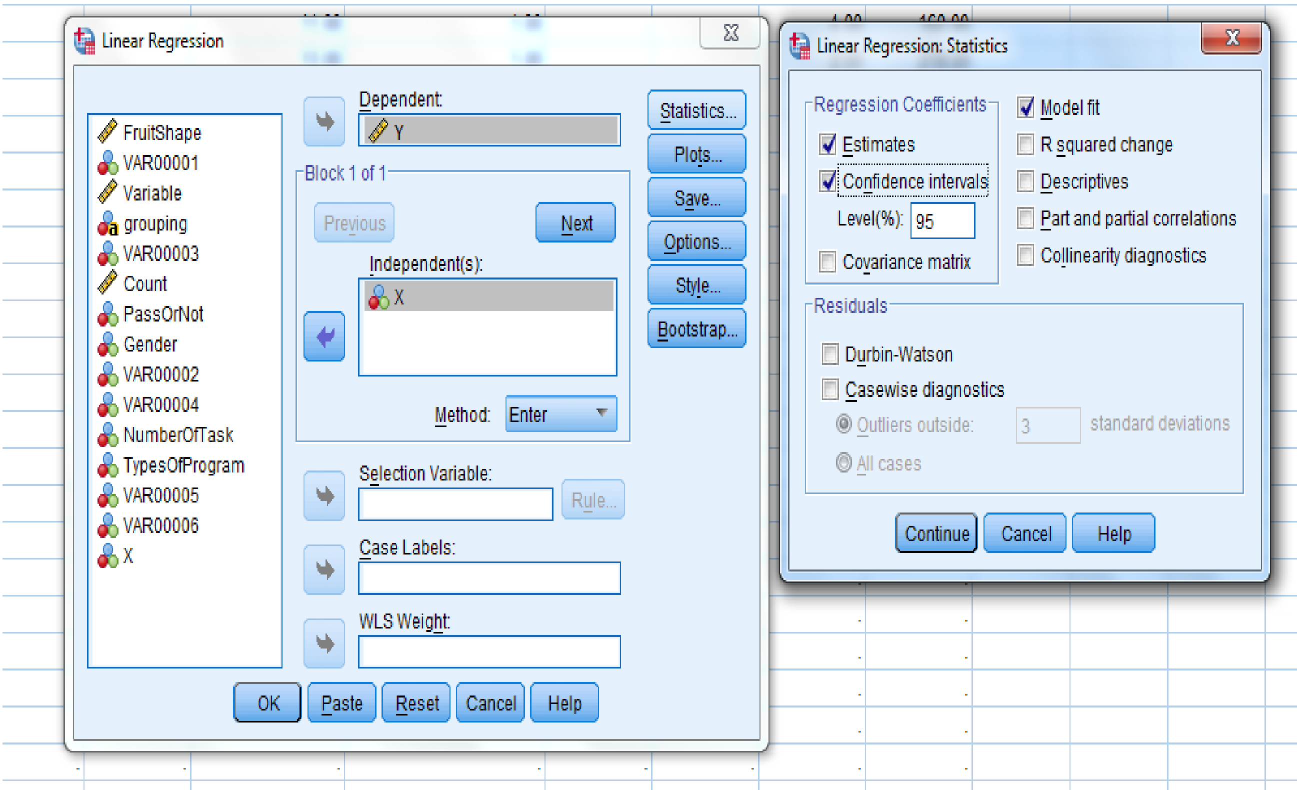
Since we eventually want to predict the price of 4-year-old Corvettes, enter the number "4" in the "x" variable column of the data window after the last row. Enter a "." for the corresponding "y" variable value (this lets SPSS know that we want a prediction for this value and not to include the value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

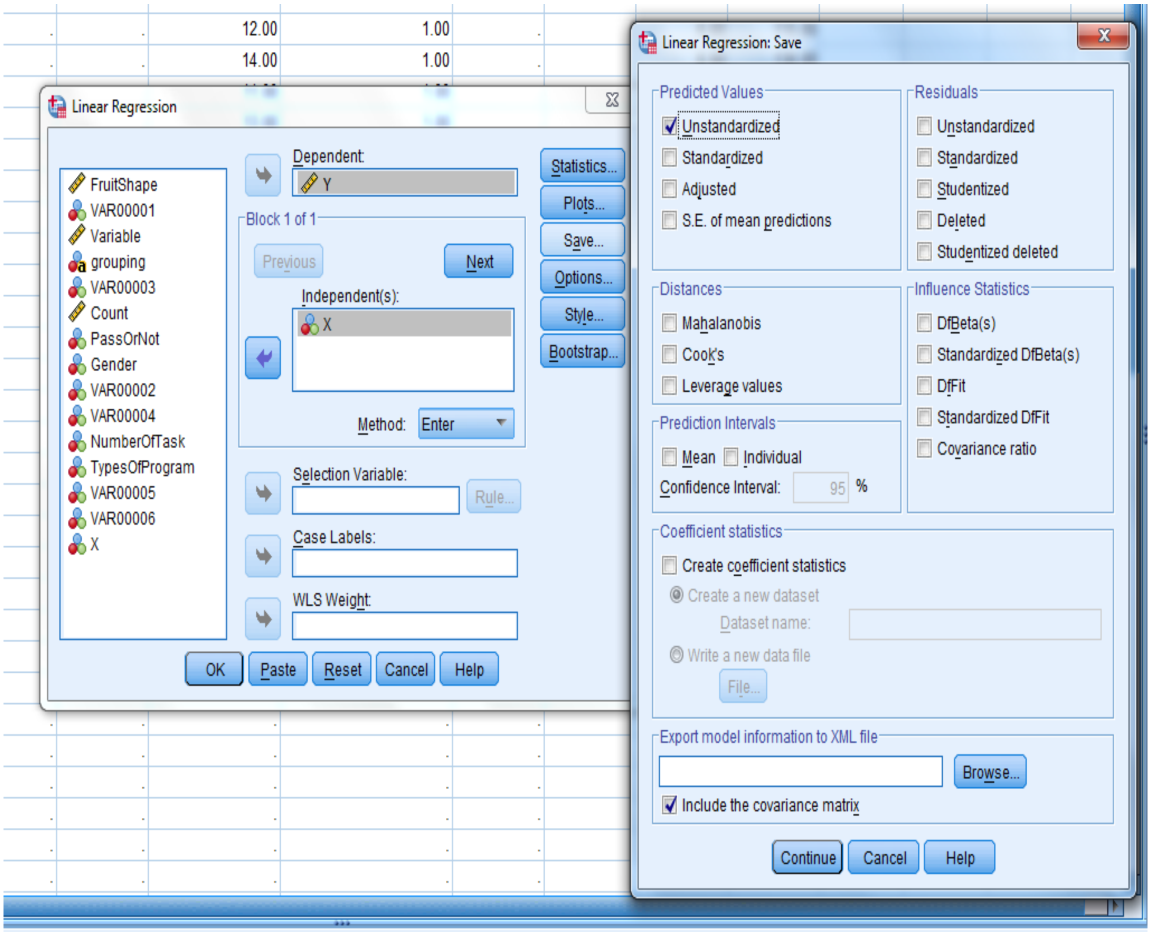
Select Analyze ◊ Regression ◊ Linear… (see figure).

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


Linear Regression
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...
FruitShape
VAR00001
Variable
grouping
VAR00003
Count
PassOrNot
Gender
VAR00002
VAR00004
NumberOfTask
TypesOfProgram
VAR00005
VAR00006
X
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.