[Academic Script]
[Econometric Problem Solving using SPSS Software]

Subject:	Business Economics
Course:	B. A. (Hons.), 5th Semester, Undergraduate
Paper No. & Title:	Paper – 502 Computational Techniques for Business Economics
Unit No. & Title:	Unit – 3 Econometric Problem Solving
Lecture No. & Title:	Lecture – 1 Econometric Problem Solving using SPSS Software

Academic Script

Econometric Problem Solving using SPSS software

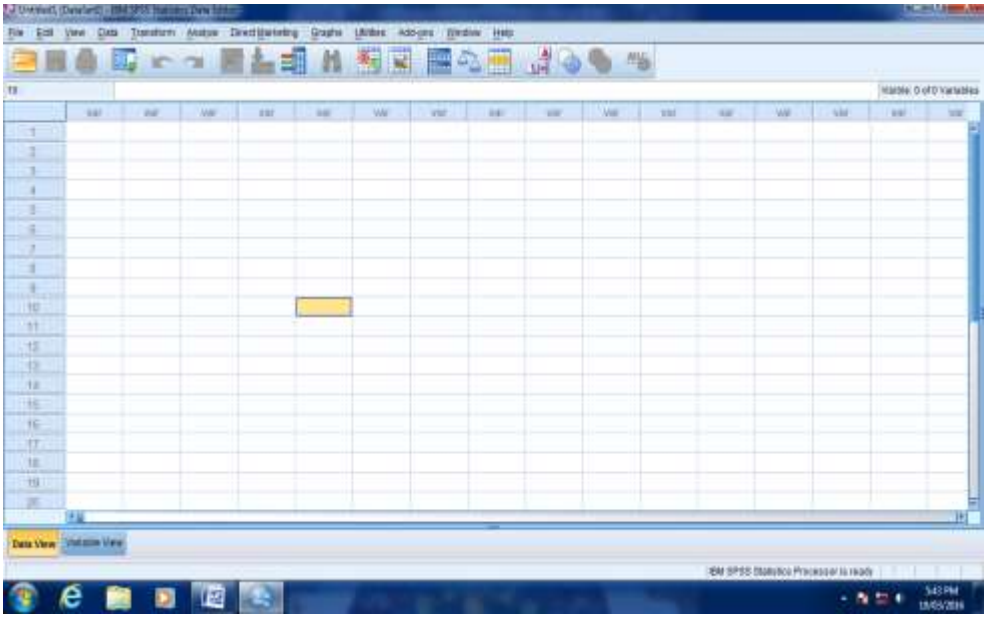
1. Introduction

SPSS is user friendly software for statistical data analysis. SPSS includes almost features of statistical methods starting from descriptive statistics, graphs, Correlation and regression, multiple regressions, testing of hypothesis, time series analysis, non parametric tests etc. All the econometric techniques studied in Paper 304 can be implemented through data with the help of SPSS.

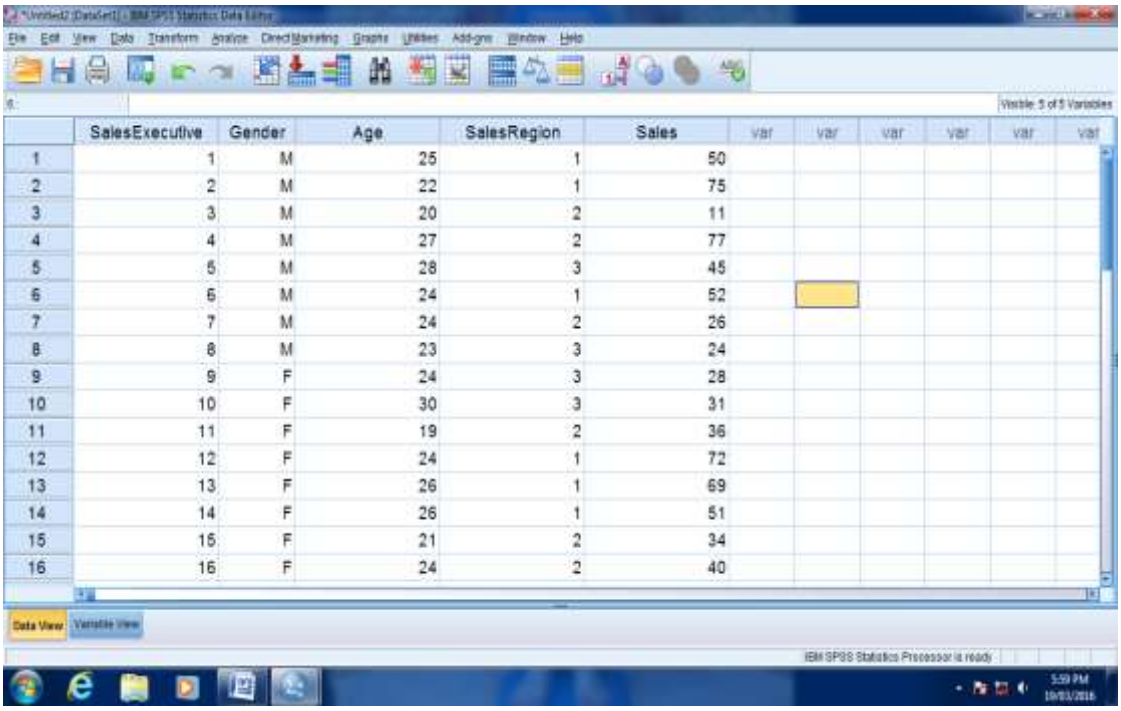
2. Data entry in SPSS

Let us see how to enter the data in SPSS. Either you open the SPSS software and directly enter the data or open the existing excel file in SPSS. To entre the data directly in SPSS adopt the following steps:

i) Click on the SPSS from the start menu, the following screen will display.



ii) Click on 'variable view' in the bottom left hand side corner and enter the list of variables stating their types whether numeric , string etc. and fill up all the information then click on 'Data View' and enter the data in columns for each variables.

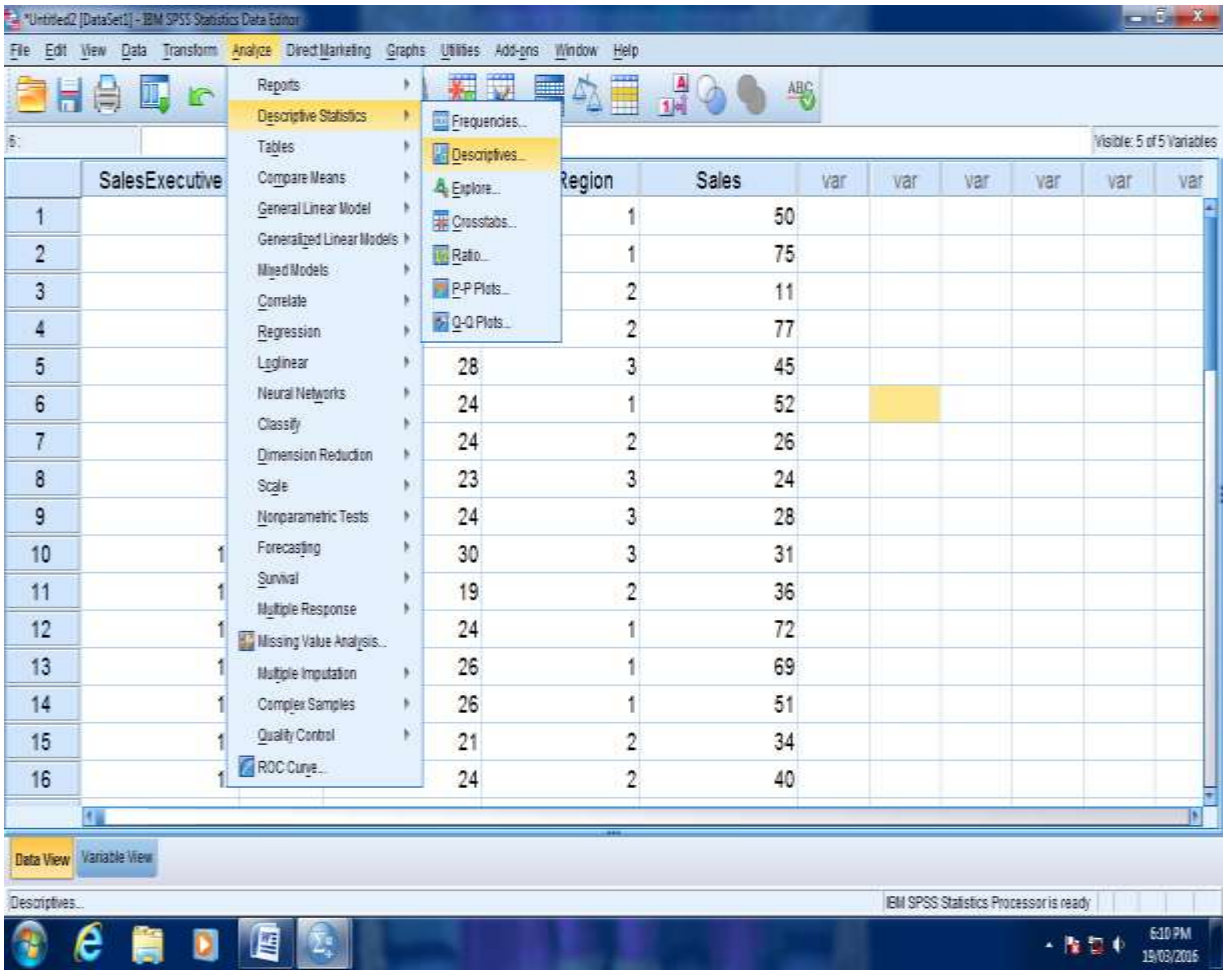


	SalesExecutive	Gender	Age	SalesRegion	Sales	var	var	var	var	var	var
1	1	M	25	1	50						
2	2	M	22	1	75						
3	3	M	20	2	11						
4	4	M	27	2	77						
5	5	M	28	3	45						
6	6	M	24	1	52						
7	7	M	24	2	26						
8	8	M	23	3	24						
9	9	F	24	3	28						
10	10	F	30	3	31						
11	11	F	19	2	36						
12	12	F	24	1	72						
13	13	F	26	1	69						
14	14	F	26	1	51						
15	15	F	21	2	34						
16	16	F	24	2	40						

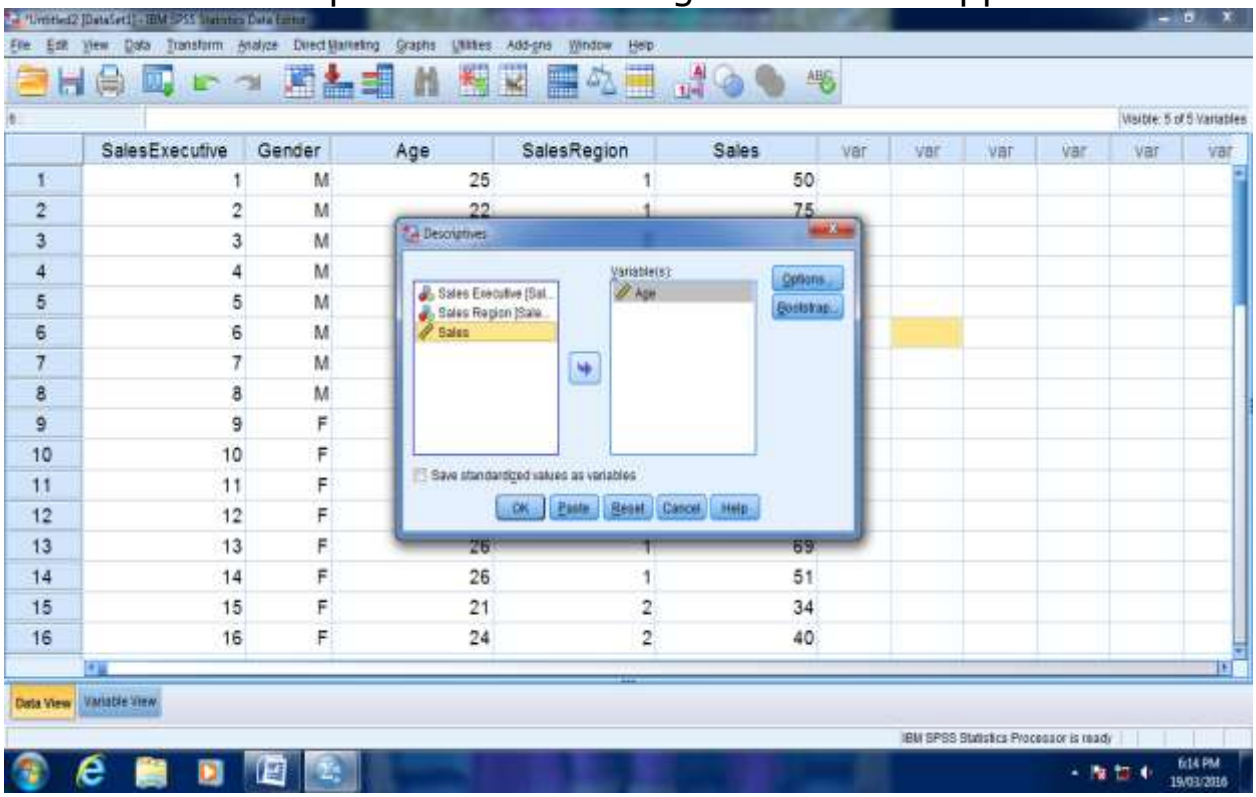
Example: there are 20 salesmen in a big company. the sales man sale the product of the company in the three regions(1,2,3).
In the data we have given gender, age and sales volume (in thousand rupees) for each sale man.

3. Descriptive Statistics

Let us find descriptive statistics for age and sales volume.



Click on ` Analyze' in the upper first raw then apply Descriptive Statistics Descriptive. The following window will appear.



Then select ` Options' and choose the necessary measures like mean, Standard deviation etc. We get the following Table of outcomes.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Age	20	19	30	24.60	2.836
Sales	20	11	77	43.55	19.975
Salary	20	10000.00	45000.00	24350.0000	9948.68412
Valid N (listwise)	20				

4. Testing of Hypothesis

In this section we will discuss how to carryout the tests like: t- test for mean, t- test for two means, F- test, and confidence interval for mean etc.

(a) Test for mean of the age of salesmen:

H: Average age of sales men = 25 years

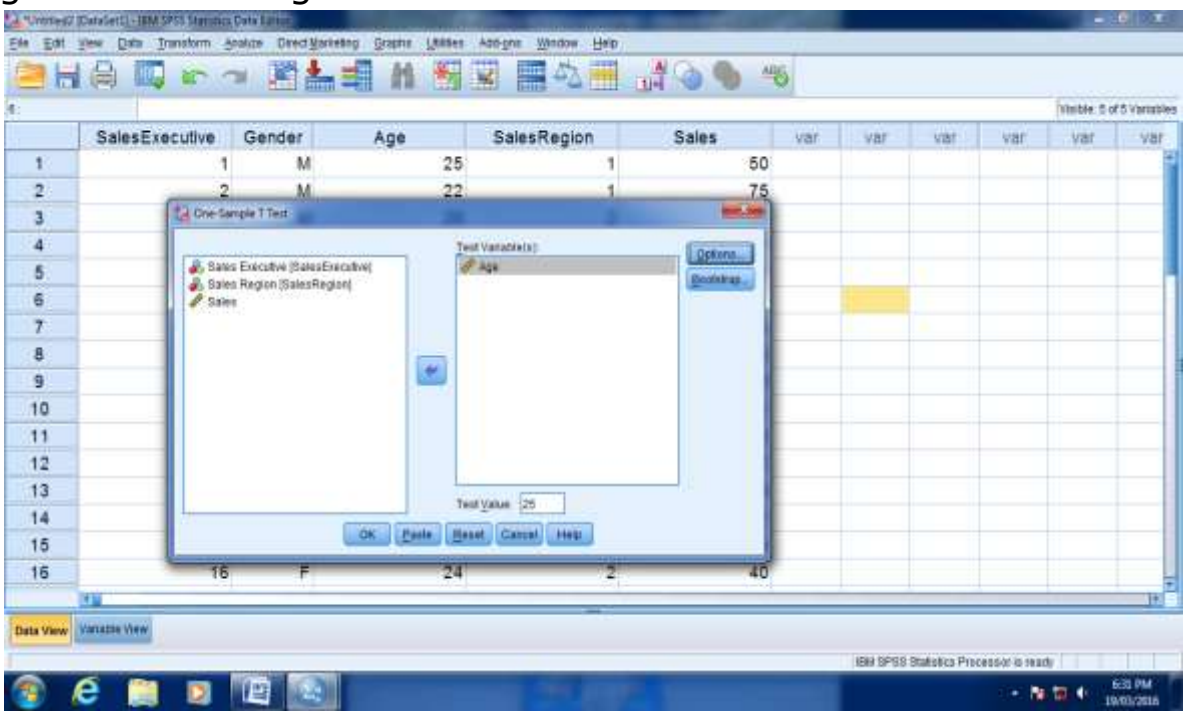
V/S

K: Average age of sales men ≠ 25 years

Perform the following steps:

Analyze ⟶ Compare Means ⟶ One Sample t-test

we get the following box.



Enter age in the test variable box and enter the test value 25, on clicking OK we get the output as

One-Sample Test

	Test Value = 25					
	T	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Age	-.631	19	.536	-.400	-1.73	.93

Here Sig(2-tailed) value (P-value) is 0.536 which is greater than 5% level of significance, hence null hypothesis can not be rejected.

Let us now perform a test for testing average age of male and female salesmen.

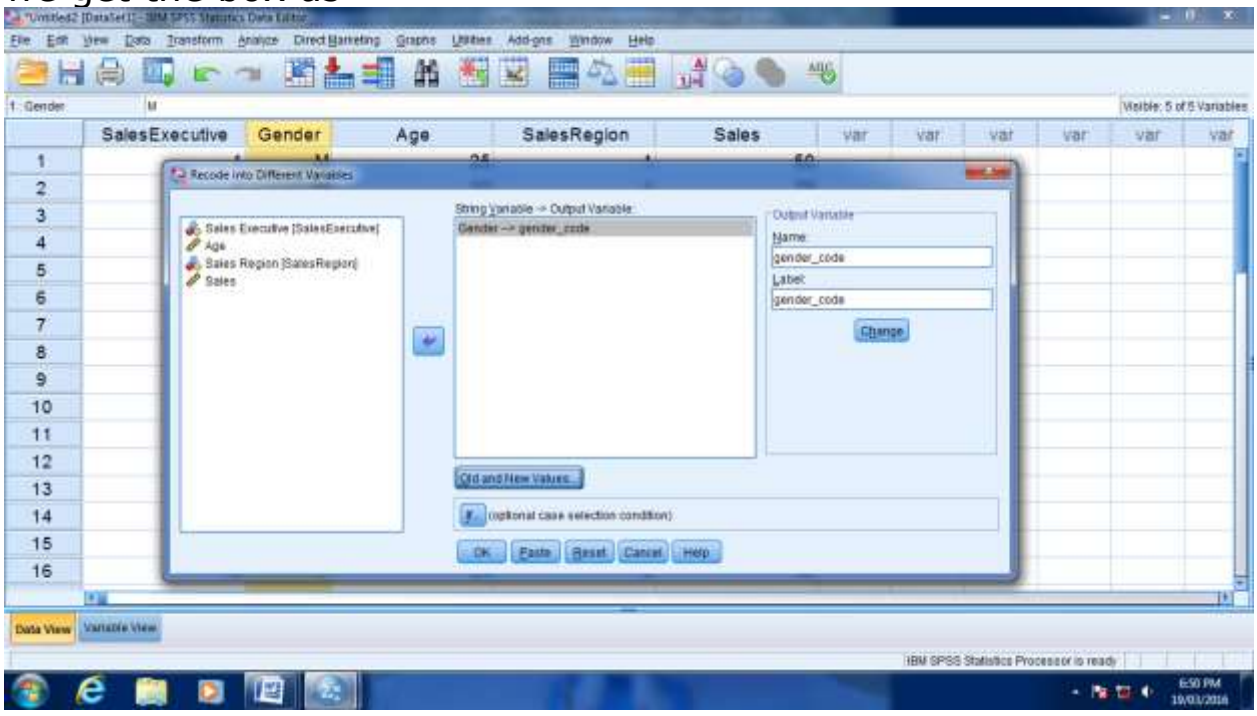
H: M1 = M2 versus K: M1 ≠ M2

Here M1 denotes the average age of Male and M2 for female.

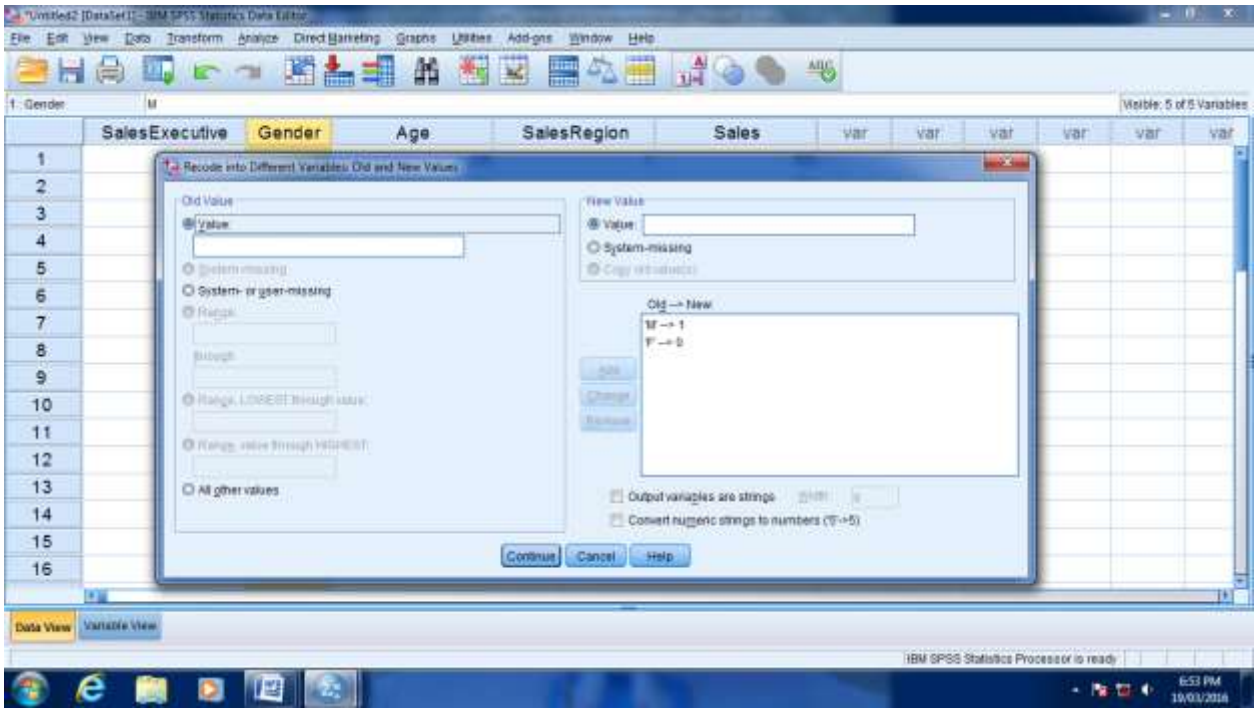
First of all we have to recode the data of gender in to code using '1' for ' Male' and '0 ' for 'female' by following steps.

Select: Transform Recode into Different Variables

we get the box as



Transform 'gender' in the RHS box and give a name 'gender_code' in output variable box then click on old and new values we get a new box as

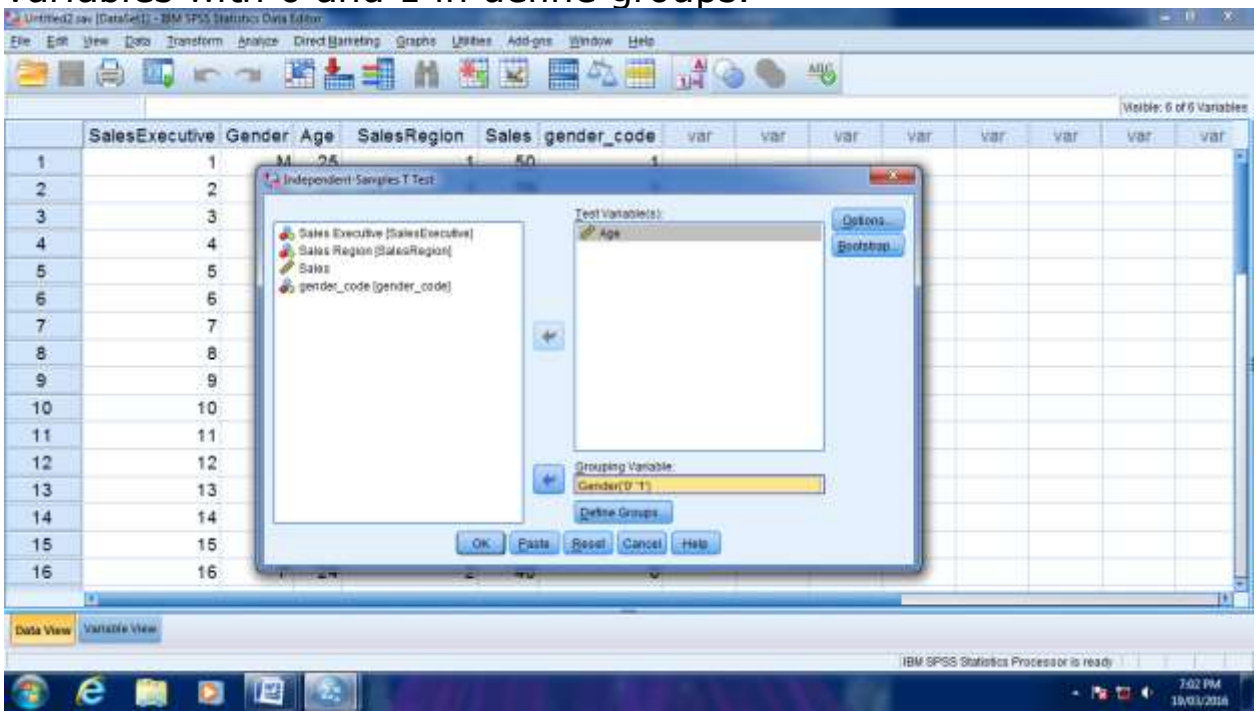


Now enter ' M' in old value box and 1 in new value box then click on " add" , similarly enter 'F' in old value box and 0 in new value box and the click continue and then OK, we get a column of gender_code in our data sheet as

Now to carryout t-test for two means do the following steps.

Analyze Compare means Independent samples t-test

Transfer 'age' in to RHS box, and enter gender_code in to grouping variables with 0 and 1 in define groups.



Clicking OK we get the following output.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means	
		F	Sig.	t	df
	Equal variances assumed	.170	.685	-.601	18

Independent Samples Test

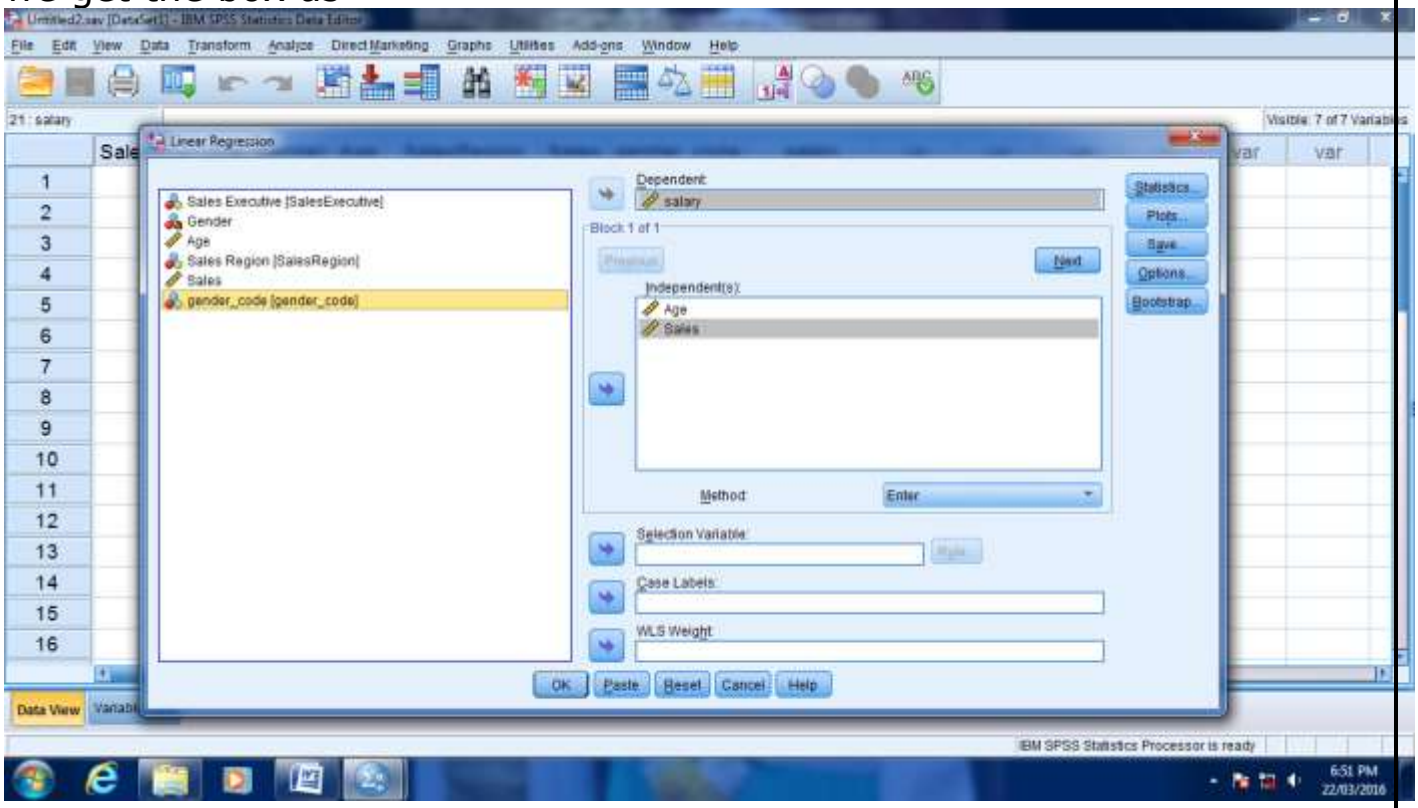
		t-test for Equality of Means			
		Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference
					Lower
Age	Equal variances assumed	.555	-.792	1.317	-3.558
	Equal variances not assumed	.542	-.792	1.272	-3.476
	Equal variances not assumed			-.623	16.825

Here first two columns denote the results to test H: $\sigma_1^2 = \sigma_2^2$ .i.e. to test whether the two groups have equal variances or not.
This test of hypothesis is necessary to test the hypothesis for equality of two means in case of two independent samples.
Here sig. value 0.685 is greater than 0.05, so we can not reject the null hypothesis of equal variances of age in male and female groups.

Now sign value for t-test is 0.555 which is also greater than 0.05, and hence the equality of mean ages of the male and female groups cannot be rejected at 5% level of significance.

5. Multiple regression equation.

Let us obtain regression equation to estimate salary of a salesman based on his age and sales volume.
Steps:
Analyze□Regression□Linear
we get the box as



Transfer salary to the first row on RHS box as ' dependent'
and age & sales volume in the second box on RHS as ' independent(s)'
Then select ' statistics from RHS upper most corner and click on ' estimates' and 'model fit' and click on continue and then click OK. We get the output as

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.986[a]	.973	.970	1731.02023

Here R square = 0.973 which is sufficiently large. Let us test it.
i.e. H: R = 0 versus K: R ≠ 0
From the following table we can get conclusion about this test.

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1829610672.638	2	914805336.319	305.298	.000[b]
	Residual	50939327.362	17	2996431.021		
	Total	1880550000.000	19			

As Significant value = 0.000, which is < 0.05, so we say that the null hypothesis H: R = 0 is rejected at 5% level of significance.
That is the multiple regression model is appropriate for the given data.
Now we prepare the multiple linear regression equation from the output. From the second column (denoted as B)of the following table we get the regression coefficients.

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2288.316	3497.126		.654	.522
	Age	27.974	141.028	.008	.198	.845
	Sales	490.781	20.022	.985	24.512	.000

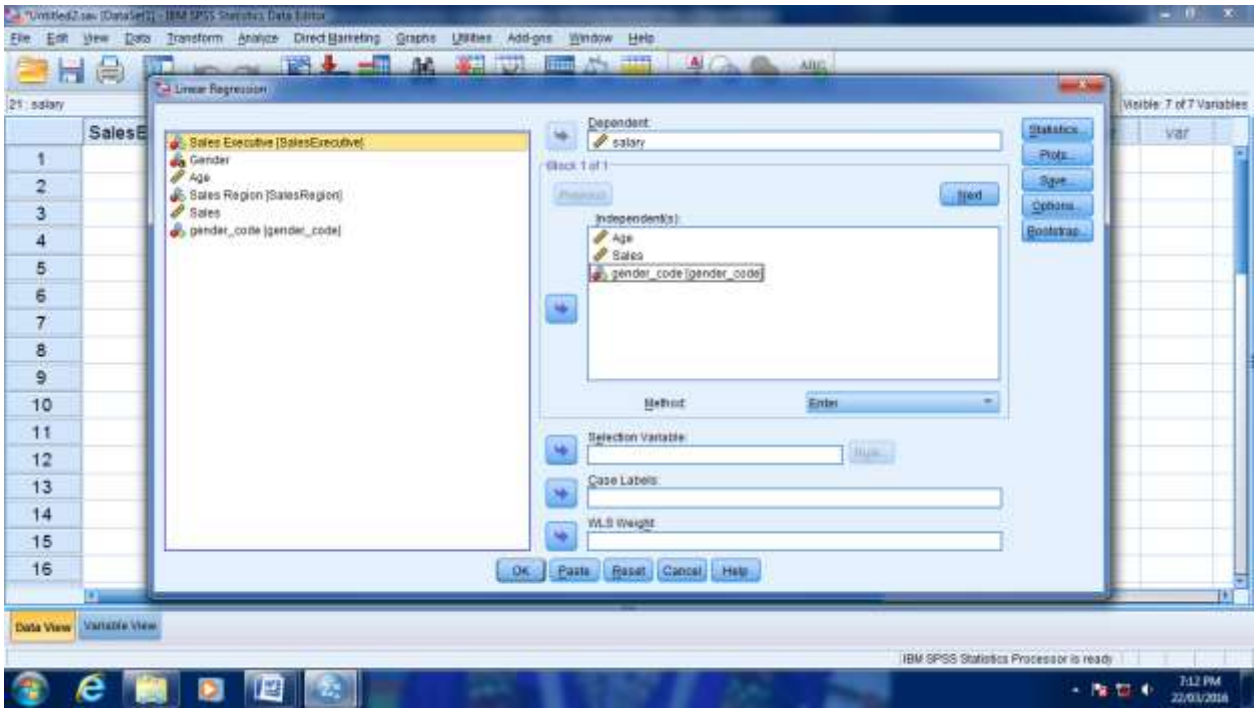
Hence the multiple regression equation can be written as

Salary = 2288.316 + 27.974(Age) +
490.781(Sales)

Now we test whether the age or sales or both have significant effect on salary or not.

Here from the last column of the table we see that significant value of age is 0.845 > 0.05, so it has no significant effect on salary, but for sales it is 0.000 < 0.05, so sales has significant effect on salary at 5% level of significant.

Let us check the effect of gender on salary , we introduce gender_code as dummy variable in the model and add gender_code as ` independent variable in the RHS box of regression and again click OK, as



We get the output as

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2368.231	3690.711		.642	.530
	Age	25.781	146.959	.007	.175	.863
	Sales	490.944	20.696	.986	23.722	.000
	gender_code	-82.705	824.856	-.004	-.100	.921

Again here significant value for gender is 0.921> 0.05, so it has no significant effect on sales volume.

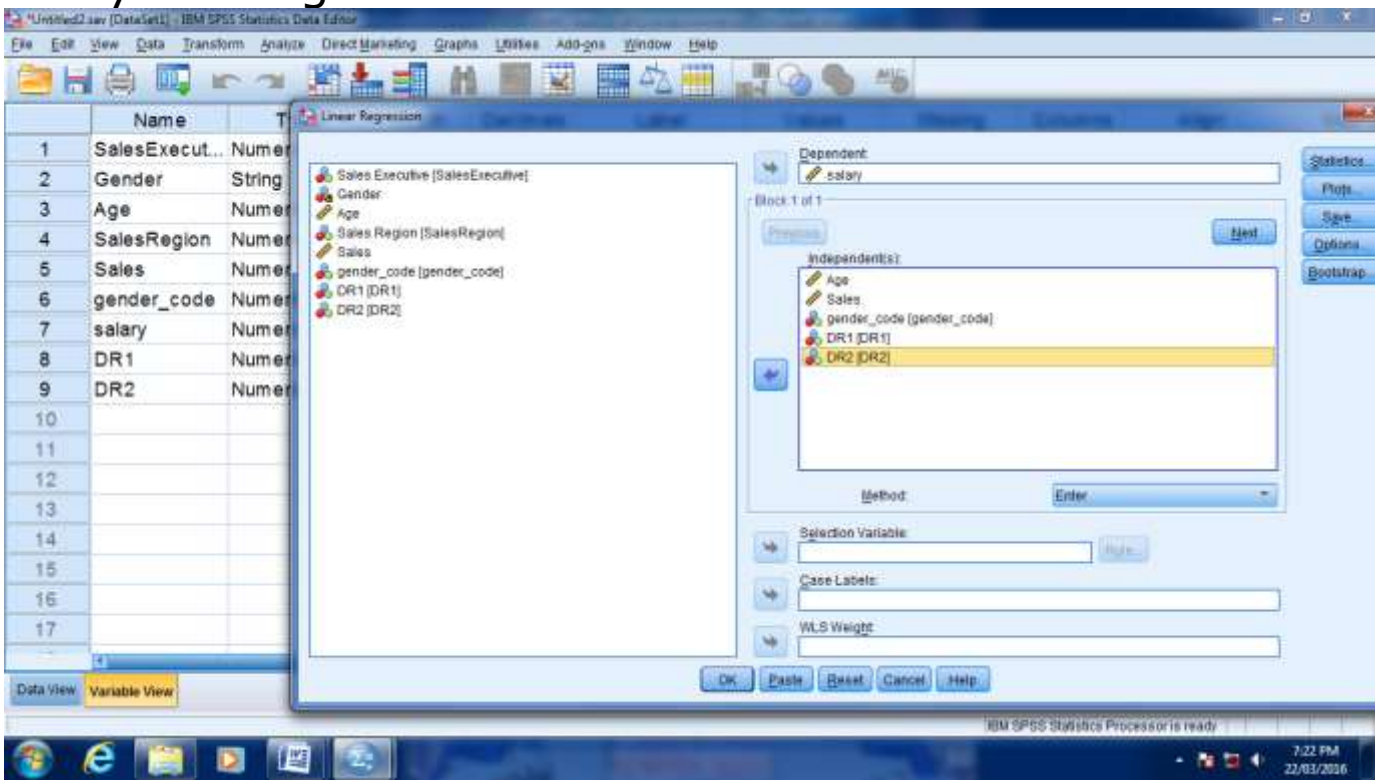
Now we add sales region as independent variable in regression box by considering dummy variables for sales region as

DR1 = 1, if Region is 1 otherwise 0

DR2 = 1, if region is 2 otherwise 0

So first of all we enter the two new columns for these dummy variables in the data sheet as DR1 and DR2 which assume the values 1 or 0.

Again select linear regression from the data analysis we get the box as



Then transfer these two dummy variables in independent box of regression in which we have already transferred age, sales and gender_code and salary in dependent box.

Then click on OK so we get the output as,

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.990[a]	.980	.972	1657.52159

Here R-squared value increases from 0.973 to 0.980 and significant value is 0.000< 0.05 that is this model becomes more appropriate than the earlier models after adding the sales regions as dummy variables.

ANOVA[a]

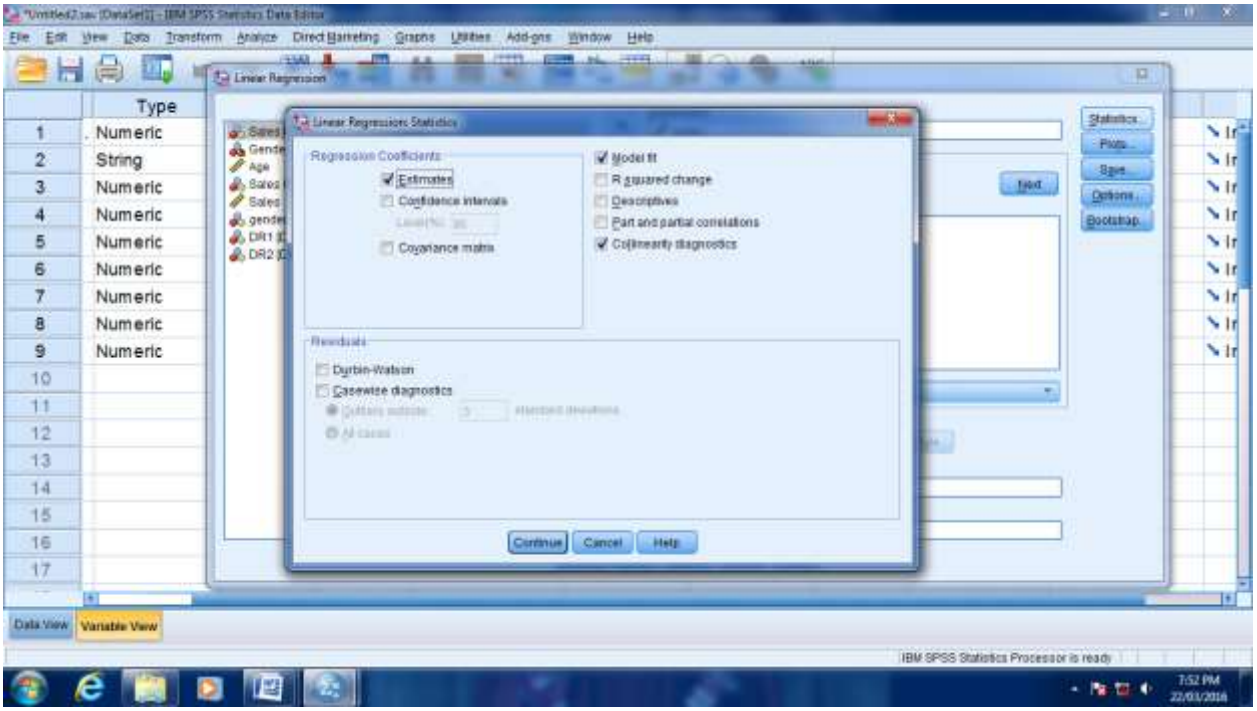
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1842086710.665	5	368417342.133	134.098	.000[b]
	Residual	38463289.335	14	2747377.810		
	Total	1880550000.000	19			

6. Collinearity Diagnostics:

In multiple regression equation sometimes independent variables may be correlated with each other. Such situation is called multicollinearity. Due to presence of multicolinearity the model wrongly gives a very high value of Multiple correlation coefficient R. It is necessary to remove if any multicollinearity in the independent variables.

The following steps are used to test multicollinearity

Steps: Select Analyze□Regression□Linear□ Statistics□Collinearity Diagnostics



we get the output as

Coefficients[a]

Model		Unstandardized Coefficients		Collinearity Statistics	
		B	Std. Error	Tolerance	VIF
1	(Constant)	204.225	4741.050		
	Age	82.717	182.528	.540	**1.853**
	Sales	513.843	26.594	.512	**1.952**
	gender_code	-274.183	772.349	.960	**1.042**
	DR1	-1184.327	1330.098	.324	**3.091**
	DR2	1054.929	1294.690	.390	**2.562**

Here VIF values are less than 5 so there is no multicollinearity amo000000ng the independent variables. If any value of VIF is greater than 5 then we remove one by one independent variable from the correlated independent variables and decide the appropriate model.

SUMMARY

In SPSS software we can easily perform, all the aspects of econometrics. First of all we have seen how to entre the data, then how to carry out the test for single mean, two mean, two variables. Next we have seen how to develop a general linear regression model with the help of number of independent variables and dependent variables. We have also see, which independent variables has a significant effect on dependent variable.

Thank You.