Instructor: Doug Golick **SPSS 2**

Outline

1. Review: Setting up data
    a. The Data Editor – Data View
    b. Computing and recoding data
    c. Summarizing data

2. Inferential statistics
    d. Chi-square
    e. T-test
    f. Correlation
    g. Analysis of Variance (ANOVA)
    h. Regression

3. Displaying data
    i. Working with tables
    j. Working with graphs

4. Other, Time permitting
    k. Q & A

Course materials supplied by Naseem Sherwani, Penn State University.

Section 1: Review

Starting SPSS

1. From the Windows, go to *Start* icon under Windows 98, Windows 2000.
2. Choose:
   All Programs
   SPSS for Windows
   SPSS for 15.0 for Windows (On your machine SPSS version 14.0)

or

You can also find an SPSS icon under the programs menu item by double-clicking to an SPSS file.

B. The Data Editor Window-Data View

To open a data file

From the menus choose
File
Open
Data

You can use the Open File button on the toolbar. 

The file we will use is in the folder (directory) in which SPSS is installed.

- Double click the *Tutorial* folder
- Double-click the *Sample-files* folder
- Click the file *demo.sav* or just demo if file extensions are not displayed.
- Click *Open* to open the SPSS data file.

Note. You can see other file formats by using the *Files of type* drop-down list.
The Data Editor Window displays the data set. The data is arranged in a spreadsheet format that contains **variables in columns and cases in rows.** There are two sheets in the window.
1. The *Data View* is the sheet that is displayed when you first open the Data Editor and contains the data.
2. You can open the second sheet by clicking on the tab labeled *Variable View*, it does not contain data, but contains information about the variables that area stored with the data set. Unlike most spreadsheet tools, the Data Editor can only have one data set open at a time. A distinction must be made between **Variable** & **Value. Variable (s)** is (are) a measure (s) or classification scheme(s) that can have several values. Variables are used to represent.

- Several values
- Different types (Numbers, Strings, Currency, and Dates).
- Response to each question on a survey

Suppose Gender is a variable.

Value(s) is (are) the number(s) or categorical classification (s) representing individual instance(s) of the variable being measured.

In the Variable View sheet, **each row indicates a different variable; each column represents a different characteristic of that variable**. Defining each variable is very important component in data analysis. The SPSS viewer is also displayed showing the data set name. You can minimize the SPSS viewer to display the data editor.

By default, you can save data in the SPSS-format files. The file extension is ".*sav*" You can activate a particular case (row) or variable (column) by clicking on the row number or column title. You can also drag and drop rows or columns to adjust row or column sequence.

The Output-SPSS Viewer
When we begin any kind of statistical analysis in SPSS in the Data Editor Window, a second window opens up automatically. This is where all the output (tables, charts, or results) from your analysis is located. This is called the Output Window. The top of window is almost identical to the Data Editor Window,

B. Computing and recoding data

SPSS has a feature to compute new a new variable. From the menu in the Data Editor window choose select compute from the *Transform* pull down menu and enter the name of the target variable. Then enter the numeric expression using the numeric pad available, specifying the function, or pasting the existing variable into your formula. You can see the new variables at the end of the spreadsheet. Suppose if we want to know the **employee's age at the time he or she started that job**. It is very simple in SPSS. You have data about the **current age** and **the number of years at current job,** which should have approximate age at which the respondent started that job**.** Right! Now, simply subtract the **current age** and **the number of years at current job** to compute the new variable

Variables can be computed conditionally For example in the above example if you were only interested in the people who have just completed their high schools, you could create a condition that would compute a new variable only if any individual has completed high school. To do this, click in *If.* Click on the button larked include *if case satisfies condition* to activate the Gray area of the dialog box. This specifies the condition for computing a new variable in the input box at the top right of the dialog box. Variables can be moved to the condition box by selecting the variable's name. Then, clicking the arrow button between the two boxes. Click the continue button to return to previous page.

SPSS allows creating new variables, not by means of calculation with already existing variables but rather by dividing a pre-existing variable into different categories and coding each category differently. To *Recoding*, go to the Transform pull down menu and then *Recode*. Type the name of the New Variable in the name dialogue box. Please notice two options, 1. Recoding into the same variable will overwrite anything already existing

in that variable; 2. Recoding into a different variable will add new information in existing variable.

Process
There are two different windows that control the recoding into different variables.
and press the *Change* button to change the name of the variable. Select the *Old and New Values* button to *change* the existing variable. After the change of each value, press the *Add* button. Press *Continue*.

Practice
From *Employee data* file
Go to File →Open → Program files → tutorial → Sample files. empoyeey.sav
Calculate a new variable called "*saldiff*" based on two old variables *salary* and *salarybegin* by calculating through formula. Do this calculation only for those who are female.
Select cases have 12 months of experience
Select cases having 36 months of experience

C. Summarizing data

You can summarize information about the variables in your data set by using the facility of Descriptive Statistics under the *Descriptives pull down menue.* You can find Frequencies, Descriptive, Crosstab, Explore and Ratio. These statistics will allow for data screening, outlier identification, description about assumption checking, and identifying differences among population. To find Descriptive Statistics such as Mean, Median, Mode, and Standard Deviation under the *Descriptives* option, this will allow you to select the variables of your choice on the left side of the dialog box will appear.

First of all let us open a data file
Descriptive Statistics
Frequencies

1. Open the File, Choose '*Analyze*'
2. Choose '*Descriptive Statistics*'
3. Choose '*Frequencies*'.
4. A box pops up. On the left side, a menu gives us the list of the variables in our data set. Choose the variable of your choice and click the right arrow in the middle.
5. Put that variable in the empty box on the right.
6. Then click on the button marked '*OK*'.
7. In the '*Output Viewer*' the first box gives us information about the name of the variable. For example, how many respondents answered the question (866) and how many did not (those people coded as missing, 3 in this case)?
    a) Column 1 gives us the value labels for the variable we are looking for.
    b) The second column gives us our actual frequencies.
    c) The third column expresses that as a percentage.

d) The fourth column, labeled 'Valid Percent' gives us the percentage of kids who are not missing.
e) The final column gives us the '*Cumulative Percent*'. This just means that the percentages are added up to 100.

However, there are several other descriptive statistics options available. Clicking on the Statistics button procedure, a dialog box with several descriptive statistics will be open for you.

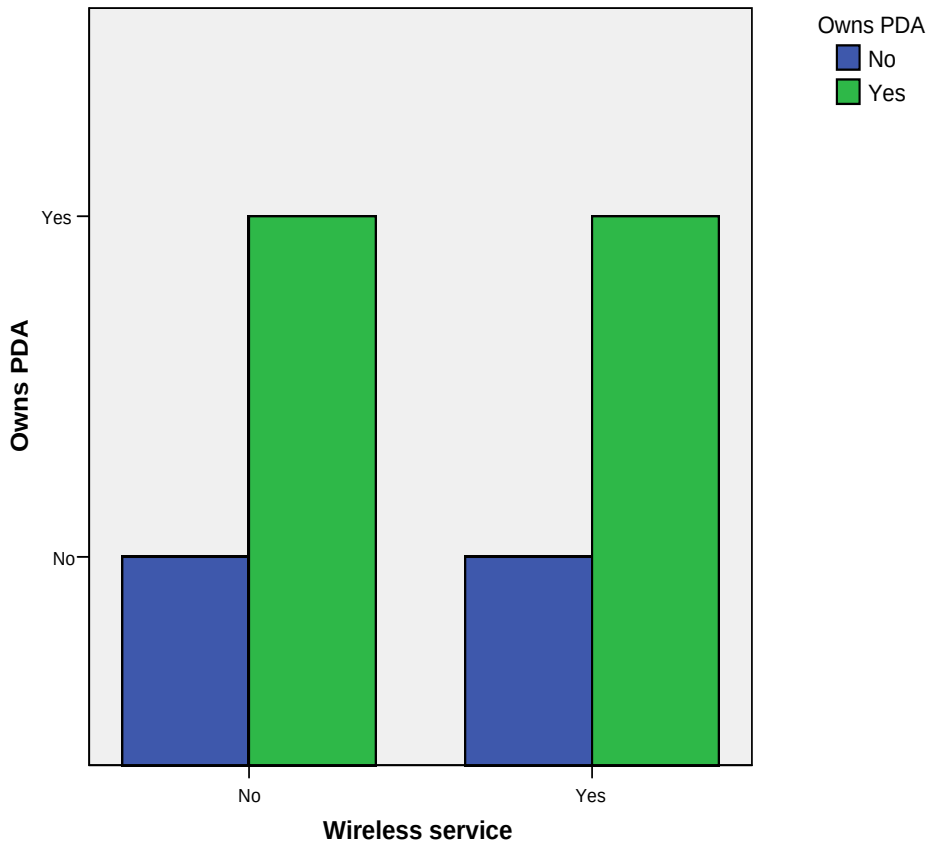
Graphs: Some statistical procedures can create high resolution charts. For this we can use the Graphs Menu to create charts. For example you can create charts that show the relationship between wireless telephones and PDA (personal digital assistance) ownership.
From the menus choose:

1. *Graphs*
2. *Chart Builder…*
3. Click *the Gallery tab* (if it is not selected)
4. Click Bar (if it is not selected)
5. Drag the Clustered Bar icon onto the *canvas*, which is the large area above the Gallery.
6. Scroll down the variable list, right click-the Wireless service [*wireless*] variable to the x axis.
7. Right click *Owns PDA [ownpda]* and choose Nominal as its measurement level.
8. Drag the *owns PDA [*ownpda*]* Variable to the cluster group drop zone in the upper right corner of the canvas.
9. Click OK to create the chart.

Wrap up
Viewer
The bar chart shows that people having wireless phone service are far more likely to have PDAs than people without wireless service.

You can edit charts and tables by double-clicking them in the contents pane of the Viewer Window, and you can Copy and Paste your results in other applications.

To Exit SPSS:
From the Menus choose:
File
Exit
You will get an alert asking if you want to save your results. Click 'Yes' if you want to save the results, otherwise click 'No'.



Descriptive

Distribution: When we collect data, we end up with a group of one or more variables. If you take the scores on one variable, and arrange them in order from lowest to highest, what you get is the ***distribution*** of scores. SPSS helps the researcher to describe the characteristics of the distribution. Such as the shape of the distribution, how spread out the scores are, and what the most common score is. There are the following sets of distribution.

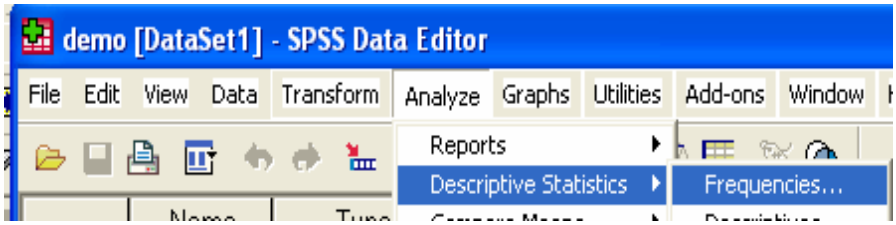
1. **Measures of Central Tendency**
2. **Measures of Variability**

Measures of Central Tendency: indicate the **location** of the distribution; they include the mean, median, and 5% trimmed mean. Follow the steps to calculate these in SPSS.

Choose from the Menus

1. Analyze

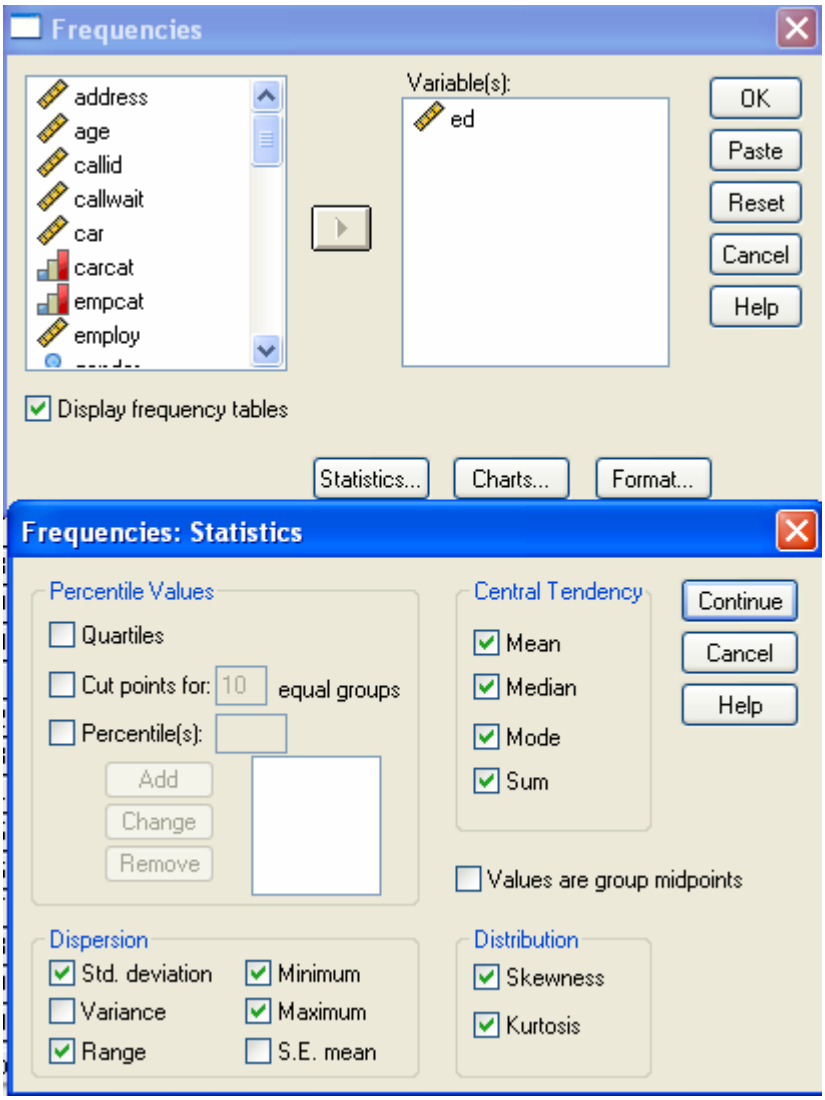
2. Descriptive Statistics
3. Frequencies



4. Select the variables of your choice,
5. Click the right arrow to add it to the List of key variables in the right box

Here we have to calculate measures of central tendency

6. Choose 'Statistics'
7. Tick the boxes for the measures you want



8. Click ‘Continue’.
9. Click ‘OK’.

Output Window: When we take all the following steps, a screen appears that gives the following options.

Here is Mode= 2 , Mean= 2.59 , and Median= 2.00

The measure of central tendency does not give us the full description of the story. It does not tell about the education level of students in the distribution. We need to know the spreads of the values around the mean or median etc.

Measures of Dispersion: show the dissimilarity of the values; these include standard error, variance, standard deviation, minimum, maximum, range, and interquartile range.

Range: Difference between the largest score and minimum score of a distribution.

Variance: The sum of squared deviations divided by the number of cases in the population, or by the number of cases minus one in the sample.

Standard Deviation: The average deviation between the individual scores in the distribution, and the means for the distribution.

To calculate the measures of ‘Spread’ follow the steps from 1 to 6 in the previous exercise, and then select the measures of Dispersion’; Standard Deviation, Variance, Range, Maximum and Minimum values, etc. Click ‘Continue’, and then ‘OK’ on next box.

Statistics

Level of education

N	**Valid**	***6400***
	Missing	***0***
Mean		***2.59***
Median		***2.00***
Mode		***2***
Std. Deviation		***1.199***
Skewness		***.274***
Std. Error of Skewness		***.031***
Kurtosis		***-.982***
Std. Error of Kurtosis		***.061***
Range		***4***
Minimum		***1***
Maximum		***5***
Sum		***16557***

Explore:

From the Menus ‘’Choose’

1. Analyze
2. Descriptive Statistics

3. Explore

In the Explore dialog box, select Both or Statistics in the Display group to enable the Statistics pushbutton.
Then click 'Statistics' or one or more Statistics.

The descriptive statistics also include measures of the **shape** of the distribution; skewness and kurtosis are displayed with their standard errors. The 95% level confidence interval for the mean is also displayed.

Crosstabs: The purpose of cross tab is to show the relationship between two or more a) categorical or nominal variables, b) a nominal and an ordinal variable, c) two ordinal variables. Cross tab can be used with continuous data only if such data is divided into separate categories such as age 0-19, 20-39, etc.

Doing Cross tabulation in SPSS

1. Go to 'Analyze'
2. Descriptive Statistics
3. Cross tabs-this is the fourth item in the pop-up list.
4. Cross-tabulation box pops up

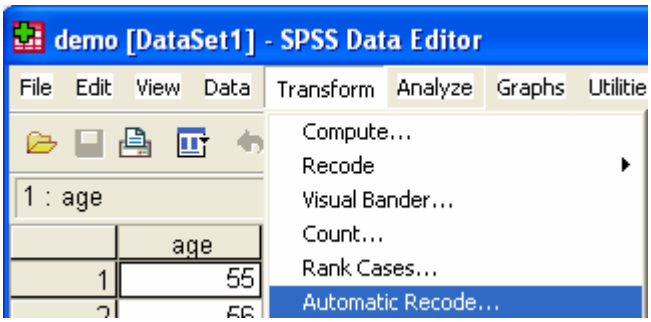
Section II. Inferential Statistics

A. Chi-Square Test: This test procedure is used to test the hypotheses about the relative proportion of cases falling into several mutually exclusive groups. You can test the hypothesis that people who participated in the survey occur in the same proportion of gender as the general population (50% males, 50% females). In this example, you need to recode the string variable Gender *[gender]* into a numeric variable before you can run a procedure.

Practice: Open *Employee_data* file

From the menu choose the following steps

Step 1. Transform
Step 2. Automatic Recode

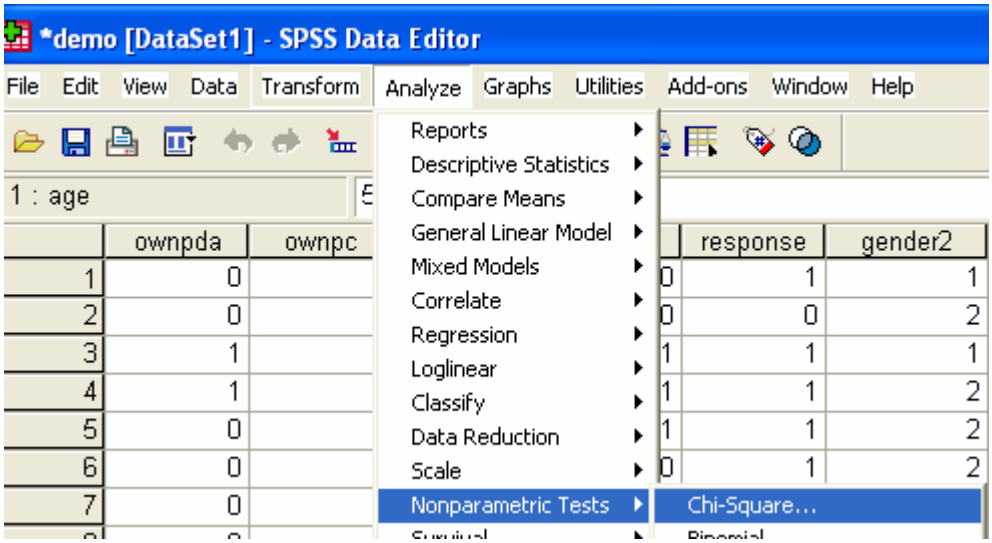


This opens the automatic Recode dialog box.
Figure
Automatic Recode dialog box

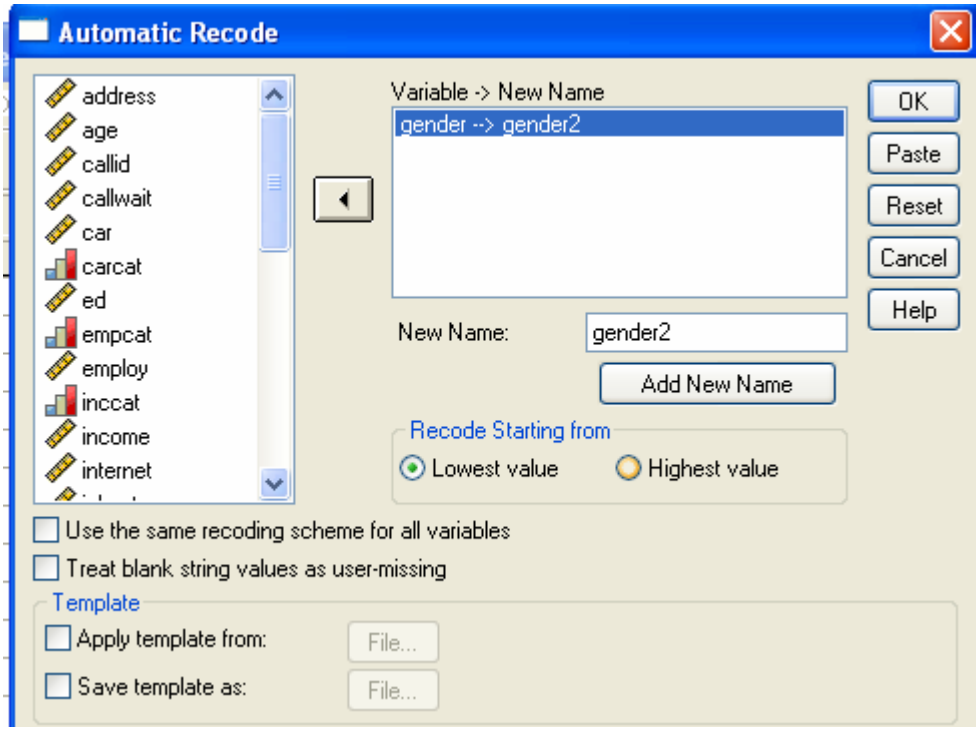
Procedure to follow
Step 3 Select the variable Gender [gender] and move it to the Variable → New Name list.

Step 4 Type gender2 in the New Name text box, and then click the Add New Name button.
Step 5 Click OK to run the procedure

This process creates a new
numeric variable called gender2, which has a value of 1 for females and a value of 2 for males. Now a chi-square test can be run with a numeric variable. From the menu Choose the following steps
Step 6 Analyze

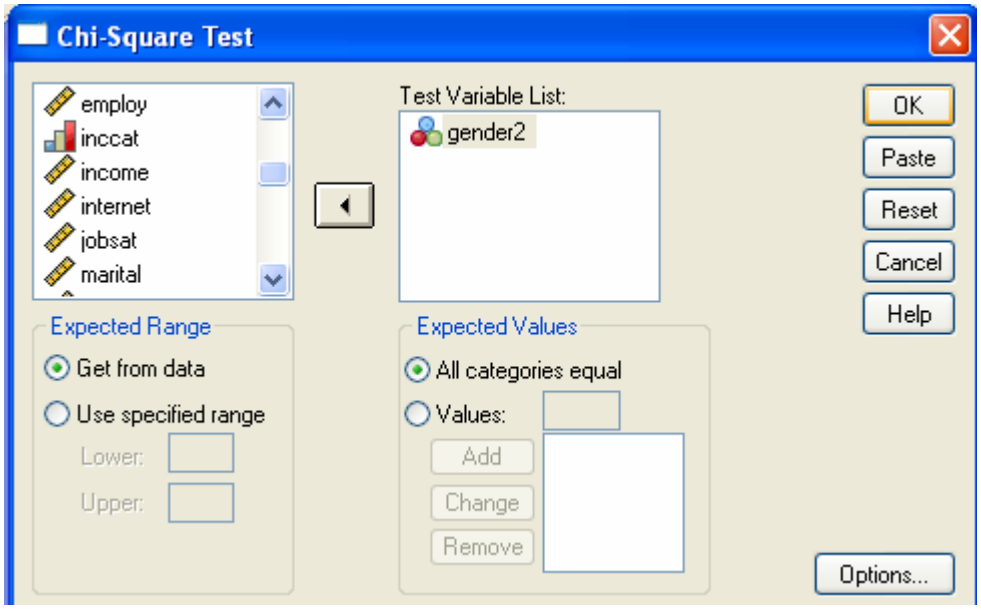


Step 7



Nonparametric Tests → Chi-Square…
(Notice: The Nonparametric tests submenu on the Analyze``` menu provides nonparametric tests for one sample or for two or more paired or independent samples. Nonparametric tests do not require assumptions about the shape of the distribution from which the data originates.).

This opens the Chi-Square Test dialog box.
Figure
Chi-Square Test dialog box



Step 8 Select Gender [*gender2*], because in the general population of working age, the number of male and female is approximately equal.
Step 9 Click OK to run the procedure.

Frequencies

Gender

	Observed N	Expected N	Residual
Female	***3179***	***3200.0***	***-21.0***
Male	***3221***	***3200.0***	***21.0***
Total	***6400***		

Test Statistics

	Gender
Chi-Square(a)	***.276***
Df	***1***
Asymp. Sig.	***.600***

a 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 3200.0.

Wrap up

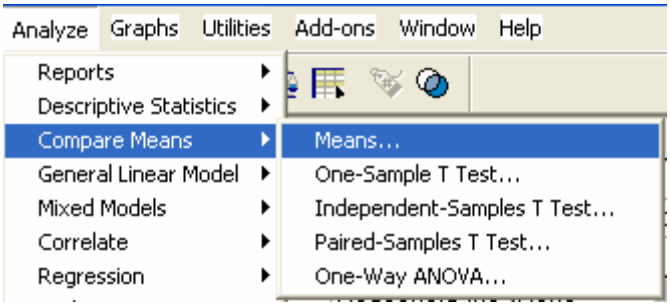
The output shows a table of the expected and residual values for the categories. The significance of the Chi-Square test is 0.6.

B. T-Tests

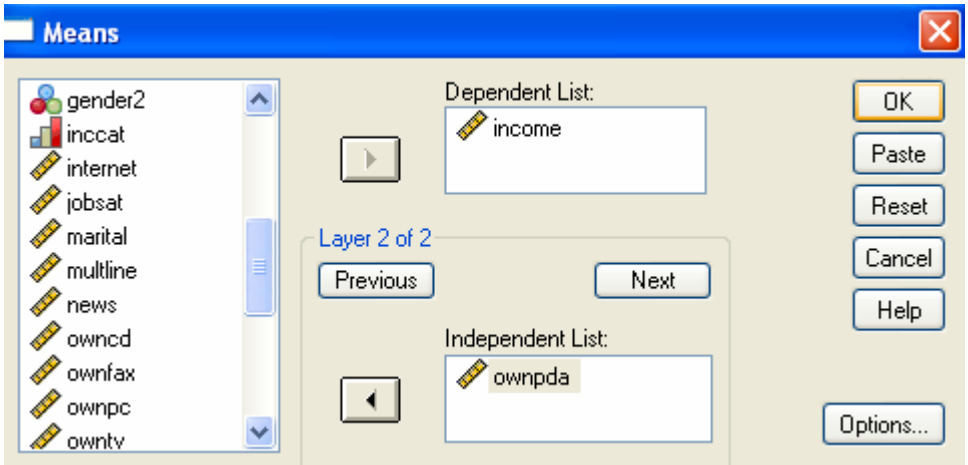
The Compare Means submenu on the Analyze menu provides techniques for displaying statistics and testing whether differences are significant between two means for both independent and paired samples. Alternatively, One-Way ANOVA procedure to test whether differences are significant among more than two independents means.

From *demo.sav* file, there are several variables available for dividing people into groups. Then you can calculate various statistics in order to compare the groups. For example, you can compute the average (mean) household income for males and females. To calculate the means, you can choose the following steps:

From the menu choose:
Analyze
Compare Means
Means…



This opens the Means dialog box.



Choose the following steps
Step 1 Select Household income in thousands *[income]* and move it to the Dependent List.
Step 2 Select Gender *[gender]* and move it to the Independent list in Layer 1.
Step 3 Click Next to create another layer.
Step 4 Select Owns PDA *[ownpda]* and move it to the Independent list in Layer 2.
Step 5 Click OK to run the procedure.

Report

Household income in thousands

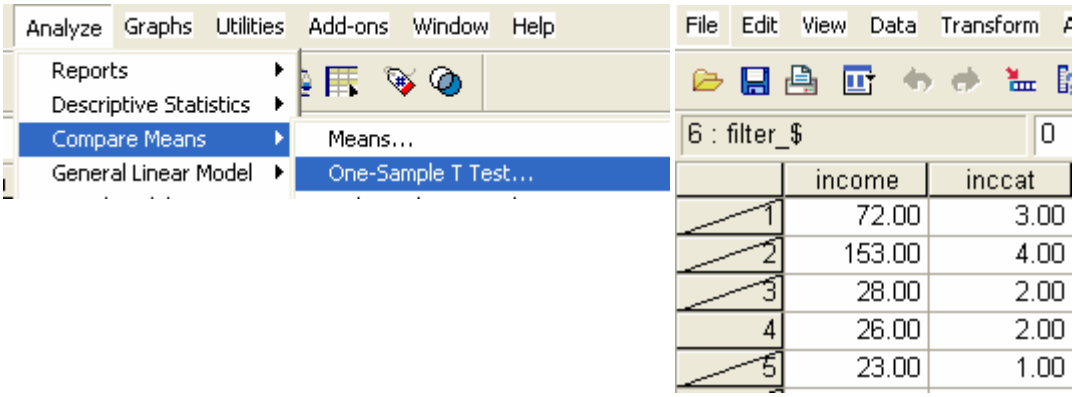
				Mean	N	Std. Deviation
Gender	Female	Owns PDA	No	*65.7480*	*2528*	*72.62241*
			Yes	*80.5530*	*651*	*85.82168*
			Total	*68.7798*	*3179*	*75.73510*
	Male	Owns PDA	No	*66.0940*	*2565*	*72.20572*
			Yes	*86.0625*	*656*	*109.43673*
			Total	*70.1608*	*3221*	*81.56216*
	Total	Owns PDA	No	*65.9222*	*5093*	*72.40594*
			Yes	*83.3183*	*1307*	*98.38646*
			Total	*69.4748*	*6400*	*78.71856*

Wrap up

What does it mean? It means the average (or mean) household income for males and females who own a PDA
One of the most frequently asked questions in data analysis is whether two means are significantly different from one another. Do people who exercise have different average serum cholesterol levels from those who do not? Does enrollment in a truancy reduction program decrease unexcused absences? Although often not explicitly stated, the question refers not to the sample means but to the underlying population means.

The statistical procedure that is most often used to test the hypotheses about two population means is called *t* test. There are three different types of t tests:

One Samples T Test: You can test whether the household income of people with college degrees differs from a national or state average. Use Select cases on the Data menu to select the cases with Level of Education [*ed]* >=4. Then, run the One-Sample T Test procedures to compute the Household income in thousands *[income]* and the test value 75.



One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Household income in thousands	***1714***	***80.4347***	***93.32004***	***2.25408***

One-Sample Test

	Test Value = 75					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Household income in thousands	***2.411***	***1713***	***.016***	***5.43466***	***1.0136***	***9.8557***

Wrap up

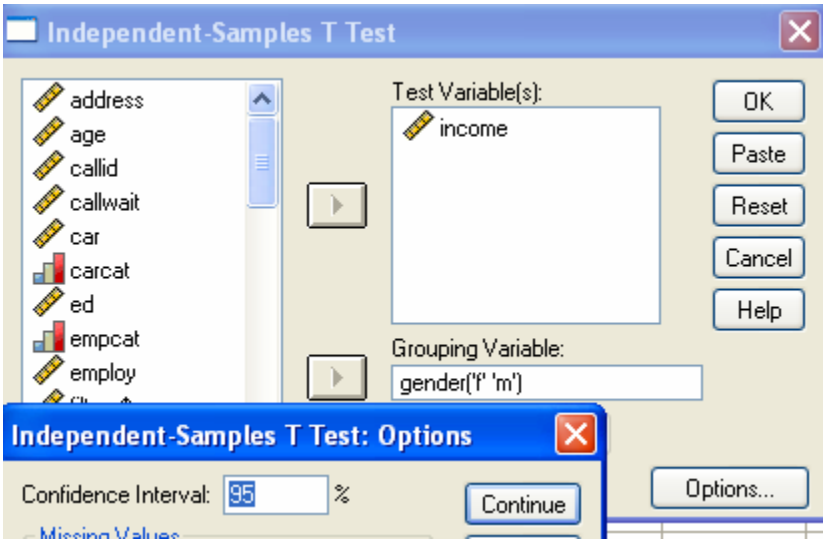
We may accept the hypothesis that the household income of people with college degrees does not differ from a national or state average level.

Independent Samples Test

Independent Samples T test: When you use a t test to compare means of one variable across independent groups, the samples are independent. Males and Females in the demo.sav can be divided into independent groups by the Variable Gender *[gender]*. You can use *t* test to determine whether the mean household incomes of males and females are the same.

Research question: *Do household incomes of men and women differ?*

Procedure
Step 1 Choose Analyze
Step 2 Select
Compare Means
Step 3 Select
Independent
Samples T Test



Group Statistics

			N	Mean	Std. Deviation	Std. Error Mean
Household income in thousands	**Gender**	**Female**	***3179***	***68.7798***	***75.73510***	***1.34323***
		Male	***3221***	***70.1608***	***81.56216***	***1.43712***

Independent Samples Test

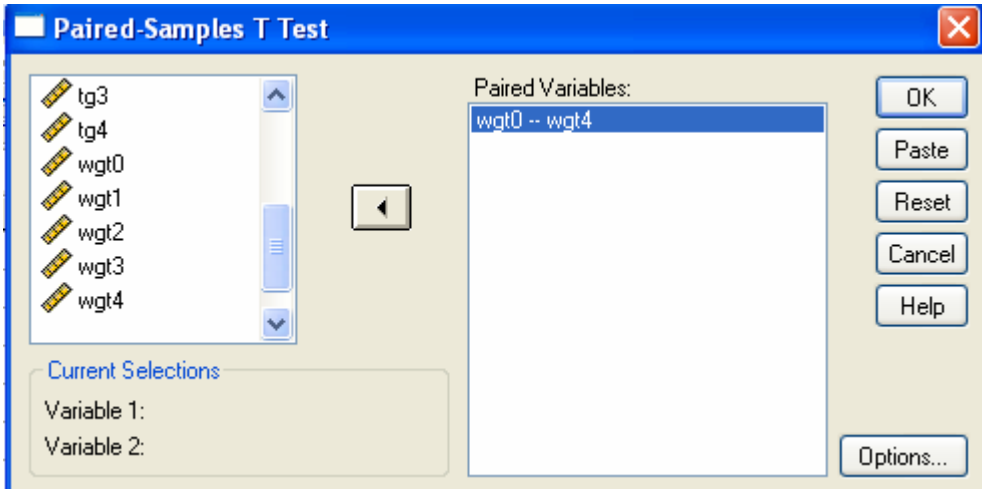
		Levene's Test for Equality of Variances		t-test for Equali			
		F	**Sig.**	**t**	**df**	**Sig. (2-tailed)**	**Mean Differen**
Household income in thousands	**Equal variances assumed**	***1.865***	***.172***	***-.702***	***6398***	***.483***	***-1.3***
	Equal variances not assumed			***-.702***	***6374.362***	***.483***	***-1.3***

Wrap up
The Output table, labeled as **Group Statistics** displays statistics. The second Output table, labeled **Independent Samples Test**, contains statistics that are helpful to answer our research question. It contains two sets of analysis, the first assumes the equal variances and the second does not. To assess whether you should use the statistics for

equal or unequal variance, use the significance level associated with the value under the heading, *Leven's Test for Equality of Variances*. It tests the hypothesis that the variances of the two groups are equal. If the Levine's test for equality of variances is significant ($p<0.05$), you should use the bottom row in the **Independent Samples Test table** to interpret the results. If the *p* value (indicated as,"**Sig**," in the *t* cells on the output) is less than specified alpha level (usually $p=.05$ or .01), then you reject the null hypothesis (μ_0) and accept the alternate hypothesis (μ_1) that population mean of female household income is different than male household income.

Paired-Sample T Test (or matched cases):
When the data are structured in such a way that there are two observations on the same individual or observations that are matched by another variable on two individuals (twins, for example). The samples are paired. In the data file dietstudy.sav. The beginning and final weights are provided for each person who participated in the study. If the diet worked, we expect that the participant's weight before and after the study would be significantly different.

Procedure
Open the data file dietstudy.sav, which can be found in the tutorial\ sample_sample_files\subdirectory of the directory in which your SPSS is installed.
Follow the following steps
Step 1 Analyze
Step 2 Compare Means
Step 3 Paired Sample T-Test
Paired Samples T Test dialog box



Step 4 Click weight [wgt0]. The variable is displayed in the current selection group in the variable list

Step 5 Click Final weight [wgt4]. The variable is displayed in the Current Selection group

Step 6 Click the arrow button to move the pair to the Paired Variable list.

Step 7 Click OK to run the procedure

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Weight	*198.38*	*16*	*33.472*	*8.368*
	Final weight	*190.31*	*16*	*33.508*	*8.377*

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	Weight & Final weight	*16*	*.996*	*.000*

Paired Samples Test

		Paired Differences					t	df
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference			
					Lower	Upper		
Pair 1	Weight - Final weight	*8.063*	*2.886*	*.722*	*6.525*	*9.600*	*11.175*	*1*

Wrap up

The results indicate that the final weight is significantly different from the beginning weight, as indicated by the small probability that is displayed in Sig. (2-tailed) column of the Paired Sample table.

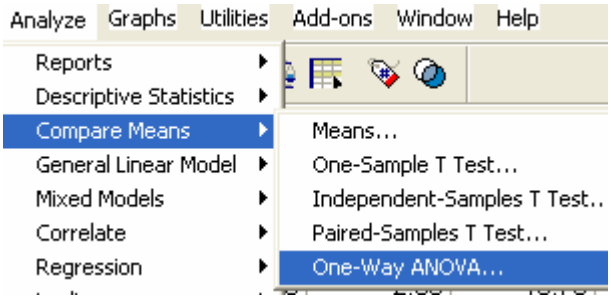
One-Way ANOVA: This procedure produces a one-way analysis of variance for a quantitative dependent variable by a single factor (independent) variable. Analysis of variance is used to test the hypothesis that several means are equal. This technique is an extension of the two-sample t test.

Procedure

Step 1 Analyze

Step 2 Compare Means

Step 3 One-Way ANOVA



The variable Level of Education *[ed]* divides employees into five independent groups by level of education. One-Way ANOVA procedure to test whether Household Income *[income]* means for the five groups are significantly different.

ANOVA

Household income in thousands

			Sum of Squares	df	Mean Square	F	Sig.
Between Groups	**(Combined)**		***376079.699***	***4***	***94019.925***	***15.309***	***.000***
	Linear Term	**Unweighted**	***294335.775***	***1***	***294335.775***	***47.924***	***.000***
		Weighted	***368053.906***	***1***	***368053.906***	***59.927***	***.000***
		Deviation	***8025.793***	***3***	***2675.264***	***.436***	***.728***
Within Groups			***39276042.251***	***6395***	***6141.680***		
Total			***39652121.950***	***6399***			

In addition to determining that differences exist among the means, you may want to know which means differ. There are two types of tests for comparing means: a priori contrasts and post hoc tests. Contrasts are tests set up before running the experiment, and post hoc tests are run after the experiment has been conducted. You can also test for trends across categories.

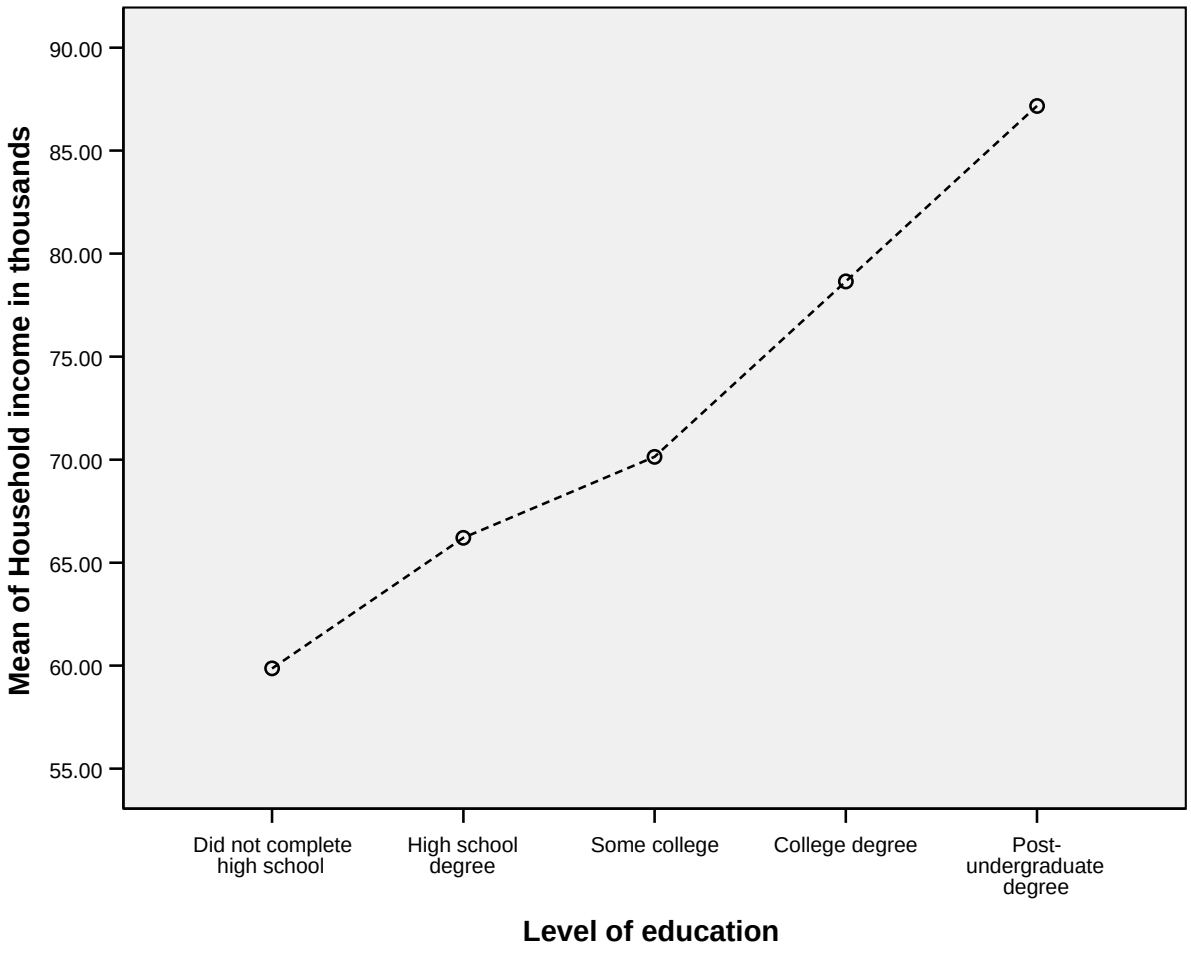
Example. Doughnuts absorb fat in various amounts when they are cooked. An experiment is set up involving three types of fat: peanut oil, corn oil, and lard. Peanut oil and corn oil are unsaturated fats, and lard is a saturated fat. Along with determining whether the amount of fat absorbed depends on the type of fat used, you could set up an a priori contrast to determine whether the amount of fat absorption differs for saturated and unsaturated fats.

Statistics. For each group: number of cases, mean, standard deviation, standard error of the mean, minimum, maximum, and 95%-confidence interval for the mean. Levene's test for homogeneity of variance, analysis-of-variance.

Data. Factor variable values should be integers, and the dependent variable should be quantitative (interval level of measurement).

Assumptions. Each group is an independent random sample from a normal population. Analysis of variance is robust to departures from normality, although the data should be symmetric. The groups should come from populations with equal variances**.**

Mean Plots



Correlation Coefficients
This is a method used to analyze the relationship between two continuous variables is called the correlation coefficient called *person's r*.
Pearson varies between -1 and +1 where +1 indicates a perfect positive correlation (A high score on variable X= a high score on variable Y
Person's r gives a number of aspects of the relationship:

- The *direction of the relationship*: a positive sign indicates a positive direction (high score on X means a high score on Y and vice versa).
- The *strength of the relationship*: the closer to 1 (+ or -) the stronger the relationship.
- A value of '0' indicates 'No' linear relationship
- The statistical significance of the relationship.

Steps to calculate the Pearson's r in SPSS

We will look at the relationship between Beginning Salary and Current Salary, both continuous variables (percentages).

1. Go to '*Analyze*'
2. In the pop-down menu, choose '*Correlate*'. A new pop-down menu appears
3. Choose '*Bivariate*', as we are looking at the relationship between two variables
4. A new box labeled 'Bivariate Correlations' now appears. Select the variables from the list on the left. We choose 'Beginning salary *[salbegin]* and click on the arrow, and then 'Current Salary [salary] and click on the arrow again. Both now appears in the right-hand box.
5. Click 'OK' and SPSS will calculate Pearson's r.
6. Output Viewer: Three important pieces of information are given:
    a) Pearson's r correlation coefficient
    b) The significance level ('Sig.')
    c) The number of cases for which we have data on both the variables.

(Note: We can put more than two variables. We can calculate correlation for up to 100 variables at once.).

Partial Correlation Coefficient: This describes the linear relationship between two variables while controlling or adjusting the effects of one or more additional variables.
Steps to calculate Partial correlation coefficient

1. Go to '*Analyze*'.
2. Select '*Correlate*'.
3. Select '*Partial*'.
4. Select two or more numeric variable for which Partial Correlations are to be computed e.g.,, Beginning Salary *[salbegin]* & Current Salary *[salary]*.
5. Select one or more numeric control variables e.g. Months since Hire *[jobtime]* & previous Experience *[prevexp]*.
6. Test of significance. You can select one or two tailed probabilities.

- If the direction of association is known in advance, select one-tailed test,
- If the direction is not known in advance, then select two-tailed test.

7. Display actual significance level. By default the probabilities and the degrees of freedom are shown for each correlation coefficient. If you deselect this item, coefficients significant at .05 levels are shown with a single asterisk, and coefficients shown at .01 levels are shown at double asterisk and the degrees of freedoms are suppressed. This setting affects both zero order correlation matrices.

Practice Exercise
Please apply the relevant Correlation Coefficient to answer the following research questions.

What is the relationship between Current Salary *[salary]* and Years of Education *[ed]*?
What is the relationship between Current Salary [salary] and Years of Education [ed] by controlling Minority Classification *[minority]* and Employment Category *[emplcateg]*.
Please interpret your results.

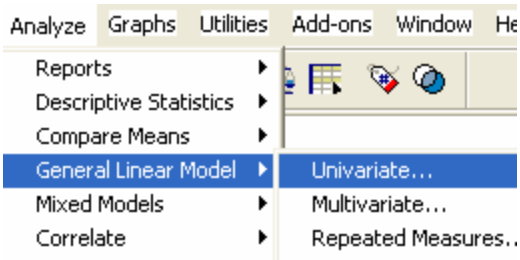
Let us create a scatterplot in a useful graphical depiction of the relationship between two continuous variables. You can create this graph by following the steps

Step 1 Go to Graphs drop down menu and select *Scatter*
Step 2 Select *Simple*
Step 3 Click on *Define* to produce a simple Scatterplot
Step 3 Choose the variables for Y axis and X axis
Step 4 Click *OK*.

D. ANOVA

Most of the procedures conducting analysis of variance (ANOVA) can be found under the General Linear Model (GLM) (If you have only one factor, you can use the One-Way ANOVA procedure on the Compare Means submenu). Under the GLM drop down menu, there are three options

The GLM command is used to determine whether there are differences between groups on the basis of one or more outcome variables..



Univariate: It is used when we have only one dependent variable but may have several independent variables that can be fixed between subject factors, random between subject factors, or covariates.

Procedure

Step 1 Go to Analyze

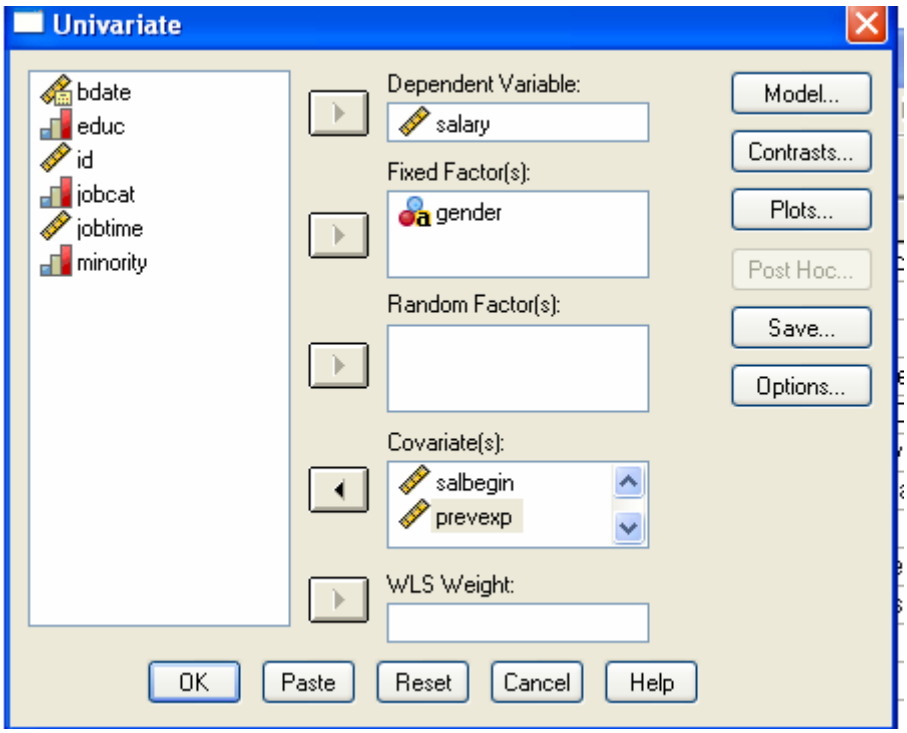
Step 2 Choose General Linear Model

Step 3 Univariate

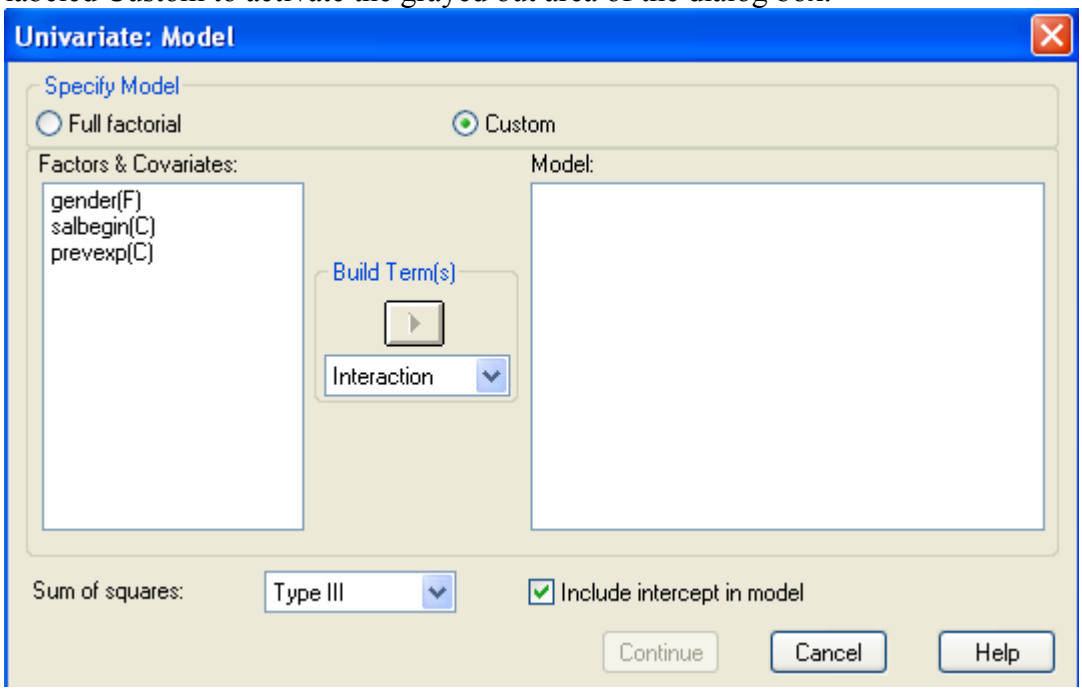
Step 4 The current salary has been fixed as dependent variable

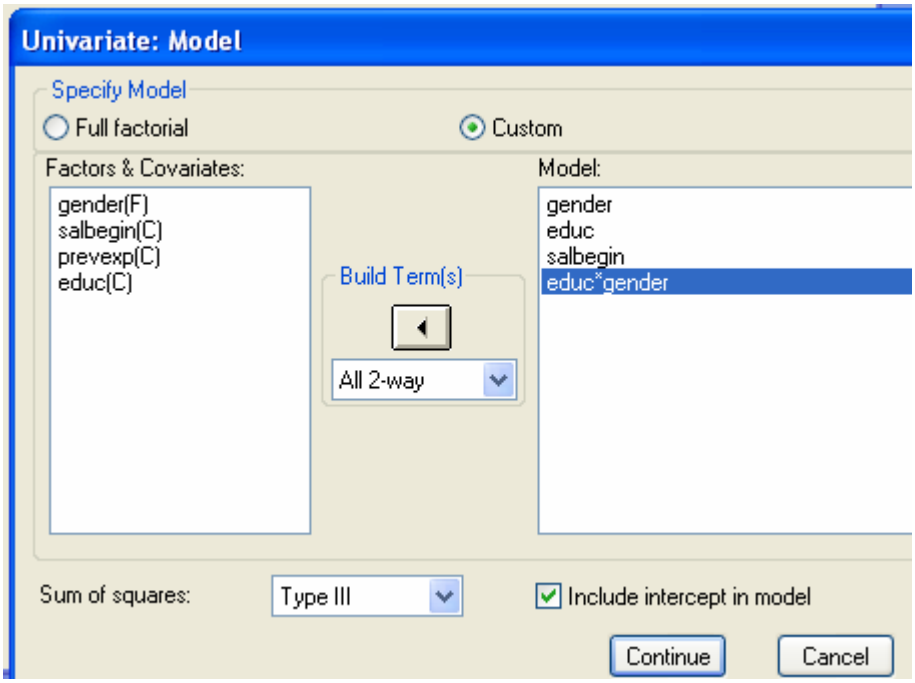
Step 5 The variable Gender has been identified as a fixed factor

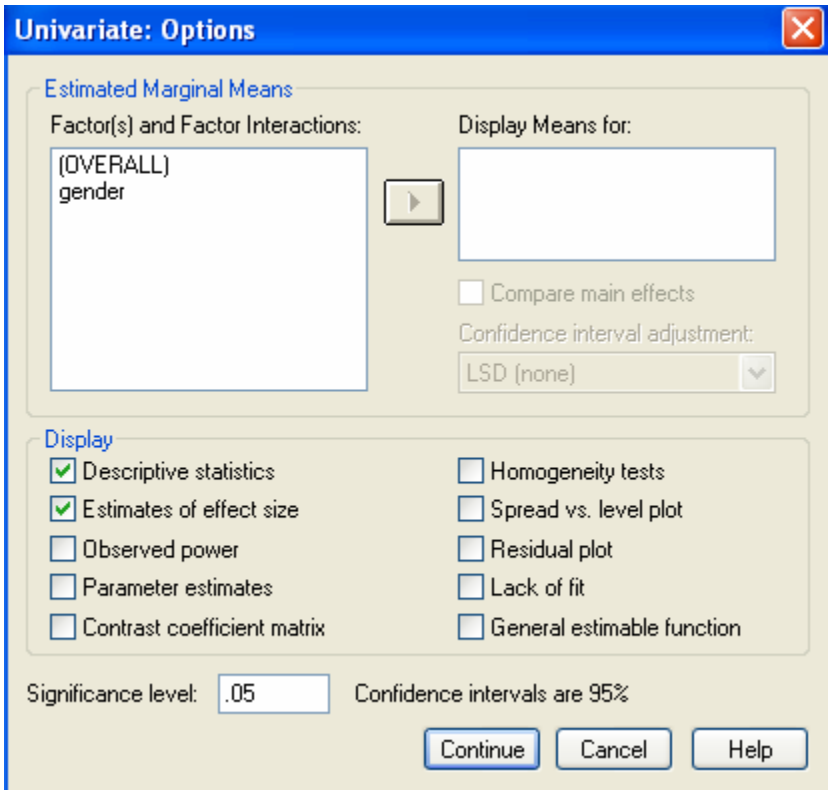
Covariate in your model represents a quantitative independent variable. Covariates are often entered in a model to reduce error variance: by removing the effects of the relationship between the covariate and the dependent variable. In this way, we get a better estimate of the amount of variance that is being accounted for by the factors in the model. Covariates can also be used to measure the linear association between the covariate and the dependent variable. In this situation, a linear relationship indicates that the dependent variable increases or decreases in value as the covariate increases or deceases in value. A covariate should be a continuous variable.



This is the default model and you are free to modify it by simply clicking on button labeled Custom to activate the grayed out area of the dialog box.







Tests of Between-Subjects Effects

Dependent Variable: Current Salary

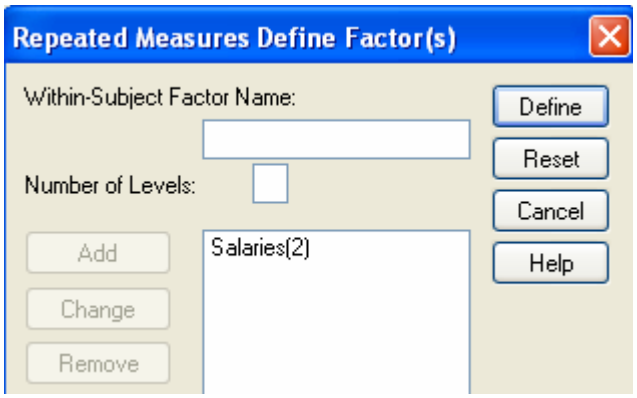
		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Source	Corrected Model	109616349797.119(a)	4	27404087449.280	454.150	.000	.795
	Intercept	673305814.465	1	673305814.465	11.158	.001	.023
	gender	40261665.287	1	40261665.287	.667	.414	.001
	educ	2135942385.341	1	2135942385.341	35.398	.000	.070
	salbegin	37431989544.697	1	37431989544.697	620.336	.000	.569
	gender * educ	95910508.051	1	95910508.051	1.589	.208	.003
	Error	28300145639.221	469	60341461.917			
	Total	699467436925.000	474				
	Corrected Total	137916495436.340	473				

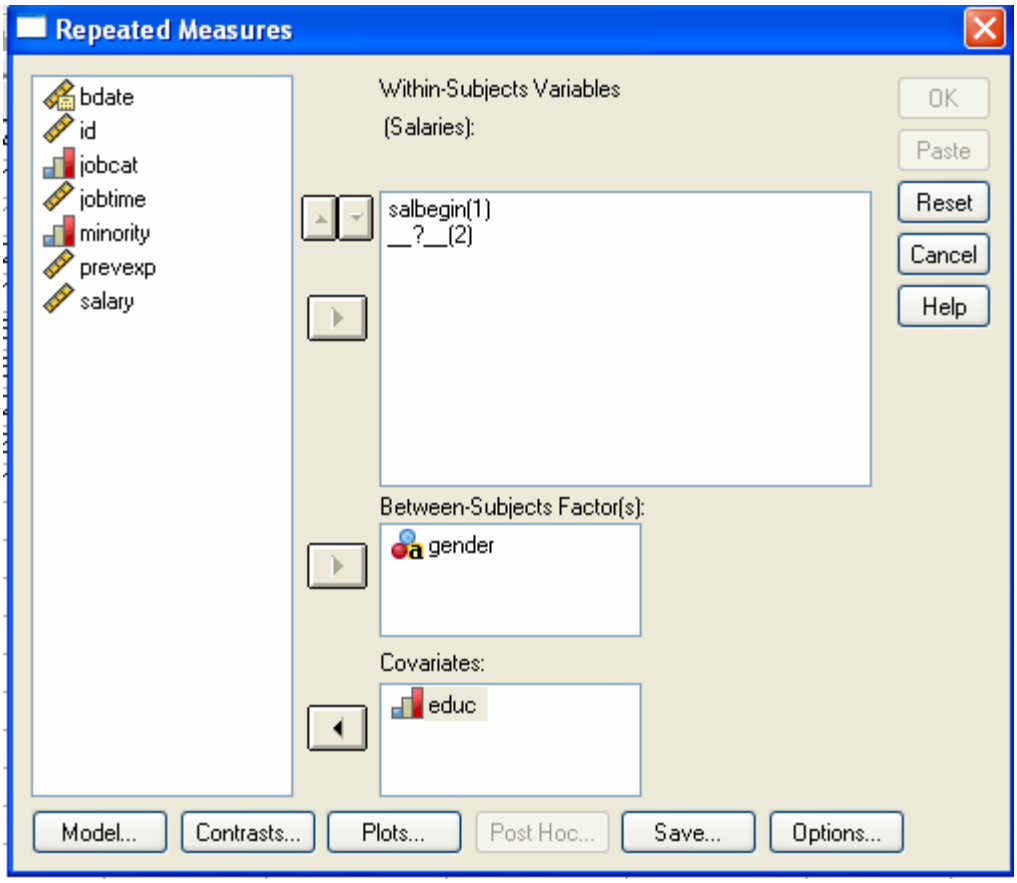
a R Squared = .795 (Adjusted R Squared = .793)

Step 1 Go to Analyze
Step 2 General Linear Model
Repeated Measures

The **Multivariate** GLM is used in situations where there is more than one dependent variables and independent variables are either fixed between subject factors or covariates.
The Repeated Measures GLM is used when you have more than one measurement occasion for a dependent variables and have fixed between-subject factors or covariates as independent variables.

Research question:
Is there any difference in salary between male and female, while controlling for the influence of previous experience and beginning salary?
Using the employee data set





Tests of Between-Subjects Effects

Dependent Variable: Current Salary

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
	Corrected Model	10961634979 7.119(a)	4	27404087449. 280	454.150	.000	.795
	Intercept	673305814.46 5	1	673305814.46 5	11.158	.001	.023
	gender	40261665.287	1	40261665.287	.667	.414	.001
	educ	2135942385.3 41	1	2135942385.3 41	35.398	.000	.070
	salbegin	37431989544. 697	1	37431989544. 697	620.336	.000	.569
	gender * educ	95910508.051	1	95910508.051	1.589	.208	.003
	Error	28300145639. 221	469	60341461.917			
	Total	69946743692 5.000	474				
	Corrected Total	13791649543 6.340	473				

a R Squared = .795 (Adjusted R Squared = .793)

Regression

This is a technique that can be used to investigate the effects of one or more predictor variables on an outcome variable. Regression allows you to make a statement about how well one or more independent variables will predict the value of a dependent variable.

Simple regression includes one independent variable

Multiple Regression includes several independent variables. Suppose we may include several independent variables to determine the beginning salary Step 1 Analyze

Step 2 Regression

Step 3 Linear

Step 4 Select *current salary* as the dependent variable

Step 5 you can add 'months since hire, education and previous experience.

Please rest assures the method for selecting or rejecting variables in this model. The box labeled Method allows you to select one out of five different options: Enter, Remove, Forward, Backward, and Stepwise. We can use the SPSS default method , Enter which is a standard approach in regression model.

Wrap up

The first table in the output includes information about the quantity of variance that is explained by the predictor variables. The first statistic R is the multiple correlation coefficient between all the predictor variables and the dependent variable. In this model, the value is .90 which indicates that there is a great deal of variance shared by the independent and dependent variables. The next value R^2 is simply the squared value of R. This is frequently used to describe the goodness-of-fit or the amount of variance explained by a given set of predictor variables.

The second table in the output is an ANOVA table that tells us whether the model overall results in a significantly good degree of prediction of the outcome variables. The F statistic tests whether the R^2 proportion of variance in the dependent variable accounted for by the predictor is zero. If the null hypothesis were true, then that would indicate that there is no relationship between the dependent variable and the predictor variables.

4. **DISPLAYING DATA**
    A. Working with tables
    B. Working with graphs

Pivoting Tables

The results from the statistical procedures are displayed in pivotal tables.

- Select the table by double clicking it.
- If the Pivoting Trays window is not visible form the menus choose:
  Pivot
    Pivoting Trays
- If you have only two variables, then select Transpose rows and columns
- If you have more than two variables, select Pivoting trays.
- A Pivoting tray will open with arrow symbols representing the variables
- Click and hold down on one symbol. The variable name will appear.
- Drag the variable to either row or column. If you have more than one variable in the row or column, the order of the symbol does not matter.

You can also change the displaying feature of a table by using **TableLook** feature under the formal pull down menu when you have a table selected. You can also change specific characteristics of your table in Table Properties form the Format pull down menu to add modify *Footnotes*, *Fonts,* Add *colors* and *borders*. Additional options include Alignment, Shading, Foreground and Background colors, and margin sizes.

Exporting table:

You can copy and paste atable into word Document or you can use **Paste Special** feature in Word. You can paste it as formatted text (regular text), a picture (a graphic form which can not be edited) and formatted text (a word table which can be easily edited). You can export your results as HTM (as text or image) files by doing the following:
Go to file and the Export
    Choose the file type as HTM
        Select the object you want to export
            Name the file you wish to give your file to which you are exporting.

Resources:

1. Geroge, D., & Mallery, P. (2005). *SPSS for Windows Step by Step: A Simple Guide and Reference 12.0 Update*. Pearson Education, Inc. USA.
2. Installation and running of software, to access the data files and other related information. Visit the web site http://www.spss.com
3. ITS. Information Technology Services. (2006, February 20). SPSS for Windows: Getting Started. The University of Texas at Austin. TX.
4. Miller, R. L., Acton, C., Fullerton, D. A., & Maltby, J. (2002). *SPSS for Social Scientists.* Covers Versions of 9, 10, & 11). Palgrave Macmillan. NY. USA.
5. Muijs, D. (2004). *Doing Quantitative Research in Education with SPSS*. SAGE Publications. London. Thousand Oaks New Delhi.
6. Norusis, M. J. (2006). SPSS 15.0 Statistical Procedures Companion. Upper Saddler River, NJ: Prentice Hall Inc.
7. SPSS 14.0 Brief Guide (2005). SPSS Inc. Chicago, IL. USA.
8. Urdan, T. C. (2005). *Statistics in Plain English.* (2^{nd}. Ed.). Lawrence Erlbaum Associates, Publishers. New Jersey, USA.