What statistical analysis should I use?

What statistical analysis should I use?

Introduction

For a useful general guide see Policy: Twenty tips for interpreting scientific claims : Nature News & Comment William J. Sutherland, David Spiegelhalter and Mark Burgman Nature Volume: 503, Pages: 335–337 Date published: (21 November 2013).

Some criticism has been made of their discussion of p values, see Replication, statistical consistency, and publication bias G. Francis, Journal of Mathematical Psychology, 57 (5) (2013), pp. 153–169.

Introduction

These examples are loosely based on a UCLA tutorial sheet. All can be realised via the syntax window, when appropriate command strokes are also indicated .

These pages show how to perform a number of statistical tests using SPSS. Each section gives a brief description of the aim of the statistical test, when it is used, an example showing the SPSS commands and SPSS (often abbreviated) output with a brief interpretation of the output.

About the A data file

Most of the examples in this document will use a data file called **A,** high school and beyond. This data file contains 200 observations from a sample of high school students with demographic information about the students, such as their gender (**female**), socio-economic status (**ses**) and ethnic background (**race**). It also contains a number of scores on standardized tests, including tests of reading (**read**), writing (**write**), mathematics (**math**) and social studies (**socst**).

About the A data file

Syntax:-

display dictionary

/VARIABLES id female race ses schtyp prog read write math science socst.

Variable	Position	Label	Value	Label
id	1			
female	2		.00 1.00	Male Female
race	3		1.00 2.00 3.00 4.00	Hispanic Asian african-amer White
ses	4		1.00 2.00 3.00	Low Middle High
schtyp	5	type of school	1.00 2.00	Public private
prog	6	type of program	1.00 2.00 3.00	general academic vocation
read	7	reading score		
write	8	writing score		
math	9	math score		
science	10	science score		
socst	11	social studies score		

About the A data file

13a[1].sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

	id	female	race	ses	schtyp	prog	read	write	math	science	socst
1	70.00	male	white	low	public	general	57.00	52.00	41.00	47.00	57.00
2	121.00	female	white	middle	public	vocation	68.00	59.00	53.00	63.00	61.00
3	86.00	male	white	high	public	general	44.00	33.00	54.00	58.00	31.00
4	141.00	male	white	high	public	vocation	63.00	44.00	47.00	53.00	56.00
5	172.00	male	white	middle	public	academic	47.00	52.00	57.00	53.00	61.00
6	113.00	male	white	middle	public	academic	44.00	52.00	51.00	63.00	61.00
7	50.00	male	african-amer	middle	public	general	50.00	59.00	42.00	53.00	61.00
8	11.00	male	hispanic	middle	public	academic	34.00	46.00	45.00	39.00	36.00
9	84.00	male	white	middle	public	general	63.00	57.00	54.00	58.00	51.00
10	48.00	male	african-amer	middle	public	academic	57.00	55.00	52.00	50.00	51.00
11	75.00	male	white	middle	public	vocation	60.00	46.00	51.00	53.00	61.00
12	60.00	male	white	middle	public	academic	57.00	65.00	51.00	63.00	61.00
13	95.00	male	white	high	public	academic	73.00	60.00	71.00	61.00	71.00
14	104.00	male	white	high	public	academic	54.00	63.00	57.00	55.00	46.00
15	38.00	male	african-amer	low	public	academic	45.00	57.00	50.00	31.00	56.00
16	115.00	male	white	low	public	general	42.00	49.00	43.00	50.00	56.00
17	76.00	male	white	high	public	academic	47.00	52.00	51.00	50.00	56.00
18	195.00	male	white	middle	private	general	57.00	57.00	60.00	58.00	56.00
19	114.00	male	white	high	public	academic	68.00	65.00	62.00	55.00	61.00
20	85.00	male	white	middle	public	general	55.00	39.00	57.00	53.00	46.00

Index End

One sample t-test

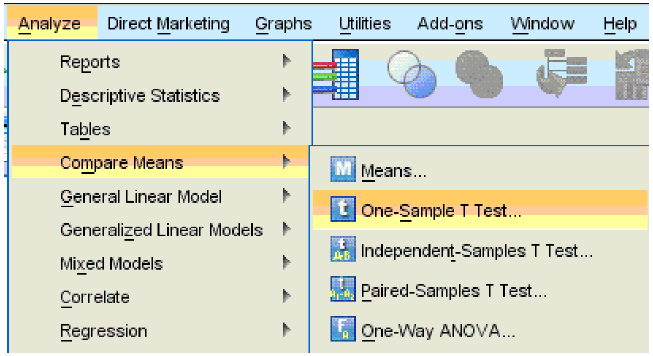
A one sample t-test allows us to test whether a sample mean (of a normally distributed interval variable) significantly differs from a hypothesized value. For example, using the A data file, say we wish to test whether the average writing score (**write**) differs significantly from 50. Test variable writing score (write), Test value 50. We can do this as shown below.

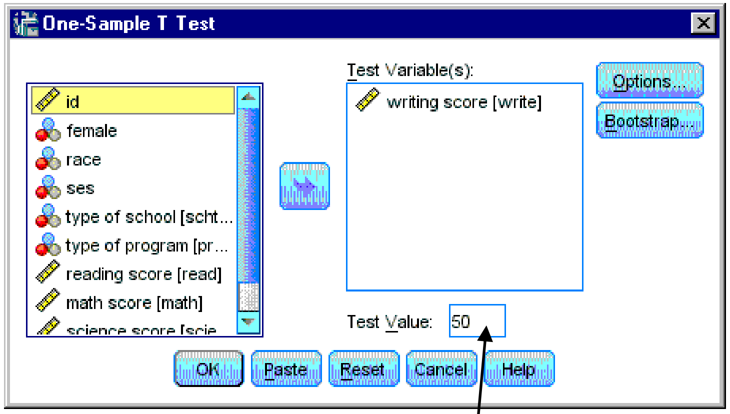
Menu selection:- Analyze > Compare Means > One-Sample T test

Syntax:-

```
t-test
 /testval = 50
 /variable = write.
```

One sample t-test





Note the test value of 50 has been selected

One sample t-test

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
writing score	200	52.7750	9.47859	.67024

One-Sample Test

	Test Value = 50			
	t	df	Sig. (2-tailed)	Mean Difference
writing score	4.140	199	.000	2.77500

One-Sample Test

	Test Value = 50	
	95% Confidence Interval of the Difference	
	Lower	Upper
writing score	1.4533	4.0967

The mean of the variable **write** for this particular sample of students is 52.775, which is statistically significantly different from the test value of 50. We would conclude that this group of students has a significantly higher mean on the writing test than 50. This is consistent with the reported confidence interval (1.45,4.10) that is (51.45,54.10) which excludes 50, of course the mid-point is the mean.

Index End

One sample median test

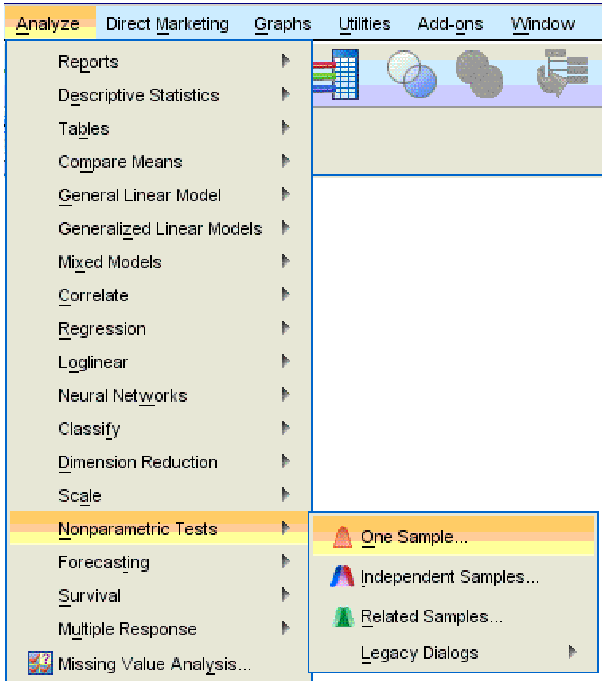
A one sample median test allows us to test whether a sample median differs significantly from a hypothesized value. We will use the same variable, **write**, as we did in the one sample t-test example above, but we do not need to assume that it is interval and normally distributed (we only need to assume that **write** is an ordinal variable).

Menu selection:- Analyze > Nonparametric Tests > One Sample

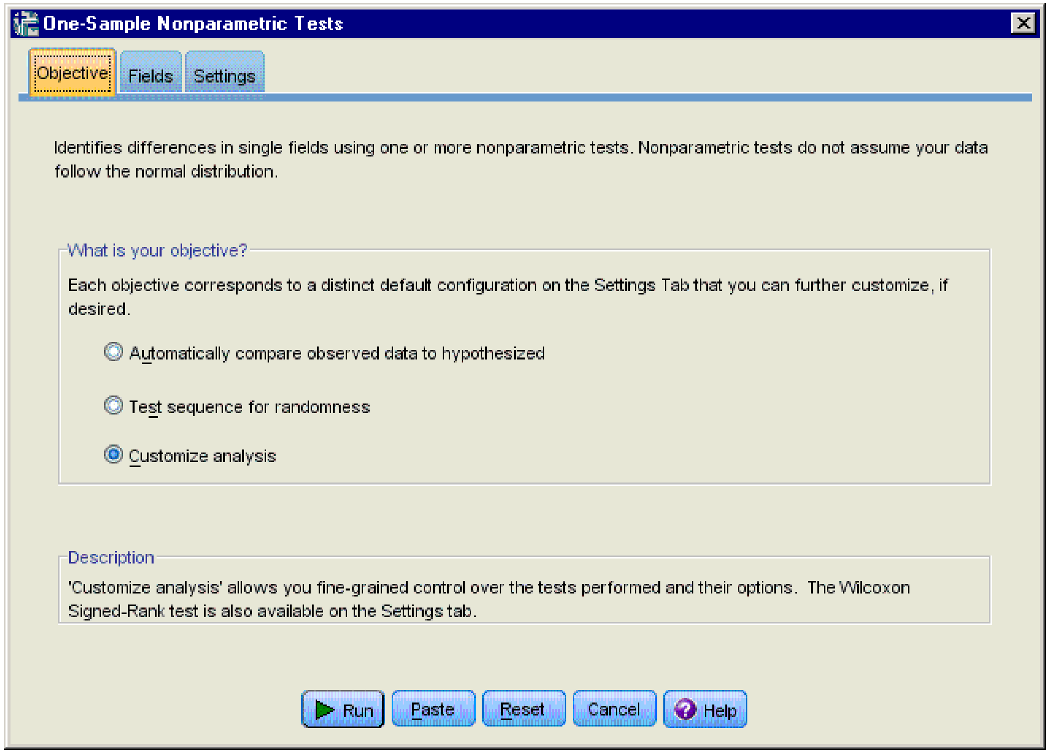
Syntax:-

```
nptests
/onesample test (write) wilcoxon(testvalue = 50).
```

One sample median test

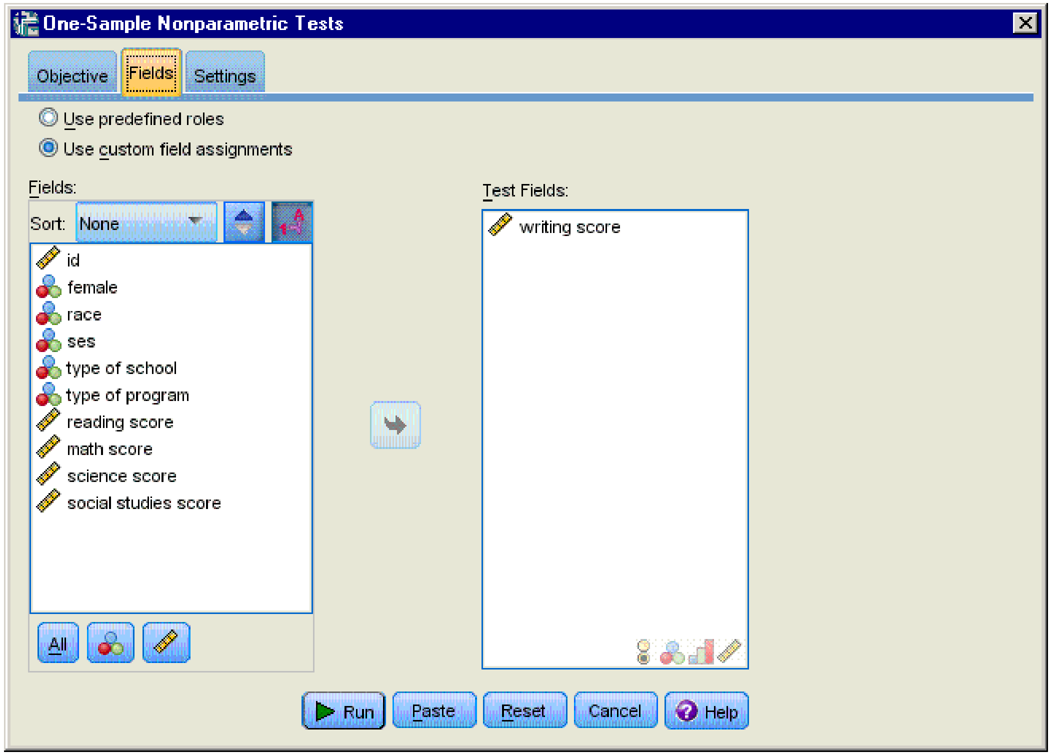


One sample median test



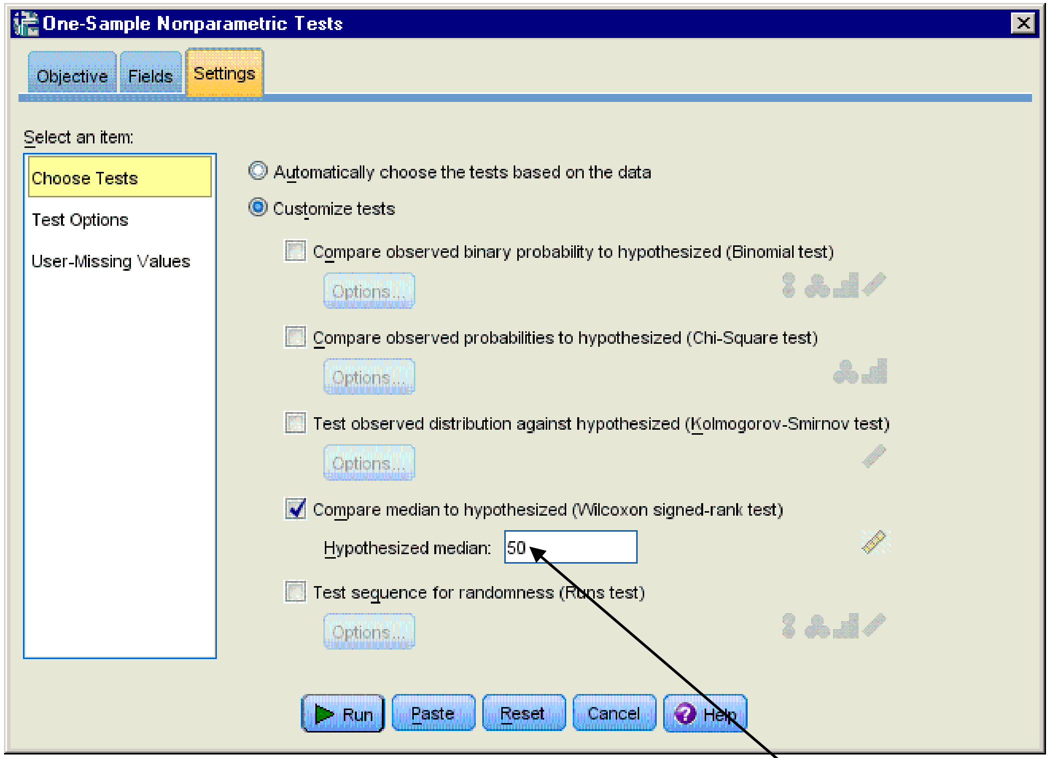
Choose customize analysis

One sample median test



Only retain writing score

One sample median test



Choose tests tick "compare median…" and enter 50 as the desired value.
Finally select the "run" button

One sample median test

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The median of writing score equals 50.00.	One-Sample Wilcoxon Signed Rank Test	.000	Reject the null hypothesis.

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

We would conclude that this group of students has a significantly higher median on the writing test than 50.

Index End

Binomial test

A one sample binomial test allows us to test whether the proportion of successes on a two-level categorical dependent variable significantly differs from a hypothesized value. For example, using the A data file, say we wish to test whether the proportion of females (**female**) differs significantly from 50%, i.e., from .5. We can do this as shown below.

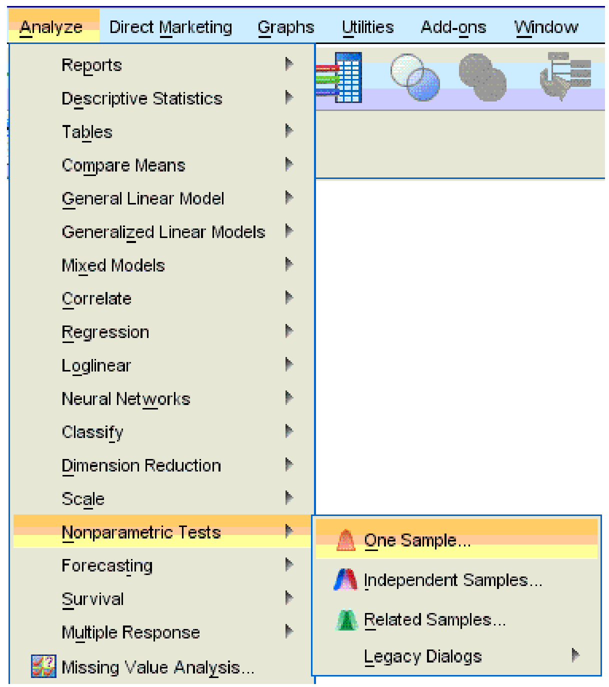
Two alternate approaches are available.

Either
Menu selection:- Analyze > Nonparametric Tests > One Sample

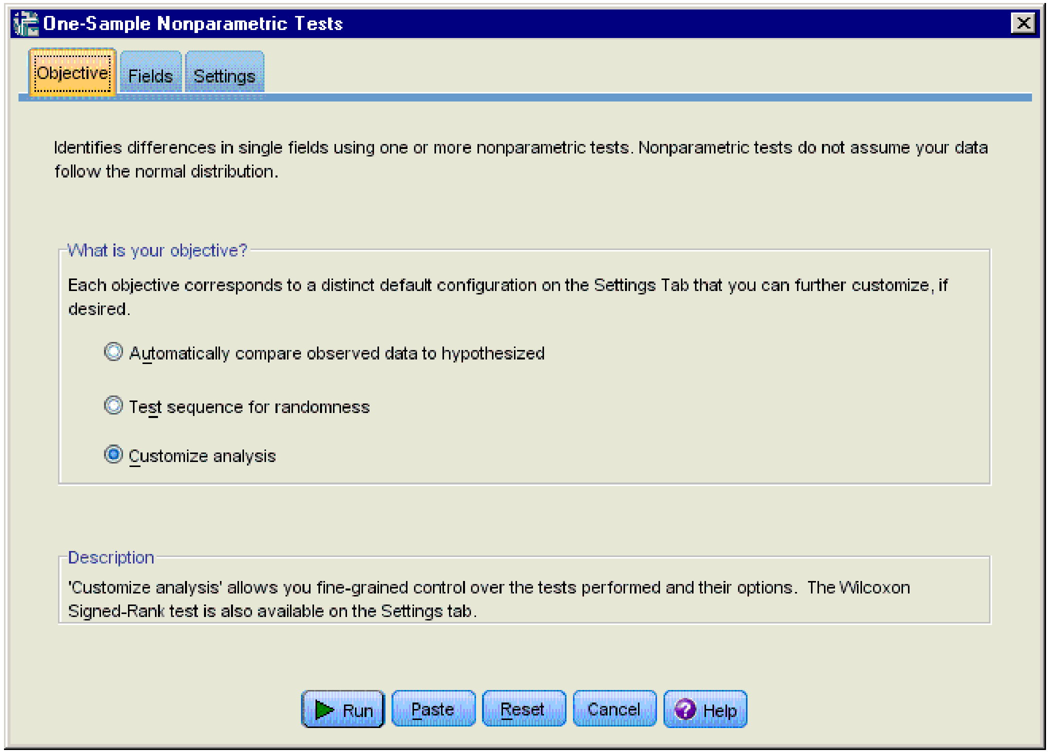
Syntax:-

```
npar tests
/binomial (.5) = female.
```

Binomial test

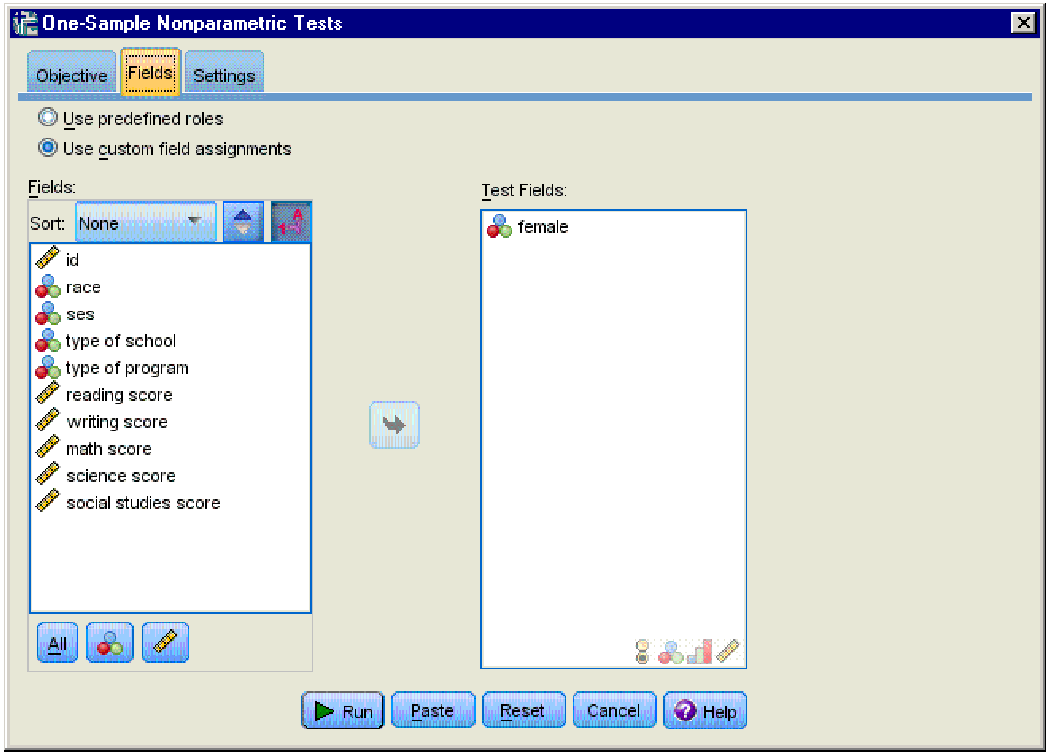


Binomial test



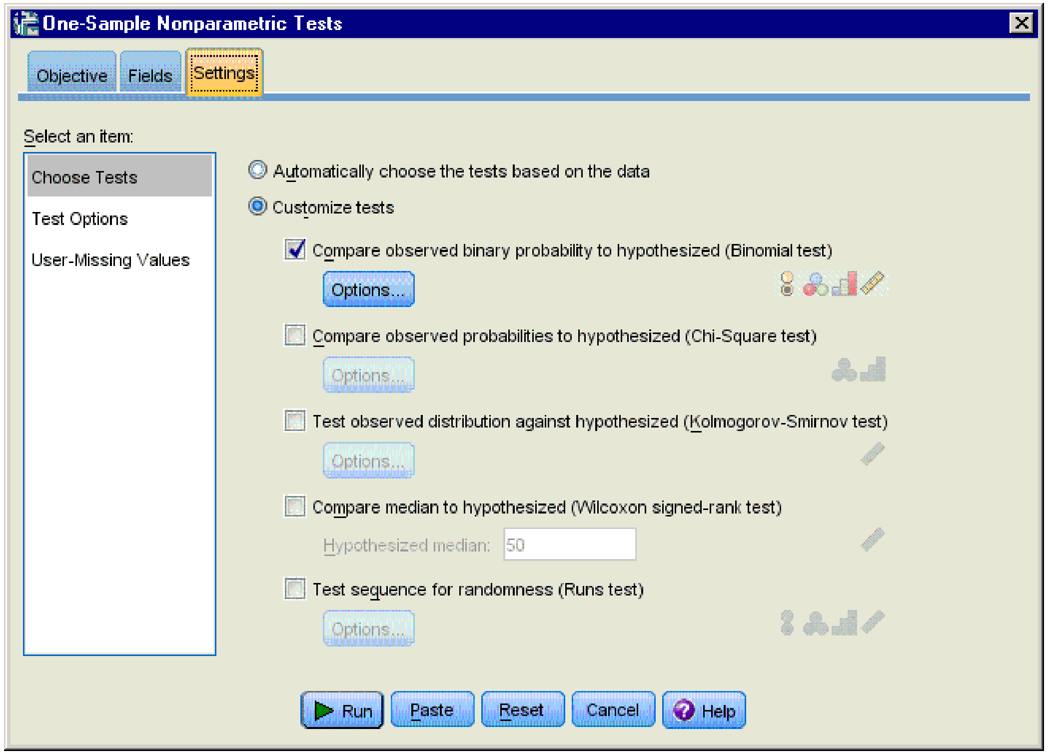
Choose customize analysis

Binomial test



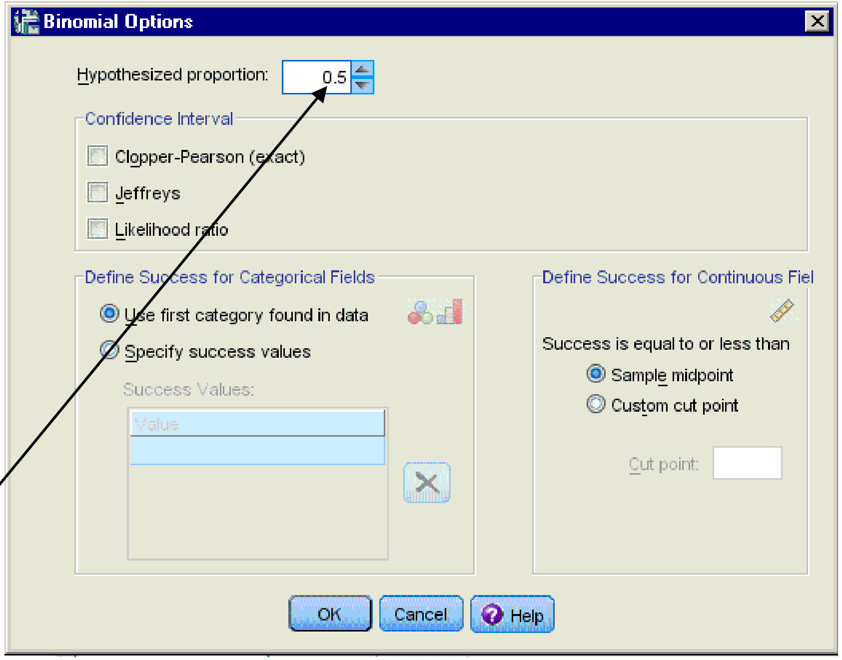
Only retain female

Binomial test



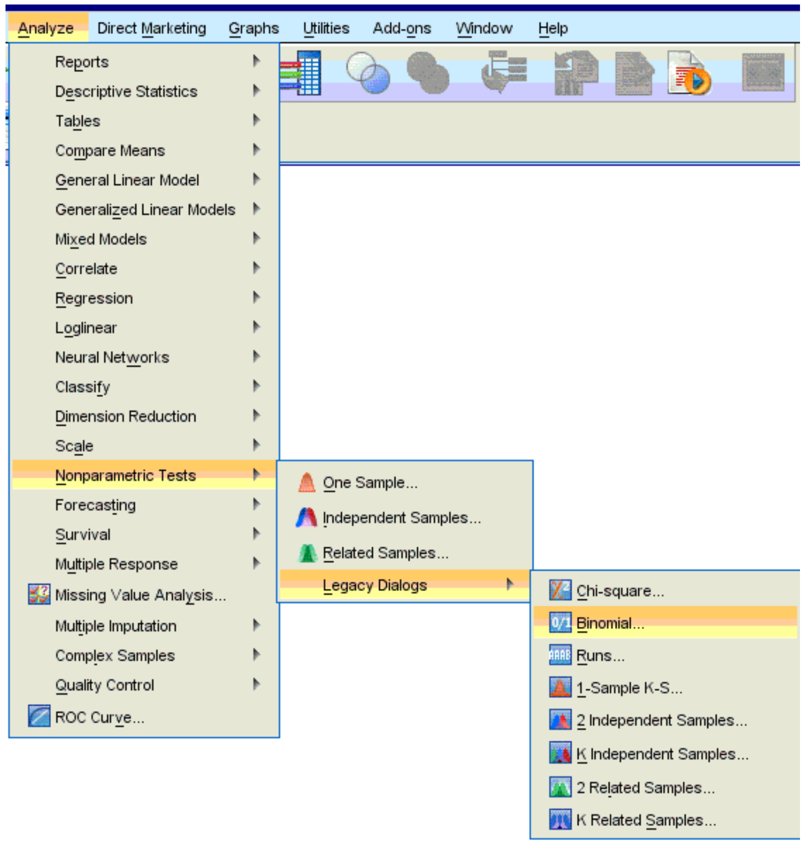
Choose tests tick "compare observed…" and under options

Binomial test



enter .5 as the desired value.
Finally select the “run” button

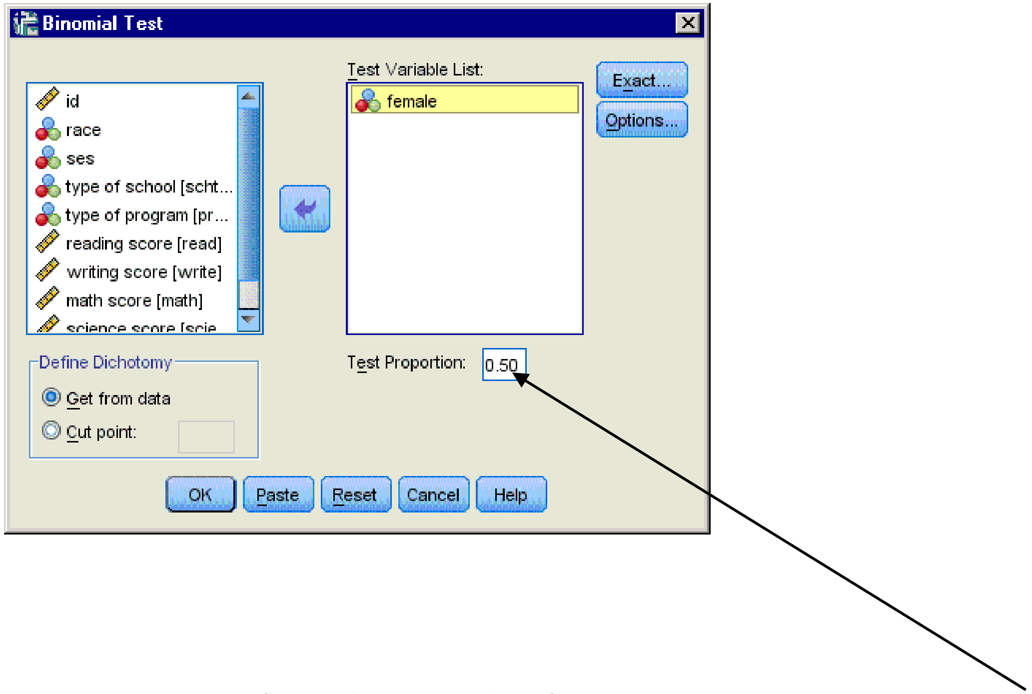
Binomial test



Or
Menu selection:- Analyze > Nonparametric Tests > Legacy Dialogs > Binomial

Syntax:- npar tests
/binomial (.5) = female.

Binomial test



Select female as the test variable, the default test proportion is .5
Finally select the "OK" button

Binomial test

Binomial Test

		Category	N	Observed Prop.	Test Prop.	Exact Sig. (2-tailed)
female	Group 1	Male	91	.46	.50	.229
	Group 2	Female	109	.54		
	Total		200	1.00		

The results indicate that there is no statistically significant difference ($p = 0.229$). In other words, the proportion of females in this sample does not significantly differ from the hypothesized value of 50%.

Index End

Chi-square goodness of fit

A chi-square goodness of fit test allows us to test whether the observed proportions for a categorical variable differ from hypothesized proportions. For example, let's suppose that we believe that the general population consists of 10% Hispanic, 10% Asian, 10% African American and 70% White folks. We want to test whether the observed proportions from our sample differ significantly from these hypothesized proportions. Note this example employs input data (10, 10, 10, 70), in addition to A.

Menu selection:- At present the drop down menu's cannot provide this analysis.

Syntax:-

```
npar test
 /chisquare = race
 /expected = 10 10 10 70.
```

Chi-square goodness of fit

race

	Observed N	Expected N	Residual
hispanic	24	20.0	4.0
asian	11	20.0	-9.0
african-amer	20	20.0	.0
white	145	140.0	5.0
Total	200		

Test Statistics

	race
Chi-Square	5.029[a]
df	3
Asymp. Sig.	.170

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 20.0.

These results show that racial composition in our sample does not differ significantly from the hypothesized values that we supplied (chi-square with three degrees of freedom = 5.029, p = 0.170).

Index End

Two independent samples t-test

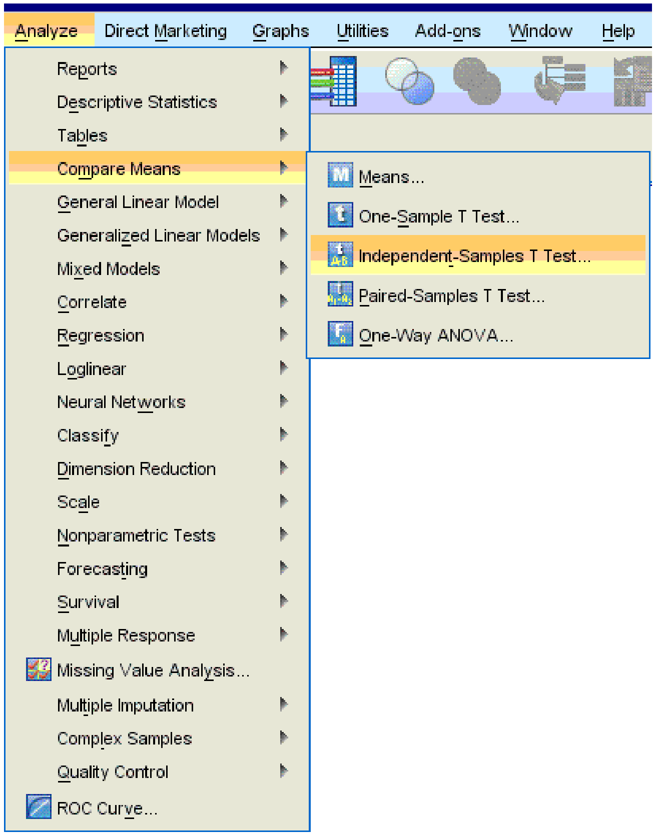
An independent samples t-test is used when you want to compare the means of a normally distributed interval dependent variable for two independent groups. For example, using the A data file, say we wish to test whether the mean for **write** is the same for males and females.

Menu selection:- Analyze > Compare Means > Independent Samples T test

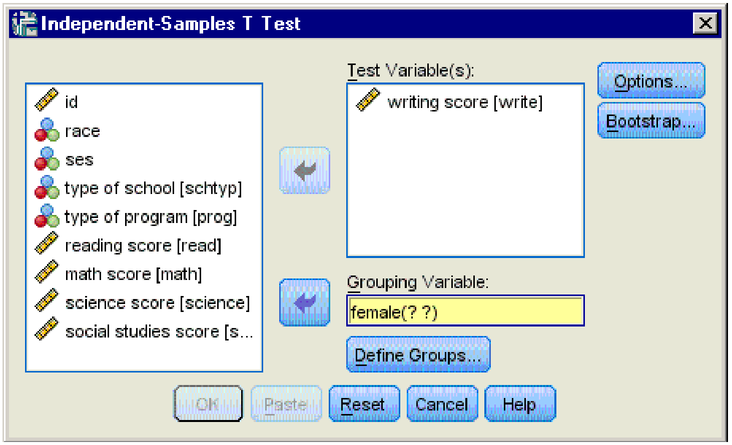
Syntax:-

```
t-test groups = female(0 1)
 /variables = write.
```

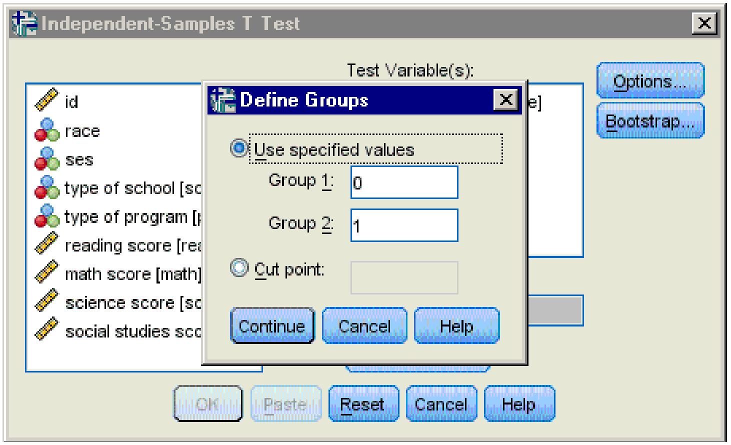
Two independent samples t-test



Two independent samples t-test



Two independent samples t-test



Do not forget to define those "pesky" groups.

Levene's test

In statistics, Levene's test is an inferential statistic used to assess the equality of variances in different samples. Some common statistical procedures assume that variances of the populations from which different samples are drawn are equal. Levene's test assesses this assumption. It tests the null hypothesis that the population variances are equal (called homogeneity of variance or homoscedasticity). If the resulting P-value of Levene's test is less than some critical value (typically 0.05), the obtained differences in sample variances are unlikely to have occurred based on random sampling from a population with equal variances. Thus, the null hypothesis of equal variances is rejected and it is concluded that there is a difference between the variances in the population.

Levene, Howard (1960). "Robust tests for equality of variances". In Ingram Olkin, Harold Hotelling, et al. Stanford University Press. pp. 278–292.

Two independent samples t-test

Group Statistics

	female	N	Mean	Std. Deviation	Std. Error Mean
writing score	male	91	50.1209	10.30516	1.08027
	female	109	54.9908	8.13372	.77907

Independent Samples Test

		Levene's Test for Equality of Variances	
		F	Sig.
writing score	Equal variances assumed	11.133	.001
	Equal variances not assumed		

Independent Samples Test

		t-test for Equality of Means			
		t	df	Sig. (2-tailed)	Mean Difference
writing score	Equal variances assumed	-3.734	198	.000	-4.86995
	Equal variances not assumed	-3.656	169.707	.000	-4.86995

Independent Samples Test

		t-test for Equality of Means		
		Std. Error Difference	95% Confidence Interval of the Difference	
			Lower	Upper
writing score	Equal variances assumed	1.30419	-7.44183	-2.29806
	Equal variances not assumed	1.33189	-7.49916	-2.24073

Because the standard deviations for the two groups are not similar (10.3 and 8.1), we will use the "equal variances not assumed" test. This is supported by the Levene's test p = .001).

The results indicate that there is a statistically significant difference between the mean writing score for males and females ($t = -3.656$, $p < .0005$). In other words, females have a statistically significantly higher mean score on writing (54.99) than males (50.12).

This is supported by the negative confidence interval (male - female).

Two independent samples t-test

Group Statistics

	female	N	Mean	Std. Deviation	Std. Error Mean
writing score	male	91	50.1209	10.30516	1.08027
	female	109	54.9908	8.13372	.77907

Independent Samples Test

		Levene's Test for Equality of Variances	
		F	Sig.
writing score	Equal variances assumed	11.133	.001
	Equal variances not assumed		

Independent Samples Test

		t-test for Equality of Means			
		t	df	Sig. (2-tailed)	Mean Difference
writing score	Equal variances assumed	-3.734	198	.000	-4.86995
	Equal variances not assumed	-3.656	169.707	.000	-4.86995

Independent Samples Test

		t-test for Equality of Means		
		Std. Error Difference	95% Confidence Interval of the Difference	
			Lower	Upper
writing score	Equal variances assumed	1.30419	-7.44183	-2.29806
	Equal variances not assumed	1.33189	-7.49916	-2.24073

Does equality of variances matter in this case?

Index End

Wilcoxon-Mann-Whitney test

The Wilcoxon-Mann-Whitney test is a non-parametric analog to the independent samples t-test and can be used when you do not assume that the dependent variable is a normally distributed interval variable (you only assume that the variable is at least ordinal). You will notice that the SPSS syntax for the Wilcoxon-Mann-Whitney test is almost identical to that of the independent samples t-test. We will use the same data file (the A data file) and the same variables in this example as we did in the independent t-test example above and will not assume that write, our dependent variable, is normally distributed.

Menu selection:- Analyze > Nonparametric Tests > Legacy Dialogs > 2 Independent Samples

Syntax:-

```
npar test
/m-w = write by female(0 1).
```

Wilcoxon-Mann-Whitney test

The Mann-Whitney U: A Test for Assessing Whether Two Independent Samples Come from the Same Distribution
Nadim Nachar
Tutorials in Quantitative Methods for Psychology 2008 **4(1)** 13-20

Wilcoxon-Mann-Whitney test

The Wilcoxon-Mann-Whitney test is sometimes used for comparing the efficacy of two treatments in trials. It is often presented as an alternative to a t test when the data are not normally distributed.

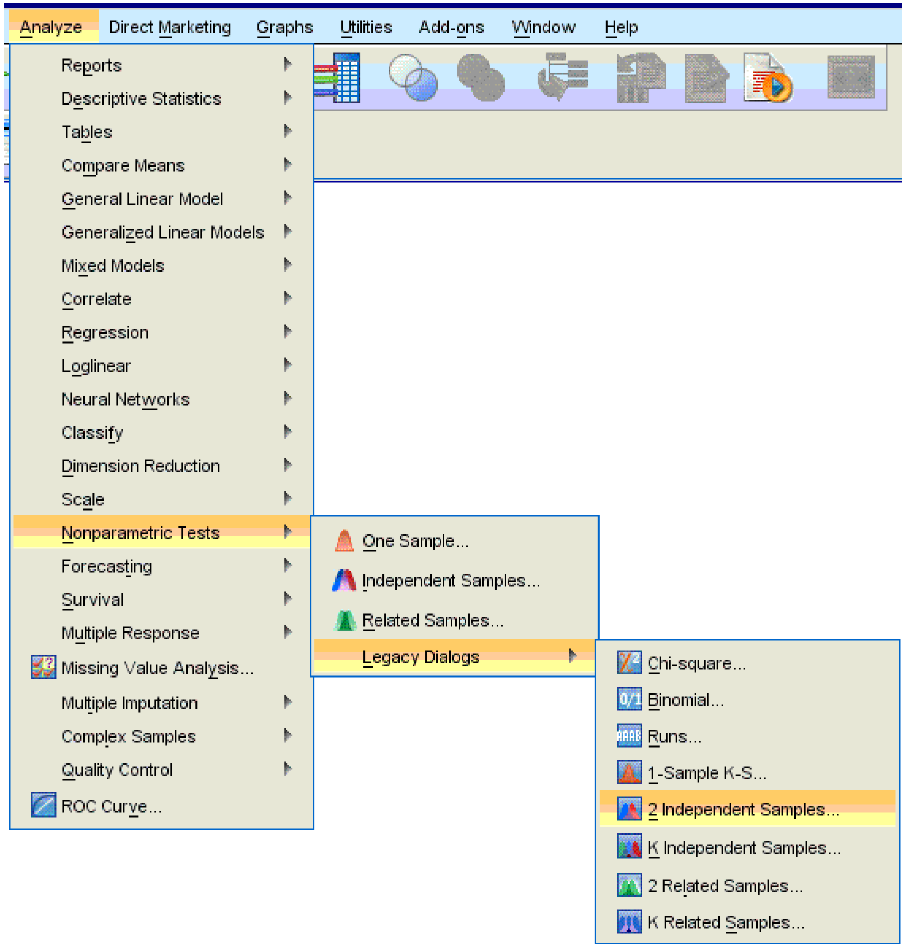
Whereas a t test is a test of population means, the Mann-Whitney test is commonly regarded as a test of population medians. This is not strictly true, and treating it as such can lead to inadequate analysis of data.

Mann-Whitney test is not just a test of medians: differences in spread can be important
Anna Hart
British Medical Journal 2001 August 18; **323(7309)**: 391–393.

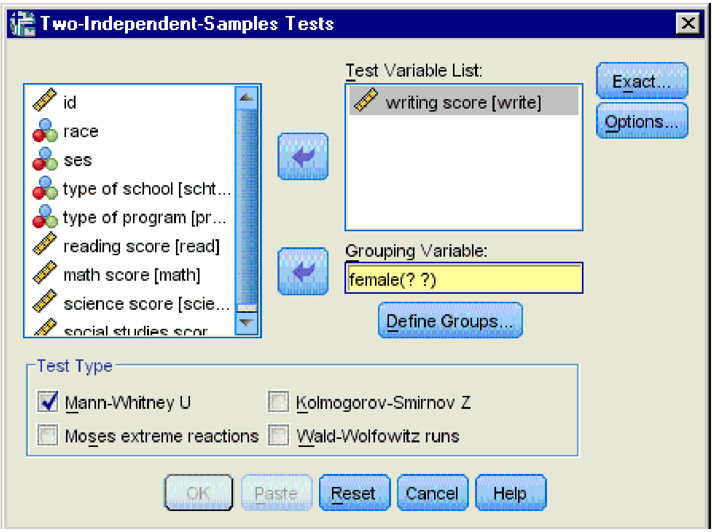
Paper

As is always the case, it is not sufficient merely to report a P value. In the case of the Mann-Whitney test, differences in spread may sometimes be as important as differences in medians, and these need to be made clear.

Wilcoxon-Mann-Whitney test

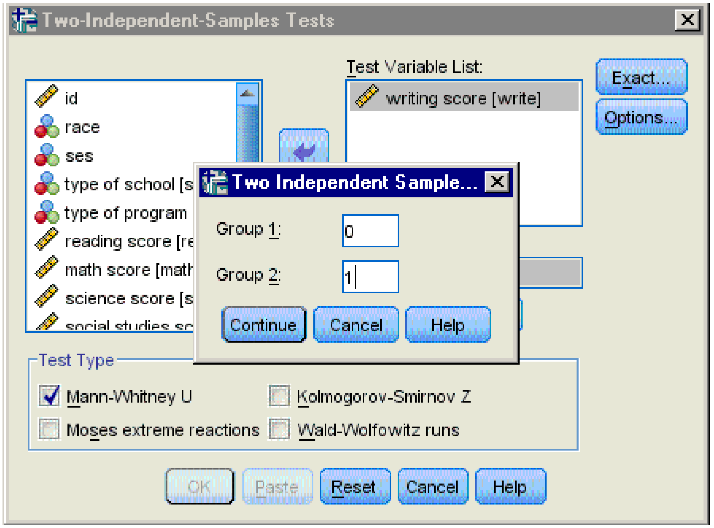


Wilcoxon-Mann-Whitney test



Note that Mann-Whitney has been selected.

Wilcoxon-Mann-Whitney test



Wilcoxon-Mann-Whitney test

Ranks

	female	N	Mean Rank	Sum of Ranks
writing score	male	91	85.63	7792.00
	female	109	112.92	12308.00
	Total	200		

Test Statistics[a]

	writing score
Mann-Whitney U	3606.000
Wilcoxon W	7792.000
Z	-3.329
Asymp. Sig. (2-tailed)	.001

a. Grouping Variable: female

The results suggest that there is a statistically significant difference between the underlying distributions of the **write** scores of males and the **write** scores of females (z = -3.329, p = 0.001).

Index End

Chi-square test (Contingency table)

A chi-square test is used when you want to see if there is a relationship between two categorical variables. In SPSS, the **chisq** option is used on the **statistics** subcommand of the **crosstabs** command to obtain the test statistic and its associated p-value. Using the A data file, let's see if there is a relationship between the type of school attended (**schtyp**) and students' gender (**female**). Remember that the chi-square test assumes that the expected value for each cell is five or higher. This assumption is easily met in the examples below. However, if this assumption is not met in your data, please see the section on Fisher's exact test, below.

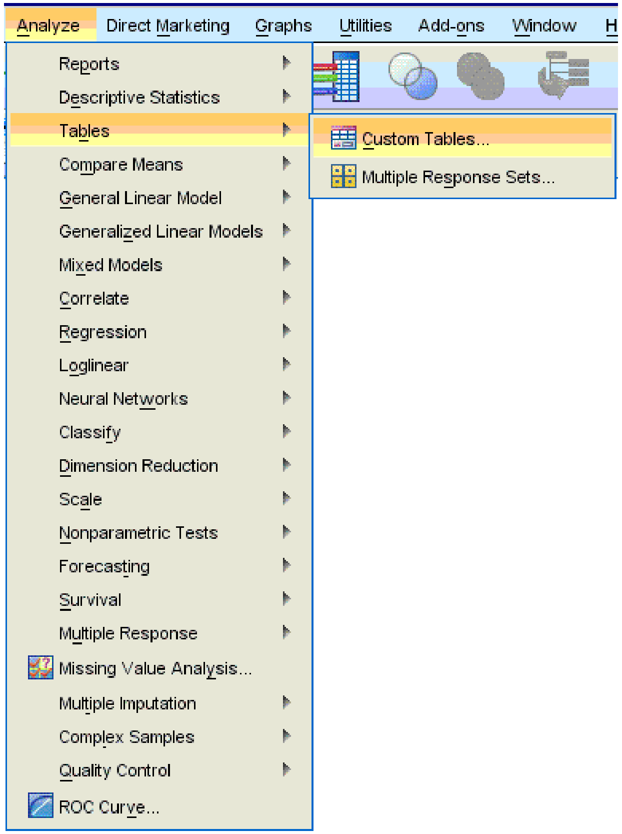
Two alternate approaches are available.

Either
Menu selection:- Analyze > Tables > Custom Tables

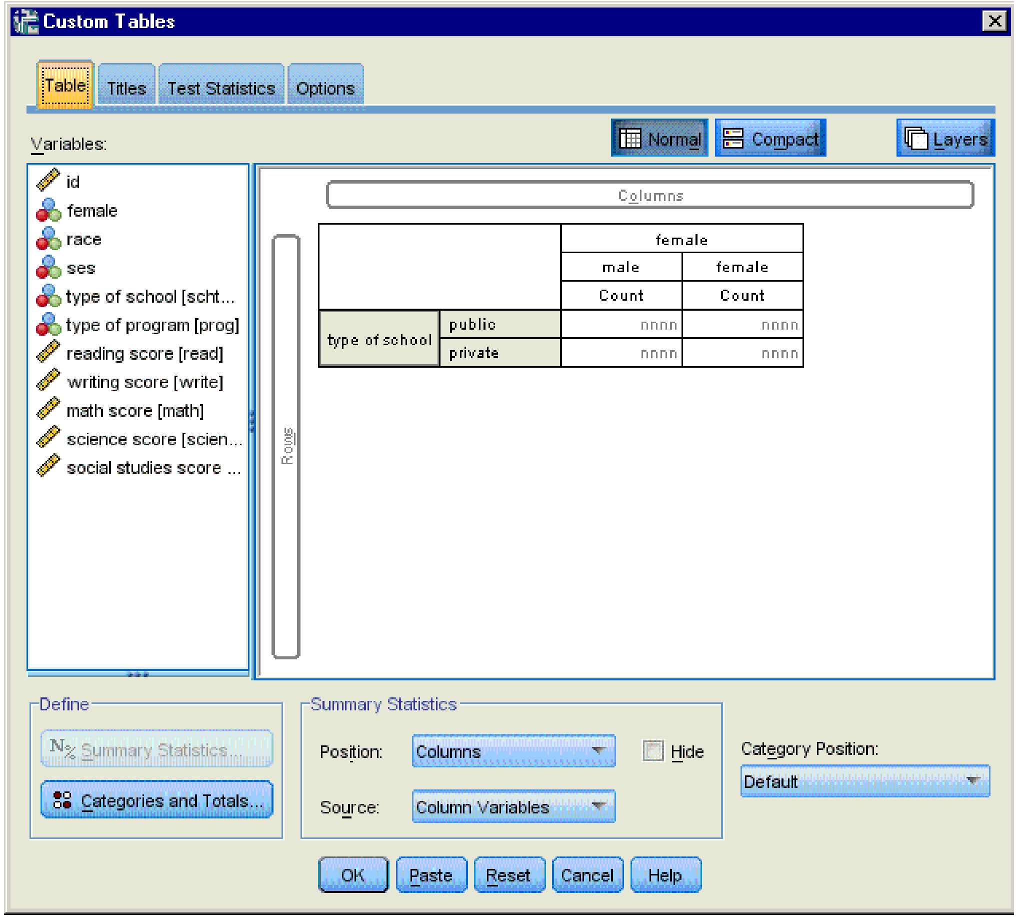
Syntax:-

```
crosstabs
/tables = schtyp by female
 /statistic = chisq.
```

Chi-square test

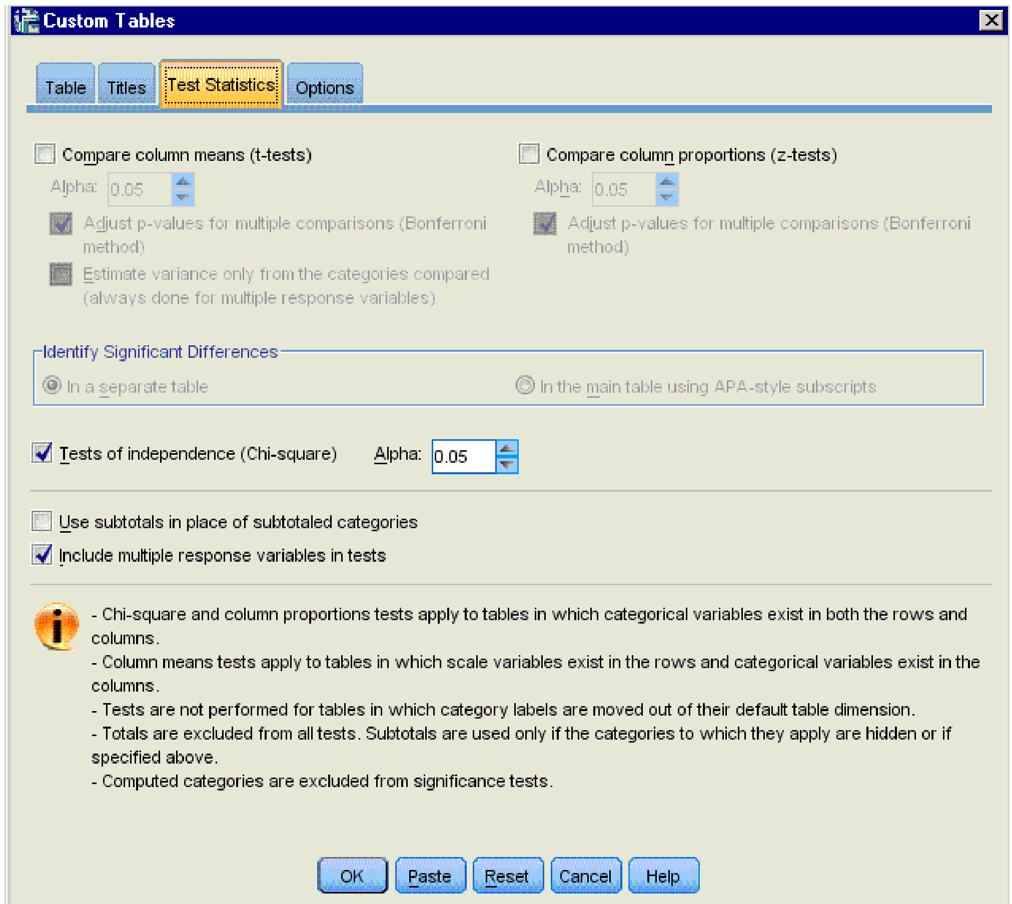


Chi-square test

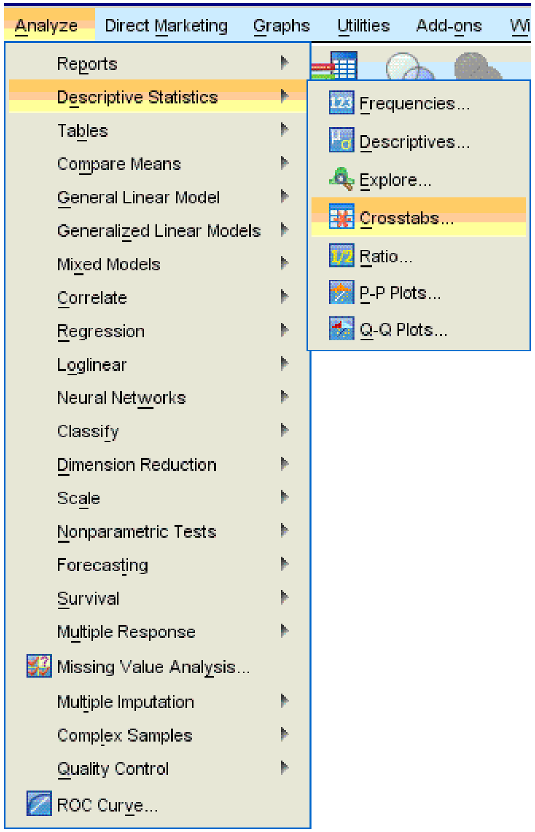


Drag selected variables to the row/column boxes

Chi-square test



Chi-square test



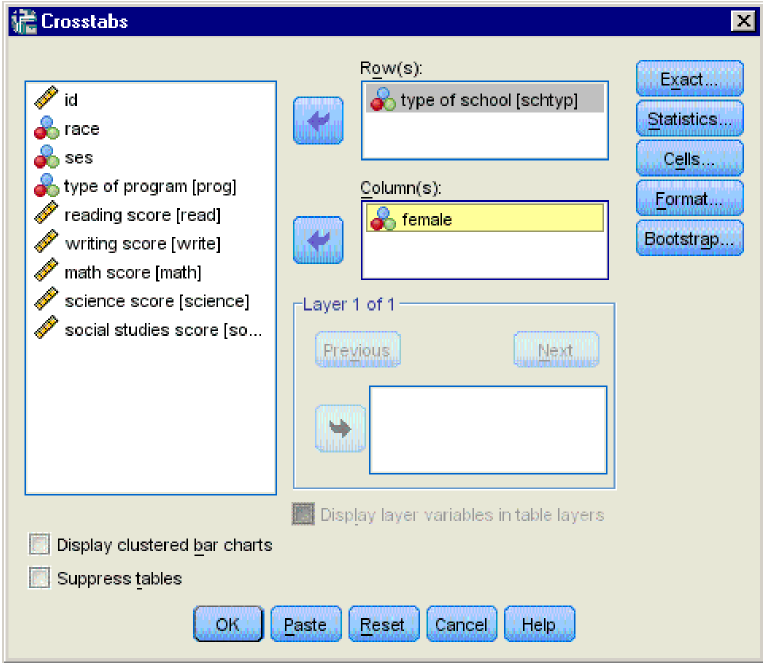
Or

Menu selection:- Analyze > Descriptive Statistics > Crosstabs

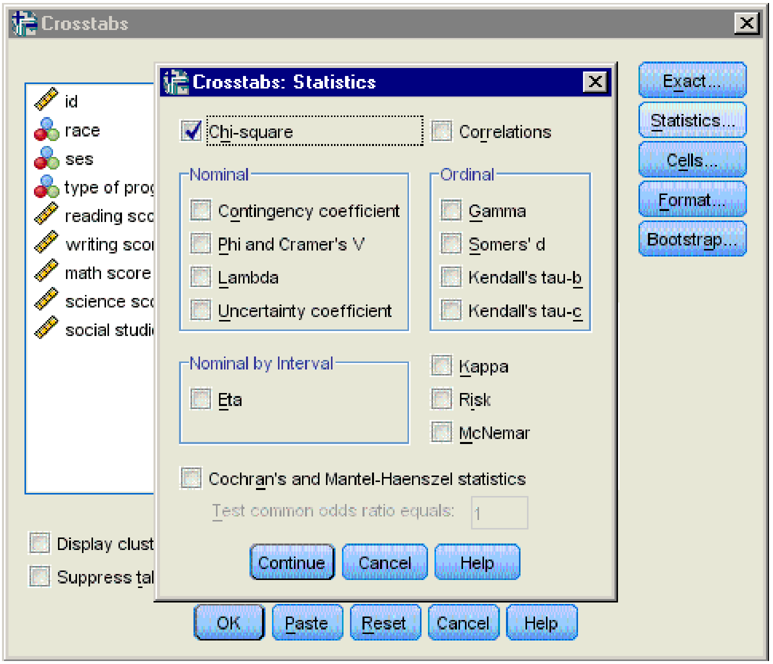
Syntax:-

```
crosstabs
/tables = schtyp by female
 /statistic = chisq.
```

Chi-square test



Chi-square test



Chi-square test

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
type of school * female	200	100.0%	0	.0%	200	100.0%

type of school * female Crosstabulation

Count

		Female		Total
		Male	female	
type of school	public	77	91	168
	private	14	18	32
Total		91	109	200

Chi-Square Tests

	Value	Df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.047[a]	1	.828		
Continuity Correction[b]	.001	1	.981		
Likelihood Ratio	.047	1	.828		
Fisher's Exact Test				.849	.492
Linear-by-Linear Association	.047	1	.829		
N of Valid Cases	200				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 14.56.

b. Computed only for a 2x2 table

These results indicate that there is no statistically significant relationship between the type of school attended and gender (chi-square with one degree of freedom = 0.047, p = 0.828).

Note 0 cells have expected count less than 5. If not use Fisher's exact test.

Chi-square test

Let's look at another example, this time looking at the relationship between gender (**female**) and socio-economic status (**ses**). The point of this example is that one (or both) variables may have more than two levels, and that the variables do not have to have the same number of levels. In this example, **female** has two levels (male and female) and **ses** has three levels (low, medium and high).

Menu selection:- Analyze > Tables > Custom Tables
Using the previous menu's.

Syntax:-

```
crosstabs
 /tables = female by ses
 /statistic = chisq.
```

Chi-square test

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
female * ses	200	100.0%	0	.0%	200	100.0%

female * ses Crosstabulation

Count

		ses			Total
		low	middle	high	
female	male	15	47	29	91
	female	32	48	29	109
Total		47	95	58	200

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	4.577[a]	2	.101
Likelihood Ratio	4.679	2	.096
Linear-by-Linear Association	3.110	1	.078
N of Valid Cases	200		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 21.39.

Again we find that there is no statistically significant relationship between the variables (chi-square with two degrees of freedom = 4.577, p = 0.101).

Note the absence of Fisher's Exact Test!

Index End

Fisher's exact test

The Fisher's exact test is used when you want to conduct a chi-square test but one or more of your cells has an expected frequency of five or less. Remember that the chi-square test assumes that each cell has an expected frequency of five or more, but the Fisher's exact test has no such assumption and can be used regardless of how small the expected frequency is. In SPSS you can only perform a Fisher's exact test on a 2x2 table, and these results are presented by default. Please see the results from the chi squared example above.

Chi-square test

Fisher's exact test

A simple web search should reveal specific tools developed for different size tables. For example

Fisher's exact test for up to 6×6 tables
For the more adventurous

For those interested in more detail, plus a worked example see.

Fisher's Exact Test or Paper only

When to Use Fisher's Exact Test
Keith M. Bower
American Society for Quality, Six Sigma Forum Magazine, **2(4)** 2003, 35-37.

Fisher's exact test

For larger examples you might try

Fisher's Exact Test

Algorithm 643

FEXACT - A Fortran Subroutine For Fisher's Exact Test On Unordered R x C Contingency-Tables
Mehta, C.R. and Patel, N.R.
ACM Transactions On Mathematical Software **12(2)** 154-161 1986.

A Remark On Algorithm-643 - FEXACT - An Algorithm For Performing Fisher's Exact Test In R x C Contingency-Tables
Clarkson, D.B., Fan, Y.A. and Joe, H.
ACM Transactions On Mathematical Software **19(4)** 484-488 1993.

Index End

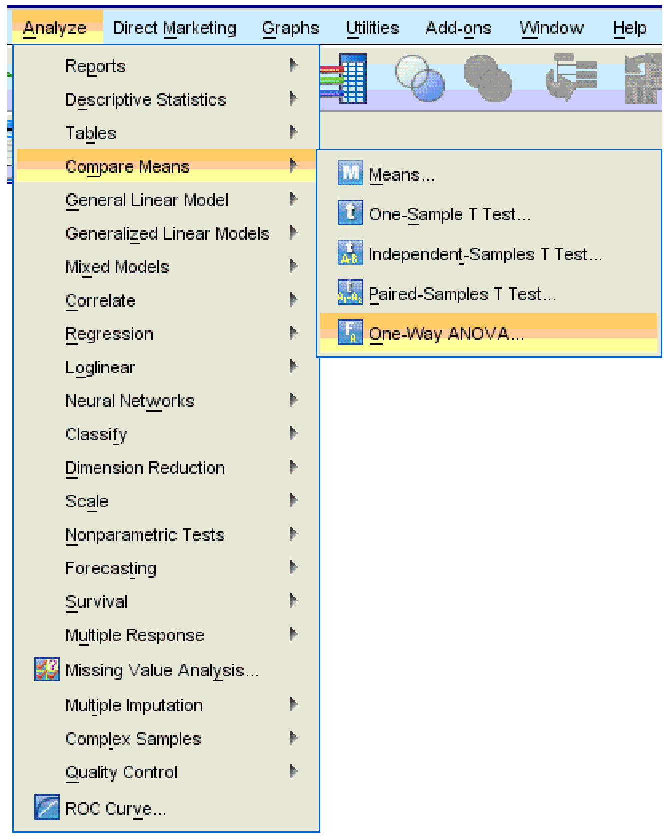
One-way ANOVA

A one-way analysis of variance (ANOVA) is used when you have a categorical independent variable (with two or more categories) and a normally distributed interval dependent variable and you wish to test for differences in the means of the dependent variable broken down by the levels of the independent variable. For example, using the A data file, say we wish to test whether the mean of **write** differs between the three program types (**prog**). The command for this test would be:

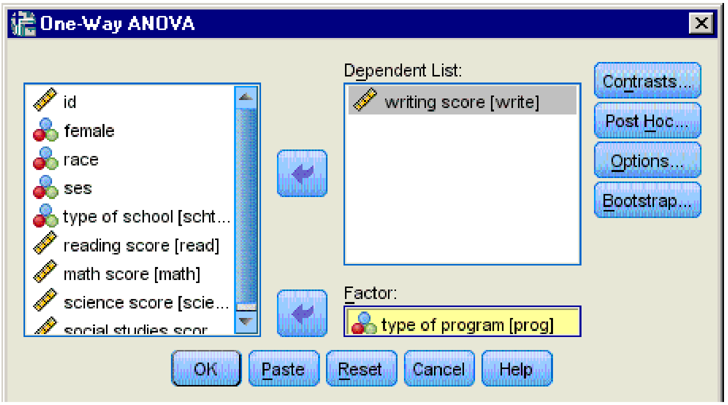
Menu selection:- Analyze > Compare Means > One-way ANOVA

Syntax:- oneway write by prog.

One-way ANOVA



One-way ANOVA



One-way ANOVA

ANOVA

writing score

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	3175.698	2	1587.849	21.275	.000
Within Groups	14703.177	197	74.635		
Total	17878.875	199			

The mean of the dependent variable differs significantly among the levels of program type. However, we do not know if the difference is between only two of the levels or all three of the levels.

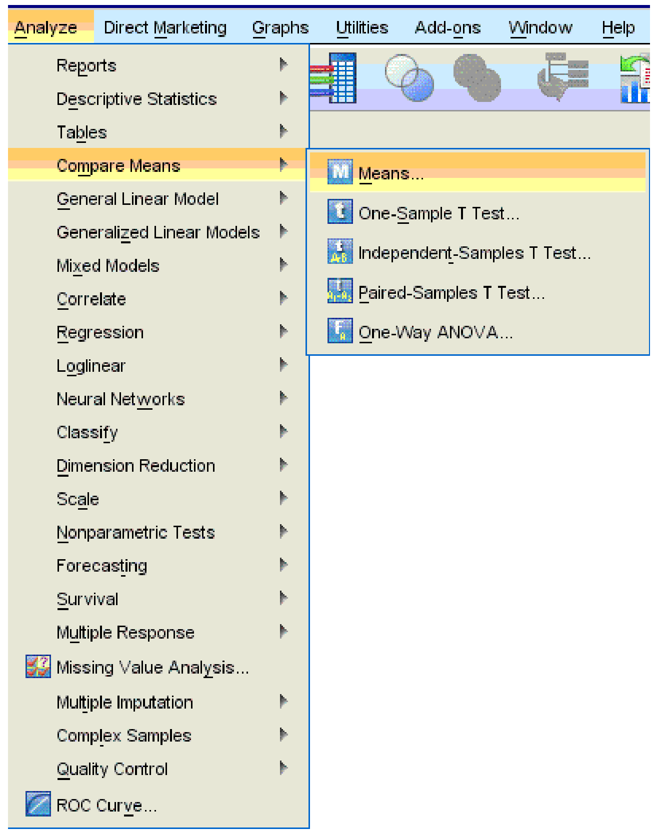
One-way ANOVA

To see the mean of **write** for each level of program type,

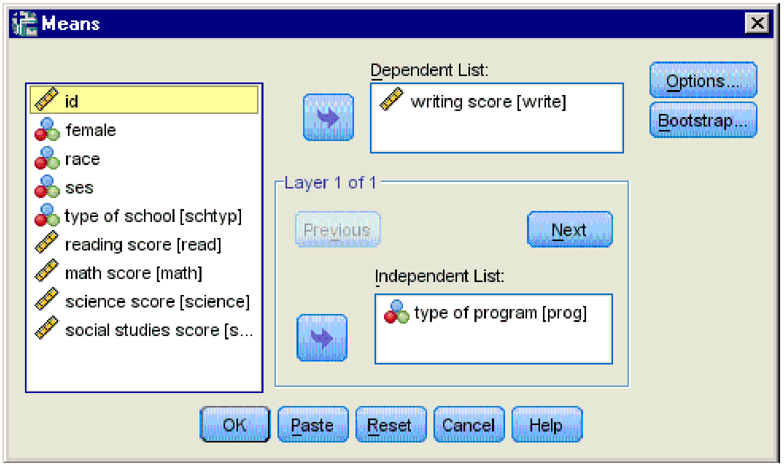
Menu selection:- Analyze > Compare Means > Means

Syntax:- means tables = write by prog.

One-way ANOVA



One-way ANOVA



One-way ANOVA

Case Processing Summary

	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
writing score * type of program	200	100.0%	0	.0%	200	100.0%

Report

writing score

type of program	Mean	N	Std. Deviation
general	51.3333	45	9.39778
academic	56.2571	105	7.94334
vocation	46.7600	50	9.31875
Total	52.7750	200	9.47859

From this we can see that the students in the academic program have the highest mean writing score, while students in the vocational program have the lowest. For a more detailed analysis refer to Bonferroni for pairwise comparisons .

Index End

Kruskal Wallis test

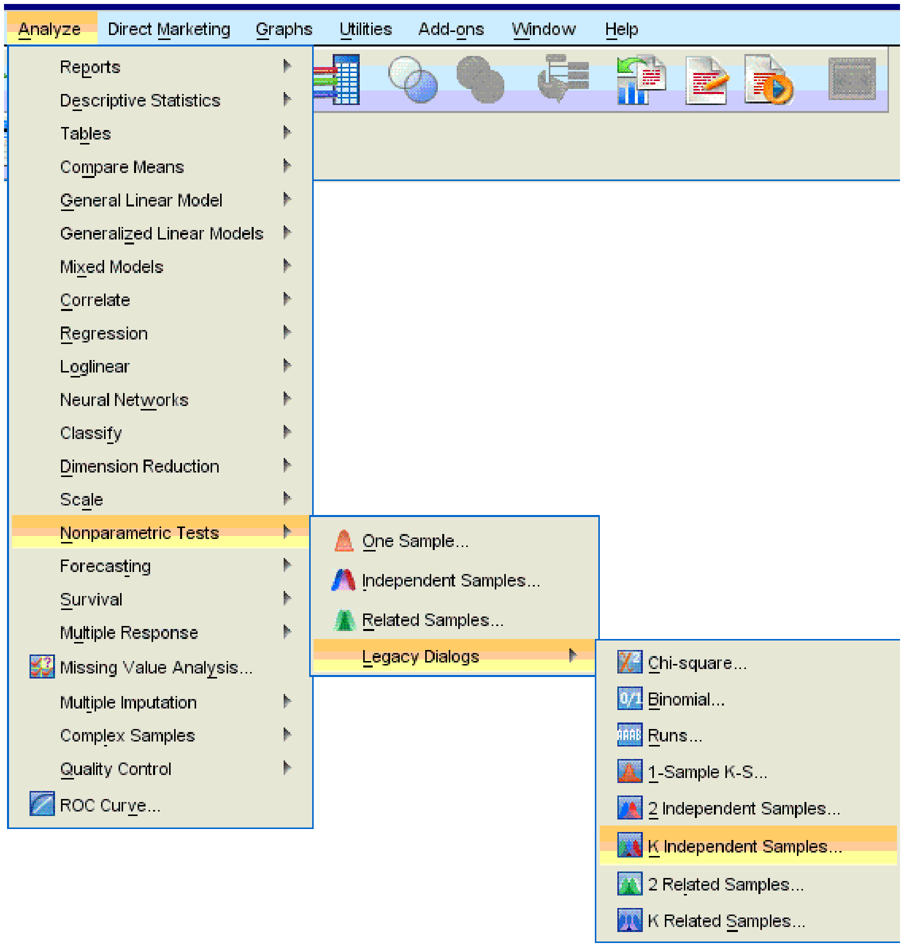
The Kruskal Wallis test is used when you have one independent variable with two or more levels and an ordinal dependent variable. In other words, it is the non-parametric version of ANOVA and a generalized form of the Mann-Whitney test method since it permits two or more groups. We will use the same data file as the one way ANOVA example above (the A data file) and the same variables as in the example above, but we will not assume that **write** is a normally distributed interval variable.

Menu selection:- Analyze > Nonparametric Tests > Legacy Dialogs > k Independent Samples

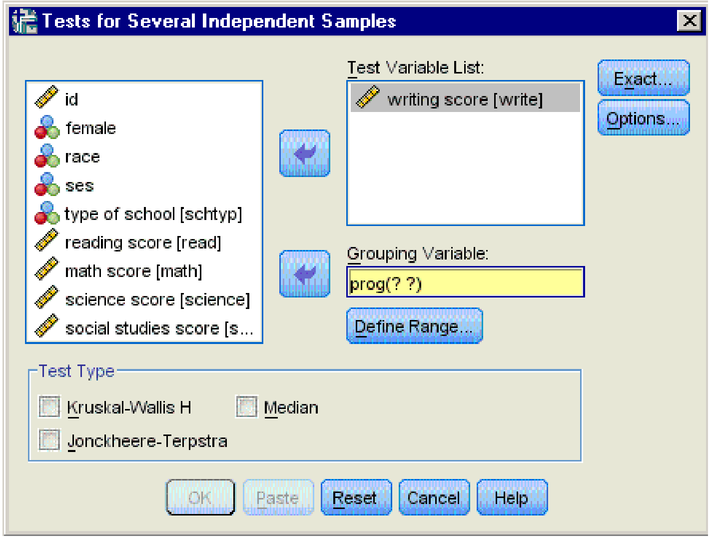
Syntax:-

```
npar tests
 /k-w = write by prog (1,3).
```

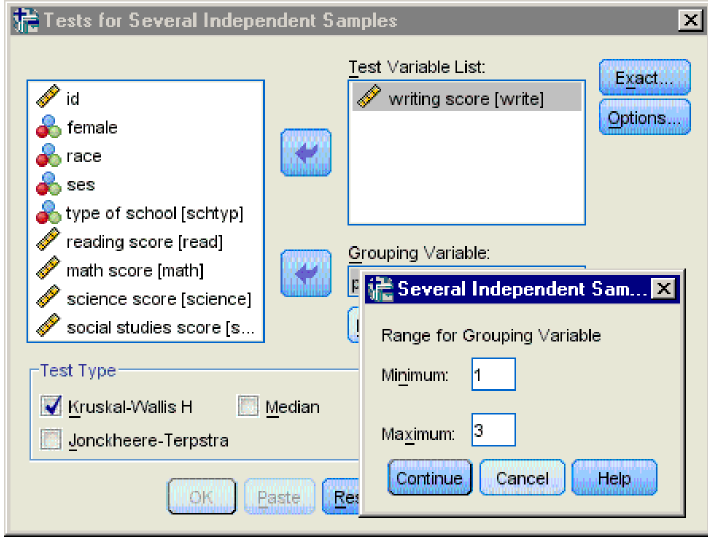
Kruskal Wallis test



Kruskal Wallis test



Kruskal Wallis test



Kruskal Wallis test

Ranks

	type of program	N	Mean Rank
writing score	general	45	90.64
	academic	105	121.56
	vocation	50	65.14
	Total	200	

Test Statistics[a,b]

	writing score
Chi-Square	34.045
df	2
Asymp. Sig.	.000

a. Kruskal Wallis Test
b. Grouping Variable: type of program

If some of the scores receive tied ranks, then a correction factor is used, yielding a slightly different value of chi-squared. With or without ties, the results indicate that there is a statistically significant difference ($p < .0005$) among the three type of programs.

Index End

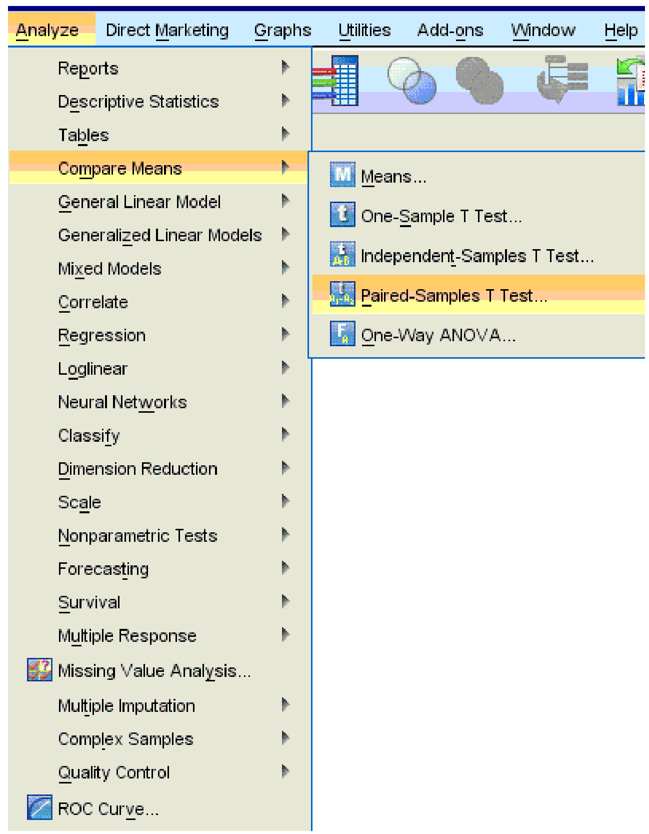
Paired t-test

A paired (samples) t-test is used when you have two related observations (i.e., two observations per subject) and you want to see if the means on these two normally distributed interval variables differ from one another. For example, using the A data file we will test whether the mean of **read** is equal to the mean of **write**.

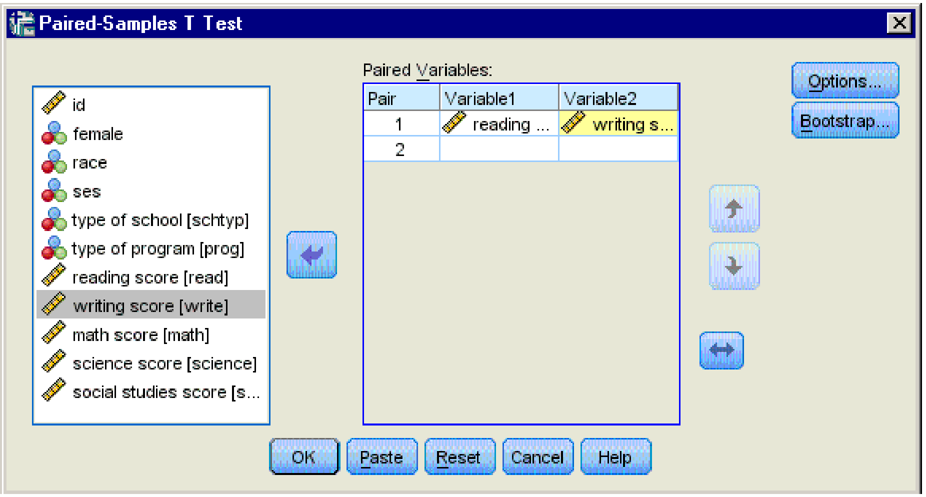
Menu selection:- Analyze > Compare Means > Paired-Samples T test

Syntax:- t-test pairs = read with write (paired).

Paired t-test



Paired t-test



Paired t-test

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	reading score	52.2300	200	10.25294	.72499
	writing score	52.7750	200	9.47859	.67024

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	reading score & writing score	200	.597	.000

Paired Samples Test

		Paired Differences		
		Mean	Std. Deviation	Std. Error Mean
Pair 1	reading score - writing score	-.54500	8.88667	.62838

Paired Samples Test

		Paired Differences: 95% Confidence Interval of the Difference		t
		Lower	Upper	
Pair 1	reading score - writing score	-1.78414	.69414	-.867

Paired Samples Test

		df	Sig. (2-tailed)
Pair 1	reading score - writing score	199	.387

These results indicate that the mean of **read** is not statistically significantly different from the mean of **write** ($t = -0.867$, $p = 0.387$).

The confidence interval includes the origin (no difference).

Index End

Wilcoxon signed rank sum test

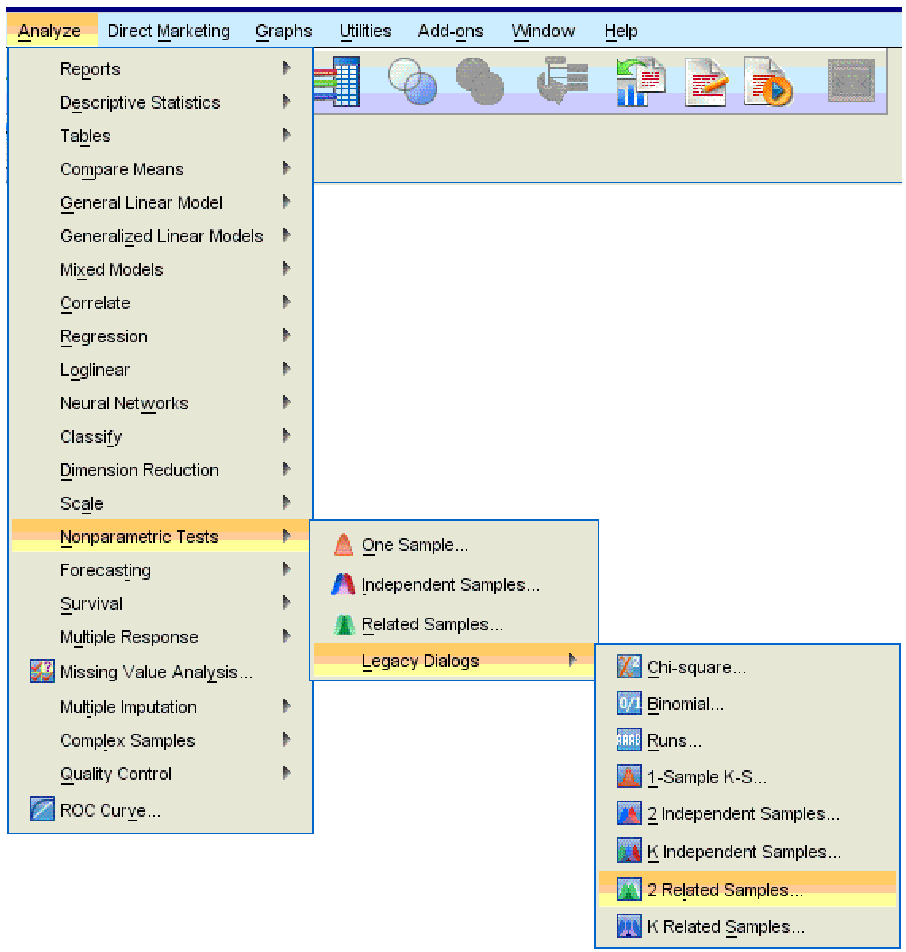
The Wilcoxon signed rank sum test is the non-parametric version of a paired samples t-test. You use the Wilcoxon signed rank sum test when you do not wish to assume that the difference between the two variables is interval and normally distributed (but you do assume the difference is ordinal). We will use the same example as above, but we will not assume that the difference between **read** and **write** is interval and normally distributed.

Menu selection:- Analyze > Nonparametric Tests > Legacy Dialogs > 2 Related Samples

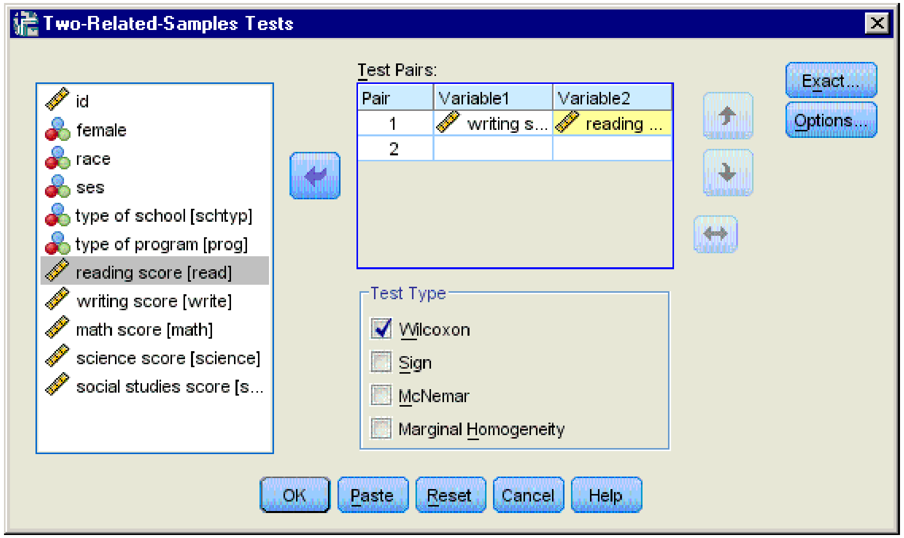
Syntax:-

```
npar test
 /wilcoxon = write with read (paired).
```

Wilcoxon signed rank sum test



Wilcoxon signed rank sum test



Wilcoxon signed rank sum test

Ranks

		N	Mean Rank	Sum of Ranks
reading score - writing score	Negative Ranks	97[a]	95.47	9261.00
	Positive Ranks	88[b]	90.27	7944.00
	Ties	15[c]		
	Total	200		

a. reading score < writing score

b. reading score > writing score

c. reading score = writing score

Test Statistics[b]

	reading score - writing score
Z	-.903[a]
Asymp. Sig. (2-tailed)	.366

a. Based on positive ranks.

b. Wilcoxon Signed Ranks Test

The results suggest that there is not a statistically significant difference ($p = 0.366$) between **read** and **write**.

Index End

Sign test

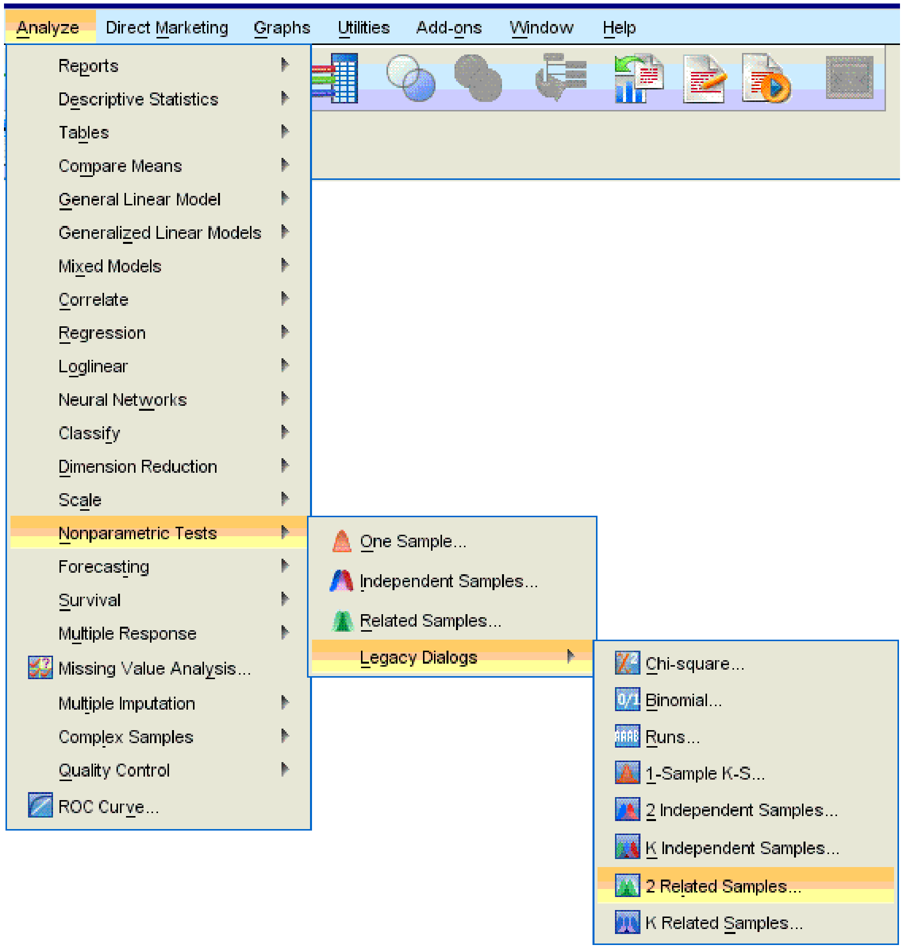
If you believe the differences between **read** and **write** were not ordinal but could merely be classified as positive and negative, then you may want to consider a sign test in lieu of sign rank test. The Sign test answers the question “How Often?”, whereas other tests answer the question “How Much?”. Again, we will use the same variables in this example and assume that this difference is not ordinal.

Menu selection:- Analyze > Nonparametric Tests > Legacy Dialogs > 2 Related Samples

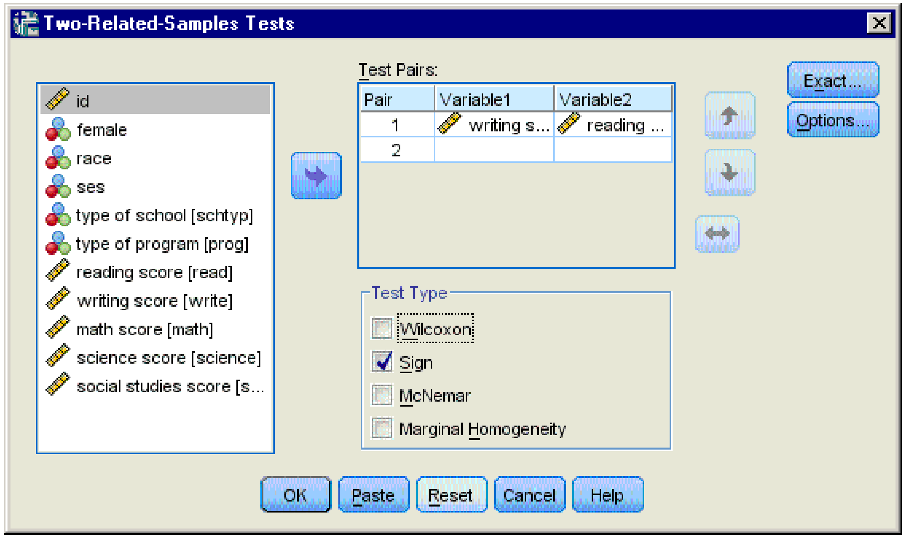
Syntax:-

```
npar test
 /sign = read with write (paired).
```

Sign test


Analyze
Direct Marketing
Graphs
Utilities
Add-ons
Window
Help
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
Quality Control
ROC Curve...
One Sample...
Independent Samples...
Related Samples...
Legacy Dialogs
Chi-square...
Binomial...
Runs...
1-Sample K-S...
2 Independent Samples...
K Independent Samples...
2 Related Samples...
K Related Samples...

Sign test



Two-Related-Samples Tests
id
female
race
ses
type of school [schtyp]
type of program [prog]
reading score [read]
writing score [write]
math score [math]
science score [science]
social studies score [s...
Test Pairs:
Pair
Variable1
Variable2
1
writing s...
reading ...
2
Test Type
Wilcoxon
Sign
McNemar
Marginal Homogeneity
Exact...
Options...
OK
Paste
Reset
Cancel
Help

Sign test

Frequencies

		N
writing score - reading score	Negative Differences[a]	88
	Positive Differences[b]	97
	Ties[c]	15
	Total	200

a. writing score < reading score
b. writing score > reading score
c. writing score = reading score

Test Statistics[a]

	writing score - reading score
Z	-.588
Asymp. Sig. (2-tailed)	.556

a. Sign Test

We conclude that no statistically significant difference was found ($p = 0.556$).

Index End

McNemar test

You would perform McNemar's test if you were interested in the marginal frequencies of two binary outcomes. These binary outcomes may be the same outcome variable on matched pairs (like a case-control study) or two outcome variables from a single group. Continuing with the A dataset used in several above examples, let us create two binary outcomes in our dataset: **himath** and **hiread**. These outcomes can be considered in a two-way contingency table.

The null hypothesis is that the proportion of students in the **himath** group is the same as the proportion of students in **hiread** group (i.e., that the contingency table is symmetric).

Menu selection:- Transform > Compute Variable

Analyze > Descriptive Statistics > Crosstabs

The syntax is on the next slide.

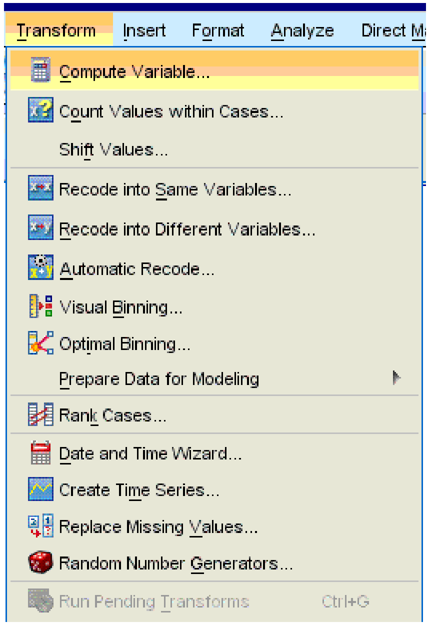
McNemar test

Syntax:-

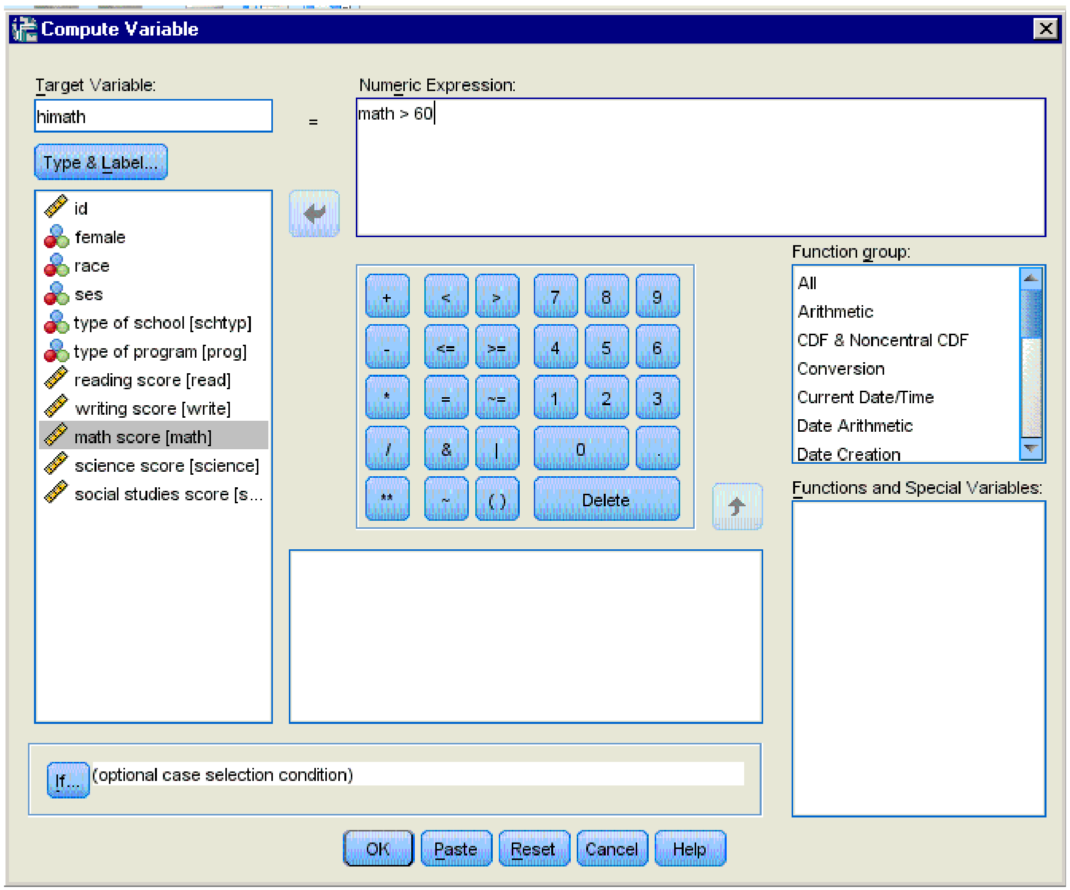
```
COMPUTE himath=math>60.
COMPUTE hiread=read>60.
EXECUTE.

CROSSTABS
/TABLES=himath BY hiread
/STATISTICS=MCNEMAR
/CELLS=COUNT.
```

McNemar test

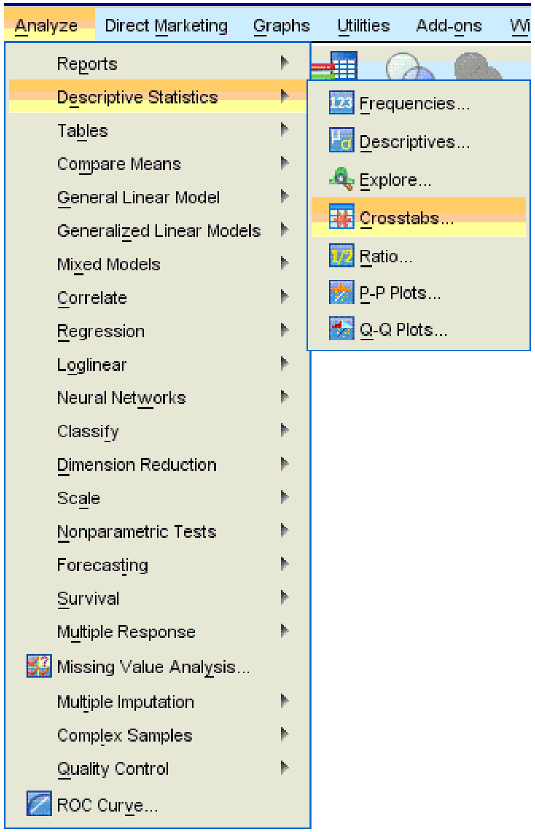


McNemar test

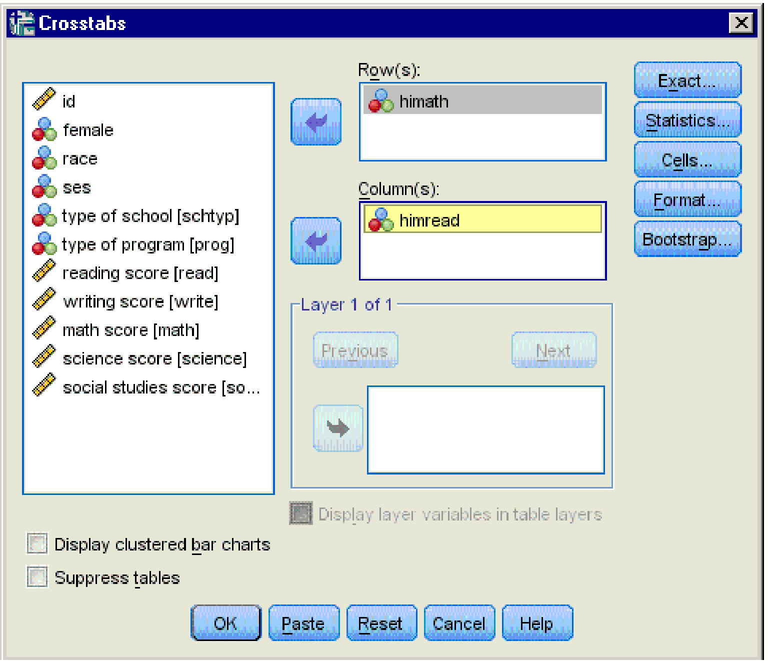


Which is utilised twice, for math and read

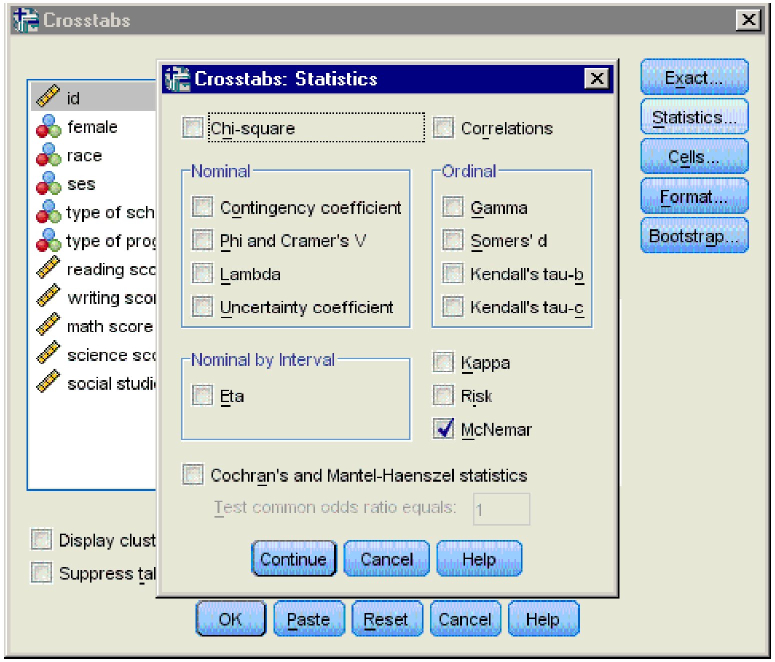
McNemar test



McNemar test



McNemar test



McNemar test

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
himath * hiread	200	100.0%	0	.0%	200	100.0%

himath * hiread Crosstabulation

Count

		hiread		Total
		.00	1.00	
himath	.00	135	21	156
	1.00	18	26	44
Total		153	47	200

Chi-Square Tests

	Value	Exact Sig. (2-sided)
McNemar Test		.749[a]
N of Valid Cases	200	

a. Binomial distribution used.

McNemar's chi-square statistic suggests that there is not a statistically significant difference in the proportion of students in the **himath** group and the proportion of students in the **hiread** group.

Index End

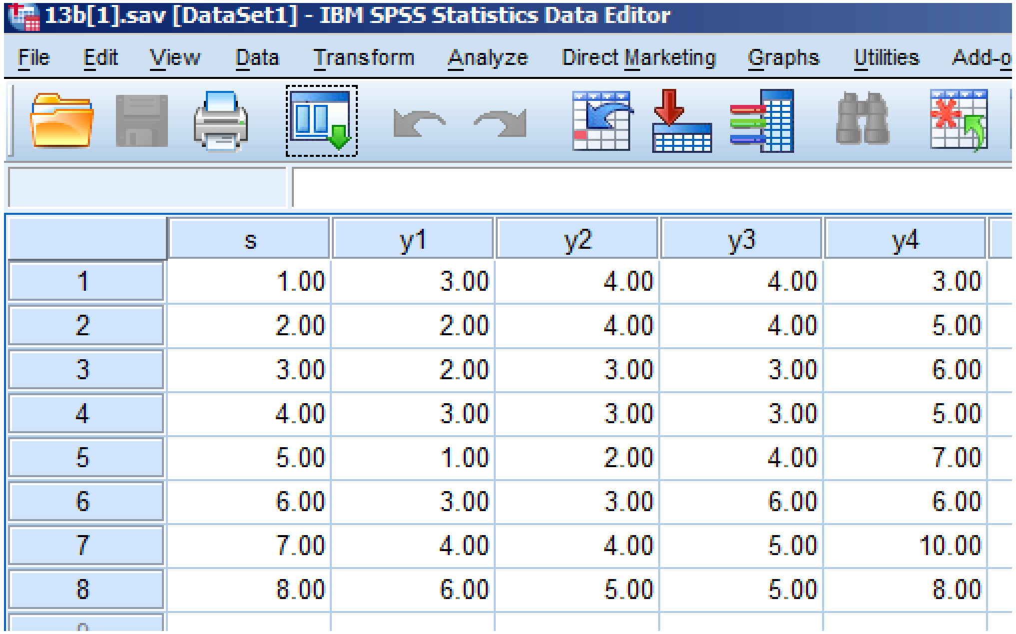
About the B data file

We have an example data set called B, which is used in Roger E. Kirk's book Experimental Design: Procedures for Behavioral Sciences (Psychology) (ISBN 0534250920).

Syntax:- display dictionary
/VARIABLES s y1 y2 y3 y4.

Variable	Position	Measurement Level
s	1	Ordinal
y1	2	Scale
y2	3	Scale
y3	4	Scale
y4	5	Scale

About the B data file



	s	y1	y2	y3	y4
1	1.00	3.00	4.00	4.00	3.00
2	2.00	2.00	4.00	4.00	5.00
3	3.00	2.00	3.00	3.00	6.00
4	4.00	3.00	3.00	3.00	5.00
5	5.00	1.00	2.00	4.00	7.00
6	6.00	3.00	3.00	6.00	6.00
7	7.00	4.00	4.00	5.00	10.00
8	8.00	6.00	5.00	5.00	8.00

Index End

One-way repeated measures ANOVA

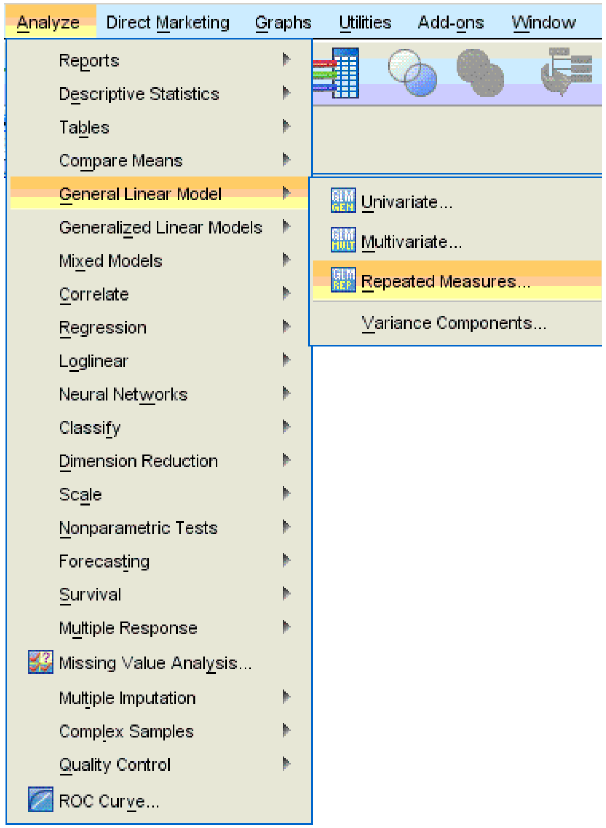
You would perform a one-way repeated measures analysis of variance if you had one categorical independent variable and a normally distributed interval dependent variable that was repeated at least twice for each subject. This is the equivalent of the paired samples t-test, but allows for two or more levels of the categorical variable. This tests whether the mean of the dependent variable differs by the categorical variable. In data set B, **y** (y1 y2 y3 y4) is the dependent variable, **a** is the repeated measure (a name you assign) and **s** is the variable that indicates the subject number.

Menu selection:- Analyze > General Linear Model > Repeated Measures

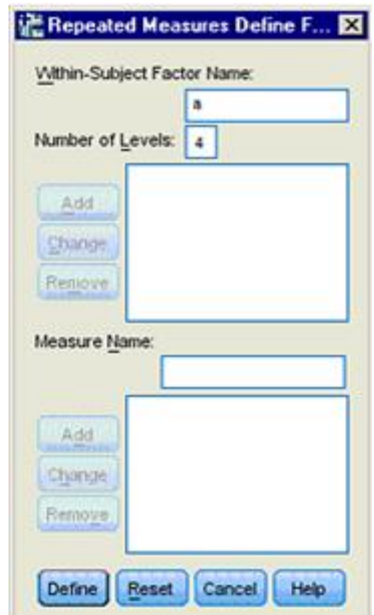
Syntax:-

```
glm y1 y2 y3 y4
 /wsfactor a(4).
```

One-way repeated measures ANOVA

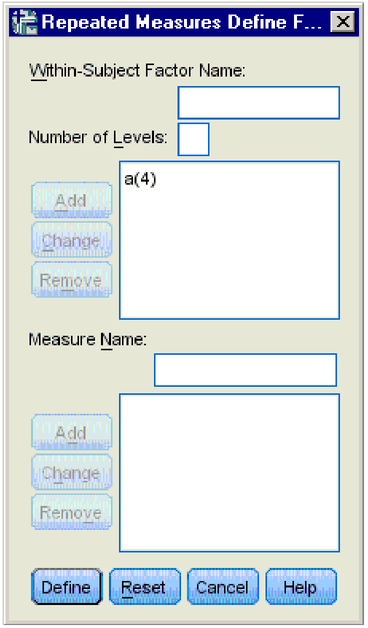


One-way repeated measures ANOVA



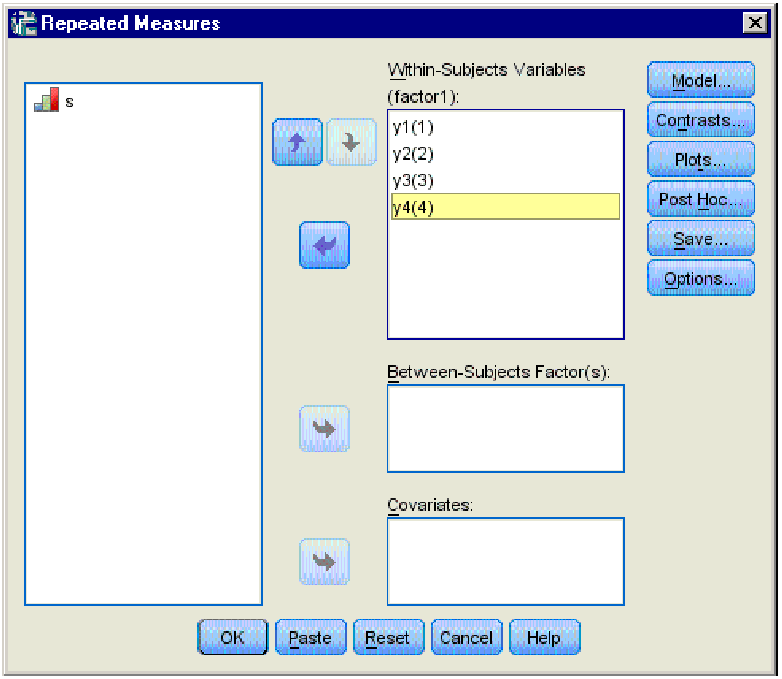
You chose the factor name a which you then “Add”.

One-way repeated measures ANOVA



You chose the factor name a which you then “Add”.

One-way repeated measures ANOVA



One-way repeated measures ANOVA

Within-Subjects Factors

Measure:MEASURE_1

a	Dependent Variable
1	y1
2	y2
3	y3
4	y4

Multivariate Tests[b]

Effect		Value	F	Hypothesis df	Error df	Sig.
a	Pillai's Trace	.754	5.114[a]	3.000	5.000	.055
	Wilks' Lambda	.246	5.114[a]	3.000	5.000	.055
	Hotelling's Trace	3.068	5.114[a]	3.000	5.000	.055
	Roy's Largest Root	3.068	5.114[a]	3.000	5.000	.055

a. Exact statistic

b. Design: Intercept
Within Subjects Design: a

Mauchly's Test of Sphericity[b]

Measure:MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.
a	.339	6.187	5	.295

Mauchly's Test of Sphericity[b]

Measure:MEASURE_1

Within Subjects Effect	Epsilon[a]		
	Greenhouse-Geisser	Huynh-Feldt	Lower-bound
A	.620	.834	.333

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Tests of Within-Subjects Effects

Measure:MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
A	Sphericity Assumed	49.000	3	16.333	11.627	.000
	Greenhouse-Geisser	49.000	1.859	26.365	11.627	.001
	Huynh-Feldt	49.000	2.503	19.578	11.627	.000
	Lower-bound	49.000	1.000	49.000	11.627	.011
Error(a)	Sphericity Assumed	29.500	21	1.405		
	Greenhouse-Geisser	29.500	13.010	2.268		
	Huynh-Feldt	29.500	17.520	1.684		
	Lower-bound	29.500	7.000	4.214		

Tests of Within-Subjects Contrasts

Measure:MEASURE_1

Source	a	Type III Sum of Squares	df	Mean Square	F	Sig.
A	Linear	44.100	1	44.100	19.294	.003
	Quadratic	4.500	1	4.500	3.150	.119
	Cubic	.400	1	.400	.800	.401
Error(a)	Linear	16.000	7	2.286		
	Quadratic	10.000	7	1.429		
	Cubic	3.500	7	.500		

Tests of Between-Subjects Effects

Measure:MEASURE_1

Transformed Variable:Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	578.000	1	578.000	128.444	.000
Error	31.500	7	4.500		

One-way repeated measures ANOVA

Tests of Within-Subjects Effects

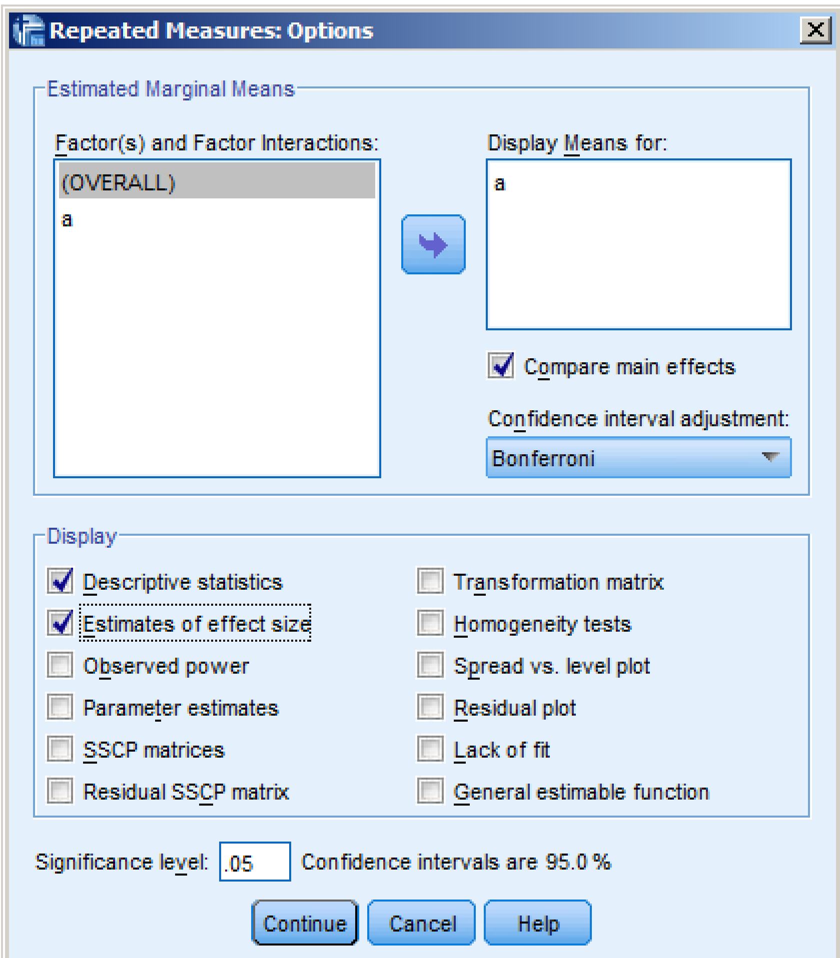
Measure:MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
A	Sphericity Assumed	49.000	3	16.333	11.627	.000
	Greenhouse-Geisser	49.000	1.859	26.365	11.627	.001
	Huynh-Feldt	49.000	2.503	19.578	11.627	.000
	Lower-bound	49.000	1.000	49.000	11.627	.011
Error(a)	Sphericity Assumed	29.500	21	1.405		
	Greenhouse-Geisser	29.500	13.010	2.268		
	Huynh-Feldt	29.500	17.520	1.684		
	Lower-bound	29.500	7.000	4.214		

You will notice that this output gives four different p-values. The output labelled "sphericity assumed" is the p-value (<0.0005) that you would get if you assumed compound symmetry in the variance-covariance matrix. Because that assumption is often not valid, the three other p-values offer various corrections (the Huynh-Feldt, H-F, Greenhouse-Geisser, G-G and Lower-bound). No matter which p-value you use, our results indicate that we have a statistically significant effect of **a** at the .05 level.

Index End

Bonferroni for pairwise comparisons



This is a minor extension of the previous analysis.

Menu selection:-
Analyze
> General Linear Model
> Repeated Measures

Syntax:-

GLM y1 y2 y3 y4
  /WSFACTOR=a 4 Polynomial
  /METHOD=SSTYPE(3)

Only the additional outputs are presented.

Bonferroni for pairwise comparisons

Descriptive Statistics

Mean	Std. Deviation	N
3.0000	1.51186	8
3.5000	.92582	8
4.2500	1.03510	8
6.2500	2.12132	8

This table simply provides important descriptive statistics for the analysis as shown below.

Bonferroni for pairwise comparisons

Estimated Marginal Means

a

Estimates

Measure:MEASURE_1

a	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	3.000	.535	1.736	4.264
2	3.500	.327	2.726	4.274
3	4.250	.366	3.385	5.115
4	6.250	.750	4.477	8.023

Using post hoc tests to examine whether estimated marginal means differ for levels of specific factors in the model.

Bonferroni for pairwise comparisons

Pairwise Comparisons

Measure:MEASURE_1

(I) a	(J) a	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
					Lower Bound	Upper Bound
1	2	-.500	.327	1.000	-1.690	.690
	3	-1.250	.491	.230	-3.035	.535
	4	-3.250*	.726	.017	-5.889	-.611
2	1	.500	.327	1.000	-.690	1.690
	3	-.750	.412	.668	-2.248	.748
	4	-2.750	.773	.056	-5.562	.062
3	1	1.250	.491	.230	-.535	3.035
	2	.750	.412	.668	-.748	2.248
	4	-2.000	.681	.131	-4.477	.477
4	1	3.250*	.726	.017	.611	5.889
	2	2.750	.773	.056	-.062	5.562
	3	2.000	.681	.131	-.477	4.477

Based on estimated marginal means

a. Adjustment for multiple comparisons: Bonferroni.

*. The mean difference is significant at the .05 level.

The results presented in the **Tests of Within-Subjects Effects** table, the Huynh-Feldt ($p < .0005$) informed us that we have an overall significant difference in means, but we do not know where those differences occurred.

This table presents the results of the Bonferroni post-hoc test, which allows us to discover which specific means differed.

Remember, if your overall ANOVA result was not significant, you should not examine the Pairwise Comparisons table.

Bonferroni for pairwise comparisons

Pairwise Comparisons

Measure:MEASURE_1

(I) a	(J) a	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
					Lower Bound	Upper Bound
1	2	-.500	.327	1.000	-1.690	.690
	3	-1.250	.491	.230	-3.035	.535
	4	-3.250*	.726	.017	-5.889	-.611
2	1	.500	.327	1.000	-.690	1.690
	3	-.750	.412	.668	-2.248	.748
	4	-2.750	.773	.056	-5.562	.062
3	1	1.250	.491	.230	-.535	3.035
	2	.750	.412	.668	-.748	2.248
	4	-2.000	.681	.131	-4.477	.477
4	1	3.250*	.726	.017	.611	5.889
	2	2.750	.773	.056	-.062	5.562
	3	2.000	.681	.131	-.477	4.477

Based on estimated marginal means

a. Adjustment for multiple comparisons: Bonferroni.

*. The mean difference is significant at the .05 level.

We can see that there was a significant difference between 1 and 4 (p = 0.017), while 2 and 4 merit further consideration.

Bonferroni for pairwise comparisons

Multivariate Tests

	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Pillai's trace	.754	5.114[a]	3.000	5.000	.055	.754
Wilks' lambda	.246	5.114[a]	3.000	5.000	.055	.754
Hotelling's trace	3.068	5.114[a]	3.000	5.000	.055	.754
Roy's largest root	3.068	5.114[a]	3.000	5.000	.055	.754

Each F tests the multivariate effect of a. These tests are based on the linearly independent pairwise comparisons among the estimated marginal means.

a. Exact statistic

The table provides four variants of the F test. Wilks' lambda is the most commonly reported. Usually the same substantive conclusion emerges from any variant. For these data, we conclude that none of effects are significant (p = 0.055).

Bonferroni for pairwise comparisons

Multivariate Tests

	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Pillai's trace	.754	5.114[a]	3.000	5.000	.055	.754
Wilks' lambda	.246	5.114[a]	3.000	5.000	.055	.754
Hotelling's trace	3.068	5.114[a]	3.000	5.000	.055	.754
Roy's largest root	3.068	5.114[a]	3.000	5.000	.055	.754

Each F tests the multivariate effect of a. These tests are based on the linearly independent pairwise comparisons among the estimated marginal means.

a. Exact statistic

Wilks lambda is the easiest to understand and therefore the most frequently used. It has a good balance between power and assumptions. Wilks lambda can be interpreted as the multivariate counterpart of a univariate R-squared, that is, it indicates the proportion of generalized variance in the dependent variables that is accounted for by the predictors.

Paper

Correct Use of Repeated Measures Analysis of Variance
E. Park, M. Cho and C.-S. Ki. Korean J Lab Med 2009 **29** 1-9

Index End

About the C data file

The C data set contains 3 pulse measurements from each of 30 people assigned to 2 different diet regiments and 3 different exercise regiments.

Syntax:- display dictionary
/VARIABLES id diet exertype pulse time highpulse.

Variable	Position
id	1
diet	2
exertype	3
pulse	4
time	5
highpulse	6

About the C data file

13c[1].sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window H

	id	diet	exertype	pulse	time	highpulse
1	1.00	1.00	1.00	85.00	1.00	.00
2	1.00	1.00	1.00	85.00	2.00	.00
3	1.00	1.00	1.00	88.00	3.00	.00
4	2.00	1.00	1.00	90.00	1.00	.00
5	2.00	1.00	1.00	92.00	2.00	.00
6	2.00	1.00	1.00	93.00	3.00	.00
7	3.00	1.00	1.00	97.00	1.00	.00
8	3.00	1.00	1.00	97.00	2.00	.00
9	3.00	1.00	1.00	94.00	3.00	.00
10	4.00	1.00	1.00	80.00	1.00	.00
11	4.00	1.00	1.00	82.00	2.00	.00
12	4.00	1.00	1.00	83.00	3.00	.00
13	5.00	1.00	1.00	91.00	1.00	.00
14	5.00	1.00	1.00	92.00	2.00	.00
15	5.00	1.00	1.00	91.00	3.00	.00
16	6.00	2.00	1.00	83.00	1.00	.00
17	6.00	2.00	1.00	83.00	2.00	.00
18	6.00	2.00	1.00	84.00	3.00	.00

Index End

Repeated measures logistic regression

If you have a binary outcome measured repeatedly for each subject and you wish to run a logistic regression that accounts for the effect of multiple measures from single subjects, you can perform a repeated measures logistic regression. In SPSS, this can be done using the **GENLIN** command and indicating binomial as the probability distribution and logit as the link function to be used in the model. In C, if we define a "high" pulse as being over 100, we can then predict the probability of a high pulse using diet regime.

Menu selection:- Analyze > Generalized Estimating Equations

However see the next slide.

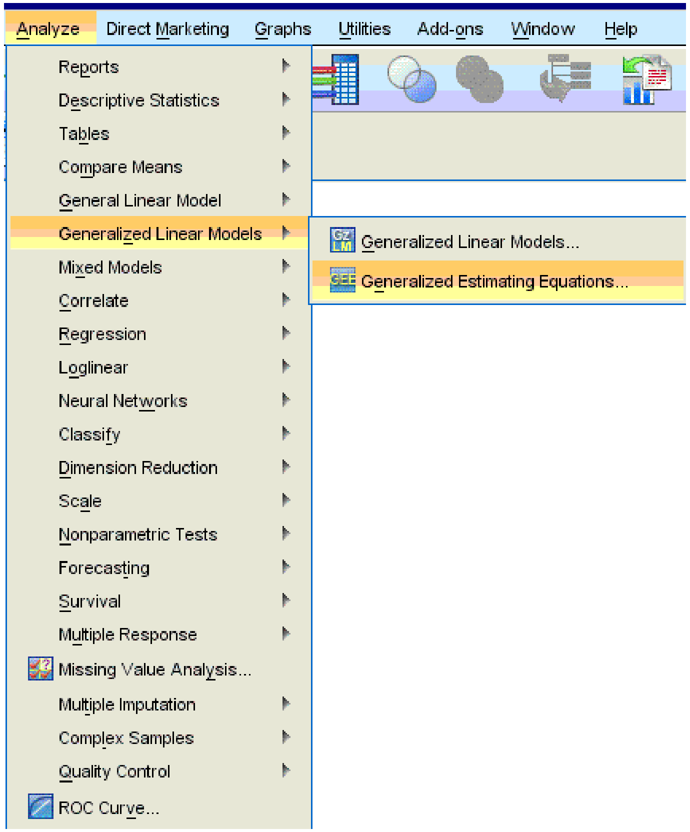
Repeated measures logistic regression

While the drop down menu's can be employed to set the arguments it is simpler to employ the syntax window.

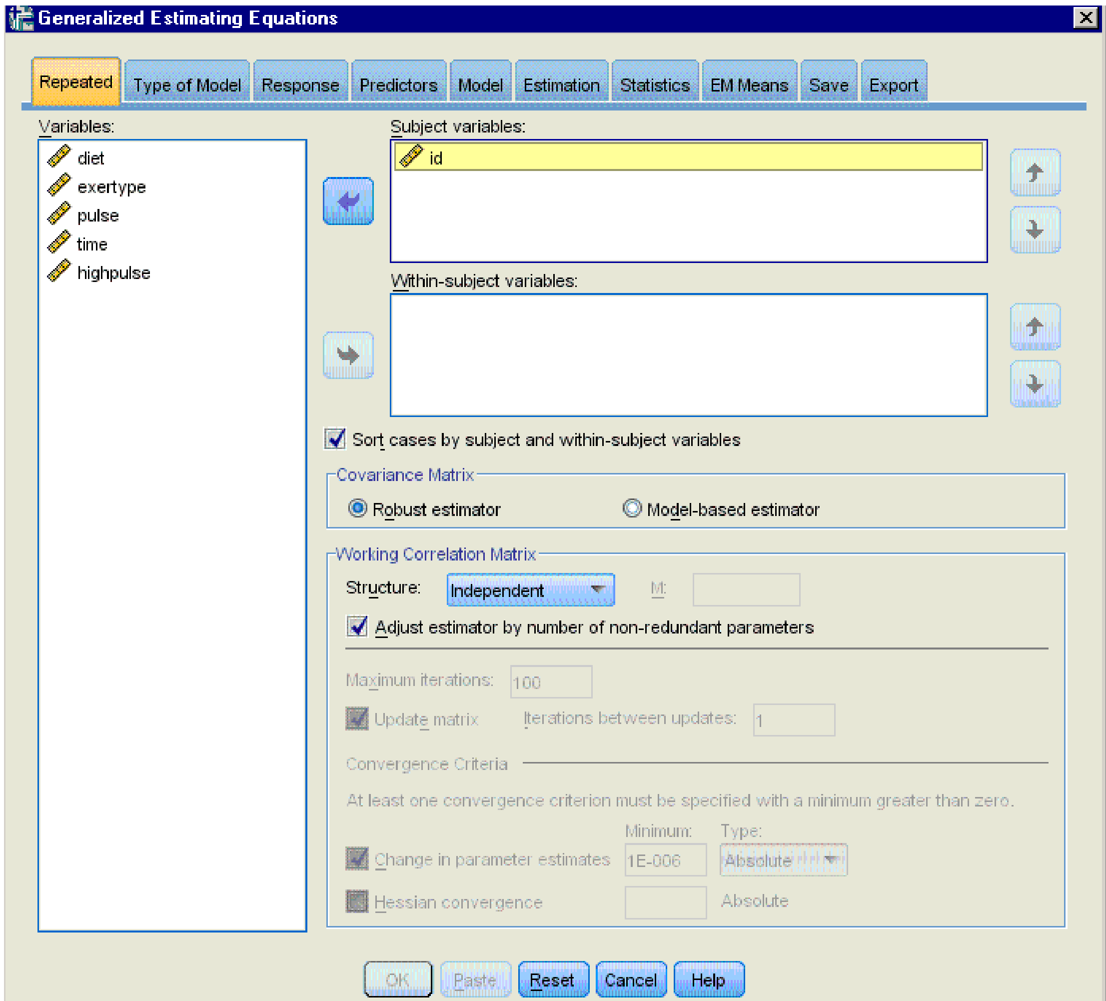
Syntax:-

```
GENLIN highpulse (REFERENCE=LAST)
  BY diet (order=DESCENDING)
 /MODEL diet
  DISTRIBUTION=BINOMIAL
  LINK=LOGIT
 /REPEATED SUBJECT=id CORRTYPE=EXCHANGEABLE.
```

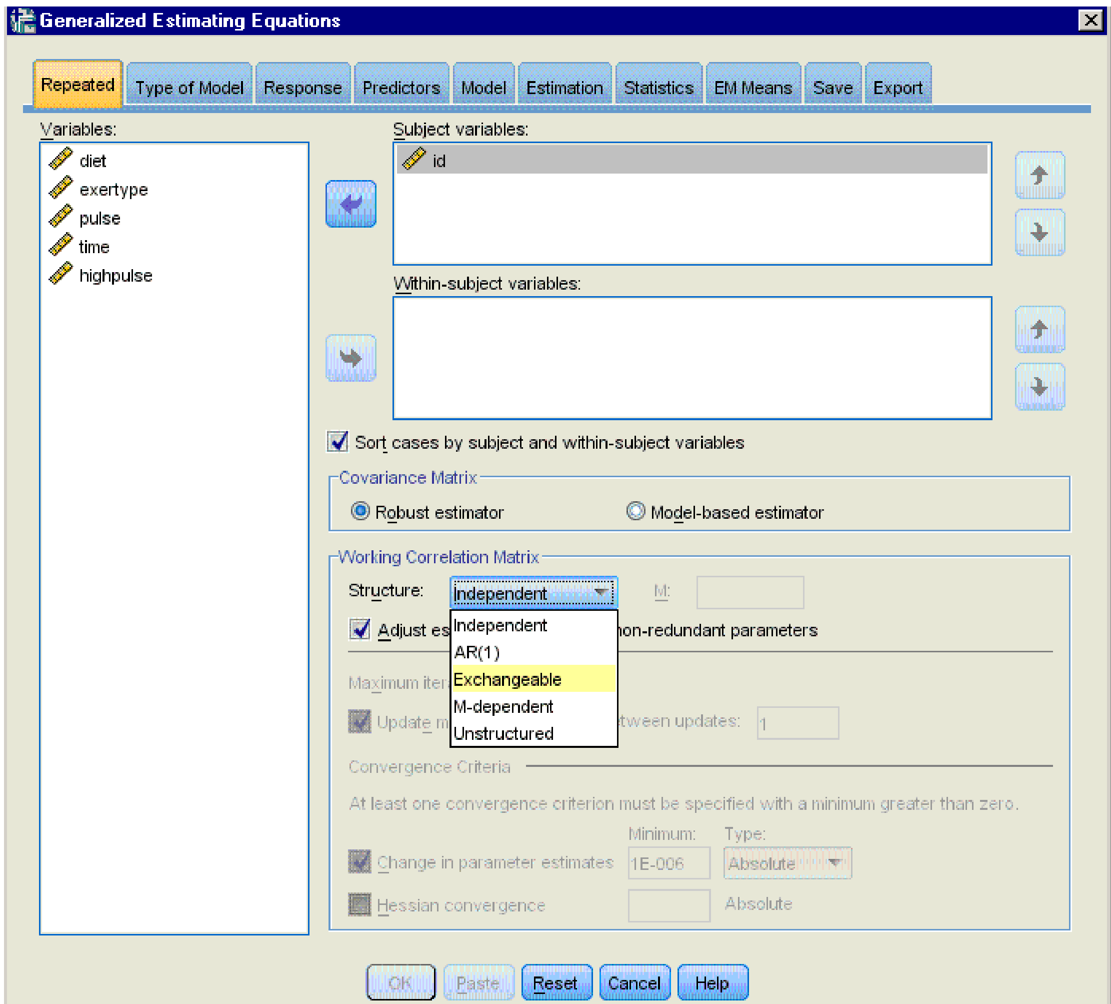
Repeated measures logistic regression



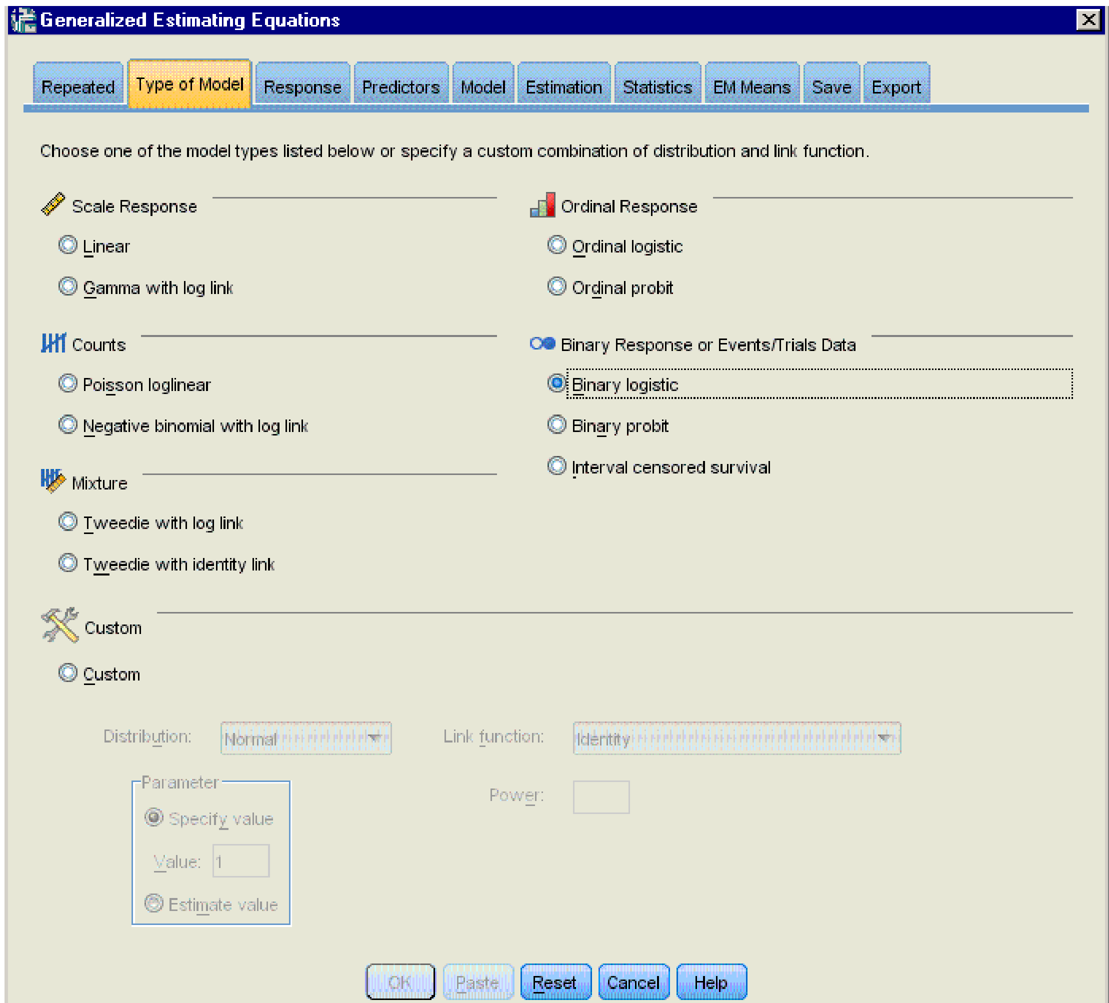
Repeated measures logistic regression



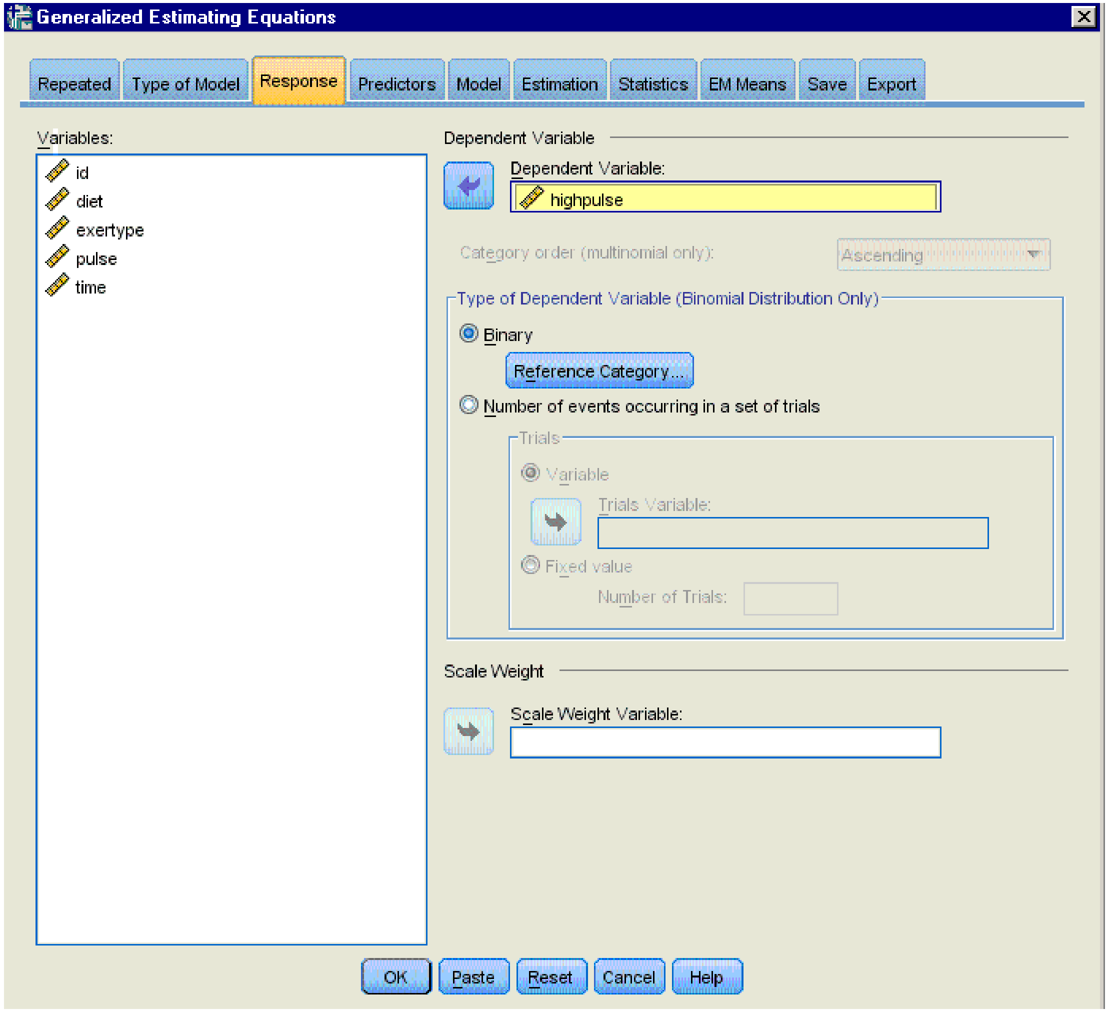
Repeated measures logistic regression



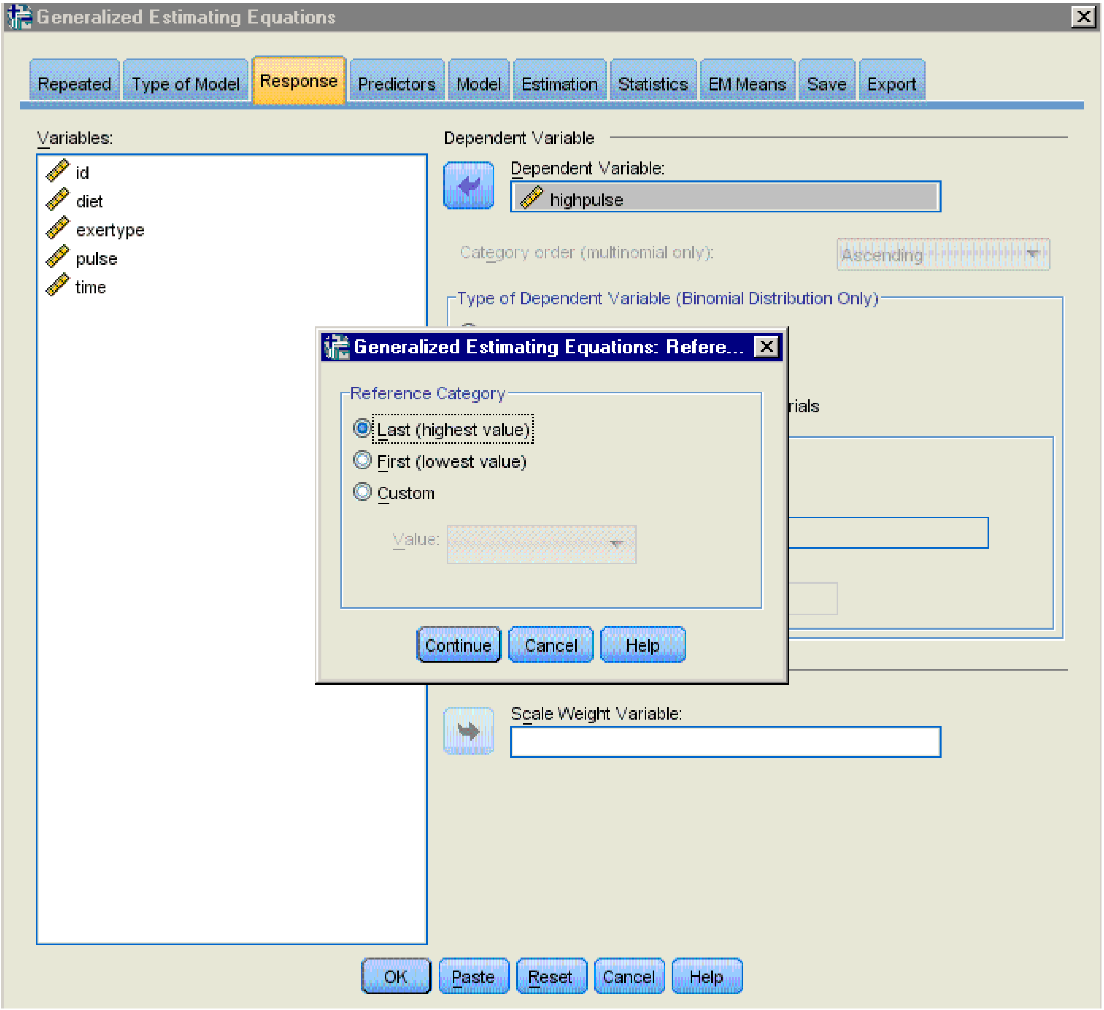
Repeated measures logistic regression



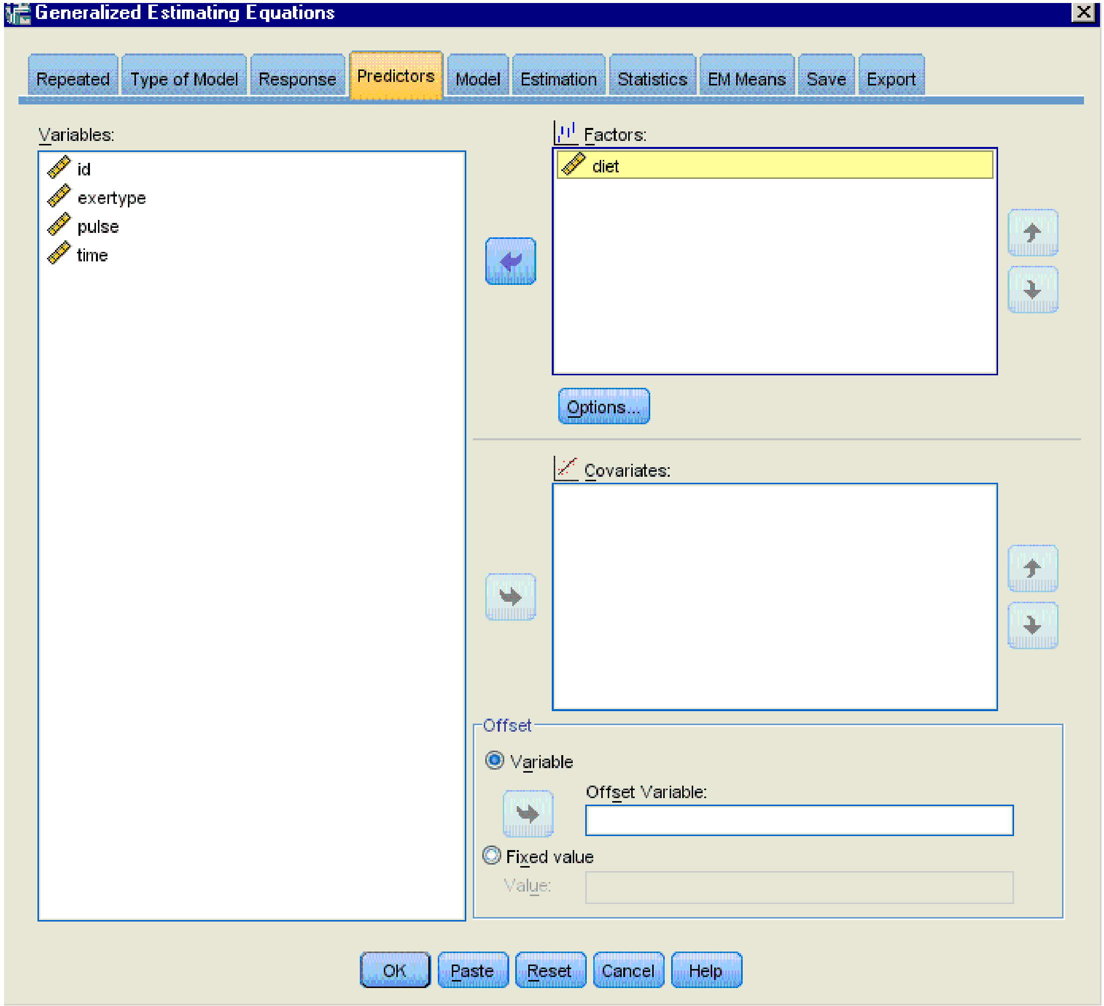
Repeated measures logistic regression



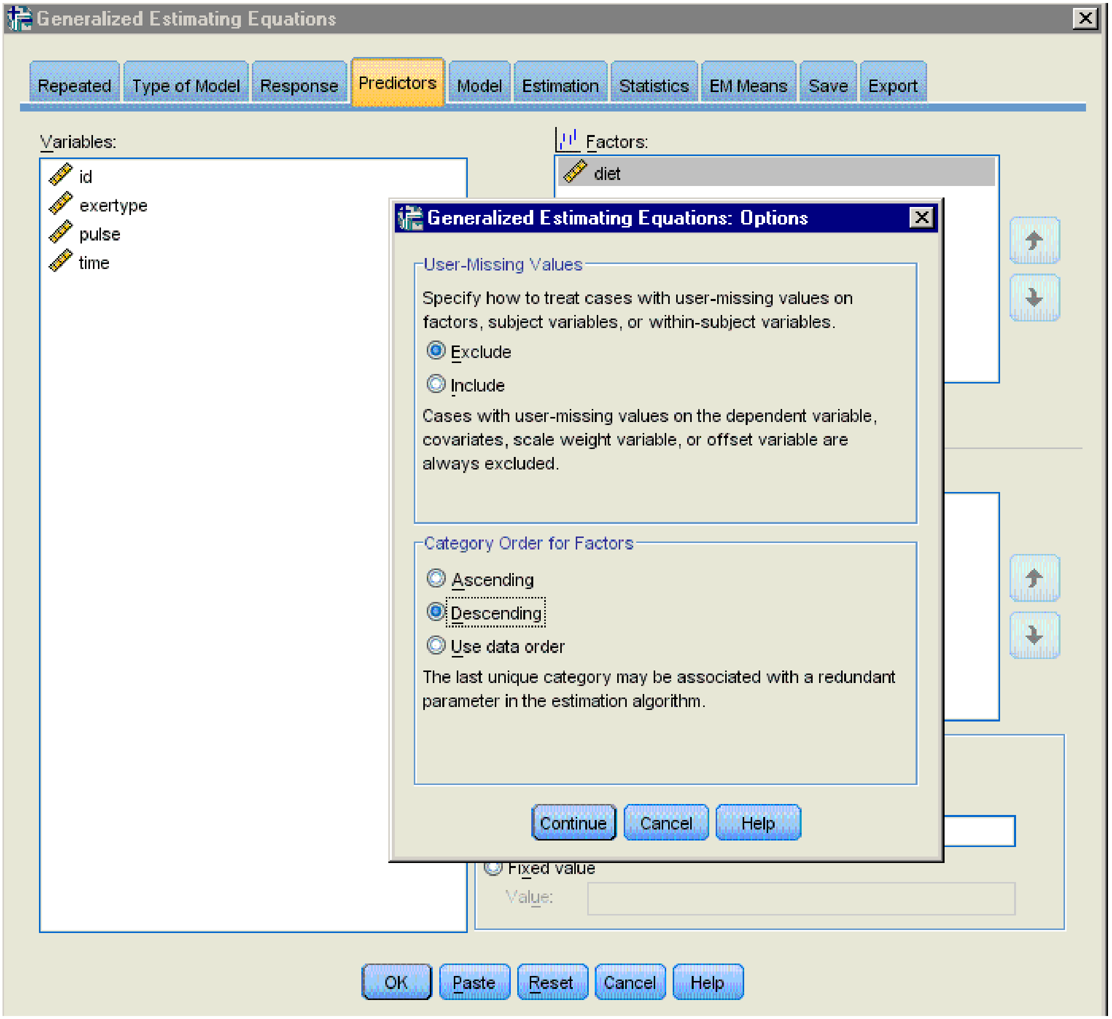
Repeated measures logistic regression



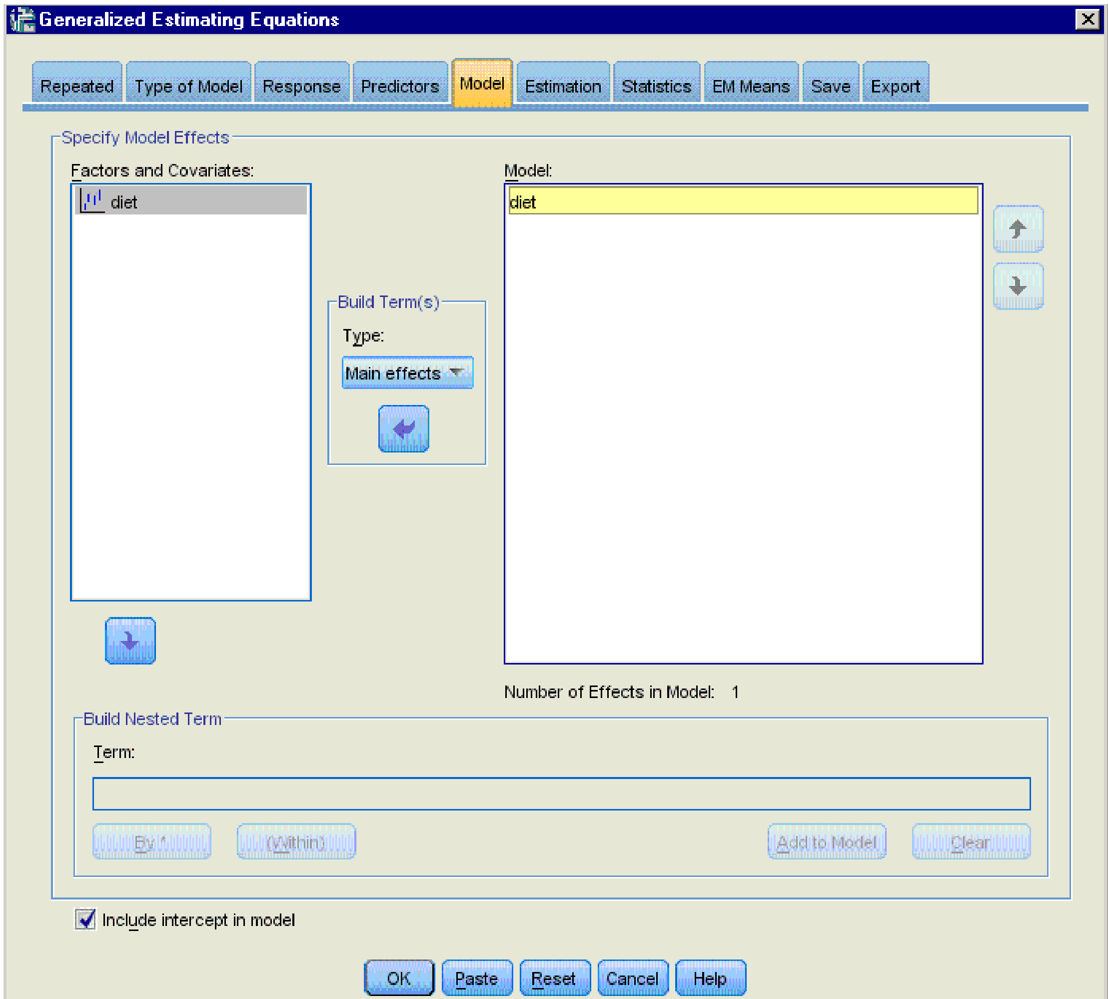
Repeated measures logistic regression



Repeated measures logistic regression



Repeated measures logistic regression



Repeated measures logistic regression

Model Information

Dependent Variable	highpulse[a]
Probability Distribution	Binomial
Link Function	Logit
Subject Effect 1	id
Working Correlation Matrix Structure	Exchangeable

a. The procedure models .00 as the response, treating 1.00 as the reference category.

Case Processing Summary

	N	Percent
Included	90	100.0%
Excluded	0	.0%
Total	90	100.0%

Correlated Data Summary

Number of Levels	Subject Effect	id	30
Number of Subjects			30
Number of Measurements per Subject	Minimum		3
	Maximum		3
Correlation Matrix Dimension			3

Categorical Variable Information

			N	Percent
Dependent Variable	highpulse	.00	63	70.0%
		1.00	27	30.0%
		Total	90	100.0%
Factor	diet	2.00	45	50.0%
		1.00	45	50.0%
		Total	90	100.0%

Goodness of Fit[b]

	Value
Quasi Likelihood under Independence Model Criterion (QIC)[a]	113.986
Corrected Quasi Likelihood under Independence Model Criterion (QICC)[a]	111.340

Dependent Variable: highpulse
Model: (Intercept), diet
a. Computed using the full log quasi-likelihood function.
b. Information criteria are in small-is-better form.

Tests of Model Effects

Source	Type III Wald Chi-Square	df	Sig.
(Intercept)	8.437	1	.004
diet	1.562	1	.211

Dependent Variable: highpulse
Model: (Intercept), diet

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval Lower	Upper
(Intercept)	1.253	.4328	.404	2.101
[diet=2.00]	-.754	.6031	-1.936	.428
[diet=1.00]	0[a]	.	.	.
(Scale)	1			

Repeated measures logistic regression

Parameter Estimates

Parameter	Hypothesis Test		
	Wald Chi-Square	df	Sig.
(Intercept)	8.377	1	.004
[diet=2.00]	1.562	1	.211
[diet=1.00]	.	.	.
(Scale)			

Dependent Variable: highpulse
Model: (Intercept), diet
a. Set to zero because this parameter is redundant.

These results indicate that **diet** is not statistically significant (Wald Chi-Square = 1.562, p = 0.211).

Index End

Factorial ANOVA

A factorial ANOVA has two or more categorical independent variables (either with or without the interactions) and a single normally distributed interval dependent variable. For example, using the A data file we will look at writing scores (**write**) as the dependent variable and gender (**female**) and socio-economic status (**ses**) as independent variables, and we will include an interaction of **female** by **ses**. Note that in SPSS, you do not need to have the interaction term(s) in your data set. Rather, you can have SPSS create it/them temporarily by placing an asterisk between the variables that will make up the interaction term(s). For the approach adopted here, this step is automatic. However see the syntax example below.

Menu selection:- Analyze > General Linear Model > Univariate

Syntax:- glm write by female ses.

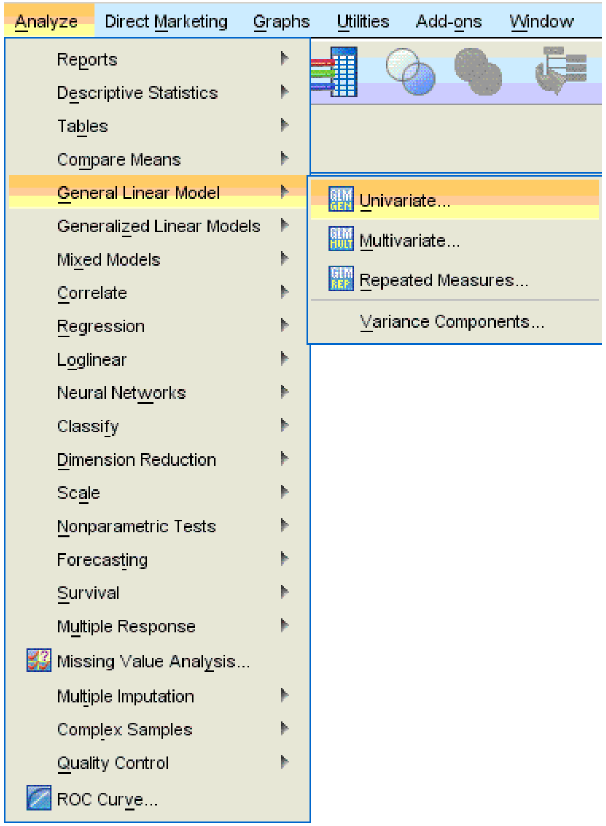
Factorial ANOVA

Alternate Syntax:-

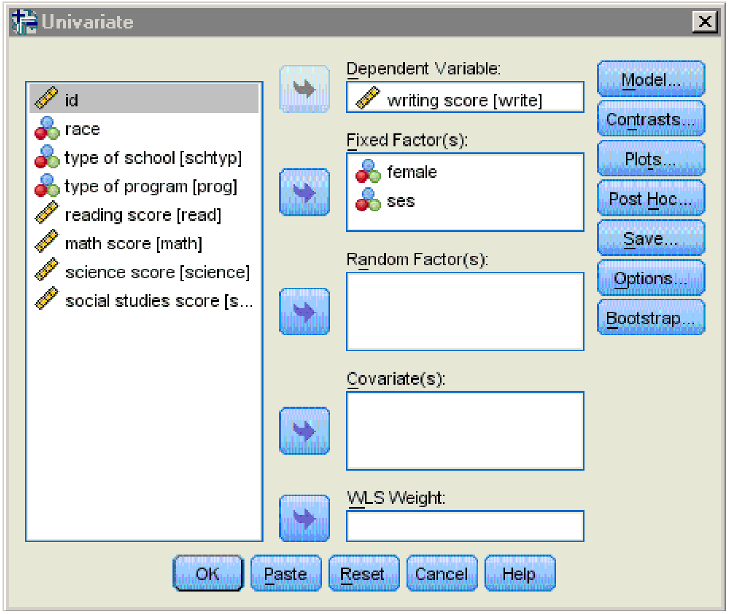
```
UNIANOVA write BY female ses
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(0.05)
  /DESIGN=female ses female*ses.
```

Note the interaction term, female*ses.

Factorial ANOVA



Factorial ANOVA



Factorial ANOVA

Between-Subjects Factors

		Value Label	N
female	.00	male	91
	1.00	female	109
ses	1.00	low	47
	2.00	middle	95
	3.00	high	58

Tests of Between-Subjects Effects

Dependent Variable:writing score

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2278.244[a]	5	455.649	5.666	.000
Intercept	473967.467	1	473967.467	5893.972	.000
female	1334.493	1	1334.493	16.595	.000
ses	1063.253	2	531.626	6.611	.002
female * ses	21.431	2	10.715	.133	.875
Error	15600.631	194	80.416		
Total	574919.000	200			
Corrected Total	17878.875	199			

a. R Squared = 0127 (Adjusted R Squared = 0105)

These results indicate that the overall model is statistically significant ($F = 5.666$, $p < 0.0005$). The variables **female** and **ses** are also statistically significant ($F = 16.595$, $p < 0.0005$ and $F = 6.611$, $p = 0.002$, respectively). However, note that interaction between **female** and **ses** is not statistically significant ($F = 0.133$, $p = 0.875$).

Index End

Friedman test

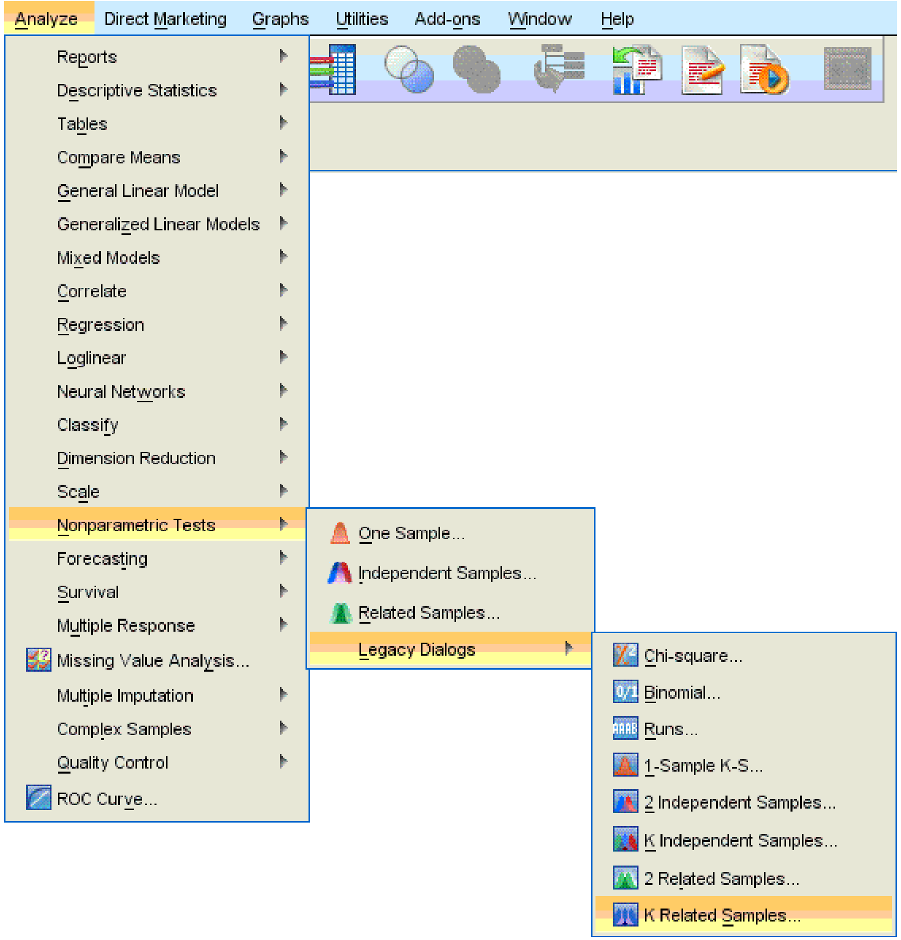
You perform a Friedman test when you have one within-subjects independent variable with two or more levels and a dependent variable that is not interval and normally distributed (but at least ordinal). We will use this test to determine if there is a difference in the reading, writing and math scores. The null hypothesis in this test is that the distribution of the ranks of each type of score (i.e., reading, writing and math) are the same. To conduct a Friedman test, the data need to be in a long format (see the next topic).

Menu selection:- Analyze > Nonparametric Tests > Legacy Dialogs > K Related Samples

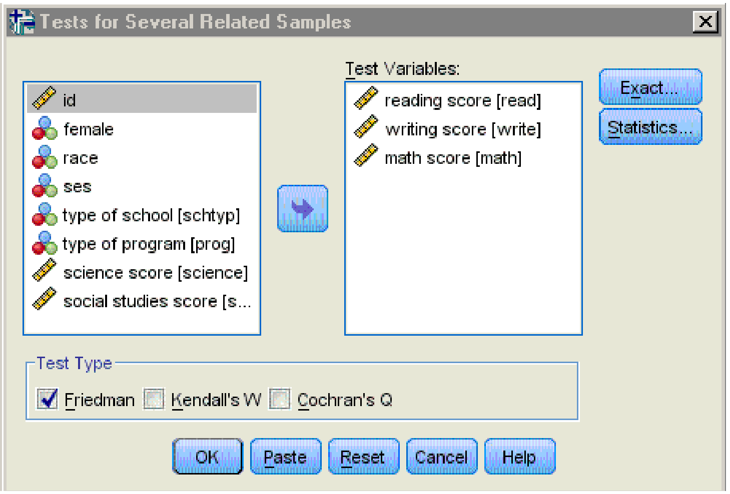
Syntax:-

```
npar tests
/friedman = read write math.
```

Friedman test



Friedman test



Friedman test

Ranks

	Mean Rank
reading score	1.96
writing score	2.04
math score	2.01

Test Statistics[a]

N	200
Chi-Square	.645
df	2
Asymp. Sig.	.724

a. Friedman Test

Friedman's chi-square has a value of 0.645 and a p-value of 0.724 and is not statistically significant. Hence, there is no evidence that the distributions of the three types of scores are different.

Index End

Reshaping data

This example illustrates a wide data file and reshapes it into long form.

Consider the data containing the kids and their heights at one year of age (ht1) and at two years of age (ht2).

FAMID	BIRTH	HT1	HT2
1.00	1.00	2.80	3.40
1.00	2.00	2.90	3.80
1.00	3.00	2.20	2.90
2.00	1.00	2.00	3.20
2.00	2.00	1.80	2.80
2.00	3.00	1.90	2.40
3.00	1.00	2.20	3.30
3.00	2.00	2.30	3.40
3.00	3.00	2.10	2.90

Number of cases read: 9   Number of cases listed: 9

This is called a wide format since the heights are wide. We may want the data to be long, where each height is in a separate observation.

Reshaping data

FAMID	BIRTH	AGE	HT
1.00	1.00	1.00	2.80
1.00	1.00	2.00	3.40
1.00	2.00	1.00	2.90
1.00	2.00	2.00	3.80
1.00	3.00	1.00	2.20
1.00	3.00	2.00	2.90
2.00	1.00	1.00	2.00
2.00	1.00	2.00	3.20
2.00	2.00	1.00	1.80
2.00	2.00	2.00	2.80
2.00	3.00	1.00	1.90
2.00	3.00	2.00	2.40
3.00	1.00	1.00	2.20
3.00	1.00	2.00	3.30
3.00	2.00	1.00	2.30
3.00	2.00	2.00	3.40
3.00	3.00	1.00	2.10
3.00	3.00	2.00	2.90

Number of cases read: 18   Number of cases listed: 18

We may want the data to be long, where each height is in a separate observation.

Data may be restructured using the point and click function in SPSS, or pre-processing with Excel.

Index End

Ordered logistic regression

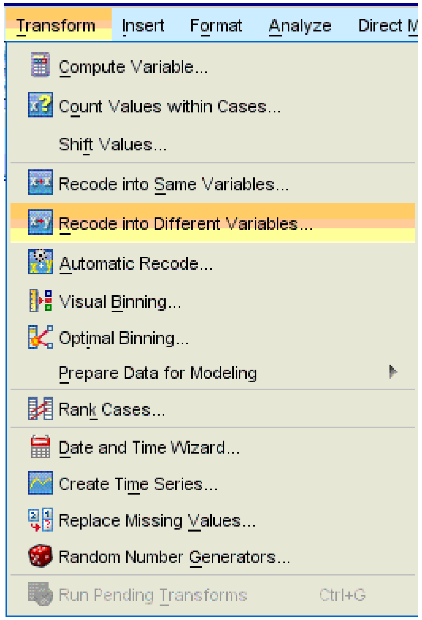
Ordered logistic regression is used when the dependent variable is ordered, but not continuous. For example, using the A data file we will create an ordered variable called **write3**. This variable will have the values 1, 2 and 3, indicating a low, medium or high writing score. We do not generally recommend categorizing a continuous variable in this way; we are simply creating a variable to use for this example.

Menu selection:- Transform > Recode into Different Variables

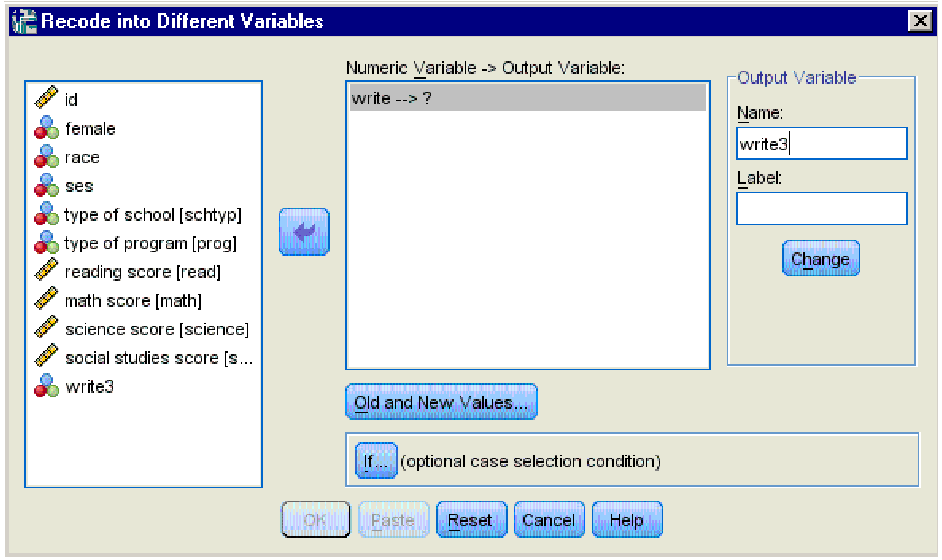
Syntax:-

```
if write ge 30 and write le 48  write3 = 1.
if write ge 49 and write le 57  write3 = 2.
if write ge 58 and write le 70  write3 = 3.
execute.
```

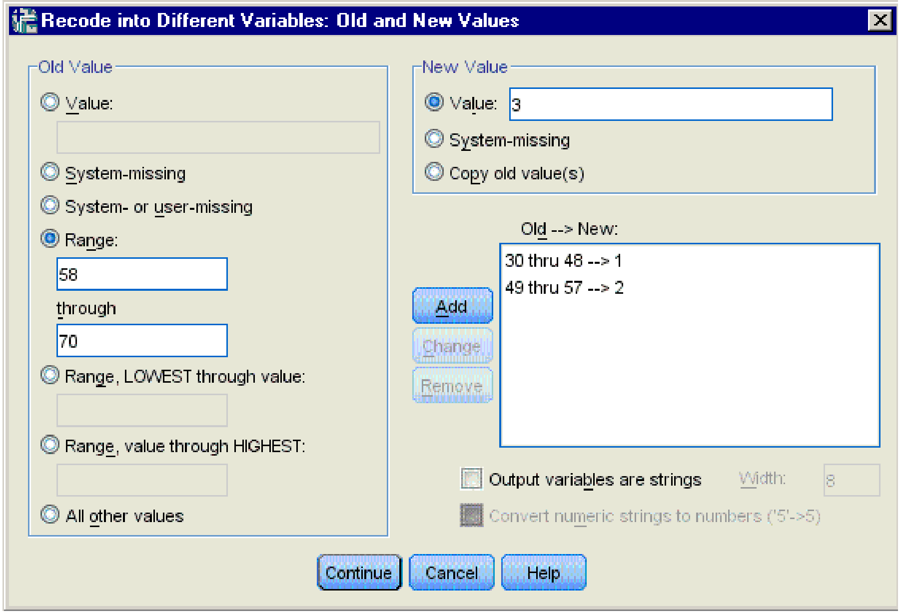
Ordered logistic regression



Ordered logistic regression

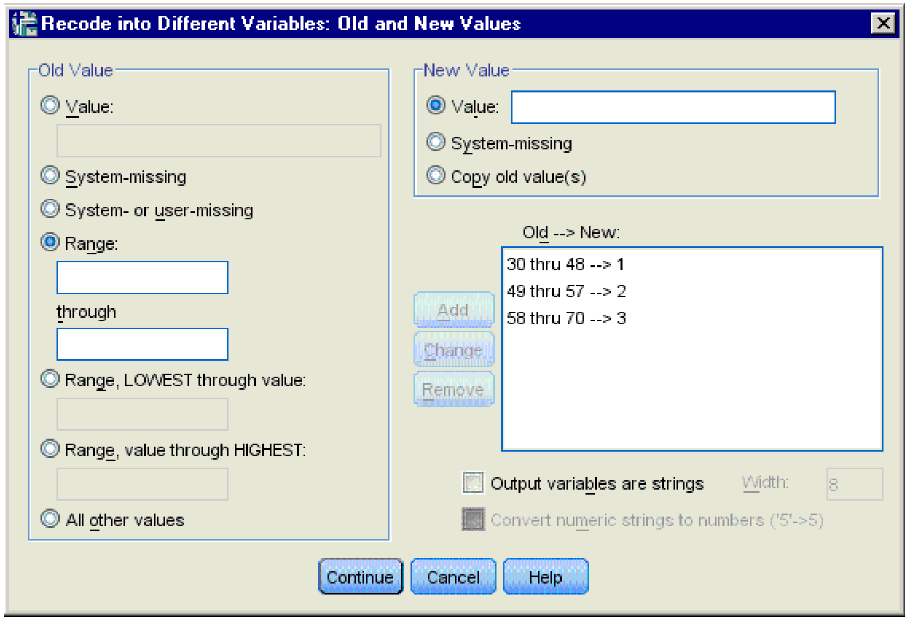


Ordered logistic regression



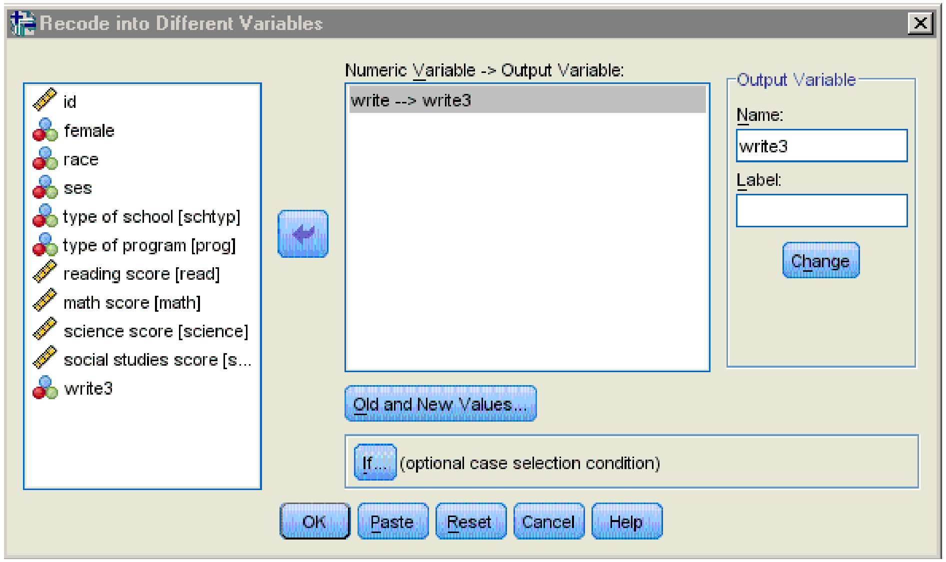
"Add" to create rules and finally "Change"

Ordered logistic regression



finally "continue"

Ordered logistic regression



use "change" to execute

Ordered logistic regression

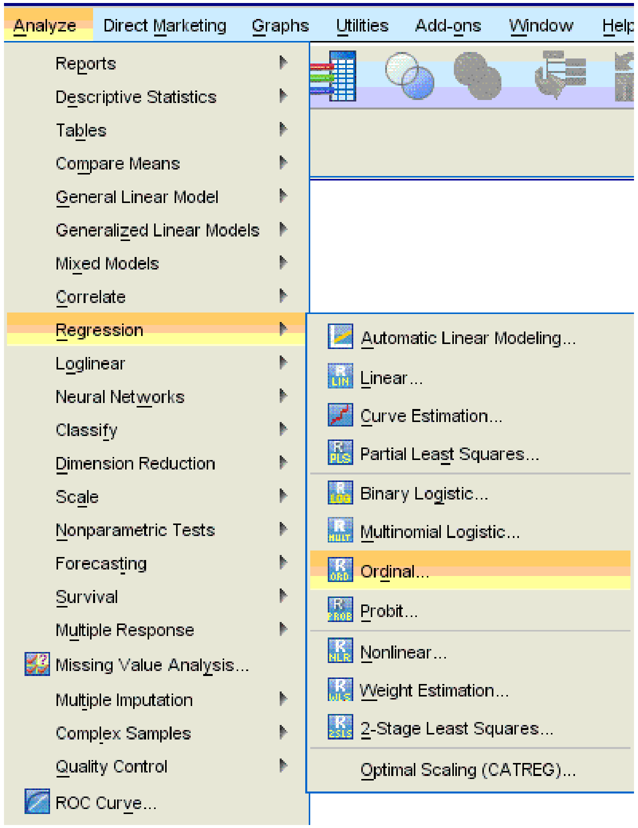
We will use gender (**female**), reading score (**read**) and social studies score (**socst**) as predictor variables in this model. We will use a logit link and on the **print** subcommand we have requested the parameter estimates, the (model) summary statistics and the test of the parallel lines assumption.

Menu selection:- Analyze > Regression > Ordinal

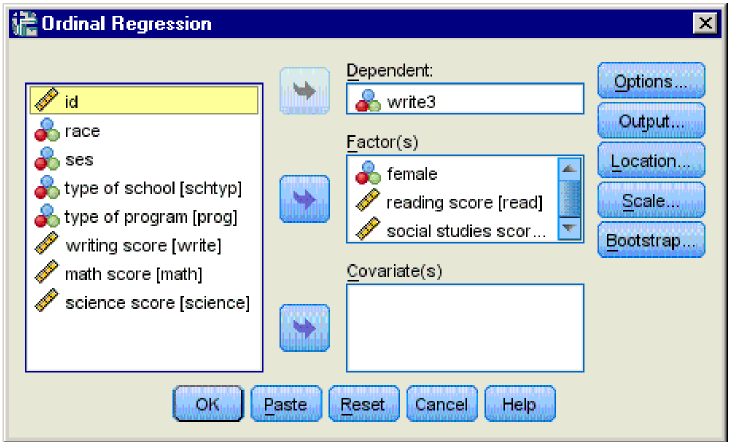
Syntax:-

```
plum write3 with female read socst
 /link = logit
 /print = parameter summary tparallel.
```

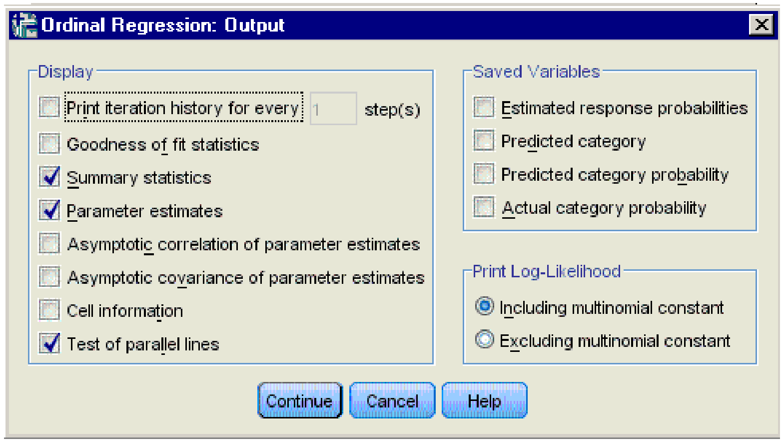
Ordered logistic regression



Ordered logistic regression



Ordered logistic regression



Ordered logistic regression

Case Processing Summary

		N	Marginal Percentage
write3	1.00	61	30.5%
	2.00	61	30.5%
	3.00	78	39.0%
Valid		200	100.0%
Missing		0	
Total		200	

Model Fitting Information

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	376.226			
Final	252.151	124.075	3	.000

Link function: Logit.

Pseudo R-Square

Cox and Snell	.462
Nagelkerke	.521
McFadden	.284

Link function: Logit.

Parameter Estimates

		Estimate	Std. Error	Wald	df	Sig.
Threshold	[write3 = 1.00]	9.704	1.203	65.109	1	.000
	[write3 = 2.00]	11.800	1.312	80.868	1	.000
Location	female	1.285	.322	15.887	1	.000
	read	.118	.022	29.867	1	.000
	socst	.080	.019	17.781	1	.000

Parameter Estimates

		95% Confidence Interval	
		Lower Bound	Upper Bound
Threshold	[write3 = 1.00]	7.347	12.061
	[write3 = 2.00]	9.228	14.372
Location	female	.653	1.918
	read	.076	.160
	socst	.043	.117

Link function: Logit.

Test of Parallel Lines[a]

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Null Hypothesis	252.151			
General	250.104	2.047	3	.563

The null hypothesis states that the location parameters (slope coefficients) are the same across response categories.

a. Link function: Logit.

The results indicate that the overall model is statistically significant ($p < .0005$), as are each of the predictor variables ($p < .0005$). There are two thresholds for this model because there are three levels of the outcome variable. We also see that the test of the proportional odds assumption is non-significant ($p = 0.563$). One of the assumptions underlying ordinal logistic (and ordinal probit) regression is that the relationship between each pair of outcome groups is the same. In other words, ordinal logistic regression assumes that the coefficients that describe the relationship between, say, the lowest versus all higher categories of the response variable are the same as those that describe the relationship between the next lowest category and all higher categories, etc. This is called the proportional odds assumption or the parallel regression assumption. Because the relationship between all pairs of groups is the same, there is only one set of coefficients (only one model). If this was not the case, we would need different models (such as a generalized ordered logit model) to describe the relationship between each pair of outcome groups.

Index End

Factorial logistic regression

A factorial logistic regression is used when you have two or more categorical independent variables but a dichotomous dependent variable. For example, using the A data file we will use **female** as our dependent variable, because it is the only dichotomous variable in our data set; certainly not because it is common practice to use gender as an outcome variable. We will use type of program (**prog**) and school type (**schtyp**) as our predictor variables. Because **prog** is a categorical variable (it has three levels), we need to create dummy codes for it. SPSS will do this for you by making dummy codes for all variables listed after the keyword **with**. SPSS will also create the interaction term; simply list the two variables that will make up the interaction separated by the keyword **by**.

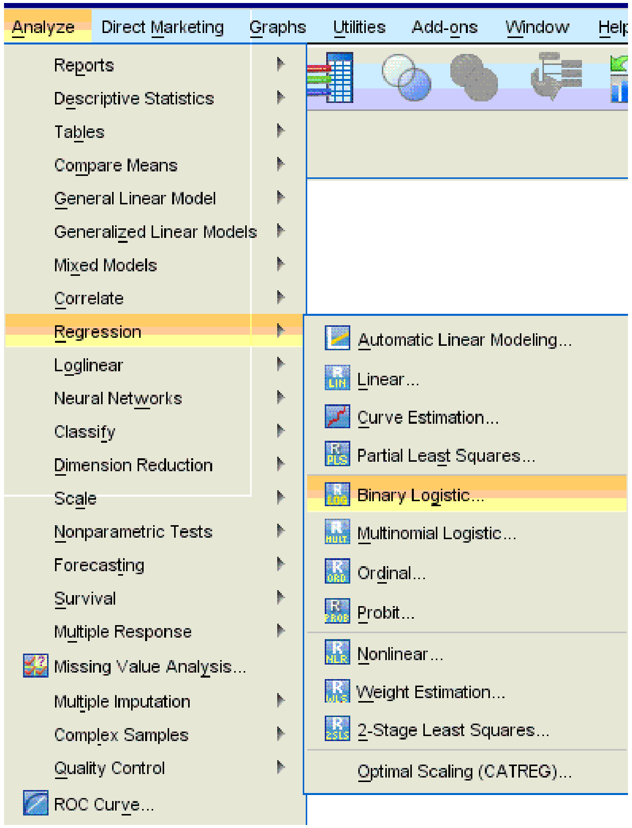
Menu selection:- Analyze > Regression > Binary Logistic

Simplest to realise via the syntax window.

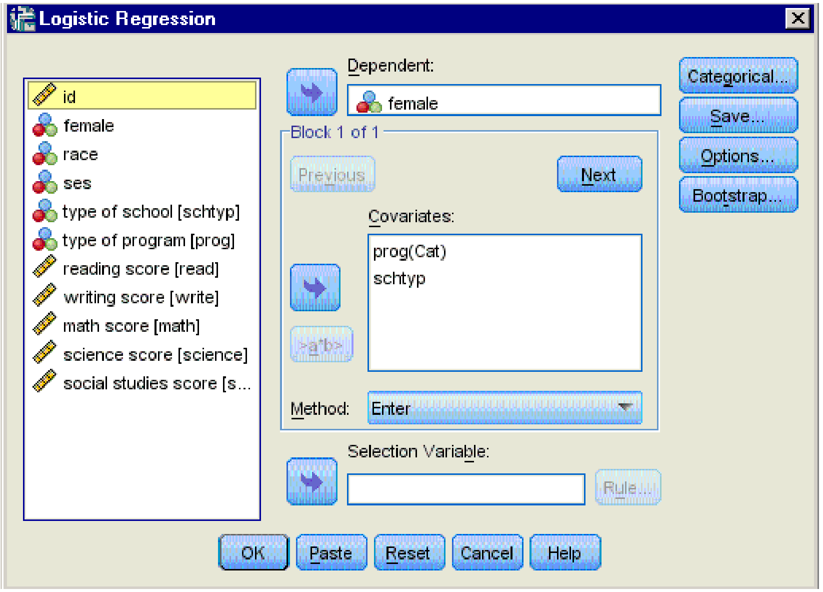
Syntax:-

```
logistic regression female with prog schtyp prog by schtyp
/contrast(prog) = indicator(1).
```

Factorial logistic regression

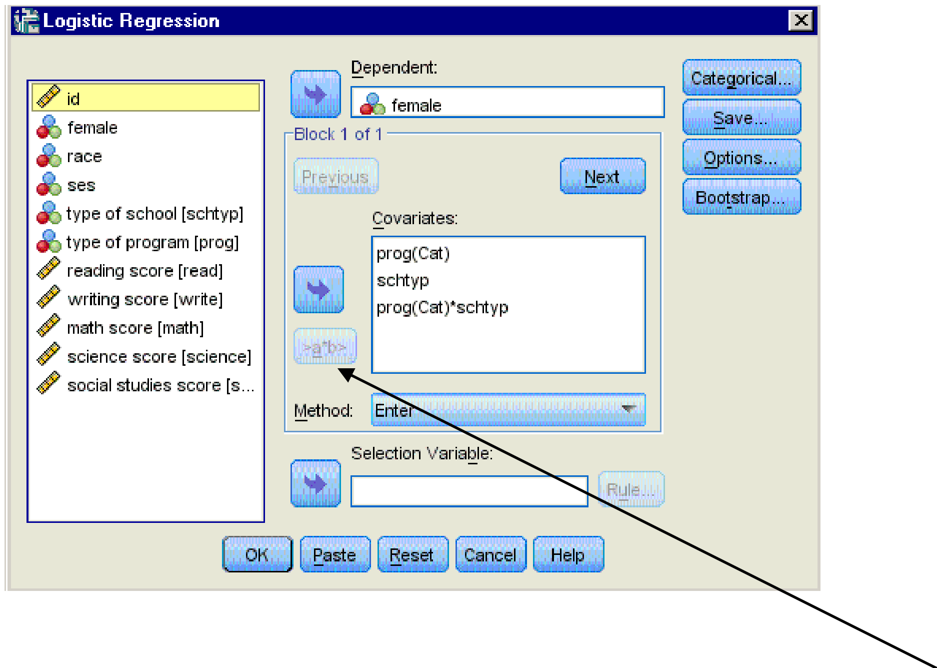


Factorial logistic regression



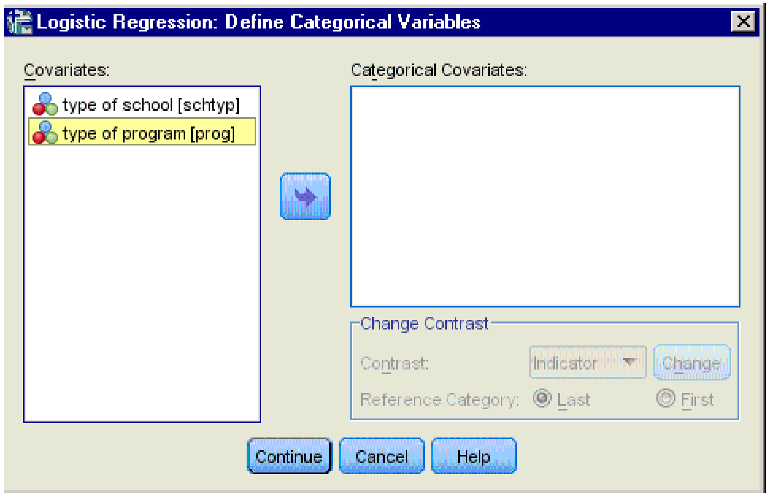
Note that the identification of prog as the categorical variable is made below.

Factorial logistic regression

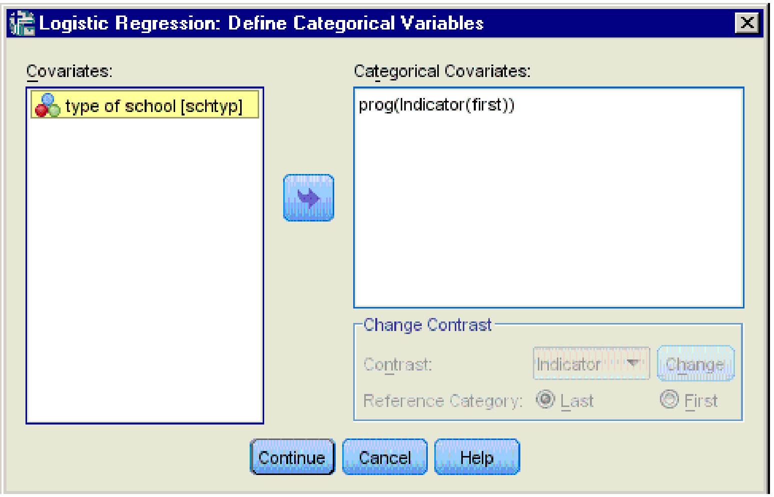


Use Ctrl with left mouse key to select two variables then >a*b> for the product term.

Factorial logistic regression



Factorial logistic regression



Indicator(1) identifies value 1 as the (first) reference category

Factorial logistic regression

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	200	100.0
	Missing Cases	0	.0
	Total	200	100.0
Unselected Cases		0	.0
Total		200	100.0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
male	0
female	1

Categorical Variables Codings

			Parameter coding	
		Frequency	(1)	(2)
type of program	general	45	.000	.000
	academic	105	1.000	.000
	vocation	50	.000	1.000

Block 0: Beginning Block

Classification Table[a,b]

			Predicted		
			female		Percentage
Observed			male	female	Correct
Step 0	female	Male	0	91	.0
		Female	0	109	100.0
	Overall Percentage				54.5

a. Constant is included in the model.

b. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	.180	.142	1.616	1	.204	1.198

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	Prog	.053	2	.974
		prog(1)	.049	1	.826
		prog(2)	.007	1	.935
		Schtyp	.047	1	.828
		prog * schtyp	.031	2	.985
		prog(1) by schtyp	.004	1	.950
		prog(2) by schtyp	.011	1	.917
	Overall Statistics		2.923	5	.712

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	3.147	5	.677
	Block	3.147	5	.677
	Model	3.147	5	.677

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	272.490[a]	.016	.021

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

Classification Table[a]

			Predicted		
			female		Percentage
Observed			male	female	Correct
Step 1	female	Male	32	59	35.2
		Female	31	78	71.6
	Overall Percentage				55.0

a. The cut value is .500

Factorial logistic regression

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	3.147	5	.677
	Block	3.147	5	.677
	Model	3.147	5	.677

The results indicate that the overall model is not statistically significant (Likelihood ratio $Chi^2 = 3.147$, $p = 0.677$). Furthermore, none of the coefficients are statistically significant either. This shows that the overall effect of **prog** is not significant.

Index End

Correlation

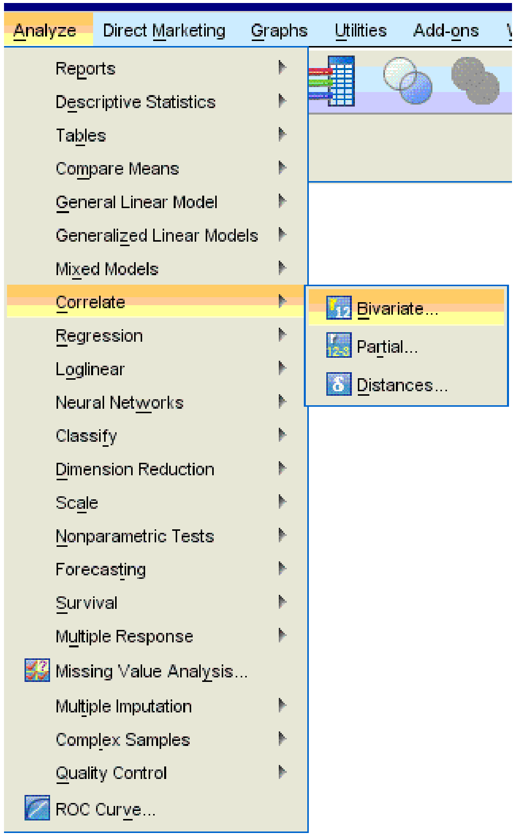
A correlation (Pearson correlation) is useful when you want to see the relationship between two (or more) normally distributed interval variables. For example, using the A data file we can run a correlation between two continuous variables, **read** and **write**.

Menu selection:- Analyze > Correlate > Bivariate

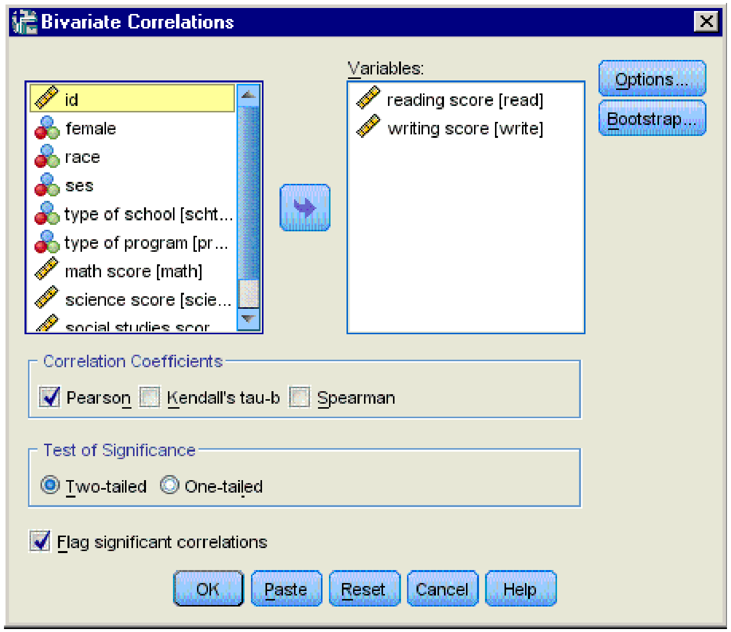
Syntax:-

```
correlations
 /variables = read write.
```

Correlation



Correlation



Correlation

Correlations

		reading score	writing score
reading score	Pearson Correlation	1	.597
	Sig. (2-tailed)		.000
	N	200	200
writing score	Pearson Correlation	.597	1
	Sig. (2-tailed)	.000	
	N	200	200

In the first example above, we see that the correlation between **read** and **write** is 0.597. By squaring the correlation and then multiplying by 100, you can determine what percentage of the variability is shared, 0.597 when squared is .356409, multiplied by 100 would be 36%. Hence **read** shares about 36% of its variability with **write**.

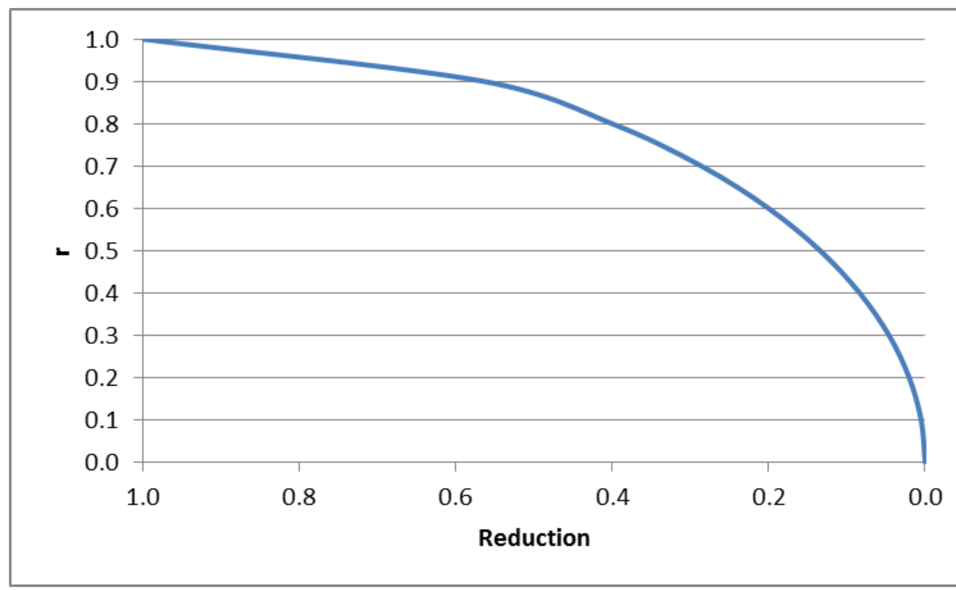
Correlation

As a rule of thumb use the following guide for the absolute value of correlation (r):

.00-.19 "very weak"
.20-.39 "weak"
.40-.59 "moderate"
.60-.79 "strong"
.80-1.0 "very strong"

Which is based on the coefficient of determination (r^2). Which indicates the proportion of variance in each of two correlated variables which is shared by both.

An index of the degree of lack of relationship is also available. It is the square root of the proportion of unexplained variance and is called the coefficient of alienation $(1-r^2)^{\frac{1}{2}}$. This in turn leads to an estimate of error reduction $1-(1-r^2)^{\frac{1}{2}}$.



A Graphic and Tabular Aid To Interpreting Correlation Coefficients
J.F. Voorhees Monthly Weather Review **54** 423 1926.

Correlation

In the second example, we will run a correlation between a dichotomous variable, **female**, and a continuous variable, **write**. Although it is assumed that the variables are interval and normally distributed, we can include dummy variables when performing correlations.

Menu selection:- Analyze > Correlate > Bivariate

Syntax:-

```
correlations
 /variables =  female write.
```

Correlation

Correlations

		female	writing score
female	Pearson Correlation	1	.256
	Sig. (2-tailed)		.000
	N	200	200
writing score	Pearson Correlation	.256	1
	Sig. (2-tailed)	.000	
	N	200	200

In the output for the second example, we can see the correlation between **write** and **female** is 0.256. Squaring this number yields .065536, meaning that **female** shares approximately 6.5% of its variability with **write**.

Index End

Simple linear regression

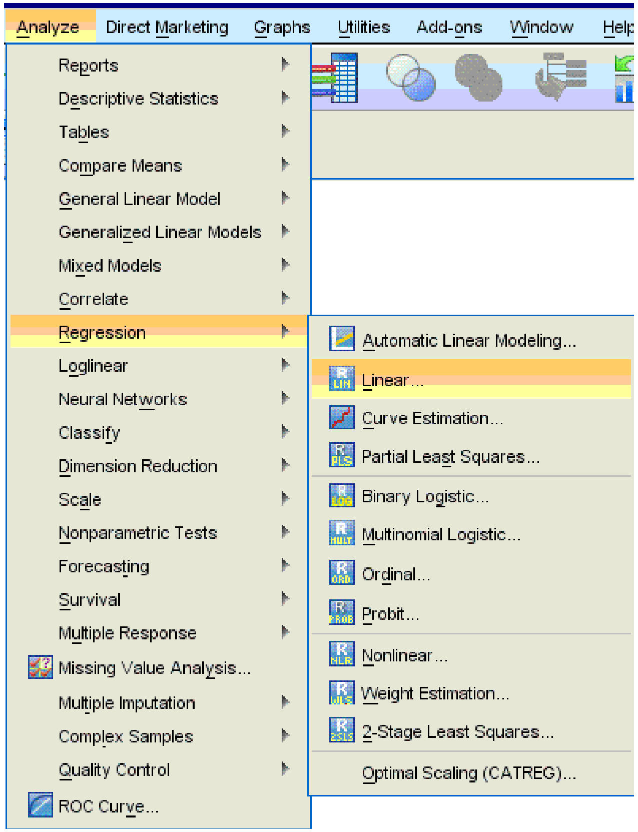
Simple linear regression allows us to look at the linear relationship between one normally distributed interval predictor and one normally distributed interval outcome variable. For example, using the A data file, say we wish to look at the relationship between writing scores (**write**) and reading scores (**read**); in other words, predicting **write** from **read**.

Menu selection:- Analyze > Regression > Linear Regression

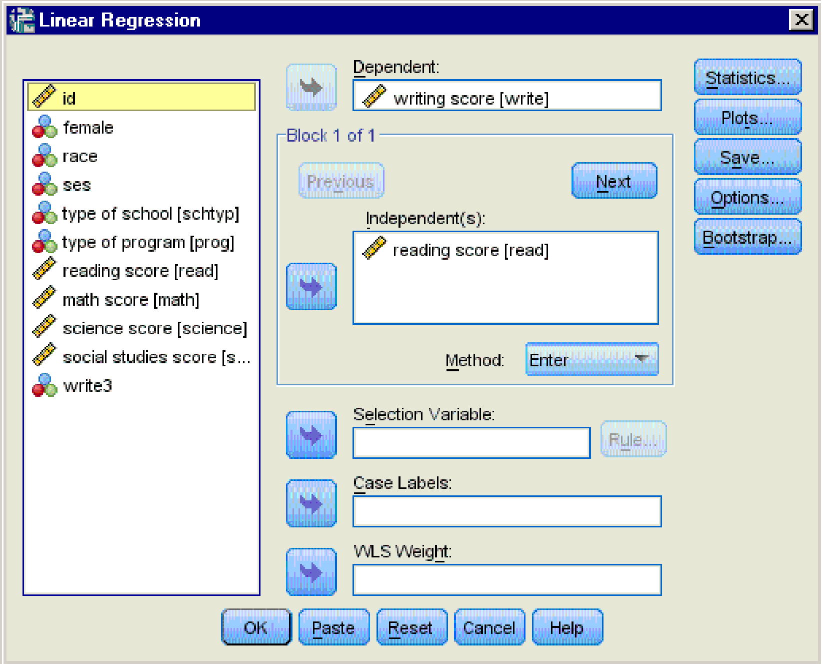
Syntax:-

```
regression variables = write read
 /dependent = write
/method = enter.
```

Simple linear regression



Simple linear regression



Simple linear regression

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	reading score	.	Enter

a. All requested variables entered.

b. Dependent Variable: writing score

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.597[a]	.356	.353	7.62487

a. Predictors: (Constant), reading score

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6367.421	1	6367.421	109.521	.000[a]
	Residual	11511.454	198	58.139		
	Total	17878.875	199			

a. Predictors: (Constant), reading score

b. Dependent Variable: writing score

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	23.959	2.806		8.539	.000
	reading score	.552	.053	.597	10.465	.000

a. Dependent Variable: writing score

We see that the relationship between **write** and **read** is positive (.552) and based on the t-value (10.47) and p-value (<0.0005), we would conclude this relationship is statistically significant. Hence, we would say there is a statistically significant positive linear relationship between reading and writing.

Take care with in/dependent assumptions.

Index End

Non-parametric correlation

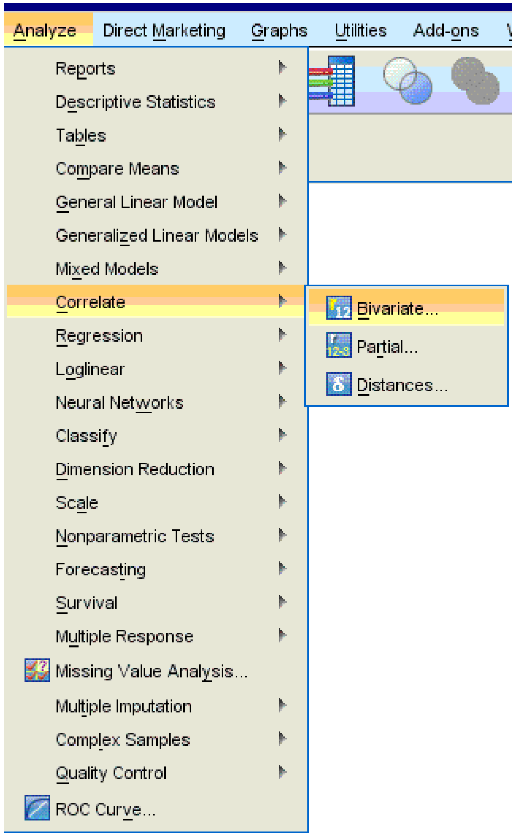
A Spearman correlation is used when one or both of the variables are not assumed to be normally distributed and interval (but are assumed to be ordinal). The values of the variables are converted to ranks and then correlated. In our example, we will look for a relationship between **read** and **write**. We will not assume that both of these variables are normal and interval.

Menu selection:- Analyze > Correlate > Bivariate

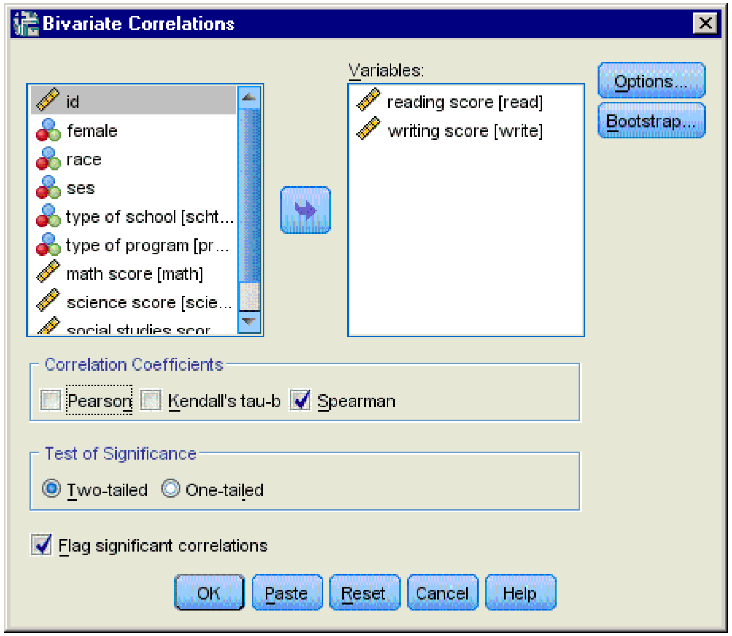
Syntax:-

```
nonpar corr
 /variables = read write
 /print = spearman.
```

Non-parametric correlation



Non-parametric correlation



Non-parametric correlation

Correlations

			reading score	writing score
Spearman's rho	reading score	Correlation Coefficient	1.000	.617
		Sig. (2-tailed)	.	.000
		N	200	200
	writing score	Correlation Coefficient	.617	1.000
		Sig. (2-tailed)	.000	.
		N	200	200

The results suggest that the relationship between **read** and **write** (ρ = 0.617, $p < 0.0005$) is statistically significant.

Non-parametric correlation

Spearman's correlation works by calculating Pearson's correlation on the ranked values of this data. Ranking (from low to high) is obtained by assigning a rank of 1 to the lowest value, 2 to the next lowest and so on. Thus the p value is only "correct" if there are no ties in the data. In the event that ties occur an exact calculation should be employed. SPSS does not do this. However the estimated value is usually reliable enough.

Comparison Of Values Of Pearson's And Spearman's Correlation Coefficients On The Same Sets Of Data
Jan Hauke, Tomasz Kossowski
Quaestiones Geographicae 30(2) 87-93 2011

Index End

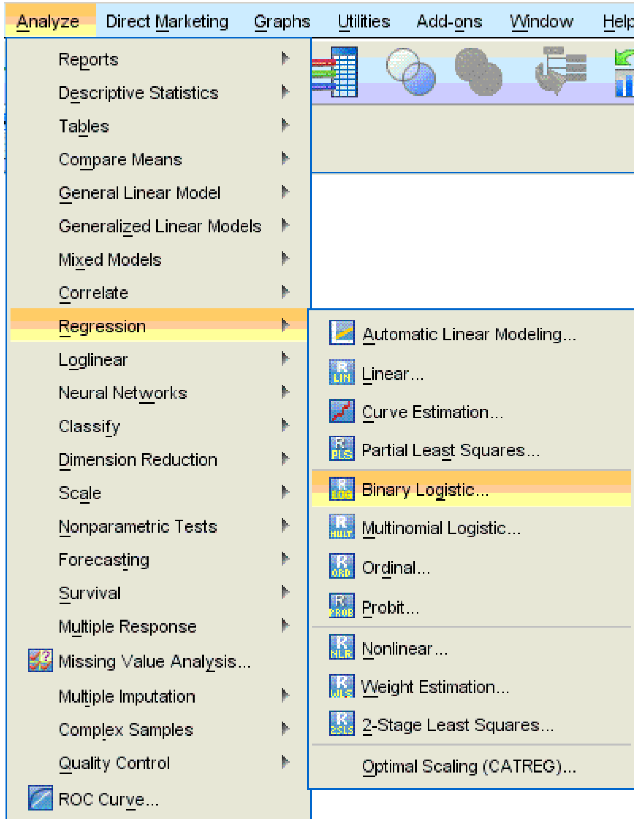
Simple logistic regression

Logistic regressior
. is binary (i.e.,
coded as 0 and 1).
data file that is
coded 0 and 1, and
t **female** is a silly
outcome variable (
is a predictor
variable), but we c
ble to illustrate
how the code for t
to interpret the
output. The first v
mand is the
outcome (or depen
f the variables
are predictor (or i
ble, **female** will be
the outcome variat
ariable. As with
ordinary least squc
bles must be
either dichotomou:
gorical.

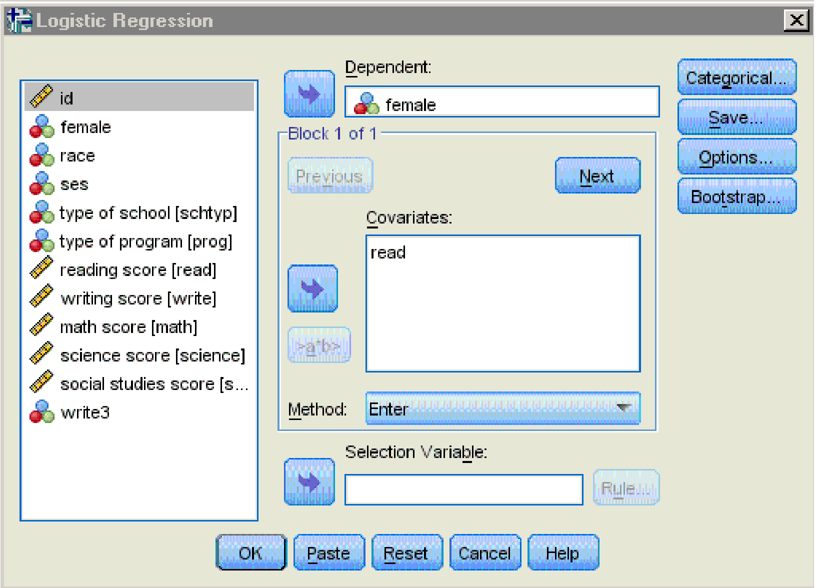
Menu selection:- Analyze > Regression > Binary Logistic

Syntax:- logistic regression female with read.

Simple logistic regression



Simple logistic regression



Simple logistic regression

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	200	100.0
	Missing Cases	0	.0
	Total	200	100.0
Unselected Cases		0	.0
Total		200	100.0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
male	0
female	1

Block 0: Beginning Block

Classification Table[a,b]

			Predicted		
			female		Percentage Correct
	Observed		male	female	
Step 0	female	male	0	91	.0
		female	0	109	100.0
	Overall Percentage				54.5

a. Constant is included in the model.

b. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	.180	.142	1.616	1	.204	1.198

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	read	.564	1	.453
	Overall Statistics		.564	1	.453

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	.564	1	.453
	Block	.564	1	.453
	Model	.564	1	.453

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	275.073[a]	.003	.004

a. Estimation terminated at iteration number 3 because parameter estimates changed by less than .001.

Classification Table[a]

			Predicted		
			female		Percentage Correct
	Observed		male	female	
Step 1	female	male	4	87	4.4
		female	5	104	95.4
	Overall Percentage				54.0

a. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	read	-.010	.014	.562	1	.453	.990
	Constant	.726	.742	.958	1	.328	2.067

a. Variable(s) entered on step 1: read.

Simple logistic regression

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	.564	1	.453
	Block	.564	1	.453
	Model	.564	1	.453

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	275.073[a]	.003	.004

a. Estimation terminated at iteration number 3 because parameter estimates changed by less than .001.

Classification Table[a]

Observed			Predicted: female – male	Predicted: female – female	Predicted: Percentage Correct
Step 1	female	male	4	87	4.4
		female	5	104	95.4
	Overall Percentage				54.0

a. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	read	-.010	.014	.562	1	.453	.990
	Constant	.726	.742	.958	1	.328	2.067

a. Variable(s) entered on step 1: read.

The results indicate that reading score (**read**) is not a statistically significant predictor of gender (i.e., being female), Wald = 0.562, p = 0.453. Likewise, the test of the overall model is not statistically significant, likelihood ratio Chi-squared = 0.564, p = 0.453.

Index End

Multiple regression

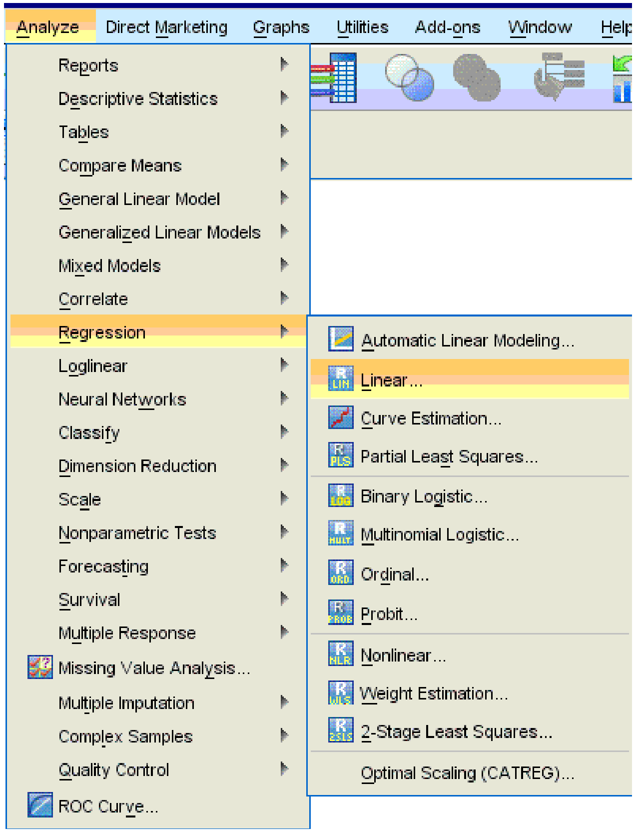
Multiple regression is very similar to simple regression, except that in multiple regression you have more than one predictor variable in the equation. For example, using the A data file we will predict writing score from gender (**female**), reading, math, science and social studies (**socst**) scores.

Menu selection:- Analyze > Regression > Linear Regression

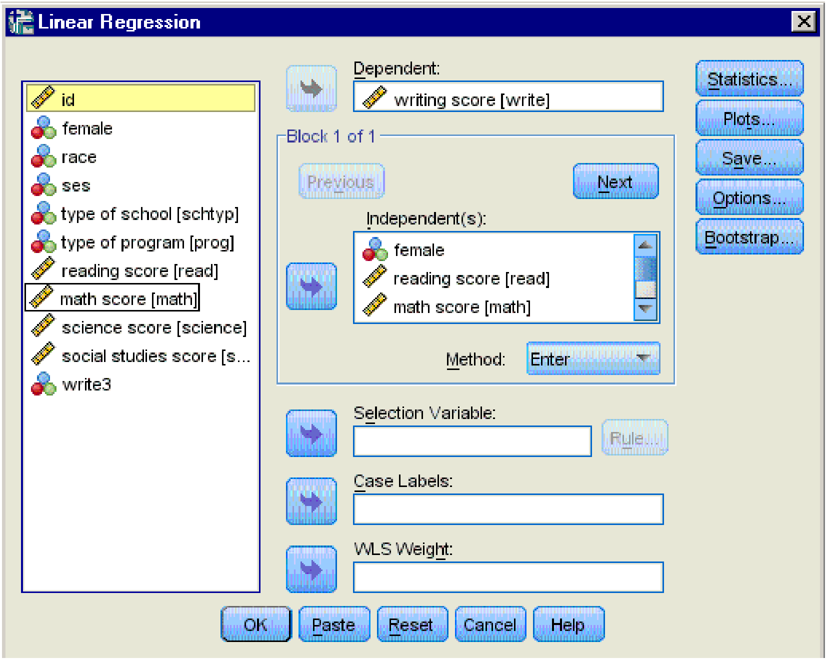
Syntax:-

```
regression variable = write female read math science socst
 /dependent = write
 /method = enter.
```

Multiple regression



Multiple regression



Note additional independent variables within box

Multiple regression

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	social studies score, female, science score, math score, reading score	.	Enter

a. All requested variables entered.
b. Dependent Variable: writing score

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.776[a]	.602	.591	6.05897

a. Predictors: (Constant), social studies score, female, science score, math score, reading score

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	10756.924	5	2151.385	58.603	.000[a]
	Residual	7121.951	194	36.711		
	Total	17878.875	199			

a. Predictors: (Constant), social studies score, female, science score, math score, reading score
b. Dependent Variable: writing score

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	6.139	2.808		2.186	.030
	female	5.493	.875	.289	6.274	.000
	reading score	.125	.065	.136	1.931	.055
	math score	.238	.067	.235	3.547	.000
	science score	.242	.061	.253	3.986	.000
	social studies score	.229	.053	.260	4.339	.000

a. Dependent Variable: writing score

The results indicate that the overall model is statistically significant ($F = 58.60$, $p < 0.0005$). Furthermore, all of the predictor variables are statistically significant except for **read**.

Multiple regression – Alternatives

There are problems with stepwise model selection procedures, these notes are a health warning.

Various algorithms have been developed for aiding in model selection. Many of them are “automatic”, in the sense that they have a “stopping rule” (which it might be possible for the researcher to set or change from a default value) based on criteria such as value of a t-statistic or an F-statistic. Others might be better termed “semi-automatic,” in the sense that they automatically list various options and values of measures that might be used to help evaluate them.

Caution: Different regression software may use the same name (e.g., “Forward Selection” or “Backward Elimination”) to designate different algorithms. Be sure to read the documentation to know find out just what the algorithm does in the software you are using - in particular, whether it has a stopping rule or is of the “semi-automatic” variety.

Multiple regression – Alternatives

The reasons for not using a stepwise procedure are as follows. There is a great deal of arbitrariness in the procedures. Forwards and backwards stepwise methods will in general give different “best models”. There are differing criteria for accepting or rejecting a variable at any stage and also for when to stop and declare the current model “best”.

The process gives a false impression of statistical sophistication. Often a complex stepwise analysis is presented, when no proper thought has been given to the real issues involved.

Multiple regression – Alternatives

Stepwise regressions are nevertheless important for three reasons. First, to emphasise that there is a considerable problem in choosing a model out of so many, so considerable that a variety of automated procedures have been devised to "help". Second to show that while purely statistical methods of choice can be constructed, they are unsatisfactory. And third, because they are fairly popular ways of avoiding constructive thinking about model selection, you may well come across them. You should know that they exist and roughly how they work.

Stepwise regressions probably do have a useful role to play, when there are large numbers of x-variables, when all prior information is taken carefully into account in inclusion/exclusion of variables, and when the results are used as a preliminary sifting of the many x-variables. It would be rare for a stepwise regression to produce convincing evidence for or against a scientific hypothesis.

Multiple regression – Alternatives

"... perhaps the most serious source of error lies in letting statistical procedures make decisions for you."

Good P.I. and Hardin J.W., Common Errors in Statistics (and How to Avoid Them), 4th Edition, Wiley, 2012, p. 3.

"Don't be too quick to turn on the computer. By passing the brain to compute by reflex is a sure recipe for disaster."

Good P.I. and Hardin J.W., Common Errors in Statistics (and How to Avoid Them), 4th Edition, Wiley, 2012, p. 152.

Multiple regression - Alternatives

"We do not recommend such stopping rules for routine use since they can reject perfectly reasonable sub-models from further consideration. Stepwise procedures are easy to explain, inexpensive to compute, and widely used. The comparative simplicity of the results from stepwise regression with model selection rules appeals to many analysts. But, such algorithmic model selection methods must be used with caution."

Cook R.D. and Weisberg S., Applied Regression Including Computing and Graphics, Wiley, 1999, p. 280.

Multiple regression – Alternatives

In a large world where parameters need to be estimated from small or unreliable samples, the function between predictive accuracy and the flexibility of a model (e.g., number of free parameters) is typically inversely U shaped. Both too few and too many parameters can hurt performance (Pitt et al. 2002). Competing models of strategies should be tested for their predictive ability, not their ability to fit already known data.

Pitt M.A., Myung I.J. and Zhang S. 2002. “Toward a method for selecting among computational models for cognition” Psychol. Rev. **109** 472–491.

Multiple regression – Alternatives

What strategies might we adopt?

Heuristics are a subset of strategies; strategies also include complex regression or Bayesian models. The part of the information that is ignored is covered by Shah and Oppenheimer's (2008) list of five aspects. The goal of making judgments more quickly and frugally is consistent with the goal of effort reduction, where "frugal" is often measured by the number of cues that a heuristic searches.

Multiple regression - Alternatives

Many definitions of heuristics exist Shah and Oppenheimer (2008) proposed that all heuristics rely on effort reduction by one or more of the following:

(a) examining fewer cues,
(b) reducing the effort of retrieving cue values,
(c) simplifying the weighting of cues,
(d) integrating less information,
(e) examining fewer alternatives.

Shah, A.K. and Oppenheimer, D.M. 2008 “Heuristics Made Easy: An Effort-Reduction Framework” Psychological Bulletin 134(2) 207-222 PsycNET

Index End

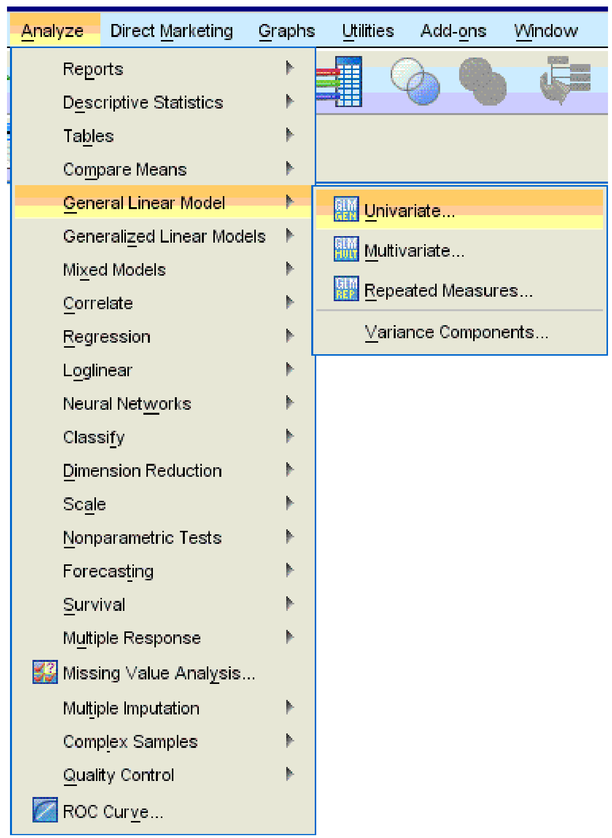
Analysis of covariance

Analysis of covariance is like ANOVA, except in addition to the categorical predictors you also have continuous predictors as well. For example, the one way ANOVA example used **write** as the dependent variable and **prog** as the independent variable. Let's add **read** as a continuous variable to this model, as shown below.

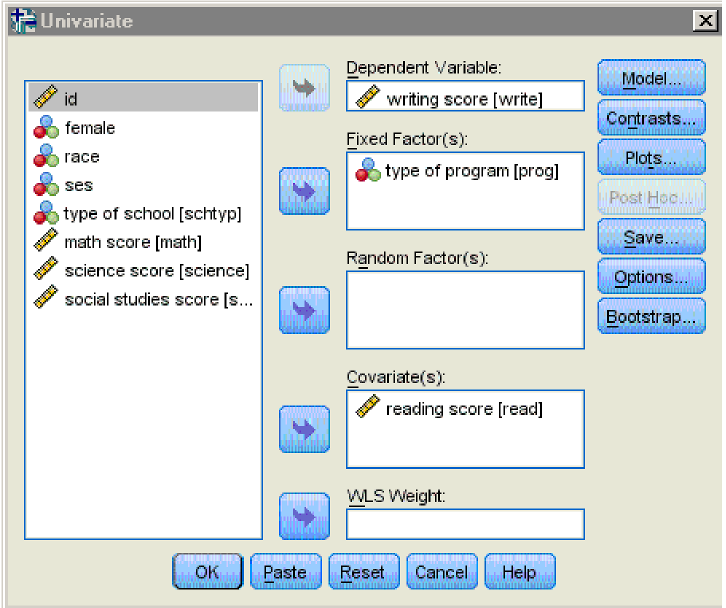
Menu selection:- Analyze > General Linear Model > Univariate

Syntax:- glm write with read by prog.

Analysis of covariance



Analysis of covariance



Analysis of covariance

Between-Subjects Factors

		Value Label	N
type of program	1.00	general	45
	2.00	academic	105
	3.00	vocation	50

Tests of Between-Subjects Effects

Dependent Variable:writing score

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	7017.681[a]	3	2339.227	42.213	.000
Intercept	4867.964	1	4867.964	87.847	.000
read	3841.983	1	3841.983	69.332	.000
prog	650.260	2	325.130	5.867	.003
Error	10861.194	196	55.414		
Total	574919.000	200			
Corrected Total	17878.875	199			

a. R Squared = 0393 (Adjusted R Squared = 0383)

The results indicate that even after adjusting for reading score (**read**), writing scores still significantly differ by program type (**prog**), $F = 5.867$, $p = 0.003$.

Index End

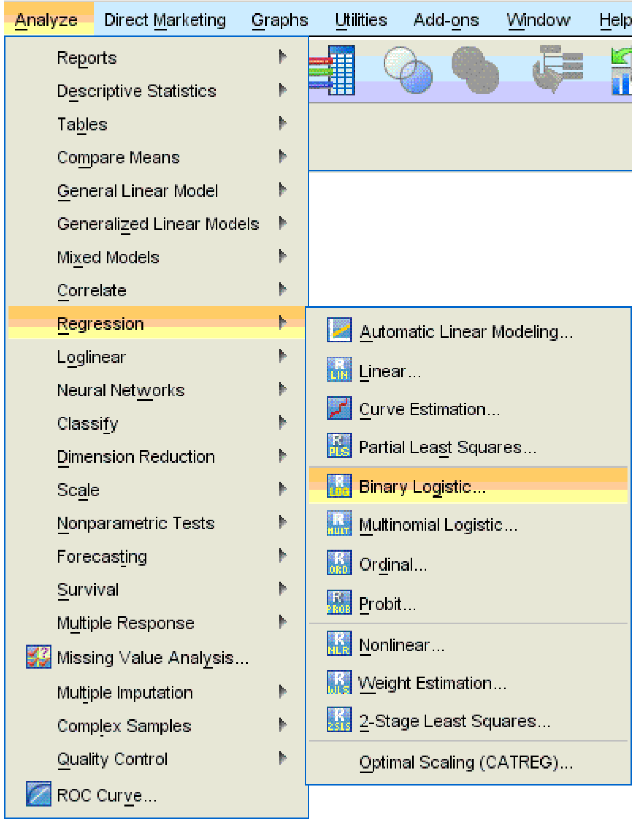
Multiple logistic regression

Multiple logistic regression is like simple logistic regression, except that there are two or more predictors. The predictors can be interval variables or dummy variables, but cannot be categorical variables. If you have categorical predictors, they should be coded into one or more dummy variables. We have only one variable in our data set that is coded 0 and 1, and that is **female**. We understand that **female** is a silly outcome variable (it would make more sense to use it as a predictor variable), but we can use **female** as the outcome variable to illustrate how the code for this command is structured and how to interpret the output. The first variable listed after the **logistic regression** command is the outcome (or dependent) variable, and all of the rest of the variables are predictor (or independent) variables (listed after the keyword **with**). In our example, **female** will be the outcome variable, and **read** and **write** will be the predictor variables.

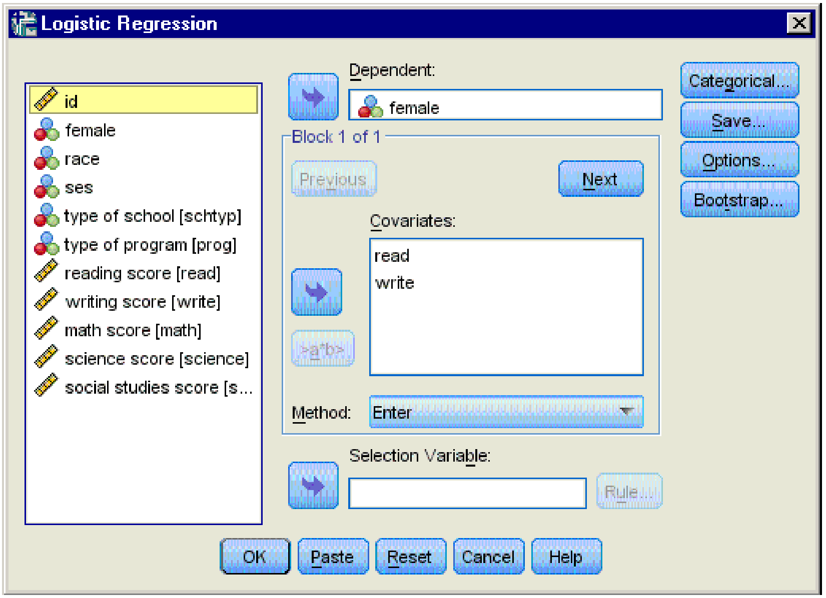
Menu selection:- Analyze > Regression > Binary Logistic

Syntax:- logistic regression female with read write.

Multiple logistic regression



Multiple logistic regression



Multiple logistic regression

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	200	100.0
	Missing Cases	0	.0
	Total	200	100.0
Unselected Cases		0	.0
Total		200	100.0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
male	0
female	1

Block 0: Beginning Block

Classification Table[a,b]

			Predicted		
			female		Percentage Correct
	Observed		male	female	
Step 0	female	male	0	91	.0
		female	0	109	100.0
	Overall Percentage				54.5

a. Constant is included in the model.

b. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	.180	.142	1.616	1	.204	1.198

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	read	.564	1	.453
		write	13.158	1	.000
	Overall Statistics		26.359	2	.000

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	27.819	2	.000
	Block	27.819	2	.000
	Model	27.819	2	.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	247.818[a]	.130	.174

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

Classification Table[a]

			Predicted		
			female		Percentage Correct
	Observed		male	female	
Step 1	female	male	54	37	59.3
		female	30	79	72.5
	Overall Percentage				66.5

a. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	read	-.071	.020	13.125	1	.000	.931
	write	.106	.022	23.075	1	.000	1.112
	Constant	-1.706	.923	3.414	1	.065	.182

a. Variable(s) entered on step 1: read, write.

Multiple logistic regression

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	read	-.071	.020	13.125	1	.000	.931
	write	.106	.022	23.075	1	.000	1.112
	Constant	-1.706	.923	3.414	1	.065	.182

a. Variable(s) entered on step 1: read, write.

These results show that both **read** and **write** are significant predictors of **female**.

Index End

Discriminant analysis

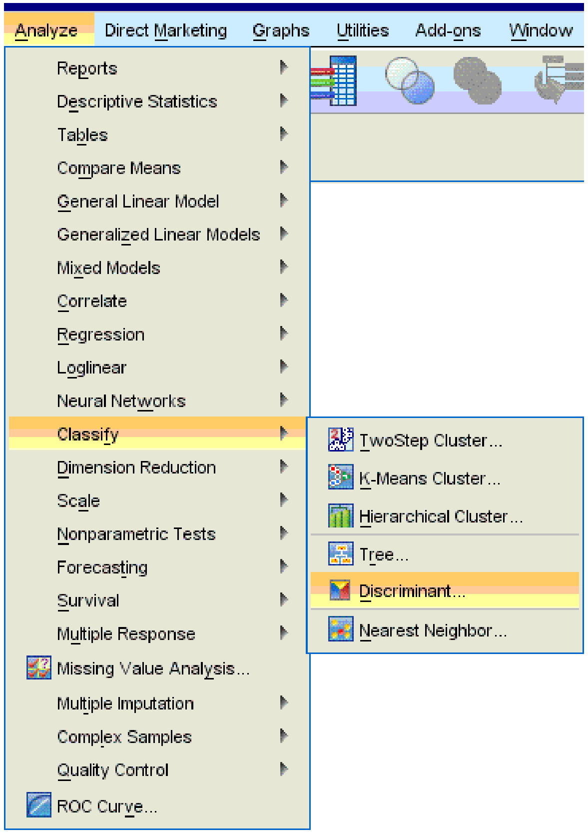
Discriminant analysis is used when you have one or more normally distributed interval independent variable(s) and a categorical dependent variable. It is a multivariate technique that considers the latent dimensions in the independent variables for predicting group membership in the categorical dependent variable. For example, using the A data file, say we wish to use **read**, **write** and **math** scores to predict the type of program a student belongs to (**prog**).

Menu selection:- Analyze > Classify > Discriminant

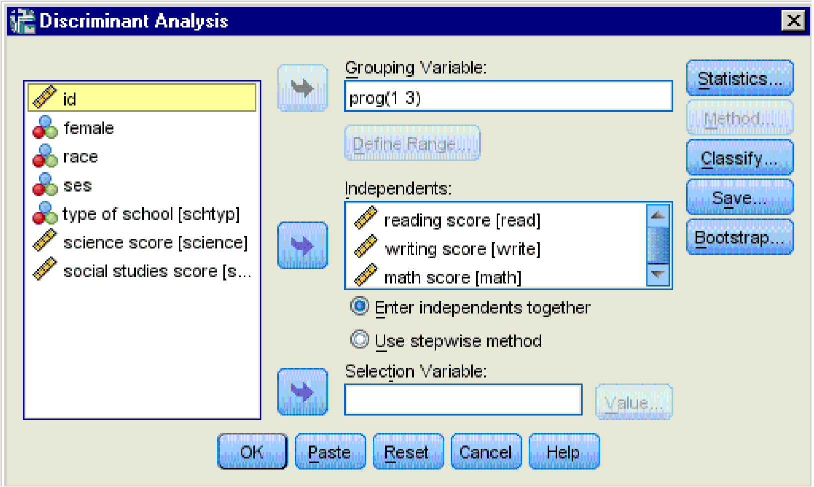
Syntax:-

```
Discriminant groups = prog(1, 3)
 /variables = read write math.
```

Discriminant analysis



Discriminant analysis



Do not forget to define the range for Prog.

Discriminant analysis

Analysis Case Processing Summary

Unweighted Cases		N	Percent
Valid		200	100.0
Excluded	Missing or out-of-range group codes	0	.0
	At least one missing discriminating variable	0	.0
	Both missing or out-of-range group codes and at least one missing discriminating variable	0	.0
	Total	0	.0
Total		200	100.0

Group Statistics

type of program		Valid N (listwise) Unweighted	Weighted
general	reading score	45	45.000
	writing score	45	45.000
	math score	45	45.000
academic	reading score	105	105.000
	writing score	105	105.000
	math score	105	105.000
vocation	reading score	50	50.000
	writing score	50	50.000
	math score	50	50.000
Total	reading score	200	200.000
	writing score	200	200.000
	math score	200	200.000

Analysis 1

Summary of Canonical Discriminant Functions

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	.356[a]	98.7	98.7	.513
2	.005[a]	1.3	100.0	.067

a. First 2 canonical discriminant functions were used in the analysis.

Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1 through 2	.734	60.619	6	.000
2	.995	.888	2	.641

Standardized Canonical Discriminant Function Coefficients

	Function 1	2
reading score	.273	-.410
writing score	.331	1.183
math score	.582	-.656

Structure Matrix

	Function 1	2
math score	.913*	-.272
reading score	.778*	-.184
writing score	.775*	.630

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions

Variables ordered by absolute size of correlation within function.

*. Largest absolute correlation between each variable and any discriminant function

Functions at Group Centroids

type of program	Function 1	2
general	-.312	.119
academic	.536	-.020
vocation	-.844	-.066

Unstandardized canonical discriminant functions evaluated at group means

Discriminant analysis

Functions at Group Centroids

type of program	Function	
	1	2
general	-.312	.119
academic	.536	-.020
vocation	-.844	-.066

Unstandardized canonical discriminant functions evaluated at group means

Clearly, the SPSS output for this procedure is quite lengthy, and it is beyond the scope of this page to explain all of it. However, the main point is that two canonical variables are identified by the analysis, the first of which seems to be more related to program type than the second.

Index End

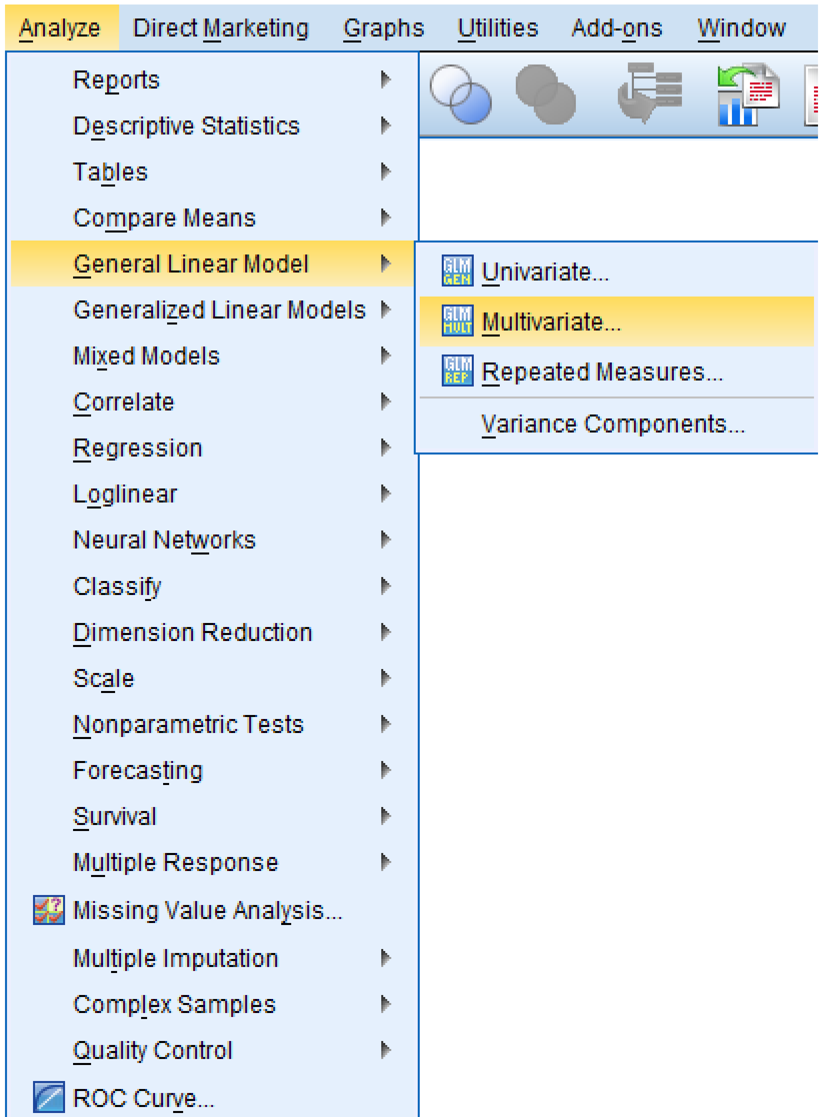
One-way MANOVA

MANOVA (multivariate analysis of variance) is like ANOVA, except that there are two or more dependent variables. In a one-way MANOVA, there is one categorical independent variable and two or more dependent variables. For example, using the A data file, say we wish to examine the differences in **read**, **write** and **math** broken down by program type (**prog**).

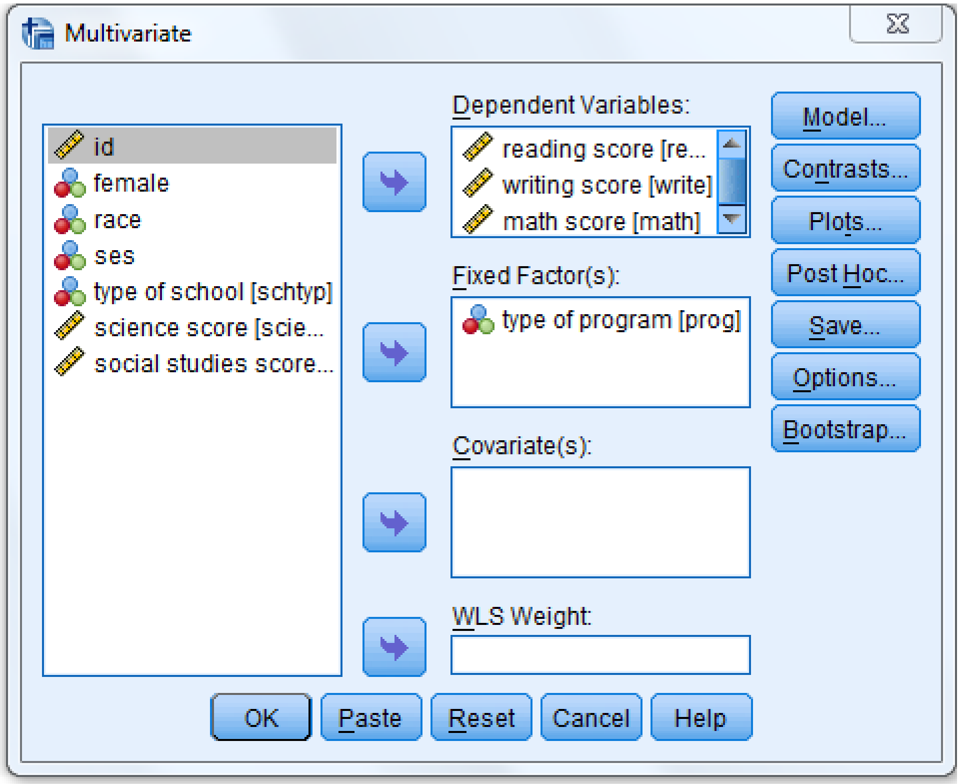
Menu selection:- Analyse > General Linear Model > Multivariate

Syntax:- glm read write math by prog.

One-way MANOVA



One-way MANOVA



One-way MANOVA

Between-Subjects Factors

		Value Label	N
type of program	1.00	general	45
	2.00	academic	105
	3.00	vocation	50

Multivariate Tests[c]

Effect		Value	F	Hypothesis df	Error df	Sig.
Intercept	Pillai's Trace	.978	2883.051[a]	3.000	195.000	.000
	Wilks' Lambda	.022	2883.051[a]	3.000	195.000	.000
	Hotelling's Trace	44.355	2883.051[a]	3.000	195.000	.000
	Roy's Largest Root	44.355	2883.051[a]	3.000	195.000	.000
prog	Pillai's Trace	.267	10.075	6.000	392.000	.000
	Wilks' Lambda	.734	10.870[a]	6.000	390.000	.000
	Hotelling's Trace	.361	11.667	6.000	388.000	.000
	Roy's Largest Root	.356	23.277[b]	3.000	196.000	.000

a. Exact statistic
b. The statistic is an upper bound on F that yields a lower bound on the significance level.
c. Design: Intercept + prog

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square
Corrected Model	reading score	3716.861[a]	2	1858.431
	writing score	3175.698[b]	2	1587.849
	math score	4002.104[c]	2	2001.052
Intercept	reading score	447178.672	1	447178.672
	writing score	460403.797	1	460403.797
	math score	453421.258	1	453421.258
prog	reading score	3716.861	2	1858.431
	writing score	3175.698	2	1587.849
	math score	4002.104	2	2001.052
Error	reading score	17202.559	197	87.323
	writing score	14703.177	197	74.635
	math score	13463.691	197	68.344
Total	reading score	566514.000	200	
	writing score	574919.000	200	
	math score	571765.000	200	
Corrected Total	reading score	20919.420	199	
	writing score	17878.875	199	
	math score	17465.795	199	

One-way MANOVA

Tests of Between-Subjects Effects

Source	Dependent Variable	F	Sig.
Corrected Model	reading score	21.282	.000
	writing score	21.275	.000
	math score	29.279	.000
Intercept	reading score	5120.994	.000
	writing score	6168.704	.000
	math score	6634.435	.000
prog	reading score	21.282	.000
	writing score	21.275	.000
	math score	29.279	.000
Error	reading score		
	writing score		
	math score		
Total	reading score		
	writing score		
	math score		
Corrected Total	reading score		
	writing score		
	math score		

a. R Squared = 0178 (Adjusted R Squared = 0169)
b. R Squared = 0178 (Adjusted R Squared = 0169)
c. R Squared = 0229 (Adjusted R Squared = 0221)

Concluding output table.

The students in the different programs differ in their joint distribution of **read**, **write** and **math**.

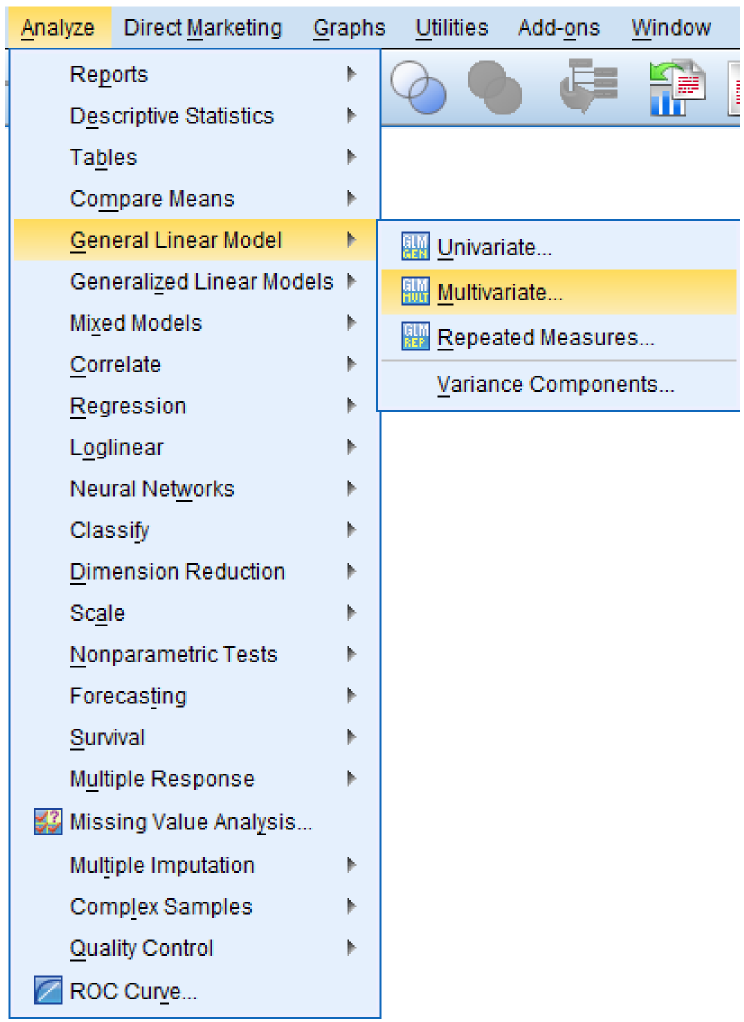
Multivariate multiple regression

Multivariate multiple regression is used when you have two or more dependent variables that are to be predicted from two or more independent variables. In our example, we will predict **write** and **read** from **female**, **math**, **science** and social studies (**socst**) scores.

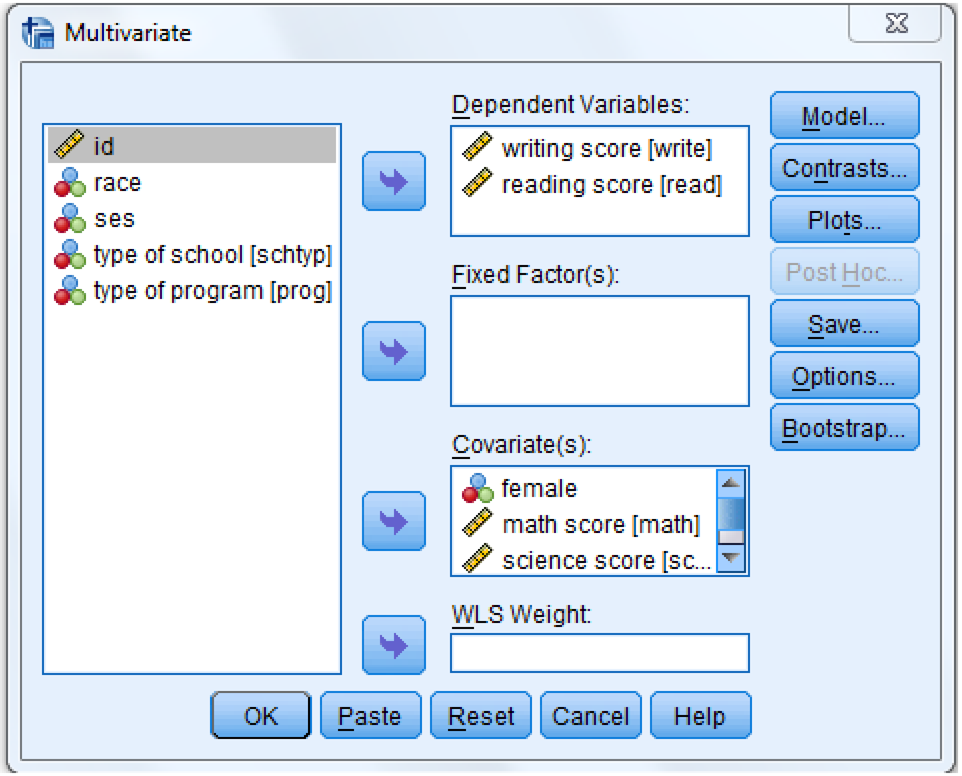
Menu selection:- Analyse > General Linear Model > Multivariate

Syntax:- glm write read with female math science socst.

Multivariate multiple regression



Multivariate multiple regression



Multivariate multiple regression

Multivariate Tests[b]

Effect		Value	F	Hypothesis df	Error df	Sig.
Intercept	Pillai's Trace	.030	3.019[a]	2.000	194.000	.051
	Wilks' Lambda	.970	3.019[a]	2.000	194.000	.051
	Hotelling's Trace	.031	3.019[a]	2.000	194.000	.051
	Roy's Largest Root	.031	3.019[a]	2.000	194.000	.051
female	Pillai's Trace	.170	19.851[a]	2.000	194.000	.000
	Wilks' Lambda	.830	19.851[a]	2.000	194.000	.000
	Hotelling's Trace	.205	19.851[a]	2.000	194.000	.000
	Roy's Largest Root	.205	19.851[a]	2.000	194.000	.000
math	Pillai's Trace	.160	18.467[a]	2.000	194.000	.000
	Wilks' Lambda	.840	18.467[a]	2.000	194.000	.000
	Hotelling's Trace	.190	18.467[a]	2.000	194.000	.000
	Roy's Largest Root	.190	18.467[a]	2.000	194.000	.000
science	Pillai's Trace	.166	19.366[a]	2.000	194.000	.000
	Wilks' Lambda	.834	19.366[a]	2.000	194.000	.000
	Hotelling's Trace	.200	19.366[a]	2.000	194.000	.000
	Roy's Largest Root	.200	19.366[a]	2.000	194.000	.000
socst	Pillai's Trace	.221	27.466[a]	2.000	194.000	.000
	Wilks' Lambda	.779	27.466[a]	2.000	194.000	.000
	Hotelling's Trace	.283	27.466[a]	2.000	194.000	.000
	Roy's Largest Root	.283	27.466[a]	2.000	194.000	.000

a. Exact statistic

b. Design: Intercept + female + math + science + socst

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square
Corrected Model	writing score	10620.092[a]	4	2655.023
	reading score	12219.658[b]	4	3054.915
Intercept	writing score	202.117	1	202.117
	reading score	55.107	1	55.107
female	writing score	1413.528	1	1413.528
	reading score	12.605	1	12.605
math	writing score	714.867	1	714.867
	reading score	1025.673	1	1025.673
science	writing score	857.882	1	857.882
	reading score	946.955	1	946.955
socst	writing score	1105.653	1	1105.653
	reading score	1475.810	1	1475.810
Error	writing score	7258.783	195	37.225
	reading score	8699.762	195	44.614
Total	writing score	574919.000	200	
	reading score	566514.000	200	
Corrected Total	writing score	17878.875	199	
	reading score	20919.420	199	

Multivariate multiple regression

Tests of Between-Subjects Effects

Source	Dependent Variable	F	Sig.
Corrected Model	writing score	71.325	.000
	reading score	68.474	.000
Intercept	writing score	5.430	.021
	reading score	1.235	.268
female	writing score	37.973	.000
	reading score	.283	.596
math	writing score	19.204	.000
	reading score	22.990	.000
science	writing score	23.046	.000
	reading score	21.225	.000
socst	writing score	29.702	.000
	reading score	33.079	.000
Error	writing score		
	reading score		
Total	writing score		
	reading score		
Corrected Total	writing score		
	reading score		

a. R Squared = 0594 (Adjusted R Squared = 0586)

b. R Squared = 0584 (Adjusted R Squared = 0576)

Concluding table.

These results show that all of the variables in the model have a statistically significant relationship with the joint distribution of **write** and **read**.

Index End

Canonical correlation

Canonical correlation is a multivariate technique used to examine the relationship between two groups of variables. For each set of variables, it creates latent variables and looks at the relationships among the latent variables. It assumes that all variables in the model are interval and normally distributed. SPSS requires that each of the two groups of variables be separated by the keyword **with**. There need not be an equal number of variables in the two groups (before and after the **with**). In this case {read, write} with {math, science}.

Canonical correlation are the correlations of two canonical (latent) variables, one representing a set of independent variables, the other a set of dependent variables. There may be more than one such linear correlation relating the two sets of variables, with each correlation representing a different dimension by which the independent set of variables is related to the dependent set. The purpose of the method is to explain the relation of the two sets of variables, not to model the individual variables.

Canonical correlation

Canonical correlation analysis is the study of the linear relations between two sets of variables. It is the multivariate extension of correlation analysis.

Suppose you have given a group of students two tests of ten questions each and wish to determine the overall correlation between these two tests. Canonical correlation finds a weighted average of the questions from the first test and correlates this with a weighted average of the questions from the second test. The weights are constructed to maximize the correlation between these two averages. This correlation is called the first canonical correlation coefficient.

You can create another set of weighted averages unrelated to the first and calculate their correlation. This correlation is the second canonical correlation coefficient. This process continues until the number of canonical correlations equals the number of variables in the smallest group.

Canonical correlation

In statistics, canonical-correlation analysis is a way of making sense of cross-covariance matrices. If we have two vectors $X = (X_1, ..., X_n)$ and $Y = (Y_1, ..., Y_m)$ of random variables, and there are correlations among the variables, then canonical-correlation analysis will find linear combinations of the X_i and Y_j which have maximum correlation with each other (Härdle and Léopold 2007). T. R. Knapp notes "virtually all of the commonly encountered parametric tests of significance can be treated as special cases of canonical-correlation analysis, which is the general procedure for investigating the relationships between two sets of variables." The method was first introduced by Harold Hotelling in 1936.

Härdle, Wolfgang and Simar, Léopold (2007). "Canonical Correlation Analysis". Applied Multivariate Statistical Analysis. pp. 321–330. Canonical Correlation Analysis - Springer ISBN 978-3-540-72243-4.
Knapp, T. R. (1978). "Canonical correlation analysis: A general parametric significance-testing system". Psychological Bulletin 85(2): 410–416. PsycNET - Display Record.
Hotelling, H. (1936). "Relations Between Two Sets of Variates". Biometrika 28 (3–4): 321–377. Relations Between Two Sets Of Variates.

Canonical correlation

The manova command is one of the SPSS commands that can only be accessed via syntax; there is not a sequence of pull-down menus or point-and-clicks that could arrive at this analysis.

Syntax:-

```
manova read write with math science
/discrim all alpha(1)
/print=sig(eigen dim).
```

Canonical correlation

* * * * * * * * * * * * * * * * * * A n a l y s i s   o f   V a r i a n c e -- Design   1 * * * * * * * * * * * * * *

EFFECT .. WITHIN CELLS Regression
Multivariate Tests of Significance (S = 2, M = -1/2, N = 97)

Test Name	Value	Approx. F	Hypoth. DF	Error DF	Sig. of F
Pillais	.59783	41.99694	4.00	394.00	.000
Hotellings	1.48369	72.32964	4.00	390.00	.000
Wilks	.40249	56.47060	4.00	392.00	.000
Roys	.59728				

Note.. F statistic for WILKS' Lambda is exact.

Eigenvalues and Canonical Correlations

Root No.	Eigenvalue	Pct.	Cum. Pct.	Canon Cor.	Sq. Cor
1	1.48313	99.96283	99.96283	.77284	.59728
2	.00055	.03717	100.00000	.02348	.00055

Dimension Reduction Analysis

Roots	Wilks L.	F	Hypoth. DF	Error DF	Sig. of F
1 TO 2	.40249	56.47060	4.00	392.00	.000
2 TO 2	.99945	.10865	1.00	197.00	.742

EFFECT .. WITHIN CELLS Regression (Cont.)
Univariate F-tests with (2,197) D. F.

Variable	Sq. Mul. R	Adj. R-sq.	Hypoth. MS	Error MS	F	Sig. of F
read	.51356	.50862	5371.66966	51.65523	103.99081	.000
write	.43565	.42992	3894.42594	51.21839	76.03569	.000

The output shows the linear combinations corresponding to the first canonical correlation. At the bottom of the output are the two canonical correlations. These results indicate that the first canonical correlation is .7728.

Canonical correlation

```
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

* * * * * * * * * * * * * * * * * * A n a l y s i s   o f   V a r i a n c e -- Design   1 * * * * * * * * * * * * * * * *

 EFFECT .. WITHIN CELLS Regression
 Multivariate Tests of Significance (S = 2, M = -1/2, N = 97 )

 Test Name             Value        Approx. F       Hypoth. DF         Error DF        Sig. of F

 Pillais                .59783        41.99694             4.00           394.00             .000
 Hotellings            1.48369        72.32964             4.00           390.00             .000
 Wilks                  .40249        56.47060             4.00           392.00             .000
 Roys                   .59728
 Note.. F statistic for WILKS' Lambda is exact.

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
 Eigenvalues and Canonical Correlations

 Root No.       Eigenvalue           Pct.      Cum. Pct.      Canon Cor.         Sq. Cor

        1          1.48313       99.96283       99.96283          .77284          .59728
        2           .00055         .03717      100.00000          .02348          .00055

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
 Dimension Reduction Analysis

 Roots              Wilks L.                F        Hypoth. DF         Error DF        Sig. of F

 1 TO 2              .40249         56.47060              4.00           392.00             .000
 2 TO 2              .99945           .10865              1.00           197.00             .742

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
 EFFECT .. WITHIN CELLS Regression (Cont.)
 Univariate F-tests with (2,197) D. F.

 Variable        Sq. Mul. R      Adj. R-sq.      Hypoth. MS        Error MS             F        Sig. of F

 read               .51356          .50862      5371.66966        51.65523     103.99081             .000
 write              .43565          .42992      3894.42594        51.21839      76.03569             .000
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
```

The F-test in this output tests the hypothesis that the first canonical correlation is not equal to zero. Clearly, F = 56.4706 is statistically significant. However, the second canonical correlation of .0235 is not statistically significantly different from zero (F = 0.1087, p = 0.742).

Index End

Factor analysis

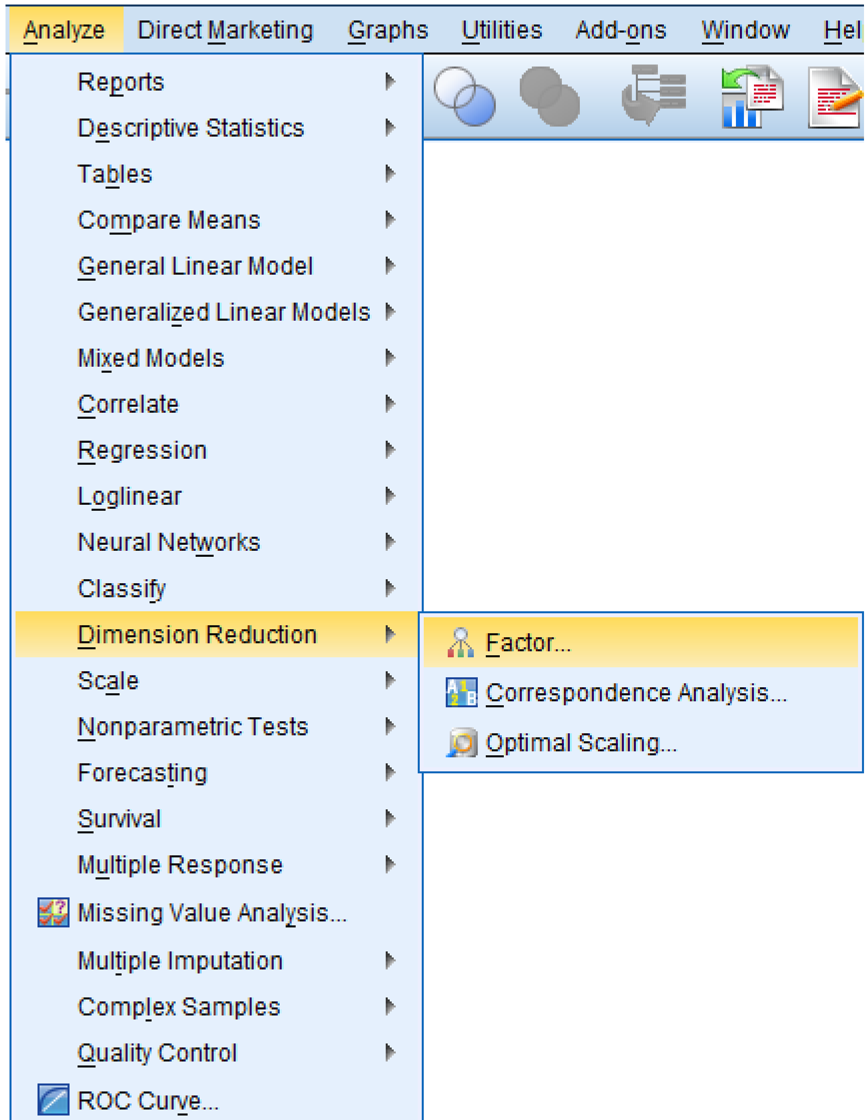
Factor analysis is a form of exploratory multivariate analysis that is used to either reduce the number of variables in a model or to detect relationships among variables. All variables involved in the factor analysis need to be interval and are assumed to be normally distributed. The goal of the analysis is to try to identify factors which underlie the variables. There may be fewer factors than variables, but there may not be more factors than variables. For our example, let's suppose that we think that there are some common factors underlying the various test scores. We will include subcommands for varimax rotation and a plot of the eigenvalues. We will use a principal components extraction and will retain two factors.

Menu selection:- Analyze > Dimension Reduction > Factor

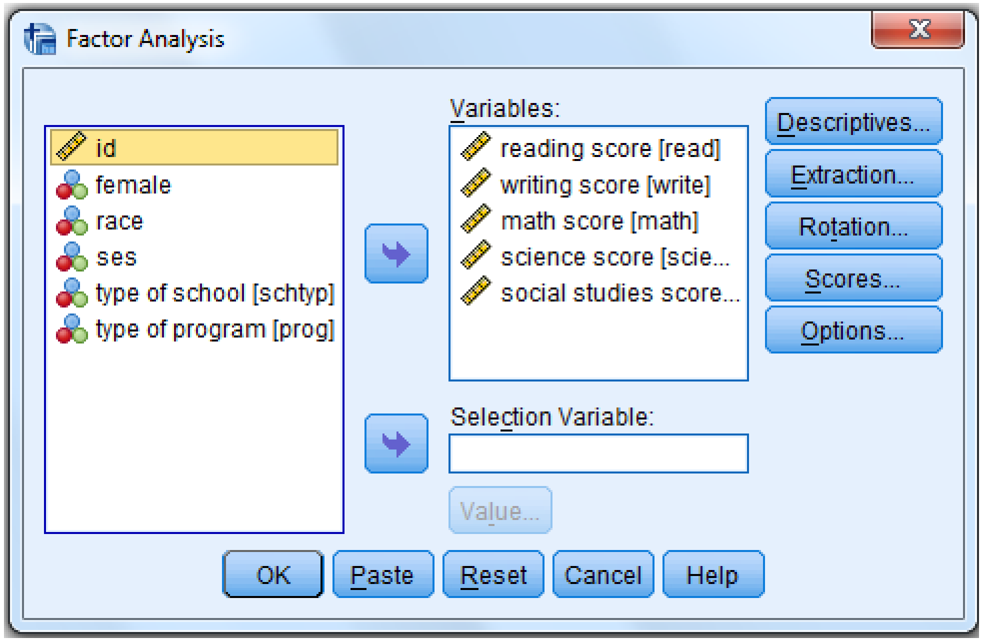
Syntax:-

```
factor
 /variables read write math science socst
 /criteria factors(2)
 /extraction pc
 /rotation varimax
 /plot eigen.
```

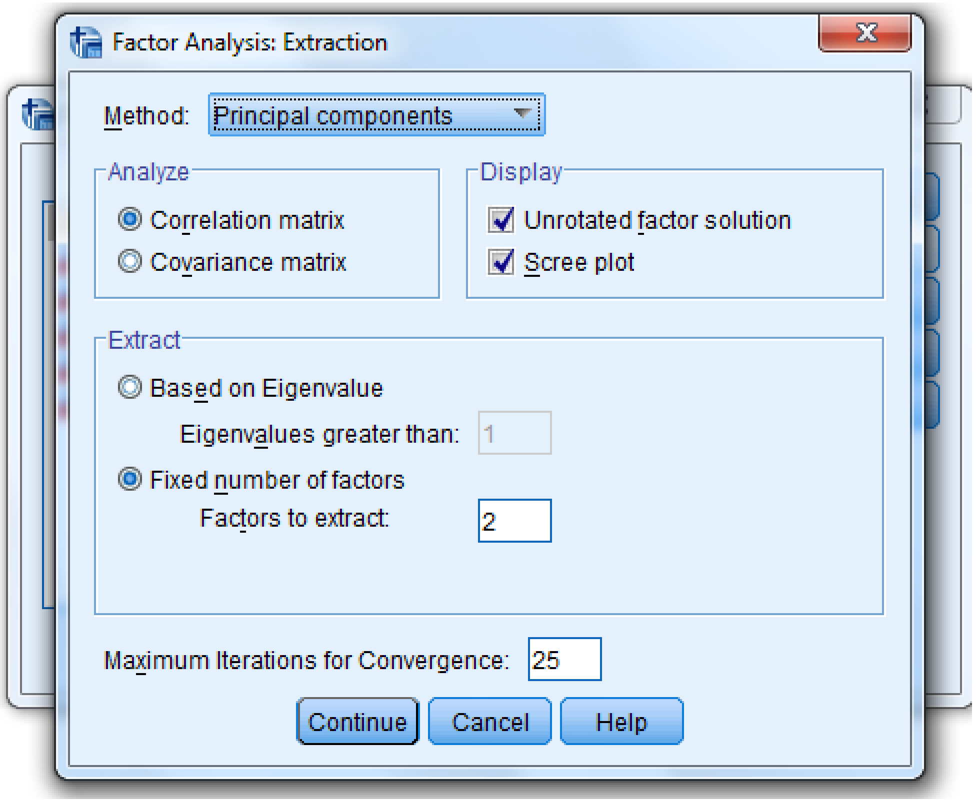
Factor analysis



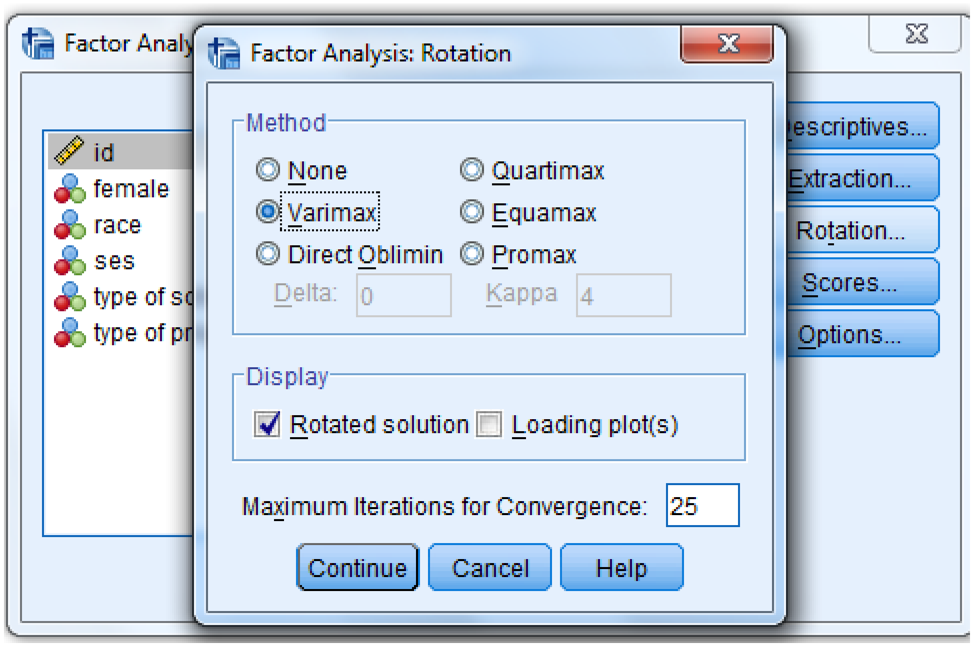
Factor analysis



Factor analysis



Factor analysis



Factor analysis

Communalities

	Initial	Extraction
reading score	1.000	.736
writing score	1.000	.704
math score	1.000	.750
science score	1.000	.849
social studies score	1.000	.900

Extraction Method: Principal Component Analysis.

Communality (which is the opposite of uniqueness) is the proportion of variance of the variable (i.e., **read**) that is accounted for by all of the factors taken together, and a very low communality can indicate that a variable may not belong with any of the factors.

Total Variance Explained

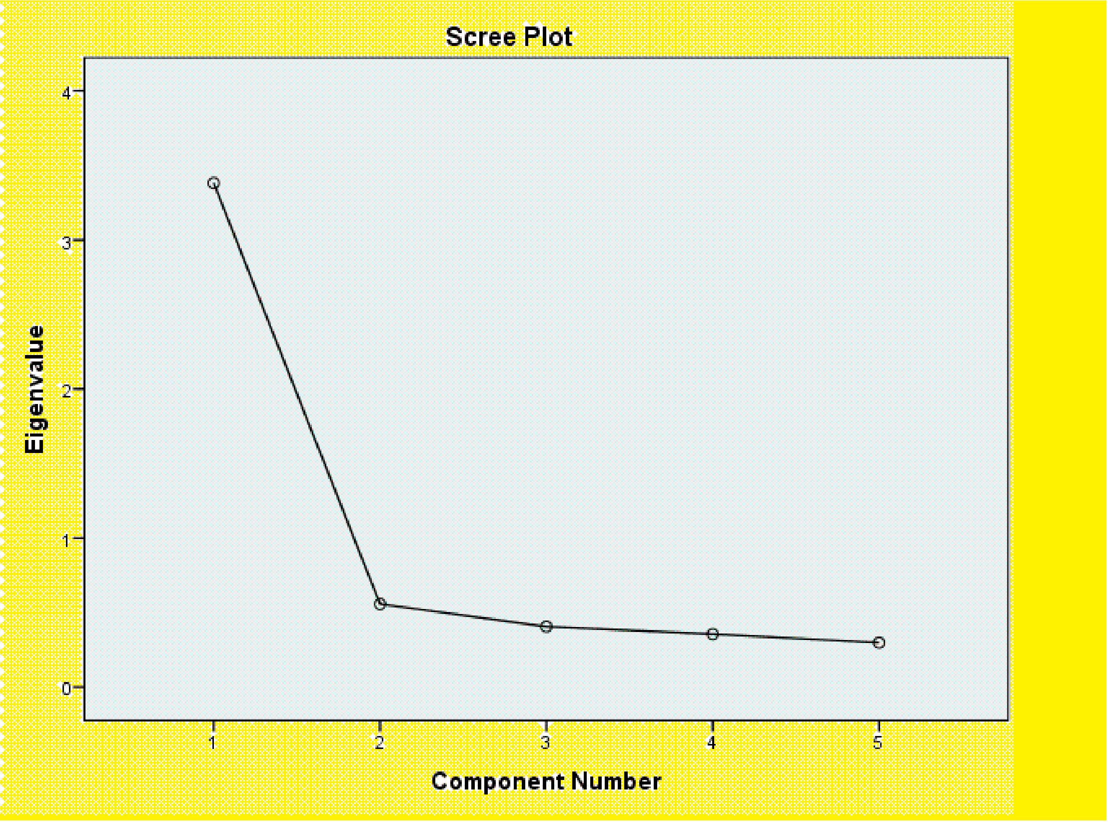
Component	Initial Eigenvalues			Extraction Sums of Squared Loadings
	Total	% of Variance	Cumulative %	Total
1	3.381	67.616	67.616	3.381
2	.557	11.148	78.764	.557
3	.407	8.136	86.900	
4	.356	7.123	94.023	
5	.299	5.977	100.000	

Total Variance Explained

Component	Extraction Sums of Squared Loadings		Rotation Sums of Squared Loadings		
	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	67.616	67.616	2.113	42.267	42.267
2	11.148	78.764	1.825	36.497	78.764
3					
4					
5					

Extraction Method: Principal Component Analysis.

Factor analysis



The scree plot may be useful in determining how many factors to retain.

Factor analysis

Component Matrix[a]

	Component	
	1	2
reading score	.858	-.020
writing score	.824	.155
math score	.844	-.195
science score	.801	-.456
social studies score	.783	.536

Extraction Method: Principal Component Analysis.
a. 2 components extracted.

Rotated Component Matrix[a]

	Component	
	1	2
reading score	.650	.559
writing score	.508	.667
math score	.757	.421
science score	.900	.198
social studies score	.222	.922

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
a. Rotation converged in 3 iterations.

Component Transformation Matrix

Component	1	2
1	.742	.670
2	-.670	.742

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

From the component matrix table, we can see that all five of the test scores load onto the first factor, while all five tend to load not so heavily on the second factor. The purpose of rotating the factors is to get the variables to load either very high or very low on each factor. In this example, because all of the variables loaded onto factor 1 and not on factor 2, the rotation did not aid in the interpretation. Instead, it made the results even more difficult to interpret.

Index End

Normal probability

Many statistical methods require that the numeric variables we are working with have an approximate normal distribution. For example, t-tests, F-tests, and regression analyses all require in some sense that the numeric variables are approximately normally distributed.

Normal probability plot

Tools for Assessing Normality include
- Histogram and Boxplot
- Normal Quantile Plot (also called Normal Probability Plot)

Goodness of Fit Tests such as
- Anderson-Darling Test
- Kolmogorov-Smirnov Test
- Lillefor's Test
- Shapiro-Wilk Test

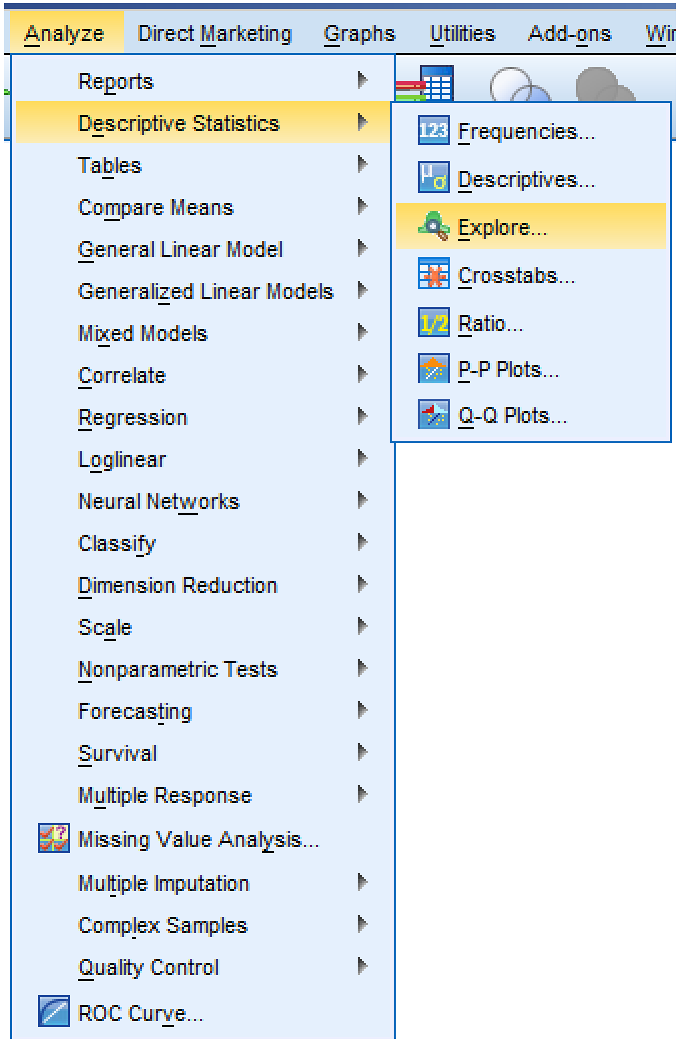
Problem: they don't always agree!

Normal probability plot

You could produce conventional descriptive statistics, a histogram with a superimposed normal curve, and a normal scores plot also called a normal probability plot.

The pulse data from data set C is employed.

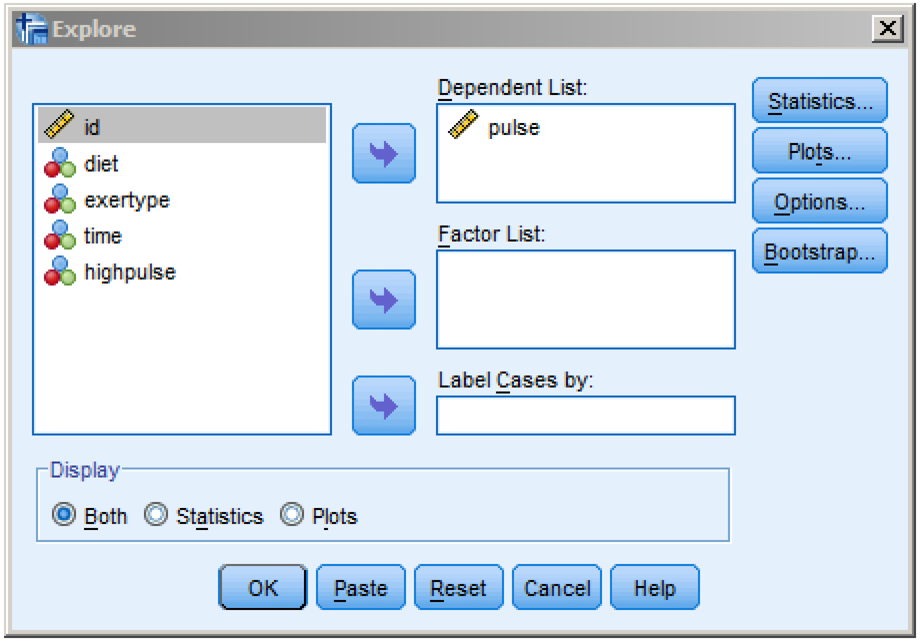
Normal probability plot



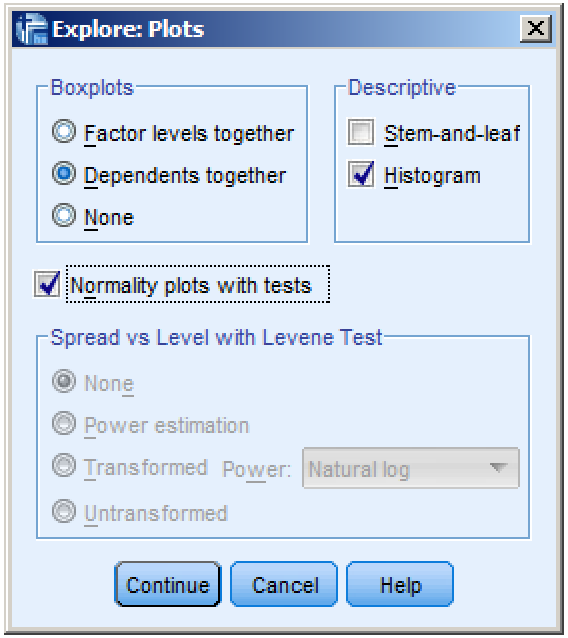
Analyze
› Descriptive Statistics
› Explore

Under plots select histogram, also normality plots with tests, descriptive statistics and boxplots are default options

Normal probability plot



Normal probability plot



Normal probability plot

Descriptives

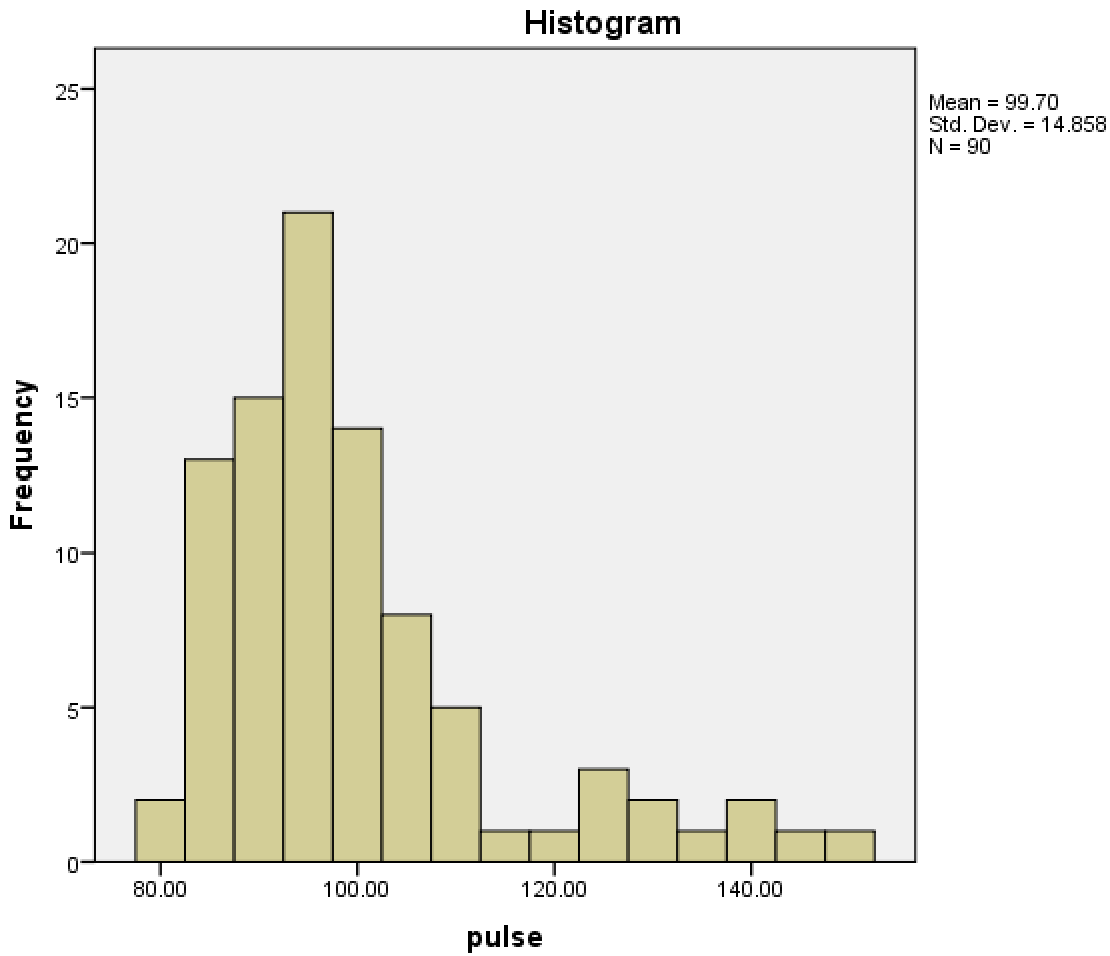
			Statistic	Std. Error
pulse	Mean		99.7000	1.56622
	95% Confidence Interval for Mean	Lower Bound	96.5880	
		Upper Bound	102.8120	
	5% Trimmed Mean		98.3086	
	Median		96.0000	
	Variance		220.774	
	Std. Deviation		14.85847	
	Minimum		80.00	
	Maximum		150.00	
	Range		70.00	
	Interquartile Range		13.00	
	Skewness		1.550	.254
	Kurtosis		2.162	.503

Tests of Normality

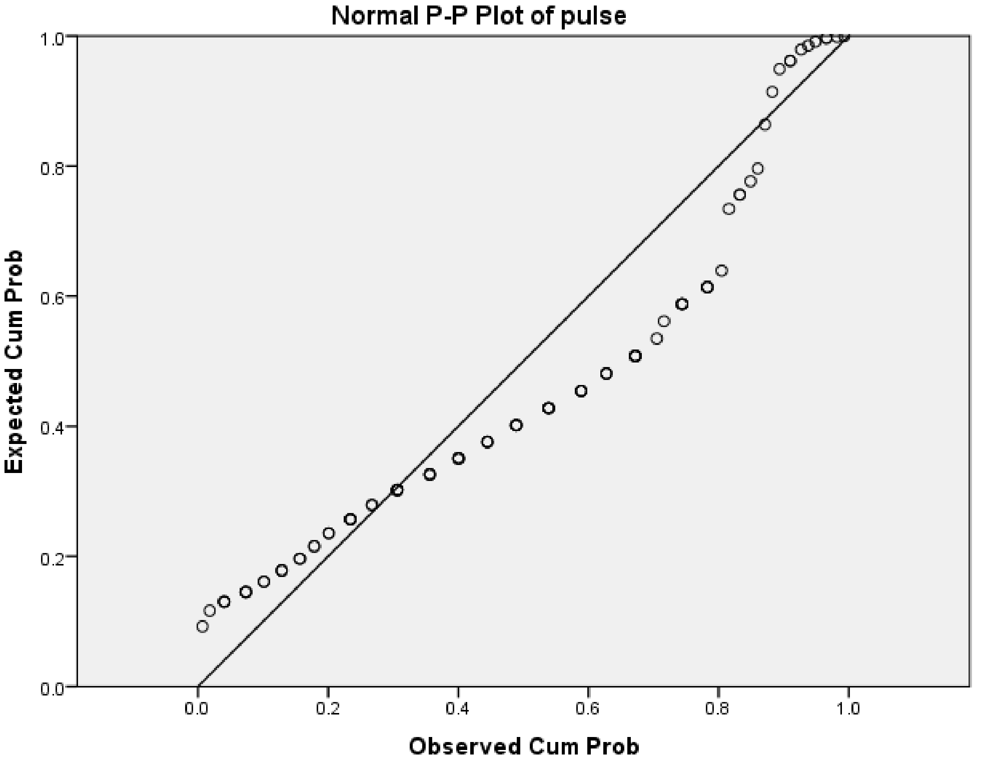
	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
pulse	.192	90	.000	.843	90	.000

a. Lilliefors Significance Correction

Normal probability plot

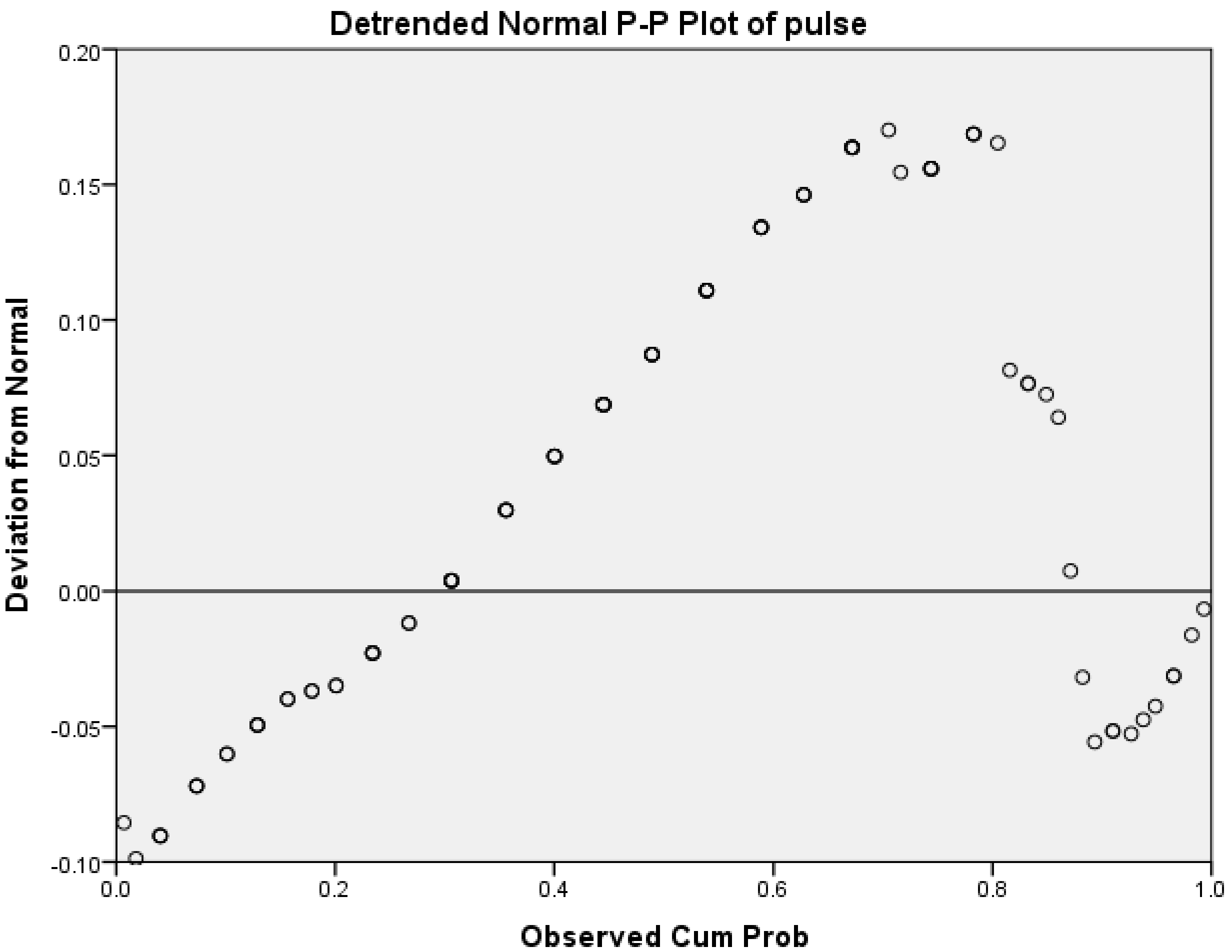


Normal probability plot



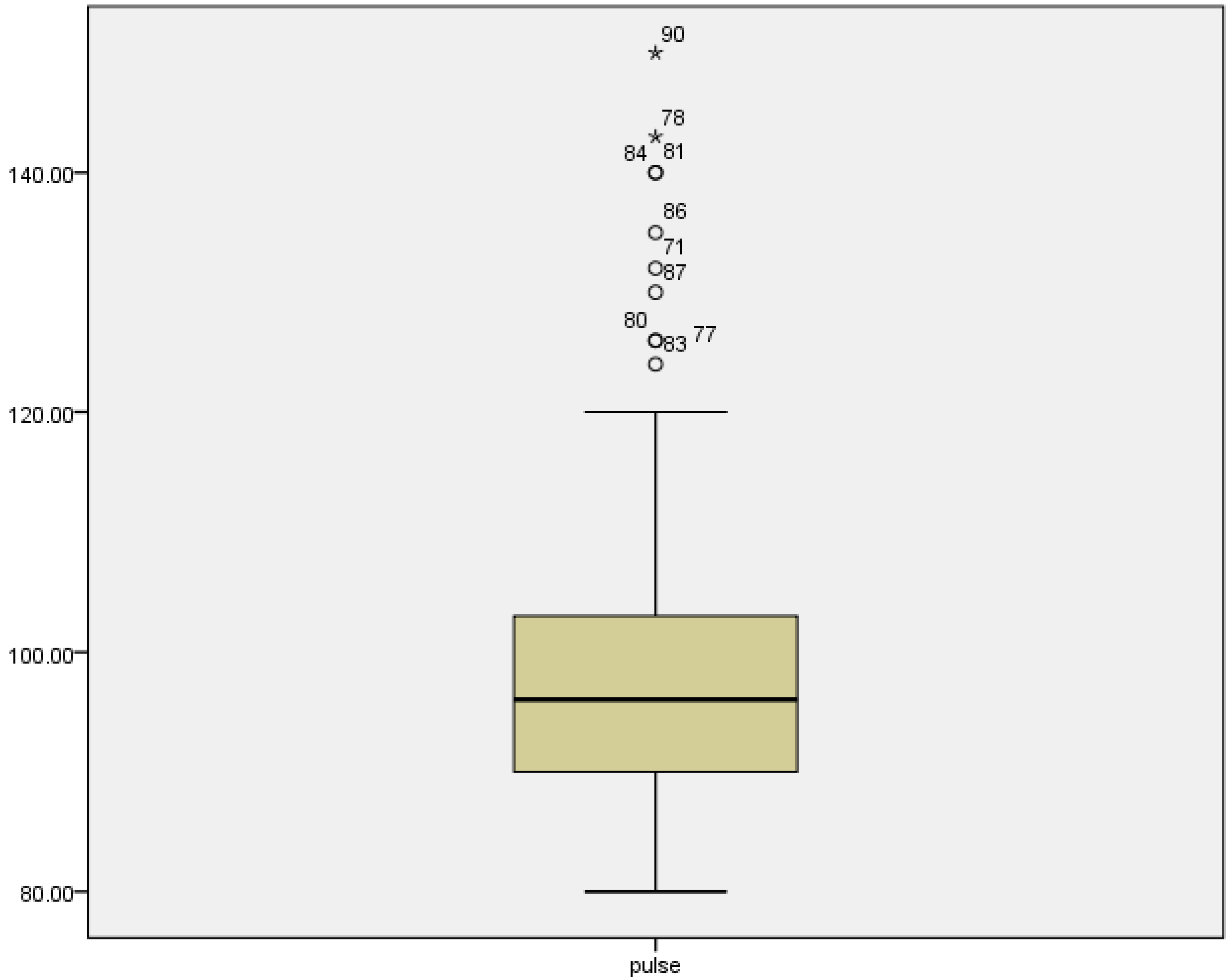
If the data is "normal" the non-linear vertical axis in a probability plot should result in an approximately linear scatter plot representing the raw data.

Normal probability plot

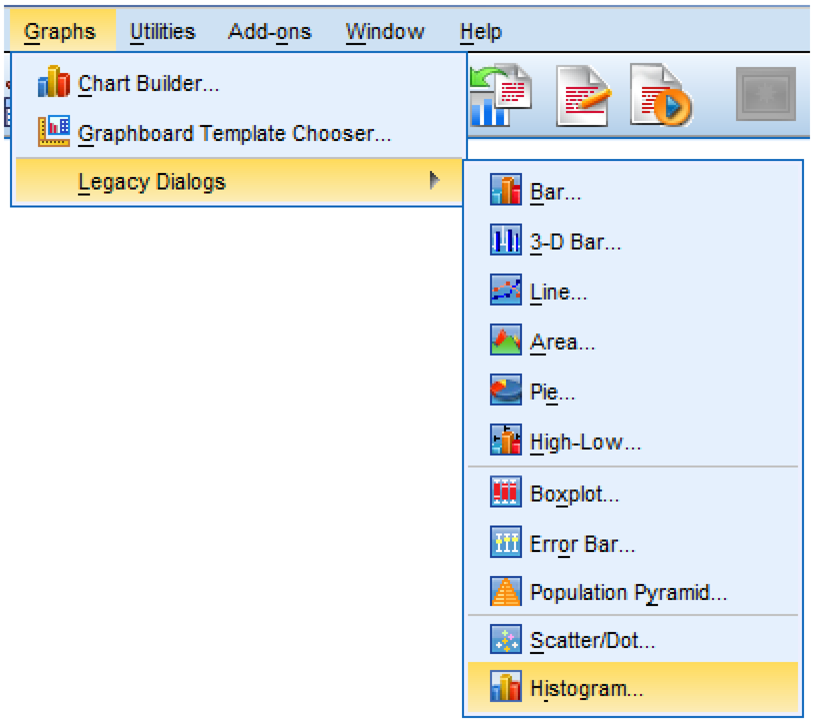


Detrended normal P-P plots depict the actual deviations of data points from the straight horizontal line at zero. No specific pattern in a detrended plot indicates normality of the variable.

Normal probability plot



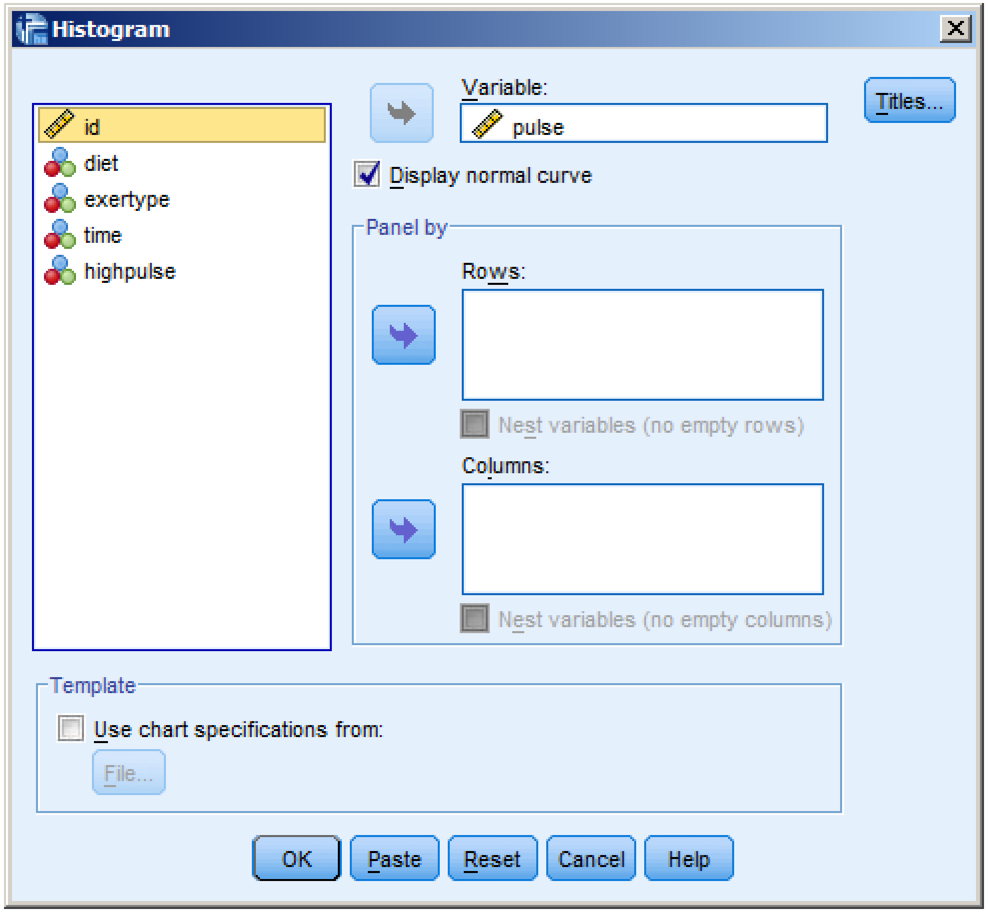
Normal probability plot



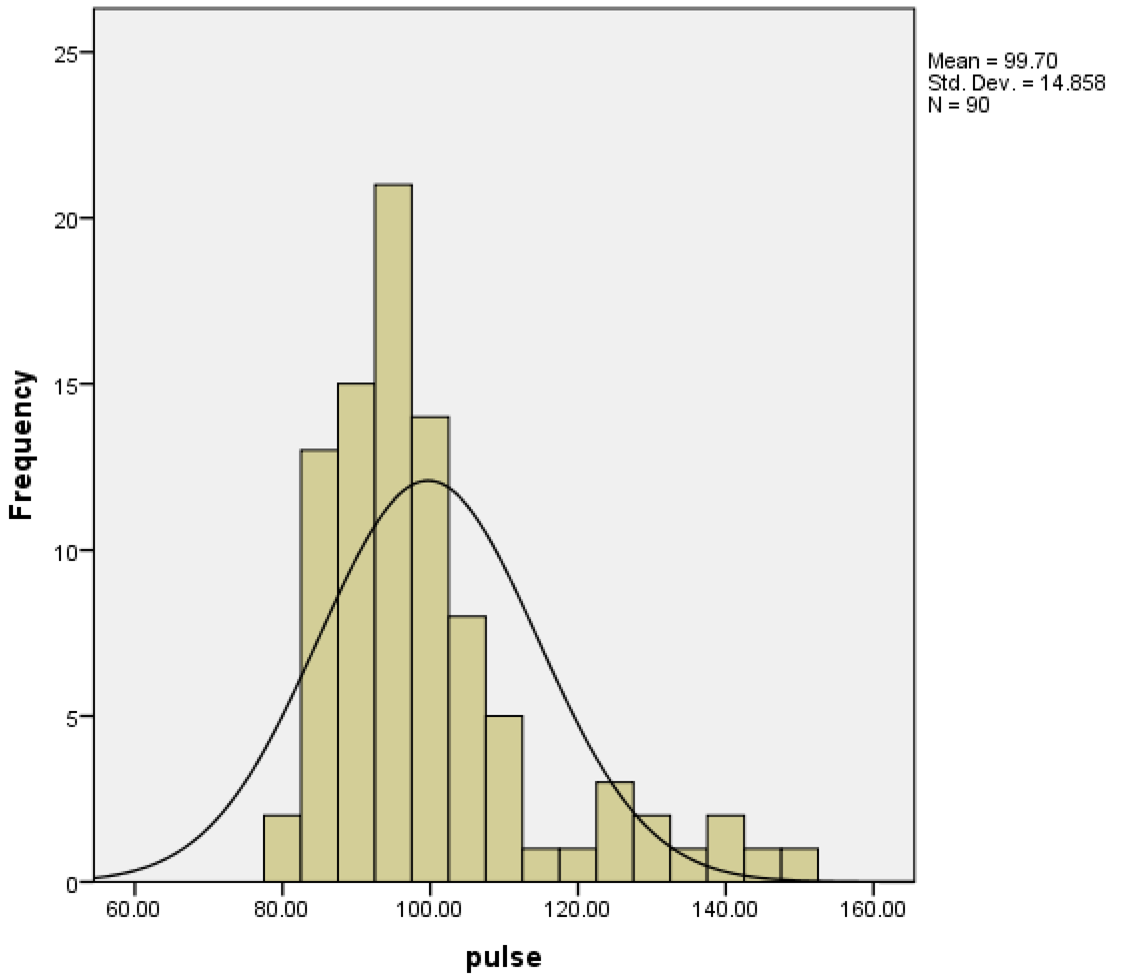
Graphs
> Legacy Dialogs
> Histogram

Tick – display normal curve

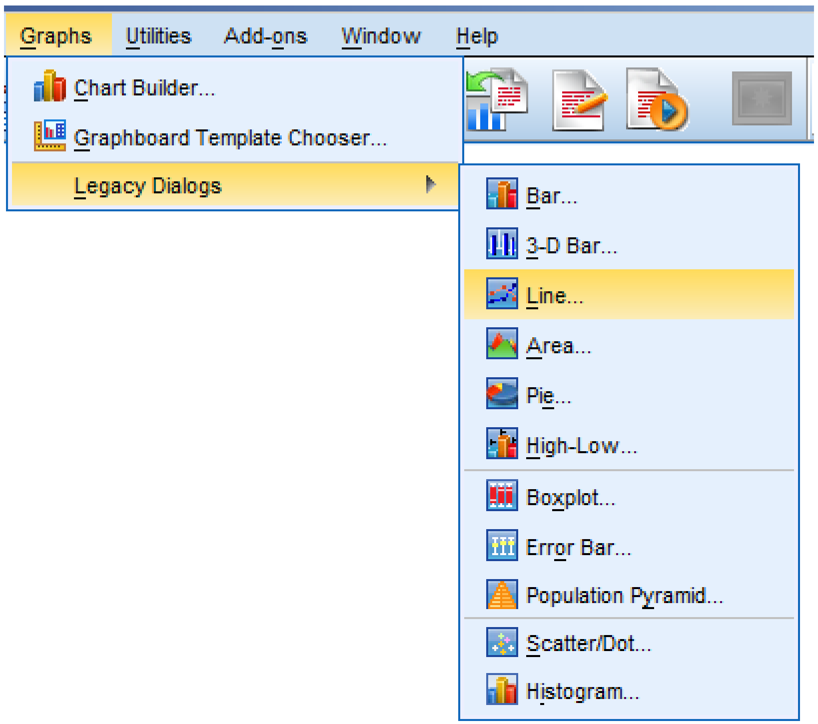
Normal probability plot



Normal probability plot



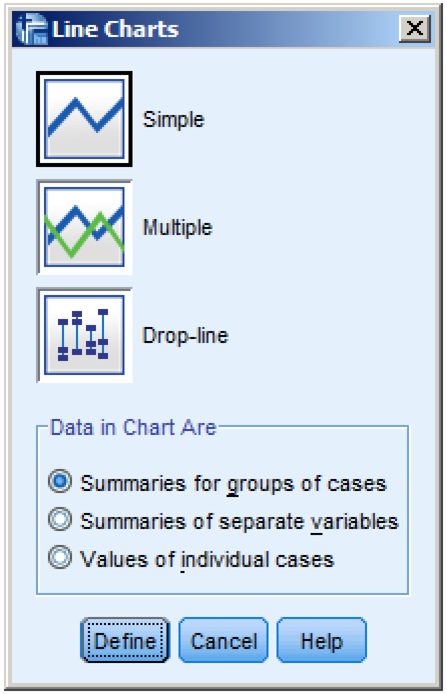
Normal probability plot



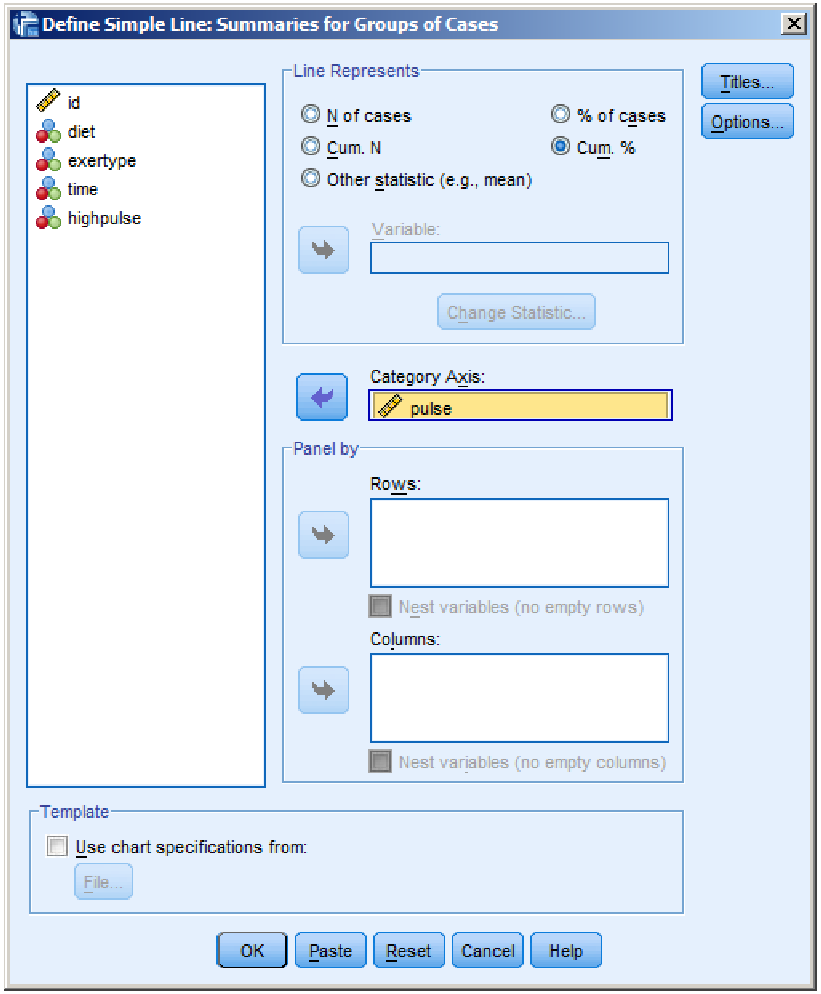
Graphs
> Legacy Dialogs
> Line

Select Simple and Groups of Cases the use Define to choose the variable and select "cum %"

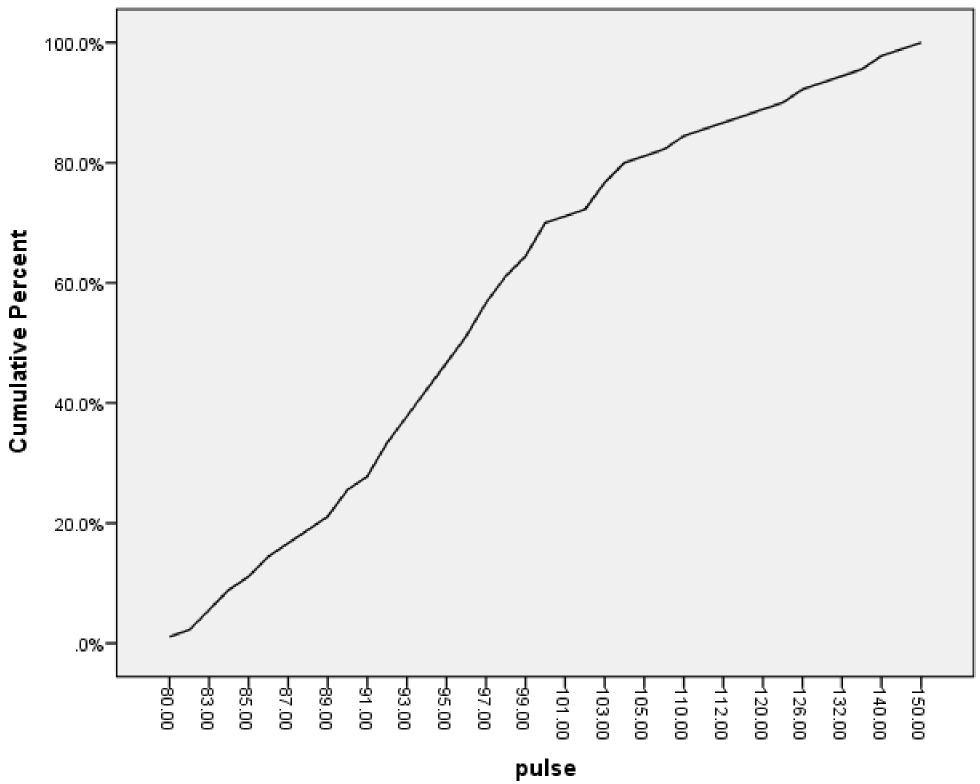
Normal probability plot



Normal probability plot



Normal probability plot



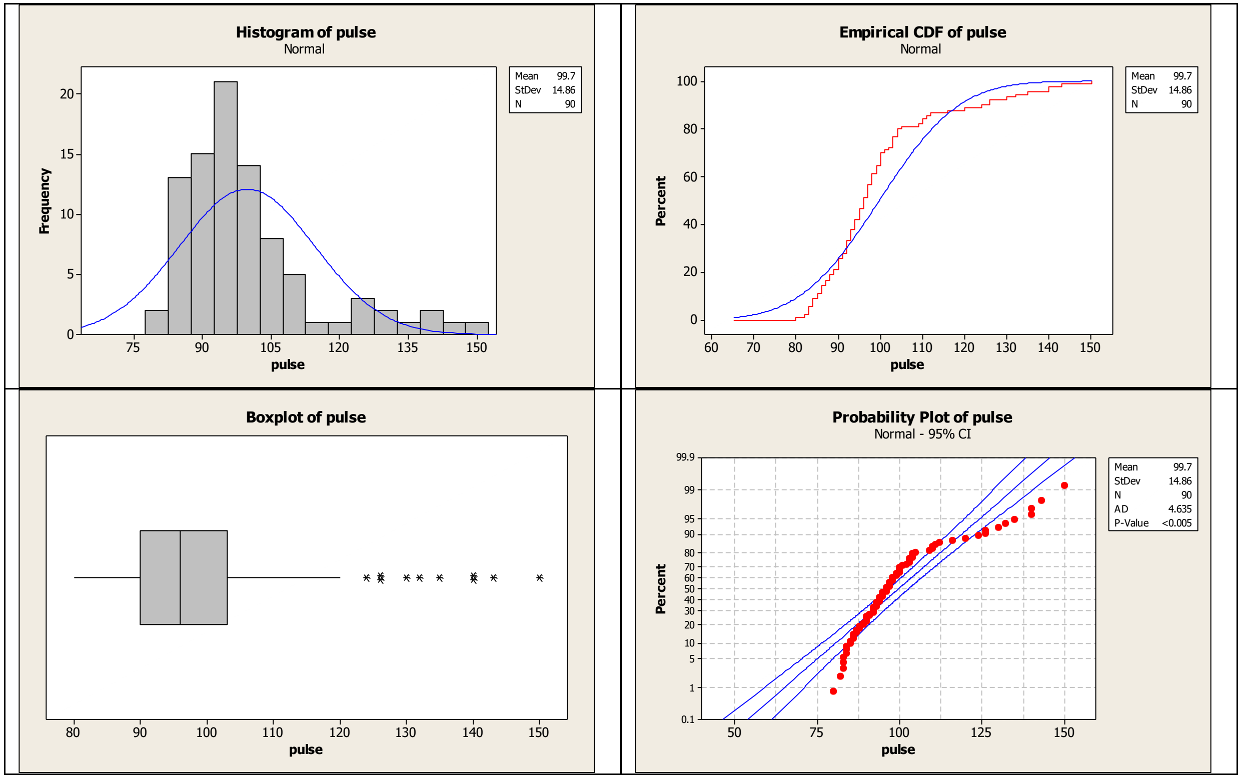
If you wish to superimpose a normal curve, it is probably simpler in Excel!

Normal probability plot

You are seeking to assess the normality of the data. The pulse data from data set C is employed.

The P-P plot is a normal probability plot with the data on the horizontal axis and the expected z-scores if our data was normal on the vertical axis. When our data is approximately normal the spacing of the two will agree resulting in a plot with observations lying on the reference line in the normal probability plot.

Normal probability plot



Does It Really Matter?

“Students t test and more generally the ANOVA F test are robust to non-normality” (Fayers 2011).

However

“Thus a clearer statement is that t tests and ANOVA are ‘robust against type-I errors’. This of course accords with the enthusiasm that many researchers have in obtaining “significant” p values.
The aim of this article (see next slide) is to show that type-II errors can be substantially increased if non-normality is ignored.” (Fayers 2011).

Does It Really Matter?

Alphas, betas and skewy distributions: two ways of getting the wrong answer Paper
Peter Fayers
Adv. Health Sci. Educ. Theory Pract. 2011 **16(3)** 291–296.

Introduction to Robust Estimation and Hypothesis Testing (2nd ed.).
Wilcox, R. R., 2005, Burlington MA:
Elsevier Academic Press. ISBN 978-0-12-751542-7.

Robustness to Non-Normality of Common Tests for the Many-Sample Location Problem Paper
Khan A. and Rayner G.D.
Journal Of Applied Mathematics And Decision Sciences, 2003, **7(4)**, 187:206

Index End

Tukey's ladder of powers

3	y^3
2	y^2
1	y^1
½	$\sqrt{y}$
0	$ln(y)$
-1	y^{-1}
-2	y^{-2}
-3	y^{-3}

Cartoon

Tukey has designed a family of power transformations (close cousin to the Box-Cox transformations, but with a visual aspect useful to find the appropriate transformation to promote symmetry and linearity relationships.

These transformations preserve order, preserve proximities and are smooth functions (not producing jumps or peaks). y^1 is the untransformed (raw) variable, y^0 is replaced by the logarithm that provides the appropriate transformation between the square root and the reciprocal.

You can also use lower and higher powers as listed, as well intermediate ones, i.e. $y^{2.5}$ will be stronger than y^2 but less than y^3.

Tukey, J. W. (1977) Exploratory Data Analysis. Addison-Wesley, Reading, MA.

Tukey's ladder of powers

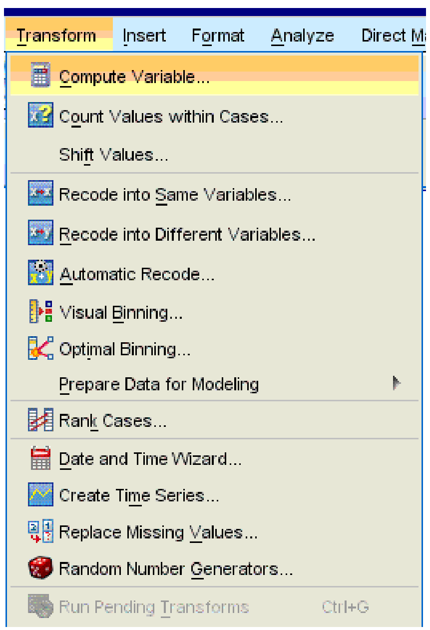
A transformation is simply a means of representing the data in a different coordinate system. In addition to restoring normality, the transformation often reduces heteroscedasticity. (Non-constancy of the variance of a measure over the levels of the factor under study.) This is important, because constant variance is often an assumption of parametric tests. Subsequent statistical analyses are performed on the transformed data; the results are interpreted with respect to the original scale of measurement.

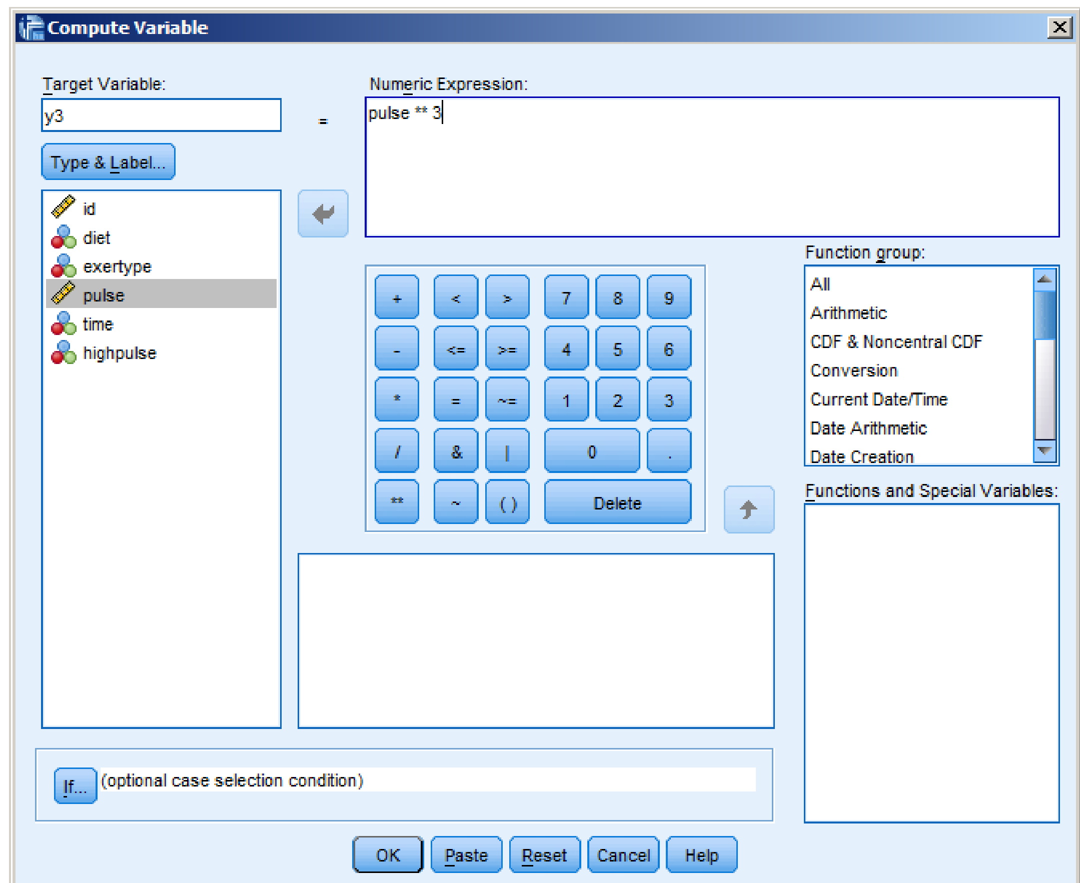
Achieving an appropriate transformation is a trial-and-error process. A particular transformation is applied and the new data distribution tested for normality; if the data are still non-normal, the process is repeated.

Nevertheless, there are certain generalities that can be used to direct your efforts, as certain types of data typically respond to particular transformations. For example Square-root transforms are often appropriate for count data, which tend to follow Poisson distributions. Arcsine ($\sin^{-1}$) transforms are used for data that are percentages or proportions, and tend to fit binomial distributions. Log and square-root transforms are part of a larger class of transforms known as the ladder of powers.

Tukey's ladder of powers

Transform > Compute Variable





See normal probability plot section for graphical options.

Tukey's ladder of powers

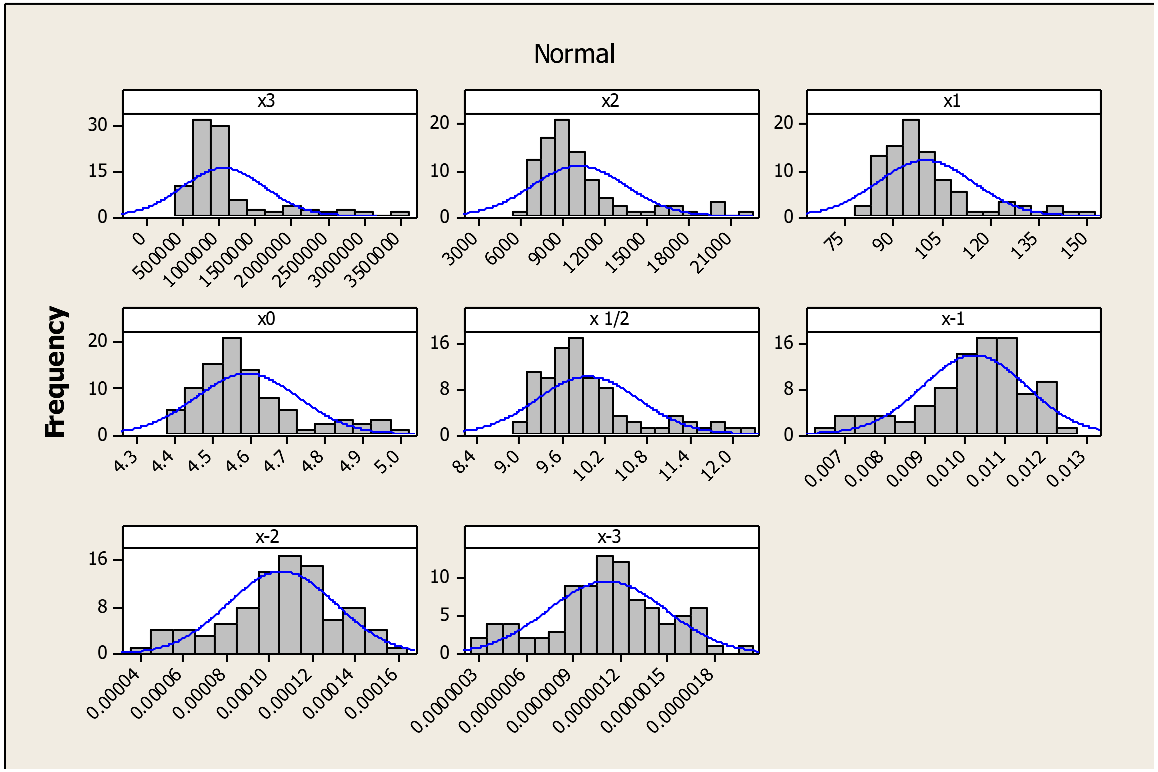
```
COMPUTE pulse1=1 + pulse/MAX(pulse).
COMPUTE y3=pulse1 ** 3.
EXECUTE.
COMPUTE y2=pulse1 ** 2.
EXECUTE.
COMPUTE y=pulse1 .
EXECUTE.
COMPUTE rt_y=SQRT(pulse1).
EXECUTE.
COMPUTE ln_y=LN(pulse1).
EXECUTE.
COMPUTE y_1=pulse1 ** -1.
EXECUTE.
COMPUTE y_2=pulse1 ** -2.
EXECUTE.
COMPUTE y_3=pulse1 ** -3.
EXECUTE.
EXAMINE VARIABLES=y3 y2 y rt_y ln_y y_1 y_2 y_3
  /COMPARE VARIABLE
  /PLOT=BOXPLOT
  /STATISTICS=NONE
  /NOTOTAL
  /MISSING=LISTWISE.
```

The pulse data from data set C is employed.

It is scaled by the first compute statement to aid interpretation of the plots.

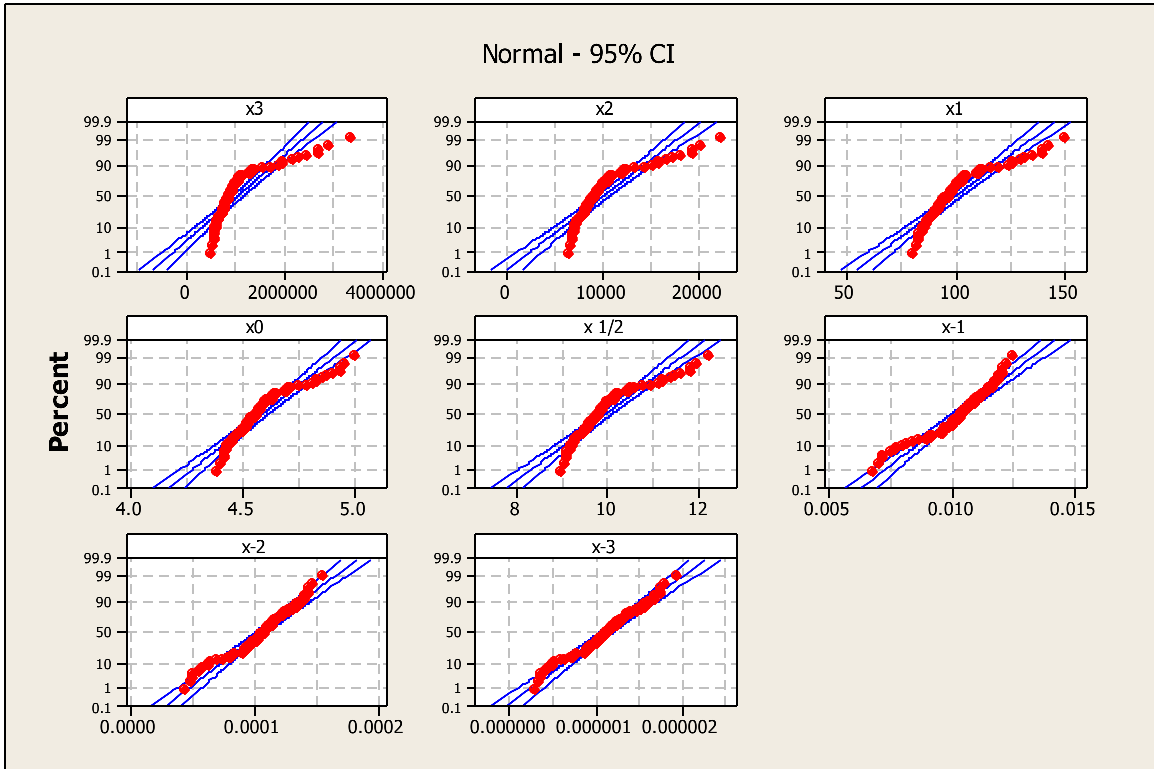
This step is non-essential, only aiding the graphical presentation.

Tukey's ladder of powers



Which appears most "normal"?

Tukey's ladder of powers



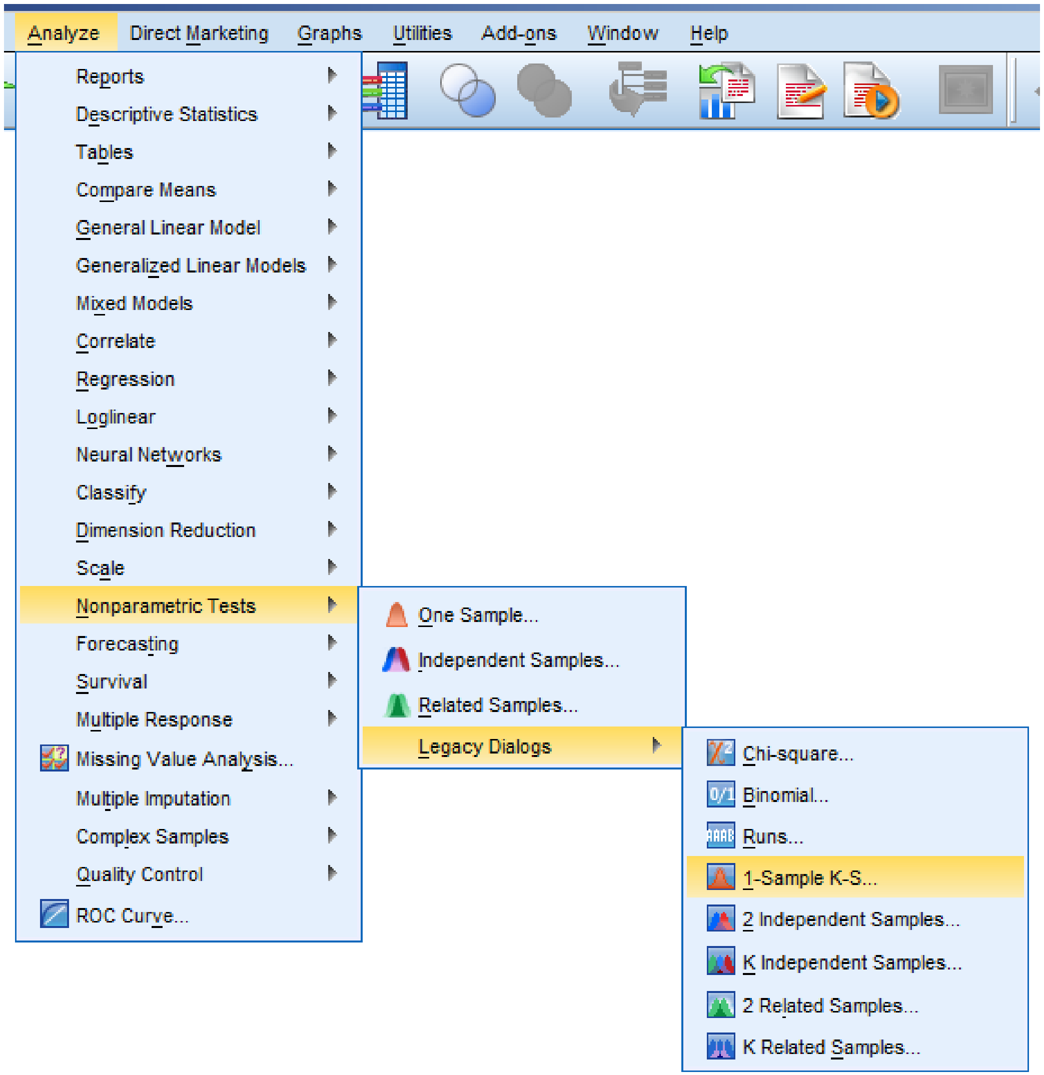
Which appears most "normal"?

Tukey's ladder of powers

	x3	*x2*	*x1*	*x0*	*x*	*x-1*	*x-2*	*x-3*
Mean	1061245	10158	99.7	4.592	9.96	0.01022	0.000106	1.12E-06
StDev	567251	3310	14.86	0.137	0.711	0.001296	2.52E-05	3.75E-07
N	90	90	90	90	90	90	90	90
AD	8.615	6.523	4.635	3.048	3.799	1.824	0.988	0.529
P-Value	<0.005	<0.005	<0.005	<0.005	<0.005	<0.005	0.013	0.172

In general if the normal distribution fits the data, then the plotted points will roughly form a straight line. In addition the plotted points will fall close to the fitted line. Also the Anderson-Darling statistic will be small, and the associated p-value will be larger than the chosen α-level (usually 0.05). So the test rejects the hypothesis of normality when the p-value is less than or equal to α.

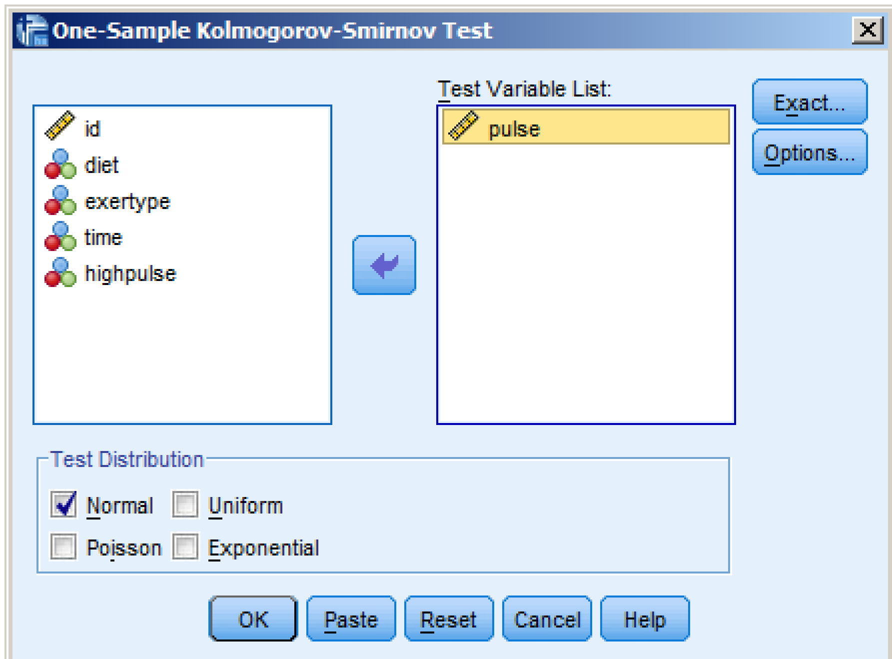
Tukey's ladder of powers



To test for normality is SPSS you can perform a Kolmogorov-Smirnov Test,

Analyze
> Nonparametric tests
> Legacy Dialogs
> 1-Sample Kolmogorov-Smirnov Test

Tukey's ladder of powers



Tukey's ladder of powers

One-Sample Kolmogorov-Smirnov Test

		pulse
N		90
Normal Parameters[a,b]	Mean	99.7000
	Std. Deviation	14.85847
Most Extreme Differences	Absolute	.192
	Positive	.192
	Negative	-.108
Kolmogorov-Smirnov Z		1.821
Asymp. Sig. (2-tailed)		.003

a. Test distribution is Normal.
b. Calculated from data.

Tukey's ladder of powers

One-Sample Kolmogorov-Smirnov Test

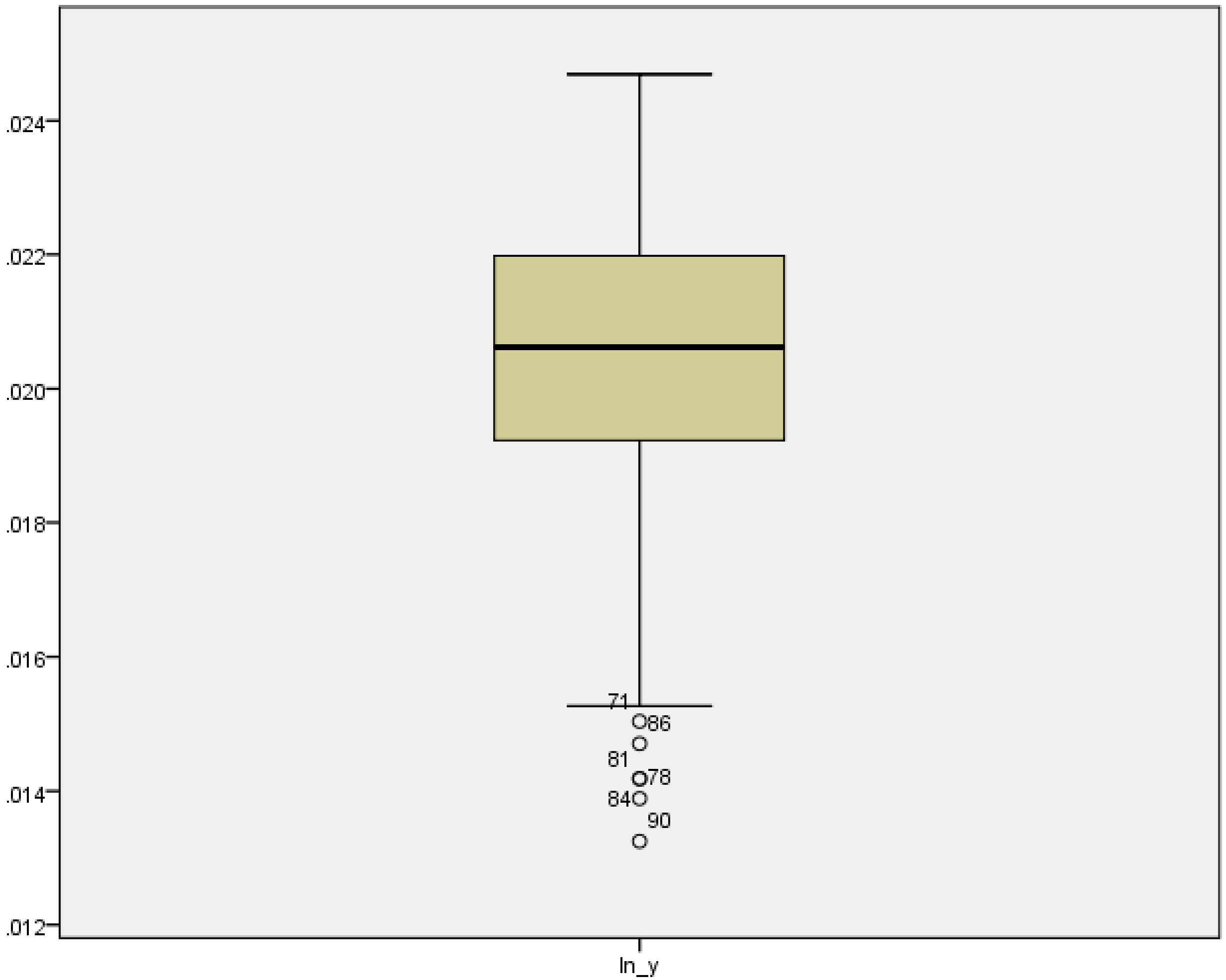
		y3	y2	y	rt_y	ln_y	y-1	y-2	y-3
N		90	90	90	90	90	90	90	90
Normal Parameters[a,b]	Mean	1061244.5667	10158.4111	99.7000	9.9599	4.5924	.0102	.0001	.0000
	Std. Deviation	567251.11996	3309.53301	14.85847	.71097	.13698	.00130	.00003	.00000
Most Extreme Differences	Absolute	.255	.221	.192	.178	.163	.133	.105	.079
	Positive	.255	.221	.192	.178	.163	.063	.060	.054
	Negative	-.173	-.139	-.108	-.094	-.080	-.133	-.105	-.079
Kolmogorov-Smirnov Z		2.422	2.099	1.821	1.684	1.544	1.263	.993	.745
Asymp. Sig. (2-tailed)		.000	.000	.003	.007	.017	.082	.278	.635

a. Test distribution is Normal.

b. Calculated from data.

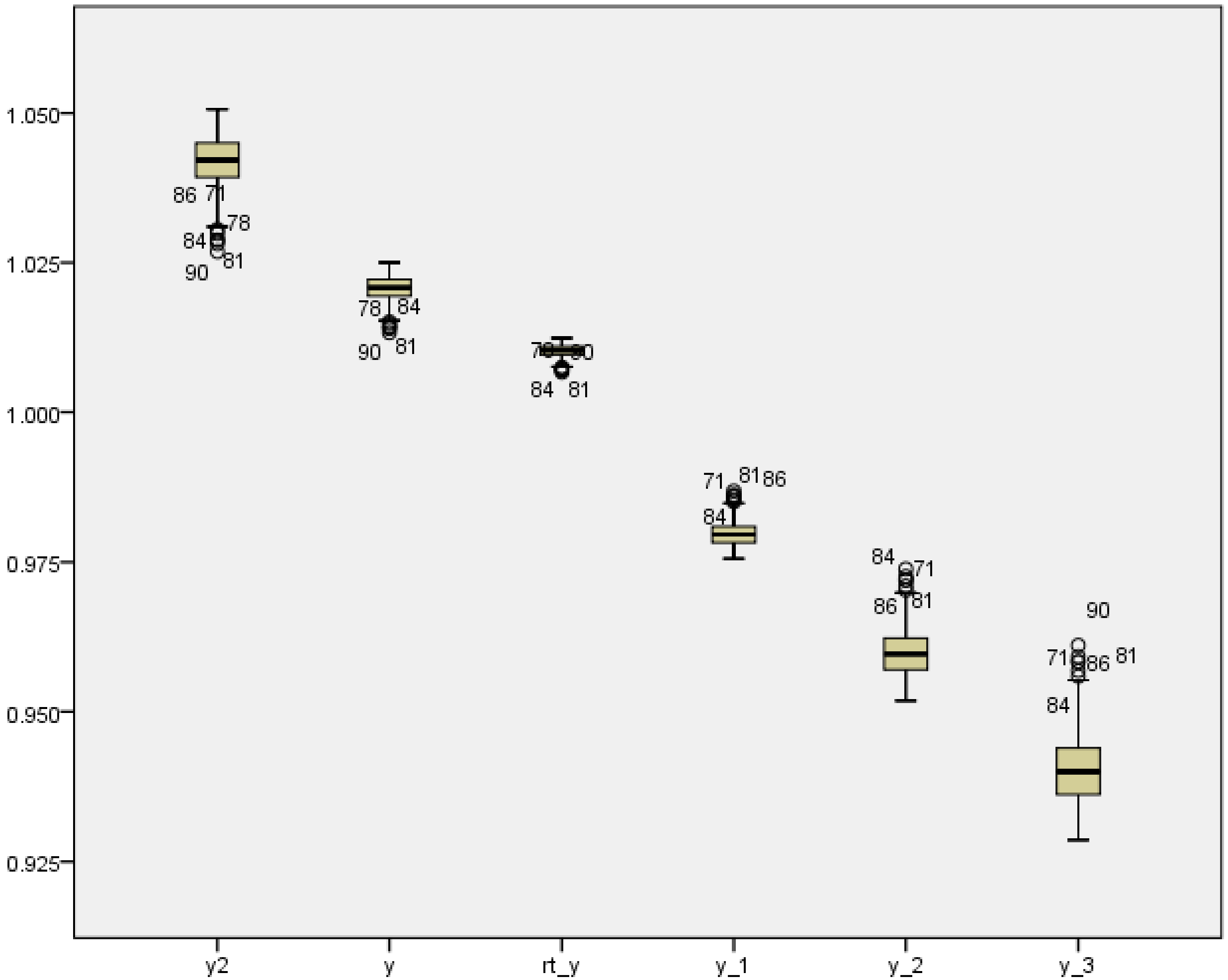
The Asymp. Sig. (2 tailed) value is also known as the p-value. This tells you the probability of getting the results if the null hypothesis were actually true (i.e. it is the probability you would be in error if you rejected the null hypothesis).

Tukey's ladder of powers



Despite the scaling the log. transform spoils the final plot.

Tukey's ladder of powers



You are seeking the most normal data visually.

Probably the square root transform.

Tukey's ladder of powers

Many statistical methods require that the numeric variables you are working with have an approximately normal distribution. Reality is that this is often times not the case. One of the most common departures from normality is skewness, in particular, right skewness.

When the data is plotted vs. the expected z-scores the normal probability plot shows right skewness by a downward bending curve.

When the data is plotted vs. the expected z-scores the normal probability plot shows left skewness by an upward bending curve.

Tukey's Ladder of Powers

Tukey (1977) describes an orderly way of re-expressing variables using a power transformation. If a transformation for x of the type x^λ, results in an effectively linear probability plot, then we should consider changing our measurement scale for the rest of the statistical analysis. There is no constraint on values of λ that we may consider. Obviously choosing $\lambda = 1$ leaves the data unchanged. Negative values of λ are also reasonable. Tukey (1977) suggests that it is convenient to simply define the transformation when $\lambda = 0$ to be the logarithmic function rather than the constant 1.

Tukey's Ladder of Powers

In general if the normal distribution fits the data, then the plotted points will roughly form a straight line. In addition the plotted points will fall close to the fitted line. Also the Anderson-Darling statistic will be small, and the associated p-value will be larger than the chosen α-level (usually 0.05). So the test rejects the hypothesis of normality when the p-value is less than or equal to α.

Tukey's Ladder of Powers

To test for normality is SPSS you can perform a Kolmogorov-Smirnov Test

Analyze > Nonparametric tests
> 1-Sample Kolmogorov-Smirnov Test

The Asymp. Sig. (2 tailed) value is also known as the p-value. This tells you the probability of getting the results if the null hypothesis were actually true (i.e. it is the probability you would be in error if you rejected the null hypothesis).

Tukey's Ladder of Powers

The hypothesis are

H_0 the distribution of x is normal
H_1 the distribution of x is not normal

If the p-value is less than .05, you reject the normality assumption, and if the p-value is greater than .05, there is insufficient evidence to suggest the distribution is not normal (meaning that you can proceed with the assumption of normality.)

In summary if the test is significant (lower than or equal to 0.05) implies the data is not normally distributed.

Median Split

There is quite a literature to suggest that, even though it is nice and convenient to sort people into 2 groups and then use a t test to compare group means, you lose considerable power. Cohen (1983) has said that breaking subjects into two groups leads to the loss of 1/5 to 2/3 of the variance accounted for by the original variables. The loss in power is equivalent to tossing out 1/3 to 2/3 of the sample.

The Cost Of Dichotomization or
Cohen, J.
Applied Psychological Measurement Volume: 7 Issue: 3
Pages: 249-253 Published: 1983

Likert Scale

These are not exhaustive notes, rather some thoughts on preparing a Likert Scale.

Statements are often rated on a five point Likert (Likert 1932) scale. There has been much research done to demonstrate that a five-point scale can lead to extremes, and therefore a seven-point scale is preferable (Payton et al. 2003).

Likert, R. (1932). A technique for the measurement of attitudes. Archives of Psychology, Vol. 22, No. 140, 55pp.

Payton, M.E., Greenstone, M.H. and Schenker, N. (2003). Overlapping confidence intervals or standard error intervals: What do they mean in terms of statistical significance?. Journal of Insect Science, 3(34), 6pp.

Likert Scale

A 7-point Likert scale is recommended to maximise the sensitivity of the scale (Allen and Seaman 2007, Cummins and Gullone 2000).

Allen, I.E., and Seaman, C.A. (2007). Likert scales and data analyses. Quality Progress, 40(7), 64-65.

Cummins, R.A., and Gullone, E. (2000). Why we should not use 5-point Likert scales: The case for subjective quality of life measurement. In Proceedings, second international conference on quality of life in cities (p74-93). Singapore: National University of Singapore.

Likert Scale

It is unclear why different scales may have been used in the same experiment, but it has been shown that data is still comparable when it has been re-scaled (Dawes 2008).

Dawes J. (2008) Do data characteristics change according to the number of scale points used? An experiment using 5 point, 7 point and 10 point scales. International Journal of Market Research. 51 (1) 61-77.

Likert Scale

It may be better if the scale contained a neutral midpoint (Tsang 2012). This decision (an odd/even scale) depends whether respondents are being forced to exclude the neutral position with an even scale.

Tsang K.K 2012 "The use of midpoint on Likert Scale: The implications for educational research" Vol. 11 121-130.

Likert Scale

An odd number of points allow people to select a middle option. An even number forces respondents to take sides. An even number is appropriate when you want to know what direction the people in the middle are leaning. However, forcing people to choose a side, without a middle point, may frustrate some respondents (Wong et al. 1993).

Wong, C.-S., Tam, K.-C., Fung, M.-Y., and Wan, K. (1993). Differences between odd and even number of response scale: Some empirical evidence. Chinese Journal of Psychology, 35, 75-86.

Likert Scale

Since they have no neutral point, even-numbered Likert scales force the respondent to commit to a certain position (Brown, 2006) even if the respondent may not have a definite opinion.

There are some researchers who prefer scales with 7 items or with an even number of response items (Cohen, Manion, and Morrison, 2000).

Brown, J.D. (2000). What issues affect Likert-scale questionnaire formats? JALT Testing and Evaluation SIG, 4, 27-30. here

Cohen, L., Manion, L. and Morrison, K. (2000). Research methods in education (5th ed.). London: Routledge Falmer.

Likert Scale

The change of response order in a Likert-type scale altered participant responses and scale characteristics. Response order is the order in which options of a Likert-type scale are offered (Weng 2000).

How many scale divisions or categories should be used (1 to 10; 1 to 7; -3 to +3)?
Should there be an odd or even number of divisions? (Odd gives neutral centre value; even forces respondents to take a non-neutral position.)
What should the nature and descriptiveness of the scale labels be?
What should the physical form or layout of the scale be? (graphic, simple linear, vertical, horizontal)
Should a response be forced or be left optional?

Li-Jen Weng 2000 Effects of Response Order on Likert-Type Scales, Educational and Psychological Measurement December vol. 60 no. 6 908-924.

Winsorize

Winsorising or Winsorization is the transformation of statistics by limiting extreme values in the statistical data to reduce the effect of possibly spurious outliers. It is named after the engineer-turned-biostatistician Charles P. Winsor (1895-1951).

The computation of many statistics can be heavily influenced by extreme values. One approach to providing a more robust computation of the statistic is to Winsorize the data before computing the statistic.

Apart from confusion about the correct spelling. There is the ambiguity about where the precise percentile sits.

Winsorize

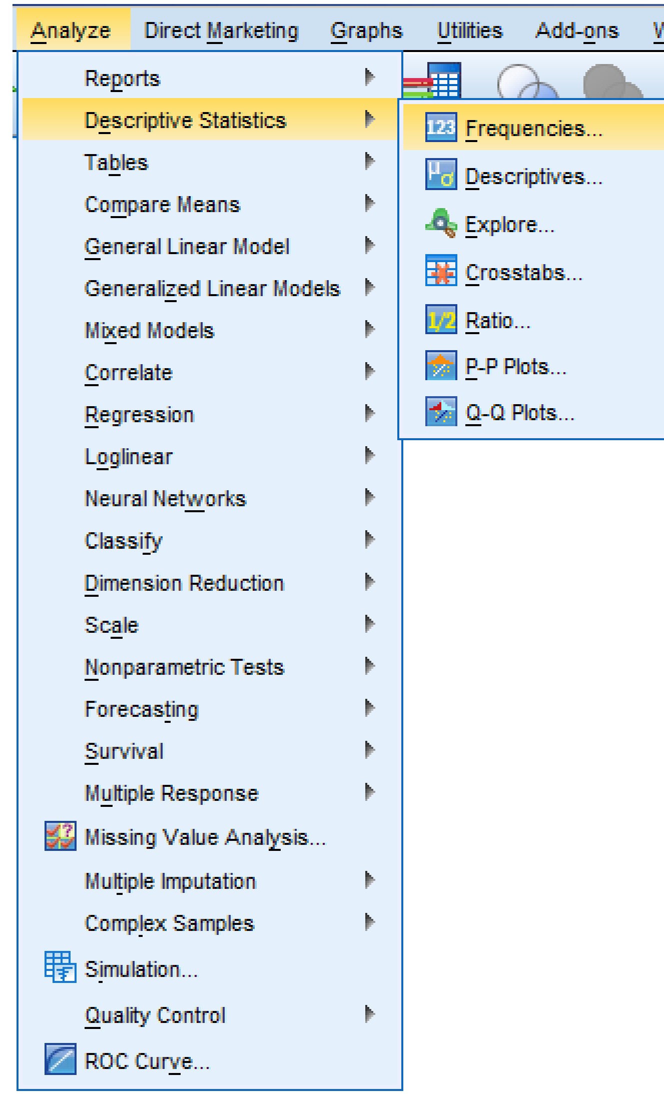
To Winsorize the data, tail values are set equal to some specified percentile of the data. For example, for a 90% Winsorization, the bottom 5% of the values are set equal to the value corresponding to the 5th percentile while the upper 5% of the values are set equal to the value corresponding to the 95th percentile.

Just because a method exists does not necessarily mean its a great idea!!

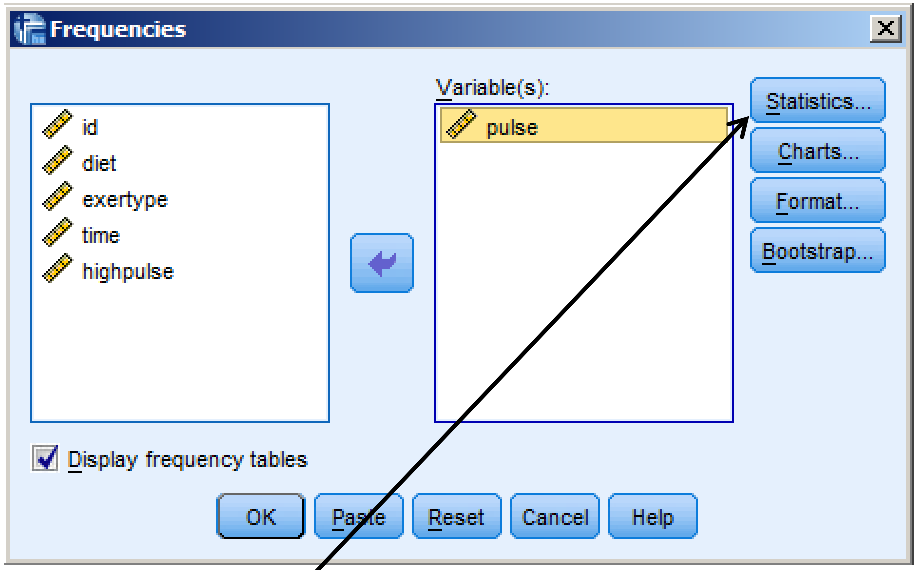
Winsorize

The pulse data from data set C is employed.

Analyze > Descriptive Statistics
> Frequencies

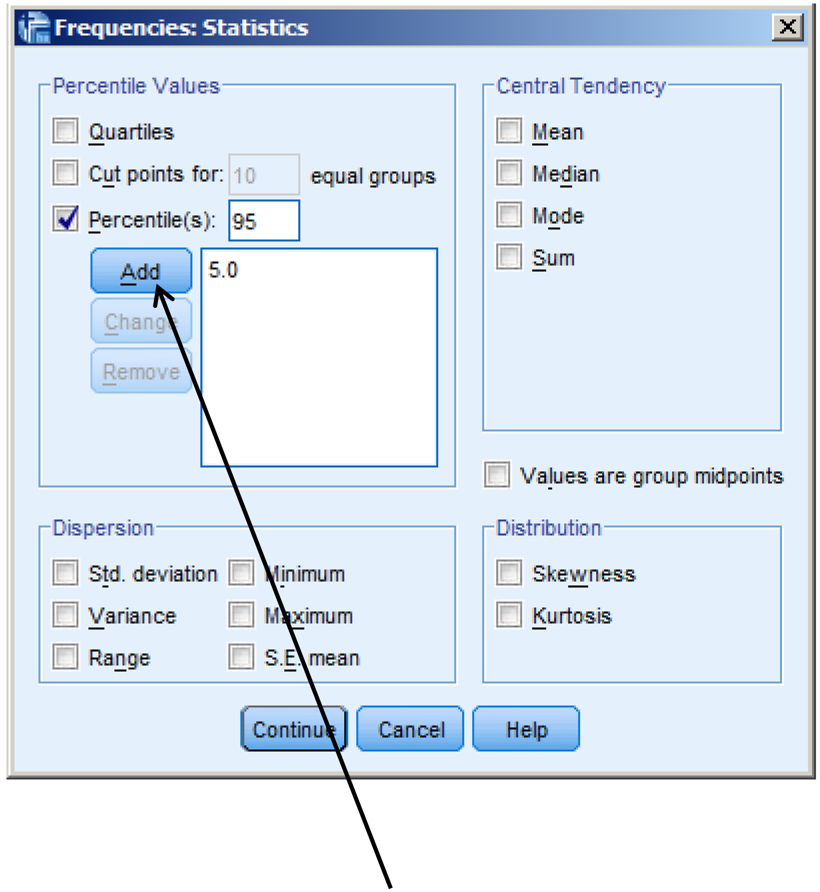


Winsorize



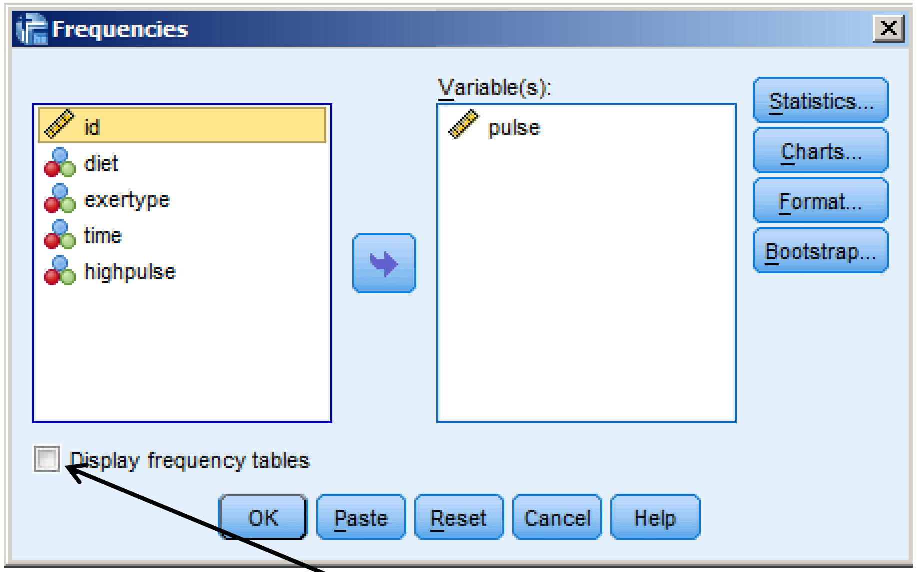
Select statistics

Winsorize



Add desired percentiles, 5 then 95

Winsorize



For brevity do not display frequency tables

Winsorize

Statistics

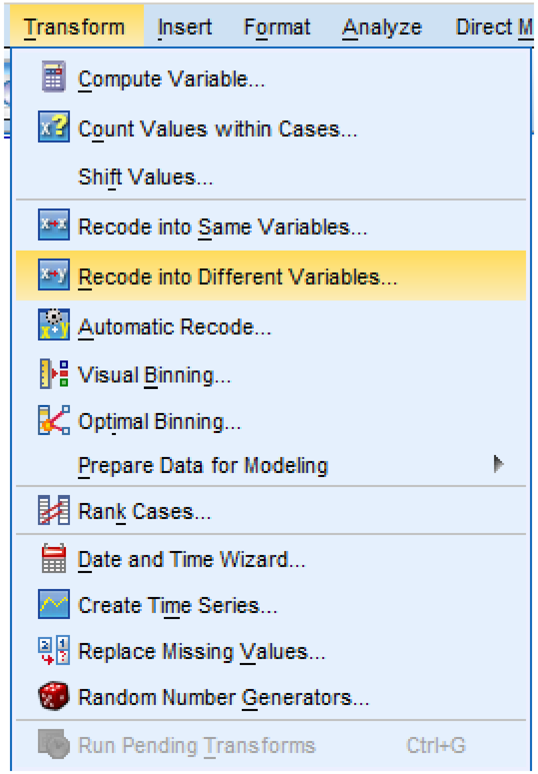
pulse

N	Valid	90
	Missing	0
Percentiles	5	83.0000
	95	137.2500

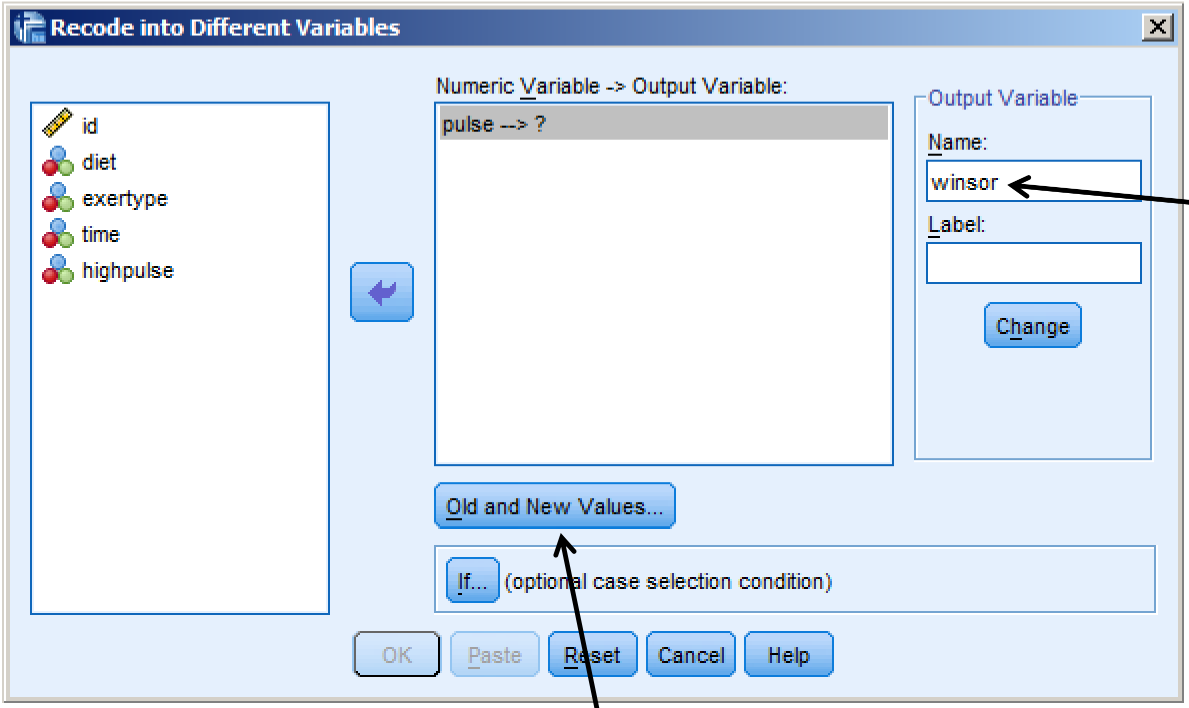
Note the percentiles and enter them into the next slide.

Winsorize

Transform > Compute Variable



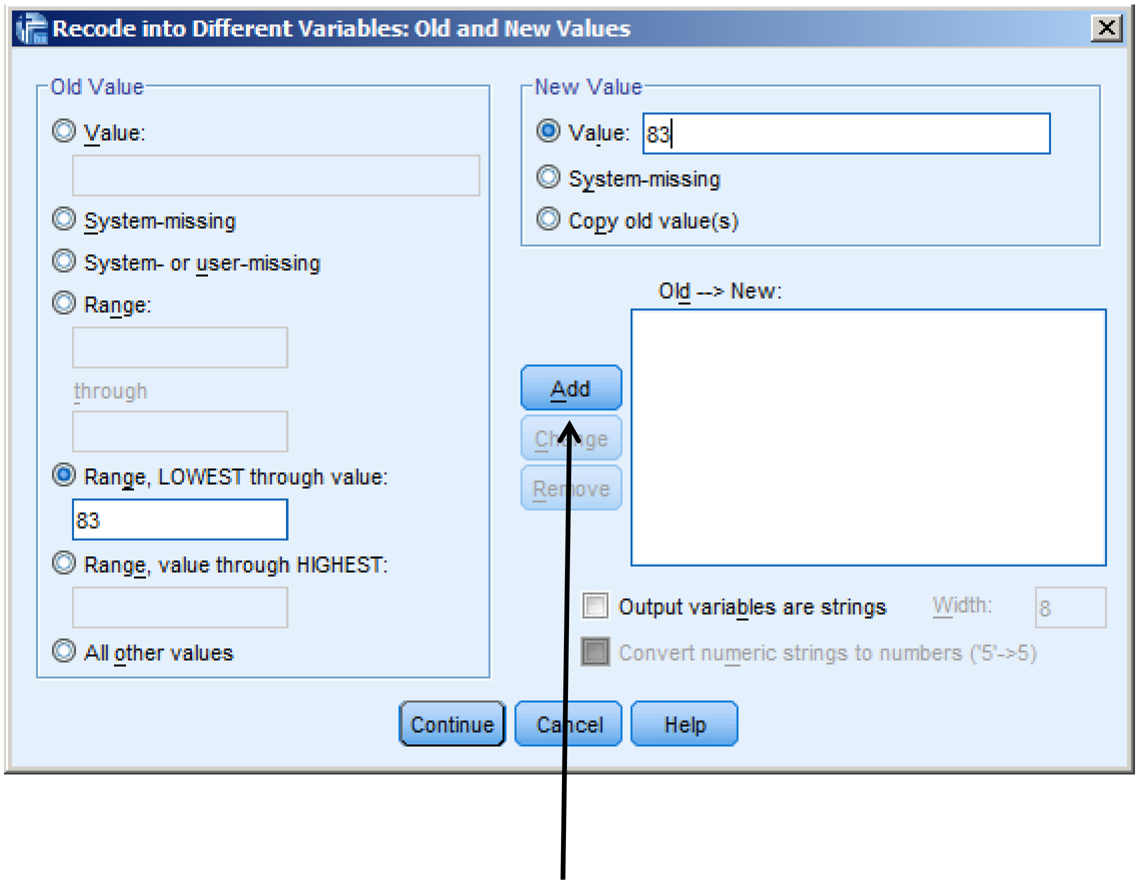
Winsorize



Choose a sensible new name

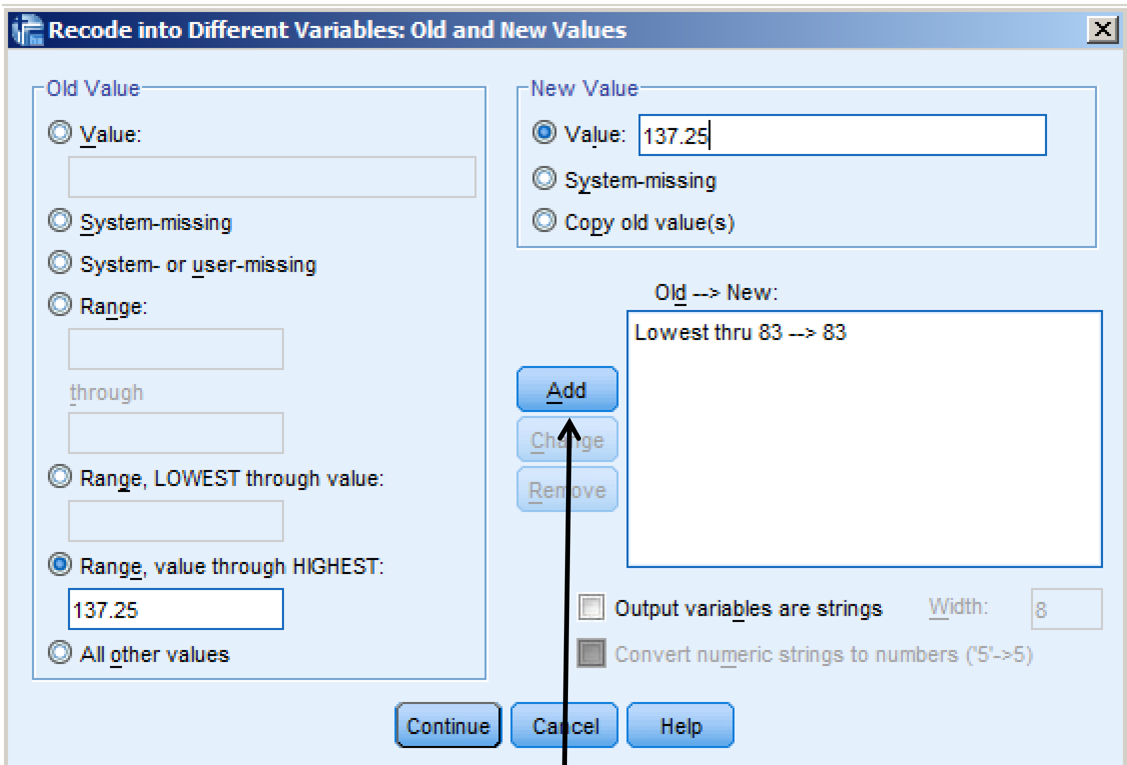
Select Old and New Values

Winsorize



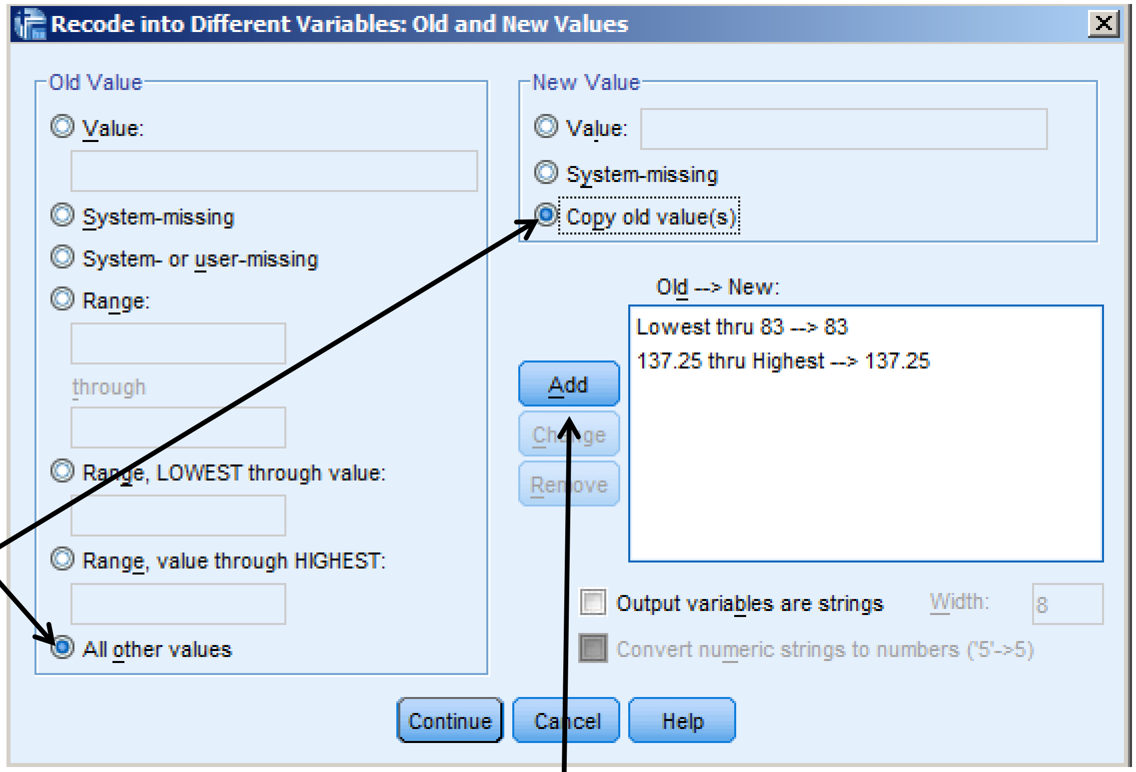
Then Add

Winsorize



Then Add

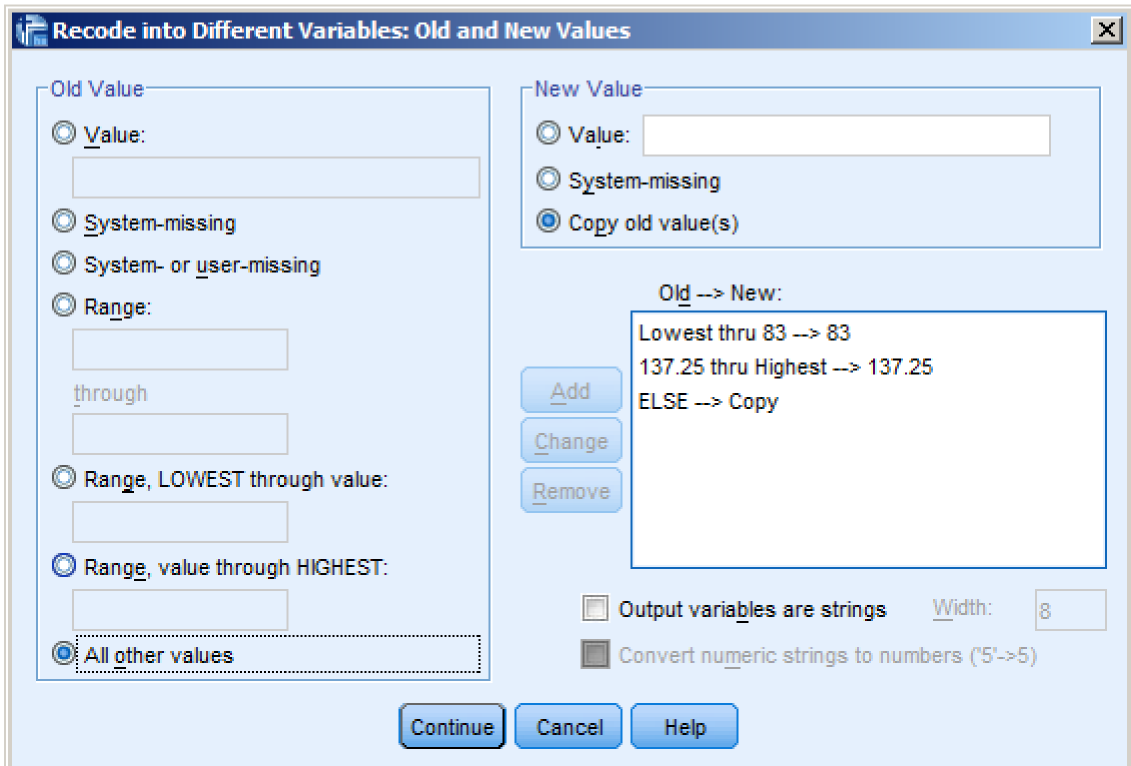
Winsorize



Retain all other values

Then Add

Winsorize

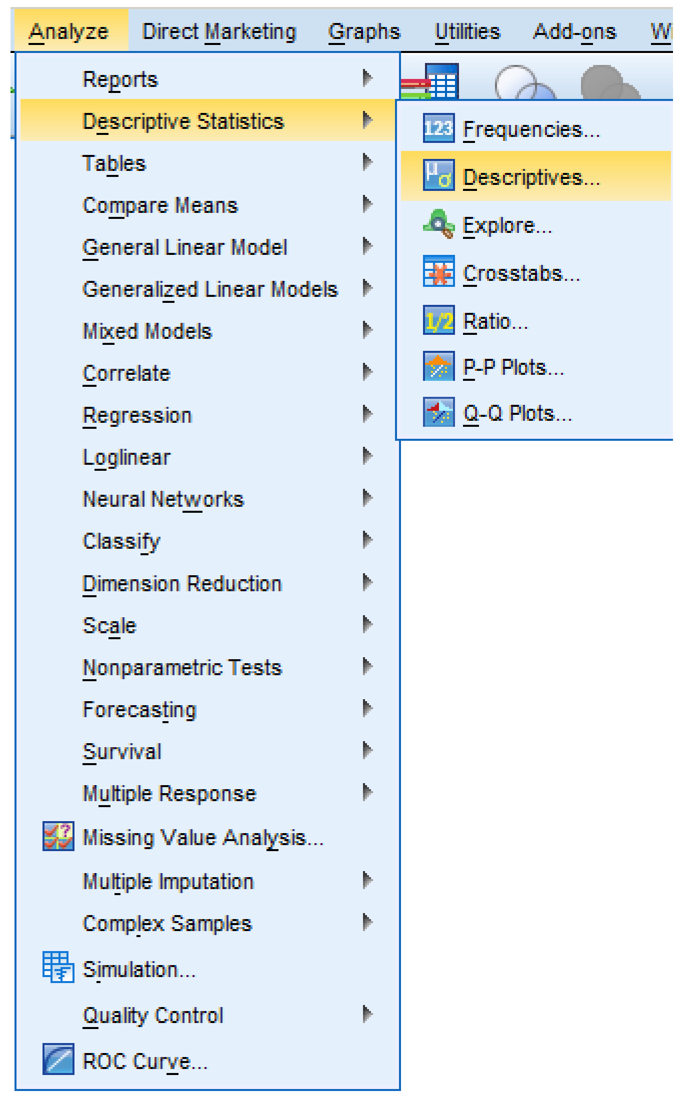


Finally, continue then OK

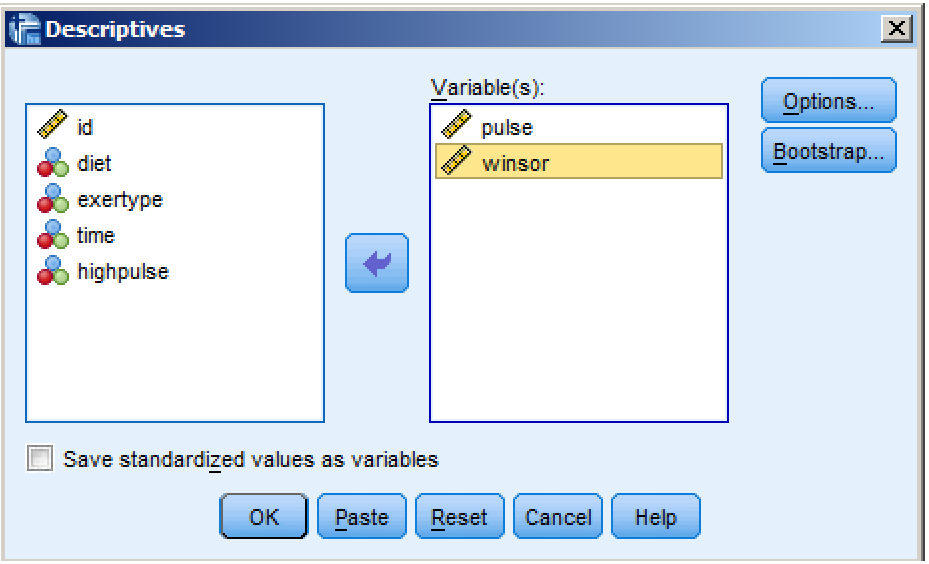
Winsorize

To check your results

Analyze > Descriptive Statistics > Descriptives



Winsorize



OK

Winsorize

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
pulse	90	80.00	150.00	99.7000	14.85847
winsor	90	83.00	137.25	99.4778	14.01753
Valid N (listwise)	90				

As desired.

Winsorize

Syntax:-

```
freq var pulse /format = notable /percentiles = 5 95.

compute winsor = pulse.
if pulse <= 83 winsor = 83.
if pulse >= 137.25 winsor = 137.25.

descriptives variables=pulse winsor
  /statistics=mean stddev min max.
```

Winsorize

Trimming and Winsorization: A review
W. J. Dixon and K. K. Yuen
Statistische Hefte
June 1974, Volume 15, Issue 2-3, pp 157-170 paper

This paper provides a literature review for robust statistical procedures trimming and Winsorization that were first proposed for estimating location, but were later extended to other estimation and testing problems. Performance of these techniques under normal and long-tailed distributions are discussed.

Winsorize

Winsorisation for estimates of change
Daniel Lewis
Papers presented at the ICES-III, June 18-21, 2007, Montreal, Quebec, Canada [paper]

Outliers are a common problem in business surveys which, if left untreated, can have a large impact on survey estimates. For business surveys in the UK Office for National Statistics (ONS), outliers are often treated by modifying their values using a treatment known as Winsorisation. The method involves identifying a cut-off for outliers. Any values lying above the cut-offs are reduced towards the cut-off. The cut-offs are derived in a way that approximately minimises the Mean Square Error of level estimates. However, for many surveys estimates of change are more important. This paper looks at a variety of methods for Winsorising specifically for estimates of change. The measure of change investigated is the difference between two consecutive estimates of total. The first step is to derive potential methods for Winsorising this type of change. Some of these methods prove more practical than others. The methods are then evaluated, using change estimates derived by taking the difference between two regular Winsorised level estimates as a comparison. The evaluation uses data from the ONS Monthly Production Inquiry. Methods are compared both by estimating Mean Squared Errors from survey data and through use of a Monte-Carlo simulation.

Winsorize

Speaking Stata: Trimming to taste
Cox, N.J.
Stata Journal 2013 **13(3)** 640-666 paper
Trimmed means are means calculated after setting aside zero or more values in each tail of a sample distribution. Here we focus on trimming equal numbers in each tail. Such trimmed means define a family or function with mean and median as extreme members and are attractive as simple and easily understood summaries of the general level (location, central tendency) of a variable. This article provides a tutorial review of trimmed means, emphasizing the scope for trimming to varying degrees in describing and exploring data. Detailed remarks are included on the idea's history, plotting of results, and confidence interval procedures. Examples are given using astronomical and medical data. The new Stata commands trimmean and trimplot are also included.

General Linear Models

Generally, the various statistical analyses are taught independently from each other. This makes it difficult to learn new statistical analyses, in contexts that differ. The paper gives a short technical introduction to the general linear model (GLM), in which it is shown that ANOVA (one-way, factorial, repeated measure and analysis of covariance) is simply a multiple correlation/regression analysis (MCRA). Generalizations to other cases, such as multivariate and nonlinear analysis, are also discussed. It can easily be shown that every popular linear analysis can be derived from understanding MCRA.

General Linear Models: An Integrated Approach to Statistics
Sylvain Chartier and Andrew Faulkner
Tutorials in Quantitative Methods for Psychology 2008 **4(2)** 65-78

Does It Always Matter?

Scientists think in terms of confidence intervals – they are inclined to accept a hypothesis if the probability that it is true exceeds 95 per cent. However within the law "beyond reasonable doubt" appears to be a claim that there is a high probability that the hypothesis – the defendant's guilt – is true.

A Story Can Be More Useful Than Maths
John Kay
Financial Times 26 February 2013

Article

Does It Always Matter?

…we slavishly lean on the crutch of significance testing because, if we didn't, much of psychology would simply fall apart. If he was right, then significance testing is tantamount to psychology's "dirty little secret."

Significance tests as sorcery: Science is empirical—significance tests are not
Charles Lambdin
Theory and Psychology 22(1) 67–90 2012

Article

Does It Always Matter?

The first rule of performing a project

1 The supervisor is always right

The second rule of performing a project

2 If the supervisor is wrong, rule 1 applies

Does It Always Matter? Probably!

Estimation based on effect sizes, confidence intervals, and meta-analysis usually provides a more informative analysis of empirical results than does statistical significance testing, which has long been the conventional choice in psychology. The sixth edition of the American Psychological Association Publication Manual now recommends that psychologists should, wherever possible, use estimation and base their interpretation of research results on point and interval estimates.

The statistical recommendations of the American Psychological Association Publication Manual: Effect sizes, confidence intervals, and meta-analysis
Geoff Cumming, Fiona Fidler, Pav Kalinowski and Jerry Lai
Australian Journal of Psychology 2012; 64: 138–146

Article

Index End

SPSS Tips

Now you should *go* and try for yourself.

Each week our cluster (5.05) is booked for 2 hours after this session. This will enable you to come and go as you please.

Obviously other timetabled sessions for this module take precedence.