

*STATISTICS SOLUTIONS*  
*PRESENTS*

# *Statistical Analysis*

---

*A Manual on Dissertation and Thesis*  
*Statistics in SPSS*

*For SPSS Statistics Gradpack Software*

## Table of Contents

Table of Contents .....	2
WELCOME MESSAGE .....	7
CHAPTER 1: FIRST CONTACT WITH SPSS .....	8
What SPSS (SPSS) looks like .....	8
Understanding the applications in the SPSS suite .....	13
SPSS Statistics Base .....	13
SPSS Regression .....	13
SPSS Advanced Statistics .....	14
AMOS .....	15
CHAPTER 2: CHOOSING THE RIGHT STATISTICAL ANALYSIS .....	16
Measurement Scales .....	16
Statistical Analysis Decision Tree .....	17
Decision Tree for Relationship Analyses .....	17
Decision Tree for Comparative Analyses of Differences .....	18
Decision Tree for Predictive Analyses .....	19
Decision Tree for Classification Analyses .....	20
How to Run Statistical Analysis in SPSS .....	20
A Word on Hypotheses Testing .....	20
CHAPTER 3: Introducing the two Examples used throughout this manual .....	22
CHAPTER 4: Analyses of Relationship .....	23
Chi-Square Test of Independence .....	23
What is the Chi-Square Test of Independence? .....	23
Chi-Square Test of Independence in SPSS .....	24
The Output of the Chi-Square Test of Independence .....	27
Bivariate (Pearson) Correlation .....	29
What is a Bivariate (Pearson's) Correlation? .....	29
Bivariate (Pearson's) Correlation in SPSS .....	30
The Output of the Bivariate (Pearson's) Correlation .....	33
Partial Correlation .....	34
What is Partial Correlation? .....	34

How to run the Partial Correlation in SPSS .....	34
The Output of Partial Correlation Analysis .....	38
Spearman Rank Correlation .....	39
What is Spearman Correlation? .....	39
Spearman Correlation in SPSS.....	40
The Output of Spearman's Correlation Analysis.....	41
Point-Biserial Correlation.....	42
What is Point-Biserial Correlation?.....	42
Point-Biserial Correlation Analysis in SPSS.....	44
The output of the Point-Biserial Correlation Analysis .....	46
Canonical Correlation.....	47
What is Canonical Correlation analysis? .....	47
Canonical Correlation Analysis in SPSS .....	49
The Output of the Canonical Correlation Analysis.....	50
CHAPTER 5: Analyses of Differences.....	55
Independent Variable T-Test .....	55
What is the independent variable t-test? .....	55
The Independent variable t-test in SPSS.....	56
Correlations.....	56
One-Way ANOVA .....	58
What is the One-Way ANOVA? .....	58
The One-Way ANOVA in SPSS.....	59
The Output of the One-Way ANOVA .....	63
One-Way ANCOVA .....	64
What is the One-Way ANCOVA? .....	64
The One-Way ANCOVA in SPSS.....	65
The Output of the One-Way ANCOVA .....	67
Factorial ANOVA .....	69
What is the Factorial ANOVA? .....	69

The Factorial ANOVA in SPSS .....	70
The Output of the Factorial ANOVA.....	72
Factorial ANCOVA .....	74
What is the Factorial ANCOVA? .....	74
The Factorial ANCOVA in SPSS .....	75
The Output of the Factorial ANCOVA .....	77
One-Way MANOVA.....	78
What is the One-Way MANOVA?.....	79
The One-Way MANOVA in SPSS.....	81
The Output of the One-Way MANOVA.....	83
One-Way MANCOVA.....	86
What is the One-Way MANCOVA? .....	86
The One-Way MANCOVA in SPSS .....	87
The Output of the One-Way MANCOVA.....	89
Repeated Measures ANOVA .....	91
What is the Repeated Measures ANOVA? .....	91
The Repeated Measures ANOVA in SPSS.....	92
The Output of the Repeated Measures ANOVA .....	95
Repeated Measures ANCOVA .....	98
What is the Repeated Measures ANCOVA?.....	98
The Repeated Measures ANCOVA in SPSS.....	99
The Output of the Repeated Measures ANCOVA .....	101
Profile Analysis .....	105
What is the Profile Analysis?.....	105
The Profile Analysis in SPSS.....	107
The Output of the Profile Analysis .....	109
Double-Multivariate Profile Analysis .....	111
What is the Double-Multivariate Profile Analysis? .....	111



The Double-Multivariate Profile Analysis in SPSS .....	112
The Output of the Double-Multivariate Profile Analysis .....	114
Independent Sample T-Test.....	118
What is the Independent Sample T-Test?.....	118
The Independent Sample T-Test in SPSS.....	119
The Output of the Independent Sample T-Test .....	123
One-Sample T-Test.....	124
What is the One-Sample T-Test? .....	124
The One-Sample T-Test in SPSS .....	125
The Output of the One-Sample T-Test.....	128
Dependent Sample T-Test.....	129
What is the Dependent Sample T-Test? .....	129
The Dependent Sample T-Test in SPSS .....	130
The Output of the Dependent Sample T-Test.....	131
Mann-Whitney U-Test .....	132
What is the Mann-Whitney U-Test? .....	132
The Mann-Whitney U-Test in SPSS .....	133
The Output of the Mann-Whitney U-Test.....	135
Wilcoxon Sign Test .....	136
What is the Wilcoxon Sign Test? .....	136
The Wilcoxon Sign Test in SPSS.....	137
The Output of the Wilcoxon Sign Test .....	138
CHAPTER 6: Predictive Analyses .....	140
Linear Regression.....	140
What is Linear Regression? .....	140
The Linear Regression in SPSS.....	140
The Output of the Linear Regression Analysis .....	143
Multiple Linear Regression .....	145
What is Multiple Linear Regression? .....	145


The Multiple Linear Regression in SPSS .....	146
The Output of the Multiple Linear Regression Analysis.....	150
Logistic Regression .....	154
What is Logistic Regression? .....	154
The Logistic Regression in SPSS.....	155
The Output of the Logistic Regression Analysis .....	157
Ordinal Regression .....	160
What is Ordinal Regression? .....	160
The Ordinal Regression in SPSS.....	161
The Output of the Ordinal Regression Analysis .....	164
CHAPTER 7: Classification Analyses .....	166
Multinomial Logistic Regression .....	166
What is Multinomial Logistic Regression? .....	166
The Multinomial Logistic Regression in SPSS.....	167
The Output of the Multinomial Logistic Regression Analysis .....	170
Sequential One-Way Discriminant Analysis.....	173
What is the Sequential One-Way Discriminant Analysis?.....	173
The Sequential One-Way Discriminant Analysis in SPSS.....	175
The Output of the Sequential One-Way Discriminant Analysis .....	176
Cluster Analysis .....	179
What is the Cluster Analysis? .....	179
The Cluster Analysis in SPSS.....	181
The Output of the Cluster Analysis .....	186
Factor Analysis .....	189
What is the Factor Analysis? .....	189
The Factor Analysis in SPSS.....	190
The Output of the Factor Analysis .....	194
CHAPTER 8: Data Analysis and Statistical Consulting Services .....	199
Terms of Use .....	200

## WELCOME MESSAGE

Statistics Solutions is dedicated to expediting the dissertation and thesis process for graduate students by providing statistical help and guidance to ensure a successful graduation. Having worked on my own mixed method (qualitative and quantitative) dissertation, and with over 18 years of experience in research design, methodology, and statistical analyses, I present this SPSS user guide, on behalf of Statistics Solutions, as a gift to you.

The purpose of this guide is to enable students with little to no knowledge of SPSS to open the program and conduct and interpret the most common statistical analyses in the course of their dissertation or thesis. Included is an introduction explaining when and why to use a specific test as well as where to find the test in SPSS and how to run it. Lastly, this guide lets you know what to expect in the results and informs you how to interpret the results correctly.

Statistics Solutions' offers a family of solutions to assist you towards your degree. If you would like to learn more or schedule your free 30-minute consultation to discuss your dissertation research, you can visit us at [www.StatisticsSolutions.com](http://www.StatisticsSolutions.com) or call us at 877-437-8622.

<b>Proposal Phase</b> Formulating hypotheses, research questions, research design Methodological review forms and all pre-dissertation paperwork	
<b>Chapter 3: Methods</b> Formulating hypotheses, research questions, research design Survey instrument help and administration, transcription, data collection, cleaning and coding Sample size calculation and justification/power analysis. Data Analysis, statistical analysis, assistance with choosing the appropriate statistical analysis. Justification of the analysis Statistics and methodology coaching Quantitative and qualitative expertise	
<b>Chapter 4: Results</b> Conduct analysis and examine assumptions Explaining the findings and statistics, formulating justification, writing the results Explaining the results in a way that shows clear support or non-support of hypotheses APA-style tables and results	
<b>APA-Editing</b> APA-formatted results and tables Proposal, dissertation and thesis editing in APA-style	
<b>Revisions and Defense</b> Assistance with committee issues and revisions Dissertation defense coaching	
<b>SPSS Software and Coaching</b> Free online webinars focused on understanding SPSS software and conducting analyses SPSS manual designed specifically for doctoral and masters-level statistics SPSS coaching and support  Statistics Solutions is an authorized partner with SPSS, an IBM Company and offers SPSS Statistics GradPack Software. Purchase online for only \$188.	

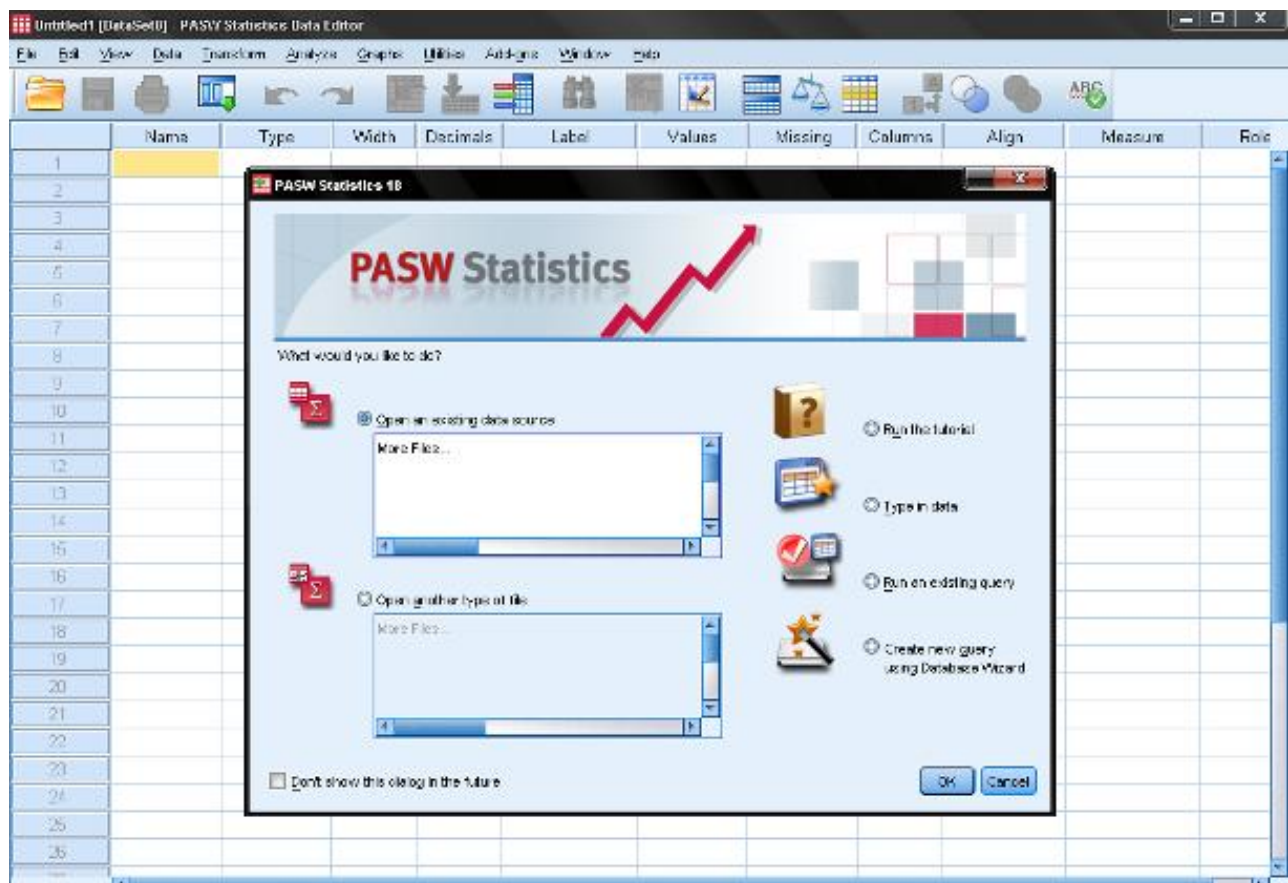
## CHAPTER 1: FIRST CONTACT WITH SPSS

SPSS stands for **Software Package for the Social Sciences** and was rebranded in version 18 to SPSS (**Predictive Analytics Software**). Throughout this manual, we will employ the rebranded name, SPSS. The screenshots you will see are taken from version 18. If you use an earlier version, some of the paths might be different because the makers of SPSS sometimes move the menu entries around. If you have worked with older versions before, the two most noticeable changes are found within the graph builder and the non-paracontinuous-level tests.

### *What SPSS (SPSS) looks like*

When you open SPSS you will be greeted by the opening dialog. Typically, you would type in data or open an existing data file.

SPSS has three basic windows: the data window, the syntax window, and the output window. The particular view can be changed by going to the Window menu. What you typically see first is the data window.



**Data Window.** The data editor window is where the data is either inputted or imported. The data editor window has two views—the data view and the variable view. These two windows can be swapped by clicking the buttons on the lower left corner of the data window. In the data view, your data is presented in a spreadsheet style very similar to Excel. The data is organized in rows and columns. Each row represents an observation and each column represents a variable.

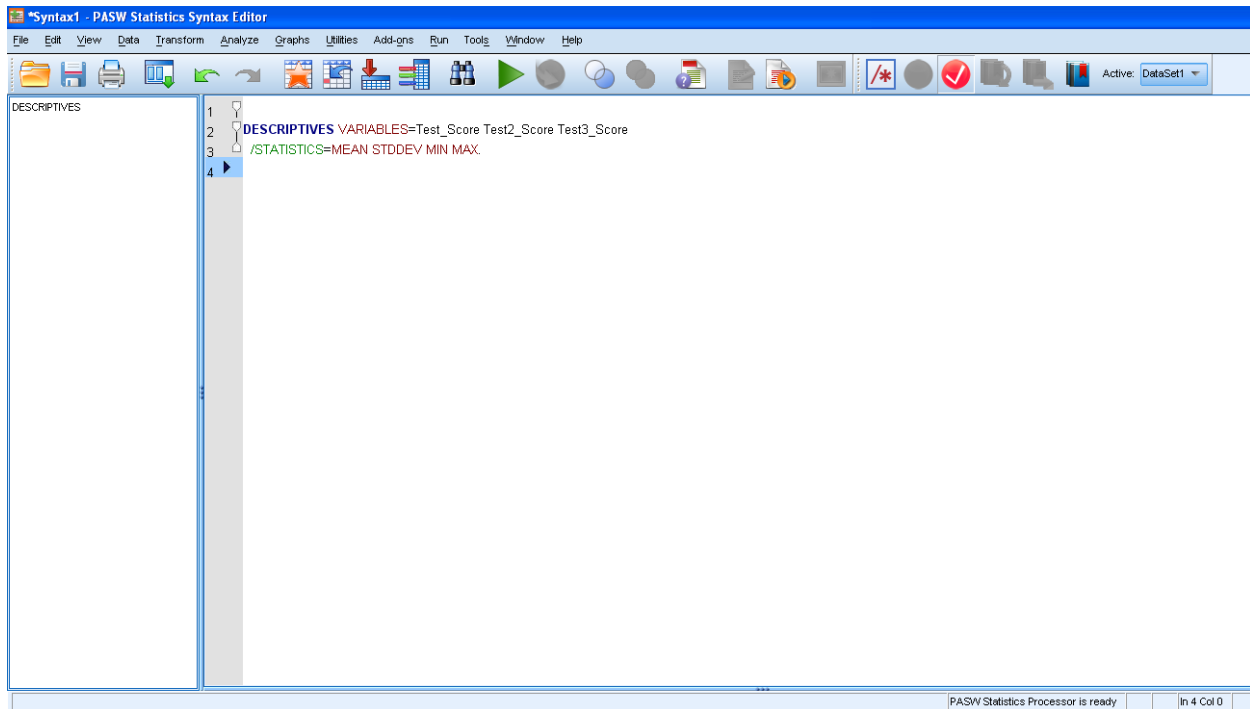
	Test_Score	Test2_Score	Test3_Score	Gender	Apt1	Apt2	Apt3	Apt4	Apt5	Exam	Grade1	Grade2	Grade3
1	66.00	32	22	1	33	16	30	25	25	1.00	2.00	2.00	3.00
2	26.00	0	0	1	11	Aptitude Test 2	3	39	37	.0	6.00	6.00	6.00
3	63.00	33	47	0	46	21	32	24	21	1.00	2.00	2.00	1.00
4	98.00	20	23	1	16	36	20	28	37	1.00	1.00	3.00	2.00
5	43.00	33	28	0	39	28	36	27	21	.0	4.00	2.00	2.00
6	83.00	28	23	1	36	19	28	31	31	1.00	1.00	2.00	2.00
7	46.00	37	20	1	44	23	23	35	23	.0	4.00	1.00	3.00
8	49.00	24	22	1	22	23	20	26	28	.0	3.00	3.00	3.00
9	81.00	37	28	1	37	17	27	33	23	1.00	1.00	1.00	2.00
10	66.00	47	20	1	45	20	14	21	19	1.00	2.00	1.00	3.00
11	53.00	19	32	1	40	7	11	41	33	1.00	3.00	4.00	1.00
12	73.00	34	33	1	35	19	33	26	36	1.00	1.00	2.00	1.00
13	48.00	46	31	1	36	9	26	26	33	.0	4.00	1.00	1.00
14	47.00	12	14	0	19	31	29	35	36	.0	4.00	5.00	4.00
15	63.00	8	23	1	31	30	23	28	28	1.00	2.00	5.00	2.00
16	91.00	36	31	1	41	31	23	13	19	1.00	1.00	1.00	1.00
17	34.00	33	20	0	44	18	17	42	27	.0	6.00	2.00	3.00
18	51.00	36	26	1	33	20	30	26	34	.0	3.00	1.00	2.00
19	49.00	2	7	0	10	38	26	40	38	.0	3.00	6.00	5.00

In the **variable view**, the logic behind each variable is stored. Each variable has a name (in the name column), a type (numeric, percentage, date, string etc.), a label (usually the full wording of the question), and the values assigned to the level of the variable in the “values” column. For example, in

the name column we may have a variable called “gender.” In the label column we may specify that the variable is the “gender of participants.” In the values box, we may assign a “1” for males and a “2” for females. You can also manage the value to indicate a missing answer, the measurement level – scale (which is metric, ratio, or interval data), ordinal or nominal, and new to SPSS 18 a pre-defined role.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	Test_Score	Numeric	8	2	Math Test	None	None	8	Right	Scale	Input
2	Test2_Score	Numeric	8	0	Reading Test	None	None	5	Right	Scale	Input
3	Test3_Score	Numeric	8	0	Writing Test	None	None	5	Right	Scale	Input
4	Gender	Numeric	8	0	Gender	{0, Male}...	None	8	Right	Nominal	Input
5	Apt1	Numeric	8	0	Aptitude Test 1	None	None	5	Right	Scale	Input
6	Apt2	Numeric	8	0	Aptitude Test 2	None	None	5	Right	Scale	Input
7	Apt3	Numeric	8	0	Aptitude Test 3	None	None	5	Right	Scale	Input
8	Apt4	Numeric	8	0	Aptitude Test 4	None	None	5	Right	Scale	Input
9	Apt5	Numeric	8	0	Aptitude Test 5	None	None	5	Right	Scale	Input
10	Exam	Numeric	8	2	Exam	{0, Fail}...	None	10	Right	Nominal	Input
11	Grade1	Numeric	8	2	Grade on Math ...	{1,00, A}...	None	10	Right	Ordinal	Input
12	Grade2	Numeric	8	2	Grade on Readin...	{1,00, A}...	None	10	Right	Ordinal	Input
13	Grade3	Numeric	8	2	Grade on Writin...	{1,00, A}...	None	10	Right	Ordinal	Input
14	Good1	Numeric	8	2	Performance on...	{0, Not goo...	None	8	Right	Ordinal	Input
15	Good2	Numeric	8	2	Performance on...	{0, Not goo...	None	8	Right	Ordinal	Input
16	Good3	Numeric	8	2	Performance on...	{0, Not goo...	None	8	Right	Ordinal	Input
17	Age	Numeric	8	2	Age	None	None	10	Right	Scale	Input
18	Final_exam	Numeric	8	2	Final Exam Sc...	{1,00, Fail}...	None	12	Right	Ordinal	Input
19	Ex1	Numeric	8	2	Mid-term Exam 1	{1,00, Low ...	None	10	Right	Ordinal	Input
20	Ex4	Numeric	8	2	Mid-term Exam 4	{1,00, Low ...	None	10	Right	Ordinal	Input
21	Ex5	Numeric	8	2	Mid-term Exam 5	{1,00, Low ...	None	10	Right	Ordinal	Input
22	Ex2	Numeric	8	2	Mid-term Exam 2	{1,00, Low ...	None	10	Right	Ordinal	Input
23	Ex3	Numeric	8	2	Mid-term Exam 3	{1,00, Low ...	None	10	Right	Ordinal	Input
24	Treatment	Numeric	8	2	Teaching Meth...	{1,00, Front...	None	11	Right	Ordinal	Input
25	Gift	Numeric	8	2	Gift chosen by ...	{1,00, Super...	None	10	Right	Nominal	Input
26											

**The Syntax Window.** In the syntax editor window you can program SPSS. Although it isn't necessary to program syntax for virtually all analyses, using the syntax editor is quite useful for two purposes: 1) to save your analysis steps and 2) to run repetitive tasks. Firstly, you can document your analysis steps and save them in a syntax file, so others may re-run your tests and you can re-run them as well. To do this you simply hit the PASTE button you find in most dialog boxes. Secondly, if you have to repeat a lot of steps in your analysis, for example, calculating variables or re-coding, it is most often easier to specify these things in syntax, which saves you the time and hassle of scrolling and clicking through endless lists of variables.



**The Output Window.** The output window is where SPSS presents the results of the analyses you conducted. Besides the usual status messages, you'll find all of the results, tables, and graphs in here. In the output window you can also manipulate tables and graphs and reformat them (e.g., to APA 6<sup>th</sup> edition style).

GET  
FILE='D:\Documents and Settings\Alexander Budzier\My Documents\StatisticsSolution.com\30 - SPSS Manual\Example.sav'.  
DATASET NAME DataSet1 WINDOW=FRONT.  
FREQUENCIES VARIABLES=Test\_Score Test2\_Score Test3\_Score  
/ORDER=ANALYSIS.

→ **Frequencies**

[DataSet1] D:\Documents and Settings\Alexander Budzier\My Documents\StatisticsSolution.com\30 -

**Statistics**

	Math Test	Reading Test	Writing Test
N Valid	107	107	107
Missing	0	0	0

**Frequency Table**

**Math Test**

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 26,00	2	1,9	1,9	1,9
27,00	1	,9	,9	2,8
28,00	1	,9	,9	3,7
29,00	1	,9	,9	4,7
30,00	3	2,8	2,8	7,5
31,00	1	,9	,9	8,4



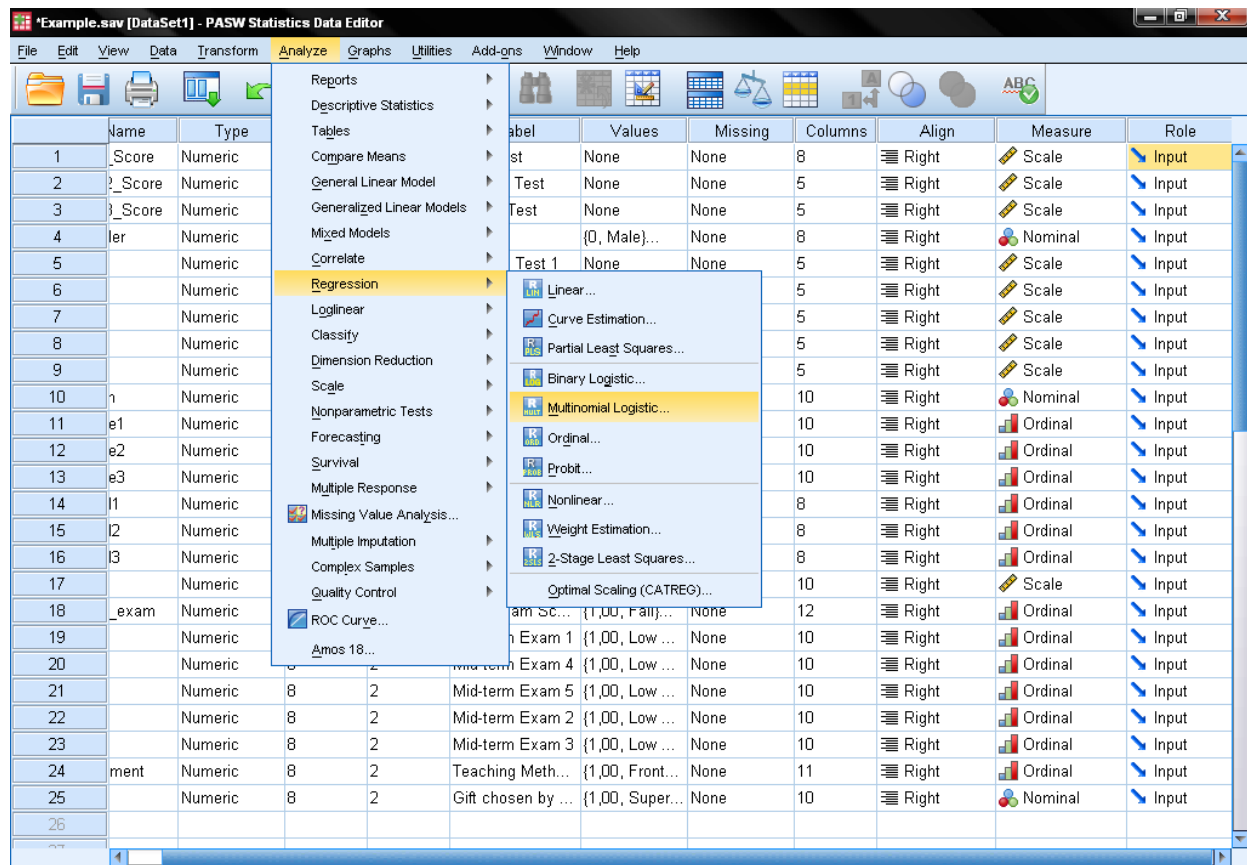
## Understanding the applications in the SPSS suite

### SPSS Statistics Base

The SPSS Statistics Base program covers all of your basic statistical needs. It includes crosstabs, frequencies, descriptive statistics, correlations, and all comparisons of mean scores (e.g., t-tests, ANOVAs, non-paracontinuous-level tests). It also includes the predictive methods of factor and cluster analysis, linear and ordinal regression, and discriminant analysis.

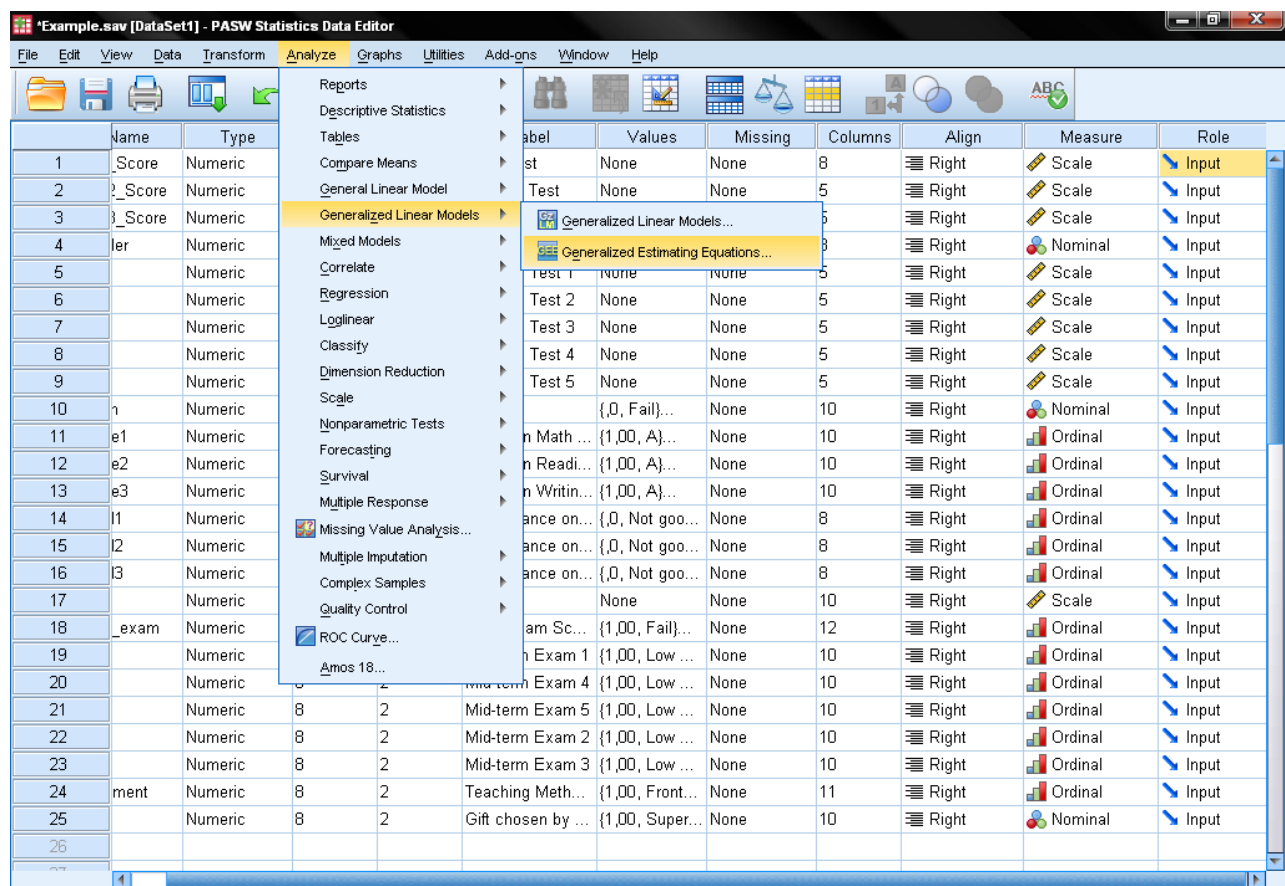
### SPSS Regression

SPSS Regression is the add-on to enlarge the regression analysis capabilities of SPSS. This module includes multinomial and binary logistic regression, constrained and unconstrained nonlinear regression, weighted least squares, and probit.



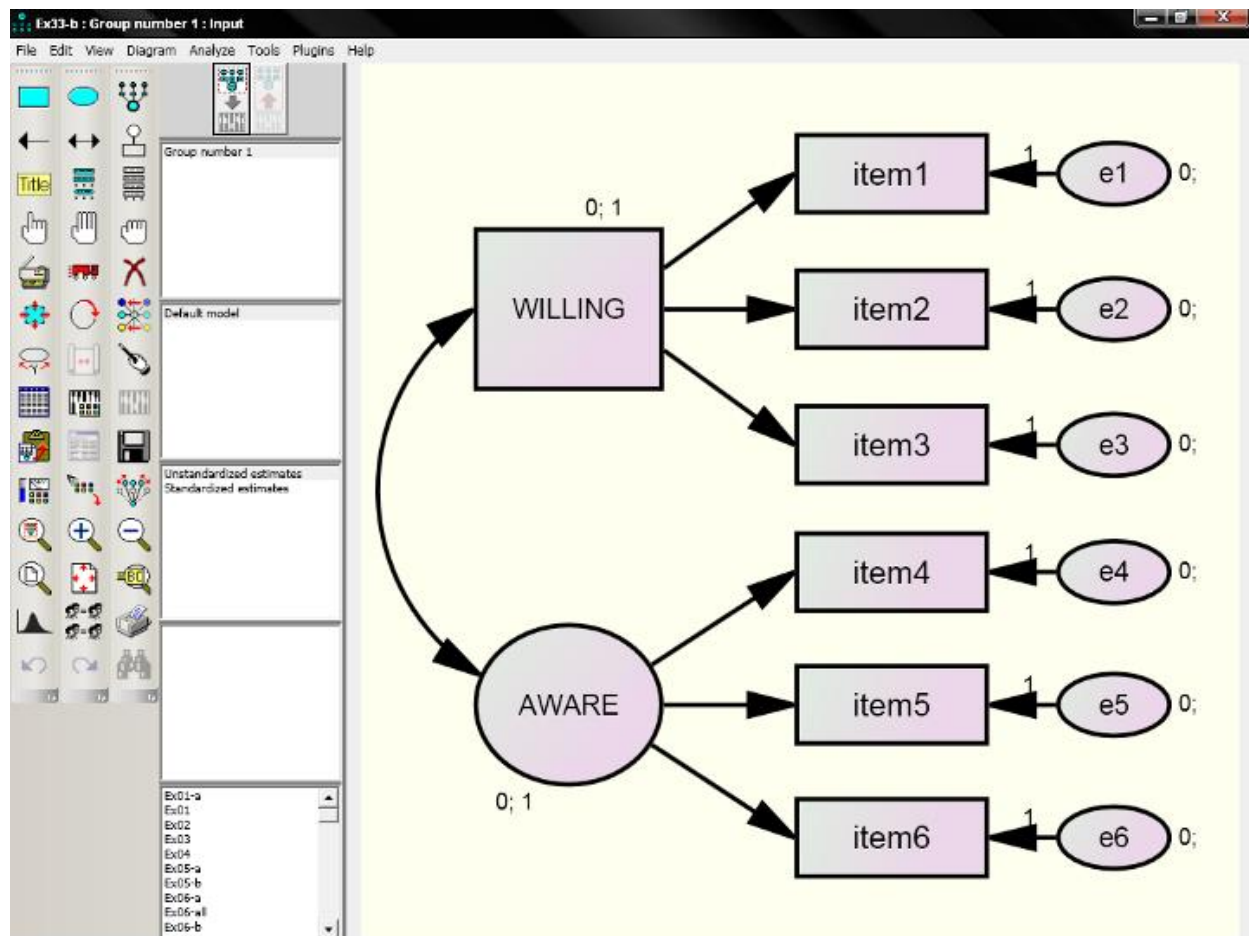
## SPSS Advanced Statistics

SPSS Advanced Statistics is the most powerful add-on for all of your regression and estimation needs. It includes the generalized linear models and estimation equations, and also hierarchical linear modeling. Advanced Statistics also includes Survival Analysis.



## AMOS

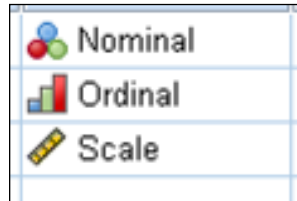
AMOS is a program that allows you to specify and estimate structural equation models. Structural equation models are published widely especially in the social sciences. In basic terms, structural equation models are a fancy way of combining multiple regression analyses and interaction effects.



## CHAPTER 2: CHOOSING THE RIGHT STATISTICAL ANALYSIS

This manual is a guide to help you select the appropriate statistical technique. Your quantitative study is a process that presents many different options that can lead you down different paths. With the help of this manual, we will ensure that you choose the right paths during your statistical selection process. The next section will help guide you towards the right statistical test for your work. However, before you can select a test it will be necessary to know a thing or two about your data.

When it comes to selecting your test, the level of measurement of your data is important. The



measurement level is also referred to as the scale of your data. The easy (and slightly simplified) answer is that there are three different levels of measurement: nominal, ordinal, and scale. In your SPSS data editor the measure column looks can have exactly those three values.

### Measurement Scales

**Nominal data** is the most basic level of measurement. All data is at least nominal. A characteristic is measured on a nominal scale if the answer contains different groups or categories like male/female; treatment group/control group; or multiple categories like colors or occupations, highest degrees earned, et cetera.

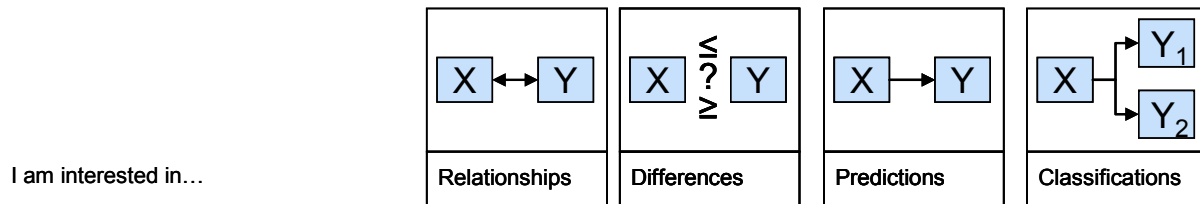
**Ordinal data** contains a bit more information than nominal data. On an ordinal scale your answer is one of a set of different possible groups like on a nominal scale, however the ordinal scale allows you to order the answer choices. Examples of this include all questions where the answer choices are grouped in ranges, like income bands, age groups, and diagnostic cut-off values, and can also include rankings (first place, second place, third place), and strengths or quantities of substances (high dose/ medium dose/ low dose).

**Scale data** also contains more information than nominal data. If your data is measured on a continuous-level scale then the intervals and/or ratios between your groups are defined. Technically, you can define the distance between two ordinal groups by either a ratio or by an interval. What is a **ratio scale**? A ratio scale is best defined by what it allows you to do. With scale data you *can* make claims such as 'first place is twice as good as second place', whereas on an ordinal scale you are unable to make these claims for you cannot know them for certain. Fine examples of scale data include the findings that a temperature of 120°K is half of 240°K, and sixty years is twice as many years as thirty years, which is twice as many years as fifteen. What is an **interval scale**? An interval scale enables you to establish intervals. Examples include the findings that the difference between 150ml and 100ml is the same as the difference between 80ml and 30ml, and five minus three equals two which is the same as twelve minus two. Most often you'll also find Likert-like scales in the interval scale category of levels of measurement. An example of a Likert-like scale would include the following question and statements: How satisfied are you with your life? Please choose an answer from 1 to 7, where 1 is completely dissatisfied, 2 is dissatisfied, 3 is somewhat dissatisfied, 4 is neither satisfied or dissatisfied, 5 is somewhat satisfied, 6 is satisfied, and 7 is completely satisfied. These scales are typically interval scales and not ratio scales because you cannot really claim that dissatisfied (2) is half as satisfied as neither (4). Similarly, logarithmic scales such as those you find in a lot of indices don't have the same intervals

between values, but the distance between observations can be expressed by ratios. [A word of caution: statisticians often become overly obsessed with the latter category; they want to know for instance if that scale has a natural zero. For our purposes it is enough to know that if the distance between groups can be expressed as an interval or ratio, we run the more advanced tests. In this manual we will refer to interval or ratio data as being of continuous-level scale.]

## Statistical Analysis Decision Tree

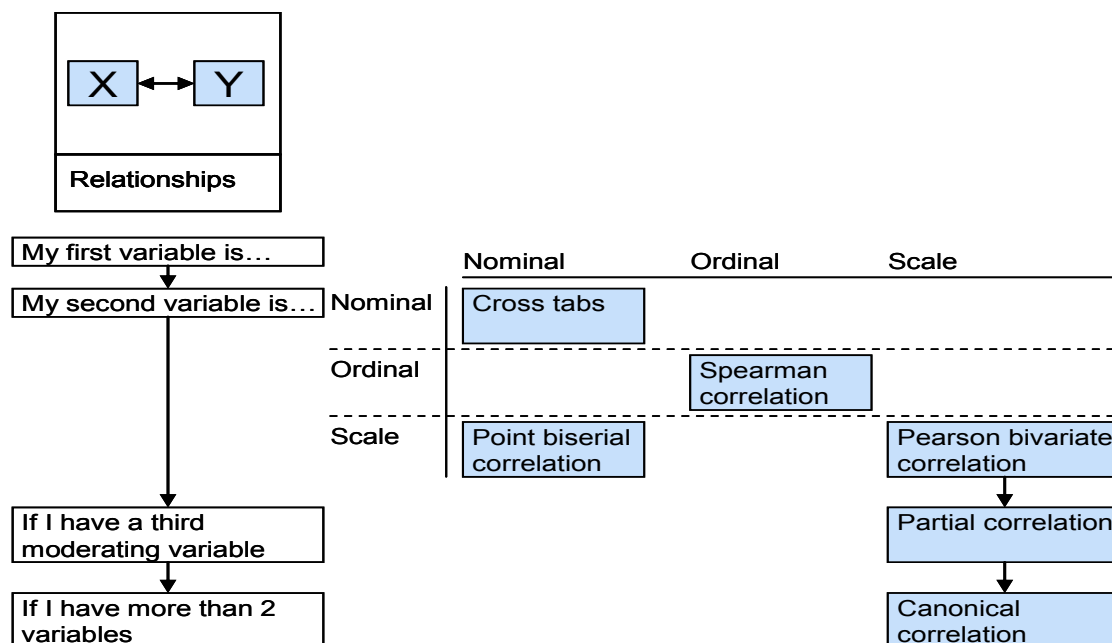
A good starting point in your statistical research is to find the category in which your research question falls.



Are you interested in the relationship between two variables, for example, the higher X and the higher Y? Or are you interested in comparing differences such as, “X is higher for group A than it is for group B?” Are you interested in predicting an outcome variable like, “how does Y increase for one more unit of X?” Or are you interested in classifications, for example, “with seven units of X does my participant fall into group A or B?”

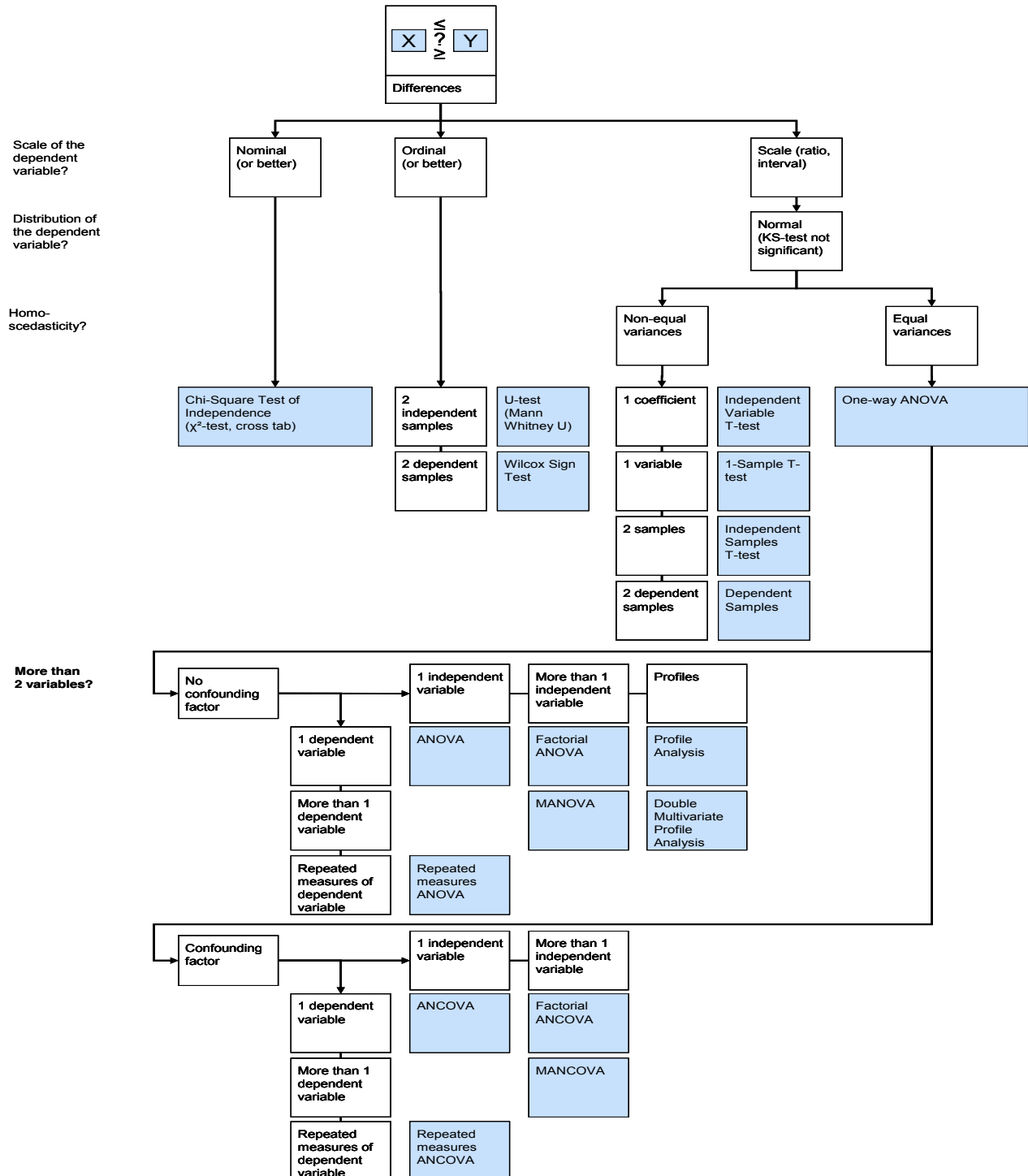
## Decision Tree for Relationship Analyses

The level of measurement of your data mainly defines which statistical test you should choose. If you have more than two variables, you need to understand whether you are interested in the direct or indirect (moderating) relations between those additional variables.



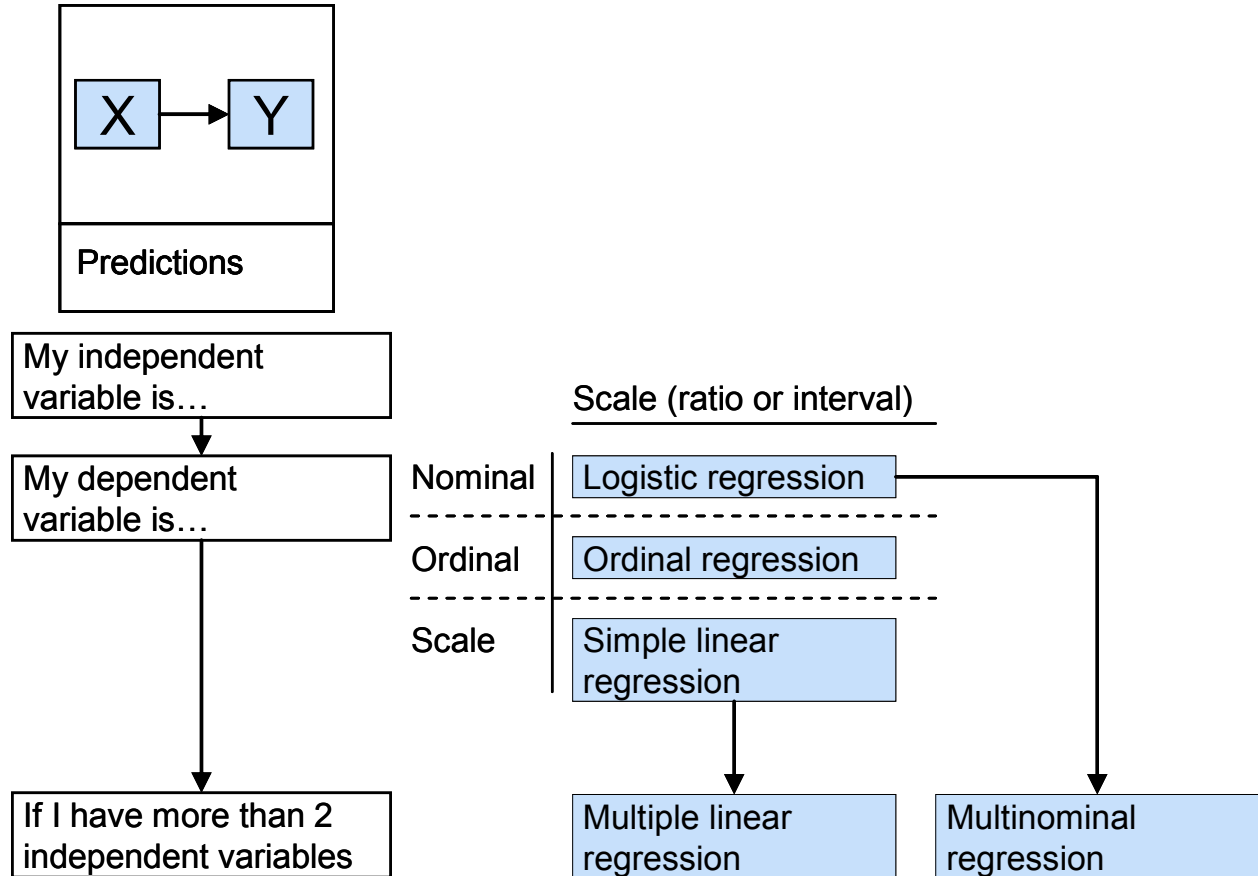
## Decision Tree for Comparative Analyses of Differences

You have chosen the largest family of statistical techniques. The choices may be overwhelming, but start by identifying your dependent variable's scale level, then check assumptions from simple (no assumptions for a Chi-Square) to more complex tests (ANOVA).



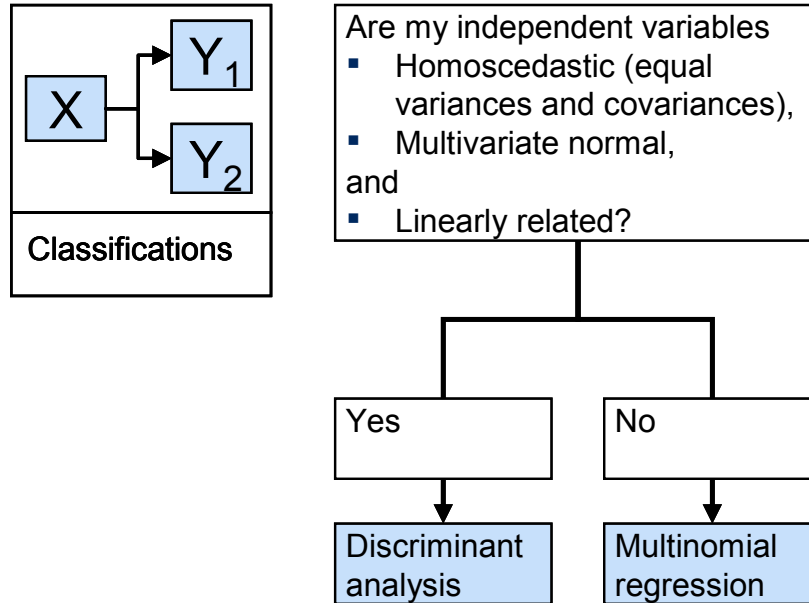
## Decision Tree for Predictive Analyses

You have chosen a straightforward family of statistical techniques. Given that your independent variables are often continuous-level data (interval or ratio scale), you need only consider the scale of your dependent variable and the number of your independent variables.

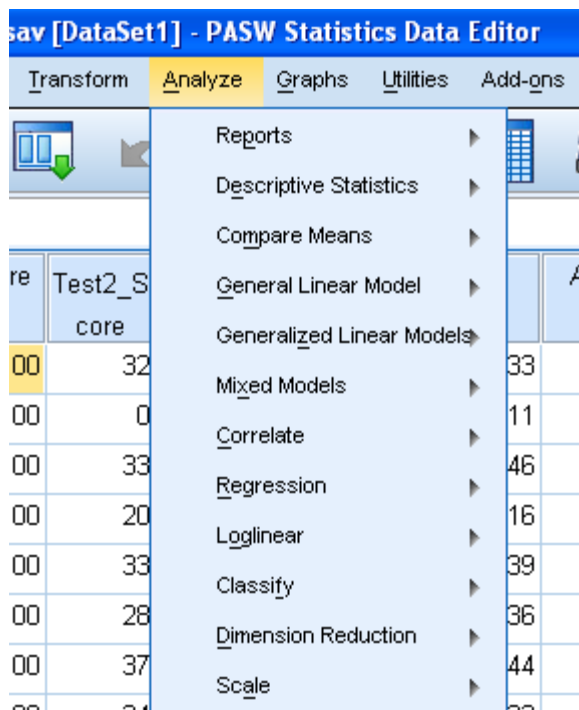


## Decision Tree for Classification Analyses

If you want to classify your observations you only have two choices. The discriminant analysis has more reliability and better predictive power, but it also makes more assumptions than multinomial regression. Thoroughly weigh your two options and choose your own statistical technique.



## How to Run Statistical Analysis in SPSS



Running statistical tests in SPSS is very straightforward, as SPSS was developed for this purpose. All tests covered in this manual are part of the *Analyze* menu. In the following chapters we will always explain how to click to the right dialog window and how to fill it correctly. Once SPSS has run the test, the results will be presented in the *Output* window. This manual offers a very concise write-up of the test as well, so you will get an idea of how to phrase the interpretation of the test results and reference the test's null hypothesis.

## A Word on Hypotheses Testing



In quantitative testing we are always interested in the question, *“Can I generalize my findings from my sample to the general population?”* This question refers to the external validity of the analysis. The ability to establish external validity of findings and to measure it with statistical power is one of the key strengths of quantitative analyses. To do so, every statistical analysis includes a hypothesis test. If you took statistics as an undergraduate, you may remember the null hypothesis and levels of significance.

In SPSS, all tests of significance give a p-value. The p-value is the statistical power of the test. The critical value that is widely used for p is 0.05. That is, for  $p \leq 0.05$  we can reject the null hypothesis; in most tests this means that we might generalize our findings to the general population. In statistical terms, generalization creates the probability of wrongly rejecting a correct null hypothesis and thus the p-value is equal to the Type I or alpha-error.

Remember: If you run a test in SPSS, if the p-value is less than or equal to 0.05, what you found in the sample is externally valid and can be generalized onto the population.

## CHAPTER 3: Introducing the two Examples used throughout this manual

In this manual we will rely on the example data gathered from a fictional educational survey. The sample consists of 107 students aged nine and ten. The pupils were taught by three different methods (frontal teaching, small group teaching, blended learning, i.e., a mix of classroom and e-learning).

Among the data collected were multiple test scores. These were standardized test scores in mathematics, reading, writing, and five aptitude tests that were repeated over time. Additionally the pupils got grades on final and mid-term exams. The data also included several mini-tests which included newly developed questions for the standardized tests that were pre-tested, as well as other variables such as gender and age. After the team finished the data collection, every student was given a computer game (the choice of which was added as a data point as well).

The data set has been constructed to illustrate all the tests covered by the manual. Some of the results switch the direction of causality in order to show how different measurement levels and number of variables influence the choice of analysis. The full data set contains 37 variables from 107 observations:

Education Example.sav [DataSet1] - PASW Statistics Data Editor										
	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	Test_Score	Numeric	8	2	Math Test	None	None	8	Right	Scale
2	Test2_Score	Numeric	8	0	Reading Test	None	None	5	Right	Scale
3	Test3_Score	Numeric	8	0	Writing Test	None	None	5	Right	Scale
4	Gender	Numeric	8	0	Gender	{0, Male}...	None	8	Right	Nominal
5	Apt1	Numeric	8	0	Aptitude Test 1	None	None	5	Right	Scale
6	Apt2	Numeric	8	0	Aptitude Test 2	None	None	5	Right	Scale
7	Apt3	Numeric	8	0	Aptitude Test 3	None	None	5	Right	Scale
8	Apt4	Numeric	8	0	Aptitude Test 4	None	None	5	Right	Scale
9	Apt5	Numeric	8	0	Aptitude Test 5	None	None	5	Right	Scale
10	Exam	Numeric	8	2	Exam	{0, Fail}...	None	10	Right	Nominal
11	Grade1	Numeric	8	2	Grade on Math Test	{1.00, A}...	None	10	Right	Ordinal
12	Grade2	Numeric	8	2	Grade on Reading Test	{1.00, A}...	None	10	Right	Ordinal
13	Grade3	Numeric	8	2	Grade on Writing Test	{1.00, A}...	None	10	Right	Ordinal
14	Good1	Numeric	8	2	Performance on Math Test	{0, Not goo...	None	8	Right	Ordinal
15	Good2	Numeric	8	2	Performance on Reading Test	{0, Not goo...	None	8	Right	Ordinal
16	Good3	Numeric	8	2	Performance on Writing test	{0, Not goo...	None	8	Right	Ordinal
17	Age	Numeric	8	2	Age	None	None	10	Right	Scale
18	Final_exam	Numeric	8	2	Final Exam Score	{1.00, Fail}...	None	12	Right	Ordinal
19	Ex1	Numeric	8	2	Grade on Mid-Term Exam 1	{1.00, A}...	None	10	Right	Ordinal
20	Ex2	Numeric	8	2	Grade on Mid-Term Exam 2	{1.00, A}...	None	10	Right	Ordinal
21	Treatment	Numeric	8	2	Teaching Methods	{1.00, Front...	None	11	Right	Ordinal
22	Gift	Numeric	8	2	Gift chosen by pupil	{1.00, Super...	None	10	Right	Nominal
23	Test1_1	Numeric	8	2		None	None	8	Right	Scale
24	Test1_2	Numeric	8	2		None	None	8	Right	Scale
25	Test1_3	Numeric	8	2		None	None	8	Right	Scale
26	Test2_1	Numeric	8	2		None	None	8	Right	Scale
27	Test2_2	Numeric	8	2		None	None	8	Right	Scale
28	Test2_3	Numeric	8	2		None	None	8	Right	Scale
29	Test3_1	Numeric	8	2		None	None	8	Right	Scale
30	Test3_2	Numeric	8	2		None	None	8	Right	Scale
31	Test3_3	Numeric	8	2		None	None	8	Right	Scale
32	Test4_1	Numeric	8	2		None	None	8	Right	Scale
33	Test4_2	Numeric	8	2		None	None	8	Right	Scale
34	Test4_3	Numeric	8	2		None	None	8	Right	Scale
35	Test5_1	Numeric	8	2		None	None	8	Right	Scale
36	Test5_2	Numeric	8	2		None	None	8	Right	Scale
37	Test5_3	Numeric	8	2		None	None	8	Right	Scale
38										

### Chi-Square Test of Independence

#### What is the Chi-Square Test of Independence?

The Chi-Square Test of Independence is also known as Pearson's Chi-Square, Chi-Squared, or  $\chi^2$ .  $\chi$  is the Greek letter Chi. The Chi-Square Test has two major fields of application: 1) *goodness of fit* test and 2) *test of independence*.

Firstly, the Chi-Square Test can test whether the distribution of a variable in a sample approximates an assumed theoretical distribution (e.g., normal distribution, Beta). *[Please note that the Kolmogorov-Smirnov test is another test for the goodness of fit. The Kolmogorov-Smirnov test has a higher power, but can only be applied to continuous-level variables.]*

Secondly, the Chi-Square Test can be used to test of independence between two variables. That means that it tests whether one variable is independent from another one. In other words, it tests whether or not a statistically significant relationship exists between a dependent and an independent variable. When used as test of independence, the Chi-Square Test is applied to a contingency table, or cross tabulation (sometimes called crosstabs for short).

Typical questions answered with the Chi-Square Test of Independence are as follows:

- **Medicine** - Are children more likely to get infected with virus A than adults?
- **Sociology** - Is there a difference between the marital status of men and woman in their early 30s?
- **Management** - Is customer segment A more likely to make an online purchase than segment B?
- **Economy** - Do white-collar employees have a brighter economical outlook than blue-collar workers?

As we can see from these questions and the decision tree, the Chi-Square Test of Independence works with nominal scales for both the dependent and independent variables. These example questions ask for answer choices on a nominal scale or a tick mark in a distinct category (e.g., male/female, infected/not infected, buy online/do not buy online).

In more academic terms, most quantities that are measured can be proven to have a distribution that approximates a Chi-Square distribution. Pearson's Chi Square Test of Independence is an approximate test. This means that the assumptions for the distribution of a variable are only approximately Chi-Square. This approximation improves with large sample sizes. However, it poses a problem with small sample sizes, for which a typical cut-off point is a cell size below five expected occurrences.

Taking this into consideration, Fisher developed an exact test for contingency tables with small samples. Exact tests do not approximate a theoretical distribution, as in this case Chi-Square distribution. Fisher's exact test calculates all needed information from the sample using a hypergeometric-level distribution.

What does this mean? Because it is an exact test, a significance value  $p$  calculated with Fisher's Exact Test will be correct; i.e., when  $p = 0.01$  the test (in the long run) will actually reject a true null hypothesis in 1% of all tests conducted. For an approximate test such as Pearson's Chi-Square Test of Independence this is only asymptotically the case. Therefore the exact test has exactly the Type I Error ( $\alpha$ -Error, false positives) it calculates as  $p$ -value.

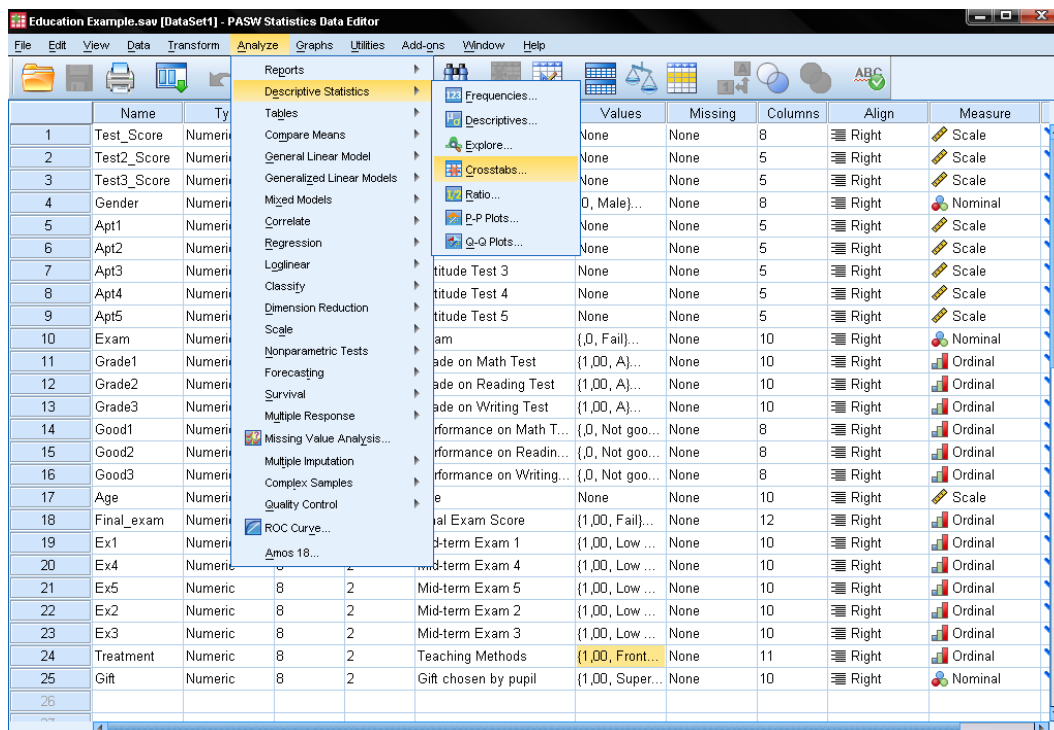
When applied to a research problem, however, this difference might simply have a smaller impact on the results. The rule of thumb is to use exact tests with sample sizes less than ten. Also both Fisher's exact test and Pearson's Chi-Square Test of Independence can be easily calculated with statistical software such as SPSS.

The Chi-Square Test of Independence is the simplest test to prove a causal relationship between an independent and one or more dependent variables. As the decision-tree for tests of independence shows, the Chi-Square Test can always be used.

### Chi-Square Test of Independence in SPSS

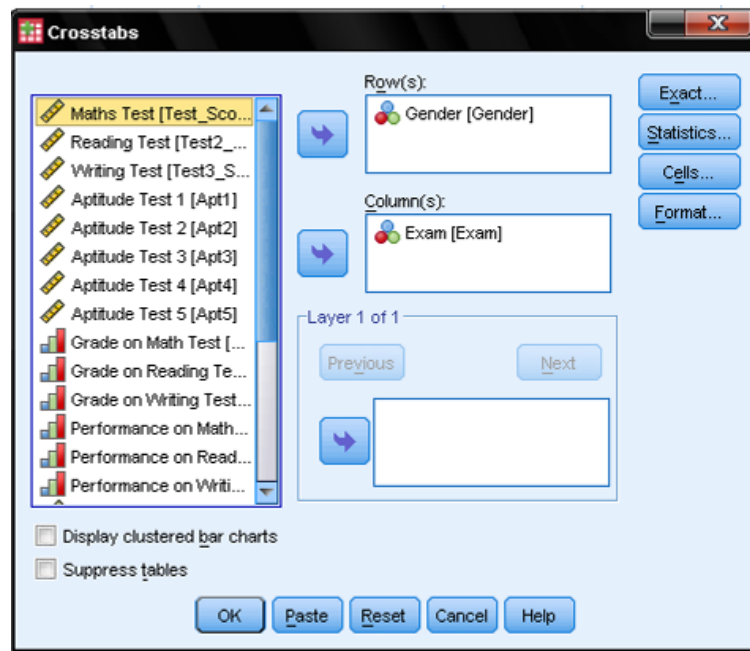
In reference to our education example we want to find out whether or not there is a gender difference when we look at the results (pass or fail) of the exam.

The Chi-Square Test of Independence can be found in *Analyze/Descriptive Statistics/Crosstabs...*

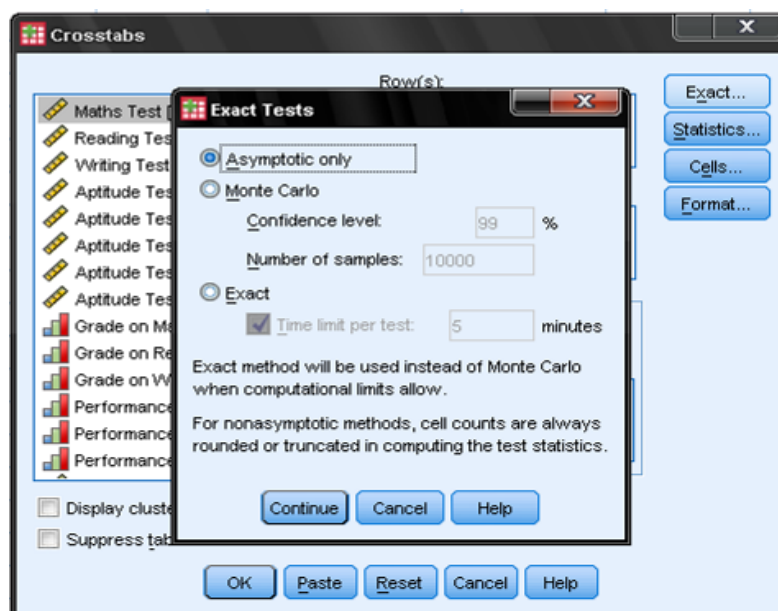


This menu entry opens the crosstabs menu. Crosstabs is short for cross tabulation, which is sometimes referred to as contingency tables.

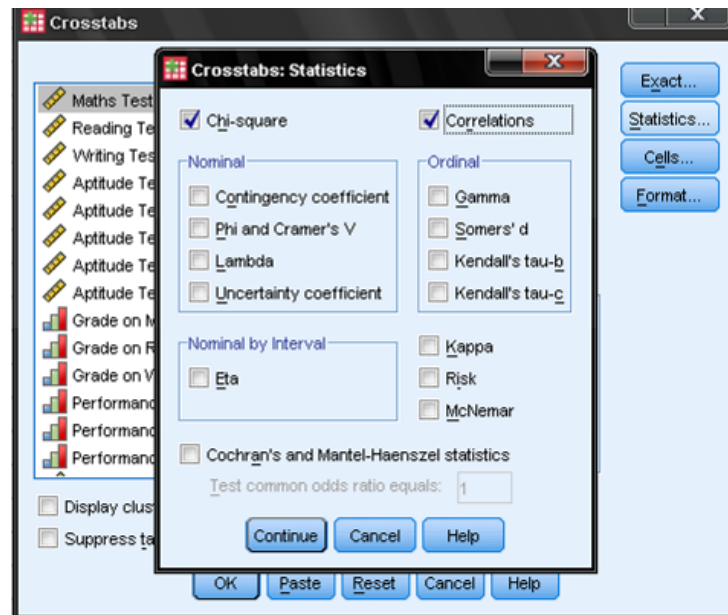
The first step is to add the variables to rows and columns by simply clicking on the variable name in the left list and adding it with a click on the arrow to either the row list or the column list.



The button *Exact...* opens the dialog for the Exact Tests. Exact tests are needed with small cell sizes below ten respondents per cell. SPSS has the choice between Monte-Carlo simulation and Fisher's Exact Test. Since our cells have a population greater or equal than ten we stick to the Asymptotic Test that is Pearson's Chi-Square Test of Independence.



The button *Statistics...* opens the dialog for the additional statics we want SPSS to compute. Since we want to run the Chi-Square Test of Independence we need to tick Chi-Square. We also want to include the contingency coefficient and the correlations which are the tests of interdependence between our two variables.



The next step is to click on the button *Cells...* This brings up the dialog to specify the information each cell should contain. Per default, only the Observed Counts are selected; this would create a simple contingency table of our sample. However the output of the test, the directionality of the correlation, and the dependence between the variables are interpreted with greater ease when we look at the differences between observed and expected counts and percentages.



## The Output of the Chi-Square Test of Independence

The output is quite straightforward and includes only four tables. The first table shows the sample description.

**Case Processing Summary**

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Gender Gender * Exam Exam	107	100.0%	0	.0%	107	100.0%

The second table in our output is the contingency table for the Chi-Square Test of Independence. We find that there seems to be a gender difference between those who fail and those who pass the exam. We find that more male students failed the exam than were expected (22 vs. 19.1) and more female students passed the exam than were expected (33 vs. 30.1).

**Gender Gender \* Exam Exam Crosstabulation**

			Exam Exam		Total
			.00 Fail	1.00 Pass	
Gender Gender	0 Male	Count	22	10	32
		Expected Count	19.1	12.9	32.0
	1 Female	Count	42	33	75
		Expected Count	44.9	30.1	75.0
Total		Count	64	43	107
		Expected Count	64.0	43.0	107.0

This is a first indication that our hypothesis should be supported—our hypothesis being that gender has an influence on whether the student passed the exam. The results of the Chi-Square Test of Independence are in the SPSS output table Chi-Square Tests:

**Chi-Square Tests**

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	1.517 <sup>a</sup>	1	.218	.283	.155
Continuity Correction <sup>b</sup>	1.033	1	.309		
Likelihood Ratio	1.546	1	.214		
Fisher's Exact Test					
Linear-by-Linear Association	1.503	1	.220		
N of Valid Cases	107				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 12.86.

b. Computed only for a 2x2 table

Alongside the Pearson Chi-Square, SPSS automatically computes several other values, Yates' continuity correction for 2x2 tables being one of them. Yates introduced this value to correct for small degrees of freedom. However, Yates' continuity correction has little practical relevance since SPSS calculates Fisher's Exact Test as well. Moreover the rule of thumb is that for large samples sizes ( $n > 50$ ) the continuity correction can be omitted.

Secondly, the Likelihood Ratio, or G-Test, is based on the maximum likelihood theory and for large samples it calculates a Chi-Square similar to the Pearson Chi-Square. G-Test Chi-Squares can be added to allow for more complicated test designs.

Thirdly, Fisher's Exact Test, which we discussed earlier, should be used for small samples with a cell size below ten as well as for very large samples. The Linear-by-Linear Association, which calculates the association between two linear variables, can only be used if both variables have an ordinal or continuous-level scale.

The first row shows the results of Chi-Square Test of Independence: the  $X^2$  value is 1.517 with 1 degree of freedom, which results in a p-value of .218. Since 0.218 is larger than 0.05 we cannot reject the null hypothesis that the two variables are independent, thus we cannot say that gender has an influence on passing the exam.

The last table in the output shows us the contingency coefficient, which is the result of the test of interdependence for two nominal variables. It is similar to the correlation coefficient  $r$  and in this case 0.118 with a significance of 0.218. Again the contingency coefficient's test of significance is larger than the critical value 0.05, and therefore we cannot reject the null hypothesis that the contingency coefficient is significantly different from zero.

Symmetric Measures					
		Value	Asymp. Std. Error <sup>a</sup>	Approx. T <sup>b</sup>	Approx. Sig.
Nominal by Nominal	Contingency Coefficient	.118			.218
Interval by Interval	Pearson's R	.119	.093	1.229	.222 <sup>c</sup>
Ordinal by Ordinal	Spearman Correlation	.119	.093	1.229	.222 <sup>c</sup>
N of Valid Cases		107			

One possible interpretation and write-up of this analysis is as follows:

*The initial hypothesis was that gender and outcome of the final exam are not independent. The contingency table shows that more male students than expected failed the exam (22 vs. 19.1) and more female students than expected passed the exam (33 vs. 30.1). However a Chi-Square test does not confirm the initial hypothesis. With a Chi-Square of 1.517 (d.f. = 1) the test can not reject the null hypothesis ( $p = 0.218$ ) that both variables are independent.*



## ***Bivariate (Pearson) Correlation***

### ***What is a Bivariate (Pearson's) Correlation?***

*Correlation* is a widely used term in statistics. In fact, it entered the English language in 1561, 200 years before most of the modern statistic tests were discovered. It is derived from the [same] Latin word *correlation*, which means relation. Correlation generally describes the effect that two or more phenomena occur together and therefore they are linked. Many academic questions and theories investigate these relationships. Is the time and intensity of exposure to sunlight related the likelihood of getting skin cancer? Are people more likely to repeat a visit to a museum the more satisfied they are? Do older people earn more money? Are wages linked to inflation? Do higher oil prices increase the cost of shipping? It is very important, however, to stress that correlation does not imply causation.

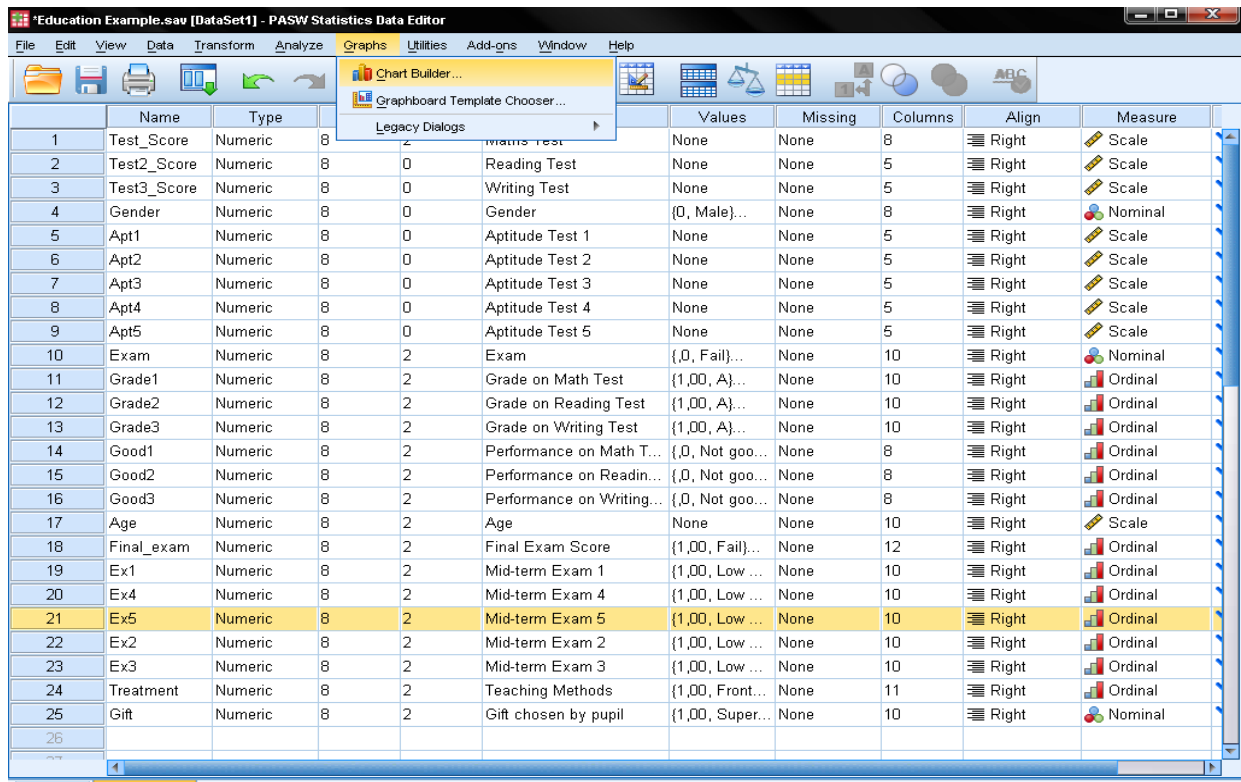
A correlation expresses the strength of linkage or co-occurrence between two variables in a single value between -1 and +1. This value that measures the strength of linkage is called *correlation coefficient*, which is represented typically as the letter  $r$ .

The correlation coefficient between two continuous-level variables is also called Pearson's  $r$  or Pearson product-moment correlation coefficient. A positive  $r$  value expresses a positive relationship between the two variables (the larger A, the larger B) while a negative  $r$  value indicates a negative relationship (the larger A, the smaller B). A correlation coefficient of zero indicates no relationship between the variables at all. However correlations are limited to linear relationships between variables. Even if the correlation coefficient is zero, a non-linear relationship might exist.

## Bivariate (Pearson's) Correlation in SPSS

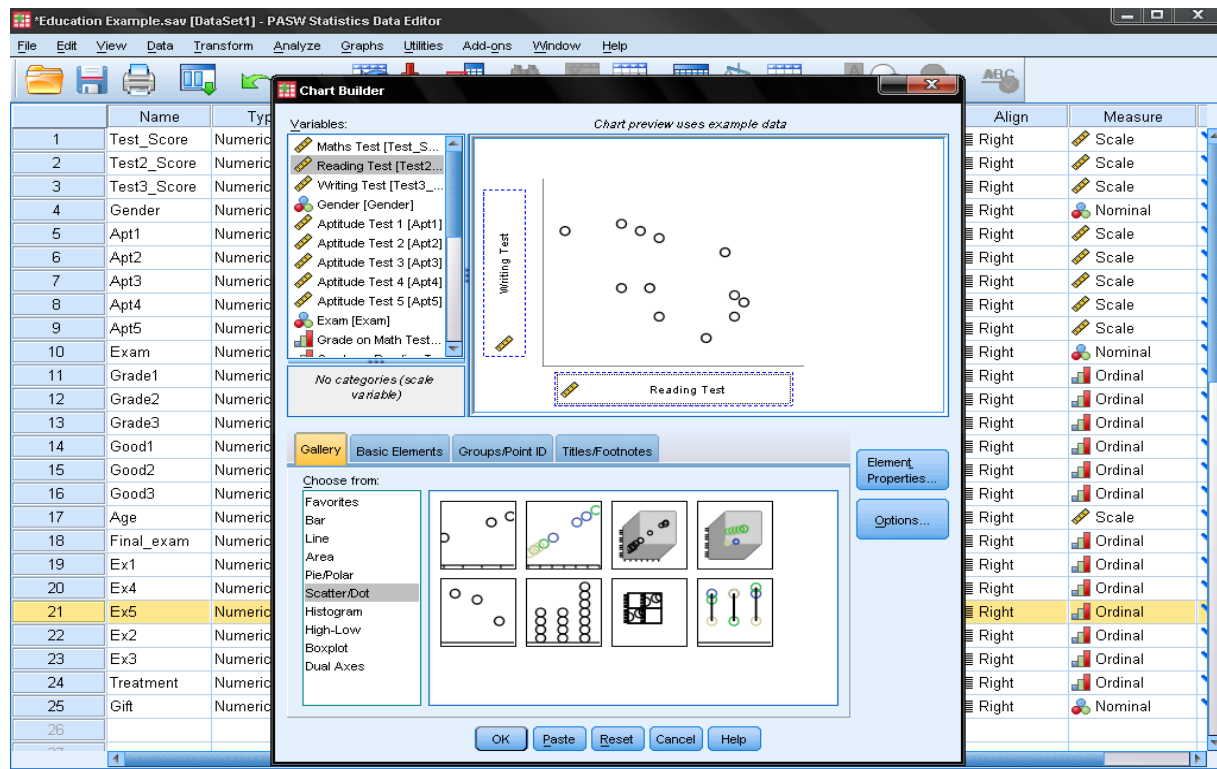
At this point it would be beneficial to create a scatter plot to visualize the relationship between our two test scores in reading and writing. The purpose of the scatter plot is to verify that the variables have a linear relationship. Other forms of relationship (circle, square) will not be detected when running Pearson's Correlation Analysis. This would create a type II error because it would not reject the null hypothesis of the test of independence ('the two variables are independent and not correlated in the universe') although the variables are in reality dependent, just not linearly.

The scatter plot can either be found in *Graphs/Chart Builder...* or in *Graphs/Legacy Dialog/Scatter Dot...*

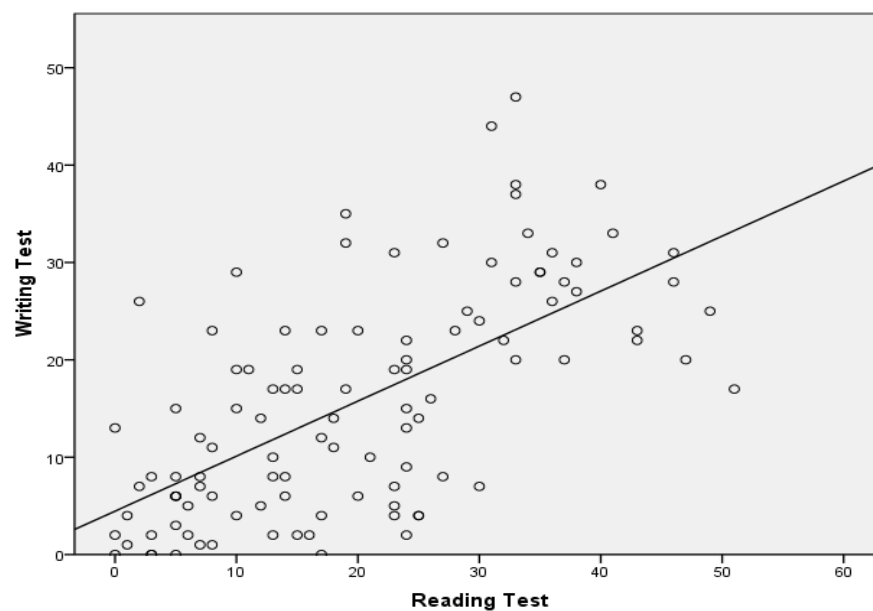


	Name	Type	Measure
1	Test_Score	Numeric	Scale
2	Test2_Score	Numeric	Scale
3	Test3_Score	Numeric	Scale
4	Gender	Numeric	Nominal
5	Apt1	Numeric	Scale
6	Apt2	Numeric	Scale
7	Apt3	Numeric	Scale
8	Apt4	Numeric	Scale
9	Apt5	Numeric	Scale
10	Exam	Numeric	Nominal
11	Grade1	Numeric	Ordinal
12	Grade2	Numeric	Ordinal
13	Grade3	Numeric	Ordinal
14	Good1	Numeric	Ordinal
15	Good2	Numeric	Ordinal
16	Good3	Numeric	Ordinal
17	Age	Numeric	Scale
18	Final_exam	Numeric	Ordinal
19	Ex1	Numeric	Ordinal
20	Ex4	Numeric	Ordinal
21	Ex5	Numeric	Ordinal
22	Ex2	Numeric	Ordinal
23	Ex3	Numeric	Ordinal
24	Treatment	Numeric	Ordinal
25	Gift	Numeric	Nominal
26			

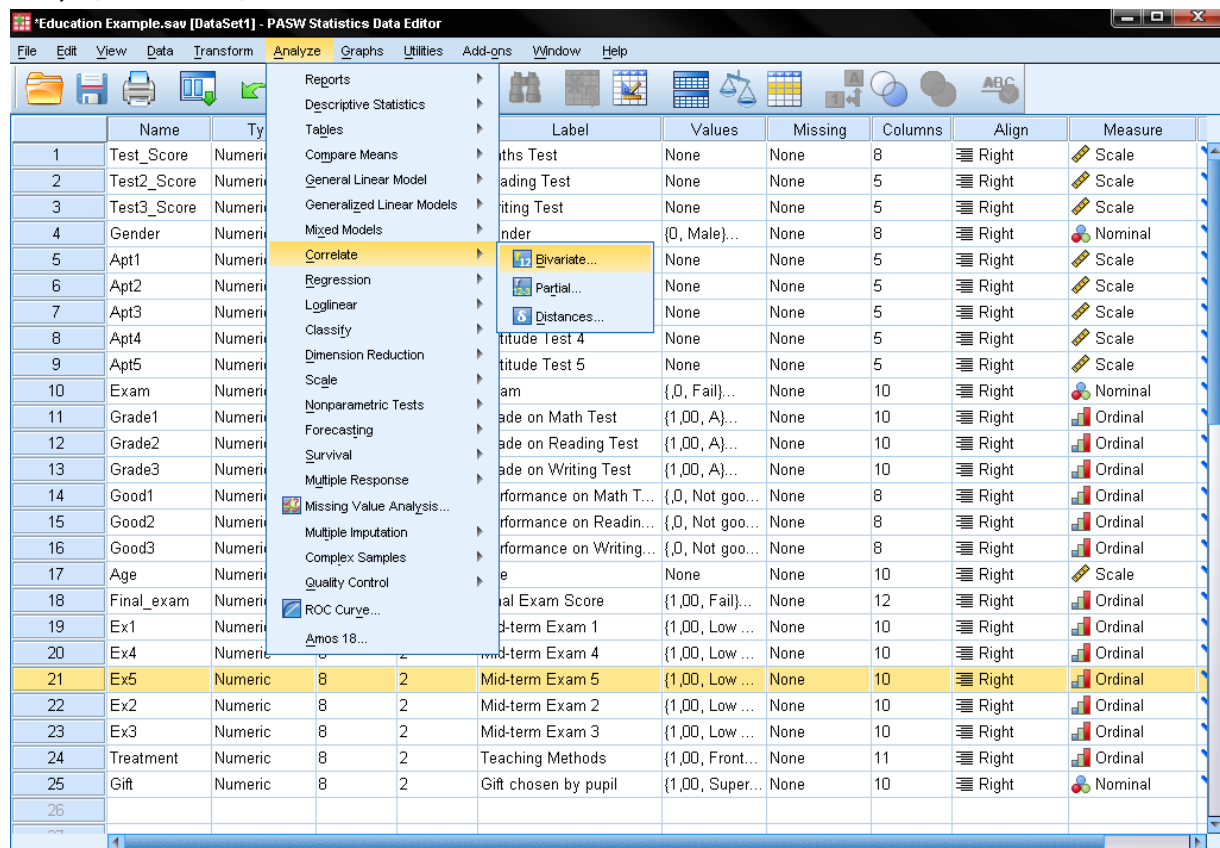
In the *Chart Builder* we simply choose in the Gallery tab the *Scatter/Dot* group of charts and drag the '*Simple Scatter*' diagram (the first one) on the chart canvas. Next we drag variable *Test\_Score* on the y-axis and variable *Test2\_Score* on the x-axis.



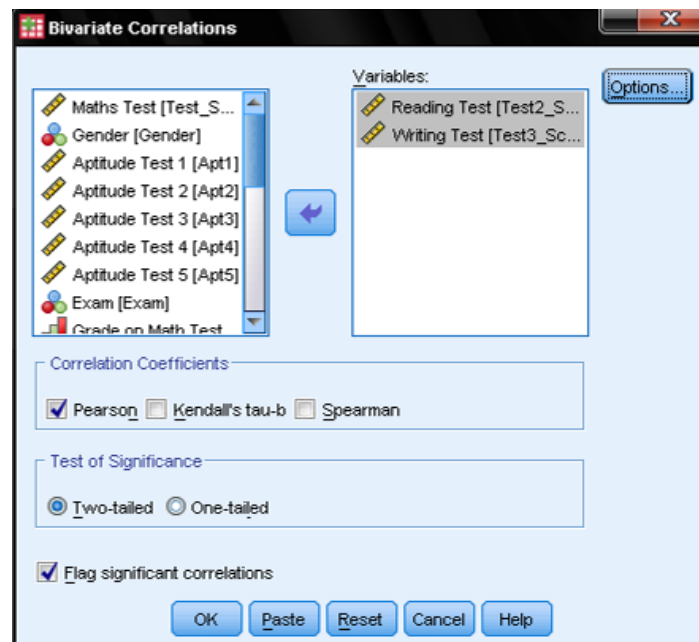
SPSS generates the scatter plot for the two variables. A double click on the output diagram opens the chart editor and a click on '*Add Fit Line*' adds a linearly fitted line that represents the linear association that is represented by Pearson's bivariate correlation.



To calculate Pearson's bivariate correlation coefficient in SPSS we have to open the dialog in *Analyze/Correlation/Bivariate...*



This opens the dialog box for all bivariate correlations (Pearson's, Kendall's, Spearman). Simply select the variables you want to calculate the bivariate correlation for and add them with the arrow.



Select the bivariate correlation coefficient you need, in this case Pearson's. For the Test of Significance we select the two-tailed test of significance, because we do not have an assumption whether it is a positive or negative correlation between the two variables *Reading* and *Writing*. We also leave the default tick mark at flag significant correlations which will add a little asterisk to all correlation coefficients with  $p < 0.05$  in the SPSS output.

### **The Output of the Bivariate (Pearson's) Correlation**

The output is fairly simple and contains only a single table - the correlation matrix. The bivariate correlation analysis computes the Pearson's correlation coefficient of a pair of two variables. If the analysis is conducted for more than two variables it creates a larger matrix accordingly. The matrix is symmetrical since the correlation between A and B is the same as between B and A. Also the correlation between A and A is always 1.

Correlations		Test2_Score Reading Test	Test3_Score Writing Test
Test2_Score Reading Test	Pearson Correlation	1	.645**
	Sig. (2-tailed)		.000
	N	107	107
Test3_Score Writing Test	Pearson Correlation	.645**	1
	Sig. (2-tailed)	.000	
	N	107	107

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

In this example Pearson's correlation coefficient is .645, which signifies a medium positive linear correlation. The significance test has the null hypothesis that there is no positive or negative correlation between the two variables in the universe ( $r = 0$ ). The results show a very high statistical significance of  $p < 0.001$  thus we can reject the null hypothesis and assume that the Reading and Writing test scores are positively, linearly associated in the general universe.

One possible interpretation and write-up of this analysis is as follows:

*The initial hypothesis predicted a linear relationship between the test results scored on the Reading and Writing tests that were administered to a sample of 107 students. The scatter diagrams indicate a linear relationship between the two test scores. Pearson's bivariate correlation coefficient shows a medium positive linear relationship between both test scores ( $r = .645$ ) that is significantly different from zero ( $p < 0.001$ ).*

## Partial Correlation

### What is Partial Correlation?

Spurious correlations have a ubiquitous effect in statistics. Spurious correlations occur when two effects have clearly no causal relationship whatsoever in real life but can be statistically linked by correlation.

A classic example of a spurious correlation is as follows:

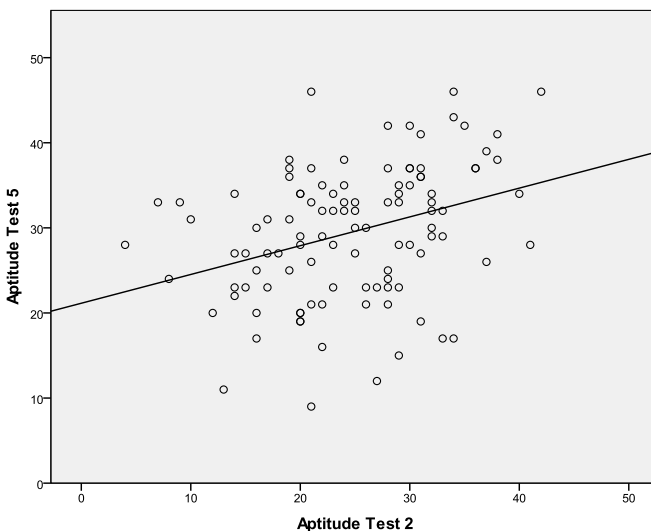
*Do storks bring Babies? Pearson's Bivariate Correlation Coefficient shows a positive and significant relationship between the number of births and the number of storks in a sample of 52 US counties.*

Spurious correlations are caused by not observing a third variable that influences the two analyzed variables. This third, unobserved variable is also called the confounding factor, hidden factor, suppressor, mediating variable, or control variable. Partial Correlation is the method to correct for the overlap of the moderating variable.

In the stork example, one cofounding factor is the size of the county – larger counties tend to have larger populations of women and storks and—as a clever replication of this study in the Netherlands showed—the cofounding factor is the weather nine months before the date of observation. Partial correlation is the statistical test to identify and correct spurious correlations.

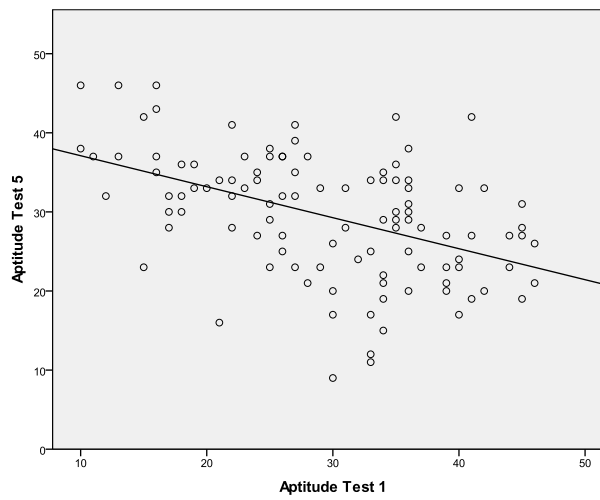
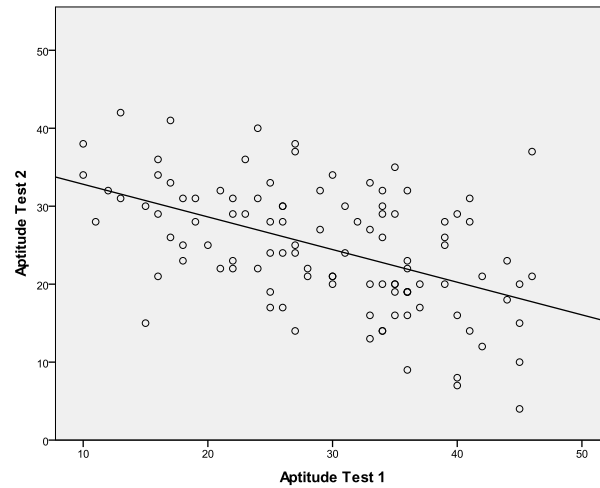
### How to run the Partial Correlation in SPSS

In our education example, we find that the test scores of the second and the fifth aptitude tests positively correlate. However we have the suspicion that this is only a spurious correlation that is caused by individual differences in the baseline of the student. We measured the baseline aptitude with the first aptitude test.



To find out more about our correlations and to check the linearity of the relationship, we create scatter plots. SPSS creates scatter plots with the menu *Graphs/Chart Builder...* and then we select *Scatter/Dot* from the Gallery list. Simply drag 'Aptitude Test 2' onto the y-axis and 'Aptitude Test 5' on the x-Axis. SPSS creates the scatter plots, which clearly shows a linear positive association between the two variables.

We can also create the scatter plots for the interaction effects of our suspected control variable Aptitude Test 1. The scatter plots show a medium, negative, linear correlation between our baseline test and the two tests in our analysis.



Partial Correlations are found in SPSS under *Analyze/Correlate/Partial...*

Education Example.sav [DataSet1] - PASW Statistics Data Editor

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

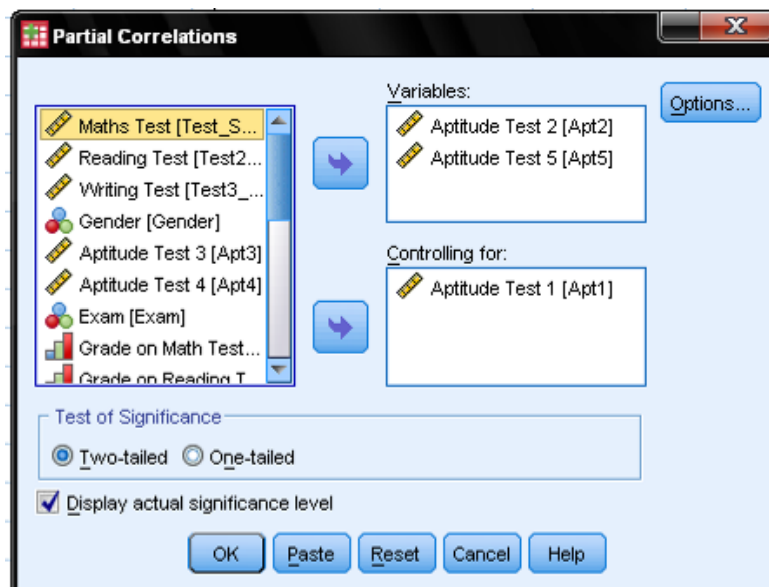
Reports  
Descriptive Statistics  
Tables  
Compare Means  
General Linear Model  
Generalized Linear Models  
Mixed Models  
**Correlate**  
Regression  
Loglinear  
Classify  
Dimension Reduction  
Scale  
Nonparametric Tests  
Forecasting  
Survival  
Multiple Response  
Missing Value Analysis...  
Multiple Imputation  
Complex Samples  
Quality Control  
ROC Curve...  
Amos 18...

Bivariate...  
**Partial...**  
Distances...

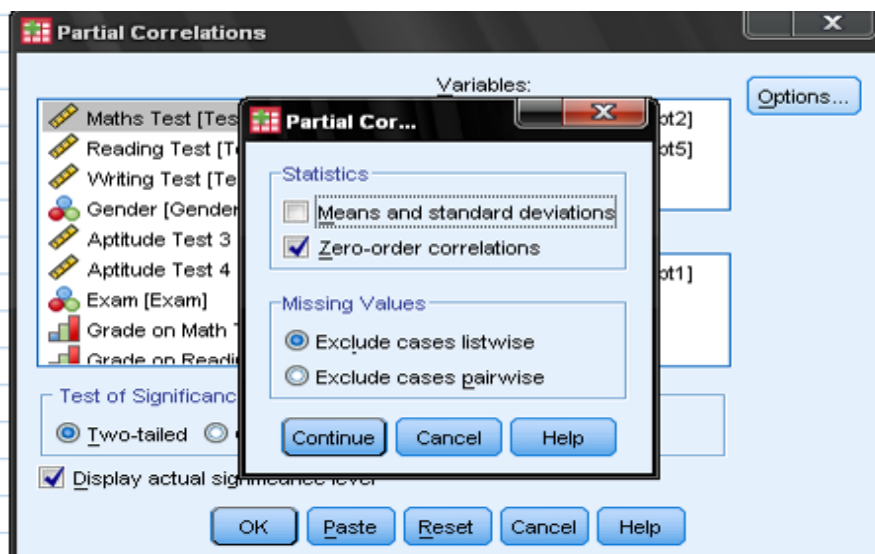
	Name	Type	Label	Values	Missing	Columns	Align	Measure	Role
1	Test_Score	Numeric	Maths Test	None	None	8	Right	Scale	Input
2	Test2_Score	Numeric	Reading Test	None	None	5	Right	Scale	Input
3	Test3_Score	Numeric	Writing Test	None	None	5	Right	Scale	Input
4	Gender	Numeric	Gender	{0, Male}...	None	8	Right	Nominal	Input
5	Apt1	Numeric	Attitude Test 4	None	None	5	Right	Scale	Input
6	Apt2	Numeric	Attitude Test 5	None	None	5	Right	Scale	Input
7	Apt3	Numeric	Exam	{0, Fail}...	None	10	Right	Nominal	Input
8	Apt4	Numeric	Grade on Math ...	{1,00, A}...	None	10	Right	Ordinal	Input
9	Apt5	Numeric	Grade on Readin...	{1,00, A}...	None	10	Right	Ordinal	Input
10	Exam	Numeric	Grade on Writin...	{1,00, A}...	None	10	Right	Ordinal	Input
11	Grade1	Numeric	Performance on...	{0, Not goo...	None	8	Right	Ordinal	Input
12	Grade2	Numeric	Performance on...	{0, Not goo...	None	8	Right	Ordinal	Input
13	Grade3	Numeric	Performance on...	{0, Not goo...	None	8	Right	Ordinal	Input
14	Good1	Numeric	Final Exam Sc...	{1,00, Fail}...	None	12	Right	Ordinal	Input
15	Good2	Numeric	Mid-term Exam 1	{1,00, Low ...	None	10	Right	Ordinal	Input
16	Good3	Numeric	Mid-term Exam 4	{1,00, Low ...	None	10	Right	Ordinal	Input
17	Age	Numeric	Mid-term Exam 5	{1,00, Low ...	None	10	Right	Ordinal	Input
18	Final_exam	Numeric	Mid-term Exam 2	{1,00, Low ...	None	10	Right	Ordinal	Input
19	Ex1	Numeric	Mid-term Exam 3	{1,00, Low ...	None	10	Right	Ordinal	Input
20	Ex4	Numeric	Teaching Meth...	{1,00, Front...	None	11	Right	Ordinal	Input
21	Ex5	Numeric	Gift chosen by ...	{1,00, Super...	None	10	Right	Nominal	Input
22	Ex2	Numeric							
23	Ex3	Numeric							
24	Treatment	Numeric							
25	Gift	Numeric							
26									



This opens the dialog of the Partial Correlation Analysis. First, we select the variables for which we want to calculate the partial correlation. In our example, these are *Aptitude Test 2* and *Aptitude Test 5*. We want to control the partial correlation for *Aptitude Test 1*, which we add in the list of control variables.



The dialog *Options...* allows to display additional descriptive statistics (mean and standard deviations) and the zero-order correlations. If you haven't done a correlation analysis already, check the zero-order correlations, as this will include Pearson's Bivariate Correlation Coefficients for all variables in the output. Furthermore we can manage how missing values will be handled.



### The Output of Partial Correlation Analysis

The output of the Partial Correlation Analysis is quite straightforward and only consists of a single table. The first half displays the zero-order correlation coefficients, which are the three Person's Correlation Coefficients without any control variable taken into account.

Correlations			Apt2 Aptitude Test 2	Apt5 Aptitude Test 5	Apt1 Aptitude Test 1
-none- <sup>a</sup>	Apt2 Aptitude Test 2	Correlation	1.000	.339	-.499
		Significance (2-tailed)	.	.000	.000
		df	0	105	105
	Apt5 Aptitude Test 5	Correlation	.339	1.000	-.468
		Significance (2-tailed)	.000	.	.000
		df	105	0	105
	Apt1 Aptitude Test 1	Correlation	-.499	-.468	1.000
		Significance (2-tailed)	.000	.000	.
		df	105	105	0
Apt1 Aptitude Test 1	Apt2 Aptitude Test 2	Correlation	1.000	.138	
		Significance (2-tailed)	.	.159	
		df	0	104	
	Apt5 Aptitude Test 5	Correlation	.138	1.000	
		Significance (2-tailed)	.159	.	
		df	104	0	

a. Cells contain zero-order (Pearson) correlations.

The zero-order correlations seem to support our hypothesis that a higher test score on aptitude test 2 increases the test score of aptitude test 5. Both a weak association of  $r = 0.339$ , which is highly significant  $p < 0.001$ . However, the variable aptitude test 1 also significantly correlates with both test scores ( $r = -0.499$  and  $r = -0.468$ ). The second part of the table shows the Partial Correlation Coefficient between the Aptitude Test 2 and Aptitude Test 5 when controlled for the baseline test Aptitude 1. The Partial Correlation Coefficient is now  $r_{xy.z} = -0.138$  and not significant  $p = 0.159$ .

One possible interpretation and write-up of this analysis is as follows:

*The observed bivariate correlation between the Aptitude Test Score 2 and the score of Aptitude Test 5 is almost completely explained by the correlation of both variables with the baseline Aptitude Test 1. The partial correlation between both variables is very weak ( $r_{xy.z} = -0.138$ ) and not significant with  $p = 0.159$ . Therefore we cannot reject the null hypothesis that both variables are independent.*

## Spearman Rank Correlation

### What is Spearman Correlation?

Spearman Correlation Coefficient is also referred to as Spearman Rank Correlation or Spearman's rho. It is typically denoted either with the Greek letter  $\rho$  ( $\rho$ ), or  $r_s$ . It is one of the few cases where a Greek letter denotes a value of a sample and not the characteristic of the general population. Like all correlation coefficients, Spearman's rho measures the strength of association of two variables. As such, the Spearman Correlation Coefficient is a close sibling to Pearson's Bivariate Correlation Coefficient, Point-Biserial Correlation, and the Canonical Correlation.

All correlation analyses express the strength of linkage or co-occurrence between two variables in a single value between -1 and +1. This value is called the *correlation coefficient*. A positive correlation coefficient indicates a positive relationship between the two variables (the larger A, the larger B) while a negative correlation coefficient expresses a negative relationship (the larger A, the smaller B). A correlation coefficient of 0 indicates that no relationship between the variables exists at all. However, correlations are limited to linear relationships between variables. Even if the correlation coefficient is zero, a non-linear relationship might exist.

Compared to Pearson's bivariate correlation coefficient, the Spearman Correlation does not require continuous-level data (interval or ratio), because it uses ranks instead of assumptions about the distributions of the two variables. This allows us to analyze the association between variables of ordinal measurement levels. Moreover, the Spearman Correlation is a non-parametric test, which does not assume that the variables approximate a multivariate normal distribution. Spearman Correlation Analysis can therefore be used in many cases where the assumptions of Pearson's Bivariate Correlation (continuous-level variables, linearity, heteroscedasticity, and multivariate normal distribution of the variables to test for significance) are not met.

Typical questions the Spearman Correlation Analysis answers are as follows:

- **Sociology:** Do people with a higher level of education have a stronger opinion of whether or not tax reforms are needed?
- **Medicine:** Does the number of symptoms a patient has indicate a higher severity of illness?
- **Biology:** Is mating choice influenced by body size in bird species A?
- **Business:** Are consumers more satisfied with products that are higher ranked in quality?

Theoretically, the Spearman correlation calculates the Pearson correlation for variables that are converted to ranks. Similar to Pearson's bivariate correlation, the Spearman correlation also tests the null hypothesis of independence between two variables. However, this can lead to difficult interpretations. Kendall's Tau-b rank correlation improves this by reflecting the strength of the dependence between the variables in comparison.

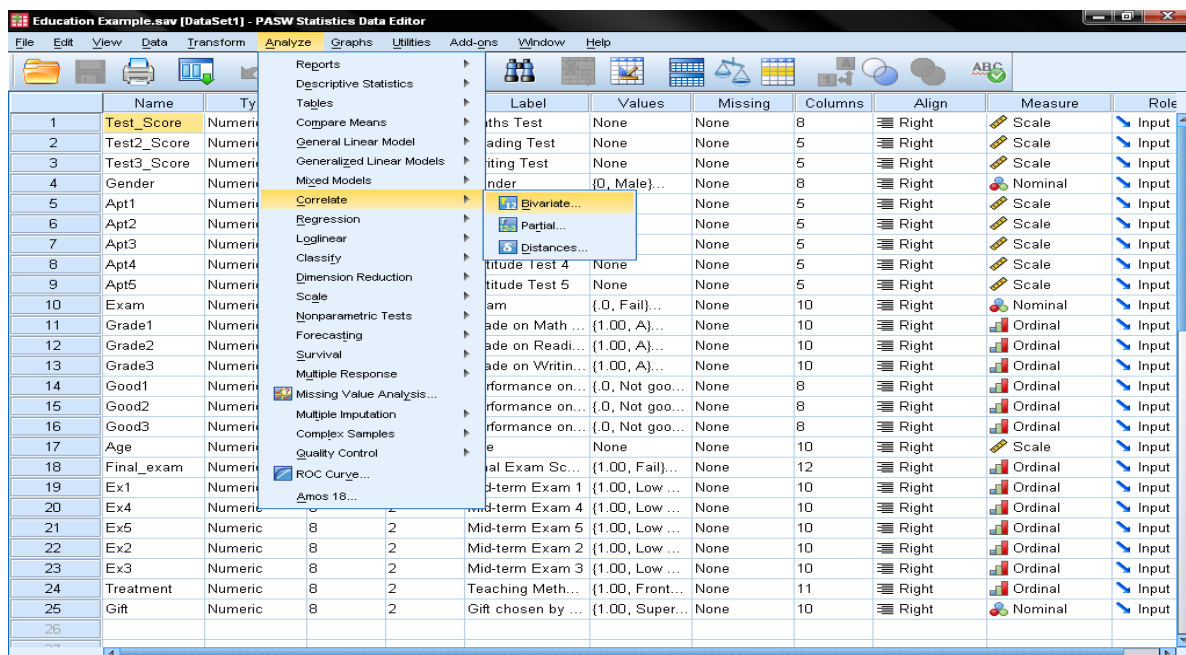
Since both variables need to be of ordinal scale or ranked data, Spearman's correlation requires converting interval or ratio scales into ranks before it can be calculated. Mathematically, Spearman correlation and Pearson correlation are very similar in the way that they use difference measurements to calculate the strength of association. Pearson correlation uses standard deviations while Spearman correlation difference in ranks. However, this leads to an issue with the Spearman correlation when tied ranks exist in the sample. An example of this is when a sample of marathon results awards two silver medals but no bronze medal. A statistician is even crueler to these runners because a rank is defined as average position in the ascending order of values. For a statistician, the marathon result would have one first place, two places with a rank of 2.5, and the next runner ranks 4. If tied ranks occur, a more complicated formula has to be used to calculate  $\rho$ , but SPSS automatically and correctly calculates tied ranks.

### Spearman Correlation in SPSS

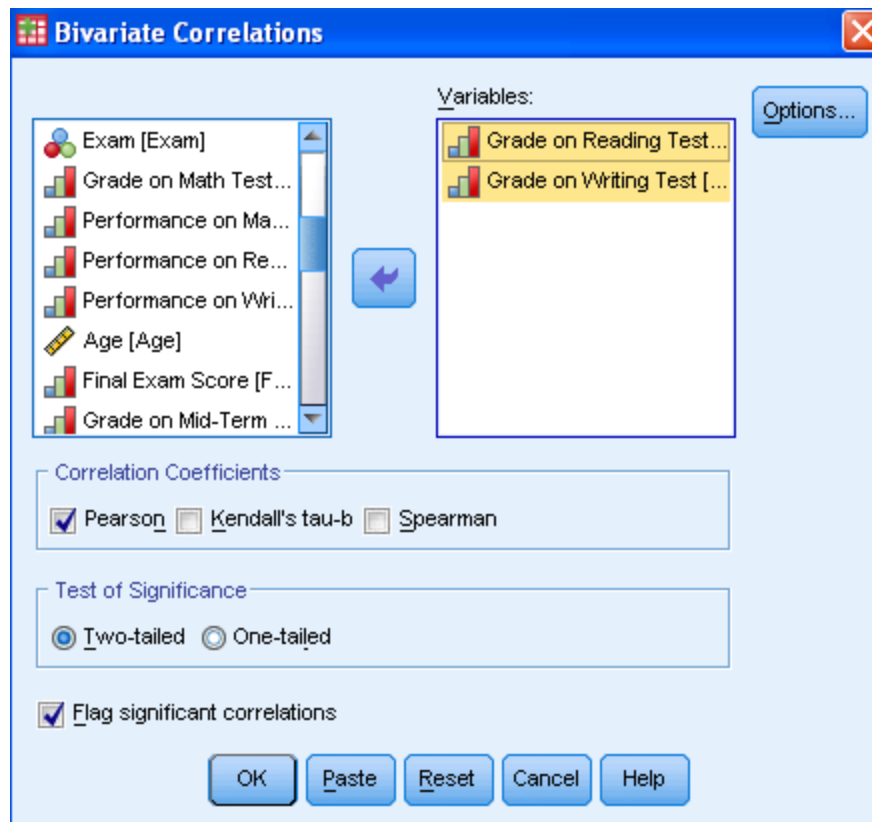
We have shown in the Pearson's Bivariate Correlation Analysis that the Reading Test Scores and the Writing test scores are positively correlated. Let us assume that we never did this analysis; the research question posed is then, *"Are the grades of the reading and writing test correlated?"* We assume that all we have to test this hypothesis are the grades achieved (A-F). We could also include interval or ratio data in this analysis because SPSS converts scale data automatically into ranks.

The Spearman Correlation requires ordinal or ranked data, therefore it is very important that measurement levels are correctly defined in SPSS. Grade 2 and Grade 3 are ranked data and therefore measured on an ordinal scale. If the measurement levels are specified correctly, SPSS will automatically convert continuous-level data into ordinal data. Should we have raw data that already represents rankings but without specification that this is an ordinal scale, nothing bad will happen.

Spearman Correlation can be found in SPSS in *Analyze/Correlate/Bivariate...*



This opens the dialog for all Bivariate Correlations, which also includes Pearson's Bivariate Correlation. Using the arrow, we add *Grade 2* and *Grade 3* to the list of variables for analysis. Then we need to tick the correlation coefficients we want to calculate. In this case the ones we want are Spearman and Kendall's Tau-b.



### ***The Output of Spearman's Correlation Analysis***

The output of Spearman's Correlation Analysis is fairly straightforward and consists of only one table.

### Correlations

			Grade2 Grade on Reading Test	Grade3 Grade on Writing Test
Kendall's tau_b	Grade2 Grade on Reading Test	Correlation Coefficient	1.000	.507**
		Sig. (2-tailed)	.	.000
		N	107	107
	Grade3 Grade on Writing Test	Correlation Coefficient	.507**	1.000
		Sig. (2-tailed)	.000	.
		N	107	107
Spearman's rho	Grade2 Grade on Reading Test	Correlation Coefficient	1.000	.634**
		Sig. (2-tailed)	.	.000
		N	107	107
	Grade3 Grade on Writing Test	Correlation Coefficient	.634**	1.000
		Sig. (2-tailed)	.000	.
		N	107	107

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

The first coefficient in the output table is Kendall's Tau-b. Kendall's Tau is a simpler correlation coefficient that calculates how many concordant pairs (same rank on both variables) exist in a sample. Tau-b measures the strength of association when both variables are measured at the ordinal level. It adjusts the sample for tied ranks.

The second coefficient is Spearman's rho. SPSS shows that for example the Bivariate Correlation Coefficient between Grade 2 and Grade 3 is  $\tau = 0.507$  and  $\rho = 0.634$ . In both cases the significance is  $p < 0.001$ . For small samples, SPSS automatically calculates a permutation test of significance instead of the classical t-test, which violates the assumption of multivariate normality when sample size is small.

One possible interpretation and write-up of Spearman's Correlation Coefficient rho and the test of significance is as follows:

*We analyzed the question of whether the grade achieved in the reading test and the grade achieved in the writing test are somewhat linked. Spearman's Correlation Coefficient indicates a strong association between these two variables ( $\rho = 0.634$ ). The test of significance indicates that with  $p < 0.001$  we can reject the null hypothesis that both variables are independent in the general population. Thus we can say that with a confidence of more than 95% the observed positive correlation between grade writing and grade reading is not caused by random effects and both variables are interdependent.*

Please always bear in mind that correlation alone does not make for causality.

### Point-Biserial Correlation

#### What is Point-Biserial Correlation?

Like all correlation analyses the Point-Biserial Correlation measures the strength of association or co-occurrence between two variables. Correlation analyses express this strength of association in a single value, the correlation coefficient.

The Point-Biserial Correlation Coefficient is a correlation measure of the strength of association between a continuous-level variable (ratio or interval data) and a binary variable. Binary variables are variables of nominal scale with only two values. They are also called dichotomous variables or dummy variables in Regression Analysis. Binary variables are commonly used to express the existence of a certain characteristic (e.g., reacted or did not react in a chemistry sample) or the membership in a group of observed specimen (e.g., male or female). If needed for the analysis, binary variables can also be created artificially by grouping cases or recoding variables. However it is not advised to artificially create a binary variable from ordinal or continuous-level (ratio or scale) data because ordinal and continuous-level data contain more variance information than nominal data and thus make any correlation analysis more reliable. For ordinal data use the Spearman Correlation Coefficient  $\rho$ , for continuous-level (ratio or scale) data use Pearson's Bivariate Correlation Coefficient  $r$ . Binary variables are also called dummy. The Point-Biserial Correlation Coefficient is typically denoted as  $r_{pb}$ .

Like all Correlation Coefficients (e.g. Pearson's  $r$ , Spearman's  $\rho$ ), the Point-Biserial Correlation Coefficient measures the strength of association of two variables in a single measure ranging from -1 to +1, where -1 indicates a perfect negative association, +1 indicates a perfect positive association and 0 indicates no association at all. All correlation coefficients are interdependency measures that do not express a causal relationship.

Mathematically, the Point-Biserial Correlation Coefficient is calculated just as the Pearson's Bivariate Correlation Coefficient would be calculated, wherein the dichotomous variable of the two variables is either 0 or 1—which is why it is also called the binary variable. Since we use the same mathematical concept, we do need to fulfill the same assumptions, which are normal distribution of the continuous variable and homoscedasticity.

Typical questions to be answered with a Point-Biserial Correlation Analysis are as follows:

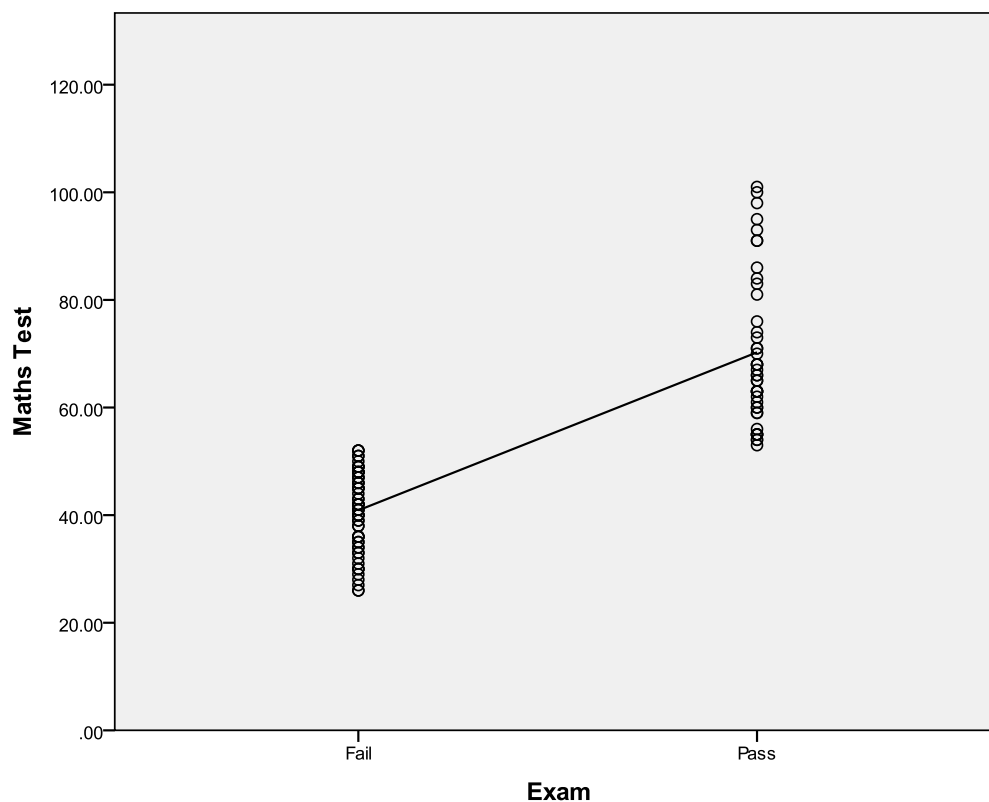
- **Biology** – Do fish react differently to red or green lighted stimulus as food signal? Is there an association between the color of the stimulus (red or green light) and the reaction time?
- **Medicine** – Does a cancer drug prolong life? How strong is the association between administering the drug (placebo, drug) and the length of survival after treatment?
- **Sociology** – Does gender have an influence on salary? Is there an association between gender (female, male) with the income earned?
- **Social psychology** – Is satisfaction with life higher the older you are? Is there an association between age group (elderly, not elderly) and satisfaction with life?
- **Economics** – Does analphabetism indicate a weaker economy? How strong is the association between literacy (literate vs. illiterate societies) and GDP?

Since all correlation analyses require the variables to be randomly independent, the Point-Biserial Correlation is not the best choice for analyzing data collected in experiments. For these cases a Linear Regression Analysis with Dummy Variables is the best choice. Also, many of the questions typically answered with a Point-Biserial Correlation Analysis can be answered with an independent sample t-Test or other dependency tests (e.g., Mann-Whitney-U, Kruskal-Wallis-H, and Chi-Square). Not only are some of these tests robust regarding the requirement of normally distributed variables, but also these tests analyze dependency or causal relationship between an independent variable and dependent variables in question.

### ***Point-Biserial Correlation Analysis in SPSS***

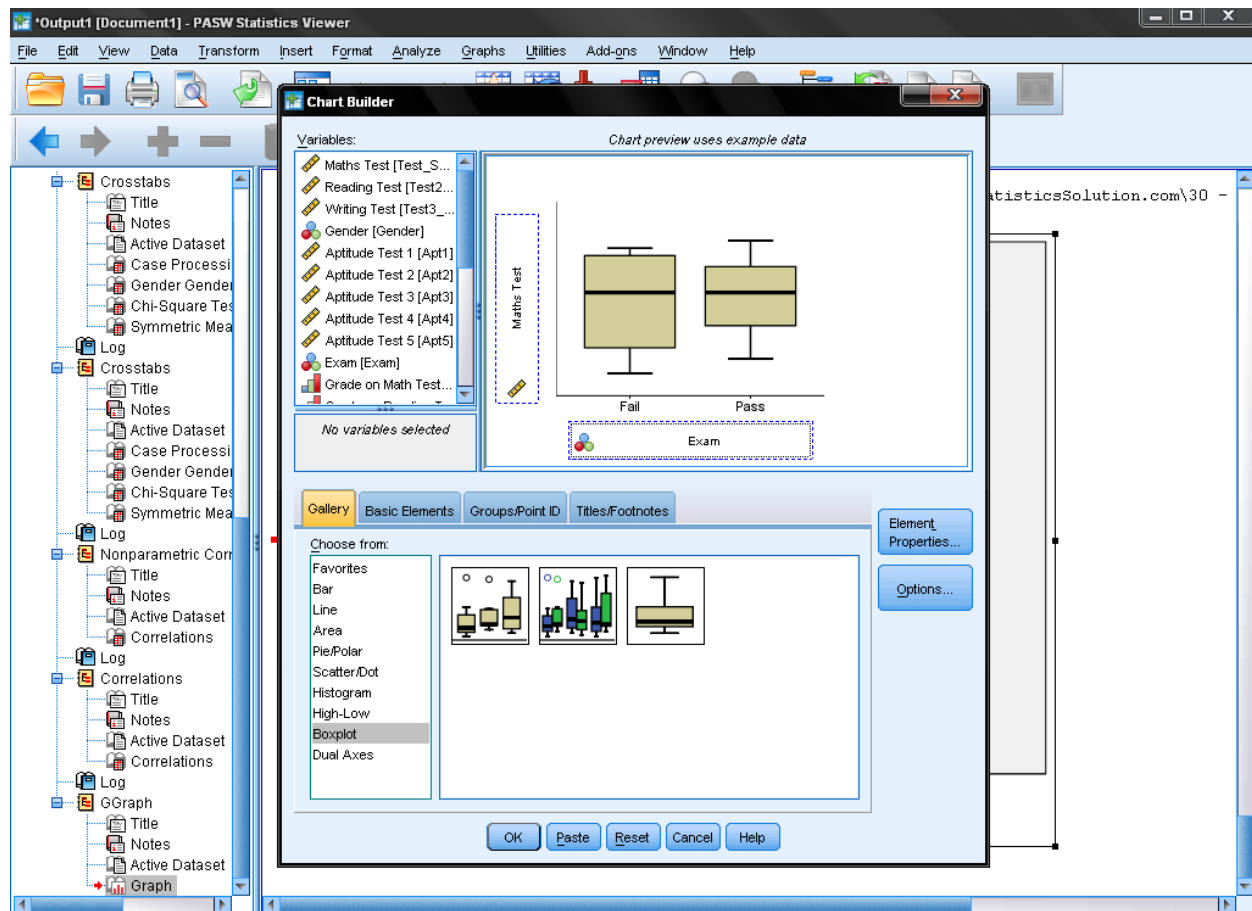
Referring back to our initial example, we are interested in the strength of association between passing or failing the exam (variable exam) and the score achieved in the math, reading, and writing tests. In order to work correctly we need to correctly define the level of measurement for the variables in the variable view. There is no special command in SPSS to calculate the Point-Biserial Correlation Coefficient; SPSS needs to be told to calculate Pearson's Bivariate Correlation Coefficient  $r$  with our data.

Since we use the Pearson  $r$  as Point-Biserial Correlation Coefficient, we should first test whether there is a relationship between both variables. As described in the section on Pearson's Bivariate Correlation in SPSS, the first step is to draw the scatter diagram of both variables. For the Point-Biserial Correlation Coefficient this diagram would look like this.

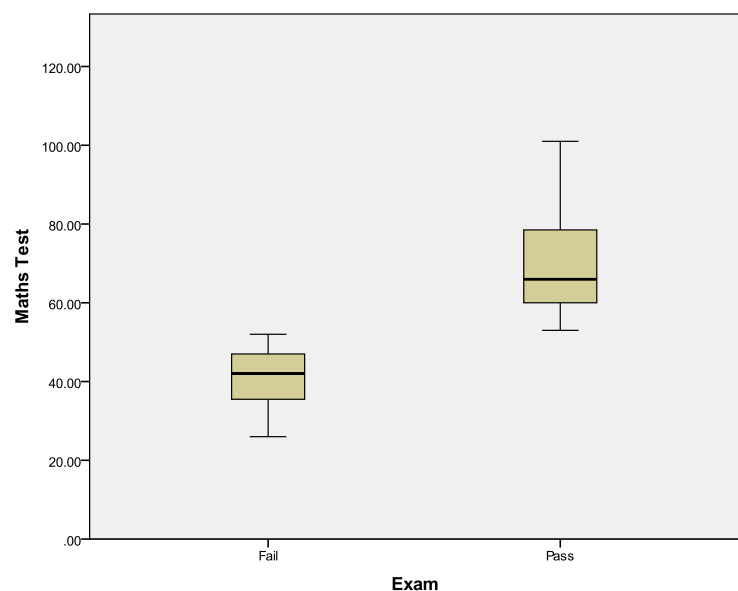




The diagram shows a positive slope and indicates a positive relationship between the math score and passing the final exam or failing it. Since our variable exam is measured on nominal level (0, 1), a better way to display the data is to draw a box plot. To create a box plot we select *Graphs/Chart Builder...* and select the *Simple Box plot* from the List in the *Gallery*. Drag *Exam* on the x-axis and *Math Test* on the y-axis.

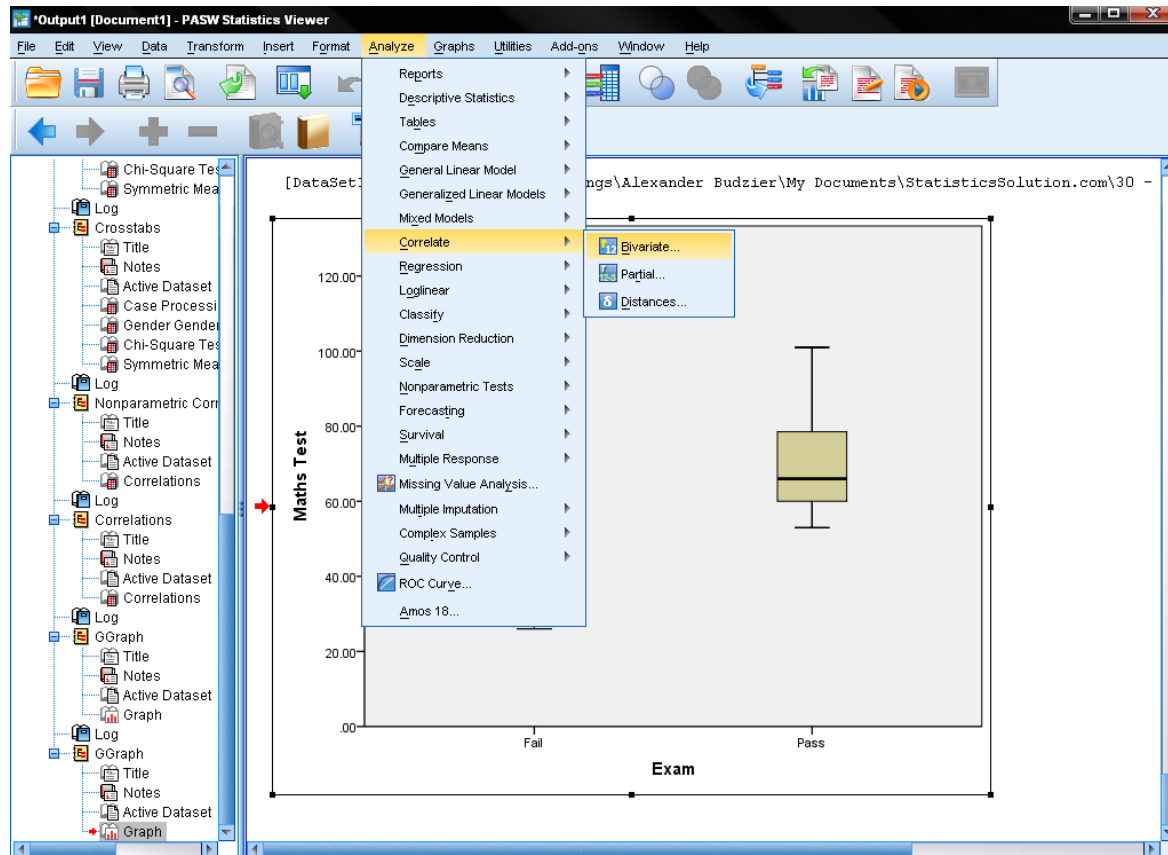


A box plot displays the distribution information of a variable. More specifically it displays the quartiles of the data. The whiskers of the box span from the 0% quartile to the 100% quartile of the data. If the sample contains outliers they are displayed as data points outside the whiskers. The box spans from the 25% quartile to the 75% quartile and the median (the 50% quartile) is displayed as a strong line inside the box.



In our example we can see in the box plot that not only are the math scores higher on average for students who passed the final exam but also that there is almost no overlapping between the two groups. Now that we have an understanding of the direction of our association between the two variables we can conduct the Point-Biserial Correlation Analysis.

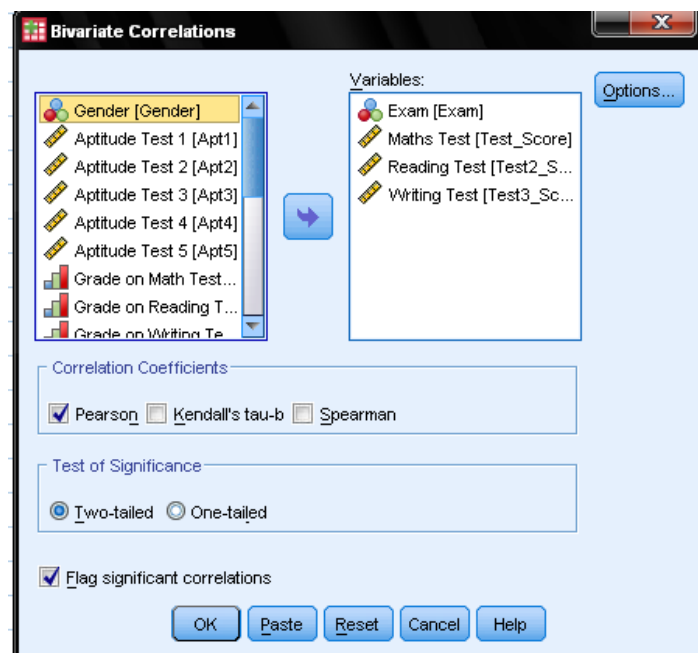
SPSS does not have a special procedure for the Point-Biserial Correlation Analysis. If a Point-Biserial Correlation is to be calculated in SPSS, the procedure for Pearson's  $r$  has to be used. Therefore we open the Bivariate Correlations dialog *Analyze/Correlate/Bivariate...*



In the dialog box we add both variables to the list of variables to analyze and select *Pearson* Correlation Coefficient and *two-tailed* Test of Significance.

### ***The output of the Point-Biserial Correlation Analysis***

SPSS calculates the output we are already familiar with. Again, it just contains one table. Please note that SPSS assumes for the calculation of the Point-Biserial Correlation that our



nominal variable is of measured on the continuous-level (ratio or interval data).

**Correlations**

		Exam Exam	Test_Score Math Test	Test2_Score Reading Test	Test3_Score Writing Test
Exam Exam	Pearson Correlation	1	.810**	.545**	.673**
	Sig. (2-tailed)		.000	.000	.000
	N	107	107	107	107
Test_Score Math Test	Pearson Correlation	.810**	1	.582**	.747**
	Sig. (2-tailed)	.000		.000	.000
	N	107	107	107	107
Test2_Score Reading Test	Pearson Correlation	.545**	.582**	1	.645**
	Sig. (2-tailed)	.000	.000		.000
	N	107	107	107	107
Test3_Score Writing Test	Pearson Correlation	.673**	.747**	.645**	1
	Sig. (2-tailed)	.000	.000	.000	
	N	107	107	107	107

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

The table shows that the correlation coefficient for math is  $r = 0.810$ , for reading is  $r = 0.545$  and for writing is  $r = 0.673$ . This indicates a strong association between the outcome of the exam and the previous test scores in reading, writing, and math. The two-tailed test of independence is significant with  $p < 0.001$ . Therefore we can reject our null hypotheses that each variable is independent in the universe with a confidence level greater than 95%.

A write-up could read as follows:

*We analyzed the relationship between passing or failing the final exam and the previously achieved scores in math, reading, and writing tests. The point-biserial correlation analysis finds that the variables are strongly and positively associated ( $r = 0.810, 0.545$ , and  $0.673$ ). That statistical test of significance confirms that the correlation we found in our sample can be generalized onto the population our sample was drawn from ( $p < 0.001$ ). Thus we might say that a higher test score increases the probability of passing the exam and vice versa.*

## Canonical Correlation

### What is Canonical Correlation analysis?

The Canonical Correlation is a multivariate analysis of correlation. As we discussed before Pearson's  $r$ , Spearman's  $\rho$ , Chi-Square, Point-Biserial are all bivariate correlation analysis, which measure the strength of association between two and only two variables.

Canonical is the statistical term for analyzing latent variables (which are not directly observed) that represent multiple variables (which are directly observed). The term can also be found in canonical regression analysis and in multivariate discriminant analysis.

Canonical Correlation analysis is the analysis of multiple-X multiple-Y correlation. The Canonical Correlation Coefficient measures the strength of association between two Canonical Variates.

A Canonical Variate is the weighted sum of the variables in the analysis. The canonical variate is denoted CV. Similarly to the discussions on why to use factor analysis instead of creating unweighted indices as independent variables in regression analysis, canonical correlation analysis is preferable in analyzing the strength of association between two constructs. This is such because it creates an internal structure, for example, a different importance of single item scores that make up the overall score (as found in satisfaction measurements and aptitude testing).

For multiple x and y the canonical correlation analysis constructs two variates  $CV_{x1} = a_1x_1 + a_2x_2 + a_3x_3 + \dots + a_nx_n$  and  $CV_{y1} = b_1y_1 + b_2y_2 + b_3y_3 + \dots + b_my_m$ . The canonical weights  $a_1 \dots a_n$  and  $b_1 \dots b_m$  are chosen so that they maximize the correlation between the canonical variates  $CV_{x1}$  and  $CV_{y1}$ . A pair of canonical variates is called a *canonical root*. This step is repeated for the *residuals* to generate additional duplets of canonical variates until the cut-off value =  $\min(n,m)$  is reached; for example, if we calculate the canonical correlation between three variables for test scores and five variables for aptitude testing, we would extract three pairs of canonical variates or three canonical roots. Note that this is a major difference from factor analysis. In factor analysis the factors are calculated to maximize between-group variance while minimizing in-group variance. They are factors because they group the underlying variables.

Canonical Variants are not factors because only the first pair of canonical variants groups the variables in such way that the correlation between them is maximized. The second pair is constructed out of the residuals of the first pair in order to maximize correlation between them. Therefore the canonical variants cannot be interpreted in the same way as factors in factor analysis. Also the calculated canonical variates are automatically orthogonal, i.e., they are independent from each other.

Similar to factor analysis, the central results of canonical correlation analysis are the canonical correlations, the canonical factor loadings, and the canonical weights. They can also be used to calculate  $d$ , the measure of redundancy. The redundancy measurement is important in questionnaire design and scale development. It can answer questions such as, *"When I measure a five item satisfaction with the last purchase and a three item satisfaction with the after sales support, can I exclude one of the two scales for the sake of shortening my questionnaire?"* Statistically it represents the proportion of variance of one set of variables explained by the variant of the other set of variables.

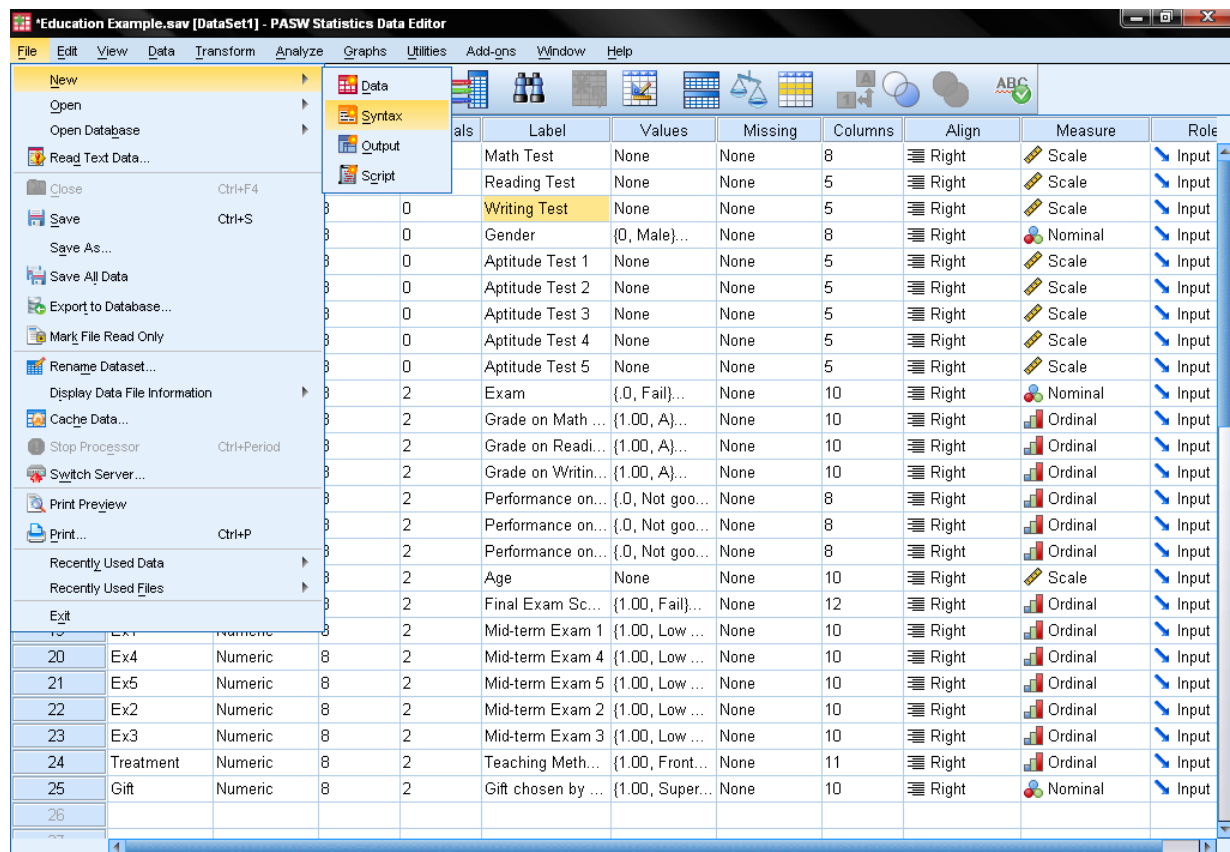
The canonical correlation coefficients test for the existence of overall relationships between two sets of variables, and redundancy measures the magnitude of relationships. Lastly Wilk's lambda (also called U value) and Bartlett's V are used as a Test of Significance of the canonical correlation coefficient.

Typically Wilk's lambda is used to test the significance of the first canonical correlation coefficient and Bartlett's V is used to test the significance of all canonical correlation coefficients.

A final remark: Please note that the Discriminant Analysis is a special case of the canonical correlation analysis. Every nominal variable with  $n$  different factor steps can be replaced by  $n-1$  dichotomous variables. The Discriminant Analysis is then nothing but a canonical correlation analysis of a set of binary variables with a set of continuous-level(ratio or interval) variables.

### Canonical Correlation Analysis in SPSS

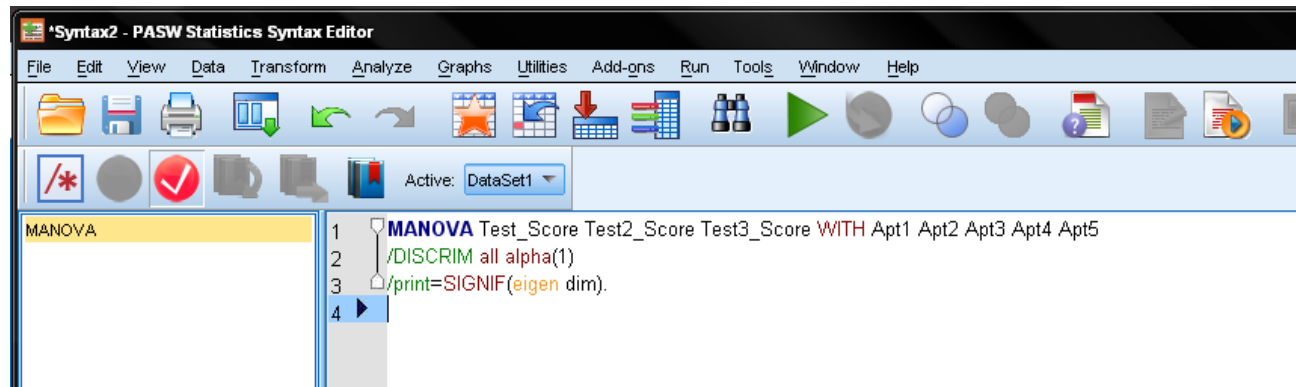
We want to show the strength of association between the five aptitude tests and the three tests on math, reading, and writing. Unfortunately, SPSS does not have a menu for canonical correlation analysis. So we need to run a couple of syntax commands. Do not worry—this sounds more complicated than it really is. First, we need to open the syntax window. Click on *File/New/Syntax*.



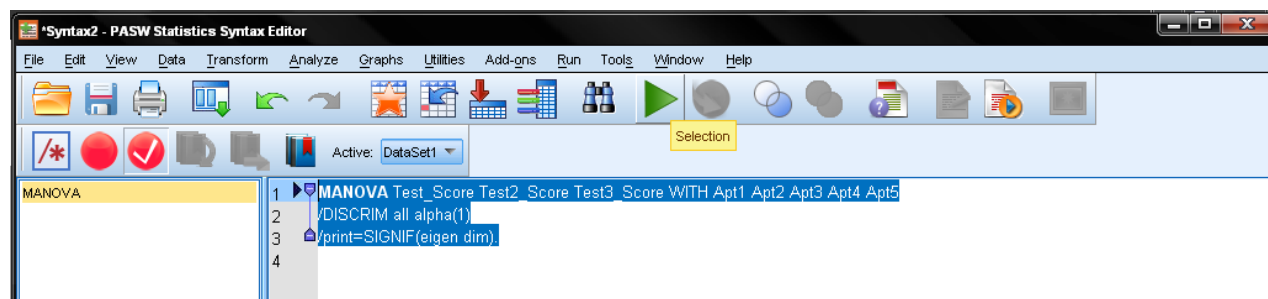
In the SPSS syntax we need to use the command for MANOVA and the subcommand */discrim* in a one factorial design. We need to include all independent variables in one single factor separating the two groups by the *WITH* command. The list of variables in the MANOVA command contains the dependent variables first, followed by the independent variables (Please do not use the command *BY* instead of *WITH* because that would cause the factors to be separated as in a MANOVA analysis).

The subcommand */discrim* produces a canonical correlation analysis for all covariates. Covariates are specified after the keyword *WITH*. *ALPHA* specifies the significance level required before a canonical

variable is extracted, default is 0.25; it is typically set to 1.0 so that all discriminant function are reported. Your syntax should look like this:



To execute the syntax, just highlight the code you just wrote and click on the big green *Play* button.



### ***The Output of the Canonical Correlation Analysis***

The syntax creates an overwhelmingly large output (see to the right). No worries, we discuss the important bits of it next. The output starts with a sample description and then shows the general fit of the model reporting Pillai's, Helling's, Wilk's and Roy's multivariate criteria. The commonly used test is Wilks's lambda, but we find that all of these tests are significant with  $p < 0.05$ .

```

.....
***** Analysis of Variance-- Design 1 *****
EFFECT .. WITHIN CELLS Regression
Multivariate Tests of Significance (S = 3, M = 1/2, N = 48 1/2)

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

Pillais        .71752      6.35012      15.00      303.00      .000
Hotellings      1.98475     12.92290     15.00      293.00      .000
Wilks          .32195      9.26289     15.00      273.70      .000
Roys           .65784
.....

```

The next section reports the canonical correlation coefficients and the eigenvalues of the canonical roots. The first canonical correlation coefficient is .81108 with an explained variance of the correlation

of 96.87% and an eigenvalue of 1.92265. Thus indicating that our hypothesis is correct – generally the standardized test scores and the aptitude test scores are positively correlated.

---

Eigenvalues and Canonical Correlations

Root No.	Eigenvalue	Pct.	Cum. Pct.	Canon Cor.	Sq. Cor
1	1.92265	96.87114	96.87114	.81108	.65784
2	.04840	2.43863	99.30977	.21486	.04617
3	.01370	.69023	100.00000	.11625	.01351

---

So far the output only showed overall model fit. The next part tests the significance of each of the roots. We find that of the three possible roots only the first root is significant with  $p < 0.05$ . Since our model contains the three test scores (math, reading, writing) and five aptitude tests, SPSS extracts three canonical roots or dimensions. The first test of significance tests all three canonical roots of significance ( $f = 9.26$   $p < 0.05$ ), the second test excludes the first root and tests roots two to three, the last test tests root three by itself. In our example only the first root is significant  $p < 0.05$ .

---

Dimension Reduction Analysis

Roots	Wilks L.	F	Hypoth. DF	Error DF	Sig. of F
1 TO 3	.32195	9.26289	15.00	273.70	.000
2 TO 3	.94094	.77260	8.00	200.00	.627
3 TO 3	.98649	.46121	3.00	101.00	.710

---

In the next parts of the output SPSS presents the results separately for each of the two sets of variables. Within each set, SPSS gives the raw canonical coefficients, standardized coefficients, correlations between observed variables and the canonical variant, and the percent of variance explained by the canonical variant. Below are the results for the 3 Test variables.

---

EFFECT .. WITHIN CELLS Regression (Cont.)  
Univariate F-tests with (5,101) D. F.

Variable	Sq. Mul. R	Adj. R-sq.	Hypoth. MS	Error MS	F	Sig. of F
Test_Sco	.23789	.20016	1607.60519	254.95838	6.30536	.000
Test2_Sc	.64316	.62549	2333.49702	64.09337	36.40777	.000
Test3_Sc	.37145	.34033	1035.16713	86.71635	11.93739	.000

---

The raw canonical coefficients are similar to the coefficients in linear regression; they can be used to calculate the canonical scores.

-----  
Raw canonical coefficients for DEPENDENT variables  
Function No.

Variable	1	2	3
Test_Sco	.00792	.02522	.08179
Test2_Sc	-.06959	-.07412	.00839
Test3_Sc	-.02168	.08361	-.11318

-----  
Easier to interpret are the standardized coefficients (mean = 0, st.dev. = 1). Only the first root is relevant since root two and three are not significant. The strongest influence on the first root is variable Test\_Score (which represents the math score).

-----  
Standardized canonical coefficients for DEPENDENT variables  
Function No.

Variable	1	2	3
Test_Sco	.14136	.45026	1.46034
Test2_Sc	-.91034	-.96965	.10970
Test3_Sc	-.24862	.95865	-1.29763

-----  
Correlations between DEPENDENT and canonical variables  
Function No.

Variable	1	2	3
Test_Sco	-.57434	.60197	.55476
Test2_Sc	-.98833	-.08947	.12326
Test3_Sc	-.72991	.66990	-.13589

-----  
Variance in dependent variables explained by canonical variables

CAN. VAR.	Pct Var DEP	Cum Pct DEP	Pct Var COV	Cum Pct COV
1	61.31476	61.31476	40.33556	40.33556
2	27.30463	88.61939	1.26055	41.59611
3	11.38061	100.00000	.15380	41.74991

-----  
The next section shows the same information (raw canonical coefficients, standardized coefficients, correlations between observed variables and the canonical variant, and the percent of variance explained by the canonical variant) for the aptitude test variables.



-----  
Raw canonical coefficients for COVARIATES  
Function No.

COVARIATE	1	2	3
Apt1	-.08024	.09133	-.06759
Apt2	.04519	.13585	.00730
Apt3	.01328	-.00559	-.05456
Apt4	.01620	-.01427	-.14681
Apt5	-.00863	.01925	-.06113

-----  
Standardized canonical coefficients for COVARIATES  
CAN. VAR.

COVARIATE	1	2	3
Apt1	-.75188	.85575	-.63335
Apt2	.35541	1.06836	.05743
Apt3	.09324	-.03924	-.38297
Apt4	.09886	-.08712	-.89597
Apt5	-.06769	.15099	-.47964

-----  
Correlations between COVARIATES and canonical variables  
CAN. VAR.

Covariate	1	2	3
Apt1	-.94346	.28824	-.06654
Apt2	.73453	.67589	.04334
Apt3	.21641	.05831	-.20545
Apt4	.39052	-.23244	-.75456
Apt5	.43147	.08796	-.41828

-----  
Variance in covariates explained by canonical variables

CAN. VAR.	Pct Var DEP	Cum Pct DEP	Pct Var COV	Cum Pct COV
1	23.88175	23.88175	36.30304	36.30304
2	.55868	24.44043	12.10159	48.40463
3	.21429	24.65472	15.85668	64.26130

-----

Again, in the table standardized coefficients, we find the importance of the variables on the canonical roots. The first canonical root is dominated by Aptitude Test 1.

The next part of the table shows the multiple regression analysis of each dependent variable (Aptitude Test 1 to 5) on the set of independent variables (math, reading, writing score).

-----  
 Regression analysis for WITHIN CELLS error term  
 --- Individual Univariate .9500 confidence intervals  
 Dependent variable .. Test\_Score      Math Test

COVARIATE	B	Beta	Std. Err.	t-Value	Sig. of t	Lower -95%	CL- Upper
Apt1	.8004598228	.4200904618	.21579	3.70937	.000	.37238	1.22854
Apt2	-.0537487173	-.0236747929	.23321	-.23047	.818	-.51638	.40889
Apt3	-.1861961490	-.0732065613	.22709	-.81993	.414	-.63668	.26428
Apt4	-.3367266635	-.1151005754	.28083	-1.19906	.233	-.89381	.22036
Apt5	.0458101909	.0201306043	.22954	.19957	.842	-.40954	.50116
Dependent variable .. Test2_Score      Reading Test							

COVARIATE	B	Beta	Std. Err.	t-Value	Sig. of t	Lower -95%	CL- Upper
Apt1	.8058566784	.5771889596	.10820	7.44812	.000	.59123	1.02049
Apt2	-.5067324452	-.3046167084	.11693	-4.33363	.000	-.73869	-.27477
Apt3	-.1481092808	-.0794727991	.11386	-1.30082	.196	-.37397	.07775
Apt4	-.1937971905	-.0904074850	.14080	-1.37638	.172	-.47311	.08552
Apt5	.0741805551	.0444878802	.11509	.64455	.521	-.15413	.30249
Dependent variable .. Test3_Score      Writing Test							

COVARIATE	B	Beta	Std. Err.	t-Value	Sig. of t	Lower -95%	CL- Upper
Apt1	.7076293351	.5783021719	.12585	5.62278	.000	.45798	.95728
Apt2	-.0838896280	-.0575402926	.13601	-.61679	.539	-.35370	.18592
Apt3	-.0895001530	-.0547959932	.13244	-.67580	.501	-.35222	.17322
Apt4	-.1069142521	-.0569090456	.16378	-.65280	.515	-.43180	.21798
Apt5	.1013971121	.0693850664	.13387	.75744	.451	-.16416	.36696

-----

The next section with the analysis of constant effects can be ignored as it is not relevant for the canonical correlation analysis. One possible write-up could be:

*The initial hypothesis is that scores in the standardized tests and the aptitude tests are correlated. To test this hypothesis we conducted a canonical correlation analysis. The analysis included three variables with the standardized test scores (math, reading, and writing) and five variables with the aptitude test scores. Thus we extracted three canonical roots. The overall model is significant (Wilk's Lambda = .32195, with  $p < 0.001$ ), however the individual tests of significance show that only the first canonical root is significant on  $p < 0.05$ . The first root explains a large proportion of the variance of the correlation (96.87%, eigenvalue 1.92265). Thus we find that the canonical correlation coefficient between the first roots is 0.81108 and we can assume that the standardized test scores and the aptitude test scores are positively correlated in the population.*

### ***Independent Variable T-Test***

#### ***What is the independent variable t-test?***

The independent variable t-test is a member of the t-test family. All tests in the t-test family compare differences in mean scores of continuous-level (interval or ratio), normally distributed data. The t-test family belongs to the bigger group of the bivariate dependency analysis. All t-tests split the analysis in a dependent and independent variable and assume that the independent variable influences the dependent variable in such a way that the influence causes the observed differences of the mean scores.

However the independent variable t-test is somewhat different. The independent variable t-test compares whether a variable is zero in the population. This can be done because many statistical calculated measurements approximate students' t-distribution when the underlying variables that go into the calculation are multivariate normal distributed.

How is that relevant? The independent variable t-test is most often used in two scenarios: (1) as the test of significance for estimated coefficients and (2) as the test of independence in correlation analyses.

To start, the independent variable t-test is extremely important for statistical analyses that calculate variable weights, for example linear regression analysis, discriminant analysis, canonical correlation analysis, or structural equation modeling. These analyses use a general linear model and an optimization mechanic, for example maximum likelihood estimates, to build a linear model by adding the weighted variables in the analysis. These estimations calculate the coefficients for the variables in the analysis. The independent variable t-test is used to check whether these variable weights exist in the general population from which our sample was drawn or whether these weights are statistical artifacts only found by chance.

Most statistical packages, like SPSS, test the weights of linear models using ANOVA and the t-test, because the ANOVA has a higher statistical power and also because the relationship is quite simple ( $t^2=F$ ). However the ANOVA makes additional assumptions, for example homoscedasticity that the t-test does not need. Most statistical packages used to estimate Structural Equation Models, e.g., AMOS, EQS, LISREL, call the independent variable t-test z-statistics. In some older versions the tests are called T-values. It is also noteworthy that the overall goodness of fit of a structural equation model when based on the covariance matrix uses the chi-square distribution. Individual path weights are based on the t-test or ANOVA, which typically can be specified when defining the model.

Once again, it is important to point out that all t-tests assume multivariate normality of the underlying variables in the sample. The test is robust for large sample sizes and in data sets where the underlying distributions are similar even when they are not multivariate normal. If the normality assumption is violated, the t-values and therefore the levels of significance are too optimistic.

Secondly, the independent variable t-test is used as the test of independence in correlation analyses, e.g., Pearson's bivariate  $r$ , point-biserial  $r$ , canonical correlation analysis, classical test of Spearman's  $\rho$ . These variables calculate a sample test value to measure independence, whose distribution approximates students'  $t$  distribution for large sample sizes. In correlation analysis the independent variable t-test assesses whether the variables are independent in the population or whether the linkage between the variables found in the sample is caused by chance.

The independent variable t-test is virtually identical to the 1-sample t-test. But where the independent variable t-test tests significance or the independence of a derived statistical measurement, the 1-sample t-test tests whether a mean score calculated from a sample equals certain hypothetically assumed value (e.g., zero, or a known population mean).

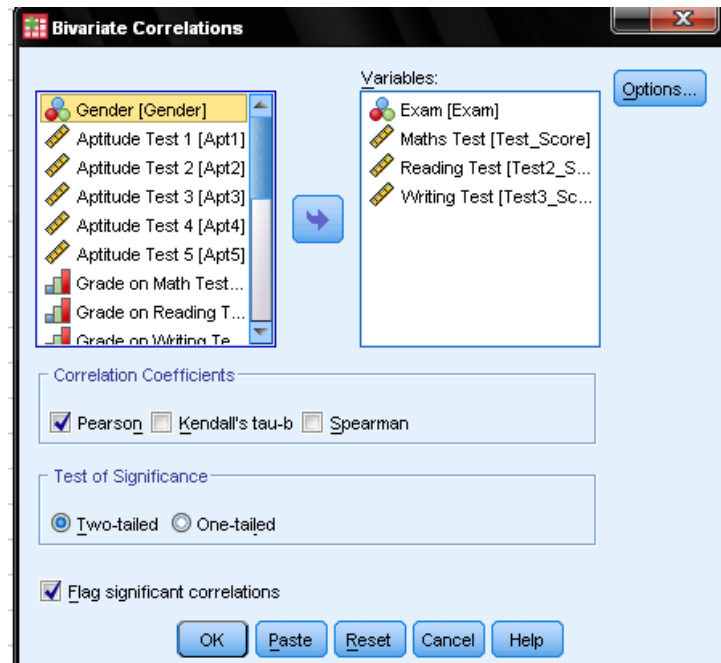
### **The Independent variable t-test in SPSS**

The independent variable t-test examines whether a statistical variable is zero in the universe from which the sample was drawn. This is typically used in correlation analysis and in coefficient estimation. In SPSS the independent variable t-test is always found in the respective menus of the analyses. See the chapters on Bivariate Correlation Analysis in SPSS, Spearman Correlation Analysis in SPSS and Regression Analysis in SPSS for details on how to calculate the analyses. These sections also include descriptions in how to interpret the full output and a sample write-up of the analysis.

### **Correlations**

In the bivariate correlation analyses it is included in the general correlation menu. It is marked per default and cannot be unchecked. The only thing that can be selected is whether a one-tailed or a two-tailed independent variable t-test shall be calculated.

The one-tailed independent variable t-test is the right test if the directionality of the correlation is known beforehand, for example when we know that the true correlation coefficient in the universe is positive. If the directionality is not known the two-tailed independent variable t-test is the right test.



The SPSS Syntax for the independent variable t-test is in the subcommand `/PRINT=TWOTAIL NOSIG`, where TWOTAIL indicates two-tailed t-test, and NOSIG that significant correlations are to be flagged.

### Correlations

		Exam Exam	Test_Score Maths Test	Test2_Score Reading Test	Test3_Score Writing Test
Exam Exam	Pearson Correlation	1	.810**	.545**	.673**
	Sig. (2-tailed)		.000	.000	.000
	N	107	107	107	107
Test_Score Maths Test	Pearson Correlation	.810**	1	.582**	.747**
	Sig. (2-tailed)	.000		.000	.000
	N	107	107	107	107
Test2_Score Reading Test	Pearson Correlation	.545**	.582**	1	.645**
	Sig. (2-tailed)	.000	.000		.000
	N	107	107	107	107
Test3_Score Writing Test	Pearson Correlation	.673**	.747**	.645**	1
	Sig. (2-tailed)	.000	.000	.000	
	N	107	107	107	107

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

The result of the test is included in the correlation matrix in the rows where it says *significance (2-tailed)*. In this example resident population and murder rate are not correlated. The correlation coefficient is relatively small and with a  $p=0.814$  the null hypothesis of the independent variable t-test cannot be rejected. Thus we assume that the coefficient is not different from zero in the universe the sample was drawn from.

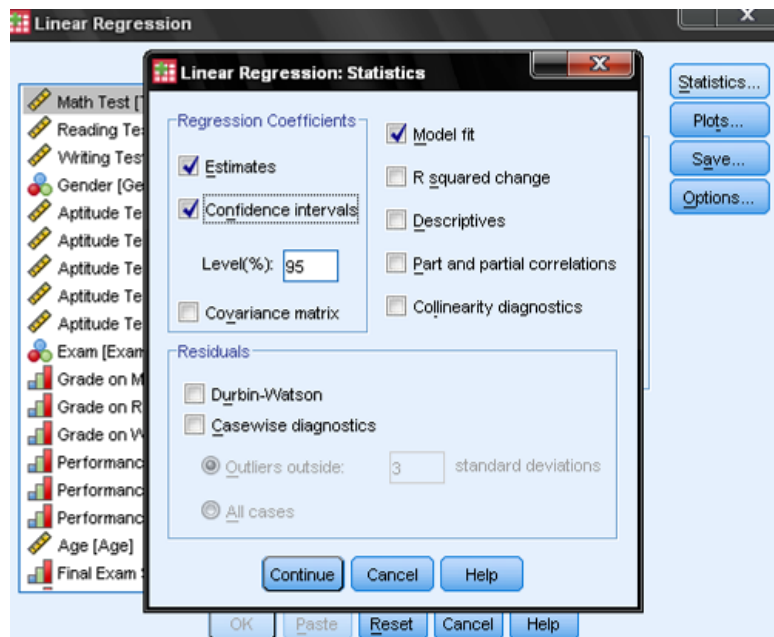
### Regression

In regression analysis the independent variable t-test can be switched on in the statistics dialog box of the linear regression menu

Analyze/ Regression/ Linear...

In the results of the analysis the independent variable t-test shows up in the table coefficients. Remember that for overall model fit the F-Test (ANOVA) is used; it however tests only if all regression coefficients are different from zero.

The SPSS syntax for generating the output related to the t-test of independence is included in the statistics subcommand of **REGRESSION/STATISTICS COEFF OUTS CI(95) R ANOVA**



This will calculate the coefficients table that includes the t-test, otherwise only the F-Test for overall significance of the model will be calculated (null hypothesis: all estimates are not zero). And the table will include the 95% confidence interval for the estimated coefficients and constants.

Coefficients <sup>a</sup>								
Model	Unstandardized Coefficients			Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
	B	Std. Error	Beta				Lower Bound	Upper Bound
1 (Constant)	36.824	2.582		14.263	.000		31.704	41.943
Test2_Score Reading Test	.795	.108	.582	7.337	.000		.580	1.009

a. Dependent Variable: Test\_Score Math Test

In this example the estimated linear regression equation is math test score = 36.824 + 0.795\* reading test score. The independent variable t-test shows that the regression constant b = 36.824 is significantly different from zero, and that the regression coefficient a = 0.795 is significantly different from zero as well. The independent variable t-test can also be used to construct the confidence interval of the coefficients—in this case the 95% confidence interval for the coefficient is [.580, 1.009].

## One-Way ANOVA

### What is the One-Way ANOVA?

ANOVA is short for **AN**alysis **Of** **VA**riance. The main purpose of an ANOVA is to test if two or more groups differ from each other significantly in one or more characteristics.

For some statisticians the ANOVA doesn't end there - they assume a cause effect relationship and say that one or more independent, controlled variables (the factors) *cause* the significant difference of one or more characteristics. The way this works is that the factors sort the data points into one of the groups and therefore they cause the difference in the mean value of the groups.

Example: Let us claim that woman have on average longer hair than men. We find twenty undergraduate students and measure the length of their hair. A conservative statistician would then claim we measured the hair of ten female and ten male students, and that we conducted an analysis of variance and found that the average hair of female undergraduate students is significantly longer than the hair of their fellow male students.

A more aggressive statistician would claim that gender has a direct relation to the length of a person's hair. Most statisticians fall into the second category. It is generally assumed that the ANOVA is an 'analysis of dependencies.' It is referred to as such because it is a test to prove an assumed cause and effect relationships. In more statistical terms it tests the effect of one or more independent variables on one or more dependent variables. It assumes an effect of  $Y = f(x_1, x_2, x_3, \dots, x_n)$ .

The ANOVA is a popular test; it is the test to use when conducting experiments. This is due to the fact that it only requires a nominal scale for the independent variables - other multivariate tests (e.g., regression analysis) require a continuous-level scale. This following table shows the required scales for some selected tests.

		Independent Variable	
		Metric	Non-metric
Dependent Variable	metric	Regression	<b>ANOVA</b>
	Non-metric	Discriminant Analysis	$\chi^2$ (Chi-Square)

The F-test, the T-test, and the MANOVA are all similar to the ANOVA. The *F-test* is another name for an ANOVA that only compares the statistical means in two groups. This happens if the independent variable for the ANOVA has only two factor steps, for example male or female as a gender.

The *T-test* compares the means of two (and only two) groups when the variances are not equal. The equality of variances (also called homoscedasticity or homogeneity) is one of the main assumptions of the ANOVA (see assumptions, Levene Test, Bartlett Test). MANOVA stands for **Multivariate Analysis of Variance**. Whereas the ANOVA can have one or more independent variables, it always has only one dependent variable. On the other hand the MANOVA can have two or more dependent variables.

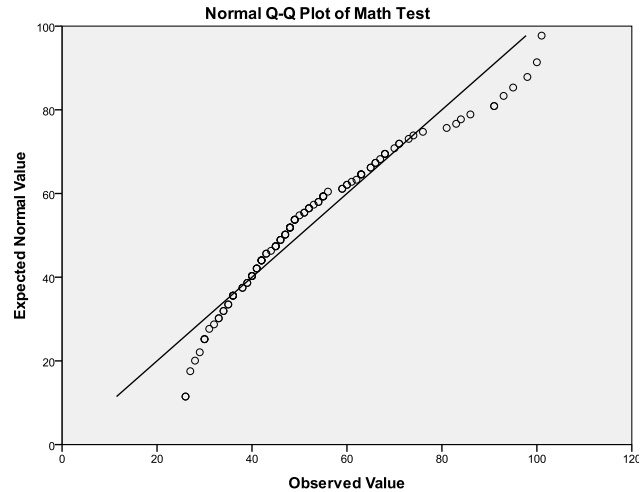
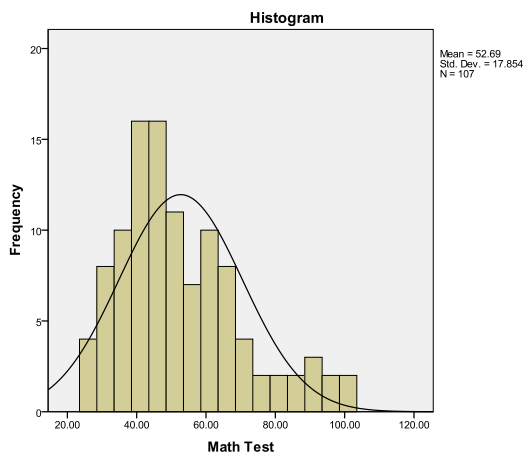
Examples for typical questions the ANOVA answers are as follows:

- **Medicine** - Does a drug work? Does the average life expectancy significantly differ between the three groups that received the drug versus the established product versus the control?
- **Sociology** - Are rich people happier? Do different income classes report a significantly different satisfaction with life?
- **Management Studies** - What makes a company more profitable? A one, three or five-year strategy cycle?

### **The One-Way ANOVA in SPSS**

Let's consider our research question from the Education studies example. Do the standardized math test scores differ between students that passed the exam and students that failed the final exam? This question indicates that our independent variable is the exam result (fail vs. pass) and our dependent variable is the score from the math test. We must now check the assumptions.

First we examine the multivariate normality of the dependent variable. We can check graphically either with a histogram (*Analyze/Descriptive Statistics/Frequencies...* and then in the menu *Charts...*) or with a Q-Q-Plot (*Analyze/Descriptive Statistics/Q-Q-Plot...*). Both plots show a somewhat normal distribution, with a skew around the mean.



Secondly, we can test for multivariate normality with the Kolmogorov-Smirnov goodness of fit test (*Analyze/Nonparacontinuous-level Test/Legacy Dialogs/1 Sample K S...*).

An alternative to the K-S test is the Chi-Square goodness of fit test, but the K-S test is more robust for continuous-level variables.

The K-S test is not significant ( $p = 0.075$ ) thus we cannot reject the null hypothesis that the sample distribution is multivariate normal. The K-S test is one of the few tests where a non-significant result ( $p > 0.05$ ) is the desired outcome.

One-Sample Kolmogorov-Smirnov Test

		Test_Score Math Test
N		107
Normal Parameters <sup>a, b</sup>	Mean	52.6916
	Std. Deviation	17.85392
Most Extreme Differences	Absolute	.124
	Positive	.124
	Negative	-.067
Kolmogorov-Smirnov Z		1.282
Asymp. Sig. (2-tailed)		.075

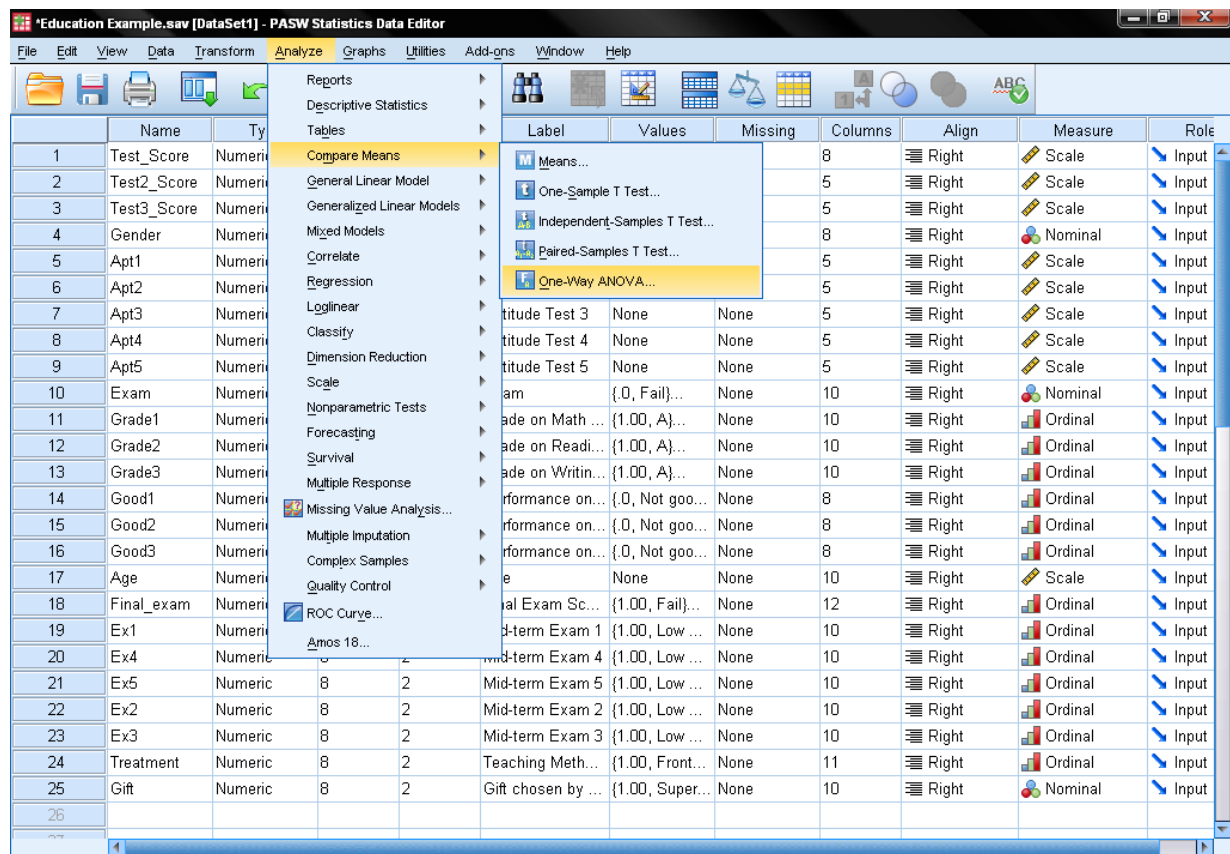
a. Test distribution is Normal.

b. Calculated from data.

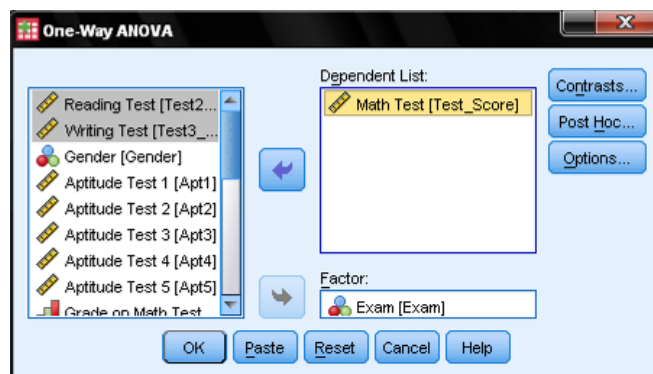
If normality is not present, we could exclude the outliers to fix the problem, center the variable by deducting the mean, or apply a non-linear transformation to the variable creating an index.

The ANOVA can be found in SPSS in *Analyze/Compare Means/One Way ANOVA*.



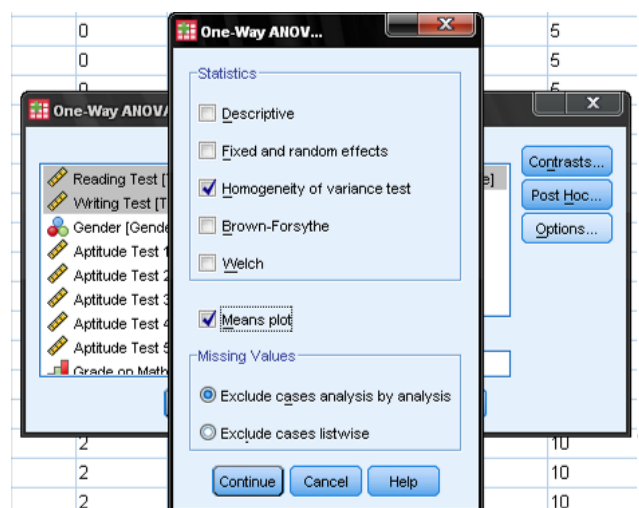


In the ANOVA dialog we need to specify our model. As described in the research question we want to test, the math test score is our dependent variable and the exam result is our independent variable. This would be enough for a basic analysis. But the dialog box has a couple more options around Contrasts, post hoc tests (also called multiple comparisons), and Options.



## Options

In the dialog box options we can specify additional statistics. If you find it useful you might include standard descriptive statistics. Generally you should select the Homogeneity of variance test (which is the Levene test of homoscedasticity), because as we find in our decision tree the outcome of this test is the



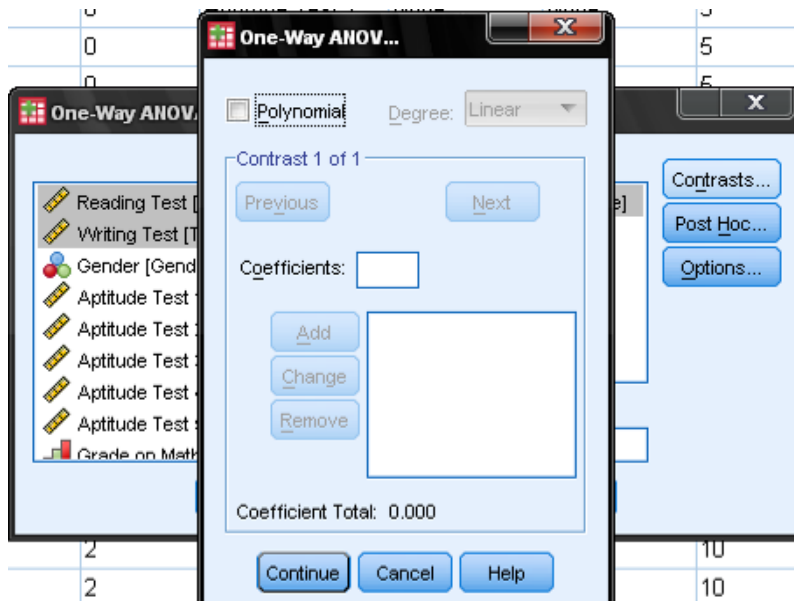
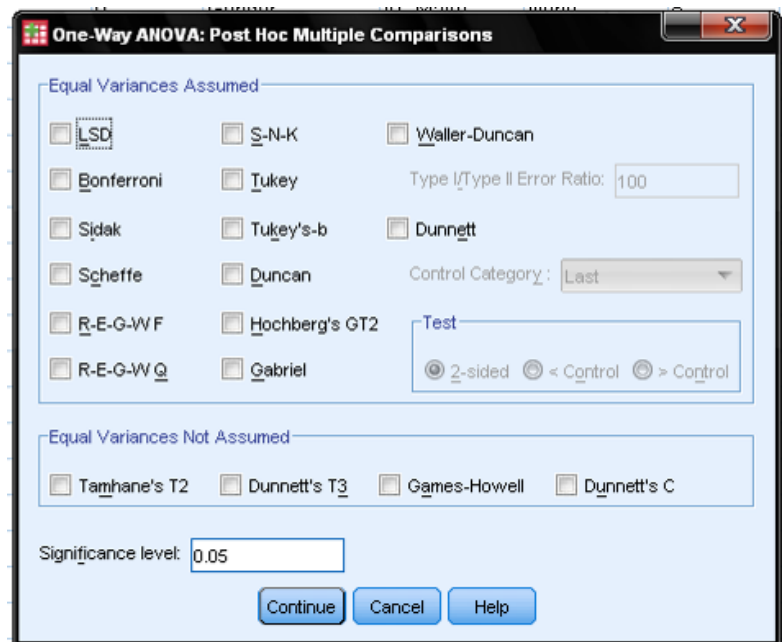
criterion that decides between the t-test and the ANOVA.

## Post Hoc Tests

Post Hoc tests are useful if your independent variable includes more than two groups. In our example the independent variable just specifies the outcome of the final exam on two factor levels – pass or fail. If more than two factor levels are given it might be useful to run pairwise tests to test which differences between groups are significant. Because executing several pairwise tests in one analysis decreases the degrees of freedom, the Bonferroni adjustment should be selected, which corrects for multiple pairwise comparisons. Another test method commonly employed is the Student-Newman-Keuls test (or short S-N-K), which pools the groups that do not differ significantly from each other. Therefore this improves the reliability of the post hoc comparison because it increases the sample size used in the comparison.

## Contrasts

The last dialog box is contrasts. Contrasts are differences in mean scores. It allows you to group multiple groups into one and test the average mean of the two groups against our third group. Please note that the contrast is not always the mean of the pooled groups! Contrast = (mean first group + mean second group)/2. It is only equal to the pooled mean, if the groups are of equal size. It is also possible to specify weights for the contrasts, e.g., 0.7 for group 1 and 0.3 for group 2. We do not specify contrasts for this demonstration.



## The Output of the One-Way ANOVA

The result consists of several tables. The first table is the Levene Test or the Test of Homogeneity of Variances (Homoscedasticity). The null hypothesis of the Levene Test is: the variances are equal. The test in our example is significant with  $p = 0.000 < 0.05$  thus we can reject the null hypothesis and cannot (!) assume that the variances are equal between the

groups with variations. Technically this means that the t-test with unequal variances is the right test to answer our research question. However, we proceed with the ANOVA.

### Test of Homogeneity of Variances

Test\_Score Math Test

Levene Statistic	df1	df2	Sig.
20.238	1	105	.000

The next table presents the results of the ANOVA. Mathematically, the ANOVA splits the total variance into explained variance (between groups) and unexplained variance (within groups), the variance is defined as  $\text{Var}(x) = \text{sum of squares}(x) / \text{degrees of freedom}(x)$ . The F-value, which is the critical test value that we need for the ANOVA is defined as  $F = \text{Var}_b / \text{Var}_w$ .

### ANOVA

Test\_Score Math Test

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	22178.402	1	22178.402	200.573	.000
Within Groups	11610.420	105	110.575		
Total	33788.822	106			

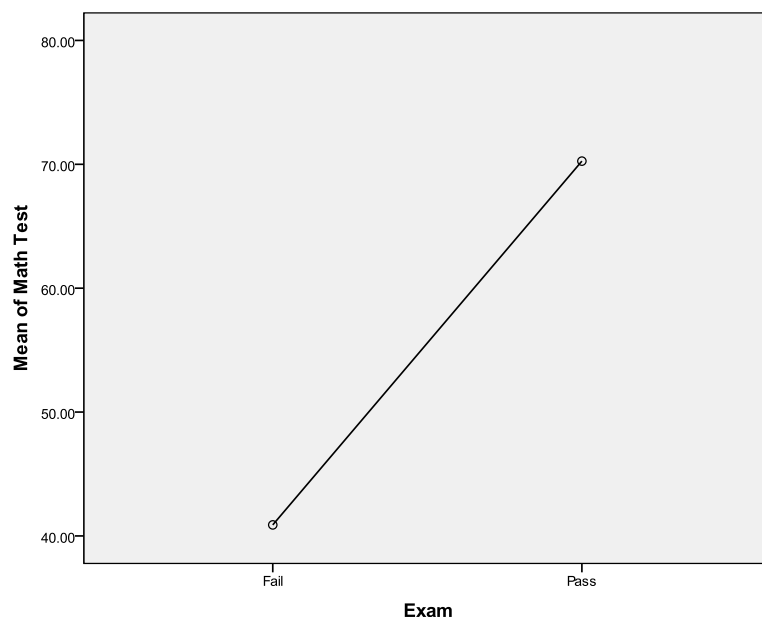
The ANOVA's F-test of significance results in  $p < 0.001$ . It tests the null hypothesis that the mean scores are equal; which is the same as saying that the independent variable has

no influence on the dependent variable. We can reject the null hypothesis and state that we can assume the mean scores to be different.

The last part of the output is the ANOVA's Means Plot. The means plot shows the marginal means for each group. If several factors are included in the analysis the slope of the line between the marginal means indicates which variable has a stronger influence on the dependent variable.

A possible write-up could be as follows:

*The ANOVA shows that a significant difference between the scores of the standardized Math Test exist between students who*



*passed and students who failed the final exam ( $p < 0.001$ ).*

## One-Way ANCOVA

### What is the One-Way ANCOVA?

ANCOVA is short for **A**nalysis of **C**ovariance. The analysis of covariance is a combination of an ANOVA and a regression analysis.

In basic terms, the ANCOVA examines the influence of an independent variable on a dependent variable while removing the effect of the covariate factor. ANCOVA first conducts a regression of the independent variable (i.e., the covariate) on the dependent variable. The residuals (the unexplained variance in the regression model) are then subject to an ANOVA. Thus the ANCOVA tests whether the independent variable still influences the dependent variable after the influence of the covariate(s) has been removed. The One-Way ANCOVA can include more than one covariate, and SPSS handles up to ten. The ANCOVA model has more than one covariate it is possible to calculate the one-way ANCOVA using contrasts just like in the ANOVA to identify the influence of each covariate.

The ANCOVA is most useful in that it (1) explains an ANOVA's within-group variance, and (2) controls confounding factors. Firstly, as explained in the chapter on the ANOVA, the analysis of variance splits the total variance of the dependent variable into:

1. Variance explained by the independent variable (also called between groups variance)
2. Unexplained variance (also called within group variance)

The ANCOVA looks at the unexplained variance and tries to explain some of it with the covariate(s). Thus it increases the power of the ANOVA by explaining more variability in the model.

Note that just like in regression analysis and all linear models, over-fitting might occur. That is, the more covariates you enter into the ANCOVA, the more variance it will explain, but the fewer degrees of freedom the model has. Thus entering a weak covariate into the ANCOVA decreases the statistical power of the analysis instead of increasing it.

Secondly, the ANCOVA eliminates the covariates effect on the relationship between independent and dependent variable that is tested with an ANOVA. The concept is very similar to the partial correlation analysis—technically it is a semi-partial regression and correlation.

The One-Way ANCOVA needs at least three variables. These variables are:

- The independent variable, which groups the cases into two or more groups. The independent variable has to be at least of nominal scale.
- The dependent variable, which is influenced by the independent variable. It has to be of continuous-level scale (interval or ratio data). Also, it needs to be homoscedastic and multivariate normal.
- The covariate, or variable that moderates the impact of the independent on the dependent variable. The covariate needs to be a continuous-level variable (interval or ratio data). The

covariate is sometimes also called confounding factor, or concomitant variable. The ANCOVA covariate is often a pre-test value or a baseline.

Typical questions the ANCOVA answers are as follows:

- **Medicine** - Does a drug work? Does the average life expectancy significantly differ between the three groups that received the drug versus the established product versus the control? This question can be answered with an ANOVA. The ANCOVA allows to additionally control for covariates that might influence the outcome but have nothing to do with the drug, for example healthiness of lifestyle, risk taking activities, or age.
- **Sociology** - Are rich people happier? Do different income classes report a significantly different satisfaction with life? This question can be answered with an ANOVA. Additionally the ANCOVA controls for confounding factors that might influence satisfaction with life, for example, marital status, job satisfaction, or social support system.
- **Management Studies** - What makes a company more profitable? A one, three or five-year strategy cycle? While an ANOVA answers the question above, the ANCOVA controls additional moderating influences, for example company size, turnover, stock market indices.

### **The One-Way ANCOVA in SPSS**

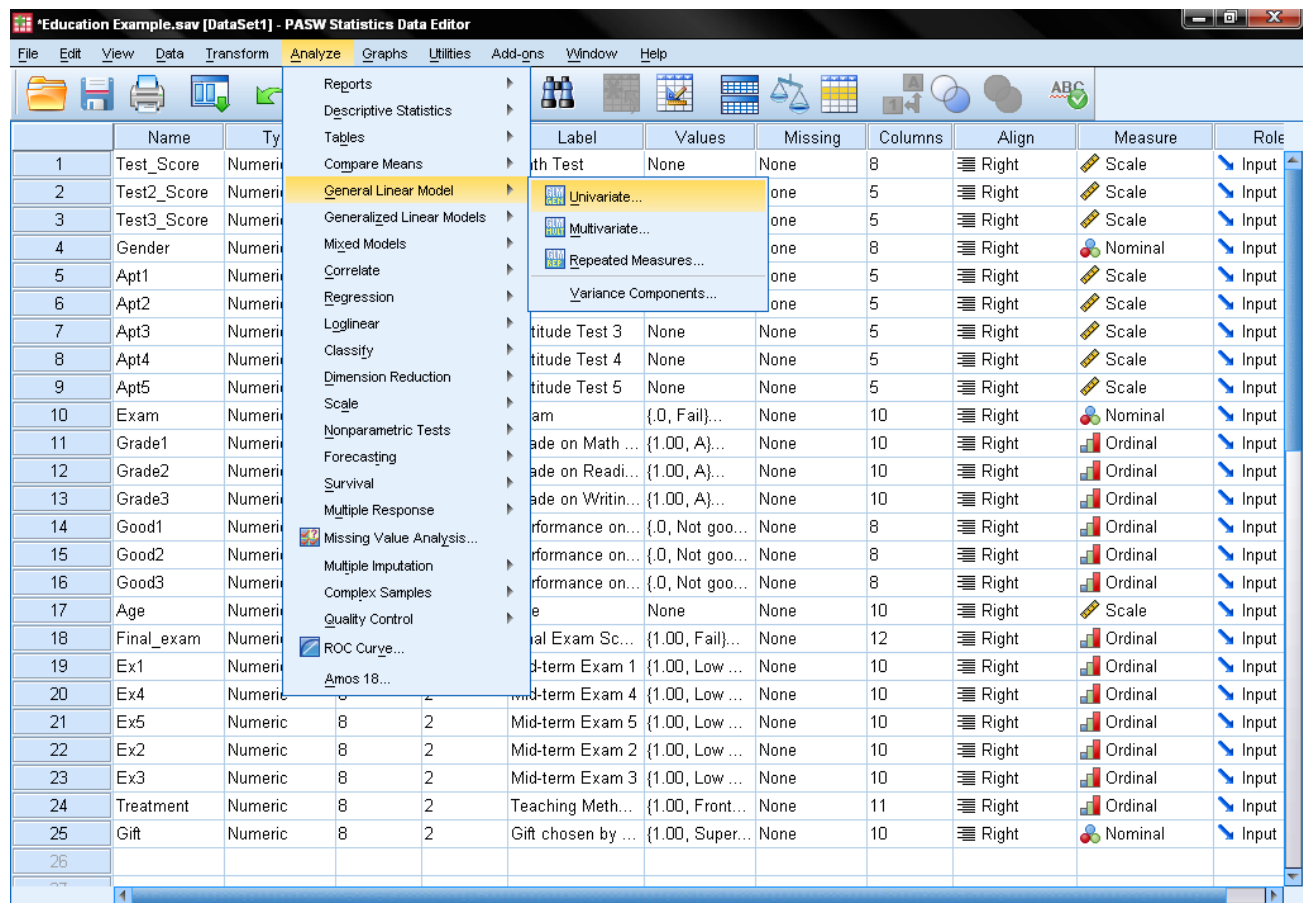
---

The One-Way ANCOVA is part of the General Linear Models (GLM) in SPSS. The GLM procedures in SPSS contain the ability to include 1-10 covariates into an ANOVA model. Without a covariate the GLM procedure calculates the same results as the ANOVA. Furthermore the GLM procedure allows specifying random factor models, which are not part of an ANCOVA design. The levels of measurement need to be defined in SPSS in order for the GLM procedure to work correctly.

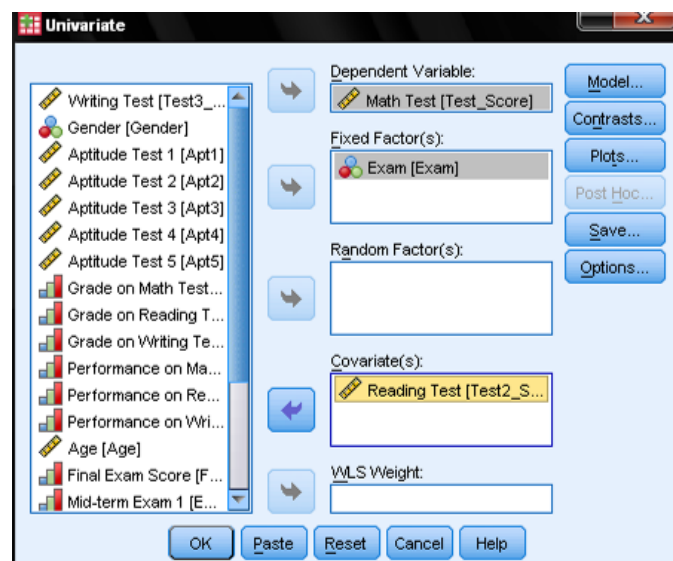
The research question for this example is as follows:

*Is there a difference in the standardized math test scores between students who passed the exam and students who failed the exam, when we control for reading abilities?*

The One-Way ANCOVA can be found in *Analyze/General Linear Model/Univariate...*



This opens the GLM dialog, which allows us to specify any linear model. For a One-Way-ANCOVA we need to add the independent variable (the factor *Exam*) to the list of fixed factors. [Remember that the factor is fixed, if it is deliberately manipulated and not just randomly drawn from a population. In our ANCOVA example this is the case. This also makes the ANCOVA the model of choice when analyzing semi-partial correlations in an experiment, instead of the partial correlation analysis which requires random data.]

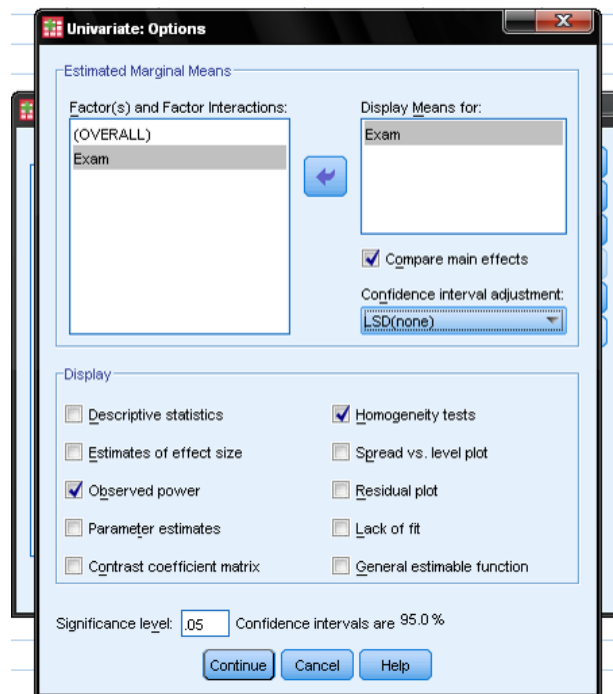


The Dependent Variable is the Students' math test score, and the covariate is the reading score.

In the dialog boxes *Model*, *Contrasts*, and *Plots* we leave all settings on the default. The field *post hocs* is disabled when one or more covariates are entered into the analysis. If it is of interest, for the factor level that has the biggest influence a contrast can be added to the analysis. If we want to compare all

groups against a specific group, we need to select *Simple* as the contrast method. We also need to specify if the first or last group should be the group to which all other groups are compared. For our example we want to compare all groups against the classroom lecture, thus we add the contrast *Exam* (Simple) first.

In the dialog *Options...* we can specify whether to display additional statistics (e.g., descriptive statistics, parameter estimates, and homogeneity tests), and which level of significance we need. This dialog also allows us to add post hoc procedures to the one-way ANCOVA. We can choose between *Bonferroni*, *LSD* and *Sidak* adjustments for multiple comparisons of the covariates.



### The Output of the One-Way ANCOVA

Once again we find that the output of the one-way ANCOVA is quite straightforward. The first table of interest includes the *Levene-Test*. The null hypothesis of the Levene test is that all variances are equal with variations across the factor levels. In our example however, the Levene test for homoscedasticity is highly significant, thus we must reject the null hypothesis and we cannot assume that the variances are equal. Strictly speaking, this violates the ANOVA assumption. Regardless, we proceed with the analysis.

The second table includes the ANCOVA

#### Levene's Test of Equality of Error Variances<sup>a</sup>

Dependent Variable: Test\_Score Math Test

F	df1	df2	Sig.
20.473	1	105	.000

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Test2\_Score + Exam



results. Note that the total sum of Type III squares is the same as in the regular ANOVA model for the same dependent and independent variable. The ANCOVA just modifies how the variance is explained.

#### Tests of Between-Subjects Effects

Dependent Variable: Test\_Score Math Test

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power <sup>b</sup>
Corrected Model	23129.643 <sup>a</sup>	2	11564.822	112.836	.000	225.672	1.000
Intercept	54237.237	1	54237.237	529.185	.000	529.185	1.000
Test2_Score	951.241	1	951.241	9.281	.003	9.281	.855
Exam	11678.362	1	11678.362	113.944	.000	113.944	1.000
Error	10659.179	104	102.492				
Total	330864.000	107					
Corrected Total	33788.822	106					

a. R Squared = .685 (Adjusted R Squared = .678)

b. Computed using alpha = .05

The results show that the covariate and the independent variable are both significant to the ANCOVA model. The next table shows the estimates for the marginal means. These are the group differences in the dependent variable after the effect of the covariate has been accounted for.

#### Estimates

Dependent Variable: Test\_Score Math Test

Exam	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
.00 Fail	42.479 <sup>a</sup>	1.369	39.765	45.193
1.00 Pass	67.892 <sup>a</sup>	1.728	64.465	71.318

a. Covariates appearing in the model are evaluated at the following values: Test2\_Score Reading Test = 19.97.

In our example the average math test score is lower for students who failed the final exam than for the students who passed after the influence of the covariate (reading test score) has been extracted.

The last table is the *Univariate Test*, which is the ANOVA test of the difference of the estimated marginal means.



### Univariate Tests

Dependent Variable: Test\_Score Math Test

	Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power <sup>a</sup>
Contrast	11678.362	1	11678.362	113.944	.000	113.944	1.000
Error	10659.179	104	102.492				

The F tests the effect of Exam. This test is based on the linearly independent pairwise comparisons among the estimated marginal means.

a. Computed using alpha = .05

A sample write-up of the ANCOVA could be as follows:

*The analysis of covariance was used to investigate the hypothesis that the observed difference in mean scores of the standardized math test is caused by differences in reading ability as measured by the standardized reading test. The ANCOVA, however, found that the marginal means of both the students who failed the exam and the students who passed the exam are highly significant ( $F = 113.944$ ,  $p < 0.001$ )—that is, after the effect of the reading score has been accounted for.*

## Factorial ANOVA

### What is the Factorial ANOVA?

ANOVA is short for **AN**alysis **Of** **V**ariance. As discussed in the chapter on the one-way ANOVA the main purpose of a one-way ANOVA is to test if two or more groups differ from each other significantly in one or more characteristics. A factorial ANOVA compares means across two or more independent variables. Again, a one-way ANOVA has one independent variable that splits the sample into two or more groups, whereas the factorial ANOVA has two or more independent variables that split the sample in four or more groups. The simplest case of a factorial ANOVA uses two binary variables as independent variables, thus creating four groups within the sample.

For some statisticians, the factorial ANOVA doesn't only compare differences but also assumes a cause-effect relationship; this infers that one or more independent, controlled variables (the factors) *cause* the significant difference of one or more characteristics. The way this works is that the factors sort the data points into one of the groups, causing the difference in the mean value of the groups.

		Independent Variables	
		1	2+
Dependent Variables	1	One-way ANOVA	<b>Factorial ANOVA</b>

Example: Let us claim that blonde women have on average longer hair than brunette women as well as men of all hair colors. We find 100

	2+	Multiple ANOVAs	MANOVA
--	----	-----------------	--------

undergraduate students and measure the length of their hair. A conservative statistician would then state that we measured the hair of 50 female (25 blondes, 25 brunettes) and 25 male students, and we conducted an analysis of variance and found that the average hair of blonde female undergraduate students was significantly longer than the hair of their fellow students. A more aggressive statistician would claim that gender and hair color have a direct influence on the length of a person's hair.

Most statisticians fall into the second category. It is generally assumed that the factorial ANOVA is an 'analysis of dependencies.' It is referred to as such because it tests to prove an assumed cause-effect relationship between the two or more independent variables and the dependent variables. In more statistical terms it tests the effect of one or more independent variables on one dependent variable. It assumes an effect of  $Y = f(x_1, x_2, x_3, \dots x_n)$ .

The factorial ANOVA is closely related to both the one-way ANOVA (which we already discussed) and the MANOVA (**M**ultivariate **A**nalysis of **V**ariance). Whereas the factorial ANOVAs can have one or more independent variables, the one-way ANOVA always has only one dependent variable. On the other hand, the MANOVA can have two or more dependent variables.

The table helps to quickly identify the right Analysis of Variance to choose in different scenarios. The factorial ANOVA should be used when the research question asks for the influence of two or more independent variables on one dependent variable.

Examples of typical questions that are answered by the ANOVA are as follows:

- **Medicine** - Does a drug work? Does the average life expectancy differ significantly between the 3 groups x 2 groups that got the drug versus the established product versus the control and for a high dose versus a low dose?
- **Sociology** - Are rich people living in the country side happier? Do different income classes report a significantly different satisfaction with life also comparing for living in urban versus suburban versus rural areas?
- **Management Studies** – Which brands from the BCG matrix have a higher customer loyalty? The BCG matrix measures brands in a brand portfolio with their business growth rate (high versus low) and their market share (high versus low). To which brand are customers more loyal – stars, cash cows, dogs, or question marks?

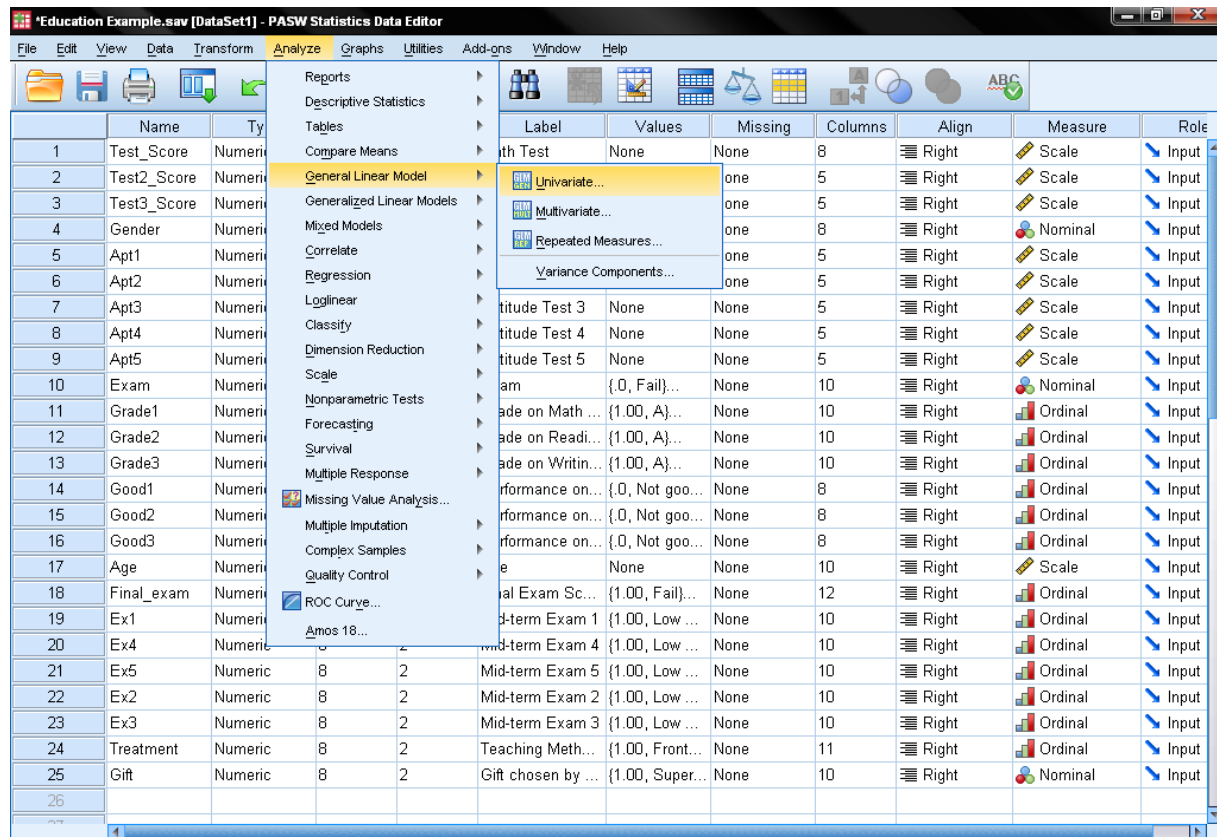
### **The Factorial ANOVA in SPSS**

Our research question for the Factorial ANOVA in SPSS is as follows:

*Do gender and passing the exam have an influence how well a student scored on the standardized math test?*

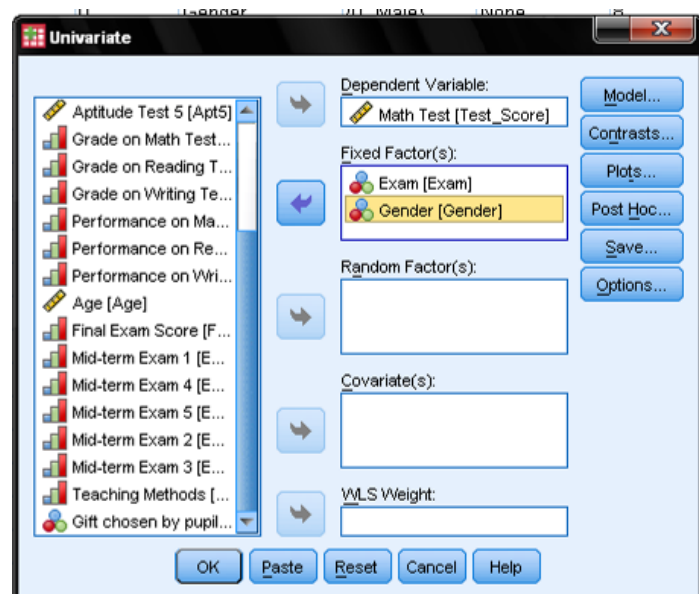
This question indicates that the dependent variable is the score achieved on the standardized math tests and the two independent variables are gender and the outcome of the final exam (pass or fail).

The factorial ANOVA is part of the SPSS GLM procedures, which are found in the menu *Analyze/General Linear Model/Univariate*.

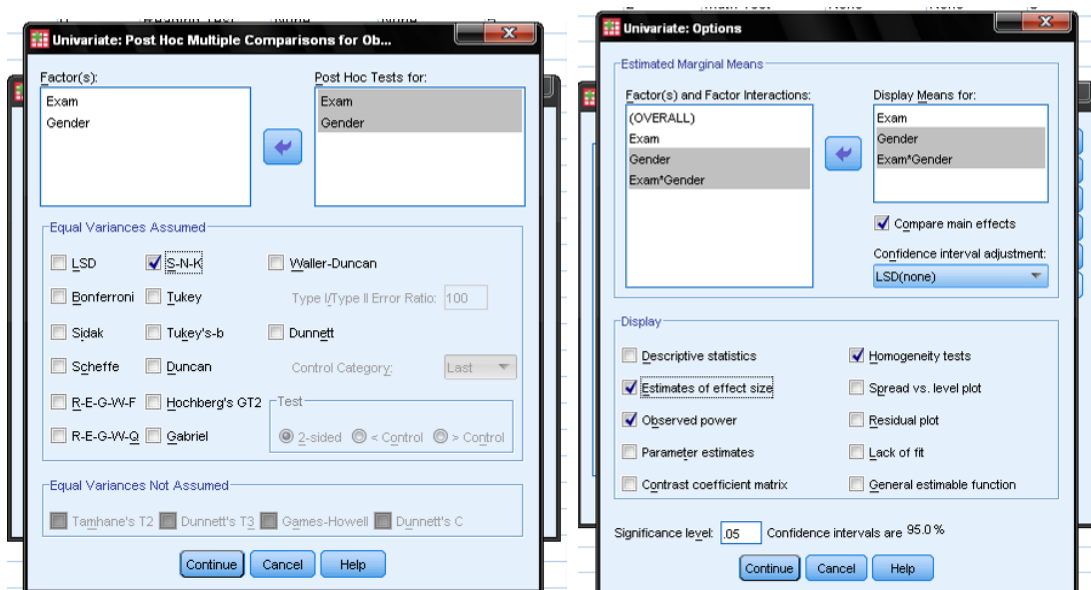


In the GLM procedure dialog we specify our full-factorial model. Dependent variable is *Math Test* with Independent variables *Exam* and *Gender*.

The dialog box *Post Hoc tests* is used to conduct a separate comparison between factor levels. This is useful if the factorial ANOVA includes factors that have more than two factor levels. In our case we included two factors of which each has only two levels. The factorial ANOVA tests the null hypothesis that all means are the same. Thus the ANOVA itself does not tell



which of the means in our design are different, or if indeed they are different. In order to do this, post hoc tests would be needed. If you want to include post hocs a good test to use is the Student-Newman-Keuls test (or short S-N-K). The SNK pools the groups that do not differ significantly from each other. Therefore it improves the reliability of the post hoc comparison by increasing the sample size used in the comparison. Another advantage is that it is simple to interpret.



The *Options* dialog allows us to add descriptive statistics, the Levene Test and the practical significance (estimated effect size) to the output and also the mean comparisons.

The Contrast dialog in the GLM procedure model us to group multiple groups into one and test the average mean of the two groups against our third group. Please note that the contrast is not always the mean of the pooled groups! Contrast = (mean first group + mean second group)/2. It is only equal to the pooled mean if the groups are of equal size. In our example we do without contrasts.

And finally the dialog *Plots...* allows us to add profile plots for the main and interaction effects to our factorial ANOVA.

## The Output of the Factorial ANOVA

The first table that is relevant is the result of the Levene Test of homogeneity of variances. In the case of a *factorial* ANOVA, the Levene Test technically tests for the homogeneity of the error variances and not the variances itself. The error variances is the variability in the error in measurement along the scale for all subgroups. As we can see the test is significant ( $p=0.000$ ) and thus we can reject the null hypothesis that the

### Levene's Test of Equality of Error Variances<sup>a</sup>

Dependent Variable: Test\_Score Math Test

F	df1	df2	Sig.
6.594	3	103	.000

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Exam + Gender + Exam \* Gender

error variance is homogenous. Although this violates the ANOVA assumption, we proceed with the test.

The next table shows the results of the factorial ANOVA. The ANOVA splits the total variance (Type III Sum of Squares) into explained variance (between groups) and unexplained variance (within groups), where the variance is  $\text{Var} = \text{sum of squares} / \text{df}$ . The F-value is then the  $F = \text{Var}_b / \text{Var}_w$ . The variance is split into explained and unexplained parts for the main effects of each factor and for the interaction effect of the factors in the analysis.

Tests of Between-Subjects Effects

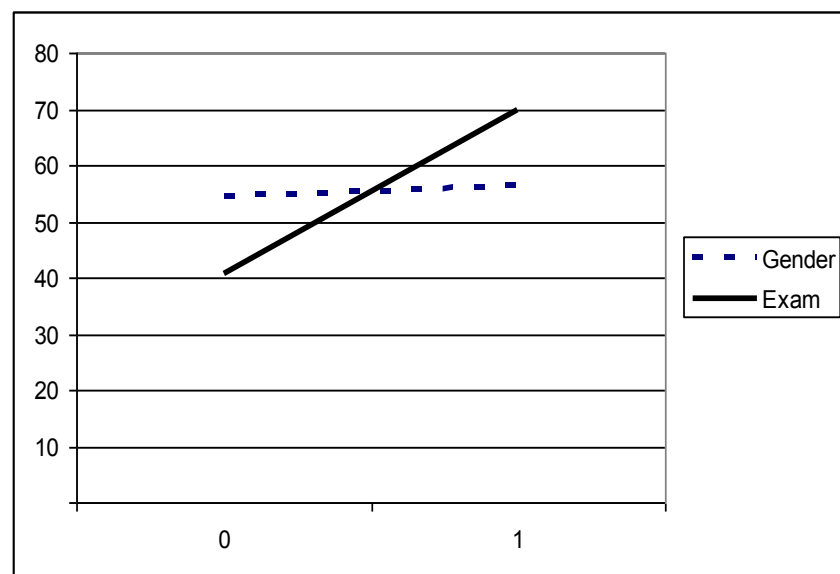
Dependent Variable: Test\_Score Math Test

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power <sup>b</sup>
Corrected Model	22258.928 <sup>a</sup>	3	7419.643	66.282	.000	.659	198.846	1.000
Intercept	243922.354	1	243922.354	2179.031	.000	.955	2179.031	1.000
Exam	16985.565	1	16985.565	151.737	.000	.596	151.737	1.000
Gender	75.018	1	75.018	.670	.415	.006	.670	.128
Exam * Gender	.172	1	.172	.002	.969	.000	.002	.050
Error	11529.894	103	111.941					
Total	330864.000	107						
Corrected Total	33788.822	106						

a. R Squared = .659 (Adjusted R Squared = .649)

b. Computed using alpha = .05

In our example the factorial ANOVA shows that the outcome of the final exam (pass vs. fail) has a significant influence on the score achieved *on the standardized math test* ( $p < 0.001$ ), but *gender* ( $p = .415$ ) and the interaction between gender and exam ( $p = .969$ ) do not. The factorial ANOVA's F-test tests the null hypothesis that the mean scores are equal; which is the same as saying that there is no effect on the dependent variable.



If both effects are significant, a marginal means plot is used to illustrate the different effect sizes. While the tables are hard to read quickly charting them in Excel helps to understand the effect sizes of our factorial ANOVA. The slope of the line of the factor *Exam* is steeper than the slope of the factor *Gender*, thus in our factorial ANOVA model exam has a larger impact than gender on the dependent variable math score.

In summary, we can conclude:

*The factorial ANOVA shows that a significant difference exists between the average math scores achieved by students who passed the exam and students who failed the final exam. However there is no significant difference between gender or the interaction effect of gender and outcome of the final exam.*

## **Factorial ANCOVA**

### **What is the Factorial ANCOVA?**

ANCOVA is short for **Analysis of Covariance**. The factorial analysis of covariance is a combination of a factorial ANOVA and a regression analysis.

In basic terms, the ANCOVA looks at the influence of two or more independent variables on a dependent variable while removing the effect of the covariate factor. ANCOVA first conducts a regression of the independent variables (the covariate) on the dependent variable. The residuals (the unexplained variance in the regression model) are then subject to an ANOVA. Thus the ANCOVA tests whether the independent variables still influence the dependent variable after the influence of the covariate(s) has been removed.

The factorial ANCOVA includes more than one independent variable and the factorial ANCOVA can include more than one covariate, SPSS handles up to ten. If the ANCOVA model has more than one covariate it is possible to run the factorial ANCOVA with contrasts and post hoc tests just like the one-way ANCOVA or the ANOVA to identify the influence of each covariate.

The factorial ANCOVA is most useful in two ways: 1) it explains a factorial ANOVA's within-group variance, and 2) it controls confounding factors.

First, the analysis of variance splits the total variance of the dependent variable into:

- Variance explained by each of the independent variables (also called between-groups variance of the main effect)
- Variance explained by all of the independent variables together (also called the interaction effect)
- Unexplained variance (also called within-group variance)

The factorial ANCOVA looks at the unexplained variance and tries to explain some of it with the covariate(s). Thus it increases the power of the factorial ANOVA by explaining more variability in the model. *[Note that just like in regression analysis and all linear models, over-fitting might occur. That is, the more covariates you enter into the factorial ANCOVA the more variance it will explain, but the fewer degrees of freedom the model has. Thus entering a weak covariate into the factorial ANCOVA decreases the statistical power of the analysis instead of increasing it.]*

Secondly, the factorial ANCOVA eliminates the covariates effect on the relationship between independent variables and the dependent variable which is tested with a factorial ANOVA. The concept is very similar to the partial correlation analysis. Technically it is a semi-partial regression and correlation.

The factorial ANCOVA needs at least four variables (the simplest case with two factors is called two-way ANCOVA):

- Two or more independent variables, which group the cases into four or more groups. The independent variable has to be at least of nominal scale.
- The dependent variable, which is influenced by the independent variable. It has to be of continuous-level scale (interval or ratio data). Also, it needs to be homoscedastic and multivariate normal
- The covariate, also referred to as the confounding factor, or concomitant variable, is the variable that moderates the impact of the independent on the dependent variable. The covariate needs to be a continuous-level variable (interval or ratio data). The ANCOVA covariate is often a pre-test value or a baseline.

Typical questions the factorial ANCOVA answers are as follows:

- **Medicine** - Does a drug work? Does the average life expectancy significantly differ between the three groups that received the drug versus the established product versus the control and accounting for the dose (high/low)? This question can be answered with a factorial ANOVA. The factorial ANCOVA allows additional control of covariates that might influence the outcome but have nothing to do with the drug, for example healthiness of lifestyle, risk taking activities, age.
- **Sociology** - Are rich people living in the countryside happier? Do different income classes report a significantly different satisfaction with life when looking where they live (urban, suburban, rural)? This question can be answered with a factorial ANOVA. Additionally the factorial ANCOVA controls for confounding factors that might influence satisfaction with life, e.g., marital status, job satisfaction, social support system.
- **Management Studies** - Which brands from the BCG matrix have a higher customer loyalty? The BCG matrix measures brands in a brand portfolio with their business growth rate (high/low) and their market share (high/ low). A factorial ANOVA answers the question to which brand are customers more loyal – stars, cash cows, dogs, or question marks? And a factorial ANCOVA can control for confounding factors, like satisfaction with the brand or appeal to the customer.

### ***The Factorial ANCOVA in SPSS***

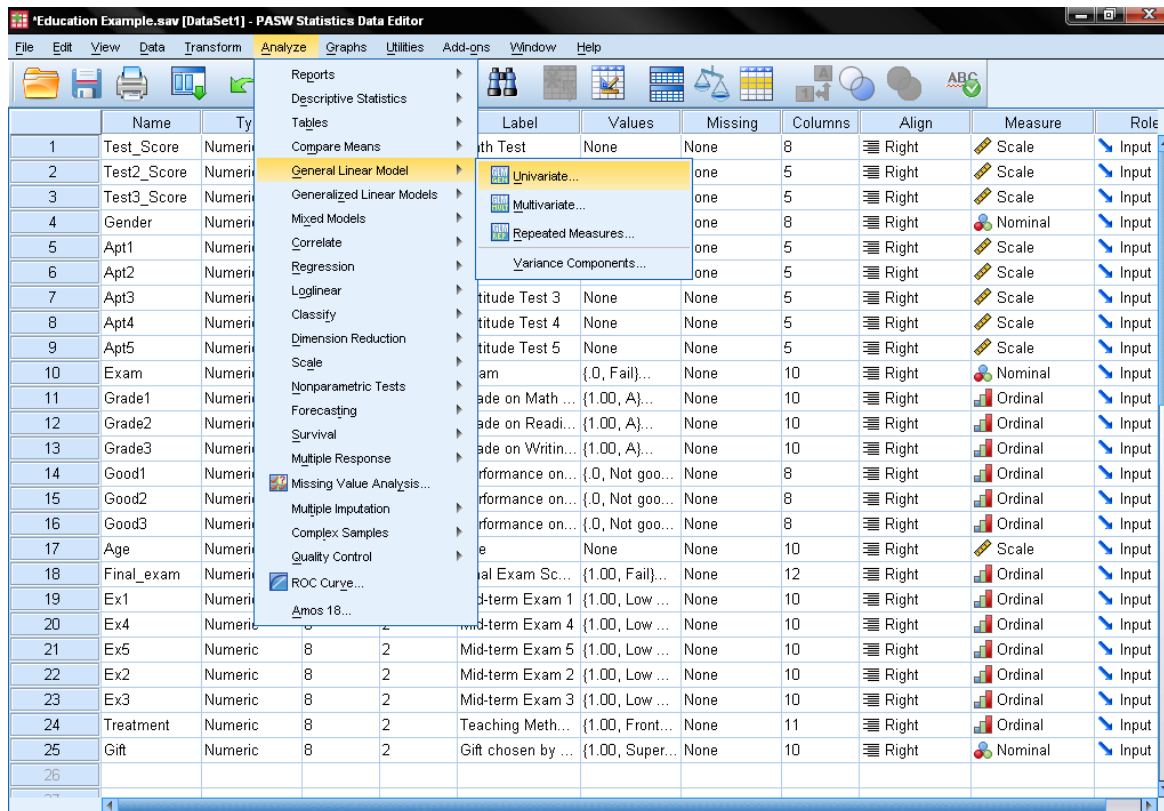
---

The Factorial ANCOVA is part of the General Linear Models in SPSS. The GLM procedures in SPSS contain the ability to include 1-10 covariates into an ANCOVA model. Without a covariate the GLM procedure calculates the same results as the Factorial ANOVA. The levels of measurement need to be defined in SPSS in order for the GLM procedure to work correctly.



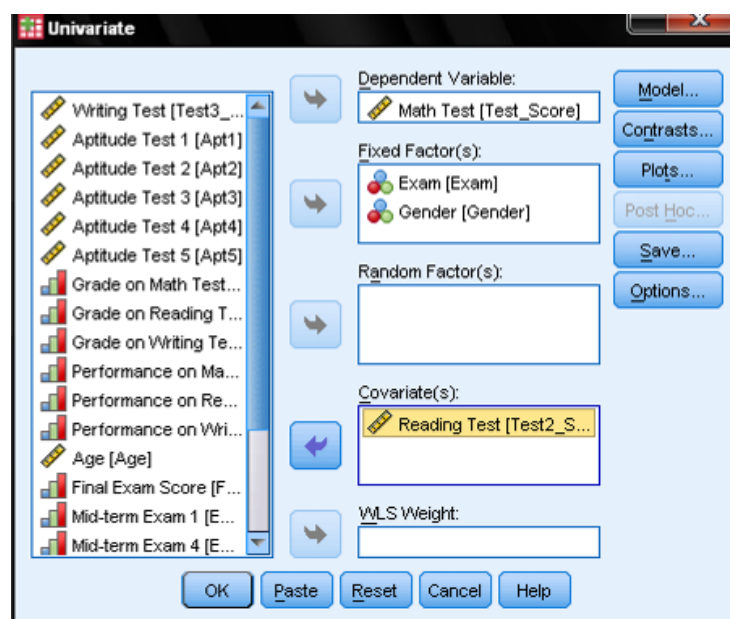
We return to the research question from the chapter on the factorial ANOVA. This time we want to know if gender and the outcome of the final exam (pass /fail) have an influence on the math score when we control for the reading ability as measured by the score of the standardized reading test.

The Factorial ANCOVA can be found in *Analyze/General Linear Model/Univariate...*



This opens the GLM dialog, which allows us to specify any linear model. To answer our research question we need to add the independent variables (*Exam* and *Gender*) to the list of fixed factors.

*[Remember that the factor is fixed if it is deliberately manipulated and not just randomly drawn from a population. In our ANCOVA example this is the case. This also makes the ANCOVA the model of choice when analyzing semi-partial correlations in an experiment, instead of the partial correlation analysis with requires random data.]*

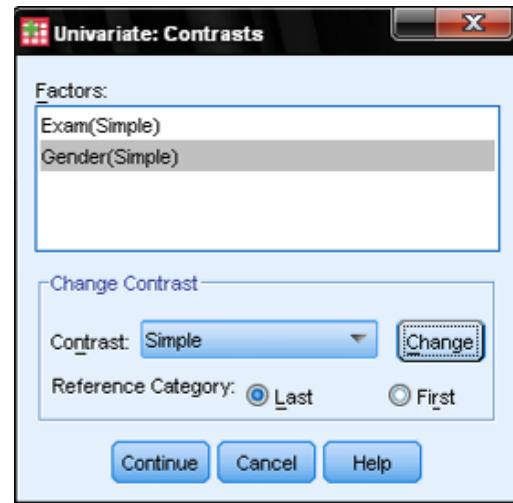




The Dependent Variable is the *Math Test*, and the Covariate is the *Reading Test*.

In the dialog box *Model...* we leave all settings on the default. The default for all GLM (including the Factorial ANCOVA) is the full factorial model. The field *post hocs* is disabled when one or more covariates are entered into the analysis. If we want to include a group comparison into our factorial ANCOVA we can add contrast to the analysis.

If we want to compare all groups against a specific group we need to select *Simple* as the Contrast Method, and we also need to specify which group (either the first or the last) should be compared against all other groups. However, since in this example both of our fixed factors only have two factor levels (male/female and pass/fail) we do not really need contrasts to answer the research question.



### The Output of the Factorial ANCOVA

The first table shows the Levene test for the equality (homoscedasticity) of error variances. The Levene test (null hypothesis= the error variances are equal) is significant, thus we can reject the null hypothesis and must assume that the error variances are not homogenous. This violates the assumption of the ANOVA. Normally you would need to stop here, however for our purposes we will dutifully note this in our findings and proceed with the analysis.

#### Levene's Test of Equality of Error Variances<sup>a</sup>

Dependent Variable: Test\_Score Math Test

F	df1	df2	Sig.
6.947	3	103	.000

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Test2\_Score + Exam + Gender + Exam \* Gender

The next table shows the Factorial ANCOVA results. The goodness of fit of the model indicates that the covariate reading test score ( $p = 0.002$ ) and the direct effects of exam ( $p < 0.001$ ) are significant, while neither the direct effect of gender ( $p = 0.274$ ) nor the interaction effect gender \* exam ( $p = 0.776$ ) are significant.

#### Tests of Between-Subjects Effects

Dependent Variable: Test\_Score Math Test

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power <sup>b</sup>
Corrected Model	23254.668 <sup>a</sup>	4	5813.667	56.293	.000	.688	225.170	1.000
Intercept	47501.253	1	47501.253	459.945	.000	.818	459.945	1.000
Test2_Score	995.740	1	995.740	9.642	.002	.086	9.642	.868
Exam	8876.780	1	8876.780	85.952	.000	.457	85.952	1.000
Gender	124.670	1	124.670	1.207	.274	.012	1.207	.193
Exam * Gender	8.393	1	8.393	.081	.776	.001	.081	.059
Error	10534.154	102	103.276					
Total	330864.000	107						
Corrected Total	33788.822	106						

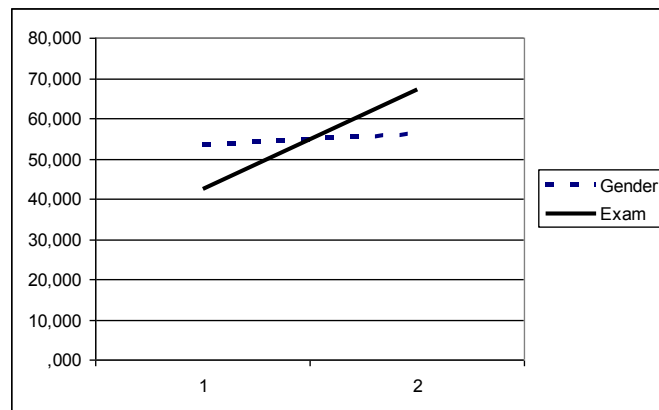
a. R Squared = .688 (Adjusted R Squared = .676)

b. Computed using alpha = .05

The practical significance of each of the variables in our factorial ANCOVA model is displayed as Partial Eta Squared, which is the partial variance explained by that variable. Eta squared ranges from 0 to 1, where 0 indicates no explanatory power and 1 indicates perfect explanatory power. Eta square is useful to compare the power of different variables, especially when designing experiments or questionnaires. It is useful to include variables on the basis of the Eta-Square Analysis.

In our example we see that the two non-significant effects, gender and gender \* exam, have a very small eta ( $< 0.01$ ). Between the two significant effects (the covariate and the outcome of the exam), the exam has a higher explanatory power than the test2 reading score.

The effect size or the power of each variable can also be seen in a partial means plot of a factorial ANCOVA. The factor with the steeper slope between factor levels typically has the higher explanatory power and thus the higher impact on the dependent variable. In our example we find that the outcome of the final exam (pass versus fail) has the bigger effect on the average score achieved in our standardized math test.



In summary, we can conclude:

*The factorial ANCOVA shows that a significant difference exists between the average math scores achieved by students who passed the exam and students who failed the final exam when controlling for the reading ability as measured by the score achieved in the standardized reading test. However there is no significant difference between gender or the interaction effect of gender and outcome of the final exam.*

#### One-Way MANOVA

## What is the One-Way MANOVA?

MANOVA is short for **M**ultivariate **A**nalysis **O**f **V**ariance. The main purpose of a one-way ANOVA is to test if two or more groups differ from each other significantly in one or more characteristics. A factorial ANOVA compares means across two or more variables. Again, a one-way ANOVA has one independent variable that splits the sample into two or more groups whereas the factorial ANOVA has two or more independent variables that split the sample in four or more groups. A MANOVA now has two or more independent variables and two or more dependent variables.

For some statisticians the MANOVA doesn't only compare differences in mean scores between multiple groups but also assumes a cause effect relationship whereby one or more independent, controlled variables (the factors) *cause* the significant difference of one or more characteristics. The factors sort the data points into one of the groups causing the difference in the mean value of the groups.

Example:

*A research team wants to test the user acceptance with a new online travel booking tool. The team conducts a study where they assign 30 randomly chosen people into 3 groups. The first group needs to book their travel through an automated online-portal; the second group books over the phone via a hotline; the third group sends a request via the online-portal and receives a call back. The team measures user acceptance as the behavioral intention to use the system, they do measure the latent construct behavioral intention with 3 variables – ease of use, perceived usefulness, effort to use.*

		Independent Variables	
		Metric	Non-metric
Dependent Variable	metric	Regression	<b>ANOVA</b>
	Non-metric	Discriminant Analysis	$\chi^2$ (Chi-Square)

In the example, some statisticians argue that the MANOVA can only find the differences in the behavioral intention to use the system. However, some statisticians argue that you can establish a causal relationship between the channel they used and the behavioral intention for future use. It is generally assumed that the MANOVA is an '*analysis of dependencies*'. It is referred to as such because it proves an assumed cause-effect relationship between two or more independent variables and two or more dependent variables. In more statistical terms, it tests the effect of one or more independent variables on one or more dependent variables.

### Other things you may want to try

When faced with a question similar to the one in our example you could also try to run a 3 *factorial* ANOVAs, testing the influence of the three independent variables (the three channels) on each of the three dependent variables (ease of use, perceived usefulness, effort to use) individually. However

running multiple factorial ANOVAs does not account for the full variability in all three dependent variables and thus the test has lesser power than the MANOVA.

Another thing you might want to try is running a factor analysis on the three dependent variables and then running a factorial ANOVA. The factor analysis reduces the variance within the three dependent variables to one factor, thus this procedure does have lesser power than the MANOVA.

A third approach would be to conduct a discriminant analysis and switch the dependent and independent variables. That is the discriminant analysis uses the three groups (online, phone, call back) as the dependent variable and identifies the significantly discriminating variables from the list of continuous-level variables (ease of use, perceived usefulness, effort to use).

Mathematically, the MANOVA is fully equivalent to the discriminant analysis. The difference consists of a switching of the independent and dependent variables. Both the MANOVA and the discriminant analysis are a series of canonical regressions. The MANOVA is therefore the best test use when conducting experiments with latent variables. This is due to the fact that it only requires a nominal scale for the independent variables which typically represent the treatment. This includes multiple continuous-level independent variables – which typically measure one latent (not directly observable) construct.

The MANOVA is much like the one-way ANOVA and the factorial ANOVA in that the one-way ANOVA has exactly one independent and one dependent variable. The factorial ANOVAs can have one or more independent variables but always has only one dependent variable. On the other hand the MANOVA can have two or more dependent variables.

The following table helps to quickly identify the right analysis of variance to choose in different scenarios.

Examples of typical questions that are answered by the MANOVA are as follows:

- **Medicine** - Does a drug work? Does the average life expectancy, perceived pain, and level of side-effects significantly differ between the three experimental groups that got the drug versus the established product, versus the control—and within each of the groups two subgroups for a high dose versus a low dose?
- **Sociology** - Are rich people living in the country side happier? Do they enjoy their lives more and have a more positive outlook on their futures? Do different income classes report a significantly different satisfaction, enjoyment and outlook on their lives? Does the area in which they live (suburbia/city/rural) affect their happiness and positive outlook?
- **Management Studies** – Which brands from the BCG matrix do have a higher customer loyalty, brand appeal, customer satisfaction? The BCG matrix measures brands in a brand portfolio with their business growth rate (high/low) and their market share (high/low). To

		Independent Variables	
		1	2+
Dependent Variables	1	One-way ANOVA	Factorial ANOVA
	2+	Multiple ANOVAs	<b>MANOVA</b>

which brand are customers more loyal, more attracted, and more satisfied with? Stars, cash cows, dogs, or question marks?

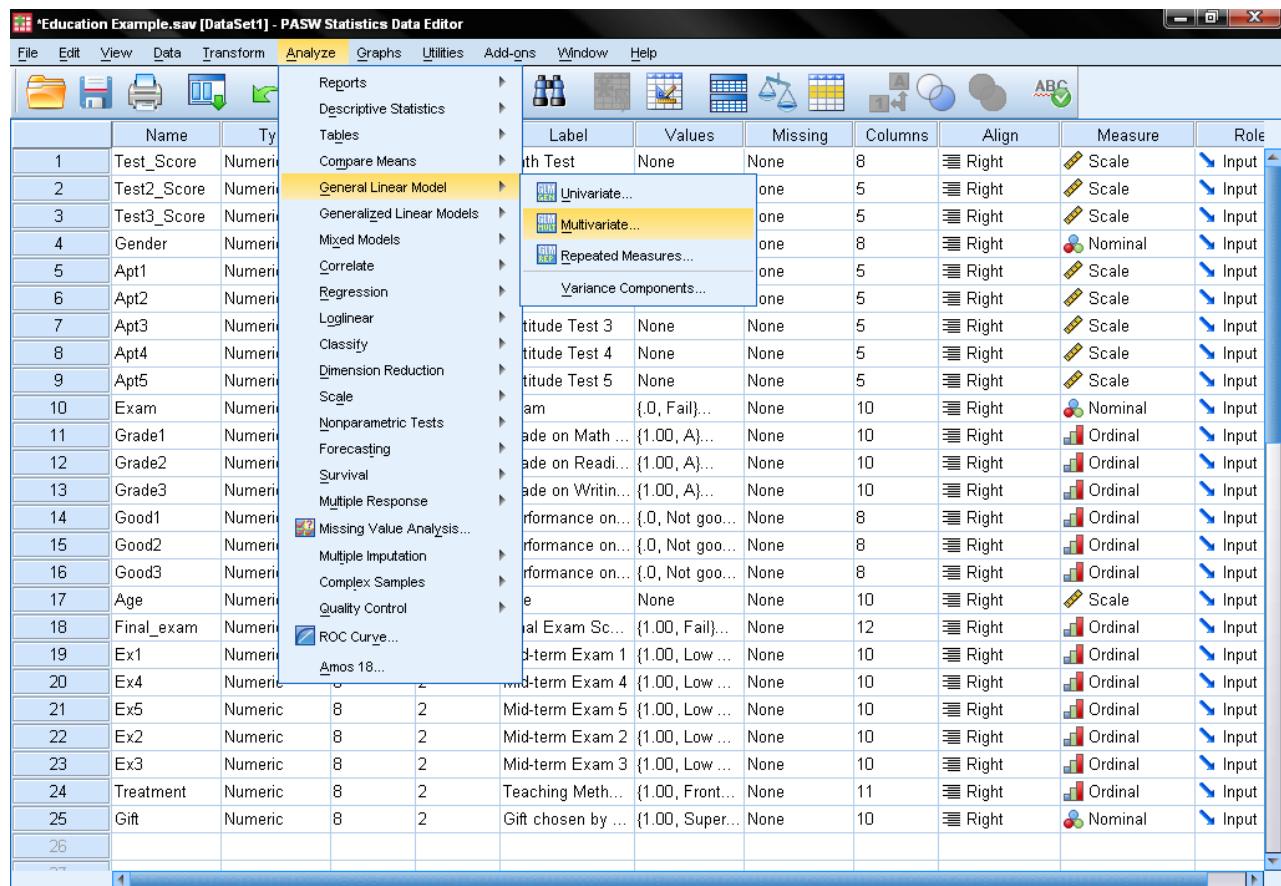
### The One-Way MANOVA in SPSS

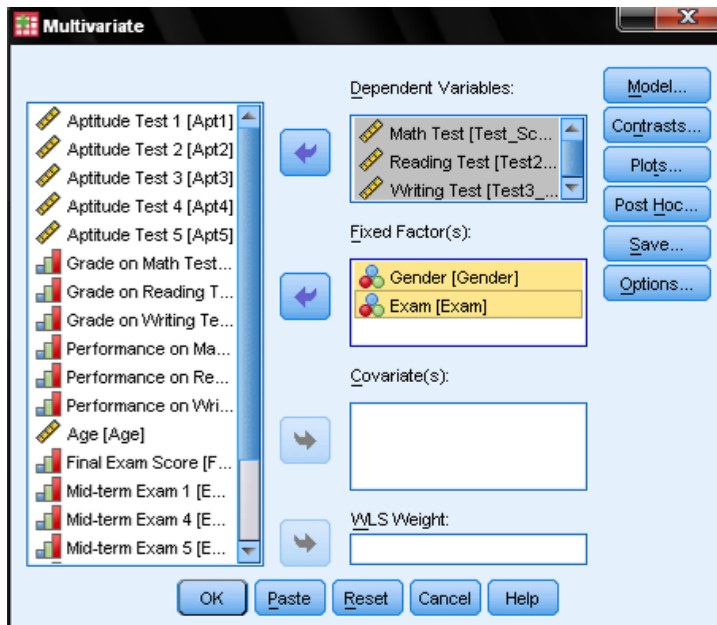
Our research question for the one-way MANOVA in SPSS is as follows:

*Do gender and the outcome of the final exam influence the standardized test scores of math, reading, and writing?*

The research question indicates that this analysis has multiple independent variables (exam and gender) and multiple dependent variables (math, reading, and writing test scores). We will skip the check for multivariate normality of the dependent variables; the sample we are going to look at has some violations of the assumption set forth by the MANOVA.

The MANOVA can be found in SPSS in *Analyze/General Linear Model/Multivariate...*, which opens the dialog for multivariate GLM procedure (that is GLM with more than one dependent variable). The multivariate GLM model is used to specify the MANOVAs.





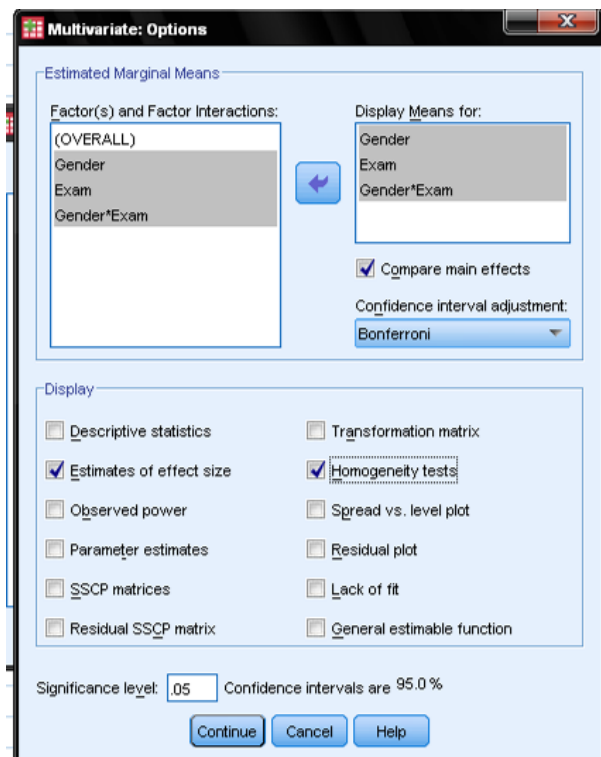
To answer our research question we need to specify a full-factorial model that includes the test scores for math, reading, and writing as dependent variable. Plus the independent variables gender and exam, which represent a fixed factor in our research design.

The dialog box *Post Hoc tests* is used to conduct a separate comparison between factor levels, this is useful if the MANOVA includes factors have more than two factor levels. In our case we select two factors and each has only two factor levels (male/female and pass/fail). The MANOVA's F-test will

test the null hypothesis that all means are the same. It does not indicate which of the means in our design are different. In order to find this information, post hoc tests need to be conducted as part of our MANOVA. In order to compare different groups (i.e., factor levels) we select the Student-Newman-Keuls test (or short S-N-K), which pools the groups that do not differ significantly from each other, thereby improving the reliability of the post hoc comparison by increasing the sample size used in the comparison. Additionally, it is simple to interpret.

The *Options...* Dialog allows us to add descriptive statistics, the Levene Test and the practical significance to the output. Also we might want to add the pairwise t-tests to compare the marginal means of the main and interaction effects. The Bonferroni adjustment corrects the degrees of freedom to account for multiple pairwise tests.

The *Contrast...* dialog in the GLM procedure model gives us the option to group multiple groups into one and test the average mean of the two groups against our third group. Please note that the contrast is not always the mean of the pooled groups! Because Contrast = (mean first group + mean second group)/2. It is only equal to the pooled mean if the groups are of



equal size. In our example we do without contrasts.

Lastly the *Plots...* dialog allows us to add profile plots for the main and interaction effects to our MANOVA. However it is easier to create the marginal means plots that are typically reported in academic journals in Excel.

### The Output of the One-Way MANOVA

The first relevant table for our One-Way MANOVA is Box's M Test of equality of covariance matrices. For the dependent variables, the MANOVA requires the covariances to be homogenous in all dependent variables. The Box Test does not confirm this since we have to reject the null hypothesis that the observed covariance matrices are equal ( $p = 0.001$ ).

#### Box's Test of Equality of Covariance Matrices<sup>a</sup>

Box's M	45.951
F	2.356
df1	18
df2	5751.928
Sig.	.001

The next table shows the overall model tests of significance. Although Wilk's Lambda is typically used to measure the overall goodness of fit of the model, SPSS computes other measures as well. In our example we see that exam has a significant influence on the dependent variables, while neither gender nor the interaction effect of gender \* exam have a significant influence on the dependent variables.

#### Multivariate Tests<sup>b</sup>

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	.956	729.739 <sup>a</sup>	3.000	101.000	.000	.956
	Wilks' Lambda	.044	729.739 <sup>a</sup>	3.000	101.000	.000	.956
	Hotelling's Trace	21.675	729.739 <sup>a</sup>	3.000	101.000	.000	.956
	Roy's Largest Root	21.675	729.739 <sup>a</sup>	3.000	101.000	.000	.956
Gender	Pillai's Trace	.037	1.283 <sup>a</sup>	3.000	101.000	.284	.037
	Wilks' Lambda	.963	1.283 <sup>a</sup>	3.000	101.000	.284	.037
	Hotelling's Trace	.038	1.283 <sup>a</sup>	3.000	101.000	.284	.037
	Roy's Largest Root	.038	1.283 <sup>a</sup>	3.000	101.000	.284	.037
Exam	Pillai's Trace	.617	54.184 <sup>a</sup>	3.000	101.000	.000	.617
	Wilks' Lambda	.383	54.184 <sup>a</sup>	3.000	101.000	.000	.617
	Hotelling's Trace	1.609	54.184 <sup>a</sup>	3.000	101.000	.000	.617
	Roy's Largest Root	1.609	54.184 <sup>a</sup>	3.000	101.000	.000	.617
Gender * Exam	Pillai's Trace	.018	.609 <sup>a</sup>	3.000	101.000	.611	.018
	Wilks' Lambda	.982	.609 <sup>a</sup>	3.000	101.000	.611	.018
	Hotelling's Trace	.018	.609 <sup>a</sup>	3.000	101.000	.611	.018
	Roy's Largest Root	.018	.609 <sup>a</sup>	3.000	101.000	.611	.018

a. Exact statistic

b. Design: Intercept + Gender + Exam + Gender \* Exam

The next table is the result of the Levene Test of homogeneity of error variances. In the case of a MANOVA the Levene Test technically tests for the homogeneity of the error variances, that is, the variability in the error in measurement along the scale. As we can see the test is not significant for two of the three dependent variables ( $p=0.000$ ,  $0.945$ , and  $0.524$ ). Therefore we cannot reject the null hypothesis that the error variance is homogenous for only reading and writing.

**Levene's Test of Equality of Error Variances<sup>a</sup>**

	F	df1	df2	Sig.
Test_Score Math Test	6.594	3	103	.000
Test2_Score Reading Test	.125	3	103	.945
Test3_Score Writing Test	.752	3	103	.524

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Gender + Exam + Gender \* Exam

The next table shows the results of the MANOVA. The MANOVA extracts the roots of the dependent variables and then basically examines a factorial ANOVAs on the root of the dependent variables for all independent variables. The MANOVA splits their total variance into explained variance (between groups) and unexplained variance (within groups), where the variance is  $\text{Var} = \text{sum of squares} / \text{df}$ . The F-value is then the  $F = \text{Var}_b / \text{Var}_w$ . This is done for the main effects each factor has on its own and the interaction effect of the factors.



Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	Test_Score Math Test	22258.928 <sup>a</sup>	3	7419.643	66.282	.000	.659
	Test2_Score Reading Test	5509.431 <sup>b</sup>	3	1836.477	14.975	.000	.304
	Test3_Score Writing Test	6450.412 <sup>c</sup>	3	2150.137	29.593	.000	.463
Intercept	Test_Score Math Test	243922.354	1	243922.354	2179.031	.000	.955
	Test2_Score Reading Test	38558.896	1	38558.896	314.418	.000	.753
	Test3_Score Writing Test	25565.231	1	25565.231	351.857	.000	.774
Gender	Test_Score Math Test	75.018	1	75.018	.670	.415	.006
	Test2_Score Reading Test	81.866	1	81.866	.668	.416	.006
	Test3_Score Writing Test	91.832	1	91.832	1.264	.264	.012
Exam	Test_Score Math Test	16985.565	1	16985.565	151.737	.000	.596
	Test2_Score Reading Test	4837.909	1	4837.909	39.449	.000	.277
	Test3_Score Writing Test	5661.856	1	5661.856	77.925	.000	.431
Gender * Exam	Test_Score Math Test	.172	1	.172	.002	.969	.000
	Test2_Score Reading Test	78.740	1	78.740	.642	.425	.006
	Test3_Score Writing Test	91.646	1	91.646	1.261	.264	.012
Error	Test_Score Math Test	11529.894	103	111.941			
	Test2_Score Reading Test	12631.485	103	122.636			
	Test3_Score Writing Test	7483.774	103	72.658			
Total	Test_Score Math Test	330864.000	107				
	Test2_Score Reading Test	60821.000	107				
	Test3_Score Writing Test	40469.000	107				
Corrected Total	Test_Score Math Test	33788.822	106				
	Test2_Score Reading Test	18140.916	106				
	Test3_Score Writing Test	13934.187	106				

a. R Squared = .659 (Adjusted R Squared = .649)

b. R Squared = .304 (Adjusted R Squared = .283)

c. R Squared = .463 (Adjusted R Squared = .447)

In our example the MANOVA shows that the exam (pass versus fail) has a significant influence on the math, reading, and writing scores, while neither gender nor the interaction effect between gender and exam have a significant influence on the dependent variables. The MANOVA's F-test tests the null hypothesis that the mean scores are equal, which is the same as saying that there is no effect on the dependent variable.

In summary, a possible conclusion could read:

*The MANOVA shows that the outcome of the final exam significantly influences the standardized test scores of reading, writing, and math. All scores are significantly higher for students passing the exam. However no significant difference for gender or the interaction effect of exam and gender could be found in our sample.*

## One-Way MANCOVA

### What is the One-Way MANCOVA?

MANCOVA is short for **M**ultivariate **A**nalysis of **C**ovariance. The words “one” and “way” in the name indicate that the analysis includes only one independent variable. Like all analyses of covariance, the MANCOVA is a combination of a One-Way MANOVA preceded by a regression analysis.

In basic terms, the MANCOVA looks at the influence of one or more independent variables on one dependent variable while removing the effect of one or more covariate factors. To do that the One-Way MANCOVA first conducts a regression of the covariate variables on the dependent variable. Thus it eliminates the influence of the covariates from the analysis. Then the residuals (the unexplained variance in the regression model) are subject to an MANOVA, which tests whether the independent variable still influences the dependent variables after the influence of the covariate(s) has been removed. The One-Way MANCOVA includes one independent variable, one or more dependent variables and the MANCOVA can include more than one covariate, and SPSS handles up to ten. If the One-Way MANCOVA model has more than one covariate it is possible to run the MANCOVA with contrasts and post hoc tests just like the one-way ANCOVA or the ANOVA to identify the strength of the effect of each covariate.

The One-Way MANCOVA is most useful for two things: 1) explaining a MANOVA's within-group variance, and 2) controlling confounding factors. Firstly, as explained in the section titled ‘*How To Conduct a MANOVA*,’ the analysis of variance splits the total variance of the dependent variable into:

- Variance explained by each of the independent variables (also called between-groups variance of the main effect)
- Variance explained by all of the independent variables together (also called the interaction effect)
- Unexplained variance (also called within-group variance)

The One-Way MANCOVA looks at the unexplained variance and tries to explain some of it with the covariate(s). Thus it increases the power of the MANOVA by explaining more variability in the model. *[Note that just like in regression analysis and all linear model's over-fitting might occur. That is, the more covariates you enter into the MANCOVA the more variance will be explained, but the fewer degrees of freedom the model has. Thus entering a weak covariate into the One-Way MANCOVA decreases the statistical power of the analysis instead of increasing it.]*

Secondly, the One-Way MANCOVA eliminates the covariates' effects on the relationship between independent variables and the dependent variables—an effect that is typically tested using a MANOVA. The concept is very similar to the concept behind partial correlation analysis; technically a MANCOVA is a semi-partial regression and correlation.

The One-Way MANCOVA needs at least four variables:

- One independent variable, which groups the cases into two or more groups, i.e., it has two or more factor levels. The independent variable has to be at least of nominal scale.
- Two or more dependent variables, which the independent variable influences. The dependent variables have to be of continuous-level scale (interval or ratio data). Also, they need to be homoscedastic and multivariate normal.
- One or more covariates, also called confounding factors or concomitant variables. These variables moderate the impact of the independent factor on the dependent variables. The covariates need to be continuous-level variables (interval or ratio data). The One-Way MANCOVA covariate is often a pre-test value or a baseline.

### The One-Way MANCOVA in SPSS

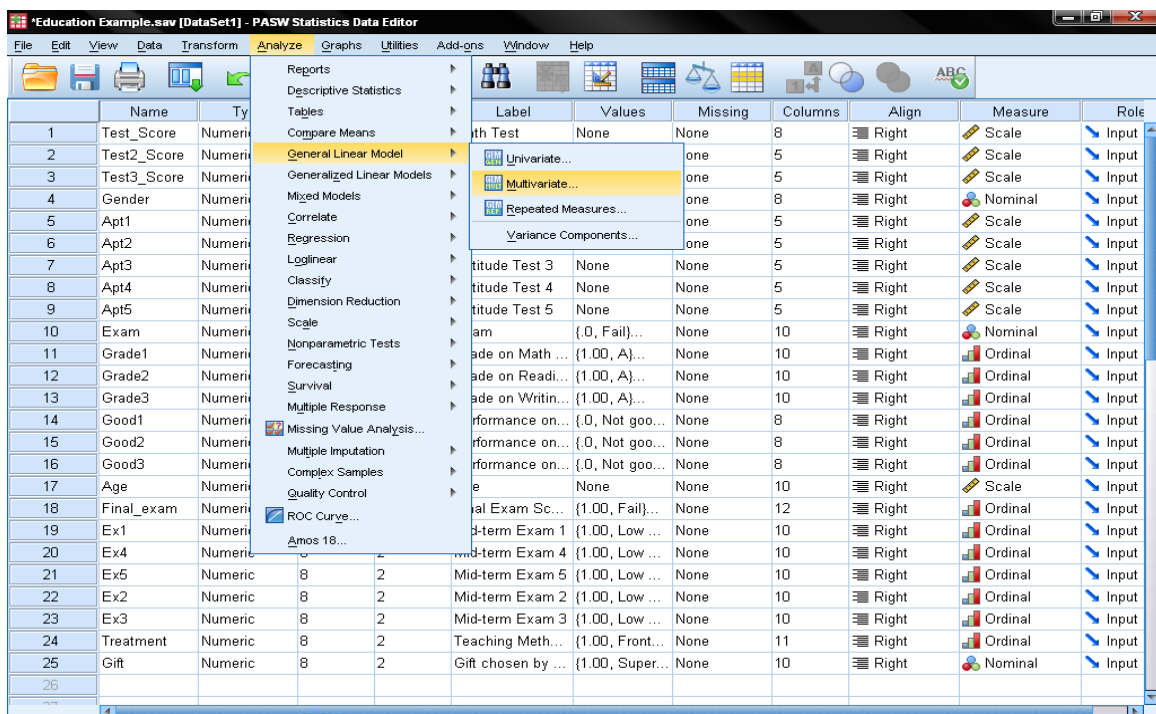
The One-Way MANCOVA is part of the General Linear Models in SPSS. The GLM procedure in SPSS has the ability to include 1-10 covariates into a MANCOVA model. Without a covariate the GLM procedure calculates the same results as the MANOVA. The levels of measurement need to be defined upfront in order for the GLM procedure to work correctly.

Let us analyze the following research question:

*Does the score achieved in the standardized math, reading, and writing test depend on the outcome of the final exam, when we control for the age of the student?*

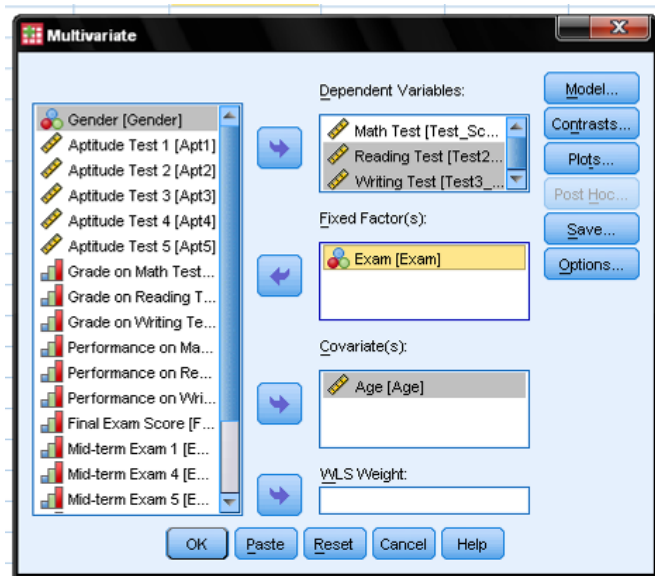
This research question means that the three test scores are the dependent variables, the outcome of the exam (fail vs. pass) is the independent variable and the age of the student is the covariate factor.

The One-Way MANCOVA can be found in *Analyze/General Linear Model/Multivariate...*

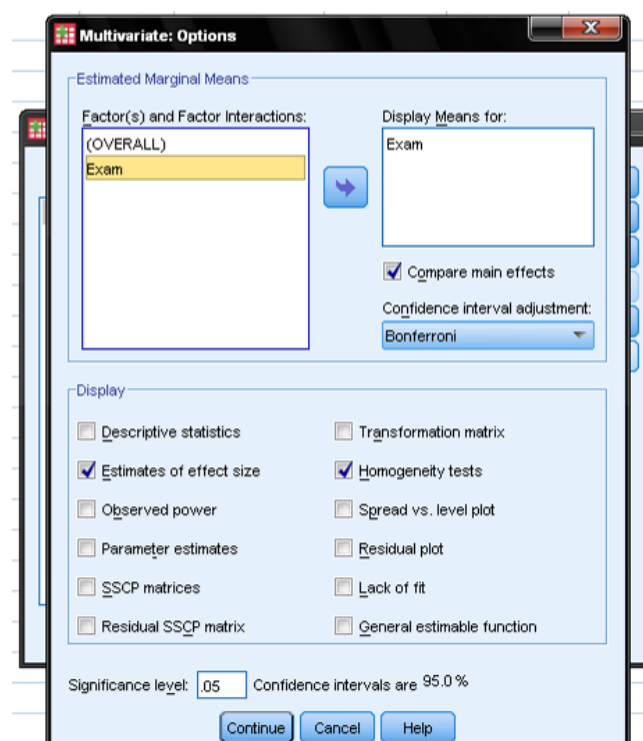


A click on this menu entry brings up the GLM dialog, which allows us to specify any linear model. For MANCOVA design we need to add the independent variable (exam) to the list of fixed factors. *[Remember that the factor is fixed, if it is deliberately manipulated and not just randomly drawn from a population. In our MANCOVA example this is the case. This also makes the ANCOVA the model of choice when analyzing semi-partial correlations in an experiment, instead of the partial correlation analysis which requires random data.]*

We need to specify a full-factorial model where the covariate is the students' age, and the dependent variables are the math, reading, and writing test scores. In the dialog box *Model...* we leave all settings on the default. The default for all GLM (including the MANCOVA) is the full factorial model.



The field *post hocs* is disabled when one or more covariates are entered into the analysis. If we want to include a group comparison into our MANCOVA we would need to add contrasts to the analysis. If you wanted to compare all groups against a specific group you would need to select *Simple* as the Contrast Method, and also need to specify which group (the first or last) should be compared against all other groups.



In the *Options...* dialog we can specify the additional statistics that SPSS is going to calculate. It is useful to include the marginal means for the factor levels and also to include the Levene test of homogeneity of error variances and the practical significance eta.

If the MANCOVA is a factorial MANCOVA and not a One-Way MANCOVA, i.e., includes more than one independent variable, you could choose to compare the main effects of those independent variables. The MANCOVA output would then include multiple ANOVAs that compare the factor levels of the independent variables. However, even if we adjust the confidence interval using the Bonferroni method, conducting multiple pairwise ANOVAs

will multiply the error terms. Thus this method of testing main effects is typically not used anymore, and has been replaced by multivariate tests, e.g., Wilk's Lambda.

### The Output of the One-Way MANCOVA

The very first table in the output just shows the sample size for each cell in the MANCOVA's design. The second table is a bit more relevant in answering our research question. The table includes the results of Box's M Test for the Equality of Covariances. Box's M test has the null hypothesis that the covariance matrices are equal across the cells. In our example the Box test is highly significant, thus we cannot reject our null hypothesis. Technically we would need to stop here, however we will proceed even though this might make our T-tests and F-tests unreliable. To rectify this problem we could also simply run a log transformation on the three dependent variables and create a math, reading and writing index. However, we can do without this step since the F-Test is robust against the violation of this assumption.

#### Box's Test of Equality of Covariance Matrices<sup>a</sup>

Box's M	33.056
F	5.329
df1	6
df2	54751.383
Sig.	.000

The next table shows the results of the multivariate tests of goodness of fit of the model. They tell us if the independent variable and the covariates have an influence on the dependent variables. In our case only the independent variable (exam) has a significant influence; the covariate (age) is not significant.

#### Multivariate Tests<sup>b</sup>

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	.093	3.494 <sup>a</sup>	3.000	102.000	.018	.093
	Wilks' Lambda	.907	3.494 <sup>a</sup>	3.000	102.000	.018	.093
	Hotelling's Trace	.103	3.494 <sup>a</sup>	3.000	102.000	.018	.093
	Roy's Largest Root	.103	3.494 <sup>a</sup>	3.000	102.000	.018	.093
Age	Pillai's Trace	.050	1.774 <sup>a</sup>	3.000	102.000	.157	.050
	Wilks' Lambda	.950	1.774 <sup>a</sup>	3.000	102.000	.157	.050
	Hotelling's Trace	.052	1.774 <sup>a</sup>	3.000	102.000	.157	.050
	Roy's Largest Root	.052	1.774 <sup>a</sup>	3.000	102.000	.157	.050
Exam	Pillai's Trace	.670	68.946 <sup>a</sup>	3.000	102.000	.000	.670
	Wilks' Lambda	.330	68.946 <sup>a</sup>	3.000	102.000	.000	.670
	Hotelling's Trace	2.028	68.946 <sup>a</sup>	3.000	102.000	.000	.670
	Roy's Largest Root	2.028	68.946 <sup>a</sup>	3.000	102.000	.000	.670

a. Exact statistic

b. Design: Intercept + Age + Exam

The next table tests the assumption of homoscedasticity. We find that the Levene Test is not significant for the reading and writing tests. Thus we cannot reject the null hypothesis that

#### Levene's Test of Equality of Error Variances<sup>a</sup>

	F	df1	df2	Sig.
Test_Score Math Test	19.551	1	105	.000
Test2_Score Reading Test	.024	1	105	.877
Test3_Score Writing Test	1.057	1	105	.306

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Age + Exam

the variances are equal. We may assume that the data of the reading and writing tests is homoscedastic in all cells. For the math tests, however, we may not assume homoscedasticity. For large samples, the T-tests and F-tests that follow are somewhat robust against this problem. Again, a log transformation or centering the data on the mean might rectify the problem.

The next table finally shows the MANCOVA results. The direct effects of the exam are significant for all three tests (math, reading, and writing). The effect of the covariate age is not significant ( $p > 0.05$  –  $p$  is 0.665, 0.07, and 0.212).

**Tests of Between-Subjects Effects**

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	Test_Score Math Test	22199.417 <sup>a</sup>	2	11099.708	99.606	.000	.657
	Test2_Score Reading Test	5785.738 <sup>b</sup>	2	2892.869	24.351	.000	.319
	Test3_Score Writing Test	6423.836 <sup>c</sup>	2	3211.918	44.477	.000	.461
Intercept	Test_Score Math Test	574.895	1	574.895	5.159	.025	.047
	Test2_Score Reading Test	957.835	1	957.835	8.063	.005	.072
	Test3_Score Writing Test	381.866	1	381.866	5.288	.023	.048
Age	Test_Score Math Test	21.015	1	21.015	.189	.665	.002
	Test2_Score Reading Test	399.306	1	399.306	3.361	.070	.031
	Test3_Score Writing Test	113.928	1	113.928	1.578	.212	.015
Exam	Test_Score Math Test	22132.733	1	22132.733	198.613	.000	.656
	Test2_Score Reading Test	5454.096	1	5454.096	45.910	.000	.306
	Test3_Score Writing Test	6347.110	1	6347.110	87.892	.000	.458
Error	Test_Score Math Test	11589.406	104	111.437			
	Test2_Score Reading Test	12355.178	104	118.800			
	Test3_Score Writing Test	7510.350	104	72.215			
Total	Test_Score Math Test	330864.000	107				
	Test2_Score Reading Test	60821.000	107				
	Test3_Score Writing Test	40469.000	107				
Corrected Total	Test_Score Math Test	33788.822	106				
	Test2_Score Reading Test	18140.916	106				
	Test3_Score Writing Test	13934.187	106				

a. R Squared = .657 (Adjusted R Squared = .650)

b. R Squared = .319 (Adjusted R Squared = .306)

c. R Squared = .461 (Adjusted R Squared = .451)

In summary, a possible write-up could be as follows:

*The MANCOVA shows that the outcome of the final exam significantly influences the standardized test scores of reading, writing, and math. All scores are significantly higher for students passing the exam than for the students failing the exam. When controlling the effect for the age of the student however, we find no significant effect of the covariate factor and no changes in the external validity of the influence of the exam outcome on the test scores.*

## Repeated Measures ANOVA

### What is the Repeated Measures ANOVA?

The repeated measures ANOVA is a member of the ANOVA family. ANOVA is short for **AN**alysis **Of** **VA**riance. All ANOVAs compare one or more mean scores with each other; they are tests for the difference in mean scores. The repeated measures ANOVA compares means across one or more variables that are based on repeated observations. A repeated measures ANOVA model can also include zero or more independent variables. Again, a repeated measures ANOVA has at least 1 dependent variable that has more than one observation.

Example:

*A research team wants to test the user acceptance of a new online travel booking tool. The team conducts a study where they assign 30 randomly chosen people into two groups. One group uses the new system and another group acts as a control group and books its travel via phone. The team measures the user acceptance of the system as the behavioral intention to use the system in the first 4 weeks after it went live. Since user acceptance is a latent behavioral construct the researchers measure it with three items – ease of use, perceived usefulness, and effort to use.*

The repeated measures ANOVA is an ‘*analysis of dependencies*’. It is referred to as such because it is a test to prove an assumed cause-effect relationship between the independent variable(s), if any, and the dependent variable(s).

When faced with a question similar to the one in our example, you could also try to run 4 MANOVAs, testing the influence of the independent variables on each of the observations of the four weeks. Running multiple ANOVAs, however, does not account for individual differences in baselines of the participants of the study.

The repeated measures ANOVA is similar to the dependent sample T-Test, because it also compares the mean scores of one group to another group on different observations. It is necessary for the repeated measures ANOVA for the cases in one observation to be directly linked with the cases in all other observations. This automatically happens when repeated measures are taken, or when analyzing similar units or comparable specimen.

The pairing of observations or making repeated measurements are very common when conducting experiments or making observations with time lags. Pairing the measured data points is typically done in order to exclude any confounding or hidden factors (cf. *partial correlation*). It is also often used to account for individual differences in the baselines, such as pre-existing conditions in clinical research. Consider for example a drug trial where the participants have individual differences that might have an impact on the outcome of the trial. The typical drug trial splits all participants into a control and the treatment group and measures the effect of the drug in month 1 -18. The repeated measures ANOVA

can correct for the individual differences or baselines. The baseline differences that might have an effect on the outcome could be typical parameter like blood pressure, age, or gender. Thus the repeated measures ANOVA analyzes the effect of the drug while excluding the influence of different baseline levels of health when the trial began.

Since the pairing is explicitly defined and thus new information added to the data, paired data can always be analyzed with a regular ANOVA as well, but not vice versa. The baseline differences, however, will not be accounted for.

A typical guideline to determine whether the repeated measures ANOVA is the right test is to answer the following three questions:

- Is there a direct relationship between each pair of observations, e.g., before vs. after scores on the same subject?
- Are the observations of the data points definitely not random (i.e., they must not be a randomly selected specimen of the same population)?
- Do all observations have to have the same number of data points?

If the answer is yes to all three of these questions the dependent sample t-test is the right test. If not, use the ANOVA or the t-test. In statistical terms the repeated measures ANOVA requires that the within-group variation, which is a source of measurement errors, can be identified and excluded from the analysis.

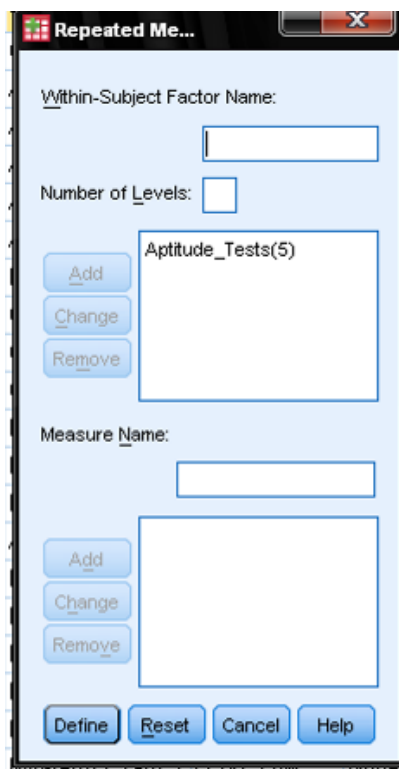
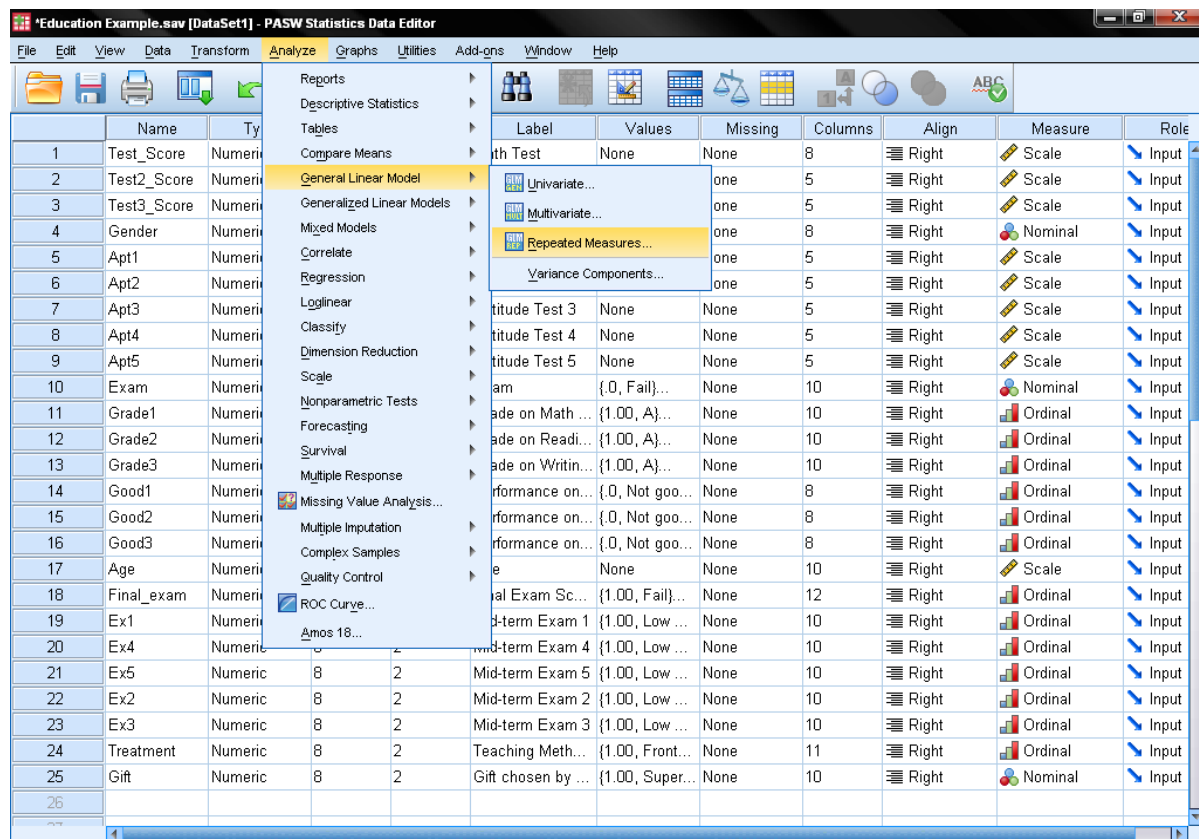
### ***The Repeated Measures ANOVA in SPSS***

---

Let us return to our aptitude test question in consideration of the repeated measures ANOVA. The question being: *“Is there a difference between the five repeated aptitude tests between students who passed the exam and the students who failed the exam?”* Since we ran the aptitude tests multiple times with the students these are considered repeated measurements. The repeated measures ANOVA uses the GLM module of SPSS, like the factorial ANOVAs, MANOVAs, and MANCOVAs.

The repeated measures ANOVA can be found in SPSS in the menu Analyze/General Linear Model/Repeated Measures...

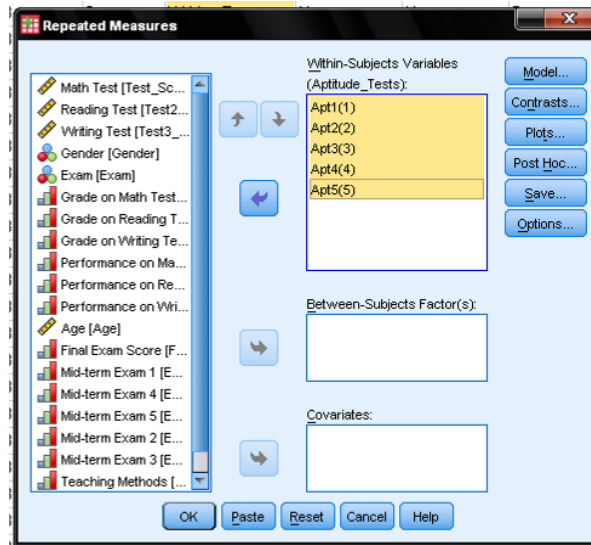




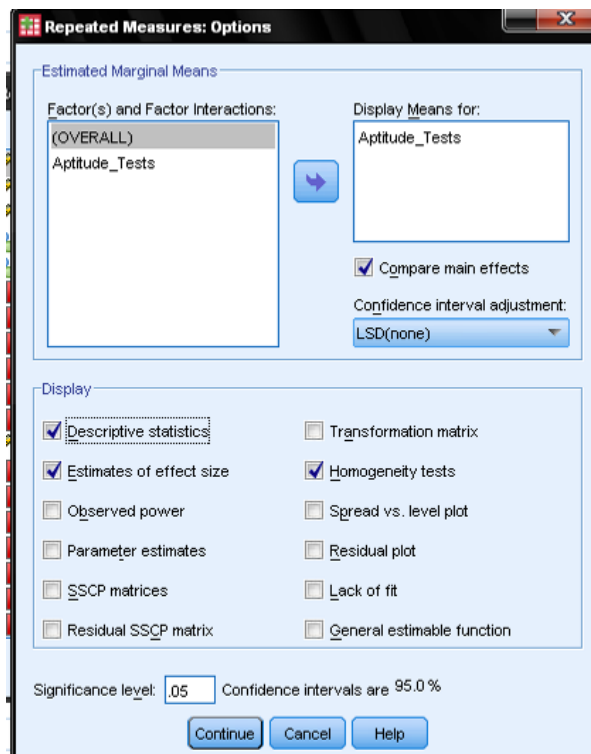
The dialog box that opens on the click is different than the GLM module you might know from the MANOVA. Before specifying the model we need to group the repeated measures.

We specify the repeated measures by creating a within-subject factor. It is called within-subject factor of our repeated measures ANOVA because it represents the different observations of one subject (so the measures are made within one single case). We measured the aptitude on five longitudinal data points. Therefore we have five levels of the within-subject factor. If we just want to test whether the data differs significantly over time we are done after we created and added the factor Aptitude\_Tests(5).

The next dialog box allows us to specify the repeated measures ANOVA. First we need to add the five observation points to the within-subject variables simply select the five aptitude test points and click on the arrow pointing towards the list of within-subject variables. In a more complex example we could also include additional dependent variables into the analysis. Plus we can add treatment/grouping variables to the repeated measures ANOVA, in such a case the grouping variable would be added as a between-subject factor.



Since our example does not have an independent variable the post hoc tests and contrasts are not needed to compare individual differences between levels of the between-subject factor. We also go with the default option of the full factorial model (in the *Model...* dialog box). If you were to conduct a post hoc test, SPSS would run a couple of pairwise dependent samples t-tests. We only add some useful statistics to the repeated measures ANOVA output in the *Options...* dialog.



Technically we only need the Levene test for homoscedasticity when we would include at least one independent variable in the sample. However it is checked here out of habit so that we don't forget to select it for the other GLM procedures we run.

It is also quite useful to include the descriptive statistics, because we have not yet compared the longitudinal development of the five administered aptitude tests.

### ***The Output of the Repeated Measures ANOVA***

The first table just lists the design of the within-subject factor in our repeated measures ANOVA. The second table lists the descriptive statistics for the five tests. We find that there is little movement within the test scores, the second test scoring lower and then the numbers picking up again.

#### **Within-Subjects Factors**

Measure: MEASURE\_1

Aptitude_Tests		Descriptive Statistics			
Dependent Variable			Mean	Std. Deviation	N
1	Apt1	Apt1 Aptitude Test 1	29.44	9.370	107
2	Apt2	Apt2 Aptitude Test 2	24.67	7.864	107
3	Apt3	Apt3 Aptitude Test 3	26.27	7.020	107
4	Apt4	Apt4 Aptitude Test 4	31.02	6.103	107
5	Apt5	Apt5 Aptitude Test 5	29.49	7.846	107

The next table shows the results of the regression modeling the GLM procedure conducts. Since our rather simple example of a simple repeated measures ANOVA does not include any regression, component we can skip this table.

#### **Multivariate Tests<sup>b</sup>**

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Aptitude_Tests	Pillai's Trace	.347	13.699 <sup>a</sup>	4.000	103.000	.000	.347
	Wilks' Lambda	.653	13.699 <sup>a</sup>	4.000	103.000	.000	.347
	Hotelling's Trace	.532	13.699 <sup>a</sup>	4.000	103.000	.000	.347
	Roy's Largest Root	.532	13.699 <sup>a</sup>	4.000	103.000	.000	.347

a. Exact statistic

b. Design: Intercept  
Within Subjects Design: Aptitude\_Tests

One of the key assumptions of the repeated measures ANOVA is sphericity. Sphericity is a measure for the structure of the covariance matrix in repeated designs. Because repeated designs violate the assumption of independence between measurements, the covariances need to be spherical. One stricter form of sphericity is compound symmetry, which occurs if all the covariances are approximately equal and all the variances are approximately equal in the samples. Mauchly's Sphericity Test tests this assumption. If there is no sphericity in the data the repeated measures ANOVA can still be done when the F-values are corrected by deducting additional degrees of freedom, e.g., Greenhouse-Geisser or Huynh-Feldt.

### Mauchly's Test of Sphericity<sup>b</sup>

Measure: MEASURE\_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon <sup>a</sup>		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Aptitude_Tests	.499	72.559	9	.000	.708	.729	.250

Mauchly's Test tests the null hypothesis that the error covariance of the orthonormalized transformed dependent variable is proportional to an identity matrix. In other words, the relationship between different observation points is similar; the differences between the observations have equal variances. This assumption is similar to homoscedasticity (tested by the Levene Test) which assumes equal variances between groups, not observations. In our example, the assumption of sphericity has not been met because the Mauchly's Test is significant. This means that the F-values of our repeated measures ANOVA are likely to be too large. This can be corrected by decreasing the degrees of freedom used. The last three columns (Epsilon) tell us the appropriate correction method to use. If epsilon is greater than 0.75 then we should use the Huynh-Feldt correction or the Greenhouse-Geisser correction. SPSS automatically corrects the F-values in the f-statistics table of the repeated measures ANOVA.

The next table shows the f statistics (also called the within-subject effects). As discussed earlier the assumption of sphericity has not been met and thus the degrees of freedoms in our repeated measures ANOVA need to be decreased. The table shows that the differences in our repeated measures are significant on a level  $p < 0.001$ . The table also shows that the Greenhouse-Geisser correction has decreased the degrees of freedom from 4 to 2.831.

### Tests of Within-Subjects Effects

Measure: MEASURE\_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Aptitude_Tests	Sphericity Assumed	2920.393	4	730.098	11.224	.000	.096
	Greenhouse-Geisser	2920.393	2.831	1031.529	11.224	.000	.096
	Huynh-Feldt	2920.393	2.917	1001.206	11.224	.000	.096
	Lower-bound	2920.393	1.000	2920.393	11.224	.001	.096
Error(Aptitude_Tests)	Sphericity Assumed	27580.807	424	65.049			
	Greenhouse-Geisser	27580.807	300.100	91.906			
	Huynh-Feldt	27580.807	309.189	89.204			
	Lower-bound	27580.807	106.000	260.196			

Thus we can reject the null hypothesis that the repeated measures are equal and we might assume that our repeated measures are different from each other. Since the repeated measures ANOVA only conducts a global F-test the pairwise comparison table helps us find the significant differences in the observations. Here we find that the first aptitude test is significantly different than the second and third; the second is only significantly different from the first, the fourth, and the fifth etc.

# Pairwise Comparisons

Measure: MEASURE\_1

(I) Aptitude_Tests	(J) Aptitude_Tests	Mean Difference (I-J)	Std. Error	Sig. <sup>a</sup>	95% Confidence Interval for Difference <sup>a</sup>	
					Lower Bound	Upper Bound
1	2	4.766 <sup>*</sup>	1.444	.001	1.903	7.630
	3	3.168 <sup>*</sup>	1.184	.009	.821	5.515
	4	-1.579	1.252	.210	-4.061	.902
	5	-.047	1.428	.974	-2.878	2.784
2	1	-4.766 <sup>*</sup>	1.444	.001	-7.630	-1.903
	3	-1.598	.932	.089	-3.446	.250
	4	-6.346 <sup>*</sup>	.906	.000	-8.143	-4.549
	5	-4.813 <sup>*</sup>	.873	.000	-6.544	-3.082
3	1	-3.168 <sup>*</sup>	1.184	.009	-5.515	-.821
	2	1.598	.932	.089	-.250	3.446
	4	-4.748 <sup>*</sup>	.945	.000	-6.621	-2.874
	5	-3.215 <sup>*</sup>	1.030	.002	-5.258	-1.172
4	1	1.579	1.252	.210	-.902	4.061
	2	6.346 <sup>*</sup>	.906	.000	4.549	8.143
	3	4.748 <sup>*</sup>	.945	.000	2.874	6.621
	5	1.533	.812	.062	-.077	3.143
5	1	.047	1.428	.974	-2.784	2.878
	2	4.813 <sup>*</sup>	.873	.000	3.082	6.544
	3	3.215 <sup>*</sup>	1.030	.002	1.172	5.258
	4	-1.533	.812	.062	-3.143	.077

Based on estimated marginal means

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

a. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

In summary, a possible write-up could be:

*During the fieldwork, five repeated aptitude tests were administered to the students. The repeated measures ANOVA shows that achieved scores on these aptitude tests are significantly different. A pairwise comparison identifies the aptitude tests 1, 2, 3 being significantly different from each other; also the tests 2 and 3 are significantly different from 4 and 5.*

## Repeated Measures ANCOVA

### What is the Repeated Measures ANCOVA?

The repeated measures ANCOVA is a member of the GLM procedures. ANCOVA is short for **Analysis of Covariance**. All GLM procedures compare one or more mean scores with each other; they are tests for the difference in mean scores. The repeated measures ANCOVA compares means across one or more variables that are based on repeated observations while controlling for a confounding variable. A repeated measures ANOVA model can also include zero or more independent variables and up to ten covariate factors. Again, a repeated measures ANCOVA has at least one dependent variable and one covariate, with the dependent variable containing more than one observation.

Example:

*A research team wants to test the user acceptance of a new online travel booking tool. The team conducts a study where they assign 30 randomly chosen people into two groups. One group that uses the new system and another group acts as a control group and books its travel via phone. The team also records self-reported computer literacy of each user. The team measures the user acceptance of the system as the behavioral intention to use the system in the first four weeks after it went live. Since user acceptance is a latent behavioral construct the researchers measure it with three items – ease of use, perceived usefulness, and effort to use. They now use the repeated measures ANCOVA to find out if the weekly measurements differ significantly from each other, if the treatment and control group differ significantly from each other all the time controlling for the influence of computer literacy.*

When faced with a question similar to the one in our example, try to run 4 MANCOVAs to test the influence of the independent variables on each of the observations of the four weeks while controlling for the covariate. Keep in mind, running multiple ANOVAs does not account for individual differences in baselines of the participants of the study. Technically, the assumption of independence is violated because the numbers of week two are not completely independent from the numbers from week one.

The repeated measures ANCOVA is similar to the dependent sample t-Test, and the repeated measures ANOVA because it also compares the mean scores of one group to another group on different observations. It is necessary for the repeated measures ANCOVA that the cases in one observation are directly linked with the cases in all other observations. This automatically happens when repeated measures are taken, or when analyzing similar units or comparable specimen.

Both strategies (pairing of observations or making repeated measurements) are very common when conducting experiments or making observations with time lags. Pairing the observed data points is typically done to exclude any confounding or hidden factors (cf. partial correlation). It is also used to account for individual differences in the baselines, for example pre-existing conditions in clinical research. Consider the example of a drug trial where the participants have individual differences that might have an impact on the outcome of the trial. The typical drug trial splits all participants into a control and the treatment group and measures the effect of the drug in months 1 - 18. The repeated

measures ANCOVA can correct for the individual differences or baselines. The baseline differences that might have an effect on the outcome could be a typical parameter like blood pressure, age, or gender. Not only does the repeated measures ANCOVA account for difference in baselines, but also for effects of confounding factors. This allows the analysis of interaction effects between the covariate, the time and the independent variables' factor levels.

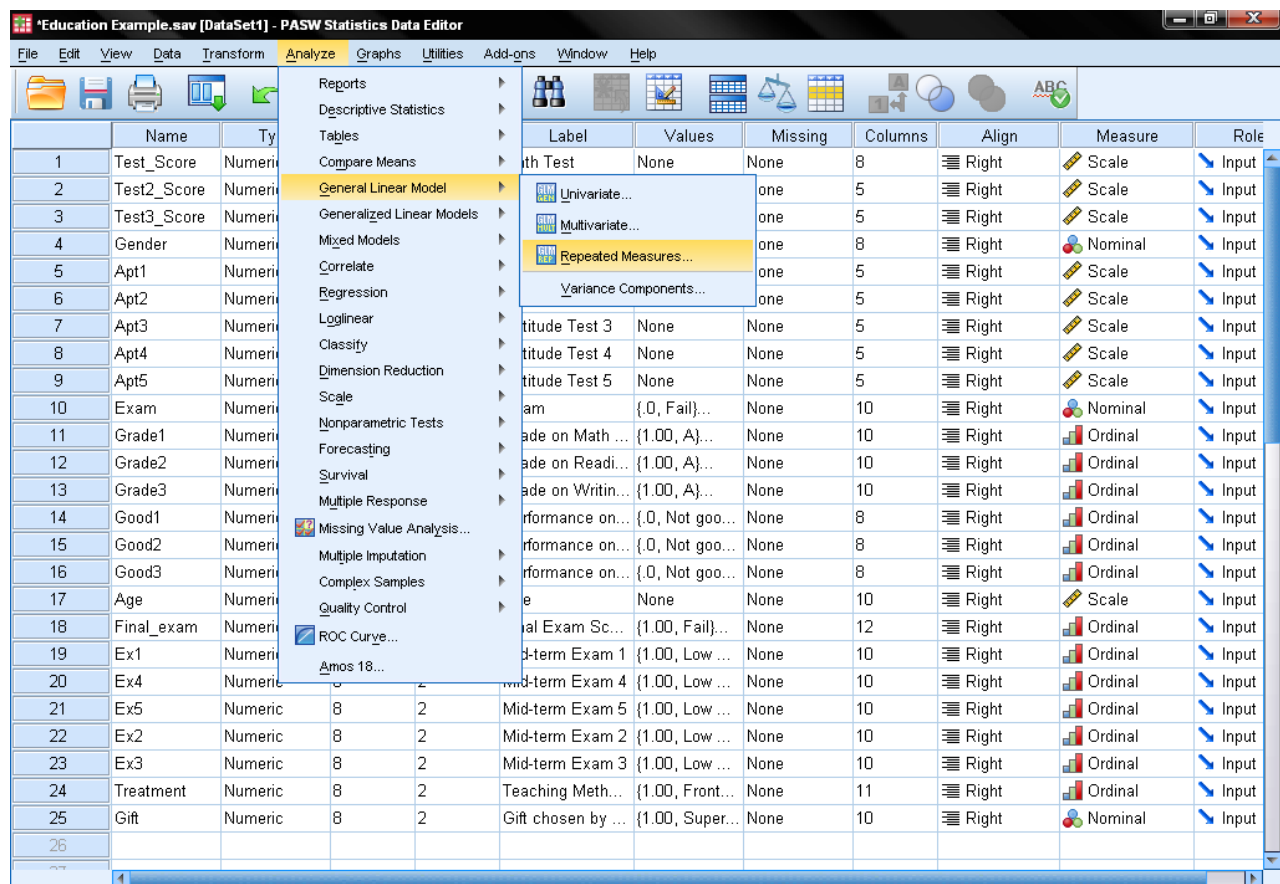
Since the pairing is explicitly defined and thus new information added to the data, paired data can always be analyzed with a regular ANCOVA, but not vice versa. The baseline differences, however, will not be accounted for.

### **The Repeated Measures ANCOVA in SPSS**

Consider the following research question:

*Are there individual differences in the longitudinal measures of aptitude between students who passed the final exam and the students who failed the final exam; when we control for the mathematical abilities as measured by the standardized test score of the math test?*

The repeated measures ANCOVA uses the GLM module of SPSS, like the factorial ANOVAs, MANOVAs, and MANCOVAs. The repeated measures ANCOVA can be found in SPSS in the menu *Analyze/General Linear Model/Repeated Measures...*



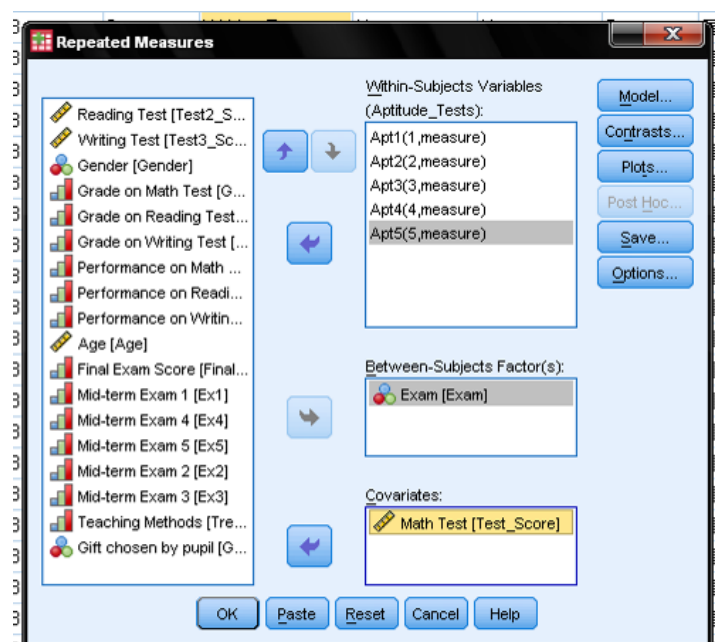
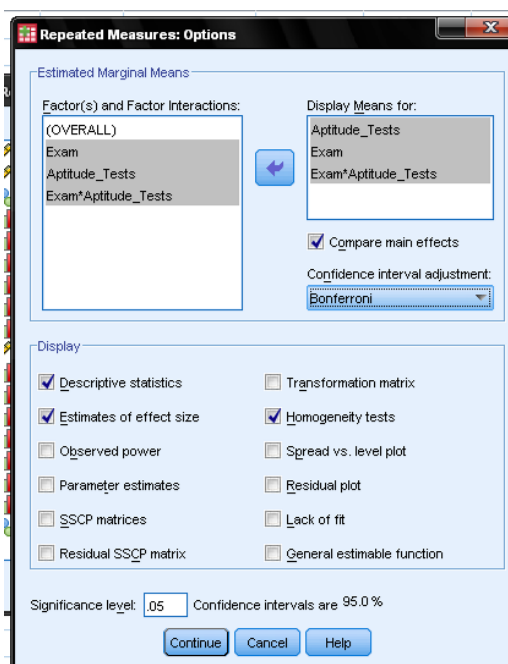
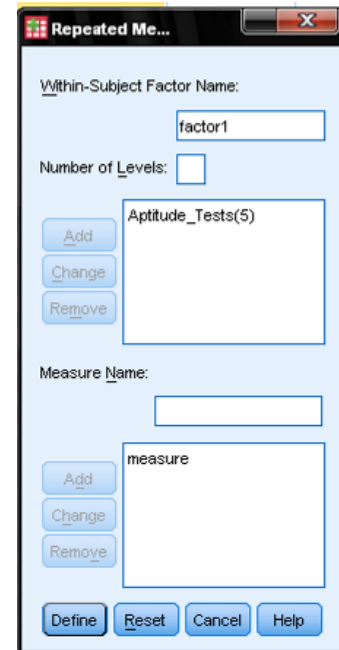
The dialog box that opens is different than the GLM module you might know from the MANCOVA. Before specifying the model we need to group the repeated measures.

This is done by creating a within-subject factor. It is called a within subject factor of our repeated measures ANCOVA because it represents the different observations of one subject. We measured the aptitude on five different data points, which creates five factor levels. We specify a factor called *Aptitude\_Tests* with five factor levels (that is the number of our repeated observations).

Since our research question also requires investigation as to the difference between the students who failed the final exam and the students who passed the final exam, we will include a measurement in the model.

The next dialog box allows us to specify the repeated measures ANCOVA. First we need to add the five observation points to the within-subject variables. Then, we need to add *Exam* (fail versus pass group of students) to the list of between-subject factors. Lastly, we add the results of the math test to the list of covariates.

As usual we go with the standard settings for the model, contrast, plots, and save the results. Also note that Post Hoc tests are disabled because of the inclusion of a covariate in the model.



We simply add some useful statistics to the repeated measures ANOVA output in the *Options...* dialog. These include the comparison of main effects with adjusted degrees of freedom, some descriptive



statistics, the practical significance  $\eta^2$ , and the Levene test for homoscedasticity since we included *Exam* as an independent variable in the analysis.

### The Output of the Repeated Measures ANCOVA

The first two tables simply list the design of the within-subject factor and the between-subject factors in our repeated measures ANCOVA.

**Within-Subjects Factors**

Measure:measure

Aptitude_Tests	Dependent Variable
1	Apt1
2	Apt2
3	Apt3
4	Apt4
5	Apt5

The second table of the repeated measures ANCOVA shows the descriptive statistics (mean, standard deviation, and sample size) for each cell in our analysis design.

**Between-Subjects Factors**

	Value Label	N
Exam Exam .00	Fail	64
1.00	Pass	43

Interestingly we see that the average aptitude

scores improve continuously for the students who failed the final exam and then drop again for the last test. And also we find that the number of students who pass the final exam starts off high, then drops by a large number and gradually increases.

**Descriptive Statistics**

	Exam Exam	Mean	Std. Deviation	N
Apt1 Aptitude Test 1	.00 Fail	25.36	8.868	64
	1.00 Pass	35.51	6.363	43
	Total	29.44	9.370	107
Apt2 Aptitude Test 2	.00 Fail	26.66	7.396	64
	1.00 Pass	21.72	7.688	43
	Total	24.67	7.864	107
Apt3 Aptitude Test 3	.00 Fail	27.13	6.482	64
	1.00 Pass	25.00	7.653	43
	Total	26.27	7.020	107
Apt4 Aptitude Test 4	.00 Fail	32.28	5.742	64
	1.00 Pass	29.14	6.205	43
	Total	31.02	6.103	107
Apt5 Aptitude Test 5	.00 Fail	30.87	8.109	64
	1.00 Pass	27.42	7.028	43
	Total	29.49	7.846	107

The next table includes the results of Box's M test, which verifies the assumption that covariance matrices of

**Box's Test of Equality of Covariance Matrices<sup>a</sup>**

Box's M	23.747
F	1.498
df1	15
df2	32481.356
Sig.	.096

each cell in our design are equal. Box's M is not significant ( $p=.096$ ) thus we can reject the null hypothesis that the covariance structures are equal and we can assume homogenous covariance structures.

The next table shows the results of the regression modeling the GLM procedure conducts. Regression is used to test the factor effects of significance. The analysis finds that the aptitude tests do not have a significant influence in the covariate regression model – that is, we cannot reject the null hypothesis that the mean scores are equal across observations. We find only that the interaction of the repeated tests with the independent variable (exam) is significant.

**Multivariate Tests<sup>b</sup>**

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Aptitude_Tests	Pillai's Trace	.046	1.225 <sup>a</sup>	4.000	101.000	.305	.046
	Wilks' Lambda	.954	1.225 <sup>a</sup>	4.000	101.000	.305	.046
	Hotelling's Trace	.049	1.225 <sup>a</sup>	4.000	101.000	.305	.046
	Roy's Largest Root	.049	1.225 <sup>a</sup>	4.000	101.000	.305	.046
Aptitude_Tests * Test_Score	Pillai's Trace	.014	.369 <sup>a</sup>	4.000	101.000	.831	.014
	Wilks' Lambda	.986	.369 <sup>a</sup>	4.000	101.000	.831	.014
	Hotelling's Trace	.015	.369 <sup>a</sup>	4.000	101.000	.831	.014
	Roy's Largest Root	.015	.369 <sup>a</sup>	4.000	101.000	.831	.014
Aptitude_Tests * Exam	Pillai's Trace	.098	2.730 <sup>a</sup>	4.000	101.000	.033	.098
	Wilks' Lambda	.902	2.730 <sup>a</sup>	4.000	101.000	.033	.098
	Hotelling's Trace	.108	2.730 <sup>a</sup>	4.000	101.000	.033	.098
	Roy's Largest Root	.108	2.730 <sup>a</sup>	4.000	101.000	.033	.098

a. Exact statistic

b. Design: Intercept + Test\_Score + Exam  
Within Subjects Design: Aptitude\_Tests

One of the special assumptions of repeated designs is sphericity. Sphericity is a measure for the structure of the covariance matrix in repeated designs. Because repeated designs violate the assumption of independence between measurements, the covariances need to be spheric. One stricter form of sphericity is compound symmetry, which occurs if all the covariances are approximately equal and all the variances are approximately equal in the samples. Mauchly's sphericity test tests this assumption. If there is no sphericity in the data the repeated measures ANOVA can still be done when the F-values are corrected by deducting additional degrees of freedom (e.g., Greenhouse-Geisser or Huynh-Feldt).

Mauchly's Test analyzes whether this assumption is fulfilled. It tests the null hypothesis that the error covariance of the orthonormalized transformed dependent variable is proportional to an identity matrix. In simpler terms, the relationship between different observation points is similar; the differences between the observations have equal variances. This assumption is similar to homoscedasticity (tested by the Levene Test) which assumes equal variances between groups, not observations.

**Mauchly's Test of Sphericity<sup>b</sup>**

Measure:measure

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon <sup>a</sup>		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Aptitude_Tests	.644	45.085	9	.000	.801	.845	.250

In our example, the assumption of sphericity has not been met, because the Mauchly's Test is highly significant. This means that the F-values of our repeated measures ANOVA are likely to be too large. This can be corrected by decreasing the degrees of freedom used to calculate F. The last three columns

(Epsilon) tell us the appropriate correction method to use. If epsilon is greater than 0.75 we should use the Huynh-Feldt correction or the Greenhouse-Geisser correction. SPSS automatically includes the corrected F-values in the f-statistics table of the repeated measures ANCOVA.

The next table shows the f-statistics. As discussed earlier the assumption of sphericity has not been met and thus the degrees of freedom in our repeated measures ANCOVA need to be decreased using the Huynh-Feldt correction. The results show that neither the aptitude test scores nor the interaction effect of the aptitude scores with the covariate factor are significantly different. The only significant factor in the model is the interaction between the repeated measures of the aptitude scores and the independent variable (exam) on a level of  $p = 0.005$ .

**Tests of Within-Subjects Effects**

Measure:measure

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Aptitude_Tests	Sphericity Assumed	258.506	4	64.627	1.140	.337	.011
	Greenhouse-Geisser	258.506	3.204	80.690	1.140	.334	.011
	Huynh-Feldt	258.506	3.381	76.458	1.140	.335	.011
	Lower-bound	258.506	1.000	258.506	1.140	.288	.011
Aptitude_Tests * Test_Score	Sphericity Assumed	110.971	4	27.743	.489	.743	.005
	Greenhouse-Geisser	110.971	3.204	34.638	.489	.702	.005
	Huynh-Feldt	110.971	3.381	32.822	.489	.712	.005
	Lower-bound	110.971	1.000	110.971	.489	.486	.005
Aptitude_Tests * Exam	Sphericity Assumed	928.778	4	232.194	4.097	.003	.038
	Greenhouse-Geisser	928.778	3.204	289.909	4.097	.006	.038
	Huynh-Feldt	928.778	3.381	274.703	4.097	.005	.038
	Lower-bound	928.778	1.000	928.778	4.097	.046	.038
Error(Aptitude_Tests)	Sphericity Assumed	23578.467	416	56.679			
	Greenhouse-Geisser	23578.467	333.184	70.767			
	Huynh-Feldt	23578.467	351.627	67.055			
	Lower-bound	23578.467	104.000	226.716			

Thus we cannot reject the null hypothesis that the repeated measures are equal when controlling for the covariate and we might unfortunately not assume that our repeated measures are different from each other when controlling for the covariate.

The last two tables we reviewed ran a global F-test. The next tables look at individual differences between subjects and measurements. First, the Levene tests is not significant for all repeated measures but the first one, thus we cannot reject our null hypothesis and might assume equal variances in all cells of our design. Secondly, we find that in our linear repeated measures ANCOVA model the covariate factor levels (Test\_Score) are not significantly different ( $p=0.806$ ), and also that the exam factor levels (pass vs. fail) are not significantly different ( $p=0.577$ ).

**Levene's Test of Equality of Error Variances<sup>a</sup>**

	F	df1	df2	Sig.
Apt1 Aptitude Test 1	5.638	1	105	.019
Apt2 Aptitude Test 2	.014	1	105	.906
Apt3 Aptitude Test 3	1.119	1	105	.293
Apt4 Aptitude Test 4	.137	1	105	.712
Apt5 Aptitude Test 5	.726	1	105	.396

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Test\_Score + Exam  
Within Subjects Design: Aptitude\_Tests

**Tests of Between-Subjects Effects**

Measure:measure  
Transformed Variable:Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	14685.976	1	14685.976	390.484	.000	.790
Test_Score	2.290	1	2.290	.061	.806	.001
Exam	11.802	1	11.802	.314	.577	.003
Error	3911.405	104	37.610			

The last output of our repeated measures ANCOVA are the pairwise tests. The pairwise comparisons between groups is meaningless since they are not globally different to begin with; the interesting table is the pairwise comparison of observations. It is here where we find that in our ANCOVA model test, 1, 2, and 3 differ significantly from each other, as well as 2 and 3 compared to 4 and 5, when controlling for the covariate.

### Pairwise Comparisons

Measure:measure

(I) Aptitude_Tests	(J) Aptitude_Tests	Mean Difference (I-J)	Std. Error	Sig. <sup>a</sup>	95% Confidence Interval for Difference <sup>a</sup>	
					Lower Bound	Upper Bound
1	2	6.065 <sup>*</sup>	1.335	.000	2.237	9.893
	3	4.244 <sup>*</sup>	1.097	.002	1.098	7.390
	4	-.610	1.145	1.000	-3.894	2.675
	5	.962	1.352	1.000	-2.915	4.839
2	1	-6.065 <sup>*</sup>	1.335	.000	-9.893	-2.237
	3	-1.821	.984	.670	-4.643	1.001
	4	-6.674 <sup>*</sup>	.961	.000	-9.430	-3.919
	5	-5.103 <sup>*</sup>	.927	.000	-7.761	-2.445
3	1	-4.244 <sup>*</sup>	1.097	.002	-7.390	-1.098
	2	1.821	.984	.670	-1.001	4.643
	4	-4.853 <sup>*</sup>	1.004	.000	-7.732	-1.974
	5	-3.282 <sup>*</sup>	1.095	.034	-6.421	-.142
4	1	.610	1.145	1.000	-2.675	3.894
	2	6.674 <sup>*</sup>	.961	.000	3.919	9.430
	3	4.853 <sup>*</sup>	1.004	.000	1.974	7.732
	5	1.572	.866	.725	-.913	4.056
5	1	-.962	1.352	1.000	-4.839	2.915
	2	5.103 <sup>*</sup>	.927	.000	2.445	7.761
	3	3.282 <sup>*</sup>	1.095	.034	.142	6.421
	4	-1.572	.866	.725	-4.056	.913

Based on estimated marginal means

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

a. Adjustment for multiple comparisons: Bonferroni.

In summary, we can conclude:

*During the fieldwork five repeated aptitude tests were administered to the students. We analyzed whether the differences between the five repeated measures are significant and whether they are significant between the students who passed the final exam and the students who failed the final exam when we controlled for their mathematical ability as measured by the standardized math test. The repeated measures ANCOVA shows that the achieved aptitude scores are not significant between the repeated measures and between the groups of students. However a pairwise comparison identifies the aptitude tests 1, 2, 3 still being significantly different from each other, when controlling for the students' mathematical abilities.*

## Profile Analysis

### What is the Profile Analysis?

Profile Analysis is mainly concerned with test scores, more specifically with profiles of test scores. Why is that relevant? Tests are commonly administered in medicine, psychology, and education studies to rank participants of a study. A profile shows differences in scores on the test. If a psychologist

administers a personality test (e.g., NEO), the respondent gets a test profile in return showing the scores on the Neuroticism, Extraversion, Agreeableness, Conscientiousness, and Openness dimensions. Similarly, many tests such as GMAT, GRE, SAT, and various intelligence questionnaires report profiles for abilities in reading, writing, calculating, and critically reasoning.

Typically test scores are used to predict behavioral items. In education studies it is common to predict test performance, for example using the SAT to predict the college GPA when graduating. Cluster analysis and Q-test have been widely used to build predictive models for this purpose.

What is the purpose of Profile Analysis? Profile Analysis helps researchers to identify whether two or more groups of test takers show up as a significantly distinct profile. It helps to analyze patterns of tests, subtests, or scores. The analysis may be across groups or across scores for one individual.

What does that mean? The profile analysis looks at profile graphs. A profile graph is simply the mean score of the one group of test takers with the other group of test takers along all items in the battery. The main purpose of the profile analysis is to identify how good a test is. Typically the tests consist of multiple item measurements and are administered over a series of time points. You could use a simple ANOVA to compare the test items, but this violates the independence assumption in two very important ways. Firstly, the scores on each item are not independent – item batteries are deliberately designed to have a high correlation among each other. Secondly, if you design a test to predict group membership (e.g., depressed vs. not depressed, likely to succeed vs. not like to succeed in college), you want the item battery to best predict the outcome. Thus item battery and group membership are also not independent.

What is the solution to this problem? Since neither the single measurements on the items nor the group membership are independent, they needed to be treated as a paired sample. Statistically the Profile Analysis is similar to a repeated measures ANOVA.

Example:

*A research team wants to create a new test for a form of cancer that seems to present in patients with a very specific history and diet. The researchers collect data on ten questions from patients that present with the cancer and a randomly drawn sample of people who do not present with the cancer.*

Profile Analysis is now used to check whether the ten questions significantly differentiate between the groups that presents with the illness and the group that does not. Profile analysis takes into account that neither items among each other nor subject assignment to groups is random.

Profile Analysis is also a great way to understand and explore complex data. The results of the profile analysis help to identify and focus on the relevant differences and help the researcher to select the right contrasts, post hoc analysis, and statistical tests when a simple ANOVA or t-test would not suffice. However profile analysis has its limitations, especially when it comes to standard error of measurement and predicting a single person's score.

Alternatives to the Profile Analysis are the Multidimensional Scaling, and Q-Analysis. In Q-Analysis the scores of an individual on the item battery are treated as an independent block (just as in Profile Analysis). The Q-Analysis then conducts a rotated factor analysis on these blocks, extracting relevant factors and flagging the items that define a factor.

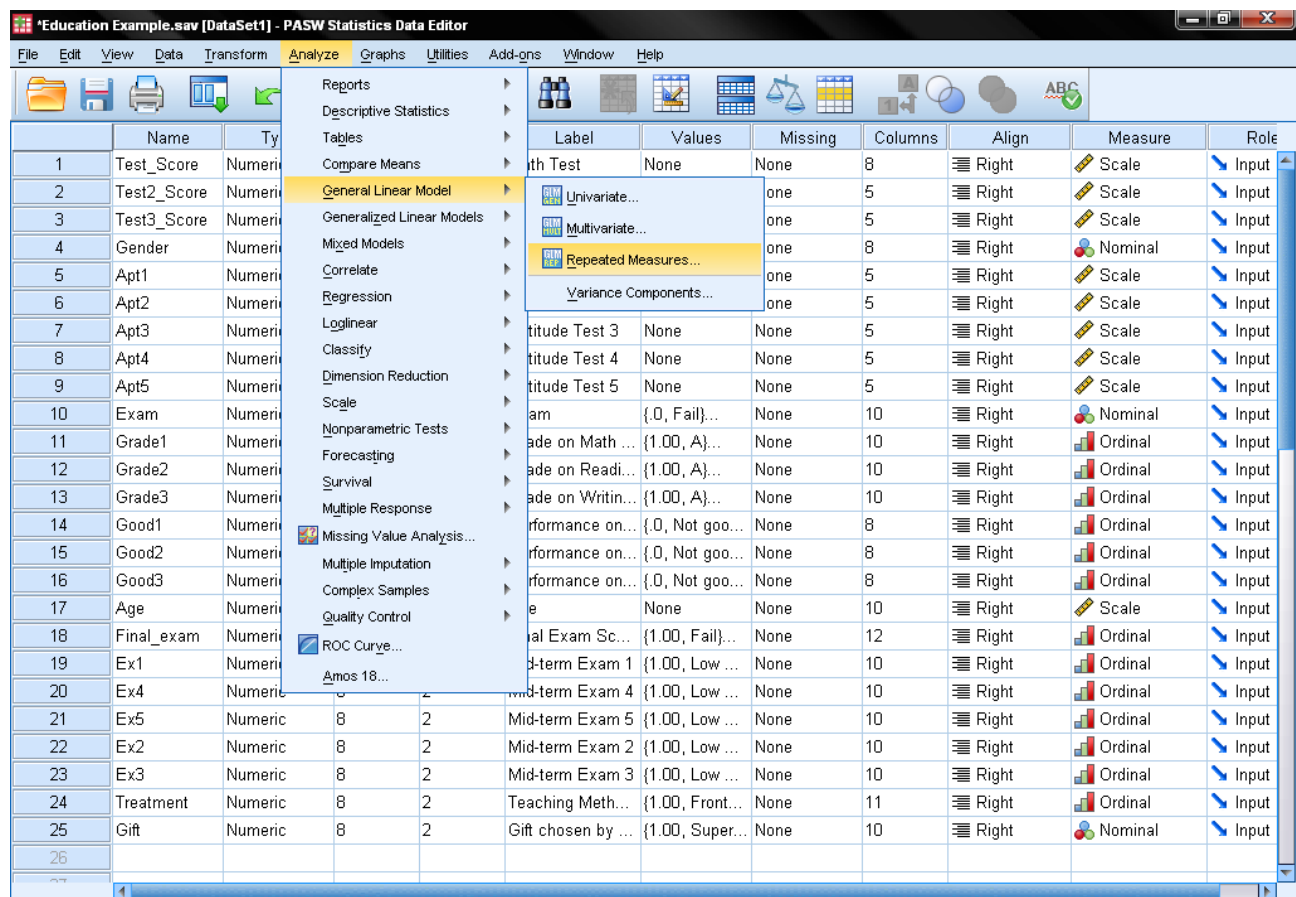
Another alternative to Profile Analysis is a two-way MANOVA (or doubly MANOVA). In this design the repeated measures would enter the model as the second dependent variable and thus the model elegantly circumvents the sphericity assumption.

## The Profile Analysis in SPSS

The research question we will examine for the Profile Analysis is as follows:

*Do the students who passed the final exam and the students who failed the final exam have a significantly different ranking in their math, reading, and writing test?*

The Profile Analysis uses the repeated measures GLM module of SPSS, like the repeated measures ANOVA and ANCOVA. The Profile Analysis can be found in SPSS in the menu *Analyze/General Linear Model/Repeated Measures...*



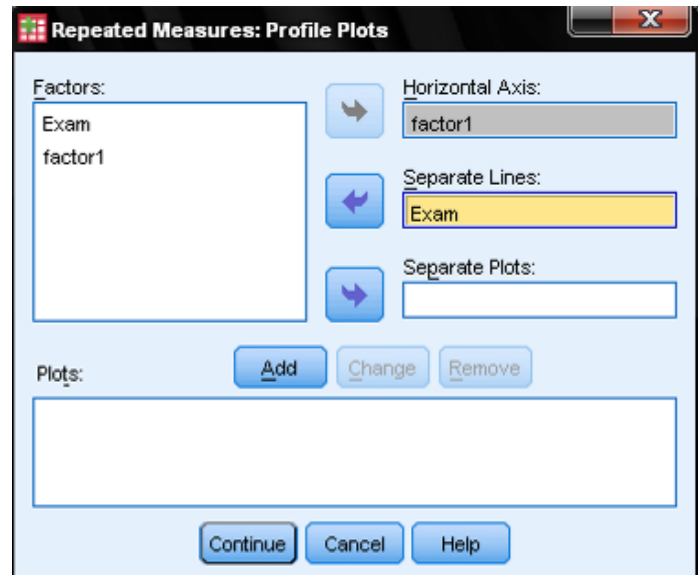
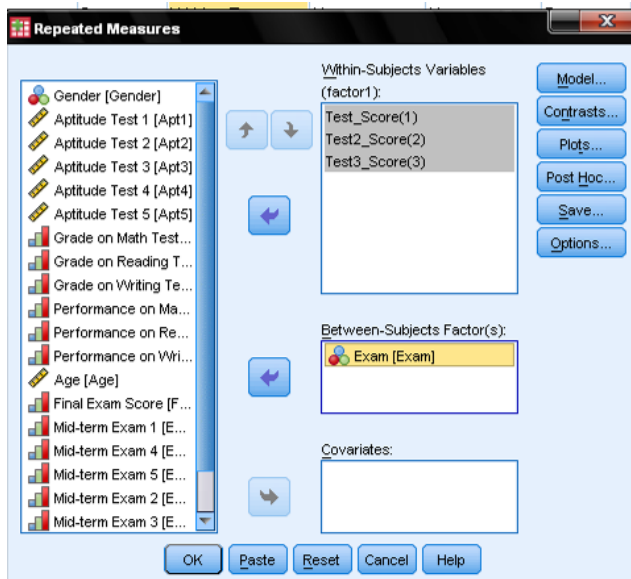
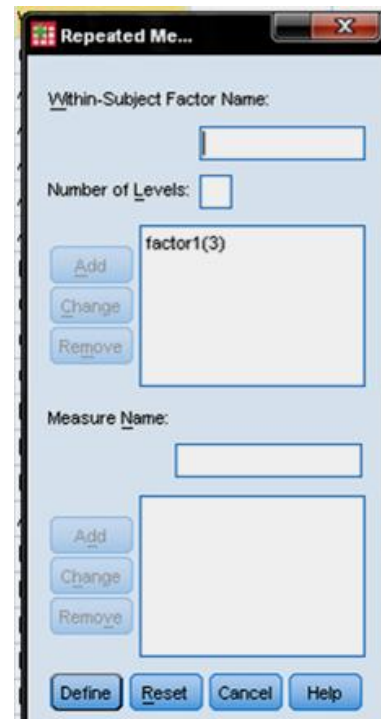
The dialog box that opens is different than the GLM module for independent measures. Before specifying the model we need to group the repeated measures—the item battery we want to test. In

our example we want to test if the standardized test, which consists of three items (math, reading, writing), correctly classifies the two groups of students that either pass or fail the final exam.

This is done by creating a within-subject factor. The item battery is called the within-subject factor of our Profile Analysis, because it represents the different observations of one subject. Our item battery contains three items – one score for math, one for reading, and one for writing. Thus we create and add a factor labeled *factor1* with three factor levels.

The next dialog box allows us to specify the Profile Analysis. First we need to add the three test items to the list of within-subjects variables. We then add the exam variable to the list of between-subjects factors. We can leave all other settings on default, apart from the plots.

To create the profile plots we want the items (or subtests) on the horizontal axis with the groups as separate lines. We also need the Levene test for homoscedasticity to check the assumptions of the Profile Analysis, the Levene Test can be included in the dialog *Options...*





## The Output of the Profile Analysis

The first tables of the output list the design of the within-subject factor and the between-subject factors in our Profile Analysis. These tables simply document our design.

Within-Subjects Factors		Box's Test of Equality of Covariance Matrices <sup>a</sup>		
Measure: MEASURE_1		Box's M	33.056	
factor1	Dependent Variable	F	5.329	
1	Test_Score	df1	6	
2	Test2_Score	df2	54751.383	
3	Test3_Score	Sig.	.000	

Between-Subjects Factors				
		Value Label	N	
Exam	Exam	.00	Fail	64
		1.00	Pass	43

Box's M Test verifies the assumption that covariance matrices of each cell in our Profile Analysis design are equal. Box's M is significant ( $p < 0.001$ ) thus we can reject the null hypothesis and we might not assume homogenous covariance structures. Also we can verify that sphericity might not be assumed, since the Mauchly's Test is significant ( $p = 0.003$ ). Sphericity is a measure for the structure of the covariance matrix in repeated designs. Because repeated designs violate the assumption of independence between measurements the covariances need to be spheric. One stricter form of sphericity is compound symmetry, which occurs if all the covariances are approximately equal and all the variances are approximately equal in the samples. Mauchly's sphericity test tests this assumption. If there is no sphericity in the data, the repeated measures ANOVA can still be done when the F-values are corrected by deducting additional degrees of freedom (e.g., Greenhouse-Geisser or Huynh-Feldt). Thus we need to correct the F-values when testing the significance of the main and interaction effects. The epsilon is greater than 0.75, thus we can work with the less conservative Huynh-Feldt correction.

**Mauchly's Test of Sphericity<sup>b</sup>**

Measure: MEASURE\_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon <sup>a</sup>		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
factor1	.893	11.822	2	.003	.903	.927	.500

The first real results table of our Profile Analysis are the within-subjects effects. This table shows that the items that build our standardized test are significantly different from each other and also that the interaction effect between passing the exam and the standardized test items is significant. However, the Profile Analysis does not tell how many items differ and in which direction they differ.

### Tests of Within-Subjects Effects

Measure: MEASURE\_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
factor1	Sphericity Assumed	90918.244	2	45459.122	726.708	.000
	Greenhouse-Geisser	90918.244	1.806	50343.849	726.708	.000
	Huynh-Feldt	90918.244	1.853	49060.921	726.708	.000
	Lower-bound	90918.244	1.000	90918.244	726.708	.000
factor1 * Exam	Sphericity Assumed	3523.441	2	1761.720	28.163	.000
	Greenhouse-Geisser	3523.441	1.806	1951.023	28.163	.000
	Huynh-Feldt	3523.441	1.853	1901.304	28.163	.000
	Lower-bound	3523.441	1.000	3523.441	28.163	.000
Error(factor1)	Sphericity Assumed	13136.516	210	62.555		
	Greenhouse-Geisser	13136.516	189.624	69.277		
	Huynh-Feldt	13136.516	194.583	67.511		
	Lower-bound	13136.516	105.000	125.110		

The Profile Analysis shows highly significant between-subjects effect. This indicates that the aptitude groups differ significantly on the average of all factor levels of the standardized test ( $p < 0.001$ ). We can conclude that the factor levels of the exam variable are

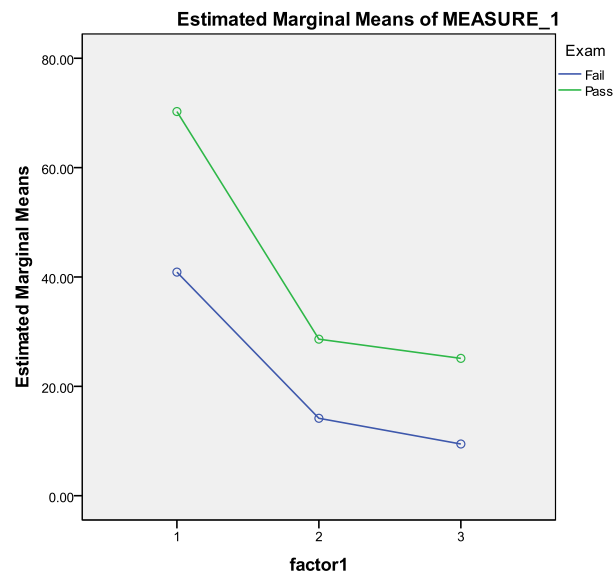
significantly different. However we cannot say which direction they differ (e.g., if failing the exam results in a lower score on the test). Also if we would have a grouping variable with more than two levels it would not tell whether all levels are significantly different or only a subset is different.

By far the most useful output of the Profile Analysis is the Profile Plot. The profile plot shows that the standardized test scores are consistently higher for the group of students that passed the exam. We could follow-up on this with a Covariate Analysis to identify the

### Tests of Between-Subjects Effects

Measure: MEASURE\_1  
Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	304625.396	1	304625.396	1696.612	.000
Exam	30351.302	1	30351.302	169.042	.000
Error	18852.667	105	179.549		



practical significance of the single items.

In summary, we may conclude as follows:

*We investigated whether the administered standardized test that measures the students' ability in math, reading, and writing can sufficiently predict the outcome of the final exam. We conducted a profile analysis and the profile of the two student groups is significantly different along all three dimensions of the standardized test, with students passing the exam scoring consistently higher.*

## **Double-Multivariate Profile Analysis**

### **What is the Double-Multivariate Profile Analysis?**

Double Multivariate Profile Analysis is very similar to the Profile Analysis. Profile Analyses are mainly concerned with test scores, more specifically with profiles of test scores. Why is that relevant? Tests are commonly administered in medicine, psychology, and education studies to rank participants of a study. A profile shows differences in scores on the test. If a psychologist administers a personality test (e.g., NEO), the respondent gets a test profile in return showing the scores on the Neuroticism, Extraversion, Agreeableness, Conscientiousness, and Openness dimensions. Similarly, many tests such as GMAT, GRE, SAT, and various intelligence questionnaires report profiles for abilities to read, write, calculate, and critical reasoning.

What is a double multivariate analysis? A double multivariate profile analysis (sometimes called doubly multivariate) is a multivariate profile analysis with more than one dependent variable. Dependent variables in Profile Analysis are the item batteries or subtests tested.

A Double Multivariate Profile Analysis can be double multivariate in two different ways: 1) two or more dependent variables are measured multiple times, or 2) two or more sets of non-commensurate measures are measured at once.

Let us first discuss the former, a set of multiple non-commensurate items that are measured two or more different times. Non-commensurate items are items with different scales. In such a case we have a group and a time, as well as an interaction effect group\*time. The double multivariate profile analysis will now estimate a linear canonical root that combines the dependent variables and maximizes the main and interaction effects. Now we can find out if the time or the group effect is significant and we can do simpler analyses to test the specific effects.

As for two or more sets of commensurate dependent variables of one subject measured one at a time, this could, for instance, be the level of reaction towards three different stimuli and the reaction time. Since both sets of measures are neither commensurate nor independent, we would need to use a double multivariate profile analysis. The results of that analysis will then tell us the main effects of our three stimuli, the reaction times, and the interaction effect between them. The double multivariate profile analysis will show which effects are significant and worth exploring in multivariate analysis with one dependent variable.

Additionally, the profile analysis looks at profile graphs. A profile graph simply depicts the mean scores of one group of test takers along the sets of measurements and compares them to the other groups of test takers along all items in the battery. Thus the main purpose of the profile analysis is to identify if non-independent measurements on two or more scales are significantly different between several groups of test takers.

Example:

*A research team wants to create a new test for a form of cardiovascular disease that seems to present in patients with a very specific combination of blood pressure, heart rate, cholesterol, and diet. The researchers collect data on these three dependent variables.*

Profile Analysis can then be used to check whether the three dependent variables differ significantly differentiate between a group that presents with the illness versus the group that does not. Profile analysis takes into account that neither items among each other nor subject assignment to groups is random.

Profile Analysis is also a great way to understand and explore complex data. The results of the profile analysis help to identify and focus on the relevant differences and help the researcher to select the right contrasts, post hoc analysis, and statistical tests, when a simple ANOVA or t-test would not be sufficient.

An alternative to profile analysis is also the double multivariate MANOVA, where the time and treatment effect are entered in a non-repeated measures MANOVA to circumvent the sphericity assumption on the repeated observations.

### ***The Double-Multivariate Profile Analysis in SPSS***

---

The Profile Analysis is statistically equivalent to a repeated measures MANOVA because the profile analysis compares mean scores in different samples across a series of repeated measurements that can either be the results of one test administered several times or subtests that make up the test.

The Double Multivariate Profile Analysis looks at profiles of data and checks a profile whether Profiles are significantly distinct in pattern and significantly different in level. Technically, Double Multivariate Profile analysis analyzes respondents as opposed to factor analysis which analyzes variables. At the same time, Double Multivariate Profile Analysis is different from cluster analysis in that cluster analysis does not take a dependent variable into account.

The purpose of Double Multivariate Profile Analysis is to check if four or more profiles are parallel. It tests four statistical hypotheses:

1. The centroids are equal;
2. The profiles are parallel (there is no interaction effect of group \* time);
3. Profiles are coincident (no group effects, given parallel profiles);
4. Profiles are level (no time effects, given parallel profiles).

In addition, the Double Multivariate Profile Analysis tests two practical hypotheses:

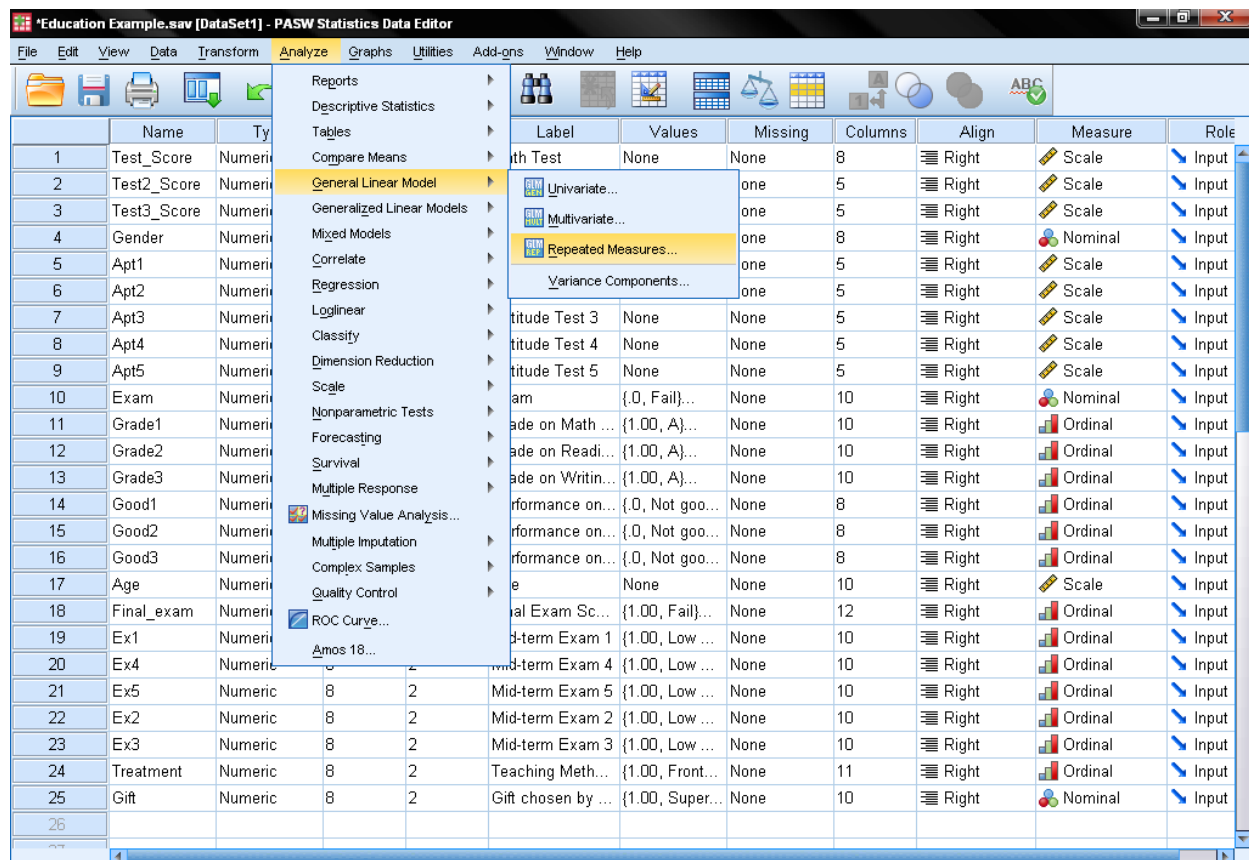
1. There are no within-subjects effects – the profile analysis tests whether the items within different batteries of subtests are significantly different, if items do not differ significantly they might be redundant and excluded.
2. There are no between-subjects effects – meaning that the subtest batteries do not produce different profiles for the groups.

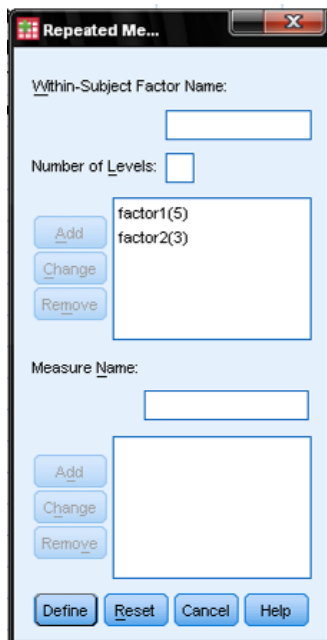
The profile analysis optimizes the covariance structure. The rationale behind using the covariance structure is that the observations are correlated and that the correlation of observations is naturally larger when they come from the same subject.

Our research question for the Doubly Multivariate Profile Analysis is as follows:

*Does the test profile for the five midyear mini-tests and snippets from the standardized tests (math, reading, and writing) differ between student who failed the final exam and students who passed the final exam? (The example is a  $3 \times 5 = (3 \text{ standardized test snippets}) \times (5 \text{ repeated mini-tests})$  design, thus the analysis requires 15 observations for each participant!)*

The Profile Analysis uses the repeated measures GLM module of SPSS, like the repeated measures ANOVA and ANCOVA. The Profile Analysis can be found in SPSS in the menu *Analyze/General Linear Model/Repeated Measures...*



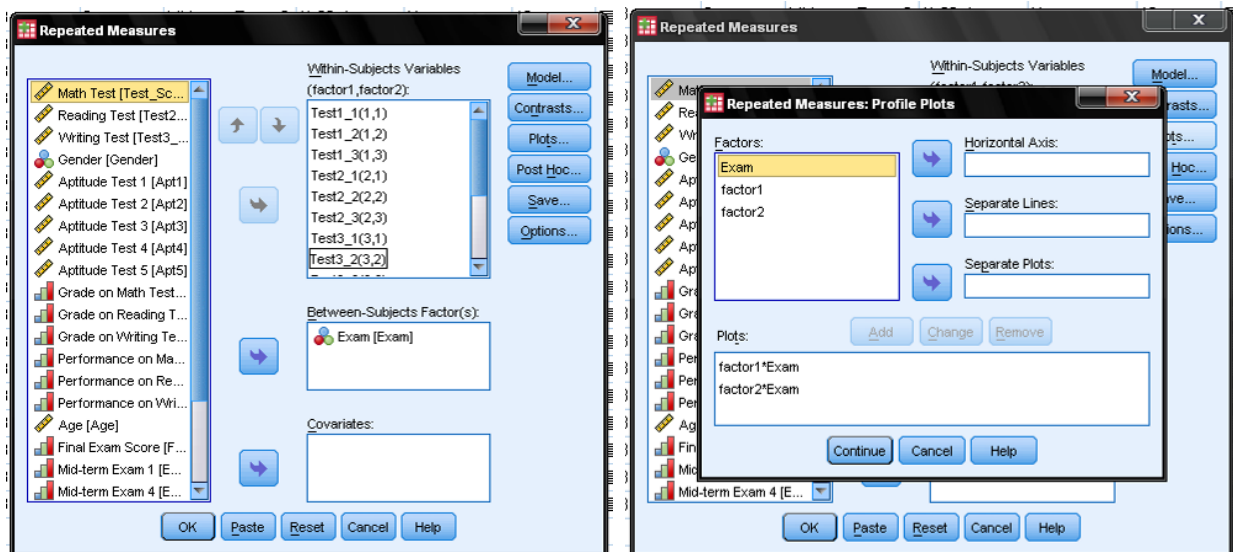


The dialog box that opens is different than the GLM module for independent measures. Before specifying the model we need to define the repeated measures, or rather, inform SPSS how we designed the study. In our example we want to test the three factor levels of standardized testing and the five factor levels of aptitude testing.

The factors are called within-subject factors of our Double Multivariate Profile Analysis because they represent the different observations of one subject, or within one subject. Thus we need to define and add two factors—one with three and one with five factor levels—to our design.

The next dialog box allows us to specify the Profile Analysis. We need to add all nested observation to the list of within-subject variables. Every factor level is explicitly marked. The first factor level on both variables is (1,1) and (3,2) is both the third level on the first factor and the second level on the second factor. Remember that this analysis needs 3x5 data points per participant! We leave all other settings on default, apart from the plots

where we add the marginal means plots.



### **The Output of the Double-Multivariate Profile Analysis**

The first table simply lists the design of the within-subject factor and the between-subject factors in our Double Multivariate Profile Analysis. These tables document our design.

Box's M test would be typically the next result to examine. However SPSS finds singularity in covariance matrices (that is perfect correlation). Usually Box M verifies the assumption that covariance matrices of each cell in our Double Multivariate Profile Analysis design are equal. It does so by testing the null hypothesis that the covariance structures are homogenous.

### Within-Subjects Factors

Measure: MEASURE\_1

factor1	factor2	Dependent Variable
1	1	Test1_1
	2	Test1_2
	3	Test1_3
2	1	Test2_1
	2	Test2_2
	3	Test2_3
3	1	Test3_1
	2	Test3_2
	3	Test3_3
4	1	Test4_1
	2	Test4_2
	3	Test4_3
5	1	Test5_1
	2	Test5_2
	3	Test5_3

### Between-Subjects Factors

	Value Label	N
Exam Exam .00	Fail	64
1.00	Pass	43

### Warnings

Box's Test of Equality of Covariance Matrices is not computed because there are fewer than two nonsingular cell covariance matrices.

The next assumption to test is sphericity. In our Double Multivariate Profile Analysis sphericity can be assumed for the main effects (Mauchly's Test is not significant  $p=0.000$ ), since we cannot reject the null hypothesis (Mauchly's Test is highly significant). Thus we need to correct the F-values when testing the significance of the interaction effects. The estimated epsilon is less than 0.75, thus we need to work with the more conservative the Greenhouse-Geisser correction.

### Mauchly's Test of Sphericity<sup>b</sup>

Measure: MEASURE\_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon <sup>a</sup>		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
factor1	.209	162.075	9	.000	.659	.684	.250
factor2	.	.	2	.	.	.	.500
factor1 * factor2	.000	.	35	.	.372	.388	.125

The first results table of our Double Multivariate Profile Analysis reports the within-subjects effects. It shows that the within-subject effects of factor 1 (the mini-test we administered) and the standardized test bits are significantly different. However the sample questions for the standardized tests (factor 2) that were included in our mini-tests are not significantly different, because of covariance structure singularities. Additionally the interaction effects factor1 \* exam (group of students who passed vs. students who failed) is significant, as well as the two-way interaction between the factors and the three-way interaction between the factors and the outcome of the exam variable. This is a good indication

that we found distinct measurements and that we do not see redundancy in our measurement approach.

The next step in our Double Multivariate Profile Analysis tests the discriminating power of our groups. It will reveal whether or not the profiles of the groups are distinct and parallel. Before we test, however, we need to verify homoscedasticity. The Levene Test (below, right) prevents us from rejecting the null hypothesis that the variances are equal, thus we might assume homoscedasticity in almost all tests.

Tests of Within-Subjects Effects

Measure: MEASURE\_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
factor1	Sphericity Assumed	25362.013	4	6340.503	254.709	.000	.708
	Greenhouse-Geisser	25362.013	2.638	9614.365	254.709	.000	.708
	Huynh-Feldt	25362.013	2.738	9263.266	254.709	.000	.708
	Lower-bound	25362.013	1.000	25362.013	254.709	.000	.708
factor1 * Exam	Sphericity Assumed	1571.793	4	392.948	15.785	.000	.131
	Greenhouse-Geisser	1571.793	2.638	595.843	15.785	.000	.131
	Huynh-Feldt	1571.793	2.738	574.084	15.785	.000	.131
	Lower-bound	1571.793	1.000	1571.793	15.785	.000	.131
Error(factor1)	Sphericity Assumed	10455.128	420	24.893			
	Greenhouse-Geisser	10455.128	276.983	37.747			
	Huynh-Feldt	10455.128	287.481	36.368			
	Lower-bound	10455.128	105.000	99.573			
factor2	Sphericity Assumed	1028.785	2	514.393	.	.	1.000
	Greenhouse-Geisser	1028.785	.	.	.	.	1.000
	Huynh-Feldt	1028.785	.	.	.	.	1.000
	Lower-bound	1028.785	1.000	1028.785	.	.	1.000
factor2 * Exam	Sphericity Assumed	.000	2	.000	.	.	.
	Greenhouse-Geisser	.000	.	.	.	.	.
	Huynh-Feldt	.000	.	.	.	.	.
	Lower-bound	.000	1.000	.000	.	.	.
Error(factor2)	Sphericity Assumed	.000	210	.000			
	Greenhouse-Geisser	.000	.	.			
	Huynh-Feldt	.000	.	.			
	Lower-bound	.000	105.000	.000			
factor1 * factor2	Sphericity Assumed	9456.598	8	1182.075	16.382	.000	.135
	Greenhouse-Geisser	9456.598	2.976	3177.639	16.382	.000	.135
	Huynh-Feldt	9456.598	3.102	3049.015	16.382	.000	.135
	Lower-bound	9456.598	1.000	9456.598	16.382	.000	.135
factor1 * factor2 * Exam	Sphericity Assumed	10102.317	8	1262.790	17.500	.000	.143
	Greenhouse-Geisser	10102.317	2.976	3394.616	17.500	.000	.143
	Huynh-Feldt	10102.317	3.102	3257.209	17.500	.000	.143
	Lower-bound	10102.317	1.000	10102.317	17.500	.000	.143
Error(factor1*factor2)	Sphericity Assumed	60613.184	840	72.159			
	Greenhouse-Geisser	60613.184	312.478	193.976			
	Huynh-Feldt	60613.184	325.660	186.124			
	Lower-bound	60613.184	105.000	577.268			

Levene's Test of Equality of Error Variances<sup>a</sup>

	F	df1	df2	Sig.
Test1_1	5.946	1	105	.016
Test1_2	.014	1	105	.907
Test1_3	1.159	1	105	.284
Test2_1	.109	1	105	.742
Test2_2	.804	1	105	.372
Test2_3	5.946	1	105	.016
Test3_1	.014	1	105	.907
Test3_2	1.159	1	105	.284
Test3_3	.109	1	105	.742
Test4_1	.804	1	105	.372
Test4_2	5.946	1	105	.016
Test4_3	.014	1	105	.907
Test5_1	1.159	1	105	.284
Test5_2	.109	1	105	.742
Test5_3	.804	1	105	.372

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Exam

Within Subjects Design: factor1 + factor2 + factor1 \* factor2

Tests of Between-Subjects Effects

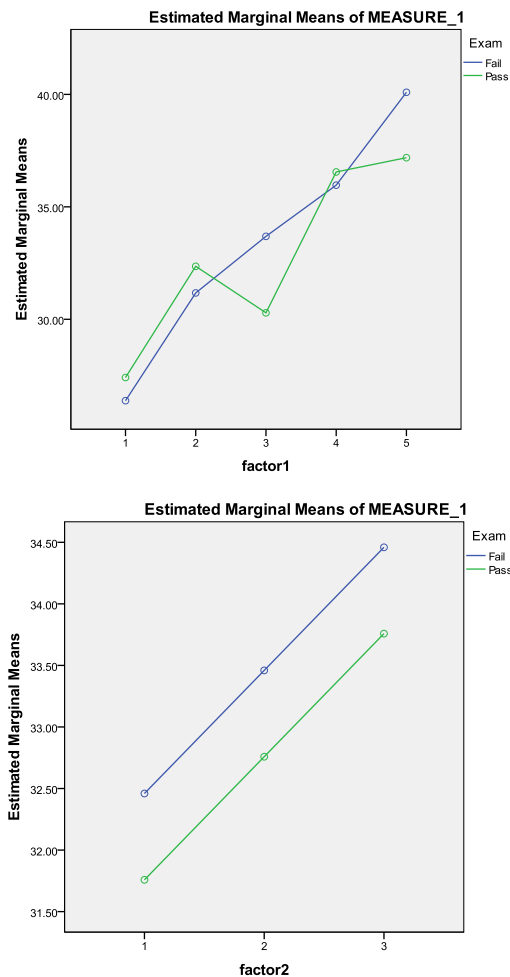
Measure: MEASURE\_1

Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	1691615.524	1	1691615.524	15128.042	.000	.993
Exam	189.707	1	189.707	1.697	.196	.016
Error	11741.085	105	111.820			



The Double Multivariate Profile Analysis shows highly significant between-subjects effect. This indicates that the student groups (defined by our external criterion of failing or passing the final exam) differ significantly across all factor levels ( $p = 0.016$ ). We can conclude that the factor levels of the tests are significantly different. However we cannot say which direction they differ, for example if the students that failed the final exam scored lower or not. Also, a grouping variable with more than two levels would not tell whether all levels are significantly different or if only a subset is different. The Profile Plots of the Double Multivariate Profile Analysis answer this question.



We find that for both repeated measures on the mini-tests and the sample questions from the standardized tests the double multivariate profiles are somewhat distinct—albeit more so for the standardized test questions. The students who failed the exam scored consistently lower than the student who passed the final exam.

In summary, a sample write-up would read as follows:

*We investigated whether five repeated mini-tests that included prospective new questions for the standardized test (three scores for math, reading, and writing) have significantly distinct profiles. The doubly multivariate profile analysis finds that the five mini-tests are significantly*

*different, and that the students who passed the exam are significantly different from the students that failed the exam on all scores measured (five mini-tests and three standardized test questions. Also all two- and three-way interaction effects are significant. However due to singularity in the covariance structures, the hypothesis could not be tested for the standardized test questions.*

## **Independent Sample T-Test**

### **What is the Independent Sample T-Test?**

The independent samples t-test is a member of the t-test family, which consists of tests that compare mean value(s) of continuous-level(interval or ratio data), normally distributed data. The independent samples t-test compares two means. It assumes a model where the variables in the analysis are split into independent and dependent variables. The model assumes that a difference in the mean score of the dependent variable is found because of the influence of the independent variable. Thus, the independent samples t-test is an analysis of dependence. It is one of the most widely used statistical tests, and is sometimes erroneously called the independent variable t-test.

The t-test family is based on the t-distribution, because the difference of mean score for two multivariate normal variables approximates the t-distribution. The t-distribution and also the t-test is sometimes also called Student's t. Student is the pseudonym used by W. S. Gosset in 1908 to publish the t-distribution based on his empirical findings on the height and the length of the left middle finger of criminals in a local prison.

Within the t-test family, the independent samples t-test compares the mean scores of two groups in a given variable, that is, two mean scores of the same variable, whereby one mean represents the average of that characteristic for one group and the other mean represents the average of that specific characteristic in the other group. Generally speaking, the independent samples t-test compares one measured characteristic between two groups of observations or measurements. It tells us whether the difference we see between the two independent samples is a true difference or whether it is just a random effect (statistical artifact) caused by skewed sampling.

The independent samples t-test is also called unpaired t-test. It is the t-test to use when two separate independent and identically distributed variables are measured. Independent samples are easiest obtained when selecting the participants by random sampling.

The independent samples t-test is similar to the dependent sample t-test, which compares the mean score of paired observations these are typically obtained when either re-testing or conducting repeated measurements, or when grouping similar participants in a treatment-control study to account for differences in baseline. However the pairing information needs to be present in the sample and therefore a paired sample can always be analyzed with an independent samples t-test but not the other way around.

Examples of typical questions that the independent samples t-test answers are as follows:

- **Medicine** - Has the quality of life improved for patients who took drug A as opposed to patients who took drug B?
- **Sociology** - Are men more satisfied with their jobs than women? Do they earn more?
- **Biology** - Are foxes in one specific habitat larger than in another?
- **Economics** - Is the economic growth of developing nations larger than the economic growth of the first world?
- **Marketing**: Does customer segment A spend more on groceries than customer segment B?

### **The Independent Sample T-Test in SPSS**

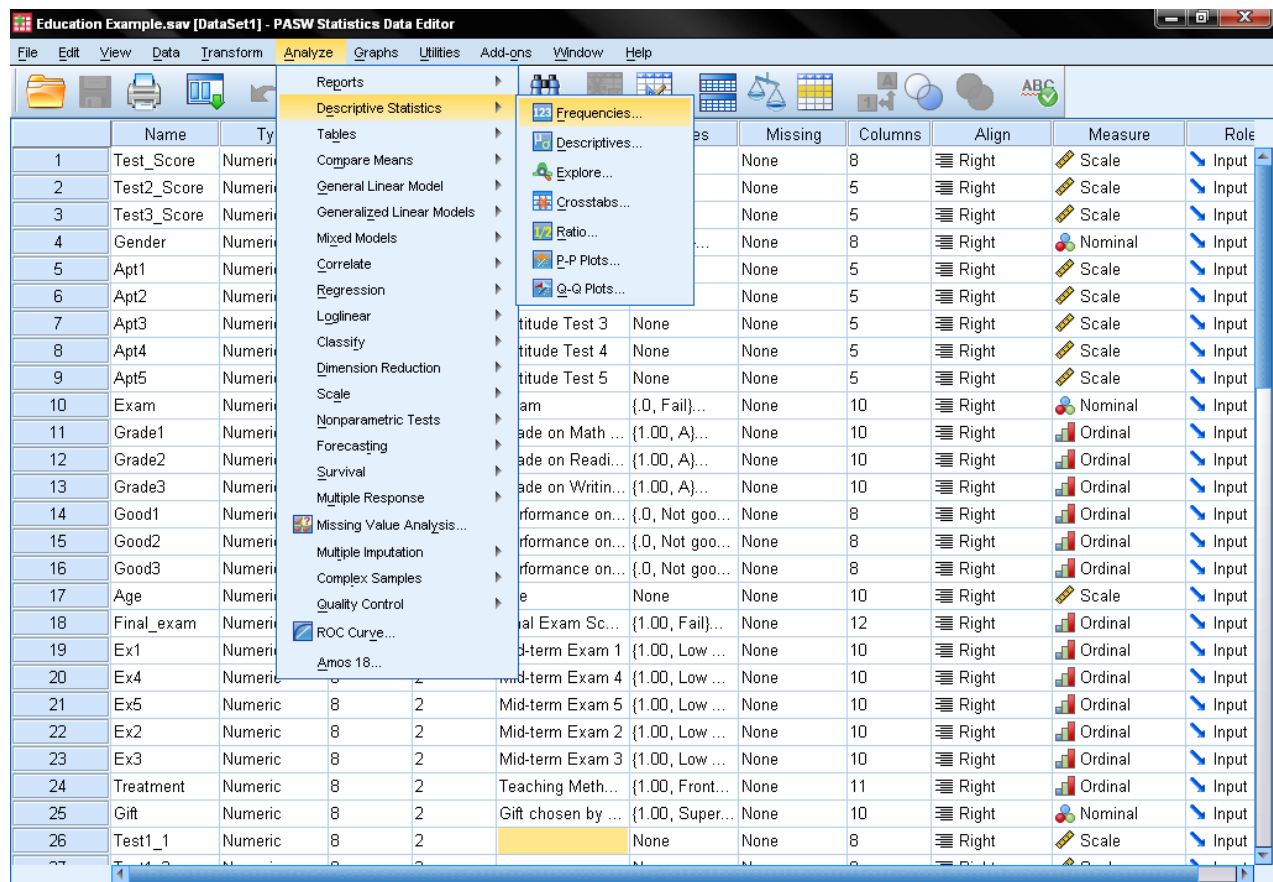
---

The independent samples t-test, or Student's t-test, is the most popular test to test for the difference in means. It requires that both samples are independently collected, and tests the null hypothesis that both samples are from the same population and therefore do not differ in their mean scores.

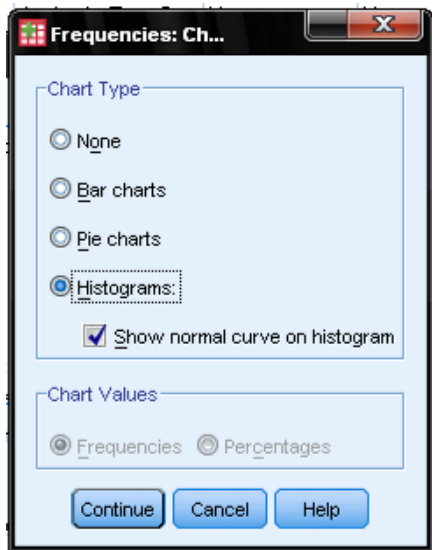
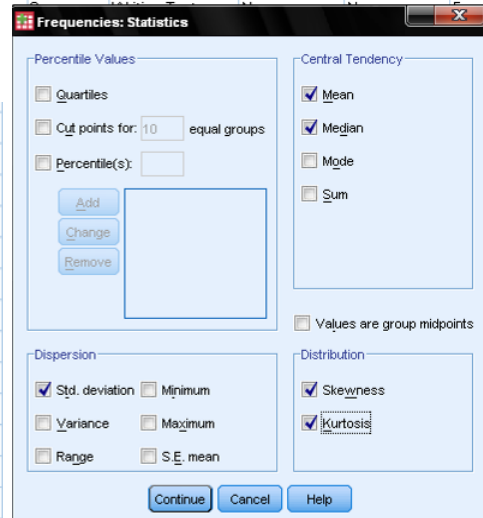
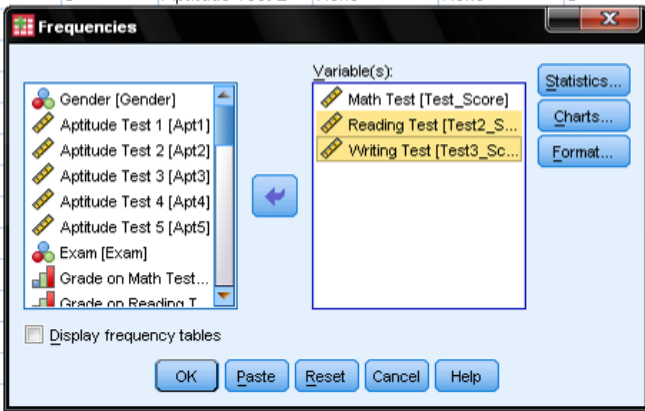
Our research question for the independent sample t-test is as follows:

*Does the standardized test score for math, reading, and writing differ between students who failed and students who passed the final exam?*

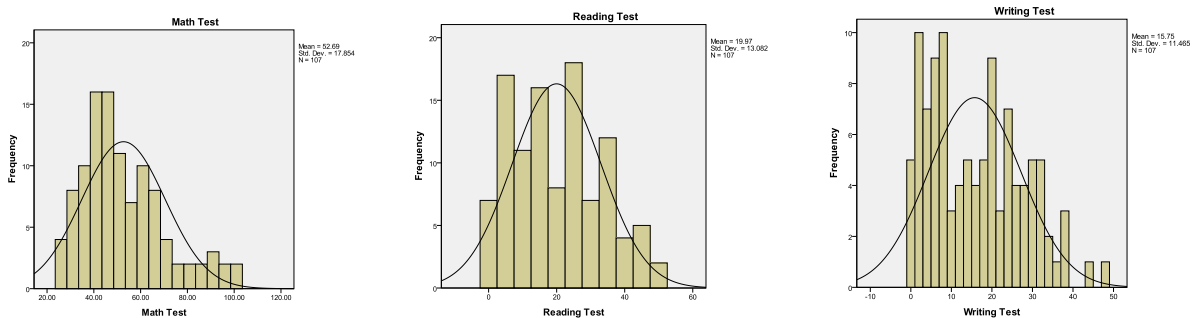
Let's start by verifying the assumptions of the t-test to check whether we made the right choices in our decision tree. First, we are going to create some descriptive statistics to get an impression of the distribution. In order to do this, we open the Frequencies menu in *Analyze/Descriptive Statistics/Frequencies...*



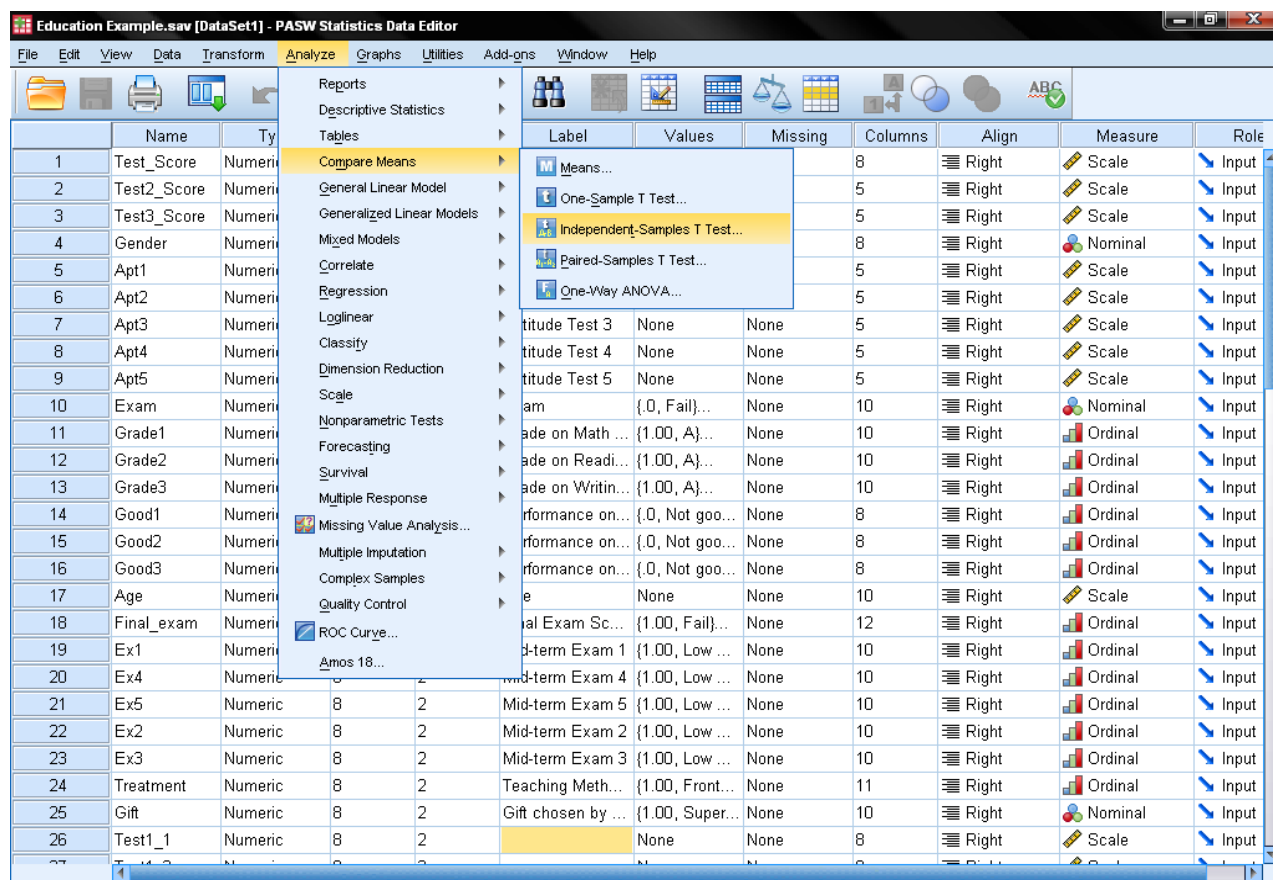
Next we add the two groups to the list of variables. For the moment our two groups are stored in the variable A and B. We deselect the frequency tables but add distribution parameters and the histograms with normal distribution curve to the output.



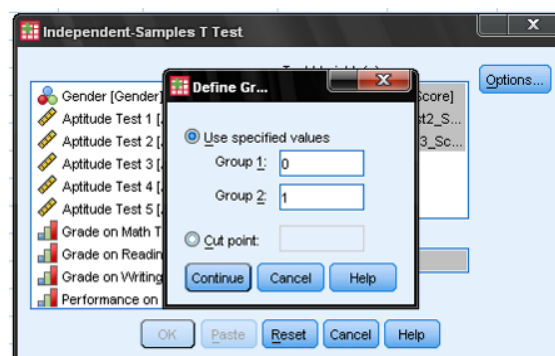
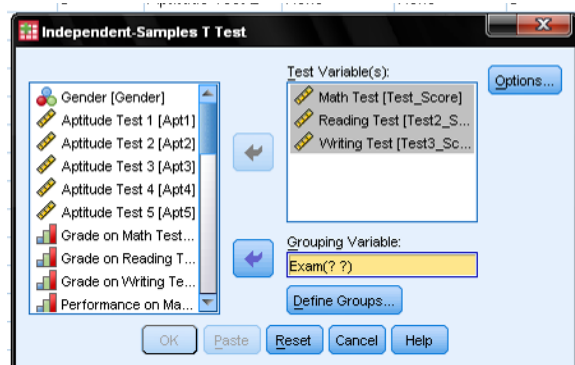
The histograms show quite nicely that the variables approximate a normal distribution and also their distributional difference. We could continue with verifying this 'eyeball' test with a K-S test, however because our sample is larger than 30, we will skip this step.



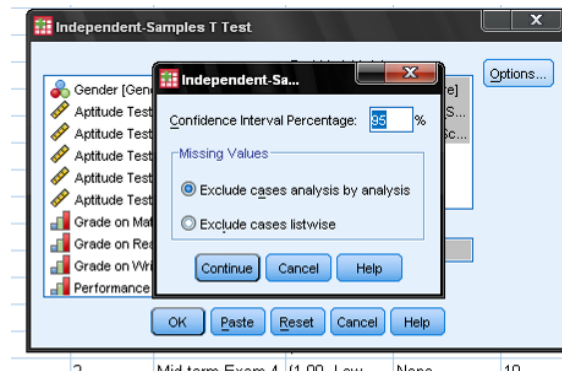
The independent samples t-test is found in *Analyze/Compare Means/Independent Samples T-Test*.



In the dialog box of the independent samples t-test we select the variable with our standardized test scores as the three test variables and the grouping variable is the outcome of the final exam (pass = 1 vs. fail = 0). The independent samples t-test can only compare two groups (if your independent variable defines more than two groups, you either would need to run multiple t-tests or an ANOVA with post hoc tests). The groups need to be defined upfront, for that you need to click on the button *Define Groups...* and enter the values of the independent variable that characterize the groups.



The dialog box *Options...* allows us to define how missing cases shall be managed (either exclude them listwise or analysis by analysis). We can also define the width of the confidence interval that is used to test the difference of the mean scores in this independent samples t-test.



### **The Output of the Independent Sample T-Test**

The output of the independent samples t-test consists of only two tables. The first table contains the information we saw before when we checked the distributions. We see that the students that passed the exam scored higher on average on all three tests. The table also displays the information you need in order to calculate the t-test manually.

**Group Statistics**

Exam Exam		N	Mean	Std. Deviation	Std. Error Mean
Test_Score Math Test	.00 Fail	64	40.8906	7.28581	.91073
	1.00 Pass	43	70.2558	14.02904	2.13941
Test2_Score Reading Test	.00 Fail	64	14.16	11.112	1.389
	1.00 Pass	43	28.63	10.885	1.660
Test3_Score Writing Test	.00 Fail	64	9.45	8.071	1.009
	1.00 Pass	43	25.12	9.155	1.396

We are not going to calculate the test manually because the second table nicely displays the results of the independent samples t-test. If you remember there is one question from the decision tree still left unanswered, and that is, *“Are the groups homoscedastic?”*

The output includes the Levene Test in the first two columns. The Levene test tests the null hypothesis that the variances are homogenous (equal) in each group of the independent variable. In our example it is highly significant for the math test and not significant for the writing and reading test. That is why we must reject the null hypothesis for the math test and assume that the variances are not equal for the math test. We cannot reject the null hypothesis for the reading and writing tests, so that we might assume that the variances of these test scores are equal between the groups of students who passed the exam and students who failed the final exam.

We find the correct results of the t-test next to it. For the math score we have to stick to the row *'Equal variances not assumed'* whereas for reading and writing we go with the *'Equal variances assumed'* row. We find that for all three test scores the differences are highly significant ( $p < 0.001$ ). The table also tells

us the 95% confidence intervals for the difference of the mean scores; none of the confidence intervals include zero. If they did, the t-test would not be significant and we would not find a significant difference between the groups of students.

Independent Samples Test										
		Levene's Test for Equality of Variances		t-Test for Equality of Means						95% Confidence Interval of the Difference
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Test_Score Math Test	Equal variances assumed	20.238	.000	-14.162	105	.000	-29.36519	2.07347	-33.47649	-25.25389
	Equal variances not assumed			-12.629	57.346	.000	-29.36519	2.32519	-34.02069	-24.70969
Test2_Score Reading Test	Equal variances assumed	.004	.953	-6.659	105	.000	-14.472	2.173	-18.781	-10.163
	Equal variances not assumed			-6.686	91.500	.000	-14.472	2.164	-18.771	-10.173
Test3_Score Writing Test	Equal variances assumed	.662	.418	-9.322	105	.000	-15.663	1.680	-18.995	-12.332
	Equal variances not assumed			-9.093	82.345	.000	-15.663	1.723	-19.090	-12.237

A possible write up could be as follows:

*We analyzed the standardized test scores for students who passed the final exam and students who failed the final exam. An independent samples t-test confirms that students who pass the exam score significantly higher on all three tests with  $p < 0.001$  ( $t = 12.629, 6.686, \text{ and } 9.322$ ). The independent samples t-test has shown that we can reject our null hypothesis that both samples have the same mean scores for math, reading, and writing.*

## One-Sample T-Test

### What is the One-Sample T-Test?

The 1-sample t-test is a member of the t-test family. All the tests in the t-test family compare differences in mean scores of continuous-level (interval or ratio), normally distributed data. The 1-sample t-test does compare the mean of a single sample. Unlike the other tests, the independent and dependent sample t-test it works with only one mean score.

The independent sample t-test compares one mean of a distinct group to the mean of another group from the same sample. It would examine the question, “*Are old people smaller than the rest of the population?*” The dependent sample t-test compares before/after measurements, like for example, “*Do pupils’ grades improve after they receive tutoring?*”

So if only a single mean is calculated from the sample what does the 1-sample t-test compare the mean with? The 1-sample t-test compares the mean score found in an observed sample to a hypothetically assumed value. Typically the hypothetically assumed value is the population mean or some other theoretically derived value.

There are some typical applications of the 1-sample t-test: 1) testing a sample against a pre-defined value, 2) testing a sample against an expected value, 3) testing a sample against common sense or expectations, and 4) testing the results of a replicated experiment against the original study.



First, the hypothetical mean score can be a generally assumed or pre-defined value. For example, a researcher wants to disprove that the average age of retiring is 65. The researcher would draw a representative sample of people entering retirement and collecting their ages when they did so. The 1-sample t-test compares the mean score obtained in the sample (e.g., 63) to the hypothetical test value of 65. The t-test analyzes whether the difference we find in our sample is just due to random effects of chance or if our sample mean differs systematically from the hypothesized value.

Secondly, the hypothetical mean score also can be some derived expected value. For instance, consider the example that the researcher observes a coin toss and notices that it is not completely random. The researcher would measure multiple coin tosses, assign one side of the coin a 0 and the flip side a 1. The researcher would then conduct a 1-sample t-test to establish whether the mean of the coin tosses is really 0.5 as expected by the laws of chance.

Thirdly, the 1-sample t-test can also be used to test for the difference against a commonly established and well known mean value. For instance a researcher might suspect that the village she was born in is more intelligent than the rest of the country. She therefore collects IQ scores in her home village and uses the 1-sample t-test to test whether the observed IQ score differs from the defined mean value of 100 in the population.

Lastly, the 1-sample t-test can be used to compare the results of a replicated experiment or research analysis. In such a case the hypothesized value would be the previously reported mean score. The new sample can be checked against this mean value. However, if the standard deviation of the first measurement is known a proper 2-sample t-test can be conducted, because the pooled standard deviation can be calculated if the standard deviations and mean scores of both samples are known.

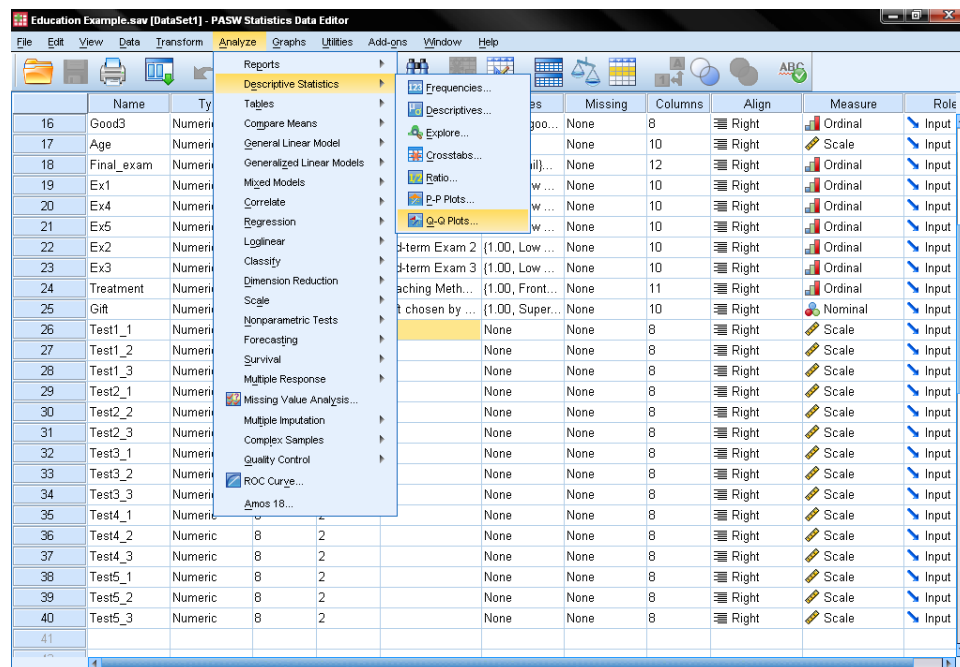
Although the 1-sample t-test is mathematically the twin brother of the independent variable t-test, the interpretation is somewhat different. The 1-sample t-test checks whether the mean score in a sample is a certain value, the independent sample t-test checks whether an estimated coefficient is different from zero.

---

### ***The One-Sample T-Test in SPSS***

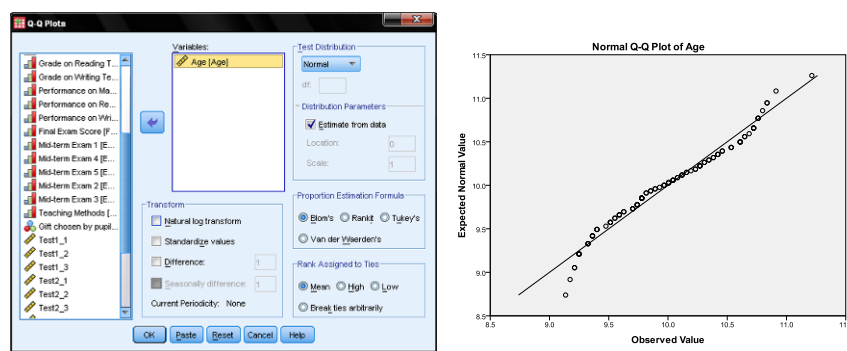
The 1-sample t-test does compare the mean of a single sample. Unlike the independent and dependent sample t-test, the 1-sample t-test works with only one mean score. The 1-sample t-test compares the mean score found in an observed sample to a hypothetically assumed value. Typically the hypothetically assumed value is the population mean or some other theoretically derived value.

The statement we will examine for the 1-sample t-test is as follows: *The average age in our student sample is 9½ years.*



Before we actually conduct the 1-sample t-test, our first step is to check the distribution for normality. This is best done with a Q-Q Plot. We simply add the variable we want to test (age) to the box and confirm that the test distribution is set to *Normal*. This will create the diagram you see below. The output

shows that small values and large values somewhat deviate from normality. As a check we can run a K-S Test to tests the null hypothesis that the variable is normally distributed. We find that the K-S Test is not significant thus we cannot reject  $H_0$  and we might assume that the variable age is normally distributed.



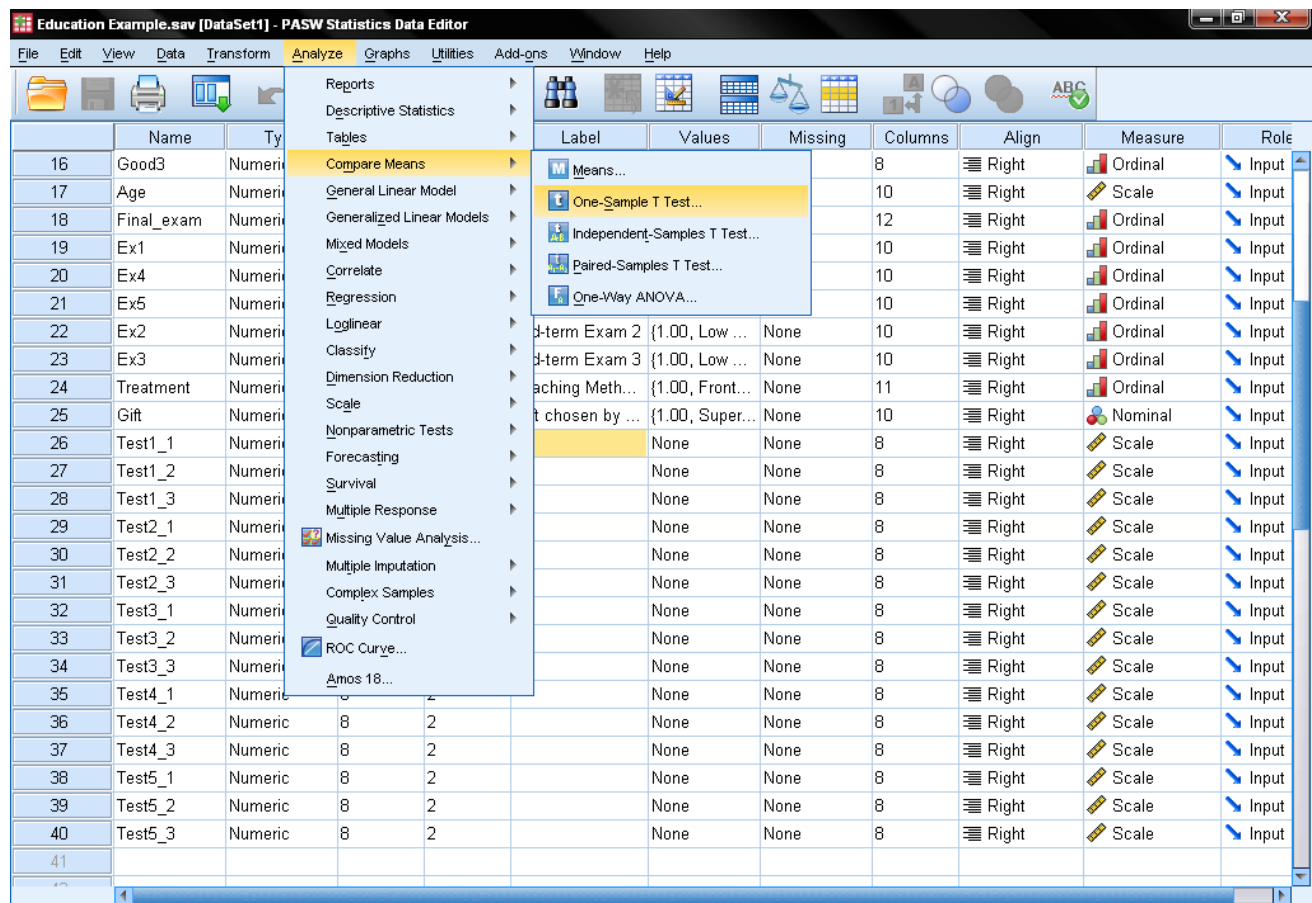
One-Sample Kolmogorov-Smirnov Test

		Age Age
N		107
Normal Parameters <sup>a, b</sup>	Mean	10.0000
	Std. Deviation	.49998
Most Extreme Differences	Absolute	.092
	Positive	.092
	Negative	-.066
Kolmogorov-Smirnov Z		.950
Asymp. Sig. (2-tailed)		.327

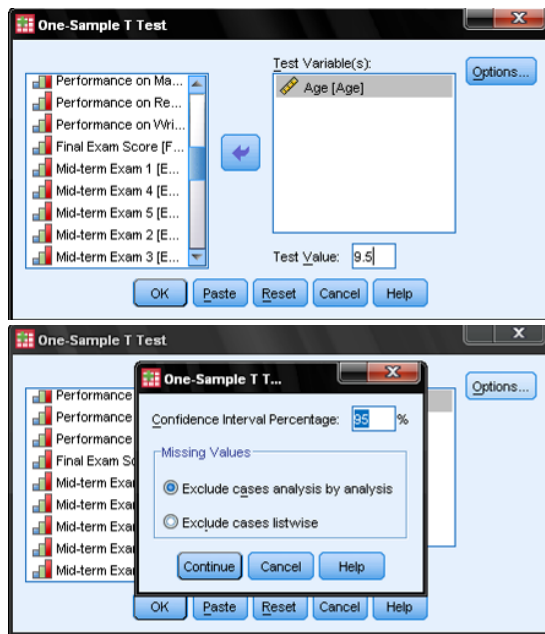
a. Test distribution is Normal.

b. Calculated from data.

Let's move on to the 1 sample t-test, which can be found in *Analyze/Compare Means/One-Sample T-Test...*



The 1-sample t-test dialog box is fairly simple. We add the test variable age to the list of *Test Variables* and enter the *Test Value*. In our case the hypothetical test value is 9.5. The dialog *Options...* gives us the setting how to manage missing values and also the opportunity to specify the width of the confidence interval used for testing.



### The Output of the One-Sample T-Test

The output of the 1-sample t-test consists of only two tables. The first table shows the descriptive statistics we examined previously.

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Age Age	107	10.0000	.49998	.04834

The second table contains the actual 1-sample t-test statistics. The output shows for each variable the sample t-value, degrees of freedom, two-tailed test of significance, mean difference, and the confidence interval.

One-Sample Test

Test Value = 9.5						
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Age Age	10.344	106	.000	.49997	.4041	.5958

The 1-sample t-test can now test our hypothesis that:

$H_1$ : The sample is significantly different from the general population because its mean score is not 9.5.

$H_0$ : The sample is from the general population, which has a mean score for age of 9.5.

In summary, a possible write up could read as follows:

*The hypothesis that the students have an average age of 9½ years was tested with a 1-sample t-test. The test rejects the null hypothesis with  $p < 0.001$  with a mean difference of .49997. Thus*

*we can assume that the sample has a significantly different mean than 9.5 and the hypothesis is not true.*

## **Dependent Sample T-Test**

### **What is the Dependent Sample T-Test?**

The dependent sample t-test is a member of the t-test family. All tests from the t-test family compare one or more mean scores with each other. The t-test family is based on the t-distribution, sometimes also called Student's t. Student is the pseudonym used by W. S. Gosset in 1908 to publish the t-distribution based on his empirical findings on the height and the length of the left middle finger of criminals in a local prison.

Within the t-test family the dependent sample T-Test compares the mean scores of one group in different measurements. It is also called the paired t-test, because measurements from one group must be paired with measurements from the other group. The dependent sample t-test is used when the observations or cases in one sample are linked with the cases in the other sample. This is typically the case when repeated measures are taken, or when analyzing similar units or comparable specimen.

Making repeated measurements or pairing observations is very common when conducting experiments or making observations with time lags. Pairing the measured data points is typically done in order to exclude any confounding or hidden factors (cf. partial correlation). It is also often used to account for individual differences in the baselines, for example pre-existing conditions in clinical research. Consider the example of a drug trial where the participants have individual differences that might have an impact on the outcome of the trial. The typical drug trial splits all participants into a control and the treatment group. The dependent sample t-test can correct for the individual differences or baselines by pairing comparable participants from the treatment and control group. Typical grouping variables are easily obtainable statistics such as age, weight, height, blood pressure. Thus the dependent-sample t-test analyzes the effect of the drug while excluding the influence of different baseline levels of health when the trial began.

Pairing data points and conducting the dependent sample t-test is a common approach to establish causality in a chain of effects. However, the dependent sample t-test only signifies the difference between two mean scores and a direction of change—it does not automatically give a directionality of cause and effect.

Since the pairing is explicitly defined and thus new information added to the data, paired data can always be analyzed with the independent sample t-test as well, but not vice versa. A typical guideline to determine whether the dependent sample t-test is the right test is to answer the following three questions:

- Is there a direct relationship between each pair of observations (e.g., before vs. after scores on the same subject)?

- Are the observations of the data points definitely not random (e.g., they must not be randomly selected specimen of the same population)?
- Do both samples have to have the same number of data points?

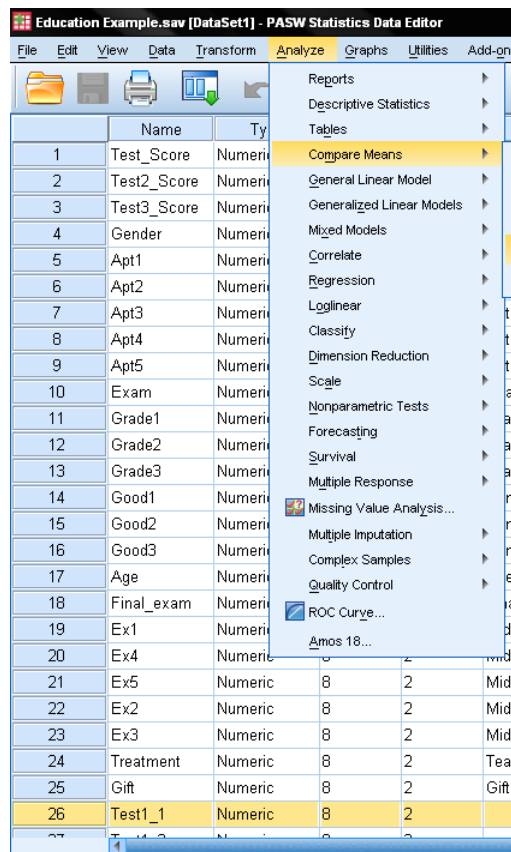
If the answer is yes to all three of these questions the dependent sample t-test is the right test, otherwise use the *independent sample t-test*. In statistical terms the dependent samples t-test requires that the within-group variation, which is a source of measurement errors, can be identified and excluded from the analysis.

### The Dependent Sample T-Test in SPSS

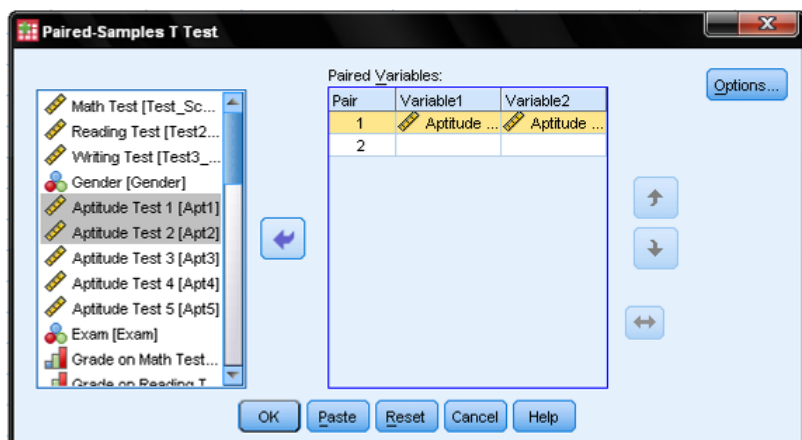
Our research question for the dependent sample t-test is as follows:

*Do students aptitude test1 scores differ from their aptitude test2 scores?*

The dependent samples t-test is found in *Analyze/Compare Means/Paired Samples T Test...*



We need to specify the paired variable in the dialog box for the dependent samples t-test. We need to inform SPSS what is the before and after measurement. SPSS automatically



assumes that the second dimension of the pairing is the case number, i.e. that case number 1 is a pair of measurements between variable 1 and 2.

Although we could specify multiple dependent samples t-test that are executed at the same time, our example only looks at the first and the second aptitude test. Thus we drag & drop 'Aptitude Test 1' into the cell of pair 1 and variable 1, and 'Aptitude Test 2' into the cell pair 1 and variable 2. The *Options...* button allows to define the width of the control interval and how missing values are managed. We leave all settings as they are.

### **The Output of the Dependent Sample T-Test**

The output of the dependent samples t-test consists of only three tables. The first table shows the descriptive statistics of the before and after variable. Here we see that on average the aptitude test score decreased from 29.44 to 24.67, not accounting for individual differences in the baseline.

**Paired Samples Statistics**

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Apt1 Aptitude Test 1	29.44	107	9.370	.906
	Apt2 Aptitude Test 2	24.67	107	7.864	.760

The second table in the output of the dependent samples t-test shows the correlation analysis between the paired variables. This result is not part of any of the other t-tests in the t-test family. The purpose of the correlation analysis is to show whether the use of dependent samples can increase the reliability of the analysis compared to the independent samples t-test. The higher the correlation coefficient the stronger the strength of association between both variable and thus the higher the impact of pairing the data compared to conducting an unpaired t-test. In our example the Pearson's bivariate correlation analysis finds a medium negative correlation that is significant with  $p < 0.001$ . We can therefore assume that pairing our data has a positive impact on the power of t-test.

**Paired Samples Correlations**

		N	Correlation	Sig.
Pair 1	Apt1 Aptitude Test 1 & Apt2 Aptitude Test 2	107	-.499	.000

The third table contains the actual dependent sample t-statistics. The table includes the mean of the differences Before-After, the standard deviation of that difference, the standard error, the t-value, the degrees of freedom, the p-value and the confidence interval for the difference of the mean scores. Unlike the independent samples t-test it does not include the Levene Test for homoscedasticity.

Paired Samples Test									
		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Apt1 Aptitude Test 1 - Apt2 Aptitude Test 2	4.766	14.939	1.444	1.903	7.630	3.300	106	.001

In our example the dependent samples t-test shows that aptitude scores decreased on average by 4.766 with a standard deviation of 14.939. This results in a t-value of  $t = 3.300$  with 106 degrees of freedom. The t-test is highly significant with  $p = 0.001$ . The 95% confidence interval for the average difference of the mean is [1.903, 7.630].

An example of a possible write-up would read as follows:

*The dependent samples t-test showed an average reduction in achieved aptitude scores by 4.766 scores in our sample of 107 students. The dependent sample t-test was used to account for individual differences in the aptitude of the students. The observed decrease is highly significant ( $p = 0.001$ ). Therefore, we can reject the null hypothesis that there is no difference in means and can assume with 99.9% confidence that the observed reduction in aptitude score can also be found in the general population. With a 5% error rate we can assume that the difference in aptitude scores will be between 1.903 and 7.630.*

## Mann-Whitney U-Test

### What is the Mann-Whitney U-Test?

The Mann-Whitney or U-test, is a statistical comparison of the mean. The U-test is a member of the bigger group of dependence tests. Dependence tests assume that the variables in the analysis can be split into independent and dependent variables. A dependence tests that compares the mean scores of an independent and a dependent variable assumes that differences in the mean score of the dependent variable are caused by the independent variable. In most analyses the independent variable is also called factor, because the factor splits the sample in two or more groups, also called factor steps.

Other dependency tests that compare the mean scores of two or more groups are the F-test, ANOVA and the t-test family. Unlike the t-test and F-test, the Mann-Whitney U-test is a non-paracontinuous-level test. That means that the test does not assume any properties regarding the distribution of the underlying variables in the analysis. This makes the Mann-Whitney U-test the analysis to use when analyzing variables of ordinal scale. The Mann-Whitney U-test is also the mathematical basis for the H-test (also called Kruskal Wallis H), which is basically nothing more than a series of pairwise U-tests.

Because the test was initially designed in 1945 by Wilcoxon for two samples of the same size and in 1947 further developed by Mann and Whitney to cover different sample sizes the test is also called Mann–Whitney–Wilcoxon (MWW), Wilcoxon rank-sum test, Wilcoxon–Mann–Whitney test, or Wilcoxon two-sample test.



The Mann-Whitney U-test is mathematically identical to conducting an independent sample t-test (also called 2-sample t-test) with ranked values. This approach is similar to the step from Pearson's bivariate correlation coefficient to Spearman's rho. The U-test, however, does apply a pooled ranking of all variables.

The U-test is a non-paracontinuous-level test, in contrast to the t-tests and the F-test; it does not compare mean scores but median scores of two samples. Thus it is much more robust against outliers and heavy tail distributions. Because the Mann-Whitney U-test is a non-paracontinuous-level test it does not require a special distribution of the dependent variable in the analysis. Thus it is the best test to compare mean scores when the dependent variable is not normally distributed and at least of ordinal scale.

For the test of significance of the Mann-Whitney U-test it is assumed that with  $n > 80$  or each of the two samples at least  $> 30$  the distribution of the U-value from the sample approximates normal distribution. The U-value calculated with the sample can be compared against the normal distribution to calculate the confidence level.

The goal of the test is to test for differences of the media that are caused by the independent variable. Another interpretation of the test is to test if one sample stochastically dominates the other sample. The U-value represents the number of times observations in one sample precede observations in the other sample in the ranking. Which is that with the two samples X and Y the  $\text{Prob}(X > Y) > \text{Prob}(Y > X)$ . Sometimes it also can be found that the Mann-Whitney U-test tests whether the two samples are from the same population because they have the same distribution. Other non-paracontinuous-level tests to compare the mean score are the Kolmogorov-Smirnov Z-test, and the Wilcoxon sign test.

### **The Mann-Whitney U-Test in SPSS**

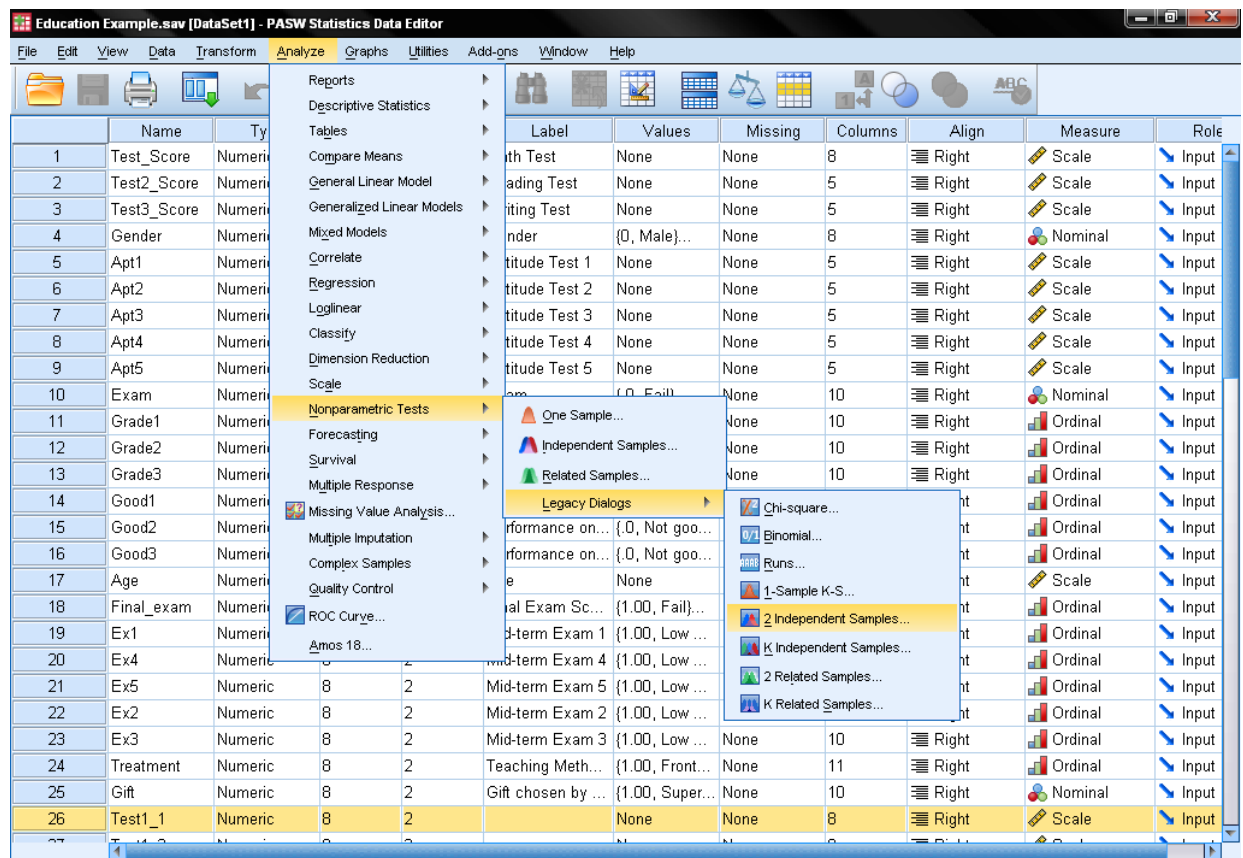
---

The research question for our U-Test is as follows:

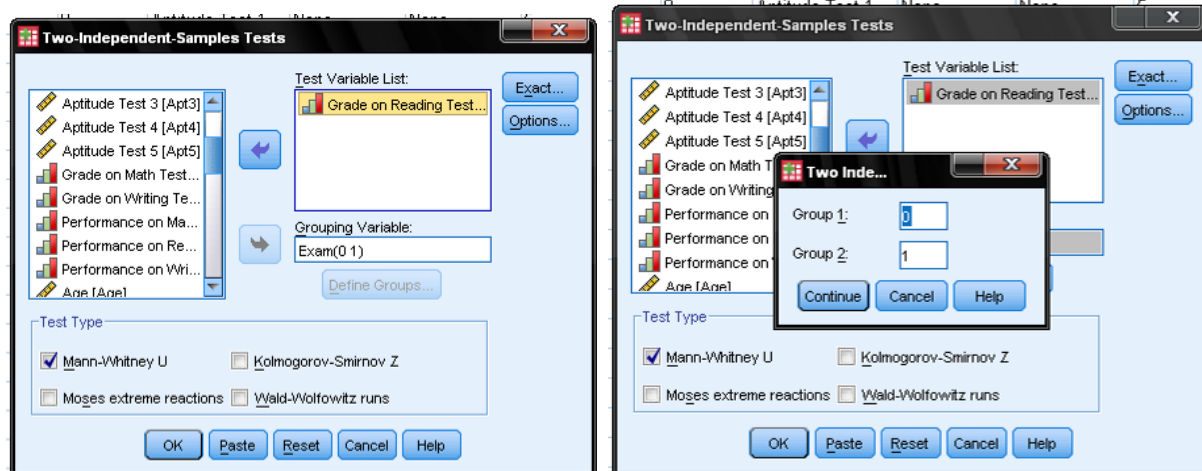
*Do the students that passed the exam achieve a higher grade on the standardized reading test?*

The question indicates that the independent variable is whether the students have passed the final exam or failed the final exam, and the dependent variable is the grade achieved on the standardized reading test (A to F).

The Mann-Whitney U-Test can be found in *Analyze/Nonparacontinuous-level Tests/Legacy Dialogs/2 Independent Samples...*



In the dialog box for the nonparacontinuous-level two independent samples test, we select the ordinal test variable '*mid-term exam 1*', which contains the pooled ranks, and our nominal grouping variable '*Exam*'. With a click on '*Define Groups...*' we need to specify the valid values for the grouping variable *Exam*, which in this case are 0 = fail and 1 = pass.



We also need to select the Test Type. The Mann-Whitney U-Test is marked by default. Like the Mann-Whitney U-Test the Kolmogorov-Smirnov Z-Test and the Wald-Wolfowitz runs-test have the null

hypothesis that both samples are from the same population. Moses extreme reactions test has a different null hypothesis: the range of both samples is the same.

The U-test compares the ranking, Z-test compares the differences in distributions, Wald-Wolfowitz compares sequences in ranking, and Moses compares the ranges of the two samples. The Kolmogorov-Smirnov Z-Test requires continuous-level data (interval or ratio scale), the Mann-Whitney U-Test, Wald-Wolfowitz runs, and Moses extreme reactions require ordinal data.

If we select Mann-Whitney U, SPSS will calculate the U-value and Wilcoxon's W, which is the sum of the ranks for the smaller sample. If the values in the sample are not already ranked, SPSS will sort the observations according to the test variable and assign ranks to each observation.

The dialog box *Exact...* allows us to specify an exact non-paracontinuous-level test of significance and the dialog box *Options...* defines how missing values are managed and if SPSS should output additional descriptive statistics.

### **The Output of the Mann-Whitney U-Test**

The U-test output contains only two tables. The first table shows the descriptive statistics for both

		Ranks			
	Exam Exam	N	Mean Rank	Sum of Ranks	
Grade2 Grade on Reading Test	.00 Fail	64	67.80	4339.50	
	1.00 Pass	43	33.45	1438.50	
	Total	107			

groups, including the sample size, the mean ranking, standard deviation of the rankings and the range of ranks. The descriptive statistics are the same for all nonparacontinuous-level 2-sample tests. Our U-test is going to compare the mean ranks, which we find are higher for the students who failed the exam. Remember that grade A = rank 1 to F = rank 6.

The second table shows the actual test results. The SPSS output contains the Mann-Whitney U, which is the sum of the sum of the ranks for both variables, plus the maximum sum of ranks, minus the sum of ranks for the first sample. In our case  $U=492.5$  and  $W = 1438.5$ , which results in a Z-Value of  $-5.695$ . The test value  $z$  is approximately normally distributed for large samples, so that  $p = 0.000$ . We know that the critical  $z$ -value for a two-tailed test is  $1.96$  and a one-tailed test  $1.645$ . Thus the observed difference in grading is statistically significant.

Test Statistics <sup>a</sup>	
	Grade2 Grade on Reading Test
Mann-Whitney U	492.500
Wilcoxon W	1438.500
Z	-5.695
Asymp. Sig. (2-tailed)	.000

a. Grouping Variable: Exam Exam

In summary, a write-up of the test could read as follows:

*In our observation, 107 pupils were graded on a standardized reading test (grades A to F). Later that year the students wrote a final exam. We analyzed the question whether the students who passed the final exam achieved a better grade in the standardized reading test than the students*

*who failed the final exam. The Mann-Whitney U-test shows that the observed difference between both groups of students is highly significant ( $p < 0.001$ ,  $U = 492.5$ ). Thus we can reject the null hypothesis that both samples are from the same population, and that the observed difference is not only caused by random effects of chance.*

## **Wilcoxon Sign Test**

### **What is the Wilcoxon Sign Test?**

The Wilcoxon Sign test or Wilcoxon Signed-Rank test is a statistical comparison of the average of two dependent samples. The Wilcoxon sign test is a sibling of the t-tests. It is, in fact, a non-parametric-level alternative to the dependent samples t-test. Thus the Wilcoxon signed rank test is used in similar situations as the Mann-Whitney U-test. The main difference is that the Mann-Whitney U-test tests two independent samples, whereas the Wilcoxon sign test tests two dependent samples.

The Wilcoxon Sign test is a test of dependency. All dependence tests assume that the variables in the analysis can be split into independent and dependent variables. A dependence test that compares the averages of an independent and a dependent variable assumes that differences in the average of the dependent variable are caused by the independent variable. Sometimes the independent variable is also called factor because the factor splits the sample in two or more groups, also called factor steps.

Dependence tests analyze whether there is a significant difference between the factor levels. The t-test family uses mean scores as the average to compare the differences, the Mann-Whitney U-test uses mean ranks as the average, and the Wilcoxon Sign test uses signed ranks.

Unlike the t-test and F-test the Wilcoxon sign test is a non-parametric-level test. That means that the test does not assume any properties regarding the distribution of the underlying variables in the analysis. This makes the Wilcoxon sign test the analysis to conduct when analyzing variables of ordinal scale or variables that are not multivariate normal.

The Wilcoxon sign test is mathematically similar to conducting a Mann-Whitney U-test (which is sometimes also called Wilcoxon 2-sample t-test). It is also similar to the basic principle of the dependent samples t-test, because just like the dependent samples t-test the Wilcoxon sign test, tests the difference of observations.

However, the Wilcoxon signed rank test pools all differences, ranks them and applies a negative sign to all the ranks where the difference between the two observations is negative. This is called the signed rank. The Wilcoxon signed rank test is a non-parametric-level test, in contrast to the dependent samples t-tests. Whereas the dependent samples t-test tests whether the average difference between two observations is 0, the Wilcoxon test tests whether the difference between two observations has a mean signed rank of 0. Thus it is much more robust against outliers and heavy tail distributions. Because the Wilcoxon sign test is a non-parametric-level test it does not require a special distribution of the dependent variable in the analysis. Therefore it is the best test to compare mean scores when the dependent variable is not normally distributed and at least of ordinal scale.

For the test of significance of Wilcoxon signed rank test it is assumed that with at least ten paired observations the distribution of the W-value approximates a normal distribution. Thus we can normalize the empirical W-statistics and compare this to the tabulated z-ratio of the normal distribution to calculate the confidence level.

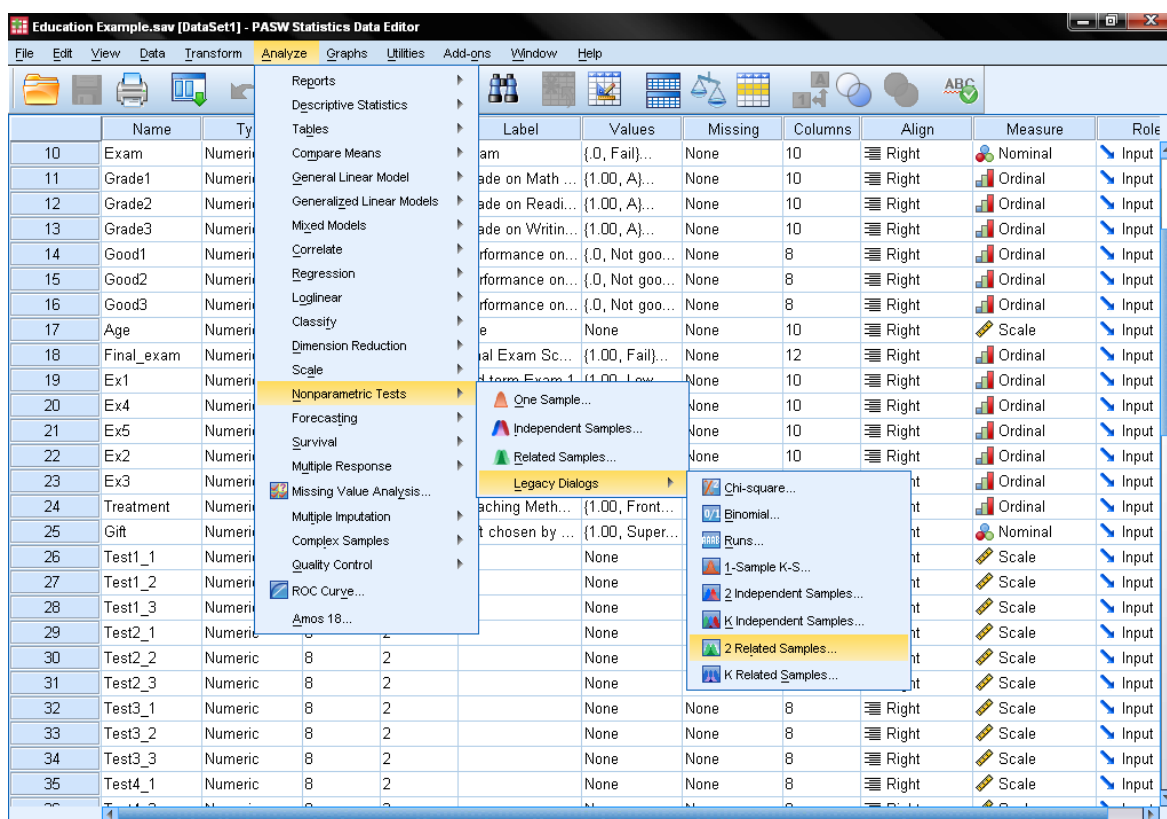
### The Wilcox Sign Test in SPSS

Our research question for the Wilcox Sign Test is as follows:

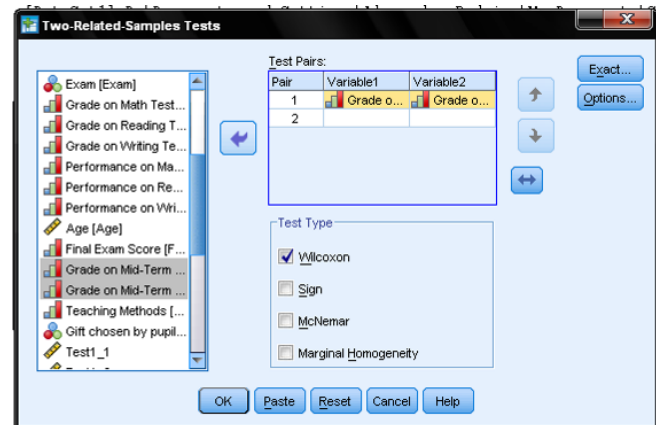
*Does the before-after measurement of the first and the last mid-term exam differ between the students who have been taught in a blended learning course and the students who were taught in a standard classroom setting?*

We only measured the outcome of the mid-term exam on an ordinal scale (grade A to F); therefore a dependent samples t-test cannot be used. This is such because the distribution is only binominal and we do not assume that it approximates a normal distribution. Also both measurements are not independent from each other and therefore we cannot use the Mann-Whitney U-test.

The Wilcox sign test can be found in *Analyze/Nonparacontinuous-level Tests/Legacy Dialog/2 Related Samples...*



In the next dialog box for the nonparacontinuous-level two dependent samples tests we need to define the paired observations. We enter 'Grade on Mid-Term Exam 1' as variable 1 of the first pair and 'Grade on Mid-Term Exam 2' as Variable 2 of the first pair. We also need to select the Test Type. The Wilcoxon Signed Rank Test is marked by default. Alternatively we could choose Sign, McNemar, or Marginal Homogeneity,



**Wilcoxon** – The Wilcoxon signed rank test has the null hypothesis that both samples are from the same population. The Wilcoxon test creates a pooled ranking of all observed differences between the two dependent measurements. It uses the standard normal distributed z-value to test of significance.

**Sign** – The sign test has the null hypothesis that both samples are from the same population. The sign test compares the two dependent observations and counts the number of negative and positive differences. It uses the standard normal distributed z-value to test of significance.

**McNemar** – The McNemar test has the null hypothesis that differences in both samples are equal for both directions. The test uses dichotomous (binary) variables to test whether the observed differences in a 2x2 matrix including all 4 possible combinations differ significantly from the expected count. It uses a Chi-Square test of significance.

**Marginal Homogeneity** – The marginal homogeneity test has the null hypothesis that the differences in both samples are equal in both directions. The test is similar to the McNemar test, but it uses nominal variables with more than two levels. It tests whether the observed differences in a n\*m matrix including all possible combinations differ significantly from the expected count. It uses a Chi-Square test of significance.

If the values in the sample are not already ranked, SPSS will sort the observations according to the test variable and assign ranks to each observation, correcting for tied observations. The dialog box *Exact...* allows us to specify an exact test of significance and the dialog box *Options...* defines how missing values are managed and if SPSS should output additional descriptive statistics.

### **The Output of the Wilcox Sign Test**

The output of the Wilcox sign test only contains two tables. The first table contains all statistics that are required to calculate the Wilcoxon signed ranks test's W. These are the sample size and the sum of ranks. It also includes the mean rank, which is not necessary to calculate the W-value but helps with the interpretation of the data.

In our example we see that 107\*2 observations were made for Exam 1 and Exam 2. The Wilcoxon Sign Test answers the question if the difference is significantly different from zero, and therefore whether the observed difference in mean ranks (39.28 vs. 30.95) can also be found in the general population.

**Ranks**

		N	Mean Rank	Sum of Ranks
Ex2 Grade on Mid-Term Exam 2 - Ex1 Grade on Mid-Term Exam 1	Negative Ranks	29 <sup>a</sup>	39.28	1139.00
	Positive Ranks	39 <sup>b</sup>	30.95	1207.00
	Ties	39 <sup>c</sup>		
	Total	107		

- a. Ex2 Grade on Mid-Term Exam 2 < Ex1 Grade on Mid-Term Exam 1
- b. Ex2 Grade on Mid-Term Exam 2 > Ex1 Grade on Mid-Term Exam 1
- c. Ex2 Grade on Mid-Term Exam 2 = Ex1 Grade on Mid-Term Exam 1

The answer to the test question is in the second table which contains the test of significance statistics. The SPSS output contains the z-value of 0.832, which is smaller than the critical test value of 1.96 for a two-tailed test. The test value z is approximately normally distributed for large samples that are  $n > 10$ , so that  $p = 0.832$ , which indicates that we cannot reject the null hypothesis. We cannot say that there is a significance difference between the grades achieved in the first and the last mid-term exam when we account for individual differences in the baseline.

**Test Statistics<sup>b</sup>**

	Ex2 Grade on Mid-Term Exam 2 - Ex1 Grade on Mid-Term Exam 1
Z	-.212 <sup>a</sup>
Asymp. Sig. (2-tailed)	.832

- a. Based on negative ranks.
- b. Wilcoxon Signed Ranks Test

In summary, a possible write-up of the test could read as follows:

*One-hundred and seven pupils learned with a novel method. A 'before and after' measurement of a standardized test for each student was taken on a classical grading scale from A (rank 1) to F (rank 6). The results seem to indicate that the after measurements show a decrease in test scores (we find more positive ranks than negative ranks). However, the Wilcoxon signed rank test shows that the observed difference between both measurements is not significant when we account for the individual differences in the baseline ( $p = 0.832$ ). Thus we cannot reject the null hypothesis that both samples are from the same population, and we might assume that the novel teaching method did not cause a significant change in grades.*

### Linear Regression

#### What is Linear Regression?

Linear regression is the most basic and commonly used predictive analysis. Regression estimates are used to describe data and to explain the relationship between one dependent variable and one or more independent variables. At the center of the regression analysis is the task of fitting a single line through a scatter plot. The simplest form with one dependent and one independent variable is defined by the formula  $y = a + b \cdot x$ .

Sometimes the dependent variable is also called endogenous variable, prognostic variable or regressand. The independent variables are also called exogenous variables, predictor variables or regressors. However Linear Regression Analysis consists of more than just fitting a linear line through a cloud of data points. It consists of 3 stages: 1) analyzing the correlation and directionality of the data, 2) estimating the model, i.e., fitting the line, and 3) evaluating the validity and usefulness of the model.

There are three major uses for Regression Analysis: 1) causal analysis, 2) forecasting an effect, 3) trend forecasting. Other than correlation analysis, which focuses on the strength of the relationship between two or more variables, regression analysis assumes a dependence or causal relationship between one or more independent and one dependent variable.

Firstly, it might be used to identify the strength of the effect that the independent variable(s) have on a dependent variable. Typical questions are what is the strength of relationship between dose and effect, sales and marketing spending, age and income.

Secondly, it can be used to forecast effects or impacts of changes. That is, regression analysis helps us to understand how much the dependent variable will change when we change one or more independent variables. Typical questions are, “How much additional Y do I get for one additional unit of X?”.

Thirdly, regression analysis predicts trends and future values. The regression analysis can be used to get point estimates. Typical questions are, “What will the price for gold be 6 month from now?” “What is the total effort for a task X?”

#### The Linear Regression in SPSS

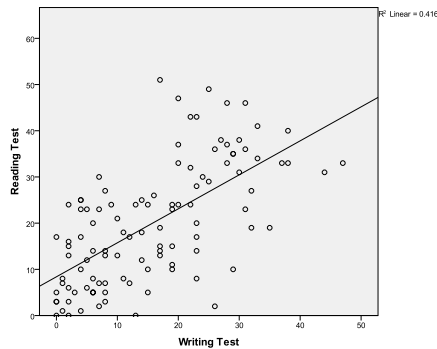
The research question for the Linear Regression Analysis is as follows:

*In our sample of 107 students can we predict the standardized test score of reading when we know the standardized test score of writing?*

The first step is to check whether there is a linear relationship in the data. For that we check the scatter plot (*Graphs/Chart Builder...*). The scatter plot indicates a good linear relationship, which allows us to conduct a linear regression analysis. We can also check the Pearson's Bivariate Correlation



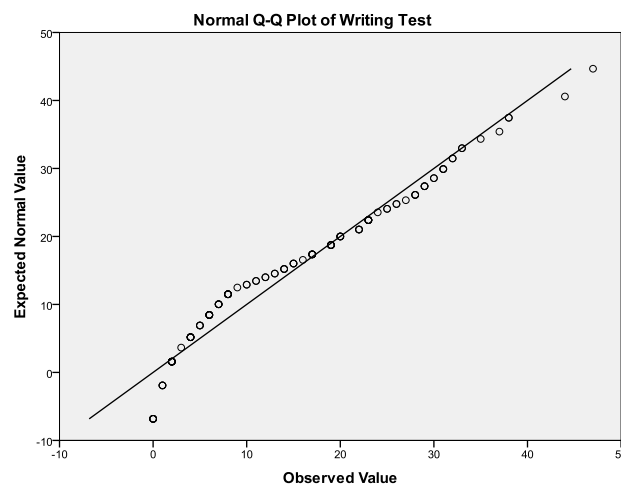
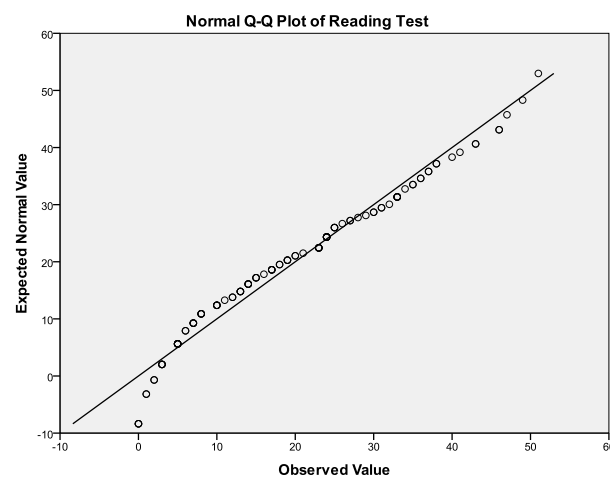
(Analyze/Correlate/Bivariate...) and find that both variables are strongly correlated ( $r = .645$  with  $p < 0.001$ ).



Correlations		Test2_Score Reading Test	Test3_Score Writing Test
Test2_Score Reading Test	Pearson Correlation	1	.645**
	Sig. (2-tailed)		.000
	N	107	107
Test3_Score Writing Test	Pearson Correlation	.645**	1
	Sig. (2-tailed)	.000	
	N	107	107

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

Secondly, we need to check for multivariate normality. We have a look at the Q-Q-Plots (Analyze/Descriptive statistics/Q-Q-Plots...) for both of our variables and see that they are not perfect, but it might be close enough.



We can check our 'eyeball' test with the 1-Sample Kolmogorov-Smirnov test (*Analyze/Non Paracontinuous-level Tests/Legacy Dialogs/1-Sample K-S...*). The test has the null hypothesis that the variable approximates a normal distribution. The results confirm that reading score can be assumed to be multivariate normal ( $p = 0.474$ )

while the writing test is not ( $p = 0.044$ ). To fix this problem we could try to transform the writing test scores using a non-linear transformation (e.g., log). However, we do have a fairly large sample in which case the linear regression is quite robust against violations of normality. It may report too optimistic T-values and F-values.

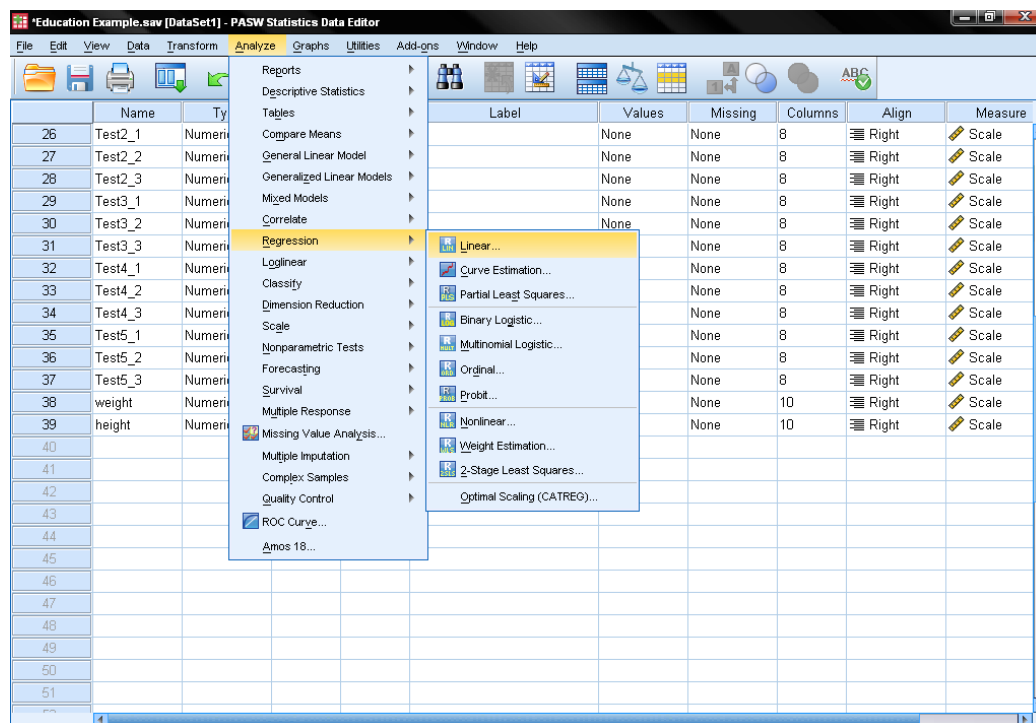
One-Sample Kolmogorov-Smirnov Test

		Test2_Score Reading Test	Test3_Score Writing Test
N		107	107
Normal Parameters <sup>a, b</sup>	Mean	19.97	15.75
	Std. Deviation	13.082	11.465
Most Extreme Differences	Absolute	.082	.134
	Positive	.082	.134
	Negative	-.063	-.085
Kolmogorov-Smirnov Z		.844	1.382
Asymp. Sig. (2-tailed)		.474	.044

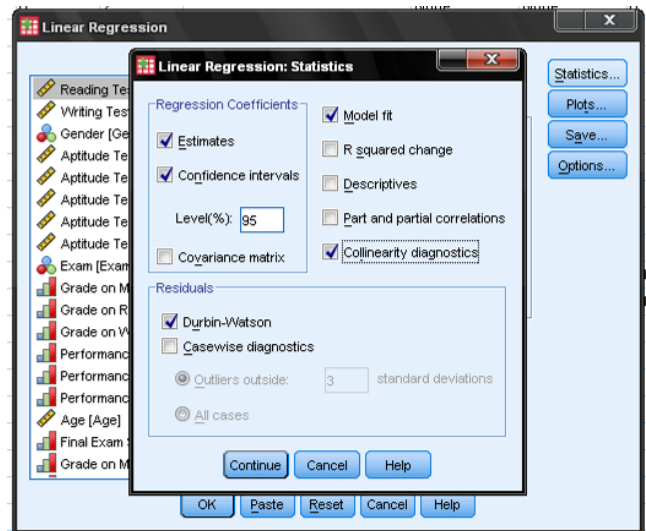
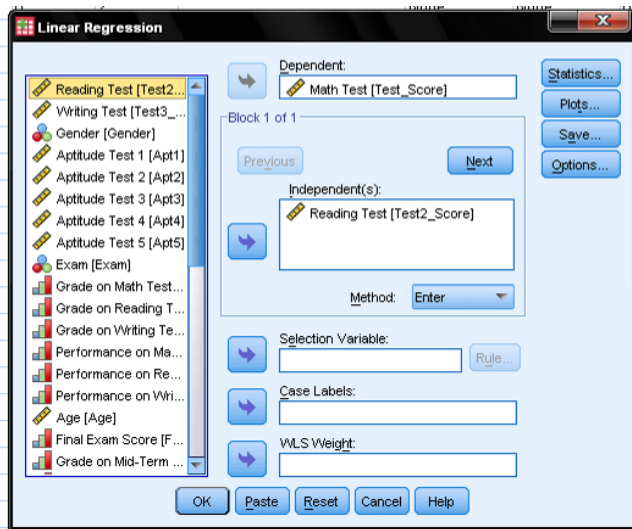
a. Test distribution is Normal.

b. Calculated from data.

We now can conduct the linear regression analysis. Linear regression is found in SPSS in *Analyze/Regression/Linear...*

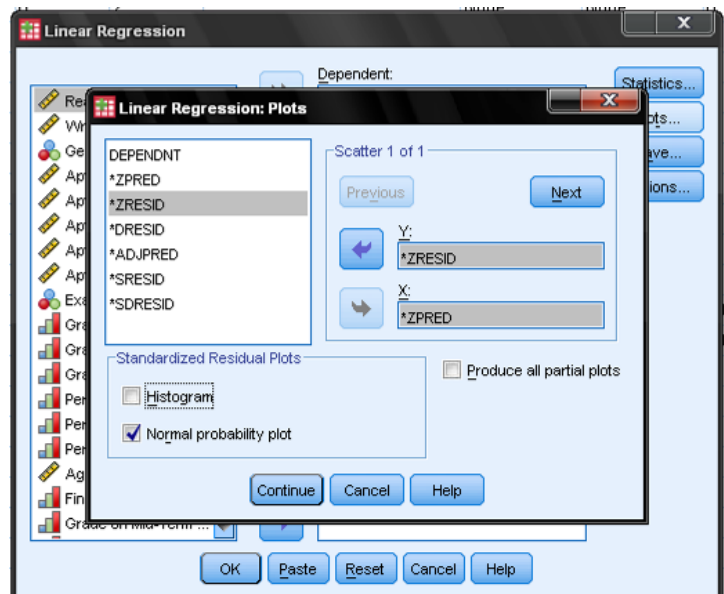


To answer our simple research question we just need to add the Math Test Score as the dependent variable and the Writing Test Score as the independent variable. The menu *Statistics...* allows us to include additional information that we need to assess the validity of our linear regression analysis. In order to assess autocorrelation (especially if we have time series data) we add the Durbin-Watson Test, and to check for multicollinearity we add the Collinearity diagnostics.



Lastly, we click on the menu *Plots...* to add the standardized residual plots to the output. The standardized residual plots chart ZPRED on the x-axis and ZRESID on the y-axis. This standardized plot allows us to check for heteroscedasticity.

We leave all the options in the menus *Save...* and *Options...* as they are and are now ready to run the test.



### The Output of the Linear Regression Analysis

The output's first table shows the model summary and overall fit statistics. We find that the adjusted  $R^2$  of our model is 0.333 with the  $R^2 = .339$ . This means that the linear regression explains 33.9% of the variance in the data. The adjusted  $R^2$  corrects the  $R^2$  for the number of independent variables in the analysis, thus it helps detect over-fitting, because every new independent variable in a regression model always explains a little additional bit of the variation, which increases the  $R^2$ . The Durbin-Watson  $d = 2.227$  is between the two critical values of  $1.5 < d < 2.5$ , therefore we can assume that there is no first order linear autocorrelation in the data.

### Model Summary<sup>b</sup>

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.582 <sup>a</sup>	.339	.333	14.58556	2.227

a. Predictors: (Constant), Test2\_Score Reading Test

b. Dependent Variable: Test\_Score Math Test

The next table is the F-test. The linear regression's F-test has the null hypothesis that there is no linear relationship between the two variables (in other words  $R^2=0$ ). With  $F = 53.828$  and 106 degrees of freedom the test is highly significant, thus we can assume that there is a linear relationship between the variables in our model.

### ANOVA<sup>b</sup>

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	11451.281	1	11451.281	53.828	.000 <sup>a</sup>
	Residual	22337.541	105	212.738		
	Total	33788.822	106			

a. Predictors: (Constant), Test2\_Score Reading Test

b. Dependent Variable: Test\_Score Math Test

The next table shows the regression coefficients, the intercept, and the significance of all coefficients and the intercept in the model. We find that our linear regression analysis estimates the linear regression function to be  $y = 36.824 + .795 * x$ . This means that an increase in one unit of  $x$  results in an increase of .795 units of  $y$ . The test of significance of the linear regression analysis tests the null hypothesis that the estimated coefficient is 0. The t-test finds that both intercept and variable are highly significant ( $p < 0.001$ ) and thus we might say that they are significantly different from zero.

### Coefficients<sup>a</sup>

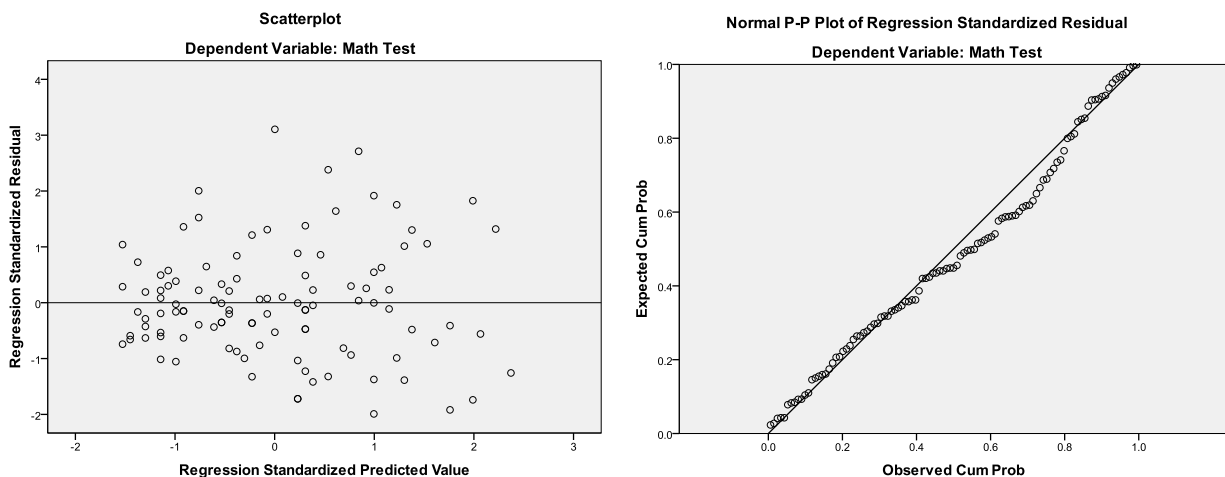
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B		Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	36.824	2.582		14.263	.000	31.704	41.943	1.000	1.000
	Test2_Score Reading Test	.795	.108	.582	7.337	.000	.580	1.009	1.000	1.000

a. Dependent Variable: Test\_Score Math Test

This table also includes the Beta weights. Beta weights are the standardized coefficients and they allow comparing of the size of the effects of different independent variables if the variables have different

units of measurement. The table also includes the collinearity statistics. However, since we have only one independent variable in our analysis we do not pay attention to neither of the two values.

The last thing we need to check is the homoscedasticity and normality of residuals. The scatterplot indicates constant variance. The P-P-Plot of  $z^*_{pred}$  and  $z^*_{presid}$  shows us that in our linear regression analysis there is no tendency in the error terms.



In summary, a possible write-up could read as follows:

*We investigated the relationship between the reading and writing scores achieved on our standardized tests. The correlation analysis found a medium positive correlation between the two variables ( $r = 0.645$ ). We then conducted a simple regression analysis to further substantiate the suspected relationship. The estimated regression model is Math Score =  $36.824 + .795 \cdot \text{Reading Score}$  with an adjusted  $R^2$  of 33.3%; it is highly significant with  $p < 0.001$  and  $F = 53.828$ . The standard error of the estimate is 14.58556. Thus we can not only show a positive linear relationship, and we can also conclude that for every additional reading score achieved the math score will increase by approximately .795 units.*

## Multiple Linear Regression

### What is Multiple Linear Regression?

Multiple linear regression is the most common form of the regression analysis. As a predictive analysis, multiple linear regression is used to describe data and to explain the relationship between one dependent variable and two or more independent variables.

At the center of the multiple linear regression analysis lies the task of fitting a single line through a scatter plot. More specifically, the multiple linear regression fits a line through a multi-dimensional cloud of data points. The simplest form has one dependent and two independent variables. The

general form of the multiple linear regression is defined as  $y = \beta_0 + \beta_1 \cdot x_{i1} + \beta_2 \cdot x_{i2} + \dots + \beta_p \cdot x_{ip}$  for  $i = 1 \dots n$ .

Sometimes the dependent variable is also called endogenous variable, criterion variable, prognostic variable or regressand. The independent variables are also called exogenous variables, predictor variables or regressors.

Multiple Linear Regression Analysis consists of more than just fitting a linear line through a cloud of data points. It consists of three stages: 1) analyzing the correlation and directionality of the data, 2) estimating the model, i.e., fitting the line, and 3) evaluating the validity and usefulness of the model.

There are three major uses for Multiple Linear Regression Analysis: 1) causal analysis, 2) forecasting an effect, and 3) trend forecasting. Other than correlation analysis, which focuses on the strength of the relationship between two or more variables, regression analysis assumes a dependence or causal relationship between one or more independent and one dependent variable.

Firstly, it might be used to identify the strength of the effect that the independent variables have on a dependent variable. Typical questions would seek to determine the strength of relationship between dose and effect, sales and marketing spend, age and income.

Secondly, it can be used to forecast effects or impacts of changes. That is to say, multiple linear regression analysis helps us to understand how much the dependent variable will change when we change the independent variables. A typical question would be "How much additional Y do I get for one additional unit X?"

Thirdly, multiple linear regression analysis predicts trends and future values. The multiple linear regression analysis can be used to get point estimates. Typical questions might include, "What will the price for gold be six months from now? What is the total effort for a task X?"

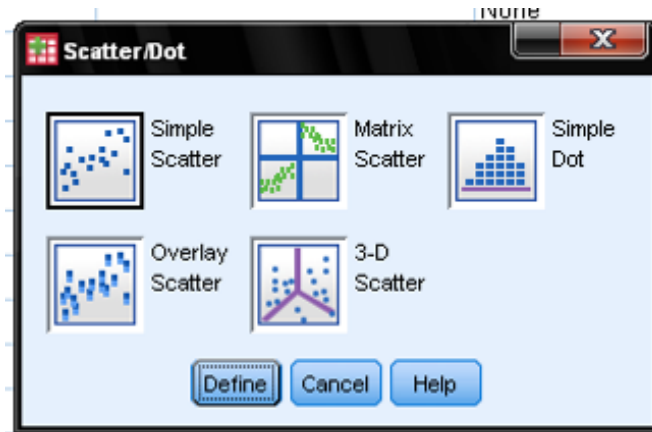
### **The Multiple Linear Regression in SPSS**

---

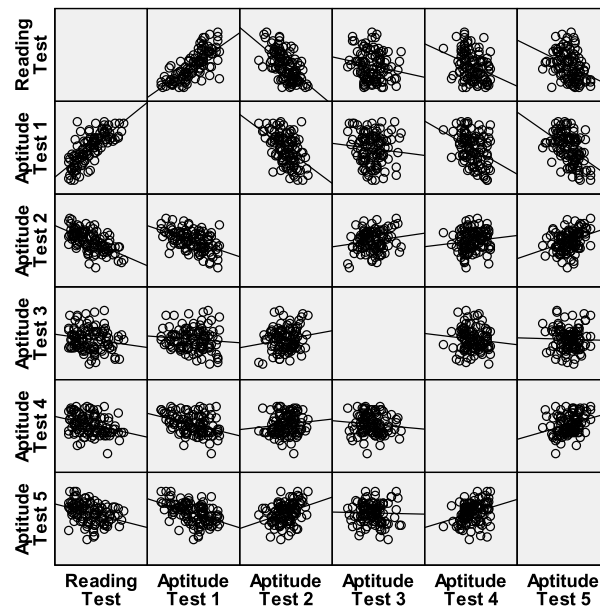
Our research question for the multiple linear regression is as follows:

*Can we explain the reading score that a student achieved on the standardized test with the five aptitude tests?*

First, we need to check whether there is a linear relationship between the independent variables and the dependent variable in our multiple linear regression model. To do so, we check the scatter plots. We could create five individual scatter plots using the *Graphs/Chart Builder...* Alternatively we can use the Matrix Scatter Plot in the menu *Graphs/Legacy Dialogs/Scatter/Dot...*



The scatter plots indicate a good linear relationship between writing score and the aptitude tests 1 to 5, where there seems to be a positive relationship for aptitude test 1 and a negative linear relationship for aptitude tests 2 to 5.



Secondly, we need to check for multivariate normality. This can either be done with an 'eyeball' test on the Q-Q-Plots or by using the 1-Sample K-S test to test the null hypothesis that the variable approximates a normal distribution. The K-S test is not significant for all variables, thus we can assume normality.

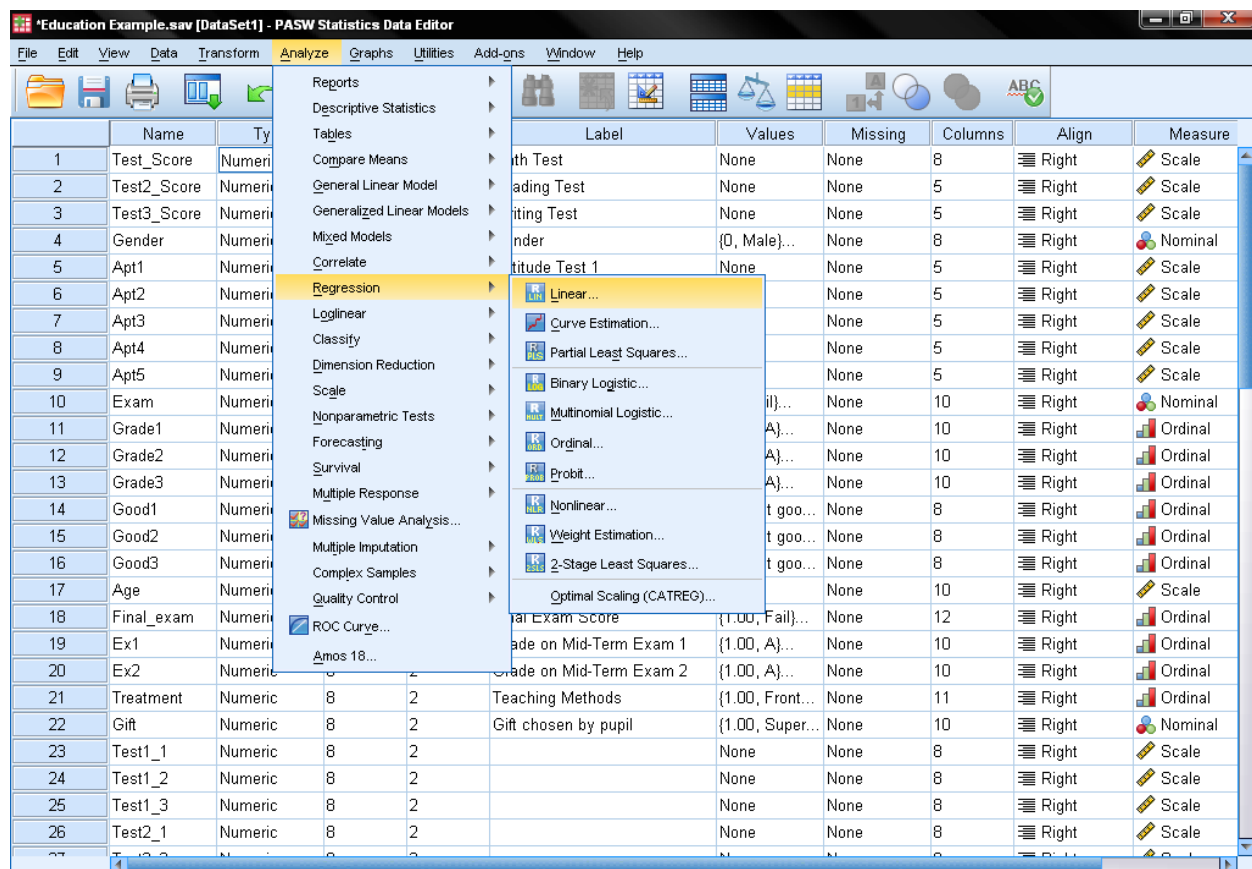
### One-Sample Kolmogorov-Smirnov Test

		Test2_Score Reading Test	Apt1 Aptitude Test 1	Apt2 Aptitude Test 2	Apt3 Aptitude Test 3	Apt4 Aptitude Test 4	Apt5 Aptitude Test 5
N		107	107	107	107	107	107
Normal Parameters <sup>a, b</sup>	Mean	19.97	29.44	24.67	26.27	31.02	29.49
	Std. Deviation	13.082	9.370	7.864	7.020	6.103	7.846
Most Extreme Differences	Absolute	.082	.106	.084	.084	.091	.084
	Positive	.082	.057	.054	.084	.091	.057
	Negative	-.063	-.106	-.084	-.047	-.061	-.084
Kolmogorov-Smirnov Z		.844	1.096	.873	.866	.938	.865
Asymp. Sig. (2-tailed)		.474	.181	.430	.442	.342	.443

a. Test distribution is Normal.

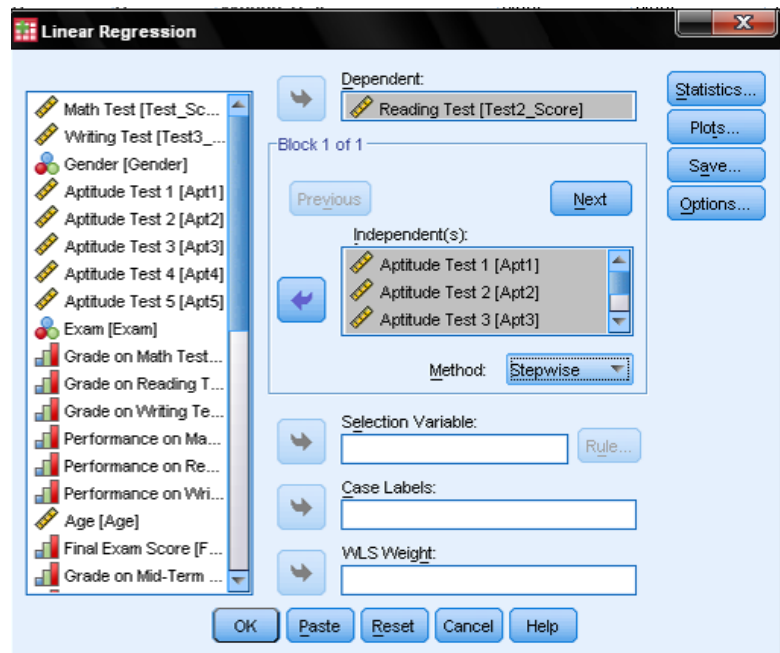
b. Calculated from data.

Multiple linear regression is found in SPSS in *Analyze/Regression/Linear...*

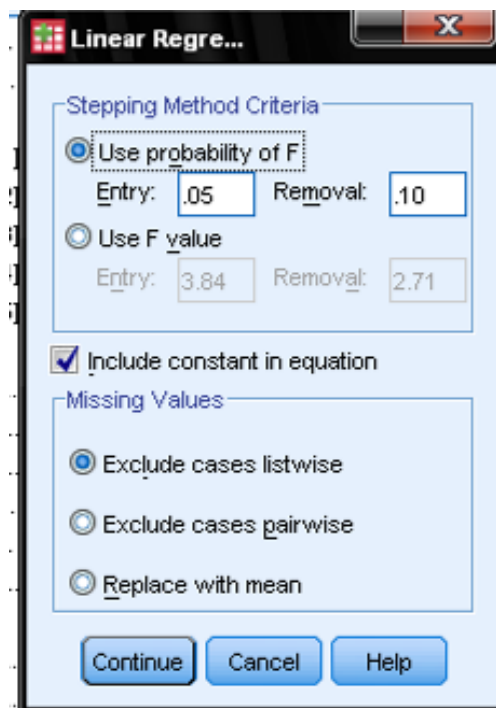




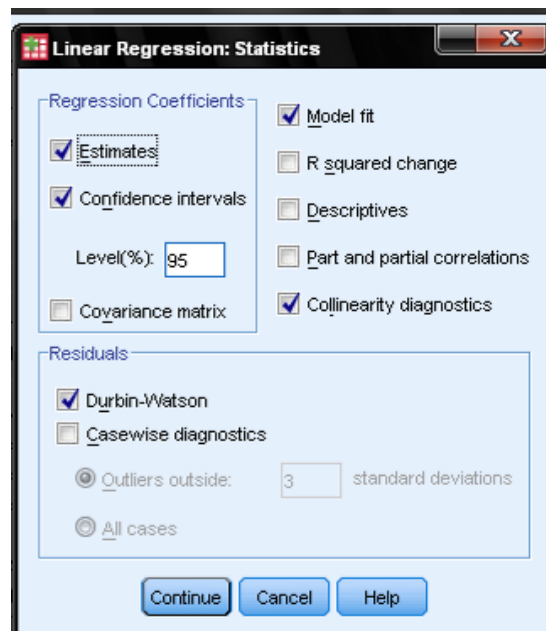
To answer our research question we need to enter the variable reading scores as the dependent variable in our multiple linear regression model and the aptitude test scores (1 to 5) as independent variables. We also select *stepwise* as the method. The default method for the multiple linear regression analysis is 'Enter', which means that all variables are forced to be in the model. But since over-fitting is a concern of ours, we want only the variables in the model that explain additional variance. Stepwise means that the variables are entered into the regression model in the order of their explanatory power.



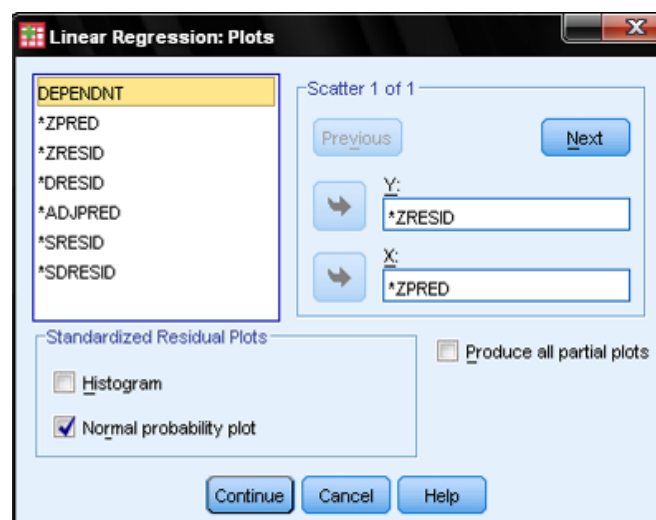
In the field *Options...* we can define the criteria for stepwise inclusion in the model. We want to include variables in our multiple linear regression model that increase F by at least 0.05 and we want to exclude them again if the increase F by less than 0.1. This dialog box also allows us to manage missing values (e.g., replace them with the mean).



The dialog *Statistics...* allows us to include additional statistics that we need to assess the validity of our linear regression analysis. Even though it is not a time series, we include Durbin-Watson to check for autocorrelation and we include the collinearity that will check for autocorrelation.



In the dialog *Plots...*, we add the standardized residual plot (ZPRED on x-axis and ZRESID on y-axis), which allows us to eyeball homoscedasticity and normality of residuals.



### ***The Output of the Multiple Linear Regression Analysis***

The first table tells us the model history SPSS has estimated. Since we have selected a stepwise multiple linear regression SPSS automatically estimates more than one regression model. If all of our five independent variables were relevant and useful to explain the reading score, they would have been

entered one by one and we would find five regression models. In this case however, we find that the best explaining variable is Aptitude Test 1, which is entered in the first step while Aptitude Test 2 is entered in the second step. After the second model is estimated, SPSS stops building new models because none of the remaining variables increases F sufficiently. That is to say, none of the variables adds significant explanatory power of the regression model.

**Variables Entered/Removed<sup>a</sup>**

Model	Variables Entered	Variables Removed	Method
1	Apt1 Aptitude Test 1	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
2	Apt2 Aptitude Test 2	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).

a. Dependent Variable: Test2\_Score Reading Test

The next table shows the multiple linear regression model summary and overall fit statistics. We find that the adjusted  $R^2$  of our model 2 is 0.624 with the  $R^2 = .631$ . This means that the linear regression model with the independent variables Aptitude Test 1 and 2 explains 63.1% of the variance of the Reading Test Score. The Durbin-Watson  $d = 1.872$ , which is between the two critical values of 1.5 and 2.5 ( $1.5 < d < 2.5$ ), and therefore we can assume that there is no first order linear autocorrelation in our multiple linear regression data.

**Model Summary<sup>c</sup>**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.750 <sup>a</sup>	.562	.558	8.697	
2	.794 <sup>b</sup>	.631	.624	8.025	1.872

a. Predictors: (Constant), Apt1 Aptitude Test 1

b. Predictors: (Constant), Apt1 Aptitude Test 1, Apt2 Aptitude Test 2

c. Dependent Variable: Test2\_Score Reading Test

If we would have forced all independent variables (Method: Enter) into the linear regression model we would have seen a little higher  $R^2 = 80.2\%$  but an almost identical adjusted  $R^2 = 62.5\%$ .

**Model Summary<sup>b</sup>**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.802 <sup>a</sup>	.643	.625	8.006	1.802

a. Predictors: (Constant), Apt5 Aptitude Test 5, Apt3 Aptitude Test 3, Apt4 Aptitude Test 4, Apt2 Aptitude Test 2, Apt1 Aptitude Test 1

b. Dependent Variable: Test2\_Score Reading Test

The next table is the F-test, or ANOVA. The F-Test is the test of significance of the multiple linear regression. The F-test has the null hypothesis that there is no linear relationship between the variables (in other words  $R^2=0$ ). The F-test of Model 2 is highly significant, thus we can assume that there is a linear relationship between the variables in our model.

**ANOVA<sup>c</sup>**

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	10198.703	1	10198.703	134.832	.000 <sup>a</sup>
	Residual	7942.213	105	75.640		
	Total	18140.916	106			
2	Regression	11443.719	2	5721.860	88.854	.000 <sup>b</sup>
	Residual	6697.197	104	64.396		
	Total	18140.916	106			

a. Predictors: (Constant), Apt1 Aptitude Test 1

b. Predictors: (Constant), Apt1 Aptitude Test 1, Apt2 Aptitude Test 2

c. Dependent Variable: Test2\_Score Reading Test

The next table shows the multiple linear regression coefficient estimates including the intercept and the significance levels. In our second model we find a non-significant intercept (which commonly happens and is nothing to worry about) but also highly significant coefficients for Aptitude Test 1 and 2. Our regression equation would be: Reading Test Score = 7.761 + 0.836\*Aptitude Test 1 – 0.503\*Aptitude Test 2. For every additional point achieved on Aptitude Test, we can interpret that the Reading Score increases by 0.836, while for every additional score on Aptitude Test 2 the Reading Score decreases by 0.503.

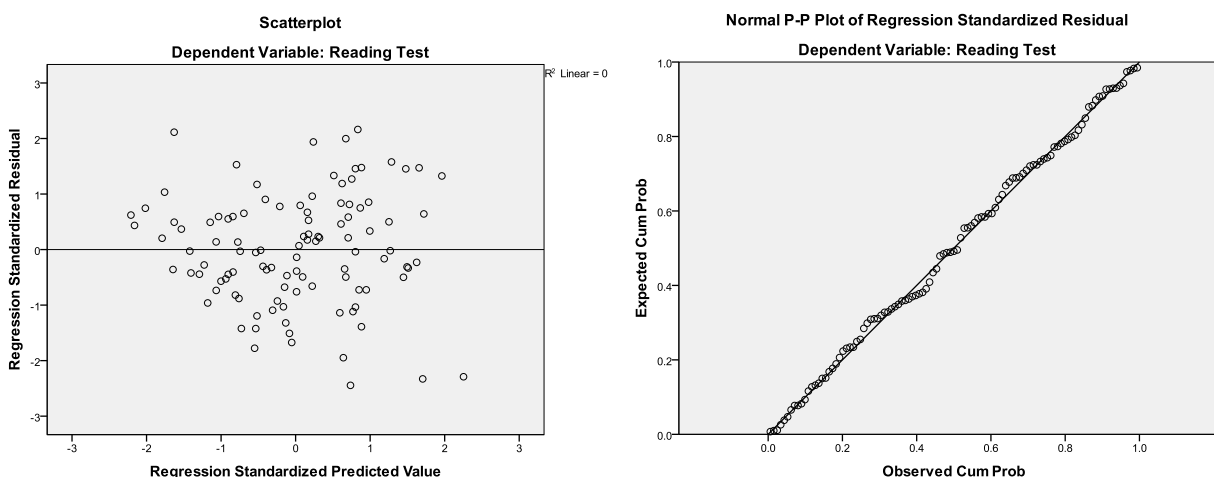
Coefficients<sup>a</sup>

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B		Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	-10.846	2.784		-3.896	.000	-16.367	-5.326		
	Apt1 Aptitude Test 1	1.047	.090	.750	11.612	.000	.868	1.226	1.000	1.000
2	(Constant)	7.761	4.951		1.568	.120	-2.056	17.579		
	Apt1 Aptitude Test 1	.836	.096	.599	8.712	.000	.646	1.027	.751	1.332
	Apt2 Aptitude Test 2	-.503	.114	-.302	-4.397	.000	-.730	-.276	.751	1.332

a. Dependent Variable: Test2\_Score Reading Test

Since we have multiple independent variables in the analysis the Beta weights compare the relative importance of each independent variable in standardized terms. We find that Test 1 has a higher impact than Test 2 (beta = .599 and beta = .302). This table also checks for multicollinearity in our multiple linear regression model. Multicollinearity is the extent to which independent variables are correlated with each other. Tolerance should be greater than 0.1 (or VIF < 10) for all variables—which they are. If tolerance is less than 0.1 there is a suspicion of multicollinearity, and with tolerance less than 0.01 there is proof of multicollinearity.

Lastly, as the Goldfeld-Quandt test is not supported in SPSS, we check is the homoscedasticity and normality of residuals with an eyeball test of the Q-Q-Plot of  $z^*$ pred and  $z^*$ presid. The plot indicates that in our multiple linear regression analysis there is no tendency in the error terms.



In summary, a possible write-up could read as follows:

*We investigated the relationship between the reading scores achieved on our standardized tests and the scores achieved on the five aptitude tests. The stepwise multiple linear regression analysis found that Aptitude Test 1 and 2 have relevant explanatory power. Together the estimated regression model (Reading Test Score =  $7.761 + 0.836 \times \text{Aptitude Test 1} - 0.503 \times \text{Aptitude Test 2}$ ) explains 63.1% of the variance of the achieved Reading Score with an adjusted  $R^2$  of 62.4%. The regression model is highly significant with  $p < 0.001$  and  $F = 88.854$ . The standard error of the estimate is 8.006. Thus we can not only show a linear relationship*

between aptitude tests 1 (positive) and 2 (negative), we can also conclude that for every additional reading score achieved the reading score will increase by approximately 0.8 (Aptitude Test 1) and decrease by 0.5 (Aptitude Test 2).

## Logistic Regression

### What is Logistic Regression?

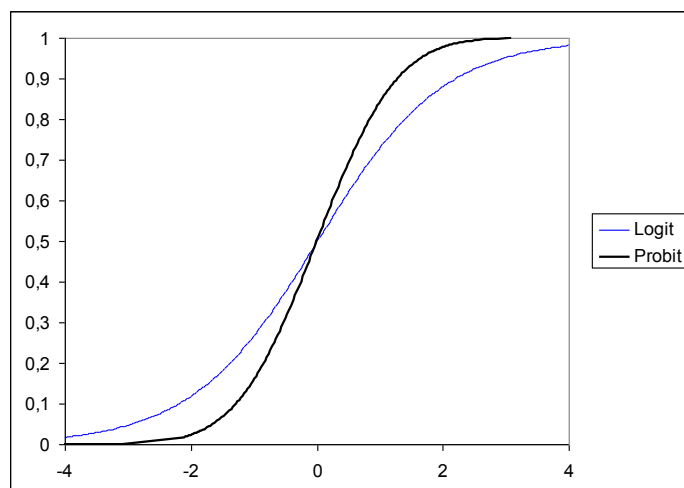
Logistic regression is the linear regression analysis to conduct when the dependent variable is dichotomous (binary). Like all linear regressions the logistic regression is a predictive analysis. Logistic regression is used to describe data and to explain the relationship between one dependent binary variable and one or more continuous-level(interval or ratio scale) independent variables.

Standard linear regression requires the dependent variable to be of continuous-level(interval or ratio) scale. How can we apply the same principle to a dichotomous (0/1) variable? Logistic regression assumes that the dependent variable is a stochastic event. For instance, if we analyze a pesticides kill rate, the outcome event is either *killed* or *alive*. Since even the most resistant bug can only be either of these two states, logistic regression thinks in likelihoods of the bug getting killed. If the likelihood of killing the bug is greater than 0.5 it is assumed dead, if it is less than 0.5 it is assumed alive.

It is quite common to run a regular linear regression analysis with dummy independent variables. A dummy variable is a binary variable that is treated as if it would be continuous. Practically speaking, a dummy variable increases the intercept thereby creating a second parallel line above or below the estimated regression line.

Alternatively, we could try to just create a multiple linear regression with a dummy dependent variable. This approach, however, has two major shortcomings. Firstly, it can lead to probabilities outside of the (0,1) interval, and secondly residuals will all have the same variance (think of parallel lines in the  $\hat{z}pred * \hat{z}resid$  plot).

To solve these shortcomings we can use a logistic function to restrict the probability values to (0,1). The logistic function is  $p(x) = 1/(1+\exp(-x))$ . Technically this can be resolved to  $\ln(p/(1-p)) = a + b \cdot x$ .  $\ln(p/(1-p))$  is also called the log odds. Sometimes instead of a logit model for logistic regression, a probit model is used. The following graph shows the difference for a logit and a probit model for different values [-4,4]. Both models are commonly used in logistic regression; in most cases a model is fitted with both functions and the function with the better fit is chosen. However, probit assumes normal distribution of the probability of the event, when logit assumes the log



distribution. Thus the difference between logit and probit is usually only visible in small samples.

At the center of the logistic regression analysis lies the task of estimating the log odds of an event. Mathematically, logistic regression estimates a multiple linear regression function defined as  $\text{logit}(p)$

$$= \log\left(\frac{p(Y=1)}{1-(p=1)}\right) = \beta_0 + \beta_1 \cdot X_{i1} + \beta_2 \cdot X_{i2} + \dots + \beta_p \cdot X_{ip} \text{ for } i = 1 \dots n.$$

Logistic regression is similar to the Discriminant Analysis. Discriminant analysis uses the regression line to split a sample in two groups along the levels of the dependent variable. Whereas the logistic regression analysis uses the concept of probabilities and log odds with cut-off probability 0.5, the discriminant analysis cuts the geometrical plane that is represented by the scatter cloud. The practical difference is in the assumptions of both tests. If the data is multivariate normal, homoscedasticity is present in variance and covariance and the independent variables are linearly related. Discriminant analysis is then used because it is more statistically powerful and efficient. Discriminant analysis is typically more accurate than logistic regression in terms of predictive classification of the dependent variable.

### **The Logistic Regression in SPSS**

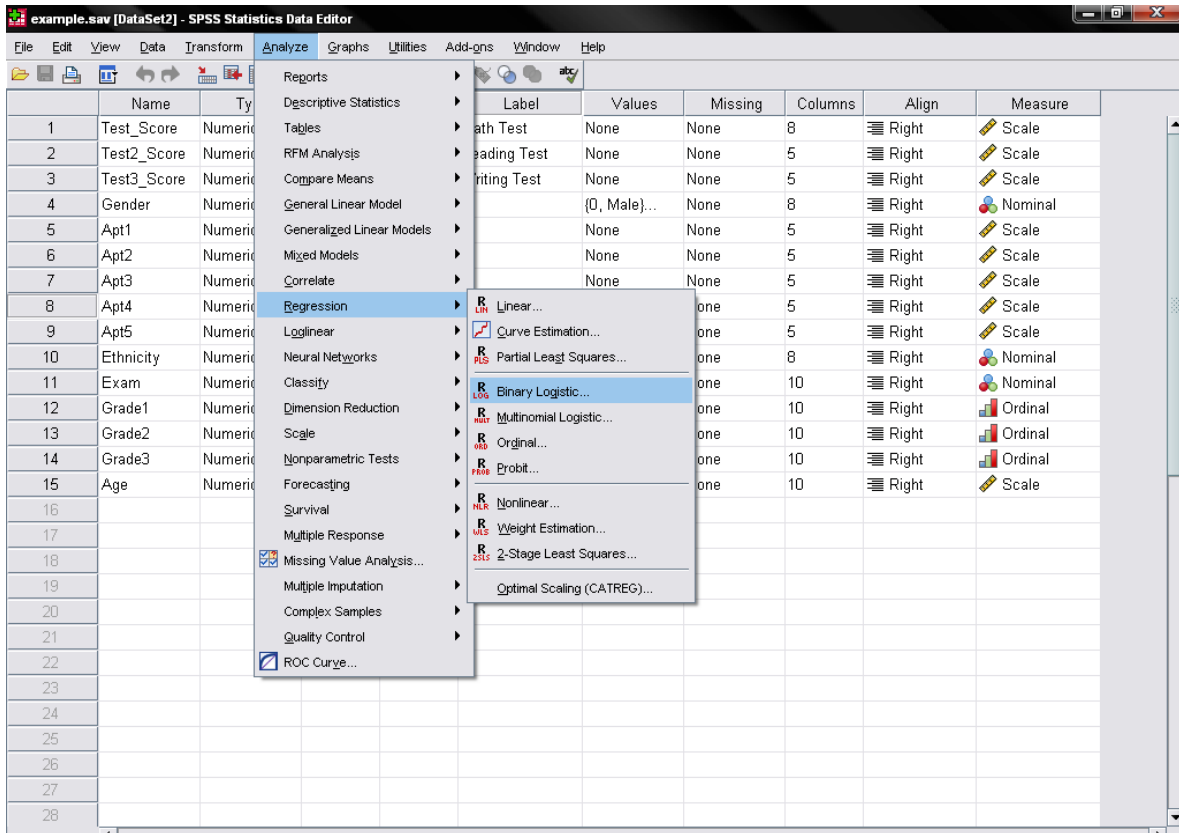
---

In terms of logistic regression, let us consider the following example:

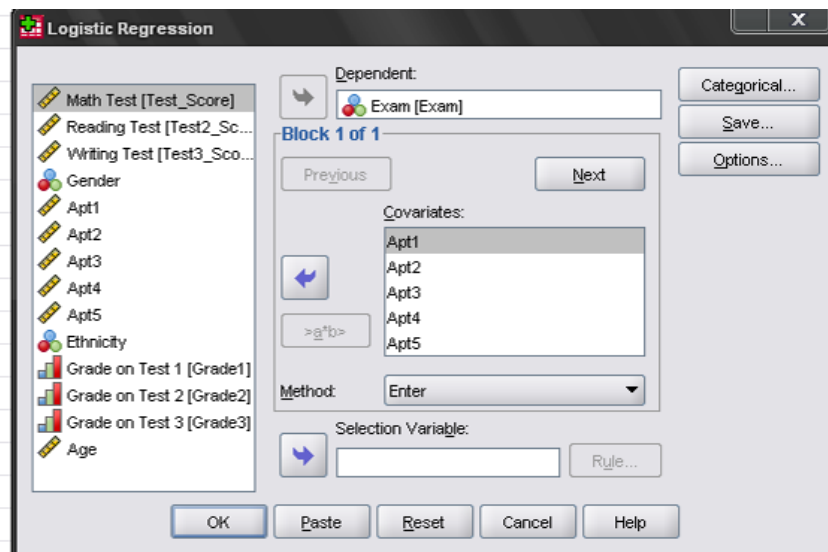
*A research study is conducted on 107 pupils. These pupils have been measured with five different aptitude tests—one for each important category (reading, writing, understanding, summarizing etc.). How do these aptitude tests predict if the pupils pass the year end exam?*

First we need to check that all cells in our model are populated. Since we don't have any categorical variables in our design we will skip this step.

Logistic Regression is found in SPSS under *Analyze/Regression/Binary Logistic...*



This opens the dialog box to specify the model. Here we need to enter the nominal variable *Exam* (pass = 1, fail = 0) into the dependent variable box and we enter all aptitude tests as the first block of covariates in the model.



The menu *Categorical...* allows to specify contrasts for categorical variables (which we do not have in our logistic regression model), and *Options...* offers several additional statistics, which don't need.



### The Output of the Logistic Regression Analysis

The first table simply shows the case processing summary, which lists nothing more than the valid sample size.

**Case Processing Summary**

Unweighted Cases <sup>a</sup>		N	Percent
Selected Cases	Included in Analysis	107	100.0
	Missing Cases	0	.0
	Total	107	100.0
Unselected Cases		0	.0
Total		107	100.0

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

The next three tables are the results for the intercept model. That is the Maximum Likelihood model if only the intercept is included without any of the dependent variables in the analysis. This is basically only interesting to calculate the Pseudo  $R^2$  that describes the goodness of fit for the logistic model.

**Variables in the Equation**

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 0 Constant	-.398	.197	4.068	1	.044	.672

**Classification Table<sup>a,b</sup>**

			Predicted		
			Exam		Percentage Correct
			Fail	Pass	
Step 0 Exam	Fail		64	0	100.0
	Pass		43	0	.0
	Overall Percentage				59.8

a. Constant is included in the model.

b. The cut value is .500

### Variables not in the Equation

	Score	df	Sig.
Step 0 Variables Apt1	30.479	1	.000
Apt2	10.225	1	.001
Apt3	2.379	1	.123
Apt4	6.880	1	.009
Apt5	5.039	1	.025
Overall Statistics	32.522	5	.000

The relevant tables can be found in the section '*Block 1*' in the SPSS output of our logistic regression analysis. The first table includes the Chi-Square goodness of fit test. It has the null hypothesis that intercept and all coefficients are zero. We can reject this null hypothesis.

### Omnibus Tests of Model Coefficients

	Chi-square	df	Sig.
Step 1 Step	38.626	5	.000
Block	38.626	5	.000
Model	38.626	5	.000

The next table includes the Pseudo  $R^2$ ; the -2 log likelihood is the minimization criteria used by SPSS. We see that Nagelkerke's  $R^2$  is 0.409, which indicates that the model is good but not great. Cox & Snell's  $R^2$  is the nth root (in our case the 107<sup>th</sup> of the -2log likelihood improvement. Thus we can interpret this as 30% probability of the event passing the exam is explained by the logistic model.

### Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	105.559 <sup>a</sup>	.303	.409

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than .001

The next table contains the classification results, with almost 80% correct classification the model is not too bad – generally a discriminant analysis is better in classifying data correctly.

**Classification Table<sup>a</sup>**

Observed			Predicted		
			Exam		Percentage Correct
			Fail	Pass	
Step 1	Exam	Fail	53	11	82.8
		Pass	11	32	74.4
Overall Percentage					79.4

a. The cut value is .500

The last table is the most important one for our logistic regression analysis. It shows the regression function  $-1.898 + .148*x_1 - .022*x_2 - .047*x_3 - .052*x_4 + .011*x_5$ . The table also includes the test of significance for each of the coefficients in the logistic regression model. For small samples the t-values are not valid and the Wald statistic should be used instead. Wald is basically  $t^2$  which is Chi-Square distributed with  $df=1$ . However, SPSS gives the significance levels of each coefficient. As we can see, only Apt1 is significant—all other variables are not.

**Variables in the Equation**

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 <sup>a</sup>						
Apt1	.148	.038	15.304	1	.000	1.159
Apt2	-.022	.036	.358	1	.549	.979
Apt3	-.047	.035	1.784	1	.182	.954
Apt4	-.052	.043	1.486	1	.223	.949
Apt5	.011	.034	.102	1	.749	1.011
Constant	-1.898	2.679	.502	1	.479	.150

a. Variable(s) entered on step 1: Apt1, Apt2, Apt3, Apt4, Apt5.

### Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	108.931 <sup>a</sup>	.281	.379

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

### Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 <sup>a</sup> Apt1	.158	.033	23.032	1	.000	1.172
Constant	-5.270	1.077	23.937	1	.000	.005

a. Variable(s) entered on step 1: Apt1.

If we change the method from *Enter* to *Forward: Wald* the quality of the logistic regression improves. Now only the significant coefficients are included in the logistic regression equation. In our case the model simplifies to Aptitude Test Score 1 and the intercept. Then we get the logistic equation

$$p = \frac{1}{1 + e^{(-5.270 + .158 Apt1)}}$$
 This equation is easier to interpret, because we know now that a score of one point higher score on the Aptitude Test 1 multiplies the odds of passing the exam by 1.17 ( $\exp(.158)$ ). We can also calculate the critical value for  $p = 50\%$ , which is  $Apt1 = -\text{intercept}/\text{coefficient} = -5.270/.158 = 33.35$ . That is if a pupil scored higher than 33.35 on the Aptitude Test 1 the logistic regression predicts that this pupil will pass the final exam.

In summary, a possible write-up could read as follows:

*We conducted a logistic regression to predict whether a student will pass the final exam based on the five aptitude scores the student achieved. The stepwise logistic regression model finds only the Aptitude Test 1 to be of relevant explanatory power. The logistic equation indicates that an additional score point on the Aptitude Test 1 multiplies the odds of passing by 1.17. Also we predict that students who scored higher than 33.35 on the Aptitude Test will pass the final exam.*

## Ordinal Regression

### What is Ordinal Regression?

Ordinal regression is a member of the family of regression analyses. As a predictive analysis, ordinal regression describes data and explains the relationship between one dependent variable and two or

more independent variables. In ordinal regression analysis, the dependent variable is ordinal (statistically it is polytomous ordinal) and the independent variables are ordinal or continuous-level (ratio or interval).

Sometimes the dependent variable is also called response, endogenous variable, prognostic variable or regressand. The independent variables are also called exogenous variables, predictor variables or regressors.

Linear regression estimates a line to express how a change in the independent variables affects the dependent variables. The independent variables are added linearly as a weighted sum of the form  $Y = \beta_0 + \beta_1 \cdot X_{i1} + \beta_2 \cdot X_{i2} + \dots + \beta_p \cdot X_{ip}$ . Linear regression estimates the regression coefficients by minimizing the sum of squares between the left and the right side of the regression equation. Ordinal regression however is a bit trickier. Let us consider a linear regression of income = 15,000 + .980 \* age. We know that for a 30 year old person the expected income is 44,400 and for a 35 year old the income is 49,300. That is a difference of 4,900. We also know that if we compare a 55 year old with a 60 year old the difference of 68,900-73,800 = 4,900 is exactly the same difference as the 30 vs. 35 year old. This however is not always true for measures that have ordinal scale. For instance if we classify the income to be low, medium, high, it is impossible to say if the difference between low and medium is the same as between medium and high, or if 3\*low = high.

There are three major uses for Ordinal Regression Analysis: 1) causal analysis, 2) forecasting an effect, and 3) trend forecasting. Other than correlation analysis for ordinal variables (e.g., Spearman), which focuses on the strength of the relationship between two or more variables, ordinal regression analysis assumes a dependence or causal relationship between one or more independent and one dependent variable. Moreover the effect of one or more covariates can be accounted for.

Firstly, ordinal regression might be used to identify the strength of the effect that the independent variables have on a dependent variable. A typical question is, *“What is the strength of relationship between dose (low, medium, high) and effect (mild, moderate, severe)?”*

Secondly, ordinal regression can be used to forecast effects or impacts of changes. That is, ordinal regression analysis helps us to understand how much will the dependent variable change, when we change the independent variables. A typical question is, *“When is the response most likely to jump into the next category?”*

Finally, ordinal regression analysis predicts trends and future values. The ordinal regression analysis can be used to get point estimates. A typical question is, *“If I invest a medium study effort what grade (A-F) can I expect?”*

### **The Ordinal Regression in SPSS**

---

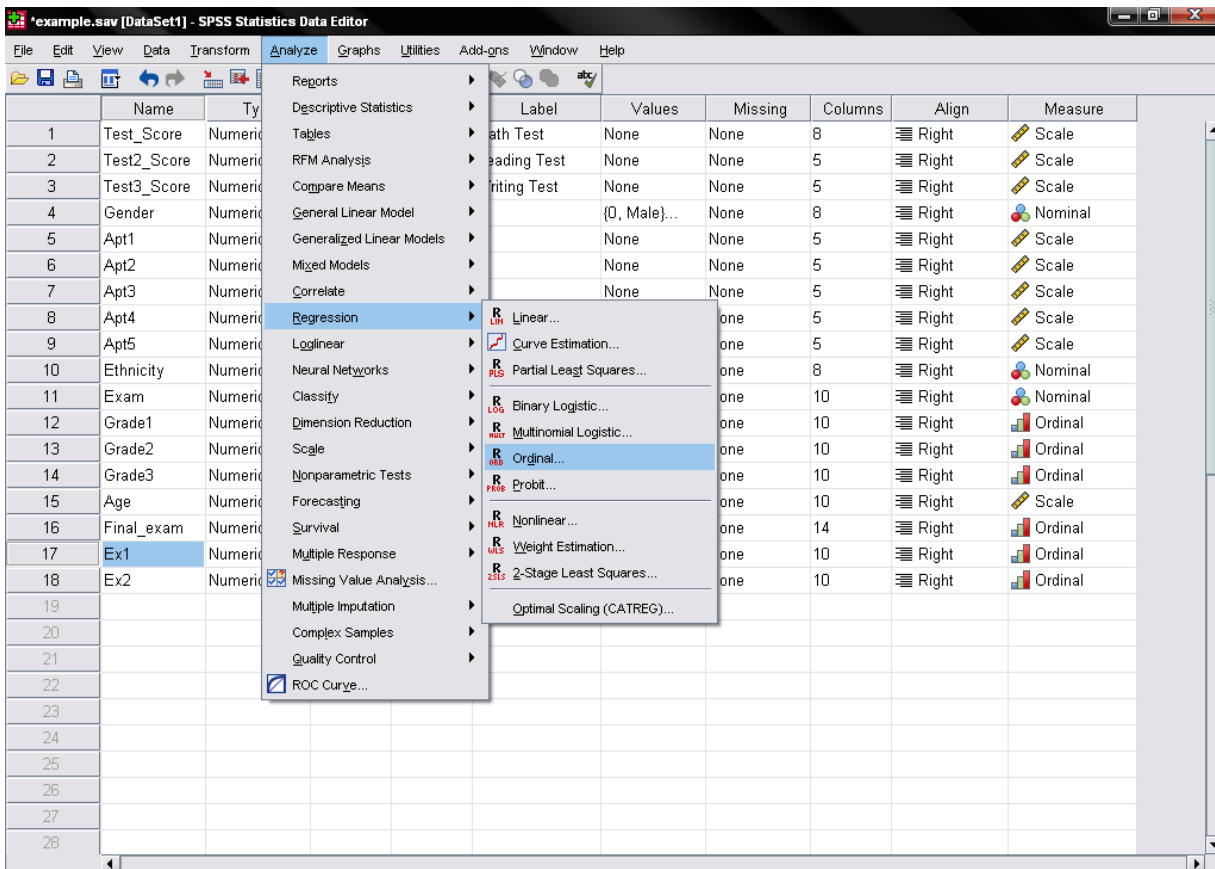
For ordinal regression, let us consider the research question:

In our study the 107 students have been given six different tests. The pupils either failed or passed the first five tests. For the final exam, the students got graded either as fail, pass, good

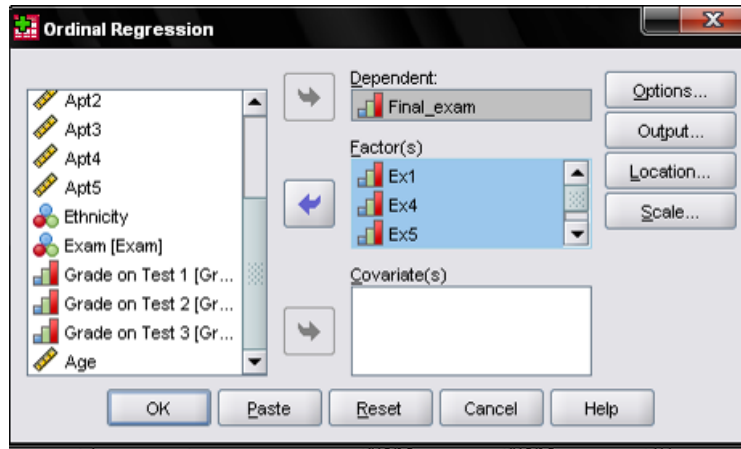
or distinction. We now want to analyze how the first five tests predict the outcome of the final exam.

To answer this we need to use ordinal regression to analyze the question above. Although technically this method is not ideal because the observations are not completely independent, it best suits the purpose of the research team.

The ordinal regression analysis can be found in *Analyze/Regression/Ordinal...*

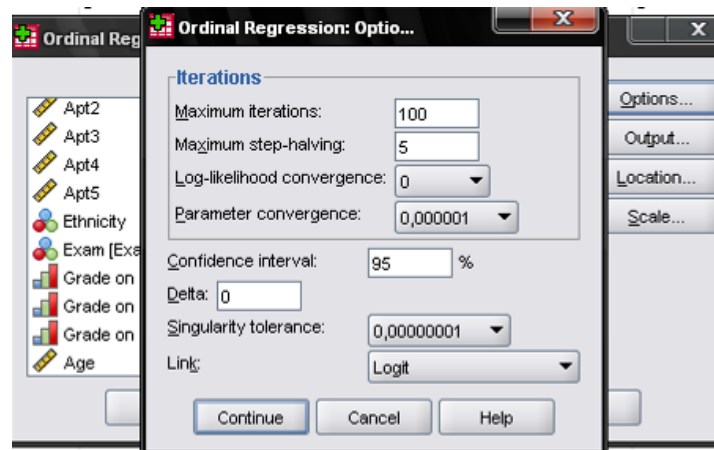


The next dialog box allows us to specify the ordinal regression model. For our example the final exam (four levels – fail, pass, good, distinction) is the dependent variable, the five factors are Ex1 ... Ex5 for the five exams taken during the term. Please note that this works correctly only if the right measurement scales have been defined within SPSS.



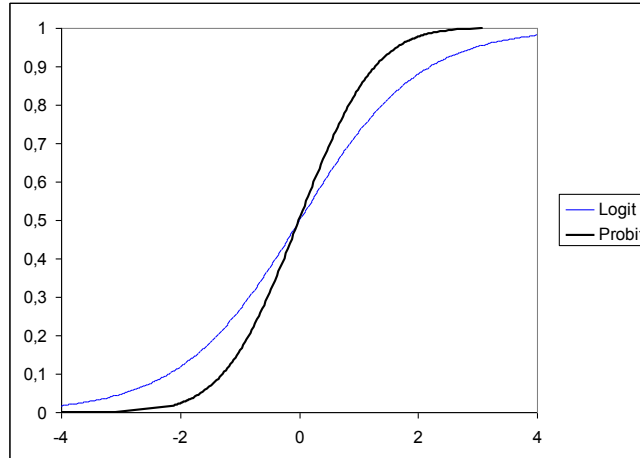
Furthermore, SPSS offers the option to include one or more covariates of continuous-level scale (interval or ratio). However, adding more than one covariate typically results in a large cell probability matrix with a large number of empty cells.

The options dialog allows us to manage various settings for the iteration solution, more interestingly here we can also change the link setting for the ordinal regression. In ordinal regression the link function is a transformation of the cumulative probabilities of the ordered dependent variable that allows for estimation of the model. There are five different link functions.



1. **Logit function:** Logit function is the default function in SPSS for ordinal regression. This function is usually used when the dependent ordinal variable has equal categories. Mathematically, logit function equals to  $p(z) = \ln(z / (1 - z))$ .

2. **Probit model:** This is the inverse standard normal cumulative distribution function. This function is more suitable when a dependent variable is normally distributed. Mathematically, the probit function is  $\rho(z) = \Phi(z)$ .



Both models (logit and probit) are most commonly used in ordinal regression, in most cases a model is fitted with both functions and the function with the better fit is chosen. However, probit assumes normal distribution of the probability of the categories of the dependent variable, when logit assumes the log distribution. Thus the difference between logit and probit is typically seen in small samples.

3. **Negative log-log:** This link function is recommended when the probability of the lower category is high. Mathematically the negative log-log is  $p(z) = -\log(-\log(z))$ .

4. **Complementary log-log:** This function is the inverse of the negative log-log function. This function is recommended when the probability of higher category is high. Mathematically complementary log-log is  $p(z) = \log(-\log(1 - z))$ .

5. **Cauchit:** This link function is used when the extreme values are present in the data. Mathematically Cauchit is  $p(z) = \tan(p(z - 0.5))$ .

We leave the ordinal regression's other dialog boxes at their default settings; we just add the test of parallel lines in the *Output* menu.

### **The Output of the Ordinal Regression Analysis**

The most interesting ordinal regression output is the table with the parameter estimates. The thresholds are the intercepts or first order effects in our ordinal regression. They are typically of limited interest. More information can be found at the Location estimates, which are the coefficients for each independent variable. So for instance if we look at the first Exam (Ex1) we find the higher the score (as in pass the first exam) the higher the score in the final exam. If we calculate the  $\exp(\text{location})$  we get  $\exp(1.886) = 6.59$ , which is our odds ratio and means that if you pass the first Exam it is 6.59 more likely to pass the final exam. This ratio is assumed constant for all outcomes of the dependent variable.



**Parameter Estimates**

		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Threshold	[Final_exam = 1,00]	-1,313	,577	5,177	1	,023	-2,445	-,182
	[Final_exam = 2,00]	,170	,560	,093	1	,761	-,927	1,267
	[Final_exam = 3,00]	1,729	,576	8,999	1	,003	,599	2,859
Location	[Ex1=1,00]	-1,886	,439	18,409	1	,000	-2,747	-1,024
	[Ex1=2,00]	0 <sup>a</sup>	.	.	0	.	.	.
	[Ex4=1,00]	,943	,397	5,656	1	,017	,166	1,720
	[Ex4=2,00]	0 <sup>a</sup>	.	.	0	.	.	.
	[Ex5=1,00]	,017	,397	,002	1	,965	-,760	,795
	[Ex5=2,00]	0 <sup>a</sup>	.	.	0	.	.	.
	[Ex2=1,00]	,455	,378	1,454	1	,228	-,285	1,195
	[Ex2=2,00]	0 <sup>a</sup>	.	.	0	.	.	.
	[Ex3=1,00]	,728	,380	3,670	1	,055	-,017	1,473
	[Ex3=2,00]	0 <sup>a</sup>	.	.	0	.	.	.

Link function: Logit.

a. This parameter is set to zero because it is redundant.

The Wald ratio is defined as  $(\text{coefficient}/\text{standard error})^2$  and is the basis for the test of significance (null hypothesis: the coefficient is zero). We find that Ex1 and Ex4 are significantly different from zero. Therefore there seems to be a relationship between pupils performing on Ex1 and Ex4 and their final exam scores.

The next interesting table is the test for parallel lines. It tests the null hypothesis that the lines run parallel. Our test is not significant and thus we cannot reject the null hypothesis. A significant test typically indicates that the ordinal regression model uses the wrong link function.

**Test of Parallel Lines<sup>a</sup>**

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Null Hypothesis	140,139			
General	134,516	5,622	10	,846

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

a. Link function: Logit.

### Multinomial Logistic Regression

#### What is Multinomial Logistic Regression?

Multinomial regression is the linear regression analysis to conduct when the dependent variable is nominal with more than two levels. Thus it is an extension of logistic regression, which analyzes dichotomous (binary) dependents. Since the SPSS output of the analysis is somewhat different to the logistic regression's output, multinomial regression is sometimes used instead.

Like all linear regressions, the multinomial regression is a predictive analysis. Multinomial regression is used to describe data and to explain the relationship between one dependent nominal variable and one or more continuous-level(interval or ratio scale) independent variables.

Standard linear regression requires the dependent variable to be of continuous-level(interval or ratio) scale. Logistic regression jumps the gap by assuming that the dependent variable is a stochastic event. And the dependent variable describes the outcome of this stochastic event with a density function (a function of cumulated probabilities ranging from 0 to 1). Statisticians then argue one event happens if the probability is less than 0.5 and the opposite event happens when probability is greater than 0.5.

How can we apply the logistic regression principle to a multinomial variable (e.g. 1/2/3)?

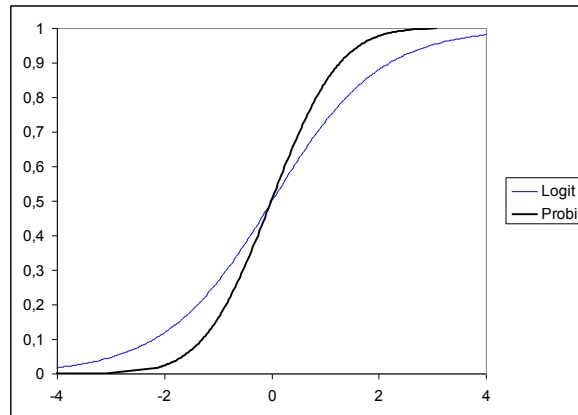
Example:

*We analyze our class of pupils that we observed for a whole term. At the end of the term we gave each pupil a computer game as a gift for their effort. Each participant was free to choose between three games – an action, a puzzle or a sports game. The researchers want to know how the initial baseline (doing well in math, reading, and writing) affects the choice of the game. Note that the choice of the game is a nominal dependent variable with more than two levels. Therefore multinomial regression is the best analytic approach to the question.*

How do we get from logistic regression to multinomial regression? Multinomial regression is a multi-equation model, similar to multiple linear regression. For a nominal dependent variable with k categories the multinomial regression model estimates k-1 logit equations. Although SPSS does compare all combinations of k groups it only displays one of the comparisons. This is typically either the first or the last category. The multinomial regression procedure in SPSS allows selecting freely one group to compare the others with.

What are logits? The basic idea behind logits is to use a logarithmic function to restrict the probability values to (0,1). Technically this is the log odds (the logarithmic of the odds of  $y = 1$ ). Sometimes a probit model is used instead of a logit model for multinomial regression. The following graph shows the difference for a logit and a probit model for different values (-4,4). Both models are commonly used as the link function in ordinal regression. However, most multinomial regression models are based on the

logit function. The difference between both functions is typically only seen in small samples because probit assumes normal distribution of the probability of the event, when logit assumes the log distribution.



At the center of the multinomial regression analysis is the task estimating the k-1 log odds of each category. In our k=3 computer game example with the last category as reference multinomial regression estimates k-1 multiple linear regression function defined as

$$\text{logit}(y=1) = \log\left(\frac{p(y=1)}{1-(p=1)}\right) = \beta_0 + \beta_1 \cdot x_{i2} + \beta_2 \cdot x_{i2} + \dots + \beta_p \cdot x_{in} \text{ for } i = 1 \dots n.$$

$$\text{logit}(y=2) = \log\left(\frac{p(y=2)}{1-(p=2)}\right) = \beta_0 + \beta_1 \cdot x_{i2} + \beta_2 \cdot x_{i2} + \dots + \beta_p \cdot x_{in} \text{ for } i = 1 \dots n.$$

Multinomial regression is similar to the Multivariate Discriminant Analysis. Discriminant analysis uses the regression line to split a sample in two groups along the levels of the dependent variable. In the case of three or more categories of the dependent variable multiple discriminant equations are fitted through the scatter cloud. In contrast multinomial regression analysis uses the concept of probabilities and k-1 log odds equations that assume a cut-off probability 0.5 for a category to happen. The practical difference is in the assumptions of both tests. If the data is multivariate normal, homoscedasticity is present in variance and covariance and the independent variables are linearly related, then we should use discriminant analysis because it is more statistically powerful and efficient. Discriminant analysis is also more accurate in predictive classification of the dependent variable than multinomial regression.

### **The Multinomial Logistic Regression in SPSS**

For multinomial logistic regression, we consider the following research question:

*We conducted a research study with 107 students. The students were measured on a standardized reading, writing, and math test at the start of our study. At the end of the study, we offered every pupil a computer game as a thank you gift. They were free to choose one of*

*three games – a sports game, a puzzle and an action game. How does the pupils' ability to read, write, or calculate influence their game choice?*

First we need to check that all cells in our model are populated. Although the multinomial regression is robust against multivariate normality and therefore better suited for smaller samples than a probit model, we still need to check. We find that some of the cells are empty. We must therefore collapse some of the factor levels. The easiest way to check is to create the contingency table (*Analyze/Descriptive Statistics/Crosstabs...*).

**Grade on Math Test \* Gift chosen by pupil Crosstabulation**

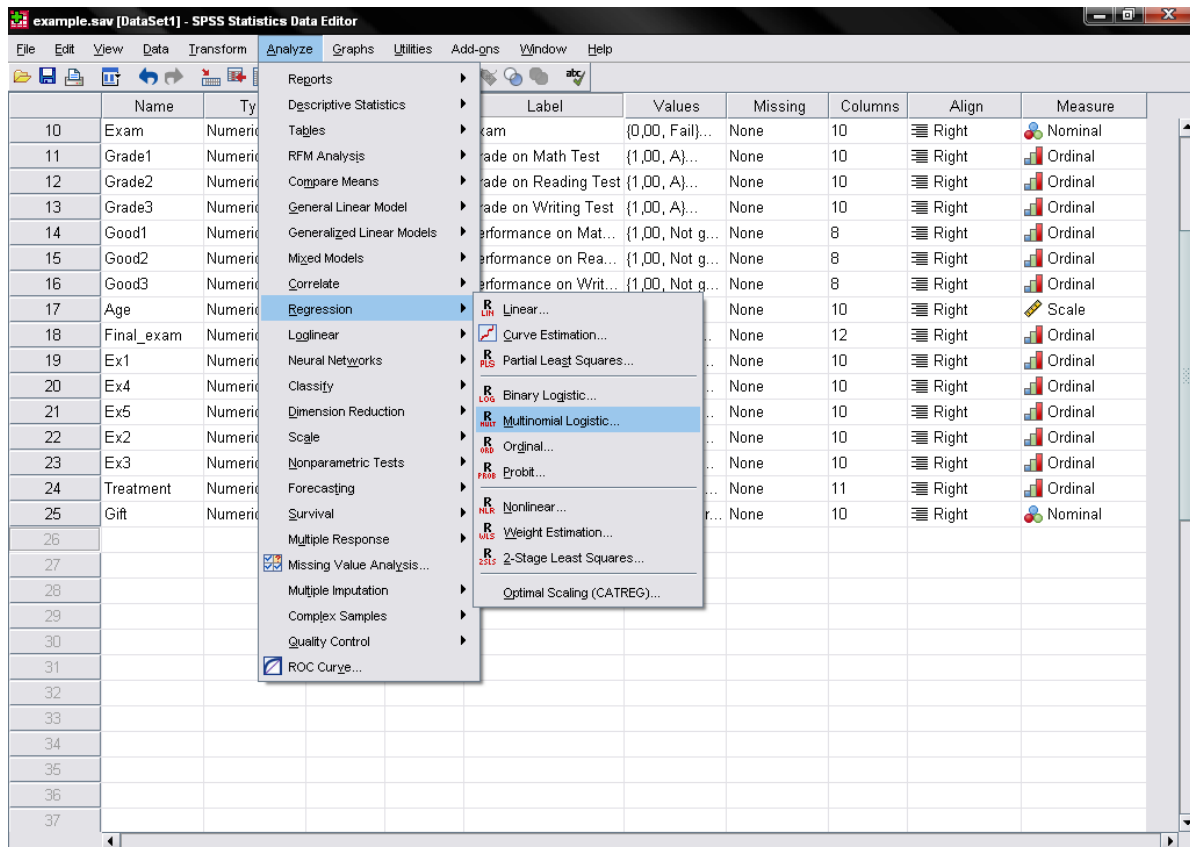
Count		Gift chosen by pupil			Total
		Superblaster	Puzzle Mania	Polar Bear Olympics	
Grade on Math Test	A	0	3	14	17
	B	0	6	12	18
	C	5	9	4	18
	D	8	5	5	18
	E	13	3	0	16
	F	14	5	1	20
Total		40	31	36	107

But even if we collapse the factor levels of our multinomial regression model down to two levels (performance good vs. not good) we observe empty cells. We proceed with the analysis regardless, noting and reporting this limitation of our analysis.

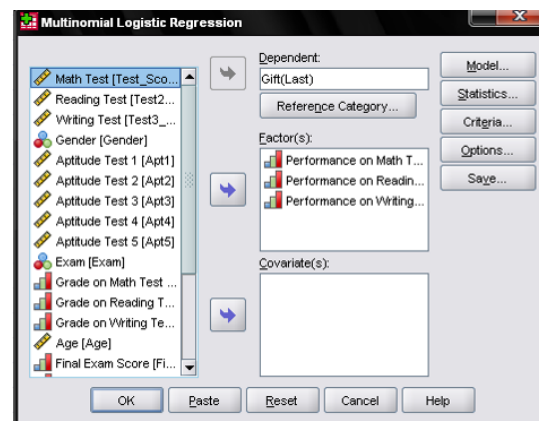
**Performance on Reading Test \* Gift chosen by pupil Crosstabulation**

Count		Gift chosen by pupil			Total
		Superblaster	Puzzle Mania	Polar Bear Olympics	
Performance on Reading Test	Not good	40	16	0	56
	Good	0	15	36	51
Total		40	31	36	107

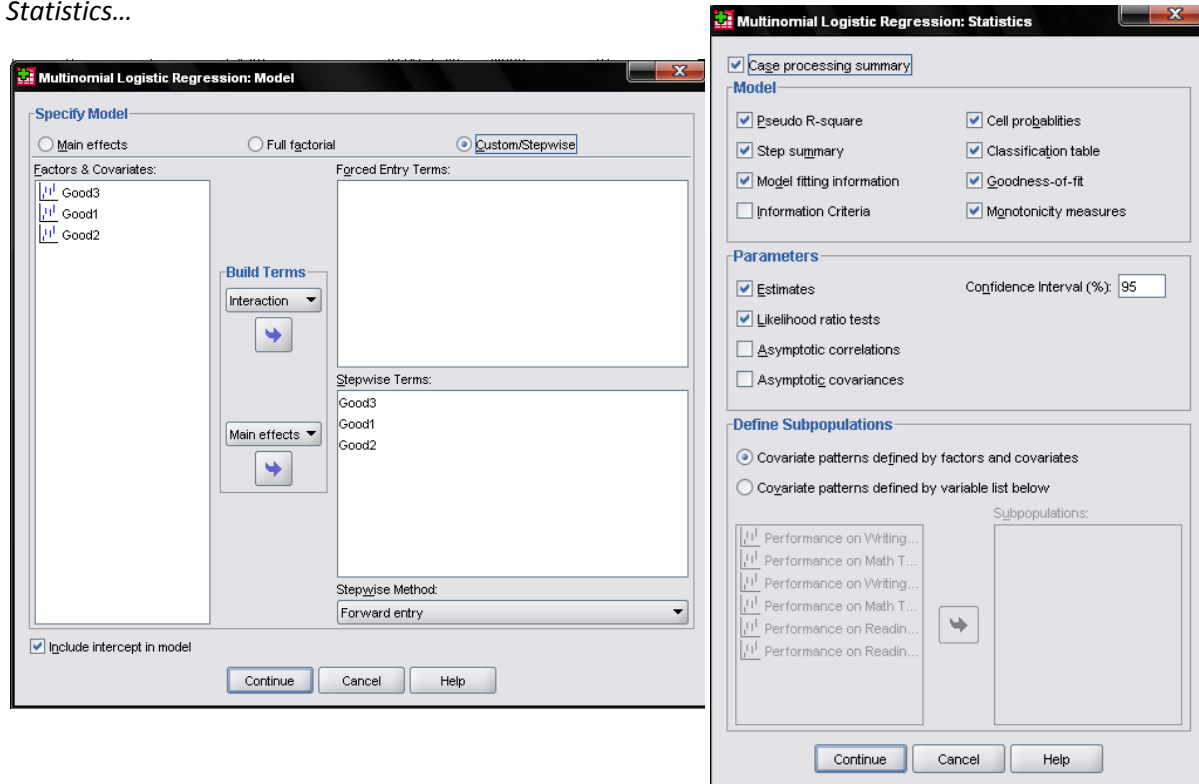
Multinomial Regression is found in SPSS under *Analyze/Regression/Multinomial Logistic...*



This opens the dialog box to specify the model. Here we need to enter the dependent variable *Gift* and define the reference category. In our example it will be the last category since we want to use the sports game as a baseline. Then we enter the three collapsed factors into the multinomial regression model. The factors are performance (good vs. not good) on the math, reading, and writing test.



In the menu *Model...* we need to specify the model for the multinomial regression. The huge advantage over ordinal regression analysis is the ability to conduct a stepwise multinomial regression for all main and interaction effects. If we want to include additional measures about the multinomial regression model to the output we can do so in the dialog box *Statistics...*



### **The Output of the Multinomial Logistic Regression Analysis**

The first table in the output of the multinomial regression analysis describes the model design.

Case Processing Summary		N	Marginal Percentage
Gift chosen by pupil	Superblaster	40	37.4%
	Puzzle Mania	31	29.0%
	Polar Bear Olympics	36	33.6%
Performance on Math Test	Not good	54	50.5%
	Good	53	49.5%
Performance on Reading Test	Not good	56	52.3%
	Good	51	47.7%
Performance on Writing test	Not good	56	52.3%
	Good	51	47.7%
Valid		107	100.0%
Missing		0	
Total		107	
Subpopulation		8 <sup>a</sup>	

a. The dependent variable has only one value observed in 5 (62.5%) subpopulations.

The next table details which variables are entered into the multinomial regression. Remember that we selected a stepwise model. In our example the writing test results (good3) and then the reading test results (good2) were entered. Also the 0 model is shown as the  $-2 \cdot \log(\text{likelihood})$  change between models is used for significance testing and calculating the Pseudo- $R^2$ s.

Step Summary						
Model	Action	Effect(s)	Model Fitting Criteria	Effect Selection Tests		
			-2 Log Likelihood	Chi-Square <sup>a</sup>	df	Sig.
0	Entered	Intercept	216.336	.		
1	Entered	Good3	111.179	105.157	2	.000
2	Entered	Good2	7.657	103.522	2	.000

Stepwise Method: Forward Entry

a. The chi-square for entry is based on the likelihood ratio test.

The next 3 tables contain the goodness of fit criteria. As we find the goodness-of-fit (chi-square test of the null hypothesis that the coefficients are different from zero) is not significant and Nagelkerke's  $R^2$  is close to 1. Remember that Cox & Snell's  $R^2$  does not scale up to 1. Cox & Snell's  $R^2$  is the nth root (in our case the 107<sup>th</sup>) of the  $-2\log$  likelihood improvement. We can interpret the Pseudo- $R^2$  as our multinomial regression model explains 85.6% of the probability that a given computer game is chosen by the pupil.

Model Fitting Information				
Model	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	216.336			
Final	7.657	208.679	4	.000

Goodness-of-Fit			
	Chi-Square	df	Sig.
Pearson	2.386	10	.992
Deviance	1.895	10	.997

Pseudo R-Square	
Cox and Snell	.858
Nagelkerke	.966
McFadden	.892

The classification table shows that the estimated multinomial regression functions correctly classify 97.2% of the events. Although this is sometimes reported, it is a less powerful goodness of fit test than Pearson's or Deviance.

Classification				
Observed	Predicted			
	Superblaster	Puzzle Mania	Polar Bear Olympics	Percent Correct
Superblaster	40	0	0	100.0%
Puzzle Mania	2	28	1	90.3%
Polar Bear Olympics	0	0	36	100.0%
Overall Percentage	39.3%	26.2%	34.6%	97.2%

The most important table for our multinomial regression analysis is the Parameter Estimates table. It includes the coefficients for the two logistic regression functions. The table also includes the test of significance for each of the coefficients in the multinomial regression model. For small samples the t-values are not valid and the Wald statistic should be used instead. However, SPSS gives the significance levels of each coefficient. As we can see, only Apt1 is significant and all other variables are not.



Parameter Estimates

Gift chosen by pupil <sup>a</sup>		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp(B)	
								Lower Bound	Upper Bound
Superblaster	Intercept	-48.123	8454.829	.000	1	.995			
	[Good3=.00]	48.030	.000	.	1	.	7.227E20	7.227E20	7.227E20
	[Good3=1.00]	0 <sup>b</sup>	.	.	0	.	.	.	.
	[Good2=.00]	48.030	.000	.	1	.	7.227E20	7.227E20	7.227E20
	[Good2=1.00]	0 <sup>b</sup>	.	.	0	.	.	.	.
Puzzle Mania	Intercept	-3.584	1.014	12.495	1	.000			
	[Good3=.00]	24.262	5978.481	.000	1	.997	3.442E10	.000	. <sup>c</sup>
	[Good3=1.00]	0 <sup>b</sup>	.	.	0	.	.	.	.
	[Good2=.00]	24.262	5978.481	.000	1	.997	3.442E10	.000	. <sup>c</sup>
	[Good2=1.00]	0 <sup>b</sup>	.	.	0	.	.	.	.

a. The reference category is: Polar Bear Olympics.

b. This parameter is set to zero because it is redundant.

c. Floating point overflow occurred while computing this statistic. Its value is therefore set to system missing.

In this analysis the parameter estimates are quite wild because we collapsed our factors to binary level for the lack of sample size. This results in the standard error either skyrocketing or dropping. The intercept is the multinomial regression estimate for all other values being zero. The coefficient for Good3 is 48.030. So, if a pupil were to increase his score on Test 3 by one unit—that is, he jumps from fail to pass because of our collapsing—the log-odds of preferring action over the sports game would decrease by -48.030. In other words, pupils that fail Test 2 and 3 (variables good2, good3) are more likely to prefer the Superblaster game.

The standard error, Wald statistic, and test of significance are given for each coefficient in our multinomial regression model. Because of our use of binary variables, the standard error is zero and thus the significance is 0 as well. This is a serious limitation of this analysis and should be reported accordingly.

### Sequential One-Way Discriminant Analysis

#### What is the Sequential One-Way Discriminant Analysis?

Sequential one-way discriminant analysis is similar to the one-way discriminant analysis. Discriminant analysis predicts group membership by fitting a linear regression line through the scatter plot. In the case of more than two independent variables it fits a plane through the scatter cloud thus separating all observations in one of two groups—one group to the "left" of the line and one group to the "right" of the line.

Sequential one-way discriminant analysis now assumes that the discriminating, independent variables are not equally important. This might be a suspected explanatory power of the variables, a hypothesis deduced from theory or a practical assumption, for example in customer segmentation studies.

Like the standard one-way discriminant analysis, sequential one-way discriminant analysis is useful mainly for two purposes: 1) identifying differences between groups, and 2) predicting group membership.

Firstly, sequential one-way discriminant analysis identifies the independent variables that significantly discriminate between the groups that are defined by the dependent variable. Typically, sequential one-way discriminant analysis is conducted after a cluster analysis or a decision tree analysis to identify the goodness of fit for the cluster analysis (remember that cluster analysis does not include any goodness of fit measures itself). Sequential one-way discriminant analysis tests whether each of the independent variables has discriminating power between the groups.

Secondly, sequential one-way discriminant analysis can be used to predict group membership. One output of the sequential one-way discriminant analysis is Fisher's discriminant coefficients. Originally Fisher developed this approach to identify the species to which a plant belongs. He argued that instead of going through a whole classification table, only a subset of characteristics is needed. If you then plug in the scores of respondents into these linear equations, the result predicts the group membership. This is typically used in customer segmentation, credit risk scoring, or identifying diagnostic groups.

Because sequential one-way discriminant analysis assumes that group membership is given and that the variables are split into independent and dependent variables, the sequential one-way discriminant analysis is a so called structure testing method as opposed to structure exploration methods (e.g., factor analysis, cluster analysis).

The sequential one-way discriminant analysis assumes that the dependent variable represents group membership the variable should be nominal. The independent variables represent the characteristics explaining group membership.

The independent variables need to be continuous-level(interval or ratio scale). Thus the sequential one-way discriminant analysis is similar to a MANOVA, logistic regression, multinomial and ordinal regression. Sequential one-way discriminant analysis is different than the MANOVA because it works the other way around. MANOVAs test for the difference of mean scores of dependent variables of continuous-level scale (interval or ratio). The groups are defined by the independent variable.

Sequential one-way discriminant analysis is different from logistic, ordinal and multinomial regression because it uses ordinary least squares instead of maximum likelihood; sequential one-way discriminant analysis, therefore, requires smaller samples. Also continuous variables can only be entered as covariates in the regression models; the independent variables are assumed to be ordinal in scale. Reducing the scale level of an interval or ratio variable to ordinal in order to conduct multinomial regression takes out variation from the data and reduces the statistical power of the test. Whereas sequential one-way discriminant analysis assumes continuous variables, logistic/ multinomial/ ordinal

regression assume categorical data and thus use a Chi-Square like matrix structure. The disadvantage of this is that extremely large sample sizes are needed for designs with many factors or factor levels.

Moreover, sequential one-way discriminant analysis is a better predictor of group membership if the assumptions of multivariate normality, homoscedasticity, and independence are met. Thus we can prevent over-fitting of the model, that is to say we can restrict the model to the relevant independent variables and focus subsequent analyses. Also, because it is an analysis of the covariance, we can measure the discriminating power of a predictor variable when removing the effects of the other independent predictors.

### **The Sequential One-Way Discriminant Analysis in SPSS**

---

The research question for the sequential one-way discriminant analysis is as follows:

*The students in our sample were taught with different methods and their ability in different tasks was repeatedly graded on aptitude tests and exams. At the end of the study the pupils go to chose from three computer game 'thank you' gifts: a sports game (Superblaster), a puzzle game (Puzzle Mania) and an action game (Polar Bear Olympics). The researchers wish to learn what guided the pupils' choice of gift.*

The independent variables are the three test scores from the standardized mathematical, reading, writing test (viz. Test\_Score, Test2\_Score, and Test3\_score). From previous correlation analysis we suspect that the writing and the reading score have the highest influence on the outcome. In our logistic regression we found that pupils scoring lower had higher risk ratios of preferring the action game over the sports or the puzzle game.

The sequential one way discriminant analysis is not a part of the graphical user interface of SPSS. However, if we want include our variables in a specific order into the sequential one-way discriminant model we can do so by specifying the order in the */analysis* subcommand of the Discriminant syntax.

The SPSS syntax for a sequential one-way discriminant analysis specifies the sequence of how to include the variables in the analysis by defining an inclusion level. SPSS accepts inclusion levels from 99...0, where variables with level 0 are never included in the analysis.

DISCRIMINANT

/GROUPS=Gift(1 3)

/VARIABLES=Test\_Score Test2\_Score Test3\_Score

/ANALYSIS Test3\_Score (3), Test2\_Score (2), Test\_Score (1)

/METHOD=WILKS

/FIN=3.84

/FOUT=2.71

/PRIORS SIZE

/HISTORY

/STATISTICS=BOXM COEFF  
/CLASSIFY=NONMISSING POOLED.

### ***The Output of the Sequential One-Way Discriminant Analysis***

The first couple of tables in the output of the sequential one-way discriminant analysis illustrate the model design and the sample size. The first relevant table is Box's M test, which tests the null hypothesis that the covariances of the dependent variable and every given pair of independent variables are equal for all groups in the independent variable. We find that Box's M is not significant therefore we cannot assume equality of covariances. The discriminant analysis is robust against the violation of this assumption.

**Test Results**

Box's M		34.739
F	Approx.	5.627
	df1	6
	df2	205820.708
	Sig.	.000

Tests null hypothesis of equal population covariance matrices.

The next table shows the variables entered in each step of the sequential one-way discriminant analysis.

**Variables Entered/Removed<sup>a,b,c,d</sup>**

Step	Entered	Wilks' Lambda							
		Statistic	df1	df2	df3	Exact F			
						Statistic	df1	df2	Sig.
1	Writing Test	.348	1	2	104.000	97.457	2	104.000	.000
2	Reading Test	.150	2	2	104.000	81.293	4	206.000	.000

At each step, the variable that minimizes the overall Wilks' Lambda is entered.

- a. Maximum number of steps is 4.
- b. Minimum partial F to enter is 3.84.
- c. Maximum partial F to remove is 2.71.
- d. F level, tolerance, or VIN insufficient for further computation.

We find that the writing test score was first entered, followed by the reading test score (based on the Wilks' Lambda). The third variable we specified, the math test score, was not entered because it did not explain anymore variance of the data. It also shows the significance of each variable by running the F-test for the specified model.

Eigenvalues				
Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	5.601 <sup>a</sup>	99.9	99.9	.921
2	.007 <sup>a</sup>	.1	100.0	.085

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

The next few tables show the variables in the analysis and the variables not in the analysis and Wilk's Lambda. All of these tables contain virtually the same data. The next table shows the discriminant eigenvalues. The eigenvalues are defined as  $\gamma = \frac{SS_b}{SS_w}$  and are maximized using ordinary least squares.

We find that the first function explains 99.9% of the variance and the second function explains the rest. This is quite unusual for a discriminant model. This table also shows the canonical correlation

coefficient for the sequential discriminant analysis that is defined as  $c = \sqrt{\frac{\gamma}{1+\gamma}}$ .

The next table in the output of our sequential one-way discriminant function describes the standardized canonical discriminant coefficient—these are the estimated Beta coefficients. Since we do have more than two groups in our analysis we need at least two functions (each canonical discriminant function can differentiate between two groups). We see that

$$Y_1 = .709 * \text{Writing Test} + .827 * \text{Reading Test}$$

$$Y_2 = .723 * \text{Writing Test} - .585 * \text{Reading Test}$$

**Standardized Canonical Discriminant  
Function Coefficients**

	Function	
	1	2
Writing Test	.709	.723
Reading Test	.827	-.585

This however has no inherent meaning other than knowing that a high score on both tests gives function 1 a high value, while simultaneously giving function 2 a lower value. In interpreting this table, we need to look at the group centroids of our one-way sequential discriminant analysis at the same time.

**Functions at Group Centroids**

Gift chosen by pupil	Function	
	1	2
Superblaster	-2.506	-.060
Puzzle Mania	-.276	.131
Polar Bear Olympics	3.023	-.045

Unstandardized canonical discriminant functions  
evaluated at group means

We find that a high score of three on the first function indicates a preference for the sports game, a score close to zero indicates a preference for the puzzle game, and a low score indicates a preference for the action game. Remember that this first function explained 99.9% of our variance in the data. We also know that the sequential one-way discriminant function 1 scored higher for high results in the writing and the reading tests, whereby reading was a bit more important than writing.

**Classification Function Coefficients**

	Gift chosen by pupil		
	Superblaster	Puzzle Mania	Polar Bear Olympics
Writing Test	.151	.403	.727
Reading Test	.206	.464	.885
(Constant)	-2.249	-8.521	-26.402

Fisher's linear discriminant functions

Thus we can say that pupils who did well on our reading and writing test are more likely to choose the sports game, and pupils who did not do well on the tests are more likely to choose the action game.

The final interesting table in the sequential one-way discriminant function output is the classification coefficient table. Fisher's classification coefficients can be used to predict group membership.

In our case we get three functions:

*Superblaster* =  $-2.249 + .151 * \text{writing} + .206 * \text{reading}$

*Puzzle Mania* =  $-8.521 + .403 * \text{writing} + .464 * \text{reading}$

*Polar Bear Olympics* =  $-26.402 + .727 * \text{writing} + .885 * \text{reading}$

If we would plug in the numbers of a new student joining class who score 40 on both tests we would get 3 scores:

*Superblaster* = 12.031

*Puzzle Mania* = 26.159

*Polar Bear Olympics* = 38.078

Thus the student would most likely choose the Polar Bear Olympics (the highest value predicts the group membership).

The table classification results show that specifically in the case where we predicted that the student would choose the sports game, 13.9% chose the puzzle game instead. This serves to alert us to the risk behind this classification function.

**Classification Results<sup>a</sup>**

			Predicted Group Membership			Total
			Superblaster	Puzzle Mania	Polar Bear Olympics	
Original	Count	Superblaster	36	4	0	40
		Puzzle Mania	1	28	2	31
		Polar Bear Olympics	0	5	31	36
	%	Superblaster	90.0	10.0	.0	100.0
		Puzzle Mania	3.2	90.3	6.5	100.0
		Polar Bear Olympics	.0	13.9	86.1	100.0

a. 88,8% of original grouped cases correctly classified.

## Cluster Analysis

### What is the Cluster Analysis?

The Cluster Analysis is an explorative analysis that tries to identify structures within the data. Cluster analysis is also called segmentation analysis or taxonomy analysis. More specifically, it tries to identify homogenous groups of cases, i.e., observations, participants, respondents. Cluster analysis is used to identify groups of cases if the grouping is not previously known. Because it is explorative it does make any distinction between dependent and independent variables. The different cluster analysis methods that SPSS offers can handle binary, nominal, ordinal, and scale (interval or ratio) data.

The Cluster Analysis is often part of the sequence of analyses of factor analysis, cluster analysis, and finally, discriminant analysis. First, a factor analysis that reduces the dimensions and therefore the number of variables makes it easier to run the cluster analysis. Also, the factor analysis minimizes multicollinearity effects. The next analysis is the cluster analysis, which identifies the grouping. Lastly, a discriminant analysis checks the goodness of fit of the model that the cluster analysis found and profiles the clusters. In almost all analyses a discriminant analysis follows a cluster analysis because the cluster analysis does not have any goodness of fit measures or tests of significance. The cluster analysis relies on the discriminant analysis to check if the groups are statistically significant and if the variables significantly discriminate between the groups. However, this does not ensure that the groups are actually meaningful; interpretation and choosing the right clustering is somewhat of an art. It is up to the understanding of the researcher and how well he/she understands and makes sense of his/her data! Furthermore, the discriminant analysis builds a predictive model that allows us to plug in the numbers of new cases and to predict the cluster membership.

Typical research questions the Cluster Analysis answers are as follows:

**Medicine** – What are the diagnostic clusters? To answer this question the researcher would devise a diagnostic questionnaire that entails the symptoms (for example in psychology standardized scales for anxiety, depression etc.). The cluster analysis can then identify groups of patients that present with similar symptoms and simultaneously maximize the difference between the groups.

**Marketing** – What are the customer segments? To answer this question a market researcher conducts a survey most commonly covering needs, attitudes, demographics, and behavior of customers. The researcher then uses the cluster analysis to identify homogenous groups of customers that have similar needs and attitudes but are distinctively different from other customer segments.

**Education** – What are student groups that need special attention? The researcher measures a couple of psychological, aptitude, and achievement characteristics. A cluster analysis then identifies what homogenous groups exist among students (for example, high achievers in all subjects, or students that excel in certain subjects but fail in others, etc.). A discriminant analysis then profiles these performance clusters and tells us what psychological, environmental, aptitudinal, affective, and attitudinal factors characterize these student groups.

**Biology** – What is the taxonomy of species? The researcher has collected a data set of different plants and noted different attributes of their phenotypes. A hierarchical cluster analysis groups those observations into a series of clusters and builds a taxonomy tree of groups and subgroups of similar plants.

Other techniques you might want to try in order to identify similar groups of observations are *Q-analysis*, *multi-dimensional scaling (MDS)*, and *latent class analysis*.

*Q-analysis*, also referred to as *Q factor analysis*, is still quite common in biology but now rarely used outside of that field. *Q-analysis* uses factor analytic methods (which rely on *R*—the correlation between



variables to identify homogenous dimensions of variables) and switches the variables in the analysis for observations (thus changing the R into a Q).

*Multi-dimensional scaling* for scale data (interval or ratio) and correspondence analysis (for nominal data) can be used to map the observations in space. Thus, it is a graphical way of finding groupings in the data. In some cases MDS is preferable because it is more relaxed regarding assumptions (normality, scale data, equal variances and covariances, and sample size).

Lastly, *latent class analysis* is a more recent development that is quite common in customer segmentations. Latent class analysis introduces a dependent variable into the cluster model, thus the cluster analysis ensures that the clusters explain an outcome variable, (e.g., consumer behavior, spending, or product choice).

### **The Cluster Analysis in SPSS**

---

Our research question for the cluster analysis is as follows:

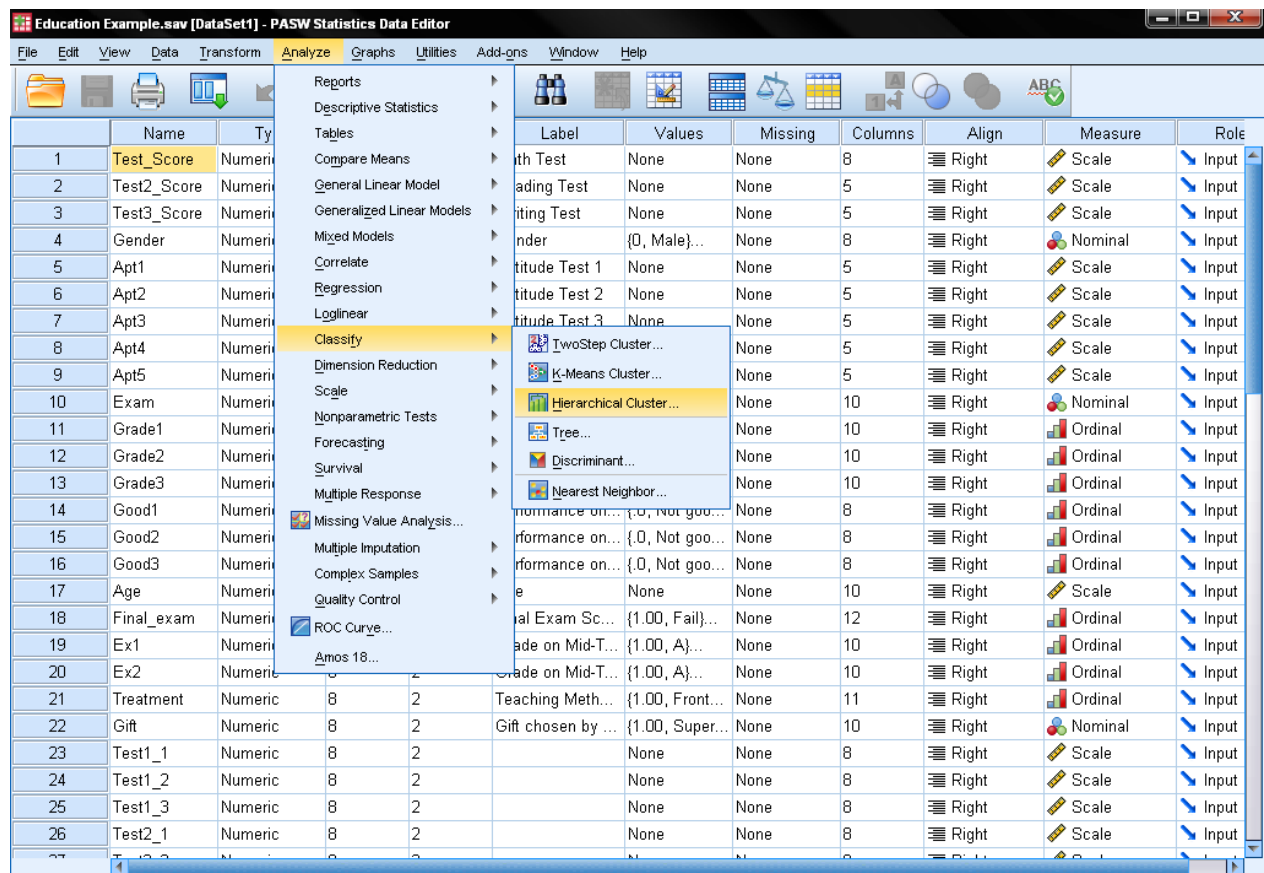
When we examine our standardized test scores in mathematics, reading, and writing, what do we consider to be homogenous clusters of students?

In SPSS Cluster Analyses can be found in *Analyze/Classify....* SPSS offers three methods for the cluster analysis: *K-Means Cluster*, *Hierarchical Cluster*, and *Two-Step Cluster*.

*K-means cluster* is a method to quickly cluster large data sets, which typically take a while to compute with the preferred hierarchical cluster analysis. The researcher must to define the number of clusters in advance. This is useful to test different models with a different assumed number of clusters (for example, in customer segmentation).

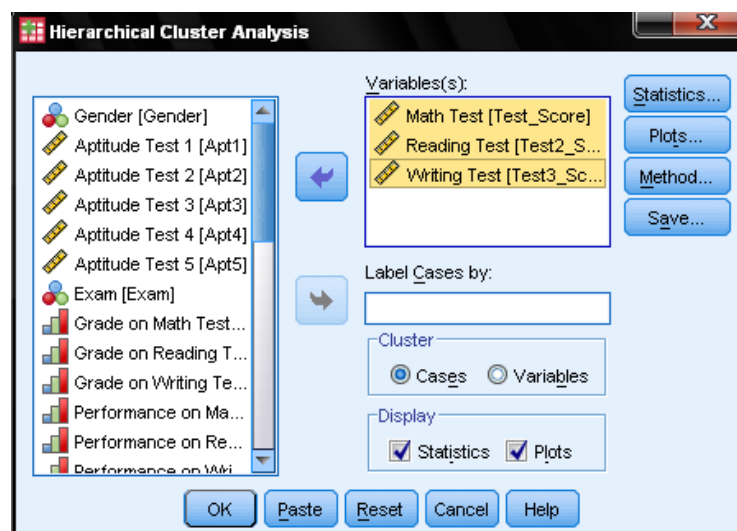
*Hierarchical cluster* is the most common method. We will discuss this method shortly. It takes time to calculate, but it generates a series of models with cluster solutions from 1 (all cases in one cluster) to  $n$  (all cases are an individual cluster). *Hierarchical cluster* also works with variables as opposed to cases; it can cluster variables together in a manner somewhat similar to factor analysis. In addition, *hierarchical cluster* analysis can handle nominal, ordinal, and scale data, however it is not recommended to mix different levels of measurement.

*Two-step cluster* analysis is more of a tool than a single analysis. It identifies the groupings by running pre-clustering first and then by hierarchical methods. Because it uses a quick cluster algorithm upfront, it can handle large data sets that would take a long time to compute with hierarchical cluster methods. In this respect, it combines the best of both approaches. Also *two-step clustering* can handle scale and ordinal data in the same model. *Two-step cluster* analysis also automatically selects the number of clusters, a task normally assigned to the researcher in the two other methods.

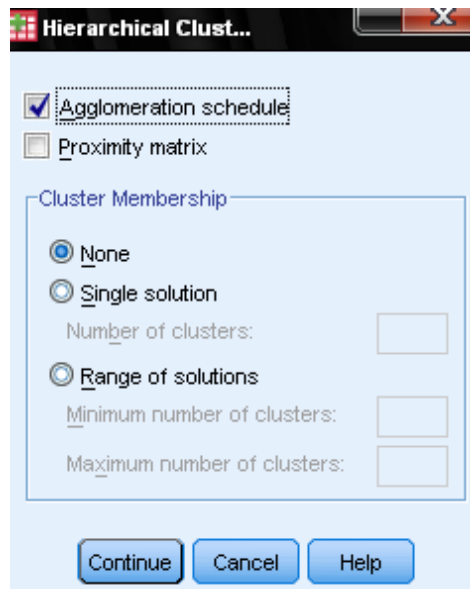


The hierarchical cluster analysis follows three basic steps: 1) calculate the distances, 2) link the clusters, and 3) choose a solution by selecting the right number of clusters.

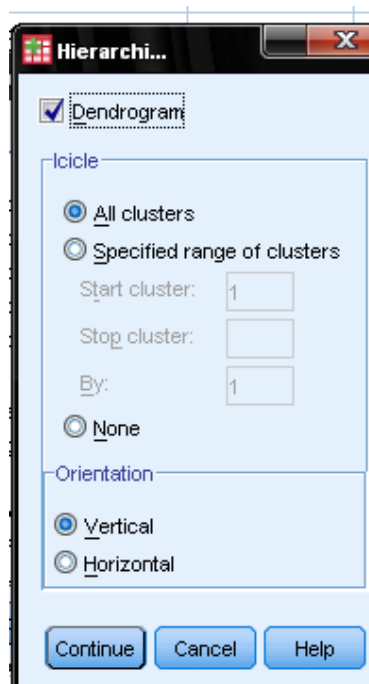
Before we start we have to select the variables upon which we base our clusters. In the dialog we add math, reading, and writing test to the list of variables. Since we want to cluster cases we leave the rest of the tick marks on the default.



In the dialog box *Statistics...* we can specify whether we want to output the proximity matrix (these are the distances calculated in the first step of the analysis) and the predicted cluster membership of the cases in our observations. Again, we leave all settings on default.

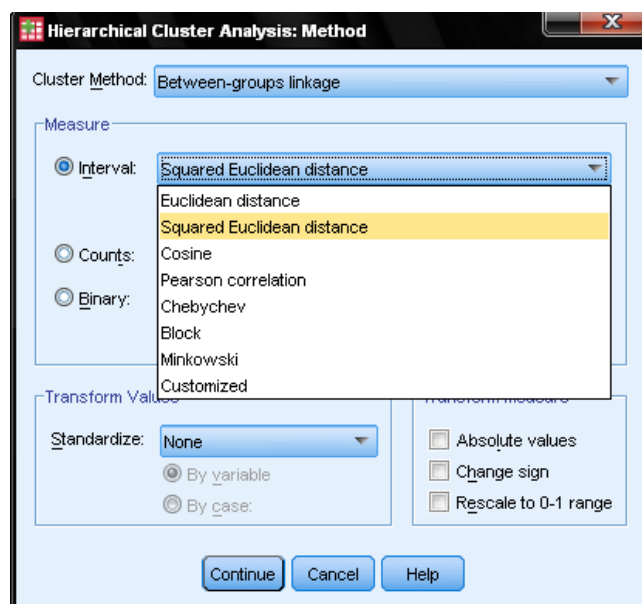


In the dialog box *Plots...* we should add the *Dendrogram*. The *Dendrogram* will graphically show how the clusters are merged and allows us to identify what the appropriate number of clusters is.

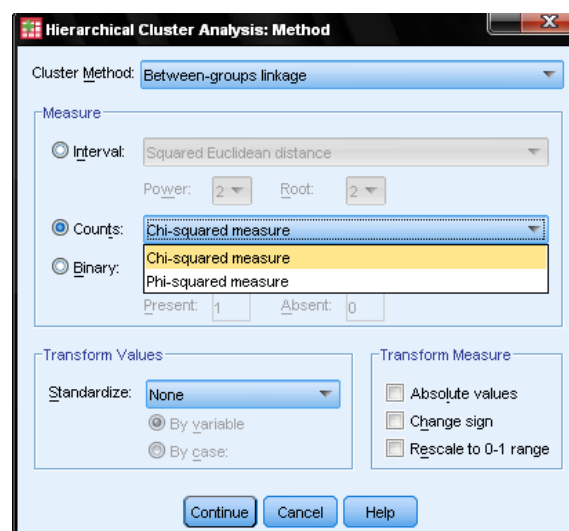


The dialog box *Method...* is very important! Here we can specify the distance measure and the clustering method. First, we need to define the correct distance measure. SPSS offers three large blocks of distance measures for interval (scale), counts (ordinal), and binary (nominal) data.

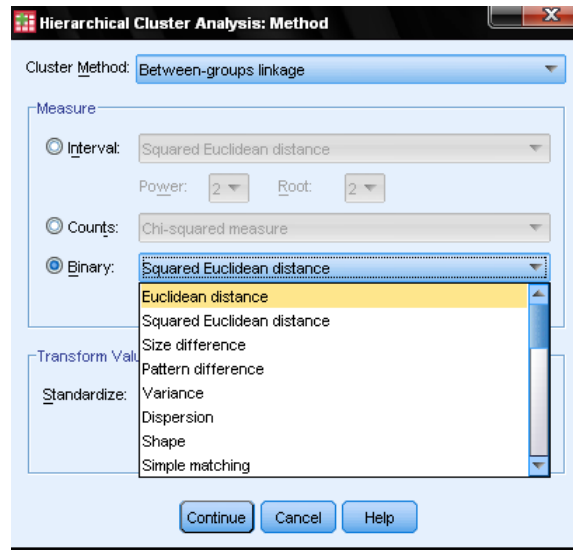
For scale data, the most common is *Square Euclidian Distance*. It is based on the Euclidian Distance between two observations, which uses Pythagoras' formula for the right triangle: the distance is the square root of squared distance on dimension x and y. The Squared Euclidian Distance is this distance squared, thus it increases the importance of large distances, while weakening the importance of small distances.



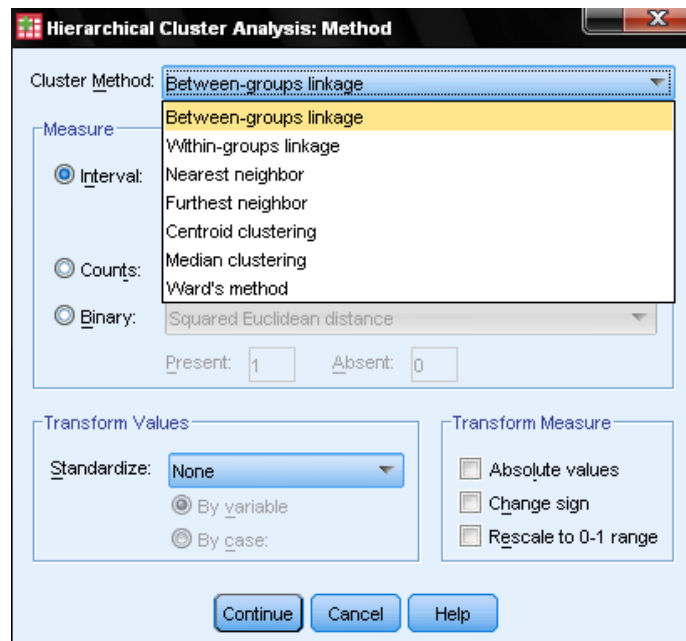
If we have ordinal data (counts) we can select between Chi-Square (think cross-tab) or a standardized Chi-Square called *Phi-Square*. For binary data SPSS has a plethora of distance measures. However, the Square Euclidean distance is a good choice to start with and quite commonly used. It is based on the number of discordant cases.



In our example we choose *Interval* and *Square Euclidean Distance*.



Next we have to choose the *Cluster Method*. Typically choices are Between-groups linkage (distance between clusters is the average distance of all data points within these clusters), nearest neighbor (single linkage: distance between clusters is the smallest distance between two data points), furthest neighbor (complete linkage: distance is the largest distance between two data points), and Ward's method (distance is the distance of all clusters to the grand average of the sample). Single linkage works best with long chains of clusters, while complete linkage works best with dense blobs of clusters and between-groups linkage works with both cluster types. The usual recommendation is to use single linkage first. Although single linkage tends to create chains of clusters, it helps in identifying outliers. After excluding these outliers, we can move onto Ward's method. Ward's method uses the F value (like an ANOVA) to maximize the significance of differences between cluster, which gives it the highest statistical power of all methods. The downside is that it is prone to outliers and creates small clusters.



A last consideration is *standardization*. If the variables have different scales and means we might want to standardize either to *Z scores* or just by centering the scale. We can also transform the values to absolute measures if we have a data set where this might be appropriate.

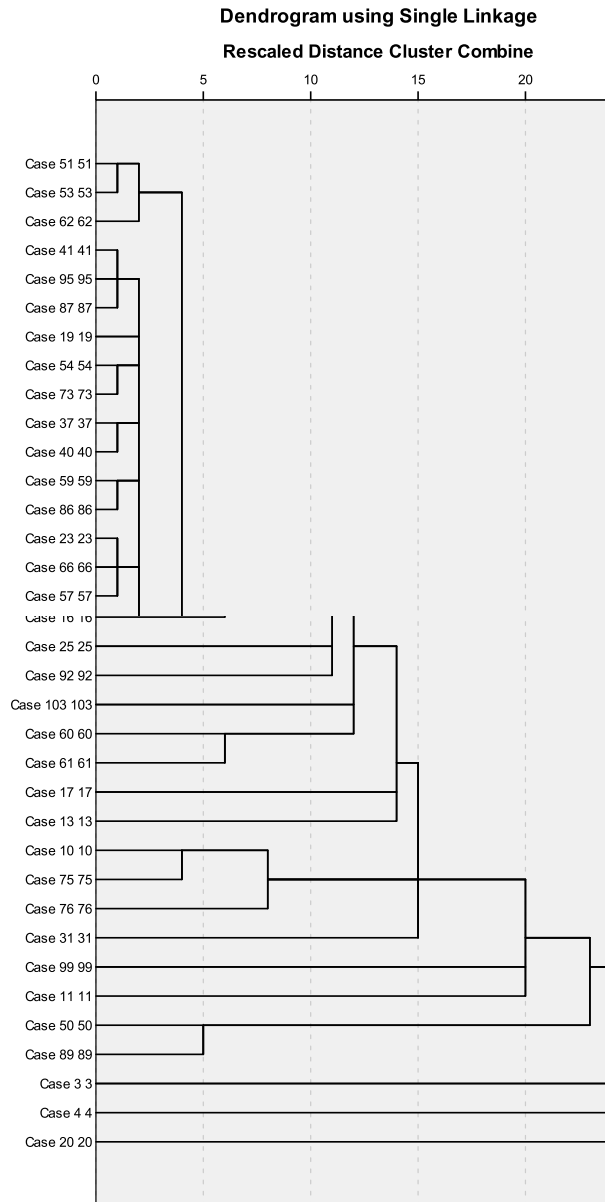
### ***The Output of the Cluster Analysis***

The first table shows the agglomeration schedule. This output does not carry a lot of meaning, but it shows the technicalities of the cluster analysis. A hierarchical analysis starts with each case in a single cluster and then merges the two closest clusters depending on their distance. In our case of single linkage and square Euclidean distance it merges student 51 and 53 as a first step. Next, this cluster's distances to all other clusters are calculated again (because of single linkage it is the nearest neighbor distance). And finally, second step merges student 91 and 105 into another cluster and so on forth, until all cases are merged into one large cluster.

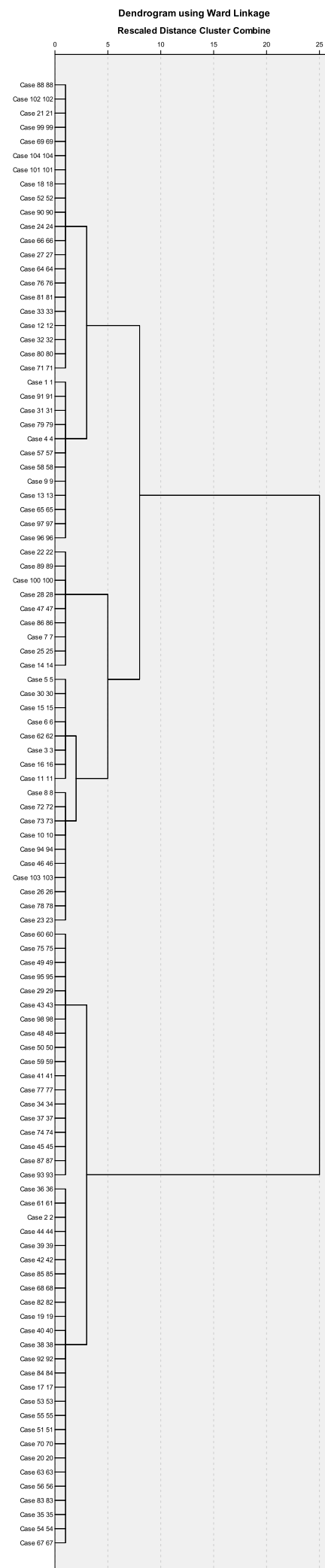
**Agglomeration Schedule**

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	51	53	.009	0	0	39
2	91	105	.009	0	0	57
3	54	73	.017	0	0	31
4	41	95	.026	0	0	6
5	45	88	.028	0	0	8
6	41	87	.031	4	0	24
7	39	64	.034	0	0	29
8	45	47	.034	5	0	13
9	37	40	.034	0	0	26
10	23	66	.036	0	0	14
11	59	86	.039	0	0	19
12	24	102	.042	0	0	62
13	2	45	.042	0	8	15
14	23	57	.043	10	0	18
15	2	71	.043	13	0	16
16	2	85	.044	15	0	28
17	32	46	.049	0	0	34
18	23	38	.050	14	0	19
19	23	59	.056	18	11	26
20	67	79	.056	0	0	21

The *icicle* and *dendrogram* plot show the agglomeration schedule in a slightly more readable format. It shows from top to bottom how the cases are merged into clusters. Since we used single linkage we find that three cases form a chain and should be excluded as outliers.



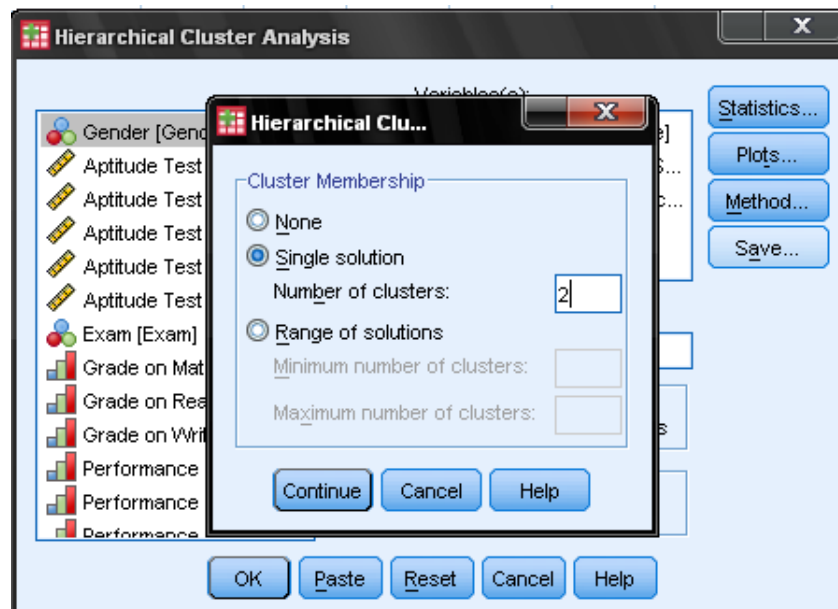
After excluding (by simply erasing) cases 3, 4, and 20 we rerun the analysis with Ward's method and we get the following dendrogram (right). The final task is to identify the correct number of clusters. A rule of thumb is to choose as few clusters as possible that explain as much of the data as possible. This rule is fulfilled at the largest step in the dendrogram. Of course, this is up for interpretation. In this case however, it is quite clear that





the best solution is a two-cluster solution. The biggest step is from 2 to 1 cluster, and is by far bigger than from 3 to 2 clusters.

Upon rerunning the analysis, this time in the dialog box *Save...* we can specify that we want to save the two cluster solutions. In this case a new variable will be added that specifies the grouping. This then would be the dependent variable in our discriminant analysis; it would check the goodness of fit and the profiles of our clusters.



## Factor Analysis

### What is the Factor Analysis?

The Factor Analysis is an explorative analysis. Much like the cluster analysis grouping similar cases, the factor analysis groups similar variables into dimensions. This process is also called identifying latent variables. Since factor analysis is an explorative analysis it does not distinguish between independent and dependent variables.

Factor Analysis reduces the information in a model by reducing the dimensions of the observations. This procedure has multiple purposes. It can be used to simplify the data, for example reducing the number of variables in predictive regression models. If factor analysis is used for these purposes, most often factors are rotated after extraction. Factor analysis has several different rotation methods—some of them ensure that the factors are orthogonal. Then the correlation coefficient between two factors is zero, which eliminates problems of multicollinearity in regression analysis.

Factor analysis is also used in theory testing to verify scale construction and operationalizations. In such a case, the scale is specified upfront and we know that a certain subset of the scale represents an independent dimension within this scale. This form of factor analysis is most often used in structural

equation modeling and is referred to as Confirmatory Factor Analysis. For example, we know that the questions pertaining to the big five personality traits cover all five dimensions N, A, O, and I. If we want to build a regression model that predicts the influence of the personality dimensions on an outcome variable, for example anxiety in public places, we would start to model a confirmatory factor analysis of the twenty questionnaire items that load onto five factors and then regress onto an outcome variable.

Factor analysis can also be used to construct indices. The most common way to construct an index is to simply sum up the items in an index. In some contexts, however, some variables might have a greater explanatory power than others. Also sometimes similar questions correlate so much that we can justify dropping one of the questions completely to shorten questionnaires. In such a case, we can use factor analysis to identify the weight each variable should have in the index.

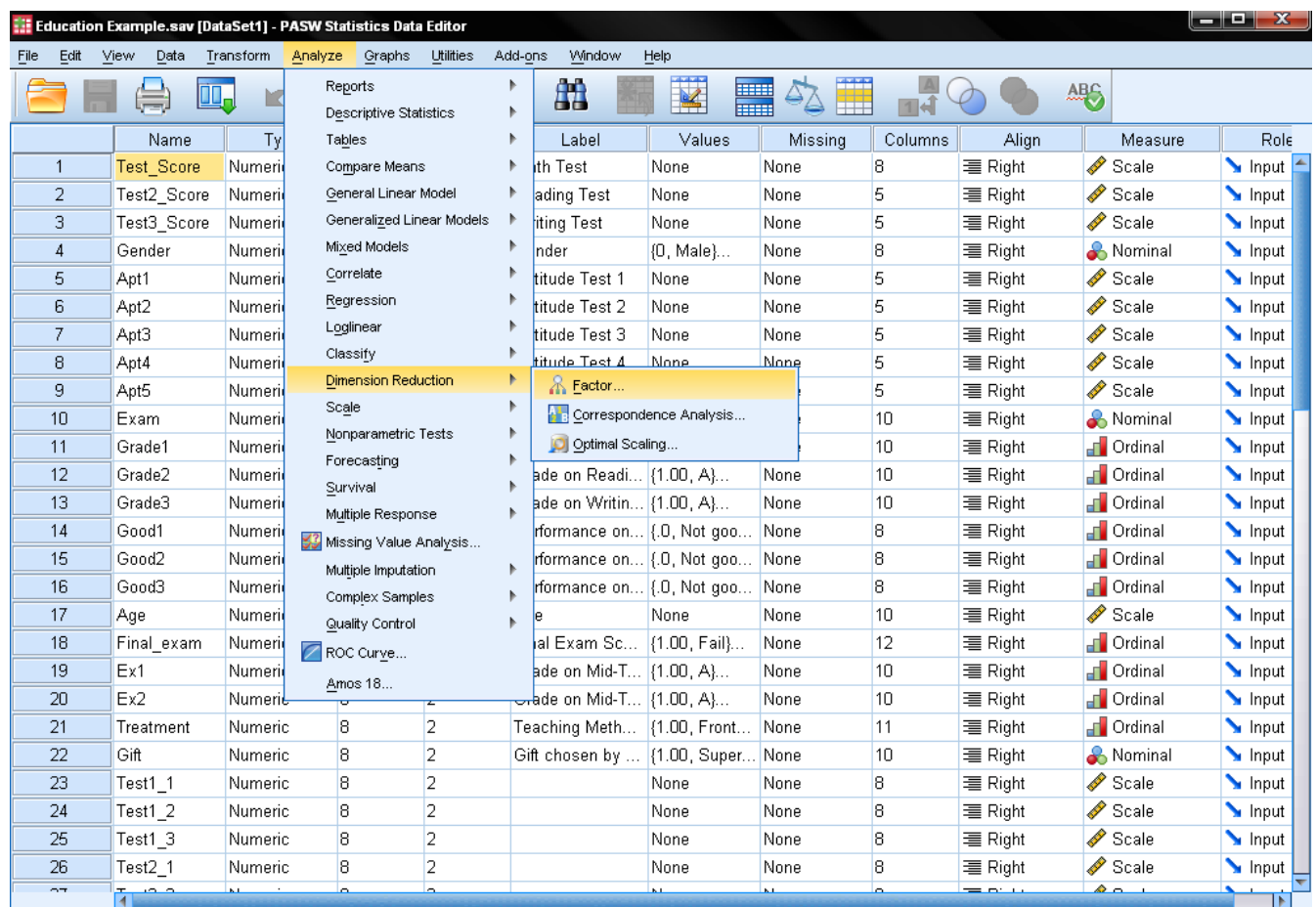
### The Factor Analysis in SPSS

The research question we want to answer with our explorative factor analysis is as follows:

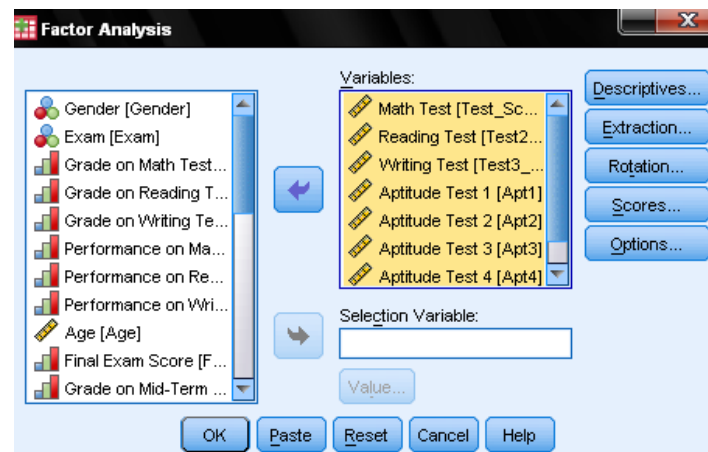
What are the underlying dimensions of our standardized and aptitude test scores?

That is, how do aptitude and standardized tests form performance dimensions?

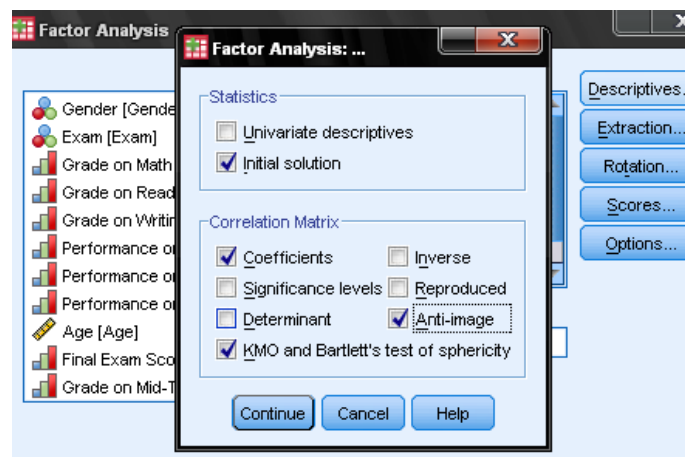
The factor analysis can be found in *Analyze/Dimension Reduction/Factor...*



In the dialog box of the *factor analysis* we start by adding our variables (the standardized tests math, reading, and writing, as well as the aptitude tests 1-5) to the list of variables.

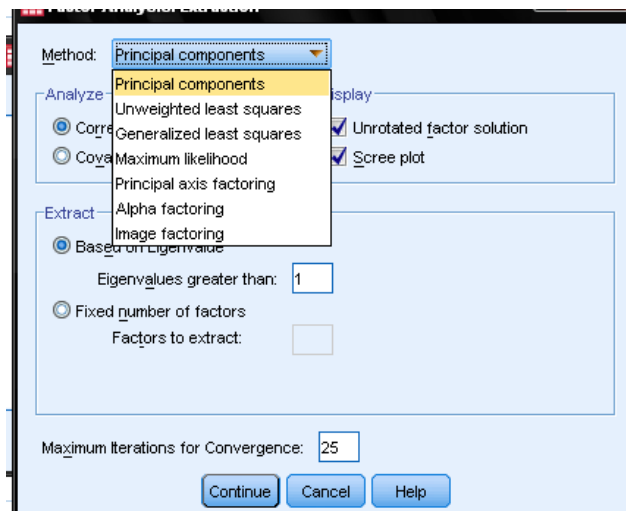


In the dialog *Descriptives...* we need to add a few statistics for which we must verify the assumptions made by the factor analysis. If you want the Univariate Descriptives that is your choice, but to verify the assumptions we need the KMO test of sphericity and the Anti-Image Correlation matrix.



The dialog box *Extraction...* allows us to specify the extraction method and the cut-off value for the extraction. Let's start with the easy one – the cut-off value. Generally, SPSS can extract as many factors as we have variables. The *eigenvalue* is calculated for each factor extracted. If the *eigenvalue* drops below 1 it means that the factor explains less variance than adding a variable would do (all variables are

standardized to have mean = 0 and variance = 1). Thus we want all factors that better explain the model than would adding a single variable.

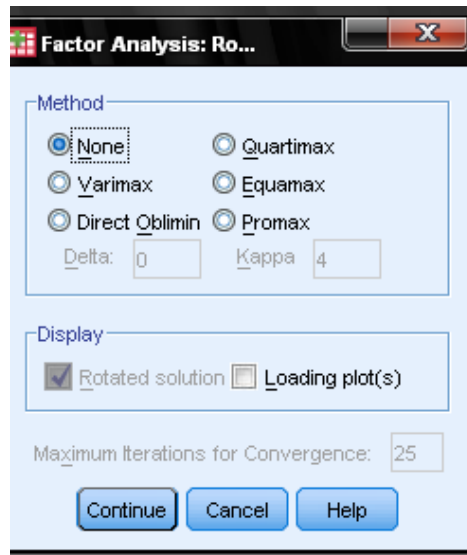


The more complex bit is the appropriate extraction method. *Principal Components* (PCA) is the standard extraction method. It does extract uncorrelated linear combinations of the variables. The first factor has maximum variance. The second and all following factors explain smaller and smaller portions of the variance and are all uncorrelated with each other. It is very similar to *Canonical Correlation Analysis*. Another advantage is that PCA can be used when a correlation matrix is singular.

The second most common analysis is *principal axis* factoring, also called *common factor analysis*, or *principal factor analysis*. Although mathematically very similar to principal components it is interpreted as that principal axis that identifies the latent constructs behind the observations, whereas principal component identifies similar groups of variables.

Generally speaking, principal component is preferred when using factor analysis in causal modeling, and principal factor when using the factor analysis to reduce data. In our research question we are interested in the dimensions behind the variables, and therefore we are going to use Principal Axis Factoring.

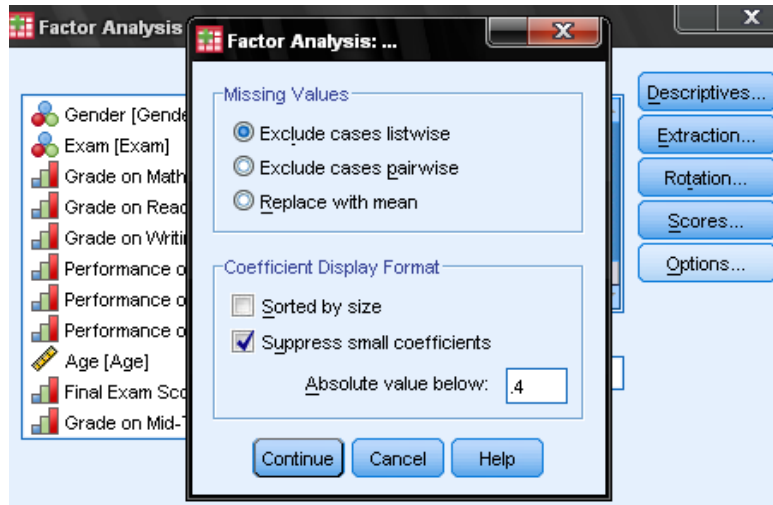
The next step is to select a rotation method. After extracting the factors, SPSS can rotate the factors to better fit the data. The most commonly used method is *Varimax*. *Varimax* is an orthogonal rotation method (that produces independent factors = no multicollinearity) that minimizes the number of variables that have high loadings on each factor. This method simplifies the interpretation of the factors.



A second, frequently used method is *Quartimax*. *Quartimax* rotates the factors in order to minimize the number of factors needed to explain each variable. This method simplifies the interpretation of the observed variables.

Another method is *Equamax*. *Equamax* is a combination of the *Varimax* method, which simplifies the factors, and the *Quartimax* method, which simplifies the variables. The number of variables that load highly on a factor and the number of factors needed to explain a variable are minimized. We choose *Varimax*.

In the dialog box *Options* we can manage how missing values are treated – it might be appropriate to replace them with the mean, which does not change the correlation matrix but ensures that we don't over penalize missing values. Also, we can specify that in the output we don't want to include all factor loadings. The factor loading tables are much easier to interpret when we suppress small factor loadings. Default value is 0.1 in most fields. It is appropriate to increase this value to 0.4. The last step would be to save the results in the *Scores...* dialog. This calculates a value that every respondent would have scored had they answered the factors questions (whatever they might be) instead. Before we save these results to the data set, we should run the factor analysis first, check all assumptions, ensure that the results are meaningful and that they are what we are looking for and then we should re-run the analysis and save the factor scores.



### The Output of the Factor Analysis

The first table shows the *correlation matrix*. This is typically used to do an eyeball test and get a feeling for which variable is strongly associated with which variable.

Correlation Matrix

	Test_Score Math Test	Test2_Score Reading Test	Test3_Score Writing Test	Apt1 Aptitude Test 1	Apt2 Aptitude Test 2	Apt3 Aptitude Test 3	Apt4 Aptitude Test 4	Apt5 Aptitude Test 5
Correlation Test_Score Math Test	1.000	.582	.747	.472	-.252	-.107	-.261	-.217
Test2_Score Reading Test	.582	1.000	.645	.750	-.601	-.178	-.319	-.354
Test3_Score Writing Test	.747	.645	1.000	.601	-.338	-.116	-.253	-.236
Apt1 Aptitude Test 1	.472	.750	.601	1.000	-.499	-.098	-.372	-.468
Apt2 Aptitude Test 2	-.252	-.601	-.338	-.499	1.000	.164	.117	.339
Apt3 Aptitude Test 3	-.107	-.178	-.116	-.098	.164	1.000	-.105	-.025
Apt4 Aptitude Test 4	-.261	-.319	-.253	-.372	.117	-.105	1.000	.295
Apt5 Aptitude Test 5	-.217	-.354	-.236	-.468	.339	-.025	.295	1.000

The next table is the *KMO* and *Bartlett test* of sphericity. The *KMO* criterion can have values between [0,1] where the usual interpretation is that 0.8 indicates a good adequacy to use the data in a factor analysis. If the *KMO* criterion is less than 0.5 we cannot extract in some meaningful way.

### KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.792
Bartlett's Test of Sphericity	Approx. Chi-Square	340.073
	df	28
	Sig.	.000

The next table is the *Anti-Image Matrices*. Image theory splits the variance into image and anti-image. Next we can check the correlation and covariances of the anti-image. The rule of thumb is that in the anti-image covariance matrix only a maximum of 25% of all non-diagonal cells should be greater than |0.09|.

The second part of the table shows the anti-image correlations the diagonal elements of that matrix are the *MSA* values. Like the *KMO* criterion the *MSA* criterion shows if each single variable is adequate for a factor analysis. A figure of 0.8 indicates good adequacy; if *MSA* is less than 0.5, we should exclude the variable from the analysis. We find that although aptitude test 5 has a *MSA* value of .511 and might be a candidate for exclusion, we can proceed with our factor analysis.

Anti-image Matrices

		Test_Score Math Test	Test2_Score Reading Test	Test3_Score Writing Test	Apt1 Aptitude Test 1	Apt2 Aptitude Test 2	Apt3 Aptitude Test 3	Apt4 Aptitude Test 4	Apt5 Aptitude Test 5
Anti-image Covariance	Test_Score Math Test	.411	-.081	-.227	.040	-.058	.015	.043	.027
	Test2_Score Reading Test	-.081	.292	-.047	-.148	.172	.054	.048	-.023
	Test3_Score Writing Test	-.227	-.047	.351	-.090	-.003	.002	-.017	-.038
	Apt1 Aptitude Test 1	.040	-.148	-.090	.354	.032	-.004	.085	.147
	Apt2 Aptitude Test 2	-.058	.172	-.003	.032	.589	-.055	.084	-.113
	Apt3 Aptitude Test 3	.015	.054	.002	-.004	-.055	.930	.126	.058
	Apt4 Aptitude Test 4	.043	.048	-.017	.085	.084	.126	.799	-.116
	Apt5 Aptitude Test 5	.027	-.023	-.038	.147	-.113	.058	-.116	.735
Anti-image Correlation	Test_Score Math Test	.747 <sup>a</sup>	-.233	-.597	.105	-.117	.024	.074	.049
	Test2_Score Reading Test	-.233	.801 <sup>a</sup>	-.146	-.459	.415	.103	.100	-.050
	Test3_Score Writing Test	-.597	-.146	.779 <sup>a</sup>	-.255	-.006	.003	-.032	-.074
	Apt1 Aptitude Test 1	.105	-.459	-.255	.815 <sup>a</sup>	.069	-.007	.159	.289
	Apt2 Aptitude Test 2	-.117	.415	-.006	.069	.797 <sup>a</sup>	-.075	.123	-.171
	Apt3 Aptitude Test 3	.024	.103	.003	-.007	-.075	.708 <sup>a</sup>	.146	.070
	Apt4 Aptitude Test 4	.074	.100	-.032	.159	.123	.146	.827 <sup>a</sup>	-.152
	Apt5 Aptitude Test 5	.049	-.050	-.074	.289	-.171	.070	-.152	.811 <sup>a</sup>

a. Measures of Sampling Adequacy(MSA)

The next table shows the *communalities*. The communality of a variable is the variance of the variable that is explained by all factors together. Mathematically it is the sum of the squared factor loadings for each variable. A rule of thumb is for all communalities to be greater than 0.5. In our example that does not hold true, however for all intents and purposes, we proceed with the analysis.

### Communalities

	Initial	Extraction
Test_Score Math Test	.589	.707
Test2_Score Reading Test	.708	.805
Test3_Score Writing Test	.649	.787
Apt1 Aptitude Test 1	.646	.729
Apt2 Aptitude Test 2	.411	.546
Apt3 Aptitude Test 3	.070	.141
Apt4 Aptitude Test 4	.201	.357
Apt5 Aptitude Test 5	.265	.345

Extraction Method: Principal Axis Factoring.

The next table shows the *total explained variance* of the model. The table also includes the *eigenvalues* of each factor. The *eigenvalue* is the sum of the squared factor loadings for each factor. SPSS extracts all factors that have an *eigenvalue* greater than 0.1. In our case the analysis extracts three factors. This table also shows us the total explained variance before and after rotation. The rule of thumb is that the model should explain more than 70% of the variance. In our example the model explains 55%.

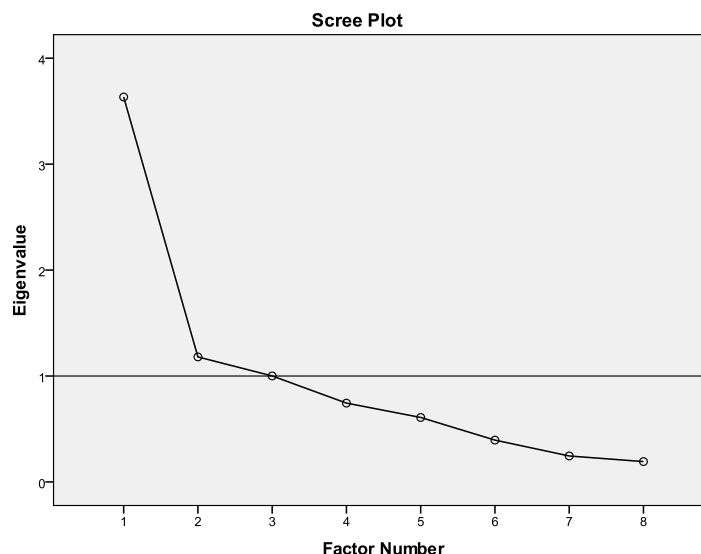
### Total Variance Explained

Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.634	45.419	45.419	3.317	41.464	41.464	1.946	24.325	24.325
2	1.180	14.752	60.170	.625	7.816	49.280	1.915	23.939	48.265
3	1.001	12.508	72.678	.476	5.950	55.229	.557	6.965	55.229
4	.744	9.303	81.981						
5	.609	7.610	89.591						
6	.395	4.939	94.530						
7	.245	3.066	97.596						
8	.192	2.404	100.000						

Extraction Method: Principal Axis Factoring.

The *eigenvalues* for each possible solution are graphically shown in the Scree Plot. As we find with the Kaiser-Criterion (eigenvalue > 1) the optimal solution has three factors. However in our case we could also argue in favor of a two factor solution, because this is the point where the explained variance makes the biggest jump (the elbow criterion). This decision rule is similar to the rule we applied in cluster analysis.





Other criteria commonly used are that they should explain 95%. In our example we would then need six factors. Yet another criterion is that there should be less than half the number of variables in the analysis, or as many factors that you can interpret plausibly and sensibly. Again, factor analysis is somewhat an art.

The next two tables show the *factor matrix* and the *rotated factor matrix*. These tables are the key to interpreting our three factors. The factor loadings that are shown in these tables are the correlation coefficient between the variable and the factor. The factor loadings should be greater than 0.4 and the structure should be easy to interpret.

Labeling these factors is quite controversial as every researcher would interpret them differently. The best way to increase validity of the findings is to have this step done by colleagues and other students that are familiar with the matter. In our example we find that after rotation, the first factor makes students score high in reading, high on aptitude test 1, and low on aptitude tests 2, 4, and 5. The second factor makes students score high in math, writing, reading and aptitude test 1. And the third factor makes students score low on aptitude test 2. However, even if we can show the mechanics of the factor analysis, we cannot find a meaningful interpretation of these factors. We would most likely need to go back and look at the individual results within these aptitude tests to better understand what we see here.

**Factor Matrix<sup>a</sup>**

	Factor		
	1	2	3
Test_Score Math Test	.713	.444	.439
Test2_Score Reading Test	.879		
Test3_Score Writing Test	.797		
Apt1 Aptitude Test 1	.826		
Apt2 Aptitude Test 2	-.582		
Apt3 Aptitude Test 3			
Apt4 Aptitude Test 4			
Apt5 Aptitude Test 5	-.454		

Extraction Method: Principal Axis Factoring.

a. 3 factors extracted. 13 iterations required.

**Rotated Factor Matrix<sup>a</sup>**

	Factor		
	1	2	3
Test_Score Math Test		.816	-.406
Test2_Score Reading Test	.664	.525	
Test3_Score Writing Test		.830	
Apt1 Aptitude Test 1	.731	.435	
Apt2 Aptitude Test 2	-.604		
Apt3 Aptitude Test 3			
Apt4 Aptitude Test 4	-.409		
Apt5 Aptitude Test 5	-.565		

Extraction Method: Principal Axis Factoring.

Rotation Method: Varimax with Kaiser Normalization.

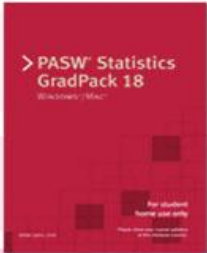
a. Rotation converged in 5 iterations.

## CHAPTER 8: Data Analysis and Statistical Consulting Services

Statistics Solutions is dedicated to facilitating the dissertation process for students by providing statistical help and guidance to ensure a successful graduation. Having worked on my own mixed method (qualitative and quantitative) dissertation, and with 18 years of experience in research design and methodology, I present this SPSS user guide, on behalf of Statistics Solutions, as a gift to you.

The purpose of this guide is to enable students with little to no knowledge of SPSS to open the program and conduct and interpret the most common statistical analyses in the course of their dissertation or thesis. Included is an introduction explaining when and why to use a specific test as well as where to find the test in SPSS and how to run it. Lastly, this guide lets you know what to expect in the results and informs you how to interpret the results correctly.

Statistics Solutions' offers a family of solutions to assist you towards your degree. If you would like to learn more or schedule your free 30-minute consultation to discuss your dissertation research, you can visit us at [www.StatisticsSolutions.com](http://www.StatisticsSolutions.com) or call us at 877-437-8622.

<b>Proposal Phase</b> Formulating hypotheses, research questions, research design Methodological review forms and all pre-dissertation paperwork	
<b>Chapter 3: Methods</b> Formulating hypotheses, research questions, research design Survey instrument help and administration, transcription, data collection, cleaning and coding Sample size calculation and justification/power analysis. Data Analysis, statistical analysis, assistance with choosing the appropriate statistical analysis. Justification of the analysis Statistics and methodology coaching Quantitative and qualitative expertise	
<b>Chapter 4: Results</b> Conduct analysis and examine assumptions Explaining the findings and statistics, formulating justification, writing the results Explaining the results in a way that shows clear support or non-support of hypotheses APA-style tables and results	
<b>APA-Editing</b> APA-formatted results and tables Proposal, dissertation and thesis editing in APA-style	
<b>Revisions and Defense</b> Assistance with committee issues and revisions Dissertation defense coaching	
<b>SPSS Software and Coaching</b> Free online webinars focused on understanding SPSS software and conducting analyses SPSS manual designed specifically for doctoral and masters-level statistics SPSS coaching and support  Statistics Solutions is an authorized partner with SPSS, an IBM Company and offers SPSS Statistics GradPack Software. Purchase online for only \$188.	

## *Terms of Use*

BY YOUR USE, YOU AGREE TO THESE TERMS OF USE.

IF YOU DO NOT AGREE TO THESE TERMS OF USE, DO NOT USE.

We make no representation or warranty about the accuracy or completeness of the materials made available.

We do not warrant that the materials are error free. You assume all risk with using and accessing the materials, including without limitation the entire cost of any necessary service, or correction for any loss or damage that results from the use and access of the materials.

Under no circumstances shall we, nor our affiliates, agents, or suppliers, be liable for any damages, including without limitation, direct, indirect, incidental, special, punitive, consequential, or other damages (including without limitation lost profits, lost revenues, or similar economic loss), whether in contract, tort, or otherwise, arising out of the use or inability to use the materials available here, even if we are advised of the possibility thereof, nor for any claim by a third party.

This material is for your personal and non-commercial use, and you agree to use this for lawful purposes only. You shall not copy, use, modify, transmit, distribute, reverse engineer, or in any way exploit copyrighted or proprietary materials available from here, except as expressly permitted by the respective owner(s) thereof.

You agree to defend, indemnify, and hold us and our affiliates harmless from and against any and all claims, losses, liabilities, damages and expenses (including attorney's fees) arising out of your use of the materials.

The terms of use shall be governed in accordance with the laws of the state of Florida, U.S.A., excluding its conflict of law's provisions. We reserve the right to add, delete, or modify any or all terms of use at any time with or without notice.