A bite-size SPSS analysis guide for common statistical tests:

For students, by students

Edited by Kelly Anne Young

Institute for Open Distance Learning (IODL)

University of South Africa (UNISA)

The following was created for students by students[1] to assist you through the process of analysing quantitative data using SPSS. It provides a basic, elementary understanding to some of the more common statistical tests. Please note, this guide is not intended to be a comprehensive analysis manual; please consult other sources when working with this document.

The datasets used herein can be found at:

https://edge.sagepub.com/field5e/student-resources/datasets

http://www.discoveringstatistics.com/docs/ds_data_files/SPSS%20Data%20Files/

---

[1] Great thanks must be given to the UNISA's Master of Arts in Research Consultation Class of 2019 for their assistance in drafting this guide.

Table of Contents

Parametric assumptions

<u>Firstly, what parametric assumptions should we test?</u>

1. <u>Normally distributed:</u> Assumes that the distribution of the data is normally distributed, forms a bell-curve. This means that it is symmetrical, with a mean score of 0 and a standard deviation of 1. The following tests / statistics can be used:
   a. Skewness and Kurtosis (and their respective standard errors)
   b. Kolmogorov-Smirnov (K-S Test)
   c. Q-Q plots
   d. Histograms
2. <u>Homoscedasticity:</u> Also known as homogeneity of variance, this principle assumes that there are equal or similar variances in the different groups being compared.
3. Data measured on the <u>interval or ratio level.</u>
4. <u>Independence</u>: Groups should be made up of different people who do not influence each other in any way.

<u>How do I test these assumptions?</u>

Normally distributed data: Use histograms (have a look at the histogram, and pay attention to the shape of your data with the normal curve overlaid, does it fit into the line drawing of the bell-curve?); Q-Q plots (dots should fall as close to the line as possible for it to be normally distributed); values of skewness and kurtosis (divided by

their respective standard errors) – check Field (2009, p. 139)[2] for these cut-offs, and the K-S test. How to assess normality using the K-S test: Analyze > Descriptive > Explore > Move variables of interest into the 'Depended list' > Plots (Select histograms and normality plots with tests). The resulting statistic and p-value should be non-significant for your data to be normally distributed.

Homoscedasticity: Use Levene's test to test this assumption. Analyze > Descriptive > Explore > Move variables of interest into the 'Dependent list' > Move grouping variable to 'Factor list' > Plots - 'Spread vs Level with Levene's test, select 'Untransformed'. The resulting statistic and p-value should be non-significant for your data to assume homogenous variances.

Eyeball **independence and interval / ratio data**. Independence implies that the group(s) of

participants were unable to influence each other, and influence the responses

provided.

If you meet all these assumptions, then move on to selecting a parametric technique to analyze your data. If you violate any of these assumptions (especially the normal distribution of data), then you need to select a non-parametric technique to analyze your data.

---

[2] Field, A. (2009). Discovering statistics using SPSS (3rd ed.). London, Sage Publications Ltd.

	Parametric	**Non-parametric**
Assumed distribution	Normal	Any
Assumed variance	Homogeneous	Any
Typical data	Ratio or interval	Ordinal or nominal
Data set relationships	Independent	Any
Statistical test		
Correlation	Pearson	Spearman / Chi-square (categorical)
Independent measures, 2 groups	Independent t-test	Mann-Whitney test
Independent measures, >2 groups	One-way (independent measures) ANOVA	Kruskal-Wallis test
Repeated measures, 2 conditions	Dependent t-test	Wilcoxon test
Repeated measures, >2 conditions	One-way (repeated measures) ANOVA	Friedman's test

Table adapted from: http://changingminds.org/explanations/research/analysis/parametric_non-parametric.html

What is a correlation and how do I perform it in SPSS? (Exam Anxiety.sav)

A correlation tests whether a relationship exists between two continuous variables. There is one exception to this rule, one variable can be categorical, granted that there are only two categories (i.e., it is a binary variable). Note that a correlation does not determine cause and effect.

Pearson's Correlation (r) – Parametric correlation

● Analyze > Correlate > Bivariate > Move variables over into the empty box > Click on the Options button > Select Means and Standard deviations > Click Continue > Make sure that Pearson's correlation is selected under the Correlation Coefficients section > Select either one-tailed or two-tailed under test of significance > One-tailed = When a direction is implied in the hypothesis; Two-tailed = When no direction is implied in the hypothesis (use this option when you are not sure about the direction of the relationship) > Click OK.

Spearman's rho (r_s) and Kendall's tau - Non-parametric correlation

● Analyze > Correlate > Bivariate > Move variables over > Make sure that you select Kendall's tau-b and Spearman under the Correlation Coefficients section > Select either one-tailed or two-tailed under test of significance > Click OK.

● When looking at the correlations table in the output it is important to remember the following: Correlation coefficients range from +1.00 to -1.00; The plus or minus sign in

front of the correlation indicates the direction of the relationship; The size of the correlation coefficient indicates the strength of the relationship as well as the effect size (can be interpreted as):

- ■ ± .1 → weak correlation and small effect
- ■ ± .3 → moderate correlation and medium effect
- ■ ± .5 → strong correlation and large effect

With that in mind, inspect your correlation coefficient and the *Sig.* number (that's your *p*-value) to determine the size of the correlation and whether this is significant.

What is a chi-square and how do I perform it in SPSS? (Cats.sav)

A chi-square test of association is used to check for associations between two categorical variables. It is a non-parametric technique.

● Analyze > Descriptive statistics > Crosstabs > Move your one variable of interest into the 'row' box and the other variable into the 'column' box > Click on the 'Statistics' button and select 'Chi-square' and 'Phi and Cramer's V' > Then click on 'Cells' and select 'Expected' > Click OK

● Inspect the output. Have a look at the Chi-square tests table and within it, have a look at the Pearson's chi-square row. Use the statistics in this row to determine if there is an association between your two categorical variables. A significant p-value (i.e. a p-value equal to or smaller than 0.05) indicates that there is a significant association

between your two variables. Your Phi and Cramer's V will give you an indication of the strength of association between two categorical variables.

● "Phi is used with 2×2 contingency tables (tables in which you have two categorical variables and each variable has only two categories). If one of the two categorical variables contain more than two categories, then Cramer's V is preferred to phi because phi fails to reach its minimum value of 0 (indicating no association) in these circumstances." (Field 2009, p. 695)

What is simple linear regression and how do I conduct it in SPSS? (Record1.sav)

A simple linear regression is a parametric technique used to predict a continuous outcome variable from one predictor variable (more than one predictor would entail a multiple linear regression).

● Analyze > Regression > Linear > Transfer your outcome variable into dependent box and your predictor variable into your independent box > Select OK.

● The model summary provides an indication of the correlation between the two variables, and it is denoted by 'R' in the 'Model summary table'. Pay attention to the *F* statistic (also called the *F*-ratio) in the ANOVA table and its associated p-value. If the *p*-value is significant, then it indicates that the regression model is significantly better at predicting the outcome than when just the mean is used. Next, have a look at the 'Coefficients' table; this table will tell you if your predictor variable is significantly

predicting the outcome (locate your variable in the table and inspect the respective *p*-value/Sig. value). Remember *p*-values smaller than 0.05 are significant.

What is binary logistic regression and how do I perform it in SPSS? (Eel.sav)

A binary logistic regression is a parametric technique used to predict a binary categorical outcome (only two possible outcomes). Predicting a variable with more than two categorical outcomes would entail a *multinomial* logistic regression.

● Analyze > Regression > Binary logistic > Transfer outcome/dependent variable in 'Dependent' box > Transfer your predictor/independent variable into the 'Covariates' box > Under the options tab, select classification plots, Hosmer-Lemeshow goodness of fit, Iteration history, and CI for exp(B) > Click OK.

● In the output, Block 0 represents the model with NO predictors > Block 1 represents a model with the predictors included > Look for 'Variables in the equation' under Block 1 = If any of the predictors have significant Wald statistic (a *p*-value below .05), then it can be concluded that the variable is a significant predictor of the outcome variable > Interpret the Exp(B) as follows: "If the value is greater than 1 then as the predictor increases, the odds of the outcome occurring increase. Conversely, a value less than 1 indicates that as the predictor increases, the odds of the outcome occurring decrease" (Field, 2009, p. 290).

What is an independent t-test and how do I perform it in SPSS? (SpiderBG.sav)

An independent t-test is used when you want to compare the means (averages) from two different groups of people (this is sometimes called an independent-measures or independent-samples t-test).

● Analyze > Compare means > Independent samples t-test > Move your dependent variable into the 'test variable' field and your independent variable into the 'grouping variable' field (Once you've moved your independent variable into this field, the 'define groups' button will be enabled. Let SPSS know what numeric values you've assigned) > Click OK.

● The first table in the output will give you an indication of the averages or means (and standard deviations) of the two variables. If there is a significant difference between your variables, you'll return to this table to determine which one is higher/lower > Your next table gives you an indication if there is a significant difference. First, determine which row you should inspect (either equal variances assumed or not) > then, inspect your p-value to determine whether there is a significant difference between the two scores. If you have a significant p-value (i.e. equal to or less than 0.05), then you can conclude that there is a significant difference between the two scores > Report the t-statistic, the degrees of freedom and the p-value. Also report the means and the corresponding standard errors.

What is a Mann-Whitney and how do I conduct it in SPSS? (Drug.sav)

A Mann-Whitney statistic assesses whether there are significant differences between two conditions and different participants have been used in each condition. It is a non-parametric test, equivalent to the parametric to the independent t-test.

● Analyze > Nonparametric test > Independent samples > Objective (Customize analysis) > Fields (check that fields are in 'Test fields' box and grouping in 'Groups' box) > Settings > Customize tests (select Mann-Whitney U) > Select Run.

● A 'Hypothesis Test Summary' table will appear in the output, containing your null hypotheses > Inspect each null hypothesis in turn to see the Independent-Sample Mann-Whitney test results (these results will provide an indication as to whether you should reject or retain the null hypotheses) > When you retain the null, this means that there is no significant difference between your two groups > When you reject the null, this means that there is a significant difference between the groups.

● If there are significant differences between your two groups, then inspect the 'Test summary' tables and the graphs beneath. Pay particular attention to the mean rank statistics in the graphs to determine which group scored significantly higher / lower.

What is a dependent t-test and how do I conduct such in SPSS? (SpiderRM.sav)

A dependent t-test is used when you want to compare the means (or average) between two sets of scores derived from the same individuals (e.g. a pre= and post-test). A significant result implies that there are significant differences between the two scores.

● Analyze > Compare means > Paired samples t-test > Input your two variables by *selecting them both* and then moving them over using the arrow button. Click 'OK'.

● The first table in the output will give you an indication of the averages or means (and standard deviations) of the two variables. If there is a significant difference between your variables, you'll return to this table to determine which one is higher/lower > Your next table gives you an indication of the correlation between the two variables (if you're interested) > The next table 'paired samples test' will give you an indication as to whether there is a significant difference between the two means > if you have a significant *p*-value (i.e. equal to or less than 0.05), then you can conclude that there is a significant difference between the two scores/means (check the first table to see which is significantly higher or lower) > Report the *t*-statistic, the degrees of freedom and the *p*-value. Also report the means and the corresponding standard errors.

What is a Wilcoxon signed-rank and how do I perform it in SPSS? (Drug.sav)

The Wilcoxon signed-rank test is used when you want to compare two sets of scores that come from the *same* participants (not like above when you wanted to compare two sets of scores from *different* participants). It is the non-parametric equivalent of the dependent t-test.

If you need to first split your data (into your related-groups), consider the following instructions in SPSS: Select the 'Data' ribbon > Split File > Select 'compare groups' (add the variables in here you would like to compare) > Click OK. This split will be

indicated in the bottom right hand corner of SPSS (simply go to split file again and choose to analyze all cases, do not create groups).

● Analyze > Nonparametric test > Related samples > Objective (Customise analysis) > Fields > Ensure what you want to analyze has been added in the Test fields > Settings > Customise tests > Select Wilcoxon matched-pair signed-rank > Click 'Run'

● A 'Hypothesis Test Summary' table will appear in the output, containing your null hypotheses > Inspect each null hypothesis in turn to see the Related-Samples Wilcoxon Signed Rank test results (these results will provide an indication as to whether you should reject or retain the null hypotheses) > When you retain the null, this means that there is no significant difference between your two scores > When you reject the null, this means that there is a significant difference between the two groups of scores. Get these statistics from the 'Related-Samples Wilcoxon Signed Rank Test summary' tables.

If there are significant differences between your two groups of scores, then inspect the histograms. Pay particular attention to the negative and positive differences. If you have more positive differences, this means that the scores on the second iteration were significantly higher. If you have more negative differences, this implies that the scores on the first iteration were significantly higher. In your write up, include both the p-value and z-scores (reported as the 'standardized test statistic' in the summary table).

What is a one-way (independent measures) ANOVA and how do I perform it in SPSS? (Viagra.sav)

A one-way (independent measures) ANOVA is a parametric test used when you want to compare *more than two* independent means (i.e., these means come from different people).

● Analyze tab > Compare means > One-way ANOVA > Move your dependent variable to the dependent box and your independent variable to your factor box > Select Options > Select Descriptive and Homogeneity of variance test > Continue > Select Post hoc > (read below and then) > Continue > OK.

> There are lots of different post hoc tests: when you have equal sample
> sizes and homogeneity of variance is met use REGWQ or Tukey's HSD.
> If sample sizes are slightly different then use Gabriel's procedure, but if
> sample sizes are very different use Hochberg's GT2. If there is any doubt
> about homogeneity of variance use the Games–Howell procedure.
> (Field, 2009, p. 388)

● Check your output > First table should contain your descriptive statistics (these will give you an indication of which means are larger, smaller) > the second table contains Levene's test (for equal variances) > Check the first row in this table; based on the mean and make note of result (equal variances assumed or not?) > Check the ANOVA table and locate the *F*-statistic and its *p*-value > Interpret result of *p*-value as follows: if significant, then there are significant differences between two (or more) of the

means, if not, then there are no significant differences in the means. If there are significant differences, move on to the Multiple Comparisons table > if equal variances are assumed, look for significant p-values in the Tukey HSD part of the table; if equal variances are not assumed, look for significant p-values in the Games-Howell part of the table. Any significant p-value denotes a significant difference in the means of those two groups. Once identified, look back at your descriptive table to ascertain which mean is significantly higher / bigger.

● Report > give details of the F-ratio and the degrees of freedom from the ANOVA table. If significant, report Levene's statistic and the output from the post-hoc comparisons (i.e., the multiple comparisons table).

What is a Kruskal-Wallis and how do I perform it in SPSS? (Soya.sav)

A Kruskal Wallis is a non-parametric test used when you want to compare *more than two* independent means (i.e., these means come from different people). It is denoted as an uppercase, italics "*H*"

● Analyze tab > Non-parametric test > Independent Samples > Select Customize analysis (in the Objective tab) > Check if your variables of interest are in the correct fields (if not, place them in the appropriate field) > Settings tab > Customize tests > Select Kruskal-Wallis 1-way ANOVA (*k* samples) > Run

● A 'Hypothesis Test Summary' table will appear in the output, containing your null hypotheses > Inspect each null hypothesis and the *p*-value (Sig.) (these results will

provide an indication as to whether you should reject or retain the null hypotheses) > When you retain the null, this means that there is no significant differences between your means > When you reject the null, this means that there are significant differences between the means.

● If there are significant differences (i.e. you have a p-value equal to or smaller than 0.05), then inspect the 'pairwise comparisons' table > Look at the p-values in the "Adj. Sig." column *only* (ignore the Sig. column) > If you have a significant 'Adj. Sig.' p-value, then you can conclude that there is a significant difference between those two particular means. Report the statistics found in the 'Independent-Samples Kruskal-Wallis Test Summary' table.

What is a one-way (repeated measures) ANOVA and how do I perform it in SPSS? (Bushtucker.sav)

● A one-way (repeated measures) ANOVA is a parametric test used when you want to compare *more than two* dependent means (i.e., these means come from the same people).

● Analyze > GLM (General Linear Model) > Repeated measures > Rename 'factor 1' to the name of your independent variable > add the number of levels which relate to your IV > click add > click define > move your IV levels, which should be your variables, into the open question marks > click on EM Means > Move your IV over to the Display means for box and select Compare main effects > From the confidence interval

adjustment box, select the dropdown arrow and choose Bonferroni > Click continue > Click on the options > choose descriptive statistics > continue > OK

● The first table in the output lists the value labels of each level of your IV > the descriptive statistics table lists the averages (i.e., the means); this table is important to come back to if there are significant differences > next, check Mauchly's test of sphericity and the p-value (you want a *ns* p-value; if significant then you've violated your assumption) > Next, check the tests of within-subjects effects table > if you violated Mauchly's test then read the statistics for Greenhouse-Geisser (if you didn't violate the assumption, then look at the first row, Sphericity assumed) > if the F-statistic (in whichever row you're reading) is significant, then there are significant differences in your mean (if not significant, then there are no significant differences in the means) > if significant, check the pairwise comparisons table > look for significant p-values in the Sig. column > if you find significant p-values, you can conclude that there is a significant difference between one mean and another (use the first table in your output to identify the label for each number in the pairwise comparisons table).

What is a Friedman's ANOVA and how do I perform it in SPSS? (Diet.sav)

● A Friedman's ANOVA is a non-parametric test used when you want to compare *more than two* dependent means (i.e., these means come from the same people). It is denoted as "F_r"

● Analyze > Non-parametric tests > Related samples > Select 'Customise analysis' in the Objective tab > Move over to the Fields tab > Check if your variables of interest

are in the 'test fields' (if not, place them there) > Move over to the Settings tab > Select 'Customize tests' and select Friedman's 2-way ANOVA by ranks (*k* samples) > Run

● A 'Hypothesis Test Summary' table will appear in the output, containing your null hypotheses > Inspect each null hypothesis and the *p*-value (Sig.) (these results will provide an indication as to whether you should reject or retain the null hypotheses) > When you retain the null, this means that there is no significant differences between your means > When you reject the null, this means that there are significant differences between the means.

● If there are significant differences (i.e. you have a *p*-value equal to or smaller than 0.05), then inspect the 'pairwise comparisons' table > Look at the *p*-values in the "Adj. Sig." column *only* (ignore the Sig. column) > If you have a significant 'Adj. Sig.' *p*-value, then you can conclude that there is a significant difference between those two particular means. Report the statistics found in the 'Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary' table.