Nonparametric Tests

In this book, we concentrated on parametric tests, which assume that the dependent variable is normally distributed. As mentioned above, parametric tests are rather robust against violations of this assumption. Specifically, when comparing means, we can assume that the dependent variable is at least asymptotically normally distributed (in non-statistical terms: close to being normally distributed) in cases where our sample size exceeds 30. Furthermore, the various t-tests that we examined require the dependent variable to be measured on an interval scale. Although researchers commonly use these tests in many practical situations on ordinal data as well, this is actually inappropriate. To cope with potential problems arising from these restrictions, we can apply nonparametric tests, which do not assume a specific distribution and can be used on dependent variables measured on a nominal or ordinal scale. Figure A6.1 provides a classification of nonparametric tests related to differences testing.

Researchers are often interested in examining nominal variables. A relevant term in this regard is χ^2 (pronounced as *chi square*). There are two types of χ^2-tests involving nominal data (note that they can also be applied to ordinal data):

1) The *one-sample χ^2-test* (also called χ^2 goodness-of-fit test) explores those cases that fall into a single variable's various categories, and compares these with an expected value.
2) The *χ^2-test for independence* is used to determine whether two nominal (or ordinal) variables are related.

Let's approach these two tests by means of an example. Suppose a mobile phone producer plans to launch a new smartphone on the market, but is not sure which color to use. 200 people are chosen from the target group, of which 40 indicate "black," 84 "silver," and 76 "red." If we want to establish whether or not these preferences differ significantly from what could be expected a priori, then this can be determined by means of a one-sample χ^2-test. In the simplest case, we would expect all three colors to be equally popular, which means that we expect 66.67

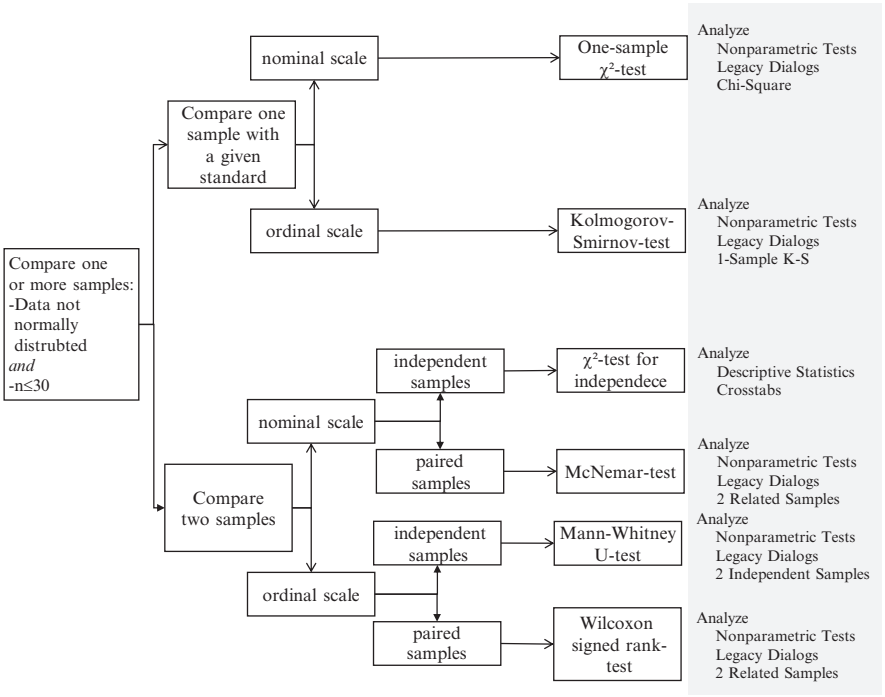

Figure A6.1 Selected nonparametric tests related to differences testing.

respondents to choose each of these colors.[1] Comparing the sample values with these expected values, we conclude that there is a difference between the two. The question is whether this difference is purely attributed to coincidence, or whether it also holds for the population. If the latter is the case, we could conclude that silver is the preferred smartphone color in the population. This is what we are going to explore by means of the one-sample χ^2-test whose null hypothesis states that there is no difference between the expected and observed values.

We can test this null hypothesis using the χ^2-test statistic, which is calculated by collecting observed values for each of the categories and examining the differences between the observed and expected values:

$$\chi^2 = \sum_{i=1}^{k} \frac{(h_i - \tilde{h}_i)^2}{\tilde{h}_i},$$

where h_i is the observed value for category i and $\tilde{h}_i$ is the expected value for the i^{th} category, and k is the total number of categories. In other words, the χ^2-test statistic

[1]We could likewise pre-specify the expected frequencies and test certain assumption regarding the proportions.

is the sum of the squared differences between the observed and expected values, divided by the expected values.

As mentioned above, we would expect there to be 200/3=66.67 respondents preferring each of the three colors, which yields the following:

$$\chi^2 = \frac{(40-66.67)^2}{66.67} + \frac{(84-66.67)^2}{66.67} + \frac{(76-66.67)^2}{66.67}$$
$$= 10.67 + 4.50 + 1.31 = 16.48$$

The test value is not directly interpretable but must be compared to the critical value obtained from the χ^2-statistic with $k - 1$ (in our example $3 - 1 = 2$) degrees of freedom (see Table A3 in the Web Appendix (→Chapters→Additional Material)). As the test statistic value (16.48) is much higher than the critical value (5.991; α=0.05), we can reject the null hypothesis and, thus, conclude that the preferences differ significantly from what could be expected. In this example, we assumed that each category has the same expected frequency (i.e. 66.67). However, we could similarly pre-specify the expected frequencies and test certain assumption regarding the proportions.

In the previous example, we considered only one sample, but researchers are frequently interested in evaluating whether there is a significant relationship between two nominal variables. Suppose that we further differentiated the survey described above by distinguishing between male and female respondents. A possible crosstab (see Chapter 5) is presented in Table A6.1 (ignore the $\tilde{h}$ values in the table for the time being):

Table A6.1 Crosstab for χ^2-test for independence

	Male	Female	Σ
Black	28	12	40
	$\tilde{h}_{11} = 20$	$\tilde{h}_{12} = 20$	
Silver	48	36	84
	$\tilde{h}_{21} = 42$	$\tilde{h}_{22} = 42$	
White	24	52	76
	$\tilde{h}_{31} = 38$	$\tilde{h}_{32} = 38$	
Σ	100	100	200

This 3x2 crosstab indicates that, in this case, 28 male respondents prefer the black smartphone, 48 the silver one, and 24 the white one. The last column (row) indicates the column (row) total indicated by the summation signs (Σ). In this sample, there are 100 males and 100 females. To answer the research question whether there is a relationship between the respondents' gender and their color preferences, we can apply the *χ^2-test for independence*. Specifically, it tests the following hypotheses:

H0: Preference for color is independent of gender

H1: Preference for color is dependent of gender

As in the one-sample case, this test examines the degree to which the observed frequencies deviate from the expected frequencies. The expected frequency of a cell $\tilde{h}_{ij}$ (i being the index of the first variable with k categories and j being the index of the second variable with m categories) is the column total times the row total divided by the number of observations. This seems complicated, but is not. For example, the expected frequency of the cell male/black $\tilde{h}_{11}$ is 100 (column total category "male") times 40 (row total of the category "black"), divided by 200 (the total number of observations), which equals 20. Similarly, the expected frequency of the cell silver/male is $\tilde{h}_{21} = \frac{100.84}{200} = 42$, and so on (see Table A6.1 for all cells' expected frequencies).

An important property of the χ^2-test is that it requires the expected frequency for each cell to be five or higher. In our example, this requirement is easily met. However, if this were not the case, we would have to revert to the Fisher's exact test, which bypasses potential problems (see Box A6.1).

Box A6.1: Fisher's exact test

The Fisher's exact test is commonly used in crosstabs when sample sizes are small. For 2x2 tables, SPSS computes the Fisher's exact test by default and we should interpret this test result instead of the χ^2-test if there is a cell with an expected frequency of less than five.[2] Unfortunately, SPSS does not routinely offer this test for tables with more than two rows or columns (only in the add-on module exact tests). Kirkman (1996), however, provides a web template for calculating the Fisher's exact test for higher-order tables (see mobile tag and URL).

http://tinyurl.com/fisher-exact

[2]As an additional statistical measure, SPSS computes the Yate's correction, which adjusts the Chi-square statistic when at least one cell in the table has an expected frequency smaller than 5 (in the SPSS output, it is labelled "Continuity Correction". However, the Yate's correction can result in an overly conservative result that fails to reject the null hypothesis when it should. We therefore do not recommend it

The computation of the χ^2-test statistic is similar to the example above, with the only exception that we have to append a second summation sign as there are now two nominal variables:

$$\chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{m} \frac{(h_{ij} - \tilde{h}ij)^2}{\tilde{h}ij} = \frac{(28-20)^2}{20} + \cdots + \frac{(52-38)}{38} = 18.43$$

The degrees of freedom are calculated as follows: $df = (k-1)\cdot(m-1)$. This example has 2 degrees of freedom and the associated critical value (for $\alpha = 0.05$) of 5.991 is much smaller than the test value (18.43). Thus, we can assume that there is a significant relationship between the respondents' gender and preference for color. A problem associated with the χ^2-test is that it only provides a weak indication of the strength of the association.[3] Fortunately, there are several measures which assist in overcoming this limitation; we introduce these in Box A6.2.

Box 6.A2: Measures of the strength of association

There are several statistical measures that provide us with information regarding the strength of the association between nominal variables. Their computation only makes sense of course if the χ^2-test or Fisher's exact test renders significant results.

- The φ *(phi)* coefficient is used to measure the strength of association in 2x2 crosstabs. In fact, it is only a correlation coefficient for nominal variables and is computed as follows:

$$\varphi = \sqrt{\frac{\chi^2}{n}}$$

  If there is no association whatsoever (which is extremely unrealistic) between the variables, φ would be zero. Conversely, a value of 1 implies that the variables are perfectly associated. Generally, a φ value below 0.30 describes a weak association, between 0.30 and 0.49 a moderate one, and above 0.50 a strong association.
- The *contingency coefficient* (CC) is the equivalent to the φ coefficient for crosstabs with more than two rows and two columns. It is very similar to φ and also varies between 0 and 1 with higher values denoting a greater strength of association:

$$CC = \sqrt{\frac{\chi^2}{\chi^2 + n}}$$

(continued)

[3]Unfortunately, the absolute value of the χ^2-test statistic does not provide any indication regarding the strength of the relationship.

CC is a very conservative measure and hardly ever reaches 1. Consequently, we should rather interpret CC in relation to its theoretical maximum value, which is defined as follows:

$$CC_{max} = \sqrt{\frac{r-1}{r}},$$

where r is the lesser of the number of rows and the number of columns. In our example, the number of columns (2) is smaller than the number of rows (3), and, consequently CC_{max} is 0.71. Using the information from our previous analysis, we compute CC as follows:

$$CC_{max} = \sqrt{\frac{18.43}{18.43 + 200}} = 0.29$$

This value indicates that the association can be considered medium when compared to CC_{max}.

- Cramer's V is a modified version of the φ coefficient and is used in crosstabs larger than 2 x 2:

$$V = \sqrt{\frac{\chi^2}{n \cdot (r-1)}} = \sqrt{\frac{18.43}{200 \cdot (2-1)}} = 0.30$$

Other measures include the *lambda, uncertainty coefficient*, and *Tshuprow's* T. While all these measures are used to measure the strength of association between nominal variables, SPSS provides different statistics such as *Kendall's τ-b* (pronounced as *tau*), *Kendall's τ-c, Somer's d*, and γ (pronounced as *gamma*) for ordinal variables. Essentially, these measures use information on the ordering of variables' categories by considering every possible pair of cases in the crosstab. They vary between −1 and +1 and thus distinguish between positive and negative relationships. Higher absolute values denote a stronger degree of association. For more information, see, for example, Fleiss et al. (2003).

While the χ^2-test for independence (as well as Fisher's exact test) helps us explore the relationship between two nominal variables from independent samples, the *McNemar test* allows us to do this when the data stem from two paired samples. More specifically, the NcNemar test is used for dichotomous variables. For example, we might carry out an experiment in which we ask respondents whether they would buy a specific smartphone before and after being exposed to an online banner. The test's null hypothesis is that the number of respondents who changed their response in one direction (i.e. buy instead of not buy) is equal to the number of those who changed in the opposite direction (i.e. not buy instead of buy). The McNemar test compares the observed data to the null expectation, using a goodness-of-fit test and is interpreted like the tests discussed before.

So far, we looked at tests related to variables that are (at least) nominally scaled, but there are also various tests that are related to ordinal data. We will only look at these tests briefly, as their computation is often rather complex and goes beyond the scope of this book.

One of the most important nonparametric tests for ordinal data is the *Kolmogorov-Smirnov test*. This tests the null hypothesis that the dependent variable under consideration follows a specific distribution. In most instances, we use it to examine whether a specific variable is normally distributed. However, in theory, we can use this test to assess any other type of distribution such as exponential or uniform.

Somewhat surprisingly, the test's null hypothesis is that the variable follows a specific distribution (e.g., the normal distribution). This means that only if the test result is insignificant, i.e. the null hypothesis is not rejected, can we assume that the data are drawn from the specific distribution against which it is tested. Technically, when assuming a normal distribution, the Kolmogorov-Smirnov test compares the sample scores with an artificial set of normally distributed scores that has the same mean and standard deviation as the sample data. However, this approach is known to yield biased results which can be corrected for by adopting the Lilliefors correction (Lilliefors 1967). The Lilliefors corrects for the fact that we do not know the true mean and standard deviation of the population. An issue with the Kolmogorov-Smirnov test is that it is very sensitive when used on very large samples and often rejects the null hypothesis if very small deviations are present.

The Shapiro-Wilk test also tests the null hypothesis that the test variable under consideration is normally distributed. Thus, rejecting the Shapiro-Wilk test provides evidence that the variable is not normally distributed. It is best used for sample sizes of less than 50. A drawback of the Shapiro-Wilk test however, is that it works poorly if the variable you are testing has many identical values, in which case you should use the Kolmogorov-Smirnov test with Lilliefors correction.

The *Mann-Whitney U-test* is a nonparametric alternative to the independent samples t-test, which can be used if the dependent variable is measured on an ordinal scale. Furthermore, it is commonly applied in situations where the dependent variable is measured on an interval scale but does not follow a normal distribution. Like the t-test, it tests the null hypothesis that the difference in the location of two populations (expressed by the median) is zero. Rather than being based on means, the Mann-Whitney U-test statistic is based on a comparison of the observations' ranks. The corresponding method for paired samples is the *Wilcoxon signed-rank test*, which tests the null hypothesis that two medians stemming from paired samples are identical.

Table A6.2 provides an overview of the steps involved when carrying out the following nonparametric tests in SPSS: One-sample χ^2-test, and the χ^2-test for independence. Compare Chapter 6 of the book for a description of the Kolmogorov-Smirnov as wee as Shapio-Wilk test.

Table A6.2: Steps involved in carrying out selected parametric and nonparametric tests in SPSS

Theory	Action
One-sample χ^2-test:	
Compare proportions of cases that fall into the various categories of a nominal or ordinal variable with a given standard	▸ Analyze ▸ Nonparametric Tests ▸ Legacy Dialogs ▸ Chi-square Test
Assumptions	
Is the test variable measured on a nominal or ordinal scale?	Check Chapter 3 to determine the measurement level of your variables
Are the observations independent?	Consult Chapter 3 to determine if observations are independent
Specification	
Select the test variable	Enter the variable into the **Test Variable List** box
Select the expected values	If the expected values are uniformly distributed across the categories, no action is necessary; otherwise, pre-specify the expected values in the **Expected Values** box
Results interpretation	
Examine the test results	Examine the χ^2-value and its significance level
χ^2-test for independent samples:	
Determine whether two nominal or ordinal variables are related	▸ Analyze ▸ Descriptive Statistics ▸ Crosstabs
Assumptions	
Is the test variable measured on a nominal or ordinal scale?	Check Chapter 3 to determine the measurement level of your variables
Are the observations independent?	Consult Chapter 3 to determine if observations are independent
Specification	
Select the test variables	Select the crosstab's row and column variables ▸ Analyze ▸ Descriptive Statistics ▸ Crosstabs
Select the test statistic	Select χ^2-test statistic ▸ Analyze ▸ Descriptive Statistics ▸ Crosstabs ▸ Statistics
Select measures of the strength of association	Go to ▸ Analyze ▸ Descriptive Statistics ▸ Crosstabs▸ Statistics and choose measures according to the scale (nominal or ordinal)
Results interpretation	
Examine the test results	If the expected frequency is five or more for all cells, examine the χ^2-value and its significance level; if the expected frequency is less than five for one cell, examine Fisher's exact test results (2x2 tables only)

(*continued*)

Table A6.2 (continued)

Theory	Action
Determine the strength of the effects	
	Examine the measures of the strength of association: ‣ Analyze ‣ Descriptive Statistics ‣Crosstabs ‣ Statistics

References for this Web Appendix

Fleiss, Joseph L., Bruce Levin and Myunghee C. Paik (2003). *Statistical Methods for Raters and Proportions*, 3rd. edition, New York et al.: Wiley.

Kirkman, T. W. (1996): Statistics to Use, http://www.physics.csbsju.edu/stats/exact_NROW_NCOLUMN_form.html.

Lilliefors, Hubert W. (1967). "On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown," *Journal of the American Statistical Association*, 62 (318), 399-402.