Nonparametric Tests

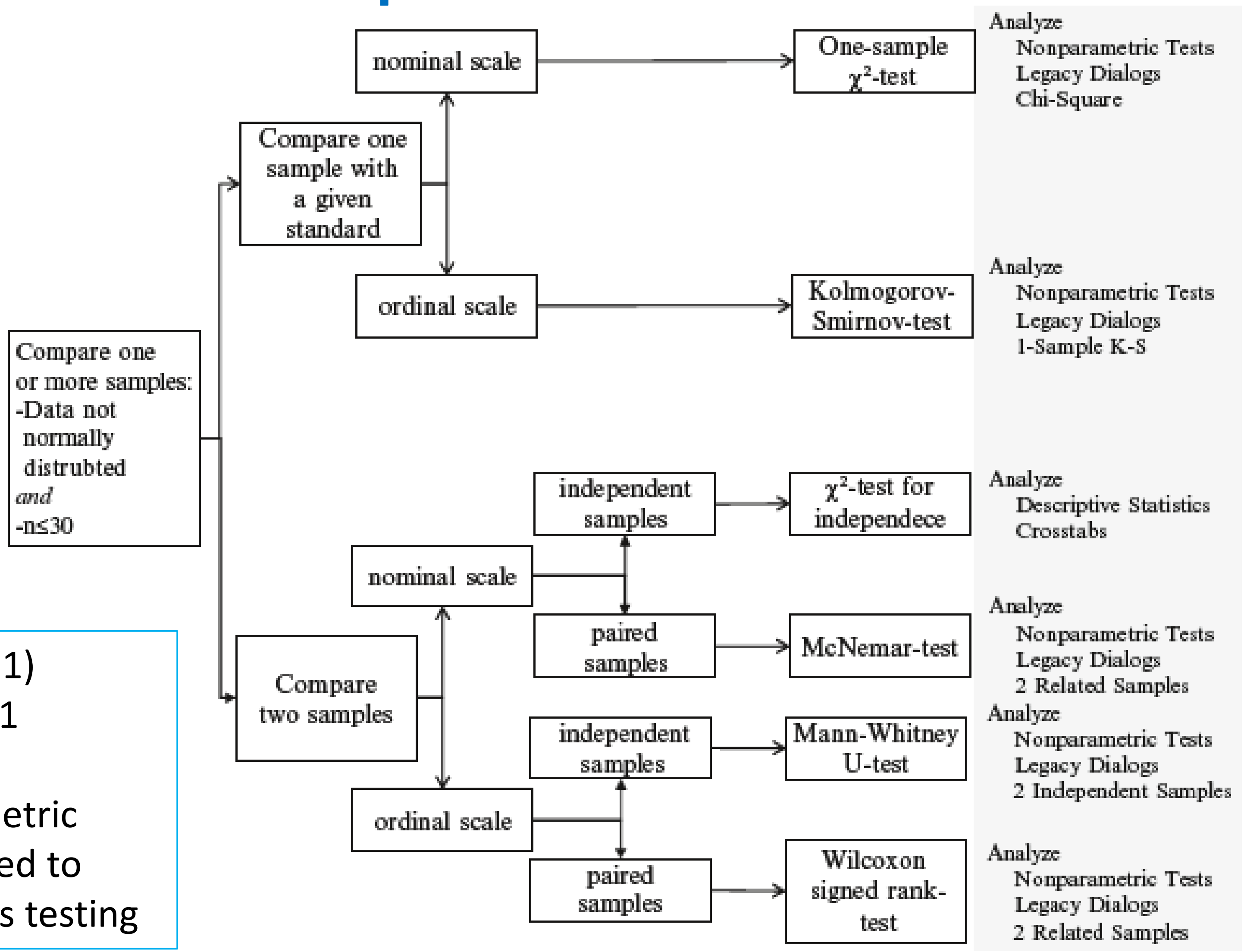
Quantitative Data Analysis

José Novais Santos

Summary

- Univariate
  - χ^2 test (chi-square test) / goodness-of-fit test
  - Kolmogorov-Smirnov test
- Bivariate
  - χ^2 test for independence
  - Mann-Whitney U-test
  - Kruskal-Wallis

Nonparametric tests



Mooi (2011)
Figure A6.1
Selected nonparametric tests related to differences testing

Univariate tests

- Open file “OH7 - survey analysis v4.sav”

χ^2 test

- Pronounced as “chi-square – test”
- Popular nonparametric test
- Do not require the variable of interest to follow a specific distribution
- Can also be used for variables measured on a nominal or ordinal scale

χ^2 test

- There are two types of χ^2 tests
  - see Mooi (2011) figure A6.1

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

χ^2 test

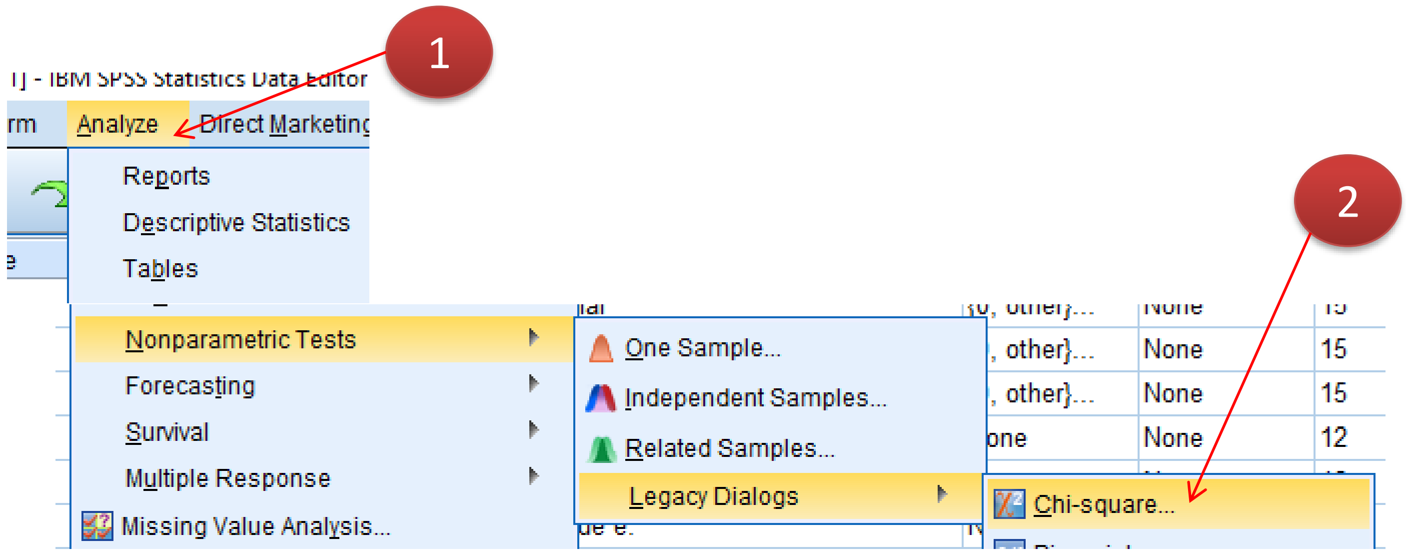
- χ^2 goodness-of-fit test
  - Comparing the sample values (observed frequencies) with these expected values (pre-determined theoretical model/distribution)
    - Do observed values/preferences/… differ significantly from what could be expected a priori?
      - Is the difference attributed to coincidence or it holds for the population
    - SPSS, by default, tests if the distribution is uniform "all categories equal"
      - Still we can pre-specify the expected frequencies and test certain assumption regarding the proportions (selecting "values" – see weight creation)

χ^2 test

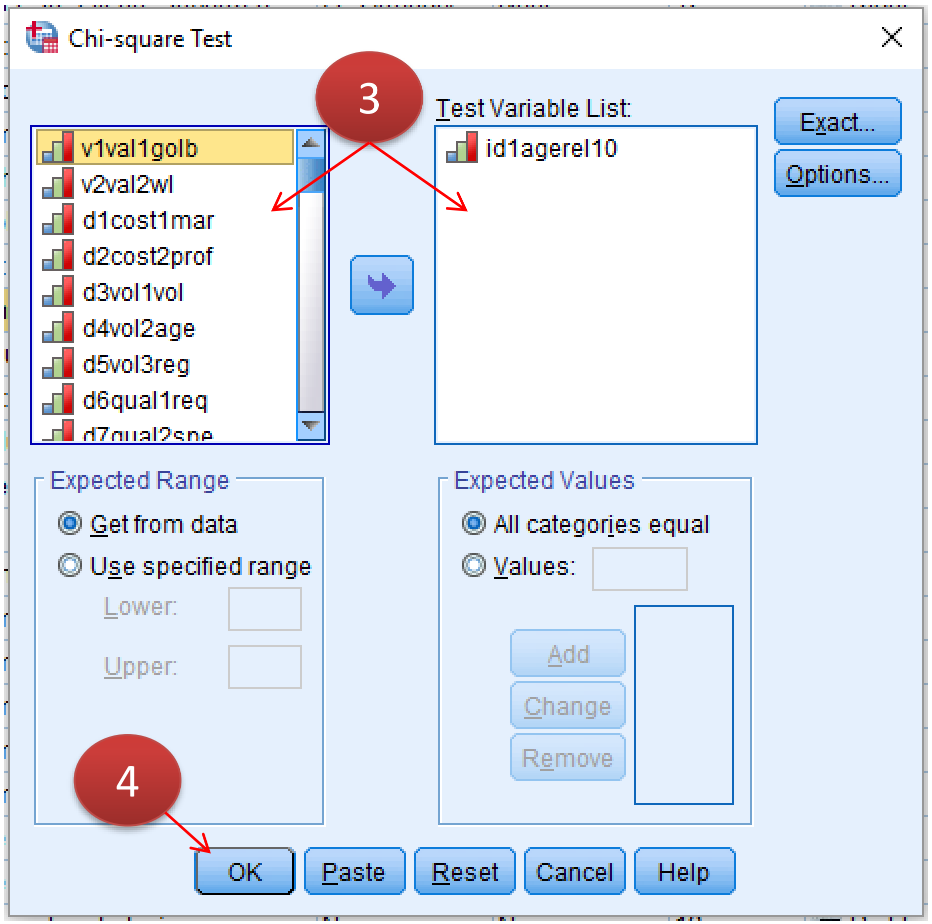
- We want to test if the relationship age (id1agerel10) has an uniform distribution
- Hypotheses
  - H_0: The relationship age (id1agerel10) has an uniform distribution in the population
  - H_1: The relationship age (id1agerel10) does not have an uniform distribution in the population
- To test these hypotheses in an nominal variable we need to preform a χ^2 goodness-of-fit test

χ^2 test

- χ^2 goodness-of-fit test



χ^2 test



relationship age 10 to 10 years

	Observed N	Expected N	Residual
0 through 9 years	123	49,5	73,5
10 through 19 years	65	49,5	15,5
20 through 29 years	2	49,5	-47,5
30 through 39 years	8	49,5	-41,5
Total	198		

Test Statistics

	relationship age 10 to 10 years
Chi-Square	194,364[a]
df	3
Asymp. Sig.	,000

a. 0 cells (0,0%) have expected frequencies less than 5. The minimum expected cell frequency is 49,5.

Results...

χ^2 test

- χ^2 goodness-of-fit test (interpretation)
  - We reject the null hypotheses. Business relationship age does not have a uniform distribution in the population ($\chi^2(3)$ 194,3; p=0,000)
  - Rejeita-se a hipótese nula. A “business relationship age” não tem uma distribuição uniforme no universo ($\chi^2(3)$ 194,3; p=0,000)

Kolmogorov-Smirnov test

- One of the most important nonparametric tests for ordinal data
- 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

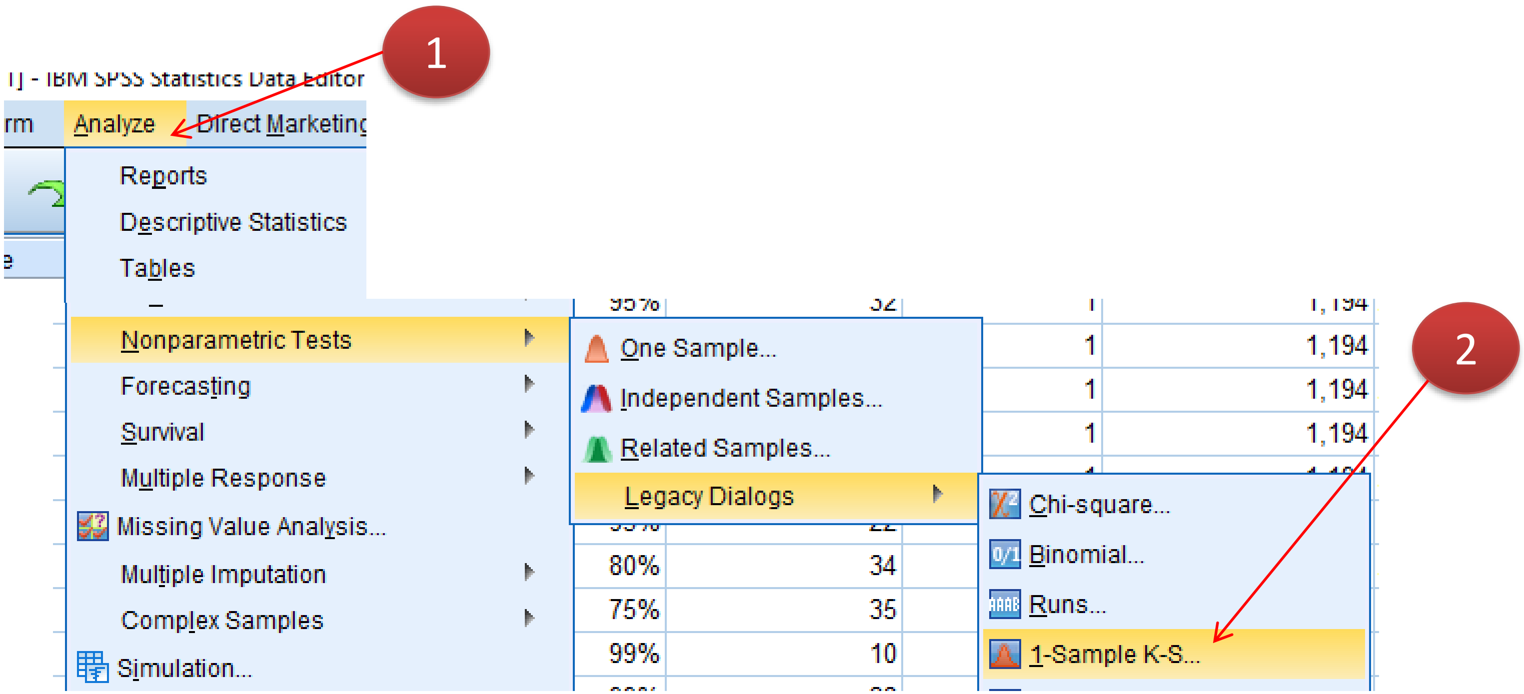
Kolmogorov-Smirnov test

- The test's null hypothesis is that the variable follows a specific distribution (e.g., the normal distribution)
  - Meaning 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
- K-S test is very sensitive when used on very large samples and often rejects the null hypothesis if very small deviations are present
  - The Lilliefors correction rectifies K-S test (included in SPSS 21.0.0.1 and later only)
  - For samples sizes of less than 50: one should use Shapiro – Wilk test

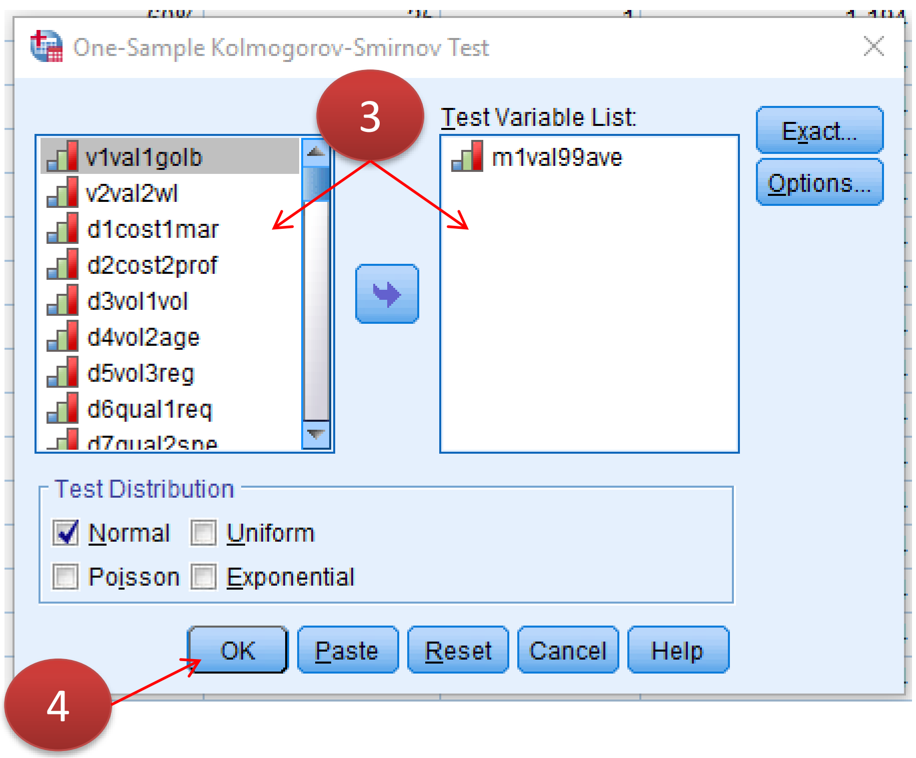
Kolmogorov-Smirnov test

- We want to test if the relationship overall value (m1val99ave) has a normal distribution in the population
- Hypotheses
  - H_0: The relationship overall value (m1val99ave) has a normal distribution in the population
  - H_1: The relationship overall value (m1val99ave) does not have a normal distribution in the population
- To test these hypotheses in an ordinal variable we need to preform a K-S test

Kolmogorov-Smirnov test



Kolmogorov-Smirnov test



Results...

One-Sample Kolmogorov-Smirnov Test

		value average
N		198
Normal Parameters[a,b]	Mean	5,24
	Std. Deviation	1,270
Most Extreme Differences	Absolute	,103
	Positive	,084
	Negative	-,103
Test Statistic		,103
Asymp. Sig. (2-tailed)		,000[c]

a. Test distribution is Normal.

b. Calculated from data.

c. Lilliefors Significance Correction.

Kolmogorov-Smirnov test

- Kolmogorov-Smirnov test (interpretation)
  - We reject the null hypotheses. The relationship overall value does not follow a normal distribution in the population (*K-S*=0,103; *p*=0,000)
  - Rejeita-se a hipótese nula. A variável "relationship overall value" não segue uma distribuição normal (*K-S*=0,103; *p*=0,000) na população

Bivariate

- Bivariate analysis
  - Involves the analysis of two variables
    - often denoted as X, Y
    - With the purpose of determining the empirical relationship between them

- Open/continue with file “OH7 - survey analysis v4.sav”

Samples

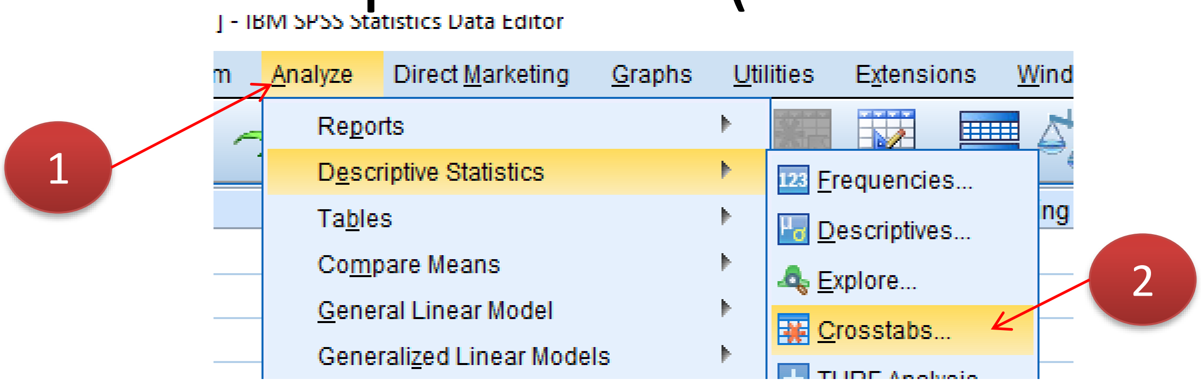
- Two samples can either be
  - Independent
    - We compare two distinct groups (e.g., males vs. females)
  - Paired
    - We can compare the same group at different points in time (e.g., customers before and after being exposed to a treatment or campaign) or compare two items within a survey

χ^2 test for independence

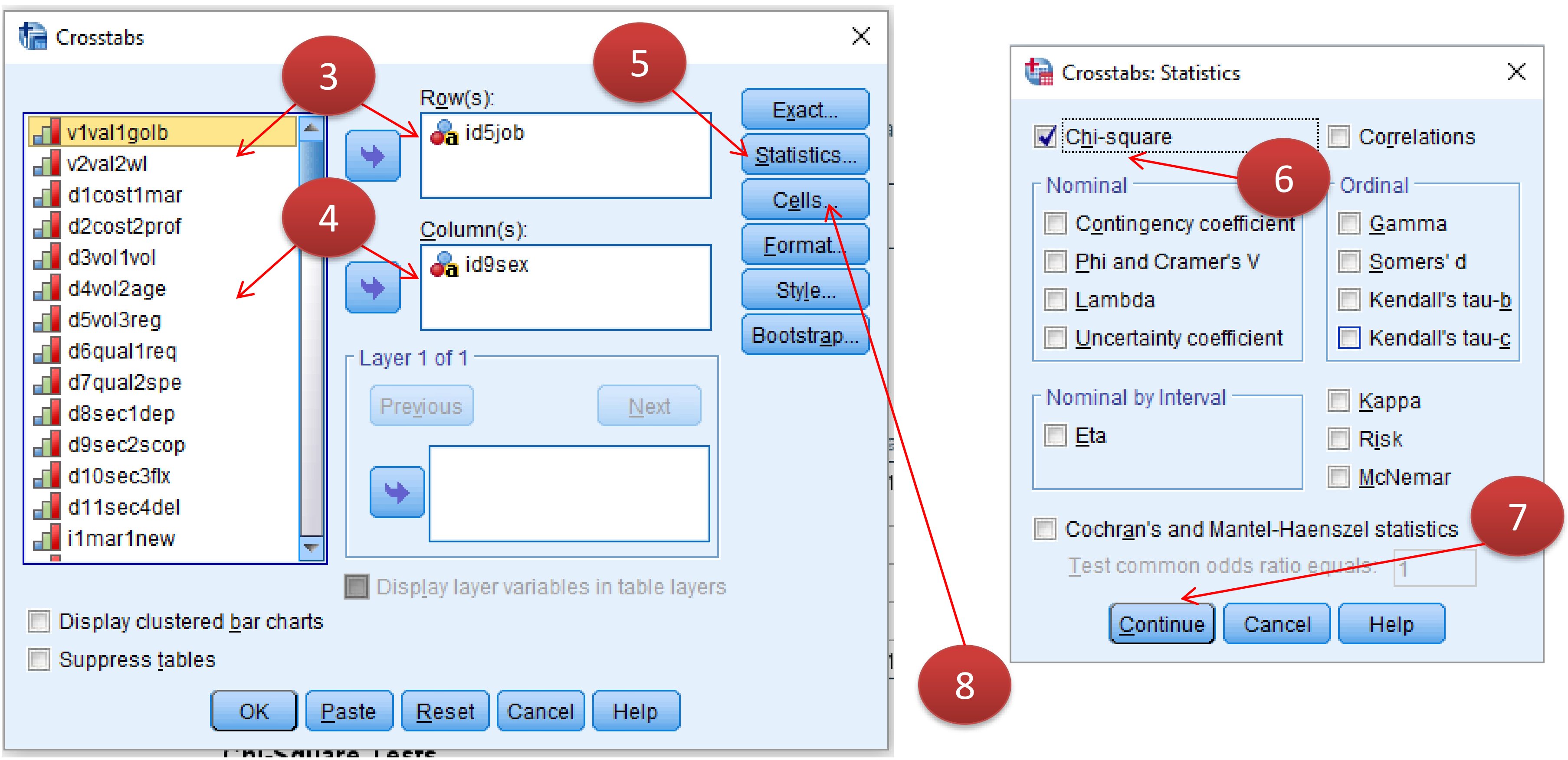
- To explore the relationship between two nominal variables from independent samples
  - If the expected frequency for a cell is less than 5: one should use the Fisher's exact test (annex)
- We can be interested in evaluating whether there is a significant relationship between two nominal variables
  - Herein, we want to know if there is a relation between the job of the respondent (id5job) and their gender(id9sex)

χ^2 test for independence

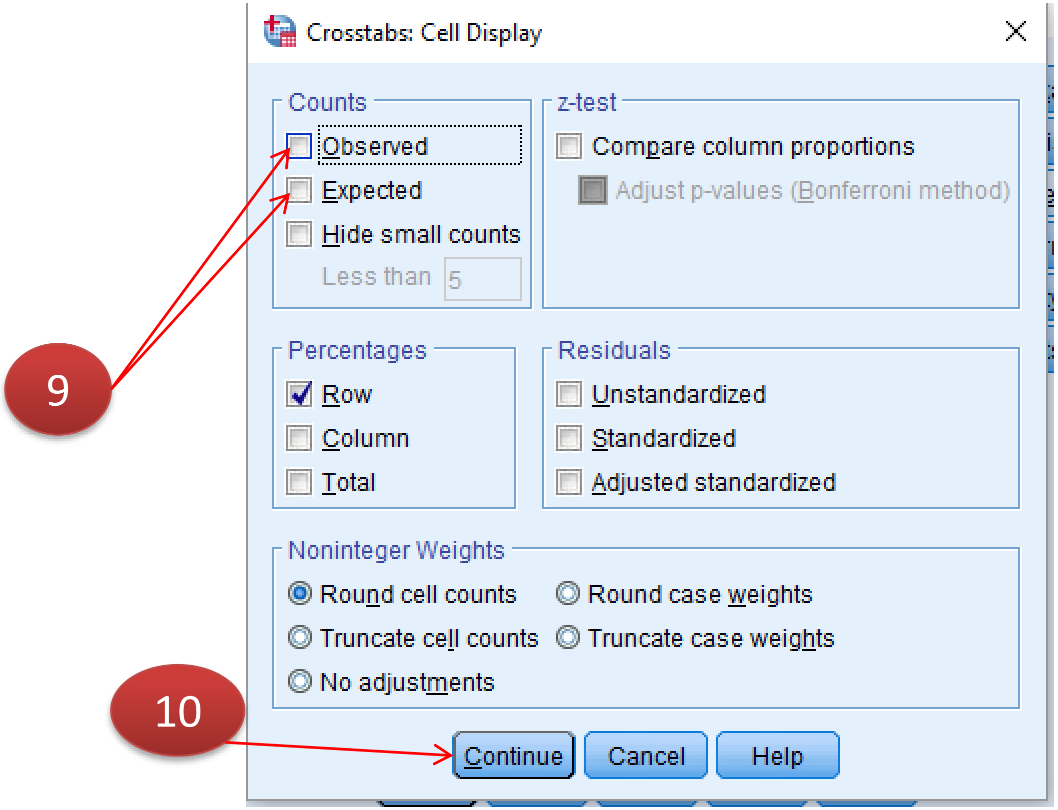
- Hypotheses
  - H_0: The job of the respondent (id5job) is independent of gender (id9sex)
  - H_1: The job of the respondent (id5job) is dependent of gender (id9sex)
- χ^2 test for independence (with crosstabs)



χ^2 test for independence



χ^2 test for independence



χ^2 test for independence

- χ^2 test for independence (crosstabs results)

id5job:A minha função é: * id9sex:O meu género é: Crosstabulation

Count

		id9sex:O meu género é: F	M	Total
id5job:A minha função é:	Administrador / Gerente / Proprietário	10	96	106
	Comercial	18	8	26
	Director	18	10	28
	Gestor	20	18	38
Total		66	132	198

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	60,762[a]	3	,000
Likelihood Ratio	64,648	3	,000
N of Valid Cases	198		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 8,67.

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")

χ^2 test for independence

- χ^2 test for independence (interpretation)
  - We reject the null hypotheses. The job of the respondent is dependent of gender, differences are statistically significant ($\chi^2(3)$ 60,8; p=0,000)
  - Rejeita-se a hipótese nula. A função desempenhada pelos inquiridos é dependente do seu genero ($\chi^2(3)$ 60,8; p=0,000)

(to assess and interpret the strength of the association: see annex)

Mann-Whitney U-test

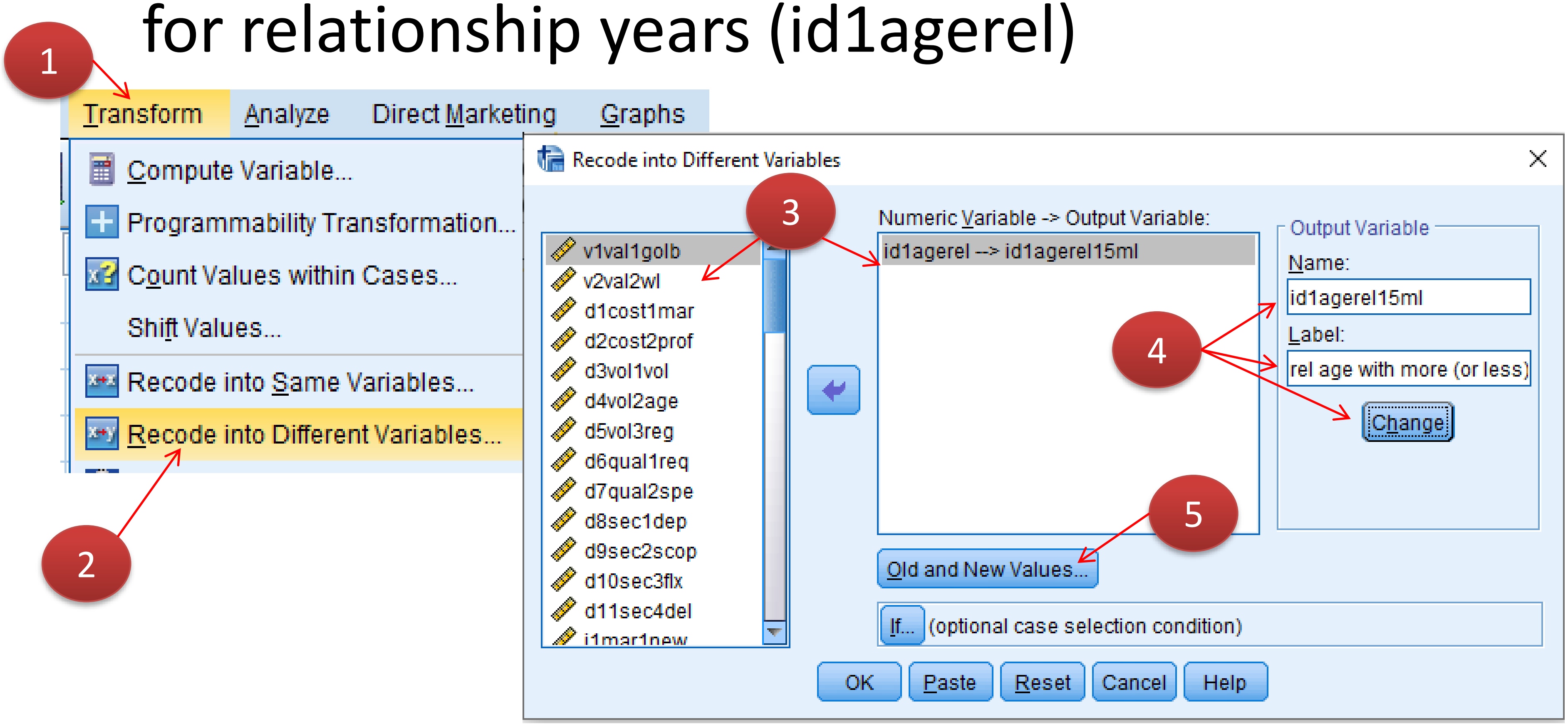
- Nonparametric alternative to the independent samples t-test
  - For two independent samples
  - Used if the dependent variable is measured on (at least) an ordinal scale
  - Commonly applied in situations where the dependent variable is measured on an interval scale but does not follow a normal distribution

Mann-Whitney U-test

- We want to test if there is a relation between the age (id1agerel15ml) and the overall value (m1val99ave) of business relationship
  - Do business relationships with less (and more) than 15 years have different overall value?
- Hypotheses (bilateral)
  - H_0: Business relationships with less (and more) than 15 years have the same overall value
  - H_1: Business relationships with less (and more) than 15 years have different overall value

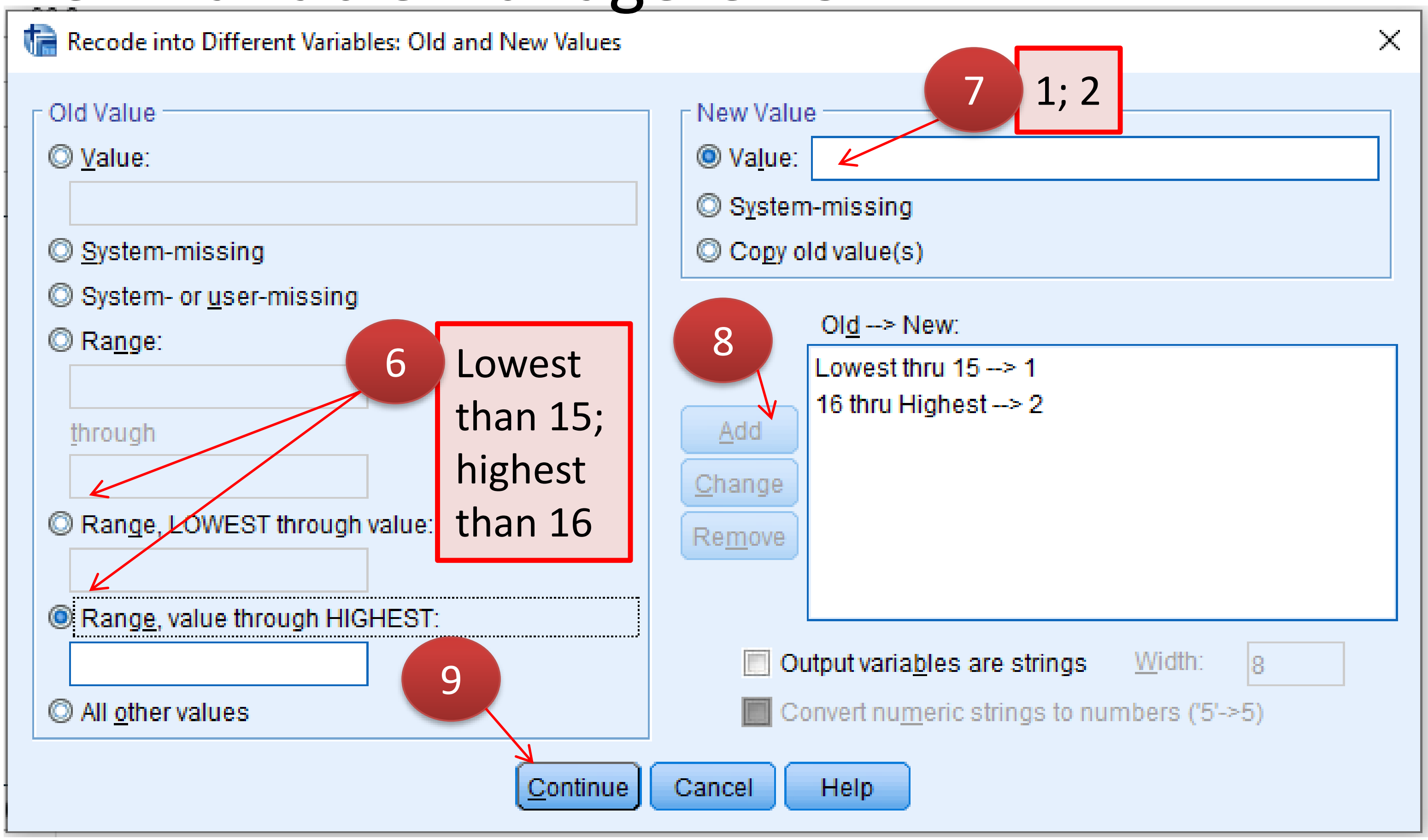
Mann-Whitney U-test

- For this test we need a new ordinal variable for relationship years (id1agerel)



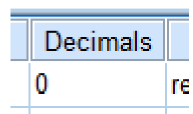
Mann-Whitney U-test

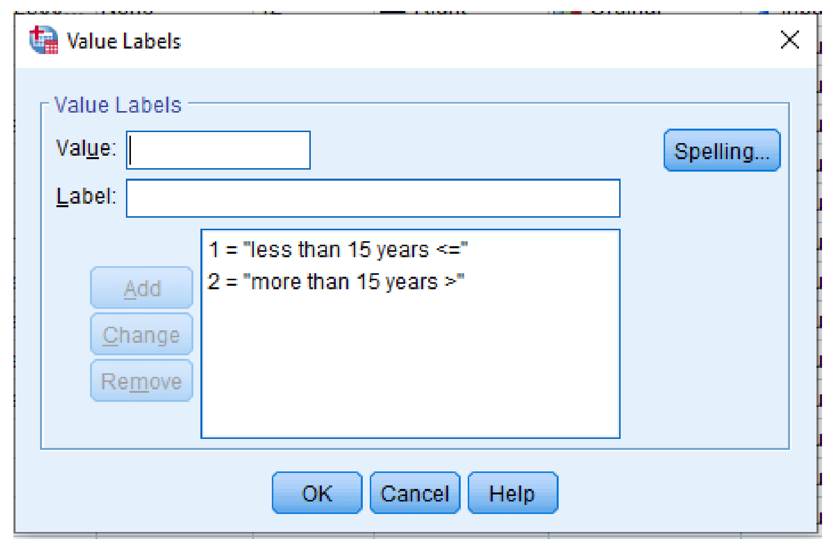
- New variable “id1agerel15ml”



Mann-Whitney U-test

- New variable “id1agerel15ml”
  - Edit variables features





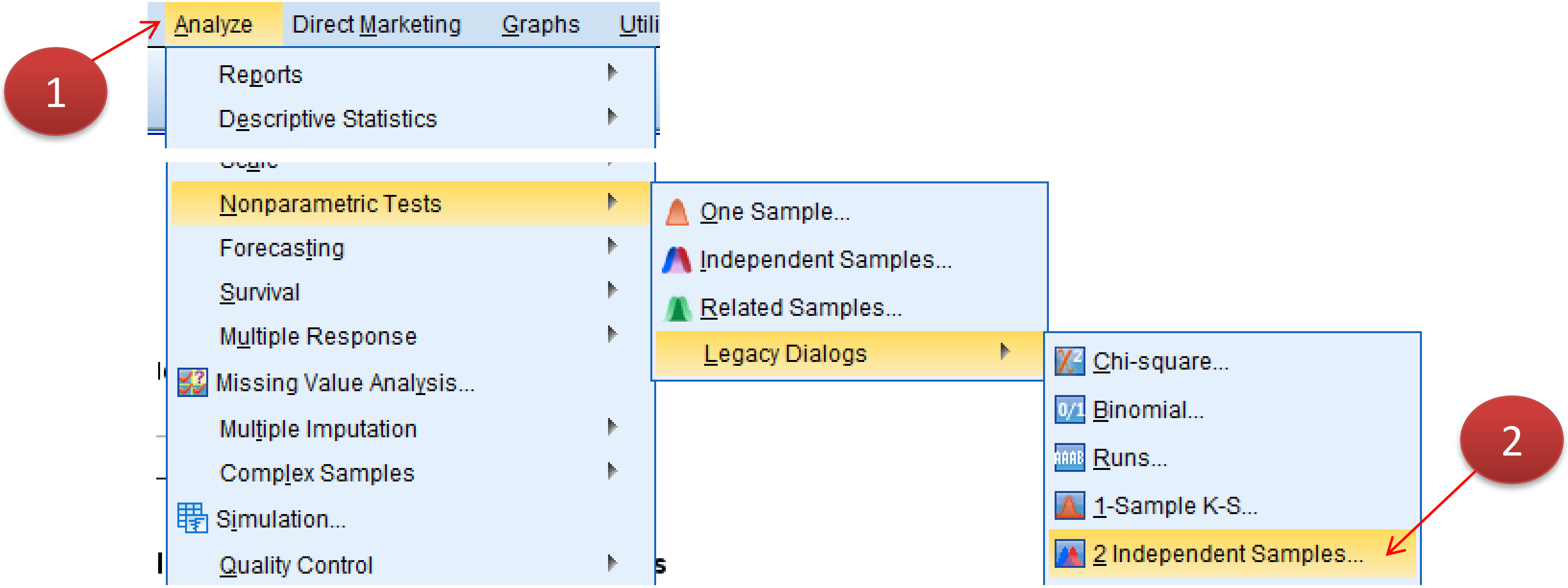
  - Variable frequencies

rel age with more (or less) than 15 years

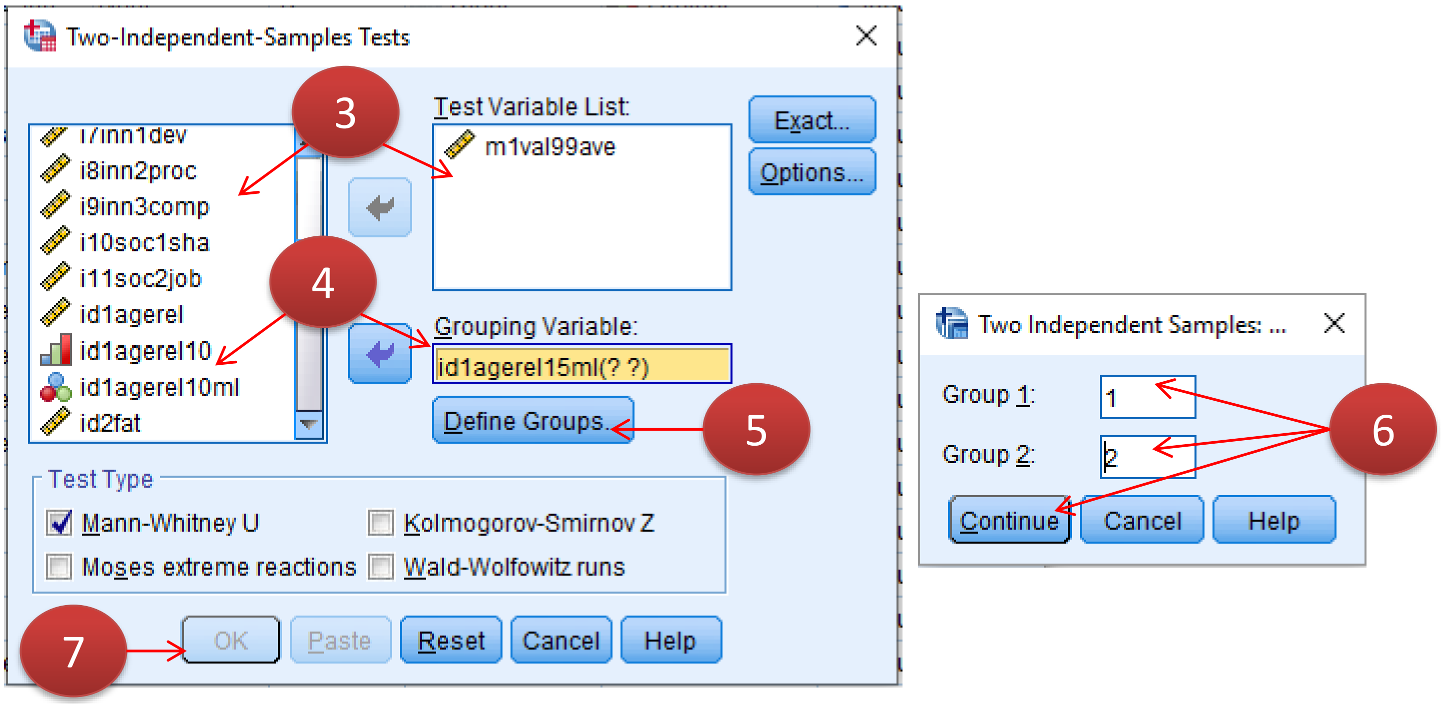
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	less than 15 years <=	171	86,4	86,4	86,4
	more than 15 years >	27	13,6	13,6	100,0
	Total	198	100,0	100,0	

Mann-Whitney U-test

- Continue with the same SPSS file or open file "OH8 - survey analysis v5.sav"



Mann-Whitney U-test



Mann-Whitney U-test

- M-W interpretation

Ranks

	rel age with more (or less) than 15 years	N	Mean Rank	Sum of Ranks
value average	less than 15 years <=	171	94,99	16243,00
	more than 15 years >	27	128,07	3458,00
	Total	198		

Test Statistics[a]

	value average
Mann-Whitney U	1537,000
Wilcoxon W	16243,000
Z	-2,813
Asymp. Sig. (2-tailed)	,005

a. Grouping Variable: rel age with more (or less) than 15 years

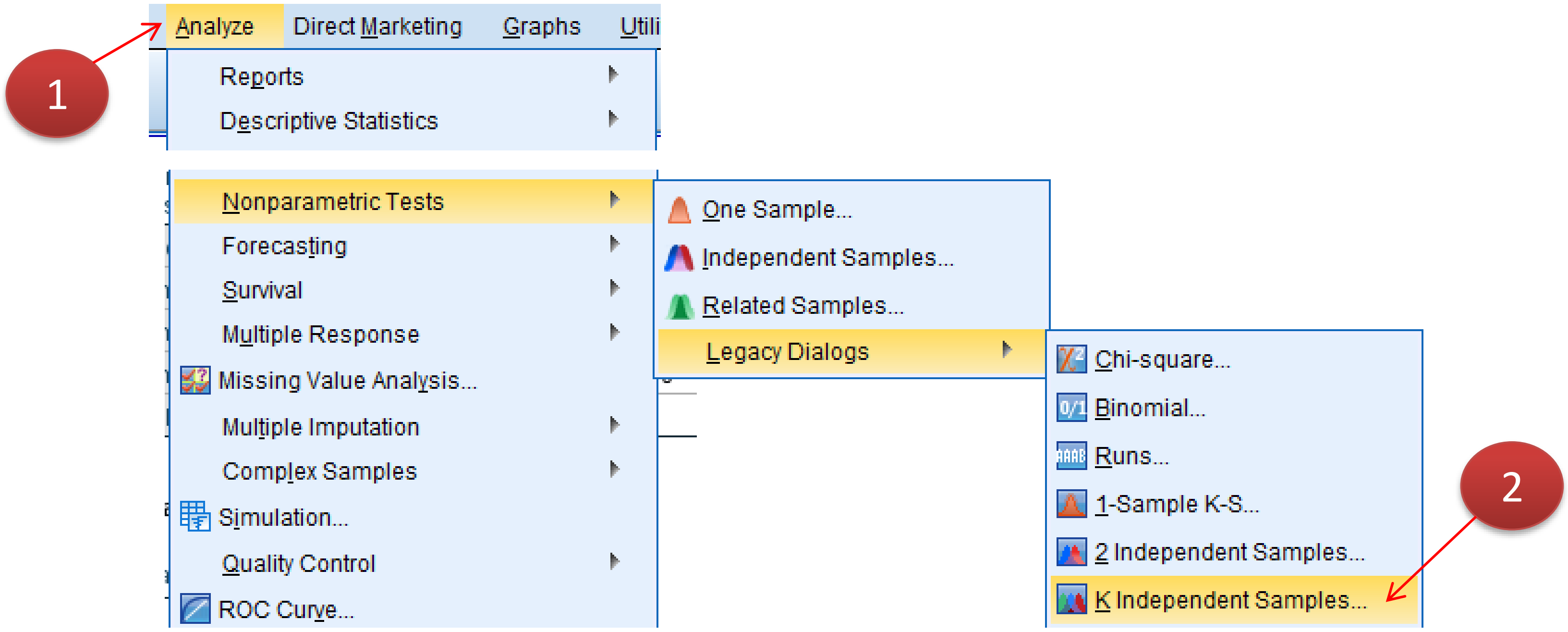
  - We reject the null hypotheses. Older business relationships are perceived as having more overall value (*Mdn*=128,07) than younger ones (*Mdn*=94,99). The Mean Rank is superior in business relationships with more than 15 years and this differences are statistically significant (*U*=1537; *z*=-2,813; *p*=0,005).
  - Rejeita-se a hipótese nula. Os relacionamentos empresariais mais longos têm maior valor. A média das ordenações é superior nos relacionamentos empresariais com mais de 15 anos e as diferenças são estatisticamente significativas (*U*=1537; *z*=-2,813; *p*=0,005).

(to assess and interpret the strength of the association: see annex)

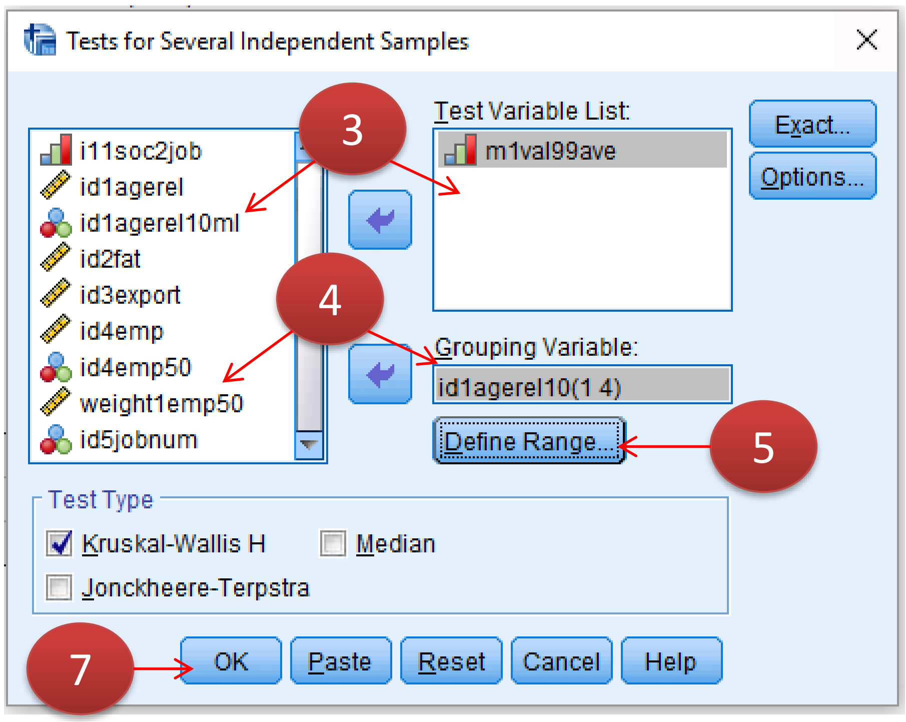
Kruskal-Wallis

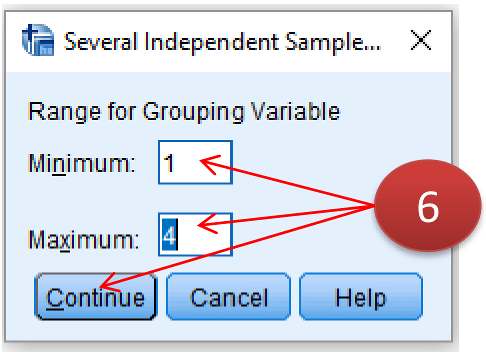
- Nonparametric alternative to the ANOVA one-way
  - For several (k) independent samples
- We want to test if there is a relation between the age (id1agerel10) and the overall value (m1val99ave) of business relationship
- Hypotheses (bilateral)
  - H_0: The business relationship value have the same distribution in the several categories of relationship age
  - H_1: The business relationship value have the different distribution in the several categories of relationship age

Kruskal-Wallis



Kruskal-Wallis





Kruskal-Wallis

- Kruskal-Wallis interpretation

Ranks

	relationship age 10 to 10 years	N	Mean Rank
value average	0 through 9 years	123	89,30
	10 through 19 years	65	117,54
	20 through 29 years	2	80,50
	30 through 39 years	8	114,50
	Total	198	

Test Statistics[a,b]

	value average
Chi-Square	11,305
df	3
Asymp. Sig.	,010

a. Kruskal Wallis Test

b. Grouping Variable: relationship age 10 to 10 years

  – We reject the null hypotheses. Business relationships between 10 and 19 years present the highest overall value (mean rank=117,54) and business relationships between 20 and 29 years present the lowest overall value (mean rank=80,50). This differences are statistically significant (K-W (3) =11,305; p=0,010).
  – Rejeita-se a hipotese nula. Os relacionamentos empresariais entre 10 a 19 anos são os que revelam maior valor global (mean rank=117,54) e os relacionamentos entre 20 a 29 anos são os que revelam menor valor global (mean rank=80,50). As diferenças são estatisticamente significativas (K-W (3) =11,305; p=0,010).

Kruskal-Wallis

- Kruskal-Wallis (post-hoc tests and effect size)
  - If we obtain a statistically significant result for your Kruskal Wallis Test, we still don't know which of the groups are statistically significantly different from one another
    - So, we will need to do some follow-up Mann-Whitney U tests between pairs of groups
    - To control for Type 1 errors, we will need to apply a Bonferroni adjustment to the alpha values if we intend to compare each group with one another (1 with 2, 1 with 3 and 2 with 3…)
  - Bonferroni adjustment involves dividing the alpha level of .05 by the number of tests that you intend to use and using the revised alpha level as your criteria for determining significance
    - Instead of doing every possible comparison… can be better to just select a few key groups to compare, keeping your alpha level at a manageable level
  - For each of the group comparisons, you can calculate an effect size statistic

    (Pallant, 2011, chapter 16)

Annex

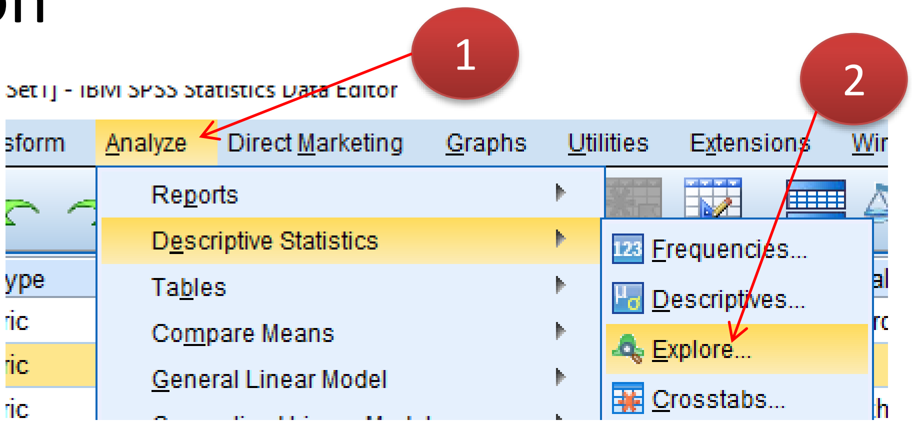
- Extension…
  - Univariate
    - Shapiro-Wilk
  - Bivariate
    - χ^2 test for independence (with "tables" menu)
    - χ^2 test for independence (effect size)
    - Fisher's exact test
    - McNemar test
    - Mann-Whitney U-test (effect size)
    - Wilcoxon signed-rank test
    - Friedman test

Shapiro-Wilk

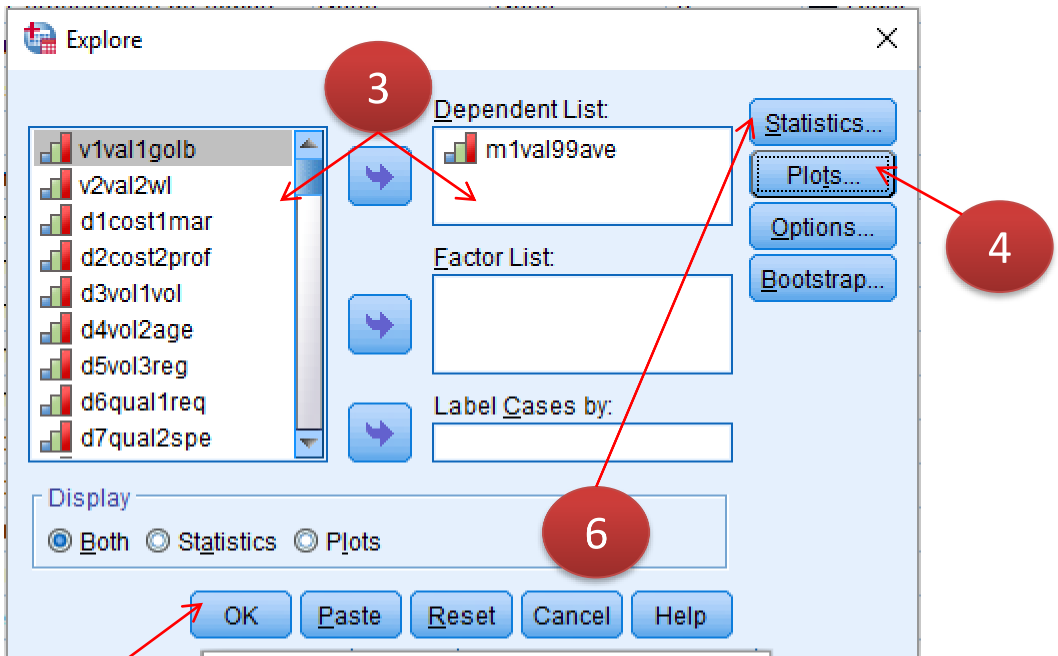
- Shapiro-Wilk
  - Tests the null hypothesis that the test variable under consideration is normally distributed (such as K-S)
    - Thus, rejecting the Shapiro-Wilk test provides evidence that the variable is not normally distributed
  - For sample sizes of less than 50
  - It works poorly if the variable you are testing has many identical values
    - In which case you should use the Kolmogorov-Smirnov test with Lilliefors correction

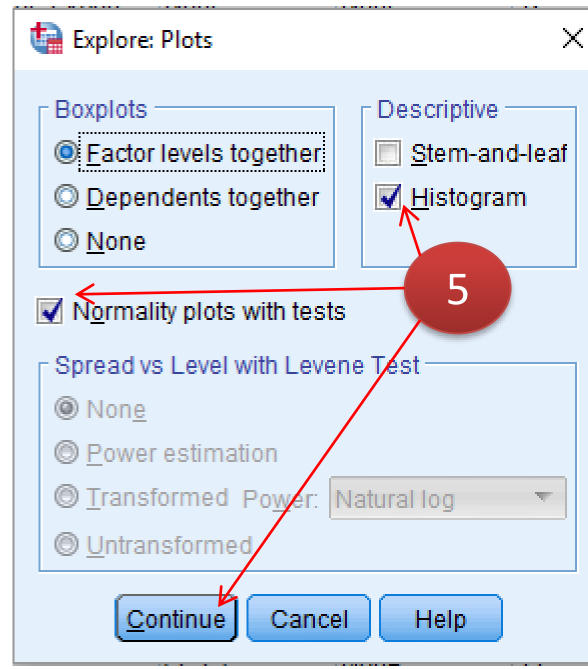
Shapiro-Wilk

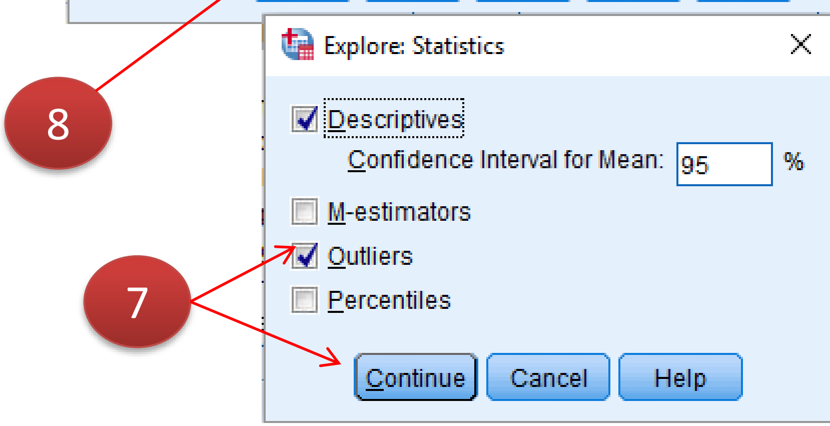
- Hypotheses
  - H_0: The relationship overall value (m1val99ave) has a normal distribution in the population
  - H_1: The relationship overall value (m1val99ave) does not have a normal distribution in the population



Shapiro-Wilk







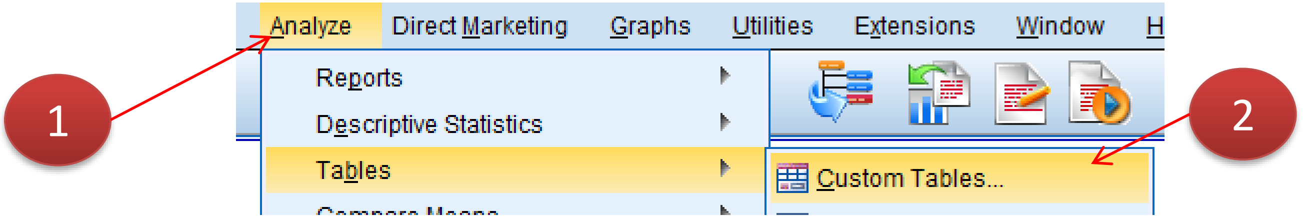
Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
value average	,103	198	,000	,937	198	,000

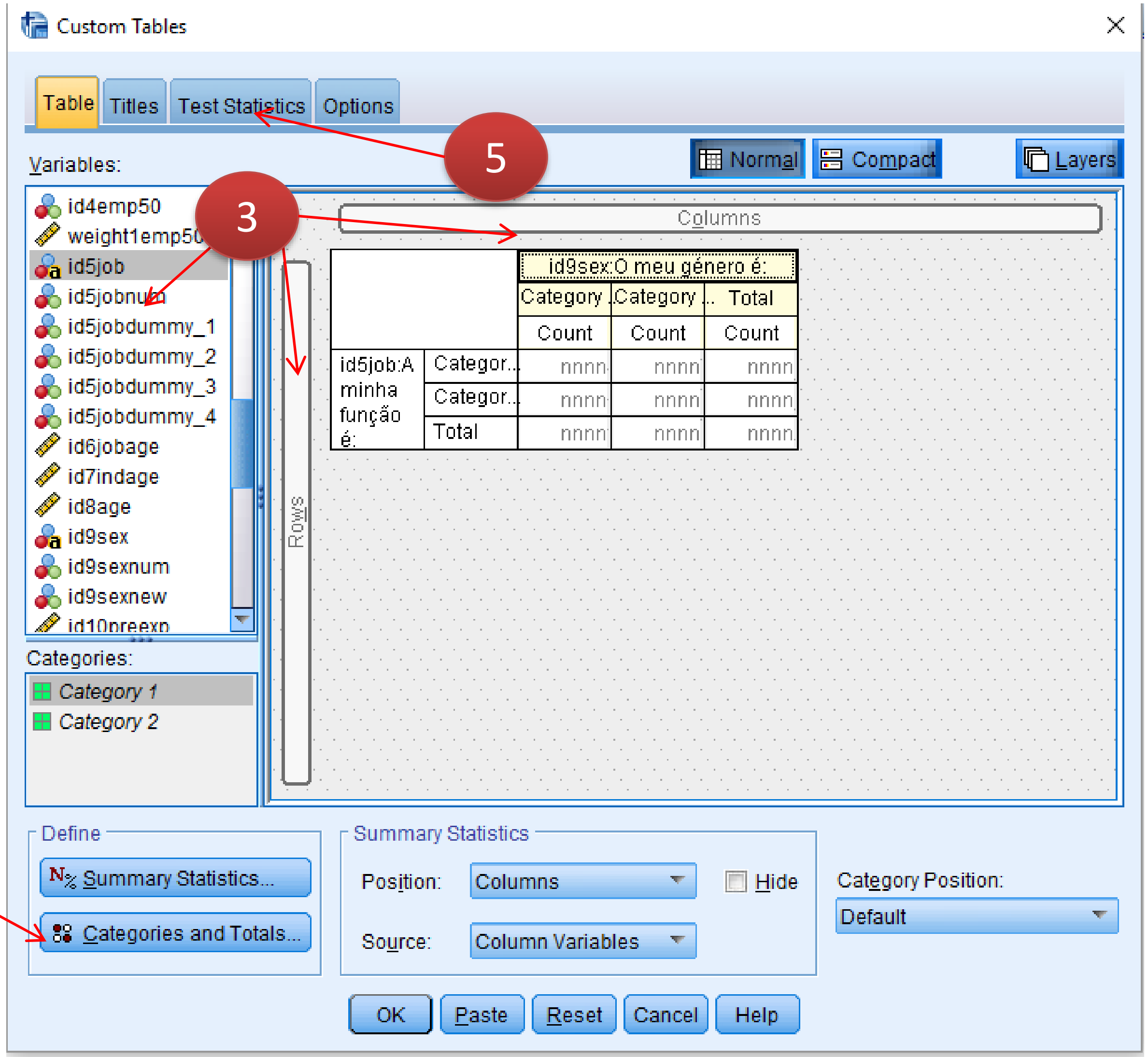
a. Lilliefors Significance Correction

χ^2 test for independence

- χ^2 test for independence
  - With tables - Another way to do the χ^2 test

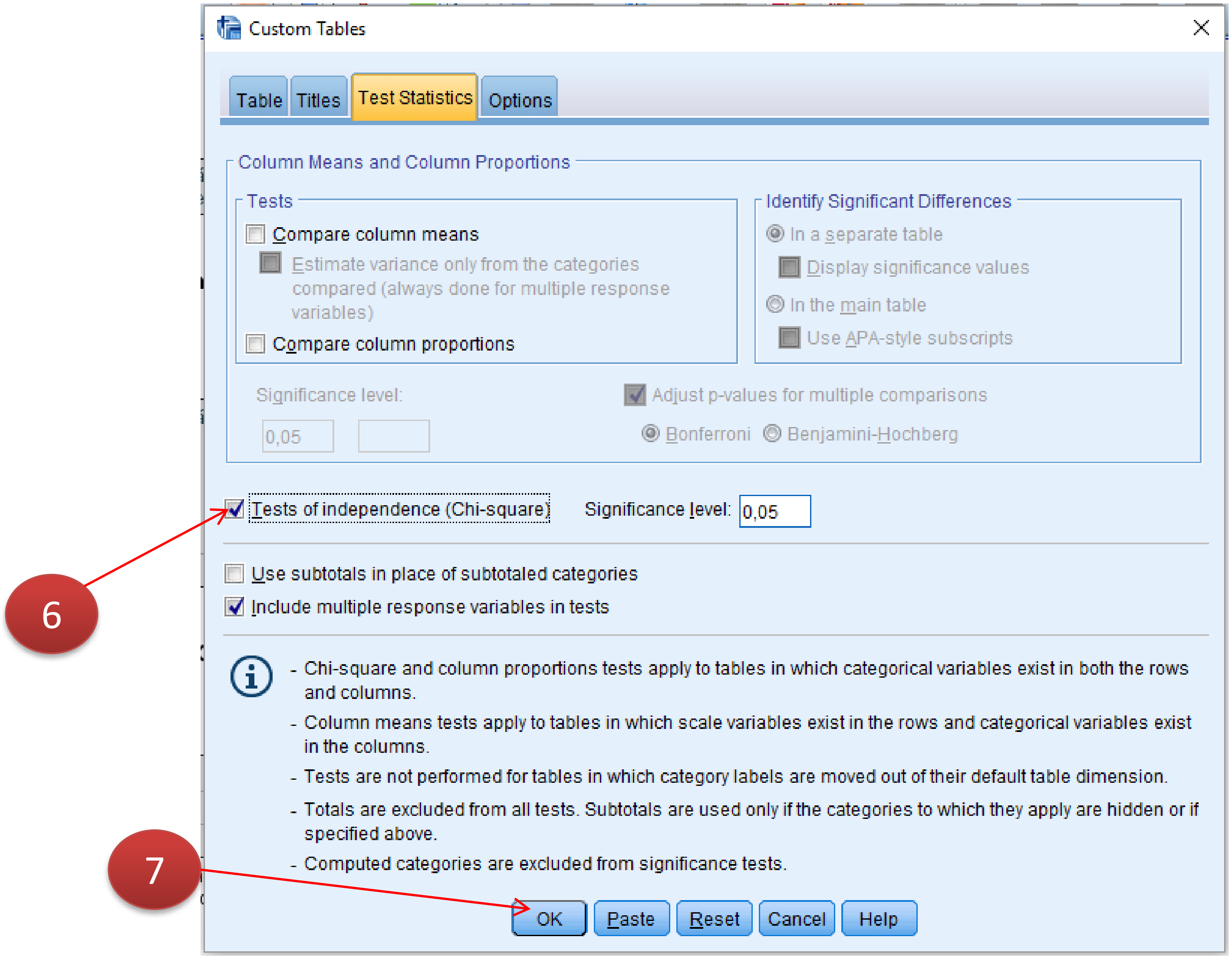


χ^2 test for independence



Add totals

χ^2 test for independence



χ^2 test for independence

- χ^2 test for independence (tables results)

		id9sex:O meu género é: F Count	M Count	Total Count
id5job:A minha função é:	Administrador / Gerente / Proprietário	10	96	106
	Comercial	18	8	26
	Director	18	10	28
	Gestor	20	18	38
	Total	66	132	198

Pearson Chi-Square Tests

		id9sex:O meu género é:
id5job:A minha função é:	Chi-square	60,762
	df	3
	Sig.	,000*

Results are based on nonempty rows and columns in each innermost subtable.

*. The Chi-square statistic is significant at the ,05 level.

Strength of association

- The χ^2 test
  - Provides a weak indication of the strength of the association
    - The absolute value of the χ^2 -test statistic does not provide any indication regarding the strength of the relationship
  - If the χ^2-test or Fisher's exact test renders significant results, several statistical measures of the strength of association can be computed

Strength of association

- Measures of the strength of association
  - For nominal variables
    - ϕ (phi) coefficient
      - 2x2 crosstabs
      - Results from 0 to 1 (above 0.50 a strong association)
    - Contingency coefficient (CC) and Cramer's V
      - Both are equivalent to the ϕ coefficient for crosstabs with more than two rows and two columns
    - Lambda, uncertainty coefficient, and Tshuprow's T
      - Other measures for nominal variables

Strength of association

- Measures of the strength of association
  - For ordinal variables
    - Kendall's ϯ-b (tau), Kendall's ϯ-c, Somer's d, and ϒ (gamma)
      - 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

Fisher’s exact test

- The χ^2 test requires the expected frequency for each cell to be 5 or higher
  - However, if this were not the case, we would have to revert to the Fisher’s exact test, which bypasses potential problems
    - Test commonly used in crosstabs when sample sizes are small
    - For 2x2 tables, SPSS computes the Fisher’s exact test by default
    - For tables bigger than 2x2 tables, see for instance: http://tinyurl.com/fisher-exact

McNemar test

- Explore the relationship between two nominal variables from paired samples
- 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)

Mann-Whitney U-test

- M-W (effect size)
  - SPSS does not provide an effect size statistic
    - But we can calculate approximate value of r (effect size) with the value of z (that is reported in the output)
  - r = z / square root of N ($r=z/\sqrt{N}$)
    - where N = total number of cases
  - In the example, z = –2,813 and N = 198; therefore the r value is 0,1999
    - This would be considered to have a small effect size using Cohen (1988) criteria of 0,1=small effect; 0,3=medium effect; 0,5=large effect
    - The results of this analysis could be presented as:
    - (U=3358; z=-3,236; p=0,001; r=0, 2)

Wilcoxon signed-rank test

- Wilcoxon signed-rank test
  - Nonparametric equivalent of the paired-samples t-test
    - Corresponding method for paired samples of the Mann-Whitney U-test
  - Tests the null hypothesis that two medians stemming from paired samples are identical
    - Is used in situations in which there are two sets of scores to compare (e.g. at time 1 and time 2), but these scores come from the same participants

Friedman test

- Friedman test
  - Is the non-parametric alternative to the one-way repeated measures analysis of variance
    - Corresponding method for paired samples of the Kurskal-Wallis
  - It is used when you take the *same* sample of participants or cases and you measure them at *three or more* points in time, or under three different conditions