Non Parametric tests

These tests all assume that underlying distribution of variables (and/or estimated variables like the residuals in a regression) follow some "parametric" distribution - the usual assumption is that the variables are distributed as a "normal" distribution. We placed a great emphasis on checking whether a variable was distributed normally.

1. Binomial test

Let's assume we have a variable whose distribution is binomial. That is, the variable can take on only one of two possible values, X and Z.

The standard example is a coin toss - the outcomes are distributed as binomial. There are two and only two possible outcomes (heads or tails) and if one occurs on a toss then the other cannot also occur on the same toss. The probability of a "tails" outcome and the probability of a "heads" outcome are the relevant parameters of the distribution. Once these are known, you can calculate the mean, standard deviation, etc.

A variable like *gender* is distributed binomially. We want to test the parameters of the distribution – the probabilities of the variable *gender* taking on the value 0 (or "male") versus the probability of it taking on the value 1 (or "female").

Go to Analyze/NONPARAMETRIC Test/ Legacy Dialogs/ Binomial.

Place the variable *gender* into the area "Test Variable List" (note: you can place more than one variable into the list).

Look at the area "Define Dichotomy." We have chosen "Get from data."

This implies that the two possible outcomes are defined in the data (i.e. – in the values of the variable *gender*). They are: 0 (for male) and 1 (for female).
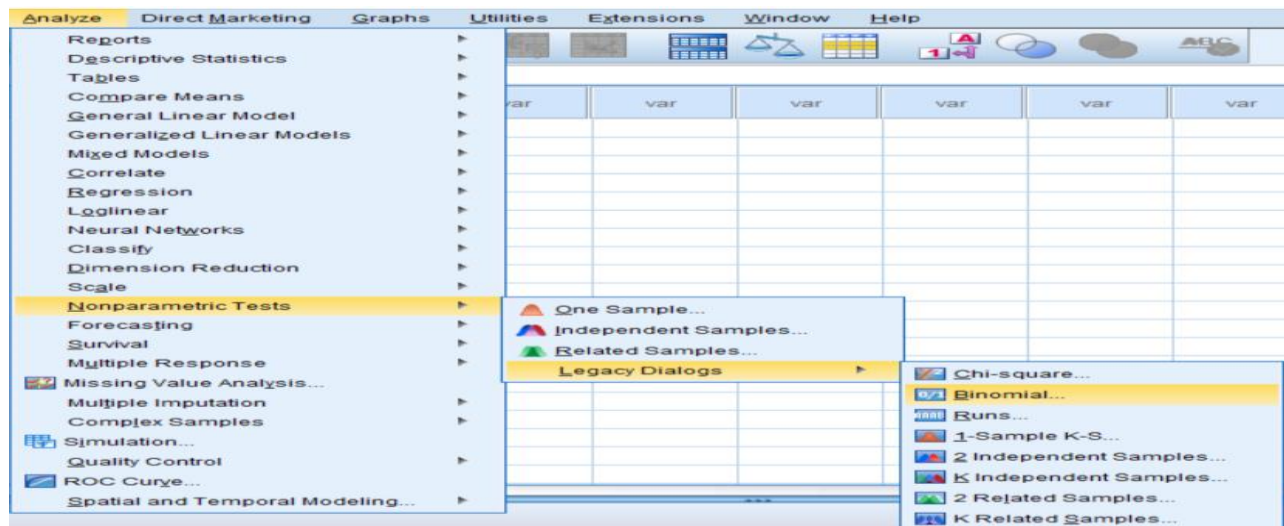
Figure (1): Binomial non parametric test

Look at the box "Test Proportion." We have chosen the default of 0.50. We are asking for a test that checks if the "Test Proportion" of .5 equals the probability of *gender* being equal to 0 ("male") for any one observation. As the probabilities have to add to 1, it follows that we are testing if the probability of *gender* being equal to 1 ("female") for any one observation =1- 0.50 = 0.50. Click on "OK."
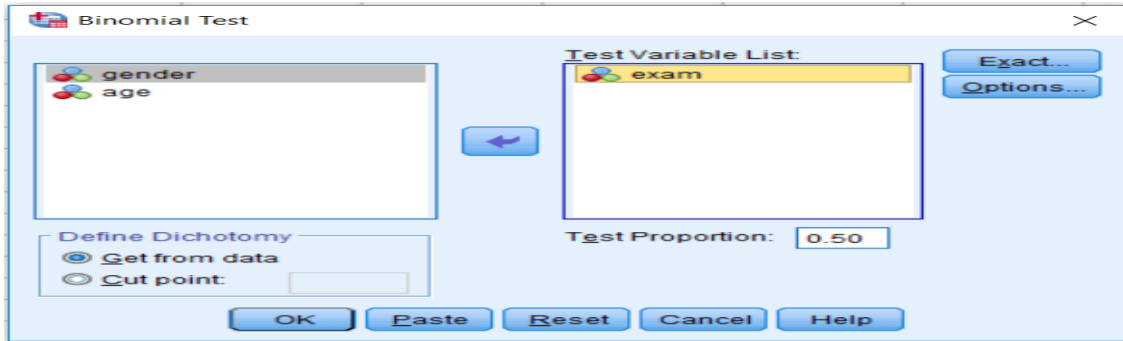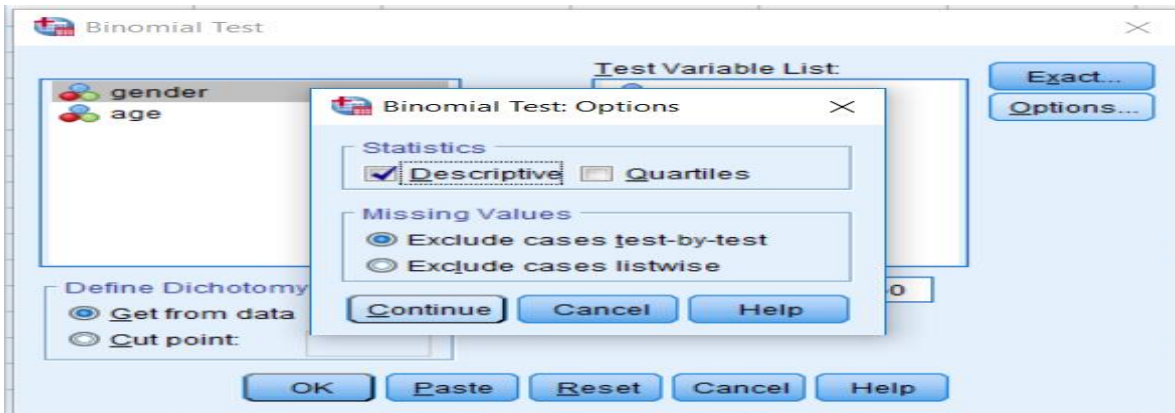

Figure (2): selecting test variables

Figure (3): Determining binomial test options

NPar Tests

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
exam	6	.50	.548	0	1

Binomial Test

		Category	N	Observed Prop.	Test Prop.	Exact Sig. (2-tailed)
exam	Group 1	no	3	.50	.50	1.000
	Group 2	yes	3	.50		
	Total		6	1.00		

Figure (4): Output of binomial test

We repeat the same procedure, but with a different "Test Proportion." We use the proportion of .80. Click on "OK" after entering the hypothesis value of ".80" into the box "Test Proportion."
The observed proportion of group 1 (females) is .8065. Because the Sig value is greater than .1, we can infer that "we cannot reject the hypothesis that the proportion (probability) of group 1 (females) does equal .80." In simpler terms, we can say with 95% confidence that .8 may be the real proportion of group1.

➡ **NPar Tests**

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
exam	6	.50	.548	0	1

Binomial Test

		Category	N	Observed Prop.	Test Prop.	Exact Sig. (1-tailed)
exam	Group 1	no	3	.5	.8	.099[a]
	Group 2	yes	3	.5		
	Total		6	1.0		

a. Alternative hypothesis states that the proportion of cases in the first group < .8.

Figure (5): Binomial test with different proportion

2. Chi-square

It could be either:
- a one-way Chi2 test, which is basically a test that compares the observed frequency of a variable in a single group with what would be the expected by chance.
- a two-way Chi2 test, the most widely used, in which the observed frequencies for two or more groups are compared with expected frequencies by chance. In other words, in this case, the Chi2 tells you whether or not there is an association between 2 categorical variables.
If you run a $\chi2$ on SPSS, it will do it in one step and will give you the level of significance of your test right away. But for you to understand what it is about, let's do it step by step.

To obtain a contingency table with SPSS, you go: Analyze > Descriptive Statistics > Crosstabs.

An important thing to know about the χ2 is that it does not tell you anything about causality; it is simply measuring the strength of the association between 2 variables and it is your knowledge of the biological system you are studying which will help you to interpret the result. Hence, you generally have an idea of which variable is acting the other.

Traditionally in SPSS, the variable which you think is going to act on the other is put in rows. This variable is called the independent variable or the predictor as, in your hypothesis, its values will predict some of the variations of the other variable. The latter, also called the outcome or the dependent variable, as it depends on the values of the predictor, is in column.

A bit of theory: the null hypothesis and the error types.

The null hypothesis (H0) corresponds to the absence of effect (e.g.: the animals rewarded by food are as likely to line dance as the ones rewarded by affection) and the aim of a statistical test is to accept or to reject H0. Traditionally, a test or a difference are said to be "significant" if the probability of type I error is: $\alpha =< 0.05$. It means that the level of uncertainty of a test usually accepted is 5%. It also means that there is a probability of 5% that you may be wrong when you say that your 2 means are different, for instance, or you can say that when you see an effect you want to be at least 95% sure that something is significantly happening.

Example 1: A basic example

The following data classify sample of 7 observations according to gender, age, smoker, and the cholesterol value as shown in Figure 6.

Can we consider that the smoker is independent from age?

	Subject	gender	age	smoker	cholestrol	var	var	var	var	var	var	v
1	1	male	44.00	smoker	2.26							
2	2	male	20.00	non-smoker	6.00							
3	3	male	43.00	non-smoker	6.15							
4	4	male	43.00	non-smoker	6.15							
5	5	female	17.00	non-smoker	6.00							
6	6	female	17.00	non-smoker	7.83							
7	7	female	48.00	smoker	7.18							

Figure (6): smoker data

Go to Analyze/ Descriptive statistics/ Cross tabs

Select gender variable and add it to Rows area, then select smoker variable and add it to columns area as shown in Figure 7.
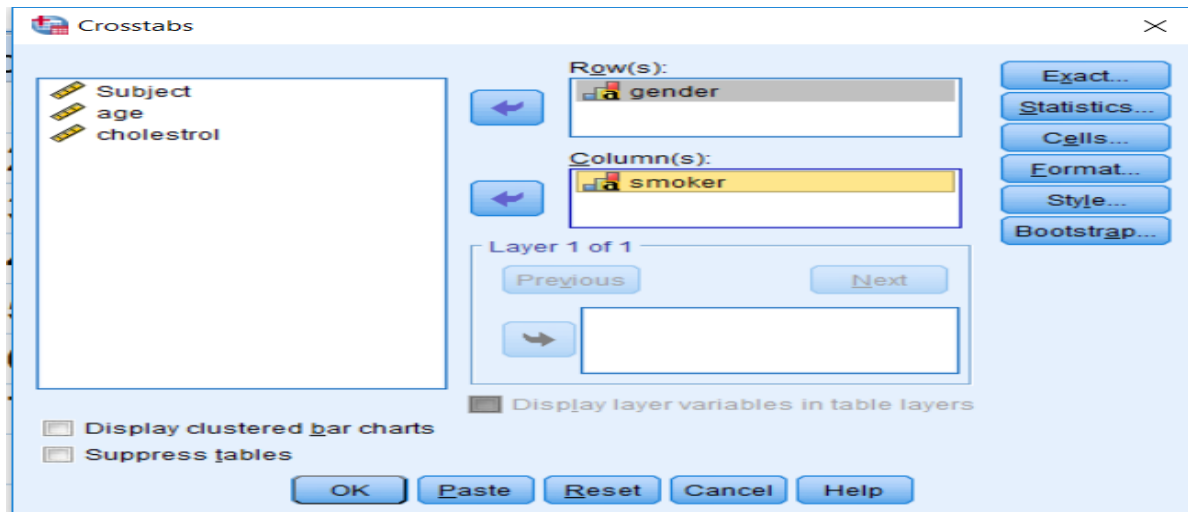

Figure (7): selecting variables for Rows and variables for Columns

From statistics button check chi square then select continue > OK, you will get the cross tabulation of gender and smoker with their related values as shown in Figure 8.

gender * smoker Crosstabulation

Count

		smoker		
		non-smoker	smoker	Total
gender	female	2	1	3
	male	3	1	4
Total		5	2	7

Figure (8): Cross tabulation of gender and smoker

The value of chi square is 0.058 which is larger than alpha value (0.05), then we accept the null hypothesis, that means that smoker is independent from age.