

# Practice 2 SPSS

*Partially based on Notes from the University of Reading: <http://www.reading.ac.uk>*

Go to (file, open, data, open data, type of file)

Archivo → Abrir → Datos

Abrir datos → Archivos de tipo → \*.xlsx

Open file

Credit\_Approval\_Data.xlsx

Recode a nominal variable

Go to (transform, compute variable, destination variable)

Transformar → Calcular Variable

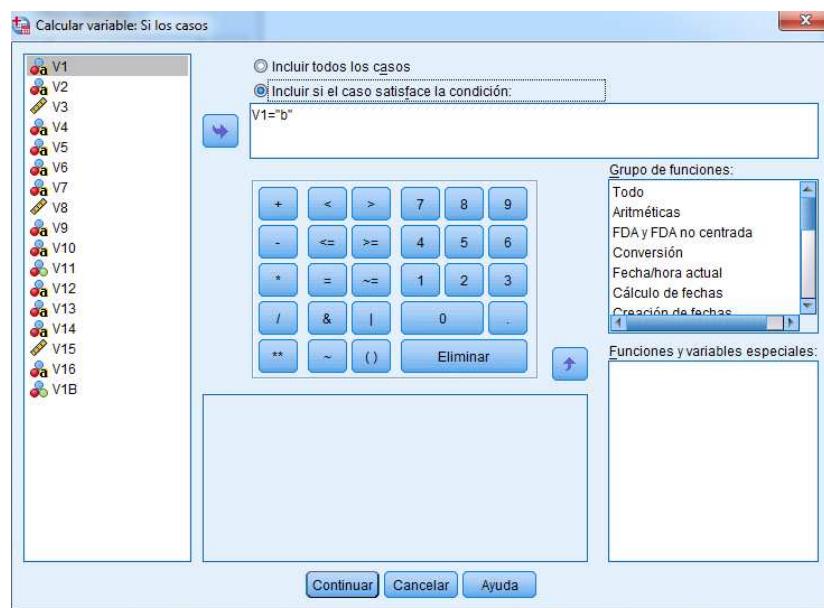
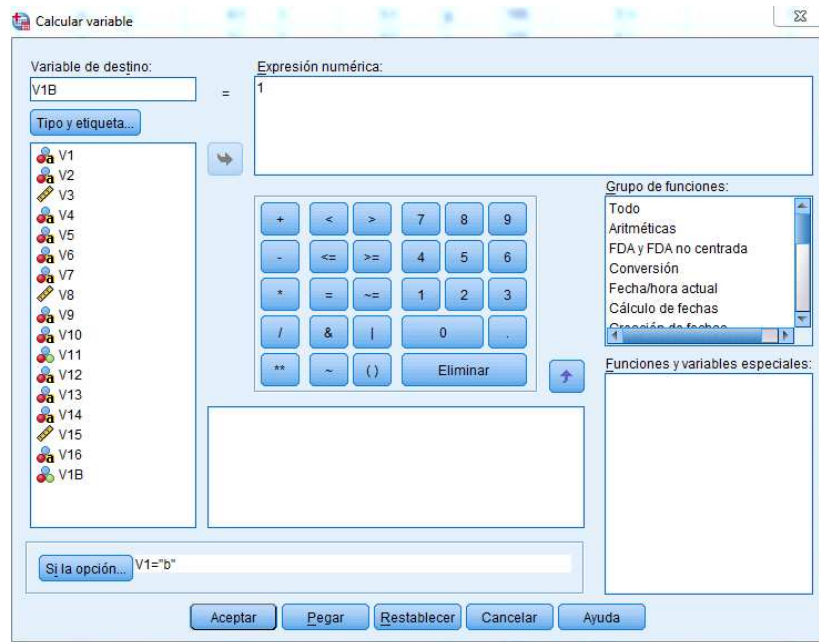
Variable destino = **V1B**

Expresion Numerica: 0

Si la opcion:

X Incluir si el caso cumple la condicion

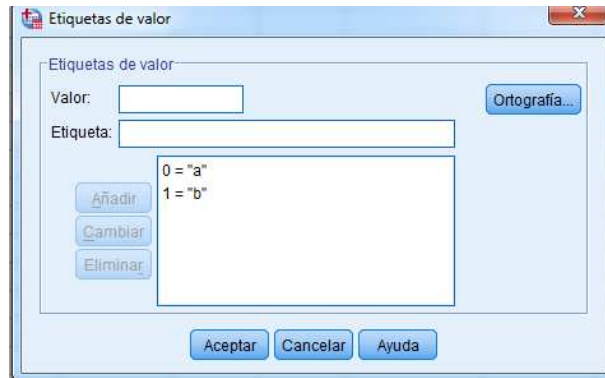
V1 = "a"



Go to (variables view)

Vista de variables:

Go to last row and change labels in **Valores**



# One proportion test

In a this database, is thought that 50% of individuals are  $a$ . You are given a sample of 690 individuals. You want to test the hypothesis that the population proportion of  $a$  is greater than 50%.

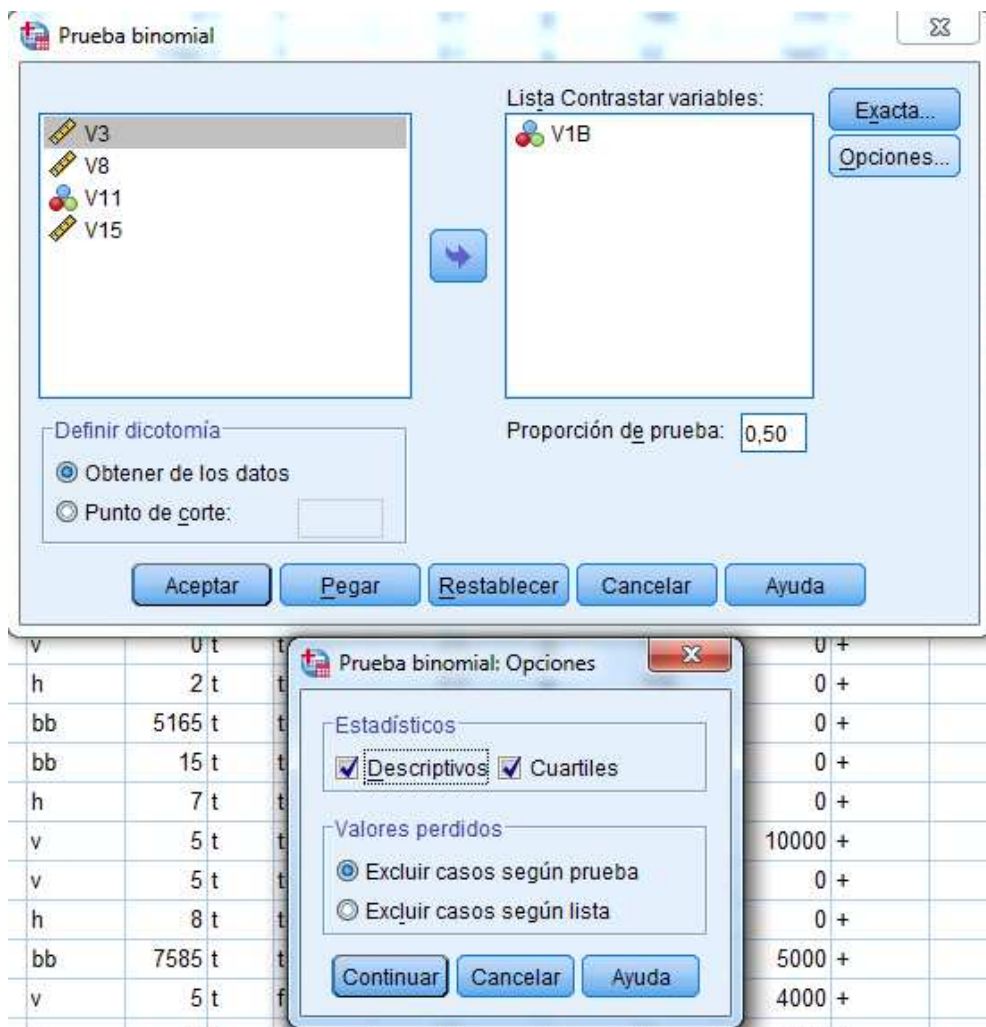
The null and alternative hypotheses are:

Null hypothesis,  $H_0: P = 0.5$

Alternative hypothesis,  $H_1: P \neq 0.5$

To perform the test select *Analyze*, then *Non-Parametric test*, then *Legacy Dialogs* then select *Binomial*.

[Analizar](#) → [Pruebas no parametricas](#) → [Cuadro de dialogos antiguo](#) → [Binomial](#)



### Interpretation:

From the results we can see the **p-value** (*Asymp. Sig.*) is significant as it is less than 0.05, meaning that we reject the null hypothesis that the proportion of **a** is equal to 50%, in favour of the alternative hypothesis that the population proportion of **a** is greater than 50%.

# One sample T-test

A one sample t-test is used when you want to test whether a population mean is equal to some hypothesised value.

When carrying out a one sample t-test there are some assumptions you have to make:

- The variables should be *Normally* distributed.
- Data should be a random and independent sample from the population.

We will use the `diabetic.txt`. The data are from a study on the effect of oral administration of glucose on diabetic patients. Plasma glucose levels were measured before and after the treatment, and age also recorded.

Now suppose we want to test the hypotheses (for change = after – before):

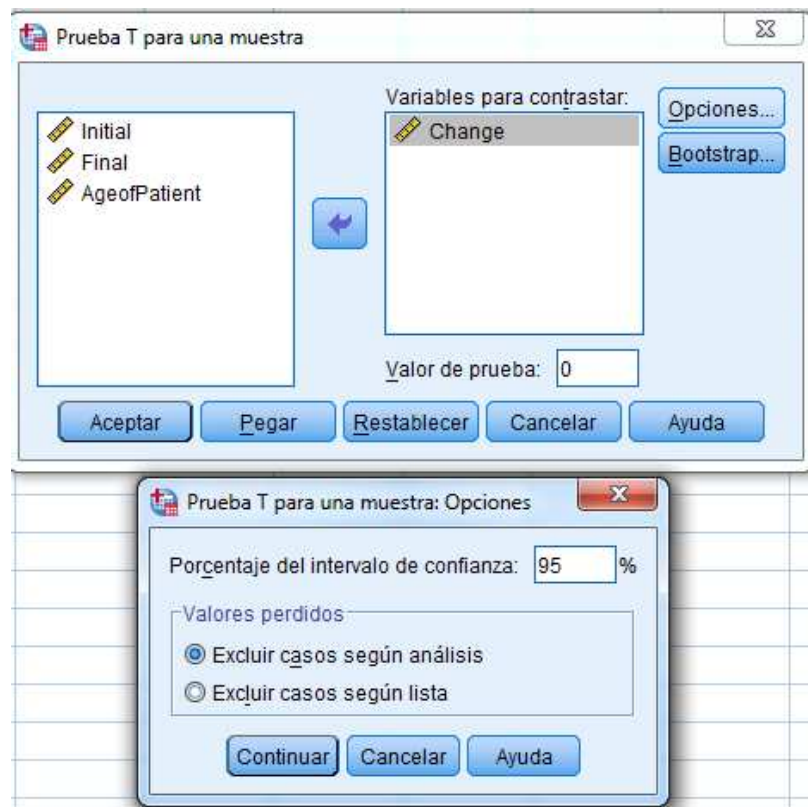
Null hypothesis ( $H_0$ ):  $\mu = 0$

Alternative hypothesis ( $H_1$ ):  $\mu \neq 0$

(**Note**: we are testing this hypothesis on the change variable because we want to know whether the glucose treatment has made a difference to the patient's diabetes.)

To do this, click on **Analyze**, and then **Compare means** followed by **One-Sample T test**.

Analizar → Comparar medias → Pruebas T para una muestra



### Interpretation:

The T statistic ( $t$ ) is a value which can be used to compare with t-tables to see whether the test is significant or not using  $N-1$  degrees of freedom ( $df$ ), in this case 11.

The **p-value** (*Sig.*) determines the significance of the test. In this example the p-value, of 0.001, is highly significant. Thus we reject the null hypothesis that the mean difference of glucose measurements is 0, in favour that it is not.

We can see that the estimated mean is 1.325, which suggests that the later measurements are greater than the earlier ones for a patient.

## Two sample T-test

In a two sample unpaired t-test the data are quantitative and randomly drawn from two different populations.

When carrying out a two sample t-test there are some assumptions you have to make:

- Variables should be normally distributed.
- Data should be a random and independent sample from each population.

For this section we will be use the [lightbulbs.txt](#). This dataset includes the number of days that a sample of light bulbs from two manufacturers lasted for.

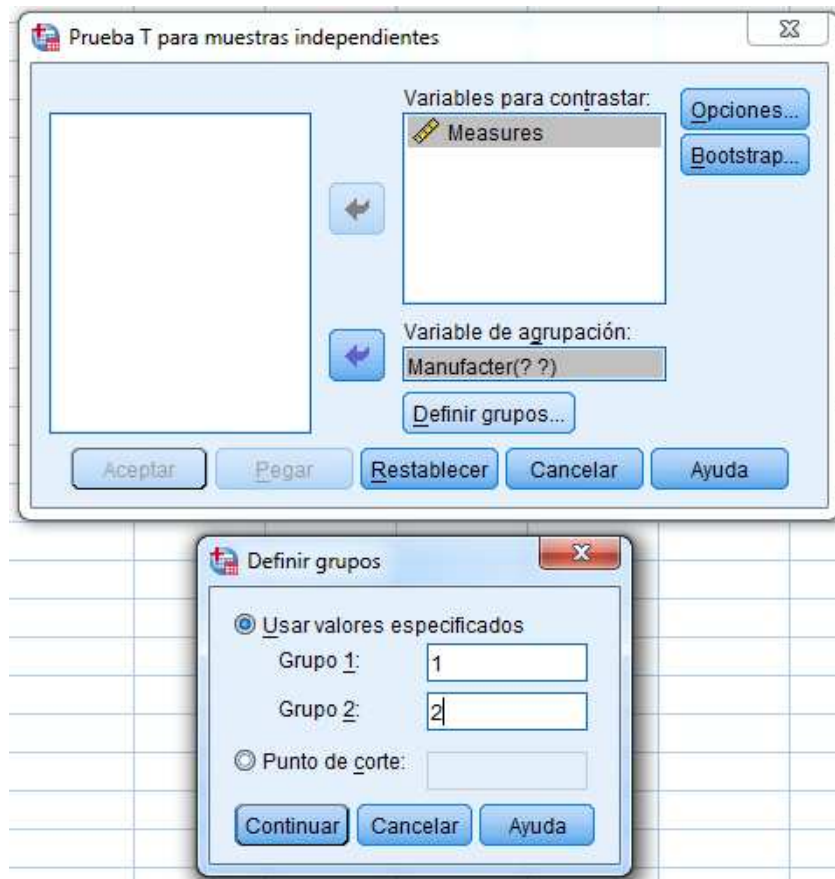
To carry out a two sample t-test in SPSS data must be stacked (i.e. one column for the measurements, and one column for the grouping or *categorical* variable).

Before carrying out a two sample t-test we must check to see whether the equal population variances assumption is reasonable, as this will effect which version of the test we use. However in SPSS this is carried out whilst the two-sample t test is calculated.

Click on **Analyze**, then **Compare Means** followed by **Independent Samples t tests**.

**Analizar** → **Comparar medias** → **Prueba T para muestras independientes**





Results contain the individual means, standard deviations and standard error of the means for each of the manufacturers. In the main output provides information about testing the equal variances assumption. The test statistic and its p-value (*Sig.*) are given.

### Interpretation:

The p-value of 0.544 in this case is large, and indicates that there is not enough evidence to reject the null hypothesis that the two populations have equal variances.

Therefore we will assume equal variances below, as this assumption does not seem to be unreasonable: we will therefore read from the **Equal variances assumed** "*Se han asumido varianzas iguales*" row for the rest of the test output.

For this example we are testing the following hypotheses in the two sample t-test:

Null hypothesis ( $H_0$ ):  $\mu_A = \mu_B$  (Manufacturer A has equal mean to Manufacturer B)

Alternative hypothesis ( $H_1$ ):  $\mu_A \neq \mu_B$  (Manufacturer A's mean is not equal to Manufacturer B).

The T-statistic of -2.095 and p-value of 0.048 on 22 degrees of freedom shows us that we can reject the null hypothesis of the two manufacturers' means being equal and instead conclude that there is a difference between the two at the 5% level.

## Paired T-test

In paired T-tests the data obtained are related in some way, so the two samples of data are not independent.

We will be use the cholesterol.txt dataset. This dataset consists of cholesterol levels in patients two and four days after a heart attack.

Note: these data are paired because the two measurements have been made on the same individuals.

For this example we will be testing the following hypotheses:

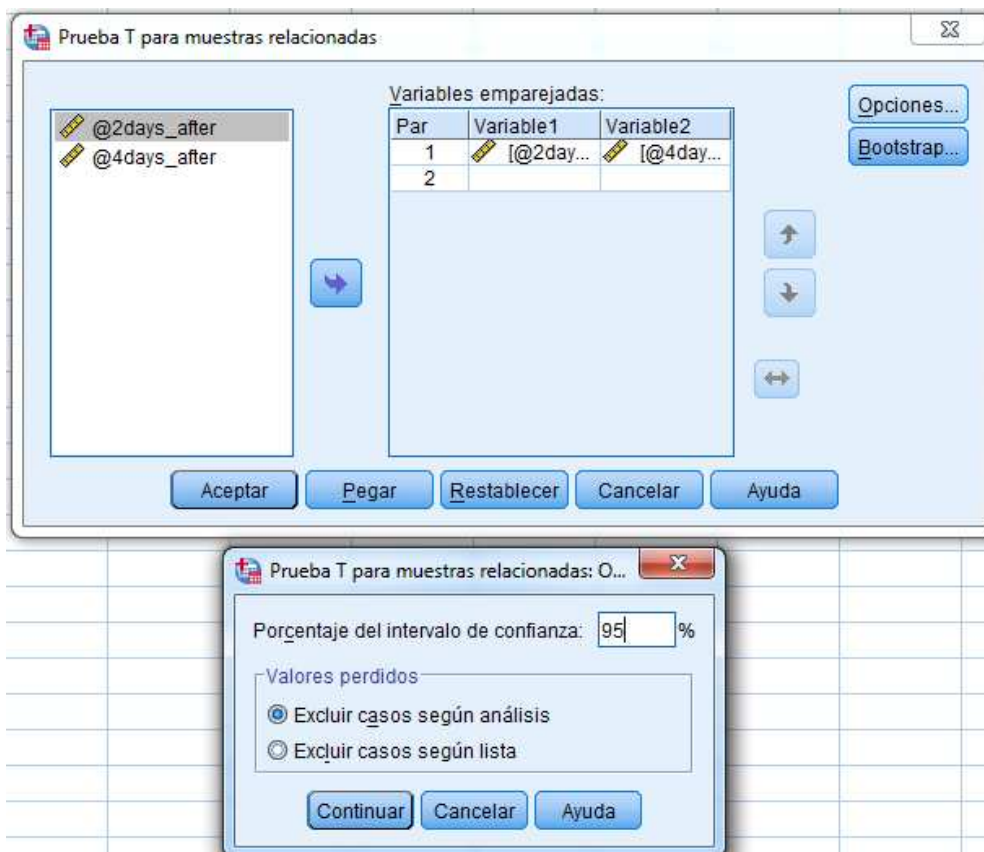
Null hypothesis ( $H_0$ ):  $\mu_A = \mu_B$  (mean cholesterol is the same after 2 days as it is after 4 days)

Alternative hypothesis ( $H_1$ ):  $\mu_A \neq \mu_B$  (The mean cholesterol level is different)

**Note:** The format for the dataset should be that the two samples are stored as different variables (two columns in the datasheet).

Go to **Analyze** followed by **Compare Means** and then click on **Paired Samples T-test**.

**Analizar** → **Comparar medias** → **Prueba T para muestras relacionadas**



Drag across @2daysafter to the Variable 1 column and then @4daysafter to the Variable 2 column. Leave the Options settings as default (for a 95% confidence interval). Click OK.

They are obtained the individual means, standard deviations and standard error of means for cholesterol levels on two and four days after the heart attack.

It is obtained how closely correlated the two variables are. In this example they are very highly correlated (which is expected as the same patients are being measured on both days).

The 95% confidence interval for the mean difference in cholesterol levels between the two time periods, T-statistic (T-value) and p-value are given.

### Interpretation:

The p-value of 0.044 shows us that we can reject the null hypothesis of the two means being equal and instead conclude that there is a difference between the mean cholesterol levels 2 days after and 4 days after, at the 5% level.

**Note:** performing a one sample t-test on the differences between two related variables, is equivalent to performing a paired t-test on the two variables.

## Notes:

- **P-value:** When interpreting the p-value for a test, if the value is less than 0.05 then the test is significant at the 5% level, and we would usually say there is evidence to reject the null hypothesis.

If the p-value is less than 0.1 but greater than 0.05 then there is weak evidence in favour of the alternative hypothesis. Finally if the p-value is greater than 0.1 then we would usually say there is no evidence to reject the null hypothesis.

**Never Accept the null hypothesis** and conclude it to be true as this will be incorrect; we always **reject or do not reject the null**.

- **Confidence Interval:** If the 95% confidence interval of a difference in means does not include zero then there is evidence that there is a difference in the population means (i.e. rejecting the null hypothesis).

- **Standard deviation and Standard Error Mean:** The standard deviation is an estimate of how variable the values are in the population. The standard error of the mean shows the uncertainty of the estimated mean as an estimate of the true unknown population mean (note: this will become smaller as the sample size becomes larger).