November 2019, Examples

Exercise 1: We want to examine to what extent the degree of ingenuity that we assign to a phrase differs when the phrase was created by a person typically considered a good person or not, and at the same time analyze whether this effect differs in university and non-university students. The experimenter selects 10 ingenious phrases. In five of them, he indicates that it was created by a person with bad press (Hitler, ...) and in the other, he indicates that they created by a person with good press (Dalai Lama, ...). 40 people between 20 and 25 years old participated in the study (20 university students, 20 non-university students). Each person had to read all the sentences, in random order, and had indicate the degree of ingenuity of each of the phrases in a Likert-type scale of 1 to 10. He measured, for each participant, the average of the ingenuity score of the five phrases of individuals with good press and the average of ingenuity of the five phrases of people with bad press.

1 - What is / are the independent variable (s)?

2 - What is / are the dependent variable (s)?

3 - What type of design was used in the study: between-subjects, intra-subjects or mixed? Justify it

4 - Is it an experimental design? Justify it

5 - How would you organize the data with SPSS (# rows, #columns? Justify it

6 - The variable "degree of ingenuity" in the study is a ….. (nominal/ordinal/interval) variable?

7 – Let’s assume that the degree of ingenuity was higher for phrases thought by people with good press than for those with bad press. Another researcher raised a concern: we need a baseline to know if the difference is due to the fact that there is a bias in favor of people with good press, or a bias against people with bad press, or the two biases. Can you think of a new study to answer that question?

8 – Let’s assume that the degree of ingenuity was higher for phrases thought by people with good press than for those with bad press. We give a talk with these data, and a researcher indicates a possible problem: Perhaps the five phrases we had for the people with bad press were actually less ingenious than the other five. How would you solve this problem in a new study?

Excercise 2. We have a set of data with different variables: introversion (4 questions, the third is inverse; each question ranged from 1 to 10), anxiety, Machiavellianism, age and if whether the individual is from a large city (more than 50,000 inhabitants; “1”) or not ("2"). The data file is at <http://www.uv.es/mperea/nov2019.sav>

(Solutions at the end)

1. Create the variable “total\_introversion” as the sum of the four introversion scores. Remember that third is inverse. Once this is done, are there differences in introversion between people who live in a big city and those who don't? You may want to use the box plot indicate all the statistical indexes that you consider relevant.

2. Dichotomize the variable “age” in “under 40 years” (value 1) and “40 or more” (value 2). Once this is done, are there differences in anxiety between the groups of younger and older individuals? You may want to use the box plot indicate all the statistical indexes that you consider relevant.

3. We want to detect the most anxious people, especially 5% of the most anxious people. What anxiety score will be the cutoff point?

4. In the sample, is the number of people in the big city similar to those not? You can obtain a chart. You can also obtain the frequency table.

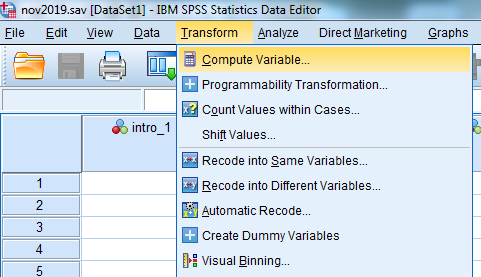
5. Does the distribution of Machiavellianism in the sample follow the normal distribution? If not, what kind of transformation would you do?

6. What T-score corresponds to the person in the first row in Anxiety? What does it mean?

7. We want to select only those individuals older than 40 years old. Can you see any differences in Machiavellianism between the people from the big city and those who don't? You may want to use the box plot indicate all the statistical indexes that you consider relevant.

S

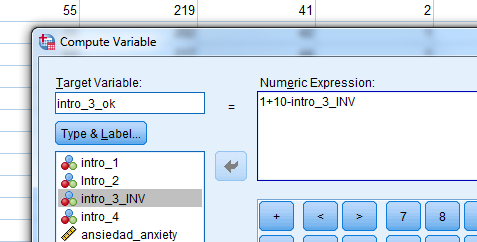
Question 1. First of all, you have to transform the third ítem of Introversion in the correct order. You just have to go to the Transform menu and then Compute Variable:



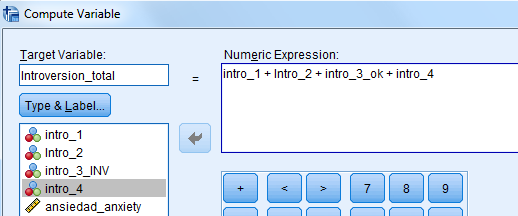
And then you just have to write down the formula (see below).

Intro\_3\_ok = 1+ 10 – Intro3INV

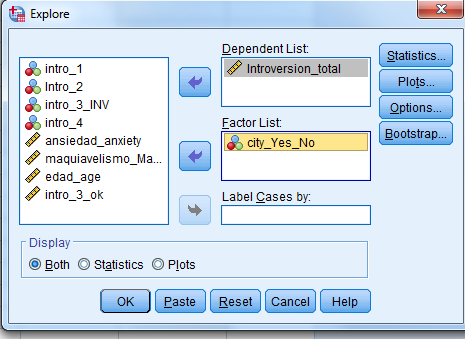
(if the scale were from 1-7 then de “10” should be a “7” etc etc)



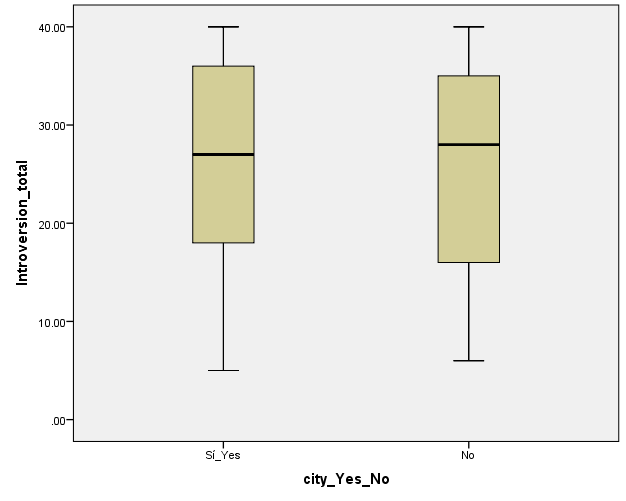
And now you just have to calculate the sum of the four items:

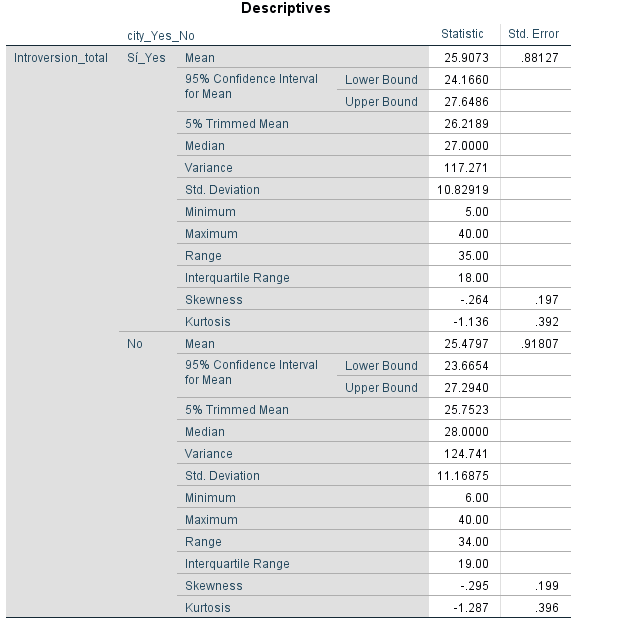


And then you just have to check whether the level of introversion differs among people living in big city and those who do not, with a box plot (Analyze / Descriptive / Explore).

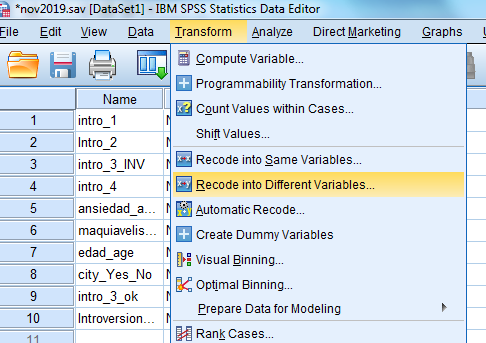


As shown in the box plot, the levels of introversion are very similar for those people who live in large cities and those who don’t. The average of introversion in both groups is very similar: 25.9 vs. 25.5 respectively. (Well, there is no reason why people who live in big city are more/less introverted than those who do not.)

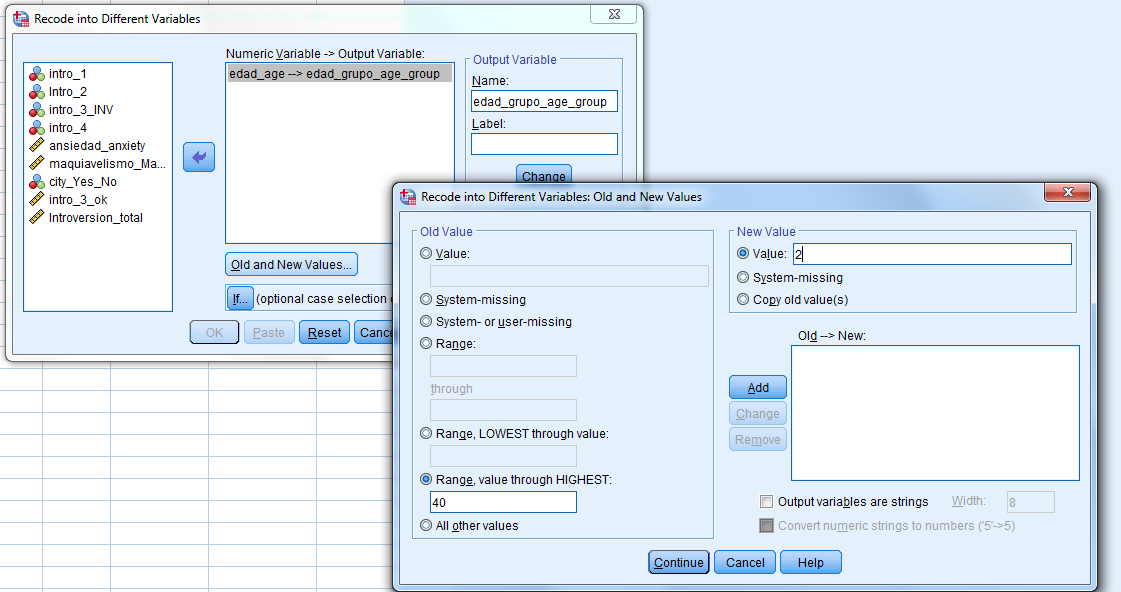


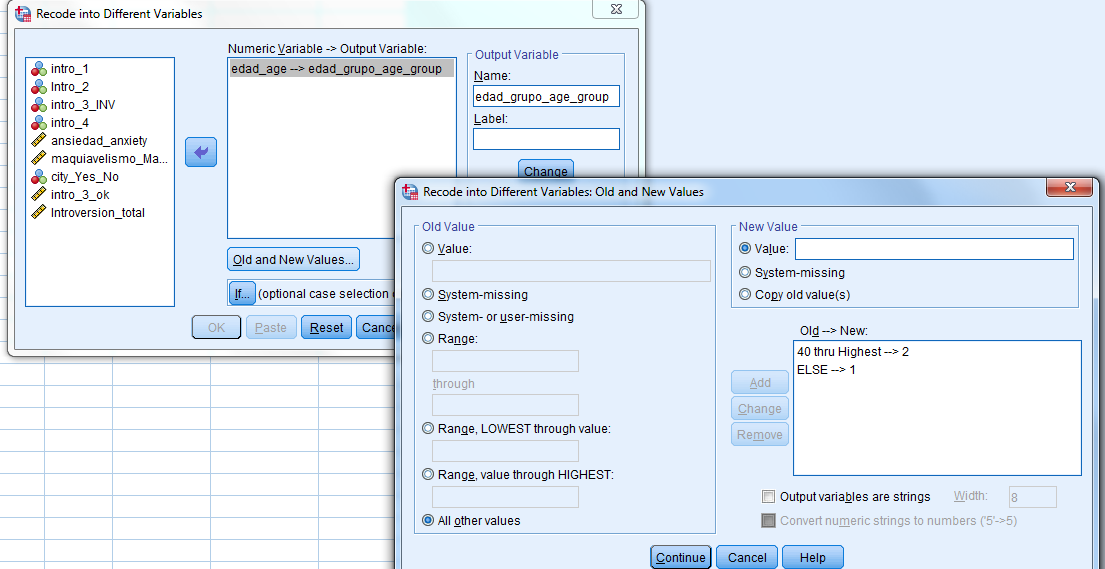


2. We must first dichotomize the variable "age" into two groups. You just have to go to the Transform menu and then to the option “Recode into Different Variables”:

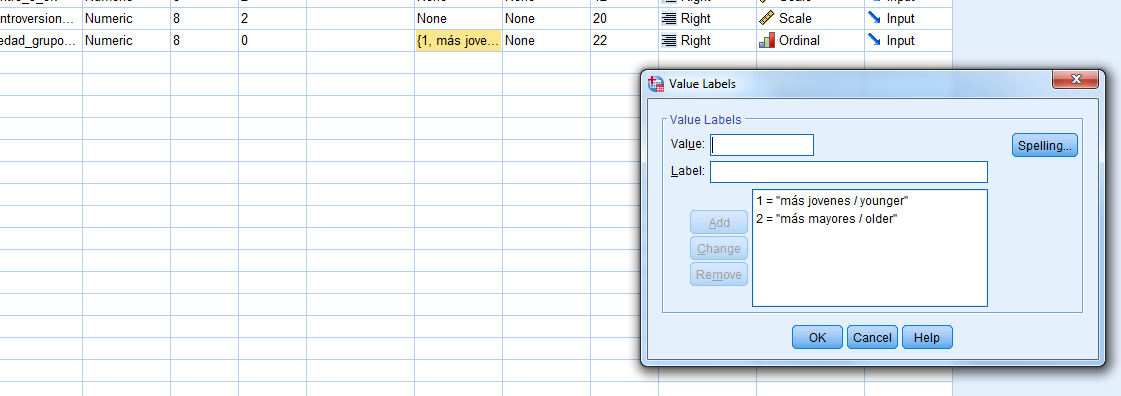


Now you have to choose a name for the new variable "edad\_groups": 40 years or more will be 2 and all others will be 1:



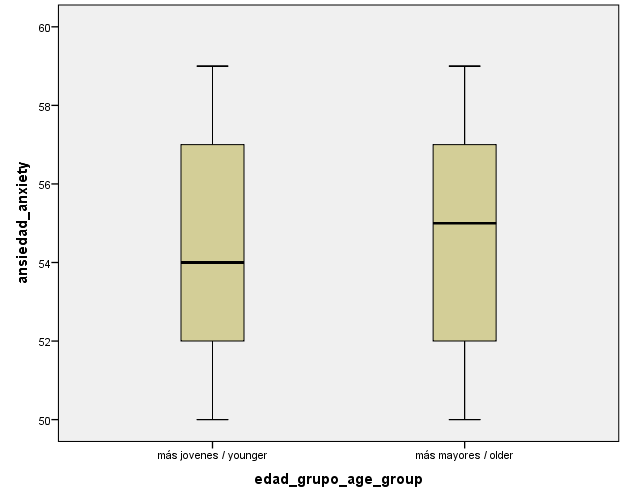


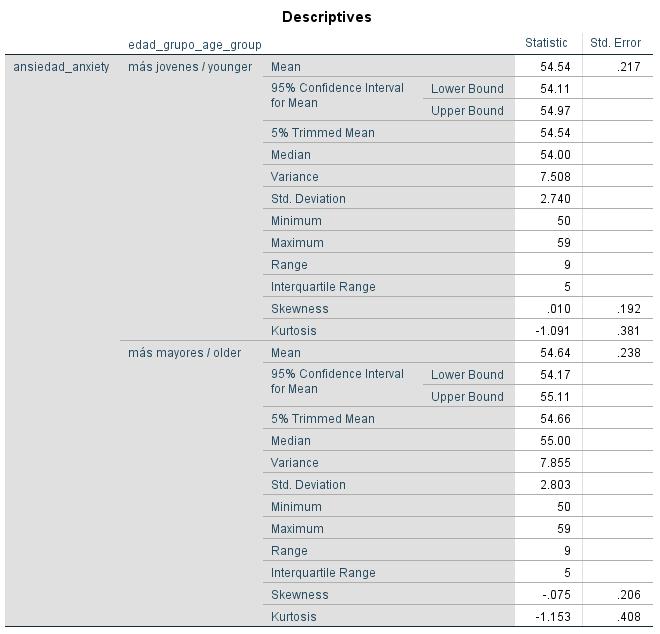
You can add in "Variable View" that 1 are “young adults” and 2 “older adults” for this new variable—this is now an "Ordinal" variable.



And we can check whether there are differences in anxiety between groups via Analyze / Descriptives/ Explore:

As can be seen in the box plot, there are no differences in anxiety between the two groups: The averages are virtually the same for younger and older adults (54.5 vs. 54.6, respectively).



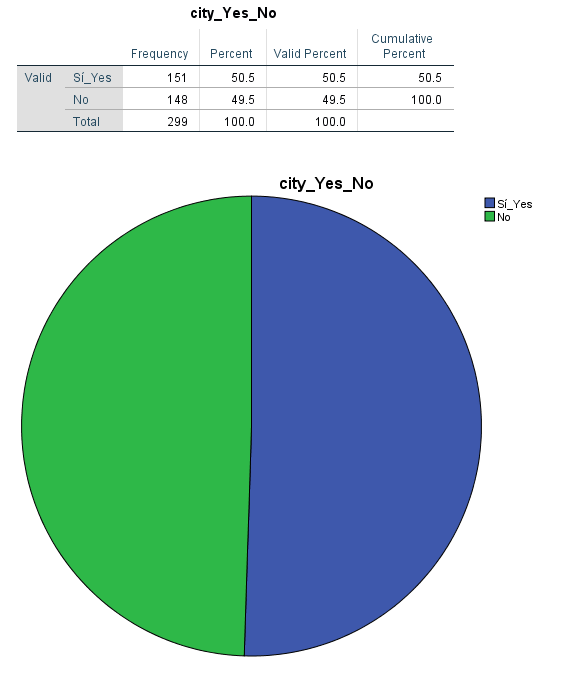


3. What we want is to know what score leaves 5% of data above it, which is parallel to say which leaves score below it the 95% of the data. That is, the cutoff point is the 95th percentile. We can obtain this value in Analyze-Descriptive-Frequencies and then click Statistics, and finally choose the 95th percentile.

The answer (using SPSS) is 59 (Percentile 95): this is the cutoff point.

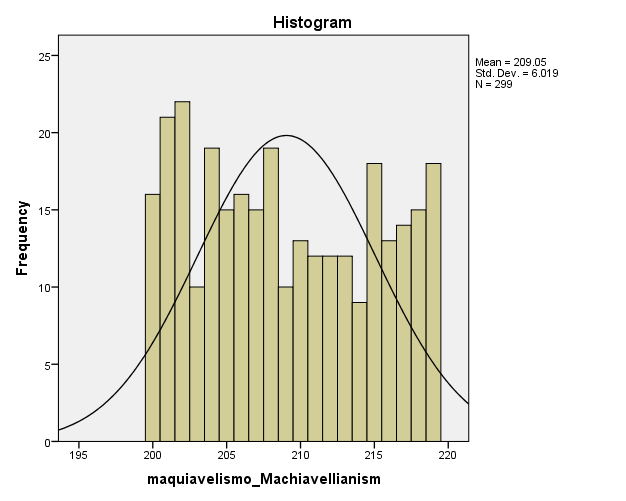
4. We asked whether, in the sample, the number of people living in a large city is similar to the number of people who don’t. This can be easily examined by looking at a pie chart or a bar chart. You can also examine this question in a frequency table. You can go to Analyze / Descriptives / Frequencies and then go to Charts. Choose one of the two graphs indicated earlier. SPSS, by default, offers the frequency table.

As can be seen in both the frequency table and the pie chart, almost half of people in the sample live in big city. So the two groups are very well balanced (i.e., it’s almost 50/50)



5. To see if the distribution of a quantitative variable follows a normal distribution, an excellent strategy is to obtain the Histogram (Analyze / Descriptives / Frequencies and choose Histogram and Show normal curve in "Charts”).

The resulting histogram is:



Clearly, the distribution of the data **is not consistent** with a normal distribution. When asked if you can do some non-linear transformation…we have only seen what to do if the distribution was skewed, this is not the case…this distribution is too far from a normal distribution…

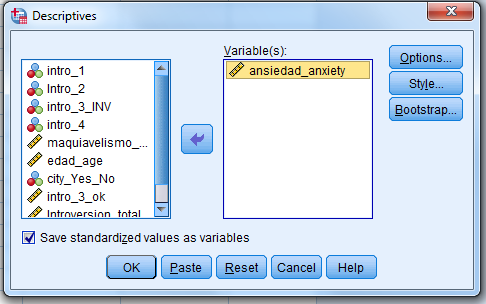
Question 6. T Scores are scales that are derived from the z-scores. Specifically:

Ti = 50 + 10 \* z*i*

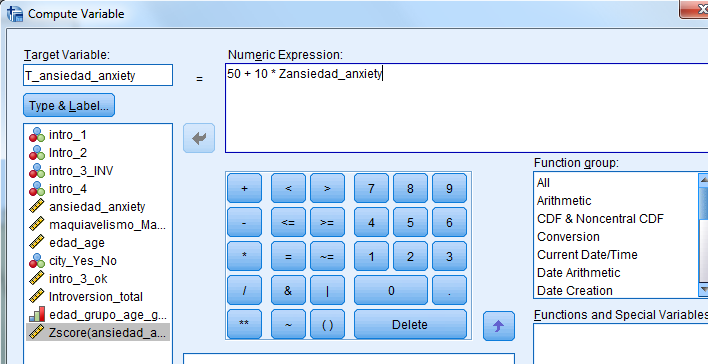
The average of the T scores is 50 and the standard deviation is 10.

An example of interpretation: A T-score of 60 would reflect that the person is one standard deviation above the average (10 points above the average, which is 50).

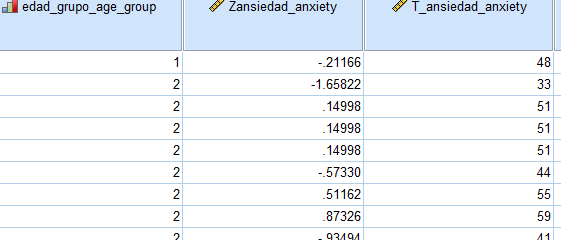
In this case we first need to get the z-scores of anxiety for all participants. Just go to Analyze / Descriptive / descriptive and click the option standardized scores/values:



And now we must apply the formula for T scores (using the Transform menu and then Compute Variable) on the z-scores:



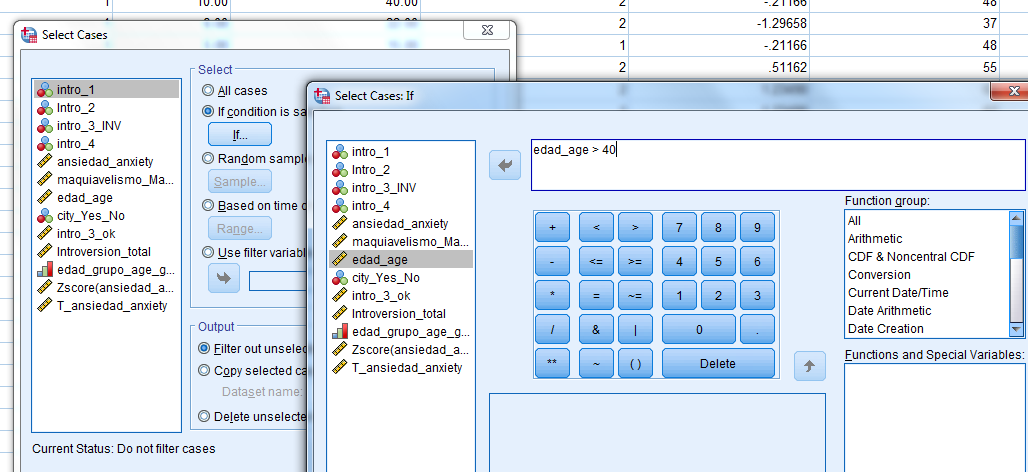
And now we can have a look at the T-score of the person in the initial row (you may want to remove decimals in the T-scores via "Variable View")



The T score of the person in the first row is 48. This is slightly below (2 units) of the mean (which is 50). Therefore, the degree of anxiety of this person is approximately average (just a bit below average).

Although this was not asked, you can check the T score of the person in row 2. It is almost two standard deviations below the mean (33 points, while the average is 50): this is a very calm person.

Question 7. You just have to select the people over 40 years in Select Cases (Data menu). And then you can get the boxplot with Machiavellianism as the dependent variable and City (large, small) as the independent variable:



As can be seen in the box plot, the levels of Machiavellianism seem to be very similar regardless of whether the individuals live in a big city or not, as deduced from the overlapping of the two boxes. There is just slightly higher Machiavellianism values ​​for the people who do not live in big city compared to the ones who live big city (averages: 209.7 vs. 208.3, respectively).

