MT9, MT10 and MT11 worksheet

Introduction

In this worksheet, we are going to give you your first opportunity to show your knowledge of applying and interpreting inferential statistics using the World Values Survey data.

Your first aim will be to wrangle and graph data for "perceptions of choice" and "life satisfaction" in the USA and Iraq. The second aim will be to test for group differences between the USA and Iraq in perceptions of choice. Finally, the third aim will be to test the simple relationship between perceptions of choice and life satisfaction in the USA.

Please write your answers and code in an R notebook and submit it as a pdf file in the Moodle submission point.

Task 1

First of all, write the code that loads the dataframe into R Studio. As always, the WVS data can be accessed here: https://github.com/thomcurran/PB130/raw/master/MT3/Workshop/WV6_Data_R_v20180912.rds. Download the file and save it to a **logical** place where it can be eaisly located. I would recommend that you save it in your PBS folder under a subfolder called Worksheets. Now load the file by clicking on the dataframe name in the files window where you saved the WVS data, change the name to WVS, and click OK. The code will appear in the console. Just copy and paste that code in your R chunk on the worksheet. For example:

```r
WVS <- readRDS("INSER FILE LOCATION HERE")

WVS$V2A <- as_character(WVS$V2A) # keep this line of code so that the country variable (V2A)
# contains the country names and not the codes
```

Task 2

Use your coding skills to load the necessary packages: "tidyverse", "mosaic", "ggplot2", "moments", "car", "supernova", "lm.beta", "ggpubr", "Hmisc", and "sjlabelled".

Task 3

Select out of the WVS dataframe the necessary variables. In this case it is the country variable (V2A), the variable for life satisfaction (V23), and the variable for perceptions of choice (V55). See the WVS codebook for the specific names and questions asked for these vairables here: https://github.com/thomcurran/PB130/blob/master/MT3/Workshop/F00007761-WV6_Codebook_v20180912.pdf

Task 4

Filter the selected variables (i.e., V23 and V55) for values of relevance (i.e., only those values that include responses to scale questions). See the WVS codebook for the specific values of the variables. Hint: typically this is values equal to or above 1.

In the same code, also filter the WVS dataframe for the countries of interest. Here, it is the "United States" and "Iraq".

Task 5

Find the mean, median, and standard deviation of life satisfaction and perceptions of choice in USA and Iraq using favstats(). Then, graph the distribution of these variables in the whole dataset using histograms.

Using the descriptive statistics and the histograms comment on; a) the mean difference between the USA and Iraq on perceptions of choice, and b) the distributional properties of the variables.

Task 6

Now I want you to build and test the linear model that examines the difference in perceptions of choice between the USA and Iraq. Our theory here is that as perceptions of choice are likely to be higher in the USA which has a less authoritarian political landscape than Iraq. Before we investigate this, though, lets just clarify the research question and null hypothesis you will be testing:

Research Question - Do percpetions of choice differ between people from from the USA and Iraq?

Null Hypothesis - The difference between perceptions of choice in the USA and Iraq will be zero.

To test this research question, there are several steps needed:

1. First, you are going to use the lm() function to build a two-parameter linear model. It is a two-parameter model becuase we have two means to estimate (i.e., USA vs Iraq). Write the code to build this linear model and save it as a new R object called "choice.model".
2. Then, request a summary of the choice.model using the summary() function. Comment on the meaning of the intercept (b0), the slope estiamte (b1), the standard error for the slope, the t-ratio for the slope, and the p value for the slope. When interpreting these estimates, remember that R will automatically code the countries 0 = Iraq and 1 = USA.
3. After the model summary, use the confint() function request the normal theory confidence intervals. Comment on the normal theory confidence interval of the slope estimate from the output.
4. Now use the Boot() function to create 1,000 resamples with replacement, estimating the linear model parameters on each occasion, and save them in a new R object called "choice.boot". Then, use the confint() function to request the bootstrap confidence intervals for the estiamates. Comment on the bootstrap confidence interval of the slope estimate from the output.
5. Finally, I want you to provide an answer to the research question - can we reject the null hypothesis that the mean difference is zero? Make sure you justify your answer using the model coeffcients and confidence intervals.

Task 7

Another, perhaps more informative, way to think about group comparisons is that we are adding an explainatory variable (country) to the empty model (i.e., intercept or mean-only model). By doing this, we are attempting to reducing the empty model error. Or, in other words, to explain or reduce the empty model variance.

We've established a mean difference and tested whether it is statistically significant using the linear model in task 6. But we didn't address how "good" this model is compared to the empty model. That is, how much variance in perceptions of choice is explained by country?

As you now know, we can answer this question using the supernova() function, which breaks down the variance in perceptions of choice due to the model (ie., country) and error (i.e., variance left over once we've subtracted out the model).

So for this task, now use the supernova() function to partition the variance in perceptions of choice between the model and the error.

Then, comment on the meaning of the F ratio and RPE (R-sqaure) from the output.

Like task 6, I want you to use the supernova() output to provide an answer to the research question - can we reject the null hypothesis that the mean difference is zero?

Task 8

Now I want you to build a linear model with the two continuous variables (i.e., a linear regression model). Here, we are we are going to examine the relationship between perceptions of choice and life satisfaction using data only from the USA. Our theory here is that perceptions of choice would share a positive relationships with life satisfaction. Given this expectation, we also expect that we can use a best fitting linear model to make relaiable estimations of life satisfaction from percieved choice.

Before we delve into this topic, lets just clarify the research question and null hypothesis we are testing:

Research Question - Is there a linear relationship between perceptions of choice and life satisfaction in the USA?

Null Hypothesis - The realtionship between perceptions of choice and life satisfaction will be zero.

There are several steps needed to conduct this analysis, so follow them closely:

1. You will need to filter the WVS to only include data from the USA, so execute the code for this first.
2. Use the rcorr() function to generate the correlation coefficient matrix. Comment on the correlation coeffcient and its associated p value.
3. Then, you are going to use the lm() function to build a two-parameter linear regression model. It is a two-parameter model becuase we an intercept (b0) and slope (b1) for the explanatory variable. Write the code to build this linear model and save it as a new R object called "usa.model".
4. Then, request a summary of the usa.model using the summary() function. Comment on the meaning of the intercept (b0), the slope estiamte (b1), the standard error for the slope, the t-ratio for the slope, and the p value for the slope.
5. After the model summary, use the confint() function request the normal theory confidence intervals. Comment on the normal theory confidence interval of the slope estimate from the output.
6. Now use the Boot() function to create 1000 resamples with replacement, estimating the linear model parameters on each occasion, and save them in a new R object called "usa.boot". Then, use the confint() function to request the bootstrap confidence intervals for the estimates. Comment on the bootstrap confidence interval of the slope estimate from the output.
7. Finally, I want you to provide an answer to the research question - can we reject the null hypothesis that the relationship is zero? Make sure you justify your answer using the model coeffcients and confidence intervals.

Task 9

Now we've tested the relationship and whether it is statistically significant using the linear model, let's address how "good" this model is compared to the empty model.

As you now know, we can answer this question using the supernova() function, which breaks down the variance in life satisfaction due to the model and error.

For this task, use the supernova() function to partition the variance in life satisfaction between the model and the error.

Then, comment on the meaning of the F ratio and RPE (R-sqaure) from the output.

Finally, use the supernova() output to provide an answer to the research question - can we reject the null hypothesis that the relationship is zero?