Introduction to statistical inference – Project work – Generation 2022/23

Name and Surname: _____ Student ID number: _____

Professor:_____ Grade: _____

Using the dataset *GenderEqualityIndex.sav* conduct the analysis, answer the questions and tasks.
The analysis can be done using SPSS or R. If the analysis is done in R, it is desirable to print the code. If the analysis is done in SPSS, there is no need to print the output. **Note: Depending on the software used and software version, the results among students might differ.**

The *GenderEqualityIndex.sav* is a dataset that provides the data on the European Gender Equality Index – a composite indicator which aims to measure the level of (in)equality among males and females in the European Union. Besides the name of the country, the overall value of the EGEI, you are provided with indicator data. Based on the pre-calculated values per pillar or dimension (Work, Money, Knowledge, Time, Power, and Health), the countries have been classified in the following manner:

Pillar or dimension	Code	Value
Work, Money, Knowledge, Time, Power, and Health	1	High
	2	Medium
	3	Low

1. Obtain the parameters of descriptive statistics for the indicators *Participation* and *Segregation and quality of work*:

Indicator	Mean	SD	Median	Minimum	Maximum
Participation					
Segregation and quality of work					

- Indicator with the higher mean is _____
- Indicator with the smaller variance is _____
- Is the median of *Participation* greater than the median of *Segregation and quality of work?* Yes No
- Range of *Segregation and quality of work* is _____
- Interquartil range of *Participation* is _____

2. How many countries belong to each group based on the calculated pillar *Time*?

Low	Medium	High

3. Is there statistically significant difference between the country groups based on *Power* for the indicator *Financial resources*?

The first step in the analysis is to conduct the _____ test.

H_0 _____

H_1 _____

Value of the statistics_____ *p* value _____

We conclude_____

To test the differences we use _____

H_0 _____

H_1 _____

Value of the statistics_____ *p* value _____

We REJECT ACCEPT (we colud not reject) the H_0

Conclusion

_____

_____

What would be your next steps if the Ho is rejected? Just explain the procedure without conducting it.

_____

_____

4. Is there difference in the values of the indicator *Not at-risk-of-poverty (%)* between countries which have medium and high *Work*?

The first step in the analysis is to conduct the _____ test.

H_0 _____

H_1 _____

Value of the statistics_____ *p* value _____

We conclude_____

To inspect whether there are differences we use _____

H_0 _____

H_1 _____

Value of the statistics_____ *p* value _____

We REJECT ACCEPT (we colud not reject) the H_0

Conclusion

_____

_____

5. Is there statistically significant correlation between the indicators *Economic situation* and *Segregation*.

The first step in the analysis is to conduct the _____ test.

H_0 _____

H_1 _____

Value of the statistics (*Economic situation*)_____ *p* value _____

Value of the statistics (*Segregation*)_____ *p* value _____

We conclude_____

To inspect whether there is significant correlation we use _____

H_0 _____

H_1 _____

Value of the statistics_____ The obtained correlation coefficient is _____ *p* value _____

We REJECT ACCEPT (we colud not reject) the H_0

Conclusion _____

_____

6. Does the *Power* level impact the *Money* level? How many countries are there with low *Power* and medium *Money* level?

H_0 _____

H_1 _____

We use the _____ test

Value of the statistics_____ *p* value _____

We REJECT ACCEPT (we colud not reject) the H_0

Conclusion _____

_____

There are _____ countries with low *Power* and medium *Money* level.

7. Is there statistically significant difference between countries with low and medium *Knowledge* regarding the values of the indicator *Attainment and participation?*.

The first step in the analysis is to conduct the _____ test.

H_0 _____

H_1 _____

Value of the statistics_____ *p* value _____

We conclude_____

Therefore, we will use the _____

As a pre-test, we conduct the _____ test.

H_0 _____

H_1 _____

Value of the statistics_____ *p* value _____

We REJECT ACCEPT (we colud not reject) the H_0 of the pre-test

We conduct the test

H_0 _____

H_1 _____

Value of the statistics_____ *p* value _____

We REJECT ACCEPT (we colud not reject) the H_0

Conclusion

_____

_____

Mean values of the indicator *Attainment and participation* per group are:

Low *Knowledge* _____ Medium *Knowledge* _____

8. Is the mean value of the indicator *Care activities* on the population 75?

The first step in the analysis is to conduct the _____ test.

H_0 _____

H_1 _____

Value of the statistics_____ *p* value _____

We conclude_____

In the next step we will use _____ test

H_0 _____

H_1 _____

Value of the statistics_____ *p* value _____

We REJECT ACCEPT (we colud not reject) the H_0

Conclsuion _____

_____

_____

9. Create a linear regression model and model how *Graduates of tertiary education* (%) impact *Attainment and participation*. Answer the following questions:

A) What is the obtained model equation?

_____

B) Is the coefficient related to the impact of *Graduates of tertiary education* (%) statistically significant?

H_0 _____

H_1 _____

Value of the statistics_____ *p* value _____

We conclude_____

_____

C) Is the overall model statistically significant?

H_0 _____

H_1 _____

Value of the statistics_____ *p* value _____

We conclude_____

_____

D) Comment on the model quality

R^2= _____ The R^2 indicates that _____

E) Can the created model be used for predictions?

Answer _____

10. Is there statistically significant difference between the *Health* groups if we observe the values of the indicator *Social activities*.

The first step in the analysis is to conduct the _____ test.

H_0 _____

H_1 _____

Value of the statistics_____ *p* value _____

We conclude_____

To test the differences we use _____

H_0 _____

H_1 _____

Value of the statistics_____ *p* value _____

We REJECT ACCEPT (we colud not reject) the H_0

Conclusion _____

_____

If the Ho is rejected, how would you proceed with the analysis? You do not have to conduct the analysis.

_____

_____

_____

11. Create a linear regression model and model how *Social, Access,* and *Political* impact *Economic*. Answer the following questions:

A) What is the obtained model equation?

_____

B) Is the coefficient related to the impact of *Social* statistically significant?

H_0 _____

H_1 _____

Value of the statistics_____ *p* value _____

We conclude_____

C) Is the coefficient related to the impact of *Access* statistically significant?

H_0 _____

H_1 _____

Value of the statistics_____ *p* value _____

We conclude_____

D) Is the coefficient related to the impact of *Political* statistically significant?

H_0 _____

H_1 _____

Value of the statistics_____ *p* value _____

We conclude_____

E) Is the overall model statistically significant?

H_0 _____

H_1 _____

Value of the statistics_____ *p* value _____

We conclude_____

_____

F) Comment on the model quality

Adjusted R^2= _____ The R^2 indicates that _____

_____

G) Is there a problem of multicollinearity in the model?

Answer _____

H) Is there a problem of autocorrelation in the model?

Value of the DW test _____

Knowing that dl=0.969 du=1.414 draw the Durbin Watson critical range and make the conclusion:

_____

Durbin Watson critical range