Technology Arts Sciences Cologne
Faculty of Economics, Business and Law
Prof. Dr. Arrenberg
Room 221, Tel. 39 14
jutta.arrenberg@th-koeln.de

Exercises Quantitative Methods in Statistics

Worksheet: Mock Exam

Problem:
Please open the file *1991 U.S. General Society Survey.sav* of the SPSS-Tutorial. This is a poll of 1 517 US-Americans of age 18 up to 89.

a) Use the sample data to test for the independence of the two variables "ethgr = Ethnic Group of Respondent" and "Happy = General Happiness".

   1. What is the name of the test?
   2. Which assumptions of the test must be checked to get an accurate result? Please verify these assumptions.
   3. Consider the p-value. What is your conclusion?

b) Does there appear to be a relationship between General Happiness and Number of Children? What does the result of your analysis tell the purchasing department?

c) Use $\alpha = 0.05$ and a goodness of fit test to see whether the data of the variable Number of Children fit a Normal distribution.

Solution

a. 1. Pearson Chi-Square test

2. df=4
   0 cells $\leq 20\%$ have expected count less than five
   minimum expected count $= 5.16 \geq 1$
   The rule of thumb is fulfilled.

3. p-value $\approx 0.000 \leq 0.05$
   General Happiness and Ethnic Group are stochastically dependent.

b) p-value of Pearson Chi-Square test $= 0.044 < 0.05$
General Happiness and Number of Children are stochastically dependent at a level $\alpha = 0.05$-test
$\gamma = +0.001$ i.e. very weak relationship that respondents with many children are dissatisfied.

c) p-Value Lilliefors test ≈ 0.000
p-value Shapiro-Wilk test ≈ 0.000
The variable Number of Children has no Normal distribution.
Remark: These two goodness-of-fit test are only sensitive for small sample sizes up to 30. In this example with $n = 1509$ you should check Normal distribution with a Plot-Point diagram or a Q-Q Plot.

Technology Arts Sciences Cologne
Faculty of Economics, Business and Law
Prof. Dr. Arrenberg
Room 221, Tel. 39 14
jutta.arrenberg@th-koeln.de

Exercises Quantitative Methods

Mock Exam

Problem

Please open the file *1991 U.S. General Society Survey.sav*.

a) Run a hierarchical cluster analysis with the variable "Years of education of the respondent " = "Letztes abgeschlossenes Schuljahr des Befragten". Please complete the following table:

Stage	Coefficient
⋮	
1 491	
1 492	
⋮	
1 508	
1 509	

How many clusters are recommended?

b) Run a K-Means cluster analysis with the following four variables:

- Years of education of the respondent = Letztes abgeschlossenes Schuljahr des Befragten
- Years of education of the father of the respondent = Letztes abgeschlossenes Schuljahr des Vaters
- Years of education of the mother of the respondent = Letztes abgeschlossenes Schuljahr der Mutter
- Years of education of the partner of the respondent = Letztes abgeschlossenes Schuljahr des Partners

where three clusters are generated.

1. Please show the average value of the four considered variables among the three clusters:

Average Value

	Cluster		
	1	2	3
Letztes abgeschlossenes Schuljahr			
Letztes abgeschlossenes Schuljahr des Vaters			
Letztes abgeschlossenes Schuljahr der Mutter			
Letztes abgeschlossenes Schuljahr des Partners			

2. Please comment the three clusters.

c) Please check with a level 0.05 test, whether the medians of the variable "General Happiness" = "Allgemeine Zufriedenheit" are the same in each cluster.

1. Name of the test?
2. p-value of the test?
3. What are the assumptions of the test?
4. Comment the decision of the test.
5. Calculate the empirical medians and the average values of the variable "General Happiness" = "Allgemeine Zufriedenheit" in each cluster. What is the cluster with the highest satisfaction?

Solution:

a) Hierarchical cluster analysis

Stage	Coefficient
⋮	
1 491	0
1 492	1
⋮	
1 508	1
1 509	3

$n - 1\,508 = 1\,510 - 1\,508 = 2$ clusters are recommended.

b) Final cluster centers of the K-Means cluster analysis:

Average Value

	Cluster		
	1	2	3
Letztes abgeschlossenes Schuljahr	16	11	13
Letztes abgeschlossenes Schuljahr des Vaters	14	6	10
Letztes abgeschlossenes Schuljahr der Mutter	16	11	13
Letztes abgeschlossenes Schuljahr des Partners	13	6	11

Cluster 1: longest years of education of the respondent, his/her father, mother and partner

Cluster 2: shortest years of education of the respondent, his/her father, mother and partner

Cluster 3: medium years of education of the respondent, his/her father, mother and partner

c)
1. Kruskal-Wallis test
2. The variable General Happiness is ordinal leveled. The three variables "General Happiness in cluster 1", "General Happiness in cluster 2" and "General Happiness in cluster 3" must be stochastically independent.
3. p-value = 0.046; i.e. rejection of H_0
4. At least two medians of general happiness differ significantly across the three clusters.
5. Empirical median and avergae value:

Cluster	General Happiness	
	Median	Average Value
1	2	1.58
2	2	1.64
3	2	1.73

In cluster 1 the satisfaction is highest.