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

Master: Quantitative Methods

Old Exams

Problem (25.01.2017)
Please open the file *telco_extra.sav*.

a) Please run a hierarchical cluster analysis with the five variables Month with service, Age in years, Years at current address, Household income in thousands, Number of people in household. How many clusters should be constructed? (Give reasons!)

b) Please run a k-means cluster analysis with the five variables Month with service, Age in years, Years at current address, Household income in thousands, Number of people in household. As the number of clusters please select the value three.

1. How many cases are in cluster 1 resp. 2 resp. 3?

	Cluster		
	1	2	3
Cases			

2. Please complete the following table:

Average Values

	Cluster		
	1	2	3
Month with service			
Age in years			
Years at current address			
Household income in thousands US-Dollar			
Number of people in household			

3. Please comment the three clusters.
4. Check with a statistical test whether the medians of the variable "Years with current employer" are the same across the three clusters. What is the name of the test? What are the assumptions of the test? What is the p-value of the test? What indicates this p-value?

5. What are the three values of the empirical medians of the variable "Years with current employer" in the clusters?

Empirical Medians

	Cluster		
	1	2	3
Years with current employer			

c) Suppose you will run an ordinal regression for the dependent variable $Y =$ "Calling card last month". What is the link function if the variable Y is transformed into an ordinal leveled variable with the three categories "0 up to 10", "more than 10 up to 20" and "more than 20"? (Give reasons!)

Telco_extra.sav **Exam QM 25/01/2017**

a) Hierarchical cluster analysis

Greatest jump oft he coefficients

Stage	Coefficient
998	197.601
999	537.970

Number of cluster=n-998=1000-998=2 cluster

b) K-means cluster analysis

1.

Number of Cases in each Cluster

Cluster	1	894,000
	2	6,000
	3	100,000
Valid		1000,000
Missing		,000

2.

Final Cluster Centers

	Cluster		
	1	2	3
Months with service	34	59	49
Age in years	40	60	53
Years at current address	11	29	17
Household income in thousands	51,70	1012,83	252,37
Number of people in household	2	1	2

3. Cluster 1: youngest average age, shortest time of service and at current address, lowest income, average 2 people in household
   Cluster 2: oldest average age, longest time of service and at current address, highest income, single household
   Cluster 3: median average age, median time of service and at current address, median income, average 2 people in household

4. Kruskal-Wallis test
   Stochastic independence of „Years with current employer“ across the three cluster
   p-value =0.000
   Rejection of H0, at least two medians of „Years with current employer“ differ significantly across the three cluster

5.

Report

Median

Cluster Number of Case	Years with current employer
1	7,00
2	32,00
3	26,00
Total	8,00

c)

Class	Cases
≤10	434
10 - ≤20	308=742-434
>20	258

Link function: negative log log