

Lösungen: Statistik in Netzwerken

Hinweis: Es wurden für die Berechnungen auch Excel-Dateien zur Verfügung gestellt, wo die Berechnung nachvollziehbar ist.

1.) Pearson Korrelation (Linearer Zusammenhang)

Sie lesen von der Behauptung, dass Geld ein wichtiger Faktor für einen hohen sozialen Status ist. Sie möchten dies mit Hilfe von Statistik und dem Wissen aus Sozialer Netzwerkanalyse überprüfen. Im Modul Soziale Netzwerkanalyse wurde erwähnt, dass das Ansehen (Prestige) zu den Social Forces zählt, wieso man sich mit jemanden verbindet. Deshalb haben Sie sich entschieden, Freunde nach ihrem Vermögen zu fragen. Gehen Sie davon aus, dass ihre Freunde ehrlich zu ihnen sind.

Datensatz	Vermögen (in CHF 1'000)	Anzahl Freunde auf Facebook
1	50	100
2	54	120
3	30	210
4	66	90
5	41	120
6	11	10
7	20	55
8	100	200

a.) Berechnen Sie den Pearson Korrelationskoeffizienten für den Datensatz, einmal mit dem Ausreisser und einmal ohne den Ausreisser. Nehmen Sie am besten Excel zur Hilfe, wie im Script angegeben.

Pearson Korrelations-Koeffizient mit Ausreisser: **0.58**

Pearson Korrelations-Koeffizient ohne Ausreisser: **0.91**

Hinweis: Excel und LibreOffice haben bereits eine Funktion PEARSON, welche den Pearson Korrelationskoeffizienten automatisch berechnen.

c.) Sie haben einen starken linearen Zusammenhang zwischen Vermögen und den Anzahl Freunden auf Facebook gefunden. Ist diese allgemein gültig? Begründen Sie.

Nein, ist sie nicht. Ihre Beispieldaten sind nicht repräsentativ, da Sie die Daten nur in ihrem Freundeskreis erhoben haben (Homogene Gruppe). Deshalb spricht die gefundene Erkenntnis nicht für die gesamte Bevölkerung, sondern nur für ihren Freundeskreis.

2.) Spearman Korrelation (Rangkorrelation)

Gegeben ist ein Datensatz¹, wo der IQ und die Anzahl Stunden, die eine Person vor dem TV verbringt, gegenüber stellt:

Datensatz	IQ	Stunden vor dem TV(pro Woche)
1	106	7
2	86	0
3	100	27
4	101	50

¹ Quelle: https://en.wikipedia.org/wiki/Spearman%27s_rank_correlation_coefficient, 15.09.2019

5	99	28
6	103	29
7	97	20
8	113	12
9	112	6
10	110	17

Berechnen Sie den Rangkorrelations-Wert mit der im Script angegebenen Formel:

$r_s = -0.18$

3.) Definition von Hypothesentests

Gegeben sei ein Co-Authorship Netzwerk, bei welchem abgebildet ist, welcher Wissenschaftler zusammen mit welchen anderen Personen Publikationen geschrieben hat.

Es handelt sich dabei um ein One-Mode Netzwerk mit den Attributen:

Knotenattribute:	Kantenattribute
Name des Autors	Anzahl gemeinsame Publikationen
Universität	
Anzahl Auszeichnungen	
Durchschnittliche Anzahl Veröffentlichungen / Jahr	

Zudem haben sie ein zweites Netzwerk, in welchem dieselben Autoren die Knoten darstellen und die Kanten die privaten Freundschaftsbeziehungen.

Definieren Sie für den gegebenen Anwendungsfall drei Hypothesentests:

- Einen Monadic Hypothesentest: (Ein Attribut soll dabei eine Netzwerk-Metrik sein)
- Einen Dyadic Hypothesentest
- Einen Mixed Monadic-Dyadic Hypothesentest. (Ein Knoten-Attribut soll dabei die Basis für ein neues Netzwerk sein.)

Monadic Hypothesentest: Umso mehr Publikationen mit anderen Autoren ((weighted Degree), desto mehr Auszeichnungen (Knotenattribut).

Dyadic Hypothesentest: Autoren, die freundschaftliche Beziehungen pflegen, veröffentlichen häufiger miteinander.

Mixed Monadic-Dyadic: Autoren an gleichen Universitäten veröffentlichen häufiger miteinander. Dazu wird ein neues Netzwerk gemacht, in welchem diejenigen Autoren miteinander verbunden werden, die an der gleichen Universität arbeiten.

4.) Permutations-Tests & QAP

Schauen Sie im Ordner der Notebooks für diesen Themenblock nach. Da befindet sich ein Unterordner mit dem Namen «Solutions».