

Deskriptive Statistik für Soziologinnen und Soziologen (Mariana Nold)

Thema: Bivariate Exploration von quantitativen und qualitativen Merkmalen: Korrelation

Abgabe: bis Freitag, den 2. Juni 2017, Briefkasten des Instituts für Soziologie, in der Nähe der Cafeteria der Carls-Zeiss-Straße (In einem Umschlag, an mich adressiert) **oder** Mittwoch, Donnerstag und Freitag von 13-15 Uhr im Sekretariat von Frau Prof. Leuze, CZ-Straße 2, R286

15. **Wortschatz von Kindern** (20 Punkte)

Aufgabe
abgeben!

(in Anlehnung an: Fahrmeir et al, Statistik Der Weg zur Datenanalyse, S .151)

Bei fünf zufällig ausgewählten Kindern wurden der Wortschatz X und die Körpergröße Y in cm gemessen. Dabei erfolgte die Messung des Wortschatzes über die Anzahl der verschiedenen Wörter, die die Kinder in einem Aufsatz über die Ergebnisse in ihren Sommerferien benutzten. Nehmen wir an, wir hätten folgende Daten erhalten:

Kind i	1	2	3	4	5
Körpergröße x_i	37	30	20	28	35
Wortschatz y_i	130	112	108	114	136

Tabelle 1: Der Wortschatz X und die Körpergröße Y in cm gemessen von 5 zufällig ausgewählten Kindern.

- (a) Zeichnen Sie ein Streudiagramm. (2 Punkte)
- (b) Erklären Sie an Hand dieses Beispiels was eine (4 Punkte)
 - positive bzw. negative lineare Korrelation
 - positive bzw. negative monotone Korrelationinhaltlich bedeuten.
- (c) Ist die folgende Aussage falsch oder richtig: Es ist im Allgemeinen möglich, dass ein positiver linearer Zusammenhang vorliegt, aber kein positiver monotoner Zusammenhang. (2 Punkte)
- (d) Schreiben Sie die Tabelle der Ränge von X und Y , berechnen die den Rangkorrelationskoeffizient r_{XY}^{SP} und interpretieren Sie diesen Wert. (4 Punkte)
- (e) Berechnen Sie nun den Korrelationskoeffizienten nach Pearson r_{XY} und interpretieren Sie diesen Wert. (4 Punkte)
- (f) Die Tabelle ?? enthält zusätzlich das Alter Z der Kinder, berechnen Sie jeweils r_{YZ} und r_{XZ} und interpretieren Sie auch diese Werte. (4 Punkte)

16. **Kreuztabellen interpretieren: Habilitationsdichte** (11 Punkte)

Aufgabe
abgeben!

Die Habilitation ist die höchstrangige Hochschulprüfung in Deutschland durch Anfertigung einer wissenschaftlichen Arbeit. In einer Untersuchung zur Habilitationsdichte

Kind i	1	2	3	4	5
Körpergröße x_i	37	30	20	28	35
Wortschatz y_i	130	112	108	114	136
Alter z_i	12	7	6	7	13

Tabelle 2: Der Wortschatz X , die Körpergröße Y in cm gemessen und das Alter Z von 5 zufällig ausgewählten Kindern.

an deutschen Hochschulen wurden u. a. die Merkmale Geschlecht und Habilitationsfach erhoben. In Tabelle ?? ist - nach Fächern aufgeschlüsselt- zusammengefasst, wieviele Habilitationen im Jahre 2015 erfolgreich abgeschlossen wurden (Quelle: Statistisches Bundesamt) Hier stellt sich die Frage, ob die Habilitationsdichte in den einzelnen Fächern im Jahr 2015 geschlechtsspezifisch ist, d. h. man interessiert sich dafür, ob zwischen den Merkmalen Geschlecht ($=:Y$) und Habilitationsfach ($=:X$) ein Zusammenhang besteht.

$$\text{Ausprägungen } X \stackrel{\wedge}{=} \begin{cases} a_1, & \text{Geisteswissenschaften} \\ a_2, & \text{Rechts-,Wirtschafts-,Sozialwiss.} \\ a_3, & \text{Mathe u. Naturwiss.} \\ a_4, & \text{Human-,Gesundheitswiss.} \\ a_5, & \text{übrige Fächer} \end{cases}$$

$$\text{Ausprägungen } Y \stackrel{\wedge}{=} \begin{cases} b_1, & \text{Frauen} \\ b_2, & \text{Männer} \end{cases}$$

		X					Σ
		a_1	a_2	a_3	a_4	a_5	
Y	b_1	77	62	66	225	32	
	b_2	159	139	181	571	115	
Σ							

Tabelle 3: Habilitationen im Jahre 2015 erfolgreich abgeschlossen wurden nach Fächern und Geschlecht aufgeschlüsselt mit Randhäufigkeiten.

- Ergänzen sie die fehlenden Randhäufigkeiten. (1 Punkt)
- Berechnen Sie die Randverteilungen. (2 Punkte)
- Wie hoch ist der Anteil der Frauen, die im Jahr 2015 ein Habilitation abgeschlossen haben. (1 Punkt)
- Wie hoch ist der Anteil an Habilitationen aus dem Fachbereich „Mathematik und Naturwissenschaften“? (1 Punkt)
- Berechnen Sie die bedingten Verteilungen gegeben dem Fachbereich und interpretieren Sie das Ergebnis. (3 Punkte)
- Berechnen Sie die bedingten Verteilungen gegeben das Geschlecht und interpretieren Sie das Ergebnis. (3 Punkte)

17. Die Korrelation mit STATA berechnen

Das Streudiagramm ?? zeigt die Spielbewertung eines neuen Spiels aufgetragen auf der Ordinate und die Mathe-Punkte (hier simuliert, nicht aus den PISA-Daten) von 300 Schülerinnen bzw. Schülern. Sie finden den entsprechenden Datensatz auf den Rechnern im Methoden-Labor im Ordner **Methoden/Statistik**.

- (a) Öffnen Sie den Datensatz und geben Sie den Befehl **summarize** in das **command-Fenster** ein. Interpretieren Sie die von **STATA** erzeugte Tabelle.
- (b) Geben Sie die Befehle **graph box x** und **graph box y** ein und interpretieren Sie die entsprechenden Boxplots.
- (c) Erzeugen Sie mit dem Befehl **scatter y x** das Streudiagramm.
- (d) Berechnen Sie mit **pwcorr x y** den Korrelationskoeffizient nach Pearson. Wie ändert sich das Ergebnis, wenn Sie den Befehl **pwcorr y x** eingeben. Interpretieren Sie diese Veränderung inhaltlich.
- (e) Berechnen Sie mit Hilfe des Befehls **spearman x y** den Wert des Rangkorrelationskoeffizienten.