20251015_ADS_AKIB3

# Beziehungen zw. den Daten.

A1. $z_x = \frac{x-\bar{x}}{\sigma_x}$ $\quad z_y = \frac{y-\bar{y}}{\sigma_y}$ Standardisierung

A2. Stichprobenkorrelation: $r_{xy} = \frac{1}{n-1} \cdot \sum_{i=1}^{n} \frac{(x_i-\bar{x})(y_i-\bar{y})}{\sigma_x \sigma_y}$

A3. Korrelationsmatrix (heatmap). s. 2. Semester.

A4. Hypothesentest & Konfidenzintervall von Pearson.

Test $H_0$ : $\rho = 0$ ✓

· Fisher-z-Transformation (Konfidenzintervall für $\rho$)

95% KI $\longrightarrow$ z $[-\sigma, \sigma]$

A5. Rangbasierte Maße: Spearman & Kendall.

· Spearman: $\rho_S$: Pearson-Korrelation der Ränge

$$\rho_S = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$$ $d_i$: Rangdifferenz

· Kendall: $\tau = \frac{C-D}{\binom{n}{2}}$ C: Anzahl konkordante
D: " diskordante

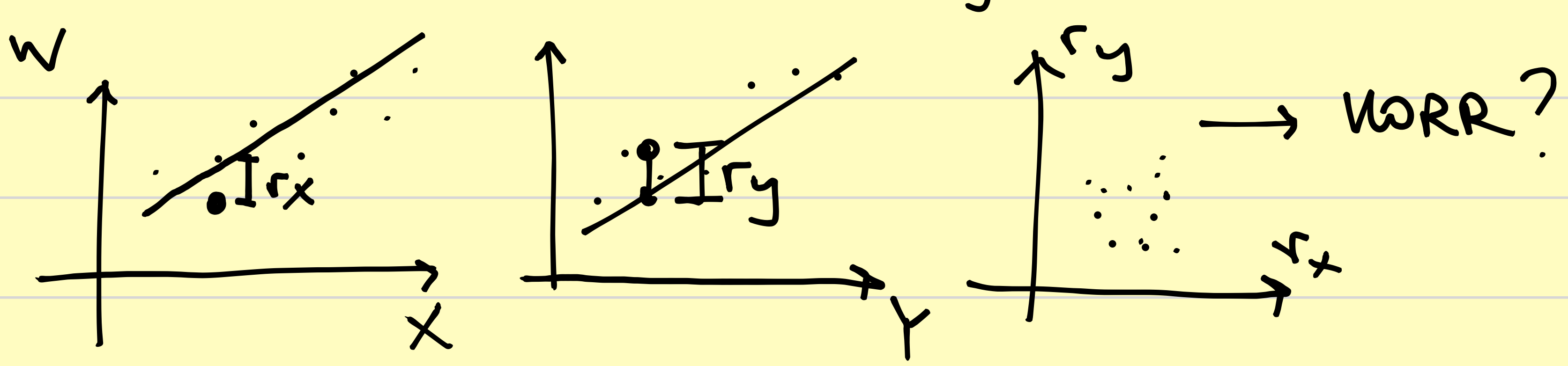
## A6. Partielle Korrelationen.

Ziel. Zusammenhang zw. X und Y gegeben weiterer Variablen W (Vektor).

(i) Residual-Methode: Regression X auf W →
→ ermittle ich Residuen $r_x$

Regression Y auf W → Residuen $r_y$

Korrelation zw. $r_x$ und $r_y$




## A8. Simpson-Paradoxon (Aggregation kann Trend umkehren).

Zwei Gruppen $G \in \{A, B\}$.

Innerhalb der Gruppen A, B gilt Y steigt mit X.

Wenn sich die Gruppen stark in Mittelwerten $E[X], E[Y]$ und Anteilen unterscheiden, kann die aggregierte Steigung negativ werden.

A9. Nicht lineare Abhängigkeiten.

- Pearson misst lineare Abhängigkeiten.
- Mutual Information:

$$I(X,Y) = E\left[\log \frac{p_{xy}}{p_x p_y}\right] \geq 0$$

$= 0 \leftrightarrow$ Unabhängig