

Skalentransformationen & Korrelationen

Skalenniveaus

	Skalenarten	Transformationen	Aussagen	Beispiele
1. Nominalskala	1.Nominalskala	Eindeutigkeits-Transformationen	Gleichheit, Verschiedenheit	Automarken, Krankheitsklassifikation, Familienstand
2. Ordinalskala	2.Ordinalskala	Monotone-Transformationen	Größer-Kleiner-Relationen	Militärische Ränge, Windstärken
3. Kardinalskala	3.Intervallskala	Lineare-Transformationen	Gleichheit von Differenzen	Temperatur, Kalenderzeit, Intensität von Einstellungen, IQ
	4.Verhältnisskala	Ähnlichkeits-Transformationen	Gleichheit von Verhältnissen	Längenmessungen, Gewichtsmessungen, Häufigkeiten pro Person

Transformation auf ein niedriges Skalenniveau:

Ordinaldaten/Nominaldaten, Ordinaldaten in Nominaldaten

Wann wird dies angewandt?

- Im Nachhinein Zweifel an höherem Skalenniveau
- Asymmetrische Merkmalsverteilung in Stichprobe
- Vergleiche von Daten, die auf unterschiedlichen Skalenniveaus gemessen wurden
- Kodierung als binäre Indikatorvariable, z.B. bestanden/nicht bestanden
- Vereinfachung für Gruppendarstellung oder Diagramm

Transformation auf ein höheres Skalenniveau:

Normalerweise ist es nicht möglich, bereits erhobene Daten von einem informationsärmeren Skalenniveau auf ein informationsreichen Niveau zu heben. Eine Ausnahme stellt - unter bestimmten Bedingungen - die Transformation ordinalskalierter Daten auf Intervallskalenniveau dar.

1.Wurden **Ordinaldaten** durch **direkte Rangordnung** erhoben, so lassen sich diese nach dem "**Law of Categorical Judgement**" in **intervallskalierte Daten** transformieren.

2.Wurden **Ordinaldaten** mittels **indirekter Rangordnung** (Dominanzpaarvergleiche) generiert, lässt sich durch das "**Law of Comparative Judgement**" diese Daten in **Intervalldaten** überführen.

3.**Eine weitere**, mathematisch anspruchsvolle Methode zur Überführung von Ordinaldaten in Kardinaldaten stellt das **Markovketten-Monte-Carlo-Verfahren** dar.

Die **Korrelation** ist ein Maß für den statistischen **Zusammenhang** zwischen Datensätzen. Sie quantifiziert so die lineare Abhängigkeit in einem Wertebereich von -1 bis +1. Bei einem Wert von 0 gibt es keinen Zusammenhang. Die Interpretation der Stärke des Zusammenhangs erfolgt nach Cohen mit: ab 0,1 klein, ab 0,3 mittel und ab 0,5 stark. Die Korrelation kann auch als **standardisierte Kovarianz** gesehen werden und ist so nicht beeinflussbar durch lineare Transformationen.

Für unterschiedliche Daten mit unterschiedlichen Skalen gibt es jeweils besondere Korrelationskoeffizienten. Die wohl bekannteste in der Psychologie ist die **Produkt-Moment-Korrelation**, die bei mindestens **intervallskalierten Datensätzen** angewendet wird. Wenn einer der beiden Datensätze nur (natürlich) **dichotom nominalskaliert** ist und der andere mindestens intervallskaliert verwendet man die **Punktbiseriale Korrelation**

$$r_{xy} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}} =$$

1. Produkt-Moment-Korrelation

Mittelwert der Y-Werte, für die X=1 Mittelwert der Y-Werte, für die X=0

$$r_{pb} = \frac{\bar{y}_1 - \bar{y}_0}{s_x} \cdot \sqrt{\frac{n_0 \cdot n_1}{n^2}}$$

← Anzahl der Fälle, für die X=0 (n_0) bzw. X=1 (n_1)

2. Punktbiseriale Korrelation

Die **Tetrachorische Korrelation** wird eingesetzt, wenn zwei Variablen künstlich dichotomisiert sind und eine Normalverteilung angenommen werden kann. Die Korrelation wird überschätzt wenn die Randverteilungen stark asymmetrisch sind oder ein $n_{xy} < 5$ ist.

$$r_{tet} = \cos \frac{\pi}{1 + \sqrt{\frac{n_{11} \cdot n_{22}}{n_{12} \cdot n_{21}}}}$$

Die künstliche Dichotomisierung von intervallskalierten Daten führt zur Setzung eines impliziten Kriteriums. Um somit entstandene beliebige Werte zu verhindern, wird die **Biseriale Korrelation** eingesetzt. Diese hat jedoch eine Normalverteilung der Daten als Voraussetzung, weswegen sie der Punktbiserialen Korrelation unterlegen ist.

$$r_{bis} = \frac{\bar{y}_1 - \bar{y}}{s_y} \cdot \frac{n_1}{n \cdot \omega}$$

Wichtig ist zu verstehen, dass man von Korrelation nicht auf Kausalität schließen kann. Um auf Kausalität schließen zu können hat Hill 1965 9 Kriterien formuliert:

1. **Stärke der Assoziation** (Effektstärke)
2. **Konsistenz**
3. **Spezifität**
4. **Zeitbedingtheit**
5. **Dosis-Wirkungs-Beziehung**
6. **(Biologische) Plausibilität**
7. **Koheränz**
8. **Experiment**
9. **Analogie**