Wahrscheinlichkeitsrechnung und mathematische Statistik
Zusammenfassung

Maximilian Springenberg

0 Univariate Daten

0.1 Skalentypen

Daten können zu verschiedenen Skalenniveaus, bzw. Typen klassifiziert werden. Je nach Skalentyp sind dann bestimmte Herangehensweisen hinsichtlich der Analyse/ Aufbereitung der Daten sinnvoll.

nominal	ordinal	intervall	verhältniss
Klassen-label	Größenordnung	Differenzen	Verhältnisse

0.2 Generelle Definitionen

0.2.1 Häufigkeiten

Absolute Häufigkeit:

$$N_j = N[x(j)] = \sum_{i=1}^{N} d_i(j), d_i(j) := I_{x(e_i)=x(j)}$$

Relative Häufigkeit:

$$f_j = \frac{N_j}{N}$$

Population:

$$M_N = \{e_1, \ldots, e_N\}$$

Quantitatives Merkmal X mit Ausprägung $x \in W_X$
Wertebereich W_X von X:

$$W_X = \{x(j)|j = 1, \ldots, J\}$$

Urliste:

$$D_N = \{x_n|n = 1, \ldots, N\}$$

Rangliste:

$$R = x_{(1)}, \ldots, x_{(N)}$$

, mit $x_{(1)} \leq \ldots \leq x_{(N)}$

1 Kennzahlen

1.1 Lage

Uns interessiert bei Daten insbesondere die Verteilung derer. Wir möchten also mit möglichst wenigen und einfachen Mitteln erfassen wie Streuung/ Mittel-/ Medianwerte sich verhalten.

1.2 Allgemeine Definitionen

1.2.1 Mittel

Arithm. Mittel:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Median:

$$med_x := \begin{cases} x_{(\frac{n+1}{2})} & \text{, n ungerade} \\ \frac{x_{(\frac{n}{2})}+x_{(\frac{n}{2}+1)}}{2} & \text{, sonst} \end{cases}$$

Modus: Häufigster Wert.

1.2.2 Quantile

Ein p-Quantil, $p \in [0,1]$ ist eine Zahl, für die $100 \cdot p\%$ kleiner-gleich sind und $100 \cdot (1-p)\%$ der Werte größer-gleich

$$Q_p := \begin{cases} x_{(j)} & , np \text{ nicht ganzzahlig, } j := \lceil np \rceil \\ \frac{x_{(j)}+x_{(j+2)}}{2} & \text{, sonst, } j := np \end{cases}$$

1.2.3 Abweichungen

absolute Abweichung:

$$\Delta_a(x) = \sum_{i=1}^{N} |x_i - x|$$

Quadratische Abweichung:

$$\Delta(x) = \sum_{i=1}^{N} (x_i - x)^2$$

Hierbei ist insbesondere interassant, dass (x) für $x = \bar{x}$ minimal ist.