

1 Beschreibende Statistik
1.1 Begriffe

1.1.1 Beschreibende/Deskriptive Statistik

Beobachtete Daten werden durch geeignete statistische Kennzahlen charakterisiert und durch geeignete Grafiken anschaulich gemacht.

1.1.2 Schließende/Induktive Statistik

Aus beobachtete Daten werden Schlüsse gezogen und diese im Rahmen vorgegebener Modelle der Wahrscheinlichkeitstheorie bewertet.

1.1.3 Grundgesamtheit

Ω : Grundgesamtheit ω : Element oder Objekt der Grundgesamtheit diskret (< 30 Ausprägungen), stetig (≥ 30 Ausprägungen), univariat ($p=1$), multivariat ($p>1$)

1.2 Lagemaße

1.2.1 Modalwerte x_{mod}

Am häufigsten auftretende Ausprägungen (insbesondere bei qualitativen Merkmalen)

1.2.2 Mittelwert

R: $mean(x)$
Schwerpunkt der Daten
Empfindlich gegenüber Ausreißern.

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

1.3 Median

R: $median(x)$
Liegt in der Mitt der sortierten Daten x_i .
Unempfindlich gegenüber Ausreißern.

$$x_{0.5} = \begin{cases} x_{\frac{n+1}{2}}, & \text{falls } n \text{ ungerade} \\ \frac{1}{2}(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}), & \text{falls } n \text{ gerade} \end{cases}$$

1.4 Streuungsmaße

1.4.1 Spannweite

$\max x_i - \min x_i$

1.4.2 Stichprobenvarianz s^2

R: $var(x)$
Verschiebungssatz:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} (\sum_{i=1}^n x_i^2 - n \bar{x}^2)$$

$n\bar{x}^2$) Gemittelte Summe der quadratischen Abweichung vom Mittelwert

1.4.3 Stichprobenstandardabweichung

R: $sd(x)$
 $s = \sqrt{s^2}$ Streuungsmaß mit gleicher Einheit wie beobachteten Daten x_i . \bar{x} minimiert die "quadratische Verlustfunktion" oder die Varianz gibt das Minimum der Fehlerquadrate an.

1.5 p-Quantile

R: $quantile(x, p)$. Teilt die sortierten Daten x_i ca. im Verhältnis p: (1-p) d.h. $\hat{F}(x_p) \approx p$. 1. Quartil = 0.25-Quantil; Median = 0.5-Quantil; 3. Quartil = 0.75-Quantil;

1.6 Interquartilsabstand I

$I = x_{0.75} - x_{0.25}$. Ist ein weiterer Streuungsparameter.

1.7 Chebyshev

$\frac{N(S_k)}{n} > 1 - \frac{1}{k^2}$, für alle $k \geq 1$ \bar{x} der Durchschnitt, $s > 0$ die Stichproben-Standardabweichung von Beobachtungswerten x_1, \dots, x_n . Sei $S_k = \{i, 1 \leq i \leq n : |x_i - \bar{x}| < k \cdot s\}$; Für eine beliebige Zahl $k \geq 1$ liegen mehr als $100 \cdot (1 - \frac{1}{k^2})$ Prozent der Daten im Intervall von $\bar{x} - ks$ bis $\bar{x} + ks$. **Speziell:** Für $k=2$ liegen mehr als 75% der Daten im 2s-Bereich um \bar{x} . Für $k=3$ liegen mehr als 89% der Daten im 3s-Bereich um \bar{x} . **Komplement Formulierung:** $\bar{S}_k = \{i || x_i - \bar{x}| \geq k \cdot s\}$; $\frac{N(\bar{S}_k)}{n} \leq \frac{1}{k^2}$; Die Ungleichheit liefert nur eine sehr grobe Abschätzung, ist aber unabhängig von der Verteilung der Daten. **Empirische Regeln** 68% der Daten im Bereich um $\bar{x} \pm s$. 95% um $\bar{x} \pm 2s$. 99.7% um $\bar{x} \pm 3s$.

1.8 Korrelation

Grafische Zusammenhang zwischen multivariaten Daten y und x durch ein Streudiagramm. Kennzahlen zur Untersuchung des Zusammenhangs:

1.8.1 Empirische Kovarians

R: $cov(x, y)$; $s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n-1} (\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y})$

1.8.2 Empirische Korrelationskoeffizient r

R: $cor(x, y)$; $r = \frac{s_{xy}}{s_x s_y}$; Näherungsweise lin. Zusammenhang zw. x und y, falls $|r| \approx 1$.

1.8.3 Regressionsgerade y

$y = mx + t$ mit $m = r \cdot \frac{s_y}{s_x}$ und $t = \bar{y} - m \cdot \bar{x}$

2.1 Begriffe

Ergebnisraum Ω : Menge aller möglichen Ergebnisse eines Experiments

Elementarereignis $\omega \in \Omega$: einzelnes Element von Ω

Ereignis $E \subseteq \Omega$: beliebige Teilmenge des Ergebnisraums Ω heißt sicheres Ereignis, \emptyset heißt unmögliches Ereignis

Vereinigung $E \cup F$: Ereignis E oder Ereignis F treten ein. $\bigcup_{i=1}^n E_i$: mindestens ein Ereignis E_i tritt ein.

Schnitt $E \cap F$: Ereignis E und Ereignis F treten ein.

$\bigcap_{i=1}^n E_i$ alle Ereignisse E_i treten ein. **Generereignis** $\bar{E} = \Omega / E$: Ereignis E tritt nicht ein (Komplement von E)

Disjunkte Ereignisse E und F: $E \cap F = \emptyset$

2.2 De Morgan'schen Regeln

$\overline{E_1 \cup E_2} = \bar{E}_1 \cap \bar{E}_2$
 $\overline{E_1 \cap E_2} = \bar{E}_1 \cup \bar{E}_2$

2.3 Wahrscheinlichkeit

$0 \leq P(E) \leq 1$; $P(\Omega) = 1$;
 $P(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i)$, falls $E_i \cap E_j = \emptyset$ für $i \neq j$

2.3.1 Satz 2.1

$P(\bar{E}) = 1 - P(E)$
 $P(E \cup F) = P(E) + P(F) - P(E \cap F)$
(Übungsaufgabe!!! Ergänzen)

2.4 Laplace-Experiment

Zufallsexperimente mit n gleich wahrscheinlichen Elementarereignissen. Dann berechnet sich die Wahrscheinlichkeit $P(E)$ für $E \subseteq \Omega$ aus:

$$P(E) = \frac{\text{Anzahl der Ereignisse}}{\text{Anzahl der möglichen Ereignisse}} = \frac{\text{Mächtigkeit von } E}{\text{Mächtigkeit von } \Omega} = \frac{|E|}{|\Omega|}$$

2.5 Kombinatorik

2.5.1 Allgemeines Zählprinzip

Anzahl der Möglichkeiten für ein k-stufiges Zufallsexperiment mit n_i Varianten im i-ten Schritt: $n_1 \cdot n_2 \cdot \dots \cdot n_k$

2.5.2 Permutationen

Anzahl einer n-elementigen Menge n-maliges Ziehen ohne Zurücklegen mit Beachtung der Reihenfolge: **n unterscheidbare Elemente:** $n! = n \cdot (n-1) \cdot \dots \cdot 2 \cdot 1$
k Klassen mit je n_i nicht unterscheidbaren Elementen $n = \sum_{i=1}^k n_i$: $\frac{n!}{n_1! \cdot n_2! \cdot \dots \cdot n_k!}$

2.5.3 Abzähl k-elementigen Teilmengen einer n-elementigen Menge k-maliges Ziehen aus einer n-elementigen Menge

ohne Zurücklegen = $k \leq n$.
mit Zurücklegen = $k > n$ möglich.

mit Beachtung der Reihenfolge, ohne Zurücklegen: $\frac{n!}{(n-k)!}$

ohne Beachtung der Reihenfolge, ohne Zurücklegen: $\binom{n}{k} = \frac{n!}{(n-k)! \cdot k!}$

mit Beachtung der Reihenfolge, mit Zurücklegen: n^k

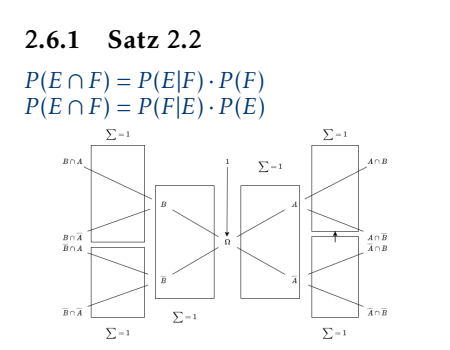
ohne Beachtung der Reihenfolge, mit Zurücklegen: $\binom{n+k-1}{k}$

2.6 Bedingte Wahrscheinlichkeit

$P(E|F) = P_F(E) = \frac{|E \cap F|}{|F|} = \frac{P(E \cap F)}{P(F)}$

2.6.1 Satz 2.2

$P(E \cap F) = P(E|F) \cdot P(F)$
 $P(E \cap F) = P(F|E) \cdot P(E)$



2.6.2 Satz der totalen Wahrscheinlichkeit

Sei $\Omega = \bigcup_{i=1}^n E_i$ mit $E_i \cap E_j = \emptyset$ für $i \neq j$ d.h. die Ereignisse bilden eine disjunkte Zerlegung bzw. eine Partition von Ω . Somit gilt:

$P(F) = \sum_{i=1}^n P(F \cap E_i) = \sum_{i=1}^n P(F|E_i) \cdot P(E_i)$

Summe der Äste des Wahrscheinlichkeitsbaums zu allen Schnitten $F \cap E_i$



2.6.3 Vierfeldertafel

$P(F) = P(F \cap E) + P(F \cap \bar{E})$
 $P(\bar{F}) = P(\bar{F} \cap E) + P(\bar{F} \cap \bar{E})$

	E	\bar{E}	
F	$P(F \cap E)$	$P(F \cap \bar{E})$	$P(F)$
\bar{F}	$P(\bar{F} \cap E)$	$P(\bar{F} \cap \bar{E})$	$P(\bar{F})$
	$P(E)$	$P(\bar{E})$	1

Satz 2.2 $P(E \cap F)P(E) \cdot P(F|E) = P(F) \cdot P(E|F)$ **Tafel** $P(F) - P(F \cap \bar{E}) = P(E) - P(\bar{F} \cap E)$; $P(\bar{F}|E) = 1 - P(F|E)$

2.6.4 Formel von Bayes

Hilfreich, wenn man $P(F|E_i)$ kennt, aber nicht $P(E_k|F)$ **Satz 2.4** $P(E_k|F) = \frac{P(F|E_k) \cdot P(E_k)}{\sum_{i=1}^n P(F|E_i) \cdot P(E_i)}$

$$\sum_{i=1}^n P(F|E_i) \cdot P(E_i)$$

Nur Nenner! $P(F)$ aus dem Satz der totalen Wahrscheinlichkeit.

2.6.5 Stochastische Unabhängigkeit

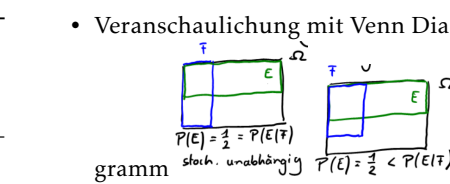
Übung Die Ereignisse E und F heißen (stochastisch) unabhängig, wenn die Information über das Eintreten des einen Ereignisses die Wahrscheinlichkeit für das Eintreten des anderen Ereignisses nicht ändert, d.h. falls $P(E|F) = P(E)$ oder $P(E \cap F) = P(E) \cdot P(F)$

$$= \frac{P(E \cap F)}{P(F)}$$

Es gilt Falls die Ereignisse E, F unabhängig sind, dann sind auch:

E, \bar{F}
 \bar{E}, F
 \bar{E}, \bar{F} unabhängig **Bemerkung**

- Stochastische Unabhängigkeit bedeutet nicht notwendigerweise eine kausale Abhängigkeit
- Veranschaulichung mit Venn Diagramm



- $A, B \neq \emptyset$ und $A \cap B = \emptyset$
 $P(A \cap B) \stackrel{?}{=} P(A) \cdot P(B)$
 $\emptyset \neq P(A) \cdot P(B)$ da $P(A) > 0$ und $P(B) > 0$
 $\Rightarrow A, B$ stochastisch abhängig

3 Zufallsvariable

Abbildung des **abstrakten** Ergebnisraums Ω auf \mathbb{R} . Eine Abbildung $X: \Omega \rightarrow \mathbb{R}$, $\omega \mapsto X(\omega)$ heißt Zufallsvariable (ZV). $x \in \mathbb{R}$ heißt Realisation der ZV X.

- Diskrete ZV: $X(\Omega) = x_1, \dots, x_2 (n \in \mathbb{N})$; z.B. X = "Augensumme beim Würfeln"
- Stetige ZV: $X(\Omega) \subseteq \mathbb{R}$; "z.B. Körpergröße eines Menschen"

3.1 Verteilungsfunktion