BeschreibendeStatistik 1.4.3 Stichprobenstandardabweichungebnisse eines Experiments 1.1 Begriffe R:sd(x) $s = \sqrt{s}$ Streuungsmaß mit gleicher Einheit wie beobachteten Daten $x_i.\bar{x}$ minimiert 1.1.1 Beschreibende/Deskriptive die "quadratische Verlustfunktionöder Statistik die Varianz gibt das Minimum der Feh-Beobachtete Daten werden durch geeiglerquadrate an. nete statistische Kennzahlen charakteri-1.5 p-Quantile siert und durch geeignete Grafiken an-R:quantile(x, p). Teilt die **sortierten** Da- **Schnitt** $E \cap F$: Ereignis E und Ereignis F schaulich gemacht. 1.1.2 Schließende/Induktive Sta-

Aus beobachtete Daten werden Schlüsse gezogen und diese im Rahmen vorgegebener Modelle der Wahrscheinlichkeitstheorie bewertet. 1.1.3 Grundgesamtheit

 Ω : Grundgesamtheit ω :Element oder Ob-

jekt der Grundgesamtheit diskret(<30 Ausprägungen), stetig(≥30 Ausprägun-

gen), univariat(p=1), mulivariat(p>1) 1.2 Lagemaße 1.2.1 Modalwerte x_{mod} Am häufigsten auftretende Ausprägun-

Hilfszettel zur Klausur von JD., Seite 1 von 2

gen (insbesondere bei qualitativen Merkmalen) 1.2.2 Mittelwert

Schwerpunkt ten. Empfindlich gegemüber Ausreißern. $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} x_i$

R:mean(x)

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

 $\frac{x_{n+1}}{2}$, falls n ungerade $\left(\frac{1}{2}(x_{\frac{n}{2}} + x_{\frac{n}{2}+1})\right)$, falls n gerade

1.4 Streuungsmaße 1.4.1 Spannweite

 $\max x_i$ - $\min x_i$

1.4.2 Stichprbenverians s^2

R:var(x)Verschiebungssatz: $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x}^2) = \frac{1}{n-1} (\sum_{i=1}^{n} x_i^2 - y = mx + t \text{ mit } m = r \cdot \frac{s_y}{s_x} \text{ und } t = \overline{y} - m \cdot \overline{x}$

schen Abweichung vom Mittelwert

ten 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-Quartil; 1.6 Interquartilsabstand I $I = x_{0.75} - x_{0.25}$. Ist ein weiterer Streu- 2.2 De Morgan'schen Regeln ungsparameter.

1.7 Chebyshev $\frac{N(S_k)}{n} > 1 - \frac{1}{k^2}$, für alle $k \ge 1 \overline{x}$ der Durchschnitt, s > 0 die Stichproben-Standardabweichung von Beobachtungswerten $x_1,...,x_n$. Sei $S_k = \{i, 1 \le i \le n : |x_i - \overline{x}| < k \cdot s\}$; Für eine beliebige Zahl

 $k \ge 1$ liegen mehr als $100 \cdot (1 - \frac{1}{L^2})$ Pro-

zent der Daten im Intervall von $\bar{x} - ks$ bis

 $\overline{x} + ks$. **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 \overline{x} . **Komplement Formulie**rung: $\overline{S}_k = \{i | |x_i - \overline{x}| \ge k \cdot s\}; \frac{N(S_k)}{n} \le \frac{1}{k^2};$ Die Ungleichheit lifert nur eine sehr grobe Abschätzung, ist aber unabhängig

von der Verteilung der Daten. Empirische Regeln 68% der Daten im Bereich um $\overline{x} \pm s$. 95% um $\overline{x} \pm 2s$. 99.7% um $\overline{x} \pm 3s$. 1.8 Korrelation Grafische Zusammenhang zwischen multivariaten Daten y und y durch ein Streudiagramm. Kennzahlen zur Unter-

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

 $\frac{1}{n-1}(\sum_{i=1}^{n}(x_iy_i-n\overline{xy}))$

suchung des Zusammenhangs:

1.8.2 Empirische Korrellationsko-

R:cor(x, x); $r = \frac{s_{xy}}{s_x x_y}$; Näherungsweise lin.

Zusammenhang zw. x und y, falls $|r| \approx 1$.

1.8.3 Regressionsgerade y

 $n\bar{x}^2$) Gemittelte Summe der quadrati- 2 Wahrscheinlichkeitsrechnung 2.1 Begriffe

Ergebnisraum Ω : Menge aller möglichen **Elementarereignis** $\omega \in \Omega$: einzelnes Element von Ω

ohne Zurücklegen = $k \le n$. mit Zurücklegen = k > n möglich. **Ereignis** $E \subseteq \Omega$: beliebige Teilmenge des Ergebnisraums Ω heißt sicheres Ereignis, Ø heißt unmögliches Ereignis **Vereinigung** $E \cup F$: Ereignis E oder Ereig-

Zurücklegen: $\frac{n!}{(n-k)!}$ ohne Beachtung der Reihenfolge, ohne nis F treten ein. $\bigcup_{i=1}^{n} E_i$: mindestens ein **Zurücklegen**: $\binom{n}{k} = \frac{n!}{(n-k)!k!}$ mit Beachtung der Reihenfolge, mit Zurücklegen: nk $\bigcap_{i=1}^{n} E_i$ alle Ereignisse E_i treten ein. **Ge**genereignis $\overline{E} = \Omega / E$: Ereignis E tritt nicht ein (Komplement von E)

ohne Beachtung der Reihenfolge, mit Zurücklegen $\binom{n+k-1}{k}$ 2.6 Bedingte Wahrscheinlichkeit **Disjunkte Ereignisse**E und F: $E \cap F = \emptyset$ $P(E|F) = P_F(E) = \frac{|E \cap F|}{|F|} = \frac{P(E \cap F))}{P(F)}$ 2.6.1 Satz 2.2

Ereignis E_i tritt ein.

 $E_1 \cup E_2 = E_1 \cap E_2$

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

2.3.1 Satz 2.1

 $P(\overline{E}) = 1 - P(E)$

2.3 Wahrscheinlichkeit

 $0 \le P(E) \le 1$; $P(\Omega) = 1$;

(Übungsaufgabe!!! Ergänzen) 2.4 Laplace-Experiment Zufallsexperimente mit n gleich wahr-Elementarereignissen.

Dann berechnet sich die Wahrscheinlich-

 $P(E \cup F) = P(E) + P(F) - P(E \cap F)$

 $P(\bigcup_{i=1}^{\infty}) = \sum_{i=1}^{\infty} P(E_i)$, falls $E_i \cap E_j = \emptyset$

$P(E) = \frac{AnzahlderfrEgnstigenEreignisse}{AnzahldermglichenEreignisse}$ $\frac{\textit{MchtigkeitvonE}}{\textit{Mchtigkeitvon}\Omega} = \frac{|E|}{\Omega} \textbf{text}$

keit P(E) für $E \subseteq \Omega$ aus:

2.5 Kombinatorik 2.5.1 Allgmeines Zählprinzip

Anzahl der Möglihckeiten für ein kstufiges Zufallsexperiment mit n_i Varianten im i-ten Schritt: $n_1 \cdot n_2 \cdot ... \cdot n_k$

2.5.2 Permutationen

maliges Ziehen ohne Zurücklgen mit Beachtung der Reihenfolge: n unterscheid**bare Elemente**: $n! = n \cdot (n-1) t ext b f ... 2 \cdot 1$ k Klassen mit je n_i nicht unterscheidbaren Elementen $n = sum_k^{i=1} n_i$: $\frac{n!}{n_1! \cdot n_2! \cdot n_k!}$

Anzahl einer n-elementigen Menge n-

Hilfreich, wenn man man $P(F|E_i)$ kennt, aber nicht $P(E_k|F)$ Satz 2.4 $P(E_k|F) =$ $P(F|E_k) \cdot P(E_k)$

 $P(F|E_i) \cdot P(E_i)$ mit Beachtung der Reihenfolge, ohne Nur Nenner!P(F) aus dem Satz der totalen Wahrscheinlichkeit.

2.6.5 Stochastische Unabhängig-

Uebung Die Ereignisse E und F heißen (stochastisch) unabhängig, wenn die Information über das Eintreten des einen

2.6.4 Formel von Bayes

Ereignisses die Wahrscheinlichkeit für das Eintreten des anderen Ereignisses nicht ändert, d.h. falls $P(E|F) = P(E)oderP(E \cap F) = P(E) \cdot P(F)$ $= \frac{P(E \cap F)}{P(F)}$

gig sind, dann sind auch: \overline{E} , \overline{F} unabhängig **Bemerkung** Stochastische Unabhängigkeit be-

Es gilt Falls die Ereignisse E, F unabhän-

deutet nicht notwendigerweise ei-

ne kausale Abhängigkeit Veranschaulichung mit Venn Dia-

 $P(E) = \frac{4}{2} = P(E(F))$

gramm stock unabhanging P(E)= 1 < P(E| F) • $A, B \neq \emptyset$ 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

=> A, B stochastisch abhängig 3 Zufallsvariable Abbildung des **abstrakte** Ergebnisraums

 Ω auf \mathbb{R} . Eine Abbildung $X: \Omega \to \mathbb{R}$, $\omega \mapsto X(\omega)$ = heißt Zufallsvariable (ZV). x ∈ R. heißt Realisation der ZV X.

2.6.2 Satz der totalen Wahrschein-

Sei $\Omega = \bigcup_{i=1}^n E_i$ mit $E_i \cap E_j = \emptyset$ für $i \neq j$

d.h. die Ereignisse bilde eine disjunkte

Zerlegung bzw. eine Partition von Ω . So-

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

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

2.5.3 Anzahl k-elementigen Teil-

mengen einer n-elementigen

Menge k-maliges Ziehen aus

einer n-elementigen Menge

2.6.3 Vierfeldertafel

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

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

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

E E

mit gilt:

P(TAE) P(TAE) P(T)

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

Würfeln

• Stetige ZV: $X(\Omega) \subseteq \mathbb{R}$; "z.B. Körpergröße eines Menschen"

3.1 Verteilungsfunktion-allg.

• Diskrete ZV: $X(\Omega) = x_1, ..., x_2 (n \in$

 \mathbb{N}); z.B. X = "Augensumme beim"

Die Wahrscheinlichkeit P(B) für ein Ereignis B in R wird zurückgefürht auf die Währscheinlichkeit der entsprechenden Ereignisse in Ω . Für jedes $X \in \mathbb{R}$ ist die Hilfszettel zur Klausur von **JD**., Seite 2 von 2

Verteilungsfunktion F: $\mathbb{R} \to [0,1]$ einer ZV X definiert durch: $F(x) = P(X \le x)$

- $0 \le F(x) \le 1$
- $\lim_{x \to -\infty} F(X) = 0 \lim_{x \to \infty} F(x) = 1$
- monoton wachsend
- P(X > x) = 1 F(x)
- $P(a < X \le b) = F(b) F(a)$

3.2 Diskrete ZVs

Für eine diskrete ZV X mit $X(\Omega) = x_1,...,x_n$ (n endlich oder abzählbar unendlich) ist die Wahrscheinlichkeitsfunktion definiert durch:

$$p(x) = \begin{cases} P(X = x_i), \text{falls } x_i \in X(\Omega) \\ 0, sonst \end{cases}$$
 (2)

Es gilt:

- $F(x) = (P(X \le x) = \sum_{x_i \le x} p(x_i)$
- F(x) ist eine rechtseitig stetige **Treppenfunktion** mit **Sprüngen** bei der Realisation von x_i .

3.3 Stegite ZVs

Stetige ZV X ist die Wahrscheinlichkeitsdichte f $f: \mathbb{R} \to [0, \infty[$ definiert durch

$$P(a < X < b) = \int_a^b f(x) dx$$

Es gilt:

- $F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)dt$ und F'(x) = f(x)
- F(x) ist stetig & $P(a < X \le b) = P(a \le X \le b)$ wegen P(X = a) = 0

3.4 Verteilungsfunktion

 $\int_{\mathbf{Untergrenze}}^{x} \mathbf{Es}$ wird normal mit - Integriert.

3.5 Zusammenfassung

3.5.1 Diskrete ZV

- Wahrscheinlichkeitsverteilung p(x) ∑_iⁿ p(x_i) = 1x_i ist Realisation der ZV.
- Verteilungsfunktion F(x) ist rechtsseitig stetige Treppenfunktion.
 Sprunghöhen:P(X = x_i) = F(x_i) lim ≠ 0
- $P(a < X \le b) = F(b) F(a) \ne P(a \le X \le b)$

3.5.2 Stetige ZV

- Dichtefunktion fx $\int_{-\infty}^{\infty} f(x)dx = 1$
- Verteilungsfunktion F(x) ist stetig mit F'(x) = f(x); $P(X = x_i) = 0$
- $P(a < X \le b) = F(b) F(a) = P(a \le X \le b) = F(a \le X < b) = P(a < X < b)$

3.6 Erwartungswert

Der Erwartungswert E[X] = einer ZV X ist der **Schwerpunkt** ihrer Verteilung or der durchschnittliche zu erwartende Wert der ZV.

- diskrete ZV: $E[X] = \sum_{i=1}^{n} x_i \cdot p(x_i)$
- stetige ZV: $E[X] = \int_{-\infty}^{\infty} x \cdot f(x) dx$

ZV ist konstant. E[X] verhält sich linear. Eigenschaften von E[X]:

- E[b] = b
- E[aX + b] = aE[X] + b
- $E[X_i + \dots + X_n] = \sum_{i=1}^n E[X_i]$
- $\sum_{i=1}^{n} x_i$

3.6.1 Satz 3.1

Sei Y = g(X) eine Funktion der ZV X. Dann gilt:

- für diskrete $\sum_{i=1}^{n} g(x) \cdot p(x_i)$ ZV:E[g(X)]
- für stetige ZV: $E[g(X)] = \sum_{i=1}^{infty} g(x) \cdot f(x)dx$. Das vertauschen von E und g nur bei **linearen**Funktionen möglich. \Rightarrow g(E[X])

3.7 Varianz

Die Varianz einer ZV X mit μ ist ein quadratisches Streungsmaß. $\sigma^2 = Var[X] = E[(X-)^2]^{\text{falls } x \text{ stetig}} \int_{-\infty}^{\infty} (x-\mu)^2 \cdot f(x)$

g(X)

Die Standardabweichung $\sigma = \sqrt{Var[X]}$ hat im Gegensatz zur Varianz die gleiche Dimension von die ZV X.

- Var[b] = 0
- $Var[aX + b] = a^2 Var[X]$

3.7.1 Satz 3.2

 $Var[X] = E[X^2] - (E[X])^2$ Beim Minuend wird beim Erwartungswert nur das einfach stehende x quadriert nicht f(x)!!!

3.8 Z-Transformation, Standardisierung

Sei X eine ZV mit μ und σ . Dann ist $Z = \frac{X - \mu}{\sigma} = \frac{x}{\sigma} - \frac{\mu(konstant)}{\sigma}$

3.9 Kovarianz

Eigenschaften:

- Cov[X, Y] = Cov[Y, X]
- Cov[X,X] = Var[X]
- Cov[aX, Y] = aCov[X, Y]

Die Kovarianz zweier ZV (X, Y) ist definiert durch Cov[X,Y] = E[(X - E[X])(Y - E[Y]) Die Kovarianz beschreibt die Abhängigkeit zweier ZV X und Y. **Je** stärker diese Korrelieren, **desto** (betragsmäßig) größer ist die Kovarianz. Falls X, Ystochastisch unabhängig $\Rightarrow Cov[X,Y] = 0$

3.10 Satz 3.3

 $Cov[X, Y] = E[XY] - E[X] \cdot E[Y]$

3.10.1 Varianz einer Summe von ZV

- $\begin{array}{lll} \bullet & Var[X_i & + & \ldots & + & X_n] & = \\ \sum_{i=1}^n \sum_{j=1}^n Cov[X_i, X_j]; & Var[X_1 & + & \\ X_2] & = & Var[X_1] & + & Var[X_2] & + \\ 2Cov[X_1, X_2] & \end{array}$
- Falls X_i, X_j paarweise unabhängig !!!: $Var[X_1 + ... + X_n] = \sum_{i=1}^{n} Var[X_i]$

3.11 Overview μ *σ* **3.11.1** E[X]

 $E[aX + b] = AE[X] + b; EX_1 + ... + E_n = \sum_{i=1}^{n} E[X_i];$ Falls X_1, X_2 unabhängig: $E[X_i] = \mu => E[\overline{X}] = E[\frac{1}{n}(X_1 + ... + X_n)] =$

$$\frac{1}{n}\sum_{i=1}^{n} \underbrace{E[x_i]}_{=} = \frac{1}{n} \cdot n \cdot \mu = \mu$$

3.11.2 Varianz

 $Var[aX + b] = a^{2}Var[X]$

Falls X_i, X_j parweise unabhängig:

$$Va[X_1 + ... + X_n] = \sum_{i=1}^n Var[X_i]$$

$$Var[X_i] = \sigma^2 \Rightarrow Var[\overline{X}] = Var[\frac{1}{n}(x_1 + ... + x_n)]$$

$$|x_n| = \frac{1}{n^2} \sum_{i=1}^n Var[X_i] = \frac{1}{n^2} \cdot n \cdot \sigma^2 = \frac{\sigma^2}{n}$$

3.12 Quantile

Sie X eine ZV mit Verteilungsfunktion F(x) und $0 . Dann ist das p-Quantil definiert als der Wert <math>x_p \in \mathbb{R}$ für den gilt:

$$F(x_p) \geq p$$
.