

Beschreibende Statistik

Ω : Grundgesamtheit

ω : Element der Grundgesamtheit

$X: \Omega \rightarrow M$: Merkmal

$X: (\omega) = x$: Ausprägung des Merkmals

Modalwert(e) x_{mod}		Am häufigsten auftretende Ausprägung
Mittelwert \bar{x}	$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$	
Median $x_{0.5}$	$x_{0.5} = \begin{cases} x_{\frac{n+1}{2}}, & n \text{ ungerade} \\ \frac{1}{2}(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}), & n \text{ gerade} \end{cases}$	Liegt in der Mitte der sortierten Daten x_i
Spannweite	$\max x_i - \min x_i$	Distanz zw. größtem und kleinstem Messwert
Stichprobenvarianz s^2	$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$ $= \frac{n}{n-1} \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right)$	Gemittelte Summe der quadratischen Abweichungen vom Mittelwert
Stichproben- standardabweichung s	$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} = \sqrt{s^2}$	Streuungsmaß mit gleicher Einheit wie beobachtete Daten x_i
p-Quantil x_p	$x_p = \begin{cases} x_{\lfloor np \rfloor + 1}, & n \cdot p \notin \mathbb{N} \\ \frac{1}{2}(x_{np} + x_{np+1}), & n \cdot p \in \mathbb{N} \end{cases}$	Teilt sortierte Daten im Verhältnis $p : (1 - p)$
Empirische Kovarianz s_{xy}	$s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$ $= \frac{1}{n-1} \left(\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y} \right)$	Für $s_{xy} > 0$ hat Punktwolke steigende, für $s_{xy} < 0$ fallende Tendenz
Empirischer Korrelationskoeffizient r	$r = \frac{s_{xy}}{s_x \cdot s_y}$	Näherungsweise linearer Zusammenhang zw. x und y falls $ r \approx 1$
Regressionsgerade	$y = mx + t$ mit $m = r \cdot \frac{s_y}{s_x}$ und $t = \bar{y} - m \cdot \bar{x}$	

Elementare Wahrscheinlichkeitsrechnung

Ω : Ergebnisraum, Menge aller möglichen Ergebnisse eines Experiments

$\omega \in \Omega$: Elementarereignis, Einzelnes Element von Ω

$E \cup F$: Ereignis E oder Ereignis F treten ein (mindestens eins)

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

$E \cap F = \emptyset$: Paarweise disjunkte Ereignisse

De Morgan'sche Regeln

$$\overline{\bigcup_{i=1}^n E_i} = \bigcap_{i=1}^n \bar{E}_i$$

$$\overline{\bigcap_{i=1}^n E_i} = \bigcup_{i=1}^n \bar{E}_i$$

Axiome von Kolmogorov

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

Folgen aus den Axiomen

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

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

Laplace Experiment

$$P(E) = \frac{\text{Anz der für } E \text{ günstigen Ergebnisse}}{\text{Anz der möglichen Ergebnisse}} = \frac{|E|}{n}$$

Zufallsexperiment mit n gleichwahrscheinlichen Elementarereignissen

Bedingte Wahrscheinlichkeit

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

Die Wahrscheinlichkeit für das Eintreten von E , wenn F eingetreten ist.

Regeln für bedingte Wahrscheinlichkeit

- Für $E, F \neq \emptyset$
1. $P(E \cap F) = P(E|F) \cdot P(F)$
 2. $P(E \cap F) = P(F|E) \cdot P(E)$

Satz der totalen Wahrscheinlichkeit

$$\Omega = \bigcup_{i=1}^n E_i \text{ mit } E_i \cap E_j = \emptyset, \text{ dann gilt:}$$

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

Alle E müssen disjunkt sein.
Allgemein die Wahrscheinlichkeit, dass F eintritt. Man nimmt nicht an, dass E_i eingetreten ist.

Vierfeldertafel

$$\text{Spezialfall: } \Omega = E \cup \bar{E}$$

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

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

Formel von Bayes

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

Hilfreich, wenn man $P(F|E_i)$ kennt, nicht aber $P(E_k|F)$

Unabhängigkeit

Gilt, wenn

$$P(E|F) = P(E) \text{ bzw. } P(E \cap F) = P(E) \cdot P(F)$$

Falls E, F unabhängig, gilt auch \bar{E}, F bzw. \bar{E}, \bar{F} unabh

Kombinatorik

Ermittlung der Mächtigkeit von Ereignissen

Allgemeines Zählprinzip	Anzahl der Möglichkeiten, für ein k-stufiges Zufallsexperiment mit n_i Varianten im i-ten Schritt: $n_1 \cdot n_2 \dots n_k$	Baum, der je nach Ebene unterschiedlich viele Kinder haben kann									
Anzahl der Permutationen einer n-elementigen Menge	n-maliges Ziehen ohne Zurücklegen mit Beachtung der Reihenfolge <ul style="list-style-type: none"> für n unterscheidbare Elemente: $n! = n \cdot (n - 1) \dots 2 \cdot 1$ für k Klassen mit je n_i nicht unterscheidbaren Elementen: $\frac{n!}{n_1! \cdot n_2! \cdot \dots \cdot n_k!}$ 										
k-maliges ziehen aus einer n-elementigen Menge	<table> <tr> <td></td><td>mit Beachtung der Reihenfolge</td><td>ohne Beachtung der Reihenfolge</td></tr> <tr> <td>ohne Zurücklegen $k \leq n$</td><td>$\frac{n!}{(n-k)!}$</td><td>$\binom{n}{k} = \frac{n!}{(n-k)!k!}$</td></tr> <tr> <td>mit Zurücklegen $k > n$ möglich</td><td>n^k</td><td>$\binom{n+k-1}{k}$</td></tr> </table>		mit Beachtung der Reihenfolge	ohne Beachtung der Reihenfolge	ohne Zurücklegen $k \leq n$	$\frac{n!}{(n-k)!}$	$\binom{n}{k} = \frac{n!}{(n-k)!k!}$	mit Zurücklegen $k > n$ möglich	n^k	$\binom{n+k-1}{k}$	
	mit Beachtung der Reihenfolge	ohne Beachtung der Reihenfolge									
ohne Zurücklegen $k \leq n$	$\frac{n!}{(n-k)!}$	$\binom{n}{k} = \frac{n!}{(n-k)!k!}$									
mit Zurücklegen $k > n$ möglich	n^k	$\binom{n+k-1}{k}$									

Zufallsvariablen

Abbildung $X: \Omega \rightarrow \mathbb{R}$, $\omega \rightarrow X(\omega) = x$, heißt Zufallsvariable

x : Realisation der ZV X

Diskrete ZV: $X(\Omega) = \{x_1, x_2, \dots, x_n\} (n \in \mathbb{N})$

Stetige ZV: $X(\Omega) \subseteq \mathbb{R}$

Eindimensionale ZV: $X: \Omega \rightarrow \mathbb{R}$

Mehrdimensionale ZV: $X: \Omega \rightarrow \mathbb{R}^p$

Verteilungsfunktion

$x \in \mathbb{R}, F: \mathbb{R} \rightarrow [0, 1]$

$$F(x) = P(X \leq x)$$

Eigenschaften:

- $0 \leq F(x) \leq 1$
- $\lim_{x \rightarrow -\infty} F(x) = 0, \lim_{x \rightarrow \infty} F(x) = 1$
- rechtsseitig stetig, also:
 $\lim_{x \rightarrow b+} F(x) = F(b)$
- monoton wachsend
- $P(X > x) = 1 - F(x)$
- $P(a < X \leq b) = F(b) - F(a)$

Diskrete Zufallsvariablen

Wahrscheinlichkeitsverteilung:

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

Eigenschaften Wahrscheinlichkeitsverteilung:

- $0 \leq p(x) \leq 1$
- $\sum_{i=1}^n p(x_i) = 1$

Es gilt:

- $F(x) = P(X \leq x) = \sum_{x_i \leq x} p(x_i)$
- $F(x)$ ist eine rechtsseitig stetige Treppenfunktion mit Sprüngen bei den Realisationen x_i

Stetige Zufallsvariablen

Wahrscheinlichkeitsdichte:

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

Eigenschaften Wahrscheinlichkeitsdichte:

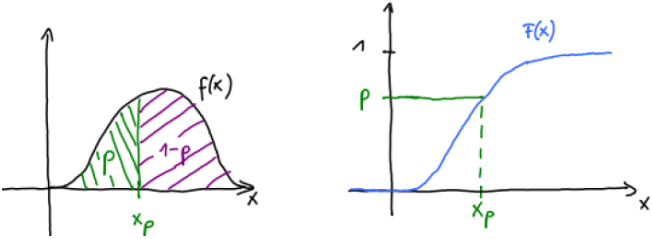
- $f(x) \geq 0$
- $\int_{-\infty}^{\infty} f(x) dx = 1$

Es gilt:

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

Wahrscheinlichkeitsdichte ist die Wahrscheinlichkeit, dass ein Zufallsexperiment ein Ergebnis in einem bestimmten Bereich liefert

Erwartungswert	$E[X] = \mu$ Für diskrete ZV: $E[X] = \sum_{i=1}^n x_i \cdot p(x_i)$ Für stetige ZV: $E[X] = \int_{-\infty}^{\infty} x \cdot f(x) dx$ Sei $g(X)$ eine Funktion der ZV X , dann gilt: Für diskrete ZV: $E[g(X)] = \sum_{i=1}^n g(x_i) \cdot p(x_i)$ Für stetige ZV: $E[g(X)] = \int_{-\infty}^{\infty} g(x) \cdot f(x) dx$ <u>Eigenschaften des Erwartungswerts:</u> <ul style="list-style-type: none"> - $E[b] = b$ - $E[aX + b] = aE[X] + b$ - $E[X_1 + \dots + X_n] = \sum_{i=1}^n E[X_i]$ 	Der Erwartungswert ist der Schwerpunkt der Verteilung einer Zufallsvariable
Varianz	$\sigma^2 = Var[X] = E[(X - \mu)^2]$ <u>Standardabweichung:</u> $\sigma = \sqrt{V[X]}$ <u>Verschiebungssatz:</u> $Var[X] = E[X^2] - (E[X])^2$ <u>Eigenschaften der Varianz:</u> <ul style="list-style-type: none"> - $Var[b] = 0$ - $Var[aX + b] = a^2 Var[X]$ - $Var[X_1 + \dots + X_n] = \sum_{i=1}^n \sum_{j=1}^n Cov[X_i, X_j]$ - $Var[X_1 + X_2] = Var[X_1] + Var[X_2] + 2Cov[X_1, X_2]$ - Falls X_i, X_j paarweise unabhängig: $Var[X_1 + \dots + X_n] = \sum_{i=1}^n Var[X_i]$ 	Die Varianz ist ein quadratisches Streuungsmaß einer ZV X
Kovarianz	$Cov[X, Y] = E[(X - E[X])(Y - E[Y])]$ Wenn X, Y stochastisch unabhängig $\Rightarrow Cov[X, Y] = 0$ <u>Verschiebungssatz:</u> $Cov[X, Y] = E[XY] - E[X] \cdot E[Y]$ <u>Eigenschaften der Kovarianz:</u> <ul style="list-style-type: none"> - $Cov[X, Y] = Cov[Y, X]$ - $Cov[X, X] = Var[X]$ - $Cov[aX, Y] = aCov[X, Y]$ 	Die Kovarianz beschreibt die Abhängigkeit zweier ZV X und Y . Je stärker diese korrelieren, desto größer die Kovarianz.

Quantile	<p>Das p-Quantil ist der kleinste Wert $x_p \in \mathbb{R}$ für den gilt:</p> $F(x_p) \geq p$  <p>Abbildung: p-Quantil einer stetigen ZV mit streng monoton wachsendem $F(x)$:</p> $x_p = F^{-1}(p)$	
Chebyshev-Ungleichung	<p>X: ZV $\mu = E[X]$ $\sigma^2: Var[X]$ Es gilt für jedes beliebige $k \in \mathbb{R}^+ \setminus \{0\}$:</p> $P(X - \mu \geq k \cdot \sigma) \leq \frac{1}{k^2}$	
Schwaches Gesetz der großen Zahlen	<p>Mit Chebyshev und $\mu = E[X_i] \ (i = 1, \dots, n) \Rightarrow \mu = E\left[\frac{X_1 + \dots + X_n}{n}\right] = E[\bar{X}]$ folgt: X_1, X_2, \dots ist eine Folge von unabhängigen, identisch verteilten ZV mit EW $E[X_i] = \mu$ und Varianz $Var[X_i] = \sigma^2$. Dann gilt für ein beliebig kleines $\epsilon > 0$:</p> $P(\bar{X} - \mu > \epsilon) \rightarrow 0 \quad \text{für } n \rightarrow \infty$ <p>d.h. der MW \bar{X} konvergiert stochastisch gegen den EW μ.</p>	