

# Formelsammlung Wahrscheinlichkeit und Statistik

## 1. Lage- und Streumaße

### a) Quantile

Für n geordnete Merkmalsausprägungen

$$x_1, x_2, x_3, \dots, x_n$$

ist das **p-Quantil**

$$Q_p = \begin{cases} x_{[np]} & \text{für } np \notin \mathbb{N} \\ \frac{x_{np} + x_{np+1}}{2} & \text{für } np \in \mathbb{N} \end{cases}$$

### b) Arithmetisches Mittel

- Hat man n Messwerte  $x_1, x_2, \dots, x_n$  eines Merkmals, dann heißt

$$\bar{x} = \frac{1}{n} (x_1 + x_2 + \dots + x_n)$$

das **arithmetische Mittel**

- Hat man k verschiedene Merkmalsausprägungen  $x_1, \dots, x_k$  mit ihren relativen Häufigkeit  $h_1, \dots, h_k$ , so berechnet sich das arithmetische Mittel

$$\bar{x} = x_1 h_1 + x_2 h_2 + \dots + x_k h_k$$

- Liegt eine Klassierung der Merkmale mit Intervallen  $[a_1, a_2), [a_2, a_3), \dots, [a_l, a_{l+1})$  und relativen Häufigkeiten  $h_1, \dots, h_l$  vor, so berechnet man das arithmetische Mittel durch

$$\bar{x} = h_1 \cdot \frac{a_1 + a_2}{2} + h_2 \cdot \frac{a_2 + a_3}{2} + \dots + h_l \cdot \frac{a_l + a_{l+1}}{2}$$

### c) Empirische Varianz und empirische Standardabweichung

- Hat man n Messwerte  $x_1, x_2, \dots, x_n$  eines Merkmals, so heißt

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

die **empirische Varianz**

- Hat man k verschiedene Merkmalsausprägungen  $x_1, \dots, x_k$  mit ihren relativen Häufigkeit  $h_1, \dots, h_k$ , so berechnet sich die empirische Varianz

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

- Liegt eine Klassierung der Merkmale mit Intervallen  $[a_1, a_2), [a_2, a_3), \dots, [a_l, a_{l+1})$  und relativen Häufigkeiten  $h_1, \dots, h_l$  vor, so berechnet man die empirische Varianz durch

$$s^2 = \frac{n}{n-1} \sum_{i=1}^l \left( \frac{a_i + a_{i+1}}{2} - \bar{x} \right)^2 \cdot h_i$$

Als **empirische Standardabweichung**, wird die positive Wurzel der empirischen Varianz  $s = \sqrt{s^2}$  bezeichnet

## 2. Korrelation

$\bar{x}, \bar{y}$  arithmetische Mittelwerte der Merkmalsausprägungen von X und Y

$s_x^2, s_y^2$  Varianz von X, bzw. Y

$s_x, s_y$  Standardabweichungen von X bzw. Y

$s_{xy}$  **Kovarianz von x,y:**  $s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$

$r_{xy}$  **Korrelationskoeffizient von x,y** (normierte Kovarianz)

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

**Regressionsgerade:**  $y = kx + d$  mit  $k = r_{xy} \frac{s_y}{s_x}$  und  $d = \bar{y} - k\bar{x}$

### 3. Wahrscheinlichkeit

Eine reelle Funktion  $P$ , definiert auf einer Ereignismenge heißt Wahrscheinlichkeit, wenn:

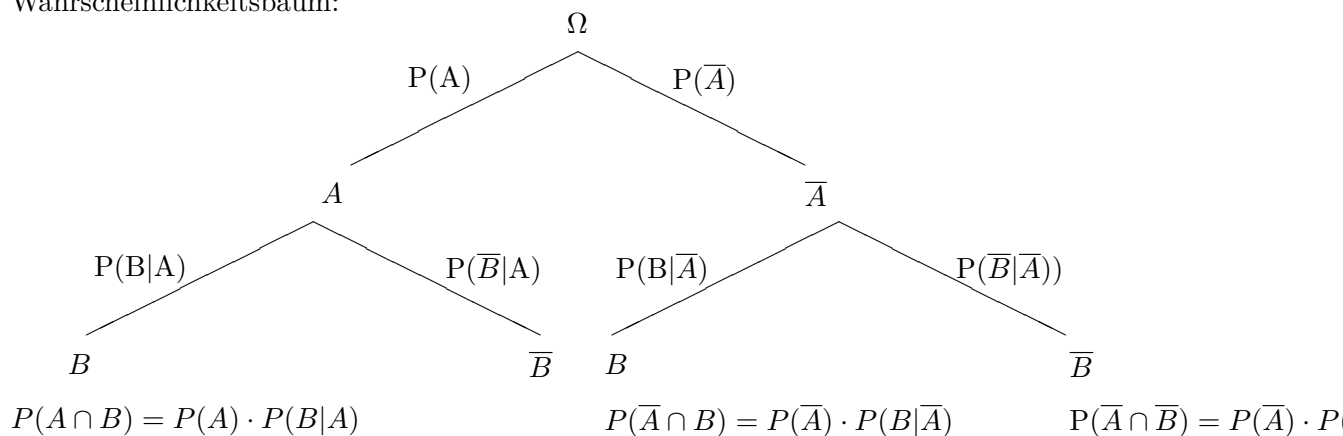
- a) Für jedes Ereignis  $A$  gilt:  $0 \leq P(A) \leq 1$
- b)  $P(\Omega) = 1$
- c) Für zwei disjunkte Ereignisse  $A$  und  $B$  gilt:  $P(A \cup B) = P(A) + P(B)$

### 4. Stochastische Unabhängigkeit

Zwei Ereignisse  $A$  und  $B$  heißen stochastisch unabhängig  $\iff P(A \cap B) = P(A) \cdot P(B)$

### 5. Bedingte Wahrscheinlichkeit $P(A|B) = \frac{P(A \cap B)}{P(B)}$ , mit $P(B) \neq 0$

Wahrscheinlichkeitsbaum:



Satz von der totalen Wahrscheinlichkeit:  $P(A) = P(A|B) \cdot P(B) + P(A|\bar{B}) \cdot P(\bar{B})$

Satz von Bayes: Für  $P(B) \neq 0$  gilt:  $P(A|B) = \frac{P(A)}{P(B)} \cdot P(B|A)$

### 6. Diskrete, stetige Zufallsvariable $X$

	diskrete Zufallsvariable					stetige Zufallsvariable
	$X$	$x_1$	$x_2$	$\dots$	$x_n$	Dichtefunktion $f(t)$
	$P(X = x_i)$	$p_1$	$p_2$	$\dots$	$p_n$	
$P(a \leq X \leq b)$	$\sum_{i: a \leq x_i \leq b} p_i$					$\int_a^b f(t) dt$
Verteilungsfunktion $F(x)$	$\sum_{i: x_i \leq x} p_i$					$\int_{-\infty}^x f(t) dt$ $F'(x) = f(x)$
Erwartungswert $E(X) = \mu$	$\sum_{i=1}^n x_i p_i$					$\int_{-\infty}^{\infty} x f(x) dx$
Varianz $Var(X) = \sigma^2$	$\sum_{i=1}^n (x_i - \mu)^2 p_i$					$\int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$

7. Lineare Kombination einer Zufallsvariable X

$$Z = aX + b \text{ mit } a, b \in \mathbb{R}$$

$$\implies E(Z) = aE(X) + b; \text{Var}(Z) = a^2 \text{Var}(X)$$

$$Z = \frac{1}{\sigma}X - \frac{\mu}{\sigma} = \frac{X - \mu}{\sigma} \text{ mit } E(X) = \mu, \text{Var}(X) = \sigma^2 \implies E(Z) = 0; \text{Var}(Z) = 1$$

8. Einige Verteilungen der Statistik

- a) stetige Gleichverteilung (keine unterschiedliche Wahrscheinlichkeit für Einzelereignisse)

$$\text{Dichtefunktion } f(t) = \begin{cases} 0 & \text{für } t < a \\ \frac{1}{b-a} & \text{für } a \leq t \leq b \\ 0 & \text{für } t > b \end{cases}$$

$$\text{Verteilungsfunktion } F(x) = \begin{cases} 0 & \text{für } t < a \\ \frac{x-a}{b-a} & \text{für } a \leq t \leq b \\ 1 & \text{für } t > b \end{cases}$$

$$E(X) = \frac{a+b}{2}; \text{Var}(X) = \frac{(b-a)^2}{12}$$

- b) Binomialverteilung (Warenausgang; konstante Trefferquote)

p konstante Trefferwahrscheinlichkeit; k Anzahl Treffer bei n Versuchen

$$\text{Wahrscheinlichkeitsfunktion: } P(X = k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \quad \text{mit} \quad \binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}$$

$$E(X) = np; \text{Var}(X) = np(1-p)$$

- c) Hypergeometrische Verteilung (Wareneingang; Waren OK/NOK)

keine konstante Trefferwahrscheinlichkeit

N: Anzahl Elemente der Grundgesamtheit; M: Anzahl Elemente mit Eigenschaft A

n: Anzahl der entnommenen Elemente; k Anzahl Treffer

$$\text{Wahrscheinlichkeitsfunktion: } P(X = k) = \frac{\binom{M}{k} \cdot \binom{N-M}{n-k}}{\binom{N}{n}}$$

$$E(X) = n \cdot \frac{M}{N}; \text{Var}(X) = n \frac{M}{N} \left(1 - \frac{M}{N}\right) \frac{N-n}{N-1}$$

- d) Poissonverteilung (bei Vorkommnisse pro (Zeit-) Einheit)

$$\text{Wahrscheinlichkeitsfunktion: } P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \text{ mit } \lambda > 0, k \in \mathbb{N}_0$$

$$E(X) = \lambda; \text{Var}(X) = \lambda$$

- e) Exponentialverteilung. (Lebensdauer; Zerfallsprozesse)

$$\text{Dichtefunktion: } f(t) = \begin{cases} 0 & \text{für } t < 0 \\ \lambda \cdot e^{-\lambda t} & \text{für } 0 \leq t \end{cases}$$

$$\text{Verteilungsfunktion: } F(x) = \begin{cases} 0 & \text{für } x < 0 \\ 1 - e^{-\lambda x} & \text{für } 0 \leq x \end{cases}$$

$$E(X) = \frac{1}{\lambda}; \text{Var}(X) = \frac{1}{\lambda^2}$$

- f) Weibullverteilung (Ausfallhäufigkeit elektronischer Bauelemente, Werkstoffe)

k > 0: Formparameter

$$\lambda > 0: \text{mittlere Ausfallrate Dichtefunktion: } f(t) = \lambda k (\lambda t)^{k-1} \cdot e^{-(\lambda t)^k} \text{ Verteilungsfunktion: } F(x) = 1 - e^{-(\lambda x)^k}$$

$$\text{Erwartungswert: } E(X) = \lambda^{-1} \cdot \Gamma\left(1 + \frac{1}{k}\right)$$

$$\text{Varianz: } \text{Var}(X) = \lambda^{-2} \left(\Gamma\left(1 + \frac{2}{k}\right) - \Gamma^2\left(1 + \frac{1}{k}\right)\right)$$

$$\text{Standardabweichung: } \sigma = \sqrt{\text{Var}(X)}$$

g) Normalverteilung (bestimmter Wert wird erwartet; kann nach oben/unten abweichen)

$$\text{Dichtefunktion: } f(t) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot e^{-\frac{(t-\mu)^2}{2\sigma^2}}$$

$$E(X)=\mu; \text{ Var}(X)=\sigma^2$$

Eine Normalverteilung mit  $\mu = 0$  und  $\sigma = 1$  heißt Standardnormalverteilung und ihre Verteilungsfunktion  $\phi$

Berechnung mit Hilfe der Standardnormalverteilung ( $Z = \frac{X - \mu}{\sigma}$ ):

$$P(X \leq x) = \Phi\left(\frac{x - \mu}{\sigma}\right)$$

$$P(a \leq X \leq b) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right)$$

$$P(|X - \mu| \leq k\sigma) = 2\Phi(k) - 1$$

## 9. Zentraler Grenzwertsatz

$X_1, X_2, \dots, X_n$  unabhängige Zufallsvariablen

a)  $X_i$  normalverteilt mit  $E(X_i) = \mu_i$  und  $\text{Var}(X_i) = \sigma_i^2$ :

$S = X_1 + \dots + X_n$  ist normalverteilt mit  $E(S) = \sum_{i=1}^n \mu_i$  und  $\text{Var}(S) = \sum_{i=1}^n \sigma_i^2$

$\bar{X} = \frac{1}{n} (X_1 + \dots + X_n)$  ist normalverteilt mit

$$E(\bar{X}) = \frac{1}{n} \sum_{i=1}^n \mu_i \text{ und } \text{Var}(\bar{X}) = \frac{1}{n^2} \sum_{i=1}^n \sigma_i^2$$

b)  $X_i$  normalverteilt mit  $E(X_i) = \mu$  und  $\text{Var}(X_i) = \sigma^2$ :

$S = X_1 + \dots + X_n$  ist normalverteilt mit  $E(S) = n\mu$  und  $\text{Var}(S) = n\sigma^2$

$\bar{X} = \frac{1}{n} (X_1 + \dots + X_n)$  ist normalverteilt mit  $E(\bar{X}) = \mu$  und  $\text{Var}(\bar{X}) = \frac{1}{n} \sigma^2$

c) (Zentraler Grenzwertsatz)  $X_i$  identisch verteilt mit  $E(X_i) = \mu$  und  $\text{Var}(X_i) = \sigma^2$ :

$S = X_1 + \dots + X_n$  ist näherungsweise normalverteilt mit  $E(S) = n\mu$  und  $\text{Var}(S) = n\sigma^2$

$\bar{X} = \frac{1}{n} (X_1 + \dots + X_n)$  ist näherungsweise normalverteilt mit

$$E(\bar{X}) = \mu \text{ und } \text{Var}(\bar{X}) = \frac{1}{n} \sigma^2$$

Ab  $n \geq 30$  kann die Näherung als hinreichend genau angesehen werden.

## 10. Punktschätzer

a)  $T$  erwartungstreuer Schätzer für  $\lambda \iff E(T) = \lambda$

b)  $T$  zudem konsistent  $\iff \lim_{n \rightarrow \infty} \text{Var}(T) = 0$

## 11. Intervallschätzung

Im Folgenden steht

$z_\beta$  für das  $\beta$ -Quantil der Standardnormalverteilung

$t_{n-1;\beta}$  für das  $\beta$ -Quantil der tstudent-Verteilung mit Stichprobengröße  $n$

$\chi_{n-1;\beta}^2$  für das  $\beta$ -Quantil der Chiquadrat-Verteilung mit Stichprobengröße  $n$

a) Konfidenzintervall für den Erwartungswert einer Normalverteilung bei bekannter Varianz

zweiseitiges Konfidenzintervall  $\left[ \bar{x} - z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \right]$

einseitiges unteres Konfidenzintervall  $\left( -\infty, \bar{x} + z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}} \right]$

einseitiges oberes Konfidenzintervall  $\left[ \bar{x} - z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}, \infty \right)$

- b) Konfidenzintervall für den Erwartungswert einer Normalverteilung bei unbekannter Varianz

zweiseitiges Konfidenzintervall  $\left[ \bar{x} - t_{n-1; 1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}, \bar{x} + t_{n-1; 1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} \right]$

einseitiges unteres Konfidenzint.  $\left( -\infty, \bar{x} + t_{n-1; 1-\alpha} \cdot \frac{s}{\sqrt{n}} \right]$

einseitiges oberes Konfidenzint.  $\left[ \bar{x} - t_{n-1; 1-\alpha} \cdot \frac{s}{\sqrt{n}}, \infty \right)$

- c) Konfidenzintervall für die Varianz einer Normalverteilung

zweiseitiges Konfidenzintervall  $\left[ \frac{(n-1) \cdot s^2}{\chi_{n-1; 1-\frac{\alpha}{2}}^2}, \frac{(n-1) \cdot s^2}{\chi_{n-1; \frac{\alpha}{2}}^2} \right]$

einseitiges unteres Konfidenzintervall  $\left[ 0, \frac{(n-1) \cdot s^2}{\chi_{n-1; \alpha}^2} \right]$

einseitiges oberes Konfidenzintervall  $\left[ \frac{(n-1) \cdot s^2}{\chi_{n-1; 1-\alpha}^2}, \infty \right)$

## 12. Hypothesentest

Wähle ein Signifikanzniveau  $\alpha$  (z.B.: 0.1, 0.05, 0.01)

- a) **Binomialtest** bei Hypothese bzgl.  $p$  einer Binomialverteilung

- Hypothesen festlegen
  - i.  $H_0 : p = p_0$  gegen  $H_1 : p \neq p_0$
  - ii.  $H_0 : p \geq p_0$  gegen  $H_1 : p < p_0$  linksseitig
  - iii.  $H_0 : p \leq p_0$  gegen  $H_1 : p > p_0$  rechtsseitig
- Ziehe eine Stichprobe vom Umfang  $n$ , bestimme die Anzahl  $c$  mit der Eigenschaft
- Bestimme das entsprechende Quantil der Binomialverteilung:
  - i.  $k_u \in \mathbb{N}$  mit  $P(X \leq k_u) \leq \frac{\alpha}{2}$  und  $k_o \in \mathbb{N}$  mit  $P(X \geq k_o) \leq \frac{\alpha}{2}$
  - ii.  $k \in \mathbb{N}$  mit  $P(X \leq k) \leq \alpha$
  - iii.  $k \in \mathbb{N}$  mit  $P(X \geq k) \leq \alpha$
- Verwerfe  $H_0$ , falls
  - i.  $c \in \{0, \dots, k_u\} \cup \{k_o, \dots, n\}$
  - ii.  $c \in \{0, \dots, k\}$
  - iii.  $c \in \{k, \dots, n\}$

b) **Gaußtest** bei Hypothese bzgl.  $\mu$  einer Normalverteilung mit bekanntem  $\sigma$

- Hypothesen festlegen
  - i.  $H_0 : \mu = \mu_0$  gegen  $H_1 : \mu \neq \mu_0$
  - ii.  $H_0 : \mu \geq \mu_0$  gegen  $H_1 : \mu < \mu_0$  linksseitig
  - iii.  $H_0 : \mu \leq \mu_0$  gegen  $H_1 : \mu > \mu_0$  rechtsseitig
- Ziehe eine Stichprobe vom Umfang  $n$ , berechne  $\bar{x}$  und den dazugehörigen standardisierten Prüfwert
$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$
- Bestimme das entsprechende Quantil der Standardnormalverteilung:
  - i.  $z_{1-\alpha/2}$
  - ii.  $z_{1-\alpha}$
  - iii.  $z_{1-\alpha}$
- Verwerfe  $H_0$ , falls
  - i.  $|z| > z_{1-\alpha/2}$
  - ii.  $z < -z_{1-\alpha}$
  - iii.  $z > z_{1-\alpha}$

c) **t-Test** bei Hypothese bzgl.  $\mu$  einer Normalverteilung mit unbekanntem  $\sigma$

- Hypothesen festlegen
  - i.  $H_0 : \mu = \mu_0$  gegen  $H_1 : \mu \neq \mu_0$
  - ii.  $H_0 : \mu \geq \mu_0$  gegen  $H_1 : \mu < \mu_0$  linksseitig
  - iii.  $H_0 : \mu \leq \mu_0$  gegen  $H_1 : \mu > \mu_0$  rechtsseitig
- Ziehe eine Stichprobe vom Umfang  $n$ , berechne  $\bar{x}$  und  $s$  und den dazugehörigen standardisierten Prüfwert
$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$
- Bestimme das entsprechende Quantil der t-Verteilung:
  - i.  $t_{n-1;1-\alpha/2}$
  - ii.  $t_{n-1;1-\alpha}$
  - iii.  $t_{n-1;1-\alpha}$
- Verwerfe  $H_0$ , falls
  - i.  $|t| > t_{n-1;1-\alpha/2}$
  - ii.  $t < -t_{n-1;1-\alpha}$
  - iii.  $t > t_{n-1;1-\alpha}$

d)  **$\chi^2$ -Test** bei Hypothese bzgl.  $\sigma^2$  einer Normalverteilung

- Hypothesen festlegen
  - i.  $H_0 : \sigma^2 = \sigma_0^2$  gegen  $H_1 : \sigma^2 \neq \sigma_0^2$
  - ii.  $H_0 : \sigma^2 \geq \sigma_0^2$  gegen  $H_1 : \sigma^2 < \sigma_0^2$  linksseitig

- iii.  $H_0 : \sigma^2 \leq \sigma_0^2$  gegen  $H_1 : \sigma^2 > \sigma_0^2$  rechtsseitig
- Ziehe eine Stichprobe vom Umfang  $n$ , berechne  $s^2$  und den dazugehörigen standardisierten Prüfwert
 
$$y = (n - 1) \frac{s^2}{\sigma_0^2}$$
- Bestimme das entsprechende Quantil der  $\chi^2$ -Verteilung:
  - i.  $\chi_{n-1;\alpha/2}^2$  und  $\chi_{n-1;1-\alpha/2}^2$
  - ii.  $\chi_{n-1;\alpha}^2$
  - iii.  $\chi_{n-1;1-\alpha}^2$
- Verwerfe  $H_0$ , falls
  - i.  $y < \chi_{n-1;\alpha/2}^2$  oder  $y > \chi_{n-1;1-\alpha/2}^2$
  - ii.  $y < \chi_{n-1;\alpha}^2$
  - iii.  $y > \chi_{n-1;1-\alpha}^2$