Beschreibende Statistik

Lageparameter

Arithmetisches Mittel

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}(x_1 + \cdots + x_n)$$

Das Arithmetische Mittel $\overline{x}$ minimiert die Funktion

$$g(t) = \sum_{i=1}^{n}(x_i - t)^2$$

Geometrisches Mittel

$$\overline{x}_{geom} = \sqrt[n]{x_1 \cdot x_2 \cdot \cdots \cdot x_n}$$

Median

$$\tilde{x} = \begin{cases} x_{\frac{n+1}{2}} & , ungerade \\ \frac{1}{2}\cdot(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}) & , gerade \end{cases}$$

Der Median $\tilde{x}$ minimiert die Funktion

$$g(t) = \sum_{i=1}^{n}|x_i - t|$$

Streungsmaße

(empirische) Varianz

$$var = \sigma^2 = {s_n}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

alternativ

$$var = \sigma^2 = \frac{n}{n-1}\cdot(\overline{x^2} - \overline{x}^2)$$

Standardabweichung

$$\sigma = s_n = \sqrt{\sigma^2}$$
$$\sigma = s_n = \sqrt{{s_n}^2}$$

mittlere absolute Abweichung

$$\frac{1}{n}\sum_{i=1}^{n}|x_i - \tilde{x}| \text{ für Median}$$

$$\frac{1}{n}\sum_{i=1}^{n}|x_i - \overline{x}| \text{ für arithmetisches Mittel}$$

Kovarianz und Korrelationskoeffizient

Kovarianz

$$cov(x,y) = S_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})\cdot(y_i - \overline{y})$$

alternativ

$$cov(x,y) = S_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i \cdot y_i - n \cdot \overline{x}\cdot\overline{y})$$

Korrelationskoeffizent

$$r_{xy} = \frac{S_{xy}}{S_x \cdot S_y}$$

Der Korrelationskoeffizent liegt immer zwischen $-1 \le r \le +1$. Je näher r_{xy} bei -1 (negative Korellation/Steigung), oder $+1$ (positive Steigung/Korrelation) liegt, desto genauer schmiegen sich die Messwerte an eine Gerade an. Bei r_{xy} nahe 0 gibt es keinen *linearen* Zusammenhang zwischen den Merkmalen.

Regressionsrechnung

Regressionsgerade

Variante 1

$$y = \overline{y} + \frac{S_{xy}}{{\sigma_x}^2}\cdot(x - \overline{x})$$

Variante 2

$$y = b + a \cdot x$$

$$a = \frac{S_{xy}}{{\sigma_x}^2} \text{ und } b = \overline{y} - a \cdot \overline{x}$$

Kleinste quadratische Abweichung

Die Parameter $a, b, c, ...$ werden so gewählt, dass

$$Q(a,b,c,...) = \sum_{i=1}^{n}(f_{a,b,c,...}(x_i) - y_i)^2$$

minimal ist

$f_{a,b,c...}(x_i)$ ist die Funktion dessen Parameter gesucht werden

Nullsetzen der partiellen Ableitungen:

$$\frac{\partial}{\partial a}Q(a,b) = 0$$
$$\frac{\partial}{\partial b}Q(a,b) = 0$$

Über die Ableitungen lassen sich die Parameter finden welche die vorgegebene Funktion am besten annähern

Vergleich ermittelter Kurven

Um Kurven zu vergleichen, einfach die ermittelten Parameter in die $Q(a,b,c,...)$ Funktion eingeben und Wert berechnen. Je kleiner der Wert desto besser passt die Kurve

Wahrscheinlichkeitstheorie

Wahrscheinlichkeitsräume

Der Wahrscheinlichkeitsbegriff

$$Ergebnismenge = \Omega$$

Beispiel Würfel $\Omega = \{1,2,3,4,5,6\}$

Ein Ereignis ist eine Teilmenge der Ergebnismenge

$\varnothing \subseteq \Omega \widehat{=}$ unmögliches Ereignis

$\Omega \subseteq \Omega \widehat{=}$ sicheres Ereignis

$A = \{1,2,3\}$ Ereignis

$\overline{A} = \{4,5,6\}$ Gegenereignis

Elementarereignis

einelementige Teilmenge von Ω

Ereignis, eine 3 werfen

$$B = \{3\}$$
$$P(\{3\}) = \frac{1}{6}$$

Laplace-Versuch

Jedes Elementarereignis ist gleich wahrscheinlich

$$P(\{\omega_i\}) = \frac{1}{|\Omega|}$$

$$P(A) = \frac{|A|}{|\Omega|} = \frac{3}{6} = \frac{1}{2}$$

Bedingte Wahrscheinlichkeit

Bedingte Wahrscheinlichkeit

Wahrscheinlichkeit für A unter der Bedingung B

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$P(\overline{A}|B) = 1 - P(A|B)$$

Formel von Bayes

$$P(A|B) = \frac{P(B|A)\cdot P(A)}{P(B)}$$

Mit ersetzung von $P(B)$ durch den Satz der totalen Wahrscheinlichkeit:

$$P(A|B) = \frac{P(B|A)\cdot P(A)}{P(B|A)\cdot P(A) + P(B|\overline{A})\cdot P(\overline{A})}$$

Satz der totalen Wahrscheinlichkeit

$$P(A) = \sum_{i=1}^{n}(P(A|B_i)\cdot P(B_i))$$

$$P(A) = P(A|B)\cdot P(B) + P(A|\overline{B})\cdot P(\overline{B})$$

Viel Felder Tafel

	A	$\overline{A}$	$\sum$
B	$P(A \cap B)$	$P(\overline{A} \cap B)$	$P(B)$
$\overline{B}$	$P(A \cap \overline{B})$	$P(\overline{A} \cap \overline{B})$	$P(\overline{B})$
$\sum$	$P(A)$	$P(\overline{A})$	1

Die Ränder sind immer die Summen der zugehörigen Zeilen oder Spalten

Allgemeine Regeln

$$P(A \cap B) = P(A|B)\cdot P(B) = P(B|A)\cdot P(A)$$
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
$$P(\overline{A}) = 1 - P(A)$$
$$P(\overline{A \cup B}) = P(\overline{A} \cap \overline{B})$$
$$P(\overline{A \cap B}) = P(\overline{A} \cup \overline{B})$$

Wenn A und B unabhängig, dann gilt

$$P(A \cap B) = P(A)\cdot P(B)$$
$$P(A|B) = P(A)$$

Zufallsvariablen

Eine Zufallsvariable ist eine Zuordnungsvorschrift die jedem möglichen Ergebnis eines Zufallsexperiments eine Größe zuordnet

$$X = k \,\widehat{=}\, \{\omega \in \Omega | X(\omega = k\}$$
$$X = 3 \,\widehat{=}\, \{\omega \in \Omega | X(\omega = 3\}$$
$$X \le k \,\widehat{=}\, \{\omega \in \Omega | X(\omega \le k\}$$

Diskrete Verteilungen

Binomialverteilung

Mit zurücklegen, Wahrscheinlichkeit für jedes Ereignis gleich

$$X \sim B(n,p)$$

$n =:$ Stichprobenumfang

$p =:$ Wahrscheinlichkeit

(p muss bei Binomialverteilung fest bleiben)

$$P(X = k) = \binom{n}{k}\cdot p^k \cdot (1-p)^{n-k}$$

$$P(X \le k) = \sum_{i=0}^{k}\binom{n}{i}\cdot p^i \cdot (1-p)^{n-i}$$

$$P(X > k) = 1 - P(X \le k)$$

Eingabe Taschenrechner

$$\binom{n}{k} \,\widehat{=}\, n\,|nCr|\,k$$

$$Mode \rightarrow 4 \rightarrow \downarrow \rightarrow 1 \rightarrow 2 \rightarrow k \rightarrow n \rightarrow p$$

Binomialverteilung approximieren

Die **Binomialverteilung** kann mit der **Poisson** Verteilung approximiert werden, dann gilt

$$\lambda = n \cdot p$$

Die **Binomialverteilung** kann auch mit der **Normalverteilung** approximiert werden, wobei gilt $n \cdot p = \mu$ und $n \cdot p \cdot (1-p) = \sigma^2$, **bedingung ist**

$$X \sim B(n,p) \approx N(n \cdot p, n \cdot p \cdot (1-p))$$

falls gilt

$$n \cdot p \cdot (1-p) > 9$$

Bei der approximation mit der Normalverteilung kann man eine **Stetigkeitskorrektur** verwenden um ein besseres Ergebnis zu erhalten

$$P(X \le k) \approx F_N(R + 0,5)$$
$$P(X < k) \approx F_N(R - 0,5)$$
$$P(a \le X \le b) \approx F_N(b + 0,5) - F_N(a - 0,5)$$

Zusammengefasst

Falls np und $n(1-p)$ groß genug sind:

$\mu = n \cdot p$ und $\sigma^2 = n \cdot p \cdot (1-p)$

$F_B(x) \approx F_N(x+0.5) = \Phi(\frac{x+0.5-np}{\sqrt{np(1-p)}})$

Hypergeometrische Verteilung

Ohne zurücklegen, Wahrscheinlichkeit ändert sich nach jedem Ereignis

$X \sim H(N, M, n)$

$n =$: Stichprobenumfang

$N =$: Gesamtzahl

$M =$: Anzahl der Elemente mit der Eigenschaft

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

$$P(X \leq k) = \sum_{i=0}^{k} \frac{\binom{M}{i} \cdot \binom{N-M}{n-i}}{\binom{N}{n}}$$

$$P(X > k) = 1 - P(X \leq k)$$

Hypergeometrische Vert. approximieren

Die hypergeometrische Verteilung kann mit der **Binomialverteilung** approximiert werden. Dabei muss folgende Bedingung gelten

$\frac{n}{N} < 0,05$

Poisson Verteilung

Schlüsselwörter sind **Ereignisse pro Zeiteinheit**, zum Beispiel Anrufe innerhalb bestimmter Zeitspanne oder Durchschnitt Ereignisse x pro y (z.B.: Staubkörner pro cm^2)

$X \sim Pois(\lambda)$

$$P(X = k) = \pi_\lambda(k) = \frac{\lambda^k}{k!} \cdot e^{-\lambda}$$

$$P(X \leq k) = \sum_{i=1}^{k} \frac{\lambda^i}{i!} \cdot e^{-\lambda}$$

$$P(X \geq k) = 1 - P(X < k)$$

Eingabe Taschenrechner

$Mode \rightarrow 4 \rightarrow \downarrow \rightarrow 3 \rightarrow 2 \rightarrow k \rightarrow \lambda$

Nährung an Normalverteilung

Wenn λ groß genug ist kann die Verteilungsfunktion $F_P(x)$ der Poissonverteilung durch die Verteilungsfunktion der Normalverteilung $F_N(x)$ mit den Parametern $\mu = \lambda$ und $\sigma^2 = \lambda$ genähert werden:

$$F_P(x) \approx F_N(x+0.5) = \Phi(\frac{x+0.5-\lambda}{sqrt\lambda})$$

Geometrische Verteilung

$X \sim Geom(n, p)$

$P(X = n) = (1-p)^{n-1} \cdot p$

Beispiel: Ein Würfel wird so lange gewürfelt bis eine 6 Auftritt. Die Zufallsvariable X ist gleich Anzahl der Würfe

Stetige Verteilungen

Dichtefunktion

Die Dichtefunktion ist ein Hilfsmittel zur **Beschreibung einer stetigen Wahrscheinlichkeitsverteilung**

Bedingungen der Dichtefunkion

$f(x) \geq 0$

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

Die Dichtefunktion muss **nicht** stetig sein

Die Dichtefunktion ist die Ableitung der Verteilungsfunktion $F(x)$

Verteilungsfunktion

Eine Verteilungsfunktion ist eine Funktion F, die jedem x einer Zufallsvariable X genau eine Wahrscheinlichkeit $P(X \leq x)$ zuordnet

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

$P(a \leq X \leq b) = F(b) - F(a)$

Bedingungen der Verteilungsfunktion

Die Verteilungsfunktion **muss** stetig sein

Die Verteilungsfunktion **muss** monoton steigend sein

$\lim_{x \to \infty} F(x) = 1$

$\lim_{x \to -\infty} F(x) = 0$

Normalverteilung

$X \sim N(\mu, \sigma^2)$

Ist $X \sim N(0,1)$ dann heißt sie Standardnormalverteilt

Jede Normalverteilung kann standardisiert werden, das heißt die Mitte der Kurve wird auf den Nullpunkt gesetzt

Wenn $X \sim N(\mu, \sigma^2)$ verteilt ist dann ist die standardisierte Zufallsvariable $Z = \frac{x-\mu}{\sigma} \sim N(0,1)$ standardnormalverteilt

Ist die Zufallsvariable standardverteilt kann die Wahrscheinlichkeit aus der Tabelle abgelesen werden

Regeln für den *Phi*-Wert

$P(X \leq k) = \Phi(k)$

$P(X \leq -k) = 1 - \Phi(+k)$

$P(X = k) = 0$ ("Integral ohne Breite!")

allgemein folgt daraus, wenn

$X \sim N(\mu, \sigma^2)$

dann gilt

$P(X \leq k) = \Phi(\frac{k-\mu}{\sigma})$

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

Additionssatz der Normalverteilung

Seien X und Y unabhängig und Normalverteilt, dann gilt:

$X + Y = N(\mu_X + \mu_Y; \sigma_X^2 + \sigma_Y^2)$

Ihre Summe ist ebenfalls Normalverteilt!

Quantile der Normalverteilung

Tabelliert ist das $\beta-$Quantil z_β der Normalverteilung $N(0,1)$

$P(X \leq z_\beta) = \beta$

$z_{1-\beta} = -z_\beta$

Beispiel

$\beta = 0.9 => z_\beta = 1.28155$

Exponentialverteilung

$X \sim Exp(\lambda)$

Eine exponentialverteilte Zufallsvariable T hat die Dichte

$$f(t) = \begin{cases} \lambda \cdot e^{-\lambda \cdot t} & , t \geq 0 \\ 0 & , t < 0 \end{cases}$$

und daraus eribt sich die Verteilungsfunktion

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

$$= \int_{-\infty}^{x} f(t)dt = \begin{cases} 1 - e^{-\lambda \cdot x} & , x \geq 0 \\ 0 & , x < 0 \end{cases}$$

Die Exponentialverteilung ist Gedächtnislos

Gleichverteilung(Rechteckverteilung)

$$f(t) = \begin{cases} \frac{1}{b-a} & , t \in [a, b] \\ 0 & , sonst \end{cases}$$

$$F(t) = \begin{cases} 0 & , t < a \\ \frac{t-a}{b-a} & , t \in [a, b] \\ 1 & , t > b \end{cases}$$

Erwartungswert und Varianz

Erwartungswert

Erwartungswert und Mittelwert sind prinzipiell gleichwertig, der Erwartungswert entspricht der theoretischen Erwartung, der Mittelwert entspricht den tatsächlichen Werten

Zufallsvariable mit diskreter Verteilung

$$\mu = E(X) = \sum_{i=0}^{n} (x_i \cdot p_i)$$

Zufallsvariable mit Dichtefunktion f

$$\mu = E(X) = \int_{-\infty}^{\infty} x \cdot f(x)dx$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 \cdot f(x)dx$$

Exponentialverteilung mit Zufallsvariable T

$E(T) = \sigma_T = \frac{1}{\lambda}$

Für Binomialverteilung

$\mu = E(X) = n \cdot p$

Für geometrische Verteilung

$\mu = E(X) = \frac{1}{p}$

Für Poissonverteilung

$\mu = E(X) = \lambda$

Für Hypergeometrischeverteilung

$E(S_n) = E(X_1 + ... + X_n) = n \cdot E(X_1) = n \cdot \frac{M}{N}$

Für Gleichverteilung(Rechteckverteilung)

$E(T_i) = \frac{a+b}{2}$

Allgemeine Regeln für den Erwartungswert

$a, b \in \mathbb{R}$

$E(aX + b) = a \cdot E(X) + b$

$E(X + Y) = E(X) + E(Y)$

$E(aX + bY) = a \cdot E(X) + b \cdot E(Y)$

Varianz

Zufallsvariable mit diskreter Verteilung

$\sigma^2 = Var(X) = \sum (x_i - \mu)^2 \cdot p_i$

Zufallsvariable mit Dichtefunktion f

$Var(X) = E(X^2) - (E(X))^2$

Varianz aus Erwartungswert berechnen

$\sigma^2 = E(X^2) - \mu^2$

Exponentialverteilung mit Zufallsvariable T

$Var(T) = \frac{1}{\lambda^2}$

Für Binomialverteilung

$$\sigma^2 = n \cdot p \cdot (1-p)$$

Für geometrische Verteilung

$$\sigma^2 = \frac{1}{p^2} - \frac{1}{p}$$

Für Poissonverteilung

$$\sigma^2 = Var(X) = E(X^2) - E(X)^2 = \lambda$$

Für Hypergeometrischeverteilung

$$Var(S_n) = n \cdot \frac{M}{N} \cdot (1 - \frac{M}{N}) \cdot \frac{N-n}{N-1}$$

Für Gleichverteilung(Rechteckverteilung)

$$Var(T_i) = \frac{(b-a)^2}{12}$$

Allgemeine Regeln für Varianz

$$Var(X+b) = Var(X)$$
$$Var(aX+b) = a^2 \cdot Var(X)$$
$$Var(X) \geq 0$$
$$Var(X+Y) = Var(X) + Var(Y) + 2 \cdot cov(X,Y)$$

wobei gilt:

$$Cov(X,Y) = E((X-\mu_X)(Y-\mu_Y)) = E(X \cdot Y) - \mu_X \mu_Y$$

bei unabhängigen Zufallsvariablen X und Y ist $Cov(X,Y) = 0$ siehe unten.

Unabhängige Zufallsvariablen

Allgemeine Regeln

$$Var(X + const) = Var(X)$$
$$Var(X+Y) = Var(X) + Var(Y)$$
$$E(X \cdot Y) = E(X) \cdot E(Y)$$

Wichtige Sätze der Stochastik

Zentraler Grenzwertsatz

n groß (Anzahl der Zufallsvariablen) $n \geq 30$

X_i unabhängig und identisch verteilt $\hat{=}$ haben die gleiche Verteilung

$$E(X_i) = \mu$$
$$Var(X_i) = \sigma^2$$

Fall a) bei Summierten Ereignisse, Fall b) für Durchschnitt

a) $\sum X_i \sim N(n \cdot \mu, n \cdot \sigma^2)$

b) $\overline{X}_n = \frac{X_1 + ... + X_n}{n} = \overline{x} \sim N(\mu, \frac{\sigma^2}{n})$

Manche Verteilungen verhalten sich in der Summe anders, zum Beispiel die Rechteckverteilung ist nicht mehr R-Verteilt. Dann wird der Zentrale Grenzwertsatz verwendet

ZGS Algorithmus

n : Anzahl der Zufallsvariablen (Stichproben) bestimmen,

σ^2: Varianz bestimmen,

μ: Erwartungswert (Mittelwert) bestimmen

Wonach wird gesucht? Summe oder Durchschnitt

1. **Summe**:
   Erwartungswert der Summe bestimmen mit $E(S) = n \cdot \mu$ alternativ und gleichwertig ist: $E(S) = \sum_{i=1}^{n} E(X_i)$
   Falls notwendig Varianz aus Erwartungswert berechnen: $Var(X) = E(X^2) - E(X)^2$
   Wahrscheinlichkeit normalverteilt berechnen mit $N(n \cdot \mu, n \cdot \sigma^2)$
2. **Durchschnitt**:
   Durchschnittserwartungswert bestimmen entwder gegeben oder $\frac{1}{n} \cdot \sum_{i=1}^{n} E(X_i)$
   Falls notwendig (nicht gegeben) Varianz berechnen.
   Wahrscheinlichkeit normalverteilt berechnen mit $N(\mu, \frac{\sigma^2}{n})$

Hinweis Normalverteilung berechnen:

$$P(X \leq k) = \Phi(\frac{k-\mu}{\sigma})$$

ZGS - Definition

Seien $X_1, ..., X_n$ unabhängige und identisch verteilte Zufallsvariablen (nicht zwangsläufig Normalverteilt) mit Erwarungswert μ und Varianz σ^2. Ihre Summe sei $S = X_1 + ... + X_n$ mit Erwarungswert $n\mu$ und Varianz $n\sigma^2$. Es gilt für die zugehörige Zufallsvariable

$$Z = \frac{S - n\mu}{\sqrt{n}\sigma} = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \text{ gilt}$$

$$\lim_{n \to \infty} P(Z \leq z) = \Phi(z)$$

Induktive Statistik - Schätztheorie

Schätzfunktionen

Maximum-Likelihood-Schätzer

Vorbereitung: Falls $F(x)$ gegen leite ab um $f(x)$ zu erhalten ($f(x_i)$ muss eine **Dichtefunktion** sein)

1. Stelle auf: $L(x_1, \ldots, x_n, \alpha) = \prod_{i=1}^{n} f(x_i)$

Tipp: Wenn möglich alles was nicht von x_i abhängt aus der Summenfunktion ziehen.

2. Wende an: $\ln(L(x_1, \ldots, x_n, \alpha))$
3. Die Funktion nach dem Parameter α ableiten und Nullsetzen: $\frac{\partial \ln L(x_1, \ldots, x_n, \alpha)}{\partial \alpha} = 0$
4. Nach α auflösen, Das Ergebnis ist der Maximum-Likelihood-Schätzer

Konfidenzintervalle

Achtung die Standartabweichung σ bezieht sich im Folgenden auf die Grundgesammtheit! Es kann also vorkommen, das eine Standartabweichung gegeben ist, dann ist allerdings noch zu Prüfen ob es sich um die der Stichprobe oder der Grundgesammtheit handelt. Mit α in den Intervallformeln ist immer die Irrtumswahrscheinlichkeit gemeint.

Intervall für $\boldsymbol{E(X)}$ einer Normalverteilung

Wenn nach Schätzintervall des **Erwartungswerts** gefragt ist. Prüfen ob Sigma bekannt oder nicht und danach Formel wählen.

Ist $X \sim N(\mu, \sigma^2)$ verteilt, dann ist

$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$$

Bei **bekannter** Standardabweichung σ

$$\left[\overline{x} - z_{1-(\alpha/2)} \cdot \frac{\sigma}{\sqrt{n}}, \overline{x} + z_{1-(\alpha/2)} \cdot \frac{\sigma}{\sqrt{n}}\right]$$

$\overline{X}$ = arithmetisches Mittel, bzw erwartungsstreuer Schätzer bei n unabhängigen Stichproben

α = Signifikanzwahrscheinlichkeit (Irrtumswahrscheinlichkeit) oder $\alpha = 1-$Konfidenzniveau

alternativ: $1 - \alpha$ = Konfidenzniveau (Vertrauensniveau)

Ist α bekannt, berechne das Quantil $z_{1-(\alpha/2)}$

Bei **unbekannter** Standardabweichung σ

$$\left[\overline{x} - t_{n-1;1-(\alpha/2)} \cdot \frac{s}{\sqrt{n}}, \overline{x} + t_{n-1;1-(\alpha/2)} \cdot \frac{s}{\sqrt{n}}\right]$$

Anstatt σ^2 wird der erwartungstreue Schätzer s verwendet

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$$

beziehungsweise

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2}$$

Intervall für $\boldsymbol{Var(X)}$ einer Normalverteilung

Wenn nach Schätzintervall der **Varianz** gefragt ist.

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

Induktive Statistik - Hypothesentest

Achtung die Standartabweichung σ bezieht sich im Folgenden auf die Grundgesammtheit! Es kann also vorkommen, das eine Standartabweichung gegeben ist, dann ist allerdings noch zu Prüfen ob es sich um die der Stichprobe oder der Grundgesammtheit handelt.

Tests für Lageparameter

Wähle Fall a) wenn Verteilung statt absoluter Wert gegeben (z.b. "jeder n-te...") und verfolge Stragie (3).

Wähle Fall b) wenn $\overline{X} \geq \mu_0$

Wähle Fall c) wenn $\overline{X} \leq \mu_0$

(1) Gauß-Test (Wählen wenn σ^2 bekannt)

Ist ein Test für den Erwartungswert einer *Normalverteilung* bei bekannter Standardabweichung σ

Wähle eine mögliche Hypothesenkombination:

a) $H_0 : \mu = \mu_0$

b) $H_0 : \mu \leq \mu_0$

c) $H_0 : \mu \geq \mu_0$

Wähle ein Signifikanzniveau α (z.B.: 0.05)

Ziehe eine Stichprobe vom Umfang n, berechne $\overline{x}$ und den zugehörigen standardisierten Prüfwert:

$$z = \sqrt{n} \cdot \frac{\overline{x} - \mu_0}{\sigma}$$

Bestimme das entsprechende Quantil der Standartnormalverteilung:

a) $z_{1-\frac{\alpha}{2}}$ bzw. b) $z_{1-\alpha}$ c) $-z_{1-\alpha}$

H_0 ist zu verwerfen, falls

a) $|z| > z_{1-\frac{\alpha}{2}}$

b) $z > z_{1-\alpha}$
c) $z < -z_{1-\alpha}$

Gauß-Test Tabellenform

	H_0	Ablehnungsbereich
	$\mu = \mu_0$	$\lvert T\rvert > Z_{1-\frac{\alpha}{2}}$
$\overline{X} \geq \mu_0$	$\mu \leq \mu_0$	$T > Z_{1-\alpha}$
$\overline{X} \leq \mu_0$	$\mu \geq \mu_0$	$T < -Z_{1-\alpha}$

(2) t-Test (Wählen wenn σ^2 **nicht** bekannt)

Wähle eine mögliche Hypothesenkombination:
a) $H_0 : \mu = \mu_0$
b) $H_0 : \mu \leq \mu_0$
c) $H_0 : \mu \geq \mu_0$

Wähle ein Signifikanzniveau α (z.B.: 0.05)
Ziehe eine Stichprobe vom Umfang n, berechne daraus $\overline{x}$ und s sowie den zugehörigen Prüfwert:

$$t = \sqrt{n} \cdot \frac{\overline{x}-\mu_0}{s}$$

Bestimme das entsprechende Quantil der t-Verteilung, wobei der n-Wert der Tabelle $= (n-1)$ ist, also die Stückzahl -1 ist:
a) $t_{n-1;1-\frac{\alpha}{2}}$ bzw. b) $t_{n-1;1-\alpha}$ c) $-t_{n-1;1-\alpha}$

H_0 ist zu verwerfen, falls
a) $|t| > t_{n-1;1-\frac{\alpha}{2}}$
b) $t > t_{n-1;1-\alpha}$
c) $t < -t_{n-1;1-\alpha}$

t-Test Tabellenform

	H_0	Ablehnungsbereich	
	$\mu = \mu_0$	$\lvert T\rvert > t_{1-\frac{\alpha}{2}}$	$(n-1)$
$\overline{X} \geq \mu_0$	$\mu \leq \mu_0$	$T > t_{1-\alpha}$	$(n-1)$
$\overline{X} \leq \mu_0$	$\mu \geq \mu_0$	$T < -t_{1-\alpha}$	$(n-1)$

(3) Wahrscheinlichkeit gegeben

$$X_i = \begin{cases} 1, & i\text{te Stichprobe hat die Eigeschaft} \\ 0, & \text{sonst} \end{cases}$$

$$X_i \sim B(1, p_0)$$
$$\Rightarrow \text{ZGS} \Rightarrow \sum X_i \overset{a}{\sim} N(np_0|np_0(1-p_0)$$
$$T = \frac{X - np_0}{\sqrt{np_0(1-p_0)}}$$

Verwerfe oder nicht nach Regeln in Variante (1)

Tests für Streuungsmaße

χ^2 - Anpassungstest

Der χ^2-Anpassungstest überprüft ob eine unbekannte Wahrscheinlichkeitsverteilung einem bestimmten Verteilungsmodell folgt

$$T = \frac{1}{n}\left(\sum_{i=1}^{r} \frac{{N_i}^2}{p_i}\right) - n$$

wobei N_i absolute Häufigkeit in Kategorie r, n ist der Stichprobenumfang. Alternativ

$$T = \sum_{i=1}^{r} \frac{(N_i - np_i)^2}{n \cdot p_i}$$

Das Ergebnis T mit $\chi^2_{r-1;1-\alpha}$ Wert aus der Tabelle vergleichen

$T < \chi^2 =$ Hypothese wird nicht verworfen

Algorithmus

1. Hypothese aufstellen
2. n und r festlegen ($n =$ Stichprobenumfang, $r =$ Anzahl der Klassen)
3. Verteilung auf welche getestet werden soll bestimmten
4. Alle p_i berechenen (zu jeder Kategorie r eines nach in 3. festgelegter Verteilung)
5. T berechnen.
6. χ^2 mit T vergleichen.

Allgemeine Matheregeln

Potenzen und Logarithmen

Potenzgesetze

$a^0 = 1$
$a^1 = a$
$a^m \cdot a^n = a^{m+n}$
$(a^n)^m = a^{n \cdot m}$
$a^n \cdot b^n = (a \cdot b)^n$
$\frac{a^n}{a^m} = a^{n-m}$
$\prod_{i=1}^{n} a^{x_i} = a^{\sum_{i=1}^{n} x_i}$
$\sqrt[n]{k} = k^{\frac{1}{n}}$

Logarithmusregeln

$x = \log_a y \Leftrightarrow y = a^x$
$\log 1 = 0$
$\log x \cdot y = \log x + \log y$
$-\log x = \log \frac{1}{x}$
$\log \frac{x}{y} = \log x - \log y$
$\log x^n = n \cdot \log x$
$\log_a x = \frac{\log x}{\log a}$
$\log\left(\prod_{i=1}^{n} x_i\right) = \sum_{i=1}^{n} \log x_i$
$ln(e^x) = x$

Ableitungen und Integrale

Grundlegende Ableitungsregeln

$f(x)$	$f'(x)$
$c = const$	0
x^n	$n \cdot x^{n-1}$
$\sqrt{x}$	$\frac{1}{2\sqrt{x}}$
e^x	e^x
a^x	$\ln a \cdot a^x$
$\ln x$	$\frac{1}{x}$
$\log_a x$	$\frac{1}{\ln a \cdot x}$
$\sin x$	$\cos x$
$\cos x$	$-\sin x$
$\tan x$	$\frac{1}{\cos^2 x}$
$\cot x$	$\frac{1}{\sin^2 x}$
$x \pm c$	1

Verknüpfte Ableitungsregeln

$f(x)$	$f'(x)$
$(f(x) + g(x))$	$(f'(x) + g'(x))$
$(f(x) \cdot g(x))$	$(f'(x) \cdot g(x)) + (f(x) \cdot g'(x))$
$\frac{f(x)}{g(x)}$	$\frac{(f'(x) \cdot g(x)) - (f(x) \cdot g'(x))}{g(x)^2}$
$f(g(x))$	$f'(g(x)) \cdot g'(x)$

wichtige Stammfunktionen

$f(x)$	$F(x)$
$x^n, n \neq 1$	$\frac{1}{n+1} \cdot x^{n+1} + c$
c	$cx + c$
$\frac{1}{x}, x \neq 0$	$\ln\lvert x\rvert + c$
$\sqrt{x}$	$\frac{2}{3} \cdot x^{\frac{3}{2}} + c$
e^x	$e^x + c$

Bestimmte Integrale

$$\int_a^b f(x)dx = [F(x) + C]_a^b = F(b) - F(a)$$

$$\int_a^b a \cdot f(x)dx = a \cdot \int_a^b f(x)dx$$

Summen und Produkte

$\prod_{i=1}^{n} a \cdot x_i = a^n \prod_{i=1}^{n} x_i$
$\sum_{i=1}^{n} a \cdot x_i = a \cdot \sum_{i=1}^{n} x_i$

Sonstiges

$\frac{b}{\alpha} = c \Leftrightarrow \alpha = \frac{b}{c}$
$\sum_{i=1}^{n} -x_i = -\sum_{i=1}^{n} x_i$
$ln(a \cdot \prod_{i=1}^{n} x_i) = ln(a) + \sum_{i=1}^{n} x_i$
$\frac{a}{b}/c = \frac{a}{b} \cdot \frac{1}{c}$

Beispiele

Zentraler Grenzwertsatz

Angabe Seien $T_1, ..., T_{102}$ unabhängige Zufallsvariablen, die alle exponentialverteilt sind mit Rate $\lambda = 3$. Berechnen Sie Mittelwert und Standartabweichung von $S := T_1 + ... + T_{102}$ und $\overline{T} := \frac{1}{102}S$. Berechnen Sie die Wahrscheinlichkeiten $P(S > 40)$ und $P(0,3 < \overline{T} < 0.4$

Lösung Fst $Y \sim Exp(\lambda)$ dann $E(Y) = \frac{1}{\lambda}$ und $Var(Y) = \frac{1}{\lambda^2}$.
Also da $T_i \sim Exp(3)$: $E(T_i) = \frac{1}{3}$ und $Var(T_i) = \frac{1}{9}$.
Da

1. T_i unabhängig und identisch verteilt sind mit $E(T_i) = \frac{1}{3}$ und $Var(T_i) = \frac{1}{9}$
2. $n = 102 \geq 30$

$\overset{ZGS}{\Rightarrow}$

1) $S = \sum_{i=1}^{102} T_i \sim N(\overbrace{102 \cdot \frac{1}{3}}^{\mu_S} | \overbrace{102 \cdot \frac{1}{9}}^{\sigma_S^2})$

2) $\overline{T} := \frac{1}{102} \cdot \sum_{i=1}^{102} T_i \sim N(\frac{1}{3} | \frac{1}{9}/102)$

also $S \sim N(34|\frac{34}{3})$
Mittelwert von S: 34
Standardabweichung von S: $\sqrt{\frac{34}{3}}$
also $\overline{T} \sim N(\frac{1}{3}|\frac{1}{918})$
Mittelwert von $\overline{T}$: $\frac{1}{3}$
Standardabweichung von $\overline{T}$: $\sqrt{\frac{1}{918}}$

$$\begin{aligned} P(S > 40) &= 1 - P(S \leq 40) \\ &= 1 - \Phi\left(\frac{40 - 34}{\sqrt{\frac{34}{3}}}\right) \\ &= 1 - 0,962 \\ &= 0,038 \end{aligned}$$

$$\begin{aligned} &P(0,3 < \overline{T} < 0,4) \\ &= \Phi\left(\frac{0,4 - \frac{1}{3}}{\sqrt{\frac{1}{918}}}\right) - \Phi\left(\frac{0,3 - \frac{1}{3}}{\sqrt{\frac{1}{918}}}\right) \\ &= \Phi(2,02) - \Phi(-1,01) \\ &= \Phi(2,02) - (1 - \Phi(1,01)) \\ &= \Phi(2,02) - 1 + \Phi(1,01) \\ &= 0,978 - 1 + 0,844 = 0,822 \end{aligned}$$