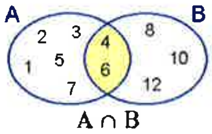
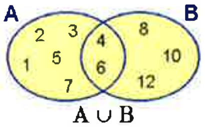
Statistik





Mengen

A ∩ B -> Schnittmenge

A ∪ B -> Vereinigungsmenge

Wahrscheinlichkeitsrechnung

LAPLACE – Wahrscheinlichkeit

$P(A) = \frac{|A|}{|\Omega|} = \frac{\textit{Anzahl der für A günstigen Fälle}}{\textit{Anzahl aller möglichen Fälle}}$ Bedingung: gleiche Wahrscheinlichkeit aller mögl. Fälle

Urnenmodell

$$\binom{n}{k} = \frac{n!}{k! * (n-k)!}$$

	ohne Wiederholung	mit Wiederholung	
Kombination k-ter Ordnung	$C(n;k) = \binom{n}{k}$	$C_W(n;k) = \binom{n+k-1}{k}$	ungeordnete Stichprobe
Variation k-ter Ordnung	$V(n;k) = \frac{n!}{(n-k)!}$	$V_w(n;k) = n^k$	geordnete Stichprobe
	Ziehung ohne Zurücklegen	Ziehung mit Zurücklegen	

Bedingte Wahrscheinlichkeit

B hängt von A ab: $P(B|A) = \frac{P(B \cap A)}{P(A)}$

Zufallsvariablen

Diskrete Zufallsvariablen

f(x) = P(X=x)

Verteilungsfunktion $F(x) = \sum_{i=1}^{n} f(x_i)$ → kumulierte Wahrscheinlichkeit → Treppe

Stetige Zufallsvariablen

Dichtefunktion $F(x) = \int_{-\infty}^{+\infty} f(t)dt$

Diskrete Verteilungen

Binomialverteilung

n-faches Ziehen mit Zurücklegen, p = Erfolgschance pro Ziehen

X ~ B(n,p) $f(x) = \binom{n}{x} * p^x * (1-p)^{n-x}$

$$f(x) = F(x) - F(x-1)$$

$p > 0{,}5$: $P(X \leq x) = 1 - F(n-x-1)$

Hypergeometrische Verteilung

n-faches Ziehen ohne Zurücklegen aus N Objekten, davon M markiert.

X ~ Hyp (N,M,n) $f(x) = \frac{\binom{M}{x} * \binom{N-M}{n-x}}{\binom{N}{n}}$ Wenn $n < \frac{N}{20} \rightarrow Hyp(N,M,n) \approx B(n,\frac{M}{N})$

Poisson Verteilung („Verteilung seltener Ereignisse“)

Approximation für B(n,p) und Hyp (N,M,n) für n -> ∞ und p -> 0

Nur geeignet wenn $n * p < 10 \; und \; n > 1500 * p$

X ~ P(μ) $\quad f(x) = \frac{\mu^x}{x!} * e^{-\mu} \quad Hyp(N,M,n) \xrightarrow{p=\frac{M}{N}} B(n,p) \xrightarrow{\mu=n\cdot p=n\cdot\frac{M}{N}} P(\mu)$

Stetige Verteilungen

Gleichverteilung

$$f(x) = \begin{cases} \frac{1}{b-a} & a \le x \le b \\ 0 & sonst \end{cases} \qquad F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{x-b} & a \le x \le b \\ 1 & x > b \end{cases}$$

Exponentialverteilung

X ~ Exp(λ)

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & falls \; x \ge 0 \\ 0 & sonst \end{cases} \qquad F(x) = \begin{cases} 1 - e^{-\lambda x} & x \ge 0 \\ 0 & sonst \end{cases}$$

Normalverteilung

X ~ N(μ,σ) $\quad f(x) = \frac{1}{\sqrt{2 * \Pi} * \sigma} e^{\frac{-(x-\mu)^2}{2 * \sigma^2}}$

Verteilungsfkt. F(x) = $\Phi(\frac{x-\mu}{\sigma}) = \Phi(z)$

In Tabelle: $\Phi(-x) = 1 - \Phi(x)$

Dichte ist symmetrisch zu μ: f(μ-x) = f(μ+x)

Maximum bei x = μ, Wendepunkte bei x = μ - σ und x = μ + σ

Vergleich stetig / diskret

	diskret	**stetig**
$f(x)$	Wahrscheinlichkeitsfunktion ($\sum = 1$)	Dichtefunktion ($\int = 1$)
$f \to F$	$F(x) = \sum_{x_i \le x} f(x_i)$	$F(x) = \int_{-\infty}^{x} f(t)dt$
$F \to f$	$f(x) = F(x) - F(x-1)$	$f(x) = F'(x)$

Verteilungsparameter

Modalwert: $f(x_{mod}) \ge f(x)$

Median: $f(x_{med}) = \frac{1}{2} \; oder \; kleinstes \; x \; mit \; F(x) = \frac{1}{2}$

Erwartungswert

$$E(X) = \begin{cases} \sum_{i=1}^{n} x_i * f(x_i), & falls \; X \; diskret \\ \int_{-\infty}^{+\infty} x * f(x)dx, & falls \; X \; stetig \end{cases}$$

Verteilung von X	E(X)	Var(X)
B(n,p)	$n * p$	$n * p(1-p)$
Hyp(N,M,n)	$n * \frac{M}{N}$	$n * \frac{M}{N} * \left(\frac{N-M}{N}\right) * \left(\frac{N-n}{N-1}\right)$
P(μ)	μ	μ
Gleichverteilung in [a,b] a<b	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exp(λ)	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
N(μ,σ)	μ	σ^2

$E(a + b * X) = a + b * E(x)$

$E(X * Y) = E(X) * E(Y)$

$E(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} E(X_i)$

Wenn f(x) symmetrisch: E(X)= x‘ $\quad$ x‘= Symmetriestelle

Varianz

$$Var(X) = E(X - E(X))^2 = E(X^2) - \big(E(X)\big)^2 = \begin{cases} \sum_{i=1}^{n} (x_i - E(X))^2 * f(x_i), & wenn\ diskret \\ \int_{-\infty}^{+\infty} (x - E(X))^2 * f(x)dx, & wenn\ stetig \end{cases}$$

Standardabweichung

$$Sta(X) = \sqrt{Var(X)}$$

Ungleichung von Tschebyscheff

$$P(|X - E(X)| \geq c) \leq \frac{Var(X)}{c^2}$$

Kovarianz

$$Cov(X,Y) = E\big[\big(X - E(X)\big) * \big(Y - E(Y)\big)\big] = E(X * Y) - E(X) * E(Y)$$

$$Cov(X,X) = Var(X)$$

$$Cov(X,Y) = Cov(Y,X)$$

$$Cov(a + bX, c + dY) = b * d * Cov(X,Y)$$

$$Cov(X + Y, Z) = Cov(X,Z) + Cov(Y,Z)$$

Summe zweier Zufallsvariablen: $Var(X + Y) = Var(X) + Var(Y) + 2 * Cov(X,Y)$

Differenz zweier Zufallsvariablen: $Var(X - Y) = Var(X) + Var(Y) - 2 * Cov(X,Y)$

Korrelationskoeffizient

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X) * Var(Y)}}$$

wenn X,Y unabhängig gilt: $E(X * Y) = E(X) * E(Y)$

- ⇨ Cov(X,Y) = 0
- ⇨ ρ (X,Y) = 0
- ⇨ Unkorreliertheit

<u>Zentraler Grenzwertsatz</u>

Summen von i.i.d Zufallsvariablen sind für große n näherungsweise <u>normalverteilt</u>.

wenn <u>n > 30</u> & Spezialfall Binomialverteilung: $n * p \geq 5\ und\ n * (1 - p) \geq p$ gilt:

$$\sum_{i=1}^{n} X_i \sim N(n * \mu, \sigma * \sqrt{n})$$

Induktive Statistik

Stichprobenfunktion	Bezeichnung	Erwartungswert	Varianz	Verteilung
$\sum_{i=1}^{n} X_i$	Merkmalssumme	$n * \mu$	$n * \sigma^2$	$N(n * \sigma, \sigma * \sqrt{n})$
$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$	Stichprobenmittel	μ	$\frac{\sigma^2}{n}$	$N(\mu, \frac{\sigma}{\sqrt{n}})$
$\frac{\bar{X} - \mu}{\sigma}\sqrt{n}$	Gauß-Statistik	0	1	$N(0,1)$
$\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$	mittlere quadratische Abweichung bzgl. μ	σ^2	-	
$\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$	mittlere quad. Abweichung	$\frac{n-1}{n} * \sigma^2$	-	
$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$	Stichprobenvarianz	σ^2	-	
$S = \sqrt{S^2}$	Stichprobenstandardabweichung	-	-	
$\frac{\bar{X} - \mu}{S}\sqrt{n}$	t-Statistik	-	-	$t(n-1)$

Punktschätzung

Schätzwert: $\hat{\vartheta}$

Schätzfunktion: $\hat{\theta} = g(X_1, \dots, X_n)$

Schätzfunktion heißt *erwartungstreu* wenn: $E(\hat{\theta}) = \vartheta$

Schätzfunktion heißt *asymptotisch erwartungstreu* wenn $\lim_{n\to\infty} E(\hat{\theta}) = \vartheta$

Wirksamkeit

$\hat{\theta}_1 \ wirksamer \ \hat{\theta}_2 \leftrightarrow Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$

Verteilung von G	ϑ	wirksamste erwartungstreue Schätzfunktion
unbekannt	μ	$\bar{X}$
B(1,p)	p (=μ)	$\bar{X}$
N(μ,σ), σ bekannt / unbekannt	μ	$\bar{X}$
N(μ,σ), μ bekannt	σ^2	$\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$
N(μ,σ), μ unbekannt	σ^2	s^2

Konsistente Schätzfunktionen

Hinreichende Konsistenzbedingung: $E(\hat{\theta}_n) = \vartheta \ und \ \lim_{n\to\infty} Var(\hat{\theta}_n) = 0$

Maximum Likelihood Prinzip

Likelihoodfunktion: $f(x_1, \dots, x_n|\vartheta)$

Wähle $\hat{\vartheta}$ so, dass für alle möglichen ϑ – Werte gilt:

$$f(x_1, \dots, x_n|\hat{\vartheta}) \geq (x_1, \dots, x_n|\vartheta)$$

Vorgehensweise ML-Schätzung:

1. Likelihoodfunktion aufstellen $\quad f(x_1, \dots, x_n|\vartheta)$
2. gegebenenfalls logarithmieren $\quad \ln(f(x_1, \dots, x_n|\vartheta))$
3. 1.Ableitung gleich null setzen $\quad \frac{\delta}{\delta\vartheta}[ln]f(x_1, \dots, x_n|\vartheta) = 0 \quad !$
4. Vorzeichen 2.Ableitung prüfen $\quad \frac{\delta^2}{\delta\vartheta^2}[ln]f(x_1, \dots, x_n|\vartheta) < 0 \quad ?$

Verteilung von G	ϑ	Ml-Schätzfunktion
B(1,p)	p (=μ)	$\overline{X}$
Exp(λ)	μ	$\overline{X}$
Exp(λ)	σ^2	$\overline{X}^2$
P(μ)	μ	$\overline{X}$
N(μ,σ), σ bekannt / unbekannt	μ	$\overline{X}$
N(μ,σ), μ bekannt	σ^2	$\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$
N(μ,σ), μ unbekannt	σ^2	$\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$

Intervallschätzung

Konfidenzintervall für μ bei Normalverteilung und σ^2 bekannt

1. Konfidenzniveau: $1 - \alpha$ bestimmen.
2. $c = 1 - \frac{\alpha}{2}$ Fraktil der N(0,1) - Verteilung bestimmen.
3. $\overline{x}$ berechnen.
4. $\frac{\sigma * c}{\sqrt{n}}$ berechnen.
5. Intervall angeben: $[\overline{x} - \frac{\sigma * c}{\sqrt{n}} \,;\, \overline{x} + \frac{\sigma * c}{\sqrt{n}}]$

Intervalllänge: $L = \frac{2 * \sigma * c}{\sqrt{n}}$

Konfidenzintervall für μ bei Normalverteilung und σ^2 unbekannt

1. Konfidenzniveau: $1 - \alpha$ bestimmen.
2. $c = 1 - \frac{\alpha}{2}$ Fraktil der t(n-1) - Verteilung bestimmen. $\quad$ n-1>30 → N(0,1) (erst ab n >100
3. $\overline{x}$ und s berechnen. $s^2 = \frac{1}{n-1} * \sum_{i=1}^{n} x_i^2 - \frac{n}{n-1} * \overline{x}^2$ $\quad$ bis auf 4. Stelle genau)
4. $\frac{s * c}{\sqrt{n}}$ berechnen.
5. Intervall angeben: $[\overline{x} - \frac{s * c}{\sqrt{n}} \,;\, \overline{x} + \frac{s * c}{\sqrt{n}}]$

Konfidenzintervall für μ bei beliebiger / dichotomer Verteilung

Voraussetzung: $n > 30$ oder $5 \leq \sum_{i=1}^{n} x_i \leq n - 5$ $falls\ G\ dichotom$

1. Konfidenzniveau: $1 - \alpha$ bestimmen.
2. $c = 1 - \frac{\alpha}{2}$ Fraktil der N(0,1)- Verteilung bestimmen.
3. $\overline{x}$ und einen Schätzwert $\hat{\sigma}$ $für\ \sigma$ bestimmen: $\hat{\sigma} = \begin{cases} \sigma & falls\ \sigma\ bekannt. \\ \sqrt{\overline{x} * (1 - \overline{x})} & falls\ G\ dichotom \\ s & sonst \end{cases}$
4. $\frac{\hat{\sigma} * c}{\sqrt{n}}$ berechnen.
5. Intervall angeben: $[\overline{x} - \frac{\hat{\sigma} * c}{\sqrt{n}}\ ;\ \overline{x} + \frac{\hat{\sigma} * c}{\sqrt{n}}]$

Intervalllänge: a) σ bekannt: siehe oben.

b) σ unbekannt: $L = \frac{2 * \hat{\sigma} * c}{\sqrt{n}}$

Konfidenzintervall für σ^2 bei Normalverteilung

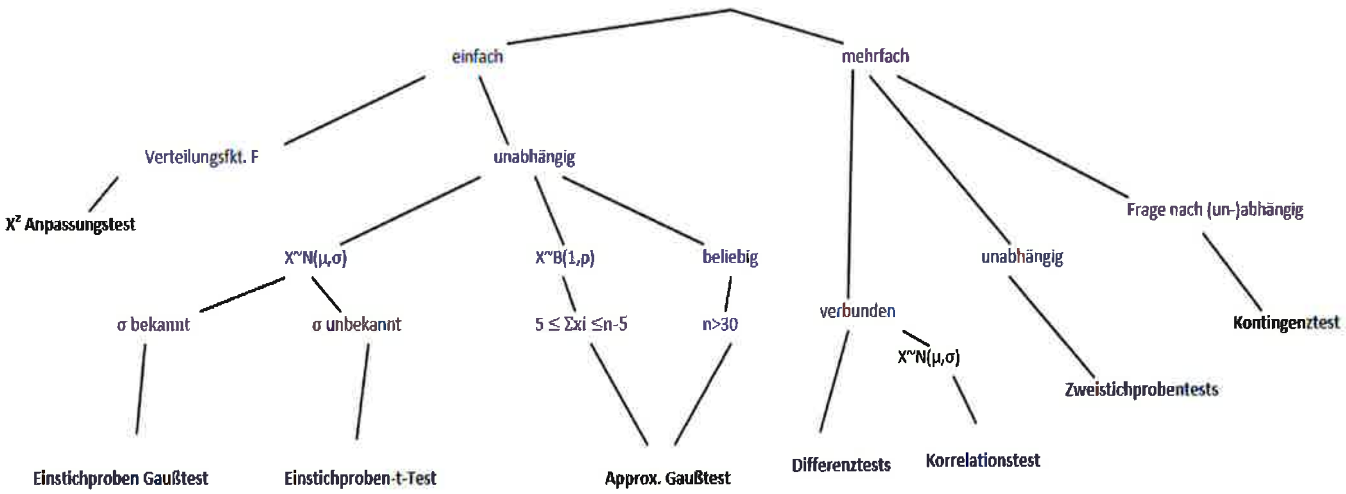
a) μ bekannt:

1. Konfidenzniveau: $1 - \alpha$ bestimmen.
2. $c_1 = \frac{\alpha}{2} und\ c_2 = 1 - \frac{\alpha}{2}$ Fraktil der $\chi^2(n)$ - Verteilung bestimmen.
3. $\sum_{i=1}^{n} (x_i - \mu)^2 = \sum_{i=1}^{n} x_i^2 - n * \mu^2$ berechnen.
4. $\vartheta_u = \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{c_2}$ und $\vartheta_o = \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{c_1}$ bestimmen.
5. Intervall angeben: $[\vartheta_u; \vartheta_o]$

b) μ unbekannt:

1. Konfidenzniveau: $1 - \alpha$ bestimmen.
2. $c_1 = \frac{\alpha}{2} und\ c_2 = 1 - \frac{\alpha}{2}$ Fraktil der $\chi^2(n-1)$ - Verteilung bestimmen.
3. $(n-1) * s^2 = \sum_{i=1}^{n} (x_i - \overline{x})^2 = \sum_{i=1}^{n} x_i^2 - n * \overline{x}^2$ berechnen.
4. $\vartheta_u = \frac{(n-1) * s^2}{c_2}$ und $\vartheta_o = \frac{(n-1) * s^2}{c_1}$ bestimmen.
5. Intervall angeben: $[\vartheta_u; \vartheta_o]$

Signifikanztests



Einstichproben-Gaußtest

- Gegeben: Einfache Stichprobe $X_1, \dots, X_n$ $\quad i.i.d.$ mit $G \sim N(\mu, \sigma)$
- Hypothesenpaar:

  H_0: $\mu = \mu_0$

- Vorgehensweise:
  1. Signifikanzniveau α festlegen.
  2. $v = \frac{\bar{x}-\mu_0}{\sigma} * \sqrt{n}$ berechnen
  3. Verwerfungsbereich B festlegen:
     a) $B = \left(-\infty, -x_{1-\frac{\alpha}{2}}\right) \cup \left(x_{1-\frac{\alpha}{2}}, \infty\right)$
     b) $B = (-\infty, -x_{1-\alpha})$ Verteilung: N(0,1)
     c) $B = (x_{1-\alpha,\infty})$
  4. H_0 verwerfen, wenn $v \in B$.

Einstichproben-t-Test und approximativer Gaußtest

- Gegeben: Einfache Stichprobe $X_1, \dots, X_n$ $\quad i.i.d.$ mit $E(X_i) = \mu \; und \; Var(X_i) = \sigma^2$
- Hypothesenpaare:
  a) H_0: $\mu = \mu_0$ $\quad H_1$: $\mu \neq \mu_0$
  b) H_0: $\mu \geq \mu_0$ $\quad H_1$: $\mu < \mu_0$
  c) H_0: $\mu \leq \mu_0$ $\quad H_1$: $\mu > \mu_0$
- Voraussetzung:
  1. $G \sim N(\mu, \sigma), \sigma \; unbekannt$

     → Einstichproben-t-Test

  2. Beliebige Verteilung mit $n > 30 \; oder \; 5 \leq \sum_{i=1}^{n} x_i \leq n-5 \; falls \; G \sim B(1,p)$

     → approximativer Gaußtest

- Vorgehensweise: $\quad s = \sqrt{\frac{1}{n-1} * \sum_{i=1}^{n}(x_i - \bar{x})^2}$
  1. Signifikanzniveau α festlegen.
  2. v berechnen. $v = \begin{cases} \frac{\bar{x}-\mu_0}{s} * \sqrt{n} & \text{unter Voraussetzung 1} \\ \frac{\bar{x}-\mu_0}{\sigma} * \sqrt{n} & \text{unter Voraussetzung 2 und } \sigma \text{ bekannt} \\ \frac{\bar{x}-\mu_0}{s} * \sqrt{n} & \text{unter Voraussetzung 2 und } \sigma \text{ unbekannt} \\ \frac{\bar{x}-p_0}{\sqrt{p_0*(1-p_0)}} * \sqrt{n} & \text{unter Voraussetzung 2 und G} \sim \text{B}(1,p) \end{cases}$
  3. Verwerfungsbereich B festlegen:
     a) $B = \left(-\infty, -x_{1-\frac{\alpha}{2}}\right) \cup \left(x_{1-\frac{\alpha}{2}}, \infty\right)$ Verteilung:
     b) $B = (-\infty, -x_{1-\alpha})$ Voraussetzung 1 und $n-1 \leq 30$ → t(n-1)
     c) $B = (x_{1-\alpha,\infty})$ Voraussetzung 2 oder $n-1 > 30$ →N(0,1)
  4. H_0 verwerfen, wenn $v \in B$.

Differenztests

- Gegeben: Zwei verbundene Stichproben $X_1, \dots, X_n$ *und* $Y_1, \dots, Y_n$ mit $E(X_i) = \mu_1$ *und* $E(Y_i) = \mu_2$
- Hypothesenpaare:

a) H_0: $\mu_1 = \mu_2$ $\quad H_1$: $\mu_1 \neq \mu_2$

b) H_0: $\mu_1 = \mu_2\ (oder\ \mu_1 \geq \mu_2)$ $\quad H_1$: $\mu_1 < \mu_2$

c) H_0: $\mu_1 = \mu_2\ (oder\ \mu_1 \leq \mu_2)$ $\quad H_1$: $\mu_1 > \mu_2$

- Trick:

Übergang zu $Z_i = X_i - Y_i\ mit\ E(Z_i) = \mu = \mu_1 - \mu_2$ Einstichprobentests mit $\mu_0 = 0$:

a) H_0: $\mu = 0$ $\quad H_1$: $\mu \neq 0$

b) H_0: $\mu = 0\ (oder\ \mu \geq 0)$ $\quad H_1$: $\mu < 0$

c) H_0: $\mu = 0\ (oder\ \mu \leq 0)$ $\quad H_1$: $\mu > 0$

- Vorgehensweise:
  1. Signifikanzniveau α festlegen.
  2. v berechnen.

Voraussetzung	anzuwendender Test	Testfunktion v
Z_i normalverteilt	Einstichproben-*t*-Test	$\frac{\overline{z}}{\sqrt{\frac{1}{n-1} * \sum_{i=1}^{n}(z_i - \overline{z})^2}} * \sqrt{n}$
$X_i, Y_i\ dichotom$ mind. 5 $z_i > 0$ und 5 $z_i < 0$	approximativer Gaußtest	$\frac{\sum_{i=1}^{n} z_i}{\sqrt{\sum_{i=1}^{n} {z_i}^2}}$
$Z_i$beliebig verteilt $n > 30$	approximativer Gaußtest	$\frac{\overline{z}}{\sqrt{\frac{1}{n-1} * \sum_{i=1}^{n}(z_i - \overline{z})^2}} * \sqrt{n}$

Alternativ wenn $X_i, Y_i\ dichotom$:

$v = \frac{h_{10} - h_{01}}{\sqrt{h_{10} + h_{01}}}$ mit

	Y=0	Y=1
X=0	h_{00}	h_{01}
X=1	h_{10}	h_{11}

Bedingung: $h_{01} \geq 5\ und\ \ h_{10} \geq 5$

  3. Verwerfungsbereich B festlegen:

a) $B = \left(-\infty, -x_{1-\frac{\alpha}{2}}\right) \cup \left(x_{1-\frac{\alpha}{2}}, \infty\right)$ $\quad$ Verteilung: siehe jew. Test

b) $B = (-\infty, -x_{1-\alpha})$

c) $B = (x_{1-\alpha,\infty})$

  4. H_0 verwerfen, wenn $v \in B$.

X² Test für die Varianz

- Gegeben: Einfache Stichprobe $X_1, \dots, X_n$ $\quad X \sim N(\mu, \sigma)$
- Hypothesenpaare:

a) H_0: $\sigma^2 = {\sigma_0}^2$ $\quad H_1$: $\sigma^2 \neq {\sigma_0}^2$

b) H_0: $\sigma^2 = {\sigma_0}^2\ (oder\ \sigma^2 \geq {\sigma_0}^2)$ $\quad H_1$: $\sigma^2 < {\sigma_0}^2$

c) H_0: $\sigma^2 = {\sigma_0}^2\ (oder\ \sigma^2 \leq {\sigma_0}^2)$ $\quad H_1$: $\sigma^2 > {\sigma_0}^2$

- Vorgehensweise:
  1. Signifikanzniveau α festlegen.
  2. v berechnen. $v = \begin{cases} \frac{1}{{\sigma_0}^2} * \sum_{i=1}^{n}(x_i - \mu)^2 & falls\ \mu\ bekannt. \\ \frac{(n-1)*s^2}{{\sigma_0}^2} = \frac{1}{{\sigma_0}^2} * \sum_{i=1}^{n}(x_i - \overline{x})^2 & falls\ \mu\ unbekannt. \end{cases}$
  3. Verwerfungsbereich B festlegen:

a) $B = [0, x_{\frac{\alpha}{2}}) \cup \left(x_{1-\frac{\alpha}{2}}, \infty\right)$ $\quad$ Verteilung: μ bekannt: $\chi^2(n)$

b) $B = [0, x_\alpha)$ $\quad$ μ unbekannt: $\chi^2(n-1)$

c) $B = (x_{1-\alpha,\infty})$

  4. H_0 verwerfen, wenn $v \in B$.

Zweistichprobentests

- Gegeben: 2 unabhängige einfache Stichproben $X_1, \dots, X_{n1}$ und $Y_1, \dots, Y_{n2}$ mit $E(X_i) = \mu_1$ und $E(Y_i) = \mu_2$
- Gesucht: Aussagen über Vergleich der Erwartungswerte μ_1 und μ_2.
- Hypothesenpaare:
  - a) $H_0:\ \mu_1 = \mu_2$ $\quad H_1:\ \mu_1 \neq \mu_2$
  - b) $H_0:\ \mu_1 = \mu_2\ (oder\ \mu_1 \geq \mu_2)$ $\quad H_1:\ \mu_1 < \mu_2$
  - c) $H_0:\ \mu_1 = \mu_2\ (oder\ \mu_1 \leq \mu_2)$ $\quad H_1:\ \mu_1 > \mu_2$
- Vorgehensweise:
  1. Signifikanzniveau α festlegen.
  2. v berechnen.

Voraussetzung	anzuwendender Test	Testfunktion v	Verteilung
$X_i \sim N(\mu_1, \sigma_1)$ $Y_i \sim N(\mu_2, \sigma_2)$ $\sigma_1, \sigma_2\ bekannt$	Zweistichproben Gaußtest	$\dfrac{\overline{x} - \overline{y}}{\sqrt{\dfrac{{\sigma_1}^2}{n_1} + \dfrac{{\sigma_2}^2}{n_2}}}$	N(0,1)
$X_i \sim N(\mu_1, \sigma_1)$ $Y_i \sim N(\mu_2, \sigma_2)$ $\sigma_1, \sigma_2\ unbekannt$ $aber\ \sigma_1 = \sigma_2$	Zweistichproben-t-Test	$\dfrac{\overline{x} - \overline{y}}{\sqrt{\dfrac{(n_1 - 1) * {s_1}^2 + (n_2 - 1) * {s_2}^2}{n_1 + n_2 - 2} * \dfrac{n_1 + n_2}{n_1 * n_2}}}$	$t(n_1 + n_2 - 2)$ für $n_1+n_2 > 32$ → N(0,1)
$X_i \sim B(1, p_1)$ $Y_i \sim B(1, p_2)$ $5 \leq \sum x_i \leq n_1 - 5$ $5 \leq \sum y_i \leq n_2 - 5$	approximativer Zweistichproben Gaußtest	$\dfrac{\bar{X} - \bar{Y}}{\sqrt{\dfrac{(\sum X_i + \sum Y_i)\cdot(n_1 + n_2 - \sum X_i - \sum Y_i)}{(n_1 + n_2)\cdot n_1 \cdot n_2}}}$	approximativ N(0,1)
$X_i, Y_i\ beliebig\ verteilt$ $n_1 > 30$, $n_2 > 30$	approximativer Zweistichproben Gaußtest	$\dfrac{\overline{x} - \overline{y}}{\sqrt{\dfrac{{s_1}^2}{n_1} + \dfrac{{s_2}^2}{n_2}}}$	approximativ N(0,1)

3. Verwerfungsbereich B festlegen:
   - a) $B = \left(-\infty, -x_{1-\frac{\alpha}{2}}\right) \cup \left(x_{1-\frac{\alpha}{2}}, \infty\right)$
   - b) $B = (-\infty, -x_{1-\alpha})$ Verteilung: Tabelle
   - c) $B = (x_{1-\alpha}, \infty)$
4. H_0 verwerfen, wenn $v \in B$.

X^2 Anpassungstest

- Gegeben: Einfache Stichprobe $X_1, \dots, X_n$ $\quad mit\ Verteilungsfunktion\ F$
- Hypothesenpaar: $H_0:\ F = F_0 \quad H_1:\ F \neq F_0$
- Voraussetzungen:

  $n * p_j \geq 5\ bzw.\ h_j \geq 5, für\ alle\ j = 1 \dots k$ → Wenn nicht erfüllt: Intervalle zusammenlegen!

- Vorgehensweise:

1. Signifikanzniveau α festlegen.
2. X-Achse in $k \geq 2$ disjunkte, aneinander angrenzende Intervalle unterteilen
   - 2.1. $A_1 = (-\infty, z_1], A_2 = (z_1, z_2], \dots, A_k = (z_k, \infty)$
   - 2.2. Anzahl h_j der in A_j liegenden Stichproben notieren $(j = 1 \dots k)$
   - 2.3. Wahrscheinlichkeit $p_j = P(X \in A_j | F_0)$ berechnen $(j = 1 \dots k)$
   - 2.4. $v = \sum_{j=1}^{k} \frac{(h_j - n*p_j)^2}{n*p_j} = \frac{1}{n} \sum_{j=1}^{k} \frac{{h_j}^2}{p_j} - n$
3. Verwerfungsbereich $B = (x_{1-\alpha}, \infty)$ festlegen. Verteilung: χ^2(k-1)
4. H_0 verwerfen, wenn $v \in B$.

Kontingenztest

- Gegeben: Zwei verbundene einfache Stichproben $X_1, \dots, X_n$ und $Y_1, \dots, Y_n$
- Hypothesenpaar:

H_0: Die Merkmale X und Y sind in G unabhängig. H_1: Die Merkmale X und Y sind in G abhängig.

- Voraussetzungen:
  1. $h_{ij} \geq 5\ bzw.\ \widetilde{h_{ij}} \leq 5\ für\ alle\ i,j$ Wenn nicht: Intervalle zusammenlegen!
  2. Wenn G diskret: pro Ausprägung ein Intervall (falls 1. erfüllt) -> Schritt 2.1 entfällt.
  3. Falls $k = l = 2$: → Schritt 2.3. entfällt, Schritt 2.4.: $v = \frac{n*(h_{11}h_{22}-h_{12}h_{21})}{h_{A1}h_{A2}h_{B1}h_{B2}}$
  4. Falls Randhäufigkeiten $p_i = P(X \in A_i), q_j = P(X \in B_j)$ bekannt:

     → Schritt 2.3.: Ersetze $\widetilde{h_{ij}}$ durch $n * p_i q_j$

     → Schritt 3: Verwende $\chi^2(k * l - 1)$ – Verteilung.

- Vorgehensweise:

1. Signifikanzniveau α festlegen.
2. X-Achse in $k \geq 2$ und Y-Achse in $l \geq 2$ disjunkte, aneinander angrenzende Intervalle unterteilen
   2.1. $A_1, \dots, A_n$ und $B_1, \dots, B_n$
   2.2. Kontingenztabelle erstellen

		Y				
		B_1	B_2	$\cdots$	B_ℓ	
X	A_1	h_{11}	h_{12}	$\cdots$	$h_{1\ell}$	h_{A_1}
	A_2	h_{21}	h_{22}	$\cdots$	$h_{2\ell}$	h_{A_2}
	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
	A_k	h_{k1}	h_{k2}	$\cdots$	$h_{k\ell}$	h_{A_k}
		h_{B_1}	h_{B_1}	$\cdots$	h_{B_ℓ}	n

h_{ij}, h_{Ai} und h_{Bj} ≙
Anzahl der beobachteten Paare (X,Y) in A_i x B_j

   2.3. $\widetilde{h_{ij}} = \frac{h_{Ai}*h_{Bj}}{n}$ berechnen.
   2.4. $v = \sum_{i=1}^{k}\sum_{j=1}^{l}\frac{(h_{ij}-\widetilde{h_{ij}})^2}{\widetilde{h_{ij}}} = \sum_{i=1}^{k}\sum_{j=1}^{l}\frac{{h_{ij}}^2}{\widetilde{h_{ij}}} - n$
3. Verwerfungsbereich $B = (x_{1-\alpha}, \infty)$ festlegen. Verteilung: $\chi^2((k-1)*(l-1))$
4. H_0 verwerfen, wenn $v \in B$

Korrelationstest

- Gegeben: Zwei verbundene einfache Stichproben $X_1, \dots, X_n$ und $Y_1, \dots, Y_n$
  mit $X_i \sim N(\mu_1, \sigma_1)$, $Y_i \sim N(\mu_2, \sigma_2)$
- Hypothesenpaare:
  a) H_0: $\rho = 0$ H_1: $\rho \neq 0$
  b) H_0: $\rho = 0\ (oder\ \rho \geq 0)$ H_1: $\rho < 0$
  c) H_0: $\rho = 0\ (oder\ \rho \leq 0)$ H_1: $\rho > 0$

  wenn v ∉ B (H_0 beibehalten):
  ➔ <u>kein</u> signifikanter Zusammenhang
- Vorgehensweise:
  1. Signifikanzniveau α festlegen.
  2. $v = \sqrt{n-2} * \frac{r}{\sqrt{1-r^2}}$ berechnen.
     $r = \frac{s_{xy}}{s_x s_y}$ $s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = \frac{1}{n-1}(\sum_{i=1}^{n} x_i y_i - n * \overline{x} * \overline{y})$
  3. Verwerfungsbereich B festlegen:
     a) $B = \left(-\infty, -x_{1-\frac{\alpha}{2}}\right) \cup \left(x_{1-\frac{\alpha}{2}}, \infty\right)$
     b) $B = (-\infty, -x_{1-\alpha})$
     c) $B = (x_{1-\alpha,\infty})$

     Verteilung: t(n-2)
  4. H_0 verwerfen, wenn $v \in B$.