Formelsammlung

Momente, Schiefe & Wölbung

nicht-zentrierte Momente	zentrierte Momente	Schiefe	Wölbung
$\widehat{m}_k = \frac{1}{n}\sum_{i=1}^n x_i^k$	$\widehat{M}_k = \frac{1}{n}\sum_{i=1}^n (x_i - \widehat{m}_1)^k$	$\gamma_{1,M} = \frac{\widehat{M}_3}{\left\lvert(\widehat{M}_2)^{\frac{3}{2}}\right\rvert}$	$\gamma_2 = \frac{\widehat{M}_4}{(\widehat{M}_2)^2}$
$\bar{x} = \widehat{m}_1 = \frac{1}{n}\sum_{i=1}^n x_i$	$s^2 = \widehat{M}_2 = \frac{1}{n}\sum_{i=1}^n x_i^2 - \bar{x}^2$		

Für gruppierte Daten: $\bar{x} = \frac{1}{n}\sum_{k=1}^K n_k \bar{x}_k \qquad s^2 = \frac{1}{n}\sum_{k=1}^K n_k s_k^2 + \frac{1}{n}\sum_{k=1}^K n_k(\bar{x}_k - \bar{x})^2$

Quantilskoeffizient der Schiefe: $\gamma_{1,Q}(p) = \frac{(z_{1-p} - \bar{x}^{\text{Med}}) - (\bar{x}^{\text{Med}} - z_p)}{z_{1-p} - z_p}$

Regeln für lineare Transformationen $y_i = a + bx_i$

$$\bar{y} = a + b\bar{x} \qquad s_y^2 = b^2 s_x^2$$

Lagemaße

$$\bar{x}^{\text{Med}} = \begin{cases} x_{[\frac{n+1}{2}]} & \text{für } n \text{ ungerade} \\ \frac{1}{2}\left(x_{[\frac{n}{2}]} + x_{[\frac{n}{2}+1]}\right) & \text{für } n \text{ gerade} \end{cases} \qquad \bar{x}^{\text{Geo}} = \left(\prod_{i=1}^n x_i\right)^{\frac{1}{n}}$$

Abhängigkeitsmaße

χ^2-Koeffizient	Kontingenzkoeffizient	normierter Kontingenzkoeffizient
$\chi^2 = \sum_{k=1}^K \sum_{l=1}^L \frac{(n_{k,l} - \tilde{n}_{k,l})^2}{\tilde{n}_{k,l}}$	$C = \sqrt{\frac{\chi^2}{\chi^2 + n}}$	$C^* = \frac{C}{\sqrt{\frac{\min\{K,L\}-1}{\min\{K,L\}}}}$

Kovarianz	Korrelationskoeffizient nach BRAVAIS-PEARSON	Korrelationskoeffizient nach SPEARMAN
$s_{x,y} = \frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$	$r_{x,y} = \frac{s_{x,y}}{s_x \cdot s_y}$	$r_{x,y}^{\text{Sp}} = \frac{s_{\text{rg}(x),\text{rg}(y)}}{s_{\text{rg}(x)} \cdot s_{\text{rg}(y)}}$
$s_{x,y} = \left(\frac{1}{n}\sum_{i=1}^n x_i y_i\right) - \bar{x}\bar{y}$		$r_{x,y}^{\text{Sp}} = 1 - \frac{6}{n(n^2-1)}\sum_{i=1}^n (\text{rg}(x_i) - \text{rg}(y_i))^2$

Lineare Regression

$\hat{y} = a + b \cdot x$	$\hat{b} = \frac{s_{x,y}}{s_x^2}$	$\hat{a} = \bar{y} - \hat{b} \cdot \bar{x}$	$R^2 = \hat{b}^2 \cdot \frac{s_x^2}{s_y^2} = r_{x,y}^2$

Momente von Zufallsvariablen

diskret	
nicht-zentrierte Momente	zentrierte Momente
$E(X^k) = \sum_{i=1}^\infty x_i^k \cdot P(X = x_i)$	$E((X - E(X))^k) = \sum_{i=1}^\infty (x_i - E(X))^k \cdot P(X = x_i)$
$E(X) = \sum_{i=1}^\infty x_i \cdot P(X = x_i)$	$Var(X) = \sum_{i=1}^\infty x_i^2 \cdot P(X = x_i) - E^2(X)$

stetig	
nicht-zentrierte Momente	zentrierte Momente
$E(X^k) = \int_{-\infty}^\infty x^k \cdot f_X(x)\,\mathrm{d}x$	$E((X - E(X))^k) = \int_{-\infty}^\infty (x - E(X))^k \cdot f_X(x)\,\mathrm{d}x$
$E(X) = \int_{-\infty}^\infty x \cdot f_X(x)\,\mathrm{d}x$	$Var(X) = \int_{-\infty}^\infty x^2 \cdot f_X(x)\,\mathrm{d}x - E^2(X)$

Binomialverteilung

$E(X)$	$Var(X)$	Wahrscheinlichkeitsfunktion
$n \cdot p$	$n \cdot p \cdot (1-p)$	$P(X=x) = \binom{n}{x} p^x (1-p)^{n-x}$

Normalverteilung

Mit $X \sim \mathcal{N}(\mu, \sigma^2)$ und $Z \sim \mathcal{N}(0,1)$ gilt $F_X(x) = F_Z\left(\frac{x-\mu}{\sigma}\right)$ und $x_p = \mu + z_p \cdot \sigma$ für $0 < p < 1$

Intervallschätzung

Erwartungswert	**GG normalverteilt**	**GG beliebig verteilt** $(n \geq 30)$
σ_X^2 bekannt	$\left[\bar{X} - z_{1-\alpha/2}\frac{\sigma_X}{\sqrt{n}}, \bar{X} + z_{1-\alpha/2}\frac{\sigma_X}{\sqrt{n}}\right]$	$\left[\bar{X} - z_{1-\alpha/2}\frac{\sigma_X}{\sqrt{n}}, \bar{X} + z_{1-\alpha/2}\frac{\sigma_X}{\sqrt{n}}\right]$
σ_X^2 unbekannt	$\left[\bar{X} - t_{1-\alpha/2}(n-1)\frac{\check{S}_X}{\sqrt{n}}, \bar{X} + t_{1-\alpha/2}(n-1)\frac{\check{S}_X}{\sqrt{n}}\right]$	$\left[\bar{X} - z_{1-\alpha/2}\frac{\check{S}_X}{\sqrt{n}}, \bar{X} + z_{1-\alpha/2}\frac{\check{S}_X}{\sqrt{n}}\right]$

Anteilswert	$\left[P - z_{1-\alpha/2}\sqrt{\frac{P(1-P)}{n}}, P + z_{1-\alpha/2}\sqrt{\frac{P(1-P)}{n}}\right]$
Varianz	$\left[\frac{(n-1)\check{S}^2}{\chi^2_{1-\alpha/2}(n-1)}, \frac{(n-1)\check{S}^2}{\chi^2_{\alpha/2}(n-1)}\right]$

Hypothesentests

Test	**Voraussetzungen**	**Prüfgröße**	**Prüfverteilung**
Approx. Binomialtest	$n\pi_0 \geq 5\,; n(1-\pi_0) \geq 5$	$G = \frac{\bar{X}-\pi_0}{\sqrt{\frac{\pi_0\,(1-\pi_0)}{n}}}$	$\mathcal{N}(0,1)$
χ^2-**Unabhängigkeitstest**		$G = \sum_{k=1}^{K}\sum_{l=1}^{L}\frac{\left(N_{k,l}-\tilde{N}_{k,l}\right)^2}{\tilde{N}_{k,l}}$	$\chi^2((K-1)(L-1))$
Prüfung einer Varianz		$G = \frac{1}{\sigma_0^2}\sum_{i=1}^{n}(X_i - \bar{X})^2$	$\chi^2(n-1)$
Vergleich zweier Varianzen		$G = \frac{\check{S}_X^2}{\check{S}_Y^2}$	$F(n-1, m-1)$
Vergl. von Anteilswerten	$n, m \geq 30$	$G = \frac{\bar{X}-\bar{Y}-\delta_0}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}+\frac{\bar{Y}(1-\bar{Y})}{m}}}$	approx. $\mathcal{N}(0,1)$
Vergl. von Anteilswerten $\delta_0 = 0$	$n, m \geq 30$	$G = \frac{\bar{X}-\bar{Y}}{\sqrt{\widehat{P}\,(1-\widehat{P})\left(\frac{1}{n}+\frac{1}{m}\right)}}$ $\widehat{P} = \frac{n\,\bar{X}+m\,\bar{Y}}{n+m}$	approx. $\mathcal{N}(0,1)$

Tests zu Lagealternativen

Voraussetzungen	**Prüfgröße**	**Prüfverteilung**
Normalverteilung Varianz σ_X^2 bekannt	$G = \frac{\bar{X}-\mu_0}{\frac{\sigma_X}{\sqrt{n}}}$	Standardnormalverteilung
Normalverteilung Varianz σ_X^2 unbekannt	$G = \frac{\bar{X}-\mu_0}{\frac{\check{S}_X}{\sqrt{n}}}$	t-Verteilung mit $(n-1)$ Freiheitsgraden
keine Verteilungsvoraussetzung Varianz σ_X^2 bekannt; $n \geq 30$	$G = \frac{\bar{X}-\mu_0}{\frac{\sigma_X}{\sqrt{n}}}$	Standardnormalverteilung
keine Verteilungsvoraussetzung approximativ Varianz σ_X^2 unbekannt; $n \geq 30$	$G = \frac{\bar{X}-\mu_0}{\frac{\check{S}_X}{\sqrt{n}}}$	Standardnormalverteilung

Tests zum Vergleich von Erwartungswerten

Voraussetzungen	Prüfgröße	Prüfverteilung
Normalverteilungen; Varianzen σ_X^2 und σ_Y^2 bekannt	$G = \dfrac{\bar{X} - \bar{Y} - \delta_0}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}}$	Standardnormalverteilung
keine Verteilungsvoraussetzung; Varianzen σ_X^2 und σ_Y^2 bekannt; $n, m \geq 30$; approximativ	$G = \dfrac{\bar{X} - \bar{Y} - \delta_0}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}}$	Standardnormalverteilung
keine Verteilungsvoraussetzung; Varianzen σ_X^2 und σ_Y^2 unbekannt; $n, m \geq 30$; approximativ	$G = \dfrac{\bar{X} - \bar{Y} - \delta_0}{\sqrt{\frac{\hat{S}_X^2}{n} + \frac{\hat{S}_Y^2}{m}}}$	Standardnormalverteilung
Normalverteilungen; Varianzen σ_X^2 und σ_Y^2 unbekannt, aber gleich	$G = \dfrac{\bar{X} - \bar{Y} - \delta_0}{\check{S}\sqrt{\frac{1}{n} + \frac{1}{m}}}$ $\check{S}^2 = \dfrac{(n-1)\check{S}_X^2 + (m-1)\check{S}_Y^2}{n+m-2}$	t-Verteilung mit $(n+m-2)$ Freiheitsgraden

Lineare Einfachregression

$\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$	$s_x^2 = \frac{1}{n}\sum_{i=1}^n x_i^2 - \bar{x}^2$	$s_{x,y} = \frac{1}{n}\sum_{i=1}^n x_i y_i - \bar{x} \cdot \bar{y}$
$\hat{\beta} = \frac{s_{x,y}}{s_x^2}$	$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$	$\hat{\varepsilon}_i = y_i - \hat{\alpha} - \hat{\beta}x_i$
$\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^n \hat{\varepsilon}_i^2 = \frac{n}{n-2}(s_y^2 - \hat{\beta}s_{x,y})$	$\hat{\sigma}_{\hat{\alpha}}^2 = \hat{\sigma}^2 \cdot \dfrac{\sum_{i=1}^n x_i^2}{n\left(\sum_{i=1}^n x_i^2 - n\cdot\bar{x}^2\right)}$	$\hat{\sigma}_{\hat{\beta}}^2 = \dfrac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n\cdot\bar{x}^2}$

Hypothesentests

$G_\alpha = \frac{\hat{\alpha} - \alpha_0}{\hat{\sigma}_{\hat{\alpha}}}$	$G_\beta = \frac{\hat{\beta} - \beta_0}{\hat{\sigma}_{\hat{\beta}}}$	$G_{\sigma^2} = (n-2) \cdot \frac{\hat{\sigma}^2}{\sigma_0^2}$
$G_\alpha \sim t(n-2)$	$G_\beta \sim t(n-2)$	$G_{\sigma^2} \sim \chi^2(n-2)$

Konfidenzintervalle

$$\left[\hat{\alpha} - \hat{\sigma}_{\hat{\alpha}} t_{1-\alpha_0/2}(n-2)\ ,\ \hat{\alpha} + \hat{\sigma}_{\hat{\alpha}} t_{1-\alpha_0/2}(n-2)\right] \qquad \left[\hat{\beta} - \hat{\sigma}_{\hat{\beta}} t_{1-\alpha_0/2}(n-2)\ ,\ \hat{\beta} + \hat{\sigma}_{\hat{\beta}} t_{1-\alpha_0/2}(n-2)\right]$$

Multiple lineare Regression

Testverfahren

t-Test: $T_j = \frac{\hat{\beta}_j - \beta_{0,j}}{\hat{\sigma}_j} \sim t(n-k-1)$

$F-$Test: $F = \frac{R^2}{1-R^2} \cdot \frac{n-k-1}{k} \sim F(k, n-k-1)$