Blatt 12

Einfache Varianzanalyse: (auch ANOVA - Analysis of Variances)

- Für die Messreihen $1, \dots, k$ gilt, dass die ZVen $Y_{i,j}$ $i = 1, \dots, k$; $j = 1, \dots, n_i$ unabhängig und $\mathcal{N}(\mu_i, \sigma^2)$-verteilt sind.

$\mu_1, \dots, \mu_k$ und σ^2 sind unbekannt.

- $H_0: \mu_1 = \dots = \mu_k$ H_1: Es gibt mind. 2 μ_i mit verschiedenen Werten

- Um die Testgröße zu berechnen, werden die folgenden Variablen benötigt

$\bar{Y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij}$, $i = 1, \dots, k$ Messreihen-Mittelwerte

$\bar{Y} = \frac{1}{n} \sum_{i=1}^{k} n_i \bar{Y}_i$ Daten-Mittelwert

$SST = \sum_{i=1}^{k} n_i (\bar{Y}_i - \bar{Y})^2$ Varianz zwischen den Messreihen

$SSE = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2$ Varianz innerhalb der Messreihen

Testgröße: $T = \frac{\frac{1}{k-1} SST}{\frac{1}{n-k} SSE}$ Verhältnis der Varianz zwischen den Messreihen zu der Varianz innerhalb der Messreihen

$$T \sim F_{k-1, n-k}$$

- H_0 ablehnen, wenn $T > F_{k-1,n-k}(1-\alpha)$

[in Sagemath mithilfe von r.anova()]

Normalverteilung (bereits auf dem Zettel Blatt 3 - Verteilungen) $\mathcal{N}(\mu;\sigma^2)$ normalverteilt mit Erwartungswert μ und Varianz σ^2

$X_1 \sim \mathcal{N}(\mu_1;\sigma_1^2)$, $X_2 \sim \mathcal{N}(\mu_2;\sigma_2^2)$, $a \in \mathbb{R}$

$\Rightarrow aX_1 \sim \mathcal{N}(a\mu_1; a^2\sigma_1^2)$

$\Rightarrow X_1 + X_2 \sim \mathcal{N}(\mu_1+\mu_2; \sigma_1^2+\sigma_2^2)$

$\Rightarrow aX_1 + b \sim \mathcal{N}(a\mu_1 + b; b^2\sigma_2^2)$

Dadurch kann $X \sim \mathcal{N}(\mu;\sigma^2)$ auch zur Standardnormalverteilung $Z \sim \mathcal{N}(0;1)$ transformiert werden:

$\frac{X-\mu}{\sigma} \sim \mathcal{N}(0;1)$

Da zudem gilt $P(Z \leq a) = \Phi(a)$ (wobei $\Phi(z)$ entweder in der Quantiltabelle der Standardnormalverteilung oder mit r.pnorm(z) in sage nachgeschaut werden kann), kann man die Transformation wie folgt anwenden:

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

Zusätzlich gilt:

Normalverteilung ist symmetrisch: $\Rightarrow P(X \leq a) = 1 - P(X > a)$

$\Rightarrow P(|X| \geq a) = 2 \cdot P(X \geq a)$

Standardnormalverteilung ist symmetrisch: $\Rightarrow \Phi(-z) = 1 - \Phi(z)$

Die Umkehrabbildung von Φ ist Φ^{-1}: (r.qnorm(u) in sage) $P(Z \leq a) = u \Leftrightarrow a = \Phi^{-1}(u)$

- Sind $Z_1, \dots, Z_n$ je $\mathcal{N}(0,1)$-verteilt

  $\Rightarrow Z_1^2 + \dots + Z_n^2 \sim \chi_n^2$ (Chi-Quadrat-verteilt mit n Freiheitsgraden)

- Ist Z $\mathcal{N}(0,1)$-verteilt, X χ_n^2-verteilt

  $\Rightarrow \frac{Z}{\sqrt{\frac{X}{n}}} \sim t_n$ (t-verteilt mit n Freiheitsgraden) (*)

- Ist $X \sim \chi_n^2$, kann $P(X \leq a)$ mit r.pchisq(a, df=n)
  berechnet werden. Die Umkehrfkt ist r.qchisq(u, df=n).

- Ist $X \sim t_n$, kann $P(X \leq a)$ mit r.pt(a, df=n) berechnet
  werden. Die Umkehrfkt ist dann r.qt(u, df=n).

- $X_1 \sim \chi_{n_1}^2$, $X_2 \sim \chi_{n_2}^2$, $\Rightarrow X_1 + X_2 \sim \chi_{n_1+n_2}^2$, $aX_1 \sim \chi_{a \cdot n_1}^2$
  $a \in \mathbb{R}$

- $X_1, \dots, X_n$ unabhängig und $\mathcal{N}(\mu, \sigma^2)$-verteilt, $\overline{X} = \frac{1}{n} \sum_{i=1}^n X_i$

  $\Rightarrow \frac{\overline{X} - \mu}{\sqrt{\frac{\sum_{i=1}^n (X_i - \overline{X})^2}{n-1}}} \sqrt{n} \sim t_{n-1}$