

2 Klassische statistische Verfahren unter Normalverteilungs-Annahme

2.1 $\chi_s^2, t_s, F_{r,s}$ -Verteilung

- a) $N_1, \dots, N_s \stackrel{uiv}{\sim} \mathcal{N}(0, 1)$
 Verteilung von $Y := N_1^2 + \dots + N_s^2$ heißt Chi-Quadrat-Verteilung mit s Freiheitsgraden (χ_s^2 -Verteilung).
 Y besitzt die Dichte

$$f(y) = \frac{1}{2^{\frac{s}{2}} \Gamma(\frac{s}{2})} e^{-\frac{y}{2}} y^{\frac{s}{2}-1}, \quad y > 0$$

Dabei:

$$\Gamma(t) = \int_0^\infty e^{-x} x^{t-1} dx, \quad t > 0$$

(Gamma-Funktion)

$$[\Gamma(x+1) = x\Gamma(x) \quad (x > 0), \Gamma(n+1) = n! \quad (n \in \mathbb{N}), \Gamma(\frac{1}{2}) = \sqrt{\pi}]$$

Es gilt:

$$EY = s, \quad \text{Var}(Y) = 2s$$

- b) Seien $N_0, N_1, \dots, N_s \stackrel{uiv}{\sim} \mathcal{N}(0, 1)$.
 Die Verteilung von

$$y := \frac{N_0}{\sqrt{\sum_{j=1}^s N_j^2 / s}}$$

heißt (studentsche) t-Verteilung mit s Freiheitsgraden (t_s -Verteilung).

Dichte:

$$f(y) = \frac{1}{\sqrt{\pi s}} \frac{\Gamma(\frac{s+1}{2})}{\Gamma(\frac{s}{2})} \left(1 + \frac{y^2}{s}\right)^{-\frac{s+1}{2}}, \quad y \in \mathbb{R}$$

$$E(Y) = 0 \quad (s \geq 2), \quad \text{Var}(Y) = \frac{s}{s-2} \quad (s \geq 3)$$

$s=1^2$:

$$f(y) = \frac{1}{\sqrt{\pi}} \frac{1}{\sqrt{\pi}} \frac{1}{1+y^2} = \frac{1}{\pi(1+y^2)}, \quad y \in \mathbb{R}$$

- c) Es seien R, S unabhängig, $R \sim \chi_r^2$, $S \sim \chi_s^2$. Die Verteilung von

$$Y := \frac{\frac{1}{r}R}{\frac{1}{s}S}$$

²Cauchy-Verteilung

heißt F(isher)-Verteilung mit r Zähler- und s Nenner-Freiheitsgraden (kurz: $Y \sim F_{r,s}$).

Dichte:

$$f(y) = \frac{\Gamma(\frac{r+s}{2}) (\frac{r}{s})^{\frac{r}{2}} y^{\frac{r}{2}-1}}{\Gamma(\frac{r}{2}) \Gamma(\frac{s}{2}) (1 + \frac{r}{s}y)^{\frac{r+s}{2}}}, \quad y > 0$$

Es gilt:

$$E(Y) = \frac{s}{s-2}, \quad s > 2, \quad \text{Var}(Y) = \frac{2s^2(r+s-2)}{r(s-2)^2(s-4)}, \quad s > 4$$

2.2 Satz

Es seien $X_1, \dots, X_n \stackrel{uiv}{\sim} \mathcal{N}(\mu, \sigma^2)$.

$$\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i, \quad \hat{\sigma}_n^2 := \frac{1}{n} \sum_{j=1}^n (X_j - \bar{X}_n)^2$$

Dann gilt:

- a) $\bar{X}_n \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$.
- b) $\frac{n}{\sigma^2} \hat{\sigma}_n^2 \sim \chi_{n-1}^2$
- c) \bar{X}_n und $\hat{\sigma}_n^2$ sind unabhängig.

Beweis

a) klar

b),c) Sei $Z_j := X_j - \mu$, $Z := (Z_1, \dots, Z_n)^T$.

Es gilt: $Z \sim \mathcal{N}_n(0, \sigma^2 \cdot I_n)$.

Sei $H = (h_{ij})$ orthogonale $n \times n$ -Matrix mit $h_{nj} = \frac{1}{\sqrt{n}}$, $1 \leq j \leq n$.

$$Y = (Y_1, \dots, Y_n)^T := HZ$$

$Y \sim \mathcal{N}_n(0, \sigma^2 H H^T) = \mathcal{N}_n(0, \sigma^2 I_n)$, d.h. $Y_1, \dots, Y_n \stackrel{uiv}{\sim} \mathcal{N}(0, \sigma^2)$.

Es gilt:

$$\sum_{j=1}^n Y_j^2 = \|Y\|^2 = \|Z\|^2 = \sum_{j=1}^n Z_j^2$$

$$Y_n = \frac{1}{\sqrt{n}} \sum_{j=1}^n Z_j = \sqrt{n}(\bar{X}_n - \mu)$$

Mit $\bar{Z}_n := \frac{1}{n} \sum_{j=1}^n Z_j$ folgt:

$$\begin{aligned}
 \sum_{i=1}^n (X_j - \bar{X}_n)^2 &= \sum_{j=1}^n (Z_j - \bar{Z}_n)^2 = \sum_{j=1}^n Z_j^2 - n\bar{Z}_n^2 \\
 &= \sum_{j=1}^n Y_j^2 - Y_n^2 \\
 &= \sum_{j=1}^{n-1} Y_j^2 \\
 &= \sigma^2 \underbrace{\sum_{j=1}^{n-1} \left(\frac{Y_j}{\sigma} \right)^2}_{\substack{\sim \mathcal{N}(0,1) \\ \sim \chi_{n-1}^2}}
 \end{aligned}$$

\Rightarrow b)

Wegen $\sum_{j=1}^n (X_j - \bar{X}_n)^2 =$ Funktion von Y_1, \dots, Y_{n-1} und $\bar{X}_n =$ Funktion von Y_n sind $\hat{\sigma}_n^2$ und \bar{X}_n unabhängig (Blockungslemma). ■

2.3 Korollar

In der Situation von 2.2 sei³

$$S_n^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{n}{n-1} \hat{\sigma}_n^2$$

Dann gilt:

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} \sim t_{n-1}$$

Beweis:

$$\sqrt{n} \frac{\bar{X}_n - \mu}{S_n} = \frac{\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}}{\sqrt{\frac{1}{\sigma^2} \cdot \underbrace{\sum_i (X_i - \bar{X}_n)^2}_{\sim \chi_{n-1}^2} / (n-1)}}$$

Zähler und Nenner sind unabhängig und der Zähler ist $\mathcal{N}(0, 1)$ -verteilt.

^{2.1b)} \Rightarrow Behauptung

³Stichprobenvarianz

2.4 Korollar

Sei $0 < \alpha \leq 1$ und sei $t_{n-1, 1-\frac{\alpha}{2}}$ das $(1 - \frac{\alpha}{2})$ -Quantil der t_{n-1} -Verteilung.⁴ Dann gilt:

$$P_{\mu, \sigma^2} \left(\left| \frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} \right| \leq t_{n-1, 1-\frac{\alpha}{2}} \right) = 1 - \alpha$$

Oder äquivalent:

$$P_{\mu, \sigma^2} \left(\bar{X}_n - \frac{S_n}{\sqrt{n}} t_{n-1, 1-\frac{\alpha}{2}} \leq \mu \leq \bar{X}_n + \frac{S_n}{\sqrt{n}} t_{n-1, 1-\frac{\alpha}{2}} \right) = 1 - \alpha$$

Sprechweise:

$[\bar{X}_n - \frac{S_n}{\sqrt{n}} t_{n-1, 1-\frac{\alpha}{2}}, \bar{X}_n + \frac{S_n}{\sqrt{n}} t_{n-1, 1-\frac{\alpha}{2}}]$ ist ein Konfidenzintervall (Vertrauensintervall) zur Konfidenzwahrscheinlichkeit (Vertrauenswahrscheinlichkeit) $1 - \alpha$ für μ . (Störparameter σ^2 tritt hier nicht auf!)

2.5 Ein-Stichproben-t-Test (1-SP-t-Test)

Seien $X_1, \dots, X_n \stackrel{uiv}{\sim} \mathcal{N}(\mu, \sigma^2)$, $\vartheta = (\mu, \sigma^2)$, $\Theta := \mathbb{R} \times \mathbb{R}_{>0}$.

Zu testen sei $H_0: \mu = \mu_0$ gegen $H_1: \mu \neq \mu_0$ (μ_0 gegebener Wert).

$$H_0/H_1 \leftrightarrow \Theta_0/\Theta_1$$

$$\Theta_0 := \{\vartheta = (\mu, \sigma^2) \in \Theta : \mu = \mu_0\}$$

$$\Theta_1 := \{\vartheta = (\mu, \sigma^2) \in \Theta : \mu \neq \mu_0\}$$

Sei $(x_1, \dots, x_n) =: x$ Realisierung von X_1, \dots, X_n .

$$\bar{x}_n := \frac{1}{n} \sum_{i=1}^n x_i$$

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

$$T(x_1, \dots, x_n) := \frac{\sqrt{n}(\bar{x}_n - \mu_0)}{s_n}$$

⁴→ Abbildung 2.1

(Zweiseitiger) 1-SP-t-Test zum Niveau α :

H_0 ablehnen, falls

$$|T(x_1, \dots, x_n)| \geq t_{n-1, 1-\frac{\alpha}{2}}$$

Kein Widerspruch zu H_0 , falls

$$|T(x_1, \dots, x_n)| < t_{n-1, 1-\frac{\alpha}{2}}$$

Sei $\vartheta \in \Theta_0$, also $\mu = \mu_0$.

$\Rightarrow T(X_1, \dots, X_n) \sim t_{n-1}$

$$\Rightarrow P(|T(X_1, \dots, X_n)| \geq t_{n-1, 1-\frac{\alpha}{2}}) = \alpha \quad \forall \vartheta \in \Theta_0$$

[Das bedeutet, wenn H_0 gilt, ist die Wahrscheinlichkeit H_0 trotzdem abzulehnen α . (\rightarrow Niveau)]

Bemerkungen:

1) Entsprechend einseitige Tests⁵, z.B. $H_0 : \mu \geq \mu_0$ gegen $H_1 : \mu < \mu_0$.
 H_0 ablehnen, falls $|T(x_1, \dots, x_n)| \geq t_{n-1, \alpha}$ ⁶

2) Sei

$$f(x, \vartheta) := \prod_{j=1}^n f_1(x_j, \vartheta) = \left(\frac{1}{\sigma \sqrt{2\pi}} \right)^n \exp\left(-\frac{1}{2\sigma^2} \sum_{j=1}^n (x_j - \mu)^2\right)$$

Die Prüfgröße $T(x_1, \dots, x_n)$ des t-Tests ergibt sich aus einem allgemeinen „Rezept“:

Bilde den (verallgemeinerten) Likelihood-Quotienten

$$\lambda(x) := \frac{\sup_{\vartheta \in \Theta_0} f(x, \vartheta)}{\sup_{\vartheta \in \Theta} f(x, \vartheta)}$$

und lehne $H_0 : \vartheta \in \Theta_0$ für zu „kleine“ Werte von $\lambda(x)$ ab.

Es gilt:

$$(n-1)(\lambda(x)^{-\frac{2}{n}} - 1) = T^2(x_1, \dots, x_n)$$

(Blatt 2, Aufgabe 6)

⁵vgl. Stochstik I (\rightarrow Wahl der Nullhypothese)

⁶etc.; Niveau α

2.6 Satz

Seien $X_1, \dots, X_m, Y_1, \dots, Y_n$ unabhängig.

$$X_i \sim \mathcal{N}(\mu, \sigma^2) \quad \forall i, \quad Y_j \sim \mathcal{N}(\nu, \sigma^2) \quad \forall j$$

$$\bar{X}_m := \frac{1}{m} \sum_{i=1}^m X_i, \quad \bar{Y}_n := \frac{1}{n} \sum_{j=1}^n Y_j$$

Dann gilt:

$$\bar{X}_m \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{m}\right), \quad \bar{Y}_n \sim \mathcal{N}\left(\nu, \frac{\sigma^2}{n}\right)$$

$$\bar{X}_m - \bar{Y}_n - (\mu - \nu) \sim \mathcal{N}\left(0, \frac{\sigma^2}{m} + \frac{\sigma^2}{n}\right) = \mathcal{N}\left(0, \frac{m+n}{mn} \sigma^2\right)$$

$$\frac{1}{\sigma^2} \sum_{i=1}^m (X_i - \bar{X}_m)^2 \sim \chi_{m-1}^2$$

$$\frac{1}{\sigma^2} \sum_{j=1}^n (Y_j - \bar{Y}_n)^2 \sim \chi_{n-1}^2$$

\bar{X}_m, \bar{Y}_n und die letzten beiden Größen sind stochastisch unabhängig!

Sei

$$S_{m,n}^2 := \frac{1}{m+n-2} \left(\sum_{i=1}^m (X_i - \bar{X}_m)^2 + \sum_{j=1}^n (Y_j - \bar{Y}_n)^2 \right)$$

Dann gilt weiter:

$$(m+n-2) \cdot S_{m,n}^2 / \sigma^2 \sim \chi_{m+n-2}^2$$

2.7 Korollar

In der Situation von 2.6 gilt:

$$\frac{\sqrt{\frac{mn}{m+n}} (\bar{X}_m - \bar{Y}_n - (\mu - \nu))}{S_{m,n}} \sim t_{m+n-2}$$

Beweis:

Wie Korollar 2.3.

2.8 Korollar

$$P_{\mu,\nu,\sigma^2} \left(\sqrt{\frac{mn}{m+n}} \frac{|\bar{X}_m - \bar{Y}_n - (\mu - \nu)|}{S_{m,n}} \leq t_{m+n-2, 1-\frac{\alpha}{2}} \right) = 1 - \alpha$$

D.h.

$$\bar{X}_m - \bar{Y}_n \pm \frac{S_{m,n}}{\sqrt{\frac{mn}{m+n}}} t_{m+n-2, 1-\frac{\alpha}{2}}$$

ist Konfidenzintervall für $\mu - \nu$ zur Konfidenzwahrscheinlichkeit $1 - \alpha$.

2.9 (Zweiseitiger) 2-SP-t-Test

Situation von 2.6.

$$H_0 : \mu = \nu \ (\mu - \nu = 0), \ H_1 : \mu \neq \nu \ (\mu - \nu \neq 0)$$

Mit

$$\Theta := \{ \vartheta = (\mu, \nu, \sigma^2) : -\infty < \mu, \nu < \infty, \sigma^2 > 0 \}$$

$$\Theta_0 := \{ (\mu, \nu, \sigma^2) \in \Theta : \mu = \nu \}$$

$$\Theta_1 := \{ (\mu, \nu, \sigma^2) \in \Theta : \mu \neq \nu \}$$

gilt: $H_0 \hat{=} \vartheta \in \Theta_0, \ H_1 \hat{=} \vartheta \in \Theta_1$.

Prüfgröße:

$$T_{m,n}(x_1, \dots, x_m, y_1, \dots, y_n) := \sqrt{\frac{mn}{m+n}} \frac{\bar{x}_m - \bar{y}_n}{s_{m,n}}$$

Testentscheidung:

H_0 ablehnen, falls $|T_{m,n}| \geq t_{m+n-2, 1-\frac{\alpha}{2}}$.

Kein Widerspruch zu H_0 , falls $|T_{m,n}| < t_{m+n-2, 1-\frac{\alpha}{2}}$.

Es gilt:

$$P(|T_{m,n}(X_1, \dots, X_m, Y_1, \dots, Y_n)| \geq t_{m+n-2, 1-\frac{\alpha}{2}}) = \alpha \ \forall \vartheta \in \Theta_0$$

D.h. Test hat Niveau α .

2.10 F-Test für den Varianz-Quotienten

Situation wie in 2.6, aber $Y_j \sim \mathcal{N}(\nu, \tau^2)$ ($\tau^2 \neq \sigma^2$ möglich).

Zu testen: $H_0 : \sigma^2 = \tau^2$ ($\frac{\sigma^2}{\tau^2} = 1$) gegen $H_1 : \sigma^2 \neq \tau^2$ ($\frac{\sigma^2}{\tau^2} \neq 1$).

Prüfgröße des F-Tests für Varianzquotienten ist

$$Q_{m,n} = \frac{\frac{1}{m-1} \sum_{i=1}^m (X_i - \bar{X}_m)^2}{\frac{1}{n-1} \sum_{j=1}^n (Y_j - \bar{Y}_n)^2}$$

Unter H_0 gilt $Q_{m,n} \sim F_{m-1,n-1}$.

Ablehnung von H_0 erfolgt für große und kleine Werte von $Q_{m,n}$

[Meist⁷: Ablehnung für $Q_{m,n} \geq F_{m-1,n-1,1-\frac{\alpha}{2}}$, $Q_{m,n} \leq F_{m-1,n-1,\frac{\alpha}{2}}$]

⁷→ Abbildung 2.2