Wahrscheinlichkeitsrechnung und Statistik

10. Vorlesung - 2017

Quantil der Ordnung α für die Verteilung des beobachteten Merkmals X ist der Wert $z_\alpha \in \mathbb{R}$ für welchen gilt

$$P(X < z_\alpha) \le \alpha \le P(X \le z_\alpha).$$

$z_{\frac{1}{2}}$ heißt **Median**.

• Falls X stetige zufällige Variable ist, dann z_α Quantil der Ordnung $\alpha \Longrightarrow P(X \le z_\alpha) = \alpha \Longrightarrow F_X(z_\alpha) = \alpha$

• $\alpha \cdot 100\%$ der Werte von X sind kleiner oder gleich mit z_α
• z.B. für $N(m, \sigma^2)$ berechnet man die Quantile z_α in Octave/Matlab mit $norminv(\alpha, m, \sigma)$, für $Student(n)$ mit $tinv(\alpha, n)$

Statistische Teste

- Eine Grundgesamtheit wird bezüglich des Merkmals X untersucht, die Verteilung von X hängt vom unbekannten Parameter θ ab.
- Statistischer Test: Überprüfung von Hypothesen bezüglich θ, anhand einer Stichprobe
- $x_1, \ldots, x_n \hookrightarrow$ *statistische Daten* (Beobachtungen, Stichprobenwerte) für das Merkmal X
- $X_1, \ldots, X_n \hookrightarrow$ *Stichprobenvariablen* sind *unabhängige* ZG mit *derselben Verteilung* wie X.

Statistische Teste

Seien gegeben $\alpha \in (0,1)$ das Signifikanzniveau (Irrtumswahrscheinlichkeit) und θ_0.

$$H_0 : \theta = \theta_0 \quad H_1 : \theta \neq \theta_0$$

$$H_0 : \theta \leq \theta_0 \quad H_1 : \theta > \theta_0$$

$$H_0 : \theta \geq \theta_0 \quad H_1 : \theta < \theta_0$$

H_0: Nullhypothese, H_1: alternative Hypothese
Man sucht einen kritischen Bereich (Ablehnungsbereich) $U \subset \mathbb{R}^n$ so dass für gegebenes Signifikanzniveau α gilt $P((X_1, \ldots, X_n) \notin U | H_0) = 1 - \alpha$.

Schlussfolgerung des Tests:
$(x_1, \ldots, x_n) \notin U \Rightarrow H_0$ wird angenommen
$(x_1, \ldots, x_n) \in U \Rightarrow$ man lehnt H_0 ab, zugunsten von H_1

Man testet eine Grundgesamtheit bezüglich des Merkmals X.

- Test für den theoretischen Erwartungswert $E(X)$
  - wenn die Varianz bekannt ist: Gauß Test (Z-Test)
  - wenn die Varianz unbekannt ist: Student Test (T-Test)
- Test für die theoretische Standardabweichung $\sqrt{V(X)}$ oder Varianz $V(X)$: χ^2-Test
- Test für den Anteilswert (approximativer Gauß Test)

Bei statistischen Tests geht man schrittweise vor:

- Welcher Parameter soll getestet werden?
- Was ist die Hypothese und was die Alternative?
- Welcher ist der Wert von α ?
- Welcher Test ist geeignet?
- Berechnen der Schätzfunktion anhand der statistischen Daten
- Schlussfolgerung

Test für den Erwartungswert $m = E(X)$ des beobachteten Merkmals X, wenn die Varianz des Merkmals $\sigma^2 = V(X)$ bekannt ist: Gauß Test, Z-Test

- Gegeben: $\alpha \in (0,1)$, $m_0 \in \mathbb{R}$, $\sigma > 0$.
- falls $X \sim N(m, \sigma^2)$ oder $n > 30$ und X hat eine beliebige Verteilung,

dann $\dfrac{\bar{X}_n - m}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$

- anhand der statistischen Daten $x_1, \dots, x_n$ berechnet man $z = \dfrac{\bar{x}_n - m_0}{\frac{\sigma}{\sqrt{n}}}$
- man berechnet das Quantil der Ordnung α der normalen Verteilung $N(0,1)$: $z_\alpha = norminv(\alpha, 0, 1)$

	H_0: $m = m_0$ H_1: $m \neq m_0$	H_0: $m \leq m_0$ H_1: $m > m_0$	H_0: $m \geq m_0$ H_1: $m < m_0$
Man akzeptiert H_0, wenn	$\lvert z\rvert < z_{1-\frac{\alpha}{2}}$	$z < z_{1-\alpha}$	$z > z_\alpha$
Man lehnt H_0 ab, zugunsten von H_1, wenn	$\lvert z\rvert \geq z_{1-\frac{\alpha}{2}}$	$z \geq z_{1-\alpha}$	$z \leq z_\alpha$

Test für den Erwartungswert $m = E(X)$ des Merkmals X, wenn die Varianz des Merkmals $\sigma^2 = V(X)$ unbekannt ist: Student Test, T-Test

- Gegeben: $\alpha \in (0,1)$, $m_0 \in \mathbb{R}$ (σ^2 unbekannt)
- falls $X \sim N(m, \sigma^2)$ oder $n > 30$ und X hat eine beliebige Verteilung,

dann $\dfrac{\bar{X}_n - m}{\frac{\tilde{S}_n}{\sqrt{n}}} \sim Student(n-1)$

- anhand der statistischen Daten $x_1, \ldots, x_n$ berechnet man $t = \dfrac{\bar{x}_n - m_0}{\frac{\tilde{s}_n}{\sqrt{n}}}$
- man berechnet das Quantil der Ordnung α der Studentverteilung mit $n-1$ Freiheitsgraden: $t_\alpha = tinv(\alpha, n-1)$

	$H_0\colon m = m_0$ $H_1\colon m \neq m_0$	$H_0\colon m \leq m_0$ $H_1\colon m > m_0$	$H_0\colon m \geq m_0$ $H_1\colon m < m_0$
Man akzeptiert H_0, wenn	$\lvert t \rvert < t_{1-\frac{\alpha}{2}}$	$t < t_{1-\alpha}$	$t > t_\alpha$
Man lehnt H_0 ab, zugunsten von H_1, wenn	$\lvert t \rvert \geq t_{1-\frac{\alpha}{2}}$	$t \geq t_{1-\alpha}$	$t \leq t_\alpha$

Test für Standardabweichung $\sigma = \sqrt{V(X)}$ des beobachteten Merkmals X: Chi-Quadrat Test

- Gegeben: $\alpha \in (0,1)$, $\sigma_0 > 0$
- wenn $X \sim N(m, \sigma^2)$, dann $\frac{n-1}{\sigma^2}\tilde{S}_n^2 \sim \chi^2(n-1)$
- anhand der statistischen Daten $x_1, \ldots, x_n$ berechnet man $q = \dfrac{n-1}{\sigma_0^2} \cdot \tilde{s}_n^2$
- man berechnet das Quantil der Ordnung α der χ^2 Verteilung mit $n-1$ Freiheitsgraden: $q_\alpha = chi2inv(\alpha, n-1)$

	H_0: $\sigma = \sigma_0$ H_1: $\sigma \neq \sigma_0$	H_0: $\sigma \leq \sigma_0$ H_1: $\sigma > \sigma_0$	H_0: $\sigma \geq \sigma_0$ H_1: $\sigma < \sigma_0$
Man akzeptiert H_0, wenn	$q_{\frac{\alpha}{2}} < q < q_{1-\frac{\alpha}{2}}$	$q < q_{1-\alpha}$	$q > q_\alpha$
Man lehnt H_0 ab, zugunsten von H_1, wenn	$q \notin (q_{\frac{\alpha}{2}}, q_{1-\frac{\alpha}{2}})$	$q \geq q_{1-\alpha}$	$q \leq q_\alpha$

Test für Anteilswert p des beobachteten Merkmals $X \sim Bernoulli(p)$:

Approximativer Gauß Test

- Gegeben: $\alpha \in (0,1)$, $p_0 \in (0,1)$.
- falls $X \sim Bernoulli(p)$ und $np(1-p) \geq 10$, dann $\frac{\bar{X}_n - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1)$
- anhand der statistischen Daten $x_1, \ldots, x_n$ berechnet man $z = \frac{\bar{x}_n - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$
- man berechnet das Quantil der Ordnung α der normalen Verteilung $N(0,1)$: $z_\alpha = norminv(\alpha, 0, 1)$
- Test kann durchgeführt werden, wenn $np_0(1-p_0) \geq 10$:

	H_0: $p = p_0$ H_1: $p \neq p_0$	H_0: $p \leq p_0$ H_1: $p > p_0$	H_0: $p \geq p_0$ H_1: $p < p_0$
Man akzeptiert H_0, wenn	$\lvert z\rvert < z_{1-\frac{\alpha}{2}}$	$z < z_{1-\alpha}$	$z > z_\alpha$
Man lehnt H_0 ab, zugunsten von H_1, wenn	$\lvert z\rvert \geq z_{1-\frac{\alpha}{2}}$	$z \geq z_{1-\alpha}$	$z \leq z_\alpha$

Beispiel 1: Ein Autohersteller behauptet, dass der Benzinverbrauch für einen neuen Autotyp im Mittel höchstens 6 Liter ist (pro 100 km). Dabei kann er davon ausgehen, dass der Verbrauch normalverteilt ist mit $\sigma = 0.3$ Liter. Eine Verbraucherzentrale vermutet, dass der Hersteller einen zu niedrigen Mittelwert angegeben hat und überprüft 20 Autos des neuen Typs auf ihren Verbrauch und berechnet einen empirischen Mittelwert von 6.1 Liter. a) Kann hiermit die Behauptung des Herstellers widerlegt werden? b) Wie groß muss der durchschnittliche Benzinverbrauch einer Stichprobe mit $n = 20$ und $\sigma = 0.3$ mindestens sein, damit die Behauptung des Herstellers widerlegt wird? ($\alpha = 0.01$)

Lösung: H_0: $m \leq 6$ mit H_1: $m > 6$, Varianz ist bekannt $\sigma^2 = 0.09$, $n = 20$, $\bar{x}_n = 6.1$

a) $z = \dfrac{\bar{x}_n - m_0}{\frac{\sigma}{\sqrt{n}}} = \dfrac{6.1 - 6}{\frac{0.3}{\sqrt{20}}} \approx 1.4907 < z_{1-\alpha} = norminv(1-\alpha) \approx 2.3263$

$\Rightarrow$ H_0 wird akzeptiert $\Rightarrow$ die Behauptung des Herstellers kann nicht widerlegt werden

b) $z = \dfrac{\bar{x}_n - m_0}{\frac{\sigma}{\sqrt{n}}} = \dfrac{\bar{x}_n - 6}{\frac{0.3}{\sqrt{20}}} \geq z_{1-\alpha} = norminv(1-\alpha) \approx 2.3263 \Rightarrow \bar{x}_n \geq 6 + z_{1-\alpha} \cdot \frac{0.3}{\sqrt{20}} \approx 6.1561$

Beispiel 2: Die Anleitungen eines Medikaments geben an, dass jede Tablette durchschnittlich 2.4 g aktive Substanzen enthält. 100 zufällig gewählte Tabletten werden untersucht und man stellt fest, dass im Mittel 2.5 g aktive Substanzen enthalten mit einer Standardabweichung von 0.2 g. Kann man behaupten, dass das Medikament die Angaben respektiert? ($\alpha = 0.01$)

Lösung: H_0: $m = 2.4$ mit H_1: $m \neq 2.4$, Varianz ist unbekannt, $n = 100$, $\bar{x}_n = 2.5, \tilde{s}_n = 0.2$

$$\Rightarrow t = \frac{\bar{x}_n - m_0}{\frac{\tilde{s}_n}{\sqrt{n}}} = \frac{2.5 - 2.4}{\frac{0.2}{\sqrt{100}}} = 5 > t_{1-\alpha/2} = tinv(1 - \alpha/2, n-1) \approx 2.6264$$

$\Rightarrow H_0$ wird abgelehnt $\Rightarrow$ die Angaben werden nicht respektiert

Beispiel 3: Eine Münze wurde 100-mal geworfen und man erhielt 61-mal "Kopf". Wir sollen testen, ob die Münze fair ist, in dem Sinne, dass die Wahrscheinlichkeit für das Auftreten von "Kopf" gleich der Wahrscheinlichkeit für das Auftreten von "Zahl" ist, d.h. man führt einen Test bzgl. dem Anteilswert p durch, zu testen ist $p = 0.5$, gegen $p \neq 0.5$. Man wählt die Irrtumswahrscheinlichkeit $\alpha = 0.05$.

Lösung: $np_0(1 - p_0) = 100 \cdot 0.5 \cdot 0.5 \geq 10$
$H_0 : p = 0.5$, $H_1 : p \neq 0.5$, Test für den Anteilswert p

$$z = \frac{\bar{x}_n - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{\frac{61}{100} - 0.5}{\sqrt{\frac{0.5(1-0.5)}{100}}} = 2.2$$

$z > z_{1-\frac{\alpha}{2}} = norminv(1 - \frac{0.05}{2}, 0, 1) = 1.96 \implies H_0$ wird abgelehnt; anhand der Daten schlußfolgert man, dass die Münze nicht fair ist.