

Einführung in die Induktive Statistik: Testen von Hypothesen

Jan Gertheiss
LMU München

Sommersemester 2011

Vielen Dank an Christian Heumann für das Überlassen von T_EX-Code!

Testen: Einführung und Konzepte

- ▶ Einführung: Gauß-, Student- und Binomialtest
- ▶ Prinzipien des Testens von Hypothesen

Ziele des Kapitels:

Exemplarische Einführung in das Testen von Hypothesen, Beschreibung der generellen Konzepte.

Gauß-Test, Student-Test und Binomialtest

Gauß- und (Student) t-Test: bekannteste Tests zum Prüfen von Hypothesen über $\mu = E(X)$.

Allgemeine Form der Hypothesen über $\mu = E(X)$:

- (a) $H_0 : \mu = \mu_0, H_1 : \mu \neq \mu_0$ zweiseitige Alternative H_1
- (b) $H_0 : \mu = \mu_0, H_1 : \mu < \mu_0$ einseitige Alternative H_1
- (c) $H_0 : \mu = \mu_0, H_1 : \mu > \mu_0$ einseitige Alternative H_1

Bemerkungen:

- ▶ Verschiedene Tests unterscheiden sich durch Annahmen über X .
- ▶ Binomialtest: Testen von analogen Hypothesen über $\pi = P(X = 1)$ bei Bernoulli-Variable.

(Exakter) Gauß-Test

- ▶ Annahmen: $X \sim N(\mu, \sigma^2)$ mit bekannter Varianz σ^2 , Stichprobenvariablen $X_1, \dots, X_n$ i.i.d. wie X .
- ▶ Hypothesen über $\mu = E(X)$: (a), (b), (c) wie angegeben.
- ▶ Idee für Test: Falls H_0 richtig ist: $E(X) = \mu_0$. Bilde arithmetisches Mittel $\bar{x}$ zu den Stichprobenwerten $x_1, \dots, x_n$. Lehne H_0 ab, falls Abweichung zwischen μ_0 und $\bar{x}$ zu groß.
- ▶ Frage: Wie groß sind die kritischen Werte für diese Abweichung zu wählen?

(Exakter) Gauß-Test

Diskussion für Hypothesenpaar (c)

$$H_0 : \mu = \mu_0, \quad H_1 : \mu > \mu_0$$

Übergang von $\bar{X}$ zu standardisierter Teststatistik

$$Z = \frac{\bar{X} - \mu_0}{\sigma} \sqrt{n}$$

$$\text{Unter } H_0 \text{ gilt: } \bar{X} \sim N(\mu_0, \frac{\sigma^2}{n}) \quad \Rightarrow \quad Z \sim N(0, 1)$$

$$\text{Testvorschrift für } Z: H_0 \text{ ablehnen} \quad \Leftrightarrow \quad Z > k$$

(Exakter) Gauß-Test

Frage: Wie ist der kritische Wert k zu wählen?

Prinzip: Die Wahrscheinlichkeit für den

Fehler 1. Art: H_0 wird abgelehnt, obwohl H_0 richtig ist

soll (höchstens) gleich einem (kleinen) vorgegebenen *Signifikanzniveau* α sein (z.B. $\alpha = 0.1, 0.05, 0.01$). D.h.

$$P(H_0 \text{ ablehnen} \mid H_0 \text{ richtig}) = \alpha$$

(Exakter) Gauß-Test

Beim exakten Gauß-Test ist dies äquivalent zu

$$P(Z > k \mid \mu = \mu_0) = \alpha$$

$$\Leftrightarrow k = z_{1-\alpha} \text{ (mit } z_{1-\alpha} = (1 - \alpha)\text{-Quantil der Standardnormalverteilung)}$$

Testvorschrift: H_0 ablehnen, falls $Z > z_{1-\alpha}$

(Exakter) Gauß-Test

Bemerkungen:

- ▶ Neben dem Fehler 1. Art gibt es den
Fehler 2. Art: H_0 wird nicht abgelehnt, obwohl H_1 richtig ist.
Es gilt (vergleiche Gütefunktion, später): Je kleiner (größer) das Signifikanzniveau α gewählt wird, desto größer (kleiner) wird die Wahrscheinlichkeit für einen Fehler 2. Art.
- ▶ Falls $H_0 : \mu \leq \mu_0$ gilt, also auch $\mu < \mu_0$ als Nullhypothese möglich ist, folgt

$$P(\text{Fehler 1. Art}) = P(Z > z_{1-\alpha} \mid \mu \leq \mu_0) \leq \alpha$$

(Exakter) Gauß-Test

Hypothesenpaar (b)

$H_0 : \mu = \mu_0, \mu < \mu_0$ symmetrisch zu (c).

$\Rightarrow$ **Testvorschrift:** H_0 ablehnen, falls $Z < -z_{1-\alpha}$

Hypothesenpaar (a)

$H_0 : \mu = \mu_0, H_1 : \mu \neq \mu_0$

Unter $H_0 : Z \sim N(0, 1)$

$\Rightarrow$ **Testvorschrift:** H_0 ablehnen, falls $|Z| > z_{1-\frac{\alpha}{2}}$

(Exakter) Gauß-Test

Alternative Formulierungen der Testentscheidungen:

- (1) Direkt über $\bar{X}$ statt Z
- (2) Mit Hilfe von Konfidenzintervallen
- (3) Mit p -Werten (“Überschreitungswahrscheinlichkeiten”)

(Exakter) Gauß-Test

(1) Für (a) $H_0 : \mu = \mu_0$, $H_1 : \mu \neq \mu_0$:

$$H_0 \text{ ablehnen, falls } |Z| > z_{1-\frac{\alpha}{2}}, \quad Z = \frac{\bar{X} - \mu_0}{\sigma} \sqrt{n}$$

$$\Leftrightarrow H_0 \text{ ablehnen, falls } |\bar{X} - \mu_0| > z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

$$\Leftrightarrow H_0 \text{ ablehnen, falls } \bar{X} < \mu_0 - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \text{ oder } \bar{X} > \mu_0 + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

(2) Testentscheidung oben offensichtlich äquivalent zu

H_0 ablehnen, falls $\mu_0 \notin \text{KI}$ für μ , d.h.

$$\mu_0 \notin \left[\bar{X} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right]$$

Bemerkung: (1) und (2) lassen sich auch für einseitige Problemstellungen (b) und (c) formulieren.

(Exakter) Gauß-Test

(3) Testentscheidungen mit Überschreitungswahrscheinlichkeiten

Der p -Wert (p -value) ist definiert als die Wahrscheinlichkeit, unter H_0 den beobachteten Prüfgrößenwert oder einen in Richtung der Alternative extremeren Wert zu erhalten.

Ist der p -Wert kleiner (oder gleich) dem vorgegebenen Signifikanzniveau α , wird H_0 verworfen. Ansonsten behält man H_0 bei.

(Exakter) Gauß-Test

Bemerkungen:

- ▶ Statistische Programmpakete geben in der Regel p -Werte für zweiseitige Tests aus.

Dann: H_0 ablehnen $\Leftrightarrow$ " p -value" $< \alpha$, α vorgegebenes Signifikanzniveau.

- ▶ Vorsicht bei einseitigen Tests zu (b) und (c)! p -Werte müssen für Testentscheidung ggf. modifiziert werden.

Approximativer Gauß-Test

Annahmen:

X beliebig verteilt mit $E(X) = \mu$; $Var(X) = \sigma^2$ bekannt.

$X_1, \dots, X_n$ i.i.d. wie X ; Faustregel: $n \geq 30$.

Wegen zentralem Grenzwertsatz: Unter $\mu = \mu_0$ gilt

$$\bar{X} \overset{a}{\sim} N\left(\mu_0, \frac{\sigma^2}{n}\right) \quad \text{bzw.} \quad Z \overset{a}{\sim} N(0, 1)$$

Testvorschrift wie beim exakten Gauß-Test.

Aber: $P(\text{Fehler 1. Art}) \overset{a}{\leq} \alpha$

(Student) t-Test

- ▶ Annahmen: Wie beim (exakten) Gauß-Test, aber: σ^2 unbekannt.
- ▶ Hypothesen: (a), (b), (c) wie bisher.
- ▶ Idee: Ersetze σ (beim Gauß-Test) durch

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2},$$

d.h. Teststatistik

$$Z = \frac{\bar{X} - \mu_0}{\sigma} \sqrt{n}$$

wird ersetzt durch Teststatistik

$$T = \frac{\bar{X} - \mu_0}{S} \sqrt{n}$$

(Student) t-Test

Man kann zeigen: $X \sim N(\mu_0, \sigma^2) \Rightarrow T \sim t(n-1)$ (Student) t-verteilt mit $n-1$ Freiheitsgraden.

$\Rightarrow$ Herleitung der Testvorschriften wie beim Gauß-Test; ersetze Z durch T und die Dichte ϕ von Z durch Dichte der $t(n-1)$ -Verteilung.

$\Rightarrow$ Ersetze in Testvorschriften Z durch T und z-Quantile durch $t(n-1)$ -Quantile.

Für $n \geq 30$: $t(n-1)$ -Quantile $\approx$ z-Quantile, und T näherungsweise normalverteilt, auch wenn X nicht normalverteilt.

Binomial-Test

Annahmen:

$X \sim B(1, \pi)$, d.h. $P(X = 1) = \pi$; $X_1, \dots, X_n$ i.i.d. wie X ;
für approximativen Binomial-Test: $n \geq 30$.

Hypothesen:

- (a) $H_0 : \pi = \pi_0, H_1 : \pi \neq \pi_0$
- (b) $H_0 : \pi \geq \pi_0, H_1 : \pi < \pi_0$
- (c) $H_0 : \pi \leq \pi_0, H_1 : \pi > \pi_0$

Binomial-Test

Teststatistik: $X = X_1 + \dots + X_n$ absolute Häufigkeit von Einsen;

unter $\pi = \pi_0$: $X \sim B(n, \pi_0)$

Alternativ: Standardisierte Teststatistik:

$$Z = \frac{X - n\pi_0}{\sqrt{n\pi_0(1 - \pi_0)}} = \frac{\hat{\pi} - \pi_0}{\sqrt{\frac{\pi_0(1 - \pi_0)}{n}}}$$

- ▶ $n < 30$: exakter Binomial-Test nötig, verwende X .
- ▶ $n \geq 30$: approximativer Binomial-Test unter Verwendung von Z ;
für $\pi = \pi_0$: $Z \stackrel{a}{\sim} N(0, 1)$ (zentraler Grenzwertsatz).
 $\Rightarrow$ gleiche Testvorschriften wie beim approximativen Gauß-Test.

Prinzipien des Testens von Hypothesen

Parameter-Tests

Generelle Problemstellung

$$X \sim F(x | \theta), \theta \in \Theta$$

θ unbekannter Parameter/Kennwert

Θ zulässiger Bereich für θ

Nullhypothese $H_0 : \theta \in \Theta_0$

Alternativhypothese $H_1 : \theta \in \Theta_1$

$$\Theta_0 \cap \Theta_1 = \emptyset$$

$X_1, \dots, X_n$ i.i.d. wie X

$T = t(X_1, \dots, X_n)$ Teststatistik

Beispiel: Gauß-Test

$$X \sim N(\mu, \sigma^2), \sigma^2 \text{ bekannt}$$

$$\theta = \mu = E(X)$$

$$\Theta = \mathbb{R}$$

$$(a) \Theta_0 = \{\mu_0\}, \Theta_1 = \mathbb{R} \setminus \{\mu_0\}$$

$$(c) \Theta_0 = (-\infty, \mu_0], \Theta_1 = (\mu_0, \infty)$$

$$X_1, \dots, X_n \text{ i.i.d. } N(\mu, \sigma^2)$$

$$T = Z = \frac{(\bar{X} - \mu_0)}{\sigma} \sqrt{n}$$

Prinzipien des Testens von Hypothesen

Nichtparametrische Tests

$H_0 : X$ normalverteilt, $H_1 : X$ nicht normalverteilt

$H_0 : X, Y$ unabhängig, $H_1 : X, Y$ abhängig

usw. ($\rightarrow$ später)

Prinzipien des Testens von Hypothesen

Struktur von Tests

1. Inhaltliches Problem als Testproblem formulieren, Annahmen über $X, Y, \dots$ festlegen. H_0 und H_1 bilden.
2. Signifikanzniveau α festlegen.
3. Prüfgröße/Teststatistik T festlegen
 - ▶ T muss für sensibel für das Testproblem sein.
 - ▶ Verteilung von T unter H_0 muss (approximativ) bekannt sein.
4. Ablehnbereich festlegen.
5. Daten aus Stichprobe erheben und Wert der Prüfgröße T berechnen.
6. Testentscheidung: H_0 ablehnen, falls T in kritischen Bereich fällt.

Prinzipien des Testens von Hypothesen

Fehlentscheidungen:

	H_0 ablehnen	H_0 nicht ablehnen
H_0 richtig	Fehler 1. Art	
H_1 richtig		Fehler 2. Art

Alle Signifikanztests garantieren:

$$P(\text{Fehler 1. Art}) \leq \alpha$$

Aber: $P(\text{Fehler 2. Art}) = P(H_0 \text{ nicht ablehnen} \mid H_1 \text{ richtig}) = ?$

Prinzipien des Testens von Hypothesen

Fehlentscheidungen:

	H_0 ablehnen	H_0 nicht ablehnen
H_0 richtig	Fehler 1. Art	
H_1 richtig		Fehler 2. Art

Alle Signifikanztests garantieren:

$$P(\text{Fehler 1. Art}) \leq \alpha$$

Aber: $P(\text{Fehler 2. Art}) = P(H_0 \text{ nicht ablehnen} \mid H_1 \text{ richtig}) = ?$

Gütefunktion:

Die Gütefunktion $g(\theta)$, $\theta \in \Theta$, fasst $P(\text{Fehler 1. Art})$ und $P(\text{Fehler 2. Art})$ in einer Funktion zusammen:

$$g(\theta) := P(H_0 \text{ ablehnen} \mid \theta)$$