Statistische Tests
Übersicht

- 1. Einführung und Übersicht
- 2. Das Einstichprobenproblem
- 3. Vergleich zweier unabhängiger Gruppen (unverbundene Stichproben)
- 4. Vergleich zweier abhängiger Gruppen (verbundene Stichproben)
- 5. Vergleich mehrerer unabhängiger Gruppen (einfache Varianzanalyse)
- 6. Vergleich mehrerer abhängiger Gruppen (einfaches Blockexperiment)
- 7. Weitere Varianzanalysemodelle
- 8. Anpassungstests
- 9. Nichtparametrische Tests

Sei X ein Merkmal (eine Zufallsvariable),
$F_X(x) = P(X \leq x) = P_\theta(X \leq x) = F_{X,\theta}(x)$
θ: Parametervektor

Beispiel: $\theta = (\mu, \sigma^2)$

μ: Erwartungswert von X
σ^2: Varianz von X

$X_1, X_2, \ldots, X_n$ Beobachtungen von X

$\mu \approx \frac{1}{n} \sum_{i=1}^{n} X_i = \overline{X}$
$\sigma^2 \approx \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2 = s^2$

D.h. die unbekannten Parameter werden geschätzt.

Statistische Tests: Einführung

Problem

Schätzungen können sehr schlecht ausfallen!

I.a. vertritt der Fachexperte gewisse Hypothesen bzgl. der (unbekannten) Parameterwerte!

Diese Hypothesen werden verworfen, wenn die erhaltenen <u>Schätzwerte</u> (z.B. $\overline{X}, s^2$) mit ihnen nicht in Einklang stehen.

Statistische Tests: Einführung

Eine verwandte Problemstellung

Elektronischer Großhandel: TV-Geräte

Händler sagt: Ausschußquote $p \leq 1\%$ $(p = 0.01)$
Käufer wäre einverstanden, prüft aber N Geräte!
Davon: N_f fehlerhaft, N_f - Teststatistik

$$\frac{N_f}{N} \cdot 100\% \gg 1\% \Rightarrow \text{ Ablehnung}$$

Zwei Fehler möglich

a) Zufällig N_f zu groß! $\qquad p < 0.01$
$\Rightarrow$ Käufer lehnt ab

b) Zufällig N_f zu klein! $\qquad p$ groß, $p \gg 0.01$
$\Rightarrow$ Käufer kauft

Risiko des Händlers

Käufer lehnt gute Ware ab (weil N_f zufällig zu groß)

Statistische Tests: Einführung

Risiken - Fehler

Risiko des Händlers

Käufer lehnt gute Ware ab (weil N_f zufällig zu groß)

Risiko des Käufers

Käufer kauft schlechte Ware (weil N_f zufällig zu klein)

Risiko des Händlers

Käufer lehnt gute Ware ab (weil N_f zufällig zu groß)

Risiko des Käufers

Käufer kauft schlechte Ware (weil N_f zufällig zu klein)

Risiken sollen quantifiziert werden:

a) $P($ Nicht kaufen $| \; p \leq 1\%)$

b) $P($ Kaufen $| \; p > 1\%)$

Beide Risiken nicht gleichzeitig zu minimieren.

Statistische Tests: Einführung

Risiken - Fehler

Risiko des Händlers

Käufer lehnt gute Ware ab (weil N_f zufällig zu groß)

Risiko des Käufers

Käufer kauft schlechte Ware (weil N_f zufällig zu klein)

Risiken sollen quantifiziert werden:

a) $P($ Nicht kaufen $\mid p \leq 1\%)$

b) $P($ Kaufen $\mid p > 1\%)$

Beide Risiken nicht gleichzeitig zu minimieren.

Lösung:

$P($ Nicht kaufen $\mid p \leq 1\%) = \alpha$ vorgeben

$P($ Kaufen $\mid p > 1\%)$ minimieren (oder es versuchen)

Hypothesentest

Beispiel: Einstichproben-Lagetest

Sei μ ein Lageparameter, z.B. der Erwartungswert.
Sei μ_0 ein vorgegebener Wert.

Nullhypothese und Alternativhypothese

a) $H_0 : \quad \mu \leq \mu_0 \qquad H_A : \mu > \mu_0$
b) $H_0 : \quad \mu \geq \mu_0 \qquad H_A : \mu < \mu_0$
c) $H_0 : \quad \mu = \mu_0 \qquad H_A : \mu \neq \mu_0$

Sei μ ein Lageparameter, z.B. der Erwartungswert.
Sei μ_0 ein vorgegebener Wert.

Nullhypothese und Alternativhypothese

a) $H_0 : \quad \mu \leq \mu_0 \qquad H_A : \mu > \mu_0$
b) $H_0 : \quad \mu \geq \mu_0 \qquad H_A : \mu < \mu_0$
c) $H_0 : \quad \mu = \mu_0 \qquad H_A : \mu \neq \mu_0$

Teststatistik

$$T(X_1, ..., X_n) = \frac{\overline{X} - \mu_0}{s} \cdot \sqrt{n}$$

T heißt auch Testgröße, Prüfgröße, Stichprobenfunktion.

Die Entscheidung für H_A oder für H_0 wird anhand einer Teststatistik

$$T = T(x_1, ..., x_n)$$

gefällt.
Liegt der Wert von T in einem vorher bestimmten Bereich K, dem sogen. <u>Ablehnungsbereich</u> oder <u>kritischen Bereich</u>, dann wird H_0 abgelehnt, anderenfalls wird H_0 nicht abgelehnt.

$T \in K \Rightarrow H_0$ ablehnen, Entscheidung für H_A
$T \notin K \Rightarrow H_0$ nicht ablehnen, Entscheidung für H_0.

a) $H_0 : \quad \mu \leq \mu_0 \qquad H_A : \mu > \mu_0$

große Werte von T sprechen für H_A.

Annahmebereich | Krit.Bereich

t_{krit}

b) $H_0 \quad \mu \geq \mu_0 \qquad H_A : \quad \mu < \mu_0$

kleine Werte von T sprechen für H_A.

Krit.B. | Annahmebereich

t_{krit}

c) $H_0 : \quad \mu = \mu_0 \qquad H_A : \quad \mu \neq \mu_0$

große Werte von $|T|$ sprechen für H_A.

Annahmebereich

$-t_{krit}$ | t_{krit}

Fehler 1.Art

Entscheidung für H_A obwohl H_0 richtig ist.

Fehler 1.Art

Entscheidung für H_A obwohl H_0 richtig ist.

Fehler 2.Art

Entscheidung für H_0 obwohl H_A richtig ist

Hypothesentest

Fehler 1. Art, Fehler 2. Art

Fehler 1.Art

Entscheidung für H_A obwohl H_0 richtig ist.

Fehler 2.Art

Entscheidung für H_0 obwohl H_A richtig ist

	Entscheidung für H_0	Entscheidung für H_A
H_0 richtig	richtig, Sicherheitswkt. $1-\alpha$	Fehler 1. Art Fehlerwkt. α.
H_A richtig	Fehler 2.Art Fehlerwkt. 1-β	richtig, Güte β

Hypothesentest

Fehler 1. Art, Fehler 2. Art

Fehler 1.Art

Entscheidung für H_A obwohl H_0 richtig ist.

Fehler 2.Art

Entscheidung für H_0 obwohl H_A richtig ist

	Entscheidung für H_0	Entscheidung für H_A
H_0 richtig	richtig, Sicherheitswkt. $1-\alpha$	Fehler 1. Art Fehlerwkt. α.
H_A richtig	Fehler 2.Art Fehlerwkt. 1-β	richtig, Güte β

Entscheidung für H_0 heißt nicht notwendig, dass H_0 richtig ist.

α und $(1-\beta)$ können nicht gleichzeitig minimiert werden.
$\Rightarrow$ Man gibt α vor (z.B. $\alpha = 0.05$), d.h. man behält α unter Kontrolle und versucht die Teststatistik so zu definieren, daß β maximal wird.

β (und manchmal auch α) hängen von wahren (i.A. unbekannten) Parametern ab.

Signifikanzniveau

$$\alpha = sup_{\theta\in\mathbf{\Theta}_0}\beta(\theta).$$

$\mathbf{\Theta}_0$: Nullhypothesenraum, also z.B. die Menge $\{\mu : \mu \geq \mu_0\}$ oder $\{\mu : \mu = \mu_0\}$.

Gütefunktion

Gütefunktion

$$\beta = \beta(\theta) = \beta(\mu) = P_\mu(T \in K)$$

K heißt Ablehnungsbereich oder Kritischer Bereich.

Beispiel: t-Test

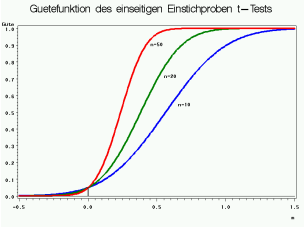
$$\begin{aligned}\beta(\mu) &= P(T \in K) \qquad K\text{: kritischer Bereich} \\ &= P(T > t_{1-\alpha,n-1} | \mu, \sigma^2) \\ &= 1 - CDF('T', t_{1-\alpha,n-1}, n-1, nc)\end{aligned}$$

$nc = \sqrt{n}\frac{\mu-\mu_0}{\sigma}$: Nichtzentralitätsparameter

$t_{1-\alpha,n-1}$: kritischer Wert

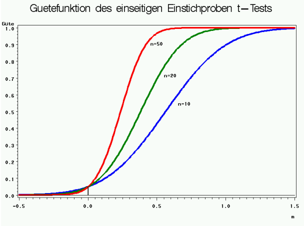
$K = [t_{1-\alpha,n-1}, \infty)$: kritischer Bereich.

Einseitiger Test

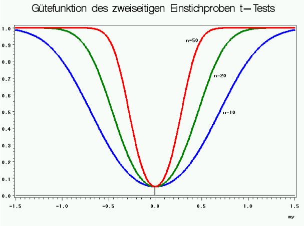


Gütefunktion

Einseitiger Test

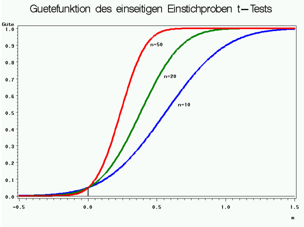


Zweiseitiger Test



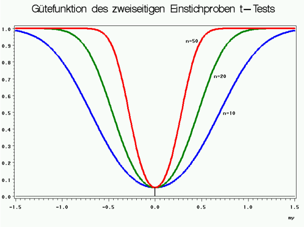
Gütefunktion

Einseitiger Test



`Test_Guete_t.sas`

Zweiseitiger Test



`Test_Guete_t2.sas`

Gütefunktion

Ideal:

Unter H_0: Güte 0 (d.h. Fehler 1. Art =0)
Unter H_A: Güte 1 (d.h. Fehler 2. Art =0)

Das ist aber nicht möglich!

Ziel:

Test mit möglichst großer Gütefunktion (unter H_A).

Wir schlagen natürlich nur solche “sinnvollen” Tests vor.

Lagetests
(bei Normalverteilungsannahme)

Einstichprobenproblem

$$H_0 : \mu \leq \mu_0 \quad H_A : \mu > \mu_0$$
$$H_0 : \mu \geq \mu_0 \quad H_A : \mu < \mu_0$$
$$H_0 : \mu = \mu_0 \quad H_A : \mu \neq \mu_0$$

Lagetests
(bei Normalverteilungsannahme)

Einstichprobenproblem

$H_0 : \mu \leq \mu_0$	$H_A : \mu > \mu_0$	Einstichproben t-Test
$H_0 : \mu \geq \mu_0$	$H_A : \mu < \mu_0$	**PROC UNIVARIATE**
$H_0 : \mu = \mu_0$	$H_A : \mu \neq \mu_0$	**PROC TTEST**

Lagetests

(bei Normalverteilungsannahme)

Einstichprobenproblem

$H_0 : \mu \leq \mu_0 \quad H_A : \mu > \mu_0$ Einstichproben t-Test

$H_0 : \mu \geq \mu_0 \quad H_A : \mu < \mu_0$ **PROC UNIVARIATE**

$H_0 : \mu = \mu_0 \quad H_A : \mu \neq \mu_0$ **PROC TTEST**

Zweistichprobenproblem

$H_0 : \mu_1 \leq \mu_2 \quad H_A : \mu_1 > \mu_2$

$H_0 : \mu_1 \geq \mu_2 \quad H_A : \mu_1 < \mu_2$

$H_0 : \mu_1 = \mu_2 \quad H_A : \mu_1 \neq \mu_2$

Lagetests
(bei Normalverteilungsannahme)

Einstichprobenproblem

$H_0 : \mu \leq \mu_0 \quad H_A : \mu > \mu_0$
$H_0 : \mu \geq \mu_0 \quad H_A : \mu < \mu_0$
$H_0 : \mu = \mu_0 \quad H_A : \mu \neq \mu_0$

Einstichproben t-Test
PROC UNIVARIATE
PROC TTEST

Zweistichprobenproblem

$H_0 : \mu_1 \leq \mu_2 \quad H_A : \mu_1 > \mu_2$
$H_0 : \mu_1 \geq \mu_2 \quad H_A : \mu_1 < \mu_2$
$H_0 : \mu_1 = \mu_2 \quad H_A : \mu_1 \neq \mu_2$

Einstichproben t-Test (verbundene Stichproben)
t-Test (unverbundene Stichproben)
PROC UNIVARIATE
PROC TTEST

c-Stichprobenproblem

$H_0: \mu_1 = \ldots = \mu_c \qquad H_A: \exists(i,j): \mu_i \neq \mu_j$

einfache Varianzanalyse

PROC ANOVA, PROC GLM

Andere Alternativen sind:

$$\mu_1 \leq \ldots \leq \mu_c$$
$$\mu_1 \geq \ldots \geq \mu_c$$

Lage- und Skalentests
(bei Normalverteilungsannahme)

c-Stichprobenproblem

$H_0 : \mu_1 = ... = \mu_c \qquad H_A : \exists(i,j) : \mu_i \neq \mu_j$

einfache Varianzanalyse

PROC ANOVA, PROC GLM

Andere Alternativen sind:

$$\mu_1 \leq ... \leq \mu_c$$
$$\mu_1 \geq ... \geq \mu_c$$

Skalentest

Zwei unverbundene Stichproben

$$H_0 : \sigma_1^2 = \sigma_2^2 \qquad H_A : \sigma_1^2 \neq \sigma_2^2$$

PROC TTEST (bei Normalverteilung)

p-Werte

bisher: "H_0 abgelehnt" oder "H_0 beibehalten"
$\Rightarrow$ wenig informativ.

Wir könnten uns auch bei jedem α fragen, ob H_0 abgelehnt wird oder nicht.

Wenn der Test bei Signifikanzniveau α ablehnt, wird er das auch für $\alpha' > \alpha$ tun.

Es gibt also ein kleinstes α, bei dem der Test H_0 ablehnt.

Der p-Wert

ist das kleinste α, bei dem wir H_0 ablehnen können.

```
Test_t_p_value
```

p-Wert

T: (zufällige) Teststatistik, t: beobachtete Teststatistik

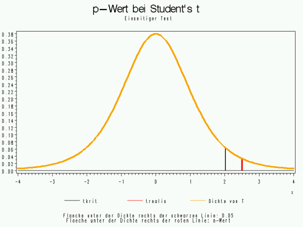
Zweiseitige Alternative: $\mu \neq \mu_0$
p-Wert $= P_0(|T| > |t|)$

Einseitige Alternative: $\mu < \mu_0$
p-Wert $= P_0(T < t)$

Einseitige Alternative: $\mu > \mu_0$
p-Wert $= P_0(T > t)$

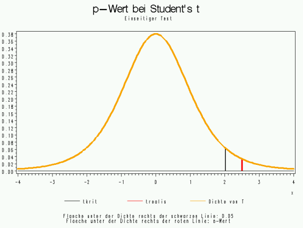
Der p-Wert heißt auch Überschreitungswahrscheinlichkeit.

Einseitiger Test

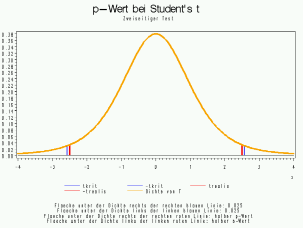


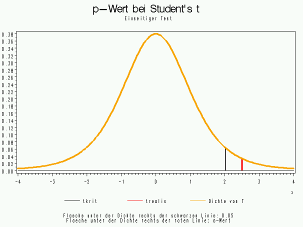
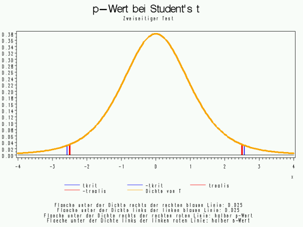
p-Wert
Illustration

Einseitiger Test



Zweiseitiger Test



Einseitiger Test	Zweiseitiger Test
	

Fäche unter der Dichte rechts der schwarzen Linie:

0.05	0.025

Fäche unter der Dichte rechts der roten Linie:

p-Wert	halber p-Wert

links entsprechend.

Bewertung von p-Werten

Der p-Wert ist also, grob, ein Maß für den Grad dafür, dass die Nullhypothese nicht zutrifft.

(vorsichtige) Interpretation

p-Wert	Grad des Nicht-Zutreffens von H_0
< 0.01	sehr streng gegen H_0
$0.01 \ldots 0.05$	streng gegen H_0
$0.05 \ldots 0.1$	schwach gegen H_0
> 0.1	wenig oder gar nichts gegen H_0

Bewertung von p-Werten

Der p-Wert ist also, grob, ein Maß für den Grad dafür, dass die Nullhypothese nicht zutrifft.

(vorsichtige) Interpretation

p-Wert	Grad des Nicht-Zutreffens von H_0
< 0.01	sehr streng gegen H_0
$0.01 \ldots 0.05$	streng gegen H_0
$0.05 \ldots 0.1$	schwach gegen H_0
> 0.1	wenig oder gar nichts gegen H_0

Warnung:

Ein großer p-Wert heisst noch lange nicht, dass H_0 zutrifft. H_0 kann zutreffen,
Der große p-Wert kann aber auch daran liegen, dass der Test niedrige Güte hat!

p-Wert und kritischer Wert

Einseitiger Test, $t_{krit} = t_{1-\alpha}$

$t \leq t_{krit} \Leftrightarrow$ p-Wert $\geq \alpha \Longrightarrow H_0$ angenommen,
$t > t_{krit} \Leftrightarrow$ p-Wert $< \alpha \Longrightarrow H_0$ abgelehnt.

Zweiseitiger Test, $t_{krit} = t_{1-\alpha/2}$

$|t| \leq t_{krit} \Leftrightarrow$ p-Wert $\geq \alpha \Longrightarrow H_0$ angenommen,
$|t| > t_{krit} \Leftrightarrow$ p-Wert $< \alpha \Longrightarrow H_0$ abgelehnt.

Ausgabe bei SAS

Wenn nicht anders vermerkt: zweiseitige p-Werte.