Statistische Tests

Testen von Hypothesen
Fehlerarten
wichtigste statistische Tests

Hypothesen

Jeder statistische Test beruht auf der Widerlegung einer zuvor aufgestellten Hypothese. Die Widerlegung ist allerdings nie zu 100% korrekt sondern mit einer bestimmten Fehlerrate behaftet.

Generelle Vorgangsweise

- Formulierung der Null- und der Alternativhypothese
- Spezifikation des Signifikanzniveaus
- Berechnung der Testgröße (engl. test statistic)
- Definition des Rückweisungsbereichs
- Selektion der passenden Hypothese

Statistische Tests funktionieren generell nach obigem Schema; der einzige Unterschied besteht in der unterschiedlichen Berechnung der Testgröße.

Hypothesen - H_0 & H_1

Null- und Alternativhypothese

zwei Hypothesen, die sich gegenseitig ausschließen und die Gesamtgrundheit {G} zur Gänze umfassen

Nullhypothese H_0

- Annahme, die überprüft werden soll
- Nullhypothese H_0 kann nur abgelehnt (widerlegt) werden, nicht jedoch angenommen

Alternativhypothese H_1

ACHTUNG: H_1 tritt nicht automatisch in Kraft falls H_0 abgelehnt wird!

$\{H_0\} \cap \{H_1\} = \{0\}$

$\{H_0\} \cup \{H_1\} = \{G\}$

Ablehnung der Nullhypothese

Eine Nicht-Ablehnung der Nullhypothese ist nicht gleichbedeutend mit der Annahme der Nullhypothese. Ein statistischer Test kann eine Nullhypothese nie annehmen, sondern (gegebenenfalls) nur ablehnen.

Ablehnung:

> "*Aufgrund der vorliegenden Daten muss die Nullhypothese abgelehnt werden*"

Nicht-Ablehnung:

> "*Aufgrund der vorliegenden Daten und des vorgegebenen Signifikanzniveaus gibt es keinen Grund zur Annahme, dass die Nullhypothese falsch ist*".

In beiden Fällen passieren Fehler, die durch das Signifikanzniveau bestimmt werden.

Fehlerarten

Wirklichkeit:

2 Möglichkeiten – H_0 wahr oder falsch

Schlussfolgerung aus Testergebnis:

wahr oder falsch

--> vier Möglichkeiten , nur zwei sind OK

		Wirklichkeit	
		H_0 = wahr	H_0 = falsch
Schluss-folgerung	H_0 = wahr	OK	Fehler vom Typ II
	H_0 = falsch	Fehler vom Typ I	OK

Fehler 1. Art

Die Nullhypothese wurde irrtümlich abgelehnt

Fehler 2. Art

Die Nullhypothese wurde irrtümlich nicht abgelehnt

[GrundStat: TYPTSTERR] Direktlink: Fehlerarten

Signifikanzniveau

Wie groß ist die Wahrscheinlichkeit, dass ein statistischer Test irrt?

Unterscheidung der beiden Fehlerarten:

Fehler 1. Art: Signifikanzniveau α bestimmt die Wahrscheinlichkeit die Nullhypothese irrtümlich abzulehnen

Fehler 2. Art: Die Wahrscheinlichkeit ß einer irrtümlichen Nicht-Ablehnung der Nullhypothese wird indirekt ebenfalls durch das Signifikanzniveau bestimmt. Allerdings lässt sich diese Wahrscheinlichkeit in der Praxis nicht berechnen, sondern bestenfalls abschätzen.

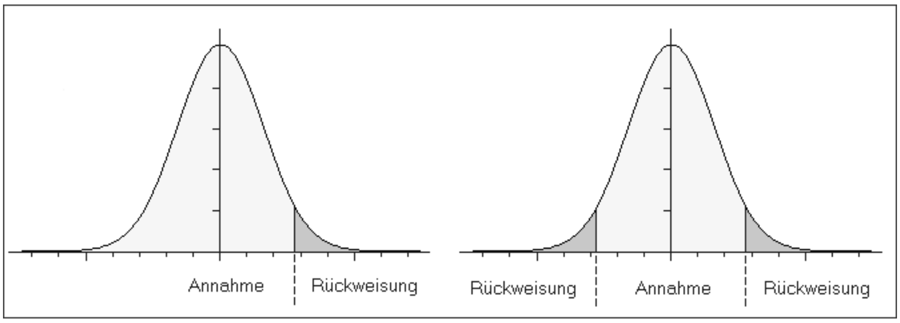
Einseitige und zweiseitige Tests

Je nach Art der Fragestellung gibt es eine oder zwei Enscheidungsgrenzen

eine Grenze = einseitiger Test

zwei Grenzen = zweiseitiger Test

ACHTUNG: Test auf Gleichheit ist zweiseitig
(zwei Grenzen die übereinander liegen)



Vergleich von Mittelwerten

Generell:

- Unterscheidung zwischen kleinen und großen Proben
- im Zweifel die Formeln für kleine Proben verwenden

Drei wichtige Fälle:

- Vergleich von Mittelwert mit bekannter vordefinierter Größe (Einstichproben-t-Test)
- Vergleich von zwei Mittelwerten (Zweistichproben-t-Test)
- Vergleich von zwei Mittelwerten bei gepaarten Experimenten (Differenzen-t-Test)

Fallunterscheidung ist wichtig, da die auftretenden Verteilungen unterschiedlich sind (Normalverteilung, t-Verteilung) und verschiedene Testgrößen verwendet werden müssen.

Einstichproben-t-Test

Vergleich von Mittelwerten mit bekannter Größe

	einseitiger Test		zweiseitiger Test
Hypothese	$H_0: \mu_1 \geq \mu_0$ $H_1: \mu_1 < \mu_0$	$H_0: \mu_1 \leq \mu_0$ $H_1: \mu_1 > \mu_0$	$H_0: \mu_1 = \mu_0$ $H_1: \mu_1 \neq \mu_0$
Testgröße (t-Verteilung)	$t = \dfrac{\bar{x} - \mu_0}{s / \sqrt{n}}$		
Freiheitsgrade	n - 1		
Rückweisung	H_0 ablehnen, falls $t < -t_\alpha$	H_0 ablehnen, falls $t > t_\alpha$	H_0 ablehnen, falls $\lvert t \rvert > t_{\alpha/2}$

Zweistichproben-t-Test

Vergleich von zwei Mittelwerten unkorrelierter Stichproben

	einseitiger Test		zweiseitiger Test
Hypothese	$H_0: \mu_1 \geq \mu_2$ $H_1: \mu_1 < \mu_2$	$H_0: \mu_1 \leq \mu_2$ $H_1: \mu_1 > \mu_2$	$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 \neq \mu_2$
Testgröße (t-Verteilung)	$t = (\bar{x}_1 - \bar{x}_2) \Big/ \sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$		$s_p^2 = \dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$
Freiheitsgrade	n_1+n_2-2		
Rückweisung	H_0 ablehnen, falls $t < -t_\alpha$	H_0 ablehnen, falls $t > t_\alpha$	H_0 ablehnen, falls $\lvert t \rvert > t_{\alpha/2}$

Differenzen-t-Test

Vergleich von zwei Mittelwerten aus gepaarten Stichproben

	einseitiger Test		zweiseitiger Test
Hypothese	$H_0: \mu_D \geq D_0$ $H_1: \mu_D < D_0$	$H_0: \mu_D \leq D_0$ $H_1: \mu_D > D_0$	$H_0: \mu_D = D_0$ $H_1: \mu_D \neq D_0$
Testgröße (t-Verteilung)	$t = \dfrac{\bar{d} - D_0}{s_D / \sqrt{n_D}}$		
Freiheitsgrade	$n_D - 1$		
Rückweisung	H_0 ablehnen, falls $t < -t_\alpha$	H_0 ablehnen, falls $t > t_\alpha$	H_0 ablehnen, falls $\lvert t \rvert > t_{\alpha/2}$

Beispiel - Luftfeuchtigkeit in Städten

Datensatz HUMIDIT2.TXT (siehe Teach/Me):
- relative Luftfeuchtigkeit in 264 Städten der USA
- Juni und September (am Morgen und am Nachmittag)
- jährliche Durchschnittswerte
- Datenquelle: **http://www.ncdc.noaa.gov/**

Überprüfe, ob es einen signifikanten Unterschied zwischen den Mittelwerten der Luftfeuchtigkeit gibt:

- Zwischen den Vormittags- und Nachmittagswerten (für Juni und September)
- Zwischen den Juni- und Septemberwerten (für die Vormittags- und Nachmittagsaufzeichnungen)

Führe die Tests bei unterschiedlichen Signifikanzniveaus durch

Vergleich von Varianzen

Zwei wichtige Fälle:

- Vergleich der Varianz mit bekannter vordefinierter Größe (Einstichproben-χ^2-Test)
- Vergleich von zwei Varianzen (Zweistichproben-F-Test)

Fallunterscheidung ist wichtig, da die auftretenden Verteilungen unterschiedlich sind (χ^2-Verteilung, F-Verteilung) und verschiedene Testgrößen verwendet werden müssen.

Einstichproben-χ^2-Test

Vergleich der Varianz mit einem vorgegebenen Wert

	einseitiger Test		zweiseitiger Test
Hypothese	$H_0: s \geq \sigma_0$ $H_1: s < \sigma_0$	$H_0: s \leq \sigma_0$ $H_1: s > \sigma_0$	$H_0: s = \sigma_0$ $H_1: s \neq \sigma_0$
Testgröße	$\chi^2 = \frac{s^2(n-1)}{\sigma_0^2}$		
Freiheitsgrade	n - 1		
Rückweisung	H_0 ablehnen, falls $\chi^2 < \chi^2_{1-\alpha}$	H_0 ablehnen, falls $\chi^2 > \chi^2_{\alpha}$	H_0 ablehnen, falls $\chi^2 < \chi^2_{(1-\alpha/2)}$ oder $\chi^2 > \chi^2_{\alpha/2}$

Zweistichproben-F-Test

Vergleich der Varianzen zweier Stichproben

	einseitiger Test		zweiseitiger Test
Hypothese	$H_0: \sigma_1^2 \geq \sigma_2^2$ $H_1: \sigma_1^2 < \sigma_2^2$	$H_0: \sigma_1^2 \leq \sigma_2^2$ $H_1: \sigma_1^2 > \sigma_2^2$	$H_0: \sigma_1^2 = \sigma_2^2$ $H_1: \sigma_1^2 \neq \sigma_2^2$
Testgröße (F-Verteilung)	$F = \frac{s_2^2}{s_1^2}$	$F = \frac{s_1^2}{s_2^2}$	$F = \frac{\text{größere Varianz}}{\text{kleinere Varianz}}$
Freiheitsgrade	$df_1 = n_1-1$ $\quad df_2 = n_2-1$		
Rückweisung	H_0 ablehnen, falls $F > F_\alpha$		H_0 ablehnen, falls $F > F_{\alpha/2}$

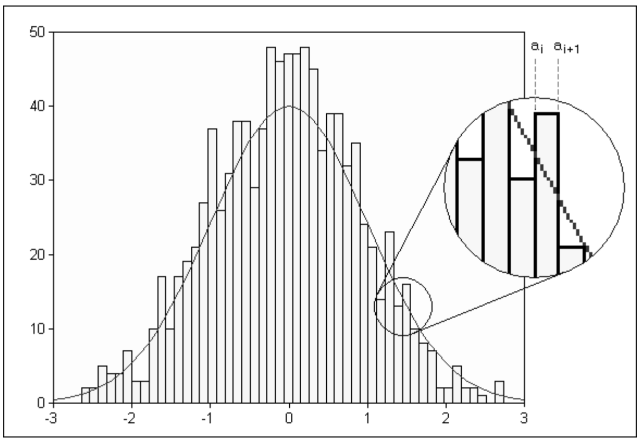
Vergleich von Verteilungen: χ^2-Test -1

Prinzip: die Summe der quadrierten Abweichungen von der vorgegebenen Wahrscheinlichkeitsdichte ist χ^2–verteilt.

Damit kann mit Hilfe der χ^2–Verteilung getestet werden, ob eine gemessene Häufigkeitsverteilung einer bestimmten Verteilung entspricht.

Nachteil:

viele Messpunkte notwendig



Vergleich von Verteilungen: χ^2-Test - 2

	einseitiger Test	zweiseitiger Test
Hypothese	———	H_0 : Verteilungen sind gleich H_1 : Verteilungen sind ungleich
Testgröße	$\chi^2 = \sum_{i=1}^{k} (F_i - E_i)^2 / E_i$	F_i ... empirische Häufigkeit der Klasse i E_i ... theoretische Häufigkeit der Klasse i k ... Zahl der Klassen
Freiheitsgrade	k-3	
Rückweisung	———	H_0 ablehnen, falls $\chi^2 > \chi^2_{\alpha,k-3}$

Alternativen zu χ^2-Test:

Kolmogorov-Smirnov-Test — einfach aber ähnlich geringe Trennschärfe wie χ^2-Test

Shapiro-Wilk-Test — bessere Trennschärfe, aber kompliziertere Anwendung

Ausreißer-Tests 1

Annahme

ein Messwert gehört zu einer bekannten Verteilung (meist Normalverteilung)

Grundprinzip

Damit lässt sich für jeden Messpunkt die Wahrscheinlichkeit angeben, dass dieser Punkt noch zur Verteilung gehört. Fällt diese Wahrscheinlichkeit unter einen bestimmten Wert, so spricht man von einem (wahrscheinlichen) Ausreißer.

Vorgangsweise

- Messwerte auf Ausreißer testen
- Ausreißer entfernen
- auf keinen Fall Ausreißertest ein zweites Mal anwenden!

Ausreißer-Tests 2

konkrete Ausreißer-Tests

z-Test

Voraussetzung: Messwerte normalverteilt
Berechnung der Standardabweichung
Werte außerhalb von +/- 2.5 Standardabweichungen sind Ausreißer
falls nicht normalverteilt: Theorem von Tschebyscheff (+/- 4s)

Interquartilsabstand

IQR = Abstand zwischen 1. und 3. Quartil
Werte unterhalb Q1-1.5*IQR oder oberhalb Q3+1.5*IQR sind Ausreißer

Dean-Dixon-Test

man sortiert alle Werte ($x_1...x_n$)
übersteigt der Quotient $|x_1 - x_2|/|x_1 - x_n|$ einen bestimmten Grenzwert, so ist x_1 ein Ausreißer