- Statistische Tests: Einführung und Übersicht
- Einstichprobenproblem
- Vergleich zweier abhängiger Gruppen
- Vergleich zweier unabhängiger Gruppen
- Vergleich von k unabhängigen Gruppen
- Vergleich k verbundener Stichproben
- Weitere Varianzanalyse-Modelle
- Anpassungstests
- Nichtparametrische Tests

5. Statistische Tests
5.1 Einführung und Übersicht

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

$X_1, X_2, ..., X_n$ 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 Schätzwerte (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 <u>zu</u> groß! $p < 0.01$
$\Rightarrow$ Käufer lehnt ab

b) Zufällig N_f <u>zu</u> klein! p groß, $p \gg 0.01$
$\Rightarrow$ Käufer kauft

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

Beide Risiken nicht gleichzeitig zu minimieren.

Lösung:

P(Nicht kaufen $| \; p \leq 1\%$) = α vorgeben
P(Kaufen $| \; p > 1\%$) 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) $H_0: \quad \mu = \mu_0 \qquad H_A: \mu \neq \mu_0$

Teststatistik

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

T heißt auch Testgröße, Prüfgröße, Stichprobenfunktion.

Hypothesentest
Allgemein

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. Ablehnungsbereich oder kritischen Bereich, 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.

Hypothesentest

Annahme- und Ablehnungsbereich

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}

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.

Hypothesentest

Fehler 1. Art, Fehler 2. Art

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

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

$\boldsymbol{\Theta}_0$: 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.

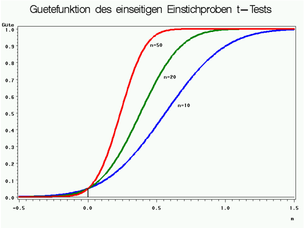
Beispiel: t-Test

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

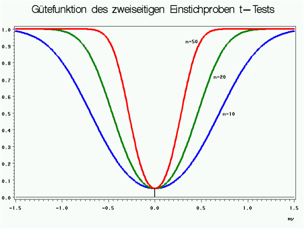
$nc = \sqrt{n}\frac{\mu-\mu_0}{\sigma}$: Nichtzentralitätsparameter
$t_{1-\alpha,n-1}$: kritischer Wert
$K = [t_{1-\alpha,n-1}, \infty)$: kritischer Bereich.

Gütefunktion

Einseitiger Test



Zweiseitiger Test



`Test_Guete_t.sas` `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

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

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

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

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

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 $= P_0(|T| > |t|)$

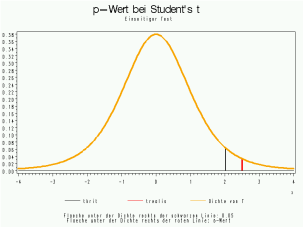
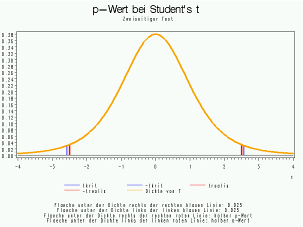
Einseitige Alternative: $\mu < \mu_0$
p-Wert $= P_0(T < t)$

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

Der p-Wert heißt auch Überschreitungswahrscheinlichkeit.

p-Wert

Illustration

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

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.

Der p-Wert ist nicht die Wahrscheinlichkeit, dass H_0 zurifft

$P(H_0|\text{Daten}) \neq$ p-Wert.

5.2 Einstichprobenproblem

	Nulhypothese	Alternative
a) H_0 :	$\mu \le \mu_0$	$H_A : \mu > \mu_0$
b) H_0 :	$\mu \ge \mu_0$	$H_A : \mu < \mu_0$
c) H_0 :	$\mu = \mu_0$	$H_A : \mu \ne \mu_0$

Teststatistik

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

'Student'

Durchführung des Tests mit
PROC UNIVARIATE MU0=μ_0; oder
PROC TTEST H0=μ_0

Einstichprobenproblem
Beispiel: Banknoten

`Test_t1_Banknote.sas`

μ_0	gr	p-Wert Pr$> \|t\|$		
215	1	0.4258	$> \alpha =$ 0.05	nosign
	2	< 0.0001	$< \alpha =$ 0.05	sign.
214.9	1	0.0784	$> \alpha =$ 0.05	nosign.
	2	0.03	$< \alpha =$ 0.05	sign.

Das sind also zweiseitge p-Werte (Alternative c)).
Was machen wir bei Alternative a) oder b)? $\rightarrow$ s.u.

vorgegeben: Fehler 1.Art α (Signifikanzniveau)
(üblich ist $\alpha = 0.05$ oder $\alpha = 0.01$)
d.h. $P_{\mu_0}(|T| > t_{krit}) = \alpha$.

Verteilung der Teststatistik T

Nehmen wir in unserem Beispiel an, die Beobachtungen

$$X_i \sim \mathcal{N}(\mu_0, \sigma^2), \quad , i = 1, \ldots, n$$

sind normal und unabhängig, dann hat die (zufällige) Teststatistik T eine t-Verteilung (Student's t),

$$T = \frac{\sqrt{n}(\overline{X} - \mu_0)}{s} \sim \frac{\mathcal{N}(0,1)}{\sqrt{\frac{1}{n-1}\chi^2_{n-1}}} =: t_{n-1}$$

(t-Verteilung mit $n-1$ Freiheitsgraden) und

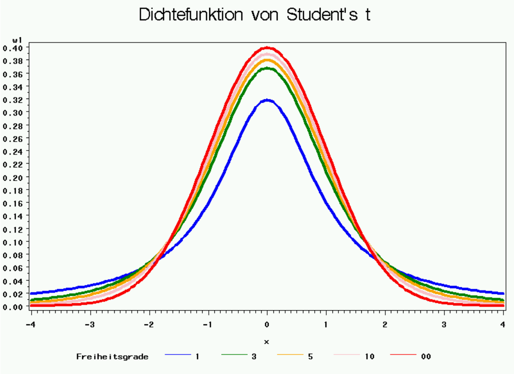
$$t_{krit} = t_{1-\frac{\alpha}{2}, n-1}$$

ist $(1-\frac{\alpha}{2})$ - Quantil einer t-Verteilung mit $n-1$ Freiheitsgraden.

Dichtefunktion einer t-Verteilung

mit $\nu (= n-1)$ Freiheitsgraden (FG)

$$f_{t_\nu}(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu \cdot \pi} \cdot \Gamma(\frac{\nu}{2})} \cdot (1 + \frac{x^2}{\nu})^{-\frac{\nu+1}{2}} \qquad PDF('T', x, \nu)$$



Test_t_Dichte.sas

Einstichprobenproblem

t-Test

a) $H_0 : \mu \leq \mu_0 \qquad H_A : \mu > \mu_0$
$\Rightarrow$ große Werte von

$$T = \frac{\overline{X} - \mu_0}{s} \cdot \sqrt{n}$$

indizieren Gültigkeit von H_A.

b) $H_0 : \mu \geq \mu_0 \qquad H_A : \mu < \mu_0$
$\Rightarrow$ kleine Werte von T indizieren H_A

c) $H_0 : \mu = \mu_0 \qquad H_A : \mu \neq \mu_0$
$\Rightarrow$ $|T|$ groß indiziert Gültigkeit von H_A.

Hypothesentest

Annahme- und Ablehnungsbereich

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}

Hypothesentest

Sei jetzt t eine Realisierung von T.

Zweiseitige Alternative $H_A : \mu \neq \mu_0$

Wenn $|t| > t_{krit} = t_{1-\frac{\alpha}{2},n-1}$ so H_0 abgelehnt.

Wenn $|t| \leq t_{krit} = t_{1-\frac{\alpha}{2},n-1}$ so H_0 nicht abgel.

Abl. Annahme Ablehnung

$-t_{krit}$ 0 t t_{krit}

Einseitige Alternative $H_A : \mu > \mu_0$

Wenn $t > t_{krit} = t_{1-\alpha,n-1}$ so H_0 abgelehnt.

Wenn $t \leq t_{krit} = t_{1-\alpha,n-1}$ so H_0 nicht abgel.

Annahme Ablehnung

0 t t_{krit}

Einseitige Alternative: $H_A : \mu < \mu_0$

Wenn $t < t_{krit} = t_{\alpha,n-1}$ so H_0 abgelehnt.

Wenn $t \geq t_{krit} = t_{\alpha,n-1}$ so H_0 nicht abgel.

Ablehnung Annahme

t_{krit} 0 t

p-Werte bei einseitigen Alternativen

Erinnerung: Der zweiseitige p-Wert ist: $P(|T| > t)$.

$$\begin{aligned}
P(|T| > t) &= P((T > t) \vee (-T > t)) \\
&= P((T > t) \vee (T < -t)) \\
&= 2 \cdot P(T > t), \qquad t > 0 \\
P(T > t) &= P(T < -t) \\
&= 1 - P(T \geq -t) \\
&= 1 - \frac{1}{2} P(|T| > -t), \qquad t \leq 0
\end{aligned}$$

(Die Verteilung von T ist stetig und symmetrisch.)

p-Werte bei einseitigen Alternativen

Fall a) $H_0 : \mu \leq \mu_0 \qquad H_a : \mu > \mu_0$

$$\text{p-Wert} = P(T > t) = \begin{cases} \frac{1}{2}P(|T| > t), & \text{falls } t > 0 \\ 1 - \frac{1}{2}P(|T| > -t), & \text{falls } t \leq 0 \end{cases}$$

Ablehnung von H_0 falls $P(T > t) < \alpha$.
Die p-Werte von SAS sind zweiseitig, sie sind also (wenn $t > 0$) durch 2 zu dividieren
(wenn $t \leq 0$ wird H_0 ohnehin nicht abgelehnt)

p-Werte bei einseitigen Alternativen

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

$$\text{p-Wert} = P(T < t) = \begin{cases} \frac{1}{2}P(|T| > |t|), & \text{falls } t \leq 0 \\ 1 - \frac{1}{2}P(|T| > -t), & \text{falls } t > 0 \end{cases}$$

Ablehnung von H_0 falls $P(T < t) < \alpha$
also wenn $t < 0$ so SAS-p-Wert durch 2 teilen!

Im Fall der zweiseitigen Alternative (c) ist der p-Wert $P(|T| > t)$ genau das was SAS ausgibt, wir brauchen also nichts zu ändern.

Zusammenfassung Einstichprobenproblem (1)

Teststatistik

$$T = \sqrt{n} \cdot \frac{\overline{X} - \mu_0}{S} \qquad \text{Realisierung } t$$

$\overline{X} = \frac{1}{n} \sum_i X_i, \quad S^2 = \frac{1}{n-1} \sum_i (X_i - \overline{X})^2$

Zweiseitige Alternative, $H_0 : \mu = \mu_0 \qquad H_A : \mu \neq \mu_0$

$|t| > t_{krit} \quad \Leftrightarrow \quad H_0$ ablehnen
p-value $< \alpha \quad \Leftrightarrow \quad H_0$ ablehnen
"Pr $> |t|$" $< \alpha \quad \Rightarrow \quad H_0$ ablehnen

Zusammenfassung Einstichprobenproblem (2)

Einseitige Alternative, $H_0 : \mu \leq \mu_0 \qquad H_A : \mu > \mu_0$

$t > 0$ und $\frac{\text{p-value}}{2} < \alpha \Leftrightarrow H_0$ ablehnen

Einseitige Alternative, $H_0 : \mu \geq \mu_0 \qquad H_a : \mu < \mu_0$

$t < 0$ und $\frac{\text{p-value}}{2} < \alpha \Leftrightarrow H_0$ ablehnen

Konfidenzbereiche

am Beispiel des t-Tests

$X \sim N(\mu, \sigma^2) \quad \Rightarrow \quad \sqrt{n} \cdot \frac{\overline{X}-\mu}{S} \sim t_{n-1}$ wenn μ der wahre (Lokations-) Parameter ist. $\Rightarrow$

$$P(\underbrace{-t_{1-\frac{\alpha}{2},n-1} \leq \sqrt{n} \cdot \frac{\overline{X}-\mu}{s} \leq t_{1-\frac{\alpha}{2},n-1}}_{(*)}) = 1-\alpha$$

Die Ungleichungenen sind äuqivalent zu

$$\begin{aligned}
(*) \quad &\leftrightarrow \quad -\frac{s}{\sqrt{n}} t_{1-\frac{\alpha}{2},n-1} \leq \overline{X}-\mu \leq \frac{s}{\sqrt{n}} t_{1-\frac{\alpha}{2},n-1} \\
&\leftrightarrow \quad -\overline{X} - \frac{s}{\sqrt{n}} t_{1-\frac{\alpha}{2},n-1} \leq -\mu \leq -\overline{X} + \frac{s}{\sqrt{n}} t_{1-\frac{\alpha}{2},n-1} \\
&\leftrightarrow \quad \overline{X} + \frac{s}{\sqrt{n}} t_{1-\frac{\alpha}{2},n-1} \geq \mu \geq \overline{X} - \frac{s}{\sqrt{n}} t_{1-\frac{\alpha}{2},n-1} \\
&\leftrightarrow \quad \overline{X} - \frac{s}{\sqrt{n}} t_{1-\frac{\alpha}{2},n-1} \leq \mu \leq \overline{X} + \frac{s}{\sqrt{n}} t_{1-\frac{\alpha}{2},n-1}
\end{aligned}$$

Konfidenzbereiche

$(1-\alpha)$ Konfidenzintervall für den (unbekannten) Parameter μ

$$\left[\overline{X} - \frac{s}{\sqrt{n}} \cdot t_{1-\frac{\alpha}{2},n-1}, \overline{X} + \frac{s}{\sqrt{n}} \cdot t_{1-\frac{\alpha}{2},n-1}\right]$$

PROC TTEST ALPHA=Wert
PROC UNIVARIATE ALPHA=Wert CIBASIC
ALPHA: Konfidenzniveau (=Signifikanzniveau)

CIBASIC: Konfidenzintervalle für μ, σ^2, σ basierend auf Normalverteilung

CIPCTLDF: verteilungsfreie Konfidenzintervalle (basierend auf empirischen Quantilen)

Konfidenzbereiche

Beispiel

`Test_t1_Banknote`

Konfidenzintervalle für den Lageparameter
$\mu = \mathbf{E}$‘laenge’:

	echt		gefälscht	
$\alpha = 0.01$	214.87	215.07	214.73	214.92
$\alpha = 0.05$	214.89	215.05	214.75	214.89
$\alpha = 0.05$ nichtparam. KI (für Median)	214.9	215.1	214.7	214.9

PROC TTEST ALPHA=Wert

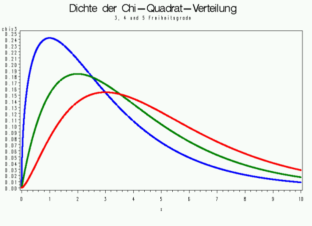
PROC UNIVARIATE ALPHA=Wert CIBASIC CIPCTLDF

Konfidenzintervalle für σ^2

$$X \sim \mathcal{N}(\mu, \sigma^2) \quad \Rightarrow \quad (n-1)\frac{S^2}{\sigma^2} \sim \chi^2_{n-1}$$

Dichte einer χ^2_ν-Verteilung

$$f_{\chi^2_\nu}(x) = \begin{cases} \frac{1}{2^{\nu/2}\Gamma(\frac{\nu}{2})} e^{-x/2} x^{\nu/2-1} & \text{falls} \quad x \geq 0 \\ 0 & \text{sonst.} \end{cases}$$



`Test_Chi2_Dichte`

Konfidenzintervall für σ^2

$$P\Big(\chi^2_{\alpha/2,n-1} \le (n-1)\frac{S^2}{\sigma^2} \le \chi^2_{1-\alpha/2,n-1}\Big) = 1-\alpha$$

auflösen nach σ^2:

$$\begin{aligned}
1-\alpha &= P\Big(\chi^2_{\alpha/2,n-1} \le (n-1)\frac{S^2}{\sigma^2} \le \chi^2_{1-\alpha/2,n-1}\Big) \\
&= P\Big(\frac{1}{\chi^2_{1-\alpha/2,n-1}} \le \frac{\sigma^2}{(n-1)S^2} \le \frac{1}{\chi^2_{\alpha/2,n-1}}\Big) \\
&= P\Big(\frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}}\Big)
\end{aligned}$$

Konfidenzintervall für σ^2

Konfidenzintervall

(Vertrauensintervall) für den (unbekannten) Parameter σ^2

$$\left[\frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}}, \frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}}\right]$$

PROC TTEST

PROC UNIVARIATE ALPHA CIBASIC CIPCTLDF

5.3 Vergleich zweier abhängiger Gruppen
(verbundene Stichproben)

$$
\begin{array}{ll}
H_0 : \mu_1 = \mu_2 & H_1 : \mu_1 \neq \mu_2 \\
H_0 : \mu_1 \geq \mu_2 & H_1 : \mu_1 < \mu_2 \\
H_0 : \mu_1 \leq \mu_2 & H_1 : \mu_1 > \mu_2
\end{array}
$$

- Gewicht einer Person zu den Zeitpunkten t_1, t_2.
- Banknoten (oben- unten, links - rechts)
- Patient nimmt Medikament 1 und 2
- Kreuz- und selbstbefruchtete Pflanzen

`Test_t2_Banknote` `Test_t2_Darwin`

Folgende Möglichkeiten:

a) Transformation $Z := X_1 - X_2$ und testen auf $\mu = 0$
PROC UNIVARIATE; VAR Z; RUN; oder
PROC TTEST H0=0; VAR Z; RUN;

b) Mit der Prozedur TTEST:

```
PROC TTEST;
   PAIRED X1*X2;
RUN;
```

5.4 Vergleich zweier unabhängiger Gruppen
(unverbundene Stichproben)

$$
\begin{array}{ll}
H_0 : \mu_1 = \mu_2 & H_1 : \mu_1 \neq \mu_2 \\
H_0 : \mu_1 < \mu_2 & H_1 : \mu_1 \geq \mu_2 \\
H_0 : \mu_1 > \mu_2 & H_1 : \mu_1 \leq \mu_2
\end{array}
$$

- Tibetische Schädel (Sikkim - Kham)
- Wasserhärte (Nord - Süd)
- Klinikaufenthalt (Klinik1 - Klinik2)
- Banknoten (echt - gefälscht)

```
Test_t2_Tibetan
Test_t2_Heroin
Test_t2_Banknote
```

Vergleich zweier unabhängiger Gruppen

$$X_1 \sim N(\mu_1, \sigma_1^2), \quad X_2 \sim N(\mu_2, \sigma_2^2)$$

Fall 1: Varianzen σ_1^2, σ_2^2 sind gleich
Fall 2: Varianzen σ_1^2, σ_2^2 sind verschieden

Fall 1:

$$T = \frac{\overline{X}_1 - \overline{X}_2}{}$$

Vergleich zweier unabhängiger Gruppen

$$X_1 \sim N(\mu_1, \sigma_1^2), \quad X_2 \sim N(\mu_2, \sigma_2^2)$$

Fall 1: Varianzen σ_1^2, σ_2^2 sind gleich
Fall 2: Varianzen σ_1^2, σ_2^2 sind verschieden

Fall 1:

$$T = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{1}{n} + \frac{1}{m}}\sqrt{\frac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2}}}$$

n, m: Umfänge Stichprobe 1 und 2

$$S_1^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_{1i} - \overline{X}_1)^2 \quad S_2^2 = \frac{1}{m-1}\sum_{i=1}^{m}(X_{2i} - \overline{X}_2)^2$$

Erläuterung des Quotienten T

$X_1 \sim \mathcal{N}(\mu_1, \sigma^2), X_2 \sim \mathcal{N}(\mu_2, \sigma^2)$

$$\overline{X}_1 \sim N(\mu_1, \sigma^2 \cdot \frac{1}{n}), \quad \overline{X}_2 \sim N(\mu_2, \sigma^2 \cdot \frac{1}{m})$$

$$\frac{(n-1)}{\sigma^2} \cdot S_1^2 \sim \chi^2_{n-1}, \quad \frac{(n-1)}{\sigma^2} \cdot S_2^2 \sim \chi^2_{m-1}$$

$$\overline{X}_1 - \overline{X}_2 \sim N(\mu_1 - \mu_2, \sigma^2 \cdot (\frac{1}{n} + \frac{1}{m}))$$

$$\frac{1}{\sigma^2}\left((n-1) \cdot S_1^2 + (m-1) \cdot S_2^2\right) \sim \chi^2_{n+m-2}$$

$$T \sim t_{n+m-2}$$

unter H_0 ($\mu_1 = \mu_2$).

Vergleich zweier unabhängiger Gruppen

T ist eine Zufallsgröße!

Werte von T werden mit gewissen Wktn. angenommen!

Die Wkt. dafür, daß T sehr große Werte annimmt (wenn H_0 richtig ist) ist also sehr klein.

Sei jetzt t eine Realisierung von T (also der Wert, der bei Ausrechnen anhand der gegebenen Daten entsteht).

Wenn jetzt t sehr groß, $|t| \in K$ (krit. Bereich)
(aber die Wkt. dafür ist sehr klein, wenn H_0 richtig ist)
$\Rightarrow H_0$ ablehnen.

Vergleich zweier unabhängiger Gruppen
ungleiche Varianzen

Fall 2: Varianzen ungleich

$$T = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{m}}}$$

$T \sim t_\nu$ approximativ. Die Zahl ν der Freiheitsgrade wird auch approximativ berechnet. (Welch-Test, 1937)

SAS bietet Tests für beide Varianten an. Satterthwaite-Approximation (1946).

```
PROC TTEST;
    CLASS Klassifikationsvariable;
    VAR auszuwertende Variable(n);  RUN;
```

Vergleich zweier unabhängiger Gruppen

Welchen Test soll man nehmen?

- Aus Vorinformation ist vielleicht bekannt, ob man gleiche Varianzen annehmen kann.

- Man kann einen Test auf gleiche Varianzen vorschalten

Problem: 2 stufiger Test

Wird das Signifikanzniveau eingehalten??

Test auf Gleichheit der Varianzen
Voraussetzung: Normalverteilung

$H_0 : \sigma_1^2 = \sigma_2^2$ $\qquad$ $H_1 : \sigma_1^2 \neq \sigma_2^2$

$$F = \frac{S_1^2}{S_2^2} \sim F_{n-1,m-1}$$

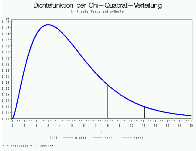
(Fisher-) F- Verteilung mit $(n-1, m-1)$ Freiheitsgraden.

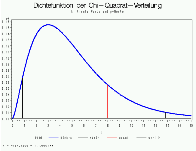
F ist Quotient zweier unabhängiger χ^2-verteilter Zufallsgrößen.
H_0 ablehnen, falls

$$\frac{s_1^2}{s_2^2} < F_{\frac{\alpha}{2},n-1,m-1} \quad \text{oder} \quad \frac{s_1^2}{s_2^2} > F_{1-\frac{\alpha}{2},n-1,m-1}$$

Test auf Gleichheit der Varianzen

F-Test





$$F_{\frac{\alpha}{2},n-1,m-1} = \frac{1}{F_{1-\frac{\alpha}{2},m-1,n-1}}$$

(beachten: Freiheitsgrade vertauschen sich)

$\Rightarrow H_0$ ablehnen, falls

$$\frac{s_1^2}{s_2^2} < \frac{1}{F_{1-\frac{\alpha}{2},m-1,n-1}} \quad \text{oder} \quad \frac{s_1^2}{s_2^2} > F_{1-\frac{\alpha}{2},n-1,m-1}$$

$\Leftrightarrow$

$$\frac{s_2^2}{s_1^2} > F_{1-\frac{\alpha}{2},m-1,n-1} \quad \text{oder} \quad \frac{s_1^2}{s_2^2} > F_{1-\frac{\alpha}{2},n-1,m-1}$$

Test auf Gleichheit der Varianzen

F-Test, prakt. Durchführung

$s_M^2 := max(s_1^2, s_2^2)$ $\qquad$ $s_m^2 := min(s_1^2, s_2^2)$

n_M, n_m: die entsprechenden Stichprobenumfänge

$\Rightarrow H_0$ ablehnen, falls

$$\frac{s_M^2}{s_m^2} > F_{1-\frac{\alpha}{2}, n_M-1, n_m-1}.$$

Formulierung mit p-Werten

H_0 ablehnen, falls

$$P(W_{n_M-1,n_m-1} > \frac{s_M^2}{s_m^2}) < \frac{\alpha}{2}$$

$W_{n_M-1,n_m-1} \sim F_{n_M-1,n_m-1}$

Zweistichprobenproblem

Output der Prozedur TTEST

- Konfidenzintervalle für μ_1, μ_2 und für $\mu_1 - \mu_2$
  Für die ersten beiden siehe Abschnitt 5.2
  Konfidenzintervalle für $\mu = \mu_1 - \mu_2$ bekommt man analog zum Einstichprobenfall, hier wird die Zweistichproben-t-Teststatistik genommen (die für gleiche Varianzen) und die entsprechenden Ungleichungen umgeformt.
- Tabelle der durchgeführten t-Tests
  - für gleiche Varianzen (pooled)
  - für ungleiche Varianzen (Satterthwaite)
- F-Test zum Vergleich der Varianzen (bitte ignorieren)

Ein- und Zweistichprobenproblem

Anmerkungen (1)

- Der F-Test (zum Skalenvergleich) ist sehr empfindlich gegenüber Abweichungen von der Normalverteilungsannahme

$\Rightarrow$ mit größter Vorsicht genießen.

- Der Einstichproben- t-Test ist nicht robust!
- Der Zweistichproben t-Test ist etwas robuster als der t-Test im Einstichprobenproblem
- Ausreißer können extremen Einfluß haben (ÜA).
- Wenn Gleichheit der Varianzen unklar

$\Rightarrow$ t-Test mit ungleichen Varianzen nehmen.
(ist bei gleichen Varianzen nur ganz wenig weniger effizient)

Ein- und Zweistichprobenproblem

Anmerkungen (2)

- Besser nicht auf das Ergebnis des F-Tests verlassen.
  (Problematik: 2-Stufentest, Nicht-Robustheit).
- Es gibt robustere Skalentests $\Rightarrow$
Levene Test und Brown-Forsythe Test.

Test auf Gleichheit der Varianzen

Levene-Test

Bilden die Werte

$$
\begin{aligned}
X_j^* &:= |X_j - \overline{X}| \\
Y_j^* &:= |Y_j - \overline{Y}|
\end{aligned}
$$

Skalenunterschiede in (X, Y) spiegeln sich jetzt in Lageunterschieden in (X^*, Y^*) wieder.
Mit den "neuen Beobachtungen" wird jetzt ein t-Test durchgeführt.
Die t-Verteilung gilt nur approximativ.

Test auf Gleichheit der Varianzen

Brown-Forsythe Test

Analog zum Levene-Test, nur hier bilden wir die Werte

$$\begin{aligned} X_j^* &:= |X_j - \text{med}_i X_i| \\ Y_j^* &:= |Y_j - \text{med}_i Y_i| \end{aligned}$$

Beide Tests sind (einigermaßen) robust gegen Abweichungen von der Normalverteilung.

Test auf Gleichheit der Varianzen

Syntax

```
PROC ANOVA;
  CLASS Klasse;
  MODEL var=Klasse;
  MEANS Klasse /
       HOVTEST=Levene (TYPE=ABS);
  MEANS Klasse / HOVTEST=BF;
RUN;
```

`Test_t2_Banknote`

Vergleich von k unabhängigen Gruppen
(einfaktorielle, einfache Varianzanalyse)

A: Faktor (Gruppenvariable) mit k Stufen (Faktorstufen)

Modell

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad i = 1...k, j = 1...n_i$$

μ:	Gesamterwartungswert
α_i:	Effekt der i-ten Stufe von A
ϵ_{ij}:	Fehler, $\epsilon_{ij} \sim (0, \sigma^2)$
Y_{ij}:	j-te Beobachtung der i-ten Faktorstufe
$\sum_{i=1}^k \alpha_i = 0$	Parametrisierungsbedingung

Einfache Varianzanalyse

$H_0 : \alpha_1 = \alpha_2 = ... = \alpha_k$
$H_1 : \alpha_i \neq \alpha_l$ (für ein $i \neq l$)

Im Fall $k = 2$ führt dieses Testproblem auf das Zweistichprobenproblem ($\rightarrow$ t-Test).

Output der Maschinen gleich?
Klausurergebnisse unterschiedlich?
Mageninhalt der Eidechsen gleich?
Cortisolgehalt unterschiedlich?

```
ANOVA_Maschinen
Varianzanalyse_Modelle\PI12erg
GLM_Eidechsen
GLM_Cortisol
```

Varianzanalyse

Varianzanalyse macht eine Streuungszerlegung:

Gesamt-varianz	=	Varianz zwischen den Faktorstufen	+	Varianz innerhalb der Faktorstufen
SST (Total)	=	SSB (Between)	+	SSW (Within) | (SSE) (Error)

$$N = \sum_{i=1}^{k} n_i$$

$$\overline{Y}_i = \frac{1}{n_i} \cdot \sum_{j=1}^{n_i} Y_{ij}, \qquad \overline{Y} = \frac{1}{N} \sum_{i,j} Y_{i,j}$$

Einfache Varianzanalyse

Satz: Es gilt

$$SSB + SSW = SST$$

wobei

$$\begin{aligned}
SSB &= \sum_{i=1}^{k} n_i(\overline{Y}_i - \overline{Y})^2 && \text{(}\underline{\text{B}}\text{etween)}\\
SSW &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_i)^2 && \text{(}\underline{\text{W}}\text{ithin)}\\
SST &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y})^2. && \text{(}\underline{\text{T}}\text{otal)}
\end{aligned}$$

Einfache Varianzanalyse

Satz: $SSB + SSW = SST$

Beweis:

$$SSB = \sum_i n_i \overline{Y}_i^2 - 2 \cdot N \cdot \overline{Y}^2 + \overline{Y}^2 \cdot N$$

$$SSW = \sum_{i,j} Y_{ij}^2 - 2 \cdot \sum_i n_i \overline{Y}_i^2 + \sum_i n_i \overline{Y}_i^2$$

$SSB + SSW =$

$$\begin{aligned} &= \sum_{i,j} Y_{ij}^2 + \sum_i n_i \overline{Y}_i^2 - N \cdot \overline{Y}^2 - \sum_i n_i \overline{Y}_i^2 \\ &= \sum_{i,j} Y_{ij}^2 - N \cdot \overline{Y}^2 = \sum_j \sum_i (Y_{ij} - \overline{Y})^2 = SST \quad \blacksquare \end{aligned}$$

Varianzanalyse

```
PROC ANOVA;
   CLASS A; /*A: Faktor */
    MODEL var=A;
    MEANS A /
          HOVTEST=Levene (TYPE=ABS);   oder
          HOVTEST=BF;
   MEANS OUT=SAS-Ausgabedatei;
RUN;
```

oder: PROC GLM anstelle von PROC ANOVA;
und weiter wie oben.
ANOVA: schneller

GLM: zusätzliche Auswertungen möglich, z.B. Ausgabe der Residuen

HOVTEST: Test auf Varianzhomogenität

Einfache Varianzanalyse

Dependent Variable: Y

Source	DF	Sum of Squares	Mean Squares	F-value	$Pr > F$
MODEL	k-1	SSB(M)	MSB	$\frac{MSB}{MSE}$	p-Wert
ERROR	N-k	SSW(E)	MSE		
Total	N-1	SST			

$$MSB = \frac{SSB}{k-1}, \qquad MSE = \frac{SSW}{N-k}$$

$$H_0: \quad \alpha_1 = \cdots = \alpha_k$$
$$H_1: \quad \exists (i,j): \alpha_i \neq \alpha_j$$

Einfache Varianzanalyse

H_0 wird getestet mit

$$F = \frac{MSB}{MSE} = \frac{\text{Mittlere Var. zwischen d. Gruppen}}{\text{Mittlere Var. innerhalb d. Gruppen}}$$

$$= \frac{N-k}{k-1} \qquad \frac{SSB}{SSW} = \frac{N-k}{k-1} \quad \frac{SST - SSW}{SSW}$$

F groß, $F > F_{1-\alpha,k-1,N-k} \Leftrightarrow H_0$ abgelehnt

Bestimmtheitsmaß

$$R^2 := \frac{SSB}{SST} = \frac{SST - SSW}{SST} = 1 - \frac{SSW}{SST}$$

Der Anteil der Varianz, der durch das Modell bestimmt wird, heißt Bestimmtheitsmaß

Einfache Varianzanalyse

Offenbar: $0 \leq R^2 \leq 1$.

$$F = \frac{MSB}{MSE} = \frac{N-k}{k-1} \cdot \frac{SSB}{SST} \cdot \frac{SST}{SSW} = \frac{N-k}{k-1} \cdot \frac{R^2}{1-R^2}$$

$$\begin{aligned} R^2 \to 0 &\quad\Longrightarrow\quad F \to 0 \\ R^2 \to 1 &\quad\Longrightarrow\quad F \to \infty. \end{aligned}$$

Schätzung der Modellstandardabweichung σ

$$RootMSE = \sqrt{MSE} = \sqrt{\tfrac{1}{N-k} SSE}$$

Variationskoeffizient

$$CV = \frac{100 \cdot RootMSE}{\overline{Y}}$$

Einfache Varianzanalyse
Anmerkungen

- • Der F-Test in der Varianzanalyse ist (einigermaßen) robust gegenüber Abweichungen von der Normalverteilungsannahme
- • Wenn man die Prozedur GLM verwendet, dann kann man die sogen. Residuen

$$\hat{\epsilon}_{ij} = Y_{ij} - \hat{\alpha}_i - \hat{\mu}$$

abspeichern (Option RESIDUAL im OUTPUT-Statement) und auf Normalität testen.
(PROC UNIVARIATE NORMAL)

Varianzanalyse

- • F-Test verlangt auch Varianzhomogenität
  Daten balanziert (gleiche Stichprobenumfänge)
  $\longrightarrow$ Abweichungen nicht so schwerwiegend.
- • Wenn die Varianzen verschieden sind, kann die Welch-Modifikation verwendet werden:
  MEANS Var/WELCH;

Einfache Varianzanalyse

Test auf Varianzhomogenität

$$H_0 : \sigma_1^2 = \sigma_2^2 = ... = \sigma_k^2$$
$$H_1 : \exists(i, l) : \sigma_i^2 \neq \sigma_l^2$$

Levene Test (1960)

HOVTEST= LEVENE im MEANS-Statement

$Z_{ij}^* = |Y_{ij} - \overline{Y}_{i.}|$

Brown-Forsythe-Test (1974)

HOVTEST = BF

$Z_{ij}^* = |Y_{ij} - medY_i|$

Einfache Varianzanalyse
Test auf Varianzhomogenität (2)

Mit diesen neuen ZV wird eine Varianzanalyse durchgeführt.

$$W = \frac{\frac{1}{k-1}\sum n_i(\overline{Z}^*_{i.} - \overline{Z}^*)^2}{\frac{1}{N-k}\sum_{i,j}(Z^*_{ij} - \overline{Z}^*_{i.})^2} \sim F_{k-1,N-k}.$$

```
GLM_Cortisol
```

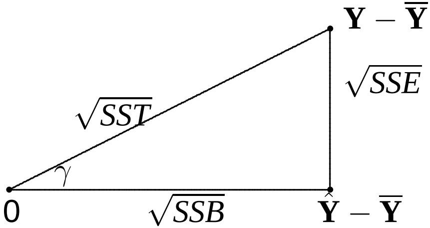
Geometrische Veranschaulichung

zur Varaianzanalyse

$$\mathbf{Y} = (Y_{11}, \ldots, Y_{kn_k}) \quad \text{Dimension } N$$

$$\hat{\mathbf{Y}} = (\underbrace{\overline{Y}_1, \ldots, \overline{Y}_1}_{n_1\text{mal}}, \ldots, \underbrace{\overline{Y}_k, \ldots, \overline{Y}_k}_{n_2\text{mal}})$$

$$\overline{\mathbf{Y}} = (\underbrace{\overline{Y}, \ldots, \overline{Y}}), \quad \overline{Y} = \frac{1}{N}\sum_{i,j} Y_{ij}$$



$$\begin{aligned} SSB + SSW &= SST \qquad\qquad R^2 = cos^2\gamma \\ ||\hat{\mathbf{Y}} - \overline{Y}||^2 + ||\mathbf{Y} - \hat{\mathbf{Y}}||^2 &= ||\mathbf{Y} - \overline{Y}||^2 \end{aligned}$$

Multiple Vergleiche

Problemstellung: H_0 abgelehnt, aber zwischen welchen Faktorstufen liegt der Unterschied?

- Idee: Alle Paarvergleiche machen.
- Problem: Wenn wir das Signifikanzniveau $\alpha (= 0.05)$ so lassen, wird das Testniveau nicht eingehalten!
- Veranschaulichung: Bei 20 gleichzeitigen Tests können wir $20 \cdot \alpha = 1$ Ablehnung erwarten, auch wenn H_0 richtig ist.

Multiple Vergleiche
Lösungsmöglichkeiten

Option BON im MEANS Statement

Signifikanzniveau für die gleichzeitigen Tests herabsetzen auf $\frac{\alpha_{nom}}{\binom{k}{2}}$,
bei $k = 4$ wäre das etwa $\frac{\alpha_{nom}}{\binom{4}{2}} = \frac{0.05}{6}$.

Begründung: Bonferroni-Ungleichung.

Option TUKEY im MEANS Statement

Bilden die $\overline{Y}_j$ und die Spannweite dazu

$$w = max_{i,j}|\overline{Y}_i - \overline{Y}_j|$$

Dazu kommt noch die empirische Standardabweichung s.

Multiple Vergleiche
Lösungsmöglichkeiten

Option TUKEY im MEANS Statement

$$t_{\text{max}} = \frac{w}{s}$$

die sogenannnte *studentisierte Spannweite*.
Diese hat (wenn die $Y_i \sim \mathcal{N}$) eine (dem SAS-Programmierer) wohlbekannte Verteilung, und entsprechende Quantile und kritische Werte.
Damit erhalten wir simultane Konfidenzintervalle für alle Paardifferenzen $\mu_i - \mu_j$. Liegt 0 nicht darin, so wird $H_{0,ij}: \ \mu_i = \mu_j$ abgelehnt zugunsten von $H_{A,ij}: \ \mu_i \neq \mu_j$.

Bem.: Es gibt eine Fülle weiterer Varianten.

5.6 Vergleich k verbundener Stichproben
2-faktorielle Varianzanalyse

Modell:

$$Y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}, \quad \epsilon_{ij} \sim (0, \sigma^2)$$

$i = 1, \ldots, a, j = 1, \ldots, b$. (eine Beobachtung je Zelle)

Das Modell ist überparametrisiert,
Bedingung: $\sum_{i=1}^{a} \alpha_i = 0, \quad \sum_{j=1}^{b} \beta_j = 0$.
Folgene Hypothesen sind zu testen:

$H_{0a}: \quad \alpha_1 = \cdots = \alpha_a = 0$ gegen
$H_{1a}: \quad \exists(i_1, i_2) : \alpha_{i_1} \neq \alpha_{i_2}$
$H_{0b}: \quad \beta_1 = \cdots = \beta_b = 0$ gegen
$H_{1a}: \quad \exists(j_1, j_2) : \beta_{j_1} \neq \beta_{j_2}$

`GLM_Synchro` `GLM_Cache`

2-faktorielle Varianzanalyse

$$\overline{Y}_{..} = \frac{1}{a \cdot b} \sum_{i=1}^{a} \sum_{j=1}^{b} Y_{ij} \quad \text{arith. Mittel aller Beob.}$$

$$\overline{Y}_{i.} = \frac{1}{b} \sum_{j=1}^{b} Y_{ij} \quad \text{Mittel der i-ten Stufe von A}$$

$$\overline{Y}_{.j} = \frac{1}{a} \sum_{i=1}^{a} Y_{ij} \quad \text{Mittel der j-ten Stufe von B}$$

$$SSA := b \sum_{i=1}^{a} (\overline{Y}_{i.} - \overline{Y}_{..})^2 \qquad SSB := a \sum_{j=1}^{b} (\overline{Y}_{.j} - \overline{Y}_{..})^2$$

$$SSE := \sum_{i=1}^{a} \sum_{j=1}^{b} (Y_{ij} - \overline{Y}_{i.} - \overline{Y}_{.j} + \overline{Y}_{..})^2$$

$$SST := \sum_{i=1}^{a} \sum_{j=1}^{b} (Y_{ij} - \overline{Y}_{..})^2$$

2-faktorielle Varianzanalyse

Quadratsummenzerlegung

Dependent Variable: Y

Source	DF	Sum Squ.	Mean Squ.	F-value	$Pr > F$ p-value
A	a-1	SSA	MSA	$\frac{MSA}{MSE}$	H_{1a}
B	b-1	SSB	MSB	$\frac{MSB}{MSE}$	H_{1b}
Model	a+b-2	SSM	MSM	$\frac{MSM}{MSE}$	H_1
Error	(a-1)(b-1)	SSE	MSE		
Total	a b - 1	SST			

$$SSM = SSA + SSB \qquad SST = SSA + SSB + SSE$$

$$MSA = \frac{SSA}{(a-1)} \qquad MSB = \frac{SSB}{(b-1)}$$

$$MSM = \frac{SSA + SSB}{a+b-2} \qquad MSE = \frac{SSE}{(a-1)(b-1)}$$

2-faktorielle Varianzanalyse

Tests

H_{0a} gegen H_{1a}:

$$F_1 = \frac{MSA}{MSE} = \frac{\text{mittl. Var. zwischen Stufen von A}}{\text{mittl. Var. innerhalb d. Gruppen}}$$

$$F_1 \quad \sim \quad F_{a-1,(a-1)(b-1)}$$

H_{0b} gegen H_{1b}:

$$F_2 = \frac{MSB}{MSE} = \frac{\text{mittl. Var. zwischen Stufen von B}}{\text{mittl. Var. innerhalb d. Gruppen}}$$

$$F_2 \quad \sim \quad F_{b-1,(a-1)(b-1)}$$

große Werte von F führen zur Ablehnung!

$$F_1 \quad > \quad F_{1-\alpha,a-1,(a-1)(b-1)} \rightarrow \text{Ablehnung von } H_{0a}$$

$$F_2 \quad > \quad F_{1-\alpha,b-1,(a-1)(b-1)} \rightarrow \text{Ablehnung von } H_{0b}$$

2-faktorielle Varianzanalyse
Tests

H_0: $\alpha_1 = \cdots \alpha_a = 0$ und $\beta_1 = \cdots \beta_a = 0$ gegen
H_1: $\exists(i_1, i_2)$: $\alpha_{i_1} \neq \alpha_{i_2} \vee \exists(j_1, j_2)$: $\beta_{j_1} \neq \beta_{j_2}$.

$$F = \frac{MSModell}{MSE} = \frac{SSA + SSB}{SSE} \cdot \frac{(a-1)(b-1)}{a+b-2}$$

$$\begin{aligned} MSModell &= \frac{SSModell}{a+b-2} \\ SSModell &= SSA + SSB. \end{aligned}$$

H_0 ablehnen, falls

$$F > F_{1-\alpha, a+b-2, (a-1)(b-1)}.$$

Zweifaktorielle Varianzanalyse

```
PROC GLM;
   CLASS A B; /*die beiden Faktoren*/
   MODEL Y = A B;
RUN;
```

Output

- Balanzierter Fall: Variante I und III identisch
- Unbalanzierter Fall: Typ III-Summen sind vorzuziehen, da der entsprechende Test unabhängig von den Stichprobenumfängen ist.

5.7 Weitere Varianzanalyse-Modelle

5.7.1 Mehrere Beobachtungen pro Kombination der Faktoren A und B

SAS-Prozedur ändert sich nicht!

Output ändert sich gegebenenfalls
a) balanzierter Fall $\rightarrow$ eindeutig
b) unbalanzierter Fall $\rightarrow$
Es gibt verschiedene Möglichkeiten die Fehlerquadratsummen zu zerlegen.
$\rightarrow$ SAS bietet die Varianten an

3 Forscher graben eine Reihe von Schädeln in 3 verschiedenen Schichten aus. Gemessen wird die Nasenlänge.
? Forschereffekt, Schichteneffekt

Weitere Varianzanalyse-Modelle

Mehrere Beobachtungen pro Kombination der Faktoren A und B

Klinische Untersuchung in mehreren Zentren

Ein Medikament zur Gewichtsreduktion soll getestet werden.

1: Medikament
0: Placebo
1-6: Zentren

Modell:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk}, \quad \epsilon_{ijk} \sim \mathcal{N}(0, \sigma^2)$$

Es interessiert nur das Medikament, nicht das Zentrum:

$$H_0: \quad \alpha_0 = \alpha_1 \qquad H_1: \alpha_0 < \alpha_1$$

Weitere Varianzanalyse-Modelle

```
PROC GLM;
   CLASS Medik Zentrum; /*die beiden Faktoren*/
   MODEL Y = Medik Zentrum;
RUN; (dieselbe Prozedur wie oben)
```

GLM_Drugeffect

Zum Output: wie bisher.

Balanzierter Fall: Variante I und III identisch.

Unbalanzierter Fall: Typ III-Summen zu bevorzugen, da der entsprechende Test unabhängig von den Stichprobenumfängen ist.

Weitere Varianzanalyse-Modelle

5.7.2 Wechselwirkungen ins Modell mit aufnehmen

$$Y_{ijk} = \alpha + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}$$

(+Reparametrisierungsbedingungen)
geht nur, wenn für jede Faktorstufenkombination mehrere Beobachtungen vorliegen.

```
PROC GLM;
   CLASS A B; /*die beiden Faktoren*/
   MODEL Y = A B A*B;
RUN;
```

`GLM_Insekten`

Weitere Varianzanalyse-Modelle

Modell mit Wechselwirkungen

Folgene Hypothesen sind zu testen:

$H_{0a}: \quad \alpha_1 = \cdots = \alpha_a = 0$ gegen
$H_{1a}: \quad \exists (i_1, i_2) : \alpha_{i_1} \neq \alpha_{i_2}$

$H_{0b}: \quad \beta_1 = \cdots = \beta_b = 0$ gegen
$H_{1a}: \quad \exists (j_1, j_2) : \beta_{j_1} \neq \beta_{j_2}$

$H_{0c}: \quad \gamma_{11} = \cdots = \gamma_{a*b} = 0$ gegen
$H_{1c}: \quad \exists (j_1, j_2) : \gamma_{j_1, j_2} \neq 0$

Weitere Varianzanalyse-Modelle

5.7.3 Faktoren (Effekte, Faktorstufen) sind zufällig

hier ist Schätzung der Varianzkomponenten interessant und evtl. ein Hypothesentest

Preisrichter seien zufällig ausgewählt.

Die Frage ist, ob die Variabilität in den Scores an den Preisrichtern liegt?

$$Y_{ij} = \mu + \underbrace{A_i}_{\text{zufällig}} + b_j + \epsilon_{ij}$$

$$A_i \sim (0, \sigma_P^2)$$

$$\epsilon_{ij} \sim (0, \sigma^2)$$

Varianzkomponentenschätzung

```
PROC VARCOMP METHOD=Type1;
   CLASS Preisrichter Wettkaempfer;
   MODEL Score = Preisricher;
RUN;
```

`GLM_syncro_zufaelligeEffekte`

METHOD=Type1: Auf den Quadratsummen beruhende Varianzschätzungen

Annahme: A_i, B_j und ϵ_{ij} unabhängig.

$$var(Y_{ij}) = var(A_i) + var(B_j) + var(\epsilon_{ij})$$

Output: Schätzungen für die Varianzkomponenten.

Weitere Varianzanalyse-Modelle

7.4 Mehr als 2 Faktoren

- höherfaktorielle VA

Frequenzspektren

Gemessen wird die Amplitude bei 35 verschiedenen Frequenzen, 4 Füllungen, 3 Richtungen,
jede Messung wird 5 mal wiederholt.
? Füllungs-, Richtungseffekt, Wiederholungseffekt?
Frequenzeffekt?
$\longrightarrow$ 4 Faktoren.

```
PROC GLM;
    CLASS A B C D;
    MODEL Y = A B C D; RUN;
```

`~/Beratung/Vogt/Glaeser1`

Weitere Varianzanalyse-Modelle

7.5 Hierarchische Modelle

Die Faktoren liegen in hierarch. Ordnung vor.

						A					
	A1			A2			A3			A4	
B11	B12	B13	B21	B22	B23	B31	B32	B33	B41	B42	B43

(mit zufäll. Effekten)

Kalzium-Gehalt verschiedener Pflanzen und von verschiedenen Blättern
4 Pflanzen werden zufällig ausgewählt
3 Blätter davon
2 Stichproben zu 100mg von jedem Blatt
Frage: Gibt es zwischen Pflanzen oder zwischen Blättern unterschiedliche CA-Konzentrationen?

Weitere Varianzanalyse-Modelle
Hierarchische Modelle (2)

Modell

$$Y_{ijk} = \mu + A_i + B_{ij} + \epsilon_{ijk}$$

$A_i \sim \mathcal{N}(0, \sigma_a^2) \quad B_{ij} \sim \mathcal{N}(0, \sigma_b^2) \quad \epsilon_{ijk} \sim \mathcal{N}(0, \sigma^2)$

hier: n=2 a=4 b=3

$$\begin{aligned} \text{var}Y_{ijk} &= \text{var}A_i + \text{var}B_{ij} + \text{var}\epsilon_{ijk} \\ &= \sigma_a^2 + \sigma_b^2 + \sigma^2 \end{aligned}$$

$H_{0a} : \sigma_a^2 = 0 \qquad H_{0b} : \sigma_b^2 = 0$

`GLM_hierarch`

Weitere Varianzanalyse-Modelle

Hierarchische Modelle (3)

```
PROC GLM;
   CLASS A B;
   MODEL Y = A B(A);  hierarch. Struktur*
   RANDOM A B(A);  Faktoren sind zufaellig*
RUN;

PROC VARCOMP;
    CLASS A B;
    MODEL Y=A B(A);
RUN;
```

5.8 Anpassungstests

Anpassungstests
Einführung

Problem

Klassische Test- und Schätzverfahren sind oft konzipiert unter der Normalverteilungsannahme.

Frage

Gilt sie überhaupt?

Gilt die Normalverteilung? (1)

Hampel, 1980, Biometr. Journal

Eine Zeitlang glaubte (fast) jeder an das ’normale Fehlergesetz’,

die Mathematiker, weil sie es für ein empirisches Faktum hielten,

und die Anwender, weil sie es für ein mathematisches Gesetz hielten.

Gilt die Normalverteilung? (2)

Geary 1947, Biometrika

Normality is a myth;

there never was,

and never will be,
a normal distribution.

Anpassungstests

$(X_1, \ldots, X_n)$ iid., $X_i \sim F$, F unbekannt.

Anpassungstest auf eine spezifizierte Verteilung:

$$H_0: \quad F = F_0 \qquad \text{gegen} \qquad H_1: \quad F \neq F_0.$$

I.A. hängt F von unbekannten Parametern ab.

Anpassungstest auf eine Normalverteilung:

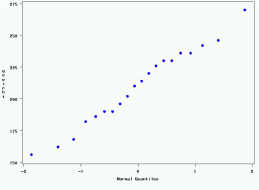
$$\begin{aligned} H_0: F(x) &= \Phi\Big(\frac{x-\mu}{\sigma}\Big) \quad (\mu, \sigma \ \text{ unbekannt}) \\ H_1: F(x) &\neq \Phi\Big(\frac{x-\mu}{\sigma}\Big) \quad \forall \mu, \sigma, \sigma > 0 \end{aligned}$$

(Φ: Verteilungsfunktion der Standardnormal.).

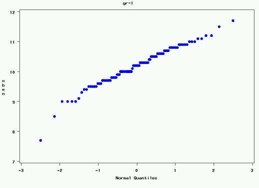
Anpassungstests

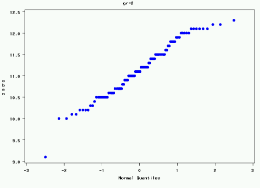
Gewicht von Hühnern

156	162	168	182	186
190	190	196	202	210
214	220	226	230	230
236	236	242	246	270



Abmessungen von Banknoten, oben (echt, falsch)





Auf der empirischen Verteilungsfunktion beruhende Tests

Empirische Verteilungsfunktion

Seien $X_1, ..., X_n$ unabh. Beobachtungen,
$X_{(1)} \leq ... \leq X_{(n)}$ die geordneten Beob.
Die Funktion

$$F_n(x) = \begin{cases} 0 & x < X_{(1)} \\ \frac{i}{n} & X_{(i)} \leq x < X_{(i+1)} \qquad i = 1...n \\ 1 & X_{(n)} \leq x \end{cases}$$

heißt empirische Verteilungsfunktion.

Satz v. Glivento-Cantelli: $F_n(x) \to F(x)$.
(Hauptsatz der math. Statistik genannt)

Anpassungstests

Auf der empirischen Verteilungsfunktion beruhende Tests

Kolmogorov-Smirnov-Test

$$D = \sup_x |F_n(x) - F_0(x)|$$

Anderson-Darling-Test

$$\text{A-sq} = n \int_{-\infty}^{\infty} \frac{(F_n(x) - F_0(x))^2}{F_0(x)(1 - F_0(x))} dF_0(x)$$

Cramer-von Mises-Test

$$\text{W-sq} = n \int_{-\infty}^{\infty} \big(F_n(x) - F_0(x)\big)^2 dF_0(x)$$

Anpassungstests auf Normalverteilung

Auf der empirischen Verteilungsfunktion beruhende Tests

hier:

$$\begin{aligned} F_0(x) &= \Phi\left(\frac{x-\hat{\mu}}{\hat{\sigma}}\right), \\ \hat{\mu} = \overline{X} &= \frac{1}{n}\sum_{i=1}^{n} X_i, \\ \hat{\sigma}^2 = s^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \end{aligned}$$

$D \sim D_n$ (Kolmogorov-Verteilung) approx.

$$\lim_{n\to\infty} P_0\left(D < \frac{x}{\sqrt{n}}\right) = 1 - 2\sum_{k=1}^{\infty}(-1)^k e^{-2k^2x^2}$$

(Kolmogorov, 1933).

Anpassungstests auf Normalverteilung

Auf der empirischen Verteilungsfunktion beruhende Tests

Modifikationen für endliche Stichproben (zur Info.)

D: $D \cdot (\sqrt{n} - 0.01 + 0.85/\sqrt{n})$

$\mathbf{A} - \mathbf{sq}$: A-sq $\cdot(1.0 + 0.75/n + 2.25/n^2)$

W-sq: W-sq $\cdot(1.0 + 0.5/n)$

Große Werte von D, A-sq und W-sq führen jeweils zur Ablehnung von H_0.
p-Werte werden vom Programm berechnet.

```
Test_GoF_Banknote.sas
Test_GoFDarwin.sas
aufg24.sas
```

Anpassungstests

Shapiro-Wilk-Test

Vorbemerkungen:

$$X_i \sim \mathcal{N}(\mu, \sigma^2), \qquad Y_i = \frac{X_i - \mu}{\sigma} \sim \mathcal{N}(0,1)$$

$i = 1, \ldots, n$.

Geordnete Beobachtungen:

$$X_{(1)} \leq \ldots \leq X_{(n)} \qquad Y_{(1)} \leq \ldots \leq Y_{(n)}.$$

Die Erwartungswerte

$$\begin{aligned} m_i &:= E(Y_{(i)}) \\ &= \frac{n!}{(i-1)!(n-i)!} \cdot \int_{-\infty}^{\infty} t\Phi^{i-1}(t)(1-\Phi(t))^{n-i}\phi(t)dt \end{aligned}$$

sind bekannt (und vertafelt).

Shapiro-Wilk-Test

Approximation (Blom, 1958)

$$m_i \approx \widetilde{m}_i = \Phi^{-1}\Big(\frac{i - 0.375}{n + 0.25}\Big)$$

$$\begin{aligned} \mathbf{E}X_{(i)} &= \mu + \sigma m_i \\ X_{(i)} &= \mu + \sigma m_i + \epsilon_i \end{aligned}$$

einfaches lineares Regressionsmodell mit Parametern μ, σ.
$\mathbf{E}\epsilon_i = 0$, aber die ϵ_i sind nicht unabhängig.

$$\mathbf{V} := \text{cov}\big(Y_{(i)}, Y_{(j)}\big), \qquad \mathbf{m}' := (m_1, \ldots, m_n)$$

$$\mathbf{X}' := \big(X_{(1)}, \ldots, X_{(n)}\big).$$

Verallgemeinerter Kleinster Quadrat-Schätzer von σ:

$$\hat{\sigma} = \frac{\mathbf{m}'\mathbf{V}^{-1}\mathbf{X}}{\mathbf{m}'\mathbf{V}^{-1}\mathbf{m}}$$

wird verglichen mit der gewöhnlichen Standardabweichung s,

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

Anpassungstests

Shapiro-Wilk-Statistik

$$W = \frac{\hat{\sigma}^2}{s^2(n-1)} \cdot \frac{(\mathbf{m}'\mathbf{V}^{-1}\mathbf{m})^2}{\mathbf{m}'\mathbf{V}^{-2}\mathbf{m}} = \frac{(\mathbf{h}'\mathbf{X})^2}{\sum_{i=1}^n (X_i - \overline{X})^2 \cdot \mathbf{h}'\mathbf{h}}$$

wobei $\mathbf{h}' = \mathbf{m}'\mathbf{V}^{-1}$ (bekannt, vertafelt).

Wegen $\sum h_i = 0$ folgt:
W ist Quadrat des (empir.) Korrelationskoeffizienten von $\mathbf{h}$ und $\mathbf{X}$:

$$W = \frac{\left(\sum_{i=1}^n (X_i - \overline{X})(h_i - \overline{h})\right)^2}{\sum_{i=1}^n (X_i - \overline{X})^2 \cdot \sum_{i=1}^n (h_i - \overline{h})^2},$$

$$W = \frac{\left(\sum_{i=1}^{n}(X_i - \overline{X})(h_i - \overline{h})\right)^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2 \cdot \sum_{i=1}^{n}(h_i - \overline{h})^2},$$

- Offenbar: $0 \leq W \leq 1$.
- $W \approx 1$ indiziert, dass $\mathbf{h}' = \mathbf{m}'\mathbf{V}^{-1}(\approx 2\mathbf{m}')$: ein Vielfaches von $\mathbf{X}$ ist. D.h. die Punkte $(m_i, X_{(i)})$ liegen etwa auf einer Geraden, was Normalverteilung indiziert.
- H_0 wird ablehnt, falls $W < W_\alpha(n)$.

`Test_GoF_Shapiro_Wilk.sas`

Shapiro-Wilk Test

Scores des 1 Wettkämpfers (5 Preisrichter)

31.2, 31.2, 31.4, 32.0, 33.1
Mit der Prozedur UNIVARIATE erhalten wir $s = 0.80747$ und mit der Prozedur GPLOT (Option REGEQN) $\hat{\sigma} = 0.805$ im Regressionsmodell $Y_i = \mu + \sigma m_i + \epsilon_i$
Für die Shapiro-Wilk Statistik bekommen wir

$$W = \frac{\hat{\sigma}^2}{s^2} \cdot c = 0.966.$$

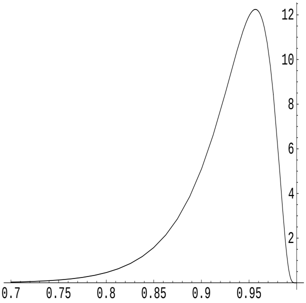
(c: Normierungsfaktor)

Shapiro-Wilk Test

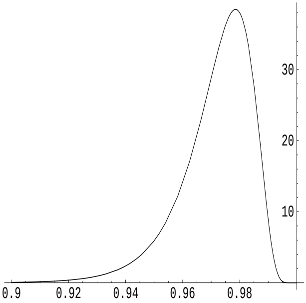
Approximative Dichtefunktion von W (unter H_0)

$$W = \frac{\left(\sum_{i=1}^{n}(X_i - \overline{X})(h_i - \overline{h})\right)^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2 \cdot \sum_{i=1}^{n}(h_i - \overline{h})^2},$$

$n = 10$



$n = 50$



Anpassungstests

- SAS verwendet eine Approximation von W.
- Der Shapiro-Wilk-Test erweist sich für kleinere, mittlere und größere Stichprobenumfänge als geeignetster Test (höchste Güte).
- Früher wurde meist der sogen. χ^2-Anpassungstest verwendet. Dieser hat jedoch geringe Güte.
- W ist etwas besser als A-sq, besser als W-sq, und viel besser als D und χ^2.
- D ist nur für sehr große Stichprobenumfänge zu empfehlen ($n \geq 2000$).

Anpassungstests

- Man sollte beim Test auf Normalverteilung das Signifikanzniveau auf $\alpha = 0.1$ hochsetzen, insbesondere wenn wenig robuste Tests (die NV verlangen) angewendet werden sollen.

Robuste Tests haben meist geringen Effizienzverlust bei NV.

Anpassungstests

Durchführung des Tests auf Normalverteilung

Unter Verwendung von $\hat{\mu}, \hat{\sigma}$:

```
PROC UNIVARIATE NORMAL;
RUN;

PROC UNIVARIATE;
HISTOGRAM variable / NORMAL;
RUN;
```

mit vorgebenen μ, σ:

```
PROC UNIVARIATE;
HISTOGRAM variable / NORMAL(mu=0, sigma=1);
RUN;
```

Bem.: Mit der Prozedur UNIVARIATE (Kommando HISTOGRAM) können Sie auch auf andere Verteilungen testen.

Anpassungstests

8.4 Anpassungstests auf weitere Verteilungen

χ^2-Anpassungstest (Pearson, 1900)

Prinzip: Daten werden in p Klassen eingeteilt.
Klassenhäufigkeiten: N_i
theoretische Klassenhäufigkeiten: np_i

$$X^2 = \sum_{i=1}^{p} \frac{(N_i - np_i)^2}{np_i}$$

$X^2 \sim \chi^2_{p-1}$ asymptotisch (bei bekannten μ, σ^2)
(Fisher, 1922)

$X^2 \sim \chi^2_{p-3}$ approx. (bei 2 zu schätzenden Parametern, ML-Schätzung mit gruppierten Daten oder Minimum-χ^2-Schätzung).

Anpassungstests
χ^2-Anpassungstest

Nachteile des χ^2-Anpassungstests

- Wert von X^2 abhängig von Klasseneinteilung.
- χ^2- Anpassungstest auf Normalverteilung hat geringe Güte.

Diskrete Verteilungen

Hier kann der χ^2-Anpassungstest genommen werden
(natürliche Klasseneinteilung)
Prozedur FREQ, Option CHISQ

Anpassungstests

χ^2-Anpassungstest

Diskrete Gleichverteilung

```
PROC FREQ;
    TABLES var1 /CHISQ;
RUN;
```

Sonstige diskrete Verteilungen

wie oben, zusätzlich sind die Einzelwktn. explizit zu formulieren, /CHISQ TESTP=($p_1, ..., p_k$);

`Test_GoF_Poisson.sas`

Anzahlen schon gegeben

Die Variablen, die Anzahlen bezeichnen, werden durch ein WEIGHT-Kommando angegeben.

Anpassungstests
EDF-Tests

Stetige Verteilungen

zugelassen sind:

Normal, Gamma, Weibull, Lognormal, Exponential

```
HISTOGRAM var1 / Gamma;
```

`Test_GoF_Darwin_1.sas`

5.9. Nichtparametrische Tests

Übersicht

Es werden die wichtigsten Rang-Analoga zu den Tests in 5.2.-5.6. behandelt.

5.9.0 Einführung

5.9.1 Einstichprobenproblem (vgl 5.2), 2 verbundene Stichproben (vgl. 5.3)
Vorzeichentest, Vorzeichen-Wilcoxon-Test

5.9.2 Zwei unverbundene Stichproben (vgl. 5.4)
Wilcoxon-Test

5.9.3 Mehrere unabhängige Stichproben (vgl. 5.5)
Kruskal-Wallis-Test

5.9.4 Mehrere verbundene Stichproben (vgl. 5.6)
Friedman-Test

Nichtparametrische Tests

5.9.0 Einführung

Was tun wenn Normalverteilung nicht vorliegt?

Nichtparametrische Tests

- sie verwenden <u>keine</u> Parameterschätzung (wie $\overline{X}, s$)
- sie halten das Signifikanzniveau (α) für jede stetige Verteilung (approx.) ein. α hängt also nicht von der zugrundeliegenden Verteilungsfunktion ab.
- sie sind relativ effizient. Der Effizienzverlust bei Normalvert. ist in vielen Fällen gering!

Annahme: Verteilungsfunktion ist stetig (wenn nicht anders vermerkt)

Nichtparametrische Tests

5.9.1 Einstichprobenproblem

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

Vorzeichentest

Wie bisher werden die Differenzen $X_i - \mu_0$ gebildet.

$$V_i := \begin{cases} 1 & \text{falls } X_i - \mu_0 > 0 \\ 0 & \text{falls } X_i - \mu_0 < 0 \end{cases}$$

$$V^+ = \sum_{i=1}^{n} V_i$$

= # Differenzen mit positivem Vorzeichen

Nichtparametrische Tests
Vorzeichentest (2)

Bem: Der Fall $X_i - \mu_0 = 0$ tritt wegen der Stetigkeit der Vf. nur mit Wkt. 0 auf.
Sollte der Wert $X_i - \mu_0 = 0$ trotzdem vorkommen (Meßungenauigkeit) so wird die entspr. Beobachtung weggelassen und der Stichprobenumfang entsprechend verringert.
(Nachteil: Es werden gerade Beob. weggelassen, die für die Nullhypothese sprechen!)

Es gilt: $V^+ \sim B(n, \frac{1}{2})$

(V^+ = # “Erfolge” bei n Versuchen mit Wkt. je $\frac{1}{2}$).

$\Rightarrow$ krit. Werte können leicht bestimmt werden:
BINV($1 - \alpha, n, \frac{1}{2}$) oder
QUANTILE(’Binomial’,$1 - \alpha, n, \frac{1}{2}$)

Nichtparametrische Tests
Vorzeichentest (3)

Teststatistik

$$\underline{M} = V^+ - \frac{n}{2} \qquad (= \frac{V^+ - V^-}{2}) \quad (zentrierteStatistik)$$

n^+: Realisierung von V^+
n^-: Realisierung von V^-

Zweiseitiger p-Wert:
$P(|\underline{M}| \geq |n^+ - \frac{n}{2}|) = P(|\underline{M}| \geq \max(n^+, n^-) - \frac{n}{2}$)=(*)

$$\text{denn } |n^+ - \frac{n}{2}| = \begin{cases} n^+ - \frac{n}{2} & n^+ > \frac{n}{2} \\ \frac{n}{2} - n^+ & n^+ < \frac{n}{2} \\ = n^- - \frac{n}{2} & \end{cases}$$

Nichtparametrische Tests

Vorzeichentest (4)

Der p-Wert ist gleich

$$(*) = P\Big(V^+ - \frac{n}{2} \geq \max(n^+, n^-) - \frac{n}{2}\Big) + P\Big(\frac{n}{2} - V^+ \geq \max(n^+, n^-) - \frac{n}{2}\Big)$$

$$= P\Big(V^+ \geq \max(n^+, n^-)\Big) + P\Big(n - V^+ \geq \max(n^+, n^-)\Big)$$

$$= 2 \sum_{j=\max(n^+,n^-)}^{n} \binom{n}{j} (\frac{1}{2})^j (\frac{1}{2})^{n-j}$$

$$= (\frac{1}{2})^{n-1} \sum_{j=\max(n^+,n^-)}^{n} \binom{n}{j}$$

$$= (\frac{1}{2})^{n-1} \sum_{j=0}^{\min(n^+,n^-)} \binom{n}{j}.$$

Nichtparametrische Tests
Vorzeichentest (5)

Die Verteilung von V^+ ist diskret, d.h. es gibt nicht zu jedem α einen entsprechenden kritischen Wert.

Aber: p-Werte gibt es immer, d.h.:

$p < \alpha \quad \Rightarrow \quad H_0$ (c) ablehnen

$M > 0 \wedge \frac{p}{2} < \alpha \quad \Rightarrow \quad H_0$ (b) ablehnen

$M < 0 \wedge \frac{p}{2} < \alpha \quad \Rightarrow \quad H_0$ (a) ablehnen

Der Vorzeichentest ist meist nicht sehr effizient
(Ausnahme: Verteilung=Doppelexponential)
besser ist der Wilcoxon-Vorzeichen-Rangtest

Nichtparametrische Tests

5.9.1.2 Wilcoxon-Vorzeichen-Rangtest

Wilcoxon-Vorzeichen-Rangtest

Bilden zu den "Beobachtungen"

$$D_i = |X_i - \mu_0|$$

die Rangzahlen, d.h. den Rang (den Platz) in der geordneten Stichprobe

$$\underbrace{D_{(1)}}_{\text{Rang 1}} \leq \ldots \qquad \ldots \leq \underbrace{D_{(n)}}_{\text{Rang n}}$$

Sei R_i^+ der Rang von D_i.

$$W_n^+ = \sum_{i=1}^{n} R_i^+ \cdot V_i$$

Summe der Ränge von D_i für die $X_i - \mu_0 > 0$.

Nichtparametrische Tests
Wilcoxon-Vorzeichen-Rangtest (2)

Erwartungswert und Varianz von W_n^+

$$\mathbf{E}_0 W_n^+ = \frac{1}{2}\sum_{i=1}^{n} R_i^+ = \frac{1}{2}\sum_{i=1}^{n} i = \frac{n\cdot(n+1)}{4} \qquad \mathbf{E}V_i = \frac{1}{2}$$

$$\text{var } W_n^+ = \mathbf{E}(W_n^+ - \mathbf{E}W_n^+)^2 = \frac{n\cdot(n+1)(2n+1)}{24} \quad \text{(ÜA)}$$

Die Berechnung der exakten Verteilung von W_n^+ kann durch Auszählen aller Permutationen erfolgen
($\rightarrow$ schon für kleinere n größere Rechenzeit!)
Deshalb verwendet man (für mittlere und große n) die asymptotische Verteilung.

Nichtparametrische Tests

Wilcoxon-Vorzeichen-Rangtest (3)

Asymtotische Verteilung

$$W_n^+ \sim \mathcal{N}(EW_n^+, \mathrm{var} W_n^+) \qquad \text{asymptotisch}$$

Große Werte von

$$\frac{|W_n^+ - EW_n^+|}{\sqrt{\mathrm{var}\ W_n^+}}$$

führen zur Ablehnung von H_0.

Nichtparametrische Tests
Wilcoxon-Vorzeichen-Rangtest (4)

SAS-Implementation (Wilcoxon-Vorzeichen-Test)

$$S = W_n^+ - EW_n^+ = \sum_{i=1}^{n} R_i^+ V_i - \frac{n(n+1)}{4}$$

R_i^+ Rang von $|X_i - \mu_0|$, Summe nur über positive $X_i - \mu_0$

$n \leq 20$: p-Werte aus der exakten Verteilung von S.

$n > 20$: Es wird eine t-Approximation angeboten:

$$t = \frac{S \cdot \sqrt{n-1}}{\sqrt{n \,\mathrm{Var}\,(S) - S^2}} \sim t_{n-1}$$

Bindungen (= Meßwertwiederholungen): Ränge werden gemittelt.

Sei t_i: $\#$ Bindungen in der i-ten Gruppe.
Korrektur in Var(S):

$$\text{var}(S) = \frac{n(n+1)(2n+1)}{24} - \frac{1}{2}\sum t_i(t_i+1)(t_i-1)$$

Nichtparametrische Tests
Wilcoxon-Vorzeichen-Rangtest (4)

IQ-Werte von Studenten (Wiwi)

99, 131, 118, 112, 128, 136, 120, 107, 134, 122

$H_0 : \mu = \mu_0 = 110 \qquad H_1 : \mu > \mu_0$

x_i	d_i	$\|d_i\|$	r_i^+	V_i
99	-11	11	5	0
131	21	21	8	1
118	8	8	3	1
112	2	2	1	1
128	18	18	7	1
136	26	26	10	1
120	10	10	4	1
107	-3	3	2	0
134	24	24	9	1
122	12	12	6	1

$d_i = x_i - 110$

Vorzeichentest:

$M = 8 - \frac{10}{2}$

p-Wert(exakt) = 0.1094

Wilcoxon-signed

$W^+ - \mathbf{E}(W^+) =$

$48 - \frac{10 \cdot 11}{4} = 20.5$.

p-Wert=0.0371.

`Test_IQ_Daten`

Nichtparametrische Tests

Wilcoxon-Vorzeichen-Rangtest (5)

- Im Gegensatz zum Vorzeichentest ist der Vorzeichen-Wilcoxon-Test (= signed rank test) sehr effizient, bei NV nur wenig schlechter, bei den meisten Vf. besser als der t-Test.
  $\Rightarrow$ Wenn NV nicht gesichert ist Vorzeichen-Wilcoxon-Test nehmen!
- Der Vorzeichentest und der Wilcoxon-Test sind sogen. Rangtests, da sie nur auf den Rangzahlen der Beobachtungen beruhen.
  Es gibt weitere Rangtests.
- Durchführung der Tests:
  PROC UNIVARIATE MU0=Wert;

Nichtparametrische Tests

Zwei verbundene Stichproben

Bilden $Z := X - Y$ und testen wie beim Einstichprobenproblem, z.B.

$$H_0 : \mu_Z = 0 \qquad\qquad H_1 : \mu_Z \neq 0$$

Banknoten: oben-unten, links-rechts
Darwin: kreuz-selbstbefruchtete Pflanzen (zur Illustration mit Prozedur RANK)

```
PROC UNIVARIATE;
  VAR Z;
RUN;
```

```
Npar_1_Banknote
Npar_1_Darwin
```

Nichtparametrische Tests
Weitere Problemstellungen im Einstichprobenfall

Binärvariablen

Sei X eine 0-1 Variable, d.h.

$$P(X=0)=p, \quad P(X=1)=1-p$$

$H_0: p = p_0$
T: Anzahl der Beobachtungen in Klasse 0.

$H_{1a} \quad p < p_0:$
p-Wert $= P(T \leq t)$ = CDF('Binomial',$t, n, p_o)$

$H_{1b} \quad p > p_0:$
p-Wert $= P(T \geq t)$

$H_{1c} \quad p \neq p_0:$
p-Wert $= P(T \leq t$ oder $T \geq n - t + 1)$

Nichtparametrische Tests
Weitere Problemstellungen im Einstichprobenfall

Binomialtest

Prozedur FREQ, Option Binomial

$$
\begin{aligned}
\hat{p} &= T/n \\
se(\hat{p}) &= \sqrt{\hat{p}(1-\hat{p})/n} = ASE \\
Z &= \frac{\hat{p}-p_0}{se(\hat{p})}
\end{aligned}
$$

Einseitige p-Werte bei SAS sind

$$
\begin{cases}
P(Z > z) & \text{falls} \quad z > 0 \\
P(Z < z) & \text{falls} \quad z \leq 0
\end{cases}
$$

Nichtparametrische Tests
Binomialtest

```
PROC FREQ;
   TABLES var / BINOMIAL(0.8);
RUN;
```

`Binomialtest_toxaemia.sas`

Konfidenzintervalle:

a) Normalapproximation: $\hat{p} \pm u_{\alpha/2} se(\hat{p})$
b) exakt: Binomialverteilung (CDF('Binomial',....))

Nichtparametrische Tests
Weitere Problemstellungen im Einstichprobenfall

Zum Vergleich, zur Erinnerung und Ergänzung

χ^2-Anpassungstest

Anpassungstest auf diskrete Gleichverteilung:

```
PROC FREQ;
   TABLES var /CHISQ;
RUN;
```

Anpassungstest auf vorgegebene diskrete Verteilung

```
PROC FREQ;
   TABLES var /CHISQ TESTP=(p1, ..., pk);
RUN;
```

Nichtparametrische Konfidenzintervalle

Option CIPCTLDF in der PROC UNIVARIATE

$(1-\alpha)$-Konfidenzintervall für p-Quantil, d.h. für x_p

Die Verteilung der j-ten Ordnungsstatistik $X_{(j)}$:

$$P(X_{(j)} < x) = \sum_{i=j+1}^{n} \binom{n}{i} F(x)^i (1-F(x))^{n-i}$$

'Erfolg" gdw. $X_i < x$, "Erfolgswkt." $F(x)$.
Insbesondere, für $x = x_p$ (das wahre p-Quantil)

$$\begin{aligned} P(X_{(j)} < x_p) &= \sum_{i=0}^{j} \binom{n}{i} F(x_p)^i (1-F(x_p))^{n-i} \\ &= \sum_{i=0}^{j} \binom{n}{i} p^i (1-p)^{n-i} \end{aligned}$$

Nichtparametrische Konfidenzintervalle

Option CIPCTLDF in der PROC UNIVARIATE (2)

$$P(X_{(j)} < x_p) = \sum_{i=j+1}^{n} \binom{n}{i} p^i (1-p)^{n-i}$$

Untere und obere Konfidengrenzen $X_{(l)}$ und $X_{(u)}$ für x_p werden so bestimmt, dass l und u (möglichst) symmetrisch um $\lfloor np \rfloor + 1$ und so dass

$$P(X_{(l)} \leq x_p < X_{(u)}) = \sum_{i=l}^{u-1} \binom{n}{i} p^i (1-p)^{n-i} \geq 1 - \alpha$$

($X_{(\lfloor np \rfloor)}$ ist Schätzung für x_p.)
PROC UNIVARIATE CIPCTLDF;

Nichtparametrische Tests

5.9.2 Zwei unverbundene Stichproben-Wilcoxon Test

Wir setzen keine Normalverteilung voraus, aber den gleichen Verteilungstyp, insbesondere gleiche Varianzen

$$H_0: \mu_1 = \mu_2 \quad H_1: \mu_1 \neq \mu_2$$
$$H_0: \mu_1 \geq \mu_2 \quad H_1: \mu_1 < \mu_2$$
$$H_0: \mu_1 \leq \mu_2 \quad H_1: \mu_1 > \mu_2$$

Wir fassen die Beobachtungen
$X_{11}, ..., X_{1n}, X_{21}, ..., X_{2m}$
zu einer Stichprobe zusammen und bilden die Rangzahlen R_{ij}, $i = 1, 2$, $j = 1 \ldots n, m$

$$\underbrace{z_{(1)}}_{\text{Rang 1}} \leq \ldots \leq \underbrace{z_{(n+m)}}_{\text{Rang n+m}}$$

Nichtparametrische Tests
Wilcoxon-Test

Summe der Ränge zur 1. bzw. 2. Stichprobe

$$S_1 = \sum_{j=1}^{n} R_{1j} \qquad S_2 = \sum_{j=1}^{m} R_{2j}$$

Die Erwartungswerte (unter H_0) sind

$$\mathbf{E}_0 S_1 = \frac{n(n+m+1)}{2} \quad \text{und} \quad \mathbf{E}_0 S_2 = \frac{m(n+m+1)}{2}$$

und die Varianzen

$$\text{var} S_1 = \text{var}\, S_2 = \frac{n \cdot m(n+m+1)}{12}.$$

Nichtparametrische Tests
Wilcoxon-Test (2)

Die Teststatistik des Wilcoxon-Tests ist

$$Z = \frac{S - E(S)}{\sqrt{varS}} \qquad \text{SAS:} Z = \frac{S - E(S) + 0.5}{\sqrt{varS}}$$

$$Z \sim \mathcal{N}(0,1) \qquad \text{approximativ}$$

(0.5 = Stetigkeitskorrektur)
bei Bindungen: korrigierte (kleinere) Varianz

```
Npar1way_Carnitinfraktion.sas
Npar1way_Banknote.sas
Npar1way_Heroin.sas
Npar1way_Tibetan.sas
```

Nichtparametrische Tests
Wilcoxon-Test (3)

- SAS gibt die Teststatistik (Z) und die ein- und zweiseitigen p-Werte an.

a) $H_0 : \mu_1 = \mu_2 \qquad H_1 : \mu_1 \neq \mu_2$

$\Rightarrow$ two-sided $Pr > |Z| = P(|Z| > Z)$

b) $H_0 : \mu_1 \leq \mu_2 \qquad H_1 : \mu_1 > \mu_2$

$\Rightarrow$ one-sided $z > 0$

$\rightarrow P(Z > z) = Pr > Z$

c) $H_0 : \mu_1 \geq \mu_2 \qquad H_1 : \mu_1 < \mu_2$

$\Rightarrow$ one-sided $z < 0$

$\rightarrow P(Z < z) = Pr < Z$

- SAS bietet die Normalapproximation und die t-Approximation an.

PROC NPAR1WAY WILCOXON; CLASS x; **RUN**;

Nichtparametrische Tests

5.9.2 Zwei unverbundene Stichproben: Kolmogorov-Smirnov Test

Wir machen gar keine Verteilungsannahmen.

$$\begin{array}{llll} H_0: & F_1 = F_2 & H_1: & F_1 \neq F_2 \\ H_0: & F_1 \leq F_2 & H_1: & F_1 > F_2 \\ H_0: & F_1 \geq F_2 & H_1: & F_1 < F_2 \end{array}$$

Kolmogorov-Smirnov Test

$$\begin{aligned} D &= max_i |F_1(x) - F_2(x)| \quad \text{(zweiseitig, EDF)} \\ D^+ &= max_i (F_1(x) - F_2(x)) \quad \text{(einseitig, D)} \\ D^- &= max_i (F_2(x) - F_1(x)) \quad \text{(einseitig, D)} \end{aligned}$$

PROC NPAR1WAY EDF D;

Zweistichprobenproblem

Allgemeine Empfehlungen

- Wenn Normalverteilung, gleiche Varianzen und keine Ausreißer: t-Test
- Wenn Normalverteilung, ungleiche oder unbekannte Varianzen und keine Ausreißer: Welch-Test (t-Test, unpooled, Satterthwaite)
- Wenn “sehr nahe” an Normalverteilung und keine Ausreißer: wie bei Normalverteilung
- keine Normalverteilung, gleiche Varianzen, und etwa gleicher Verteilungstyp (Ausreißer in begrenztem Maße erlaubt): Wilcoxon Test

  oder: Adaptiver Test (von SAS nicht angeboten)
- keine Normalverteilung, Verteilungstypen verschieden, ungleiche Varianzen: K-S Test

  oder: Brunner-Munzel Test (von SAS nicht angeboten)

Nichtparametrische Tests

5.9.3. Mehrere unverbundene Stichproben

Modell:

$$Y_{ij} = \mu_i + \epsilon_{ij}, \quad \epsilon_{ij} \sim (0, \sigma^2), \quad j = 1, ..., n_i, i = 1, .., k$$

$H_0 : \mu_1 = ... = \mu_k \qquad H_1 : \exists(\mu_{i_1}, \mu_{i_2}) \;\; \mu_{i_1} \neq \mu_{i_2}$

Wir fassen alle Beobachtungen
$X_{11}, \ldots, X_{1n_1}, \ldots, X_{k1}, \ldots, X_{kn_k}$
zusammen und bilden die Rangzahlen R_{ij}, $i = 1...k, j = 1...n_i$.

Mit den Rangzahlen führen wir eine
einfaktorielle Varianzanalyse durch
= Kruskal-Wallis Test

Nichtparametrische Tests

Mehrere unverbundene Stichproben

Kruskal-Wallis Test

$$KW = \frac{\sum_{i=1}^{k}(T_i - E_0(T_i))^2 \cdot n_i}{S^2}, \quad \text{wobei}$$

$$T_i = \frac{1}{n_i}\sum_{j=1}^{n_i} R_{ij} \quad \text{mittl. Rangsumme der i-ten Gruppe}$$

Kruskal-Wallis	Varianzanalyse
T_i	$\overline{Y}_{i.}$
$E_0 T_i = \frac{N+1}{2}$	$\overline{Y}_{..} = \overline{Y}$
Zähler	SSB
$S^2 = \frac{(N-1)N(N+1)}{12}$	SST
$= \sum_i \sum_j (R_{ij} - \frac{N+1}{2})^2)$	

$N = \sum_{i=1}^{k} n_i$ Gesamtstichprobenumfang

Nichtparametrische Tests
Kruskal-Wallis-Test

$$
\begin{aligned}
S^2 &= \sum_i \sum_j (R_{ij} - \frac{N+1}{2})^2 = \\
&= \sum_{k=1}^{N} (k - \frac{N+1}{2})^2 \\
&= \sum_k k^2 - (N+1) \sum_k k + \frac{(N+1)^2}{4} \cdot N \\
&= \frac{N(N+1)(2N+1)}{6} - \frac{N(N+1)^2}{2} + \frac{(N+1)^2}{4} \cdot N \\
&= \frac{(N+1) \cdot N}{12} \big(4N + 2 - 6N - 6 + 3N + 3\big) \\
&= \frac{N(N+1)}{12} \cdot (N-1) = \frac{(N-1) \cdot N \cdot (N+1)}{12}.
\end{aligned}
$$

Nichtparametrische Tests
Kruskal-Wallis-Test

Vorteil: S^2 ist nicht zufällig, hängt nur vom Stichprobenumfang ab.

$KW \sim \chi^2_{k-1}$ (asymptotisch)

H_0 ablehnen, falls p-value = ”Pr $>$ Chi Square” $< \alpha$

SAS-Output

Mean Score:	T_i
Chi-Square:	realisierte KW
DF=$k-1$:	Freiheitsgrade.

```
Npar1way_Maschinen.sas
~\Varianzanalyse_Modelle\PI12erg.sas
```

Nichtparametrische Tests
Kruskal-Wallis-Test

- Bei Bindungen erfolgt eine Korrektur der Statistik
- KW-Test ist relativ effizient bei NV. Bei Nicht-NV meist besser als der VA-F-Test.
- KW-Test hält (wie alle nichtparam. Tests) asymptotisch das Signifikanzniveau ein.
- kleine Stichproben ($N \leq 20$): Option EXACT möglich

```
PROC NPAR1WAY WILCOXON;
    CLASS Faktor;
   VAR var;
RUN;
```

Nichtparametrische Tests

5.9.4 Mehrere verbundene Stichproben-Friedman Test

Modell, wie bei der 2-faktoriellen Varianzanalyse

$$Y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij} \quad \epsilon_{ij} \sim (0, \sigma^2),$$

$$j = 1...k, i = 1...n$$

$H_0 : \beta_1 = ... = \beta_k (= 0) \qquad H_1 : \exists (j_1, j_2) : \beta_{j1} \neq \beta_{j2}$

Ränge werden zeilenweise gebildet, $Y_{1(1)} \leq ... \leq Y_{1(k)}$
R_{ij} der Rang von Y_{ij} in der i-ten Zeile.

Nichtparametrische Tests

Friedman Test

	Behandlung				
Block	1	2	...	k	Zeilensummme
1	R_{11}	R_{12}	...	R_{1k}	$\frac{k(k+1)}{2}$
.					
.					
n	R_{n1}	R_{n2}	...	R_{nk}	$\frac{k(k+1)}{2}$
	$R_{.1}$	$R_{.2}$	...	$R_{.k}$	$\frac{nk(k+1)}{2}$
	$n\overline{R}_{.1}$	$n\overline{R}_{.2}$	...	$n\overline{R}_{.k}$	

$$F_k = \frac{n^2 \sum_{j=1}^{k} (\overline{R}_{.j} - E(\overline{R}_{.j}))^2}{n \cdot k(k+1)/12}$$

Nichtparametrische Tests

Friedman Test

$$F_k = \frac{n^2 \sum_{j=1}^{k} (\overline{R}_{.j} - E(\overline{R}_{.j}))^2}{n \cdot k(k+1)/12}$$

$\overline{R}_{.j} = \frac{1}{n} \sum_{i=1}^{n} R_{ij}$ Spaltenmittel der j-ten Spalte (Vergleiche mit $\overline{Y}_{.j}$)

$E\overline{R}_{.j} = \frac{1}{n} \cdot \frac{n(k+1)}{2} = \frac{k+1}{2}$ (Vergleiche mit $\overline{Y}_{..}$)

Unter$H_0 : F_k \sim \chi^2_{k-1}$(asympt.)

H_0 ablehnen, falls $F_k > \chi^2_{1-\alpha,k-1}$
oder falls p-value $< \alpha$.

Nichtparametrische Tests
Friedman-Test

- Bei Bindungen Korrektur des Nenners.
- Für kleinere n ist Friedman-Test (asy.) meist etwas konservativ.
- Für größere k (etwa $k \geq 5$) ist der Friedman-Test (bei NV) einigermaßen effizient.
- Für $k = 2$ ist der Friedman-Test zum Vorzeichentest äquivalent (also nicht besonders effizient).

Friedman-Test

Durchführung des Friedman-Tests

```
PROC FREQ;
   TABLES Faktor A * Faktor B * Y
      /CMH2 SCORES=RANK NOPRINT;
RUN;
```

NOPRINT: unterdrückt den Druck von Kontingenztafeln

SCORES=RANK: Ränge werden (zeilenweise) gebildet.

CMH2: Cochran-Mantel-Haenszel

`Test_Friedman_Hypnose.sas`
`Test_Friedman_Synchro.sas`

Hier ist nur die folgende Zeile interessant:
Row Mean Scores Differ