Evaluation
&
Forschungsstrategien

Seminar

WS2011/12

Prof. Dr. G. Meinhardt
Johannes Gutenberg Universität Mainz

Seminar

Anwendung statistischer Verfahren in

- Überblick
- Grundprinzip
- wichtigsten mathematischen Beziehungen
- Anwendungsbeispiele in Excel & Statistica

HA/Tut

- Vertiefung mit Anwendungsbeispielen
- Aufgabenbearbeitung mit Excel & Statistica

Prüfung

Mündliche Modulabschlußprüfung

Stud.Leistung

Bearbeitung eines Project Files

Einführung

Evaluationsproblem am Beispiel der Wirksamkeitsprüfung einer therapeutischen Maßnahme: Grundprobleme und Prüfstrategien

Verfahren

- ANOVA, Hotelling's T^2
- ANOVA – Messwiederholungsdesigns/Trendanalyse
- Effektstärkenprüfung
- Faktoranalyse
- Fallklassifikation

Versuchspläne

Typische Designs aus der klinischen Psychologie

Ziele

- Wissen über statistische Verfahren
- Wissen über Untersuchungsstrategien
- Umsetzung mit Software

Evaluations-problem

Problemstellung

Eine Psychologin leitet eine Therapie-Evaluationsstudie zur Wirksamkeit einer neuen Behandlungsmethode für depressive Verstimmungen. Dazu werden 22 Patienten über 8 Wochen lang mit der neuen Methode behandelt, und die Werte in 3 Kontrollvariablen V1: erlebte Beanspruchung, V2: Ohnmacht/Hilflosigkeit, V3: Körperbeschwerden jede Woche erhoben (Vx: Mehritemskalen, die einen intervallskalierten Score pro Variable liefern). Zugleich existiert eine Kontrollgruppe, an der die Variablen ebenfalls erhoben werden. Alle 3 Variablen sind für die Beurteilung des Behandlungserfolges gleichermaßen wichtig, und sollten in das Urteil über die Eignung der Behandlungsmethode eingehen.

Design

ANOVA mit einem **Gruppierungsfaktor** und mind. einem **Messwiederholungsfaktor**

	t1	t2	…	tk
Control	v1 v2 v3	v1 v2 v3	…	v1 v2 v3
Therapy	v1 v2 v3	v1 v2 v3	…	v1 v2 v3

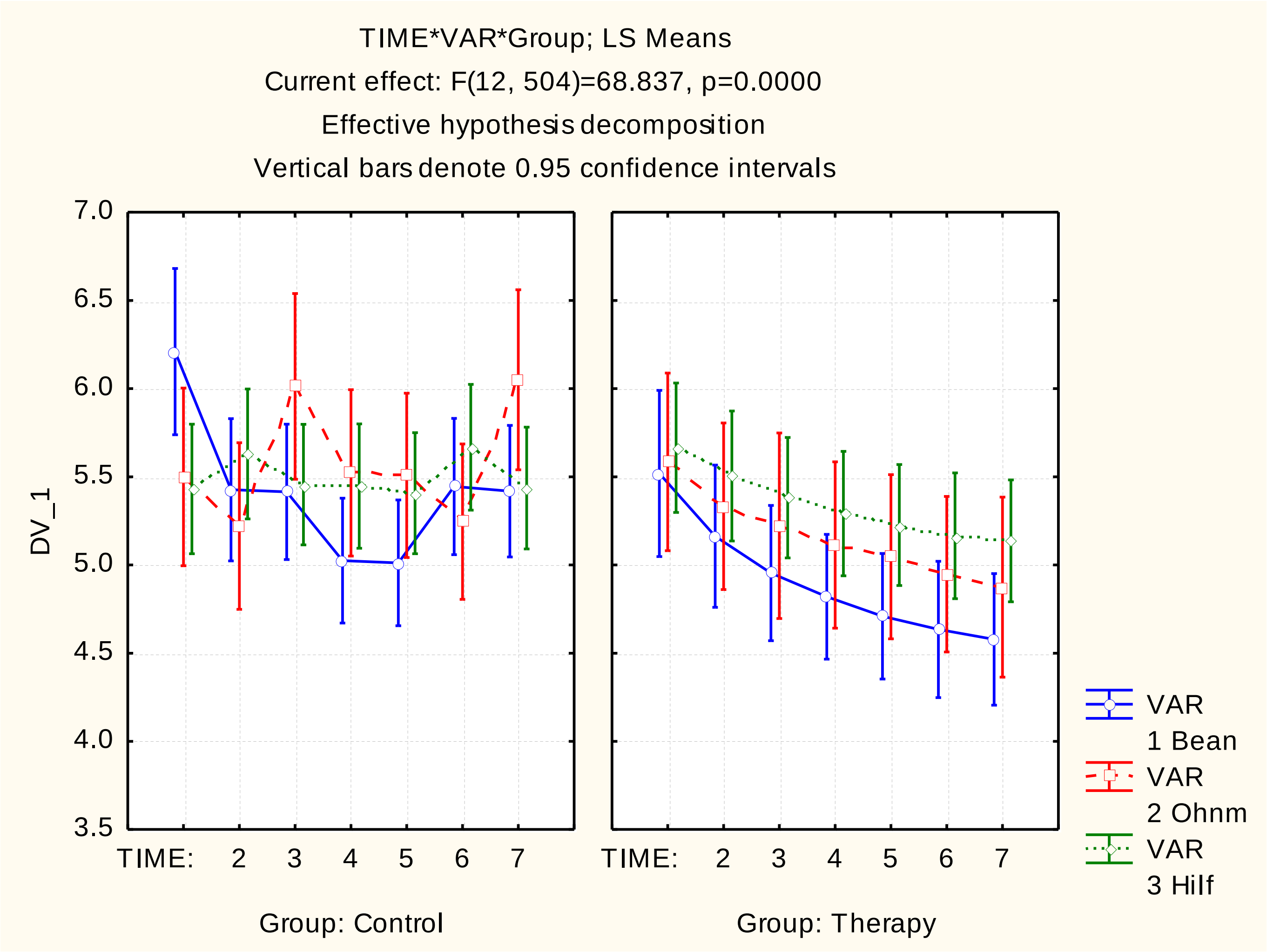
Hierbei wird jede Variable V_j auf n_i Versuchspersonen gemessen.

i = 1: Kontrollgruppe

i = 2: Therapiegruppe

(die Stichprobenumfänge beider Gruppen dürfen verschieden sein)

Daten

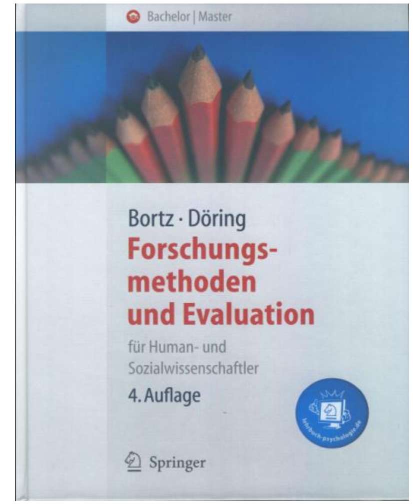

TIME*VAR*Group; LS Means
Current effect: F(12, 504)=68.837, p=0.0000
Effective hypothesis decomposition
Vertical bars denote 0.95 confidence intervals
DV_1
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
TIME: 1 2 3 4 5 6 7
Group: Control
TIME: 1 2 3 4 5 6 7
Group: Therapy
VAR
1 Bean
VAR
2 Ohnm
VAR
3 Hilf

Fragen

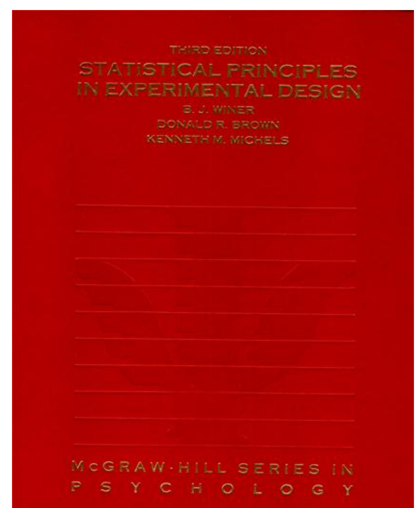
Literatur

a) **Bortz**

b) **Bortz/Döring**



c) **Winer**



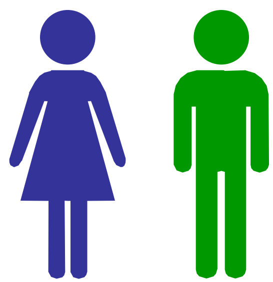
Univariate Mittelwertevergleiche - *Problemstellung*

Problem

Gruppierungsvariable → Messgröße (metrisch)

Beispiel

Geschlecht → x



M J

Anzahl der gefundenen Zielelemente in einem Konzentrationsleistungstest

---

Frage

Unterscheidet sich die Leistung von Mädchen und Jungen im statistischen Mittelwert ?

Univariate Mittelwertevergleiche - *Problemstellung*

Stichprobe

Wir untersuchen 40 Mädchen und 45 Jungen

Beispieldaten

Geschlecht

M	J
$\bar{x}_M$	$\bar{x}_J$
23.7	17.2

$$\bar{x}_M - \bar{x}_J = \Delta\bar{x}$$

$$23.7 - 17.2 = 6.5$$

Frage

Gibt es wirkliche Leistungsunterschiede zwischen Jungen und Mädchen, oder ist der gefundene Unterschied „rein zufällig“ ?

Univariate Mittelwertevergleiche - *Prüfstrategie*

Strategie

Ermittle die Wahrscheinlichkeit für den beobachteten Mittelwertsunterschied unter der Annahme, dass beide Gruppen in der Population denselben Mittelwert besitzen

Annahme

Die Populationsmittelwerte von Jungen und Mädchen sind gleich

Null-Hypothese

Alternativ-Hypothese

$$H_0 : \mu_J = \mu_M$$

$$H_1 : \mu_J \neq \mu_M$$

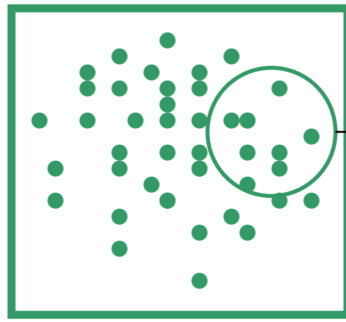
Urteil

Ist der beobachtete Mittelwertsunterschied unter der H_0 sehr unwahrscheinlich (höchstens 5%), so lehnen wir die H_0 ab, und sehen die H_1 als die bessere Alternative an.

Theoretische Verteilung – *Sampling Distribution*

Sampling

Population der Jungen



Stichprobe des Umfangs N_J

$\overline{x}_J$

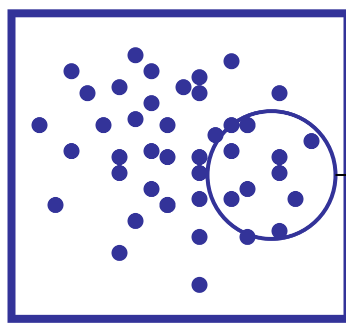
Mittelwertsdifferenz

$$\Delta\overline{x} = \overline{x}_M - \overline{x}_J$$

Tue dies k - mal:

$$\Delta\overline{x}_1 = \overline{x}_{M1} - \overline{x}_{J1}$$
$$\Delta\overline{x}_2 = \overline{x}_{M2} - \overline{x}_{J2}$$
$$\vdots \qquad \vdots$$
$$\Delta\overline{x}_k = \overline{x}_{Mk} - \overline{x}_{Jk}$$

Population der Mädchen



$\overline{x}_M$

Stichprobe des Umfangs N_M

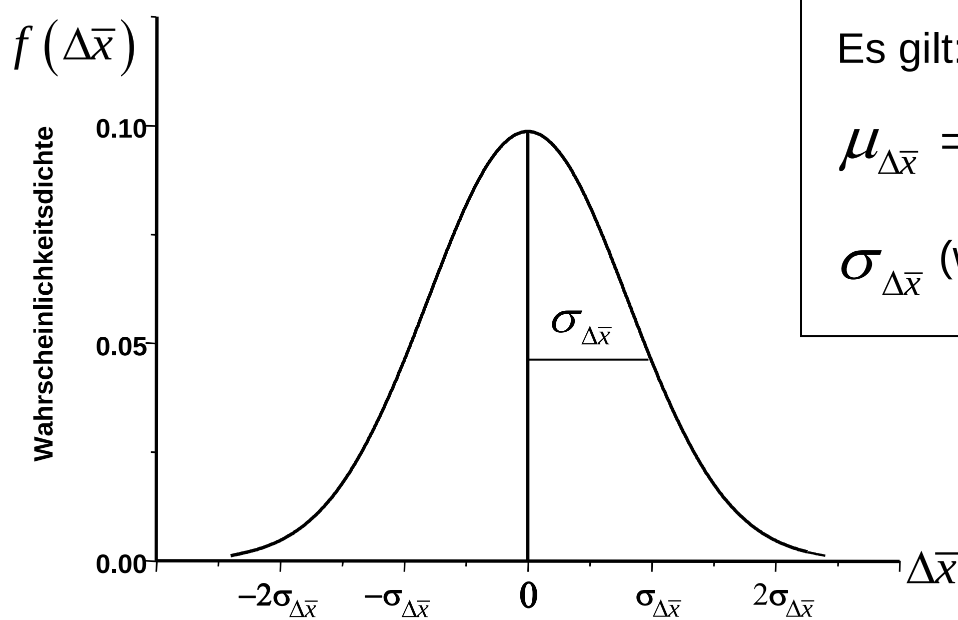
Verteilung der Differenzen von Mittelwerten

$$\left(\Delta\overline{x}_1 \quad \Delta\overline{x}_2 \quad \ldots \quad \Delta\overline{x}_i \quad \ldots \quad \Delta\overline{x}_k\right)$$

Theoretische Verteilung – *Sampling Distribution*

Central Limit Theorem

Die Verteilung von Differenzen von Mittelwerten nähert sich mit wachsendem Umfang der Sample-Stichproben einer Normalverteilung. Für N > 30 ist die Approximation gut.



Es gilt:

$$\mu_{\Delta\bar{x}} = 0$$

$\sigma_{\Delta\bar{x}}$ (wird geschätzt)

Inferenzstat. Schluss

In der theoretischen Verteilung der Differenzen von Mittelwerten wird die Wahrscheinlichkeitsbestimmung vorgenommen. Sie liegt dem inferenzstatistischen Schluss zugrunde.

Sampling Distribution – *Bestimmung des Standardfehlers*

Unabhängigkeit

Ist die Messvariable eine in beiden Populationen unabhängige ZV:

$$\sigma_{\Delta\bar{x}} = \sqrt{\frac{\sigma_M^2}{N_M} + \frac{\sigma_J^2}{N_J}}$$

Gleichheit der Populations-varianz

Jungen und Mädchen kommen aus derselben Population

$$\sigma_M^2 = \sigma_J^2 = \sigma^2$$

$$\Downarrow$$

Standardfehler

$$\sigma_{\Delta\bar{x}} = \sqrt{\sigma^2\left(\frac{1}{N_M} + \frac{1}{N_J}\right)}$$

Sampling Distribution – *Schätzung des Standardfehlers*

Schätzung aus Stichproben

Für die Populationsvarianz verwendet man eine Schätzung aus den Daten beider Stichproben:

"Pooling"

$$\hat{\sigma}^2 = \frac{N_M \cdot s_M^2 + N_J \cdot s_J^2}{N_M + N_J - 2} = \frac{SAQ_M + SAQ_J}{df_M + df_J}$$

wobei s_M^2 und s_J^2 die Stichprobenvarianzen sind

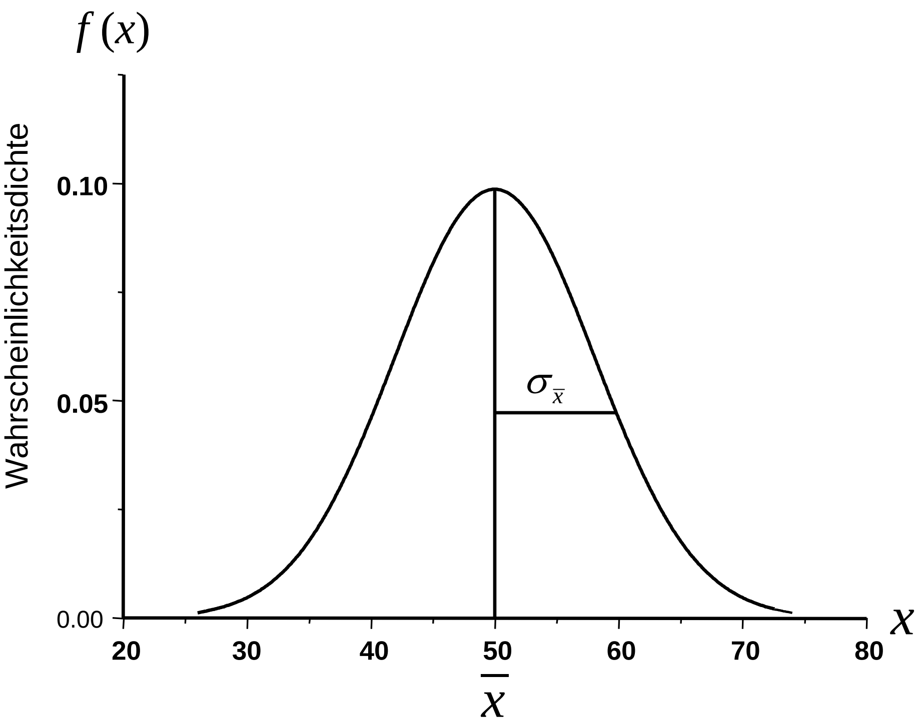
Dann gilt

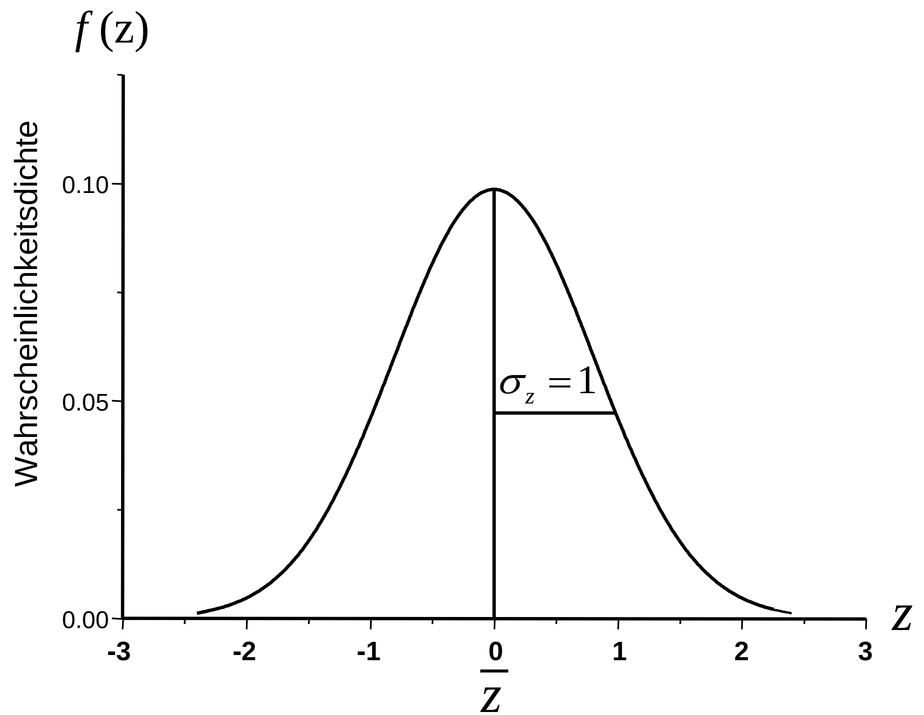
Schätzformel

$$\hat{\sigma}_{\Delta\bar{x}} = \sqrt{\frac{N_M \cdot s_M^2 + N_J \cdot s_J^2}{N_M + N_J - 2}} \cdot \sqrt{\frac{1}{N_M} + \frac{1}{N_J}}$$

(Beste Schätzung des Standardfehlers aus Stichprobendaten)

Normalverteilung – *z*–*Standardnormalverteilung*





Normalverteilung

Standard-Normalverteilung

$$z = \frac{x - \bar{x}}{s}$$

Die z- Transformation übersetzt die Rohdatenskala in die Standardskala ($\bar{z} = 0$, $\sigma_z = 1$)

Sampling Distribution – *Prüfgrösse*

z- Skala der Differenzen von Mittelwerten

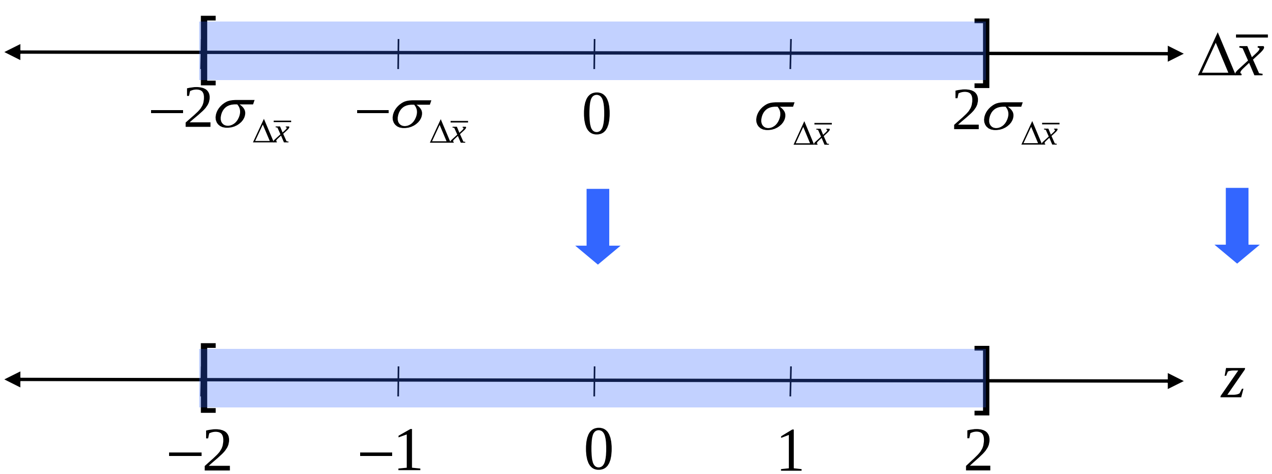
$$z = \frac{\Delta\bar{x} - \mu_{\Delta\bar{x}}}{\sigma_{\Delta\bar{x}}}$$

Unter der H0 gilt $\mu_{\Delta\bar{x}} = 0$

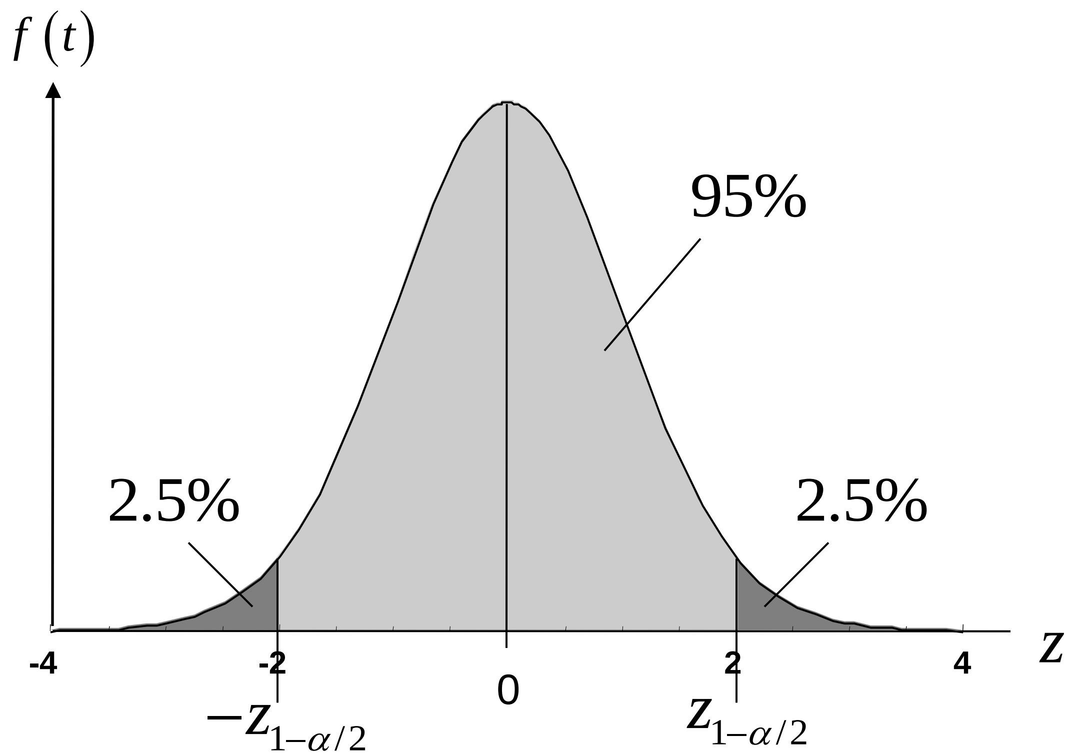
Prüfgrösse

Dann gilt: $z = \dfrac{\Delta\bar{x}}{\sigma_{\Delta\bar{x}}}$ ist standardnormalverteilt

Transformation



Entscheidung über Prüfgrösse mit Standardnormalverteilung



Prüfgrösse

$$z = \frac{\Delta\bar{x}}{\sigma_{\Delta\bar{x}}}$$

Signifikanzniveau

$$\alpha = 0.05$$

$$P\left(|z| > z_{1-\alpha/2}\right) = \alpha$$

Ablehnungsbereich	Annahmebereich	Ablehnungsbereich
$z < z_{1-\alpha/2}$	$\lvert z\rvert \leq z_{1-\alpha/2}$	$z > z_{1-\alpha/2}$

Testen zum Signifikanzniveau α: Ist $|z| > z_{1-\alpha/2}$?

Entscheidung über Signifikanz des Mittelwerteunterschieds

1. Prüfgrösse

Berechne $$z = \frac{\Delta\bar{x}}{\sigma_{\Delta\bar{x}}}$$

2. Kritischer z - Wert

Ermittle kritischen z - Wert $z_{1-\alpha/2}$ für ein α– Fehlerniveau

3. Entscheide

A. Gilt $|z| > z_{1-\alpha/2}$ → Ablehnung von H_0

(die Mittelwerte der J. und M. sind signifikant verschieden)

B. Gilt $|z| \leq z_{1-\alpha/2}$ → Beibehalten von H_0

(die Mittelwerte der J. und M. unterscheiden sich zufällig)

Numerisches Beispiel

Differenz der Mittelwerte

$\bar{x}_M$	$\bar{x}_J$
23.7	17.2
s_M^2	s_J^2
173	106

$$\bar{x}_M - \bar{x}_J = \Delta\bar{x}$$

$$23.7 - 17.2 = 6.5$$

Standardfehler

$$\hat{\sigma}_{\Delta\bar{x}} = \sqrt{\frac{40 \cdot 173 + 45 \cdot 106}{40 + 45 - 2}} \cdot \sqrt{\frac{1}{40} + \frac{1}{45}} = 2.58$$

Prüfgrösse und Kritischer Wert

$$z = \frac{6.5}{2.58} = 2.52 \qquad z_{1-\alpha/2} = z_{0.975} = 1.96$$

Entscheidung

$2.52 > 1.96$ ➡ d.h. $|z| > z_{1-\alpha/2}$ ➡ H_0 ablehnen

Die Mittelwerte entstammen nicht derselben Population (unterscheiden sich signifikant)

Voraussetzungen der Prüfung

Varianz-homogenität

a. Die Populationsvarianzen die beiden Stichproben zu Grunde liegen, müssen gleich (homogen) sein. (Prüfung mit geeignetem Verfahren)

Unabhängigkeit

b. Die Messeinheiten innerhalb jeder Stichprobe müssen unabhängig sein.
c. Die Messeinheiten beider Stichproben dürfen nicht teilweise paarweise zuzuordnen sein.

Verletzungen

Der Test ist relativ robust gegen Verletzungen der Varianzhomogenität. Verletzungen der Unabhängigkeit (b.) führen zur Ungültigkeit der Prüfgrösse, der Unabhängigkeit (c.) je nach Höhe der Korrelationen zu progressiven (kleine Korr.) oder zu konservativen Entscheidungen (hohe Korr.).

Mittelwertsprüfung bei mehreren Variablen

Beispiel

Lebenszufriedenheit

10 Variablen

Arbeit

X_1: Gehalt

X_2: Entscheidungsfreiheit

X_3: Qualität der Kommunikation

Privatsphäre

X_4: Ehe

X_5: Freunde/Beziehungen

X_6: Sexualität

Person

X_7: Lebensansprüche

X_8: Sinnhaftigkeit

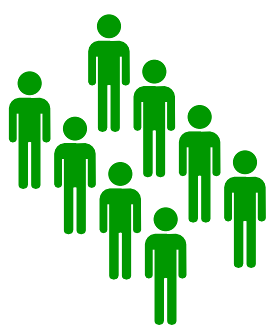
Aktivität

X_9: Hobbies

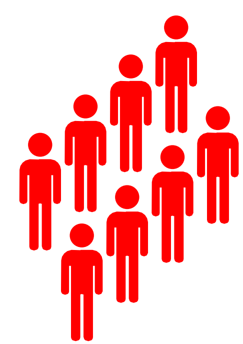
X_{10}: Sport/Fitness

$$\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{10}$$

2 Gruppen



Gesunde



Herzinfarktpatienten

Multivariate Mittelwertsvergleiche - Einzeltestungen

Frage

Unterscheiden sich Gesunde und Patienten im Variablen-komplex **Lebenszufriedenheit**?

Teststrategie

Wir testen auf jeder der 10 Skalen den Gruppenunterschied mit einem t- Test. Wenn **irgend einer** der Tests signifikant wird, sehen wir die Gruppen als verschieden an.

Probleme

1. **Multiples Testen:** Dieselbe Hypothese wird 10 mal geprüft.
2. **Unterstellte Unabhängigkeit:** Man behandelt die einzelnen Skalen als unabhängig voneinander.
3. **Fehlendes Konstrukt:** Lebenszufriendenheit wird nicht als Variablenkomplex mit Binnenstruktur behandelt.
4. **Mangelnde Teststärke:** Man nutzt nicht die Korrelations-struktur der Variablen für einen leistungsfähigen Test.

Ausweg

Verwendung **eines** multivariaten Tests, der die Information **aller** 10 Variablen und ihrer Korrelationsstruktur in **eine** statistische Prüfgrösse einfliessen lässt.

Einzeltestungen - Bonferronikorrektur

α – Fehler Kumulierung

Bei simultanen Einzeltestungen „kumuliert“ sich das α – Risiko:

$$\hat{\alpha} = P(\text{mind. 1 falsch}) = 1 - P(\text{keinen 1 falsch})$$
$$= 1 - P(T_1 \cap T_2 \cap \ldots \cap T_m)$$
$$= 1 - (1-\alpha)(1-\alpha)\ldots(1-\alpha)$$
$$= 1 - (1-\alpha)^m$$

Overall α

Setzt man das overall $\hat{\alpha}$ -Niveau fest und löst nach auf, folgt

$$\alpha = 1 - (1-\hat{\alpha})^{1/m}$$
$$\approx \frac{\hat{\alpha}}{m}$$

Bonferroni Approximation

Um alle m Tests auf einem konventionellen Alpha Niveau abzusichern, muss dieses durch die Anzahl der Tests geteilt werden. Bei 10 Tests muss man für ein overall Alpha = 5% ein Test-Alpha von 0.5% verwenden.

Multivariate Mittelwertsvergleiche - Verfahren

Variablen-komplex

$$\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{10}$$

Multivariates Testkonstrukt

Multivariate Distanz (Mahalanobisdistanz)

Multivariate Quadratsummen (SSCP-Matrizen-Zerlegung)

Optimale Linearkombination (Linear Discriminant Function)

Verfahren

Hotelling‘s T^2

MANOVA

Diskriminanz-Analyse

Alle Verfahren entscheiden über den Gruppenunterschied im **gesamten Variablenkomplex** mit **einem** statistischen Test