$\alpha=0.05$

$1-\beta=0.90$

Grundlagen der Biometrie

Beschreibende und schließende Statistik in klinischen Studien

PD Dr. Thomas Sudhop & Dr. med. Dipl. chem. Michael Reber
Abteilung für Klinische Pharmakologie
Universität Bonn

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

"Jede mathematische Formel reduziert die Anzahl der Zuhörer um 50%"

Wie viele Formeln werden benötigt, um den Saal zu leeren?

"Statistik"

Lehre von den Verteilungen

- **Deskriptive Statistik = empirische Verteilungen von Merkmalen**
- **Induktive/Analytische Statistik = Schließen von einer Stichprobe auf die Grundgesamtheit**
- **Wahrscheinlichkeitstheorie = Verteilungen von Zufallsvariablen**

Deskriptive Statistik

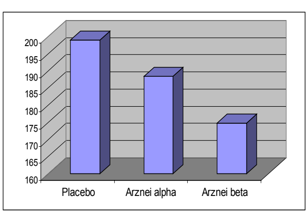
Aufgabe:

Strukturierung der Rohdaten

Deskriptive Statistik

- **Tabellen / Graphische Darstellung**

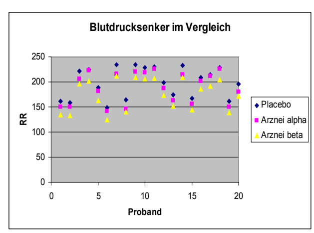
Patient	Placebo	Arznei alpha	Arznei beta
1	161	150	135
2	158	150	133
3	222	206	196
4	225	223	201
....			
18	228	226	204
19	162	150	139
20	196	180	172
Mittelwerte	198,95	188,4	174,75



Deskriptive Statistik

- **Tabellen / Graphische Darstellung**

Patient	Placebo	Medisan alpha	Medisan beta
1	161	150	135
2	158	150	133
3	222	206	196
4	225	223	201
....			
18	228	226	204
19	162	150	139
20	196	180	172
Mittelwerte	198,95	188,4	174,75



Deskriptive Statistik

- **Was?**

  Strukturierung der Rohdaten
- **Wie?**

  Verwendung mathematischer Methoden zur standardisierten Erfassung bestimmter Merkmale der erhobenen Daten
- **Warum?**

  Hervorheben wesentlicher Zusammenhänge durch Datenreduktion und graphische Darstellung um anderen Personen ohne Kenntnisse der Einzeldaten die erhobenen Beobachtungen vermitteln zu können

Population

- **Population (Grundgesamtheit)**

  Die **Grundgesamtheit** sind alle Individuen, für welche Schlussfolgerungen gezogen werden sollen.

  - Alle Einwohner eines Bundeslandes
  - Alle Autos in Deutschland
  - Alle Typ II Diabetiker (Zielpopulation)
- Populationen weisen einen großen Umfang (=Menge der Elemente) auf und können daher nicht vollständig untersucht werden.

Stichprobe

- **Stichprobe**

  Eine **Stichprobe** aus einer Population stellt die Anzahl von Individuen dar, welche tatsächlich beobachtet werden.
- Der **Stichprobenumfang** (Elemente der Stichprobe = Fallzahl) muss ausreichend groß sein
- **Stichproben** sollten **repräsentativ** für die Population sein

Repräsentative Stichprobe

- **Stichprobe sollte Elemente aus allen Bereichen der Population umfassen**

  ✓ Alle PKW, welche an einem Stichtag zugelassen wurden

  ✗ Alle roten PKW in Berlin sind **nicht** repräsentativ für alle PKW

Univariante deskriptive Statistik

- **Kurze und prägnante Charakterisierung der Daten einer Stichprobe**
  - **Statistische Kennwerte**
    - **Lagemaße**
    - **Streumaße**
  - **Graphische Darstellung**

Lagemaße

- Mittelwerte
  - **Arithmetisches Mittel**
  - **Geometrisches Mittel**
  - **Harmonisches Mittel**
  - **Getrimmtes Mittel**
- Median

Lagemaße

? Wo liegt das Zentrum der Daten

? Was ist ein typischer mittlerer Wert

Arithmetisches Mittel

- Der Mittelwert beschreibt das Verhalten der Daten „im Mittel" (Σ = Summe)
- Er ist der durchschnittliche Wert aller Elemente einer Menge
- Nachteil: empfindlich gegenüber Extremen
- Berechnung:

Mittelwert = Summe aller Element : Anzahl aller Elemente

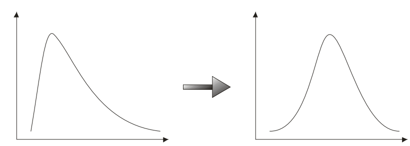
$$\bar{x} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n}$$

Geometrisches Mittel

- Findet häufig Anwendung in der Pharmakokinetik
- ⊕ Weniger empfindlich gegen Extremwerte
- Berechnung erfordert log.-Transformation
- Berechnung:

$$\bar{x} = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdot \ldots \cdot x_n}$$

Log - Transformation



statistische Verfahren beruhen auf der Annahme, dass Versuchsdaten sich der ***Normalverteilung*** annähern

Log - Transformation

- Anpassung der Transformation durch Auswahl des Logarithmus
- Anwendung bei rechtschiefer Verteilung (Es liegen mehr Werte rechts vom Mittelwert)

$$\bar{x} = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdot \ldots \cdot x_n}$$

$$\ln(\bar{x}) = \frac{\ln(x_1) + \ln(x_2) + \ldots + \ln(x_n)}{n}$$

$$e^{\ln(\bar{x})} = \text{Geometrisches Mittel}$$

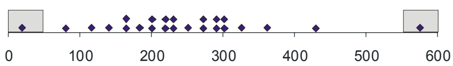
Harmonisches Mittel

- Es dient als Lagemaß, wenn die Beobachtungswerte Verhältniszahlen sind (z.B. zur Berechnung einer durchschnittlichen Geschwindigkeit oder Überlebenszeit). Bsp.: Ohmsches Gesetz
- Berechnung:

$$\bar{x}_H = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

Getrimmtes Mittel

- Entspricht einem Arithmetischen Mittel
- Vor der Berechnung werden an beiden Enden der Verteilung die Extremwerte gekappt (grau unterlegt)



19

Median

- **Der Median beschreibt den mittleren Wert in einer sortierten Stichprobe**

Berechnung:

- **Stichprobe aufsteigend sortieren**
- **Bei ungeradem Stichprobenumfang**
  - **⇒ Mittleres Element ist der Median**
- **Bei geradem Stichprobenumfang**
  - **⇒ Median ist der Mittelwert aus den beiden mittleren Elementen**

20

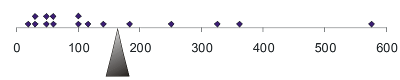
Median Beispiel

- Bestimmung des Alters-Medians von 6 Patienten
  - Alter der Patienten: 48, 50, 46, 52, 47, 48
- 1. Schritt: aufsteigend sortieren
  - 46, 47, 48, 48, 50, 52
- 2. Schritt: Mittelwert der beiden mittleren Werte bilden
  - 46, 47, 48, 48, 50, 52
  - (48 + 48) ÷ 2 = 48
- Der Alters-Median der Patienten beträgt 48 Jahre

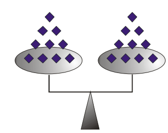
21

Mittelwert versus Median

- Der **Mittelwert** ist derjenige Wert, der die Daten auf einer Waage ausbalanciert. Entfernte Werte besitzen eine große Hebelkraft.



- Beim **Median** spielt der Abstand der Beobachtung keine Rolle. Der Median ist *robust* gegen Ausreißer.



22

Mittelwert versus Median

Die Wahl zwischen Mittelwert und Median ist:

- Abhängig davon, ob ein typischer oder ein mittlerer Wert gesucht wird
- Abhängig von der Verteilung (Normal, Schief oder „Gibt es Ausreißer?“)
- Abhängig davon, ob Präzision oder Robustheit im Vordergrund steht

23

Praktisches Beispiel Lagemaße

- Klinische Studie mit ACE-Hemmern
- 360 Probanden
- Randomisiert auf drei Behandlungsarme

24

Streumaße

- Streumaße liefern Informationen zur Zusammensetzung (Streuung) von Stichproben
- **Stichprobe A:** { 2, 2, 3, 3, 4, 4, 4, 5, 5, 6 }
- **Stichprobe B:** { 2, 2, 2, 5, 6, 9, 9, 19, 19, 21 }

25

Streumaße - Übersicht

- Range
- Standardabweichung
- Varianz
- Standardfehler
- Quantile / Perzentile

26

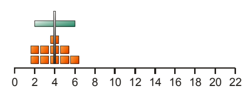
Range (Spannweite)

- Definition: Differenz aus größtem und kleinstem Element einer Stichprobe
- **Stichprobe A:** { 2, 2, 3, 3, 4, 4, 4, 5, 5, 6 }
  - Range: 6 - 2 = 4
- **Stichprobe B:** { 2, 2, 2, 5, 6, 9, 9, 19, 19, 21 }
  - Range: 21 - 2 = 19

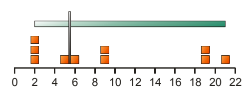
27

Range / Median

- Median und Range beschreiben Stichprobe
- **Stichprobe A:** { 2, 2, 3, 3, 4, 4, 4, 5, 5, 6 }
  - Median: 4
  - Range: 4



- **Stichprobe B:** { 2, 2, 2, 4, 5, 6, 9, 19, 19, 21 }
  - Median: 5,5
  - Range: 19



28

Streumaße - Übersicht

- Range
- Standardabweichung
- Varianz
- Standardfehler
- Quantile / Perzentile

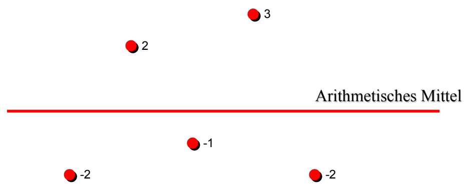
29

Standardabweichung

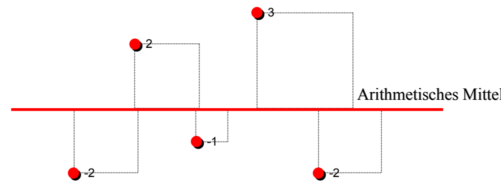
- Standardabweichung (engl. Standard deviation, SD) wird meist in Verbindung mit dem Mittelwert angegeben
  - Mittelwert ± Standardabweichung (Mean ± SD)
- Sie stellt ein Maß für die Streuung um den Mittelwert dar.
- Grobe Vorstellung: gibt den „durchschnittlich" Abstand des Einzelwertes vom Mittelwert an.

30

Standardabweichung



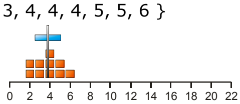
Standardabweichung



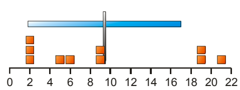
$$SD = \sqrt{\frac{(\bar{x}-x_1)^2 + (\bar{x}-x_2)^2 + (\bar{x}-x_3)^2 + \ldots + (\bar{x}-x_n)^2}{n-1}}$$

Standardabweichung

- Stichprobe A: { 2, 2, 3, 3, 4, 4, 4, 5, 5, 6 }
  - Mittelwert: 3.8 ± 1.3



- Stichprobe B: { 2, 2, 2, 5, 6, 9, 9, 19, 19, 21 }
  - Mittelwert: 9.4 ± 7.6



Standardabweichung

Proband	Blutdruck (syst.)	
	Tablette A	Tablette B
1	140	150
2	125	141
3	120	110
4	130	107
5	135	152
6	115	105
Mittelwert	**127,5**	**127,5**
SD	**9,4**	**22,5**

Streumaße - Übersicht

- Range
- Standardabweichung
- Varianz
- Standardfehler
- Quantile / Perzentile

Varianz

- Varianz = Standardabweichung²
- „Mittleres Abstandsquadrat“ der Elemente vom Mittelwert der Stichprobe
- Berechnung:

$$Varianz = \frac{(\bar{x}-x_1)^2 + (\bar{x}-x_2)^2 + (\bar{x}-x_3)^2 + \ldots + (\bar{x}-x_n)^2}{n-1}$$

Standardabweichung / Varianz

- Standardabweichung ist das meistgebrauchte Streuungsmaß
- Vorteil der Standardabweichung - gleiche Einheit wie die ursprünglichen Messwerte.

37

Streumaße - Übersicht

- Range
- Standardabweichung
- Varianz
- Standardfehler
- Quantile / Perzentile

38

Standardfehler des Mittelwerts (SEM)

- Standardfehler
  standard error of the mean = SEM
- Abgeleitet aus Standardabweichung (SD) und Stichprobenumfang (n)
- Immer kleiner als Standardabweichung

$$SEM = \frac{SD}{\sqrt{n}}$$

39

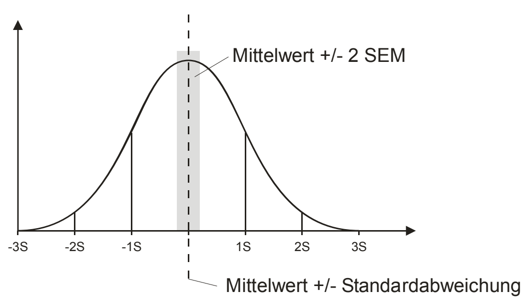
Standardfehler des Mittelwerts (SEM)

- Der Standardfehler beschreibt **nicht** die Daten.
- SEM gibt die Genauigkeit des Mittelwertes als Schätzwert an.
- CAVE: Häufig wird SEM anstelle des Standard-Abweichung verwandt. Die kleinere Maßzahl für SEM soll eine bessere Wirkung suggerieren.
- Nährung 95%-KI des Mittelwert:

  Mittelwert +/- 2 SEM

40

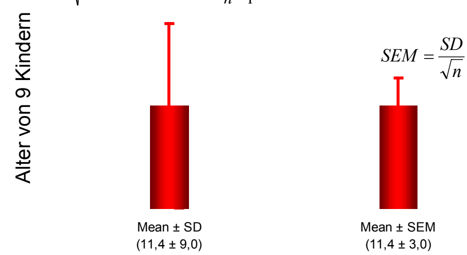
SD ⇔ SEM



41

SD > SEM

$$SD = \sqrt{\frac{(\bar{x}-x_1)^2 + (\bar{x}-x_2)^2 + (\bar{x}-x_3)^2 + \ldots + (\bar{x}-x_n)^2}{n-1}}$$



42

Streumaße - Übersicht

- Range
- Standardabweichung
- Varianz
- Standardfehler
- Quantile / Perzentile

Rang

- Definition
  - Position innerhalb der aufsteigend sortierten (Rang-)Liste einer Stichprobe
- Beispiel
  - Platzierungen im Sport
- Berechnung
  - Elemente aufsteigend sortieren
  - Beginnend bei „1" nummerieren

Meßwert	Rang
57	1
77	2
80	3
82	**4**
90	5
90	6
91	7
115	8
116	9
116	10
121	11
124	12
130	13
132	14
135	15
136	16
140	17
143	18
145	19
148	20

Perzentile

- Als x%-Perzentile wird derjenige Wert einer Stichprobe bezeichnet, der kleiner oder gleich x% aller Werte ist

Meßwert	Rangplatz	Perzentile
57	1	
77	2	10%
80	3	
82	4	20%
90	5	
90	6	30%
91	7	
115	8	40%
116	9	
116	10	50%
121	11	
124	12	60%
130	13	
132	14	70%
135	15	
136	16	80%
140	17	
143	18	90%
145	19	
148	20	100%

Perzentile - BMI

Jungen Alter (Jahre)	P3	P10	P25	P50 (M)	P75	P90	P97	P99.5
0	10,20	11,01	11,81	12,68	13,53	14,28	15,01	15,86
0,5	14,38	15,06	15,80	16,70	17,69	18,66	19,72	21,09
1	14,58	15,22	15,93	16,79	17,76	18,73	19,81	21,25
1,5	14,31	14,92	15,60	16,44	17,40	18,37	19,47	20,95
2	14,00	14,58	15,25	16,08	17,03	18,01	19,14	20,69
2,5	13,73	14,31	14,97	15,80	16,76	17,76	18,92	20,51
3	13,55	14,13	14,79	15,62	16,59	17,62	18,82	20,51
3,5	13,44	14,01	14,67	15,51	16,50	17,56	18,80	20,61
4	13,36	13,94	14,60	15,45	16,46	17,54	18,83	20,68
4,5	13,30	13,88	14,55	15,42	16,45	17,56	18,90	20,87
5	13,24	13,83	14,51	15,40	16,46	17,61	19,02	21,17
5,5	13,20	13,80	14,50	15,40	16,50	17,71	19,19	21,52
6	13,18	13,79	14,51	15,45	16,59	17,86	19,44	21,92
6,5	13,19	13,82	14,56	15,53	16,73	18,07	19,76	22,40
7	13,23	13,88	14,64	15,66	16,92	18,34	20,15	23,07
7,5	13,29	13,96	14,76	15,82	17,14	18,65	20,60	23,81
8	13,37	14,07	14,90	16,01	17,40	19,01	21,11	24,62

Quartile

- Bezeichnen die 25%, 50%, 75% und 100% - Perzentile

Meßwert	Rangplatz	Perzentile	Quartil
57	1		
77	2		
80	3		
82	4		
90	**5**	**25%**	**1. Quartil**
90	6		
91	7		
115	8		
116	9		
116	**10**	**50%**	**2. Quartil**
121	11		
124	12		
130	13		
132	14		
135	**15**	**75%**	**3. Quartil**
136	16		
140	17		
143	18		
145	19		
148	**20**	**100%**	**4. Quartil**

Inter-Quartil-Spannweite

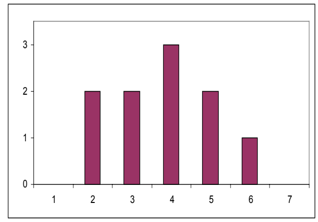
- „interquartile range"
- Bezeichnen die Differenz aus 3. und 1. Quartil
- 50% aller Werte einer Stichprobe liegen innerhalb dieses Bereichs

Meßwert	Rangplatz	Perzentile	Quartil
57	1		
77	2		
80	3		
82	4		
90	**5**	**25%**	**1. Quartil**
90	6		
91	7		
115	8		
116	9		
116	10	50%	2. Quartil
121	11		
124	12		
130	13		
132	14		
135	**15**	**75%**	**3. Quartil**
136	16		
140	17		
143	18		
145	19		
148	20	100%	4. Quartil

Graphische Darstellung

49

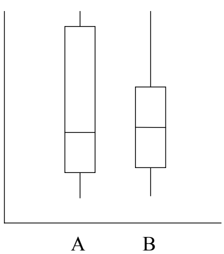
Grafik - Histogramm



Stichprobe A: { 2, 2, 3, 3, 4, 4, 4, 5, 5, 6 }

50

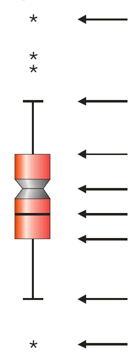
Quartile



- „Box" – Bereich von der 25. zur 75. Perzentile
- Stäbe (whiskers) sind nicht einheitlich definiert
  - Minimum / Maximum (SPSS)
  - 10% / 90% Perzentile

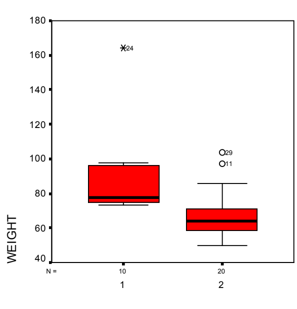
51

Grafik - Boxplots



52

Boxplots – Bsp. ACE-Hemmer



53

Zusammenfassung

- Die **deskriptive Statistik** beschreibt mathematische Eigenschaften des erhobene Datenmaterials anhand von **Stichproben**
- Es werden **Lagemaße** (Mittelwert, Median, 95%-Perzentile) von **Streumaßen** (Standardabweichung, Varianz, SEM, range, interquartile range) unterschieden.
- Anhand dieser Parameter können Untersuchungs-ergebnisse **standardisiert** berichtet werden, so dass es anderen gelingt, die Ergebnisse einer Untersuchung nachzuvollziehen, ohne alle Einzeldaten zu kennen.

54

Grundlagen der Biometrie

Beschreibende und schließende Statistik in klinischen Studien

PD Dr. med. Thomas Sudhop & Dr. med. Dipl. chem. Michael Reber
Abteilung für Klinische Pharmakologie
Universität Bonn

Wahrscheinlichkeit

- Verhältnis „Anzahl aller günstigen Ereignisse" zu „Anzahl aller möglichen Ereignisse"

$$p = \frac{\textit{Anzahl aller günstigen Ereignisse}}{\textit{Anzahl aller möglichen Ereignisse}}$$

- Wahrscheinlichkeit, mit einem Würfel im nächsten Wurf eine „6" zu werfen:

$$p = \frac{\{6\}}{\{1,2,3,4,5,6\}} = \frac{1}{6} = 0.1\overline{66666} \cong 16{,}7\%$$

- p liegt immer im Intervall [0; 1] (0-100%)

56

Chance (Odd)

- Verhältnis „Anzahl aller günstigen Ereignisse" zu „Anzahl aller ungünstigen Ereignisse"

$$p = \frac{\textit{Anzahl aller günstigen Ereignisse}}{\textit{Anzahl aller ungünstigen Ereignisse}}$$

- Chance, mit einem Würfel im nächsten Wurf eine „6" zu werfen:

$$p = \frac{\{6\}}{\{1,2,3,4,5\}} = \frac{1}{5} = 0{,}2 \cong 20\%$$

57

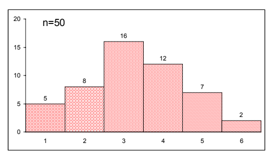
Absolute und relative Häufigkeit

Absolute Häufigkeit

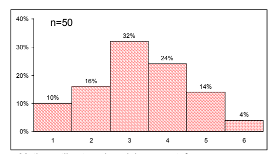
- Angabe, wie oft ein bestimmter Datenwert in der Stichprobe enthalten ist

Relative Häufigkeit

- Angabe, wie oft ein bestimmter Datenwert in der Stichprobe relativ zum Stichprobenumfang enthalten ist



Mathematiknoten einer Jahrgangsstufe



Mathematiknoten einer Jahrgangsstufe

58

Zufallsvariable (Random variable)

- Variable in einer Studie, die auf einer Zufallsstichprobe basiert
  - Alter
  - systolischer Blutdruck
  - ....
  - Zielgröße in einer Studie
- Zufallsvariable unterliegt einer bestimmten Verteilung

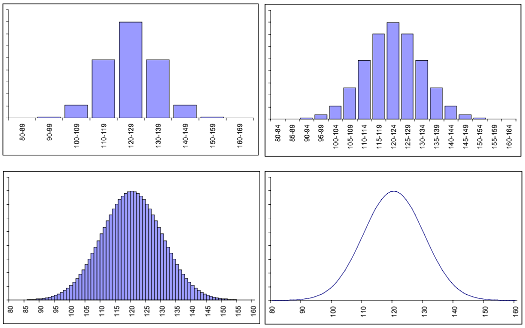
59

Skalen für Zufallsvariablen

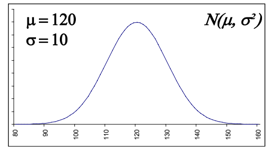
- diskret / kategorial
  - Nominalskaliert: keine lineare Ordnung
    - Beispiel: Farben, ja/nein
  - Ordinalskaliert: Ausprägung kann geordnet werden
    - Beispiel: Schulnoten
- stetig / kontinuierlich
  - intervallskaliert: Differenzen sind einheitlich interpretierbar
    - Beispiel: Temperatur in Grad Celsius
  - verhältnisskaliert: Verhältnisse sind einheitlich interpretierbar
    - Beispiel: Luftdruck, etc.

60

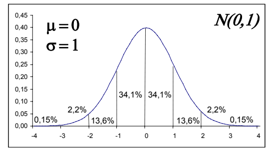
Histogramm -> Verteilung



Normalverteilung
Standard-Normalverteilung (z)



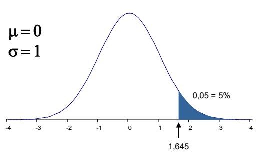
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



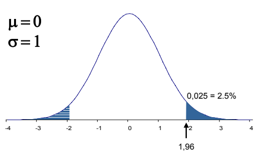
$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

$$z = \frac{x-\mu}{\sigma}$$

Z-Verteilung
„Kritische Werte“

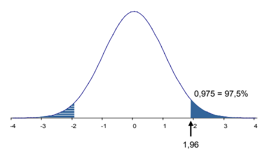


< 5% der Werte sind >1,645



< 5% der Werte sind >1,96 bzw. < -1,96

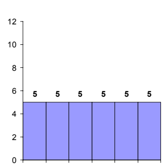
Z-Transformation
„Kritische Werte“

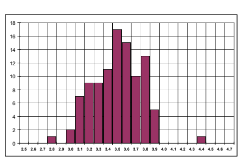


$$X = z\sigma + \mu$$

- Durch Transformation können die kritischen Werte der z-Verteilung auf jede Normalverteilung angepasst werden
  - RR in der Normalbevölkerung: μ=120, σ=10
  - krit. Grenze ($z_{97,5\%}$) = 1,96*10+120 = 139,6 mmHg

Central limit Theorem





Verteilung der Einzelwerte: Uniform

Verteilung der Stichprobenmittelwerte: Normal

„Central Limit Theorem“

- Der Mittelwert der Stichproben-Mittelwerte entspricht dem Mittelwert der Population
- Ist die Population normal verteilt, so ist auch der Mittelwert der Stichproben-Mittelwerte normal verteilt
- Ist die Population nicht normal verteilt, so ist der Mittelwert der Stichproben-Mittelwerte dennoch annähernd normal verteilt*

*für große Stichproben

Standardabweichung und Standardfehler

Standardabweichung

- SD ist die Standard-abweichung der Einzelwerte

Standardfehler

- SEM entspricht der Standardabweichung der **Mittelwerte**

$$SEM = \frac{SD}{\sqrt{n}} = \frac{\sigma}{\sqrt{n}}$$

$$SEM^2 = \frac{\sigma^2}{n}$$

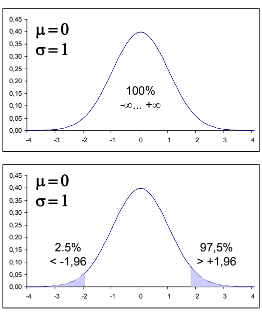
67

Konfidenzintervall / Vertrauensbereich des Mittelwerts

- Der x%-Vertrauensbereich eines Mittelwerts einer Stichprobe ($\bar{x}$) bezeichnet das Intervall, das mit x%iger Wahrscheinlichkeit den Mittelwert der Population (μ) enthält
  - Beispiel: $\bar{x}$=122 mmHg, 95%-CI [118; 124]
- 2 Konstellationen sind zu unterscheiden
  - Varianz/SD der Population ist bekannt
  - Varianz/SD der Population ist <u>un</u>bekannt

68

Vertrauenbereich für z-Verteilung N(μ,σ²) = N(0, 1)



$$X = z\sigma + \mu$$

$$[\bar{x} - z_{2,5\%} \cdot \sigma ; \bar{x} + z_{97,5\%} \cdot \sigma]$$

$$[\bar{x} - z_{2,5\%} \cdot \frac{\sigma}{\sqrt{n}} ;\ \bar{x} + z_{97,5\%} \cdot \frac{\sigma}{\sqrt{n}}]$$

$$[\bar{x} - 1{,}96 \cdot SEM ; \bar{x} + 1{,}96 \cdot SEM]$$

$$[\bar{x} - 1{,}96 \cdot \frac{\sigma}{\sqrt{n}} ; \bar{x} + 1{,}96 \cdot \frac{\sigma}{\sqrt{n}}]$$

69

Beispiel: 95%-CI bei bekannter SD der Population

- Systolischer Blutdruck der Normalpopulation (SD=10 mmHg)
- Stichprobe mit n=25 liefert einen Mittelwert von 122 mmHg

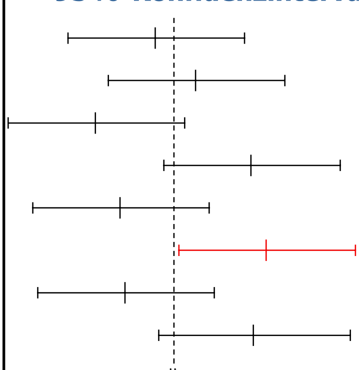
$$95\%CI = \bar{x} \pm 1{,}96 \cdot \frac{10}{\sqrt{25}}$$

$$95\%CI = 122 \pm 1{,}96 \cdot 2 = 122 \pm 3{,}92$$

$$95\%CI = [\,118{,}078 \ ; \ 125{,}92\,]$$

70

95%-Konfidenzintervall

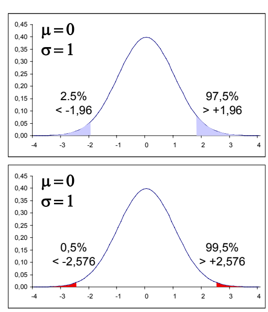


95% aller Stichproben beinhalten mit ihrem 95%-CI den Populations-mittelwert μ

Nur 5% aller Stichproben beinhalten mit ihrem 95%-Vertrauensintervall <u>nicht</u> den Populations-mittelwert μ

71

Irrtumswahrscheinlichkeit α



$\alpha = 5\%$

$z_{\alpha/2} = -1{,}96 \qquad z_{1-\alpha/2} = +1{,}96$

$\alpha = 1\%$

$z_{\alpha/2} = -2{,}576 \qquad z_{1-\alpha/2} = +2{,}576$

72

Konfidenzintervall bei bekannter SD

$$CI_{1-\alpha} = [\ \bar{x} - z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\ ;\ \bar{x} + z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\ ]$$

α = Irrtumswahrscheinlichkeit
σ = Standardabw. der Population
$\bar{x}$ = Mittelwert der Stichprobe
n = Umfang der Stichprobe

95%-Vertrauensbereich bei unbekannter SD

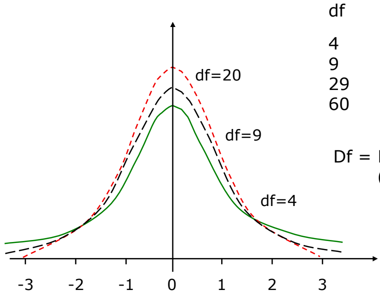
- Bei unbekanntem Populations-SD müssen anstelle von $z_{1-\alpha/2}$ die entsprechenden Werte der t-Verteilung eingesetzt werden

$$95\%CI = \bar{x} \pm \cancel{1{,}96} \cdot SEM$$

$$95\%CI = \bar{x} \pm \cancel{z_{0{,}975}} \cdot SEM$$

$$CI_{1-\alpha} = [\ \bar{x} - t_{n-1,1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\ ;\ \bar{x} + t_{n-1,1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\ ]$$

t-Verteilung (Student-t)



df	$t_{df;0,975}$	$z_{0,975}$
4	2,776	1,96
9	2,262	1,96
29	2,045	1,96
60	2,000	1,96

Df = Degree of Freedom (Freiheitsgrade)

Konfidenzintervall in der analytischen Statistik

- Klinische Studie
  - Patienten mit Grenzwerthypertonie (n=15)
  - Zielgröße: systolischer Blutdruck
  - Design: 1-armig, intraindividueller Vergleich
  - Systolischer Blutdruck vor Therapie ($RR_{t=0}$) und nach 4 Wochen ($RR_{t=28}$) kontinuierlicher Intervention
  - Fragestellung: Ist durch die Intervention eine Blutdruckänderung nachweisbar?
    - Zufallsvariable: $RR_{t=28}$ - $RR_{t=0}$

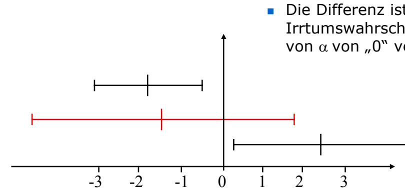
Beispiel Zufallsvariable: $RR_{t=28}$ - $RR_{t=0}$

	Vorher	Nachher	Differenz		Konfidenzintervalle	$t_{14,1-\alpha/2}$	Linke Grenze	Rechte Grenze	p
	140	136	-4		95%	2.14	-6.06	-2.75	0.05
	135	132	-3		97%	2.41	-6.26	-2.54	0.03
	141	134	-7		99%	2.98	-6.70	-2.11	0.01
	140	139	-1		99.90%	4.14	-7.59	-1.21	0.001
	140	133	-7		99.95%	4.50	-7.87	-0.93	0.0005
	135	127	-8		99.99%	5.36	-8.54	-0.27	0.0001
	141	136	-5						
	140	136	-4						
	144	146	2						
	143	137	-6						
	140	132	-8						
	138	130	-8						
	120	119	-1						
	124	118	-6						
	137	135	-2						
x	**137.20**	**132.80**	**-4.40**						
SD	**6.70**	**7.22**	**2.99**						
SEM	**1.73**	**1.86**	**0.77**						

- Da das 95%-Konfidenzintervall nicht die „0" umfasst, ist die Behandlungsdifferenz von „0" verschieden
- Simplifiziert: Es liegt ein signifikanter Behandlungseffekt mit Irrtumswahrscheinlichkeit von α = 0,05 vor

Konfidenzintervall für Differenzen

- Beinhaltet ein 1-α Konfidenzintervall für eine Differenz die „0", so kann keine „signifikante Differenz" angenommen werden.
- Ist die „0" nicht im 1-α Konfidenzintervall für eine Differenz enthalten, so kann von einem signifikanten Unterschied ausgegangen werden
  - Die Differenz ist mit einer Irrtumswahrscheinlichkeit von α von „0" verschieden



Statistischer Test
Hypothesen

- Einfluss der Intervention
  - H_0: hat keinen Einfluss
  - H_1: hat einen Einfluss
- Bezogen auf gemessene Differenz der Stichprobe
  - H_0: Differenz ist nicht „0" verschieden
  - H_1: Differenz ist von „0" verschieden

79

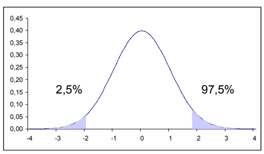
Aufbau der Hypothesen

- Die Null-Hypothese (H_0) geht von keinem systematischen Unterschied aus. Gefundene Unterschiede sind zufällig und nicht systematisch
- Die Alternativ-Hypothese (H_1 / H_A) ist die logische Umkehrung der Null-Hypothese, d.h. es existiert ein systematischer Unterschied. Gefundene Unterschiede sind nicht zufällig, sondern systematisch
- Null- und Alternativ-Hypothesen müssen sich gegenseitig ausschließen und alle Möglichkeiten abdecken.
  - Wenn H_0 falsch ist, muss H_1 wahr sein
  - Wenn H_0 wahr ist, muss H_1 falsch sein

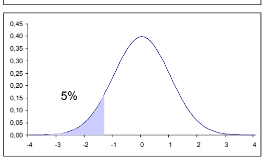
80

Ein- und zweiseitige Fragestellung

- Ungerichteter Effekt
  - H_0: $RR_{t=28} - RR_{t=0} = 0$
  - H_1: $RR_{t=28} - RR_{t=0} \neq 0$
  - Zweiseitiger Test



- Gerichteter Effekt
  - H_0: $RR_{t=28} - RR_{t=0} = 0$
  - H_1: $RR_{t=28} - RR_{t=0} < 0$
  - Einseitiger Test



81

Testergebnis und Wirklichkeit
Statistische Fehler

- 4 Möglichkeiten, wie Testergebnis und Wirklichkeit zusammentreffen können
  - H_0 wird akzeptiert, H_0 ist in Wirklichkeit wahr
  - H_0 wird akzeptiert, H_1 ist in Wirklichkeit wahr
  - H_0 wird abgelehnt, H_1 ist in Wirklichkeit wahr
  - H_0 wird abgelehnt, H_0 ist in Wirklichkeit wahr

82

Statische Fehler
Fehler I. Art und II. Art

Testentscheidung	Wirklichkeit: Differenz<>0 (H_1 ist wahr)	Wirklichkeit: Differenz=0 (H_0 ist wahr)
Differenz<>0 (H_0 ablehnen)	Richtig positiv (Power = 1-β)	Falsch positiv (Fehler I. Art α-Fehler)
Differenz=0 (H_0 beibehalten)	Falsch negativ (Fehler II. Art β-Fehler)	Richtig negativ

83

Testergebnis und Wirklichkeit
Statistische Fehler

- α-Fehler
  - H_0 wird abgelehnt, obwohl H_0 in Wirklichkeit wahr ist
  - Ein Effekt wird angenommen, wo keiner ist
- β-Fehler
  - H_0 wird akzeptiert, obwohl H_1 in Wirklichkeit wahr ist
  - Ein vorhandener Effekt wird nicht erkannt
- Welcher Fehler ist „schlimmer" und daher eher zu vermeiden?

84

Signifikanz-Niveau

- Konsequenzen eines falsch-positiven Tests
  - uneffektive Behandlung
  - Risiko ohne Nutzen („Nihil nocere")
  - Kosten ohne Nutzen
- Fazit
  - Das Risiko eines falsch positiven Tests sollte bekannt sein und durch vorherige Festlegung eines α-Niveaus kontrolliert werden
  - Übliche Werte für α
    - 0,05 (5%), 0,01 (1%), 0,001 (0,1%) ...
  - Das Signifikanz-Niveau muss vor Testbeginn festgelegt werden

Gepaarter t-Test

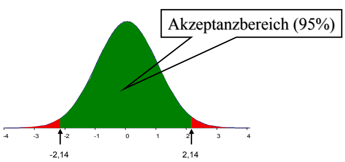
- Testet, ob eine Differenz zwischen unabhängigen Beobachtungspaaren von „0" verschieden ist
- Verteilung der Differenz entspricht einer t-Statistik der Form:

$$t = \frac{\bar{d}}{SE_{\bar{d}}}$$

  - mit n-1 Freiheitsgraden

Vorher	Nachher		Differenz
140	138		-2
135	131		-4
141	135		-6
140	136		-4
140	134		-6
135	136		1
141	138		-3
140	134		-6
144	140		-4
143	141		-2
140	142		2
138	140		2
120	121		1
124	117		-7
137	131		-6
		d	**-2,93**
		SD_d	**3,09**
		SE_d	**0,80**
		t	**-3,68**
		$t_{krit; 14; 2,5\%}$	**-2,14**
		$t_{krit; 14; 97,5\%}$	**2,14**

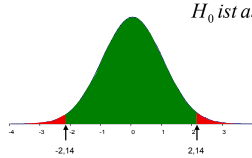
Gepaarter t-Test „Kritische Werte"



$$t = \frac{\bar{d}}{SE_{\bar{d}}}$$

- Ist der gefundene t-Wert kleiner als der untere kritische Wert oder größer als der obere kritische Wert, muss die Nullhypothese H_0 auf dem α-Signifikanzniveau abgelehnt werden
- Einfacher: Ist der Betrag des gefundenen t-Wertes größer als der positive (obere) kritische Wert, muss H_0 abgelehnt werden: $|t| > t_{krit,n-1,1-\alpha/2}$

Gepaarter t-Test Beispiel

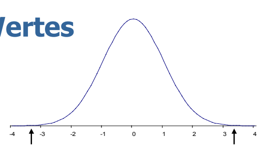


$$H_0 \text{ ist abzulehnen, wenn } |t| > t_{krit,n-1,1-\alpha/2}$$

$$t = \frac{\bar{d}}{SE_{\bar{d}}} = \frac{-2,93}{0,8} = -3,68$$

- Da |t|=3,43 größer als der kritische Wert für die t-Verteilung bei 14 Freiheitsgraden und dem 0,975-Quantil ist (2,14), muss die H_0-Hypothese auf dem Signifikanz-Niveau α=0,05 verworfen werden

Gepaarter t-Test Bedeutung des p-Wertes



α	1-α/2	$t_{krit,14,1-\alpha/2}$
0,05	0,9750	2,14
0,02	0,9900	2,62
0,01	0,9950	2,98
0,005	0,9975	3,33
0,004	0,9980	3,44
0,003	0,9985	3,58
0,0025	0,9988	3,67
0,0024	0,9988	3,70

$$t = \frac{\bar{d}}{SE_{\bar{d}}} = \frac{-2,93}{0,8} = -3,68$$

← P-Wert

P-Wert eines statistischen Tests

Vorher	Nachher		Differenz
140	138		-2
135	131		-4
141	135		-6
140	136		-4
140	134		-6
135	136		1
141	138		-3
140	134		-6
144	140		-4
143	141		-2
140	142		2
138	140		2
120	121		1
124	117		-7
137	131		-6
		d	**-2,93**
		SD_d	**3,09**
		SE_d	**0,80**
		t	**-3,68**
		$t_{krit; 14; 97,5\%}$	**2,14**
		p	**0,0025**

- P bezeichnet die Wahrscheinlichkeit eine solche Differenz oder noch extremere wie die gefundene zu erhalten, wenn die Null-Hypothese wahr wäre
- Alternativ: Die Wahrscheinlichkeit, dass eine solche Differenz zufällig beobachtet wird (ohne das ein signifikanter Unterschied vorhanden wäre)
- Wenn p<α, muss die H_0-Hypothese abgelehnt werden

Klinische Studie
"Z99 a new compound lowering BP"

- "Z99" wurde zur Behandlung der systoloischen Hypertonie etwickelt
- Phase II Studie über 7 Tage an 50 Therapie-naiven milden Hypertonikern (130 < $RR_{sys.}$ < 160 mmHg)
- Design
  - Randomisiert
  - Doppel-blind
  - Placebo-kontrolliert
  - 2-armige Parallelgruppenstudie (1:1)

Clinical Trial Example
Hypotheses

H_0: Eine 7-tägige Behandlung mit Z99 beeinflusst den systolischen Blutdruck im Vergleich zu Placebo nicht

$$x_{Z99} = x_{PBO}$$

H_1: Eine 7-tägige Behandlung mit Z99 beeinflusst den systolischen Blutdruck im Vergleich zu Placebo

$$x_{Z99} \neq x_{PBO}$$

Wenn H_0 wahr ist, muss H_1 falsch sein

U N D

Wenn H_0 falsch ist, muss H_1 wahr sein

Klinische Studie
Statistischer Plan

- Voraussetzung
  - Beide Behandlungsgruppen weisen bedingt durch vorherige Randomisierung vergleichbare Ausgangswerte auf
- Statistischer Test
  - Vergleich der beiden Gruppenmittelwerte nach 7 Tagen Behandlung mittels ***t*-test** für unabhängige Stichproben
- Signifikanz-Niveau wird auf α = 0,05 gesetzt

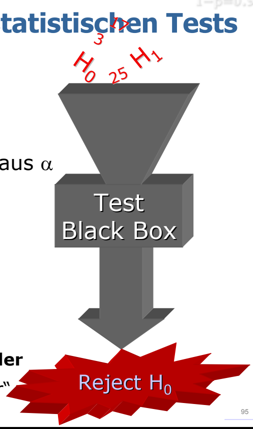
Klinische Studie
Ergebnisse

- n = 2 x 25 Patienten
- Ausgangswerte
  - x_{PBO}: 142 ± 15 mmHg (MW ± SD)
  - x_{Z99}: 142 ± 16 mmHg
- Nach 7 Tagen
  - x_{PBO}: 142 ± 15 mmHg
  - x_{Z99}: 129 ± 17 mmHg
  - t-test: p = 0.0078

	PBO	Z99
	150	120
	160	130
	145	110
	133	133
	166	115
	120	140
	157	157
	158	120
	120	100
	120	155
	145	145
	132	132
	122	122
	145	145
	120	120
	143	150
	120	110
	140	100
	150	110
	145	130
	148	148
	171	130
	151	151
	140	130
	145	130
Mean	142	129
SD	15	17
p	0,0078	

Durchführung eines statistischen Tests
"Operating the Black Box"

- Festlegung von H_0 und H_1
- Wahl des Signifikanz-Niveaus α
- Testdurchführung
- In Abhängigkeit vom Testergebnis (p)
  - H_0 ablehnen: H_1 ist wahr **oder**
  - H_0 beibehalten: H_0 ist "wahr"



Voraussetzungen für t-Test

- Intervallskalierte Daten
- Normalverteilung der Gruppen
- Varianzhomogenität der Gruppen
  - kann verletzt werden, wenn n1=n2
  - wenn n1<>n2 und Varianzhomogenität nicht gegeben, spezielle Anpassung der Freiheitsgrade möglich

Test auf Normalverteilung

- Verfahren in SPSS (explorative Datenanalyse)
- Kolmogorov-Smirnov Test
  - H_0: Stichprobe ist normalverteilt
  - H_1: Stichprobe ist nicht normalverteilt
- Shapiro-Wilk Test
  - H_0: Stichprobe ist normalverteilt
  - H_1: Stichprobe ist nicht normalverteilt

Test auf Varianzhomogenität

- Verfahren in SPSS (t-Test für unverbundene Stichproben)
- Levene`s Test (F-Test auf Varianzhomogenität)
  - H_0: $\sigma_1^2 = \sigma_2^2$
  - H_1: $\sigma_1^2 \neq \sigma_2^2$
  - Wenn H_1 wahr, spezieller heteroskedastischer t-Test mit Anpassung der Freiheitsgrade

Nichtparametrischer Test: 2 unabhängige Stichproben

- Mann-Whitney U-Test
  - aka Wilcoxon Rank-Sum Test
  - aka Mann-Whitney-Wilcoxon Rank-Sum Test
- Bildet aus den Werten Ränge und berechnet modifizierte t-Statistik für die Ränge (robuster gegen Ausreißer)
- Trennschärfer als t-Test, wenn Voraussetzungen für t-Test verletzt sind

Nichtparametrischer Test: 2 verbundene Stichproben

- Wilcoxon signed-ranks
  - Sortiert Differenzen nach absolutem Betrag und bildet entsprechende Ränge
  - Modifizierte t-Statistik für Ränge

2-Stichproben-Tests

	Parametrisch	Nicht-parametrisch
Verbundene Daten (gepaart)	Gepaarter t-Test	Wilcoxon signed-ranks Test
unverbundene Daten	t-Test für unverbundene Daten	Mann-Whitney U Test

Einfluss der Fallzahl
"Weniger ist mehr?"

- Gleiche Studie aber nur die ersten n = 2 x 13 Patienten werden ausgewertet
- Ausgangswerte
  - x_{PBO}: 142 ± 15 mmHg
  - x_{Z99}: 142 ± 16 mmHg
- Ergebnis nach 7 Tagen Behandlung
  - x_{PBO}: 141 ± 17 mmHg
  - x_{Z99}: 129 ± 17 mmHg
  - t-test: $p = 0.0987$
    - da $p > \alpha$ (0.05) kann H_0 nicht verworfen werden
    - "Z99" hat keinen Einfluss auf den systolischen Blutdruck

	PBO	Z99
	150	120
	160	130
	145	110
	133	133
	166	115
	120	140
	157	157
	158	120
	120	100
	120	155
	145	145
	132	132
	122	122
Mean	141	129
SD	17	17
p	0,0987	

Einfluss der Fallzahl

- Eine zu geringe Fallzahl kann falsch negative Ergebnisse bewirken (Fehler II. Art/β-Fehler)
- Experimente müssen die notwendige statistische Power aufweisen, um signifikante Ergebnisse liefern zu können
- Fazit: Beim Design eines Experiments ist eine Fallzahlabschätzung notwendig!

103

β Fehler und statistische Power

- **β Fehler**
  - Definition: Wahrscheinlichkeit H_0 nicht zu verwerfen, obwohl H_0 falsch ist
  - z.B.: Obwohl $\mu_{PBO} \neq \mu_{Z99}$ liefert der Test $x_{PBO} = x_{Z99}$ **(falsch negatives Ergebnis)**
- **Statistische Power (1-β)**
  - Definition: Wahrscheinlichkeit H_0 zu verwerfen, wenn H_0 falsch ist, d.h. die Wahrscheinlichkeit eine "reale" Differenz auch als solche zu entdecken
  - Vereinfacht: Wahrscheinlichkeit ein signifikantes Testergebnis zu erhalten (wenn ein signifikanter Unterschied besteht)

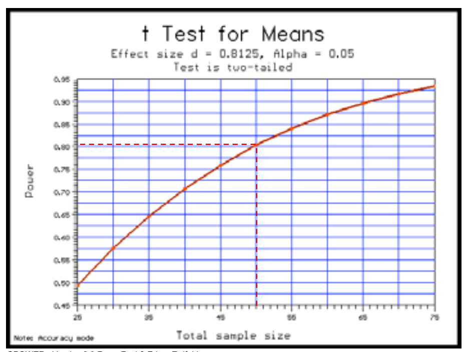
104

Vermeidung von β Fehlern: Power-Schätzung/Berechnung

- Vergleich der beiden "Z99"-Experimente
  - 1. Experiment: n = 2x25 ⇒ Power ~ 80%
  - 2. Experiment: n = 2x13 ⇒ Power ~ 38%
- Power-Schätzung
  - Wenn die stat. Power eines Studiendesigns nur 50% beträgt, wird jede 2. Studie mit diesen Parametern keine signifikanten Unterschiede anzeigen
  - Konfirmatorische Studien: Power ≥ 80%
  - Große Phase III Studien: 85-95%

105

Power & Fallzahl

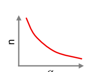


GPOWER - Version 2.0 Franz Faul & Edgar Erdfelder

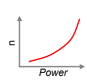
106

Faktoren, die die Fallzahl beeinflussen

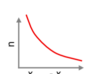
- Signifikanz-Niveau (α)
  - Je niedriger das angestrebte α, um so höher die erforderliche Fallzahl



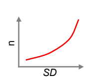
- Power (1-β)
  - Je größer die gewünschte Power, um so höher die erforderliche Fallzahl



- Geschätzte Differenz
  - Je kleiner die nachzuweisende Differenz, um so höher die erforderliche Fallzahl



- Geschätzte Standardabweichung
  - Je größer die Standardabweichung, um so höher die erforderliche Fallzahl



107

Fallzahlberechnung

- 1. Festlegung von α und gewünschter Power
  - z.B. α = 0.05 (5%), power = 80%
- 2. Schätzung der nachzuweisenden Differenz
  - Ist die Schätzung klinisch relevant?
- 3. Schätzung der erwarteten Varianz/Standardabweichung
  - Möglichst realistische Werte aus vorangegangenen Experimenten oder der Literatur verwenden
- 4. Fallzahlberechnung durchführen (oder durchführen lassen!)
  - Ist die geschätzte Fallzahl klinisch realisierbar?
  - Ist die geschätzte Fallzahl adäquat zum klinischen Problem?
  - Anpassung der Fallzahl an die geschätzte Drop-Out-Rate

108

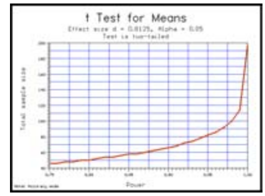
Anpassung der Fallzahlschätzung
„Drop out" Rate

- Faktoren, die die "Drop out" Rate beeinflussen
  - Studiendauer
  - Krankheitsbezogene Verschlechterung
  - Studienbedingte Unannehmlichkeiten, Adverse Events ...
- Die Fallzahlschätzung sollte immer auch die antizipierte Drop out Rate beinhalten
  - n = 50 & antizipierte "drop out" Rate 11% $\Rightarrow$ n = 56

109

Praktische Fallzahlschätzung
1. Beispiel

- α = 5%
- Power = 80%
- Geschätzte Differenz & SD
  - x_{PBO} - x_{Z99} ~ 13 mmHg
  - SD_{pooled} ~ 16
- Fallzahlberechnung
  - 2 x n = 50
  - Antizipierte Drop out Rate: 0%
  - 25 Patienten pro Gruppe benötigt



GPOWER - Version 2.0 Franz Faul & Edgar Erdfelder

110

Power: A priori & Post-hoc

"A priori" Power

- Schätzung, basierend auf
  - geschätzte Differenz
  - geschätzte SD
  - kalkulierte Fallzahl

"Post-hoc" Power

- Berechnung, basierend auf
  - beobachteter Differenz
  - beobachteter SD
  - echter Fallzahl

"Post-hoc Power" kann größer aber auch kleiner als die "a priori Power" sein!

111

Tipps & Tricks
"Oder, warum Studien scheitern?"

- Frühzeitige Einbindung des Statistikers in die Studienplanung
- Verwendung realistischer Schätzer für die erwartete Differenz und Varianz/SD
- Strikte Protokolleinhaltung
- Exakte Messung
- Vermeidung von Drop outs

112

Literatur

- Bücher
  - Rossner B. Fundamentals of Biostatistics. Duxberry Press
  - Dawson-Saunders B. & Trapp R.G. Basics and Clinical Biostatistics. Prentice Hall International Inc.
  - Motulsky, H. Intuitive Biostatistics, Oxford University Press
- Software
  - SPSS - www.spss.com
  - SAS - www.sas.com
  - Buchner A., Faul F., Erdfelder E. GPOWER 2.0 - Computer program for power- and sample size calculation, http://www.psycho.uni-duesseldorf.de/aap/projects/gpower/ (Freeware) [MS-DOS/Windows and Macintosh]

113