



Institut für Biometrie und
Klinische Forschung

Vorlesung

Biometrie für Zahnmediziner

Vorlesungsinhalt

- Deskriptive Statistik
 - Darstellung 1es Merkmals
 - Zusammenhang 2er Merkmale
- Statistisches Testen
 - Prinzipien
 - Spezielle Teste
 - Multivariat
- Föderierung / Adaptive Designs
- Multiple Regression / logistische Regression
- Überlebenszeiten

Literatur



**Weiß, C.: Basiswissen Medizinische Statistik.
6. Auflage, Springer Berlin 2013.**



**Schumacher, M., Schulgen, G.: Methodik
klinischer Studien. 3. Auflage, Springer Berlin
2008.**

Literatur

EN DE Home Produkte Hilfe Kontakt Portal

Thieme DMW

Zeitschrift Autoren Abonnement Service DMW - DGIM - Gesellschaftsausgaben

nicht eingeloggt Login

Volltextsuche >

Universität und Landesbibliothek der Universität Zürichbibliothek Medizin (ZBMed)

DMW Deutsche Medizinische Wochenschrift

DOI: 10.1055/s-00000011

Ausgabe S 01 · Volume 132 · 2007 DOI: 10.1055/s-002-6752

Jahr
2007

Ausgabe

51/52: Weihnachtsheft
50: 2677-2714
49: 2619-2654 | p26
48: 2557-2594 | p25
47: 2495-2534
46: Hypertonie
45: 2367-2396
44: 2311-2344

> aktuelle Ausgabe
> Probeausgabe (01/2016)

Statistik | Statistics

e1 Lange, S.; Bender, R.: Median oder Mittelwert? Median or mean? [Volltext](#) [HTML](#) [PDF \(85 kb\)](#)

e3 Lange, S.; Bender, R.: Quantile, empirische Verteilungsfunktion und Box Plot Quantiles, cumulative distribution function, and box-plot [Volltext](#) [HTML](#) [PDF \(169 kb\)](#)

e5 Lange, S.; Bender, R.: Variabilitätsmaße Measures of variability [Volltext](#) [HTML](#) [PDF \(82 kb\)](#)

e7 Lange, S.; Bender, R.: Das Histogramm The Histogram [Volltext](#) [HTML](#) [PDF \(137 kb\)](#)

e9 Lange, S.; Bender, R.: Lineare Regression und Korrelation Linear regression and correlation [Volltext](#) [HTML](#) [PDF \(173 kb\)](#)

DMW WALTER SIEGENTHALER PREIS

> Weggefährten – Platz vergangenen Vergangenheitsrichten

Deutsches Ärzteblatt: Serie zur Bewertung wissenschaftlicher Publikationen

<https://www.aerzteblatt.de/dae-plus/serie/35/Bewertung-wissenschaftlicher-Publikationen>

Raphael Koch / Robert Kwiecien / Jeanette Köppe

Biometrie in der medizinischen Forschung?

- Woher kommt unser medizinischen Wissen?
- Wie kann ich den Vorteil einer neuen Therapie beweisen?
- Halbwertszeit d. Erlernten ist klein, wird kleiner
- Mehr als 500.000 Fachartikel pro Jahr
- Es gibt viel zu tun: Krebs, AIDS, Parodontitis, MS, ...
- Sprache der wissenschaftlichen Medizin = Biometrie

Docetaxel plus Prednison oder Mitoxantron plus Prednison bei Prostatakarzinom im fortgeschrittenen Stadium

Docetaxel plus Prednisone or Mitoxantrone plus Prednisone for Advanced Prostate Cancer

Ian F. Tannock, and others

• **Hintergrund** Mitoxantron plus Prednison wirkt schmerzlindernd und verbessert die Lebensqualität bei Männern mit fortgeschrittenem hormonrefraktärem Prostatakarzinom, aber es verbessert nicht das Überleben. Wir verglichen eine derartige Behandlung mit Docetaxel plus Prednison bei Männern mit dieser Erkrankung.

• **Methoden** Von März 2000 bis einschließlich Juni 2002 erhielten 1006 Männer mit metastasierendem hormonrefraktärem Prostatakarzinom 5 mg Prednison zweimal täglich und wurden auf die Gabe von 12 mg Mitoxantron/m² Körperoberfläche in dreiwöchigem Abstand, 75 mg Docetaxel/m² in dreiwöchigem Abstand bzw.

30 mg Docetaxel/m² wöchentlich über 5 von jeweils 6 Wochen randomisiert. Primärer Endpunkt war das Gesamtüberleben. Sekundäre Endpunkte waren Schmerzen, die PSA-Spiegel (prostataspezifisches Antigen) und die Lebens-

jeweils 6 Wochen randomisiert. Primärer Endpunkt war das Gesamtüberleben. Sekundäre Endpunkte waren Schmerzen, die PSA-Spiegel

Was ist „randomisiert“?
Warum wird randomisiert?

Was ist ein primärer Endpunkt?
Warum betrachtet man primäre und sekundäre Endpunkte?

•**Ergebnisse** Im Vergleich zu den Männern der Mitoxantron-Gruppe ergab sich für die Männer der Gruppe, die Docetaxel im Abstand von 3 Wochen erhalten hatte, eine Hazard Ratio für Tod von 0,76 (95%- Konfidenzintervall 0,62 bis 0,94; $p=0,009$ im stratifizierten Log-Rank-Test);

Hazard-Ratio
Konfidenzintervall
 $p=0,009$
Log-Rank-Test

Biometrie in der medizinischen Forschung!

- **Aktiv:**
 - Promotion
 - Habilitation
 - Wissenschaftler
- **Passiv:**
 - Prüfarzt
 - Fachzeitschriften
 - Pharmareferent
 - Patienten – Laienartikel

Beispiel: NB 2004 Studie

<http://www.kinderkrebsinfo.de/fachinformationen/studienportal/pohkinderkrebsinfotherapiestudien/nb2004>

 **kinderkrebsinfo.de**
Informationsportal zu Krebs- und Bluterkrankungen bei Kindern und Jugendlichen

Home | Über uns | Erkrankungen | Patienten | **Fachinformationen** | Aktuelles | Services | GPOH |

Home / Fachinformationen / Studienportal / Studien und Regist... / NB2004

NB2004

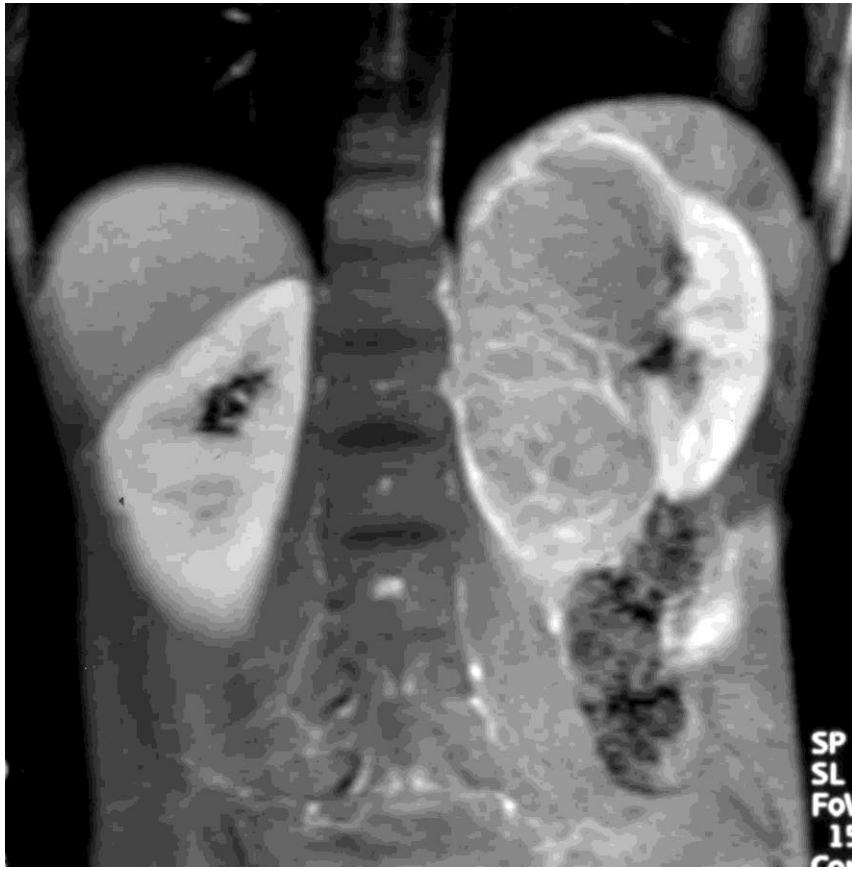
Titel	NB2004 Trial Protocol for Risk Adapted Treatment of Children with Neuroblastoma
Erkrankung	Neuroblastom
Art	Kooperative, multizentrische Therapieoptimierungsstudie für die Behandlung von Säuglingen, Kindern und Jugendlichen mit Neuroblastom
Fragestellung / Ziel	<p>Beobachtungsgruppe -die Zahl der Patienten in der Beobachtungsgruppe, die postoperativ ohne Chemotherapie auskommen, durch Erweiterung der Einschlusskriterien zu erhöhen -tumorbedingte Symptome oder Progress durch eine milde Chemotherapie zu behandeln</p> <p>Mittlere Risikogruppe -eine Verbesserung des Event-free-survivals durch eine Intensivierung der Therapie im Vergleich zur Vorstudie</p> <p>Hochrisikogruppe Risikopatienten werden in der Studie B'NB 2004 HR behandelt.</p>
Therapie / Studienarme	<p>Patienten mit einem sicher diagnostiziertem Neuroblastom werden in eine von drei Risikogruppen stratifiziert.</p> <p>Beobachtungsgruppe: in diese Gruppe kommen alle Patienten mit Stadium 1 ohne MYCN-Amplifikation oder Patienten mit Stadium 2 im Alter von 0-21 Jahren oder Patienten mit Stadium 3 im Alter von 0-2 Jahren oder Patienten mit Stadium 4S im Alter von 0-1 Jahr, bei denen weder eine MYCN-Amplifikation noch eine 1p Aberration nachgewiesen wurde. Die Größe des Resttumors bei Patienten mit Stadium 2 spielt keine Rolle mehr. Diese Gruppe erhält keine Chemotherapie. Säuglinge werden bis zum Ende des 2. Lebensjahrs entsprechend den Protokollvorgaben beobachtet, Kinder, die zum Diagnosezeitpunkt älter als ein Jahr waren, werden für 12 Monate nach Diagnosestellung beobachtet. Die Patienten sollen alle sechs Wochen zu Kontrolluntersuchungen einbestellt werden. Danach werden die Abstände zwischen den Kontrollen länger. Im Falle eines Rezidivs oder eines Tumorprogresses sollte für das weitere Vorgehen mit der Studienzentrale Kontakt aufgenommen werden. In der Regel wird eine milde Chemotherapie verabreicht.</p> <p>Medium-Risk-Gruppe: in diese Gruppe kommen alle Patienten mit Stadium 2 im Alter von 0-21 Jahren mit einer 1p oder 1p Aberration und ohne MYCN-Amplifikation oder Patienten mit Stadium 3 im Alter >= 2-21 Jahren ohne 1p Aberration oder MYCN-Amplifikation oder</p>

NB 2004 – Neuroblastomstudie bei Kindern

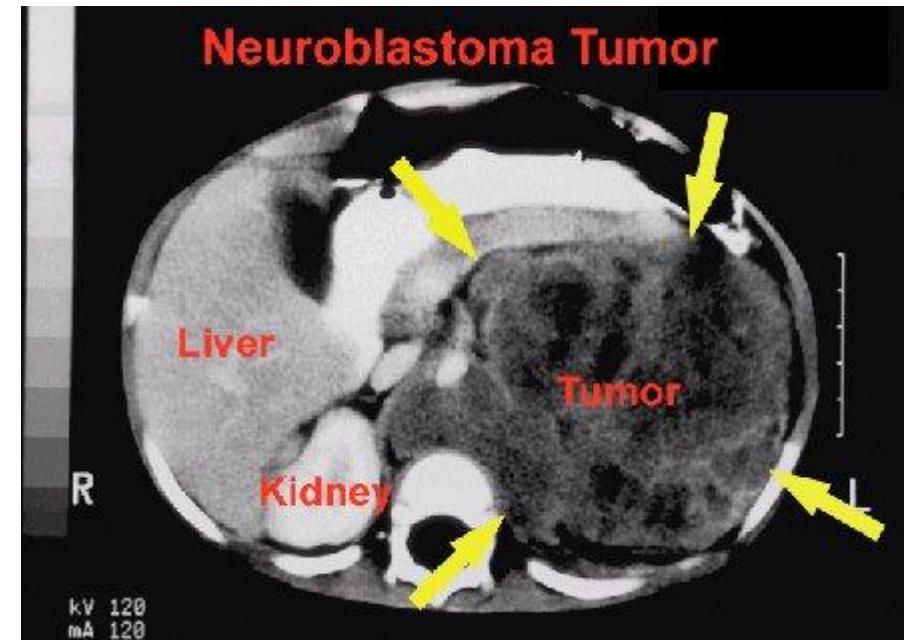
- Maligner Tumor des sympathischen Nervensystems
- Häufigster solider Tumor im Kindesalter (7,3% der Malignome)
1 Fall / 100.000 Kinder unter 15 Jahren pro Jahr, ca. 130 jährlich
- **Symptome:** Schmerz, Fieber, tastbarer Bauchtumor, Diarröhö
- **Prognose:** vor 25 Jahren nur 48% der Kinder
heute 75% der Kinder nach 5 Jahren am Leben
- **NB 2004:** Deutschlandweite Therapieoptimierungsstudie
- **Studienleitung:** Prof. Berthold, Prof. Simon, PD Hero (Köln)
- Rasche Metastasierung aber auch Spontanheilungen möglich

Beispiel: NB 2004 Studie

Linker Grenzstrang



Tumor Nebenniere



Studienidee – NB 2004

Risikostratifizierte Therapie nach

- Tumorstadium: 1, 2, 4s, 3, 4
- Alter: < 1, 1 – 2, 2 – 21 Jahre
- Laktatdehydrogenase (LDH): normal vs. erhöht
- MYCN: Onkogen amplifiziert ja vs. nein, Chrom. 2
- 1p (neu): normal vs. deletiert oder imbalanciert

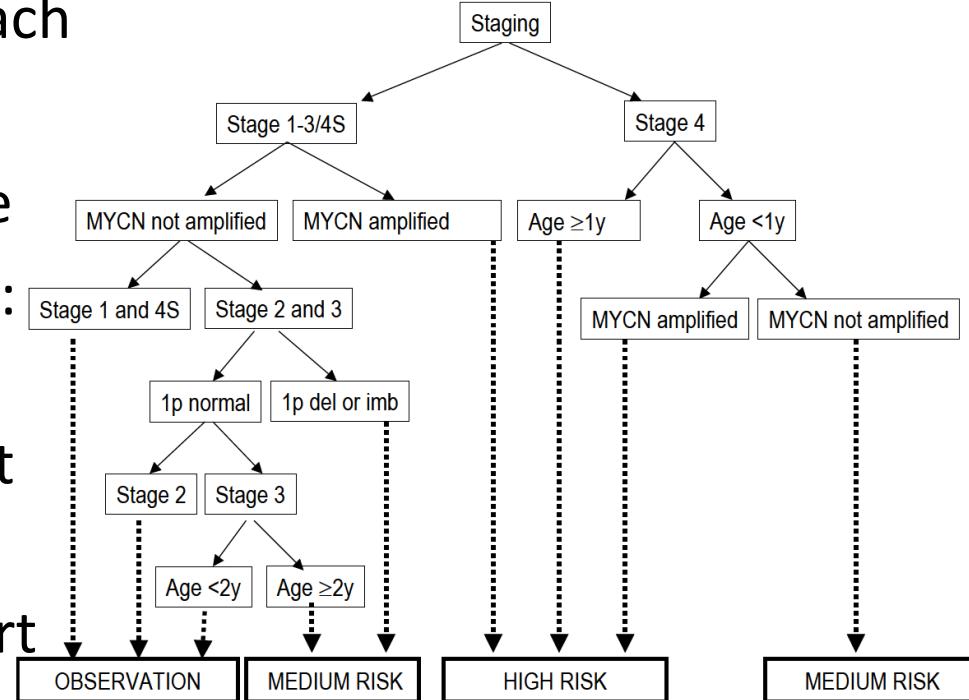
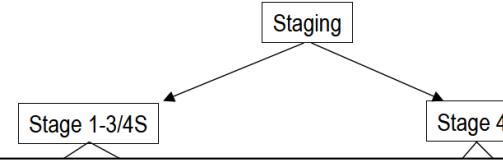


Figure 2: Risk group classification in NB2004

Studienidee – NB 2004

Risikostratifizierte Therapie nach

- **Tumorstadium:** 1, 2, 4s, 3, 4



- 1:** Lokalisierte Tumor auf Ursprungsort begrenzt
- 2:** Lokalisierte Tumor infiltriert Umgebung, ohne die Mittellinie zu überschreiten
- 3:** Lokalisierte Tumor überschreitet Mittellinie, regionale Lymphknoten können beidseits befallen sein
- 4:** Hämatogene Fernmetastasen
- 4s:** wie 1 oder 2, aber mit Fernmetastasen in Leber, Haut; nur im Säuglingsalter

- **1p (neu):** normal vs. deletiert oder imbalanciert

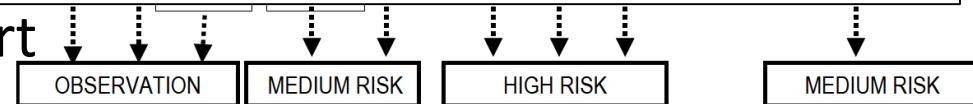


Figure 2: Risk group classification in NB2004

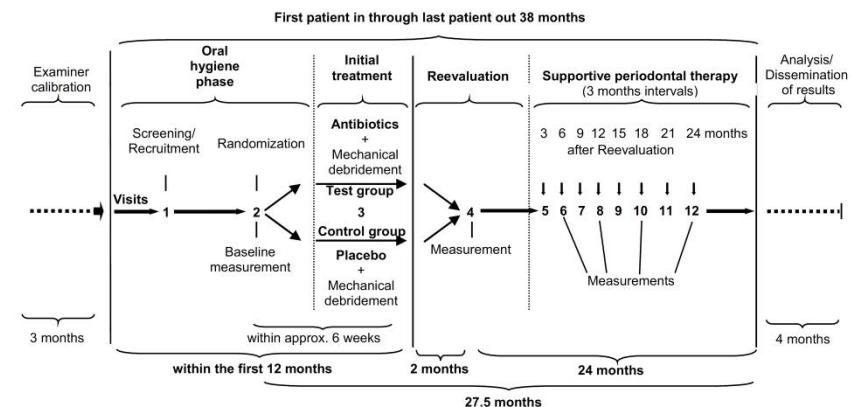
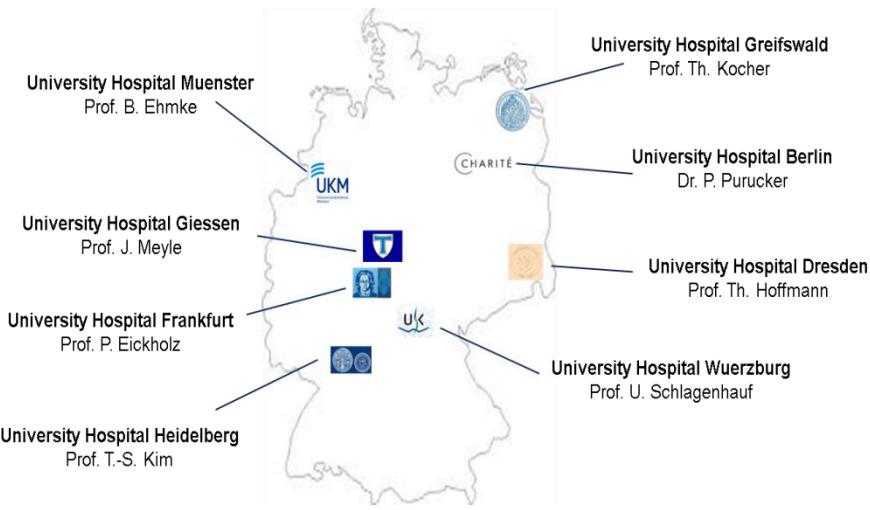
Studienidee – NB 2004

- Beobachtungsgruppe: Wait and see für erweiterte Gruppe
- Mittlere Risikogruppe: Intensivierung der Therapie
- Hochrisikogruppe: beste Chemotherapie der Vorgängerstudie
NB $97 \pm$ neuer Chemotherapiezyklus (wirksam bei Rezidiven)
- Genexpressionsprofile aller Tumore

ABParo Studie

- Adjunctive Antimicrobial Therapy of Periodontitis: Long-Term Effects on Disease Progression and Oral Microbiological Colonization
- Prospektiv, randomisiert, stratifiziert, doppelblind, multizentrisch (Clinical Trials.gov NCT00707369)
- Parallelgruppendesign: Antibiotika (Amoxicillin + Metronidazole (3x/Tag, 7 Tage) vs. Placebo)
- 506 Patienten randomisiert
- Rekrutierungsstart bis Ende Nachbeobachtung: Oktober 2008 – Januar 2012

Participating Centres



Vorlesungsinhalt

- **Deskriptive Statistik**
 - Darstellung 1es Merkmals
 - Zusammenhang 2er Merkmale
- Statistisches Testen
 - Prinzipien
 - Spezielle Tests
 - Multiples Testen
- Fallzahlplanung / Adaptive Designs
- Lineare Regression / logistische Regression
- Überlebenszeiten

Deskriptive Statistik – Darstellung 1es Merkmals

- Grundbegriffe
- kategoriale Variablen
- stetige Variablen

Deskriptive Statistik – Darstellung 1es Merkmals

- Grundbegriffe
- kategoriale Variablen
- stetige Variablen

INRG a 08-11-15d.sav [DataSet2] - IBM SPSS Statistics Daten-Editor

File Edit View Data Transform Analyze Direct Marketing Charts Options Windows Help

11 : fer_ln 3,63230910262554

	nummer	ferritin	LDH	mycn	aber_1p	alter	fer_ln	ldh_ln	inss	var
1	3213	1435	934	nein	ja	2,7	5,0	2,2	4	
2	3254	1239	524	nein	ja	3,6	4,8	1,7	4	
3	3203	730	970	nein	ja	4,7	4,3	2,3	4	
4	2789	665	1064	nein	ja	5,0	4,2	2,4	4	
5	3172	481	713	nein	ja	2,7	3,9	2,0	4	
6	3296	461	2036	ja	ja	2,5	3,8	3,0	4	
7	3120	427	1058	nein	ja	4,6	3,8	2,4	4	
8	3304	417	2069	ja	ja	1,6	3,7	3,0	4	
9	1309	404	8251	nein	ja	1,9	3,7	4,4	3	
10	3342	390	394	ja	ja	5,2	3,7	1,4	4	
11	3247	377	507	nein	ja	3,2	3,6	1,6	4	
12	3242	368	2530	ja	ja	3,0	3,6	3,2	3	
13	3312	332	1811	ja	ja	2,2	3,5	2,9	4	
14	3335	274	2851	ja	ja	3,2	3,3	3,4	4	
15	3256	265	2700	ja	ja	1,9	3,3	3,3	4	
16	3294	263	2822	ja	ja	1,7	3,3	3,3	4	
17	3273	256	1584	ja	ja	1,7	3,2	2,8	4	
18	3306	205	632	ja	ja	1,8	3,0	1,8	4	
19	3319	177	2470	ja	ja	1,6	2,9	3,2	4	
20	3157	172	.	nein	ja	2,3	2,9	.	4	
21	2838	152	4090	ja	ja	1,2	2,7	3,7	4	

Studieneinheiten

- **Beobachtungseinheit**
 - Definition: **kleinste Einheit**, die beobachtet wird
 - *NB 2004*: An einem Neuroblastom erkrankte/r **PatientIn**
 - *ABParo*: Zahnstelle, Zahn oder Patient
- **Grundgesamtheit**
 - Definition: Alle **potentiellen Beobachtungseinheiten**
 - *NB 2004*: Alle an einem **Neuroblastom** erkrankten **PatientInnen** unter 21 Jahren
 - *ABParo*: Alle **Parodontitispatienten**

Studieneinheiten

- **Stichprobe**
 - Definition: Alle **untersuchten Beobachtungseinheiten**
 - *NB 2004*: Teilnehmer der NB 2004 Studie
- **Stichprobegröße, -umfang**

Anzahl der untersuchten Beobachtungseinheiten = **Fallzahl**

Merkmale

Bekannte Risikofaktoren der NB 2004 Studie

- LDH: in U/l
- Alter: in Tagen
- International Neuroblastoma Staging System (INSS):
Stadium 1 - 4
- MYCN-Amplifikation: ja oder nein

Merkmal = Variable

- Definition: Jede an einer Beobachtungseinheit
erhobene Eigenschaft
- NB 2004: LDH, Alter, INSS, MYCN

Merkmale

Bekannte Risikofaktoren der NB 2004 Studie

- LDH: in U/l
- Alter: in Tagen
- International Neuroblastoma Staging System (INSS):
Stadium 1 - 4
- MYCN-Amplifikation: ja oder nein

Ausprägung

- Definition: Jeder mögliche Merkmalswert
- NB 2004: 350 U/l, 183 Tage, Stadium 4, MYCN-amplifiziert

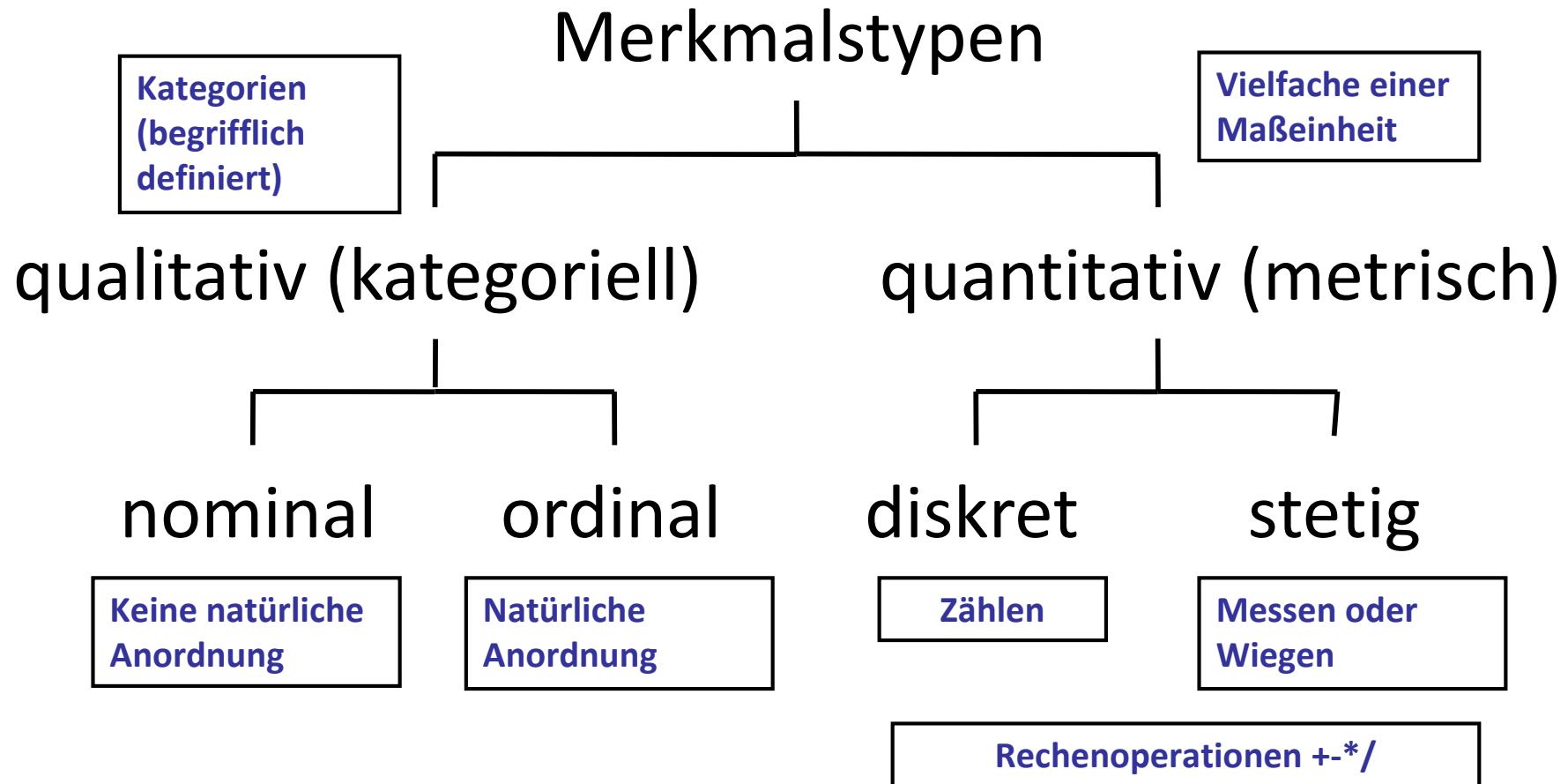
Merkmale

Bekannte Risikofaktoren der NB 2004 Studie

- LDH: in U/l
- Alter: in Tagen
- International Neuroblastoma Staging System (INSS):
Stadium 1 - 4
- MYCN-Amplifikation: ja oder nein

Merkmalstyp

- Definition: **Klasse** von **Ausprägungen**
- **Quantitativ**: stetig, diskret
- **Qualitativ = kategorial**: nominal, ordinal



Merkmale

Bekannte Risikofaktoren der NB 2004 Studie

- LDH: in U/l stetig
- Alter: in Tagen diskret
- International Neuroblastoma Staging System (INSS):
Stadium 1 - 4 ordinal
- MYCN-Amplifikation: ja oder nein nominal: binär

Merkmalstyp

- Definition: Klasse von Ausprägungen
- Quantitativ: stetig, diskret
- Qualitativ = kategorial: nominal, ordinal

Merkmalstypen

- Quantitativ: Geordnete Werte, Vielfache einer Einheit
 - Stetig: Prinzipiell sind alle Zwischenwerte beobachtbar
Beispiele: Gewicht, Größe, Blutdruck
 - Diskret: Nicht alle Zwischenwerte sind möglich, z.B. Anzahl
Beispiele: Leukozytenanzahl, Zahl der Geburten
- Kategorial: wenige Ausprägungen, ungleichmäßige Zuwächse
 - Nominal: Ausprägungen ohne Ordnung
Beispiele: Blutgruppe, Haarfarbe, Wohnort
 - Ordinal: geordnete Ausprägungen
Beispiele: Schweregrad, Schmerzstärke, Schulnoten
 - Sonderfall mit 2 Kategorien: binär = dichotom

Deskriptive Statistik – Darstellung 1es Merkmals

- Grundbegriffe
- kategoriale Variablen
- stetige Variablen

INSS-Stadium

Relative Häufigkeit:

Absolute Häufigkeit / Gesamtzahl
mit einer beobachteten
Ausprägung

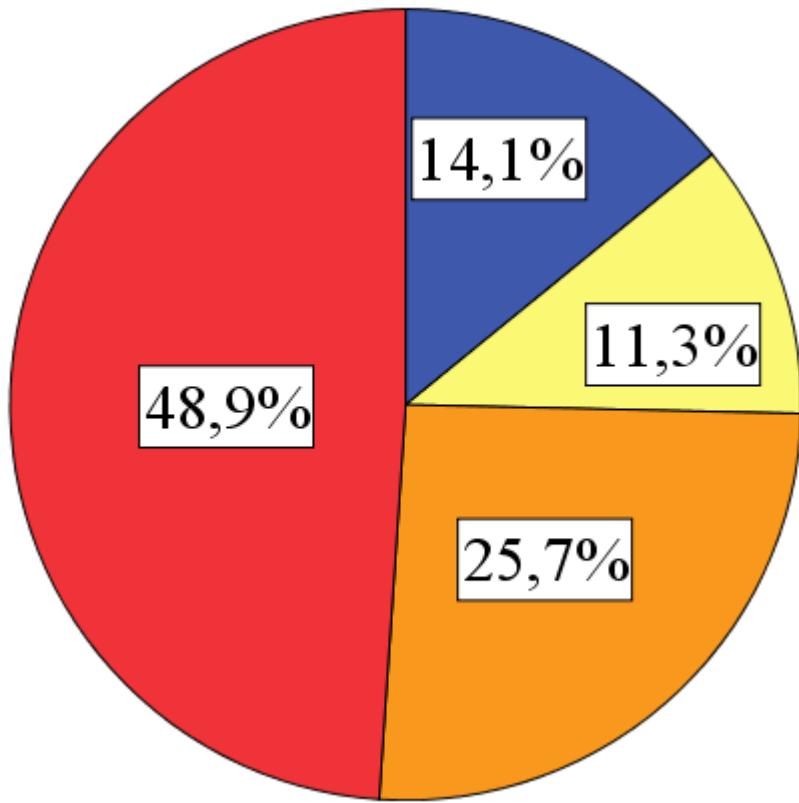
		Häufigkeit	Prozent	Gültige Prozente
Gültig	1	356	10,6	14,1
	2	284	8,5	11,3
	3	649	19,4	25,7
	4	1235	36,9	48,9
	Gesamt	2524	75,5	100,0
Fehlend	System	821	24,5	
	Gesamt	3345	100,0	

Faustregel:

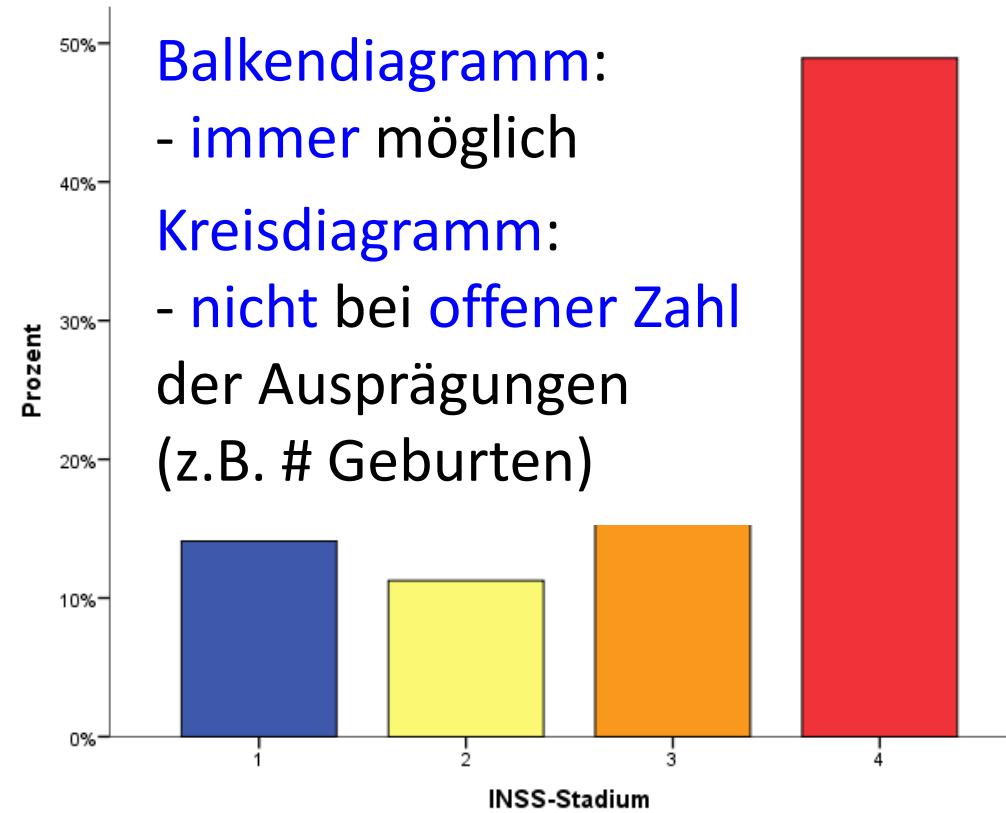
- Gesamtzahl ≤ 30 :
keine Prozentzahl
- $30 < \text{Gesamtzahl} \leq 300$:
Prozentzahl ohne
Nachkommastelle
- $300 < \text{Gesamtzahl} \leq 3000$:
Prozentzahl mit
1 Nachkommastelle
- ...
- **Fehlende Werte angeben!**

INSS-Stadium

Kreisdiagramm



Balkendiagramm



Deskriptive Statistik – Darstellung 1es Merkmals

- Grundbegriffe
- kategoriale Variablen
- stetige Variablen

Lagemaße bei stetigen Variablen

- Frage: Um welchen Wert konzentrieren sich die Daten?
- Mittelwert:
 - schätzt den Erwartungswert μ
 - (Summe aller Werte) / (Anzahl der beobachteten Werte)
 - reagiert empfindlich auf extreme Werte
- Median:
 - Zentrum der geordneten Werte $x_{(1)}, x_{(2)}, \dots, x_{(n)}$
 - $x_{((n+1)/2)}$ für ungerade n & $(x_{(n/2)}+x_{(n/2+1)})/2$ für gerade n
 - robust bei extremen Werten
- Modalwert = Modus: häufigster beobachteter Wert

Beispiel für Lagemaße bei stetigen Variablen

- **Messwerte:** 2, 10, 2, 7, 14, (85)
- **Mittelwert:**
 - mit Extremwert (85): $(2+10+2+7+14+85)/6 = 120/6 = 20$
 - ohne Extremwert (85): $(2+10+2+7+14)/5 = 35/5 = 7$
 - reagiert empfindlich auf extreme Werte
- **Median:**
 - geordneten Werte $x_{(1)}, x_{(2)}, \dots, x_{(n)}$: 2, 2, 7, 10, 14, (85)
 - mit Extremwert (85): $(7 + 10) / 2 = 8,5$
 - ohne Extremwert (85): 7
- **Modalwert = Modus:** 2 (weniger wichtig)

Lagemaße bei stetigen Variablen

- **Minimum:**
 - kleinste beobachtete Ausprägung
 - fällt monoton mit Stichprobenumfang \Rightarrow ungünstiges Maß
- **Maximum:**
 - größte beobachtete Ausprägung
 - steigt monoton mit Stichprobenumfang \Rightarrow ungünstiges Maß
- **Beispiel:** Messwerte = 2, 10, 2, 7, 14, (85)

Minimum = 2,

Maximum = 14 (85),

Lagemaße bei stetigen Variablen

1. Quartil Q1 (25% Quantil):

- mindestens 25% der Werte sind kleiner oder gleich
- mindestens 75% der Werte sind größer oder gleich

2. Quartil Q2 (50% Quantil):

- mindestens 50% der Werte sind kleiner oder gleich
- mindestens 50% der Werte sind größer oder gleich

3. Quartil Q3 (75% Quantil):

- mindestens 75% der Werte sind kleiner oder gleich
- mindestens 25% der Werte sind größer oder gleich

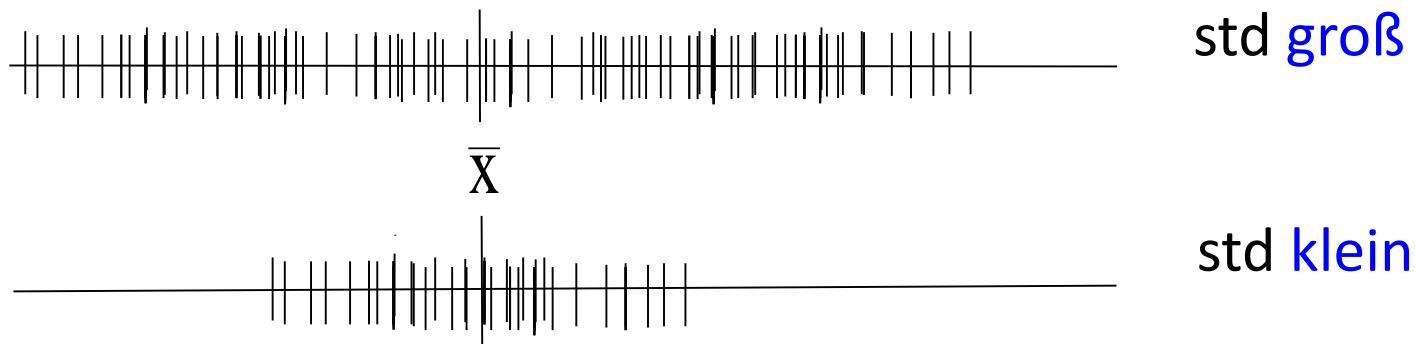
- Werte **auf den Quartilen** werden in **beide Richtungen** mitgezählt

Lagemaße bei stetigen Variablen

- Konzept der Quartile:
 - Median: teilt Beobachtungen 50% : 50% auf
 - Quartile: teilen Beobachtungen $\approx 25\% : 25\% : 25\% : 25\%$
- Messwerte: 3, 8, 4, 11, 10, 1, 7, 12, 9, 2, 5, 6
 - Geordnete Messwerte: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12
 - Median = $(6 + 7) / 2 = 6,5$
 - $Q1 = 3,5; Q2 = 6,5; Q3 = 9,5$ oder
 - $Q1 = 3; Q2 = 6,2; Q3 = 10$ oder ...
- Quartile sind nicht eindeutig aber robust bei Extremwerten

Streuungsmaße bei stetigen Variablen

- Frage: Wie weit liegen die Werte auseinander?
- Varianz (Var): Quadratische Abweichungen vom Mittelwert
- $Var = \frac{\sum_{Stichprobe} (Einzelwert - Mittelwert)^2}{Stichprobenumfang - 1} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$
- Standardabweichung (std) = \sqrt{Var}



Streuungsmaße bei stetigen Variablen

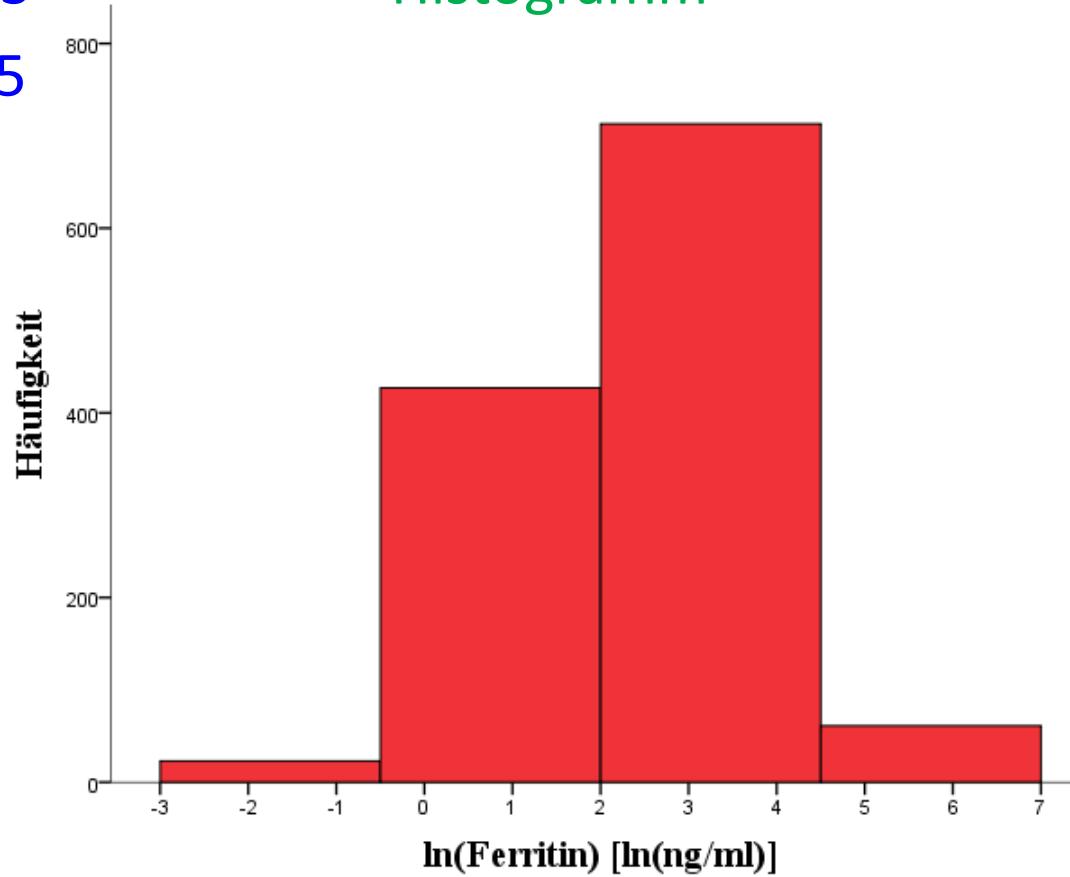
- Frage: Wie weit liegen die Werte auseinander?
- Varianz (Var): Quadratische Abweichungen vom Mittelwert
- $$Var = \frac{\sum_{Stichprobe} (Einzelwert - Mittelwert)^2}{Stichprobenumfang - 1} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$
- Standardabweichung (std) = \sqrt{Var}
- Varianz und Standardabweichung: reagieren empfindlich auf extreme Werte, da mittelwertbasiert & Mittelwert empfindlich
- Alternativ:
 - Spannweite: Maximum – Minimum
 - Interquartilsabstand: 3. Quartil – 1. Quartil

Neuroblastom - logarithmiertes Serum Ferritin (ng/ml)

Minimum: -2,3

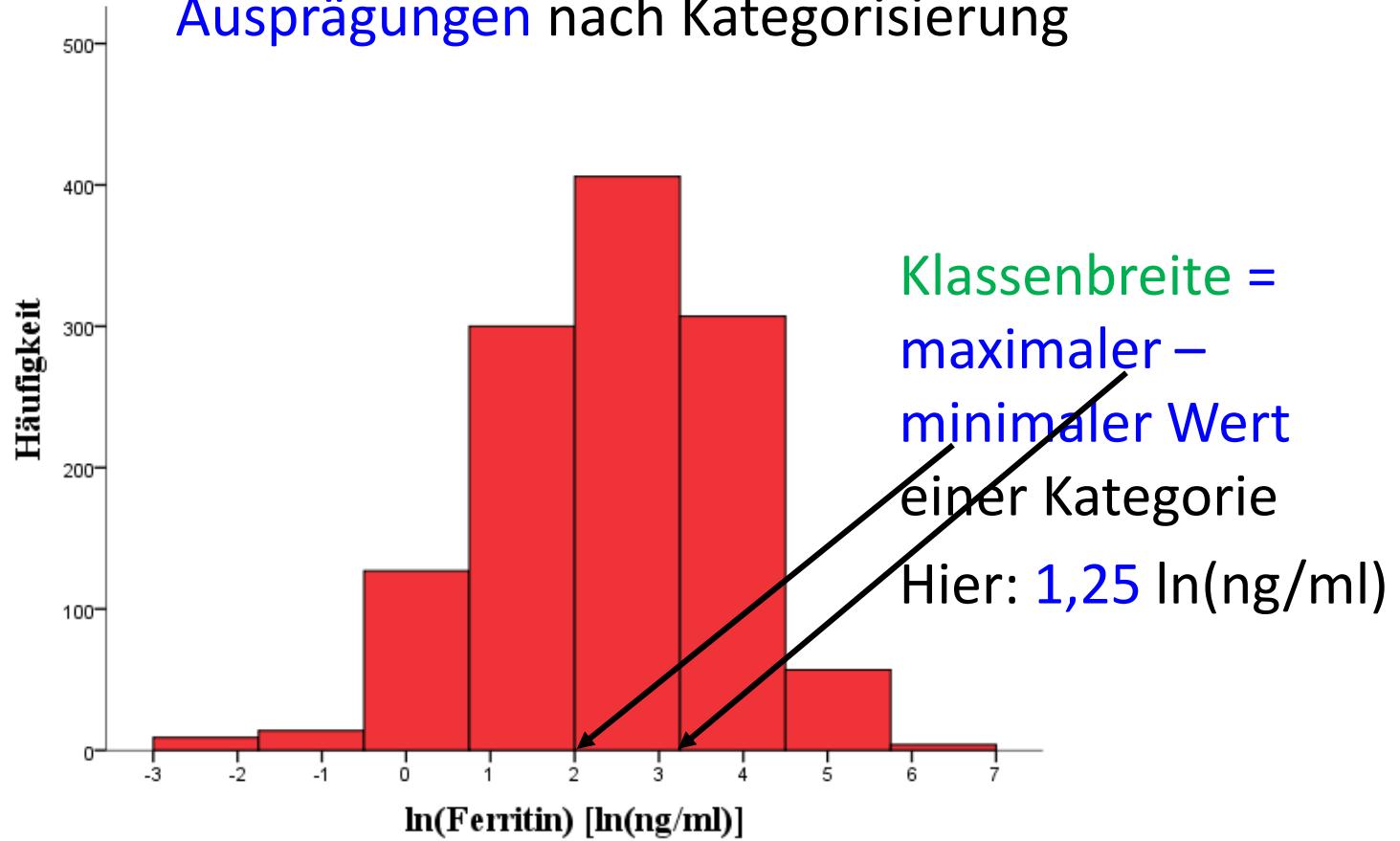
Maximum: 6,5

Histogramm



Neuroblastom - logarithmiertes Serum Ferritin (ng/ml)

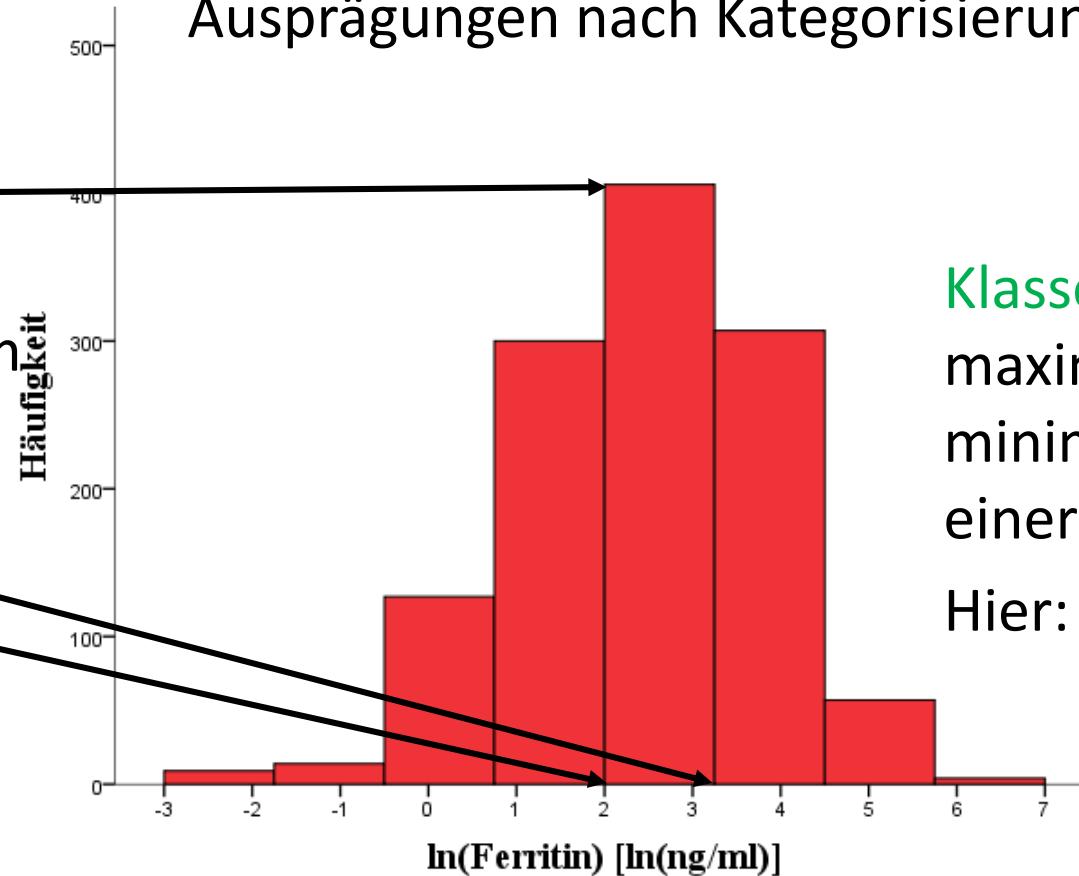
Klassenzahl = 8 = Zahl der möglichen Ausprägungen nach Kategorisierung



Neuroblastom - logarithmiertes Serum Ferritin (ng/ml)

Klassenzahl = 8 = Zahl der möglichen Ausprägungen nach Kategorisierung

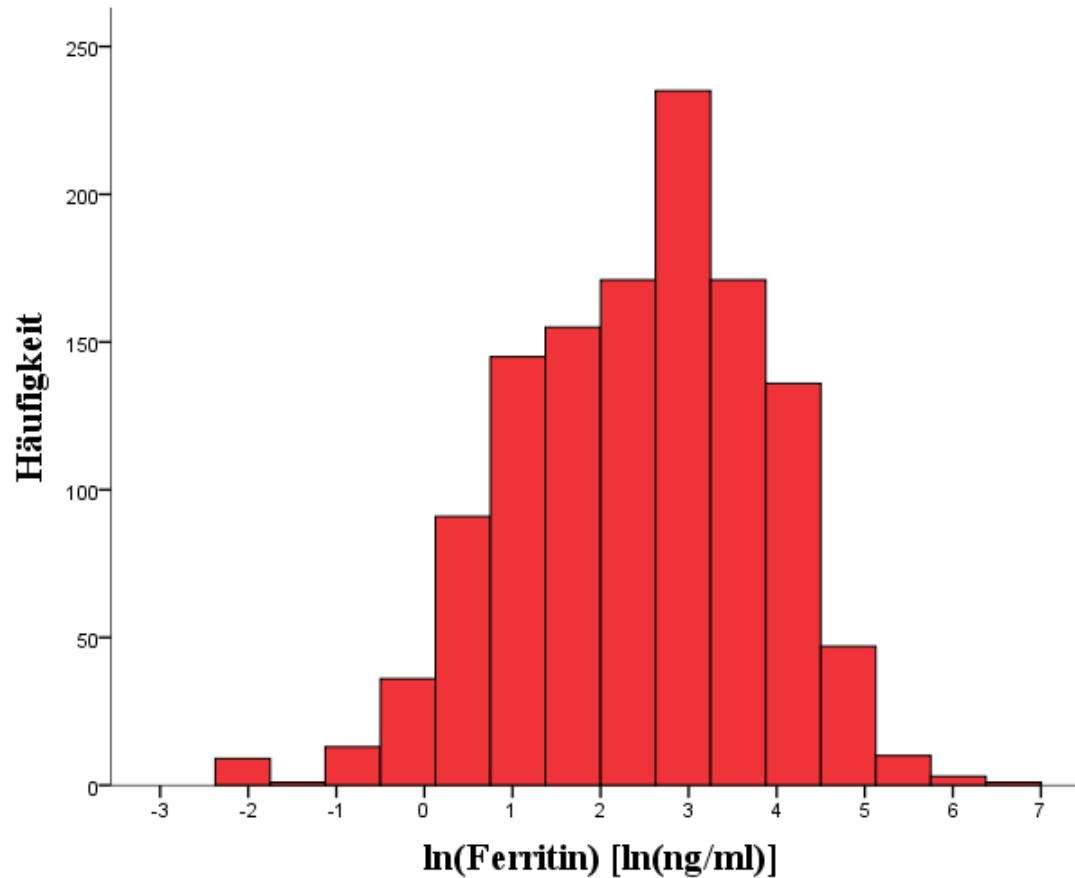
Ca. 400 Patienten mit logarithmiertem Serum Ferritin- wert zwischen 2,00 und 3,25 $\ln(\text{ng/ml})$



Klassenbreite = maximaler – minimaler Wert einer Kategorie
Hier: 1,25 $\ln(\text{ng/ml})$

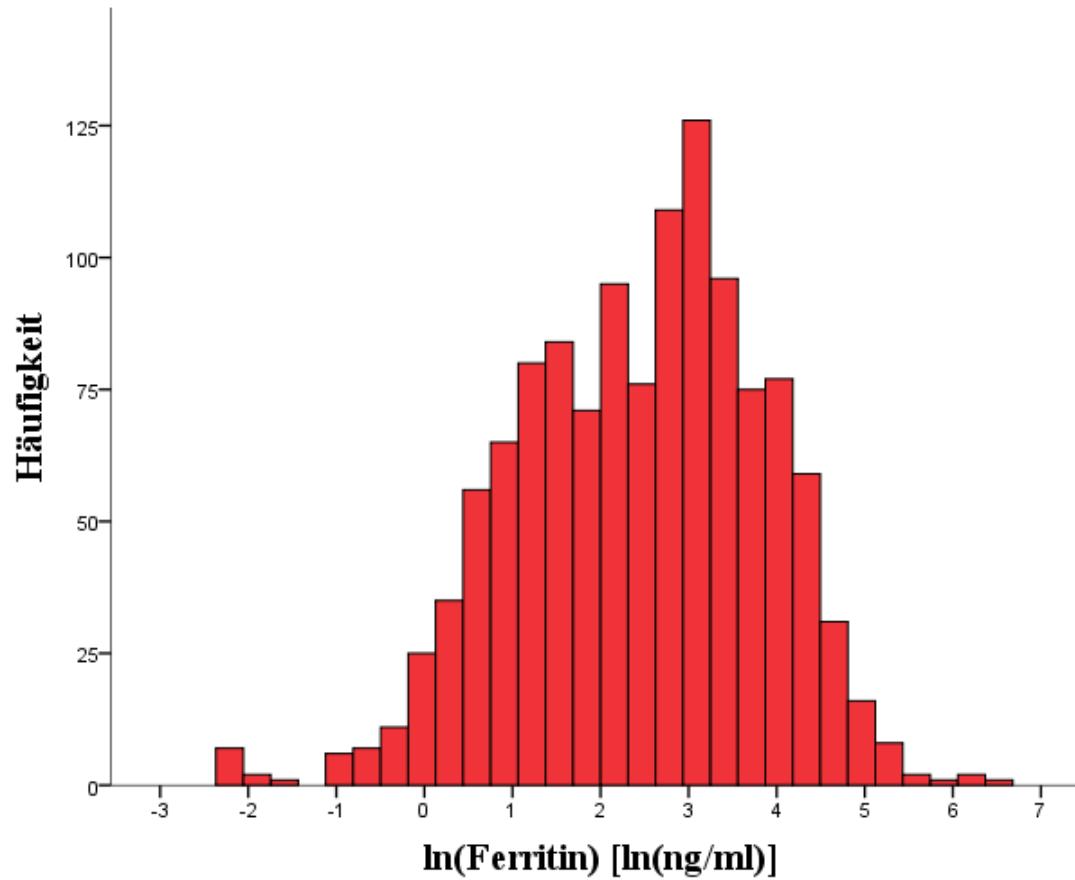
Neuroblastom - logarithmiertes Serum Ferritin (ng/ml)

Klassenzahl = 16



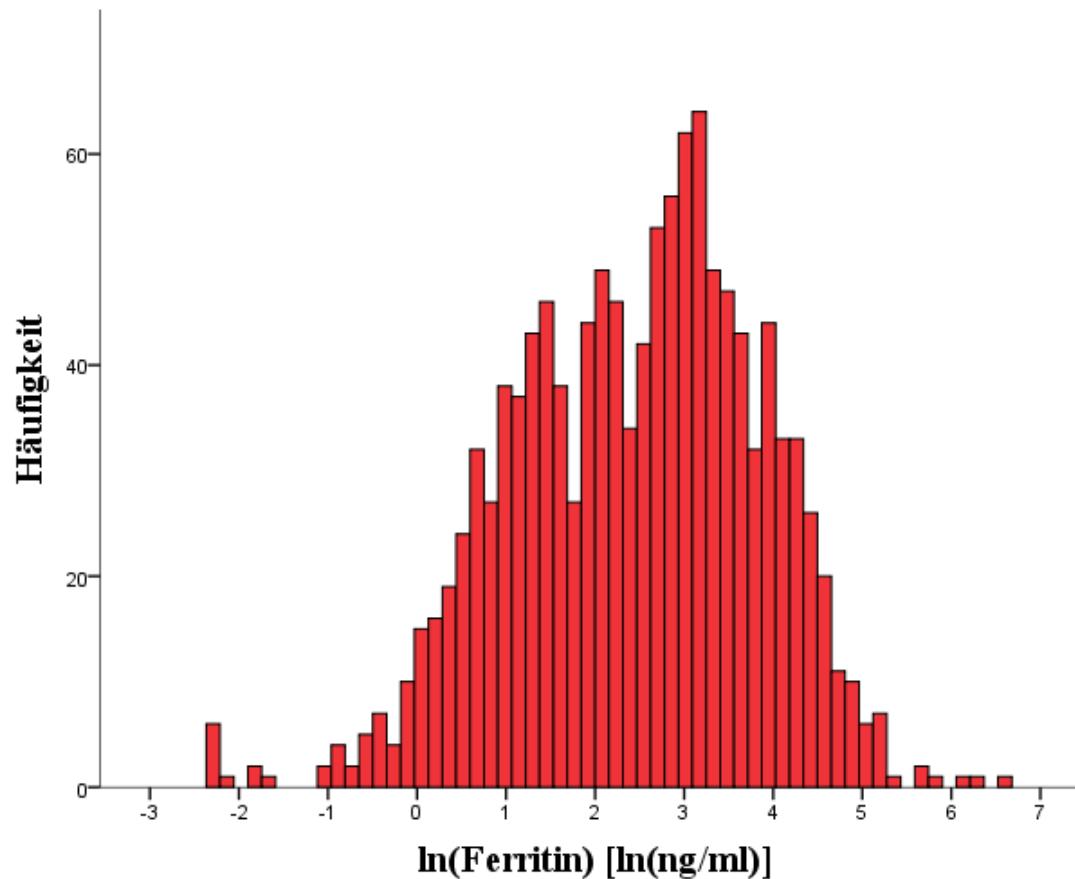
Neuroblastom - logarithmiertes Serum Ferritin (ng/ml)

Klassenzahl = 32



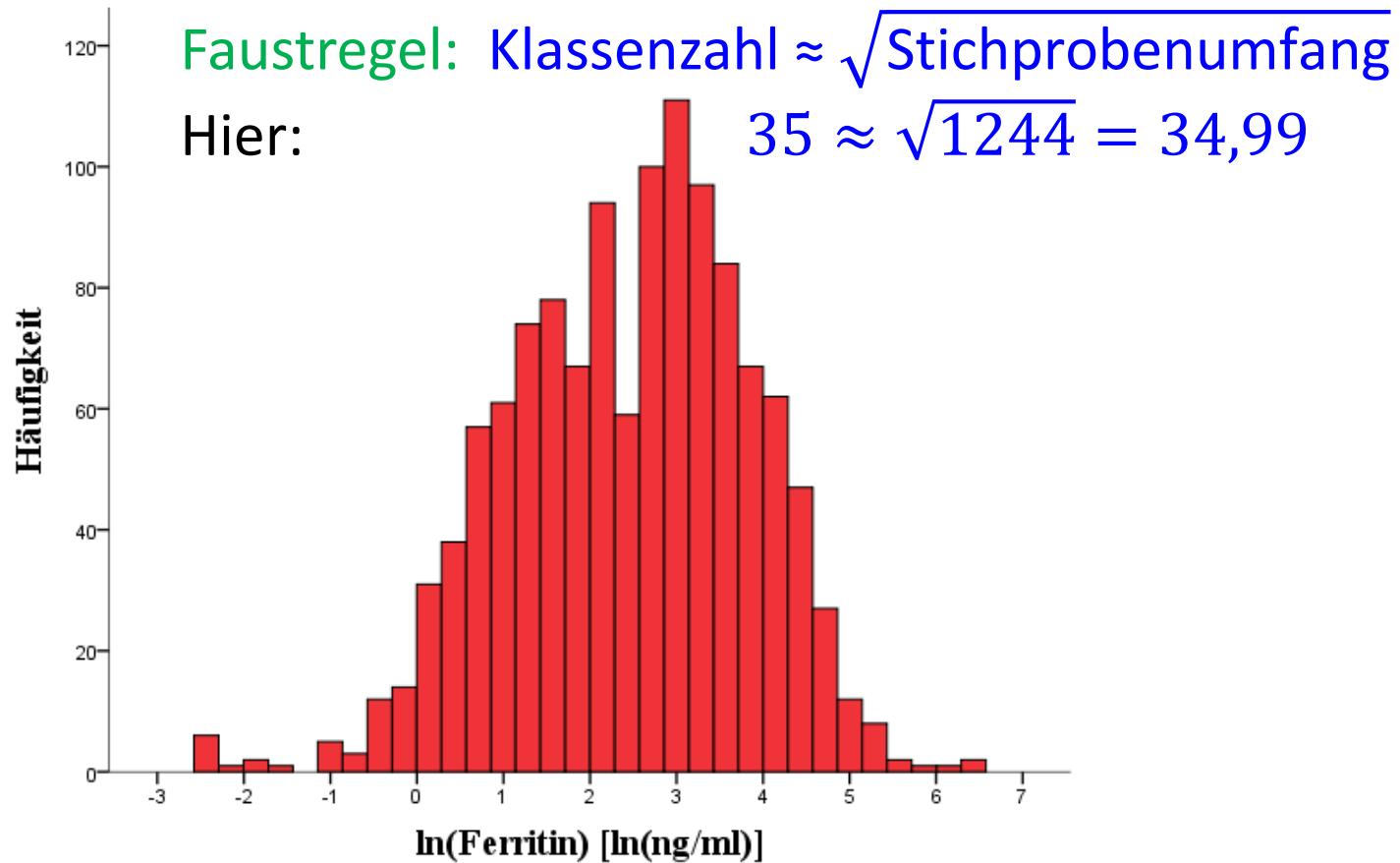
Neuroblastom - logarithmiertes Serum Ferritin (ng/ml)

Klassenzahl = 64



Neuroblastom - logarithmiertes Serum Ferritin (ng/ml)

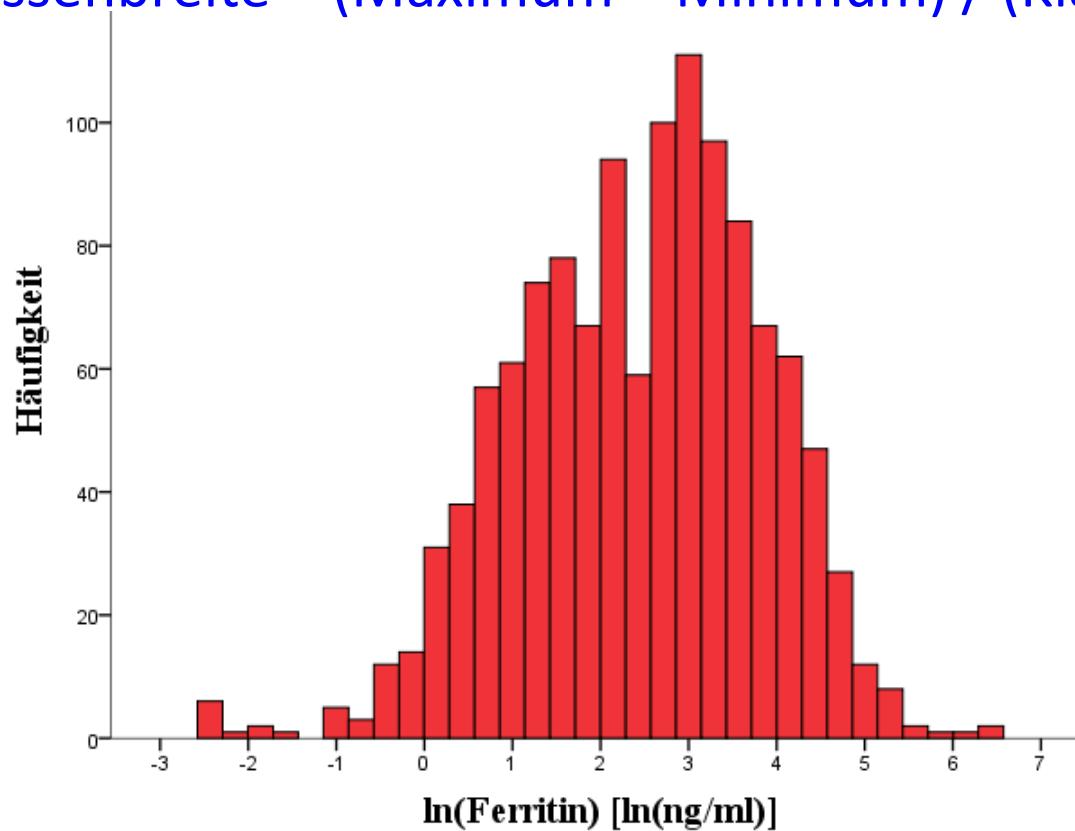
Klassenzahl = 35



Neuroblastom - logarithmiertes Serum Ferritin (ng/ml)

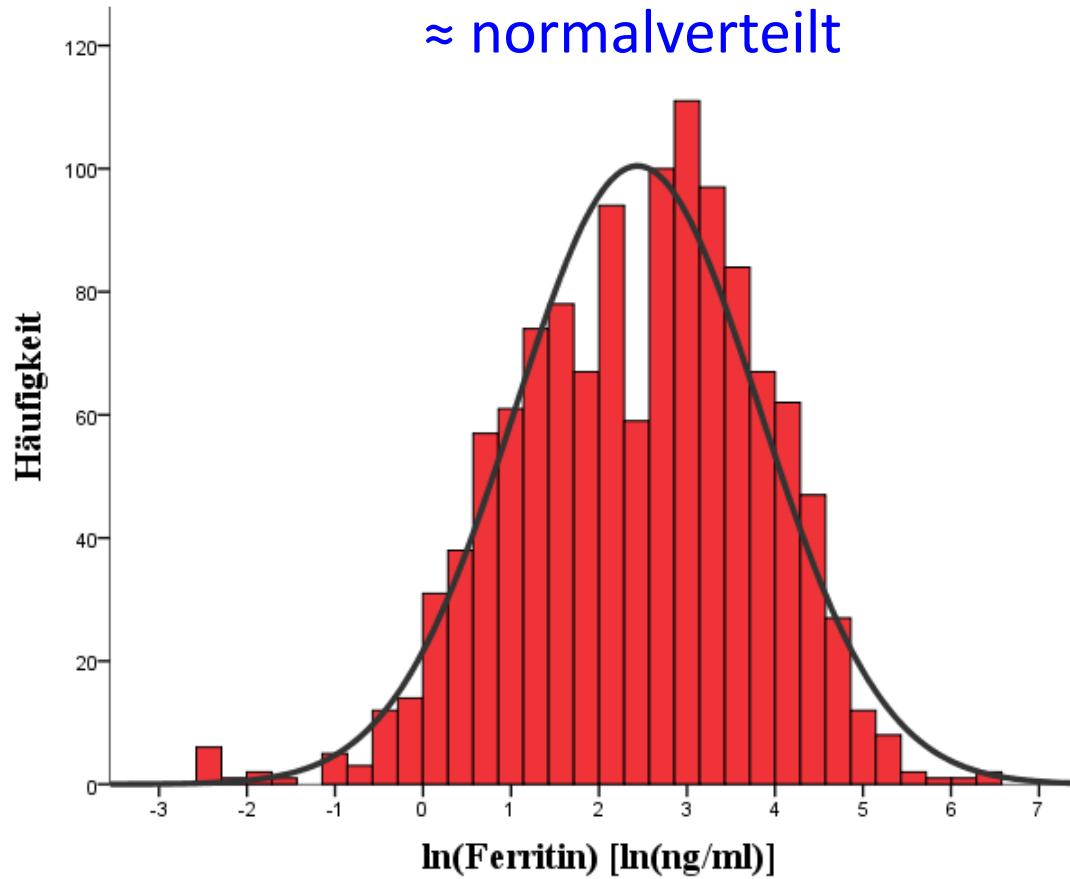
Klassenzahl = 35

Faustregel: Klassenbreite \approx (Maximum – Minimum) / (Klassenzahl - 1)



Neuroblastom - logarithmiertes Serum Ferritin (ng/ml)

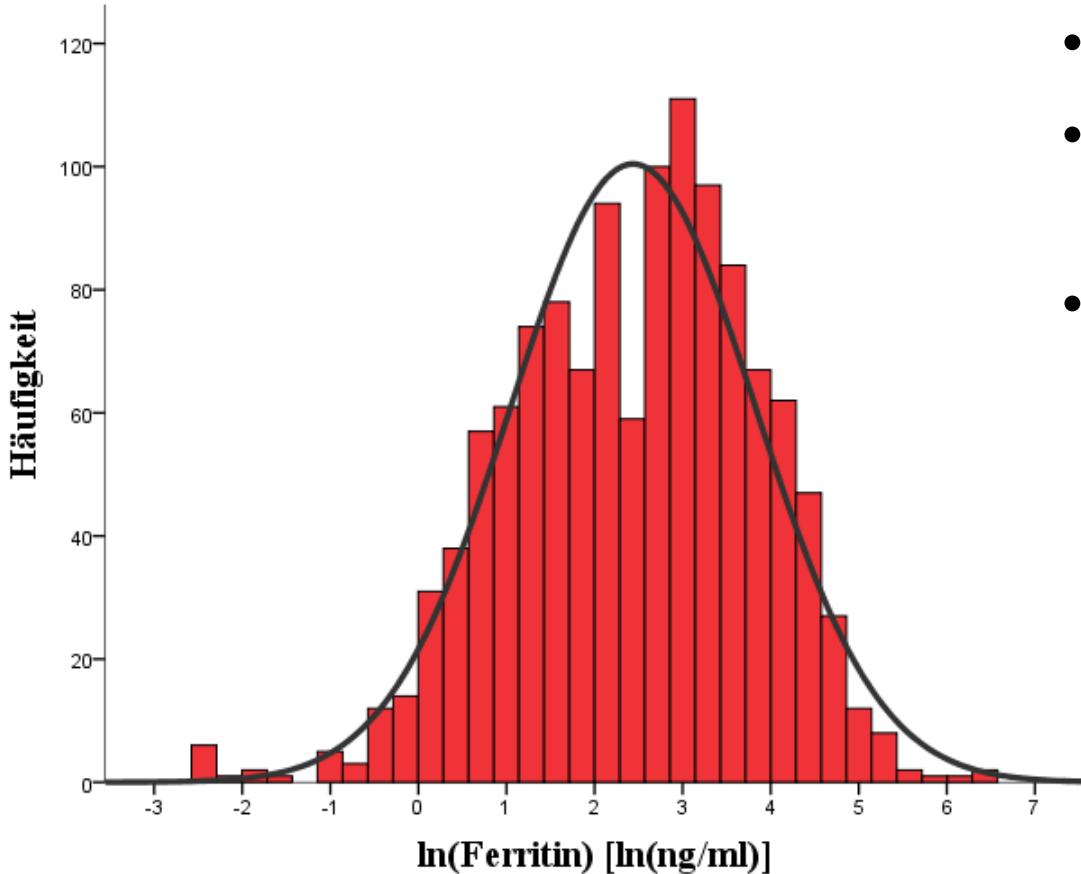
symmetrisch
≈ normalverteilt



Neuroblastom - logarithmiertes Serum Ferritin (ng/ml)

Normalverteilung:

- eingipflig, symmetrisch, u.a.
- schöne Eigenschaften bzgl. Auswertbarkeit
- bei großen Fallzahlen Methoden der Normalverteilung anwendbar



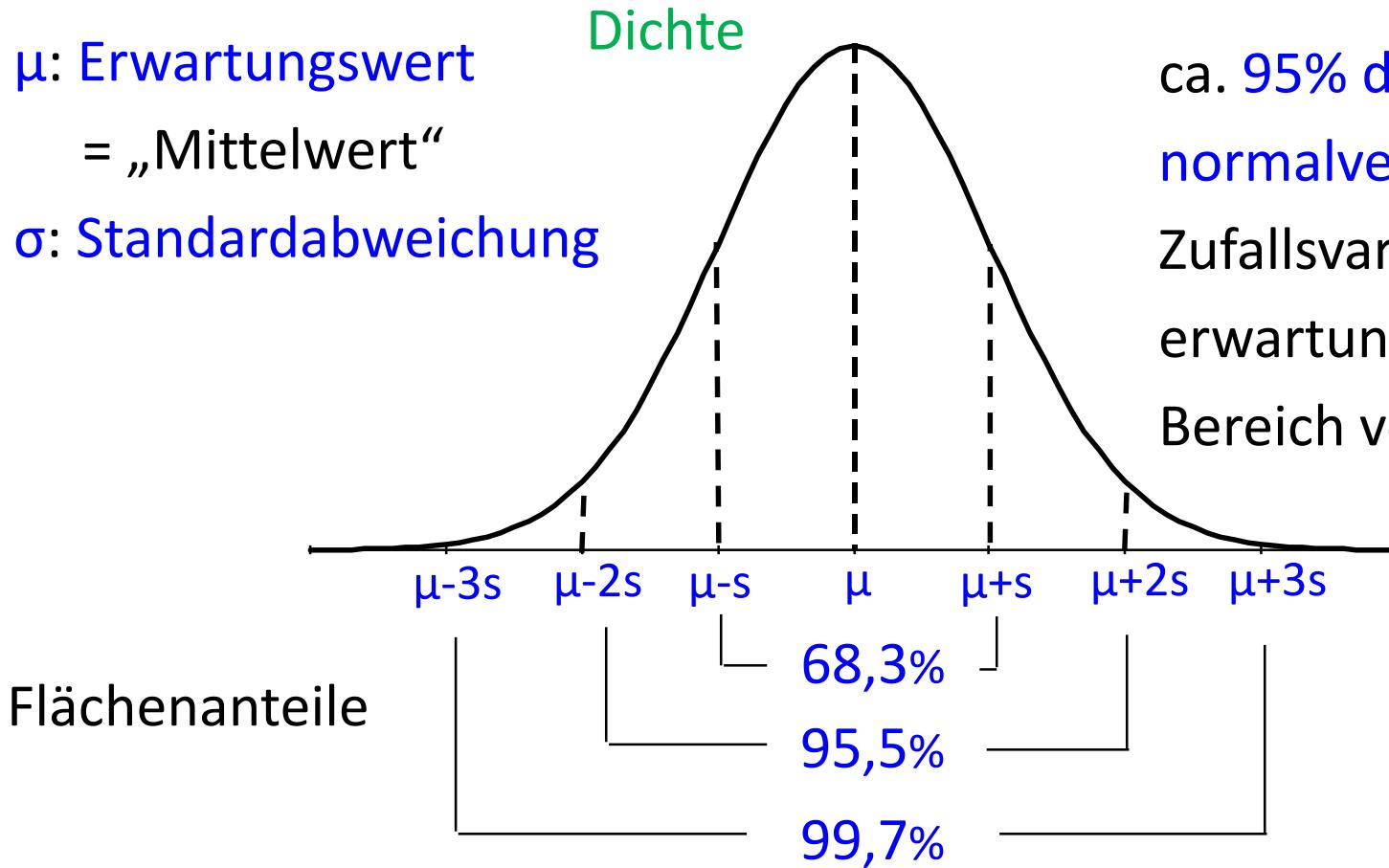
$$f(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{1}{2} \left(\frac{x-\mu}{\sigma} \right)^2}$$

Normalverteilung

μ : Erwartungswert
= „Mittelwert“

σ : Standardabweichung

ca. 95% der Werte einer normalverteilten Zufallsvariablen liegen erwartungsgemäß im Bereich von $\mu \pm 1,96 \cdot \sigma$.



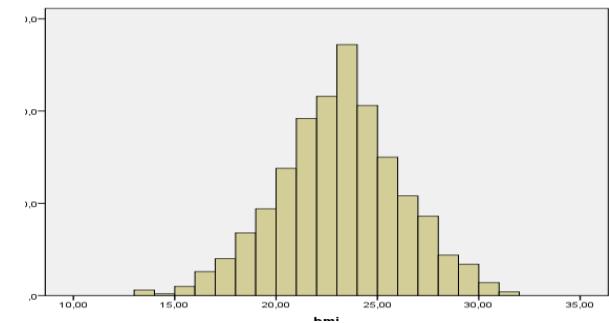
Verteilungsmaße und -formen bei stetigen Variablen

Frage: Streben die Werte in eine Richtung?

Schiefe (g):

$$\hat{g} = \frac{n}{(n-1)(n-2)} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s} \right)^3$$

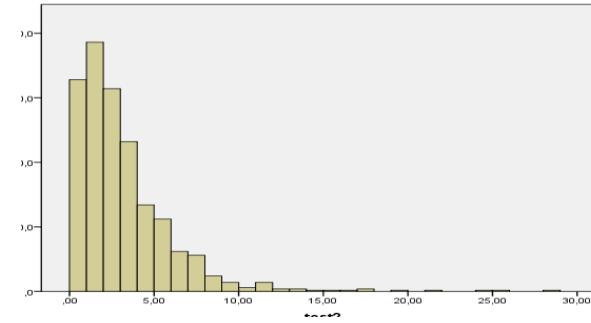
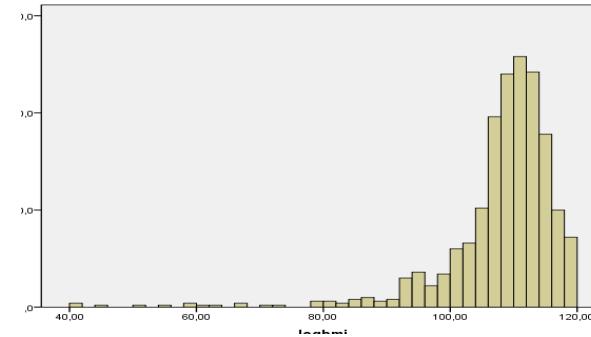
- $g = 0$: perfekt symmetrische Verteilung, z.B. Gaußverteilung = Normalverteilung
- $-1 \leq g \leq +1$: akzeptable symmetrische Verteilung, Mittelwert \approx Median



Verteilungsmaße und -formen bei stetigen Variablen

Frage: Streben die Werte in eine Richtung?

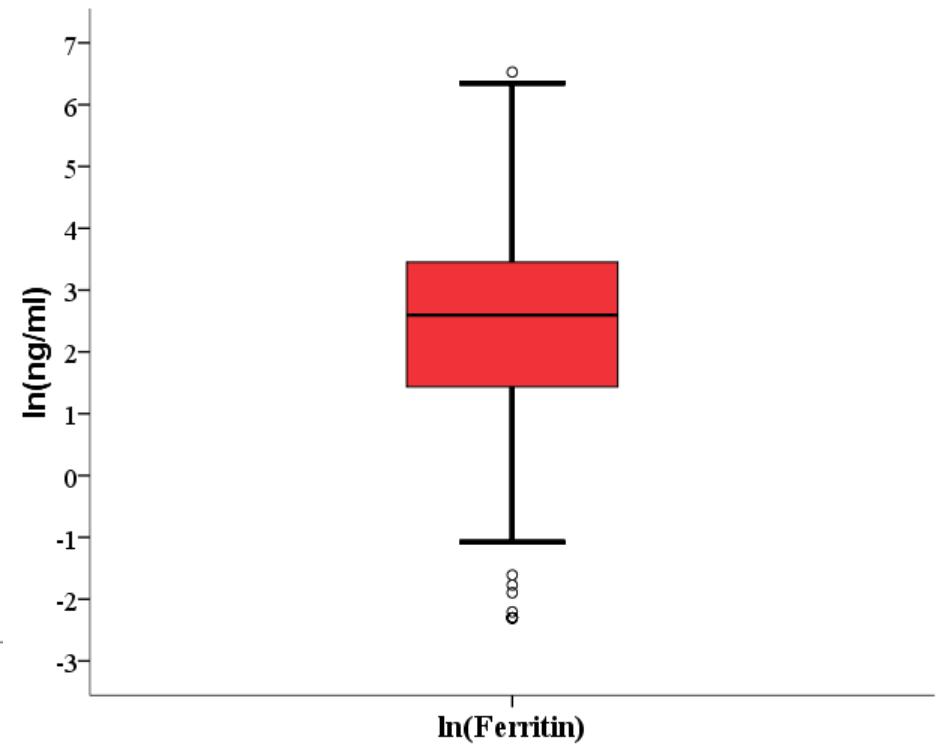
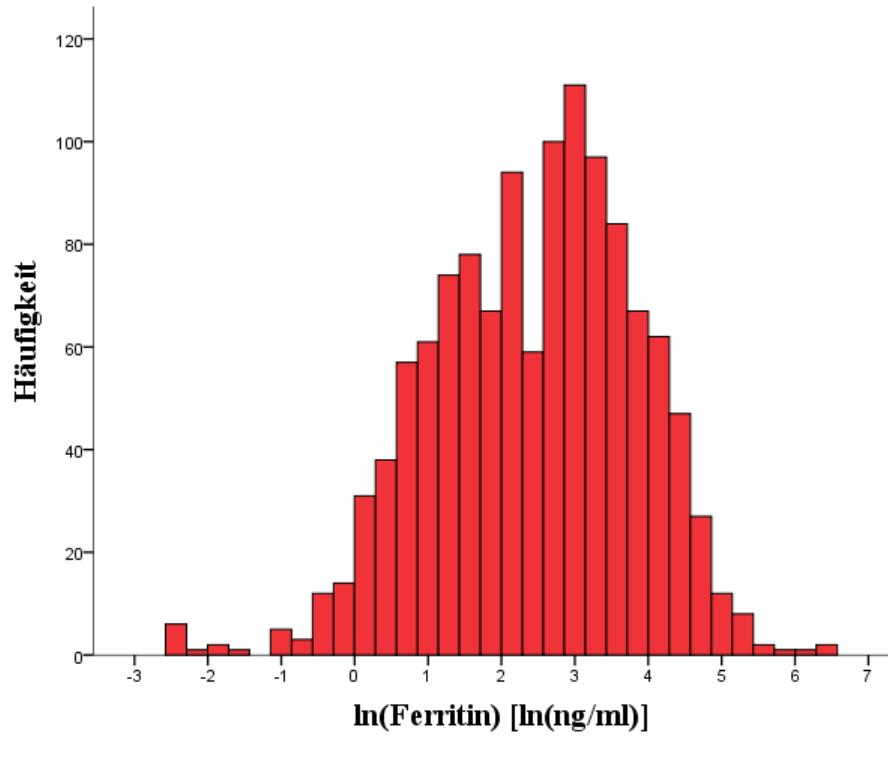
- Schiefe (g): $g < -1$: Verteilung ist linksschief = rechtsgipflig, Mittelwert < Median
- Schiefe (g): $g > +1$: Verteilung ist rechtsschief = linksgipflig, Mittelwert > Median

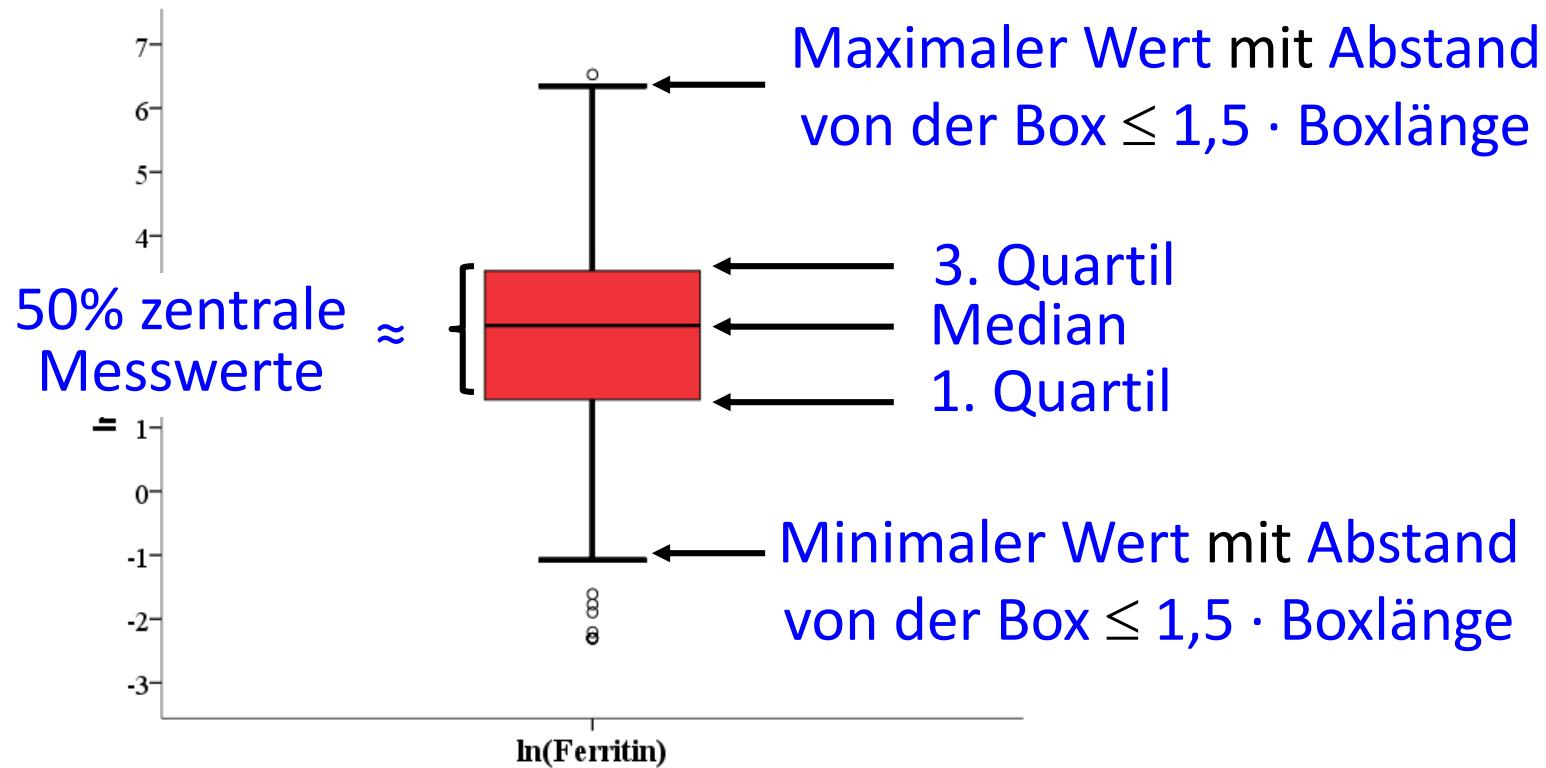


Welche Maße wann publizieren?

<i>Ausnahmen bestätigen die Regel!</i>	$ \text{Schiefe} \leq 1$	$ \text{Schiefe} > 1$
Mittelwert	sinnvoll	oft unsinnig
Varianz, Standardabweichung	sinnvoll	oft unsinnig
Median	sinnvoll	sinnvoll
Quartile, Interquartilsabstand	sinnvoll	sinnvoll
Minimum, Maximum, Spannweite	(sinnvoll)	(sinnvoll)

Graphische Darstellung stetiger Variablen



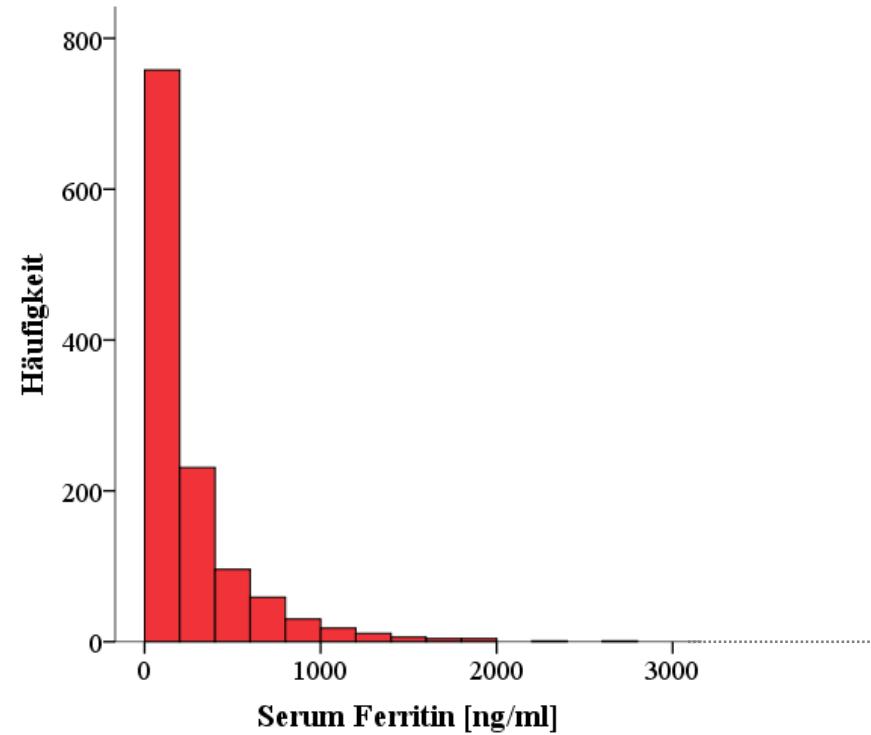


$1,5 \cdot$ Boxlänge < Abstand von der Box $\leq 3 \cdot$ Boxlänge: Ausreißer o

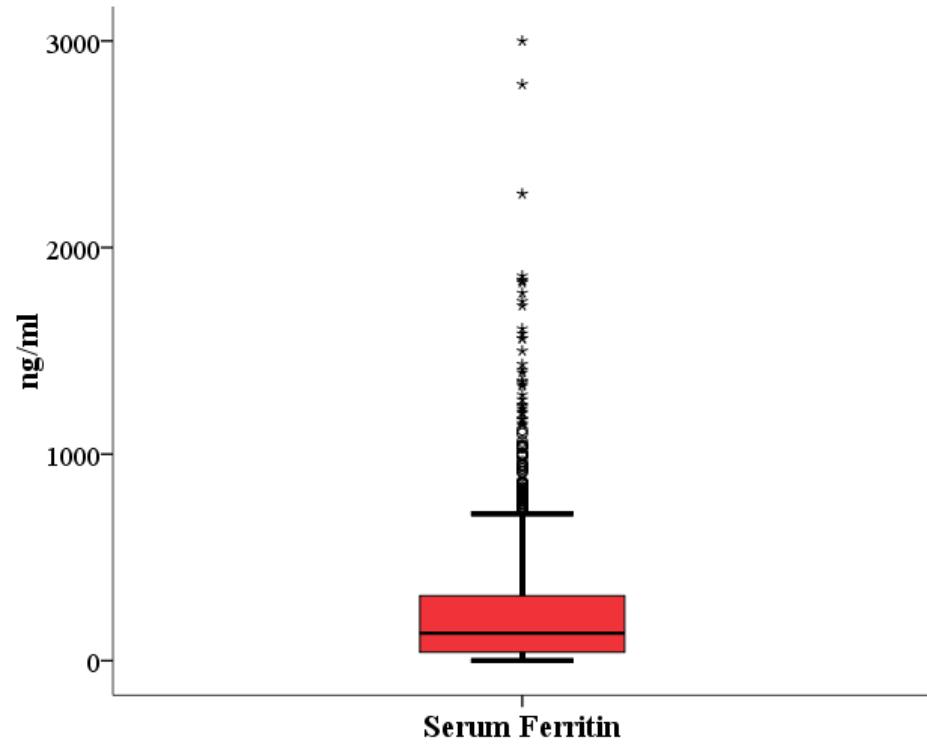
Abstand von der Box $> 3 \cdot$ Boxlänge: Extremwert *

Graphische Darstellung stetiger Variablen

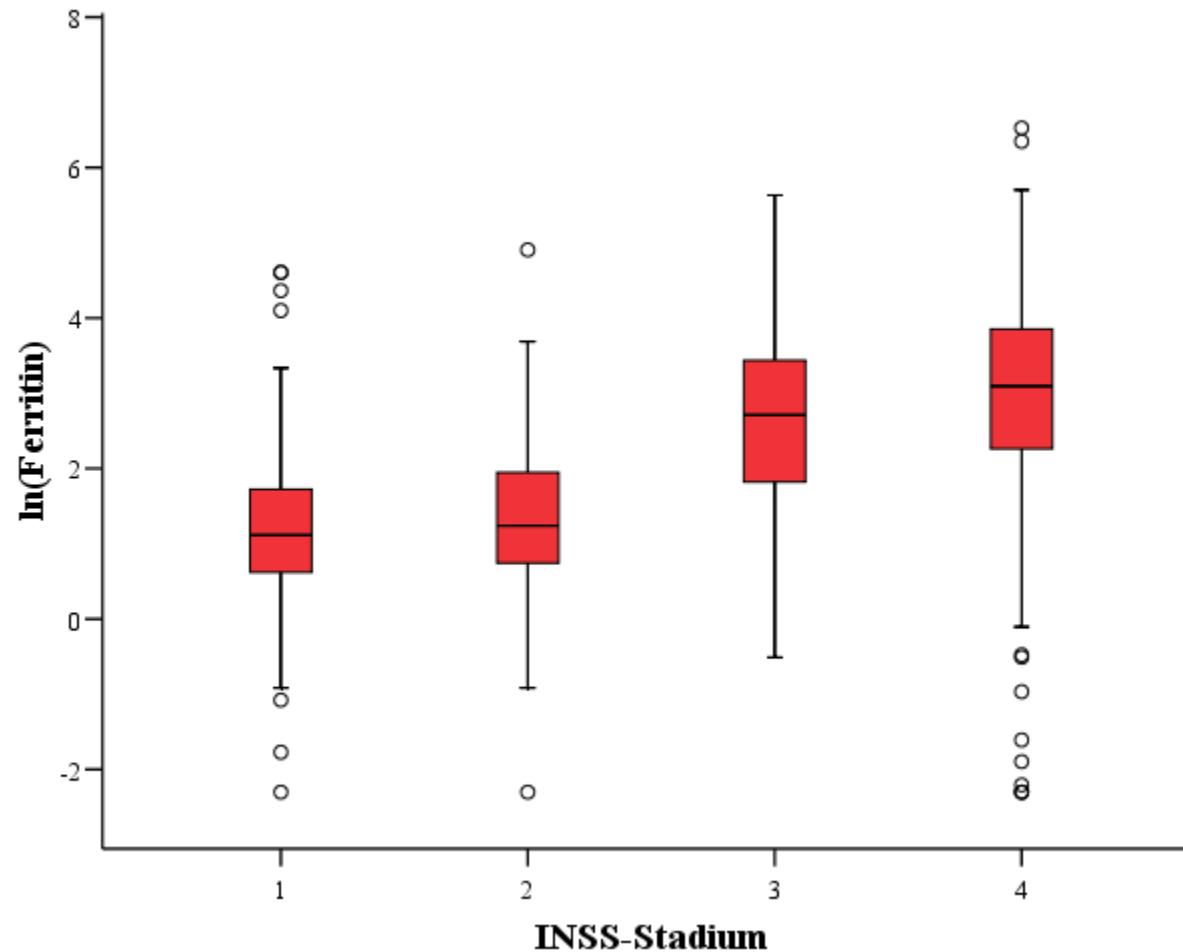
Histogramm



Box-and-Whisker Plot
(Boxplot)



Gruppiertter Boxplot

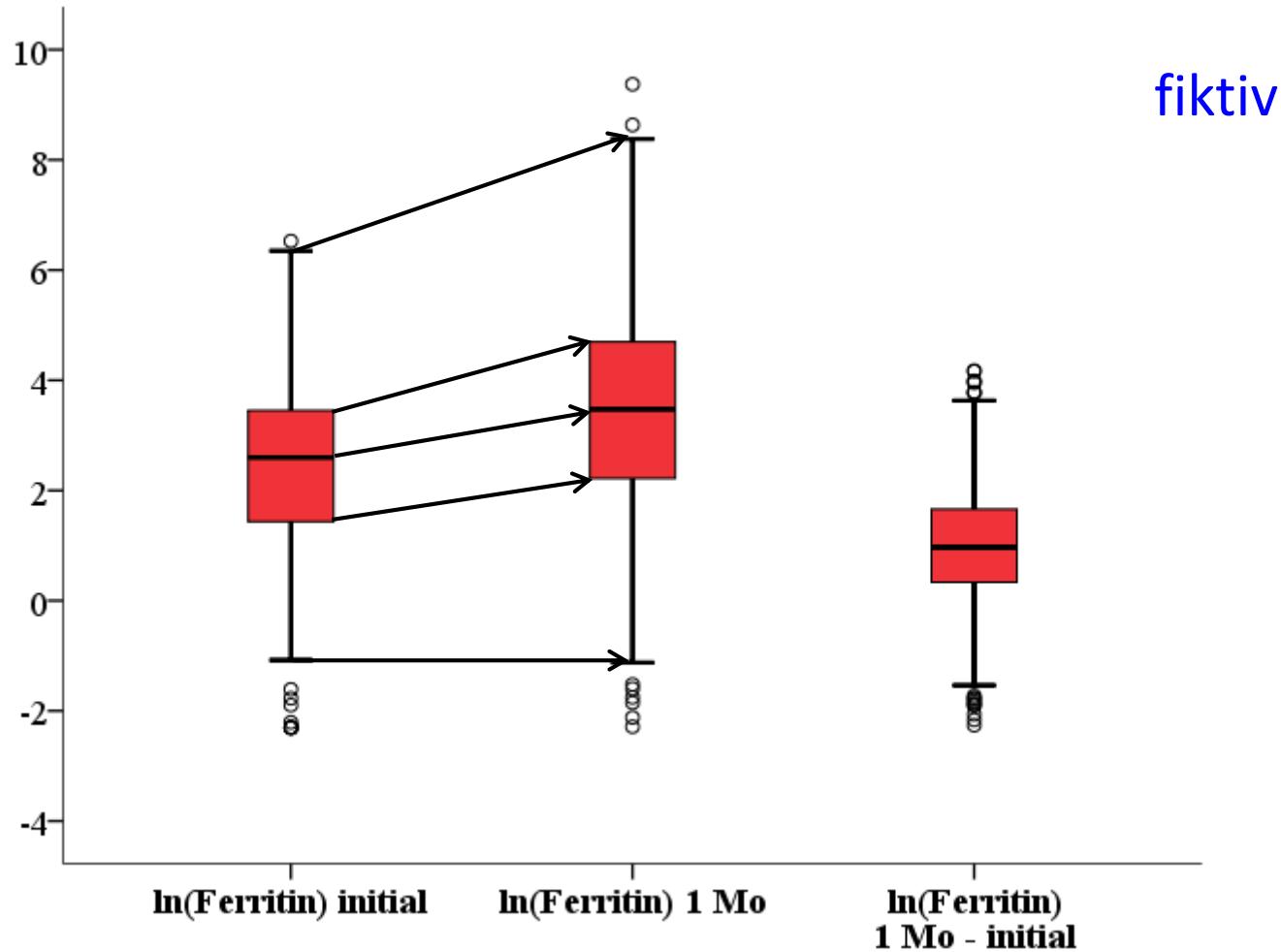


Verbundenes Design

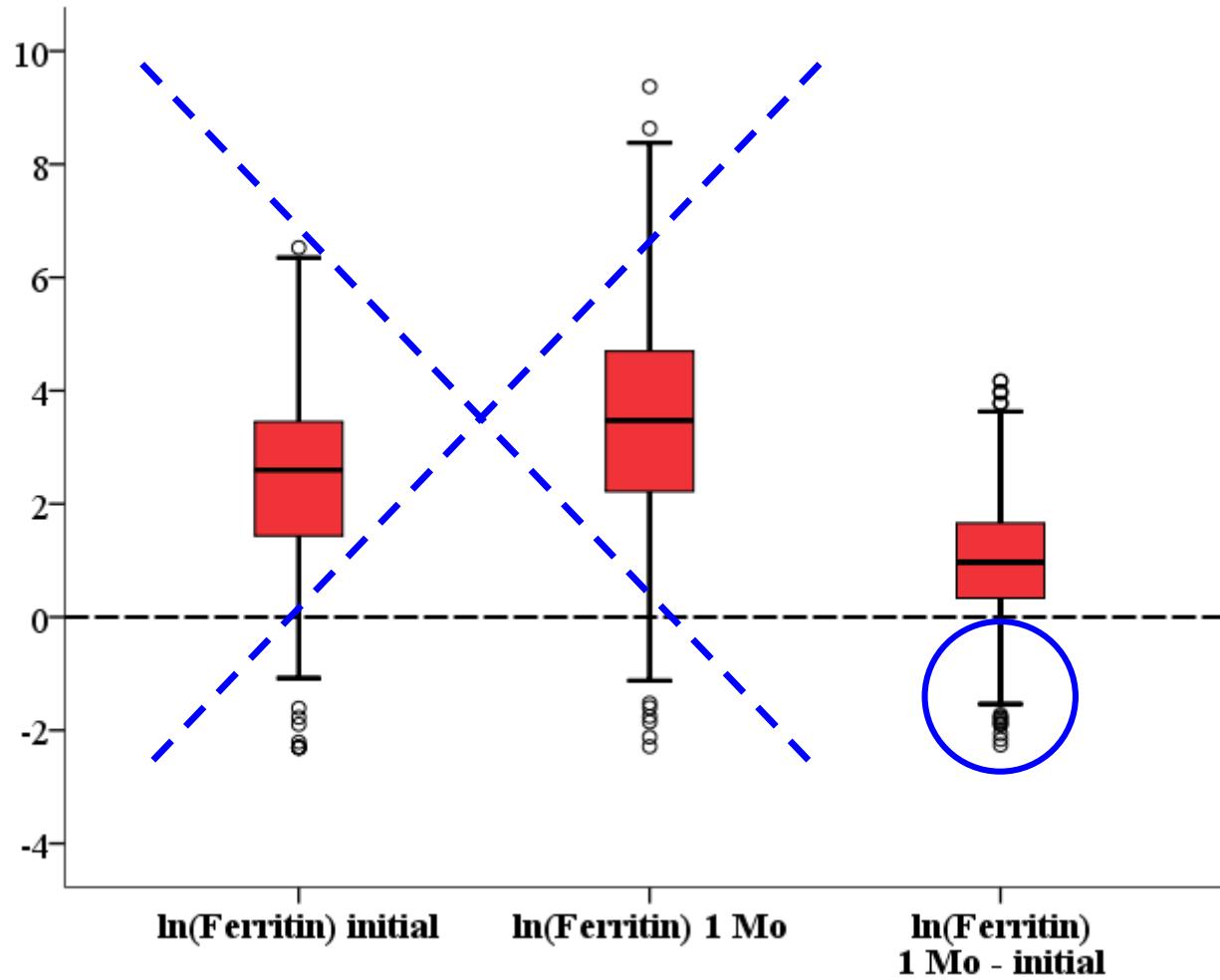
- Dieselbe Beobachtungseinheit:
 - z. B. prä- und postoperativer Blutdruck
 - Unterschied eines Tests am rechten und linken Auge
- Beobachtungspaare:
 - 1 Fall and 1 Kontrolle mit ähnlichen Eigenschaften
 - z. B. 1 leukämiekrankes Kind und 1 Kind gleichen Alters
- Vorteil: Reduktion der Variabilität
- Vorgehensweise:

Boxplot der Differenzen der gepaarten Werte statt
2 Einzelboxplots

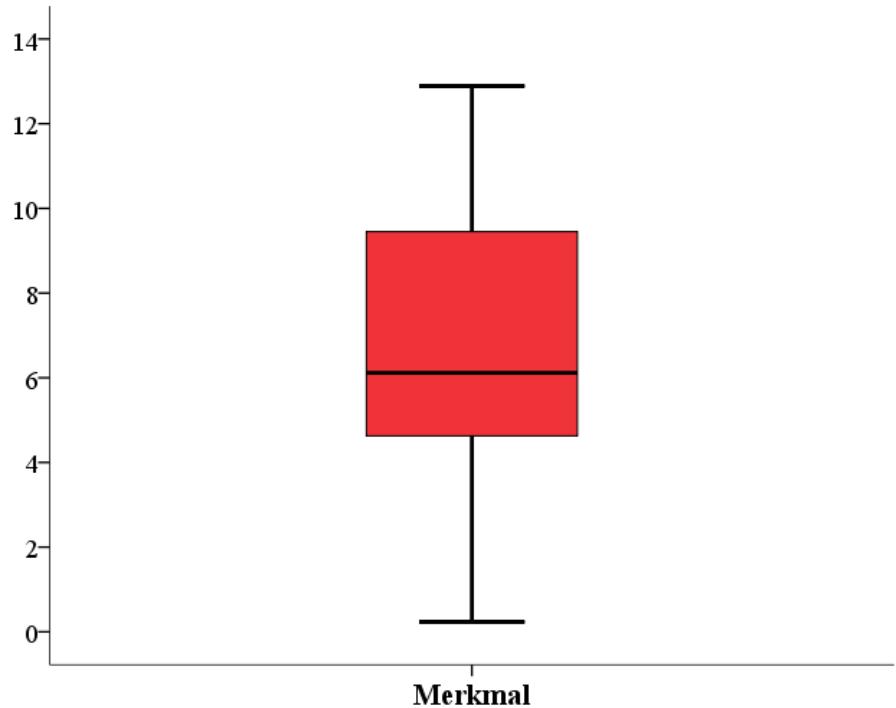
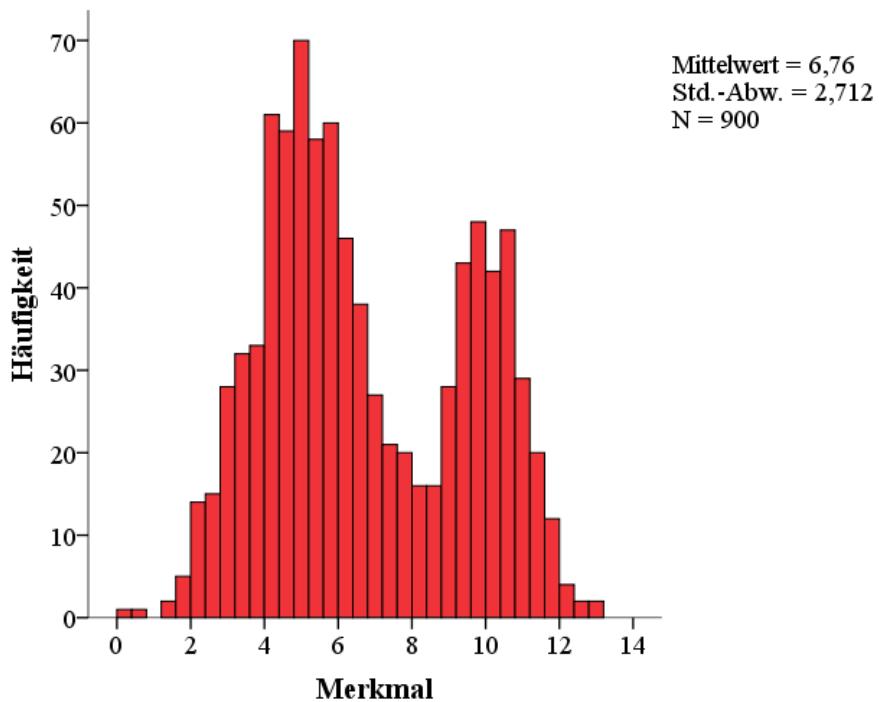
Differenzen-Boxplot



Differenzen-Boxplot



Histogramm vs. Boxplot



Boxplot zeigt nicht die Zahl der Gipfel

Histogramm vs. Boxplot

- **Vorteile** des Histogramms gegenüber dem Boxplot:
 - informativer
 - mehrere Gipfel sind zu erkennen
z.B. inhomogene Grundgesamtheit
- **Nachteile** des Histogramms gegenüber dem Boxplot:
 - willkürliche Wahl der Klassenzahl und –breite
 - mehrere Messreihen sind schwer vergleichbar
- Tendenz zugunsten des Boxplots
- **Mittlerweile unüblich: Mittelwertbalken + Standardfehler**

Vorlesungsinhalt

- **Deskriptive Statistik**
 - Darstellung 1es Merkmals
 - **Zusammenhang 2er Merkmale**
- **Statistisches Testen**
 - Prinzipien
 - Spezielle Tests
 - Multiples Testen
- **Fallzahlplanung / Adaptive Designs**
- **Lineare Regression / logistische Regression**
- **Überlebenszeiten**

NB 2004 - Vergleich der Risikofaktoren

Folgende Fragen sollen geklärt werden:

1. Gibt es einen Zusammenhang zwischen **LDH** und **Ferritin**?
2. Hängen **MYCN**-Amplifikation und **1p-Anomalie** zusammen?
3. Kommt die **MYCN**-Amplifikation in einigen **INSS-Stadien** häufiger vor?
4. Unterscheidet sich die **Ferritin**-Konzentration in den verschiedenen **INSS-Stadien**?

NB 2004 - Vergleich der Risikofaktoren

Folgende Fragen sollen geklärt werden:

1. Gibt es einen Zusammenhang zwischen LDH und Ferritin?
stetig vs. *stetig*
2. Hängen MYCN-Amplifikation und 1p-Anomalie zusammen?
3. Kommt die MYCN-Amplifikation in einigen INSS-Stadien häufiger vor?
4. Unterscheidet sich die Ferritin-Konzentration in den verschiedenen INSS-Stadien?

NB 2004 - Vergleich der Risikofaktoren

Folgende Fragen sollen geklärt werden:

1. Gibt es einen Zusammenhang zwischen LDH und Ferritin?
stetig vs. stetig
2. Hängen **MYCN**-Amplifikation und **1p-Anomalie** zusammen?
binär vs. **binär**
3. Kommt die MYCN-Amplifikation in einigen INSS-Stadien
häufiger vor?
4. Unterscheidet sich die Ferritin-Konzentration in den
verschiedenen INSS-Stadien?

NB 2004 - Vergleich der Risikofaktoren

Folgende Fragen sollen geklärt werden:

1. Gibt es einen Zusammenhang zwischen LDH und Ferritin?
stetig vs. stetig
2. Hängen MYCN-Amplifikation und 1p-Anomalie zusammen?
binär vs. binär
3. Kommt die MYCN-Amplifikation in einigen INSS-Stadien
häufiger vor?
binär vs. > 2 Kategorien
4. Unterscheidet sich die Ferritin-Konzentration in den
verschiedenen INSS-Stadien?

NB 2004 - Vergleich der Risikofaktoren

Folgende Fragen sollen geklärt werden:

1. Gibt es einen Zusammenhang zwischen LDH und Ferritin?
stetig vs. stetig
2. Hängen MYCN-Amplifikation und 1p-Anomalie zusammen?
binär vs. binär
3. Kommt die MYCN-Amplifikation in einigen INSS-Stadien
häufiger vor?
binär vs. > 2 Kategorien
4. Unterscheidet sich die **Ferritin**-Konzentration in den
verschiedenen **INSS-Stadien**?
stetig vs. **> 2 Kategorien**

	kategorial		stetig	
	binär	> 2 Kategorien	normalverteilt	nicht normalverteilt
binär				
> 2 Kategorien				
kategorial				
stetig	normal-verteilt	nicht norm.		

		kategorial		stetig
	binär	> 2 Kategorien	normalverteilt	nicht normalverteilt
binär			2 Boxplots / Histogramme 2*stetige Maßzahlen mit Mittelwert, Std.	?
kategorial	> 2 Kategorien			
stetig	normal-verteilt			
	nicht norm.			

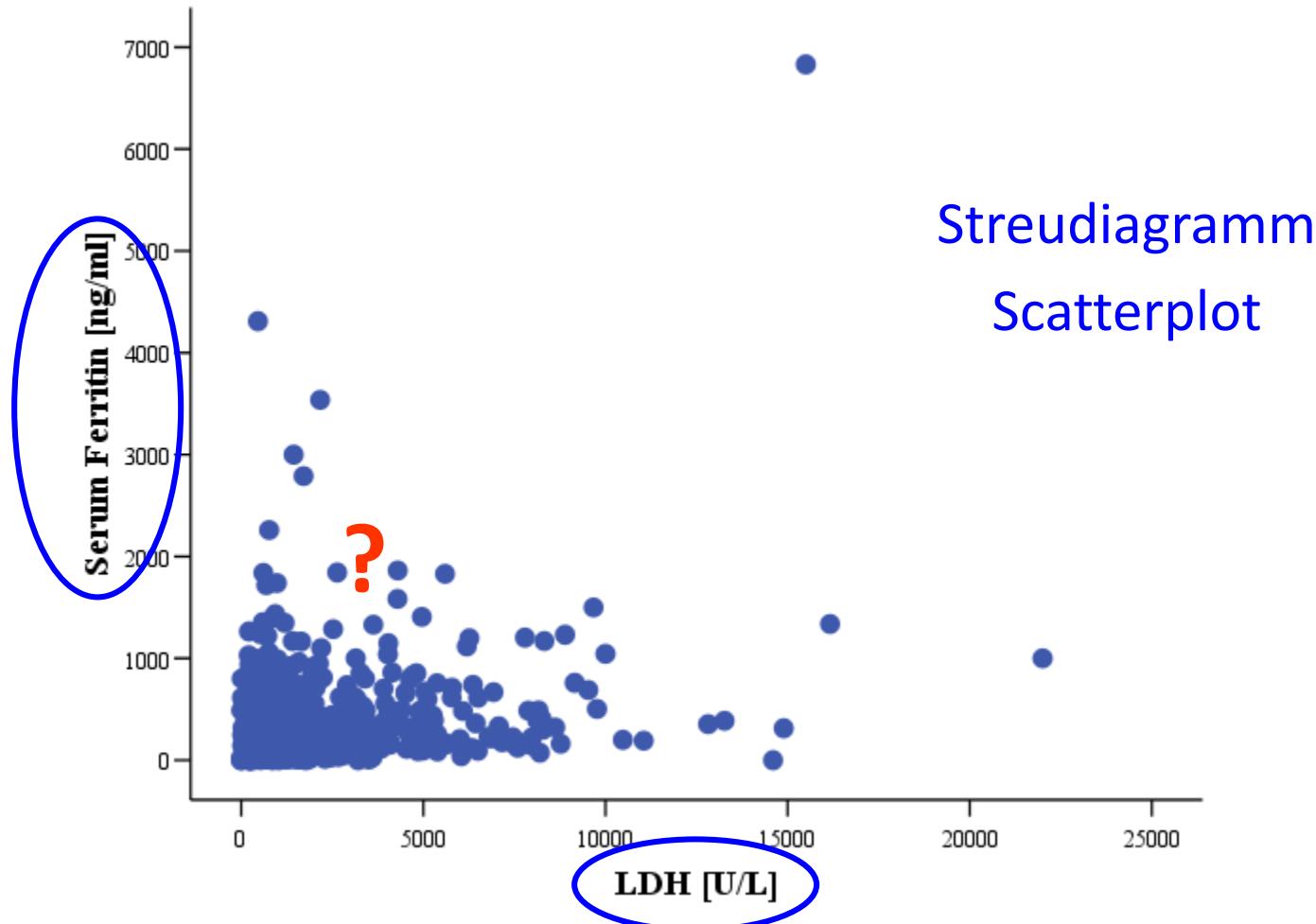
		kategorial		stetig
	binär	> 2 Kategorien	normalverteilt	nicht normalverteilt
binär	?	?	2 Boxplots / Histogramme 2*stetige Maßzahlen mit Mittelwert, Std.	2 Boxplots / Histogramme 2*stetige Maßzahlen ohne Mittelwert, Std.
kategorial	> 2 Kategorien		?	?
stetig	normal-verteilt		?	?
	nicht norm.			

NB 2004 - Vergleich der Risikofaktoren

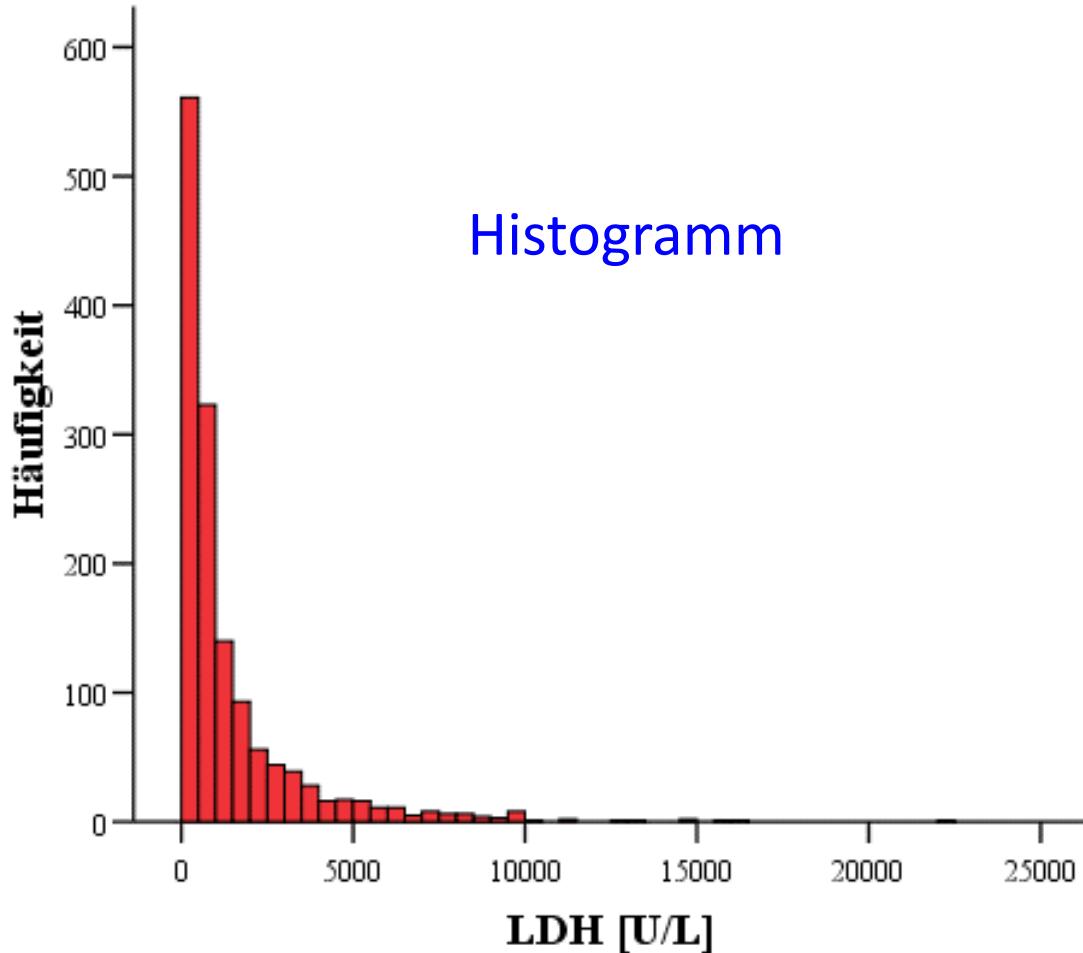
Folgende Fragen sollen geklärt werden:

1. Gibt es einen Zusammenhang zwischen LDH und Ferritin?
2. Hängen MYCN-Amplifikation und 1p-Anomalie zusammen?
3. Kommt die MYCN-Amplifikation in einigen INSS-Stadien häufiger vor?
4. Unterscheidet sich die Ferritin-Konzentration in den verschiedenen INSS-Stadien?

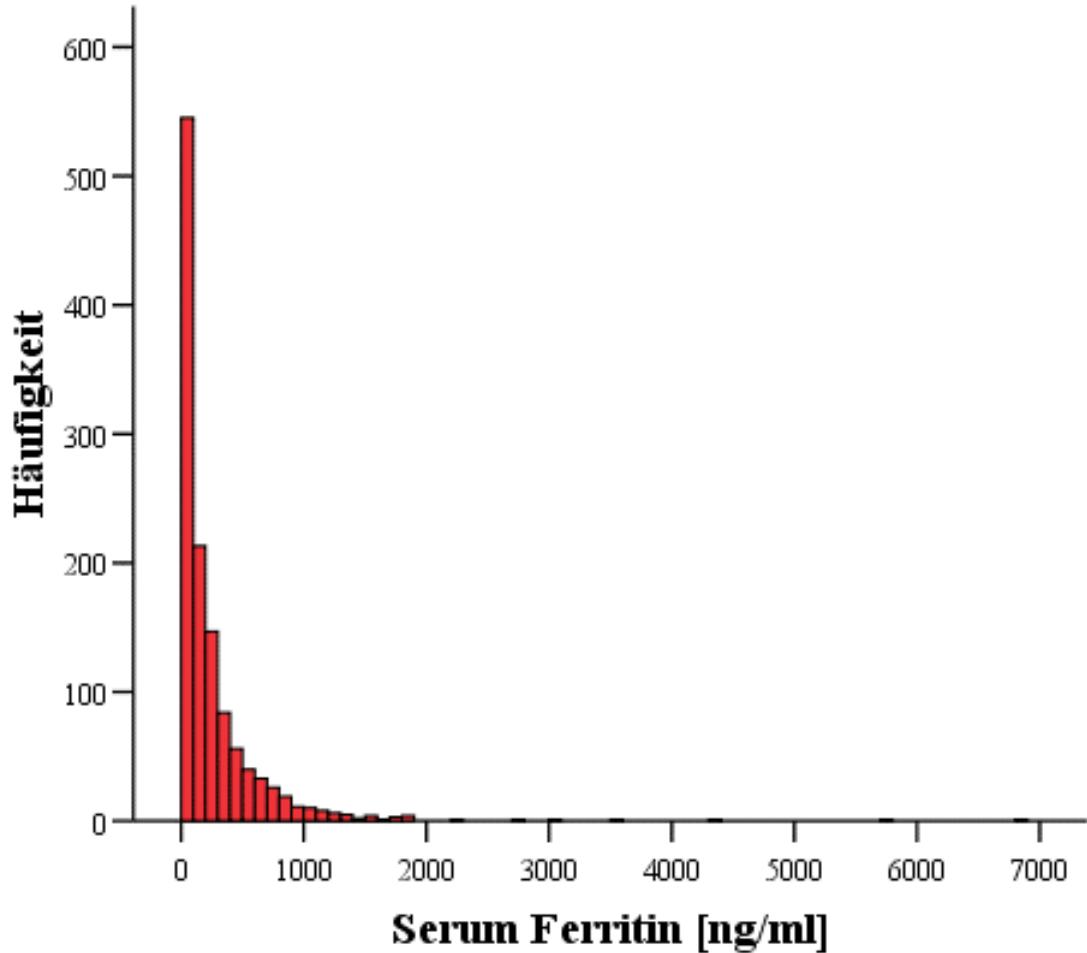
Zusammenhang zwischen LDH und Ferritin



LDH

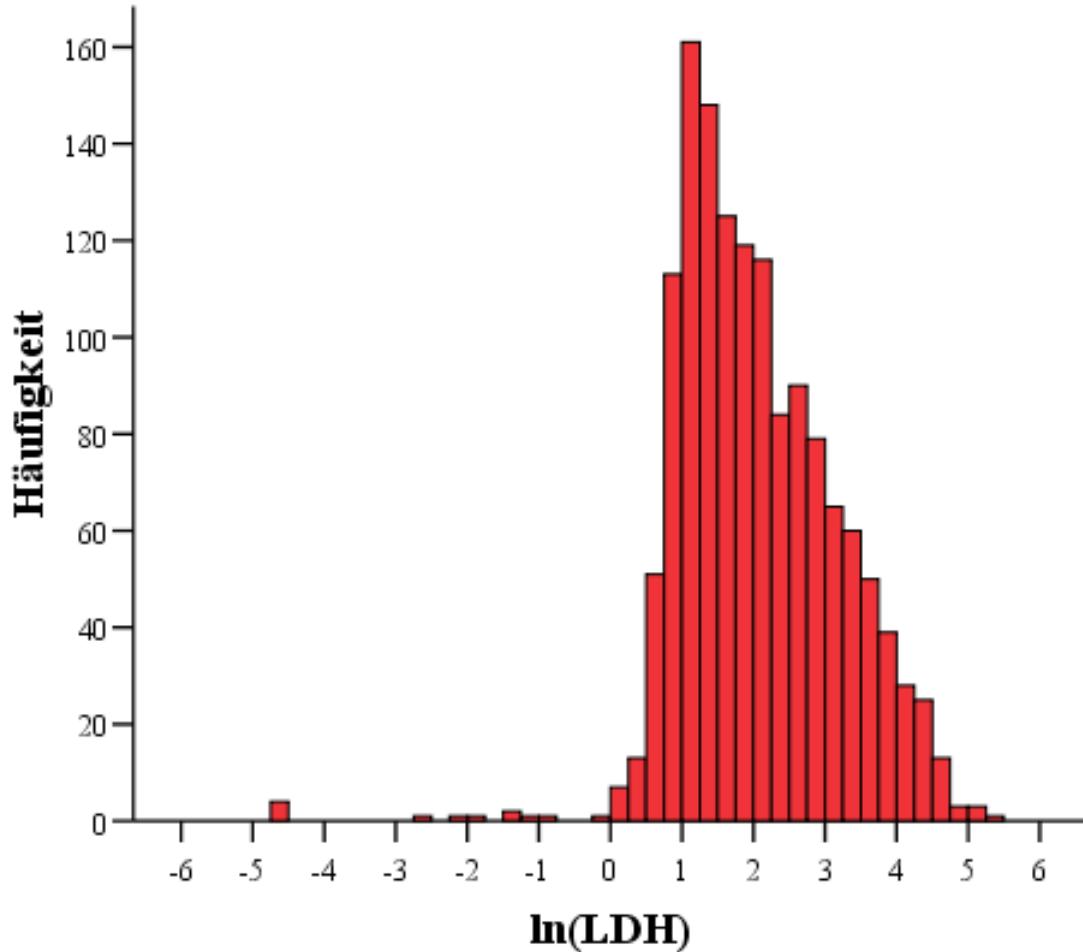


Serum Ferritin



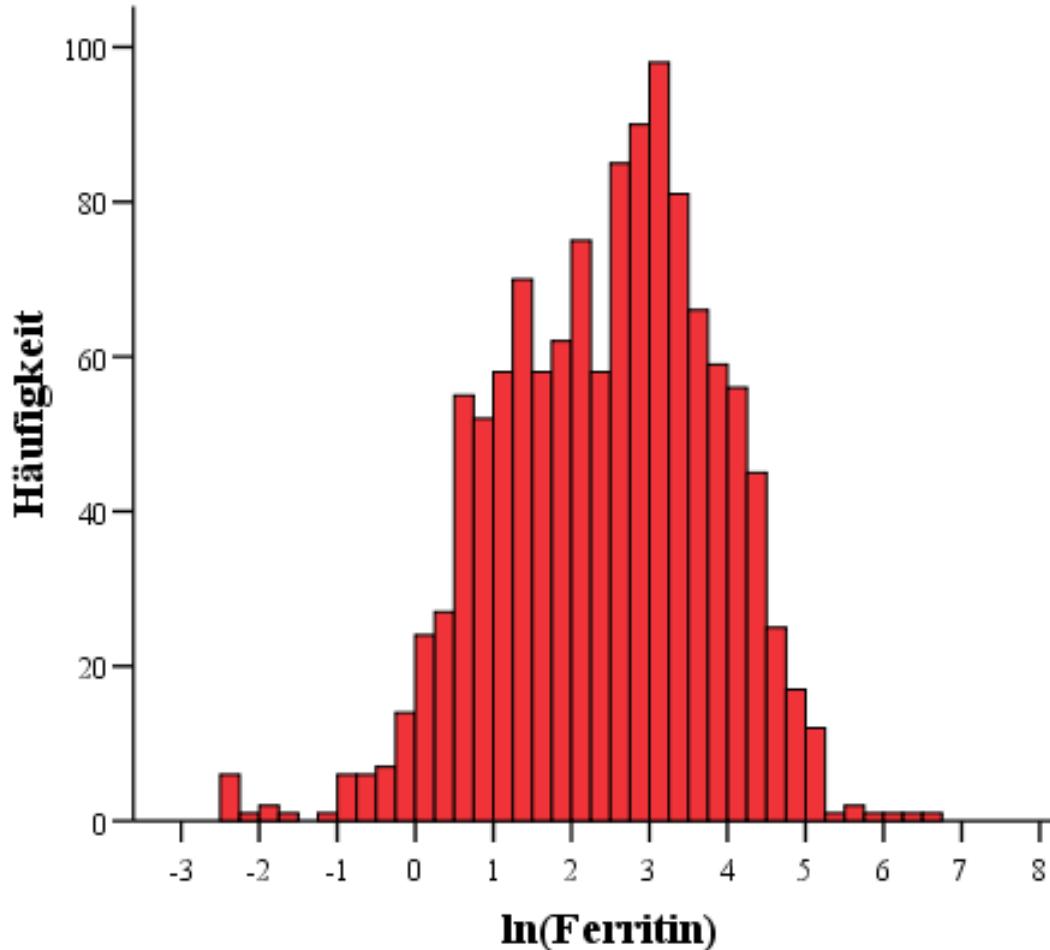
	Serum Ferritin [ng/ml]
N	1224
Gültig	2121
Fehlend	
Mittelwert	261
Median	133
Standardabweichung	432
Schiefe	7
Minimum	0
Maximum	6831
Perzentile	
25	41
50	133
75	315

Logarithmus von LDH



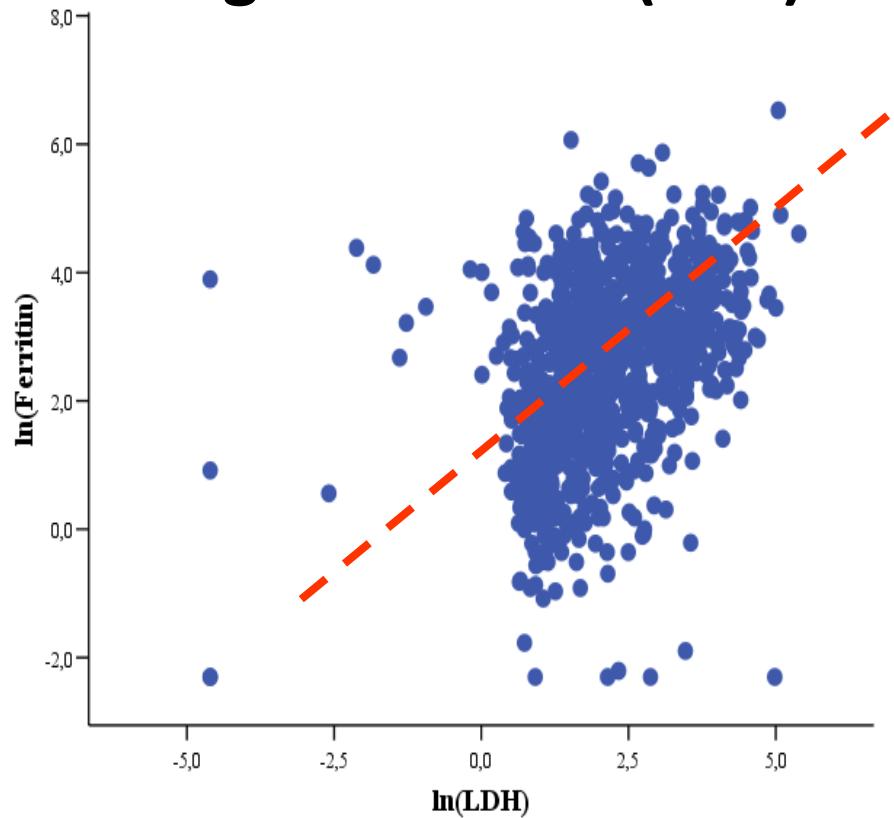
	$\ln(\text{LDH})$
N	1405
Gültig	
Fehlend	1940
Mittelwert	2,1
Median	1,9
Standardabweichung	1,1
Schiefe	-,3
Minimum	-4,6
Maximum	5,4
Perzentile	
25	1,2
50	1,9
75	2,8

Logarithmus von Ferritin



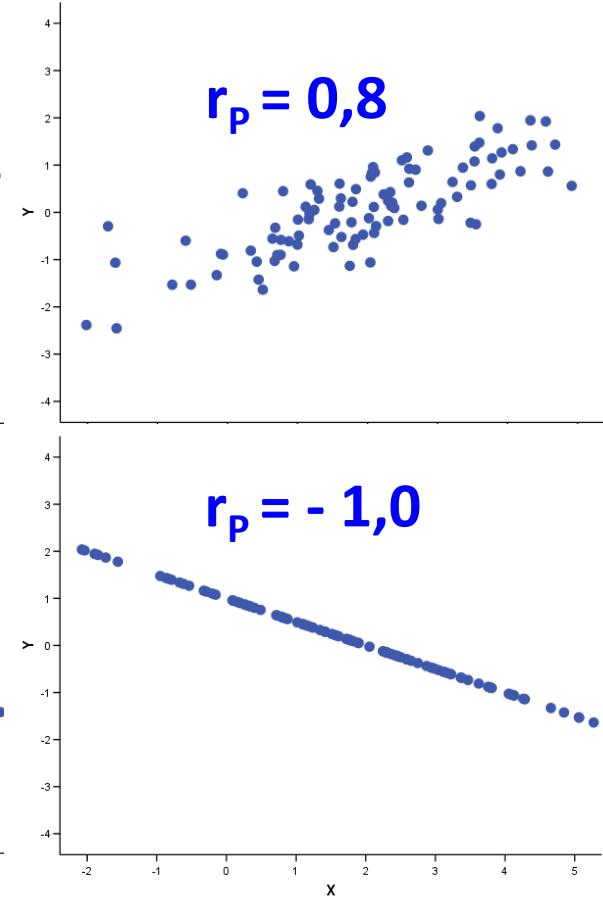
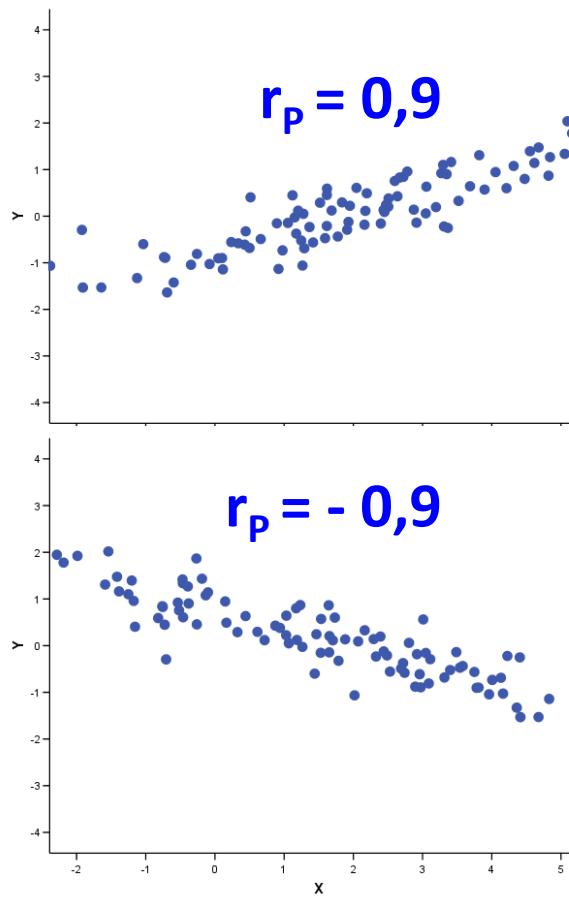
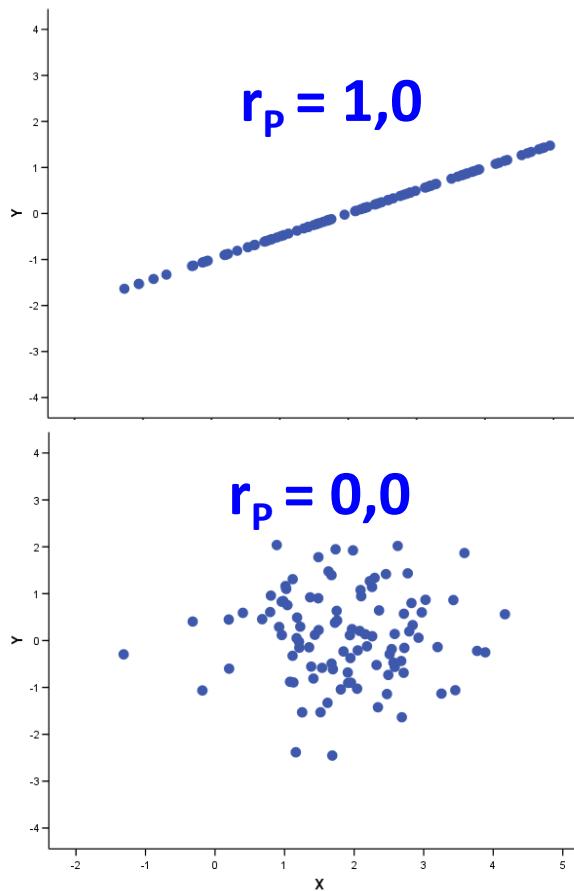
	$\ln(\text{Ferritin})$
N	1224
Gültig	2121
Fehlend	
Mittelwert	2,4
Median	2,6
Standardabweichung	1,4
Schiefe	-,3
Minimum	-2,3
Maximum	6,5
Perzentile	25 1,4
	50 2,6
	75 3,5

Zusammenhang zwischen $\ln(\text{LDH})$ und $\ln(\text{Ferritin})$



Wie kann man diesen Zusammenhang bewerten?

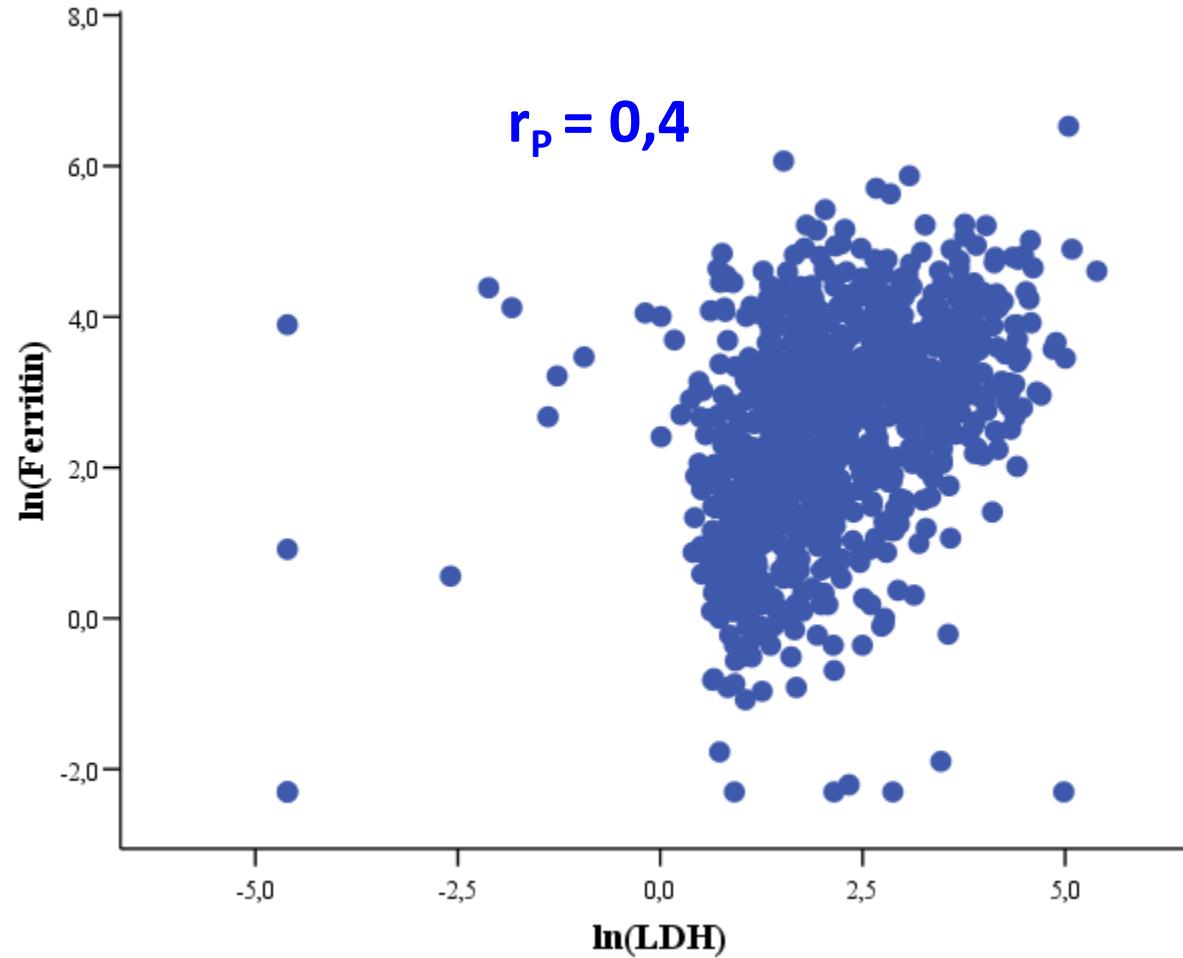
Korrelationskoeffizient und Streudiagramm



Korrelationskoeffizient nach Pearson r_p

- **Vorzeichen** von r_p : **+** bedeutet einen **positiven** und **-** bedeutet einen **negativen linearen Zusammenhang**
- $r_p = 1$: Perfekter **positiver linearer Zusammenhang**, d.h. alle **Punkte liegen** auf einer **steigenden Geraden**
- $r_p = -1$: Perfekter **negativer linearer Zusammenhang**, d.h. alle **Punkte liegen** auf einer **fallenden Geraden**
- $r_p = 0$: Kein **linearer Zusammenhang**
- Je **höher** $|r_p|$ ist, desto **deutlicher** ist der **lineare Zusammenhang**
- **Vertauschung** von X und Y ändert r_p **nicht**
- r_p ist **nur** für zwei **normalverteilte** Merkmale **aussagekräftig**

Korrelation zwischen $\ln(\text{LDH})$ und $\ln(\text{Ferritin})$

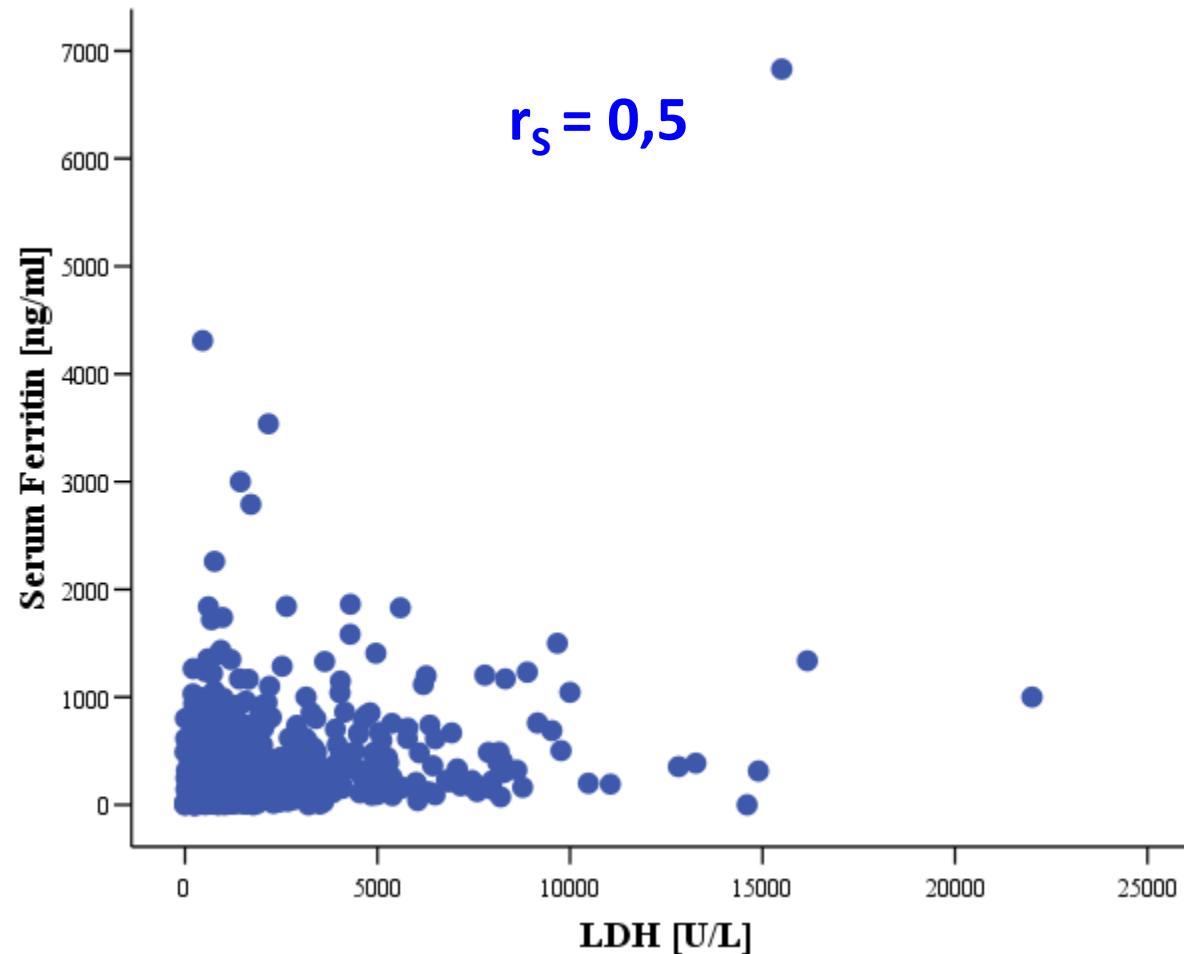


		kategorial	stetig	
	binär	> 2 Kategorien	normalverteilt	nicht normalverteilt
binär			2 Boxplots / Histogramme 2*stetige Maßzahlen mit Mittelwert, Std.	2 Boxplots / Histogramme 2*stetige Maßzahlen ohne Mittelwert, Std.
kategorial	> 2 Kategorien			
stetig	normal-verteilt		Streudiagramm Korrelationskoeffizient nach Pearson r_p	?
	nicht norm.			

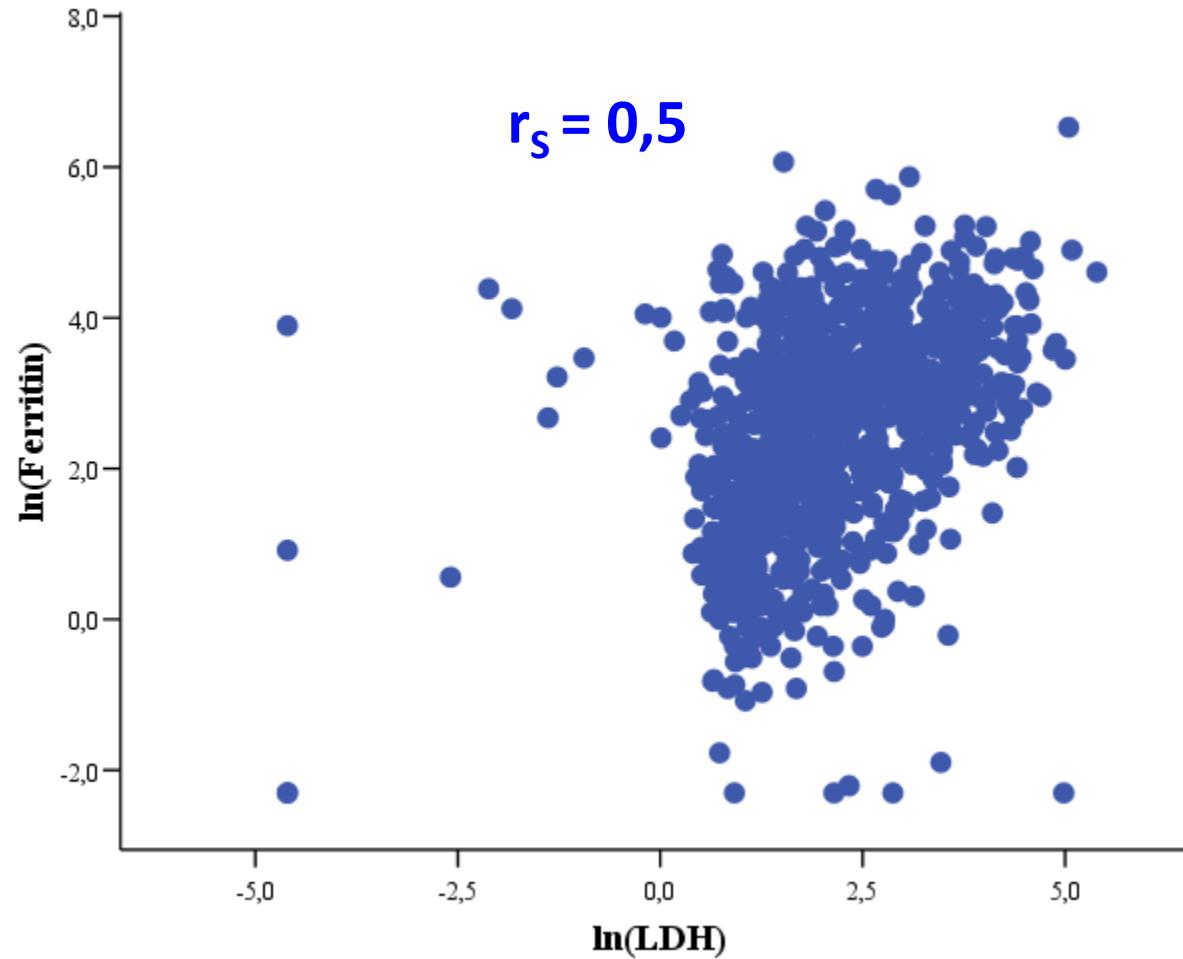
Zusammenhang nicht normalverteilter stetiger Größen

- Korrelationskoeffizient nach Spearman r_s
- Originalwerte werden durch ihre Ränge ersetzt:
aus 23, 100, 2004, ... wird 1, 2, 3, ...
- Berechnung nach der Formel für den Korrelationskoeffizienten nach Pearson mit Rängen statt Originalwerten
- Insbesondere: r_s liegt zwischen -1 und 1
- $r_s = 1$: Perfekter positiver monotoner Zusammenhang zwischen X und Y, d.h. X wächst genau dann, wenn Y wächst
- $r_s = -1$: Perfekter negativer monotoner Zusammenhang zwischen X und Y, d.h. X wächst genau dann, wenn Y fällt
- $r_s = 0$: Kein monotoner Zusammenhang

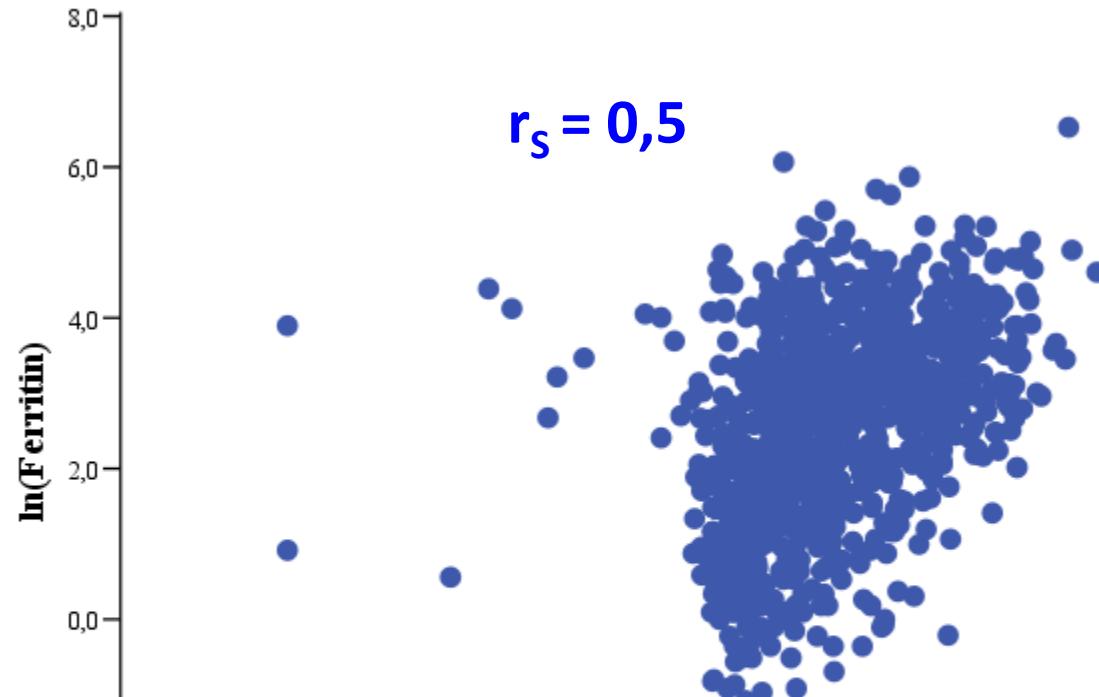
Zusammenhang zwischen LDH und Ferritin



Zusammenhang zwischen $\ln(\text{LDH})$ und $\ln(\text{Ferritin})$



Zusammenhang zwischen $\ln(\text{LDH})$ und $\ln(\text{Ferritin})$



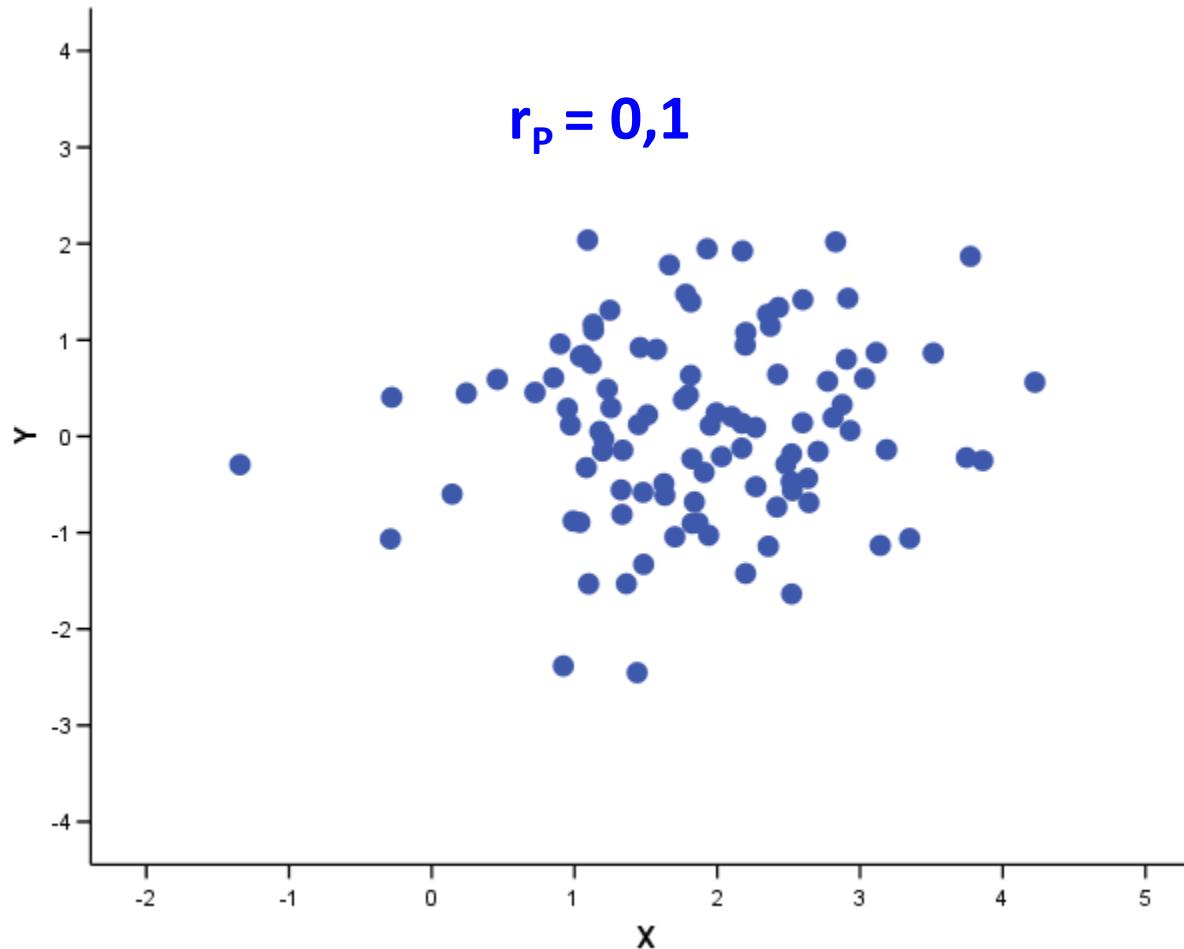
Logarithmieren ist eine monotone Transformation:

$$a < b \Leftrightarrow \ln(a) < \ln(b)$$

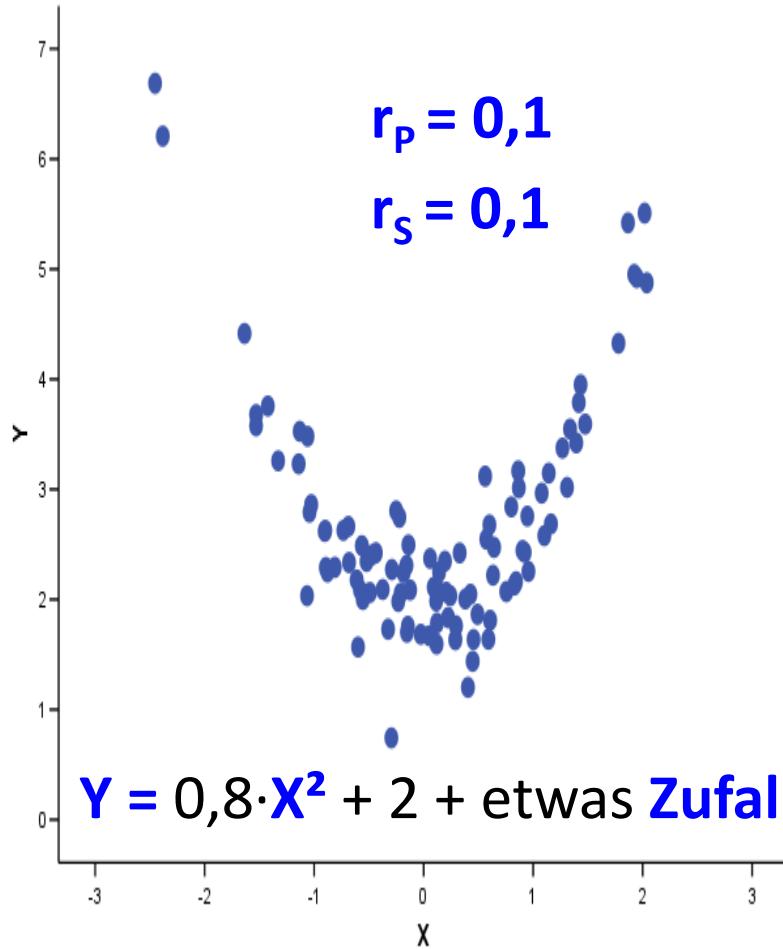
\Rightarrow Ränge von X und $\ln(X)$ bzw. Y und $\ln(Y)$ sind gleich!

		kategorial		stetig
	binär	> 2 Kategorien	normalverteilt	nicht normalverteilt
binär			2 Boxplots / Histogramme 2*stetige Maßzahlen mit Mittelwert, Std.	2 Boxplots / Histogramme 2*stetige Maßzahlen ohne Mittelwert, Std.
kategorial	> 2 Kategorien			
stetig	normal-verteilt		Streudiagramm Korrelationskoeffizient nach Pearson r_p	Streudiagramm Korrelationskoeffizient nach Spearman r_s
stetig	nicht norm.			

Vorsicht: Korrelationskoeffizient und Streudiagramm



Vorsicht: Korrelationskoeffizient und Streudiagramm



Wichtig:

- Immer erst das Streudiagramm betrachten!
- Korrelationskoeffizient nach Pearson bewertet den linearen Zusammenhang!
- Korrelationskoeffizient nach Spearman bewertet den monotonen Zusammenhang!
- Viele andere Zusammenhänge sind möglich!

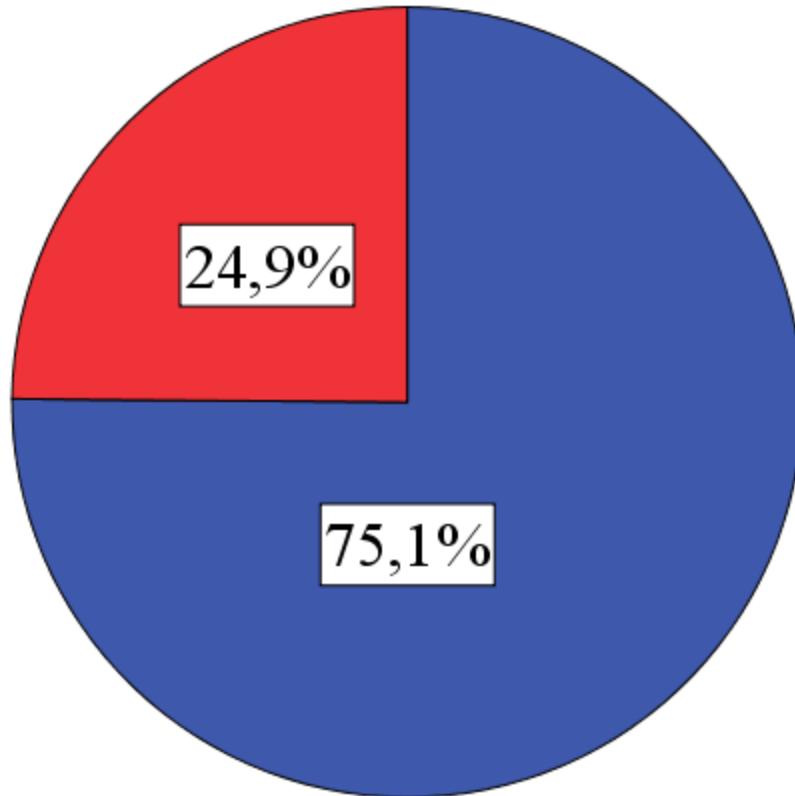
		kategorial	stetig	
	binär	> 2 Kategorien	normalverteilt	nicht normalverteilt
binär	?		2 Boxplots / Histogramme 2*stetige Maßzahlen mit Mittelwert, Std.	2 Boxplots / Histogramme 2*stetige Maßzahlen ohne Mittelwert, Std.
kategorial	> 2 Kategorien			
stetig	normal-verteilt		Streudiagramm Korrelationskoeffizient nach Pearson r_p	Streudiagramm Korrelationskoeffizient nach Spearman r_s
norm.	nicht norm.			

NB 2004 - Vergleich der Risikofaktoren

Folgende Fragen sollen geklärt werden:

1. Gibt es einen Zusammenhang zwischen LDH und Ferritin?
2. Hängen **MYCN**-Amplifikation und **1p**-Anomalie zusammen?
3. Kommt die MYCN-Amplifikation in einigen INSS-Stadien häufiger vor?
4. Unterscheidet sich die Ferritin-Konzentration in den verschiedenen INSS-Stadien?

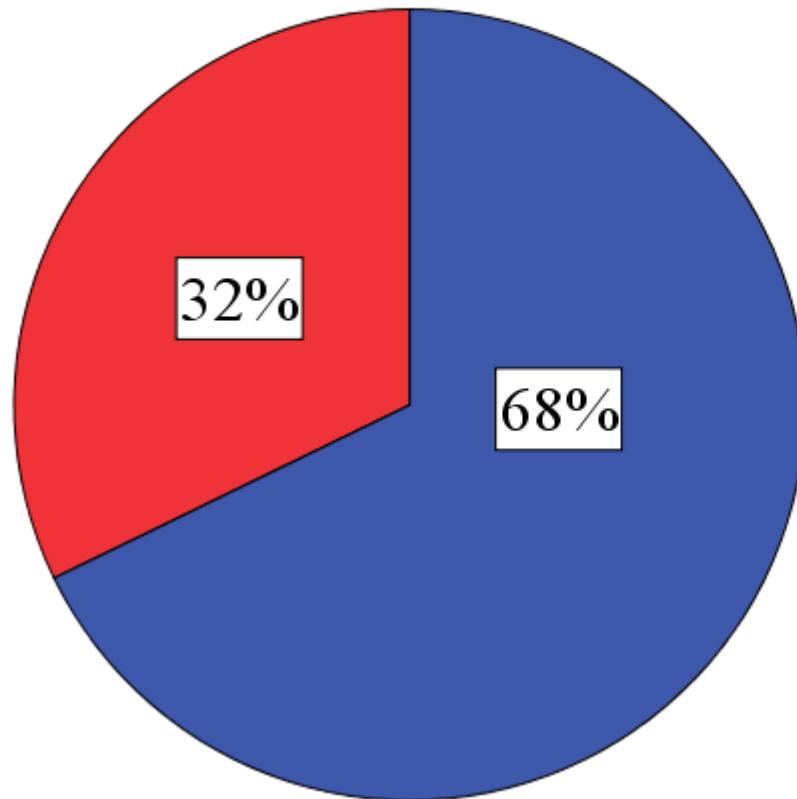
MYCN-Amplifikation



nein
ja

		Häufigkeit	Prozent	Gültige Prozente
Gültig	nein	1538	46,0	75,1
	ja	509	15,2	24,9
Fehlend	Gesamt	2047	61,2	100,0
	System	1298	38,8	
	Gesamt	3345	100,0	

1p-Anomalie



nein
ja

		Häufigkeit	Prozent	Gültige Prozente
Gültig	nein	141	4	68
	ja	67	2	32
Fehlend	Gesamt	208	6	100
	System	3137	94	
Gesamt		3345	100	

Zusammenhang zwischen MYCN und 1p-Anomalie

MYCN-Amplifikation * 1p-Anomalie Kreuztabelle

MYCN-Amplifikation	nein	Anzahl	1p-Anomalie		Gesamt
			nein	ja	
nein			114	28	142
ja			5	32	37
Gesamt		Anzahl	119	60	179

Zusammenhang zwischen MYCN und 1p-Anomalie

MYCN-Amplifikation * 1p-Anomalie Kreuztabelle

		1p-Anomalie		Gesamt
		nein	ja	
MYCN-Amplifikation	nein	Anzahl	114	28
		% von MYCN-Amplifikation	80%	20%
	ja	Anzahl	5	32
		% von MYCN-Amplifikation	14%	86%
Gesamt		Anzahl	119	60
		% von MYCN-Amplifikation	66%	34%
				179
				100%

Zeilenprozente

Zusammenhang zwischen MYCN und 1p-Anomalie

MYCN-Amplifikation * 1p-Anomalie Kreuztabelle

		1p-Anomalie		Gesamt
		nein	ja	
MYCN-Amplifikation	nein	Anzahl	114	28
	ja	% von 1p-Anomalie	96%	47%
Gesamt	ja	Anzahl	5	32
	Gesamt	% von 1p-Anomalie	4%	53%
		119	60	179
		% von 1p-Anomalie	100%	100%
				100%

Spaltenprozente

Zusammenhang zwischen MYCN und 1p-Anomalie

MYCN-Amplifikation * 1p-Anomalie Kreuztabelle

		1p-Anomalie		Gesamt
		nein	ja	
MYCN-Amplifikation	nein	114	28	142
		80%	20%	100%
		96%	47%	79%
	ja	5	32	37
		14%	86%	100%
		4%	53%	21%
Gesamt		119	60	179
		66%	34%	100%
		100%	100%	100%

Kreuztabelle = Kontingenztafel

Regeln für Kreuztabellen

- Einflussgröße = unabhängiges Merkmal: Zeile
- Zielgröße = abhängiges Merkmal: Spalte
- Kohortenstudien:
Bei Beginn der Studie Auswahl der Kohorten nach der Einflussgröße \Rightarrow Zeilenprozente
- Fall-Kontroll-Studien:
Bei Beginn der Studie Auswahl von Fällen und Kontrollen nach der Zielgröße \Rightarrow Spaltenprozente
- In unserem Beispiel: Keine natürliche Reihenfolge vorhanden
 \Rightarrow Angabe von Zeilen- und Spaltenprozenten sinnvoll

Regeln für Kreuztabellen

	„geheilt“	„nicht geheilt“	
Therapie A	80 (80%) (40%)	20 (20%) (20%)	100 (100%)
Therapie B	120 (60%) (60%)	80 (40%) (80%)	200 (100%)
	200 (100%)	100 (100%)	300

Was erscheint hier sinnvoller?

- Zeilenprozente oder Spaltenprozente?

Risikomaße

- **Risiko:** relative Häufigkeit oder Wahrscheinlichkeit, dass ein Ereignis auftritt (Nebenwirkung, Heilung, ...)
- **Risikodifferenz:** Differenz der Risiken in 2 Gruppen
- **Odds = Chance:**
Quotient aus Risiko für ein Ereignis und Risiko für kein Ereignis

Bspl.: Wahrscheinlichkeit für eine 6 beim Würfeln = 1/6

Chance auf eine 6 beim Würfeln = 1/6 : 5/6 = 1 : 5

- **Relatives Risiko (RR):** Quotient der Risiken zwischen 2 Gruppen
- **Odds Ratio = Chancenverhältnis (OR):**
Quotient der Odds zwischen 2 Gruppen

Risikomaße

MYCN-Amplifikation * 1p-Anomalie Kreuztabelle

% innerhalb von MYCN-Amplifikation

		1p-Anomalie		Gesamt
		nein	ja	
MYCN-Amplifikation	nein	80%	20%	100%
	ja	14%	86%	100%
Gesamt		66%	34%	100%

- Risiko: relative Häufigkeit oder Wahrscheinlichkeit, dass ein Ereignis auftritt (Nebenwirkung, Heilung, ...)
- Risiko einer 1p-Anomalie
 - ohne MYCN-Amplifikation: 20% = 0,20
 - mit MYCN-Amplifikation: 86% = 0,86
- Risikodifferenz einer 1p-Anomalie zwischen PatientInnen mit und ohne MYCN-Amplifikation: $0,86 - 0,20 = 0,66$

MYCN-Amplifikation * 1p-Anomalie Kreuztabelle

% innerhalb von MYCN-Amplifikation

		1p-Anomalie		Gesamt
		nein	ja	
MYCN-Amplifikation	nein	80%	20%	100%
	ja	14%	86%	100%
Gesamt		66%	34%	100%

- **Relatives Risiko:** Quotient der Risiken zwischen 2 Gruppen
- **Relatives Risiko einer 1p-Anomalie**
von PatientInnen mit MYCN-Amplifikation im Vergleich zu PatientInnen ohne MYCN-Amplifikation: $0,86 / 0,20 = 4,3$
- Sprechweisen: Das **Risiko einer 1p-Anomalie** ist
 - um **Faktor 4,3** höher,
 - um $(4,3 - 1) \cdot 100\% = 330\%$ erhöht,wenn **MYCN amplifiziert** ist (vs. ohne MYCN-Amplifikation).

MYCN-Amplifikation * 1p-Anomalie Kreuztabelle

% innerhalb von MYCN-Amplifikation

		1p-Anomalie		Gesamt
		nein	ja	
MYCN-Amplifikation	nein	80%	20%	100%
	ja	14%	86%	100%
Gesamt		66%	34%	100%

- **Relatives Risiko:** Quotient der Risiken zwischen 2 Gruppen
- **Relatives Risiko einer 1p-Anomalie**
von PatientInnen **ohne MYCN-Amplifikation** im Vergleich zu
PatientInnen **mit MYCN-Amplifikation:** $0,20 / 0,86 = 0,23$
- Sprechweisen: Das **Risiko einer 1p-Anomalie** ist
 - um **Faktor 0,23 niedriger**,
 - um $(1 - 0,23) \cdot 100\% = 77\%$ **gesenkt**,wenn MYCN **nicht amplifiziert** ist (vs. MYCN-Amplifikation).

Risikomaße

MYCN-Amplifikation * 1p-Anomalie Kreuztabelle

% innerhalb von MYCN-Amplifikation

		1p-Anomalie		Gesamt
		nein	ja	
MYCN-Amplifikation	nein	80%	20%	100%
	ja	14%	86%	100%
Gesamt		66%	34%	100%

- Odds = Chance:
Quotient aus Risiko für ein Ereignis und Risiko für kein Ereignis
„Wahrscheinlichkeit geteilt durch Gegenwahrscheinlichkeit“)
- Chance einer 1p-Anomalie:
 - ohne MYCN-Amplifikation: $20\% / 80\% = 1 : 4 = 0,25$
 - mit MYCN-Amplifikation: $86\% / 14\% = 6,14$
- Ps.: Chance steigt genau dann, wenn Wahrscheinlichkeit steigt

MYCN-Amplifikation * 1p-Anomalie Kreuztabelle

% innerhalb von MYCN-Amplifikation

		1p-Anomalie		Gesamt
		nein	ja	
MYCN-Amplifikation	nein	80%	20%	100%
	ja	14%	86%	100%
Gesamt		66%	34%	100%

- **Odds Ratio:** Quotient der Chancen zwischen 2 Gruppen
- **Odds Ratio einer 1p-Anomalie**
von PatientInnen **ohne MYCN-Amplifikation** im Vergleich zu
PatientInnen **mit MYCN-Amplifikation**: $0,25 / 6,14 = 0,04$
- Sprechweisen: Das **Chance** einer **1p-Anomalie** ist
 - um **Faktor 0,04** niedriger,
 - um $(1 - 0,04) \cdot 100\% = 96\%$ gesenkt,wenn **MYCN nicht amplifiziert** ist (vs. **MYCN-Amplifikation**).

Interpretation der Risikomaße

- Relatives Risiko (RR) ist **anschaulicher** als Odds Ratio (OR)
- $RR \approx 1 \Leftrightarrow OR \approx 1$: kein Risikounterschied zwischen 2 Gruppen
- $RR > 1 \Leftrightarrow OR > 1$: Risikoerhöhung
- $RR < 1 \Leftrightarrow OR < 1$: Risikosenkung
- Bei Fall-Kontroll-Studien:
 - kein RR berechenbar
 - OR gültig
- Bei kleinen Risiken:
 - Risiko \approx Chance, da Wahrscheinlichkeit für kein Ereignis ≈ 1
 - $RR \approx OR$

Interpretation der Risikomaße

Bei Fall-Kontroll-Studien:

- kein RR berechenbar / OR aber gültig

	erkrankt	Kontrolle	
Biomarker vorhanden	5 (83%)	1	6 (100%)
Biomarker nicht vorhanden	5 (36%)	9	14 (100%)
	10	10	
	erkrankt	Kontrolle	
Biomarker vorhanden	5 (4.7%)	100	105 (100%)
Biomarker nicht vorhanden	5 (0.6%)	900	905 (100%)
	10	1000	

$$RR = 83/36 = 2,33$$

$$OR = (5/1)/(5/9) = 9$$

$$RR = 4.7/0.6 = 8,62$$

$$OR = (5/100)/(5/900) = 9$$

- Ggf. inkonsistente Resultate für RR bei Fall-Kontroll-Studien
(wobei RR für Spaltenprozente hier wieder konsistent wäre)
- Zeilenprozente hier nicht aussagekräftig

		kategorial		stetig
	binär	> 2 Kategorien	normalverteilt	nicht normalverteilt
binär	Kreuztabelle Zeilen/Spalten% RR, RD, OR	?	2 Boxplots / Histogramme 2*stetige Maßzahlen mit Mittelwert, Std.	2 Boxplots / Histogramme 2*stetige Maßzahlen ohne Mittelwert, Std.
kategorial	> 2 Kategorien			
stetig	normal-verteilt		Streudiagramm Korrelationskoeffizient nach Pearson r_p	Streudiagramm Korrelationskoeffizient nach Spearman r_s
	nicht norm.			

NB 2004 - Vergleich der Risikofaktoren

Folgende Fragen sollen geklärt werden:

1. Gibt es einen Zusammenhang zwischen LDH und Ferritin?
2. Hängen MYCN-Amplifikation und 1p-Anomalie zusammen?
3. Kommt die **MYCN**-Amplifikation in einigen **INSS**-Stadien häufiger vor?
4. Unterscheidet sich die Ferritin-Konzentration in den verschiedenen INSS-Stadien?

Zusammenhang zwischen INSS und MYCN

INSS-Stadium * MYCN-Amplifikation Kreuztabelle

		MYCN-Amplifikation		Gesamt
		nein	ja	
INSS-Stadium	1	304	9	313
		97,1%	2,9%	100,0%
		21,4%	1,9%	16,5%
	2	223	18	241
		92,5%	7,5%	100,0%
		15,7%	3,8%	12,7%
	3	320	142	462
		69,3%	30,7%	100,0%
		22,6%	29,8%	24,4%
	4	572	307	879
		65,1%	34,9%	100,0%
		40,3%	64,5%	46,4%
Gesamt		1419	476	1895
		74,9%	25,1%	100,0%
		100,0%	100,0%	100,0%

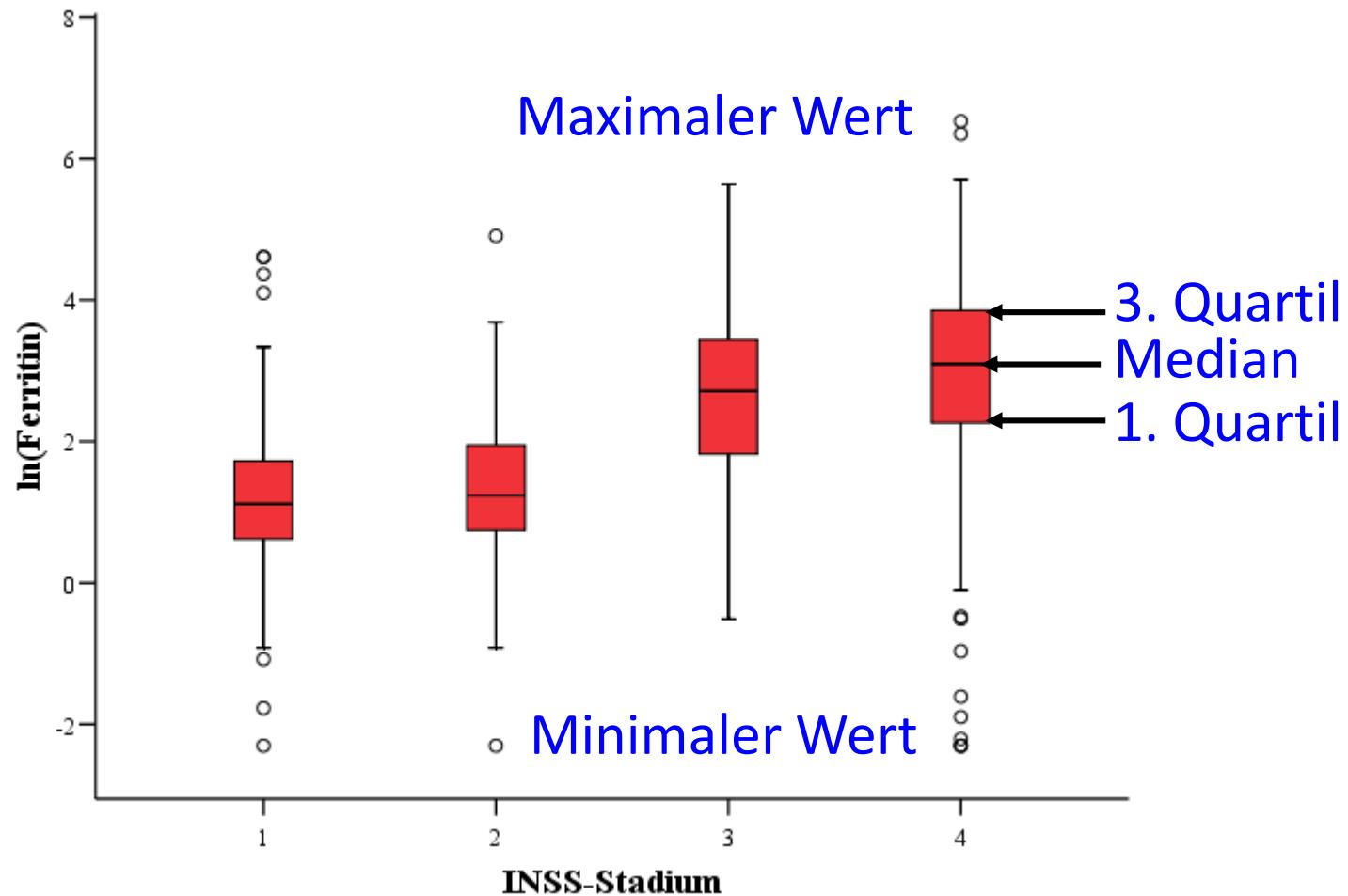
		kategorial		stetig
	binär	> 2 Kategorien	normalverteilt	nicht normalverteilt
binär	Kreuztabelle Zeilen/Spalten% RR, RD, OR	Kreuztabelle Zeilen% bzw. Spalten%	2 Boxplots / Histogramme 2*stetige Maßzahlen mit Mittelwert, Std.	2 Boxplots / Histogramme 2*stetige Maßzahlen ohne Mittelwert, Std.
Kategorial	> 2 Kategorien		?	?
stetig	normal-verteilt		Streudiagramm Korrelationskoeffizient nach Pearson r_p	Streudiagramm Korrelationskoeffizient nach Spearman r_s
	nicht norm.			

NB 2004 - Vergleich der Risikofaktoren

Folgende Fragen sollen geklärt werden:

1. Gibt es einen Zusammenhang zwischen LDH und Ferritin?
2. Hängen MYCN-Amplifikation und 1p-Anomalie zusammen?
3. Kommt die MYCN-Amplifikation in einigen INSS-Stadien häufiger vor?
4. Unterscheidet sich die **Ferritin**-Konzentration in den verschiedenen **INSS**-Stadien?

Zusammenhang zwischen INSS und ln(Ferritin)



Zusammenhang zwischen INSS und ln(Ferritin)

INSS-Stadium 1

INSS-Stadium 2

INSS-Stadium 3

INSS-Stadium 4

Statistiken^b

In(Ferritin)	
N	Gültig
	168
	Fehlend
	188
Mittelwert	1,142
Median	1,115
Modus	,6 ^a
Standardabweichung	1,0633
Varianz	1,131
Schiefe	,297
Standardfehler der Schiefe	,187
Spannweite	6,9
Minimum	-2,3
Maximum	4,6
Perzentile	25 ,610
	50 1,115
	75 1,723

Statistiken^a

In(Ferritin)	
N	Gültig
	112
	Fehlend
	172
Mittelwert	1,362
Median	1,238
Modus	,6
Standardabweichung	1,0419
Varianz	1,086
Schiefe	,232
Standardfehler der Schiefe	,228
Spannweite	7,2
Minimum	-2,3
Maximum	4,9
Perzentile	25 ,742
	50 1,238
	75 1,953

Statistiken^b

In(Ferritin)	
N	Gültig
	324
	Fehlend
	325
Mittelwert	2,569
Median	2,712
Modus	,1 ^a
Standardabweichung	1,2175
Varianz	1,482
Schiefe	-,273
Standardfehler der Schiefe	,135
Spannweite	6,1
Minimum	-,5
Maximum	5,6
Perzentile	25 1,817
	50 2,712
	75 3,439

Statistiken^a

In(Ferritin)	
N	Gültig
	514
	Fehlend
	721
Mittelwert	2,970
Median	3,096
Modus	2,3
Standardabweichung	1,2421
Varianz	1,543
Schiefe	-,986
Standardfehler der Schiefe	,108
Spannweite	8,8
Minimum	-2,3
Maximum	6,5
Perzentile	25 2,262
	50 3,096
	75 3,855

a. Mehrere Modi vorhanden. Der kleinste Wert wird angezeigt.

b. INSS-Stadium = 1

a. INSS-Stadium = 2

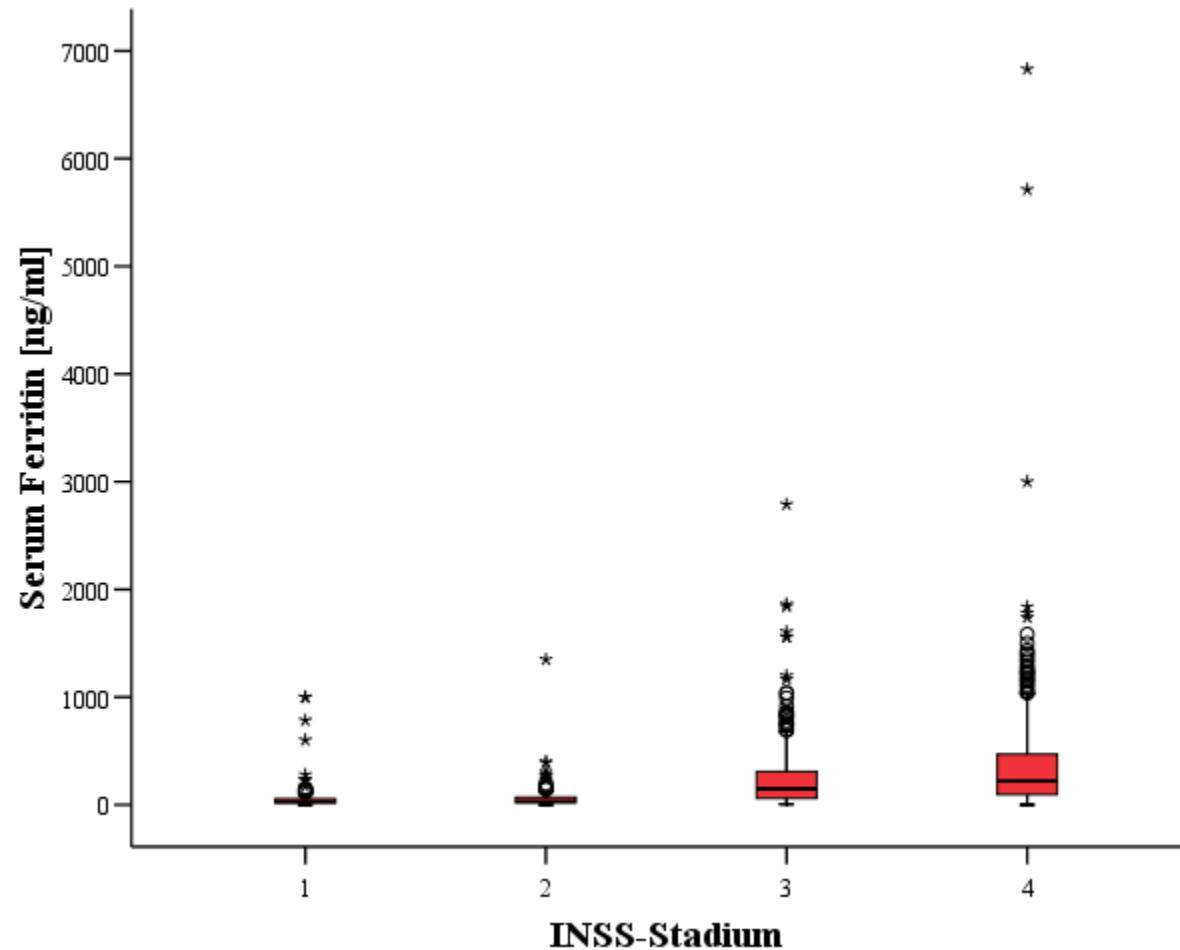
a. Mehrere Modi vorhanden. Der kleinste Wert wird angezeigt.

b. INSS-Stadium = 3

a. INSS-Stadium = 4

Mittelwert?, Standardabweichung?, Varianz?, Schiefe, ✓
 Median, Quartile, Minimum, Maximum, (Modus)

Zusammenhang zwischen INSS und Ferritin



Zusammenhang zwischen INSS und Serum Ferritin

INSS-Stadium 1 INSS-Stadium 2 INSS-Stadium 3 INSS-Stadium 4

Statistiken^b

Serum Ferritin [ng/ml]		
N	Gültig	168
	Fehlend	188
Mittelwert		61,70
Median		29,50
Modus		18 ^a
Standardabweichung		133,364
Varianz		17785,904
Schiefe		5,643
Standardfehler der Schiefe		,187
Spannweite		1000
Minimum		0
Maximum		1000
Perzentile	25	17,40
	50	29,50
	75	55,00

Statistiken^a

Serum Ferritin [ng/ml]		
N	Gültig	112
	Fehlend	172
Mittelwert		72,46
Median		33,50
Modus		18
Standardabweichung		143,142
Varianz		20489,624
Schiefe		6,832
Standardfehler der Schiefe		,228
Spannweite		1349
Minimum		0
Maximum		1349
Perzentile	25	20,00
	50	33,50
	75	69,50

Statistiken^b

Serum Ferritin [ng/ml]		
N	Gültig	324
	Fehlend	325
Mittelwert		248,01
Median		149,65
Modus		10 ^a
Standardabweichung		321,696
Varianz		103488,150
Schiefe		3,443
Standardfehler der Schiefe		,135
Spannweite		2785
Minimum		5
Maximum		2790
Perzentile	25	60,55
	50	149,65
	75	310,50

Statistiken^a

Serum Ferritin [ng/ml]		
N	Gültig	514
	Fehlend	721
Mittelwert		355,90
Median		220,00
Modus		95
Standardabweichung		506,381
Varianz		256421,323
Schiefe		7,076
Standardfehler der Schiefe		,108
Spannweite		6831
Minimum		0
Maximum		6831
Perzentile	25	95,00
	50	220,00
	75	471,25

a. Mehrere Modi vorhanden. Der kleinste Wert wird angezeigt.
b. INSS-Stadium = 1

a. INSS-Stadium = 2

a. Mehrere Modi vorhanden. Der kleinste Wert wird angezeigt.
b. INSS-Stadium = 3

a. INSS-Stadium = 4

~~Mittelwert?~~, ~~Standardabweichung?~~, ~~Varianz?~~, ~~Schiefe~~,
~~Median~~, ~~Quartile~~, ~~Minimum~~, ~~Maximum~~, ~~(Modus)~~

		kategorial		stetig
	binär	> 2 Kategorien	normalverteilt	nicht normalverteilt
binär	Kreuztabelle Zeilen/Spalten% RR, RD, OR	Kreuztabelle Zeilen% bzw. Spalten%	2 Boxplots / Histogramme 2*stetige Maßzahlen mit Mittelwert, Std.	2 Boxplots / Histogramme 2*stetige Maßzahlen ohne Mittelwert, Std.
Kategorial	> 2 Kategorien		<p>k Boxplots / Histogramme</p> <p>k*stetige Maßzahlen mit Mittelwert, Std.</p>	<p>k Boxplots / Histogramme</p> <p>k*stetige Maßzahlen ohne Mittelwert, Std.</p>
stetig	normal-verteilt		<p>Streudiagramm Korrelationskoeffizient nach Pearson r_p</p>	<p>Streudiagramm Korrelationskoeffizient nach Spearman r_s</p>
	nicht norm.			

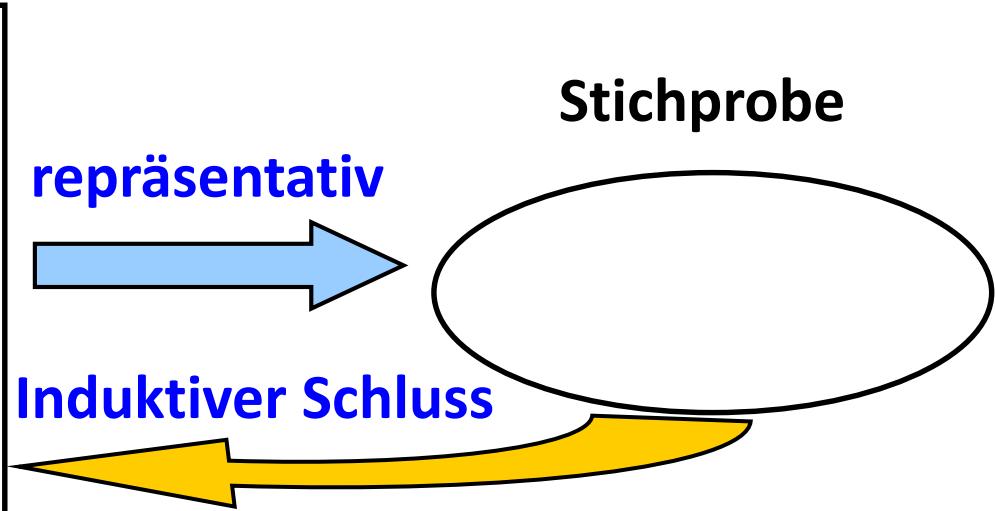
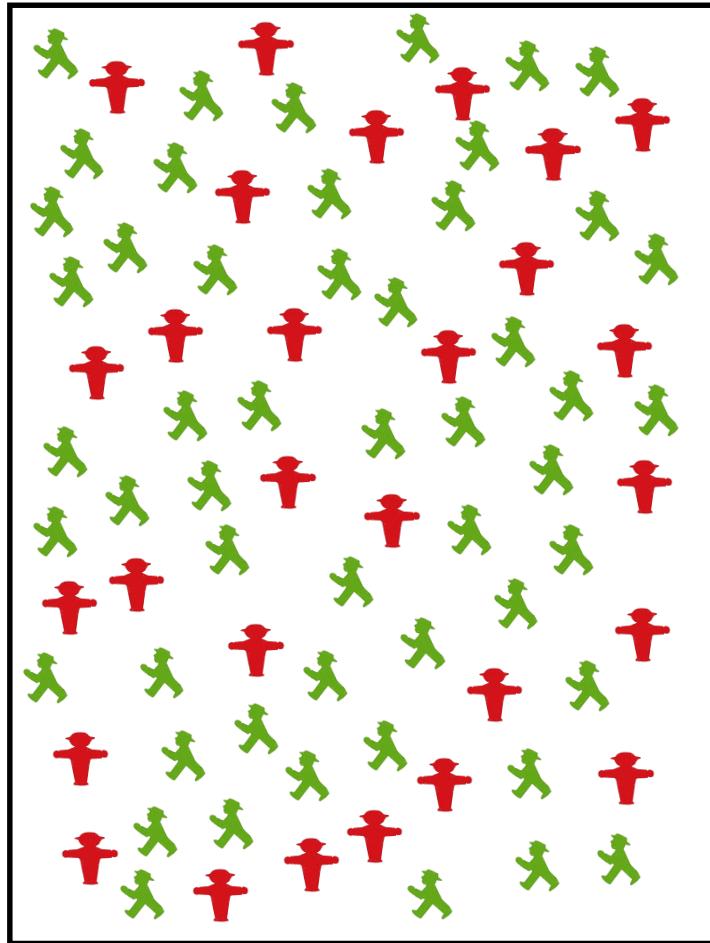
Vorlesungsinhalt

- Deskriptive Statistik
 - Darstellung 1es Merkmals
 - Zusammenhang 2er Merkmale
- **Statistisches Testen**
 - Prinzipien
 - Spezielle Tests
 - Multiples Testen
- Fallzahlplanung / Adaptive Designs
- Lineare Regression / logistische Regression
- Überlebenszeiten

Theorie statistisches Testen

Deskriptive Statistik vs. Konfirmatorische Statistik

Grundgesamtheit



Deskriptive Statistik:

Beschreibung der Stichprobe

Konfirmatorische Statistik:

Schlussfolgerung auf Grundgesamtheit

Wie kann man „beweisen“, dass Therapie A besser ist als B?

Problem:

- Stichprobe liefert nur Teilinformation
- Daher ist ein Beweis, also ein Nachweis unter Ausschluss jedweden Irrtums, nicht möglich.
- Dieser Irrtum soll daher quantifiziert / eingegrenzt werden (= die „nächstbeste“ Alternative zum Ausschluss jedweden Irrtums).

Dies gelingt mittels

- statistischer Tests
- Konfidenzintervallen

Beispiel: Statistische Tests (Signifikanztests)

Aussagekraft hängt vom Stichprobenumfang ab:

		Unerwünschtes Ereignis Adverse event, AE		
		Ja	Nein	
Therapie A	20 (20%)	80 (80%)	100 (100%)	
	10 (10%)	90 (90%)	100 (100%)	

**H0: Beide AE-Raten
sind identisch**

$$RR = 20/10 = 2, OR=2.25$$

$$p = 0.048$$

„Unter Therapie A ist die AE-Rate signifikant erhöht.“

		Unerwünschtes Ereignis Adverse event, AE		
		Ja	Nein	
Therapie A	10 (20%)	40 (80%)	50 (100%)	
	5 (10%)	45 (90%)	50 (100%)	

$$RR = 10/5 = 2, OR=2.25$$

$$p = 0.161$$

„Zwischen beiden Therapien besteht kein signifikanter Unterschied in der AE-Rate“

Beispiel: Statistische Tests (Signifikanztests)

Obige, zugehörige These (die nachgewiesen werden soll) ist von folgender Form:

Es gibt eine Assoziation / Zusammenhang zwischen

- dem Aufkommen eines Adverse-Events (= **Endpunkt / Zielgröße**)
- und der Therapie (= **Einflussgröße**) .

Bemerkung: Endpunkt / Zielgröße sollte so gewählt sein, dass dieser

- **relevant**
- **objektiv**

ist.

Konzept des statistischen Testens als „Widerspruchsbeweis“

In der Mathematik, folgende formale Vorgehensweise:

1. Stelle eine Behauptung / **These** auf, nenne diese **Behauptung „A“**
2. Nehme formal die Richtigkeit vom Gegenteil von „A“ an, **nehme an „Nicht A“ wäre richtig.**
3. Leite aus der Annahme von „Nicht A“ per Schlussfolgerungskette eine **widersprüchliche (also falsche) Aussage „W“** her.
4. Da „W“ falsch ist, muss die ursprüngliche Annahme „Nicht A“ falsch sein.
5. Folglich ist Behauptung „A“ richtig (und damit bewiesen).

Konzept des statistischen Testens als „Widerspruchsbeweis“

Beim **statistischen Testen** ist es ähnlich:

1. Stelle eine Behauptung / **These** auf, nenne diese Behauptung „**H1**“
2. Nehme formal die Richtigkeit vom **Gegenteil von „H1“** an, nenne dieses Gegenteil „**H0**“, und nehme an „**H0**“ wäre richtig.
3. Erhebe statistische Daten, und berechne die Wahrscheinlichkeit für das Zustandekommen dieser Daten unter Annahme „**H0**“. Diese Wahrscheinlichkeit wird als „**p-Wert**“ bezeichnet.
4. Ist diese Wahrscheinlichkeit (= p-Wert) „klein“, ist die Richtigkeit von „**H0**“ **unplausibel (anstatt „widersprüchlich“)**, die beobachteten Daten passen also nicht zu „**H0**“.
5. Folglich ist die Behauptung „**H1**“ plausibel.

Konzept des statistischen Testens als „Widerspruchsbeweis“

1. Stelle eine Behauptung / These auf, nenne diese Behauptung „H1“
 - **Beispiel:** Therapie A und Therapie B weisen tendenziell unterschiedlich viele Adverse Events auf.
2. Nehme formal die Richtigkeit vom Gegenteil von „H1“ an, nenne dieses Gegenteil „H0“, und nehme an „H0“ wäre richtig.
 - **Beispiel:** Nehme an, Therapie A und Therapie B weisen tendenziell gleich viele Adverse Events auf.

Konzept des statistischen Testens als „Widerspruchsbeweis“

3. Erhebe statistische Daten, und berechne die Wahrscheinlichkeit für das Zustandekommen dieser Daten unter Annahme „H0“.
 - Das Problem ist hier offensichtlich, die richtigen Daten zu erheben, den richtigen Endpunkt zu wählen, und die oben genannte Wahrscheinlichkeit zu berechnen ->
Wahrscheinlichkeitstheorie, Statistik, Testtheorie
4. Ist diese Wahrscheinlichkeit „klein“, ist die Richtigkeit von „H0“ unplausibel, da die beobachteten Daten zur Annahme „H0“ nicht passen.
 - Hier ist zu quantifizieren, was „klein“ bedeutet. Konventionell wird „klein“ als „kleiner als 5%“ festgelegt.

Beispiel – Binomialtest, einseitig



- Eine Impfung soll Infektionen verhindern, Erfolg (E := Impferfolg) oder Misserfolg (V := Impfversagen) sein.
- Stichprobe mit 10 ProbandInnen
 - Ausgang könnte beispielsweise sein (E,V,V,E,E,V,E,V,V)
- These (Alternativhypothese) H_1 : Erfolgsrate der Impfung ist $> 50\%$
 - Bei mehr als 50% Erfolgsrate (in der Grundgesamtheit) definiere Impfung als klinisch sinnvolle Maßnahme
- Nullhypothese H_0 : Erfolgsrate der Impfung = 50%

Beispiel – Binomialtest, einseitig

Mögliche 2^{10} Ergebnisse: (V,V,...,V), (E,V,...V), ..., (E,E,...,E)

- Definiere $\#E$:= Anzahl der Erfolge
- also bei (E,E,E,V,V,V,V,V,V) ist $\#E = 3$
- Unter H_0 sind alle 2^{10} Studienergebnisse gleich wahrscheinlich
- Also (unter H_0): Wahrscheinlichkeit für jedes Studienergebnis
 $= 1/2^{10} = 1/1024 \approx 0,001$

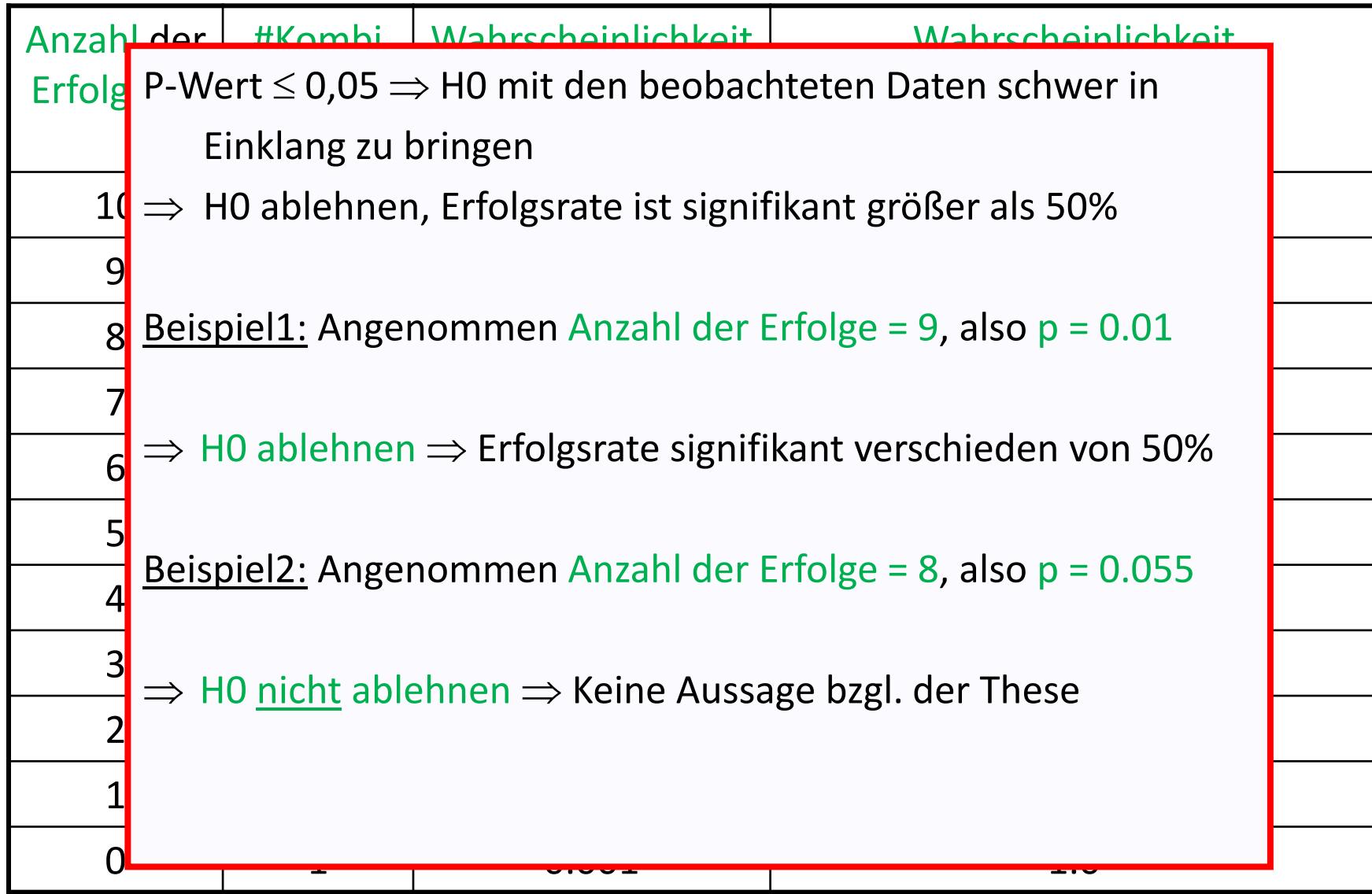
Aus der Kombinatorik: Anzahl von Kombinationen mit $\#E$ bei

$$\text{Stichprobe } n = 10: \binom{10}{\#E} := \frac{10 \cdot (10 - 1) \cdot \dots \cdot (10 - (\#E - 1))}{\#E \cdot (\#E - 1) \cdot \dots \cdot 2 \cdot 1}$$

Anzahl der Erfolge #E	#Kombi = $\binom{10}{#E}$	Wahrscheinlichkeit unter H_0 für genau #E Erfolge	Wahrscheinlichkeit unter H_0 für (mind.) #E
10	$\binom{10}{10} = 1$	1 * 0,001	1 * 0,001
9			
8			
7			
6			
5			
4			
3			
2			
1			
0			

Anzahl der Erfolge #E	#Kombi = $\binom{10}{#E}$	Wahrscheinlichkeit unter H_0 für genau #E Erfolge	Wahrscheinlichkeit unter H_0 für <u>mind.</u> #E
10	1	$1 * 0,001$	$1 * 0,001$
9	$\binom{10}{9} = 10$	$10 * 0,001$	$1 * 0,001 + 10 * 0,001 = 0,01$
8	(V,E,E,E,E,E,E,E,E)		
7	(E,V,E,E,E,E,E,E,E)		
6	(E,E,V,E,E,E,E,E,E)		
5	(E,E,E,V,E,E,E,E,E)		
4	(E,E,E,E,V,E,E,E,E)		
3	(E,E,E,E,E,V,E,E,E)		
2	(E,E,E,E,E,E,V,E,E)		
1	(E,E,E,E,E,E,E,V,E)		
0	(E,E,E,E,E,E,E,E,V)		

Anzahl der Erfolge #E	#Kombi = $\binom{10}{#E}$	Wahrscheinlichkeit unter H_0 für genau #E Erfolge	Wahrscheinlichkeit unter H_0 für <u>mind.</u> #E
10	1	0.001	0.001
9	10	0.01	0.01
8	45	0.044	0.055
7	120	0.117	0.172
6	210	0.205	0.377
5	252	0.246	0.623
4	210	0.205	0.828
3	120	0.117	0.945
2	45	0.044	0.989
1	10	0.01	0.999
0	1	0.001	1.0



Beispiel – Fisher-test, zweiseitig, Simulation

	Adverse event, AE		
	Ja	Nein	
Therapie A	40 (20%)	160 (80%)	200 (100%)
Therapie B	20 (10%)	180 (90%)	200 (100%)

← beobachtete Daten, OR = 2.25

Simulation1: OR = $(28/xx)/(32/yy)$

	Ja	Nein	
Therapie A	28	xx	200
Therapie B	32	yy	200

Teile 60 (= 40 + 20) AEs auf

1. Zufällig
2. Mit 50%-50% für ThA & ThB
(„=“ H₀ - Mechanismus)
3. Berechne OR

Simulation2: OR = $(33/xx)/(27/yy)$...

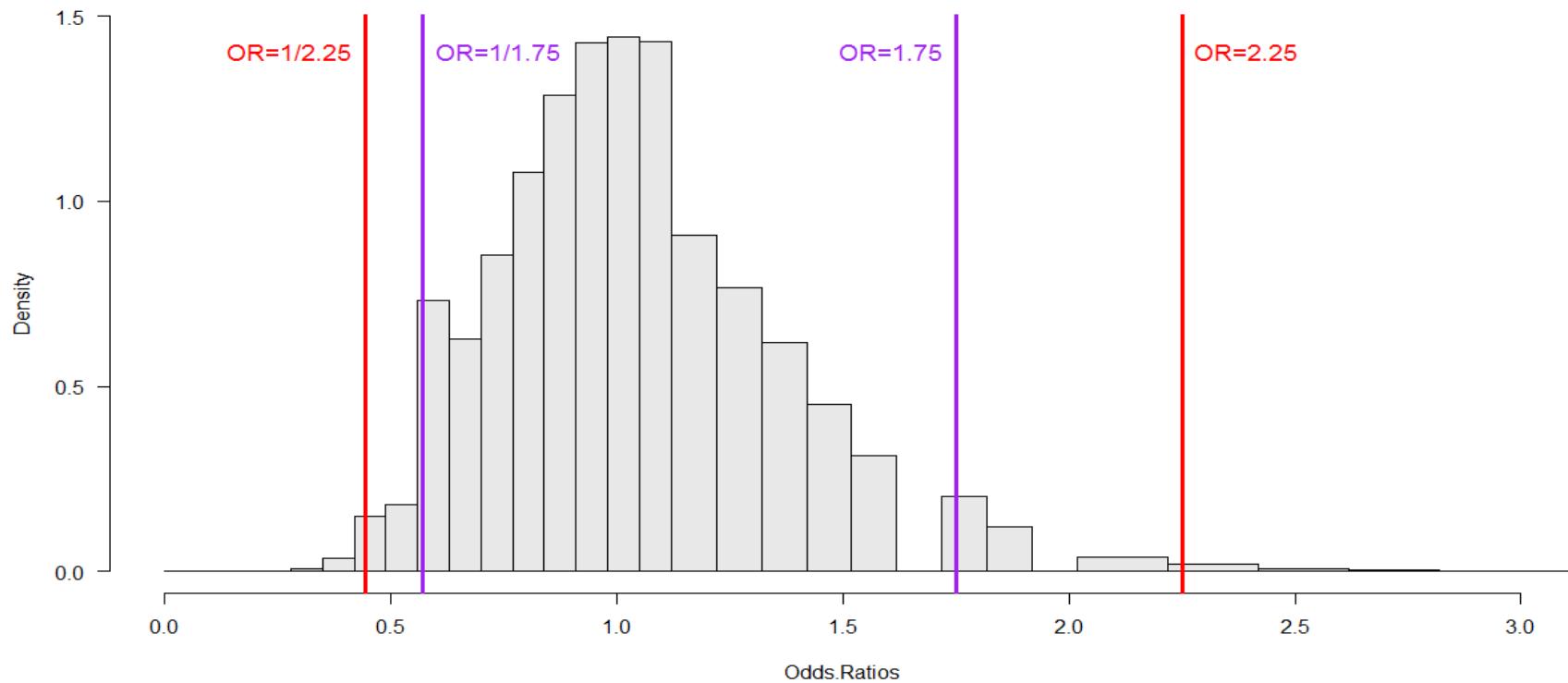
	Ja	Nein	
Therapie A	33	xx	200
Therapie B	27	yy	200

4. Wiederhole Schritte 1-3
tausende Male ...
5. Betrachte Verteilung der ORs

Beispiel – Fisher-test, zweiseitig, Simulation

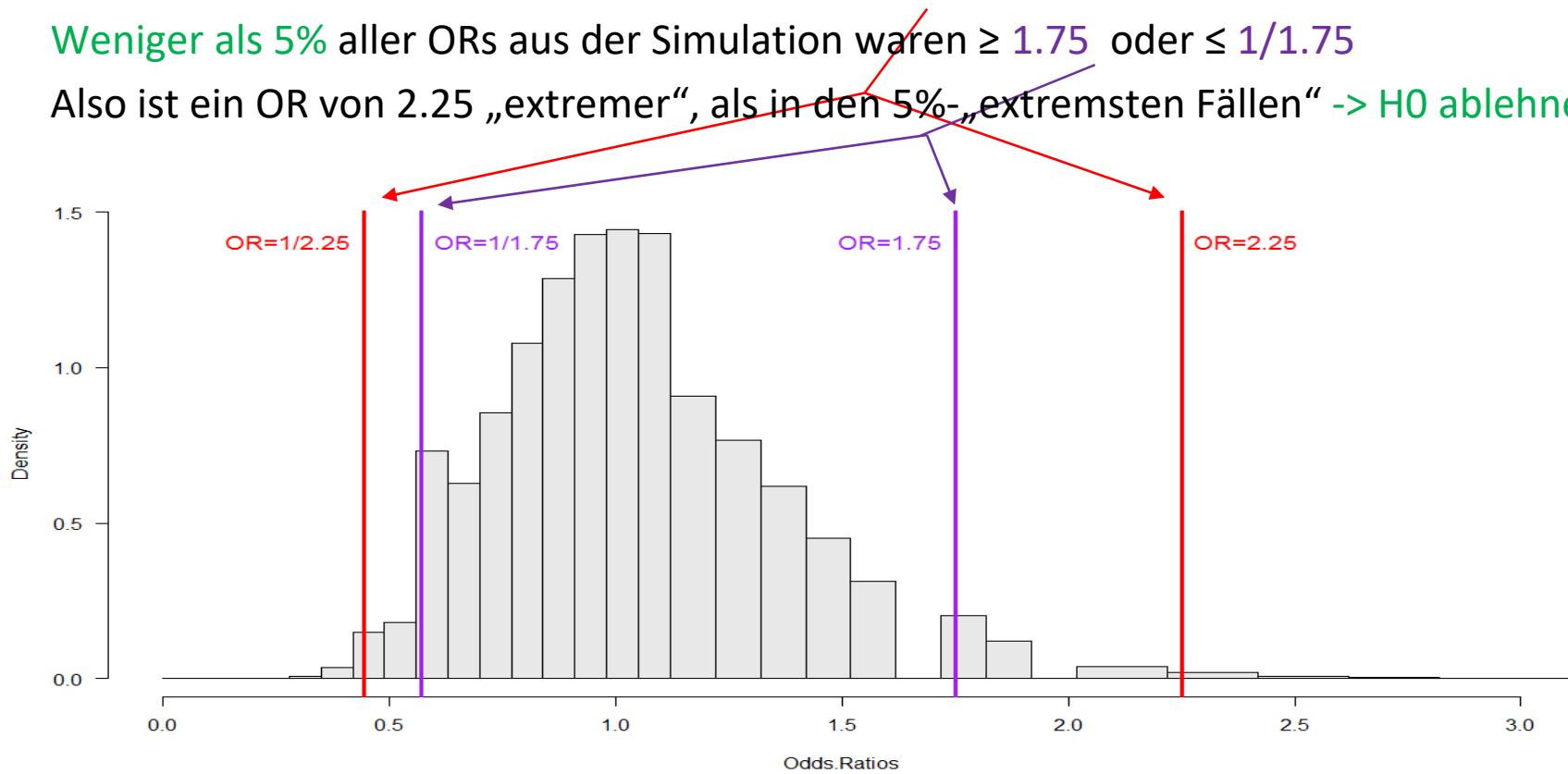
Obiges Verfahren 100 000 Mal wiederholt (mit einem Computerprogramm)

Histogramm der Odds-Ratios (ORs):



Beispiel – Fisher-test, zweiseitig, Simulation

- Beobachtetes OR = **2.25**
- „extremeres Ereignis“ bedeutet $OR \geq 2.25$ oder $OR \leq 1/2.25$ („weiter weg von 1“)
- **Weniger als 5%** aller ORs aus der Simulation waren ≥ 1.75 oder $\leq 1/1.75$
- Also ist ein OR von 2.25 „extremer“, als in den 5%-„extremsten Fällen“ -> H_0 ablehnen



Begriffe des statistischen Testens

- Nullhypothese H_0 vs. Alternativhypothese H_1
- Signifikanzniveau α
- Power
- P-Wert
- zweiseitiges Testproblem / einseitiges Testproblem
- Signifikanz

Statistische Hypothesen

- **Alternativhypothese** = H_1 =
die Vermutung zum Studieneffekt, die nachgewiesen werden soll = Gegenteil der Nullhypothese
- **Nullhypothese** = H_0 =
die Vermutung zum Studieneffekt, die widerlegt werden soll
- **Vorsicht:** Kläre genau ab, was „besser“ oder „überlegen“ bei einer Therapie bedeutet. **Beispiel:**
 - Pat A lebte länger als Pat B bei stärkeren leidvollen Beinträchtigungen.
 - Was ist oder bedeutet hier „besser“ ?

Typisierung der Fehler des statistischen Testens

Zu untersuchender Effekt	kein Effekt bei den Daten sichtbar	Effekt untermauert anhand der Daten
In Wirklichkeit gibt es keinen Effekt = H_0	Richtige Studienaussage	Fehler 1. Art
In Wirklichkeit gibt es einen Effekt = H_1	Fehler 2. Art	Richtige Studienaussage

Maximal tolerierte Wahrscheinlichkeit für einen Fehler 1. Art:
 α = Signifikanzniveau (= 5% üblich, aber auch 0,1%, 1% , 10%):

Typisierung der Fehler des statistischen Testens

Zu untersuchender Effekt	Entscheidung: H_0 beibehalten	Entscheidung für H_1
H_0 liegt vor	Richtige Studienaussage	Fehler 1. Art
H_1 liegt vor	Fehler 2. Art	Richtige Studienaussage

Signifikanzniveau – Interpretation und Wahl

- Signifikanzniveau = Niveau = α
Maximal zulässige Wahrscheinlichkeit, irrtümlich einen vermeintlichen Studieneffekt nachzuweisen, wenn in Wirklichkeit dieser Studieneffekt gar nicht existiert.
(falsche Entscheidung im Fall dass in Wirklichkeit kein Studieneffekt vorliegt)
- Übliche Wahl:
ICH-E9: “ $\alpha = 5\% \text{ or less ...}$ ”, aber Ausnahmen sind möglich.

Power – Interpretation und Wahl

- Power =
Wahrscheinlichkeit, Studieneffekt nachzuweisen, wenn der angenommene Studieneffekt auch zutrifft.
(richtige Entscheidung bei richtiger Vermutung)
- Power = $1-\beta$ (β = Wahrscheinlichkeit Fehler 2. Art)
- Übliche Wahl: $\geq 80\%$
- Die Power beeinflusst maßgeblich den notwendigen Stichprobenumfang, je mehr Power gewünscht, desto hoher der Stichprobenumfang

P-Wert

- Definition: Der p-Wert ist die Wahrscheinlichkeit, das beobachtete Studienergebnis (oder extremer) zu erhalten, obwohl die Nullhypothese H_0 stimmt.
- Ergebnis eines statistischen Tests (komplizierte Berechnung)
- $0 \leq p\text{-Wert} \leq 1$
- Vorsicht: p-Wert \neq Wahrscheinlichkeit der Nullhypothese H_0
 $p\text{-Wert} = P(\text{Daten} \mid H_0) \neq P(H_0 \mid \text{Daten})$
- Trotzdem oft fälschlicherweise als Maß für die Plausibilität von H_0 benutzt

P-Wert

Kriterium zur Entscheidung für oder gegen H_0

- p-Wert **klein**
 - ⇒ Studienergebnis mit H_0 schwer vereinbar
 - ⇒ H_0 ablehnen
 - p-Wert **groß**
 - ⇒ H_0 beibehalten
- Aber bei großem p-Wert: **H_0 nicht bewiesen**

Mit einem p-Wert lässt sich H_0 widerlegen, aber nicht beweisen.

Arbeitsschritte beim statistischen Testen

1. Fragestellung / These (**H1**) aufstellen
 - Zielgröße / Endpunkt aufstellen (z. B. Impferfolg ja / nein)
 - ggf. Einflussgröße (z. B. geimpft mit Präparat A oder B)
2. Nullhypothese (**H0** aufstellen) = Negation von H1
3. Wähle aus: **Test** und **Signifikanzniveau α** (=5% in der Regel)
 - Beim Impfbeispiel wäre der Binomialtest ein entsprechend korrekter Test
4. Wähle **Signifikanzniveau α** (=5% in der Regel)
5. Datenerhebung
6. $p\text{-Wert} \leq \text{Signifikanzniveau } \alpha$? → **Testentscheidung**

Arbeitsschritte beim statistischen Testen

6. $p\text{-Wert} \leq \text{Signifikanzniveau } \alpha$? → **Testentscheidung**

Falls $p\text{-Wert} \leq \alpha$:

- H_0 ablehnen und damit H_1 annehmen
- Das Ergebnis ist dann **statistisch signifikant**.
- Das Studienergebnis mit H_0 kaum vereinbar

Falls $p\text{-Wert} > \alpha$:

- H_0 beibehalten, H_1 (noch) nicht annehmen
- Das Ergebnis ist dann **nicht statistisch signifikant**.
- Das Studienergebnis widerspricht nicht H_0

Ein- und zweiseitige Nullhypothesen H0

- Nullhypothese 1: Therapie B (innovativ) führt nicht zu einer **geringeren Rate** von ‚adverse events‘ als Therapie A.
Diese Nullhypothese ist **einseitig**, weil eine Richtung vorgegeben wird.
- Nullhypothese 2: Therapie B (innovativ) führt nicht zu einer **anderen Rate** von ‚adverse events‘ als Therapie A.
Diese Nullhypothese ist **zweiseitig**, weil keine Richtung vorgegeben wird.
- **Standard** = zweiseitige Hypothese
- **Einseitige Hypothese \Rightarrow ICH E9 α halbieren:** z.B. 2,5% statt 5%

Ein- und zweiseitige Nullhypothesen H_0

Formal betrachtet:

- **Zweiseitiges** Testproblem

H_0 : Heilungsrate Therapie A = Heilungsrate Therapie B
(„kein Effekt“)

VS.

H_1 : Heilungsrate Therapie A \neq Heilungsrate Therapie B

- **Einseitiges** Testproblem

H_0 : Heilungsrate Therapie A \leq Heilungsrate Therapie B
(„kein Effekt oder A schlechter als B“)

VS.

H_1 : Heilungsrate Therapie A $>$ Heilungsrate Therapie B

Fallstricke beim statistischen Testen

!Achtung!

- p-Wert = „Maß der Plausibilität der Daten unter H_0 “
= $P(\text{Daten} \mid H_0) \neq P(H_0 \mid \text{Daten})$
 - D. h. p-Wert ist nicht die Wahrscheinlichkeit für die Gültigkeit von H_0
- H_0 kann nicht „bewiesen“ werden, nur „nicht ablehnt“ (bei $p>\alpha$)
 - D. h. ein induktiver Schluss ist nur zugunsten der Alternativhypothese H_1 möglich
- Statistisch signifikant sagt aus, dass ein Studieneffekt vorhanden aber nicht wie groß dieser ist, oder ob dieser klinisch relevant ist
 - Schätzung der Größe des Effekts, am besten mittels Konfidenzintervall
 - klinische Relevanz muss der Kliniker entscheiden!

Signifikanz versus klinische Relevanz

- Statistische Signifikanz \neq medizinische Relevanz / klinische Relevanz
- Statistische Signifikanz
 - beruht auf einem großen Studieneffekt oder
 - wurde durch eine große Fallzahl belegt
- Medizinische Relevanz
 - bewertet die Größe des beobachteten Studieneffektes
 - Beurteilung dessen ist eine klinische Fragestellung

Konfidenzintervalle

Konfidenzintervall – Parameterschätzung in der Inferenzstatistik

Aufgabe: **Schätzung des wahren Effektes** (Parameterschätzer)

Vorgehen bei Parameterschätzungen:

1. Kläre, **welcher Parameter** geschätzt werden soll
(Erfolgsrate oder Erwartungswert, oder ...)
2. Berechne die Schätzung, wie z. B. eine relative Rate
(**Punktschätzer**)
3. Berechne einen Bereich, im welchen der wahre Parameter mit hoher Wahrscheinlichkeit liegt (**Konfidenzintervall**)

Konfidenzintervall – Parameterschätzung in der Inferenzstatistik

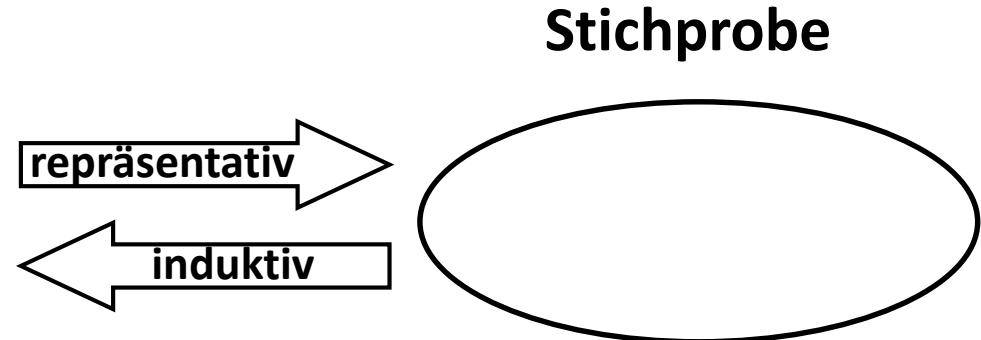
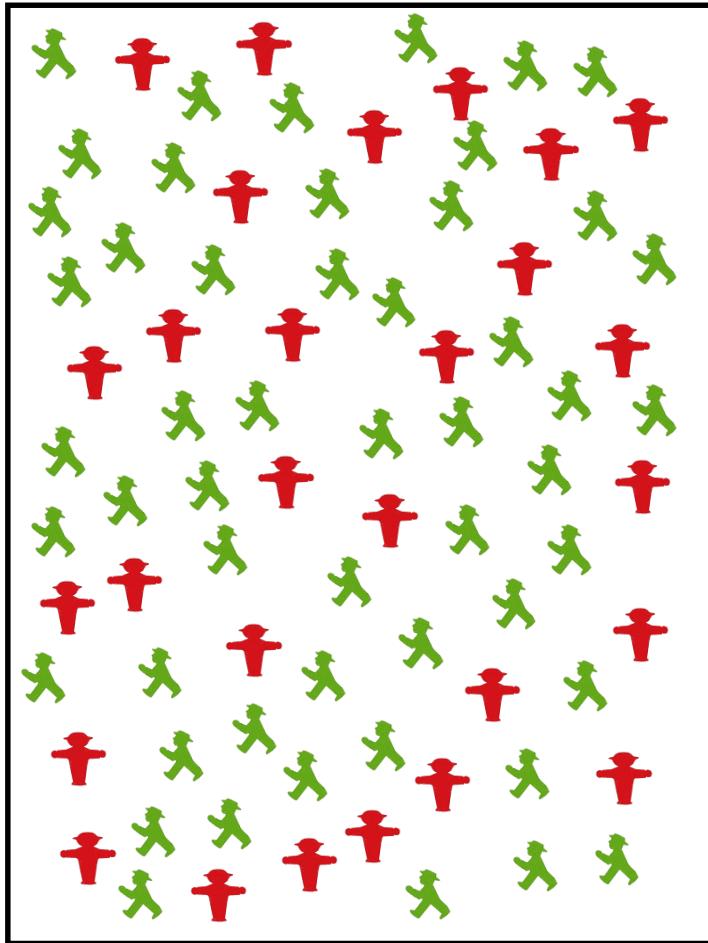
Konfidenzintervalle werden z. B. verwendet, um **Signifikanz** und **klinische Relevanz kombiniert** darzustellen.

Konfidenzintervall

- **Konfidenzintervall:** Bereich/Intervall (möglichst klein), der den **wahren Parameter** mit **hoher Wahrscheinlichkeit** (in der Regel 95%) enthält
- **Definition:** Intervall $[U, O]$ heißt **$(1-\alpha)$ -Konfidenzintervall** für Parameter $\theta \Leftrightarrow P_\theta(U \leq \theta \leq O) \geq 1-\alpha$
 - Parameter θ könnte beispielsweise der prozentuale Frauenanteil in einer Patientenpopulation sein.
- **Interpretation:** $(1-\alpha)$ -Konfidenzintervall = $(1-\alpha)$ -KI = Bereich um Effektschätzer der Studie, welcher den wahren Wert mindestens mit Wahrscheinlichkeit $1-\alpha$ enthält

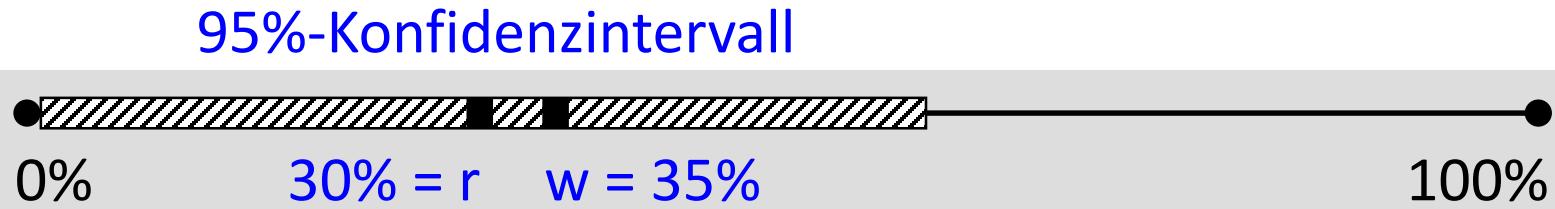
Beispiel: Konfidenzintervall

Grundgesamtheit



Gesucht: Geschlechteranteil (w) in
Grundgesamtheit: $28/80 = 35\%$
Stichprobe: relative Häufigkeit = $3/10$
Schätzer (r = relative Rate): 30%
-> (komplizierte) Formel ->
95%-Konfidenzintervall: $[2\%; 58\%]$

Beispiel: Konfidenzintervalle



Gesucht: Geschlechteranteil (w)  in
Grundgesamtheit: $28/80 = 35\%$
Stichprobe: relative Häufigkeit = $3/10$
Schätzer (r = relative Rate): 30%
-> (komplizierte) Formel ->
95%-Konfidenzintervall: $[2\%; 58\%]$

Spezielle Konfidenzintervalle

Konfidenzintervalle existieren für folgende Punktschätzer:

- Relative Häufigkeit
- Differenz relativer Häufigkeiten
- Relatives Risiko
- Odds Ratio
- Mittelwert
- Differenzen von Mittelwerten
- Median
- Varianz
- ...

Konfidenzintervall binärer Merkmale - Berechnungsalgorithmus

- Sei X eine **binäre Zufallsvariable** (Merkmal) mit **Ausprägung ja** oder **nein**
- w = wahre Wahrscheinlichkeit für $X = \text{ja}$ (= wahrer Parameter)
- r = **beobachtete, relative Häufigkeit** für $X = \text{ja}$ in der Stichprobe
- n = Stichprobenumfang
- **95%-KI** (Konfidenzintervall) für w

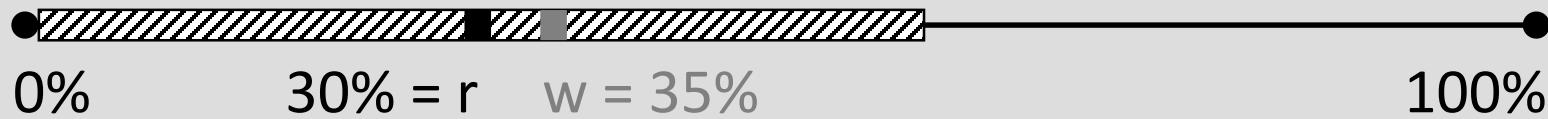
$$\left[r - 1.96 \cdot \left(\frac{r(1-r)}{n} \right)^{1/2}, r + 1.96 \cdot \left(\frac{r(1-r)}{n} \right)^{1/2} \right] = r \pm 1.96 \cdot \left(\frac{r(1-r)}{n} \right)^{1/2}$$

- $|95\%-KI| \leq 2/\sqrt{n}$; $n \uparrow \Rightarrow |95\%-KI| \downarrow$; $n \cdot 100 \Rightarrow |95\%-KI| / 10$
- Gültig falls $r \cdot (1-r) \cdot n \geq 9$, z. B. $r = 10\%$, $n = 100$, $|95\%-KI| = 12\%$

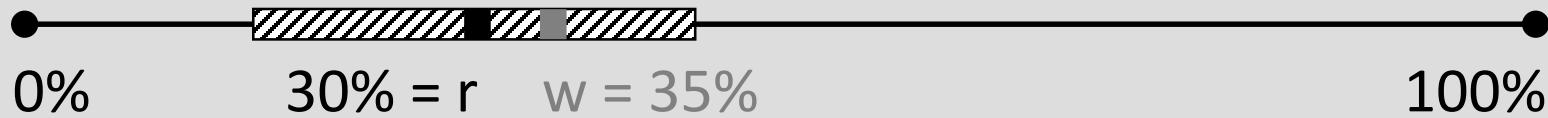
$1-\alpha \downarrow$ oder $|Fallzahl| \uparrow \Rightarrow |(1-\alpha)\%-KI| \downarrow$

95%-Konfidenzintervall

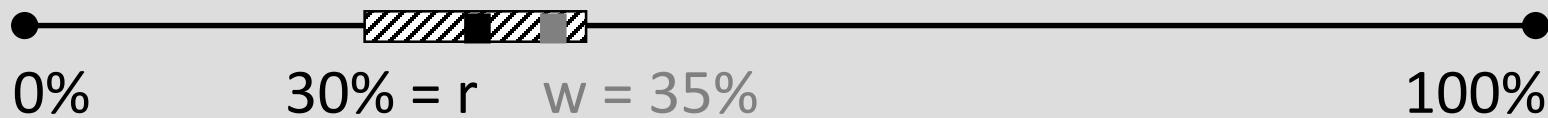
$n = 10$



$n = 40$



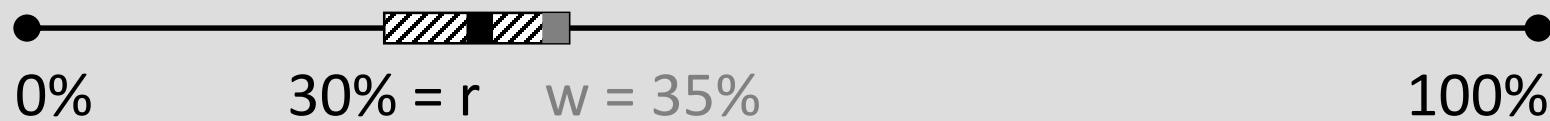
$n = 160$



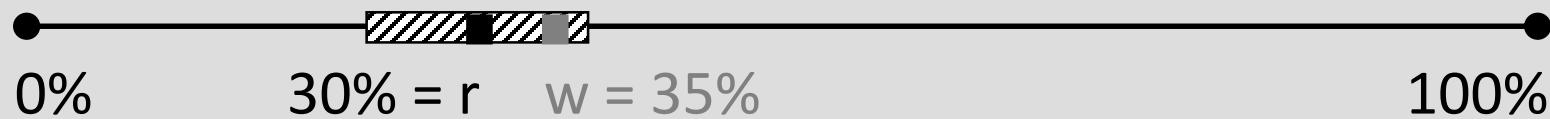
$1-\alpha \downarrow$ oder $| \text{Fallzahl} | \uparrow \Rightarrow |(1-\alpha)\%-KI| \downarrow$

$n = 160$

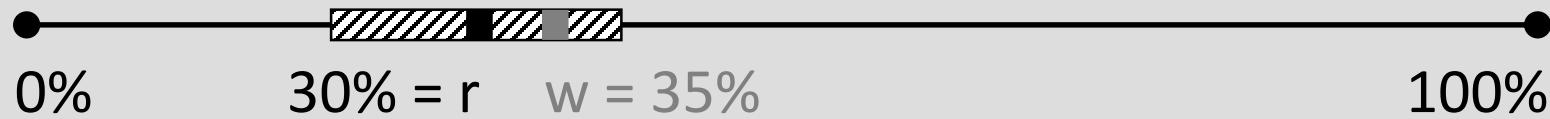
90%-Konfidenzintervall



95%-Konfidenzintervall



99%-Konfidenzintervall



Konfidenzintervall vs. Statistischer Test

- $H_0: \theta = \theta_0$ (θ_0 = zu widerlegender Wert, θ = interessierender Parameter)
- $\theta_0 \notin (1-\alpha)\text{-KI} \Leftrightarrow H_0 \text{ ablehnen} \Leftrightarrow p\text{-Wert} \leq \alpha$
- Beispiel 1: Wahrscheinlichkeit eines Bronchialkarzinoms (BC)
 - $H_0: BC \text{ bei Rauchern} = BC \text{ bei Nichtrauchern}$ (d.h. Relatives Risiko = 1)
 - $(1-\alpha)\text{-KI}$ des **Relativen Risikos** für BC (Raucher vs. Nichtraucher)
 - $1 \notin (1-\alpha)\text{-KI} \Leftrightarrow \text{sign. Risikounterschied zum Niveau } \alpha$



Vorlesungsinhalt

- Deskriptive Statistik
 - Darstellung 1es Merkmals
 - Zusammenhang 2er Merkmale
- **Statistisches Testen**
 - Prinzipien
 - **Spezielle Tests**
 - Multiples Testen
- Fallzahlplanung / Adaptive Designs
- Lineare Regression / logistische Regression
- Überlebenszeiten

Spezielle statistische Tests

Fall 1: Zusammenhang zwischen INSS und MYCN

INSS-Stadium * MYCN-Amplifikation Kreuztabelle

INSS-Stadium		MYCN-Amplifikation		Gesamt
		nein	ja	
		Anzahl	% von INSS-Stadium	
1	Anzahl	304	9,9%	313
	% von INSS-Stadium	97,1%	2,9%	100,0%
	% von MYCN-Amplifikation	21,4%	1,9%	16,5%
2	Anzahl	223	18	241
	% von INSS-Stadium	92,5%	7,5%	100,0%
	% von MYCN-Amplifikation	15,7%	3,8%	12,7%
3	Anzahl	320	142	462
	% von INSS-Stadium	69,3%	30,7%	100,0%
	% von MYCN-Amplifikation	22,6%	29,8%	24,4%
4	Anzahl	572	307	879
	% von INSS-Stadium	65,1%	34,9%	100,0%
	% von MYCN-Amplifikation	40,3%	64,5%	46,4%
Gesamt		1419	476	1895
		74,9%	25,1%	100,0%
		100,0%	100,0%	100,0%

Test auf Zusammenhang zweier kategorialer Größen

- **Frage:** Gibt es einen **Zusammenhang** zwischen dem INSS-Stadium (=Einflussgröße) und der MYCN-Amplifikation (=Zielgröße)?
- **Nullhypothese:** Es gibt **keinen Zusammenhang** zwischen dem INSS-Stadium und der MYCN-Amplifikation.
- **Alternativhypothese:** Es gibt **einen Zusammenhang** zwischen dem INSS-Stadium und der MYCN-Amplifikation.
- **Signifikanzniveau:** z.B. $\alpha=0,05$
- **Statistischer Test:** **Chiquadrat-Unabhängigkeitstest** = χ^2 -Test
- **Zweiseitiger P-Wert:**
SPSS-Ausgabe: $p = 0,000 \Rightarrow p < 0,001$

Test auf Zusammenhang zweier kategorialer Größen – ordentliche Interpretation

- Sachaussage:
 - Studienkollektiv (sollte beschrieben werden)
 - Signifikanzniveau (sollte genannt werden)
 - Signifikant: ja oder nein
 - Merkmale / Endpunkte (sollten genannt werden)
 - Falls signifikant: **Richtung** der Aussage (bei 2-seitigem Test)

Angenommen, P-Wert des χ^2 -Tests < 0,001

- „Bei **Kindern** und **Jugendlichen** mit einem **Neuroblastom** besteht zum **5%-Niveau** ein **signifikanter Zusammenhang** zwischen **INSS-Stadium** und **MYCN-Amplifikation**. Patienten in **höheren Stadien** haben **häufiger** eine **MYCN-Amplifikation**.“

Test auf Zusammenhang zweier kategorialer Größen (unverbunden)

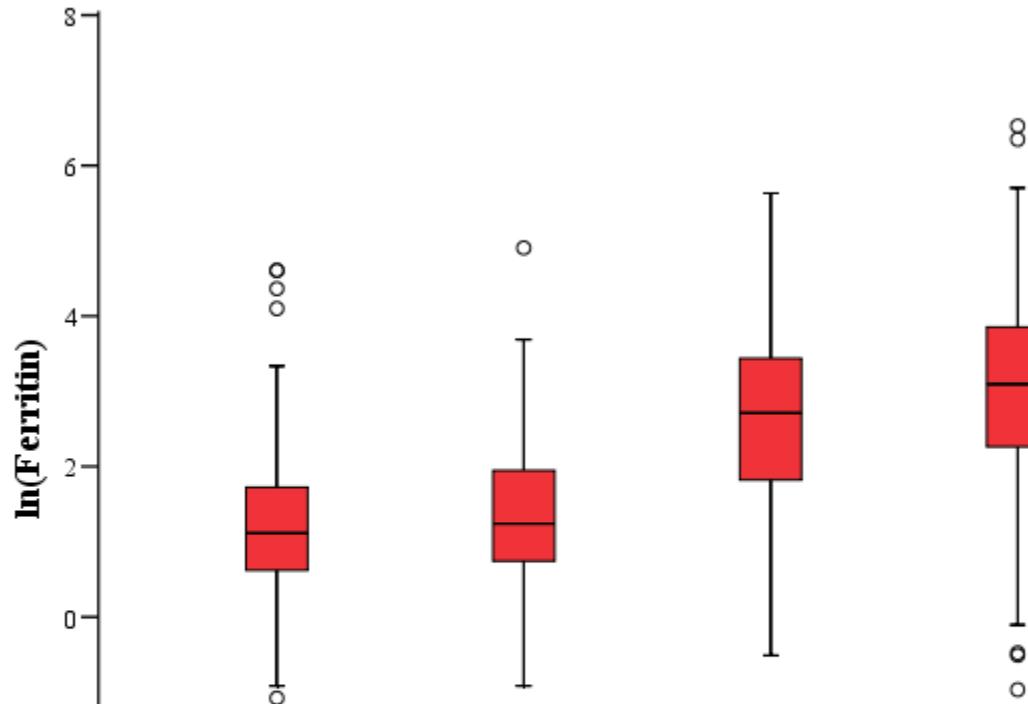
Chiquadrat-Unabhängigkeitstest Test

- Voraussetzungen: Gesamtfallzahl: $n \geq 60$
Minimale erwartete Zellhäufigkeit ≥ 5 (SPSS-Output)
- Beispiel: $n = 1895$ und min. erwartete Zellhäufigkeit = 60
- Falls die Voraussetzungen nicht erfüllt sind:
 - Kategorien sinnvoll zusammenfassen
Falls Voraussetzungen jetzt erfüllt sind: χ^2 -Test
 - Vorsicht: Bei konfirmatorischem Test datenabhängige Entscheidung über die Zusammenfassung nicht möglich!

Exakter Fisher-Test

- benötigt höhere Rechenkapazität

Fall 2: Zusammenhang zwischen INSS und $\ln(\text{Ferritin})$



- $|\text{Schiefe}|$ von $\ln(\text{Ferritin}) \leq 1$ in allen INSS-Stadien
- Histogramm: eingipflig in allen INSS-Stadien
- Normalverteilung in allen INSS-Stadien erscheint plausibel

Assoziation zwischen einer normalverteilten Größe und mehreren Gruppen

- **Frage:** Unterscheidet sich die logarithmierte Ferritin-Konzentration in den verschiedenen INSS-Stadien?
- **Nullhypothese:** Die logarithmierte Ferritin-Konzentration in den verschiedenen INSS-Stadien unterscheidet sich nicht.
- **Alternativhypothese:** Die logarithmierte Ferritin-Konzentration in den verschiedenen INSS-Stadien unterscheidet sich.
- **Signifikanzniveau:** z.B. $\alpha=0,05$
- **2 Gruppen:** t-Test für unverbundene Stichproben nach Welch
- **> 2 Gruppen:** Einfaktorielle ANalysis Of Variance
- **Voraussetzung:** Normalverteilte stetige Größe in allen Gruppen

Assoziation zwischen einer normalverteilten Größe und mehreren Gruppen – ordentliche Interpretation

- Sachaussage:
 - Studienkollektiv (sollte beschrieben werden)
 - Signifikanzniveau (sollte genannt werden)
 - Signifikant: ja oder nein
 - Merkmale / Endpunkte (sollten genannt werden)
 - Falls signifikant: **Richtung** der Aussage (bei 2-seitigem Test)

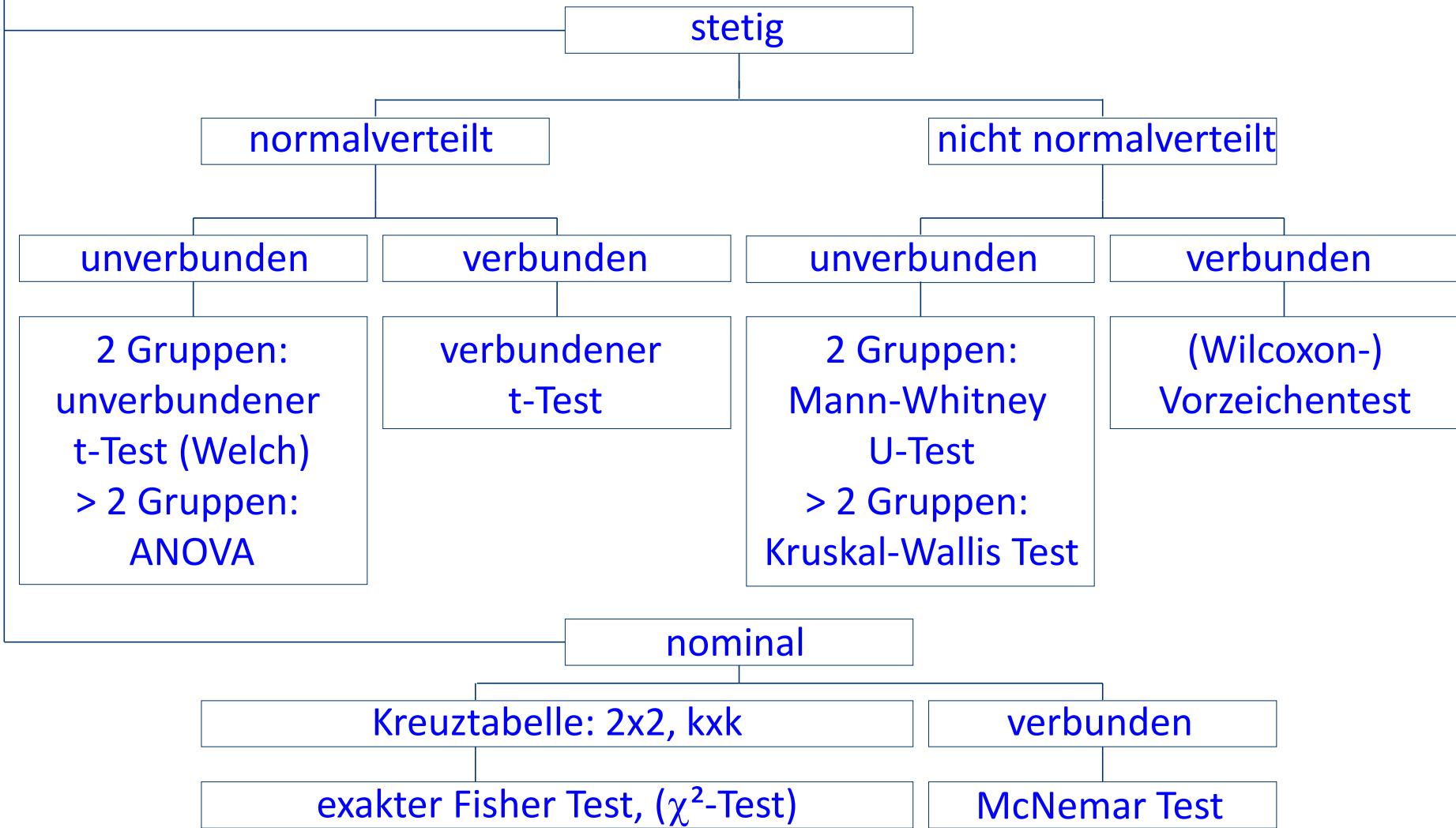
Angenommen, P-Wert der einfaktoriellen ANOVA $< 0,001$

- „Die **logarithmierte Ferritin-Konzentration** in den verschiedenen **INSS-Stadien unterscheidet** sich bei Kindern und Jugendlichen mit einem **5%-Niveau signifikant**. Mit **steigenden Stadien wächst die Konzentration**.“

Weitere statistische Tests

- Es gibt sehr **viele verschiedene Tests** (t-Test, Wilcoxon-Test, Fisher-Test, χ^2 -Test, Logrank-Test, Likelihood-Ratio Test, ...)
- **Transparenz:** Tests genau im Protokoll spezifizieren
- Auch **statistische Programme** im Protokoll spezifizieren
- Wahl des richtigen Tests hängt ab von
 - Zielgröße: **Stetig** (z.B. Ferritin) oder **nominal** (z.B. MYCN)
 - Studiendesign: **unverbunden** (mehrere Gruppen) oder **verbunden** (Paare von Beobachtungen)
 - **Zahl** der **Gruppen** (2 oder > 2)
 - **Methode:** **Normalverteilt** oder **nicht normalverteilt**

Zielgröße



Zielgröße

stetig

Es gibt weitere ztausende Tests

Holen Sie sich den Rat Ihres Biometrikers!

unverbunden

verbunden

unverbunden

verbunden

2 Gruppen:
unverbundener
t-Test (Welch)
> 2 Gruppen:
ANOVA

verbundener
t-Test

2 Gruppen:
Mann-Whitney
U-Test
> 2 Gruppen:
Kruskal-Wallis Test

(Wilcoxon-)
Vorzeichentest

nominal

Kreuztabelle: 2x2, kxk

exakter Fisher Test, (χ^2 -Test)

verbunden

McNemar Test

Vorlesungsinhalt

- Deskriptive Statistik
 - Darstellung 1es Merkmals
 - Zusammenhang 2er Merkmale
- Statistisches Testen
 - Prinzipien
 - Spezielle Tests
 - Multiples Testen
- Fallzahlplanung / Adaptive Designs
- Lineare Regression / logistische Regression
- Überlebenszeiten



Einführung Überlebenszeitanalyse

Beispiele für „Überlebenszeit“

- **Gesamtüberlebenszeit**

Zeit bis Ereignis = Tod jedweder Art

- **Progressionsfreie Überlebenszeit**

Zeit bis Ereignis = Progression oder Tod jedweder Art

- **Ereignisfreie Überlebenszeit**

Zeit bis Ereignis = Therapiezwischenfall oder Progression oder Tod

- **Krankheitsdauer**

Zeit bis Ereignis = Heilung bzw. Remission

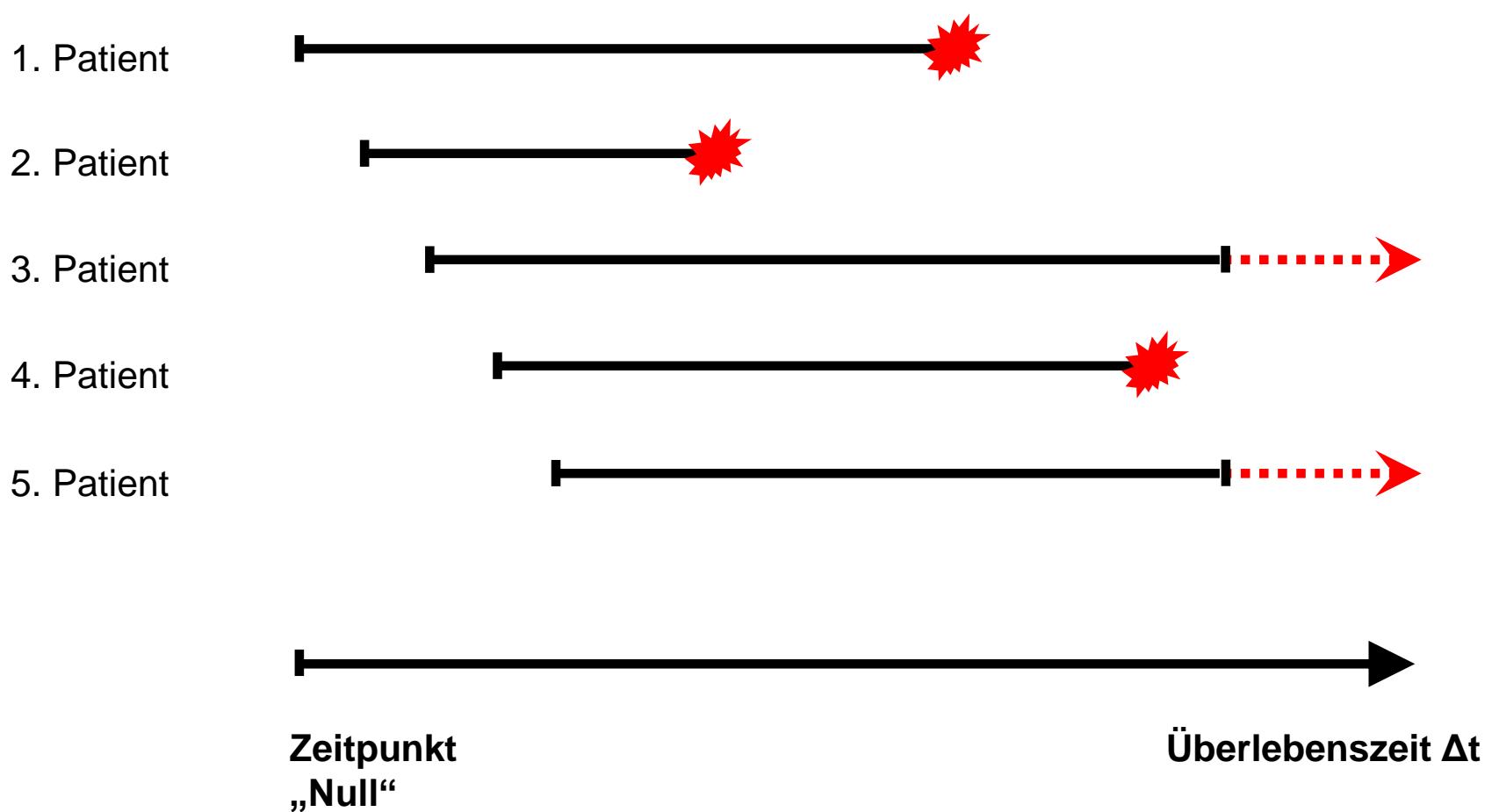
- **Gehstrecke**

Distanz bis Ereignis = Gehstopp durch Claudicatio intermittens

Definition von Überlebenszeit

- Definiere „Anfang“: Zeitpunkt / Startpunkt
(z. B. Diagnose, Aufnahme in die Studie, Randomisierung, Therapiestart, Startpunkt der Laufstrecke, ...)
- Definiere Ereignis (z. B. Progression oder Tod)
- Es gibt 2 Arten von Beobachtungen („beobachtet“ und „zensiert“):
 - a. Ereignis wird beobachtet \Rightarrow Dauer von Startzeitpunkt bis Ereignis bekannt
 - b. Ereignis wird nicht beobachtet \Rightarrow Dauer von Startzeitpunkt bis letzte Beobachtung ist bekannt
= (rechts-)zensierte Überlebenszeiten

Zensierte Ereigniszeiten



Kaplan-Meier Schätzung

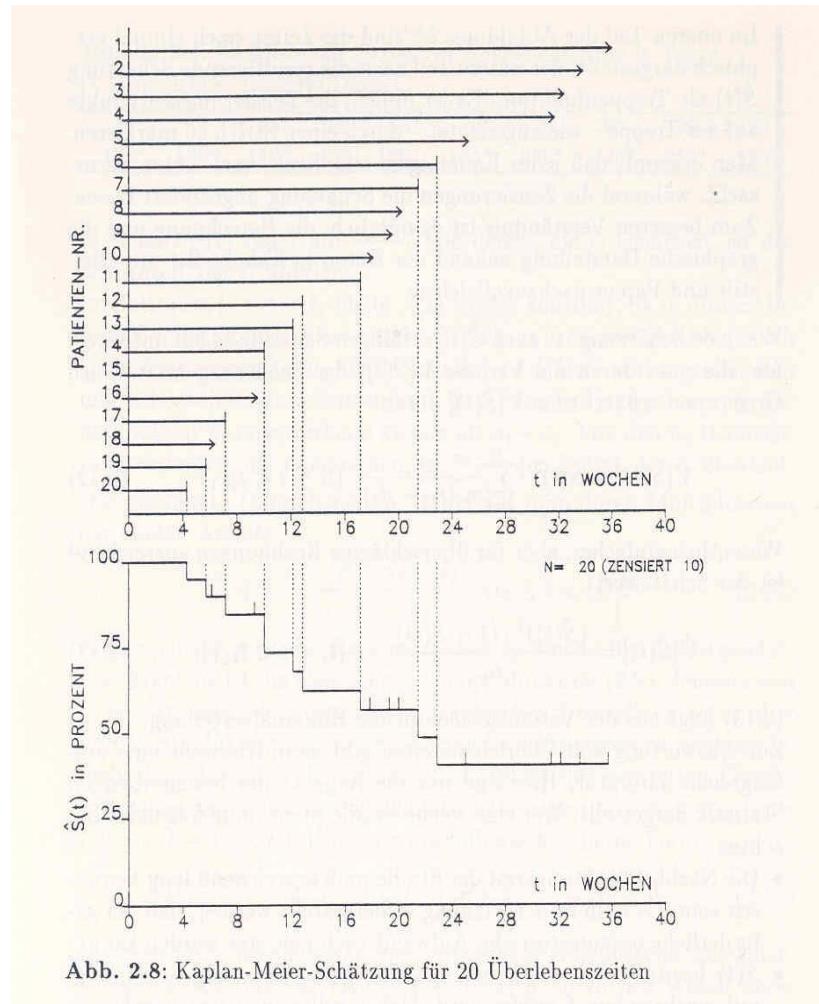
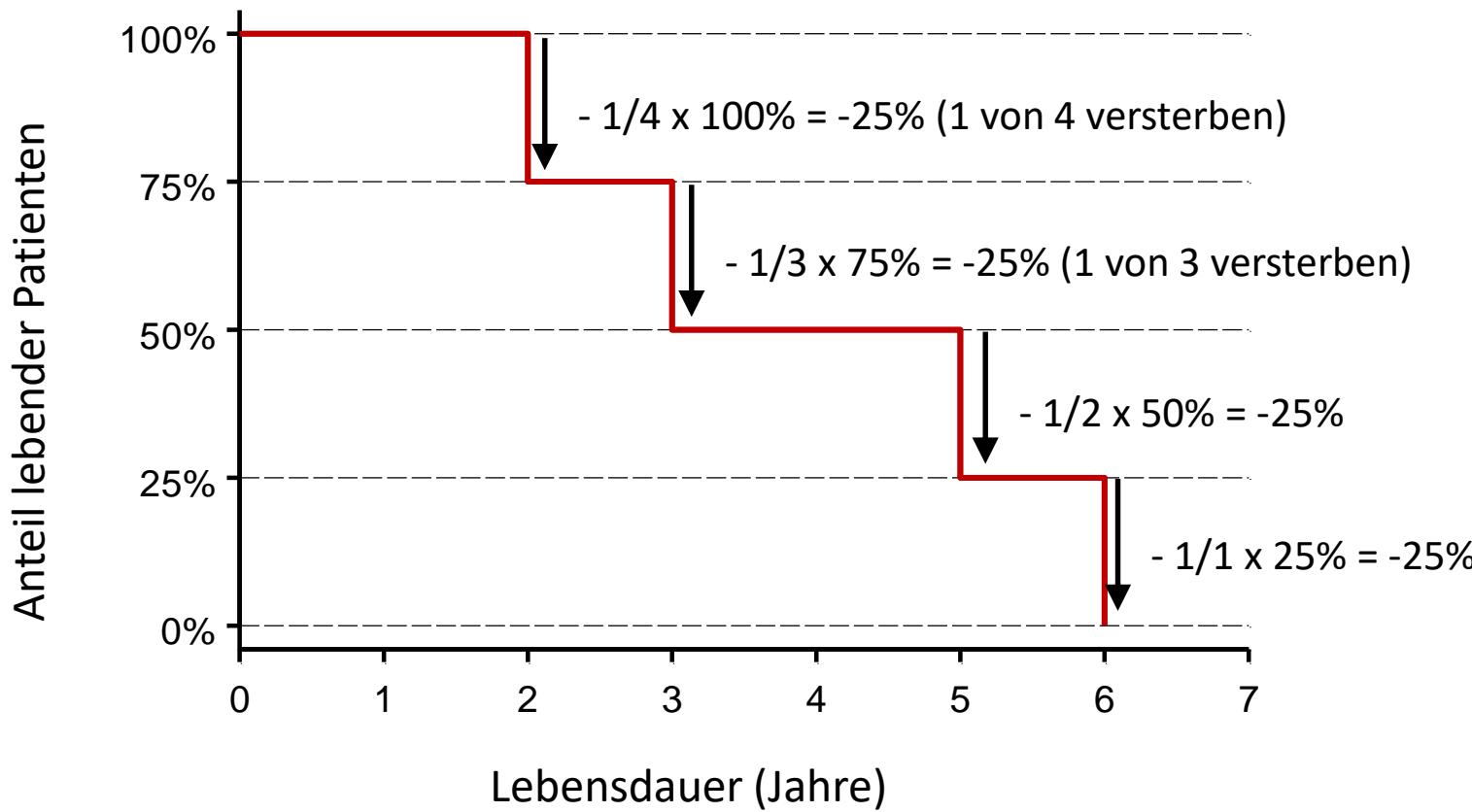


Abb. 2.8: Kaplan-Meier-Schätzung für 20 Überlebenszeiten

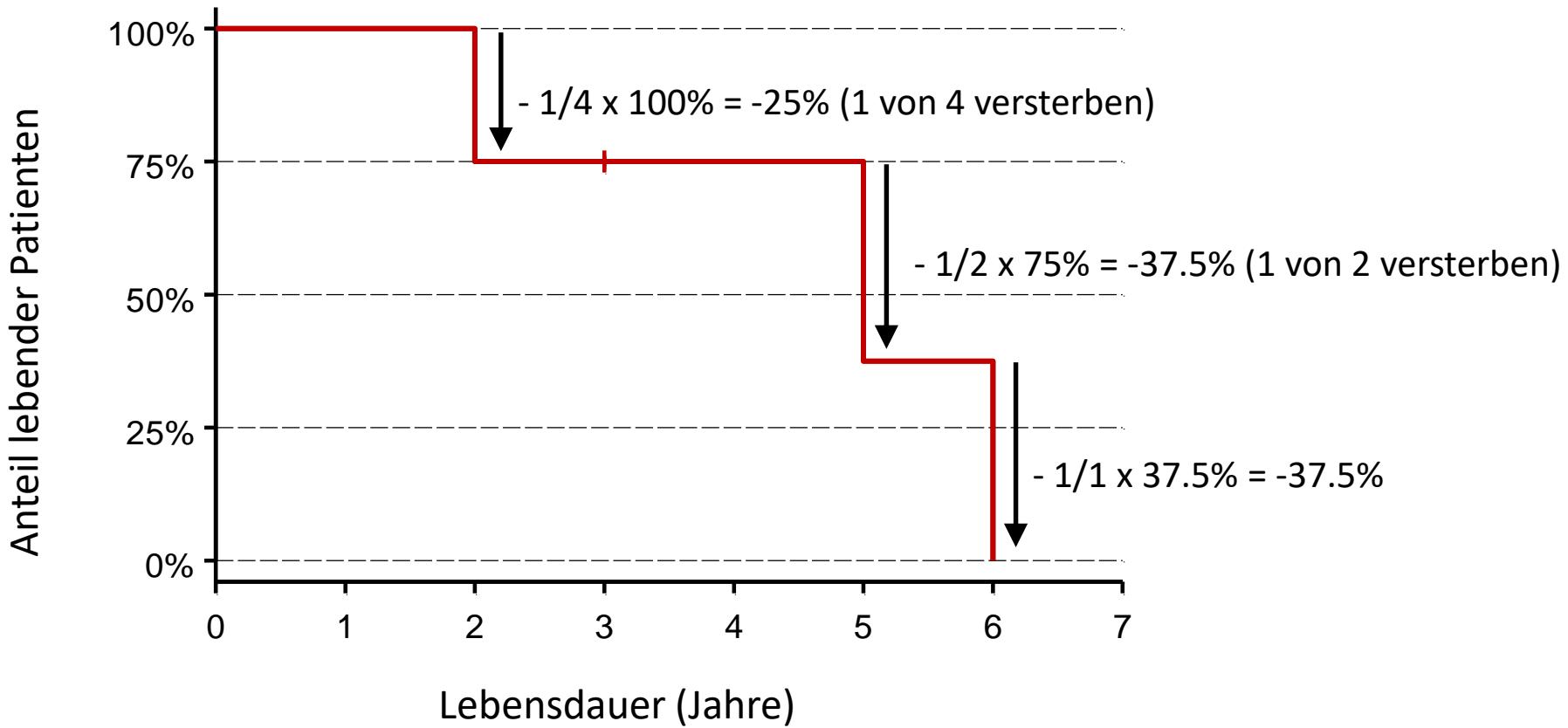
Kaplan-Meier Schätzung

Survtimes: 2, 3, 5, 6



Kaplan-Meier Schätzung

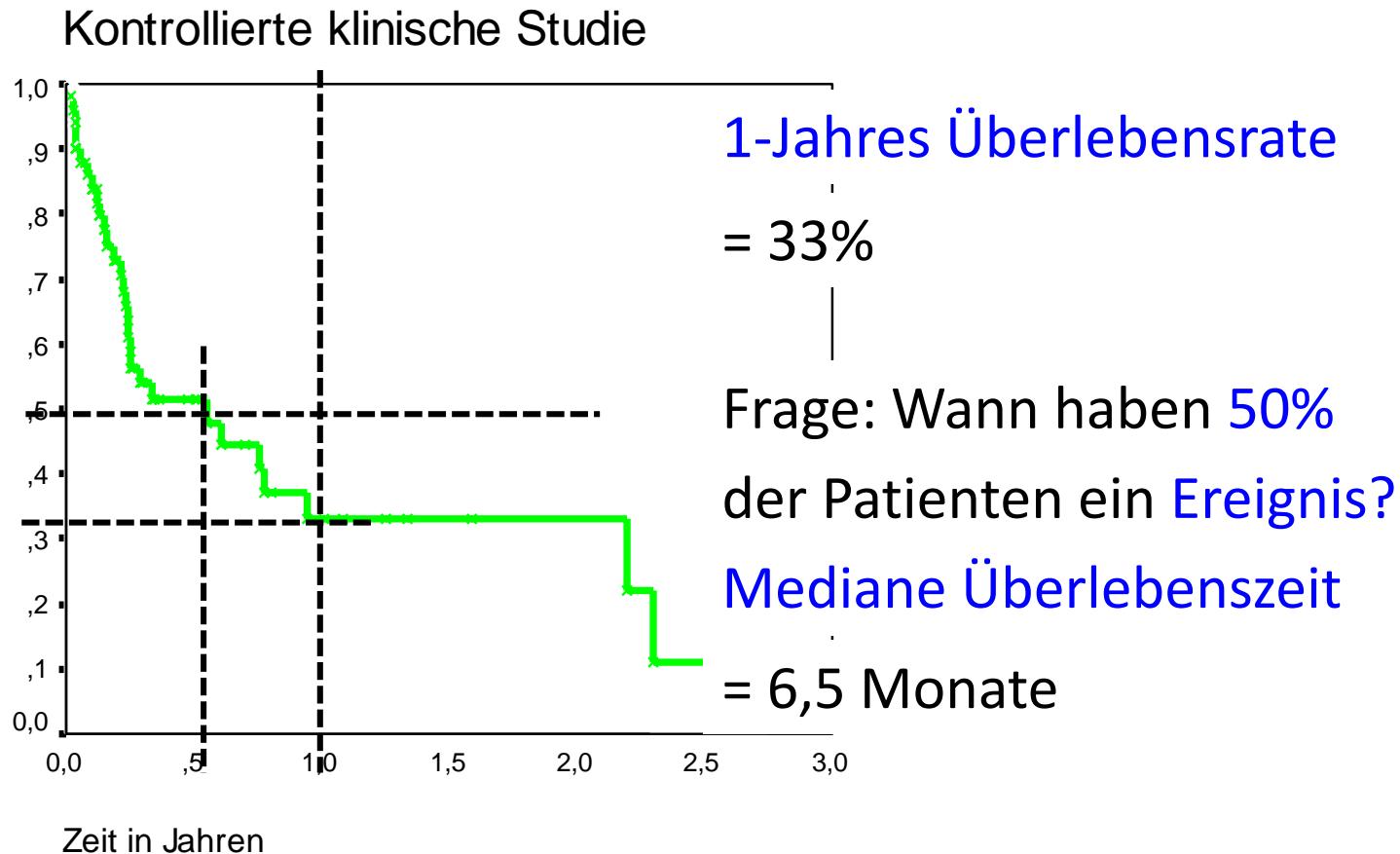
Survtimes: 2, 3+, 5, 6



Wie stellt man Überlebenszeiten dar?

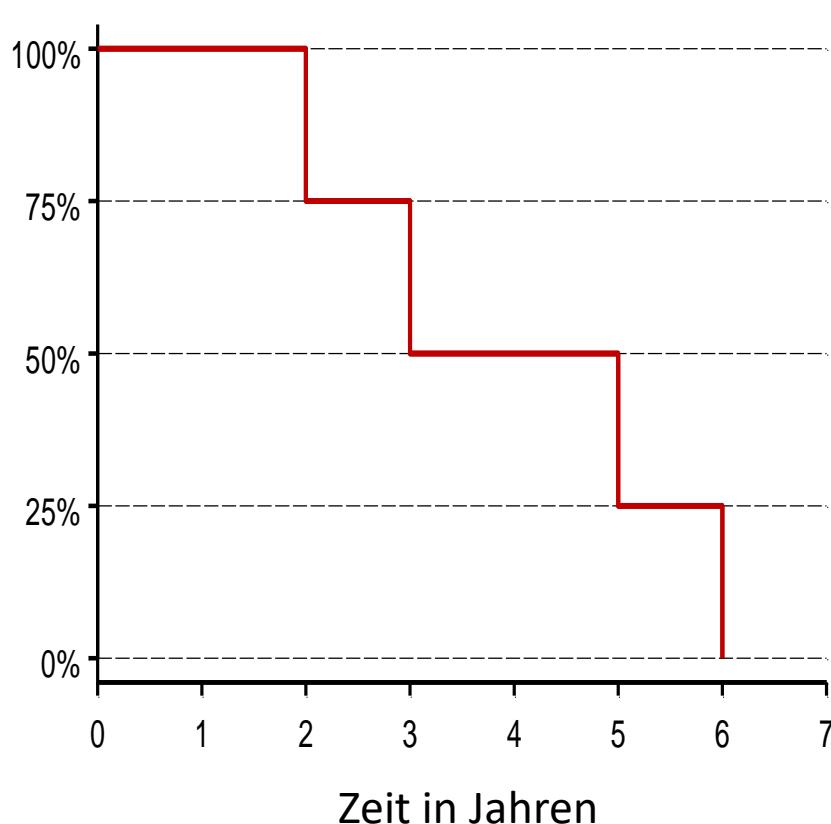
- Überlebensrate:
 - Anteil von Patienten ohne Ereignis zu einem festen Zeitpunkt t
 - Beispiel: 1-Jahres Überlebensrate von 35% = 65% Patienten mit Ereignis erwartet (nach 1 Jahr)
- Überlebenszeitkurve
 - Überlebenszeitraten zu jedem Zeitpunkt t
 - Vorteil: volle Information
- Wichtig: Berücksichtigung von Zensierungen mit der Methode von Kaplan und Meier

Überlebenszeitkurve nach Kaplan-Meier



Überlebenszeitkurve nach Kaplan-Meier

Survtimes: 2, 3, 5, 6



„Nicht eindeutige“ mediane Überlebenszeit:

Mediane Überlebenszeit wird in diesem Fall mit 3 Jahre definiert, nicht mit 4 oder 5 Jahren.

Vergleich von Überlebenszeiten

- Szenario:
 - Vergleich: Neuen Therapie vs. Standardtherapie (bzgl. Überleben)
 - Vergleich von Merkmalen (z. B. Geschlecht, Schweregrad)
- Studienfrage: Wie **beeinflusst**
 - die Therapie
 - das Merkmaldie Überlebenszeit?
- Studienfrage: Wie ist die Überlebenszeit mit dem Merkmal **assoziiert** (ohne **Kausalität**)?

Hazard-Rate und Hazard-Ratio

Hazard: Das „Risiko“ eines Ereignisses zum Zeitpunkt t
oder die „Ereignisintensität“ zum Zeitpunkt t.

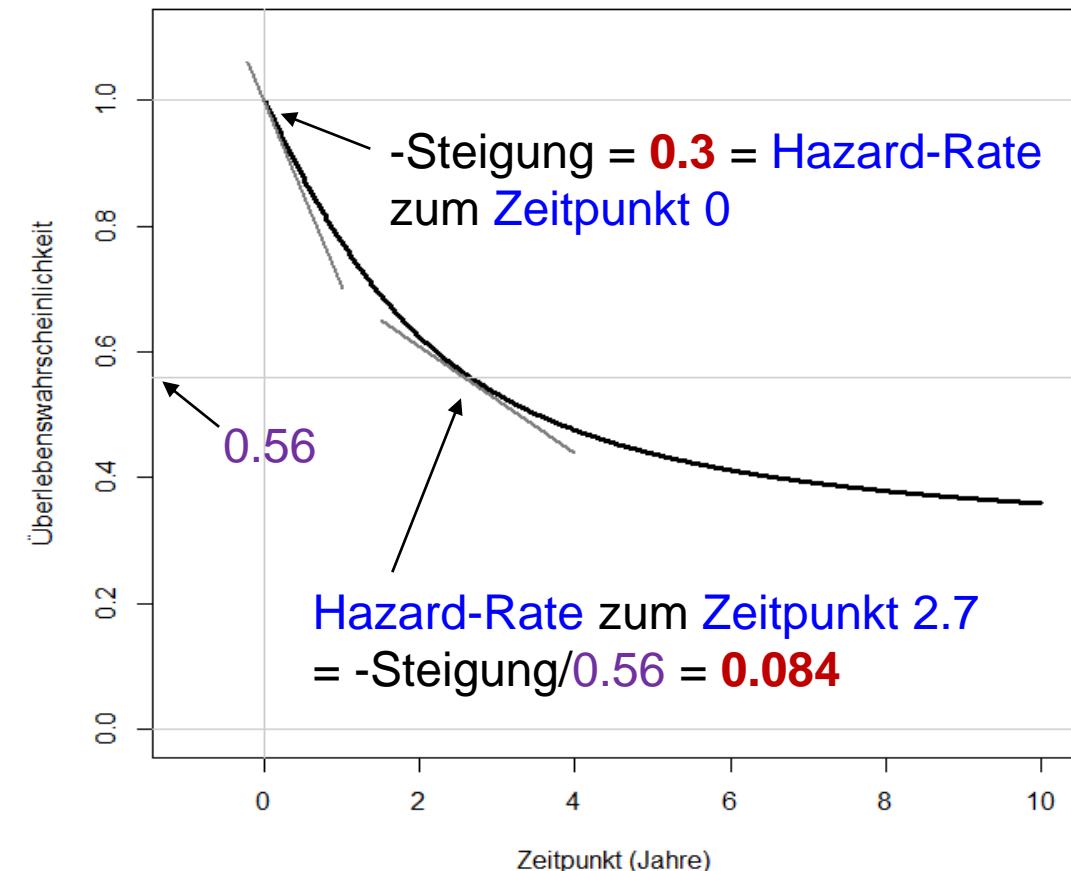
$$\text{hazard}(t) := h(t) := \lim_{\Delta \searrow 0} \frac{P(\text{Ereignis in } [t, t + \Delta] \mid \text{kein Ereignis in } [0, t])}{\Delta}$$

Hazard Ratio zum Zeitpunkt t:

$$HR(t) := \frac{\text{hazard}(t) \text{ unter der neuen Therapie}}{\text{hazard}(t) \text{ unter der Standardtherapie}}$$

$$HR(t) \left\{ \begin{array}{lll} < 1 & & \text{besser} \\ = 1 & \Leftrightarrow \text{die neue Therapie ist} & \text{gleich gut} \\ > 1 & & \text{zum Zeitpunkt t} \end{array} \right. \text{schlechter}$$

Überlebenszeitkurve und Hazard-Rate

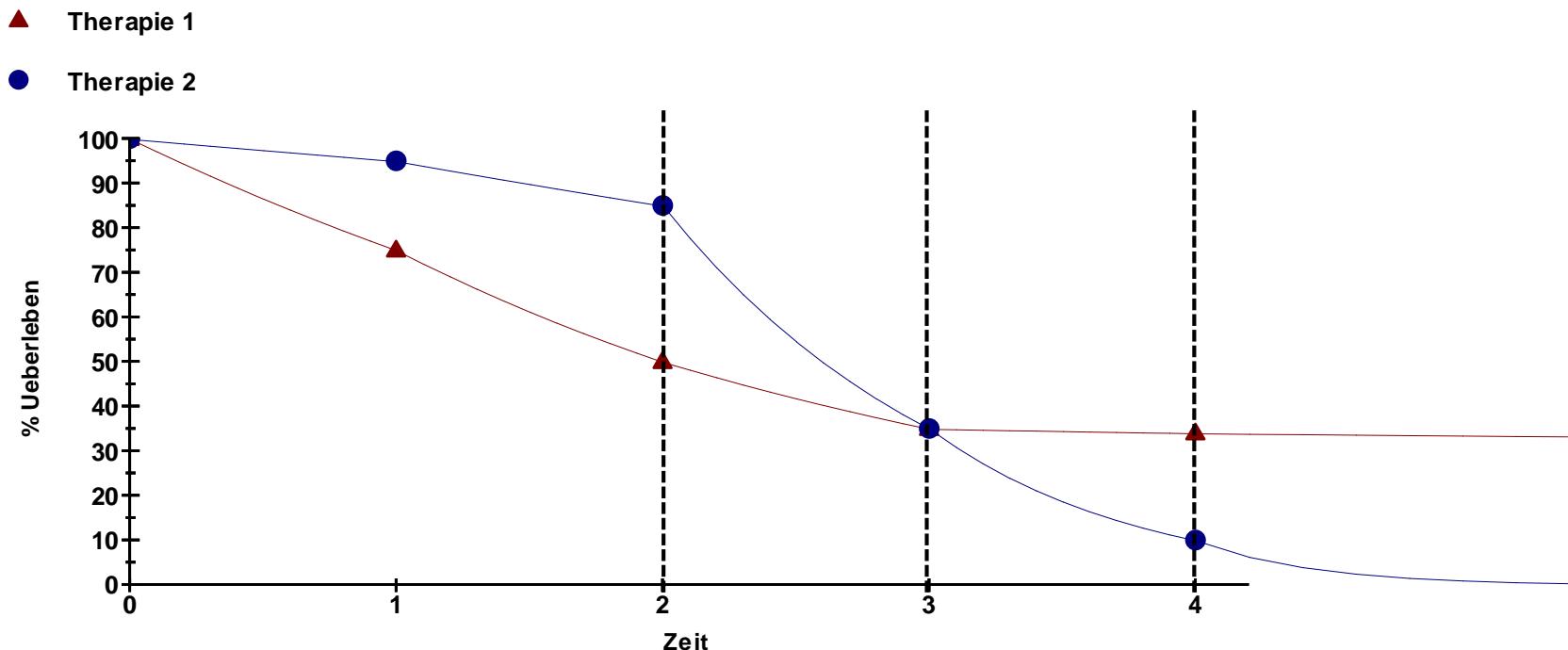


Hazard-Rate = **0.3** bedeutet:
In 0.1 Jahren sterben etwa
0.3 * 0.1 *100% = 3%

Hazard-Rate = **0.084** bedeutet:
In 0.1 Jahren sterben etwa
0.084 * 0.1 *100% = 0.8%
(0.8% von den Verliebenden)

Prozentuale Sterberate \approx Hazard-Rate * (kurzer) Zeitraum * 100%

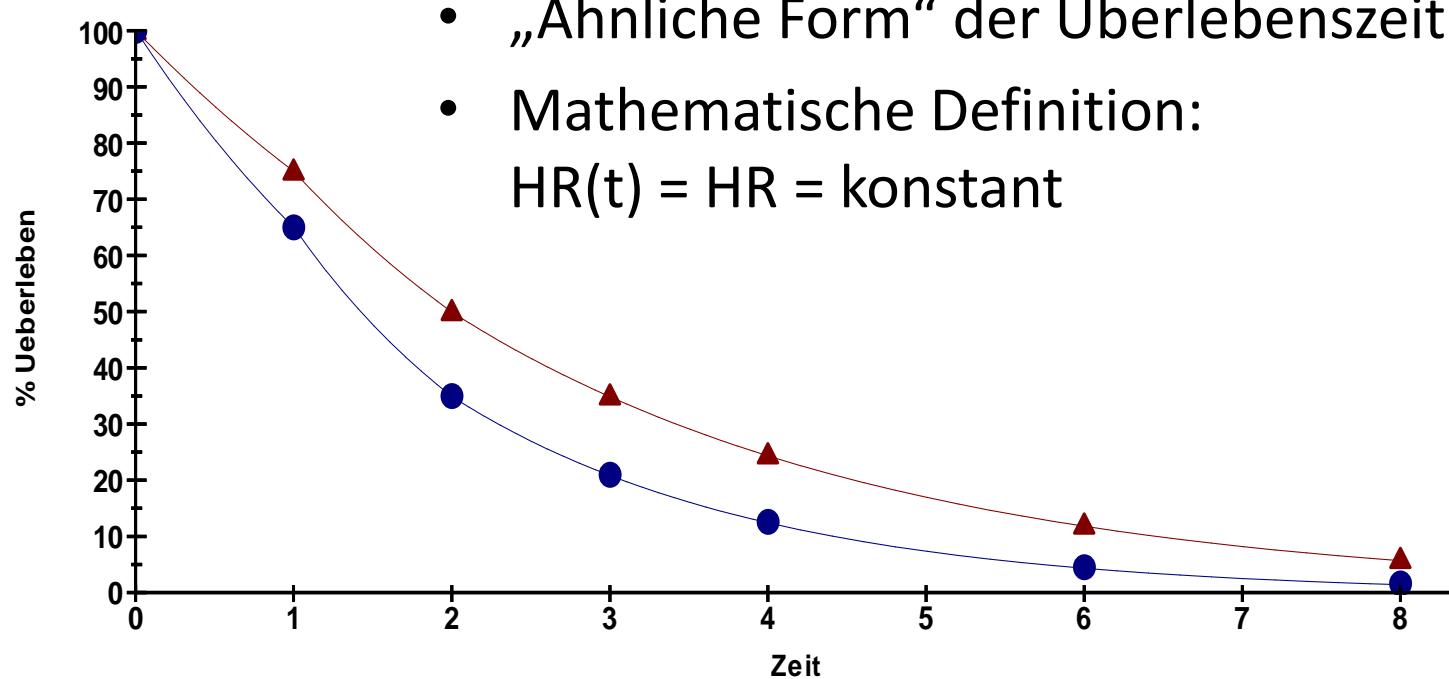
Wann sind Überlebenszeiten vergleichbar?



Der Begriff von Überlegenheit ist hier unklar!

Proportional Hazard

- ▲ Therapie 1
- Therapie 2



Proportional Hazard:

- „Ähnliche Form“ der Überlebenszeitkurven
- Mathematische Definition:
 $HR(t) = HR = \text{konstant}$

Statistische Tests

- Frage: Sind die Überlebenszeitkurven unterschiedlich zwischen 2 Gruppen (so dass Unterschied kaum noch mit dem Zufall zu erklären)?
- Standardtest: Log-Rank Test
- Voraussetzung: Überlebenszeitkurven schneiden sich nicht
- Alternativen:
 - Breslow Test (= verallgemeinerter Wilcoxon Test)
Gewicht auf frühe Ereignisse
 - Tarone-Ware Test, Kompromiss zwischen Logrank und Breslow Test



Fallbeispiel: Hypothesen und Tests – NB 2004

- **Vergleich:** Intensivierte Induktionstherapie vs. Standardinduktionstherapie bzgl. EFS (= Endpunkt bzw. Zielgröße)
- **H0 (Zweiseitige Nullhypothese):**
Die intensivierte Induktionstherapie führt zu **keinem anderen EFS** als die Standardinduktionstherapie.
- **H1 (Alternativhypothese):** Die intensivierte Induktionstherapie führt zu **einem anderen EFS** als die Standardinduktionstherapie.
- **Test:** Logrank-Test (zweiseitig)
- **Deskription:** Kaplan-Meier Kurven, Quartile, EFSraten

Vorlesungsinhalt

- Deskriptive Statistik
 - Darstellung 1es Merkmals
 - Zusammenhang 2er Merkmale
- **Statistisches Testen**
 - Prinzipien
 - Spezielle Tests
 - **Multiples Testen**
- Fallzahlplanung / Adaptive Designs
- Lineare Regression / logistische Regression
- Überlebenszeiten

Multiplicität

Multiples Testen

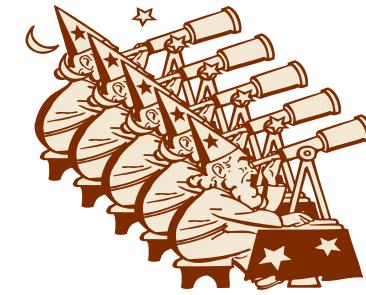
Variable	Verum N=50	Placebo N=50	p-Wert
Anzahl (%) an Patienten			
Tod	2 (4%)	1 (2%)	0.98
Progression	6 (12%)	12 (24%)	0.19
Remission	30 (60%)	20 (40%)	0.07
Schmerz	10 (20%)	10 (20%)	1.00
Gute Lebensqualität	14 (28%)	5 (10%)	0.04

Multiples Testen

Problem:

Zielvariable	Signifikanz- niveau
Tod	5%
Progression	5%
Remission	5%
Schmerz	5%
Lebensqualität	5%
Insgesamt bis zu 25%	

je mehr man anschaut ...



... desto größer die Gefahr,
einen zufälligen
Unterschied zu finden
(„ α -Inflation“)

α - Adjustierung

Problem:

Zielvariable	Signifikanz- niveau
Tod	5% 1%
Progression	5% 1%
Remission	5% 1%
Schmerz	5% 1%
Lebensqualität	5% 1%
Insgesamt bis zu 25% 5%	

Lösungen:

- A) Bonferroni-Methode:
Signifikanzniveau / #Tests
⇒ höhere Fallzahl !

α - Adjustierung

Problem:

Zielvariable	Signifikanzniveau
Tod	5%
Progression	5%
Remission	5%
Schmerz	5%
Lebensqualität	5%

~~– Insgesamt bis zu 25%~~

Lösungen:

- A) Bonferroni-Methode:
Signifikanzniveau / #Tests
⇒ höhere Fallzahl !
- B) **Auswahl eines Hauptziels**
für konfirmatorischen Test
und explorative Prüfung
der **Nebenziele**

α - Adjustierung

Problem:

Zielvariable	Signifikanz- niveau	
Tod	5%	1.
Progression	5%	2.
Remission	5%	3.
Schmerz	5%	4.
Lebensqualität	5%	5.
Insgesamt bis zu 25%		5%

Lösungen:

- A) Bonferroni-Methode:
Signifikanzniveau / #Tests
⇒ höhere Fallzahl !
- B) Auswahl eines Hauptziels
für konfirmatorischen Test
und explorative Prüfung
der Nebenziele
- C) Hierarchisches Testen
nacheinander Testen bis
ein Ergebnis nicht sign. ist
...

Typen von Fragestellungen

- **Deskriptiv:** Beschreibung der untersuchten Studieneinheiten (Patienten, Probanden, Tiere ...)
- **Explorativ:**
 - Erkennen von Besonderheiten in der Studie
 - Bewertung der Besonderheiten (wie zufällig ist der Effekt?)
 - Kein Beweis (nur Indiz) der Besonderheiten oder Thesen
- **Konfirmatorisch:**
 - Stellenwert eines **Beweises** von Thesen
 - Cave: **Wenige konfirmatorische Fragestellungen**, sonst sehr kleines lokales Signifikanzniveau

Haupt- und Nebenziele

- **Hauptziele (konfirmatorisch):**
 - Rechtfertigen die Durchführung der Studie
 - **Fallzahlplanung** richtet sich nach diesen Zielen
 - Faustregel: So viele wie nötig, so wenige wie möglich
- **Nebenziele (explorativ):**
 - weitere erfasste, interessante Parameter
 - Generierung neuer Fragen
 - u.a. für Subgruppen
- **Aber:**
p-Wert (explorativ) ist nicht gleich p-Wert (konfirmatorisch)

Multiples Testen – NB 2004

- Signifikanzniveau: 5%
- Nur eine Hauptfragestellung: Therapie innovativ vs. Therapie standard bzgl. Ereignisfreiem Überleben (EFS)
- Daher keine α -Adjustierung nötig
- Nebenfragestellungen: circa 50

Vorlesungsinhalt

- Deskriptive Statistik
 - Darstellung 1es Merkmals
 - Zusammenhang 2er Merkmale
- Statistisches Testen
 - Prinzipien
 - Spezielle Tests
 - Multiples Testen
- **Fallzahlplanung** / Adaptive Designs
- Lineare Regression / logistische Regression
- Überlebenszeiten

Fallzahlplanung

Grundidee des statistischen Testens

Zu untersuchender Effekt	Studie entscheidet für H0	Studie entscheidet für H1
H0 liegt vor	Richtige Studienaussage	Fehler 1. Art = Typ I – Fehler
H1 liegt vor	Fehler 2. Art = Typ II – Fehler	Richtige Studienaussage

Wahrscheinlichkeit eines Fehlers 2. Art β :

$1-\beta = \text{Power}$, Geplant: Power = 80% (= üblichste Wahl)

Fallzahlplanung – JUMBO Münsteraner Biometrie-Oberfläche

<https://campus.uni-muenster.de/fileadmin/einrichtung/imib/lehre/skripte/biomathe/bio/fallb.html>

Bspl.: p_1 : Erfolgsrate unter Plazebo
 p_2 : Erfolgsrate unter aktiver Therapie
 $H_0: p_1 = p_2$ gegen $H_1: p_1 \neq p_2$

Fallzahlschätzung für den Vergleich von Häufigkeiten zweier unverbundener Stichproben

Ende Neustart Hilfe!

- Fallzahlberechnung für vorgegebene Power
- Powerberechnung für vorgegebene Fallzahl
- Berechnung von p_2 für vorgegebene Fallzahl und Power

Eingabe von p_1 :
Eingabe von p_2 :

- einseitiger Test
- zweiseitiger Test

Eingabe von α (Standard ist 0.05):

Eingabe der Power (Standard ist 0.80):

Die Fallzahl für jede Gruppe ist:

Falls es **keinen** Unterschied gibt:
5%ige Wahrscheinlichkeit eines **falsch positiven** Testergebnisses

Falls es einen **Unterschied** der o.a. Größe gibt: 80%ige
Wahrscheinlichkeit eines **richtig positiven** Testergebnisses

=> Notwendige Fallzahl
Anzahl **auswertbarer** Patienten

Alternative Software:
G*Power

Berechne

Erfolgreich mit der richtigen Zahl von Patienten

Aussagesicherheit ~
Signifikanzniveau (5%, 1%)

Konfirmatorische
Fragestellungen

Erwarteter
Studieneffekt

Fallzahl

Studiendesign
(Länge der Studie,...)

Streuung d. Studieneffekte

Gewünschte Erfolgswahrscheinlichkeit
= Power (80%, 90%)

Wann wird die Fallzahl groß?

Signifikanzniveau

Konfirmatorische
Fragestellungen

Erwarteter
Studieneffekt

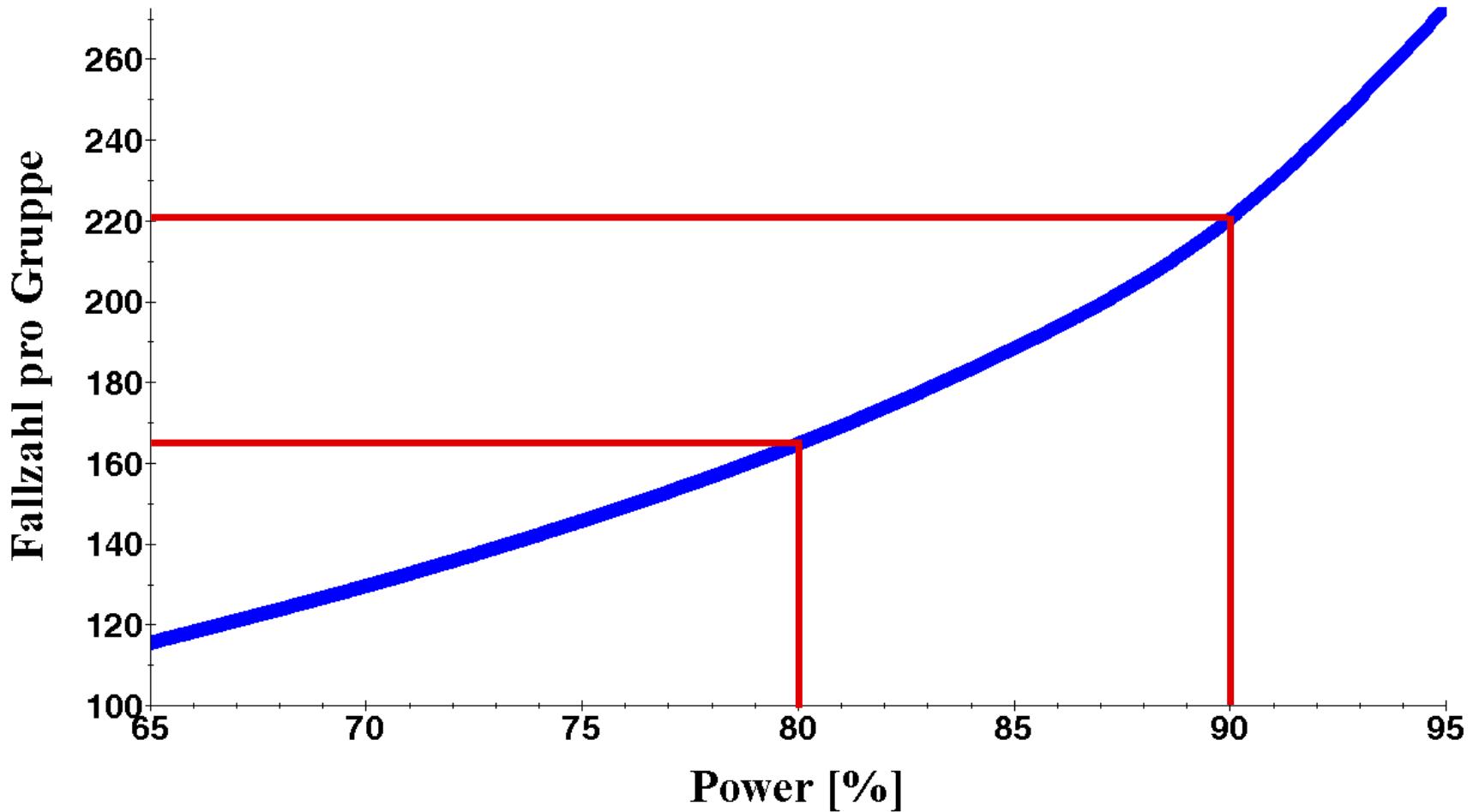
Studiendesign
(Länge der Studie,...)

Fallzahl

Streuung d. Studieneffekte

Gewünschte Erfolgswahrscheinlichkeit
= Power

NB 2004: Power vs. Fallzahl



Tipps für die Fallzahlplanung

- Lassen Sie sich Ihre benötigte Fallzahl von Ihrem Biometriker berechnen (so früh wie möglich)
- Beginnen Sie keine Studie, die 1.000.000 Patienten benötigt
- Sparen Sie Zeit: Warum 200 Patienten, falls 100 reichen
- Vorsicht vor zu großer Euphorie:
Studieneffekte werden meistens überschätzt.
- Die Rekrutierungsrate muss realistisch sein: Vorstudien
- Die vermuteten Effekte müssen belegt werden: Z. B. Pubmed
- Bei großem Vorteil der neuen Therapie:
Ist die Standardtherapie ethisch noch vertretbar?
Aber: Der Vorteil der neuen Therapie ist bisher nicht bewiesen!

Literatursuche – PubMed

<http://www.ncbi.nlm.nih.gov/sites/entrez>

neuroblastoma “high risk” children

neuroblastoma "high risk" children - PubMed result - Microsoft Internet Explorer

Search: PubMed

neuroblastoma "high risk" children

Results: 1 to 20 of 291

Results: 1 – 20 of 291

Filter your results:

All (291)
Review (48)
Free Full Text (51)
Manage Filters

Titles with your search terms:

Phase I trial of oral irinotecan and temozolamide for children with relapsed high-risk neuroblastoma [J Clin Oncol. 2009]
Outcomes of the POG 9340/9341/9342 trials for children with high-risk neuroblastoma [Pediatr Blood Cancer. 2008]
A case series of children with high-risk metastatic neuroblastoma [Pediatr Hematol Oncol. 2008]
See more...

16 free full-text articles in PubMed Central

Bcl6 is expressed in neuroblastoma: tumor cell-type-specific expression [Transl Oncol. 2009]
High-dose chemotherapy and autologous stem cell rescue in patients with high-risk neuroblastoma [J Korean Med Sci. 2009]
Interleukin-6 and soluble interleukin-6 receptor levels as markers of disease [Clin Cancer Res. 2008]
See all (16)...

Find related data

Database: Select

Find items

Search details

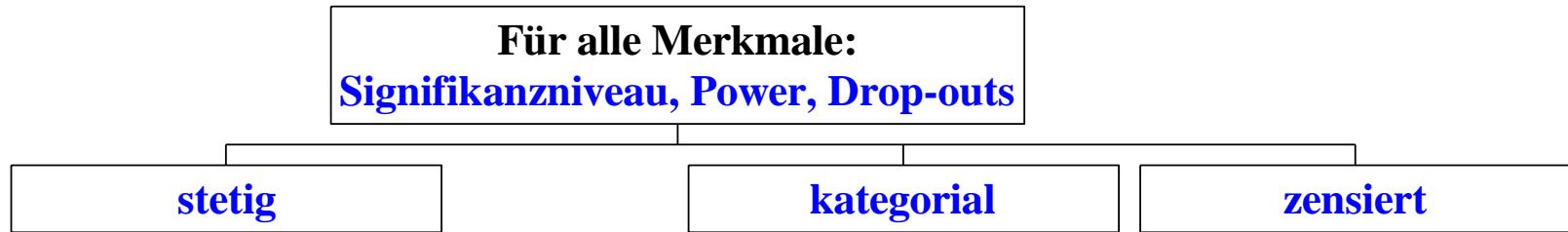
("neuroblastoma" [MeSH Terms])

Fertig

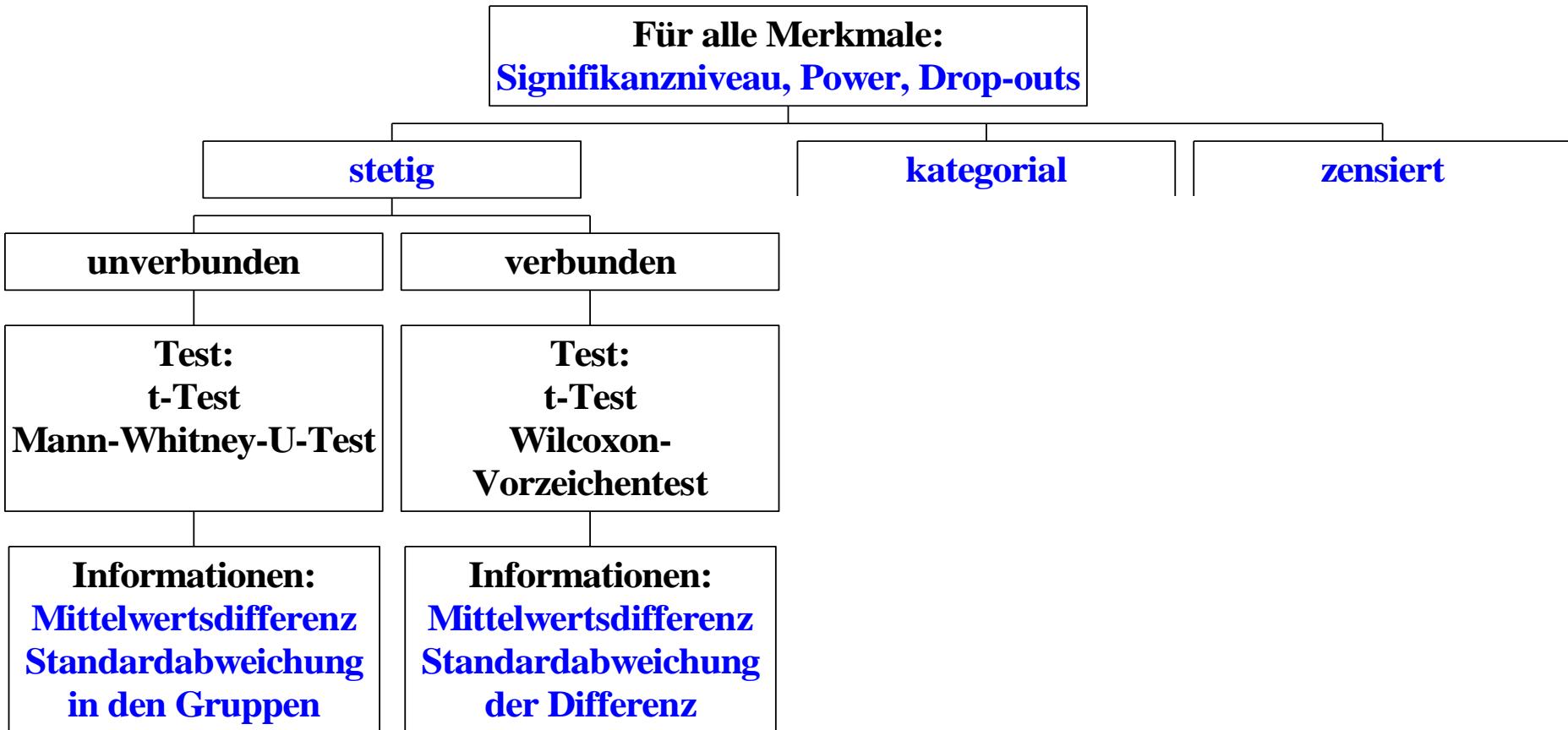
Notwendige Kennzahlen für die Fallzahlplanung

Für alle Merkmale:
Signifikanzniveau, Power, Drop-outs

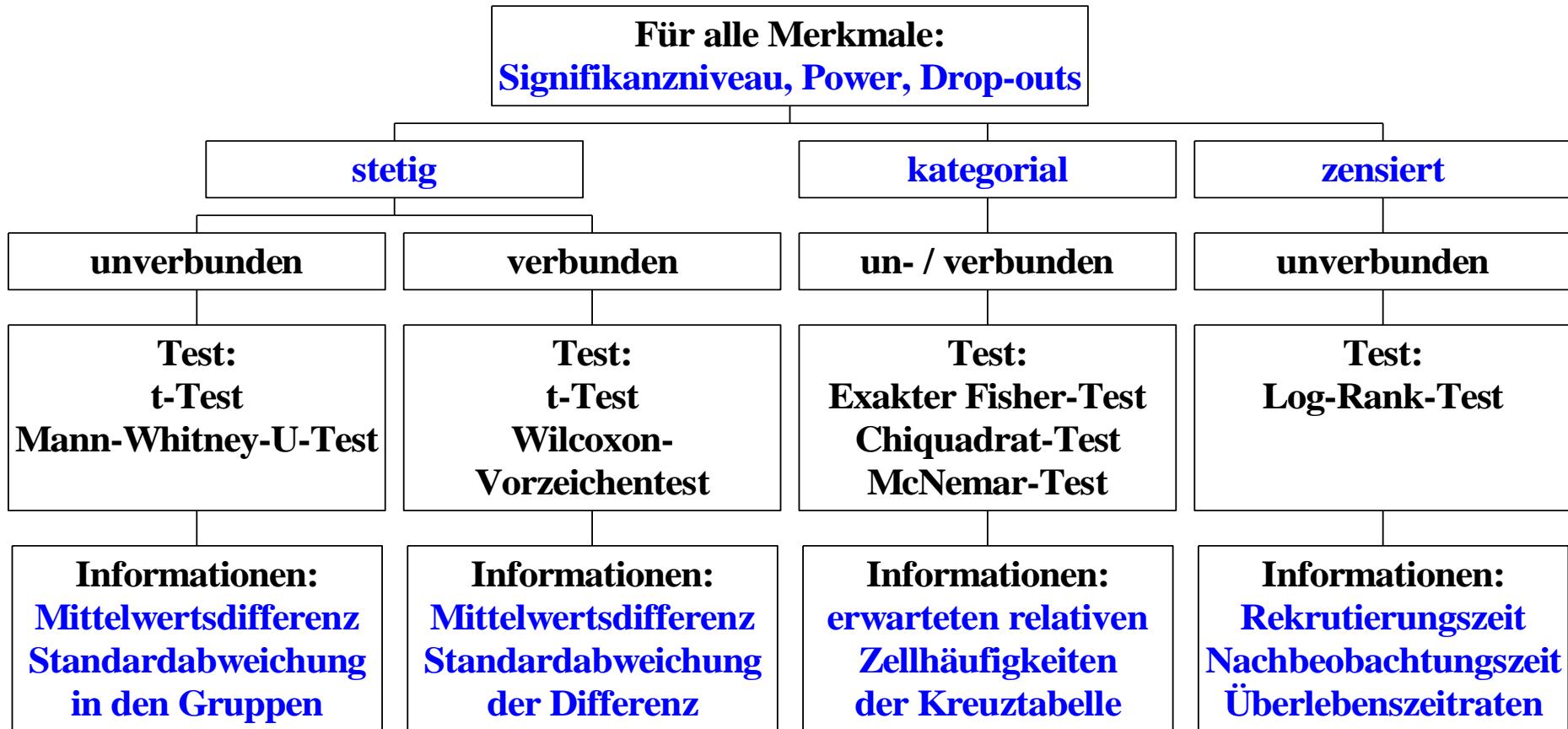
Notwendige Kennzahlen für die Fallzahlplanung



Notwendige Kennzahlen für die Fallzahlplanung



Notwendige Kennzahlen für die Fallzahlplanung



Vorlesungsinhalt

- Deskriptive Statistik
 - Darstellung 1es Merkmals
 - Zusammenhang 2er Merkmale
- Statistisches Testen
 - Prinzipien
 - Spezielle Tests
 - Multiples Testen
- Fallzahlplanung / **Adaptive Designs**
- Lineare Regression / logistische Regression
- Überlebenszeiten

Adaptive Designs

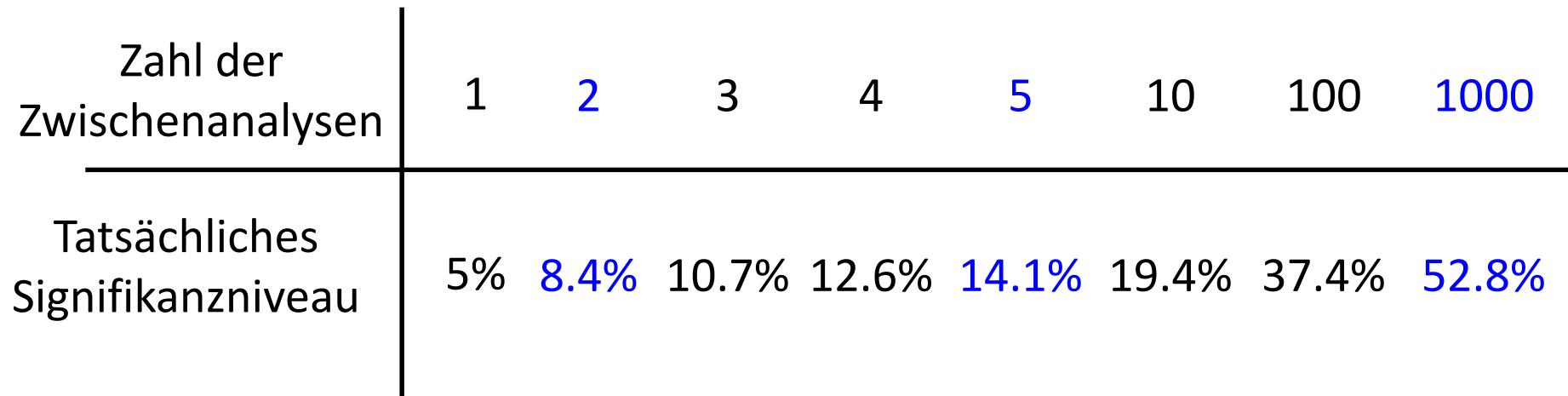
Warum brauchen wir Flexibilität / Zwischenanalysen?

- **Informationen** vor Beginn der Studie sind häufig **unsicher**
 - D.h. Planungskorrektur im Laufe der Studie erwünscht (auf Basis von Zwischenanalysen)
- **Ethische und Ökonomische Gründe**
 - a. Effekt **kaum mehr nachweisbar** ⇒ neuer Therapieansatz, Vermeidung weiterer randomisierter Zuteilungen einer ineffektiven Therapie
 - b. Effekt **frühzeitig nachweisbar** ⇒ innovativen Therapieansatz etablieren, Vermeidung weiterer randomisierter Zuteilungen der schlechteren Standard-Therapie

In beiden Fällen a. und b. : Einsparung von Ressourcen

Warum brauchen wir Regeln für Zwischenanalysen?

Mehrere Zwischenauswertungen zum Signifikanzniveau 5% \Rightarrow
Wahrscheinlichkeit einer falschen „Signifikanz“ $>> 5\% !!$

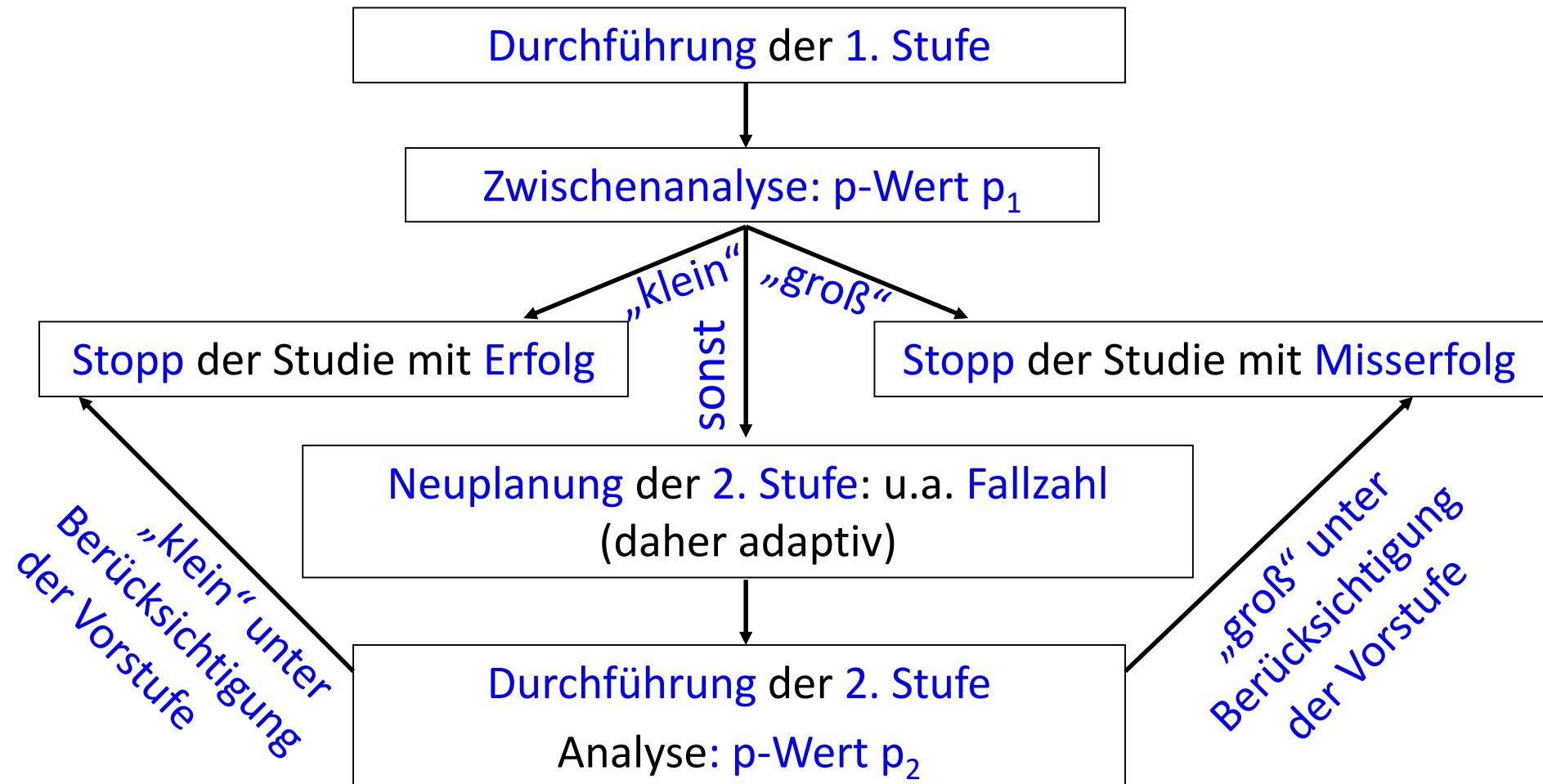


Inflation des Signifikanzniveaus (Variante von multiplem Testen)!

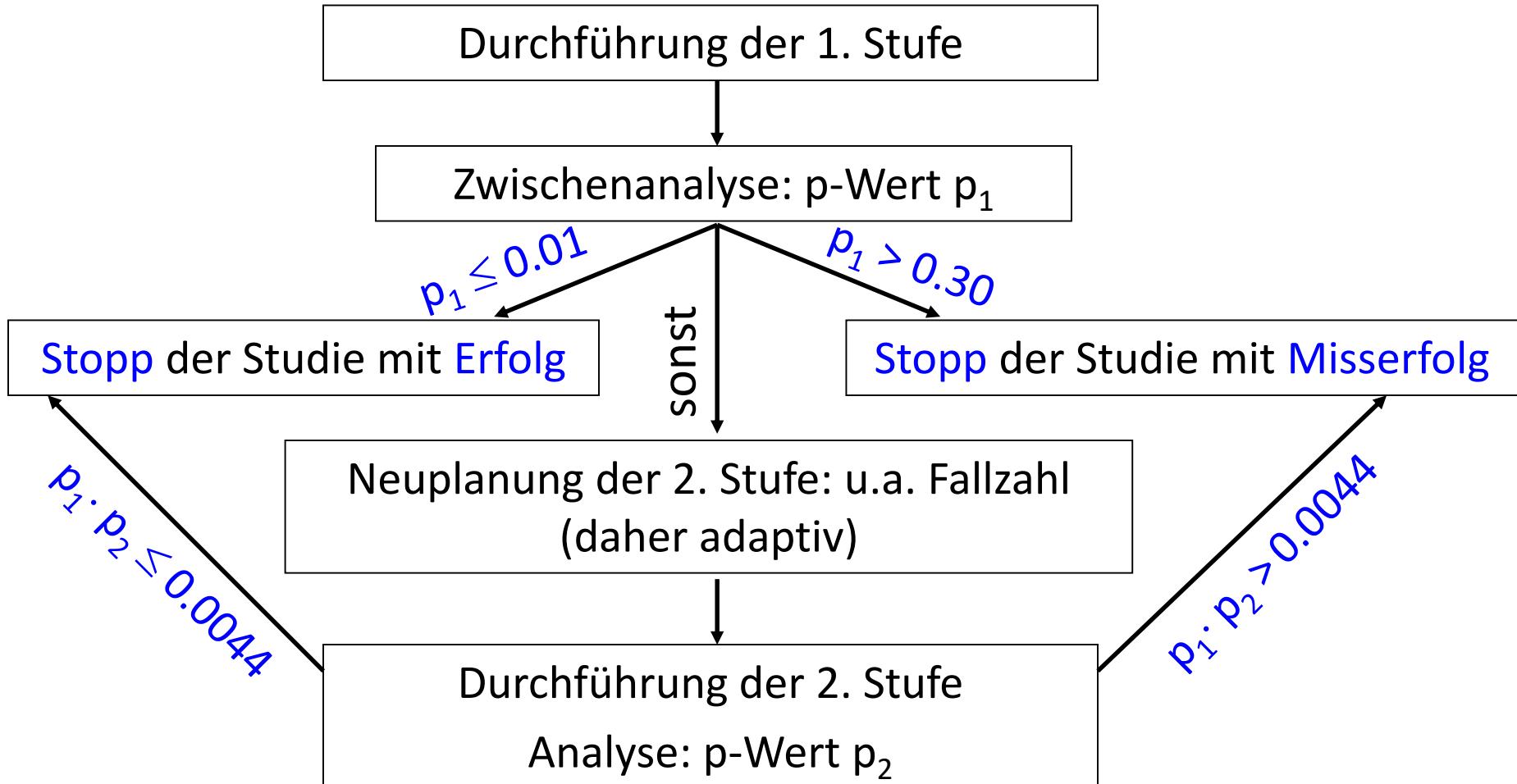
Typen von Zwischenanalysen (Weiterentwicklungen der Methodik)

- Pilotstudie
 - Ziel: Schätzung des Effektes
 - Patienten für Hauptstudie verloren
- Gruppensequentielles Design (seit 70er)
 - Gesamtfallzahl wird vor Studienbeginn festgelegt
 - Zwischenanalyse nach festgelegter Zeit, Patientenzahl, ...
 - Entscheidung über Erfolg, Misserfolg oder Weiterführung
- Adaptives Design (seit 90er)
 - Studie wird nach jeder Zwischenanalyse neu geplant (update der Effektschätzungen)
 - Ergebnisse der bisherigen Studie werden genutzt

Prinzip einer adaptiv-sequentiellen Studie



Beispiel - Bauer und Köhne Design (1994) mit $\alpha = 0.025$



Was können adaptive Designs leisten?

- Frühe Beendigung der Studie, falls das Studienziel erreicht, oder Erfolg bei Weiterführung fast aussichtslos
- Anpassung des Studiendesigns (insbesondere Fallzahl) an neue Erkenntnisse oder beobachtete Studieneffekte
- Anpassung der Studienziele an neue Erkenntnisse
- Flexible Zeitpunkte (z. B. Studienende) der Analyse
- Schnellere Ergebnisse im Mittel
- Geringere Fallzahlen bzw. höhere Power im Mittel

Auch höhere Fallzahlen möglich, falls nötig bei kleinem Studieneffekt

Wichtig: Transparenz

Festlegung der Regeln für Zwischenanalysen im Protokoll !



Zwischenanalysen – NB 2004

- 2 Zwischenanalysen: nach 1/3 & 2/3 der erwarteten Ereignisse
- Entscheidung über Beendigung oder Fortführung der Studie, falls das Studienziel erreicht wurde oder unrealistisch ist
- Durch Zwischenanalysen max. Fallzahl 360 statt 330 Patienten (nach NB 97: 60 Patienten / Jahr, 6 Jahre Rekrutierung)
- Im Mittel aber kürzere Studienzeit und weniger Patienten
- Möglichkeit das gesamte Studiendesign während der Studie jederzeit ungeplant adaptiv zu verändern



Vorlesungsinhalt

- Deskriptive Statistik
 - Darstellung 1es Merkmals
 - Zusammenhang 2er Merkmale
- Statistisches Testen
 - Prinzipien
 - Spezielle Tests
 - Multiples Testen
- Fallzahlplanung / Adaptive Designs
- **Lineare Regression** / logistische Regression
- Überlebenszeiten



Regressionsmodelle

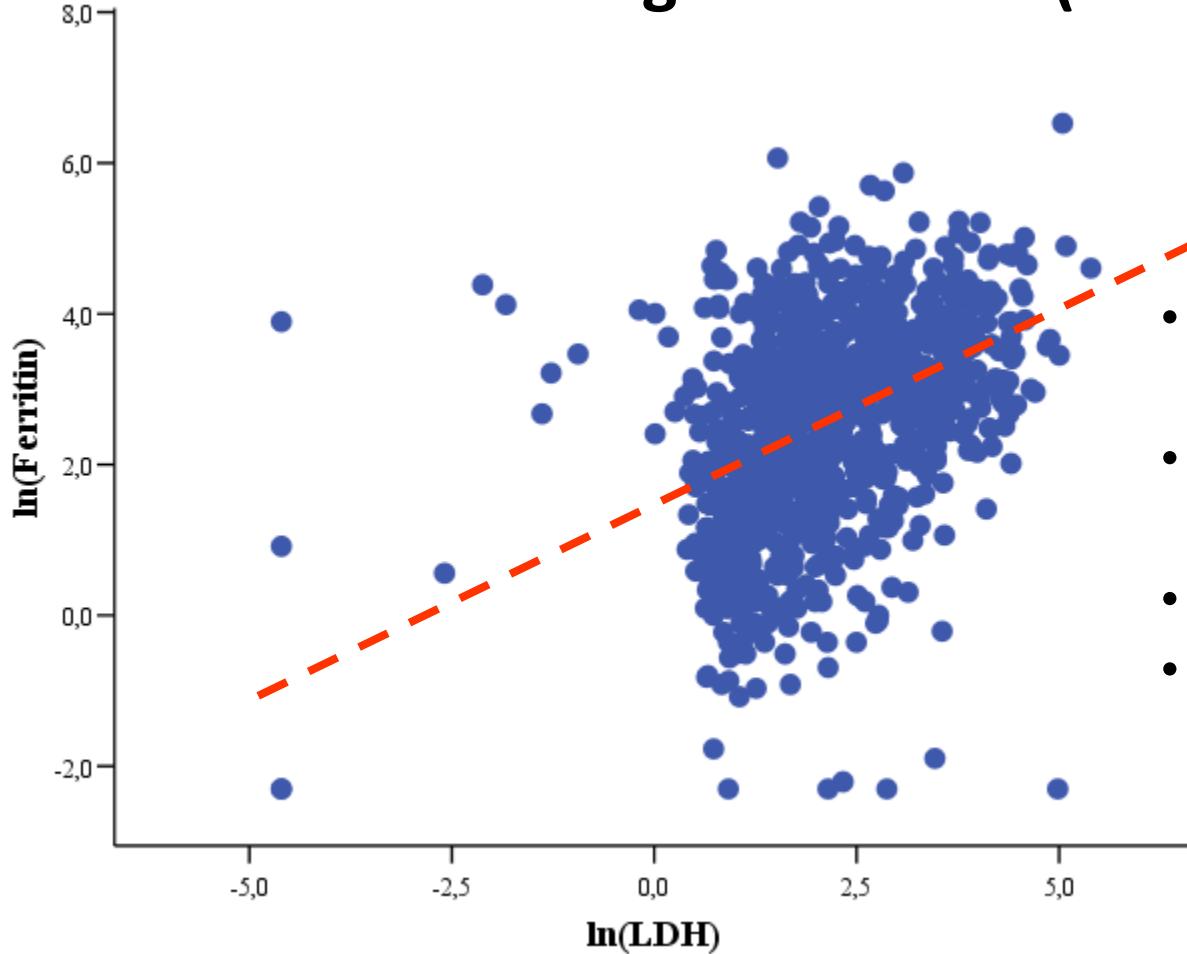
- Einflussgrößen → Zielgröße (Alter, Geschlecht → Blutdruck)
- Zielgröße entscheidet über das Regressionsmodell
 - stetige Zielgröße → lineare Regression
 - binäre Zielgröße → logistische Regression
 - zensierte Zielgröße (Überlebenszeiten) → Cox-Regression
 - Anzahl: 1
- Einflussgrößen entscheiden über die Komplexität
 - 1 → einfache Regression
 - > 1 → multivariable / multiple (/ multivariate) Regression
 - quantitative oder qualitative Einflussgrößen möglich

Regressionsmodelle - Ziele

- Vorhersage = **Prognose** der **Zielgröße** durch **Einflussgrößen**
- Identifikation **prognostisch relevanter Einflussgrößen**
- Identifikation **prädiktiver Einflussgrößen**:
 - bei welchen Ausprägungen wirkt die **Therapie**
 - **Mamma**.: Östrogenrezeptorstatus → Hormontherapie
- Wie wirken mehrere Einflüsse gleichzeitig auf eine Zielgröße?
 - **Korrektur heterogener Stichproben**
 - Ohne Alterskorrektur: **ältere Frauen, jüngere Männer** ⇒ **Frauen** haben **vermeintlich** den **höheren Blutdruck**
- Identifikation **unabhängiger Einflussgrößen**: prognostischer oder prädiktiver Wert zusätzlich zu bekannten Einflüssen

Einfache lineare Regression

Zusammenhang zwischen $\ln(\text{LDH})$ und $\ln(\text{Ferritin})$



- Y = abhängige Variable
Zielgröße
- X = unabhängige Variable
= Einflussgröße
- $Y = a + b \cdot X$
- Regressionskoeffizienten a, b :
 a : Achsenabschnitt
(Intercept) b : Steigung
(Slope)

Einfache lineare Regression

- Beispiel NB 2004: **hängt** $\log(\text{Ferritin})$ von $\log(\text{LDH})$ ab
- Allgemein:
 - **Y = abhängige Variable** = Zielgröße (z. B. $\log(\text{Ferritin})$)
 - **X = unabhängige Variable** = Einflussgröße (z. B. $\log(\text{LDH})$)
- Modell:
 - **deterministisch**: $Y = a + b \cdot X$
 - **stochastisch**: $Y = a + b \cdot X + \varepsilon$
 ε : normalverteilt, Erwartungswert = 0, gleiche Varianz
- Begriffe: **Regressionskoeffizienten a und b**
 - **a** heißt **Achsenabschnitt** = Intercept
 - **b** heißt **Steigung** = Slope



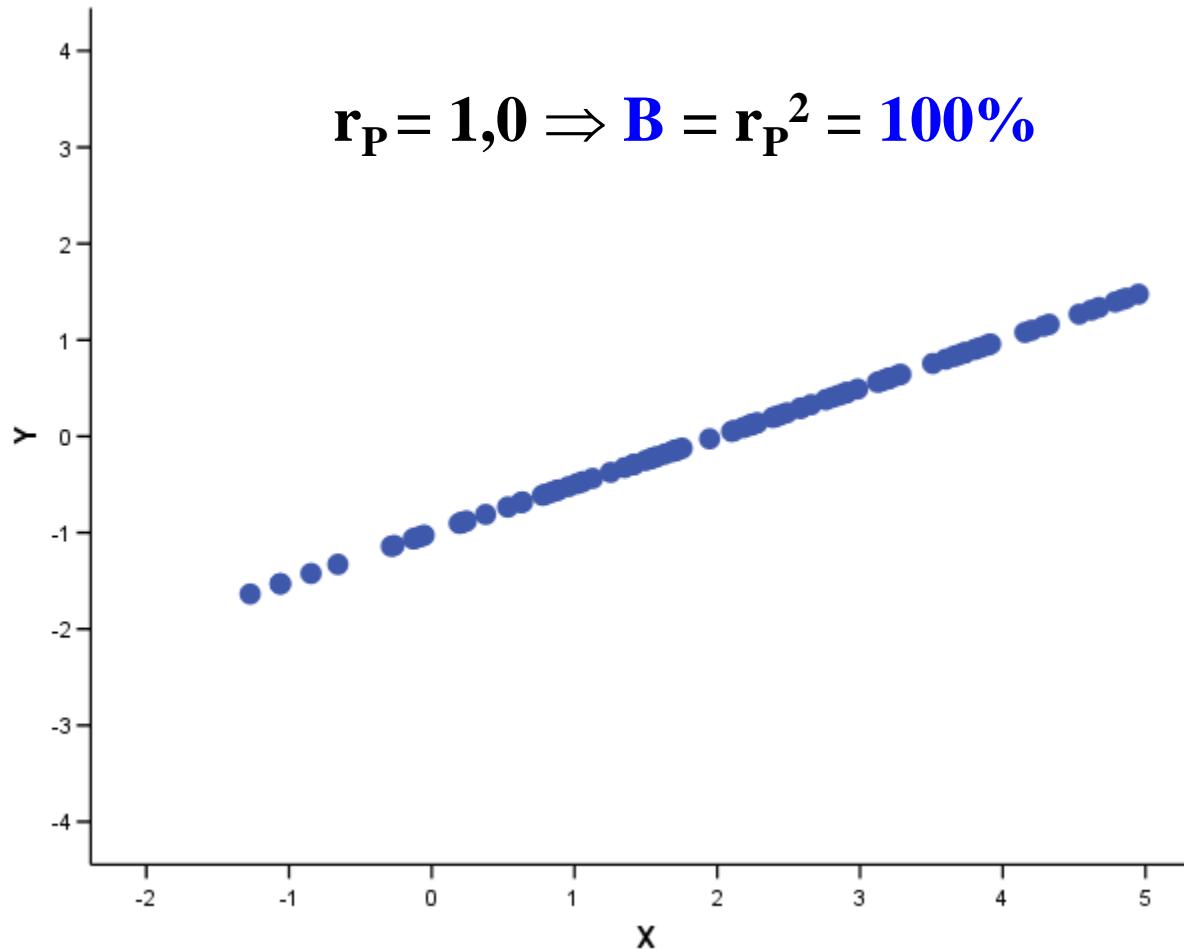
Einfache lineare Regression

- Modell: $Y = a + b \cdot X + \varepsilon \Rightarrow$
Ist Einflussgröße $X = x$, so ist der erwartete Y -Wert $y = a + b \cdot x$
- Beispiel NB 2004: $\log(\text{Ferritin})$ [$\log(\text{ng/ml})$] =
 $1,4 \log(\text{ng/ml}) + 0,5 \log(\text{ng/ml})/\log(\text{U/l}) \cdot \log(\text{LDH})$ [$\log(\text{U/l})$]
- Interpretation des Achsenabschnitts $1,4 \log(\text{ng/ml})$:
bei einem $\log(\text{LDH})$ -Wert von $0 \log(\text{U/l})$ wird
ein $\log(\text{Ferritin})$ -Wert von $1,4 \log(\text{ng/ml})$ erwartet
- Interpretation der Steigung $0,5 \log(\text{ng/ml})/\log(\text{U/l})$:
steigt der $\log(\text{LDH})$ -Wert um $1 \log(\text{U/l})$,
so steigt der erwartete $\log(\text{Ferritin})$ -Wert um $0,5 \log(\text{ng/ml})$

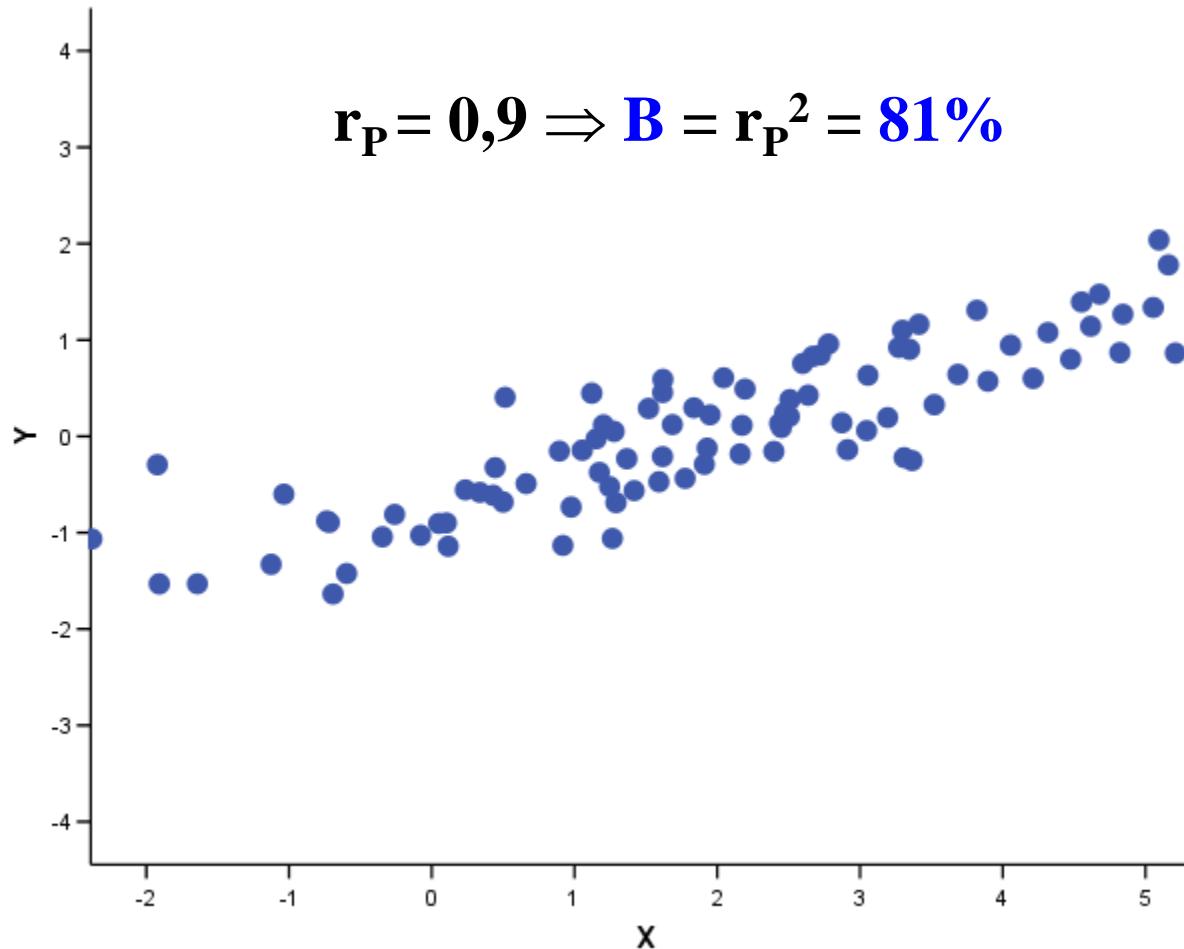
Einfache lineare Regression

- Einheiten nicht vergessen
- Keine Extrapolationen: Regressionsgleichung nur gültig für Wertebereich -4,6 bis 5,4 $\log(U/I)$ von $\log(LDH)$
- Frage: Wie stark ist der Zusammenhang zwischen X und Y
- Antwort: Bestimmtheitsmaß B
 - $B = 1 - \sum(Y_i - a - b \cdot X_i)^2 / \sum(Y_i - Y_M)^2$; Y_M = Mittelwert v. Y
 - Anteil der Varianz von Y, der durch $a + b \cdot X$ erklärt wird
 - B: vorhersagbare Streuung, 1-B: zufällige Streuung
 - Oft: guter linearer Zusammenhang $\leftrightarrow B > 50\%$
 - $B = r_p^2 \in [0; 1] \Rightarrow$ zeigt keine Richtung d. Zusammenhangs

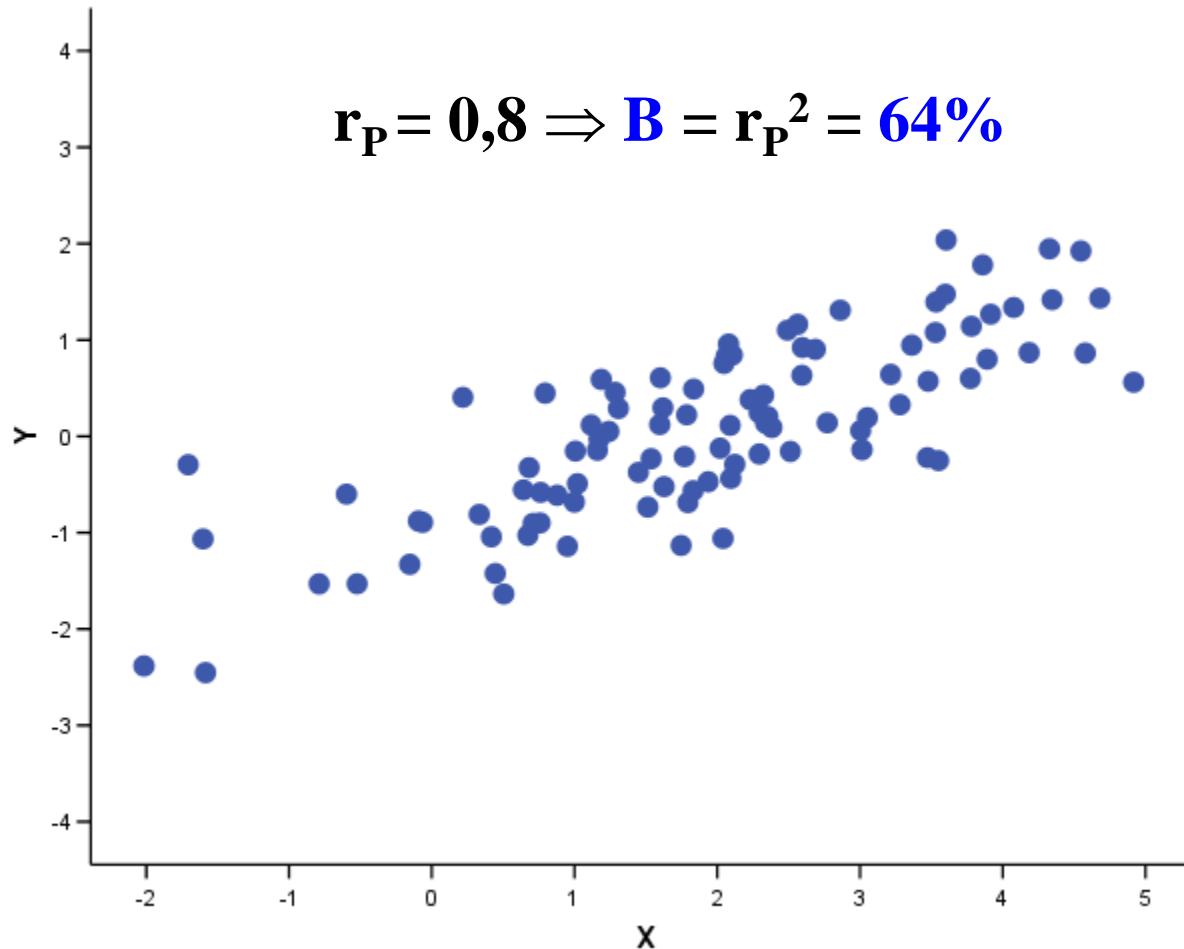
Korrelationskoeffizient und Streudiagramm



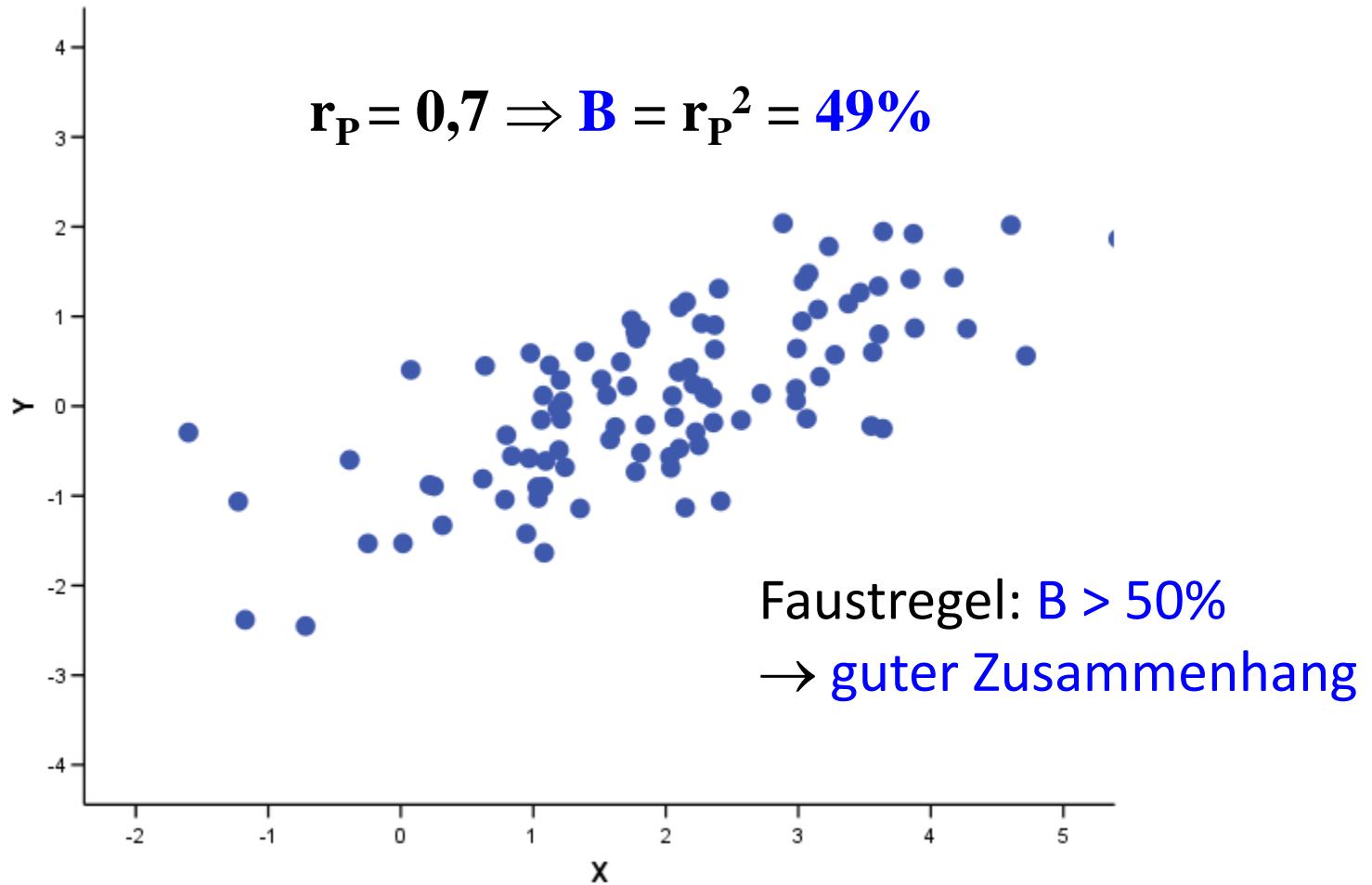
Korrelationskoeffizient und Streudiagramm



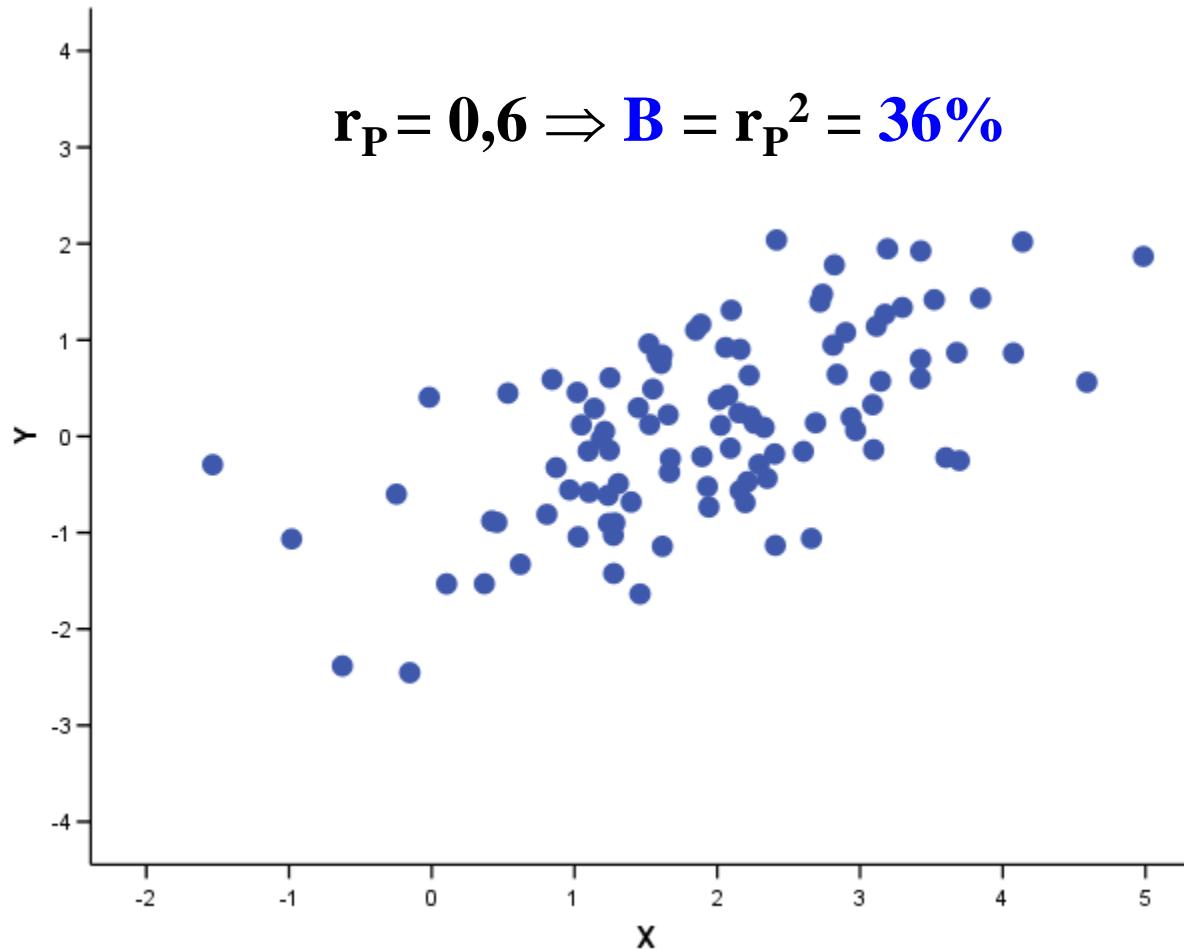
Korrelationskoeffizient und Streudiagramm



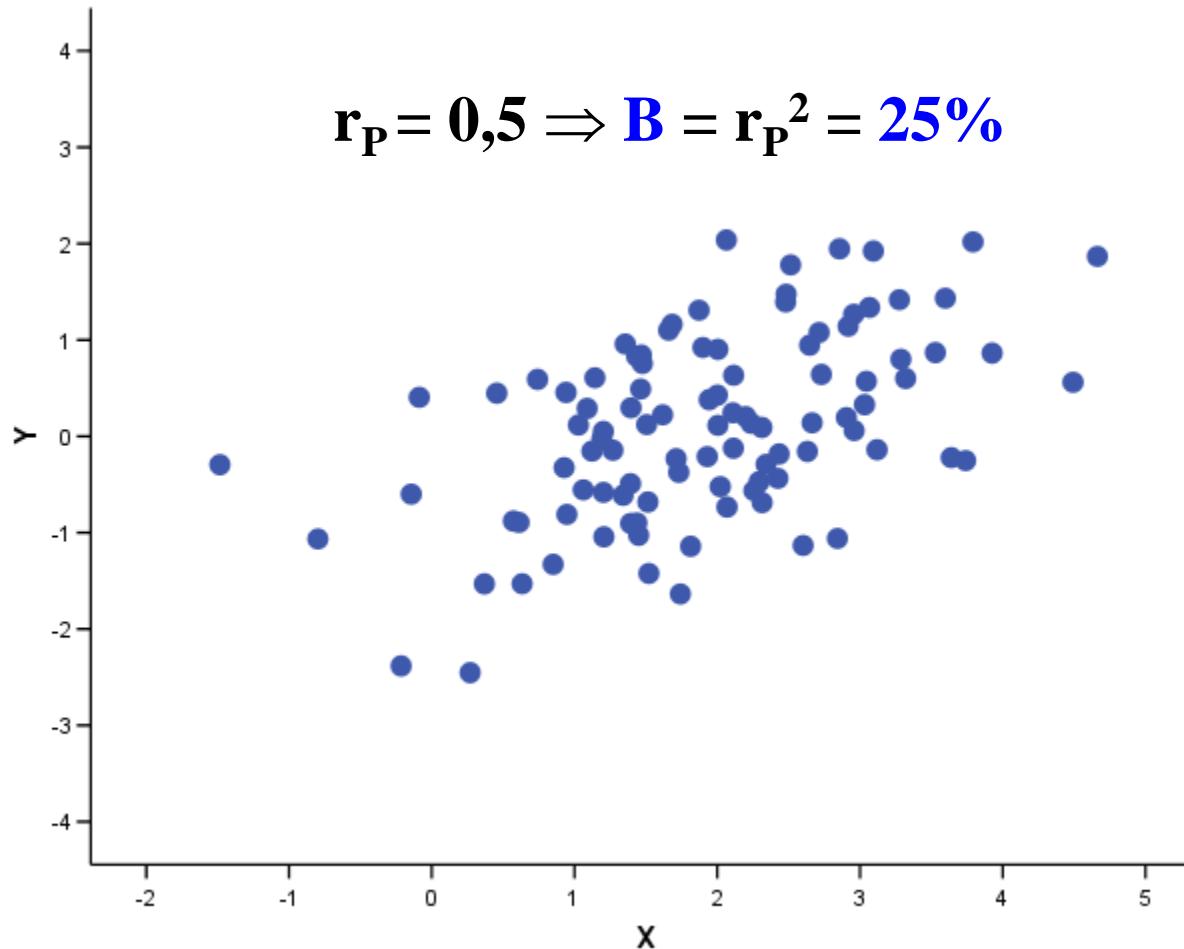
Korrelationskoeffizient und Streudiagramm



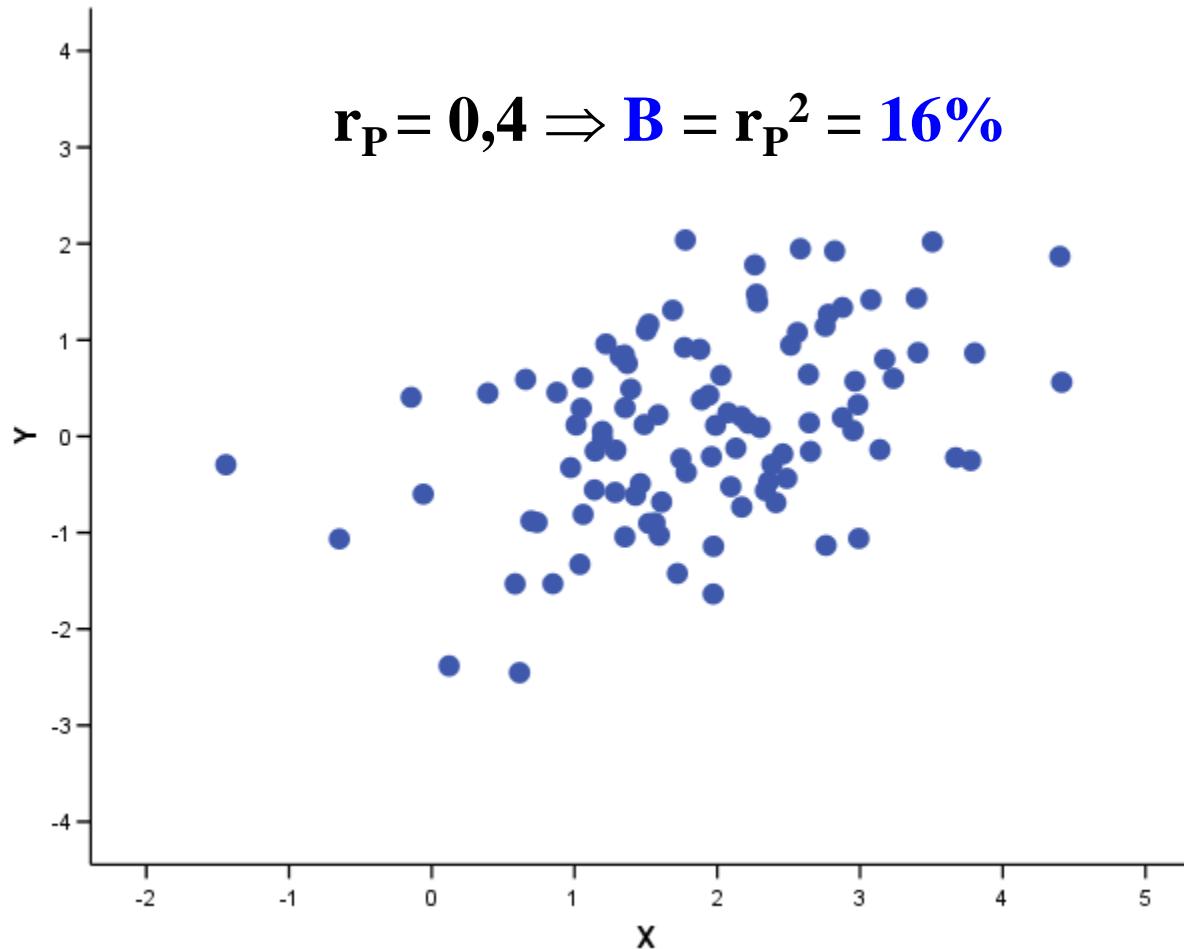
Korrelationskoeffizient und Streudiagramm



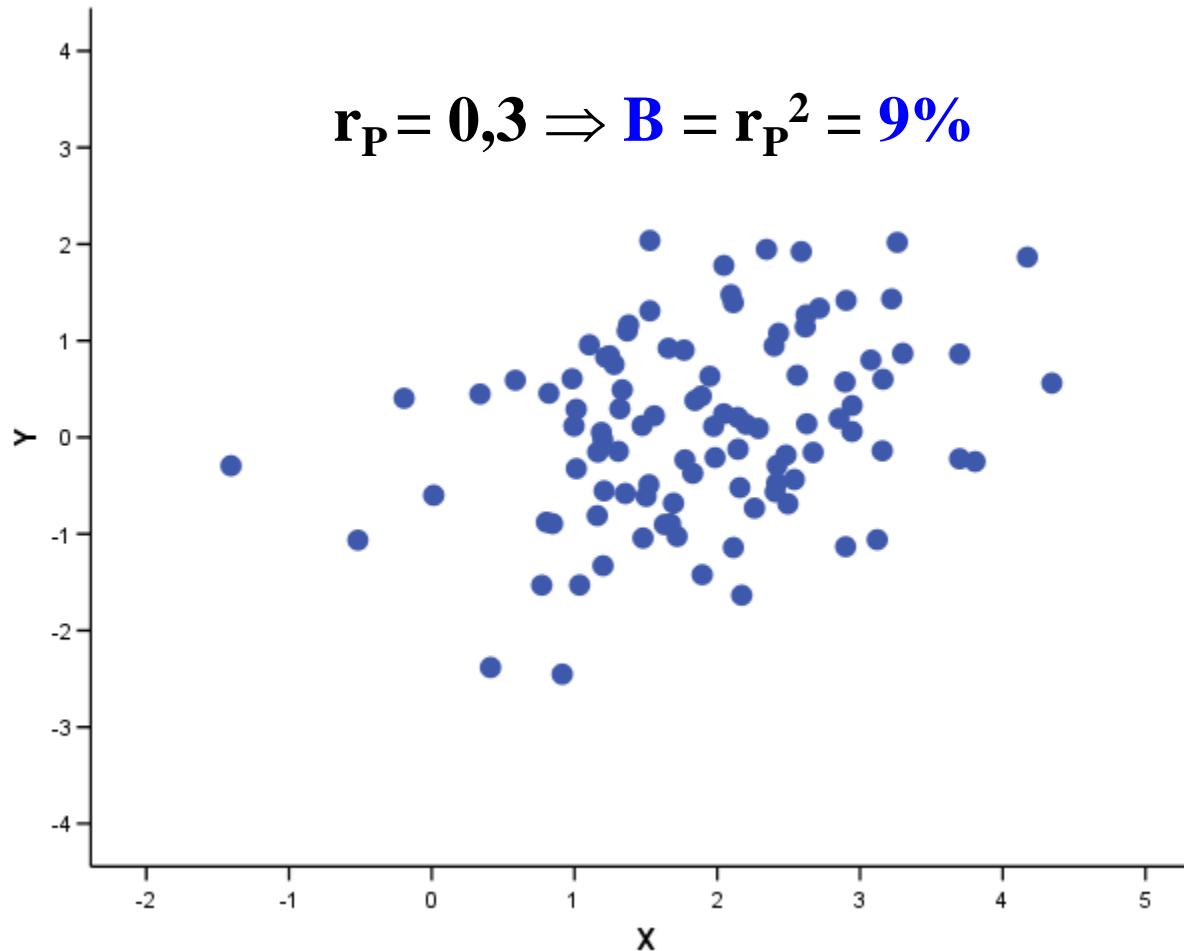
Korrelationskoeffizient und Streudiagramm



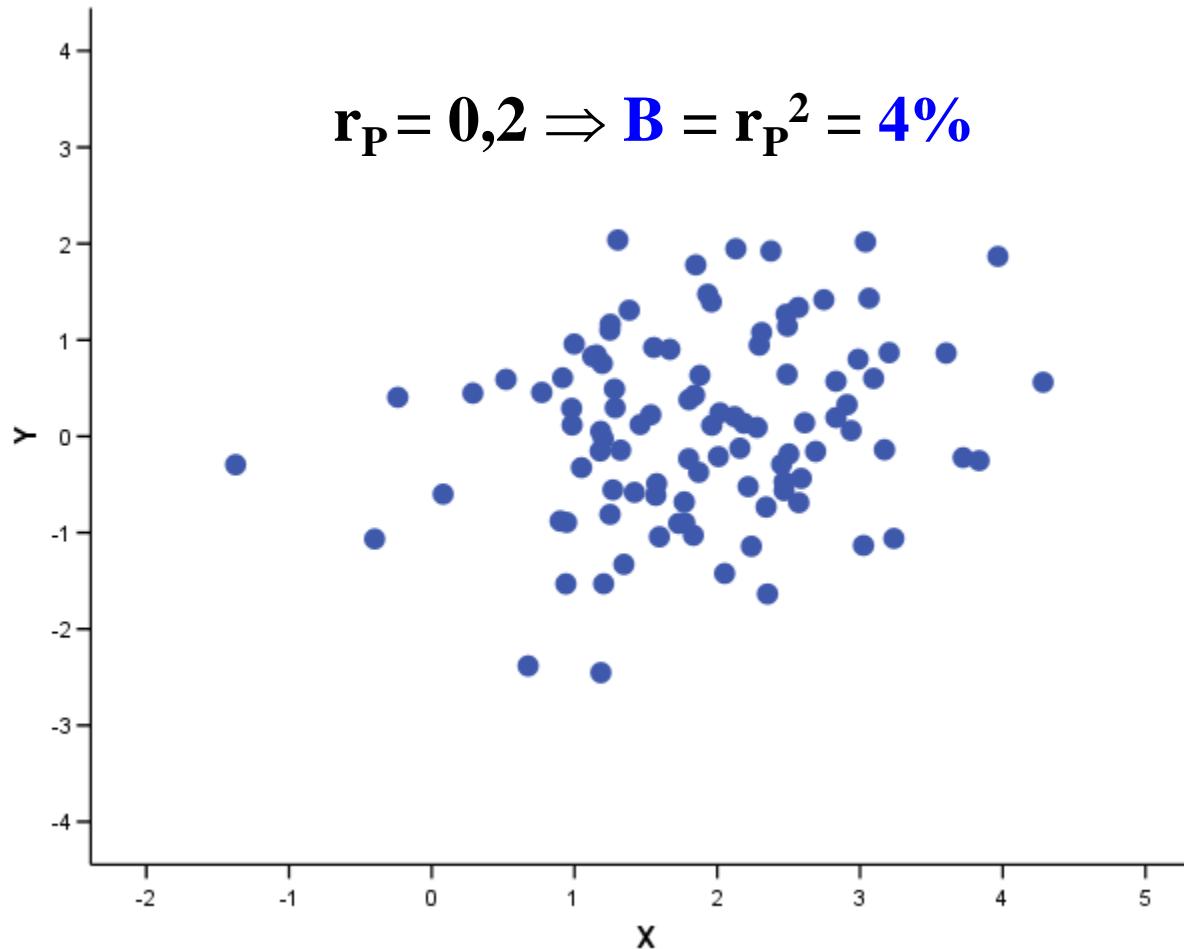
Korrelationskoeffizient und Streudiagramm



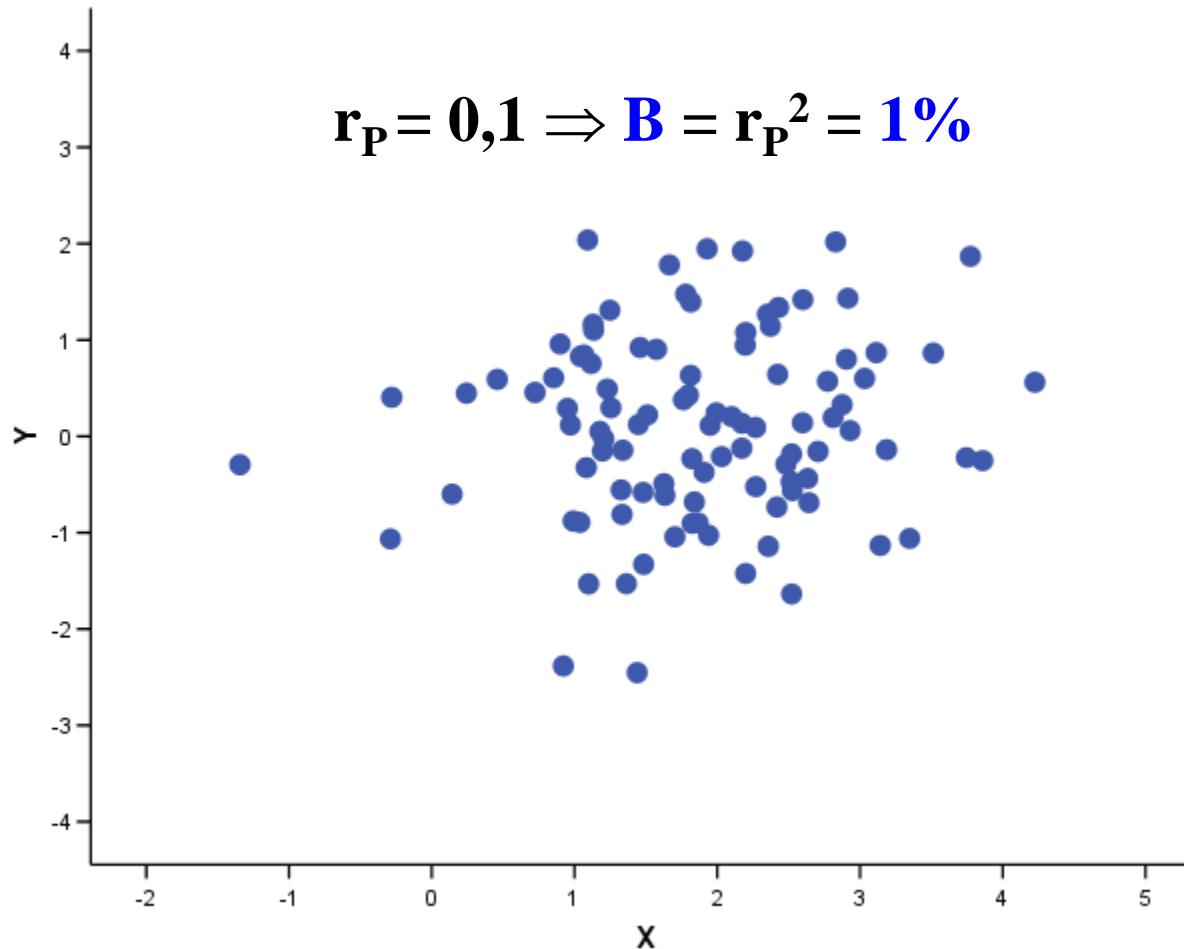
Korrelationskoeffizient und Streudiagramm



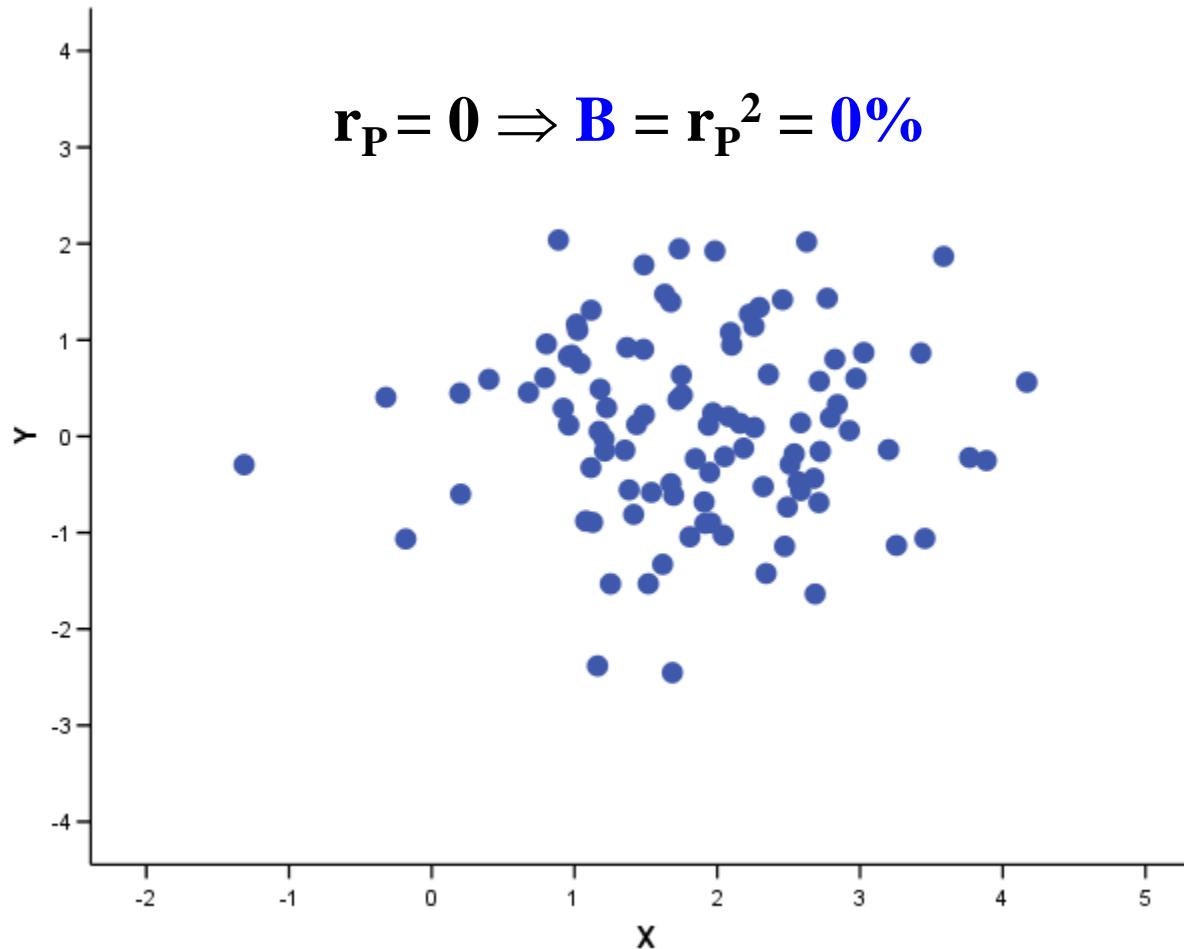
Korrelationskoeffizient und Streudiagramm



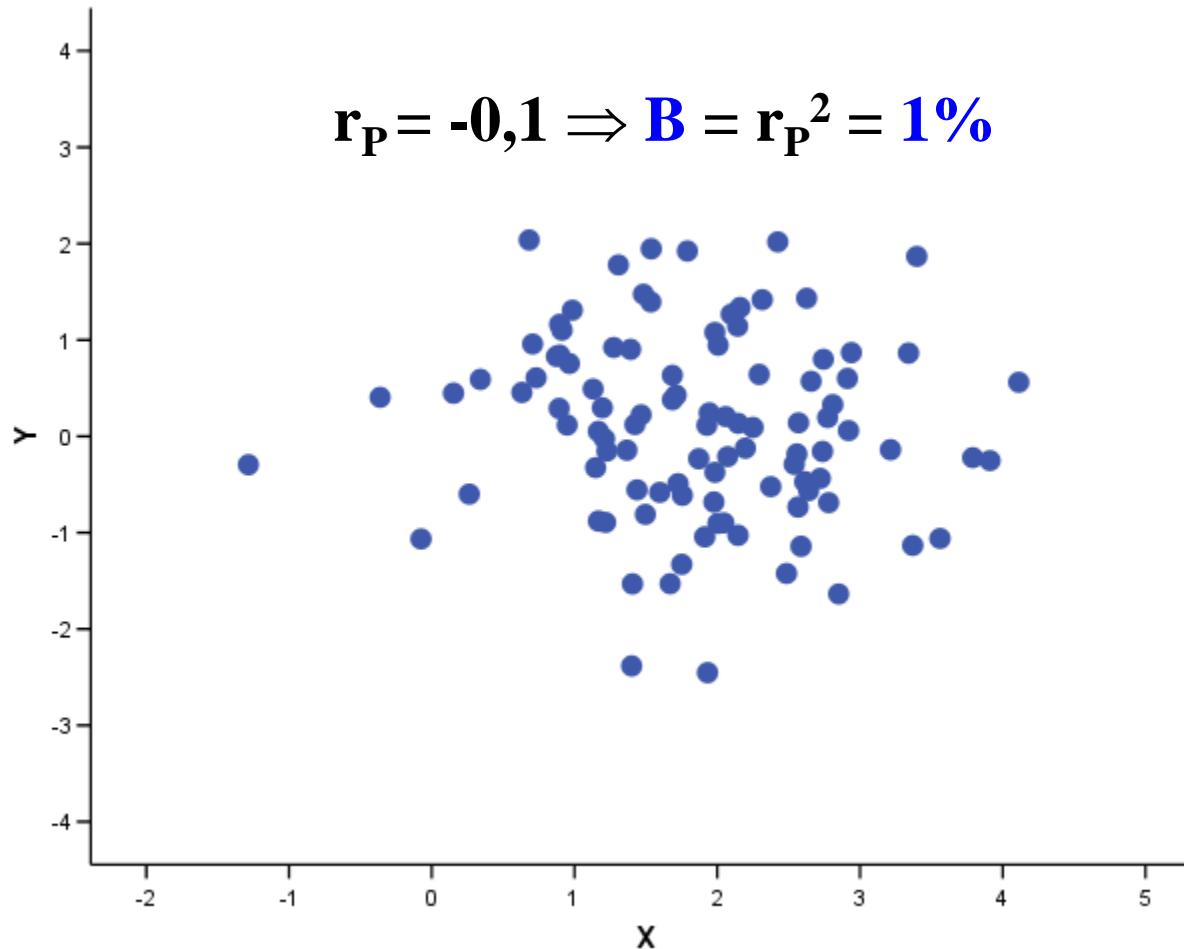
Korrelationskoeffizient und Streudiagramm



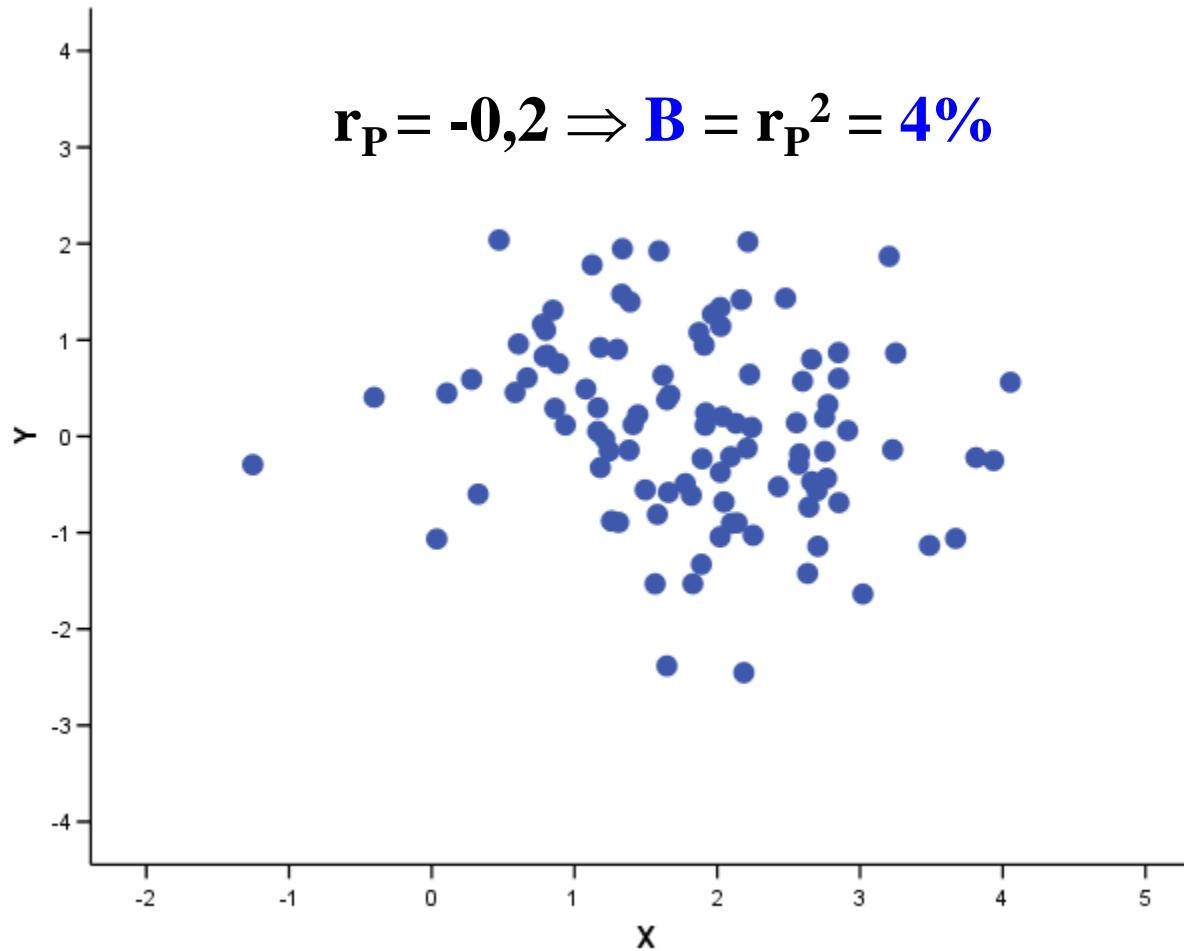
Korrelationskoeffizient und Streudiagramm



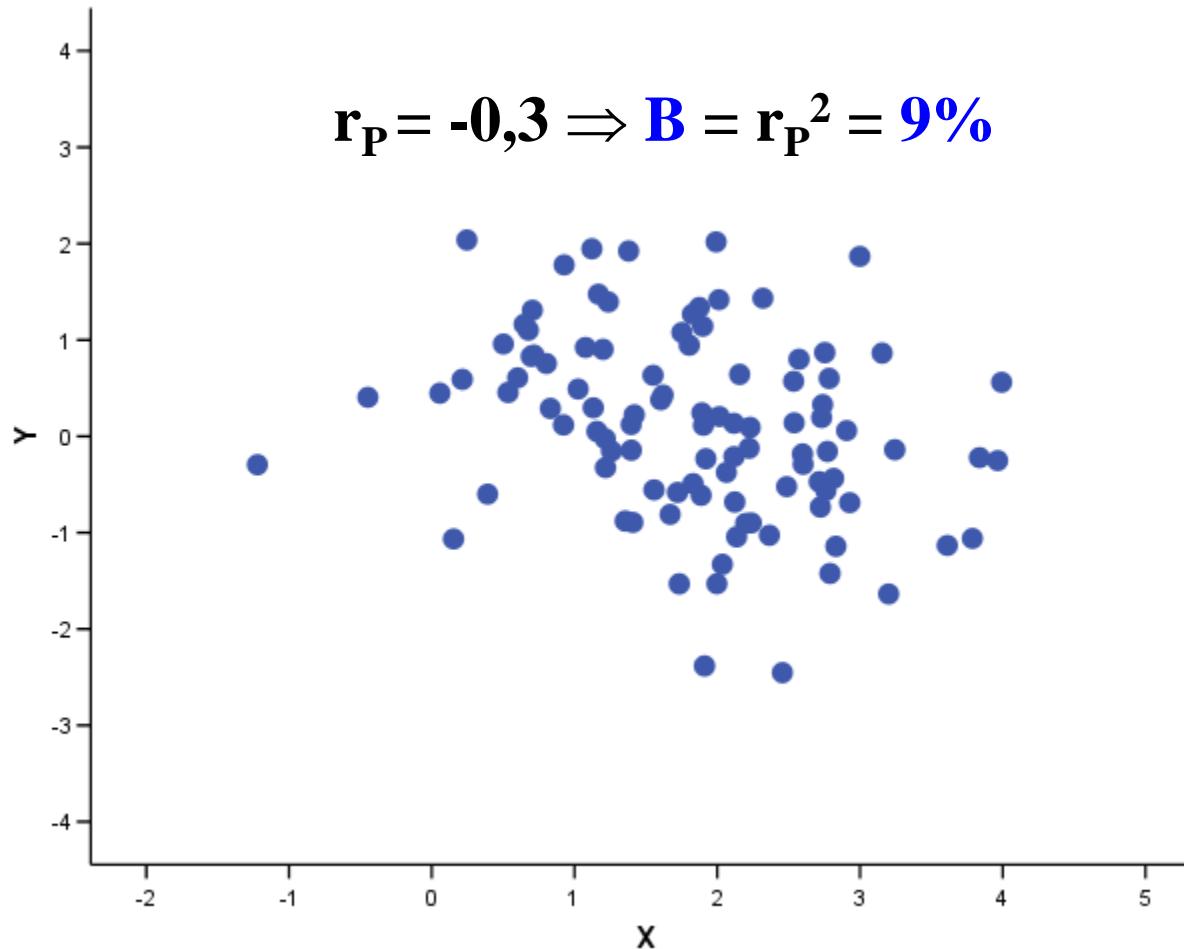
Korrelationskoeffizient und Streudiagramm



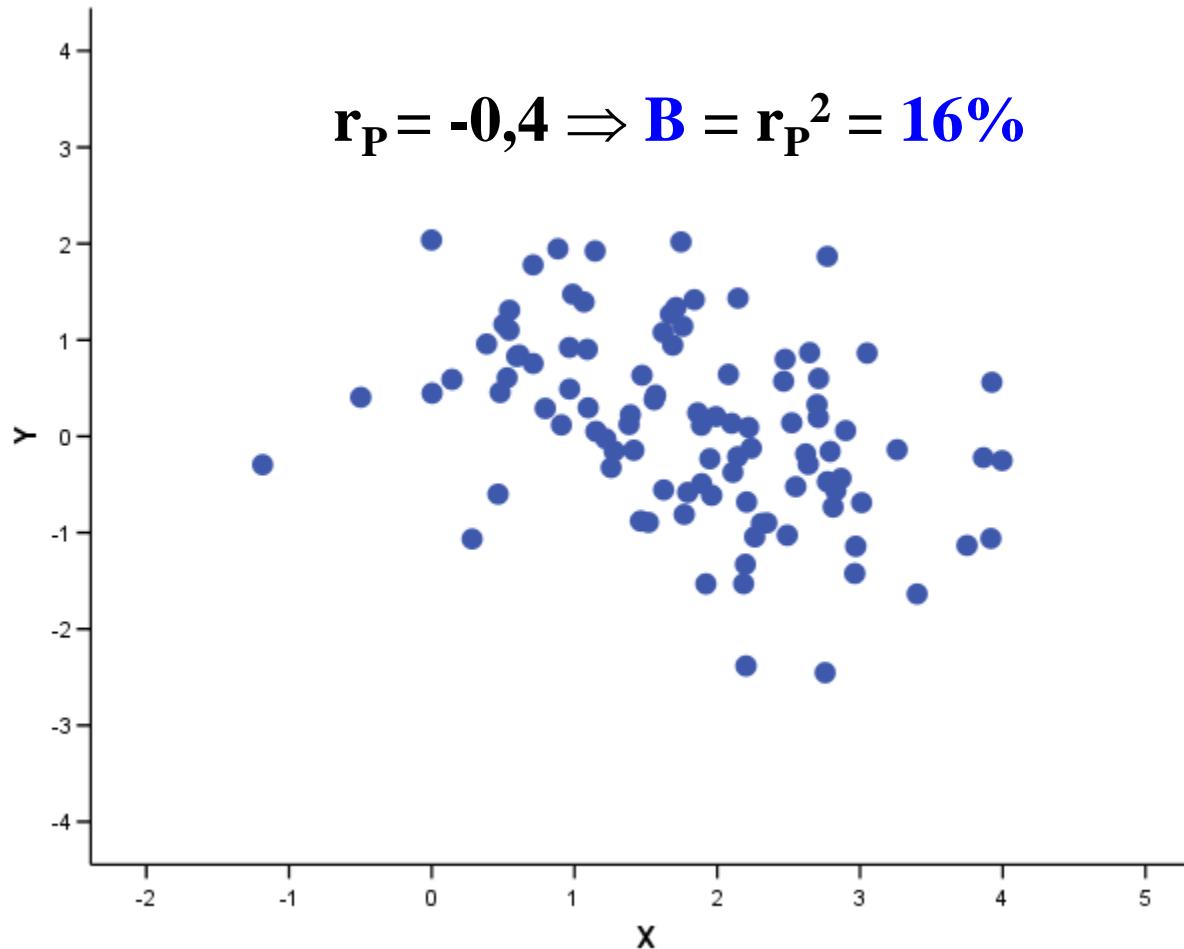
Korrelationskoeffizient und Streudiagramm



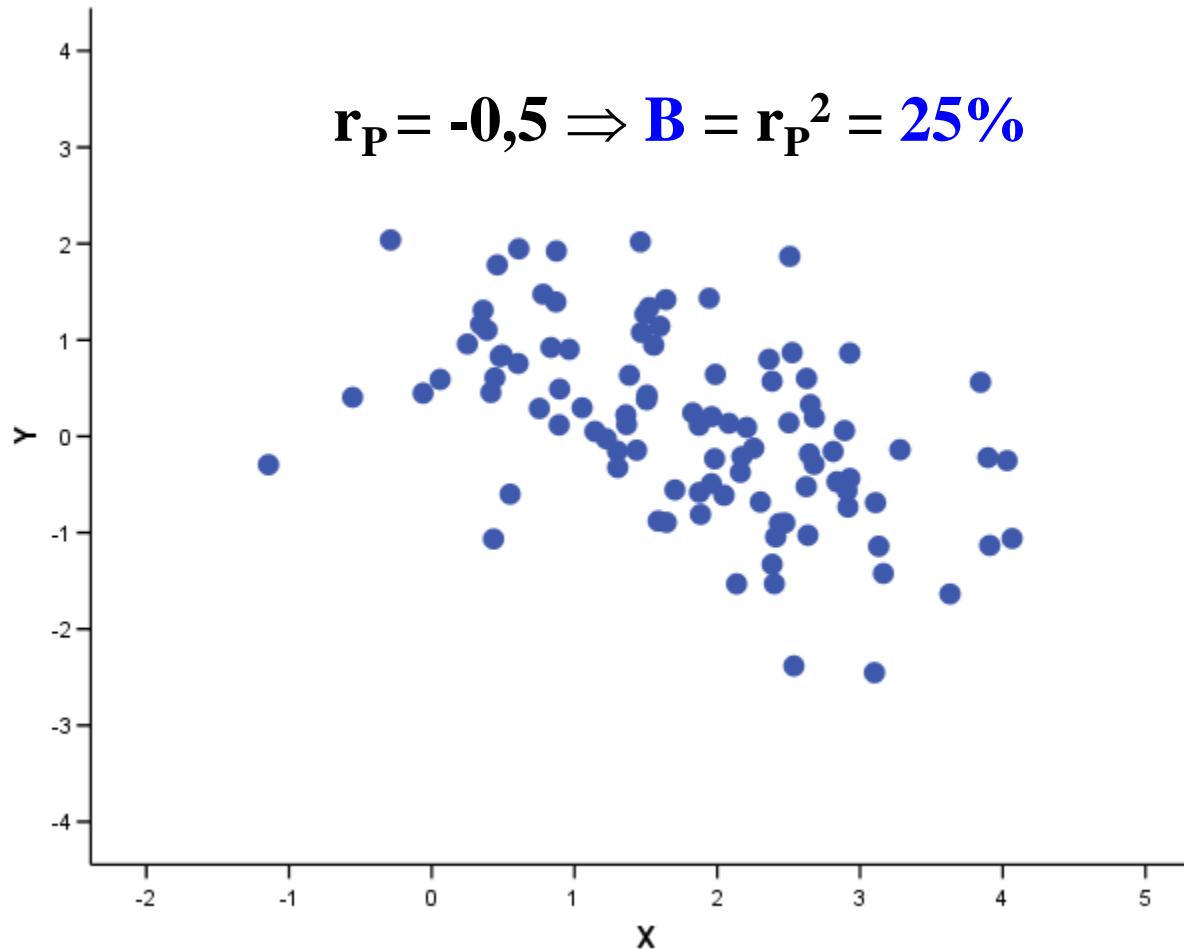
Korrelationskoeffizient und Streudiagramm



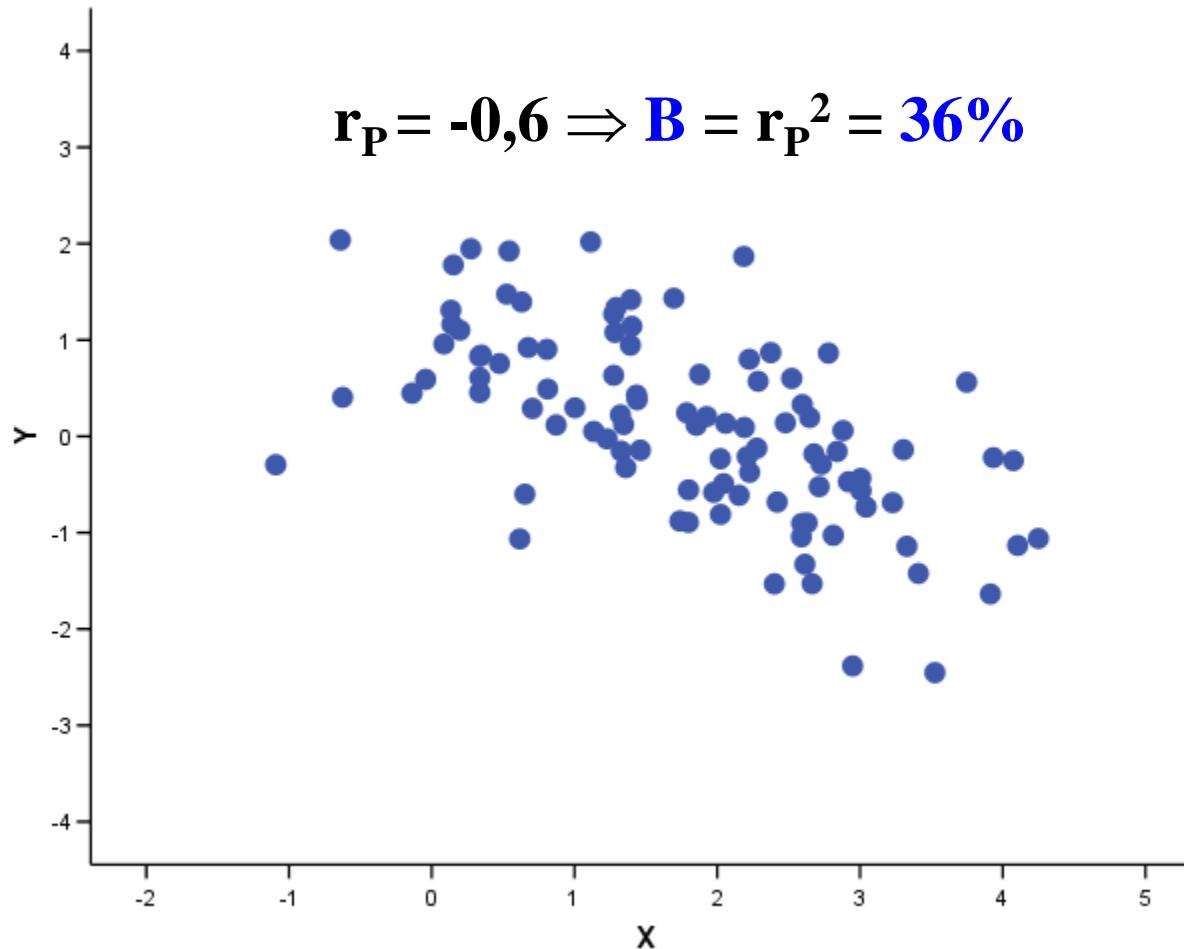
Korrelationskoeffizient und Streudiagramm



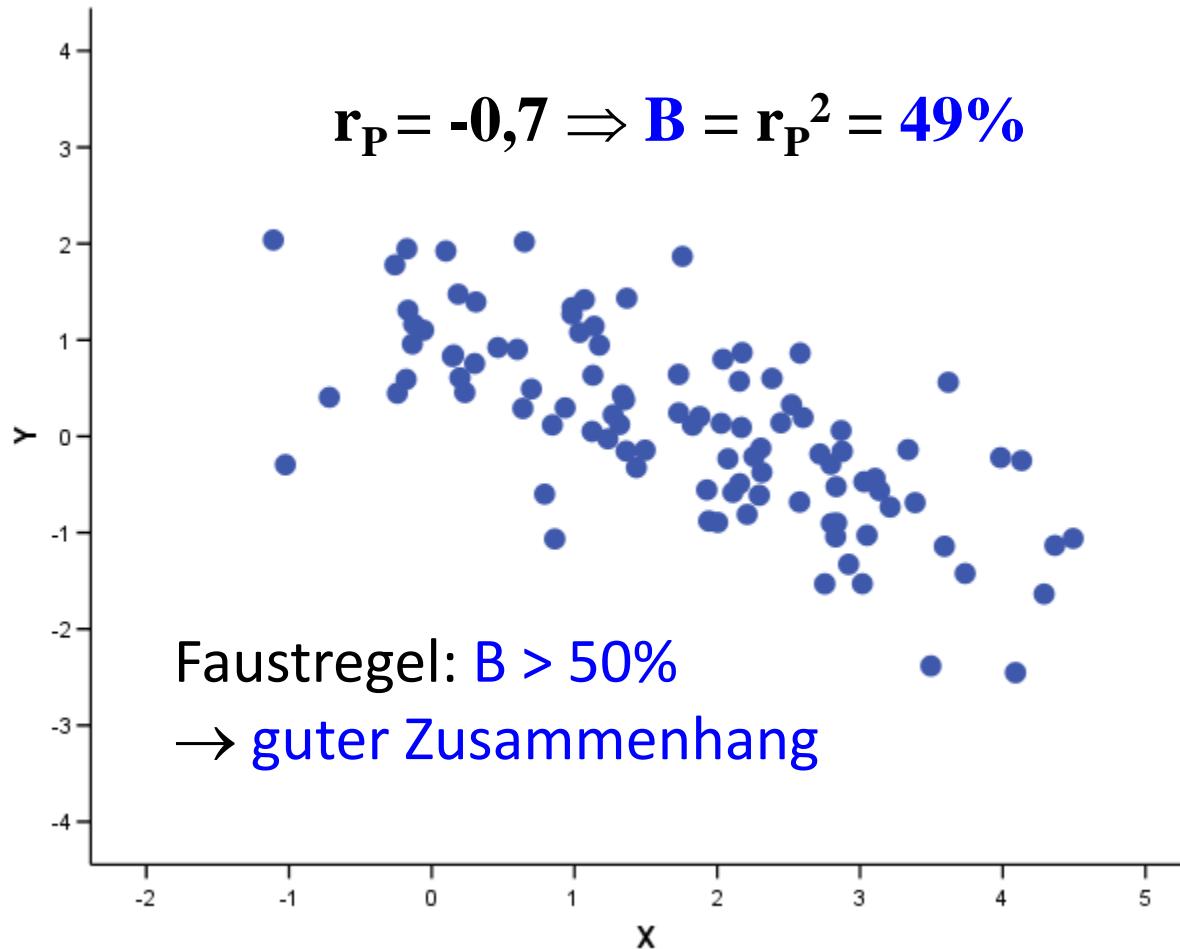
Korrelationskoeffizient und Streudiagramm



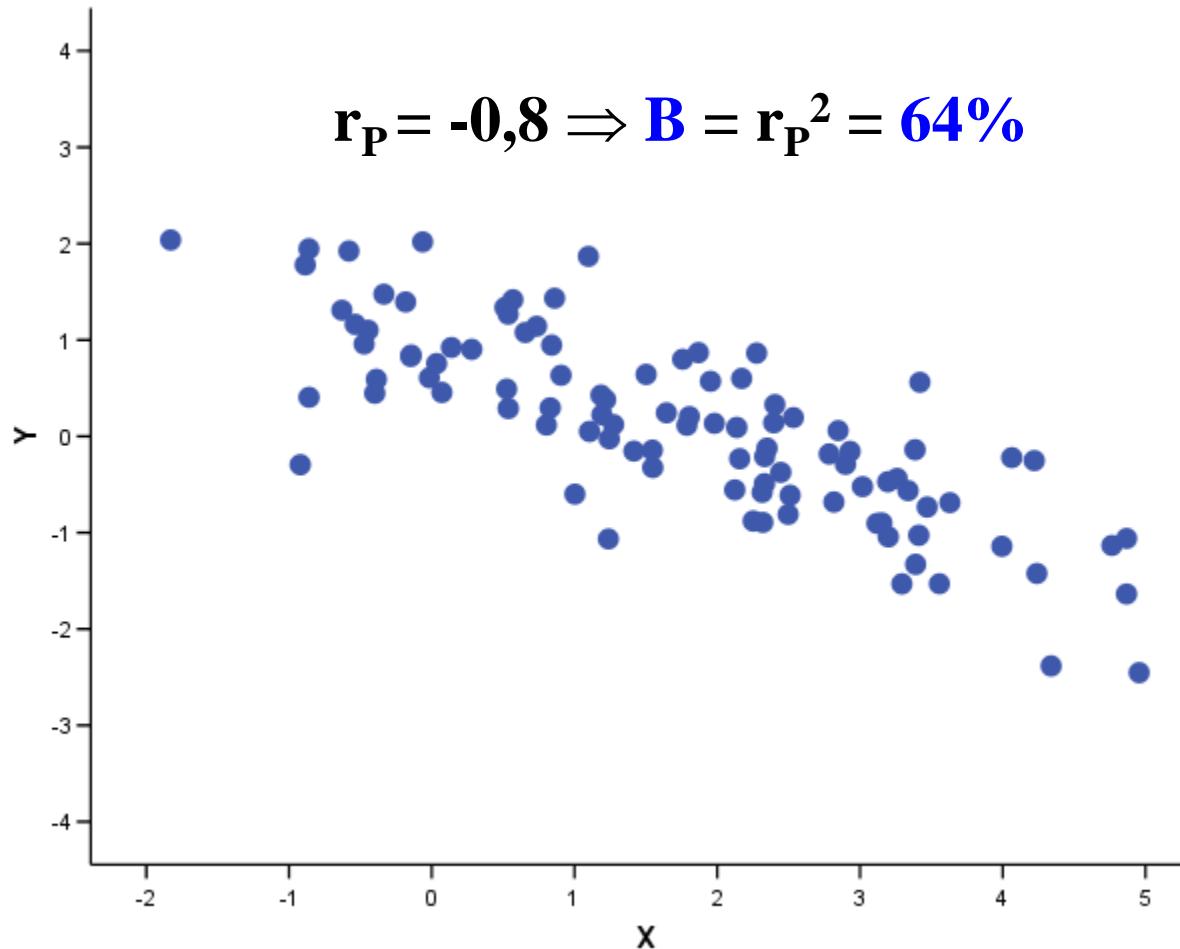
Korrelationskoeffizient und Streudiagramm



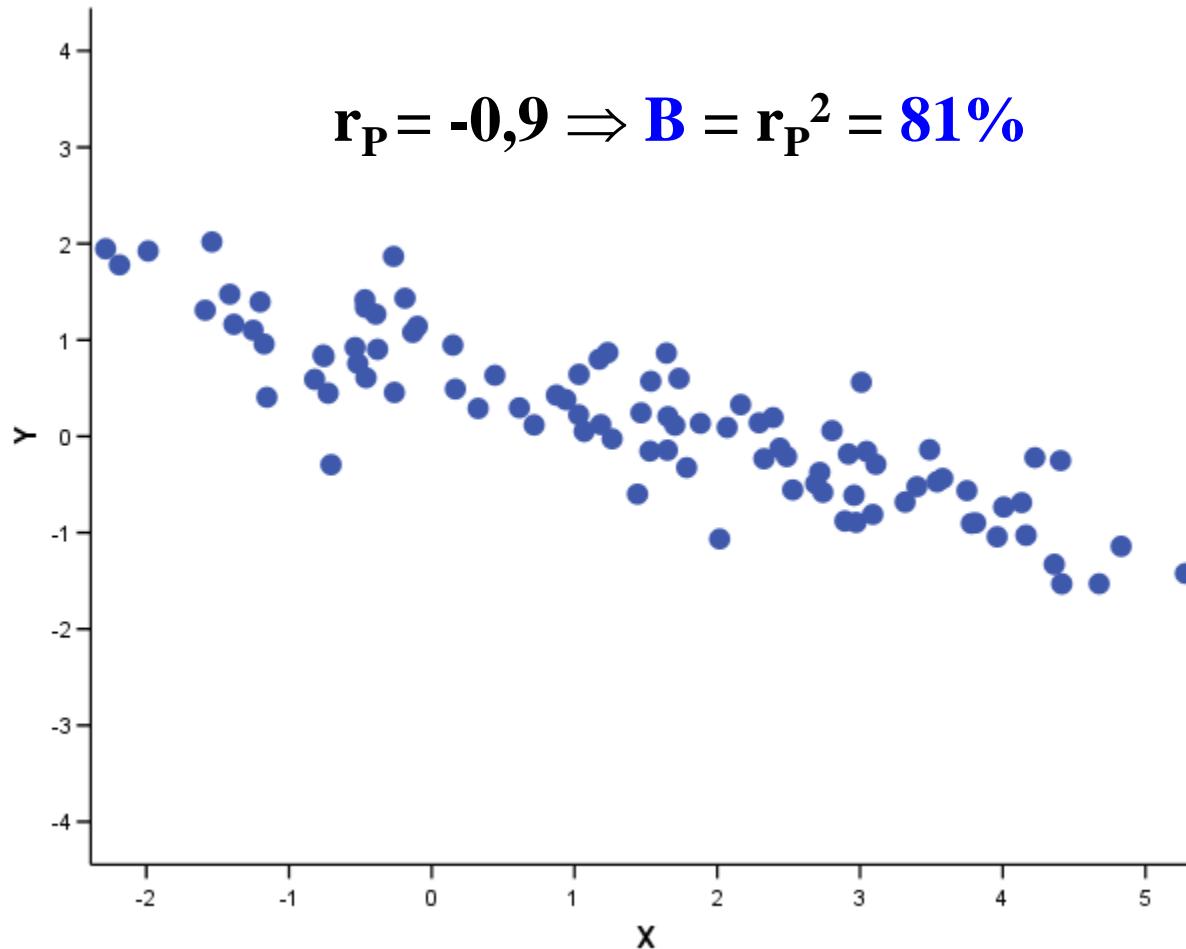
Korrelationskoeffizient und Streudiagramm



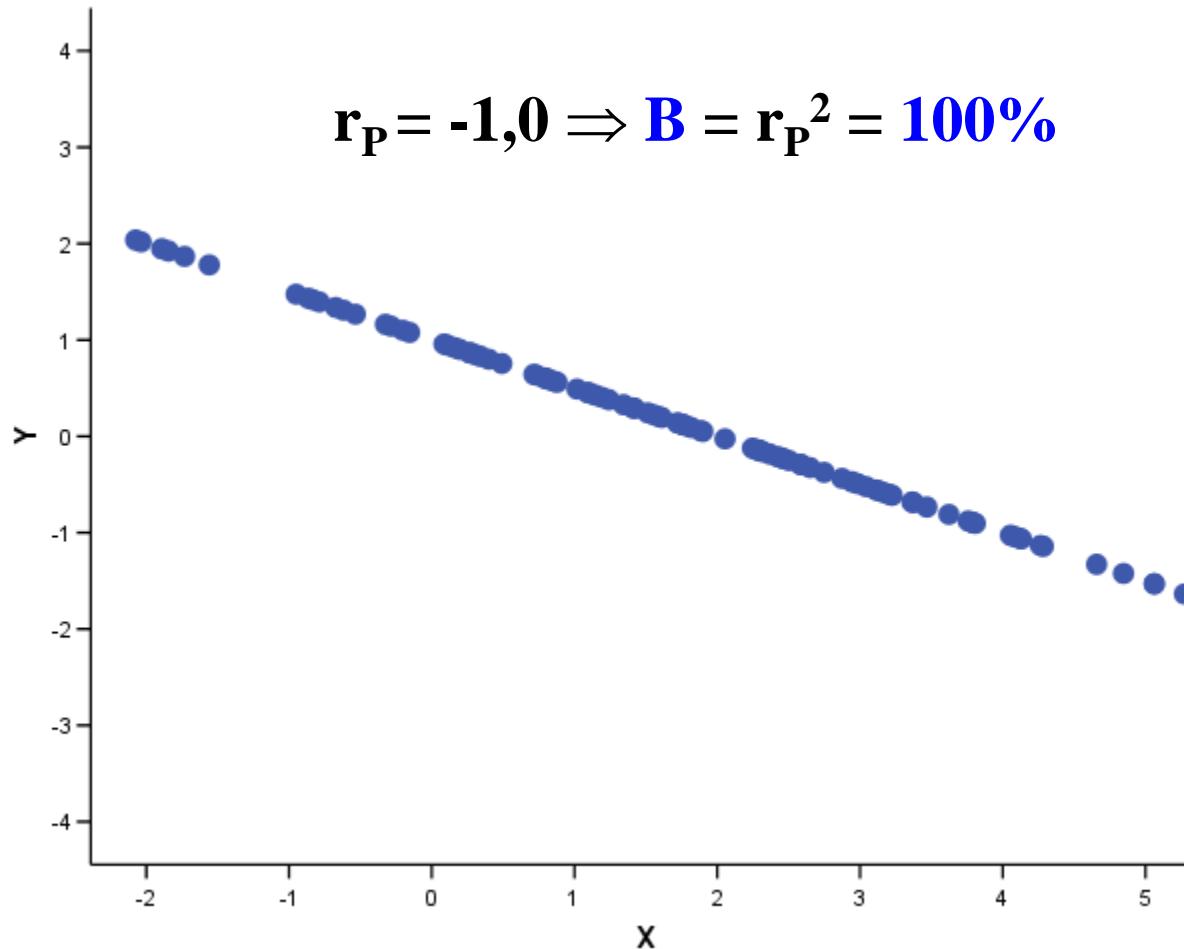
Korrelationskoeffizient und Streudiagramm



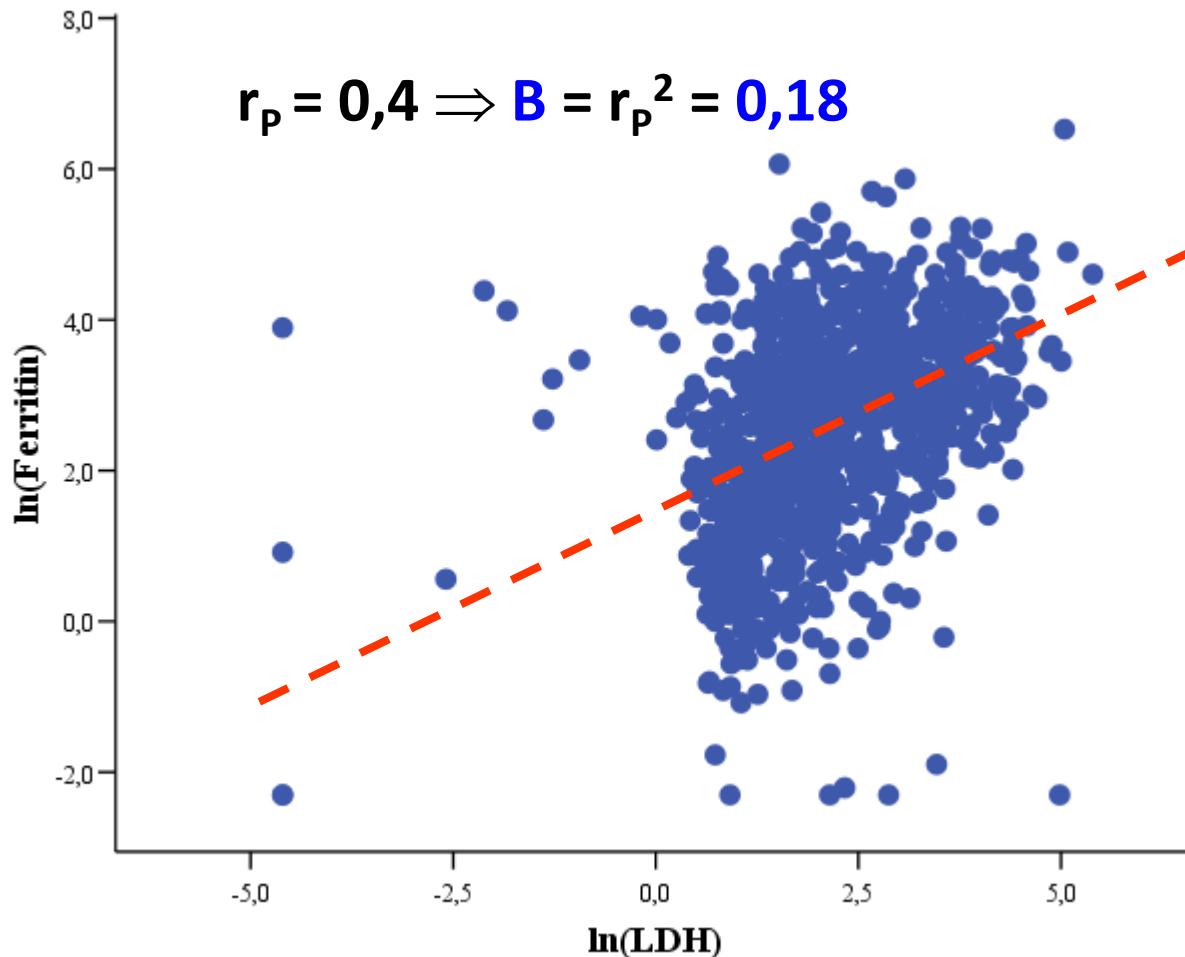
Korrelationskoeffizient und Streudiagramm



Korrelationskoeffizient und Streudiagramm



Zusammenhang zwischen $\ln(\text{LDH})$ und $\ln(\text{Ferritin})$



Multiple lineare Regression

- Beispiel NB 2004: **hängt** $\log(\text{Ferritin})$ von $\log(\text{LDH})$, **Alter**, **1p-Anomalie**, **MYCN** und **INSS** ab
- **Y** = **abhängige Variable** = Zielgröße (z. B. $\log(\text{Ferritin})$)
- X_1, \dots, X_k = **unabhängige Variablen** = Einflussgrößen (z. B. $\log(\text{LDH})$, Alter, 1p-Anomalie, MYCN, INSS)
- Lineares Modell: $Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + \varepsilon$
 - ε : normalverteilt, Erwartungswert = 0, gleiche Varianz
 - β_0 (**Intercept**), β_1, \dots, β_k heißen **Regressionskoeffizienten**
- Sind die **Einflussgrößen** (x_1, \dots, x_k), so ist der **erwartete** Y-Wert $y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$



Multiple lineare Regression - Beispiel

$\log(\text{Ferritin})$ [log(ng/ml)] =

0,7 log(ng/ml)

+0,8 log(ng/ml)/log(U/l) · log(LDH) [log(U/l)]

+0,1 log(ng/ml)/Jahre · Alter [Jahre]

-0,5 log(ng/ml) [falls 1p nicht anomali ist]

+0,7 log(ng/ml) [für MYCN = negativ]

-1,3 log(ng/ml) [für INSS = 1]

-0,7 log(ng/ml) [für INSS = 2]

-0,8 log(ng/ml) [für INSS = 3]

log(LDH): +1 log(U/l) →

log(Ferritin): +0,8 log(ng/ml)

Alter: +1 Jahr →

log(Ferritin): +0,1 log(ng/ml)

falls 1p anomali →
für MYCN = positiv

log(Ferritin): ±0 log(ng/ml)

für INSS = 4 →

log(Ferritin): ±0 log(ng/ml)

Multiple lineare Regression

- Einheiten nicht vergessen
- Keine Extrapolationen: Regressionsgleichung nur gültig für den Wertebereich der Einflussgrößen
- Bestimmtheitsmaß B auch hier gültig
 - $B = 1 - \sum(Y_i - \beta_0 - \beta_1X_1 - \dots - \beta_kX_k)^2 / \sum(Y_i - Y_M)^2$;
 Y_M = Mittelwert von Y
 - Anteil der Varianz von Y, der durch die Regressionsgerade erklärt wird
 - Im Beispiel: B = 61% → gut!

Lineare gemischte Modelle

- Lineares Modell: $Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + \varepsilon$
 - Y = abhängige Variable = Zielgröße (z. B. $\log(\text{Ferritin})$)
 - X_1, \dots, X_k = unabhängige Variablen = Einflussgrößen (z. B. $\log(\text{LDH})$, Alter, 1p-Anomalie, MYCN, INSS)
 - β_0 (Intercept), β_1, \dots, β_k heißen Regressionskoeffizienten
 - ε : Fehlerterm (unerklärtes stochastisches Rauschen)
- Bisherige Annahme:
 - Keine wiederholte Messung an der gleichen statistischen Einheit.
 - Keine Messungen an Clustern verwandter statistischer Einheiten.



Lineare gemischte Modelle

- In der Zahnheilkunde ist diese Annahme oft nicht erfüllt.
- Beispiele:
 - Zielgröße: Plaquerückstand vor/nach Behandlung von Zahnfehlstellungen.
 - Zielgröße: Zahntaschentiefe vor/nach Behandlung von Parodontitis.
- Hier erfolgt die Messung der Zielgröße
 - Separat für jeden Zahn
 - Wiederholt über mehrere Zeitpunkte hinweg
- Folglich: Mehrfachmessung von Clustern von Daten pro Patient



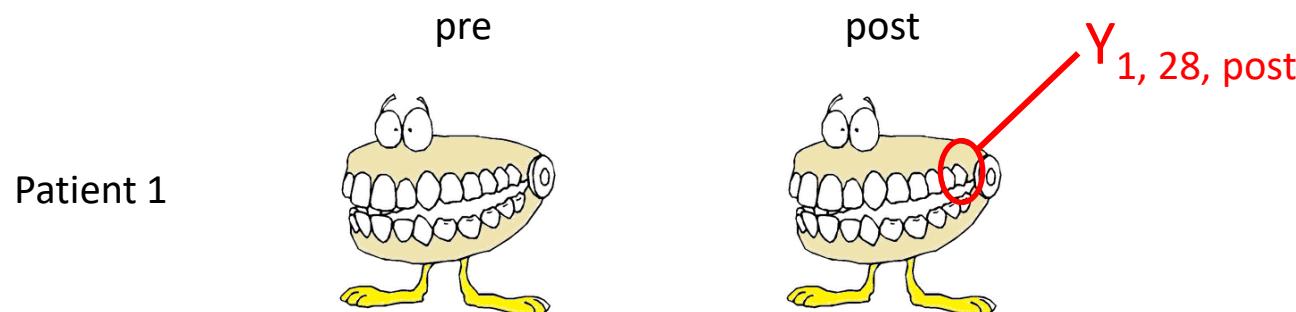
Lineare gemischte Modelle

- Beispiel: Behandlung von Zahnfehlstellung
 - Beurteilung des Ausmaßes der Zahnfehlstellung durch Plaque-Rückstand nach erfolgter Zahnreinigung
 - Versuchsablauf: Bei jedem Patienten wird zu jedem Messzeitpunkt ein Gipsabdruck des Gebisses angefertigt.
 - Im Abdruck werden die Zähne mit Titandioxid belegt und von einem Putzroboter gereinigt.
 - Zielgröße: TDC = Relativer Anteil der Zahnfläche, die nach erfolgter Reinigung noch mit Titandioxid bedeckt ist.
- Fragestellung: Ist ein neues Verfahren (N) dem Standardverfahren (S) zur Behandlung von Zahnfehlstellung überlegen?



Lineare gemischte Modelle

- Zielgröße wird pro Patient mehrfach erhoben: Beurteilung des $Y_{i, j, k}$ = TDC bei Zahn j von Patient i zum Zeitpunkt k
- $Y_{1, 28, \text{post}}$ = gemessene TDC bei
 - Zahn 28 (gemäß FDI-Zahnschema nach Viohl)
 - Von Patient Nr. 1 nach Therapie-Ende



Lineare gemischte Modelle

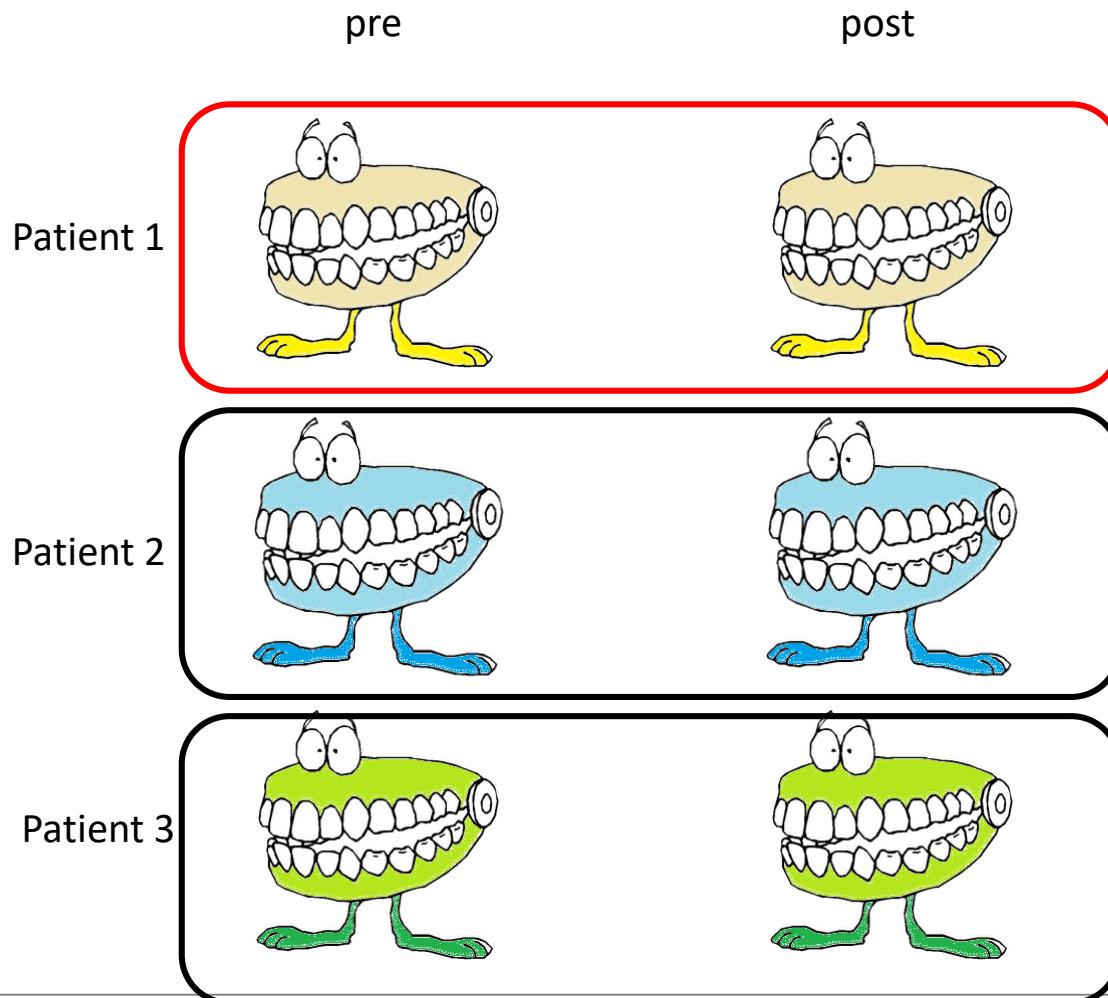
- Zielgröße wird pro Patient mehrfach erhoben:
 - $Y_{i,j,k}$ = TDC bei Zahn j von Patient i zum Zeitpunkt k
- Einflussgröße:
 - X_i Therapiegruppen-Indikator von Patient i
 - $X_i = 1$, wenn Patient i aus Therapiegruppe N; $X_i = 0$, sonst.
 - Hängt nicht ab von j und k.
- Marginales Modell: $Y_{i,j,k} = \beta_0 + \beta_1 X_i + \varepsilon_{i,j,k}$
- Zentraler Unterschied zu bisherigen Modellannahme:
 - Fehlerterme $\varepsilon_{i,j,k}$ und $\varepsilon_{i,j',k'}$ nicht mehr alle unabhängig!!!
 - Informationen vom selben Patienten sind abhängig!!!



Lineare gemischte Modelle

Diagram illustrating the independence of patient information. Three rows of cartoon teeth characters (Patient 1: yellow, Patient 2: blue, Patient 3: green) show 'pre' and 'post' states. A blue arrow points from the 'post' state of Patient 1 to a text box stating: 'Informationen von verschiedenen Patienten sind unabhängig.'

Lineare gemischte Modelle

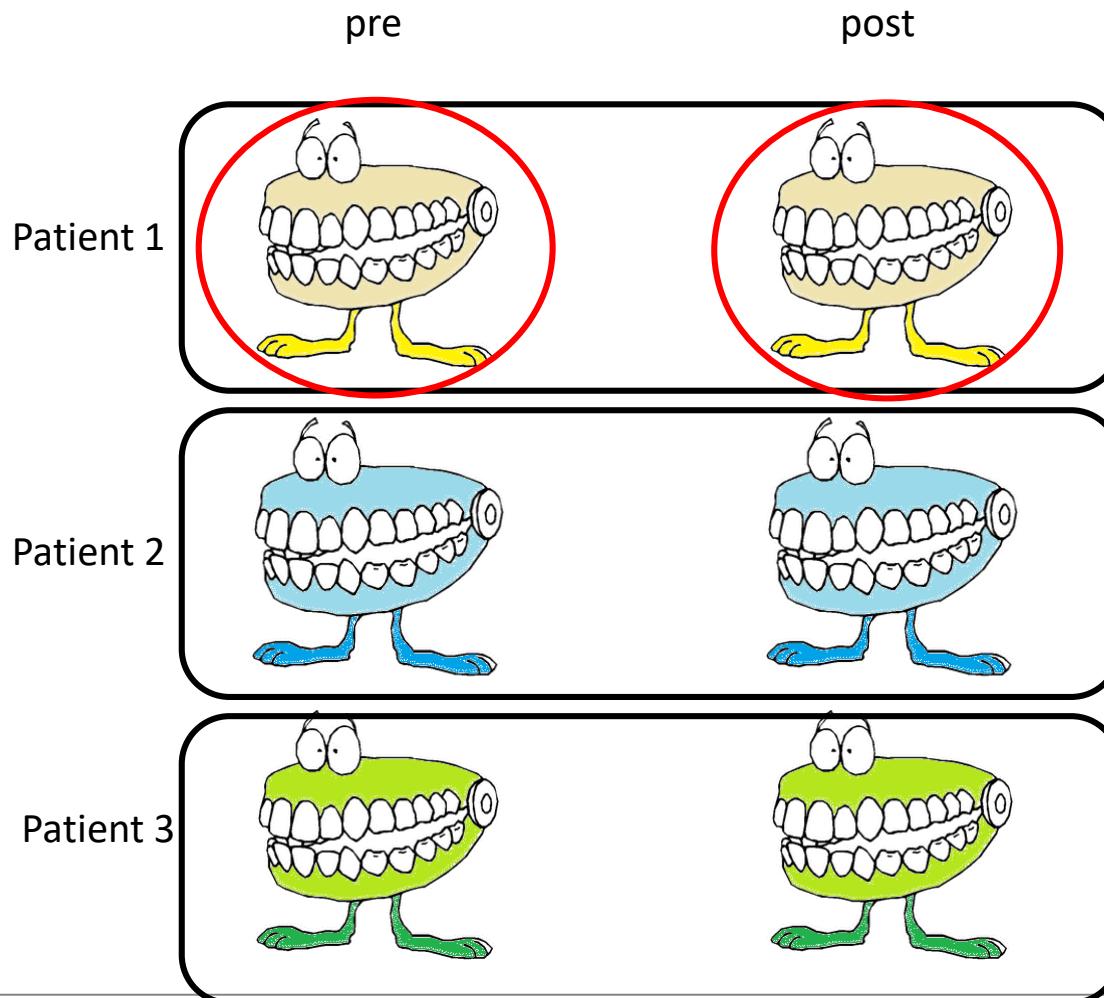


Aber:

Informationen, die
vom selben
Patienten stammen,
sind i.A. abhängig!!!



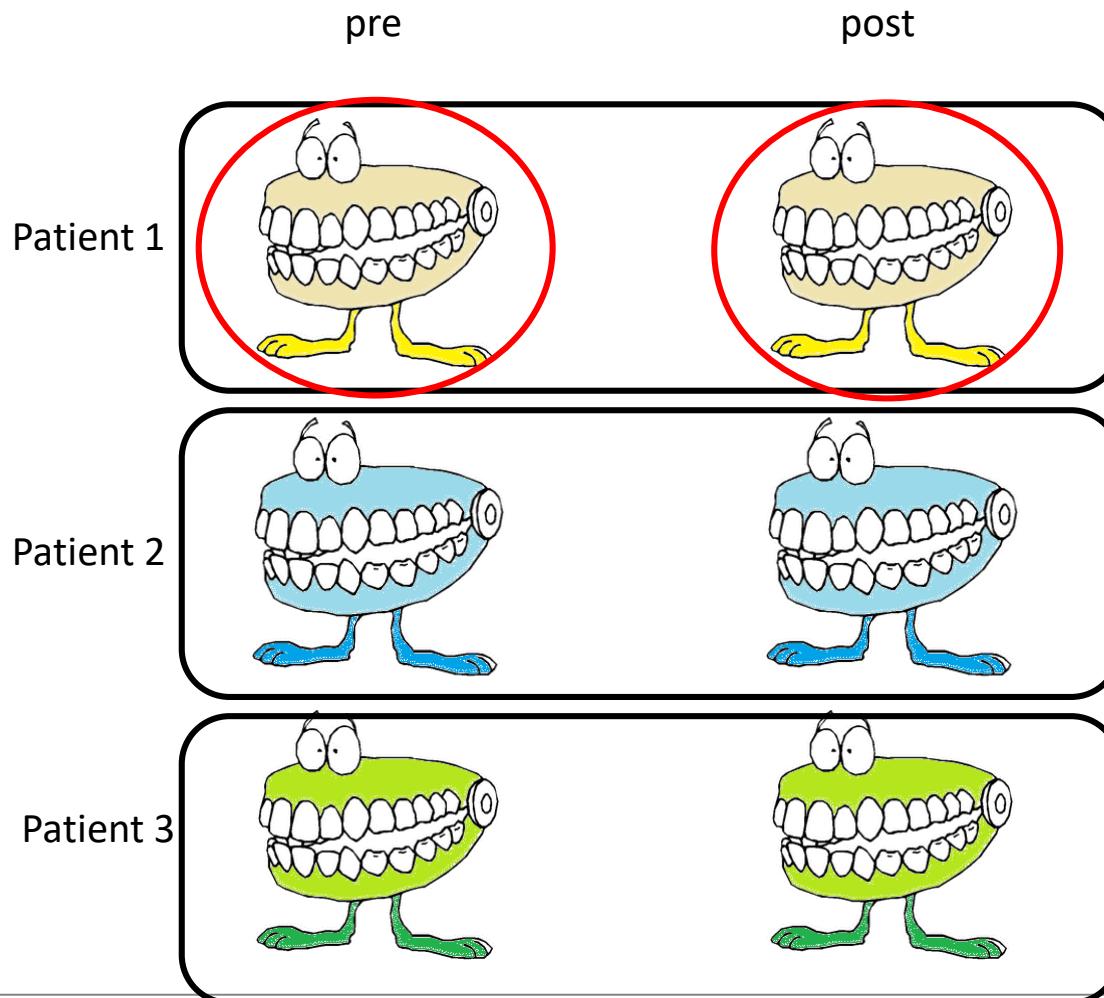
Lineare gemischte Modelle



Der Zustand des Gebisses zu einem späteren Zeitpunkt hängt i.A. noch vom Ausgangszustand ab.



Lineare gemischte Modelle



Clusterstruktur der Daten:

Zähne von Patient 1
pre-treatment
Und

Zähne von Patient 1
Post-treatment

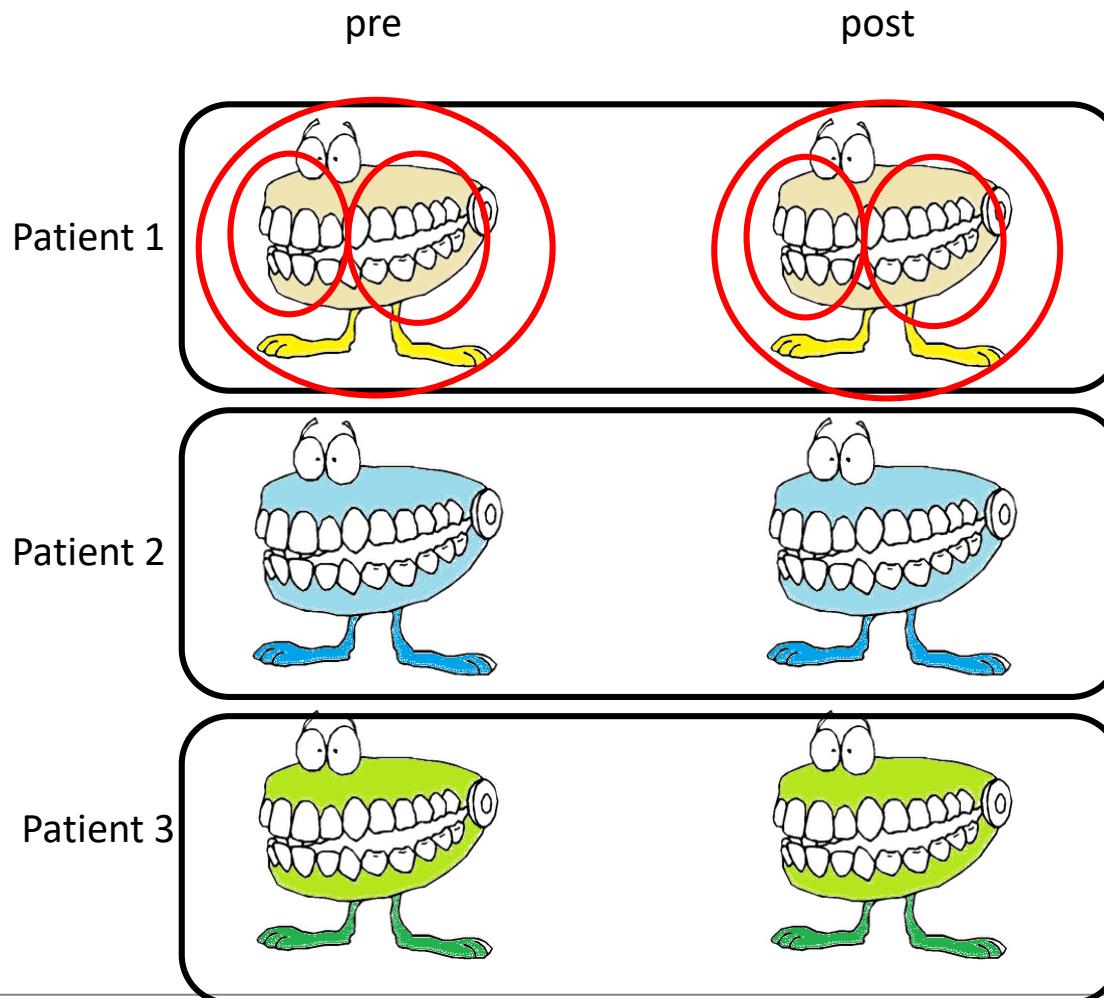
sind sich ähnlicher als

Zähne von Patient 1 pre-
treatment
Und

Zähne von Patient 2
Post-treatment



Lineare gemischte Modelle



Cluster-Struktur lässt sich hierarchisch fortsetzen:

Cluster innerhalb von Clustern:

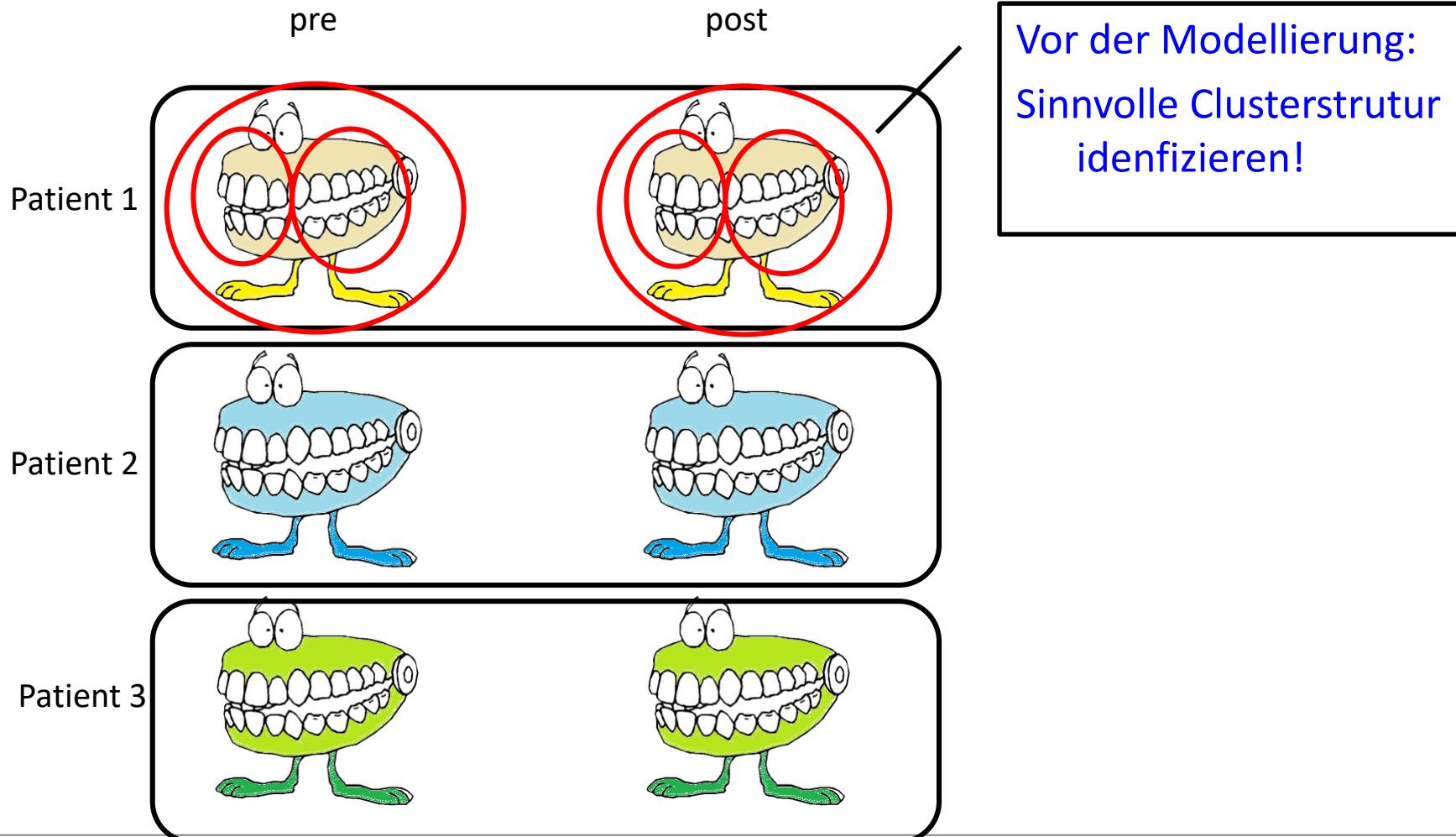
Zu jedem Zeitpunkt gilt:

Schneidezähne

sind sich ähnlicher als

Schneidzahn Und Backenzahn

Lineare gemischte Modelle



Lineare gemischte Modelle

- Fazit:
 - Bei wiederholter Messung an derselben statistischen Einheit...
 - Bei Messungen an Clustern verwandter statistischer Einheiten...
- ... ist Abhängigkeitsstruktur zu bedenken.
- Informationsgehalt der Daten ist entsprechend zu gewichten:
 - Hochkorrelierte Messungen steuern weniger Information bei als unabhängige Messungen
- Wechselseitige Einflussprozesse innerhalb der Cluster ist zu berücksichtigen
- Vernachlässigung dessen kann zur Fehlinterpretation von empirischen Ergebnissen führen.



Modellierungsansätze bei Abhängigkeiten

- Einfachste Lösung (nicht immer möglich, oft ineffizient) : Aggregation der Daten
 - Betrachte nur Änderungen über Zeitintervalle:
 - Bilde $\Delta Y_{i,j} = Y_{i,j, \text{pre}} - Y_{i,j, \text{post}}$
 - Bildung von Mittelwerten innerhalb von Clustern:
 - Bilde $\Delta Y_i = \sum_j \Delta Y_{i,j} / |j|$
- Betrachte aggregiertes Outcome ΔY_i statt $Y_{i,j,k}$ als Zielgröße:
 - Dann: Wieder nur noch eine Beobachtung pro Patient, d.h. alle Beobachtungen sind wieder unabhängig!
 - Können ein gewöhnliches lineares Modell verwenden.



Modellierungsansätze bei Abhängigkeiten

- Einfachste Lösung: Aggregation der Daten
- Cave:
- Mittelwertwertbildung $\Delta Y_i = \sum_j \Delta Y_{i,j} / |j|$ erfordert:
 - Gleiche Clustergröße bei jedem Patient (z.B. nicht gegeben im Falle von fehlenden Zähne)
 - Ist nur sinnvoll innerhalb eines Clusters, aber nicht möglich über Cluster unterschiedlicher Hierarchie hinweg.



Lineare gemischte Modelle

- Beste Lösung: Anwendung eines gemischten linearen Modells
- Erlaubt valide Interpretation von empirischen Ergebnissen in Anwesenheit von Abhängigkeiten.
- Ist robust gegenüber Fehlspezifikation der Abhängigkeitsstruktur (sandwich estimator, relevant für die Praxis!)
- Bietet maximale Flexibilität. Erlaubt:
 - Unterschiedliche Clustergröße von Patient zu Patient (z.B. im Falle fehlender Zähne)
 - Erlaubt (geschachtelte) Cluster unterschiedlicher Hierarchie.
 - Erlaubt beliebig viele Messzeitpunkte (unterschiedlicher Anzahl von Patient zu Patient).



Lineare gemischte Modelle

- Beispiel:
 - Beurteilung des Ausmaßes der Zahnfehlstellung durch Plaque-Rückstand nach erfolgter Zahncleaning
 - Versuchsablauf: Bei jedem Patienten wird zu jedem Messzeitpunkt ein Gipsabdruck des Gebisses angefertigt.
 - Im Abdruck werden die Zähne mit Titandioxid belegt und von einem Putzroboter gereinigt.
 - TDC = Relativer Anteil der Zahnoberfläche, die nach erfolgter Reinigung noch mit Titandioxid bedeckt ist.



Lineare gemischte Modelle

- Beispiel: Behandlung von Zahnfehlstellung
- Fragestellung: Ist ein neues Verfahren (N) dem Standardverfahren (S) zur Behandlung von Zahnfehlstellung überlegen?
- Quantifizierung des Behandlungserfolges mittels:
 - Zielgröße: $\Delta Y_{i,j} = Y_{i,j,pre} - Y_{i,j,post}$: Änderung der TDC bei Zahn j von Patient i im Vergleich „vor versus nach“ Behandlung
- Einflussgröße:
 - $X_i = 1\{\text{Patient } i \text{ in Gruppe } N\}$: Therapiegruppen-Indikator
 - Marginales Modell: $\Delta Y_{i,j} = \beta_0 + \beta_1 X_i + \varepsilon_{i,j}$
- Nullhypothese: $H_0 : \beta_1 = 0$.



Lineare gemischte Modelle

- Erforderliche Datenstruktur für gemischte lineare Modelle:

	PatID	Behandlungsgruppe	Plaque	FDI_Zahnschema	Messzeitpunkt
1	1	S	12,65	21	before treatment
2	1	S	18,10	22	before treatment
3	1	S	17,50	23	before treatment
4	1	S	20,90	24	before treatment
5	1	S	21,60	25	before treatment
6	1	S	22,30	26	before treatment
7	1	S	30,80	27	before treatment
8	1	S	32,40	28	before treatment
9	1	S	15,70	21	after treatment
10	1	S	13,11	22	after treatment
11	1	S	11,92	23	after treatment
12	1	S	23,10	24	after treatment
13	1	S	16,20	25	after treatment
14	1	S	15,80	26	after treatment
15	1	S	26,80	27	after treatment
16	1	S	34,40	28	after treatment

Long format



Lineare gemischte Modelle

- Ergebnis unter Berücksichtigung der Abhängigkeitsstuktur:

Estimates of Fixed Effects^a

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	18,536531	1,362637	26,034	13,603	,000	15,735766	21,337295
[Behandlungsgruppe=1, 00]	3,469899	1,926393	25,997	1,801	,083	-,489881	7,429679
[Behandlungsgruppe=2, 00]	0 ^b	0					

Estimates of Covariance Parameters^a

Parameter	Estimate	Std. Error
Residual	55,393677	4,094832
Intercept [subject = PatID] Variance	22,038216	7,210932

- Ergebnis bei Ignorierung der Abhängigkeitsstuktur:

Estimates of Fixed Effects^a

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	18,536531	,622513	392	29,777	,000	17,312649	19,760412
[Behandlungsgruppe=1, 00]	3,429075	,878140	392	3,905	,000	1,702622	5,155529
[Behandlungsgruppe=2, 00]	0 ^b	0					



Lineare gemischte Modelle

- Ergebnis unter Berücksichtigung der Abhängigkeitsstuktur:

Estimates of Fixed Effects ^a						
Parameter	Estimate	Std. Error	df	t	Sig.	95%
						Lower Upper
Intercept	18,536531	1,362637	26,034	13,603	,000	15,87
[Behandlungsgruppe=1, 00]	3,469899	1,926393	25,997	1,801	,083	-
[Behandlungsgruppe=2, 00]	0 ^b	0				

Estimates of Covariance Parameters ^a		
Parameter	Estimate	Std. Error
Residual	55,393677	4,094832
Intercept [subject = PatID] Variance	22,038216	7,210932

Kein „signifikanter“
Behandlungseffekt

- Ergebnis bei Ignorierung der Abhängigkeitsstuktur:

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	18,536531	,622513	392	29,777	,000	17,312649	19,760412
[Behandlungsgruppe=1, 00]	3,429075	,878140	392	3,905	,000	1,702622	5,155529
[Behandlungsgruppe=2, 00]	0 ^b	0					



Lineare gemischte Modelle

- Ergebnis unter Berücksichtigung der Abhängigkeitsstuktur:

Estimates of Fixed Effects ^a						
Parameter	Estimate	Std. Error	df	t	Sig.	
Intercept	18,536531	1,362637	26,034	13,603	,000	95
[Behandlungsgruppe=1, 00]	3,469899	1,926393	25,997	1,801	,083	15
[Behandlungsgruppe=2, 00]	0 ^b	0				-

Estimates of Covariance Parameters ^a		
Parameter	Estimate	Std. Error
Residual	55,393677	4,094832
Intercept [subject = PatID] Variance	22,038216	7,210932

Kein „signifikanter“
Behandlungseffekt

- Ergebnis bei Ignorierung der Abhängigkeitsstuktur:

Estimates of Fixed Effects ^a						
Parameter	Estimate	Std. Error	df	t	Sig.	
Intercept	18,536531	,622513	392	29,777	,000	95
[Behandlungsgruppe=1, 00]	3,429075	,878140	392	3,905	,000	17
[Behandlungsgruppe=2, 00]	0 ^b	0				1

Fälschlicherweise
„signifikanter“
Behandlungseffekt



Lineare gemischte Modelle

- Ergebnis unter Berücksichtigung der Abhängigkeitsstuktur:

Estimates of Fixed Effects ^a						
Parameter	Estimate	Std. Error	df	t	Sig.	
Intercept	18,536531	1,362637	26,034	13,603	,000	95
[Behandlungsgruppe=1, 00]	3,469899	1,926393	25,997	1,801	,083	15
[Behandlungsgruppe=2, 00]	0 ^b	0				-

Estimates of Covariance Parameters ^a		
Parameter	Estimate	Std. Error
Residual	55,393677	4,094832
Intercept [subject = PatID] Variance	22,038216	7,210932

Kein „signifikanter“ Behandlungseffekt

- Ergebnis bei Ignorierung der Abhängigkeitsstuktur:

Überbewertung des Informationsgehaltes der Daten → Fehlinterpretation

Parameter	Estimate	Std. Error	t	Sig.	
Intercept	18,536531	1,362637	29,777	,000	95
[Behandlungsgruppe=1, 00]	3,469899	1,926393	3,905	,000	17

Fälschlicherweise „signifikanter“ Behandlungseffekt



Vorlesungsinhalt

- Deskriptive Statistik
 - Darstellung 1es Merkmals
 - Zusammenhang 2er Merkmale
- Statistisches Testen
 - Prinzipien
 - Spezielle Tests
 - Multiples Testen
- Fallzahlplanung / Adaptive Designs
- Lineare Regression / **logistische Regression**
- Überlebenszeiten

Fragestellung

- Wie hängt **eine binäre** Zielgröße von mehreren Einflussgrößen ab?
- Beispiel (B.W. Brown 1980 aus [1]):
Population: Patienten mit Prostatakarzinom
Zielgröße = Befall regionaler Lymphknoten (ja/nein)
Einflussgrößen = Grading (niedrig/hoch),
saure Phosphatase im Serum (King-Armstrong Einheiten),
Röntgenbefund (negativ/positiv),
Tumorgröße (klein/groß)

Fragestellung

- Wie hängt **eine binäre** Zielgröße von mehreren Einflussgrößen ab?
- Beispiel (B.W. Brown 1980 aus [1]):
Population: Patienten mit Prostatakarzinom
Zielgröße = **Befall regionaler Lymphknoten (ja/nein)**
Einflussgrößen = Grading (niedrig/hoch),
saure Phosphatase im Serum (King-Armstrong Einheiten),
Röntgenbefund (negativ/positiv),
Tumorgröße (klein/groß)

Fragestellung

- Wie hängt eine binäre Zielgröße von mehreren Einflussgrößen ab?
- Beispiel (B.W. Brown 1980 aus [1]):
Population: Patienten mit Prostatakarzinom
Zielgröße = Befall regionaler Lymphknoten (ja/nein)
Einflussgrößen = Grading (niedrig/hoch),
saure Phosphatase im Serum (King-Armstrong Einheiten),
Röntgenbefund (negativ/positiv),
Tumorgröße (klein/groß)

Fragestellung

- Wie hängt eine binäre Zielgröße von mehreren Einflussgrößen ab?
- Beispiel (B.W. Brown 1980 aus [1]):
Population: Patienten mit Prostatakarzinom
Zielgröße = Befall regionaler Lymphknoten (ja/nein)
Einflussgrößen = Grading (**niedrig/hoch**),
saure Phosphatase im Serum (King-Armstrong Einheiten),
Röntgenbefund (**negativ/positiv**),
Tumorgröße (**klein/groß**)

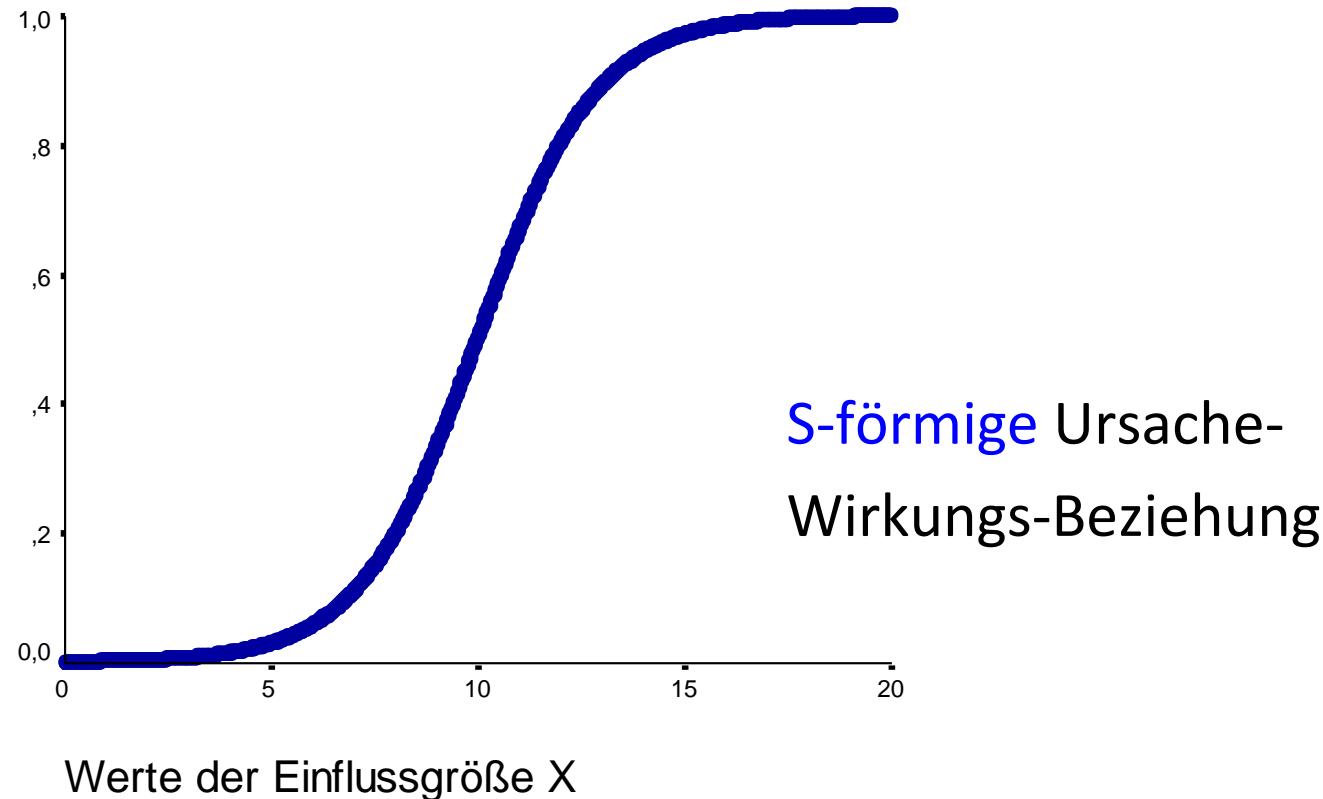
Fragestellung

- Wie hängt eine binäre Zielgröße von mehreren Einflussgrößen ab?
- Beispiel (B.W. Brown 1980 aus [1]):
Population: Patienten mit Prostatakarzinom
Zielgröße = Befall regionaler Lymphknoten (ja/nein)
Einflussgrößen = Grading (niedrig/hoch),
saure Phosphatase im Serum (King-Armstrong Einheiten),
Röntgenbefund (negativ/positiv),
Tumorgröße (klein/groß)

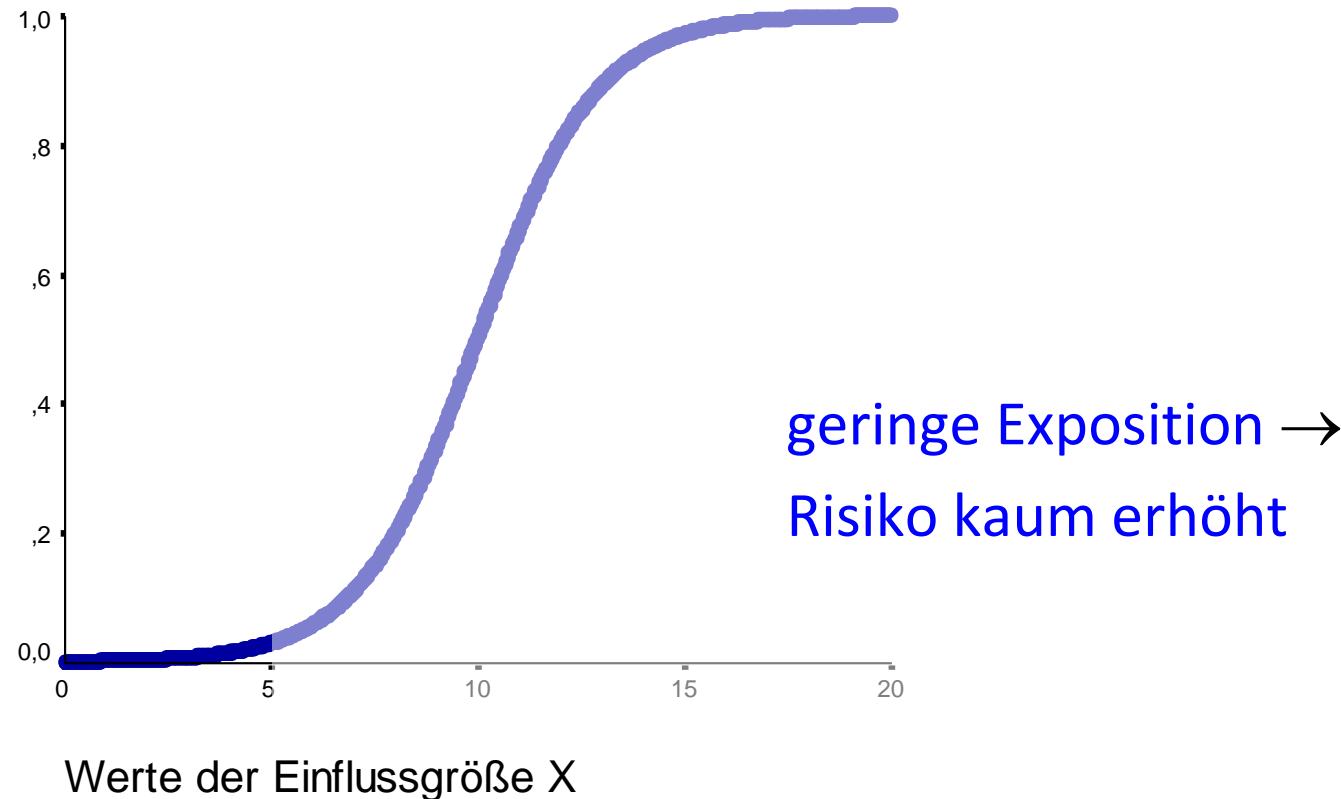
Wünsche an die Modellierung

- Das Modell soll die Ursache-Wirkungsbeziehung zwischen der Ziel- und den Einflussgrößen gut beschreiben
- Mehrere Möglichkeiten
- Populärste: logistische Transformation
- Wahrscheinlichkeit \rightarrow Logarithmus des Odds (Chance)
- $\text{logit}(p) = \log(\text{Odds}) = \log(p/(1-p)) = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k$
- $\text{logit}(p)$ ist nicht beschränkt
- $\text{logit}(p)$ ist streng monoton wachsend
- $p = 1 / (1 + e^{-\text{logit}(p)})$
- $p = 0,5 \Leftrightarrow \text{logit}(p) = 0$

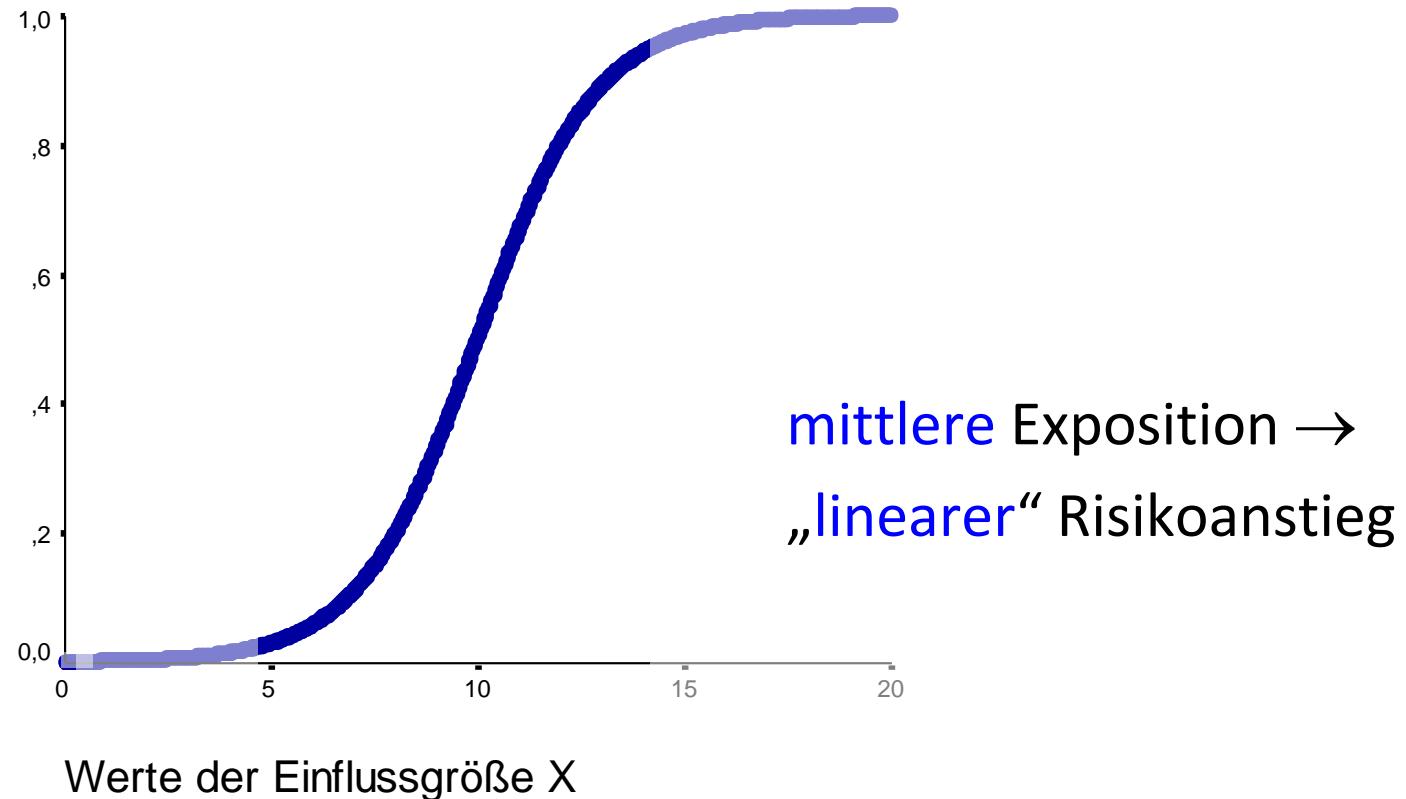
Logistische Ursache-Wirkungs-Beziehung



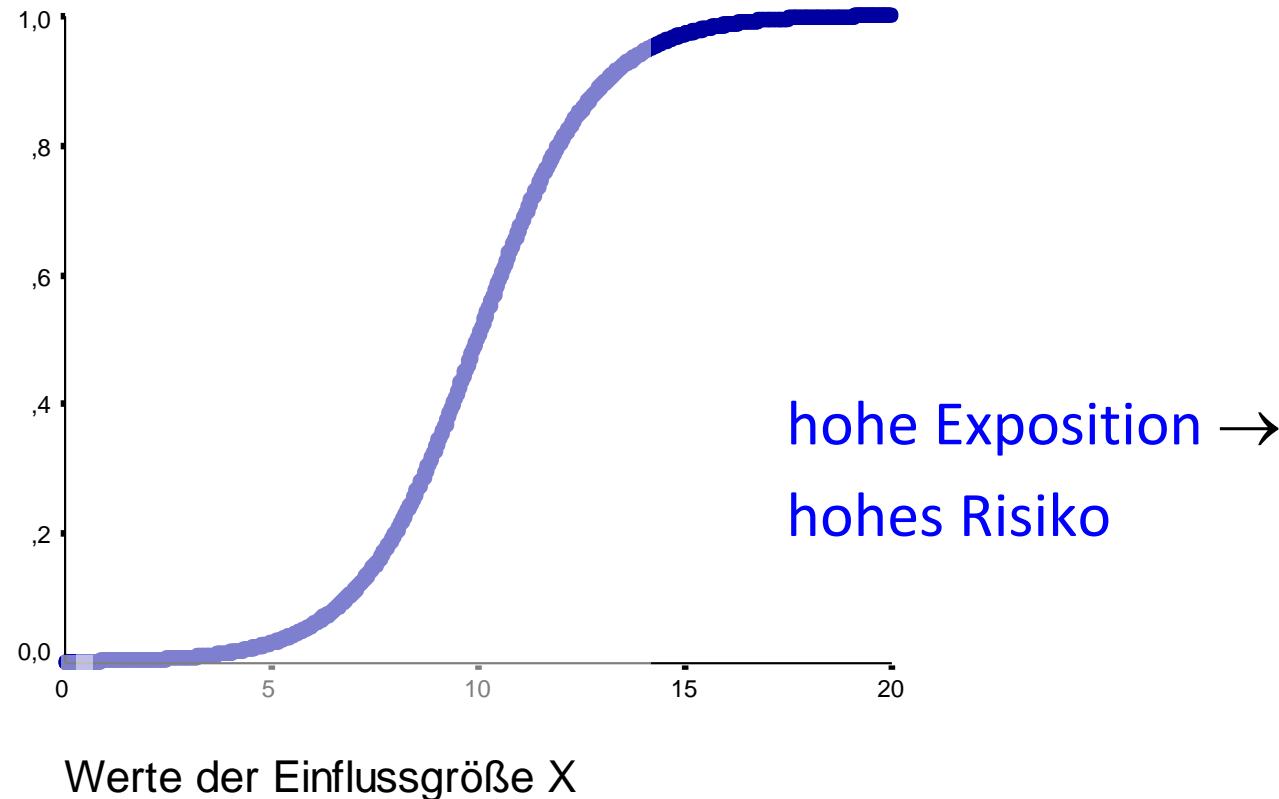
Logistische Ursache-Wirkungs-Beziehung



Logistische Ursache-Wirkungs-Beziehung



Logistische Ursache-Wirkungs-Beziehung



Interpretation der Regressionsparameter

- X eine Einflussgröße
- p die Wahrscheinlichkeit für ein Ereignis
- $\text{logit}(p) = \log(p/(1-p)) = \text{log(Odds)} = \beta_0 + \beta_1 X$
- $\text{Odds} = p/(1-p) = \exp(\beta_0 + \beta_1 X)$
- Das Odds verändert sich um Faktor $\exp(\beta_1)$, falls x um eine Einheit steigt – unabhängig von x :

$$\begin{aligned}\text{Odds}(x+1) &= \exp(\beta_0 + \beta_1(x+1)) = \exp(\beta_0 + \beta_1x + \beta_1) \\ &= \exp(\beta_0 + \beta_1x) \cdot \exp(\beta_1) = \text{Odds}(x) \cdot \exp(\beta_1) \\ \Rightarrow \text{OR}(x+1, x) &= \frac{\text{Odds}(x+1)}{\text{Odds}(x)} = \exp(\beta_1)\end{aligned}$$

Interpretation der Regressionsparameter

$$\text{OR}(x+1, x) = \frac{\text{Odds}(x+1)}{\text{Odds}(x)} = \frac{\exp(\beta_0 + \beta_1(x+1))}{\exp(\beta_0 + \beta_1 x)} = \exp(\beta_1).$$

- $\beta_1 = 0$: $\text{OR}(x+1, x) = \exp(\beta_1) = 1 \Rightarrow \text{Odds}(x+1) = \text{Odds}(x) \Leftrightarrow \text{kein Einfluss von } X$
- $\beta_1 > 0$: $\text{OR}(x+1, x) = \exp(\beta_1) > 1 \Rightarrow \text{Odds}(x+1) > \text{Odds}(x) \Leftrightarrow \text{mit } X \text{ steigende Chance}$
- $\beta_1 < 0$: $\text{OR}(x+1, x) = \exp(\beta_1) < 1 \Rightarrow \text{Odds}(x+1) < \text{Odds}(x) \Leftrightarrow \text{mit } X \text{ fallende Chance}$



Binäre Einflussgrößen

- Die Differenz der beiden Werte einer binären Einflussgröße sollte 1 sein
- Beispiel einer ungünstigen Codierung
 - Geschlecht: 0 für männlich, 2 für weiblich
 - $\exp(\beta_1)$ ist das Odds Ratio zwischen 0 und 1
 - aber 1 kommt nicht vor
- Beispiel für eine gute Codierung
 - Geschlecht: 1 für weiblich, 2 für männlich
 - Exposition: 0 für nichtexponiert, 1 für exponiert

Kategoriale Einflussgrößen > 2 Ausprägungen

- Beispiel: Blutgruppe A, B, AB, 0
- Eine Kategorie wird Referenzkategorie
- Referenzkategorie =
 - Inhaltlichen Grund: Referenzpatient
 - häufigste Ausprägung → kleinste Varianz der Schätzer
 - In Europa: Blutgruppe A = Referenz
- Andere Kategorien (Blutgruppe B, AB, 0):
umkodiert in eigene (Dummy-) Variable (SPSS: automatisiert)
Blutgruppe B: ja vs. nein, AB: ja vs. nein, 0: ja vs. nein
- Odds Ratio: andere Kategorien vs. Referenz (B/A, AB/A, 0/A)

Codierung von stetigen Einflussgrößen

- $\text{logit}(p) = \log(\text{Odds}) = \beta_0 + \beta_1 X_1$
- Problem: Modell setzt **Linearität** des **Einflusses** voraus
- Annahme ist oft **fraglich**
- 1. Lösung: **Transformation** in einem **theoretischen Modell**
quadratischer Einfluss, höhere Potenzen, Logarithmieren, ...
Vorwissen ist **selten**
- 2. Lösung: **Kategorisieren** → **Informationsverlust** →
weniger Trennschärfe → besser **vermeiden**
- 3. Lösung: **Modellierung** in der logistischen Regression
fast kein Vorwissen nötig - **datenabhängig**

Beispiel (Prostatakarzinom - LK)

- Prostatakarzinom - Lymphknoten (B.W. Brown 1980):
Population: Patienten mit Prostatakarzinom
- Zielgröße = Befall regionaler Lymphknoten (ja/nein)
- Einflussgröße =
GRADE: Grading (niedrig/hoch)
- SPSS liefert folgende Lösung für den Regressionsparameter:

Beispiel (Prostatakarzinom - LK)

Univariable logistische Regression

Variablen in der Gleichung

	RegressionskoeffizientB	Standardfehler	Wald	df	Sig.	Exp(B)	95,0% Konfidenzintervall für EXP(B)	
							Unterer Wert	Oberer Wert
Schritt 1	GRADE(1) Konstante	1,181 -,981	,596 ,391	3,934 6,297	1 1	,047 ,012	3,259 ,375	1,014 10,475

a. In Schritt 1 eingegebene Variablen: GRADE.

β_i s.e.(β_i)

OR(Grading hoch
versus niedrig)

Einflussgröße

95%-KI des OR



Beispiel (Prostatakarzinom - LK)

- Prostatakarzinom - Lymphknoten (B.W. Brown 1980):
Population: Patienten mit Prostatakarzinom
- Zielgröße = Befall regionaler Lymphknoten (ja/nein)
- Einflussgrößen =
 - ACID: saure Phosphatase im Serum (King-Armstrong Einheiten)
 - XRAY: Röntgenbefund (negativ/positiv)
 - SIZE: Tumorgröße (klein/groß)
 - GRADE: Grading (niedrig/hoch)
- SPSS liefert folgende Lösung für die Regressionsparameter:

Beispiel (Prostatakarzinom - LK)

Multivariable logistische Regression

Variablen in der Gleichung

		Regressions koeffizientB	Standardf ehler	Wald	df	Sig.	Exp(B)	95,0% Konfidenzintervall für EXP(B)	
								Unterer Wert	Oberer Wert
Schritt 1	ACID	2,283	1,299	3,087	1	,079	9,808	,768	125,211
	XRAY(1)	2,012	,794	6,417	1	,011	7,480	1,577	35,481
	SIZE(1)	1,550	,768	4,070	1	,044	4,712	1,045	21,243
	GRADE(1)	,772	,755	1,046	1	,306	2,165	,493	9,508
	Konstante	-3,901	1,261	9,572	1	,002	,020		

a. In Schritt 1 eingegebene Variablen: ACID, XRAY, SIZE, GRADE.

$$\beta_i \quad s.e.(\beta_i)$$

Einflussgrößen

$$OR(x_i+1, x_i)$$

95%-KI des $OR(x_i+1, x_i)$



Univariable und multivariable logistische Regression

Variablen in der Gleichung

	RegressionskoeffizientB	Standardfehler	Wald	df	Sig.	Exp(B)	95,0% Konfidenzintervall für EXP(B)	
							Unterer Wert	Oberer Wert
Schritt 1	GRADE(1)	1,181	,596	3,934	1	,047	3,259	1,014
	Konstante	-,981	,391	6,297	1	,012	,375	10,475

a. In Schritt 1 eingegebene Variablen: GRADE.

Grading allein = Einflussgröße multipel: Grading hat keinen Einfluss

Variablen in der Gleichung

	RegressionskoeffizientB	Standardfehler	Wald	df	Sig.	Exp(B)	95,0% Konfidenzintervall für EXP(B)	
							Unterer Wert	Oberer Wert
Schritt 1	ACID	2,283	1,299	3,087	1	,079	9,808	,768
	XRAY(1)	2,012	,794	6,417	1	,011	7,480	1,577
	SIZE(1)	1,550	,768	4,070	1	,044	4,712	1,045
	GRADE(1)	,772	,755	1,046	1	,306	2,165	,493
	Konstante	-3,901	1,261	9,572	1	,002	,020	9,508

a. In Schritt 1 eingegebene Variablen: ACID, XRAY, SIZE, GRADE.



Univariable und multivariable logistische Regression

- Univariables OR(Grading hoch versus niedrig) = 3,259, p=0,047
- Multivariables OR(Grading hoch versus niedrig) = 2,165, p=0,306

Univariables OR

- Vergleich aller Studienpatienten mit hohem Grading versus aller Patienten mit niedrigem Grading
ungeachtet weiterer Einflussgrößen

Multivariables OR

Vergleich fiktiver Patienten mit hohem Grading versus Patienten mit niedrigem Grading
mit gleichem ACID, XRAY, SIZE („Konstanthaltung“/“Adjustierung“)

Welche Einflussgrößen sind wichtig?

Modellwahl

- Logistisches Modell:
Ereigniswahrscheinlichkeit \leftrightarrow $\log(\text{Odds}) \leftarrow$ Einflussgrößen
- Naive Idee: Einschluss von vielen Einflussgrößen ins Modell
 - Vorteil: bessere Anpassung des Modells an die Daten
 - Nachteile:
 - i. die Schätzungen werden i. d. R. unpräziser
 - ii. Fehlender Wert einer Einflussgröße \rightarrow Fallausschluss
- Daher: nur Einflussgrößen einschließen, die zu einer relevanten Erhöhung der Wahrscheinlichkeit der Daten führen
- Frage: Was heißt relevant?

A priori Bewertung von Einflussgrößen

1. Klinisch etablierte Einflüsse, Confounder → Berücksichtigung im Modell: immer
2. Einflüsse mit plausibel vermutetem kausalem Zusammenhang → Berücksichtigung im Modell: möglichst
3. Identifizierung neuer fraglich wichtiger Einflüsse → Berücksichtigung: erst zum Schluss
 - A priori Bewertung: inhaltlich, nicht statistisch
 - Falls Einflüsse inhaltlich gleiches Gewicht besitzen → statistische Modellwahl

3 Statistische Tests

- Alle 3 führen asymptotisch zu gleichen Ergebnis (große n)
- Cave: Unterschiede bei kleineren Datensätzen
- Likelihood-Ratio Test → Modellanpassung
- Wald Test → Präzision der Schätzer
- Score Test → schnell
- Tukey: „lieber richtige als präzise Schätzer“
- Likelihood-Ratio Test = Königstest
- Keine Bewertung des kausalen Zusammenhangs

Automatisierte Modellwahl

- Motivation: großer Aufwand bei der Modellwahl
- Automatisierte Regeln: Quick and dirty
- 3 Grundverfahren:
 1. Variablen gleichzeitig einschließen
 2. Forward Stepwise Selection = Bottom Up:
Variablen Zug um Zug ins Modell aufnehmen
 3. Backward Stepwise Elimination = Top Down:
Variablen Zug um Zug aus der Modell ausschließen

Blockweise

- Kombination verschiedener Verfahren
- Automatisierte Blöcke möglich
- Nach jedem Block: kein Ausschluss bereits eingeschlossener Einflussgrößen möglich
- Vorschlag:
 1. Block: Einschluss von etablierten Einflüssen und Confoundern
 2. Block: Backward Verfahren zur Auswahl von Einflüssen mit plausibel vermutetem kausalen Zusammenhang
 3. Block: Forward Verfahren zur Identifizierung neuer fraglich wichtiger Einflüsse

Beispiel (Prostatakarzinom - LK)

Einschluss

Variablen in der Gleichung

Schritt	ACID	RegressionskoeffizientB	Standardfehler	Wald	df	Sig.	Exp(B)	95,0% Konfidenzintervall für EXP(B)	
								Unterer Wert	Oberer Wert
1	ACID	2,283	1,299	3,087	1	,079	9,808	,768	125,211
	XRAY(1)	2,012	,794	6,417	1	,011	7,480	1,577	35,481
	SIZE(1)	1,550	,768	4,070	1	,044	4,712	1,045	21,243
	GRADE(1)	,772	,755	1,046	1	,306	2,165	,493	9,508
	Konstante	-3,901	1,261	9,572	1	,002	,020		

a. In Schritt 1 eingegebene Variablen: ACID, XRAY, SIZE, GRADE.

Beispiel (Prostatakarzinom - LK)

Backward Stepwise Selection

Variablen in der Gleichung

		Regressions koeffizientB	Standardf ehler	Wald	df	Sig.	Exp(B)	95,0% Konfidenzintervall für EXP(B)	
								Unterer Wert	Oberer Wert
Schritt 1	acid	2,283	1,299	3,087	1	,079	9,808	,768	125,214
	xray(1)	2,012	,794	6,417	1	,011	7,480	1,577	35,481
	size(1)	1,550	,768	4,070	1	,044	4,712	1,045	21,243
	grade(1)	,772	,755	1,046	1	,306	2,165	,493	9,508
	Konstante	-3,901	1,261	9,572	1	,002	,020		
Schritt 2	acid	2,063	1,265	2,660	1	,103	7,869	,660	93,891
	xray(1)	2,062	,778	7,029	1	,008	7,860	1,712	36,091
	size(1)	1,756	,739	5,643	1	,018	5,787	1,359	24,633
	Konstante	-3,576	1,181	9,164	1	,002	,028		
Schritt 3	xray(1)	2,119	,747	8,054	1	,005	8,326	1,926	35,989
	size(1)	1,588	,700	5,148	1	,023	4,895	1,241	19,304
	Konstante	-2,045	,610	11,236	1	,001	,129		

a. In Schritt 1 eingegebene Variablen: acid, xray, size, grade.

Beispiel (Prostatakarzinom - LK)

Forward Stepwise Selection

Variablen in der Gleichung

		Regressions koeffizientB	Standardf ehler	Wald	df	Sig.	Exp(B)	95,0% Konfidenzintervall für EXP(B)	
Schritt								Unterer Wert	Oberer Wert
Schritt 1	xray(1)	2,182	,697	9,783	1	,002	8,861	2,258	34,770
	Konstante	-1,170	,382	9,403	1	,002	,310		
Schritt 2	xray(1)	2,119	,747	8,054	1	,005	8,326	1,926	35,989
	size(1)	1,588	,700	5,148	1	,023	4,895	1,241	19,304
	Konstante	-2,045	,610	11,236	1	,001	,129		

- a. In Schritt 1 eingegebene Variablen: xray.
- b. In Schritt 2 eingegebene Variablen: size.

Forward Stepwise Selection = Backward Stepwise Selection



Modellbildung - Bewertung

- Viele Auswahlverfahren für Modelle existieren
- Auswahlverfahren werden oft kontrovers diskutiert
- Automatisierte Verfahren:
 - in klinischen Studien oft benutzt
 - I. d. R. akzeptiert, falls: Forward = Backward
- Immer: aus inhaltlicher Sicht Konzept erforderlich
- Interaktionen bedenken
- Wichtig: in Publikationen Verfahren angeben und begründen!
- Mehrere Modelle: Alternativen diskutieren

Vorlesungsinhalt

- Deskriptive Statistik
 - Darstellung 1es Merkmals
 - Zusammenhang 2er Merkmale
- Statistisches Testen
 - Prinzipien
 - Spezielle Tests
 - Multiples Testen
- Fallzahlplanung / Adaptive Designs
- Lineare Regression / logistische Regression
- Überlebenszeiten



Beispiele für „Überlebenszeit“

- **Gesamtüberlebenszeit**

Zeit bis Ereignis = Tod jedweder Art

- **Progressionsfreie Überlebenszeit**

Zeit bis Ereignis = Progression + Tod jedweder Art

- **Ereignisfreie Überlebenszeit**

Zeit bis Ereignis = Therapiezwischenfall + Progression + Tod

- **Krankheitsdauer**

Zeit bis Ereignis = Heilung bzw. Remission

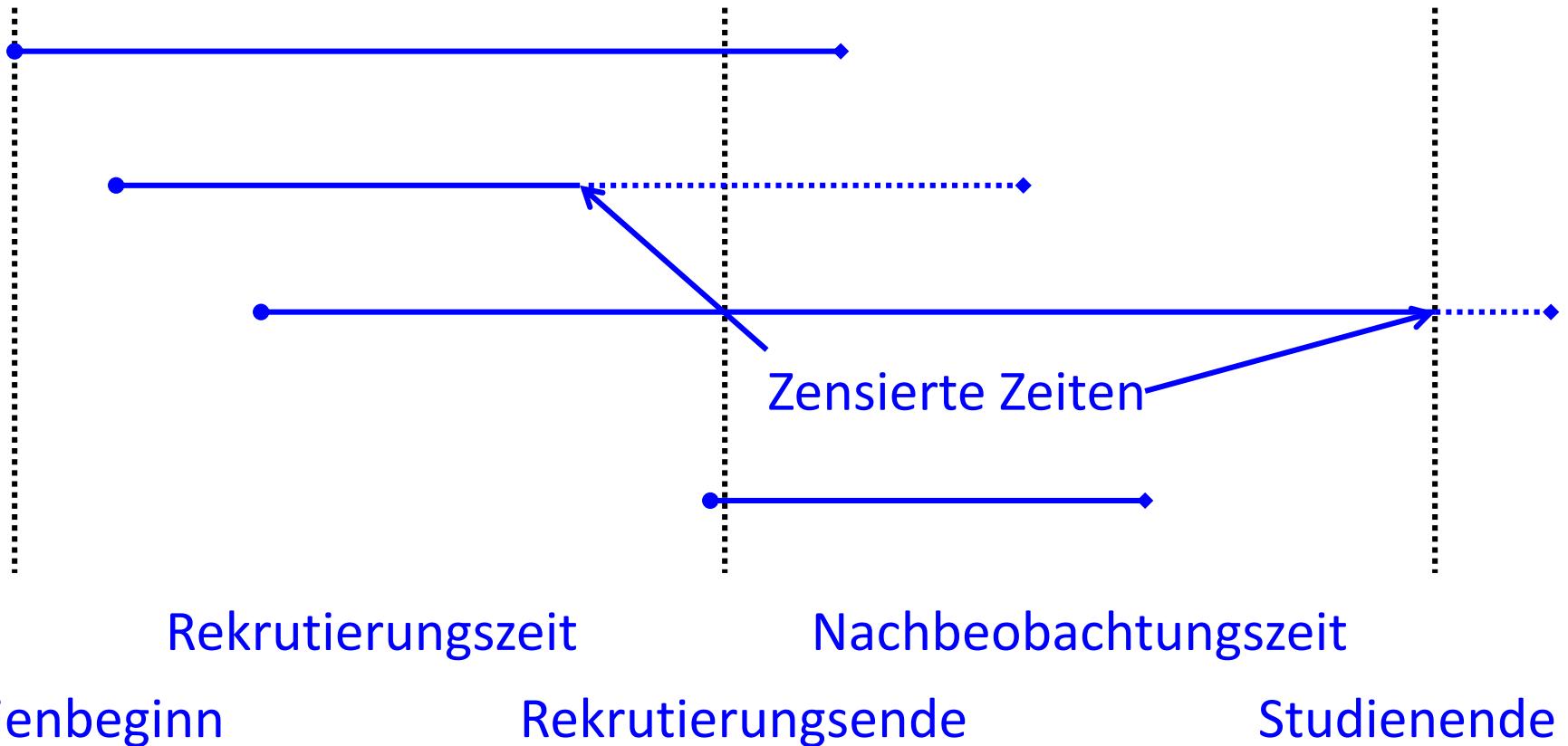
- **Gehstrecke**

Distanz bis Ereignis = Gehstopp durch **Claudicatio intermittens**

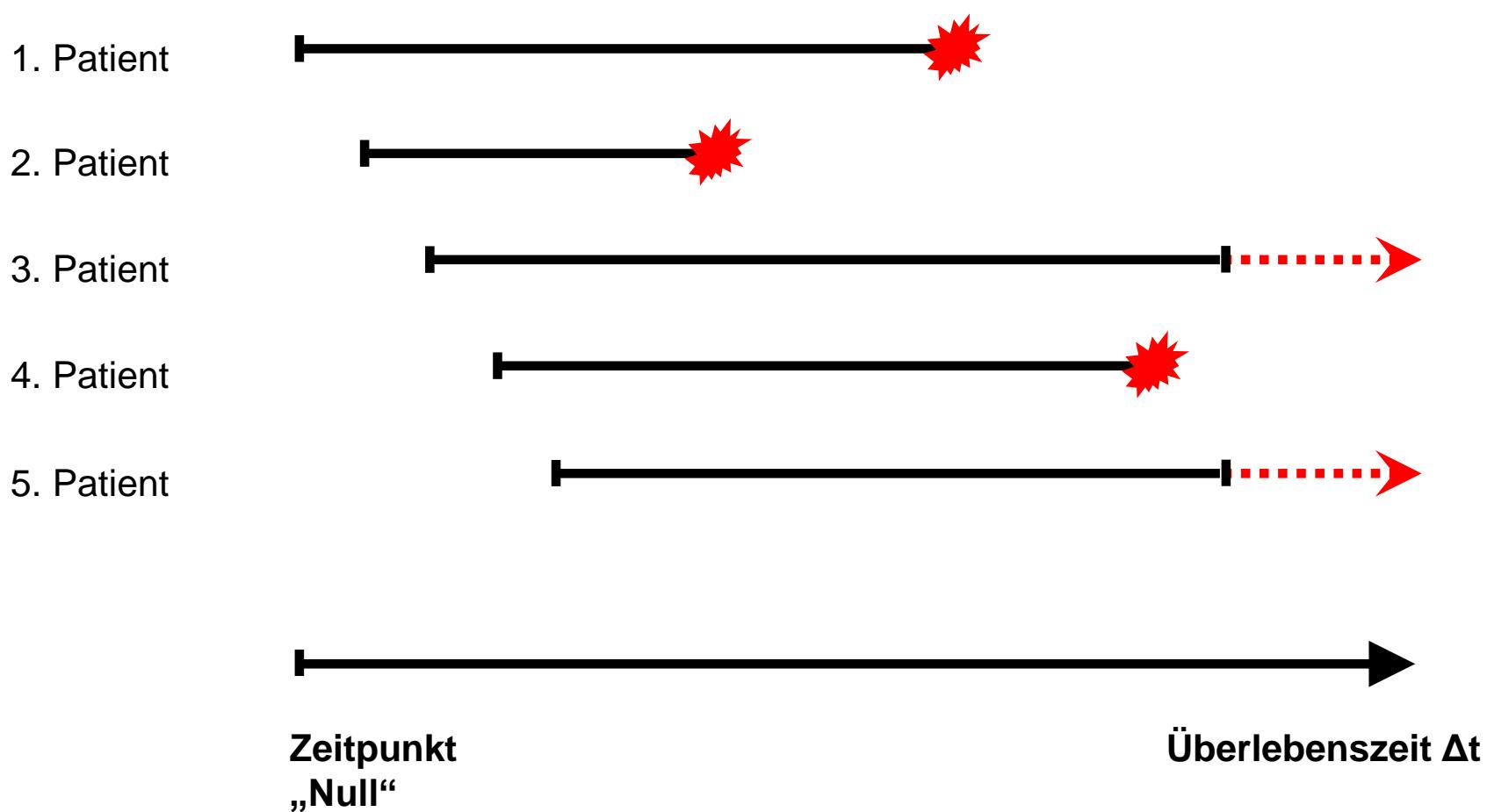
Definition von Überlebenszeit

- Definiere **Zustand** (z. B. stabiles Krankheitsstadium)
- **Anfang**: Beginn des Zustands ist **bekannt**
(z. B. **Diagnose**, Aufnahme in die Studie, **Randomisierung**,
Therapiestart, Startpunkt der Laufstrecke, ...)
- Definiere **Ereignis** das Zustand **beendet** (z. B. Progression + Tod)
- 2 Arten von **Beobachtungen**:
 - a. Ereignis wird **beobachtet** \Rightarrow Dauer des Zustands **bekannt**
 - b. Ereignis wird **nicht beobachtet** \Rightarrow **Mindestdauer** des Zustands = Beginn **bis** letzte Beobachtung **ist bekannt**
= (rechts-)**zensierte Überlebenszeiten**

Überlebenszeitstudie



Zensierte Ereigniszeiten



Kaplan-Meier Schätzung

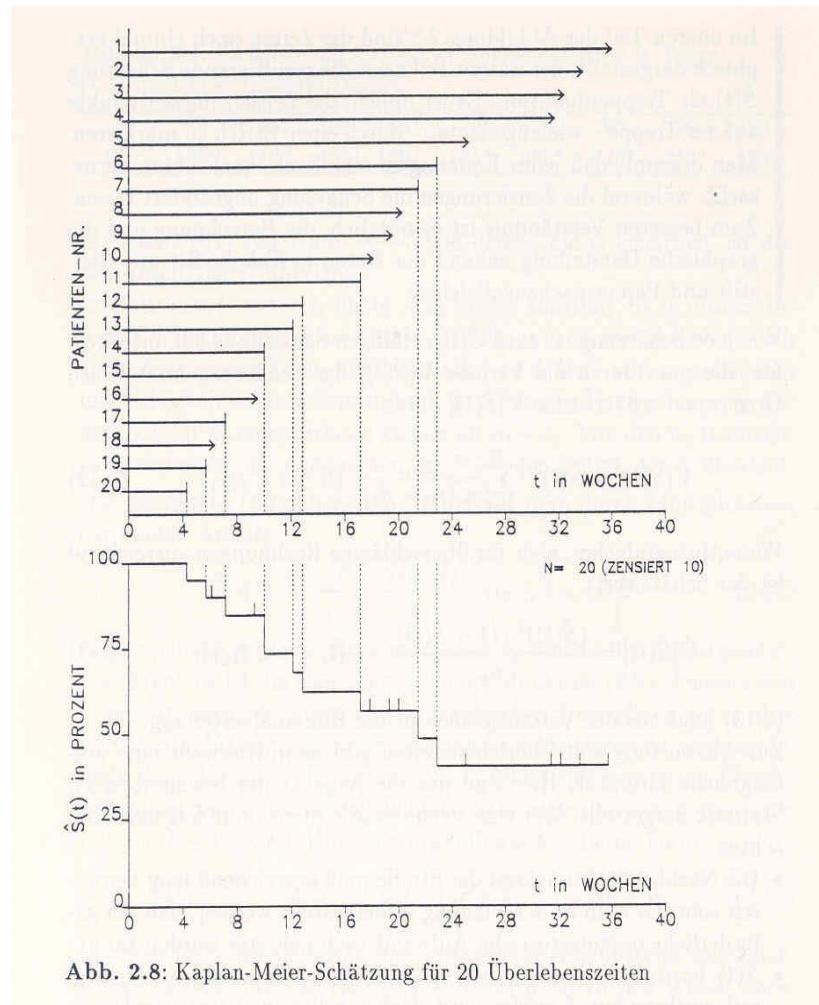
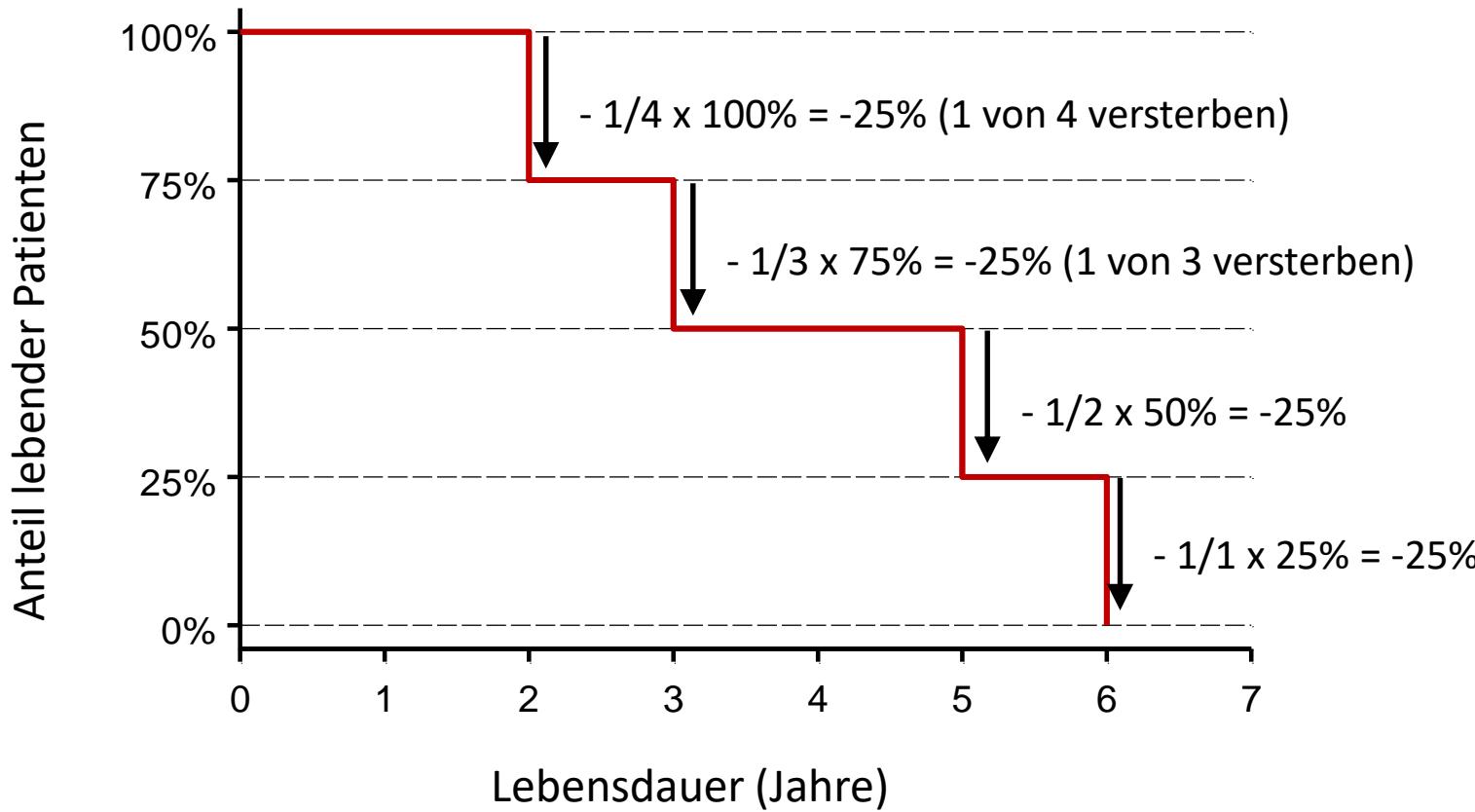


Abb. 2.8: Kaplan-Meier-Schätzung für 20 Überlebenszeiten

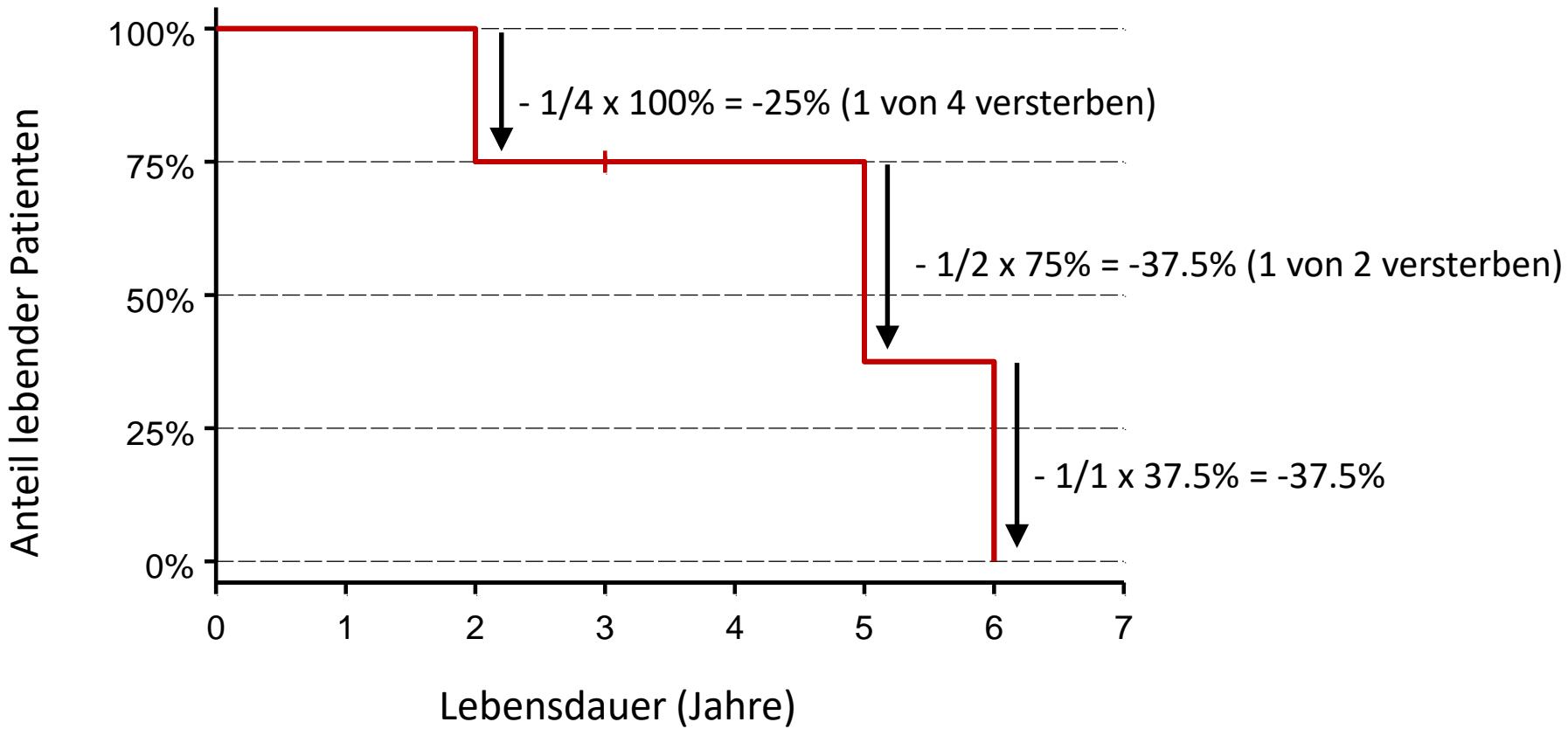
Kaplan-Meier Schätzung

Survtimes: 2, 3, 5, 6



Kaplan-Meier Schätzung

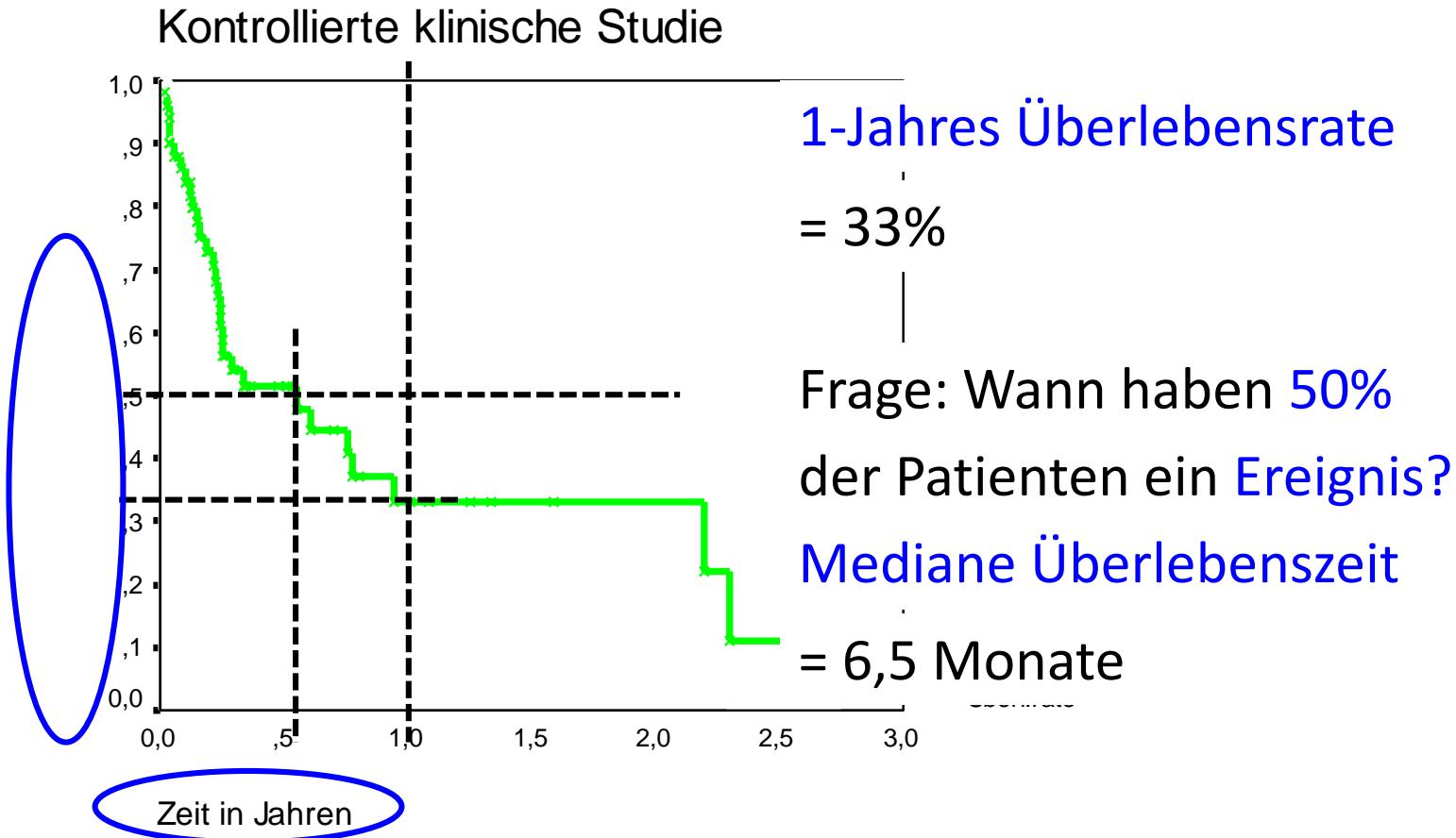
Survtimes: 2, 3+, 5, 6



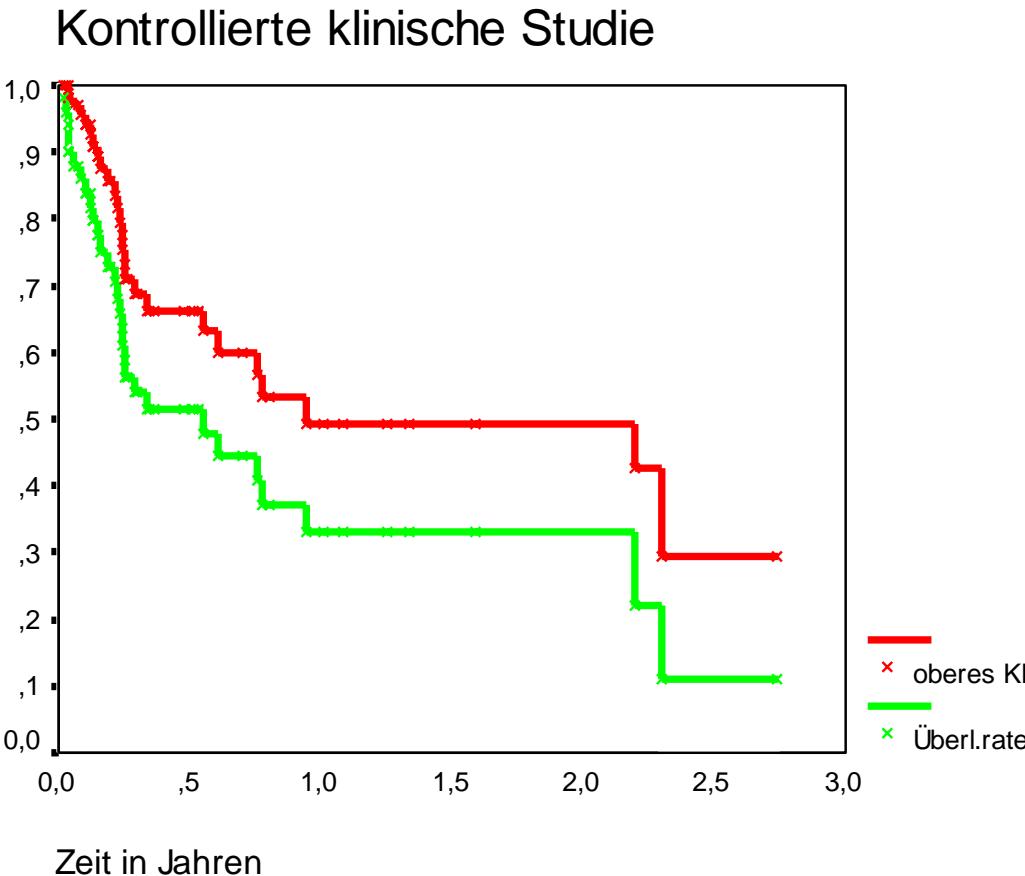
Wie stellt man Überlebenszeiten dar?

- Überlebensrate:
 - Anteil von Patienten ohne Ereignis zu einem festen Zeitpunkt
 - Beispiel: 1-Jahres Überlebensrate von 35%
Nach 1 Jahr: 65% Patienten mit Ereignis erwartet
- Überlebenszeitkurve
 - Überlebenszeitraten zu jedem Zeitpunkt
 - Vorteil: volle Information
- Wichtig: Berücksichtigung von Zensierungen
Methode von Kaplan und Meier

Überlebenszeitkurve nach Kaplan-Meier

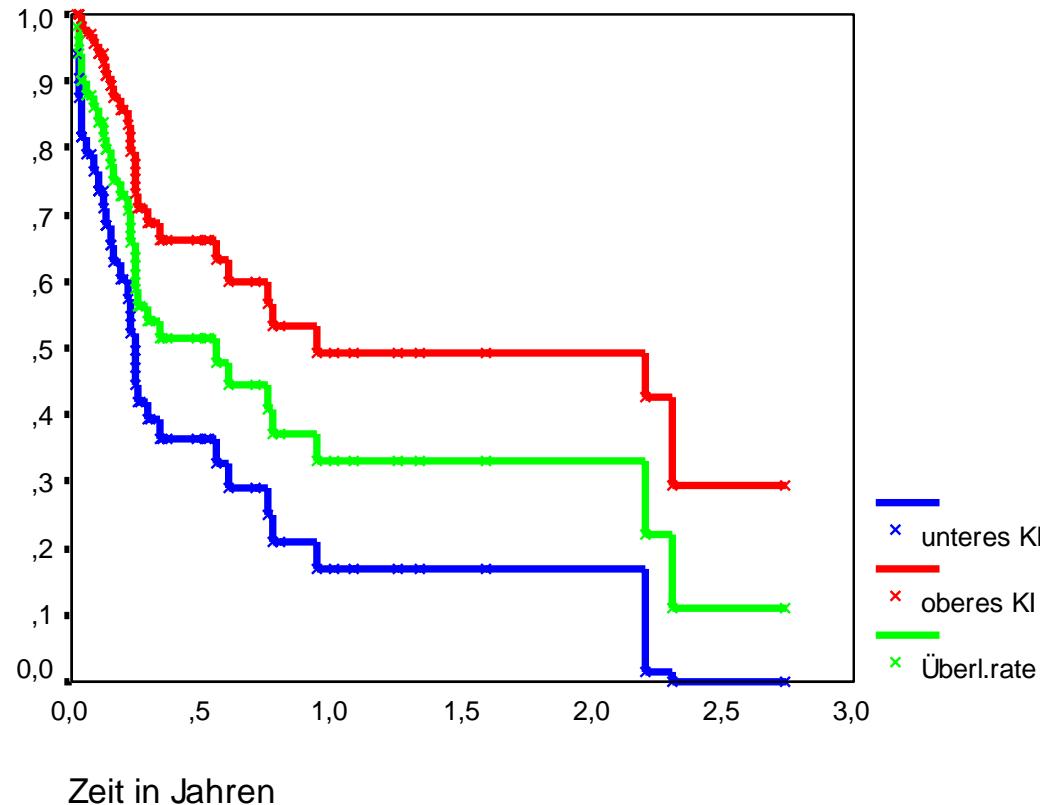


Überlebenszeitkurve mit Konfidenzintervall



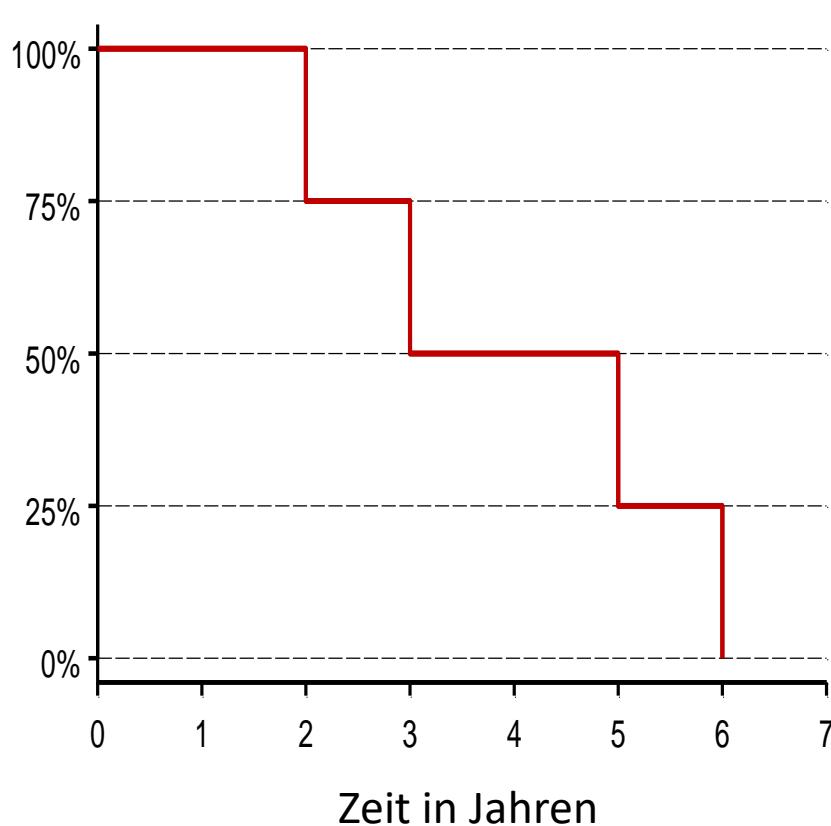
Überlebenszeitkurve mit Konfidenzintervall

Kontrollierte klinische Studie



Überlebenszeitkurve nach Kaplan-Meier

Survtimes: 2, 3, 5, 6



„Nicht eindeutige“ mediane Überlebenszeit:

Mediane Überlebenszeit wird in diesem Fall mit 3 Jahre definiert, nicht mit 4 oder 5 Jahren.

Vergleich von Überlebenszeiten

- Szenario:
 - Vergleich einer neuen Therapie mit einer Standardtherapie
 - Vergleich von Merkmalen (z. B. Geschlecht, Schweregrad)
- Studienfrage:

Wie beeinflusst

 - die Therapie
 - das Merkmal

die Überlebenszeit

Hazard-Rate und Hazard-Ratio

Hazard: Das „Risiko“ eines Ereignisses zum Zeitpunkt t
oder die „Ereignisintensität“ zum Zeitpunkt t.

$$\text{hazard}(t) := h(t) := \lim_{\Delta \searrow 0} \frac{P(\text{Ereignis in } [t, t + \Delta] \mid \text{kein Ereignis in } [0, t])}{\Delta}$$

Hazard Ratio zum Zeitpunkt t:

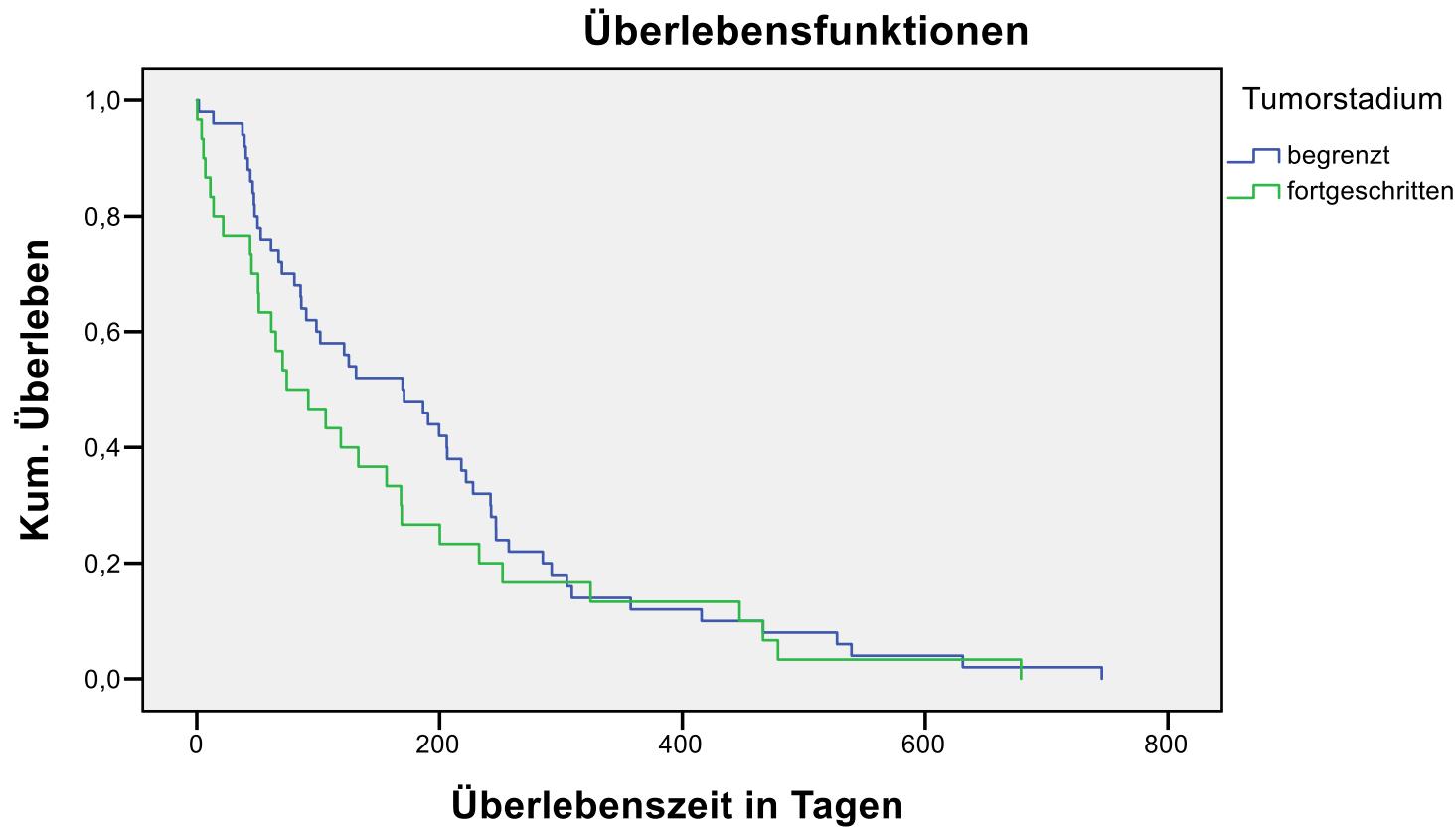
$$HR(t) := \frac{\text{hazard}(t) \text{ unter der neuen Therapie}}{\text{hazard}(t) \text{ unter der Standardtherapie}}$$

$$HR(t) \left\{ \begin{array}{ll} < 1 & \text{besser} \\ = 1 & \Leftrightarrow \text{die neue Therapie ist gleich gut zum Zeitpunkt t} \\ > 1 & \text{schlechter} \end{array} \right.$$

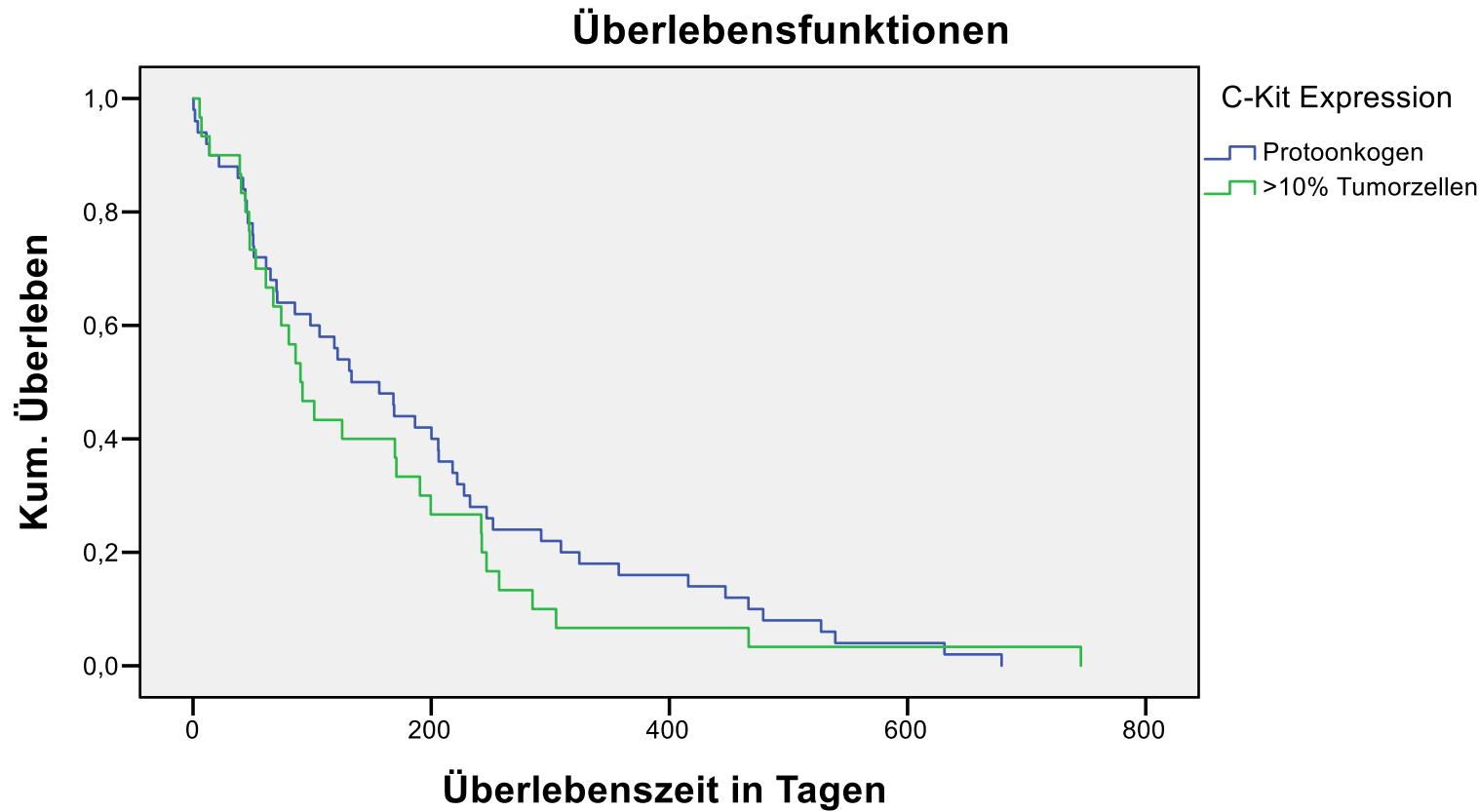
Explorative multivariable Überlebenszeitanalyse

- Ziele:
 - Erstellen einer individuellen Prognose
 - Welche Einflussgrößen sind relevant und unabhängig?
- Beispiel einer Beobachtungsstudie
 - Grundgesamtheit: Patienten ohne Chemotherapieresponse bei kleinzelligem Bronchialkarzinom
 - Ziel: Identifikation von Hochrisikogruppen
 - Mögliche Einflussgrößen:
 - a. Tumorstadium (Stage: begrenzt vs. fortgeschritten)
 - b. C-Kit Expression (Protoonkogen, >10% Tumorzellen)

Einfluss des Tumorstadiums auf die Überlebenszeit



Einfluss der C-Kit Expression auf die Überlebenszeit



Fragen

- Wie hoch ist der Einfluss von Tumorstadium und C-Kit Expression?
- Welchen gemeinsamen Einfluss haben Tumorstadium und C-Kit Expression?
- Gibt es Wechselwirkungen?
- Wird der Einfluss der C-Kit Expression durch die imbalancierte Verteilung des Tumorstadiums über- oder unterbewertet?

Coxregression

- D.R. Cox, Regression models and life tables, 1972 J. R. Statist. Soc. B
- Bewertung mehrerer Einflussgrößen gleichzeitig
- Berücksichtigung von Wechselwirkungen
- Auswahl erklärender Einflüsse/Wechselwirkungen
- Aufdecken verborgener Einflüsse
- Angabe des Hazard Ratios = Risikoerhöhung
- Modell der Überlebenswahrscheinlichkeit
- Konfidenzintervall und p-Wert
- nur 1 Einflussgröße (Therapie) \cong Log-Rank Test

Einfluss von Tumorstadium und C-Kit Expression auf das Überleben

	B	SE	Wald	df	Signifikanz	Exp(B)	95,0% Konfidenzinterv. für Exp (B)	
							Untere	Obere
Tumorstadium	,263	,233	1,278	1	,258	1,301	,824	2,055

	B	SE	Wald	df	Signifikanz	Exp(B)	95,0% Konfidenzinterv. für Exp (B)	
							Untere	Obere
C-Kit Expression	,192	,236	,660	1	,416	1,211	,763	1,922

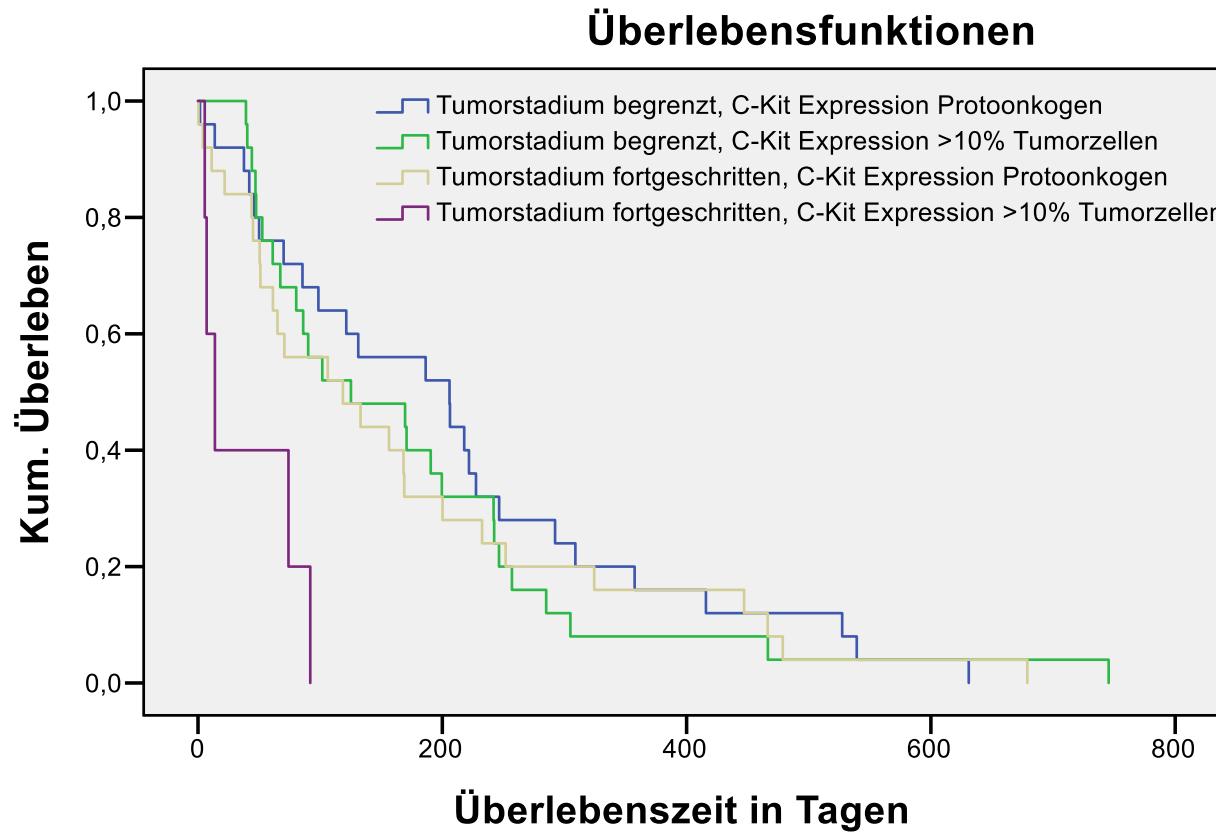
	B	SE	Wald	df	Signifikanz	Exp(B)	95,0% Konfidenzinterv. für Exp (B)	
							Untere	Obere
Tumorstadium	,438	,267	2,684	1	,101	1,549	,918	2,616
C-Kit Expression	,390	,270	2,081	1	,149	1,477	,869	2,509

	B	SE	Wald	df	Signifikanz	Exp(B)	95,0% Konfidenzinterv. für Exp (B)	
							Untere	Obere
Tumorstadium	,205	,285	,518	1	,472	1,227	,702	2,144
C-Kit Expression	,156	,288	,291	1	,589	1,168	,664	2,057
C-Kit Expression*Tumorstadium	1,272	,584	4,747	1	,029	3,569	1,136	11,208

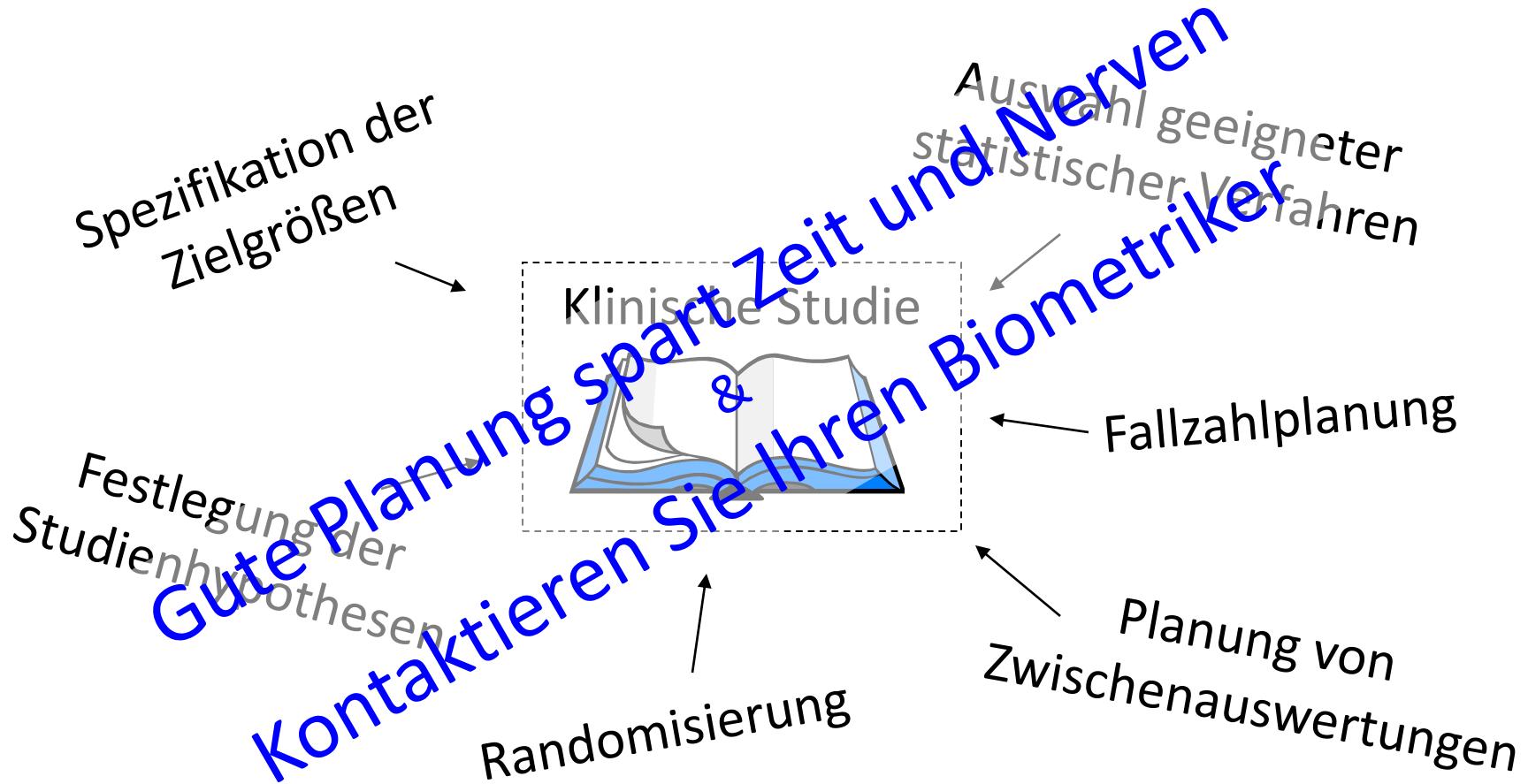
Einfluss von Tumorstadium und C-Kit Expression auf das Überleben

- Das Tumorstadium alleine hat keinen signifikanten Einfluss auf das Überleben („Haupteffekt des Tumorstadiums“).
- Die C-Kit Expression alleine hat keinen signifikanten Einfluss auf das Überleben („Haupteffekt der C-Kit Expression“).
- 3,6faches Hazard (Risiko) bei fortgeschrittenem Tumorstadium und gleichzeitiger positiver C-Kit Expression („Interaktion“=„Wechselwirkung“).

Einfluss von Tumorstadium und C-Kit Expression auf das Überleben



Take Home Message



Anhang – Liste von Synonymen im Statistikvokabular

Merkmals / Variable	Wahrscheinlichkeit für Fehler 1.	verbundene Stichproben /
Merkmaltyp / Skalenniveau	Art / Rate für Fehler 1. Art / falsch positiv Rate (seltener)	abhängige Stichproben
binär / dichotom		unabhängige Variable /
stetige Variable / kontinuierliche Variable	Streudiagramm / Scatterplot / Punktewolke	Einflussgröße
Lokationsmaß / Lagemaß	Kreuztabelle / Kontingenztafel /	abhängige Variable / Zielgröße /
Spannweite / Range	4-Felder Tafel (bei 2x2)	Endpunkt
Quantil / Perzentil	Normalverteilung / Gaußsche	Hauptzielgröße / primärer
Erwartungswert / erwarteter Wert	Verteilung	Endpunkt
Varianz / Streuung	Cox-Modell / proportional hazard	Zwischenanalyse /
relative Häufigkeit / prozentuale Häufigkeit	Modell	Interimsanalyse
rechtsschief / linksgipflig	zensierte Daten / trunkierte Daten	
linksschief / rechtsgipflig	y-Achsen Abschnitt / Intercept	
Zusammenhang / Assoziation		



Institut für Biometrie und
Klinische Forschung

Raphael Koch

Dr. rer. medic.

Diplom-Mathematiker

Robert Kwiecien

Dr. rer. nat.

Diplom-Mathematiker

Jeanette Köppe

Dr. rer. nat.

M.Sc. Physikerin

ibkf@ukmuenster.de

<http://ibkf.uni-muenster.de>

Raphael Koch / Robert Kwiecien / Jeanette Köppe