Institut für Biometrie und
Klinische Forschung

Biometrie

Vorlesung

Querschnittsfach
Epidemiologie, Medizinische Biometrie und Medizinische Informatik

Münster
Wintersemester 2026/27

medizinische
fakultät
Universität Münster

IBKF
Institut für Biometrie und Klinische Forschung

Andreas Faldum

Univ.-Prof. Dr. rer. nat. et med. habil.
Diplom-Mathematiker und Arzt

Geschäftsführender Direktor des Instituts für Biometrie und Klinische Forschung

faldum.andreas@ukmuenster.de
http://ibkf.uni-muenster.de

medizinische fakultät
Universität Münster

Institut für Biometrie und
Klinische Forschung

Moritz Fabian Danzer

Dr. rer. medic.
Master of Science in Mathematics

Sprecher der internationalen Arbeitsgruppe „Adaptive Designs und Multiples Testen“

moritzfabian.danzer@ukmuenster.de
http://ibkf.uni-muenster.de

medizinische
fakultät
Universität Münster

ibkf.uni-muenster.de



Geschäftsf. Direktor:
Univ.-Prof. Dr. rer. nat. et med. habil. Andreas Faldum
Schmeddingstraße 56
48149 Münster

Telefon: +49 (0)2 51 / 83 - 5 06 60
Fax: +49 (0)2 51 / 83 - 5 52 77
E-Mail: faldum.andreas@ukmuenster.de

Instituts-E-Mail: ibkf@uni-muenster.de

Publikationsliste des Instituts

Video "20 Fragen an einen Biometriker"



Lehre

- Vorlesung der Medizinischen Biometrie (Pflichtvorlesung zum Querschnittsfach Epidemiologie, Biometrie und Med. Informatik) (V - 1 SWS) Blockveranstaltung
- Praktikum der Medizinischen Biometrie
- Seminar über Neuere Methoden der Biometrie
- Statistik mit SPSS
- Veranstaltungen für Studierende der Experimentellen Medizin
- Methodik Klinischer Studien
- Systems Medicine - Systembiologische Methoden in der Medizin

Beratung

Statistische Beratung für Promotions- und Habilitationsarbeiten sowie Vorträge und Veröffentlichung

Eine Beratung erfolgt nur nach Vorliegen eines vollständig ausgefüllten Anmeldeformulars. Der/die Ratsuchende muss (persönlich oder telefonisch) Herrn Oleg Choubine (Tel.: (0251) 83-55272 oder ✉ beratung-ibkf@ukmuenster.de) zur Terminabsprache kontaktieren.

Eine telefonische Erst- oder "Kurz"-Beratung erfolgt nicht.

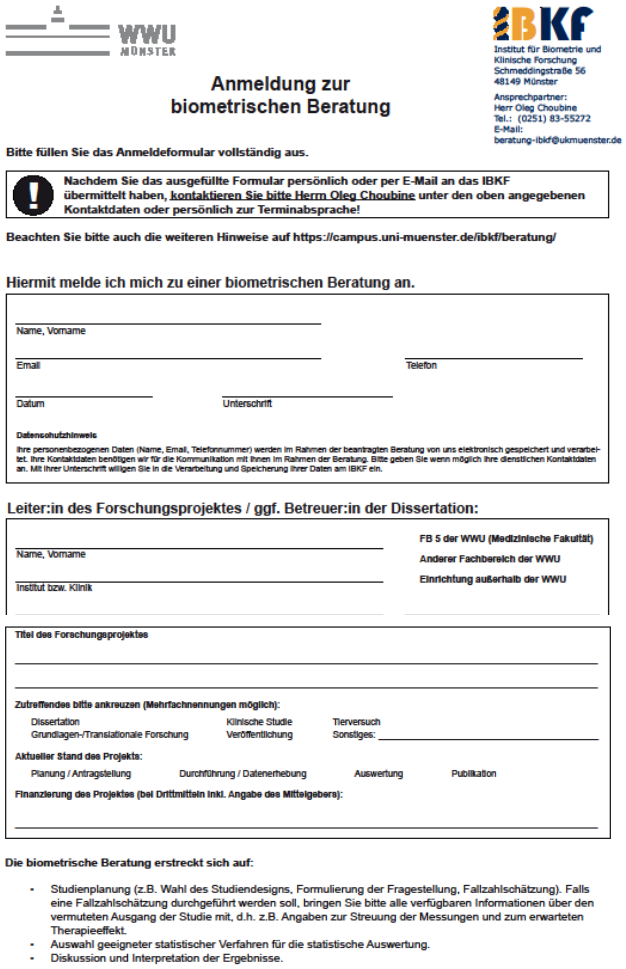
Antragsformular zur statistischen Beratung

Mit dem Firefox-Explorer gibt es im Moment Probleme bei Ausfüllen des Beratungsformulars! Bitte benutzen Sie, wenn möglich Google Chrome, Microsoft Edge oder den Internetexplorer.

Das hier verlinkte Dokument beschreibt, welche inhaltlichen und technischen Bedingungen klinische Datensätze erfüllen sollen, um möglichst problemlose statistische Analysen zu ermöglichen.

Der Umgang von Wissenschaftler/innen mit Patientendaten unterliegt strikten Regelungen, welche natürlich auch an der Medizinischen Fakultät berücksichtigt werden. Dabei bedarf es aus datenschutzrechtlichen Gründen für Wissenschaftler/innen im Zugriff auf patientenbezogene Daten im Rahmen eines Forschungsprojektes bzw. Promotion/Habilitation/Veröffentlichung besonderer Regelungen, wenn die/der Wissenschaftler/in zu den Patienten in keinem direkten Behandlungszusammenhang steht.

Deshalb bitten wir Sie das Dokument "Datenschutzerklärung zur biometrischen Beratung", das dem Anmeldeformular angefügt ist, auszufüllen und zur Beratung mitzubringen. Hier wird es dann von unseren Wissenschaftlern abgezeichnet und entsprechend der Aufbewahrungsfrist archiviert.



WWU MÜNSTER

IBKF
Institut für Biometrie und Klinische Forschung
Schmeddingstraße 56
48149 Münster
Ansprechpartner:
Herr Oleg Choubine
Tel.: (0251) 83-55272
E-Mail:
beratung-ibkf@ukmuenster.de

Anmeldung zur biometrischen Beratung

Bitte füllen Sie das Anmeldeformular vollständig aus.

Nachdem Sie das ausgefüllte Formular persönlich oder per E-Mail an das IBKF übermittelt haben, kontaktieren Sie bitte Herrn Oleg Choubine unter den oben angegebenen Kontaktdaten oder persönlich zur Terminabsprache!

Beachten Sie bitte auch die weiteren Hinweise auf https://campus.uni-muenster.de/ibkf/beratung/

Hiermit melde ich mich zu einer biometrischen Beratung an.

Name, Vorname

Email

Telefon

Datum

Unterschrift

Datenschutzhinweis

Ihre personenbezogenen Daten (Name, Email, Telefonnummer) werden im Rahmen der beantragten Beratung von uns elektronisch gespeichert und verarbeitet. Ihre Kontaktdaten benötigen wir für die Kommunikation mit Ihnen im Rahmen der Beratung. Bitte geben Sie wenn möglich Ihre dienstlichen Kontaktdaten an. Mit Ihrer Unterschrift willigen Sie in die Verarbeitung und Speicherung Ihrer Daten am IBKF ein.

Leiter:in des Forschungsprojektes / ggf. Betreuer:in der Dissertation:

Name, Vorname

Institut bzw. Klinik

FB 5 der WWU (Medizinische Fakultät)
Anderer Fachbereich der WWU
Einrichtung außerhalb der WWU

Titel des Forschungsprojektes

Zutreffendes bitte ankreuzen (Mehrfachnennungen möglich):

Dissertation | Klinische Studie | Tierversuch
Grundlagen-/Translationale Forschung | Veröffentlichung | Sonstiges:

Aktueller Stand des Projekts:

Planung / Antragstellung | Durchführung / Datenerhebung | Auswertung | Publikation

Finanzierung des Projektes (bei Drittmitteln inkl. Angabe des Mittelgebers):

Die biometrische Beratung erstreckt sich auf:

- Studienplanung (z.B. Wahl des Studiendesigns, Formulierung der Fragestellung, Fallzahlschätzung). Falls eine Fallzahlschätzung durchgeführt werden soll, bringen Sie bitte alle verfügbaren Informationen über den vermuteten Ausgang der Studie mit, d.h. z.B. Angaben zur Streuung der Messungen und zum erwarteten Therapieeffekt.
- Auswahl geeigneter statistischer Verfahren für die statistische Auswertung.
- Diskussion und Interpretation der Ergebnisse.

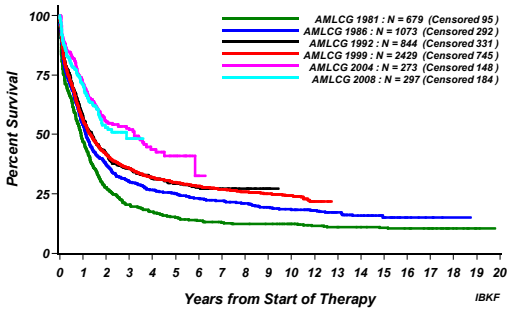
Studienkooperationen

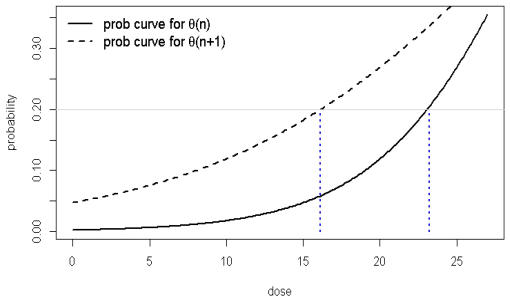
- 5-ALA
- ABParo
- ADAM
- Aided
- AXADIA
- BARGRU
- BARGRU-II
- BIBF 1120 in AML
- B-NHL 2013
- ClaiMS
- ELIAS
- EPIS-AKI
- ES-RiP
- Fabry Disease
- GD2+ IL18 CART
- GenderVasc
- Glacé-Trial
- HandiCAP
- InMediValue
- ISRT 20 Gy
- NOA 11-Studie
- OptAHF
- oVID
- PREDICT-IVI
- PREVENT-JIA
- PSA-Inform
- RICH-TRIAL
- RIT in GBM
- SOLKID-GNR
- SoLKID-Studie
- TOSYMA
- TRABTRAP
- WIN-SAB

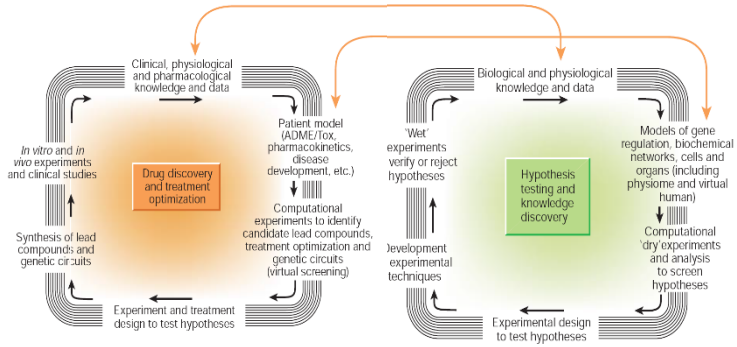
Biometrisches Zentrum für Kinderkrebsstudien

- AT/RT
- Kraniopharyngeom
- EU-RHAB-Register
- EWING
- HGG
- LOGGIC
- Neuroblastome
- Neuropsychologie
- PNET
- RB-Registry
- T-HAPLO-SCT for SCD
- YCMB-LR

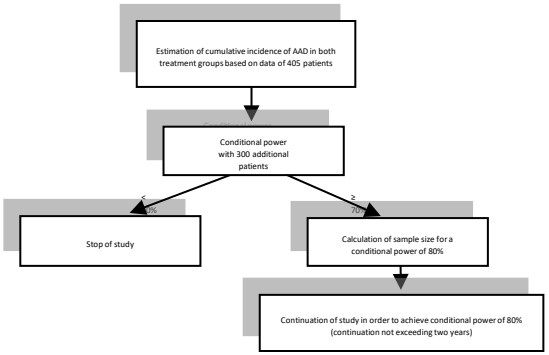
Methodische Forschungsprojekte

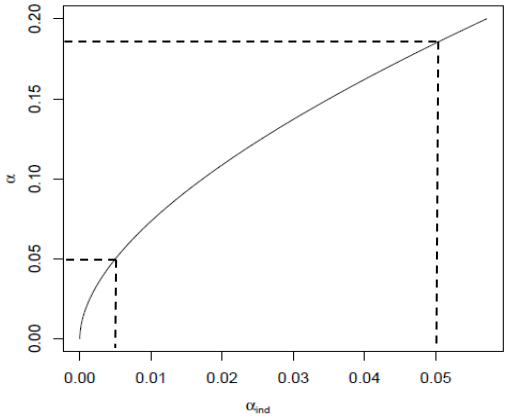


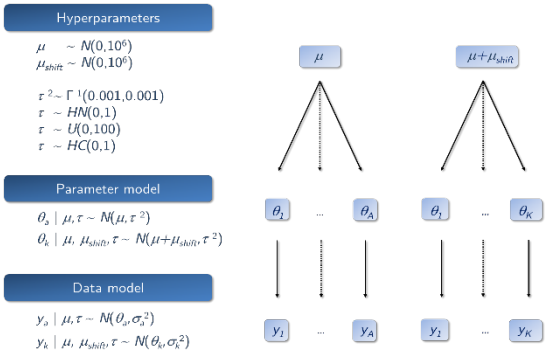




IBKF: Biometrische Forschung + Kooperation in medizinischen Projekten







Vorlesung und Praktikum

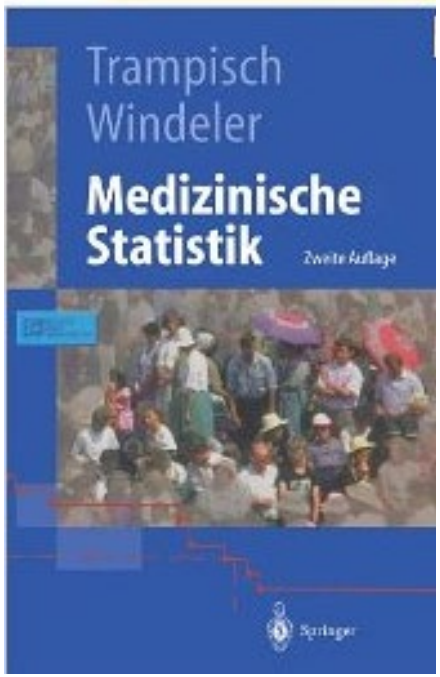
- Vorlesung
  9 Stunden
- Praktikum
  - 4 Doppelstunden
  - Anwesenheitspflicht
- Leistungskontrolle
  - Single-Choice-Prüfungsfragen
  - Stoff der Klausur:
    Vorlesung und Praktikum

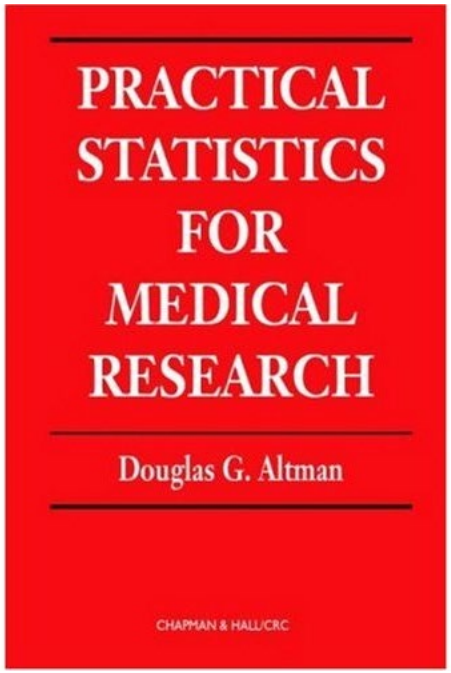

Bushaltestelle Hautklinik
Von-Esmarch-Straße
Bushaltestelle Coesfelder Kreuz
Einsteinstraße
Coesfelder Kreuz
Institut für Biometrie und Klinische Forschung
Schmeddingstraße 56
Albert-Schweitzer-Straße
Rishon-le-Zion-Ring
Bushaltestelle Uni-Klinikum
Domagkstraße
Haupt-zufahrt
Zentralklinikum
Bushaltestelle Schmeddingstr.
Junge-blodt-platz
Schmeddingstraße
Waldeyerstraße
Kardinal-von-Galen-Ring

Vorlesungsinhalt

- Deskriptive Statistik
- Überlebenszeiten
- Statistisches Testen
  - Prinzipien
  - Spezielle Tests
  - Multiples Testen
- Fallzahlplanung / Adaptive Designs
- Regressionsmodelle
  - Logistische Regression
  - Coxregression

Literatur





- Trampisch, H.J., Windeler, J. und Ehle, B.: Medizinische Statistik. 2. Auflage Springer Berlin 2000.
- Altman, D.G.: Practical Statistics for Medical Research. 1. Auflage Chapman & Hall London 1990.

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
- ½ Millionen Fachartikel pro Jahr
- Es gibt viel zu tun: Krebs, AIDS, Chron. Polyarthritis, 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!

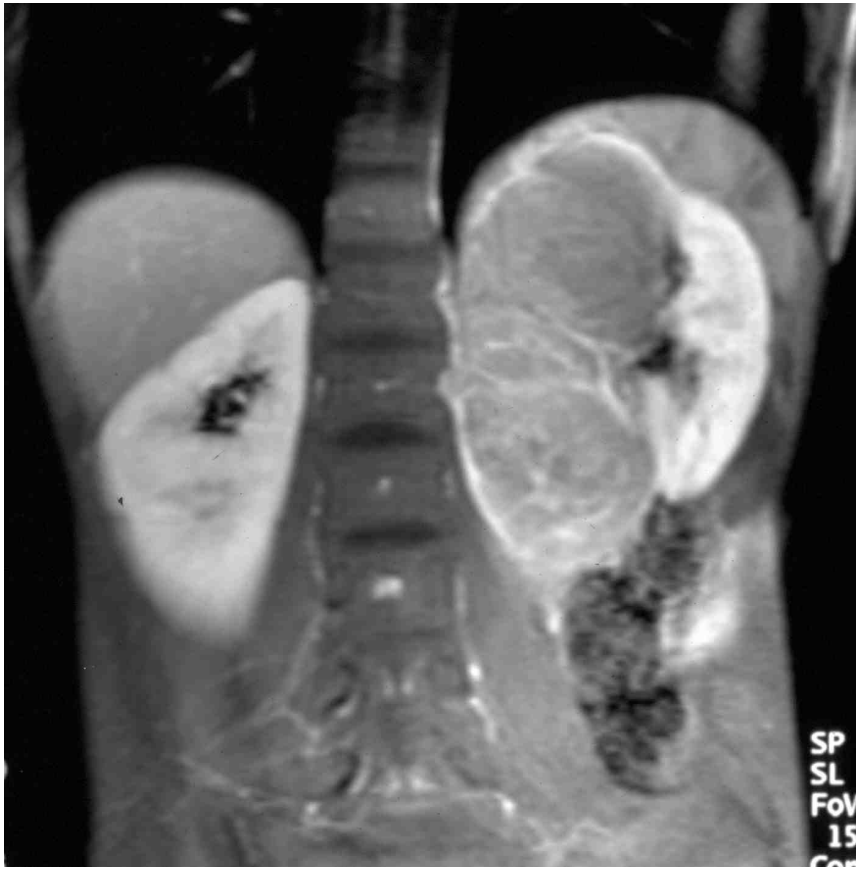
- Aktiv:
  - Promotion
  - Habilitation
  - Wissenschaftler
- Passiv:
  - Prüfarzt
  - Fachzeitschriften
  - Pharmareferent
  - Patienten – Laienartikel

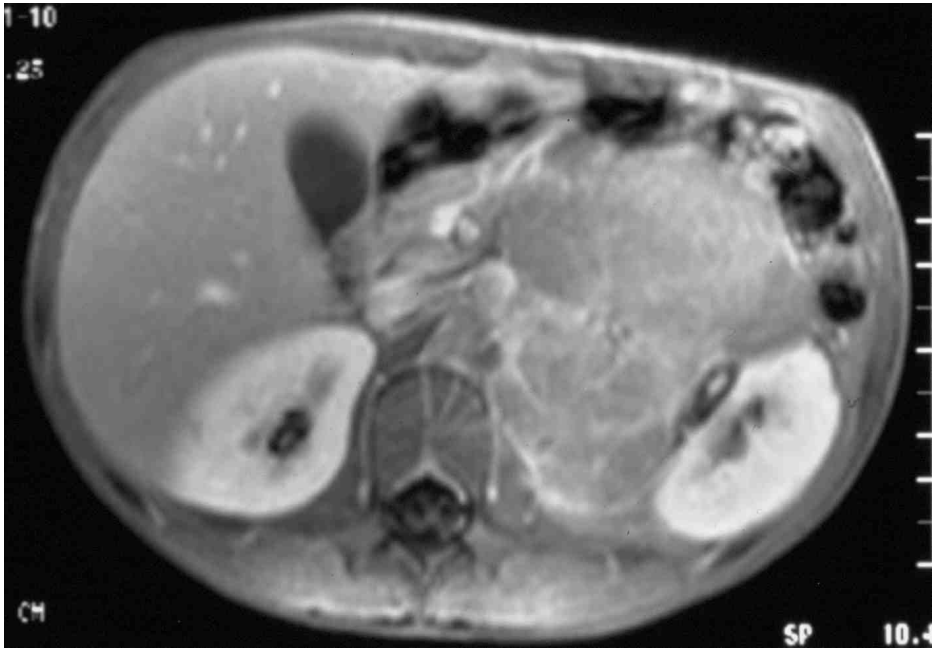


Beispiel:
NB 2004
Studie

Neuroblastom
Stadium 4S

Beispiel: NB 2004 Studie
Tumor linker Grenzstrang





NB 2004 – Neuroblastomstudie bei Kindern

- Maligner Tumor des sympathischen Nervensystems
- Häufigster solider Tumor im Kindesalter (7,3% der Malignome)
  1 Fall / 100000 Kinder unter 15 Jahren pro Jahr
- Symptome: Schmerz, Fieber, tastbarer Bauchtumor, Diarrhö
- 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

Studienidee – NB 2004

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

Vorlesungsinhalt

- Deskriptive Statistik
- Überlebenszeiten
- Statistisches Testen
  - Prinzipien
  - Spezielle Tests
  - Multiples Testen
- Fallzahlplanung / Adaptive Designs
- Regressionsmodelle
  - Logistische Regression
  - Coxregression

Deskriptive Statistik

- Grundbegriffe
- kategoriale Variablen
- stetige Variablen
- Zusammenhang 2er Variablen

INRG a 08-11-15d.sav [DatenSet2] - IBM SPSS Statistics Daten-Editor

Datei Bearbeiten Ansicht Daten Transformieren Analysieren Direktmarketing Diagramme Extras Fenster Hilfe

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
- Grundgesamtheit
  - Definition: Alle potentiellen Beobachtungseinheiten
  - NB 2004: Alle an einem Neuroblastom erkrankte PatientInnen unter 21 Jahren
- Stichprobe
  - Definition: Alle untersuchten Beobachtungseinheiten
  - NB 2004: Teilnehmer der NB 2004 Studie

Studieneinheiten

- Beobachtungseinheit
  - Definition: kleinste Einheit, die beobachtet wird
  - NB 2004: An einem Neuroblastom erkrankte/r PatientIn
- Grundgesamtheit
  - Definition: Alle potentiellen Beobachtungseinheiten
  - NB 2004: Alle an einem Neuroblastom erkrankte PatientInnen unter 21 Jahren
- 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 — 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

- Grundbegriffe
- kategoriale Variablen
- stetige Variablen
- Zusammenhang 2er Variablen

INSS-Stadium

Relative Häufigkeit:

Absolute Häufigkeit / Gesamtzahl mit einer beobachteten Ausprägung

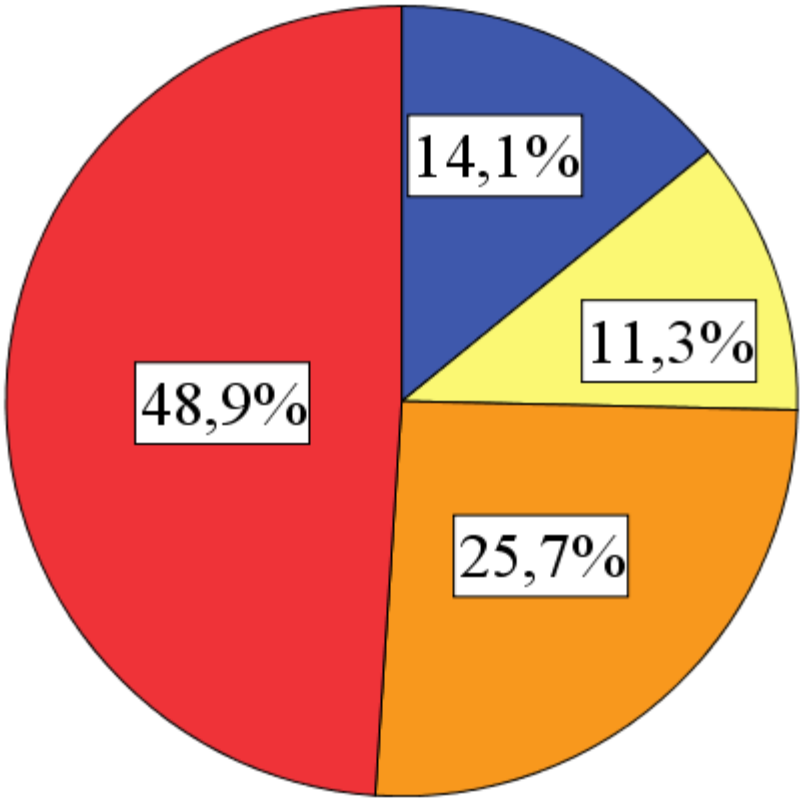
		Häufigke	Prozent	Gültige Prozente
Gültig	1	356	10,6	14,1
	2	284	8,5	11,3
	3	649	19,4	25,7
	4	1235	36,9	48,9
	Gesamt	2524	75,5	100,0
Fehlend	System	821	24,5	
Gesamt		3345	100,0	

Faustregel:

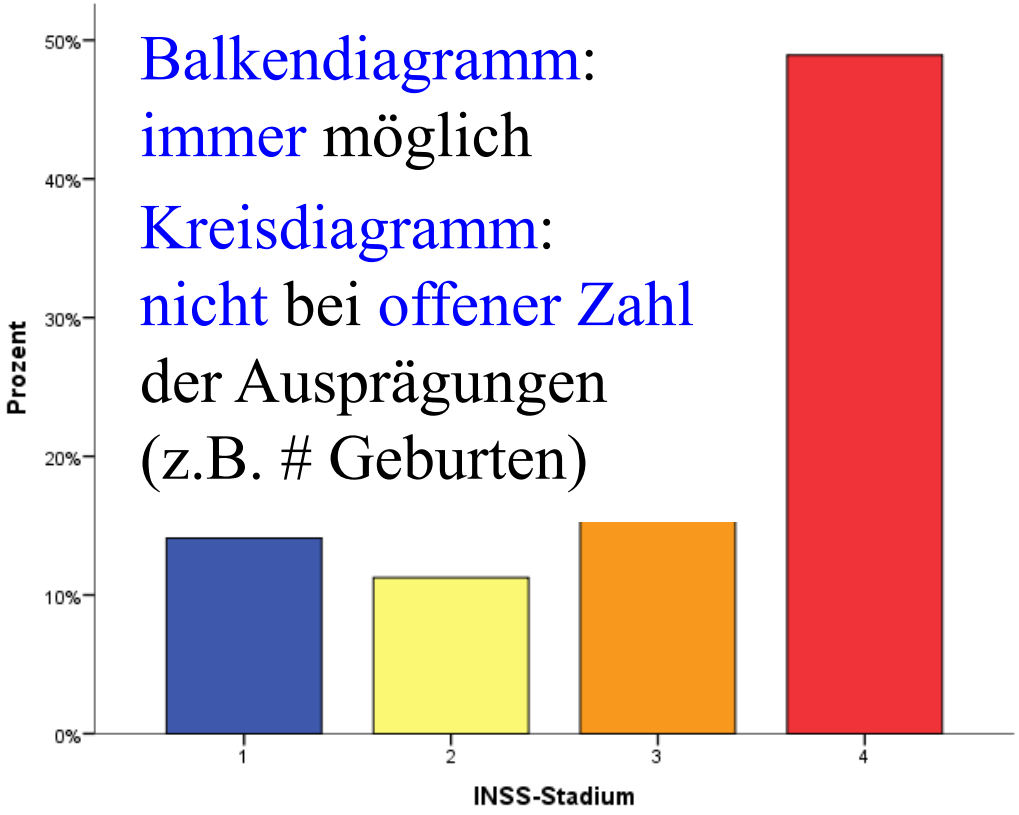
- Gesamtzahl ≤ 30: keine Prozentzahl
- 30 < Gesamtzahl ≤ 300: Prozentzahl ohne Nachkommastelle
- 300 < Gesamtzahl ≤ 3000: Prozentzahl mit 1 Nachkommastelle
- …
- Fehlende Werte angeben!

INSS-Stadium

Kreisdiagramm



Balkendiagramm



Balkendiagramm:
immer möglich

Kreisdiagramm:
nicht bei offener Zahl
der Ausprägungen
(z.B. # Geburten)

Deskriptive Statistik

- Grundbegriffe
- kategoriale Variablen
- stetige Variablen
- Zusammenhang 2er 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

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

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

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

Lagemaße bei stetigen Variablen

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

  Minimum = 2,

  Maximum = 14 (85),

Lagemaße bei stetigen Variablen

1. Quartil Q1:
   - mindestens 25% der Werte sind kleiner oder gleich
   - mindestens 75% der Werte sind größer oder gleich
2. Quartil Q2:
   - mindestens 50% der Werte sind kleiner oder gleich
   - mindestens 50% der Werte sind größer oder gleich
3. Quartil Q3:
   - 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

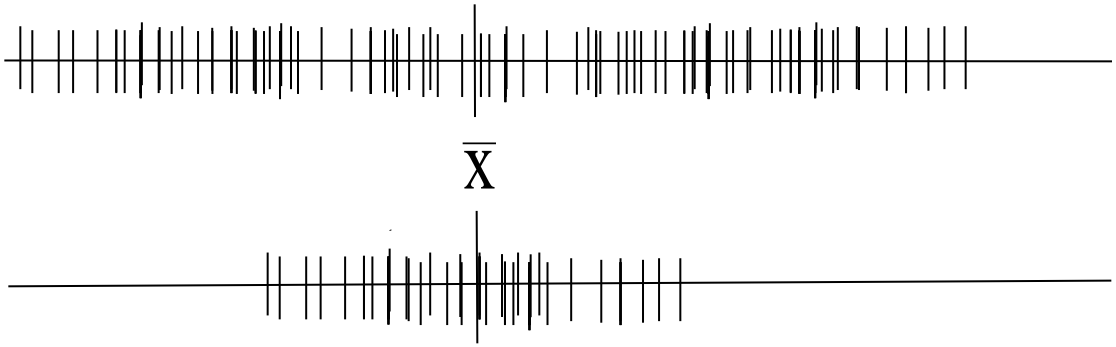
- Konzept der Quartile:
  - Median: teilt Beobachtungen 50% : 50% auf
  - Quartile: teilen Beobachtungen ~ 25% : 25% : 25% : 25%
- Messwerte: 3, 8, 4, 11, 10, 1, 7, 12, 9, 2, 5, 6
  - Geordnete Messwerte: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12
  - Median = (6 + 7) / 2 = 6,5
  - Q1 = 3,5; Q2 = 6,5; Q3 = 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

$$\text{Var} = \frac{\sum_{\text{Stichprobe}} (\text{Einzelwert} - \text{Mittelwert})^2}{\text{Stichprobenumfang} - 1}$$

- Standardabweichung (std) = $\sqrt{\text{Var}}$



std groß

std klein

Streuungsmaße bei stetigen Variablen

- Frage: Wie weit liegen die Werte auseinander?
- Varianz (Var): Quadratische Abweichungen vom Mittelwert

$$\text{Var} = \frac{\sum_{\text{Stichprobe}} (\text{Einzelwert} - \text{Mittelwert})^2}{\text{Stichprobenumfang} - 1}$$

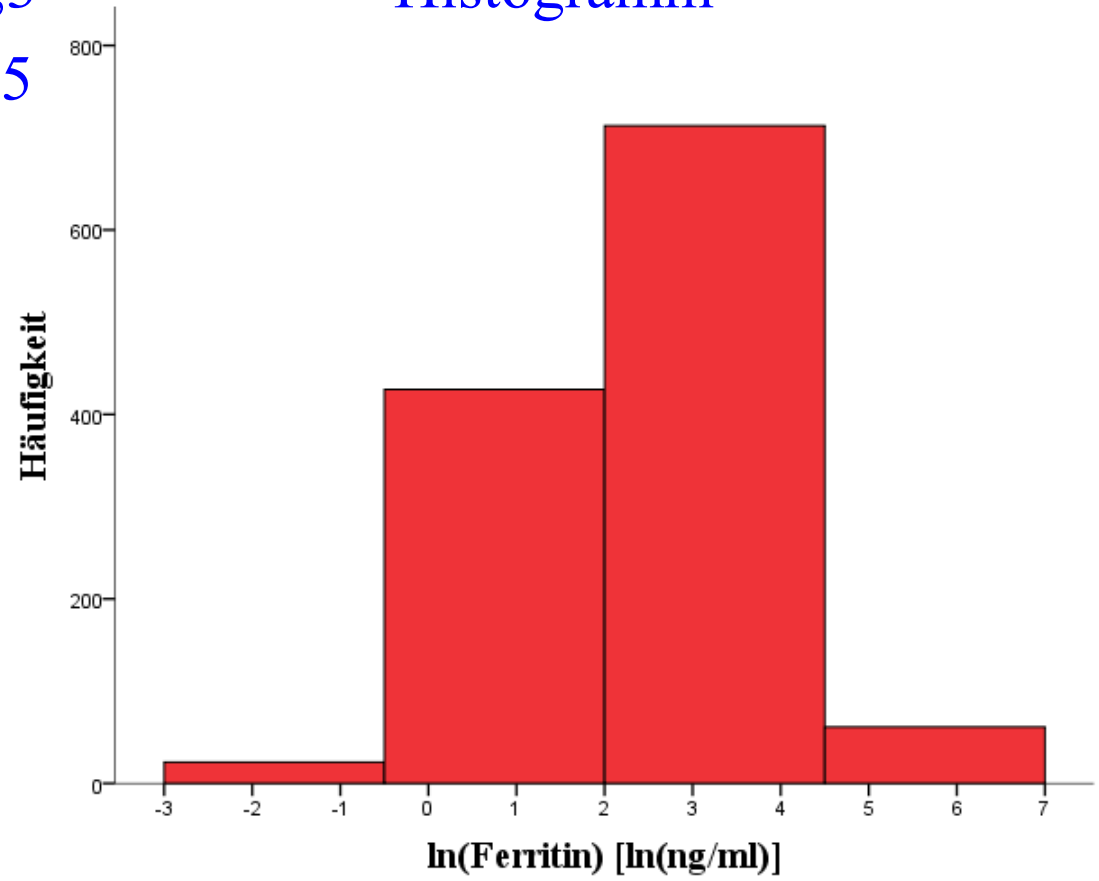
- Standardabweichung (std) = $\sqrt{\text{Var}}$
- Varianz und Standardabweichung: reagieren empfindlich auf extreme Werte, da mittelwertbasiert & Mittelwert empfindlich
- Alternativ: - Spannweite: Maximum – Minimum
  - Interquartilsabstand: 3. Quartil – 1. Quartil

Neuroblastom - logarithmiertes Serum Ferritin (ng/ml)

Minimum: -2,3

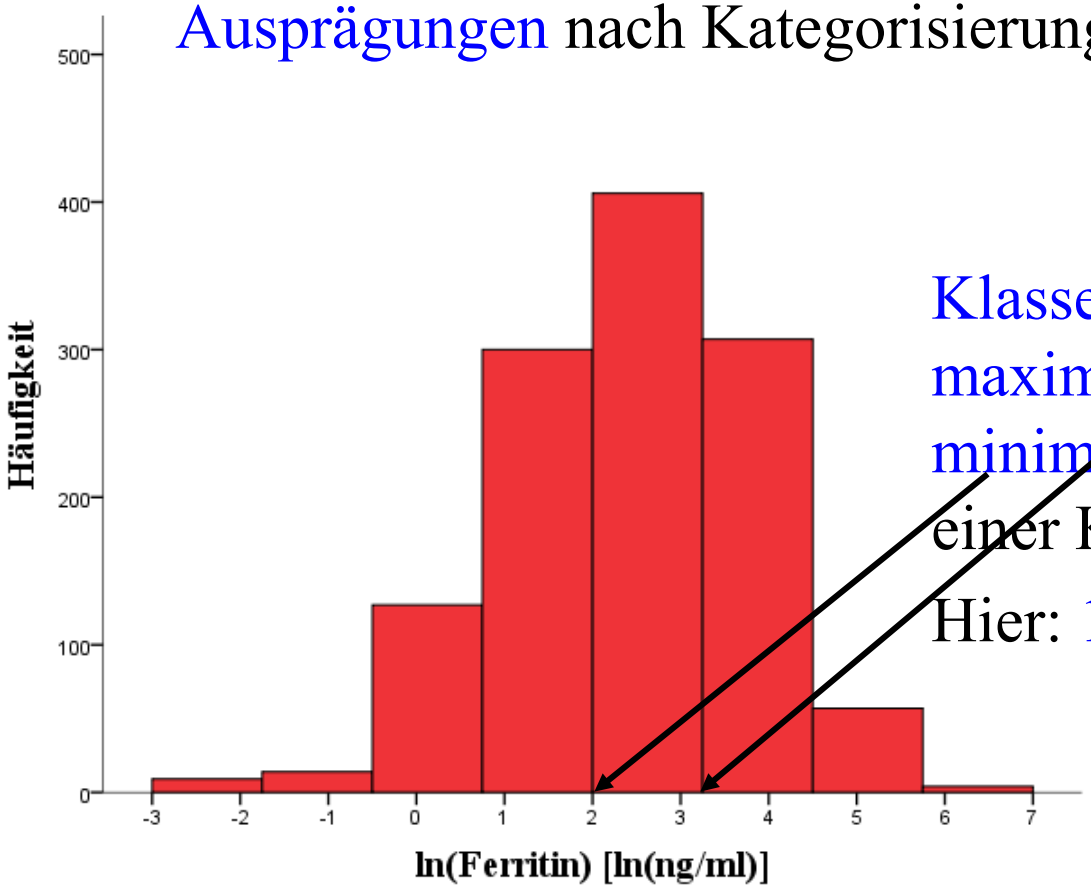
Maximum: 6,5

Histogramm



Neuroblastom - logarithmiertes Serum Ferritin (ng/ml)

Klassenzahl = 8 = Zahl der möglichen Ausprägungen nach Kategorisierung



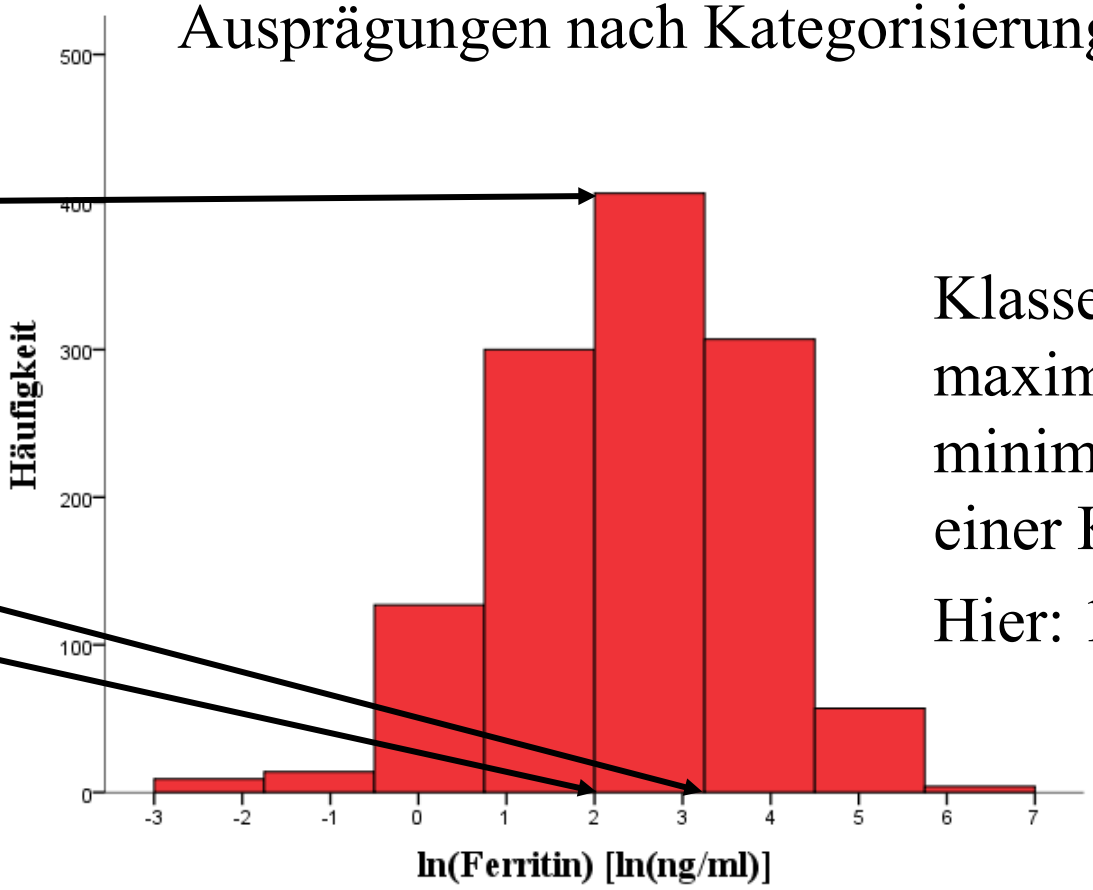
Klassenbreite = maximaler – minimaler Wert einer Kategorie

Hier: 1,25 ln(ng/ml)

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(ng/ml)

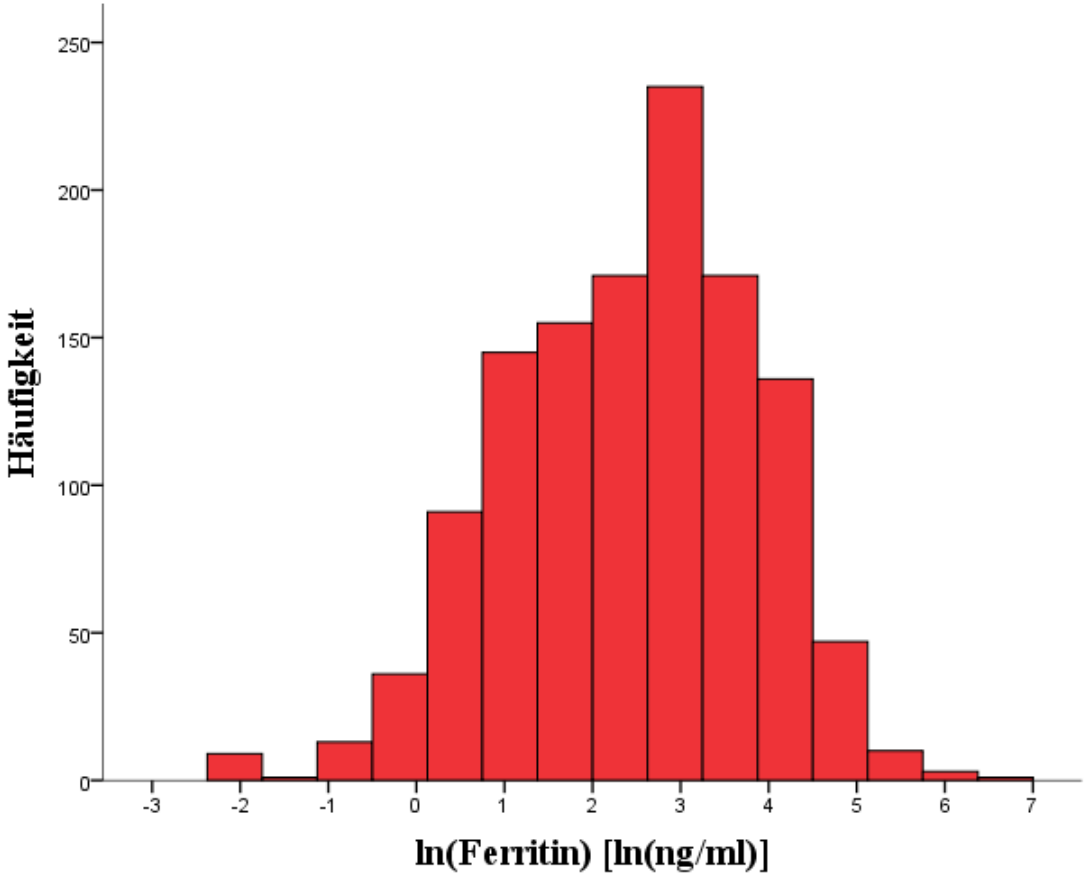


Klassenbreite = maximaler – minimaler Wert einer Kategorie

Hier: 1,25 ln(ng/ml)

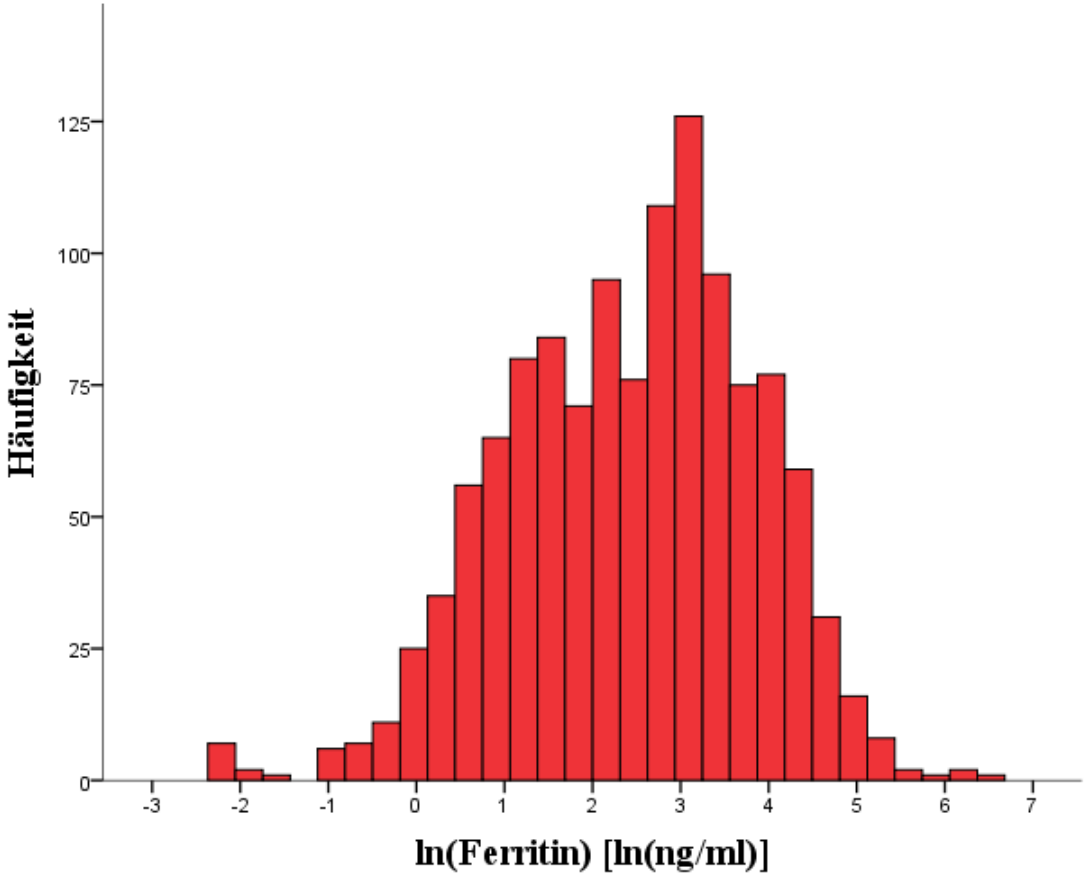
Neuroblastom - logarithmiertes Serum Ferritin (ng/ml)

Klassenzahl = 16



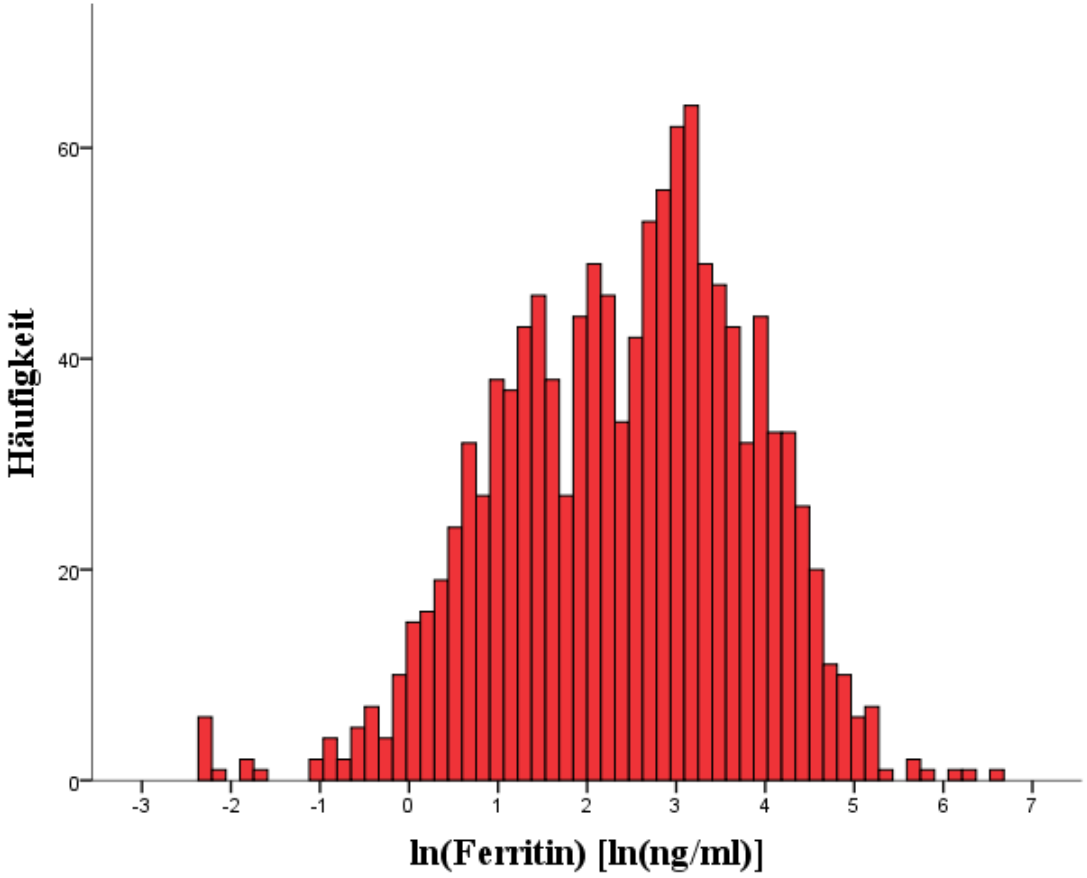
Neuroblastom - logarithmiertes Serum Ferritin (ng/ml)

Klassenzahl = 32



Neuroblastom - logarithmiertes Serum Ferritin (ng/ml)

Klassenzahl = 64

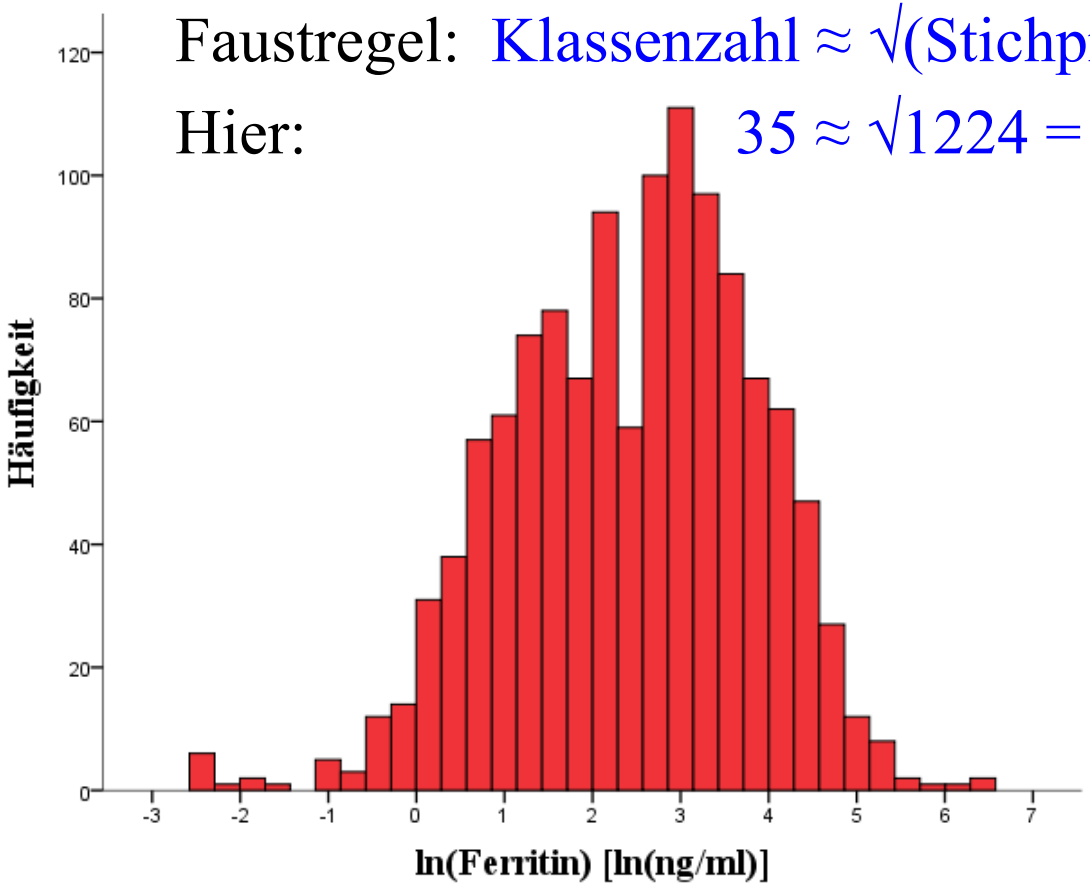


Neuroblastom - logarithmiertes Serum Ferritin (ng/ml)

Klassenzahl = 35

Faustregel: Klassenzahl ≈ √(Stichprobenumfang)

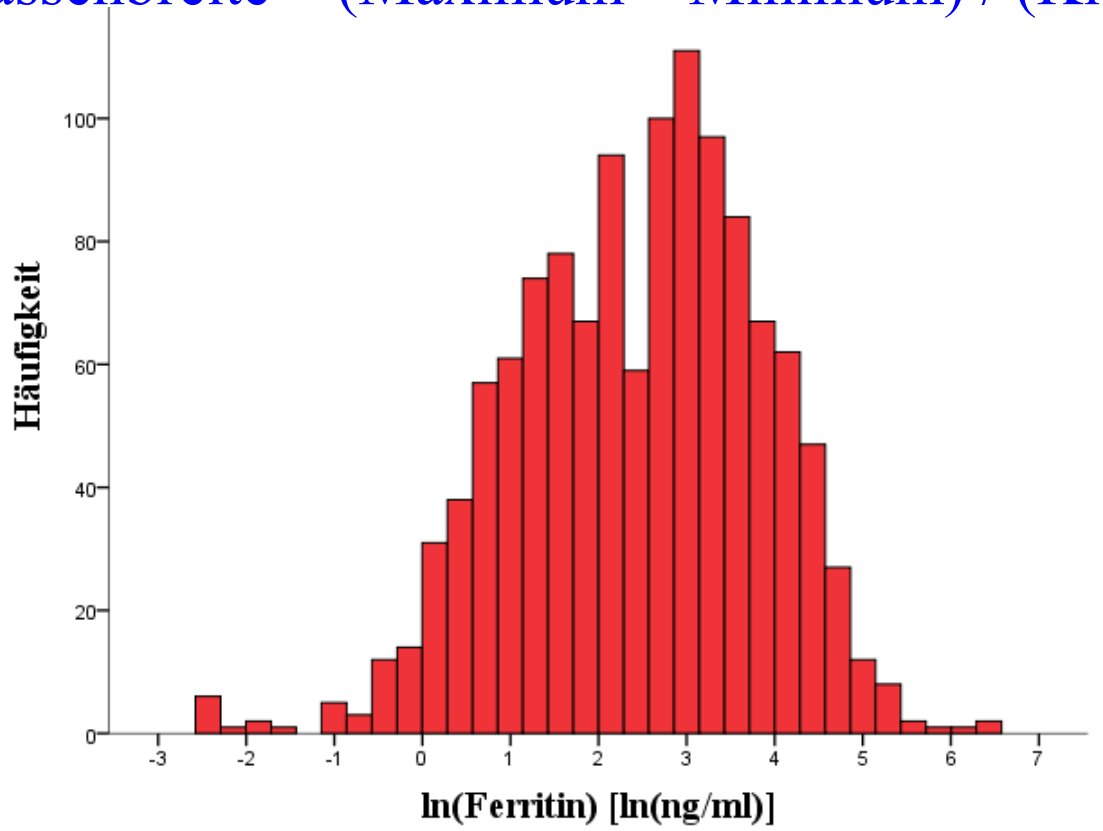
Hier: $35 \approx \sqrt{1224} = 34{,}99$



Neuroblastom - logarithmiertes Serum Ferritin (ng/ml)

Klassenzahl = 35

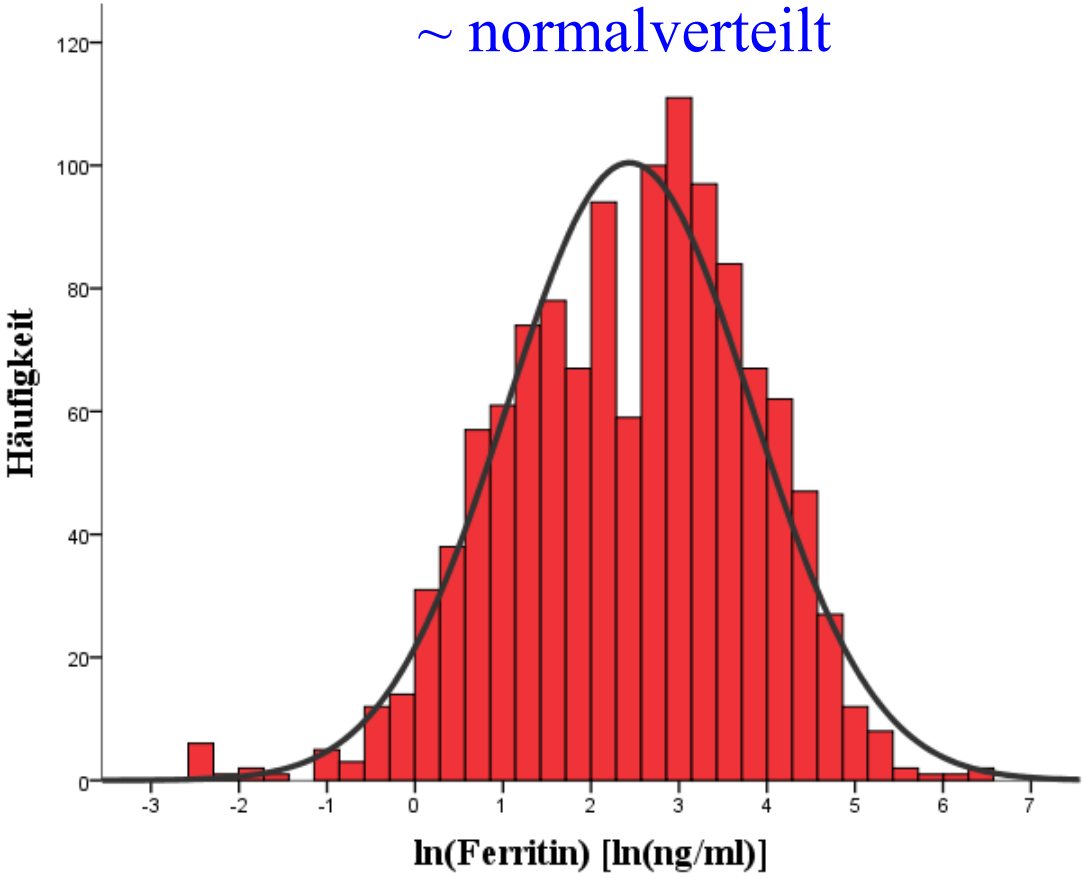
Faustregel: Klassenbreite ≈ (Maximum – Minimum) / (Klassenzahl - 1)



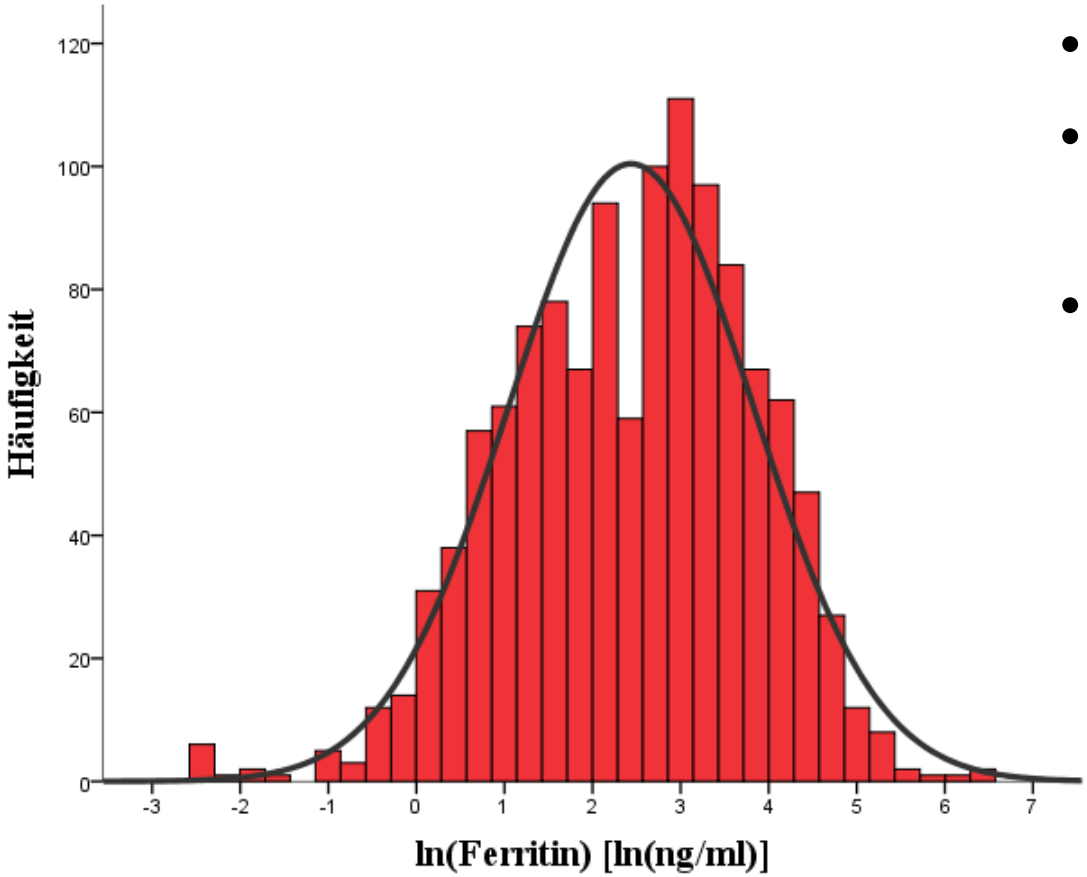
Neuroblastom - logarithmiertes Serum Ferritin (ng/ml)

symmetrisch

~ normalverteilt



Neuroblastom - logarithmiertes Serum Ferritin (ng/ml)



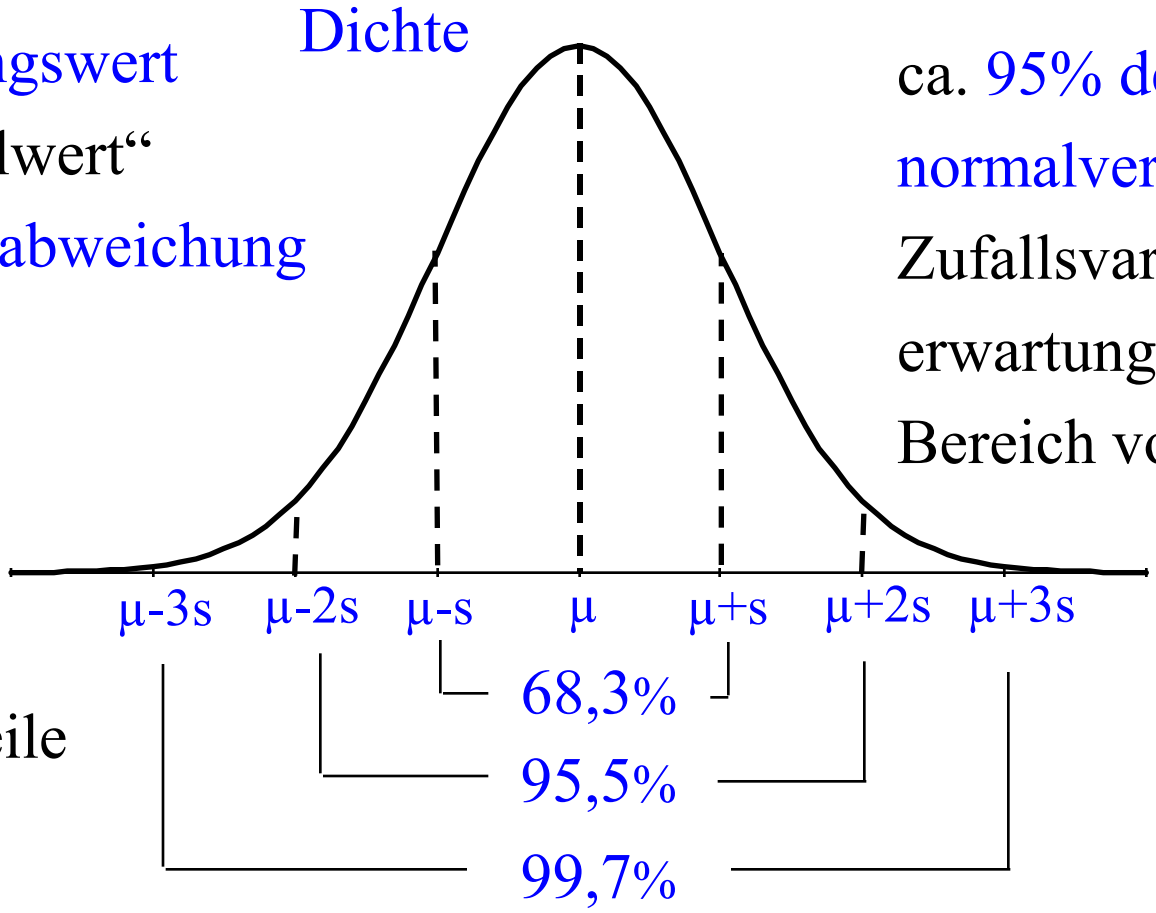
Normalverteilung:

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

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

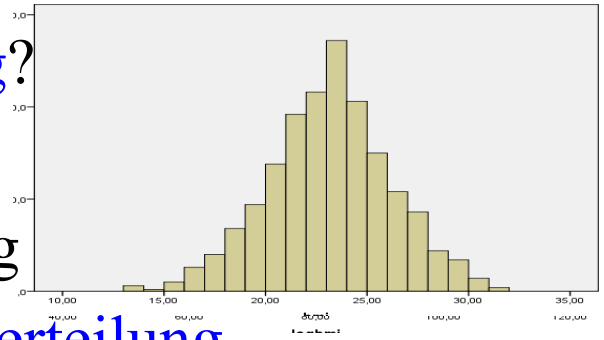
Normalverteilung

μ: Erwartungswert
= „Mittelwert“
σ: Standardabweichung



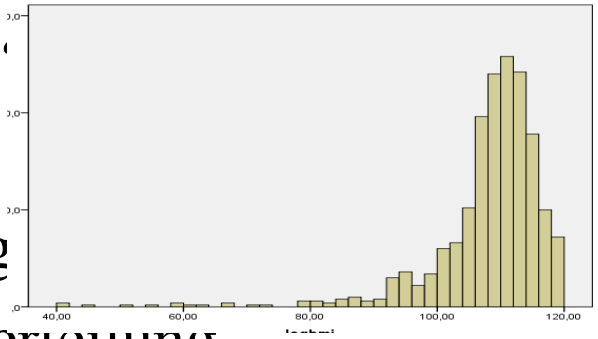
ca. 95% der Werte einer normalverteilten Zufallsvariablen liegen erwartungsgemäß im Bereich von $\mu \pm 1{,}96 \cdot \sigma$.

Verteilungsmaße und -formen bei stetigen Variablen



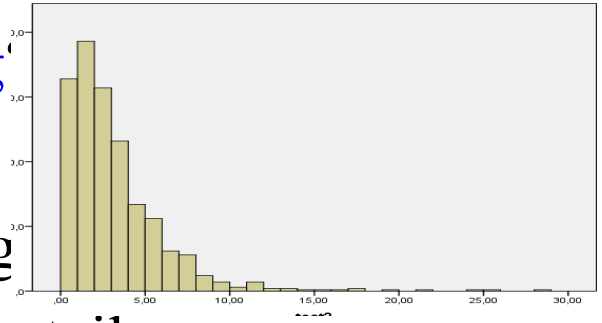
- Frage: Streben die Werte in eine Richtung?
- Schiefe (g):
  - g = 0: perfekt symmetrische Verteilung
    z.B. Gaußverteilung = Normalverteilung
  - -1≤ g ≤ +1: akzeptable symmetrische Verteilung
    Mittelwert ≈ Median
  - g < -1: Verteilung ist linksschief = rechtsgipflig
    Mittelwert << Median
  - g > +1: Verteilung ist rechtsschief = linksgipflig
    Mittelwert >> Median

Verteilungsmaße und -formen bei stetigen Variablen



- Frage: Streben die Werte in eine Richtung?
- Schiefe (g):
  - g = 0: perfekt symmetrische Verteilung
    z.B. Gaußverteilung = Normalverteilung
  - -1≤ g ≤ +1: akzeptable symmetrische Verteilung
    Mittelwert ≈ Median
  - g < -1: Verteilung ist linksschief = rechtsgipflig
    Mittelwert << Median
  - g > +1: Verteilung ist rechtsschief = linksgipflig
    Mittelwert >> Median

Verteilungsmaße und -formen bei stetigen Variablen



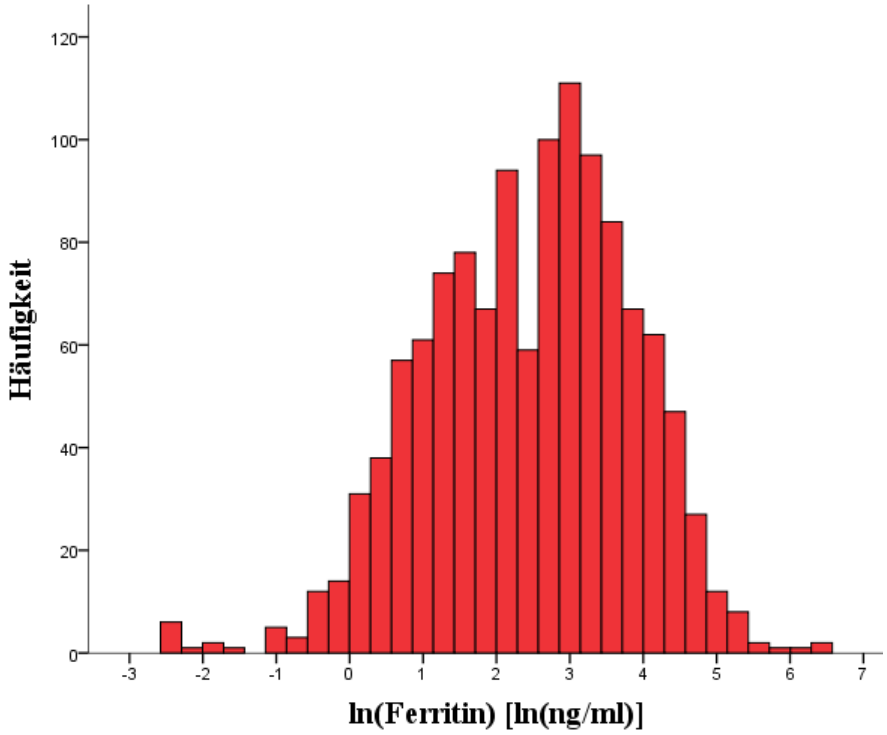
- Frage: Streben die Werte in eine Richtung?
- Schiefe (g):
  - g = 0: perfekt symmetrische Verteilung
    z.B. Gaußverteilung = Normalverteilung
  - -1≤ g ≤ +1: akzeptable symmetrische Verteilung
    Mittelwert ≈ Median
  - g < -1: Verteilung ist linksschief = rechtsgipflig
    Mittelwert << Median
  - g > +1: Verteilung ist rechtsschief = linksgipflig
    Mittelwert >> Median

Welche Maße wann publizieren?

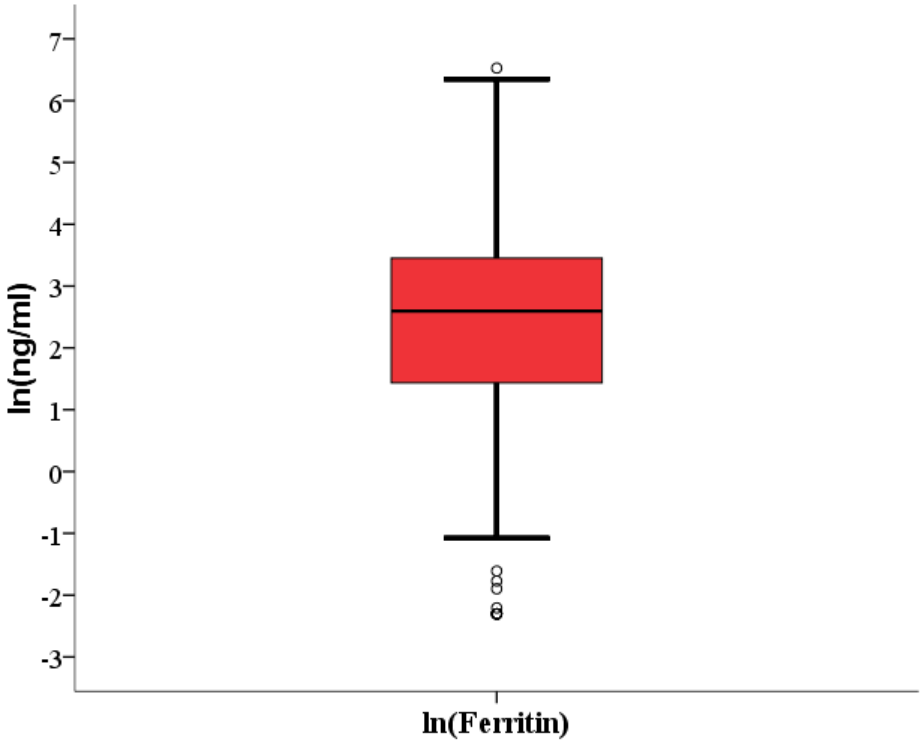
Ausnahmen bestätigen die Regel!	|Schiefe| ≤ 1	|Schiefe| > 1
Mittelwert	sinnvoll	oft unsinnig
Varianz, Standardabweichung	sinnvoll	oft unsinnig
Median	sinnvoll	sinnvoll
Quartile, Interquartilsabstand	sinnvoll	sinnvoll
Minimum, Maximum, Spannweite	(sinnvoll)	(sinnvoll)

Graphische Darstellung stetiger Variablen

Histogramm



Box-and-Whisker Plot

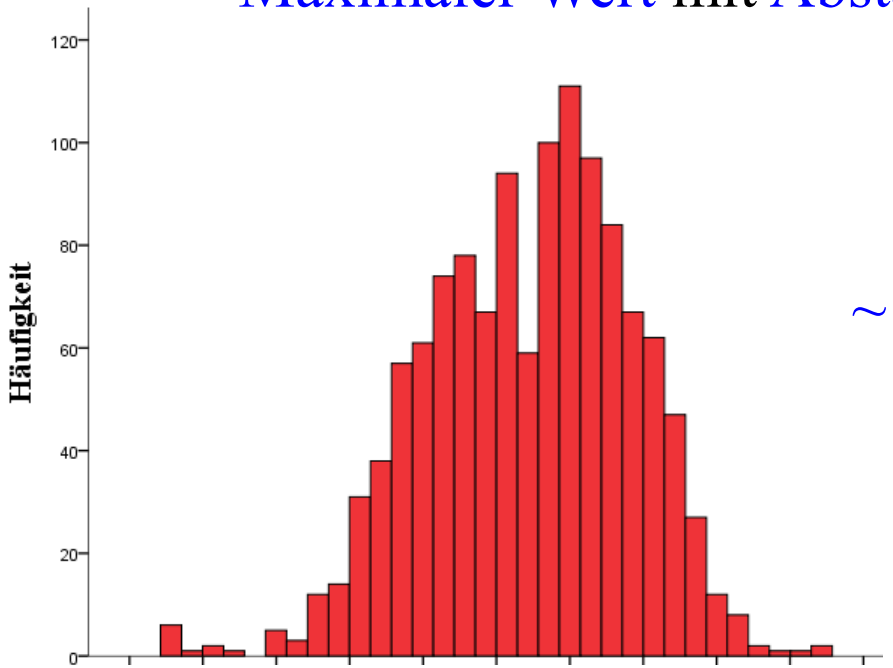


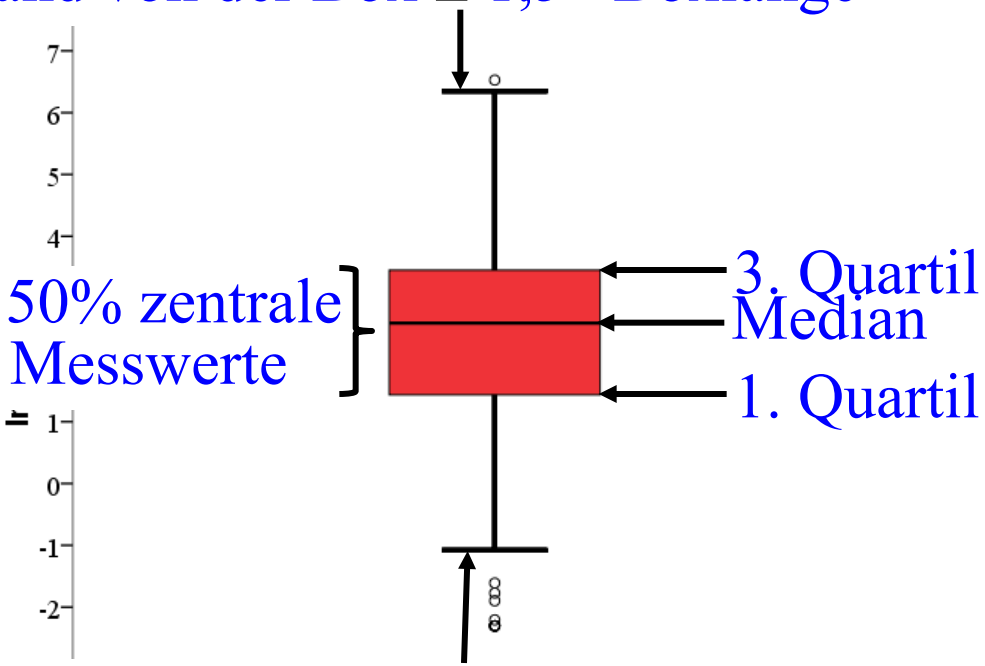
Graphische Darstellung stetiger Variablen

Histogramm

Box-and-Whisker Plot

Maximaler Wert mit Abstand von der Box ≤ 1,5 · Boxlänge

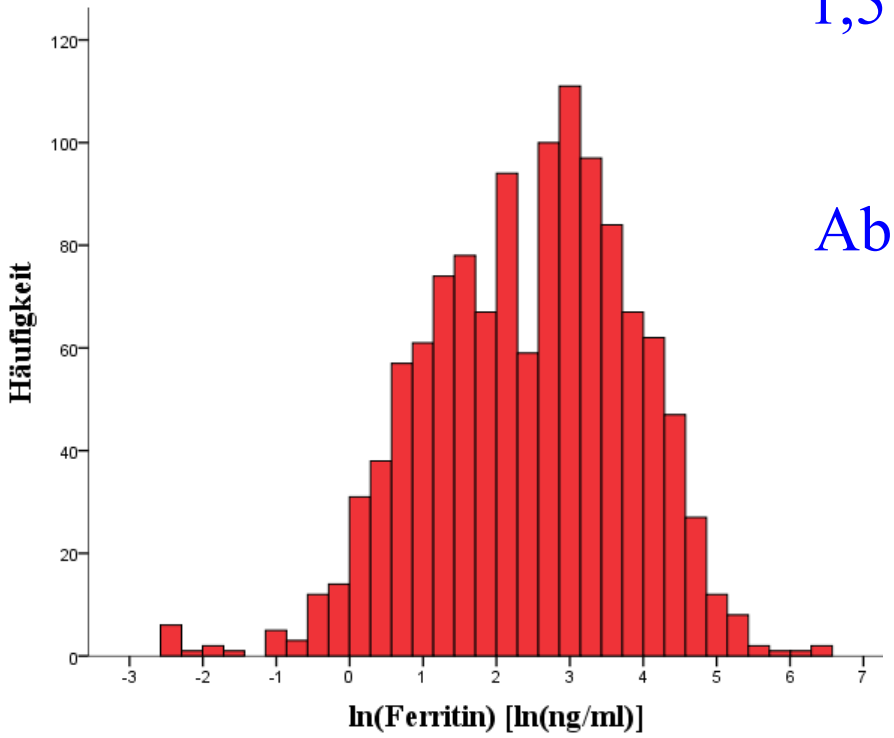




Minimaler Wert mit Abstand von der Box ≤ 1,5 · Boxlänge

Graphische Darstellung stetiger Variablen

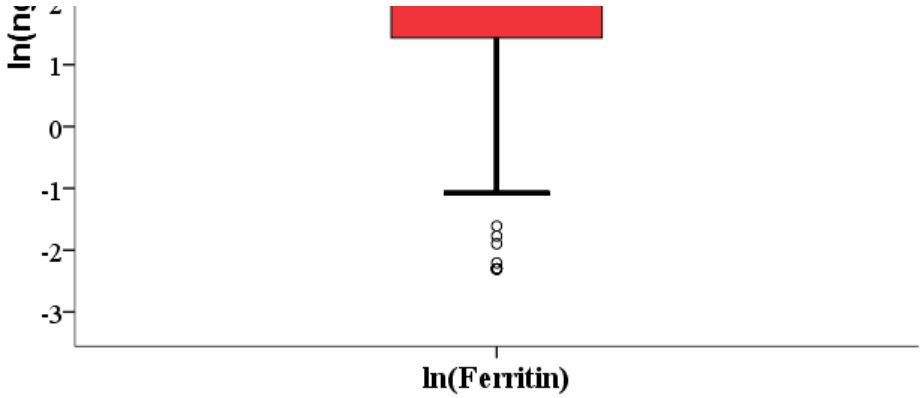
Histogramm



Box-and-Whisker Plot

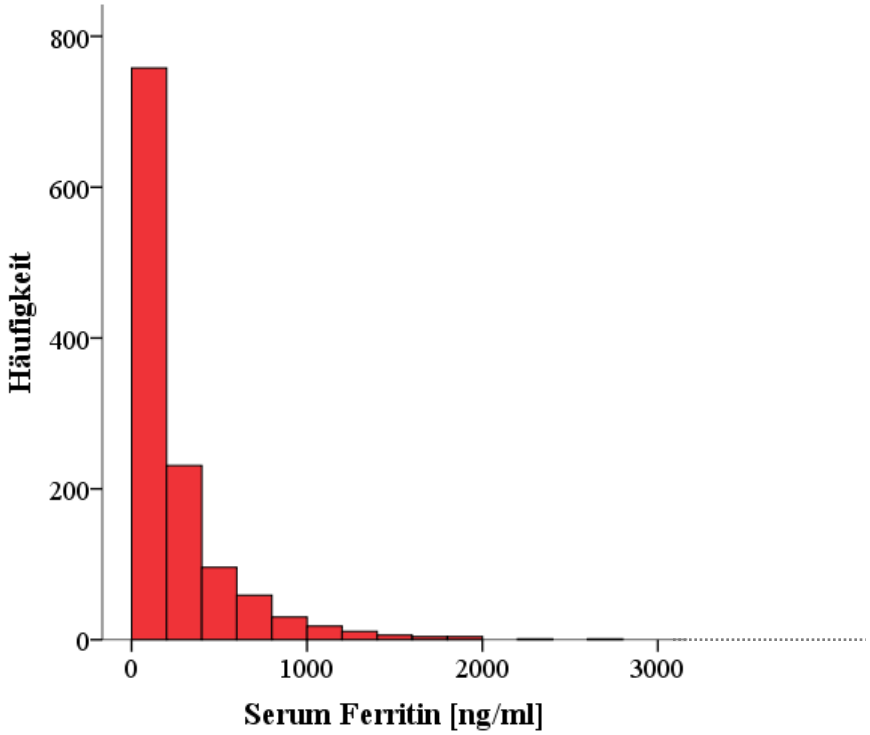
1,5 · Boxlänge < Abstand von der Box ≤ 3 · Boxlänge: Ausreißer o

Abstand von der Box > 3 · Boxlänge: Extremwert *

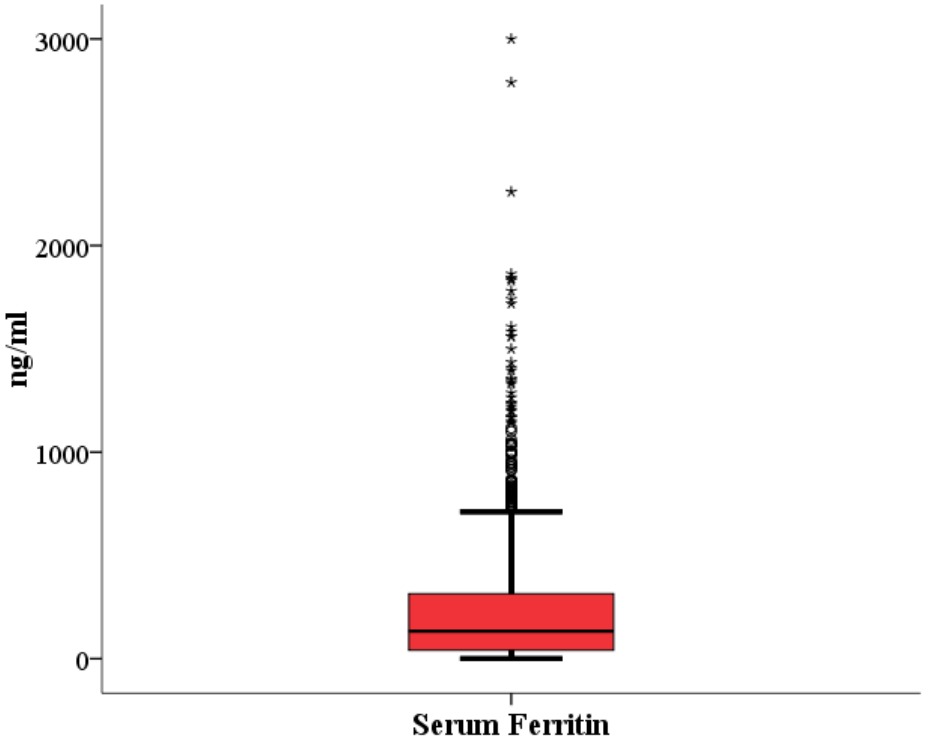


Graphische Darstellung stetiger Variablen

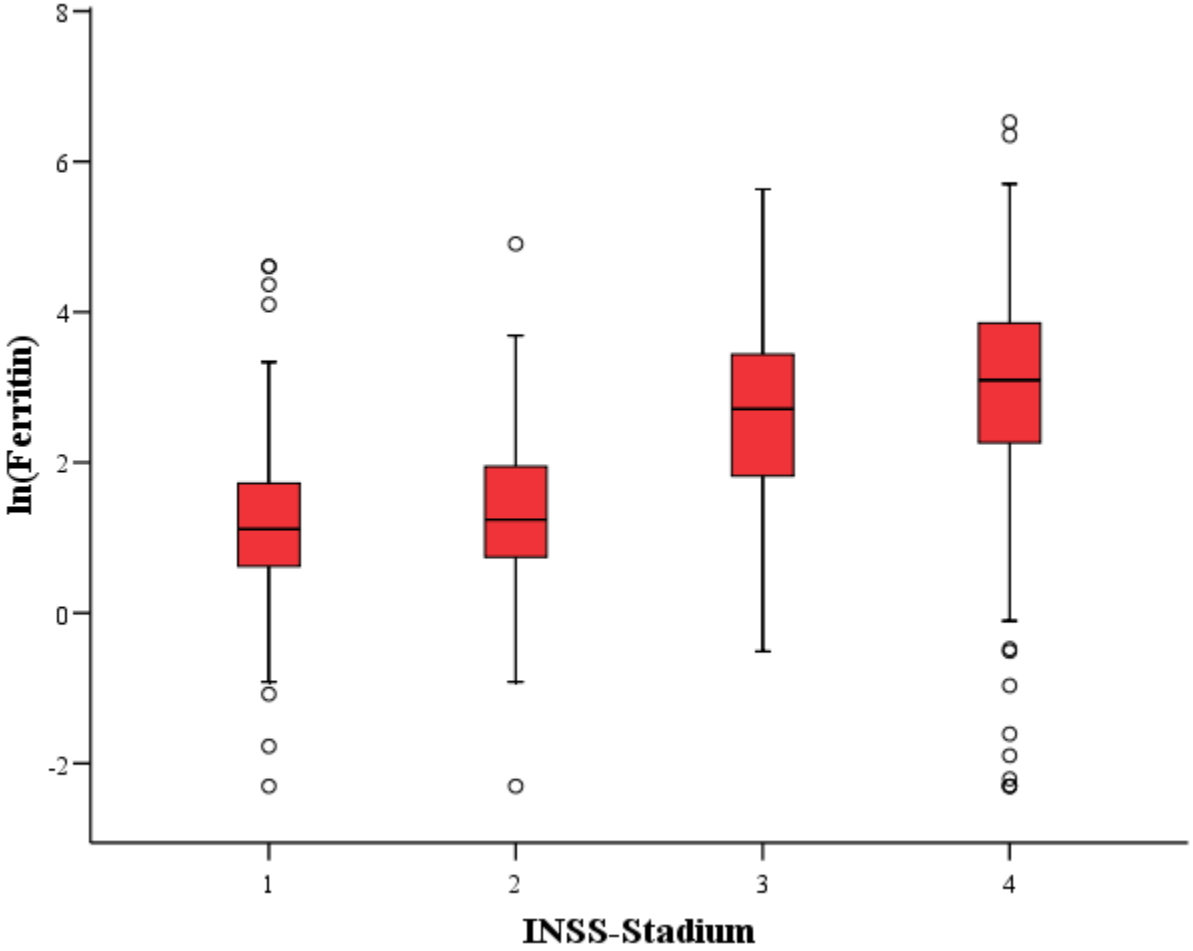
Histogramm



Box-and-Whisker Plot



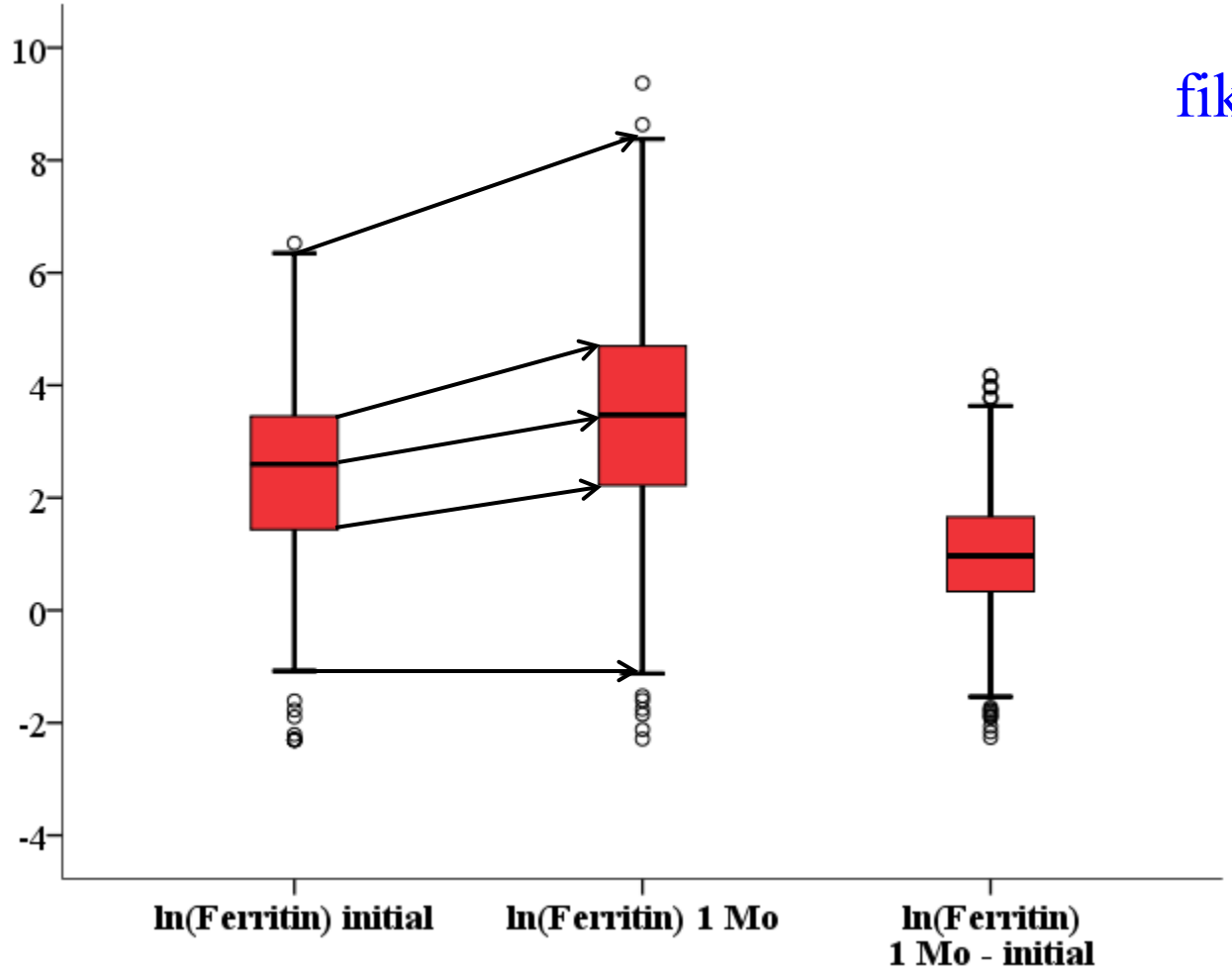
Gruppierter Boxplot



Verbundenes Design

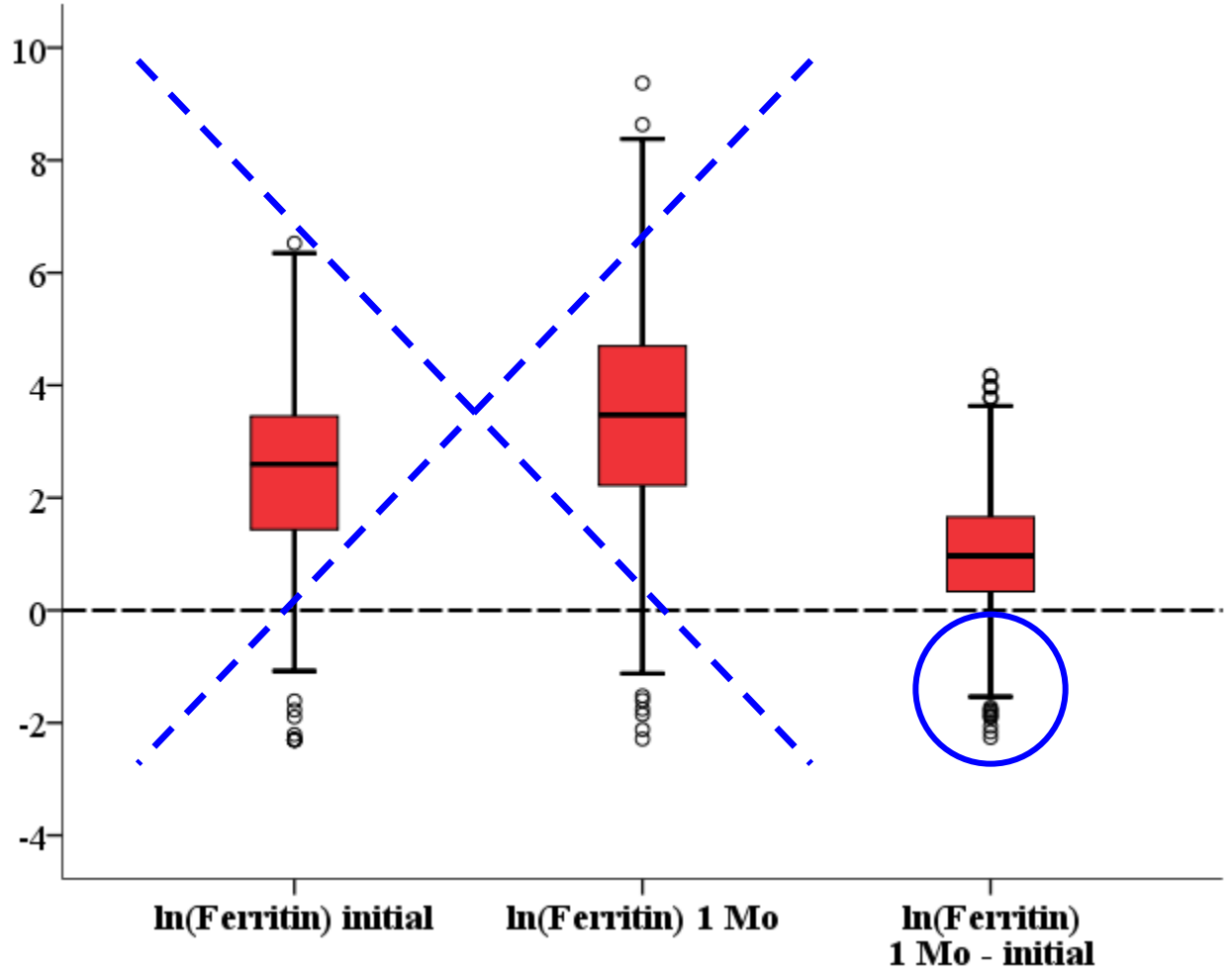
- Verbundenes Design
  - Dieselbe Beobachtungseinheit:
    z. B. prä- und postoperativer Blutdruck
    Unterschied eines Tests am rechten und linken Auge
  - Beobachtungspaare:
    1 Fall and 1 Kontrolle mit ähnlichen Eigenschaften
    z. B. 1 leukämiekrankes Kind und 1 Kind gleichen Alters
  - Vorteil: Reduktion der Variabilität
- Vorgehensweise:
  Boxplot der Differenzen der gepaarten Werte statt
  2 Einzelboxplots

Differenzen-Boxplot

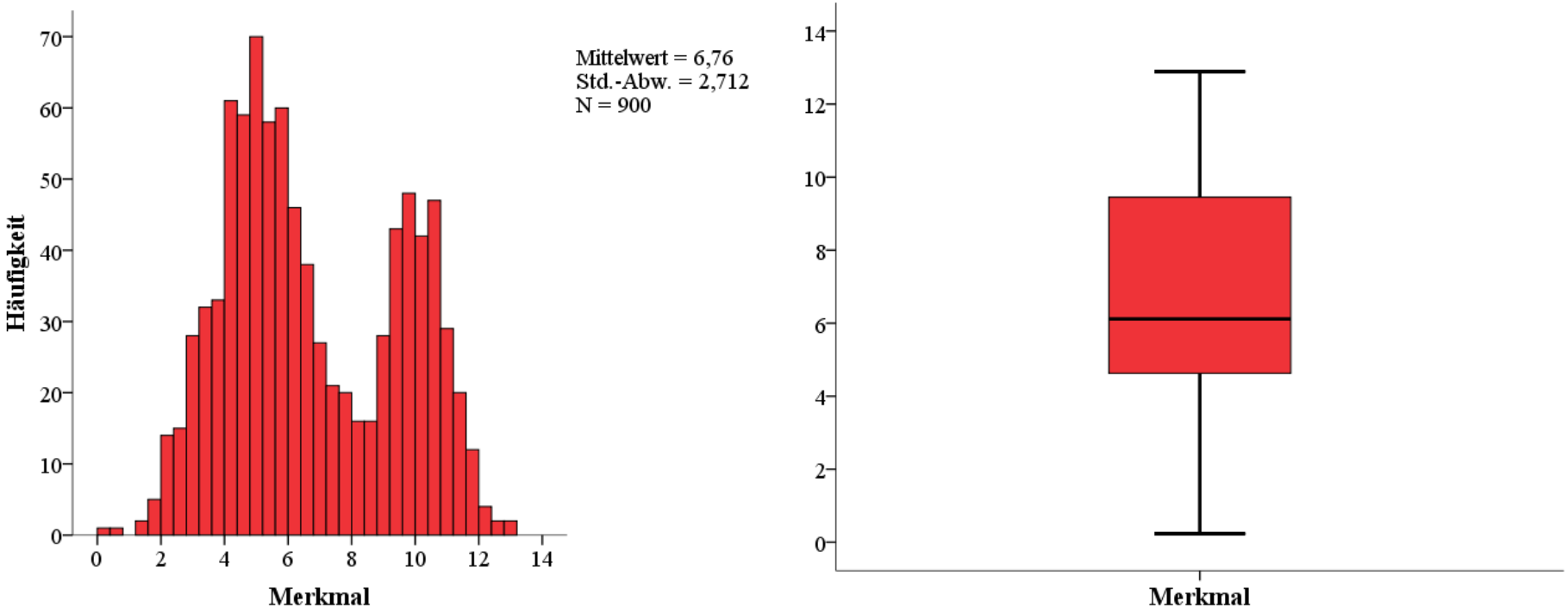


fiktiv

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: Mittelwertsbalken + Standardfehler

Deskriptive Statistik

- Grundbegriffe
- kategoriale Variablen
- stetige Variablen
- Zusammenhang 2er Variablen

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?

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?

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?

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?

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

Vergleich der Risikofaktoren

Folgende Fragen sollen geklärt werden:

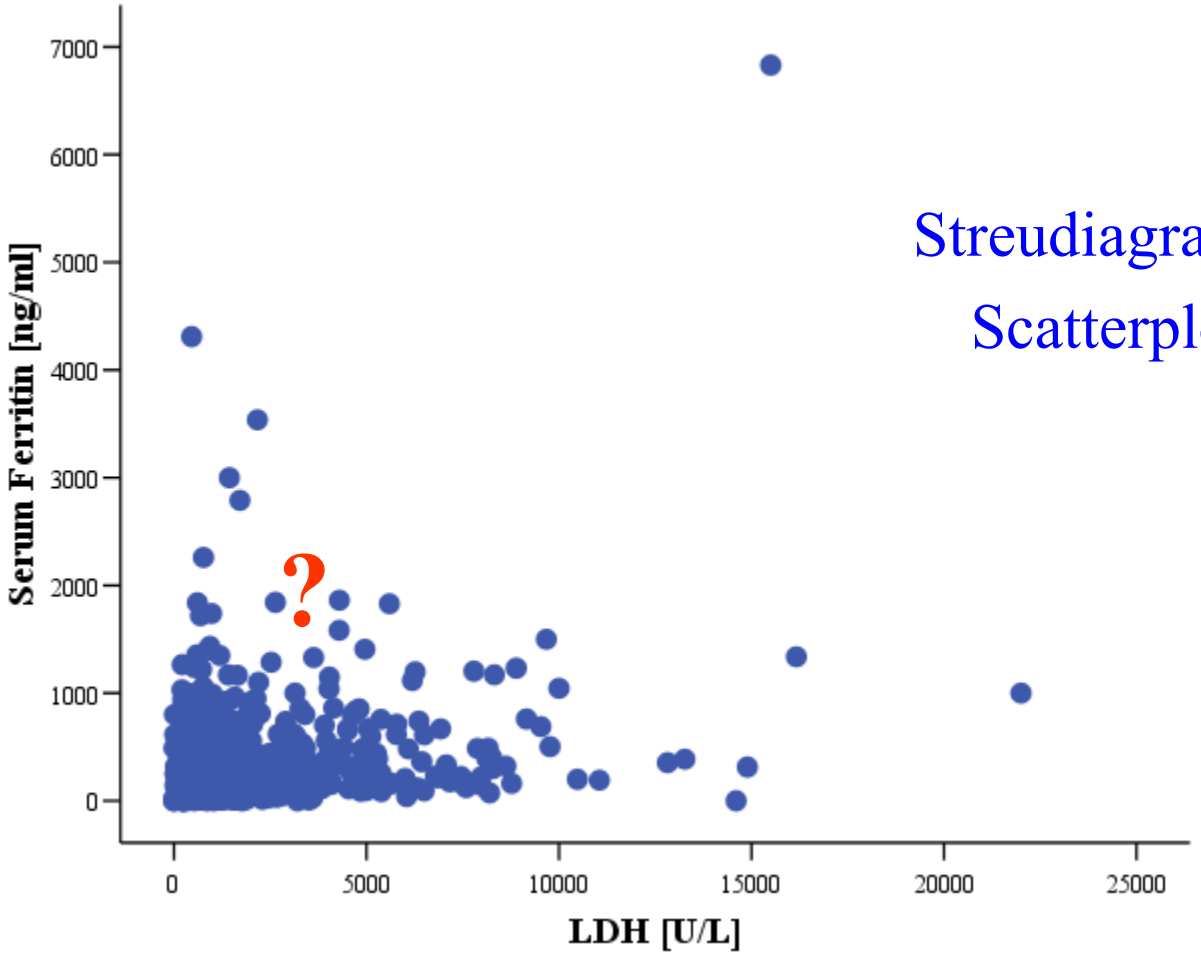
1. Gibt es einen Zusammenhang zwischen LDH und Ferritin?

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?

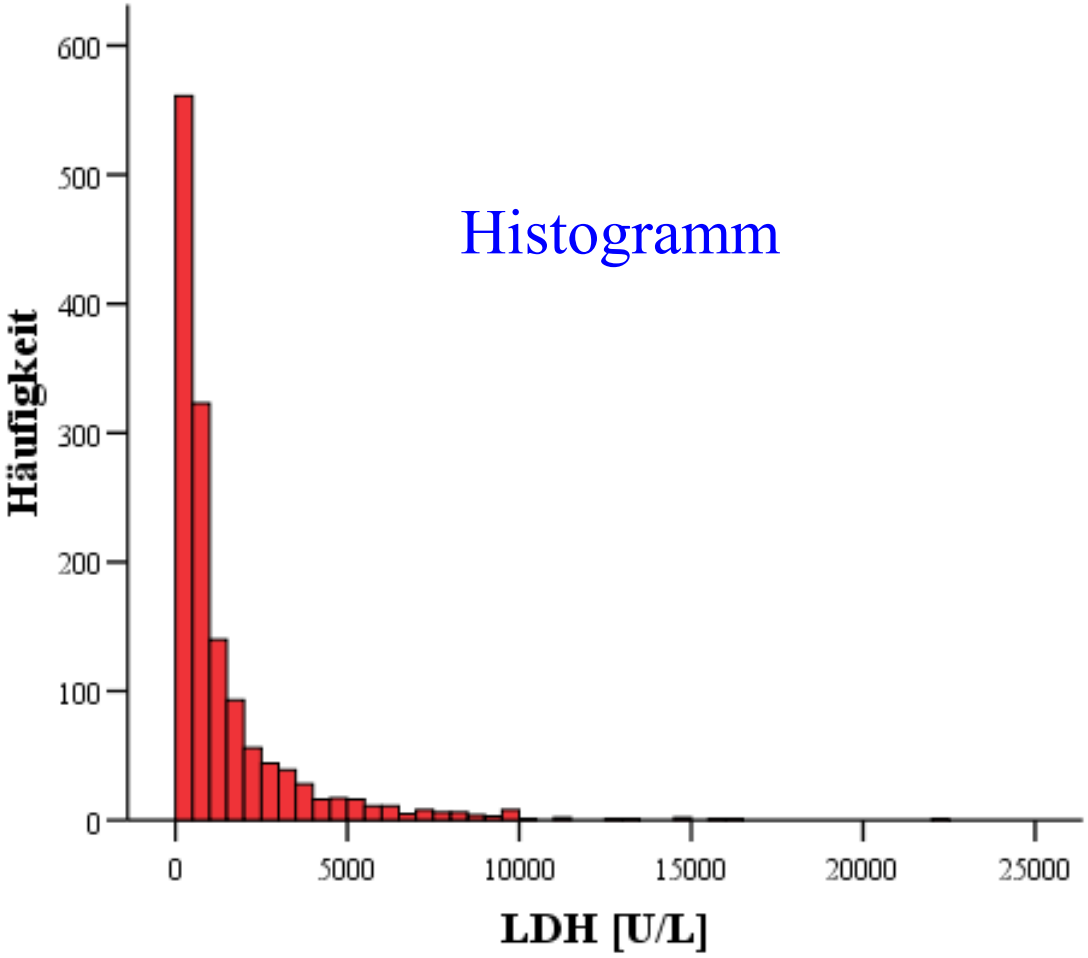
Zusammenhang zwischen LDH und Ferritin



Streudiagramm

Scatterplot

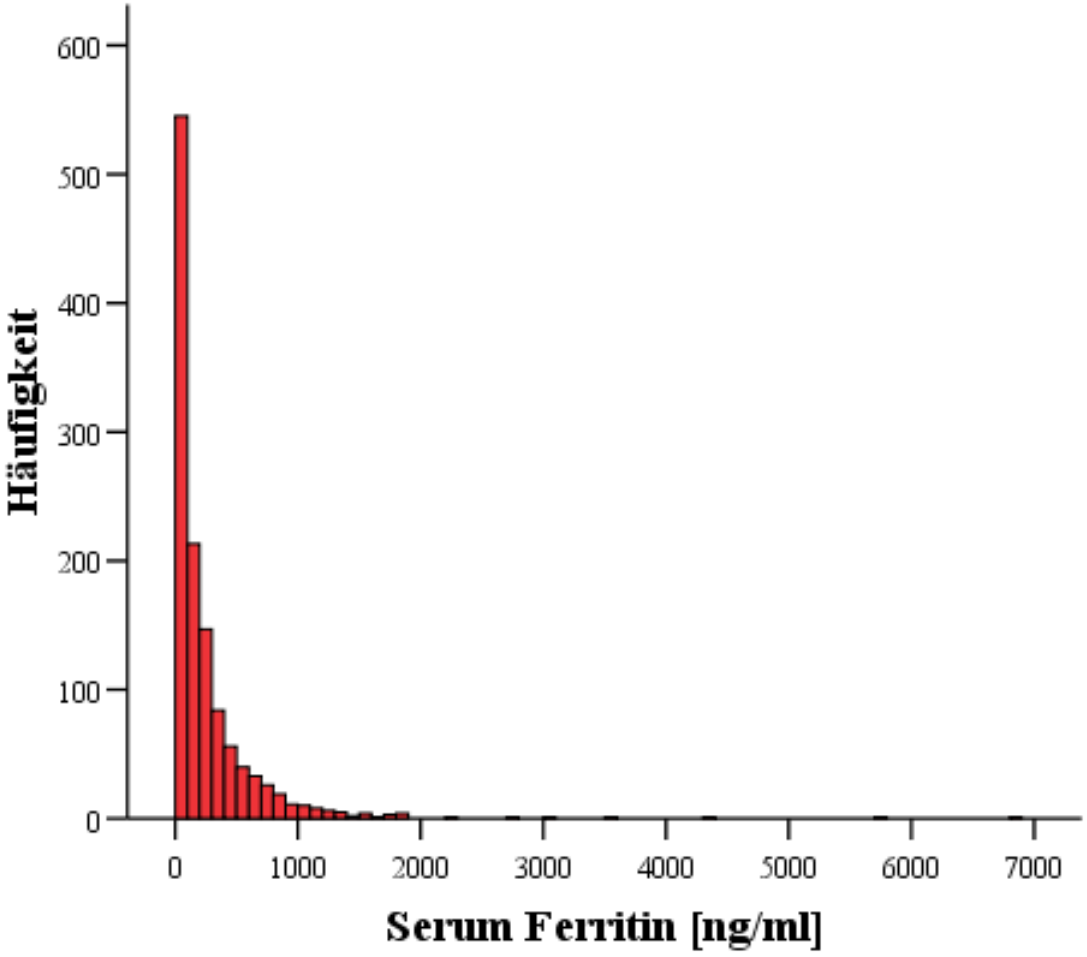
LDH



Histogramm

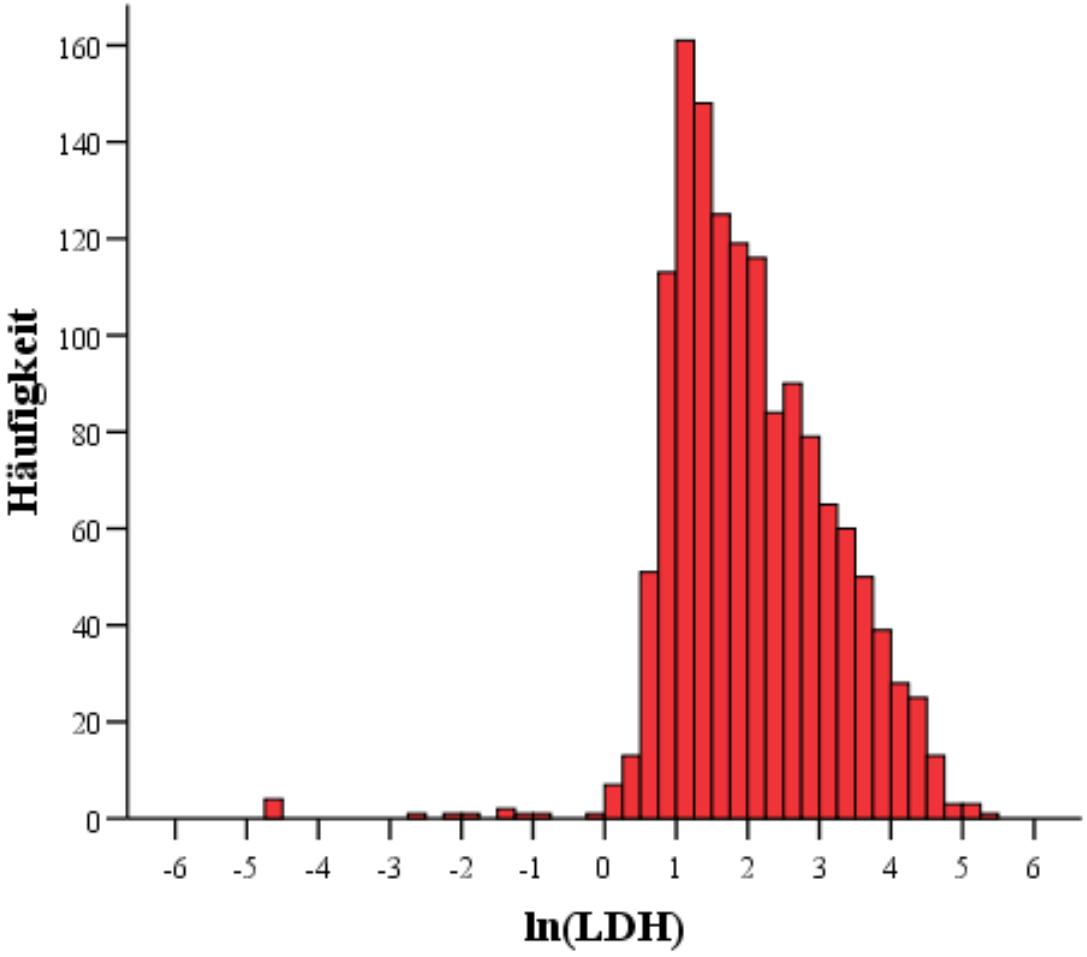
		LDH [U/L]
N	Gültig	1405
	Fehlend	1940
Mittelwert		1456
Median		669
Standardabweichung		2052
Schiefe		3
Minimum		0
Maximum		22000
Perzentile	25	344
	50	669
	75	1606

Serum Ferritin



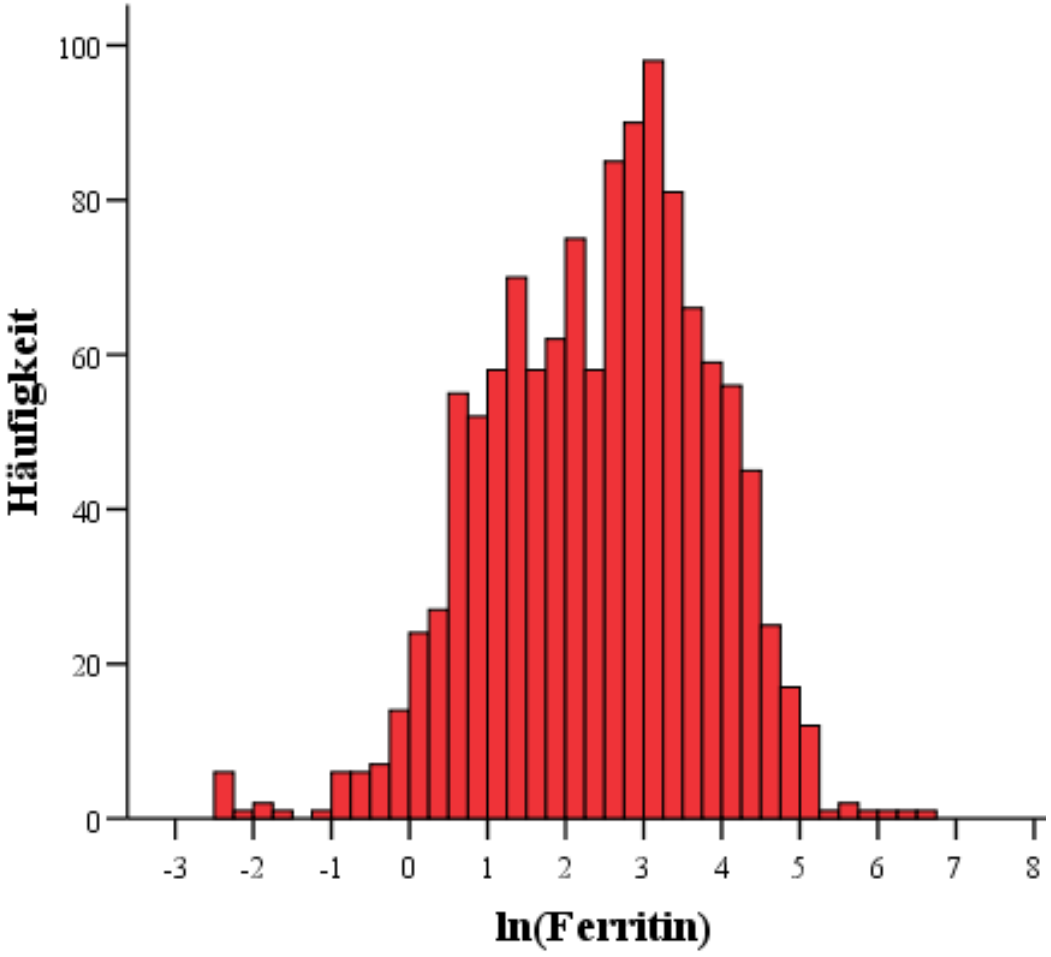
		Serum Ferritin [ng/ml]
N	Gültig	1224
	Fehlend	2121
Mittelwert		261
Median		133
Standardabweichung		432
Schiefe		7
Minimum		0
Maximum		6831
Perzentile	25	41
	50	133
	75	315

Logarithmus von LDH



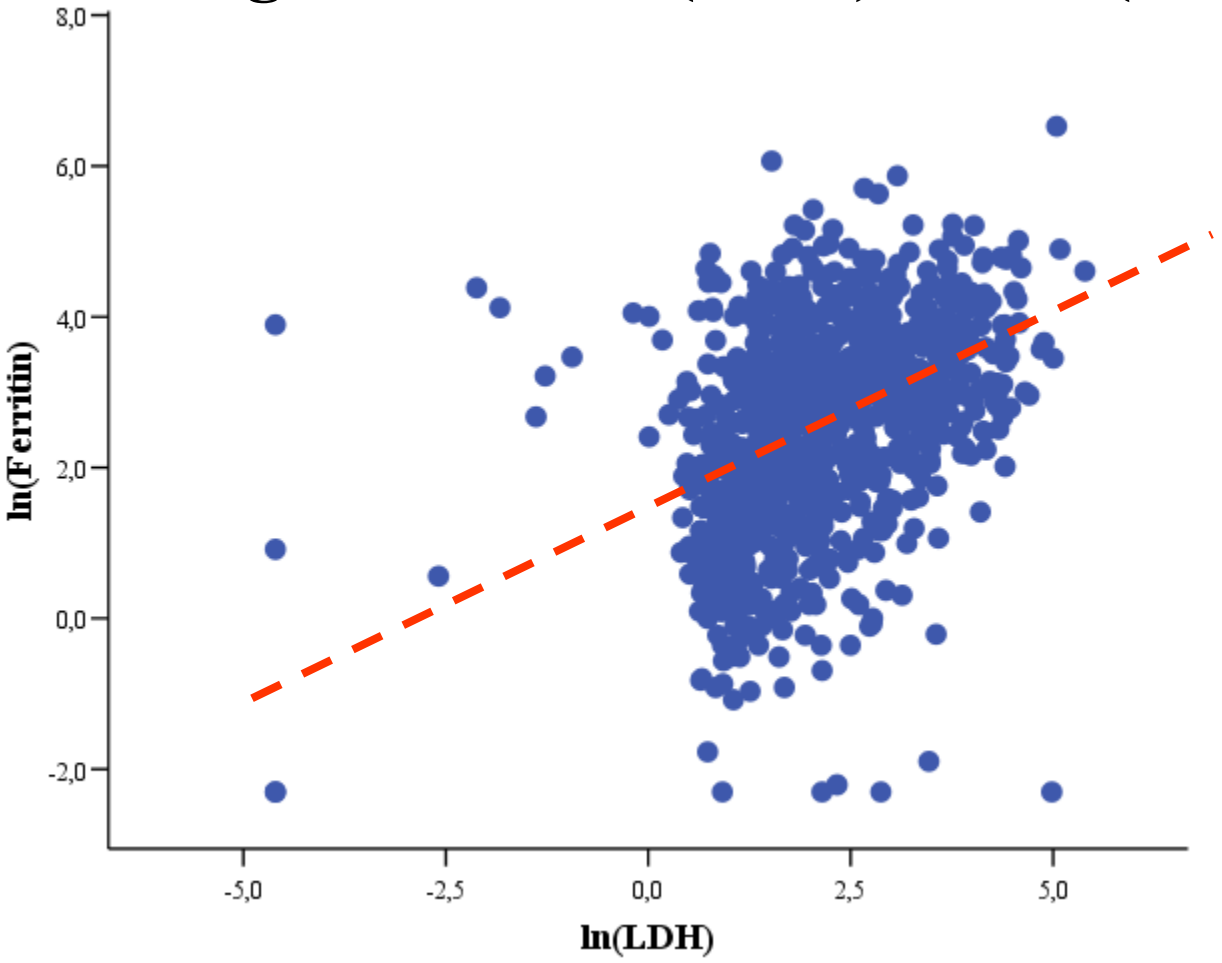
		ln(LDH)
N	Gültig	1405
	Fehlend	1940
Mittelwert		2,1
Median		1,9
Standardabweichung		1,1
Schiefe		-,3
Minimum		-4,6
Maximum		5,4
Perzentile	25	1,2
	50	1,9
	75	2,8

Logarithmus von Ferritin

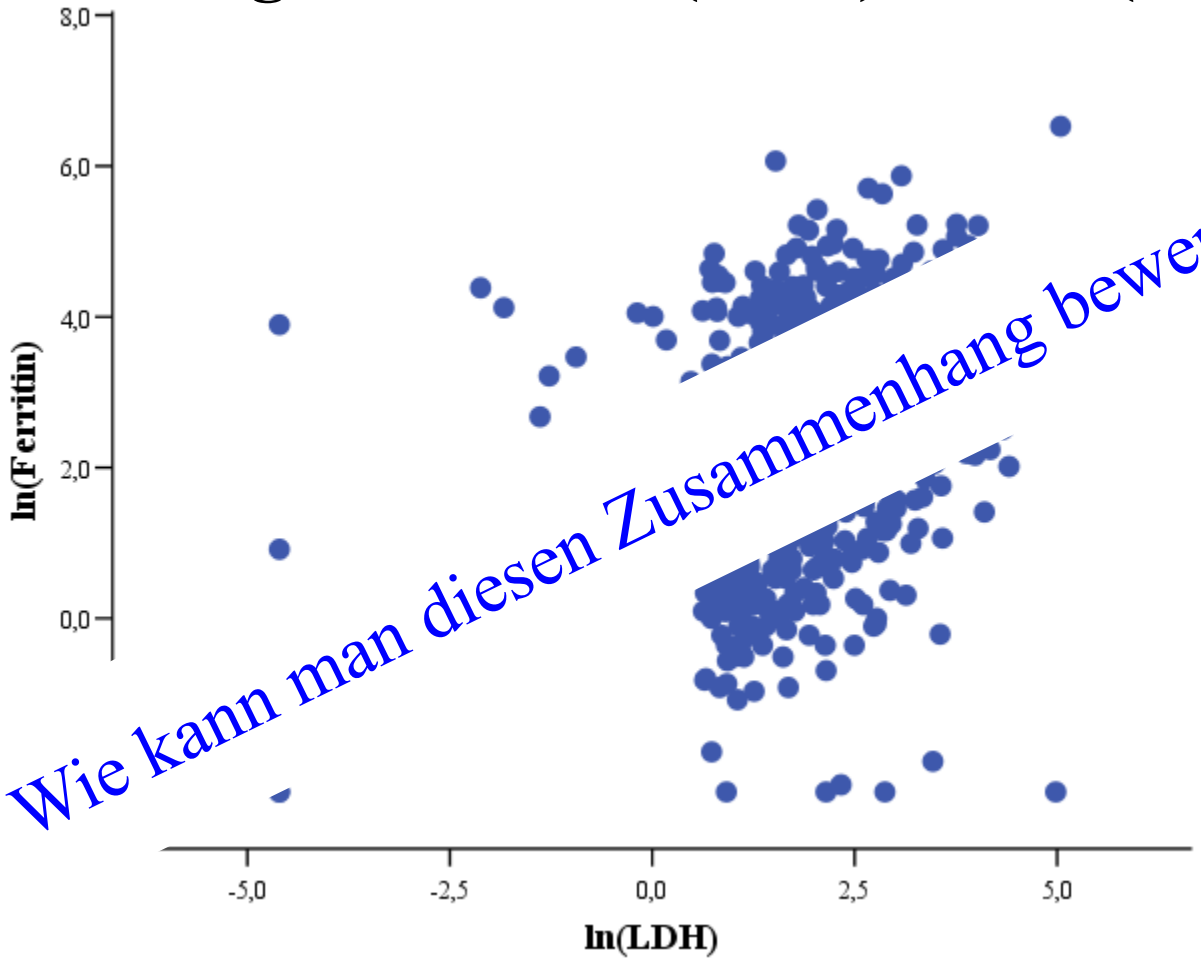


		ln(Ferritin)
N	Gültig	1224
	Fehlend	2121
Mittelwert		2,4
Median		2,6
Standardabweichung		1,4
Schiefe		-,3
Minimum		-2,3
Maximum		6,5
Perzentile	25	1,4
	50	2,6
	75	3,5

Zusammenhang zwischen ln(LDH) und ln(Ferritin)

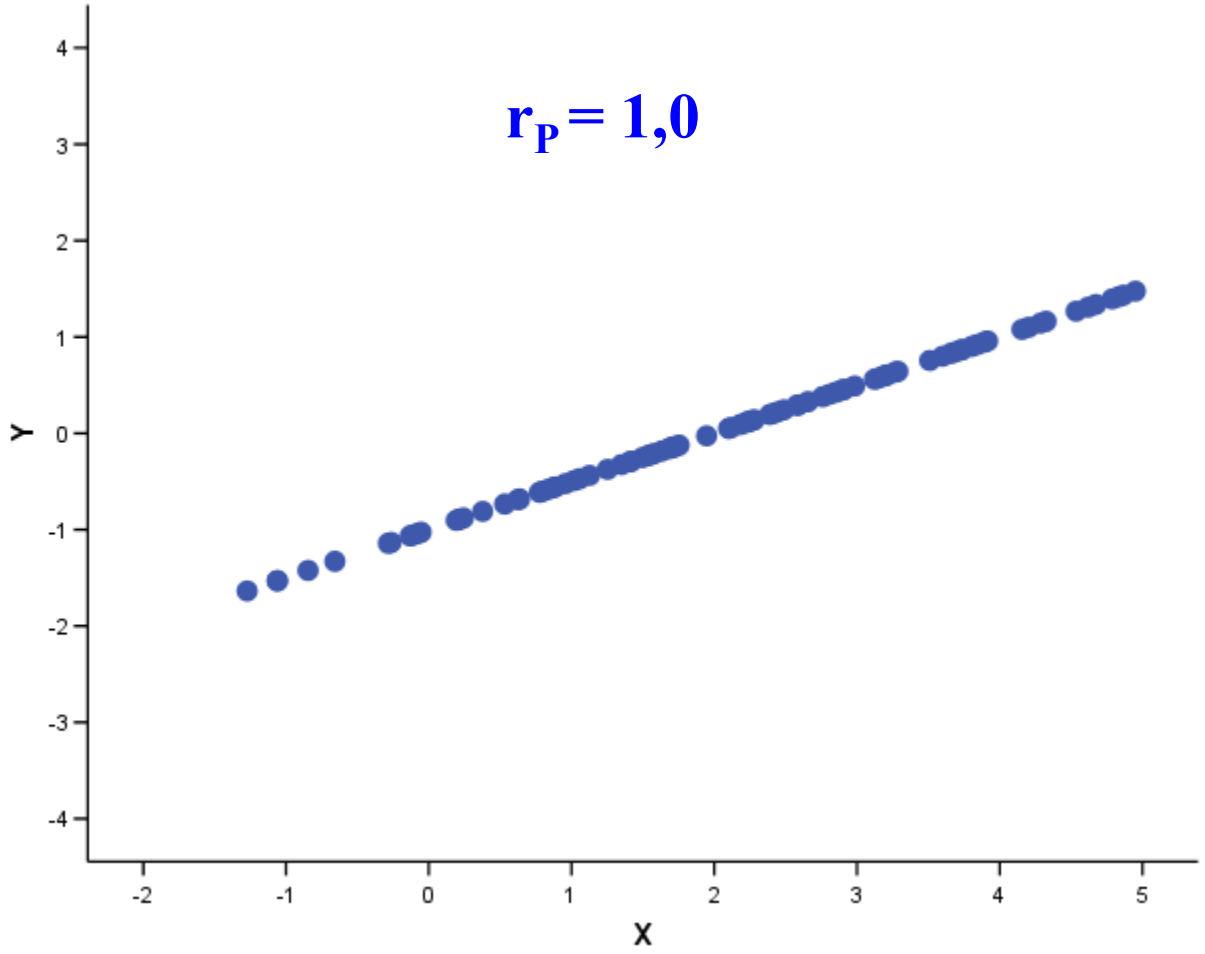


Zusammenhang zwischen ln(LDH) und ln(Ferritin)



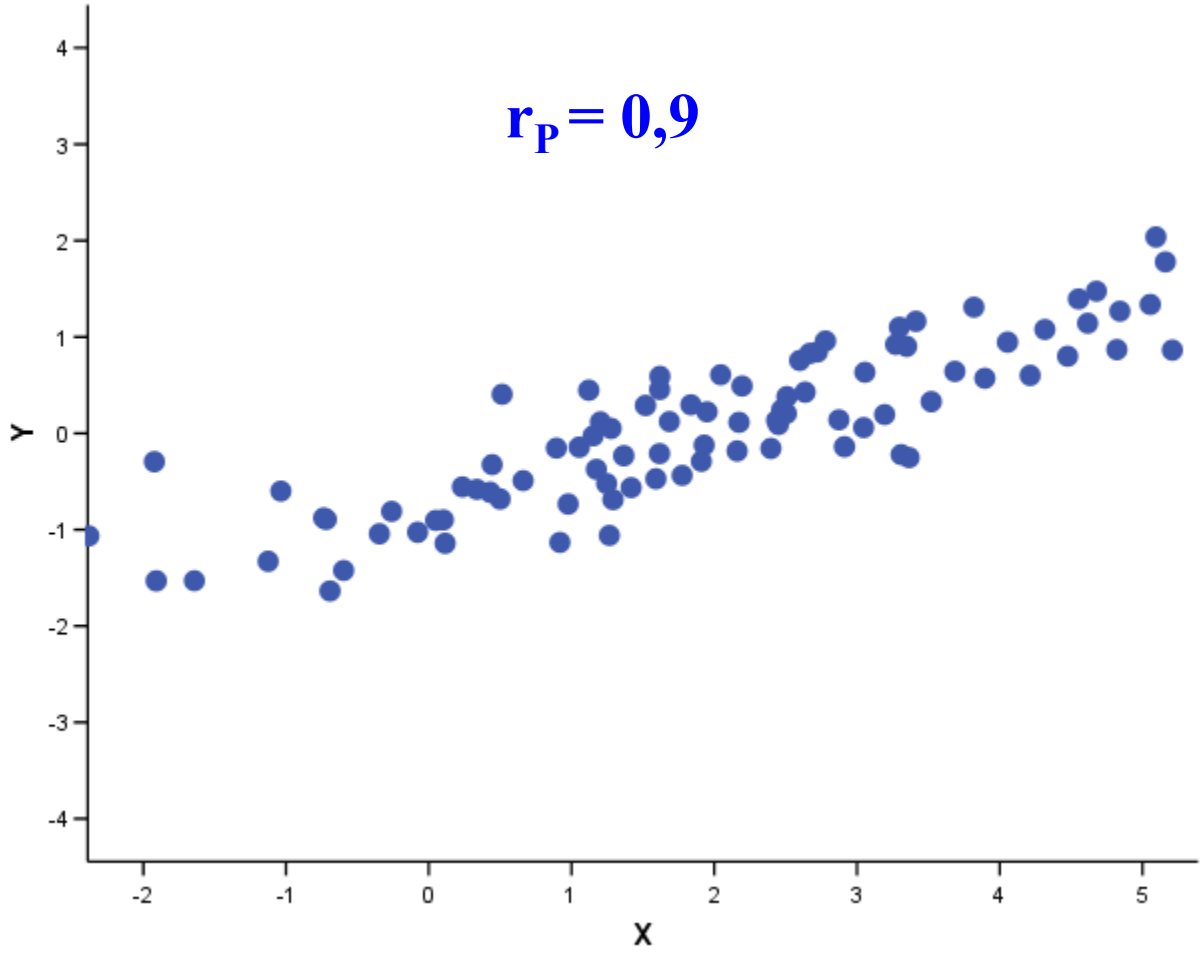
Korrelationskoeffizient und **Streudiagramm**

$r_P = 1{,}0$



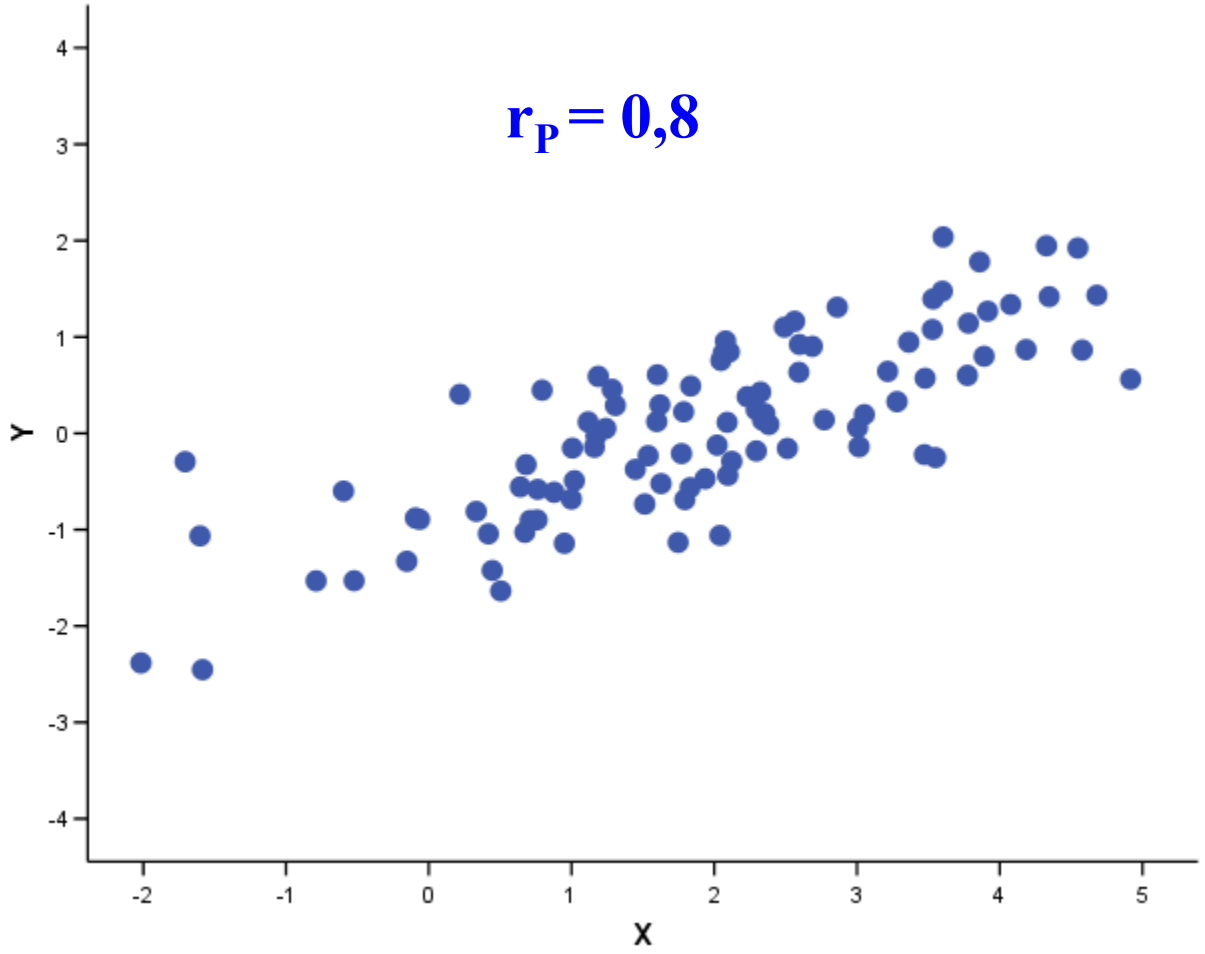
Korrelationskoeffizient und **Streudiagramm**

$r_P = 0{,}9$

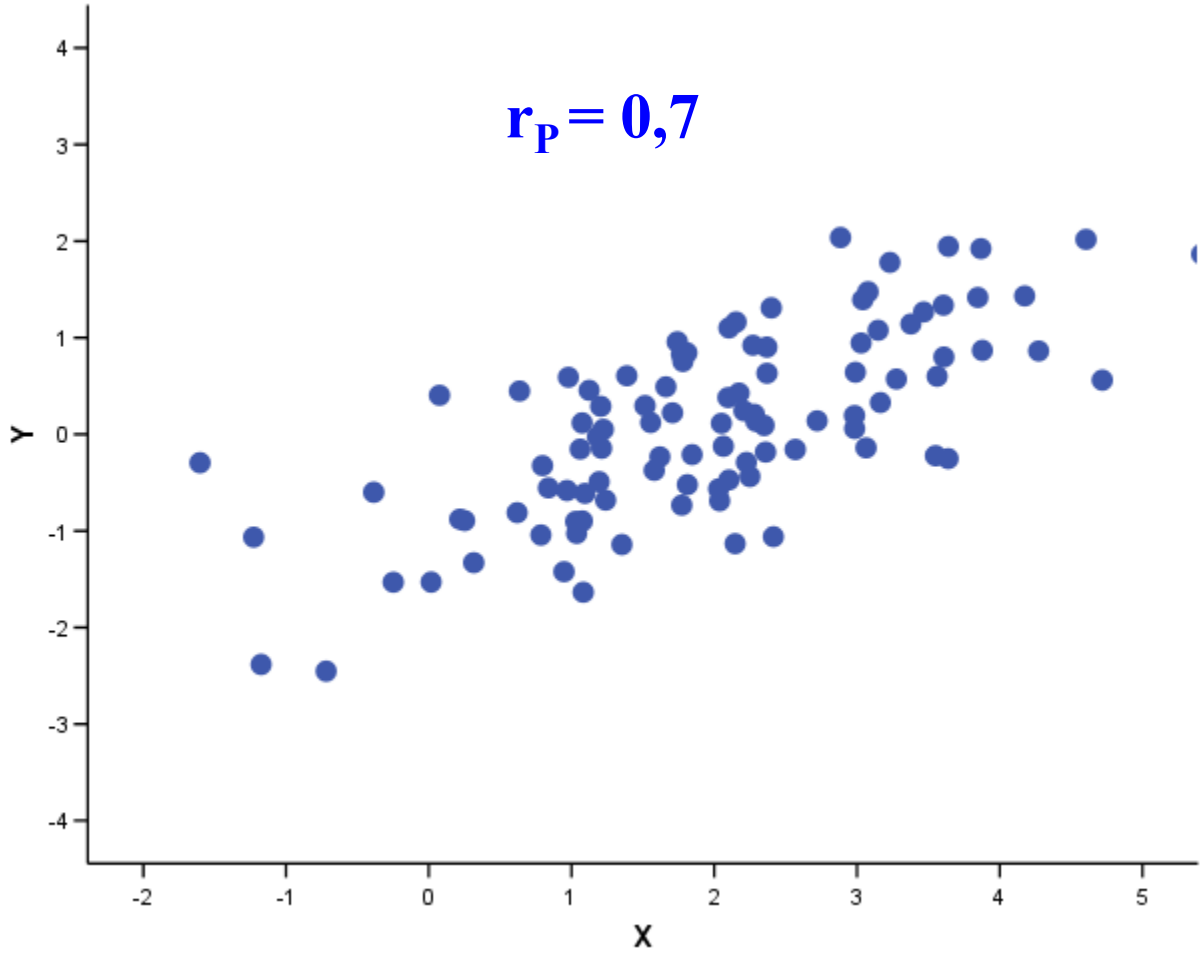


Korrelationskoeffizient und **Streudiagramm**

$r_P = 0{,}8$

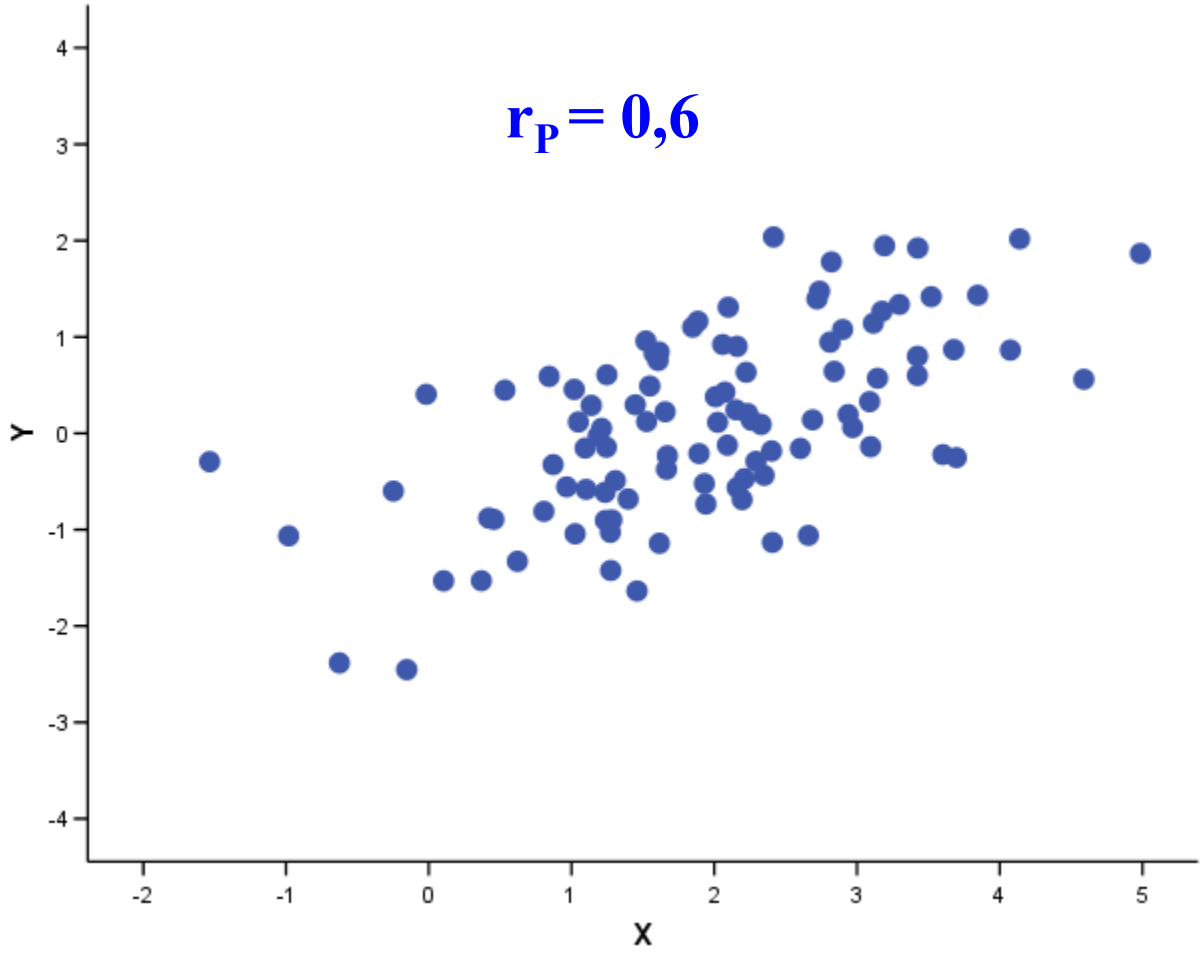


Korrelationskoeffizient und **Streudiagramm**

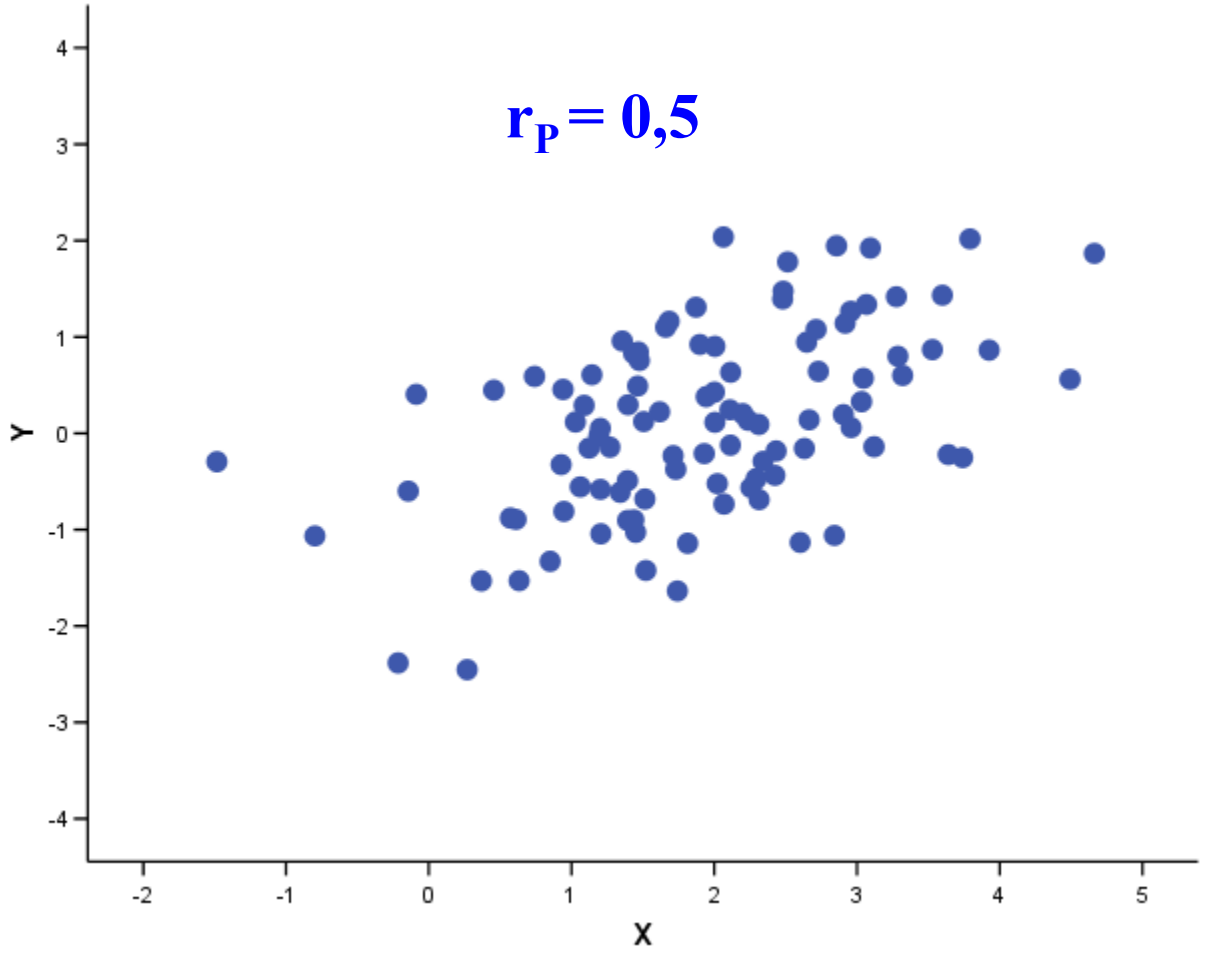


Korrelationskoeffizient und **Streudiagramm**

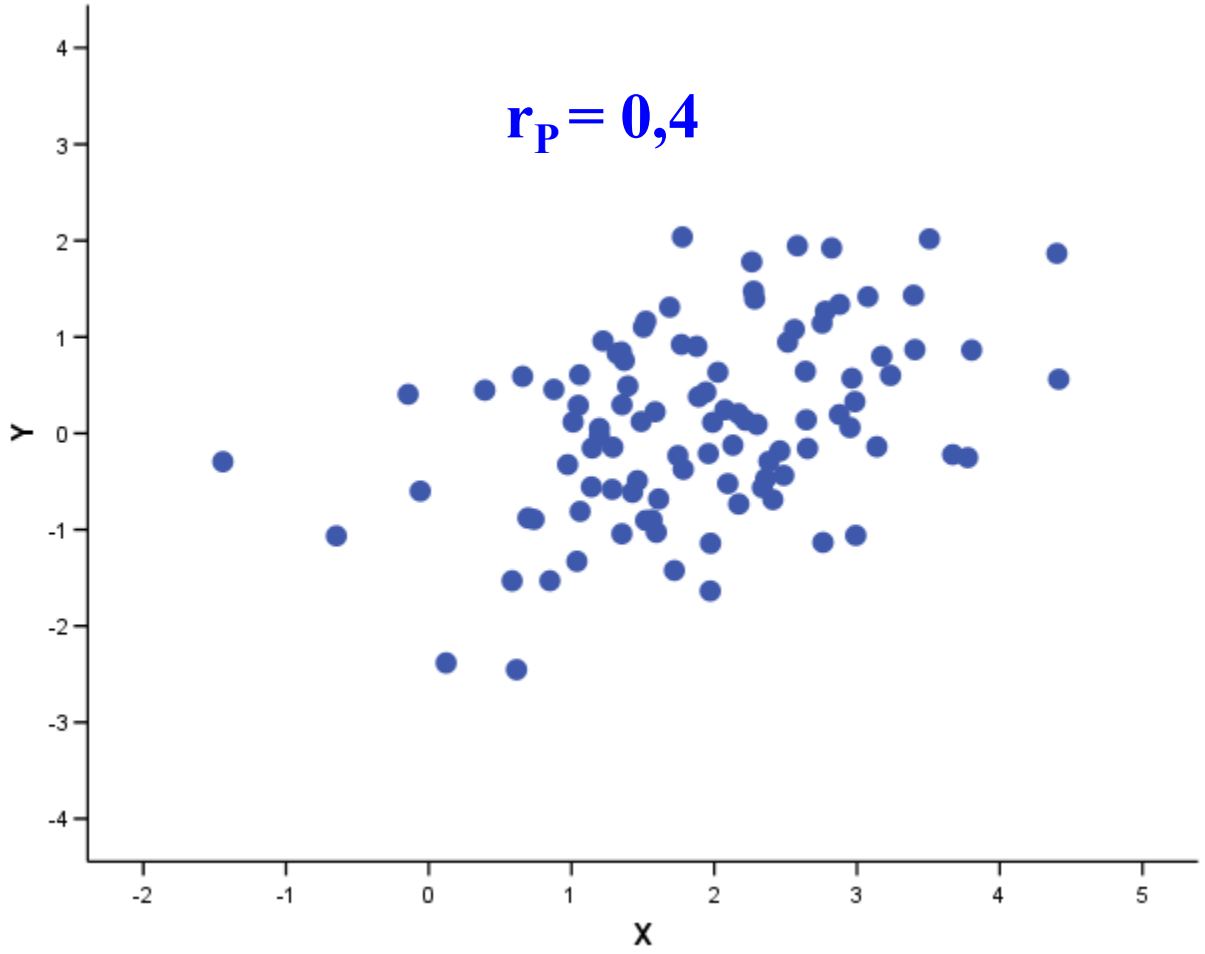
$r_P = 0{,}6$



Korrelationskoeffizient und **Streudiagramm**

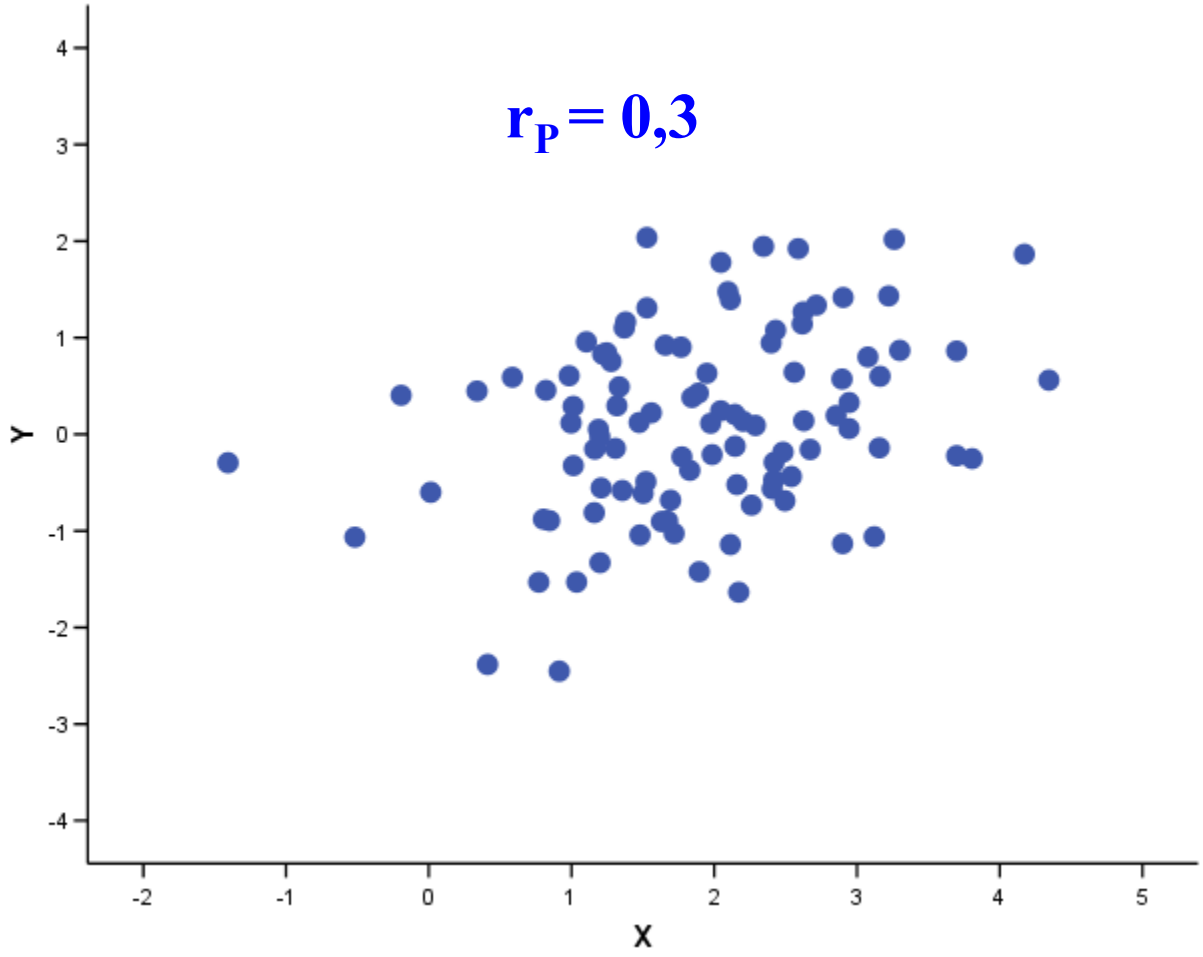


Korrelationskoeffizient und **Streudiagramm**

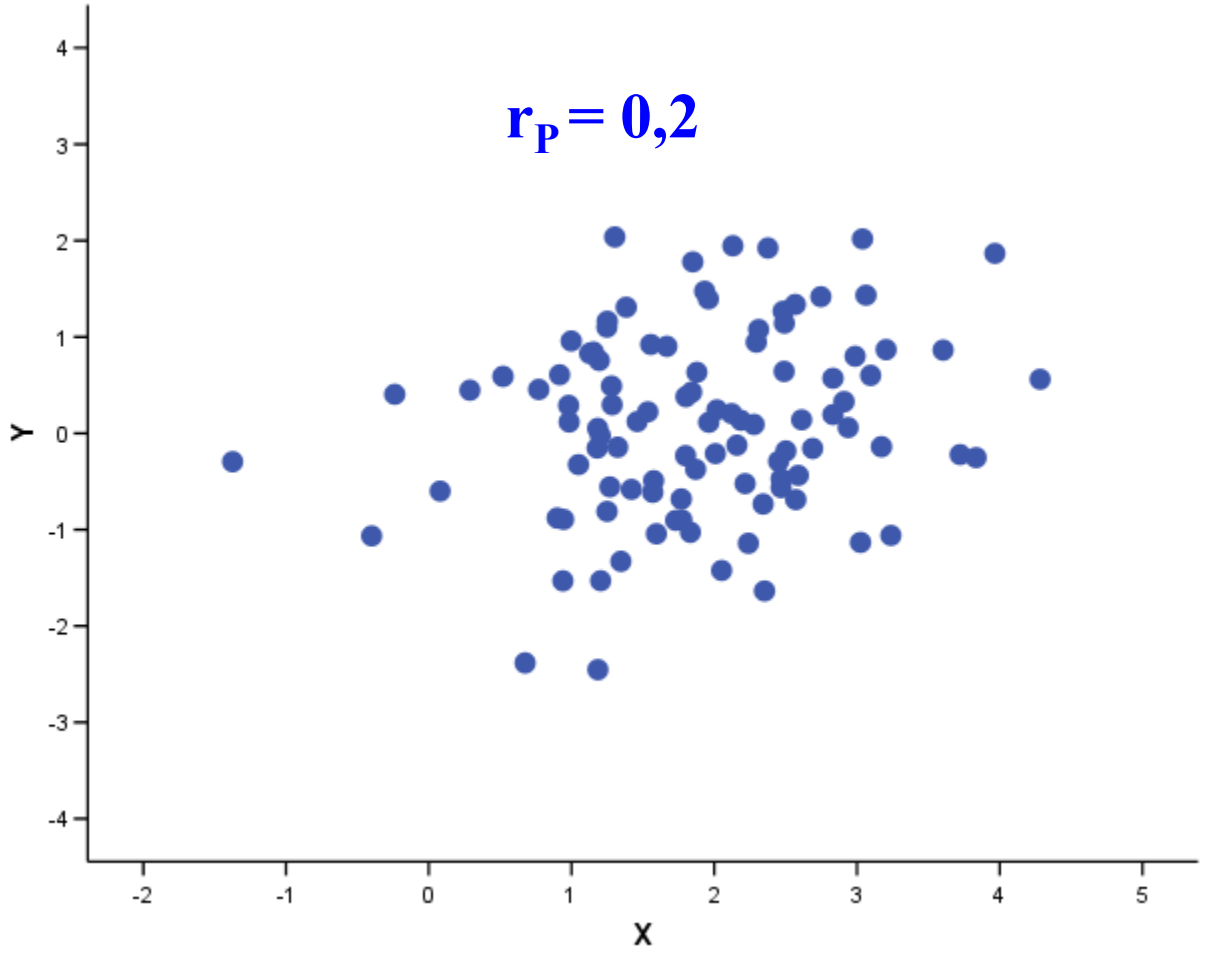


Korrelationskoeffizient und **Streudiagramm**

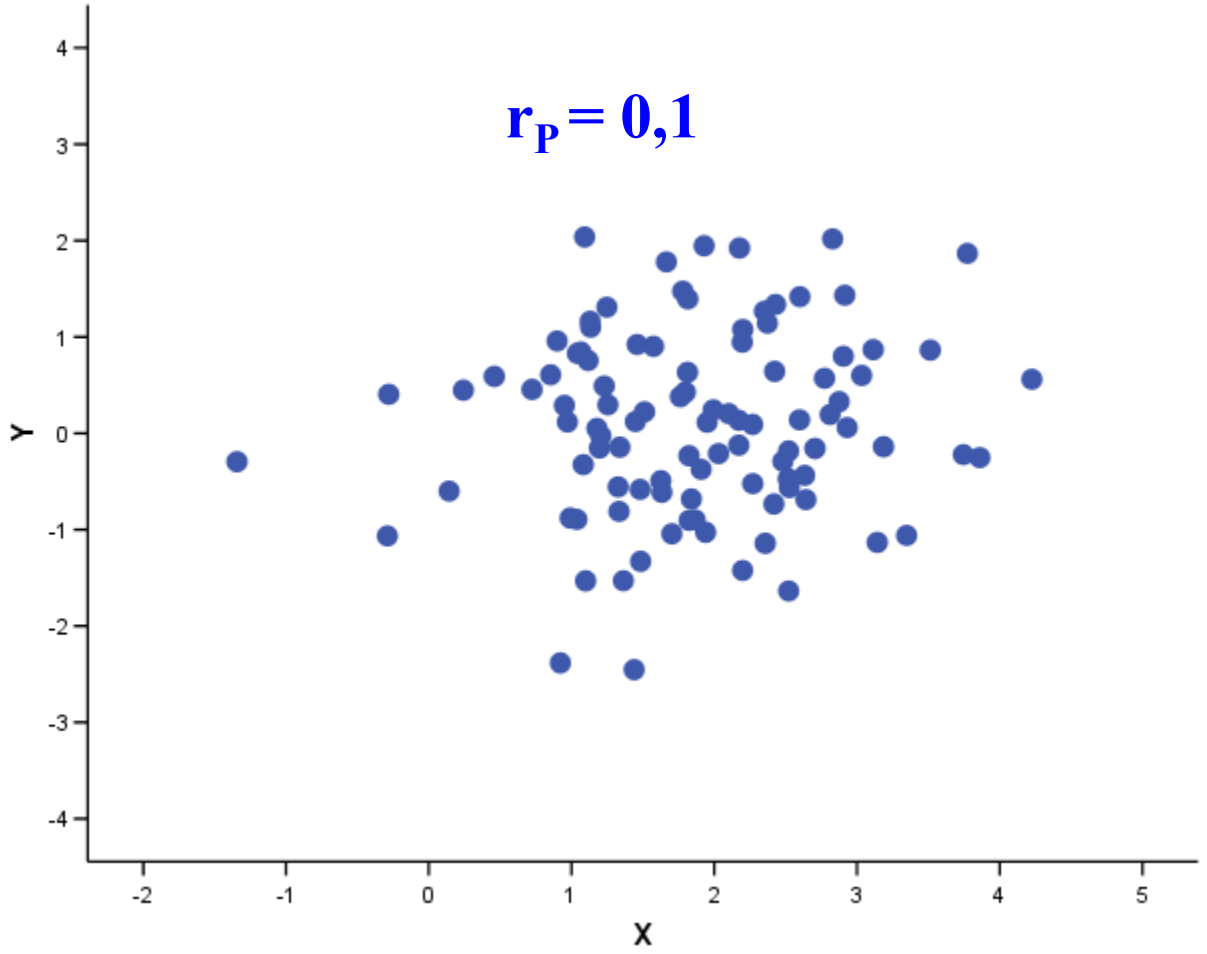
$r_P = 0{,}3$



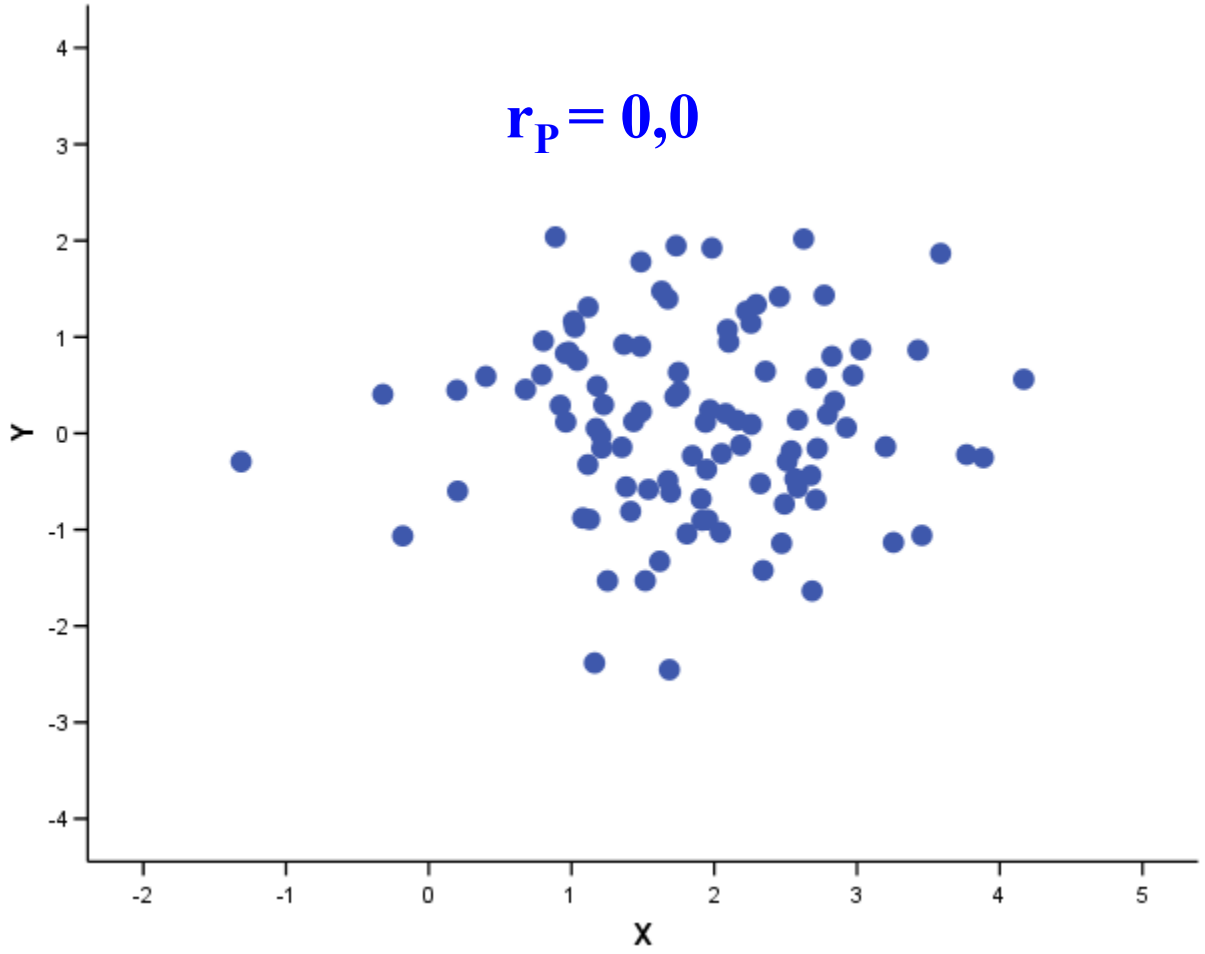
Korrelationskoeffizient und **Streudiagramm**



Korrelationskoeffizient und **Streudiagramm**

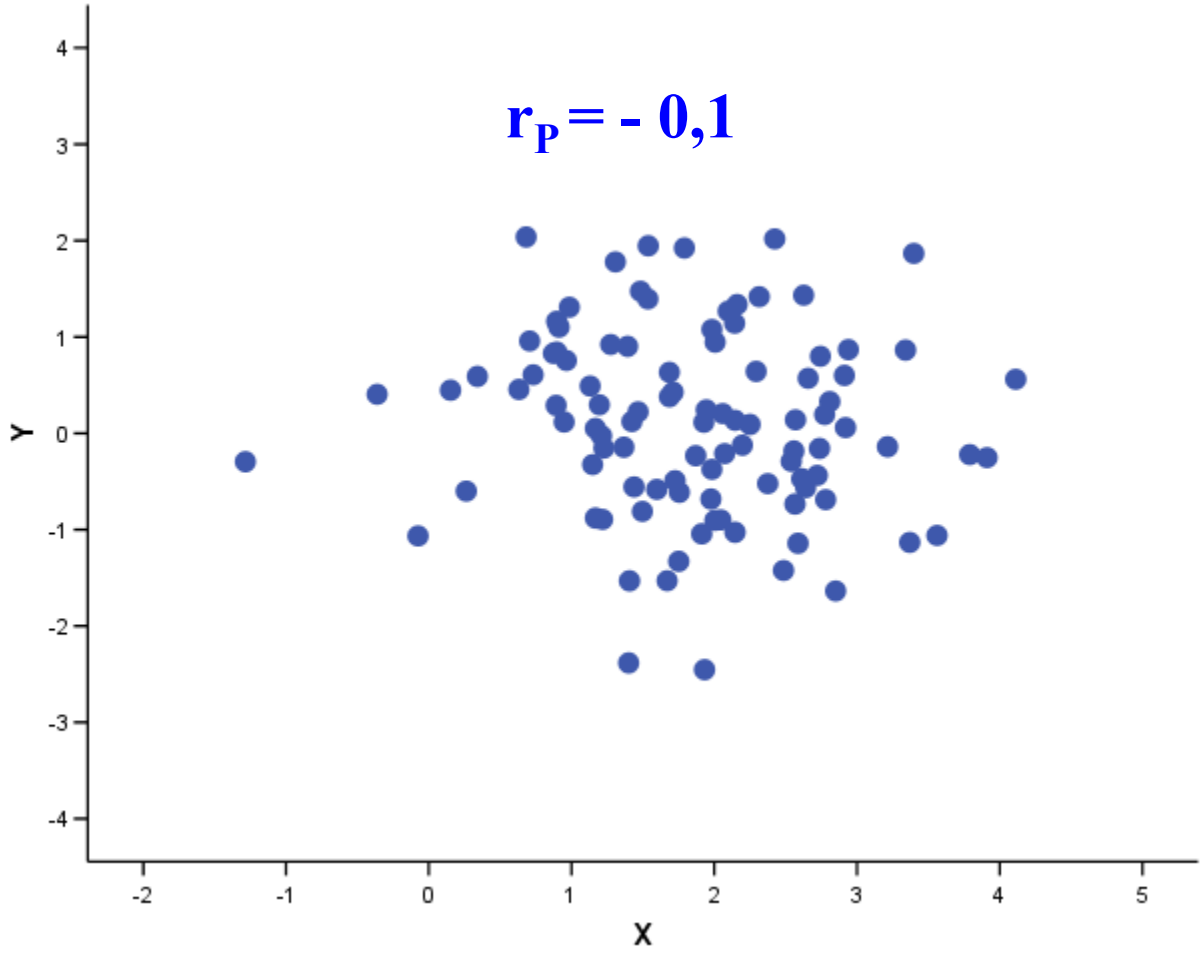


Korrelationskoeffizient und **Streudiagramm**

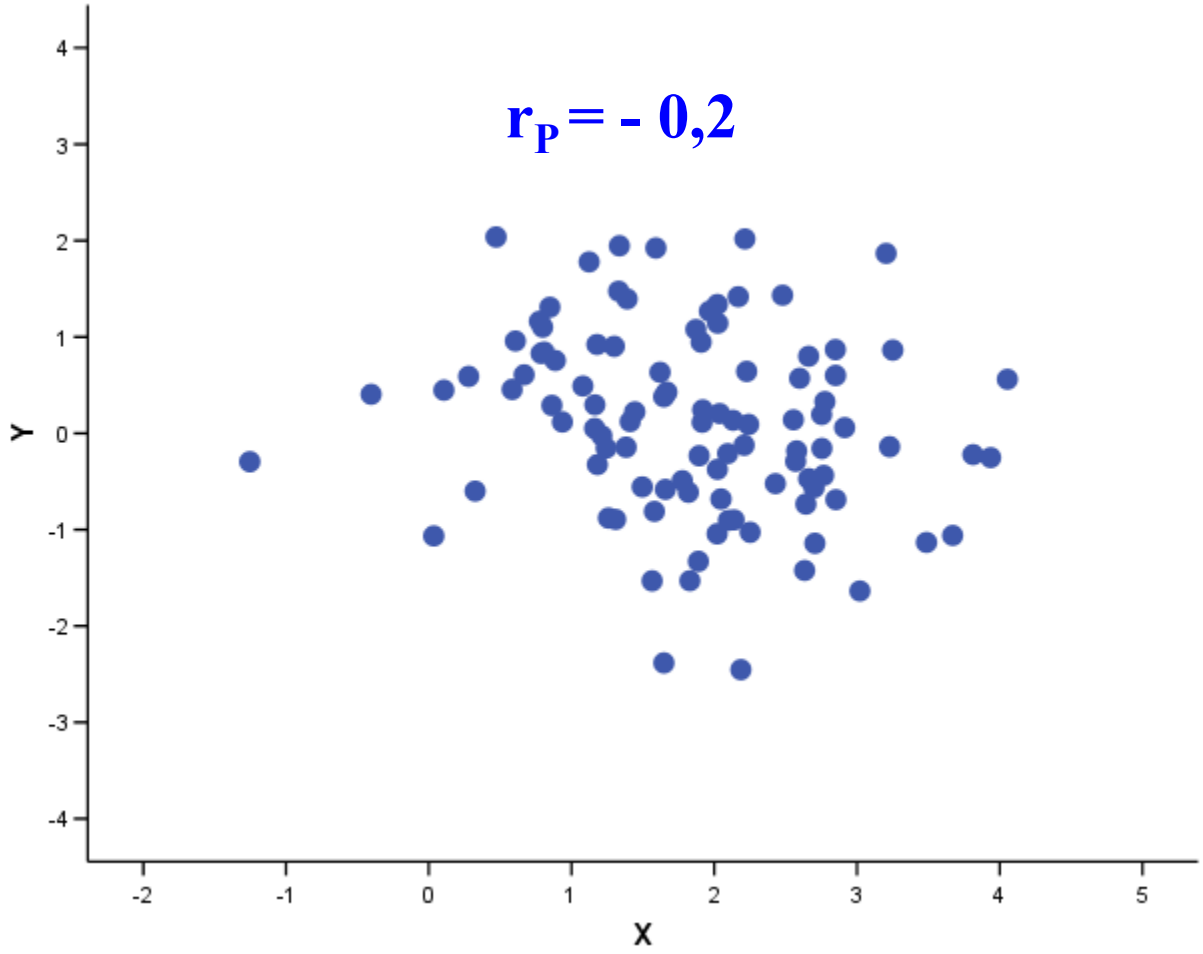


Korrelationskoeffizient und **Streudiagramm**

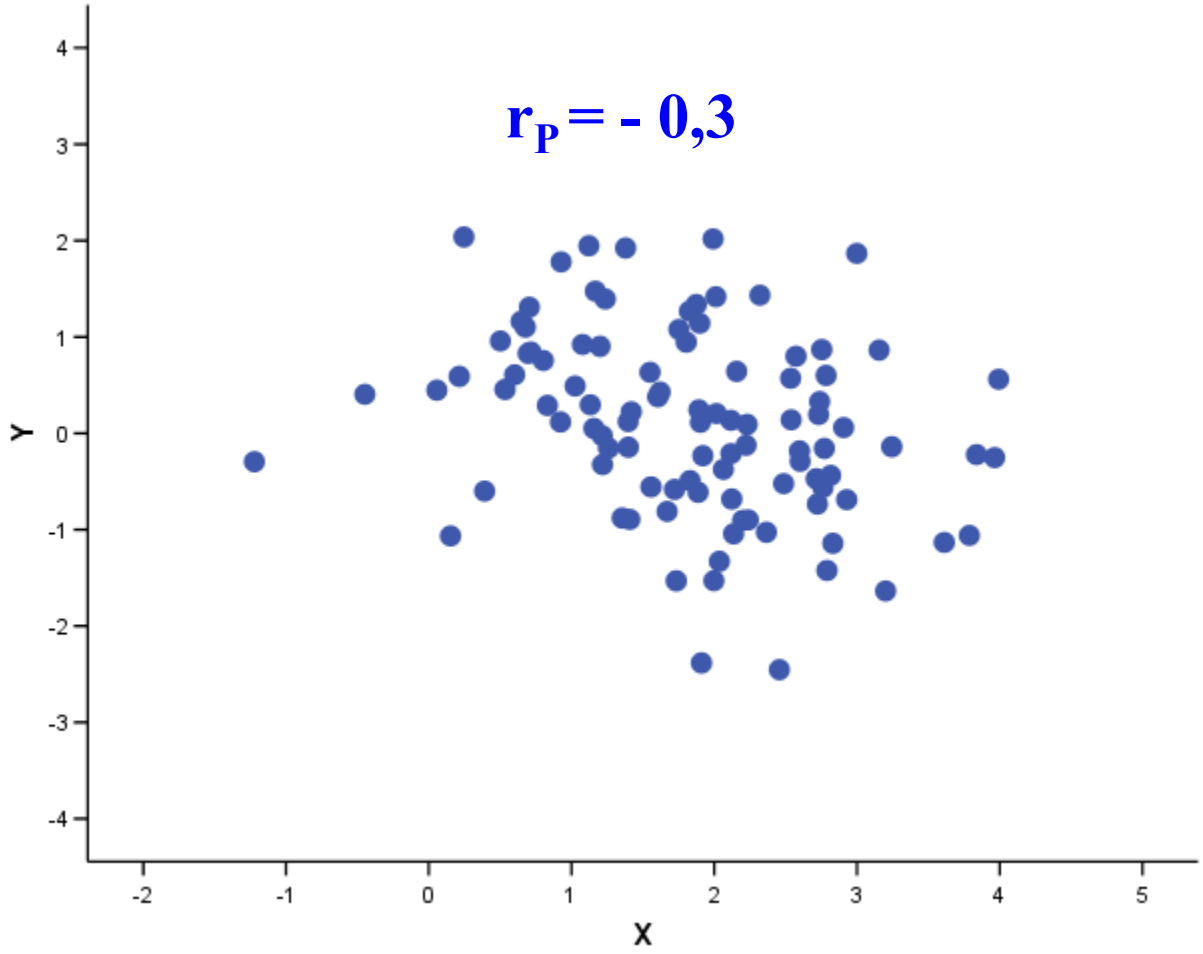
$r_P = -0{,}1$



Korrelationskoeffizient und **Streudiagramm**

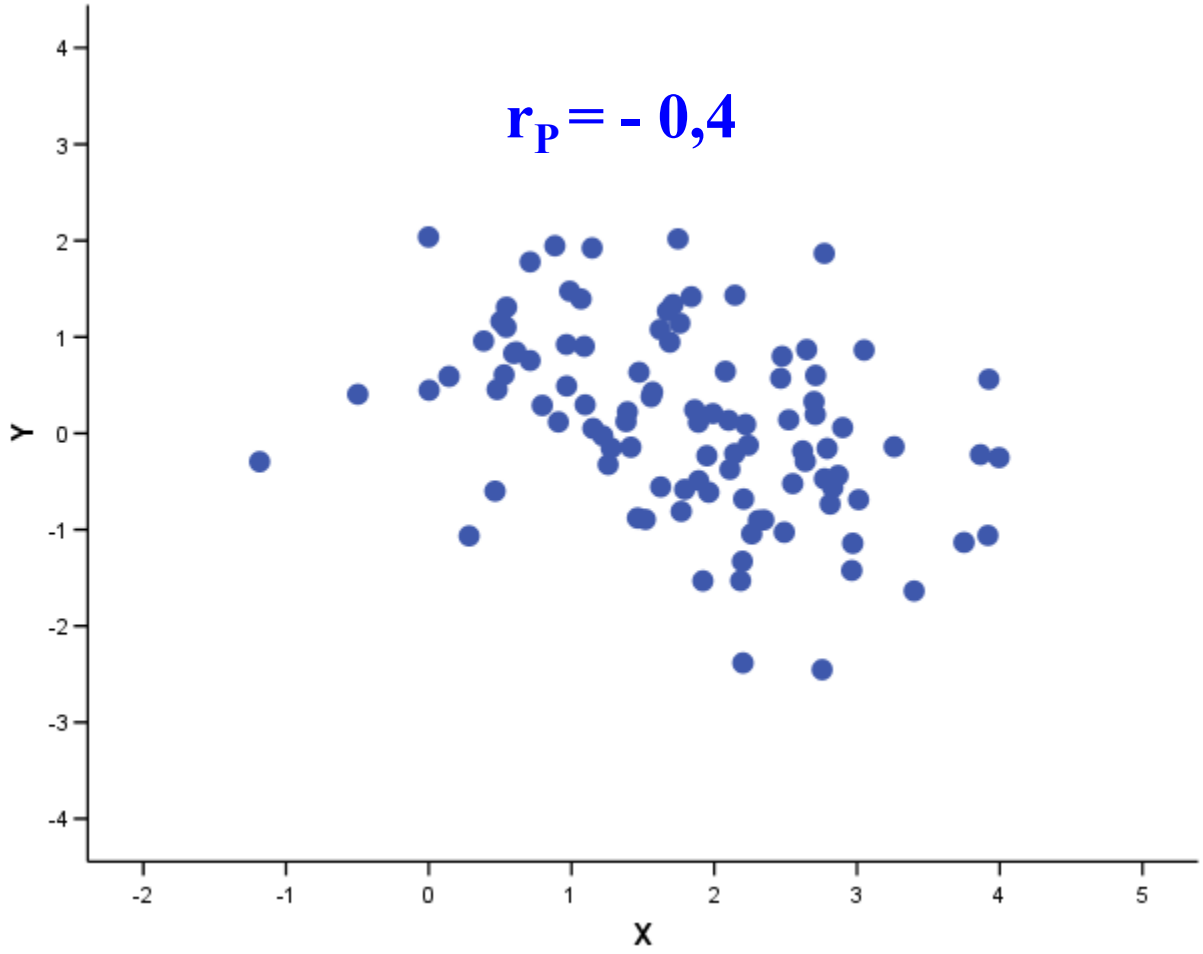


Korrelationskoeffizient und **Streudiagramm**

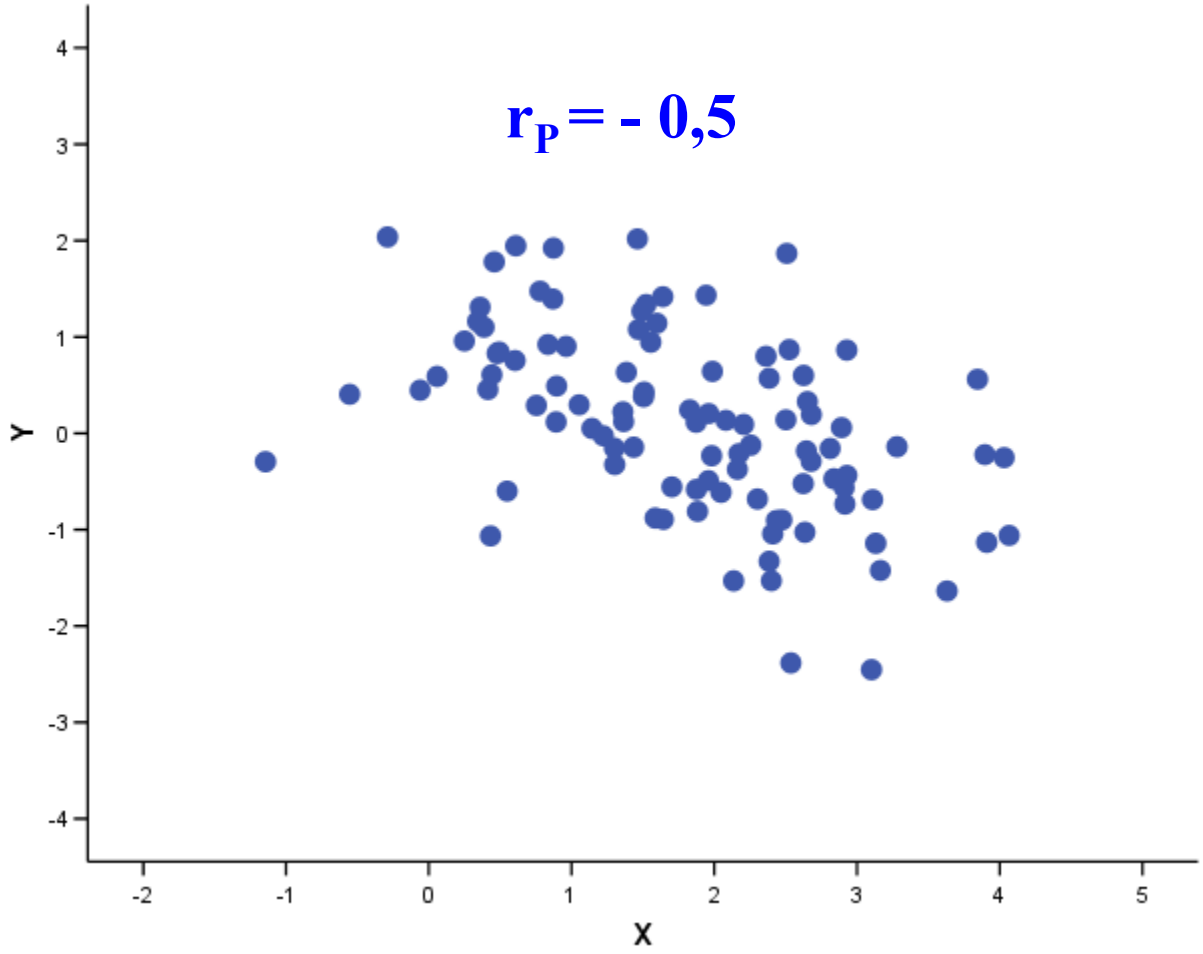


Korrelationskoeffizient und **Streudiagramm**

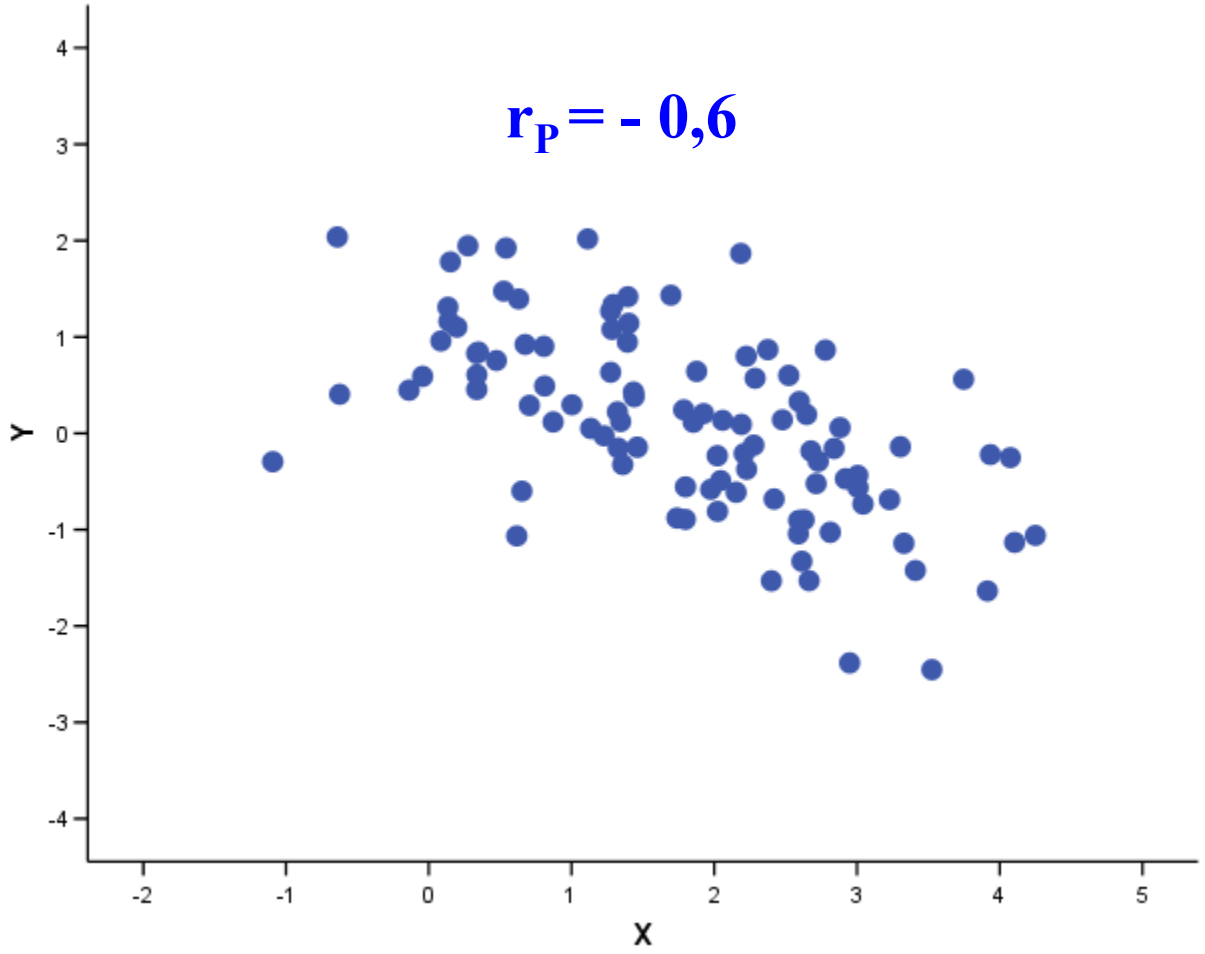
$r_P = -0{,}4$



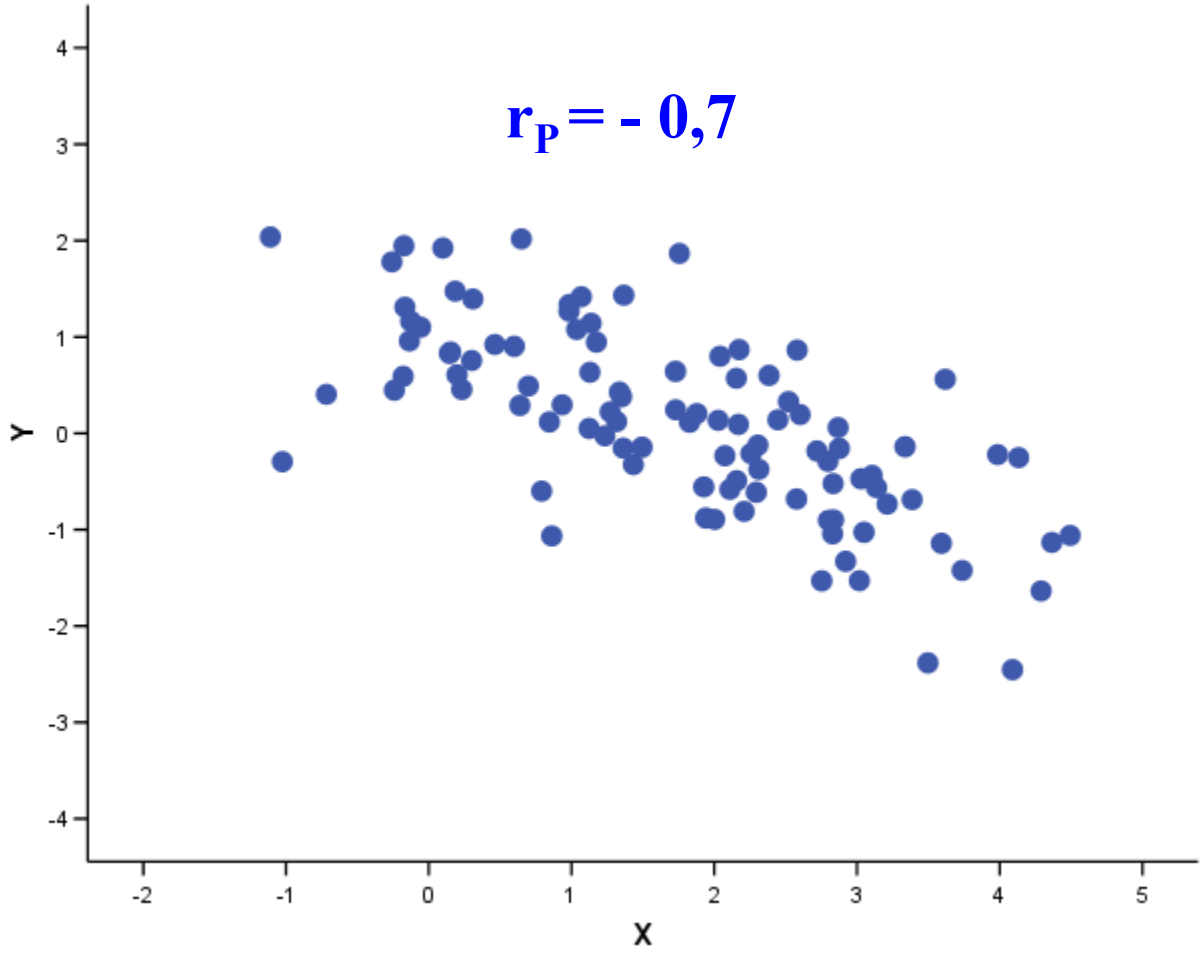
Korrelationskoeffizient und **Streudiagramm**



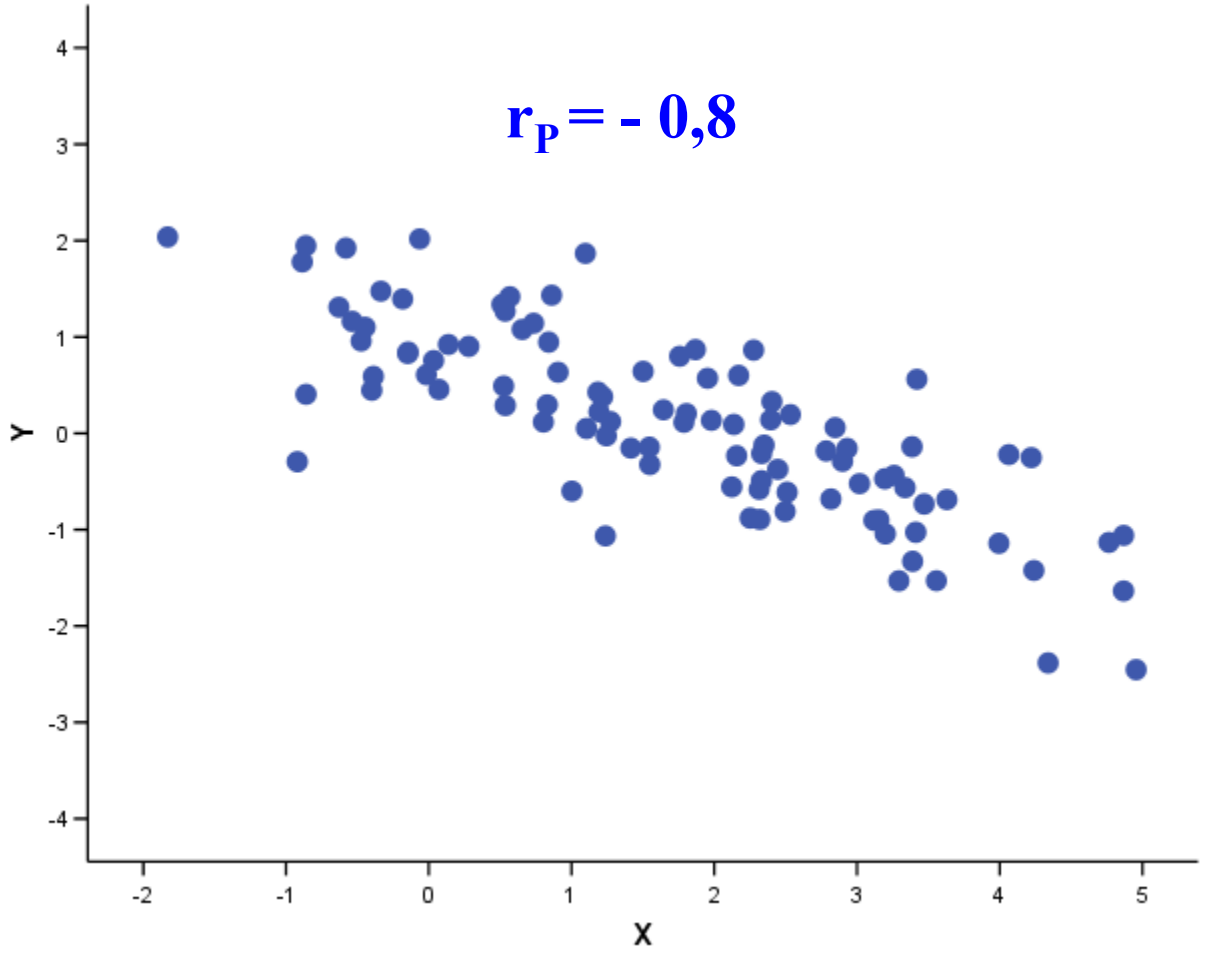
Korrelationskoeffizient und **Streudiagramm**



Korrelationskoeffizient und **Streudiagramm**

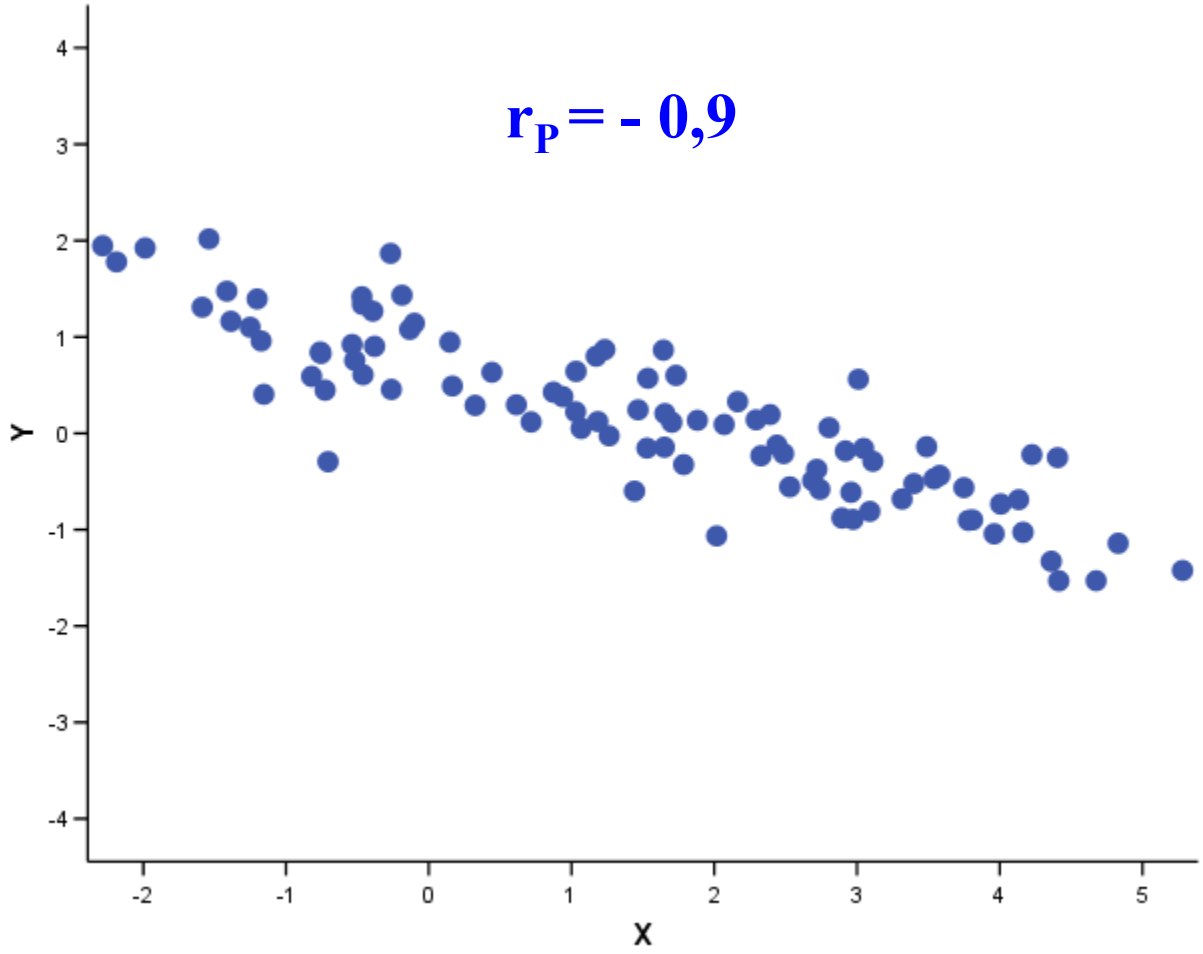


Korrelationskoeffizient und **Streudiagramm**



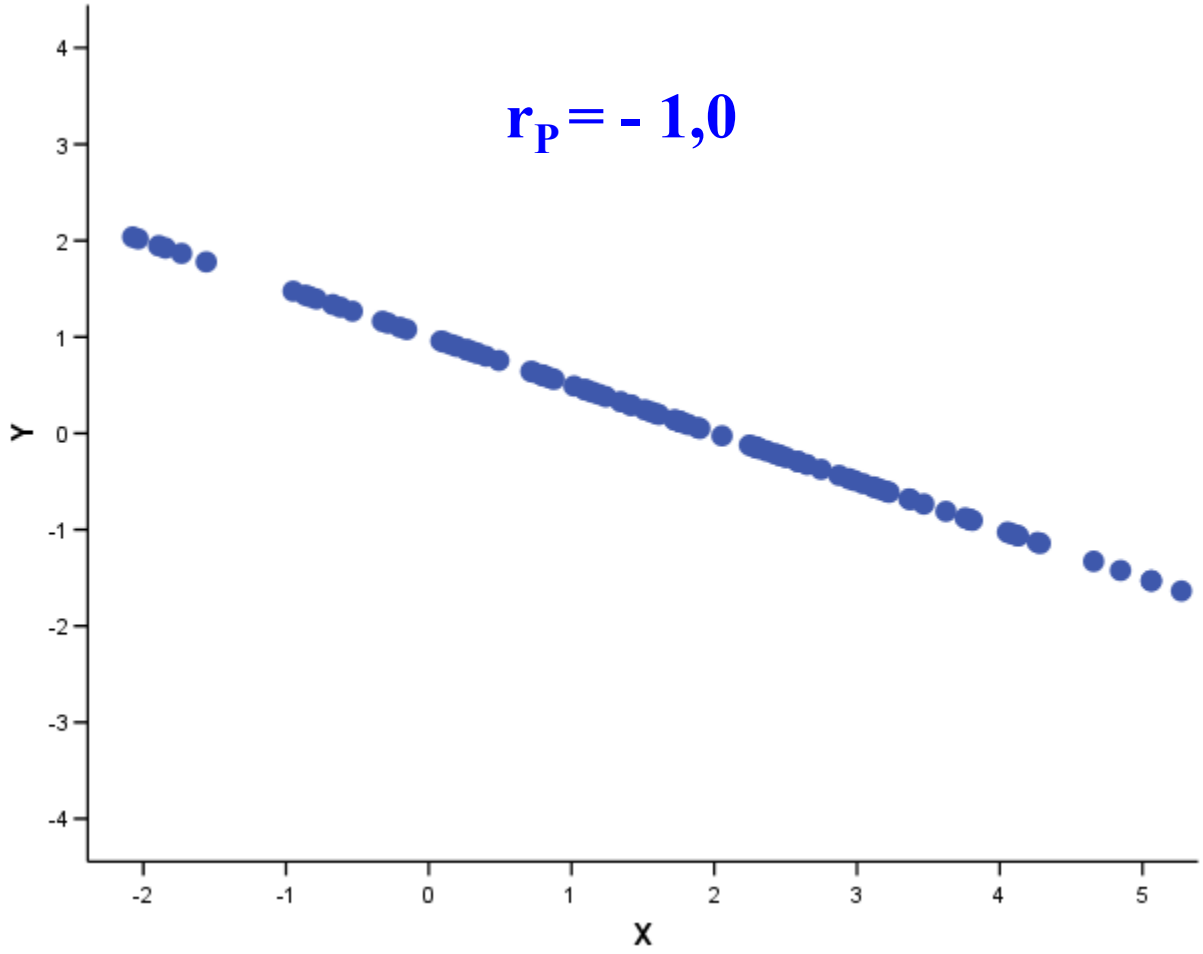
Korrelationskoeffizient und **Streudiagramm**

$r_P = -0{,}9$



Korrelationskoeffizient und **Streudiagramm**

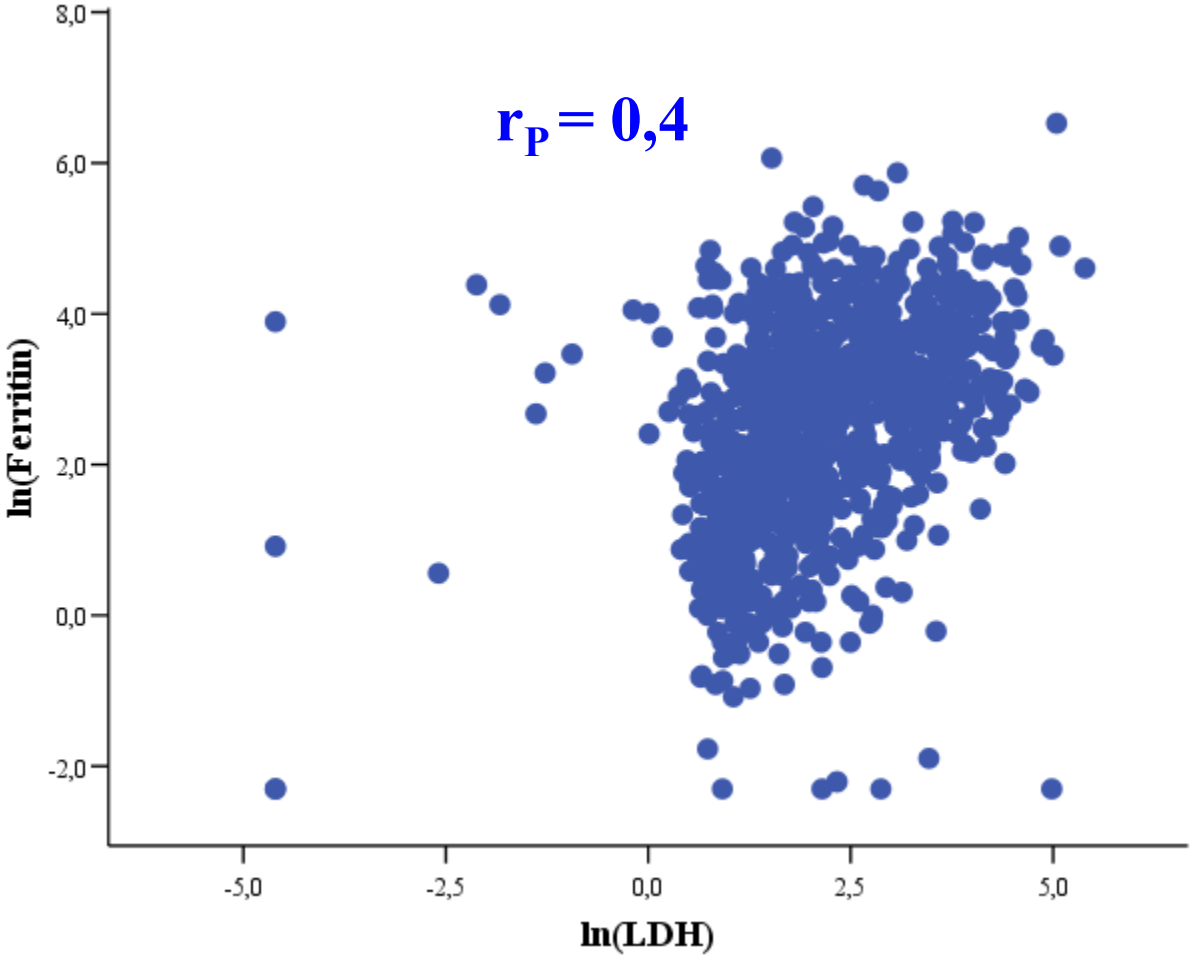
$r_P = -1{,}0$



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(LDH) und ln(Ferritin)

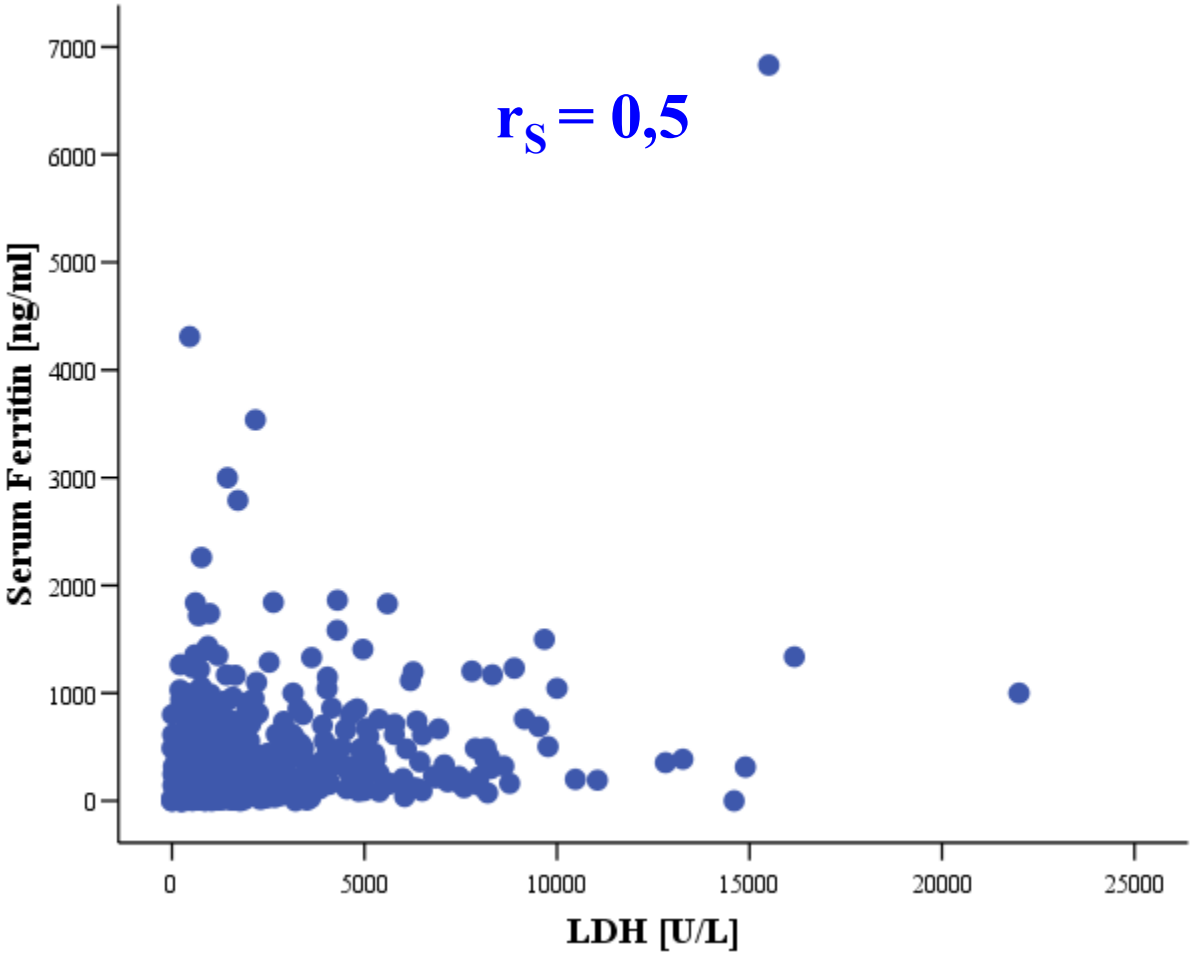


		kategorial		stetig	
		binär	**> 2 Kategorien**	**normalverteilt**	**nicht normalverteilt**
kategorial	**binär**			2 Boxplots / Histogramme 2*stetige Maßzahlen mit Mittelwert, Std.	2 Boxplots / Histogramme 2*stetige Maßzahlen ohne Mittelwert, Std.
	> 2 Kategorien				
stetig	**normal-verteilt**			Streudiagramm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

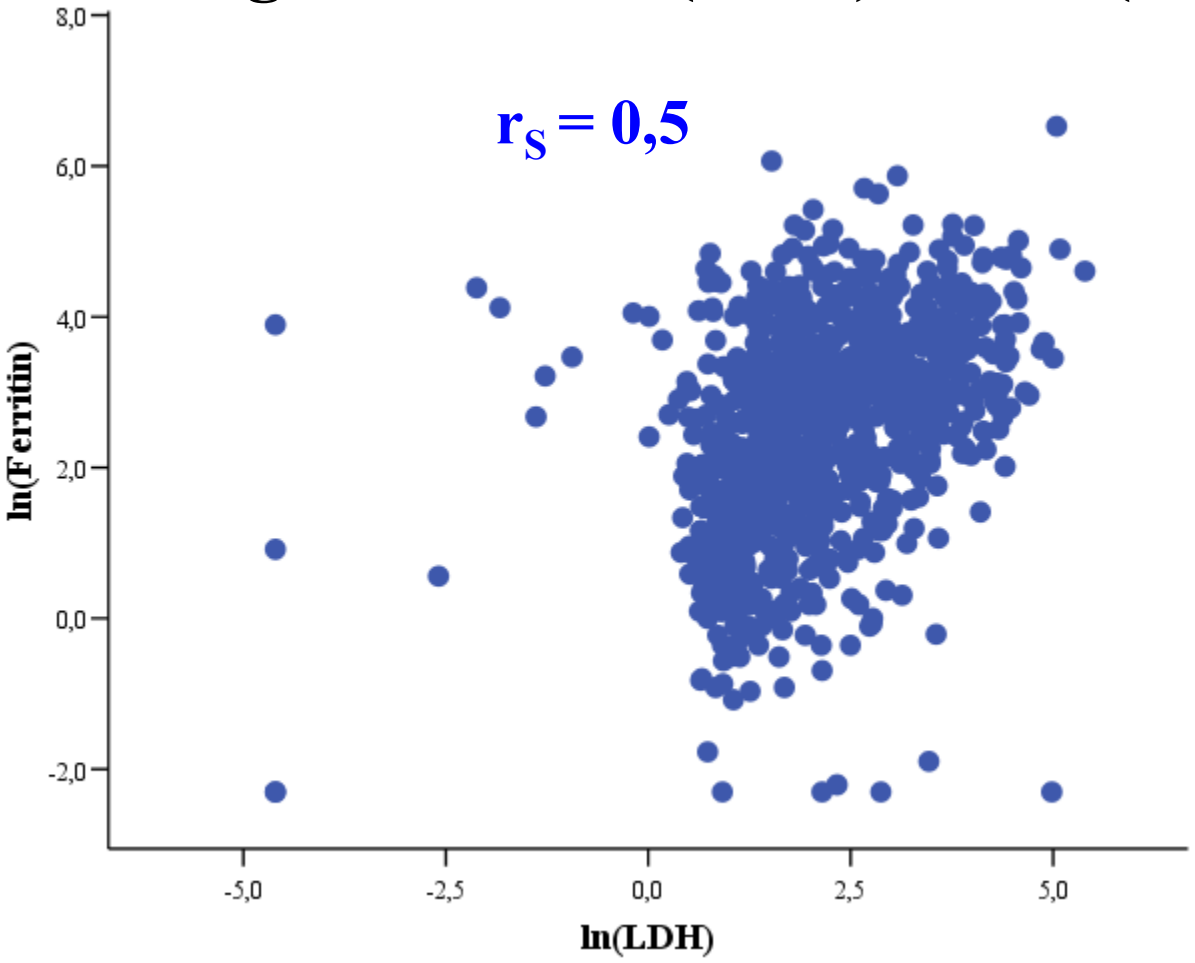
Zusammenhang nicht normalverteilter stetiger Größen

- 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

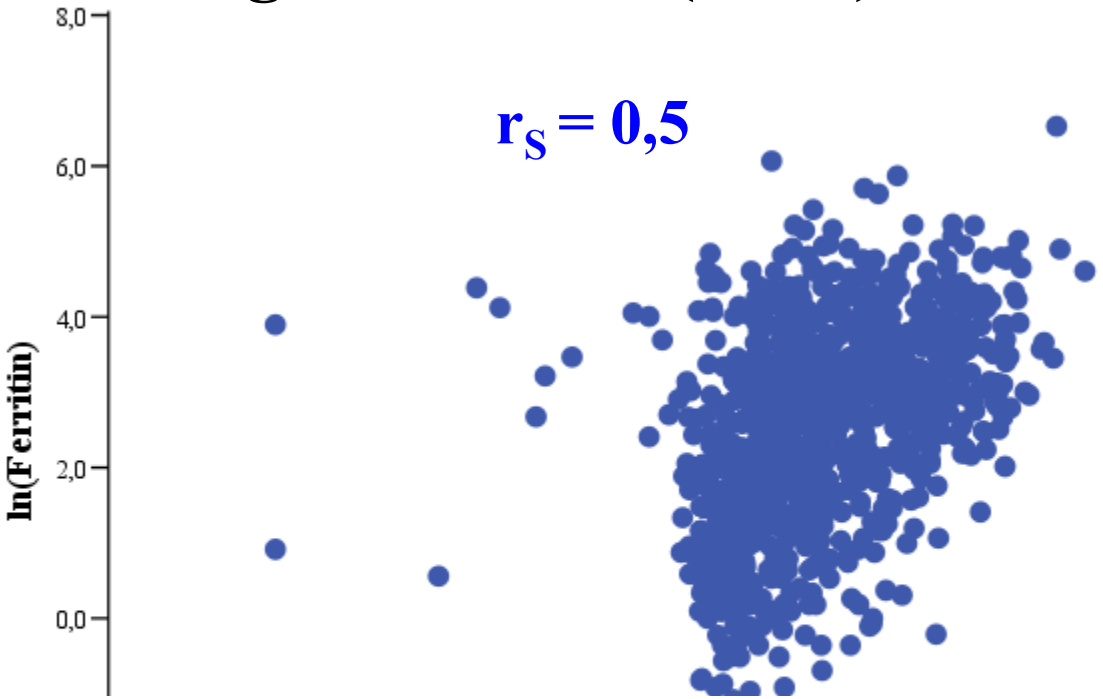
Zusammenhang zwischen LDH und Ferritin



Zusammenhang zwischen ln(LDH) und ln(Ferritin)



Zusammenhang zwischen ln(LDH) und ln(Ferritin)



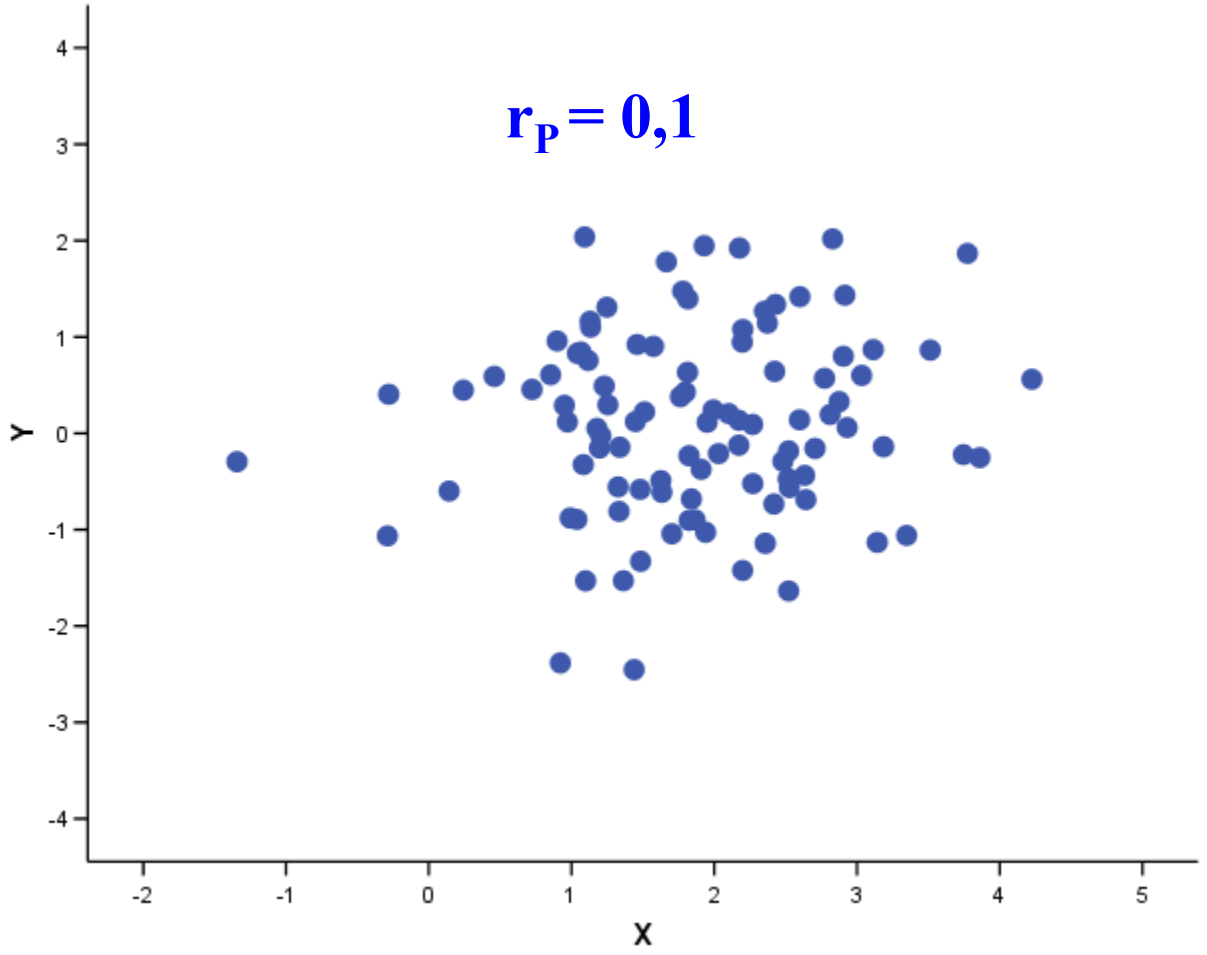
Logarithmieren ist eine monotone Transformation:

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

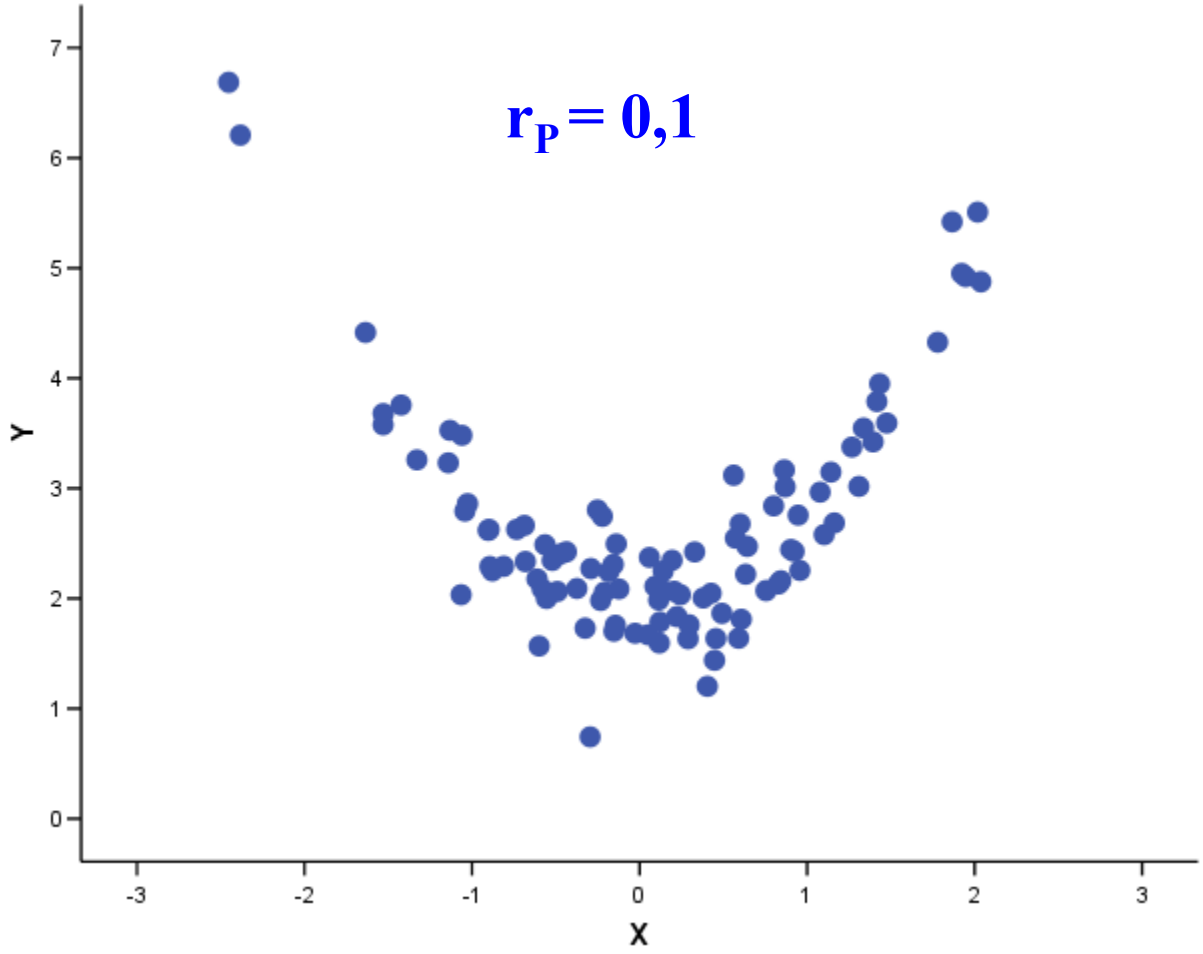
$\Rightarrow$ Ränge von X und ln(X) bzw. Y und ln(Y) sind gleich!

		kategorial		stetig	
		binär	**> 2 Kategorien**	**normalverteilt**	**nicht normalverteilt**
kategorial	**binär**			2 Boxplots / Histogramme 2*stetige Maßzahlen mit Mittelwert, Std.	2 Boxplots / Histogramme 2*stetige Maßzahlen ohne Mittelwert, Std.
	> 2 Kategorien				
stetig	**normal-verteilt**			Streudiagramm Linearer Zshg.: r_P Monotoner Zshg.: r_S	Streudiagramm Monotoner Zusammenhang: Korrelationskoeffizient nach Spearman r_S
	nicht norm.				

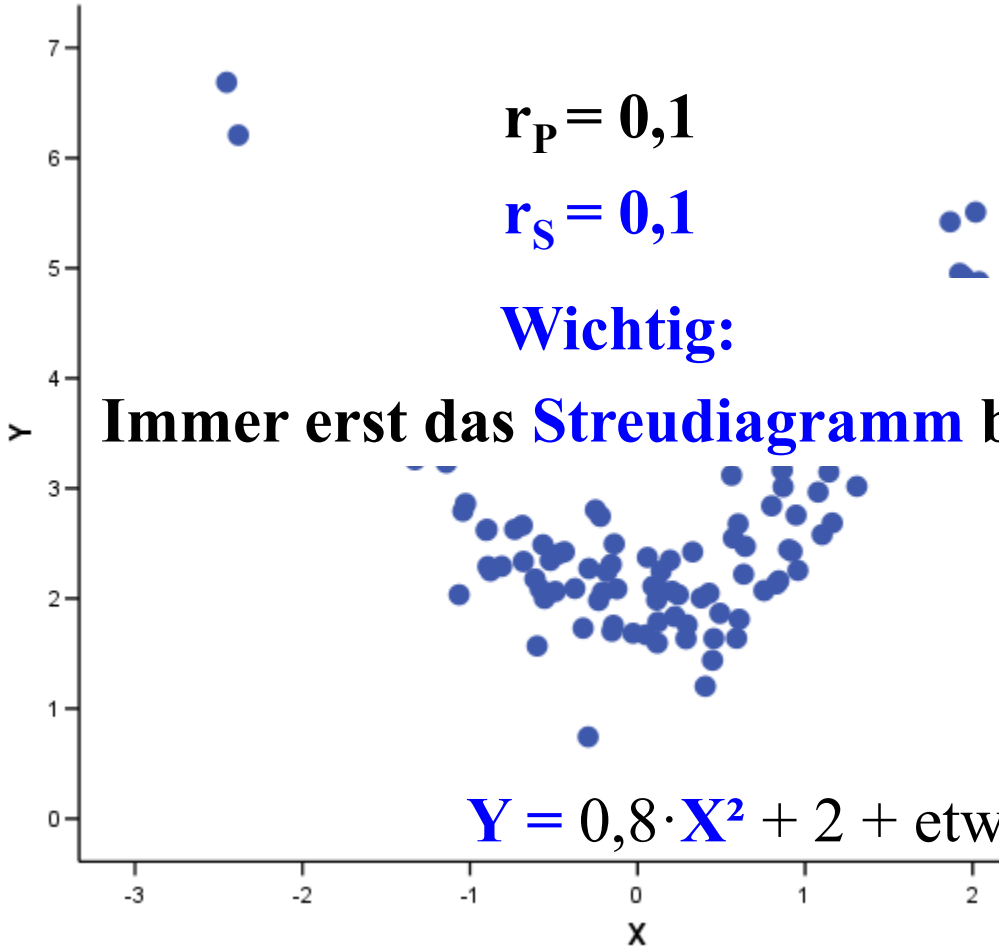
Vorsicht: Korrelationskoeffizient und Streudiagramm



Vorsicht: Korrelationskoeffizient und Streudiagramm



Vorsicht: Korrelationskoeffizient und Streudiagramm



$r_P = 0{,}1$

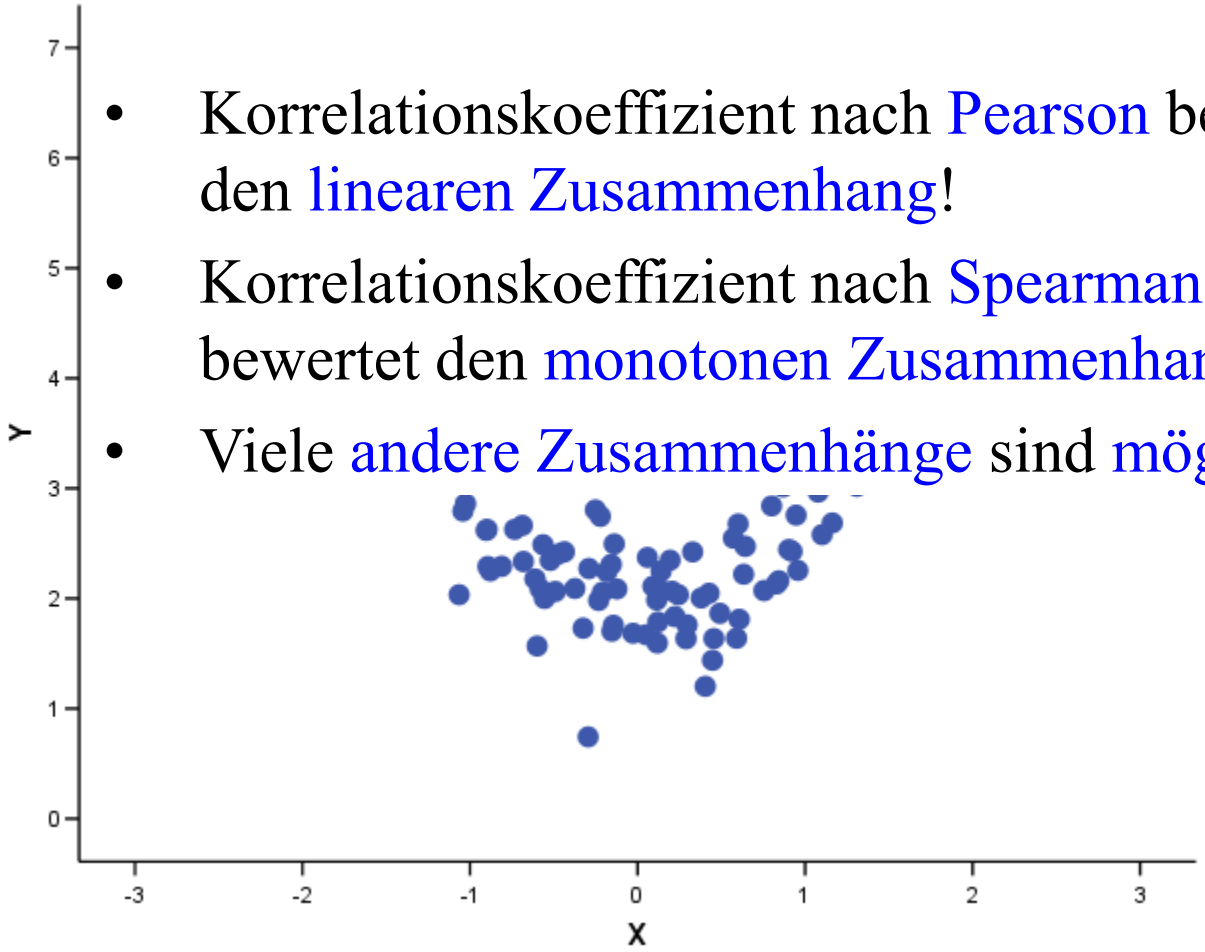
$r_S = 0{,}1$

Wichtig:

Immer erst das Streudiagramm betrachten!

$Y = 0{,}8 \cdot X^2 + 2 +$ etwas **Zufall**

Vorsicht: Korrelationskoeffizient und **Streudiagramm**



- 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**
kategorial	**binär**	?		2 Boxplots / Histogramme 2*stetige Maßzahlen mit Mittelwert, Std.	2 Boxplots / Histogramme 2*stetige Maßzahlen ohne Mittelwert, Std.
	> 2 Kategorien				
stetig	**normal-verteilt**			Streudiagramm Linearer Zshg.: r_P Monotoner Zshg.: r_S	Streudiagramm Monotoner Zusammenhang: Korrelationskoeffizient nach Spearman r_S
	nicht norm.				

Vergleich der Risikofaktoren

Folgende Fragen sollen geklärt werden:

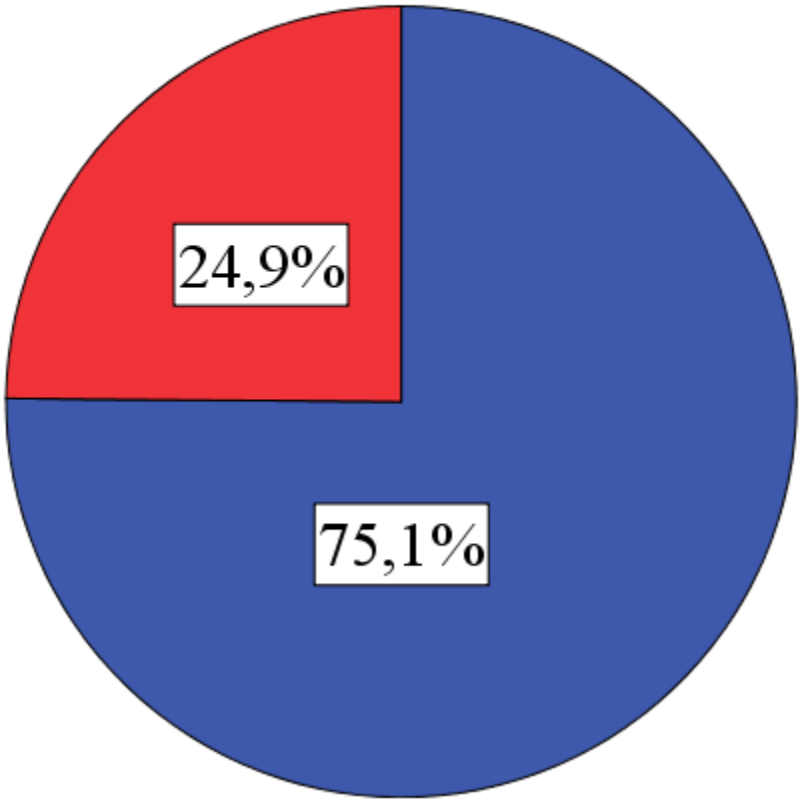
1. Gibt es einen Zusammenhang zwischen LDH und Ferritin?

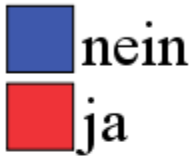
2. Hängen MYCN-Amplifikation und 1p-Anomalie zusammen?

3. Kommt die MYCN-Amplifikation in einigen INSS-Stadien häufiger vor?

4. Unterscheidet sich die Ferritin-Konzentration in den verschiedenen INSS-Stadien?

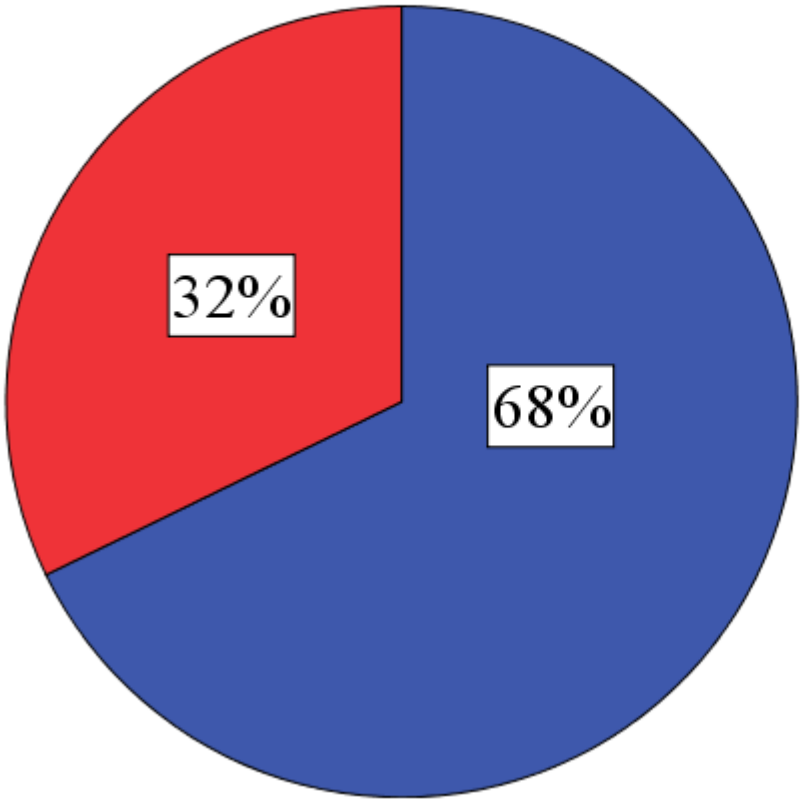
MYCN-Amplifikation

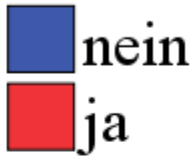




		Häufigke	Prozent	Gültige Prozente
Gültig	nein	1538	46,0	75,1
	ja	509	15,2	24,9
	Gesamt	2047	61,2	100,0
Fehlend	System	1298	38,8	
Gesamt		3345	100,0	

1p-Anomalie





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

Zusammenhang zwischen MYCN und 1p-Anomalie

MYCN-Amplifikation * 1p-Anomalie Kreuztabelle

			1p-Anomalie		Gesamt
			nein	ja	
MYCN-Amplifikation	nein	Anzahl	114	28	142
	ja	Anzahl	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	142
		% von MYCN-Amplifikation	80%	20%	100%
	ja	Anzahl	5	32	37
		% von MYCN-Amplifikation	14%	86%	100%
Gesamt		Anzahl	119	60	179
		% von MYCN-Amplifikation	66%	34%	100%

Zeilenprozente

Zusammenhang zwischen MYCN und 1p-Anomalie

MYCN-Amplifikation * 1p-Anomalie Kreuztabelle

			1p-Anomalie		
			nein	ja	Gesamt
MYCN-Amplifikation	nein	Anzahl	114	28	142
		% von MYCN-Amplifikation	80%	20%	100%
	ja	Anzahl	5	32	37
		% von MYCN-Amplifikation	14%	86%	100%
Gesamt		Anzahl	119	60	179
		% von MYCN-Amplifikation	66%	34%	100%

Zeilenprozente

Zusammenhang zwischen MYCN und 1p-Anomalie

MYCN-Amplifikation * 1p-Anomalie Kreuztabelle

			1p-Anomalie		
			nein	ja	Gesamt
MYCN-Amplifikation	nein	Anzahl	114	28	142
		% von MYCN-Amplifikation	80%	20%	100%
	ja	Anzahl	5	32	37
		% von MYCN-Amplifikation	14%	86%	100%
Gesamt		Anzahl	119	60	179
		% von MYCN-Amplifikation	66%	34%	100%

Zeilenprozente

Zusammenhang zwischen MYCN und 1p-Anomalie

MYCN-Amplifikation * 1p-Anomalie Kreuztabelle

			1p-Anomalie		Gesamt
			nein	ja	
MYCN-Amplifikation	nein	Anzahl	114	28	142
		% von MYCN-Amplifikation	80%	20%	100%
	ja	Anzahl	5	32	37
		% von MYCN-Amplifikation	14%	86%	100%
Gesamt		Anzahl	119	60	179
		% von MYCN-Amplifikation	66%	34%	100%

Zeilenprozente

Zusammenhang zwischen MYCN und 1p-Anomalie

MYCN-Amplifikation * 1p-Anomalie Kreuztabelle

			1p-Anomalie		
			nein	ja	Gesamt
MYCN-Amplifikation	nein	Anzahl	114	28	142
		% von 1p-Anomalie	96%	47%	79%
	ja	Anzahl	5	32	37
		% von 1p-Anomalie	4%	53%	21%
Gesamt		Anzahl	119	60	179
		% von 1p-Anomalie	100%	100%	100%

Spaltenprozente

Zusammenhang zwischen MYCN und 1p-Anomalie

MYCN-Amplifikation * 1p-Anomalie Kreuztabelle

			1p-Anomalie		
			nein	ja	Gesamt
MYCN-Amplifikation	nein	Anzahl	114	28	142
		% von 1p-Anomalie	96%	47%	79%
	ja	Anzahl	5	32	37
		% von 1p-Anomalie	4%	53%	21%
Gesamt		Anzahl	119	60	179
		% von 1p-Anomalie	100%	100%	100%

Spaltenprozente

Zusammenhang zwischen MYCN und 1p-Anomalie

MYCN-Amplifikation * 1p-Anomalie Kreuztabelle

			1p-Anomalie		
			nein	ja	Gesamt
MYCN-Amplifikation	nein	Anzahl	114	28	142
		% von 1p-Anomalie	96%	47%	79%
	ja	Anzahl	5	32	37
		% von 1p-Anomalie	4%	53%	21%
Gesamt		Anzahl	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	Anzahl	114	28	142
		% von MYCN-Amplifikation	80%	20%	100%
		% von 1p-Anomalie	96%	47%	79%
	ja	Anzahl	5	32	37
		% von MYCN-Amplifikation	14%	86%	100%
		% von 1p-Anomalie	4%	53%	21%
Gesamt		Anzahl	119	60	179
		% von MYCN-Amplifikation	66%	34%	100%
		% von 1p-Anomalie	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 ⇒ Zeilenprozente
- Fall-Kontroll-Studien:
  Bei Beginn der Studie Auswahl von Fällen und Kontrollen nach der Zielgröße ⇒ Spaltenprozente
- In unserem Beispiel: Keine natürliche Reihenfolge vorhanden
  ⇒ Angabe von Zeilen- und Spaltenprozenten sinnvoll

Risikomaße

- Risiko: relative Häufigkeit oder Wahrscheinlichkeit,
  dass ein Ereignis auftritt (Nebenwirkung, Heilung, …)
- Risikodifferenz: Differenz der Risiken in 2 Gruppen
- Relatives Risiko (RR): Quotient der Risiken zwischen 2 Gruppen
- Odds = Chance:
  Quotient aus Risiko für ein Ereignis und Risiko für kein Ereignis
  Wahrscheinlichkeit für eine 6 beim Würfeln = 1/6
  Chance auf eine 6 beim Würfeln = 1/6 : 5/6 = 1 : 5
- Odds Ratio = Chancenverhältnis (OR):
  Quotient der Odds zwischen 2 Gruppen

Risikomaße

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

MYCN-Amplifikation * 1p-Anomalie Kreuztabelle

% innerhalb von MYCN-Amplifikation

		1p-Anomalie		
		nein	ja	Gesamt
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) ·100% = 330% erhöht,

  wenn MYCN amplifiziert ist (vs. ohne MYCN-Amplifikation).

MYCN-Amplifikation * 1p-Anomalie Kreuztabelle

% innerhalb von MYCN-Amplifikation

		1p-Anomalie		
		nein	ja	Gesamt
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) ·100% = 77% gesenkt,

  wenn MYCN nicht amplifiziert ist (vs. MYCN-Amplifikation).

MYCN-Amplifikation * 1p-Anomalie Kreuztabelle

% innerhalb von MYCN-Amplifikation

		1p-Anomalie		
		nein	ja	Gesamt
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
- Chance einer 1p-Anomalie:
  - ohne MYCN-Amplifikation: 20% / 80% = 1 : 4 = 0,25
  - mit MYCN-Amplifikation: 86% / 14% = 6,14

MYCN-Amplifikation * 1p-Anomalie Kreuztabelle

% innerhalb von MYCN-Amplifikation

		1p-Anomalie		
		nein	ja	Gesamt
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) ·100% = 96% gesenkt,

  wenn MYCN nicht amplifiziert ist (vs. MYCN-Amplifikation).

Interpretation der Risikomaße

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

		kategorial		stetig	
		binär	**> 2 Kategorien**	**normalverteilt**	**nicht normalverteilt**
kategorial	**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.
	> 2 Kategorien				
stetig	**normal-verteilt**			Streudiagramm Linearer Zshg.: r_P Monotoner Zshg.: r_S	Streudiagramm Monotoner Zusammenhang: Korrelationskoeffizient nach Spearman r_S
	nicht norm.				

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		
			nein	ja	Gesamt
INSS-Stadium	1	Anzahl	304	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		Anzahl	1419	476	1895
		% von INSS-Stadium	74,9%	25,1%	100,0%
		% von MYCN-Amplifikation	100,0%	100,0%	100,0%

		kategorial		stetig	
		binär	**> 2 Kategorien**	**normalverteilt**	**nicht normalverteilt**
kategorial	**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.
	> 2 Kategorien			?	?
stetig	**normal-verteilt**			Streudiagramm Linearer Zshg.: r_P Monotoner Zshg.: r_S	Streudiagramm Monotoner Zusammenhang: Korrelationskoeffizient nach Spearman r_S
	nicht norm.				

Vergleich der Risikofaktoren

Folgende Fragen sollen geklärt werden:

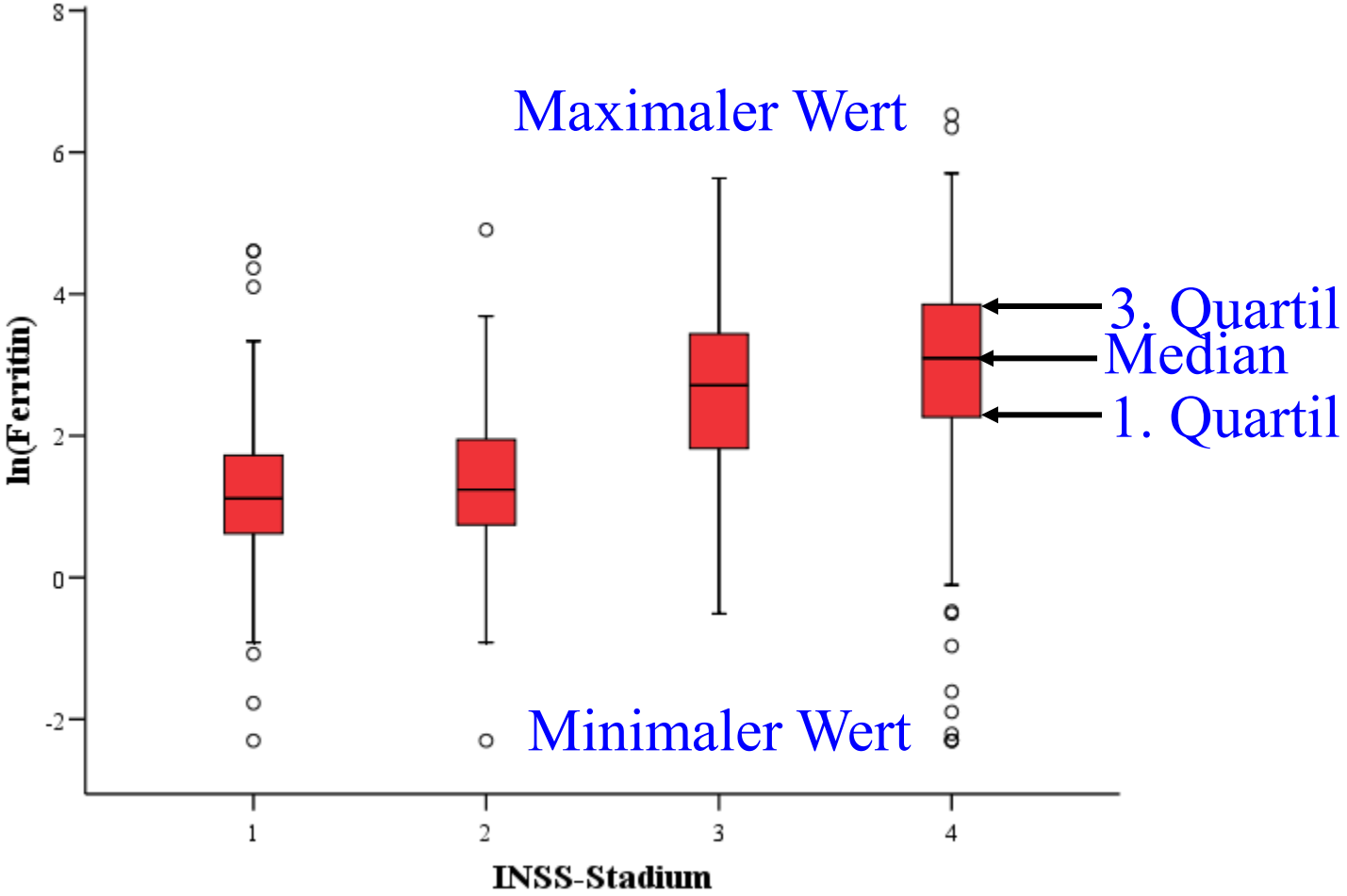
1. Gibt es einen Zusammenhang zwischen LDH und Ferritin?

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)

Statistiken[b]

ln(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

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

Statistiken[a]

ln(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

a. INSS-Stadium = 2

Statistiken[b]

ln(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

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

Statistiken[a]

ln(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

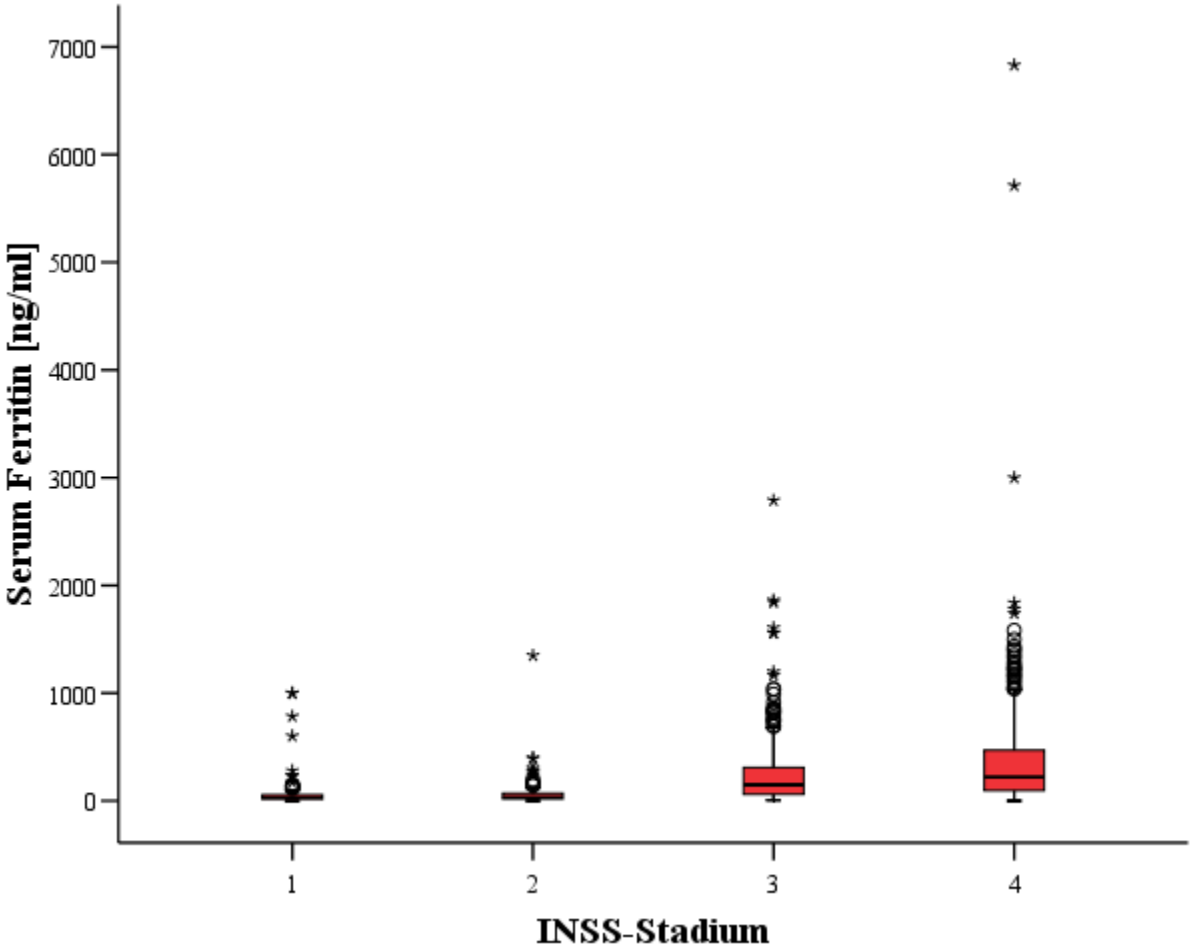
a. INSS-Stadium = 4

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

Mittelwert[?], Standardabweichung[?], Varianz[?], Schiefe, ✓

Median, Quartile, Minimum, Maximum, (Modus)

Zusammenhang zwischen INSS und Ferritin



Zusammenhang zwischen INSS und Serum Ferritin

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

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

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

a. INSS-Stadium = 2

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

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

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. INSS-Stadium = 4

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

~~Mittelwert~~[?], ~~Standardabweichung~~[?], ~~Varianz~~[?], Schiefe,

Median, Quartile, Minimum, Maximum, (Modus)

		kategorial		stetig	
		binär	**> 2 Kategorien**	**normalverteilt**	**nicht normalverteilt**
kategorial	**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.
	> 2 Kategorien			k Boxplots / Histogramme k*stetige Maßzahlen mit Mittelwert, Std.	k Boxplots / Histogramme k*stetige Maßzahlen ohne Mittelwert, Std.
stetig	**normal-verteilt**			Streudiagramm Linearer Zshg.: r_P Monotoner Zshg.: r_S	Streudiagramm Monotoner Zusammenhang: Korrelationskoeffizient nach Spearman r_S
	nicht norm.				

Vorlesungsinhalt

- Deskriptive Statistik
- Überlebenszeiten
- Statistisches Testen
  - Prinzipien
  - Spezielle Tests
  - Multiples Testen
- Fallzahlplanung / Adaptive Designs
- Regressionsmodelle
  - Logistische Regression
  - Coxregression

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

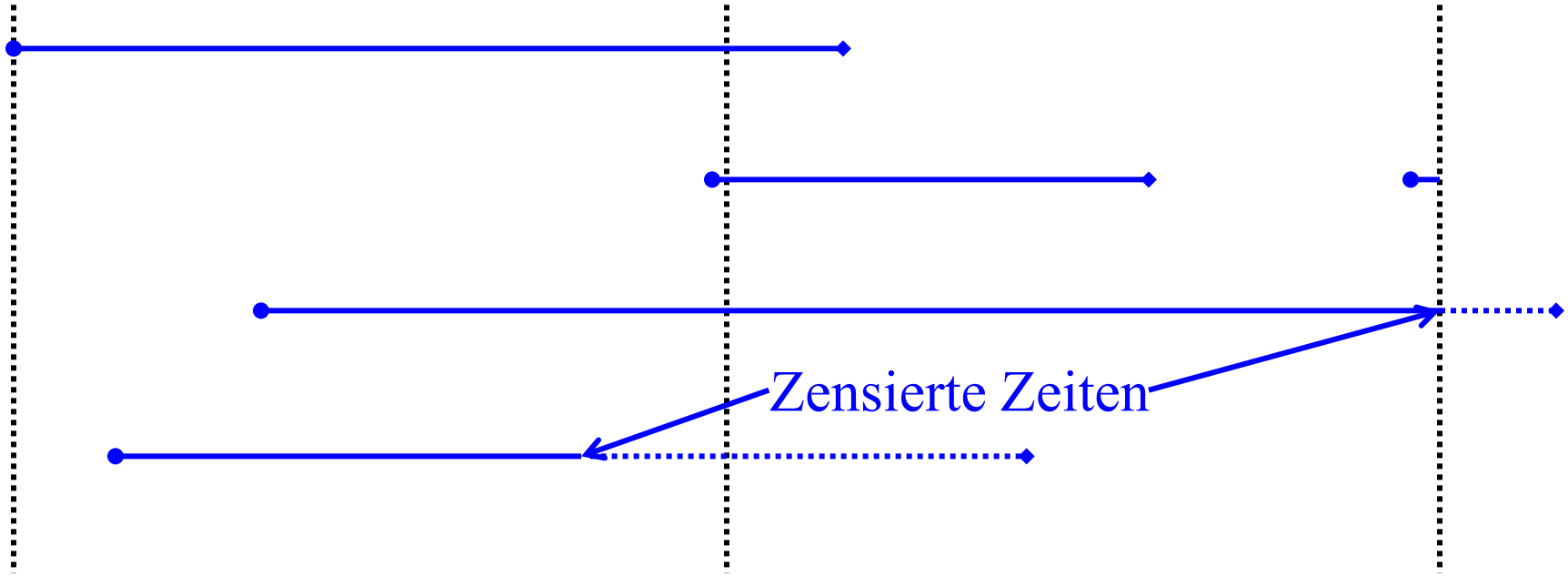
  Zeit bis Ereignis = Heilung bzw. Remission
- Gehstrecke

  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 ⇒ Dauer des Zustands bekannt
  b. Ereignis wird nicht beobachtet ⇒ Mindestdauer des Zustands = Beginn bis letzte Beobachtung ist bekannt
     = (rechts-)zensierte Überlebenszeiten

Überlebenszeitstudie



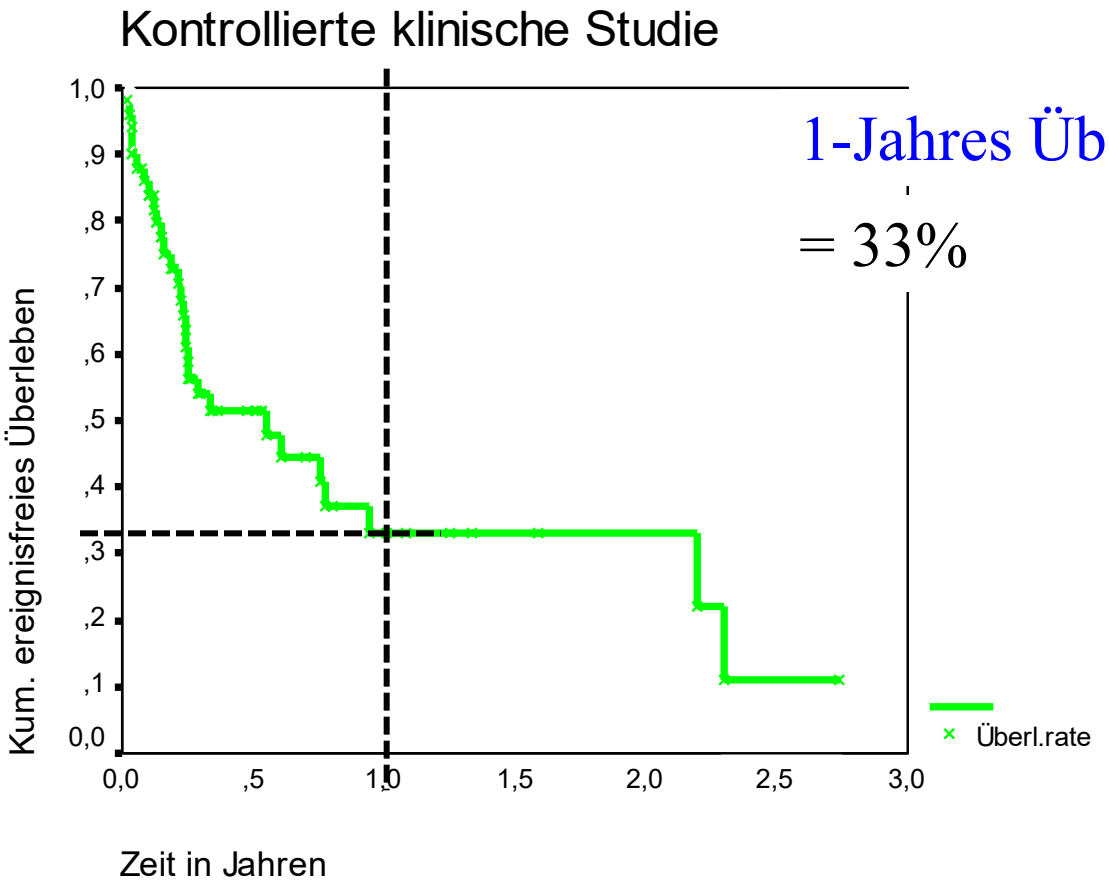
Wie stellt man Überlebenszeiten dar?

- 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

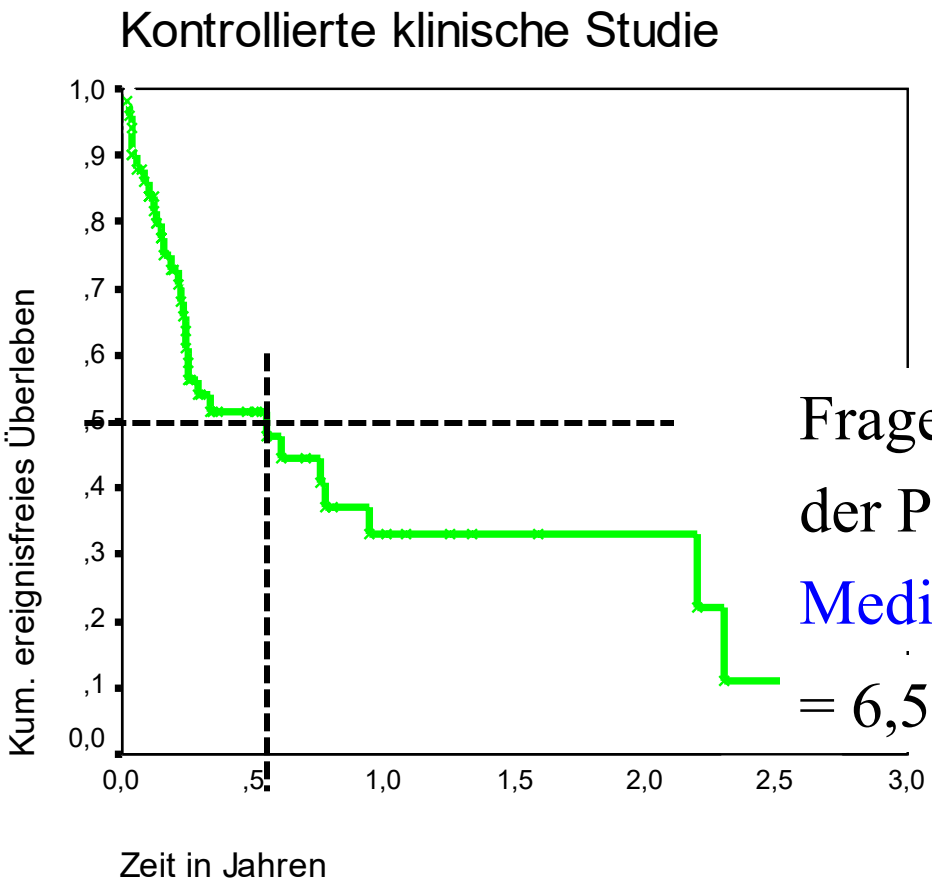
Überlebenszeitkurve nach Kaplan-Meier



1-Jahres Überlebensrate

= 33%

Überlebenszeitkurve nach Kaplan-Meier



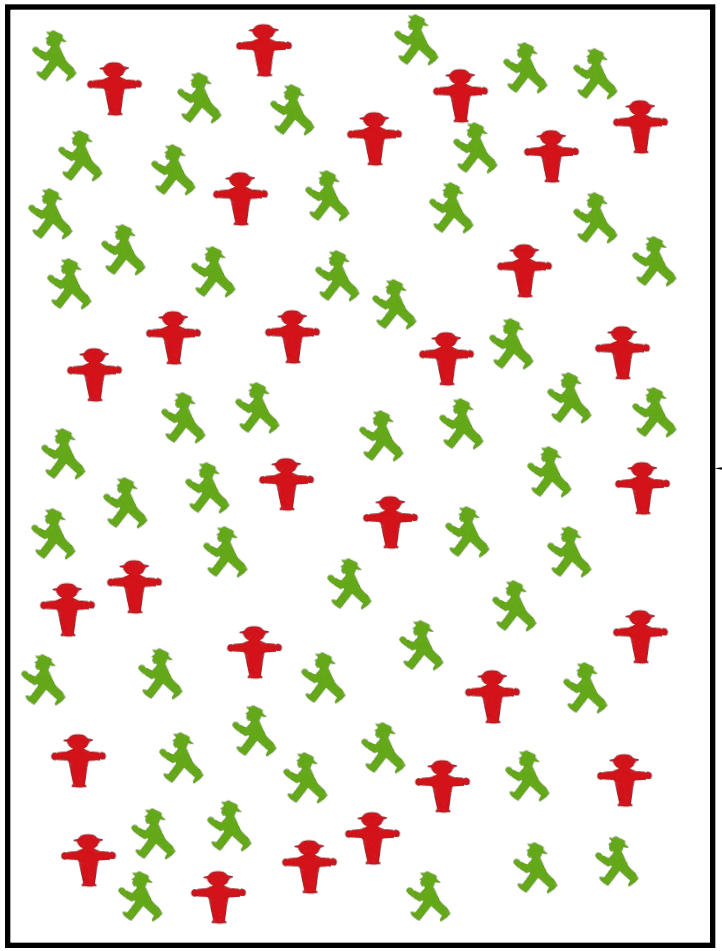
Frage: Wann haben 50% der Patienten ein Ereignis?

Mediane Überlebenszeit = 6,5 Monate

Vorlesungsinhalt

- Deskriptive Statistik
- Überlebenszeiten
- Statistisches Testen
  - Prinzipien
  - Spezielle Tests
  - Multiples Testen
- Fallzahlplanung / Adaptive Designs
- Regressionsmodelle
  - Logistische Regression
  - Coxregression

Grundgesamtheit und Stichprobe



Grundgesamtheit

Stichprobe

repräsentativ

Induktiver Schluss

Deskriptive Statistik:
Beschreibung der Stichprobe

Konfirmatorische Statistik:
Schlussfolgerung auf Grundgesamtheit

Probleme beim induktiven Schluss

- Wie kann man beweisen, dass Therapie A besser ist als B?
- Problem: Stichprobe liefert Teilinformation
  $\Rightarrow$ systematischer Fehler
  $\Rightarrow$ Überinterpretation von Zufallseffekten
- Vermeidung von systematischen Fehlern
  - Stichprobe (Repräsentativität, Randomisierung, …)
  - Zielgröße (Relevanz, Objektivität, Verblindung, …)
- Eingrenzung der Überinterpretation von Zufallseffekten
  - Statistische Tests (Signifikanztests)
  - Punktschätzer mit Konfidenzintervall

Fast sicherer Widerspruchsbeweis

- Hauptfrage: Wie wahrscheinlich ist ein beobachteter Effekt in der Stichprobe ohne einen Effekt in der Grundgesamtheit?
- Beispiel: 2 Therapien sind gleich gut, aber durch unglückliche Patientenverteilung erscheint Therapie A in der Studie besser.
- Idee: Stelle die (Null-)Hypothese auf, dass es keinen Effekt gibt
- Finde Methode, die Wahrscheinlichkeit für den beobachteten Studieneffekt zu berechnen, falls die Nullhypothese gilt (p-Wert)
- Wähle Grenze, ab der die Wahrscheinlichkeit so klein ist, dass die Nullhypothese unglaubwürdig wird (Signifikanzniveau)
- Führe die Studie durch $\Rightarrow$ lehne Hypothese ab oder nicht

Beispiel

- Frage: Eine Impfung soll Infektionen verhindern.
  Die Impfung wird als Erfolg angesehen, wenn mehr als 50% der Infektionen verhindert werden können.
- Stichprobe mit 10 ProbandInnen
- Nullhypothese H_0: Erfolgsrate der Impfung = 50% = ½
- Bezeichne E = Impferfolg, V = Impfversagen
  Mögliche 2^{10} Ergebnisse: (V,V,…,V), (E,V,…V), …, (E,E,…,E)
- Unter H_0 sind alle 2^{10} Studienergebnisse gleich wahrscheinlich
  $\Rightarrow$ Wahrscheinlichkeit für jedes Studienergebnis = $2^{-10} \approx 0{,}001$
- Setze Grenze für die Glaubwürdigkeit von H_0 auf 5%

Anzahl der Erfolge #E	#Studien $= \binom{10}{\#E}$	Wahrscheinlichkeit unter H_0 für genau #E Erfolge	Wahrscheinlichkeit unter H_0 für mindestens #E Erfolge	P-Wert 2-seitig
10	1	0,001	0,001	
9	10	0,010	0,011	
8	45	0,044	0,055	
7	120	0,117	0,172	

Anzahl der Erfolge #E	#Studien $= \binom{10}{\#E}$	Wahrscheinlichkeit unter H_0 für genau #E Erfolge	Wahrscheinlichkeit unter H_0 für mindestens #E Erfolge	P-Wert 2-seitig
10	1	Beachte: Viele V widersprechen H_0 genauso wie viele E		
9	10			
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	0,010	0,999	
0	1	0,001	1,000	

Anzahl der Erfolge #E	#Studien $= \binom{10}{\#E}$	Wahrscheinlichkeit unter H_0 für genau #E Erfolge	Wahrscheinlichkeit unter H_0 für mindestens #E Erfolge	P-Wert 2-seitig
10	1	0,001	0,001	
9	10	0,010	0,011	
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	0,010	0,999	
0	1	0,001	1,000	

Anzahl der Erfolge #E	#Studien $= \binom{10}{\#E}$	Wahrscheinlichkeit unter H_0 für genau #E Erfolge	Wahrscheinlichkeit unter H_0 für mindestens #E Erfolge	P-Wert 2-seitig
10	1	+ **0,001**	0,001	0,002
9	10	0,010	0,011	
8				
7				
6				
5				
4				
3	120	0,117	0,945	
2	45	0,044	0,989	
1	10	0,010	0,999	
0	1	+ **0,001**	1,000	0,002

P-Wert = 0,002 = 0,2% bedeutet:
Wenn die wahre Erfolgsrate der Impfung 50% beträgt (H_0), ist die Wahrscheinlichkeit bei 10 Versuchen 10 Erfolge oder 10 Misserfolge zu erzielen 0,2%.

Anzahl der Erfolge #E	#Studien $= \binom{10}{\#E}$	Wahrscheinlichkeit unter H_0 für genau #E Erfolge	Wahrscheinlichkeit unter H_0 für mindestens #E Erfolge	P-Wert 2-seitig
10	1	+ **0,001**	0,001	0,002
9	10	+ **0,010**	0,011	0,021
8				
7				
6				
5				
4				
3	120	0,117	0,945	
2	45	0,044	0,989	
1	10	+ **0,010**	0,999	0,021
0	1	+ **0,001**	1,000	0,002

P-Wert = 0,021 = 2,1% bedeutet:
Wenn die wahre Erfolgsrate der Impfung 50% beträgt (H_0), ist die Wahrscheinlichkeit bei 10 Versuchen mindestens 9 Erfolge oder 9 Misserfolge zu erzielen 2,1%.

Anzahl der Erfolge #E	#Studien $= \binom{10}{\#E}$	Wahrscheinlichkeit unter H_0 für genau #E Erfolge	Wahrscheinlichkeit unter H_0 für mindestens #E Erfolge	P-Wert 2-seitig
10	1	0,001	0,001	0,002
9	10	0,010	0,011	0,021
8	45	0,044	0,055	0,109
7	120	0,117	0,172	0,344
6	210	0,205	0,377	0,754
5	252	0,246	0,623	1,000
4	210	0,205	0,828	0,754
3	120	0,117	0,945	0,344
2	45	0,044	0,989	0,109
1	10	0,010	0,999	0,021
0	1	0,001	1,000	0,002

Anzahl der Erfolge #E	#Studien $= \binom{10}{\#E}$	Wahrscheinlichkeit unter H_0 für genau #E Erfolge	Wahrscheinlichkeit unter H_0 für mindestens #E Erfolge	P-Wert 2-seitig
10	1	0,001	0,001	**0,002**
9	10	0,010	0,011	**0,021**
8	P-Wert $\leq 0{,}05 \Rightarrow$ Studie oder H_0 unglaubwürdig $\Rightarrow$ H_0 ablehnen, Erfolgsrate ist signifikant größer als 50%			0,109
7				0,344
6				0,754
5	252	0,246	0,623	1,000
4	210	0,205	0,828	0,754
3	120	0,117	0,945	0,344
2	45	0,044	0,989	0,109
1	10	0,010	0,999	0,021
0	1	0,001	1,000	0,002

Anzahl der Erfolge #E	#Studien $= \binom{10}{\#E}$	Wahrscheinlichkeit unter H_0 für genau #E Erfolge	Wahrscheinlichkeit unter H_0 für mindestens #E Erfolge	P-Wert 2-seitig
10	1	0,001	0,001	**0,002**
9	10	0,010	0,011	**0,021**
8	P-Wert $\leq 0,05 \Rightarrow$ Studie oder H_0 unglaubwürdig $\Rightarrow$ H_0 ablehnen, Erfolgsrate ist signifikant größer als 50%			0,109
7				0,344
6				0,754
5	252	0,246	0,623	1,000
4	P-Wert $\leq 0,05 \Rightarrow$ Studie oder H_0 unglaubwürdig $\Rightarrow$ H_0 ablehnen, Erfolgsrate ist signifikant kleiner als 50%			0,754
3				0,344
2				0,109
1	10	0,010	0,999	**0,021**
0	1	0,001	1,000	**0,002**

Grundidee des statistischen Testens

Zu untersuchender Effekt	Studie zeigt keinen Effekt	Studie zeigt einen Effekt
In Wirklichkeit gibt es keinen Effekt	Richtige Studienaussage	
In Wirklichkeit gibt es einen Effekt		Richtige Studienaussage

Grundidee des statistischen Testens

Zu untersuchender Effekt	Studie zeigt keinen Effekt	Studie zeigt einen Effekt
In Wirklichkeit gibt es keinen Effekt	Richtige Studienaussage	Fehler 1. Art = Typ I – Fehler
In Wirklichkeit gibt es einen Effekt		Richtige Studienaussage

Maximale Wahrscheinlichkeit für einen Fehler 1. Art:

α = Signifikanzniveau: festgesetzt auf 0,1%, 1%, 5% oder 10%

Methode des statistischen Testens

- Fragestellung → Zielgröße (Impfbeispiel: Impferfolg ja / nein)
- Alternativhypothese $\rightarrow_{\text{Negation}}$ Nullhypothese H_0
- Wähle Teststatistik = Zusammenfassung der Studienergebnisse (Impfbeispiel: Zahl der Impferfolge)
- Bestimme Wahrscheinlichkeit der Teststatistik unter H_0
- Wähle Signifikanzniveau α (5%)
- Führe Studie durch
- p-Wert = Wahrscheinlichkeit für das beobachtete oder ein noch extremeres Studienergebnis (Teststatistik) unter H_0
- Vergleich p-Wert mit Signifikanzniveau → Testentscheidung

Signifikanzniveau

- Studienergebnis ist auch Zufallsprodukt
- Signifikanzniveau = Niveau = α
  Maximal zulässige Wahrscheinlichkeit, irrtümlich einen vermeintlichen Studieneffekt nachzuweisen, wenn in Wirklichkeit dieser Studieneffekt gar nicht existiert.
- Übliche Wahl:
  ICH-E9:“Conventionally the probability of type I error is set at 5% or less …
  ...or as dictated by any adjustments made necessary for multiplicity considerations”

Statistische Hypothesen

- Formale Beschreibung des gewünschten Studieneffekts
- Nullhypothese = H_0 =
  die Vermutung zum Studieneffekt, die widerlegt werden soll
- Alternativhypothese = H_1 =
  die Vermutung zum Studieneffekt, die nachgewiesen werden soll = Gegenteil der Nullhypothese
- Wichtig: korrekte Entsprechung zur Fragestellung
- Eindeutigkeit
- Vollständigkeit

Ein- und zweiseitige Nullhypothesen H_0

- Nullhypothese 1: Die intensivierte Induktionstherapie führt zu keiner längeren ereignisfreien Überlebenszeit (EFS) als die Standardinduktionstherapie.
- Nullhypothese 2: Die intensivierte Induktionstherapie führt zu keiner anderen EFS als die Standardinduktionstherapie.
- Nullhypothese 1: Einseitig = eine Richtung wird vorgegeben. Frage, ob die intensivierte Induktion besser ist.
- Nullhypothese 2: Zweiseitig = keine Richtung wird vorgegeben. Frage, ob intensivierte Induktion besser oder schlechter ist.
- Standard = zweiseitige Hypothese
- Einseitige Hypothese $\Rightarrow^{\text{ICH E9}}$ α halbieren: z.B. 2,5% statt 5%

P-Wert

- Definition: Vorausgesetzt die Nullhypothese trifft zu,
  d.h. der gewünschte Effekt existiert nicht in der Grundgesamtheit
  (z.B. kein Unterschied zwischen neuer und alter Therapie).
  Wahrscheinlichkeit, ein Studienergebnis wie das tatsächlich beobachtete zu erhalten oder eines,
  dass der Nullhypothese mindestens genauso widerspricht?
- Ergebnis eines statistischen Tests
- $0 \leq$ p-Wert ≤ 1
- Cave: p-Wert $\neq$ Wahrscheinlichkeit der Nullhypothese H_0
  p-Wert = $P(\text{Daten} \mid H_0) \neq P(H_0 \mid \text{Daten})$
- Trotzdem oft als Maß für die Plausibilität von H_0 benutzt

P-Wert

Kriterium zur Entscheidung für oder gegen H_0

- p-Wert klein

  $\Rightarrow$ Studienergebnis unplausibel oder H_0 falsch

  Bei Vertrauen in die Studie $\Rightarrow$ H_0 ablehnen
- p-Wert groß

  $\Rightarrow$ Studienergebnis plausibel $\Rightarrow$ H_0 beibehalten

  Aber H_0 nicht bewiesen, da p-Wert unter H_0 berechnet wird

  Ringschluss: Wenn H_0 falsch, dann ev. p-Wert falsch
- Ergo: Mit einem p-Wert lässt sich H_0 fast sicher widerlegen, aber nicht beweisen.

P-Wert und Signifikanzniveau

- Grenze für „klein“ bzw. „groß“ ist das Signifikanzniveau α
- Falls p-Wert $\leq \alpha$

  $\Rightarrow$ H_0 ablehnen, H_1 annehmen

  Das Ergebnis ist dann statistisch signifikant.

  Das Studienergebnis kam aus Sicht von H_0 überzufällig zustande
- Falls p-Wert $> \alpha$

  $\Rightarrow$ H_0 beibehalten, H_1 (noch) nicht annehmen

  Das Ergebnis ist dann nicht statistisch signifikant.

  Das Studienergebnis widerspricht nicht H_0

Signifikanz - Interpretation

- Formulierung wichtig: nicht signifikant ≠ signifikant nicht
  - Nicht signifikant ⇔ p-Wert hoch ⇒ keine sichere Aussage: H_0 richtig oder H_0 mit zugehörigem p-Wert ist falsch
  - Signifikant nicht: Das Gegenteil einer Aussage (z.B. die Therapien sind nicht gleich) wurde signifikant nachgewiesen
- Ein Signifikanztest auf Unterschied eignet sich
  - zum Nachweis von Unterschieden
  - nicht zum Nachweis von Gleichheit
- Statistisch signifikant sagt aus, dass ein Studieneffekt vorhanden aber nicht wie groß dieser ist.

Signifikanz versus Relevanz

- Statistische Signifikanz ≠ medizinische Relevanz
- Statistische Signifikanz = Verallgemeinerung vom Studienergebnis mit kontrollierter Irrtumswahrscheinlichkeit
- Statistische Signifikanz
  - beruht auf einem großen Studieneffekt
  - oder wurde durch eine große Fallzahl belegt, d.h. der eventuell kleine Studieneffekt ist stabil
- Medizinische Relevanz
  - bewertet die Größe des beobachteten Studieneffektes
  - abhängig von der jeweiligen Fragestellung

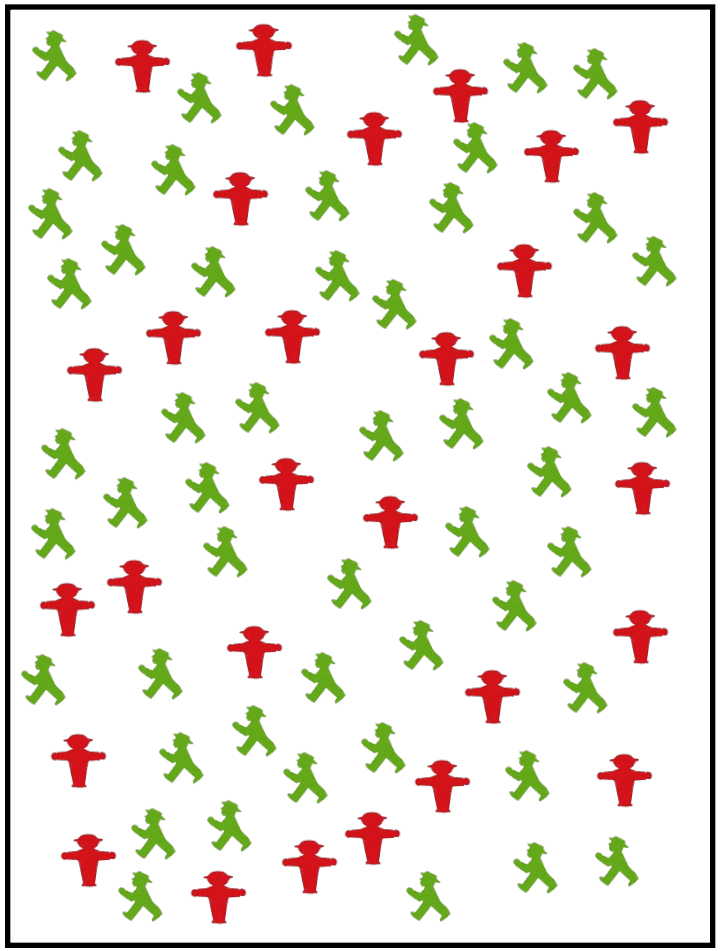
Kombination von Signifikanz und Relevanz

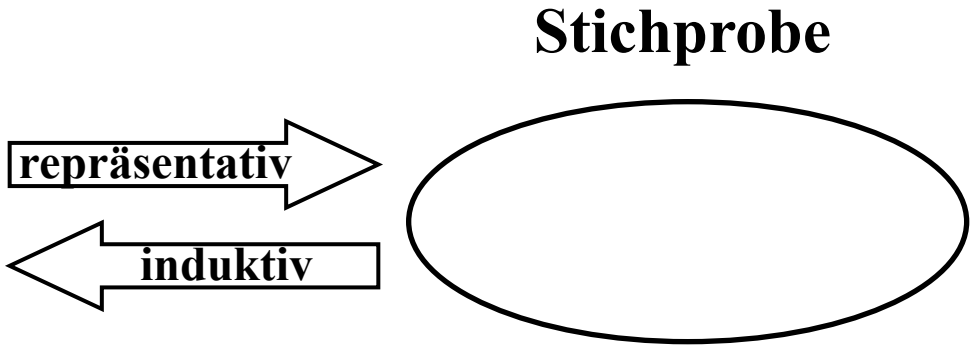
- Aufgabe: Schätzung des wahren Effektes (Parameterschätzer)
  z. B. für das Risiko, den Erwartungswert, …
- Fragen für den wahren Effekt:
  1. Was ist die beste Schätzung?
  2. In welchem Bereich liegt er mit hoher Wahrscheinlichkeit?
- Antworten:
  1. Gütekriterien für einen Effekt-/Parameterschätzer:
     u. a. unverzerrt, maximale Wahrscheinlichkeit für die beobachteten Daten (relative Häufigkeit, Mittelwert, …)
  2. Konfidenzintervall

Konfidenzintervall

- Bisher: Schätzung eines Parameters nahe dem wahren Wert
- Idee: Wahl eines möglichst kleinen Bereichs/Intervalls, der den wahren Wert mit hoher Wahrscheinlichkeit enthält
- Definition: $S(\delta) \in$ (1-α)-Konfidenzintervall [U(x), O(x)] $\Leftrightarrow P_\delta(U(x) \leq S(\delta) \leq O(x)) \geq 1-\alpha$
- Interpretation: (1-α)-Konfidenzintervall = (1-α)-KI Bereich um Effektschätzer der Studie, welcher den wahren Wert mindestens mit Wahrscheinlichkeit 1-α enthält
- Oft: $\alpha = 5\% \Rightarrow$ Konfidenzintervall mit 95% Wahrscheinlichkeit
- $\alpha \uparrow$ oder |Fallzahl| $\uparrow \Rightarrow$ |(1-α)-KI| $\downarrow$

Beispiel: Konfidenzintervall





Gesucht: Wahrscheinlichkeit für

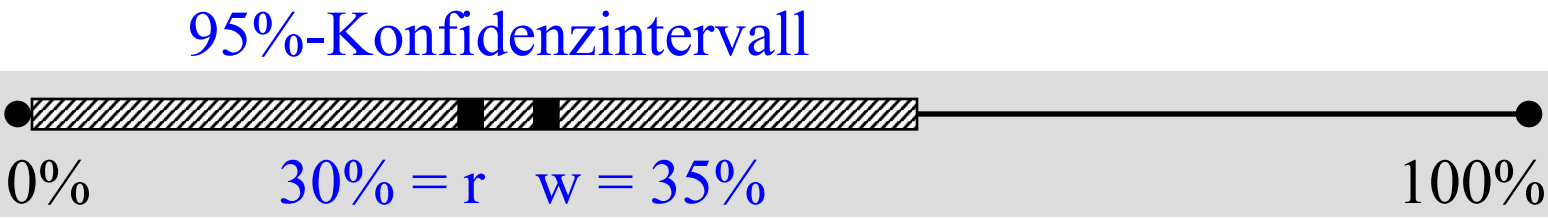
Grundgesamtheit (w): 28/80 = 35%

Stichprobe: relative Häufigkeit = 3/10

Schätzer (r): 30%

95%-Konfidenzintervall: [2%; 58%]

Beispiel: Konfidenzintervalle



Gesucht: Wahrscheinlichkeit für

Grundgesamtheit (w): 28/80 = 35%

Stichprobe: relative Häufigkeit = 3/10

Schätzer (r): 30%

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
- …

Spezielle Konfidenzintervalle

Konfidenzintervalle existieren für folgende Punktschätzer:

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

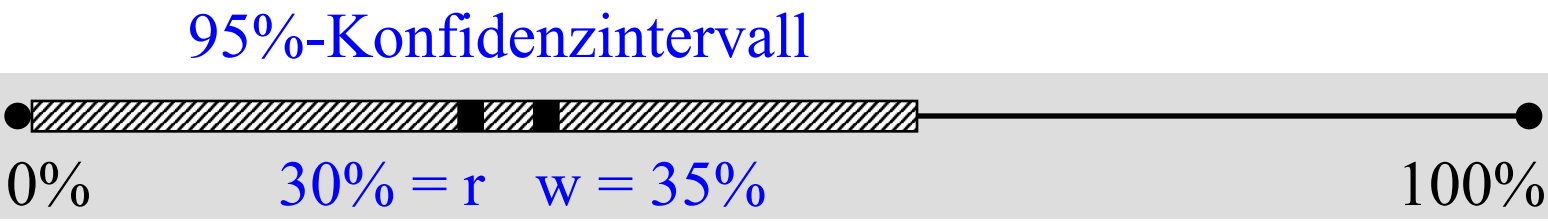
Konfidenzintervall binärer Merkmale

- Sei X eine binäre Zufallsvariable (Merkmal) mit Ausprägung: Ereignis ja oder nein
- w = wahre Wahrscheinlichkeit für X = ja
- r = beobachtete relative Häufigkeit einer Stichprobe für X = ja
- n = Stichprobenumfang
- 95%-KI (Konfidenzintervall) für w

$$\left[r - 1{,}96 \cdot \sqrt{\frac{r \cdot (1-r)}{n}}, r + 1{,}96 \cdot \sqrt{\frac{r \cdot (1-r)}{n}} \right] = r \pm 1{,}96 \cdot \sqrt{\frac{r \cdot (1-r)}{n}}$$

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

Beispiel: Konfidenzintervall



Gesucht: Wahrscheinlichkeit für

Grundgesamtheit (w): 28/80 = 35%

Stichprobe: relative Häufigkeit = 3/10

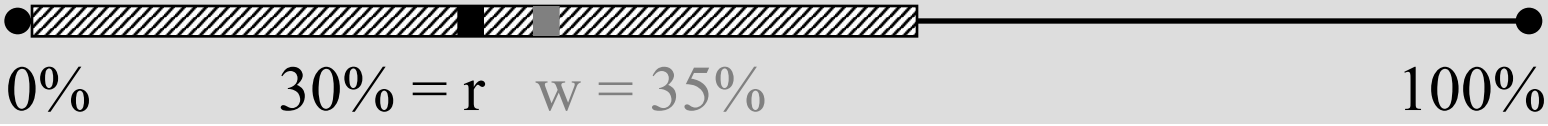
Schätzer (r): 30%

95%-Konfidenzintervall: [2%; 58%]

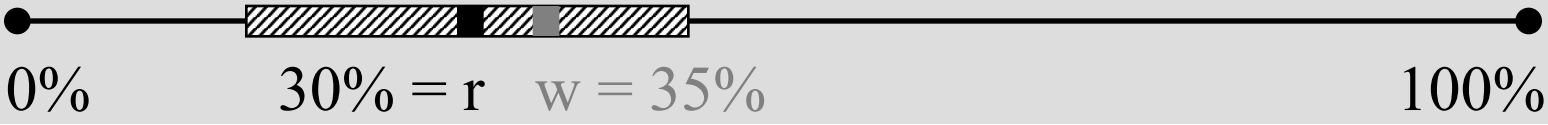
Beispiel: Konfidenzintervall

95%-Konfidenzintervall

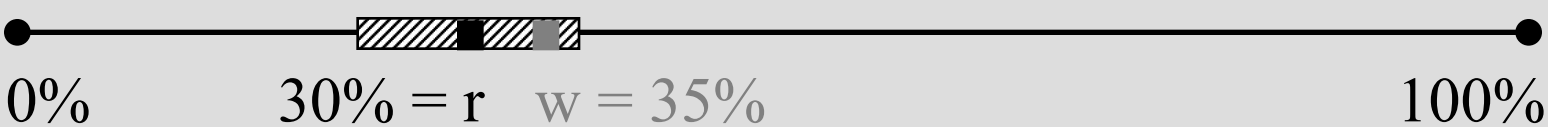
n = 10



n = 40



n = 160



Beispiel: Konfidenzintervall

n = 160

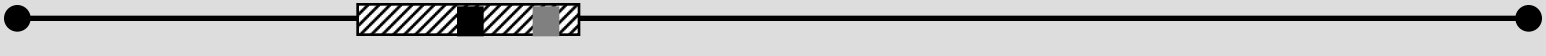
90%-Konfidenzintervall



99%-Konfidenzintervall

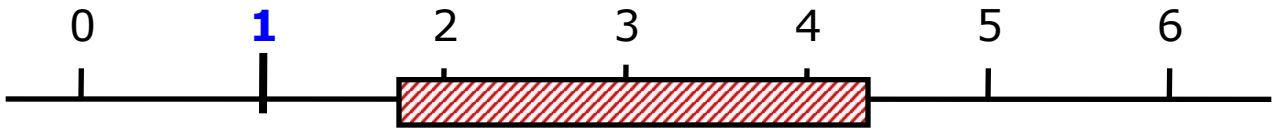


95%-Konfidenzintervall



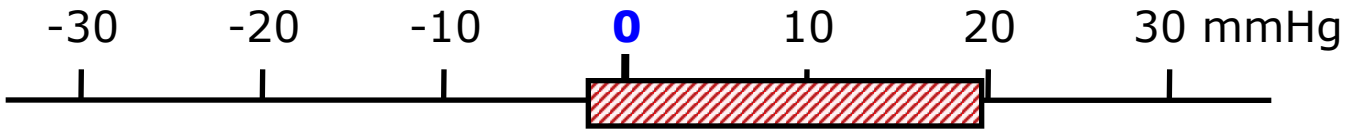
Konfidenzintervall versus Signifikanzniveau

- I. d. R.: Statistischer Test ≅ Konfidenzintervall
- Sei δ der interessierende Parameter
- H_0: $\delta = \delta_0$ (zu widerlegender Wert)
- $\delta_0 \notin$ (1-α)-KI $\Leftrightarrow$ H_0 ablehnen $\Leftrightarrow$ p-Wert $\leq \alpha$
- Beispiel 1: Wahrscheinlichkeit eines Bronchialkarzinoms (BC)
  - H_0: BC bei Rauchern = BC bei Nichtrauchern
  - (1-α)-KI des Relativen Risikos für BC (Rauchen: ja vs. nein)
  - 1 $\notin$ (1-α)-KI $\Leftrightarrow$ sign. Risikounterschied zum Niveau α



Konfidenzintervall versus Signifikanzniveau

- I. d. R.: Statistischer Test ≅ Konfidenzintervall
- Sei δ der interessierende Parameter
- H_0: $\delta = \delta_0$ (zu widerlegender Wert)
- $\delta_0 \notin$ (1-α)-KI $\Leftrightarrow$ H_0 ablehnen $\Leftrightarrow$ p-Wert $\leq \alpha$
- Beispiel 2: erwartete Blutdrucksenkung (BS)
  - H_0: BS unter Therapie A = BS unter Therapie B
  - Bilde (1-α)-KI für (BS unter A) – (BS unter B)
  - $0 \in$ (1-α)-KI $\Leftrightarrow$ kein sign. Unterschied der BS zum Niveau α



Vorlesungsinhalt

- Deskriptive Statistik
- Überlebenszeiten
- Statistisches Testen
  - Prinzipien
  - Spezielle Tests
  - Multiples Testen
- Fallzahlplanung / Adaptive Designs
- Regressionsmodelle
  - Logistische Regression
  - Coxregression

Zusammenhang zwischen INSS und MYCN

INSS-Stadium * MYCN-Amplifikation Kreuztabelle

			MYCN-Amplifikation		
			nein	ja	Gesamt
INSS-Stadium	1	Anzahl	304	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		Anzahl	1419	476	1895
		% von INSS-Stadium	74,9%	25,1%	100,0%
		% von MYCN-Amplifikation	100,0%	100,0%	100,0%

Test für den Vergleich zweier kategorialer Größen

- Frage: Gibt es einen Zusammenhang zwischen dem INSS-Stadium und der MYCN-Amplifikation?
- 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. 0,05
- Statistischer Test: Chiquadrat-Unabhängigkeitstest = χ^2-Test
- Zweiseitiger P-Wert:
  SPSS-Ausgabe: $p = 0{,}000 \Rightarrow p < 0{,}001$

Test für den Vergleich zweier kategorialer Größen

- Sachaussage:
  - Studienkollektiv
  - Signifikanzniveau
  - Signifikant: ja oder nein
  - Merkmale
  - Falls signifikant: Richtung der Aussage
- 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.

Chiquadrat-Unabhängigkeitstest Test

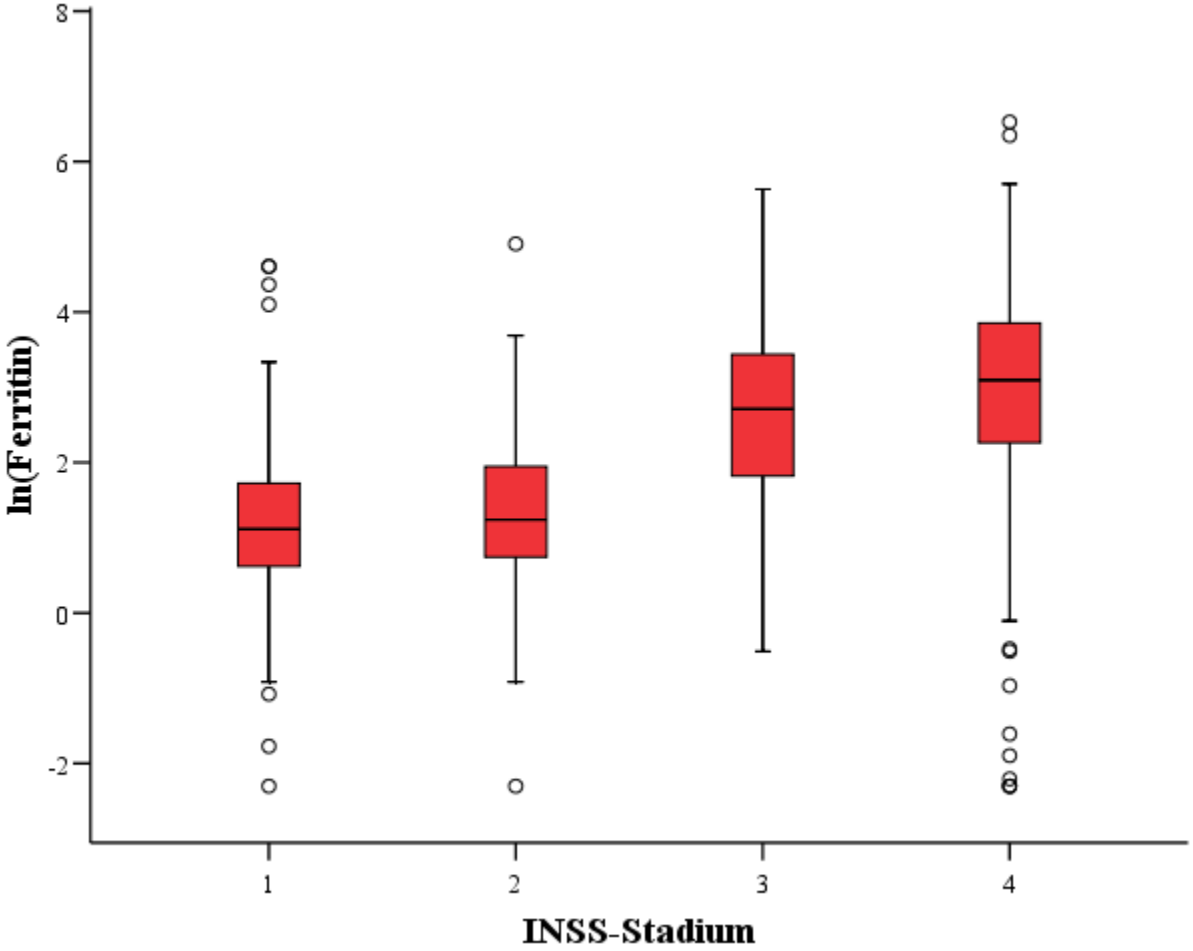
- Voraussetzungen: Gesamtfallzahl: $n \geq 60$
  Minimale erwartete Zellhäufigkeit ≥ 5 (SPSS-Output)
- Beispiel: $n = 1895$ und min. erwartete Zellhäufigkeit = 60
- Falls die Voraussetzungen nicht erfüllt sind:
  - Kategorien sinnvoll zusammenfassen
    Falls Voraussetzungen jetzt erfüllt sind: χ^2-Test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

Zusammenhang zwischen INSS und MYCN

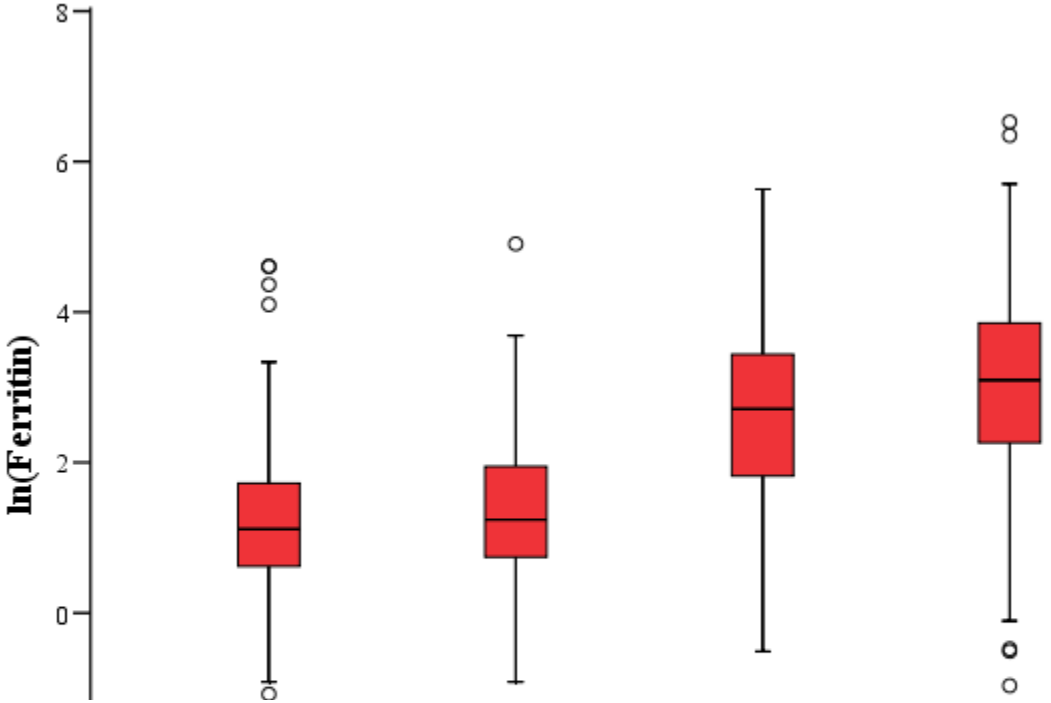
INSS-Stadium * MYCN-Amplifikation Kreuztabelle

			MYCN-Amplifikation		
			nein	ja	Gesamt
INSS-Stadium	1	Anzahl	304	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		Anzahl	1419	476	1895
		% von INSS-Stadium	74,9%	25,1%	100,0%
		% von MYCN-Amplifikation	100,0%	100,0%	100,0%

Zusammenhang zwischen INSS und ln(Ferritin)



Zusammenhang zwischen INSS und ln(Ferritin)



- |Schiefe| von ln(Ferritin) ≤ 1 in allen INSS-Stadien
- Histogramm: eingipflig in allen INSS-Stadien
- Normalverteilung in allen INSS-Stadien erscheint plausibel

Vergleich einer normalverteilten Größe zwischen mehreren Gruppen

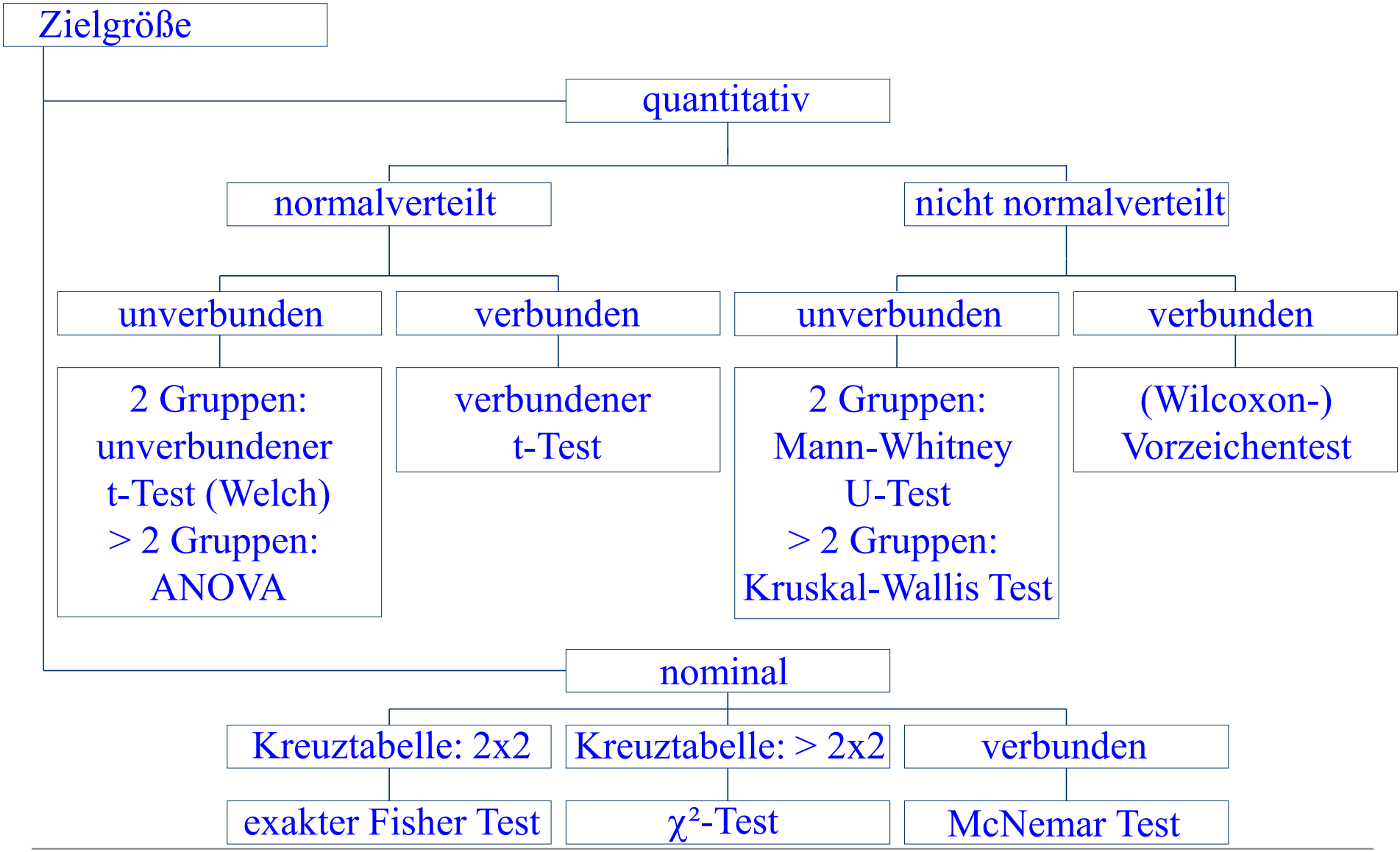
- Frage: Unterscheidet sich die logarithmierte Ferritin-Konzentration in den verschiedenen INSS-Stadien?
- Nullhypothese: Die logarithmierte Ferritin-Konzentration in den verschiedenen INSS-Stadien unterscheidet sich nicht.
- Alternativhypothese: Die logarithmierte Ferritin-Konzentration in den verschiedenen INSS-Stadien unterscheidet sich.
- Signifikanzniveau: z.B. 0,05
- 2 Gruppen: t-Test für unverbundene Stichproben nach Welch
- > 2 Gruppen: Einfaktorielle **AN**alysis **O**f **Va**riance
- Voraussetzung: Normalverteilte stetige Größe in allen Gruppen

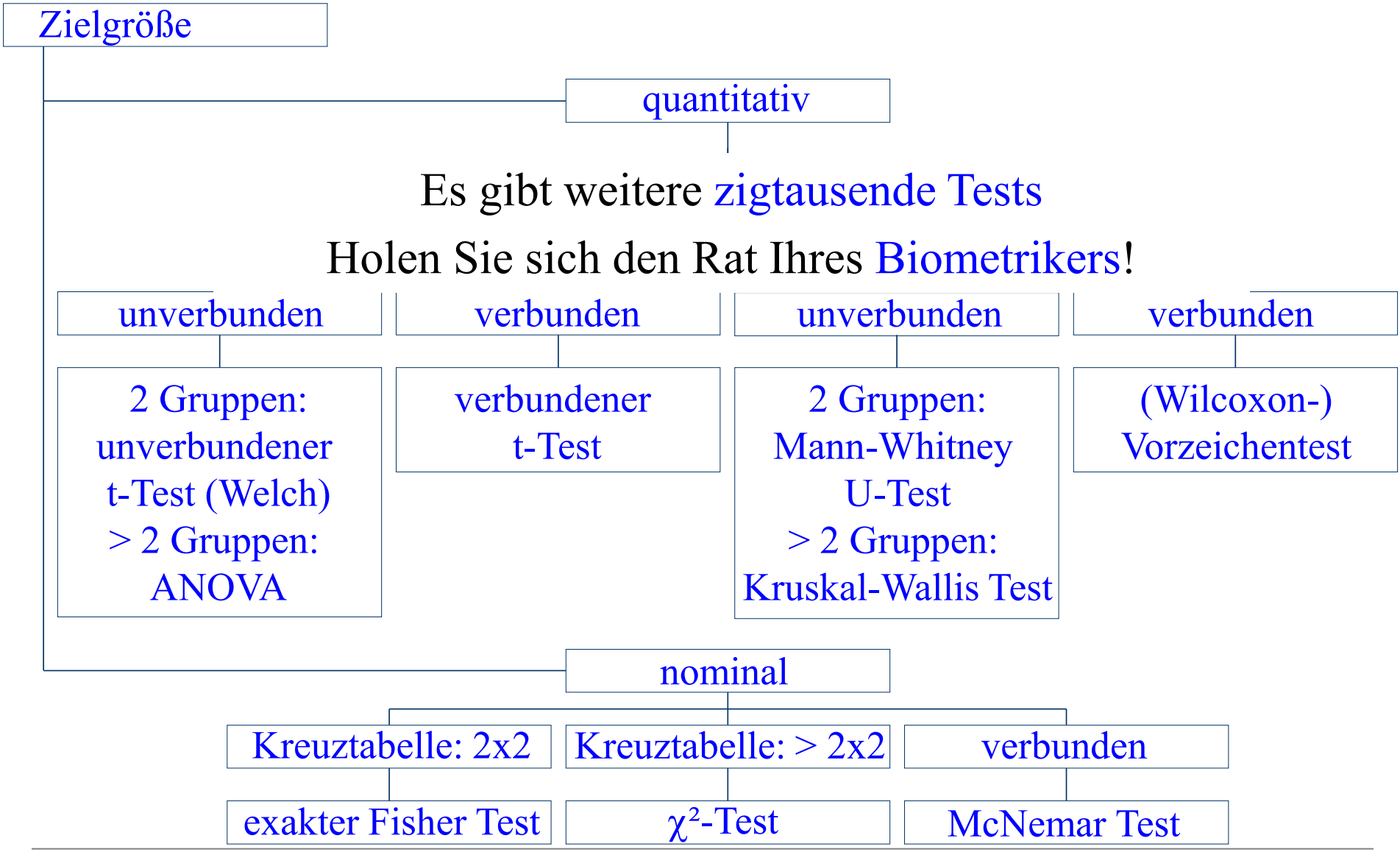
Vergleich einer normalverteilten Größe zwischen mehreren Gruppen

- P-Wert des F-Tests der einfaktoriellen ANOVA < 0,001
- Sachaussage:
  - Studienkollektiv
  - Merkmale
  - Signifikanzniveau
  - Signifikant: ja oder nein
  - Falls signifikant: Richtung der Aussage
- Die logarithmierte Ferritin-Konzentration in den verschiedenen INSS-Stadien unterscheidet sich bei Kindern und Jugendlichen mit einem Neuroblastom zum 5%-Niveau signifikant. Mit steigenden Stadien wächst die Konzentration.

Weitere statistische Tests

- Es gibt sehr viele verschiedene Tests (t-Test, Wilcoxon-Test, Fisher-Test, X^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
quantitativ
normalverteilt
nicht normalverteilt
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
Kreuztabelle: > 2x2
verbunden
exakter Fisher Test
χ²-Test
McNemar Test


Zielgröße
quantitativ
Es gibt weitere zigtausende 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
Kreuztabelle: > 2x2
verbunden
exakter Fisher Test
χ^2-Test
McNemar Test

Vergleich von Überlebenszeiten

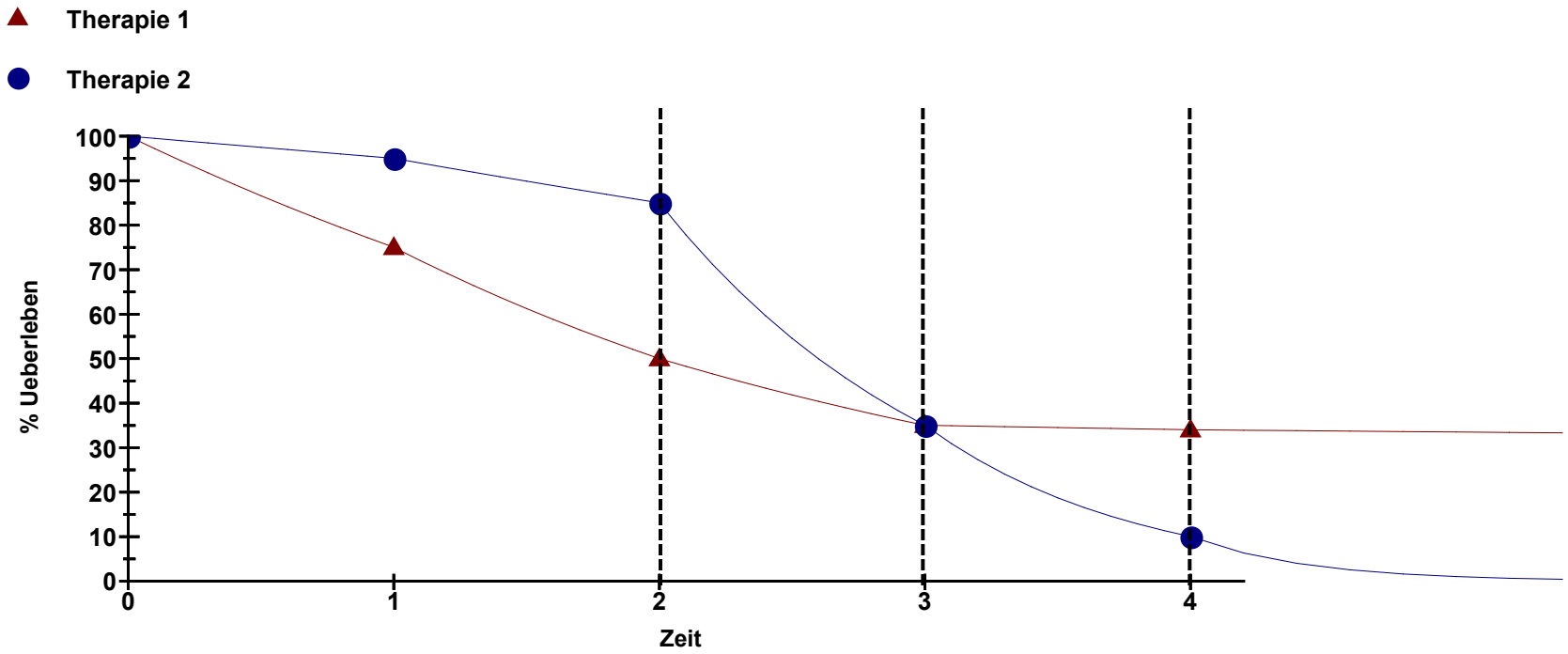
- Szenario:
  - Vergleich einer neuen Therapie mit einer Standardtherapie
  - Vergleich von Merkmalen (z. B. Geschlecht, Schweregrad)

- Studienfrage:

  Wie beeinflusst
  - die Therapie
  - das Merkmal

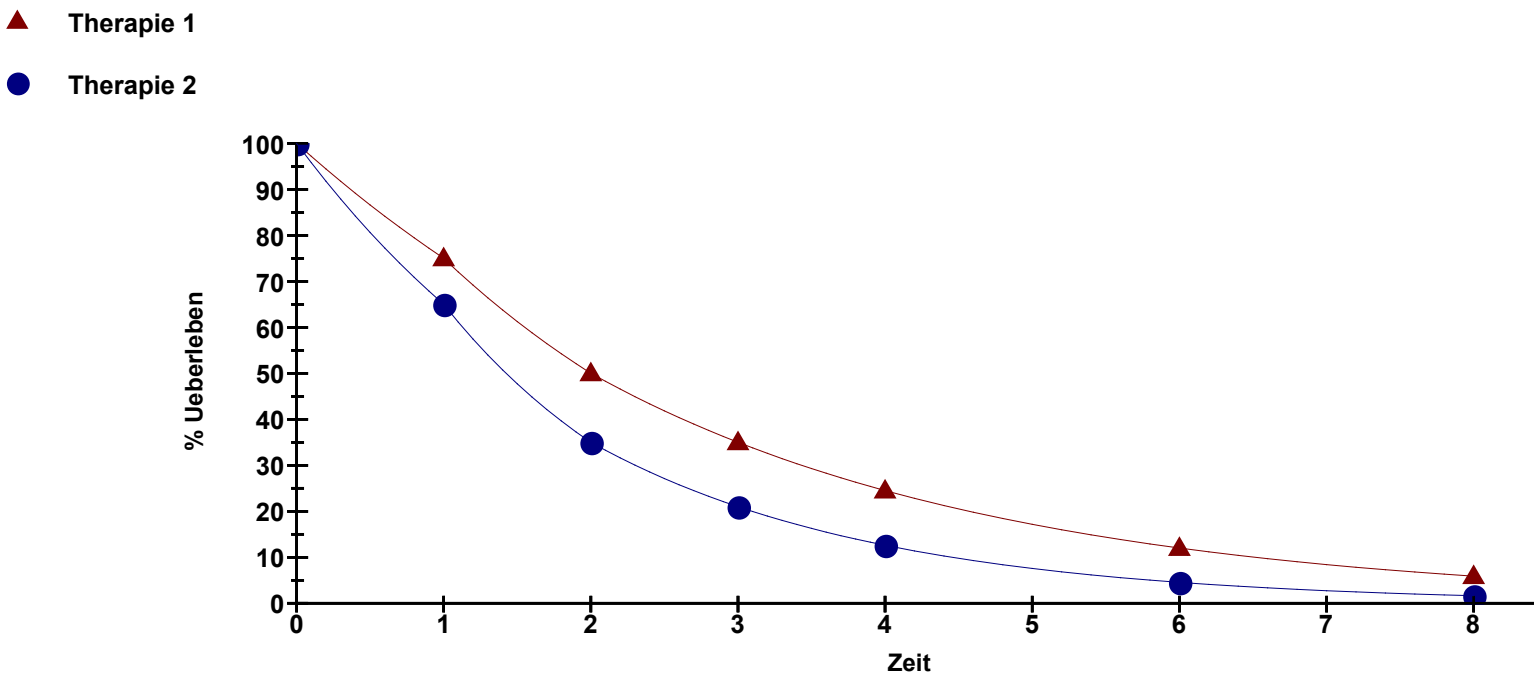
  die Überlebenszeit

Wann sind Überlebenszeiten vergleichbar?



Kein Vergleich von Äpfeln und Birnen!

Proportional Hazard



Ähnliche Form der Überlebenszeitkurven

Proportional Hazard

- Hazard: das „Risiko“ eines Ereignisses zum Zeitpunkt t

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

- Hazard Ratio zum Zeitpunkt t:

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

< 1: besser

- $hr(t) = 1 \Leftrightarrow$ die neue Therapie ist gleich gut zum Zeitpunkt t

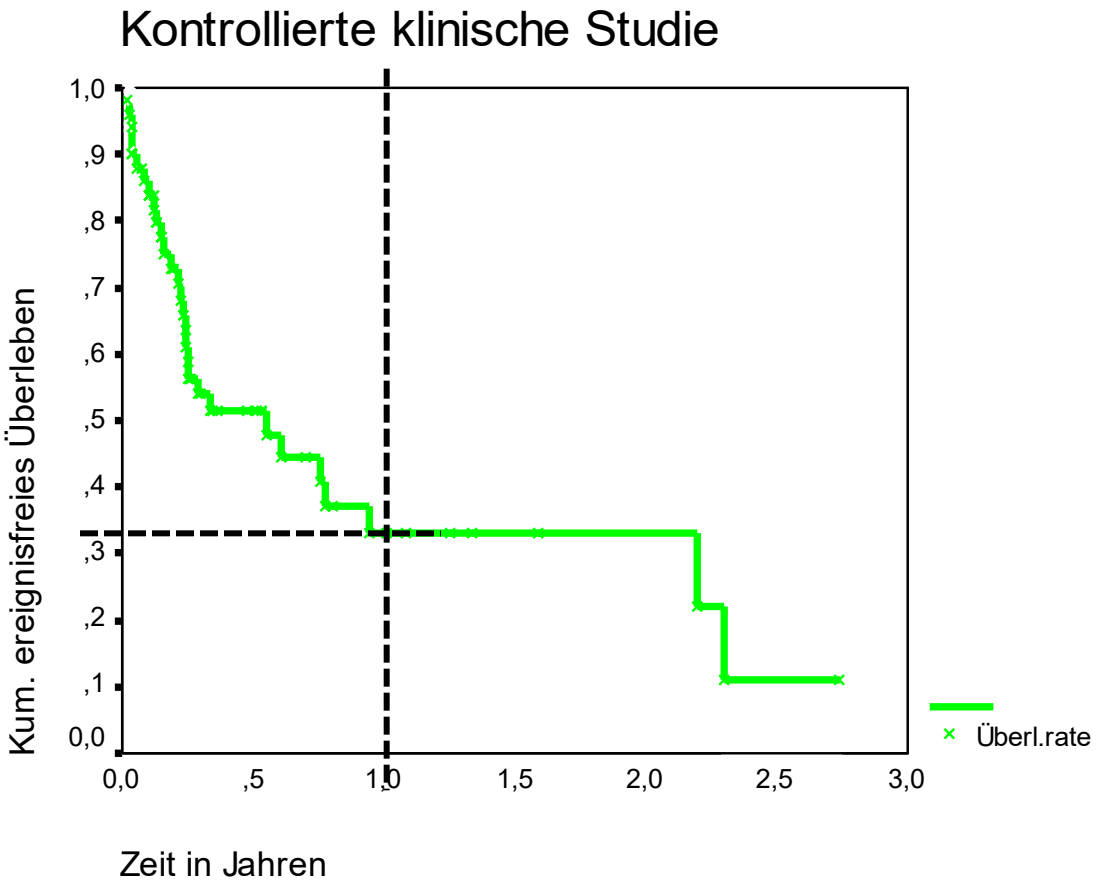
> 1: schlechter

- Proportional Hazard: $hr(t) = hr = \text{konstant}$
  die Formen der Überlebenszeitkurven sind ähnlich

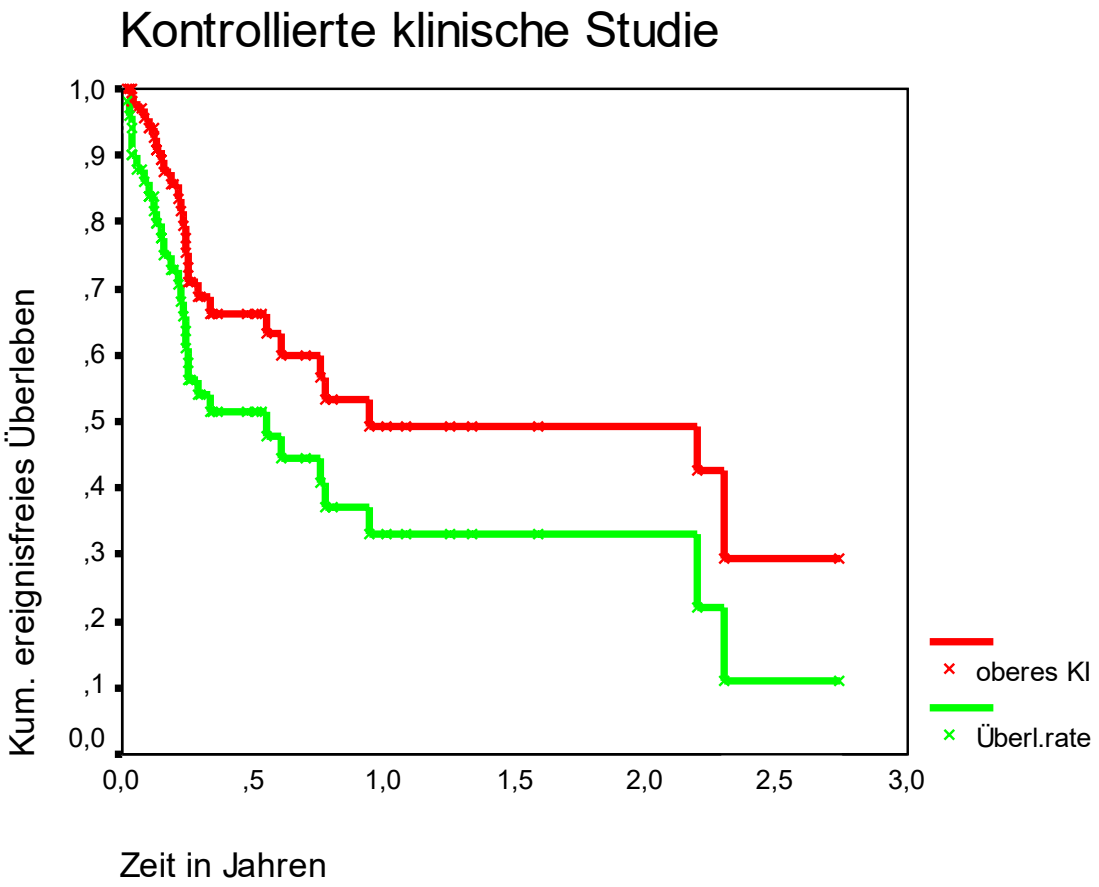
Statistische Tests

- Frage: Sind die Überlebenszeitkurven zufällig unterschiedlich?
- Königstest: Log-Rank Test
- Idee: erwartete – beobachtete Anzahl von Ereignissen
- Voraussetzung: Überlebenszeitkurven schneiden sich nicht
- Optimal: Proportional Hazard
- Alternativen:
  - verallgemeinerter Wilcoxon = Breslow Test
    Gewicht auf frühe Ereignisse
  - Tarone-Ware Test
    Kompromiss zwischen Logrank und Breslow Test

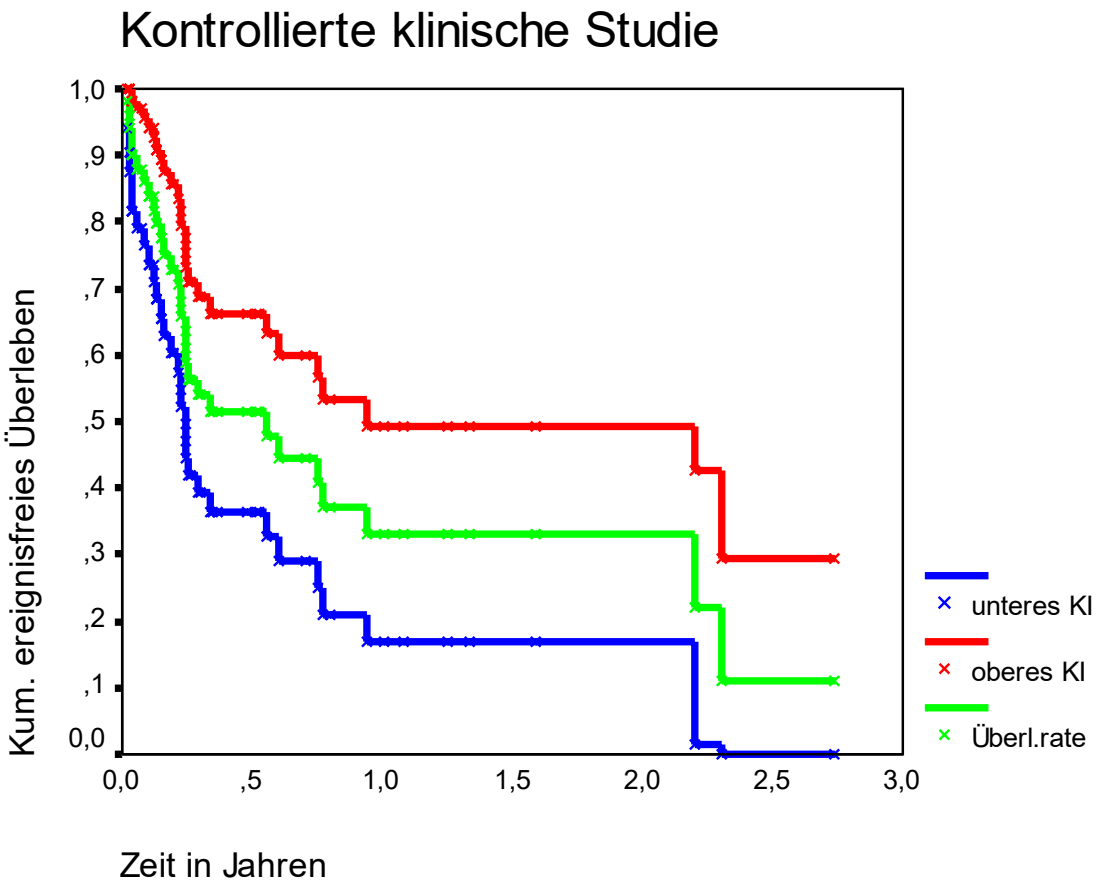
Überlebenszeitkurve nach Kaplan-Meier



Überlebenszeitkurve mit Konfidenzintervall



Überlebenszeitkurve mit Konfidenzintervall



Hypothesen und Tests – NB 2004

- Zielgröße: Ereignisfreies Überleben (EFS)
- Fragestellung: führt die intensivierte Induktionstherapie zu einem anderen EFS als die Standardinduktionstherapie?
- Zweiseitige Nullhypothese:
  die intensivierte Induktionstherapie führt zu keinem anderen EFS als die Standardinduktionstherapie.
- Alternativhypothese: die intensivierte Induktionstherapie führt zu einem anderen EFS als die Standardinduktionstherapie.
- Test: zweiseitiger Logrank-Test
- Deskription: Kaplan-Meier Kurven + KI, Quartile, EFSraten
- Software: SPSS

Vorlesungsinhalt

- Deskriptive Statistik
- Überlebenszeiten
- Statistisches Testen
  - Prinzipien
  - Spezielle Tests
  - Multiples Testen
- Fallzahlplanung / Adaptive Designs
- Regressionsmodelle
  - Logistische Regression
  - Coxregression

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	Signifikanz-niveau
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

α - Adjustierung

Problem:

Zielvariable	Signifikanz-niveau
Tod	5%
Progression	5%
Remission	5%
Schmerz	5%
Lebensqualität	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 der Besonderheiten oder Thesen
- Konfirmatorisch:
  - Beweis von Thesen
  - Cave: wenige konfirmatorische Fragestellungen, sonst sehr kleines lokales Signifikanzniveau

Haupt- und Nebenziele

- Hauptziele (konfirmatorisch):
  - rechtfertigen die Durchführung der Studie
  - Erfolg der Studie wird für diese Ziele geplant
  - 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 Hauptzielgröße: Ereignisfreies Überleben (EFS)
- Keine α-Adjustierung nötig
- Nebenfragestellungen: circa 50

Vorlesungsinhalt

- Deskriptive Statistik
- Überlebenszeiten
- Statistisches Testen
  - Prinzipien
  - Spezielle Tests
  - Multiples Testen
- Fallzahlplanung / Adaptive Designs
- Regressionsmodelle
  - Logistische Regression
  - Coxregression

Grundidee des statistischen Testens

Zu untersuchender Effekt	Studie zeigt keinen Effekt	Studie zeigt einen Effekt
In Wirklichkeit gibt es keinen Effekt	Richtige Studienaussage	Fehler 1. Art = Typ I – Fehler
In Wirklichkeit gibt es einen Effekt	Fehler 2. Art = Typ II – Fehler	Richtige Studienaussage

Wahrscheinlichkeit eines Fehlers 2. Art β:

1-β = Power: geplant 70%, 80% oder 90% mit richtiger Fallzahl

Erfolgreich mit der richtigen Zahl von Patienten

Aussageunsicherheit ~
Signifikanzniveau (5%, 1%)

Konfirmatorische Fragestellungen

Tests

Studiendesign
(Länge der Studie,...)

Erwarteter ≥
klinisch relevanter
Studieneffekt

Fallzahl

Streuung d. Studieneffekte

Gewünschte Erfolgswahrscheinlichkeit
= Power (80%, 90%)

Wann wird die Fallzahl groß?

Aussageunsicherheit ~
Signifikanzniveau

Konfirmatorische Fragestellungen

Tests

Fallzahl

Studiendesign (Länge der Studie,...)

Erwarteter ≥ klinisch relevanter Studieneffekt

Streuung d. Studieneffekte

Gewünschte Erfolgswahrscheinlichkeit

= Power

Wann wird die Fallzahl klein?

Signifikanzniveau

Konfirmatorische Fragestellungen

Tests

Studiendesign (Länge der Studie,...)

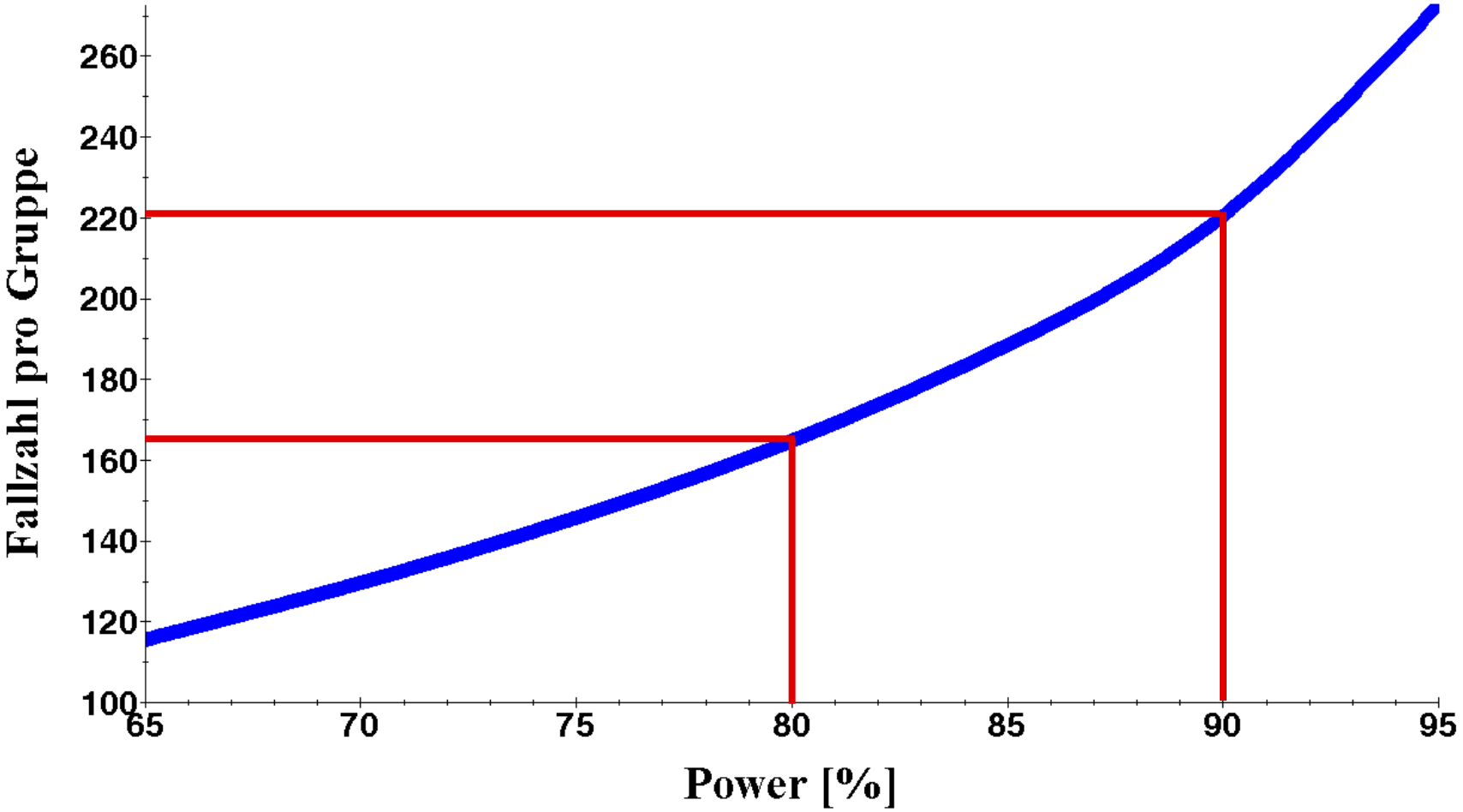
Erwarteter ≥ klinisch relevanter Studieneffekt

Fallzahl

Streuung d. Studieneffekte

Gewünschte Erfolgswahrscheinlichkeit = Power

NB 2004: Power vs. Fallzahl



Tipps für die Fallzahlplanung

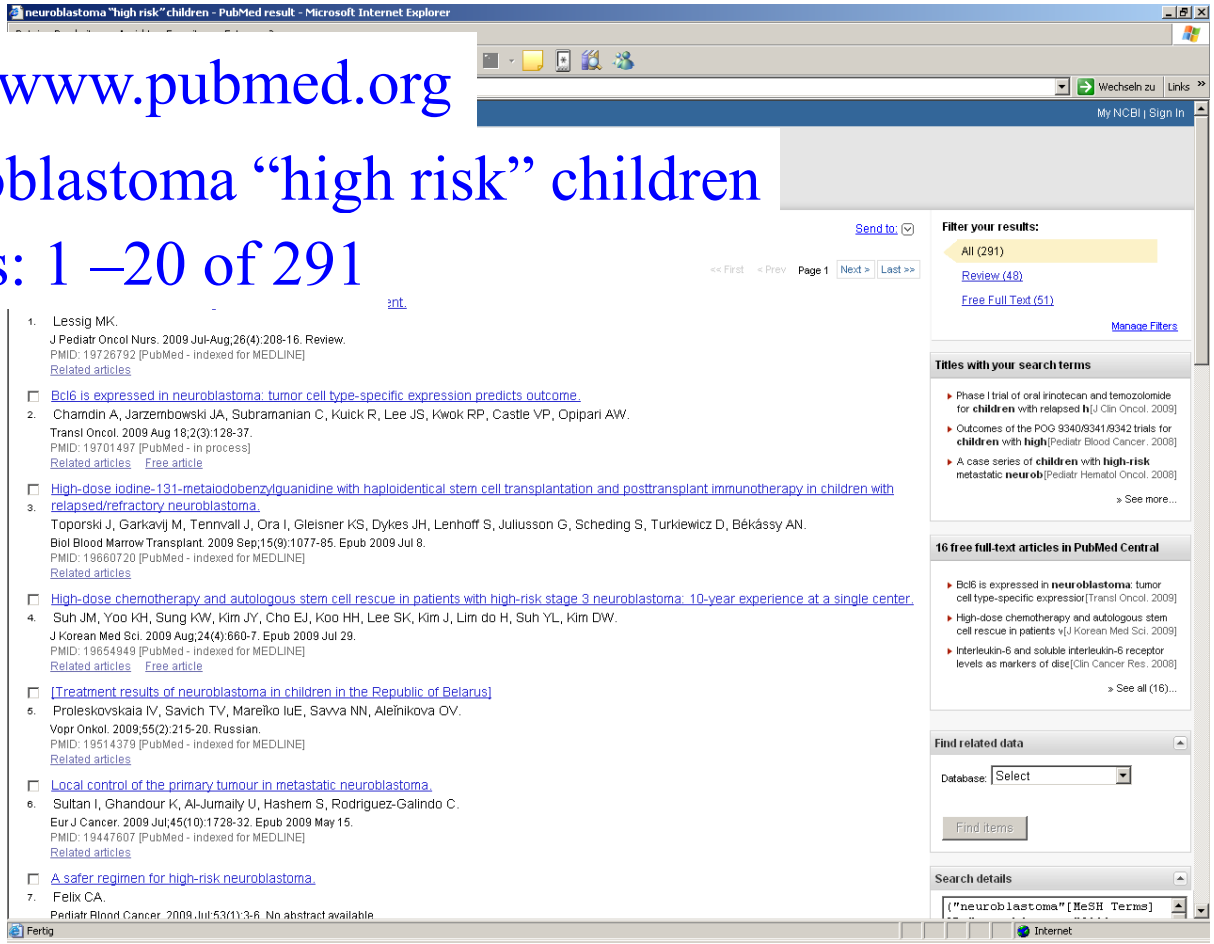
- Kontaktieren Sie Ihren Biometriker so früh wie möglich
- Lassen Sie sich Ihre benötigte Fallzahl berechnen
- Beginnen Sie keine Studie, die 1.000.000 Patienten benötigt
- Sparen Sie Zeit: warum 200 Patienten, falls 100 reichen
- Vorsicht vor zu großer Euphorie:
  Studieneffekte werden praktisch immer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

Literatursuche – PubMed

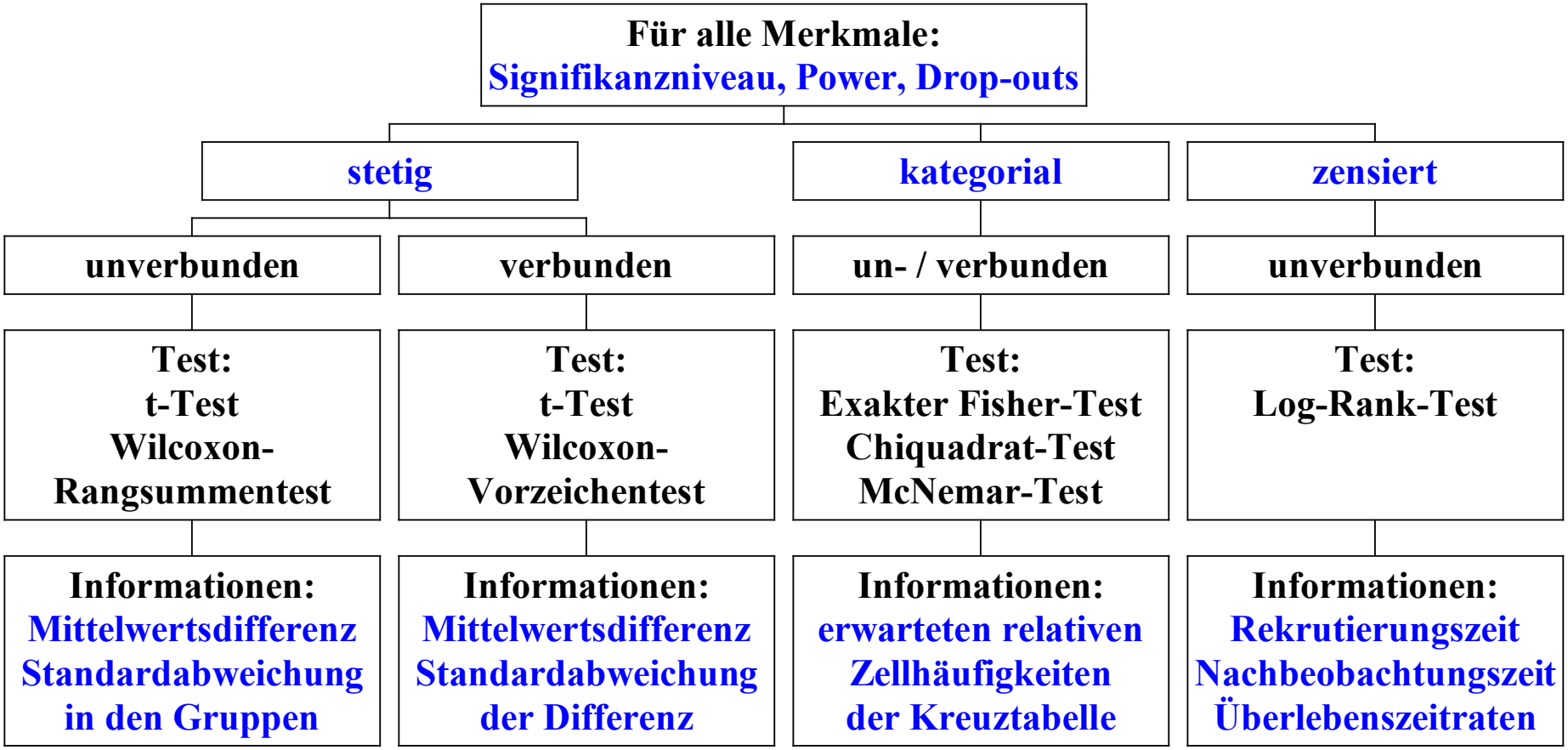
http://www.pubmed.org

neuroblastoma “high risk” children

Results: 1 –20 of 291



Wichtige Informationen für die Fallzahlplanung



Vorlesungsinhalt

- Deskriptive Statistik
- Überlebenszeiten
- Statistisches Testen
  - Prinzipien
  - Spezielle Tests
  - Multiples Testen
- Fallzahlplanung / Adaptive Designs
- Regressionsmodelle
  - Logistische Regression
  - Coxregression

Warum brauchen wir Flexibilität?

- Informationen vor Beginn der Studie
  sind häufig unsicher und verzerrt

- Zwischenanalysen sind oft notwendig bei langer Studiendauer

- Ethische Gründe: evidente Überlegenheit einer Therapie
  „… it is recognised that safety must be monitored in all trials and therefore the need for formal procedures to cover early stopping for safety reasons should always be considered.“
  (ICH Guideline, 05.02.98)

- Ökonomische Gründe
  kein Effekt mehr nachweisbar $\Rightarrow$ neuer Therapieansatz

Warum brauchen wir Regeln für Zwischenanalysen?

Mehrere Zwischenauswertungen zum Signifikanzniveau 5% ⇒
Wahrscheinlichkeit einer falschen „Signifikanz“ >> 5% !!

Zahl der Zwischenanalysen	1	2	3	4	5	10	100	1000
Tatsächliches Signifikanzniveau	5%	8.4%	10.7%	12.6%	14.1%	19.4%	37.4%	52.8%

Warum brauchen wir Regeln für Zwischenanalysen?

Mehrere Zwischenauswertungen zum Signifikanzniveau 5% ⇒
Wahrscheinlichkeit einer falschen „Signifikanz“ >> 5% !!

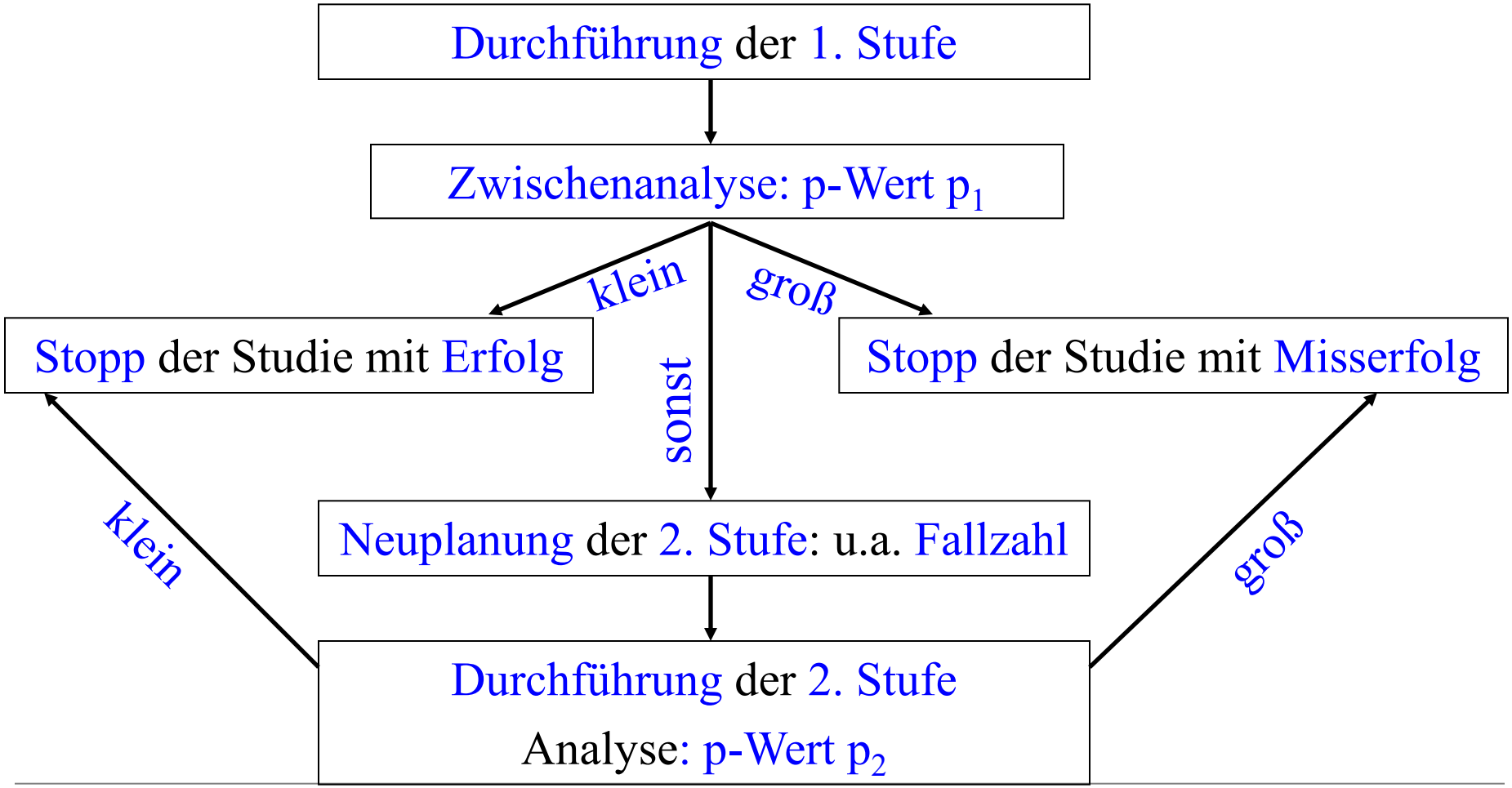
Zahl der Zwischenanalysen	1	2	3	100	1000
Tatsächliches Signifikanzniveau	5%			14.1% 19.4% 37.4%	52.8%

Inflation des Signifikanzniveaus

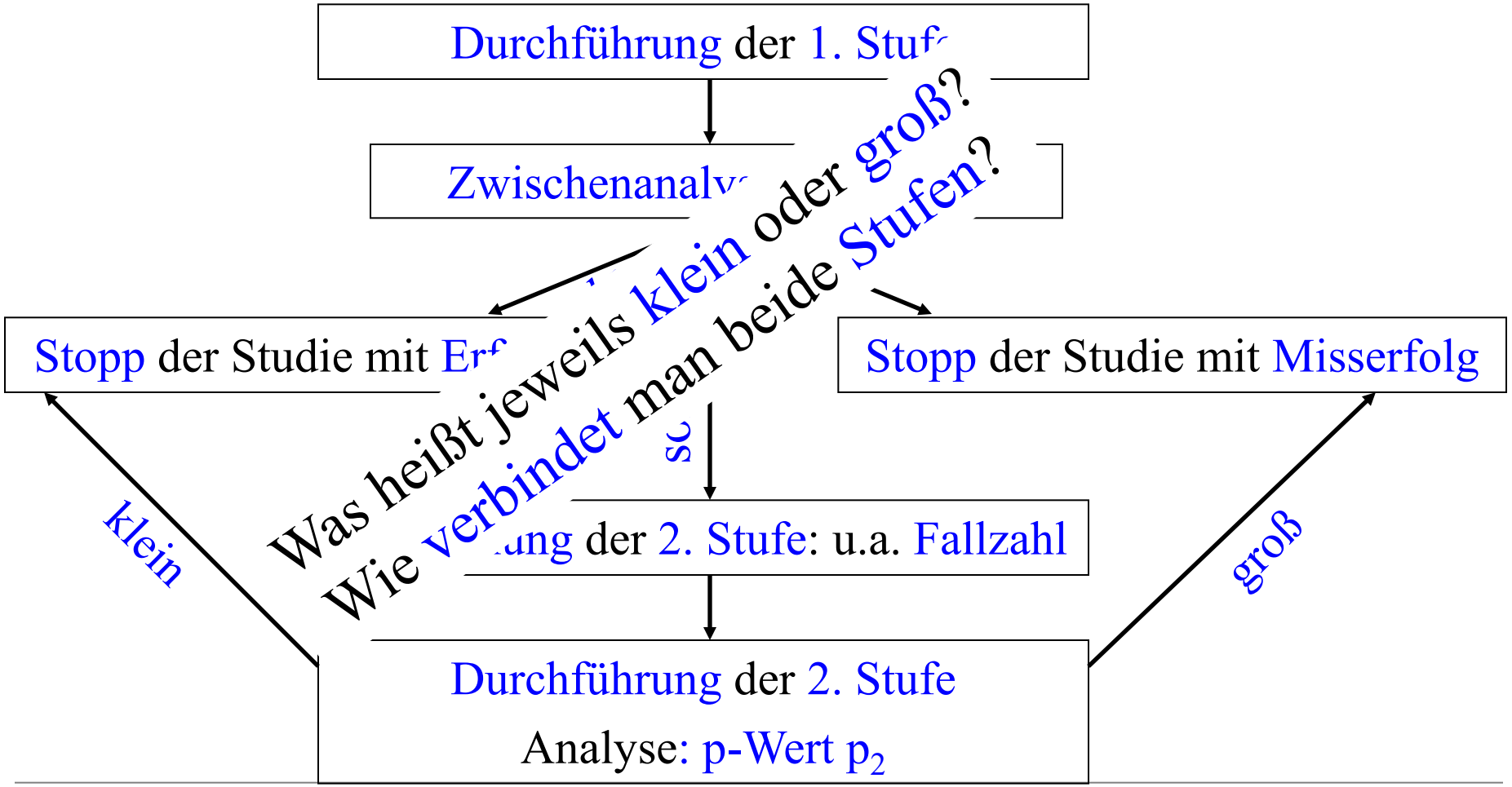
Typen von Zwischenanalysen

- 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
  - Ergebnisse der bisherigen Studie werden genutzt

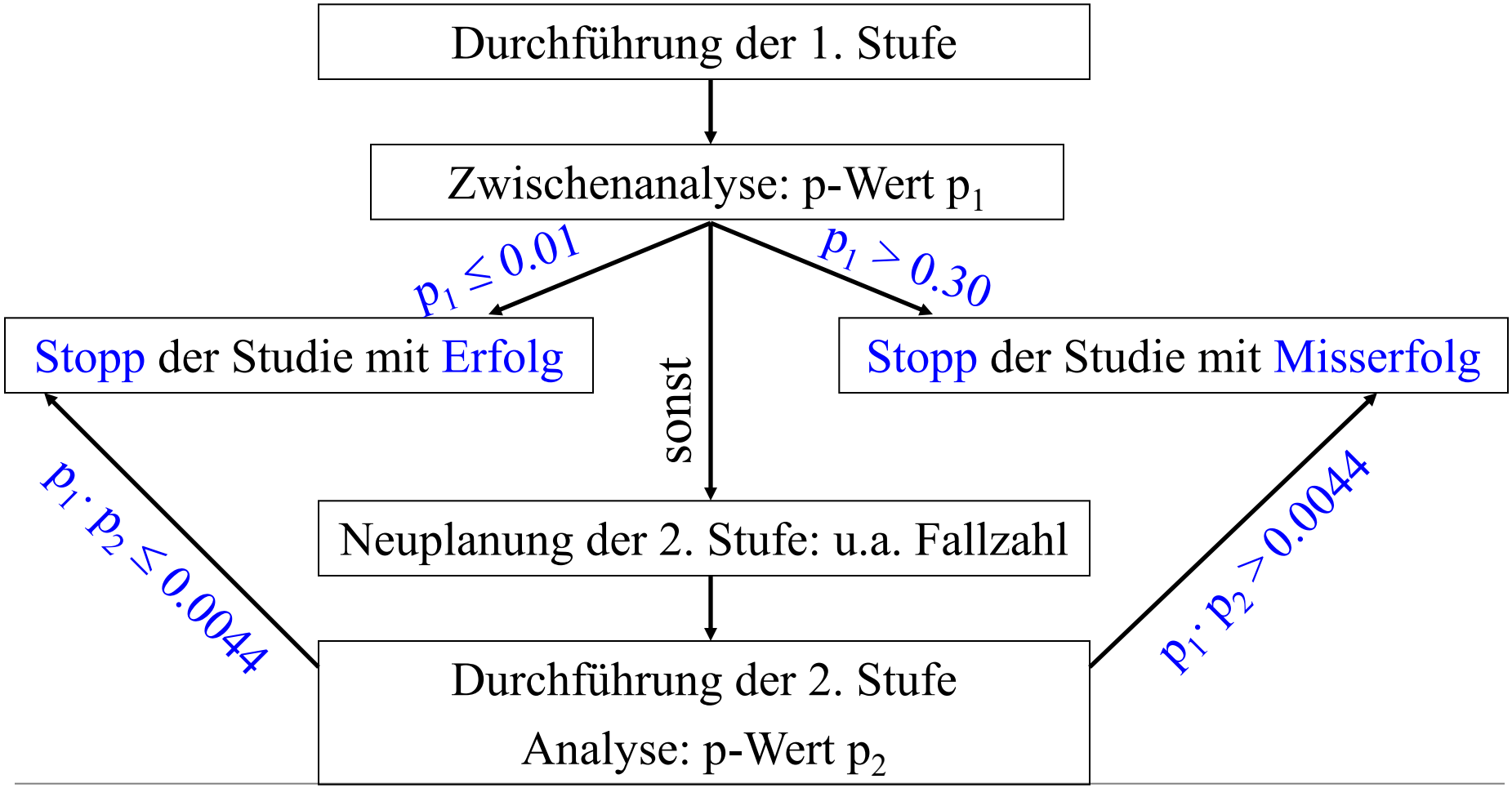
Prinzip einer adaptiv-sequentiellen Studie



Prinzip einer adaptiv-sequentiellen Studie



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



Was können adaptive Designs leisten?

- Beendigung der Studie, falls das Studienziel erreicht wurde
- Beendigung der Studie, falls die Weiterführung zwecklos ist
- Anpassung der Fallzahl an beobachtete Studieneffekte
- Anpassung des Studiendesigns an neue Erkenntnisse
- Anpassung der Studienziele an neue Erkenntnisse
- Flexible Zeitpunkte der Analyse
- Flexibles Studienende
- Schnellere Ergebnisse im Mittel
- Geringere Fallzahlen bzw. höhere Power im Mittel

Im schlimmsten Fall höhere Fallzahlen nötig bei kleinem Studieneffekt

Was können adaptive Designs leisten?

- Beendigung der Studie, falls das Studien-
- Beendigung der Studie, falls die W
- Anpassung der Fallzahl an h
- Anpassung des Studie
- Anpassung der S
- Flexible Z
- Fle
- gebnisse im Mittel

re Fallzahlen bzw. höhere Power im Mittel

Im s mmsten Fall höhere Fallzahlen 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 kürzere Studienzeit
- Möglichkeit das gesamte Studiendesign während der Studie jederzeit adaptiv zu verändern

Vorlesungsinhalt

- Deskriptive Statistik
- Überlebenszeiten
- Statistisches Testen
  - Prinzipien
  - Spezielle Tests
  - Multiples Testen
- Fallzahlplanung / Adaptive Designs
- Regressionsmodelle
  - Logistische Regression
  - Coxregression

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
  - Mammaca.: 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

- Beispiel NB 2004: hängt log(Ferritin) von log(LDH) ab
- Allgemein:
  - Y = abhängige Variable = Zielgröße (z. B. log(Ferritin))
  - X = unabhängige Variable = Einflussgröße (z. B. log(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

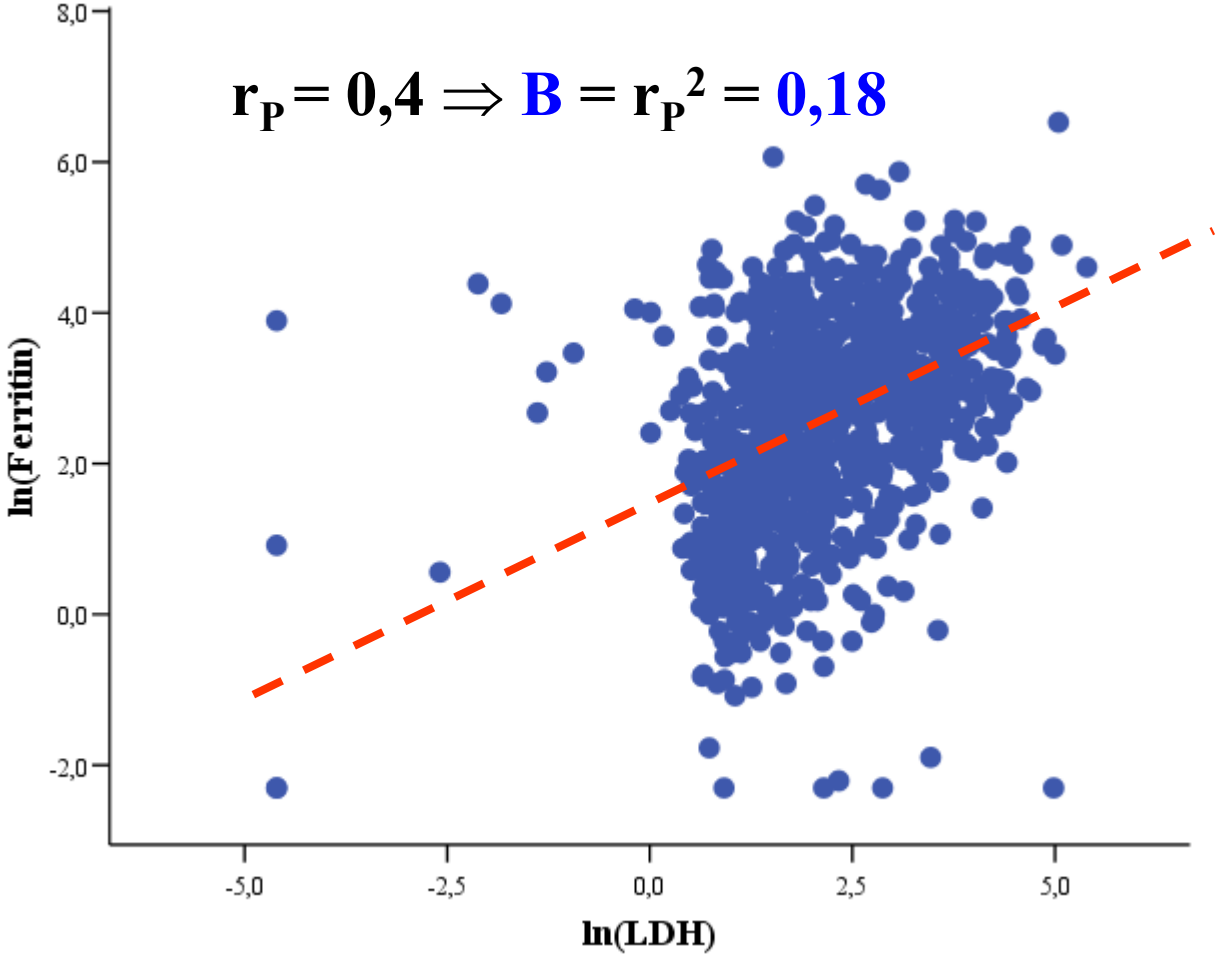
Einfache lineare Regression

- 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(Ferritin) [log(ng/ml)] =
  1,4 log(ng/ml) + 0,5 log(ng/ml)/log(U/l) · log(LDH) [log(U/l)]
- Interpretation des Achsenabschnitts 1,4 log(ng/ml):
  bei einem log(LDH)-Wert von 0 log(U/l) wird
  ein log(Ferritin)-Wert von 1,4 log(ng/ml) erwartet
- Interpretation der Steigung 0,5 log(ng/ml)/log(U/l):
  steigt der log(LDH)-Wert um 1 log(U/l),
  so steigt der erwartete log(Ferritin)-Wert um 0,5 log(ng/ml)

Einfache lineare Regression

- Einheiten nicht vergessen
- Keine Extrapolationen: Regressionsgleichung nur gültig für Wertebereich -4,6 bis 5,4 log(U/l) von log(LDH)
- Frage: Wie stark ist der Zusammenhang zwischen X und Y
- Antwort: Bestimmtheitsmaß B
  - $B = 1 - \Sigma(Y_i - a - b \cdot X_i)^2 / \Sigma(Y_i - Y_M)^2$; 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

Zusammenhang zwischen ln(LDH) und ln(Ferritin)



Bestimmtheitsmaß:

$0 \leq B \leq 1$

Korrelationskoeff.:

$-1 \leq r_P \leq 1$

Steigung:

$-\infty < b < \infty$

$b < 0 \Leftrightarrow r_P < 0$

$b > 0 \Leftrightarrow r_P > 0$

Vorlesungsinhalt

- Deskriptive Statistik
- Überlebenszeiten
- Statistisches Testen
  - Prinzipien
  - Spezielle Tests
  - Multiples Testen
- Fallzahlplanung / Adaptive Designs
- Regressionsmodelle
  - Logistische Regression
  - Coxregression

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)

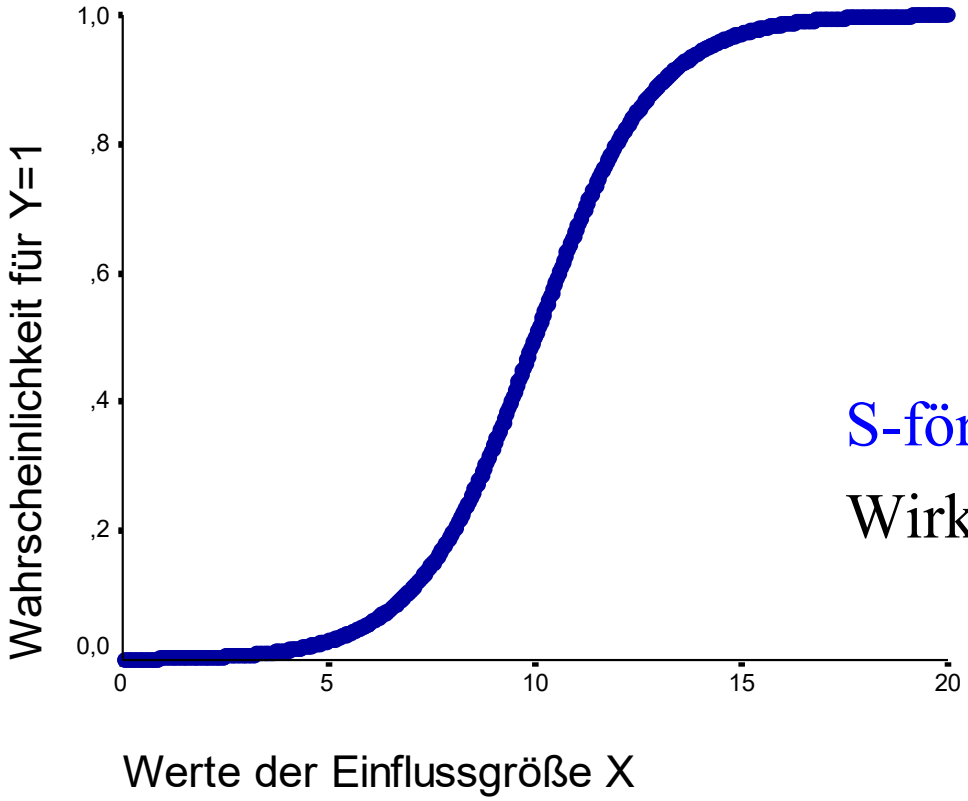
Fragestellung

- 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

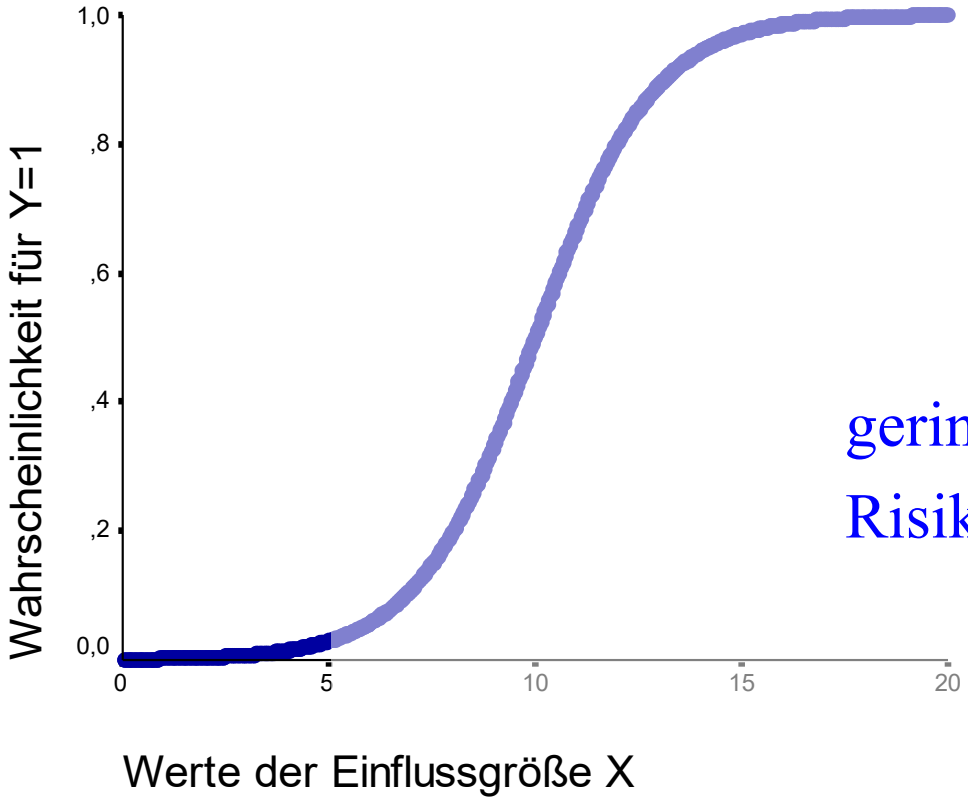
- Das Modell soll die Ursache-Wirkungsbeziehung zwischen der Ziel- und den Einflussgrößen gut beschreiben
- Mehrere Möglichkeiten
- Populärste: logistische Transformation
- Wahrscheinlichkeit $\rightarrow$ Logarithmus des Odds (Chance)
- $\text{logit}(p) = \log(\text{Odds}) = \log(p/(1-p)) = \beta_0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



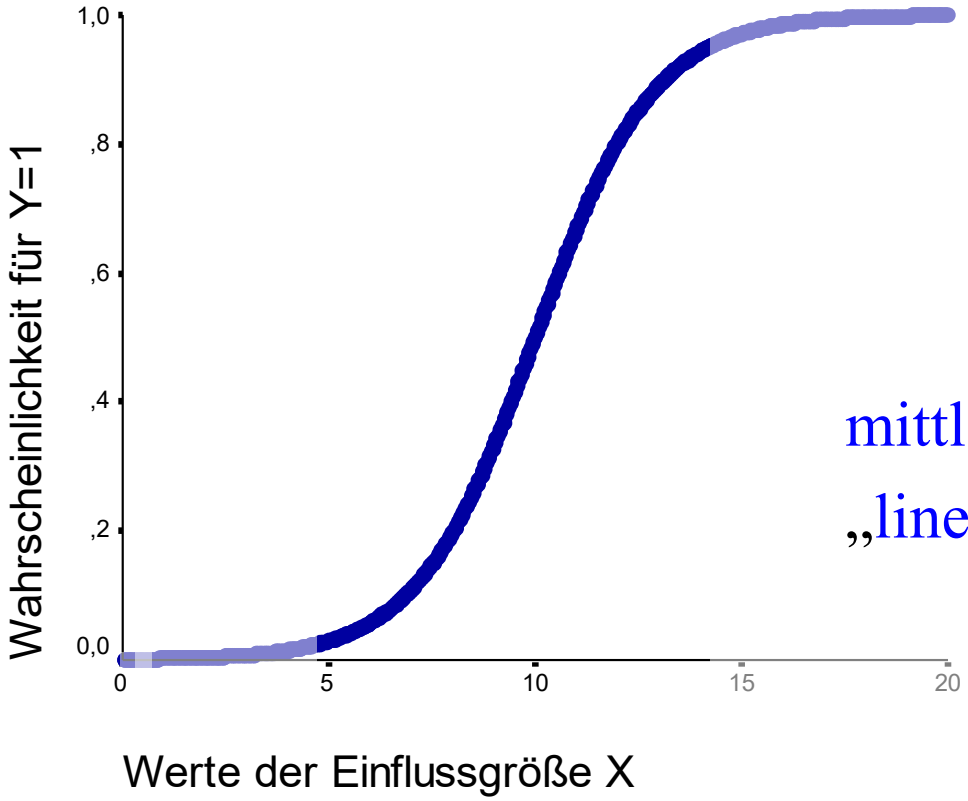
S-förmige Ursache-Wirkungs-Beziehung

Logistische Ursache-Wirkungs-Beziehung



geringe Exposition →
Risiko kaum erhöht

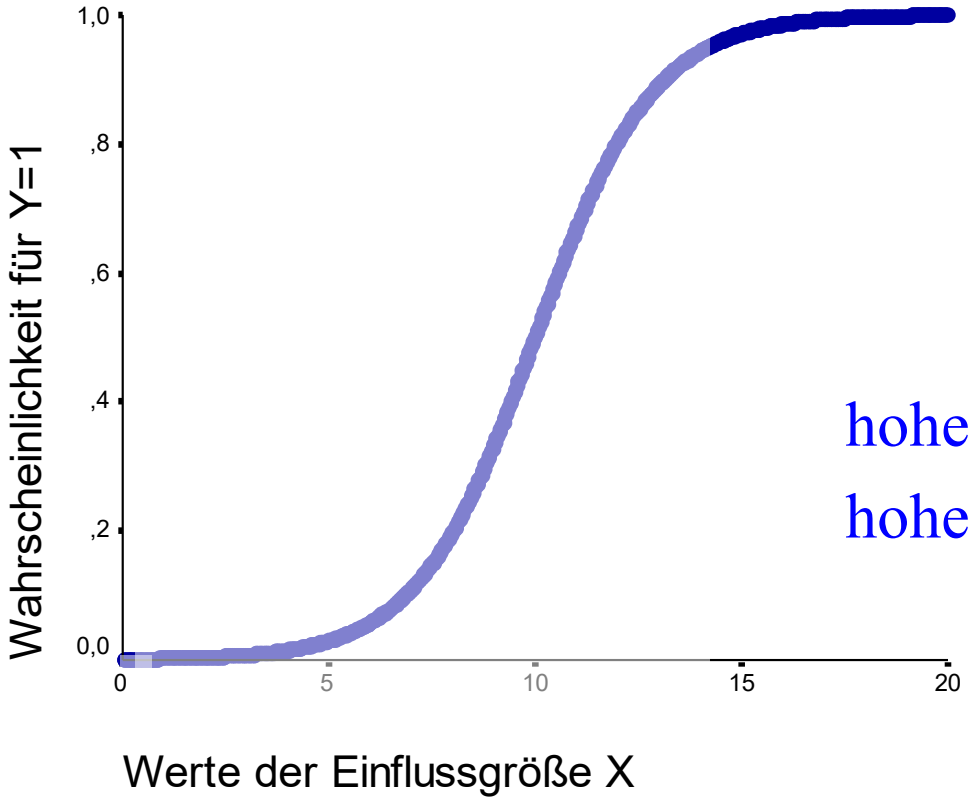
Logistische Ursache-Wirkungs-Beziehung



mittlere Exposition →

„linearer“ Risikoanstieg

Logistische Ursache-Wirkungs-Beziehung



hohe Exposition →
hohes Risiko

Interpretation der Regressionsparameter

- X eine Einflussgröße
- p die Wahrscheinlichkeit für ein Ereignis
- logit(p) = log(p/(1-p)) = log(Odds) = $\beta_0 + \beta_1 X$
- Odds = p/(1-p) = exp($\beta_0 + \beta_1 X$)
- Das Odds verändert sich um Faktor exp(β_1),
  falls x um eine Einheit steigt – unabhängig von x:

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

Interpretation der Regressionsparameter

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

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

Welche Einflussgrößen sind wichtig?

- Logistisches Modell:
  Ereigniswahrscheinlichkeit ↔ log(Odds) ← 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 nur einer Einflussgröße → Fallausschluss
- Daher: nur Einflussgrößen einschließen, die zu einer relevanten Erhöhung der Wahrscheinlichkeit der Daten führen
- Frage: Was heißt relevant?

Modellwahl - Beispiel

- 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)

Variablen in der Gleichung

		Regressionskoeffizient B	Standardfehler	Wald	df	Sig.	Exp(B)	95,0% Konfidenzintervall für EXP(B)	
								Unterer Wert	Oberer Wert
Schritt 1[a]	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.

β_i $s.e.(\beta_i)$ $OR(x_i+1,x_i)$

Einflussgrößen

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

Modellbildung – Beispiel (Prostatakarzinom)

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[a]	GRADE(1)	1,181	,596	3,934	1	,047	3,259	1,014	10,475
	Konstante	-,981	,391	6,297	1	,012	,375		

a. In Schritt 1 eingegebene Variablen: GRADE.

Grading allein = Einflussgröße multipel: Grading hat keinen Einfluss

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[a]	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.

A priori Bewertung von Einflussgrößen

1. Klinisch etablierte Einflüsse, Confounder → Berücksichtigung im Modell: immer
2. Einflüsse mit plausibel vermutetem kausalen 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

		Regressionskoeffizient B	Standardfehler	Wald	df	Sig.	Exp(B)	95,0% Konfidenzintervall für EXP(B)	
								Unterer Wert	Oberer Wert
Schritt 1[a]	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[a]	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[a]	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[a]	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

		Regressionskoeffizient B	Standardfehler	Wald	df	Sig.	Exp(B)	95,0% Konfidenzintervall für EXP(B)	
								Unterer Wert	Oberer Wert
Schritt 1[a]	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[b]	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

Odds Ratio bei multivariabler logistischer Regression

- Bisher wurde die Änderung nur einer Einflussgröße betrachtet
- Frage: Wie ändert sich das Odds bei gleichzeitige Änderung mehrerer Einflussgrößen?
- Seien $X_1,...X_k$ Einflussgrößen (Original, Dummies, Transform.)
- $\log(\text{Odds}) = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k$
- OR 2er Patienten mit $(x_{11}, x_{21}, \ldots, x_{k1})$ und $(x_{12}, x_{22}, \ldots, x_{k2})$:

$$\text{OR}(x_2, x_1) = \frac{\text{Odds}(x_2)}{\text{Odds}(x_1)} = \frac{\exp(\beta_0 + \beta_1 x_{12} + \cdots + \beta_k x_{k2})}{\exp(\beta_0 + \beta_1 x_{11} + \cdots + \beta_k x_{k1})} =$$

$$\exp(\beta_1[x_{12} - x_{11}] + \cdots + \beta_k[x_{k2} - x_{k1}]) = \prod_{x_{i2} \neq x_{i1}} \exp(\beta_i)^{\lfloor x_{i2} - x_{i1} \rfloor}$$

Beispiel einer multivariabler logistischen Regression

- Prostatakarzinom - Lymphknoten (B.W. Brown 1980):
  Einflussgrößen: X = XRAY: Röntgenbefund (negativ/positiv),
  S = SIZE: Tumorgröße (klein/groß)
- SPSS: $\log(\text{Odds}) = -2{,}0 + 2{,}1 \cdot X + 1{,}6 \cdot S$
- OR 2er Patienten: (Röntgenbefund: negativ, Tumorgröße: groß)$_1$
  versus (Röntgenbefund: positiv, Tumorgröße: klein)$_2$
- $OR = \prod_{x_{i2} \neq x_{i1}} \exp(\beta_i)^{|x_{i2} - x_{i1}|} = \exp(2{,}1)^{[1-0]} \cdot \exp(1{,}6)^{[0-1]} = 1{,}70$
- Die Chance auf einen LK-Befall ist bei Patienten 2
  im Vergleich zu Patient 1 um 70% höher
- Vorteil kleiner Tumorgröße << Vorteil negativer Röntgenbefund

Modellbildung - Bewertung

- Viele Auswahlverfahren für Modelle existieren
- Auswahlverfahren werden oft kontrovers diskutiert
- Automatisierte Verfahren:
  - wenig akzeptiert
  - 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
- Überlebenszeiten
- Statistisches Testen
  - Prinzipien
  - Spezielle Tests
  - Multiples Testen
- Fallzahlplanung / Adaptive Designs
- Regressionsmodelle
  - Logistische Regression
  - Coxregression

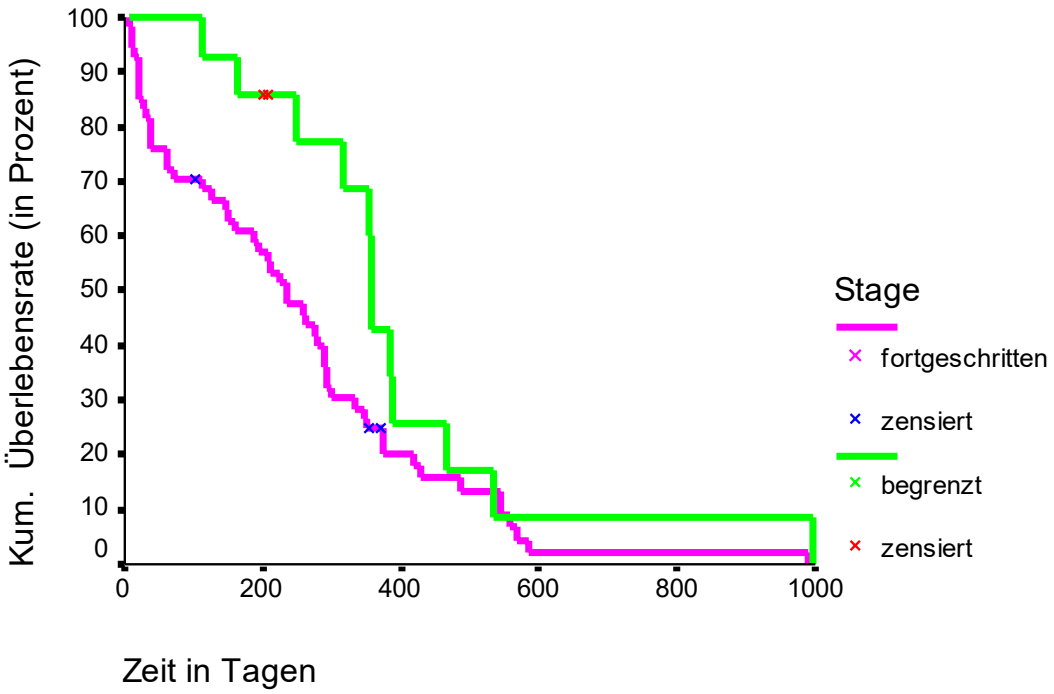
Explorative multivariable Überlebenszeitanalyse

- 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

Patienten mit kleinzelligem Bronchialkarzinom

ohne Chemotherapieresponse



Mediane Ü-zeit (Tage):

fortgeschritten:	233
begrenzt:	358

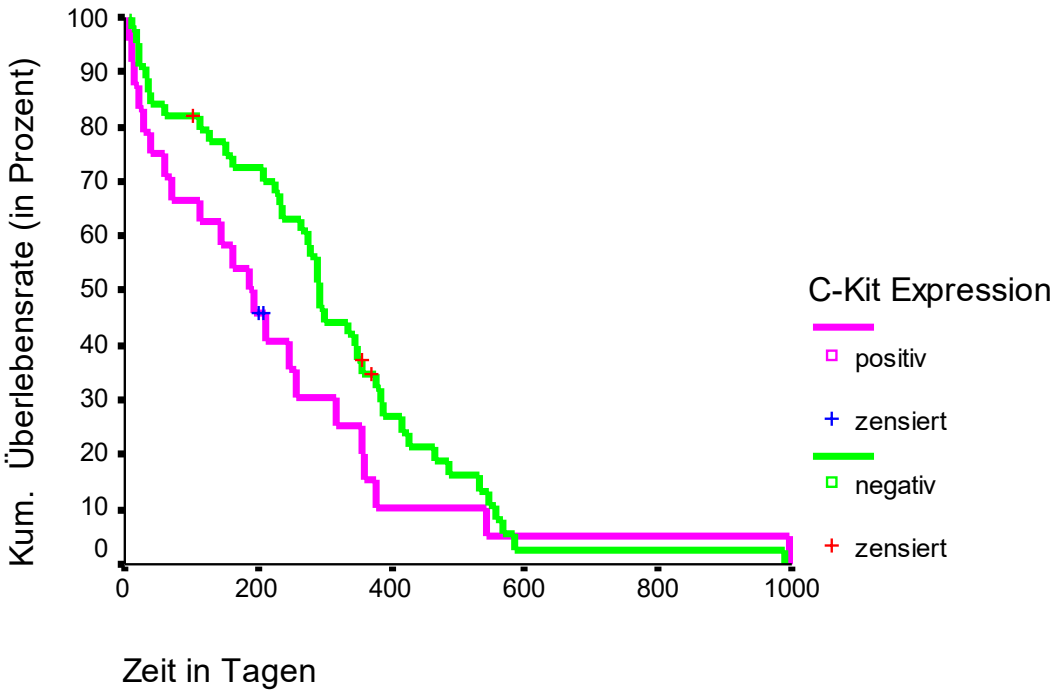
p-Werte

Log-Rank:	0,076
Breslow:	0,021
Tarone:	0,035

Einfluss der C-Kit Expression auf die Überlebenszeit

Patienten mit kleinzelligem Bronchialkarzinom

ohne Chemotherapieresponse



Mediane Ü-zeit (Tage):

positiv: 187

negativ: 291

p-Werte

Log-Rank:	0,1588
Breslow:	0,0390
Tarone:	0,0532

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) ≅ Log-Rank Test

Gemeinsamer Einfluss von Tumorstadium und C-Kit Expression

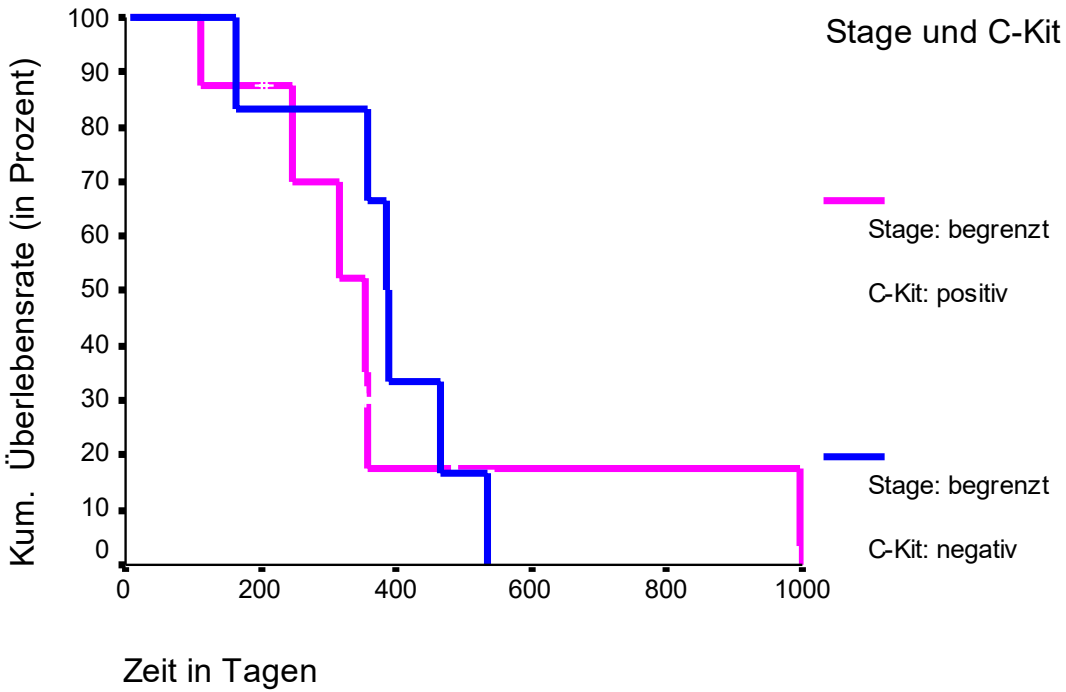
Finales Modell der Coxregression			
frag. Einfluss	p-Wert	Hazard Ratio	95%-KI
Stadium	0,262	ohne Einfluss	entfällt
C-Kit	0,314	ohne Einfluss	entfällt
Wechselw.	0,002	2,7	[1,5; 4,9]

- erklärender Einfluss: nur Wechselwirkung
- 2,7faches Hazard (Risiko) bei fortgeschrittenem Tumorstadium und gleichzeitiger positiver C-Kit Expression

Gemeinsamer Einfluss von Tumorstadium und C-Kit Expression

Patienten mit kleinzelligem Bronchialkarzinom

ohne Chemotherapieresponse



Coxregression

Erklärender Einfluss:

Fortgeschrittenes Tumorstadium

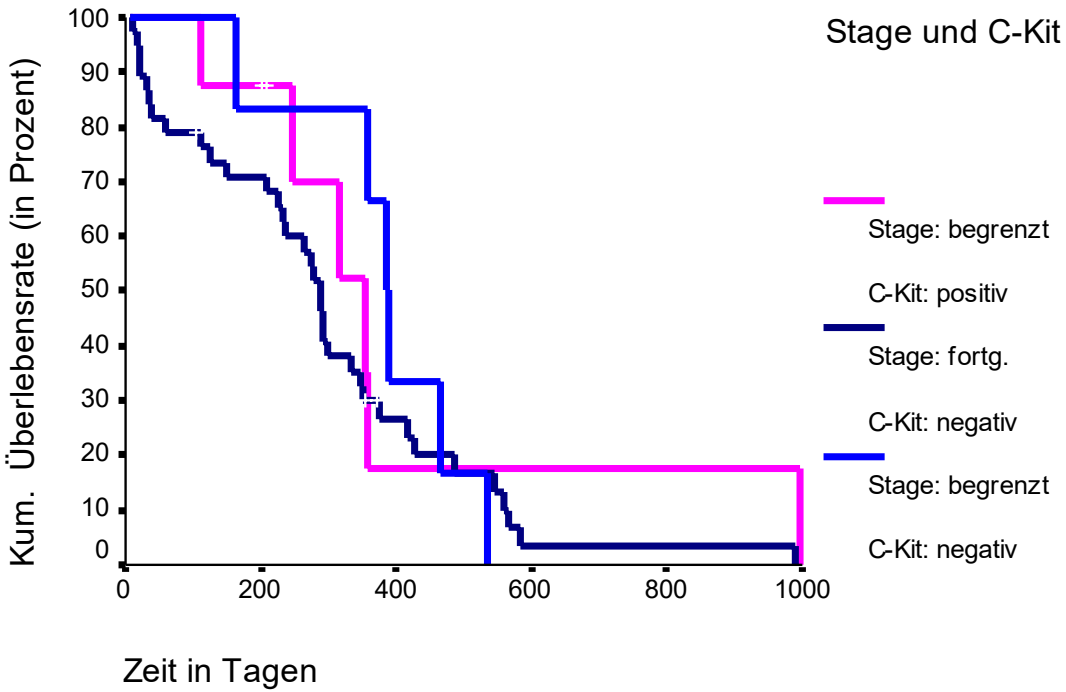
und gleichzeitig

positive C-Kit Expression

Gemeinsamer Einfluss von Tumorstadium und C-Kit Expression

Patienten mit kleinzelligem Bronchialkarzinom

ohne Chemotherapieresponse



Coxregression

Erklärender Einfluss:

Fortgeschrittenes Tumorstadium

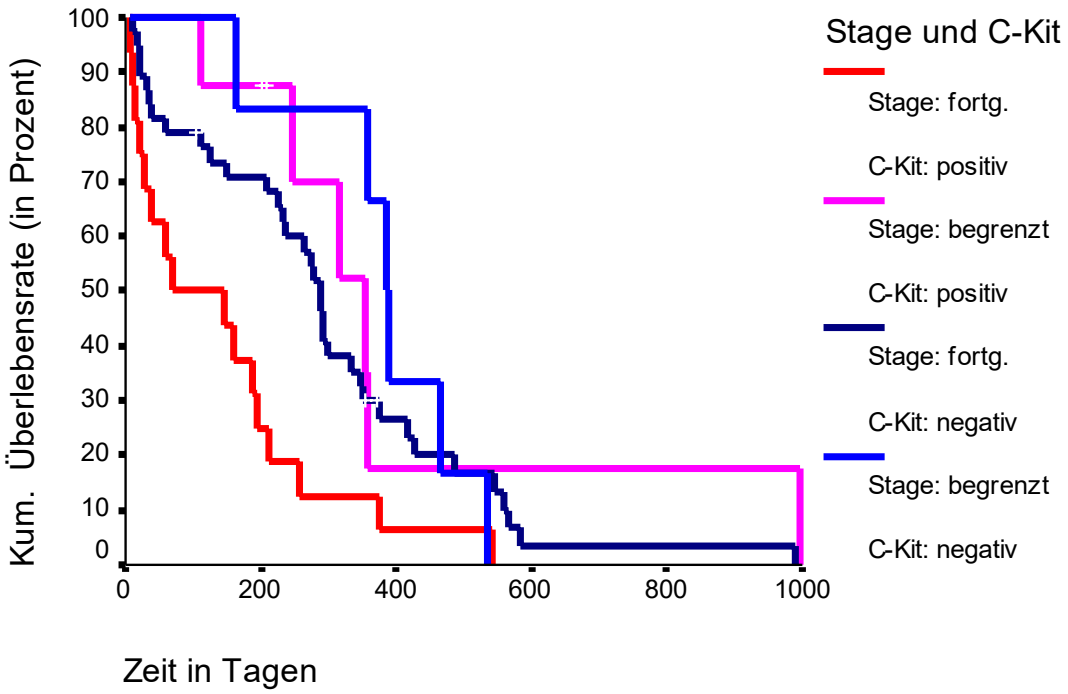
und gleichzeitig

positive C-Kit Expression

Gemeinsamer Einfluss von Tumorstadium und C-Kit Expression

Patienten mit kleinzelligem Bronchialkarzinom

ohne Chemotherapieresponse



Coxregression

Erklärender Einfluss:

Fortgeschrittenes Tumorstadium

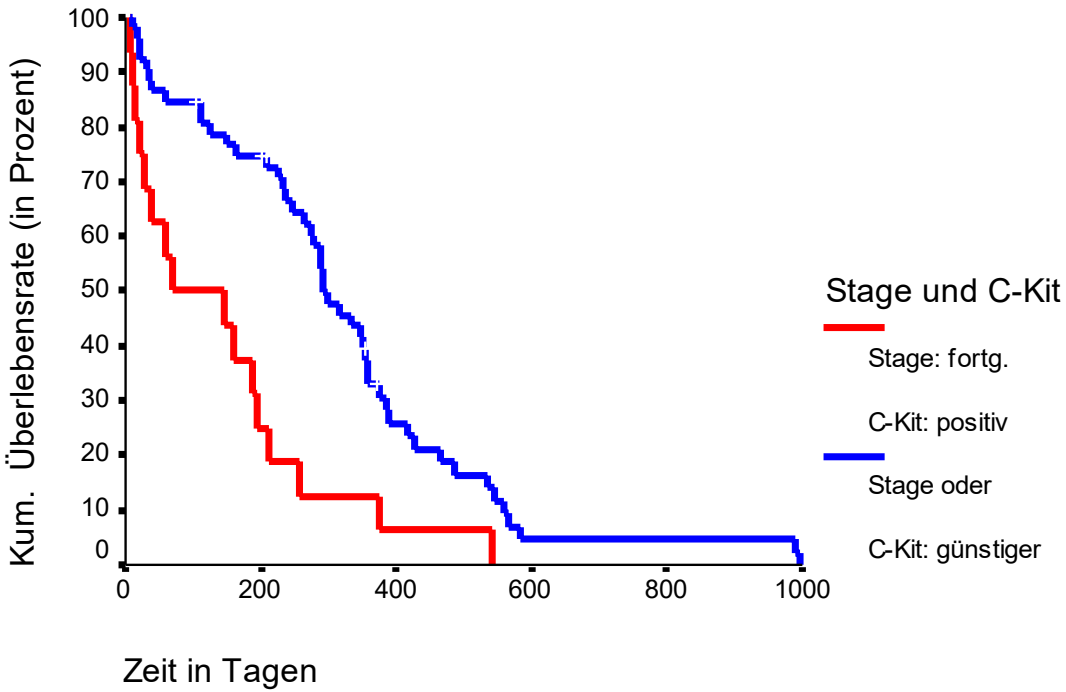
und gleichzeitig

positive C-Kit Expression

Gemeinsamer Einfluss von Tumorstadium und C-Kit Expression

Patienten mit kleinzelligem Bronchialkarzinom

ohne Chemotherapieresponse



Coxregression

Erklärender Einfluss:

Fortgeschrittenes Tumorstadium

und gleichzeitig

positive C-Kit Expression

hr = 2,7: Risikoerhöhung um 170% gegenüber günstigerer Ausgangslage

Take Home Message

