Biomathematik

Klinische Studien

Phasen der klinischen Studie

Tierversuche

I **Medikament bei gesunden Probanden**

Pharmakokinetik
Verträglichkeit

II **Medikament bei einzelnen Patienten**

Wirksamkeit, Verträglichkeit
Pilotstudie

III **Klinische Therapiestudie**

Wirksamkeit

Zulassung durch Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM)

IV **Breite des Anwendungsbereichs** (Wirkung, Nebenwirkung)

I

n = 8 - 24

biometrische Stichprobenumfangschätzung nicht erforderlich

Medikament an gesunden Probanden evtl. auch an Patienten

Dosierung ***(schrittweise Erhöhung...evtl cross-over)***

Verträglichkeit ***(Nebenwirkungen)***

biologische Verfügbarkeit ***(AUC, C_{max} ,t_{max})***

Kriterien für Anfangsdosis
Kriterien für Dosissteigerung
Placebo

Zielgrößen

Berechnungsgrundlagen

Deskriptive Statistik
Median, Perzentile,
Mittelwert, Standardabweichung
(Geometrisches Mittel)

II

n = 8 - 20 pro Gruppe

biometrische Stichprobenumfangschätzung nicht erforderlich

Medikament an einzelnen Patienten (Pilotstudie)

Überprüfung der in Phase I festgelegten Dosis
Wirkung, Verträglichkeit
evtl. Vergleichsmedikation (Placebo)

Biometrie

Explorative Datenanalyse
Stichprobenparameter
Konfidenzintervalle ...

Design
Zielgrößen

Grundlage einer biometrischen Stichprobenumfangschätzung für Phase III

III

Klinische Therapiestudie

-Design — **prospektiv, kontrolliert, randomisiert, verblindet, offen**
parallel, cross-over ...
Anzahl der Zentren und Gruppen

- Ziel — **Definition von Haupt- und Nebenzielgrößen**
biometrische Stichprobenumfangschätzung (*α, power*)

- Patienten — Ein- und Ausschlußkriterien
Dosierungsschema, Behandlungs- bzw. Untersuchungsplan
Patientenaufklärung, Patientenversicherung
Compliance
drop-outs
Randomisierungsplan

- Daten — **Monitoring**
Datenmanagement

- Auswertung — **statistische Testverfahren** *(Name, Signifikanzniveau α)*
explorative Datenanalyse

IV

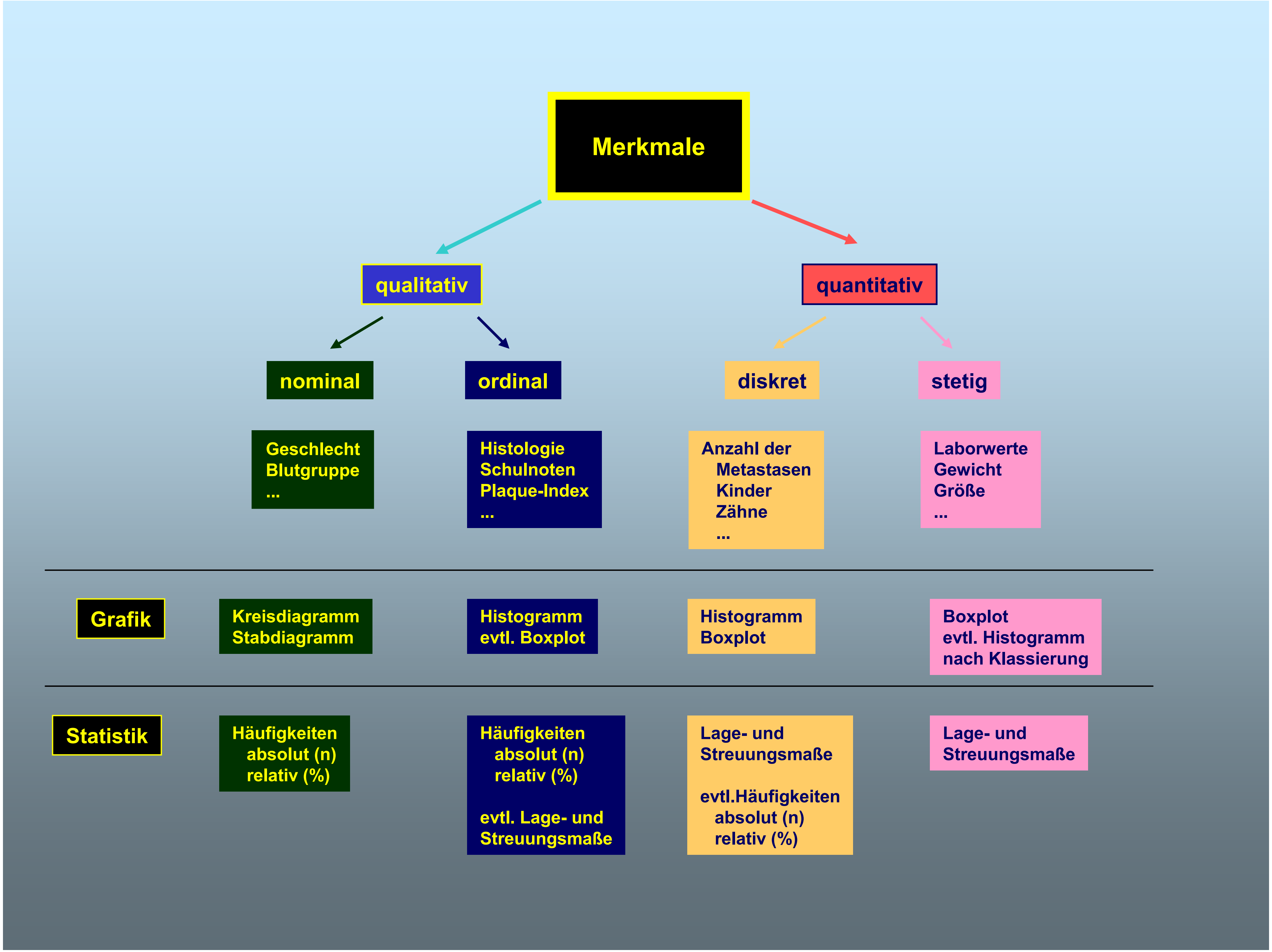
Breite des Anwendungsbereichs - Nebenwirkungen

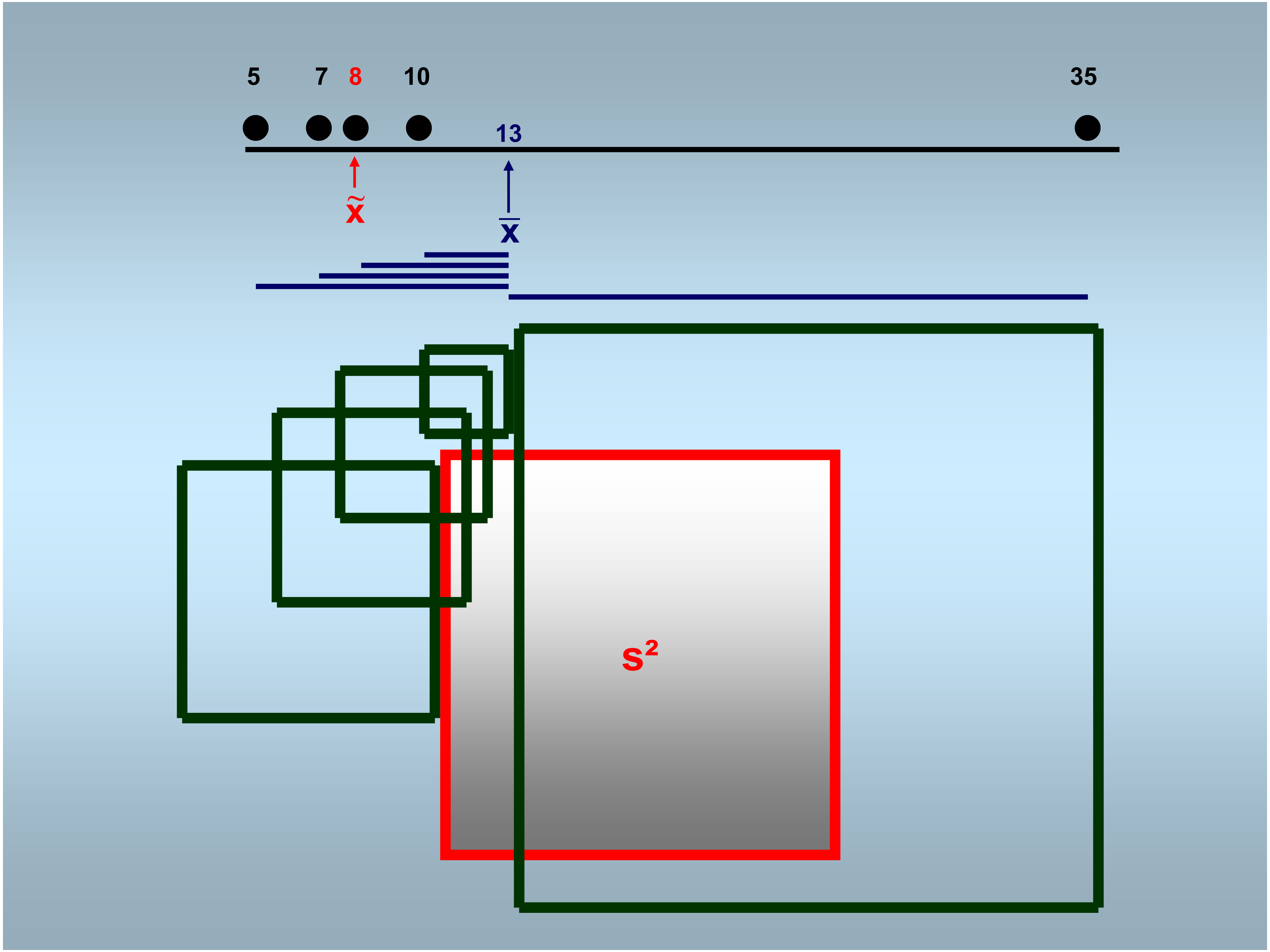
Kontrollierte Studien wie Phase III

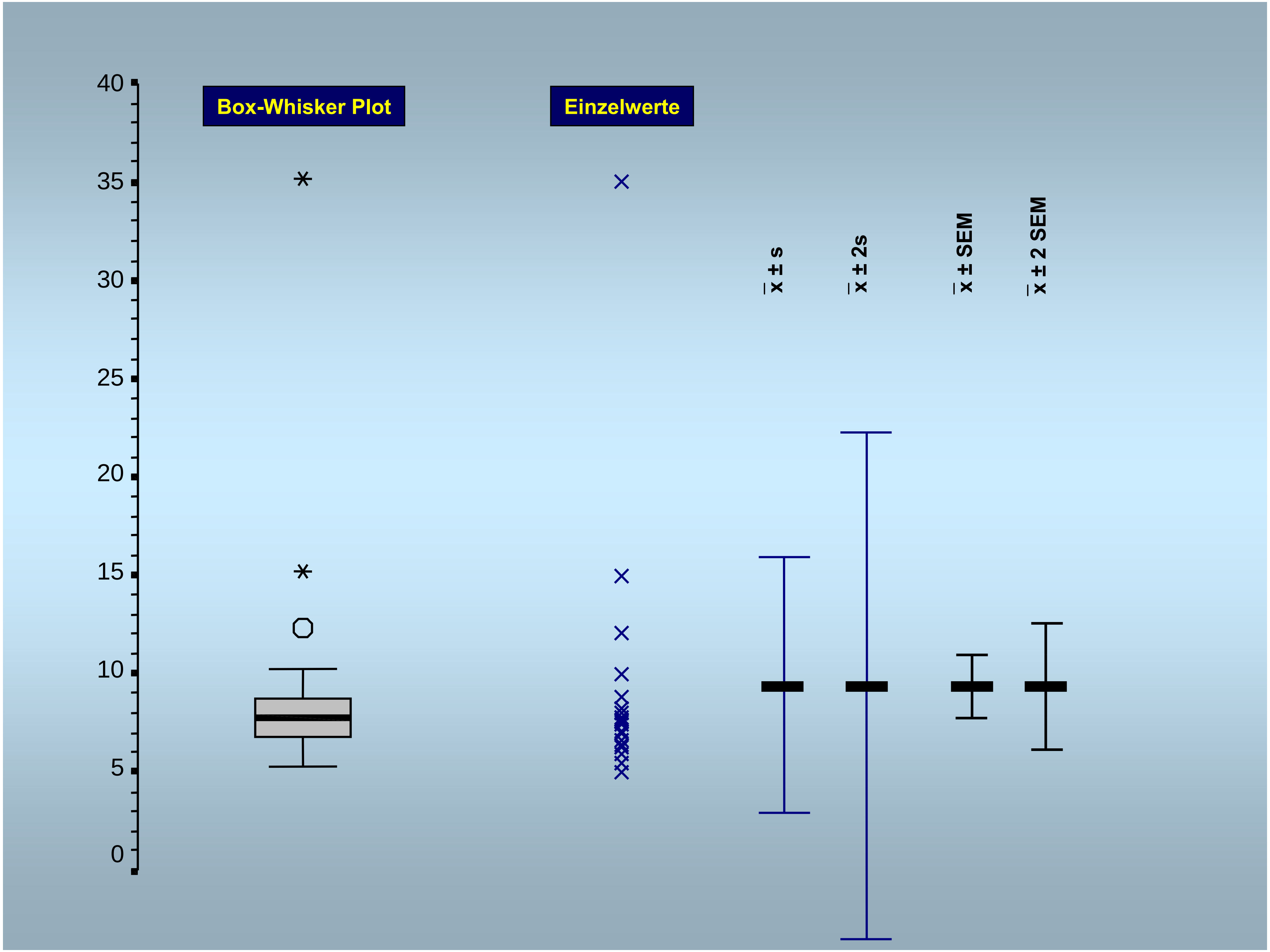
Postmarketingstudien
Anwendungsbeobachtungen

Q1 - Biometrie

Deskriptive Statistik


Merkmale
qualitativ
quantitativ
nominal
ordinal
diskret
stetig
Geschlecht
Blutgruppe
...
Histologie
Schulnoten
Plaque-Index
...
Anzahl der
Metastasen
Kinder
Zähne
...
Laborwerte
Gewicht
Größe
...
Grafik
Kreisdiagramm
Stabdiagramm
Histogramm
evtl. Boxplot
Histogramm
Boxplot
Boxplot
evtl. Histogramm
nach Klassierung
Statistik
Häufigkeiten
absolut (n)
relativ (%)
Häufigkeiten
absolut (n)
relativ (%)
evtl. Lage- und
Streuungsmaße
Lage- und
Streuungsmaße
evtl.Häufigkeiten
absolut (n)
relativ (%)
Lage- und
Streuungsmaße


5
7
8
10
35
13
x̃
x̄
s²



40
35
30
25
20
15
10
5
0
Box-Whisker Plot
Einzelwerte
x̄ ± s
x̄ ± 2s
x̄ ± SEM
x̄ ± 2 SEM

Deskriptive Statistik

Problem:

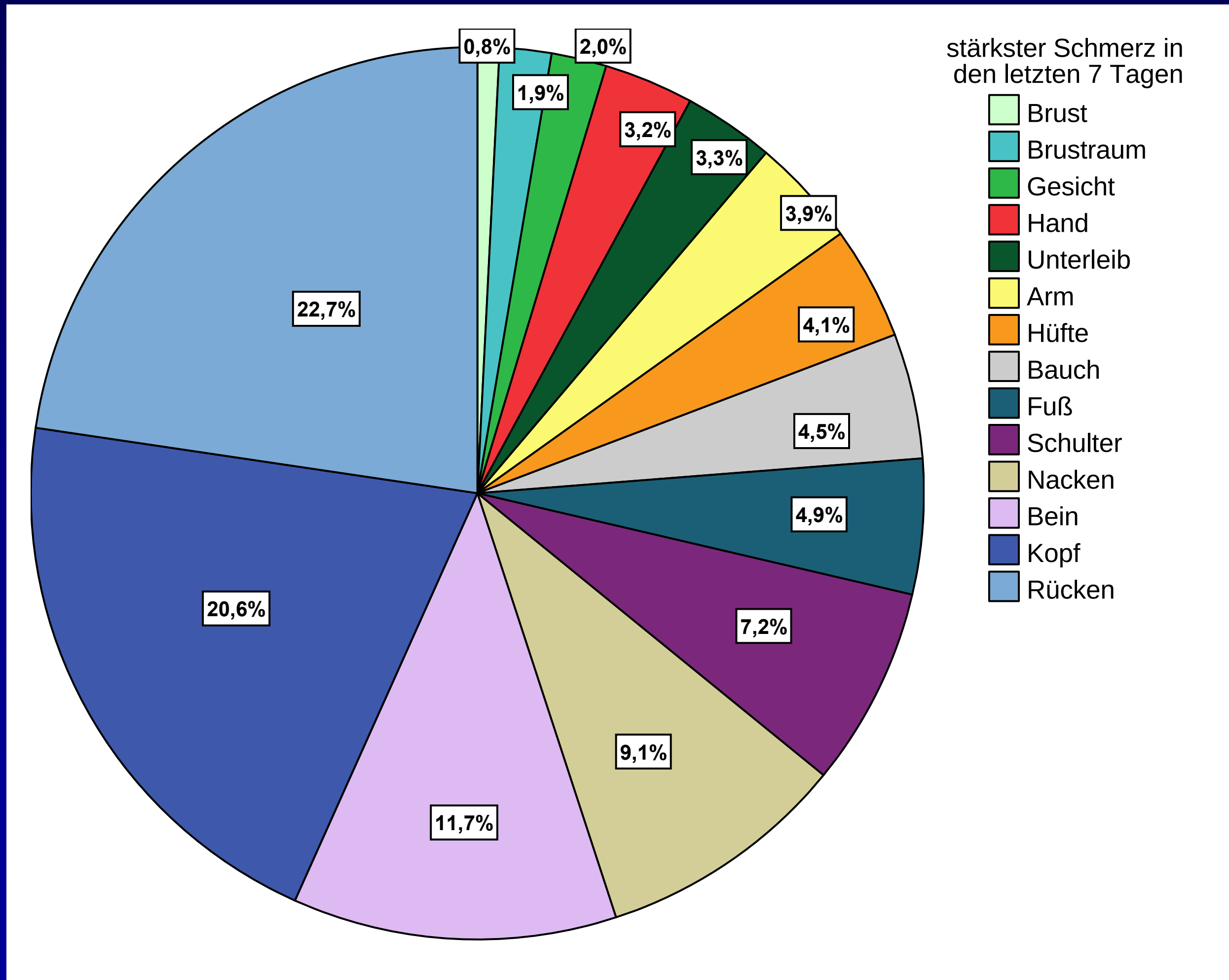
Wie lassen sich die erhobenen Daten übersichtlich darstellen und beschreiben?

Lösung:

durch geeignete Grafiken und Maßzahlen, die die Daten repräsentieren und zusammenfassen

Kategoriale Variablen: Kreisdiagramm

- Bei Personen mit Schmerzen innerhalb der letzten 7 Tage wurde nach dem am stärksten schmerzenden Körperteil gefragt (n=4656)



Zeilenprozente beziehen sich auf die Zahl der Merkmalsträger in der jeweiligen Zeile:

Blutgruppe / **Ethnie**	*A*	*B*	*AB*	*0*	***Gesamt***
Europäische Amerikaner	45	8	4	43	100
Afroamerikaner	29	17	4	50	100
Deutsche	42	11	4	43	100
Südamerikanische Indianer	0	0	0	100	100
Australische Aborigines	56	0	0	44	100
Gesamt	34	7	2	56	100[1]

[1] Abweichung aufgrund von Rundungsfehlern

Spaltenprozente beziehen sich auf die Zahl der Merkmalsträger in der jeweiligen Spalte:

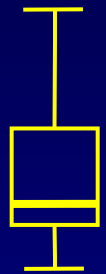
Blutgruppe **Ethnie**	*A*	*B*	*AB*	*0*	***Gesamt***
Europäische Amerikaner	26	22	33	15	20
Afroamerikaner	17	47	33	18	20
Deutsche	24	31	33	15	20
Südamerikanische Indianer	0	0	0	36	20
Australische Aborigines	32	0	0	16	20
Gesamt	100[1]	100	100[1]	100	100

[1] Abweichung aufgrund von Rundungsfehlern

In diesem Beispiel ist nur die Angabe von Zeilenprozenten von Interesse, weil die anderen Prozentwerte davon abhängen, wie groß die Stichproben in den einzelnen Bevölkerungen gewählt werden.

Lagemaße

Streuungsmaße



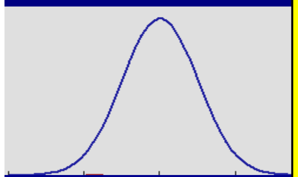
Median (**50.Percentil oder Zentralwert))**

Percentile 25 und 75 (Quartile)

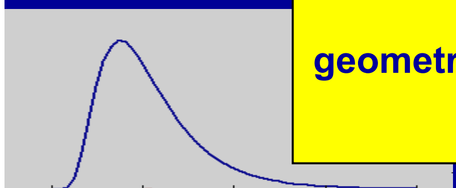
10 und 90, 5 und 95

3 und 97

Minimum Maximum



Arithmetischer Mittelwert $\frac{1}{n}\sum x_i$



geometrisches Mittel

$$\sqrt[n]{x_1 \cdot x_2 \cdots \quad \cdot x_n}$$

Interpercentilbereiche

Box-Whisker Plot

Referenzbereich

Spannweite (range)

Standardabweichung $\sqrt{\frac{\sum (x_i - m)^2}{n-1}}$

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

Variationskoeffizient (rel. Std-abw.)

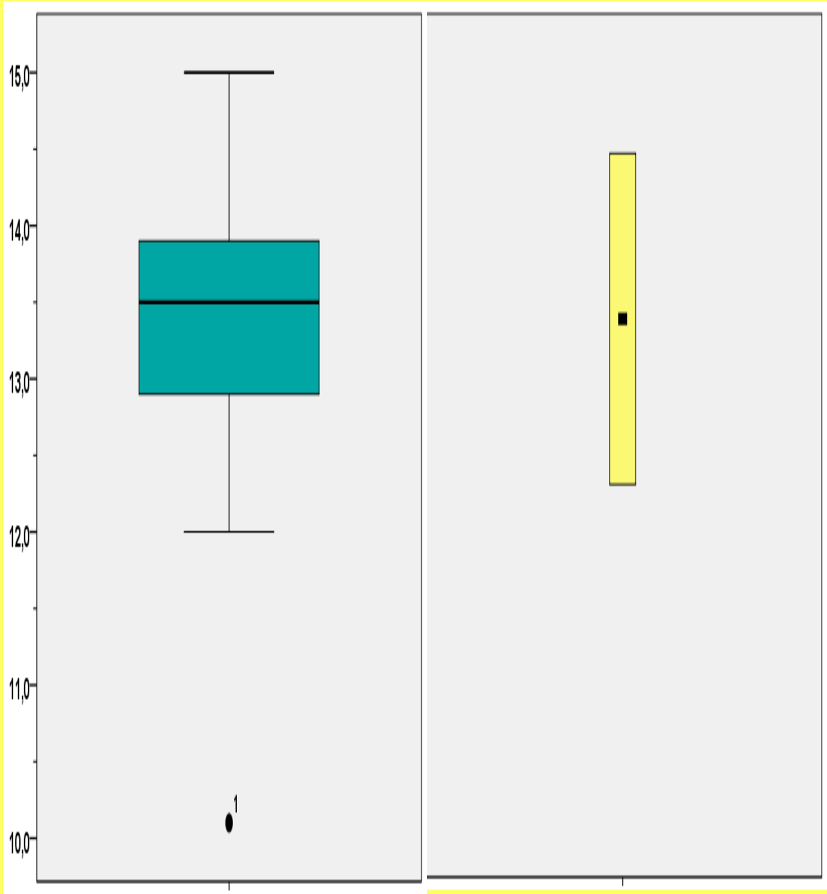
$$V_k = \frac{s}{m} \cdot 100\ [\%]$$

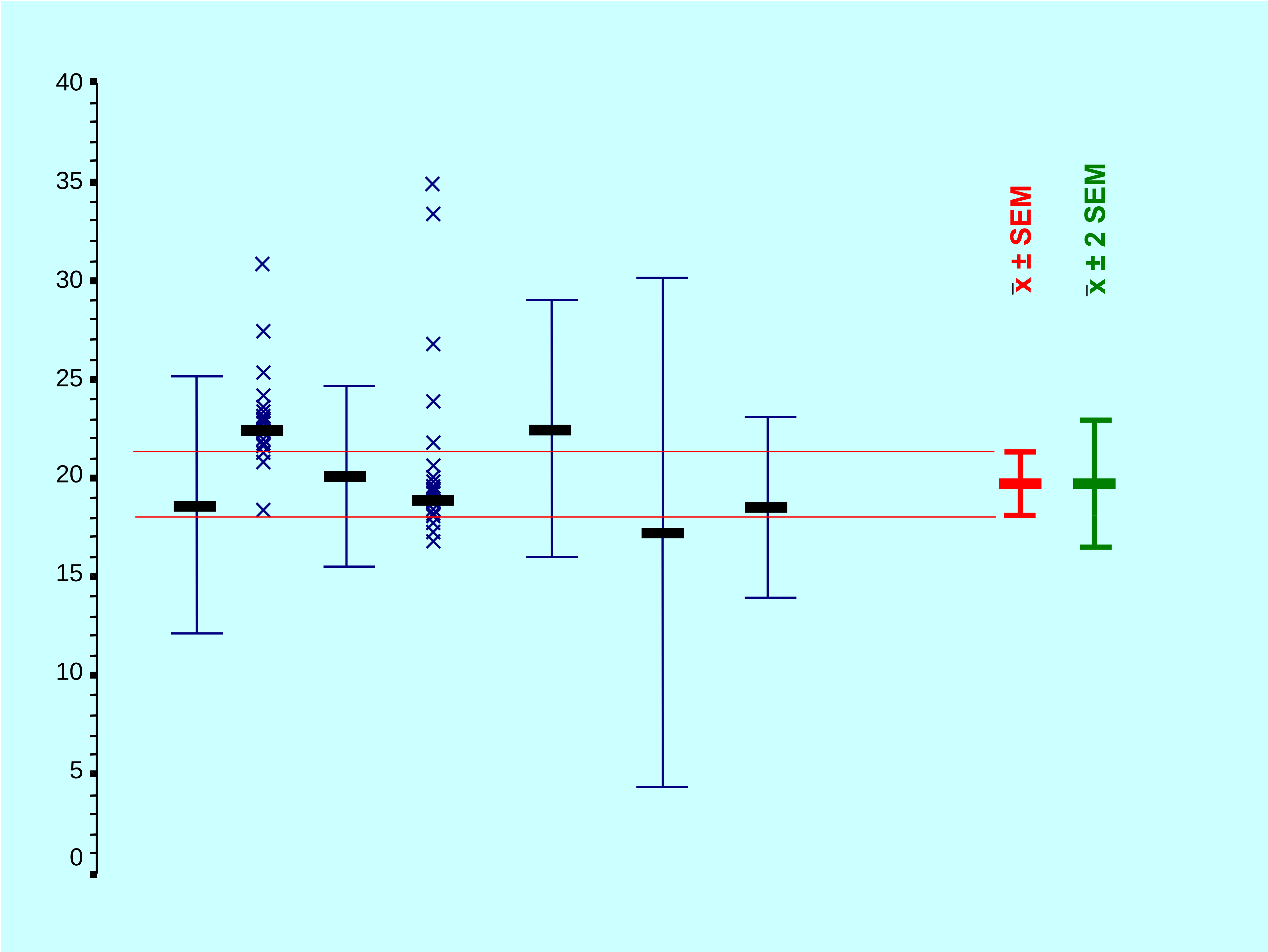
Hb
10,1
12,0
12,4
12,8
12,8
12,9
13,0
13,0
13,3
13,4
13,5
13,6
13,7
13,7
13,8
13,9
14,3
14,4
14,5
14,9
15,0

Statistiken

Hb

N	Gültig	21
	Fehlend	0
Mittelwert		13,381
Standardabweichung		1,0866
Minimum		10,1
Maximum		15,0
Perzentile	25	12,850
	50	13,500
	75	14,100





40
35
30
25
20
15
10
5
0
x̄ ± SEM
x̄ ± 2 SEM

Normbereiche / Referenzbereiche

- Im Allgemeinen ist der Referenzbereich der Bereich, in dem 95% aller gesunden Menschen liegen. Jeder 20. Gesunde liegt also bei einem bestimmten Laborwert nicht im Referenzbereich.
- Beispiel BMI:

  Als normalgewichtig gelten Männer mit einem BMI zwischen 20 und 25 und

  Frauen mit einem BMI zwischen 19 und 24.

Klassifikation	Männer	Frauen
Untergewicht	<20	<19
Normalgewicht	20-25	19-24
Übergewicht	25-30	24-30
Adipositas	30-40	30-40
Massive Adipositas	>40	>40

Skalentransformation

- **Problem:**

  Viele statistische Verfahren setzen normalverteilte Daten voraus. In der Praxis hat man es aber oft mit nicht normalverteilten Daten zu tun.

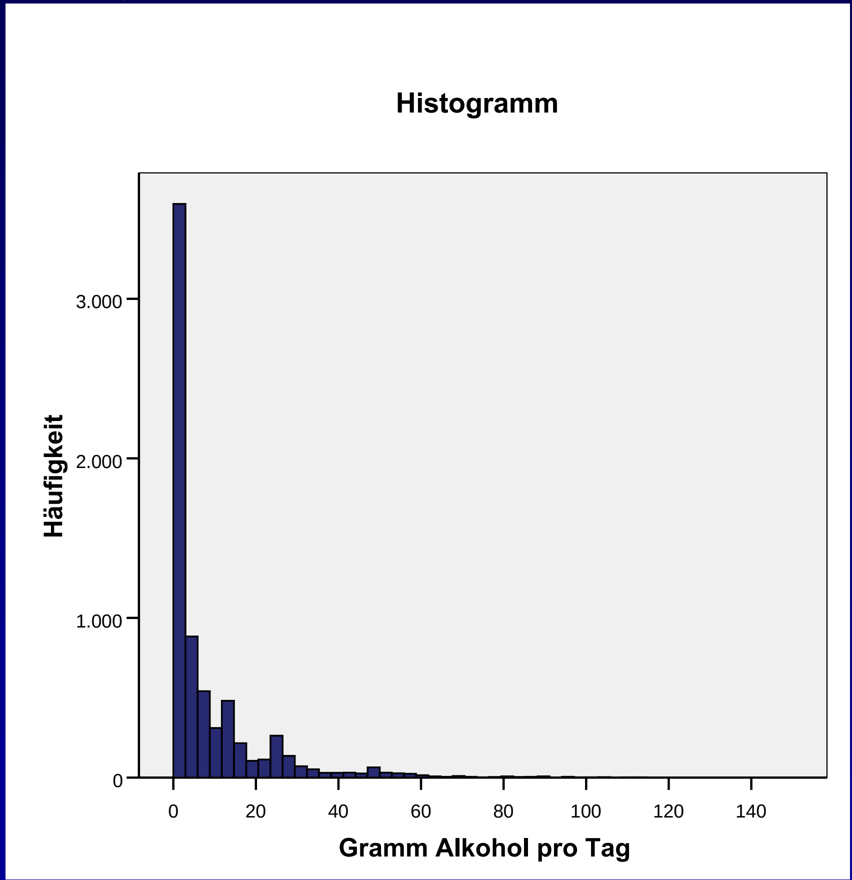
- **Lösung:**

  Skalentransformation zum Beispiel durch Logarithmierung der Daten.

Quantitative Variablen: Histogramm

durchschnittlicher Alkoholkonsum pro Tag in Gramm reinem Alkohol (nicht normalverteilt) :

N		7124
Mittelwert		8,72
Median		2,81
Varianz		201,34
Standardabweichung		14,19
Minimum		0
Maximum		164,27
Spannweite		164,27
Perzentile	25	0,25
	50	2,81
	75	11,88



8,72 g /Tag entspricht etwa 1,5 Liter Bier pro Woche,

2,81 g /Tag entspricht etwa 0,5 Liter Bier pro Woche.

Skalentransformation

- **Beispiel:**

  Alkoholkonsum in Gramm pro Tag

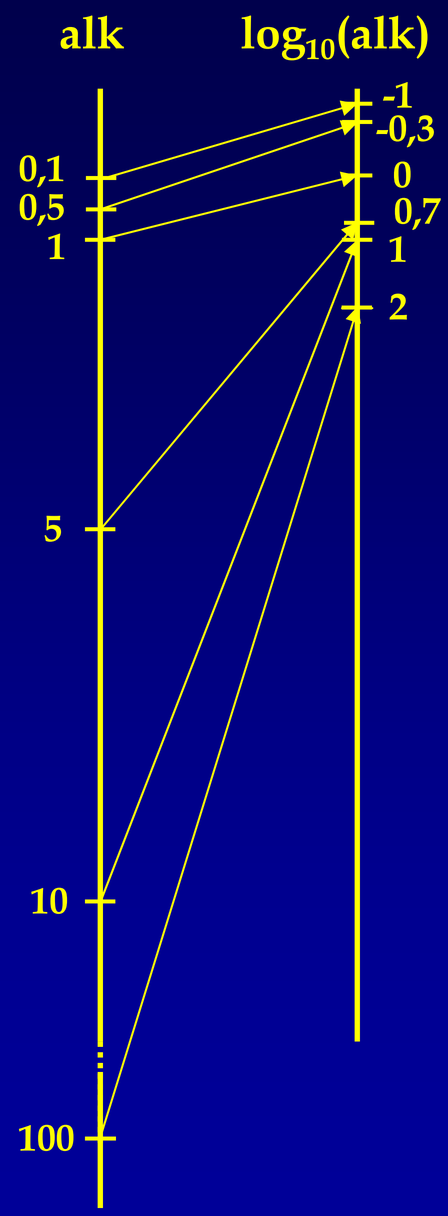
  (bei Personen die Alkohol trinken, n=5881)

  => Viele Personen trinken sehr wenig Alkohol und nur wenige Personen trinken viel Alkohol.

  => Die Verteilung der Daten ist schief.

  => log(alk)= $\log_{10}$(alk) (für alk>0)

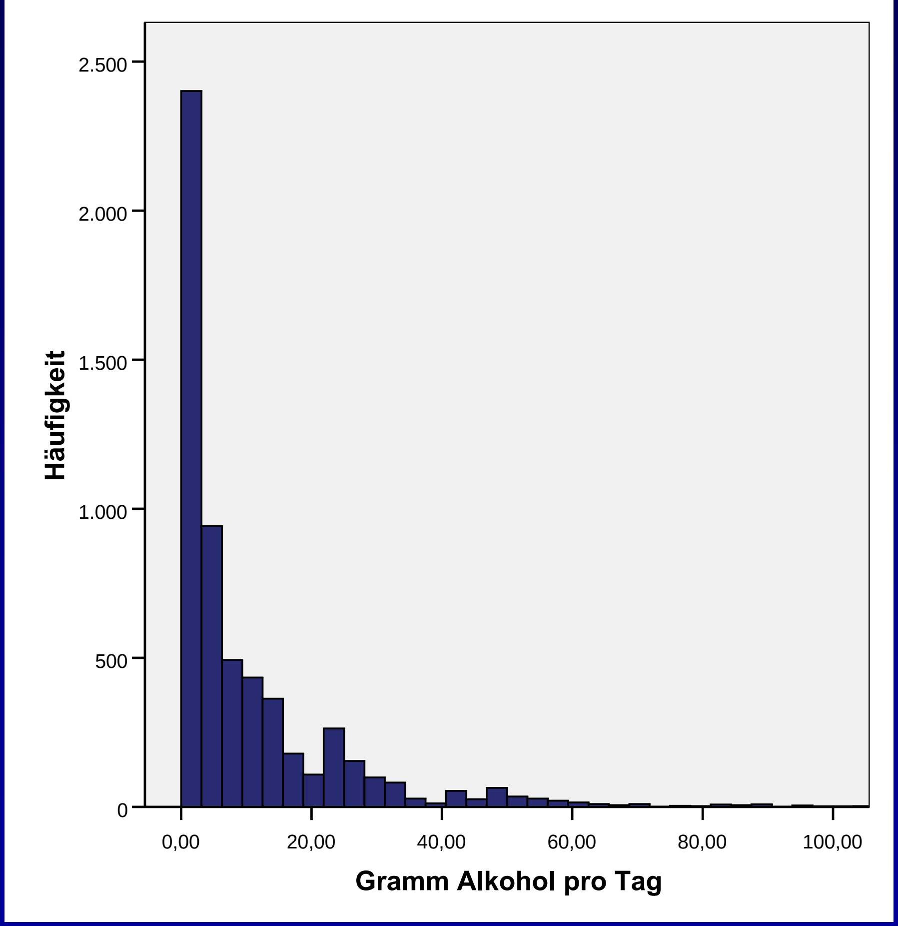
alk	**$\log_{10}$(alk)**
0,1	-1
0,5	-0,3
1	0
5	0,7
10	1
100	2



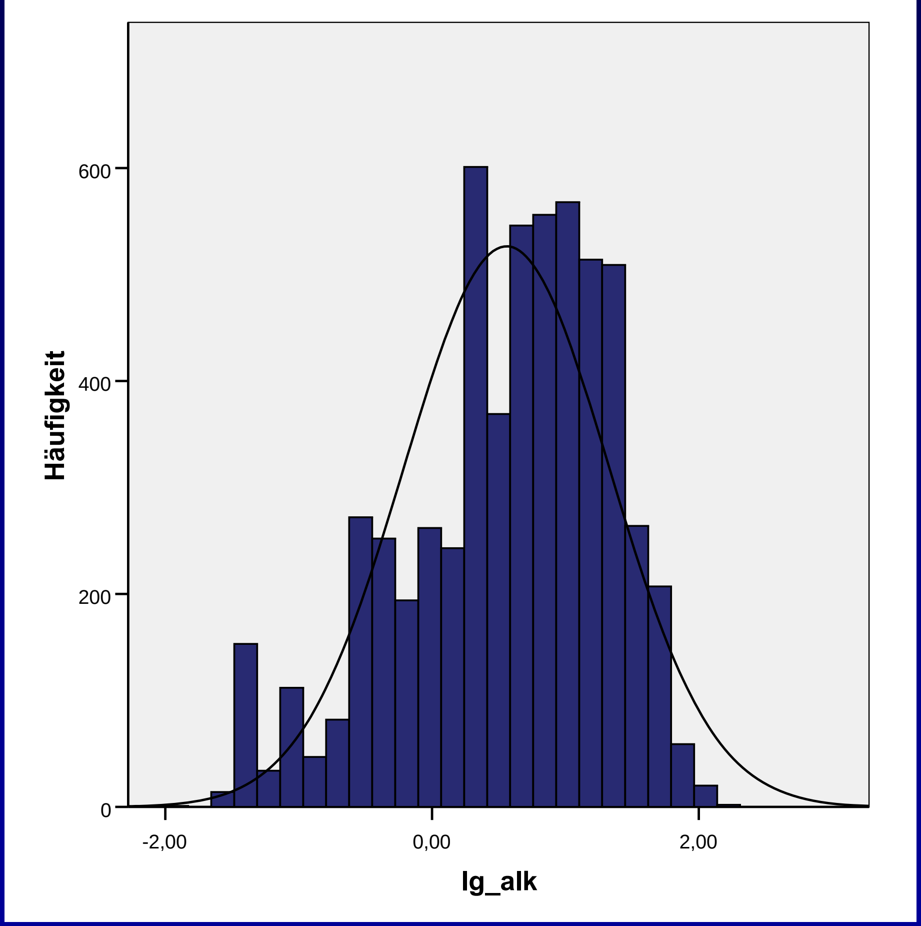
Skalentransformation

Bsp. Alkoholkonsum in Gramm pro Tag für konsumierende Personen

„Rohe“ Werte



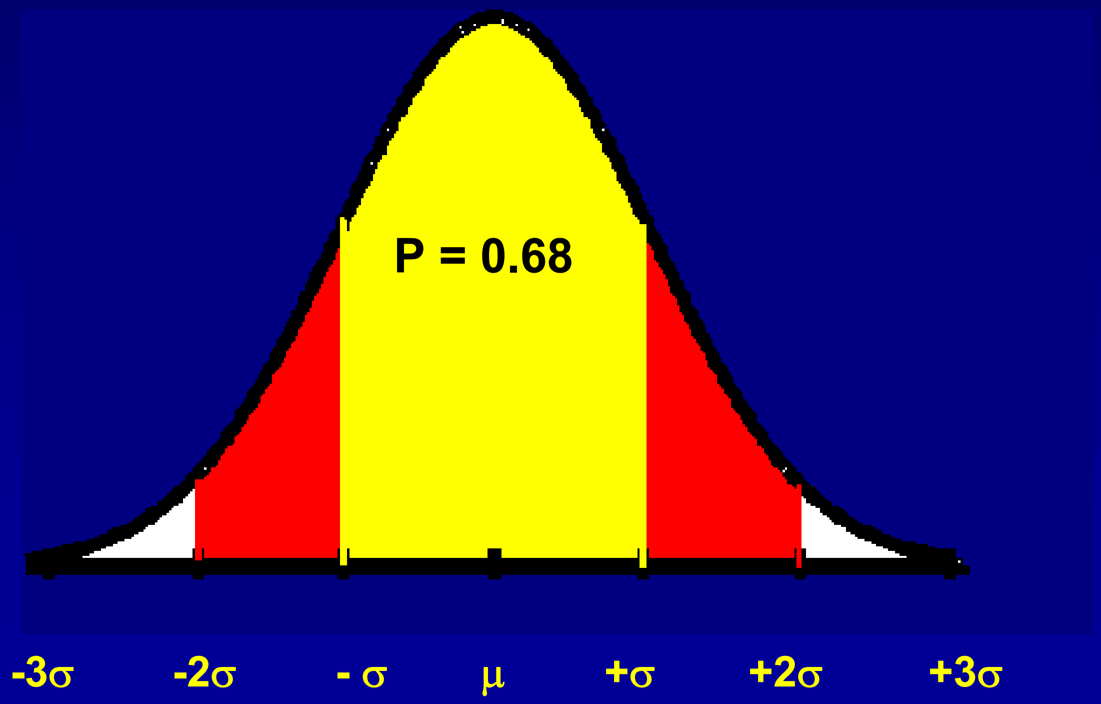
logarithmierte Werte



Messfehler sind in der Regel normalverteilt
Gauß‘sche Glockenkurven beschrieben werden

Die Wahrscheinlichkeit, dass ein Versuchsergebnis in einen Bereich fällt, ist identisch mit der Fläche

P = 0.955



Biomathematik

Korrelations- und Regressionsanalyse

Lernziele: Korrelation und Regression

1. linearer Korrelationskoeffizient r (nach Pearson)
   (Berechnung und Interpretation)
2. Bestimmtheitsmaß
3. Spearman Rangkorrelation
4 Interpretation der Regressionsgleichung
5. Diagramm nach ‚Bland-Altman'

Korrelation und Regression

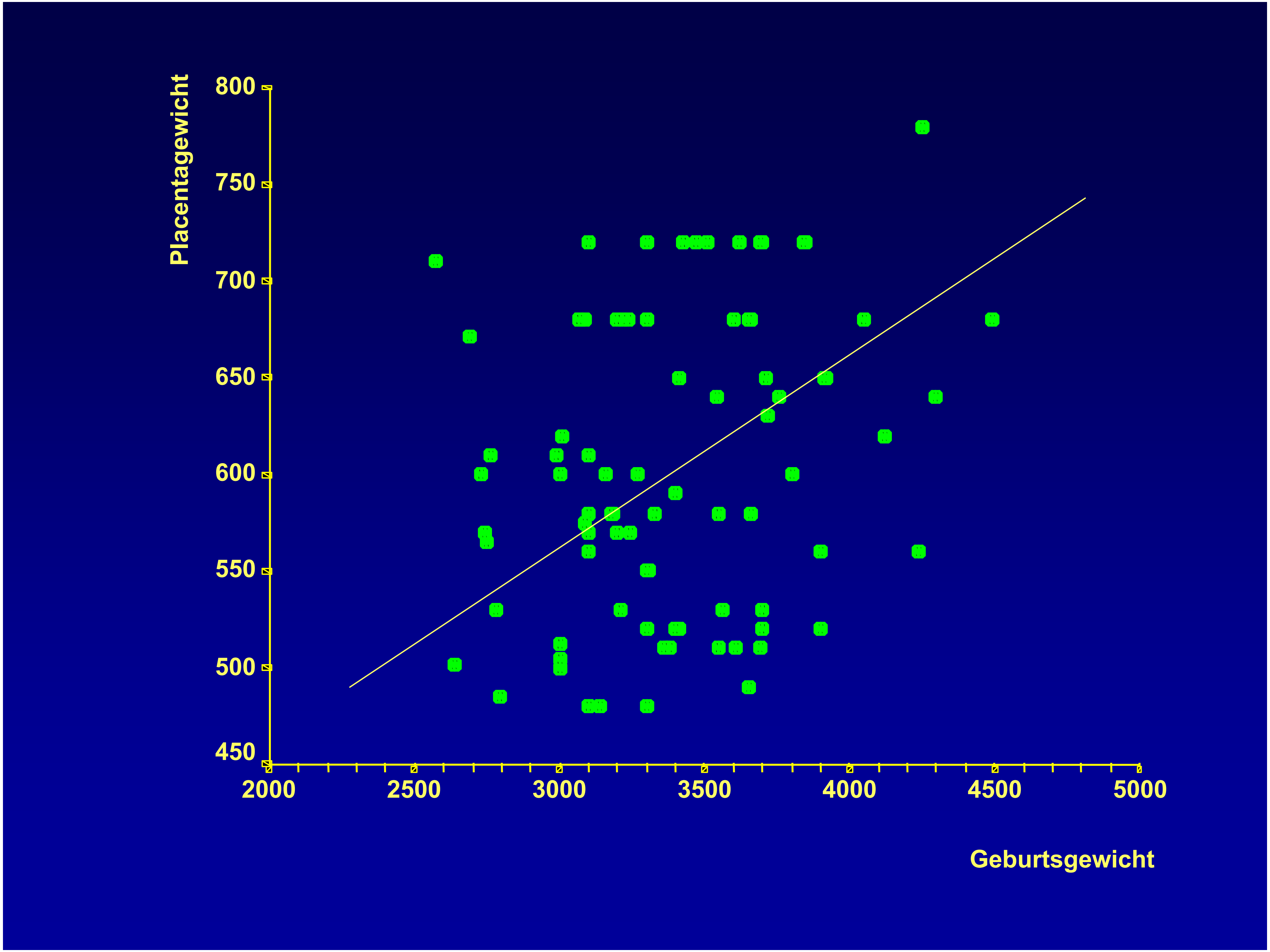
Problem:

Wie lassen sich lineare Abhängigkeit (Übereinstimmung) zweier Merkmale beschreiben?

Lösung:

Korrelation (Analyse des Zusammenhangs)

Regression (formelmäßige Erfassung des Zusammenhangs)


Placentagewicht
800
750
700
650
600
550
500
450
2000
2500
3000
3500
4000
4500
5000
Geburtsgewicht

Korrelations- und Regressionsanalyse

In empirischen Daten wird man praktisch **nie** eine **exakt** lineare Abhängigkeit zweier Messgrößen beobachten können, vielmehr werden die Daten eine Punktwolke bilden, die nur „mehr oder weniger gut“ zu einer Geraden passt.

Daraus ergeben sich die folgenden Fragestellungen:

Welche Gerade passt am besten zu einer beobachteten Punktwolke und wie wird diese Gerade bestimmt? - lineare Regressionsanalyse

Wie gut ist diese Anpassung? - Korrelationsanalyse

Lineare Zusammenhänge und Geradengleichung

Eine Größe Y hängt *linear* von einer Größe X:

Änderungen von X auf X+ΔX führen immer zu Änderungen von Y auf Y + **b • ΔX**

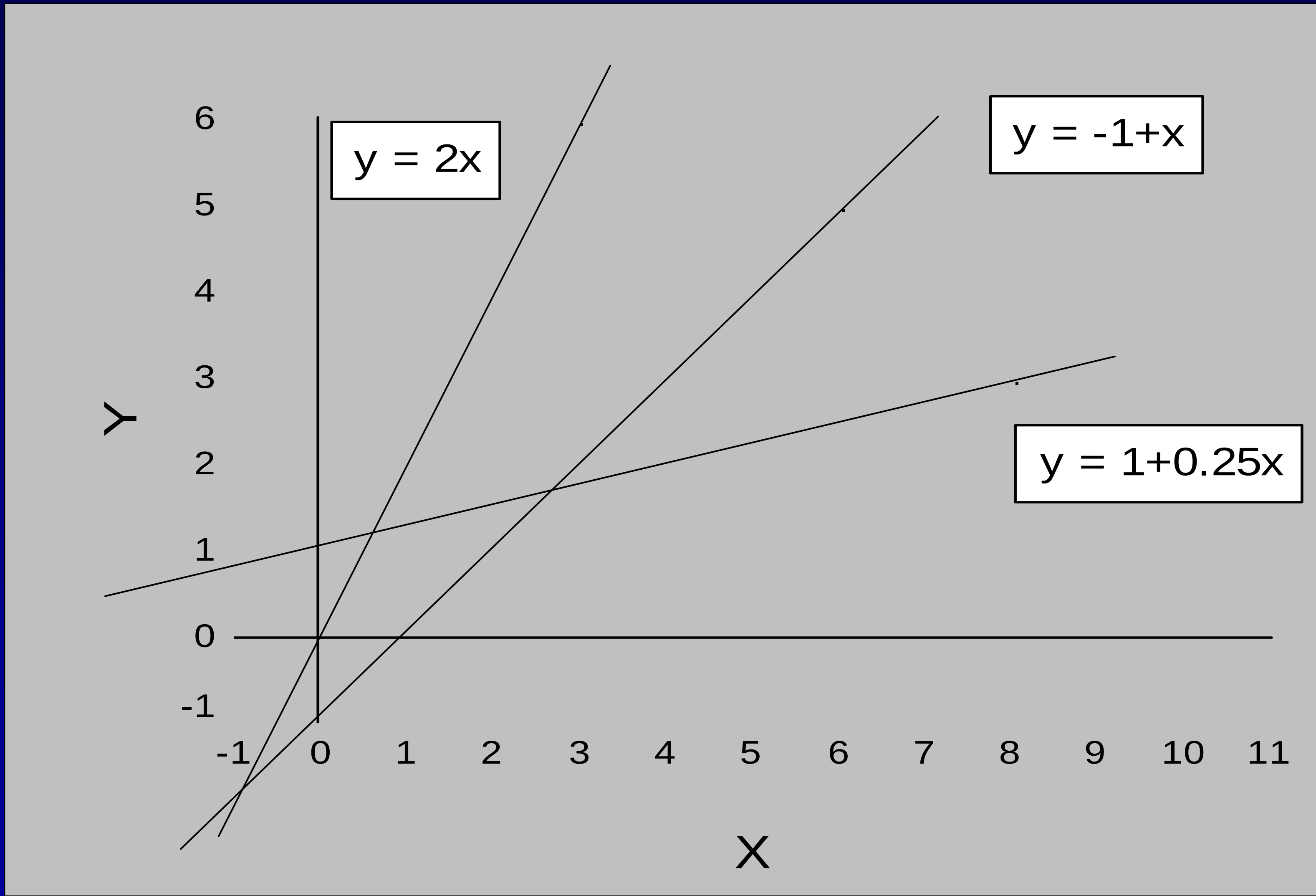
wobei b für alle X und ΔX identisch ist.

Jede derartige lineare Abhängigkeit kann durch eine Gleichung

$$y = a + b \cdot x$$

beschrieben und durch eine Gerade grafisch dargestellt werden.

Beispiele für Geraden



Bezeichnungen:

b heißt „Koeffizient“ von X oder Steigung der Geraden.
a heißt „Konstante“ oder y-Achsenabschnitt der Geraden.

Identifikation von a und b in der Grafik:

a ist der Schnittpunkt der Geraden mit der y-Achse.
b lässt sich durch Steigungsdreiecke bestimmen.

Sonderfälle:

a = 0, b beliebig	Gerade durch den Ursprung (0,0) mit Steigung b
b = 0, a beliebig	Waagerechte Gerade durch (0,a)
a = 0, b = 1	Winkelhalbierende mit y = x

Senkrechte Geraden?? Nicht darstellbar!!

Korrelation und Regression: Beispieldatensatz

- Lebenserwartung von Männern und Frauen

  Datei‘ WELT.SAV‘ aus SPSS - Achim Bühl und Peter Zöfel
- N=109 Länder
- Die Datei enthält Informationen zu :

  Lebenserwartung (Männer und Frauen)

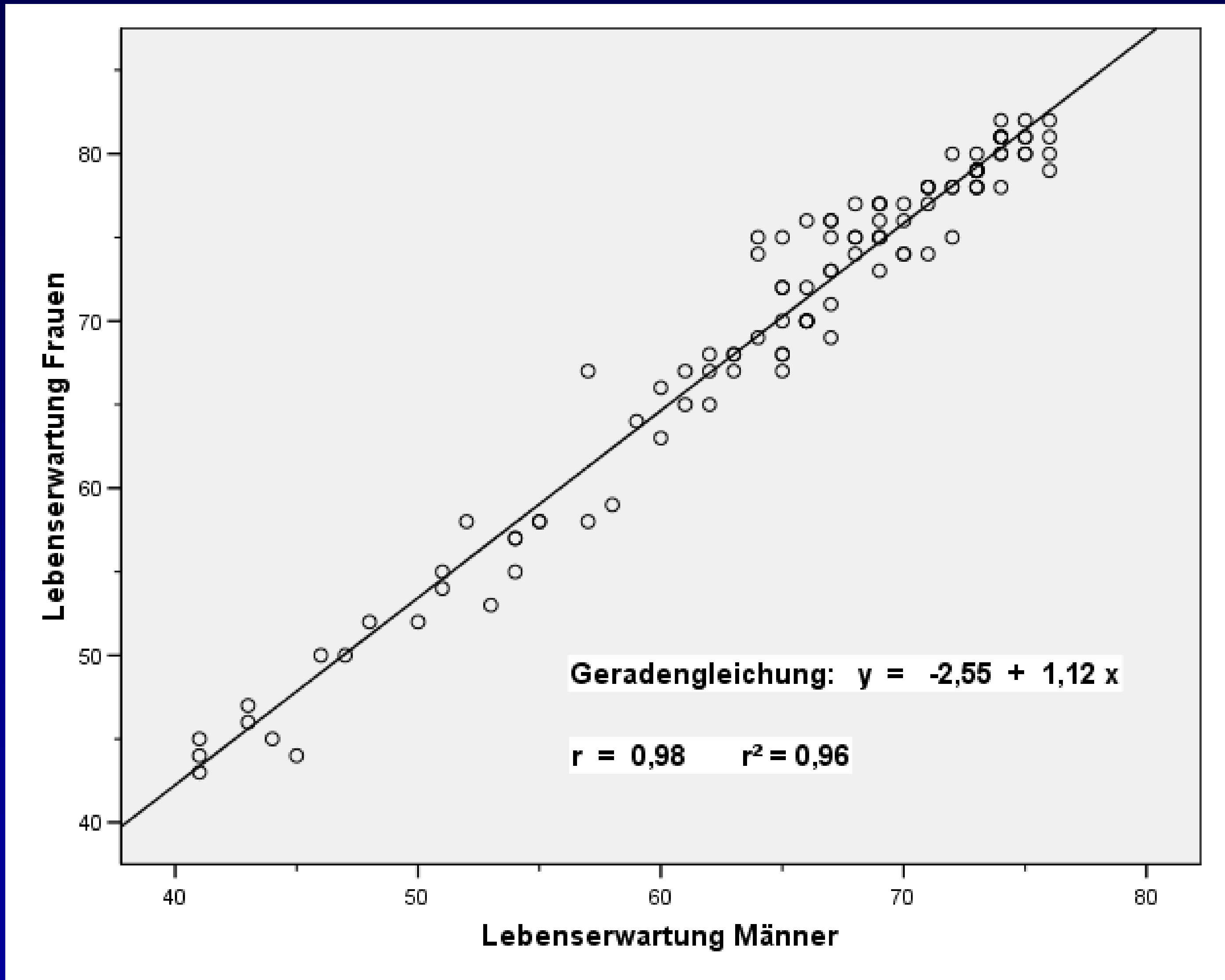
  Bevölkerungszuwachs (Prozent)

  Proz. Anteil der Stadtbevölkerung

  Tägliche Kalorienaufnahme

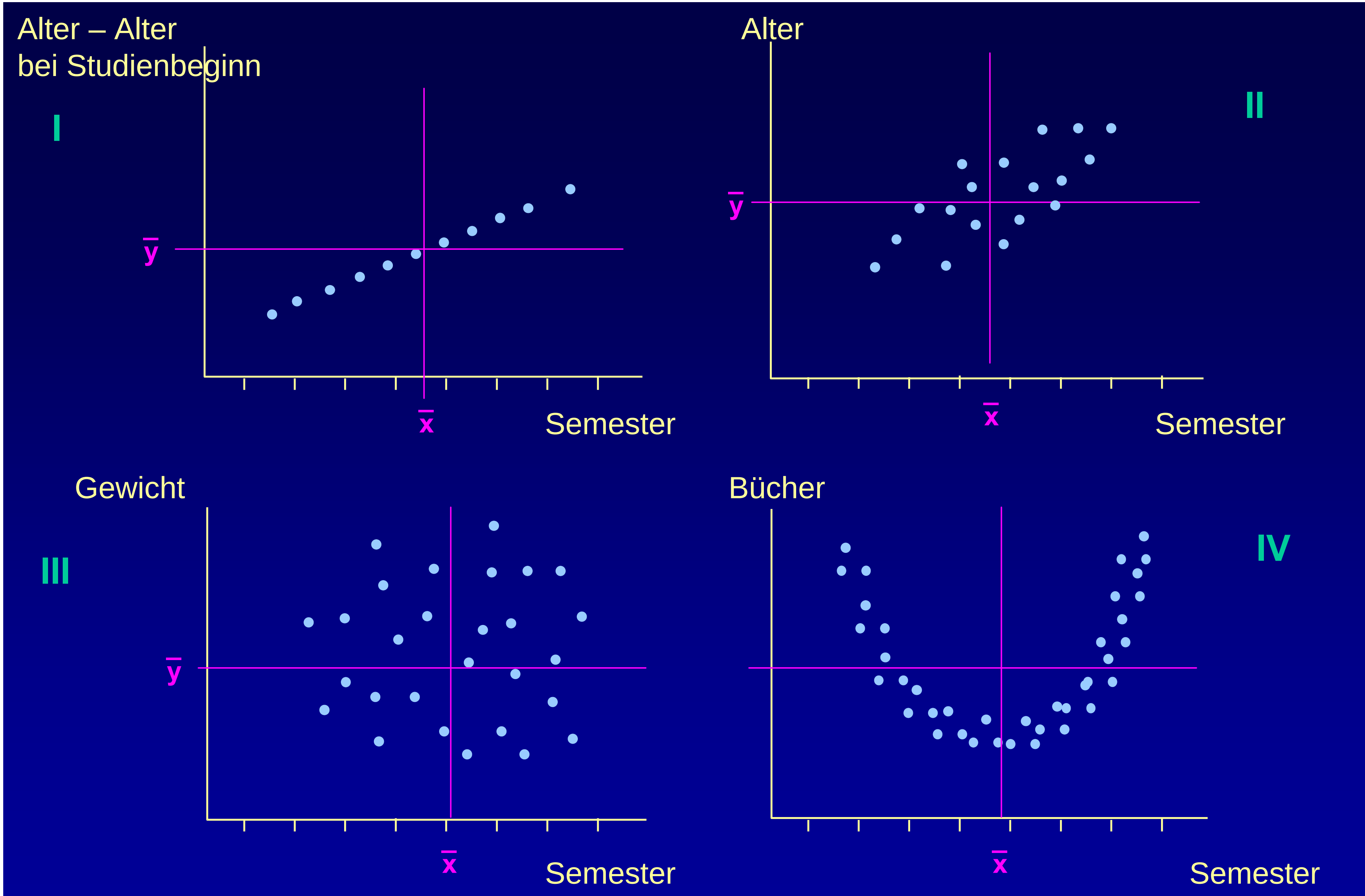
  Einwohnerdichte

Punktwolke mit Korrelation und Regression



$$r = \frac{\sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}}$$

$$= \frac{\Sigma x_i y_i - \frac{1}{n} (\Sigma x_i)(\Sigma y_i)}{\sqrt{(\Sigma x_i^2 - \frac{1}{n} (\Sigma x_i)^2) * (\Sigma y_i^2 - \frac{1}{n} (\Sigma y_i)^2)}}$$



Punkt (x_i / y_i) $(x_i - \bar{x} / y_i - \bar{y})$ Koordinatenprodukt $(x_i - \bar{x}) \bullet (y_i - \bar{y})$

$\Sigma \, (x_i - \bar{x}) \bullet (y_i - \bar{y})$

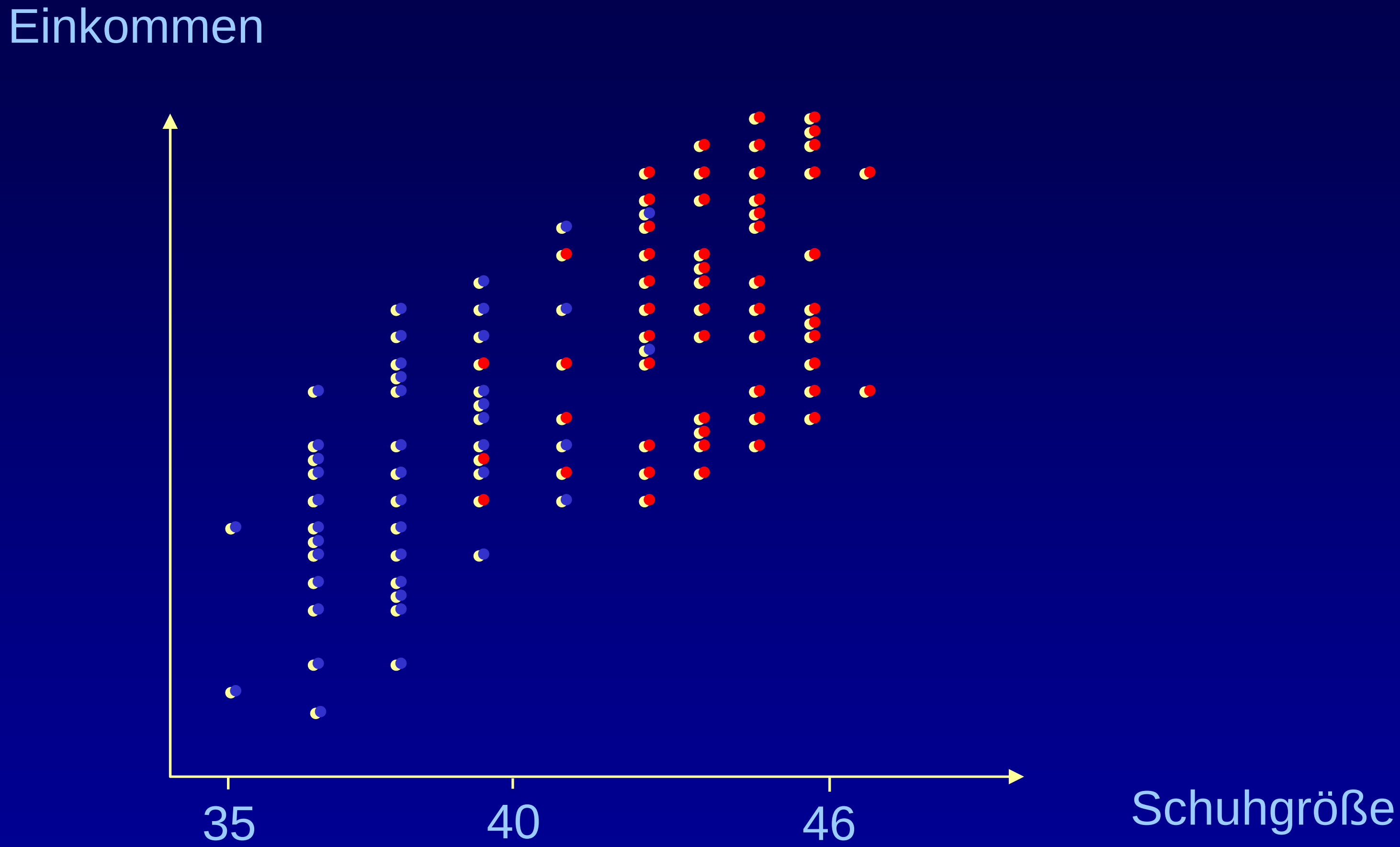
$$r = \frac{\sum (x_i - \overline{x}) \cdot (y_i - \overline{y})}{\sqrt{\sum (x_i - \overline{x})^2} \sqrt{\sum (y_i - \overline{y})^2}}$$

Qualitative Beurteilung von r (bei wenigstens 30 Wertepaaren)

r^2		
$\leq$ 4 %	$\lvert r \rvert \leq 0{,}2$	keine statistische Bedeutung
4 % - 16 %	$0{,}2 < \lvert r \rvert \leq 0{,}4$	schwache Korrelation
16 % - 50 %	$0{,}4 < \lvert r \rvert \leq 0{,}7$	mittlere Korrelation
> 50 %	$\lvert r \rvert > 0{,}7$	starke Korrelation

Scheinkorrelationen

Gemeinsamkeitskorrelation	* Hb - Erythrozytendurchmesser * Schuhgröße - Einkommen * Gehälter der Pastoren - Alkoholpreise * Storchennester - Geburtenrate * Temperatur in Ungarn - Ölverbrauch in der Schweiz * Güte der Wohngegend - Krebsmortalität
Korrelation zwischen Zeitreihen	* Zahl der Rundfunkgenehmigungen - Belegung von Nervenheilanstalten * Krebsmortalität - Orangenimport, Strumpfproduktion Konservennahrung
Korrelation zwischen regionalen Einheiten	* Anzahl der Krebsfälle - Anzahl der Einwohner von Städten


Einkommen
35
40
46
Schuhgröße

Rangkorrelation nach Spearman

$$r_s = 1 - \frac{6 \cdot \sum d_i^{\,2}}{n \cdot (n^2 - 1)}$$

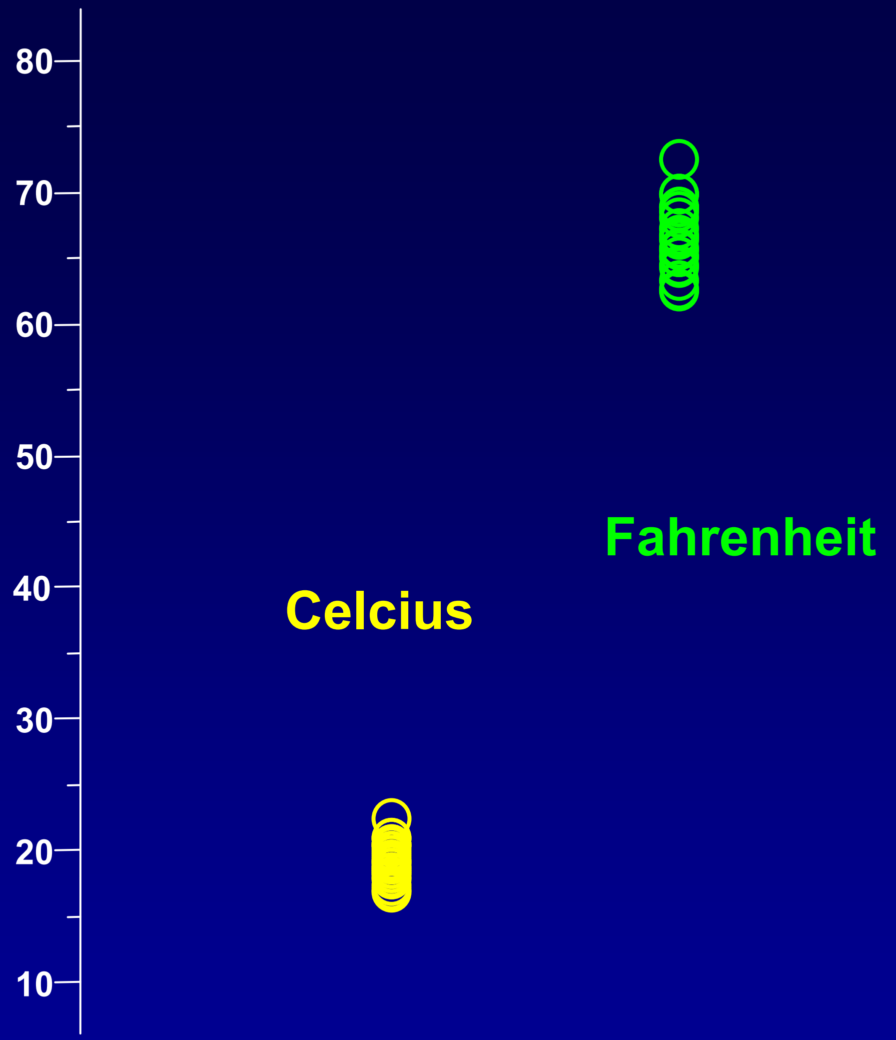
Vorteile von r_s :

- r_s kann auch bei Ausreißern benutzt werden
- r_s kann bei ordinalskalierten Merkmalen eingesetzt werden
- r_s erkennt auch nichtlineare monotone Zusammenhänge
- (Sättigungskurven, exponentieller Abfall)

Die qualitativen Beurteilungen von r und r_s sind identisch.

Temperatur

°F = °$C \bullet$ $9/5$ + 32



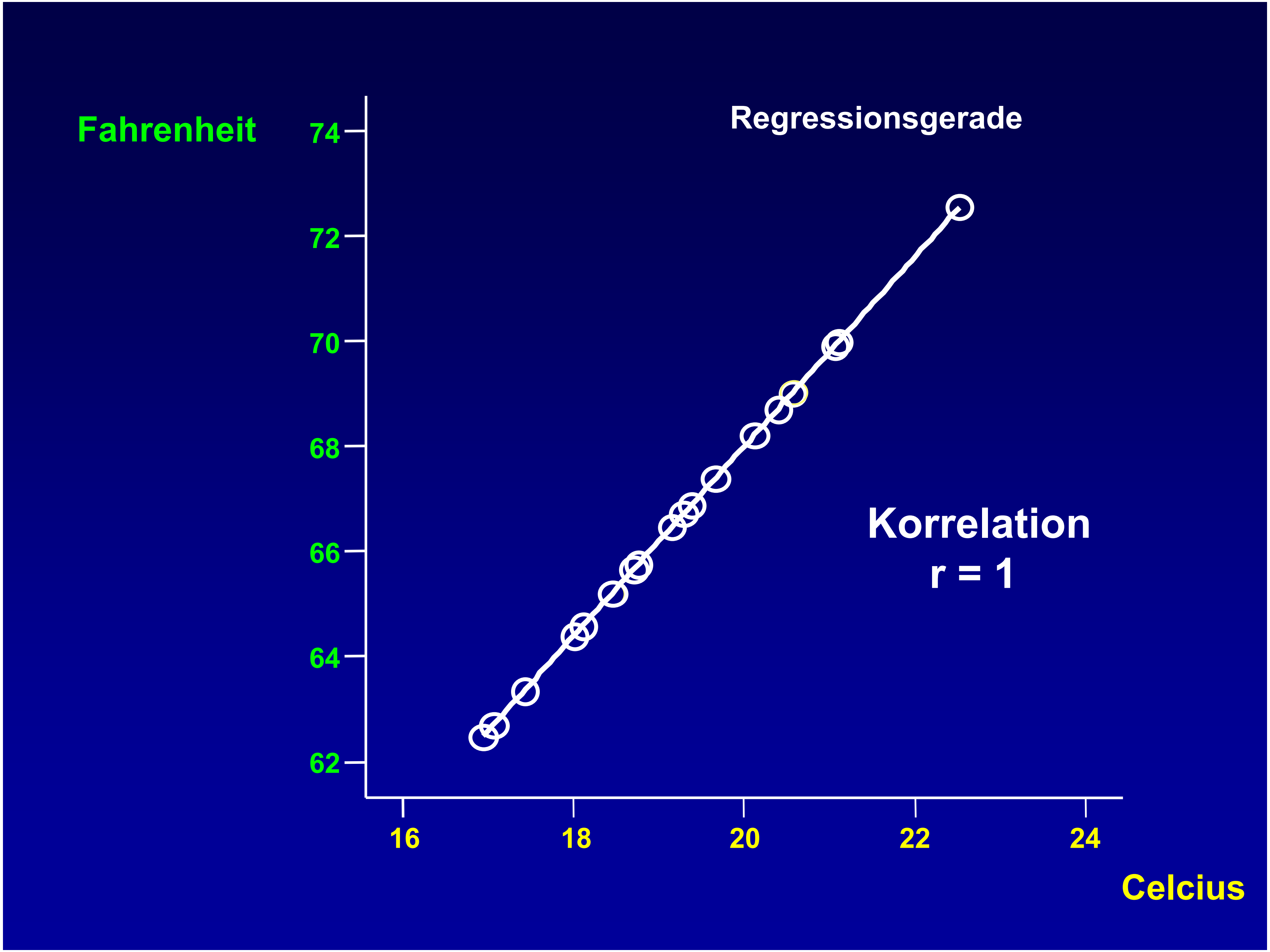
m = 66.7 s = 2.6

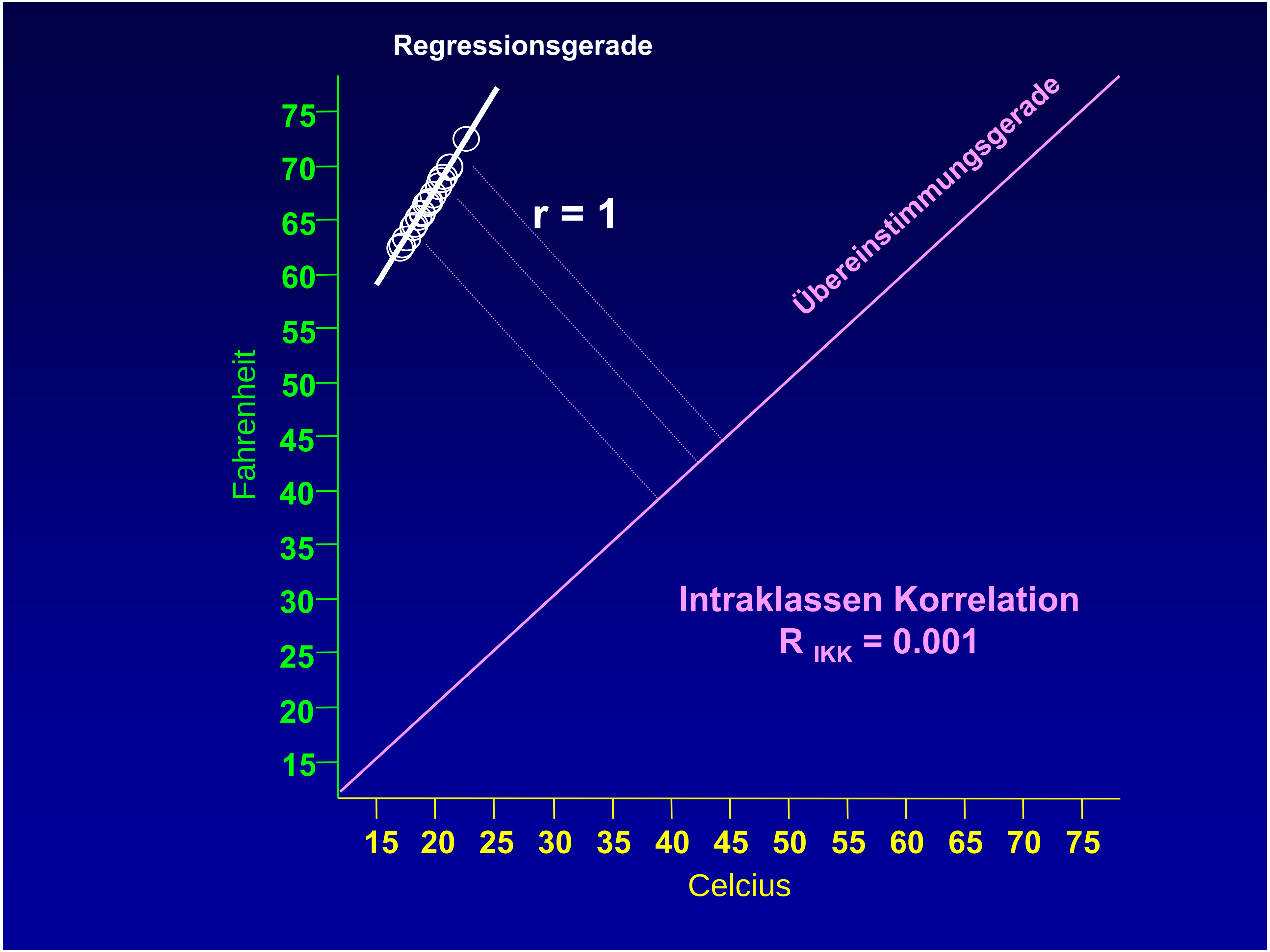
m = 19.3 s = 1.5

V_k

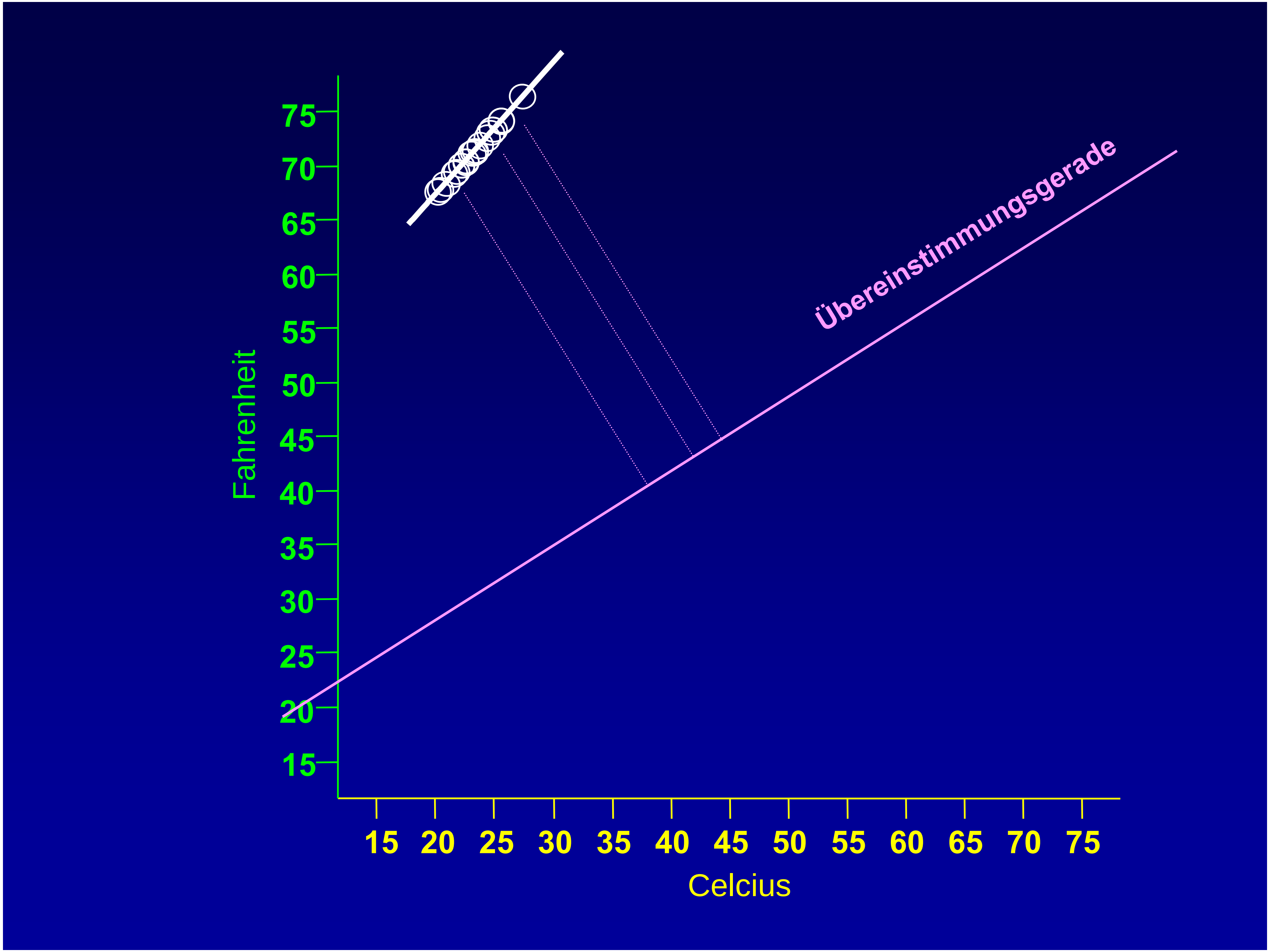
$$\frac{1.5}{19.3} = 7.7\ \%$$

$$\frac{2.6}{66.7} = 4.0\ \%$$

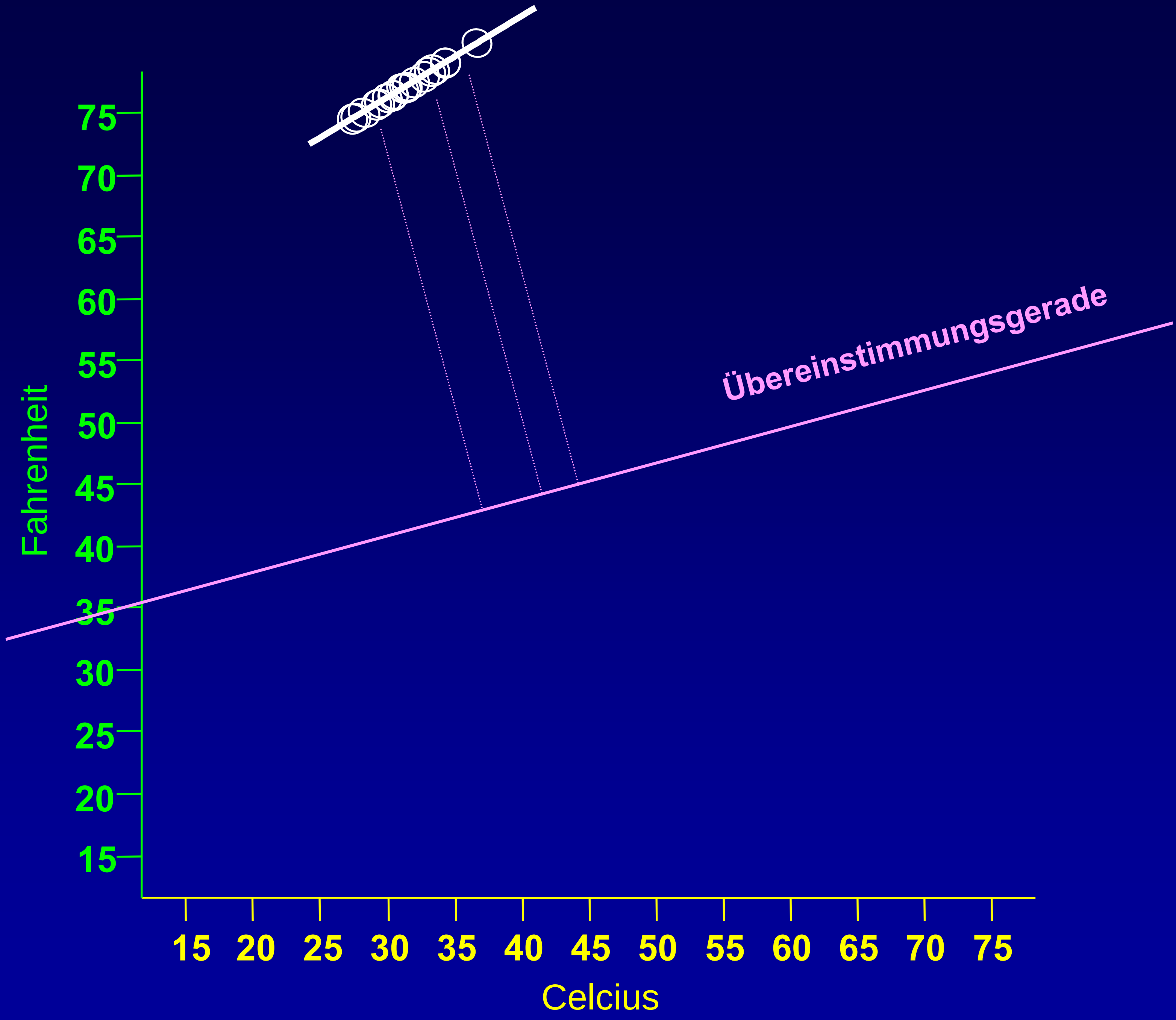

Fahrenheit
74
72
70
68
66
64
62
Regressionsgerade
Korrelation
r = 1
16
18
20
22
24
Celcius

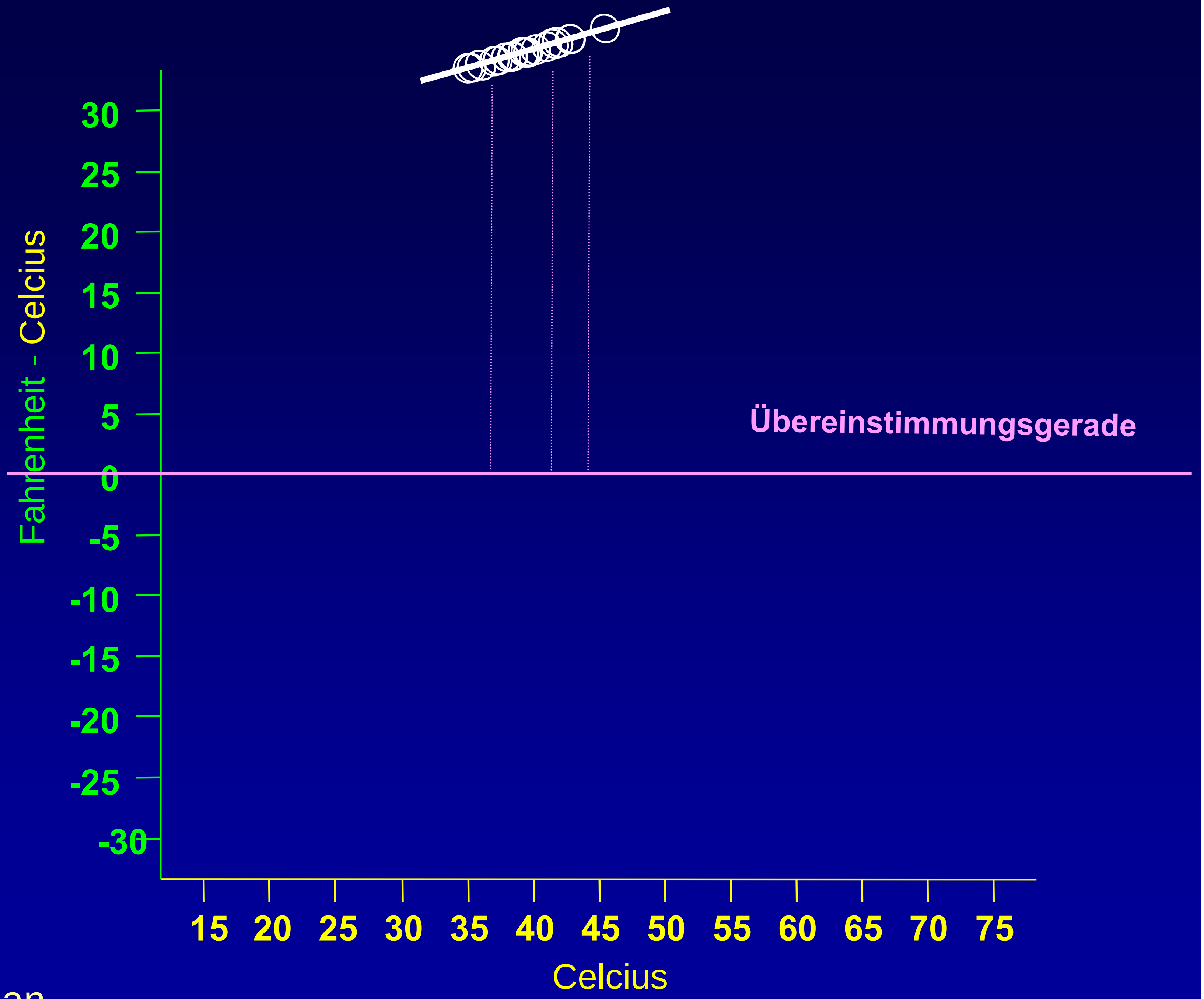


Regressionsgerade
r = 1
Übereinstimmungsgerade
Intraklassen Korrelation
R IKK = 0.001
Fahrenheit
75
70
65
60
55
50
45
40
35
30
25
20
15
15 20 25 30 35 40 45 50 55 60 65 70 75
Celcius



75
70
65
60
55
50
45
40
35
30
25
20
15
Fahrenheit
Übereinstimmungsgerade
15 20 25 30 35 40 45 50 55 60 65 70 75
Celcius


Übereinstimmungsgerade
Fahrenheit
75
70
65
60
55
50
45
40
35
30
25
20
15
15 20 25 30 35 40 45 50 55 60 65 70 75
Celcius



Bland Altman

Punktwolke mit Korrelation und Regression

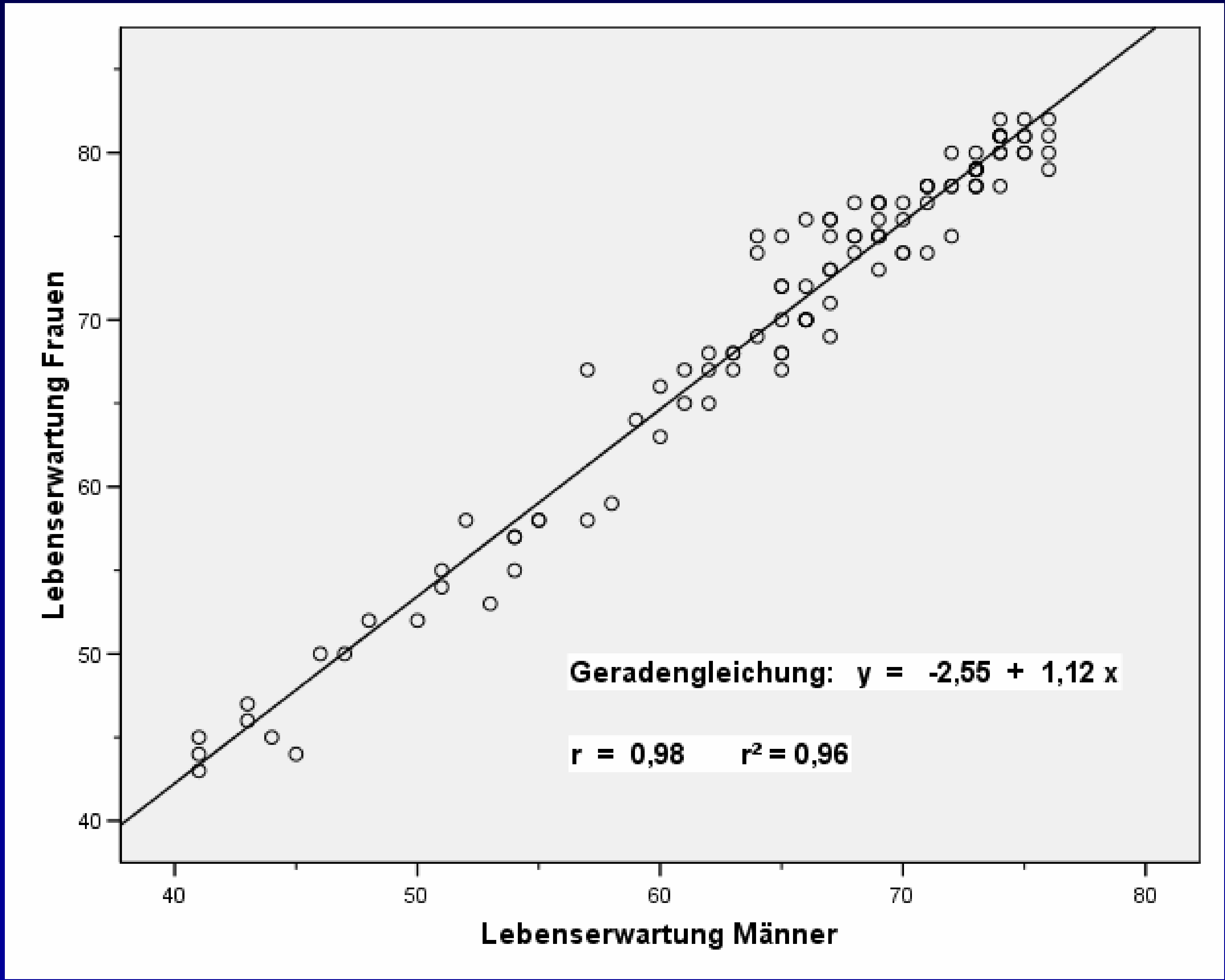
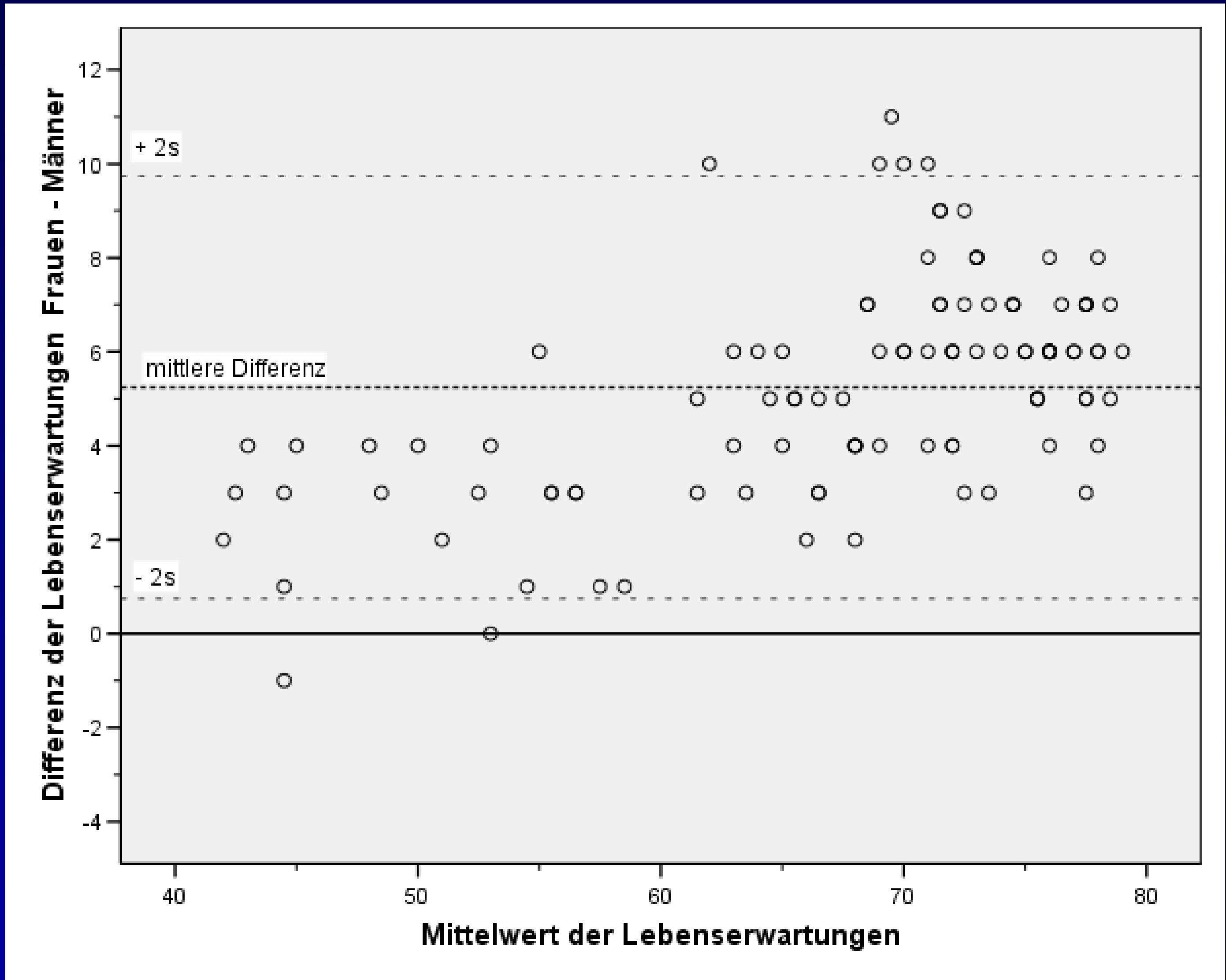


Diagramm nach Bland und Altman



Von *Altman* und *Bland* wurde eine graphische Methode entwickelt, die dem herkömmlichen Streudiagramm deutlich überlegen ist, allerdings vorausssetzt, dass beide Messungen auf derselben Skala erfolgen:

x-Achse: Mittelwert der Messungen ***y-Achse:*** Differenz der Messungen

Vorteile der Darstellung von Altman und Bland:

- Die Fläche des Diagramms wird besser genutzt
- Tatsächliche Abweichungen sind besser ablesbar
- Eine mögliche Abhängigkeit der Abweichungen von der Größenordnung der Werte ist besser erkennbar

Q1 Biometrie

Einführungsvorlesung

Wahrscheinlichkeitsrechnung

Beschreibende Statistik

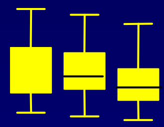
Wahrscheinlichkeitsrechnung

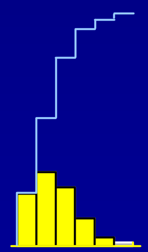
Stichprobe

Grundgesamtheit

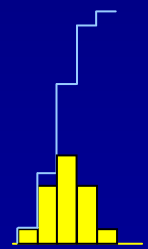
Stichprobe		Grundgesamtheit	
Lagemaße	$\bar{x}$ $\tilde{x}$	μ $\tilde{\mu}$	Erwartungswerte
Streuungsmaße	s, s^2	σ, σ^2	
rel.Häufigkeit	$0 \leq h(x) \leq 1$	$0 \leq p(x) \leq 1$	Wahrscheinlichkeit
Summenhäufigkeit	$0 \leq H(x) \leq 1$ Empirische Verteilungsfunktion	$0 \leq F(x) \leq 1$	Verteilungsfunktion

Box Plots
Error bars

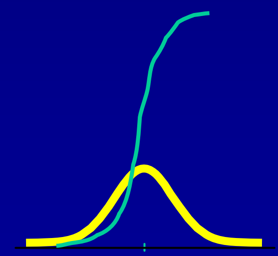




Histogramm



Wahrscheinlichkeits-
funktion



Dichtefunktion

Wahrscheinlichkeitsrechnung

Wenn das Ergebnis eines Experiments nicht genau vorhersagbar ist, spricht man von einem *Zufallsexperiment*.

Beispiele: Wurf einer Münze (Ergebnisse „Kopf“ oder „Zahl“)
Würfel (Ergebnisse 1,2,3,4,5,6).

Bei Zufallsexperimenten kann das genaue Ergebnis nicht vorhergesagt werden, man weiß aber, welche Ergebnisse überhaupt möglich sind. Die Ergebnisse werden in der Wahrscheinlichkeitsrechnung als *„Ereignisse“* bezeichnet. Die Zusammenfassung aller möglichen Ereignisse heißt *„Wahrscheinlichkeitsraum“*.

Wenn man dasselbe Zufallsexperiment unter exakt gleichen Bedingungen sehr oft wiederholt, wird man feststellen, dass jedes Ereignis prozentual mit einer bestimmten (relativen) Häufigkeit auftritt. Diese Häufigkeit wird dann als *Wahrscheinlichkeit* für dieses Ereignis angesehen. Die Zusammenfassung aller Wahrscheinlichkeiten heißt *Wahrscheinlichkeitsverteilung*.

In Rahmen der mathematischen Wahrscheinlichkeitsrechnung werden diese Annahmen durch die *Axiome von Kolmogoroff* formalisiert.

Grundlagen der Wahrscheinlichkeitsrechnung

Bezeichnungen

Ω	Zusammenfassung aller Ereignisse (Wahrscheinlichkeitsraum)
A,B,C, ...	Ereignisse (Teilmengen von Ω)
$A \cup B$	Vereinigung von A und B (A *oder* B wird beobachtet)
$A \cap B$	Schnitt von A und B (A *und* B werden beobachtet)
P	Wahrscheinlichkeitsverteilung
P (A)	Wahrscheinlichkeit von Ereignis A

Axiome der Wahrscheinlichkeitsrechnung
A1

Definitionsbereich der Wahrscheinlichkeit: $0 \leq p(E) \leq 1$

jedem zufälligen Ereignis E wird eine Zahl p(E) zugeordnet,
die zwischen 0 und 1 liegt

p(E) wird als Wahrscheinlichkeit des Ereignisses E bezeichnet

Axiome der Wahrscheinlichkeitsrechnung
A2

Richtung des Definitionsbereiches $p(S) = 1$

S = sicheres Ereignis

Daraus folgt unmittelbar, dass das unmögliche Ereignis die Wahrscheinlichkeit 0 hat.

Komplementäres Ereignis

p (Diabetes) = 0.2 d.h. p (kein Diabetes) = 1 – 0.2 = 0.8

Axiome der Wahrscheinlichkeitsrechnung A3

Additionssatz

p (A oder B) = p (A) + p (B) - p (A und B)

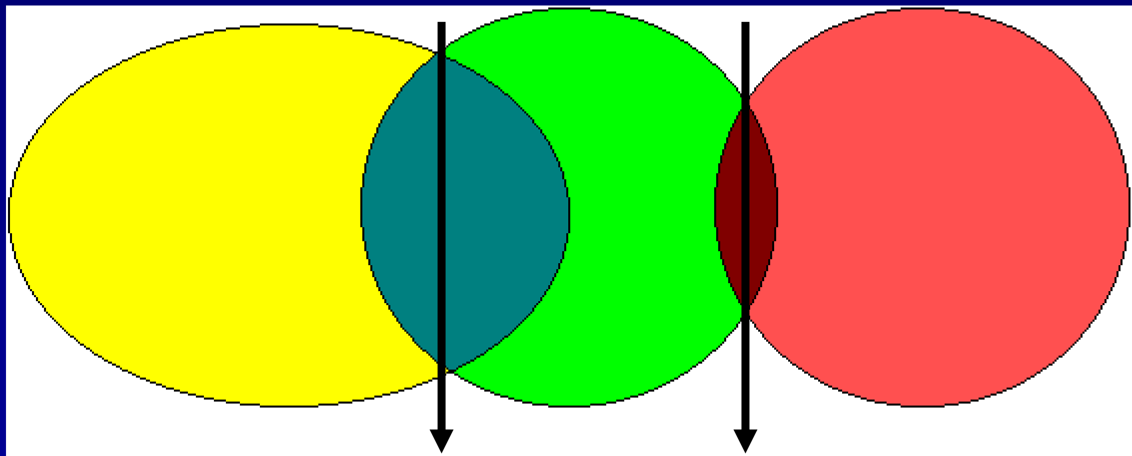
$\cup$ $\cap$

Zentrales Problem der histologischen Schilddrüsendiagnostik

Hyperplastischer Knoten

Adenom

Karzinom



Problemfälle in den Überschneidungsflächen

Multiplikationssatz

p (A und B) = p (A) · p (B)

$\cap$

A und B sind genau dann unabhängig, wenn diese Gleichung gilt.

p (Streptokokken im Abstrich) = 0.8

p (S) =0,8 d.h p (1-S) = p(s) = 0,2

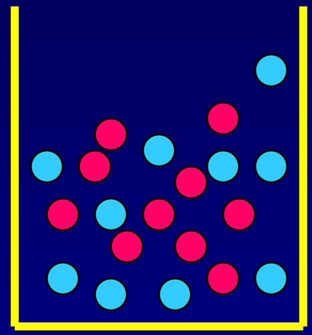
Wie groß ist die Wahrscheinlichkeit bei 2 Abstrichen ?

	SS	Ss	sS	ss
p	$0{,}8 \cdot 0{,}8$	$0{,}8 \cdot 0{,}2$	$0{,}2 \cdot 0{,}8$	$0{,}2 \cdot 0{,}2$
	0,64	**0,16**	**0,16**	**0,04**

0,96

Wahrscheinlichkeit

10 10



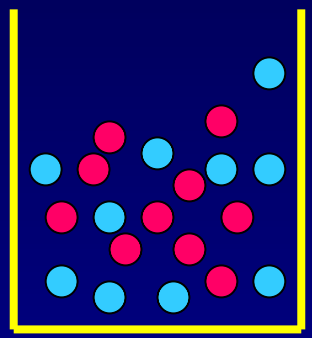
$$P\ (\text{rot}) = \frac{10}{20}$$

$$\text{Chance} = \frac{10}{10}$$

(odds)

Wahrscheinlichkeit

10 10



Modell mit Zurücklegen

$$P(\text{rot}) = \frac{10}{20}$$

Modell ohne Zurücklegen

$$P(\text{rot} / \text{rot}) = \frac{9}{19}$$

$$P(\text{rot} / \text{blau}) = \frac{10}{19}$$

Die Prävalenz ist der Anteil der Erkrankten
zu einem bestimmte Zeitpunkt (Punktprävalenz)
oder innerhalb eines gewissen Zeitraums (Periodenprävalenz)

$$\textbf{Prävalenz} = \frac{\textbf{Anzahl der Erkrankten}}{\textbf{Gesamtpopulation}}$$

- längere Krankheitsdauer
- Lebensverlängerung von Patienten die nicht geheilt werden
- Zunahme der Inzidenz
- Zuwanderung von erkrankten oder anfälligen Personen
- Abwanderung von Gesunden
- Bessere diagnostische Methoden

- kürzere Krankheitsdauer
- Hohe Letatlitätarate
- Abnahme der Inzidenz
- Abwanderung von erkrankten oder anfälligen Personen
- Zuwanderung von Gesunden
- höhere Heilungsrate der Erkrankten

Die Inzidenz zählt die Neuerkrankungen innerhalb eines vorgegebenen Zeitraums und setzt sie in Bezug zur nicht erkrankten Bevölkerung am Anfang des Zeitraums

$$\textbf{Inzidenz} = \frac{\textbf{Anzahl der Neuerkrankungen}}{\textbf{Bezugsbevölkerung}}$$

Wahrscheinlichkeit, an verschiedenen Tagen Geburtstag zu haben

n = 2 $$p_2 = \frac{365}{365} \cdot \frac{364}{365} = 0.9972$$

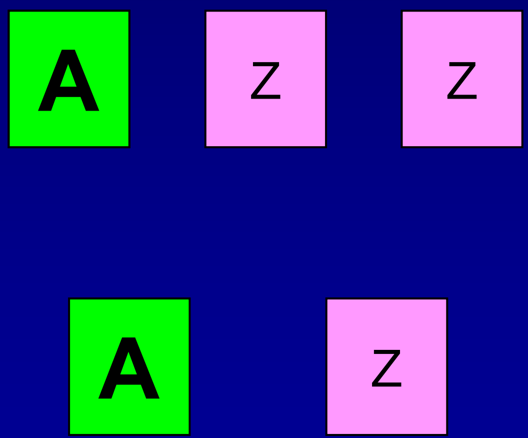
n = 3 $$p_3 = \frac{365}{365} \cdot \frac{364}{365} \cdot \frac{363}{365} = 0.9917$$

n = 4 $$p_4 = \frac{365}{365} \cdot \frac{364}{365} \cdot \frac{363}{365} \cdot \frac{362}{365} = 0.9836$$

n = 25 $$p_{25} = \frac{365}{365} \cdot \frac{364}{365} \cdot \frac{363}{365} \cdot \frac{362}{365} \cdots \quad \cdot\cdot \frac{342}{365} \cdot \frac{341}{365} = 0.4321$$

n = 50 $$p_{50} = 0.03$$

Monty Hall und Marilyn vos Savant - Lets make a deal



1.Wahl: 1 von 3

2.Wahl: 1 von 2

1	2	3	Tor	bleiben	wechseln
A	Z	Z	1	1	0
A	Z	Z	2	0	1
A	Z	Z	3	0	1
			Chance	**1**	**2**
			p	**1/3**	**2/3**
				P (A | B)	P (A | W)

Zitate von Wissenschaftlern an Universitäten

Es gibt genügend mathematischen Unverstand in der Welt, und die Inhaberin des höchsten Intelligenzquotienten braucht seiner Verbreitung nicht noch Vorschub zu leisten. Schämen Sie sich.

Sie haben Unsinn verzapft. Als Mathematiker bin ich sehr besorgt über das verbreitete Unwissen in mathematischen Dingen.
Geben Sie Ihren Fehler zu und seien Sie in Zukunft vorsichtiger.

Ihre Lösung steht klar im Widerspruch zur Wahrheit

Darf ich den Vorschlag machen, dass Sie zunächst einmal in ein Standard- Lehrbuch Der Wahrscheinlichkeitsrechnung schauen, bevor Sie das nächste Mal versuchen, Ein derartiges Problem zu lösen.

Wie viele entrüstete Mathematiker braucht es, bis Sie endlich Ihre Meinung ändern.

Vielleicht gehen Frauen mathematische Probleme anders an als Männer.

Sie haben Unrecht. Bedenken Sie doch: Wenn sich alle diese Professoren und Doktoren irren würden, stünde es sehr schlecht um unser Land.

Wirkung bei Frauen

	ja	nein		p (Wirkung)
A	20	5	25	20 / 25 oder 80%
B	70	30	100	70 / 100 oder 70%

Wirkung bei allen

	ja	nein		p (Wirkung)
A	50	75	125	50 / 125 oder 40%
B	75	50	125	75 / 125 oder 60%

Wirkung bei Männern

	ja	nein		p (Wirkung)
A	30	70	100	30 / 100 oder 30%
B	5	20	25	5 / 25 oder 20%

Wirkung bei Frauen

	ja	nein		Chance - odds
A	20	5	25	20 / 5
B	70	30	100	70 / 30

Odds ratio A/B

$$\frac{20}{5} : \frac{70}{30} = \frac{20}{5}\ \frac{30}{70} = 1{,}7$$

Wirkung bei allen

	ja	nein		Chance - odds
A	50	75	125	50 / 75
B	75	50	125	75 / 50

Odds ratio A/B

$$\frac{50}{75} : \frac{75}{50} = \frac{50}{75}\ \frac{50}{75} = 0{,}33$$

Wirkung bei Männern

	ja	nein		Chance - odds
A	30	70	100	30 / 70
B	5	20	25	5 / 20

Odds ratio A/B

$$\frac{30}{70} : \frac{5}{20} = \frac{30}{70}\ \frac{20}{5} = 1{,}7$$

Wahrscheinlichkeitsrechnung und Medizin

Diagnose und *Therapie* sind natürlich keine Zufallsexperimente wie der Wurf einer Münze oder eines Würfels. Dennoch weiß man, dass nicht jede Diagnose korrekt gestellt wird oder jede Therapie erfolgreich verläuft.

Ursachen für *falsche diagnostische Entscheidungen:*
Messfehler, Fehlinterpretation, unterschiedlich ausgeprägte Krankheitsbilder ...

Ursachen für *Misserfolge in der Therapie:*
fehlerhafte Durchführung der Therapie, unterschiedliche Response der Patienten ...

Diese Unsicherheit lassen sich mit Hilfe der *Wahrscheinlichkeitsrechnung* quantifizieren.
Im folgenden Beispiel geht man davon aus, dass eine Therapie bei *allen Patienten* die gleiche Heilungswahrscheinlichkeit p hat. Dann kann man auch ausrechnen, wie groß die Wahrscheinlichkeit ist, dass in einer Gruppe von Patienten ein bestimmter Prozentsatz dieser Patienten geheilt wird.

Die Binomialverteilung

Situation:
Ein Medikament hat Heilungswahrscheinlichkeit **p**
Das Medikament wird **n** Patienten verabreicht.

Mit welcher Wahrscheinlichkeit werden **k** *= 0, 1, ...n Patienten geheilt ?*

Beispiel: n = 4 Patienten, p = 0.8 Heilungswahrscheinlichkeit
Die *möglichen Ergebnisse* sind 0, 1, 2, 3 oder 4 Heilungen.
Als *Wahrscheinlichkeitsraum* ist daher $\Omega_1 = \{0, 1, 2, 3, 4\}$ geeignet.

Es zeigt sich aber, dass man zur Berechnung der Wahrscheinlichkeiten für Ω_1 einen Umweg gehen muss, bei dem man auch berücksichtigt, *welche Patienten* geheilt werden und welche nicht.

Im folgenden Ω_2 bedeutet beispielsweise die Sequenz 1011, dass der erste, der dritte und der vierte Patient geheilt wurden. Die Wahrscheinlichkeit für eine derartige Sequenz *unabhängiger Ereignisse* berechnet sich jeweils als Produkt der Einzelereignisse.

$\Omega_2 = \{0\,0\,0\,0$ 0 Heilungen $p = 0.2 \cdot 0.2 \cdot 0.2 \cdot 0.2$

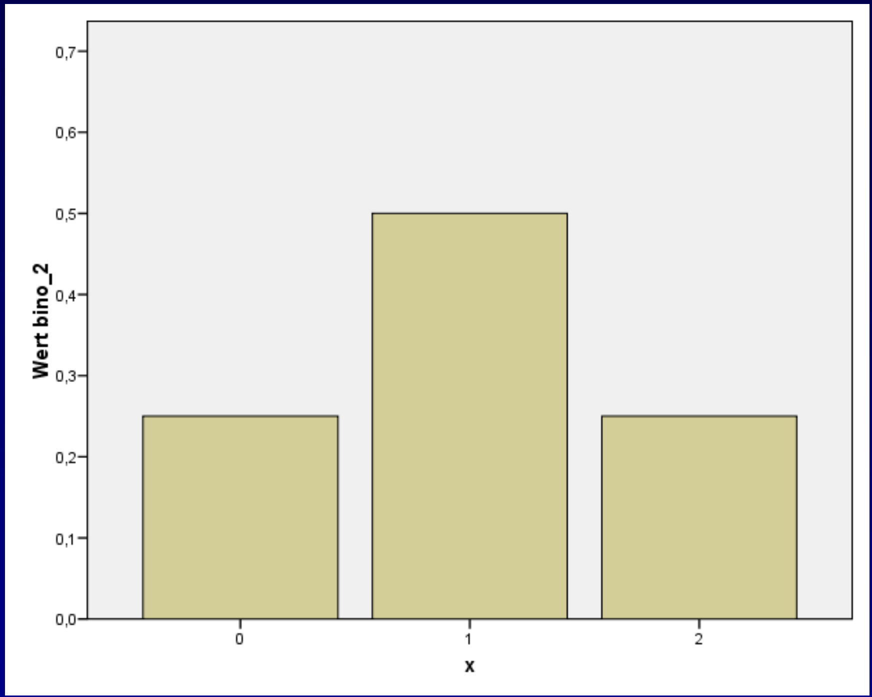
0 0 0 1
0 0 1 0
0 1 0 0
1 0 0 0 1 Heilung, p *jeweils* $0.8 \cdot 0.2 \cdot 0.2 \cdot 0.2$

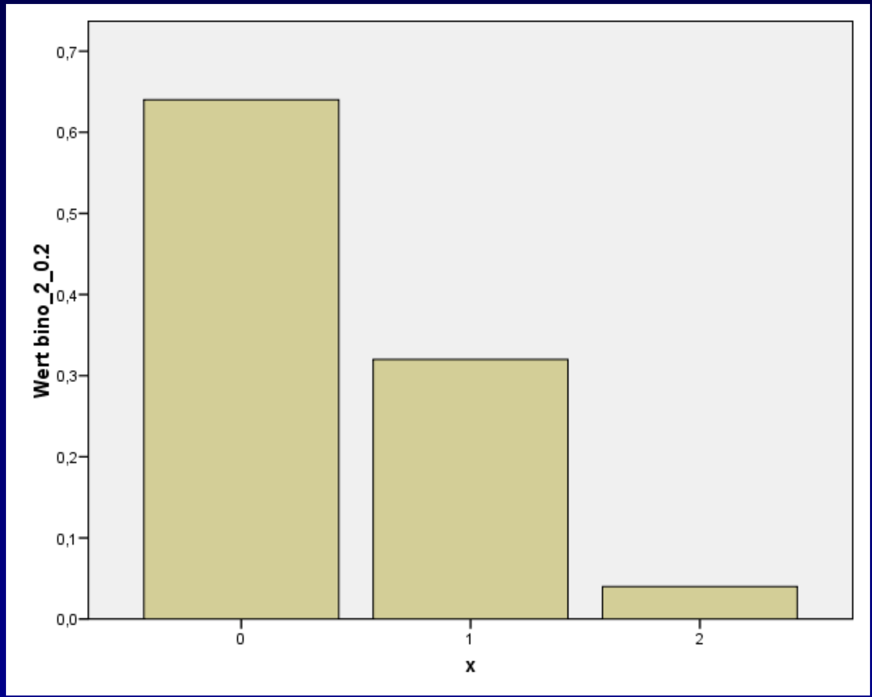
0 0 1 1
0 1 0 1
1 0 0 1
0 1 1 0
1 0 1 0
1 1 0 0 2 Heilungen, p *jeweils* $0.8 \cdot 0.8 \cdot 0.2 \cdot 0.2$

0 1 1 1
1 0 1 1
1 1 0 1
1 1 1 0 3 Heilungen, p *jeweils* $0.8 \cdot 0.8 \cdot 0.8 \cdot 0.2$

1 1 1 1 $\}$ 4 Heilungen $p = 0.8 \cdot 0.8 \cdot 0.\ 8 \cdot\ 0.8$

$$B(n/k,p) = \binom{n}{k} p^k \ (1-p)^{(n-k)}$$





$$P(MM) = P(M)\ P(M) = \frac{1}{2}\cdot\frac{1}{2} = 0.25$$

$$P(JM) = P(J)\ P(M) + P(M)\ P(J) = \frac{1}{2}\cdot\frac{1}{2} + \frac{1}{2}\cdot\frac{1}{2} = 0.5$$

$$P(JJ) = P(J)\ P(J) = \frac{1}{2}\cdot\frac{1}{2} = 0.25$$

$$P(\overline{D}\overline{D}) = P(\overline{D})\ P(\overline{D}) = 0{,}8\cdot 0{,}8 = 0.64$$

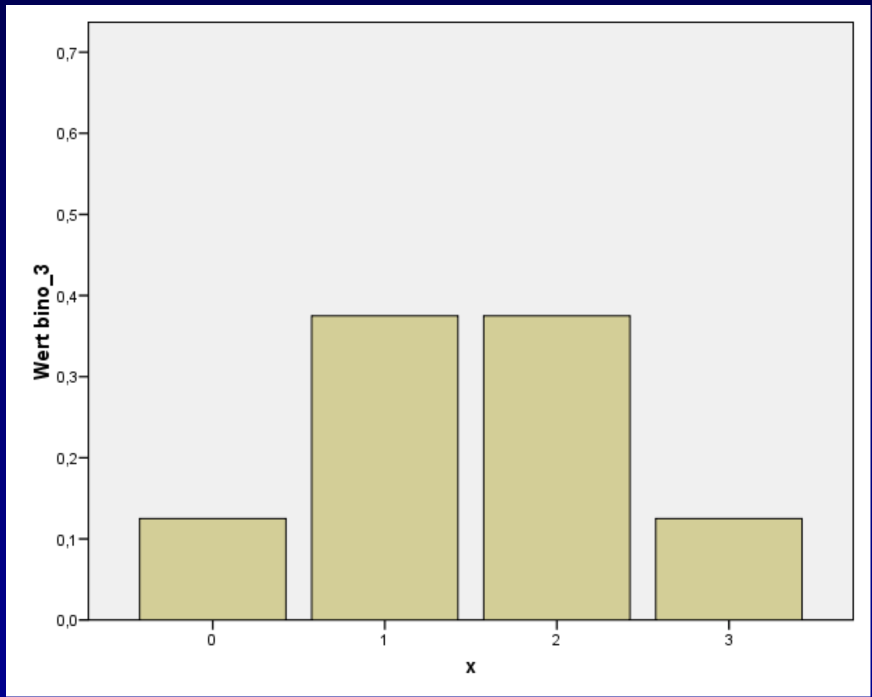
$$P(D\overline{D}) = P(D)\ P(\overline{D}) + P(\overline{D})\ P(D) = 0{,}2\cdot 0{,}8 + 0{,}8\cdot 0{,}2 = 0.32$$

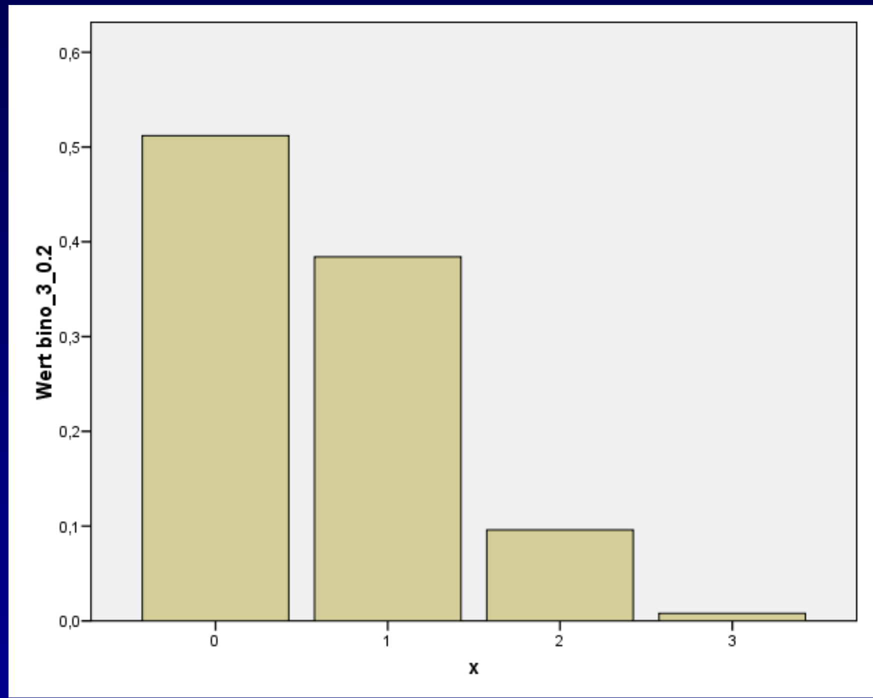
$$P(DD) = P(D)\ P(D) = 0{,}2\cdot 0{,}2 = 0{,}04$$

$$B(2/k, \tfrac{1}{2}) = \binom{2}{k}\left(\frac{1}{2}\right)^2$$

$$B(2/k, 0.2) = \binom{2}{k} 0.2^k \ 0.8^{2-k}$$

$$B(n/k,p) = \binom{n}{k} p^k \; (1-p)^{(n-k)}$$

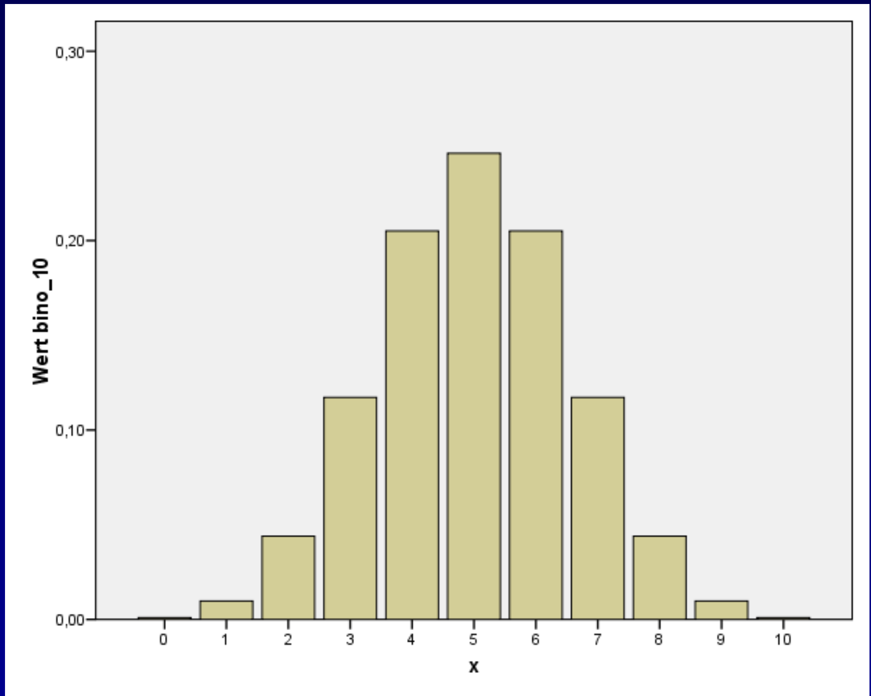


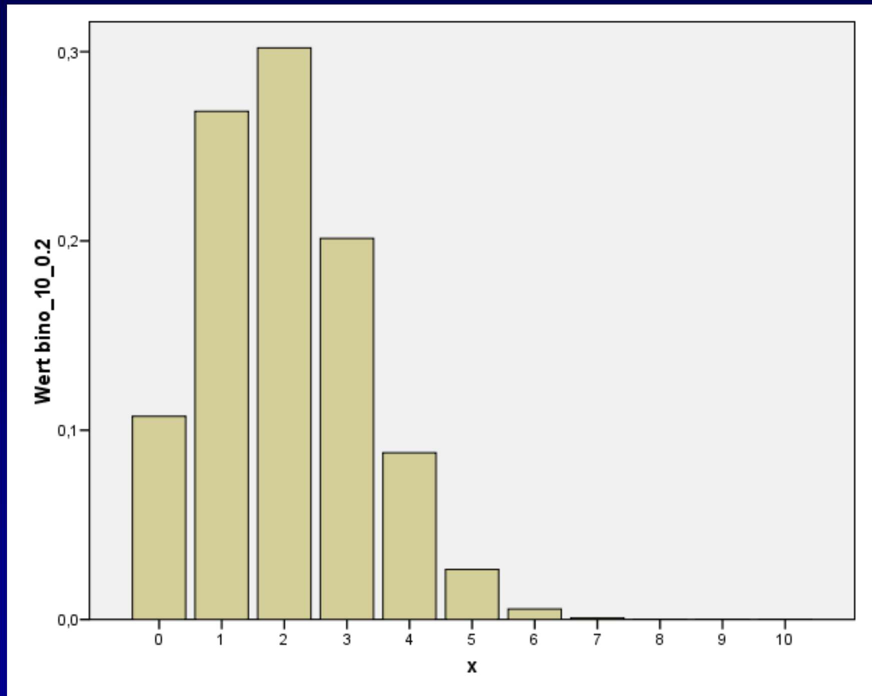


$$B(3/k,\tfrac{1}{2}) = \binom{3}{k} \left(\frac{1}{2}\right)^3$$

$$B(3/k,0.2) = \binom{3}{k} 0.2^k \; 0.8^{3-k}$$

$$B(n/k,p)=\binom{n}{k}\,p^k\,(1-p)^{(n-k)}$$

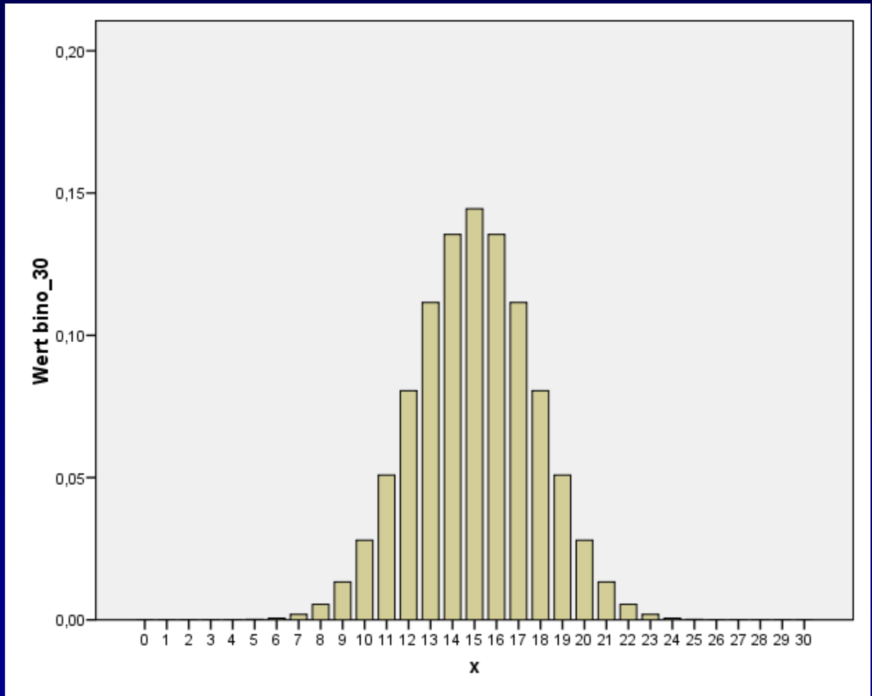


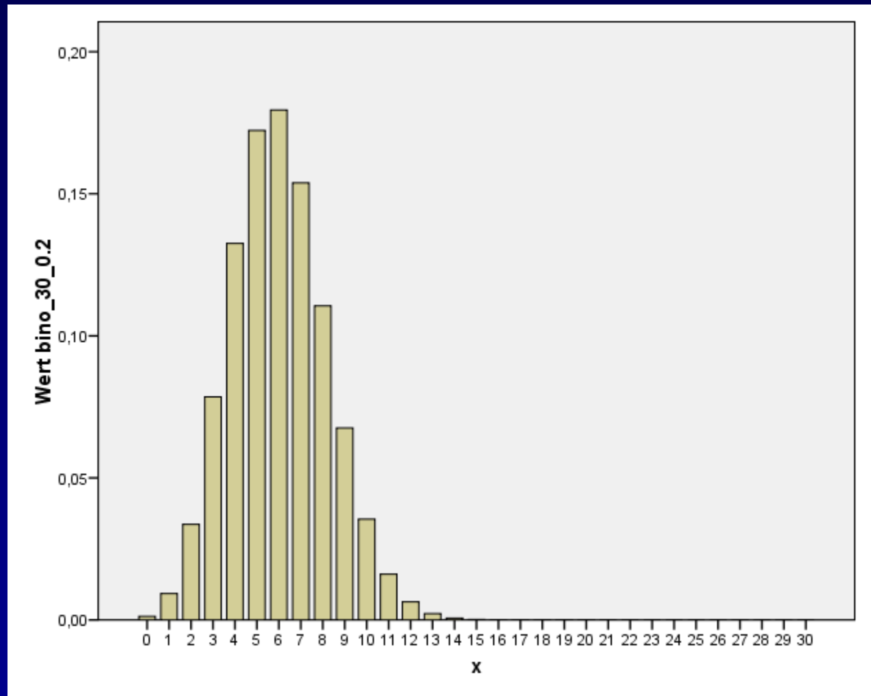


$$B(10/k,\tfrac{1}{2})=\binom{10}{k}\left(\frac{1}{2}\right)^{10}$$

$$B(10/k,0.8)=\binom{10}{k}\,(0{,}2)^k\,(0{,}8)^{10-k}$$

$$B(n/k,p)=\binom{n}{k}\,p^k\;(1-p)^{(n-k)}$$





$$B(30/k,\tfrac{1}{2})=\binom{30}{k}\left(\frac{1}{2}\right)^{30}$$

$$B(30/k,0.8)=\binom{30}{k}\,(0,2)^k\;(0,8)^{30-k}$$

Feste Reihenfolge, Wahrscheinlichkeit für

0 Heilungen:	$0.2 \cdot 0.2 \cdot 0.2 \cdot 0.2$	=	0.0016
1 Heilung:	$0.8 \cdot 0.2 \cdot 0.2 \cdot 0.2$	=	0.0064
2 Heilungen:	$0.8 \cdot 0.8 \cdot 0.2 \cdot 0.2$	=	0.0256
3 Heilungen:	$0.8 \cdot 0.8 \cdot 0.8 \cdot 0.2$	=	0.1024
4 Heilungen:	$0.8 \cdot 0.8 \cdot 0.8 \cdot 0.8$	=	0.4096

Beliebige Reihenfolge, Wahrscheinlichkeit für

0 Heilungen:	$\mathbf{1} \cdot 0.0016$	=	0.0016
1 Heilung:	$\mathbf{4} \cdot 0.0064$	=	0.0256
2 Heilungen:	$\mathbf{6} \cdot 0.0256$	=	0.1536
3 Heilungen:	$\mathbf{4} \cdot 0.1024$	=	0.4096
4 Heilungen:	$\mathbf{1} \cdot 0.4096$	=	0.4096

Gesamtsumme: **1.000**

Die Binomialverteilung

Die Wahrscheinlichkeit,

genau **k** Patienten zu heilen,
wenn **n** Patienten behandelt werden
und die Heilungsrate für einen Patienten **p** beträgt,

ist

$$\boxed{\binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}}$$

Zugehöriges Ω: **{0, 1, 2, ..., n}**

Allgemeine Formel der Möglichkeiten für k Heilungen bei n Patienten

Binomialkoeffizient:

$$\boxed{\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}}$$

Fakultät:

$$\boxed{n! = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 1}$$

Beispiel:

$$\boxed{4! = 4 \cdot 3 \cdot 2 \cdot 1 = 24}$$

Pascal Dreieck

1

1 1

1 2 1

1 3 3 1

1 4 6 4 1

1 5 10 10 5 1

1 6 15 20 15 6 1

1 7 21 35 35 21 7 1

1 8 28 56 70 56 28 8 1

1 9 36 84 126 126 84 36 9 1

1 10 45 120 210 252 210 120 45 10 1

$$\binom{0}{0}$$

$$\binom{1}{0} \quad \binom{1}{1}$$

$$\binom{2}{0} \quad \binom{2}{1} \quad \binom{2}{2}$$

$$\binom{3}{0} \quad \binom{3}{1} \quad \binom{3}{2} \quad \binom{3}{3}$$

$$\binom{4}{0} \quad \binom{4}{1} \quad \binom{4}{2} \quad \binom{4}{3} \quad \binom{4}{4}$$

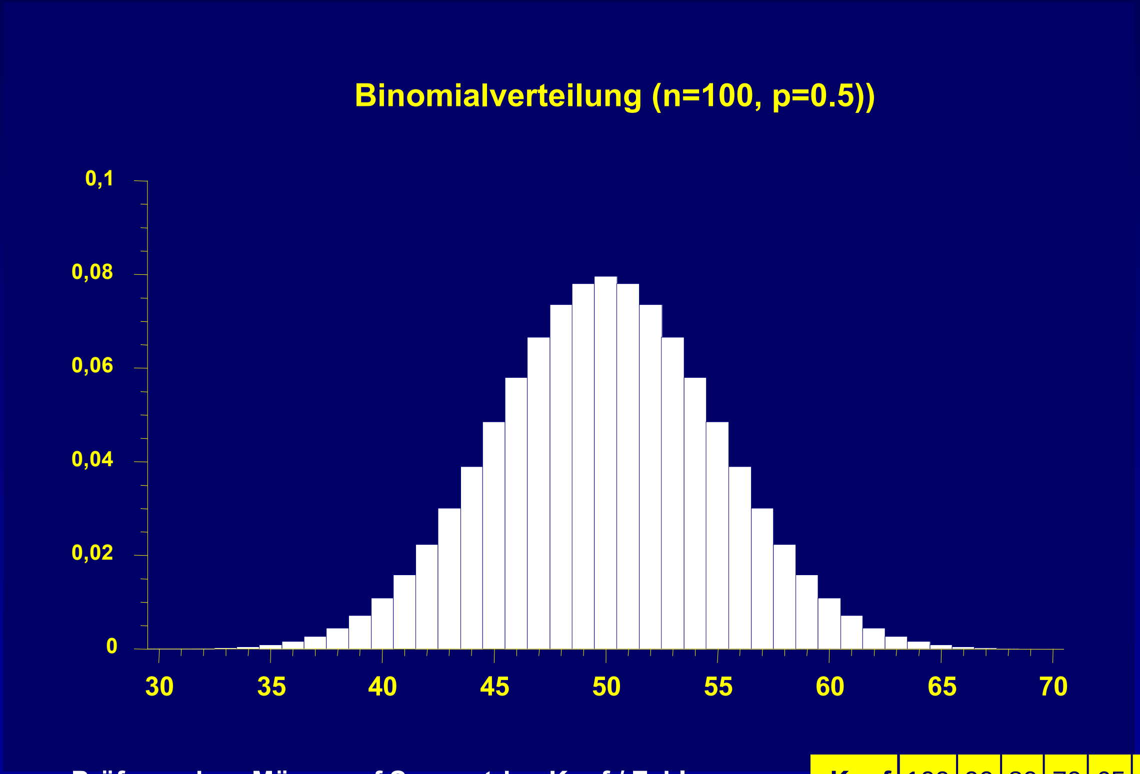
$$\binom{5}{0} \quad \binom{5}{1} \quad \binom{5}{2} \quad \binom{5}{3} \quad \binom{5}{4} \quad \binom{5}{5}$$

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$$

Binomialkoeffizienten

$$B(n/k,p) = \binom{n}{k} p^k \; (1-p)^{(n-k)}$$

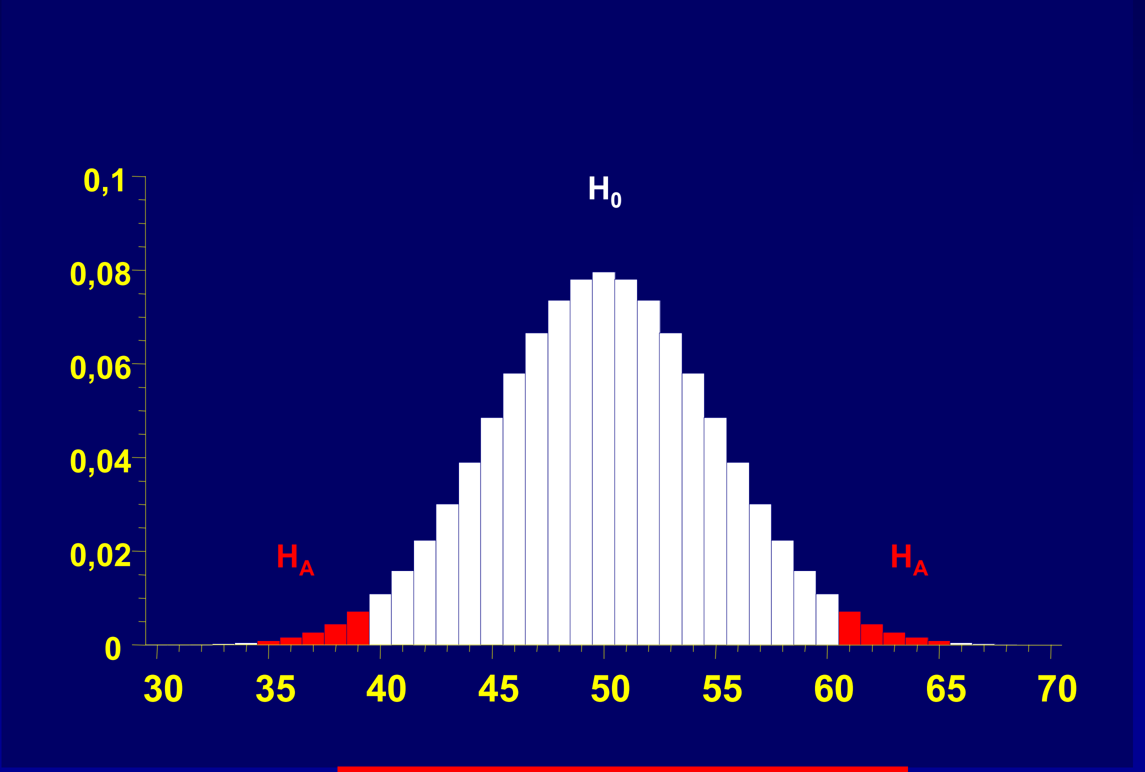
$$B(100/k, \tfrac{1}{2}) = \binom{100}{k} \left(\frac{1}{2}\right)^k \left(1-\frac{1}{2}\right)^{100-k} = \binom{100}{k} \left(\frac{1}{2}\right)^{100}$$



Prüfung einer Münze auf Symmetrie: Kopf / Zahl
Prüfung einer blutdrucksenkenden Wirkung: ja / nein

Kopf	100	90	80	70	65	60	55	50 ...
Zahl	0	10	20	30	35	40	45	50 ...

Binomialverteilung (n=100, p=0.5)



α
Signifikanzniveau
Fehler 1.Art
α = 5% oder 1%

Q1 Biometrie

Einführungsvorlesung

Konfirmatorische Statistik

Konfirmatorische Statistik

In der *konfirmatorischen* Statistik werden aus den Ergebnissen einer Stichprobe *Rückschlüsse* auf die Grundgesamtheit gezogen. Dies geschieht auf zwei verschiedene Arten:

- **„Konfidenzintervall“**: Angabe eines Wertebereichs für die Grundgesamtheit z.B. „Die Rezidivrate beim Kolorektalen Karzinom liegt nach zwei Jahren zwischen 12% und 28%.“
- **„Hypothesentest“**: Angabe einer ja/nein Entscheidung für die Grundgesamtheit z.B. „Die Rezidivrate liegt sicher unter 40%.“

Beide Rückschlüsse sind nicht fehlerfrei! Die wahre Rezidivrate kann außerhalb des Intervalls 12%-28% liegen, sie kann sogar über 40% liegen. Dann wäre es aber sehr unwahrscheinlich, in einer Studie mit 150 Patienten nur 20% Rezidive zu finden. In der konfirmatorischen Statistik präzisiert man, was mit „unwahrscheinlich“ gemeint ist.

Dafür benötigt man die Methoden der Wahrscheinlichkeitsrechnung.

Klinische Studie mit 100 Patienten

In einer Studie werden *100 Patienten* mit einem neuen Medikament behandelt. Das herkömmliche Präparat hat eine Heilungsrate von *50%*. Ist das neue Medikament besser?

Dieses Beispiel wurde bewusst sehr einfach gewählt, um das Prinzip des statistischen Testens zu erläutern. Es ist in folgender Hinsicht nicht realistisch:

- Man sollte die Heilungsrate des herkömmlichen Medikaments nicht als gegeben annehmen, sondern eine *Kontrollgruppe* mitführen, in der die herkömmliche Therapie verabreicht wird.
- Die Zuordnung zum neuen Medikament bzw. zur Kontrollgruppe sollte nach dem *Zufallsprinzip* (randomisiert) erfolgen.
- Die *Fallzahl* der Studie ist möglicherweise nicht realistisch.
- Das *Zielkriterium* „Heilung ja/nein“ müsste genauer präzisiert werden.

Berechnung der Heilungswahrscheinlichkeiten

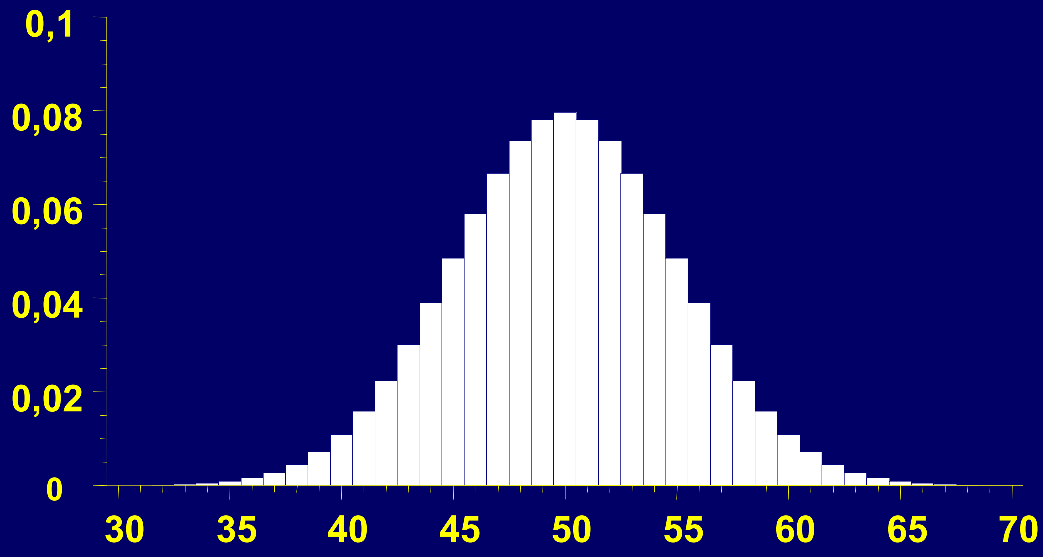
In der folgenden Graphik sind die Wahrscheinlichkeiten dafür angegeben, dass mit dem herkömmlichen Präparat von den 100 Patienten exakt 0, 1, ..., 100 Patienten geheilt werden.

Auf der x-Achse steht die Zahl der geheilten Patienten, auf der y-Achse entspricht die Höhe des Balkens der jeweiligen Wahrscheinlichkeit.

Die Berechnungen erfolgten mit Hilfe der Formel für die Binomialverteilung

Die größten Einzelwahrscheinlichkeiten (ca. 8%) werden für 50 Patienten berechnet, was einleuchtet, da 50% von 100 genau 50 ergibt.

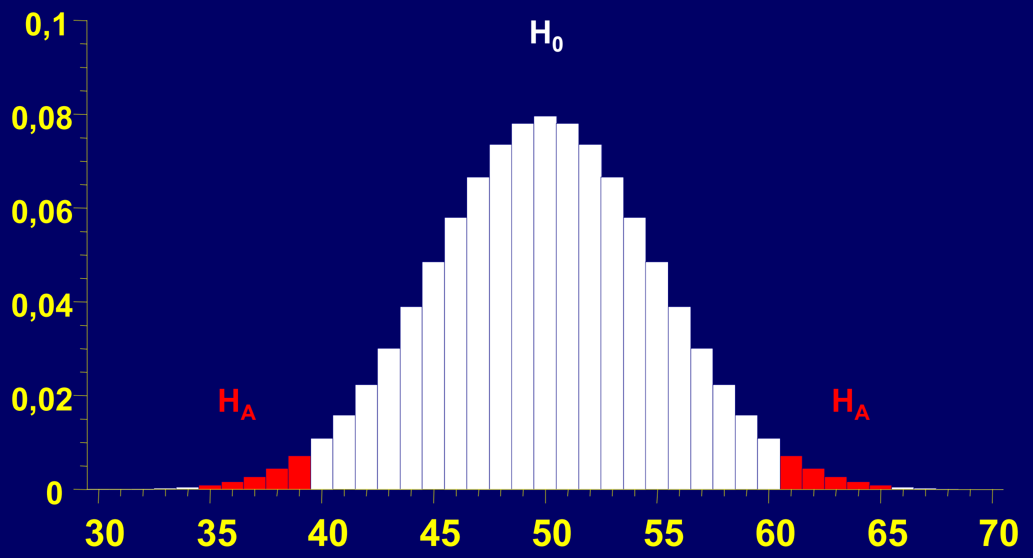
Binomialverteilung (n=100, p=0.5))



ja	100	90	80	70	60	50
nein	0	10	20	30	40	50

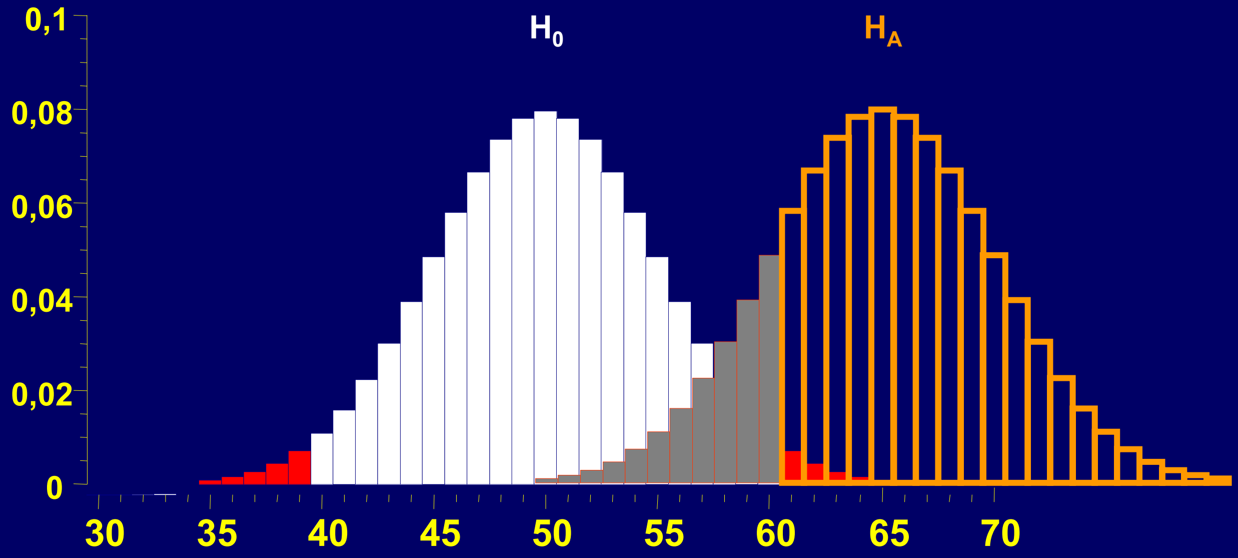
Prüfung einer Wirksamkeit: geheilt ja / nein

Binomialverteilung   (n=100, p=0.5)



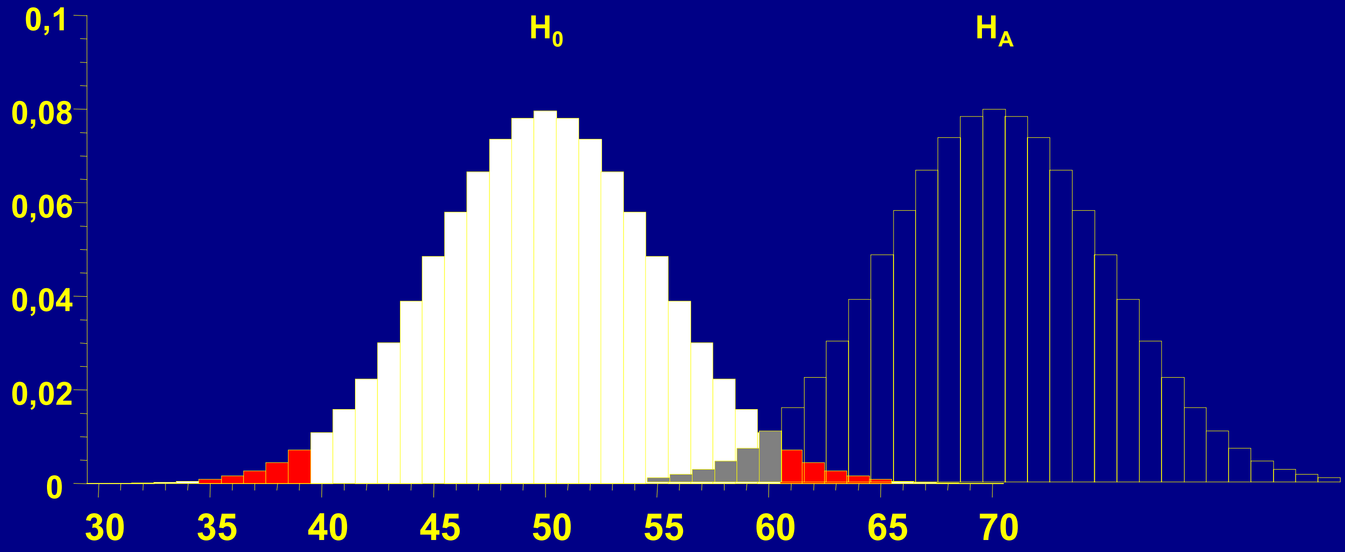
α
Signifikanzniveau
Fehler 1.Art
α = 5% oder 1%

Binomialverteilung   (n=100, p=0.5)



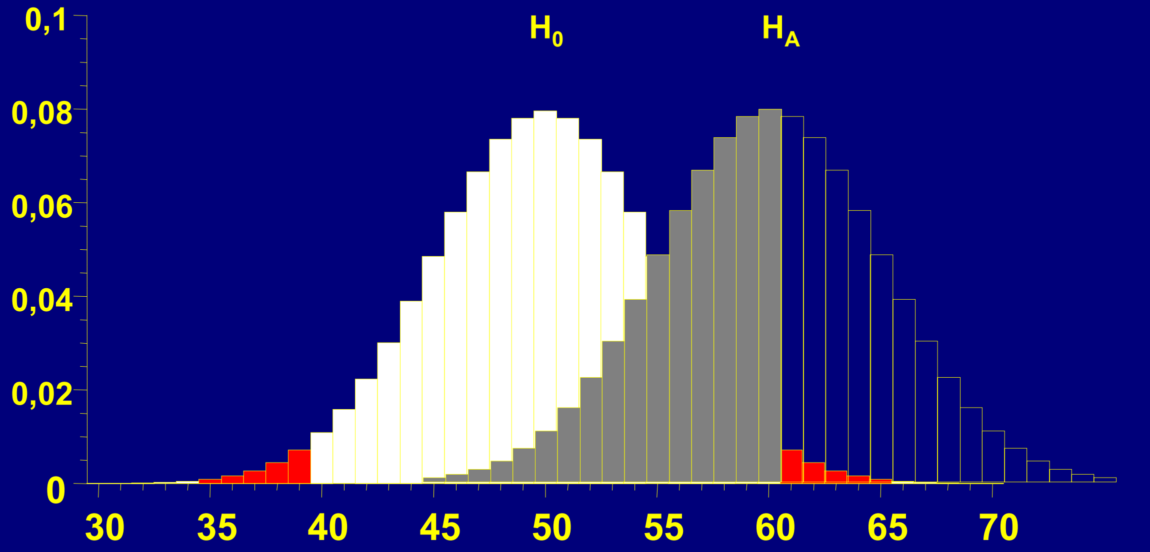
Fehler 2.Art   β

Binomialverteilung (n=100, p=0.5)



Fehler 2.Art β

Binomialverteilung (n=100, p=0.5))



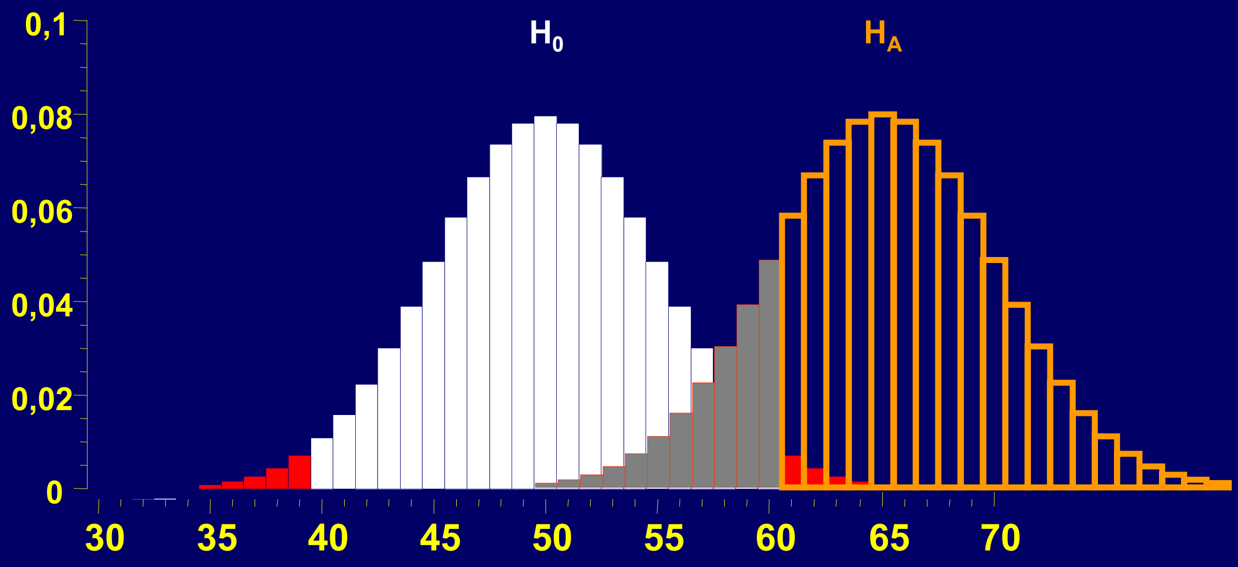
Fehler 2.Art β

Die Erfolgsaussichten einer Studie hängen von der Fallzahl ab

Mit 15 Patienten hätte man nur geringe Chancen gehabt, die Nullhypothese p=0.5 abzulehnen, wenn die tatsächliche Alternative „nur“ p=0.6 beträgt. Bei 1ßß Patienten ist diese Chance schon größer. Um den Unterschied 60% vs 50% nachweisen zu können, wird eine *größere Fallzahl* benötigt.

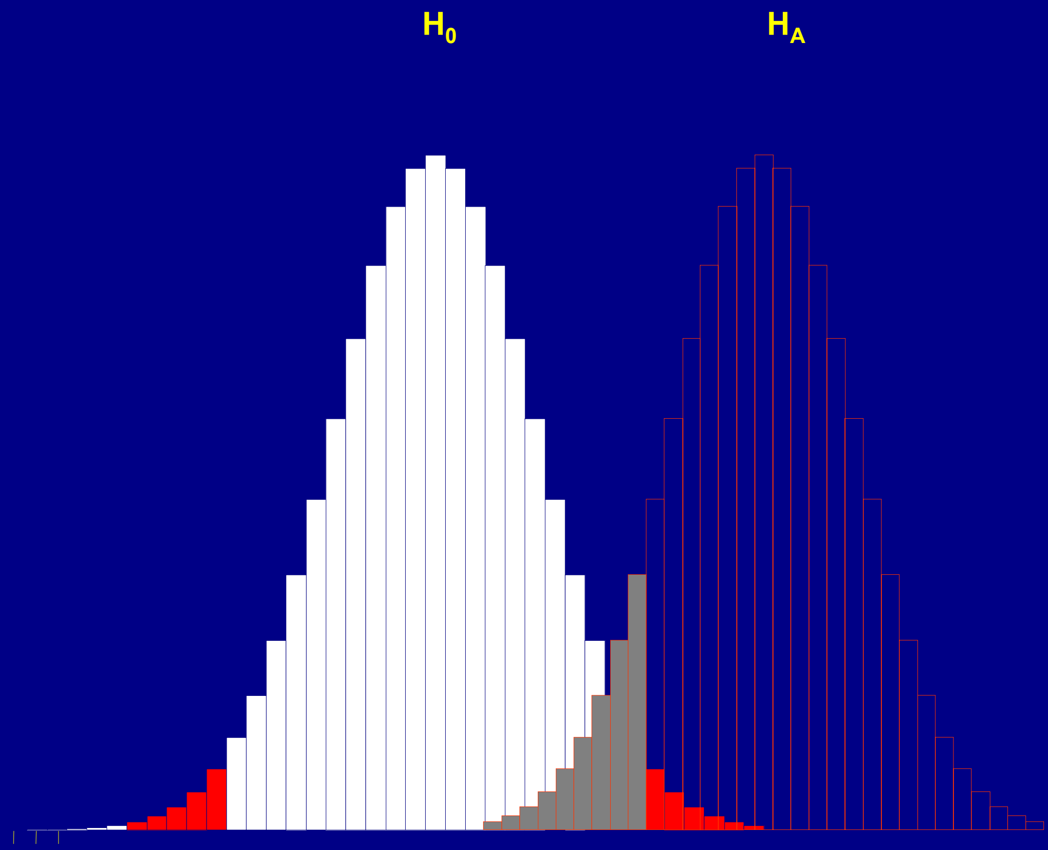
In der Medizin verlangt man üblicherweise, dass der Fehler zweiter Art bei *20%* oder bei *10%* liegt.

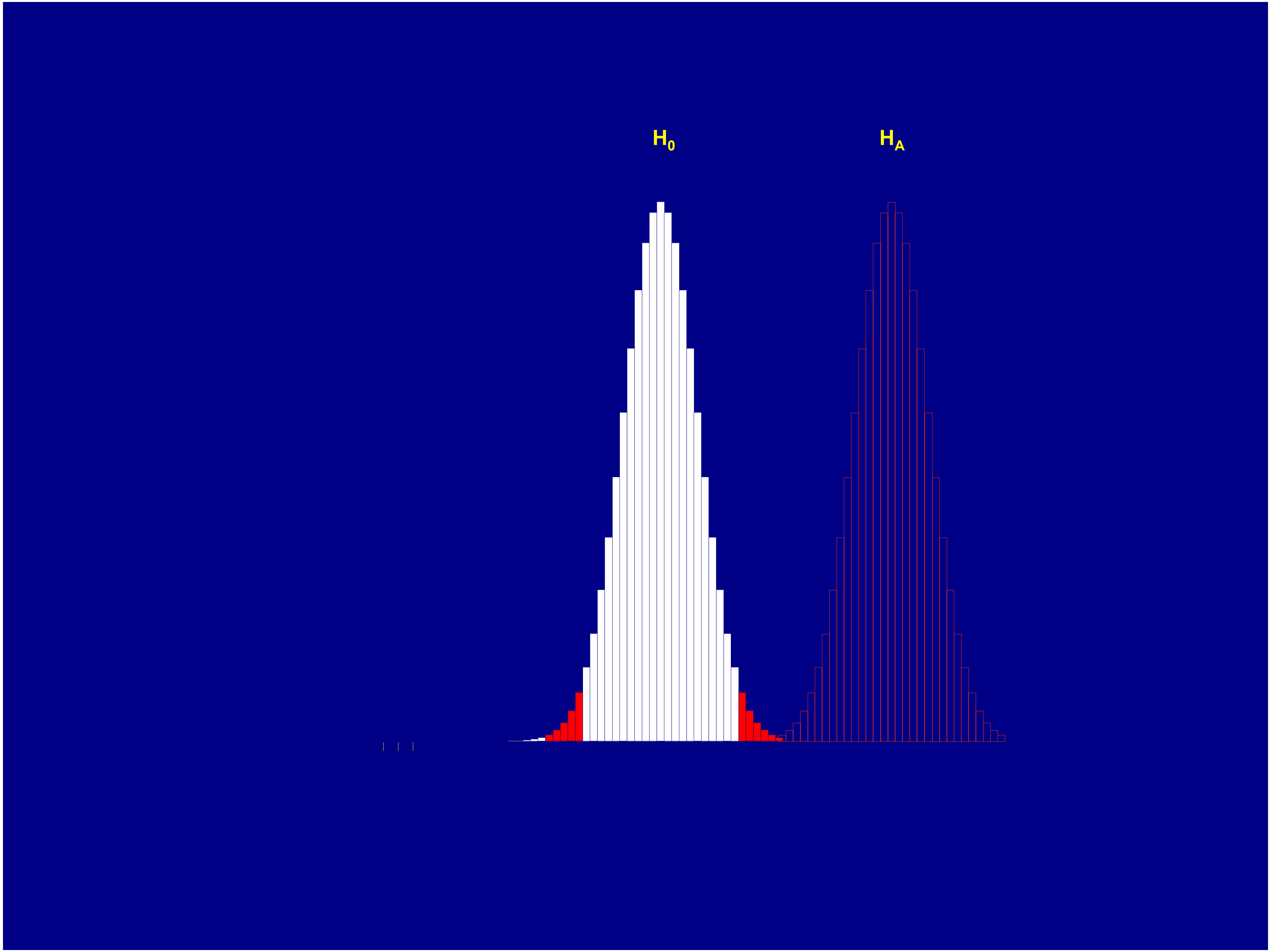
Binomialverteilung (n=100, p=0.5)



Fehler 2.Art β

Binomialverteilung (n >> 100, p=0.5))




H_0
H_A

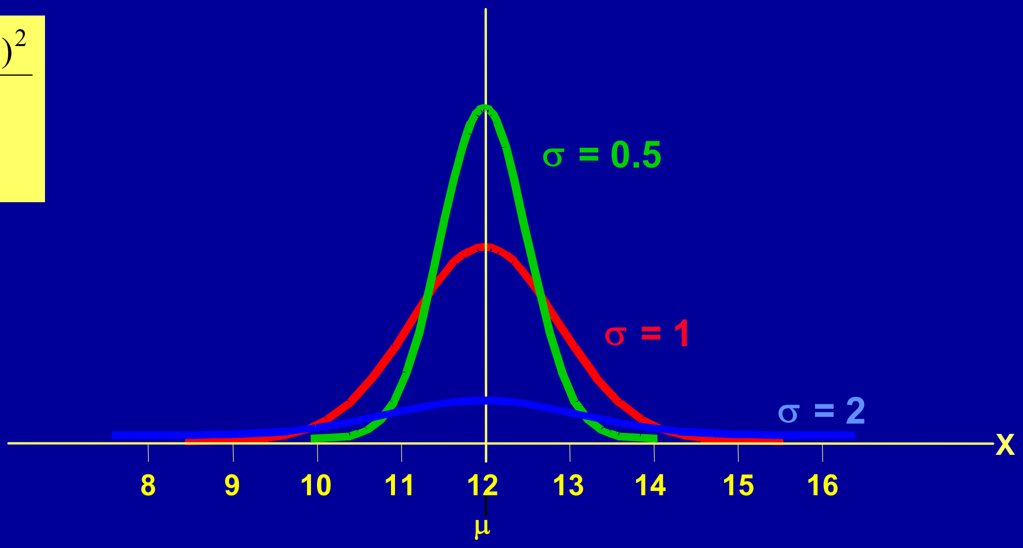
- 2 Stichproben A , B

- Hypothesen H_0: $A = B$ (Nullhypothese)
  H_A: $A \neq B$ (Alternativhypothese)

- Entscheidung für H_0 oder H_A

Testentscheidung	Realität: H_0 – *Med. nicht wirksam*	Realität: H_A – *Med. wirksam*
H_0 – *Med. nicht wirksam*	richtig $1 - \alpha$	β falsch negative Entscheidung *Produzentenrisiko*
H_A – *Med. wirksam*	α falsch positive Entscheidung *Konsumentenrisiko*	richtig $1 - \beta$ power

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



Normalverteilung N(μ,σ^2)

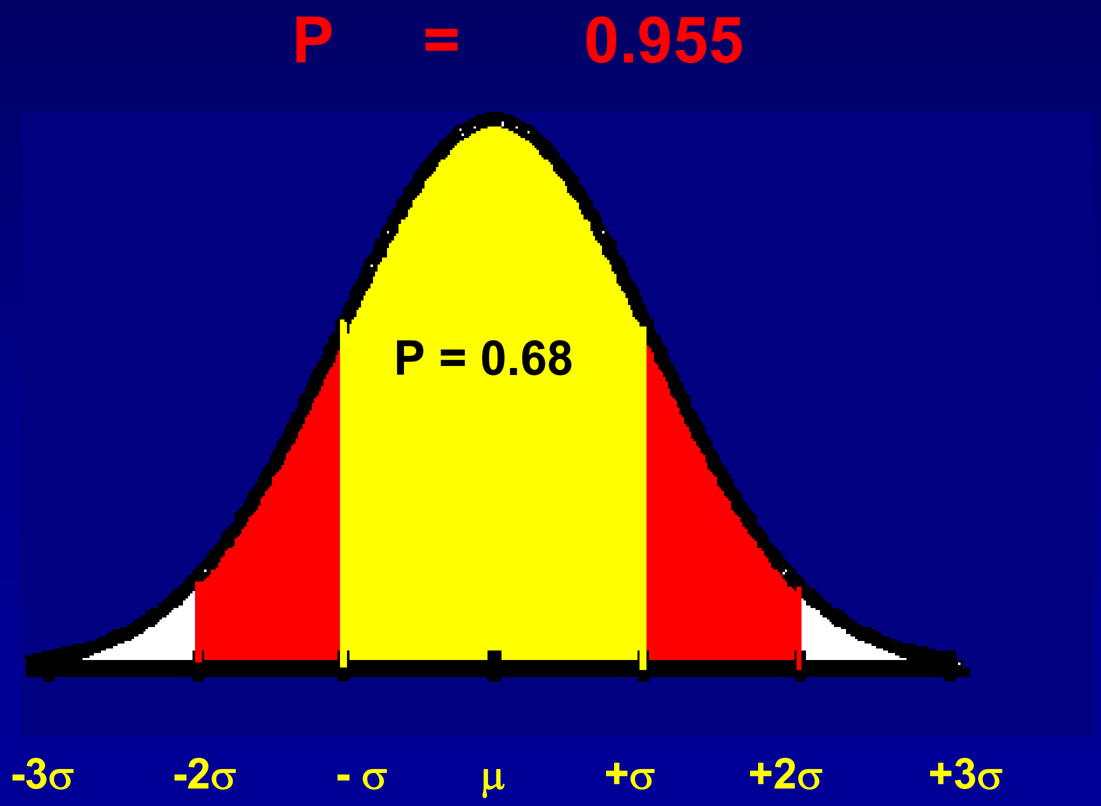
Mittelwert μ

Standardabweichung σ

Standard - Normalverteilung N (0 / 1)

Messfehler sind in der Regel normalverteilt
Gauß'sche Glockenkurven beschrieben werden

Die Wahrscheinlichkeit, dass ein Versuchsergebnis in einen Bereich fällt, ist identisch mit der Fläche



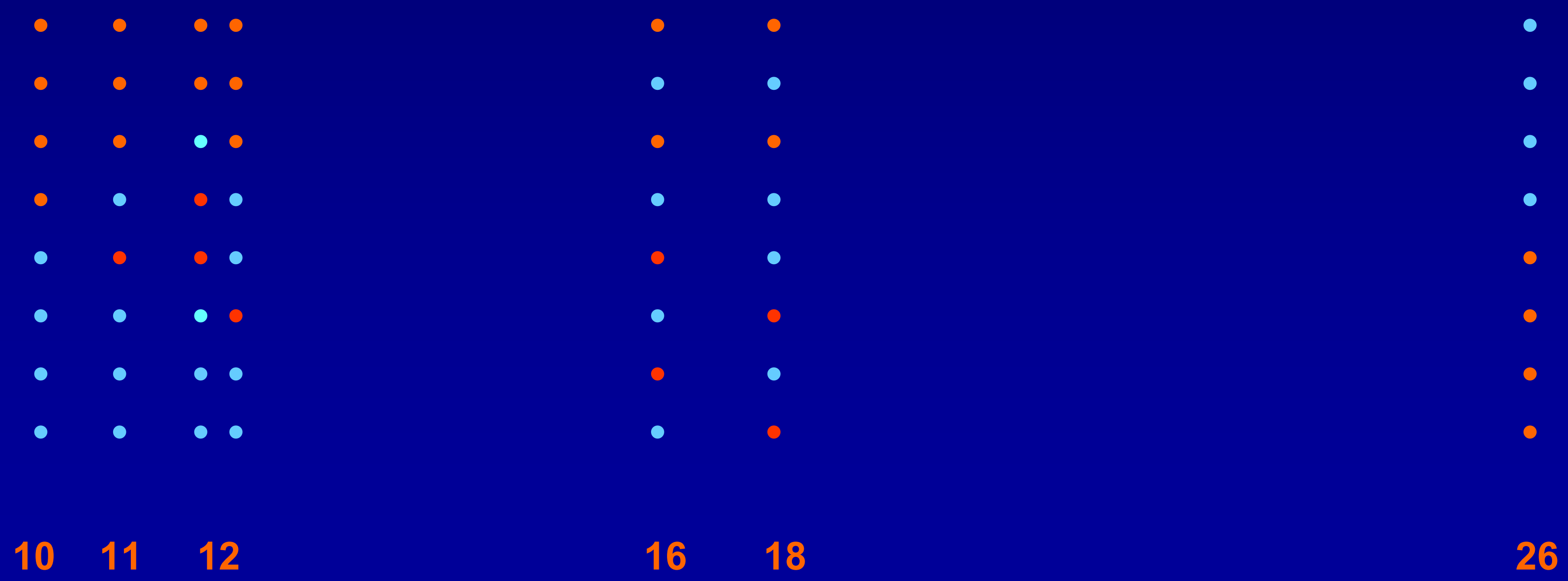
Mann - Whitney Test

unabhängige Stichproben

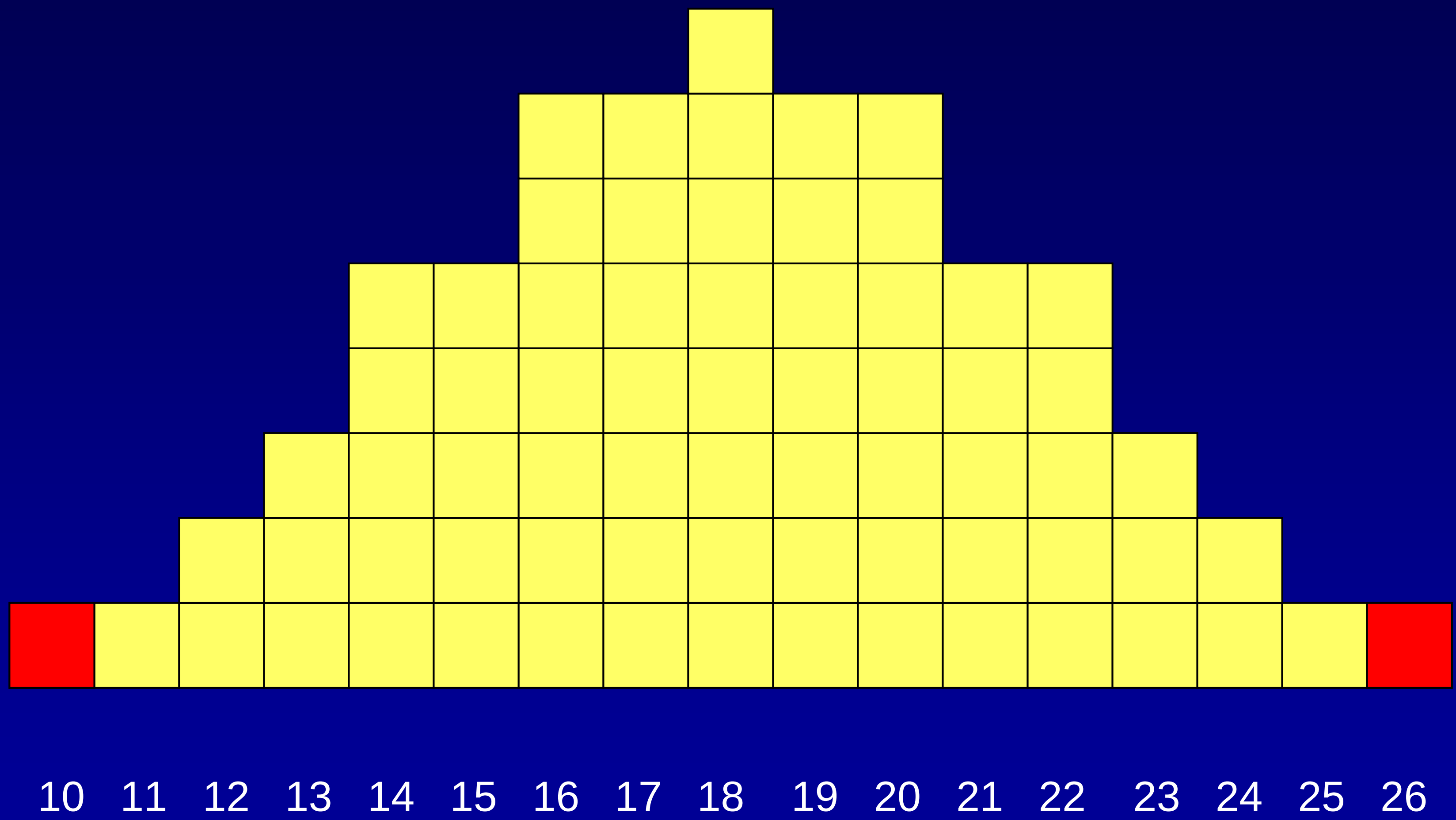
13,7	12,5	10,9	1
15,2	10,9	11,8	2
14,8	11,8	12,5	3
14,1	13,6	13,6	4
		13,7	5
		14,1	6
		14,8	7
		15,2	8

$\Sigma = 10$

$\Sigma = 26$



Verteilung der Rangzahlsummen (Anzahl der Möglichkeiten $\binom{8}{4}$ ‚8 über 4‘ = 70)



rote Fläche: p = 2/70 = 0,0286

Mann-Whitney Test - Vertrauensgrenzen für die Rangsumme T, α = 0.05
(Tabellenauszug)

n_1

n_2	4	5	6	7	8	9	10	11	12	13	14	15	16
4	**10-26**	16-34	23-43	31-53	40-64	49-77	60-90	72-104	85-119	99-135	114-152	130-170	147-189
5	11-29	**17-38**	24-48	33-58	42-70	52-83	63-97	75-112	89-127	103-144	118-162	134-181	151-201
6	12-32	18-42	**26-52**	34-64	44-76	55-89	66-104	79-119	92-136	107-153	122-172	139-191	157-211
7	13-35	20-45	27-57	**36-69**	46-82	57-96	69-111	82-127	96-144	111-162	127-181	144-201	162-222
8	14-38	21-49	29-61	38-74	**49-87**	60-102	72-118	85-135	100-152	115-171	131-191	149-211	167-233
9	14-42	22-53	31-65	40-79	51-93	**62-109**	75-125	89-142	104-160	119-180	136-200	154-221	173-243
10	15-45	23-57	32-70	42-84	53-99	65-115	**78-132**	92-150	107-169	124-188	141-209	159-231	178-254
11	16-48	24-61	34-74	44-89	55-105	68-121	81-139	**96-157**	111-177	128-197	145-219	164-241	183-265
12	17-51	26-64	35-79	46-94	58-110	71-127	84-146	99-165	**115-185**	132-206	150-228	169-251	189-275
13	18-54	27-68	37-83	48-99	60-116	73-134	88-152	103-172	119-193	**136-215**	155-237	174-261	195-285
14	19-57	28-72	38-88	50-104	62-122	76-140	91-159	106-180	123-201	141-223	**160-246**	179-271	200-296
15	20-60	29-76	40-92	52-109	65-127	79-146	94-166	110-187	127-209	145-232	164-256	**184-281**	206-306
16	21-63	30-80	42-96	54-114	67-133	82-152	97-173	113-195	131-217	150-240	169-265	190-290	**211-317**
17	21-67	32-83	43-101	56-119	70-138	84-159	100-180	117-202	135-225	154-249	174-274	195-300	217-327
18	22-70	33-87	45-105	58-124	72-144	87-165	103-187	121-209	139-233	159-257	179-283	200-310	223-337
19	23-73	34-91	46-110	60-129	74-150	90-171	107-193	124-217	143-241	163-266	184-292	205-320	228-348
20	24-76	35-95	48-114	62-134	77-155	93-177	110-200	128-224	147-249	167-275	188-302	211-329	234-358

Wilcoxon Test

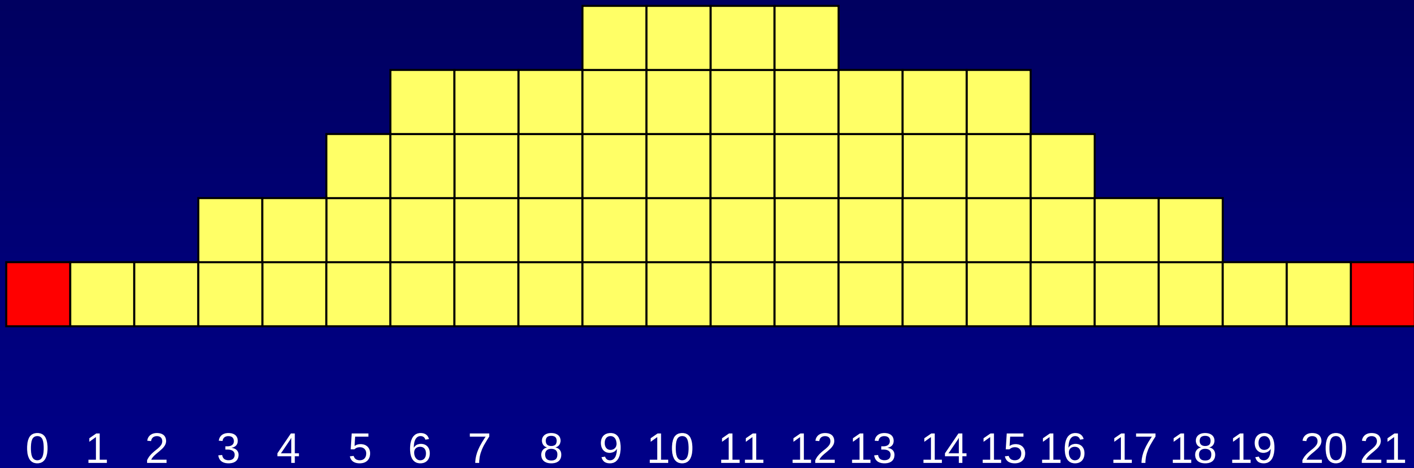
abhängige Stichproben

Proband	Bd_{sys} vorher	Bd_{sys} nachher	Differenz	Rangzahlen
1	132	128	4	2
2	144	137	7	4
3	129	132	-3	1
4	136	125	9	5
5	131	125	6	3
6	140	130	10	6

Rangzahlsumme der negativen Differenzen: 1

Rangzahlsumme der positiven Differenzen: 20

Verteilung der Rangzahlsummen (Anzahl der Möglichkeiten 2^6)



rote Fläche: $p = 2/\ 2^6 = 0{,}0325$

Tabelle: Wilcoxon-Test für Paardifferenzen (n=Anzahl der Paare)

n	α 0.10	α 0.05	α 0.01
5	0 – 15		
6	2 – 19	0 – 21	
7	3 – 25	2 – 26	
8	5 – 31	3 – 33	0 – 36
9	8 – 37	5 – 40	1 – 44
10	10 – 45	8 – 47	3 – 52
11	13 – 53	10 – 56	5 – 61
12	17 – 61	13 – 65	7 – 71
13	21 – 70	17 – 74	9 – 82
14	25 – 80	21 – 84	12 – 93
15	30 – 90	25 – 95	15 – 105
16	35 – 101	29 – 107	19 – 117
17	41 – 112	34 – 119	23 – 130
18	47 – 124	40 – 131	27 – 144
19	53 – 137	46 – 144	32 – 158
20	60 – 150	52 – 158	37 – 173
21	67 – 164	58 – 173	42 – 189
22	75 – 178	66 – 187	48 – 205
23	83 – 193	73 – 203	54 – 222
24	91 – 209	81 – 219	61 – 239
25	100 – 225	89 – 236	68 – 257

2. t-Tests für stetige Daten

Für normalverteilte Daten stehen je nach Stichprobenstruktur *drei verschiedene t-Tests* zur Verfügung: Der t-Test für *eine* Stichprobe, für *zwei verbundene* und für *zwei unverbundene* Stichproben.

Die *Grundidee* aller drei t-Tests besteht darin, dass man eine beobachtete *Mittelwertsdifferenz* ins Verhältnis setzt zur *Streuung* in den jeweiligen Stichproben:

Bei *großen Streuungen* werden auch *große Zufallsschwankungen* der Mittelwerte für möglich gehalten, bei *kleinen Streuungen* können bereits *geringe Unterschiede* der Mittelwerte auf *systematische Effekte* rückschließen lassen.

3. Der t-Test für unverbundene Stichproben

Situation:

Vergleich der Mittelwerte $\bar{x}_1$ und $\bar{x}_2$ eines stetigen, normalverteilten Merkmals aus zwei unabhängigen Stichproben

Fragestellung:

Sind die beiden zugrundeliegenden Erwartungswerte μ_0 und μ_1 *verschieden?*

Nullhypothese:

Die Abweichung der Mittelwerte beruht auf Zufall, $\mu_0 = \mu_1$

Alternative:

Die Abweichung der Mittelwerte beruht auf systematischen Unterschieden, $\mu_0 \neq \mu_1$

t-Test für 2 unverbundene Stichproben

- - 2 unverbundene Stichproben der Umfänge n_1 und n_2
- die Werte beider Stichproben sind normalverteilt mit ähnlichen Standardabweich

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}}}$$

$$s^2 = \frac{(n_1 - 1)\, s_1^{\,2} + (n_2 - 1)\, s_2^{\,2}}{n_1 + n_2 - 2}$$

Die Nullhypothese wird abgelehnt, falls

$$|t| > t_{n_1 + n_2 - 2;\, 1 - \frac{\alpha}{2}}$$

1. Statistische Tests für stetige Daten aus verbundenen Stichproben

Verbundene Stichproben entstehen, wenn am *selben Merkmalsträger* das *gleiche Merkmal* unter *verschiedenen Bedingungen* gemessen wird.

Beispiel: Zwei *verschiedene Therapien* (=Bedingungen) werden am *selben Patienten* (=Merkmalsträger) mit dem *selben Zielkriterium* (=Merkmal) verglichen.

Für dichotomes Zielkriterium (Erfolg ja/nein) wird bei unverbundenen Stichproben der **Chi-Quadrat-Test** (Fallzahl < 20: **Exakter Fisher Test)**, bei verbundenen Stichproben der **Mc Nemar Test** (Fallzahl < 10: **Binomialtest**) verwendet (Vorlesung II und III).

Für ordinale oder stetige, nicht normalverteilte Zielkriterien wird bei unverbundenen Stichproben der **U-Test von Mann-Whitney** verwendet (Vorlesung IV). An seine Stelle tritt für verbundene Stichproben der **Wilcoxon Test**.

Analog wird der **t-Test für unverbundene Stichproben** (Vorlesung IV) durch den **t-Test für verbundene Stichproben** ersetzt, wenn die Daten normalverteilt sind.

Der t-Test für verbundene Stichproben

Situation:

Vergleich der Mittelwerte $\bar{x}_1$ und $\bar{x}_2$ eines stetigen, normalverteilten Merkmals aus zwei **abhängigen** Stichproben

Fragestellung:

Sind die beiden zugrundeliegenden Erwartungswerte μ_0 und μ_1 *verschieden?*

Nullhypothese:

Die Abweichung der Mittelwerte beruht auf Zufall, $\mu_0 = \mu_1$

Alternative:

Die Abweichung der Mittelwerte beruht auf systematischen Unterschieden, $\mu_0 \neq \mu_1$

t-Test für 2 verbundene Stichproben

- - 2 paarige Stichproben des Umfangs n mit Wertepaaren (x_i,y_i)
- Differenzen $d_i = y_i - x_i$ (oder $x_i - y_i$), die normalverteilt sind

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

$$s_{\bar{d}} = \frac{s_d}{\sqrt{n}}$$

Die Nullhypothese wird abgelehnt, falls

$$|t| > t_{n-1;1-\frac{\alpha}{2}}$$

Durchführung statistischer Tests

Formulierung der inhaltlichen Fragestellung

Formulierung der statistischen Hypothesen H_O und H_A

Festlegung der Signifikanzniveaus α

Wahl der Prüfgröße bzw. des statistischen Verfahrens

Entscheidung für eine der beiden Hypothesen
Statistische Signifikanz - medizinische Relevanz

Annahme H_0 ist richtig (A = B)

Anzahl Tests	Wahrscheinlichkeit für H_0		p	adjustiert
1	0.95			0.05
2	$0.95^2 = 0.903$	$0.975^2 = 0.95$	0.097	0.025
3	$0.95^3 = 0.857$	$0.983^3 = 0.95$	0.143	0.017
10	$0.95^{10} = 0.599$	$0.995^{10} = 0.95$	0.401	0.005
50	$0.95^{50} = 0.077$	$0.999^{50} = 0.95$	0.923	0.001

Quellen und Methoden zur Vermeidung von Bias

Strukturungleichheit bzgl. unbekannter Variablen:
- Randomisierung

Strukturungleichheit bzgl. bekannter Variablen:
- Stratifikation und Poststratifikation
- Adjustierung(Kovarianzanalyse, logistische oder Cox-Regression)

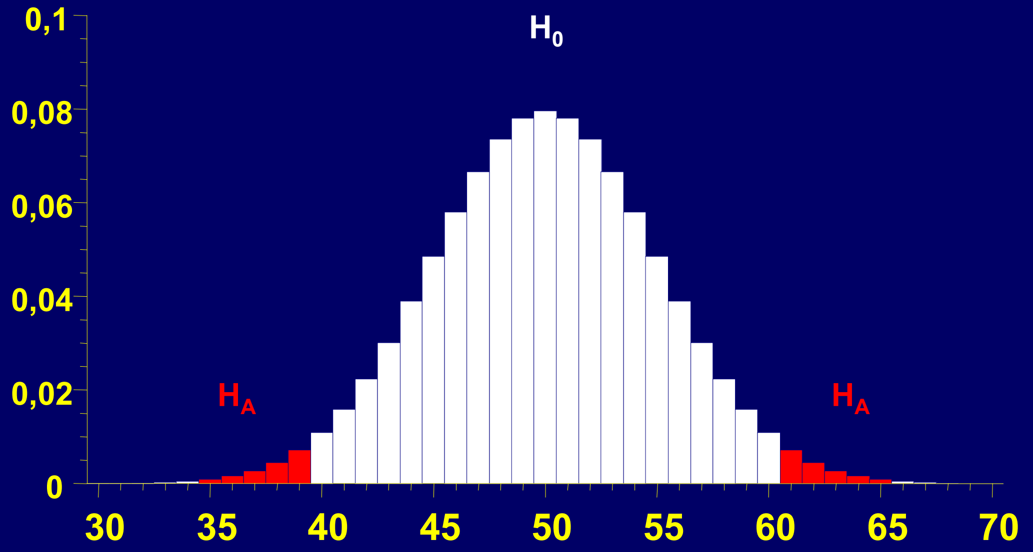
Beobachtungsungleichheit
- Verblindung, Standardisierung, Laborringversuche

Die Methoden zur Verhinderung von Strukturungleichheit dienen auch der Verringerung der Fehlervarianz

Statistische Testverfahren in der Medizin

		2 Stich proben		mehr als 2 Stichproben	
		unverbunden	verbunden	unverbunden	verbunden
stetig	normal-verteilt	t-Test	t-Test paarig	Varianz-analyse	Varianz-analyse
	nicht normal	Mann-Whitney Test	Wilcoxon Test	Kruskal-Wallis Test	Friedman Test
diskret		χ^2-Test Exakter Fisher Test	χ^2-Test nach McNemar	χ^2-Kontingenz-test	Q-Test

Binomialverteilung (n=100, p=0.5)



α
Signifikanzniveau
Fehler 1.Art
α = 5% oder 1%

Aufbau des Chi-Quadrat-Tests

Situation:

Vergleich der Häufigkeiten h_0 vs h_1 eines binären, also kategorialen Merkmals aus zwei unabhängigen Stichproben

Fragestellung:

Sind die beiden zugrundeliegenden Wahrscheinlichkeiten p_0 und p_1 unterschiedlich?

Nullhypothese:

Die Abweichungen der Häufigkeiten beruhen auf Zufall, $p_0 = p_1$

Alternative:

Die Abweichungen der Häufigkeiten beruhen auf systematischen Unterschieden, $p_0 \neq p_1$

Beispiel zum Chi-Quadrat Test

Bei *jeweils 100 Patienten* wird die herkömmliche Therapie A mit der neuen Therapie B verglichen.
Bei Therapie A werden *60/100 Erfolge*, bei Therapie B werden *80/100 Erfolge* beobachtet.

	Erfolg		
	Ja	nein	Gesamt
Therapie A	**60**	**40**	100
Therapie B	**80**	**20**	100
Gesamt	140	60	200

Kann dieser Unterschied auf Zufall beruhen?

Vorgehen beim Chi-Quadrat Test

1.Schritt: Berechnung der Gesamtheilungsrate (140/200 = 70%).

2.Schritt: Berechnung der in beiden Gruppen aufgrund der Gesamt-Heilungsrate zu erwartenden Erfolge bzw. Misserfolge (jeweils 70 Erfolge und 30 Misserfolge).

	Erfolg Ja	nein	Gesamt
Therapie A	**60** **70**	**40** **30**	100
Therapie B	**80** **70**	**20** **30**	100
Gesamt	140	60	200

Therapie * Erfolg Kreuztabelle

			Erfolg		Gesamt
			ja	nein	
Therapie	A	Anzahl	60	40	100
		Erwartete Anzahl	70,0	30,0	100,0
	B	Anzahl	80	20	100
		Erwartete Anzahl	70,0	30,0	100,0
Gesamt		Anzahl	140	60	200
		Erwartete Anzahl	140,0	60,0	200,0

Chi-Quadrat-Tests

	Wert	df	Asymptotische Signifikanz (2-seitig)
Chi-Quadrat nach Pearson	9,524[b]	1	,002
Anzahl der gültigen Fälle	200		

a. Wird nur für eine 2x2-Tabelle berechnet

b. 0 Zellen (,0%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist 30,00.

Die Prüfgröße des Chi-Quadrat-Tests auf Unabhängigkeit für Vier-Felder-Tafeln

$$\chi^2 = \frac{(h_{11}-e_{11})^2}{e_{11}} + \frac{(h_{12}-e_{12})^2}{e_{12}} + \frac{(h_{21}-e_{21})^2}{e_{21}} + \frac{(h_{22}-e_{22})^2}{e_{22}}$$

$$= \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{(h_{ij}-e_{ij})^2}{e_{ij}}$$

$$\sum \frac{(B-E)^2}{E}$$

h_{11}: Beobachtete Häufigkeit Therapie A: Erfolg (1.Zeile, 1.Spalte)
h_{12}: Beobachtete Häufigkeit Therapie A: Misserfolg (1.Zeile, 2.Spalte)
h_{21}: Beobachtete Häufigkeit Therapie B: Erfolg (2.Zeile, 1.Spalte)
h_{22}: Beobachtete Häufigkeit Therapie B: Misserfolg (2.Zeile, 2.Spalte)

e_{11}, e_{12}, e_{21} e_{22} bezeichnet die entsprechenden erwarteten Häufigkeiten.

Stichprobenumfangschätzung
Häufigkeiten (Chi-Quadrat Test)
α =0.05 1- β = 90 %

$h_1=60\%$ $h_2 = 20\%$	$\Delta = 40\%$	n = 30
$h_1=60\%$ $h_2 = 30\%$	$\Delta = 30\%$	n = 56
$h_1=60\%$ $h_2 = 40\%$	$\Delta = 20\%$	n = 86
$h_1=60\%$ $h_2 = 45\%$	$\Delta = 15\%$	n = 231
$h_1=60\%$ $h_2 = 50\%$	$\Delta = 10\%$	n = 519
$h_1=60\%$ $h_2 = 55\%$	$\Delta = 5\%$	n = 2053

Ergänzung (kein Prüfungsstoff):

Die *mathematischen Gründe*, warum der Chi-Quadrat Test auf einer *Chi-Quadrat verteilten Prüfgröße* mit *einem Freiheitsgrad* basiert, sind nicht Thema der Vorlesung. Folgende heuristische Erklärung ist der Kern des mathematischen Beweises:

Beim Chi-Quadrat Test gehen *vier beobachtete Häufigkeiten* (Erfolg unter Therapie A bzw. B, Misserfolg unter A bzw. B) in den Test ein, zur Bestimmung der *erwarteten Häufigkeiten* werden *drei Randbedingungen* festgehalten (Gesamtzahl aller Patienten, Prozent Erfolg unter A, Prozent Erfolg unter B).

Beim Chi-Quadrat Test hat man also 4-3=1 Freiheitsgrade.

Anmerkungen zum Chi-Quadrat Test

Bei kleinen Fallzahlen (*mindestens eine* erwartete Häufigkeit < *5*)

muss der *Chi-Quadrat Test* durch

den *Exakten Fisher Test,* ersetzt werden.

Dieser Test benutzt die hypergeometrische Verteilung,
die nicht Thema der Vorlesung ist.

1. Statistische Tests für stetige Daten

In den bislang behandelten Beispielen hatte die Zielgröße nur zwei Ausprägungen (Therapieerfolg ja/nein), war also *kategoriell*. In vielen Studien wird die Zielgröße aber als *stetiges* Merkmal gemessen. Dann hängt die Frage, welcher statistische Test verwendet werden kann, davon ab, ob das entsprechende Merkmal *normalverteilt* ist oder nicht. In der Praxis entscheidet man diese Frage anhand von zwei Kriterien:

- Hat das *Histogramm* die Form einer Glockenkurve?
- Liegt die *Schiefe* zwischen -1 und +1?

Sind *beide* Kriterien erfüllt, geht man von normalverteilten Daten aus. Bei vielen Autoren wird zusätzlich noch die *Varianzengleichheit* in den beteiligten Stichproben gefordert. Allerdings gibt es inzwischen Verfahren, die es erlauben, auch Stichproben mit unterschiedlicher Streuung zu vergleichen.

Konfidenzintervalle (Vertrauensbereiche)

Lernziel:

Fähigkeit, Konfidenzintervalle zu errechnen und ihre Bedeutung zu erkennen

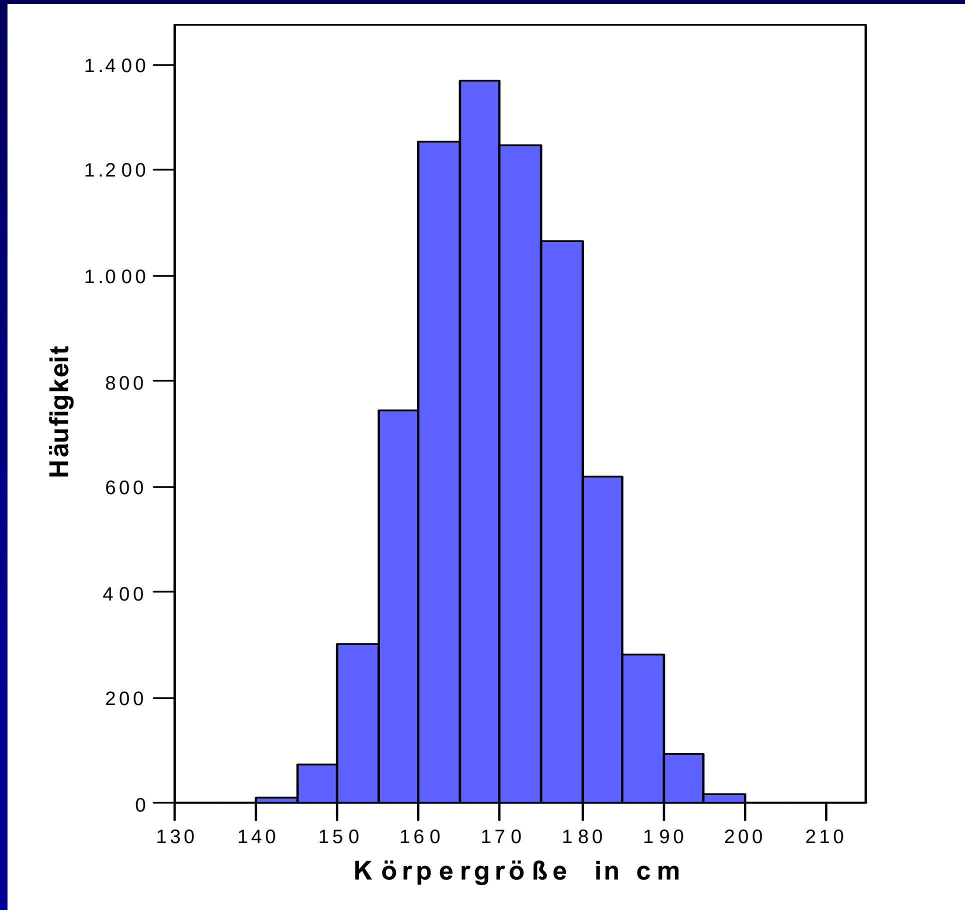
Fragestellung:

Mit welcher Sicherheit kann ein Parameter geschätzt werden und welches Vertrauen kann man zu der Schätzung haben?

Gesucht: ein Bereich mit abschätzbarem Sicherheitsgrad

Konfidenzintervalle (Vertrauensbereiche)

Beispiel Körpergröße (cm)
für ca. 7000 Personen
(normalverteilt)



Normbereich
(Referenzbereich):

MW ± 2 St.abw.
169,4 ± 18,9

Konfidenzintervall:
[169,2;169,7]

Bundesgesundheitssurvey 1998

Konfidenzintervalle (Vertrauensbereiche)

Definition Konfidenzintervall

Für einen Sicherheitsgrad von 95% (1-α) werden die untere und die obere Grenze gesucht:

$$P(UG \leq \mu \leq OG) = 1\text{-}\alpha$$

Der Erwartungswert μ wird aus dem arithmetischen Mittelwert $\bar{x}$ geschätzt und bei bekannter Standardabweichung σ ergibt sich:

$$\bar{x} - 1{,}96\,\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + 1{,}96\,\frac{\sigma}{\sqrt{n}}$$

Konfidenzintervalle (Vertrauensbereiche)

Der Wert 1,96 ist der Standardnormalverteilung für ein 95%-Konfidenzintervall entnommen. Er kann als Näherung auf 2 gerundet werden.

$$SEM = \frac{\sigma}{\sqrt{n}}$$ Standardfehler des Mittelwertes

Untere Grenze

$$\bar{x} - 2 * SEM$$

Obere Grenze

$$\bar{x} + 2 * SEM$$

Konfidenzintervalle (Vertrauensbereiche)

Beispiel Körpergröße (cm) aus BGS98
getrennt nach dem Geschlecht:

	n	Normbereich [cm] MW ± 2 Standardabw.	Konfidenzintervall [cm] [untere Grenze;obere Grenze]
gesamt	7084	169,4 ± 18,9	[169,2 ; 169,7]
weiblich	3647	163,3 ± 13,6	[163,1 ; 163,5]
männlich	3437	176,0 ± 14,6	[175,7 ; 176,3]

Konfidenzintervalle (Vertrauensbereiche)

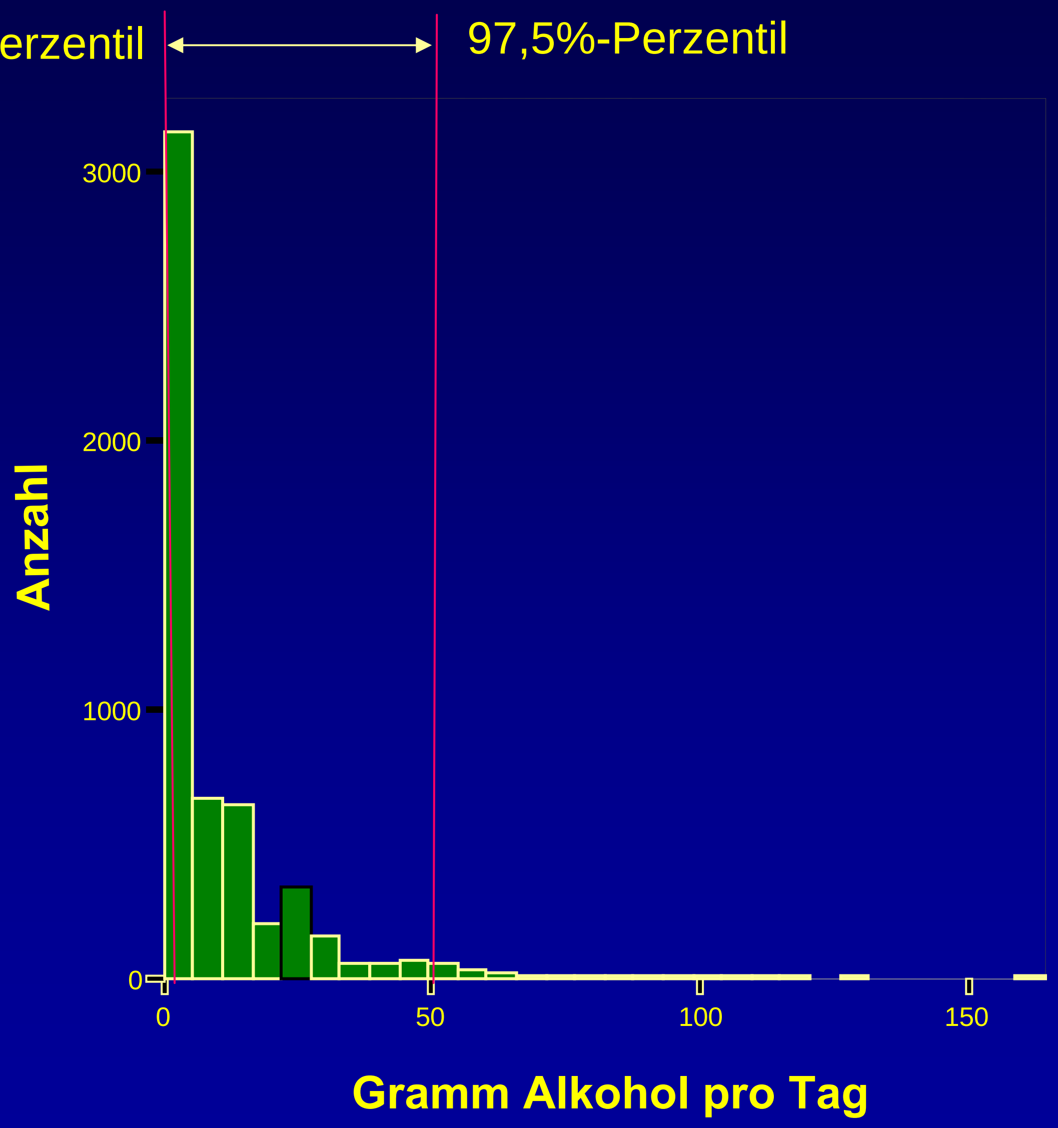
Durchschnittlicher Konsum von reinem Alkohol pro Tag (BGS98)

(nicht normalverteilt)

ca. 5500 Personen

Normbereich:
[2,5;97,5]-Perzentil

=[0g ; 50,4g]



Konfidenzintervalle (Vertrauensbereiche)

Durchschnittlicher Konsum von reinem Alkohol pro Tag

(nicht normalverteilt)

Normbereich:

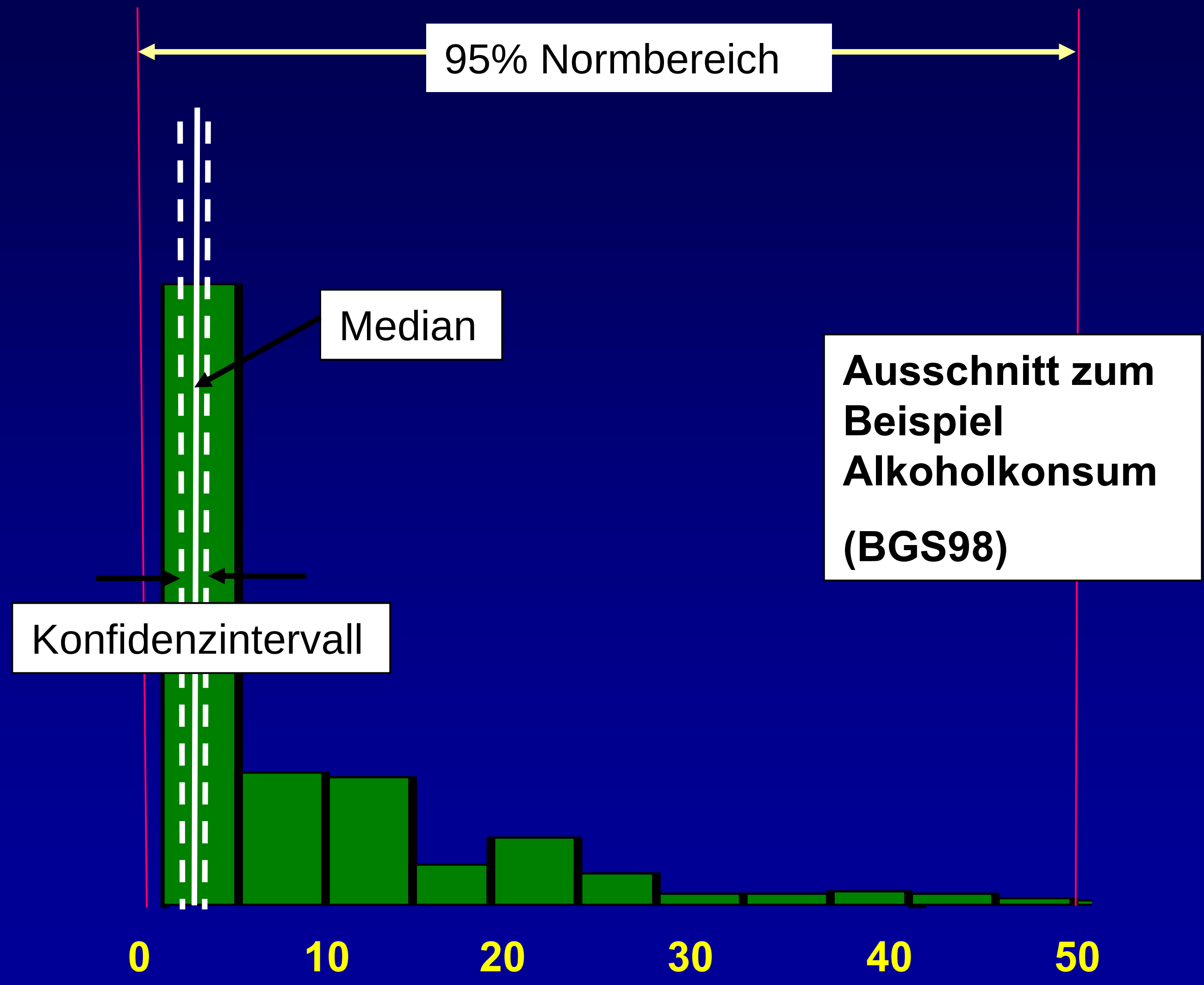
[2,5;97,5]-Perzentil =

[0 ; 50,4]

Median = 3,82

u = 3,43

s = 3,92



Konfidenzintervalle (Vertrauensbereiche)

Für <u>nicht normalverteilte</u> Stichproben ist der Median die geeignete statistische Kenngröße. Daher wird das Konfidenzintervall bestimmt durch die ganzzahlige Rundung auf die Perzentile, die sich berechnen als:

$$u = \frac{n}{2} - (1{,}96 * \frac{\sqrt{n}}{2})$$

untere Grenze

und

$$o = 1 + \frac{n}{2} + (1{,}96 * \frac{\sqrt{n}}{2})$$

obere Grenze

Konfidenzintervalle (Vertrauensbereiche)

Konfidenzintervalle für relative Häufigkeiten

Fragestellung:

Gesucht ist der unbekannte relative Anteil p für eine Grundgesamtheit, die binomialverteilt ist.

Für k= Ereignisses in einer Stichprobe vom Umfang n, ist

$$\boxed{\hat{p} = \frac{k}{n}}$$

der Punktschätzer.

Konfidenzintervalle (Vertrauensbereiche)

Beispiel für ein Ereignis, das jeweils bei einem Viertel der Stichprobe auftritt

$$\hat{p} = 0{,}25$$

Das Konfidenzintervall wird mit zunehmender Stichprobengröße schmaler

n	25% aufgetretene Ereignisse	95% Konfidenz-intervall (gerundet)
4	1	[1% ; 81%]
20	5	[9% ; 49%]
40	10	[13% ; 41%]
80	20	[16% ; 36%]
120	30	[17% ; 33%]
200	50	[19% ; 31%]

Konfidenzintervalle (Vertrauensbereiche)

Beispiel Rauchgewohnheiten (BGS98)
Gesamte Stichprobe: 6963 Probanden

Rauchgewohnheit	Anzahl	in Prozenten (gerundet)	95% Konfidenzintervall (gerundet)
täglich	1853	26,6	[25,6% ; 27,6%]
gelegentlich	442	6,3	[5,7% ; 6,9%]
habe früher geraucht	1363	19,6	[18,7% ; 20,5%]
in den letzten 12 Mon. aufgehört	120	1,7	[1,4% ; 2,0%]
nie geraucht	3185	45,7	[44,5% ; 46,9%]

Konfidenzintervalle (Vertrauensbereiche)

Das Konfidenzintervall wird mit der Approximation durch die Normalverteilung geschätzt.

$$\left[\hat{p} - 1{,}96 * \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} ; \hat{p} + 1{,}96 * \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

Dabei ist $\hat{p}$ der Punktschätzer und *n* die Größe der Stichprobe bei einem 95%-Sicherheitsgrad.

Konfidenzintervalle (Vertrauensbereiche)

Beispiel Rauchgewohnheiten (BGS98)

Von 6963 Personen mit Angaben haben 3185 noch nie geraucht: k=3185; $\hat{p}$ = 45,7%.

Das Konfidenzintervall wird annähernd bestimmt durch:

$$\left[45,7 - 1,96 * \sqrt{\frac{45,7(1-45,7)}{6963}};45,7 + 1,96 * \sqrt{\frac{45,7(1-45,7)}{6963}}\right]$$

= [44,5; 46,9]

Konfidenzintervalle (Vertrauensbereiche)

Die Breite des Konfidenzintervalles ist abhängig von:

- Stichprobenumfang n

  Je größer der Stichprobenumfang ist, desto genauer ist die Schätzung und das Konfidenzintervall ist schmaler

- Streuung der Stichprobe

  Je kleiner die Streuung ist, desto genauer ist die Schätzung und das Konfidenzintervall ist schmaler

- Sicherheitsgrad $1-\alpha$

  Je größer der Sicherheitsgrad $1-\alpha$ wird, je höher die Zuverlässigkeit der Aussage wird, desto breiter wird das Konfidenzintervall

Biomathematik

Methoden zur Evaluierung von Diagnoseverfahren

Diagnosestellung aus statistischer Sicht

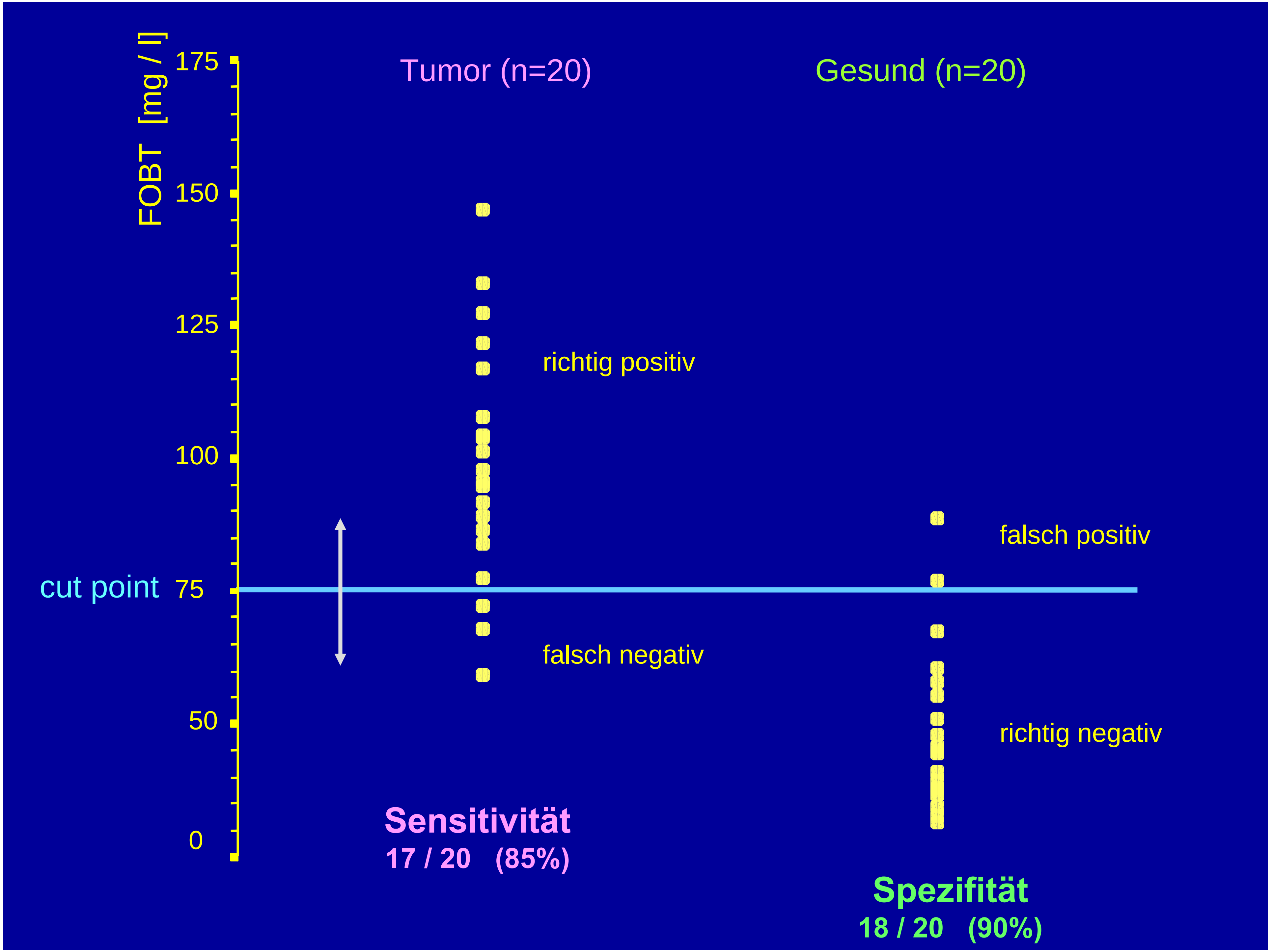
Die *Treffsicherheit* diagnostischer Aussagen kann stark variieren:
Durch Färbung histologischer Schnitte sind Tumorzellen u.U. *zweifelsfrei* nachzuweisen.
Bei negativer Färbung können Tumorzellen in anderen Schnitten desselben Patienten aber *nicht ausgeschlossen* werden.

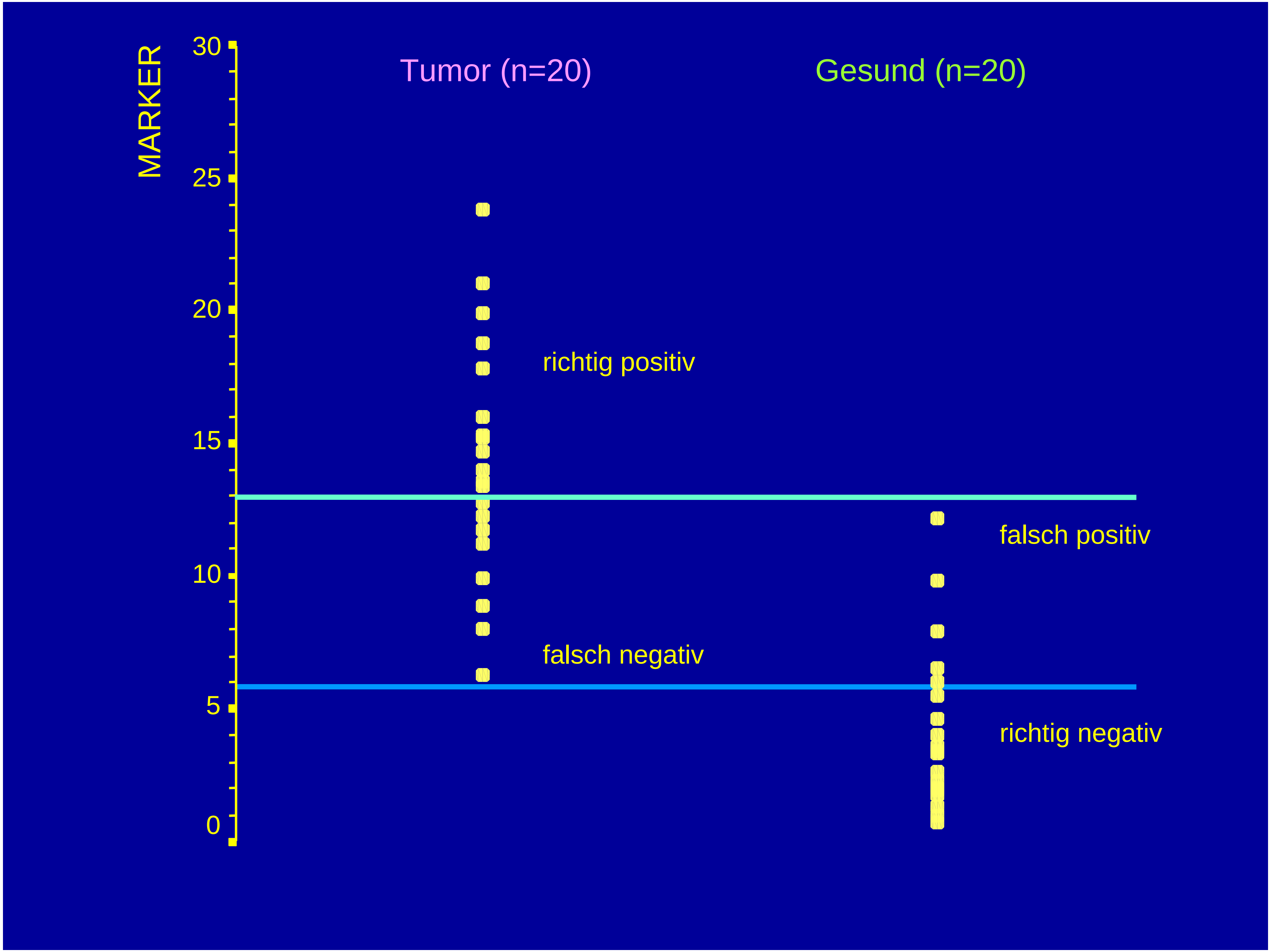
In der Bildgebung ist das Risiko für eine falsche Entscheidung noch größer.

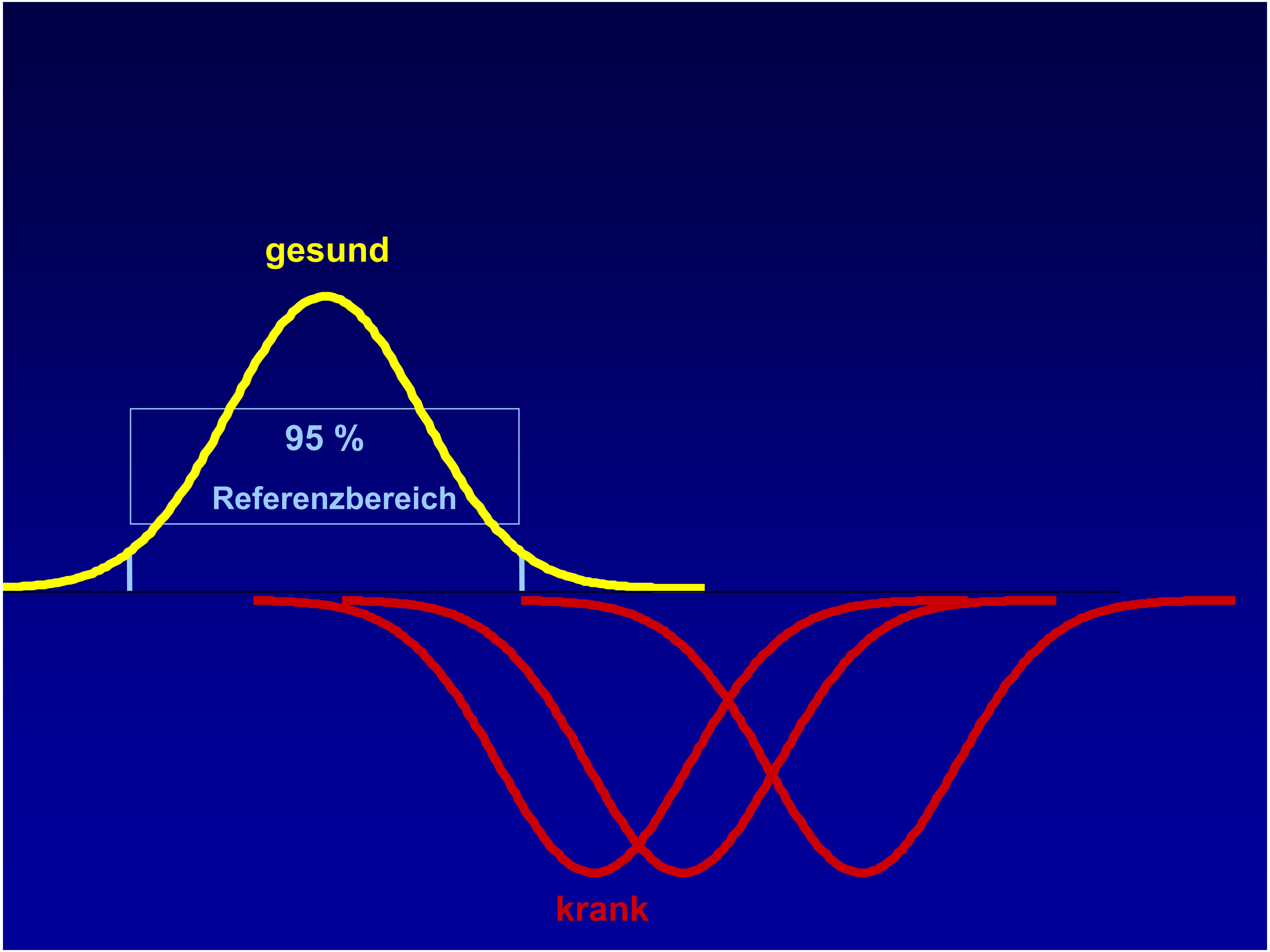
Deswegen betrachtet man in Diagnosestudien und im klinischen Alltag die histologische Sicherung als *Referenzkriterium* (Goldstandard).

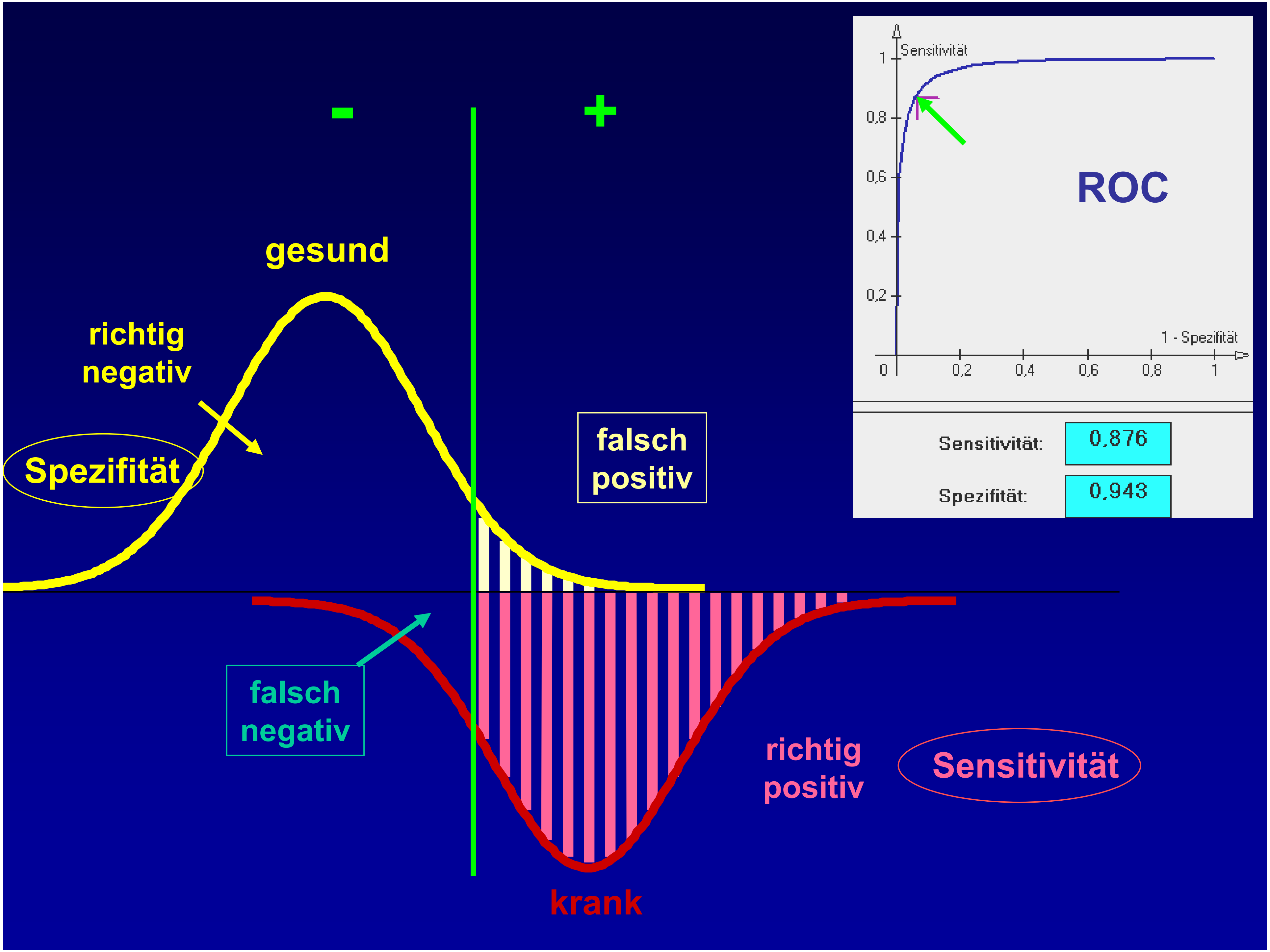
Wenn kein fehlerfreies Referenkriterium existiert,
betrachtet man häufig ein *etabliertes* Diagnosekriterium als Goldstandard, obwohl dies eigentlich nicht gerechtfertigt ist.
Für die Diagnose des Glaukoms (grüner Star) wird z.B. die Perimetrie als sensorischer Goldstandard verwendet, obwohl sie erst *spät* Schädigungen entdeckt.

Die *statistische Analyse* zur Bewertung von Diagnoseverfahren hängt davon ab, ob eines der untersuchten Verfahren als *fehlerfrei* betrachtet wird oder nicht.

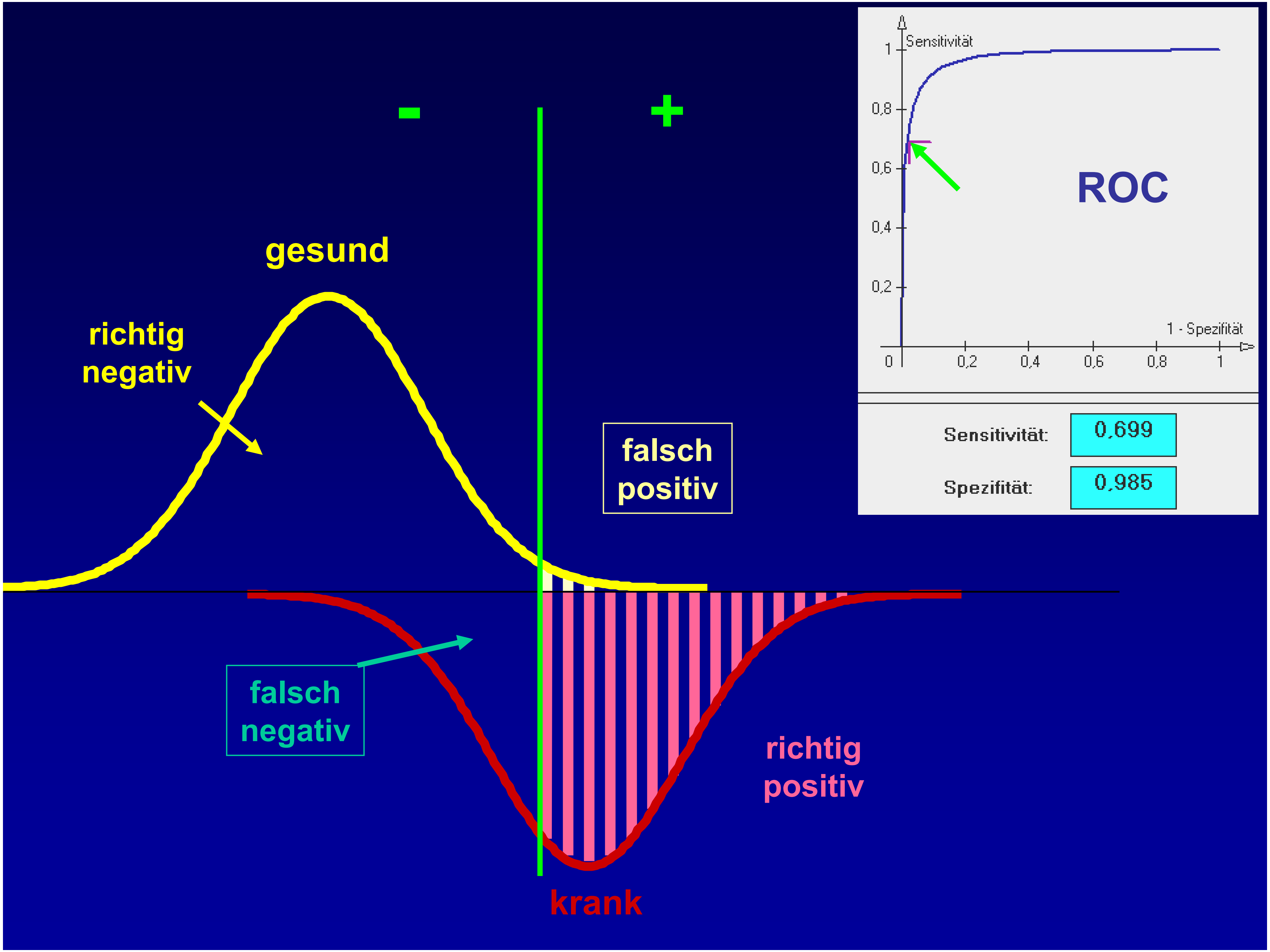

FOBT [mg / l]
175
150
125
100
75
50
0
Tumor (n=20)
Gesund (n=20)
richtig positiv
falsch positiv
cut point
falsch negativ
richtig negativ
Sensitivität
17 / 20 (85%)
Spezifität
18 / 20 (90%)


MARKER
30
25
20
15
10
5
0
Tumor (n=20)
Gesund (n=20)
richtig positiv
falsch negativ
falsch positiv
richtig negativ

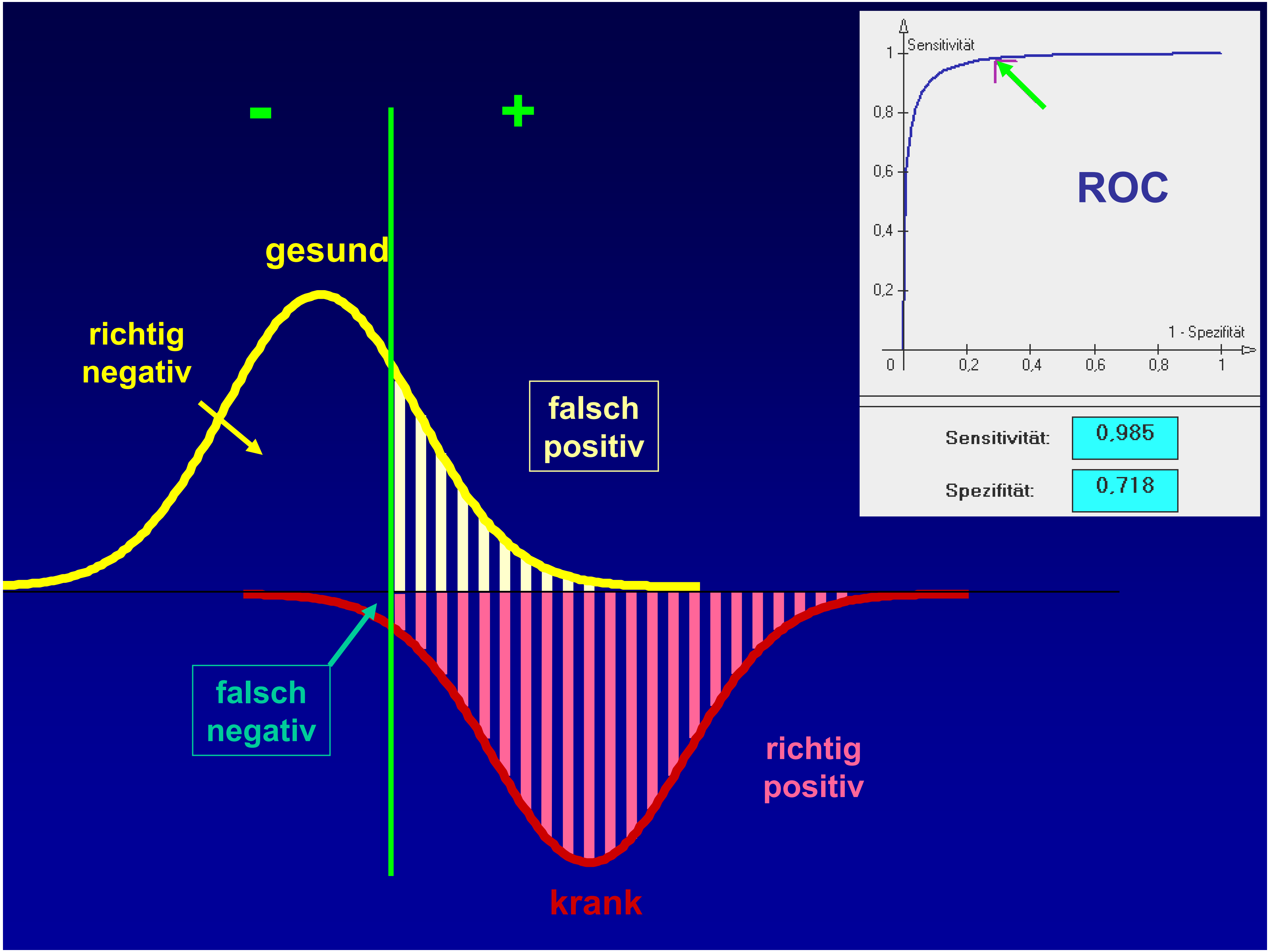

gesund
95 %
Referenzbereich
krank

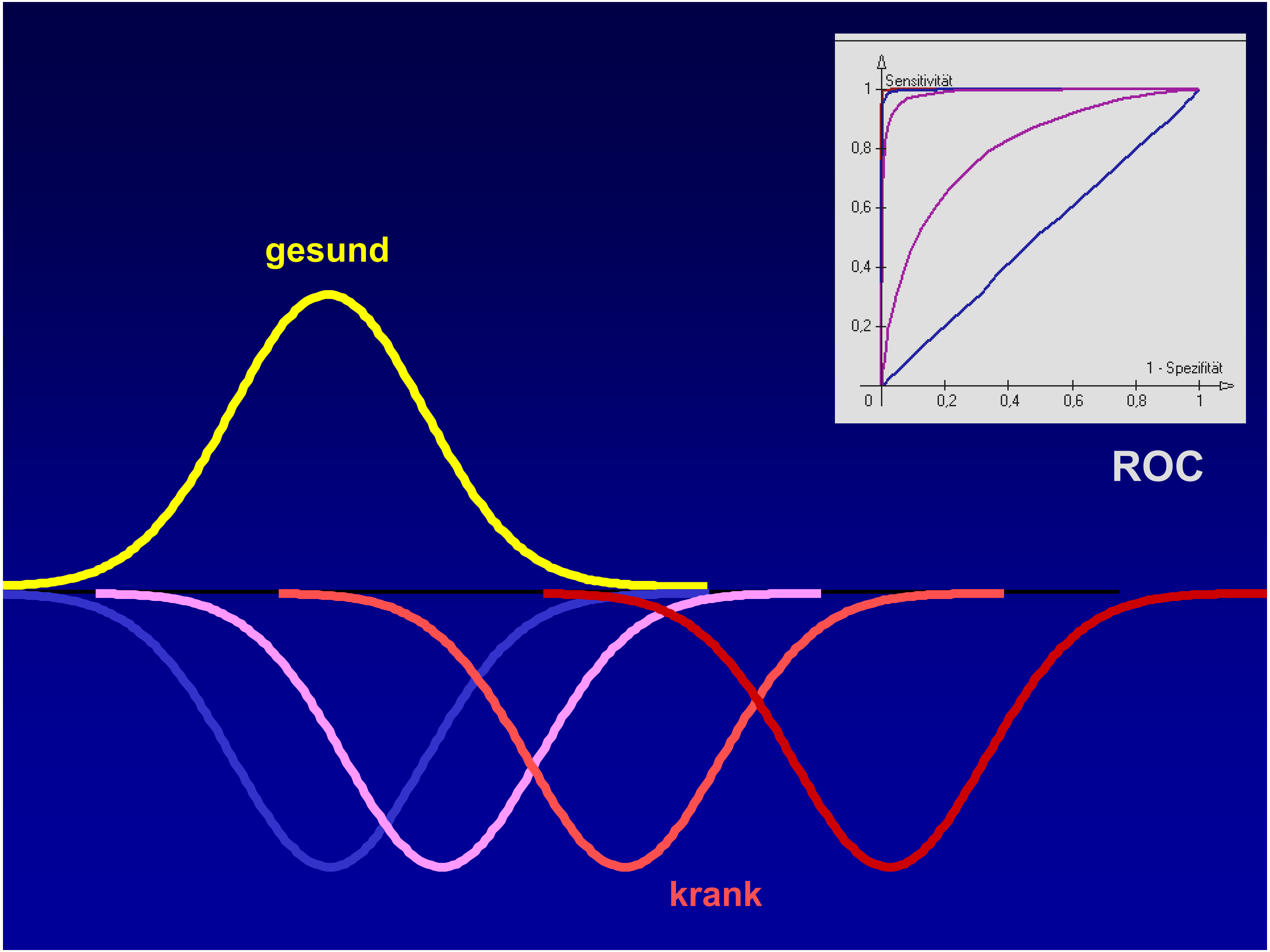


-
+
gesund
richtig negativ
Spezifität
falsch positiv
falsch negativ
richtig positiv
Sensitivität
krank
ROC
Sensitivität
1
0,8
0,6
0,4
0,2
0
0,2
0,4
0,6
0,8
1
1 - Spezifität
Sensitivität: 0,876
Spezifität: 0,943

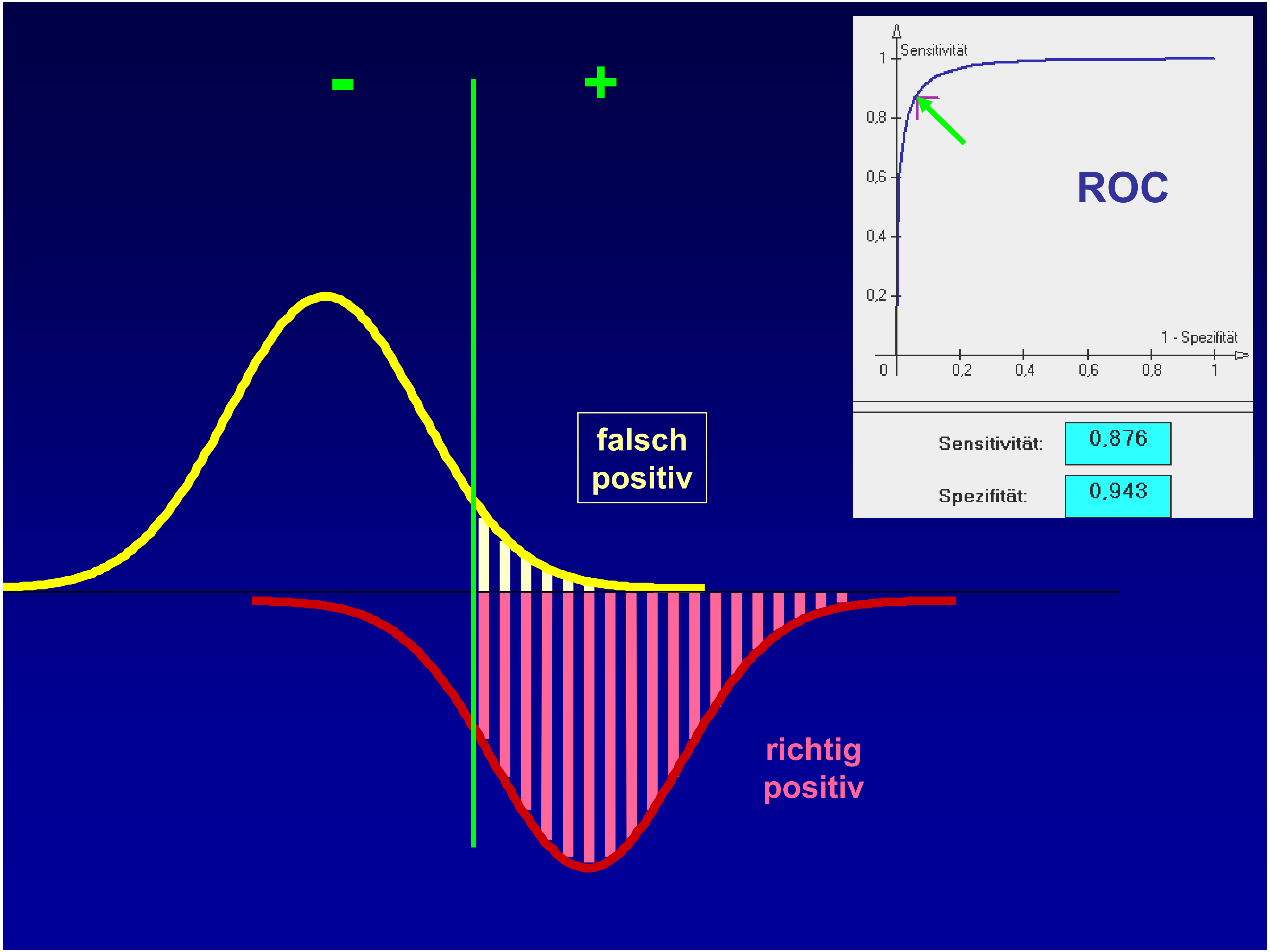


-
+
gesund
richtig
negativ
falsch
positiv
falsch
negativ
richtig
positiv
krank
Sensitivität
1
0,8
0,6
0,4
0,2
0
0,2
0,4
0,6
0,8
1
1 - Spezifität
ROC
Sensitivität:
0,699
Spezifität:
0,985

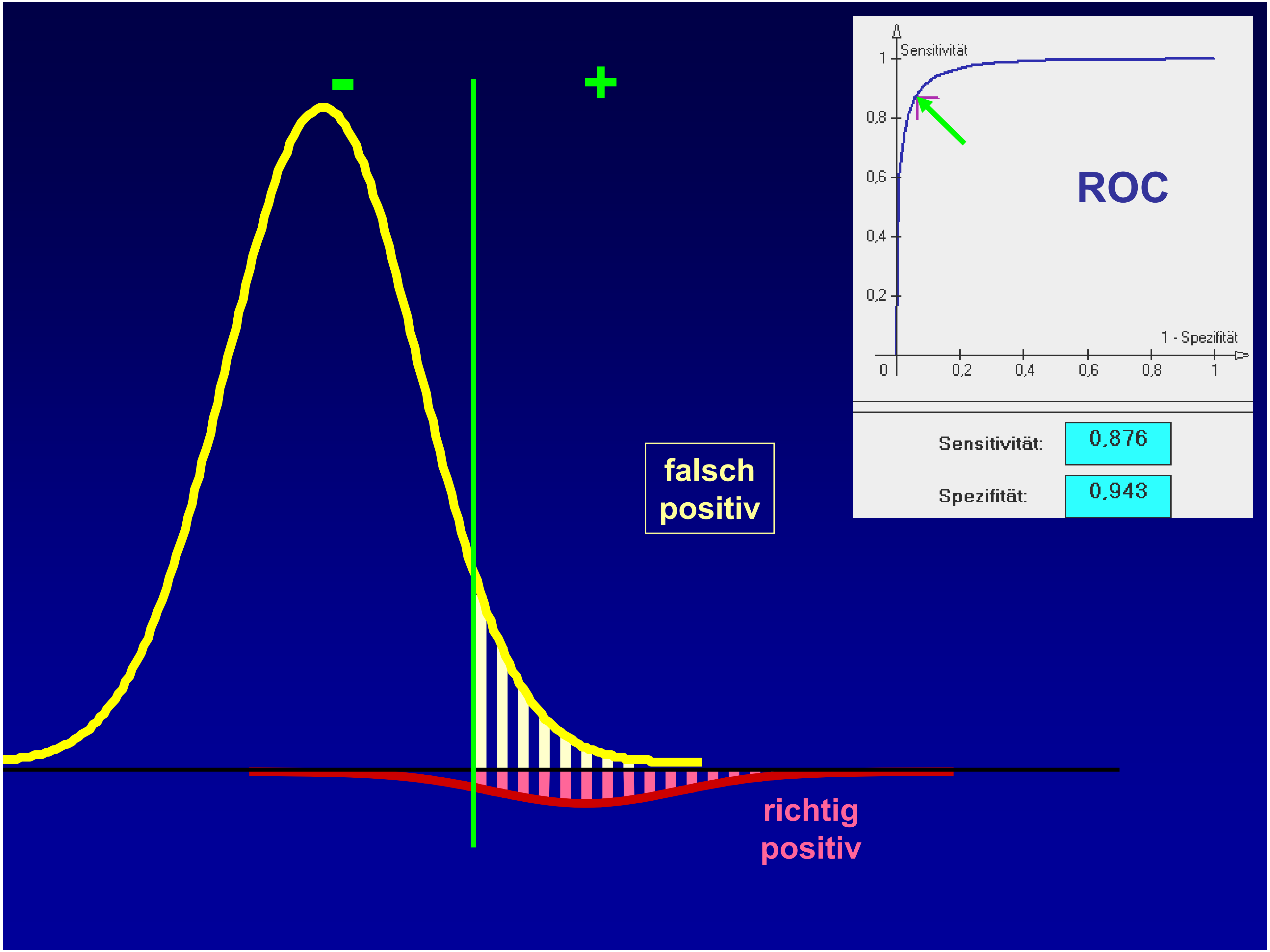

-
+
gesund
richtig negativ
falsch positiv
falsch negativ
richtig positiv
krank
Sensitivität
1
0,8
0,6
0,4
0,2
0
0,2
0,4
0,6
0,8
1
1 - Spezifität
ROC
Sensitivität: 0,985
Spezifität: 0,718



Sensitivität
1
0,8
0,6
0,4
0,2
0
0,2
0,4
0,6
0,8
1
1 - Spezifität
ROC
gesund
krank



-
+
falsch positiv
richtig positiv
ROC
Sensitivität
1
0,8
0,6
0,4
0,2
0
0,2
0,4
0,6
0,8
1
1 - Spezifität
Sensitivität: 0,876
Spezifität: 0,943



-
+
Sensitivität
1
0,8
0,6
0,4
0,2
0
0,2
0,4
0,6
0,8
1
1 - Spezifität
ROC
Sensitivität:
0,876
Spezifität:
0,943
falsch
positiv
richtig
positiv

Sensitivität

p (Test + / krank)

p (krank / Test +)

positiv prädiktiver Wert

Thomas Bayes

Thomas Bayes (1702 – 1761)

Mathematische Theorie der bedingten Wahrscheinlichkeiten

$$p(A/B) = \frac{p(A)\cdot p(B/A)}{p(A)\cdot p(B/A) + p(\bar{A})\cdot p(B/\bar{A})}$$

$$p(K/T^{pos}) = \frac{p(K)\cdot p(T^{pos}/K)}{p(K)\cdot p(T^{pos}/K) + p(\bar{K})\cdot p(T^{pos}/\bar{K})}$$

$$\textbf{pos. präd. Wert} = \frac{\textbf{Prävalenz} \cdot \textbf{Sensitivität}}{\textbf{Prävalenz} \cdot \textbf{Sensitivität} + (1 - \textbf{Prävalenz})\,(1 - \textbf{Spezifität})}$$

A und B Ereignisse:

Multiplikationsregel (MR) :

$$P(A \cap B) = P(A) \quad P(B|A) = P(B) \quad P(A|B)$$

Anmerkung:

1) $$P(B|A) = \frac{P(A \cap B)}{P(A)} \overset{MR}{=} \frac{P(B) * P(A|B)}{P(A)}$$

2) $$P(A) \overset{3.\text{Axiom}}{=} P(A \cap B) + P(A \cap B) \overset{MR}{=} P(B) * P(A|B) + P(\bar{B}) * P(A|\bar{B})$$

Theorem von BAYES

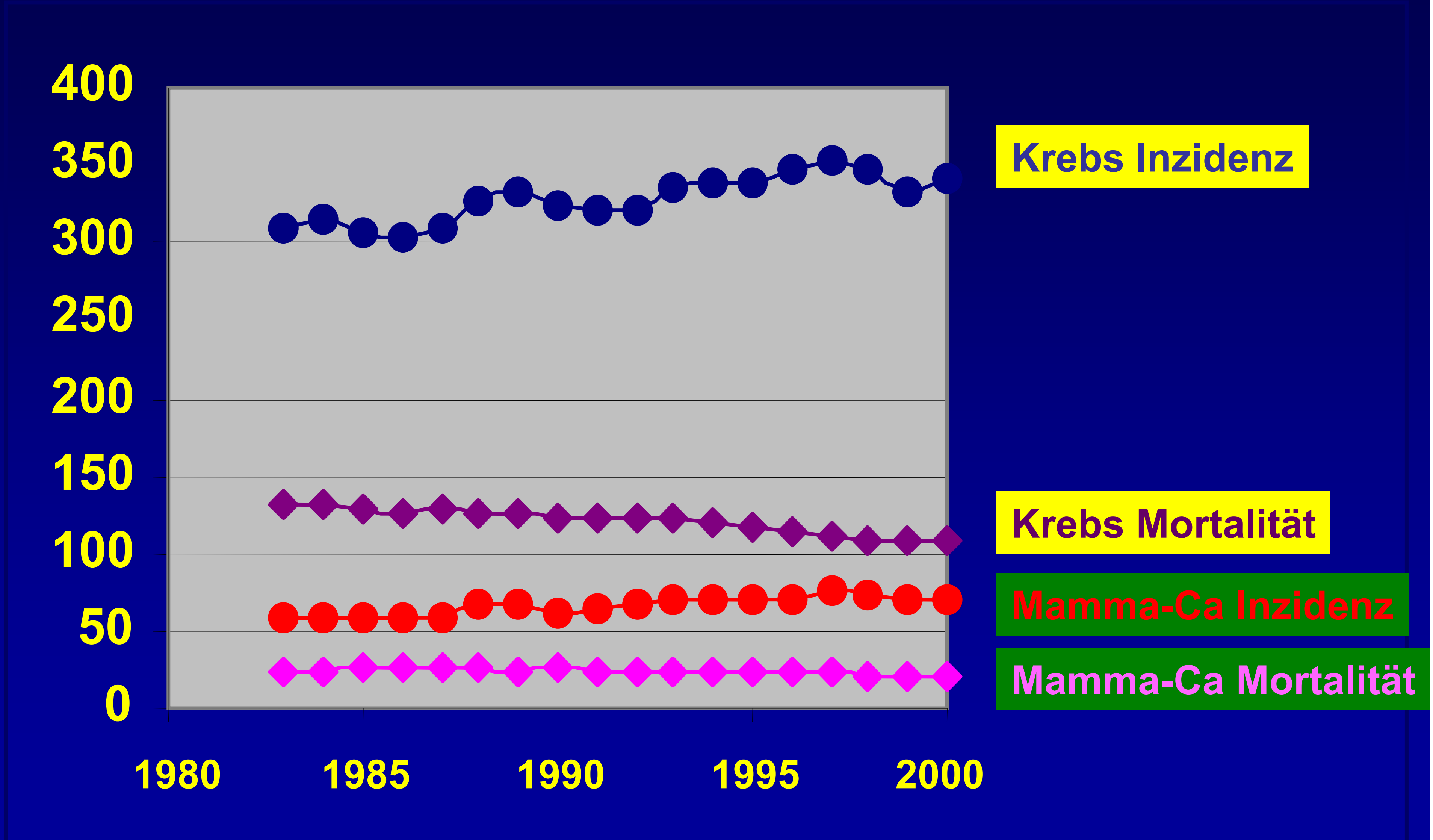
$$\boxed{P(B|A) = \frac{P(B) * P(A|B)}{P(B) * P(A|B) + P(\bar{B}) * P(A|\bar{B})}}$$

Österreichisches Krebsregister und Todesursachenstatistik

	Jahr	Inzidenz maligner Tumore	Mortalität maligner Tumore	Inzidenz Mammakarzinom	Mortalität Mammakarzinom
1	1983	16012	9459	3391	1462
2	1984	15955	9663	3582	1550
3	1985	15868	9526	3544	1618
4	1986	15510	9392	3473	1636
5	1987	15933	9665	3588	1675
6	1988	16848	9578	3997	1678
7	1989	16902	9514	4051	1669
8	1990	16480	9664	3796	1736
9	1991	16455	9522	3939	1671
10	1992	16488	9689	4127	1746
11	1993	17090	9642	4396	1743
12	1994	17159	9546	4396	1728
13	1995	17008	9530	4364	1737
14	1996	17003	9205	4511	1712
15	1997	17650	9214	4852	1651
16	1998	17310	9032	4671	1621
17	1999	16353	9100	4453	1562
18	2000	16901	9225	4577	1671

Jahrbuch der Gesundheitsstatistik

Zeitliche Entwicklung in Österreich
Altersstandardisierte Raten auf 100.000 Frauen



Radiologischer Alltag

		Ca	Kein Ca	
Mammographie	+	**80**	**4** falsch pos	84
	-	**20** falsch neg	**96**	116
Prävalenz 50 %		**100**	**100**	200

Sensitivität 80 / 100 ≈ **80%** P (+ | Ca)

Spezifität 96/ 100 ≈ **96 %** P (- | kein Ca)

pos.präd.Wert 80 / 84 ≈ **95,2 %** **P (Ca | +)**

neg.präd.Wert 96 / 116 ≈ **82,8 %** **P (kein Ca | -)**

Screening

		Ca	Kein Ca	
Mammographie	+	**640**	**3.968**	4.608
	-	**160**	**95.232**	95.392
Prävalenz	**0.8 %**	**800**	**99.200**	100.000

Sensitivität 640 / 800 ≈ **80%** P (+ | Ca)

Spezifität 95.232/99.200 ≈ **96 %** P (- | kein Ca)

pos.präd.Wert 640 / 4.608 ≈ **13,9 %** **P (Ca | +)**

neg.präd.Wert 95.232 / 95.392 ≈ **99,8 %** **P (kein Ca | -)**

		Ca	Kein Ca	
Haemocult	+	richtig positiv 60 %	falsch positiv 3 %	?
	-	falsch negativ 40 %	richtig negativ 97 %	?
		?	?	?

		Ca	Kein Ca	Klinik
Haemocult	+	60	3	63
	-	40	97	137
Prävalenz 50 %		100	100	200

Sensitivität 60 / 100 = **60%** P (+ | Ca)

Spezifität 97/ 100 = **97 %** P (- | kein Ca)

pos.präd.Wert 60 / 63 ≈ **95,2 %** **P (Ca | +)**

neg.präd.Wert 97 / 137 ≈ **70,8 %** **P (kein Ca | -)**

		Ca	Kein Ca	Screening
Haemocult	+	120	3.000	3.120
	-	80	97.000	97.080
Prävalenz 0.2 % Männer 75-80		200	100.000	100.200

Sensitivität 120 / 200 ≈ **60%** P (+ | Ca)

Spezifität 97.000 / 100.000 = **97 %** P (- | kein Ca)

pos.präd.Wert 120 / 3.120 ≈ **3,8 %** **P (Ca | +)**

neg.präd.Wert 97.000 / 97.080 ≈ **99,9 %** **P (kein Ca | -)**

		Ca	Kein Ca	Screening
Haemocult	+	200	3.000	3.200
	-	0	97.000	97.000
Prävalenz 0.2 % Männer 75-80		200	100.000	100.200

Sensitivität 200 / 200 ≈ **100%** **P (+ | Ca)**

Spezifität 97.000 / 100.000 **= 97 %** **P (- | kein Ca)**

pos.präd.Wert 200 / 3.200 ≈ **6,25 %** **P (Ca | +)**

neg.präd.Wert 97.000 / 97.000 = **100 %** **P (kein Ca | -)**

Vorhersagewerte in Abhängigkeit von Se, Sp und Prävalenz

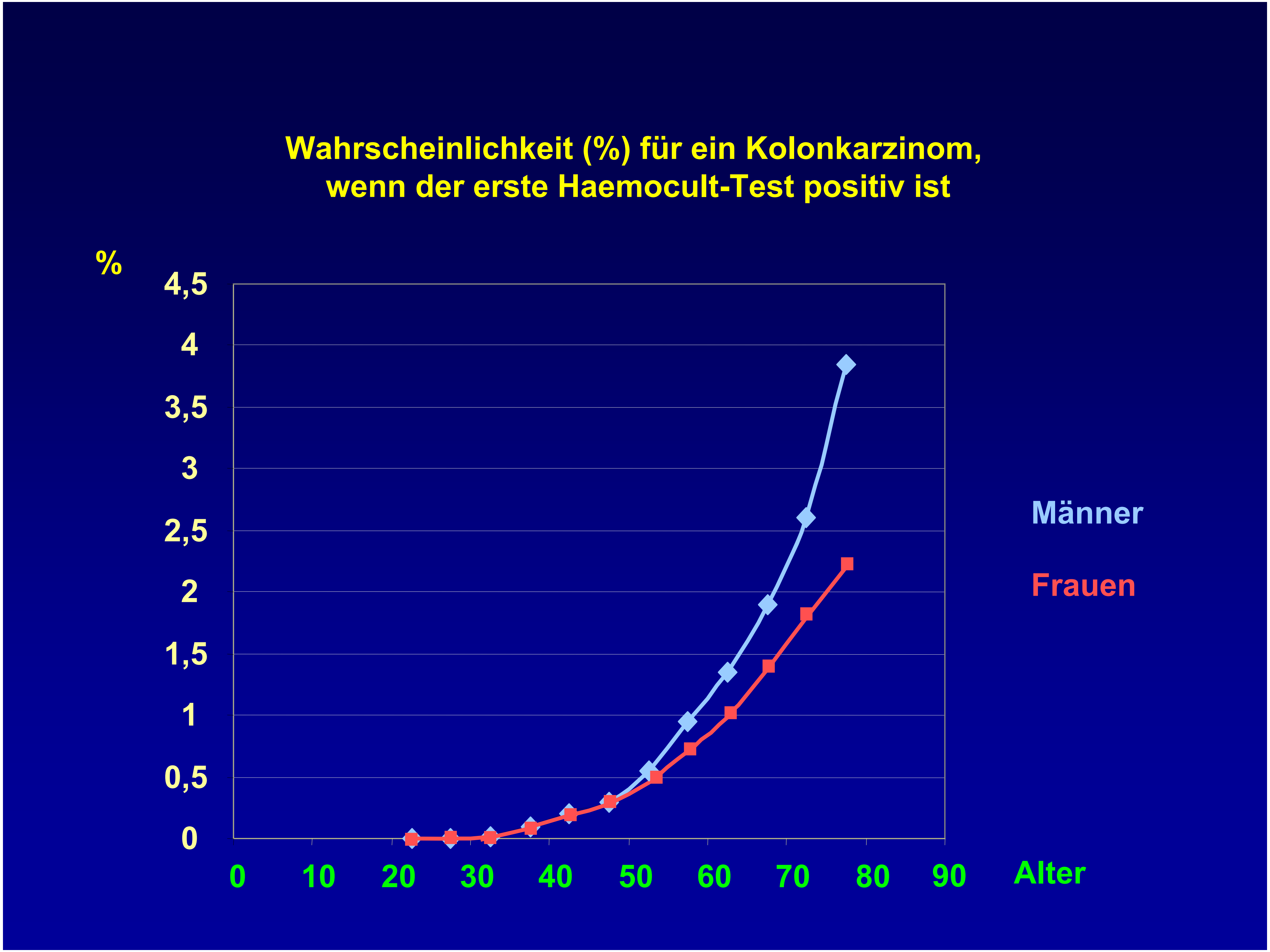
Prävalenz				
	Se	.80	.60	.80
	Sp	.95	.95	.80
.5		.94	.92	.80
.4		.91	.89	.73
.3		.87	.84	.63
.2		.80	.75	.50
.1		.64	.57	.31
.05		.46	.38	.17
.01		.14	.11	.04
.001		.016	.012	.004
.0001		.0016	.0012	.0004

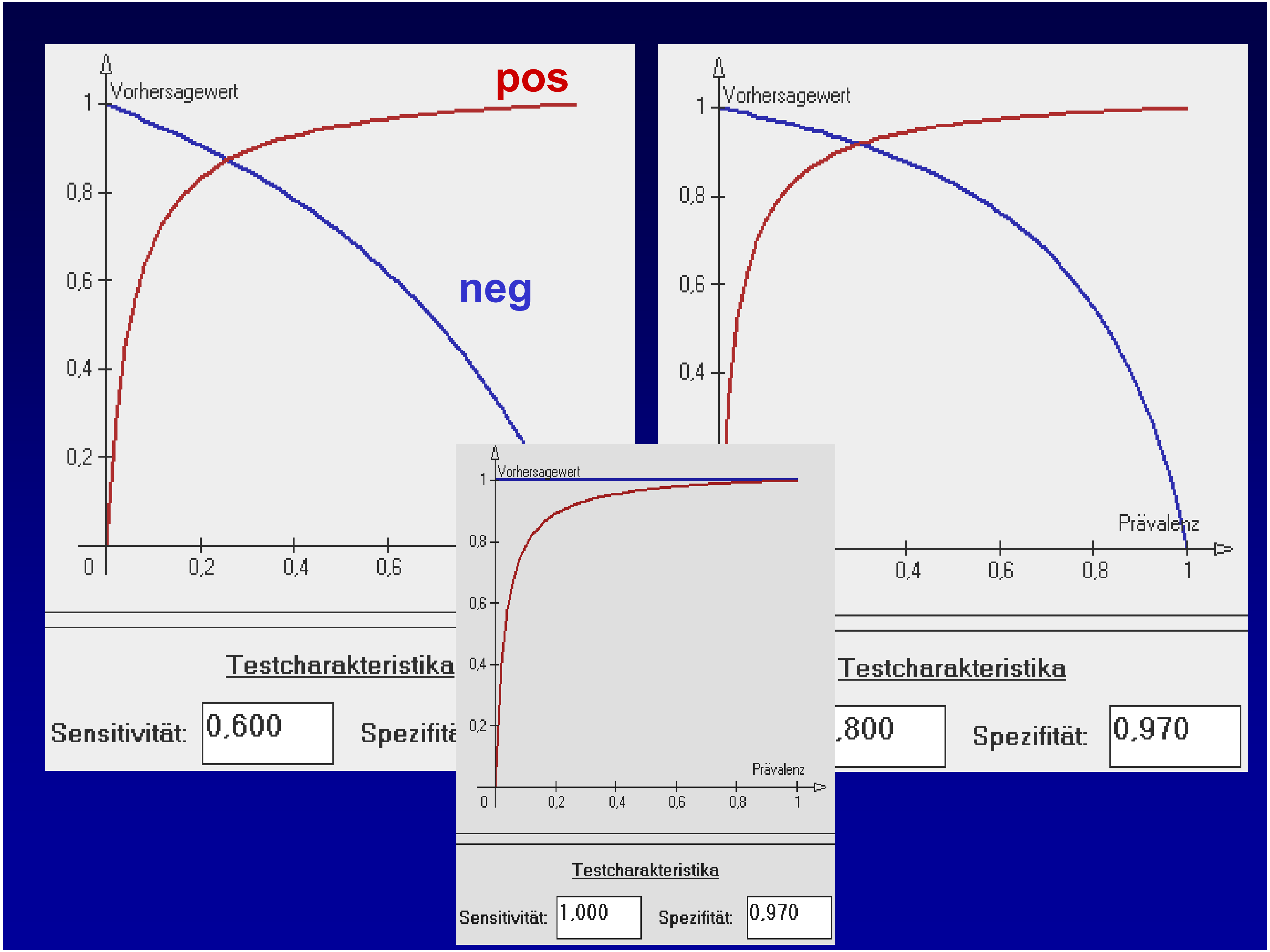
Positiver
Prädiktiver
Wert

Rate falsch positiver Befunde bei einem HIV-ELISA in Abhängigkeit von der HIV-Seroprävalenz und der Güte des Tests.

Sensitivität	.999	.999	.999
Spezifität	.985	.995	.998
Prävalenz			
.50	1.5%	0.5%	0.2%
.25	5.7%	1.5%	0.6%
.10	11.9%	4.3%	1.8%
.01	59.8%	33.1%	16.5%
.001	93.8%	83.3%	66.6%
.0001	99.3%	98.0%	95.2%

Derzeitiges Zulassungskriterium für einen HIV-Antikörper-Test: die Sensitivität beim Suchtest und die Spezifität beim Bestätigungstest müssen über 99% (.99) liegen.

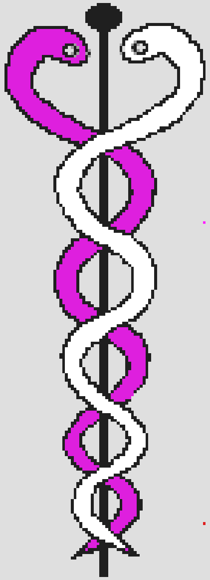

Wahrscheinlichkeit (%) für ein Kolonkarzinom,
wenn der erste Haemocult-Test positiv ist
%
4,5
4
3,5
3
2,5
2
1,5
1
0,5
0
0
10
20
30
40
50
60
70
80
90
Alter
Männer
Frauen



pos
neg
Vorhersagewert
1
0,8
0,6
0,4
0,2
0
0,2
0,4
0,6
Testcharakteristika
Sensitivität:
0,600
Vorhersagewert
1
0,8
0,6
0,4
Prävalenz
0,4
0,6
0,8
1
Testcharakteristika
,800
Spezifität:
0,970
Vorhersagewert
1
0,8
0,6
0,4
0,2
Prävalenz
0
0,2
0,4
0,6
0,8
1
Testcharakteristika
Sensitivität:
1,000
Spezifität:
0,970

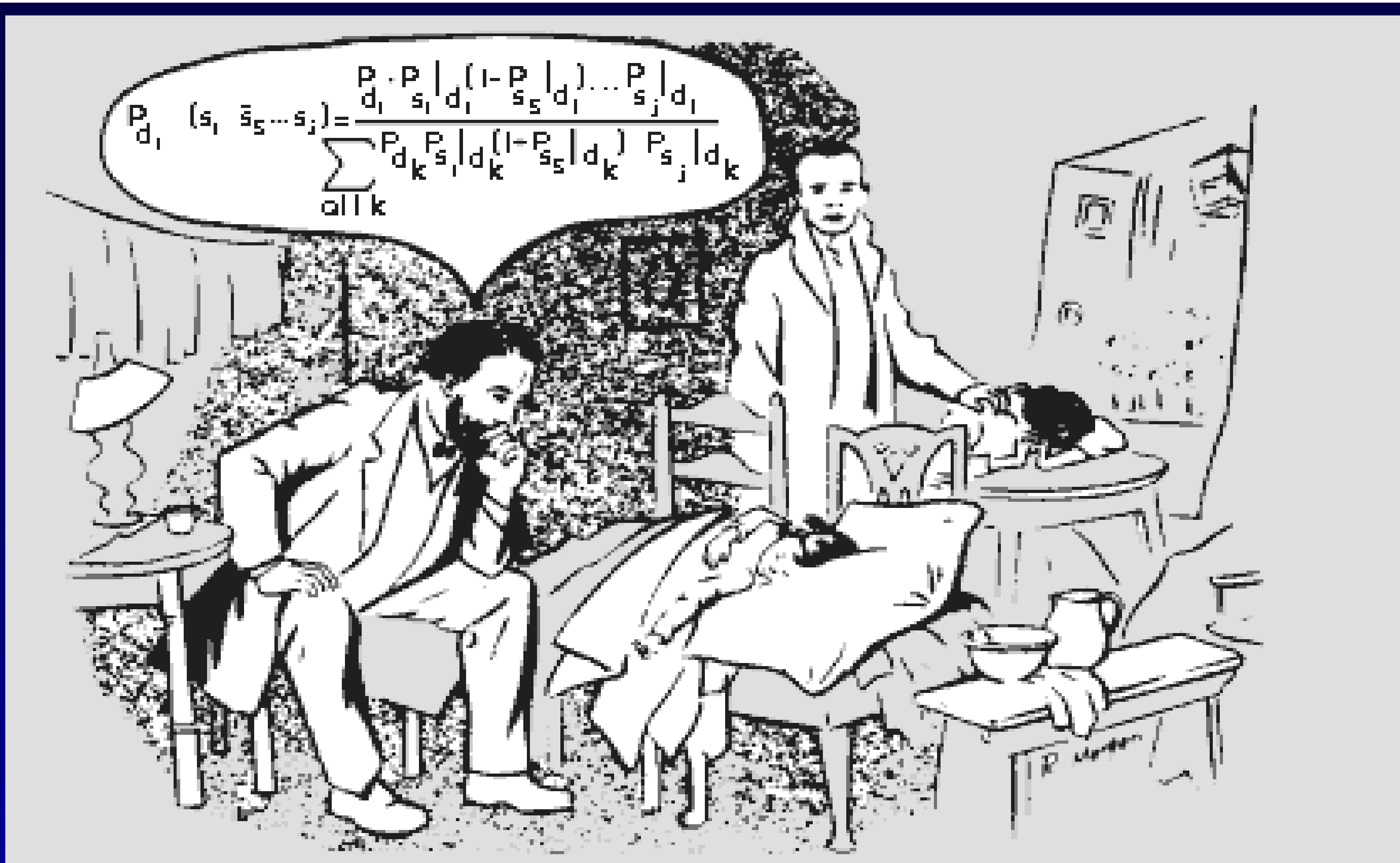
Prävalenzen einiger Krankheiten in der Bevölkerung (Europa) - Screening

Hypertonie (über 50-jährige)	0,2	Multiple Sklerose	0,0003
Diabetes mellitus	0,04	Adrenogenitales Syndrom	0,0002
Asthma bronchiale	0,03	Morbus Cushing	0,0001
Schizophrenie	0,01	Lupus erythematodis	0,0001
Herzfehler	0,01	Phenylketonurie	0,0001
Skoliose	0,01	Leukämie	0,00005
Epilepsie	0,005	Alkaptonurie	0,00004
Down-Syndrom	0,0015	Myasthenia gravis	0,00003
Klinefelter-Syndrom	0,0013	Morbus Hodgkin	0,00002
Klumpfuß	0,0010	Pfaundler-Hurler	0,00001
Mukoviszidose	0,0005	Wilson'sche Krankheit	0,000003
Turner-Syndrom	0,0004	Brucellose	0,000001



Prävalenz klein - Anteil an Falsch-Positiven groß

Liegen aber bereits bestimmte Symptome einer Erkrankung vor, dann wird die a-priori-Wahrscheinlichkeit größer sein.



Die klinische Problemlösung mit der strengen Logik quantitativer Methoden zu durchleuchten, lohnt sich.

Dennoch bleibt die klinische Entscheidungsfindung ein schwer durchschaubarer Prozess der auf Erfahrung und Lernen, induktivem und deduktivem Denken und auf einer kaum definierbaren Intuition beruht.

		Ca	Kein Ca	
Haemocult	+	richtig positiv 60 %	falsch positiv 3 %	?
	-	falsch negativ 40 %	richtig negativ 97 %	?
		?	?	?

		Med wirkt	Med wirkt nicht	
Statistischer Test	signifikant	**80**	**5**	85
	nicht sign.	**20**	**95**	115
wiss. Riecher **50 %**		**100**	**100**	200

power 80 / 100 = **80%** **P (sign | Med wirkt)**

95 / 100 = **95 %** **P (n.s. | Med wirkt nicht)**

β = 20 % **P (nicht sign | Med wirkt)**

Signifikanzniveau 5 % **P (sign | Med wirkt nicht)**

pos.präd.Wert 80 / 85 ≈ 94,1 % P (Med wirkt | Test sign.)

neg.präd.Wert 95 / 115 ≈ 82,6 % P (Med wirkt nicht | Test nicht sign.)

		Med wirkt	Med wirkt nicht	
Statistischer Test	signifikant	**8**	**10**	18
	nicht sign.	**2**	**180**	182
wiss. Riecher **5 %**		**10**	**190**	200

power 8 / 10 **= 80%** **P (sign | Med wirkt)**

Signifikanzniveau 5 % **P (sign | Med wirkt nicht)**

pos.präd.Wert 8 / 18 = 44,4 % P (Med wirkt | Test sign.)

neg.präd.Wert 180 / 182 ≈ 98,9 % P (Med wirkt nicht | Test nicht sign.)

Vergleich diagnostischer Entscheidungen ohne Referenzkriterium

Wenn kein Referenzkriterium zur Verfügung steht, wird die diagnostische Basismatrix mit anderen statistischen Methoden evaluiert als bei Vorliegen eines Goldstandards.

Mögliche Ergebniskonstellationen		*Ergebnis des zweiten Diagnostischen Tests*	
		positiv	**negativ**
Ergebnis des ersten Diagnostischen Test	**positiv**	**positiv** **positiv**	**positiv** **negativ**
	negativ	**negativ** **positiv**	**negativ** **negativ**

Beispiel:

		Diagnostischer Test 2		
		positiv	negativ	Gesamt
Diagnostischer Test 1	positiv	**40**	**40**	80
	negativ	**20**	**900**	920
	Gesamt	60	940	**1000**

Fragestellungen:

1. Liefert eines der beiden Diagnoseverfahren *systematisch* häufiger ein positives Testergebnis?
   Lösung: **Mc. Nemar Test**, Vergleich 40 (positiv/negativ) vs 20 (negativ/positiv), ergibt: p=0.014.

2. Wie groß ist der Grad an *Übereinstimmung* zwischen beiden Untersuchern?
   Lösung: **Kappa-Maß**

Das Kappa-Maß

Im vorausgehenden Beispiel könnte man von **94%** Übereinstimmung sprechen:

$$\frac{\text{Anzahl beide positiv} \quad + \quad \text{Anzahl beide negativ}}{\text{Gesamtzahl}}$$

$$= \quad (40 + 900) / 1000 \quad = \; 94\%$$

Dieser Wert wäre zu *optimistisch.*

Würde z.B. das zweite Diagnoseverfahren für alle Gesunden und alle Kranken immer ein negatives Ergebnis liefern,
dann berechnet sich die Übereinstimmung: 920 / 1000 = **92%**

		Diagnostischer Test 2		
		positiv	negativ	Gesamt
Diagnostischer Test 1	positiv	**40**	**40**	80
	negativ	**20**	**900**	920
	Gesamt	60	940	**1000**

		Diagnostischer Test 2		
		positiv	negativ	Gesamt
Diagnostischer Test 1	positiv	**0**	**80**	80
	negativ	**0**	**920**	920
	Gesamt	0	1000	**1000**

Man muss daher berücksichtigen, dass es *zufällig* zu gleichen Diagnoseentscheidungen kommen kann.

Im Beispiel sind 920*940/1000 = 865 negative und 40*60/1000 = 2 positive Übereinstimmungen zufällig zu erwarten.

Bei Berechnung von kappa werden diese nicht berücksichtigt.
Das heißt, sowohl im Zähler als auch im Nenner wird die beobachtete Häufigkeit um die zu erwartenden Übereinstimmungen verringert.

$$\frac{\text{Beobachtete } \textbf{minus} \text{ zufällig zu erwartende Übereinstimmungen}}{\text{Gesamtzahl } \textbf{minus} \text{ zufällig zu erwartende Übereinstimmungen}}$$

= (940-867) / [1000-867] = 55%

Dieser Wert ist realistischer als die „naiv“ berechneten 94%.

		Diagnostischer Test 2		
		positiv	negativ	Gesamt
Diagnostischer Test 1	positiv	**40**	**40**	80
	negativ	**20**	**900**	920
	Gesamt	60	940	**1000**

Kriterien für die Güte der Übereinstimmung

(nach Landis und Koch)

Kappa	Übereinstimmung
< 0	grauslich
0 – 0.20	schlecht
0.21 – 0.40	schwach
0.41 – 0.60	mäßig
0.61 – 0.80	stark
> 0.8	so gut wie perfekt

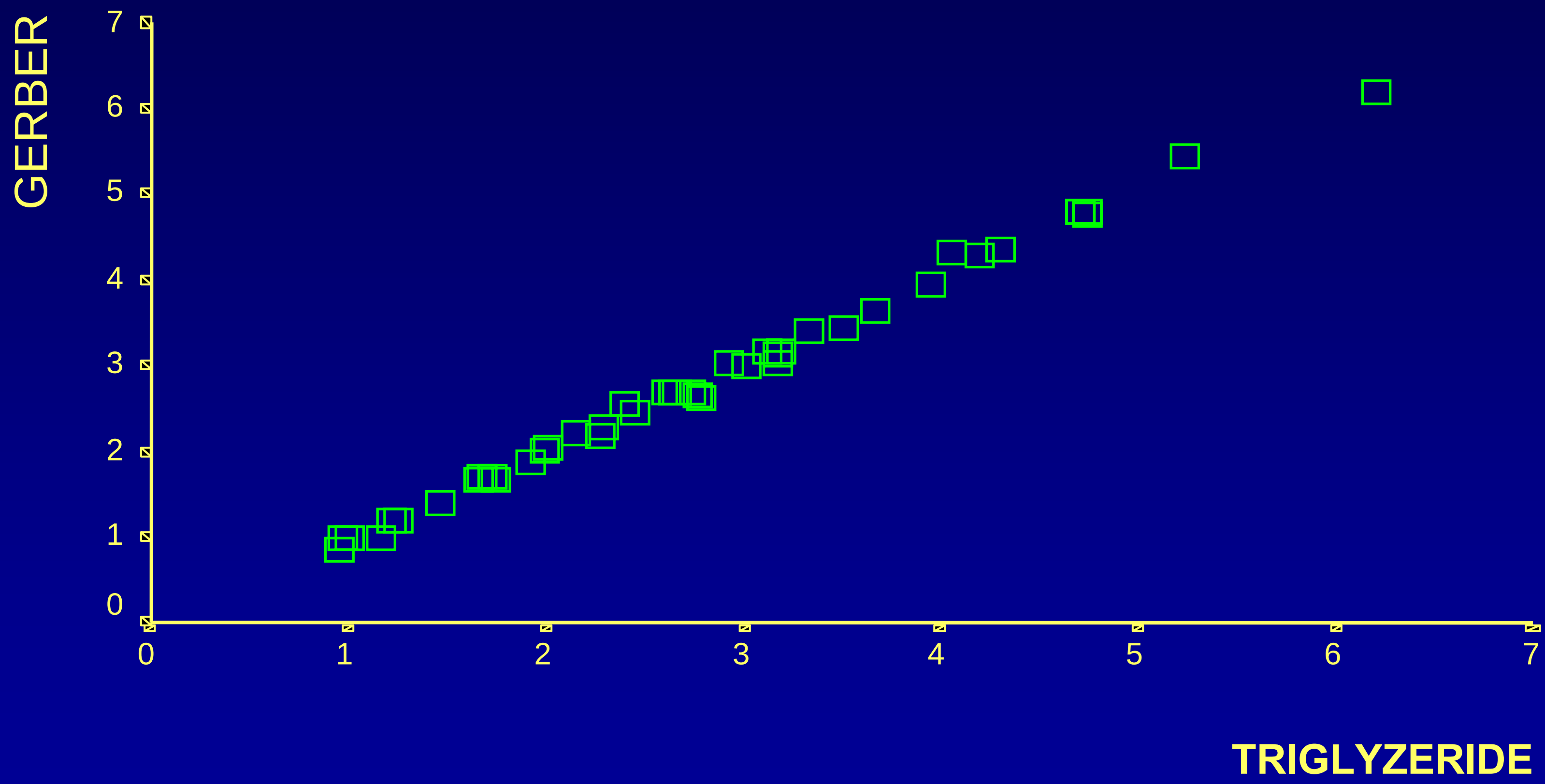
Die Methode von Altman und Bland

Liefern Diagnoseverfahren Messergebnisse auf *stetigen* Skalen, müssen diese für diagnostische Entscheidungen *klassifiziert* werden (z.B. „krank“ vs „gesund“, oder „Therapie“ vs „Zusatzdiagnostik“ vs „keine Therapie“).

Für die Quantifizierung der Übereinstimmung kann man aber auch die *ursprünglichen* Daten verwenden. Eine Möglichkeit der graphischen Darstellung wäre ein *Streudiagramm*: Auf der x-Achse würde das erste Verfahren, auf der y-Achse das zweite Verfahren aufgetragen.

Bei zwei hoch korrelierenden Verfahren würde sich allerdings ein Diagramm ergeben, in dem die Einzelpunkte sehr eng an einer Geraden liegen, das Diagramm würde also überwiegend aus *weißer Fläche* bestehen. In der folgenden Graphik ist dies für zwei Messungen des Fettgehalts in der Muttermilch dargestellt:

Vergleich von zwei Messverfahren für den Fettgehalt der Muttermilch

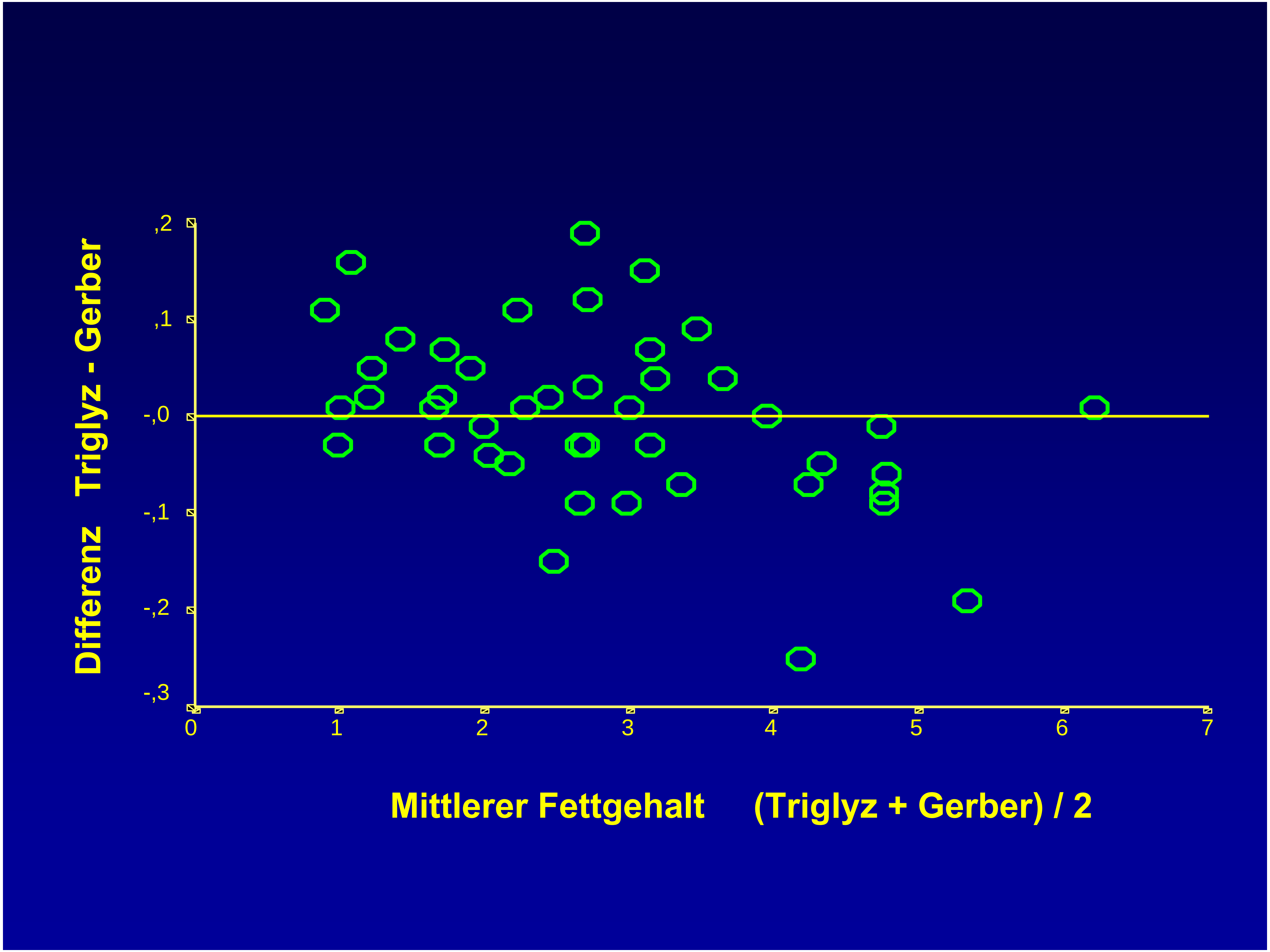


Von *Altman* und *Bland* wurde eine graphische Methode entwickelt, die dem herkömmlichen Streudiagramm deutlich überlegen ist, allerdings voraussetzt, dass beide Messungen auf derselben Skala erfolgen:

x-Achse: Mittelwert der Messungen ***y-Achse:*** Differenz der Messungen

Vorteile der Darstellung von Altman und Bland:

- Die Fläche des Diagramms wird besser genutzt
- Tatsächliche Abweichungen sind besser ablesbar
- Eine mögliche Abhängigkeit der Abweichungen von der Größenordnung der Werte ist besser erkennbar



,2
,1
-,0
-,1
-,2
-,3
0
1
2
3
4
5
6
7
Differenz Triglyz - Gerber
Mittlerer Fettgehalt (Triglyz + Gerber) / 2

Die Phasen der Klinischen Prüfung

Präklinik (Tierversuche, Toxikologie)

I **Medikament bei gesunden Probanden**
Dosis, Verträglichkeit, Pharmakokinetik

II **Medikament bei ausgewählten Patienten**
Wirkungsnachweis

III **Klinische Therapiestudie**
Wirksamkeitsnachweis

Zulassung durch Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM)

IV **Breite des Anwendungsbereichs**

Fall-Kontroll-Studie

- Design retrospektiv, vergleichend

- Prinzip erkrankte Personen
  Bildung einer geeigneten Kontrollgruppe
  (matched pairs)

- Ziel Klärung ätiologischer Fragestellungen

- ! Erhebungsschwierigkeiten
  Selektion in der Stichprobe der Erkrankten

- Analyse odds ratio

Kohortenstudie (Prospektiv)

- Design — prospektiv
- Prinzip — exponierte Personen
  Folgebeobachtungen
- Ziel — Erfassung von unerwünschten Ereignissen
  (Nebenwirkungen)
- ! — Zeitaufwand und drop-out Probleme
  unterschiedliche Beobachtungsintensitäten
  in den Kohorten
- Analyse — Erkrankungsinzidenzen
  relatives Risiko

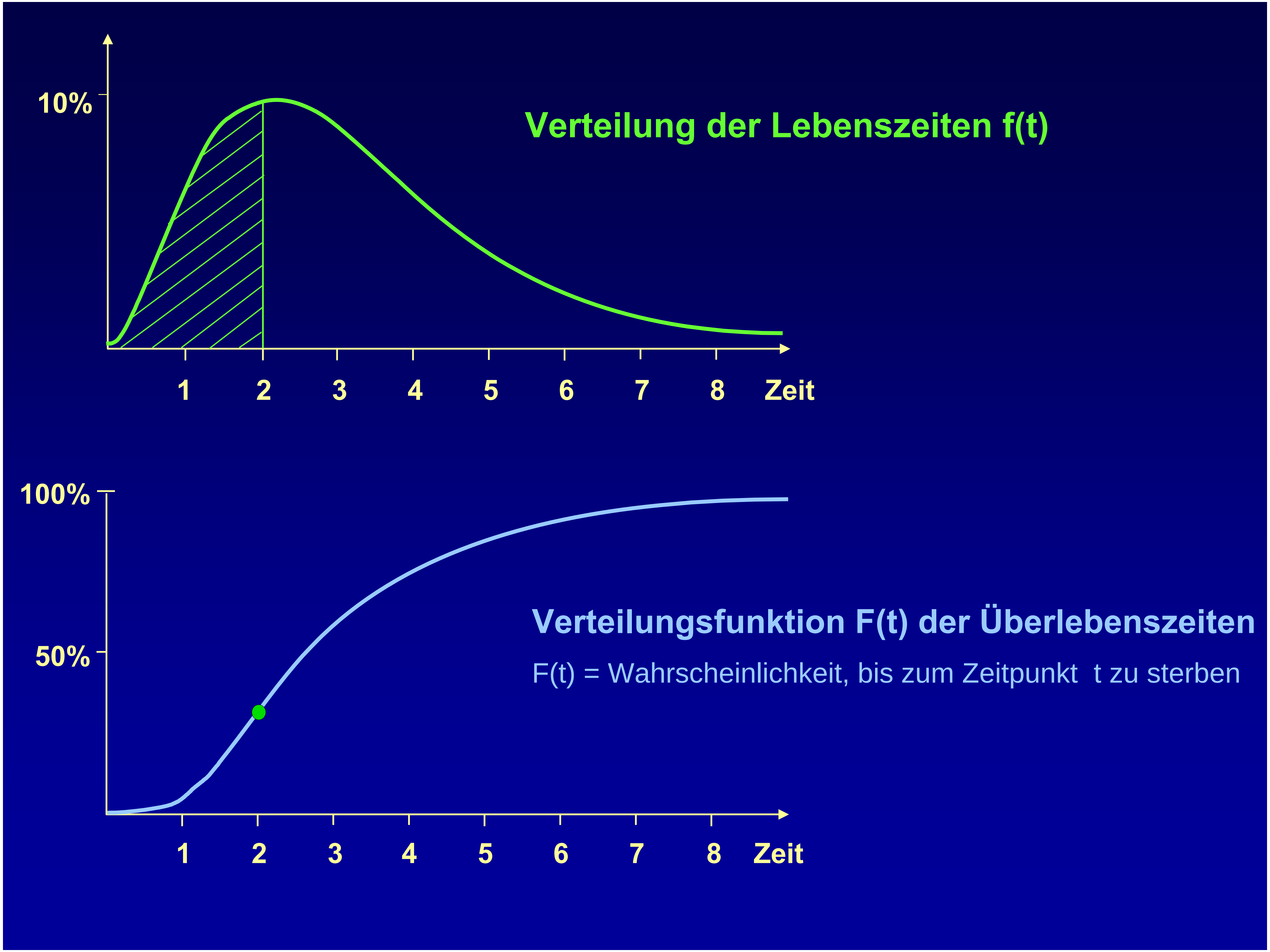
III

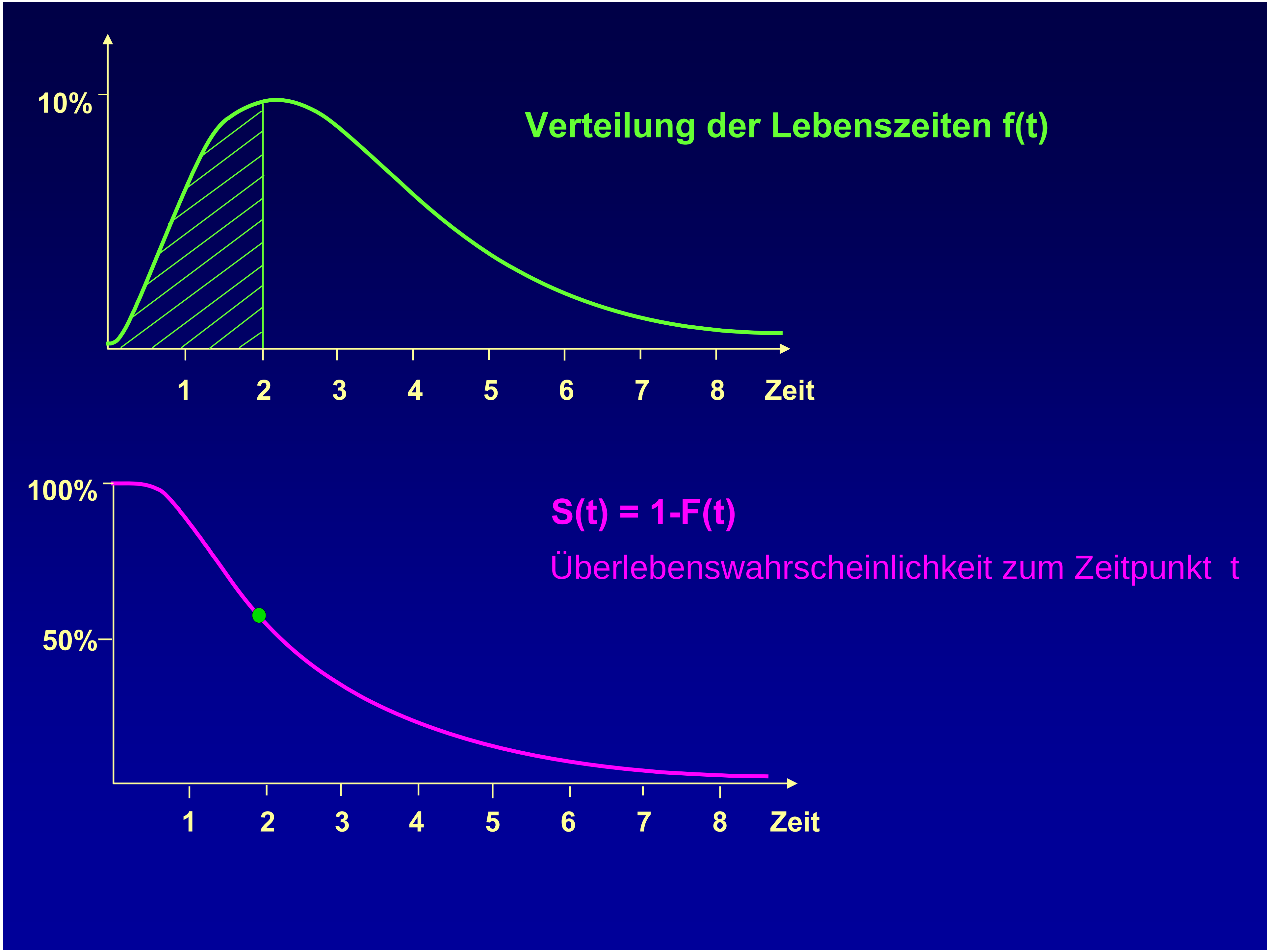
Klinische Therapiestudie

-Design	**prospektiv, kontrolliert, randomisiert, verblindet, offen** **parallel, cross-over ...** **Anzahl der Zentren und Gruppen**
- Ziel	**Definition von Haupt- und Nebenzielgrößen**
- Patienten	Ein- und Ausschlußkriterien Dosierungsschema, Behandlungs- bzw. Untersuchungsplan *Patientenaufklärung, Patientenversicherung* **Compliance** **drop-outs** **Randomisierungsplan** **Stichprobenumfangschätzung**
- Daten	**Monitoring** **Datenmanagement**
- Auswertung	**statistische Testverfahren** *(Name, Signifikanzniveau α)* **explorative Datenanalyse**

Biomathematik

Überlebensanalyse


10%
Verteilung der Lebenszeiten f(t)
1
2
3
4
5
6
7
8
Zeit
100%
50%
Verteilungsfunktion F(t) der Überlebenszeiten
F(t) = Wahrscheinlichkeit, bis zum Zeitpunkt t zu sterben
1
2
3
4
5
6
7
8
Zeit


10%
Verteilung der Lebenszeiten f(t)
1
2
3
4
5
6
7
8
Zeit
100%
50%
S(t) = 1-F(t)
Überlebenswahrscheinlichkeit zum Zeitpunkt t
1
2
3
4
5
6
7
8
Zeit

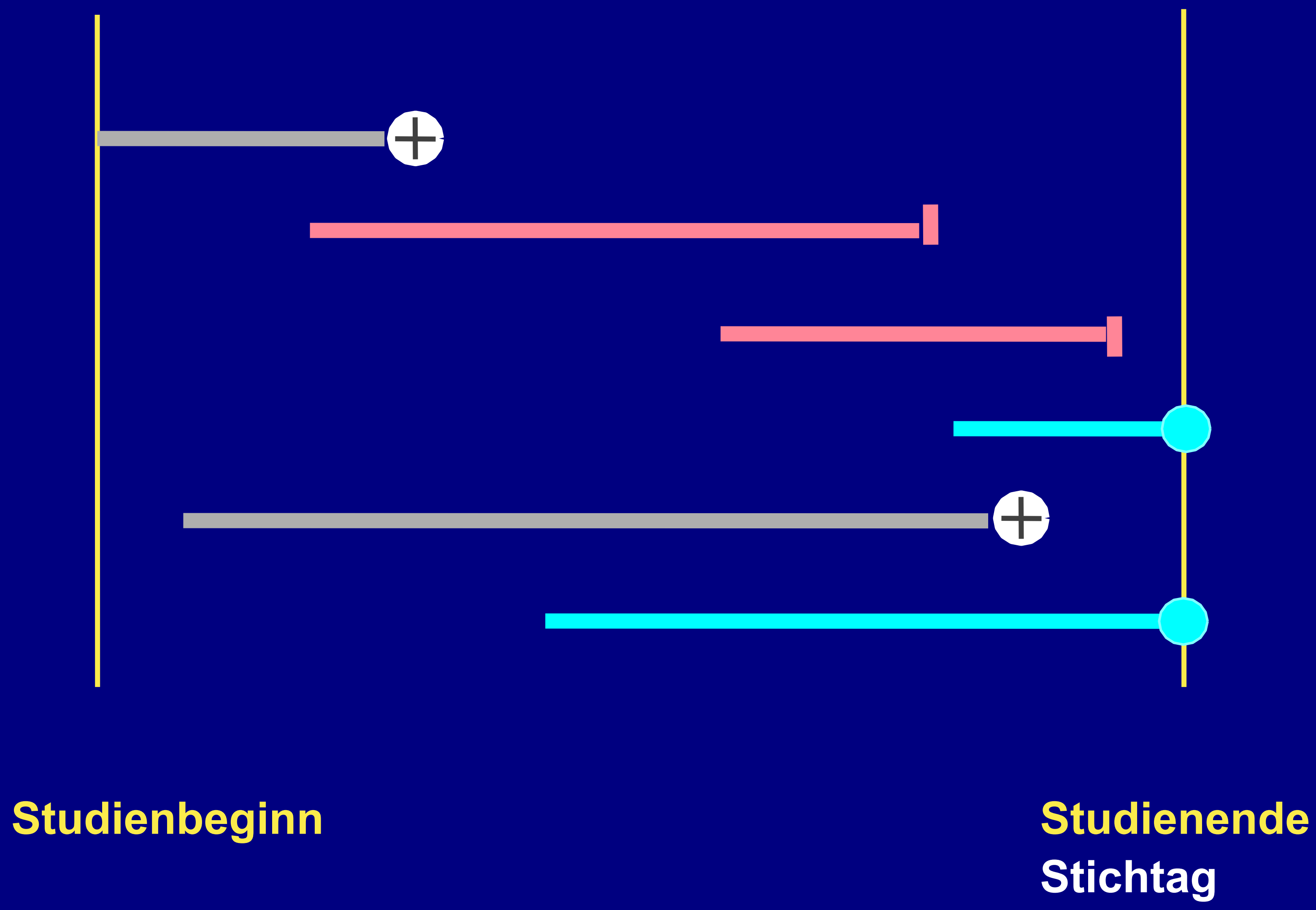
Überlebensanalysen in der Medizin

Die statistische Beschreibung und Analyse von Überlebenszeiten wirft spezielle Probleme auf, die bei anderen stetigen Merkmalen, wie z.B. systolischer Blutdruck [mmHg] oder Linsenstärke [dpt] nicht auftreten.

Eine Überlebenszeitstudie besteht aus zwei Zeitabschnitten: Im ersten werden die Patienten aufgenommen (Rekrutierungszeit), im zweiten werden sie nachbeobachtet (Follow Up). Für Patienten, die am Anfang der Rekrutierungszeit aufgenommen werden, ist das Follow Up maximal, für Patienten, die erst am Ende der Rekrutierungszeit aufgenommen werden, wird im allgemeinen ein Mindest-Follow Up verlangt.

In Bezug auf einen gemeinsamen zeitlichen Nullpunkt (Synchronisierung, z.B. Diagnose) erhält man also unterschiedliche Follow Up Zeiten für die einzelnen Patienten. Dies gilt analog auch für andere Ereignisse in der Zeit, z.B. für die Analyse von Rezidiven oder Spätkomplikationen.

Zeitabhängige Beobachtungen



life-tables - Verweildaueranalysen -

Erfolg Nichteintreffen eines oder mehrerer Zielereignisse

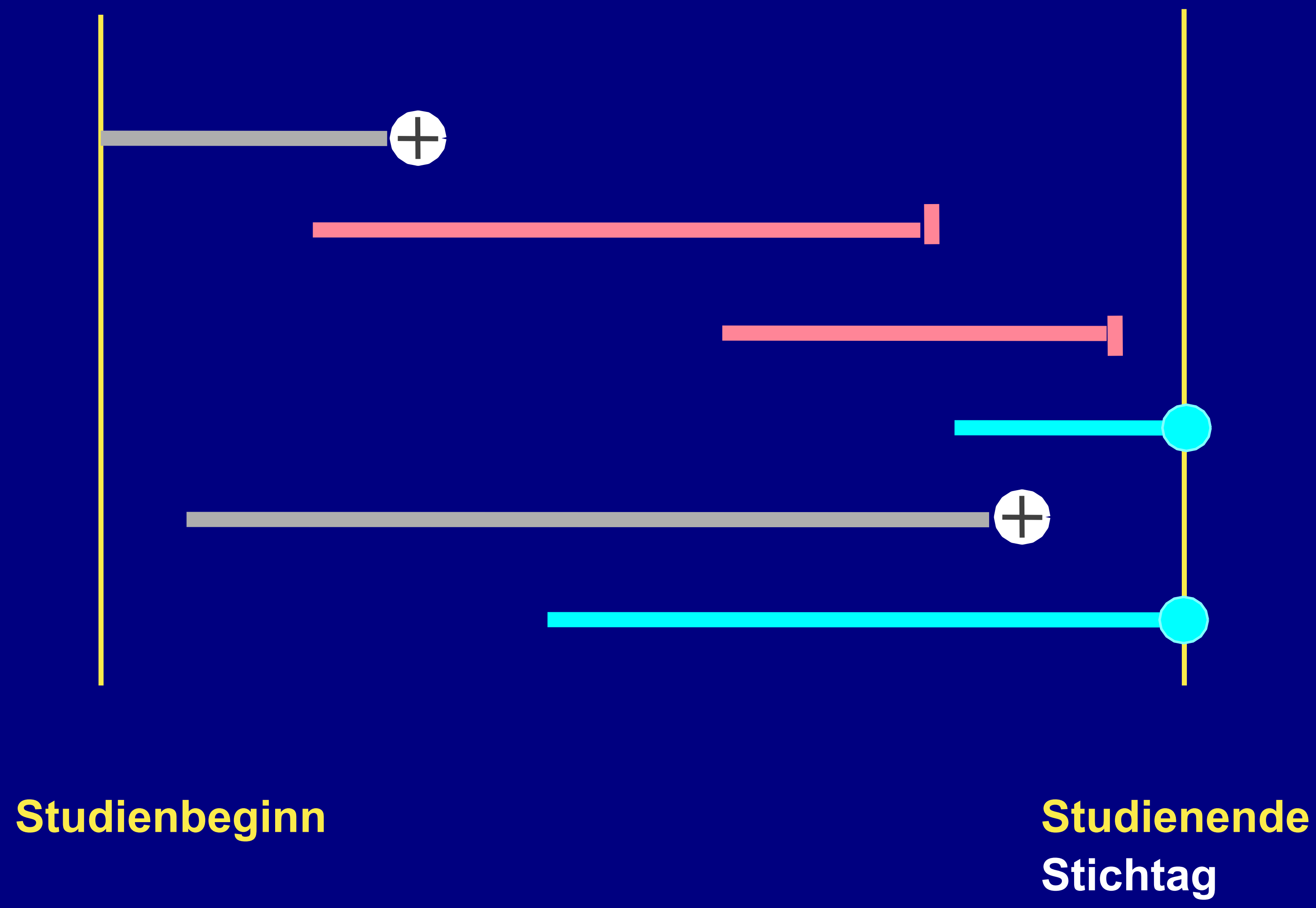
- Implantatverlust,
- Erreichen eines Lockerungsgrades
- vertikaler Knochenverlust > 0.2 mm/Jahr
- Schmerzen, Infektionen
- Suprastruktur defekt oder nicht ästhetisch
- negative Beurteilung durch den Patienten

Verweildauer: Zeiteinheit zwischen Implantation bzw Eingliederung und dem Eintreffen des Ereignisses

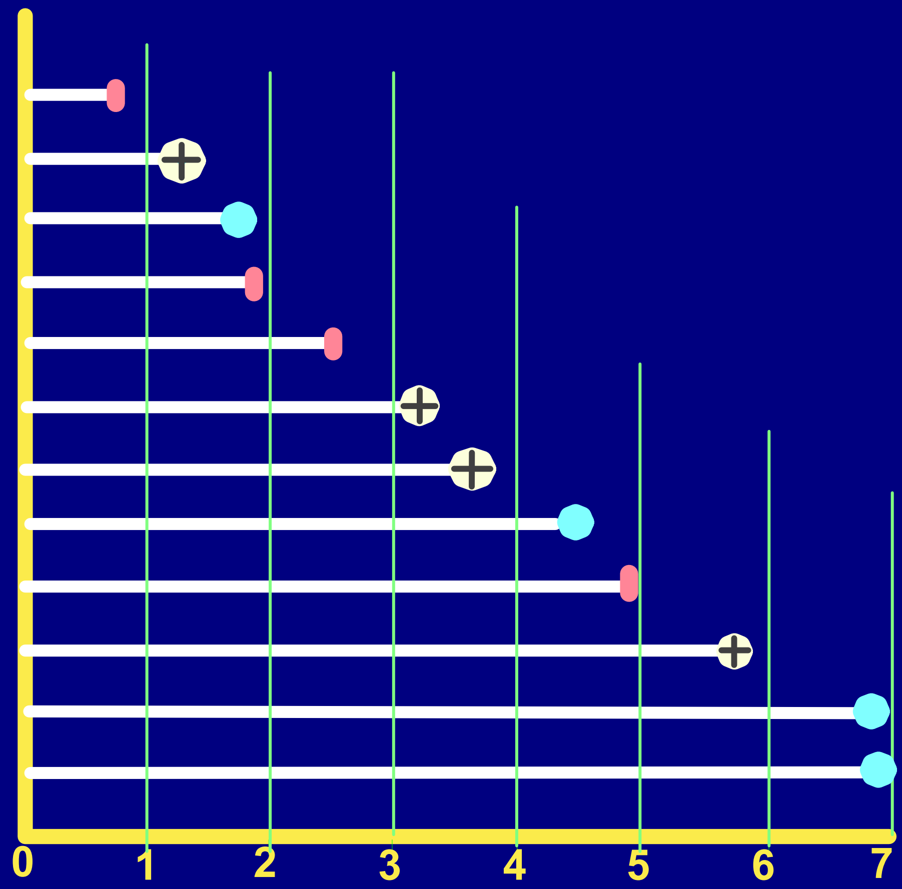
Zensierte Verweildauer: Ereignis wird nicht beobachtet

lost case	Ausscheiden wegen anderer Ursachen (Umzug, Trauma, Tod)
withdrawal	Beobachtungszeit zu kurz

Zeitabhängige Beobachtungen

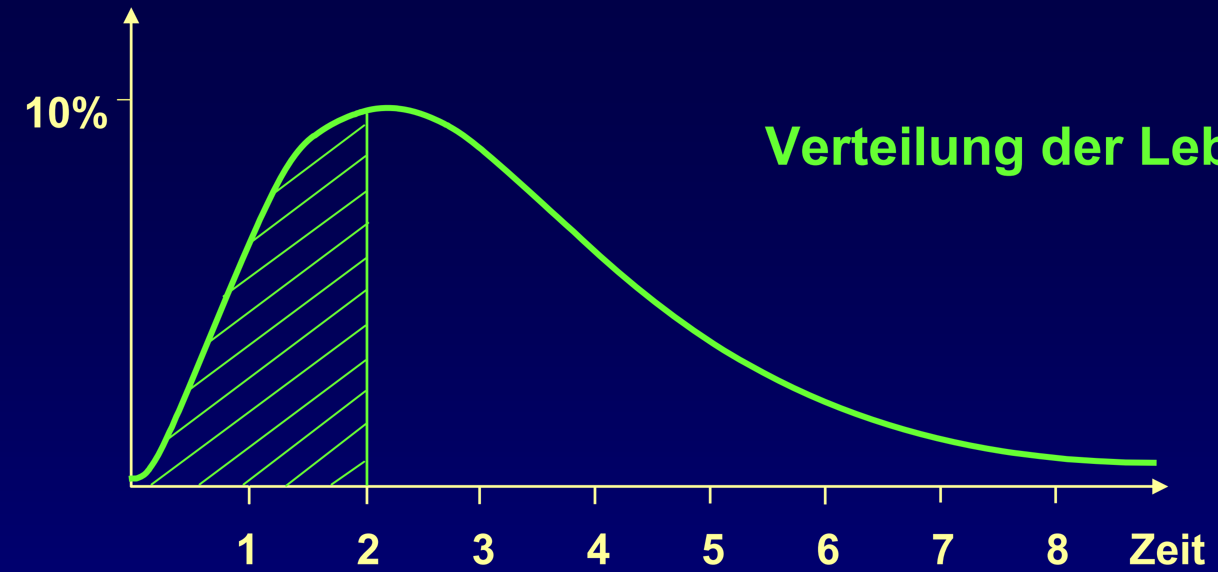


Verweildaueranalyse
n=12

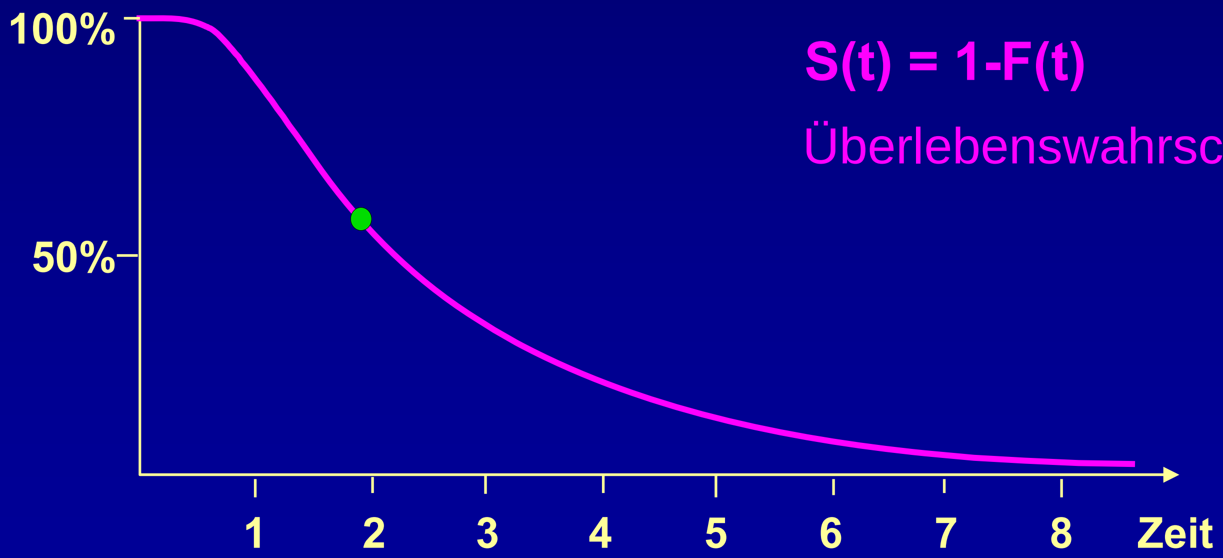


Jahre nach Implantation

d_i		1		2		1		Mißerfolg
u_i	1	1	1		1			lost case
w_i		1			1		2	withdrawal
l_i	11	8	7	5	3	2	0	Erfolg



Verteilung der Lebenszeiten f(t)



S(t) = 1-F(t)

Überlebenswahrscheinlichkeit zum Zeitpunkt t

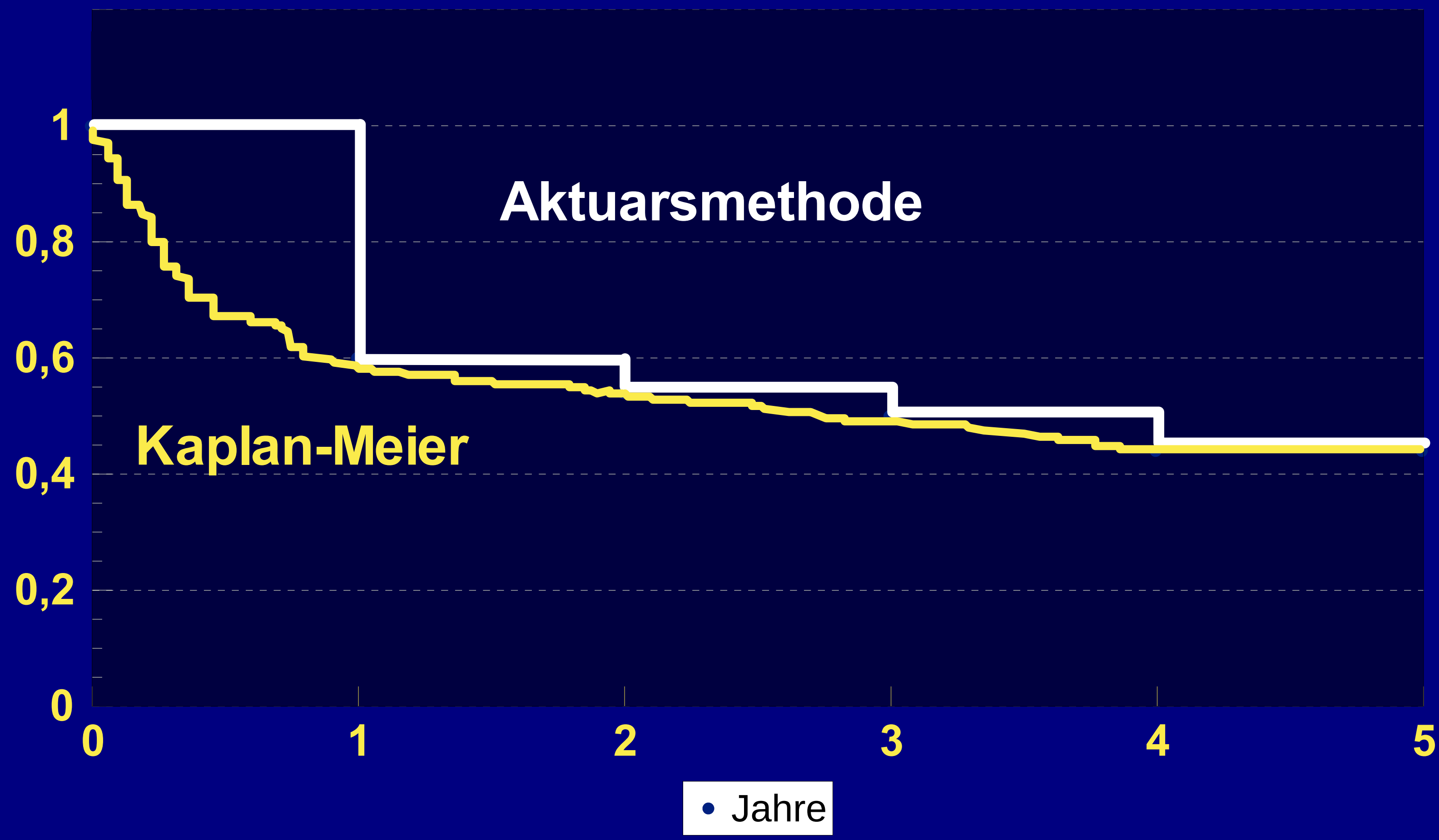
Berechnung der Erfolgsrate

Zeit	Impl. unter Beob.	Miß-erfolg	with-drawal	lost cases	Impl. unter Risiko	Korrig. Mißerfolg HAZARD	Erfolg	kum. Erfolgs-rate
t	l	d	w	u	l'	q	p	S
to - t1	126	47	15	4	116,5	0,4	0,6	0,6
t1 - t2	60	5	11	6	51,5	0,1	0,9	0,54
t2 - t3	38	2	15	0	30,5	0,07	0,93	0,5
t3 - t4	21	2	7	2	16,5	0,12	0,88	0,44
t4 - t5	10	0	6	0	7,0	0,0	1,0	0,44

$$\frac{47}{126} \qquad \frac{47}{116{,}5} \qquad \frac{47}{107}$$

0.373 0,403 0,439

Life-tables - Beispieldaten



Mathematische Grundlagen der Life-tables

Mißerfolgswahrscheinlichkeit = $\frac{\text{Mißerfolge}}{\text{Implantate}}$ — allgemeiner Ansatz

$$q = \frac{d}{l}$$

zu optimistisch: **q zu klein**
(withdrawals und lost in l)

$$q = \frac{d + q\ 1/2\,(w+u)}{l}$$

w und u stehen im Mittel bis zur Intervallmitte unter Beobachtung

.
.
.

mathematische Umformung

$$= \frac{d}{l - 1/2(w+u)}$$

Korr. Wahrsch. (HAZARD)
Mißerfolge bezogen auf die "effektiv Risiko"-Implantate

$$q = \frac{d}{l - (w+u)}$$

zu pessimistisch: **q zu groß**
(withdrawal und lost cases unberücksichtigt)

Die Graphische Darstellung von Überlebensdaten

Die Darstellung von Überlebenswahrscheinlichkeiten erfolgt durch sogenannte „Überlebenskurven" (Die alternative Darstellung als Hazard-funktion wird hier nicht behandelt)

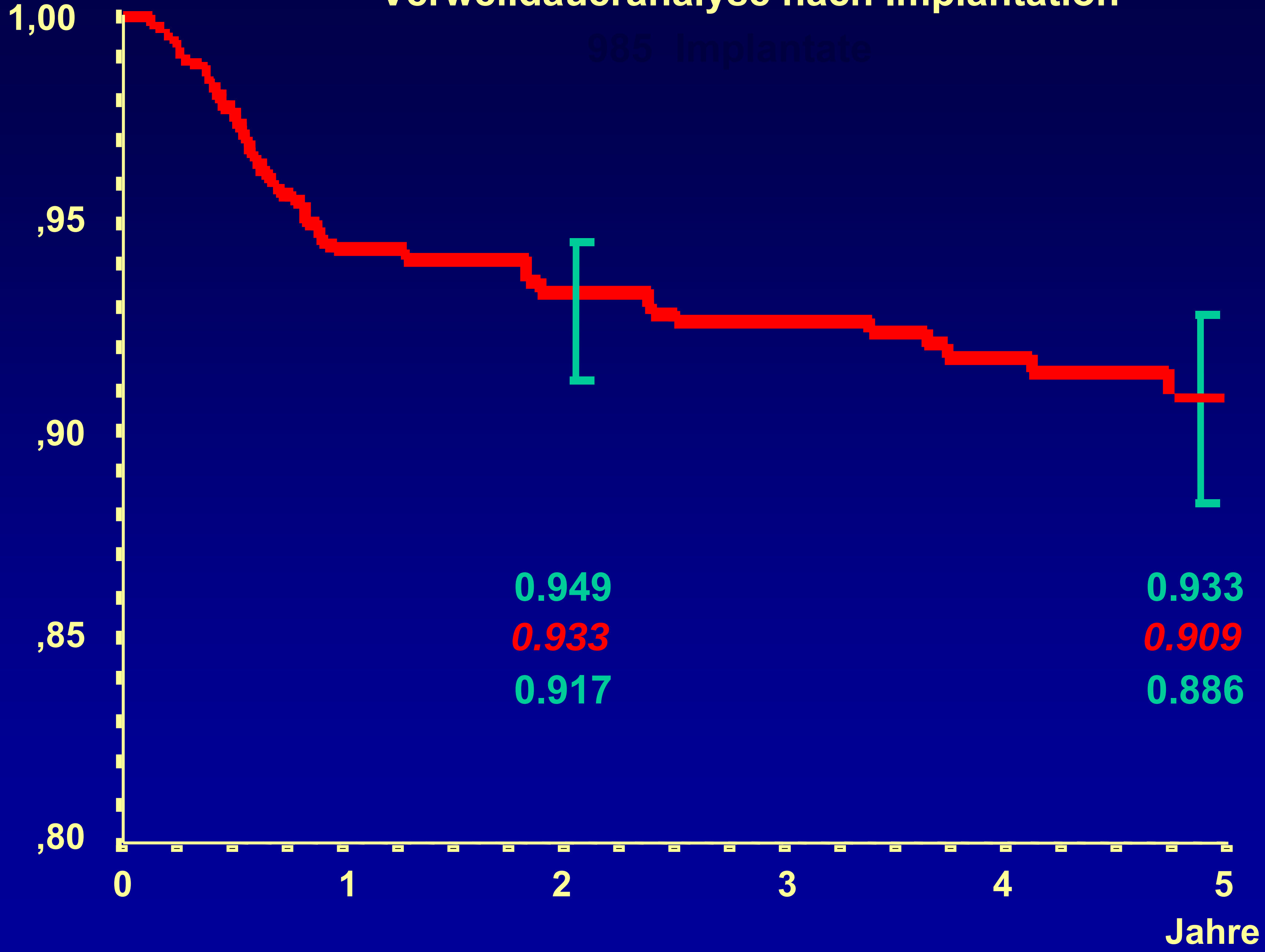
Überlebenskurven sind Treppenfunktionen, bei denen auf der x-Achse die Nachbeobachtungszeit abgetragen wird. Auf der y-Achse wird die geschätzte Gesamtüberlebenswahrscheinlichkeit (vorletzte Spalte der Tabelle) in % oder als Dezimalzahl aufgetragen.

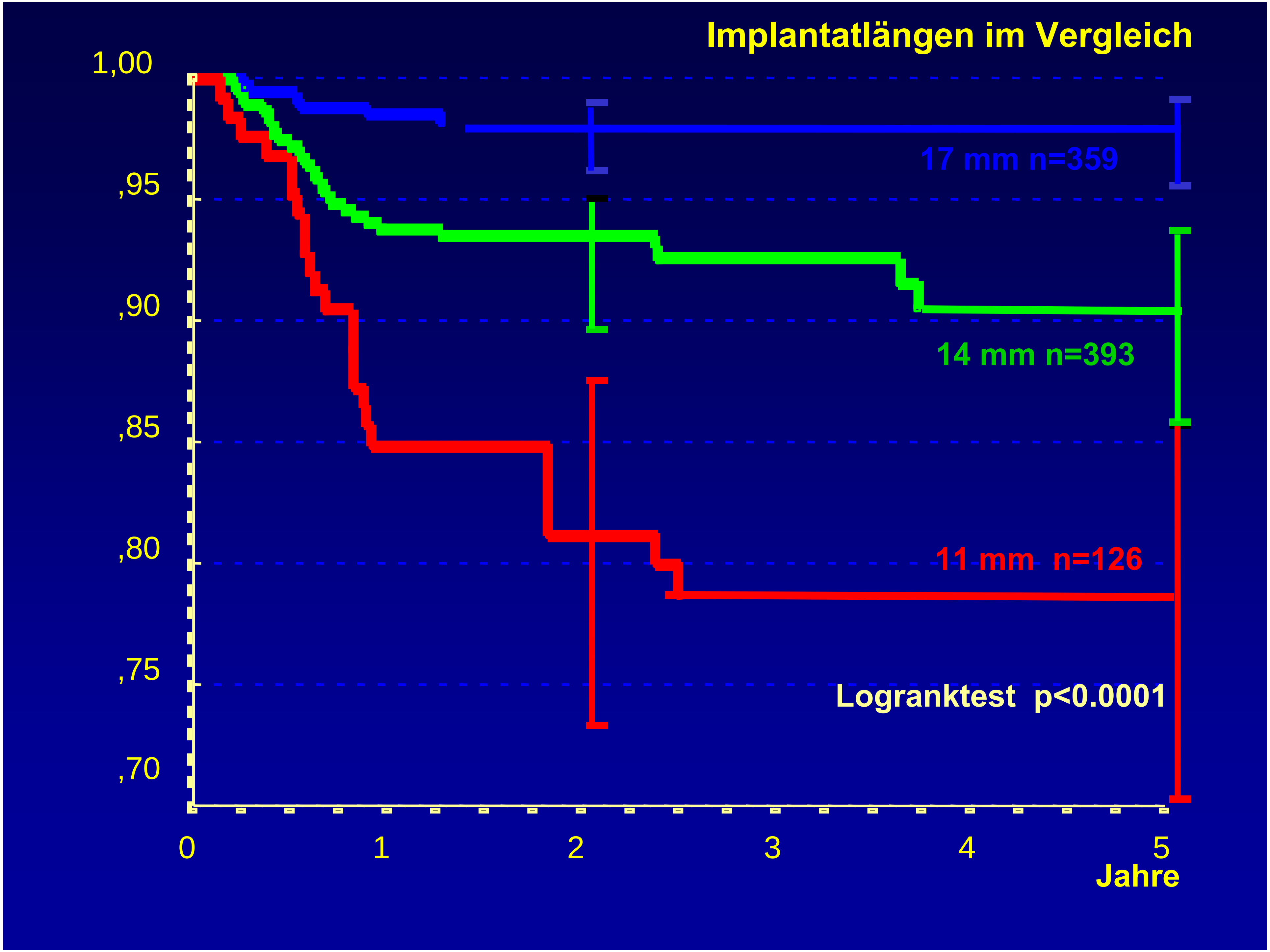
Jeder Todesfall führt zu einer Stufe der Überlebenskurve. An Zeitpunkten mit Zensierungen verläuft die Kurve dagegen waagerecht. Zensierungen haben aber den Effekt, dass sich die nächste Stufe der Überlebenskurve vergrößert.

Die Schätzungen der Überlebenswahrscheinlichkeiten sind für das Ende der Nachbeobachtungszeit sehr ungenau. Man kann dies durch Konfidenz-intervalle für Überlebenswahrscheinlichkeiten quantifizieren (im SPSS aber sehr umständlich in die Überlebenskurve zu integrieren).

Verweildaueranalyse nach Implantation

985 Implantate




Implantatlängen im Vergleich
1,00
,95
,90
,85
,80
,75
,70
0
1
2
3
4
5
Jahre
17 mm n=359
14 mm n=393
11 mm n=126
Logranktest p<0.0001

Der Vergleich von Überlebenskurven mittels Log-Rank Test

Der Log Rank Test ermöglicht den Vergleich von Überlebens-kurven zweier oder mehrerer Populationen ähnlich wie beim χ^2 -Test:

Für jeden Zeitpunkt wird festgehalten, wie viele Patienten in jeder der beiden Gruppen noch leben und wie viele insgesamt verstorben sind .

Dann wird berechnet, wie viele Todesfälle in jeder Gruppe an diesem Tag zu erwarten gewesen wären, wenn sich die Todesfälle anteilsmäßig auf beide Gruppen verteilt hätten.

Diese Zahl wird mit der beobachteten in beiden Gruppen verglichen.

Die Abweichungen werden über alle Zeitpunkte aufsummiert.

Berechnungen zum Log Rank Test

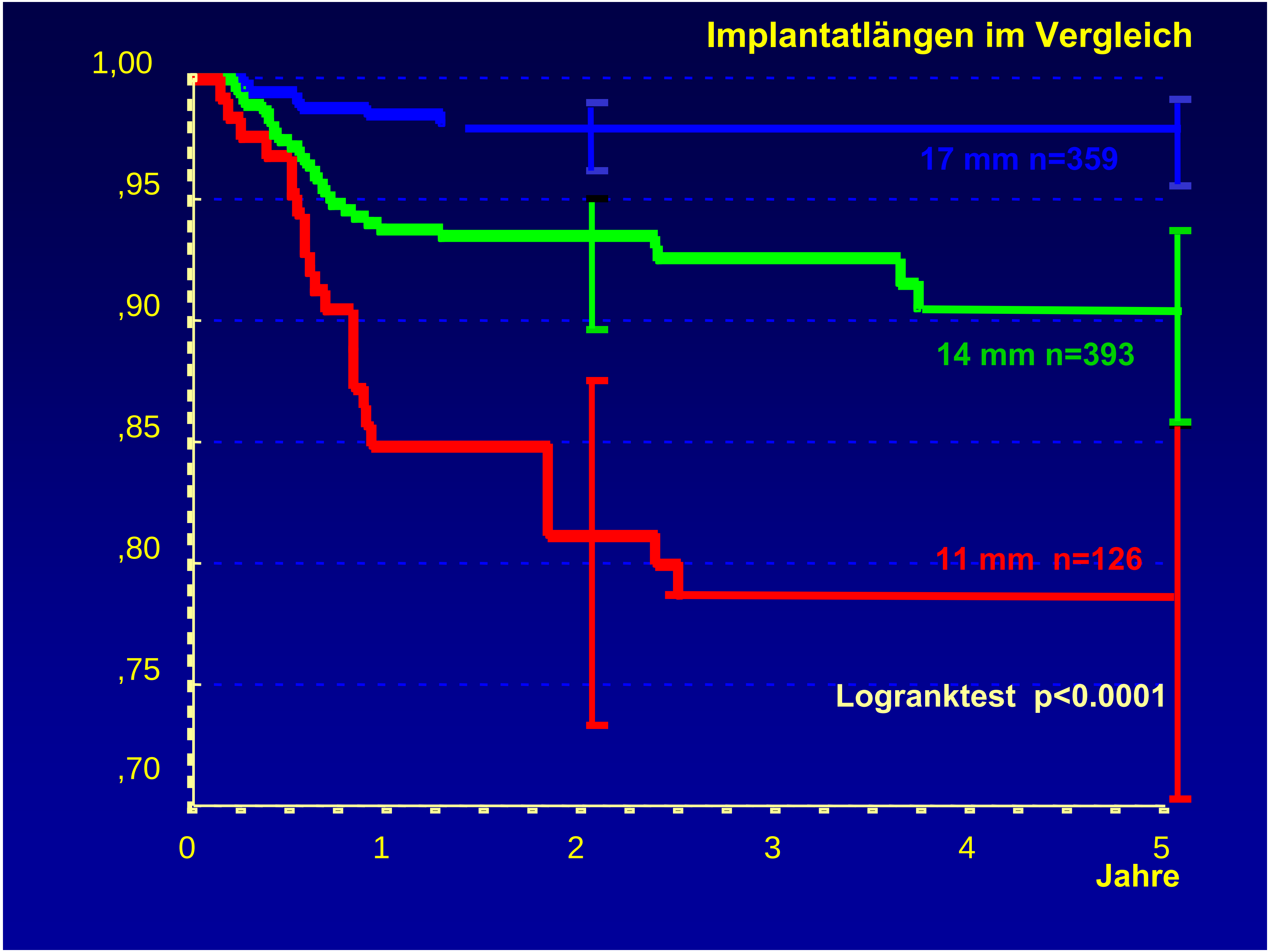
Wie beim χ^2 -Test werden Quotienten aus quadrierten Differenzen und beobachteten Häufigkeiten summiert, die Prüfgröße ist χ^2 verteilt mit einem Freiheitsgrad.

Beim Logrank Test wird dies allerdings nur für die beobachteten und erwarteten Sterbefälle, nicht jedoch für die Überlebenden durchgeführt!

$$\chi^2 = \frac{(h_A - e_A)^2}{e_A} + \frac{(h_B - e_B)^2}{e_B}$$

Das Ergebnis ist signifikant zum 5% Niveau (zweiseitig).

Ein Unterschied zwischen zwei Gruppen A und B wird nachgewiesen, wenn $p < 0.05$



Implantatlängen im Vergleich
1,00
,95
,90
,85
,80
,75
,70
0
1
2
3
4
5
Jahre
17 mm n=359
14 mm n=393
11 mm n=126
Logranktest p<0.0001

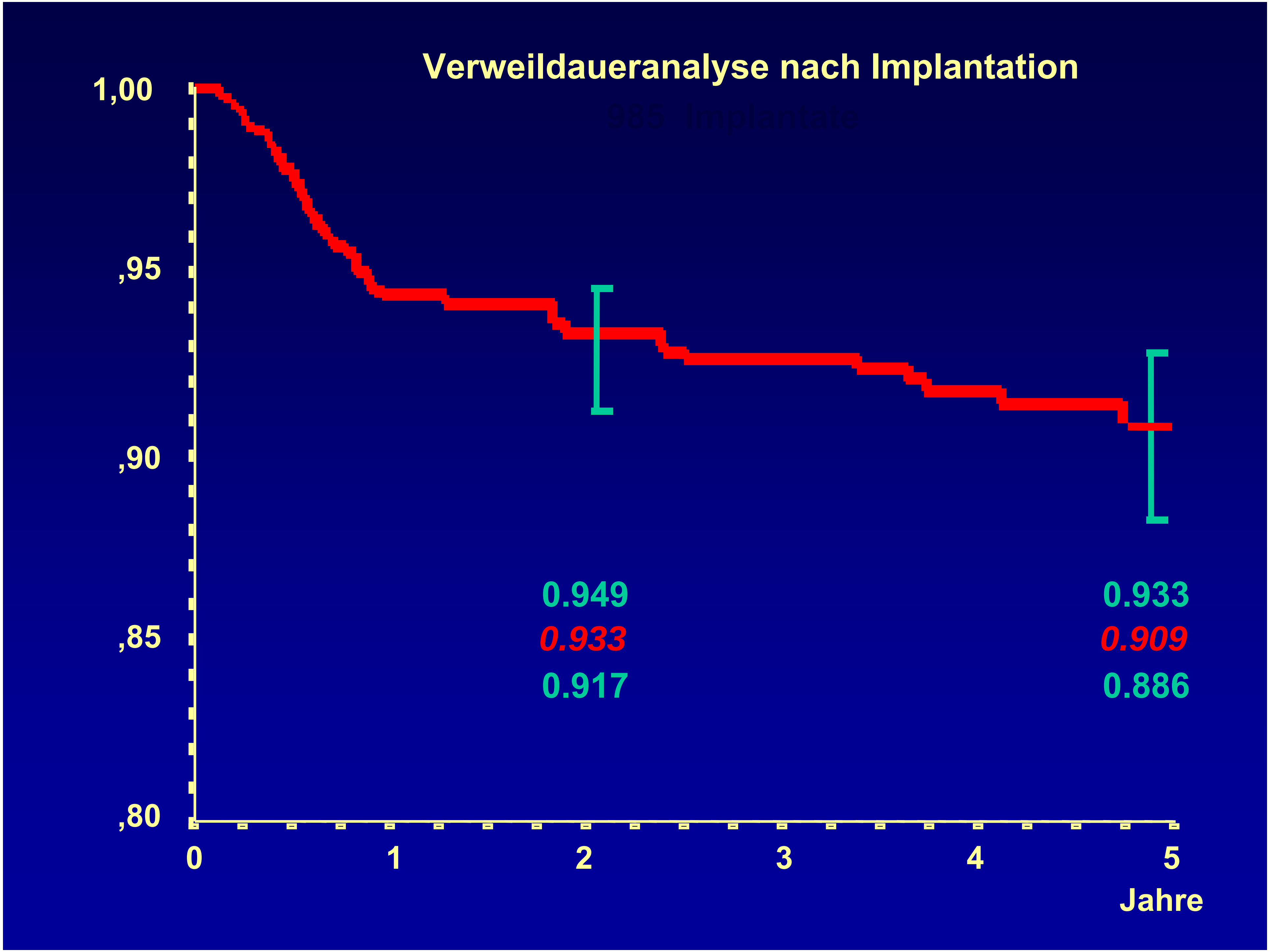
Beispiel für logrank-Test

	Gruppe 1		Gruppe 2		Gesamt			
Zeit	*n1*	*d1*	*n2*	*d2*	*n1 + n2*	*d1 + d2*	*e1*	*e2*
1	30	0	50	1	80	1	0,38	0,63
2	30	0	46	1	76	1	0,39	0,61
1	29	4	41	0	70	4	1,66	2,34
2	25	3	41	2	66	5	1,89	3,11
3	22	2	37	3	59	5	1,86	3,14
4	20	4	33	4	53	8	3,01	4,97
5	15	3	26	3	41	6	2,21	3,81
6	12	2	22	0	34	2	0,71	1,29
7	10	1	18	2	28	3	1,07	1,93
8	6	2	15	0	21	2	0,56	1,43
9	3	1	10	1	13	2	0,46	1,55
10	1	0	9	1	10	1	0,11	0,89
Σ		22		18			14,31	25,7

$$\text{logrank } \chi^2 = \frac{(\Sigma\, d1 - \Sigma\, e1)^2}{\Sigma\, e1} + \frac{(\Sigma\, d2 - \Sigma\, e2)^2}{\Sigma\, e2} = \frac{(22-14{,}31)^2}{14{,}31} + \frac{(18-25{,}7)^2}{25{,}7}$$

$$= 4{,}13 + 2{,}3 = 6{,}43$$

$(\chi^2_{FG=1;\ p=0.05} = 3{,}84)$


Verweildaueranalyse nach Implantation
985 Implantate
1,00
,95
,90
,85
,80
0.949
0.933
0.917
0.933
0.909
0.886
0
1
2
3
4
5
Jahre

Überlebensanalysen in der Medizin

Das entscheidende Problem ist aber, dass i.a. nicht alle Patienten *während* einer Studie das „Zielereignis“ erleben. Es ist zwar evident, dass jeder Patient zu einem bestimmten Zeitpunkt versterben wird. Der Tod des Patienten kann aber nur beobachtet werden, wenn dieser Zeitpunkt vor Ende des Follow Ups liegt.

Für die eine Gruppe von Patienten lässt sich also die Zeit bis zum Zielereignis angeben („events“). Für die restlichen Patienten lässt sich nur eine Mindestzeit angeben, in der das Zielereignis nicht eingetreten ist („zensierte Fälle“). Die Nachbeobachtungszeiten der zensierten Fälle unterscheiden sich zum einen aufgrund der unterschiedlichen Rekrutierungszeitpunkte, aber auch dadurch, dass manche Patienten frühzeitig aus der Studie ausscheiden (lost to follow up).

Die Kaplan Meier Methode (s.u.) basiert auf der Annahme, dass Zensierungen am Studienende und Zensierungen durch lost to follow up als gleichbedeutend angesehen werden und unabhängig von der Prognose der Patienten sind.

Analyse von Implantatsystemen "Ereignisse"

- Implantatverlust

- keine absolute Immobilität
- erreichen eines Lockerungsgrads
- Osteolysen
- vertikaler Knochenverlust > 0.2 mm/Jahr
- Knochenabbauvorgänge
- Schmerzen
- Infektionen
- Neuropathien
- andere klinische Kriterien

- Funktionstüchtigkeit des Implantats
- Funktionstüchtigkeit der Suprastruktur

- Komfort der Behandlung
- Ästhetik der Behandlung
- subjektive Faktoren
- prothetische Faktoren

Biomathematik

Alte Klausuraufgaben

Bei Patienten wird das Merkmal Glucosekonzentration im Blut gemessen.

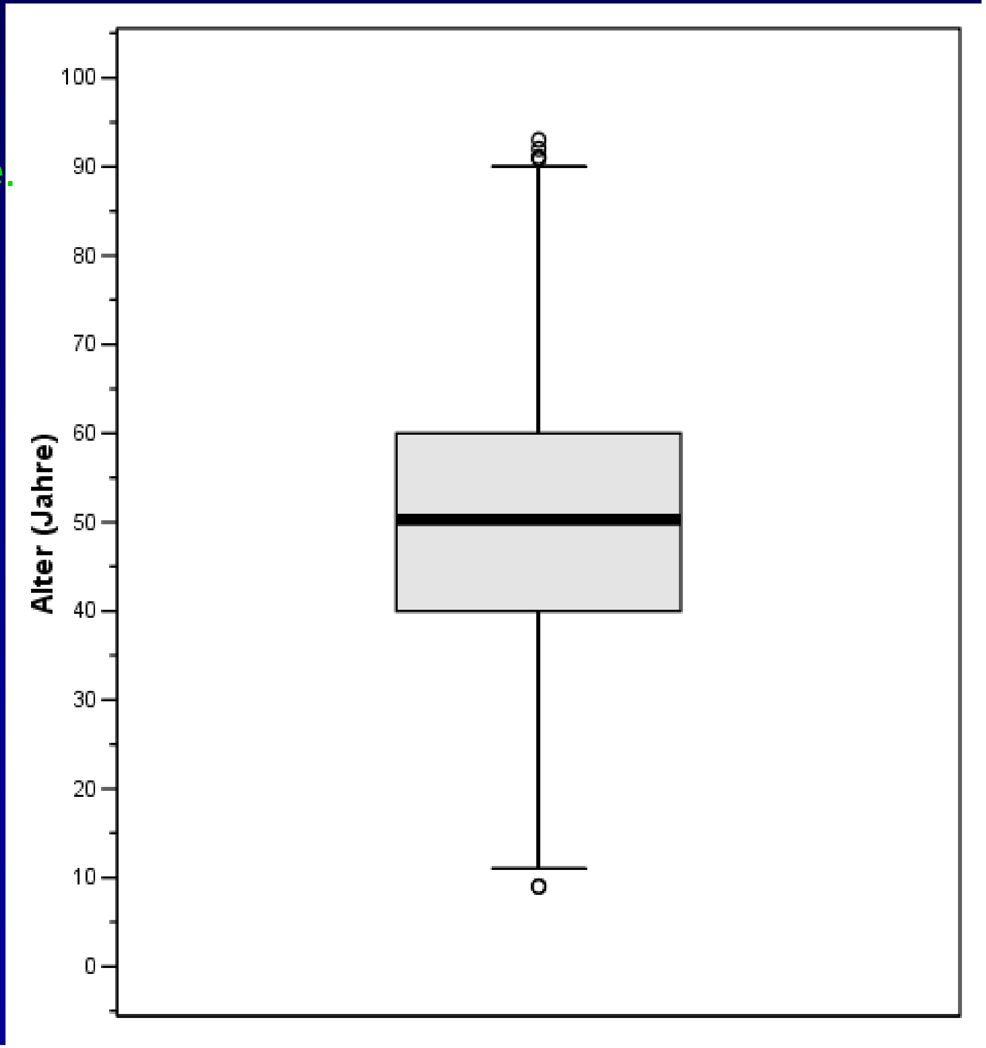
Dieses Merkmal ist

- A nominal
- B ordinal
- C dichotom
- D stetig
- E kategoriell

Innerhalb einer Studie mit ca. 6000 pathologischen Präparaten von Schilddrüsenpatienten wurde folgende Abbildung erstellt:

Welche Aussage ist richtig ?

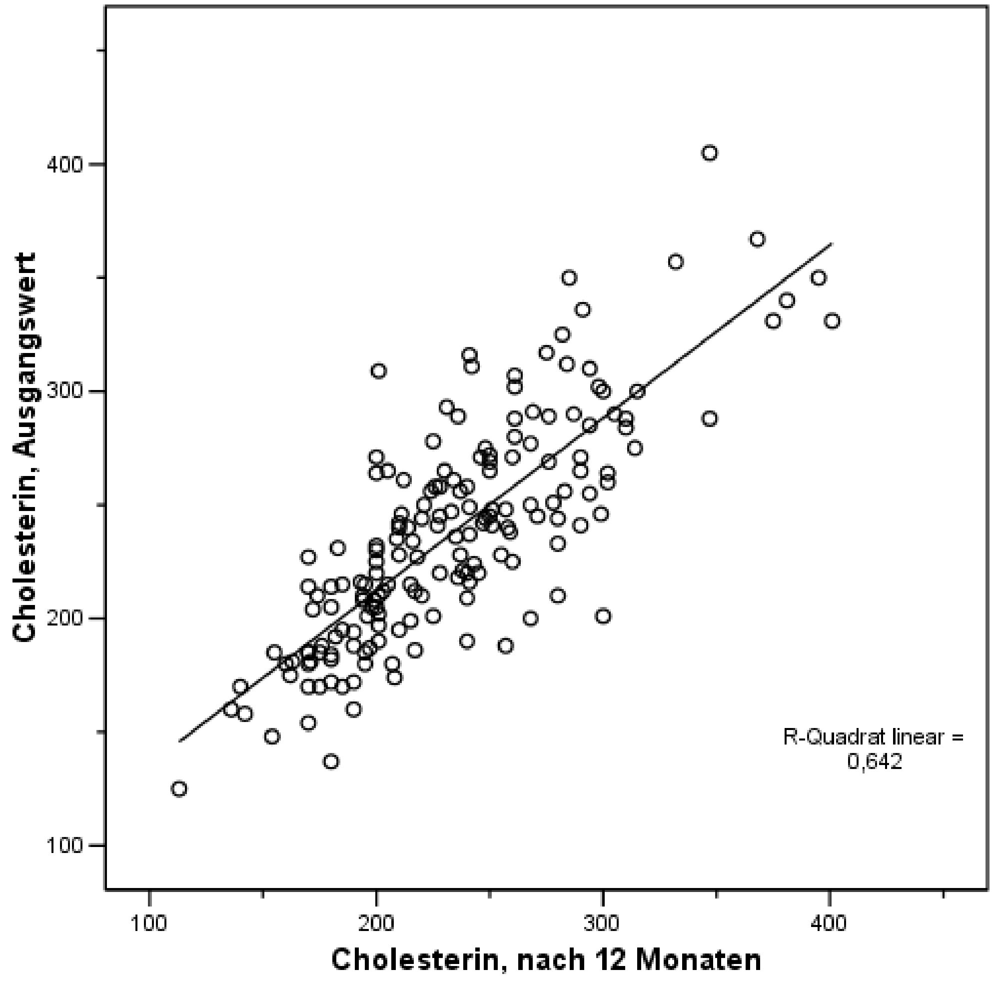
A Die waagrechte Linie in der Box ist der arithmetische Mittelwert
B Das 50% Quantil liegt bei 50 Jahren.
C Die Standardabweichung ist 40 Jahre.
D Das erste Quartil liegt bei 60 Jahren.
E 75% der Patienten sind jünger als 40 Jahre.



Im Rahmen einer Studie mit 174 Hochdruckpatienten wurde u.a. das Gesamtcholesterin in mg/dl im Serum bestimmt.
In der folgenden Abbildung sind die Cholesterinwerte zu Beginn der Studie (Ausgangswert) und nach 12 Monaten gegenübergestellt.

Der Korrelationskoeffizient r nach Pearson ist ein Maß für die lineare Beziehung zwischen zwei quantitativen Merkmalen.
Geben Sie für die Grafik einen Wert für r an (Genauigkeit 2 Kommastellen).

A	r = 0,36
B	r = 0,41
C	r = 0,64
D	r = 0,80
E	r kann nicht bestimmt werden



Um für die obige Punktwolke die Gleichung der Regressionsgerade zu bestimmen, wird mit Hilfe von SPSS eine lineare Regressionsanalyse durchgeführt.
U.a. erscheint im SPSS-Ausgabefenster folgende Tabelle:

Koeffizienten[a]

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Signifikanz
		B	Standard fehler	Beta		
1	(Konstante)	33,083	11,619		2,847	,005
	chol.0 (Cholesterin, Ausgangswert)	,843	,048	,802	17,581	,000

a. Abhängige Variable: chol.12 (Cholesterin, nach 12 Monaten)

Welche Aussage ist richtig?

A die Steigung der Regressionsgerade ist 33,083

B die Regressionsgleichung ist: chol.12 = 33,083 + 0,843 · chol.0

C die Regressionsgleichung ist: chol.0 = 33,083 + 0,843 · chol.12

D für das Bestimmtheitsmaß ergibt sich $r^2 = 0.843$

E die Steigung der Regressionsgerade ist nicht im Ausdruck enthalten

In der gleichen Studie wird für das Cholesterin zwischen Ausgangswert und dem Wert nach 6 Monaten ein Korrelationskoeffizient von 0.75 gefunden.

Ein amerikanischer Wissenschaftler transformiert die in mg/dl gemessenen Originaldaten in mmol/l
(Umrechnungsfaktor: mg/dl · 0,02586 = mmol/l).

Welchen Wert hat der Korrelationskoeffizient jetzt?

A 0,02586 / 0,75

B 0,75 / 0,02586

C 0,75 * 0,02586

D 0,75

E r kann aus den Angaben nicht errechnet werden.

Zusätzlich wird in der Studie der Erfolg der Hochdruckbehandlung nach einem bekannten Responderkriterium festgelegt:
Bei Senkung des diastolischen Blutdruck um mindestens 10 mmHg gilt der Patient als ‚Responder', andernfalls als ‚Nonresponder'.

Um zu prüfen, ob sich die Reponderraten zwischen den Zeitpunkten 6 und 12 Monaten statistisch signifikant unterscheiden, wird ein McNemar-Test durchgeführt.
Mit Hilfe von SPSS erhält man dabei folgenden Ausdruck:

Responder 6 Monate * Responder 12 Monate Kreuztabelle

Anzahl

		Responder 12 Monate		Gesamt
		nein	ja	
Responder 6 Monate	nein	7	33	40
	ja	7	127	134
Gesamt		14	160	174

Chi-Quadrat-Tests

	Wert	Exakte Signifikanz (2-seitig)
McNemar-Test		,000[a]
Anzahl der gültigen Fälle	174	

a. Verwendete Binomialverteilung

Welche Interpretation des Testergebnisses ist richtig?

A A Der Test erkennt keine signifikante Veränderung der Responderrate zwischen dem 6. und 12. Monat.

B B Die Responderrate hat sich zwischen 6. und 12. Monat statistisch signifikant verschlechtert (p<0.001)

C C Die Responderrate hat sich zwischen 6. und 12. Monat statistisch signifikant verbessert (p<0.001)

D D Der McNemar-Test ist nicht sinnvoll, da die angegebenen Häufigkeiten in Zeilen und Spalten unabhängig sind.

Welche Aussage über Kontingenztafeln trifft <u>nicht</u> zu?

A In Kontingenztafeln werden Häufigkeiten eingetragen

B Die Vierfeldertafel ist ein Spezialfall der Kontingenztafel

C Beobachtungen von zwei kategoriellen Merkmalen können in einer Kontingenztafel dargestellt werden

D Zur Prüfung der Unabhängigkeit der Merkmale in einer Kontingenztafel, benutzt man den χ^2 - Test

E Die Summe der Spaltenprozente in jeder Zeile ist immer 100%

Medikament A steht im Verdacht Leberschäden zu verursachen.
Innerhalb einer klinischen Studie wurden 20 Patienten mit der Therapie A und 20 Patienten mit der Standardtherapie B nach randomisiert unter Dopppelblindbedingungen behandelt.

Gemessen wurde das Ferment γ-GT im Serum.
Die Verteilung dieser Messgröße in der Grundgesamtheit ist nicht normalverteilt.

Welcher statistische Test ist für den Therapievergleich bzgl. γ-GT geeignet?

A t - Test für paarige Stichproben

B Test auf lineare Regression

C Mann Whitney – Test (für unverbundene Stichproben)

D Wilcoxon – Test (für paarige Stichproben)

E χ^2 - Test

Der t-Test für unabhängige Stichproben und der χ^2 - Vierfeldertest zum Vergleich zweier Stichproben setzen <u>beide</u> voraus:

A normalverteilte Beobachtungen
B qualitative Beobachtungen
C bekannte Varianzen
D Verteilungsunabhängigkeit
E unpaarige Beobachtungen

Im Rahmen einer Untersuchung zur Prophylaxe von Mamma-Carcinomen wurde folgende Kontingenztafel erstellt:

			Mamma-Carcinom liegt vor		
			ja	nein	Gesamt
Mammographie Screening	positiv	Anzahl	81	41	122
		% von Zeile	66,4%	33,6%	100,0%
		% von Spalte	81,0%	4,1%	11,1%
	negativ	Anzahl	19	959	978
		% von Zeile	1,9%	98,1%	100,0%
		% von Spalte	19,0%	95,9%	88,9%
Gesamt		Anzahl	100	1000	1100
		% von Zeile	9,1%	90,9%	100,0%
		% von Spalte	100,0%	100,0%	100,0%

In der Kontingenztafel finden Sie als relative Häufigkeit den Wert 95,9 % . Dieser Wert steht für

A die Spezifität

B die Sensitivität

C den positiven prädiktiven Wert

D den negativen prädiktiven Wert

E die Prävalenz

Im Rahmen einer Untersuchung zur Prophylaxe von Mamma-Carcinomen wurde folgende Kontingenztafel erstellt:

			Mamma-Carcinom liegt vor		
			ja	nein	Gesamt
Mammographie Screening	positiv	Anzahl	81	41	122
		% von Zeile	66,4%	33,6%	100,0%
		% von Spalte	81,0%	4,1%	11,1%
	negativ	Anzahl	19	959	978
		% von Zeile	1,9%	98,1%	100,0%
		% von Spalte	19,0%	95,9%	88,9%
Gesamt		Anzahl	100	1000	1100
		% von Zeile	9,1%	90,9%	100,0%
		% von Spalte	100,0%	100,0%	100,0%

In der Kontingenztafel finden Sie als relative Häufigkeit den Wert 66,4 % . Dieser Wert steht für

A die Spezifität

B die Sensitivität

C den positiven prädiktiven Wert

D den negativen prädiktiven Wert

E die Prävalenz

Welche der folgenden Kenngrößen eines Mammographie-Screening Tests hängt <u>nicht</u> von der Prävalenz in der Population der Frauen ab ?

A der Anteil der positiven Testergebnisse an der Gesamtzahl.

B der negative prädiktive Wert

C der positive prädiktive Wert

D die Spezifität des Mammographie-Tests

E die jährliche Inzidenzrate der Mammakarzinome bei 100.000 Frauen

Man will mit einem statistischen Test zweiseitig prüfen, ob sich ein Medikament A in seiner Wirkung von Placebo unterscheidet.

Was ist die Nullhypothese in diesem Test?

- **A** Die Hypothese, die widerlegt werden soll.
- **B** A ist besser als Placebo.
- **C** Die Größe des sog. Placeboeffekts.
- **D** A und Placebo haben unterschiedliche Wirkung.
- **E** Die Hypothese, die bewiesen werden soll.

Man will mit einem statistischen Test zweiseitig prüfen,
ob sich ein Medikament A in seiner Wirkung von Placebo unterscheidet.

Was ist die Alternativhypothese in diesem Test?

A	Die Hypothese, die widerlegt werden soll.
B	A ist schlechter als Placebo.
C	Die Größe des sog. Placeboeffekts.
D	A und Placebo haben keine unterschiedliche Wirkung.
E	Die Hypothese, die statistisch nachgewiesen werden soll.

Die Prävalenz eines Mammakarzinoms bei einer 40-jährigen Frau betrage 2 pro 1000.

Die erste ‚Vorsorgeuntersuchung' mittels Tastbefund habe eine Sensitivität von 50% und eine Spezifität von 90%.

Etwa wie viele Testpositive müssen durchschnittlich zum definitiven Nachweis bzw. Ausschluss eines Karzinoms mammographiert werden, um bei 40-jährigen Frauen mittels Früherkennungsuntersuchung ein Mammakarzinom zu finden ?

A	**2**
B	**4**
C	**10**
D	**50**
E	**100**

	Ca	Kein Ca	
+	1	100	101
-	1	900	901
	2	1000	1002

Methoden der Verblindung

offen:	Keine Verblindung
einfach blind:	Patient
	Arzt (Therapie und/oder Befunderhebung)
doppelt blind:	Patient und Arzt (behandelnd/bewertend)

Quellen und Methoden zur Vermeidung von Bias

Strukturungleichheit bzgl. unbekannter Variablen:

Randomisierung

Strukturungleichheit bzgl. bekannter Variablen:

Stratifikation und Poststratifikation

Adjustierung(Kovarianzanalyse, logistische oder Cox-Regression)

Beobachtungsungleichheit

Verblindung, Standardisierung, Laborringversuche

Die Methoden zur Verhinderung von Strukturungleichheit dienen auch der Verringerung der Fehlervarianz

Fall-Kontroll-Studie
Case control study

- Design — retrospektiv, vergleichend

- Prinzip — erkrankte Personen
  Bildung einer geeigneten Kontrollgruppe (matched pairs)

- Ziel — Klärung ätiologischer Fragestellungen

- ! — Erhebungsschwierigkeiten
  Selektion in der Stichprobe der Erkrankten

- Auswertung — explorative Datenanalyse
  relatives Risiko, odd‘s ratio ...

Kohortenstudie
Anwendungsbeobachtung
follow-up study

- **Design** prospektiv
- **Prinzip** exponierte Personen
  Folgebeobachtungen
- **Ziel** Erfassung von unerwünschten Ereignissen (Nebenwirkungen)
  bei Arzneimitteln, Genußmitteln, ...
- **!** Zeitaufwand
  drop-out Probleme
  unterschiedliche Beobachtungsintensität in den Kohorten
- **Auswertung** explorative Datenanalyse
  Erkrankungsinzidenzen
  Überlebenraten (survival, life-table)
  relatives Risiko, odd's ratio ...