

UNIVERSITÄT LEIPZIG

Medizinische Fakultät

QSB1
Epidemiologie, Medizinische Biometrie und Medizinische Informatik

Lösungen

der Aufgaben in den Vorlesungen

imise.

V 1, Folien 42 - 44

- Diast-CHF-Studie:

  prospektive, nicht kontrollierte Beobachtungsstudie im Längsschnitt

- Schülersportstudie:

  prospektive, randomisierte, kontrollierte Interventionsstudie im Längsschnitt

- Endokarditis-Studie:

  retrospektive, kontrollierte, Interventionsstudie im Längsschnitt

V 2 Folie 11: Stichprobe

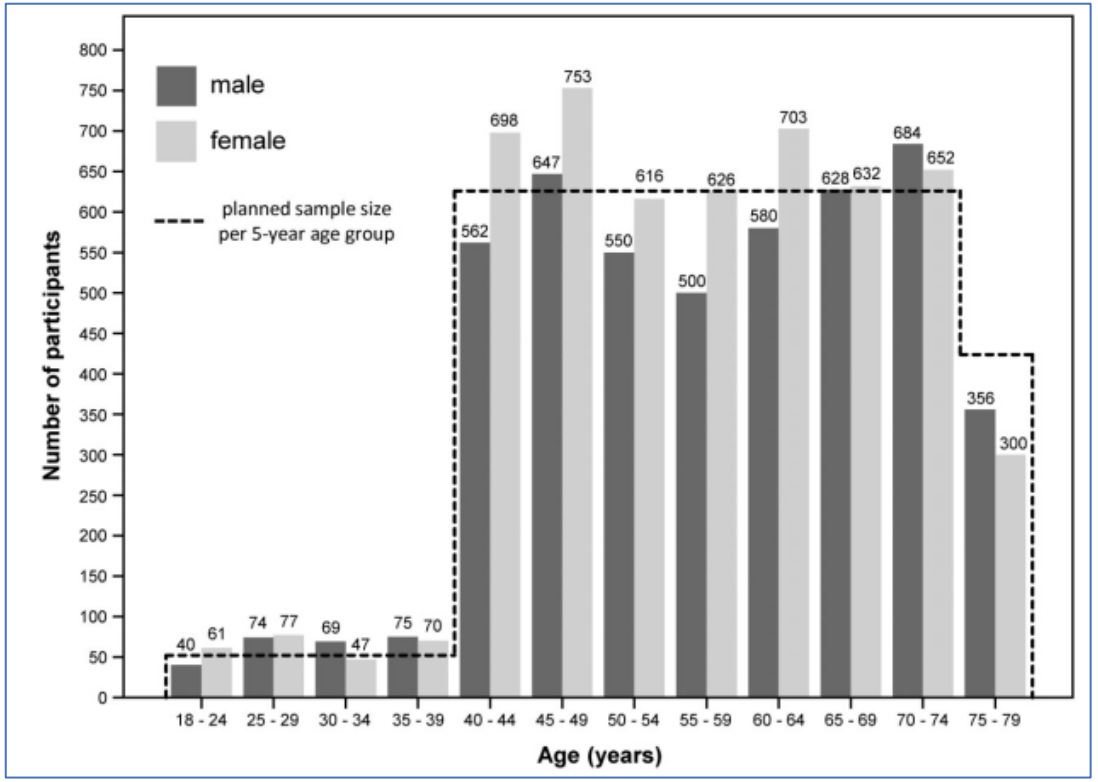
Da es i. d. R. zu aufwändig ist, alle Probanden der GGS zu messen, erfasst man eine (zufällige) Auswahl, eine Stichprobe.

Szenario: Bundestagswahl
Was sind Grundgesamtheit:

- alle Wahlberechtigten

und evtl. Stichprobe

- z. B. Befragte bei Wahl-umfragen (Sonntagsfrage)



V 2, Folie 27 Übung

Welches Skalenniveau (welchen Merkmalstyp) haben die folgenden Variablen der Ex-DHF-Studie:

- Body-Maß-Index (BMI) — kontinuierlich / quantitativ metrisch
- Raucher (ja / nein) — binär
- Wie viele Monate Diagnose Herzinsuffizienz? — kontinuierlich
- LV-EF — kontinuierlich
- Anzahl Geschwister — quantitativ diskret
- NYHA-Klasse (I, II, III, IV) — ordinal
- Endpunkt nach 1 Jahr (worsened – unchanged – improved) — ordinal

V 3, Folie 22: Median III

Median bei ungerader Anzahl von Beobachtungen

z. B.: Ex-DHF-Studie, NT-proBNP baseline, n = 11

189, 51, 141, 36, 173, 130, 95, 26, 41, 103, 43 ng / ml

Schritt 1: Beobachtungen ordnen:

26, 36, 41, 43, 51, 95, 103, 130, 141, 173, 189

1. 2. 3. 4. 5. 6. 7. 8. 9. 10. 11.

Schritt 2: Mittlere Zahl bestimmen: 6. Beobachtung

Median: 95 ng / ml

V 3, Folie 23 Median IV

Median bei gerader Anzahl von Beobachtungen

z. B.: Ex-DHF-Studie, NT-proBNP baseline, n = 10

189, 51, 141, 36, 173, 130, 95, 26, 41, 103 ng / ml

Schritt 1: Beobachtungen ordnen:

26, 36, 41, 51, 95, 103, 130, 141, 173, 189

1. 2. 3. 4. 5. 6. 7. 8. 9. 10.

Schritt 2: Zwei mittlere Zahlen bestimmen: 5. und 6. Beobachtung,

Schritt 3: Daraus der Mittelwert

Median: 99 ng / ml

V 3, Folie 25 Quartile II

z. B.: Ex-DHF, NT-proBNP baseline, n = 12

1. Ordnen (Rangplatz zuordnen)

26	36	41	43	51	67	95	103	130	141	173	189
$X_{(1)}$	$X_{(2)}$	$X_{(3)}$	$X_{(4)}$	$X_{(5)}$	$X_{(6)}$	$X_{(7)}$	$X_{(8)}$	$X_{(9)}$	$X_{(10)}$	$X_{(11)}$	$X_{(12)}$

2. Für Q1: $12 \cdot 0{,}25 = 3$ → Mittel aus Wert des 3. und 4. Elementes

$$Q1 = {}^{(41+43)}/_{2} = 42\,\frac{ng}{ml}$$

3. Für Q3: $12 \cdot 0{,}75 = 9$ → Mittel aus Wert des 9. und 10. Elementes

$$Q3 = {}^{(130+141)}/_{2} = 135{,}5\,\frac{ng}{ml}$$

CAVE! Statistische Programme (SAS, SPSS, R, ...)
nutzen verschiedene Berechnungsformeln



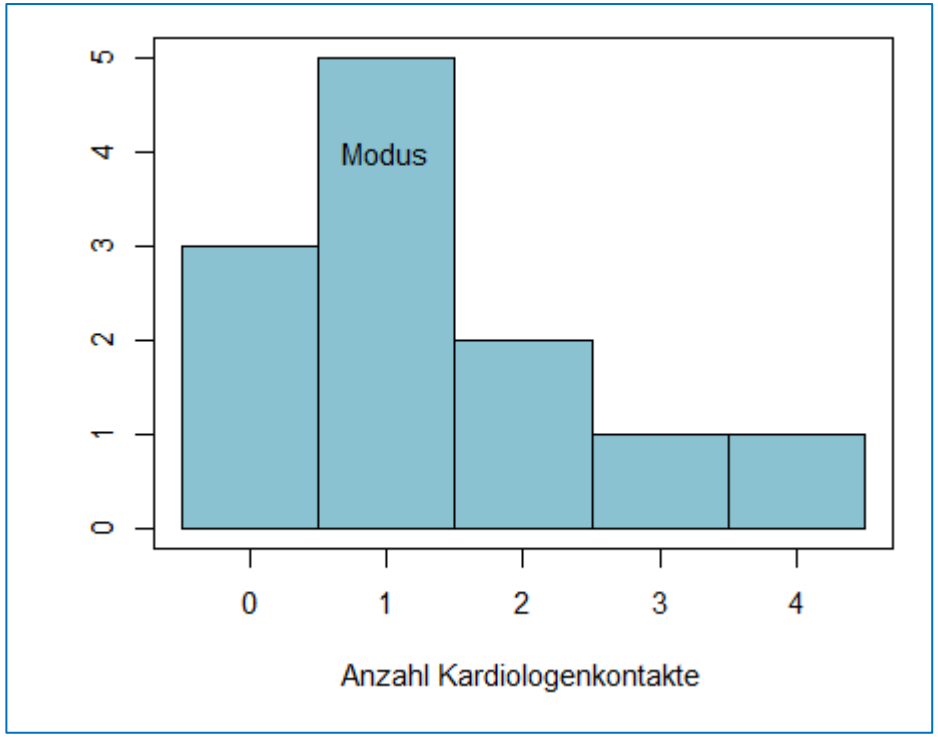
Jotquadrat: Grafitto in Bonn. commons. wikipedia.org

V 3, Folie 28 Lagemaß II: Modalwert

Der Modalwert (syn.: Modus) ist der Wert, welcher in der Datenerhebung eines Merkmales am häufigsten vorkommt.

- Beispiel 3.3: Ex-DHF, Anzahl der Kardiologen-kontakte im letzten Jahr:

- Modalwert = 1



V 3, Folie 29 3.2.2 Parametrische Kenngrößen

Wurde ein Merkmal X mit den Beobachtungen x_i $(i = 1, \dots, n)$ gemessen, so ist der

- arithmetische Mittelwert: $\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n} \sum_{i=1}^{n} x_i$

Beispiel 3.2 DIAST-CHF, NT-proBNP, n = 11

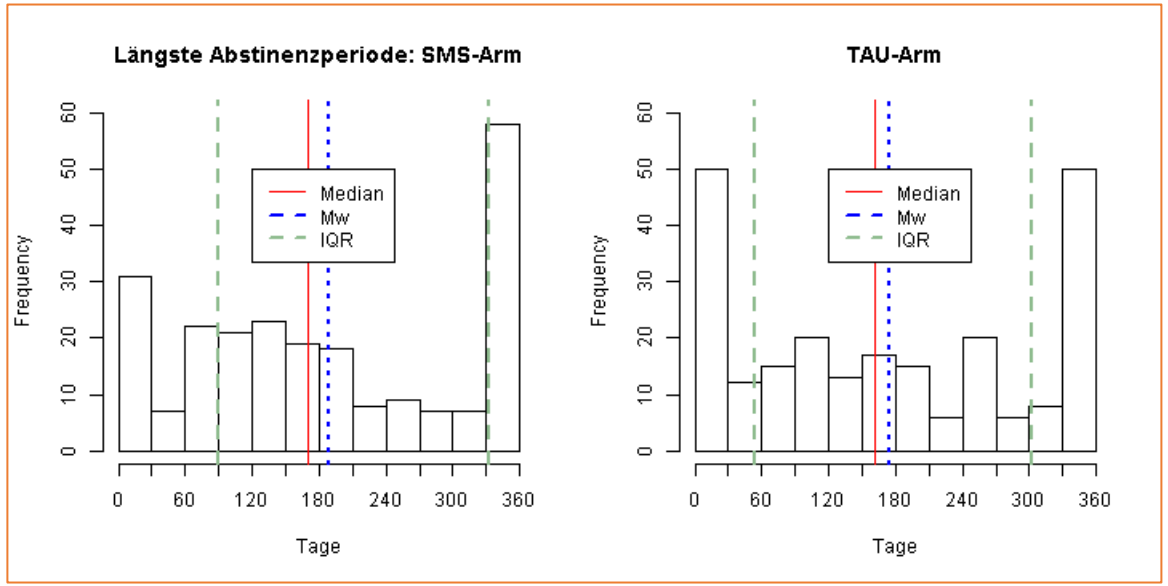
189, 51, 141, 36, 173, 130, 95, 26, 41, 103, 43 ng / ml

$$\bar{x} = \frac{189+51+141+36+173+130+95+26+41+103+43}{11} = 93{,}5 \ \frac{ng}{ml}$$

V 3, Folie 48 Beispiel Histogramm: CAPS-Studie

Lucht M et al. Addiction 2021;116(6):1431-1442

- Treatment as usual (TAU)
- SMS-Intervention



Längste Abstinenzperiode:

Weil sich Werte an den Rändern häufen, sind weder Mittelwert noch Median repräsentativ.

Welcher Studienarm erscheint besser?

der SMS-Arm, dort gibt es

- mehr Probanden mit 360 Abstinenztagen und
- deutlich weniger Probanden mit 0 Abstinenztagen

V 4, Folie 13 Ereignis

Ein Ereignis ist eine Menge von Ergebnissen.

z. B.: Ereignis = {Augensumme gleich 4 bei einem Wurf mit zwei Würfeln}

wird erzielt in folgenden Fällen: {(1, 3), (2, 2), (3, 1)}

Ein Ereignis, welches nur ein mögliches Ergebnis enthält, nennt man Elementarereignis.

Die Menge aller möglichen Ergebnisse wird als
sicheres Ereignis S bezeichnet.

Die Menge, die kein mögliches Ergebnis enthält, wird als
unmögliches Ereignis bezeichnet.

V 4, Folie 18 Klassische Definition der Wahrscheinlichkeit

Für ein beliebiges Ereignis A ist die Wahrscheinlichkeit (Laplace)

$$P(A) = \frac{\textit{Anzahl der für A günstigen Ergebnisse}}{\textit{Anzahl der möglichen Ergebnisse}}$$

Voraussetzung: Alle möglichen Ergebnisse austauschbar (gleichberechtigt) (Elementarereignisse).

Zum Beispiel: Wahrscheinlichkeit für gerade Augenzahl bei 1 x Würfeln:

- $P(gerade\ Augenzahl) = \frac{3}{6} = 0{,}5$

V 4, Folie 33 Subtraktionsregel

$$P(A) = 1 - P(\bar{A})$$

wird angewandt, wenn $P(\bar{A}) = P(nicht\ A)$ einfacher zu berechnen ist als $P(A)$.

Z. B.:

Wie hoch ist die Wahrscheinlichkeit des Ereignisses A, mit zwei Würfeln zwei unterschiedliche Augenzahlen zu würfeln?

$$P(A) = 1 - P(nicht\ A)$$
$$= 1 - P(\text{Pasch würfeln})$$
$$= 1 - 1 \cdot \frac{6}{36} = \frac{5}{6} = 0{,}833 = 83{,}3\%$$

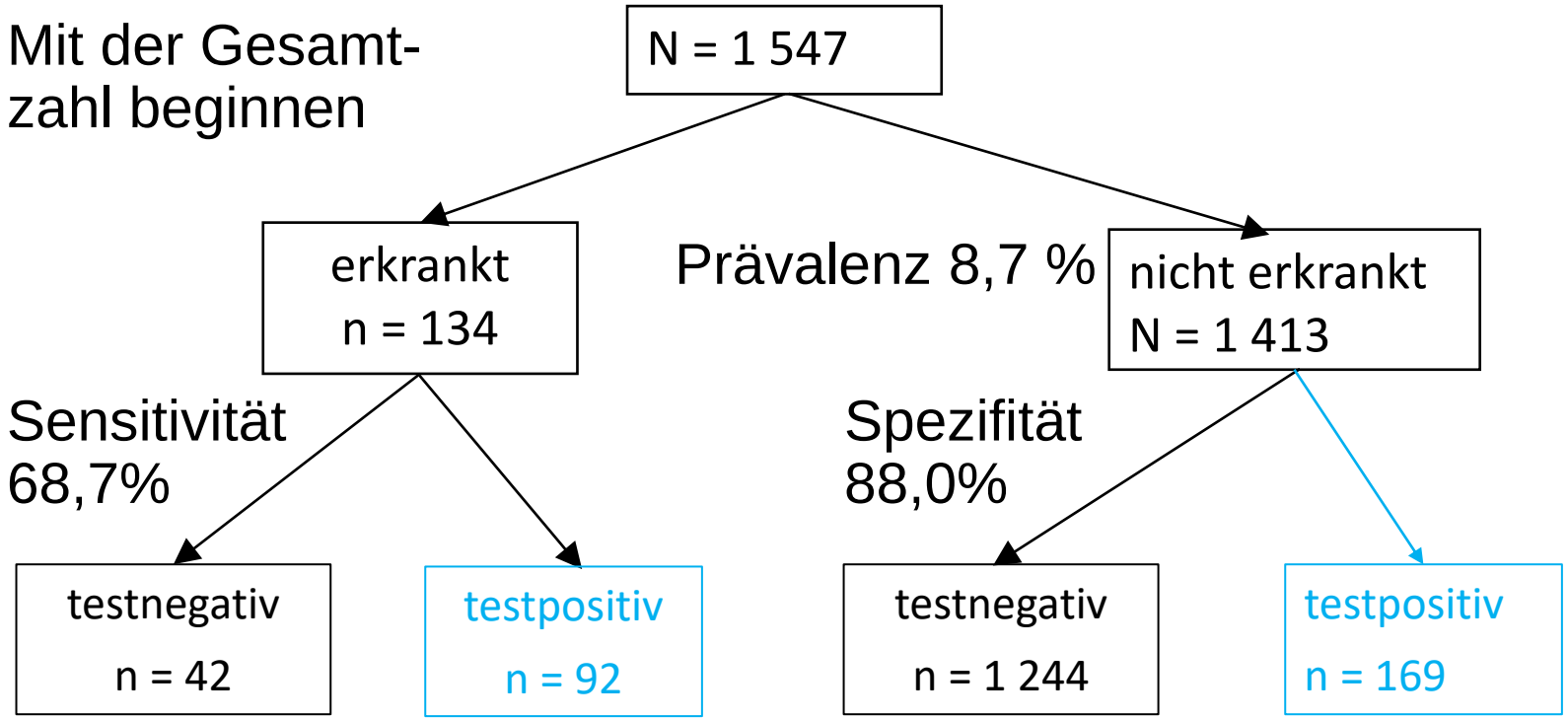
V 4, Folie 37 de Mérés Frage

Können Sie de Mérés Frage (Folie 7) nun beantworten?

$$P(\text{mindestens eine 6 in vier Würfen mit einem Würfel}) =$$
$$1 - P(\text{keine 6 in vier Würfen mit einem Würfel}) =$$
$$1 - \left(1 - \frac{1}{6}\right)^4 = 1 - \left(\frac{5}{6}\right)^4 = 0{,}5177 \approx 52\%$$

$$P(\text{mindestens ein 6er}-\text{Pasch in 24 Würfen mit zwei Würfeln}) =$$
$$1 - P(\text{kein 6er}-\text{Pasch in 24 Würfen mit zwei Würfeln}) =$$
$$1 - \left(1 - \frac{1}{36}\right)^{24} = 0{,}4914 \approx 49\%$$

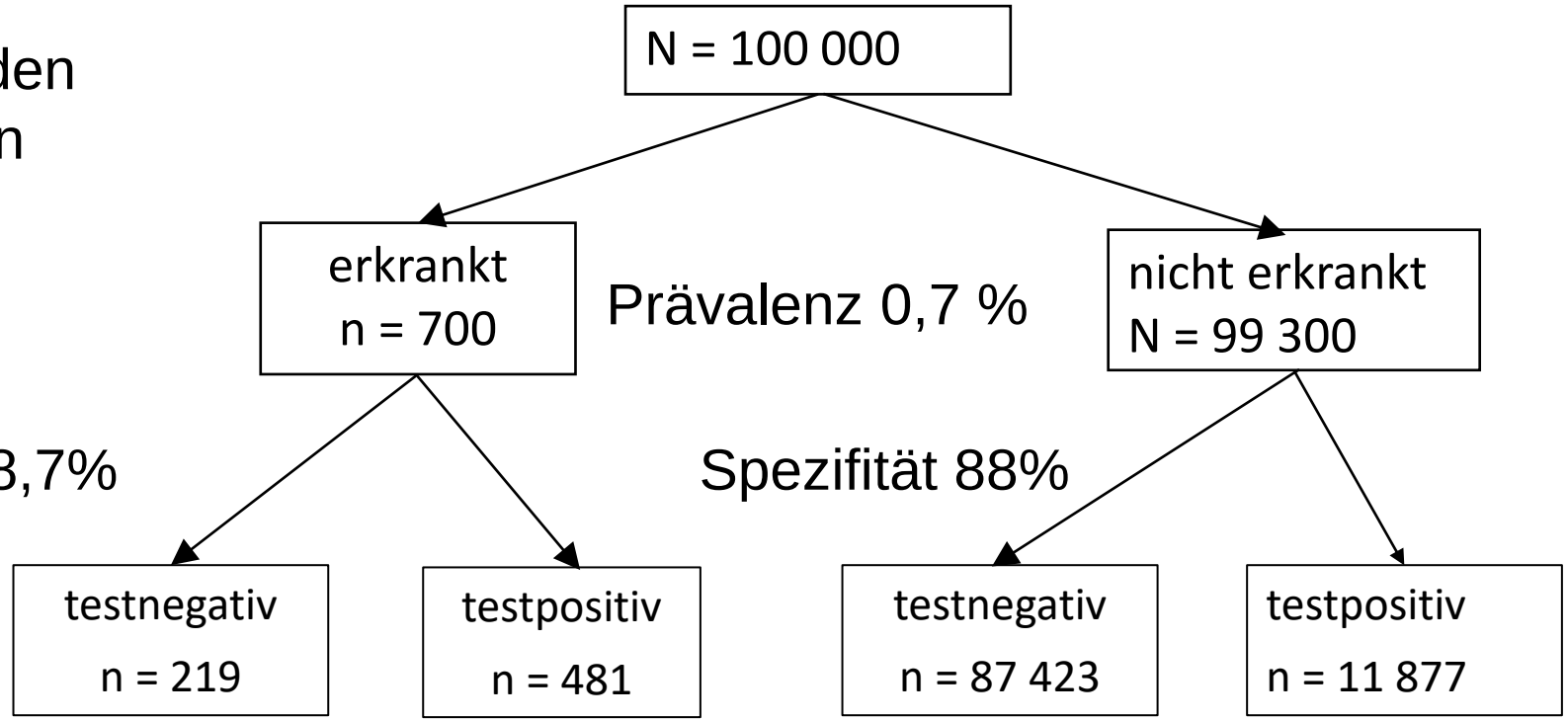
V 4, Folie 66 Lösung 3: PPW mit Ereignisbaum



Positiv-prädiktiver Wert $:= \frac{\#\, richtig\ Positive}{\#\, Testpositive} = \frac{92}{92+169} \approx 0{,}352 =$ 35,2%

V 4, Folie 72 Lösung 3a: PPW mit Ereignisbaum

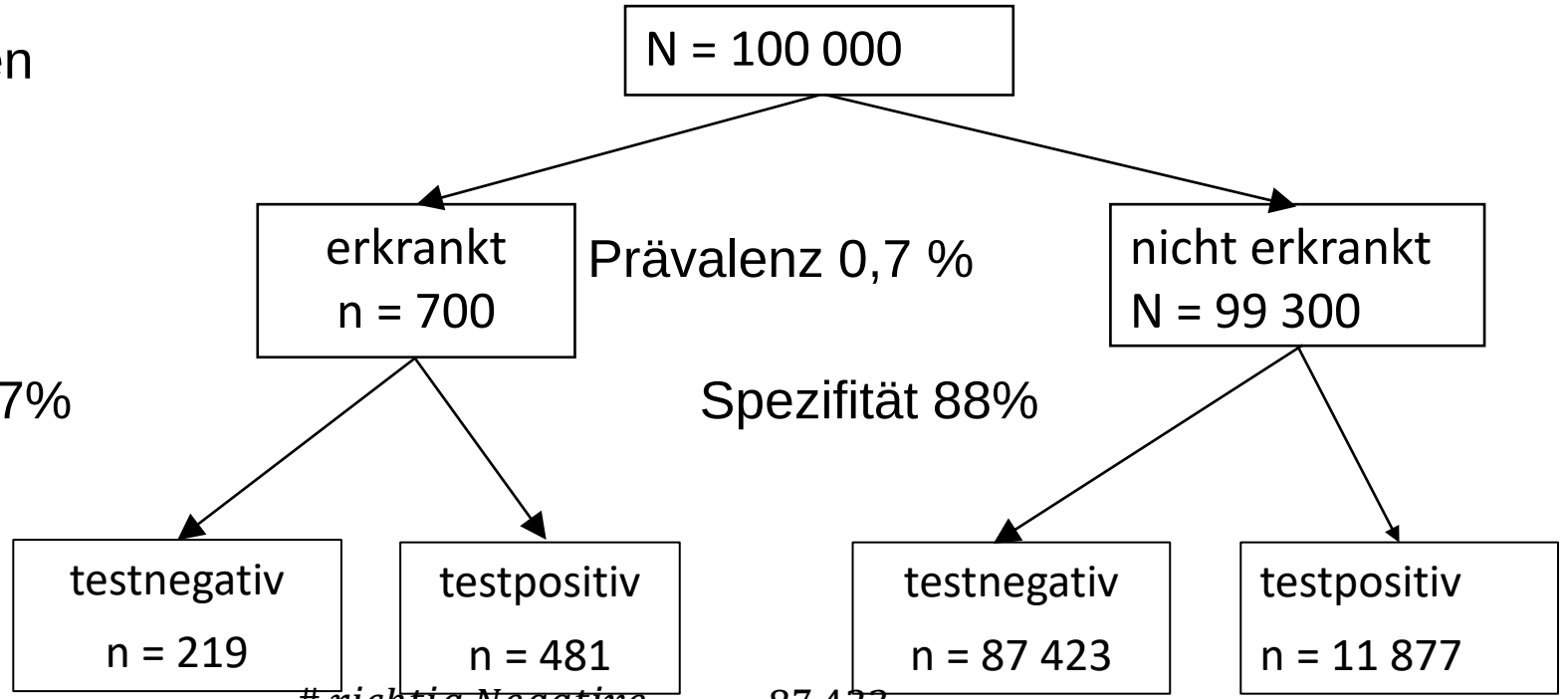
Mit einer runden Zahl beginnen



Positiv-prädiktiver Wert $:= \frac{\#\, richtig\ Positive}{\#\, Testpositive} = \frac{481}{481+11\,877} = 0{,}039 = 3{,}9\%$

V 4, Folie 73 Lösung 3b: NPW mit Ereignisbaum

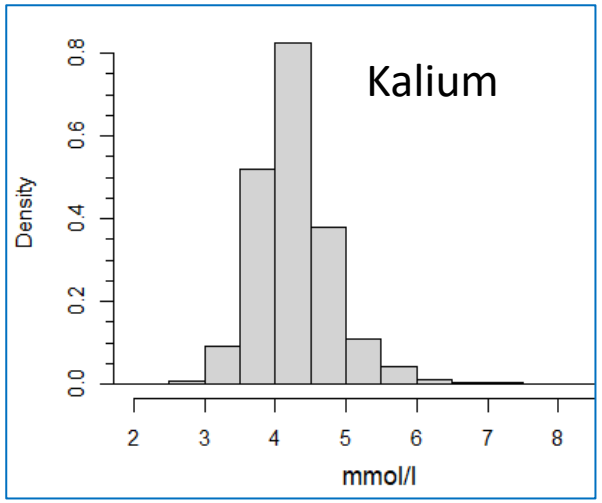
Mit einer runden Zahl beginnen

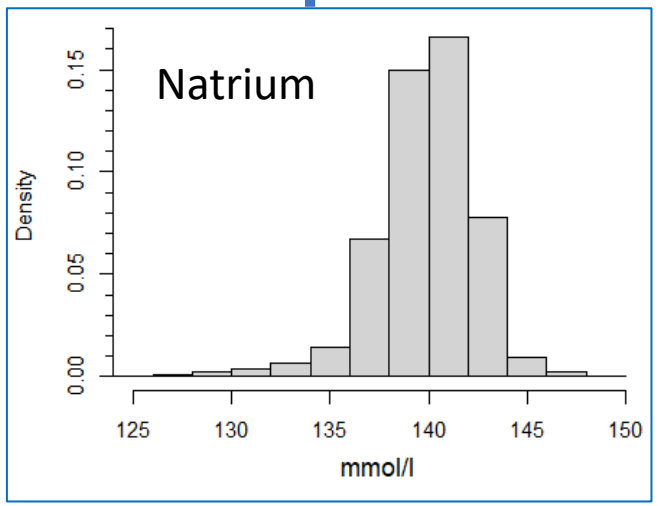


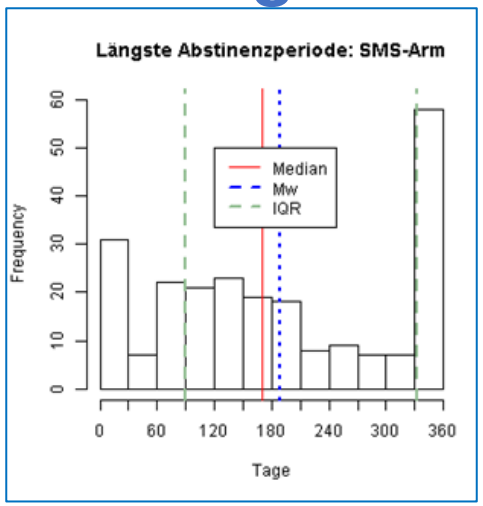
Negativ-prädiktiver Wert $:= \frac{\#\, richtig\ Negative}{\#\, Testnegative} = \frac{87\,423}{87\,423 + 219} = 0{,}9975 = 99{,}75\%$

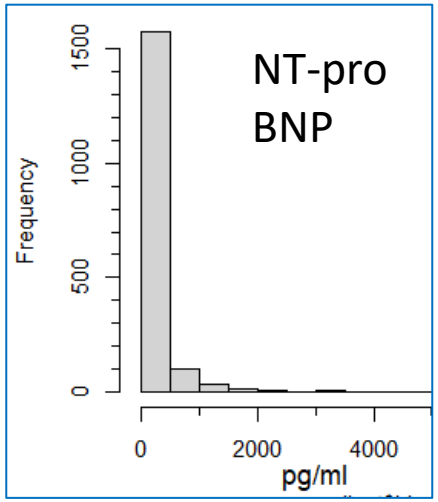
Bei negativem Testausgang kann sich Frau Hof ziemlich sicher sein, kein AF zu haben.

V 5, Folie 16 Verschiedene empirische Verteilungen









Verteilungen unterscheiden sich hinsichtlich

- Wertebereich
- Symmetrie bzw. Schiefe
- extremen Werten

V 5, Folie 18 Lineare Funktionen von ZV

Ist Y eine lineare Funktion einer Zufallsvariablen X. Dann gilt:

- $E[Y] = E[a \cdot X + b] = a \cdot E[X] + b$
- $V[Y] = V[a \cdot X + b] = a^2 \cdot V[X]$

Beispiel 5.2: Sie setzen 55 € und gewinnen 100 €, falls beim Münzwurf Zahl fällt. Wie hoch ist der erwartete Gewinn?

X – Zufallsvariable Münzwurf [0 = Wappen, 1 = Zahl], dann ist:

$$Gewinn = -55 + 100 \cdot X$$

- $p(x) = 0{,}5$;
- $E[X] = 0 \cdot 0{,}5 + 1 \cdot 0{,}5 = 0{,}5$
- Erwarteter Gewinn:

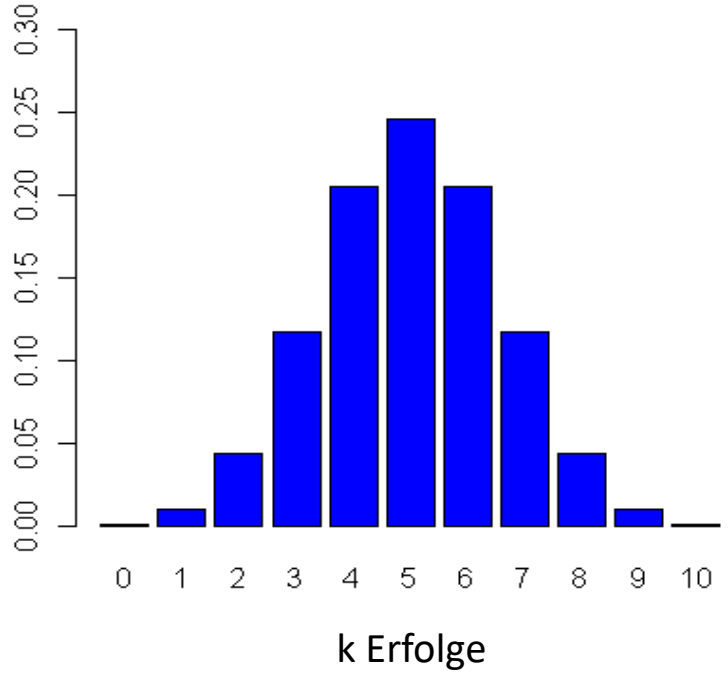
$$E[G] = E[-55 + 100 \cdot X] = -55 + 100 \cdot E[X] = -5\ €$$

V 5, Folie 30 Binomialverteilung VIII

(Stichprobe $n = 10$)

$p = 0{,}5$

$P(X = k)$



$\mu = 0{,}5 \cdot 10 = 5$

$\sigma^2 = 10 \cdot 0{,}5 \cdot (1 - 0{,}5)$

$\sigma = \sqrt{2{,}5} = 1{,}58$



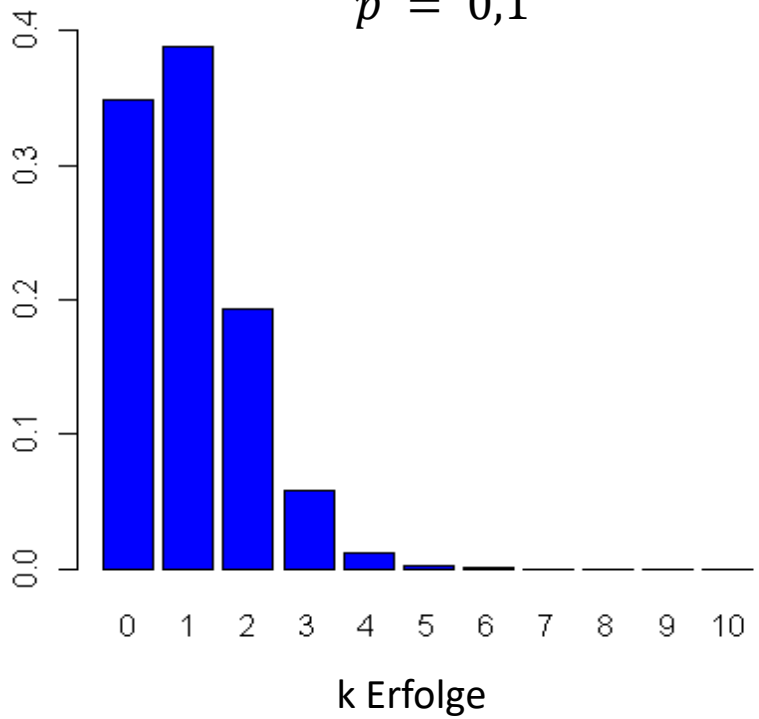
Gonick & Smith: Statistik in Cartoons

V 5, Folie 31 Binomialverteilung IX

(Stichprobe $n = 10$)

$p = 0{,}1$

$P(X = k)$



$\mu = 10 \cdot 0{,}1 = 1$

$\sigma^2 = 10 \cdot 0{,}1 \cdot 0{,}9$

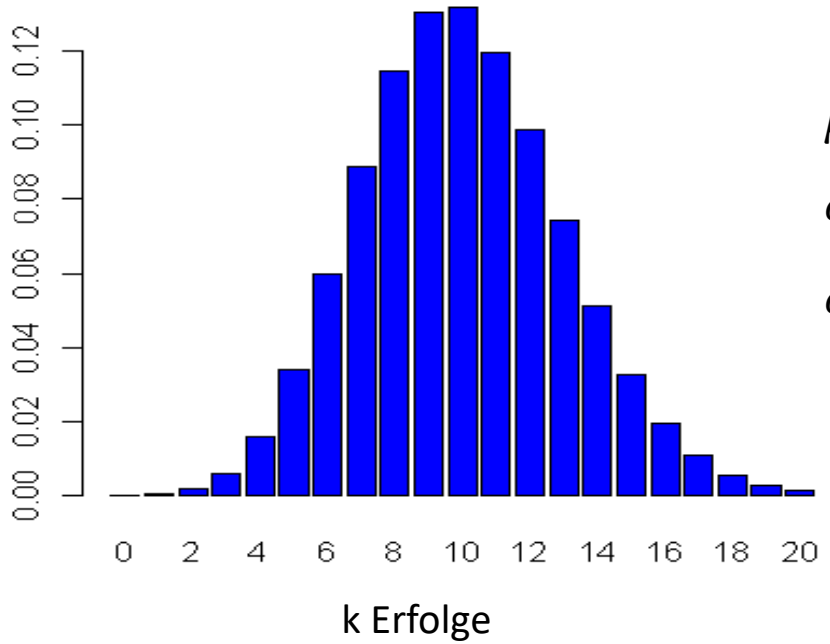
$\sigma = \sqrt{0{,}9} = 0{,}95$

V 5, Folie 32 Binomialverteilung X

(Stichprobe $n = 100$)

$p = 0{,}1$

$P(X = k)$



$\mu = 100 \cdot 0{,}1 = 10$

$\sigma^2 = 100 \cdot 0{,}1 \cdot 0{,}9 = 9$

$\sigma = \sqrt{9} = 3$

V 5, Folie 33 Binomialverteilung – Beispiel B 5.3

30 % der Sprechstunden-Patienten mit einer Diagnose haben Fieber.
Wie groß ist die Wahrscheinlichkeit, dass von 10 Patienten genau zwei Fieber haben?

Die Wahrscheinlichkeit, dass von 10 Patienten genau zwei Fieber haben, beträgt ca. 22%.

$p = 0{,}3; n = 10; k = 2$

$$P(X = k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

$$P(X = 2) = \binom{10}{2} \cdot 0{,}3^2 \cdot (1-0{,}3)^{10\text{-}2} = 45 \cdot 0{,}09 \cdot 0{,}7^8 = 0{,}223$$



Jotquadrat: Grafitto in Bonn. commons. wikipedia.org

Warum könnte diese Rechnung nicht ganz aufgehen?
Das Bernoulli-Schema setzt unabhängige Versuche voraus. Das Auftreten von Fieber könnte u. U. nicht unabhängig sein, z. B. unter Familienangehörigen.

V 6, Folie 44 Lösungen 6.1

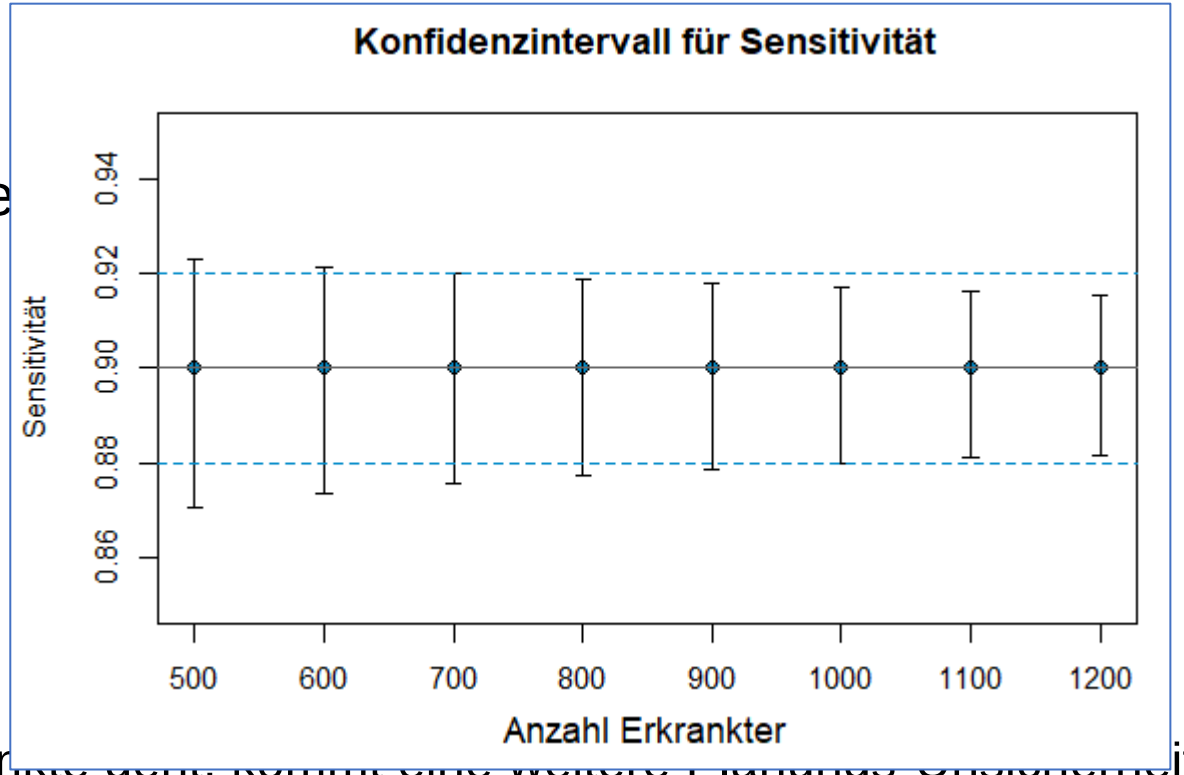
- Sie haben ein 95%-Konfidenzintervall berechnet. Eine Kollegin verlangt ein 99%-Konfidenzintervall. Wird dieses
  ○ kleiner ⊙ größer oder ○ genauso breit wie das vorhandene sein?

- Sie wollen ein zweiseitiges 99%-Konfidenzintervall für den Mittelwert bei gegebener Varianz berechnen. Welches Quantil müssen Sie dafür heraussuchen? $z_{0,995}$

- Es soll ein zweiseitiges 90%-Konfidenzintervall für den Mittelwert berechnet werden. Die Standardabweichung bestimmen Sie aus der Stichprobe von 101 Probanden. Welches Quantil verwenden Sie? $t_{100;\,0,95}$ oder $z_{0,95}$

V 6, Folie 55 Lösung 1

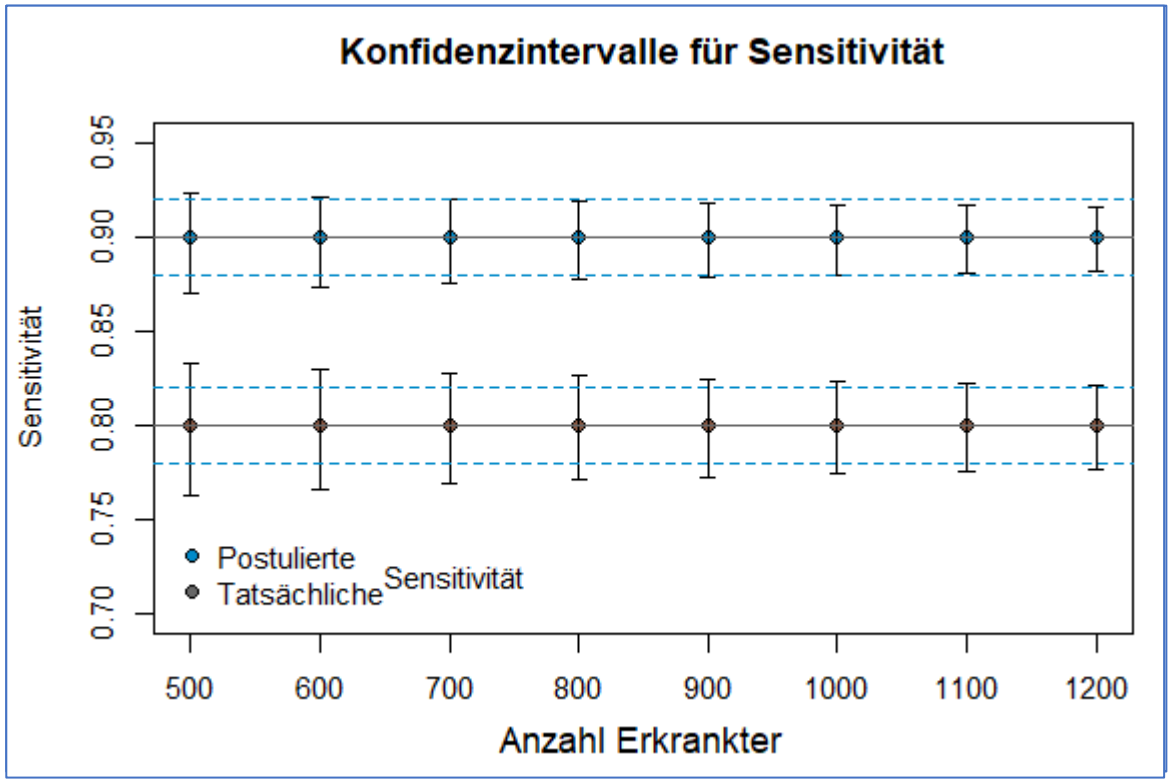
- Nach Formel von Wilson müsste die Anzahl Erkrankter erhöht werden.
- Grafik zeigt, dass 1 100 Probanden *sicher* ausreichen würden: →
  - 2 Jahre Pat. rekrutieren oder
  - multizentrische Studie



Hinweise:

1. Da es bei der Sensitivität um Erkrankte geht, kommt eine weitere Planungs-Unsicherheit hinzu:
   Es ist unklar, wie viele im Zeitraum überhaupt als Erkrankte in die Tabelle eingehen!
2. Die Breite 4% des Konfidenzintervalls wird bereits für 1 000 Erkrankte erreicht. Die untere Grenze des Intervalls läge aber mit 87,98% unter den angestrebten 88%.

V 6, Folie 56 Lösung 2



Hinweise:

3. Die Planung ist dennoch unsicher, da offen bleibt, ob der Zielwert 90% erreicht wird. Wäre die tatsächliche Sensitivität z. B. nur 80%, würden die Konfidenzintervalle deutlich breiter!
4. Zielt man auf eine Mindest-Sensitivität von 90%, könnte man auch einseitige Konfidenzintervalle berechnen

V 6, Folie 56 Konfidenzintervall – Zusammenfassung

Die Breite eines Konfidenzintervalls ist abhängig von:

- der Anzahl der Beobachtungen in der Stichprobe,
- der Varianz des betrachteten Merkmals
- dem Konfidenzniveau α

Ergänzen Sie diese drei Sätze:

- Erhöht man den Stichprobenumfang, wird das Konfidenzintervall … schmaler
- Ist die Varianz größer (als erwartet), wird das Konfidenzintervall … breiter
- Erhöht man das Konfidenzniveau, wird das Konfidenzintervall … breiter