

UNIVERSITÄT LEIPZIG

Medizinische Fakultät

QSB 1: Epidemiologie, Medizinische Biometrie und Medizinische Informatik

Übung 4
Konfidenzintervalle für Anteile und Mittelwerte

Wintersemester 2026/27

imise.

Lernziele

Die Studierenden sollen

- das Konzept der Konfidenzintervalle verstehen und anwenden,
- Konfidenzintervalle für die Schätzung von Wahrscheinlichkeiten (Anteile) und Erwartungswerten (Mittelwerte) berechnen,
- die Voraussetzungen für die Gültigkeit der Näherungsformel prüfen und
- Quantile der t- und Normalverteilung ablesen können.

Konfidenzintervalle (1)

- **Punktschätzer**

  Funktionen, die einen Verteilungsparameter berechnen, nennt man Punktschätzer.

- **Intervallschätzer**

  Punktschätzer sagen nichts über die Genauigkeit der Schätzung aus. Deshalb konstruiert man um den Punktschätzer ein Intervall (Konfidenzintervall, kurz KI).

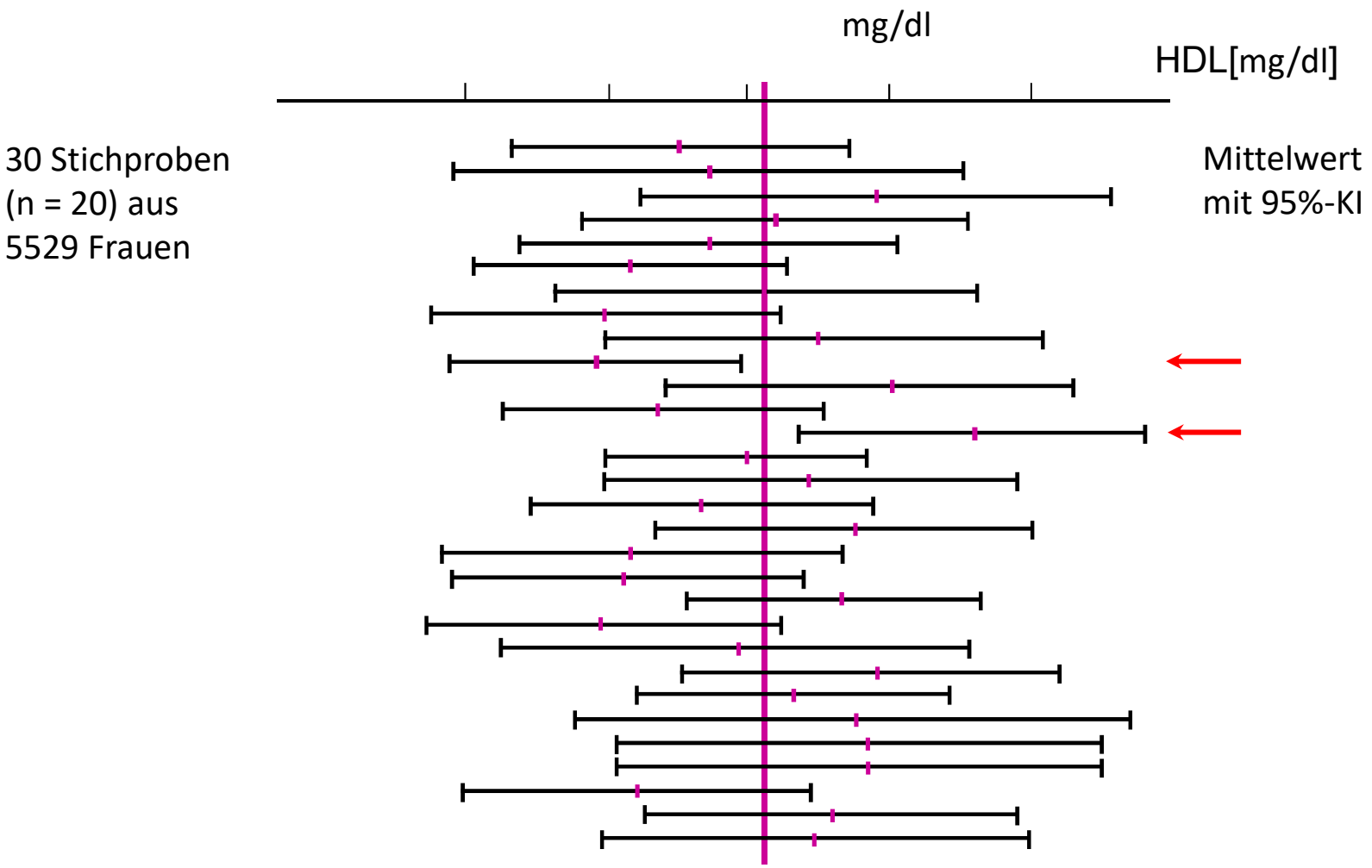
Konfidenzintervalle (2)

Bedeutung des Konfidenzintervalls

- Die Grenzen des 95%-Konfidenzintervalls sind so gewählt, dass das zufällige Intervall mit ca. 95%-iger Wahrscheinlichkeit den wahren Wert überdeckt.

- Modellvorstellung:
  Bei unendlich häufiger Stichprobenziehung werden 95% der so berechneten (festen) Intervalle den wahren Wert enthalten.

Konfidenzintervalle (3)



1. Konfidenzintervall für Anteile

(Synonym: relative Häufigkeiten)

Konfidenzintervalle für Anteile (1)

Aufgabe 1

	Deckel gewölbt (T^+)	Deckel nicht gewölbt (T^-)	Summe
verdorben (D^+)	160	60	220
nicht verdorben (D^-)	120	60	180
Summe	280	120	400

Zur Hausaufgabe ("Diagnose" Joghurt)

- Berechnen Sie das 95%-Konfidenzintervall für die

  Sensitivität $Sens = \frac{160}{220} = 0{,}727$ (also 72,7%)!

Konfidenzintervalle für Anteile (2)

Formel

Anteil in Stichprobe: $\hat{p} = -$ (Ereignisse / Stichprobenumfang)

Anteil in Grundgesamtheit: p (= Wahrscheinlichkeit)

Voraussetzung für Näherungsformel: $n \cdot \hat{p} \cdot (1 - \hat{p}) \geq 9$

Formel: $\underbrace{\hat{p} - z \quad \cdot \sqrt{\frac{(\quad)}{\quad}}}_{\text{untere Grenze}} \leq p \leq \underbrace{\hat{p} + z \quad \cdot \sqrt{\frac{(\quad)}{\quad}}}_{\text{obere Grenze}}$

95%-Konfidenzintervall:

$$\hat{p} - z \quad \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}} \leq p \leq \hat{p} + z \quad \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}}$$

Konfidenzintervalle für Anteile
Tabelle 11.1, Skriptenheft Seite 176

Werte der Standardnormalverteilung $\Phi(z)$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6338	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9416	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936

Konfidenzintervalle für Anteile
Tabelle 11.2, Skriptenheft Seite 177

α	0,1000	0,0500	0,020	0,010	0,005
einseitig: $z_{1-\alpha}$	1,2815	1,645	2,053	2,326	2,575
zweiseitig: $z_{1-\alpha/2}$	1,6450	1,960	2,326	2,575	2,810

Konfidenzintervalle für Anteile (4)

Aufgabe 2

	Deckel gewölbt (T^+)	Deckel nicht gewölbt (T^-)	Summe
verdorben (D^+)	40	15	55
nicht verdorben (D^-)	30	15	45
Summe	70	30	100

Angenommen, im Beispiel aus Aufgabe 1 wären nur 100 statt 400 Joghurtbecher untersucht worden.

- **Berechnen Sie das 95%-Konfidenzintervall für die Sensitivität.**
- **Vergleichen Sie die Ergebnisse aus Aufgabe 1 und 2.**

Konfidenzintervalle für Anteile (6)

Aufgabe 2 - Lösung

$$\hat{p} - z \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}} \leq p \leq \hat{p} + z \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}}$$

Voraussetzung: $n \cdot \hat{p} \cdot (1-\hat{p}) = 55 \cdot 0{,}727 \cdot 0{,}273 = 10{,}9 > 9$

Die Bedingung für die Anwendung der Näherungsformel ist erfüllt.

$$0{,}727 - 1{,}96 \cdot \sqrt{\frac{0{,}727 \cdot 0{,}273}{55}} \leq p \leq 0{,}727 + 1{,}96 \cdot \sqrt{\frac{0{,}727 \cdot 0{,}273}{55}}$$

$$0{,}609 \leq p \leq 0{,}845$$

95%-KI: [60,9; 84,5]%

Konfidenzintervalle für Anteile (7)

Aufgabe 2 - Lösung

$n = 220$:

95%-KI: $[66{,}8; 78{,}6]\%$

$n = 55$:

95%-KI: $[60{,}9; 84{,}5]\%$

Breite des Konfidenzintervalls:

$n = 220$: $11{,}8\%$

$n = 55$: $23{,}6\%$;

d. h., man braucht hier den vierfachen Stichprobenumfang, um die Breite des Konfidenzintervalls zu halbieren.

Konfidenzintervalle für Anteile (8)

Aufgabe 3 - Historisches Ereignis

Bei dem Giftgasunfall im Raum Seveso 1976 war hochgiftiges Dioxin ausgetreten.

Später fand man heraus, dass die Männer, welche einer Exposition mit Dioxin ausgesetzt waren, in den Folgejahren 88 Jungen und 103 Mädchen gezeugt hatten.

Die Geschlechterverteilung in der Normalbevölkerung ist wie 106 Jungen zu 100 Mädchen.

- Eine Schlagzeile in der Presse behauptete, dass Dioxin die Zeugung von Jungen verringert. Prüfen Sie diese Behauptung mit einem Konfidenzintervall!

Konfidenzintervalle für Anteile (9)

Aufgabe 3 - Vorüberlegung

- Zufallsereignis: Zeugung eines Jungen
- Idee:
  - Berechne Wahrscheinlichkeit p_0 für Jungengeburten in der Allgemeinbevölkerung!
  - Berechne den Anteil an Jungengeburten für Seveso-Arbeiter inklusive 99%-Konfidenzintervall!
  - Prüfe, ob dieses Konfidenzintervall die Wahrscheinlichkeit p_0 enthält! Wenn ja, dann ist die Häufung von Mädchen-geburten aus dem Zufall erklärbar.

2. Konfidenzintervall für einen Mittelwert

Konfidenzintervalle für Mittelwerte (1)

Formel

1. Standardabweichung σ in der Grundgesamtheit bekannt:

$$\bar{x} - z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

2. Standardabweichung in der Grundgesamtheit unbekannt:

$$\bar{x} - t_{f;1-\alpha/2} \cdot \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{f;1-\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

s - aus Stichprobe geschätzte Standardabweichung

f - Freiheitsgrad (f = n-1)

für n > 100 gilt $t \sim z$, weshalb z genutzt werden kann

Konfidenzintervalle für Mittelwerte (2)

Vergleich mit KI für Anteile

Konfidenzintervalle für Anteile und Mittelwerte

- Anteile

$$\hat{p} - z_{1-\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{1-\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}}$$

Punkt-schätzer

$(1 - \quad / \,)$-Quantil

$\frac{empir.Standardabweichung}{\sqrt{n}}$

$$\bar{x} - t_{f;1-\alpha/2} \cdot \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{f;1-\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

- Mittelwerte

Konfidenzintervalle für Mittelwerte (3)

Aufgabe 4

- Dr. X hatte bei 25 Frauen Brustkrebs diagnostiziert. Die Frauen waren im Mittel $\bar{x} = 55{,}0$ Jahre alt bei einer Standardabweichung von $s = 10{,}0$.

  Aus der Literatur ist ihm bekannt, dass das Erkrankungsalter bei früheren Untersuchungen im Populationsmittel bei 62,0 Jahren liegt.

- Dr. X fragt sich, ob seine Stichprobe mit der Patientengruppe der Grundgesamtheit in Bezug auf das Alter vergleichbar ist?

→ **Berechnen Sie zur Beantwortung dieser Frage das 95%-Konfidenzintervall.**

Hausaufgabe (1)
Konfidenzintervall

- Berechnen Sie für beide Therapiearme das 99%-Konfidenzintervall für den Mittelwert des Alters.
- Berechnen Sie für beide Therapiearme das 99%-Konfidenzintervall für den Anteil weiblicher Studienteilnehmer.

Table 1. Demographic and Clinical Characteristics of the Patients at Baseline.*

Characteristic	Canagliflozin (N=2202)	Placebo (N=2199)	All Patients (N=4401)
Age — yr	62.9±9.2	63.2±9.2	63.0±9.2
Female sex — no. (%)	762 (34.6)	732 (33.3)	1494 (33.9)