# Forschungsmethodik: Interventions- und Evaluationsforschung theoretisch begründen

*Vorlesung Biostatistik, WS 2026/2027*

*Dr. Marcus Wagner, IMISE*

# Vorlesung II am 20.10.2026

---

**Inhaltsverzeichnis**

*2. Deskriptive Statistik*

*2.4. Wiederholung*

*2.5. Gesamtpopulation: metrische Variablen*

*2.6. Anpassen eines Verteilungsmodells*

*3. Schließen von der Stichprobe auf die Gesamtpopulation*

*3.1. Schätzung von Kenngrößen*

*3.2. Konfidenzintervall: Grundlagen*

*3.3. Konfidenzintervall: Beispiele*

*4. Exkurs: Studien als "statistische Experimente"*

*4.1. Studiendesigns*

*4.2. Elemente der Studienplanung*

---

# Aus Vorlesung I: Erstes Grundkonzept der Statistik

**... für Merkmale mit vollständiger Information**

- *Deskriptive (beschreibende) Statistik*
- *Merkmale werden immer als Zufallsgrößen betrachtet und in der Terminologie der Wahrscheinlichkeitstheorie beschrieben: Verteilung des Merkmals, relative Häufigkeiten, Verteilungsfunktion, Verteilungsdichte, Mittelwert, Standardabweichung, ...*
- *Mathematik bleibt jedoch im Hintergrund; wir verlassen uns auf Angemessenheit und Widerspruchsfreiheit der Begriffe und benutzen die fertigen Werkzeuge ("Denkzeuge")*
- *Datenreduktion: wir wollen die ursprüngliche Verteilung der Merkmale mit wenigen Begriffen und Kenngrößen beschreiben und darstellen. Achtung: Informationsverlust!*

# Aus Vorlesung I: Deskriptive Statistik

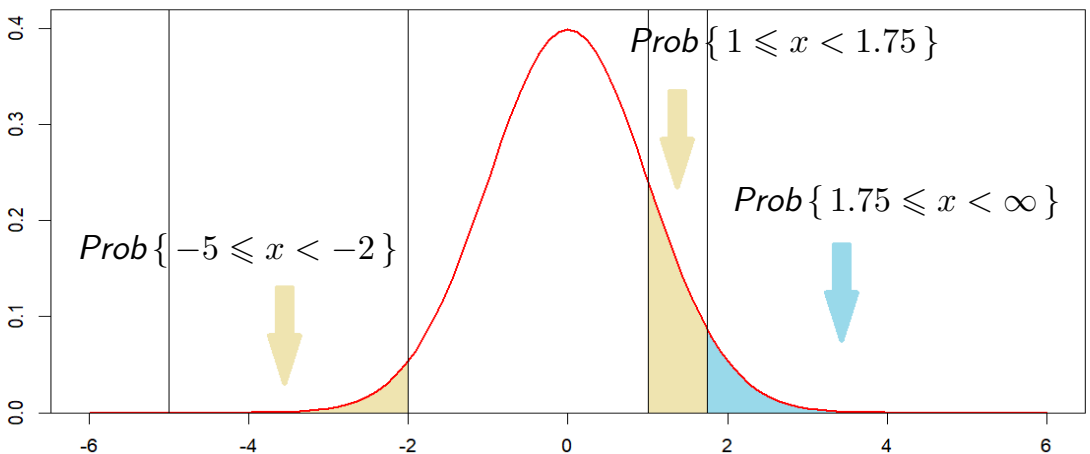
• **Kenngrößen für metrische Variablen in einer Stichprobe:** *ergeben sich aus den empirischen Häufigkeiten (sog. diskrete Wahrscheinlichkeitsverteilung)*

| | *"parametrisch"* | *"nichtparametrisch"* |
|---|---|---|
| *Definition* | *aufgrund der Werte* | *aufgrund der Ränge* |
| *Lagemaß* | *Stichproben-Mittelwert* $\widehat{\mu}$ | *Stichproben-Median* $\widehat{m}$ |
| *Streuungsmaße* | *emp. Standardabweichung* $\widehat{\sigma}$, *(Varianz* $\widehat{\sigma}^2$, *Spannweite* $\widehat{R}$ *)* | *Quartile* $\widehat{Q}_{25}$ *und* $\widehat{Q}_{75}$, *Interquartil-Range IQR* |
| *Robustheit gegen Ausreißer* | *nicht robust* | *robust geg. wenige Ausreißer* |

• **Kenngrößen für metrische Variablen in der Gesamtpopulation:** *Hier sind neue Berechnungsvorschriften nötig, weil die Merkmale in der Gesamtpopulation nicht mehr mit diskreten, sondern mit stetigen Wahrscheinlichkeitsverteilungen beschrieben werden.*

# Stetige Wahrscheinlichkeitsverteilung: wie beschreiben?

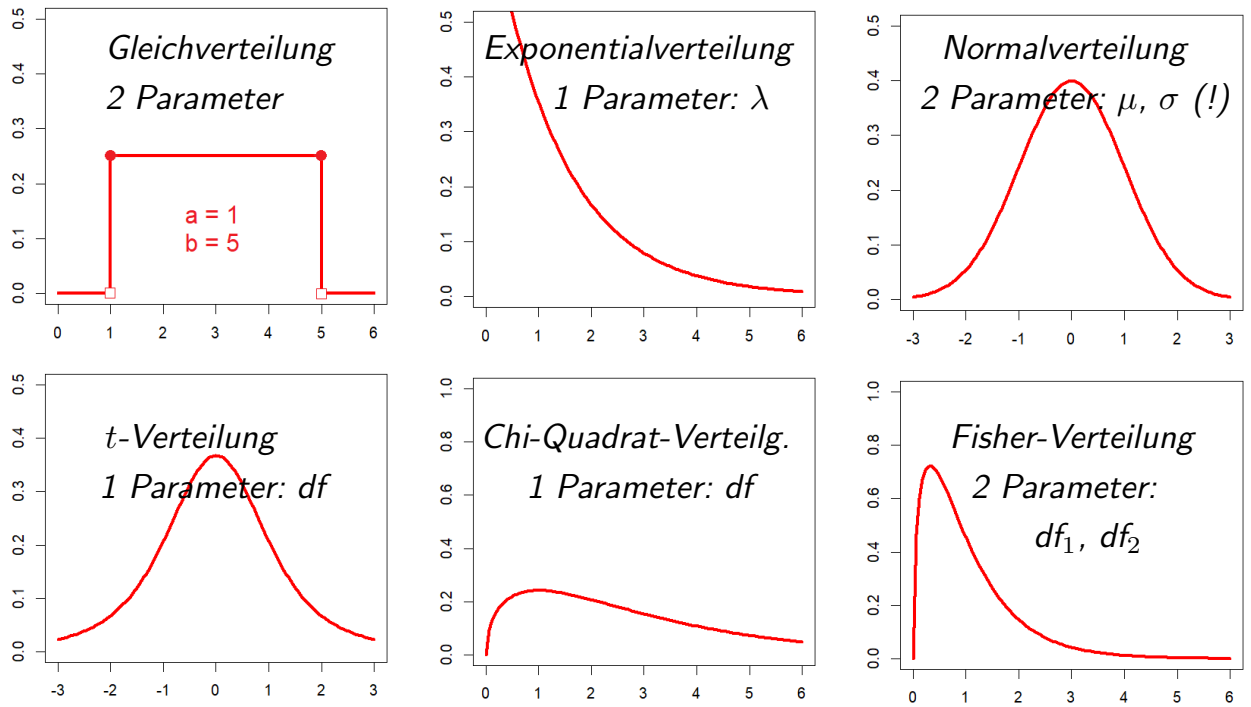
• **Grundsatz:** *Alle Informationen über eine stetige Wahrscheinlichkeitsverteilung kommen aus ihrer Verteilungsdichte* $f$.




• **Definition 2.1.:** *(Verteilungsdichte) Das ist eine nichtnegative Funktion* $f$*, aus der man alle Wahrscheinlichkeiten dafür, daß die metrische Variable* $x$ *in ein halboffenes Intervall* $[\,a\,,\,b\,)$ *fällt, durch Integration erhalten kann:*

$$\text{Prob}\{\,a \leqslant x < b\,\} = \int_a^b f(x)\,dx \quad \textit{(Wahrscheinlichkeit gleich Fläche unter der Kurve)}$$

# Stetige, parameterabhängige Verteilungen



# Kenngrößen aus der Dichte

- *$f$ sei die Verteilungsdichte für eine metrische Variable $x$ in der Gesamtpopulation.*
- **Definition 2.2.:** *(Mittelwert $\mu$) — Integral über die gesamte reelle Achse:*

$$\mu = \int_{-\infty}^{\infty} x \cdot f(x)\, dx\,.$$

- **Definition 2.3.:** *(Standardabweichung $\sigma$) — Integral über die gesamte reelle Achse:*

$$\sigma \;=\; \sqrt{\int_{-\infty}^{\infty} \big(\, x - \mu \,\big)^2 \cdot f(x)\, dx}\;,$$ *darin ist $\mu$ der Mittelwert von $x$.*

- **Definition 2.4.:** *(Verteilungsfunktion $F$) — unbestimmtes Integral der Verteilungsdichte $f$:*

$$F(x_0) = \mathit{Prob}\,\{\, -\infty \leqslant x < x_0 \,\} = \int_{-\infty}^{x_0} f(x)\, dx\,.$$

- *Diese Integrale sind für alle wichtigen Verteilungen in Abhängigkeit von ihren Parametern vertafelt. (Um die Details der Integration müssen wir uns also nicht kümmern!)*

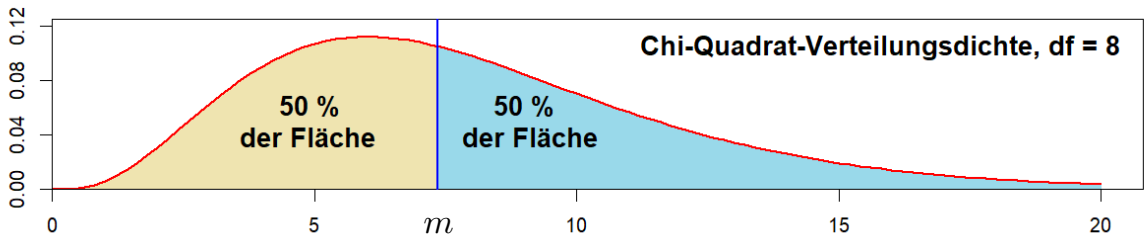
# Kenngrößen aus der Dichte: Median

• **Definition 2.5.:** *(Median $m$) Für eine metrische Variable $x$ in der Gesamtpopulation sei die Verteilungsdichte $f$ gegeben. Der Median $m$ der Variablen ist derjenige $x$-Wert, für den die Fläche unter der Verteilungsdichte halbiert wird.*

• **In Zahlen:** *Der Median $m$ von $x$ ist die Lösung der folgenden Integralgleichung:*

$$Prob\{-\infty \leqslant x < m\} = F(m) = \int_{-\infty}^{m} f(x)\,dx = 0.5\,.$$

• **Beispiel 2.6.:** *Median einer Chi-Quadrat-Verteilung*



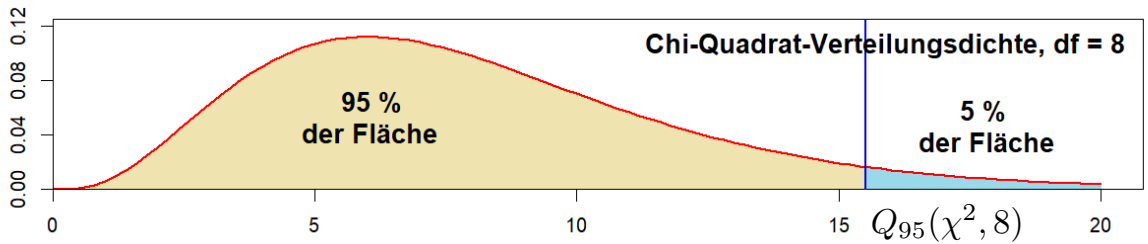
# Kenngrößen aus der Dichte: Quantile

• **Definition 2.7.:** *(beliebiges $p$-Quantil $Q_p(f)$) Für eine metrische Variable $x$ in der Gesamtpopulation sei die Verteilungsdichte $f$ gegeben. Das $p$-Quantil $Q_p(f)$, $0 < p < 100$, ist derjenige $x$-Wert, der die Fläche unter der Verteilungsdichte so aufteilt, daß $p$ Prozent der Gesamtfläche links von $Q_p(f)$ liegen, und die restlichen $(100 - p)$ Prozent der Gesamtfläche rechts von $Q_p(f)$ liegen.*

• **In Zahlen:** *Das Quantil $Q_p(f)$ von $x$ ist die Lösung der folgenden Integralgleichung:*

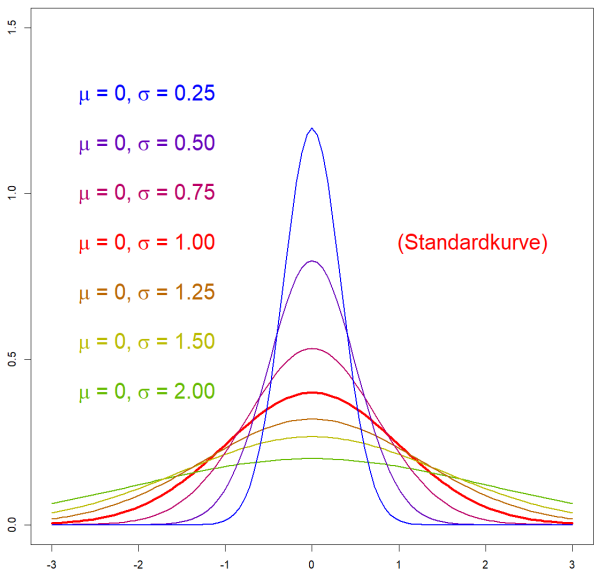
$$\mathit{Prob}\{\, -\infty \leqslant x < Q_p(f)\,\} = F(\,Q_p(f)\,) = \int_{-\infty}^{Q_p(f)} f(x)\,dx \;=\; p/100\,.$$

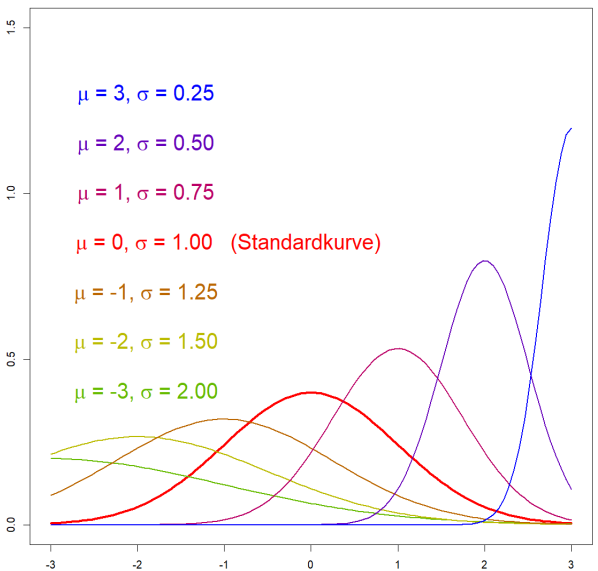
• **Beispiel 2.8.:** *95-%-Quantil einer Chi-Quadrat-Verteilung*



# Parameter-Abhängigkeit der Normalverteilung

- **"Kurvendiskussion":** *Bei der Normalverteilung hängt die Gestalt der Verteilungsdichte unmittelbar vom Mittelwert $\mu$ und der Standardabweichung $\sigma$ ab.*

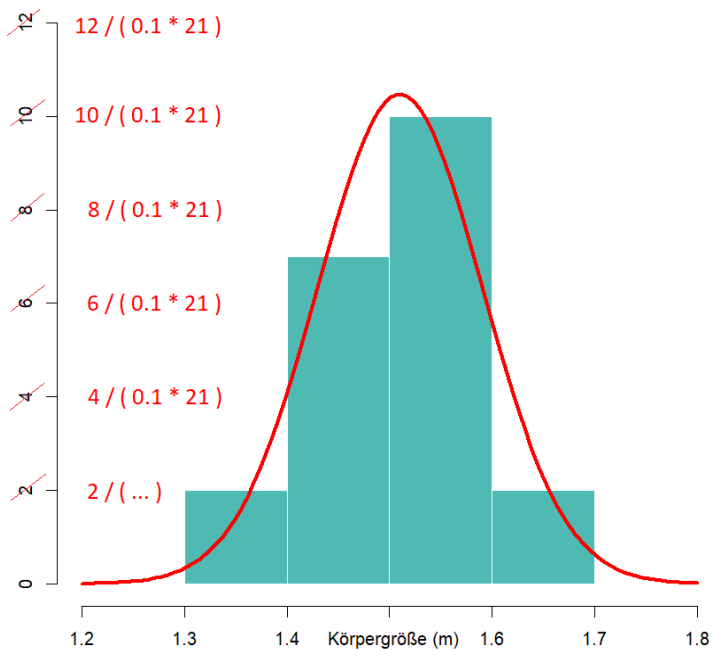



- *$\mu$ bestimmt Position der Symmetrieachse / $\sigma$ bestimmt Steilheit der Glockenkurve*

# Anpassen ans Histogramm

• *Beispiel 1.15.: Approximation fürs Histogramm durch die Verteilungsdichte einer Normalverteilung* $f(x, \mu, \sigma)$



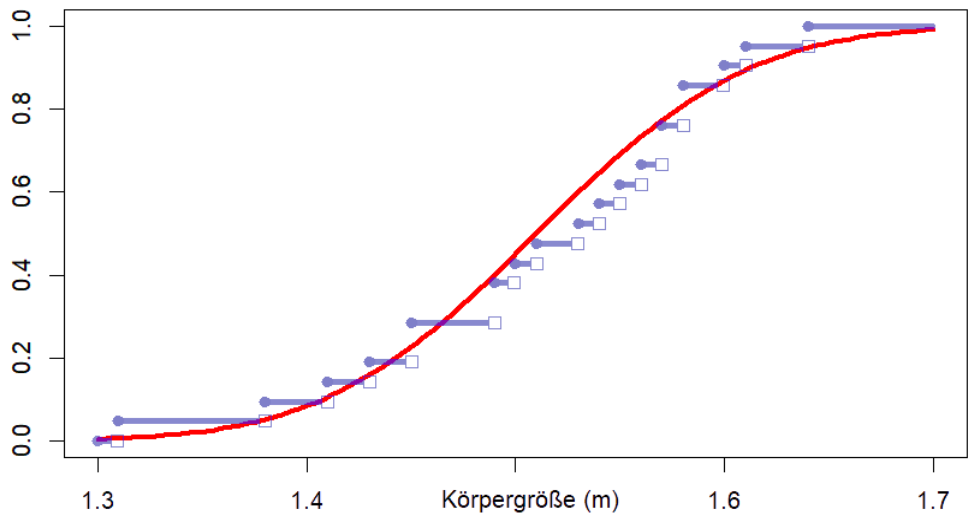

• *Skalierung der* $y$*-Achse auf relative statt absolute Häufigkeiten und Breite der Balken (Division durch 0.1 und 21): Fläche des Histogramms wird auf Eins normiert*

• *2 Parameter für Normalverteilung übernehmen:*
$\mu \approx \widehat{\mu} = 1.51, \sigma \approx \widehat{\sigma} = 0.08$

• *Ergebnis: angepaßte Verteilungsdichte* $f(x, 1.51, 0.08)$

# Anpassen an die empirische Verteilungsfunktion

• *Beispiel 1.15.: Approximation für die empirische Verteilungsfunktion* $\widehat{F}(x)$ *durch die Verteilungsfunktion* $F(x, \mu, \sigma)$ *einer Normalverteilung*




• *2 Parameter für Normalverteilung:* $\mu = 1.51$, $\sigma = 0.08$*; keine weitere Skalierung nötig*

• *Ergebnis: angepaßte Verteilungsfunktion* $F(x, 1.51, 0.08)$

# Gesamtpopulation und Stichprobe

• *Kenngrößen der Stichprobe als Schätzungen für die Kenngrößen der Gesamtpopulation*

| **Gesamtpopulation** $\mathcal{P}$ | **Stichprobe** $\mathcal{S}$ | **Definitionen** |
|---|---|---|
| *Gesamtumfang* $N(\mathcal{P})$ | *Stichproben-Umfang* $N(\mathcal{S}) = N$ | |
| *metrische Variable* $x$ | *metrische Variable* $x$ | *identisch* |
| *Mittelwert* $\mu$ | *Stichproben-Mittelwert* $\widehat{\mu}$ | *2.2. / 1.16.* |
| *Standardabweichung* $\sigma$ | *empirische Standardabweichung* $\widehat{\sigma}$ | *2.3. / 1.17.* |
| *Median* $m$ | *Stichproben-Median* $\widehat{m}$ | *2.5. / 1.19.* |
| *Verteilungsfunktion* $F$ | *empirische Verteilungsfunktion* $\widehat{F}$ | *2.4. / 1.22.* |

• **Entscheidende Frage:** *Wie genau sind diese Schätzungen?*

• **Zitat aus Vorlesung 1:** *"Alle Schlüsse von der Stichprobe auf die Merkmalsverteilung in der Gesamtpopulation sind mit prinzipieller, unvermeidlicher Unsicherheit behaftet. Diese Unsicherheit können wir jedoch zahlenmäßig beschreiben."*

# Verteilung der Stichproben-Mittelwerte

• **Idee:** *Ziehe aus der Gesamtpopulation $\mathcal{P}$ immer wieder (unabhängig voneinander, mit Zurücklegen) neue Stichproben $\mathcal{S}^{(1)}$, $\mathcal{S}^{(2)}$, $\mathcal{S}^{(3)}$, ... , alle vom gleichen Umfang $N$, und bilde dazu die Stichproben-Mittelwerte $\widehat{\mu}^{(1)}$, $\widehat{\mu}^{(2)}$, $\widehat{\mu}^{(3)}$, ... Nun fassen wir die Stichproben-Mittelwerte selbst als eine neue metrische Variable in $\mathcal{P}$ auf.*

• **Frage:** *Wie sind die Stichproben-Mittelwerte verteilt?*

• **Antwort** *aus der Wahrscheinlichkeitstheorie:*

**Lehrsatz 2.9.:** *(Verteilung der Stichproben-Mittelwerte) Eine metrische Variable $x$ in der Gesamtpopulation $\mathcal{P}$ möge den Mittelwert $\mu$ und die Standardabweichung $\sigma$ besitzen. Unabhängig davon, auf welche Weise $x$ in $\mathcal{P}$ verteilt ist, sind die Stichproben-Mittelwerte dieser Variablen (für unabhängige Stichproben vom gleichen Umfang $N \geqslant 30$) in sehr guter Näherung normalverteilt, und zwar mit Mittelwert $\mu$ und Standardabweichung $\sigma/\sqrt{N}$.*

• *Mit Hilfe dieser Aussage können wir die Abweichung der Mittelwertschätzung $\widehat{\mu}$ vom wahren Mittelwert $\mu$ zahlenmäßig erfassen.*

# Konfidenzintervall für den Stichproben-Mittelwert

- *Die Stichproben-Mittelwerte sind normalverteilt mit Parametern $\mu$ und $\sigma/\sqrt{N}$: 95% aller $\widehat{\mu}$-Werte liegen im Intervall $[\,Q_{2.5}(\mathcal{N})\,,\,Q_{97.5}(\mathcal{N})\,]$ mit Mittelpunkt $\mu$.*

$\Longrightarrow$ *(i) das Intervall hat die Darstellung*

$$\left[\,\mu - \frac{Q_{97.5}(\mathcal{N}) - Q_{2.5}(\mathcal{N})}{2}\,,\,\mu + \frac{Q_{97.5}(\mathcal{N}) - Q_{2.5}(\mathcal{N})}{2}\,\right]$$ *mit Mittelpunkt $\mu$,*

$\Longrightarrow$ *(ii) ein einzelner Stichproben-Mittelwert $\widehat{\mu}$ fällt mit einer Wahrscheinlichkeit von 95% in das Intervall*

$$\left[\,\mu - \frac{Q_{97.5}(\mathcal{N}) - Q_{2.5}(\mathcal{N})}{2}\,,\,\mu + \frac{Q_{97.5}(\mathcal{N}) - Q_{2.5}(\mathcal{N})}{2}\,\right]$$ *mit Mittelpunkt $\mu$,*

$\Longrightarrow$ *(iii) der wahre Mittelwert $\mu$ fällt mit einer Wahrscheinlichkeit von 95% in das verschobene Intervall*

$$\left[\,\widehat{\mu} - \frac{Q_{97.5}(\mathcal{N}) - Q_{2.5}(\mathcal{N})}{2}\,,\,\widehat{\mu} + \frac{Q_{97.5}(\mathcal{N}) - Q_{2.5}(\mathcal{N})}{2}\,\right]$$ *mit Mittelpunkt $\widehat{\mu}$!*

# Konfidenzintervall für den Stichproben-Mittelwert (II)

• *Die 2.5-%- und 97.5-%-Quantile zur Normalverteilung $\mathcal{N}$ mit Parametern $\mu$ und $\sigma/\sqrt{N}$ sind bekannt:* $Q_{2.5}(\mathcal{N}) = \mu - 1.96 \cdot \sigma/\sqrt{N}$, $Q_{97.5}(\mathcal{N}) = \mu + 1.96 \cdot \sigma/\sqrt{N}$.

• **Definition 2.10.:** *(95-%-Konfidenzintervall für den Stichproben-Mittelwert) Für eine metrische Variable $x$ in der Gesamtpopulation $\mathcal{P}$ sei eine Stichprobe vom Umfang $N$ mit Stichproben-Mittelwert $\widehat{\mu}$ gegeben. Dann erklären wir das 95-%-Konfidenzintervall für $\widehat{\mu}$ durch*

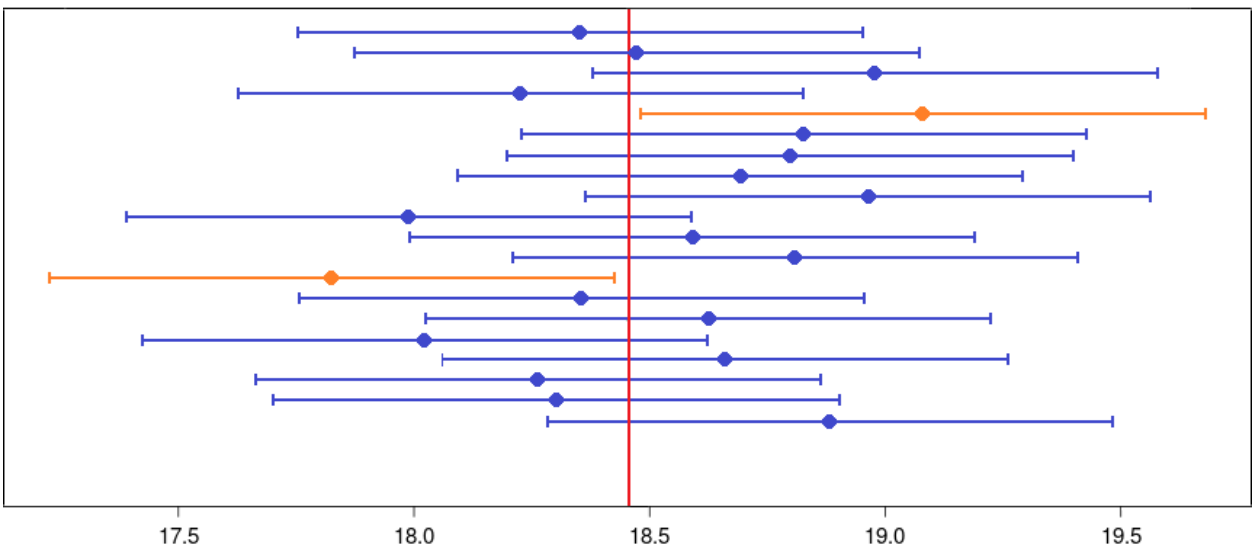
$$\left[\, \widehat{\mu} - \frac{1.96 \cdot \sigma}{\sqrt{N}} \,,\, \widehat{\mu} + \frac{1.96 \cdot \sigma}{\sqrt{N}} \,\right],$$

*worin $\sigma$ die Standardabweichung von $x$ in $\mathcal{P}$ ist.*

• **Interpretation ("in the long run"):** *Man stellt sich vor, daß das Ziehen der Stichprobe vom Umfang $N$ unendlich oft wiederholt werden kann. Dann überdecken die 95-%-Konfidenzintervalle für die jeweiligen Stichproben-Mittelwerte in 95% aller Fälle den (unbekannten) wahren Mittelwert $\mu$ der Variablen $x$ in der Gesamtpopulation.*

# Konfidenzintervall für den Stichproben-Mittelwert (III)

• **Beispiel 2.11.:** *"Boston Housing Dataset", metrische Variable $x$: durchschnittliche Klassenstärke im Stadtbezirk,* $N(\mathcal{P}) = 506$, $\mu = 18.46$, $\sigma = 2.16$




• *20 Stichproben mit $N = 50$; von 20 zugehörigen 95-%-Konfidenzintervallen für die Stichproben-Mittelwerte überdecken 18 Intervalle den wahren Mittelwert $\mu$ in $\mathcal{P}$*

# Konfidenzintervall: Zusätze und Verallgemeinerungen

• **Zusatz zu Definition 2.10.:** *Wenn die Stichprobe groß genug ist ($N \geqslant 30$), darf bei der Bildung des 95-%-Konfidenzintervalls anstelle der (möglicherweise unbekannten) wahren Standardabweichung $\sigma$ die empirische Standardabweichung $\widehat{\sigma}$ benutzt werden.*

• **Konfidenzintervalle zu beliebigen Prozentzahlen:** *Zum Beispiel lautet das 90-%-Konfidenzintervall für den Stichproben-Mittelwert $\widehat{\mu}$:*

$$\left[\, \widehat{\mu} - \frac{Q_{95}(\mathcal{N}) - Q_5(\mathcal{N})}{2} \,,\, \widehat{\mu} + \frac{Q_{95}(\mathcal{N}) - Q_5(\mathcal{N})}{2} \,\right] \;=\; \left[\, \widehat{\mu} - \frac{1.645 \cdot \sigma}{\sqrt{N}} \,,\, \widehat{\mu} + \frac{1.645 \cdot \sigma}{\sqrt{N}} \,\right],$$

*und allgemein das $p$-%-Konfidenzintervall für den Stichproben-Mittelwert $\widehat{\mu}$:*

$$\left[\, \widehat{\mu} - \frac{Q_{(100-p/2)}(\mathcal{N}) - Q_{p/2}(\mathcal{N})}{2} \,,\, \widehat{\mu} + \frac{Q_{(100-p/2)}(\mathcal{N}) - Q_{p/2}(\mathcal{N})}{2} \,\right].$$

• **Interpretation ("in the long run"):** *Wird das Ziehen der Stichprobe vom Umfang $N$ unendlich oft wiederholt, so überdecken die 90-%-Konfidenzintervalle ($p$-Prozent-Konfidenzintervalle) für die jeweiligen Stichproben-Mittelwerte in 90% aller Fälle ($p$% aller Fälle) den (unbekannten) wahren Mittelwert $\mu$ der Variablen $x$ in der Gesamtpopulation.*

# Konfidenzintervalle bei wachsendem Stichproben-Umfang

• **Beispiel 2.12.:** *BMI-Messung in der LIFE-ADULT-Studie (Baseline), vergleiche Beispiel 1.3.: Wir ziehen drei "Stichproben aus der Stichprobe" und bilden drei 95-%-Konfidenzintervalle zu den Stichproben-Mittelwerten.*

$\mathcal{S}^{(1)}$ *mit* $N = 200$, $\widehat{\mu}^{(1)} = 27.10$, $\widehat{\sigma}^{(1)} = 5.26$ *:*

$$\mathcal{I}^{(1)} = \left[\widehat{\mu}^{(1)} - 1.96 \cdot \widehat{\sigma}^{(1)}/\sqrt{200}\ ,\ \widehat{\mu}^{(1)} + 1.96 \cdot \widehat{\sigma}^{(1)}/\sqrt{200}\right] = [26.37\,,\,27.83]\,.$$

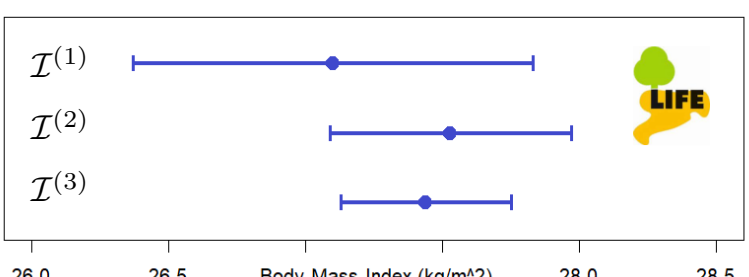
$\mathcal{S}^{(2)}$ *mit* $N = 500$, $\widehat{\mu}^{(2)} = 27.53$, $\widehat{\sigma}^{(2)} = 5.03$ *:*

$$\mathcal{I}^{(2)} = \left[\widehat{\mu}^{(2)} - 1.96 \cdot \widehat{\sigma}^{(2)}/\sqrt{500}\ ,\ \widehat{\mu}^{(2)} + 1.96 \cdot \widehat{\sigma}^{(2)}/\sqrt{500}\right] = [27.09\,,\,27.97]\,.$$

$\mathcal{S}^{(3)}$ *mit* $N = 1000$, $\widehat{\mu}^{(3)} = 27.44$, $\widehat{\sigma}^{(3)} = 5.07$ *:*

$$\mathcal{I}^{(3)} = \left[\widehat{\mu}^{(3)} - 1.96 \cdot \widehat{\sigma}^{(3)}/\sqrt{1000}\ ,\ \widehat{\mu}^{(3)} + 1.96 \cdot \widehat{\sigma}^{(3)}/\sqrt{1000}\right] = [27.13\,,\,27.75]\,.$$

• *Mit wachsendem Stichproben-Umfang $N$ werden die 95-%-Konfidenzintervalle (gesetzmäßig) immer enger, sind jedoch nicht ineinander verschachtelt.*



# Konfidenzintervalle zu abnehmenden Prozentzahlen

• **Beispiel 2.13.:** *(vergleiche Beispiele 1.3. und 2.12.) Zum Stichproben-Mittelwert* $\widehat{\mu}^{(1)}$ *für die Stichprobe* $\mathcal{S}^{(1)}$ *aus dem vorigen Beispiel bilden wir das 95-%-, das 90-%- und das 85-%-Konfidenzintervall.*

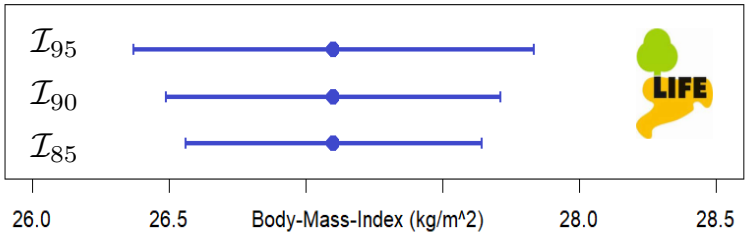
$\mathcal{S}^{(1)}$ *mit* $N = 200$, $\widehat{\mu}^{(1)} = 27.10$, $\widehat{\sigma}^{(1)} = 5.26$ *:*

$$\mathcal{I}_{85} = \left[\, \widehat{\mu}^{(1)} - 1.440 \cdot \widehat{\sigma}^{(1)}/\sqrt{200} \;,\; \widehat{\mu}^{(1)} + 1.440 \cdot \widehat{\sigma}^{(1)}/\sqrt{200} \,\right] = [\,26.56\,,\,27.64\,]\,.$$

$$\mathcal{I}_{90} = \left[\, \widehat{\mu}^{(1)} - 1.645 \cdot \widehat{\sigma}^{(1)}/\sqrt{200} \;,\; \widehat{\mu}^{(1)} + 1.645 \cdot \widehat{\sigma}^{(1)}/\sqrt{200} \,\right] = [\,26.49\,,\,27.71\,]\,;$$

$$\mathcal{I}_{95} = \left[\, \widehat{\mu}^{(1)} - 1.960 \cdot \widehat{\sigma}^{(1)}/\sqrt{200} \;,\; \widehat{\mu}^{(1)} + 1.960 \cdot \widehat{\sigma}^{(1)}/\sqrt{200} \,\right] = [\,26.37\,,\,27.83\,]\,;$$

• *Mit abnehmender Prozentzahl* $p$ *werden die* $p$*-%-Konfidenzintervalle (gesetzmäßig) immer enger und sind dabei ineinander verschachtelt.*



# Konfidenzintervalle zu anderen Kenngrößen

• **Grundprinzip:** *Wenn man in einer Stichprobe eine Kenngröße $\widehat{k}$ ermittelt und damit den (unbekannten) Wert der Kenngröße $k$ in der Gesamtpopulation schätzt, kann man (fast) immer ein $p$-%-Konfidenzintervall für $\widehat{k}$ berechnen. Gleiche Interpretation wie bisher: "In the long run" überdecken $p$% dieser Intervalle den wahren Wert $k$.*

• **Rechenvorschrift:** *Die Formeln fürs Konfidenzintervall hängen davon ab, wie das Gegenstück zu Lehrsatz 2.9. für $\widehat{k}$ und $k$ aussieht: Welcher Verteilung unterliegen die Kenngrößen $\widehat{k}$, wenn man immer wieder unabhängige Stichproben vom Umfang $N$ zieht?*

• **Definition 2.14.:** *(95-%-Konfidenzintervall für die relative Häufigkeit) Für eine kategoriale, binäre Variable $x$ in der Gesamtpopulation $\mathcal{P}$ sei eine Stichprobe vom Umfang $N$ gegeben, in der $x$ = "TRUE" mit der relativen Häufigkeit $0 < \widehat{p} < 1$ auftritt. Es gelte $\widehat{p} \cdot (1 - \widehat{p}) \cdot N \geqslant 9$. Dann erklären wir das 95-%-Konfidenzintervall für $\widehat{p}$ durch*

$$\left[\, \widehat{p} - 1.96 \cdot \sqrt{\frac{\widehat{p} \cdot (1 - \widehat{p})}{N}} \,,\, \widehat{p} + 1.96 \cdot \sqrt{\frac{\widehat{p} \cdot (1 - \widehat{p})}{N}} \,\right].$$

# Nochmals der Body-Mass-Index

• **Zurück zu Beispiel 1.3.:** *LIFE-ADULT-Studie (Baseline):*
*Gesamtzahl der Probanden mit BMI-Messung: 10 289, davon Anzahl der Probanden mit BMI $\geqslant$ 25: 6 716, das sind 65.3 Prozent. Gesucht ist das zugehörige 95-%-Konfidenzintervall.*

• *Die Eigenschaft "BMI $\geqslant$ 25" ist eine kategoriale, binäre Variable $x$. Dafür haben wir eine Stichprobe $\mathcal{S}$ mit $N = 10\,289$ und relativer Häufigkeit $\widehat{p} = 6\,716 \,/\, 10\,289 = 0.653$ für $x =$ "TRUE".*

• *Test der Voraussetzung:* $\widehat{p} \cdot (1 - \widehat{p}) \cdot N = 0.653 \cdot 0.347 \cdot 10\,289 = 2331.395 \geqslant 9$

• *Also lautet das 95-%-Konfidenzintervall für $\widehat{p}$:*

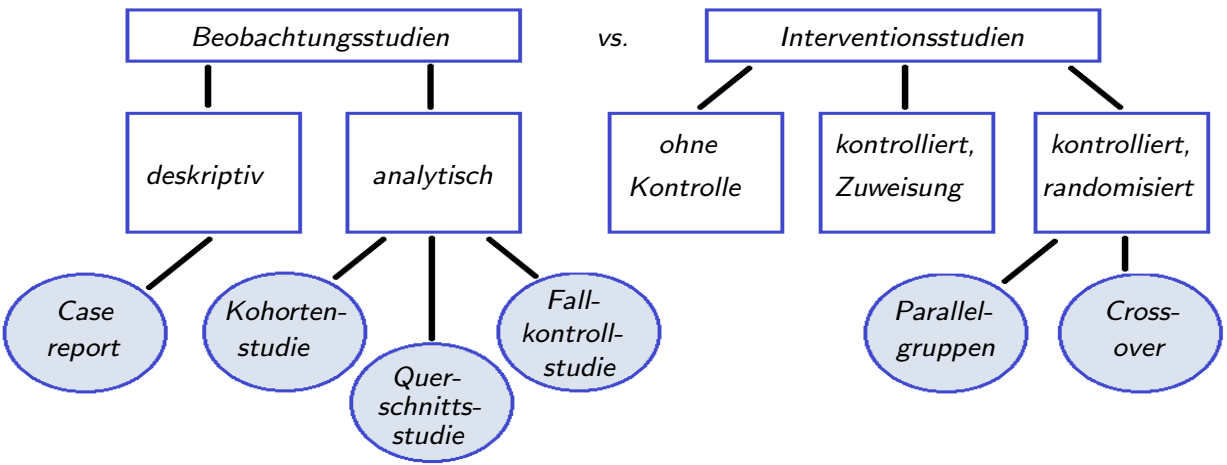
$$\left[\, \widehat{p} - 1.96 \cdot \sqrt{\frac{\widehat{p} \cdot (1 - \widehat{p})}{10\,289}} \,,\, \widehat{p} + 1.96 \cdot \sqrt{\frac{\widehat{p} \cdot (1 - \widehat{p})}{10\,289}} \,\right] = [\, 0.644 \,,\, 0.662 \,] \,,$$

*das entspricht [ 64.4 Prozent , 66.2 Prozent ].*

• **Konfidenzintervalle für weitere Kenngrößen:** *siehe Anmerkungen zur Vorlesung*

# Studien als "statistische Experimente"

- **Grundlegende Studiendesigns**




- *Kontrollierte Studien können offen oder verblindet sein (einfach, zweifach, dreifach).*
- *Bei der Planung und Auswertung einer Studie werden i.a. alle vier Grundkonzepte der Statistik benötigt.*

# Grundlegende Studiendesigns (I)

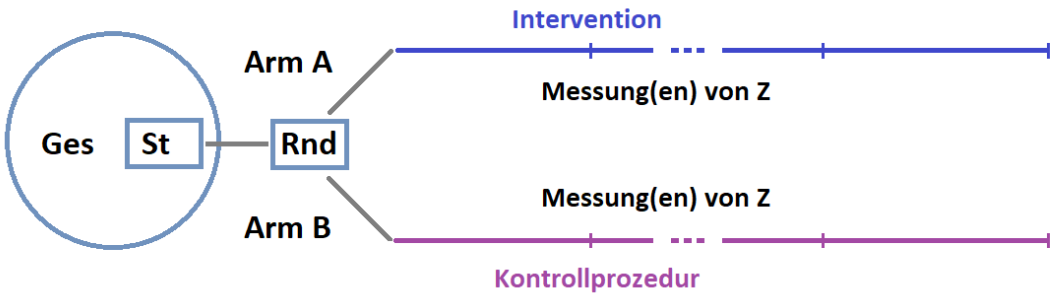
• **Querschnittsstudie:** *Aus der Gesamtpopulation wird eine möglichst repräsentative Stichprobe ausgewählt. Das Zielmerkmal und die Begleitmerkmale ("Expositionen") werden zu einem einzigen Zeitpunkt als "Momentaufnahme" bestimmt.*

• **Kohortenstudie:** *Aus der Gesamtpopulation wird als Stichprobe eine sog. Kohorte ausgewählt: Personen, die sich in ähnlichen zeitlichen oder örtlichen Lebensumständen befinden (gleiches Geburtsjahr, gleiches Schuljahr, gleicher Wohnort, ...). Zielmerkmal und Begleitmerkmale werden normalerweise zu mehreren aufeinanderfolgenden Zeitpunkten bestimmt ("prospektives, longitudinales Design").*

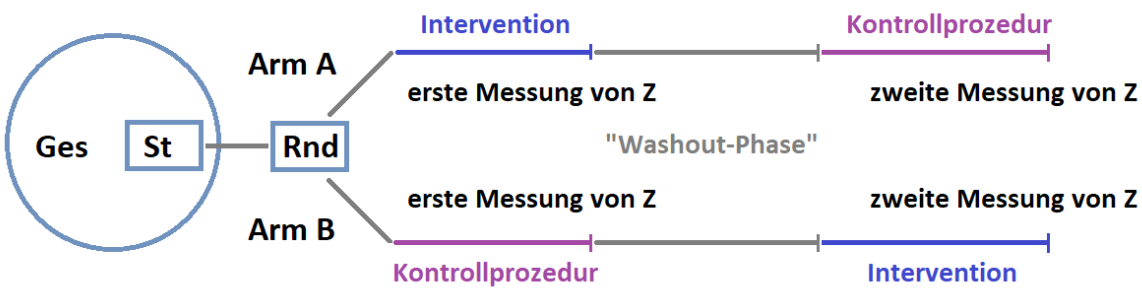
• **Fallkontrollstudie:** *Retrospektives Design: Die verfügbaren Probanden werden nach einem Ja-Nein-Begleitmerkmal sortiert. Jedem Probanden der Ja-Gruppe ("Cases") wird ein Proband der Nein-Gruppe ("Controls") fest zugeordnet, der sich in allen anderen Begleitmerkmalen möglichst wenig von dem betreffenden Case-Probanden unterscheidet ("Matching"). Das Zielmerkmal wird in beiden Gruppen bestimmt; die statistische Auswertung muß die feste Paar-Zuordnung der Probanden berücksichtigen.*

# Grundlegende Studiendesigns (II)

- *Randomisierte Interventionsstudie, Parallelgruppendesign (vgl. Beispiel 2.15. unten)*




- *Randomisierte Interventionsstudie, Crossover-Design*



# Elemente der Studienplanung: Vom Problem zur Lösung

**Wie wird ein "statistisches Experiment" geplant?**

*(1) Humanwissenschaftliche Fragestellung und Studienidee*

*(2a) Praktische Planung: umfaßt z.B.*

*— Zielgröße und deren Messung ("primärer Endpunkt")*
*— Studiendesign (z.B. Anzahl der "Arme")*
*— Studienmaßnahmen ("Intervention")*
*— zahlreiche ethische und juristische Fragen*

*(2b) Statistische Planung: umfaßt z.B.*

*— Übersetzung der Frage bzw. Hypothese in ein mathematisch-statistisches Modell*
*— Zielmerkmal und wesentliche Einflußfaktoren ("Kovariaten")*
*— notwendiger Mindestumfang der Stichproben ("Fallzahlplanung")*

# Vom Problem zur Lösung (II)

*(3) Durchführung:*

— *Rekrutierung und ggf. Randomisierung der Probanden, d.h. praktisches Ziehen der Stichprobe(n)*
— *Studienintervention*
— *Erfassung der Daten, laufendes Monitoring*

*(4) Auswertung der Daten:*

— *Datenbereinigung*
— *eigentliche statistische Analyse, z.B. ein Hypothesentest*
— *Darstellung und korrekte Visualisierung der Ergebnisse*

*(5) Interpretation der Ergebnisse im Kontext des jeweiligen Fachgebietes*

**Vorbild:** *Orientierung am vergleichsweise hohen Standard klinischer Studien*

# Beispiel Studienplanung: ausgeweiteter Sportunterricht

**Beispiel 2.15. "Schülersportstudie"** [Walther et al. 2011]

- *(1) Studienidee und Fragestellung: Vorgabe im Lehrplan sind zwei Sportstunden pro Schulwoche. Würde sich eine Ausweitung des Sportunterrichts auf fünf Sportstunden pro Schulwoche (also täglich) positiv auf die Fitneß der Schüler auswirken?*
- *(2a) Praktische Hauptfrage: Wie kann man Fitneß quantifizieren?*

- *Spiroergometrischer Test:*
  *Person kommt aufs Laufband, die Belastung wird phasenweise gesteigert, bis man außer Atem kommt. Die Sauerstoffaufnahme wird mit Spirometer kontrolliert.*
  *Objektive Meßgröße ist die maximale $O_2$-Aufnahme, noch aufs Körpergewicht normiert (Einheit: $ml/(min \cdot kg)$): wer fitter ist, atmet tiefer.*

# Beispiel Studienplanung (II)

**Studiendesign** *(2a)*

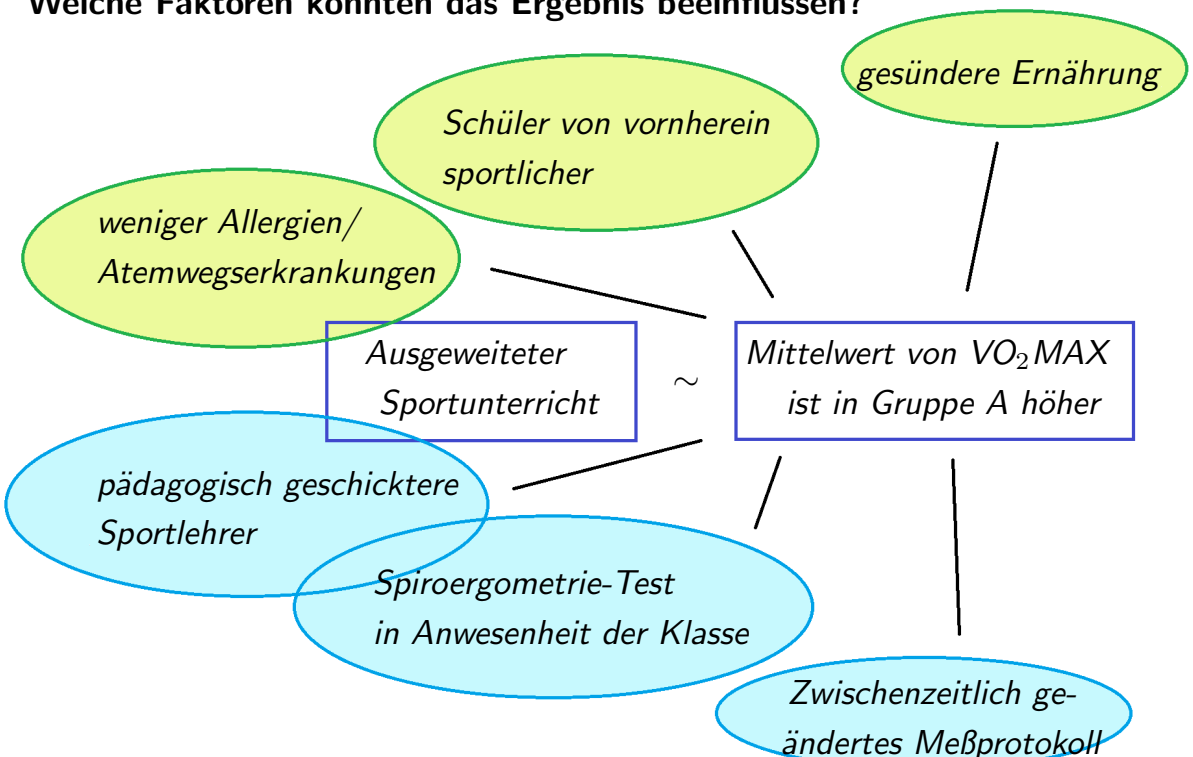
- *zufällige Zuordnung von Schülern zur Interventions- (A) oder Kontrollgruppe (B)*
- *Maßnahmen: zwei Schuljahre lang fünf Sportstunden pro Woche (A) oder zwei Sportstunden pro Woche (B)*
- *Primärer Endpunkt: maximale $O_2$-Aufnahme $VO_2MAX$, aufs Körpergewicht normiert, Zuwachs gemessen nach zwei Jahren*

**Ein Gedankenexperiment** *(2a)*

- *Angenommen, wir beobachten beim primären Endpunkt einen Gruppenunterschied, z.B. erhöhten Mittelwert von $VO_2MAX$ in der Gruppe A.*
- *Wunsch-Interpretation: Ausgeweiteter Sportunterricht steigert die Fitneß der Schüler*
- *Wissenschaftliche Integrität verlangt jedoch, möglichst alle wesentlichen Faktoren zu berücksichtigen, die das Ergebnis beeinflussen könnten. Welche sind das?*

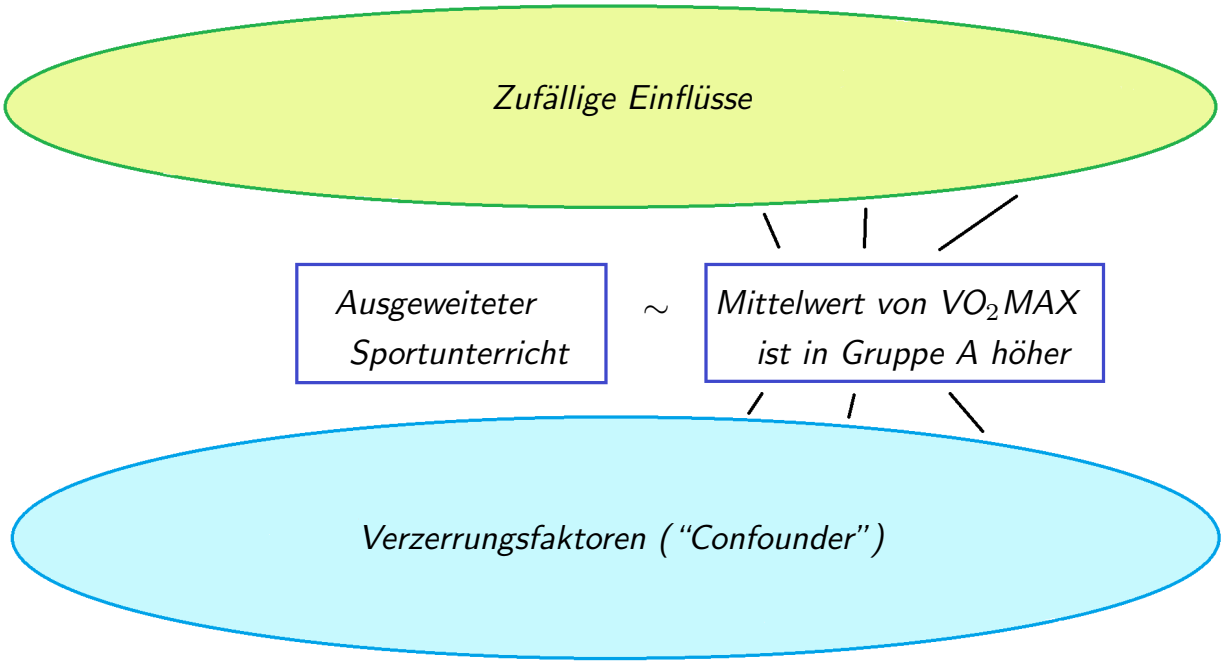
# Beispiel Studienplanung (III)

**Welche Faktoren könnten das Ergebnis beeinflussen?**



# Beispiel Studienplanung (III)

**Welche Faktoren könnten das Ergebnis beeinflussen?**



# Beispiel Studienplanung (IV)

**Wie behandeln wir zufällige Einflüsse?**

• **Randomisierung** *(2a), (2b)*
*Zufall als Helfer: zufällige Zuordnung zu Interventions- oder Kontrollgruppe erzeugt annähernd gleiche Ausgangssituation für beide Gruppen*

• **Fallzahl** *(2b)*
*beide Gruppen als Stichproben müssen einerseits genug Variation aufweisen und andererseits genug Information enthalten*

*beides zusammen* $\Longrightarrow$ *Strukturgleichheit*

• **Statistische Auswertung** *(2b), (4)*
*— Vorgehen nach Grundkonzept 3 und 4: Hypothesentest, Modellgleichung für Einflüsse der Gruppenzuordnung und der wichtigsten Kovariaten auf den primären Endpunkt*
*— prinzipielle Unsicherheit der Auswertung kann zahlenmäßig beschrieben werden*

# Beispiel Studienplanung (V)

**Wie schalten wir Confounder aus?**

• **Sorgfältige Planung und Durchführung der Studie** *(2a), (3)*
*sollte absichern:*

$\Longrightarrow$ *Behandlungsgleichheit*

$\Longrightarrow$ *Beobachtungsgleichheit*

$\Longrightarrow$ *Auswertungsgleichheit*

• **Statistische Auswertung** *(2b), (4)*
*Modelle auf mögliche Verzerrungsfaktoren adjustieren*

**Fazit:** *Mit prospektiven, randomisierten, kontrollierten Interventionsstudien hat man ein methodisches Werkzeug geschaffen, das Verzerrungsfaktoren minimiert.*

# Zum Schluß: Was wurde realisiert?

- *"Cluster-Randomisierung" (klassenweise)* $\Longrightarrow$ *Balance der Merkmale prüfen!*

| *Merkmal* | *Gruppe A* | *Gruppe B* |
|---|---|---|
| *Anzahl Teilnehmer* | $N = 142$ *(6 Klassen)* | $N = 90$ *(5 Klassen)* |
| *Anteil Jungen* | *53 %* | *49 %* |
| *Alter (Jahre)* | $11.6 \pm 0.6$ | $11.7 \pm 0.5$ |
| *BMI* $(kg/m^2)$ | $18.3 \pm 2.8$ | $18.0 \pm 2.6$ |
| *Herzfrequenz in Ruhe (1/min)* | $100 \pm 15$ | $101 \pm 11$ |

- *Ausgangsposition: Mittelwert von* $VO_2MAX$ *zu Baseline*

$39.6 \pm 13.7$ *ml/(min* $\cdot$ *kg) (Gruppe A) /* $38.8 \pm 10.2$ *ml/(min* $\cdot$ *kg) (Gruppe B)*

- **Ergebnis:** *mittlerer Zuwachs nach 2 Jahren*

$\Delta = +10.1$ *ml/(min* $\cdot$ *kg) (Gruppe A) /* $\Delta = +7.9$ *ml/(min* $\cdot$ *kg) (Gruppe B)*