Grundlagen der Diagnostik

Lerneinheit 10

Einzelfalldiagnostik II

1. Überblick über wichtige Aspekte bei der Einzelfalldiagnostik
2. Berechnung eines frequentistischen Konfidenzintervalls
3. Beurteilung von Normstichproben
4. Normwerte und Prozentränge
5. Verbalisierung des Konfidenzintervalls & Rückmeldung
6. Beispiel
7. **Vorwissen berücksichtigen mit Bayesianischer Statistik**

Lernziele

Bayesianische Statistik als alternativer statistischer Ansatz zur frequentistischen Statistik:

- Frequentistisches 95% KI: Wenn man unendliche viele Zufallsstichproben ziehen würde (bzw. in der Einzelfalldiagnostik eine Person unendlich oft testen würde, ohne dass diese sich an die vorherigen Testungen erinnert), enthalten 95% aller gebildeten KIs den wahren Wert.
- Bayesianisches 95% KI: Gegeben meiner Vorannahme (Prior), befindet sich der wahre Wert mit 95% Wahrscheinlichkeit zwischen den errechneten Intervallgrenzen.

→ Ermöglicht den Einbezug von Vorwissen, was in der Einzelfalldiagnostik sehr praktisch sein kann!

Grundprinzipien Bayesianische Statistik

„Im Satz von Bayes wird eine bestehende Erkenntnis über den zu untersuchenden Parameter (die *A-priori* Verteilung, kurz **Prior**) mit den neuen Erkenntnissen aus den Daten kombiniert (**Likelihood**), woraus eine neue, verbesserte Erkenntnis (*A-posteriori* Verteilung, kurz **Posterior**) resultiert.“ (nach Wikipedia)

Posterior ∝ Likelihood x Prior

Lies: „Die Posterior ist proportional (∝) zum Produkt aus Likelihood und Prior“

Grundprinzipien Bayesianische Statistik

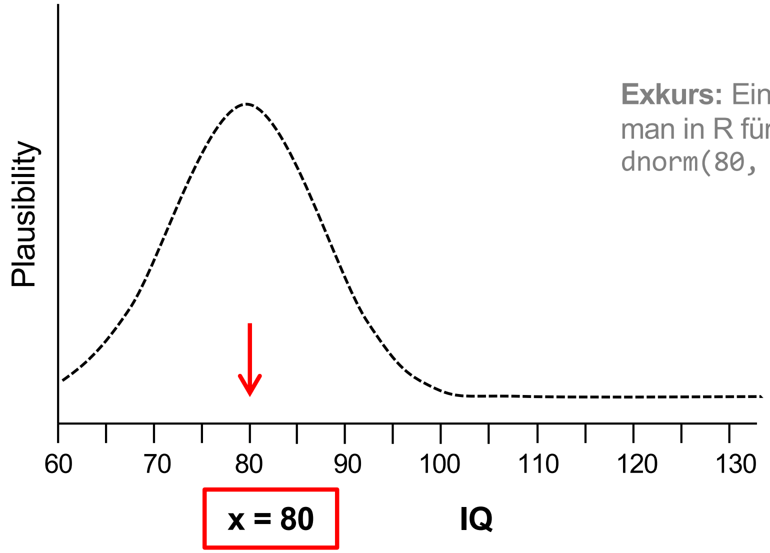
Beispiel:

- Bob hat einen Testwert von 80 IQ-Punkten im Intelligenztest (x = 80)
- Testwertverteilung: X ~ N(IQ, $\sigma_x^2 \cdot (1 - Rel)$)
- Wobei in der Formel…
  - „IQ“ für die wahre Intelligenz von Bob steht die uns eigentlich interessiert
  - „Rel“ für die Reliabilität des Testwerts steht, die als bekannt vorausgesetzt wird
  - „σ_x^2“ für die Varianz des Testwerts steht, in diesem Beispiel 225 wegen IQ-Normwerten
- Die Testwertverteilung entspricht wieder dem vereinfachten Testmodell, welches wir auch zur Berechnung der approximativen frequentistischen Konfidenzintervalle herangezogen haben (siehe LE9).

Grundprinzipien Bayesianische Statistik

Likelihood

- Verteilungsfunktion, die einer beobachteten Variable zugeordnet ist
  → „Wie plausibel ist es die tatsächlich vorliegenden Daten zu beobachten gegeben bestimmter Werte für die Modellparameter?"
- Plausibilität einen Testwert von x = 80 zu beobachten in Abhängigkeit von verschiedenen wahren IQ-Werten



Exkurs: Einen einzelnen Wert auf der Likelihood Kurve könnte man in R für das vorliegende Beispiel wie folgt berechnen:

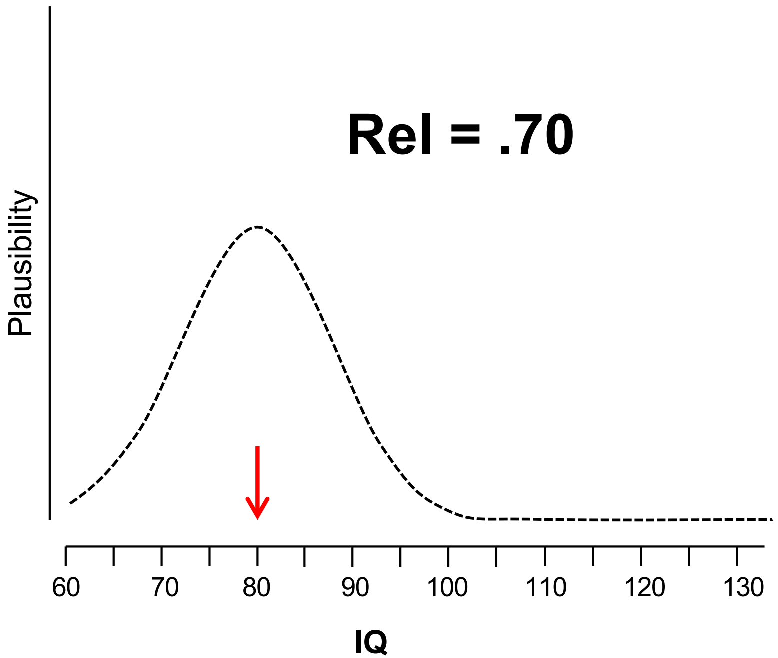
```
dnorm(80, mean = IQ, sd = sqrt(225*(1-0.7)))
```

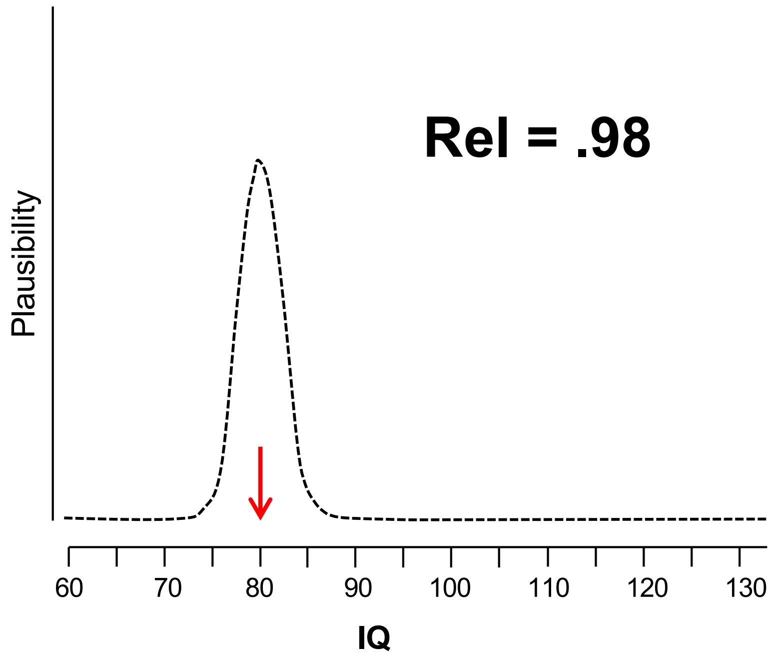
(vgl. Wagenmakers, Morey, & Lee, 2016; McElreath, 2020: Kap. 2+3)

Grundprinzipien Bayesianische Statistik

Likelihood

- Formalisiert das Wissen, das wir durch die Daten erlangen
- Hängt von der Reliabilität ab: Je unreliabler der Test, desto breiter ist die Likelihood (d.h. desto plausibler sind wahre Werte weit weg vom Testwert)

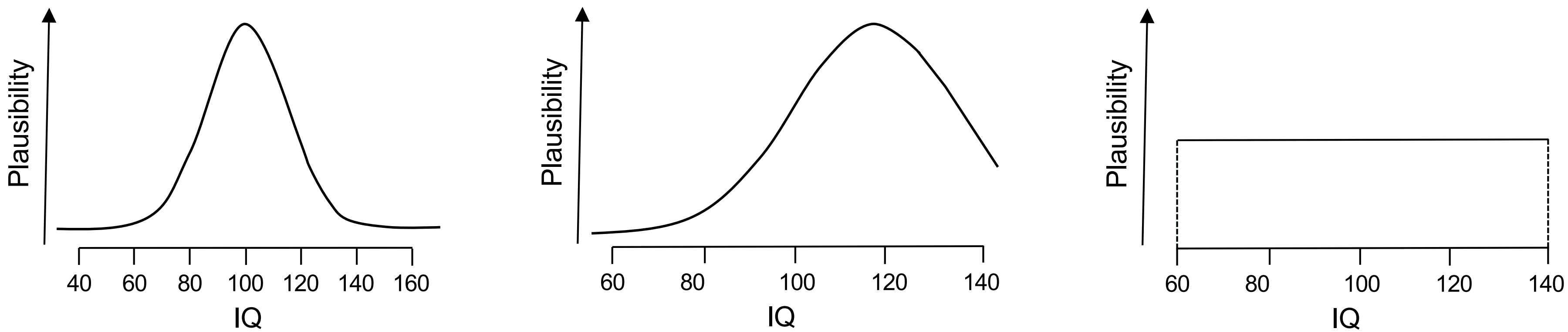




Grundprinzipien Bayesianische Statistik

Prior

- Die Prior-Verteilung ist eine Vorannahme, die definiert, wie plausibel mögliche wahre Werte in der Population a priori sind (d.h., bevor man die Daten beobachtet hat)
- Diese Vorannahmen können sehr vage sein („uninformativ"), oder substantielles Vorwissen enthalten („informativ")
- Je sicherer man sich vorher schon ist, desto schmalgipfliger ist die Prior um den erwarteten Wert herum (→ geringere Varianz der Verteilung)

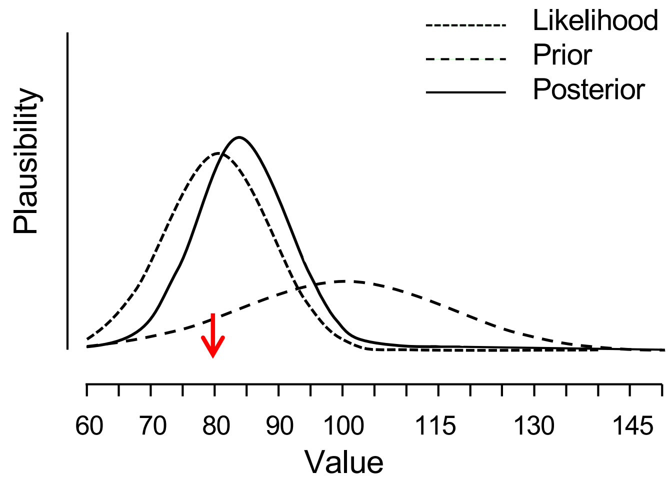


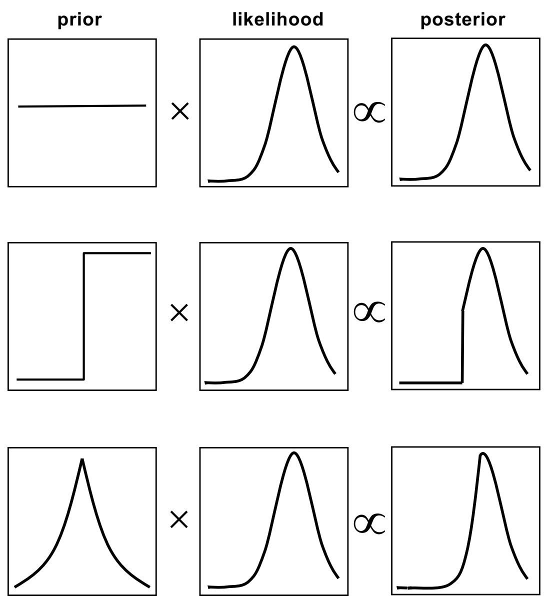
Grundprinzipien Bayesianische Statistik

Posterior

- **„Updating process“:**
  → Wir aktualisieren unser Vorwissen mit Hilfe der erhobenen Daten
  → daraus resultiert eine (verbesserte) Wahrscheinlichkeitsverteilung

- Die Posterior-Verteilung quantifiziert die Plausibilität möglicher wahrer Werte *nachdem* man die Testwerte beobachtet hat



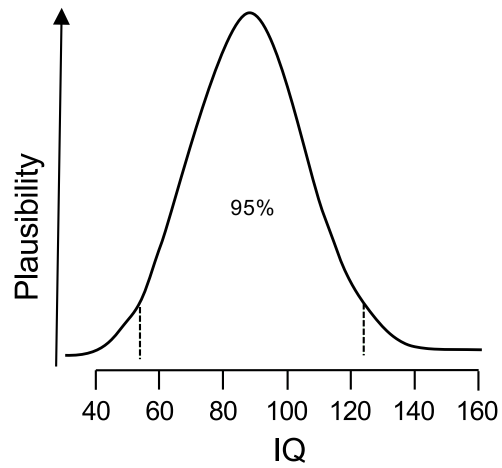


Grundprinzipien Bayesianische Statistik

Highest Densitiy Interval (HDI)

- Ein bestimmtes Intervall definierter Masse der Posterior stellt das **bayesianische KI-Äquivalent** dar
  - z.B. 95% der Fläche → 95% HDI

- Unter Annahme der Prior kann man mit diesen bayesianischen KIs dann Interpretationen über den wahren Wert vornehmen
  - z.B. „Der wahre Wert liegt mit 95% Wahrscheinlichkeit zwischen...“



Exkurs: Im Gegensatz zu HDIs gibt es auch „equal-tailed intervals“ (ETIs). Die genaue Definition eines HDI haben wir hier nicht besprochen und ist in der Praxis meist irrelevant. HDIs und ETIs unterscheiden sich stark nur bei sehr schiefen Posteriori Verteilungen.

McElreath, 2018: Kap. 1-3

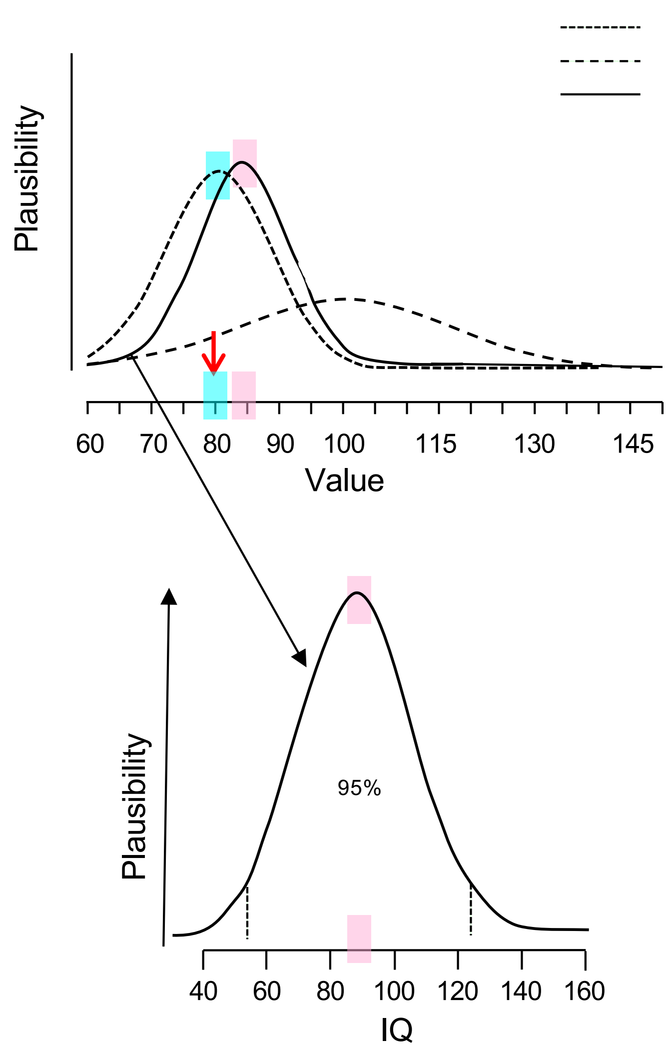
Grundprinzipien Bayesianische Statistik

Bester Punktschätzer

- Frequentismus: beobachteter Wert ist die beste Punkt-Schätzung für den wahren Wert

- Bayes: Modus der Posterior d.h., der Gipfel der Verteilung) ist die beste Punkt-Schätzung für den wahren Wert

Hinweis: Alternativ zum Modus wird häufig auch der Erwartungswert oder der Median der Posterior als bayesianischer Punkt-Schätzer verwendet.

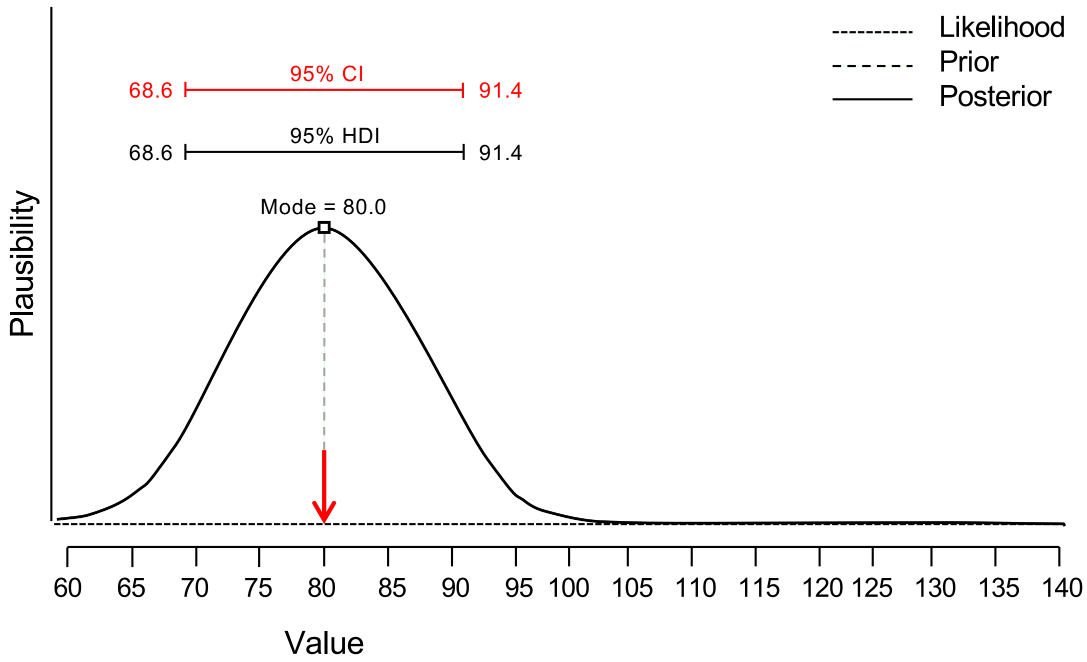


(vgl. Wagenmakers, Morey, & Lee, 2016; McElreath, 2018: Kap. 1-3)

Beispiel 1: kein Vorwissen

- Testwert = 80 IQ-Punkte, Reliabilität = .85
- flache („uninformative“) Prior

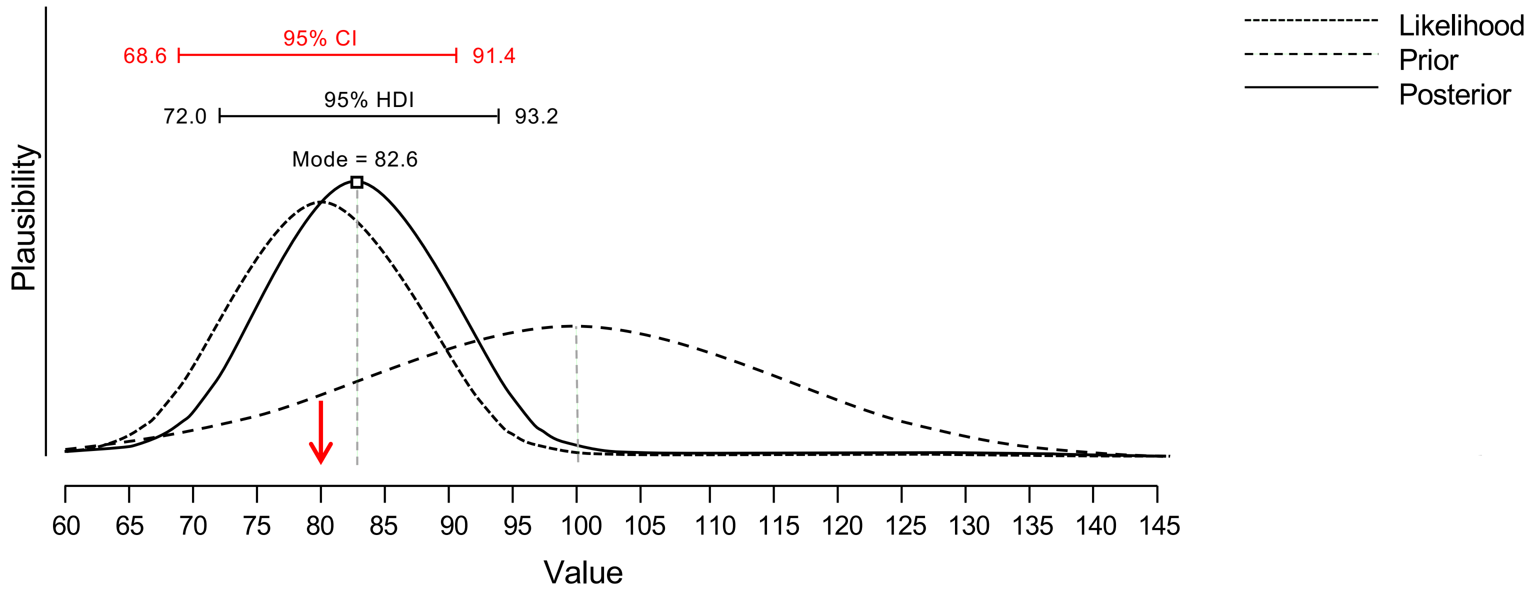
  → Jeder IQ-Wert von –∞ bis +∞ ist a priori gleich wahrscheinlich



→ Frequentistisches KI und Bayesianisches HDI (und Punktschätzer) fallen zusammen

Beispiel 2: Vorwissen vorhanden

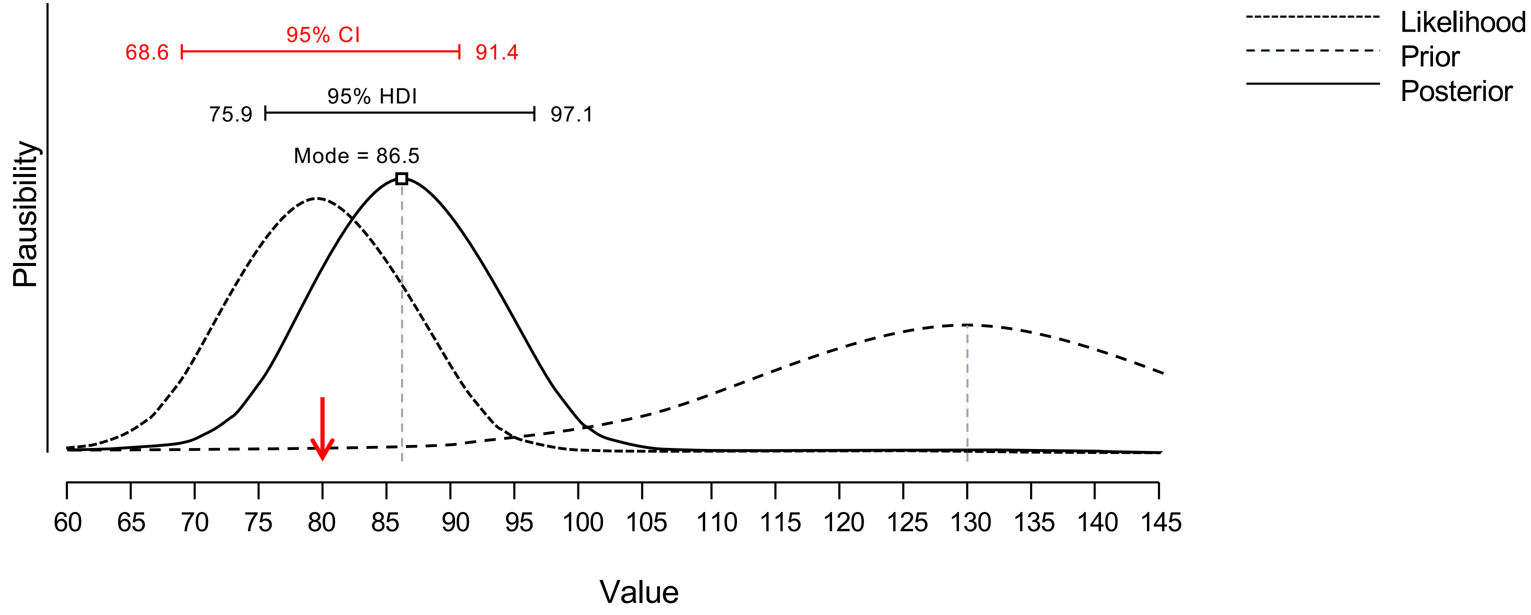
- Testwert = 80 IQ-Punkte, Reliabilität = .85
- Prior: IQ ~ $N(100, 15^2)$ → die Prior ist normalverteilt mit einem Mittelwert von 100 und einer Varianz von 15^2 bzw. SD von 15



→ Der wahrscheinlichste Bayes-Punktschätzer ist *nicht* der Testwert von 80, sondern liegt bei 82.6

Beispiel 3: Spezifisches Vorwissen vorhanden I

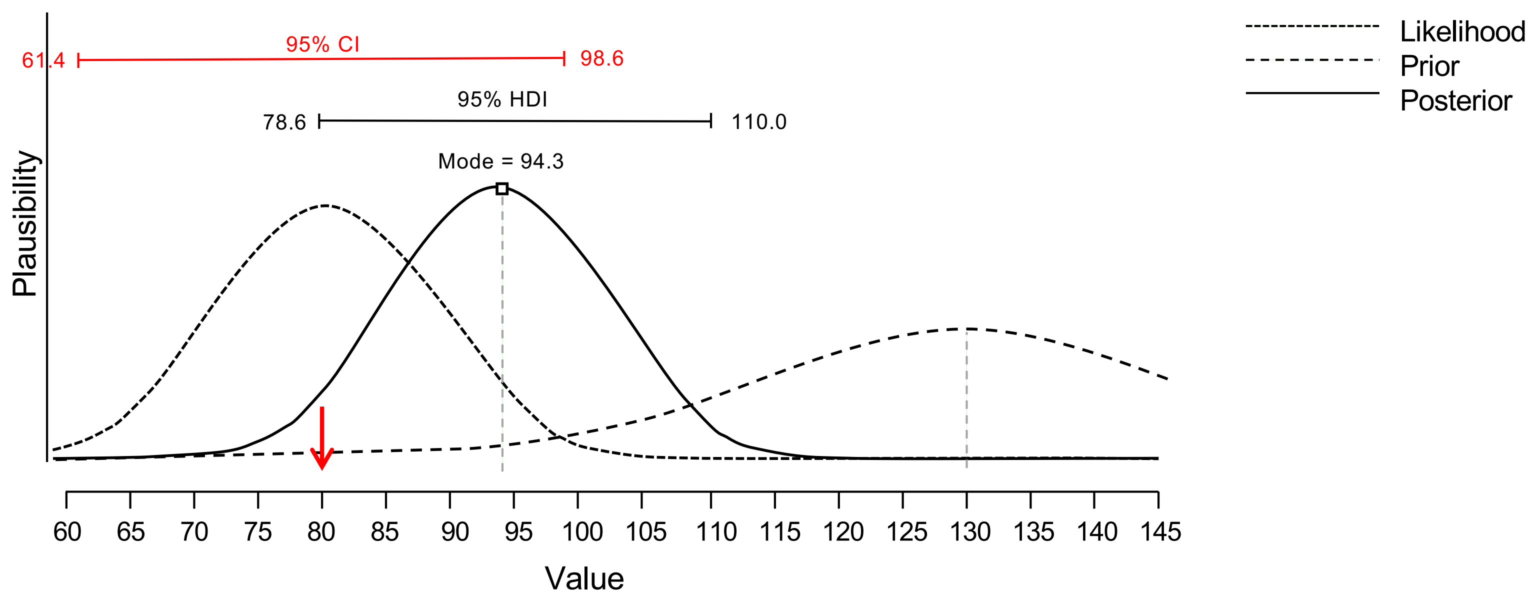
- Testwert = 80 IQ-Punkte, Reliabilität = .85
- Prior: IQ ~ $N(130, 15^2)$ → die Prior ist normalverteilt mit einem Mittelwert von 130 und einer Varianz von 15^2 bzw. SD von 15 (Annahme, weil wir wissen, dass die begutachtete Person aus der Population Mathematikstudierender kommt)



→ Der wahrscheinlichste Bayes-Punktschätzer ist liegt in diesem Fall bei 86.5

Beispiel 3: Spezifisches Vorwissen vorhanden II

- Testwert = 80 IQ-Punkte, **Reliabilität = .60**
- Prior: IQ ~ $N(130, 15^2)$ → die Prior ist normalverteilt mit einem Mittelwert von 130 und einer Varianz von 15^2 bzw. SD von 15 (wie auf der Folie vorher)



→ Je unreliabler das Messinstrument, desto stärker der Einfluss der Prior und desto breiter ist das bayesianische HDI (und auch das frequentistische KI).

Fazit: Bayes in der Einzelfalldiagnostik

- Das Bayes-Theorem stellt eine „Berechnungsvorschrift“ dar, wie man Vorwissen mit neuen Daten verrechnen kann. Das bisher Bekannte (die Prior) wird mit den neuen Daten aktualisiert, um so den aktualisierten Wissensstand zu erhalten (die Posterior).
- Damit kann etwas formalisiert werden, was in der Praxis ohnehin implizit gemacht wird: Urteile mithilfe externer Informationen anzupassen, abhängig davon, wie stark mein Vertrauen in meine diagnostische Messung ist.
- Wenn man ein reliables Messinstrument hat, kann man dem Messwert relativ stark vertrauen, und das Vorwissen ist relativ irrelevant
- Wenn das Messinstrument schlecht ist, kann es ratsam sein, vorhandenes Wissen einzubeziehen:
  - Extremfall: Das Messinstrument liefert im Grunde nur Rauschen. Dann sollte man rationaler Weise nur das Vorwissen nutzen (unter der Voraussetzung, dass das Vorwissen mehr als nur Rauschen kodiert)
  - Sensitivitätsanalysen überprüfen die Ergebnisse bei Heranziehen verschiedener (plausibler) Priors

Fazit: Bayes in der Einzelfalldiagnostik

- Ob die Berücksichtigung von **spezifischem** Vorwissen für meinen diagnostischen Fall sinnvoll ist, hängt vom Kontext ab
- Dabei stellen sich ähnliche Fragen, wie bei der Auswahl der „interessierenden“ Normstichprobe (siehe LE9)

Es gibt Situationen, da möchte ich…

- **…spezifisches Vorwissen über die von mir getestete Person berücksichtigen, um die individuelle diagnostische Entscheidung zu verbessern**, z.B. wenn ich in einem neuropsychologischen Setting herausfinden will, ob eine Person kognitive Defizite durch eine degenerative Erkrankung aufweist und Vorwissen über ihre frühere Leistungsfähigkeit oder aktuelle Beeinträchtigungen im Alltag vorliegt
- **…für alle getesteten Personen das gleiche Vorwissen verwenden, um alle Personen gleich (“fair“) zu behandeln**, z.B. wenn ich in einem personalpsychologischen Setting Bewerberinnen basierend auf einem standardisierten Leistungstest auswählen will und es nicht vertretbar ist, dass Personen mit dem gleichen Testwert unterschiedliche Beurteilungen bekommen

Fazit: Bayes & Frequentismus

- Unter flachen, uniformativen Priors (Extremfall: Gleichverteilung) sind die Intervallgrenzen von frequentistischen Konfidenzintervallen und bayesianischen HDIs sehr nah beisammen (Extremfall: identisch)
  - Es stellt sich dann aber die Frage, ob die Prior, bei der eine numerische Übereinstimmung entsteht, überhaupt plausibel ist!
- Heißt umgekehrt: Man kann nicht automatisch von einer numerischen Äquivalenz beider Intervalle ausgehen
- Die Abweichung zwischen den Intervallen ist besonders groß, …
  a. wenn der Testwert in einen dünn besiedelten Bereich der Prior fällt
  b. wenn der Test sehr unreliabel ist

Praxis-Tipp: Shiny-App zur Berechnung von HDIs

http://shinyapps.org/apps/Bobs_IQ/

Bayesian Credible Interval for an IQ Test Score

Properties of the test instrument

Test reliability:

0 — 0.8 — 0.99

0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 0.99

Between-person standard deviation of the test scores (e.g. 15 for IQ scores, or 1 for z scores)

15

Obtained test score

Observed test score

0 — 80 — 180

0 18 36 54 72 90 108 126 144 162 180

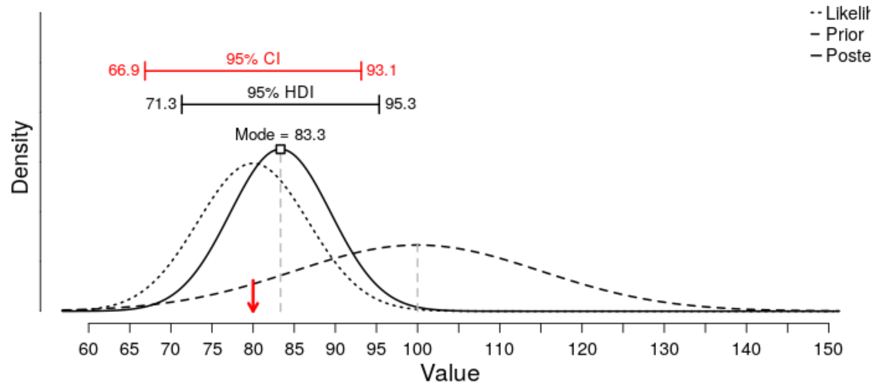
Prior

Mean of prior

100

SD of prior (enter large value, such as 999, for flat prior)

15



This app extends code from Quentin Gronau and Richard Morey.
"Bob's IQ" is an example from the paper: Wagenmakers, E.-J., Morey, R. D., & Lee, M. D. (2016). Bayesian benefits for the pragmatic researcher. Current Directions in Psychological Science, 25, 169–176. doi:10.1177/0963721416643289.
Technical details: The app assumes a known (fixed) variance for the single data point, which is derived from the reliability of the measurement instrument.

Gesamtfazit

- Nicht nur Punktschätzer betrachten, sondern über KI (frequentistisch oder auch bayesianisch) immer die Unsicherheit in der Messung mitberücksichtigen und auch kommunizieren!
- Die Reliabilität des Tests bestimmt u.a. die Präzision der Messung
  - Unreliable Tests können ein so breites KI/HDI ergeben, dass alle Werte von unterdurchschnittlich bis überdurchschnittlich darin enthalten sind (→ d.h. fast alle Werte wären plausibel!)
  - Dies betont noch einmal die Wichtigkeit der Verwendung reliabler Tests in der Einzelfalldiagnostik!

Das war‘s schon fast für heute …

- Nächste Sitzung Fragestunde am Donnerstag um 10.15 per ZOOM
- Link folgt über Moodle
- https://lmu-munich.zoom-x.de/j/69552116495?pwd=LXTLo0bXB8eXboT8BlXHwRObdWtNxq.1

Fragen zur Nachbereitung

- Welche Aspekte sind wichtig bei der Einzelfalldiagnostik? Welche Verfahren zur Testauswertung kann man unterscheiden?
- Wie berechnet man ein frequentistisches KI? Von welchen Faktoren hängt dessen Breite ab?
- Was bedeutet „interessierende" Normstichprobe und welche Eigenschaften der Normstichprobe sind relevant?
- Wie bildet man Normwerte? Was sind Prozentränge? Was muss man hier jeweils bei der Interpretation beachten?
- Wie kann man Testwerte klassifizieren?
- Wie sollte die mündliche/schriftliche Rückmeldung jeweils gestaltet sein?
- Welchen Vorteil hat ein bayesianischer Ansatz in der Einzellfalldiagnostik?
- Nach welchen Grundprinzipien funktioniert Bayes?

Quellen zu Bayes

- Hoekstra, R., Morey, R. D., Rouder, J. N., & Wagenmakers, E. J. (2014). Robust misinterpretation of confidence intervals. *Psychonomic Bulletin & Review*, *21*(5), 1157–1164. http://doi.org/10.3758/s13423-013-0572-3
- McElreath, R. (2020). Statistical rethinking: A Bayesian course with examples in R and Stan. Chapman and Hall/CRC. https://github.com/rmcelreath/stat_rethinking_2024
- Morey, R. D., Hoekstra, R., Rouder, J. N., Lee, M. D., & Wagenmakers, E.-J. (2015). The fallacy of placing confidence in confidence intervals. *Psychonomic Bulletin & Review*, *23*, 103–123. http://doi.org/10.3758/s13423-015-0947-8
- Wagenmakers, E. J., Morey, R. D., & Lee, M. D. (2016). Bayesian Benefits for the Pragmatic Researcher. *Current Directions in Psychological Science*, *25*(3), 169–176. http://doi.org/10.1177/0963721416643289