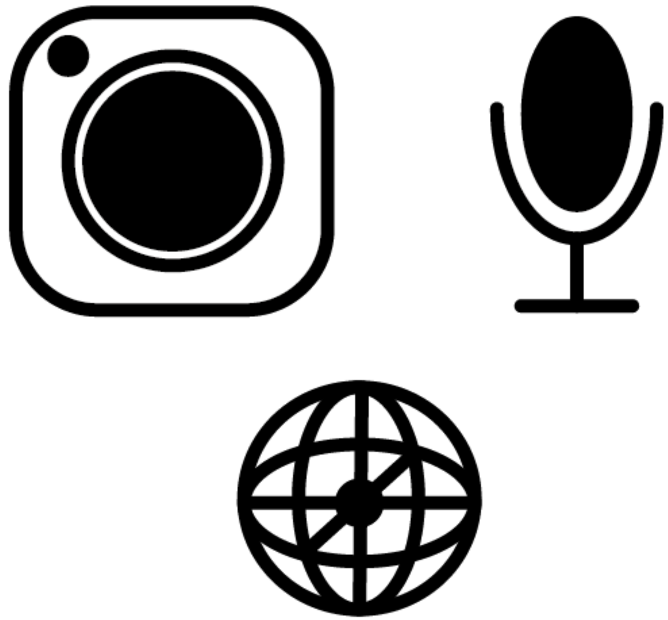


Sensordatenverarbeitung **(ab 13.1.25)**

ENTWICKLUNG UND EVALUA-TION VON SDV-SYSTEMEN (12C)

Deskriptive Analyse: Verteilung

- Information wie Stichprobe um den zentralen Wert verteilt ist
- *Stichprobenvarianz*: $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$
- Warum n-1 statt n? → Unbiased (erwartungstreu, unvoreingenommen) Schätzer der echten Varianz
  - Vergleiche Werte für n und n-1 für kleine n
  - "n" für finite echte Population
  - "n-1" für geschätzte Varianz von Stichproben der echten Population
- Standardabweichung wird oft verwendet
  - Gleiche Einheit wie Stichprobe
  - $s = \sqrt{s^2}$
- Varianz kann nur im Zusammenhang mit Mittelwert interpretiert werden

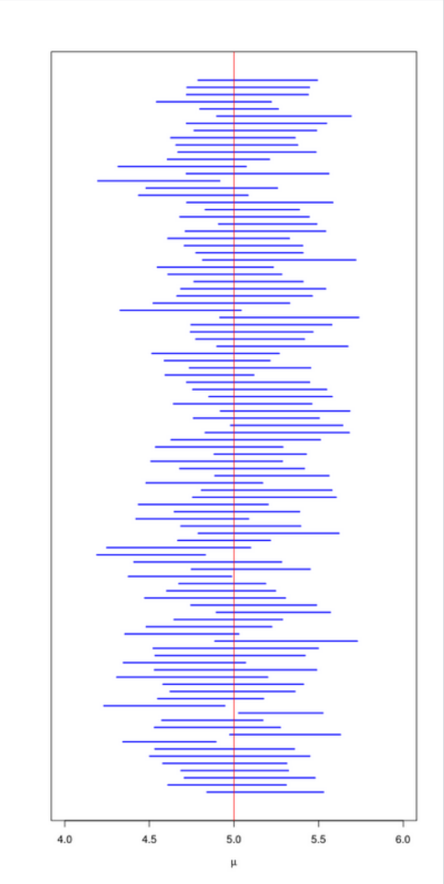
Deskriptive Analyse: Standardfehler

AG MSIS

CSL Cognitive Systems Lab

- Stichproben Mittelwert ist nur eine Approximation des echten Mittelwertes der Population
  - Kann man schätzen wie nah man am echten Wert ist?

- Betrachte Stichproben Mittelwert als Zufallsvariable
- $\bar{X} = \sum_{i=1}^{n} X_i$ / n
  - X_i sind unabhängige, gleichverteilte Zufallsvariablen, jedes X_i repräsentiert ein Element der Stichprobe
  - Standardabweichung von $\bar{X}$ kann berechnet werden, wenn wir die Standardabweichung der X_i kennen
  - Die Standardabweichung von $\bar{X}$ heißt Standardfehler *standard error* (SE)
  - $SE = \sqrt{var(X)/n} = sd(X)/\sqrt{n}$

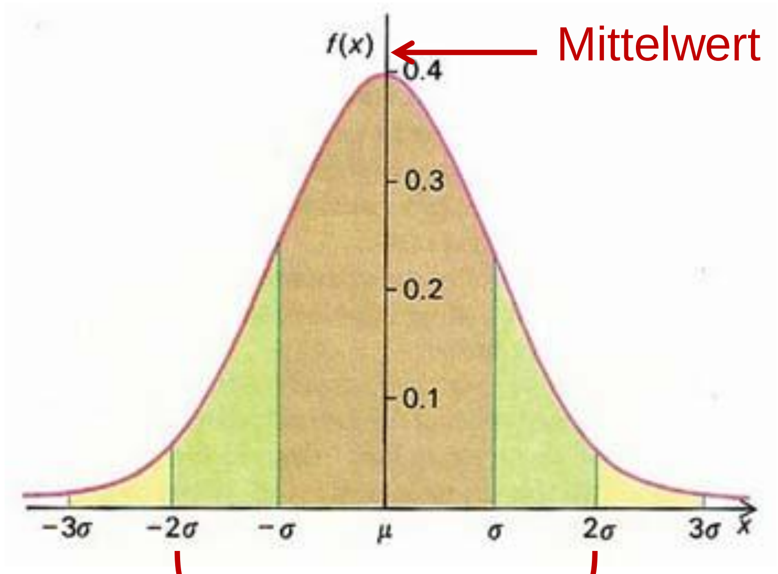


Konfidenzintervalle zum Niveau 95 % für 100 Stichproben vom Umfang 30 aus einer normalverteilten Grundgesamtheit. Davon überdecken 94 Intervalle den exakten Erwartungswert μ = 5; die übrigen 6 tun das nicht.

Wikipedia

Standardfehler

- Stichprobenmittelwert als Summe von unabhängigen, identisch verteilen Zufallsvariablen folgt einer Normalverteilung
  - Zentraler Grenzwertsatz (Central limit theorem)
- Für eine Normalverteilung fallen 68.2% (95.4%) aller möglichen Werte innerhalb von einer (zwei) Standardabweichung vom Mittelwert
  → ***Konfidenzintervall* des Stichprobenmittelwert**



Der echte Mittelwert fällt mit W-keit 95.4% in dieses Intervall

Deskriptive Statistik: Grenzen

- Deskriptive Statistik erlaubt Stichprobenanalyse
- **Standardfehler** gibt Auskunft über die mittlere Abweichung des Mittelwerts einer Stichprobe vom tatsächlichen Mittelwert der Grundgesamtheit

- Aber:
  – Vergleicht nicht zwei verschiedene Stichproben
  – Hilft nicht Hypothese zu evaluieren
  – Erlaubt nicht zu bewerten, ob Fehler in der Stichprobe noch verlässliche Antwort erlaubt

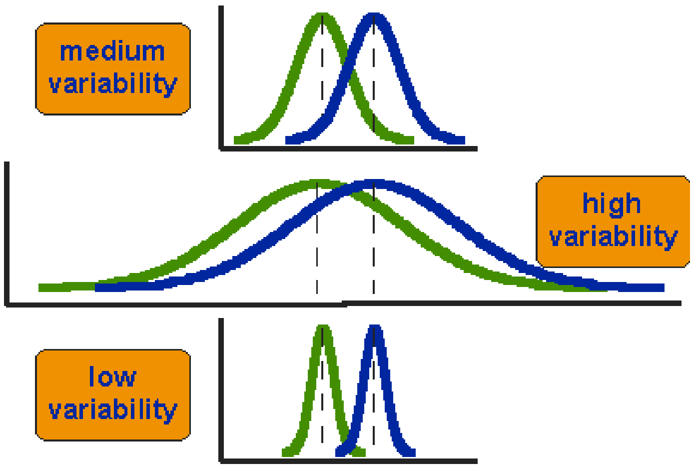
→ Inferenzstatistik
  – Ziel: Schlüsse aus Prozessen ziehen, die Zufallsvariablen enthalten

Überblick

- Entwicklung von SdV-Systemen
- Evaluation von SdV-Systemen
- **Statistische Tests**

Wozu braucht man statistische Tests

- Wir sammeln Daten von zwei Populationen. Sind diese unterschiedlich?
- Errechne Mittelwerte der Stichproben und untersuche, ob sich diese unterscheiden.
- Caveat: Damit schauen wir uns nur **ein Sample** der Population an
  - Unterschiede im Mittelwert könnten von Zufallsprozessen kommen
- Je größer die Varianz von Verteilungen, desto größer der Einfluss von Zufallseffekten



- Wann bedeutet also ein Unterschied im Stichproben-Mittelwert einen tatsächlichen Unterschied zwischen den Populationen?

t-Test

- Wir haben die Stichproben X und Y von zwei Population, welche nach zwei Verteilungen A und B verteilt sind
  - A und B sind normalverteilt
  - A und B haben die gleiche Varianz
  - X und Y haben die gleiche Stichprobengröße
- Frage: Sind die Mittelwerte von A und B unterschiedlich
- Wir wissen, das hängt ab von:
  - Den Stichproben-Mittelwerten
    (größerer Unterschied → Unterschied zwischen A und B wahrscheinlicher)
  - Den (unbekannten) Varianzen von A und B
    (große Varianz → Unterschied zwischen A und B unwahrscheinlicher)

Details des t-test

- Definiere T-Statistik:

$$T = \frac{\overline{X}\text{-}\overline{Y}}{\sigma_{X-Y}}$$

schätze diese von Stichproben:

$$t = \frac{\overline{x}\text{-}\overline{y}}{\sqrt{\frac{s_x^{\;2}}{n} + \frac{s_y^{\;2}}{n}}}$$

- Der Divisor ist der kombinierte Standardfehler der Mittelwerte
  - Quantifiziert wie weit der Stichproben-Mittelwert vom echten Mittelwert entfernt liegt
- T ist eine Zufallsvariable, von welcher wir den Wert t beobachten

- Wenn A und B gleiche Mittelwerte haben, folgt t der Student's *t-distribution*
- → Wir können die Wahrscheinlichkeit p errechnen $p = P(|T| \geq |t|)$
- Den Wert p erhält man durch Nachschlagen in Tabelle, diese ist bereits integriert in Statistiktools wie z.B. R, SAS, SPSS, Excel, OpenOffice
- Wenn p < kritischen Wert („Signifikanzniveau", α)
  → Mittelwerte von A und B unterscheiden sich

t-Test: Beispiel

- Ein Heim-Entertainment System mit Spracheingabe (A) wird erweitert durch ein zusätzliches Gesten-Interface (B)
- 40 Testpersonen wählen Musik entweder per Sprache (A) oder mit Sprache + Geste (B) aus. Die Auswahlzeit wird gemessen
- Annahme: Beide Populationen sind normalverteilt mit gleicher Varianz
- Mittelwert von Gruppe A ist 9.75 Sekunden, Gruppe B ist 11.1 Sekunde

- Unterscheiden sich diese beiden Mittelwerte signifikant?
  (kann man mit Sicherheit behaupten, dass System A eine schnellere Auswahlzeit als System B erlaubt?)
- Zu diesem Zweck setzt man ein „Signifikanzniveau“
  - in der Informatik in der Regel 5%,
  - in der Medizin eher 1% oder 0,1%
    in lebenswichtigen Entscheidungen möchte man kein Risiko eingehen!

t-Test: Beispiel

- Ein Heim-Entertainment System mit Spracheingabe (A) wird erweitert durch ein zusätzliches Gesten-Interface (B)
- 40 Testpersonen wählen Musik entweder per Sprache (A) oder mit Sprache + Geste (B) aus. Die Auswahlzeit wird gemessen.
- Annahme: Beide Populationen sind gleich groß (je 20 Personen) und normalverteilt mit gleicher Standardabweichung 2.56.
- Mittelwert von Gruppe A ist 9.75 Sekunden, Gruppe B ist 11.1 Sekunde

- Ist dieser Unterschied signifikant?
  - $t = \frac{11.1 - 9.75}{\sqrt{\frac{2.56^2}{20} + \frac{2.56^2}{20}}} \approx 1.6676$
  - p laut Tabelle: 0.1036
  - Wenn man behauptet würde, dass System A schneller als System B irrt man sich in 10% der Fälle, d.h. wenn man eine „sichere Aussage“ machen möchte, würde man in diesem Fall zu dem Schluss kommen, dass sich System A und B in Geschwindigkeit NICHT signifikant unterscheiden.

Zusammenfassung

- Daten / Ergebnisse analysieren und interpretieren, nicht nur präsentieren
  - Beschreibung der Daten, Verteilung (Mittelwert, Standardabweichung, Outlier)
  - Warum funktioniert ein Ansatz/Methode nicht?
  - Warum funktioniert ein Ansatz/Methode.
  - Statistische Tests zur abschließenden Aussage über signifikante/nicht signifikante Verbesserungen

Tipps für Bachelorarbeiten

- Nicht nur auflisten, sondern Bemerkenswertes identifizieren
  - Ziel ist Erkenntnisgewinn
- Fakten von eigenen Einschätzungen trennen
  - Z.B.: "Der grüne Legostein wurde nicht erkannt, vermutlich wegen des Glanzlichtes." (eine plausible Vermutung)
  - Z.B.: "Der grüne Legostein wurde nicht erkannt, weil das Glanzlicht nicht der Region zugehörig klassifiziert wurden und dadurch die Fläche zu klein wurde." (muss man überprüft haben; dann die wissenschaftlichere Aussage)
- Warum funktioniert etwas nicht?
  - Zwischenergebnisse anschauen und analysieren
  - Einsichten auf Entwurfsebene oft besonders interessant
- Zusammenfassung (engl. Conclusion) soll bewerten
  - "Das Verfahren erreicht eine Precision von 98% und ist unserer Meinung nach für den praktischen Einsatz geeignet.“
  - „Mit dem durch Sprache gesteuerten Heim-Entertainment System kann man signifikant schneller Musik auswählen, als nur mit Gesten.“