Analyse experimenteller Daten

-

Grundlagen

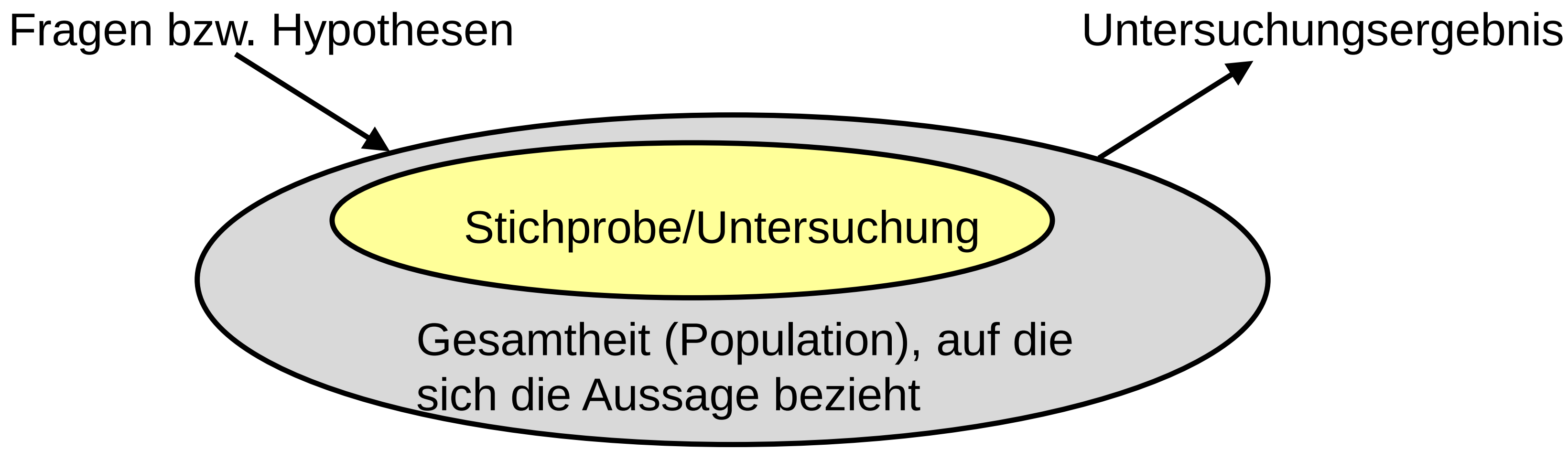
Human Centered Multimedia
Institute of Computer Science
Augsburg University
Universitätsstr. 6a
86159 Augsburg, Germany

Aufgaben der Datenanalyse

- **Deskriptive bzw. Beschreibende Statistik**
  - Aufbereiten der Datengrundlage:
    - Behandlung fehlender Werte
    - Ausreißerkontrolle
  - Beschreiben und Darstellen der Hauptmerkmale erhobener Daten
- **Explorative Statistik**
  - Finden von Strukturen in den Daten
  - Aufstellen von Hypothesen, ausgehend von den Daten
- **Schließende Statistik bzw. Inferenzstatistik**
  - Prüfen von Hypothesen
  - Schließen von Werten einer Stichprobe auf die Werte in einer Population

Grundlagen:

- **Population:** Gesamtheit, über die eine Aussage getroffen werden soll (z.B. die gesamte Bevölkerung in Deutschland).
- Erforschung einer Fragestellung für eine ganze Populationen oft zu aufwändig => **Stichprobe**
- Mit Hilfe der Stichprobe wird versucht, die Verteilung des Merkmals in der Population - mit Hilfe statistischer Verfahren - zu schätzen.

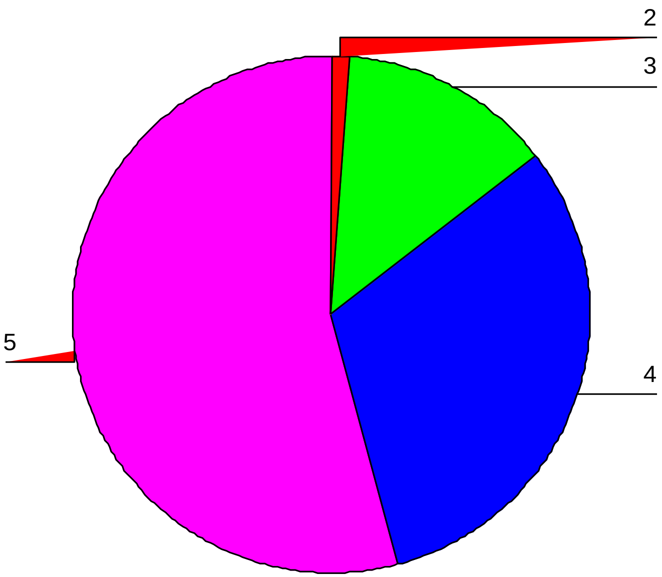


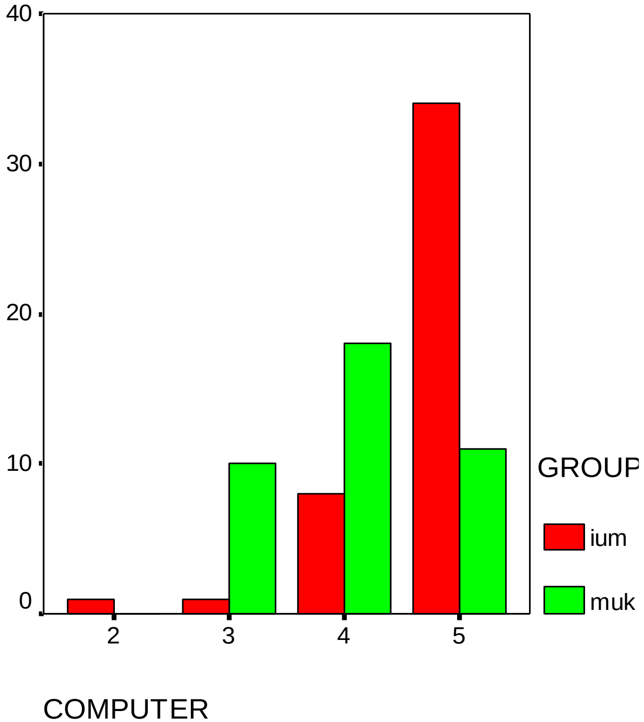
Arten experimenteller Daten

- **Nominalskalierte Daten**
  - Klassifikation von Individuen, Zuständen oder Ereignissen je nach Vorhandensein oder Abwesenheit eines oder mehrerer Merkmale
  - Beispiel: Geschlecht, Herkunft, Beruf, Vorlieben
- **Ordinalskalierte Daten**
  - Werden Gegenstände oder Sachverhalte einer Kategorie daraufhin beurteilt, ob die Merkmalsausprägung ,,größer", „kleiner" oder „gleich" ist (Größe, Stärke, Intensität), dann entsteht eine Ordinalskala.
  - Beispiel: Platzierungen in Wettkämpfen, Grad der Usability
- **Intervallskalierte Daten**
  - Wird der Abstand (das Intervall) zwischen zwei benachbarten Skalenpunkten für die gesamte Skala einheitlich definiert, dann entsteht eine Intervallskala. Hier sind sowohl Ordnungen (größer als, kleiner als) als auch Differenzen bestimmbar.
  - Beispiel: Temperatur, IQ, Dauer der Nutzung von X

Deskriptive Statistik

- Erster Eindruck für alle Arten von Daten: Punktzahlen, Zeiten, Bewertungen etc.
- Ausgabe der deskriptiven Statistik





Mittelwert:

$$\bar{x} = {}^{1}\!/_{n} \cdot \sum_{i=1}^{n} x_i$$

- Welche Gruppen unterscheiden sich stärker?

Experiment 1

Gruppe 1	Gruppe 2
Mittel: 7	Mittel: 10
1,10,10	3,6,21

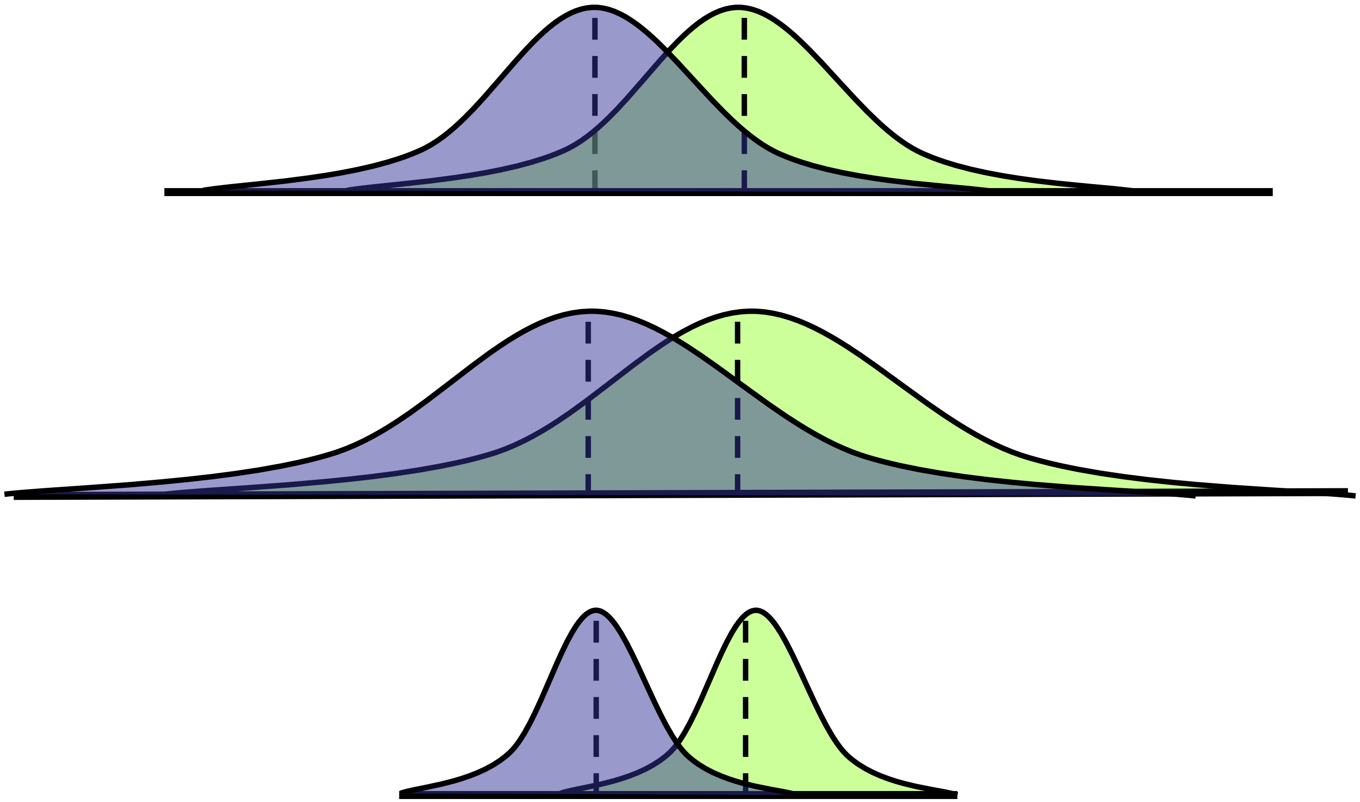
Experiment 2

Gruppe 1	Gruppe 2
Mittel: 7	Mittel: 10
6,7,8	8,11,11

Universität Augsburg University

Mittelwerte:

- Welche Gruppen unterscheiden sich stärker?

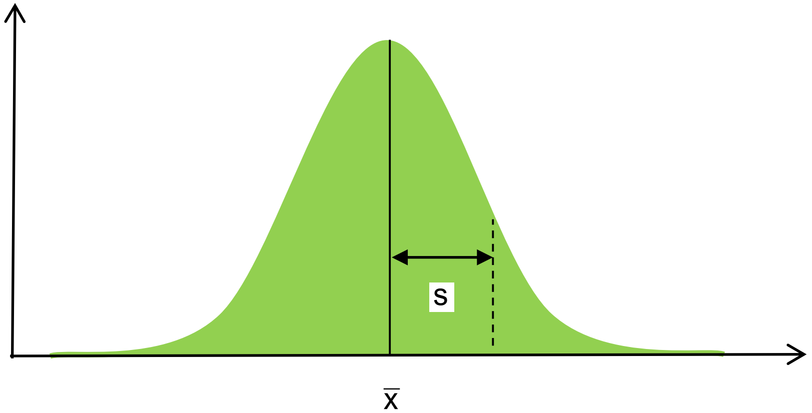


- **Varianz:** Maß wie stark eine Messgröße „streut“, d.h. was der durchschnittliche Fehler zum Mittelwert ist, der entstehen kann

$$v = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$

- **Standardabweichung:** Maß für den Abstand des Mittelwerts zum Wendepunkt einer Normalverteilung an. Damit ist die Standardabweichung ein Maß für die Breite einer Normalverteilung.

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}$$



- Varianz und Standardabweichung messen
  - die Genauigkeit der Schätzung des Mittelwerts einer Population
  - die Schwankungen der Daten

- **Hohe Werte** zeigen, dass es **sehr starke Schwankungen** im vorliegenden Datensatz gibt. Der Mittelwert ist dann **kein** repräsentativer Wert für alle Testdaten.

Statistischer Kennwert	Stichprobe	Population	Schätzwert
Arithmetisches Mittel	$\bar{x}$	μ	$\widehat{\mu}$
Varianz	s^2	σ^2	$\hat{\sigma}^2$
Standard-abweichung	s	σ	$\hat{\sigma}$

- Häufiges Problem: Daten aus unterschiedlichen Tests sollen miteinander verglichen werden.
- Testwerte aus unterschiedlichen Verteilungen mit unterschiedlichen Mittelwerten können nicht anhand der Absolutwerte verglichen werden

➢ **z-Standardisierung** $z_i = \frac{x_i - \mu}{\sigma}$

- Die transformierten Werte gehören einer z-Verteilung an.
- Mittelwert = 0 und Standardabweichung = 1
- Vorteil:
  - Werte sind unabhängig von Mittelwert und Streuung der ursprünglichen Verteilung vergleichbar
  - z-Werte aus unterschiedlichen Tests lassen sich zur Bildung eines arithmetischen Mittels aus verschiedenen Testwerten nutzen.

Universität Augsburg University

- Beispiel:
  - In zwei Tests T-1 und T-2, die Verteilungen mit unterschiedlichen Mittelwerten und Streuungen entstammen

$$(\mu_1 = 100; \quad \sigma_1 = 15; \quad \mu_2 = 50; \quad \sigma_2 = 5)$$

erzielt eine Versuchsperson die folgende Werte:

$$x_1 = 110 \qquad x_2 = 53$$

  - Frage: Hat sie in beiden Tests gleich abgeschnitten?
  - Transformation der Testwerte in z-Werte:

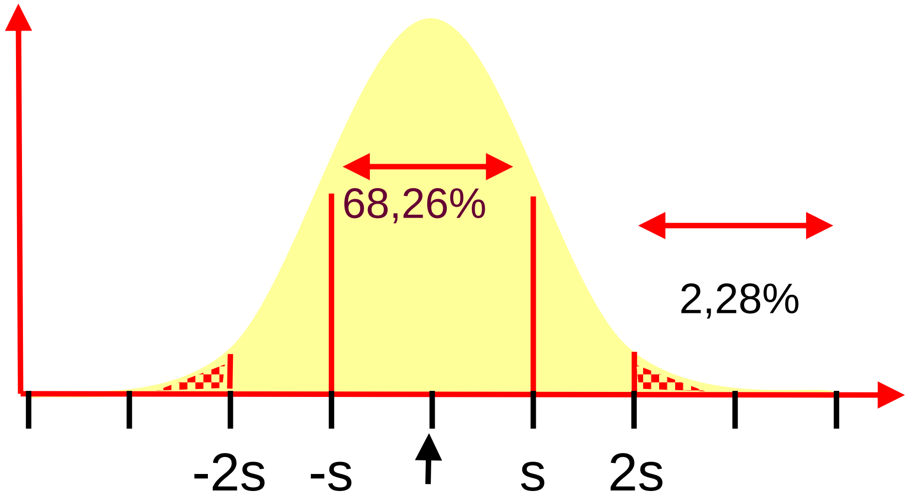
$$z_1 = \frac{110 - 100}{15} = {}^{2}\!/_{3} \approx 0{,}67 \qquad z_2 = \frac{53 - 50}{5} = {}^{3}\!/_{5} = 0{,}6$$

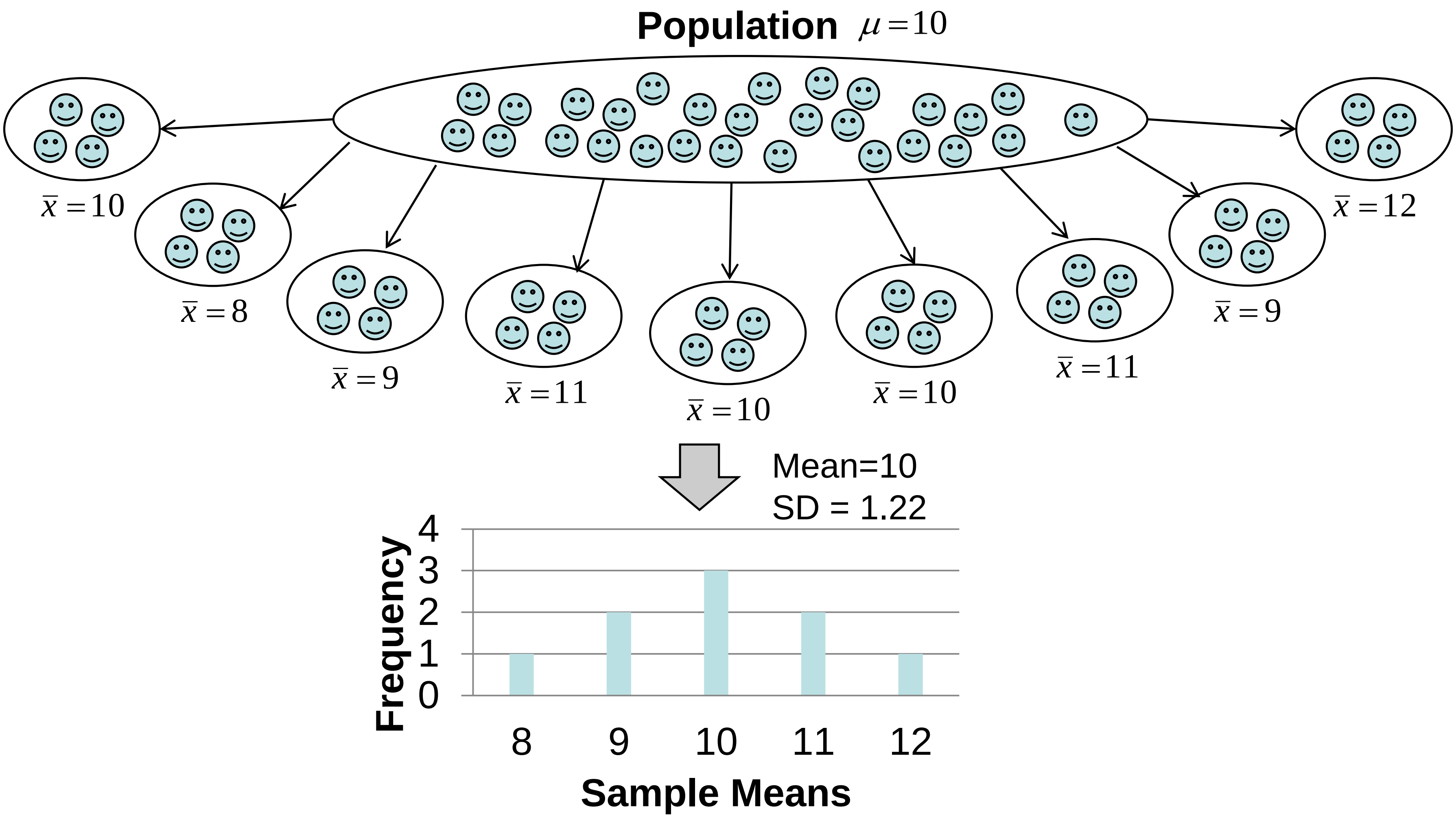
  - Aus den beiden z-Werten lässt sich anschließend ein arithmetisches Mittel wie folgt bilden:

$$\bar{z} = \frac{0{,}67 + 0{,}6}{2} = 0{,}635$$

Normalverteilung:

- Voraussetzung für viele statistische Verfahren
- Symmetrische glockenförmige Kurve, die sich der x-Achse asymptotisch nähert ohne sie jemals zu berühren
- Wird durch zwei Merkmale vollständig beschrieben: arithmetisches Mittel (=Median) und Streuung
- Unter unendlich vielen Normalverteilungen gibt es eine mit Mittelwert 0 und der Streuung 1: Standardnormalverteilung (z-Verteilung)




Population $\mu = 10$
$\bar{x} = 10$
$\bar{x} = 8$
$\bar{x} = 9$
$\bar{x} = 11$
$\bar{x} = 10$
$\bar{x} = 10$
$\bar{x} = 11$
$\bar{x} = 9$
$\bar{x} = 12$
Mean=10
SD = 1.22
Frequency
0
1
2
3
4
8
9
10
11
12
Sample Means

- Der Standardfehler der Mittelwerte ist definiert als die Streuung in einer Verteilung von Mittelwerten aus gleich großen Zufallsstichproben einer Population.

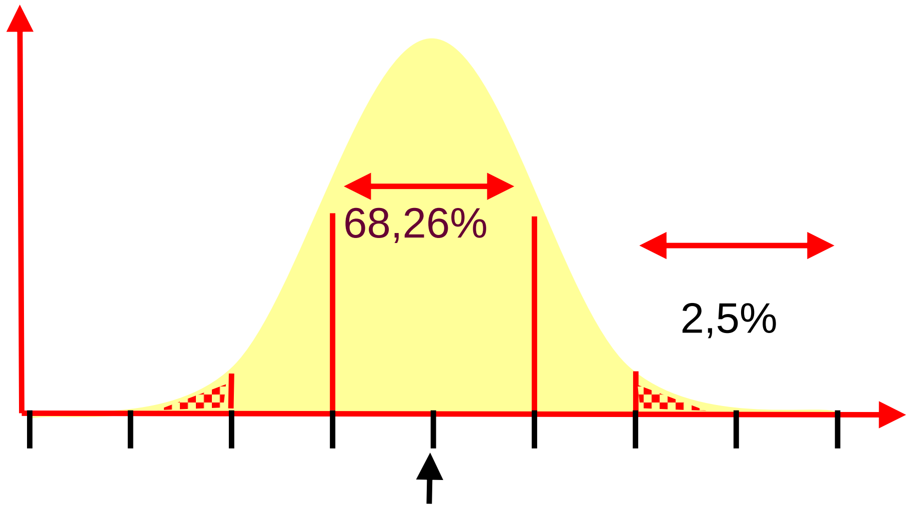
$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$$

σ_x Standardabweichung der Stichprobe,
n Größe der Stichprobe

- Der Standardfehler liefert eine Aussage über die Güte des ermittelten Mittelwertes.
- Je größer die Streuung der Population, desto größer ist der Standardfehler.
- Je größer die Stichprobe ist, desto kleiner ist der Standardfehler.

- Standardfehler eignet sich zur Konstruktion von Konfidenzintervallen.
- Ein Konfidenzintervall gibt an, in welchem Bereich der Großteil aller Mittelwerte aller Stichproben bzw. der wahre Populationsparameter sich mit einer bestimmten Wahrscheinlichkeit befindet.
- Es berechnet sich wie folgt:
  - Untere Grenze: $\bar{X} - z \cdot \sigma_{\bar{x}}$
  - Obere Grenze: $\bar{X} + z \cdot \sigma_{\bar{x}}$
- Für ein 95%-Konfidenzintervall gilt: z=1,96
- Für ein 99%-Konfidenzintervall gilt: z=2.58
- Der „wahre" Mittelwert liegt in der Mitte des Intervalls

- Wenn wir beispielsweise ein 95%-Konfidenzintervall bilden, sagen wir damit, dass in 95% aller Fälle (also in 95 von 100 gezogenen Stichproben) der wahre Populationsparameter, den wir mit Hilfe der Stichprobe schätzen, in diesem Intervall liegt.



Evaluierung der Bedienfreundlichkeit einer Nutzerschnittstelle

- Bewertungen von 20 Studenten der Universität Augsburg
- Annahme: Studenten einigermaßen repräsentativ für Studenten der Uni
- Mittlere Bewertung: 3.22 (Standardabweichung von 0.48)
- Messung des Mittelwerts der Population mit Hilfe der Stichprobe durch ein 95%-Intervall
- Standardfehler: $$\sigma_{\bar{x}} = s/\sqrt{n} = 0.48/\sqrt{20} = 0.11$$
- Konfidenzintervall:
  - Untere Grenze: 3.22 – 1.96*0.11= 3.00
  - Obere Grenze: 3.22+1.96*0.11= 3.44

➢ Stichproben werden zu 95%-iger Sicherheit einen Mittelwert zwischen 3 und 3.44 erhalten

Deskriptive Statistiken –
Darstellung

Wichtig:

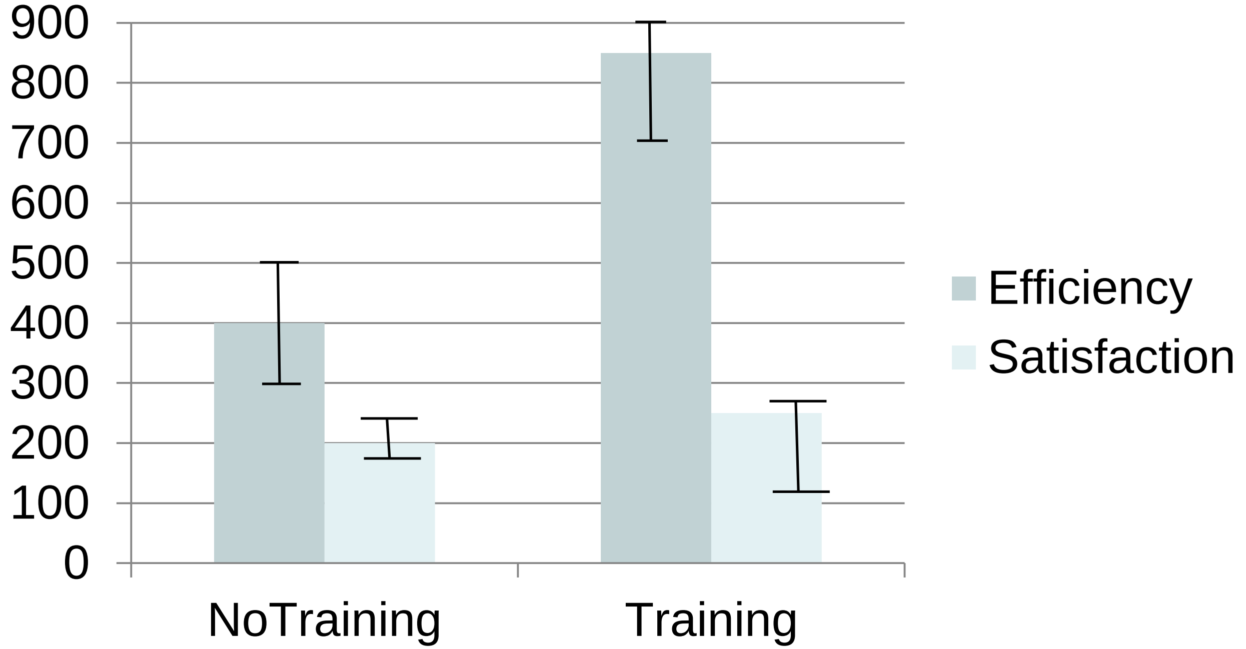
- Nicht nur Mittelwerte, sondern auch Standardabweichung bzw. Standardfehler angeben.

Beispiele aus wissenschaftlichen Papieren:

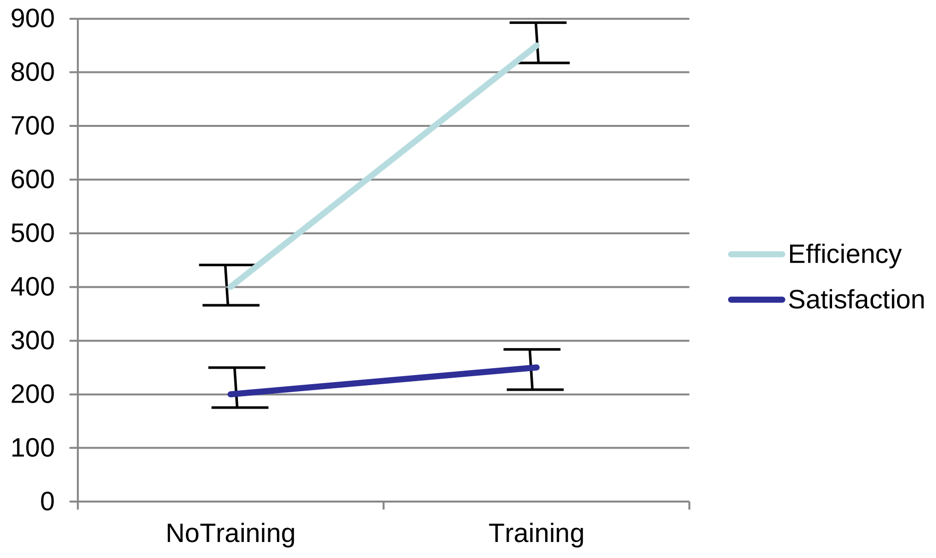
- Seventy persons, ranging from 17 to 48 years age (M=24.09; SD=5.717), participated in the study.
- Participants on average laughed 8.76 times (SD=7.696).
- Participants frequently laughed while interacting with the program (M=8.76; SE=1.2).

Deskriptive Statistiken – Darstellung

- Balkendiagramm mit Konfidenzintervallen für jede Gruppe und jedes Merkmal



- Liniendiagramm mit Konfidenzintervallen für jede Gruppe und jedes Merkmal



Signifikanz

- Wenn die Parameter einer Population bekannt sind, dann lässt sich mit Hilfe von statistischen Verfahren auch bestimmen, ob der entsprechende Mittelwert einer Stichprobe bedeutsam oder nur zufällig größer oder kleiner ist als der Mittelwert der Population.

- **Beispiele:**
  - Der mittlere IQ von Kindern in der 3. Klasse beträgt 100.
  - Bei einer 3. Klasse mit 20 Kindern eines Sportgymnasiums wurde ein mittlerer IQ von 110 ermittelt?
  - Sind Kinder in der 3. Klasse eines Sportgymnasiums nun besonders intelligent?

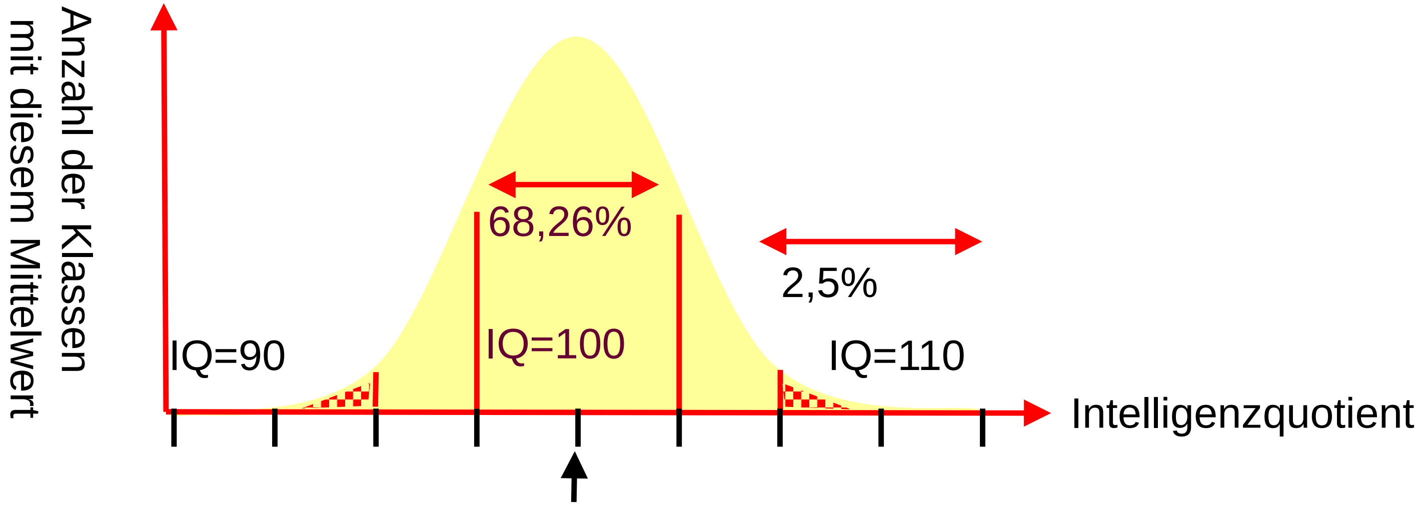
Formulierung von Hypothesen

- **Nullhypothese H0**
  - Die beobachteten Daten lassen sich durch zufallsbedingte Abweichungen erklären.
- **Alternativhypothese H1**
  - Die erhobenen Daten sprechen für die Vermutung, dass es einen wahren Effekt gibt.

- **bezogen auf das Beispiel:**
  - **Hypothese H0:** Der Mittelwertsunterschied der geprüften Schulklasse und der Population ist zufällig zustande gekommen.
  - **Hypothese H1:** Der Mittelwertsunterschied ist nicht zufällig zustande gekommen. Er ist **signifikant**.

Signifikanz (Beispiel)

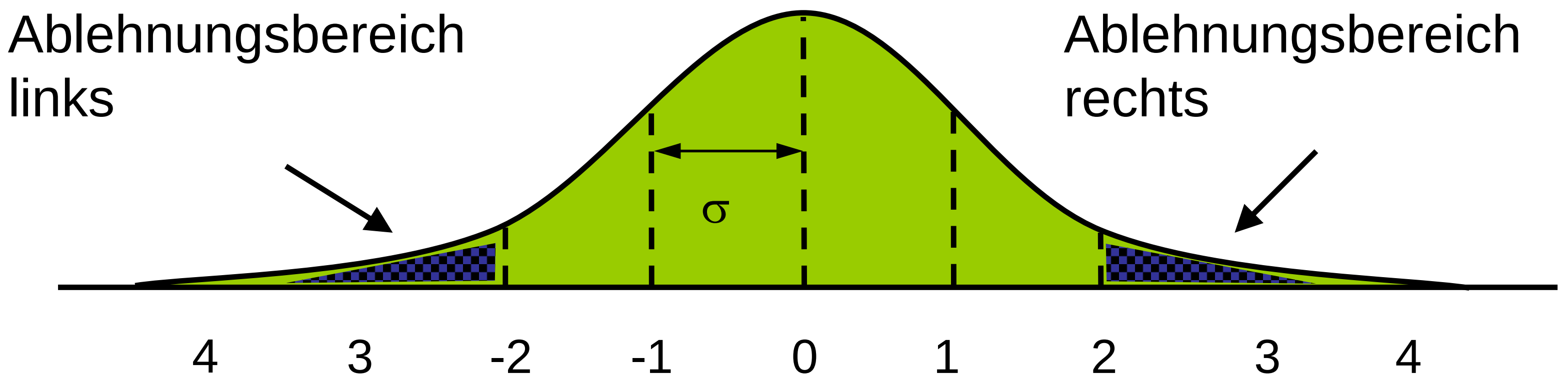
- **Annahme:** Schulklasse mit einem durchschnittlichen IQ von 110.
- **Frage:** Ist die Abweichung vom allgemeinen Mittelwert Zufall?
- Häufigkeitsverteilung der Mittelwerte des IQs von 100 Schulklassen:



- Wahrscheinlichkeit, dass der mittlere IQ einer Klasse 110 oder mehr beträgt, ist äußerst gering.

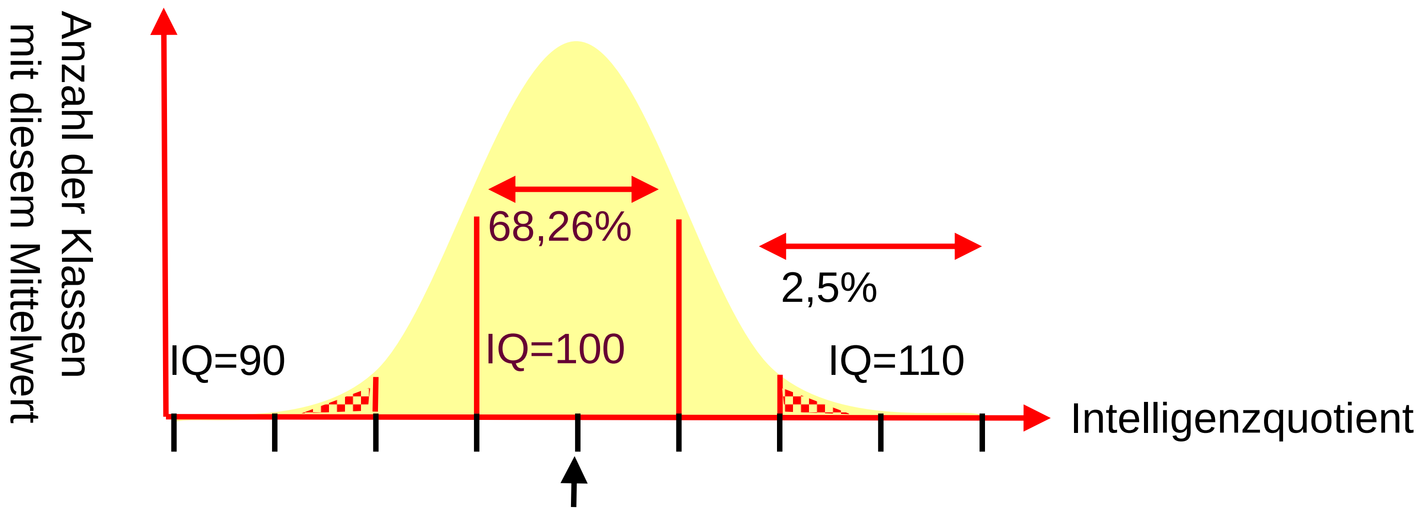
Signifikanz

- Die kritischen Werte des Signifikanztests ergeben sich aus den Grenzen zwischen
  - Annahmebereich und
  - Ablehnungsbereich
  - (Annahme bzw. Ablehnung der Null-Hypothese)



Signifikanz (Beispiel)

- **Annahme:** Schulklasse mit einem durchschnittlichen IQ von 110.
- **Frage:** Ist die Abweichung vom allgemeinen Mittelwert Zufall?
- Häufigkeitsverteilung der Mittelwerte des IQs von 100 Schulklassen:



- Wahrscheinlichkeit, dass der mittlere IQ einer Klasse 110 oder mehr beträgt, ist äußerst gering.

➢ Mit hinreichender Sicherheit kann man sagen, dass der Unterschied nicht nur durch Zufall zustande gekommen ist.

Fehlentscheidungen beim Testen

- **Fehler 1. Art (Signifikanzniveau): das unberechtigte Ablehnen der Nullhypothese**

  p(Fehler 1. Art) = α

- **Fehler 2. Art: das unberechtigte Beibehalten der Nullhypothese**

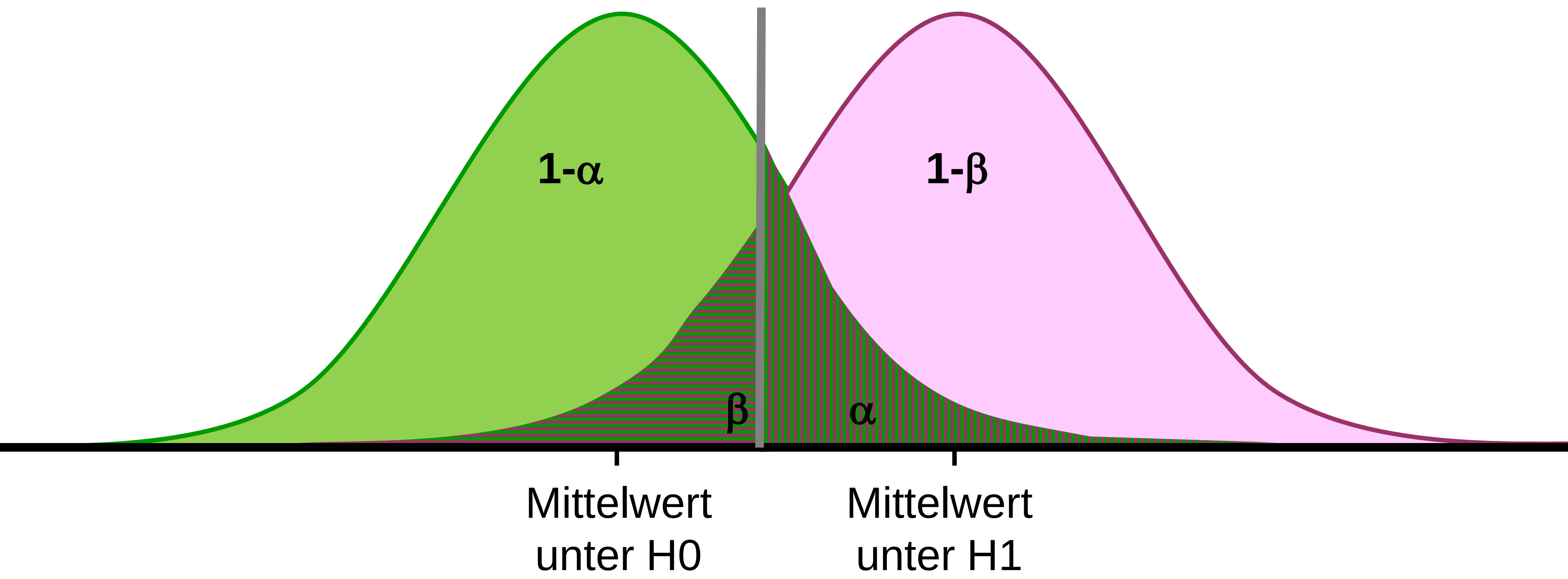
  p(Fehler 2. Art) = β

α	Nicht existierender Unterschied wird als Effekt ausgegeben
1-α	Nicht existierender Unterschied wird auch als solcher erkannt
β	Vorhandener Effekt wird nicht entdeckt
1- β	Vorhandener Effekt wird auch entdeckt

Die 4 Möglichkeiten des Entscheidungsproblems

Fehlentscheidungen beim Testen

- Falls die Alternativhypothese gilt, dann machen wir in α der Fälle einen Fehler, in 1- β der Fälle liegen wir richtig.
- Falls die Nullhypothese gilt, dann machen wir in β der Fälle einen Fehler, in 1- α der Fälle liegen wir richtig.



Signifikanzstufen

- Ein Testergebnis heißt statistisch signifikant, wenn der p-Wert unterhalb des vorgegebenen Fehlers 1. Art α liegt ($p \leq \alpha$)

- Dabei gibt es klassischerweise drei Signifikanzniveaus:
  - $p \leq 0{,}05$ signifikant *
  - $p \leq 0{,}01$ sehr signifikant **
  - $p \leq 0{,}001$ höchst signifikant ***

Signifikanzprüfung:
Vergleich eines empirischen Testwerts mit einem kritischen Wert, der anhand einer theoretischen Verteilung bestimmt wird und für den eine bestimmte Irrtumswahrscheinlichkeit gilt.

- **Einige Verfahren:**
  - **Unterschiedtests:** Zwei Messreihen werden hinsichtlich ihrer Mittelwerte miteinander verglichen.
  - **Varianzanalyse:** Mehrere Messreihen werden hinsichtlich ihrer Mittelwerte miteinander verglichen.
  - **Zusammenhangtests:** Der Zusammenhang zwischen zwei Messreihen wird analysiert.

Prinzipielle Vorgehensweise

- Formulierung der Hypothesen
  - Nullhypothese - Alternativhypothese
- Wahl des Signifikanzniveaus
- Wahl des Testverfahrens
- Durchführung des Tests und Entscheidung