# 1 Daten

## 1.1 Klassifizierung

Statistisches Lernen: - allgemeine Methodik, wie kann man Daten bzw. empirische Befunde beschreiben, auswerten und verallgemeinern (→ generalizzare) kann.
- erlaubt eine evidenzbasierte Entscheidungsfindung mittels Schätzen (Stime) und Testen.

**Deskriptive Statistik**: Darstellung und Zusammenfassung von beobachteten Daten, Description and representation of Data.

La statistica descrittiva utilizza metodi numerici e grafici per trovare modelli nei dati, informazioni ricevute e per presentare tali informazioni in modo significativo.

**Induktive Statistik** (→ process of using the data): Analyse von Daten mithilfe mathematischer Modelle. Wir versuchen Rückschlüsse zu ziehen und definieren konkrete Messwerte (Verallgemeinerungen, Gesetzmässigkeiten)

(schliessende, Inferenz-Statistik)

La statistica induttiva usa i dati per fare stime, decisioni, previsioni o altre generalizzazioni sull'ambiente da cui i dati sono stati ottenuti.

Mit Statistik kann man Wissen aus Daten generieren, das ist unmöglich ohne Modelle, Vereinfachungen, Abstraktionen.

### Klassifizierung von Datentypen

- **Nach Skala**: kategorisch (→ auch Merkmale mit Ja oder Nein (z.B: Kreditausfall)) (Haarfarbe, Geschlecht) (qualitativ), keine Rechenoperation, 1) metrisch (quantitativ) (Metrische Daten: → Rechenoperationen möglich) diskret (nach Jahren), abzählbare Werte/Ausprägungen, Noten, Anzahl Produkte, 2) metrisch stetig (Alter, Lohn) (→ genaue Körpergrösse, Breite des Produktes) (quantitativ) nicht abzählbare Ausprägungen, non hanno un numero contato di caratteristiche

  Oftmals treten metrische und kategorische Daten auf (appaiono spesso assieme/z.B $CO_2$ Emissions graph, pg. 10/33 Kap. 1)

| Class (Highest Degree) — qualitativ | Frequency (Number of Executives) — quantitativ | Relative Frequency (Proportion of Total) |
|---|---|---|
| None | 2 | 0.1 |
| Bachelors | 9 | 0.45 |
| Masters | 5 | 0.25 |
| Doctorate | 4 | 0.2 |
| Total | 20 | 1.0 |

Table 1.2: Summary of Table 1.1

- **Nach Herkunft** (→ origine):
  - Primärdaten (→ raccolti per la prima volta in relazione diretta con un soggetto di indagine.) (Rohdaten, dati grezzi), lassen sich in Häufigkeitstabellen zusammenfassen.*, abgeleitete Daten (→ Sekundärdaten) (↳ prozessierte Primärdaten), Daten über Daten (Metadaten) (→ sie enthalten Informationen zu anderen Daten, sie haben eine übergeordnete (sovraordinata) Funktion.
  - Prozessdaten, administrative Daten (vom Staat; Steuererklärung), Erhebungen (→ rilevazioni), Experimente

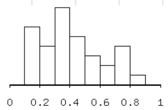
  * Von Vorteil, wenn der gleiche Wert mehrmals auftritt. ↳ bei kategorische und metrisch diskrete Daten.
  Le tabelle non sono adatte per dati metrisch stetig.

- **Nach Verfügbarkeit**: öffentlich, frei verfügbar (→ die meisten sind gratis) (↳ Können von allen Menschen ohne Einschränkungen verwendet werden.) / proprietär (in Eigentum) und kommerziell (↳ appartengono a qualcuno (privati o organizzazioni) e non sono quindi accessibili. (es. dati di Google))

- **Nach Struktur**:
  - Querschnittsdaten (→ same time period, → dati raccolti in un momento/periodo temporale (zu einem Zeitpunkt)) (↳ Relevant ist die Veränderung der Daten im Zeitablauf (scorrere del tempo)), Zeitreihen (→ Dati osservati e registrati in diversi periodi di tempo (e non solo uno come per i querschnittsdaten)) (→ Veränderung der Daten über die Zeit) (Wetter, Temperatur) (↳ Different units of time, variazione dei dati nel tempo), Paneldaten (wiederholte Querschnittsdaten; Querschnitt + Zeitreihe) (↳ osservazione di diverse variabili in una serie di periodi di tempo)
  - Unstrukturiert (↳ nicht formalisierte Form, können in einen Datenbank nicht gespeichert werden (Word-Datei, E-mail Verkehr)) oder komplexe Strukturen (Text) (↳ Strukturiert ↳ formalisierte Form, können in einen Datenbank gespeichert werden (Excel-Tabelle, CSV-Datei))

## 1.2 Graphische Darstellung

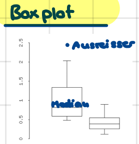
### Histogramm




für metrisch stetige Merkmale (z.B. Lohnverteilung) quantitative Daten / numerisch

Eigenschaften: symmetrisch, asymmetrisch, Unimodal (ein Gipfel), Bimodal (zwei Gipfel), Rechtsschief (linke Seite höher), $\bar{x} > m$
Linksschief (rechte Seite höher), $\bar{x} < m$, Uniform/Gleichverteilt

### Boxplot




• Ausreisser → extremer Datenwert

rechtsschief: Boxplot zieht nach links/unten

linksschief: Boxplot zieht nach rechts/oben

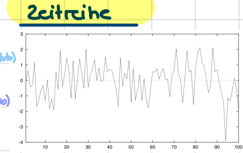
### Häufigkeitstabelle

| Class | Frequency M | Frequency F | Frequency Total | Relative Frequency M | Relative Frequency F | Relative Frequency Total |
|---|---|---|---|---|---|---|
| 0.0 – 0.5 | 1 | 23 | 24 | .06 | .70 | .48 |
| 0.5 – 1.0 | 10 | 10 | 20 | .58 | .30 | .40 |
| 1.0 – 1.5 | 4 | 0 | 4 | .24 | .00 | .08 |
| 1.5 – 2.0 | 1 | 0 | 1 | .06 | .00 | .02 |
| 2.0 – 2.5 | 1 | 0 | 1 | .06 | .00 | .02 |
| Total | 17 | 33 | 50 | 1.00 | 1.00 | 1.00 |

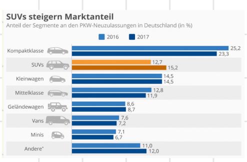
Absolute Anzahl $h_k$ (numero assoluto)

Relative Anzahl $p_k$ (numero relativo)

### Zeitreihe



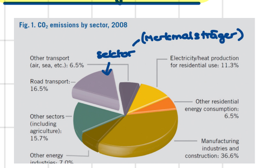
### Balkendiagramm




kategorische Daten

(qualitativ)

### Kuchendiagramm




metrische Daten

(qualitativ)

**Geometrisches Mittel**: $\bar{X}_{geom} = \sqrt[N]{X_1 \times X_2 \times \ldots \times X_N}$

Beispiel: Der Wertzuwachs eines Portfolios betrage im ersten Jahr 5%, im zweiten Jahr 25%. Es wäre falsch zu sagen, dass der ∅-Zuwachs 15% beträgt, da

$$100 \times 1.15 \times 1.15 = 132.25 \neq 100 \times 1.05 \times 1.25 = 131.25$$

Die echte ∅-Rendite ist $\bar{r} = \sqrt{1.05 \times 1.25} - 1 = 0.146 = 14.6\%$

→ verwendet, um eine durchschnittliche Wachstumsrate zu berechnen.

## 1.3 Kennzahlen

Le cifre chiave (Kennzahlen) riassumono le informazioni quantitative e le rendono comparabili.

### Arithmetisches Mittel (Mittelwert) $\mu$

Durchschnittswert der $x_i$; aus Rohdaten berechnet

↳ Definition: Statistischer Lageparameter für metrische Merkmale, der den Durchschnitt alle betrachteten Werte zeigt

$$\boxed{\bar{x} = \frac{1}{n}(x_1 + x_2 + \ldots + x_n)}$$

Summe alle Werte / Anzahl Werte

$$\bar{X} = \frac{X_1 + X_2 + \ldots + X_N}{N} = \frac{1}{N}\sum_{i=1}^{N} X_i$$

**Media aritmetica complessiva**

Aus Häufigkeitstabellen mit $K$ verschiedenen Werten $X_k$ wird das arithm. Mittel wie folgt berechnet:

$$\bar{X} = f_1 X_1 + \ldots + f_K X_K = \sum_{k=1}^{K} f_k X_k$$

wobei $f_k = h_k / \sum_{k=1}^{K} h_k = h_k / N$.

$f$ ist die Frequenz

↳ $\bar{x} = \sum_{g=1}^{G} f_g \bar{x}_g$

### Median

m: Datenwert, der genau die Hälfte der Datenwerte teilt ⚠ Zuerst muss man die Daten ordnen! robuster Lageparameter

ist das mittlere Ergebnis bei ungeradem n (sonst Mittelwert der beiden mittleren Ergebnissen)

Bei einer ungeraden Anzahl an Datenwerten ist der Median der Wert in der Mitte.
Bei einer geraden Anzahl an Datenwerten entspricht der Median dem Durchschnitt der beiden mittleren Werte.

| $x$ | $g$ |
|---|---|
| 0.03 | 0 |
| 0.07 | 1 |
| 0.14 | 0 |
| 0.27 | 0 |
| 0.32 | 0 |
| 0.35 | 1 |
| 0.35 | 1 |
| 0.56 | 1 |
| 0.87 | 1 |
| 0.88 | 1 |

Median

m

$\boxed{Med(x) = (0.32 + 0.35) : 2}$

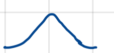
$Min(x) = 0.03$  $Max(x) = 0.88$

Symmetrie: liegen Mittelwert + Median weit auseinander, gibt es wenig Symmetrie

$\bar{x} < m$ : linksschief 

$\bar{x} > m$ : rechtsschief 

| $p_k$ |
|---|
| 1/12 |
| 2/12 |
| 3/12 |
| 3/12 |
| 2/12 |
| 1/12 |

$\bar{x} = m$ : symmetrisch 

↳ Symmetrische Verteilung: Median und Mittelwert stimmen überein.

**Perzentile** Verallgemeinerung des Medians

- $Q_1$ das erste Quartil 25% - Perzentil
- $Q_2 = m$ Median 50% - P
- $Q_3$ das dritte Quartil 75% - P

→ **Interquartilabstand:** $\boxed{IQA = Q_3 - Q_1}$

robustes Streuungsmass

↳ kann nie negativ sein! (Al minimo 0 quando non c'è varianza tra i dati (sono tutti uguali)) → $\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2$

**Varianz** mittlere quadratische Abweichung

→ Streuung einer Verteilung
↳ sparpimento
→ misura la deviazione quadratica media dal valore medio.

$$\boxed{\sigma^2 = \frac{1}{N}\left((x_1-\bar{x})^2 + \dots + (x_n-\bar{x})^2\right)}$$

Anzahl ist bekannt (Grundverteilung)

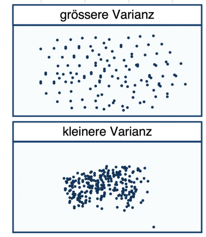
$$\boxed{s^2 = \frac{1}{n-1}\left((x_1-\bar{x})^2 + \dots + (x_n-\bar{x})^2\right)}$$

Anzahl ist unbekannt (Stichprobe)

**Standardabweichung** Deviazione standard

Sample Variance

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N} x_i^2 - \left[\frac{1}{N}\sum_{i=1}^{N} x_i\right]^2$$



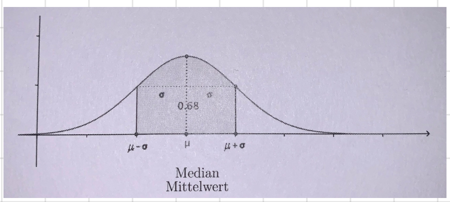

$\boxed{\sigma = \sqrt{Var(x)}}$ (Grundverteilung) oder $s = \sqrt{Var(x)}$ (Stichprobe)

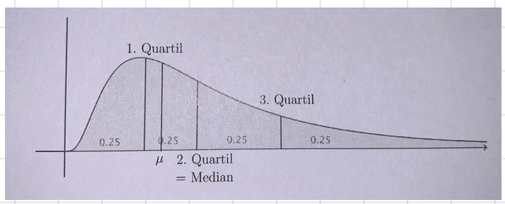
↳ $\sigma = \sqrt{\sum_{g=1}^{?} P_g (x_g - \bar{x})^2}$ → per gruppe oder $\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2}$.

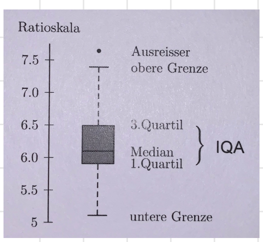
**Standardisierter Datensatz**

$$\boxed{x_i = \frac{x_i - \bar{x}}{\sigma}}$$

**Allgemein zu den Kennzahlen:**




**Ausreisser:** extremer Datenwert in einer Stichprobe

- $x_i > Q_3 + 1.5 \cdot IQA$ → Ausreisser nach oben
- $x_i < Q_1 + 1.5 \cdot IQA$ → Ausreisser nach unten

## 1.4 Gruppierte Daten

Sei $g \in G$ ein Indikator der Gruppenzugehörigkeit, $\bar{x}_g$ der Mittelwert in Gruppe $g$, und $f_g$ die relative Häufigkeit der Gruppe $g$.

Dann gilt:

$\bar{x} = f_1\bar{x}_1 + \ldots + f_G\bar{x}_G$ (gewichteter Mittelwert)

Beispiele:

- Altersklassen (G basierend auf metrischem Merkmal)
- Autotyp (kategorisches G)
- Experiment (binäres G): $\bar{x}_1$ Mittel in Versuchsgruppe; $\bar{x}_0$ Mittel in Kontrollgruppe

**Kompositionseffekt**

Der Kompositionseffekt beschreibt den Anteil des Gesamtunterschiedes, der sich aus der unterschiedlichen Zusammensetzung der Gesamtmenge ergibt. → che risulta dalla diversa composizione dell'insieme totale.
↱ parte della differenza totale

### Gruppierte Daten: Beispiel

| | Kompakt | Mittelklasse | SUV | Andere |
|---|---|---|---|---|
| Marktanteil bei Neuzulassungen in 2014 | 27% | 17% | 16% | 40% |
| Durchschnittliches g CO2/km | 105 | 140 | 152 | 135 |
| Marktanteil bei Neuzulassungen in 2017 | 22% | 17% | 24% | 37% |
| Durchschnittliches g CO2/km | 104 | 136 | 150 | 130 |

Der ∅ Ausstoss in 2014: $105 \times .27 + 140 \times .17 + 152 \times .16 + 135 \times .40 = 130.4$ → $\bar{x}_{2014}$

Der ∅ Ausstoss in 2017: $104 \times .22 + 136 \times .17 + 150 \times .24 + 130 \times .37 = 130.1$ → $\bar{x}_{2017}$

Hypothetische Entwicklung bei gleichbleibendem Anteil:
$104 \times .27 + 136 \times .17 + 150 \times .16 + 130 \times .40 = 127.2$

Dieses Beispiel illustriert die Bedeutung von Kompositionseffekten.
↳ betrachtet, wie sich der gesamte Mittel über die verschiedenen Gruppen verändert

**Laspeyre-Index**

### Konsumentenpreisindex

Ein ähnliche Kompositionsanalyse liegt dem Schweizer Konsumentenpreisindex (CPI) zugrunde. Betrachtet seien die Konsumausgaben eines repräsentativen Haushaltes zum Zeitpunkt $t = 0, 1, 2, \ldots$.

Der Laspeyre Index ist definiert als: $\boxed{P_t = \frac{\sum_{i=1}^{I} p_{ti} q_{0i}}{\sum_{i=1}^{I} p_{0i} q_{0i}}}$

$q_{0i}$ ist ein "Warenkorb" von $i = 1, \ldots, I$ verschiedenen Gütern in der Basisperiode.

$\sum_{i=1}^{I} p_{0i} q_{0i}$ sind die Gesamtausgaben für diesen Warenkorb in der Basisperiode.

$\sum_{i=1}^{I} p_{ti} q_{0i}$ sind die (hypothetischen) Ausgaben für die gleichen Güter, aber jetzt bewertet zu aktuellen Preisen $p_{ti}$.

"Hypothetisch", weil sich in der Realität natürlich auch die Konsummuster verschieben. Zudem reagieren Leute auf Preisveränderungen und nehmen Substitutionen vor.

| | **Welche der folgenden Aussagen trifft nicht zu?** |
|---|---|
| 1: | Haarfarbe, Studienfach und Ehestand sind qualitative Merkmale. |
| 2: | Die Varianz kann nie negativ sein. |
| 3: X | Wenn Frauen in jeder Prüfung die besseren Noten haben als Männer, dann liegt der Notendurchschnitt von Frauen über dem der Männer. → es hängt von dem Anteil ab! |
| 4: | Der Median teilt die geordneten Daten in zwei gleichgrosse Hälften. |

# 2 Wahrscheinlichkeitstheorie

## 2.1 Glücksspiele / Zufallsexperiment

→ disjunkt aber nicht unabhängig (K und Zahl sind disjunkt aber abhängig, denn wenn Kopf eintritt kann nicht Zahl eintreten (wenn unabhängig ist $P(A \cap B) = P(A) \cdot P(B)$))

**Beispiele:**
- Münzwurf $\{H, Z\}$ (qualitatives Merkmal) → nach einer bestimmten Vorschrift ausgeführt
- würfeln $\{1, 2, 3, 4, 5, 6\}$ → Beliebig oft wiederholbar
- Roulette $\{0, 1, 2, \dots, 36\}$ → liefert zufallsabhängig ein Ergebnis (Ergebnisse schliessen sich gegenseitig aus)
- Kugel ziehen aus Urne → wohldefinierter Ergebnisraum (z.B. $S = \{H, Z\}$)

## 2.2 Ergebnisse und Ereignisse

→ hat ein Ausgang, der nicht exakt vorhergesehen werden kann, denn er ist vom Zufall beeinflusst. (Wiederholung desselben Zufallexperiments = unterschiedliche Ergebnisse)

→ risultato

**Ergebnis:** Jedes mögliche Ergebnis eines Zufallsexperiments. Ergebnisse schliessen sich gegenseitig aus (sind disjunkt)
↳ z.B. Münzwurf, Würfeln, Roulette.
↳ non possono accadere contemporaneamente ($A \cap B = \emptyset$)
↳ o vince Fermat o Pascal, non possono vincere entrambi.

(Elementarereignis). Aber genau eines passiert mit Sicherheit
↳ Elementare Elemente des Ergebnisraumes

Ergebnis: z.B. (Pascal gewinnt, oder Fermat gewinnt)

Ereignis: z.B. H oder Z bei Münzwurf.

(1 o 2 o 3 o 4 o 5 o 6 giocato, due non posson accadere contemporaneamente)

Zwei Ereignisse können gleichzeitig auftreten (z.B. rosso-pari roulette)

Die # der möglichen Ereignisse ist viel höher als den Anzahl der Ergebnisse.

**Ergebnisraum:** Menge aller möglichen Ergebnisse des Experiments

$S = \{1, 2, 3, 4, 5, 6\}$

→ evento

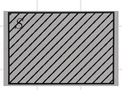
Mengen von Ergebnissen $A \subseteq S$

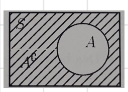
**Ereignis:** Teilmenge des Ergebnisraums (besteht aus keinem, einem oder mehreren Ergebnissen). Zwei Ereignisse können gleichzeitig auftreten. (nicht notwendigerweise disjunkt)

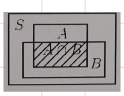
z.B. würfeln einer geraden Zahl $A = \{2, 4, 6\}$ (versch. Ereignisse sind idR. nicht gleich wahrscheinlich)

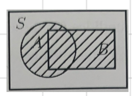
### Spezielle Ereignisse

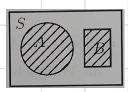

unmögliches Ereignis $A = \emptyset$


sicheres Ereignis $A = S$

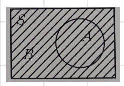

komplementäres Ereignis $A^c = S \setminus A$

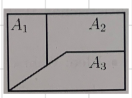

Durchschnitt $A \cap B$


Vereinigung $A \cup B$


Disjunkte Ereignisse $A \cap B = \emptyset$, Beispiel: $A \cap A^c = \emptyset$
(unverträglichkeit hat nichts mit unabhängigkeit zu tun)

disjunkt: $P(A) \cup P(B) = P(A) + P(B)$


Ausschöpfend $A \cup B = \Omega$ $\quad A \cup A^c = \Omega$


Partition des Ergebnisraums: disjunkt + ausschöpfend

## 2.3 Wahrscheinlichkeitsbegriffe

**Laplace Wahrscheinlichkeit (klassisch)**

Prinzip der "Fairness" → alle Ergebnisse sind gleich wahrscheinlich

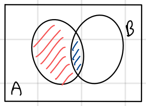
$$\boxed{P(A) = \frac{\text{Anzahl Ergebnisse in } A}{\text{Anzahl aller Ergebnisse}} = \frac{\# A}{\# S}}$$

$P(K) = P(Z) = 0.5$

**Wahrscheinlichkeits-Axiome von Kolmogoroff**

Gesetz der Totalen Wahrscheinlichkeit

$$P(A) = \underbrace{P(A|B) \cdot P(B)}_{P(A \cap B)} + \underbrace{P(A|B^c) \cdot P(B^c)}_{P(A \cap B^c)}$$



$\text{m} + \text{m} = A$

Gemäss Kolmogoroff muss ein Wahrscheinlichkeitsmass folgende drei Eigenschaften (Axiome) erfüllen:

- $P(A)$ ist eine reelle Zahl mit $0 \leq P(A) \leq 1$
- Das sichere Ereignis hat Wahrscheinlichkeit eins: $P(S) = 1$ (Das unmögliche: $P(\emptyset) = 0$)
- Für zwei disjunkte Ereignisse gilt: $P(A \cup B) = P(A) + P(B)$
  ≠ unabhängig
  Spesso se sono disjunkt allora non sono unabhängig (ci sono però eccezioni) → Disjunkte Ereignisse sind nicht unbedingt unabhängig!

$P(\emptyset) = 0$

| | |
|---|---|
| | **Eine Münze wird 9 mal hintereinander geworfen. Welche der folgenden Sequenzen (Ergebnisse) tritt mit höherer Wahrscheinlichkeit auf?** |
| 1: | HHHHHHHHZ |
| 2: | HHHHHHHHH |
| 3: | HHZHZZZHH |
| 4: | Alle Sequenzen sind gleichwahrscheinlich ✗ |

Aus diesen Eigenschaften lassen sich alle wichtigen Rechenregeln ableiten:

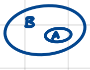
- Das Komplement hat die Wahrscheinlichkeit $\boxed{P(A^c) = 1 - P(A)}$

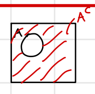
"Odds" = $\frac{P(A) \rightarrow \text{Wahrscheinlichkeit des Eintretens}}{P(A^c) \rightarrow \text{Gegenwahrscheinlichkeit } (1 - P(A))}$

**Beispiel Würfel**:
Bestimmen Sie die Odds, eine 6 zu würfeln.

Odds $= \frac{\frac{1}{6}}{\frac{5}{6}} = \frac{1}{5}$

- Es gilt Monotonie: $A \subseteq B \rightarrow P(A) \leq P(B)$




- Zählregel: $\boxed{P(A) = \sum_{s \in A} P(s)}$ Beispiel: $P(\text{mind. zwei } Z) = 1 - P(ZHHH) - P(HZHH) - P(HHZH) - P(HHHZ) - P(HHHH)$

- Additionssatz: $\boxed{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$
(nicht disjunkte Ereignisse, sonst Schnitt $= 0$)

**Beispiel Würfel**:
Berechnen Sie die Wahrscheinlichkeit, eine gerade Zahl, oder eine Zahl grösser als 4 zu würfeln.

A Gerade Zahl $= 3/6$ $\{2, 4, 6\}$
B grösser als 4 $= 2/6$ $\{5, 6\}$
$A \cap B = 1/6$ $\{6\}$
$P(A \cup B) = \frac{1}{2} + \frac{1}{3} - \frac{1}{6} = \frac{4}{6} = 2/3$

$(A \cup B) = \{2, 4, 5, 6\}$

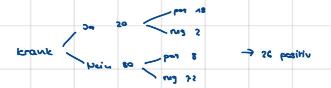
## 2.4 Bedingte Wahrscheinlichkeit

**Abhängigkeit**

Unbedingte Wahrscheinlichkeit: $P(A|B) = \frac{P(A) \cdot P(B)}{P(B)} = P(A)$

z.B. Kugeln aus Urne ziehen ohne zurücklegen

| | | |
|---|---|---|
| | Ein Bluttest sei 90% akkurat: bei 10 getesteten Personen mit Krankheit ist der Test 9 mal positiv; bei 10 getesteten Personen ohne Krankheit ist der Test einmal positiv. Angenommen, 100 Personen werden getestet, von denen 20 wirklich krank sind. | |
| 1: | Dann ist die Anzahl der positiven Tests 18 | |
| 2: ✕ | Dann ist die Anzahl der positiven Tests 26 | |
| 3: | Dann ist die Anzahl der positiven Tests 74 | |

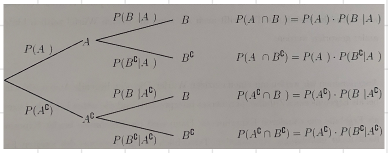



**bedingte Wahrscheinlichkeit:** $\boxed{P(B|A) = \frac{P(A \cap B)}{P(A)}}$ → Wahrsch. für B wenn A bereits eingetreten ist

oder $P(A|B) = \frac{P(A \cap B)}{P(B)}$ → Wahrsch. von A wenn B schon eingetreten ist.

- Ereignis A hat keinen Einfluss auf B: $P(B|A) = P(B)$
  ↳ ist unabhängig

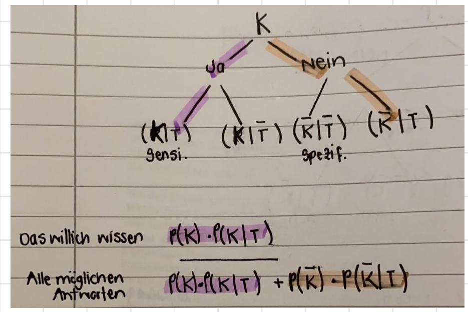
**Wahrscheinlichkeitsbaum** Gute Darstellung der Abhängigkeiten




→ Multiplikationssatz $\boxed{P(A \cap B) = P(A) \cdot P(B|A)}$

Für 3 Ereignisse: $P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2)$

**Satz von Bayes**




Bayes'sche Formel: $\boxed{P(A|B) = \frac{P(A) \cdot P(B|A)}{P(B)}}$
↳ Angaben B (B hat schon stattgefunden)

Umgekehrte bedingte Wahrscheinlichkeit: dass A eintritt, falls B bereits eingetreten ist

**Absolute Wahrscheinlichkeit berechnen**

$P(A) = P(B) \cdot P(A|B) + P(\bar{B}) \cdot P(A|\bar{B})$

## 2.6 Wetten und erwarteter Gewinn

6. Am Zürcher Flughafen bietet man Ihnen für CHF 15.- eine Versicherung für Ihr Gepäckstück im Wert von CHF 250.- an. Die Wahrscheinlichkeit, dass Ihr Gepäck geklaut wird beträgt 0.001. Weiter beträgt die Wahrscheinlichkeit, dass das Flughafenpersonal den Koffer nicht in das Flugzeug einlädt 0.1. Schliessen Sie die Versicherung ab?

$-15.-$ mit Versicherung

$-250 \times 0.001 + (-250) \times 0.1 = \underline{-25.25}$ ohne Versicherung

→ Ja, ich schliesse eine Versicherung ab

Der erwartete Gewinn ergibt sich aus der Summe der möglichen Zahlungen $X(\omega)$ multipliziert mit der jeweiligen Wahrscheinlichkeit.

**Berechnung erwarteter Gewinn**:

$$\text{Erwarteter Gewinn} = \sum_{\omega \in \Omega} X(\omega) \times P(\omega)$$

## 2.7 Unabhängigkeit / unabhängige Ereignisse

Dies ergibt sich aus der Anforderung nach Fairness/Gleichwahrscheinlichkeit beim erweiterten Spiel

Ereignisse A und B sind unabhängig falls $\boxed{P(A \cap B) = P(A) \cdot P(B)}$ Multiplikationssatz (Produktregel)

$\Rightarrow P(A|B) = P(A)$ oder $P(B|A) = P(B)$ $\quad P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$

Beispiel: • Zweifacher Münzwurf;
Die Wahrscheinlichkeit für den zweiten Wurf ist unabhängig vom ersten

• Zweimaliges Würfeln: es gibt $6^2 = 36$ verschiedene Kombinationen

Auch wenn einzelne Ergebnisse eine hohe Wahrscheinlichkeit haben (z.B. 1/2) ist die Wahrscheinlichkeit des *gemeinsamen* Auftretens solcher Ergebnisse bei zusammengesetzten Zufallsexperimenten sehr klein (z.B. $1/2^4 = 0.0625$ bei 4 Experimenten) *vorausgesetzt sie sind auch wirklich unabhängig*. Bei positiver Abhängigkeit/Korrelation (mehr dazu später) kann die Wahrscheinlichkeit viel höher liegen, bis zu 1/2.

Beispiel 1:

Bei einem 4-stelligen Smartphone PIN gibt es an jeder Stelle 10 Möglichkeiten $0, 1, \dots, 9$, bei zufälliger Auswahl je mit Wahrscheinlichkeit 1/10; damit ist die Wahrscheinlichkeit einer bestimmten Kombination $(1/10)^4$, also 0.0001 oder 1 in 10'000.

Beispiel 2: Die Wahrscheinlichkeit, dass ein Triebwerk während eines Transatlantik-Fluges ausfällt, ist 1 pro 100'000, ein unakzeptables Risiko. Bei zwei Triebwerken sinkt die Wahrscheinlichkeit, dass beide ausfallen, auf $(1/100'000)^2$, oder 1 in 10 Milliarden. Das klingt o aber sind die Ereignisse wirklich unabhängig?

Beispiel 3: Die Folgen der Vernachlässigung von positiven Korrelationen (oder "systemischen" Risiken) wurde in der Finanzkrise von 2008 schmerzhaft deutlich.

**Gamblers Fallacy**

Wegen Unabhängigkeit haben die Sequenzen HHHHHHHHZ und HHHHHHHHH die gleiche Wahrscheinlichkeit

Dies widerspricht der Intuition "wenn H so häufig aufgetreten ist, muss Z folgen"

Eine andere Version dieses Irrtums: die meisten Menschen empfinden, dass die Sequenz *HHZHZZZHH* "zufälliger" aussieht, als *HZHZHZHZH* oder *HHHHHHHHH*; dabei haben alle die gleiche Wahrscheinlichkeit! (nämlich $0.5^9 = 0.00195$) [1]

**Hot hand fallacy**

Die Erfolgen haben nach einer Sequenz mehrere Erfolgen die gleiche Wahrscheinlichkeit. Dies widerspricht der Intuition "Wenn ich mehrere Erfolgen mache, dass ist es wahrscheinlicher, dass ich einen weiteren Erfolg mache."

[1] Es ist korrekt, dass *irgendeine* Sequenz mit $5 \times H$ wahrscheinlicher ist, als $9 \times H$; dazu später mehr.

# 3.1 Diskrete Zufallsvariabeln

Meist sind wir nicht an den Ergebnissen direkt interessiert, sondern an einem Zahlenwert der von dem Ausgang abhängig ist.

**Beispiele:**
- Augensumme mehrerer geworfenen Würfel
- Anzahl Würfe die man braucht, bis man eine Zwei würfelt

Dieses Konzept wird durch die Zufallsvariable realisiert

**Zufallsvariable:** eine Funktion $X : S \to \mathbb{R}$ die Elementen des Ergebnisraumes $S$ reelle Zahlen zuordnet

↳ variabile casuale

## 3.1 Diskrete Zufallsvariabeln

**Realisationen:** Die zugeordneten Werte $(x)$ in $\mathbb{R}$

Wir sprechen von der Wahrscheinlichkeit $P(X=x)$ dass die Zufallsvariable $X$ den konkreten Wert $x$ annimmt

$(X = x \;\; = \;\; A \subseteq S)$

**diskrete Zufallsvariable:** nimmt endlich oder abzählbare Werte an (Augensumme zweifacher Würfelwurf) ⇒ Schuhgrösse, Alter in Jahren, Anzahl Tore bei einem Fussballspiel, Studenten an der UZH.

**stetige Zufallsvariable:** kann alle Werte in einem Intervall annehmen (Gewicht eines Studenten)

## 3.2 Wahrscheinlichkeitsfunktion und Verteilungsfunktion

**Wahrscheinlichkeitsfunktion:** $\boxed{P(X=x) = \sum P(\{s\})}$

- Wahrscheinlichkeiten nehmen immer Werte zwischen 0 und 1 an: $0 \leq P(X = x_i) \leq 1$
- Summe der Wahrscheinlichkeiten für alle Werte ist 1: $\sum_i P(X=x_i) = 1$

**Beispiel:** Anzahl $X$ der Würfe bei denen Kopf erscheint

Ergebnisraum des Experiments: $S = \{KK, KZ, ZK, ZZ\}$

Die Zufallsvariable ordnet diesen Ergebnissen folgende Zahlenwerte zu:

- $X(KK) = 2$ • $X(ZK) = 1$
- $X(KZ) = 1$ • $X(ZZ) = 0$

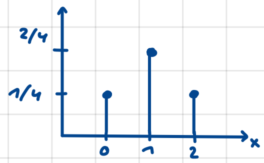
Jedes Ergebnis hat die gleiche Wahrscheinlichkeit $1/4$. Die Wahrscheinlichkeitsverteilung von $X$ wird aber durch die Zahlregel bestimmt und ist nicht gleichverteilt.

- $P(X=0) = 1/4 \Rightarrow (1-p)^2$
- $P(X=1) = 2/4 \Rightarrow p(1-p) + p(1-p)$ $\longrightarrow \sum_{i=0}^{2} (X=i) = P(X=0) + P(X=1) + P(X=2) = 1$
- $P(X=2) = 1/4 \Rightarrow p^2$

$$\quad 1/4 \; + \; 2/4 \; + \; 1/4$$

**Stabdiagramm**

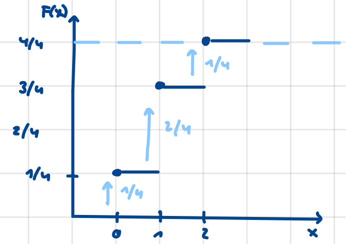
Mit ihm kann die Wahrscheinlichkeitsfunktion dargestellt werden




Wahrscheinlichkeiten nicht gleichverteilt!

**Verteilungsfunktion**

Häufig interessant: Zufallsvariable nimmt Werte an, die nicht grösser als ein vorgegebener Wert sind.




Dazu wird die Verteilungsfunktion definiert: $F(x) = P(X \leq x) = \sum_{x_i \leq x} P(X = x_i)$

## 3.3 Erwartungswert und Varianz

Erwartungswert und Varianz charakterisieren eine Zufallsvariable anhand der Lage- und Streuparameter. Sie sind nicht vom Zufall abhängig, sondern konkrete Zahlen, die das Verhalten von X beschreiben.

### Erwartungswert

Mittel aller möglichen Werte $x_i$

$$\boxed{\mu_X = E(X) = \sum_i x_i \cdot P(X = x_i)}$$

**Beispiel** (Münzwurf): $\sum_{i=0}^{2} i\, P(X=i) = 0 \cdot 1/4 + 1 \cdot 2/4 + 2 \cdot 1/4 = 1$

Der Erwartungswert ist linear:

- $E(X+Y) = E(X) + E(Y)$ (Beispiel Summe zweier Münzwürfe: $1+1=2$)
- $E(a+bX) = a + b\,E(X)$

Sie zahlen zuerst CHF 11, um spielen zu dürfen. Dann werfen Sie einen Würfel und bekommen die dreifache Augenzahl ausgezahlt. (Netto-)Gewinn = $G$; Gewürfelte Augenzahl = $X$.

$a = -11$, $b = 3$ $\rightarrow$ $-11 + 3 \cdot E(X) = -11 + 3 \cdot 3.5 = -0.5$

### Varianz

Quadrierte Differenz zwischen Zufallsvariabeln X und dem Mittelwert $\mu_X$:

$$\boxed{\sigma_X^2 = Var(X) = E((X-\mu)^2) = \sum_i (x_i - \mu_X)^2 \cdot P(X = x_i)}$$

oder

$$\boxed{\sigma^2 = E(X^2) - (E(X))^2}$$

**Beispiel**: $\sigma^2 = \sum_{i=0}^{2} (i-\mu)^2\, P(X=i)$
$= (0-1)^2 \cdot 1/4 + (1-1)^2 \cdot 2/4 + (2-1)^2 \cdot 1/4$
$= 1/4 + 0 + 1/4$
$= 2/4$

Die Varianz ist quadratisch:

- $Var(X+Y) = Var(X) + Var(Y)$ ! Nur falls X und Y unabhängig sind $Var(X-Y) = Var(X) + Var(Y)$
- $Var(a+bX) = b^2 \cdot Var(X)$

**Allgemeines Beispiel:**

- $E(aX + bY + c) = a\,E(X) + b\,E(Y) + c$
- $Var(aX + bY + c) = a^2\, Var(X) + b^2\, Var(Y)$ ($\sigma^2$, $\sigma^2$)
- $SA(aX + bY + c) = \sqrt{a^2\, Var(X) + b^2\, Var(Y)}$ ($\sigma^2$, $\sigma^2$)

### Standardabweichung

- $SA(X+Y) = \sqrt{2/4 + 2/4} = 1$
- $SA(a+bX) = |b| \cdot SA(X)$

$$\boxed{\sigma_X = SA(X) = \sqrt{Var(X)}}$$

### Gesetz der grossen Zahl

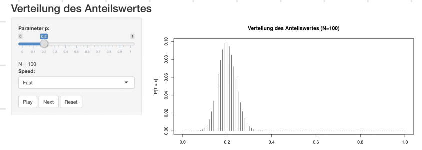
Die Varianz und die SA sind nicht von der konkreten Lage der Zufallswerte abhängig.

Die Kenngrössen ändern sich nicht: $\boxed{\sigma^2(X+a) = \sigma^2(X)}$ $\boxed{\sigma(X+a) = \sigma(X)}$

dies ist durch das Gesetz der Grossen Zahl garantiert

Più è grande N, più la probabilità si avvicina a quella reale.

Verteilung des Anteilswertes




Sei $X_i$ unabhängig $Bernoulli(p)$-verteilt. Dieses Beispiel zeigt die theoretische Verteilung des Stichprobenanteils

$$\hat{p} = \frac{1}{N}\sum_{i=1}^{N} X_i$$

für $N = 1, 2, \ldots, 100$.

Wir sehen, wie die Verteilung von $\hat{p}$ für steigendes $N$ gegen das wahre $p$ konvergiert.

Download R code (estimatordistribution.R)

Se p è 0.5 dopo 100 tentativi esce picco più grande a 0.5

## 3.4 Bernoulliverteilung

Voraussetzung ist, dass ein Zufallsexperiment nur zwei mögliche Ausgänge besitzt:

- Erfolg $= 1 = p$ → Wahrscheinlichkeit
- Misserfolg $= 0 = 1-p$ → Wahrscheinlichkeit

| 0 | 1 |
|---|---|
| $1-p$ | $p$ |

$X \sim$ Bernoulli $(p)$

$$\boxed{P(X=x) = p^x (1-p)^{1-x} \qquad x = 0,1}$$

**Erwartungswert:** $E(X) = \mu_X = p$

**Varianz:** $Var(X) = \sigma_X^2 = p(1-p)$

**Standardabweichung:** $SA(X) = \sigma_X = \sqrt{p(1-p)}$

| $x$: | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $P(x)$: | .2 | .3 | .2 | .1 | .1 | .1 |

The expected value of $X$ is

$$E\{X\} = (1)(.2) + (2)(.3) + (3)(.2) + (4)(.1) + (5)(.1) + (6)(.1) = .2 + .6 + .6 + .4 + .5 + .6 = 2.9.$$

where $\sigma^2\{\ \}$ is called the *variance operator*. The calculation of the variance of the length of hospital stay can be organized in the table below:

| $x$: | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $P(x)$: | .20 | .30 | .20 | .10 | .10 | .10 |
| $x - E\{X\}$: | -1.90 | -.90 | .10 | 1.10 | 2.10 | 3.10 |
| $(x - E\{X\})^2$: | 3.61 | .81 | .01 | 1.21 | 4.41 | 9.61 |

$1-2.9$ $2-2.9$

from which

$$\sigma^2\{X\} = (3.61)(.2) + (.81)(.3) + (.01)(.2) + (1.21)(.1) + (4.41)(.1) + (9.61)(.1) = .722 + .243 + .002 + .121 + .441 + .961 = 2.49.$$

### "Fairness" der Münze

Beeinflusst den Erwartungswert und die Streuung

→ Je unfairer die Münze, desto weniger die Streuung

Die Extremfälle $p=0$ und $p=1$ sind auch zulässig, dann ist X sicher vorhersagbar mit $\sigma_X = \sigma_X^2 = 0$

## 3.5 Binomialverteilung → immer eingipflig!

z.B. Münzwurf, Würfeln

Diese Verteilung beschreibt die Anzahl der Erfolge, wenn man das Bernoulli-Experiment mehrfach ausführt (Anzahl Erfolge wird gezählt)

- $n$ : Anzahl unabhängiger, gleichartiger Versuche
- $p$ : Erfolgswahrscheinlichkeit pro Einzelversuch
- $X$ : Anzahl der Erfolge nach $n$ Versuchen

$k \rightarrow$

| 1 | 2 | 3 | ... | N |
|---|---|---|---|---|
| $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ |

$X \sim$ Binomial $(N, p)$

Werteberech $S$ umfasst $N+1$ Elemente

$\hookrightarrow W = \{0, 1, \ldots, n\}$

$$\boxed{P(X=k) = \binom{N}{k} p^k (1-p)^{N-k}}$$

$k$ Erfolge

**Binomialkoeffizient** :

$$\boxed{\binom{n}{k} = \frac{n!}{k!(n-k)!}}$$

$n$ → Anzahl Versuche, $k$ → Anzahl Erfolge

$\binom{10}{0} = 1$ $\binom{10}{1} = 10$

Hohe Binomialkoeffizienten sollte man kürzen:

$$\binom{50}{48} = \frac{50!}{48! \cdot 2!} = \frac{50 \cdot 49 \cdot \overbrace{48 \cdots 2 \cdot 1}^{\text{kürzen}}}{\underbrace{(48 \cdot 47 \cdots 2 \cdot 1)}_{\text{kürzen}} \cdot (2 \cdot 1)} = \frac{50 \cdot 49}{2} = 1'225$$

**Erwartungswert** : $\mu_X = E(X) = np$

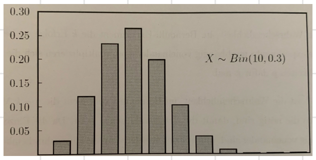
**Verteilungsfunktion** : $F(x) = P(X \leq x) = \sum_{k=0}^{x} \binom{n}{k} p^k (1-p)^{n-k}$

**Varianz** : $\sigma_X = Var(X) = np(1-p)$

Wertebereich $S$ umfasst die Werte $N+1$

**Standardabweichung** : $\sigma_X^2 = SA(X) = \sqrt{np(1-p)}$

**Grafisch**




Bei dem Mittelwert 3 ist die Wahrscheinlichkeit am grössten

- Wenn der Parameter $p$ vergrössert wird, verschieben sich die maximalen Werte weiter nach rechts
- Wenn die Anzahl Versuche erhöht wird, nähert sich die Form einer Glocke an

## 3.6 Poissonverteilung → es gibt keine natürliche Obergrenze (Mögliche Ergebnisse von 0 bis $\infty$, $x \in \mathbb{N}_0$)

Variable can take any integer from 0 to + infinity

Diese Verteilung wird verwendet, wenn die Erfolgswahrscheinlichkeit pro Einzelexperiment sehr klein ist und beliebig viele Erfolge möglich sind (Geburte, Todesfälle, Unfälle, wobei jeder Tag einen Versuch darstellt)

- $\mu$ : Anzahl Erfolge im Schnitt $> 0$ (parameter)
- Mögliche Ergebnisse : $0, 1, 2, 3, \ldots$

$X \sim$ Poisson $(\mu)$

$$\boxed{P(X=x) = \frac{\mu^x \cdot e^{-\mu}}{x!}}$$

**Erwartungswert** : $\mu_X = E(X) = \mu$ → Lage (→ posizione) der Verteilung der Zufallsvariablen $X$

$E(X) = \sum_{x \in W} x P(X=x)$

$x$: Möglicher Wert, $W$: Wertebereich, $P(X=x)$: Wahrscheinlichkeit des Wertes $x$

**Varianz** : $\sigma_X^2 = Var(X) = \mu$ → Streuung (sparpimento) der Verteilung der Zufallsvariablen $X$. $Var(X) = \sigma_X^2 = E(X - E(X))^2 = \sum_{x \in W} (x - \mu_X)^2 P(X=x)$

**Standardabweichung** : $\sigma_X = SA(X) = \sqrt{\mu}$ $SA(X) = \sigma_X = \sqrt{\sigma_X^2}$

- $S = X + Y$ ($X, Y$ gleichverteilt) gilt nicht bei Poisson-Verteilung.
- $E(a + bX) = a + bE(X)$ • $Var(a + bX) = b^2 Var(X)$
- Falls $X, Y$ unabhängig: $Var(X+Y) = Var(X) + Var(Y)$

**Gegenereignis-Regel**

Die Versuche brechen nicht ab.

Deshalb sollte man bei $P(X \geq x)$ die Gegenereignis-Regel verwenden :

$$\boxed{P(X \geq x) = 1 - P(X < x)}$$

Für die bekannten Verteilungen ergeben sich folgende Erwartungswerte und Varianzen:

→ Bernoulli Verteilung mal Nummer

| | Bernoulli | Binomial | Poisson |
|---|---|---|---|
| Erwartungswert | $p$ | $n * p$ | $\mu$ |
| Varianz | $p(1-p)$ | $n * p(1-p)$ | $\mu$ |

↳ kann maximal 0.25 sein

| | Sei X Poisson-verteilt mit Mittelwert (und Varianz) 2; Y Bernoulli-verteilt mit p = 0.5. Welche der folgenden Aussagen trifft NICHT zu? | |
|---|---|---|
| 1: | X + Y ist Poisson-verteilt. | X |
| 2: | E(X + Y) = 2.5 | |
| 3: | E(2 + Y) = 2.5 | |
| 4: | Var(2 + 2X) = 8 | |

| | Welche der folgenden Aussagen trifft zu? | |
|---|---|---|
| 1: | Der Wertebereich der Poisson-Verteilung sind die ganzen Zahlen. | |
| 2: | Die Binomialverteilung ist immer symmetrisch. | |
| 3: | Die Binomialverteilung kann zweigipflig sein. | |
| 4: | Für jede beliebige Zufallsvariable gilt: P(X<=a) = 1 - P(X>a). | X |

| Bei welchem der folgenden Fälle liegt eine Binomial-Verteilung vor? | |
|---|---|
| Die Anzahl der Versuche, die es beim Würfeln braucht, bis erstmals die 6 erscheint. | |
| Unter 20 zufällig ausgewählten Personen die Anzahl derjenigen, die grösser als 1 Meter 80 sind. | X |
| Die Anzahl der Tage, die ein Arbeitnehmer während einer Woche fehlt. | |

# 3.2 Stetige Zufallsvariablen

Partielle Integration

$$\int_a^b \underbrace{g(x)}_{x}\,\underbrace{\ell'(x)}_{\lambda\exp(-\lambda x)}\,dx = \left[\ell(x)\,g(x)\right]_a^b - \int_a^b \ell(x)\,g'(x)\,dx$$

## 3.7 Stetige Zufallsvariablen

Eine **stetige Zufallsvariable** kann jeden Wert in einem bestimmten Intervall annehmen

z.B. Grössen mit Zeit, exaktes Gewicht.

Sie ist eine Funktion $X: S \to \mathbb{R}$ von einem Ergebnisraum $S$ in die Reellen Zahlen

$a \le x \le b$ → $b$: obere Grenze, $a$: untere Grenze

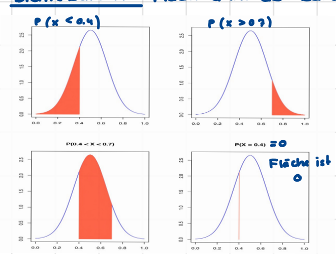
## 3.8 Dichtefunktion und Verteilungsfunktion

Es gibt zu viele mögliche Ergebnisse, um sie aufzulisten → es ist unmöglich eine Wahrscheinlichkeitsfunktion zu definieren, da die Wahrscheinlichkeit für jeden einzelnen Wert Null ist $P(X=x)=0$

**Dichtefunktion** $f$ — Verteilung (von stetige ZV) wird deshalb durch eine Dichtefunktion $f$ beschrieben

$f(x) \ge 0$ und $\boxed{\int_{-\infty}^{\infty} f(x)\,dx = 1}$

Dichtekurven: Fläche unter der Kurve ist die Wahrscheinlichkeit vom Ereignis




Die Wahrscheinlichkeit entspricht der Fläche:

$$\boxed{P(\underbrace{a}_{\text{untere Grenze}} \le x \le \underbrace{b}_{\text{obere Grenze}}) = \int_a^b f(x)\,dx = F(b) - F(a)}$$

**Verteilungsfunktion** Die Verteilungsfunktion beschreibt die Integralwerte an den Grenzen des Intervalls

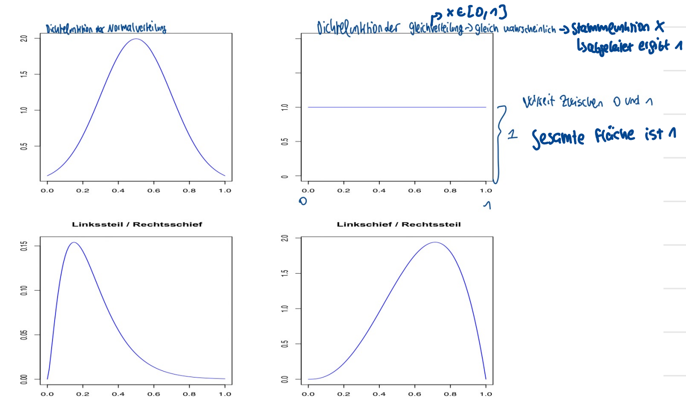
$$\boxed{F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)\,dt} = F(b) - F(a) \qquad f = F'$$

kennen wir die Stammfunktion, genügt es die Werte an den Rändern zu bestimmen

## 3.9 Erwartungswert und Varianz

**Erwartungswert** Mittelwert einer stetigen Zufallsvariable $X$ ist der gewichtete Mittelwert aller möglichen Werte $x$

$$\boxed{\mu_x = E(x) = \int_{-\infty}^{\infty} x\,f(x)\,dx}$$




**Varianz**

$$\boxed{\sigma_x^2 = Var(x) = \int_{-\infty}^{\infty} (x-\mu_x)^2 f(x)\,dx}$$

vereinfacht:

$$\boxed{Var(x) = E(x^2) - \mu_x^2}$$

**Standardabweichung**

$$\boxed{\sigma_x = SA(x) = \sqrt{\sigma_x^2}}$$

**Perzentile** → z.B 0.01, 0.02, ..., 0.49, 0.50, ... 0.99, 1.00

Das $100 \times p$te Perzentil $x_p$ erfüllt folgende Bedingung: $P(X \le x_p) = p$ $\quad p \in (0,1)$

gibt ein Wert an, es ist keine Wahrscheinlichkeit !

Es teilt die Dichtefunktion in zwei Bereiche:

- Wahrscheinlichkeit links von $x_p$ beträgt $p$
- Wahrscheinlichkeit rechts von $x_p$ beträgt $1-p$

$$\boxed{F(x_p) = \int_{-\infty}^{x_p} f(x)\,dx = p}$$

**Spezialfall Median:** Der Median ist das 50te Perzentil → $P(x \le \text{Median}) = 50\% = 1/2$

- Der Erwartungswert ist der Balance-Punkt der Fläche unter $f(x)$
- Der Median teilt die Fläche 50-50 auf → er ist das 50.te Perzentil einer Verteilung

**Exponentialverteilung**

Die Stammfunktion der Dichte $f(x)$ ist gegeben als $\tilde{F}(x) = -\exp(-\lambda x)$. Es folgt

$$\int_0^\infty f(x)\mathrm{d}x = \int_0^\infty \lambda\exp(-\lambda x)\mathrm{d}x = -\exp(-\lambda x)\Big|_0^\infty = \lim_{x\to\infty}[-\exp(-\lambda x)] - (-1) = 1.$$

The Poisson probability distribution applies to the number of occurrences in a time interval. The *exponential* probability distribution applies to the amount of time between occurrences. For this reason it is often called the *waiting-time distribution.* It is a continuous distribution because time is measured along a continuum. An *exponential random variable* $X$ is the time between occurrences of a random event. The probability density function is

$$f(x) = \lambda e^{-\lambda x} = \lambda\exp(-\lambda x), \qquad (x > 0). \tag{3.41}$$

It turns out that the probability that $X \ge x$ is

$$P(X \ge x) = e^{-\lambda x}. \tag{3.42}$$

The mean and variance of an exponential distribution are

$$E\{X\} = \frac{1}{\lambda}$$

$$SA(X) = \frac{1}{\lambda}$$

and

$$\sigma^2\{X\} = \frac{1}{\lambda^2}.$$

Für ein $c > 0$ gilt:

$$P(x \le c) = \int_0^c \lambda\exp(-\lambda x)\,dx = -\exp(-\lambda x)\Big|_0^c = 1 - \exp(-\lambda c)$$

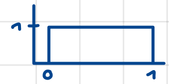
The shape of the exponential distribution is governed by the single parameter $\lambda$. As indicated in the plots of some exponential distributions in Figure 3.11, the exponential probability density function declines as $x$ increases from zero, with the decline being sharper the greater the value of $\lambda$. The probability density function intersects the y-axis at $\lambda$.

## 3.10 Die Gleichverteilung → probability of all occurences in the sample space are the same (probability distributions can be either discrete or continuous)

Beschreibt Zufallsexperimente mit einer reellen Zahl als Ergebnis, die stets in einem Intervall $[a, b]$ liegt und dort jeden Wert mit gleicher Wahrscheinlichkeit annimmt $X \sim U[a,b]$

**Dichtefunktion** $\boxed{f(x) = \frac{1}{b-a}}$ (für $a \le x \le b$) sonst $f(x) = 0$

Fläche unter der Dichtefunktion ist ein Rechteck mit Flächeninhalt 1:




Consider a computer random number generator that cranks out random numbers between 0 and 9. By construction of the computer program, the probability that any one of the 10 numbers will be turned up is 1/10 or 0.1. The probability distribution for this process is therefore

| $x$: | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| $P(x)$: | .1 | .1 | .1 | .1 | .1 | .1 | .1 | .1 | .1 | .1 |

This random variable is called a *discrete uniform random variable* and its probability distribution is a *discrete uniform probability distribution*. The discrete probability function is

$$P(x) = \frac{1}{s} \tag{3.33}$$

Erwartungswert: $\mu_x = \frac{1}{2}(b+a)$

Varianz: $\sigma_x^2 = \frac{1}{12}(b-a)^2$

Standardabweichung: $\sigma_x = \sqrt{\frac{1}{12}(b-a)^2}$

**Beispiel:** $[1,3]$ $P(x \le 2)$?

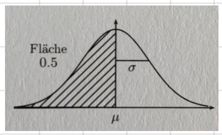
$P(x \le 2) = \int_{-\infty}^{2} f(x)dx = \int_{1}^{2} \frac{1}{2} dx = 1 - 1/2 = 0.5$

→ most important

## 3.11 Normalverteilung und Standardnormalverteilung → beschreibt z.B. Aktienrendite, Studienergebnisse, Geburtsgewicht, IQ von Personen

Beschreibt die meisten natürlichen Zufallsphänomene und Stichprobenvariablen, die den Grenzwertsatz erfüllen (Beschreibung Variation von Studienergebnissen / Schätzresultate) $X \sim Normal(\mu, \sigma^2)$

**Dichtefunktion** $\boxed{f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}$




Form einer Glockenkurve, Funktion bestimmt durch $\mu$ und $\sigma^2 > 0$

$\mu$: Lage des Maximums

$\sigma$: Breite der Glocke

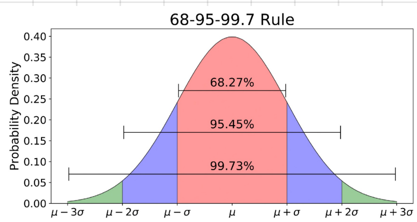
**Standardnormalverteilung**

$\mu = 0$ $\sigma^2 = 1$

↳ = Median (wegen Symmetrie)

**Die 68 - 95 - 99.7 % - Regel**

- Mit Wahrscheinlichkeit 68% liegt $X$ in dem Interval $\mu \pm \sigma$.
- Mit Wahrscheinlichkeit 95% liegt $X$ in dem Interval $\mu \pm 2\sigma$.
- Mit Wahrscheinlichkeit 99.7% liegt $X$ in dem Interval $\mu \pm 3\sigma$




**Berechnung der Wahrscheinlichkeiten**

Gesucht: $P(x < a)$ ⟶ $F(a) = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$ ⟶ Jedoch Integration von Hand ist unmöglich!

Deswegen werden die Werte per Computer approximiert und in Tabellen erfasst

Um die Tabellen zu benutzen muss man die normalverteilte Zufallsvariable $X \sim N(\mu, \sigma^2)$ in die standardisierte Zufallsvariable $Z \sim N(0,1)$ umwandeln

Mit Standardisierung können wir jede beliebige ZV so umwandeln, dass sie den Erwartungswert 0 und die Varianz 1 ist. $E(Z)=0$ $Var(Z)=1$ gilt immer!

$\boxed{Z = \frac{X-\mu}{\sigma}}$ → standardnormal $Z \sim N(0,1)$

$\boxed{Z = -\frac{\mu}{\sigma} + \frac{1}{\sigma}X \quad \text{and} \quad X = \mu + \sigma Z}$

$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{N}}} = \frac{N\hat{p} - Np}{\sqrt{Np(1-p)}} \sim Normal(0,1)$

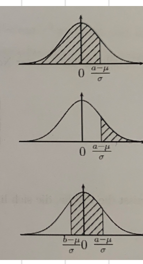
$E\left(\frac{X-E(X)}{\sqrt{Var(X)}}\right) = 0 = E\left(\frac{X-\mu}{\sigma}\right)$; $Var\left(\frac{X-E(X)}{\sqrt{Var(X)}}\right) = 1 = Var\left(\frac{X-\mu}{\sigma}\right)$

Verteilungsfunktion ist tabelliert: $P(X < a) = \left(\frac{X-\mu}{\sigma} < \frac{a-\mu}{\sigma}\right) = P\left(Z < \frac{a-\mu}{\sigma}\right)$

$\boxed{P(X \le a)} = P\left(Z \le \frac{a-\mu}{\sigma}\right) = \Phi\left(\frac{a-\mu}{\sigma}\right)$

$\boxed{P(X \ge a)} = 1 - P(X \le a) = 1 - \Phi\left(\frac{a-\mu}{\sigma}\right)$

$\boxed{P(b \le X \le a)} = P\left(\frac{b-\mu}{\sigma} \le Z \le \frac{a-\mu}{\sigma}\right) = \Phi\left(\frac{a-\mu}{\sigma}\right) - \Phi\left(\frac{b-\mu}{\sigma}\right)$




Beispiel: $X \sim Normal(750, \underbrace{25}_{\sigma^2})$ [Inhalt einer Flasche Wein ...]

$$P(X < 744) = P\left(\frac{X-750}{5} < \frac{744-750}{5}\right) = P(Z < -1.2) \overset{\text{Tab.}}{=} 0.115$$

$\boxed{P(X \le -n) = 1 - P(X \le n)}$

↳ wegen Symmetrie der Normalverteilung

$P(|X-\mu| \le 5.6) = P(X \le 77.6) = P(Z \le 0.35)$ → 5.6 + μ → Standardisierung

**Berechnung Perzentile**

$\boxed{P(Z \ge -x) = P(Z \le x)}$ $\boxed{1 - P(Z \le -x) = P(Z \le x)}$

$\boxed{\begin{aligned} P(-x \le Z \le y) &= P(Z \le y) - P(Z \le -x) \\ &= P(Z \le y) - (1 - P(Z \le x)) \end{aligned}}$

$\boxed{P(x \le Z \le y) = P(Z \le y) - P(Z \le x)}$

Man berechnet die standardisierte Variable $Z \sim N(0,1)$, berechnet deren Perzentil und rechnet es auf $X$ zurück

Das $100 \times p$te Perzentil von Normal$(0,1)$ $z_p$ findet sich in einer Tabelle:

$p = P(Z < z_p) = P\left(\frac{X-\mu}{\sigma} < z_p\right) = P(X < \mu + \sigma z_p)$ ⟶ $\boxed{X_p = \mu + \sigma z_p}$ → von Normalverteilung zu Perzentil

Beispiel: Was ist $x_{0.90}$ von $X \sim Normal(750, 25)$?

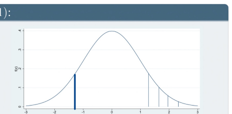
- Tabelle: $z_{0.90} = 1.282 \Longrightarrow x_{0.90} = 750 + 5 \times 1.282 = 756.41$
- Somit: 90% aller Flaschen enthalten höchstens 756.41 ml

Die wichtigsten Perzentile von *Normal*$(0,1)$:

| $p$ | 0.90 | 0.95 | 0.975 | 0.99 |
|---|---|---|---|---|
| $z_p$ | 1.282 | 1.645 | 1.960 | 2.326 |

Wegen Symmetrie:

| $p$ | 0.10 | 0.05 | 0.025 | 0.01 |
|---|---|---|---|---|
| $z_p$ | −1.282 | −1.645 | −1.960 | −2.326 |




Wegen 68% Regel: $z_{0.16} = ?$ −1 $z_{0.84} = ?$ +1

| | Welche der folgenden Aussagen trifft zu? | |
|---|---|---|
| 1: | Das Alter in einer Bevölkerung ist annäherungsweise normalverteilt. | |
| 2: | Bei einer stetigen Zufallsvariablen ist die Wahrscheinlichkeit P(X=a) gleich Null. → jedes Individuellen Ergebnis hat W'keit null. | X |
| 3: | Eine normalverteilte Zufallsvariable liegt mit 95 Prozent Wahrscheinlichkeit in dem Intervall [-1.96, +1.96] → muss um μ und σ kümmern. Nur wenn E(X)=0 und σ=1 ∈[-2;2] | |
| 4: | Das 25ste Perzentil gibt die Wa~~h~~rscheinlichkeit an, mit der eine normalverteilte Zufallsvariable einen Wert kleiner als 0.25 annimmt. ↳ gibt nicht eine Wahrscheinlichkeit, gibt einen Wert | |

# 4 Stichprobenverteilung

## 4.1 Zufallsstichproben

→ bei kategorischen Merkmal mit zwei Ausprägungen: $E(X)=p$. Bei einem metrischen Merkmal: $E(x)=\mu$ ($p$ und $\mu$: unbekannt, sind Parameter der Populationsverteilung)

Eine Zufallsstichprobe ist die wiederholte Zufallsauswahl (unabhängige "Ziehung") aus der Population. Alle Elemente der Population ("Merkmalsträger") haben die gleiche Wahrscheinlichkeit, ausgewählt zu werden

Da die Stichproben zufällig ausgewählt werden, sind auch $\hat{p}$ und $\bar{x}$ zufällig

In einer Stichprobe der Grösse N ist:

<u>Stichprobe</u>: ein Teil der Population ist betrachtet und zufällig ausgewählt

$x_1$ : das gemessene Merkmal bei der ersten Merkmalsträgerin (repräsentiert die Ziehung des ersten Elements)

$x_2$ : das gemessene Merkmal beim zweiten Merkmalsträger (Ziehung des zweiten Elements)

...

| | Stichproben: welche Aussage trifft zu? | |
|---|---|---|
| 1: | Eine Stichprobe mit weniger als 10 Beobachtungen ist nicht aussagekräftig. → falsch, nicht genau/eine Aussage kräftig aber man kann nicht sagen dass sie gar nicht aussagekräftig ist | |
| 2: | Die Verallgemeinerung von der Stichprobe auf die Population nennt man 'Deduktion'. → wäre Induktion | |
| 3: | Eine Stichprobe, in der es zweimal so viele Männer wie Frauen gibt, kann nicht zufällig sein. → falsch, kann zufällig sein | |
| 4: | Bei einer ZSP haben alle Elemente der Population die gleiche Wahrscheinlichkeit, ausgewählt zu werden. richtig | X |

→ $x_1, x_2, \dots, x_N$ (Zufallsvariablen) → diese haben die gleiche Verteilung und sind unabhängig voneinander

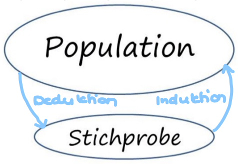
→ Alle Elemente der Population haben die gleiche Wahrscheinlichkeit ausgewählt zu werden

Zufallsstichproben stellen Repräsentativität sicher. Aber im Einzelfall kann die Verteilung eines Merkmals in der Stichprobe aber stark von der Population abweichen.

### Deduktion vs. Induktion

<u>Deduktion</u>: Aus Aussagen über die Population, folgen Aussagen über die Stichprobe

z.B Verteilung des Stichprobenmittels (ZGS $N \geq 50$, $\bar{x} \sim Normal(\mu, \frac{\sigma^2}{N})$)




<u>Induktion</u>: Aus Aussagen über die Stichprobe folgt eine Verallgemeinerung auf eine Population

z.B. Hypothesentest, Schätzen

## 4.2 Verteilung von Stichprobenmittelwert (SPM) und Stichprobenanteil (SPA)

SPM → = Mittelwert in der Stichprobe, um $\mu$ zu schätzen

SPA → oder Anteilswert in der Stichprobe (SPA), $\hat{p}$ um $p$ zu schätzen

### Verteilung des arithmetischen Mittelwerts

$$\boxed{\bar{x} = \frac{1}{N} \sum_{i=1}^{n} X_i = \frac{1}{N}(x_1, x_2, \dots, x_n)}$$

↳ Summe = Anzahl Erfolge bei unabhängigen Bernoulli-Experiment

$\bar{x}$ bezeichnet man als:

- Stichprobenmittel (SPM)
- Stichprobenanteil (SPA), wenn $X \in \{0,1\}$
  ↳ $N \times \hat{p} \sim Bin(N, p)$

Uns interessiert nun, wie SPM und SPA verteilt sind, wenn man wiederholt neue Stichproben ziehen würde, für die jeweils ein $\bar{x}$ berechnet werden kann

**bekannte Verteilung:** Solange die Verteilung von X bekannt ist, lässt sich die Verteilung von $\bar{x}$ immer berechnen.

Einfache Resultate für die Verteilung von $\bar{X}$ ergeben sich, wenn X eine Bernoulli- oder Normalverteilung hat

**unbekannte Verteilung:** Der zentrale Grenzwertsatz besagt, dass die Verteilung von $\bar{X}$ näherungsweise normalverteilt ist, solange N genügend gross ist. Dies gilt für beliebige Verteilung von X, auch für unbekannte.

### Beispiel

N = 2

An einem Fliessband werden Schuhe produziert. Sei $X \in \{0, 1, 2\}$ die Anzahl der defekten Schuhe bei einem Paar, wobei $P(X = 0) = 0.81$, $P(X = 1) = 0.18$, und $P(X = 2) = 0.01$. Erwartungswert und Varianz lassen sich berechnen als $\mu_X = 0.2$ und $\sigma_X^2 = 0.18$.

| Stichprobe | $\bar{X}$ | $P(\bar{X})$ | Stichprobe | $\bar{X}$ | $P(\bar{X})$ |
|---|---|---|---|---|---|
| (0,0) | 0.0 | $(.81)^2 = .6561$ | (1,2) | 1.5 | (.18)(.01) = .0018 |
| (0,1) | 0.5 | (.81)(.18) = .1458 | (2,0) | 1.0 | (.01)(.81) = .0081 |
| (0,2) | 1.0 | (.81)(.01) = .0081 | (2,1) | 1.5 | (.01)(.18) = .0018 |
| (1,0) | 0.5 | (.18)(.81) = .1458 | (2,2) | 2.0 | $(.01)^2 = .0001$ |
| (1,1) | 1.0 | $(.18)^2 = .0324$ | | | |

Angenommen, es werden zwei Paare zufällig entnommen. Was ist dann die Verteilung von $\bar{X}$, der mittleren Anzahl der defekten Schuhe?

Somit gilt:

| $\bar{X}$ | 0 | .5 | 1 | 1.5 | 2 |
|---|---|---|---|---|---|
| $P(\bar{X})$ | .6561 | .2916 | .0486 | .0036 | .0001 |

$\bar{x} \in \{0, 0.5, 1, 1.5, 2\}$

Wir können zeigen: $E(\bar{X}) = 0.2$ und $Var(\bar{X}) = 0.09$, also gerade die halbe Populationsvarianz.

Wir könnten das Beispiel fortführen, und die Verteilung von $\bar{X}$ für $N = 3$ oder $N = 4$ berechnen, aber es ist offensichtlich, dass das sehr mühsam wäre.

Wenn N vervierfacht und s konstant bleibt, dann schrumpft der Standardfehler von $\bar{x}$ auf die Hälfte

## 4.3 Erwartungswert und Varianz

**Stichprobenmittel (SPM)**

$\boxed{E(SPM) = \mu}$ → Erwartungswert des Stichprobenmittels = Erwartungswert der Population
$\hookrightarrow \bar{x}$ $E(\bar{x}) = E(x)$

mit $\bar{x}$ teilen wir die Varianz durch N (sonst nicht)

$\boxed{Var(SPM) = \frac{\sigma_x^2}{N}}$ → $Var(\bar{x})$ halbiert sich, wenn sich N verdoppelt.

wenn N grösser + grösser wird tendiert $Var(\bar{x})$ gegen Null

⚠ wenn es um $\bar{x}$ geht, dann $N \geq 50$, wenn es um $\hat{p}$ geht, dann $\min\{Np, N(1-p)\} \geq 10$

**Stichprobenanteil (SPA)**

$\boxed{E(SPA) = p}$ $\hookrightarrow \hat{p}$

$\boxed{Var(SPA) = \frac{p(1-p)}{N}}$
$\hookrightarrow \hat{p}$

$Var(\hat{p})$ halbiert sich, wenn N sich verdoppelt

Grosse Stichproben können sehr genaue Informationen über p liefern.

## 4.4 Der zentrale Grenzwertsatz → besagt, dass sich der Mittelwert unabhängig und identisch verteilter Zufallsvariablen der Normalverteilung annähert, wenn der Stichprobenumfang zunimmt

**Probleme:**
- In der Regel kennen wir die Parameter $\mu, \sigma^2, p$ nicht
- Oft kennen wir die Verteilungsform von $\bar{x}$ nicht

Lösung

**Grenzwertsatz:** → Falls der Stichprobenumfang N gross ist, ähnelt die Verteilung von $\bar{x}$ einer Normalverteilung schon ab $N \geq 50$

"näherungsweise" Populationsvarianz

$\boxed{\bar{x} \overset{\sim}{\sim} Normal\left(\mu_x, \frac{\sigma_x^2}{N}\right)}$ → Je grösser N desto mehr ähnelt die Verteilung einer Glockenkurve (Normalverteilung) und desto kleiner ist die Streuung von $\bar{x}$
$E(x)$

$\boxed{z = \frac{x - \mu}{\sqrt{\sigma^2 : N}}}$ → Standardnormal $z \sim N(0,1)$

of the *Central Limit Theorem.* The Central Limit Theorem says that when the sample size is sufficiently large the sample mean $\bar{X}$ will become approximately normally distributed with mean equal to the population mean and variance equal to the population variance divided by the sample size. And

Der zentrale Grenzwertsatz gilt für beliebige Verteilungen von X. Er setzt nur voraus, dass die Elemente der Summe gleichverteilt und unabhängig voneinander sind.

**Beispiel:** $X \in \{1, 2, .., 6\}$ $E(x) = 3.5$ $Var(x) = 2.92$ SPM $\overset{\sim}{\sim}$ Normal $\left(3.5, \frac{2.92}{100}\right)$

- Sie werfen einen Würfel 100 mal hintereinander
- Was ist die ungefähre Wahrscheinlichkeit, dass die Summe der Augenzahlen 400 übersteigt?
- Gesucht ist also $P(SPM > 4)$.

$P(\bar{x} > 4) = P\left(z > \frac{4 - 3.5}{\sqrt{2.92 : 100}}\right) = 1 - P\left(z \leq \frac{0.5 \cdot 10}{\sqrt{2.92}}\right) = 1 - P(z \leq 2.9) \approx 0$

## 4.5 Approximation der Binomialverteilung

- Das zugrundeliegende Zufallsexperiment ist Bernoulli → Daher $\mu_x = p$ / $\sigma_x^2 = p(1-p)$
- $\bar{x}$ ist der Anteil (oder Prozentsatz) der "Erfolge" in den N Wiederholungen des Experiments
$\hookrightarrow \hat{p}$ (Stichprobenanteil)

| diskret | stetig |
|---|---|
| x = c | $c - 0.5 < x < c + 0.5$ |
| x < c | $x < c - 0.5$ |
| x ≤ c | $x < c + 0.5$ |
| x > c | $x > c + 0.5$ |
| x ≥ c | $x > c - 0.5$ |

Populationsvarianz im Bernoulli Fall

**Zentraler Grenzwertsatz:** $\boxed{\hat{p} \overset{\sim}{\sim} Normal\left(p, \frac{p(1-p)}{N}\right)}$ $\min\{Np, N(1-p)\} \geq 10$

| | Nehmen Sie an, die Anforderungen zur Anwendung des Zentralen Grenzwertsatze seien erfüllt. Welche der folgenden Aussagen trifft NICHT zu? | |
|---|---|---|
| 1: | Die Wahrscheinlichkeit, dass der Mittelwert von X in dem Intervall mu ± 1.96 SA(X) liegt, beträgt 95% man muss $\bar{x}$ stehen | X |
| 2: | Wenn sich die Stichprobengrösse verdoppelt, halbiert sich die Varianz vom SPM. → | |
| 3: | Mittelwerte in wiederholten Zufallsstichproben streuen symmetrisch um den Erwartungswert der Population. richtig | |
| 4: | Der Erwartungswert des Stichprobenmittels ist gleich dem Erwartungswert der Population. richtig | |

Für grosse N nähert sich die Binomialverteilung der Form der Normalverteilung an.

Für die Normal-Approximation benötigen wir: $Np \geq 10$ $N(1-p) \geq 10$

Unter der Relation $X = Np$ gilt: $\boxed{X \overset{\sim}{\sim} Normal(Np, Np(1-p))}$

Erfolge bezeichnet als X

**Beispiel** Gesucht: $P(\hat{p} \geq 0.2)$ $\hat{p} \to$ SPA $N \cdot p = \frac{100}{6} \geq 10$ ✓ $N(1-p) = \frac{100 \cdot 5}{6} \geq 10$ ✓ $\hat{p} \overset{\sim}{\sim} Normal\left(1/6, \frac{1/6 \cdot (1 - 1/6)}{100}\right)$

- Sie werfen einen Würfel 100 mal hintereinander
- Was ist die ungefähre Wahrscheinlichkeit, dass Sie mindestens 20 Sechsen werfen?

$P(\hat{p} \geq 0.2) = P\left(z \geq \frac{0.2 - 1/6}{\sqrt{\frac{1/6 \cdot 5/6}{100}}}\right) = 1 - P(z < 0.894) = 18.6\%$

## 4.6 Stetigkeitskorrektur → ⚠ nur wenn explizite Annahmen gemisst werden müssen!

Betrachten Sie das Beispiel der vorherigen Seite. Wir können die Wahrscheinlichkeit äquivalent beschreiben als:

(a) $P(\text{Anzahl} > 19)$ (b) $P(\text{Anzahl} \geq 20)$

Aber wenn wir die Normal-Approximation anwenden, dann bekommen wir je nachdem zwei unterschiedliche Antworten:

(a) $P(\text{Anzahl} > 19) = P(\hat{p} > 0.19) \approx 0.266$
(b) $P(\text{Anzahl} \geq 20) = P(\hat{p} \geq 0.2) \approx 0.186$

Dies ist unbefriedigend, aber es gibt eine Lösung: wir wenden die Normal-Approximation auf $P(\text{Anzahl} > 19.5)$ an, und die Antwort ist dann $P(\text{Anzahl} > 19.5) = P(\hat{p} > 0.195) \approx 0.224$.

Die exakte W-keit basierend auf *Binomial*(100, 1/6) beträgt 0.220

$P(X = x) \approx P(x - 0.5 \leq Y \leq x + 0.5) = P\left(\frac{x - 0.5 - \mu}{\sigma} \leq z \leq \frac{x + 0.5 - \mu}{\sigma}\right)$

- Wobei z standardnormalverteilt ist
- Das Punktereignis $X = x$ wird durch das Intervallereignis $Y \in [x - 0.5, x + 0.5]$ ersetzt

# 5 Schätzen mit Konfidenz

↳ Stime

## 5.1 Grundlagen der induktiven Statistik

**Induktive Statistik** Hat zum Ziel, aus einer gegebenen Stichprobe auf die Bevölkerung zu schliessen (Aussagen von der Stichprobe verallgemeinern und etwas über die Eigenschaften der zugrundeliegenden Population zu lernen)

Eine "Vollerhebung" ist in allen diesen Fällen unmöglich oder teuer.

Verallgemeinerungen durch Schätzung sind immer mit Schätzfehler verbunden. Um diese zu quantifizieren benutzen wir:

- Konfidenzintervalle
- Hypothesentests

Diese Methoden der Wahrscheinlichkeitsrechnung liefern die Verbindung zwischen der Deskriptiven und der Induktiven Statistik

**Parameter und Schätzer**

Vorgehen um von einer Stichprobe auf die Bevölkerung zu schliessen: Dies geschieht durch Schätzung eines Parameters der Verteilung durch einen Stichprobenwert.

**Parameter $\theta$:** Zahl, die die Population beschreibt. Er ist eine Kenngrösse der Grundgesamtheit oder einer Verteilung (z.B.: $\mu, \sigma^2, \sigma, p, m$)

**Schätzer $\hat{\theta}$:** Eine Zufallsvariable, die aus der Stichprobenzufallsvariable abgeleitet wird → Wenn wir die Grundgesamtheit (Population) nicht kennen, müssen wir den Parameter schätzen. Durch Auswerten einer Stichprobe, erhalten wir Schätzwerte für den Parameter (z.B: $\bar{x}, s^2, s, \hat{p}$) ↳ Stichprobenmittelwert; ↳ Stichproben Anteil

↳ Bevor Stichprobe: ZV mit $\mu, \sigma^2$ usw.

Danach: Zahl (Schätzwert) → eine Methode, die jeder (Zufalls-)Stichprobe einen Schätzwert für den Populationsparameter zuordnet.

$\hat{\theta} \neq \theta$ es gibt einen Schätzfehler $\hat{\theta} - \theta$

Zahl, die den Populationsparameter schätzt

## 5.2 Schätzfunktionen und ihre Eigenschaften

**Verzerrung** Betrifft die Lage (das Zentrum) der Stichproben-Verteilung

$\boxed{\text{Bias} = E(\hat{\theta}) - \theta}$ Wenn der Bias nicht gleich Null ist, wurde der Schätzer nicht gut konstruiert, d.h. er ist systematisch verzerrt

Falls Bias = 0 ist der Schätzer unverzerrt (erwartungstreu) (unbiased), da sein Erwartungswert mit dem Parameter übereinstimmt: $E(\hat{\theta}) = \theta$

Für jede ZSF gilt: $E(\bar{x}) = \mu$, $E(\hat{p}) = p$

→ SPM $\bar{x}$ ist ein unverzerrter Schätzer für $\mu$
→ SPA $\hat{p}$ ist ein unverzerrter Schätzer für $p$

Mittleren quadratischen Fehler: $MSE = E(\hat{\theta} - \theta)^2$. Bei unverzerrten (non distorto) Schätzer ist dies einfach $Var(\hat{\theta})$. Gerne → Schätzer mit möglichst kleinem MSE

↳ umso kleiner, je: 1) kleiner die Populationsvarianz $\sigma^2$ oder 2) entfernt der Populationsanteil von 0.5 liegt oder 3) grösser die Stichprobengrösse N

**Variabilität** Betrifft die Streuung der Stichproben-Verteilung

kann durch $Var(\hat{\theta})$ oder $SA(\hat{\theta})$ gemessen werden

→ Die Variabilität gibt an, wie stark der Schätzwert variiert, wenn verschiedene Stichproben gezogen werden

| | Ein Schätzer wird als unverzerrt bzw. erwartungstreu bezeichnet, wenn |
|---|---|
| 1: | bei steigendem Stichprobenumfang die Varianz gegen Null geht. |
| 2: | der Erwartungswert des Parameters gleich der Schätzfunktion ist. |
| 3: | sein Erwartungwert gleich der Varianz ist. bei poisson-... |
| 4: | sein Erwartungswert gleich dem zu schätzenden Parameter ist. |

→ Parameter hat keinen EW

→ je grösser N, desto more likely dass $E(\hat{\theta}) = \theta$

**Konsistenz** Für grosses N verschwindet der ~~Bias (wenn vorhanden)~~ MSE ~~und~~ (die Varianz ~~geht gegen~~ bei unverzerrten Schätzer) gleich Null

Beispiele: konsistenter Schätzer: $\hat{\mu} = \bar{x} + \frac{\hat{1}}{N}$ → Varianz geht gegen Null / inkonsistenter Schätzer: $\hat{\mu} = X_1$ Varianz geht nicht gegen Null

↳ Sowohl SPM wie auch SPA sind konsistente Schätzer für $\mu$ bzw. $p$

Konsistent wenn die Wahrsch. dass der Schätzer beliebig nahe beim Populationsparameter liegt, mit steigendem N gegen eins geht. Genauigkeit der Schätzung beliebig gross, MSE geht gegen 0

**Effizienz** Ein Schätzer hat die kleinste Varianz unter allen möglichen unverzerrten Schätzern $\sigma^2\{S_1\} < \sigma^2\{S_2\}$ and $E\{S_1\} = E\{S_2\} = \theta$

**Schätzfunktionen**

**Für den Erwartungswert $E(x)$**

Ist das Stichprobenmittel $\bar{x}$

die richtige Schätzfunktion:

$$\boxed{\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i}$$

**Für die Varianz $\sigma^2(x)$**

Ist die Varianz der Stichprobe $s^2$

die richtige Schätzfunktion:

$$\hat{\sigma}^2 = \boxed{s^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2}$$

Teilen durch N−1: $E(s^2) = \sigma^2$ (unverzerrtheit)

↳ Erwartungswert der geschätzte Varianz ist gleich die wahre Varianz

**Für die Erfolgswahrscheinlichkeit p**

eines Bernoulliexperiments ist der Anteilswert $\hat{p}$

die richtige Schätzfunktion.

$$\boxed{\hat{p} = \frac{X}{N} = \frac{\text{Anzahl Erfolge}}{\text{Grösse der Stichprobe}}}$$

Geschätzten Varianzen: $\widehat{Var}(\bar{x}) = \hat{\sigma}^2/N$ ; $\widehat{Var}(\hat{p}) = \hat{p}(1-\hat{p})/N$ → Wurzel ist Standardfehler ($SF(\bar{x}) = \sqrt{\widehat{Var}(\bar{x})}$ bzw. $SF(\hat{p}) = \sqrt{\widehat{Var}(\hat{p})}$) "Geschätzte Standardabweichung"

**Wichtigste Beispiele:**

| unverzerrt: | Parameter | Schätzer | Varianz |
|---|---|---|---|
| $E(\bar{x}) = \mu$ | Mittelwert $\mu$ | Stichprobenmittel $\bar{x}$ | $Var(\bar{x}) = \frac{\sigma^2}{N}$ |
| $E(s^2) = \frac{1}{N-1}\sum(x_i - \bar{x})^2$ | Varianz $\sigma^2$ | Stichprobenvarianz $s^2$ | $Var(s^2) = \frac{\sigma^2}{N}$ |
| $E(\hat{p}) = p$ | Anteil $p$ | Stichprobenanteil $\hat{p}$ | $Var(\hat{p}) = \frac{p(1-p)}{N}$ |

→ All diese Schätzer sind unverzerrt!

(A) $E(X - E(X))^2$ (B) $\frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})^2$ (C) $E(\bar{X} - \mu)^2$

(D) $\widehat{Var}(\bar{X}) = s^2/N$ (E) $s/\sqrt{N}$

- B ( ) Stichprobenvarianz — in ZF ∇ o
- D ( ) Geschätzte Varianz des SPM
- C ( ) Mittlerer Quadratischer Fehler des SPM
- E ( ) Standardfehler des SPM $= s/\sqrt{N} = \sqrt{\frac{s^2}{N}} = \sqrt{\text{Stichprobenvarianz}/N}$
- A ( ) Populationsvarianz

Schätzfunktion ist unverzerrt (erwartungstreu) wenn ihr Erwartungswert dem zu schätzenden Parameter entspricht

$\hat{\mu} = X_1$ $\hat{\mu} = \frac{X_1 + X_N}{2}$ ABER nicht $\hat{\mu} = \bar{x} + \hat{\bar{x}}$

Ein 90% KI enthält mit 90%-iger Wahrsch. den gesuchten Parameter (Mit 10%-iger Wahrsch. ist er nicht enthalten) ↳ falsch!

$\hat{p} \sim Normal(p, \sqrt{\hat{p}(1-\hat{p})/N})$

Das $100 \times (1-\alpha)\%$ KI für $p$ ist:

$$100 \times (1-\alpha)\% KI = \hat{p} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{N}} \quad SF(\hat{p})$$

Nicht vergessen: wir benötigen $\mathrm{Min}\{N\hat{p}, N(1-\hat{p})\} \geq 10$

## 5.3 Konfidenzintervalle

Je breiter das Intervall, desto ungenauer die Schätzung

Ein Konfidenzintervall ist ein Intervall um den Schätzwert $\hat{\theta}$, in dem der wahre Parameter $\theta$ mit einer gewissen Konfidenz enthalten ist.

Grundlage des Konfidenzintervalls ist der zentrale Grenzwertsatz. Dieser garantiert: mehr Stichproben ⟶ kleineres Intervall

| $\alpha$ | $1-\alpha$ | $(1-\alpha/2)$ | $z$ | $s_{\bar{x}}$ | $zs_{\bar{x}}$ | $\bar{X}$ | $\bar{X}+zs_{\bar{x}}$ | $\bar{X}-zs_{\bar{x}}$ |
|---|---|---|---|---|---|---|---|---|
| 0.1 | .90 | .950 | 1.645 | .32 | .5264 | 8.31 | 8.84 | 7.78 |
| 0.05 | .95 | .975 | 1.960 | .32 | .6272 | 8.31 | 8.94 | 7.68 |
| 0.01 | .99 | .995 | 2.576 | .32 | .8243 | 8.31 | 9.13 | 7.48 |

$$100 \times (1-\alpha)\% KI = \hat{\theta} \pm z_{1-\frac{\alpha}{2}} \times SF(\hat{\theta})$$

$$100 \times (1-\alpha)\% \text{ KI } = \hat{p} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{N}}$$

$\alpha = 0.05$:

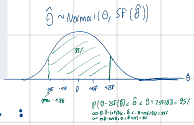
$$95\% \text{ KI } = \bar{X} \pm 1.96 \times \frac{s}{\sqrt{N}}$$

### Konfidenzgleichung

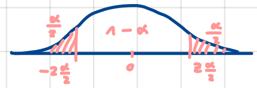
Man gibt sich ein Konfidenzniveau $\underline{1-\alpha}$ ($100 \cdot$) vor (üblicherweise in Prozent)

In Abhängigkeit von der Stichprobenzahl erhält man dann ein Intervall um den Schätzwert:

$\hat{\theta} \sim Normal(\theta, SF(\hat{\theta}))$




Wir dürfen nicht sagen, dass die W'keit dass $\theta$ in dem Intervall liegt, beträgt $100 \times (1-\alpha)$ Prozent → weil es nicht wer zufällig ist.
→ Unbekannte Parameter ist gegeben, Parameter kann auch nicht im Intervall sein (es gibt keinen Zufall)




↳ immer Hälfte von Konfidenzintervall angeben: [18.6, 24.8] → D = 6.2 → FM = 3.1

$$\boxed{\begin{aligned}(1-\alpha)\% \text{ Konfidenzintervall} &= \text{Schätzer} \pm \text{Fehlermarge (FM)} \\ &= \text{Schätzer} \pm z_{1-\frac{\alpha}{2}} \times \text{Standardfehler (SF)}\end{aligned}}$$

→ Ein Konfidenzintervall ist umso grösser, je höher das Konfidenzniveau

→ hohes $\alpha$ bedeutet: Intervall ist kurz, hängt stark von Stichprobe ab

→ KI umso kleiner, je kleiner der SF. Und umso kleiner, je kleiner das Konfidenzniveau (grösseres $\alpha$ führt zu kleinem KI) z.B. $\alpha = 0.05$ Konfidenzniveau $= 1 - 0.05$ $z(\alpha) = 0.95$

### Standardfehler (SF)

Standardabweichung der zu schätzenden Variable

Mittelwertschätzung: $\boxed{SF = \frac{\sigma}{\sqrt{N}}}$

Anteilsschätzung: $\boxed{SF(\hat{p}) = \sqrt{\frac{p(1-p)}{N}}}$

→ Je höher SF desto ungenauer die Schätzung

)). Somit

$$\boxed{Z = \frac{\hat{\theta} - \theta}{SF(\hat{\theta})} \dot{\sim} Normal(0, 1)}$$

$$P(-1.96 < (\hat{\theta} - \theta)/SF(\hat{\theta}) < 1.96) = 0.95$$

$$P(\hat{\theta} - 1.96SF(\hat{\theta}) < \theta < \hat{\theta} + 1.96SF(\hat{\theta})) = 0.95$$

### Fehlermarge (FM)

Halbe Breite eines Konfidenzintervalls

$$\boxed{FM = z_{1-\frac{\alpha}{2}} \times \frac{s}{\sqrt{N}}}$$

$s$: Standardabweichung der Stichprobe $\alpha$: Signifikanz

$z$: Perzentilfunktion der Standardnormalverteilung. Dabei ist $z = z_{1-\frac{\alpha}{2}}$ der Wert, so dass $P(-z < Z \leq z) = 1-\alpha$ die Konfidenz ist

- $1-\alpha \uparrow$ : FM $\uparrow$
  Intervall breiter (mehr Sicherheit durch Unbestimmtheit)
- $s \uparrow$ : FM $\uparrow$
  Intervall breiter (unsicheres Experiment)
- $N \uparrow$ : FM $\downarrow$
  Intervall kürzer (Bessere Schätzung)

## 5.4 Planung der Stichprobengrösse

kann kleiner sein wenn Konfidenzniveau sehr klein ist

| | Welche der folgenden Aussagen trifft zu? | |
|---|---|---|
| 1: | Sei SPM = 0, SF=1, und N>50; dann liegt der Populationsmittelwert mit 95% Wahrscheinlichkeit in dem Intervall [-1.96, +1.96]. | |
| 2: | Die Fehlermarge ist immer grösser als der Standardfehler. | |
| 3: | Ein Konfidenzintervall ist umso grösser, je höher das Konfidenzniveau. | X |
| 4: | Ein Konfidenzintervall gibt den wahrscheinlichen Bereich an, in dem eine neue Beobachtung liegen wird. | |

Gesucht ist die Minimalzahl von Stichproben die zu nehmen sind damit eine vorgegebene Konfidenz erreicht wird.

Vorgegeben sind dazu die Intervallbreite bzw. Fehlermarge FM, $s$ oder $s^2$ sowie die Konfidenz $1-\alpha$

$$\boxed{N \geq \left(\frac{z_{1-\frac{\alpha}{2}} \times s}{FM}\right)^2}$$

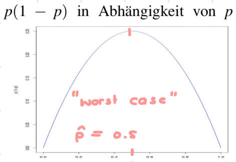
muss ggf. auf die nächst grössere natürliche Zahl aufgerundet werden

oder $\cancel{N \geq \left(\frac{z_{1-\frac{\alpha}{2}} \times \hat{p}(1-\hat{p})}{FM}\right)^2}$

Problem: Wir kennen $s$ nicht → man kann eine "worst case scenario" Schätzung für $s$ verwenden

### Stichprobenumfang für gewünschte Fehlermarge

$p(1-p)$ in Abhängigkeit von $p$




Wir wollen ein 90% Konfidenzintervall für $p$ mit vorgegebener Fehlermarge FM von höchstens ± 5%

→ Wie gross muss die Stichprobe mind. sein? $1-\alpha = 90\%$ $\alpha = 0.1$ $z_{1-\frac{\alpha}{2}} = z_{0.95}$ → $z = 1.645$

Lösung: $FM \leq 1.645 \times \sqrt{\frac{\hat{p}(1-\hat{p})}{N}} \xrightarrow{\text{nach N auflösen}} N^* = \left(\frac{1.645}{0.05}\right)^2 \times \hat{p}(1-\hat{p})$

$$\boxed{N = \left(\frac{z}{\hat{p}}\right)^2 \hat{p}(1-\hat{p})}$$

nun benutzen wir das "worst case" $\hat{p}$: 0.5 $N^* = 270$

### Endliche und unendliche Populationen

Mit jedem Ziehen ohne Zurücklegen ändert sich die Struktur der Population, und daher sind die Züge nicht unabhängig voneinander. Man kann zeigen, dass die Varianz von $\hat{p}$ kleiner ist als die Binomial-Varianz:

$$\boxed{Var(\hat{p}) = \frac{p(1-p)}{N_S} \times \frac{N_P - N_S}{N_P - 1}}$$

$N_P$: Grösse der Population
$N_S$: Grösse der Stichprobe

Der zweite Bruch ist die Endlichkeitskorrektur beim Modell ohne Zurücklegen. Der Bruch tendiert gegen 1 wenn $\frac{N_P}{N_S}$ gross wird

→ Daher ergibt sich bei kleineren Populationen ($N_S, \alpha$ gegeben) eine kleinere FM / ($FM^*, \alpha$ gegeben) eine kleinere notwendige Beobachtungsanzahl $N_S$

### Kleine Stichproben

Wie lässt sich das Konfidenzintervall bestimmen, wenn der ZGS nicht erfüllt ist?

→ $N\hat{p} \sim$ Binomial $(N, p)$ wenn $X \sim$ Bernoulli $(p)$

→ "t-Verteilung" wenn $X \sim$ Normal $(\mu, \sigma^2)$ !

$$\boxed{t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{N}}}} \rightarrow [a, b] = \bar{x} \pm t_{N-1, 1-\frac{\alpha}{2}} \times \frac{s}{\sqrt{N}}$$

Dieses Intervall ist etwas breiter aber für N = 101 schon fast identisch

| | 90% | 95% |
|---|---|---|
| $t$ für $N = 21$ | 1.729 | 2.093 |
| $t$ für $N = 51$ | 1.677 | 2.010 |
| $t$ für $N = 101$ | 1.660 | 1.984 |
| $t$ für $N = 1'001$ | 1.646 | 1.962 |
| $z$ | 1.645 | 1.960 |

## 5.5 Konfidenzrechnung

### Schätzung des Mittelwerts $\mu$ $\quad X \overset{\cdot}{\sim} \text{Normal}\left(\mu, \frac{s^2}{N}\right)$

$$[a,b] = \left[\bar{x} - \frac{\sigma \cdot z_{1-\alpha/2}}{\sqrt{n}}, \bar{x} + \frac{\sigma \cdot z_{1-\alpha/2}}{\sqrt{n}}\right]$$

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

erfüllt, ist gerade $1-\alpha$.

Gesucht ist der Mittelwert $\mu$ einer Verteilung

Gegeben sind: die geforderte Konfidenz $1-\alpha$, $N$, $s$, Mittelwert $\bar{x}$

auszurechnen ist dann das Konfidenzintervall: $[a,b] = \bar{x} \pm z_{1-\frac{\alpha}{2}} \cdot \underbrace{\frac{s}{\sqrt{N}}}_{SA}$

**Beispiel:** Geschätzt werden soll das Durchschnittsgewicht $\mu$ gefangener Fliegen

$n = 100$ gewogene Fliegen, Mittelwert $\bar{x} = 3.2$ $\quad SA\ (s) = 2.1$, es soll ein 95%-Konfidenzintervall erstellt werden

① zuerst wird der Schätzer aufgestellt

da $\mu$ gefragt ist nehmen wir $\bar{x} = 3.2$

② SF ausrechnen: $SF = \frac{s}{\sqrt{N}} = \frac{2.1}{\sqrt{100}} = 0.21$

③ Die geforderte Konfidenz ist $1-\alpha = 95\%$ $\rightarrow$ also $\alpha = 0.05$

$\rightarrow z_{1-\frac{\alpha}{2}} =$

$1 - 0.025 = z_{0.975} \rightarrow z = 1.96$

in der Mitte der Tabelle die Werte am Rand ablesen

④ $FM = z \cdot SF$

$= 1.96 \cdot 0.21 = 0.41$

⑤ Jetzt können wir das Intervall aufstellen: $[a,b] = 3.2 \pm 0.41 = [2.79, 3.61]$

$\Rightarrow$ unter Annahme, dass $\mu = 3.2$ der tatsächliche Mittelwert für das Gewicht ist, liefert eine zufällige Stichprobe mit $n = 100$ Fliegen ein Durchschnittsgewicht in diesem Intervall mit Wahrscheinlichkeit $\geq 95\%$.

### Prognose-Intervalle $\quad X_1, \ldots, X_{T+1} \overset{iid}{\sim} \text{Normal}(\mu, \sigma^2)$

Neben dem Konfidenzintervall gibt es auch das Prognoseintervall. Es gibt einen Bereich für eine neue einzelne Beobachtung $X_{neu}$ an

Der beste Prognosewert für $X_{neu}$ ist das Stichprobenmittel $\bar{x}$, $E\bar{x} = \mu$

$Var(X_{neu}) = E(X_{neu} - \bar{X})^2 + E(\bar{x} - \mu)^2$ $\quad \rightarrow$ Prognoseintervall: $[a,b] = \bar{x} \pm z_{1-\frac{\alpha}{2}} \times \sqrt{\left(1+\frac{1}{N}\right)s^2}$ oder $\times \sqrt{\frac{s^2}{T} + s^2}$

schrumpft mit zunehmendem $\alpha$

$\hat{SA}(X_{neu}) = \sqrt{\left(1+\frac{1}{N}\right)s^2}$

- Da $\hat{SA}(X_{neu}) > \frac{s}{\sqrt{N}}$ gilt ist das Prognoseintervall immer grösser als (breiter) das Konfidenzintervall
- Beim Prognoseintervall müssen wir annehmen, dass $X$ in der Population normalverteilt ist
- ! ZGS können wir nicht verwenden

**Beispiel:**

Die tägliche Nachfrage nach Brötchen sei normalverteilt. Messungen in den vergangenen 100 ($N = 100$) Tagen ergaben folgende Werte: $\bar{x} = 500$, $s = 20$. Was ist das 95%-Prognoseintervall für die Nachfrage morgen?

$1-\alpha = 95\% \rightarrow \alpha = 0.05$

$\hat{SA}(X_{neu}) = \sqrt{\left(1+\frac{1}{100}\right) \cdot 20^2} = 20.1$

$z_{1-\frac{\alpha}{2}} = z_{0.975}$

$z = 1.96$

$500 \pm 1.96 \cdot 20.1 =$

$500 \pm 4.116 = [495.884, 504.116]$

Wegen ZGS gilt: $\bar{x} \overset{\cdot}{\sim} \text{Normal}\left(\mu, \frac{s^2}{N}\right)$ für $N \geq 50$

$\hat{p} \overset{\cdot}{\sim} \text{Normal}\left(p, \frac{\hat{p}(1-\hat{p})}{N}\right)$ für $\min\{N\hat{p}, N(1-\hat{p})\} \geq 10$

KI für $\mu$: $SF(\bar{x}) = \frac{s}{\sqrt{n}}$, $SF(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{N}}$

$100 \cdot (1-\alpha)\%\ KI = \bar{x} \pm \overbrace{z_{1-\frac{\alpha}{2}} \cdot SF(\bar{x})}^{FM}$ -> für $\bar{x}$ -> Stichprobenmittelwert

$100 \cdot (1-\alpha)\%\ KI = \hat{p} \pm z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{N}}$ -> für $\hat{p}$ -> Stichprobenanteil

### Stichprobenanteilswerte (Schätzung eines Anteils einer Eigenschaft)

$$[a,b] = \left[\hat{p} - z_{1-\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{1-\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

Gesucht ist eine Schätzung des Anteils $p$ der Individuen, die eine gewisse Eigenschaft erfüllen

Gegeben sind: geforderte Konfidenz $1-\alpha$, Anzahl $N$ in der ein Anteil von $\hat{p}$ Kandidaten die Eigenschaft erfüllt

Konfidenzintervall: $[a,b] = \hat{p} \pm z_{1-\frac{\alpha}{2}} \times \underbrace{\sqrt{\frac{\hat{p}(1-\hat{p})}{N}}}_{SA}$

! $N\hat{p} \geq 10$, $N(1-\hat{p}) \geq 10$

**Beispiel:** Gefragt ist der Anteil Linkshänder in der Gesamtbevölkerung. In einer Stichprobe mit $n = 20$ befragten Personen, waren vier linkshändig

→ Anteil von $\hat{p} = 0.2 = 20\%$ $\quad$ Daraus soll ein Konfidenzintervall zum Niveau 99% für den Anteil $p$ erstellt werden.

① Schätzer aufstellen: $\hat{p} = 0.2 = 20\%$

② SF ausrechnen: $s = \sqrt{\hat{p}(1-\hat{p})} = \sqrt{0.2 \cdot 0.8} = 0.4$

$SF = \frac{s}{\sqrt{N}} = \frac{0.4}{\sqrt{100}} = 0.09$

③ Fehlermarge ausrechnen: $1-\alpha = 99\% \rightarrow \alpha = 0.01 \rightarrow z_{1-\frac{\alpha}{2}} = z_{0.995}$

$FM = 2.575 \times 0.09 = 0.23$ $\quad z = 2.575$

④ Konfidenzintervall: $0.2 \pm 0.23 = [-0.03, 0.43]$

→ Die hohe Konfidenz des Intervalls (bei kleinem N) wurde dadurch erkauft, dass es sehr breit ist, was die eigentliche Aussage über $p$ verwässert.

## Schätzung einer Differenz von Mittelwerten (Vergleich von Mittelwerten) → unabhängige Stichproben

Gesucht ist nun die Schätzung der Differenz zweier Mittelwerte $\Delta = \mu_1 - \mu_2$

Gegeben sind dazu zwei Stichproben mit Anzahlen $N_1, N_2$ und Mittelwerten $\bar{X}, \bar{Y}$ sowie den Varianzen $s_1^2, s_2^2$ und der geforderten Konfidenz $1-\alpha$

Es ergibt sich der gemischte Standardfehler : $\boxed{SF = \sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}}$ → Konfidenzintervall : $[a,b] = \bar{X} - \bar{Y} \pm z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N}}$

**Beispiel:**

Gesucht sei eine Schätzung der mittleren Gewichtsabnahme einer Person nach einer Diät, wobei nur eine niedrige Konfidenz von $1-\alpha = 80\%$ angesetzt wird. Vor Beginn der Diät wurden $n_1 = 60$ Personen gewogen, mit einem Mittelwert von $\bar{x}_1 = 75$ Kilogramm und einer Standardabweichung von $s_1 = 2.5$. Während der Diät haben leider einige Kandidaten aufgegeben, am Schluss waren noch $n_2 = 56$ Personen beteiligt, mit Mittelwert $\bar{x}_2 = 62$ Kilogramm und Standardabweichung $s_2 = 1.2$. Daraus soll ein 80%-Konfidenzintervall für die Gewichtsmitteldifferenz $\Delta = \mu_1 - \mu_2$ erstellt werden.

① Schätzer : $\bar{X}_1 - \bar{X}_2 = 75 - 62 = \underline{13}$

② $SF = \sqrt{\frac{2.5^2}{60} + \frac{1.2^2}{56}} = \underline{0.36}$

③ $1-\alpha = 80\%$   $\alpha = 0.2$   $z_{1-\frac{\alpha}{2}} = z_{0.9}$ → $z = \underline{1.285}$

④ $13 \pm \overset{0.4626}{1.285 \times 0.36} = \underline{\underline{[12.53, 13.46]}}$

→ Schätzung ist genau (kleines Intervall), jedoch ist die Konfidenz niedrig

| | Betrachtet sei die Schätzung der Differenz, Delta, von Mittelwerten in zwei Populationen. Welche der folgenden Aussagen trifft NICHT zu? | |
|---|---|---|
| 1: | SPM1-SPM2 ist ein unverzerrter Schätzer von Delta. | |
| 2: | Für den Standardfehler bei unabhängigen Stichproben gilt: SF(SPM1-SPM2) = SF(SPM1) + SF(SPM2). Wir dürfen Standardfehler nicht addieren | X |
| 3: | Der Standardfehler bei verbundenen Stichproben ist meist kleiner als der bei unverbundenen Stichproben. | |
| 4: | Die Fehlermarge beim 95% KI ist FM=1.96*SF(SPM1-SPM2) | |

## Datenpaare (verbundene Stichproben) → abhängige Stichproben

Es macht oft mehr Sinn, Datenpaare $(X_i, Y_i)$ zu sammeln. Wenn innerhalb eines Paares Abhängigkeit besteht, dann sind die Stichproben verbunden

→ $\bar{X}$ und $\bar{Y}$ sind nicht mehr unabhängig $(s_1 = s_2,\ N_1 = N_2)$

→ SF bei verbundenen Stichproben ist meist kleiner als bei unverbundenen

$D = X_i - Y_i$ stellen eine Zufallsstichprobe dar → Konfidenzintervall : $[a,b] = \bar{D} \pm z_{1-\frac{\alpha}{2}} \times \frac{s_D}{\sqrt{N}}$

$\bar{D} = \bar{x}_1 - \bar{x}_2$

$\bar{D}$: Stichproben-Mittelwert von $D_1, D_2, \dots, D_N$

$s_D$: Stichproben-SA von $D_1, D_2, \dots, D_N$

$s_D = s_1 - s_2$

**Beispiele:**

Fragestellung:

- Ein Medikament soll (zu) hohes Cholesterol senken bei Patienten, die unter diesem Problem leiden
- $\mu_1$ = mittlerer Cholesterol-Spiegel vor Medikament
- $\mu_2$ = mittlerer Cholesterol-Spiegel nach Medikament

Natürliche Vorgehensweise:

- Wähle eine Stichprobe von Patienten und messe ihren Cholesterol-Spiegel vor und nach Einnahme des Medikamentes
- Der $i$te Patient liefert uns dann das Datenpaar $(X_i, Y_i)$

'Problem':

- Hoher $X_i$-Wert bedeutet generell auch hoher $Y_i$-Wert
- $\Longrightarrow$ Abhängigkeit innerhalb eines Paares

Unterscheidet sich beim Fussball tendenziell die erwartete Anzahl der Tore in erster und zweiter Halbzeit? Wir betrachten die Ergebnisse der Vorrunde zweier Fussballweltmeisterschaften (denjenigen von 2002 in Japan/Südkorea und 2018 in Russland) als eine geeignete ZSP. In diesem Fall erhalten wir $(N = 96)$:

$\bar{x}_1 = 1.03, SA_1 = 1.11 \qquad \bar{x}_2 = 1.58, SA_2 = 1.04 \qquad \bar{D} = -0.55, SA_D = 1.44$

Somit ergibt sich das 95% KI für $\mu_1 - \mu_2$ basierend auf der verbundenen Stichproben als

$$-0.55 \pm 1.96 \times 1.44/\sqrt{96} \quad \Longrightarrow \quad [-0.838, -0.262]$$

Welches KI hätte sich unter der Annahme von unverbundenen Stichproben ergeben? Warum ist diese Annahme hier nicht sinnvoll?

weil die 1. und 2. Halbzeit nicht unabhängig sind (sind verbunden)

## Schätzung einer Differenz von Anteilen (Vergleich von Anteilswerten)

| | Welche Aussage trifft NICHT zu? Beim 95%-Konfidenzintervall für den Populationsanteil p | |
|---|---|---|
| 1: | beträgt die maximale Fehlermarge ca. 1/Wurzel(N) | |
| 2: | wird die Fehlermarge in Prozent punkte ausgedrückt | X |
| 3: | sinkt die Fehlermarge, wenn bei gegebener Grösse der ZSP die Grösse der Population abnimmt. | |
| 4: | sollte gelten: min[Np,N(1-p)] > 10 | |

Gesucht ist die Schätzung der Differenz zweier Anteile $\Delta = p_1 - p_2$

Gegeben sind dazu die Konfidenz $1-\alpha$, zwei Stichproben mit Anzahlen $N_1, N_2$, sowie Anteilen in der Stichprobe $\hat{p}_1, \hat{p}_2$

- Schätzer : $\hat{\Delta} = \hat{p}_1 - \hat{p}_2$
- $SA(\hat{\Delta}) = \sqrt{\frac{p_1(1-p_1)}{N_1} + \frac{p_2(1-p_2)}{N_2}}$ → $\boxed{SF = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{N_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{N_2}}}$

→ Konfidenzintervall : $[a,b] = \hat{p}_1 - \hat{p}_2 \pm z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{N_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{N_2}}$

**Beispiel:**

Laut Angaben der Schweizer Gesundheitsbefragung 2017 waren von den $N = 2562$ Befragten in der Altergruppe der 25-34 Jährigen 34.7 Prozent trainiert[‡]. In der Befragung von 2012 waren es nur 30.5 Prozent gewesen $(N = 2499)$.

Somit ist der Standardfehler der Differenz der SPAs

$$\sqrt{\frac{0.347(1-0.347)}{2562} + \frac{0.305(1-0.305)}{2499}} = 0.013$$

und das 95% Konfidenzintervall für den Anstieg beim Anteil der Trainierten von 2012 bis 2017 ist gegeben als ...

$\overset{0.042}{0.347 - 0.305} \pm \overset{0.02548}{1.96 \cdot 0.013} = [0.017, 0.0067]$

Benötigte Grösse der Stichprobe, damit bei gegebenem $\alpha$ die $FM \le FM^*$, wobei $FM^*$ eine vorgegebene Obergrenze ist:

$$N \ge \left(\frac{z_{1-\alpha/2} \times s}{FM^*}\right)^2$$

$N \ge \left(\frac{z_{1-\alpha/2} \cdot \sqrt{\hat{p}(1-\hat{p})}}{FM}\right)^2$

- Problem: wir kennen $s$ nicht
- Lösung: um auf der sicheren Seite zu sein, kann man eine "worst-case-scenario" Schätzung für $s$ verwenden.

# 6 Hypothesentests

**Hypothese:** Aussage über den Wert eines Parameters in der Gesamtbevölkerung → Sie haben keine Wahrscheinlichkeit, sind entweder wahr oder falsch

**Hypothesentest:** Bestätigt oder widerlegt die Hypothese zu gegebenem Konfidenzniveau durch Untersuchung gegebener Daten,
(statistischer Test) die aus einer Stichprobe stammen. Dazu gibt es ein allgemeines Verfahren:

## 6.1 Grundlagen

Wir stellen eine Nullhypothese $H_0$ auf, die auf objektiven oder subjektiven Informationen über den Parameter basiert, und die wir deshalb für wahrscheinlich halten (deren fälschliche Ablehnung ist mit hohen Kosten verbunden).
Ihr stellen wir eine Alternativhypothese $H_1$ entgegen, die wir akzeptieren müssen, falls unser Test die Nullhypothese widerlegt.
($H_1$ wollen wir "beweisen")

Dann geben wir ein Konfidenzniveau $1-\alpha$ bzw. ein **Signifikanzniveau** $\alpha$ vor: Signifikanzniveau $\alpha$ ist die maximal tolerierte Wahrscheinlichkeit für einen Typ-1 Fehler
Ist $H_0$ richtig, so wird er mit der Konfidenzwahrscheinlichkeit $1-\alpha$ die Nullhypothese beibehalten.

Dann berechnen wir ein Konfidenzintervall für den Schätzer des Parameters und ziehen eine Stichprobe.

- Falls die Stichprobe einen Schätzer produziert, der nicht in unserem Intervall liegt, verwerfen wir die Nullhypothese $H_0$ zugunsten der Alternativhypothese $H_1$
- Produziert die Stichprobe einen Schätzer im Konfidenzintervall, behalten wir die Nullhypothese bei

Wir erhalten so eine Ja/Nein-Entscheidung, welche vom Signifikanzniveau $\alpha$ abhängt:
→ Hohes $\alpha$: Stichprobe hat hohe Signifikanz/Bedeutung. Schon bei geringer Abweichung zur Hypothese vertrauen wir eher der Stichprobe und lehnen $H_0$ zugunsten $H_1$ ab.
→ Niedriges $\alpha$: Stichprobe hat nicht viel Bedeutung. Sie muss weit von $H_0$ abweichen damit wir $H_0$ verwerfen

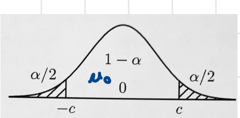
Die konkrete Testentscheidung trifft man mit der **standardisierten Teststatistik**:

$$\boxed{z = \frac{\text{Schätzwert} - \text{Vermutung}}{\text{Standardfehler}}}$$

$z$ misst wie stark die Stichprobe der Nullhypothese widerspricht (liegt $z$ im Annahmebereich wird $H_0$ beibehalten)

### Einseitige und zweiseitige Tests

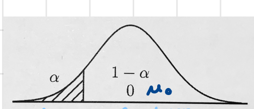
**Zweiseitiger Test**




Nullhypothese $H_0$: $\mu = \mu_0$
Alternativhypothese $H_1$: $\mu \neq \mu_0$
Annahmebereich: $[-c, c]$

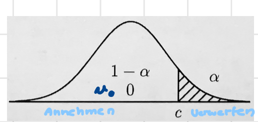
Man berechnet einen Schätzwert für $\mu$ und lehnt ab, falls dieser zu weit von $\mu_0$ entfernt ist. Für die standardisierte Teststatistik ist der Annahmebereich ein Intervall, das die Wahrscheinlichkeit $1-\alpha$ abdeckt.

**Linksseitiger Test**




Nullhypothese $H_0$: $\mu \geq \mu_0$
Alternativhypothese $H_1$: $\mu < \mu_0$
Annahmebereich: $\geq -c$
$[-c, \infty)$

**Rechtsseitiger Test**




Nullhypothese $H_0$: $\mu \leq \mu_0$
Alternativhypothese $H_1$: $\mu > \mu_0$
Annahmebereich: $\leq c$
$(-\infty, c]$

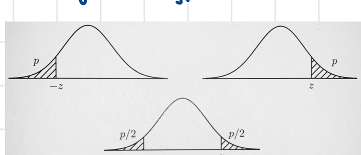
### p-Wert einer Stichprobe

Ist der p-Wert negativ ist er $\approx 0$

Der p-Wert ändert sich wenn $H_0$ sich ändert

Der p-Wert ist die Wahrscheinlichkeit dafür, bei richtiger Nullhypothese die Teststatistik $z$ (oder höher) zu erhalten

$$\text{p-Wert} = \begin{cases} P(Z \geq z) & \text{falls } H_A: \mu > \mu_0 \\ P(Z \leq z) & \text{falls } H_A: \mu < \mu_0 \\ 2P(Z \geq |z|) & \text{falls } H_A: \mu \neq \mu_0 \end{cases}$$




Ein ausgewerteter p-Wert sagt uns, wie hoch unser Signifikanzniveau $\alpha$ gewählt werden muss, damit der Test $H_0$ ablehnt

- PW $\leq \alpha$ → Wir verwerfen $H_0$ zugunsten $H_1$. Die Stichprobe widerlegt $H_0$
- PW $> \alpha$ → Wir behalten $H_0$ bei. Die Stichprobe widerspricht der Hypothese nicht

## Testentscheidungen

$H_0$ kann nicht bewiesen werden: was wir beweisen wollen, wählen wir als $H_1$.

Hypothesen haben keine Wahrscheinlichkeiten (es gilt mit Sicherheit: entweder $\mu \leq 4$, oder aber $\mu > 4$)

Der $p$-Wert ist daher nicht die Wahrscheinlichkeit, dass $H0$ stimmt.

In einer gegebenen Anwendung ist unsere Entscheidung, die Nullhypothese wegen $p$-Wert $< \alpha$ zu verwerfen, entweder richtig oder falsch. Bei einem kleinen $p$-Wert gilt: "Entweder H0 stimmt und etwas sehr Unwahrscheinliches ist passiert, oder H0 ist falsch.".

Die Anwendung dieser Methode stellt sicher, dass wir im Durchschnitt, wenn viele Hypothesen getestet werden, oder wenn die gleiche Hypothese mit verschiedenen Daten getestet wird, der Anteil der fälschlich verworfenen Nullhypothesen (bzw. der Anteil der fälschlich angenommenen Alternativhypothesen) $100 \times \alpha\%$ nicht übersteigt.

Schwerwiegenderen Fehler, Wahrscheinlichkeit seines Auftretens ist begrenzt

| | $H_0$ ist richtig | $H_0$ ist falsch | |
|---|---|---|---|
| $H_0$ verworfen | Fehler Typ I<br>$\alpha$-Fehler | kein Fehler<br>$1-P$ (Typ II) | $\alpha = P(\text{Fehler I})$ |
| $H_0$ beibehalten | kein Fehler | Fehler Typ II<br>$\beta$-Fehler<br>$1-\alpha$ | $\beta = P(\text{Fehler II})$ |

Wahrsch. eines Typ II Fehlers wird kleiner, je grösser $N$ ist

- Der Typ 1 Fehler wird als gravierender betrachtet und seine Wahrscheinlichkeit wird im voraus durch das Signifikanzniveau $\alpha$ kontrolliert
- Das Gegenstück zur Signifikanz ist die Güte eines Tests
- $\beta$ ist die Wahrscheinlichkeit, dass die Teststatistik nicht im Ablehnungsbereich liegt, wenn $H_1$ gilt.

**Güte** Die Güte beschreibt die Wahrscheinlichkeit $H_0$ zu verwerfen, wenn diese in der Tat falsch ist

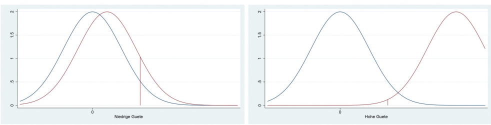
$\boxed{\text{Güte} = 1 - P(\text{Typ II})}$

$\boxed{P(\text{Typ II}) = 1 - \text{Güte}}$

Diese können wir nur in Spezialfällen direkt ausrechnen. Es gilt:

- Grosses $\alpha$ $\longrightarrow$ Grosse Güte (weniger Typ II Fehler, dafür aber mehr Typ I Fehler)
  - Je kleiner $\alpha$, desto kleiner der Ablehnungsbereich und desto grösser $\beta$.
- Je weiter $\mu$ von $\mu_0$ entfernt ist, desto grösser ist die Güte (Differenz ist leichter zu erkennen)
- Je kleiner die Standardabweichung $s$, desto grösser die Güte (Stichproben sind sicherer)
- Je grösser der Stichprobenumfang $N$, desto grösser die Güte (Stichproben sind aussagekräftiger)
- Einseitige Tests haben höhere Güte als zweiseitige (da Fehler nur in eine Richtung)

- Wert von $\mu \Longrightarrow$ je weiter weg von $\mu_0$ (in der Richtung von $H1$), desto besser




- Falls die Differenz zwischen dem Populationsparameter $\mu$ und dem Wert des Parameters $\mu_0$ unter $H_0$ steigt, steigt die Güte des Tests

## Beispiel Berechnung Güte:

Wir stellen uns nach die Frage nach der W'keit, $H0$ korrekterweise zu verwerfen, bevor $\bar{x}$ beobachtet wird.

Problem: diese W'keit hängt von $\mu$ ab, aber wir kennen $\mu$ ja gar nicht.

Lösung:

Können Erfahrungswert aus der Literatur nehmen;

oder wir setzen einen Wert ein, ab dem die Alternativhypothese auch ökonomisch relevant wird.

Teste $H0 : \mu \leq 0$ gegen $H1 : \mu > 0$

$\alpha = 0.05$, $s = 1.6$ und $N = 64$

Die Güte wird an einem Punkt $\mu \neq 0$ berechnet. Z.B., sei $\mu = 0.5$

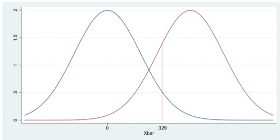
Güteberechnungen werden mit Vorteil im $X$-Raum durchgeführt (nicht mit $Z$)

$H0$ wird immer dann verworfen, wenn

$\bar{x} > 1.645 \times SF = 1.645 \times 0.2 = 0.329$

Gesucht ist somit Güte = $P(\bar{X} > 0.329)$ wenn $\mu = 0.5$.

$$P\left(\frac{\bar{X} - 0.5}{0.2} > \frac{0.329 - 0.5}{0.2}\right) = P(Z > -0.855) = 0.80$$




**Interpretation:**

Wir testen eine einseitige Hypothese $H0 : \mu \leq 0$ gegen $H1 : \mu > 0$ auf dem 5% Signifikanzniveau. Der Standardfehler ist 0.2.

- Wenn $H0$ stimmt, dann verwerfen wir $H0$ irrtümlich mit Wahrscheinlichkeit 5% (Typ-I Fehler).
- Wenn H0 nicht stimmt (weil $\mu = 0.5 \neq 0$), dann verwerfen wir $H0$ korrekterweise mit (in diesem Fall) Wahrscheinlichkeit 80%. Das ist die Güte des Tests.
- Der Typ-II Fehler, $H0$ nicht zu verwerfen obwohl falsch, hat daher eine Wahrscheinlichkeit von $\beta$ =20%.

In diesem Fall ist die Wahrscheinlichkeit eines Typ-II Fehlers "nur" 4-mal höher als die des Typ-I Fehlers.

In vielen praktischen Anwendungen ist die Güte viel kleiner, und $\beta$ damit viel höher als 0.2.

Wir können einen Hypothesentest in 6 Schritten durchführen.

**6 Schritten Hypothesentest**

1. Bestimmung von $H_0$ und $H_1$.
2. Das Signifikanzniveau $\alpha$ festlegen, z.B. $\alpha = 0.05$.
3. Prüfen, ob die Bedingungen für den ZGS erfüllt sind.
4. Den Wert der Teststatistik in Abhängigkeit von $\bar{X}$, $s$ und $N$ unter $H_0$ berechnen.
5. Entweder die Ablehnungsregion bestimmen, oder den den $p$-Wert berechnen.
6. Wir verwerfen $H_0$ auf dem $100 \times \alpha\%$ Signifikanzniveau, wenn die Teststatistik in die Ablehnungsregion fällt, oder wenn der $p$-Wert kleiner ist als $\alpha$.

## Warum niedrige Güte ein Problem ist

Durch verwerfen von $H_0$ beweisen wir $H_1$. $H_0$ nicht zu verwerfen beweist nichts (wegen der potentiell hohen wahrsch des Typ II Fehlers.

$\rightarrow$ wir wollen also $H_0$ verwerfen. Eine verworfene Nullhypothese nennen wir ein Resultat

$\rightarrow$ Was ist die wahrsch. dass ein Resultat korrekt ist?

Überlegung: Betrachte ein Wissensgebiet (z.B. Umweltökonomie, Sozialpsychologie, Pharmakologie), in dem viele Hypothesen getestet werden.

Sei $\phi$ der Anteil der Alternativhypothesen, die stimmen.

Dann gibt es zwei Möglichkeiten ein Resultat zu haben (also $H0$ zu verwerfen)

- Falsche $H0$s (davon gibt es $\phi$) werden mit W'keit $(1 - \beta)$ verworfen
- Korrekte $H0$s (davon gibt es $1 - \phi$) werden mit W'keit $\alpha$ verworfen

Insgesamt beträgt die W'keit eines Resultates damit $(1 - \beta)\phi + \alpha(1 - \phi)$

Davon sind $(1 - \beta)\phi$ "korrekte Resultate". Also ist die bedingte W'keit

$$P(\text{Resultat ist korrekt}|\text{es gibt ein Resultat}) = \frac{(1-\beta)\phi}{(1-\beta)\phi + \alpha(1-\phi)}$$

**Beispiel:**

Sei $\phi = 0.5$: die Hälfte aller getesteten Nullhypothesen sind falsch, die Hälfte der Alternativhypothesen $H1$ sind korrekt.

Dann gilt:

$$P(\text{Resultat ist korrekt}|\text{es gibt ein Resultat}) = \frac{(1-\beta)}{(1-\beta) + \alpha}$$

bzw.

$$P(\text{Resultat ist falsch}|\text{es gibt ein Resultat}) = \frac{\alpha}{(1-\beta) + \alpha}$$

Zwei Fälle:

- Hohe Güte: $1 - \beta = 0.9$
- Tiefe Güte: $1 - \beta = 0.1$

Bei tiefer Güte sind Resultate mit erheblicher W'keit falsch (weil der Anteil der fälschlich verworfenen H0s an allen verworfenen H0s mit sinkender Güte zunimmt).

Dies betrifft insbesondere Wissensgebiete, bei denen kleine Stichproben, hohe Variabilität des Merkmals, kleine "Effektgrössen" ($\mu$ nahe bei $\mu_0$), und überraschende Resultate ($\phi$ klein) überwiegen.

## 6.2 Hypothesentests für Mittelwert $\mu$

❗ Solange Min $N \geq 50$ ist ZGS erfüllt : wir können Standardnormalverteilung nehmen

Abhängig von der Aufgabenstellung formulieren wir $H_0: \mu = \mu_0$ und $H_1 \to$ (einseitig $\mu < \mu_0$, $\mu > \mu_0$ oder beidseitig $\mu \neq \mu_0$)

Dadurch wird die Aufstellung des Konfidenzintervalls bestimmt

① Wir stellen die **Testvariable** auf:

$$\boxed{Z = \frac{\text{Schätzer} - \text{Vermutung}}{\text{Standardfehler}} = \frac{\bar X - \mu_0}{S/\sqrt{N}}}$$

und berechnen die **Teststatistik**:

$$\boxed{z = \frac{\text{konkretes Ergebnis} - \text{Vermutung}}{\text{Standardfehler}} = \frac{\bar x - \mu_0}{s/\sqrt{N}}} \sim N(0,1)$$

$\bar x$ : Stichprobenmittelwert

$\mu_0$ : vermuteter Mittelwert

$s$ : Standardabweichung

② Berechnung des **p-Werts**:

$$\text{p-Wert} = \begin{cases} P(Z \geq z) & \text{falls} \quad H_A: \mu > \mu_0 \\ P(Z \leq z) & \text{falls} \quad H_A: \mu < \mu_0 \\ 2P(Z \geq |z|) & \text{falls} \quad H_A: \mu \neq \mu_0 \end{cases}$$

$$\boxed{z = \frac{\bar x_1 - \bar x_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}}}$$

③ Wir treffen die **Testentscheidung**: falls PW $\leq \alpha$ verwerfen wir $H_0$

Um den Test durchzuführen, müssen folgende Werte vorliegen: vermuteter Mittelwert $\mu_0$, Stichprobenmittelwert $\bar x$, $N$, $s$, $\alpha$

## 6.3 Hypothesentests für Anteil $p$

❗ Solange Min $Np_0, N(1-p_0) \geq 10$ ist ZGS erfüllt

Wir formulieren $H_0: p = p_0$ und $H_1 \to$ (einseitig: $p > p_0$, $p < p_0$ oder zweiseitig $p \neq p_0$)

① **Testvariable**:

$$\boxed{Z = \frac{\text{Schätzer} - \text{Vermutung}}{\text{Standardfehler}} = \frac{\bar P - p_0}{\sqrt{\frac{p_0(1-p_0)}{N}}}}$$

**Teststatistik**:

$$\boxed{z = \frac{\text{Schätzer} - \text{Vermutung}}{\text{Standardfehler}} = \frac{\hat p - p_0}{\sqrt{\frac{p_0(1-p_0)}{N}}}} \sim N(0,1)$$

$\bar P$: Zufallsvariable (beschreibt die Schätzung)

$\hat p$ : konkret gemessener Anteil in einer Stichprobe

$p_0$ : vermuteter Anteil in der Nullhypothese

② **p-Wert**:

$$\text{PW} = \begin{cases} P(Z \geq z) & \text{falls} \quad H_A: p > p_0 \\ P(Z \leq z) & \text{falls} \quad H_A: p < p_0 \\ 2P(Z \geq |z|) & \text{falls} \quad H_A: p \neq p_0 \end{cases}$$

$$\boxed{z = \frac{\hat p_1 - \hat p_2}{\sqrt{\frac{\hat p_1(1-\hat p_1)}{N_1} + \frac{\hat p_2(1-\hat p_2)}{N_2}}}}$$

③ Wir treffen die **Testentscheidung**: falls PW $\leq \alpha$ verwerfen wir $H_0$

Um den Test durchzuführen, müssen folgende Werte vorliegen: Vermuteter Anteilswert $p_0$, Stichprobenanteilswert $\hat p$, SF $\sigma_p^2 = \frac{p_0(1-p_0)}{N}$, $\alpha$

## 6.4 Hypothesentests für Differenzen von Mittelwerten

❗ Solange Min $N \geq 50$ ist ZGS erfüllt

Hier ist $\Delta = \mu_1 - \mu_2$ die unbekannte Differenz zweier Mittelwerte

Wir formulieren $H_0: \Delta = \underset{\Delta_0}{0}$ und $H_1 \to$ (einseitig: $\Delta < \underset{\Delta_0}{0}$, $\Delta > \underset{\Delta_0}{0}$ oder $\Delta \neq \underset{\Delta_0}{0}$)

① **Testvariable**:

$$\boxed{Z = \frac{\text{Schätzer} - \text{Vermutung}}{\text{Standardfehler}} = \frac{\bar X - \bar Y - \Delta_0}{\sqrt{\frac{S_1^2}{N_1} + \frac{S_2^2}{N_2}}}}$$

**Teststatistik**:

$$\boxed{z = \frac{\text{gemessene Differenz} - \text{Vermutung}}{\text{Standardfehler}} = \frac{\bar x_1 - \bar y_2 - \Delta_0}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}}} \sim N(0,1)$$

$$z = \frac{(\bar x_1 - \bar x_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}}$$

$\bar x_1, \bar x_2$ : Stichprobenmittelwerte

$\Delta_0$ : vermutete Differenz aus der Nullhypothese

② **p-Wert**:

$$\text{PW} = \begin{cases} P(Z \geq z) & \text{falls} \quad H_A: \Delta > \Delta_0 \\ P(Z \leq z) & \text{falls} \quad H_A: \Delta < \Delta_0 \\ 2P(Z \geq |z|) & \text{falls} \quad H_A: \Delta \neq \Delta_0 \end{cases}$$

③ Wir treffen die **Testentscheidung**: falls PW $\leq \alpha$ verwerfen wir $H_0$

Um den Test durchzuführen, müssen folgende Werte vorliegen: Vermutete Differenz der Mittelwerte $(\mu_1 - \mu_2)_0$, $\bar x_1, \bar x_2$, $N_1, N_2$, $s_1, s_2$, $\alpha$

## 6.5 Hypothesentests für Differenzen von Anteilen

❗ Solange Min $Np_0, N(1-p_0) \geq 10$ ist ZGS erfüllt

Hier ist $\Delta = p_1 - p_2$ die unbekannte Differenz zweier Mittelwerte

Wir formulieren $H_0: \Delta = 0$ ($\Delta_0$) und $H_1 \rightarrow$ (einseitig: $\Delta < 0$ ($\Delta_0$), $\Delta > 0$ ($\Delta_0$) oder $\Delta \neq 0$ ($\Delta_0$))

① **Test variable:**

$$\boxed{Z = \frac{\text{Schätzer} - \text{Vermutung}}{\text{Standardfehler}} = \frac{\bar{p}_1 - \bar{p}_2 - \Delta_0}{\sqrt{\frac{\bar{p}_1(1-\bar{p}_1)}{N_1} + \frac{\bar{p}_2(1-\bar{p}_2)}{N_2}}}}$$

$\bar{p}_1, \bar{p}_2$: Stichproben-Zufallsvariablen (nehmen in der Stichprobe die Werte $\hat{p}_1, \hat{p}_2$ an)

$\hat{p}_1, \hat{p}_2$: Stichprobenmittelwerte

$\Delta_0$: vermutete Differenz aus der Nullhypothese

**Teststatistik:**

$$\boxed{z = \frac{\text{gemessene Differenz} - \text{Vermutung}}{\text{Standardfehler}} = \frac{\hat{p}_1 - \hat{p}_2 - \Delta_0}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{N_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{N_2}}}} \sim N(0,1)$$

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)_0}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{N_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{N_2}}}$$

② **p-Wert:**

$$\text{PW} = \begin{cases} P(Z \geq z) & \text{falls} \quad H_A: \Delta > \Delta_0 \\ P(Z \leq z) & \text{falls} \quad H_A: \Delta < \Delta_0 \\ 2P(Z \geq |z|) & \text{falls} \quad H_A: \Delta \neq \Delta_0 \end{cases}$$

③ Wir treffen die **Testentscheidung:** falls $PW \leq \alpha$ verwerfen wir $H_0$

Um den Test durchzuführen, müssen folgende Werte vorliegen: Vermutete Differenz der Anteile $(p_1 - p_2)_0$, $\hat{p}_1$, $\hat{p}_2$, $N_1$, $N_2$, $\alpha$

## 6.6 Güte berechnen

von $H_0$ — wahrer Wert

**Güte linksseitig:** $P\left(Z < \frac{\mu_0 - \mu_1}{\sigma/\sqrt{N}} - z_{1-\alpha}\right)$ oder $P\left(Z < \frac{p_0 - p_1}{\sqrt{\frac{p_1(1-p_1)}{N}}} - z_{1-\alpha}\right)$

**Güte rechtsseitig:** $P\left(Z > \frac{\mu_0 - \mu_1}{\sigma/\sqrt{N}} + z_{1-\alpha}\right)$ $\quad P\left(Z > \frac{p_0 - p_1}{\sqrt{\frac{p_1(1-p_1)}{N}}} + z_{1-\alpha}\right)$

**Güte zweiseitig:** $P\left(Z < \frac{\mu_0 - \mu_1}{\sigma/\sqrt{N}} - z_{1-\alpha/2}\right) + P\left(Z > \frac{\mu_0 - \mu_1}{\sigma/\sqrt{N}} + z_{1-\alpha/2}\right)$ $\quad P\left(Z < \frac{p_0 - p_1}{\sqrt{\frac{p_1(1-p_1)}{N}}} - z_{1-\frac{\alpha}{2}}\right) + P\left(Z > \frac{p_0 - p_1}{\sqrt{\frac{p_1(1-p_1)}{N}}} + z_{1-\frac{\alpha}{2}}\right)$

ist einer der Werte negativ ist er $\approx 0$

---

## Zusatz

- Falls $H_0$ auf einem 10%-Signifikanzniveau verworfen wird, wird sie nicht unbedingt auf einem 5% Signifikanzniveau verworfen

  Bsp. P-Wert von 0.07 $\quad 0.07 < 0.1 \rightarrow$ verworfen $\quad 0.07 > 0.05 \rightarrow$ nicht verworfen

- Signifikanzniveau von $\alpha = 0.05$ bedeutet:
  - → Wahrsch. $H_0$ zu verwerfen obwohl sie richtig ist beträgt 5%.
  - → Wahrsch. $H_0$ nicht zu verwerfen, wenn sie in der Tat richtig ist beträgt 95%

- **Parameter:** alle statistischen Kennzahlen, die zur Grundgesamtheit gehören (KEINE Kennzahlen die die Verteilung in einer Stichprobe beschreiben)

  z.B. Mittelwert $\mu$, Varianz $\sigma^2$, Median $m$

- Das 95% KI ist breiter als das 80% KI → Mehr Sicherheit wird durch Unbestimmtheit erkauft
- Das 95% KI schliesst das 90% KI mit ein (und nicht umgekehrt!)

$H_0$ wird verworfen, wenn $\mu_0$ nicht im $100 \times (1-\alpha)\%$ KI für $\mu$ liegt. gilt nur für zweiseitige Tests ! (nicht für einseitige) da KIs immer zweiseitig sind !

| | Die Einführung eines Produktes erzeugt einen hohen Verlust, wenn der zu erwartende Absatz nicht mehr als 50 beträgt. Aus der Nichteinführung ergeben sich keine Kosten. Also testen Sie | | % |
|---|---|---|---|
| 1: | H0: $\mu \cancel{=} 0$ gegen H1: $\mu \cancel{>} 0$ | | 3 |
| 2: | H0: ~~SPM~~ = 50 gegen H1: ~~SPM~~ > 50 | | 5 |
| 3: | H0: $\mu \cancel{<} 50$ gegen H1: $\mu >= 50$ | | 32 |
| 4: | H0: $\mu <= 50$ gegen H1: $\mu > 50$ ✓ | | 61 |

"Einseitiger Test"