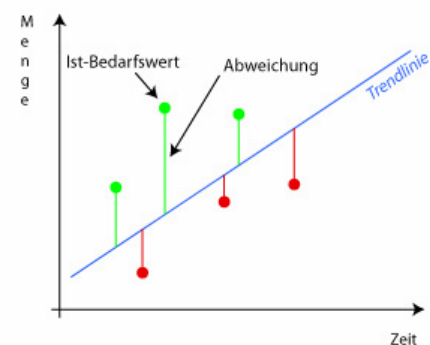
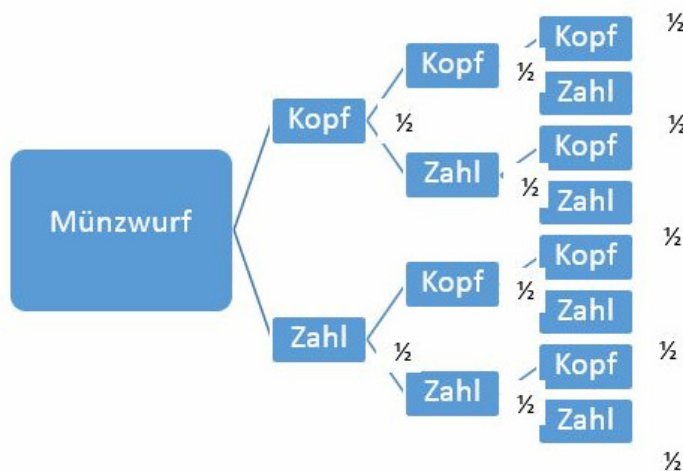
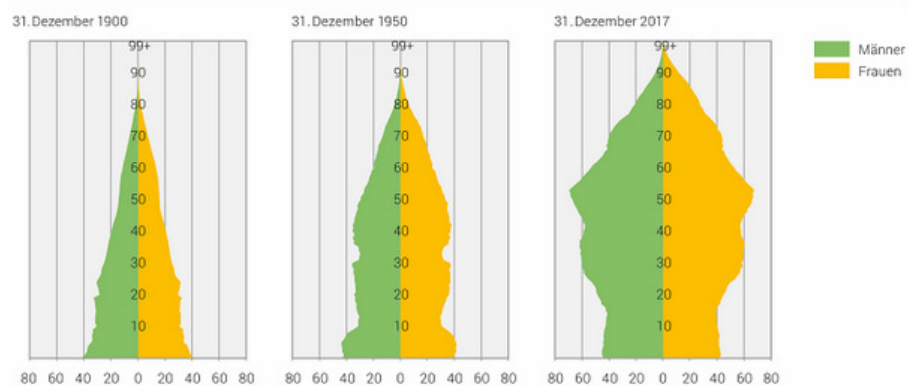


Stochastik

Altersaufbau der Bevölkerung

Anzahl Personen in 1 000



Statistik & Wahrscheinlichkeitsrechnung

Definitionen

Stochastik kommt vom griechischen Wort $\sigma\tau\omega\chi\omega\varsigma$ (stochos) und bedeutet «klug vermuten, die Kunst des Vermutens». Damit werden Vorgänge untersucht, deren Ausgang nicht zum Vornherein bestimmt ist. Sie besteht aus

- **Statistik:**
Der Begriff kommt vom lateinischen Wort status (Zustand). Die Anfänge der **Statistik** gehen auf die Volkszählungen vor unserer Zeitrechnung zurück.
Erst im 18. Jahrhundert wurde sie zur selbständigen Wissenschaft, und ist eigentlich Teilgebiet der Mathematik; und
- **Wahrscheinlichkeitstheorie:**
Sie untersucht Gesetzmässigkeiten zufälliger Ereignisse.
Damit sollen auch Voraussagen für noch nicht aufgetretene Ereignisse getroffen werden können.

Zweck und Inhalt des Kapitels

- Grundbegriffe der Statistik kennen und anwenden
- Erhebung von ausgewählten Daten
- Aufbereitung
- Rückschlüsse auf die Gesamtheit ziehen und entsprechend handeln
- Daten ordnen, also zu betrachten, wo sie in der bestimmenden Umgebung angesammelt sind, wie sie sich dabei verteilen und was daraus an sinnvoller Information herausgelesen werden kann.
- Grundformel der Wahrscheinlichkeitsrechnung
- einfache Wahrscheinlichkeiten und deren Verknüpfungen betrachten.
- Berechnungsmöglichkeiten mit Werkzeugen wie dem Baumdiagramm oder mathematischen Formeln wie der Binomialverteilung und den Bernoulli - Experimente.

Auch für dieses Kapitel ist die Beherrschung der Bedienung Ihres Taschenrechners ein Muss.

Suchen Sie ein Manual und machen Sie sich ein Bild, ob der Rechner eingebaute Statistikfunktionen hat.

Beispiel 1 für Statistik: Datenerhebung von Körpergrössen

- Annahme: die Regierung hat beschlossen, für Lernende der GIBB Schul-Uniformen einzuführen.
- Sie haben den Auftrag, innert kurzer Zeit eine Abschätzung der Körpergrössen der Lernenden zu erstellen.

Statistische Begriffe:

Die etwa **7000 Berufsschüler** bilden die **Grundgesamtheit** oder **Population** Ihrer Umfrage.

Da Sie nicht alle befragen können, treffen Sie eine Auswahl. Sie beschränken sich aus Zeitgründen auf 2 Informatikerklassen pro Lehrjahr, das sind etwa 140 Lernende. In der Statistik nennt man **die Informatikern Lernenden** die **Stichprobe** und die Zahl **140** ist den **Umfang der Stichprobe**.

Die Grösse (nicht etwa das Geschlecht oder die Korpulenz) stellt das interessierte **Merkmal** dar, und die **Ausprägung des Merkmals** ist die **Zahl in cm**. Das Ganze notieren Sie auf einem **Erhebungsformular**, die sogenannte **Urliste**.

Übung 1

Stellen Sie die obigen Begriffe in der Tabelle zusammen!

Wer oder was	Statistischer Begriff
Erhebungsformular, Fragebogen	
Alle Berufsschüler	oder
Die ausgewählte Klasse(n) Informatiker	
140 (Anzahl befragte Lernende)	
Körpergrösse	
z.B. 175 cm	

Vor jeder Datenerhebung muss klar sein, was damit bezweckt wird! Eine zweckmässige Struktur der **Urliste** kann viel nachträgliche Arbeit ersparen. Die Urliste kann die Form einer Strichliste haben. Sinnvoll ist es heute bestimmt, grössere Datenbestände elektronisch zu erfassen, ev. direkt in eine Datenbank aufzunehmen. Die Daten können einfacher verarbeitet werden, z.B. ordnen oder sortieren. In einer Umfrage können **mehrere** Merkmale vorkommen.

Sorgfältig muss ebenfalls die **Stichprobe** ausgewählt werden. Sie soll **repräsentativ** (stellvertretend ähnlich) für die Grundgesamtheit sein. Eine umfassende Stichprobe erfasst **zwischen 600 und 1200 Werte** (deshalb wird in öffentlichen Umfragen meist gesagt, dass z.B. 1150 Leute befragt wurden). Damit ist die Aussage in der Regel präzise genug, um ein Bild der Grundgesamtheit zu erhalten. Je weniger Werte erfasst werden, umso grösser ist die Gefahr, daneben zu liegen, also eine ganze Gruppe nicht berücksichtigt zu haben.

Aus Effizienz- und Kostengründen wird die **Grundgesamtheit** selten erfasst. Nur etwa bei Volkszählungen, die aber nicht jedes Jahr neu durchgeführt werden.

Sorgfältig ausgewählte und **gut vorbereitete Stichproben** sind besonders wichtig und sparen Aufwand, Kosten und Ressourcen.

Es ist einleuchtend, dass z.B. eine Autofirma nicht mit allen Neuwagen Crashtests durchführt...

Nach der Erfassung der Daten ordnet man sie meistens, falls möglich oder sinnvoll. Einfach zu ordnen sind zum Beispiel Zahlen, Datum, Ranglisten, Preise usw. Schwieriger wird hingegen die Bestimmung einer Ordnung bei Lieblingsfarben, Essensgewohnheiten, Reisezielen usw. Doch auch bei diesen Themen kann man bestimmte Reihenfolgen erstellen, je nachdem, was damit bezweckt wird.

Nehmen wir bei unserer Umfrage nach den Schuluniformen folgende Daten für **eine** Klasse an. Sie haben zusätzlich zur Körpergrösse noch das Merkmal Geschlecht (**m/w**) erfasst. 22 Personen:

178m 164w 176w 188m 173m 160w 176m 181m 177m 158w 169m
 173m 178w 180m 173m 173w 179m 183m 168m 171w 173w 176m

Absolute und relative Häufigkeit

Die absolute Häufigkeit ist die Anzahl der Ereignisse, sog. Elementarereignisse.

Sie gibt an, wie viele Elemente mit jeweils dem **gleichen** Merkmal gezählt worden sind.

In unserem Beispiel der Schuluniformen kommt das Elementarereignis «Körpergrösse 158 cm» nur 1x vor. Die **absolute** Häufigkeit für dieses Ereignis ist also 1.

Die **absolute Häufigkeit** wird mit dem **Formelzeichen** H_n angegeben. Beispiel:

$H_n(A) = 1$ bedeutet: Die absolute Häufigkeit des Ereignisses A (Körpergrösse 158 cm) ist 1 (1x).

Die relative Häufigkeit ist der **Anteil** des Ereignisses A an der **Grundgesamtheit** oder an der Stichprobe. Die relative Häufigkeit wird mit dem **Formelzeichen** h_n angegeben.

$$h_n(A) = \frac{H_n(A)}{n} \quad n = \text{Gesamtzahl aller Elemente in der Menge.}$$

Bei unserem Beispiel ist die relative Häufigkeit des Ereignisses A (Körpergrösse 158 cm) folgende:

$$h_n(A) = \frac{H_n(A)}{n} = \frac{1}{22} = 0,04545 \text{ oder } 4,545 \%$$

Übung 2

Ordnen Sie die Daten nach Grösse und Geschlecht.

Erstellen Sie eine Strichliste (**linke** Hälfte der Tabelle auf nächster Seite).

Lassen Sie die rechte Tabelle noch leer.

Übung 3

Füllen Sie jetzt die **rechte** Tabelle der Übung 2 aus. Nehmen Sie die Geschlechter m und w nicht mehr auseinander. Die Häufigkeiten sollen sich nur auf die Körpergrösse beziehen.

Körpergrösse in cm	Stichliste: weiblich	Strichliste: männlich
158	I	
Befragte 22	Total w	Total m

Strichliste: Total m+w	Absolute Häufigkeit (Anzahl)	Relative Häufigkeit %
22	22	100 %

Sobald mehr als nur einige Daten erhoben sind, wird das Sortieren etwas mühsam und auch fehleranfällig. Der Gebrauch von Anfang an einer Software oder zumindest einer Strichliste ist sehr empfehlenswert. Nachdem unsere Daten erhoben und sortiert sind, können wir sie weiterverarbeiten.

Wie haben Sie Übung 2+3 empfunden?

☐ Spannend, möchte alle 140 Werte der Stichprobe sortieren

☐ Spannend, aber es genügt☐ Ziemlich langweilig

☐ Muss ich nicht machen, das beherrsche ich blind

Wie beurteilen Sie die Gefahr, dass Fehler auftreten?

☐ Alles stimmte auf Anhieb☐ Ich musste 2-3x etwas korrigieren☐ Mein Ergebnis stimmt nicht☐ Habe aufgegeben

