

# Deskriptive (Beschreibende) Statistik

## 1 Einleitung

### 1.1 Grundlagen

Eine grundlegende Aufgabe der Statistik oder statistischen Datenanalyse besteht darin, Eigenschaften von bestimmten gleichartigen Objekten - der sogenannten **Grundgesamtheit** - zu gewinnen. Man möchte feststellen, wie die Ausprägungen eines Merkmals in der Grundgesamtheit verteilt sind.

Aus praktischen Gründen ist es meistens nicht möglich, alle Objekte zu untersuchen. Eine sogenannte Vollerhebung bedeutet oft einen hohen Zeit- oder Kostenaufwand oder die Zerstörung der untersuchten Objekte (z.B. von Neonröhren). Man entnimmt deshalb aus der Grundgesamtheit eine sogenannte **Stichprobe vom Umfang  $n$**  und untersucht ihre Eigenschaften mit den Methoden der

**Beschreibenden Statistik**. Die Daten werden geordnet, graphisch dargestellt und mit charakteristischen Kennzahlen beschrieben.

Sofern jedes Element der Grundgesamtheit die gleiche Chance hat, in die Untersuchung einbezogen zu werden, m.a.W. wenn die Auswahl repräsentativ ist (was in der Praxis gar nicht immer einfach ist), so ermöglicht die

**Wahrscheinlichkeitsrechnung** Rückschlüsse auf die Eigenschaften der Grundgesamtheit. Solche Rückschlüsse sind das Thema der so genannten

**Schliessenden Statistik** (auch Inferenzstatistik, induktive Statistik). Jede Vorhersage über die Grundgesamtheit ist allerdings mit einer gewissen Unsicherheit, dem Stichprobenfehler, behaftet. Als Begründer der Schliessenden Statistik gelten K. Pearson und R.A. Fisher.



R.A. Fisher



K. Pearson

Historisches:

Bereits in der Bibel ist eine Volkszählung erwähnt. Als eigentliche Geburtsstunde der statistischen Tätigkeit wird der Zeitpunkt angesehen, an dem Joh. Graunt (1642 -1674) die Geburts- und Sterberegister zur Schätzung der Altersstruktur der Londoner Bevölkerung verwendete. Die buchhalterische Führung der Register war nicht Selbstzweck, sondern die statistische Arbeit bestand darin, Folgerungen aus den Daten zu ziehen.

## 1.2 Anwendungsbereiche der Statistik

Einige Beispiele:

Politik:

Hochrechnung von Abstimmungsergebnissen, Volkszählung als Planungsgrundlage für die Altersvorsorge.

Wirtschaft:

Entwicklung der Lebenshaltungskosten, der Teuerung oder von Aktienindizes.

Medizin:

Erprobung der Wirksamkeit neuer Medikamente.

Sport:

Vergleich von Trainingsmethoden.

Öffentlicher Verkehr:

Fahrgastbefragungen als Grundlage für die Fahrplangestaltung.

Industrie:

Qualitätskontrolle bei der Herstellung von Massenartikeln, Produktionsplanung.

Naturwissenschaften:

Studien über die Umweltbelastung, Versuchsplanung bei Experimenten.

Psychologie und Soziologie:

Marktforschung.

## 1.3 Datentypen

Folgende Datentypen werden unterschieden:

**Metrische Daten** können auf dem Zahlenstrahl abgetragen werden, wobei die Abstände zwischen den Daten relevant sind.

Beispiele:

Längen, Massen, gemessenen Zeiten (stetig), täglicher Zigarettenkonsum, Anzahlen (diskret).

Bei **ordinalen Daten** ist eine Rangordnung möglich, die Abstände sind ohne Bedeutung.

Beispiele:

Qualifikation (sehr gut, gut, genügend, ..), Rang bei einem Wettkampf, IQ(?)

Bei **nominalen Daten** ist keine Rangordnung möglich:

Beispiele:

Geschlecht, Antwort (ja, nein, weiss nicht, ...), Blutgruppe, Haarfarbe, Wohnort.

## 1.4 Methoden zur Gewinnung von Daten

Bei einem **Experiment** befinden sich die Wirkungen der Störfaktoren unter Kontrolle oder sie können ausgeschaltet werden.

Beispiele:

Dosierung eines Medikaments, Leistungsmessung auf einem Fahrradergometer.

Bei einer **Beobachtung** sind die Störfaktoren nicht oder nur zum Teil bekannt.

Beispiel:

Spielverhalten von Kindern.

Bei einer **Erhebung oder Teilerhebung** auf Stichprobenbasis resultieren die Störfaktoren vor allem aus der Heterogenität der untersuchten Grundgesamtheit.

## 1.5 Datenschutz

Persönliche Daten sind schützenswert und dürfen nur unter bestimmten Voraussetzungen erhoben werden!

## 1.6 Vereinfachtes Musterbeispiel aus der Qualitätskontrolle für das Vorgehen in der Statistik (Quelle: Statistikkurs. M. Steiner FHNW)

Bei der Herstellung von Massenartikeln z.B. Neonröhren werden nicht alle Stücke gleich ausfallen, sondern mehr oder weniger stark von einem Sollwert abweichen. Der Hersteller wird deshalb angeben, welche Abweichungen nach oben oder unten auftreten dürfen. Er muss deshalb die Produktion periodisch überwachen, um sicher zu gehen, dass die Ware den Anforderungen des Kunden entspricht. Dazu wird aus der Grundgesamtheit (z.B. der Tagesproduktion) eine Zufallsstichprobe gezogen. Diese ermöglicht mit Hilfe der Wahrscheinlichkeitsrechnung gewisse Rückschlüsse auf die Grundgesamtheit.

a)

Formulierung des Problems:

Welche mittlere Lebensdauer haben die in einem Betrieb verwendeten Neonröhren?

b)

Planung und Durchführung des Experiments:

Es wird zufällig eine Stichprobe von Neonröhren mit dem Stichprobenumfang  $n$  gewählt und ihre Lebensdauer untersucht

c)

Beschreibung der experimentellen Daten in übersichtlicher Form (Tabellen, grafische Darstellung, Boxplot)

d)

Schluss von der Stichprobe auf die Grundgesamtheit

Auf Grund der Stichprobe kann man entscheiden, ob die mittlere Lebensdauer in der Stichprobe signifikant von der mittleren Lebensdauer in der Grundgesamtheit abweicht oder nicht. Die Theorie der Wahrscheinlichkeitsrechnung ermöglicht es, die Genauigkeit dieser Entscheidung abzuschätzen. Eine Verbesserung der Schätzung könnte erreicht werden, indem der Umfang der Stichprobe vergrößert wird, was allerdings zu grösserem Zeitaufwand und grösseren Kosten führt.