

## Statistik

### Definition

Statistik ist das methodische Vorgehen bei der Beschaffung von Informationen, die man braucht um vernünftige Entscheidungen treffen zu können.

Wichtige Gesichtspunkte

- Vielzahl von Informationen
- Straffung der Informationen, damit sie überschaubar werden
- Nicht nur nützliches Hilfsmittel, sondern Voraussetzung um sinnvolle Ergebnisse zu erzielen
- Hat stets einer bestimmten konkreten Aufgabe zu dienen

### Statistische Daten

Grundsätzlich 2 Kategorien

- Verbale Informationen (z. B. heute ist es kalt)
- Quantitative, zahlenmäßige Informationen (Es herrschen  $-10^{\circ}$ )

Der Begriff Statistik hat eine doppelte Bedeutung.

- Methoden zur Datenbeschaffung und Verarbeitung
- Ergebnis des Methodeneinsatzes (die zusammengefaßten statistischen Daten)

### Statistisches Material

Zahl der Einheiten = Umfang der Gesamtheit

Angaben, die für eine bestimmte Gesamtheit richtig sind, lassen sich nicht ohne weiteres auf andere Gesamtheiten übertragen.

Zur präzisen Bestimmung

3 Grundfragen (Seite 29)

1. Wie wird die statistische Masse abgegrenzt ? (Wen frag ich?)
2. Welche Merkmale der stat. Masse sollen untersucht werden?
3. Welche Erfassungsmethoden müssen angewendet werden?

1. Abrenzungskriterien

- Sachliche Abgrenzung (z.B. lebende Personen)
- Räumliche/örtliche Abgrenzung (z.B. in der BRD wohnen)
- Zeitliche Abgrenzung
  - Angabe eines Zeitpunktes (genaues Datum – z.B. 25.06.1986)
  - Angabe eines Zeitraumes

Beispiele für Bestandsmassen (Menge aller):

- Einwohner in der BRD
- Unternehmen in NRW
- Kraftfahrzeuge in München
- Schiffe auf dem Rhein

### Bewegungsmassen

Kurzlebige Gesamtheiten, aus Ereignissen

Können nur sinnvoll für einen Zeitraum erfaßt werden.

- Umsätze von Unternehmen
- Regenfälle in einem bestimmten Gebiet
- Geburten
- Todesfälle

### Totalerhebung

Ergebnisse von Totalerhebungen großer Gesamtheiten sind oft fehlerhaft (systematische Fehler)

Durchführung der Erhebung und Kontrolle wird aus Kostengründen weniger sorgfältig

### Teilerhebung

Teilerhebung (Auswählerhebung) wird auch als Stichprobe bezeichnet

Ergebnisse einer Teilerhebung werden auf die Ausgangsgesamtheit hochgerechnet d. h. verallgemeinert. (Hochrechnung)

Vorteile:

- Auch machbar in Fällen, in denen eine Totalerhebung praktisch unmöglich oder sachlich unsinnig ist
- Billiger
- Schnellere Ergebnisse
- Ausschließen von systematischen Fehlern

### Statistische Merkmale

Unterschiedsmerkmale (nominal)

Zwischen den Ausprägungen besteht lediglich ein Unterschied

z. B. Geschlecht, Familienstand, Staatsangehörigkeit, Beruf, Augenfarbe, Konfession etc.

Rangmerkmale (ordinal)

Zeigen Unterschiede an

z. B. Schulnoten, Güteklassen, Härtegrad

Abstandsmerkmale

kontinuierliche Merkmale (beliebe Werte innerhalb eines bestimmten Intervalls)

z. B. Körpergröße, Alter, Gewicht

theoretisch unendlich viele Ausprägungen

diskrete Merkmale (nur einzelne Werte innerhalb eines bestimmten Intervalls)

z. B. Kinderzahl (nur Werte von 0,1,2,3,4 ...), Einwohnerzahl, Zahl der Kühe, Einkommen

werden durch einen Zählvorgang erfaßt

oft nur wenige Werte

### Auswertung statistischer Daten (Buch S. 46)

$n$  = Umfang der Gesamtheit, Anzahl der Einheiten

$x$  = Merkmal

$f$  = Häufigkeit

$x_i$  = Merkmalsausprägung der  $i$ -ten Einheit

$f_i$  = Häufigkeit der  $i$ -ten Merkmalsausprägung

Beispiel:

$X_3 = 1,5$  Punkte

$F_3 = 6$  (1,5 Punkte kommt 6x vor)

$$\sum_{i=1}^n X_i \quad \text{Laufindex von 1 bis n}$$

Beispiel:  $\sum_{i=5}^{10} X_i$

Summe des Laufindex

Laufindex von 5 bis 10, also

$$X_5 + X_6 + X_7 + X_8 + X_9 + X_{10}$$

### Häufigkeitsverteilung (S. 52)

- Absolute Häufigkeit (Zahl der Einheiten die eine bestimmte Ausprägung aufweisen)
- Relative Häufigkeit (Anteil der Einheiten an der Gesamtheit – meist in Prozent)

### (Häufigkeits-)Summenverteilung = kumulieren

Ermittlung Summenverteilung für Merkmalswerte, die nicht mit Klassengrenzen zusammenfallen.

Annahme: gleichmäßige Verteilung über die gesamte Klassenbreite.

Aufenthalt unter ... Jahren	Anzahl
1	340
4	578
6	223

Frage: Wie viele Ausländer leben weniger als drei Jahre in der BRD?

340 Ausländer lebten weniger als 1 Jahr in der Bundesrepublik.  
Der Wert der Summenverteilung muß also mindestens 340 sein.

578 Ausländer lebten zwischen 1 und 4 Jahren in der Bundesrepublik. Verteilt man diese 578 auf die beiden Zeitabschnitte, und zwar entsprechend deren Länge, entfallen auf den Zeitraum von 1 bis unter 3 Jahren 385 und auf den Zeitraum von 3 bis unter 4 Jahren

----- 1 -----2----- 3 -----4----- 5  
340 -- |-----578 -- -----|

$$\frac{578}{3} \times 2 = 385$$

Wert der Summenverteilung  $340 + 385 = 725$

### Maßzahlen (S. 65)

Maßzahlen nach Auswahl  
z. B. häufigste Wert

Maßzahlen durch Addition

Durch Addition der Häufigkeiten ergibt sich die Gesamtzahl der erfaßten Einheiten (Umfang der Gesamtheit)  
Durch Addition der Merkmalswerte erhält man den Totalwert (z. B. Volkseinkommen als Summe aller Einzeleinkommen)

Maßzahlen durch Subtraktion

Subtrahiert man zwei Merkmalswerte voneinander, erfährt man den Größenunterschied.

### Maßzahlen durch Multiplikation

Multiplikation der Werte von zwei Merkmalen entstehen zusammengesetzte Größen.

z. B. Umsatz eines Gutes aus dem Produkt von Menge und Preis

### Maßzahlen durch Division

Die meisten statistischen Maßzahlen entstehen durch Division.

z.B. Gesamtbetrag eines Merkmals (den Totalwert) durch die Zahl der Einheiten, spricht man von Durchschnittsbildung.

## Darstellung statistischer Daten (S. 71)

### Aufbau einer Tabelle

- Titel bzw. Überschrift
- Hauptteil
- Quellenangabe

### Üblich, verwendete Zeichen in Tabellen

- = nichts ist vorhanden, der Zahlenwert ist 0

0 = Zahlenwert ist von 0 verschieden, jedoch kleiner als die Hälfte der kleinsten Einheit.

z.B. Bevölkerung in Tausend, alle Gruppen die 499 oder weniger haben, werden nach der üblichen Rundungsregel abgerundet, also auf 0

- = kein Nachweis vorhanden, meist weil der Zahlenwert unbekannt ist

## Grafische Darstellungen

für Unterschieds- und Rangmerkmale = Stab-/Säulendiagramm

diskrete Merkmale werden mit einer Lücke zwischen den Säulen gezeichnet

Bei einem Vergleich von zwei Häufigkeitsverteilungen, empfiehlt es sich die Basislinie um 90° zu drehen und auf beiden Seiten die Stäbe waagrecht darzustellen.

## Verhältniszahlen

$$\text{Verhältniszahl} = \frac{\text{Berichtsgröße}}{\text{Basisgröße}}$$

Bei Messzahlenreihen Bezug zum Basisjahr

Ergebnis ist meistens eine Prozentzahl.

### 3 Arten von Verhältniszahlen

- Gliederungszahlen
- Meßzahlen
- Beziehungszahlen

## Gliederungszahlen (S 116)

Gliederungszahlen, entstehen, wenn Teilbeträge eines Merkmals auf den Totalwert bezogen werden. Berichtsgröße ist ein Teil der Basisgröße.

z.B. Anteil einer Gruppe  $\frac{\text{Umsatz der Gruppe}}{\text{Gesamtumsatz}}$

## Meßzahlen

Sachlich gleiche, aber zeitlich oder örtlich verschiedene Merkmalswerte

z. B. wenn man die Einwohnerzahlen der deutschen Großstädte jeweils auf die Einwohnerzahl Berlins bezieht.

## Beziehungszahlen

Sachlich verschiedenartige Maßzahlen

$$\text{z. B. } PkwDichte = \frac{\text{AnzahlPkw}}{\text{Anzahl Einwohner}}$$

## Mittelwerte

Häufigster Wert

Ist die Ausprägung die die größte Häufigkeit hat.

Je stärker eine Ausprägung dominiert, desto Aussagekräftiger ist sie

Bei mehreren gleich häufig vorkommenden Werten, ist eine Angabe

Wenig aussagekräftig.

Zentralwert

Ist der Merkmalswert, der eine Häufigkeitsverteilung in zwei gleichgroße Teile trennt.

Orientierung an der Zahl der Werte. Diese müssen der Größe nach sortiert sein.

Arithmetischer Mittelwert (stinknormaler Durchschnitt)

$$\text{ArithmetischerMittelwert} = \frac{\text{Totalwert}}{\text{AnzahlderEinheiten}}$$