

3. Gewinnung von Daten

Warum heißt es "Statistischer Fragebogen"? gebogen werden doch erst die Antworten!

Anonymus

3.1. Allgemeines

Statistische Datenerhebungen unterscheidet man nach der Art der Datengewinnung in **primärstatistische** und **sekundärstatistische** Datenerhebungen:

- bei der Primärerhebung werden die Daten eigens für die Untersuchung *neu gewonnen* ⇒ welche Arten der Datengewinnung gibt es?
- bei der Sekundärerhebung verwendet man schon vorhandene, *für andere Zwecke gewonnene* Daten ⇒ welche Organisationen liefern welche Daten? wie ist ihre Qualität?

Desweiteren unterscheidet man statistische Erhebungen nach dem Umfang in **Voll-** und **Teilerhebungen** ⇒ Vor und Nachteile? Wann wendet man Teilerhebungen an? Ist bei Teilerhebungen die Auswahl der Teilmenge repräsentativ?

3.2 Primärerhebung: Erhebungsarten

1. Befragung

- (a) Mündlich durch Interviewer: Hohe Qualität, aber aufwändig und teuer,
- (b) schriftlich (postalisch durch Fragebogen oder elektronisch über's Internet): Geringer Aufwand, aber die Güte ist geringer, die Repräsentativität ist fraglich, und der zeitliche Aufwand höher.

2. Beobachtung und Messung.

- Standard bei naturwissenschaftlichen Daten,
- aber auch in Verkehr&Wirtschaft anwendbar: Verkehrszählungen, Beobachtung von Kassenschlangen, Zeitnehmer bei der Arbeitsvorbereitung etc.

Vorteil: unverfälscht ("objektiv")

Findet die Messung automatisch statt (Induktions-Zählschleifen unter Autobahnen, Stromzähler etc), spricht man auch von **automatischer Erhebung**

3. Experiment. Typische Anwendungen:

- Qualitätssicherung,
- Produkttest ("Stiftung Warentest"),
- psychologische Experimente, z.B. zum Kaufverhalten.

3.3 Sekundärerhebung: Datenquellen

Volks- und betriebswirtschaftliche Daten liegen i.A. schon vor. Die Datenproduzenten (“Träger der Statistik”) lassen sich in zwei Gruppen einteilen:

1. Träger der amtlichen Statistik: staatliche oder vom Staat (bzw. von mehreren Staaten) gebildete und kontrollierte Organisationen:

- Statistisches Bundesamt, z.B. “Statistisches Jahrbuch”, “Verkehr in Zahlen”
- Statistische Landesämter und Ämter der Städte
- Ministerien
- Bundesanstalt für Arbeit
- EU-Behörden.

Von vergleichbarer Qualität sind die Daten einiger globaler Institutionen:

- UNO und ihre Tochterorganisationen wie WHO (Gesundheit), UNDP (Entwicklung)
- Weltbank, IWF, OECD (Wirtschaft)

2. Träger der nichtamtlichen Statistik:

- Wirtschaftsverbände
- wirtschaftswissenschaftliche Institute
- Markt- und Meinungsforschungsinstitute (IFO, Al-lensbach)
- Die Unternehmen selbst

3.3(b) Sekundärerhebung: Datenquellen

Vorteile der amtlichen Statistik

- Objektivität
- Vollständigkeit
- Daten häufig kostenlos und/oder schnell verfügbar
- Datenqualität meist über alle Zweifel erhaben (obwohl auch die primären Träger viele Datenerhebungen von sekundäre Trägern durchführen lassen!)

Vorteile der nichtamtlichen Statistik

- flexibler
- individueller

Aber Gefahr der Verfälschung durch Eigeninteressen!

Frage: Wie kann man häufig sekundärstatistische Daten schnell gewinnen?

3.4 Teilerhebung: Methoden

Ziel aller Methoden ist, einen **möglichst repräsentativen** Teil der statistischen Masse als **Stichprobe** zu erfassen. Im Rahmen der **induktiven Statistik** wird die Stichprobe dann auf die ursprüngliche Masse **hochgerechnet**. Die nun behandelte deskriptive Statistik betrachtet grundsätzlich nur die gesamte statistische Masse. Häufig verwendete Methoden sind:

- Zufallsauswahl
- Abschneideverfahren (v.a. amtliche Statistik): Nur die wichtigsten Merkmalsträger werden untersucht
- Quotenverfahren (v.a. Markt- und Meinungsforschungsinstitute): Bestimmte Merkmalsausprägungen sollen in der Teilmenge genau dieselbe relative Häufigkeit haben wie in der Gesamtheit, z.B. Geschlecht, Altersgruppe, Wohnortgröße etc.

3.5. Urliste und Datenaufbereitung

Die **Urliste** ist in der Regel das direkte Ergebnis einer primären Datenerhebung (die Daten einer sekundären Erhebung liegen meist schon aggregiert bzw. klassifiziert vor, s.u.):

Es wurde das Merkmal X für eine statistische Masse von n Einheiten erhoben. Die Zusammenstellung der Ergebnisse $x_i, i = 1, \dots, n$ in der Reihenfolge der Erfassung heißt **Urliste**

- Sie stellt den höchstmöglichen Informationsgehalt dar, ist aber auch am unübersichtlichsten.
- ordnet man die Elemente der Urliste, erhält man eine **statistische Reihe**
 - Falls das Ordnungsmerkmal die Zeit ist: **Zeitreihe** (\Rightarrow Kap. 10 ff)
 - Falls das Ordnungsmerkmal der Ort oder ein sachliches Kriterium ist: **Querschnittsreihe**

Beispiel: Hypothetische Klausurergebnisse

Name	Ergebnis (Punkte)	Name	Ergebnis (Punkte)
W	24	X	14
Q	19	R	9
E	12	F	14
I	20	J	10
K	18	L	22
C	14	D	11
M	17	N	4
A	17	B	7
U	15	V	15
O	19	P	7
G	2	H	14
S	12	T	16