

Beschreibende Statistik

In der beschreibenden Statistik geht es um die Erfassung, Auswertung und Darstellung von experimentell oder empirisch gewonnenen Daten. Dabei werden endliche Mengen an Objekten hinsichtlich bestimmter Eigenschaften untersucht. Dabei werden allgemein folgende Schritte durchlaufen:

- Zunächst müssen in der beschreibenden Statistik alle für die Analyse relevanten Daten vollständig erhoben werden.
- Das bei der Datenerhebung gewonnene, oftmals sehr umfangreiche Datenmaterial muss als nächstes in eine übersichtliche Form gebracht werden, üblicherweise in eine Tabelle oder eine Graphik.
- Anschließend kann mit der Analyse der Daten begonnen werden. Hierbei lassen sich die Daten beispielsweise mittels wichtiger Kennzahlen wie Mittelwert und Streuungsmaß charakterisieren, ebenso können beispielsweise zeitliche Trends oder Abhängigkeiten zwischen mehreren Größen untersucht werden.
- Zuletzt können die Ergebnisse der Analyse interpretiert werden.

system-message

ERROR/3 in `index.rst`, line 30
Unknown directive type "index".

```
.. index:: Merkmal, Merkmalsträger, Ausprägung
```

backrefs:

Merkmale, Merkmalsträger und Grundgesamtheit

Als (Untersuchungs-)Merkmal wird die interessierende statistische Information bezeichnet. Ein einzelnes Objekt, das dieses Merkmal besitzt, nennt man Merkmalsträger. Die möglichen Werte, die ein Merkmal annehmen kann, heißen Merkmalswerte oder Ausprägungen dieses Merkmals.¹

system-message

ERROR/3 in `index.rst`, line 44
Unknown directive type "index".

```
.. index:: Grundgesamtheit
```

backrefs:

Die Menge an Objekten G , die hinsichtlich einem oder mehrerer zu untersuchender Merkmale gleichwertig sind, wird als "Grundgesamtheit" oder "Population" bezeichnet. Bei der Festlegung der Grundgesamtheit werden müssen klare Abgrenzungen getroffen werden, beispielsweise müssen räumliche oder zeitliche Einschränkung vorliegen; die Mitglieder der Grundgesamtheit müssen somit nicht nur Träger des Untersuchungsmerkmals sein, sondern auch übereinstimmende Abgrenzungsmerkmale besitzen.

Beispiel:

- Bei einem naturwissenschaftlichen Experiment sind die einzelnen Messungen die Merkmalsträger, die ihrerseits Messdaten als Merkmale enthalten.
- Bei einer Inventur werden zu einem bestimmten Zeitpunkt alle Objekte eines räumlich abgegrenzten Bereichs beispielsweise hinsichtlich ihrer Funktionsfähigkeit als Merkmal untersucht.

Die Mächtigkeit $n = |G|$ der Grundgesamtheit ist gleich der Anzahl ihrer Objekte. In Tabellen werden die einzelnen zu untersuchenden Merkmale häufig einem Buchstaben A, B ... zugeordnet, die einzelnen zu einem jeweiligen Merkmalsträger gehörenden Merkmalswerte werden zeilenweise durchnummeriert und in der jeweiligen Spalte eingetragen.

Meist ist bei einer Datenerhebung nicht möglich, alle Mitglieder der Grundgesamtheit zu untersuchen ("Vollerhebung"). In diesem Fall muss sich die Statistik mit einer kleineren, möglichst repräsentativen Stichprobe auskommen und von dieser auf die Gesamtheit schließen.

Qualitative und quantitative Merkmale

Merkmale können allgemein in zwei Gruppen unterteilt werden:

- *Qualitative* Merkmale lassen sich nur verbal beschreiben, es können nur Namen oder Klassenbezeichnungen als Werte vorkommen.

Handelt es sich bei den Merkmalswerten um Namen, so spricht man auch von artmäßigen Merkmalen. Beispiele für derartige Merkmale sind Familiennamen, Geschlecht, Farbbezeichnungen, usw.

Handelt es sich bei den Merkmalswerten um Klassenbezeichnungen, so spricht man auch von intensitätsmäßig abgestuften Merkmalen. Ein Beispiele hierfür sind Schulnoten ("sehr gut", "gut", usw.).

Qualitative Merkmale lassen sich zudem in "häufbare" und "nicht häufbare" Merkmale unterscheiden. Ein qualitatives Merkmal ist häufbar, wenn ein Merkmalsträger mehrere Merkmalswerte gleichzeitig aufweisen kann; beispielsweise kann eine Person gegebenenfalls mehrere Berufsausbildungen absolviert haben. Ein qualitatives Merkmal ist nicht häufbar, wenn ein Merkmalsträger nur genau einen Merkmalswert aufweisen kann; beispielsweise hat jede Person genau eine Augenfarbe.

- *Quantitative* Merkmale können als Vielfaches einer Einheit ausgedrückt werden, beispielsweise Zeitdauer, Energiebedarf, usw.

Können bei einem quantitativen Merkmal nur ganzzahlige Werte auftreten, so spricht man von einem diskreten Merkmal. Ein Beispiel hierfür sind Stückzahlen.

Können bei einem quantitativen Merkmal beliebige Werte auftreten, so spricht man von einem stetigen oder kontinuierlichen Merkmal. Beispiele hierfür sind Zeitdauern, Längenangaben, usw.

Um eine Vielzahl unterschiedlicher quantitativer Messwerte abzubilden, können diese in einzelne Intervalle zusammengefasst werden. Anstelle (sehr) viele Einzelergebnisse aufzulisten, genügt es damit, die Anzahl an Werten in den einzelnen Intervallen anzugeben. Üblicherweise werden zwischen 5 und 20 einzelne Intervallen mit jeweils gleich großen Intervallen und eindeutig zuzuordnenden Intervallgrenzen gewählt. Durch diese Methode gehen zwar einerseits die statistischen Informationen der Einzelmessungen teilweise verloren, andererseits werden dafür die Ergebnisse "komprimiert" und somit übersichtlicher.

Statistische Mess-Skalen

Mittels einer Mess-Skala können die möglichen Merkmalswerte nach bestimmten Ordnungsprinzipien dargestellt werden. Für qualitative Merkmale werden Nominal- oder Ordinalskalen verwendet, für quantitative Merkmale kommen oftmals Intervall- oder Verhältnisskalen zum Einsatz. Im folgenden Abschnitt werden diese Skalen näher beschrieben.

Nominalskala

Eine Nominalskala hat die möglichen Namen eines quantitativen Merkmals als Skalenwerte. Diese werden gleichberechtigt nebeneinander angeordnet. Die einzelnen Namen können zur Unterscheidung von artmäßigen

Merkmale genutzt werden, entsprechen jedoch keiner Rangordnung. Nehmen die einzelnen Namen zu viel Platz ein, so können ihnen auch Abkürzungen oder Nummern als Schlüsselwerte zugewiesen werden.

Ordinalskala

Eine Ordinalskala hat die Klassenbezeichnungen eines quantitativen Merkmals als Skalenwerte. Im Gegensatz zu einer Nominalskala sind die einzelnen Klassenbezeichnungen nicht gleichwertig, sondern entsprechen einer Rangordnung in auf- oder absteigender Folge.

Intervall- und Verhältnisskala

Bei diesen beiden Skalentypen handelt es sich um metrische Skalen, vergleichbar mit einem Meterstab. Als Skalenwerte werden Vielfache einer Grundeinheit abgetragen.

Eine metrische Skala heisst Intervallskala, wenn der Nullpunkt willkürlich gewählt ist; in diesem Fall können zwar Differenzen zwischen zwei Werten sinnvoll interpretiert werden, Quotienten hingegen nicht; Beispielsweise entsprechen 20°C nicht einer doppelt so hohen Temperatur wie 10°C , wenn man vom absoluten Temperaturnullpunkt $T_0 = -273^{\circ}\text{C}$ ausgeht.

Ist der Nullpunkt einer Skala eindeutig festgelegt, so spricht man von einer Verhältnisskala. In diesem Fall sind auch Quotienten von einzelnen Werten sinnvoll interpretierbar. Beispiele hierfür sind Gewichtsangaben, Geldmengen, Stückzahlen, absolute Temperaturangaben usw.

Graphische Darstellungen

Bisweilen ist es praktisch, statistische Informationen als Diagramme graphisch darzustellen; diese müssen einerseits eindeutig beschriftet sein und sollten andererseits möglichst übersichtlich gestaltet werden.

- Bei einem Histogramm werden auf der waagrechten Achse die einzelnen Intervall- oder Klassengrenzen abgetragen. Über den einzelnen Intervallen werden Rechtecke gezeichnet, deren Höhe die absoluten oder relativen Häufigkeiten des jeweiligen Intervalls oder der jeweiligen Klasse darstellen.
- Todo: Liniendiagramm

Umgang mit ungenauen Messwerten

Als Messfehler werden Differenzen zwischen gemessenen Werten und den unbekannten wahren Werten der jeweiligen Messgrößen bezeichnet. Sie lassen sich grundsätzlich in zwei Arten unterteilen -- in systematische und statistische (zufällige) Fehler.

Systematische Fehler

Systematische Fehler entstehen durch mangelhafte Messverfahren, beispielsweise durch defekte Messgeräte, falsche Eichungen, oder Vernachlässigung von störenden Einflussgrößen. Je nach Fehler weichen die gemessenen Werte entweder nach oben oder nach unten von den tatsächlichen Werten ab.

Systematische Fehler werden "reproduzierbar" genannt, denn bei erneuten Messvorgängen treten sie unter gleichen Bedingungen erneut auf. Wird der Fehler gefunden, so kann er berücksichtigt und eventuell korrigiert werden.