

Ziel der deskriptiven Statistik

- Ziel der deskriptiven Statistik ist es Untersuchungsergebnisse darzustellen.
- Dazu ist es nicht zweckmäßig und oft auch nicht möglich alle Messwerte oder Beobachtungen im Einzelnen mitzuteilen.
- Es gilt vielmehr, die Ergebnisse sinnvoll zusammenzufassen, um das Wesentliche klar verständlich zum Ausdruck bringen bzw. schnell erfassen zu können.



Aufgaben der deskriptiven Statistik

- Übersichtliche Darstellung vorliegender Daten
- Charakterisierung von Daten mit Hilfe von Durchschnittswerten
- Bestimmung von Zusammenhängen von Merkmalen bzw. Unterschiede zwischen Gruppen



Formen statistischer Beschreibung

- Tabellen
- Graphische Darstellungen
- Kenngrößen (charakteristische Maßzahlen)



Tabellen: Strichliste

- Das Zählen von Merkmalsausprägungen ist bereits der deskriptiven Statistik zuzuordnen.
- Zählen heißt das Einsortieren der beobachtbaren Ausprägungen in bereits festgelegte Merkmalsklassen.
- Beispiel einer Strichliste: Befragung zum Familienstand (n=86)

		Häufigkeit (f)
ledig		32
verheiratet		37
verwitwet		5
geschieden		12



Tabelle: Häufigkeitstabelle

Ziel: Zusammenfassen der Strichliste in eine Häufigkeitstabelle:

Häufigkeitstabelle einer Befragung zum Familiensand (N=86)

	ledig	verheiratet	verwitwet	geschieden	Σ
Häufigkeiten (f)	32	37	5	12	86
Rel. Häufigkeiten (f/n)	32/86 =0.37	37/86 =0.43	5/86 =0.06	12/86 =0.14	1.00
Kum. Häufigkeiten (cum f)	32	69	74	86	
Kum. Prozent- häufigkeiten (cum f%)	37.0	80.0	86.0	100.0	



Begriffe der Häufigkeitstabelle

- **Häufigkeit:**

Unter einer Häufigkeit versteht man die Anzahl, mit der ein bestimmtes Ergebnis bzw. Ereignis eingetreten ist. Bei der Angabe dieser Häufigkeit unterscheidet man verschiedene Formen:

- **Absolute Häufigkeit:**

Sie gibt an, wie viele Merkmalsträger zu einer bestimmten Merkmalsausprägung in einem Datensatz ab. Sie hängt somit vom Umfang des betrachteten Datensatzes ab. Für den Vergleich unterschiedlicher Datensätze wird ein normiertes Maß, die relative Häufigkeit verwendet.

- **Relative Häufigkeit:**

Die relative Häufigkeit f/n ist die absolute Häufigkeit f dividiert durch die Anzahl aller Beobachtungswerte



Begriffe der Häufigkeitstabelle

- **Kumulierte Häufigkeit:**

Die kumulierte Häufigkeit cum f ist die sukzessive summierte absolute Häufigkeit f.

- **Prozentwert:**

Ein Prozentwert ist definiert als die Anzahl der Beobachtungen, die einen bestimmten Wert annehmen, dividiert durch die Anzahl aller Beobachtungen, multipliziert mit 100. Kurz: die relative Häufigkeit $f/n \times 100$.

- **Prozentränge:**

Prozentränge cum $f\%$ sind kumulierte (aufsummierte) Prozentwerte.



Graphiken: Graphische Darstellungsformen

- **Histogramm:**

Das Histogramm ist ein spezielles Säulendiagramm, wobei die Höhe der Säule den Messwert des jeweiligen Merkmals repräsentiert.

- **Kreisdiagramm:**

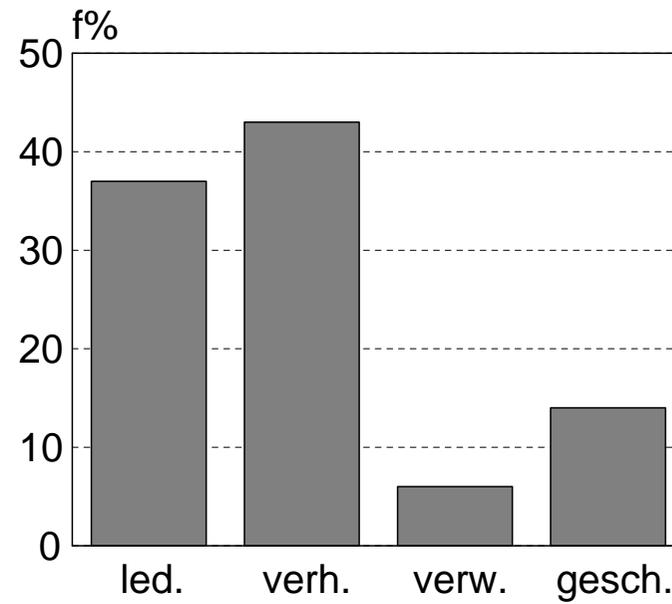
Das Kreisdiagramm eignet sich vor allem zur Darstellung von relativen Zahlenverhältnissen, wobei die Kreissegmente den jeweiligen Messwert des Merkmals repräsentieren.

- **Polygon:**

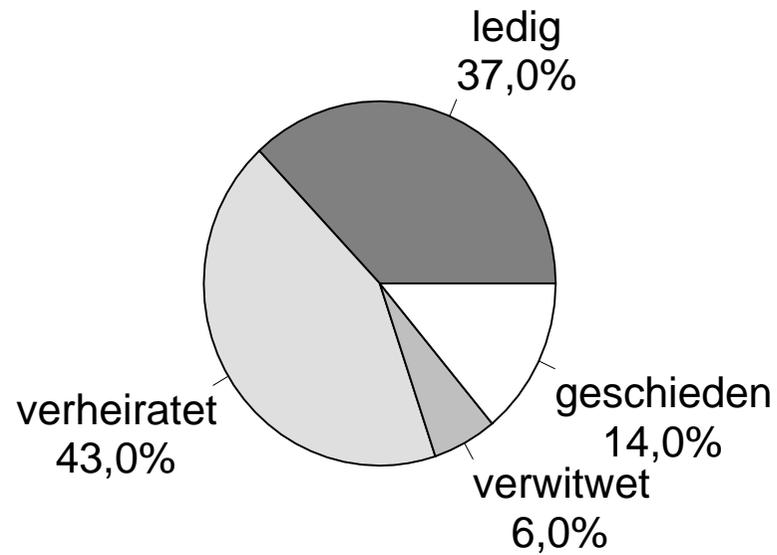
Ein Polygon dient der Darstellung einer Messgröße in ihrer Abhängigkeit von einem Faktor, wobei die Lage der einzelnen Punkte den Messwert des jeweiligen Merkmals bei verschiedenen Bedingungen repräsentieren.



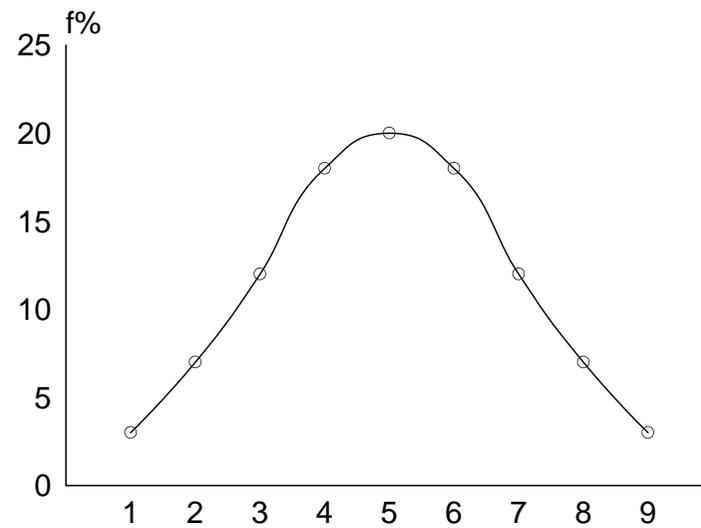
Graphiken: Histogramm



Graphiken: Kreisdiagramm



Graphiken: Polygon

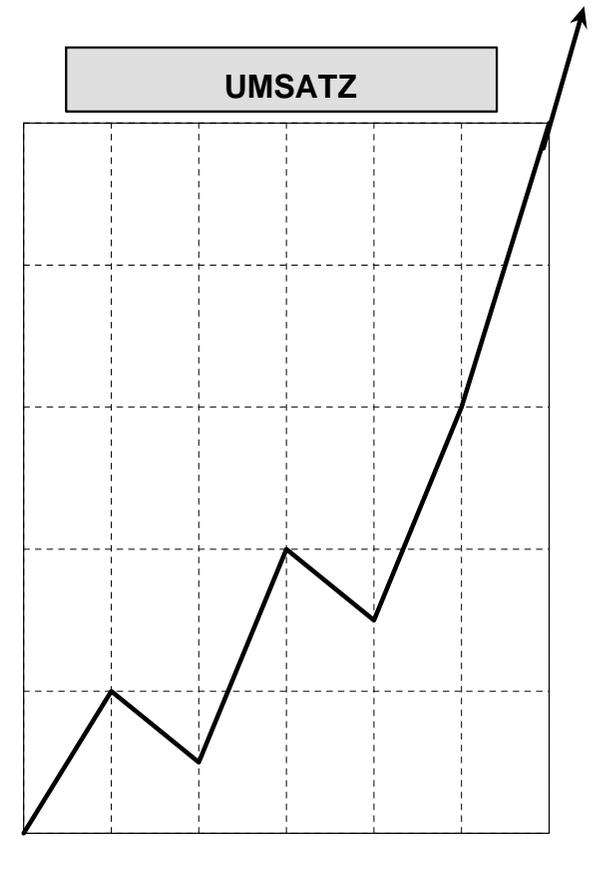
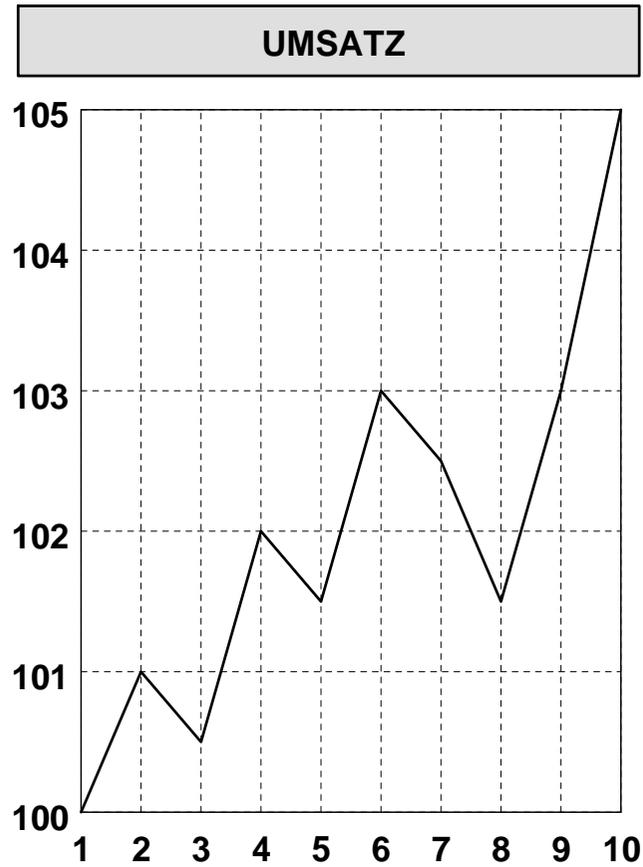
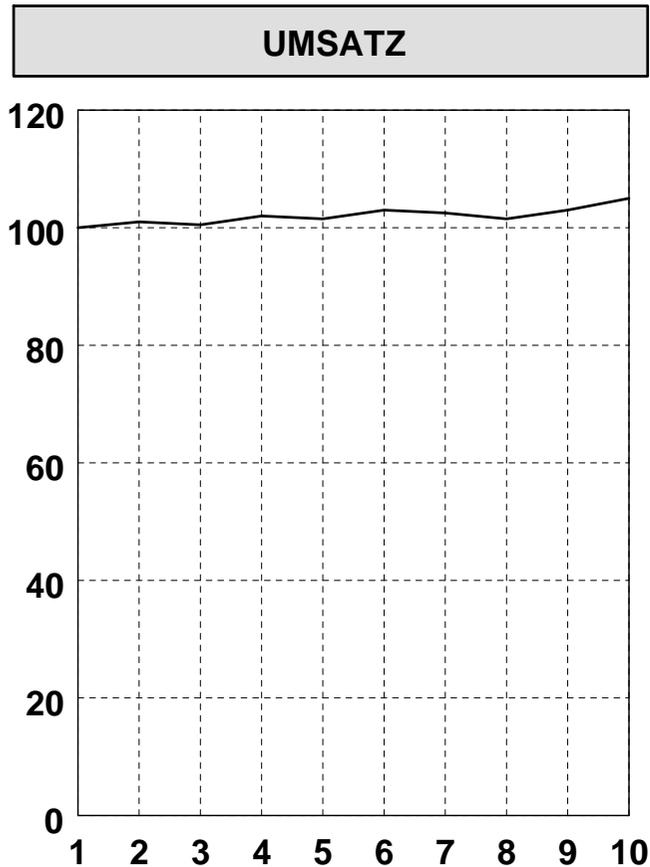


Verfälschungen bei Graphiken

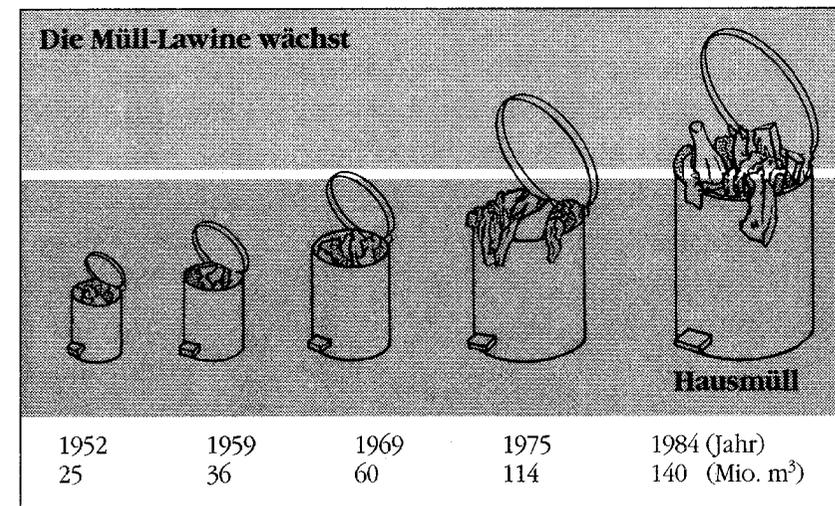
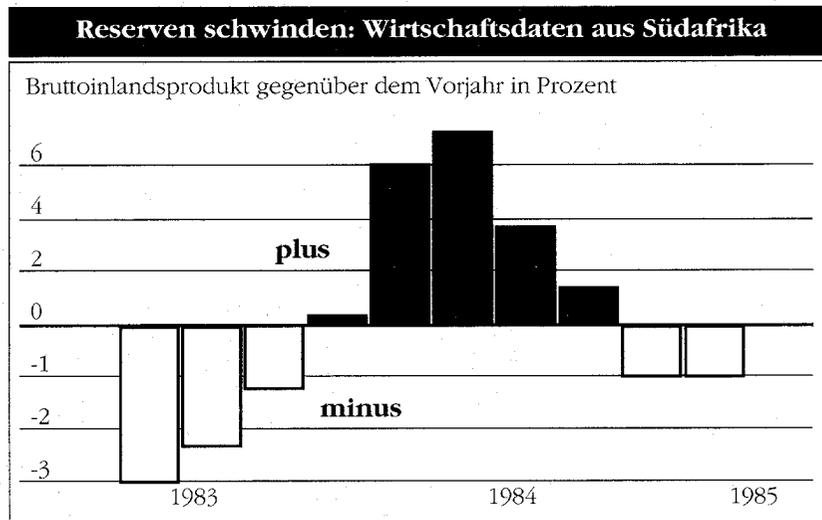
- Daten werden aus Gründen der Übersicht zusammengefasst dargestellt.
- Ziel dieser Art der Visualisierung ist die Fokussierung bestimmter Aspekte, die sich aus den Daten ergeben.
- Dabei kann es zu beabsichtigten oder unbeabsichtigten Verzerrungen kommen.



Verfälschungen zwischen Graphiken



Verfälschungen innerhalb einer Graphik



Kenngrößen: Typen von Kenngrößen

- **Maße der zentralen Tendenz** (Lokationsmaße):
Beschreibung der Verteilung bzgl. ihres Schwerpunktes
(Mittelwert, Modus, Median)
- **Streuungsmaße** (Dispersionsmaße):
Aussagen über die Streuung von Häufigkeitsverteilungen
(Standardabweichung, Varianz, Range, Quartilmaße)
- **Formmaße:**
Aussagen über die Anpassung der Verteilung an die Normalverteilung
(Schiefe, Exzess)



Überblick über die wichtigsten Kenngrößen in Abhängigkeit vom Skalenniveau

Kenngrößen	Nominalskalen	Ordinalskalen	Intervallskalen
Maße der zentralen Tendenz	Modus	Modus Median	Modus Median Mittelwert
Streuungsmaße	Keine Berechnung möglich	Prozentränge Dezildifferenz Quartilabstand	Prozentränge Dezildifferenz Quartilabstand Range Varianz Standardabweichung
Formmaße	Keine Berechnung möglich	Keine Berechnung möglich	Schiefe Exzess



Modus/Modalwert

- **Definition:**
Der Modus ist die Kategorie mit der größten Häufigkeit.
- **Information:**
Kategorie der Messwertklasse mit dem häufigsten Vorkommen.
- **Keine Information:**
 - die Häufigkeit selbst
 - die Größe des Unterschieds zu den Häufigkeiten der restlichen Messwertklassen
- **Anwendung:**
Nominal-, ordinal-, und intervallskalierten Daten



Beispiel Modus

- Befragung zum Familienstand (N=86):

	ledig	verheiratet	verwitwet	geschieden	Σ
Häufigkeiten (f)	32	37	5	12	86
Rel. Häufigkeiten (f/n)	32/86 =0.37	37/86 =0.43	5/86 =0.06	12/86 =0.14	1.00
Kum. Häufigkeiten (cum f)	32	69	74	86	
Kum. Prozent- häufigkeiten (cum f%)	37.0	80.0	86.0	100.0	

Der Modus ist in diesem Beispiel die Kategorie „verheiratet“



Median

- **Definition:**

Der Median ist derjenige Wert einer Häufigkeitsverteilung, der die Messwertreihe halbiert, so dass oberhalb und unterhalb dieses Messwertes 50% der Messwerte liegen. Er halbiert also die der Größe nach geordnete Reihe.

- **Information:**

Die Größe des Messwertes, der die Gesamtheit aller Messwerte in zwei gleich große Hälften teilt. Oberhalb und unterhalb liegen dann gleich viele Werte, unabhängig davon wie groß diese sind.

- **Keine Information:**

Wertebereich, den die Messwerte in den jeweiligen Hälften abdecken.

- **Anwendung:**

Ordinal- und intervallskalierten Daten



Formel und Beispiel Median (N=gerade)

- Median, wenn die Anzahl der Messwerte ungerade ist:

$$Md = x_{(n+1)/2}$$

- Beispiel:

Eine Schülergruppe (N=11) wurde hinsichtlich ihrer Sportmotivation in eine geordnete Reihenfolge gebracht:

17, 21, 22, 24, 25, 28, 31, 43, 35, 36, 39

$$Md = x_{(n+1)/2} = x_{(11+1)/2} = x_{(6)} = 28$$



Formel und Beispiel Median (N=ungerade)

- Median, wenn die Anzahl der Messwerte gerade ist:

$$Md = [x_{(n/2)} + x_{((n/2)+1)}] / 2$$

- Beispiel:

Eine Studentengruppe (N=6) wurde hinsichtlich ihrer Konzentrationsfähigkeit in eine geordnete Reihenfolge gebracht:

3, 7, 11, 13, 15, 17

$$Md = [x_{(n/2)} + x_{((n/2)+1)}] / 2 = [x_{(6/2)} + x_{((6/2)+1)}] / 2 = [x_{(3)} + x_{(4)}] / 2 = 24 / 2 = 12$$



Arithmetisches Mittel/Mittelwert

- **Definition:**
Das arithmetische Mittel ist definiert als die Summe aller Messwerte dividiert durch deren Anzahl
- **Information:**
Die Größe der durchschnittlichen Messwertes.
- **Keine Information:**
 - über den Wertebereich, den die Messwerte abdecken und damit
 - wie sehr der Mittelwert durch extreme Messwerte beeinflusst wird
- **Anwendung:**
Intervallskalierte Daten



Formel und Beispiel arithmetisches Mittel

- Formel:

$$\bar{x} = \frac{1}{N} \sum x_i = x_1 + x_2 + \dots + x_n$$

- Beispiel:

Messwerte (N=9): 5, 7, 12, 14, 17, 21, 26, 29, 31

$$\bar{x} = \frac{1}{9} \sum (5 + 7 + 12 + 14 + 17 + 21 + 26 + 29 + 31) = \frac{162}{9} = 18$$



Verwendung der Maße für die zentrale Tendenz

- **Modus:**
 - wenn lediglich ein grober Überblick über eine Verteilung gefragt ist
 - wenn „durchschnittlich im Sinne von „am häufigsten“ benutzt wird
- **Median:**
 - wenn mindestens ordinalskalierte Variablen vorliegen
 - wenn man den exakten Mittelpunkt einer Verteilung kennen möchte
 - wenn Extremwerte das arithmetische Mittel stark verzerren würden
 - wenn eine deutliche Abweichung von der Normalverteilung vorliegt
- **Arithmetisches Mittel:**
 - wenn mindestens eine Intervallskalierung vorliegt
 - wenn die Verteilung in etwa symmetrisch ist

