

Graphische Darstellung einer univariaten Verteilung:

Die graphische Darstellung einer univariaten Verteilung hängt von dem Messniveau der Variablen ab. Bei einer graphischen Darstellung wird die Häufigkeit jeder Kategorie/Merkmalausprägung einer Variablen proportional abgebildet.

Nominales Messniveau:

1. Streifendiagramm (Balken- bzw. Säulendiagramm)
 2. Kreisdiagramm
- Die Abstände und Reihenfolge der Säulen sind beliebig wählbar.

Ordinales Messniveau:

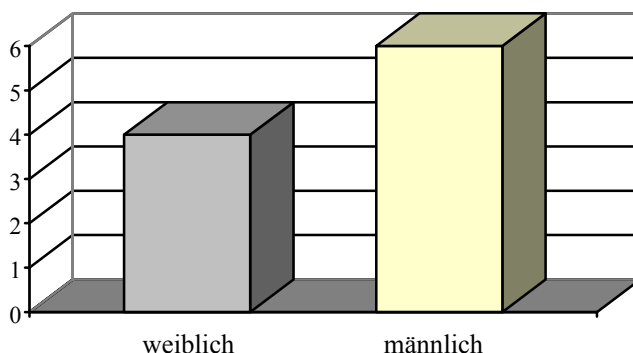
1. Streifendiagramm (Balken- bzw. Säulendiagramm)
 2. Kreisdiagramm
- Bei ordinalen Daten ist die Reihenfolge zu beachten. Die Anordnung der Säulen folgt der Reihenfolge der rangmäßig geordneten Kategorien.

Metrisches Messniveau:

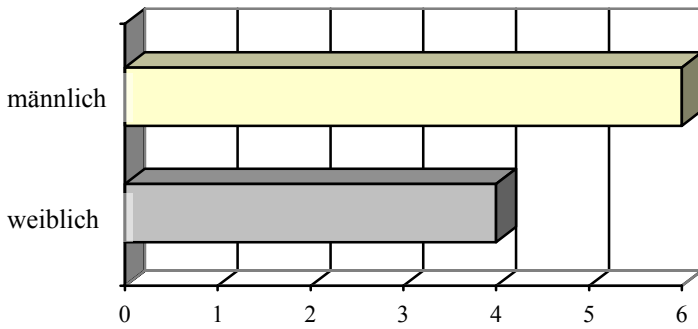
1. Histogramm
2. Polygon (auch Häufigkeitspolygon bzw. Polygonzug genannt)

Beispiel:

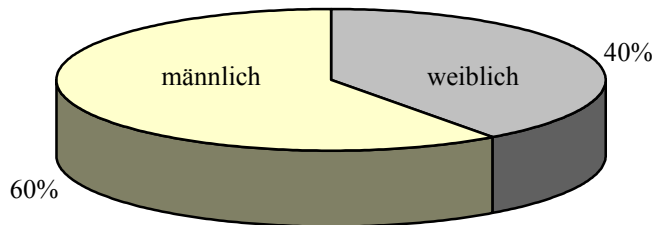
1. Darstellungsmöglichkeit eines Säulendiagramms der Variable „Geschlecht“:



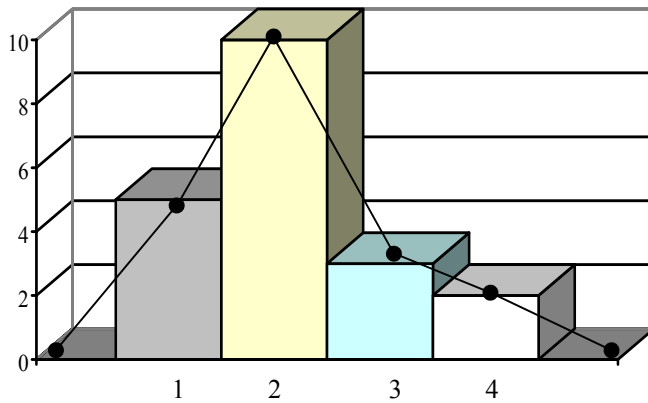
2. Darstellungsmöglichkeit eines Säulendiagramms der Variable „Geschlecht“:



Kreisdiagramm der Variable „Geschlecht“:



Histogramm und Polygonzug der Variable „Anzahl bisheriger Arbeitgeber“:



- Bei einem Histogramm wird jedes Rechteck so gezeichnet, dass die eine Seite die exakte untere Grenze des Intervalls und die andere Seite die exakte obere Grenze des Intervalls markiert. Die Intervalle besitzen die gleiche Breite. Die Häufigkeiten werden durch die Höhe der Rechtecke zum Ausdruck gebracht.

Maßzahlen der zentralen Tendenz:

- 1) Modus (h) \Rightarrow ab nominalem Messniveau
- 2) Median (\tilde{x}) \Rightarrow ab ordinalem Messniveau
- 3) Arithmetische Mittel (\bar{x}) \Rightarrow ab metrischem Messniveau

1) Modus (h):

- ist der dichteste Wert einer Verteilung
- ist der am häufigsten vorkommende Wert einer Verteilung
- ist der Wert, der eine größere Häufigkeit als andere Werte in seiner Nachbarschaft aufweist

Eine Verteilung kann **unimodal** (einen Modalwert besitzen), **bimodal** (zwei Modalwerte besitzen), **trimodal** (drei Modalwerte besitzen) und **multimodal** (mehr als drei Modalwerte besitzen) sein.

Beispiel für Unimodalität:

x_i	f_i
1	1
2	2
3	3
4	2
5	1
6	1

Modus = $h = 3$, da dieser Messwert $x_3 = 3$ eine größere Häufigkeit (nämlich $f_3 = 3$) aufweist als alle anderen Werte in seiner Nachbarschaft.

Wichtig: Der Modus entspricht immer dem Messwert (x_i), nicht dem f_i - Wert.

Beispiel Spezialfall der Unimodalität - gleiche Häufigkeiten nebeneinander liegender Werte:

x_i	f_i
1	1
2	2
3	3
4	3
5	1
6	1

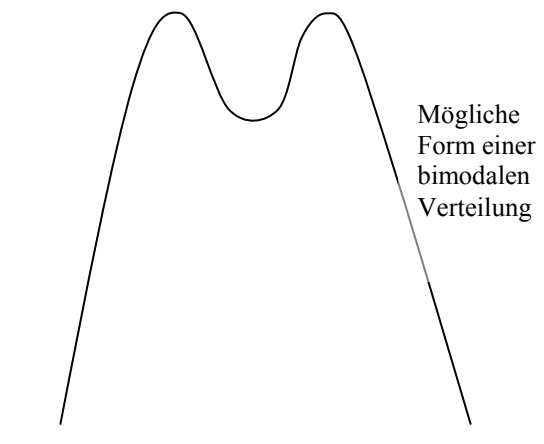
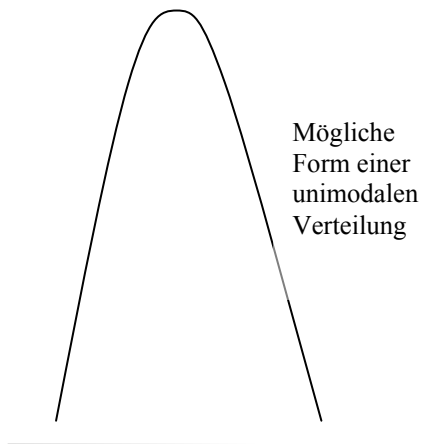
$$h = \frac{(x_3 + x_4)}{2}, \text{ also : } h = \frac{(3+4)}{2} = \frac{7}{2} = 3,5$$

Beispiel für Bimodalität:

x_i	f_i	f_i
1	1	1
2	2	2
3	3	9
4	2	2
5	1	1
6	3	8

$$h_1 = 3 \text{ und } h_2 = 6$$

Mögliche Verteilungsformen:



2) Median (\tilde{x}):

- ist der Wert, der eine nach ihrer Größe geordnete Reihe von Messwerten halbiert

Mathematische Errechnung des Medians:**1) ungerade Anzahl von Fällen (N):**

$$\text{Median} = \frac{(N+1)}{2} = \text{der (Mess-) Wert des Falles}$$

Beispiel:

x_i	f_i
1	1
2	1
3	1
4	1
5	1
6	1
7	1
Σ	$N = 7$

$$\text{Median} = \frac{(7+1)}{2} = 4 \quad (\text{ist der Messwert des } \mathbf{4. \text{ Falles}}, \text{ also } \mathbf{\text{Messwert } 4})$$

- Es liegen ebenso viele Fälle unterhalb wie oberhalb des 4. Falles.

2) gerade Anzahl von Fällen (N):

- Median ist der halbierte Wert der mittleren beiden Fälle (fiktiver Wert)

$$\frac{N}{2} = \text{Fall } a \quad \frac{N}{2} + 1 = \text{Fall } b$$

$$\text{Median} = \frac{(\text{Fall } a + \text{Fall } b)}{2}$$

Beispiel:

x_i	f_i
1	1
2	1
3	1
4	1
-----	-----
5	1
6	1
7	1
8	1
Σ	$N = 8$

$$\frac{8}{2} = 4. \text{ Fall} \quad \frac{8}{2} + 1 = 5. \text{ Fall}$$

$$\text{Median} = \frac{(4+5)}{2} = 4,5 \quad (\text{ist der halbierte Messwert des } \mathbf{4. \text{ und } 5. \text{ Falles, also Messwert } 4,5})$$

Allgemein:

- Er ist nützlich für Verteilungen mit sehr extremen Werten, d.h. für sehr stark asymmetrische Verteilungen (z.B. einer j - förmigen Verteilung).
- Der Median kann für Verteilungen mit offenen Enden (unvollständigen Verteilungen) mit Hilfe des **Medianintervalls** berechnet werden.

Medianintervall (Mediankategorie, Eingriffsspielraum):

- Das Medianintervall wird für **Verteilungen mit offenen Enden** angewendet, aber auch, wenn die Messwerte einer Verteilung als **gruppierte Daten** vorliegen.

Vorgehensweise:

- 1) Man bildet die kumulierten Häufigkeiten.
 - Man bestimmt mit Hilfe der kumulierten Häufigkeitsverteilung das Intervall, in das der Median fällt (Medianintervall).
- 2) Man halbiert zunächst die Anzahl der Fälle: $\frac{N}{2}$. Damit ist die Anzahl der Fälle bekannt, die unterhalb und oberhalb des Medians liegen.
- 3) Alsdann ermittelt man das Intervall, in das der Median fällt. Das ist das Intervall mit einer kumulierten Häufigkeit **gleich** oder (**nächst**) **größer** als $\frac{N}{2}$.
 - Wenn die kumulierte Häufigkeit gleich $\frac{1}{2} N$ ist, dann stellt die exakte obere Grenze des Intervalls den Wert des Medians dar.
- 4) Man formuliert die exakten Grenzen des (Median-) Intervalls.
- 5) Man berechnet den Median nun nach der Formel:

$$\text{Median} = U + \left[\frac{\frac{1}{2} \cdot N - F_u}{F_m} \right] \cdot h$$

U = exakte untere Grenze des Medianintervalls

N = Anzahl der Fälle

F_u = kumulierte Häufigkeit unterhalb des Medianintervalls

F_m = Häufigkeit im Medianintervall

h bzw. K_b = Breite des Medianintervalls

3) Arithmetische Mittel (\bar{x}):

- wird auch Durchschnittswert genannt
- wird vorzugsweise für symmetrische, unimodale und lineare Verteilungen berechnet, die keine offenen Enden aufweisen.
- ist definiert, als die Summe der Messwerte (multipliziert mit den Häufigkeiten), geteilt durch ihre Anzahl:

$$\bar{x} = \frac{\sum f_i \cdot x_i}{N} = \frac{f_1 \cdot x_1 + f_2 \cdot x_2 + f_3 \cdot x_3 + \dots + f_k \cdot x_k}{N}$$

Beispiel:

x_i	f_i	$f_i \cdot x_i$
6	1	6
7	2	14
8	3	24
9	2	18
Σ	$N = 8$	62

$$\bar{x} = \frac{62}{8} = 7,75$$

Drei wichtige Eigenschaften des arithmetischen Mittels sind zu beachten:

- 1) **Die Summe der Abweichungen aller Messwerte von ihrem arithmetischen Mittel ist gleich Null:**

$$\sum (x_i - \bar{x}) = 0$$

Beispiel:

x_i	f_i	$f_i x_i$	$x_i - \bar{x}$
3	1	3	3 - 4 = -1
4	1	4	4 - 4 = 0
5	1	5	5 - 4 = 1
Σ	N = 3	12	0 = Summe der Abweichungen

$$\bar{x} = \frac{12}{3} = 4$$

- 2) **Die Summe der quadrierten Abweichungen von ihrem arithmetischen Mittel ist kleiner als die Summe der quadrierten Abweichungen aller Messwerte von einem beliebigen anderen Wert (x_0):**

$$\sum (x_i - \bar{x})^2 = \min \quad \text{bzw.} \quad \sum (x_i - \bar{x})^2 < \sum (x_i - x_0)^2$$

Beispiel:

- $x_0 = 3$ (beliebiger Messwert der Verteilung)

x_i	f_i	$f_i x_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - x_0)^2$
3	1	3	3 - 4 = -1	$-1^2 = 1$	$(3 - 3)^2 = 0^2 = 0$
4	1	4	4 - 4 = 0	$0^2 = 0$	$(4 - 3)^2 = 1^2 = 1$
5	1	5	5 - 4 = 1	$1^2 = 1$	$(5 - 3)^2 = 2^2 = 4$
Σ	N = 3	12	0	2	5

$$\bar{x} = \frac{12}{3} = 4$$

3) Die Addition (Subtraktion) einer bestimmten Zahl (C) zu allen Einzelwerten einer Verteilung vergrößert das arithmetische Mittel um diese Zahl:

$$(x_i + C) = \bar{x} + C \quad \text{bzw.} \quad (x_i - C) = \bar{x} - C$$

Beispiel:

- $C = 2$

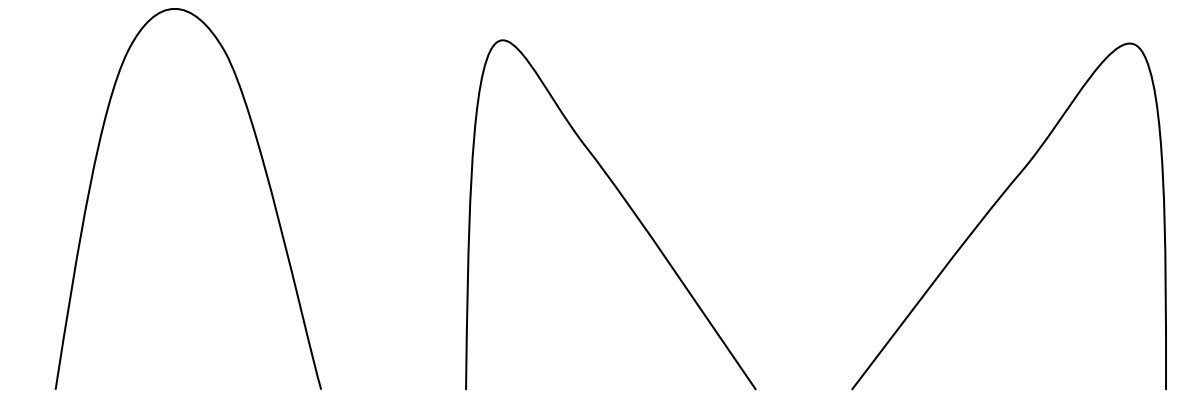
x_i	f_i	$f_i x_i$	$f_i \cdot (x_i + C)$
3	1	3	$3 + 2 = 5$
4	1	4	$4 + 2 = 6$
5	1	5	$5 + 2 = 7$
Σ	$N = 3$	12	18

$$\bar{x}_{alt} = \frac{12}{3} = 4$$

$$\bar{x}_{neu} = \frac{18}{3} = 6, \text{ also } \bar{x}_{alt} + C = 4 + 2$$

Typische Verteilungsformen:

(die Lage der Maßzahlen der zentralen Tendenz h, \tilde{x}, \bar{x} in verschiedenen Verteilungen)



$$h = \tilde{x} = \bar{x}$$

- unimodal
- **symmetrisch**

$$h < \tilde{x} < \bar{x}$$

- unimodal
- **linkssteil bzw. rechts-schief**

$$h > \tilde{x} > \bar{x}$$

- unimodal
- **rechtssteil bzw. linksschief**

Zusätzliche Verteilungsformen:

- schmalgipflig (steil)
- breitgipflig (flach)
- u - förmig
- j - förmig

Beispiel für 2 Verteilungen mit **gleichen Maßzahlen**, aber **ungleicher Streuung**:

