

Univariate deskriptive Statistik

1 Deskriptive vs. Inferenzstatistik

2 Formale Grundlagen

3 Darstellung von Verteilungen

(Häufigkeiten, kumulierte Häufigkeiten, primäre vs. sekundäre Häufigkeitsverteilung, Histogramm, Kuchendiagramm, Polygonzug)

4 Maße der zentralen Tendenz

(Arithmetischer Mittelwert, Modalwert, Median)

Literatur

-  Bortz, J. & Schuster, Ch. (2010). *Statistik für Human- und Sozialwissenschaftler* (7. Auflage). Berlin: Springer. [Kap. 2 & 3]
-  Eid, M., Gollwitzer, M. & Schmitt, M. (2010). *Statistik und Forschungsmethoden*. Weinheim: Beltz. [Kap. 6]

Univariate deskriptive Statistik

- In der **deskriptiven Statistik** geht es darum, die Daten, die man an einer Stichprobe von Personen (oder anderen Untersuchungseinheiten) erhoben hat, zu beschreiben (durch Tabellen, Grafiken und statistische Kennwerte).
- In der Regel sind wir aber daran interessiert, Aussagen zu machen, die über die Beschreibung der Daten der untersuchten Stichprobe hinausgehen. Wir wollen vielmehr allgemeine Aussagen über die **Grundgesamtheit (Population)** aller Personen machen, aus der wir die Stichprobe gezogen haben (z.B. alle Studierenden, alle 3-jährigen Kinder, alle Menschen)
- Die **Inferenzstatistik** (oder schließende Statistik) beschäftigt sich u.a. damit, wie man solche Schlüsse von der Stichprobe auf die Grundgesamtheit vornehmen kann und wie man Hypothesen (über Populationen) an den Daten testen kann.
- Wir werden uns zunächst mit der deskriptiven Statistik beschäftigen, und dort zunächst mit der **univariaten** Deskription, also der Beschreibung einzelner Variablen und dann mit der Beschreibung der Beziehung zwischen zwei Variablen (**bivariat**).

Festlegungen

➤ Notation:

- Wir bezeichnen **Variablen** mit großen, kursiven Buchstaben, z.B. X, Y
- und ihre **Ausprägungen** mit kleinen, kursiven Buchstaben, wobei Subskripte angeben, von welchem Objekt die Ausprägung stammt: z.B. x_i, y_j

➤ Schreibweise mit **Summenzeichen**:

- Man schreibt verkürzt auch $\sum_{i=1}^n x_i = x_1 + x_2 + \dots + x_n$

mit x_i als i -ter Summand und i als Laufindex, der von 1 bis n läuft ($i = 1 \dots n$).

Man sagt: „Summe über alle x - i -Werte für i gleich 1 bis n “

- Einige einfache Rechenregeln:

$$\sum_{i=1}^n (x_i + a) = (x_1 + a) + (x_2 + a) + \dots + (x_n + a) = \sum_{i=1}^n x_i + n \cdot a = n \cdot a + \sum_{i=1}^n x_i$$

$$\sum_{i=1}^n b \cdot x_i = b \cdot x_1 + b \cdot x_2 + \dots + b \cdot x_n = b \cdot \sum_{i=1}^n x_i$$

$$\sum_{i=1}^n (x_i + y_i) = \sum_{i=1}^n x_i + \sum_{i=1}^n y_i$$

Univariate deskriptive Statistik: Häufigkeiten

- **Datenbeispiel 1:** 40 Schüler einer Schulklasse haben die folgenden Noten in einer Mathematikarbeit geschrieben:

1	4	1	3	4	5	3	1	3	2
4	3	4	2	3	4	5	3	4	3
5	3	3	2	4	2	5	4	3	4
3	2	1	4	2	3	3	2	5	4

Urliste: Liste der Daten in der (beliebigen) Reihenfolge, in der sie angefallen sind.

- Zur Beschreibung der Daten kann man aus der Urliste eine (diskrete) **Häufigkeitsverteilung** erstellen, indem man die verschiedenen Messwerte mit ihrer Auftretenshäufigkeit auflistet; diese werden auch als (**absolute**) **Häufigkeiten** h_k mit $k = 1 \dots 5$ bezeichnet (z.B. $h_2 = 7$):

Note	Strichliste	h_k
1		4
2	 	7
3	 	13
4	 	11
5	 	5

Univariate deskriptive Statistik: Häufigkeiten

- Aus den absoluten Häufigkeiten h_k können die **relativen Häufigkeiten** f_k bestimmt werden, in dem die Häufigkeiten aller $k = 1 \dots c$ **Kategorien (Klassen)** an der Gesamtzahl von Personen (Beobachtungen) n relativiert werden:

$$f_k = \frac{h_k}{n}$$

- Statt der relativen Häufigkeiten, die immer Werte zwischen 0 und 1 annehmen, werden häufig die **prozentualen Häufigkeiten** p_k angegeben, die man erhält, wenn man die relativen Häufigkeiten mit 100 multipliziert:

$$p_k = \frac{h_k}{n} \cdot 100 = f_k \cdot 100$$

- Bsp. für Ergebnisse bei **Datenbeispiel 1**:
- $h_2 = 7$ Schüler haben als Note eine „2“.
 - $p_4 = 27.5\%$ aller Schüler haben mit der Note „4“ abgeschlossen.

Note	h_k	f_k	p_k
1	4	0.100	10.0
2	7	0.175	17.5
3	13	0.325	32.5
4	11	0.275	27.5
5	5	0.125	12.5
Summe:	40	1	100

} $k = 1 \dots c$
 $c = 5$

↖ n

Univariate deskriptive Statistik: Häufigkeiten

- Die absoluten, relativen oder prozentualen Häufigkeiten lassen sich bei Vorliegen von (mindestens) Ordinalskalenniveau sinnvollerweise aufsteigend kumulieren. Auf diese Weise kann man die folgenden **kumulierten** oder **Summenhäufigkeiten** bestimmen:

$$hc_k = \sum_{q=1}^k h_q$$

$$fc_k = \sum_{q=1}^k f_q$$

$$pc_k = \sum_{q=1}^k p_q$$

Note	h_k	f_k	p_k	hc_k	fc_k	pc_k
1	4	0.100	10.0	4	0.100	10.0
2	7	0.175	17.5	11	0.275	27.5
3	13	0.325	32.5	24	0.600	60.0
4	11	0.275	27.5	35	0.875	87.5
5	5	0.125	12.5	40	1.000	100.0

- Bsp. für Ergebnisse bei **Datenbeispiel 1**:

- $hc_2 = 11$ Schüler haben eine Note „2“ oder besser erhalten.
- $pc_3 = 60\%$ der Schüler haben mindestens eine „3“ bekommen und $100 - pc_3 = 40\%$ der Schüler haben schlechtere Noten erhalten als eine „3“.

Univariate deskriptive Statistik: Häufigkeiten

- **Datenbeispiel 2** aus Bortz und Schuster (2010, S. 25): 90 Patienten mit hirnorganischen Schäden wurden hinsichtlich der Fähigkeit, aus einzelnen Teilstücken eine vorgegebene Figur zusammzusetzen, untersucht. Die interessierende Variable ist die Bearbeitungszeit in Sekunden.

131.8	106.7	116.4	84.3	118.5	93.4	65.3	113.8	140.3
119.2	129.9	75.7	105.4	123.4	64.9	80.7	124.2	110.9
86.7	112.7	96.7	110.2	135.2	134.7	146.5	144.8	113.4
128.6	142.0	106.0	98.0	148.2	106.2	122.7	70.0	73.9
78.8	103.4	112.9	126.6	119.9	62.6	116.6	84.6	101.0
68.1	95.9	119.7	122.0	127.3	109.3	95.1	103.1	92.4
103.0	90.2	136.1	109.6	99.2	76.1	93.9	81.5	100.4
114.3	125.5	121.0	137.0	107.7	69.0	79.0	111.7	98.8
124.3	84.9	108.1	128.5	87.9	102.4	103.7	131.7	139.4
108.0	109.4	97.8	112.2	75.6	143.1	72.4	120.6	95.2

- Ist für solche Daten ebenfalls eine Bestimmung der Häufigkeiten sinnvoll?

Univariate deskriptive Statistik: Häufigkeiten

- Weisen die Variablen sehr viele Kategorien auf oder handelt es sich um eine kontinuierliche Variable, so ist es sinnvoll, die Häufigkeiten für **Bereiche** oder **Intervalle** von Messwerten anzugeben, auch bezeichnet als **sekundäre Häufigkeitsverteilung** (andernfalls werden die Tabellen sehr lang mit vielen kleinen Häufigkeiten).
- Wie breit sollten die Intervalle gewählt werden (bzw. wie viele Kategorien sollten sich durch die Zusammenfassung ergeben)? Grundsätzlich sollten die Intervalle gleich breit sein und breiter sein, wenn weniger Daten (Personen) vorliegen und der Wertebereich der Variablen größer ist.
- Bsp. für Ergebnisse bei **Datenbeispiel 2** (mit Wertebereich von 62.2 bis 148.2) in Abhängigkeit der Intervallbreiten:

Intervall	h_k
60.0 - 89.9	20
90.0 - 119.9	44
120.0 - 149.9	26

Intervallbreite 30, 3 Kat.

Intervall	h_k
60.0 - 69.9	5
70.0 - 79.9	8
80.0 - 89.9	7
90.0 - 99.9	12
100.0 - 109.9	17
110.0 - 119.9	15
120.0 - 129.9	13
130.0 - 139.9	7
140.0 - 149.9	6

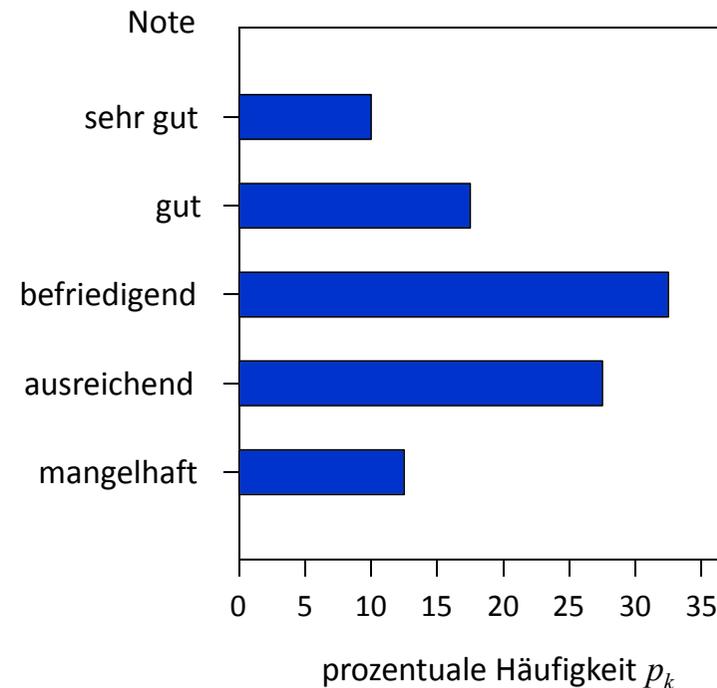
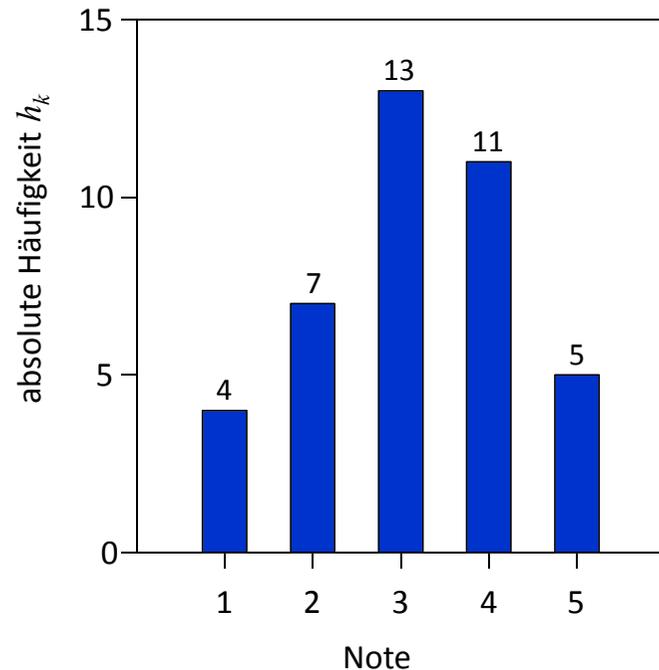
Intervallbreite 10, 9 Kat.

Intervall	h_k
60.0 - 61.9	0
62.0 - 63.9	1
64.0 - 65.9	2
66.0 - 67.9	0
...	...
142.0 - 143.9	2
144.0 - 145.9	1
146.0 - 147.9	1
148.0 - 149.9	1

Intervallbreite 2, 45 Kat.

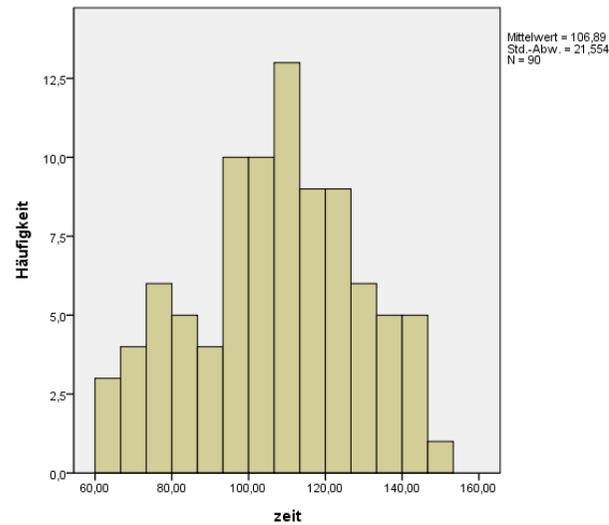
Univariate deskriptive Statistik: Grafische Darstellung von Häufigkeiten

- Die tabellierten Häufigkeiten können zur besseren Veranschaulichung auch grafisch dargestellt werden.
- Im **Histogramm (Säulendiagramm)** werden die Häufigkeiten von Kategorien (Klassen) durch die dazu proportionale Höhe von Säulen dargestellt (i.d.R. mit konstanter Breite der Säulen). Wird die Grafik um 90 Grad gedreht, spricht man von einem **Balkendiagramm**. Ggf. müssen bei vielen Ausprägungen wieder Bereiche zusammengefasst werden.

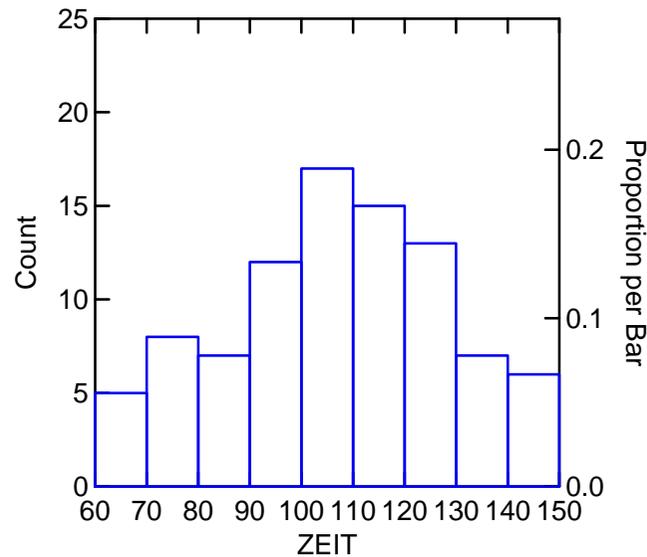


Univariate deskriptive Statistik: Grafische Darstellung von Häufigkeiten

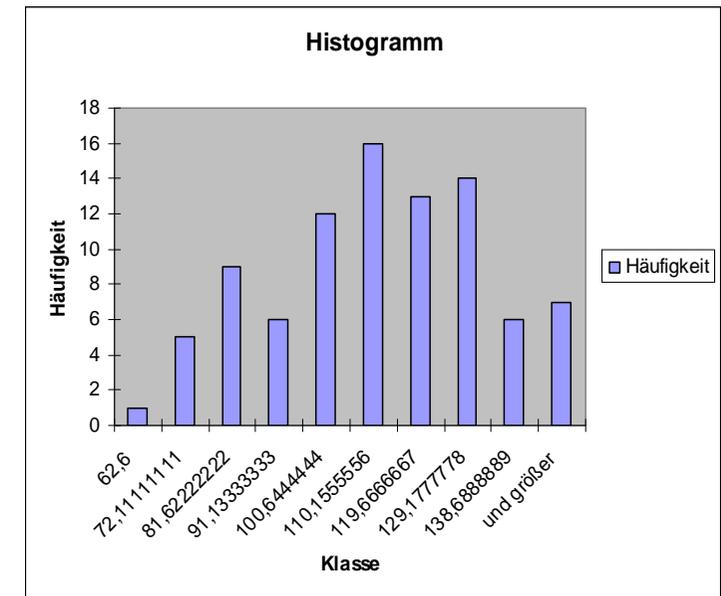
- Standard-Ausgaben der Histogramme für den **Beispielsdatensatz 2** (Reaktionszeiten):



SPSS



Systat

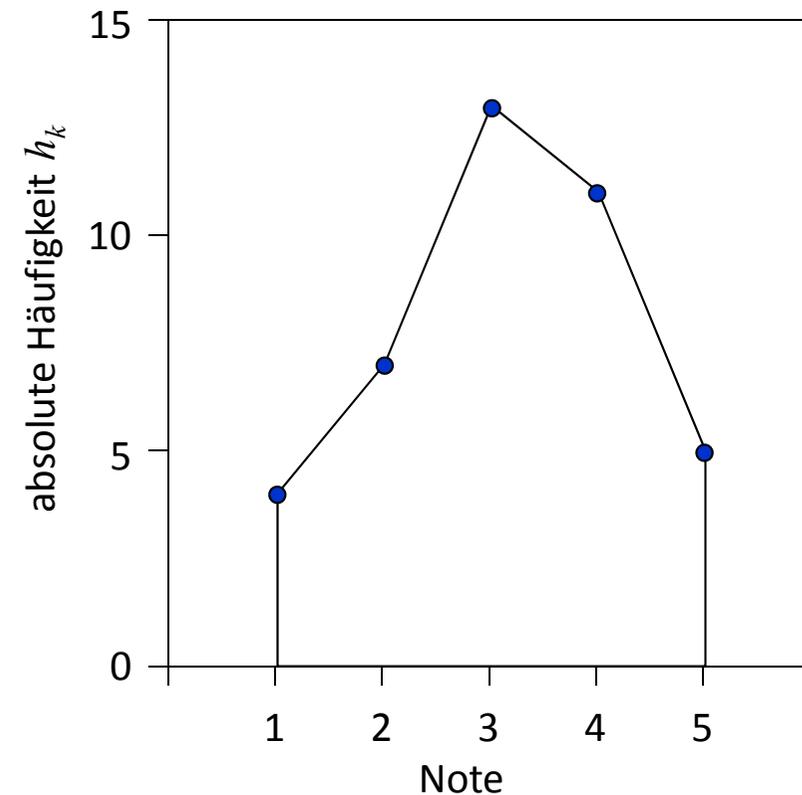
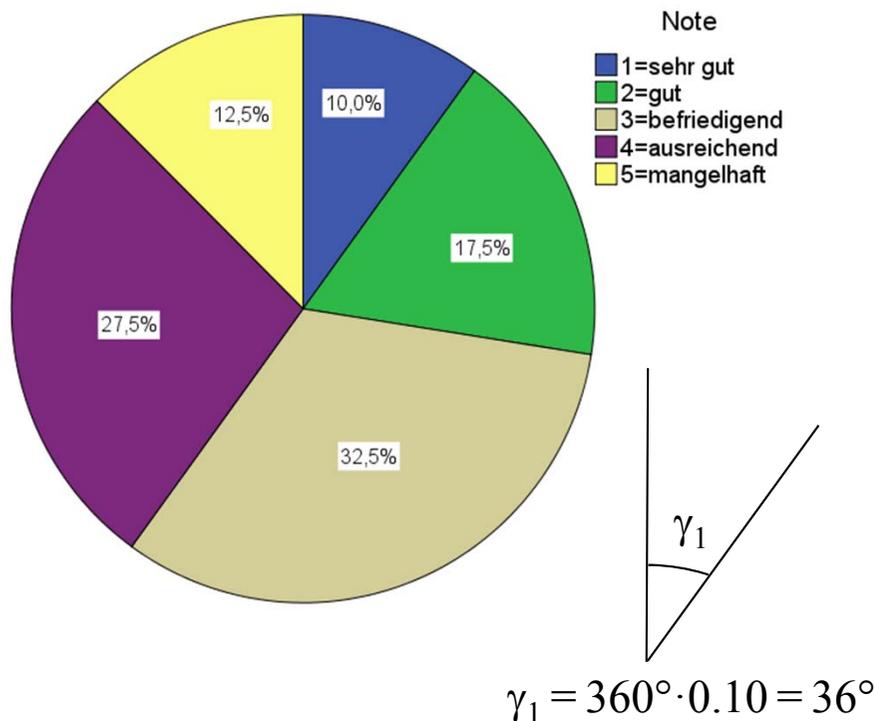


Excel

- Die meisten Computerprogramme legen die Intervallbreiten und die Zahl der daraus resultierenden neuen Kategorien automatisch fest (die besseren erlauben auch eine Veränderung durch den Nutzer).
- Eine Editierung zur optischen Verbesserung ist tw. innerhalb der Programme möglich (oder auch außerhalb mittels Grafikprogrammen wie z.B. Adobe Illustrator)

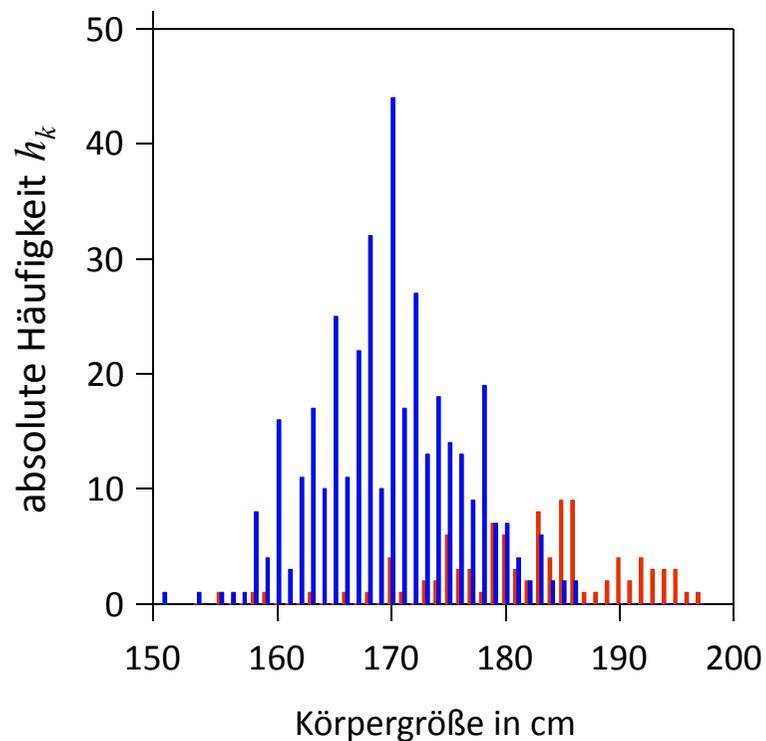
Univariate deskriptive Statistik: Grafische Darstellung von Häufigkeiten

- Bei einem **Kuchendiagramm (pie-chart)** werden die Häufigkeiten von Kategorien als Kreissegmente dargestellt. Der Winkel γ_k eines Tortenstücks ist dabei proportional zur relativen Häufigkeit $\gamma_k = 360^\circ \cdot f_k$.
- Der **Polygonzug** ist ähnlich wie ein Histogramm aufgebaut, nur werden hier die Säulen durch Punkte ersetzt, die durch Geraden verbunden werden. Ggf. kann die Darstellung auch ohne die Punkte erfolgen.

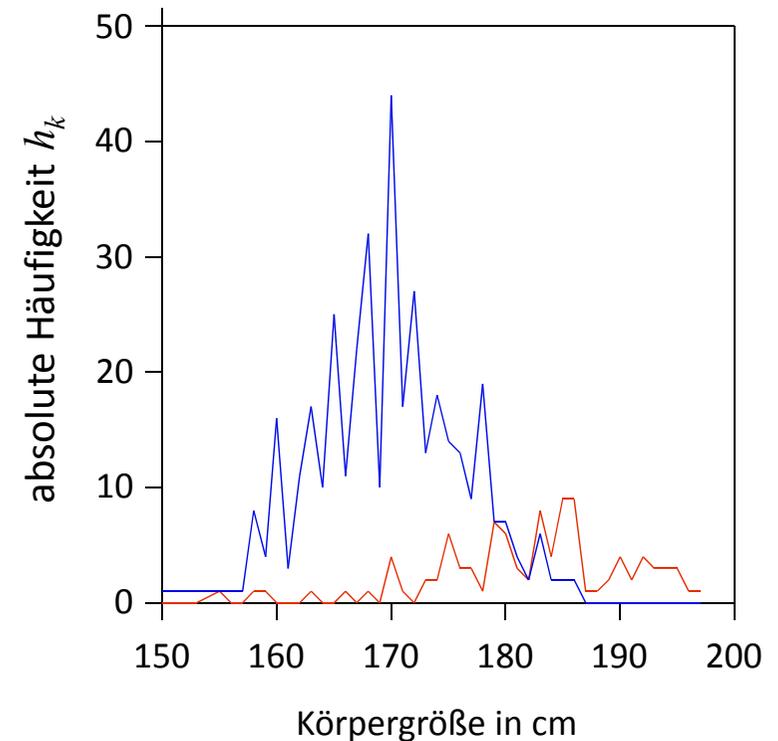


Univariate deskriptive Statistik: Grafische Darstellung von Häufigkeiten

- Wenn man die Verteilungen getrennt für zwei Gruppen von Personen ausgeben will, kann man dies in einer grafischen Darstellung überlagern.
- **Beispiel:** Körpergrößen der Psychologie-Erstsemeser (im Zeitraum WS 03/04 bis WS 09/10; $n = 485$, $n_{\text{♀}} = 381$, $n_{\text{♂}} = 102$), getrennt nach Geschlecht:

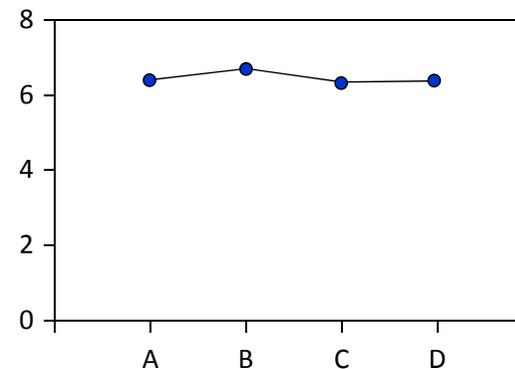
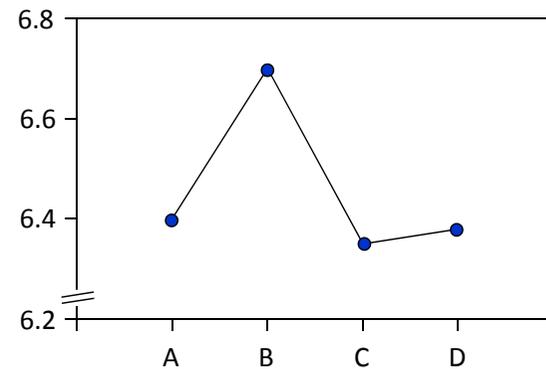
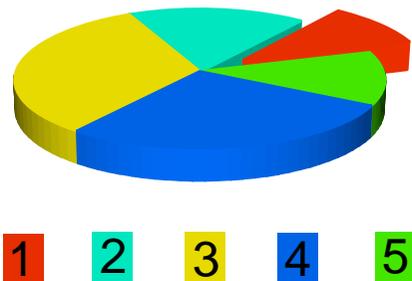
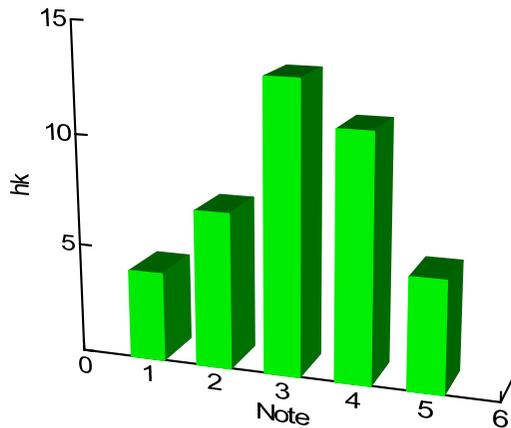


■ männlich
■ weiblich



Univariate deskriptive Statistik: Grafische Darstellungen

- Grafiken sollten so eingesetzt werden, dass die Kommunikation der relevanten Informationen möglichst ohne Missverständnisse gelingt.
- Aus Sicht des Nutzers sind Grafiken immer sorgfältig zu betrachten.



← gleiche Daten
←

1 Deskriptive vs. Inferenzstatistik

2 Formale Grundlagen

3 Darstellung von Verteilungen

(Häufigkeiten, kumulierte Häufigkeiten, primäre vs. sekundäre Häufigkeitsverteilung, Histogramm, Kuchendiagramm, Polygonzug)

4 Maße der zentralen Tendenz

(Arithmetischer Mittelwert, Modalwert, Median)

Univariate deskriptive Statistik: Kennwerte

- Wie kann man eine bestimmte Verteilung näher beschreiben? Zwei wichtige Informationen sind die Lage der Verteilung und die Streubreite der Verteilung.
- Maße (**Kennwerte, Statistiken**), die die Lage der Verteilung (das Niveau) charakterisieren, bezeichnet man als Maße der **zentralen Tendenz**; solche, die die Variabilität der Messwerte quantifizieren als **Streuungsmaße** (oder Dispersionsmaße).
- Welche Maße der zentralen Tendenz und der Variabilität zur Beschreibung einer Verteilung geeignet sind, hängt auch vom Skalenniveau ab.
- Durch die Angabe dieser Statistiken (ohne die Verteilung) erhält man verdichtete Informationen über wichtige Charakteristika der Variablen, es gehen aber auch Informationen verloren (z.B. über die genaue Form der Verteilung).

Maße der zentralen Tendenz: Mittelwert

- Der **arithmetische Mittelwert** \bar{x} ist das gebräuchlichste Maß zur Kennzeichnung der zentralen Tendenz der Verteilung einer intervallskalierten Variablen. Er wird berechnet als die Summe aller Messwerte, dividiert durch die Zahl der Werte n (Zahl der Personen):

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_i + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n} = \frac{1}{n} \cdot \sum_{i=1}^n x_i$$

- **Datenbeispiel 3:** Fünf Vpn einer Arbeitsgruppe haben in einem Intelligenztest die Messwerte (IQ-Werte) 104, 119, 97, 113 und 95 erzielt.

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i = \frac{1}{5} \cdot (104 + 119 + 97 + 113 + 95) = \frac{1}{5} \cdot 528 = 105.6$$

Ergebnis: Der (arithmetische) Mittelwert der Intelligenz in der Gruppe beträgt 105.6.

- Liegen die Daten in einer Häufigkeitstabelle vor (gruppiert in c Kategorien), so ergibt sich

$$\bar{x} = \frac{1}{n} \cdot \sum_{k=1}^c h_k \cdot x_k$$

- Der arithmetische Mittelwert weist folgende Eigenschaften auf:
- Er hängt von allen Messwerten ab (auch von extremen Messwerten, sog. **Ausreißern**).
 - Er muss als Messwert gar nicht auftreten.
 - Wird zu allen Messwerten dieselbe Zahl addiert, vergrößert sich der Mittelwert um eben diese Zahl, d.h. wenn $y_i = x_i + a$ für alle i , dann ist $\bar{y} = \bar{x} + a$.
 - Werden alle Messwerte mit der gleichen Zahl multipliziert, dann erhält man den neuen Mittelwert, indem man ihn mit dieser Zahl multipliziert, d.h. wenn $y_i = b \cdot x_i$ für alle i , dann ist $\bar{y} = b \cdot \bar{x}$.

- Die Summe der Abweichungen aller Messwerte vom Mittelwert ergibt 0 :

$$\sum_{i=1}^n (x_i - \bar{x}) = 0$$

- Die Summe der quadrierten Abweichungen aller Messwerte vom Mittelwert ist ein Minimum, d.h. es gibt keinen anderen Kennwert, bei dem diese kleiner sind:

$$\sum_{i=1}^n (x_i - \bar{x})^2 = \min$$

Maße der zentralen Tendenz: Mittelwert

- Beweis der Behauptung, dass die Summe der Abweichungen aller Werte vom Mittelwert 0 ergibt:

$$\begin{aligned}\sum_{i=1}^n (x_i - \bar{x}) &= \sum_{i=1}^n x_i - \sum_{i=1}^n \bar{x} \\ &= \sum_{i=1}^n x_i - n \cdot \bar{x} \\ &= \sum_{i=1}^n x_i - n \cdot \frac{\sum_{i=1}^n x_i}{n} \\ &= \sum_{i=1}^n x_i - \sum_{i=1}^n x_i \\ &= 0\end{aligned}$$

w.z.b.w.
q.e.d.

- Neben dem arithmetischen Mittelwert gibt es noch andere Mittelwerte, z.B. den **geometrischen Mittelwert**:

$$GM = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_i \cdot \dots \cdot x_n} = (x_1 \cdot x_2 \cdot \dots \cdot x_i \cdot \dots \cdot x_n)^{1/n} \quad (\text{alle } x_i > 0)$$

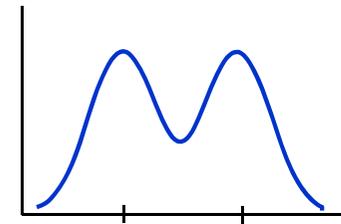
- und den **harmonischen Mittelwert**:

$$HM = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$$

- Beide finden nur in besonderen Fällen Einsatz (für Beispiele vgl. Bortz & Schuster, 2010, S. 29). Im Folgenden meinen wir, wenn wir von „Mittelwert“, „im Mittel“, „Durchschnitt“ etc. sprechen, immer die arithmetische Mittelung.

Maße der zentralen Tendenz: Modalwert

- Der **Modalwert M_o** einer Verteilung ist der Wert, der am häufigsten vorkommt. Um ihn sinnvoll deuten zu können, muss die Variable nur nominalskaliert sein.
- Probleme beim Modalwert ergeben sich, wenn ...
 - die Werte einer kontinuierlichen Variablen sehr fein gestuft sind, da dann im Extremfall jeder Wert nur einmal vorkommt. Evtl. macht es dann Sinn den Modalwert einer sekundären Häufigkeitsverteilung zu bestimmen.
 - wenn es mehrere Werte gibt, die gleich häufig vorkommen (z.B. bei einer bimodalen Verteilung)



Note	Häufigkeit	h_k
1		4
2	 	7
3	 	13
4	 	11
5	 	5

➤ **Datenbeispiel 1** (Mathematik-Noten):

$$M_o = 3$$

Die am häufigsten auftretende Note ist die „3“.

Maße der zentralen Tendenz: Median

- Der **Median** Md ($= \tilde{x}$) ist derjenige Wert einer Verteilung, der die Messwerte in zwei gleich große Teile zerlegt; d.h. mindestens 50% der Werte sind kleinergleich und mindestens 50% größergleich dem Median. Seine Berechnung erfordert mindestens Ordinalskalenniveau.

- Vorgehensweise bei der Bestimmung des Medians

- Alle Messwerte in der Variablen werden aufsteigend sortiert: $x_{(1)} \leq x_{(2)} \dots \leq x_{(n)}$
- Falls n **ungerade** ist, ist der Median dann der Wert in der Mitte der Rangordnung,

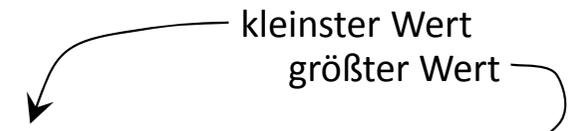
formal: $Md = x_{\left(\frac{n+1}{2}\right)}$

- **Datenbeispiel 3:** Bei $n = 5$ Messwerten ist es also der $(n + 1)/2 = 3$, d.h. der dritte Messwert in der Rangordnung:

$$Md = x_{\left(\frac{n+1}{2}\right)} = x_{\left(\frac{5+1}{2}\right)} = x_{(3)} = 104$$

$x_i:$	104	119	97	113	95
$x_{(i)}:$	95	97	104	113	119

$x_i \neq x_{(i)}$



Maße der zentralen Tendenz: Median

- Vorgehensweise bei der Bestimmung des Medians (Fs.)
 - Falls n gerade ist, ist der Median der Mittelwert zwischen den beiden in der Mitte der Rangordnung liegenden Werten, formal:

$$Md = (x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}) / 2$$

- **Datenbeispiel 3:** Bei $n = 6$ Messwerten ist der Median also der Mittelwert aus dem dritten und vierten Messwert in der Rangordnung (siehe rechts).

x_i :	104	119	97	113	95	120
$x_{(i)}$:	95	97	104	113	119	120

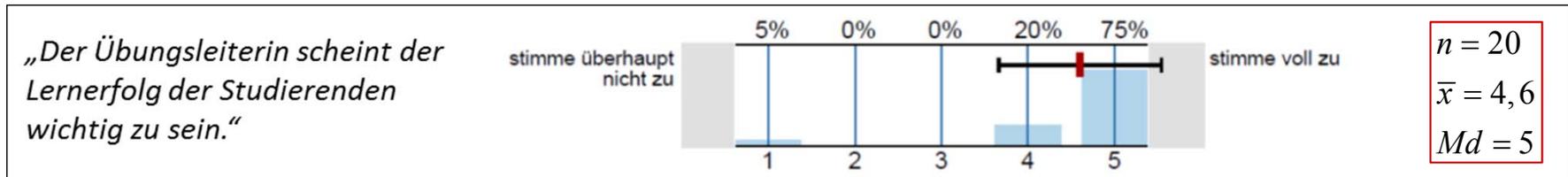
$$\begin{aligned} Md &= (x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}) / 2 = (x_{\left(\frac{6}{2}\right)} + x_{\left(\frac{6}{2}+1\right)}) / 2 \\ &= (x_{(3)} + x_{(4)}) / 2 = (104 + 113) / 2 = 108.5 \end{aligned}$$

- In diesem Beispiel erfüllen auch alle anderen Werte zwischen 104 und 113 die obige Definition des Medians; die Mittelung stellt nur eine Konvention dar. (Tatsächlich unterscheiden sich Computerprogramme auch geringfügig darin, wie sie in diesem Fall die Berechnung vornehmen.)

Maße der zentralen Tendenz: Median

➤ Eigenschaften des Medians

- Er hängt nicht von allen Werten ab (z.B. nicht von Ausreißern).



- Die Summe der absoluten Abweichungen aller Werte vom Median ist ein Minimum:

$$\sum_{i=1}^n |x_i - Md| = \min$$

5 Variabilitätsmaße

(Varianz, Standardabweichung, Range, Quantile, Interquartilabstand)

6 Verteilungsformen

(Modalität, Symmetrie, Gipfligkeit, Schiefe, Exzess, boxplots, stem-and-leaf plots)

7 Ausreißer

8 Transformationen von Messwerten

(z-Werte, Prozentwerte)

- Als ein Maß für die **Variabilität der Messwerte** könnte man sich zunächst die mittlere Abweichung der Messwerte vom Mittelwert vorstellen:

$$\frac{\sum_{i=1}^n (x_i - \bar{x})}{n}$$

- Dieses Maß ist aber ungeeignet, weil sich positive und negative Abweichungen ausgleichen/ausmitteln (s.o). Eine Möglichkeit, um dies zu verhindern, ist stattdessen die Abweichungen zunächst zu quadrieren und dann zu summieren.
- Das dann resultierende wichtige Variabilitätsmaß für intervallskalierte Variablen ist die **Varianz s^2** . Sie ist definiert als die Summe der quadrierten Abweichungen der Messwerte von dem Mittelwert, dividiert durch $n - 1$:

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

- Warum durch $n - 1$ statt durch n dividiert wird, werden wir später behandeln.

Maße der Variabilität: Varianz

- Berechnung anhand von **Datenbeispiel 3**: Fünf Vpn einer Arbeitsgruppe haben die Intelligenzwerte 104, 119, 97, 113, 95 erzielt. Oben bereits berechnet: $\bar{x} = 105.6$

$$\frac{\sum_{i=1}^n (x_i - \bar{x})}{n} = \frac{(104 - 105.6) + (119 - 105.6) + \dots + (95 - 105.6)}{5}$$
$$= \frac{(-1.6) + (13.4) + (-8.6) + (7.4) + (-10.6)}{5} = 0$$



$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} = \frac{(104 - 105.6)^2 + (119 - 105.6)^2 + \dots + (95 - 105.6)^2}{5-1}$$
$$= \frac{(-1.6)^2 + (13.4)^2 + (-8.6)^2 + (7.4)^2 + (-10.6)^2}{4}$$
$$= \frac{2.56 + 179.56 + 73.96 + 54.76 + 112.36}{4}$$
$$= \frac{423.2}{4} = 105.8$$



➤ Eigenschaften der Varianz:

- Sie hängt von allen Messwerten ab.
- Sind alle Messwerte gleich, besteht also keine Variabilität, so weist sie den Wert 0 auf. Je größer die Variabilität, desto größer wird s^2 .
- Durch das Quadrieren fallen größere Abweichungen überproportional ins Gewicht.
- Durch das Quadrieren weist s^2 nicht mehr die Einheiten der Messwerte auf (z.B. cm bei der Körpergröße).
- Wird zu den Messwerten eine Konstante addiert, so verändert dies die Varianz nicht: Wenn $y_i = x_i + a$, dann ist $s_Y^2 = s_X^2$. Anders geschrieben, $s^2(X + a) = s^2(X)$.
- Werden die Messwerte mit einer Konstanten b multipliziert, so verändert sich die Varianz um den Faktor b^2 : Wenn $y_i = b \cdot x_i$, dann ist $s_Y^2 = b^2 \cdot s_X^2$. Oder: $s^2(b \cdot X) = b^2 \cdot s^2(X)$.

➤ Man kann die Varianz alternativ auch nach folgender Formel berechnen (die besser zur Programmierung geeignet ist):

$$s^2 = \frac{\sum_{i=1}^n x_i^2 - \frac{1}{n} \cdot \left(\sum_{i=1}^n x_i \right)^2}{n-1}$$

- Ein weiteres häufig verwendetes Maß ist die **Standardabweichung** s . Sie ergibt sich als Wurzel aus der Varianz:

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

- Alle auf der vorherigen Seite beschriebenen Eigenschaften der Varianz gelten auch für die Standardabweichung, es gibt lediglich zwei Unterschiede:
- Durch das Wurzelziehen wird der Effekt des Quadrieren wieder neutralisiert und s weist wieder die Einheiten der Messwerte auf (z.B. cm bei der Körpergröße)
 - Werden die Messwerte mit einer Konstanten b multipliziert, so verändert sich die Standardabweichung um den Faktor b : Wenn $y_i = b \cdot x_i$, dann ist $s_Y = b \cdot s_X$.
- Berechnung anhand von **Datenbeispiel 3**: Fünf Vpn einer Arbeitsgruppe haben die Intelligenzwerte 104, 119, 97, 113 und 95 erzielt. Oben bereits berechnet: $s^2 = 105.8$.

$$s = \sqrt{s^2} = \sqrt{105.8} = 10.3$$

Maße der Variabilität: Range

- Als weiteres Maß für die Variabilität gibt die **Variationsbreite (Range, Spannweite)** an, wie breit der Bereich der Variablen ist, indem sich Messwerte befinden. Er ist entsprechend definiert als die Differenz zwischen dem größten und dem kleinsten Messwert:

$$\text{Range} = x_{(n)} - x_{(1)} = x_{\max} - x_{\min}$$

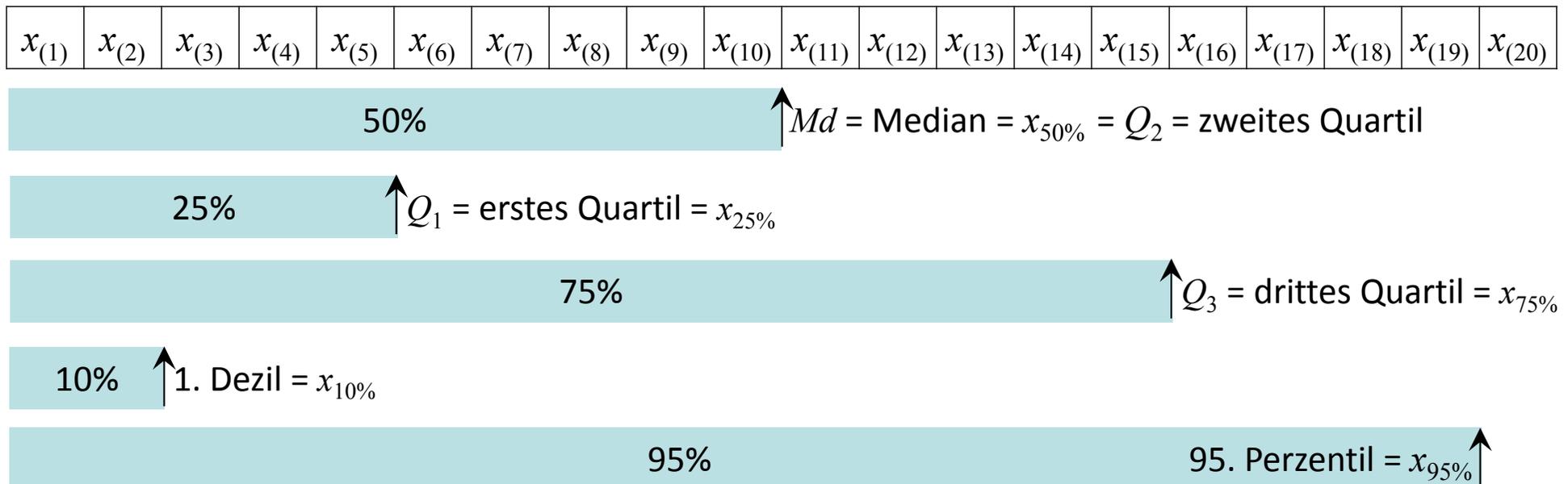
- Eigenschaften des Ranges:
 - Der Range ist nur abhängig von beiden Extremwerten der Verteilung und reagiert damit sehr sensibel auf Ausreißer.
- Berechnung anhand von **Datenbeispiel 3**: Fünf Vpn einer Arbeitsgruppe haben die Intelligenzwerte 104, 119, 97, 113 und 95 erzielt.

$$\text{Range} = x_{\max} - x_{\min} = 119 - 95 = 24$$

Größter und kleiner Messwert unterscheiden sich um 24 Punkte, der Streubereich der Werte beträgt 24.

Maße der Variabilität: Exkurs Quantile

- Eine Verallgemeinerung des Medians stellen **Quantile** dar. Quantile sind Kennzahlen, die die geordneten Messwerte einer Variablen in einem bestimmten Verhältnis teilen.
- Ein Quantil ist derjenige Wert $x_{p\%}$ ($0 < p < 100$), für den gilt, dass mindestens $p\%$ der Messwerte kleiner oder gleich $x_{p\%}$ und mindestens $(1 - p)\%$ der Messwerte größer oder gleich $x_{p\%}$ sind.
- Je nach Anzahl der Unterteilungen unterscheidet man die Spezialfälle Median (2er Teilung), **Quartile** (4er Teilung), **Dezentile/Dezile** (10er) und **Perzentile/Centile** (100er).



Maße der Variabilität: Exkurs Quantile

➤ Ein **Quantil** wird z.B. wie folgt berechnet:

• Falls $q = n \cdot p/100$ ganzzahlig: $x_{p\%} = (x_{(q)} + x_{(q+1)})/2$

• Falls q nicht ganzzahlig: $x_{p\%} = x_{(\lceil q \rceil)}$

mit $\lceil q \rceil$ als nächstgrößere ganze Zahl, die auf q folgt.

Spezialfall Md mit $q = n/2$

$$x_{50\%} = (x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)})/2 \quad (\text{gerade})$$

$$x_{50\%} = x_{(\frac{n+1}{2})} \quad (\text{ungerade})$$

➤ **Datenbeispiel 3***: $Q_3 = x_{75\%}$ für Stichprobe mit $n = 9$ Intelligenzwerten.

$q = 9 \cdot 75/100 = 6.75$ nicht ganzzahlig

$\lceil q \rceil = 7 \quad Q_3 = x_{75\%} = x_{(7)} = 104$

$x_{(1)}$	$x_{(2)}$	$x_{(3)}$	$x_{(4)}$	$x_{(5)}$	$x_{(6)}$	$x_{(7)}$	$x_{(8)}$	$x_{(9)}$
91	95	97	101	103	103	104	113	119

➤ **Datenbeispiel 3***: 20. Perzentil $x_{20\%}$ für Stichprobe mit $n = 10$ Intelligenzwerten.

$q = 10 \cdot 20/100 = 2$ ganzzahlig

$x_p = (x_{(2)} + x_{(2+1)})/2 = (95 + 97)/2 = 96$

$x_{(1)}$	$x_{(2)}$	$x_{(3)}$	$x_{(4)}$	$x_{(5)}$	$x_{(6)}$	$x_{(7)}$	$x_{(8)}$	$x_{(9)}$	$x_{(10)}$
91	95	97	101	103	103	104	113	115	119

➤ Wie schon beim Median unterscheiden sich auch bei Quantilen Computerprogramme darin, wie sie die Berechnung vornehmen.

Maße der Variabilität: Interquartilabstand

- Auf der Basis der Quartile lässt sich als weiteres Streuungsmaß der **Interquartilsabstand** (oder **-bereich**) bestimmen. Er gibt an, wie stark die mittleren 50% der Verteilung streuen:

$$IQA = Q_3 - Q_1$$

- Eigenschaft des Interquartilabstands:

- Durch die Betrachtung der Streubreite des mittleren Bereichs ist der *IQA* robust gegen Ausreißer.

- **Datenbeispiel 3***: Stichprobe mit $n = 9$ Intelligenzwerten.

$$Q_3 = x_{75\%} = 104 \text{ (s.o.)}$$

$$Q_1 = x_{25\%} ?$$

$$q = 9 \cdot 25 / 100 = 2.25 \text{ nicht ganzzahlig; } \lceil q \rceil = 3$$

$$Q_1 = x_{25\%} = x_{(\lceil q \rceil)} = 97$$

$x_{(1)}$	$x_{(2)}$	$x_{(3)}$	$x_{(4)}$	$x_{(5)}$	$x_{(6)}$	$x_{(7)}$	$x_{(8)}$	$x_{(9)}$
91	95	97	101	103	103	104	113	119

↑
 Q_1

↑
 Q_3

$IQA = 7$, d.h. die mittleren 50%

der Stichprobe streuen 7 IQ-Punkte

- $IQA = Q_3 - Q_1 = 104 - 97 = 7$

Streuungsmaße (Eugen Roth)

*Ein Mensch der von Statistik hört,
denkt dabei nur an den Mittelwert.
Er glaubt nicht dran und ist dagegen,
ein Beispiel soll es gleich belegen:
Ein Jäger auf der Entenjagd
hat einen ersten Schuß gewagt.
Der Schuß zu hastig aus dem Rohr,
lag eine gute Handbreit vor.
Der zweite Schuß mit lautem Krach
lag eine gute Handbreit nach.
Der Jäger spricht ganz unbeschwert
voll Glauben an den Mittelwert:
"Statistisch ist die Ente tot."
Doch wär er klug und nähme Schrot-
dies sei gesagt, Ihn zu bekehren
würde seine Chancen mehren:
Der Schuß geht ab, die Ente stürzt,
weil Streuung ihr das Leben kürzt.*

5 Variabilitätsmaße

(Varianz, Standardabweichung, Range, Quantile, Interquartilabstand)

6 Verteilungsformen

(Modalität, Symmetrie, Gipfligkeit, Schiefe, Exzess, boxplots, stem-and-leaf plots)

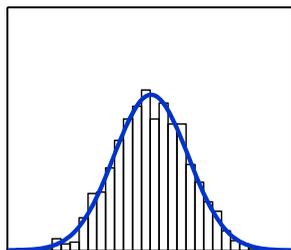
7 Ausreißer

8 Transformationen von Messwerten

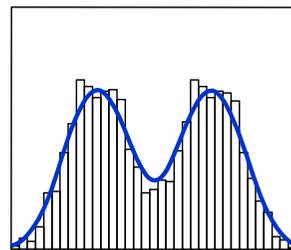
(z-Werte, Prozentwerte)

Univariate deskriptive Statistik: Verteilungsformen

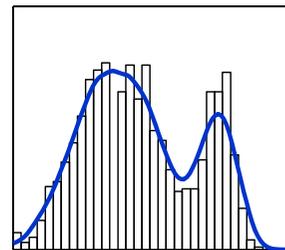
- Wir haben oben gesehen, dass wir die Verteilung grafisch darstellen können (z.B. als Histogramm oder Polygonzug). Wir wollen nun einige Eigenschaften der Form von Verteilungen näher betrachten.
- Verteilungen können **unimodal** = **eingipflig** sein (d.h. einen Gipfel/Peak aufweisen) oder bimodal (zweigipflig), trimodal usw. sein (allg. **multimodal**).
- Verteilungen können **symmetrisch** oder mehr oder weniger **asymmetrisch** (=schief) sein.
 - Ist eine Verteilung unimodal und asymmetrisch, so kann der Schwerpunkt (z.B. der Modalwert) links (**linkssteile** = **linksgipflig** = **rechtsschief** Verteilung; positively skewed) oder rechts (**rechtssteil** = **rechtsgipflig** = **linksschief**; negatively skewed) vom Mittelwert liegen.
 - Ist sie symmetrisch, so ist $\bar{x} = Md$; ist sie zudem eingipflig, so gilt $\bar{x} = Md = Mo$.



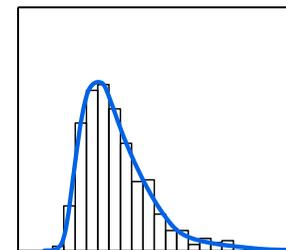
unimodal
symmetrisch



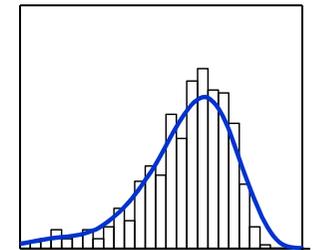
bimodal
symmetrisch



bimodal
asymmetrisch



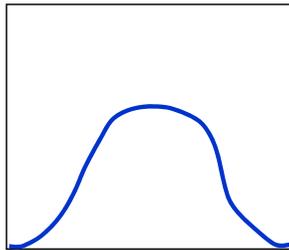
unimodal
linkssteil



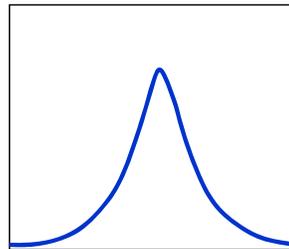
unimodal
rechtssteil

Univariate deskriptive Statistik: Verteilungsformen

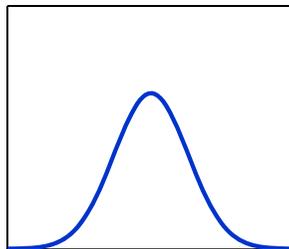
- Verteilungen können sich in der **Form des Gipfels** unterscheiden. Liegen viele Werte in der Nähe des Modalwerts, so bezeichnet man die Verteilung als **stumpf-** oder **breitgipflig** (**platykurtisch**); andernfalls **schmal-** oder **steilgipflig** (**leptokurtisch**). Bezugspunkt ist dabei die Normalverteilung.



platykurtisch



leptokurtisch



normalverteilt



paranormal
verteilt

Univariate deskriptive Statistik: Verteilungsformen

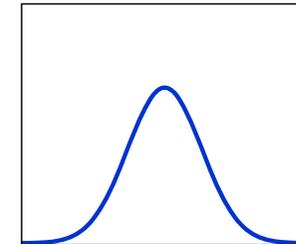
- Das Ausmaß der Asymmetrie und der Breitgipfligkeit lässt sich durch statistische Kennzahlen quantifizieren. Diese charakterisieren Verteilungen wie die bereits bekannten Kennzahlen Mittelwert oder Varianz.
- Die **Schiefe** (**skewness**) ist ein Maß für Asymmetrie der Verteilung:

$$Sch = \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{n \cdot s^3}$$

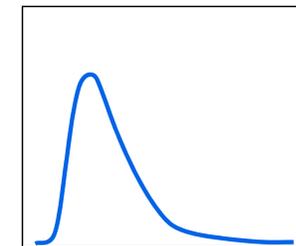
- **Eigenschaften** der Schiefe-Statistik

- Bei symmetrischen Verteilungen ist $Sch = 0$.
- Bei linkssteilen Verteilungen ist $Sch > 0$.
- Bei rechtssteilen Verteilungen ist $Sch < 0$.

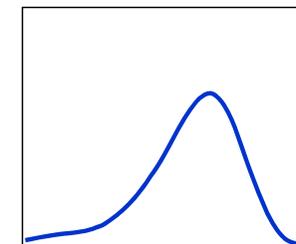
- (Bei der Schiefe und auch dem auf der folgenden Folie dargestellten Exzess verwenden einige Programme auch etwas abweichende Formeln.)



$Sch = 0$



$Sch > 0$



$Sch < 0$

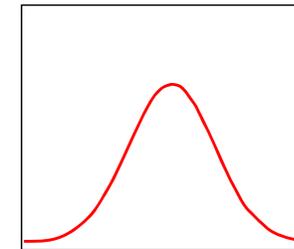
- Der **Exzess (Kurtosis)** quantifiziert das Ausmaß an Breitgipfligkeit (bei einer unimodalen Verteilung):

$$Ex = \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{n \cdot s^4} - 3$$

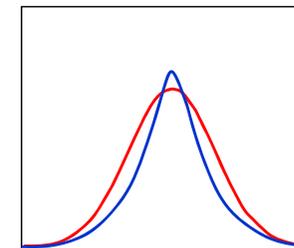
- Als Bezugsgröße für die Breite der Gipfform wird eine theoretische, die **Normalverteilung**, herangezogen (normalgipflig, mesokurtisch).

- **Eigenschaften** der Exzess-Statistik:

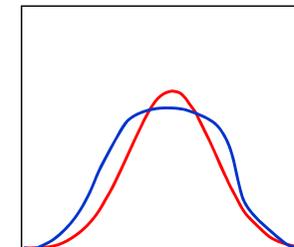
- Bei einer Normalverteilung ist $Ex = 0$.
- Ist die Verteilung steilgipfliger als die Normalverteilung (leptokurtisch), so ist $Ex > 0$.
- Ist die Verteilung breitgipfliger als die Normalverteilung (platykurtisch), so ist $Ex < 0$.



$$Ex = 0$$



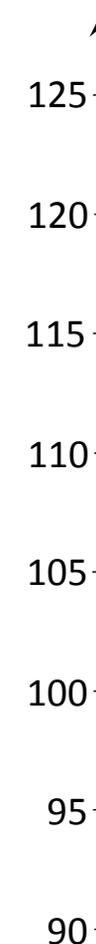
$$Ex > 0$$



$$Ex < 0$$

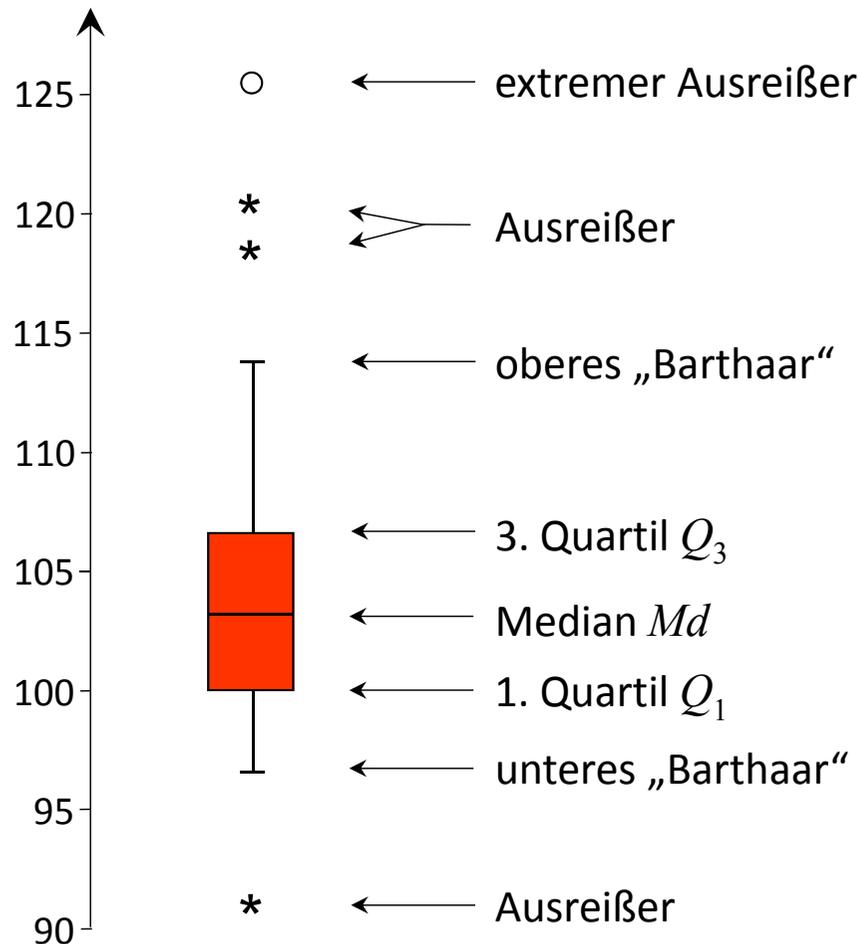
Verteilungsformen: Boxplots

- Eine informative grafische Darstellungsform von Verteilungen sind **Box-Whisker Diagramme** (**Boxplots**; Tukey, 1977). Sie stellen nicht die gesamte Verteilung, aber simultan eine Reihe von (gegenüber Ausreißern robusten) Kenngrößen dar.
- Die Variable wird auf der Ordinate dargestellt.
 - Der Kasten (box) umschließt die mittleren 50 Prozent der Verteilung und enthält als Unterteilung die Kennzeichnung des Medians.
 - An die Box angeheftet finden sich oben und unten Barthaare (whiskers), die den „weiteren“ Streubereich der Werte anzeigen.
 - Fälle außerhalb dieses Bereichs werden als Ausreißerwerte durch Kreise bzw. Sterne markiert.



Verteilungsformen: Boxplots

- Im Detail sind die Elemente eines Boxplots (meist) wie folgt definiert:



Als extreme Ausreißer werden Werte bezeichnet, die mehr als $3 \cdot IQA$ außerhalb der Box liegen.

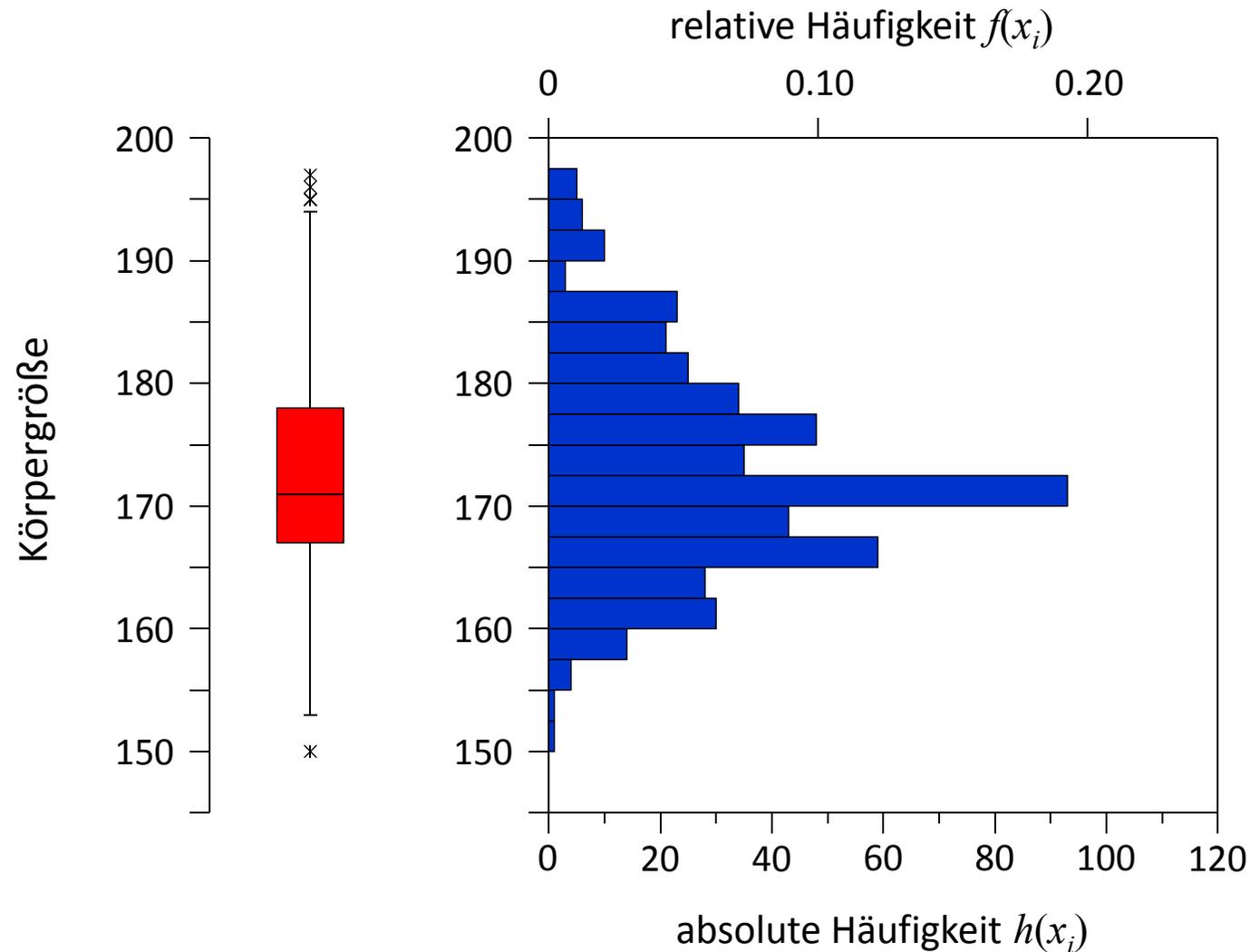
Die Barthaare (whiskers) enden jeweils am letzten Wert, der sich innerhalb des Bereichs von $1.5 \cdot IQA$ außerhalb der Box befindet.

Die Box reicht von Q_1 bis Q_3 (hinges) und enthält den Median.

Als Ausreißer werden Werte bezeichnet, die zwischen $1.5 \cdot IQA$ und $3 \cdot IQA$ außerhalb der Box liegen.

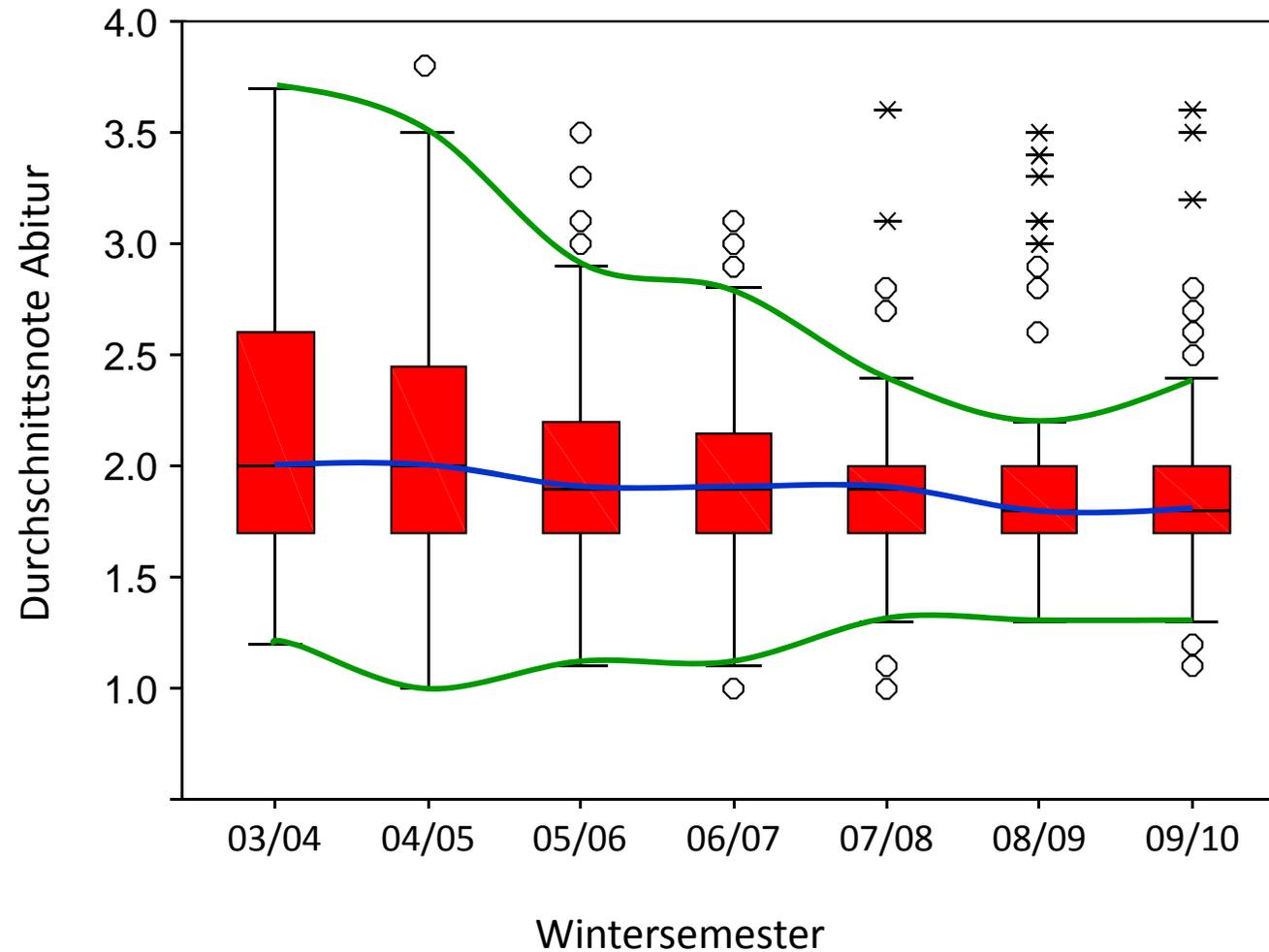
Verteilungsformen: Boxplots

- **Beispiel:** Körpergrößen der Psychologie-Erstsemester (im Zeitraum WS 03/04 bis WS 09/10):



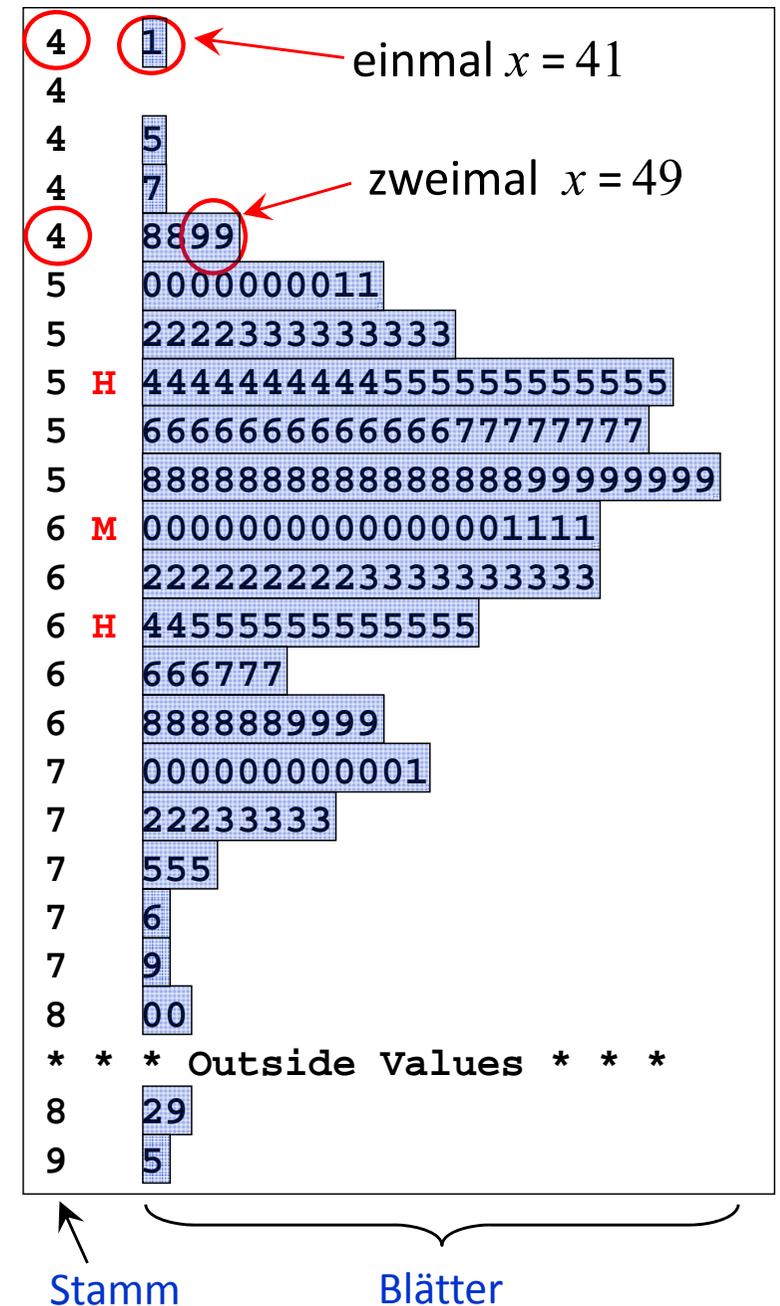
Verteilungsformen: Boxplots

- **Beispiel:** Abiturdurchschnittsnoten von $n = 481$ Psychologie-Erstsemester, getrennt für Kohorten WS 03/04 bis WS 09/10:



Verteilungsformen: Stem-and-leaf Plots

- Ähnlich können Statistiken ohne Grafik in einer textlichen Darstellung in sog. **stem-and-leaf Diagrammen/Plots** (Tukey, 1977) eingezeichnet werden.
- Dabei werden die ersten Stellen des Messwertes als **Stamm** (stem) links dargestellt und die letzte Stelle als **Blatt** (leaf), z.B. der Wert 41 mit 4 als Stamm und 1 als Blatt.
- **Beispiel:** Körpergewicht von $n = 195$ Psychologie-Erstsemesterstudentinnen im Wertebereich von 41 bis 95 kg (nur WS 03/04 bis WS 06/07).
- In dieser Darstellungsform des Programms Systat werden zusätzlich der Median (**M** = 60) und die Quartile (**H** = hinges 55 und 65) dargestellt sowie die Ausreißer speziell markiert (hier die Werte 82, 89 und 95 kg).



5 Variabilitätsmaße

(Varianz, Standardabweichung, Range, Quantile, Interquartilabstand)

6 Verteilungsformen

(Modalität, Symmetrie, Gipfligkeit, Schiefe, Exzess, boxplots, stem-and-leaf plots)

7 Ausreißer

8 Transformationen von Messwerten

(z-Werte, Prozentwerte)

Ausreißer

- In den Daten können extrem hohe oder niedrige, für die Verteilung untypische Datenwerte auftreten, die auch als **Ausreißer** (**outlier**) bezeichnet werden.
- Ausreißer können „per Augenschein“ in der Häufigkeitsverteilung oder anhand von statistischen Kriterien (wie beim Boxplot) festgestellt werden. Sie verdienen besondere Aufmerksamkeit.
- Ausreißer können ganz verschiedene Ursachen haben, z.B.
 - machen Vpn versehentlich einen Fehler, z.B. bei der Beantwortung einer Frage.
 - haben Vpn eine Instruktion oder Frage falsch verstanden.
 - geben Vpn mutwillig falsche Antworten.
 - passieren Fehler durch Messgeräte.
 - passieren Fehler bei der Dateneingabe (z.B. wird als Gewicht statt 67.5 der Wert 675 eingegeben).
 - weisen die Vpn korrekter Weise extreme Merkmalsausprägungen auf (z.B. ein Gedächtniskünstler in einer Studie zum Gedächtnis).
- (Viele der Ursachen können, müssen aber nicht zu Ausreißern führen.)

Ausreißer

- Ausreißer können die Ergebnisse statistischer Analysen, insbesondere bei kleinem n , deutlich verändern.
- Was man mit Ausreißern sinnvollerweise tun sollte, hängt von der Ursache ab:
 - Falls es sich nachweislich um einen fehlerhaften Wert handelt und man den korrekten Wert noch feststellen kann, kann man korrigieren.
 - Kann man den korrekten Wert in diesem Fall nicht mehr feststellen, so kann man den einzelnen fehlerhaften Wert löschen = auf „**fehlenden Wert**“ (=missing data) setzen.
 - In einzelnen begründeten Fällen (z.B. bei völlig unsinnigen Antworten der Vp) kann man diese Vp ganz aus dem Datensatz löschen (und dies in der Studie berichten).
 - Man kann Ausreißer in dem Datensatz belassen, berichten, und die Analysen mit und ohne Ausreißer durchführen und analysieren, inwieweit sich die Ergebnisse unterscheiden.
 - Man kann robustere Statistiken verwenden, also solche die weniger stark von Ausreißern beeinflusst werden (also z.B. den Median statt den Mittelwert).

„Der Übungsleiterin scheint der Lernerfolg der Studierenden wichtig zu sein.“



$n = 20$
 $\bar{x} = 4,6$
 $Md = 5$
 $Mo = 5$

5 Variabilitätsmaße

(Varianz, Standardabweichung, Range, Quantile, Interquartilabstand)

6 Verteilungsformen

(Modalität, Symmetrie, Gipfligkeit, Schiefe, Exzess, boxplots, stem-and-leaf plots)

7 Ausreißer

8 Transformationen von Messwerten

(z-Werte, Prozentwerte)

z-Transformation

- Wie kann man beispielsweise eine Punktzahl von $x_i = 32$ in einer Klausur (oder eine Reaktionszeit von 180 msec etc.) bewerten?
- Eine Möglichkeit, die Leistung zu bewerten, besteht darin, sich die maximal erreichbare Punktzahl anzusehen oder zu schauen, ob die Person damit die Klausur bestanden hat (**kriteriumsorientierte Bewertung**). Man wüsste dann beispielsweise, dass die Person 32 von maximal 40 Punkten erreicht hat oder dass sie bestanden hat. Ob diese Leistung schwierig oder leicht zu erreichen war, kann man daraus nicht schließen.
- Eine andere Möglichkeit besteht darin, die Leistung der Person in Bezug zu setzen zur Leistung der Anderen (**normorientierte Bewertung**).
- Dazu könnte man die Position der Person relativ zum Mittelwert in der Stichprobe betrachten, z.B. in dem man einen **Abweichungswert** a_i bestimmt:

$$a_i = x_i - \bar{x}$$

Beträgt der Mittelwert etwa $\bar{x} = 28$ und damit $a_i = 32 - 28 = 4$, so erkennt man, dass die Leistung über dem Durchschnitt liegt. Wir wissen aber nicht, ob 4 Punkte über (oder auch unter) dem Durchschnitt viel oder wenig sind.

z-Transformation

- Um dies einschätzen zu können, müssen wir die Variabilität der Werte in der Stichprobe zusätzlich betrachten.
- Wir können dazu die Abweichungswerte noch durch die Standardabweichung dividieren und erhalten dadurch sogenannte **z-transformierte** Messwerte (für $s > 0$):

$$z_i = \frac{a_i}{s} = \frac{x_i - \bar{x}}{s}$$

- Falls in unserem Beispiel die Standardabweichung der Klausurergebnisse in der Stichprobe $s = 2$ beträgt, ergibt sich ein deutlich höherer z-Wert von $z_i = (32 - 28)/2 = 2$ als wenn die Streuung mit $s = 4$ größer ist: $z_i = (32 - 28) / 4 = 1$.
- Man bezeichnet die Berechnung von Abweichungswerten auch als **Zentrierung** und die von z-Werten auch als **(z-) Standardisierung**.
- Die Zentrierung führt zu einer Verschiebung der Verteilung auf den Mittelwert 0 (additive Konstante) und die Division durch s zu einer Stauchung oder Streckung der Verteilung, so dass sie eine Streuung von 1 aufweist (multiplikative Konstante). Die Form der Verteilung (d.h. ihre Symmetrie, Schiefe) ändert sich durch die z-Transformation nicht.

z-Transformation

- Datenbeispiel 3: $n = 5$ Intelligenzwerte.

Person i	Messwert x_i	Abweichungswert a_i	z-Wert z_i
1	104	$104 - 105.6 = -1.6$	$-1.6/10.3 = -0.16$
2	119	$119 - 105.6 = 13.4$	$13.4/10.3 = 1.30$
3	97	-8.6	-0.84
4	113	7.4	0.72
5	95	-10.6	-1.03
Mittelwert:	105.6	0	0
Standardabw.:	10.3	10.3	1

- Mittels z-Werten kann man verschiedene Messungen miteinander vergleichen. **Beispiel:** Ist die Leistung der obigen Person A mit einem Punktwert von 32 (bei einem Mittelwert von $\bar{x} = 28$ und einer Standardabweichung von $s = 4$) höher zu bewerten als die einer Person B , die in einer veränderte Klausur ein Jahr später 30 Punkte erzielt hat (mit $\bar{x} = 24$ und $s = 4$)?

$$z_A = \frac{x_i - \bar{x}}{s} = \frac{32 - 28}{4} = 1$$

$$z_B = \frac{x_i - \bar{x}}{s} = \frac{30 - 24}{4} = 1.5$$

Antwort: nein

z-Transformation

➤ Eigenschaften der z-Werte:

- Die z-Werte weisen (wie die Abweichungswerte) einen Mittelwert von 0 auf.
- Die z-Werte weisen eine Standardabweichung von 1 auf.
- Der z-Wert gibt an, wie viele Standardabweichungen ein Wert über dem Mittelwert liegt.
- Die z-Werte weisen nicht die Einheiten der Messwerte (z.B. cm bei Körpergröße) auf, sondern sind sog. dimensionslose Größen.
- Die z-Werte stellen eine lineare Transformation der Messwerte dar:

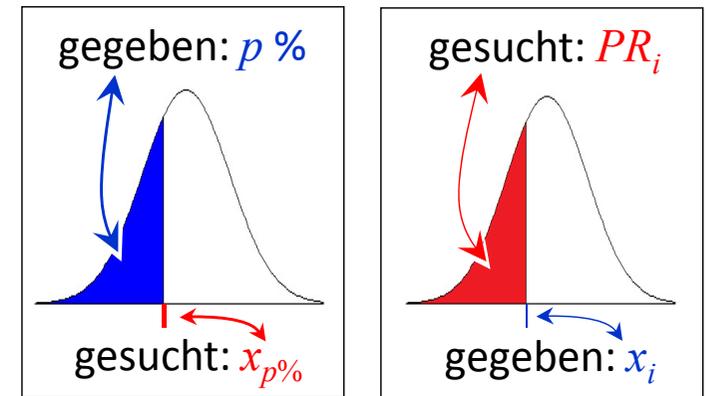
$$z_i = \frac{x_i - \bar{x}}{s} = \frac{1}{s} \cdot x_i - \frac{\bar{x}}{s}$$

The diagram illustrates the decomposition of the z-score formula. The term $\frac{1}{s}$ is highlighted in blue and labeled as the 'multiplikative Konstante' (multiplicative constant). The term $-\frac{\bar{x}}{s}$ is highlighted in red and labeled as the 'additive Konstante' (additive constant). Arrows point from these labels to their respective terms in the equation.

- ## ➤ z-Werte sind Statistiken, die an vielen Stellen in der Statistik, Testtheorie etc. von Bedeutung sind.

Prozentränge

- Bei den Quartilen bestimmen wir den Messwert $x_{p\%}$, der p Prozent der Verteilung abschneidet.
- Umgekehrt können wir auch fragen, wieviel Prozent der Verteilung einen kleineren oder gleichen Messwert wie eine Person i aufweisen.
- Der **Prozentrang** PR_i (percentil score) einer Person i in einer mindestens ordinalskalierten Variablen entspricht dem Anteil der Personen in der Stichprobe, die gleiche oder kleinere Messwerte aufweisen. Er stellt keine lineare Transformation der Messwerte dar.
- Der Prozentrang für einen in der Stichprobe aufgetretenen Wert x_i entspricht der kumulierten prozentualen Häufigkeiten $pc(x_i)$ der primären Häufigkeitsverteilung.
- **Datenbeispiel 1** (Schulnoten). Z.B. ist der Prozentrang der Person 4 mit einer Note $x_4 = 3$ $PR_4 = pc(3) = 60.0$. Deutung: 60 Prozent der Personen in der Stichprobe haben also gleich oder besser in der Mathematikarbeit als Person 4 abgeschnitten.



Note	$h(x_i)$	$p(x_i)$	$pc(x_i)$
1	4	10.0	10.0
2	7	17.5	27.5
3	13	32.5	60.0
4	11	27.5	87.5
5	5	12.5	100.0

Zitierte Quellen

- Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.

Änderungen am 31.10.2016

- Folie 28: Zahlendreher bei Berechnung der Varianz: 432.2 durch 423.2 ersetzt (Ergebnis ist korrekt).