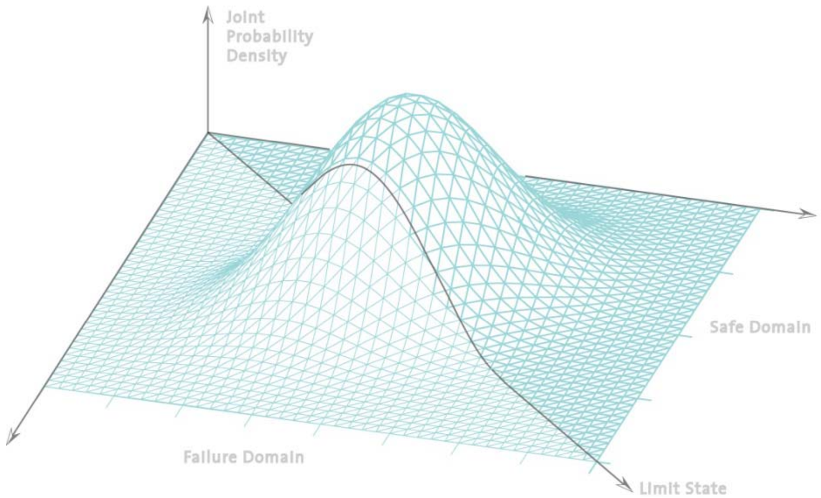


Statistik und Wahrscheinlichkeitsrechnung

Dr. Jochen Köhler

Warum Statistik und Wahrscheinlichkeits-rechnung im Ingenieurwesen?

- Zusammenfassung der letzten Vorlesung

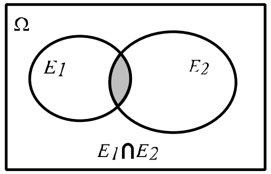
Statistik und Wahrscheinlichkeitsrechnung ist im Ingenieurwesen notwendig, um:

- Unsicherheiten im Zusammenhang mit Ingenieurmodellen zu quantifizieren.
- die Ergebnisse von Experimenten zu dokumentieren und zu bewerten.
- die Wichtigkeit von unsicheren Einflussgrössen beurteilen zu können.
- effiziente Entscheidungen treffen zu können.

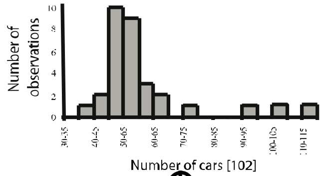
Aufbau der Vorlesung

Module A
Entscheidungsfindung im Ingenieurswesen unter Unsicherheit

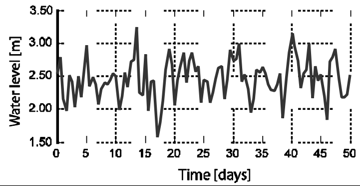
Module B - Grundlagen der Wahrscheinlichkeitstheorie



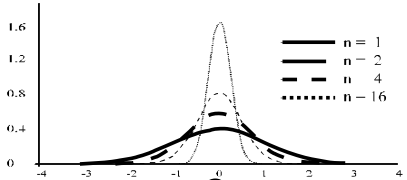
Module C - Beschreibende Statistik



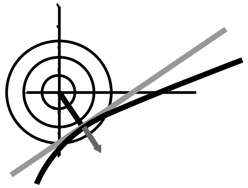
Module D - Modellierung von Unsicherheiten



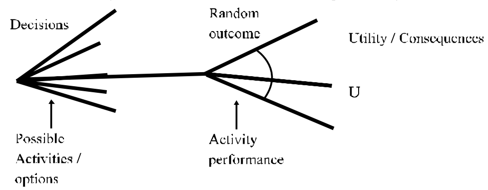
Module E - Abschätzung und Modelerstellung



Module F - Methoden der strukturellen Zuverlässigkeit



Module G - Bayes'sche Entscheidungsanalyse



Karten Warm-up

Was studieren Sie?

Bauwesen

Geodäsie

Umwelt

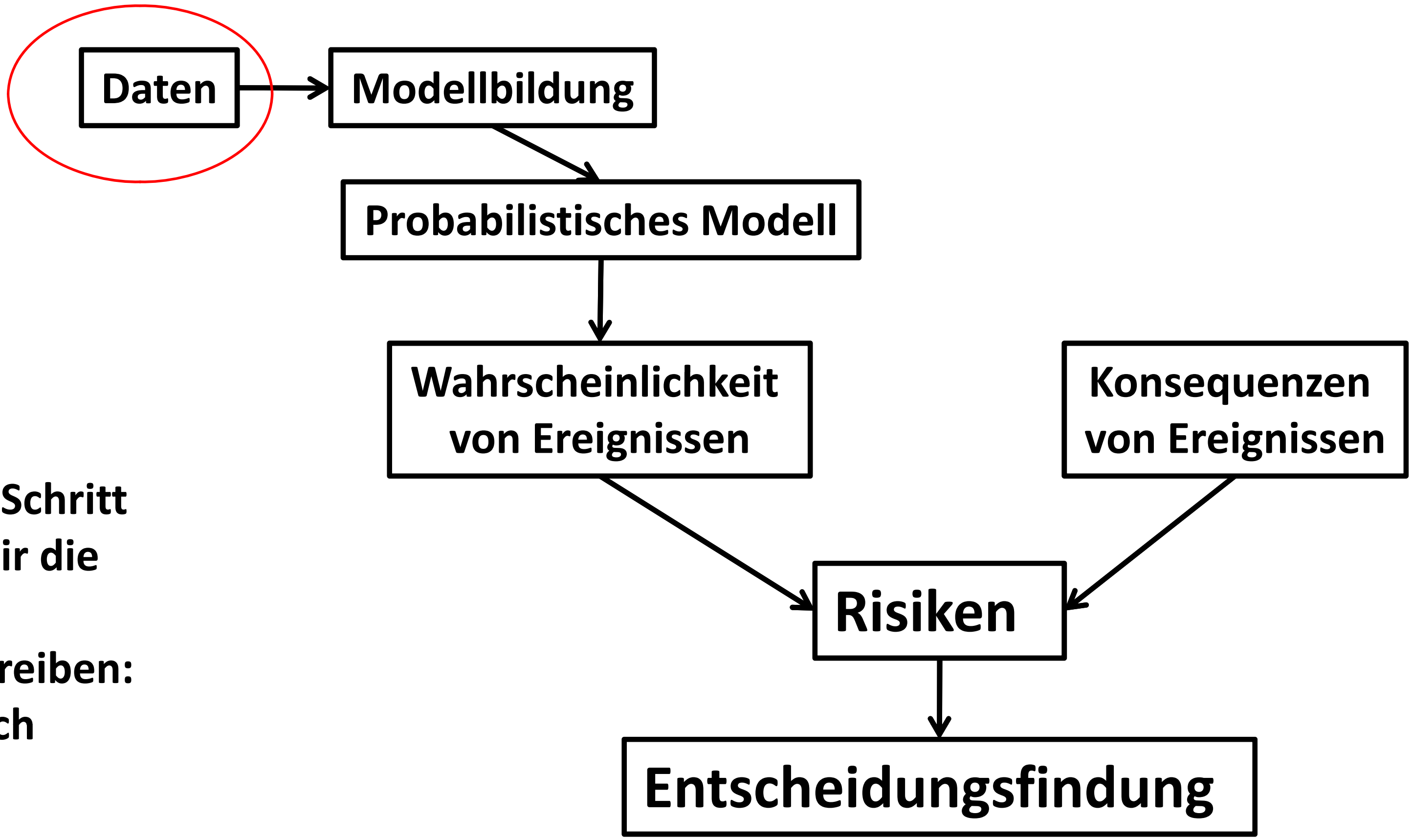
Kleine Denkaufgabe 2.1

Welchen Nutzen hat
Statistik und Wahrscheinlichkeitsrechnung
im Ingenieurwesen?

- (rot) Im Ingenieurwesen keinen – nützt höchstens dem Verständnis von Wahlergebnissen.
- (grün) Ermöglicht Entscheidungsfindung bei aussergewöhnlichen Fragestellungen.
- (orange) Weiss nicht…

Inhalte der heutigen Vorlesung

- Ziel:



Im ersten Schritt werden wir die Daten nur beschreiben:
- numerisch
- grafisch

Inhalte der heutigen Vorlesung

- Überblick der beschreibenden Statistik

- **Numerische Zusammenfassungen**

  Mit welchen einfachen Zahlen können Datenmengen charakterisiert werden?

- **Grafische Darstellung von Datenmengen**

  Wie werden Datenmengen informativ in Grafiken umgesetzt?

Ziel der beschreibenden Statistik

- Beschreiben von Datenmengen

Körpergrösse

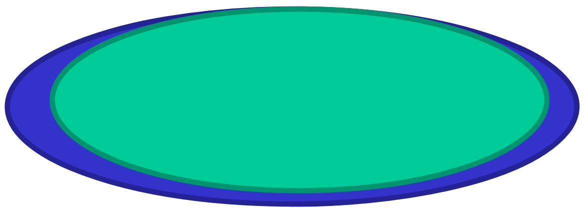
170	191	184	184	182	176	183	164	178	183
190	176	175	170	171	180	176	177	187	170
190	171	183	182	178	180	185	175	180	184
175	169	183	176	179	172	176	180	183	182
173	165	175	190	160	189	174	184	191	165
170	165	178	180	176	185	187	174	187	184
183	166	177	189	197	174	166	186	184	174
178	183	180	176	185	178	185	185	184	183
190	186	183	183	178	188	185	181	175	171
175	170	168	178	185	184	187	162	170	183
175	174	187	176	184	183	184	195	180	178
183	187	160	200	170	179	160	179	180	
164	172	175	181	170	179	189	182	183	
176	164	175	176	188	185	190	179	175	
169	176	162	175	187	175	173	180	174	
178	180	175	185	182	182	168	183	170	
188	178	158	177	186	176	184	182	170	
187	191	158	173	158	183	178	165	174	
164	174	187	175	172	177	187	186	181	
183	178	172	183	176	173	187	175	175	

Vorbemerkung

- Stichprobe und Grundgesamtheit
  - Die statistischen Eigenschaften einer Grundgesamtheit werden anhand von Stichproben untersucht.

    Z.B.: Die Grundgesamtheit aller Studierenden, welche für Statistik und Wahrscheinlichkeitsrechnung eingeschrieben sind, ist m = 199.

    Stichprobe von letzter Woche, n = 191.

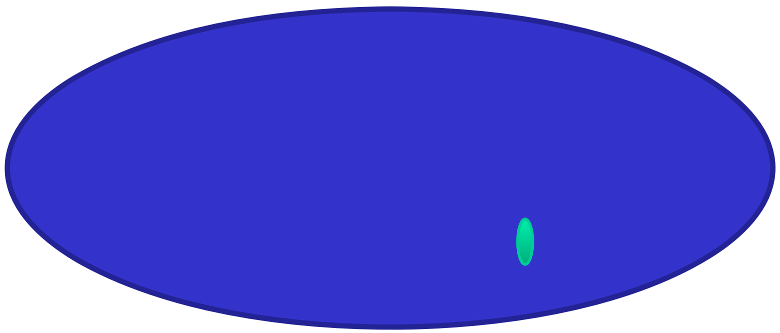


Vorbemerkung

- Stichprobe und Grundgesamtheit
  - Die statistischen Eigenschaften einer Grundgesamtheit werden anhand von Stichproben untersucht.

    Z.B.: Biegezähigkeit von Büroklammern, $m = \infty$.

    Stichprobe, $n = 190$

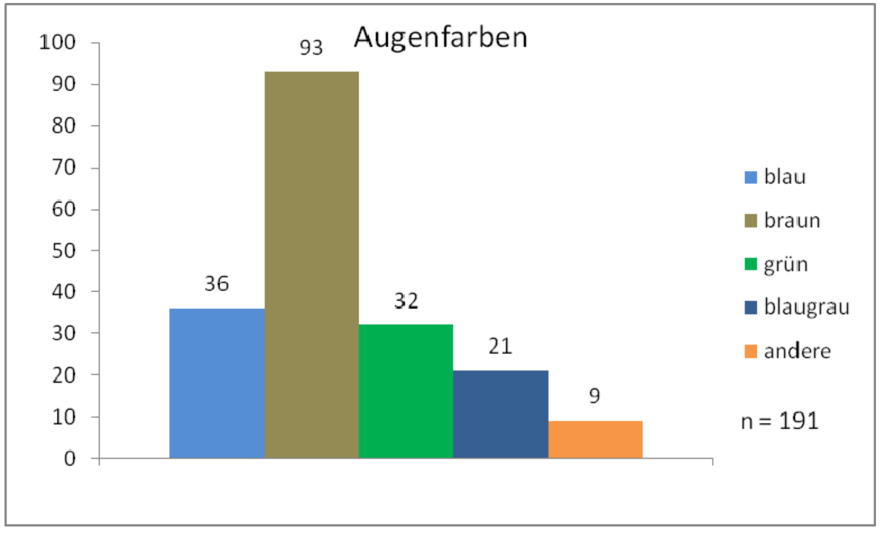


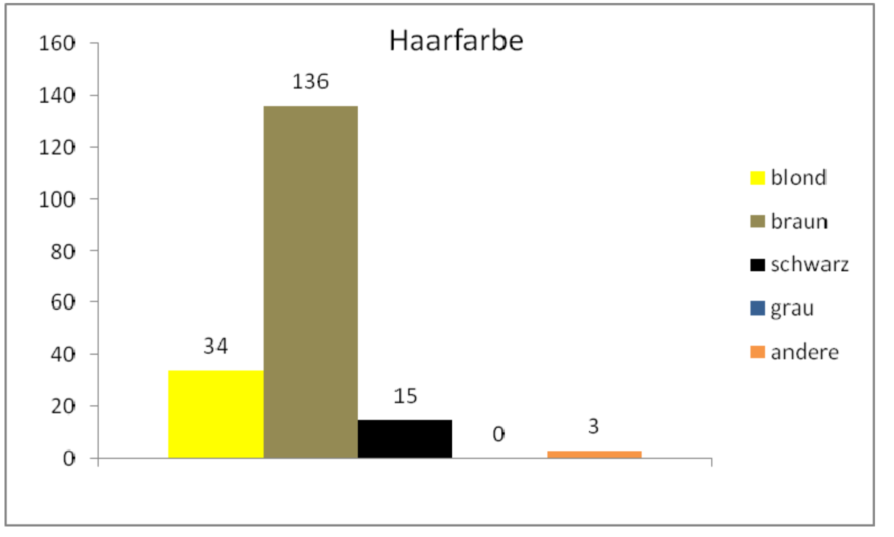
Vorbemerkung

- Stichprobe und Grundgesamtheit
  - Die statistischen Eigenschaften einer Grundgesamtheit werden anhand von Stichproben untersucht.
  - Damit die Stichprobe die Grundgesamtheit repräsentiert, müssen die Stichproben **zufällig** aus der Grundgesamtheit entnommen werden.

Vorbemerkung

- Skalenniveau
  - Nominalskala: Qualitative Eigenschaften, welche nicht der Grösse nach sortiert werden können.





Vorbemerkung

- Skalenniveau
  - Ordinalskala: Qualitative Eigenschaften, welche der Grösse nach sortiert werden können – über den Abstand zwischen den Eigenschaften lässt sich nichts aussagen.
    (Schulnoten, Ligatabelle)
  - Intervallskala: Quantitative Eigenschaften, sortierbar, der Abstand zwischen zwei Werten lässt sich sachlich begründen. Nullpunkt willkürlich festgelegt.
    (Temperatur in °C, Jahreszahlen)
  - Verhältnisskala: Wie Intervallskala, aber mit absolutem Nullpunkt
    (Temperatur in Kelvin, Festigkeit, Körpergrösse)

Ziel der beschreibenden Statistik

- Beschreiben von Datenmengen

Körpergrösse

170	191	184	184	182	176	183	164	178	183
190	176	175	170	171	180	176	177	187	170
190	171	183	182	178	180	185	175	180	184
175	169	183	176	179	172	176	180	183	182
173	165	175	190	160	189	174	184	191	165
170	165	178	180	176	185	187	174	187	184
183	166	177	189	197	174	166	186	184	174
178	183	180	176	185	178	185	185	184	183
190	186	183	183	178	188	185	181	175	171
175	170	168	178	185	184	187	162	170	183
175	174	187	176	184	183	184	195	180	178
183	187	160	200	170	179	160	179	180	
164	172	175	181	170	179	189	182	183	
176	164	175	176	188	185	190	179	175	
169	176	162	175	187	175	173	180	174	
178	180	175	185	182	182	168	183	170	
188	178	158	177	186	176	184	182	170	
187	191	158	173	158	183	178	165	174	
164	174	187	175	172	177	187	186	181	
183	178	172	183	176	173	187	175	175	

⇨ **Zahlen**

⇨ **Grafiken**

Keine Annahmen – nur Beschreibung !!

Datenbeschreibung

- Zusammenfassen zu nur einer Zahl

Arithmetisches Mittel: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

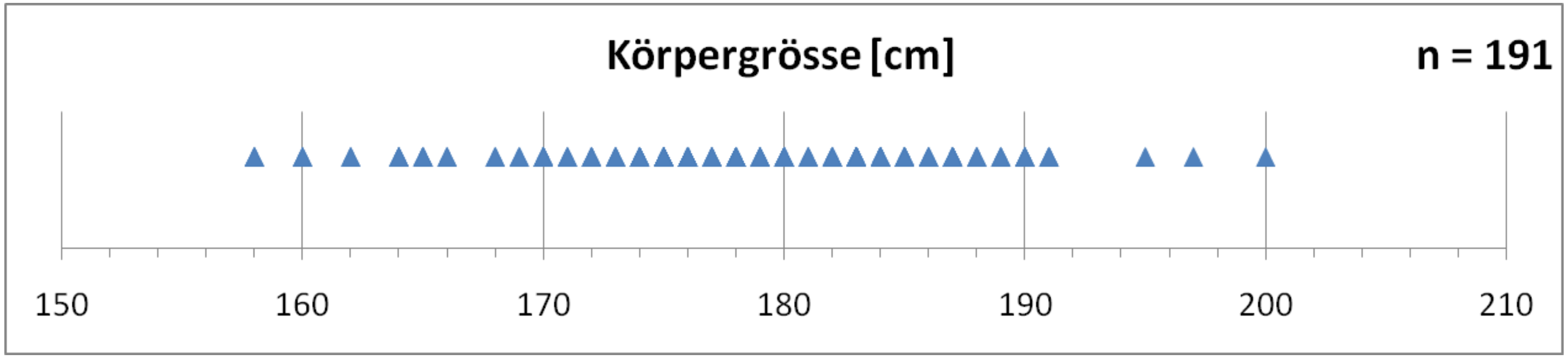
Für einen Datensatz: $\mathbf{x} = \left(x_1, x_2, \ldots, x_n\right)^T$

Um eine Stichprobe nur mit Hilfe einer Zahl zu beschreiben, wird normalerweise der Stichproben-Mittelwert verwendet.

Datenbeschreibung

- Einfache graphische Darstellung von Stichproben

Eindimensionales Streudiagramm:



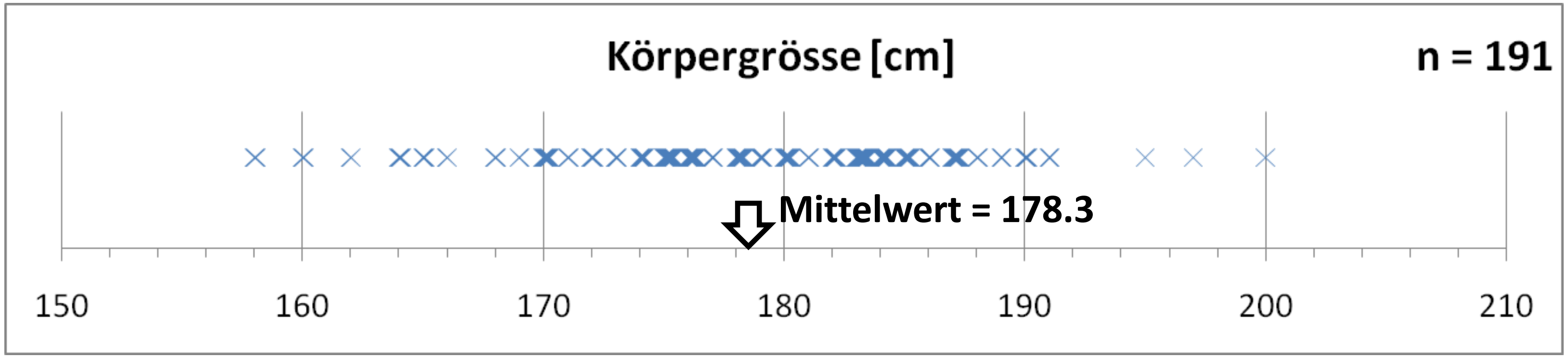
Guter Datenüberblick (Maximum, Minimum).

Vorsicht bei diskret verteilten Daten !

Datenbeschreibung

- Einfache graphische Darstellung von Stichproben

Eindimensionales Streudiagramm:



Der Stichprobenmittelwert $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ entspricht dem „Schwerpunkt" der Daten.

Datenbeschreibung

- Einfache graphische Darstellung von Stichproben

Histogramm:

Einteilung der Datenreihe in Klassen.

Darstellung der Grösse der Klassen.

z.B. die Körpergrösse

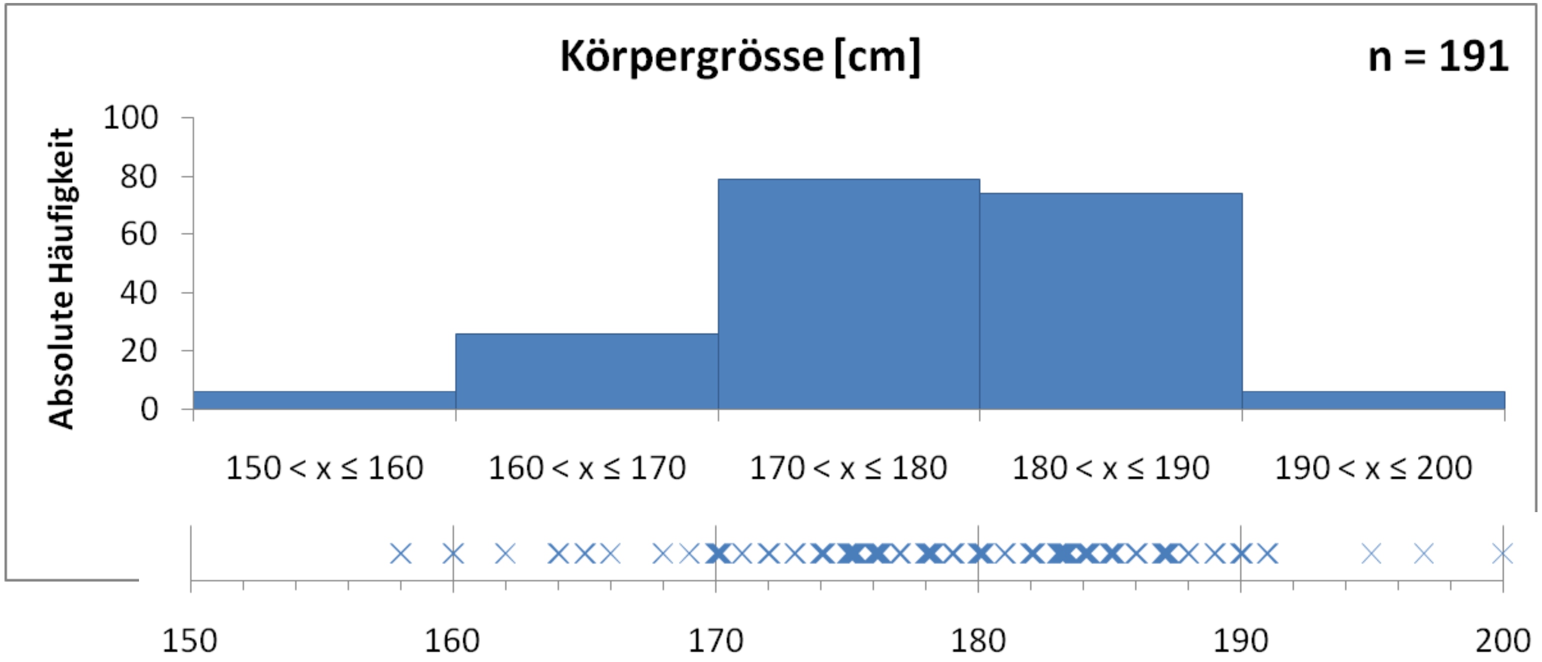
Klassen	Anzahl
$150 < x \leq 160$	6
$160 < x \leq 170$	26
$170 < x \leq 180$	79
$180 < x \leq 190$	74
$190 < x \leq 200$	6
n =	191

Datenbeschreibung

- Einfache graphische Darstellung von Stichproben

Klassen	Anzahl
150 < x ≤ 160	6
160 < x ≤ 170	26
170 < x ≤ 180	79
180 < x ≤ 190	74
190 < x ≤ 200	6
n =	191

Histogramm:



Datenbeschreibung

- Neben dem Mittelwert gibt es noch andere sog. Lagemasse:
  - Der **Median** oder Zentralwert $\tilde{x}$ ist der mittlere Wert einer nach der Grösse geordneten Stichprobe $x_1^o \leq x_2^o \leq \ldots \leq x_n^o$.

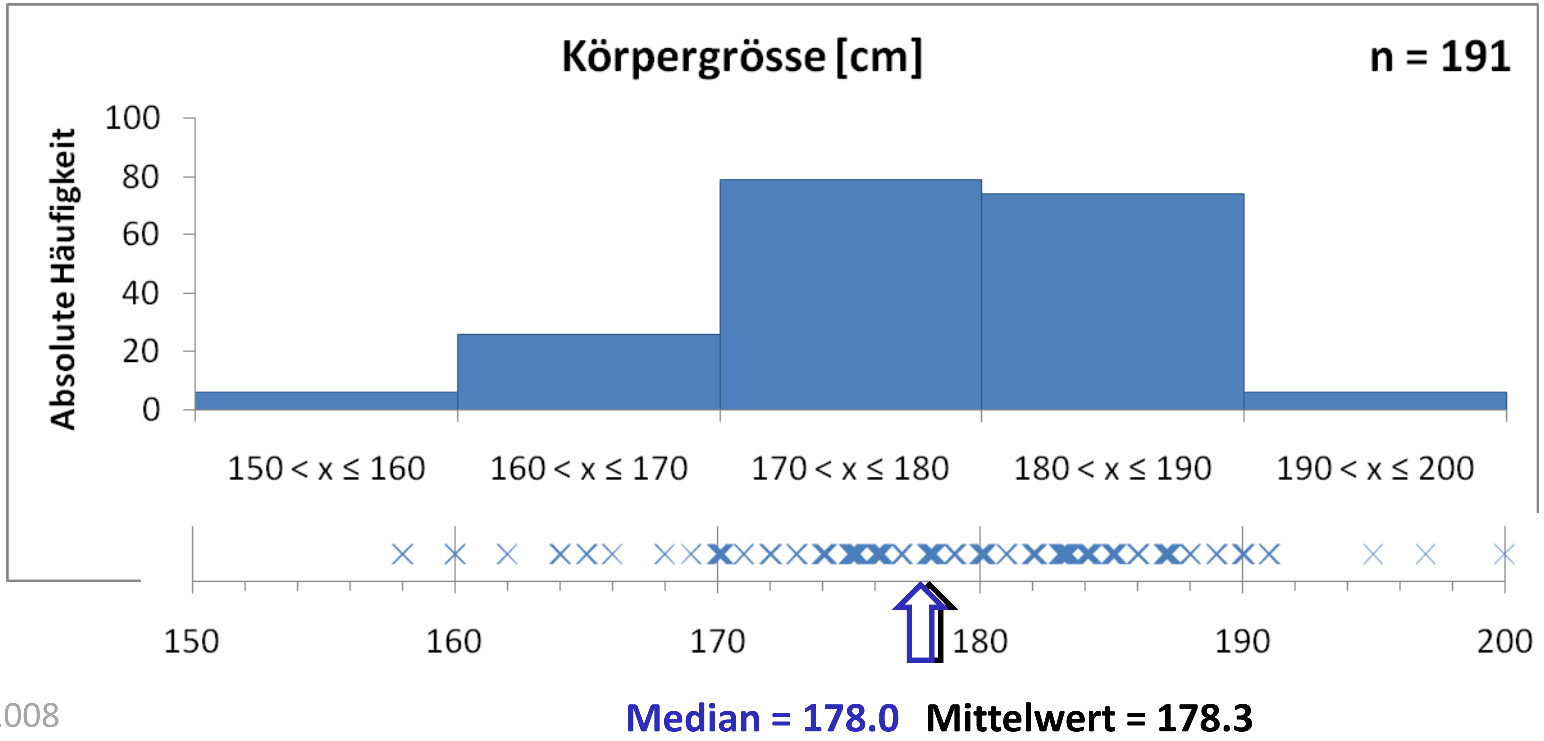
$$\tilde{x} = \begin{cases} x_{\frac{n+1}{2}} & \text{n ungerade} \\ \frac{1}{2}\left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}\right) & \text{n gerade} \end{cases}$$

  - Beispiele: [23 30 31 33 120]

    [23 30 31 33]

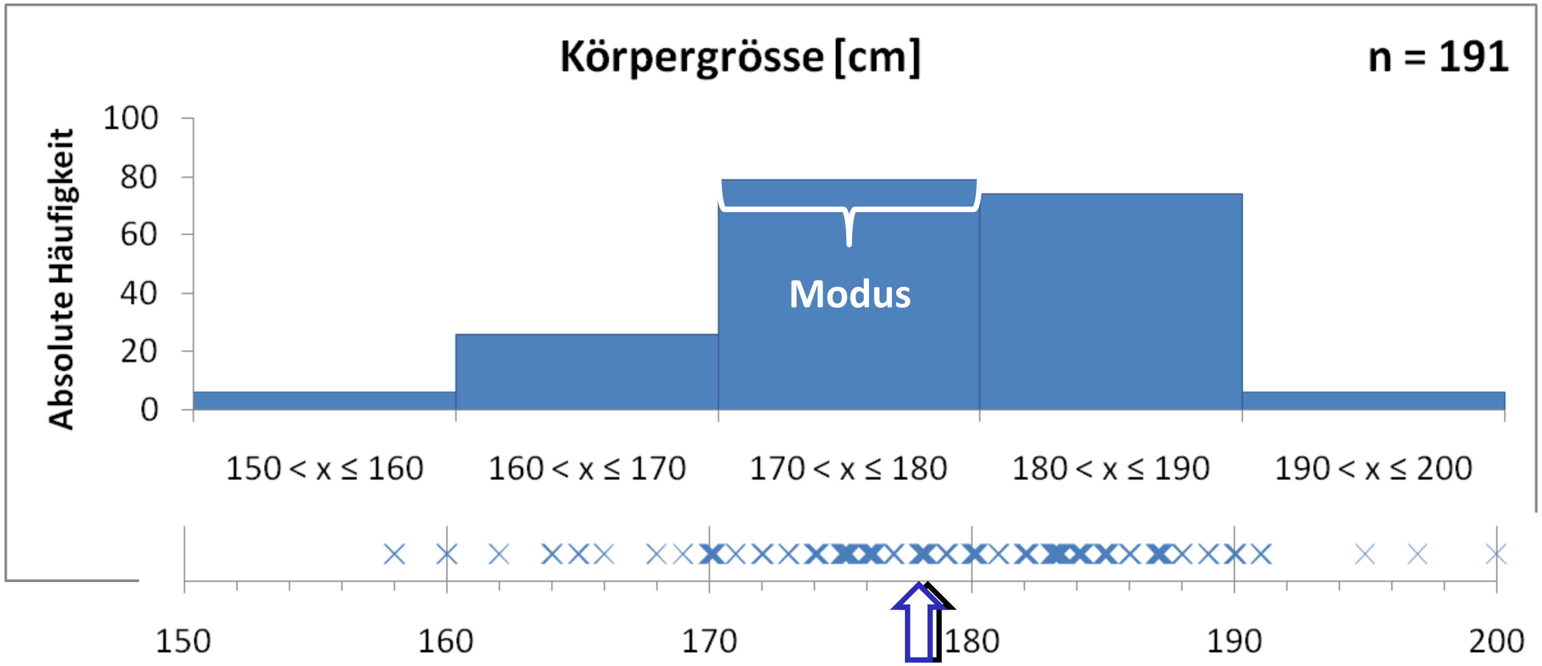
Datenbeschreibung

- Neben dem Mittelwert gibt es noch andere sog. Lagemasse:
  - Der Median oder Zentralwert $\tilde{x}$ ist der mittlere Wert einer nach der Grösse geordneten Stichprobe $x_1^o \leq x_2^o \leq \ldots \leq x_n^o$.



Datenbeschreibung

- Neben dem Mittelwert gibt es noch andere sog. Lagemasse:
  - Der **Modus** oder Modalwert ist der am häufigsten auftretende Wert – bei kontinuierlichen Wertemengen u.a. aus Histogramm ersichtlich.



Median = 178.0 **Mittelwert = 178.3**

Datenbeschreibung

- Streumasse – Streuung um den Mittelwert

  - Die Varianz der Stichprobe $$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

  - Die Standardabweichung der Stichprobe $$s = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

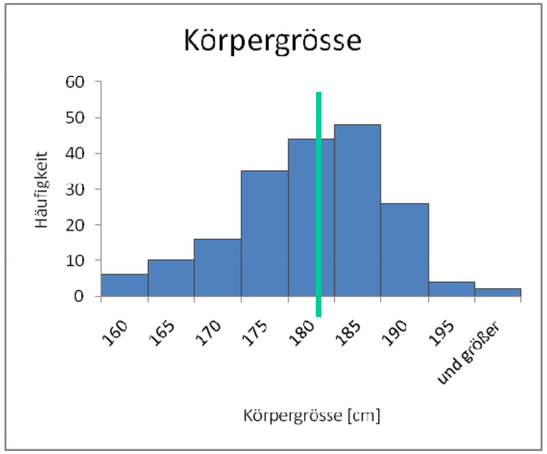
  - Der Variationskoeffizient der Stichprobe (relative Streuung) $$v = \frac{s}{\overline{x}}$$

Datenbeschreibung

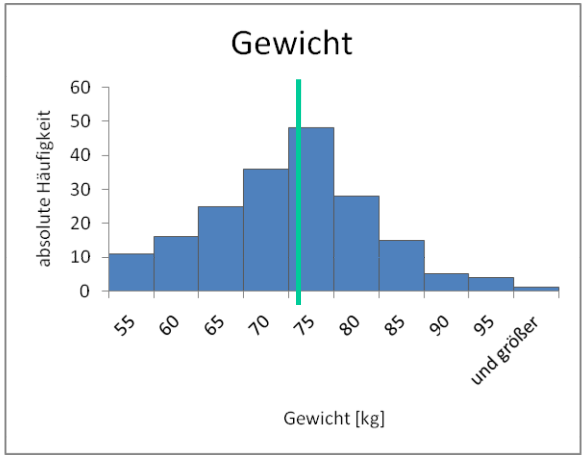
- Streumasse – Streuung um den Mittelwert

Varianz $s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$ Standardabweichung $s = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}$ COV $v = \frac{s}{\overline{x}}$

Beispiel



$\overline{x} = 178.3$ [cm]
$s^2 = 61.65$ [cm^2]
$s = 7.85$ [cm]
$v = 0.04$ [-]



$\overline{x} = 71.2$ [kg]
$s^2 = 86.11$ [kg^2]
$s = 9.28$ [kg]
$v = 0.13$ [-]

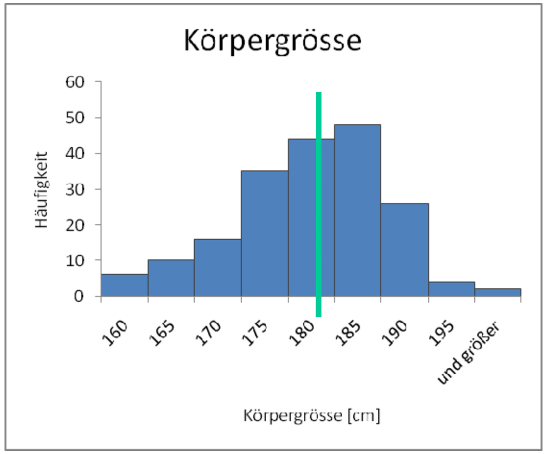
Datenbeschreibung

- Streumasse – Streuung um den Mittelwert

  - Der Schiefekoeffizient der Stichprobe
    -> Mass für die Asymmetrie

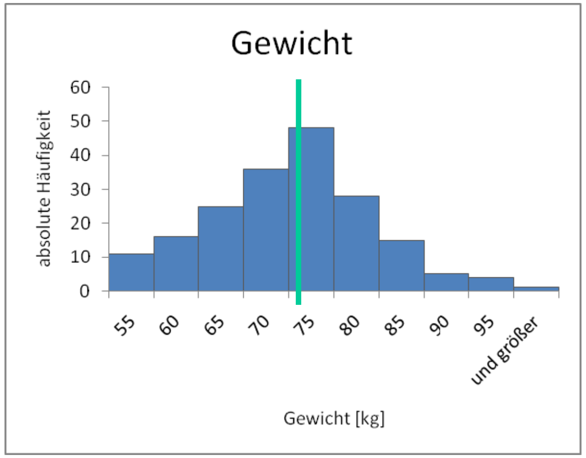
$$\eta = \frac{1}{n} \cdot \frac{\sum_{i=1}^{n} (x_i - \bar{x})^3}{s^3}$$

Beispiel



$\eta = -0.36$

Linksschief



$\eta = 0.1$

Rechtsschief

Datenbeschreibung

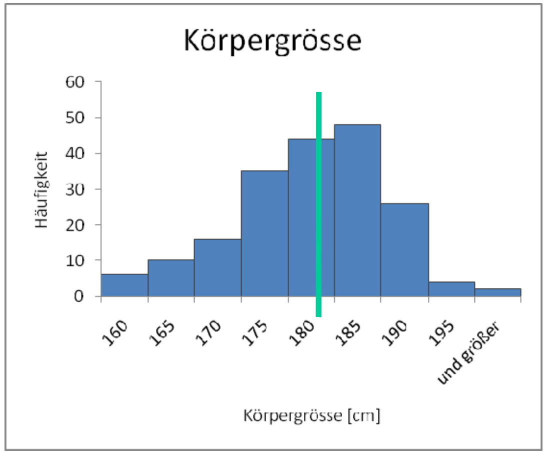
- Streumasse – Streuung um den Mittelwert

  - Kurtosis der Stichprobe:

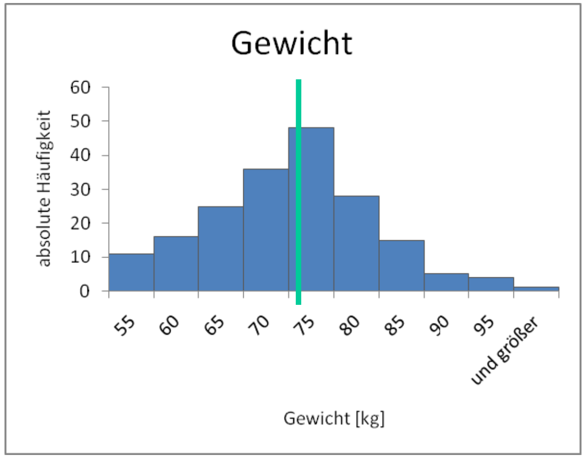
    -> Mass für die Wölbung

$$\kappa = \frac{1}{n} \cdot \frac{\sum_{i=1}^{n} (x_i - \overline{x})^4}{s^4}$$

Beispiel



$\kappa = 3.05$



$\kappa = 3.04$

Datenbeschreibung

- Beschreibung von paarweise beobachteten Eigenschaften

$$\mathbf{x} = \left(x_1, x_2, x_3, \ldots, x_n\right)^T$$

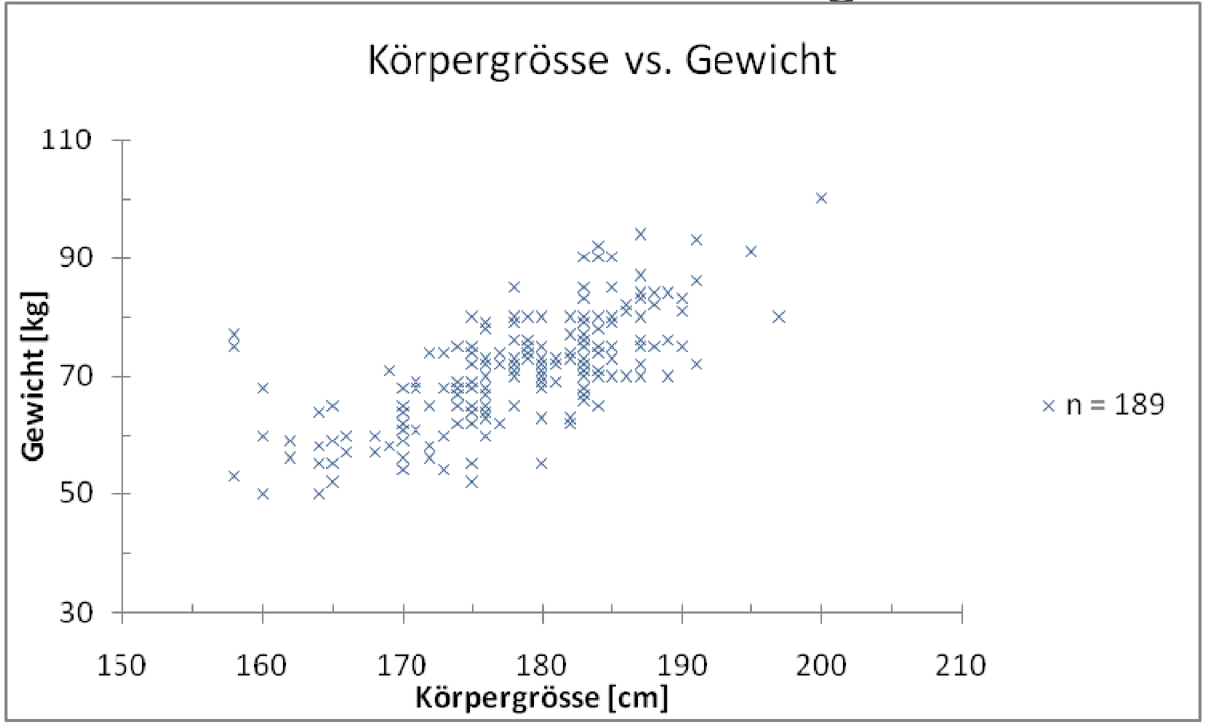
$$\begin{array}{cccc} \updownarrow & \updownarrow & \updownarrow & \quad \updownarrow \end{array}$$

$$\mathbf{y} = \left(y_1, y_2, y_3, \ldots, y_n\right)^T$$

Datenbeschreibung

- Beschreibung von paarweise beobachteten Eigenschaften

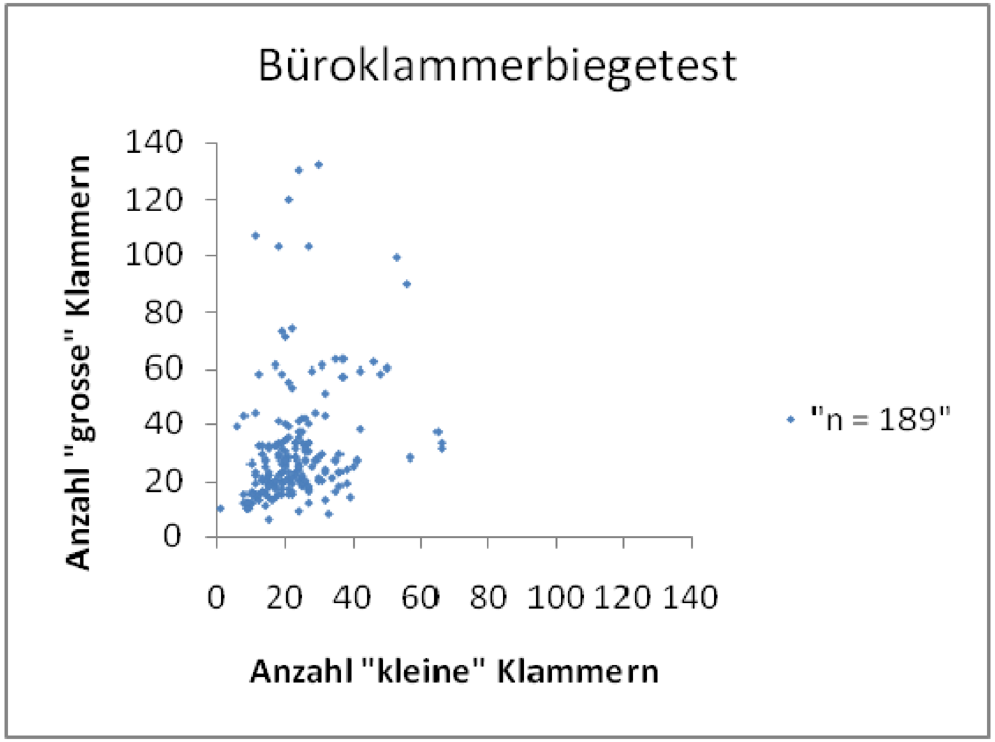
Das zweidimensionale Streudiagramm



Datenbeschreibung

- Beschreibung von paarweise beobachteten Eigenschaften

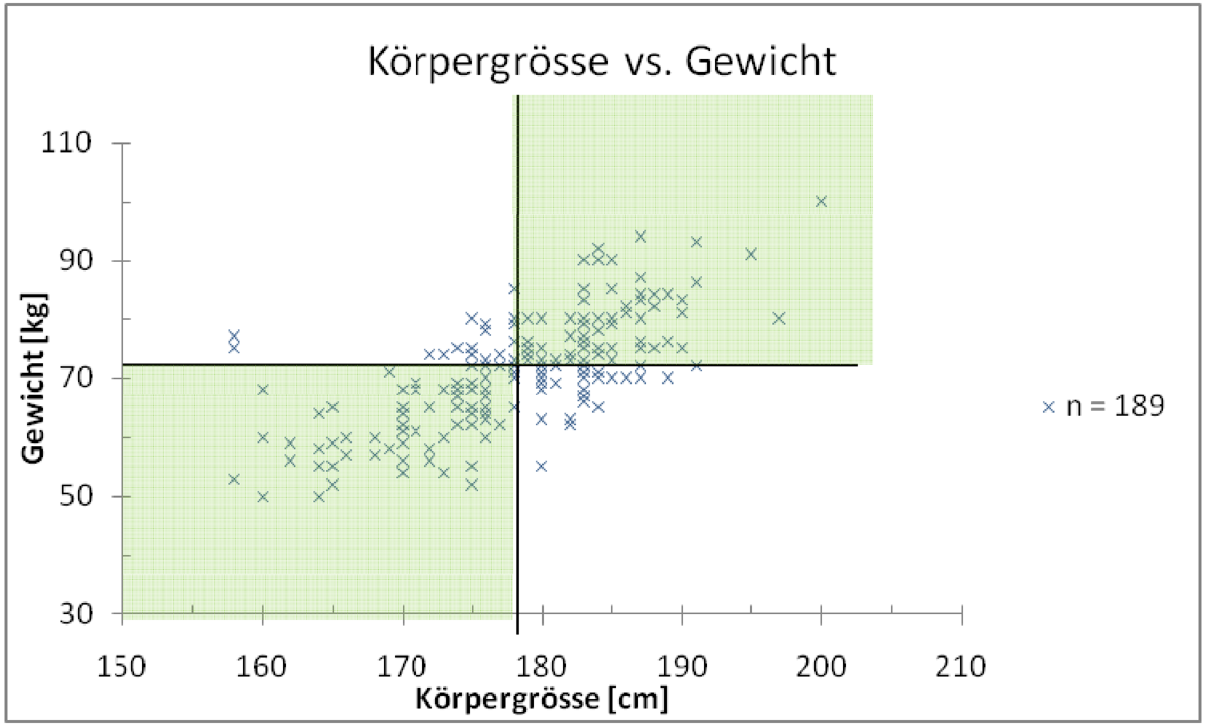
Das zweidimensionale Streudiagramm



Datenbeschreibung

- Beschreibung von paarweise beobachteten Eigenschaften

- Die Kovarianz: $$s_{XY} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})\cdot(y_i - \bar{y})$$



$x \triangleq$ Körpergrösse

$\bar{x} = 178.3$ cm

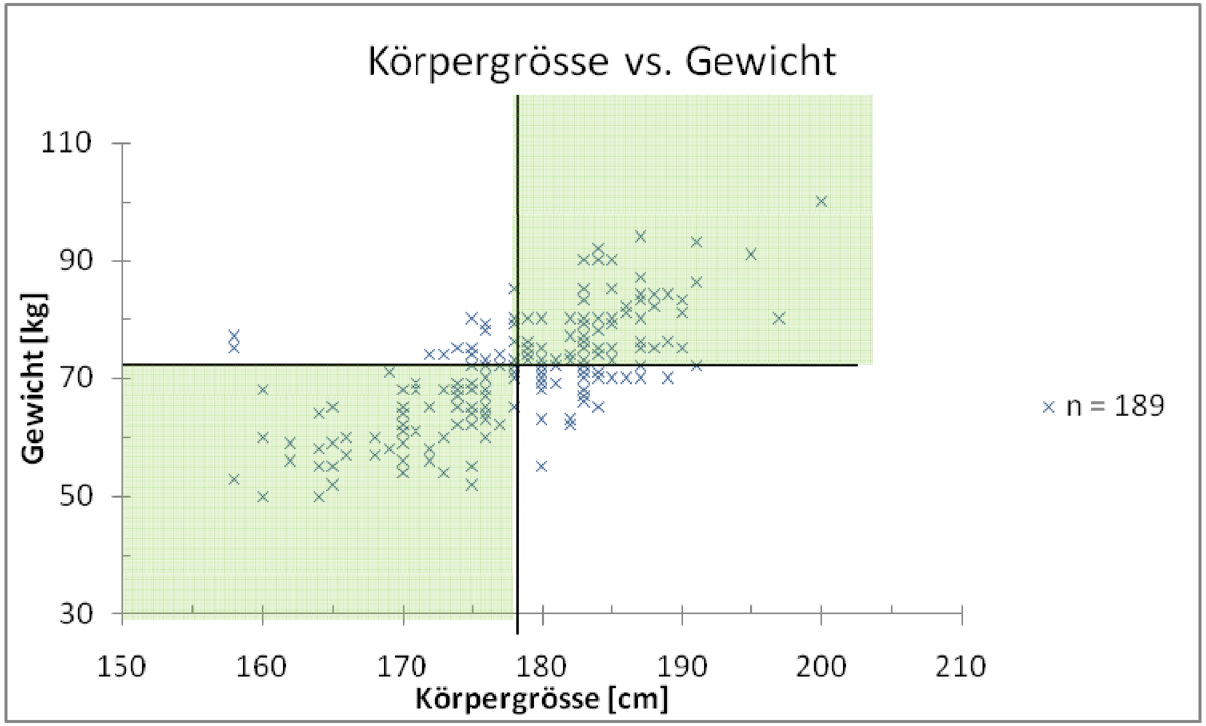
$y \triangleq$ Gewicht

$\bar{y} = 71.2$ kg

Datenbeschreibung

- Beschreibung von paarweise beobachteten Eigenschaften

- Die Kovarianz: $s_{XY} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})\cdot(y_i - \overline{y}) = 50.8$



$x \triangleq$ Körpergrösse

$\overline{x} = 178.3$ cm

$y \triangleq$ Gewicht

$\overline{y} = 71.2$ kg

Datenbeschreibung

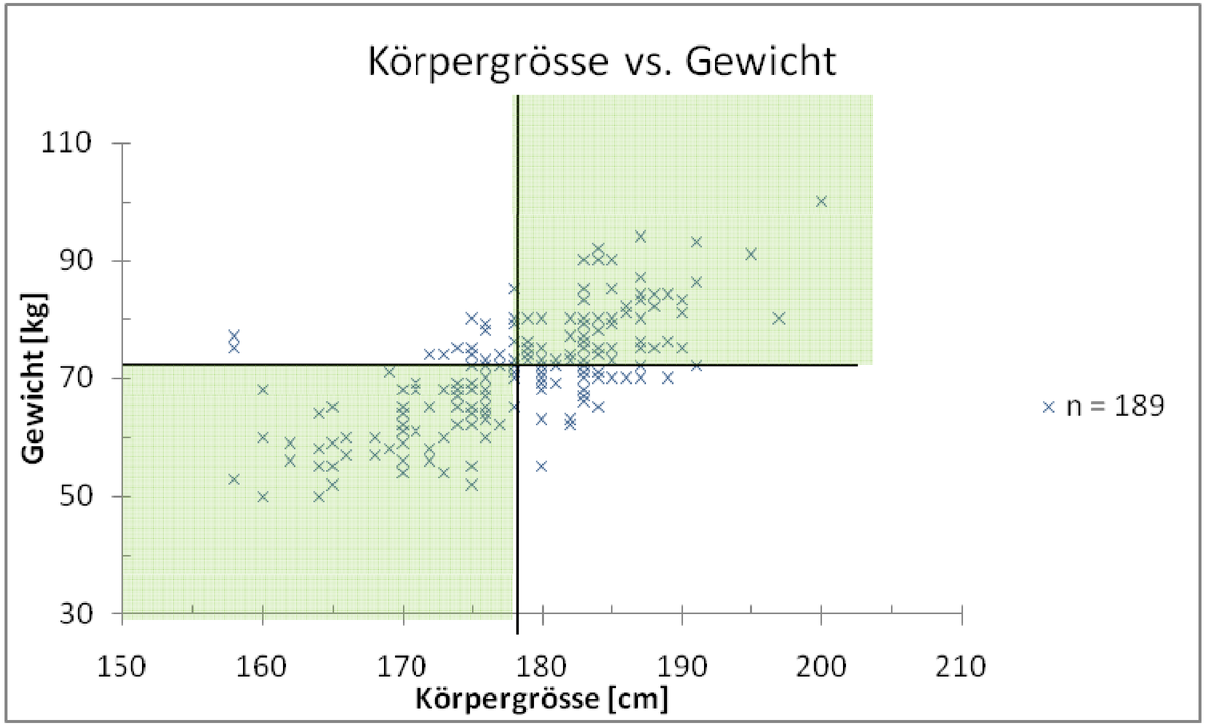
- Beschreibung von paarweise beobachteten Eigenschaften
- Die Kovarianz: $s_{XY} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})\cdot(y_i - \overline{y})$
- Der Korrelationskoeffizient:

$$r_{XY} = \frac{1}{n}\frac{\sum_{i=1}^{n}(x_i - \overline{x})\cdot(y_i - \overline{y})}{s_X \cdot s_Y}$$

ist limitiert auf das Interval $[-1,1]$

Datenbeschreibung

- Beschreibung von paarweise beobachteten Eigenschaften
- Der Korrelationskoeffizient: $r_{XY} = \frac{1}{n} \frac{\sum_{i=1}^{n} (x_i - \overline{x}) \cdot (y_i - \overline{y})}{s_X \cdot s_Y} = 0.697$



$x \triangleq$ Körpergrösse

$\overline{x} = 178.3$ cm

$y \triangleq$ Gewicht

$\overline{y} = 71.2$ kg

Nummerische Zusammenfassungen

Mittelwerte:	
Arithmetisches Mittel:	Schwerpunkt der Stichprobe
Median:	mittlerer Wert einer Stichprobe
Modalwert:	am häufigsten vorkommender Wert
Streuungsmasse:	
Varianz / Standardabweichung:	Verteilung um den Mittelwert
Variationskoeffizient :	Variabilität relativ zum Mittelwert
Andere Masse:	
- Schiefekoeffizient:	Schiefe relativ zum Mittelwert
- Kurtosis:	Wölbung um den Mittelwert
Masse für Korrelation:	
- Kovarianz:	Tendenz für paarweise beobachtete Eigenschaften
- Korrelations-koeffizient :	Normalisierter Koeffizient zwischen -1 und +1

Weitere graphische Darstellungsformen

- Histogramm Teil II
- Quantile Plots
- Tukey Boxplots

Histogramm

- Prinzip:
  - Aufteilung der Stichprobe in *k* Grössenklassen
  - Auftragen der Häufigkeit je Klasse

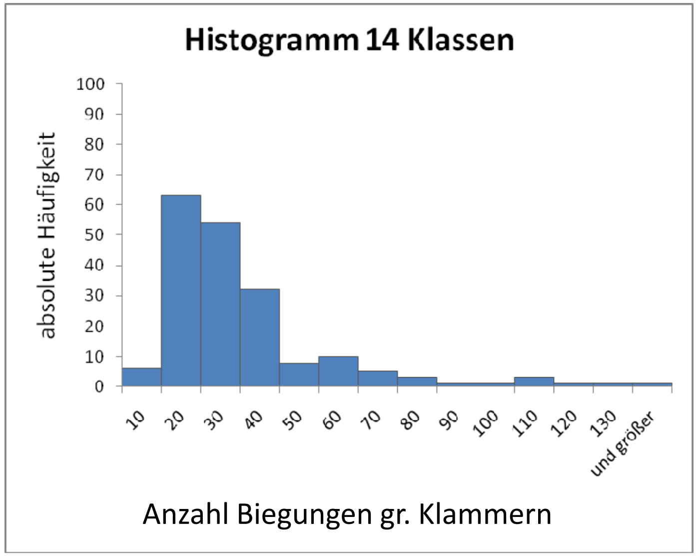
- Beispiel: Ihre Büroklammerdaten vom letzten Mal
  „grosse“ Klammern, Stichprobenumfang n = 190,
  Maximalwert 132, Minimalwert 6.

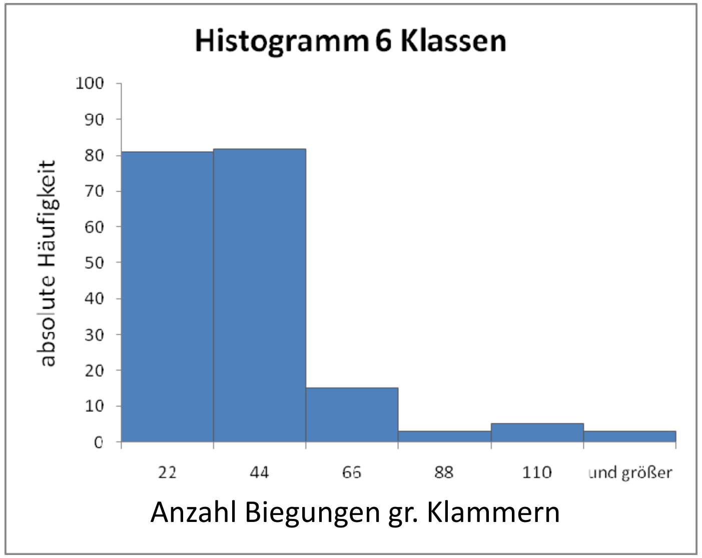
  Einteilung in 14 Klassen; (0,10]; (10,20]; (20,30];… ; (130,140]

Histogramm

- Prinzip:
  - Aufteilung der Stichprobe in *k* Grössenklassen
  - Auftragen der Häufigkeit je Klasse

- Beispiel:





Aussage abhängig von der Anzahl Klassen !!!!

Histogramm

- Prinzip:
  - Aufteilung der Stichprobe in *k* Grössenklassen
  - Auftragen der Häufigkeit je Klasse
  - Faustregel für die Anzahl Klassen: $k = 1 + 3.3\log(n)$

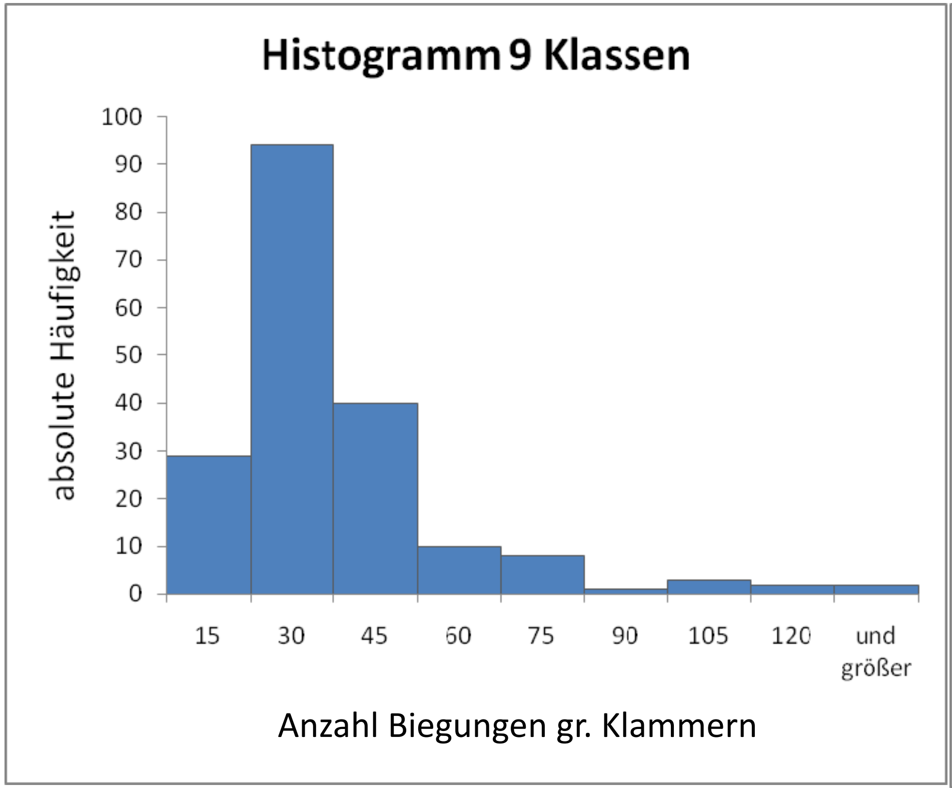
- Beispiel: Büroklammerdaten „grosse" Klammern,
  Stichprobenumfang n = 190, Wertebereich [6, 132]

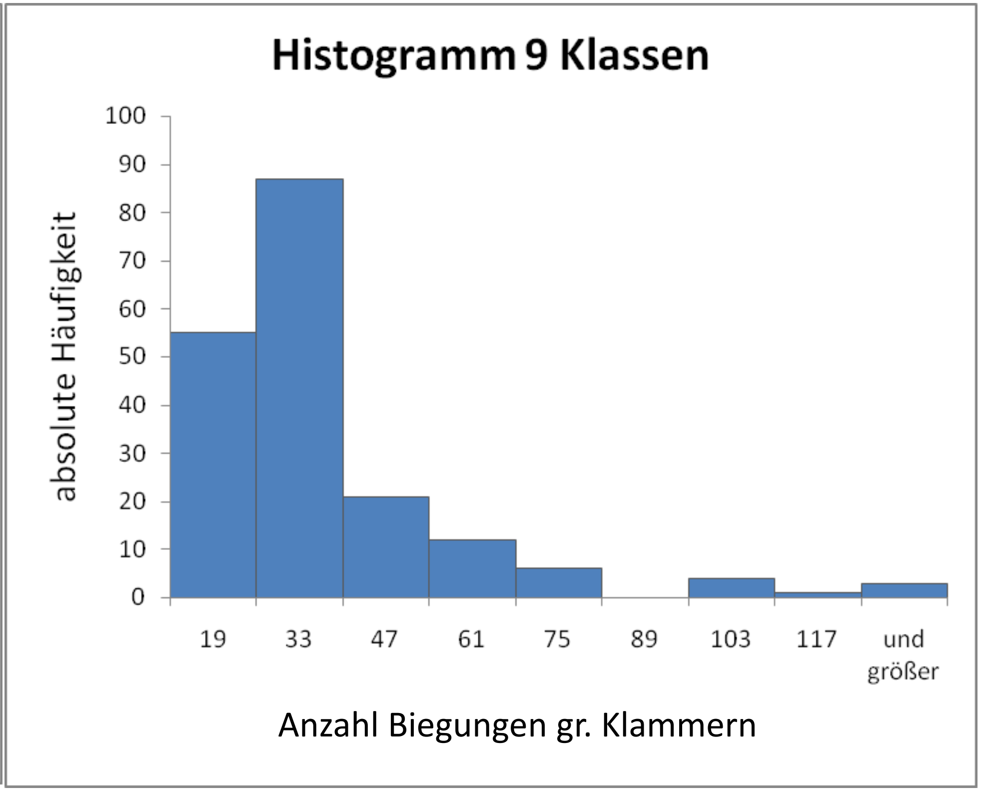
$$k = 1 + 3.3\log(190) = 8.52 \cong 9 \text{ Klassen}$$

(0,15]; (15,30]; (30,45];… ; (120,135]

oder (5,19]; (19,33]; (36,50];… ; (117,131] ?

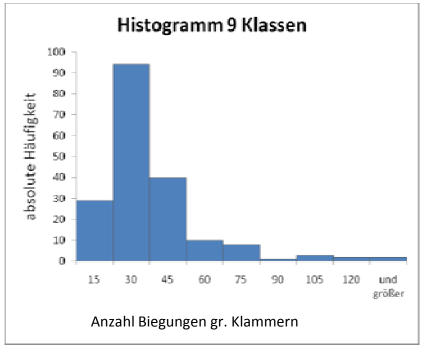
Histogramm

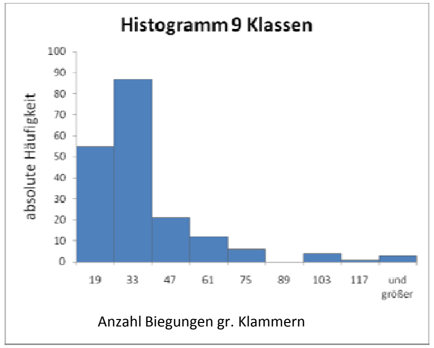

Histogramm 9 Klassen
absolute Häufigkeit
100
90
80
70
60
50
40
30
20
10
0
15
30
45
60
75
90
105
120
und größer
Anzahl Biegungen gr. Klammern

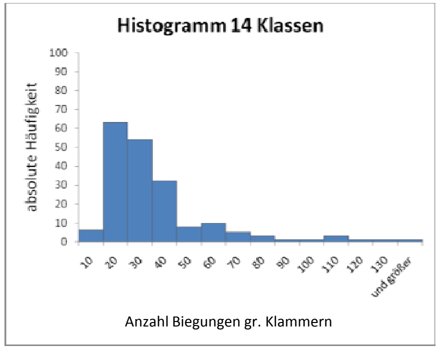

Histogramm 9 Klassen
absolute Häufigkeit
100
90
80
70
60
50
40
30
20
10
0
19
33
47
61
75
89
103
117
und größer
Anzahl Biegungen gr. Klammern

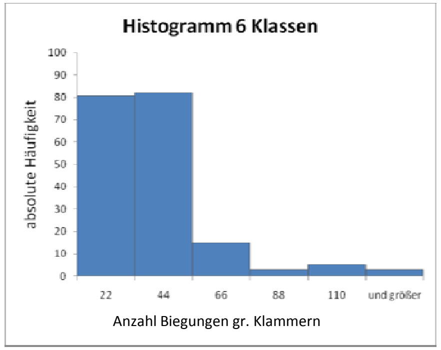
Histogramm

- Die Form des Histogramms hängt ab von
  - der Anzahl Klassen.
  - der Wahl des Startpunktes.



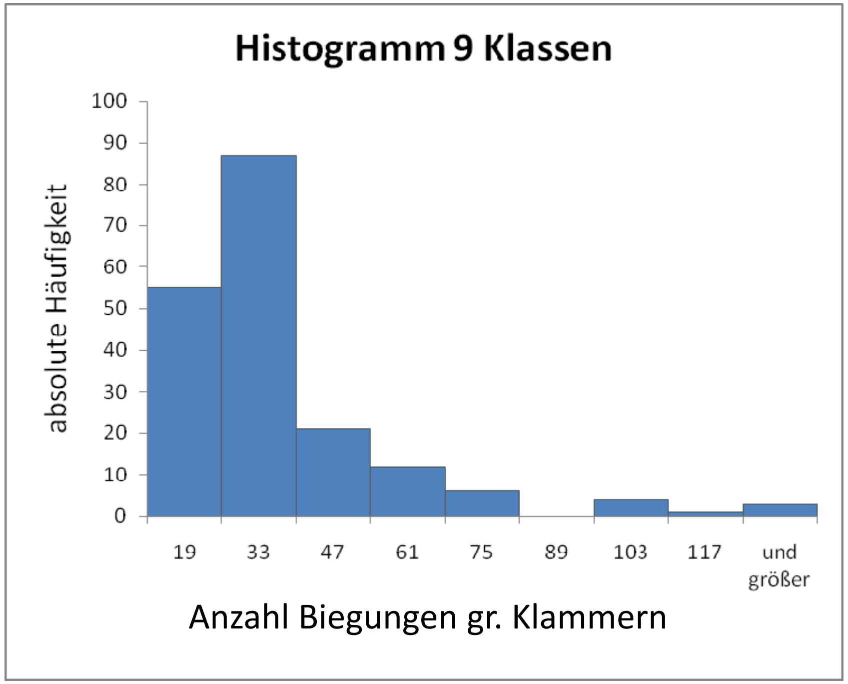


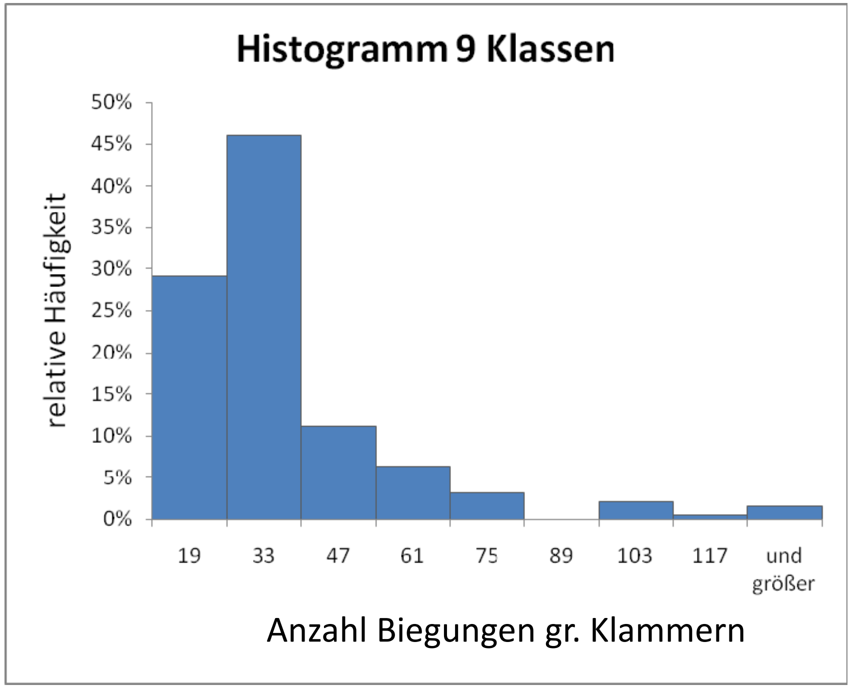




Histogramm

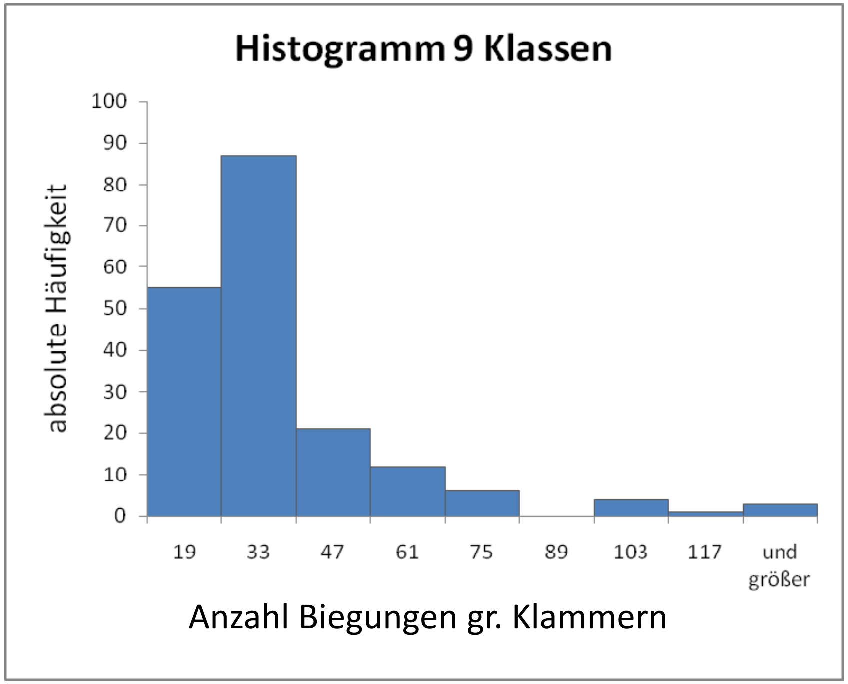
- Bisher betrachteten wir die absolute Häufigkeit.
- In der Regel wird die Häufigkeit relativ, also normiert betrachtet.

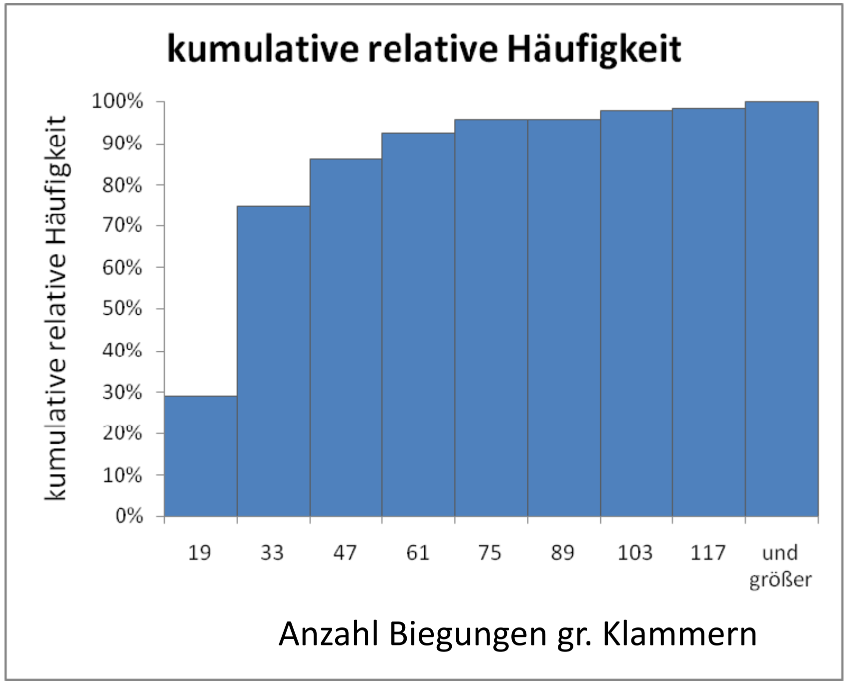




Histogramm

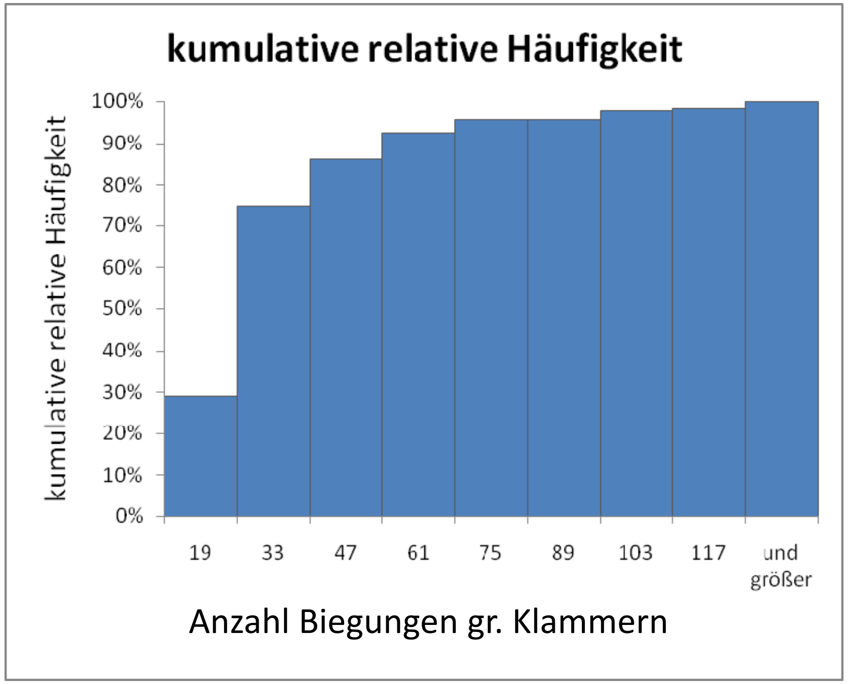
- Eine Spielart des Histogramms ist das kumulative Häufigkeits-diagramm.

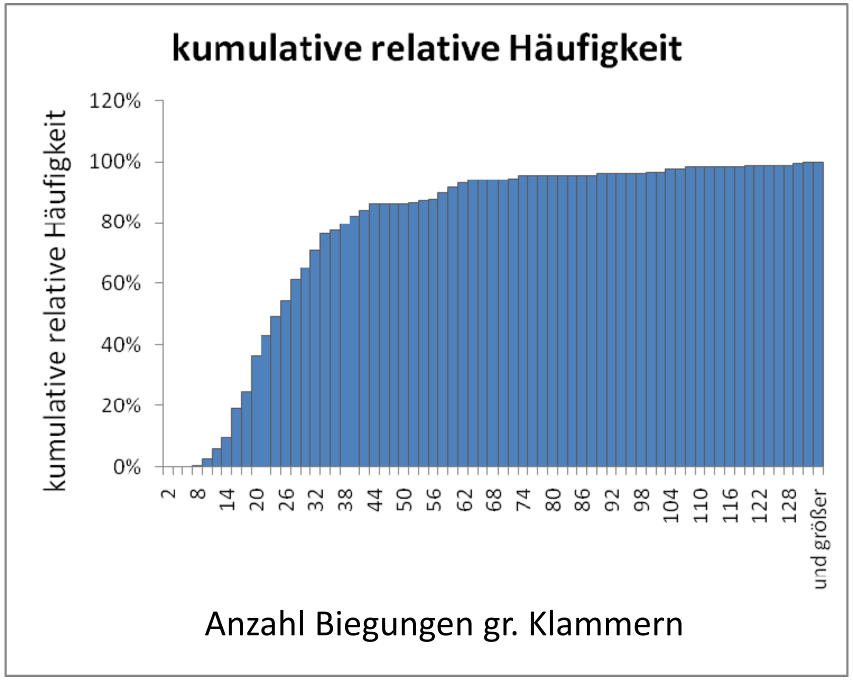




Histogramm

- Eine Spielart des Histogramms ist das kumulative Häufigkeits-diagramm.
- Hier kann die Klasseneinteilung beliebig klein sein!





Weitere graphische Darstellungsformen

- Histogramm Teil II.
- Quantile Plots
- Tukey Boxplots

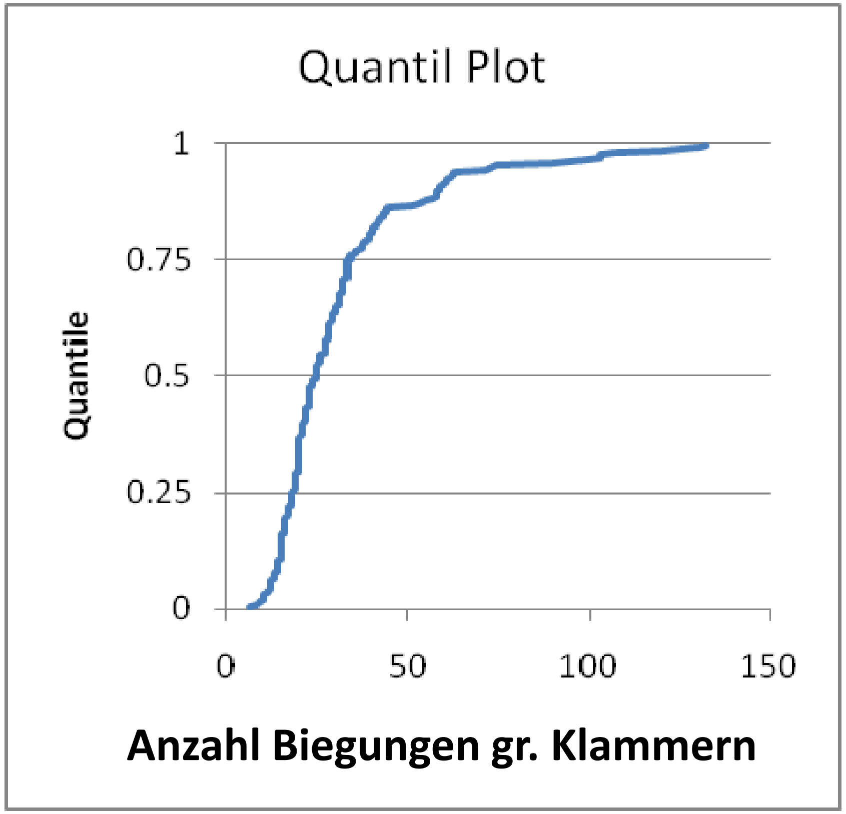
Quantil - Plot

- Definition :
  - Die Q-Quantile korrespondiert mit dem Wert der Stichprobe, welcher mit dem Wert 100% - Q x 100% überschritten wird.
  - D.h. zum Beispiel: das 0.75-Quantil wird von 100% - 0.75 x 100% = 25% der Daten überschritten.
  - Die Quantile werden von der geordneten Stichprobe berechnet: $x_1^o \leq x_2^o \leq \ldots \leq x_n^o$

$$Q_i = \frac{i}{1+n}$$

Quantil - Plot

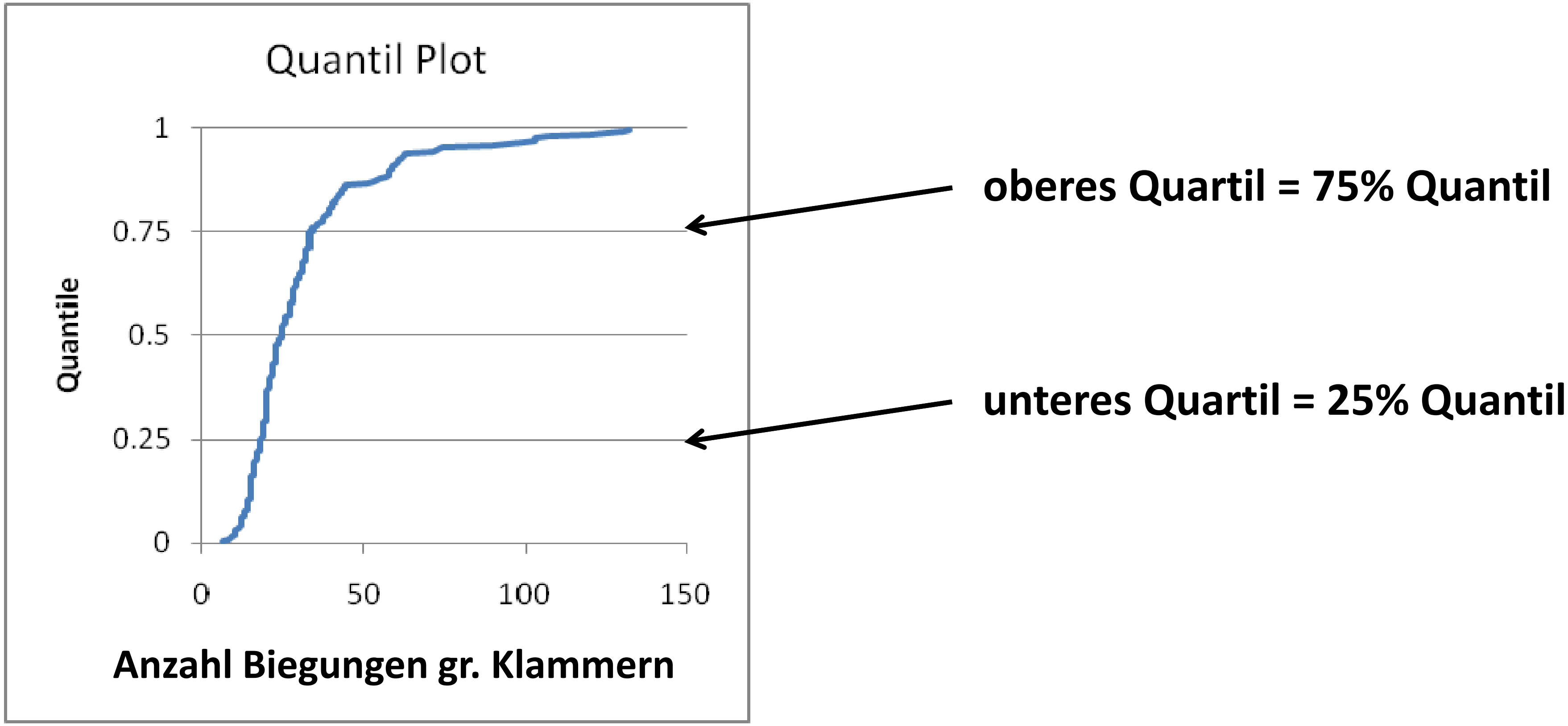
- Quantile-Plots werden durch Auftragen der Daten und dem Quantilwert gebildet.



i	i/n+1	x_i
1	0.005236	6
2	0.010471	8
3	0.015707	9
4	0.020942	10
5	0.026178	10
6	0.031414	10
7	0.036649	11
8	0.041885	12
9	0.04712	12
10	0.052356	12
11	0.057592	12
12	0.062827	12
13	0.068063	13
14	0.073298	13
15	0.078534	13
16	0.08377	14
17	0.089005	14

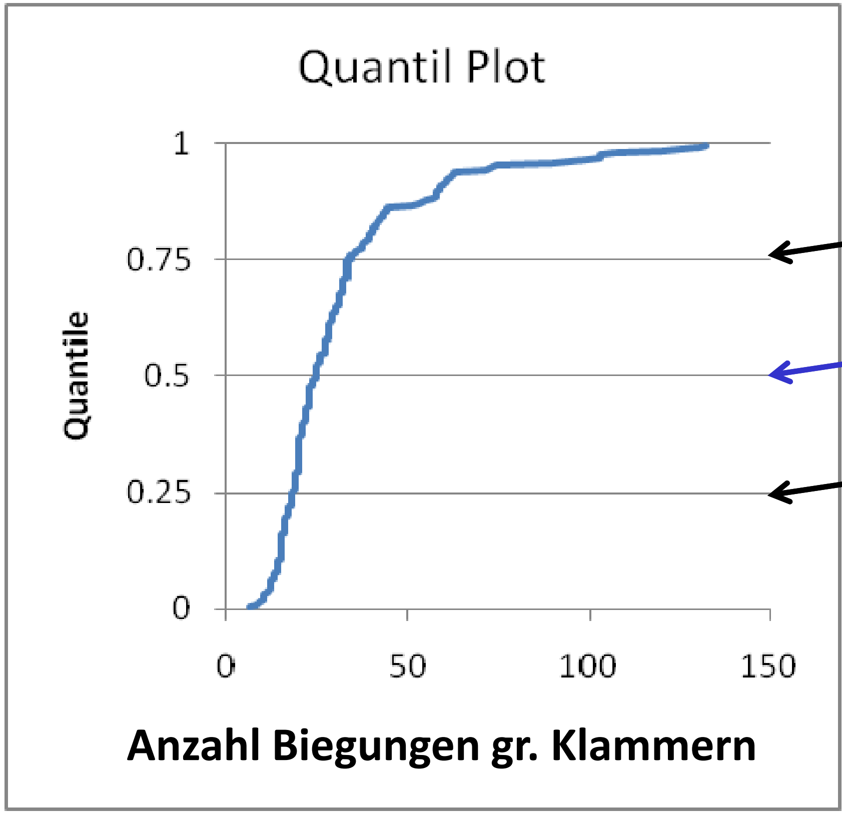
Quantil - Plot

- Quantile-Plots werden durch Auftragen der Daten und dem Quantilwert gebildet.



Quantil - Plot

- Quantile-Plots werden durch Auftragen der Daten und dem Quantilwert gebildet.



oberes Quartil = 75% Quantil

Und was ist das ??

unteres Quartil = 25% Quantil

Median

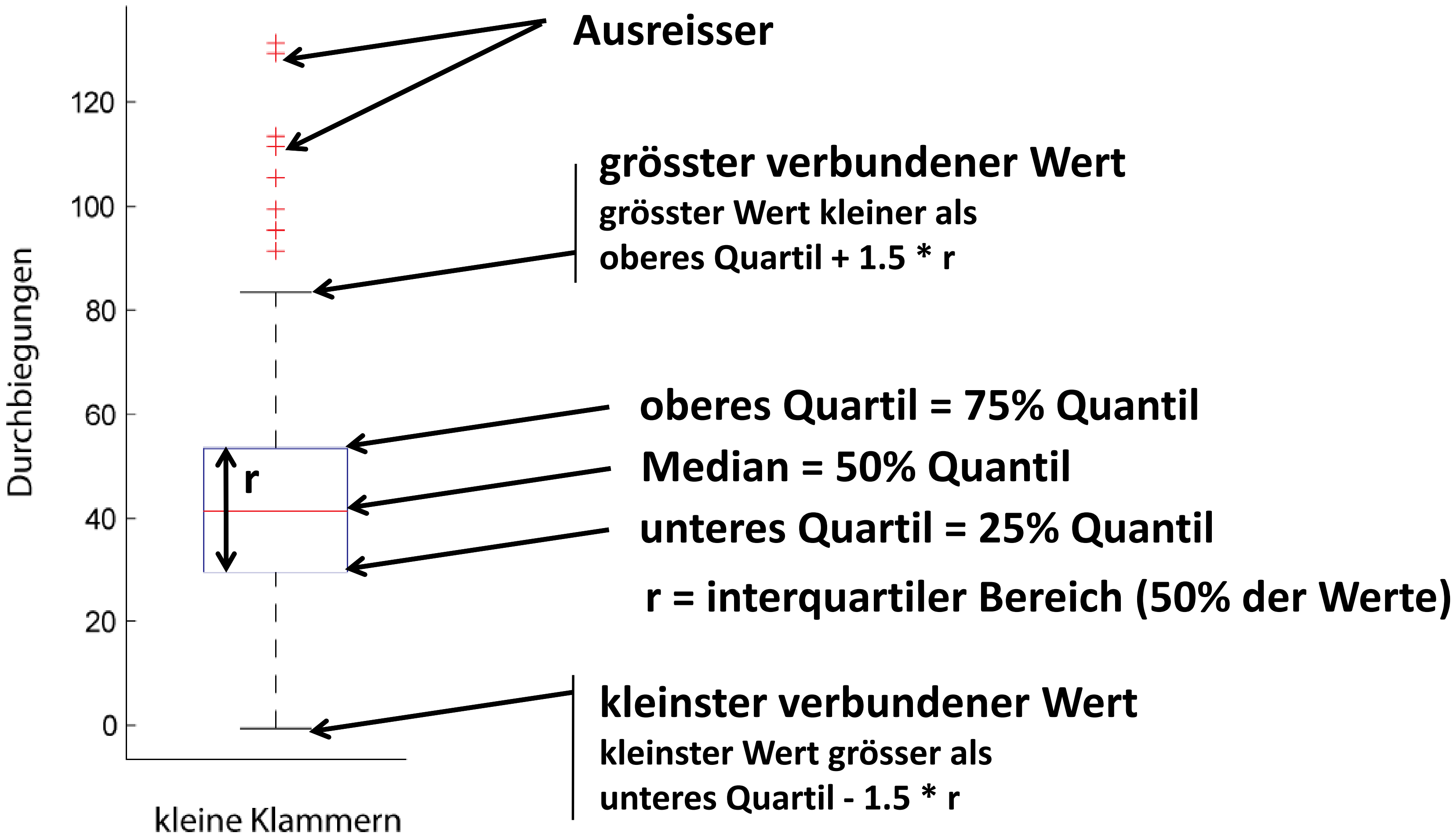
Mittelwert

Weiss nicht…

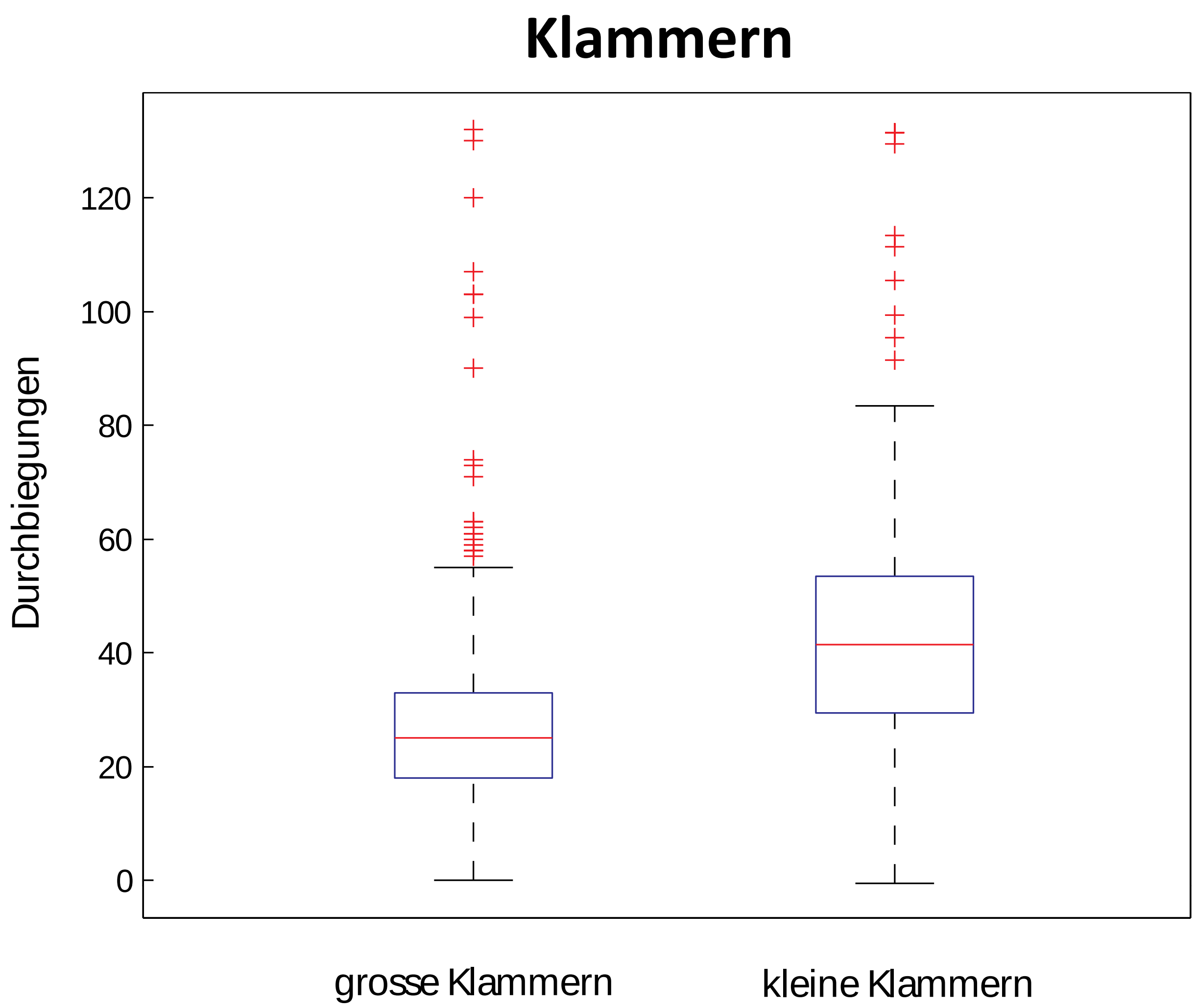
Tukey Boxplot

- Der Tukey Boxplot illustriert:
  - Median
  - untere und obere Quartilwerte
  - Streubreite
  - Ausreisser

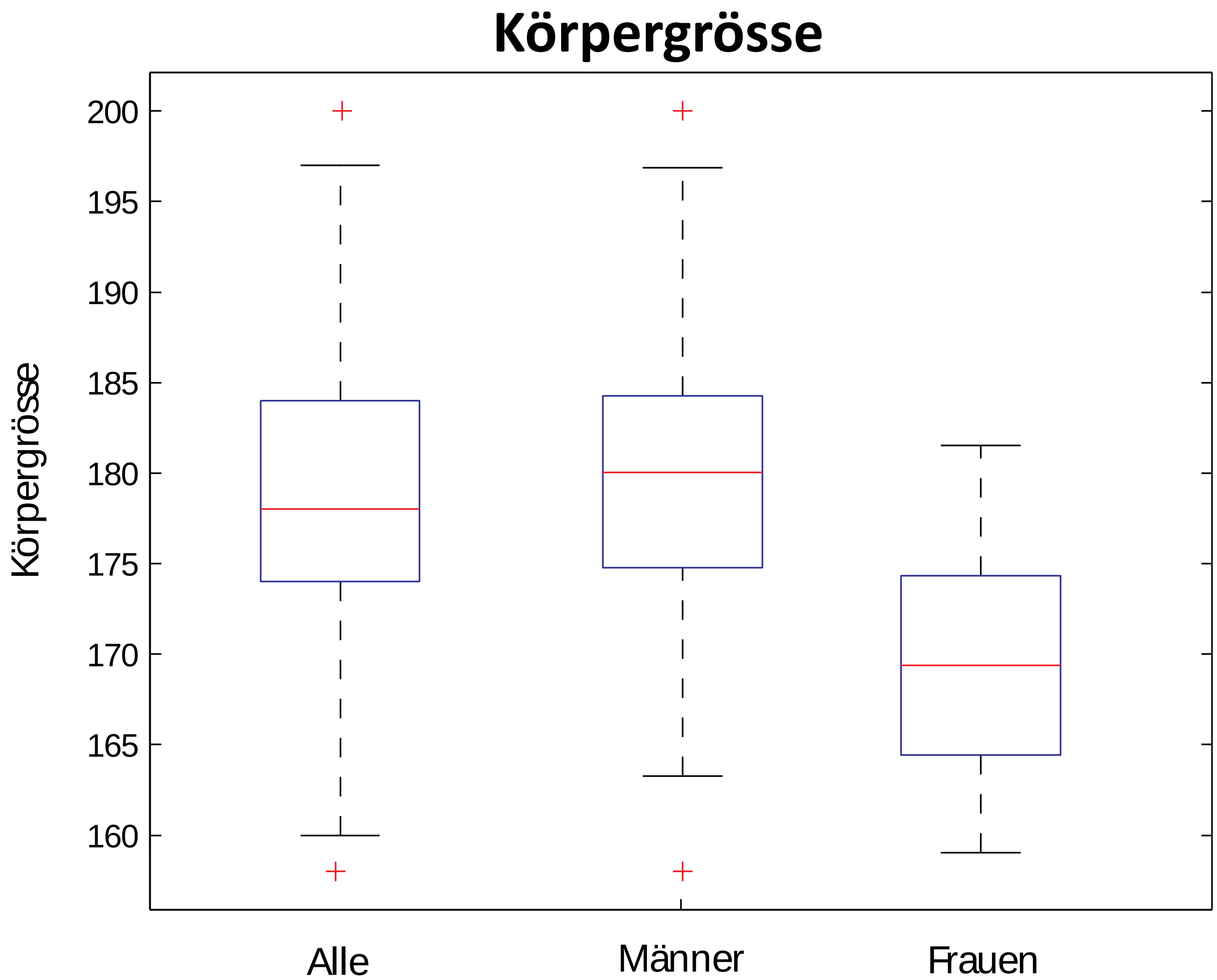
Tukey Boxplot



Tukey Boxplot

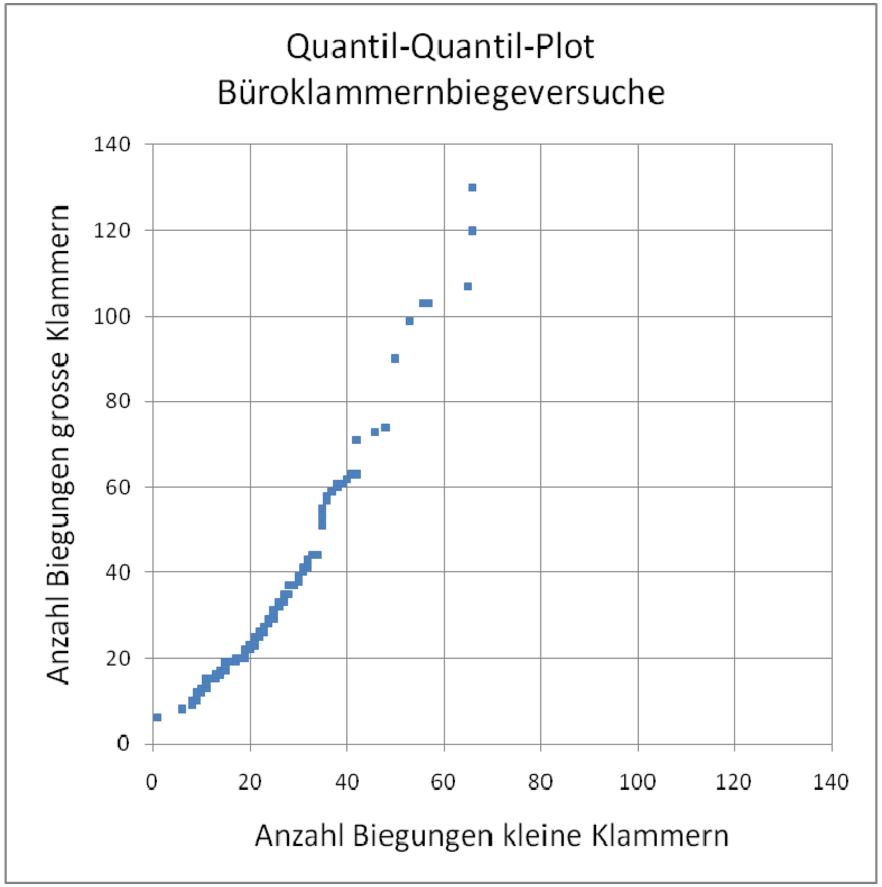


Tukey Boxplot



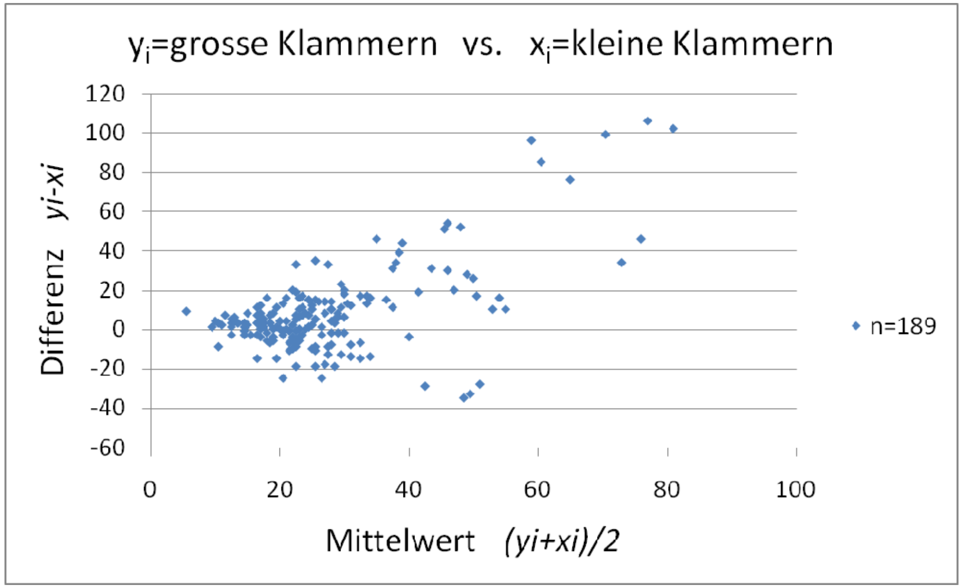
Q-Q Plots

- Q-Q plots dienen zur Darstellung und Vergleich von 2 Datenreihen.
- Datenpunkte der beiden Datenreihen mit demselben Quantilwert werden aufgetragen.



Mittel-über-Differenz Plots

- Mittel-über-Differenz Plots dienen zur Darstellung und dem Vergleich von zwei Datenreihen.
- Das Mittel $(y_i + x_i)/2$ wird über die Differenz $y_i - x_i$ aufgetragen.



Zusammenfassung Graphische Darstellung

Ein-dimensionales Streudiagramm	Veranschaulicht den Bereich und die Verteilung von Datenreihen entlang einer Achse, und zeigt Symmetrie.
Zwei-dimensionales Streudiagramm	Veranschaulicht den paarweisen Zusammenhang von Daten.
Histogramm	Stellt die Verteilung von Daten über einem Bereich von Datenreihen dar, zeigt Modalwert und Symmetrie.
Quantile Plot	Stellt Median, Verteilung und Symmetrie dar.
Tukey – Boxplot	Stellt Median, obere/untere Quartile, Symmetrie und Verteilung dar.
Q-Q Plot	Vergleicht zwei Datenreihen, relatives Bild.
Mittel-über-Differenz Plot	Vergleicht zwei Datenreihen, relatives Bild.