2. Beschreibende Statistik

2.1. Grundlagen

2.1.1. Grundbegriffe

Deskriptive Statistik (Beschreibende Statistik):
- Einordnung und graphische Darstellung von Daten

Induktive Statistik (Schließende Statistik):
- Schätzen von Kenngrößen einer Datenmenge aus Stichproben

Wahrscheinlichkeitstheorie:
- Modellierung von Zufallsprozessen mit Hilfe von Wahrscheinlichkeitsverteilungen und Zufallsvariablen

Statistische Elemente:
- Träger von Informationen.

Grundgesamtheit (auch Stichprobenraum oder Merkmalsraum, engl. sample space):
- die Menge aller Objekte, über die eine Aussage getroffen werden soll.

Stichprobe (engl. sample):
- Tatsächlich untersuchte Teilmenge der Grundgesamtheit
  Umfang der Stichprobe: Anzahl statistischer Elemente in der Stichprobe.

Merkmale (engl. variables oder measurements):
- Größen, die an den statistischen Elementen beobachtet oder gemessen werden können.

Ausprägungen (engl. values):
- Mögliche Werte eines Merkmals.

Univariate Daten:
- Daten mit genau einem beobachteten oder gemessenen Merkmal

Multivariate Daten:
- Daten mit mehreren gleichzeitig beobachteten oder gemessenen Merkmalen

Beispiel: Serverauslastung

• Statistische Elemente:	Datenpakete
• Grundgesamtheit:	Anfragen an einen Datenbankserver
• Stichprobe:	alle Anfragen am 28.09.2012
• Merkmale:	Paketgröße, Antwortzeit
• Ausprägungen:	z.B. 752Bit, 12ms

2.1.2. Merkmalstypen

Definition 2.1 (Merkmalstypen)
Ein Merkmal heißt
- ***diskret***, wenn es endlich viele oder abzählbar unendlich viele Ausprägungen besitzt;
- ***stetig***, wenn es überabzählbar unendlich viele Ausprägungen besitzt.

Beispiel: Merkmalstypen

• Diskrete Merkmale:	Geschlecht, Alter in Jahren, Partei, Einwohnerzahl
• Stetige Merkmale:	Körpergröße, Zeitpunkt
• *Quasi-stetige* Merkmale:	Geldbeträge in Euro und Cent (sehr feine Abstufung)

Definition 2.2 (Skalenniveaus von Merkmalen)
Ein Merkmal heißt

- ***nominal skaliert*,** wenn seine Ausprägungen Namen sind, für die es keine natürliche Rangfolge gibt (die also gleichberechtigt sind);
- ***ordinal skaliert***, wenn seine Ausprägungen **Namen** oder **Zahlenwerte** sind, die auf natürliche Weise **geordnet** werden können (im Sinne von „größer“, „besser“ o.ä.), deren Abstände aber nicht interpretierbar sind;
- ***kardinal skaliert*** (oder ***metrisch***), wenn seine Ausprägungen Zahlenwerte sind, deren Abstände interpretiert werden können. Kardinal skalierte Merkmale werden weiter unterschieden in
  - ***intervallskaliert***, wenn die Zahlenwerte einen **willkürlichen Nullpunkt** besitzen;
  - ***verhältnisskaliert***, wenn die Zahlenwerte einen **absoluten Nullpunkt** besitzen.

Nominal oder ordinal skalierte Merkmale werden auch *kategorial* genannt.

Skala	Zusätzliche mathematische Relationen/Operationen	Zusätzliche messbare Eigenschaften	Beispiel
Nominal	$=, \neq$	Häufigkeit (Gruppierung)	Postleitzahl
Ordinal	$<, >$	Rangfolge	Tabellenplatz (Bundesliga)
Intervall	$+, -$ („Merkmal + Merkmalsdifferenz“)	Abstand	Datum
Verhältnis	$\cdot, :$ („Faktor $\cdot$ Merkmal“)	Natürlicher Nullpunkt	Alter (in Jahren)

Beispiel: Merkmalstypen
- Nominal skalierte Merkmale: Geschlecht, Partei
- Ordinal skalierte Merkmale: Dienstgrad, Tarifstufe, Klausurnote
- Intervallskalierte Merkmale: Zeitpunkt, Temperatur in Grad Celsius oder Fahrenheit
- Verhältnisskalierte Merkmale: Körpergröße, Zeitdauer, Temperatur in Kelvin

Definition 2.3 (Merkmalstypen)
Ein Merkmal heißt

- ***qualitativ***, wenn seine Ausprägungen eine ***Qualität*** wiedergeben
  (dies ist für nominal und ordinär skalierte Merkmale der Fall);
- ***quantitativ***, wenn seine Ausprägungen ein ***Ausmaß*** wiedergeben
  (dies ist für kardinal skalierte Merkmale der Fall).

2.2. Häufigkeitsverteilung

2.2.1. Ordnungsstatistik

Urliste (Rohdaten):

- Auflistung der beobachteten oder gemessenen Merkmalsausprägungen der statistischen Elemente einer Stichprobe in der Reihenfolge ihrer Erhebung

Ordnungsstatistik:

- Auflistung der ihrer Größe nach geordneten Merkmalsausprägungen

Beispiel 2.4 Anzahl der Studiensemester von 60 Bachelorabsolvent*innen

(Urliste)

9	8	7	7	8	10	6	8	8	7	9	7
10	8	8	9	7	8	9	10	6	10	8	9
9	7	7	8	8	7	8	7	7	8	8	8
10	7	10	9	8	6	9	7	8	7	9	12
9	8	9	6	12	8	7	8	9	7	8	7

(Ordnungsstatistik)

6	7	7	7	8	8	8	8	9	9	9	10
6	7	7	7	8	8	8	8	9	9	9	10
6	7	7	7	8	8	8	8	9	9	10	10
6	7	7	7	8	8	8	8	9	9	10	12
7	7	7	7	8	8	8	8	9	9	10	12

2.2.2. Absolute und relative Häufigkeit

Definition 2.6 (Absolute Häufigkeit)

- Sei X ein Merkmal der statistischen Elemente einer Stichprobe vom Umfang n und seien x_i, $i = 1,\ldots,n$, die Merkmalsausprägungen der Urliste. $x_1, \ldots, x_n$ 66~~8~~8
- Sei $m \leq n$ die Anzahl unterschiedlicher Ausprägungen des Merkmals X und seien a_j, $j = 1,\ldots,m$ die (geordneten) unterschiedlichen Ausprägungen. $a_1 \ldots a_7$
- Die ***absolute Häufigkeit*** (engl. *absolute frequency* oder *frequency*) n_j der Ausprägung a_j ist die Anzahl statistischer Elemente mit $x_i = a_j$, $8 \simeq 2$

$$n_j = |\{i : x_i = a_j,\ 1 \leq i \leq n\}|, \quad j = 1,\ldots,m.$$

Definition 2.7 (Relative Häufigkeit)
Seien die gleichen Voraussetzungen gegeben wie in Def. 2.6.

Die ***relative Häufigkeit*** (engl. *relative frequency* oder *empirical probability*) f_j der Ausprägung a_j ist definiert als

$$f_j = \frac{n_j}{n}, \quad j = 1,\ldots,m.$$

Beispiel: Absolute und relative Häufigkeit

a_j	n_j	f_j
6	4	0.07
7	16	0.27
8	20	0.33
9	12	0.20
10	6	0.10
12	2	0.03
Σ	60	1.00

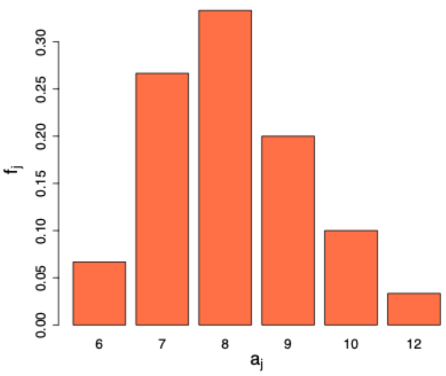
Folgerung 2.8 (Absolute und relative Häufigkeit)
Aus den Definitionen der Häufigkeiten folgt unmittelbar

(i) für die absoluten Häufigkeiten

$$\forall j \in \{1,\ldots,m\}: \quad 0 \leq n_j \leq n \quad \text{und insgesamt} \quad \sum_{j=1}^{m} n_j = n,$$

(ii) für die relativen Häufigkeiten

$$\forall j \in \{1,\ldots,m\}: \quad 0 \leq f_j \leq 1 \quad \text{und insgesamt} \quad \sum_{j=1}^{m} f_j = 1.$$



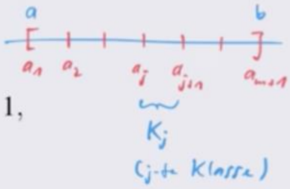
2.2.3. Klassierte Häufigkeitsverteilungen

Definition 2.9 (Absolute Häufigkeit einer Klasse)

Sei X ein Merkmal mit Ausprägungen x_i, $i = 1, \dots, n$, im Intervall $[a, b]$, sei $m > 1$ die Anzahl der Klassen und seien

$$a = a_1 < a_2 < \dots < a_m < a_{m+1} = b$$

eine Folge reeller Zahlen. Dann heißt das Intervall



$$K_j = \begin{cases} [a_1, a_2] & \text{falls} \quad j = 1, \\ (a_j, a_{j+1}] & \text{sonst} \end{cases}$$

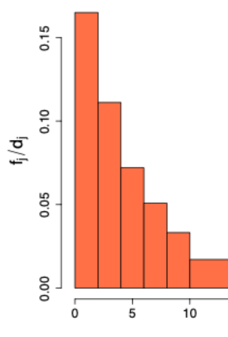
die j-te Klasse (engl. *category* oder *bin*). Die absolute Häufigkeit n_j einer Klasse K_j ist definiert durch

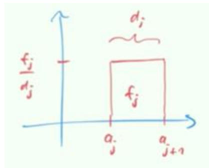
$$n_j = |\{i : x_i \in K_j,\ 1 \le i \le n\}|, \quad j = 1, \dots, m.$$

Beispiel 2.11 (Klassierte Häufigkeiten von Telefondaten)
Dauer von 5000 Telefongesprächen in Minuten im Intervall [0,30)

K_j	d_j	m_j	n_j	f_j
$[0,2]$	2	1.0	1650	0.330
$(2,4]$	2	3.0	1111	0.222
$(4,6]$	2	5.0	720	0.144
$(6,8]$	2	7.0	508	0.102
$(8,10]$	2	9.0	332	0.066
$(10,15]$	5	12.5	427	0.085
$(15,30]$	15	22.5	252	0.050
Σ	30		5000	0.999

Beispiel : Histogramm der Telefondaten





Definition 2.10 (Relative Häufigkeit einer Klasse)
Seien die gleichen Voraussetzungen gegeben wie in Def. 2.9.

Die ***relative Häufigkeit f_j einer Klasse K_j*** ist definiert durch

$$f_j = \frac{n_j}{n}.$$

Außerdem werden definiert

$$\textbf{\textit{die Klassenbreite}} \quad d_j = a_{j+1} - a_j,$$

$$\textbf{\textit{die Klassenmitte}} \quad m_j = \frac{a_j + a_{j+1}}{2}.$$

2.2.4. Kumulierte Häufigkeiten

Definition 2.12 (Kumulierte Häufigkeiten)
Gegeben seien die absoluten Häufigkeiten n_j und die relativen Häufigkeiten f_j der Merkmalsausprägungen a_j bzw. der Klassen K_j, $j = 1, \dots, m$.

Dann ist die ***kumulierte absolute Häufigkeit*** N_j (engl. *cumulative frequency*) definiert durch

$$N_j = \sum_{k=1}^{j} n_k, \quad j = 1, \dots, m.$$

Die ***kumulierte relative Häufigkeit*** F_j ist definiert durch

$$F_j = \sum_{k=1}^{j} f_k, \quad j = 1, \dots, m.$$

Folgerung 2.13 (Kumulierte Häufigkeiten)
Aus den Definitionen der kumulierten absoluten und relativen Häufigkeiten folgt unmittelbar

$$\forall j \in \{1, \dots, m-1\}: \quad N_j \le N_{j+1}, \qquad N_m = n,$$
$$\forall j \in \{1, \dots, m-1\}: \quad F_j \le F_{j+1}, \qquad F_m = 1.$$

Beispiel: Kumulierte Häufigkeiten der Telefondaten

K_j	f_j	F_j
$[0,2]$	0.330	0.330
$(2,4]$	0.222	0.552
$(4,6]$	0.144	0.696
$(6,8]$	0.102	0.798
$(8,10]$	0.066	0.864
$(10,15]$	0.085	0.949
$(15,30]$	0.050	0.999

Frage:
Wie viel Prozent der Gespräche dauern unter 6 Minuten?
Antwort: F3 = 69.6%

2.3. Kenngrößen univariater Daten

2.3.1. Lageparameter

Arithmetisches Mittel

- der **durchschnittliche Wert** (oder Mittelwert) einer Stichprobe

Definition 2.14 (Arithmetisches Mittel)
Sei $(x_1, \dots, x_n)$ eine Stichprobe vom Umfang n, deren Ausprägungen $a_1, \dots, a_m$ mit den absoluten Häufigkeiten $n_1, \dots, n_m$ bzw. den relativen Häufigkeiten $f_1, \dots, f_m$ auftreten.

Dann ist das ***arithmetische Mittel*** (engl. *arithmetic mean* oder *mean*) $\bar{x}$ der Stichprobe definiert als

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\sum_{j=1}^{m} n_j a_j = \sum_{j=1}^{m} f_j a_j.$$

Beispiel 2.15 (Arithmetisches Mittel)
Das arithmetische Mittel der Stichprobe $(3,6,4,1,5,7,2,4,9)$ beträgt

$$\bar{x} = \frac{3+6+4+1+5+7+2+4+9}{9} = \frac{41}{9} \approx 4.56.$$

Beispiel 2.16 (Arithmetisches Mittel)
Das arithmetische Mittel der Stichprobe aus Bsp. 2.4 kann unter Nutzung der relativen Häufigkeiten f_j berechnet werden:

$$\begin{aligned}\bar{x} = \sum_{j=1}^{m} f_j a_j &= 0.07 \cdot 6 + 0.27 \cdot 7 + 0.33 \cdot 8 + 0.2 \cdot 9 + 0.1 \cdot 10 + 0.03 \cdot 12 \\ &\approx 8.11.\end{aligned}$$

geometrisches Mittel

- Für Werte, die multiplikativ (faktoren) verknüpft werden (z. B. Zinssätze)

Definition 2.17 (Geometrisches Mittel)
Sei $(x_1, \dots, x_n)$ eine Stichprobe vom Umfang n mit nicht-negativen Werten $x_i \geq 0$, $1 \leq i \leq n$.

Dann ist das ***geometrische Mittel*** (engl. *geometric mean*) $\bar{x}_{\text{geom}}$ der Stichprobe definiert als

$$\bar{x}_{\text{geom}} = \sqrt[n]{\prod_{i=1}^{n} x_i}.$$

Beispiel 2.18 (Geometrisches Mittel)
Ein Kapital K_0 werde im ersten Jahr mit einem Zinssatz von 3 % (Faktor $q_1 = 1.03$), im zweiten Jahr mit 7 % (Faktor $q_2 = 1.07$) angelegt. Nach zwei Jahren ist das Kapital angewachsen auf

$$K_2 = q_2 q_1 K_0 = 1.1021\, K_0.$$

Das geometrische Mittel der Faktoren beträgt

$$\bar{q}_{\text{geom}} = \sqrt[2]{q_1 q_2} \approx 1.0498,$$

das Kapital nach zwei Jahren kann damit berechnet werden als

$$K_2 = \bar{q}_{\text{geom}}^2 K_0 = 1.1021\, K_0.$$

Satz 2.19 (Arithmetisches und geometrisches Mittel)
Sei $(x_1, \dots, x_n)$ eine Stichprobe vom Umfang n.

Dann ist das geometrische Mittel stets kleiner oder gleich dem arithmetischen Mittel,

$$\bar{x}_{\text{geom}} \leq \bar{x}.$$

harmonisches Mittel

- Das selten benötigte harmonische Mittel wird hier nicht weiter behandelt.

Median (oder Zentralwert)

Der Median einer geordneten Stichprobe des Umfangs n ist also das mittlere Element.
Der Median ist eine robuste Kenngröße für den Durchschnitt, robust gegen Ausreisen (20 in dem Fall)

Definition 2.20 (Median)
Sei $(x_1, \ldots, x_n)$ eine geordnete Stichprobe vom Umfang n mit $x_1 \leq x_2 \leq \ldots \leq x_n$.

Dann ist der ***Median*** (oder ***Zentralwert***, engl. *median*) $\tilde{x}$ der Stichprobe definiert als

$$\tilde{x} = \begin{cases} x_{(n+1)/2} & \text{falls} \quad \frac{n}{2} \notin \mathbb{N} \quad (n \text{ ungerade}), \\ \frac{1}{2}(x_{n/2} + x_{n/2+1}) & \text{falls} \quad \frac{n}{2} \in \mathbb{N} \quad (n \text{ gerade}). \end{cases}$$

$$(6,6,6,6,6,6,7,7,7,8,8,8,9,9,20).$$

Das arithmetische Mittel und der Median ergeben sich als

$$\bar{x} = 7.93, \quad \tilde{x} = 7.$$

Ohne Berücksichtigung des Ausreißers $x_{15} = 20$ erhalten wir

$$\bar{x}' = 7.07, \quad \tilde{x}' = 7.$$

Modus (oder Modalwert) - Für nominal skalierte Merkmale

- Die Ausprägung die am häufigsten vorkommt
- Der Modalwert ist nur dann eine sinnvolle Kenngröße, wenn die zugehörige Ausprägung **deutlich häufiger** auftritt als alle anderen Ausprägungen.

Definition 2.23 (Modalwert)
Sei $(x_1, \ldots, x_n)$ eine Stichprobe vom Umfang n, deren Ausprägungen $a_1, \ldots, a_m$ mit den absoluten Häufigkeiten $n_1, \ldots, n_m$ bzw. den relativen Häufigkeiten $f_1, \ldots, f_m$ auftreten.

Dann ist der ***Modalwert*** (oder ***Modus***, engl. *mode*) $\bar{x}_{\text{mod}}$ definiert als diejenige Ausprägung, die am häufigsten auftritt.

$$\bar{x}_{\text{mod}} = \{a_j : n_j = \max(n_1, \ldots, n_m), 1 \leq j \leq m\}.$$

Beispiel 2.24 (Modalwert)
Gegeben sei eine Stichprobe von Beobachtungen der Haarfarbe einer Gruppe Studierender,

(braun, schwarz, blond, blond, rot, blond, braun, blond, blond).

Der Modalwert ergibt sich als $\bar{x}_{\text{mod}} = \text{blond}$.

p-Quantil und Quartile (Verallgemeinerung des Medians)

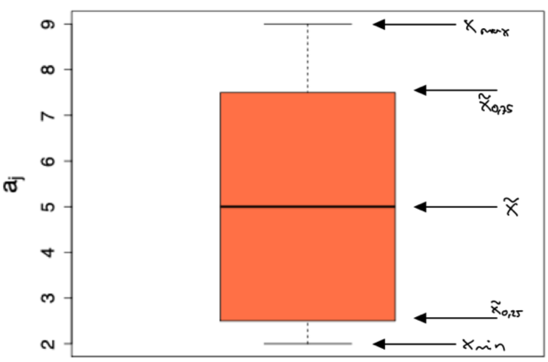
Quantile stellen keine Mittelwerte dar, sondern Niveaus, unterhalb derer bestimmte Anteile aller beobachteten Werte liegen. Gleichwohl zählen sie zu den Lageparametern.

Definition 2.25 (p-Quantile)
Sei $(x_1, \ldots, x_n)$ eine geordnete Stichprobe vom Umfang n mit $x_1 \leq x_2 \leq \ldots \leq x_n$. Sei $p \in \mathbb{R}$ mit $0 < p < 1$.

Dann ist das ***p-Quantil*** $\tilde{x}_p$ der Stichprobe definiert als

$$\tilde{x}_p = \begin{cases} x_{\lfloor np \rfloor + 1} & \text{falls} \quad np \notin \mathbb{N}, \\ \frac{1}{2}(x_{np} + x_{np+1}) & \text{falls} \quad np \in \mathbb{N}. \end{cases}$$

Das 0.5-Quantil entspricht dem Median; die 0.25- und 0.75-Quantile werden auch als ***untere und obere Quartile*** bezeichnet.



Folgerung 2.26 (p-Quantile)
Das p-Quantil $\tilde{x}_p$ ist so definiert, dass

- mindestens ein Anteil p der beobachteten Werte kleiner oder gleich $\tilde{x}_p$ ist und
- mindestens ein Anteil $1 - p$ der beobachteten Werte größer oder gleich $\tilde{x}_p$ ist

Beispiel 2.27 (p-Quantile)
Für die geordnete Stichprobe $(2,2,3,5,7,8,9)$ vom Umfang $n = 7$ ergeben sich die Quantile

$$\tilde{x}_{0.1} = x_{\lfloor 0.7 \rfloor + 1} = x_1 = 2,$$
$$\tilde{x}_{0.25} = x_{\lfloor 1.75 \rfloor + 1} = x_2 = 2,$$
$$\tilde{x}_{0.5} = x_{\lfloor 3.5 \rfloor + 1} = x_4 = 5,$$
$$\tilde{x}_{0.75} = x_{\lfloor 5.25 \rfloor + 1} = x_6 = 8,$$
$$\tilde{x}_{0.9} = x_{\lfloor 6.3 \rfloor + 1} = x_7 = 9$$

$\tilde{x}_{0,1} = x_{\lfloor n \cdot 0,1 \rfloor + 1} = x_{\lfloor 7 \cdot 0,1 \rfloor + 1} = x_1 = 2$

Quartile : p: 25% , (50%) , 75%
untere obere Quartil

Dezile : p: 10%, 20%, [...] 90%

Perzentil : p: 1%, 2%, [...] 99%

Streuungsparameter (Streuungsmaße)

Neben dem Durchschnitt der beobachteten Werte ist auch deren Streuung, also die Breite ihrer Verteilung, von Interesse. Diese wird durch Streuungsparameter angegeben.

Spannweite

- Ein sehr einfaches Maß für die Streuung ist durch die Spannweite gegeben. Falls die Daten **Ausreißer enthalten**, wird die Spannweite jedoch **ausschließlich von diesen bestimmt !**

Definition 2.30 (Spannweite)
Sei $(x_1, \ldots, x_n)$ eine Stichprobe vom Umfang n.

Dann ist die ***Spannweite*** (engl. *range*) R definiert als die Differenz zwischen dem größten und dem kleinsten beobachteten Wert,

$$R = \max_{1 \leq i \leq n} x_i - \min_{1 \leq i \leq n} x_i.$$

Varianz und Standardabweichung

Definition 2.31 (Empirische Varianz)
Sei $(x_1, \ldots, x_n)$ eine Stichprobe vom Umfang n mit arithmetischem Mittel $\bar{x}$.

Dann ist die ***empirische Varianz*** (oder ***Stichprobenvarianz***, engl. *sample variance*) s_x^2 (oder kurz s^2) definiert als

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

- In der Praxis wird der quadratische Abstand vom Mittelwert bevorzugt. Dadurch werden große Abweichungen überproportional stark in das Ergebnis eingehen.

Definition 2.32 (Empirische Standardabweichung)
Seien die gleichen Voraussetzungen gegeben wie in Def. 2.31.

Die ***empirische Standardabweichung*** (oder ***Stichprobenstandardabweichung***, engl. *sample standard deviation*) s_x (oder kurz s) ist definiert als

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

Beispiel 2.33 (Varianz und Standardabweichung - Ausreißer)
Für die Stichprobe aus Bsp. 2.22,

$$(6,6,6,6,6,6,7,7,7,8,8,8,9,9,20),$$

erhalten wir als empirische Varianz und Standardabweichung

$$s_x^2 \approx 12.4, \quad s_x \approx 3.51.$$

Ohne Berücksichtigung des Ausreißers $x_{15} = 20$ erhalten wir

$$s_x^2 \approx 1.30, \quad s_x \approx 1.14.$$

Für die praktische Berechnung der empirischen Varianz ist die folgende Beziehung hilfreich.

Satz 2.34 (Varianz und Mittelwert)
Sei $(x_1, \ldots, x_n)$ eine Stichprobe vom Umfang n mit arithmetischem Mittel $\bar{x}$.

Dann gilt

$$s_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{1}{n-1} \left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2 \right).$$

$$\sum_i x_i c = \sum_i (x_i c)$$

$$\sum_i x_i - c \neq \sum_i (x_i - c)$$

$$\sum_i x_i - c = \left(\sum_i x_i \right) - c$$

Definition 2.35 (Empirischer Variationskoeffizient)
Sei $(x_1, \ldots, x_n)$ eine Stichprobe vom Umfang n mit positiven Ausprägungen $x_i > 0$, arithmetischem Mittel $\bar{x}$ und Standardabweichung s_x.

Dann ist der ***empirische Variationskoeffizient*** (engl. *coefficient of variation*) CV_x (oder kurz CV) definiert als

$$\mathrm{CV}_x = \frac{s_x}{\bar{x}}.$$

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

Interquartilsabstand

Eine robuste Kenngröße für die Streuung ist der Interquartilsabstand. Er ist weitgehend unab- hängig von Ausreißern, was je nach Anwendung gewünscht sein kann oder auch nicht.

Definition 2.36 (Interquartilsabstand)
Sei $(x_1, \dots, x_n)$ eine Stichprobe vom Umfang n.

Dann ist der ***Interquartilsabstand*** d_Q definiert als die Differenz zwischen dem oberen und dem unteren Quartil,

$$d_Q = \tilde{x}_{0.75} - \tilde{x}_{0.25}.$$

Beispiel 2.37 (Interquartilsabstand)
Für die Stichprobe aus Bsp. 2.22,

$$(6,6,6,6,6,6,7,7,7,8,8,8,9,9,20),$$

erhalten wir mit $n = 15$ die Quartile

$$\tilde{x}_{0.25} = x_4 = 6, \quad \tilde{x}_{0.75} = x_{12} = 8, \quad d_q = 8 - 6 = 2.$$

Ohne Berücksichtigung des Ausreißers $x_{15} = 20$ erhalten wir mit $n = 14$

$$\tilde{x}'_{0.25} = x_4 = 6, \quad \tilde{x}'_{0.75} = x_{11} = 8, \quad d'_q = 8 - 6 = 2.$$

2.3.3. Skalenniveaus und Kenngrößen

Die folgende Tabelle gibt an, welche Kenngrößen in Abhängigkeit vom Skalenniveau eines Merkmals berechnet werden können. Es sind auch einige Kenngrößen aufgeführt, die erst später oder gar nicht in der Vorlesung behandelt werden.

Skalenniveau	Kenngrößen
Nominal	Modalwert, Chi-Quadrat-(χ^2-)Anpassungstest
Ordinal	zusätzlich Median, Quantile, Rangkorrelation
Intervall	zusätzlich Arithmetisches Mittel, Standardabweichung, Varianz, Interquartilabstand, Varianzanalyse, Korrelation, Regression
Verhältnis	zuätzlich Geometrisches Mittel, Harmonisches Mittel