Univariate Statistik

⊛ örtlich/zeitlich/sachlich

Grundgesamtheit & Erhebungsmerkmale

Grundgesamtheit := Menge der Elemente/Untersuchungseinheiten mit festgelegten⊛ Eigenschaften (Identifikationskriterien)

Bsp: Ω = {Studierenden an der HNEE 2025 im Studiengang NOEM}

Erhebungsmerkmale := Variable⊛ Eigenschaften der Untersuchungseinheiten

Bsp:
- Geschlecht: männlich, weiblich, ...
- Note: 1, 2, 3, 4, 5
- Einkommen: 800€, 1200€, 1337€, ...

Aufgabe: Die zuständige Abteilung für Finanzen des Senats von Berlin lässt sich für eine statistische Analyse von allen Berliner Bühnen die Höhe der Einnahmen und der Ausgaben im Jahr 2024 melden.

(a) Was ist hier die Grundgesamtheit, was sind die einzelnen statistischen Einheiten?

(b) Welche statistischen Merkmale werden betrachtet? Welche Merkmale sind Identifikationskriterien und welche sind Erhebungsmerkmale?

(c) Nennen Sie weitere mögliche erfassbare Merkmale in dieser Gesamtheit!

(a) $\Omega = \{$ Berliner Bühnen 2024 $\}$ Bsp: Volksbühne, Deutsche Oper, Theater des Westens, ...

(b) örtlich: Berlin (festgelegt)

zeitlich: 2024 (-"-)

sachlich: Einnahmen & Ausgaben (Variablen)

(c) • Besucherzahl

• Anzahl Mitarbeitenden

• Bühnengröße

Mächtigkeit von Merkmalen

diskret = Anzahl der Ausprägungen ist zählbar

stetig = Anzahl der Ausprägungen ist überabzählbar

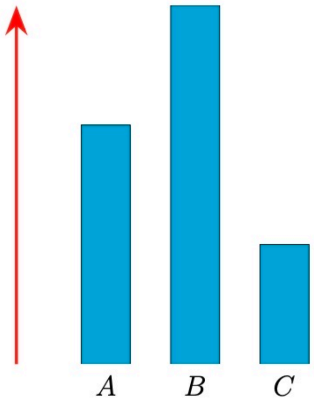
(quasistetig = diskret, aber als stetig behandelt

diskretisiert = stetig, aber nur in diskreter Form erfassbar)

Skalenniveaus von Merkmalen

„Name"

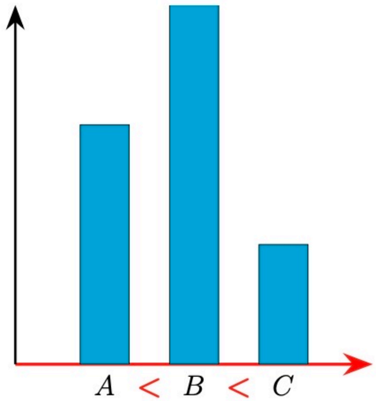
Nominalskala



- Säulen-, Balken-, Kreisdiag.
- Abs. & Rel. Häufigkeiten
- Modus
- χ^2-basierte Kontingenzkoeffizienten

„Ordnung"

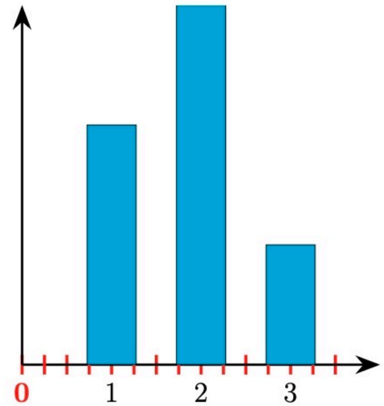
Ordinalskala



- Boxplot
- Emp. Verteilungsfkt.
- p-Quantile (Bsp. Median)
- Spannweite
- Quartilabstand
- ~~Dispersionskoeffizient~~
- ~~Konkordanz, Diskordanz~~
- ~~γ, τ_a, τ_b, τ_c~~, r_S - Korr.-koeffs.

„+,- rechnen ist sinnvoll"

Kardinalskala



- Histogramm
- Artihm./~~Harm.~~/Geom. Mittel
- Varianz/Standardabweichung
- Variationskoeffizient
- ~~Konzentrationsrate~~
- ~~Konzentrationskurve~~
- ~~Herfindahl-Index~~
- ~~Lorenzkurve~~
- ~~Gini-Koeffizient~~
- Kovarianz
- r_B - Korr.-koeff.
- Regressionskoeffizienten

Aufgabe: Sind die folgenden Merkmale diskret oder stetig? Welche Skalierung haben sie?

(a) Farbe eines Autos → diskret, nominal

(b) Punkte in einer Klausur → diskret, kardinal

(c) Lebensalter → stetig, kardinal

(d) Lebensalter in Jahren → diskret, kardinal

(e) Kleidergröße (XS, S, M, L, XL) → diskret, ordinal

(f) Anzahl Neugeborener Kinder im Jahr 2024 → diskret, kardinal

Empirische Verteilungsfunktion

Def.: Die empirische Verteilungsfunktion ist die Verteilungsfunktion der Stichprobe. $F: \mathbb{R} \to [0,1]$ mit

unklassierte Daten:
(geordnete Gruppen $j = 1, \dots, k$)

$$F(x) = \begin{cases} 0 & x < x_1 \\ \sum_{i=1}^{j} f_i & x_j \leq x < x_{j+1} \\ 1 & x \geq x_k \end{cases}$$

klassierte Daten:
(innerhalb gleichverteilt)

Anteil der aktuellen Klasse

$$F(x) = \begin{cases} 0 & x < x_1 \\ \sum_{i=1}^{j-1} f_i + \boxed{\frac{x - x_j^u}{x_j^o - x_j^u}} f_j & x_j^u \leq x < x_j^o \\ 1 & x \geq x_k \end{cases}$$

$n :=$ Stichprobengröße

$n_i :=$ absolute Häufigkeit der x_i

$f_i := \frac{n_i}{n}$ relative Häufigkeit der x_i

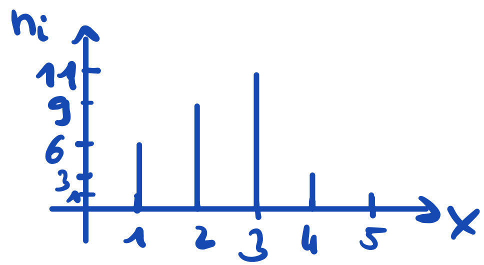
Aufgabe: In einer Statistik Klausur des letzten Schuljahres haben die 30 Schüler:innen der Klasse 10b eines Hamburger Gymnasiums folgende Noten erzielt:

Note x_i	1	2	3	4	5
Häufigkeit n_i	6	9	11	3	1

(a) Ermitteln Sie die absolute und relative Häufigkeit für jede Ausprägung des Merkmals „Note“ und wählen Sie eine passende grafische Darstellung der Daten.

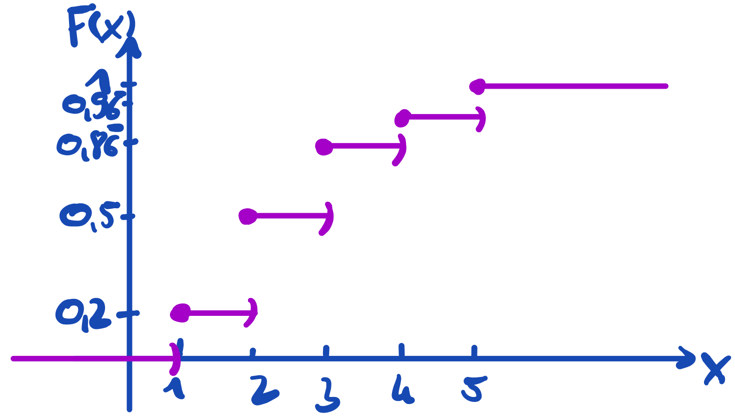
(b) Geben Sie die empirische Verteilungsfunktion an und stellen Sie diese grafisch dar.

x_i	n_i	$f_i := \frac{n_i}{n}$	$F_i = F(x_i) = \sum_{x \leq x_i} f_i$
1	6	0,2	0,2
2	9	0,3	0,5
3	11	$0{,}3\overline{6}$	$0{,}8\overline{6}$
4	3	0,1	$0{,}9\overline{6}$
5	1	$0{,}0\overline{3}$	1
$\sum$	$n=30$	1	/



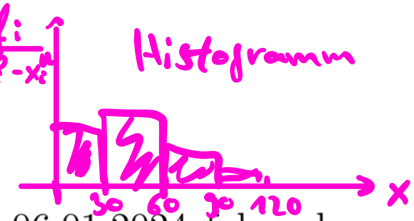
Stabdiagramm

$$F(x) = \begin{cases} 0 & x < 1 \\ 0{,}2 & 1 \leq x < 2 \\ 0{,}5 & 2 \leq x < 3 \\ 0{,}8\overline{6} & 3 \leq x < 4 \\ 0{,}9\overline{6} & 4 \leq x < 5 \\ 1 & x \geq 5 \end{cases}$$



emp. Vert.-Fkt. bei unklassierten Daten ist Treppenfkt.

$\hat{f}_i = \frac{f_i}{x_i^o - x_i^u}$ Histogramm



Aufgabe: An einer Bahnschranke der Bahnstrecke von A nach B wurden am 06.01.2024 folgende Abstände der Zugfolge in Minuten gemessen (und aufsteigend sortiert):

5, 5, 6, 6, 9, 11, 12, 14, 16, 18, 23, 23, 26, 27, 29, 33, 36, 37, 40, 43, 43, 43, 43, 44, 44, 46, 46, 47, 52, 52, 57, 59, 63, 69, 77, 80, 81, 112, 113, 117

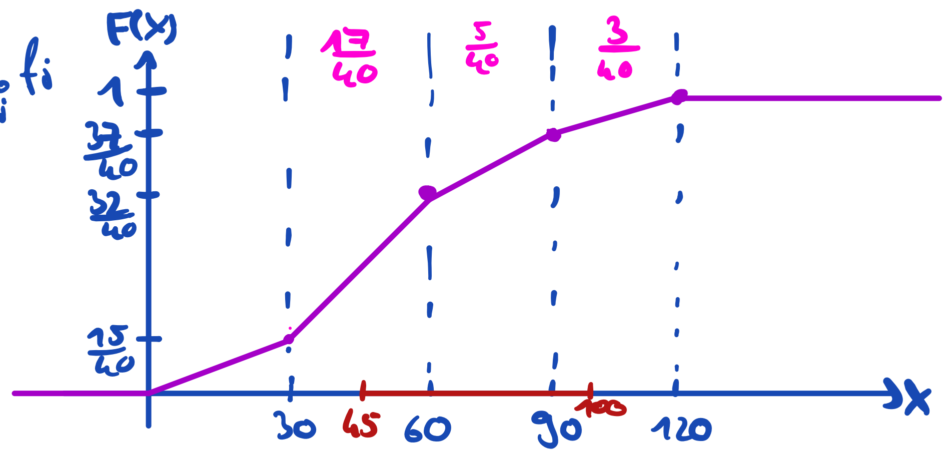
(a) Ermitteln Sie die absoluten und relativen Klassenhäufigkeiten unter Verwendung von Klassen der Breite 30 Minuten. (Stellen Sie die Daten grafisch in einem zugehörigen Histogramm dar.)

(b) Geben Sie die empirische Verteilungsfunktion für die klassierten Daten an und stellen Sie diese grafisch dar.

(c) Wieviele Zeitabstände von 45 bis 100 Minuten konnten beobachtet werden? Wie groß ist der Anteil in den klassierten Daten und wie lässt sich die Abweichung begründen?

(a)

$[x_j^u, x_j^o)$	n_j	f_j	$F_j = F(x_j^o) = \sum_{x \le x_j^o} f_i$
$[0,30)$	15	$\frac{15}{40}$	$\frac{15}{40}$
$[30,60)$	17	$\frac{17}{40}$	$\frac{32}{40}$
$[60,90)$	5	$\frac{5}{40}$	$\frac{37}{40}$
$[90,120)$	3	$\frac{3}{40}$	1
Σ	$n=40$	1	/



(c) unklassierte Daten:

$P(45 \le X \le 100) = \frac{12}{40} = 0{,}3$

klassierte Daten:

$P(45 \le X \le 100) = \underbrace{\frac{60-45}{60-30}}_{=\frac{1}{2}} \cdot \frac{17}{40} + \frac{5}{40} + \underbrace{\frac{100-90}{120-90}}_{=\frac{1}{3}} \cdot \frac{3}{40} = 0{,}3625$

Modalwert / Modus

Def.: Der Modus ist der (Ausprägungs-) Wert, der am häufigsten auftritt.

unklassierte Daten:
(geordnete Gruppen $j = 1, \dots, k$)

$$x_{mod} = \text{Häufigste Ausprägung}$$

klassierte Daten:
(innerhalb gleichverteilt)

Anteil

$$x_{mod} = x_M^u + \boxed{\frac{f_M^* - f_{M-1}^*}{2 \cdot f_M^* - f_{M-1}^* - f_{M+1}^*}} (x_M^o - x_M^u)$$

$$f_i^* := \frac{f_i}{x_i^o - x_i^u} \quad \text{(Dichte der Klasse } i\text{)}$$

M = Index der Klasse mit der größten Dichte

Aufgabe: Im Folgenden betrachten wir einen Ausschnitt aus dem Mietspiegel München 2003. Der Datensatz umfasst Wohnungen ohne zentrale Warmwasserversorgung und mit einer Wohnfläche von höchstens 71 m^2 eines bestimmten Gebiets. Die (gerundeten) Nettokaltmieten (in €) betragen:

x_i	80	110	130	160	170	180	210	230	240	$\boxed{260}$	270	280	310	350	$\boxed{360}$	400	530
n_i	1	1	1	2	2	2	2	1	1	$\boxed{3}$	1	2	1	1	$\boxed{3}$	1	1

Tabelle 1: Gruppierte Daten

$(x_i^u, x_i^o]$	$(0, 100]$	$(100, 200]$	$(200, 300]$	$(300, 400]$	$(400, 500]$	$(500, 600]$	Σ
n_i	1	8	10	5	1	1	$n = 26$
f_i	$\frac{1}{26}$	$\frac{8}{26}$	$\frac{10}{26}$	$\frac{5}{26}$	$\frac{1}{26}$	$\frac{1}{26}$	1
f_i^*	$\frac{1}{2600}$	$\frac{8}{2600}$	$\frac{10}{2600}$	$\frac{5}{2600}$	$\frac{1}{2600}$	$\frac{1}{2600}$	/

Tabelle 2: Klassierte Daten

Bestimmen Sie die Modalwerte der unklassierten und der klassierten Daten. Wie ist der Unterschied zu erklären?

unklassiert: $x_{mod} = \underline{\underline{260}}$ $\underline{\text{und}}$ $x_{mod} = \underline{\underline{360}}$

weil alle Klassen die gleiche Breite haben

klassiert: $$x_{mod} = x_M^u + \frac{f_M^* - f_{M-1}^*}{2 \cdot f_M^* - f_{M-1}^* - f_{M+1}^*}(x_M^o - x_M^u) = x_M^u + \frac{n_M - n_{M-1}}{2 n_M - n_{M-1} - n_{M+1}} \cdot (x_M^o - x_M^u)$$

$M = 3$

(Klasse $(200, 300]$)

$$= 200 + \frac{10 - 8}{2 \cdot 10 - 8 - 5}(300 - 200)$$

$$= 200 + \frac{2}{7} \cdot 100 \approx \underline{\underline{228{,}5714}}$$

P-Quantile

<u>Def.:</u> Sei $p \in (0,1)$. P-Quantile sind die Werte, bei denen mindestens p $[\cdot 100\%]$ aller Ausprägungen drunter liegen <u>und</u> mindestens $(1-p)[\cdot 100\%]$ aller Ausprägungen drüber liegen.

unklassierte Daten:
(geordnete Gruppen $j=1,\ldots,k$)

$$x_p = \begin{cases} x_{(\lceil np \rceil)} & np \notin \mathbb{Z} \\ \frac{1}{2}\left(x_{(np)} + x_{(np+1)}\right) & np \in \mathbb{Z} \end{cases}$$

klassierte Daten:
(innerhalb gleichverteilt)

$$x_p = x_Q^u + \overbrace{\boxed{\frac{p - F_{Q-1}}{f_Q}}}^{\text{Anteil}} \cdot \left(x_Q^o - x_Q^u\right)$$

Median = (mittlere) 50%-Quantil

Q = Index der Klasse, in der die (emp.) Verteilungsfunktion den Wert p erreicht

Aufgabe: Um festzustellen, wie viele Stunden pro Spieltag ein Fußball-Fan die WM-Berichterstattung im Fernsehen während der letzten Fußball-WM verfolgte, wurden 20 Fußball-Fans in A-Dorf nach ihrem Fernsehkonsum während der WM befragt. Die Befragung brachte folgendes Ergebnis:

Stunden	0	1	2	3	4
rel. Häufigkeit	0.05	0.1	0.35	0.25	0.25

(a) Bestimmen Sie tabellarisch die absoluten Häufigkeiten und die empirische Verteilungsfunktion.

(b) Geben Sie an, wieviele Stunden 85% der Befragten mindestens fernsehen.

(c) Wieviele Stunden sehen 50% der Befragten höchstens fern?

(d) Bestimmen Sie den Median. Worin liegt der Unterschied zum Ergebnis aus (c)?

(e) Bestimmen Sie das untere und obere Quartil.

(a)

x_i	n_i	f_i	F_i
0	1	0,05	0,05
1	2	0,1	0,15
2	7	0,35	0,5
3	5	0,25	0,75
4	5	0,25	1
Σ	n=20	1	/

(b) f_i summieren (von max Stunden nach unten) bis 85% erreicht sind: $\underset{4h}{0{,}25}+\underset{3h}{0{,}25}+\underset{2h}{0{,}35}=0{,}85$

$\underline{\underline{x=2}}$ Stunden

(c) f_i summieren (von min Stunden nach oben) bis 50% erreicht sind: $\underset{0h}{0{,}05}+\underset{1h}{0{,}1}+\underset{2h}{0{,}35}=0{,}5$

$\underline{x=2 \text{ Stunden}}$

(d) $x_{med}=x_{0,5}=\frac{2+3}{2}=\underline{\underline{2{,}5}}$

(e) $x_{0,25}=\underline{\underline{2}}$

$x_{0,75}=\frac{3+4}{2}=\underline{\underline{3{,}5}}$

0, 1, 1, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 4, 4, 4, 4, 4

Aufgabe: Herr Meier besitzt einen Gartenzwerg-Großhandel mit drei Filialen: Berlin, New York und Flensburg. Am Ende des Geschäftsjahres möchte er einen Überblick über die Geschäftslage erhalten und fordert deshalb in allen drei Filialen Informationen über die innerhalb des letzten Jahres eingegangenen Aufträge an. Seine Berliner Filiale übermittelt ihm folgende Informationen:

Median-Klasse Q

Auftragshöhe in € (von … bis unter …)	Anzahl der Aufträge n_i	f_i	F_i
0 – 20.000	20	0,2	0,2
20.000 – 50.000	35	0,35	0,55
50.000 – 150.000	40	0,4	0,95
150.000 – 300.000	5	0,05	1
Σ	$n = 100$	1	/

(a) Bestimmen Sie den Median der Umsätze!

(b) Wie groß war der Umsatz der 20% defizitärsten Aufträge?

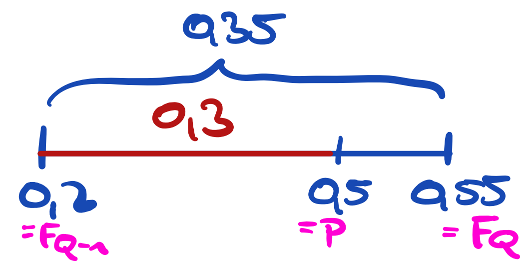
(c) Wie groß war der Umsatz der 20% lukrativsten Aufträge?

$\frac{p - F_{Q-1}}{F_Q - F_{Q-1}}$

(a) $$x_{med} = x_{0,5} = x_Q^u + \boxed{\frac{p - F_{Q-1}}{f_Q}} \cdot (x_Q^o - x_Q^u)$$

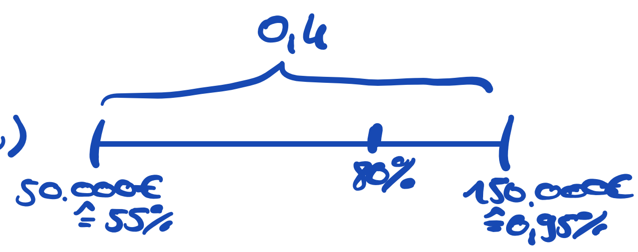
$$= 20.000 + \frac{0,5 - 0,2}{0,35} \cdot (50.000 - 20.000)$$

$$= 20.000 + \underbrace{\frac{0,3}{0,35}}_{=\frac{6}{7}} \cdot 30.000 \approx 45.714,2857\,€$$



(b) $x_{0,2} = 20.000\,€$

(c) $$x_{0,8} = 50.000\,€ + \frac{0,8 - 0,55}{0,4} \cdot (150.000 - 50.000) = 112500\,€$$



Arithmetisches Mittel

Def.: Das arithmetische Mittel ist die Summe aller Beobachtungen im Verhältnis zur Stichprobengröße: $\overline{x} := \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$

unklassierte Daten:
(geordnete Gruppen $j=1,\dots,k$)

$$\overline{x} = \frac{1}{n} \cdot \sum_{i=1}^{k} n_i \cdot x_i = \sum_{i=1}^{k} \underbrace{f_i}_{=\frac{n_i}{n}} \cdot x_i$$

klassierte Daten:
(innerhalb gleichverteilt)

$$\overline{x} = \frac{1}{n} \cdot \sum_{j=1}^{k} n_j \cdot \overline{x}_j = \sum_{j=1}^{k} f_j \cdot \overline{x}_j$$

$\overline{x}_j$:= Klassenmittelwerte

Aufgabe: In der Klasse 12c des Geschwister-Scholl-Gymnasiums in Fürstenwalde wurden vor 2 Monaten die Kinder nach der Anzahl ihrer Geschwister befragt. In der nachstehenden Häufigkeitstabelle finden Sie die Ergebnisse der Umfrage:

Anzahl der Geschwister	Häufigkeit
0	2
1	12
2	13
3	2
4	1
Σ	$n=30$

Berechnen Sie die durchschnittliche Anzahl der Geschwister der befragten Schüler:innen.

$$\bar{x} = \frac{1}{30} \cdot (0 \cdot 2 + 1 \cdot 12 + 2 \cdot 13 + 3 \cdot 2 + 4 \cdot 1) = \underline{\underline{1{,}6}}$$

Aufgabe: Im Sinne einer medizinischen Untersuchung wurde eine schwedische Gruppe von Studierenden zu Beginn der Corona Pandemie nach ihrer Körpergröße (in Meter) befragt. Die Größenverteilung wurde in folgender Tabelle zusammengefasst:

$$m_1 = \frac{1{,}55 + 1{,}65}{2} = 1{,}6$$
$$m_2 = \frac{1{,}65 + 1{,}75}{2} = 1{,}7$$
$$m_3 = \frac{1{,}75 + 1{,}85}{2} = 1{,}8$$

Körpergröße	Häufigkeit
$[1.55, 1.65)$	14
$[1.65, 1.75)$	18
$[1.75, 1.85)$	25
$\sum$	$n = 57$

Berechnen Sie die Durchschnittsgröße der Studierenden mit Hilfe des arithmetischen Mittels. Warum ist dieses Lagemaß hier besser geeignet als der Modus oder der Median?

$$\bar{x} = \frac{1}{57} \cdot (1{,}6 \cdot 14 + 1{,}7 \cdot 18 + 1{,}8 \cdot 25) \approx \underline{\underline{1{,}72\,m}}$$

Empirische Varianz

Def.: Die empirische Varianz beschreibt die mittlere quadratische Abweichung der einzelnen Beobachtungen vom arith. Mittel.

$$\tilde{s}^2 := \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2 \text{ (unkorrigiert)} \qquad s^2 := \frac{n}{n-1} \cdot \tilde{s}^2 \text{ (korrigiert)}$$

unklassierte Daten:
(geordnete Gruppen $j=1,\dots,k$)

$$\tilde{s}^2 = \frac{1}{n} \cdot \sum_{j=1}^{k} n_j \cdot (x_j - \bar{x})^2$$

$$= \sum_{j=1}^{k} f_j \cdot (x_j - \bar{x})^2$$

$$= \overline{x^2} - \bar{x}^2$$

$\overline{x^2}$:= Mittelwert der quadr. Beobachtungen

klassierte Daten:
(innerhalb gleichverteilt)

$$\tilde{s}^2 = \tilde{s}^2_{innerhalb} + \tilde{s}^2_{zwischen}$$

$$= \frac{1}{n} \cdot \sum_{j=1}^{k} n_j \cdot \tilde{s}_j^2 + \frac{1}{n} \cdot \sum_{j=1}^{k} n_j \cdot (\bar{x}_j - \bar{x})^2$$

$$= \sum_{j=1}^{k} f_j \cdot \tilde{s}_j^2 + \sum_{j=1}^{k} f_j \cdot (\bar{x}_j - \bar{x})^2$$

Aufgabe: In der Boulder Weltrangliste der Frauen (Stand 21. November 2022) stehen folgende 10 Sportlerinnen an der Spitze:

Weltrang #	Name	Alter in Jahren
1	Alex Puccio	33
2	Michaela Kiersch	28
3	Staša Gejo	24
4	Janja Garnbret	23
5	Alizée Dufraisse	35
6	Ashima Shiraishi	21
7	Nolwen Berthier	28
8	Laura Rogora	21
9	Matilda Söderlund	30
10	Lucy Watson	34

(a) Bestimmen Sie die empirische Varianz des Lebensalters für diese Stichprobe.

(b) Ordnen Sie die Daten in zwei Klassen ein. Die Klassenbreiten sollen dabei den Spannweiten entsprechen, die sich durch die Aufteilung in „Sportlerinnen in ihren 20er Jahren“ und „Sportlerinnen in ihren 30er Jahren“ ergibt.

(c) Überprüfen Sie die Berechnung aus (a), indem Sie eine Streuungszerlegung durchführen und die Varianzen innerhalb und zwischen den Klassen berechnen.

(d) Wie groß ist die Varianz der klassierten Daten aus (b), wenn die Originaldaten verloren gegangen sind?

(e) Bestimmen Sie den Variationskoeffizienten.

Weltrang #	Name	Alter in Jahren
1	Alex Puccio	33
2	Michaela Kiersch	28
3	Staša Gejo	24
4	Janja Garnbret	23
5	Alizée Dufraisse	35
6	Ashima Shiraishi	21
7	Nolwen Berthier	28
8	Laura Rogora	21
9	Matilda Söderlund	30
10	Lucy Watson	34

(a) $\tilde{s}^2 = \overline{x^2} - \bar{x}^2 = 792{,}5 - 27{,}7^2 = \underline{\underline{25{,}21}}$

$\bar{x} = \frac{1}{10} \cdot (33 + 28 + \ldots + 34) = 27{,}7$

$\overline{x^2} = \frac{1}{10} \cdot (33^2 + 28^2 + \ldots + 34^2) = 792{,}5$

(b) <u>Klasse 1</u>: 20er Jahre

21, 21, 23, 24, 28, 28

<u>Klasse 2</u>: 30er Jahre

30, 33, 34, 35

$\bar{x}_j$	k_j	n_j	$f_j = \frac{n_j}{n}$
$\frac{145}{6}$	[20, 30)	6	$3/5 = 0{,}6$
33	[30, 40)	4	$2/5 = 0{,}4$
	Σ	$n = 10$	1

$\bar x_j$	k_j	n_j	$f_j = \frac{n_j}{n}$
$\frac{145}{6}$	$[20,30)$	6	$\frac{3}{5} = 0{,}6$
33	$[30,40)$	4	$\frac{2}{5} = 0{,}4$
Σ		$n=10$	1

(c)

INNERHALB DER KLASSEN

$$\bar x_1 = \frac{1}{6}(21+\ldots+28) = \frac{145}{6} \; (=24{,}1\overline{6}) \qquad \bar x_2 = \frac{1}{4}(30+\ldots+35) = 33$$

$$\overline{x_1^2} = \frac{1}{6}\cdot(21^2+\ldots+28^2) = \frac{1185}{2} \qquad \overline{x_2^2} = \frac{1}{4}(30^2+\ldots+35^2) = \frac{2185}{2}$$

$$\tilde s_1^2 = \overline{x_1^2} - \bar x_1^2 = \frac{305}{36} \qquad \tilde s_2^2 = \overline{x_2^2} - \bar x_2^2 = \frac{7}{2}$$

Mittelwert der einzelnen Varianzen

$$\tilde S^2_{\text{innerhalb}} = \frac{1}{n}\cdot\sum_{j=1}^{k} n_j\cdot\tilde s_j^2 = \frac{1}{10}\cdot\left(6\cdot\frac{305}{36} + 4\cdot\frac{7}{2}\right) = \frac{389}{60} \; (=6{,}48\overline{3})$$

Streuungszerlegung: $\tilde S^2 = \tilde S^2_{\text{innerhalb}} + \tilde S^2_{\text{zwischen}}$

$25{,}21 = \frac{389}{60} + \frac{2809}{150}$ ✓

ZWISCHEN DEN KLASSEN

$$\bar x = \frac{1}{10}\left(6\cdot\frac{145}{6} + 4\cdot 33\right) = 27{,}7 \quad \text{(siehe (a))}$$

Mittelwert aus den Klassenmittelwerten = Mittelwert aus den Originalwerten

$$\overline{x^2} = \frac{1}{10}\left(6\cdot\left(\frac{145}{6}\right)^2 + 4\cdot 33^2\right) = \frac{47161}{60} \; (=786{,}01\overline{6})$$

$$\tilde S^2_{\text{zwischen}} = \frac{1}{n}\cdot\sum_{j=1}^{k} n_j\cdot(\bar x_j - \bar x) = \overline{x^2} - \bar x^2 = \frac{47161}{60} - 27{,}7^2 = \frac{2809}{150} \; (=18{,}72\overline{6})$$

ANMERKUNG:

Würden die Originaldaten bei der Klassenbildung „verloren gehen", werden die Klassenmittelwerte geschätzt durch die Klassenmitten.

echte Mittelwerte $\bar{x}_j$	geschätzte Mittelwerte $\bar{x}_j$	k_j	n_j	$f_j = \frac{n_j}{n}$
$\frac{145}{6}$	25	$[20,30)$	6	$3/5 = 0{,}6$
33	35	$[30,40)$	4	$2/5 = 0{,}4$
		Σ	$n=10$	1

(d)

$$\bar{x} = \frac{1}{n}\cdot\sum_{j=1}^{k} n_j \cdot \bar{x}_j = \frac{1}{10}\cdot(6\cdot 25 + 4\cdot 35) = 29$$

$$\overline{x^2} = \frac{1}{n}\cdot\sum_{j=1}^{k} n_j \cdot \bar{x}_j^2 = \frac{1}{10}(6\cdot 25^2 + 4\cdot 35^2) = 865$$

$$s^2 = \frac{1}{n}\cdot\sum_{j=1}^{k} n_j \cdot (\bar{x}_j - \bar{x})^2 = \overline{x^2} - \bar{x}^2 = 865 - 29^2 = \underline{\underline{24}} \text{ Jahre}^2$$