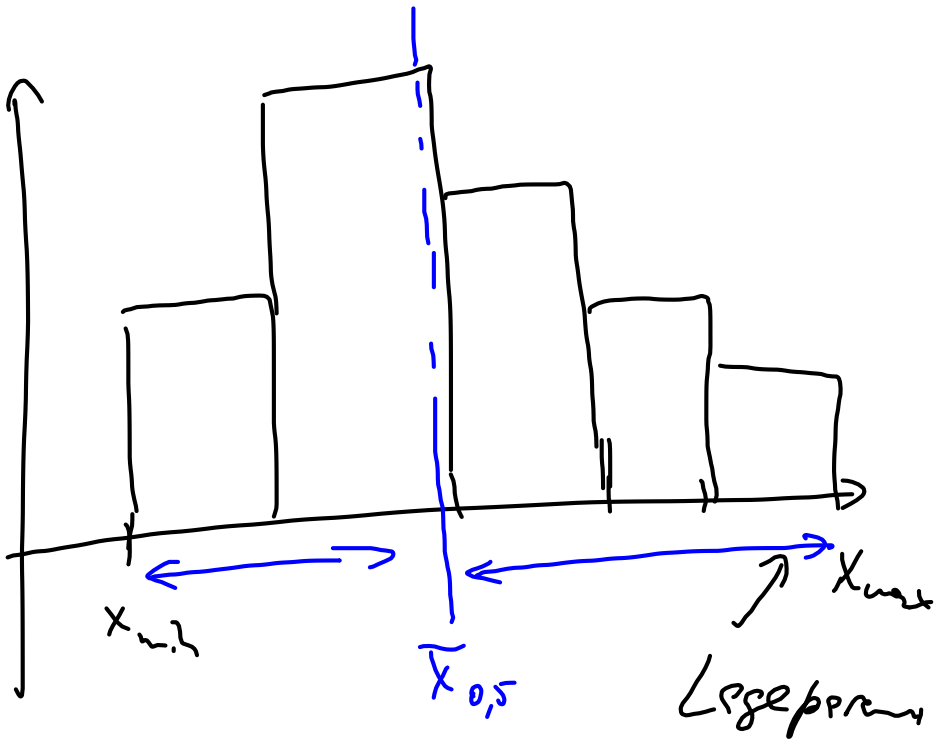
2. Stichprobenkennwerte



Lageparameter

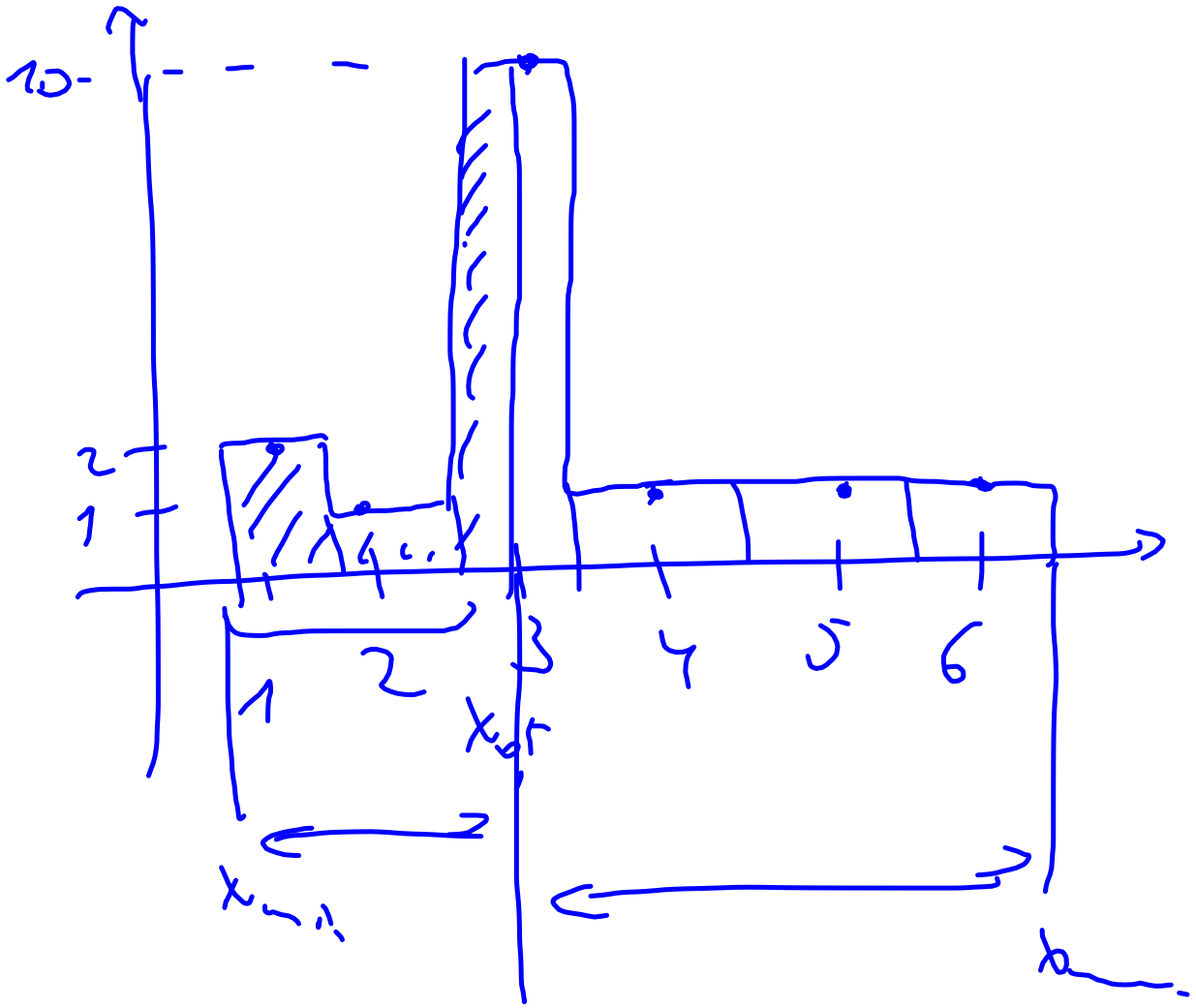
Streuungsmaße

$\tilde{x}_{0,5}$ → 50% links,
50% rechts

↑
Lagerparameter

$|x_{min} - \tilde{x}_{0,5}|$ — streuungsparameter

$|x_{max} - \tilde{x}_{0,5}|$


10
2
1
1
2
3
4
5
6
$x_{śr}$
x_{m+1}
b

2.1. Lageparameter

(Mittelwert, Median, Quantile, Modalwert)

2.1.1. Das arithmetische Mittel.

Sei X ein zufälliges Merkmal; $x_1, x_2, \ldots, x_n$ sei eine Stichprobe von X vom Umfang n.

Def! $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$

heißt arithmetisches Mittelwert.

Bsp: Werkzeit:

Mo	Di:	Mi:	Do	Fr
10	15	20	15	12

$$\bar{x} = \frac{10+15+20+15+12}{5} = 14{,}4 \text{ min}$$

Median: 10, 12, 15, 15, 20

$n = 5$

$$x_{0,5} = x_{[5+1/2]} = x_{[3]} = 15$$

Bsp 2:

P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}
40K	40K	40K	40K	40K	40K	40K	40K	40K	40K

$$\bar{x} = \frac{9 \cdot 40000 + 40000000}{10} =$$

$$= 436000$$

Nachteil vom ar. Mittel

"Ausreißer"

Median: $n = 10$

$$x_{0,5} = \frac{x_{[5]} + x_{[6]}}{2} = \frac{40K + 40K}{2} =$$

$$= 40000.$$

2.1.2. Der Median

$x_1, x_2, \ldots, x_n$ werde aufsteigend geordnet:

$x_{[1]}, x_{[2]}, x_{[3]}, \ldots, x_{[n]}$

Def! 1: Der Median x einer Stichprobe ist der kleinste Wert der Stichprobe, für den gilt:

$F_n(x) \geq 0.5$. Er wird $x_{0,5}$ bezeichnet.

Bsp:

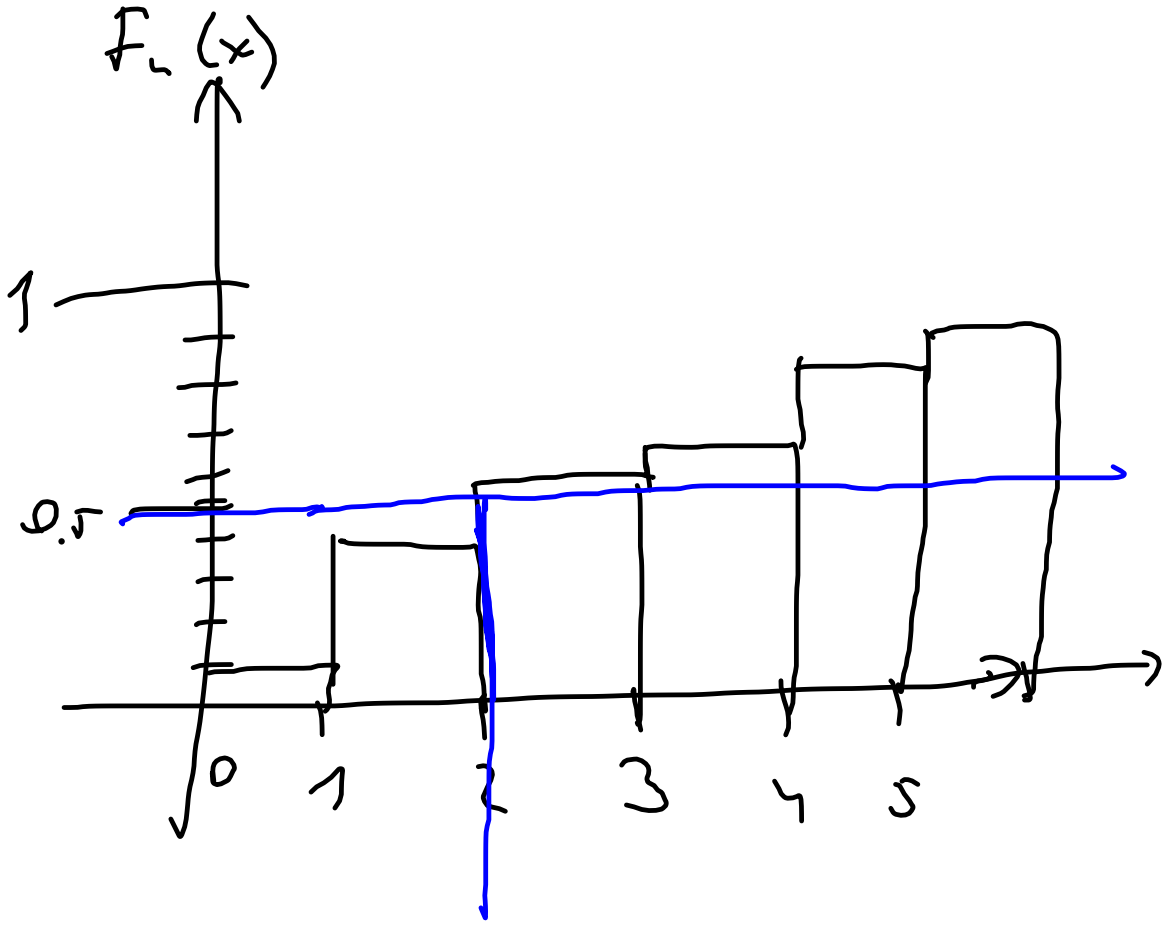
Geordnete Stichprobe

0, 1, 1, 1, 2, 2, 3, 4, 4, 5

(Umfang $n = 10$) $x_{[5]}$ $x_{[6]}$

Häufigkeitsverteilung:

0	1	2	3	4	5
0.1	0.3	0.2	0.1	0.2	0.1



Bsp. zur Def. 2.1 $n = 10$

→ gerade

$$\Downarrow \quad x_{0.5} = \frac{x_{[\frac{10}{2}]} + x_{[\frac{12}{2}]}}{2} =$$

$$= \frac{x_{[5]} + x_{[6]}}{2} = 2.$$

Def. 2: Der Median der Stichprobe $x_1, x_2, \ldots, x_n$ mit der geordneten Stichprobe $x_{[1]}, x_{[2]}, \ldots, x_{[n]}$ (aufsteig. von links nach rechts):

$$x_{0,5} = \begin{cases} x_{\left[\frac{n+1}{2}\right]} & \text{für } n \text{ ungerade} \\ \dfrac{x_{\left[\frac{n}{2}\right]} + x_{\left[\frac{n+2}{2}\right]}}{2} & \text{für } n \text{ gerade} \end{cases}$$

Diese Variante für Berechnungen verwenden!

2.1.3. Das α-Quantil einer Beobachtungsreihe:

$x_1, \dots, x_n$

$x_\alpha \quad (0 < \alpha < 1)$

$(n-\alpha)$ Beobachtungen x_α nicht überschreiten.

(oder $\alpha \cdot 100\%$ der Beobachtungen überschreiten x_α nicht)

Sei:

$$x_{[1]} \leq x_{[2]} \leq x_{[3]} \leq \ldots \leq x_{[n]}$$

Def 1: Das α-Quantil ist der kleinste Wert der Stichprobe, für den gilt: $F_n(x) \geq \alpha$. Er wird x_α bezeichnet.

Def. 2: Der Wert

$$x_\alpha = \begin{cases} \dfrac{x_{[k]} + x_{[k+1]}}{2}, & \text{falls } k = n\cdot\alpha \text{ ganze Zahl ist} \\ x_{[k]}, & \text{falls } (n\cdot\alpha) \text{ keine ganze Zahl ist,} \end{cases}$$

und k die auf $(n\cdot\alpha)$ folgende (nat.) Zahl ist

heißt α-Quantil der Beobachtungsreihe $x_1, \ldots, x_n$

Median ist 0,5 - Quantil

0,25 - Quantil heißt
unteres Quartil

0,75 - Quantil heißt
oberes Quartil.

0,1 - 0,2 - ..... 0,9 - Quantile
heißen Dezentile.

Bsp: 0, 1, 1, 1, 2, 2, 3, 4, 4, 5

↑

Oberes Quartil: $x_{[8]}$

$\alpha = 0,75$

$n = 10$

$n \cdot \alpha = 7,5$ ist nicht ganz

↓

$k = 8$

$\Rightarrow x_{0,75} = x_{[8]} = 4$

unteres Quartil:

$\alpha = 0{,}25$

$n \cdot \alpha = 2{,}5 \longrightarrow k = 3$

$x_{0{,}25} = 1$

0,2 – Quantil (0,2-Dezil)

$\alpha = 0{,}2$

$(\alpha \cdot n) = 2 \in \mathbb{N} \Rightarrow k = 2$

$$x_{0{,}2} = \frac{x_{[2]} + x_{[3]}}{2} = \frac{1+1}{2} = 1$$

0,7 - Dezentil

$$k = (k \cdot n) = 10 \cdot 0,7 = 7$$

$$x_{0,7} = \frac{x_{[7]} + x_{[8]}}{2} = 3,5$$

2.1.4. Der Modalwert

Bei bestimmte Daten wie Berg, Geschlecht, Bezeichnung (Name) kann man keinen Mittelwert, Median u.s.f. bilde, da die Date

nominal skaliert sind.

Def: Der Modalwert (auch "Modus") x_{mod} ist der häufigste Wert einer Verteilung.

Ist X diskret mit $X = \{a_1, \ldots, \ldots\}$ so gilt

$$h_n(x_{mod}) \geq h_n(a_i) \quad \forall i.$$

Bei gruppierten Daten ist der Modalwert die Klassenmitte der am dichtesten besetzten Klasse; Bezeichnung: x_{mod} statt x_{mod}.

Bsp: 12, 10, 16, 12, 17.

$x_{mod} = 12$.