Vorlesung vom 25.10.2012:

- Lage- und Streuungsparameter
- empirische Verteilungsfunktion
- Standardisierung von Daten

---

1) **Lageparameter**: bestimmen das Zentrum der erhobenen Daten

Wichtige Parameter: Modalwert, Median, Mittelwert (quantitative Daten)

arithmetisch — wichtigste!!, geometrisch, harmonisch

a) Modalwert(e): x_{mod} als der Wert / die Werte, die in der Datenmenge am häufigsten vorkommt / vorkommen.

b) Median: geht nur für ordinal bzw. kardinal skalierte Merkmale.

„offensichtlicher" „Zentralwert"

x_{med} teilt die Datenmenge in zwei gleich große Hälften von der Anzahl her.

c) α-Quartil: hat zu tun mit der Verteilungsfunktion $F(x)$

↳ wichtige Spezialfälle: $\boxed{\alpha = 0{,}25 = \frac{1}{4}}$ mit $\boxed{1-\alpha = 1-0{,}25 = 0{,}75 = 3/4}$

ergibt das untere Quartil

$\alpha = 0{,}5 = \frac{1}{2}$, $x_{\frac{1}{2}} = x_{med}$

ergibt das obere Quartil

$\alpha \in]0,1[$, also $0 < \alpha < 1$

Achtung, Teil ist leicht überarbeitet!

x_α ist das „Datum" aus der Datenmenge, so dass für die Anzahl $N \in \mathbb{N}$ der Daten mit $x_i \le x_\alpha$ gilt: $\boxed{\frac{N}{n} = \alpha}$. Oder anders

$x_\alpha = x_N$, wenn $\boxed{\frac{N-1}{n} < \alpha \le \frac{N}{n}}$ und n die Gesamtdatenanzahl ist.

Dazu Ü4) vom 1. Aufgabenblatt:

Teil b),

i	1	2	3	4	5	6	7	8
x_i	34	45	11	42	49	33	27	11

ordinal skalieren!

$\alpha = 1/4$, $\alpha = 1/2$, $\alpha = 3/4$

i	1	2	3	4	5	6	7	8
$\tilde{x}_i$	11	11	27	33	34	42	45	49

30 60 60

$\boxed{x_{mod} = 11}$, da als einziger Wert doppelt auftretend

$x_{med} = \frac{1}{2}(\tilde{x}_4 + \tilde{x}_5) = \frac{33+34}{2} = 33{,}5$

arithmetischer Mittelwert,

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} \tilde{x}_i = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{1}{8} \cdot \sum_{i=1}^{8} x_i = \frac{1}{8}(x_1 + x_2 + \ldots + x_8)$$

griech. Sigma für Summe

$$= \frac{34+45+\ldots+11}{8} = \frac{252}{8} = 31{,}5$$

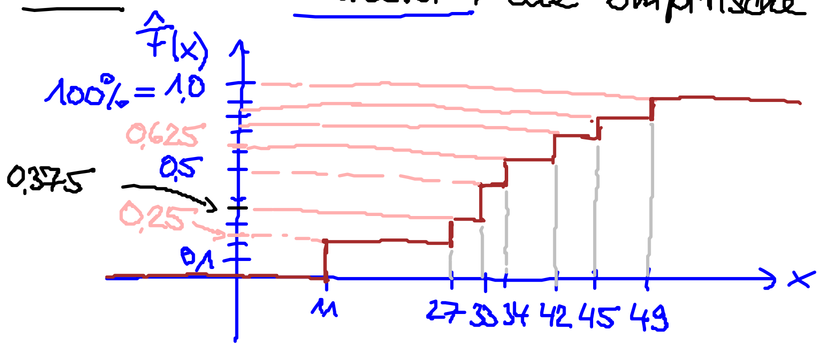
Achtung, Teil wurde noch einmal überarbeitet.

Hier: $x_{mod} = 11 < \bar{x} = 31{,}5 < x_{med} = 33{,}5$

Quartile = $\frac{1}{4}$-Quantile: $n = 8$, $\alpha = 0.25$ sowie $1-\alpha = 0.75$. Die Forderung $\frac{N-1}{8} < \alpha \le \frac{N}{8}$ ergibt für $\alpha = 0.25 = \frac{1}{4}$: $N = 2$ und für $\alpha = 0.75 = \frac{3}{4}$: $N = 6$

Also: $x_{0.25} = \tilde{x}_2 = 11$, $x_{0.75} = \tilde{x}_6 = 42$

Teil a), Wir skizzieren die empirische Verteilungsfunktion



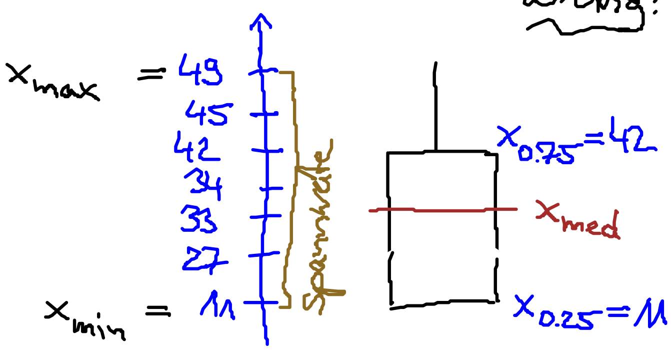
$n = 8$ Daten

$x = 11$ wird 2x angenommen, das entspricht einer relativen Häufigkeit von $\frac{2}{n} = \frac{2}{8} = 0{,}25$ ‖ „Sprunghöhe" von $\hat{F}(x)$ bei $x = 11$!

Alle anderen Werte werden einmal angenommen, also relative Häufigkeit von $\frac{1}{8} = 0{,}125$

Wir haben jetzt die (empirische) Verteilungsfunktion und die Lageparameter beoproben.

2) Streuungsparameter / Vorher: Boxplot zu Aufgabe 4



Wichtig: x_{min}, x_{max}, x_{med}, $x_{0.25}$ und $x_{0.75}$

unteres / oberes Quartil

* für „Ausreißer“

Quartilsabstand $QA = x_{0.75} - x_{0.25} = 42 - 11 = 31$

1) $x_{0.25} - 1{,}5 \cdot QA = 11 - \frac{3}{2} \cdot 31 = -35{,}5 < x_{min}$

2) $x_{0.75} + 1{,}5 \cdot QA = 42 + \frac{3}{2} \cdot 31 = 88{,}5 > x_{max}$

keine Ausreißer nach unten und nach oben!

Achtung! Modifizierter Boxplot!

Nach dem Statistikbuch von Kockelkorn sind α-Quantile – und damit auch die Quartile – nicht eindeutig bestimmt. Die Festlegung $\boxed{x_\alpha = \min\{x \mid F(x) \geq \alpha\}}$ ist daher etwas „willkürlich“.

Übrigens: Ordnet man eine Datenmenge $x_1, x_2, \dots, x_n$ der Größe nach, schreibt man statt $\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n$ auch gern $\boxed{x_{(1)}, x_{(2)}, \dots, x_{(n)}}$ Indizes in runden Klammern.
Dann gilt automatisch:

$$\boxed{x_{min} = x_{(1)}, \quad x_{max} = x_{(n)}, \quad x_\alpha = x_{(N)} \text{ mit } \frac{N-1}{n} < \alpha \leq \frac{N}{n}}.$$

ENDE der Vorlesung!