

Empirische Quantile

Definition. Sei $p \in (0, 1)$ und

$$Q_p = X_{(\lfloor np \rfloor + 1)}.$$

- Dann bezeichnet Q_p das **empirische p -Quantil** der Stichprobe.
- Anteil von $p * 100\%$ ist $\leq Q_p$
- Median
 - 0.5-Quantil
- Quartile

$$Q_{0.25}, Q_{0.5}, Q_{0.75}$$

- Perzentile

$$Q_{0.1}, Q_{0.2}, \dots, Q_{0.9}$$

- Beispiel

Wir betrachten die Zeit (s), die eine CPU für die Ausführung eines Befehls benötigt und notieren $n = 15$ Werte:

1.86	3.49	2.63	3.49	1.69
1.83	0.81	4.70	0.85	4.24
3.49	2.75	1.65	0.92	0.62.

Gesucht sind die empirischen Quartile $q_{0.25}$, $q_{0.5}$ und $q_{0.75}$.

Handwritten calculations for empirical quantiles:

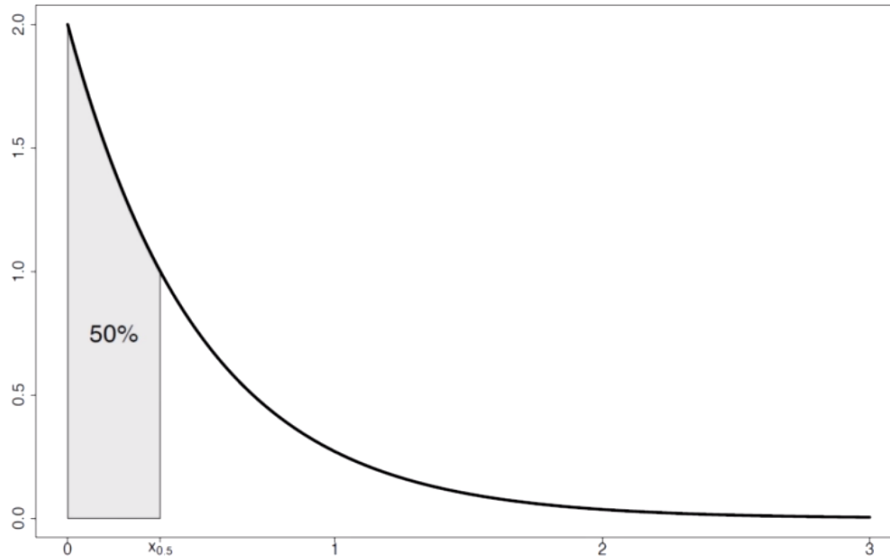
$$\begin{aligned} n &= 15 \\ q_{0.25} &= X_{(\lfloor n \cdot 0.25 \rfloor + 1)} \\ &= X_{(3+1)} = X_{(4)} = 0.92 \\ q_{0.5} &= X_{(\lfloor n \cdot 0.5 \rfloor + 1)} = X_{(8)} = 1.86 \\ q_{0.75} &= X_{(\lfloor n \cdot 0.75 \rfloor + 1)} = X_{(12)} = 3.49 \end{aligned}$$

Theoretische Quantile

Definition. Sei $p \in (0, 1)$ und X eine Zufallsvariable mit stetiger Verteilungsfunktion F . Gilt für $x_p \in \mathbb{R}$, dass

$$F(x_p) = p,$$

- dann heißt x_p **theoretisches p -Quantil** der Verteilung F .
- Wahrscheinlichkeitsmasse von 50% links von x_{50}



- Empirische Quantile
 - Schätzer von theoretischen Quantilen
- Beispiel

Gegeben sei eine exponentialverteilte Zufallsvariable $X \sim \text{Exp}(2)$.
Die entsprechende Verteilungsfunktion ist gegeben durch

$$F(x) = \begin{cases} 1 - e^{-2x} & x \geq 0, \\ 0 & \text{sonst.} \end{cases}$$

Bestimmen sie den Median $x_{0.5}$.

$$\begin{aligned}
 F(x_{0.5}) &\stackrel{!}{=} 0.5 \\
 \Leftrightarrow 1 - e^{-2x_{0.5}} &= 0.5 \\
 \Leftrightarrow -e^{-2x_{0.5}} &= -0.5 \\
 \Leftrightarrow e^{-2x_{0.5}} &= 0.5 \\
 \Leftrightarrow -2x_{0.5} &= \log(0.5) \\
 \Leftrightarrow x_{0.5} &= -\frac{\log(0.5)}{2} \approx 0.35
 \end{aligned}$$

Lokationsmaße

Mittelwert \bar{X}

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i,$$

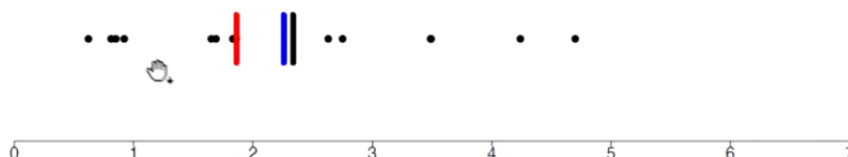
Median $\text{med} = Q_{0.5},$

α -getrimmtes Mittel \bar{X}_α für $\alpha \in (0, 0.5)$

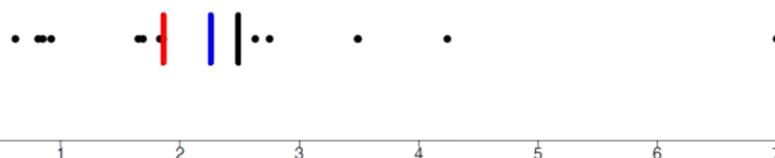
$$\bar{X}_\alpha = \frac{1}{n - 2\lfloor n\alpha \rfloor} \sum_{i=\lfloor n\alpha \rfloor + 1}^{n - \lfloor n\alpha \rfloor} X_{(i)},$$

Modalwert mode: Der am häufigsten vorkommende Wert in der Stichprobe (i.A. nicht eindeutig).

•



•



•

Abbildung: Grafische Veranschaulichung des CPU-Datensatzes mit Mittelwert (schwarz), 0.25-getrimmtem Mittel (blau) und Median (rot).

•

In der zweiten Abbildung wurde der Wert 4.70 durch 7.00 ersetzt.

Streuungsmaße

Standardabweichung S

$$S = \left(\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \right)^{1/2},$$

Likelihood Standardabweichung S_L

$$S_L = \left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \right)^{1/2},$$

iqr-Standardabweichung S_q

$$S_q = \frac{\text{iqr}}{1.349} = \frac{Q_{0.75} - Q_{0.25}}{1.349},$$

iqr ... Interquartilsabstand (interquartile range).

mad-Standardabweichung S_{mad}

$$\begin{aligned} S_{\text{mad}} &= \frac{1}{0.674} \text{mad} \\ &= \frac{1}{0.674} \text{med}\{|X_i - Q_{0.5}| : i = 1, \dots, n\}. \end{aligned}$$

mad ... median absolute deviation.

Schiefe

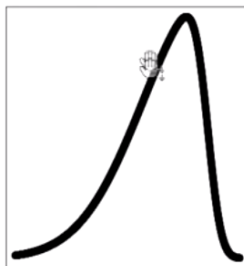
Das skalierte dritte zentrale Moment γ_3 einer Zufallsvariablen X nennt man **Schiefe**,

$$\gamma_3 = E \left[\left(\frac{X - \mu}{\sigma} \right)^3 \right].$$

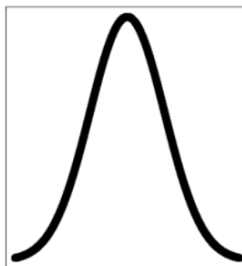
Die empirische Schiefe ist gegeben durch

$$V = \frac{1}{n} \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{S_L} \right)^3.$$

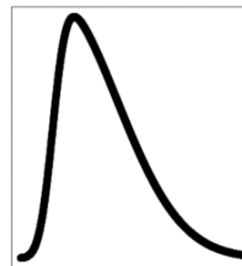
-
- beschreibt (A)Symmetrie



linksschief



symmetrisch



rechtsschief

-
- Normalverteilung
 - Schiefe null, somit symmetrisch
 - weicht die empirische Schiefe von null ab
 - Annahme der Normalverteilung womöglich falsch

Wölbung

Das skalierte vierte zentrale Moment γ_4 einer Zufallsvariablen X heißt **Wölbung** oder **Kurtosis**,

$$\gamma_4 = E \left[\left(\frac{X - \mu}{\sigma} \right)^4 \right].$$

Die empirische Wölbung ist

$$W = \frac{1}{n} \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{S_L} \right)^4.$$

-
- misst Ausprägung der Extrema
 - große Wölbung → heavy tails
- Normalverteilung

Die Wölbung einer Normalverteilung ist gleich 3. Man nennt daher

$$\gamma_4 - 3 = E \left[\left(\frac{X - \mu}{\sigma} \right)^4 \right] - 3$$

den **Exzess**. Der empirische Exzess $W - 3$ misst also die Abweichung der Stichprobe zur Normalverteilung.

–

Konsistenz

- unter Annahme des [[Gesetz der großen Zahlen]]

$$\bar{X}_n \xrightarrow{p} \mu,$$

$$S_n \xrightarrow{p} \sigma,$$

$$V_n \xrightarrow{p} \gamma_3,$$

$$W_n \xrightarrow{p} \gamma_4,$$

-
- \bar{X}_n , S_n , V_n und W_n **konsistente Schätzer** sind.