STAT

Lageparameter

Arithmetisches Mittel ("Durchschnitt") / .mean(): $\bar{x} = \frac{x_1 + x_2 + x_3 \dots x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$ ↗ mean sagt nichts über die Streuung der Daten aus
↳ Mittelwert

Median ("Mitte") = Daten sortieren
↗ Bei gerader Anzahl: mean() von den mittleren zwei
↘ Bei ungerader Anzahl: den Wert in der Mitte
↳ Vorteil: Robustheit bei extremen Beobachtungen
Trick: $k = np.round(0.5 \cdot n + 0.5) - 1$
pandas: median()

Quantile
Unteres Quartil: Wert, wo 25% kleiner oder gleich und 75% grösser oder gleich sind
Oberes Quartil: Wert, wo 75% kleiner oder gleich und 25% grösser oder gleich sind

xxx.quantile(q=.25 / .75, interpolation="midpoint") = q^{75}, q^{25} = xxx.quartile(q=[.75, .25], interpolation="midpoint")

Streuungsparameter:

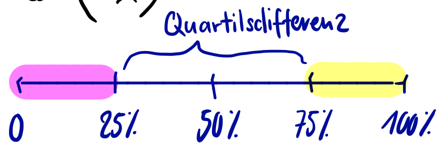
Empirische Varianz / Empirische Standardabweichung
↱ Pandas ⇒ .var() ↱ Pandas ⇒ .std()
↗ Durch $\sqrt{\ }$ kommen die Daten wieder in dieselbe Einheit.

$$Var(x) = \frac{(x_1-\bar{x})^2 + (x_2-\bar{x})^2 + \dots + (x_n-\bar{x})^2}{n-1} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

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

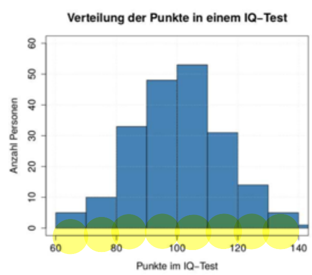
- Var(x) & s_x gross ⇒ so ist die Streuung der Messwerte gross. (+ $\bar{x}$)

Quartilsdifferenz: oberes Quartil − unteres Quartil

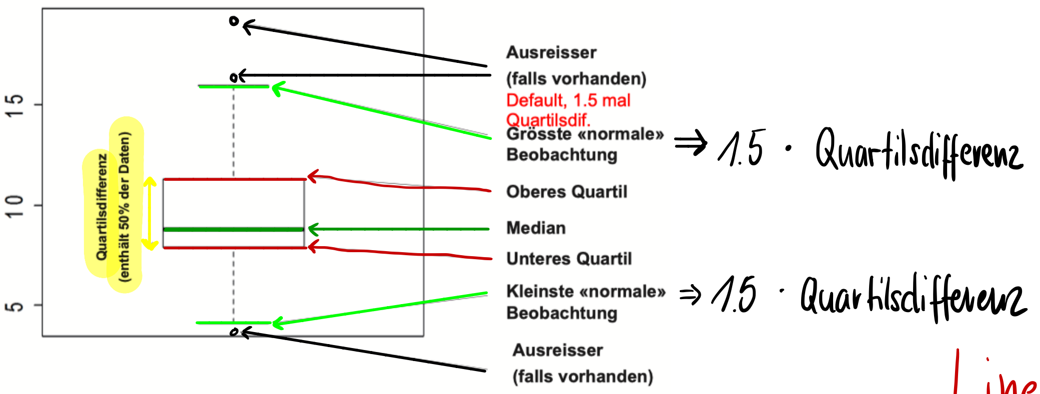


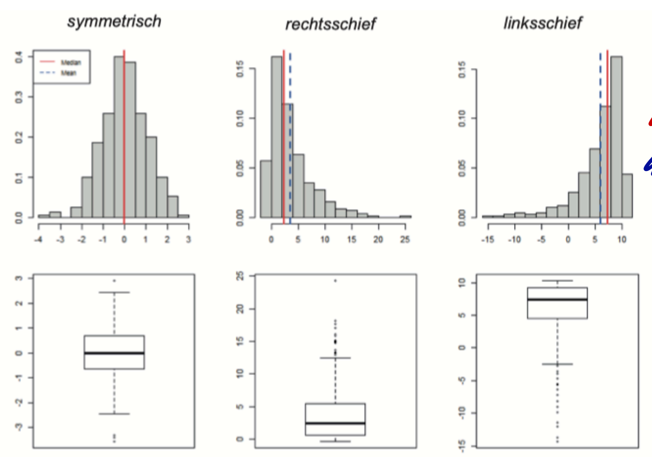
Graphische Darstellung

Histogramm: Aufteilung in n-Klassen (Intervall) <
< 50 Messungen = Klasse 5 bis 7
> 250 Messungen = Klasse 10 bis 20



Boxplot



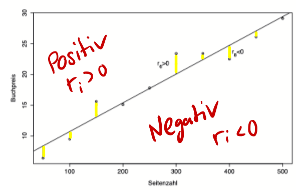


Schiefe von Daten

rot = Median
blau = Mean

Lineare Regression

Residumm: Abstand der Messpunkte zur Regressionsgerade. $r_i = y_i - a - bx_i$



$$r = \sum_{i=1}^{n}(r_i)^2$$

a+b von linearer Regression

$$b = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$a = \bar{y} - b\bar{x}$$

Empirische Korrelation

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{(\sum_{i=1}^{n}(x_i - \bar{x})^2) \cdot (\sum_{i=1}^{n}(y_i - \bar{y})^2)}}$$

Numerische Zusammenfassung der lin. Abhängigkeiten von zwei Grössen.

eine Zahl $r \Rightarrow [-1;1]$

$r = +1 \Rightarrow$ dann liegen Punkte auf steigender Geraden $b > 0$

$r = -1 \Rightarrow$ dann liegen Punkte auf fallender Geraden $b < 0$

$r = 0 \Rightarrow$ x und y unabhängig, keinen Zusammenhang

Stetige W'keitsverteilung

Jass bsw: $\omega = As \Rightarrow X(\omega) = 11 \Rightarrow x = 11$

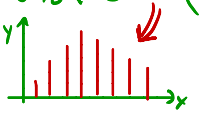
Zufallsvariable X ist eine Funktion (kann auch Y, Z sein)

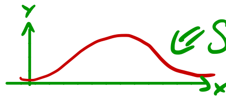
Elementarereignis ω

Zahlenwert / konkreten Wert : x (x eine Realisierung der Zufallsvariable (X))

Grundraum Ω

diskret (endlich abzählbare Menge)





stetig

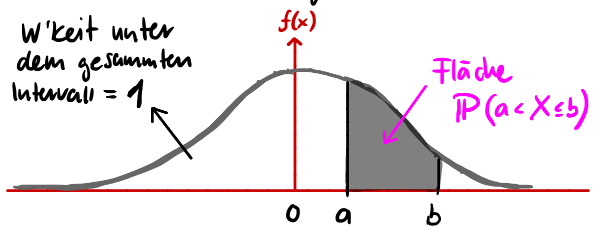
Die Werte einer Zufallsvariablen X (die möglichen Realisationen von X) treten mit gewissen Wahrscheinlichkeiten auf. Die Wahrscheinlichkeit, dass X den Wert x annimmt, berechnet sich wie folgt:

$$P(X = x) = P(\{\omega \mid X(\omega) = x\}) = \sum_{\omega; X(\omega) = x} P(\omega)$$

W'keit von allen Realisierungen $\Rightarrow$ W'keitsverteilung $\sum_{\text{alle möglichen } x} P(X = x) \Rightarrow 1$

Eigenschaften Wahrscheinlichkeitsdichte

1. $f(x) \geq 0$ für alle x (da $F(\cdot)$ monoton wachsend ist)
2. $P(a < X \leq b) = F(b) - F(a) = \int_a^b f(x)dx$ (Fläche zwischen a und b unter $f(x)$)
3. $\int_{-\infty}^{\infty} f(x)dx = 1$ (wegen 2.)



Erwartungswert : $E(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx$

Varianz : $Var(X) = \sigma_X^2 = E(X^2) - E(X)^2$

Kumulative Verteilung : $P(X \leq x) \Rightarrow$ cdf