Stochastik Zusammenfassung

Jorit Geurts, jgeurts@student.ethz.ch

Version: 24. Januar 2022

1 Grundlagen

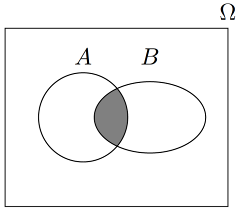
1.1 Grundbegriffe

- **Elemntarereigniss** ω: *Ein* möglicher Ausgang eines Zufallexperiment
- **Grundraum** Ω: Die Menge *aller* möglichen Elementarereignisse
- **Ereignis** $A \subseteq \Omega$: Eine Kollektion von *gewissen* Elementarereignissen
- **Norm** $|\Omega|$: Anzahl Elemente in Ω.

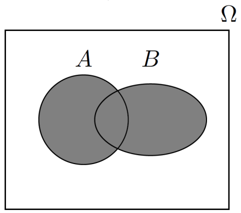
1.1.1 Venn Diagramme

Durchschnitt und Vereinigung

Durchschnitt (Schnitt):
$A \cap B$, A und B

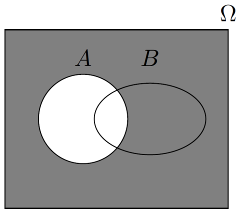


Vereinigung:
$A \cup B$, A und/oder B

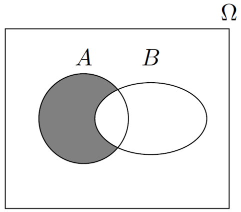


Komplement und Differenz

Komplement:
A^c, nicht A



Differenz:
$A \setminus B = A \cap B^c$, A ohne B



De Morgan'sche Regeln

$$(A \cap B)^c = A^c \cup B^c$$
$$(A \cup B)^c = A^c \cap B^c$$

Disjunkt

A und B sind **disjunkt** wenn die Schnittmenge die leere Menge ist. Sprich sie können nicht zusammen auftreten.

$$A \cap B = \emptyset$$

1.2 Wahrscheinlichkeit

Wahrscheinlichkeit dass ein Ereignis eintritt:

$$\mathbb{P}(A) = \text{Wahrscheinlichkeit}$$

1.2.1 Interpretationsarten

Frequentistische Interpretation:
Idealisierung der relativen Häufigkeit bei vielen unabhängingen Wiederhohlungen.

Bayes'sche Interpretation:
Mass für den Glauben dass ein Ereignis eintreten wird.

1.3 Axiome (Kolomogrov)

(A1) $0 \leq \mathbb{P}(A) \leq 1$

(A2) $\mathbb{P}(\Omega) = 1$

(A3) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$, falls A und B disjunkt

$\mathbb{P}(A_1 \cup A_2 \cup \ldots) = \sum_{i \geq 1} \mathbb{P}(A_i)$, für $A_k \cap A_l = \emptyset, k \neq l$

1.3.1 Rechenregeln

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \underbrace{\mathbb{P}(A \cap B)}_{=0, A \cap B = \emptyset}$$

$$\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$$

$$\mathbb{P}(A_1 \cup \ldots \cup A_n) \leq \sum_{i=1}^{n} \mathbb{P}(A_i)$$

$$\mathbb{P}(B) \leq \mathbb{P}(A), \quad \forall B \subseteq A \subseteq \Omega$$

$$\mathbb{P}(A \setminus B) = \mathbb{P}(A) - \mathbb{P}(B), \quad \forall B \subseteq A \subseteq \Omega$$

2 Diskrete Wahrscheinlickeitsmodelle

2.1 Unabhängigkeit von Ereignissen

A und B sind stochastisch Unabhängig, wenn:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$$

Allgemeiner Fall: alle Kombinationen müssen Unabhängig sein.

$$\mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \ldots \mathbb{P}(A_{i_k})$$

2.2 Bedingte Wahrscheinlichkeit

Bedingte Wahrscheinlichkeit: ($A|B$ = A gegeben B)

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

2.3 Rechenregeln:

$0 \leq \mathbb{P}(A \mid B) \leq 1, \qquad \mathbb{P}(B \mid B) = 1$

$\mathbb{P}(A_1 \cup A_2 \mid B) = \mathbb{P}(A_1 \mid B) + \mathbb{P}(A_2 \mid B)$, A_1, A_2 disjunkt

$\mathbb{P}(A^c \mid B) = 1 - \mathbb{P}(A \mid B)$

$\mathbb{P}(A \cap B) = \mathbb{P}(A \mid B)\,\mathbb{P}(B) = \mathbb{P}(B \mid A)\,\mathbb{P}(A)$

Für A, B unabhängig gilt:
$\mathbb{P}(A \mid B) = \mathbb{P}(A), \mathbb{P}(B|A) = \mathbb{P}(B)$

i.A. gilt:
$\mathbb{P}(A \mid B) \neq \mathbb{P}(B \mid A)$; $\mathbb{P}(A \mid B^c) \neq 1 - \mathbb{P}(A \mid B)$

2.4 Satz von Bayes

Für zwei Ereginisse A und B mit $\mathbb{P}(A), \mathbb{P}(B) > 0$ gilt:

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A)}$$

Kombination mit dem Satz der totalen Wahrscheinlichkeit

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i)\mathbb{P}(B_i)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B_i)\mathbb{P}(B_i)}{\sum_{i=1}^{k} \mathbb{P}(A|B_i)\mathbb{P}(B_i)}$$

2.5 Satz der totalen Wahrscheinlichkeit

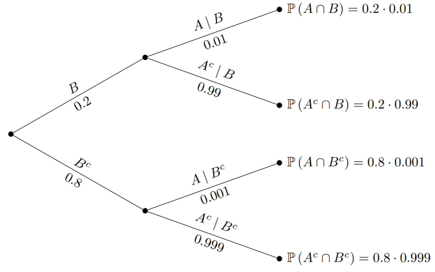
Wir haben k disjunkte Ereignisse $B_1, \ldots, B_k$.

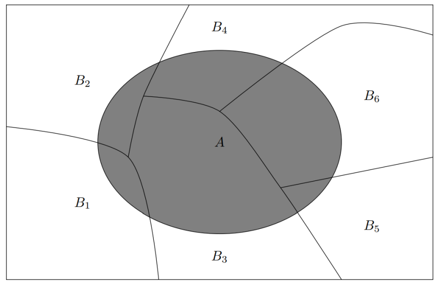
$$B_1 \cup \ldots \cup B_k = \Omega$$

Dann gilt

$$\mathbb{P}(A) = \sum_{i=1}^{k} \mathbb{P}(A \cap B_i) = \sum_{i=1}^{k} \mathbb{P}(A|B_i)\mathbb{P}(B_i)$$

Graphische Interpretation:





2.6 Definition

Ω besitzt endlich viele oder unendlich abzählbare Elemente.

$$\Omega = \{\omega_1, \omega_2, \ldots\}$$

Alle Elemente sind per Definition **disjunkt**.

$$\mathbb{P}(A) = \sum_{k:\omega_k \in A} \mathbb{P}(\{\omega_k\}), \quad \rightarrow \quad \mathbb{P}(\Omega) = \sum_{\Omega} \mathbb{P}(\{\omega_k\}) = 1$$

2.7 Laplace Modell

Alle Elementarereignisse sind *gleich wahrscheinlich*

$$\mathbb{P}(\{\omega_k\}) = \frac{1}{|\Omega|}$$

$$\mathbb{P}(A) = \frac{\#\text{ günstige Fälle}}{\#\text{ mögliche Fälle}}$$

$$\mathbb{P}(A) = \sum_{k:\omega_k \in A} \mathbb{P}(\{\omega_k\}) = \sum_{k:\omega_k \in A} \frac{1}{|\Omega|} = \frac{|A|}{|\Omega|}$$

3 Wahrscheindlichkeitsverteilungen

- **Zufallsvariable:** $X \in \mathbb{R}$
- X ist im voraus nicht bekannt
- **Realisierter Wert:** $x \in \mathbb{R}$
- **Zufallsvariable X nimmt den Wert x an:** $\{X = x\}$
- **Wertebereich:** $W = \Omega \subseteq \mathbb{R}$

Unabhängigkeit

Zwei Zufallsvariablen sind unabhängig wenn gilt:

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$$

$\{X \in A, Y \in B\} = \{X \in A\} \cap \{Y \in B\}$.

3.1 Diskrete Verteilungen

Eine Zufallsvariable X heisst **diskret**, wenn ihr Wertebereich W endlich oder unendlich abzählbar ist.

$$W = \{x_1, x_2, \ldots\}$$

Beispiel:

$$W = \mathbb{N}_0 = \{0, 1, 2, \ldots\}, \quad W = \{0, 1, 2, \ldots, 100\}$$

Wahrscheinlichkeitsfunktion: $p(x_k)$

$$p(x_k) = \mathbb{P}(X = x_k), \quad k \geq 1$$

$$\sum_{W} = 1$$

$$\mathbb{P}(X \in A) = \sum_{k:x_k \in A} p(x_k)$$

Kumulative Verteilungsfunktion: $F(x)$

$$F(x) = \mathbb{P}(X \leq x), \quad x \in \mathbb{R}$$

Rechenregeln

$\mathbb{P}(a < X \leq b) = \mathbb{P}(X \in (a, b])$
$= \mathbb{P}(X \in (-\infty, b]) - \mathbb{P}(X \in (-\infty, a]) = F(b) - F(a)$

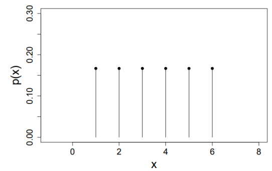
$\mathbb{P}(X > x) = 1 - F(x)$

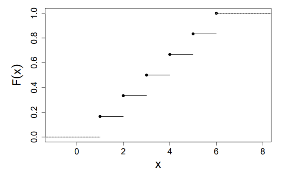
"$\leq$" oder "$<$" macht einen Unterschied

F ist monoton steigend

$\lim_{x \to -\infty} F(x) = 0$; $\lim_{x \to \infty} F(x) = 1$

F ist rechts-stetig, d.h. $\lim_{x \downarrow a} F(x) = F(a)$





3.1.1 Kennzahlen

Erwartungswert:

$$\mu_X := \mathbb{E}[X] := \sum_{k \geq 1} x_k p(x_k)$$

Transformation: $Y = g(X), \quad g : \mathbb{R} \to \mathbb{R}$

$$\mathbb{E}[Y] = \mathbb{E}[g(x)] = \sum_{k \geq 1} g(x_k)\, p(x_k)$$

Varianz:

$$\sigma_X^2 := \text{Var}(X) := \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \sum_{k \geq 1} (x_k - \mu_X)^2\, p(x_k)$$

Standartabweichung:

$$\sigma_X := \sqrt{\text{Var}(X)}$$

3.1.2 Verteilungen

Bernoulli Verteilung: $X \sim$ Bernoulli(p), $\quad p \in (0,1)$

$$X = \begin{cases} 1 & \text{Wahrscheinlichkeit } p \\ 0 & \text{Wahrscheinlichkeit } 1-p \end{cases}$$

$$\mathbb{E}[X] = p, \quad \text{Var}(X) = p \cdot (1-p)$$

T_1 und T_2 unabhängig und beide p

$$\min(T_1, T_2) \sim \text{Bernoulli}(p^2)$$

$X \sim$ Bernoulli(p) dann ist $X^n \sim$ Bernoulli(p)

Binominalverteilung: $X \sim$ Bin(n,p), $\quad n \in \mathbb{N}, p \in (0,1)$

Verteilung der Anzahl "Erfolge" bei n (unabhängigen) Wiederholungen und "Erfolgswahrscheinlichkeit" p.

$$\mathbb{P}[X = x] = p(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x \in W$$

$$\mathbb{P}[X \leq x] = \sum_{k=0}^{x} \mathbb{P}[X = k] = \sum_{k=0}^{x} p(k)$$

Binominalkoeffizient:

$$\binom{n}{x} = \frac{n!}{x!(n-x)!} \overset{TR}{=} n \text{ nCr } x$$

$$\mathbb{E}[X] = np, \quad \text{Var}(X) = n \cdot p \cdot (1-p)$$

Fakten:
$X_1 \sim$ Bin(n_1, p) und $X_2 \sim$ Bin(n_2, p) unabhängig

$$X_1 + X_2 \sim \text{Bin}(n_1 + n_2, p)$$

Geometrische Verteilung: $X \sim$ Geom(p), $\quad p \in (0,1)$

Anzahl Wiederholungen von unabhängigen Bernoulli(p) Experimenten bis zum ersten Erfolg. $p(x) = p(1-p)^{x-1}$;

$$p(x) = p \cdot (1-p)^{x-1}, \quad W = \mathbb{N}$$

$$\mathbb{E}[X] = \frac{1}{p}, \quad \text{Var}(X) = \frac{1-p}{p^2}$$

Kumulative Verteilungsfunktion:

$$F(x) = \sum_{i=1}^{x} p(1-p)^{i-1} = 1 - (1-p)^x$$

C, V unabhängig mit $\sim$ Geom(p).

$$\mathbb{P}[C \cup V] \sim \text{Geom}(2p - p^2)$$

Gedächtnisslos:

$$\mathbb{P}[X > s + t | Y > s] = \mathbb{P}[X > t]$$

Poissonverteilung: $X \sim$ Pois(λ), $\quad \lambda > 0$

$$p(x) = e^{-\lambda} \frac{\lambda^x}{x!}, \quad x \in W$$

$$\mathbb{E}[X] = \lambda, \quad \text{Var}(X) = \lambda$$

Summe zweier unabhänigen Poissonverteilungen:

$$X \sim \text{Pois}(\lambda_1), Y \sim \text{Pois}(\lambda_2) \Rightarrow X + Y \sim \text{Pois}(\lambda_1 + \lambda_2)$$

(*Achtung:* $\frac{1}{2}(X+Y) \sim$ Pois $\left(\frac{1}{2}(\lambda_1 + \lambda_2)\right)$ stimmt nicht!)
(*Achtung:* $(X - Y) \sim$ Pois $((\lambda_1 - \lambda_2))$ stimmt nicht!)

Poissonapproximation der Binominalverteilung:

$$X \sim \text{Bin}(n,p) \approx \text{Pois}(\lambda = np), \quad p \ll 1 \ll n$$

Summe:

$$\mathbb{P}[X_1 + X_2 = K | X_2 = T] = \mathbb{P}[X_1 = K - T]$$

3.2 Stetige Verteilungen

Eine Zufallsvariable X heisst **stetig**, wenn ihr Wertebereich W aus einem oder mehreren Intervallen besteht.
Beispiel:

$$W = [0,1], \quad W = \mathbb{R}$$

Wahrscheinlichkeitsdichte (Dichte): $f(x) > 1$ möglich

$$f(x) = \lim_{h \to 0} \frac{\mathbb{P}(x < X \leq x + h)}{h} = F'(x)$$

Approximation:

$$\mathbb{P}(x > X \leq x + h) \approx h f(x), \quad h \ll 1$$

Kumulative Verteilungsfunktion: $F(x)$

$$F(x) = \mathbb{P}[X \leq x] = \int_{-\infty}^{x} f(u) du$$

$$\tilde{F}(x) = \mathbb{P}[X \geq x] = \int_{x}^{\infty} f(u) du$$

Weitere Zusammenhänge:

$\mathbb{P}(a < X \leq b) = F(b) - F(a) = \int_a^b f(x)dx$

$\int_{-\infty}^{\infty} f(x)dx = 1$, $\mathbb{P}[X = x] = 0$, $f(x) \geq 0$

3.2.1 Kennzahlen

Erwartungswert:

$$\mathbb{E}[X] = \mu_X = \int_{-\infty}^{\infty} x f(x) dx$$

Transformation: $Y = g(X): \quad \mathbb{R} \to \mathbb{R}$

$$\mathbb{E}[Y] = \mathbb{E}[g(x)] = \int_{-\infty}^{\infty} g(x) f(x) dx$$

Varianz:

$$\text{Var}(X) = \sigma_x^2 = \mathbb{E}[(X - \mu_X)^2] = \int_{-\infty}^{\infty} (x - \mu_X)^2 f(x) dx$$

Gleiche Rechenregeln wie im diskreten Fall!!!

Quantile:

Das $(\alpha \cdot 100\%)$ Quantil q_α beschreibt den Wert, der mit der Wahrscheinlichkeit $(\alpha \cdot 100\%)$ unterschritten wird.

$$\alpha = \mathbb{P}(X \leq q_\alpha) = F(q_\alpha), \quad \alpha \in (0,1) \quad \to \quad q_\alpha = F^{-1}(\alpha)$$

Median: 50% Quantil.
Achtung q_α müssen im Bereich der kumulativen Verteilungsfunktion sein. Daher immer die Option wählen die das Erfüllt.

3.2.2 Verteilungen

Uniforme Verteilung: $X \sim$ Uni(a,b), $\quad a < b \in \mathbb{R}$

$$f(x) = \begin{cases} \frac{1}{b-a} & x \in [a,b] = W \\ 0 & \text{sonst} \end{cases}$$

$$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & x \in [a,b] \\ 1 & x > b \end{cases}$$

$$\mathbb{E}[X] = \frac{a+b}{2}, \quad \text{Var}(X) = \frac{(b-a)^2}{12}$$

Fakten:
$X \sim$ Uni$(a,b): c + d \cdot X \sim$ Uni$(c + da, c + db)$

Exponentialverteilung: $X \sim$ Exp(λ), $\quad \lambda > 0$

Stetige Version der Geometrischen Verteilung. λ = Ausfallrate

$$f(x) = \begin{cases} 0 & x < 0 \\ \lambda e^{-\lambda x} & x \geq 0 \end{cases}$$

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0 \end{cases}$$

$$\mathbb{E}[X] = \frac{1}{\lambda}, \quad \text{Var}(X) = \frac{1}{\lambda^2}$$

Fakten:
$X \sim$ Exp$(\lambda): c \cdot X \sim$ Exp $\left(\frac{\lambda}{c}\right)$
$\mathbb{E}[XY] = \frac{1}{\lambda_1 \lambda_2}$, wenn unabhängig
Gedächtnisslos: $\mathbb{P}[X \leq R | X > T] = \mathbb{P}[X \leq R - T]$

Normalverteilung: $X \sim \mathcal{N}(\mu, \sigma^2)$, $\quad \mu \in \mathbb{R}, \sigma > 0$

Häufigste Verteilung von Messwerten.

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{x - \mu}{\sigma}\right)^2\right), \quad x \in \mathbb{R}$$

$$\mathbb{E}[X] = \mu, \quad \text{Var}(X) = \sigma^2$$

Summe: X_1 und X_2 sind unabhängig

$$X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

Standartnormalverteilung: $X \sim \mathcal{N}(0,1)$

$$f(x) = \varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

$$F(x) = \Phi(x) = \int_{-\infty}^{x} \varphi(u) du$$

Werte sind **Tabelliert**.
Quantile:

$$z_\alpha = \Phi^{-1}(\alpha), \quad \alpha \in (0,1)$$

Verteilungsfunktion einer Normalverteilung:

$$F(x) = \Phi\left(\frac{x - \mu}{\sigma}\right), \quad x \in \mathbb{R}$$

Falls α nicht im Wertebereich ist:

$$\Phi(A) \leq \alpha \to 1 - \Phi(-A) \leq \alpha$$

Standardisierung:

Transformation: $\mu = 0$ und Var $= 1$ erreichen

$$Z = g(x) = \frac{x - \mu_x}{\sigma_x}$$

$$\mathbb{E}[Z] = 0, \quad \text{Var}(Z) = 1$$

3.2.3 Transformationen

Linearer Fall: $g(x) = a + bx$

Dichte: $b \neq 0$

$$f_Y(y) = \frac{1}{|b|} f_X\left(\frac{y - a}{b}\right)$$

Verteilungsfunktion: $b > 0$

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(a + bX \leq y)$$
$$= \mathbb{P}\left(X \leq \frac{y-a}{b}\right) = F_X\left(\frac{y-a}{b}\right)$$

Verteilungsfunktion: $b < 0$

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(a + bX \leq y)$$
$$= \mathbb{P}\left(X \geq \frac{y-a}{b}\right) = 1 - F_X\left(\frac{y-a}{b}\right)$$

Allgemeiner streng monotoner Fall: $g(x)$ differenzierbar

Dichte:

$$f_Y(y) = \left|\frac{1}{g'(g^{-1}(y))}\right| f_X\left(g^{-1}(y)\right)$$

Wertebereich:

$$W_Y = g(W_X) = \{g(x), x \in W_X\}$$

Erwartungswert transformiert nicht mit, Jensen'sche Ungleichung:
Für g konvex $g'' \geq 0$.

$$\mathbb{E}[g(X)] \geq g(\mathbb{E}[X])$$

Quantile Transformieren mit:

$$\alpha = \mathbb{P}(X \leq q_\alpha) = \mathbb{P}(g(X) \leq g(q_\alpha)) = \mathbb{P}(Y \leq g(q_\alpha))$$

3.3 Simulation von Zufallsvariablen

$U \sim$ Uni$(0,1)$, F beliebig und $X = F^{-1}(U)$ dann gilt:

$$\mathbb{P}(X \leq x) = \mathbb{P}(F^{-1}(U) \leq x) = \mathbb{P}(U \leq F(x)) = F(s)$$

Vorgehen:

1. Kreiere $U \sim$ Uni$(0,1)$
2. Berechne $x = F^{-1}(u)$. u ist dann mit F verteilt.

4 Rechenregeln $\mathbb{E}(X)$ & Var(X)

Rechenregeln:

$\mathbb{E}[a + bX + cY] = a + b\mathbb{E}[X] + c\mathbb{E}[Y]$, $a, b, c \in \mathbb{R}$

Var$(x) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$

Var$(a + bX) = b^2$Var(X)

Var$(a) = 0$, $\quad a \in \mathbb{R}$

$\mathbb{E}[a + bX + cY] = a + b\mathbb{E}[X] + c\mathbb{E}[Y]$

Unabhängig:
Var$(XY) = \mathbb{E}[X]^2\mathbb{V}[Y] + \mathbb{E}[Y]^2\mathbb{V}[X] + \mathbb{V}[X]\mathbb{V}[Y]$

Kovarianz und Korrelation

Var$(X) =$ Cov(X, X)

Unabhängigkeit:

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y], \quad \mathbb{E}[X/Y] \neq \mathbb{E}[X]/\mathbb{E}[Y]$$

Bilinearität:

$$\text{Cov}\left(\sum_{i=1}^{n} a_i X_i, \sum_{j=1}^{m} b_j Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j \text{Cov}(X_i, Y_j)$$

Symmetrie:
Cov$(X, Y) =$ Cov(Y, X)

Cov$(a + bX, c + dY) = bd$Cov(X, Y)

Corr$(a + bX, c + dY) =$ sign(b) sign(d) Corr(X, Y)

Var$\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n}$ Var$(X_i) + 2\sum_{i<j}^{n}$ Cov(X_i, Y_j)
$n = 2$: Var$(X + Y) =$ Var$(X) +$ Var$(Y) + 2$Cov(X, Y)

Für $X_1, \ldots, X_n$ unabhängig gilt:
Var$\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n}$ Var(X_i)

5 Deskriptive Statistik

Daten Darstellen und Zusammenfassen.

5.1 Kennzahlen

Arithmeitshces Mittel: (Durchschnitt, Mittelwert)

$$\bar{x} = \frac{1}{n}(x_1 + \ldots + x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Emprirische Varianz:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{x})^2$$

Empirische Standardabweichung:

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{x})^2}$$

Geordnete Werte: Aufsteigende Reihenfolge der Daten

$$x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)}$$

Klammer bedeutet Geordnet.
Emprirische Quantile:

$$q_\alpha = \begin{cases} \frac{1}{2}\left(x_{(\alpha \cdot n)} + x_{\alpha \cdot n+1}\right) & \text{falls } \alpha \text{ eine ganze Zahl ist} \\ x_{(\lceil \alpha \cdot n \rceil)} & \text{sonst} \end{cases}$$

$\lceil \cdot \rceil$ = Aufrunden
Empririscher Median: 50% Quantil

$$q_{0.5} = \begin{cases} \frac{1}{2}\left(x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}\right) & n \text{ gerade} \\ x_{\left(\frac{n+1}{2}\right)} & n \text{ ungerade} \end{cases}$$

Quartile:
unteres Quartil: $q_{0.25}$
oberes Quartil: $q_{0.75}$
Quartildifferenz: $q_{0.75} - q_{0.25}$

5.2 Graphische Darstellung

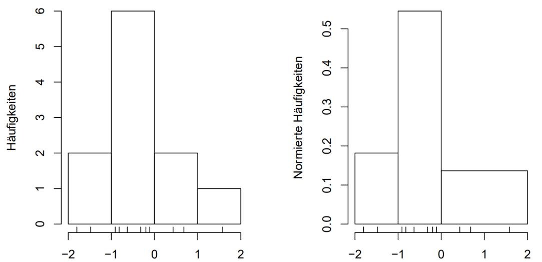
5.2.1 Histogramm

Einteilung der Daten in Klassen (Intervalle) $(c_{k-1}, c_k]$.
Ermitteln der Anzahl Beobachtungen im Intervall/Klasse h_k.
Höhe der Balken proportional zu

$$H = \frac{h_k}{c_k - c_{k-1}}$$

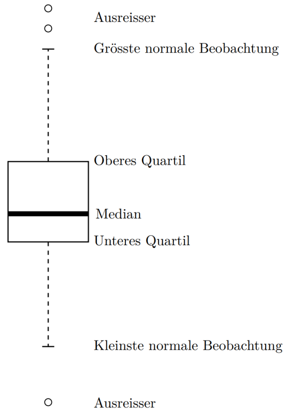
Mögliche Fausteregel zur Wahl der Anzahl Klassen: (Sturges Rule)

$$N = \lceil 1 + \log_2(n) \rceil$$



Guter Überblick über die Verteilung.

5.2.2 Boxplot

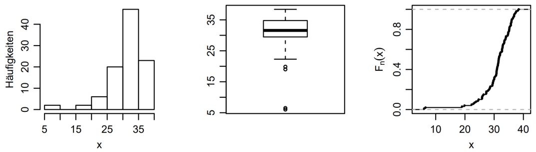


Normale Beobachtung: Maximal 1.5-Mal Quartalsdifferenz von den Quartilen entfernt

Emprische kumulative Verteilungsfunktion

$$F_n(x) = \frac{1}{n}\text{Anzahl}\{i | x_i \le x\} \in [0, 1]$$

Darstellung der Daten (rechts: kumulative Verteilungsfunktion):



5.3 Mehrere Messgrössen

Es liegen Datenpaare vor $(x_1, y_1), \ldots, (x_n, y_n)$.
Am Zusammenhang der Datenpaare interessiert.
Emprirische Kovarianz:

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

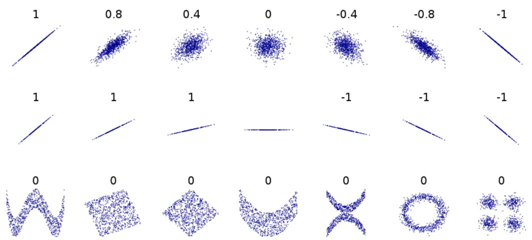
Emprirische Korrelation:

$$r = \frac{s_{xy}}{s_x s_y} \in [-1, 1]$$

Interpretation der empirischen Korrelation:

$$r = +1 \Leftrightarrow y_i = a + bx_i, \quad a \in \mathbb{R}, b > 0$$
$$r = -1 \Leftrightarrow y_i = a + bx_i, \quad a \in \mathbb{R}, b < 0$$

Nur **Linearer** Zusammenhang wird geprüft. Daher sollte man neben der Berechnung auch immer die Plots anschauen.



6 Merhdimensionale Verteilungen

6.1 Verteilungen

6.1.1 Diskreter Fall

Gemeinsame Wahrscheinlichkeitsfunktion:

$$\mathbb{P}(X = x, Y = y), \quad x \in W_X, y \in W_Y$$

Randverteilung: (Satz der totalen Wahrscheinlichkeit)

$$\mathbb{P}(X = x) = \sum_{y \in W_Y} \mathbb{P}(X = x, Y = y), \quad x \in W_X$$

$$\mathbb{P}(Y = y) = \sum_{x \in W_X} \mathbb{P}(X = x, Y = y), \quad y \in W_Y$$

Gesamte Verteilung im Unabhängigen Fall:

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y), x \in W_X, y \in W_Y$$

Bedingte Verteilung: UA = Unabhängig

$$\mathbb{P}(X = x | Y = y) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)} \stackrel{UA}{=} \mathbb{P}(X = x)$$

$$\mathbb{P}(Y = y | X = x) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(X = x)} \stackrel{UA}{=} \mathbb{P}(Y = y)$$

Randverteilung aus bedingter Verteilung:

$$\mathbb{P}(X = x) = \sum_{y \in W_Y} \mathbb{P}(X = x | Y = y)\mathbb{P}(Y = y), x \in W_X$$

$$\mathbb{P}(Y = y) = \sum_{x \in W_X} \mathbb{P}(Y = y | X = x)\mathbb{P}(x = x), y \in W_Y$$

Bedingter Erwartungswert:

$$\mathbb{E}[Y | X = x] = \sum_{y \in W_Y} y \, \mathbb{P}(Y = y | X = x)$$

$$\mathbb{E}[X | Y = y] = \sum_{x \in W_X} x \, \mathbb{P}(X = x | Y = y)$$

6.1.2 Stetiger Fall

Gemeinsame Dichte:

$$\mathbb{P}\left(X \in [x, x + dx], Y \in [y, y + dy]\right) = f_{X,Y}(x, y)\, dxdy$$

Gemeinsame Wahrscheinlichkeit:

$$\mathbb{P}\left((X, Y) \in A\right) = \iint_A f_{X,Y}(x, y)\, dxdy$$

Unabhängigkeit:

$$f_{X,Y}(x, y) = f_X(x)\, f_Y(y)$$

Randdichte:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx$$

Bedingte Verteilung: UA = Unabhängig

$$f_{Y|X=x}(y) = f_Y(y \mid X = x) = \frac{f_{X,Y}(x, y)}{f_X(x)} \stackrel{UA}{=} f_Y(y)$$

$$f_{X|Y=y}(x) = f_X(x \mid Y = y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} \stackrel{UA}{=} f_X(x)$$

Bedingter Erwartungswert:

$$\mathbb{E}[Y \mid X = x] = \int_{-\infty}^{\infty} y \, f_{Y|X=x}(y)\, dy$$

$$\mathbb{E}[X \mid Y = y] = \int_{-\infty}^{\infty} x \, f_{X|Y=y}(x)\, dx$$

Unabhängigkeit Folgen:

$$f_{Y|X=x}(y) = f_Y(y), \quad \text{bzw.} \quad f_{X|Y=y} = f_X(x)$$

$$F_{X,Y}(x, y) = f_{Y|X=x}(y) f_X(x) = f_{X|Y=y} f_Y(y)$$

6.2 Erwartungswert

Transformation: $Z = g(X, Y): \quad \mathbb{R}^2 \to \mathbb{R}$
Stetiger Fall:

$$\mathbb{E}[g(X, Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y) f_{X,Y}(x, y) dxdy$$

Diskreter Fall:

$$\mathbb{E}[g(X, Y)] = \sum_{x \in W_X}\sum_{y \in W_y} g(x, y)\mathbb{P}(X = x, Y = y)$$

Linear Kombination:

$$\mathbb{E}\left[a_0 + \sum_{i=1}^{n} a_i X_i\right] = a_0 + \sum_{i=1}^{n} a_i \mathbb{E}[X_i]$$

6.3 Kovarianz und Korrelation

Kovarianz:

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

Korrelation:

$$\text{Corr}(X, Y) = \rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}, \quad \in [-1, 1]$$

Interpretation der Korrelation: rein linearer Zusammenhang

$$\text{Corr}(X, Y) = +1 \Leftrightarrow Y = a + bX, \quad a \in \mathbb{R}, b > 0$$
$$\text{Corr}(X, Y) = -1 \Leftrightarrow Y = a + bX, \quad a \in \mathbb{R}, b < 0$$

X und Y unabhängig $\Rightarrow$ Corr $(X, Y) = 0 \Rightarrow$ unkorreliert

6.4 Zweidimensionale Normalverteilung

Definition:

$$f_{X,Y}(x, y) = \frac{\exp\left(-\frac{1}{2}(x - \mu_X \quad y - \mu_Y)\Sigma^{-1}\begin{pmatrix} x - \mu_X \\ y - \mu_Y \end{pmatrix}\right)}{2\pi\sqrt{\det(\Sigma)}}$$

Kovariazmatrix:

$$\Sigma = \begin{pmatrix} \text{Cov}(X, X) & \text{Cov}(X, Y) \\ \text{Cov}(Y, X) & \text{Cov}(Y, Y) \end{pmatrix}$$
$$= \begin{pmatrix} \sigma_X^2 & \text{Cov}(X, Y) \\ \text{Cov}(X, Y) & \sigma_Y^2 \end{pmatrix}$$

6.5 Dichte einer Summe

Sei $S = X + Y$ eine Zufallsvariable mit $X, Y \sim f_{X,Y}$.
Dichte von S

$$f_S(s) = \int_{-\infty}^{\infty} f_{X,Y}(x, s - x) dx$$

Unabhäniger Fall: (Faltung)

$$f_S(s) = \int_{-\infty}^{\infty} f_X(x) F_y(s - x) dx$$

7 Grenzwertsätze

7.1 Die i.i.d. Annahme

X_i: die i-te Wiederhohlung eins Zufallexperiments $i = 1, \ldots, n$
Alle X_i sind unabhängig und gleichverteilt.
independent and **i**dentically **d**istributed (i.i.d.)

7.2 Summen und arithmetische Mittel

Summe von n i.i.d. Zufallsvariablen:

$$S_n = X_1 + \ldots + X_n = \sum_{i=1}^{n} X_i$$

Arithmetisches Mittel von n i.i.d. Zufallsvariablen:

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n} S_n$$

Spezialfälle: $\bar{X}_n$ ist nicht stetig

1. $X_i \in \{0,1\} \Rightarrow S_n \sim \text{Bin}(n,p),\ p = \mathbb{P}(X_i = 1)$
2. $X_i \sim \text{Pois}(\lambda) \Rightarrow S_n \sim \text{Pois}(n\lambda)$
3. $X_i \sim \mathcal{N}(\mu, \sigma^2)$:
   $\Rightarrow S_n \sim \mathcal{N}(n\mu, n\sigma^2),\ \overline{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$

Erwartungswert und Varianz der Summe:

$$\mathbb{E}[S_n] = n\mathbb{E}[X_i], \quad \text{Var}(S_n) = n\text{Var}(X_i), \quad \sigma_{S_n} = \sqrt{n}\sigma_{X_i}$$

Erwartungswert und Varianz des arithmetischen Mittel:

$$\mathbb{E}[\overline{X}_n] = \mathbb{E}[X_i], \quad \text{Var}(\overline{X}_n) = \frac{1}{n}\text{Var}(X_i), \quad \sigma_{\overline{X}_n} = \frac{\sigma_{X_i}}{\sqrt{n}}$$

7.3 Gesetz der Grossen Zahlen

Seien X_i i.i.d. Zufallsvariablen mit Erwartungswert μ dann gilt:

$$\overline{X}_n \xrightarrow{n\to\infty} \mu$$

Relative Häufigkeit:

$$f_n(A) \xrightarrow{n\to\infty} \mathbb{P}(A)$$

7.4 Zentraler Grenzwertsatz

Seien X_i i.i.d. Zufallsvariablen mit Erwartungswert μ und Varianz σ^2 dann gilt:

$$S_n \approx \mathcal{N}(n\mu, n\sigma^2), \qquad \overline{X}_n \approx \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right), \quad n \gg 1$$

Für grosse n liegt approximativ eine Normalverteilung vor.
Wir verwenden dann:

$$\Phi\left(\frac{X - \mu_N}{\sqrt{N}\sigma_N}\right), \quad \Phi\left(\frac{\sqrt{N}(X-\mu)}{\sigma}\right)$$

Normalapproximation der Binominalverteilung:
Wenn normal Binomial verteilt:

$$X \sim \text{Bin}(n,p), \quad \mathbb{E}[X] = np, \quad \text{Var}(X) = np(1-p)$$

Für grosse n ($n \gg 1$) gilt dann:

$$\mathbb{P}(X \le x) = \Phi\left(\frac{x - np}{\sqrt{np(1-p)}}\right)$$

Erst Approximieren, dann Einsetzten:

$$\mathbb{P}[X \ge x] \approx 1 - \mathbb{P}_{\mathcal{N}}[X < x]$$

8 Parameterschätzung

8.1 Wahl der Verteilungsfamilie

Q-Q-Plot(Quantile-Quantile-Plot):

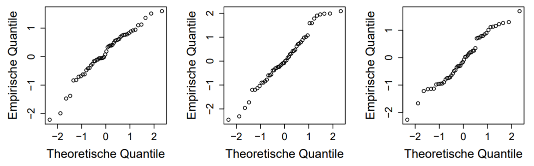
Vergleichen der Theoretischen und Empirischen Quantilen.
Betrachten der

$$\alpha_k = \frac{k - 0.5}{n}, k = 1, \ldots, n$$

empirische Quantile.
Die Beobachtung $x_{(k)}$ ist das entprechende Quantil.
Geplottet gegen die theoretischen Quantile:

$$q_\alpha = F^{-1}(\alpha_k)$$

$F(x)$ ist die kumulative Verteilungsfunktion.
Plot: $\{F^{-1}(\alpha_k), x_{(k)}\}, k = 1, \ldots, n$



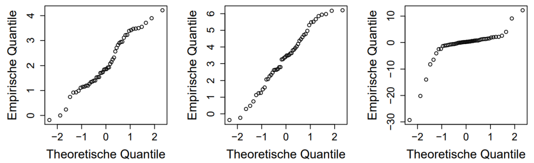
Wenn die Verteilungen übereinstimmt, liegen die Punkte ungefähr auf der **Winkelhalbierenden**.

Normalplot:

Q-Q-Plot mit der Standartnormalverteilung für $F(x) = \Phi(x)$. Wenn die Punkt auf einer Gerade liegen, dann sind die Daten

$$X \sim \mathcal{N}(\mu, \sigma^2), \quad \text{da} \quad q_\alpha = \mu + \sigma z_\alpha$$

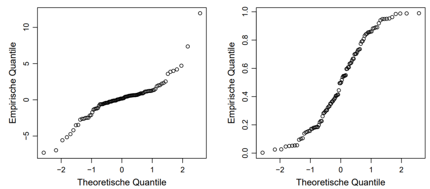
verteilt.
Achsenabschnitt der Gerade $= \mu$
Steigung der Gerade $= \sigma$



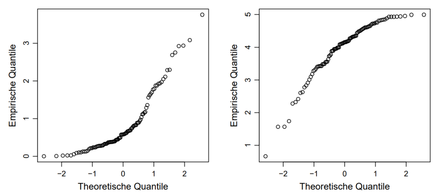
Links und Mitte erfüllt, Rechts nicht erfüllt.

Lang bzw. kurzschwänzige Verteilung:

Langschwänzig: Extremere Werte im Randbereich (links)
Kurzschwänzig: Moderatere Werte im Randbereich (rechts)



Schiefe Verteilung:



Wenn ausreisser Vorhanden, ist es nicht Normalverteilt.

8.2 Methoden der Parameterschätzung

Parameter:

Wir wollen den unbekannten Parametervektor finden.

$$\theta \in \mathbb{R}^r$$

Beispiele:
Normalverteilung:

$$\theta = \left(\mu, \sigma^2\right) \in \mathbb{R}^2$$

Poissonverteilung:

$$\theta = \lambda \in \mathbb{R}$$

Schätzer:

Möglicher "guter" Wert aus den Daten.

$$\hat{\theta} : \mathbb{R}^n \to \mathbb{R}^r, \quad \hat{\theta} = \hat{\theta}(x_1, \ldots, x_n)$$

$\hat{\theta}$ entspricht meist nie genau dem Wahren θ.

8.2.1 Momentenmethode

Das $k-$te Moment:

$$\mu_k = \mathbb{E}\left[X^k\right]$$

Das $k-$te empirische Moment:

$$m_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k$$

$m_2 = s^2 + \bar{x}^2$ Das empirische und theoretische Moment müssen jetzt gleich sein.
Gleichungssystem:

$$\mu_k(\hat{\theta}) = m_k, \quad k = 1, \ldots, r$$

Nicht immer die optimale Methode. Zudem kann es zu Unsinnigen Lösungen kommen.

8.2.2 Maximum-Likelihood Methode

Likelihoodfunktion:

Beschreibt die Wahrscheinlichkeit, dass das gemesse Ereignis tatsächlich eintritt (für gegebenes θ) (diskreter Fall).

$$L(\theta) = p_{X_1,\ldots,X_n}(x_1, \ldots, x_n \mid \theta) = \prod_{i=1}^{n} p_X(x_i \mid \theta) \stackrel{!}{=} \max$$

Diese Funktion wollen wir Maximieren, idem wir θ finden.
Uniformverteilung: Maximales Intervall erhalten!!!

Log-Likelihoodfunktion:

Da der Logarithmus monoton wachsend ist können wir den Logarithmus der Funktion nehmen und diesen maximieren.

$$l(\theta) = \log(L(\theta)) = \sum_{i=1}^{p} log(p(x_i|\theta))$$

Maximieren durch Ableiten und Null setzten:

$$\frac{\partial}{\partial\theta} l(\theta; x_i) = 0 \quad \rightarrow \quad \hat{\theta}$$

Für den stetigen Fall ersetzt man $p(x)$ durch $f(x)$.
Hut bei der Lösung nicht vergessen!

8.3 Allgeimene Schätzer für $\hat{\mu}$ und $\hat{\sigma}^2$

Allgemeines Schätzen des Erwartungswert und der Varianz.
Erwartungswert anhand des arithmetischen Mittels:

$$\hat{\mu}_X = \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Varianz anhand der empirischen Varianz:

$$\hat{\sigma}_X^2 = S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

Genauigkeit der Schätzer:
Erwartungswert ist Erwartungstreu:

$$\mathbb{E}[\hat{\mu}_X] = \mathbb{E}[X] = \mu$$

$$\text{Var}(\hat{\mu}_X) = \frac{1}{n}\sigma_X^2$$

Varianz ist Erwartungstreu:

$$\mathbb{E}[\hat{\sigma}_X^2] = \sigma_X^2$$

Beispiel

$Y \sim Bin(N, p^*),\ p^* = p(2x-1)1 - x),\ \mathbb{E}[Y] = Np^*$

$$\hat{p} = \frac{\frac{Y}{N} - 1 + x}{2x - 1}$$

Erwartungstreu:

$$\mathbb{E}[\hat{p}] = \frac{\frac{\mathbb{E}[Y]}{N} - 1 + x}{2x - 1} = p$$

wegen den Rechenregeln der Varianz gilt dann

$$\text{Var}[\hat{p}] = \frac{\text{Var}[Y]}{N^2(2x-1)^2} = \frac{Np^*(1-p^*)}{N^2(2x-1)^2}$$

8.4 Genauigkeit von Schätzern

Annahme: Daten x_i sind Normalverteilt $\mathcal{N}(\mu, \sigma^2)$.
Als Schätzer erhalten wir also:

$$\hat{\mu}_X = \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Der Schätzer ist auch Normalvertielt:

$$\hat{\mu}_X \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

Standardfehler = Standardabweichung vom Schätzer.

$$srd(\hat{\mu}) = \frac{\sigma}{\sqrt{n}}$$

Falls wir σ kennen dann liegt $\hat{\mu}$ mit Wahrscheinlichkeit 0.95 im Intervall

$$\mu \pm z_{0.975}\frac{\sigma}{\sqrt{n}}$$

mit $z_{0.975} = 97.5\%$-Quantil der Standardnormalverteilung
95%-**Vertauensintervall:** ($z_{0.975} = 1.96 \approx 2$)

$$I \approx \hat{\mu}_X \pm 2\frac{\sigma}{\sqrt{n}}$$

9 Statistische Test

9.1 Grundsätzliches Vorgehen

1. Wähle ein Modell für die Daten
2. Nullhypothese aufstellen: $H_0 : \theta = \theta_0$
3. Alternativhypothese aufstellen:

$$H_A : \begin{cases} \theta \neq \theta_0 & \text{zweiseitig} \\ \theta > \theta_0 & \text{einseitig nach oben} \\ \theta < \theta_0 & \text{einseitig nach unten} \end{cases}$$

4. Signifikatzniveau festlegen: $\alpha = 0.05, 0.01$
5. Verwerfungsbereich K Konstruieren so dass:

$$\mathbb{P}(\text{Fehler 1. Art}) \leq \alpha$$

6. Teststatistik auswerten und Entscheiden:
   - Beobachtung $x \in K \rightarrow$ Verwerfe H_0 für H_A
   - Beobachtung $x \notin K \rightarrow$ Behalte H_0

		Entscheidung H_0	Entscheidung H_A
Wahrheit	H_0	kein Fehler	Fehler 1. Art
Wahrheit	H_A	Fehler 2. Art	kein Fehler

Vorsicht: Falls man H_0 nicht verwierft heisst das nicht dass H_0 stimmt.

9.2 Binominaltest

1. Annahme: $X \sim \text{Bin}(n, p)$, n bekannt, p unbekannt
2. Nullhypothese: $H_0 : p = p_0$
3. Alternativhypothese:

$$H_A : \begin{cases} p \neq p_0 & \text{zweiseitig} \\ p > p_0 & \text{einseitig nach oben} \\ p < p_0 & \text{einseitig nach unten} \end{cases}$$

4. Signifikatzniveau festlegen: α
5. Finde Kleinstes $c(\alpha)$ so das gilt:
   - Einseitig nach oben:

$$\mathbb{P}_{p_0}(X \geq c) = \sum_{k=c}^{n} \binom{n}{k} p_0^k (1 - p_0)^{n-k} \leq \alpha$$

   Verwerfungsbereich: $K = \{c, c+1, \ldots, n\}$
   - Einseitig nach unten:

$$\mathbb{P}_{p_0}(X \leq c) = \sum_{k=c}^{n} \binom{n}{k} p_0^k (1 - p_0)^{n-k} \leq \alpha$$

   Verwerfungsbereich: $K = \{0, 1, \ldots, c\}$
   - Beidseitig:

$$\mathbb{P}_{p_0}(X \leq c_1) = \sum_{k=c}^{n} \binom{n}{k} p_0^k (1 - p_0)^{n-k} \leq \frac{\alpha}{2}$$

$$\mathbb{P}_{p_0}(X \geq c_2) = \sum_{k=c}^{n} \binom{n}{k} p_0^k (1 - p_0)^{n-k} \leq \frac{\alpha}{2}$$

   $K = \{0, 1, \ldots, c\} \cup \{c, c+1, \ldots, n\}$
6. Testentschied: $x \in K$?

9.3 Test für normalverteilte Daten

Normalverteilte Daten $X \sim \mathcal{N}(\mu, \sigma^2)$
Als unbekannten Parameter nehmen wir die Schätzer:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

$$\hat{\sigma}^2 = S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \hat{\mu})^2$$

Nullhypothese:

$$H_0 : \mu = \mu_0 = \hat{\mu}$$

Alternativhypothesen:

$$H_A : \begin{cases} \mu \neq \mu_0 & \text{zweiseitig} \\ \mu > \mu_0 & \text{einseitig nach oben} \\ \mu < \mu_0 & \text{einseitig nach unten} \end{cases}$$

9.3.1 Z-Test (σ bekannt)

Mit der Annahme wissen wir:

$$\overline{X}_n \sim \mathcal{N}\left(\mu_0, \frac{\sigma^2}{n}\right)$$

Teststatistik:

$$Z = \frac{\overline{X}_n - \mu_0}{\sigma / \sqrt{n}} = \frac{\text{beobachtet} - \text{erwartet}}{\text{Standardfehler}} \sim \mathcal{N}(0, 1)$$

Verwerfungsbereich:

$\mu \neq \mu_0$	$K = \left(-\infty, -z_{1-\frac{\alpha}{2}}\right] \cup \left[z_{1-\frac{\alpha}{2}}, \infty\right)$
$\mu > \mu_0$	$K = [z_{1-\alpha}, \infty)$
$\mu < \mu_0$	$K = (-\infty, z_\alpha] = (-\infty, -z_{1-\alpha}]$

9.3.2 t-Test (σ unbekannt)

Annahme:

$$\overline{X}_n \sim \mathcal{N}\left(\mu_0, \frac{\sigma^2}{n}\right)$$

Teststatistik:

$$T = \frac{\overline{X}_n - \mu_0}{S_n / \sqrt{n}} = \frac{\text{beobachtet} - \text{erwartet}}{\text{Standardfehler}} \sim t_{n-1}$$

$n - 1$ Freiheitsgrade. Pro Beobachtung erhalten wir einen, pro Parameter verlieren wir einen.
Verwerfungsbereich:

$\mu \neq \mu_0$	$K = \left(-\infty, -t_{n-1,1-\frac{\alpha}{2}}\right] \cup \left[t_{n-1,1-\frac{\alpha}{2}}, \infty\right)$
$\mu > \mu_0$	$K = [t_{n-1,1-\alpha}, \infty)$
$\mu < \mu_0$	$K = (-\infty, t_{n-1,\alpha}] = (-\infty, -t_{n-1,1-\alpha}]$

9.3.3 Fehler 2. Art bei Z-Test

Formel ist für Durchschnitt. Bei Summe der die anderer Formel für die Annäherung nehmen.

$$Z \sim \mathcal{N}\left(\sqrt{N}\frac{\mu - \mu_0}{\sigma}, 1\right)$$

Daraus folgt dann mit Kr = kritisches Niveau (einseitig)

$$\beta = \mathbb{P}(Z \leq Kr) = \Phi\left(\sqrt{N}\frac{\mu - \mu_0}{\sigma} - Kr\right)$$

9.4 Eigenschaften von statistischen Test

9.4.1 Macht

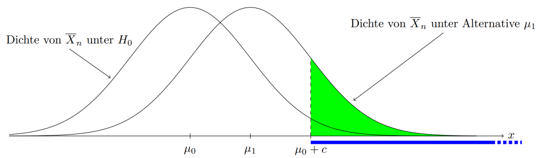
Macht:

$$M = \mathbb{P}(\text{Test verwirft richtigerweise } H_0 \text{ für } \theta \in H_A) = 1 - \beta(\theta)$$

Wahrscheinlichkeit eines Fehlers 2. Art:

$$\beta(\theta) = \mathbb{P}(\text{Test akzeptiert } H_0, \text{ obwohl } \theta \in H_A \text{ stimmt})$$

Grüne Fläche ist die Macht



Macht einer Binominalverteilung: c = kritischer Wert

$$\text{Macht} = \mathbb{P}_{real}(X \geq c) = \mathbb{P}_{real}(X \in K)$$

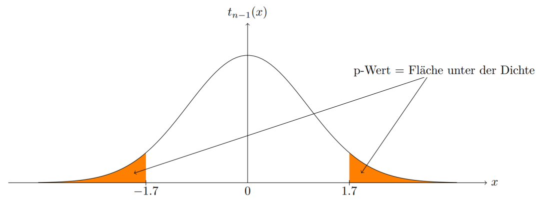
9.4.2 P-Wert

Definition:

Der **p-Wert** ist die Wahrscheinlichkeit, unter der Nullhypothese, einen mindestens so extremen Wert der Teststatistik zu erhalten. Die Alternativhypothese definiert was "extrem" ist.

Testentscheid mithilfe des P-Wert:

- $p - Wert < \alpha$: H_0 verwerfen
- $p - Wert \geq \alpha$: H_0 nicht verwerfen



Man kann sogar sagen wie "stark" wir H_0 verwerfen.

Konkrete Beispiele

P-Wert Binominaltest: Q = Beobachteter Wert
einseitig oben:

$$\text{p-Wert} = \mathbb{P}_{p_0}[X \geq Q]$$

einseitig unten:

$$\text{p-Wert} = \mathbb{P}_{p_0}[X \leq Q]$$

P-Wert Z-Test:
Teststatistik auswerten:

$$Z = \frac{\overline{X}_n - \mu_0}{\sigma / \sqrt{n}} = \frac{\text{beobachtet} - \text{erwartet}}{\text{Standardfehler}} \sim \mathcal{N}(0, 1)$$

In der Tabelle ablesen: (einseitig), zweiseitig anders Intervall

$$\text{p-Wert} = 1 - \Phi(Z)$$

P-Wert t-Test:
Teststatistik auswerten:

$$T = \frac{\overline{X}_n - \mu_0}{S_n / \sqrt{n}} = \frac{\text{beobachtet} - \text{erwartet}}{\text{Standardfehler}} \sim t_{n-1}$$

In der Tabelle rückwärts ablesen!

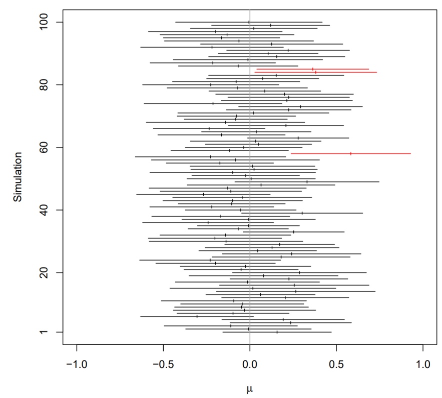
9.5 Vertrauensintervalle

Vertrauensintervall (Konfidenzintervall):
Besteht aus allen Parameter die mit der Beobachtung (Daten) zum α Signifikatzniveau verträglich sind.

$$I = \{\theta_0 : \text{Nullhypothese } H_0 : \theta = \theta_0 \text{ wird nicht verworfen}\}$$

Wenn ein Parameter θ im Vertrauensintervall liegt, wird der Test H_0 *nicht verwerfen*.
Andere Intepretation:

$$\mathbb{P}(\theta \in I) = 1 - \alpha$$



Das Vertaueensintervall enthält mehr Informationen als der p-Wert. Breite macht eine Aussage über die Genauigkeit der Schätzung.

$$\text{Testeintscheid} \prec \text{P-Wert} \prec \text{Vertrauensintervall}$$

Vertrauensintervalle einiger Tests

Z-**Test:** Alle haben gleiche Breite

$$I = \overline{X}_n \pm \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}$$

t-**Test:** Breite von den Daten abhängig

$$I = \overline{X}_n \pm \frac{S_n}{\sqrt{n}} t_{n-1,1-\frac{\alpha}{2}}$$

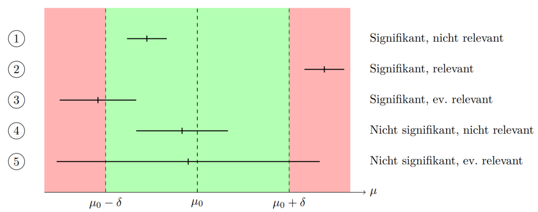
Achtung:
Das Vertrauensintervall ist **nicht** das gleiche wie der Annahmebereich eines Tests, da man beim Annahmebereich von einer konkreten Nullhypothese ausgeht. Das Vertrauensintervall hingegen basiert auf den Daten!!!

9.5.1 Statistische vs. Fachliche Relevanz

Allgemein gilt:

$$\text{Statistisch Relevant} \neq \text{Fachlich Relevant}$$

Wir müssen Fachwissen anwenden und definieren was Relevant ist.



Grüner Bereich ist falchlich nicht relevant.

9.6 Test für nicht normalverteilte Daten

Wenn die Daten nicht normalverteilt sind, müssen wir Tests mit weniger Annahmen verwenden.

9.6.1 Vorzeichen-Test

Daten beliebig *stetig* verteilt mit $X_i \sim \mathcal{F}(\mu)$ und Median μ.
Für symmetrische Verteilungen ist der Median gleich dem Erwartungswert.
Hypothesen:

- Nullhypothese: $H_0 : \mu = \mu_0$
- Alternativhypothese: $H_A : \mu \neq \mu_0$

Wir erwarten 50% der Daten über bzw. unter dem Median.
$\rightarrow$ Zählen der Anzahl positiver $X_i - \mu_0$.

$$V_i = \begin{cases} 1 & X_i - \mu_0 > 0 \\ 0 & \text{sonst} \end{cases}$$

Die Anzahl positiver V_i ist Binomialverteilt:

$$Q = \sum_{i=1}^{n} V_i \sim \text{Bin}(n, 0.5)$$

$\Rightarrow$ Binominaltest durchführen.

9.6.2 Wilcoxon-Test

Weniger Annahmen als der t-Test, aber besser als Vorzeichen-Test.

Daten sind *symmetrisch* und *stetig* mit $X_i \sim \mathcal{F}(\mu)$ und Erwartungswert (Median) μ verteilt.
Hypothesen:

- Nullhypothese: $H_0 : \mu = \mu_0$
- Alternativhypothese: $H_A : \mu \neq \mu_0$

Wir geben den Werten einen Rang: (beginnt bei 1)

$$\text{Rang}\left(|X_i - \mu_0|\right) = k$$

Teststatistik: V_i vom Vorzeichen-Test

$$W = \sum_{i=1}^{n} \text{Rang}\left(|X_i - \mu_0|\right) V_i$$

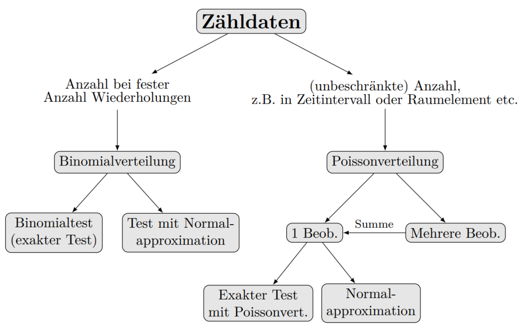
Wir schaune jetzt ob der Wert für W in einem gewissen Intervall (Tabelliert) liegt.
Kleines W: μ_0 zu gross
Mittiges W: μ_0 stimmt
Grosses W: μ_0 zu klein

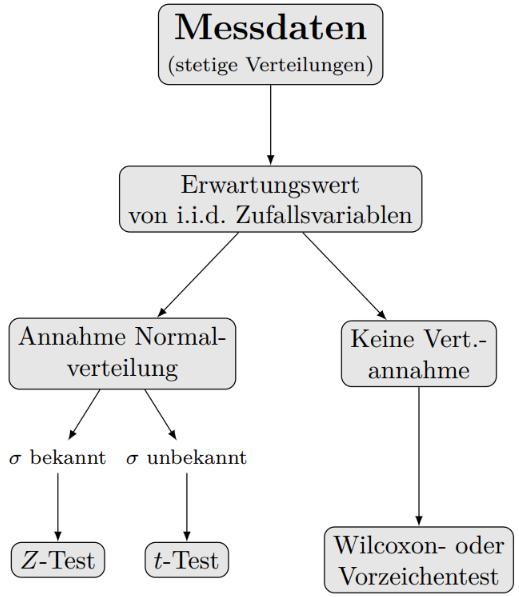
Falls der Rang für N Werte gleich ist

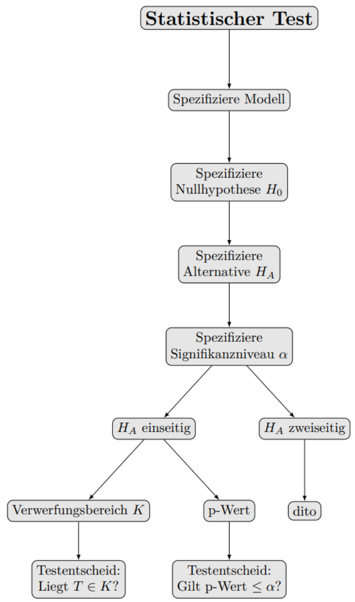
$$R = \frac{1}{N} \sum R_i$$

R_i sind die theoretischen Ränge falls sie nicht gleich wären.

9.7 Test Zusammengefasst







Vertrauensintervall
(meist zweiseitig)
Alle Nullhypothesen, die beim entsprechenden Test nicht verworfen werden können.

9.7.1 Unterschied zu ungepaart

Da wir nun mehr Information über die Testbedingungen haben (das unterschiedliche "Niveau" aufgrund der Versuchsbedingungen kürzt sich nun weg), sollte es leichter möglich sein, ein signifikantes Testergebnis zu erhalten. Insbesondere erwarten wir, dass der p-Wert kleiner wird.

10 Vergleich zweier Stichproben

10.1 Gepaarte/Ungepaarte Stichproben

Gepaarte Stichprobe:
Versuchsbedingungen an *derselben* Versuchseinheit:

$$\begin{aligned} &x_1, \ldots, x_n, \text{ unter Versuchsbedingung 1} \\ &y_1, \ldots, y_n, \text{ unter Versuchsbedingung 2} \end{aligned}$$

Jede Versuchseinheit einer Gruppe kann genau einer Versuchseinheit der anderen Gruppe zu ordnen. Daher müssen die Gruppen auch gleichgross sein.
Beispiele:

- Zwei verschiedene Reifen am gleichen Auto
- Prüfkörper werden von 2 Labors gemessen (nondestruktiv)

Ungepaarte Stichprobe:
Die Versuchseinheiten der Gruppen haben nichts miteinander zu tun:

$$\begin{aligned} &x_1, \ldots, x_n, \text{ unter Versuchsbedingung 1} \\ &y_1, \ldots, y_m, \text{ unter Versuchsbedingung 2} \end{aligned}$$

Wobei gelten kann $n \neq m$.
Beispiele:

- Zugfestigkeit eines Drahtes aus zwei Werken (destruktiv)
- Zufällige Placebo/Medikamenten Tests

10.2 Versuchsplanung

Wichtigste Punkte:

- Randomisierung: damit der einzige systematische Unterschied die Versuchsbedingung ist
- Kontrollgruppe: Vergleich zum Nullwert
- Doppelblind: Wenn Menschen involviert sind, sollten nicht mal die Tester wissen welche Versuchsbedingung die Propanden haben
- Beobachtungsstuden: Wenn keine Experimente durchführbar sind, müssen wir sehr gut aufpassen bei Ursache-Wirkung Beziehungen

10.3 Gepaarte Vergleiche

Bei gepaarten Stichproben betrachten wir immer die Differenz:

$$u_i = x_i - y_i, \quad i = 1, \ldots, n$$

Falls die Versuchsbedingungen identisch waren erwarten wir

$$\mathbb{E}[U_i] = 0$$

Wir nehmen als Nullhypothese $\mathbb{E}[U_i] = 0$ und führen einen t-Test, Vorziechen-Test oder Wilcoxon-Test durch.

10.4 Zwei-Stichproben Tests

Bei ungepaarten Stichproben brauchen wir andere Test.
Annahmen:

- X_i und Y_i sind unabhängig
- X_i i.i.d. $\sim \mathcal{N}(\mu_X, \sigma^2)$
- Y_i i.i.d. $\sim \mathcal{N}(\mu_Y, \sigma^2)$
- σ^2 ist gleich

Wir vergleichen die Erwartungswerte. Hypothesen:

- Nullhypothese: $H_0 : \mu_X = \mu_Y$
- Alternativhypothese: $H_A : \mu_X \neq \mu_Y$

Teststatistik:

$$Z = \frac{(\overline{X}_n - \overline{Y}_n) - (\mu_X - \mu_Y)}{\sigma \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim \mathcal{N}(0,1)$$

σ meist unbekannt:

$$\begin{aligned} S_{pool}^2 &= \frac{1}{n+m-2} \left(\sum_{i=1}^{n} (X_i - \bar{X}_n)^2 + \sum_{i=1}^{m} (Y_i - \bar{Y}_i)^2 \right) \\ &= \frac{1}{n+m-2} \left((n-1) S_X^2 + (m-1) S_Y^2 \right) \end{aligned}$$

Neue Teststatistik:

$$T = \frac{\bar{X}_n - \bar{Y}_m - (\mu_X - \mu_Y)}{S_{pool} \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n+m-2}$$

Für die Nullhypothese $\mu_X = \mu_Y$ testen wir:

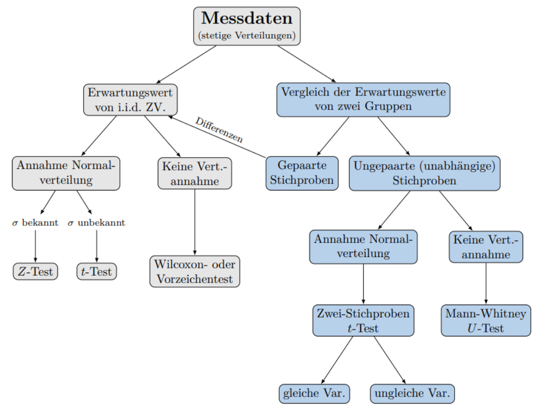
$$t = \frac{\bar{x}_n - \bar{y}_n}{s_{pool} \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

Mit Verwerfungsbereich:

H_0	$K =$
$\mu_X \neq \mu_Y$	$\left(-\infty, -t_{n+m-2,1-\frac{\alpha}{2}}\right] \cup \left[t_{n+m-2,1-\frac{\alpha}{2}}, \infty\right)$
$\mu > \mu_0$	$[t_{n+m-2,1-\alpha}, \infty)$
$\mu < \mu_0$	$(-\infty, t_{n+m-2,\alpha}] = (-\infty, -t_{n+m-2,1-\alpha}]$

Vertrauensintervall:

$$\begin{aligned} I &= \left\{ d : \left| \frac{\bar{X}_n - \bar{Y}_m - d}{S_{pool} \sqrt{\frac{1}{n} + \frac{1}{m}}} \right| < t_{n+m-2,1-\frac{\alpha}{2}} \right\} \\ &= \bar{X}_n - \bar{Y}_m \pm S_{pool} \sqrt{\frac{1}{n} + \frac{1}{m}} t_{n+m-2,1-\frac{\alpha}{2}} \end{aligned}$$



11 Lineare Regression

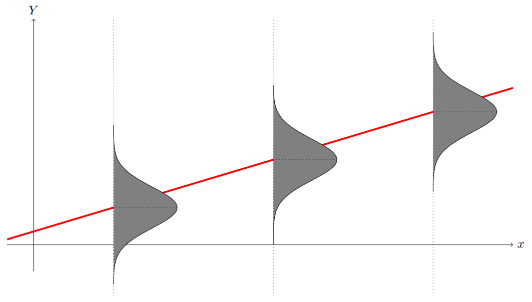
11.1 Einfache Lineare Regression

Datenpaare (x_i, y_i) gegeben.
Linearer Zusammenhang: (erklärende Variable x_i)

$$Y_i = \beta_0 + \beta_1 x_i + E_i, \quad i = 1, \dots, n$$

Fehler ist i.i.d. $E_i \sim \mathcal{N}(0, \sigma^2)$.
Zielvariable:

$$Y_i \sim \mathcal{N}(\beta_0 + \beta_1 x_i, \sigma^2)$$



11.1.1 Parameterschätzung

Kleinste Quadrate Schätzer:
Minimieren des Fehlers:

$$e = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

Minimierung:

$$\hat{\beta}_0, \hat{\beta}_1 = \text{argmin}_{\beta_0, \beta_1} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

Lösung:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = r \frac{s_y}{s_x}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

r = empirische Korrelation
s_i = empirische Standardabweichung
Residuen:

$$r_i = y_i - \hat{y}_i, \quad \hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

Geschätzte Varianz:

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^{n} r_i^2$$

Maximum-Likelihood Schätzer:
Gleiches Resultat wie kleinste Quadrate Methode.
Modell:

$$Y_i \sim \mathcal{N}\left(\beta_0 + \beta_1 x_i, \sigma^2\right)$$

Likelihoodfunktion:

$$L(\beta_0, \beta_1) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{y_i - \beta_0 - \beta_1 x_i}{\sigma}\right)^2\right)$$

Log-Likelihoodfunktion:

$$l(\beta_0, \beta_1) = c - \frac{1}{2} \sum_{i=1}^{n} \left(\frac{y_i - \beta_0 - \beta_1 x_i}{\sigma}\right)^2$$

11.1.2 Vertrauensintervalle

Wir kennen die Genauigkeit der Parameter:

$$\hat{\beta}_0 \sim \mathcal{N}\left(\beta_0, \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{SS_X}\right)\right)$$

$$\hat{\beta}_1 \sim \mathcal{N}\left(\beta_1, \frac{\sigma^2}{SS_X}\right)$$

mit

$$SS_X = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

Test für $\hat{\beta}_1$:
Teststatistik:

$$\frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}/\sqrt{SS_X}} \sim t_{n-2}$$

Hypothesen:

- Nullhypothese: $H_0 : \beta_1 = 0$
- Alternativhypothese: $H_A : \beta_1 \neq 0$

Wir verwerfen wenn

$$\left|\frac{\hat{\beta}_1}{\hat{\sigma}/\sqrt{SS_X}}\right| \geq t_{n-2, 1-\frac{\alpha}{2}}$$

Vertrauensintervall:

$$I = \hat{\beta}_1 \pm t_{n-2, 1-\frac{\alpha}{2}} \cdot \frac{\hat{\sigma}}{\sqrt{SS_X}}$$

Für $\hat{\beta}_0$ gehen die Berechnungen analog.

11.1.3 Residuen Analyse

Modellannahme:

1. $\mathbb{E}[E_i] = 0$
2. $E_1, \dots, E_n$ sind i.i.d.
3. $E_1, \dots, E_n \sim \mathcal{N}(0, \sigma^2)$

Je deutlicher diese verletzt sind, desto weniger vertrauenswürdig sind die Resultate.
Aussagen könne über verschieden Plots gemacht werden.

Tukey-Anscombe Plot (TA-Plot): r_i gegen $\hat{y}_i$ (oder x_i)

- Starke Abweichung von $\hat{y}_i$-Achse widerspricht Annahme.
- Stark unterschiedliche Streuung widerspricht Annahme.
- Es kann Ausreisser geben.

Normalplot der Residuen:

- Starke Abweichung von einer Geraden widerspricht Annahme ??.

Serial Correlation Plot: E_i gegen i plotten

Vor allem sinnvoll wenn Daten in chronologischer Reihenfolge

- Idealfall: keine Regionen mit ähnlichen Residuen (z.B. gleiches Vorzeichen)

11.2 Multiple Lineare Regression

Mehrere Variablen $x^{(1)}, \dots, x^{(m)}, m > 1$.
Modell:

$$Y_i = \beta_0 + \sum_{j=1}^{m} \beta_j x_i^{(j)} + E_i$$

Fehler ist i.i.d. $E_i \sim \mathcal{N}(0, \sigma^2)$.
Wir haben $p = m + 1$ verschiedene β-Parameter.
$x^{(j)}$ können auch Funktionen sein (x_i^2 oder $\log(x_i)$).
Gleichungssystem:

$$\underbrace{\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}}_{Y} = \underbrace{\begin{pmatrix} 1 & x_1^{(1)} & \cdots & x_1^{(m)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n^{(1)} & \cdots & x_n^{(m)} \end{pmatrix}}_{\text{Designmatrix } X} \underbrace{\begin{pmatrix} \beta_0 \\ \vdots \\ \beta_m \end{pmatrix}}_{\beta} + \underbrace{\begin{pmatrix} E_1 \\ \vdots \\ E_n \end{pmatrix}}_{E}$$

Lösung: falls X vollen Rang hat

$$\hat{\beta} = \left(X^T X\right)^{-1} X^T y$$

Standardfehler:

$$\hat{\sigma}^2 = \frac{1}{n-m} \sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \sum_{j=1}^{m} \hat{\beta}_j x_i^{(j)}\right)^2$$

11.2.1 Tests und Vertrauensintervall

Individuelle Tests:
Wir Testen jeden Parameter ob er weggelassen werden kann:

- Nullhypothese: $H_{0,j}$: $\beta_j = 0$
- Alternativhypothese: $H_{A,j}$: $\beta_j \neq 0$

Als Resultat erhalten wir ein t-Verteilung $\to t_{n-m}$

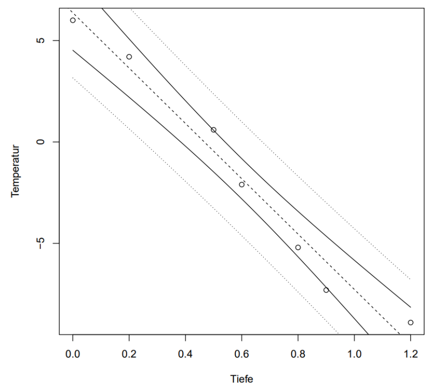
F-Test:
Testen ob alle Parameter weggelassen werden können:

- Nullhypothese: H_0: $\beta_j = 0 \; \forall j \in \{1, \dots, m\}$
- Alternativhypothese: H_A: $\exists j \in \{1, \dots, m\} : \beta_j \neq 0$

$\to$ F-Verteilung (aus Computerprogramm)

11.3 Prognoseintervall

Das Prognoseintervall ist ein Intervall, dass mit hoher Wahrscheinlichkeit neu gesammelte Daten einfängt. Das Prognoseintervall ist immer grösser als das Vertrauensintervall. Beide sind gekrümmt.



12 Anhang

12.1 Sonstige Fakten

- Die Macht eines Test ist grösser, desto weiter weg der wahre Wert von der Nullhypothese liegt.
- $X \sim$ Bernoulli(p), $Y \sim$ Bernoulli(q) und $\mathbb{E}[XY] = K$, dann ist $\mathbb{P}[X = 1, Y = 1] = K$.
- Stichprobe mit reelen Zahlen kann der empirische Median und Mittelwert berechnet werden.
- Falls Cor$(X, Y) \neq 0$ sind X und Y abhängig ($\sim \mathcal{N}$)
- Falls A und B unabhängig auch A und B^c unabhängig
- Aus den Randverteilungen kann man die gesamte Verteilung nur berechnen, falls X, Y unabhängig sind.

12.2 Multiples Testen

m statistische Tests mit Nullhypothesen $H_{0,j}$.
Auch wenn H_0 stimmt werden $\alpha \cdot 100\%$ der Tests H_0 verwerfen. Anders ausgedrückt wird es fast immer ein statistisch siginifikanter Testeintschied geben wenn wir genügen oft Testen !!!

Unabhängige Tests:

$$\begin{aligned} &\mathbb{P}(\text{mindestens ein } H_{0,j} \text{ wird verworfen}) \\ &= 1 - \mathbb{P}(\text{kein ein } H_{0,j} \text{ wird verworfen}) \\ &= 1 - \prod_{j=1}^{m} \mathbb{P}\left(H_{0,j} \text{ wird nicht verworfen}\right) \\ &= 1 - (1 - \alpha)^m \end{aligned}$$

Bereits für $\alpha = 0.05$ und $m = 50$ bei 92%.
Abhängige Tests:

$$\begin{aligned} &\mathbb{P}(\text{mindestens ein } H_{0,j} \text{ wird verworfen}) \\ &\leq \sum_{j=1}^{m} \mathbb{P}\left(H_{0,j} \text{ wird verworfen}\right) \\ &= \alpha \cdot m \end{aligned}$$

Bonferroni-Korrektur: strikteres Niveua

$$\alpha^* = \frac{\alpha}{m} \to \mathbb{P}(\text{mindestens ein } H_{0,j} \text{ wird verworfen}) \leq \alpha$$

Falls die Nullhypothese stimmt und wir mit dem Niveau α testen, müssen wir im Schnitt $\frac{1}{\alpha}$ mal wiederholen, bis wir die Nullhypothese das erste mal verwerfen.

Beispiel mit Tabelle

$X \setminus Y$	1	2	3	4	Σ
1	0.080	0.015	0.003	0.002	0.1
2	0.050	0.350	0.050	0.050	0.5
3	0.030	0.060	0.180	0.030	0.3
4	0.001	0.002	0.007	0.090	0.1
Σ	0.161	0.427	0.240	0.172	1

Randverteilung von X (orange) ist einfach die zeilenweise summierten Wahrscheinlichkeiten.
Randverteilung von Y (pink) ist einfach die spaltenweise summierten Wahrscheinlichkeiten.
X und Y sind nicht unabhängig, da Wahrscheinlichkeiten nicht produkt der jeweiligen Randdichten sind.

12.3 Berechnung der Macht

Zwei-Stichproben Z-Test

Zwei Datensätze vergleichen mit μ_A und μ_B.

$$H_0 : \mu_A = \mu_B, \quad H_A : \mu_A \neq \mu_B$$

Teststatistik unter H_0:

$$Z = \frac{(\overline{X}_n - \overline{Y}_n) - (\mu_A - \mu_B)}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim \mathcal{N}(0,1)$$

Daten von H_0 mit der Verteilung $H_A : \mu_A - \mu_B = \Delta\mu$ ist:

$$Z_A = \frac{(\overline{X}_n - \overline{Y}_n) - \Delta\mu}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim \mathcal{N}\left(-\frac{\Delta\mu}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}}, 1\right)$$

Daher ist $Z_A = Z - \frac{\Delta\mu}{\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}}$

Die Macht is dann:
Kr = kritischer Wert, $\frac{\Delta\mu}{\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}} = D$

$$\begin{aligned}\mathbb{P}[|Z_A| > Kr] &= 1 - \mathbb{P}[-Kr < Z_A < Kr] \\ &= 1 - \mathbb{P}[-Kr + D < Z\mu < Kr + D] \\ &= 1 - \Phi(-Kr + D) - \Phi(Kr + D)\end{aligned}$$

Achtung bei einsitigen Tests nicht den Betrag nehmen!

Z-Test:

Nullhypothese:

$$Z = \frac{\overline{X}_n - \mu_0}{\sigma\sqrt{\frac{1}{n}}} \sim \mathcal{N}(0,1)$$

Alternativhypothese:

$$Z_A = \frac{\overline{X}_n - \mu_A}{\sigma\sqrt{\frac{1}{n}}} \sim \mathcal{N}\left(\frac{\mu_0 - \mu_A}{\sigma\sqrt{\frac{1}{n}}}, 1\right)$$

Macht: Kr = Kritischer Wert
$D = \frac{\mu_0 - \mu_A}{\sigma\sqrt{\frac{1}{n}}}$

$$\begin{aligned}\mathbb{P}[|Z_A| > Kr] &= 1 - \mathbb{P}[-Kr < Z_A < Kr] \\ &= 1 - \mathbb{P}[-Kr + D < Z\mu < Kr + D] \\ &= 1 - \Phi(-Kr + D) - \Phi(Kr + D)\end{aligned}$$

Achtung bei einsitigen Tests nicht den Betrag nehmen!

t-Test

Datensatz mit (beidseitig)

$$H_0 : \mu_0 = 0, \quad H_A : \mu_A \neq 0 \sim t_{n-1}$$

Teststatistik unter H_0:

$$T = \frac{\bar{X}_n - \mu_0}{s_n/\sqrt{n}} \sim t_{n-1}$$

Wenn nun $H_A : \mu_A$ stimmen würde: nicht $\sim t_{n-1}$

$$T_A = \frac{\bar{X}_n - \mu_A}{s_n/\sqrt{n}}$$

Die Macht ist dann:
Kr = Kritischer Wert, $T_A = T + \frac{\mu_0 - \mu_A}{s_n/\sqrt{n}}$

$$\begin{aligned}\mathrm{M} &= \mathbb{P}[T_A > Kr] + \mathbb{P}[T_A < Kr] \\ &= \mathbb{P}\left[T > Kr - \frac{\mu_A}{s_n/\sqrt{n}}\right] + \mathbb{P}\left[T < -Kr - \frac{\mu_A}{s_n/\sqrt{n}}\right]\end{aligned}$$

Die Werte für die Wahrscheinlichkeiten rückwerts ablesen.

12.4 Kochrezepte

12.4.1 Z-Test

Kochrezept

1. Bestimme Nullhypothese:
$$H_0 : \mu = \mu_0$$
2. Bestimme Alternativhypothese:
   (a) $H_A : \mu \neq \mu_0$
   (b) $H_A : \mu > \mu_0$
   (c) $H_A : \mu < \mu_0$
3. Berechne realisierte Teststatistik:
$$z = \frac{\overline{x}_n - \mu_0}{\sigma/\sqrt{n}}$$
4. Bestimme $z_{1-\alpha}$ ($(1-\alpha)$-Quantil der Stdnormverteilung):
$$1 - \alpha = \mathbb{P}(Z \leq z_{1-\alpha}) = F(z_{1-\alpha}) = \Phi(z_{1-\alpha})$$
5. Bestimme den Verwerfungsbereich K:
   Siehe Tabelle Oben!
6. Bestimme ob $z \in K$
   ja $\Rightarrow$ Nullhypothese kann verworfen werden
   nein $\Rightarrow$ Nullhypotese kann nicht verworfen werden

12.4.2 t-Test

Kochrezept

1. Bestimme Nullhypothese:
$$H_0 : \mu = \mu_0$$
2. Bestimme Alternativhypothese:
   (a) $H_A : \mu \neq \mu_0$
   (b) $H_A : \mu > \mu_0$
   (c) $H_A : \mu < \mu_0$
3. Berechne realisierte Teststatistik:
$$t = \frac{\overline{x}_n - \mu_0}{s_n/\sqrt{n}}$$
4. Bestimme Teststatistik
   Unter H_0 ist $T \sim t_{n-1}$ Beispiel: $n = 10 \to t_9$
5. Bestimme den Verwerfungsbereich K:
   Mit Tabelle Oben und mit Tabelle für t_{n-1}
6. Bestimme ob $t \in K$
   ja $\Rightarrow$ Nullhypothese kann verworfen werden
   nein $\Rightarrow$ Nullhypotese kann nicht verworfen werden

Für bestimmung über P-Wert muss ab Schritt 4 folgendermassen vorgegangen werden:

5. Je nach H_A ist der p-Wert nun folgendermassen definiert:
   (a) $p = \mathbb{P}(T \geq |t|)$
   (b) $p = \mathbb{P}(T \geq t) = 1 - \mathbb{P}(T \leq t) = 1 - F(t)$
   (c) $p = \mathbb{P}(T \leq t) = F(t)$
6. Diese beziehungen können folgendermassen umgeformt werden:
   (a) $|t| = t_{n-1,1-0.5p}$
   (b) $t = t_{n-1,1-p}$
   (c) $t = t_{n-1,p}$
7. Nun kann anhand der Tabelle der p-Wert bestimmt werden
8. Testentscheid:
   $p < \alpha$ $\Rightarrow$ Nullhypothese kann verworfen werden
   $p \geq \alpha$ $\Rightarrow$ Nullhypotese kann nicht verworfen werden

12.4.3 Vorzeichen Test

Kochrezept

1. Setze: $\mu = \text{median}(X_i)$
2. Bestimme Nullhypothese:
$$H_0 : \mu = \mu_0$$
3. Bestimme Alternativhypothese:
   (a) $H_A : \mu \neq \mu_0$
   (b) $H_A : \mu > \mu_0$
   (c) $H_A : \mu < \mu_0$
4. Bestimme alle Vorzeichen $q_i = |x_i - \mu_0|$
5. Bestimme realisierte Teststatistik (zähle pos.Vorzeichen)
$$q = \sum_{i=1}^{n} 1 \text{ if } q_i > 0$$
6. Je nach Alternativhypothese kann nun Verwerfungsbereich bestimmt werden:
   (a) $\mathbb{P}(Q \geq c_1) \leq \alpha/2$ und $\mathbb{P}(Q \leq c_2) \leq \alpha/2$
   (b) $\mathbb{P}(Q \geq c) \leq \alpha$
   (c) $\mathbb{P}(Q \leq c) \leq \alpha$
7. Bestimme ob $q \in K$
   ja $\Rightarrow$ Nullhypothese kann verworfen werden
   nein $\Rightarrow$ Nullhypotese kann nicht verworfen werden

Für bestimmung über P-Wert muss ab Schritt 5 folgendermassen vorgegangen werden:

6. Je nach H_A ist der p-Wert nun folgendermassen definiert:
   (a) $p = \mathbb{P}(Q \geq q) + \mathbb{P}(Q \leq (n-q))$
   (b) $p = \mathbb{P}(Q \geq q)$
   (c) $p = \mathbb{P}(Q \leq q)$
7. Ist nun p-Wert $< \alpha$?
   ja $\Rightarrow$ Nullhypothese kann verworfen werden
   nein $\Rightarrow$ Nullhypotese kann nicht verworfen werden

12.5 Unsicherheit von Schätzern

Wir haben eine Schätzer in linearer Abhängigkeit von einer Zufallsvariable gegeben:

$$\hat{\theta} = a \cdot X + b$$

Dann ist der Erwartungswert und die Varianz des Schätzers folgendermassen gegeben:

$$\mathbb{E}[\hat{\theta}] = a \cdot \mathbb{E}[X] + b \qquad Var(\hat{\theta}) = a^2 Var(X)$$

Der Standardfehler ist nun gegeben als die Standardabweichung:

$$\text{Standardfehler} = \sigma = \sqrt{Var(\hat{\theta})}$$

Spezialfall Normalverteilung

Wir wissen nun, dass X normalverteilt ist:

$$X \sim \mathcal{N}(\mu_X, \sigma_X^2)$$

Daher:

$$\mathbb{E}[aX + b] = a\mu_X + b \qquad Var(aX + b) = a^2\sigma_X^2$$

Also ist unser Parameter auch normalverteilt:

$$\hat{\theta} \sim \mathcal{N}\Big(\underbrace{a\mu_X + b}_{\mu_{\hat{\theta}}}, \underbrace{a^2\sigma_X^2}_{\sigma_{\hat{\theta}}^2}\Big)$$

Das $\beta \cdot 100\%$ Vertauensintervall ergibt sich nun als:

$$\hat{\theta} \pm \Phi^{-1}\left(1 - \frac{1-\beta}{2}\right) \cdot \sigma_{\hat{\theta}}$$

Was nichts anderes ist als:

$$\hat{\theta} \pm \Phi^{-1}\left(1 - \frac{1-\beta}{2}\right) \cdot a \cdot \sigma_X$$

12.6 Integrale

$$\int u \cdot v' \, dx = u \cdot v - \int u' \cdot v \, dx$$

$$\int \frac{1}{x} dx = \ln|x| + C$$

$$\int (ax+b)^s \, dx = \frac{(ax+b)^{s+1}}{a(s+1)} + C,\ s \neq -1$$

$$\int \frac{1}{ax+b} dx = \frac{\ln|ax+b|}{a} + C$$

$$\int \left(ax^p+b\right)^s x^{p-1} dx = \frac{\left(ax^p+b\right)^{s+1}}{ap(s+1)} + C,\ s \neq -1,\ a, p \neq 0$$

$$\int \left(ax^p+b\right)^{-1} x^{p-1} dx = \frac{\ln|ax^p+b|}{ap} + C,\ a \neq 0,\ p \neq 0$$

$$\int \frac{ax+b}{cx+d} dx = \frac{ax}{c} - \frac{ad-bc}{c^2} \ln|cx+d| + C$$

$$\int \frac{ax+b}{x^2+cx+d} dx = \frac{a}{2} \ln\left|x^2+cx+d\right| + \frac{b-\frac{ac}{2}}{\sqrt{d-\frac{c^2}{4}}} \arctan\left(\frac{x+\frac{c}{2}}{\sqrt{d-\frac{c^2}{4}}}\right) + C,\ c^2-4d<0$$

$$\int \frac{1}{x^2+a^2} dx = \frac{\arctan\frac{x}{a}}{a} + C$$

$$\int \frac{1}{x^2-a^2} dx = \frac{\ln\left|\frac{x-a}{x+a}\right|}{2a} + C$$

$$\int \sqrt{a^2+x^2} dx = \frac{x}{2}\sqrt{a^2+x^2} + \frac{a^2}{2} \ln\left(x+\sqrt{a^2+x^2}\right) + C$$

$$\int \sqrt{a^2-x^2} dx = \frac{x}{2}\sqrt{a^2-x^2} + \frac{a^2}{2} \arcsin\frac{x}{|a|} + C$$

$$\int \sqrt{x^2-a^2} dx = \frac{x}{2}\sqrt{x^2-a^2} - \frac{a^2}{2} \ln\left|x+\sqrt{x^2-a^2}\right| + C$$

$$\int \frac{1}{\sqrt{x^2+a^2}} dx = \ln\left(x+\sqrt{a^2+x^2}\right) + C$$

$$\int \frac{1}{\sqrt{x^2-a^2}} dx = \ln\left|x+\sqrt{x^2-a^2}\right| + C$$

$$\int \frac{1}{\sqrt{a^2-x^2}} dx = \arcsin\frac{x}{|a|} + C$$

$$\int a^{kx} dx = \frac{a^{kx}}{k \cdot \ln(a)} + C$$

$$\int e^{ax} p(x) \, dx = e^{ax} \sum_{i=0}^{n} (-1)^i a^{-i-1} p^{(i)} + C,\ a \neq 0$$

$p(x)$: Polynom n-ten Grades

12.7 Integrale

$$\int \ln|x| \, dx = x(\ln|x|-1) + C$$

$$\int \log_a |x| \, dx = x\left(\log_a |x| - \log_a e\right) + C$$

$$\int x^k \ln x dx = \frac{x^{k+1}}{k+1}\left(\ln x - \frac{1}{k+1}\right) + C,\ k \neq -1$$

$$\int \frac{\ln x}{x} dx = \frac{1}{2}(\ln x)^2 + C$$

$$\int \tan x dx = -\ln|\cos x| + C$$

$$\int \sin^2 x dx = \frac{1}{2}(x - \sin x \cos x) + C$$

$$\int \cos^2 x dx = \frac{1}{2}(x + \sin x \cos x) + C$$

$$\int \tan^2 x dx = \tan x - x + C$$

$$\int \frac{1}{\sin x} dx = \ln\left|\tan\frac{x}{2}\right| + C$$

$$\int \frac{1}{\cos x} dx = \ln\left|\tan\left(\frac{x}{2}+\frac{\pi}{4}\right)\right| + C$$

$$\int \frac{1}{\tan x} dx = \ln|\sin x| + C$$

$$\int \arcsin x dx = x \arcsin x + \sqrt{1-x^2} + C$$

$$\int \arccos x dx = x \arccos x - \sqrt{1-x^2} + C$$

$$\int \arctan x dx = x \arctan x - \frac{1}{2}\ln\left(1+x^2\right) + C$$

$$\int \tanh x dx = \ln(\cosh x) + C$$

$$\int \operatorname{arsinh} x dx = x \operatorname{arsinh} x - \sqrt{x^2+1} + C$$

$$\int \operatorname{arcosh} x dx = x \operatorname{arcosh} x - \sqrt{x^2-1} + C$$

$$\int \operatorname{artanh} x dx = x \operatorname{artanh} x + \frac{1}{2}\ln\left(1-x^2\right) + C$$

$$\int_0^{2\pi} \sin(mx)\sin(nx) \, dx = \int_0^{2\pi} \cos(mx)\cos(nx) \, dx = \begin{cases} 0 & m \neq n \\ \pi & m = n \neq 0 \end{cases};\ m, n \in \mathbb{Z}$$

$$\int_0^{2\pi} \sin(mx)\cos(nx) \, dx = 0,\ m, n \in \mathbb{Z}$$

$$\int_0^{\infty} \frac{\sin(ax)}{x} dx = \frac{\pi}{2},\ a > 0$$

12.8 Integrale

$$\int_0^{\infty} \sin\left(x^2\right) dx = \int_0^{\infty} \cos\left(x^2\right) dx = \frac{1}{2}\sqrt{\frac{\pi}{2}}$$

$$\int_0^{\infty} e^{-ax} x^n dx = \frac{n!}{a^{n+1}},\ a > 0$$

$$\int_0^{\infty} e^{-ax^2} dx = \frac{1}{2}\sqrt{\frac{\pi}{a}},\ a > 0$$

12.9 Ableitungen

$$\frac{d}{dx}(\ln|x|) = \frac{1}{x}$$

$$\frac{d}{dx}(\log_a |x|) = \frac{\log_a e}{x} = \frac{1}{x \ln a}$$

$$\frac{d}{dx}\left(a^x\right) = a^x \cdot \ln a$$

$$\frac{d}{dx}(\sin x) = \cos x$$

$$\frac{d}{dx}(\cos x) = -\sin x$$

$$\frac{d}{dx}(\tan x) = \frac{1}{\cos^2 x} = 1 + \tan^2 x$$

$$\frac{d}{dx}(\arcsin x) = \frac{1}{\sqrt{1-x^2}}$$

$$\frac{d}{dx}(\arccos x) = -\frac{1}{\sqrt{1-x^2}}$$

$$\frac{d}{dx}(\arctan x) = \frac{1}{1+x^2}$$

$$\frac{d}{dx}(\sinh x) = \cosh x$$

$$\frac{d}{dx}(\cosh x) = \sinh x$$

$$\frac{d}{dx}(\tanh x) = \frac{1}{\cosh^2 x} = 1 - \tanh^2 x$$

$$\frac{d}{dx}(\operatorname{arsinh} x) = \frac{1}{\sqrt{x^2+1}}$$

$$\frac{d}{dx}(\operatorname{arcosh} x) = \frac{1}{\sqrt{x^2-1}}$$

$$\frac{d}{dx}(\operatorname{artanh} x) = \frac{1}{1-x^2}$$