Beschreibende Statistik

Modalwert(e) x_{mod}
- am häufigsten auftretende Ausprägung

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

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

Spannweite
$$\max x_i - \min x_i$$

Stichprobenvarianz s^2
Auch für ZV möglich
$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{n}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right)$$

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

p-Quantil x_p
$$x_p = \begin{cases} x_{floor(np)+1}, & n\cdot p \notin \mathbb{N} \\ \frac{1}{2}\left(x_{np} + x_{np+1}\right), & n\cdot p \in \mathbb{N} \end{cases}$$

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

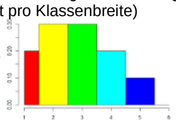
Empirischer Korrelationskoeffizient r
$$r = \frac{S_{xy}}{S_x \cdot S_y}$$

Regressionsgerade
$y = mx + t$ mit

$m = r \cdot \frac{s_y}{s_x}$ und $t = \bar{y} - m \cdot \bar{x}$

Histogramm

Flächentreue Darstellung der Häufigkeitsverteilung
Höhe = Dichte (Häufigkeit pro Klassenbreite)
Fläche =abs. Häufigkeit



Empirische Verteilungsfunktion

Klasse	Intervall	absolute Häufigkeit	relative Häufigkeit	relative Summen-häufigkeit
i		h_i	f_i	$F_i = \sum_{k<i} f_k$
1	[1.0, 1.5[	1	0.1	0
2	[1.5, 2.5[	3	0.3	0.1
3	[2.5, 3.5[	3	0.3	0.4
4	[3.5, 4.5[	2	0.2	0.7
5	[4.5, 5.5[	1	0.1	0.9
6	[5.5, 6.0[	0	0	1
Summe		$n = 10$	1	

Elementare Wahrscheinlichkeitsrechnung

De Morgan
$$\overline{\bigcup_{i=1}^{n} E_i} = \bigcap_{i=1}^{n} \bar{E}_i \qquad \overline{\bigcap_{i=1}^{n} E_i} = \bigcup_{i=1}^{n} \bar{E}_i$$

Axiome
1. $0 \le P(E) \le 1$
2. $P(\Omega) = 1$
3. $P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=n}^{\infty} P(E_i)$, falls $E_i \cap E_j = \emptyset\ für\ i \neq j$

$P(\bar{E}) = 1 - P(E)$
$P(E \cup F) = P(E) + P(F) - P(E \cap F)$

Laplace Exeriment
Zufallsexperiment mit n gleichwahrscheinlichen Elementarereignissen
$$P(E) = \frac{Anz\ der\ für\ E\ günstigen\ Ergebnisse}{Anz\ der\ möglichen\ Ergebnisse} = \frac{|E|}{n}$$

Bedingte Wahrscheinlichkeit
Die Wahrscheinlichkeit für das Eintreten von E, wenn F eingetreten ist.
$$P(E|F) = P_F(E) = \frac{|E|}{|F|} = \frac{P(E \cap F)}{P(F)}$$
Für $E, F \neq \emptyset$:
1. $P(E \cap F) = P(E|F) \cdot P(F)$
2. $P(E \cap F) = P(F|E) \cdot P(E)$

Satz der totalen Wahrscheinlichkeit
Alle E müssen disjunkt sein. Allgemein die Wahrscheinlichkeit, dass F Eintritt. Man nimmt nicht an, dass E_i eingetreten ist!
$$\Omega = \bigcup_{i=1}^{n} E_i\ mit\ E_i \cap E_j = \emptyset, dann\ gilt:$$
$$P(F) = \sum_{i=1}^{n} P(F \cap E_i) = \sum_{i=1}^{n} P(F|E_i) \cdot P(E_i)$$

Vierfeldertafel
$Spezialfall: \Omega = E \cup \bar{E}$
$P(F) = P(F \cap E) + P(F \cap \bar{E})$
$P(\bar{F}) = P(\bar{F} \cap E) + P(\bar{F} \cap \bar{E})$
$P(\bar{E}|F) = 1 - P(E|F)$
$P(E \cap \bar{F}) = P(E) - P(E \cap F) = P(E) - P(E \cap F)$
$= P(E|F) \cdot P(F) = P(F|E) \cdot P(E)$

Formel von Bayes
Hilfreich, wenn man $P(F|E_i)$ kennt, nicht aber $P(E_k|F)$
$$P(E_k|F) = \frac{P(F|E_k) \cdot P(E_k)}{\sum_{i=1}^{n} P(F|E_i) \cdot P(E_i)}$$

Unabhängigkeit
Bedeutet nicht unbedingt kausale Unabhängigkeit!
Gilt, wenn
$$P(E|F) = P(E) \quad bzw. \quad P(E \cap F) = P(E) \cdot P(F)$$
Falls E, F unabhängig, gilt auch $\bar{E}, F$ bzw. $\bar{E}, \bar{F}$ unabh

Kombinatorik

Allgemeines Zählprinzip
Anzahl der Möglichkeiten, für ein k-stufiges Zufallsexperiment mit n_i Varianten im i-ten Schritt:
$$n_1 \cdot n_2 \dots n_k$$

Anzahl der Permutationen einer n-elementigen Menge
n-maliges Ziehen **ohne Zurücklegen mit Beachtung der Reihenfolge**

- für n unterscheidbare Elemente:
  $n! = n \cdot (n-1) \dots \cdot 2 \cdot 1$
- für k Klassen mit je n_i nicht unterscheidbaren Elementen:

$$\frac{n!}{n_1! \cdot n_2! \cdot \dots \cdot n_k!}$$

k-maliges ziehen aus einer n-elementigen Menge

	mit Beachtung der Reihenfolge	ohne Beachtung der Reihenfolge
ohne Zurücklegen $k \le n$	$\frac{n!}{(n-k)!}$	**Lotto 6 aus 49** $\binom{n}{k} = \frac{n!}{(n-k)!k!}$
mit Zurücklegen $k > n$ möglich	n^k	$\binom{n+k-1}{k}$

$TR: \binom{n}{k} = n\ nCr\ k$

Ableitungsregeln

Produktregel
$y = u \cdot v$
$y' = u' \cdot v + v' \cdot u$

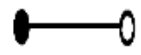
Quotientenregel
$y = \frac{u}{v}$
$y' = \frac{u' \cdot v - v' \cdot u}{v^2}$

Kettenregel
$\left(\boldsymbol{u(v(x))}\right)' = \boldsymbol{u'(v(x)) \cdot v'(x)}$

$\mathbf{\ln(x)' = \frac{1}{x}}$

Lotto 6 aus 49

Zufallsvariablen

Verteilungsfunktion F(x)
$x \in \mathbb{R}, F: \mathbb{R} \to [0, 1]$
$$F(x) = P(X \le x)$$
Eigenschaften:
- $0 \le F(x) \le 1$
- $\lim_{x\to-\infty} F(x) = 0, \lim_{x\to\infty} F(x) = 1$
- rechtsseitig stetig, also:
  $\lim_{x\to b+} F(x) = F(b)$ 
- monoton wachsend
- $P(X > x) = 1 - F(x)$
- $P(a < X \le b) = F(b) - F(a)$

Diskrete Zufallsvariablen

Wahrscheinlichkeitsverteilung:
$$p(x) = \begin{cases} P(X = x_i), falls\ x = x_i \in X(\Omega) \\ 0, \quad sonst \end{cases}$$

Eigenschaften Wahrscheinlichkeitsverteilung:
- $0 \le p(x) \le 1$
- $\sum_{i=1}^{n} p(x_i) = 1$

Es gilt:
- $F(x) = P(X \le x) = \sum_{x_i \le x} p(x_i)$

$F(x)$ist eine rechtsseitig **stetige Treppenfunktion** mit **Sprüngen** bei den Realisationen x_i
$P(X = 1) = P(X \le 1) - P(X < 1)$
$= Sprunghöhe, wenn\ 1\ bei\ Sprung. sonst\ 0$

Stetige Zufallsvariablen

Wahrscheinlichkeitsdichte f(x):
$$P(a < X < b) = \int_a^b f(x)\,dx$$

Eigenschaften Wahrscheinlichkeitsdichte:
- $f(x) \ge 0$
- $\int_{-\infty}^{\infty} f(x)\,dx = 1$

Es gilt:
- $F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)\,dt$
- $F'(x) = f(x)$
- $F(x)$ ist stetig und
- $P(a < X \le b) = P(a \le X \le b) =$
  $P(a \le X < b) = P(a < X < b)$
  $= F(b) - F(a)$

Zufallsvariablen

Erwartungswert $E[X] = \mu$

Für diskrete ZV:
$$E[X] = \sum_{i=1}^{n} x_i \cdot p(x_i)$$
Für stetige ZV:
$$E[X] = \int_{-\infty}^{\infty} x \cdot f(x)\,dx$$
Sei g(X) eine Funktion der ZV X, dann gilt:
Für diskrete ZV:
$$E[g(X)] = \sum_{i=1}^{n} g(x_i) \cdot p(x_i)$$
Für stetige ZV:
$$E[g(X)] = \int_{-\infty}^{\infty} g(x) \cdot f(x)\,dx$$

Eigenschaften des Erwartungswerts:
- $E[b] = b$
- $E[aX + b] = aE[X] + b$

$$E[X_1 + \dots + X_n] = \sum_{i=1}^{n} E[X_i]$$

Varianz σ^2
$$\sigma^2 = Var[X] = E[(X - \mu)^2]$$
Standardabweichung:
$$\sigma = \sqrt{Var[X]}$$
Verschiebungssatz:
$$Var[X] = E[X^2] - (E[X])^2$$
Eigenschaften der Varianz:
- $Var[b] = 0$
- $Var[aX + b] = a^2 Var[X]$
- $Var[X_1 + \dots + X_n] = \sum_{i=1}^{n}\sum_{j=1}^{n} Cov[X_i, X_j]$
- $Var[X_1 + X_2] = Var[X_1] + Var[X_2] + 2Cov[X_1, X_2]$
- Falls X_i, X_j *paarweise unabhängig*:

$$Var[X_1 + \dots + X_n] = \sum_{i=1}^{n} Var[X_i]$$

Kovarianz $Cov[X, Y]$
$$Cov[X, Y] = E[(X - E[X])\,(Y - E[Y])]$$
Wenn X, Y stochastisch unabhängig ⇒ $Cov[X, Y] = 0$
Verschiebungssatz:
$$Cov[X, Y] = E[XY] - E[X] \cdot E[Y]$$
Eigenschaften der Kovarianz:
- $Cov[X, Y] = Cov[Y, X]$
- $Cov[X, X] = Var[X]$
- $Cov[aX, Y] = aCov[X, Y]$

Zufallsvariablen

Quantile
Das p-Quantil ist der kleinste Wert $x_p \in \mathbb{R}$ für den gilt:
$$F(x_p) \ge p$$

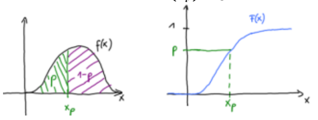


Abbildung: p-Quantil einer stetigen ZV mit streng monoton wachsendem $F(x)$:
$$x_p = F^{-1}(p)$$
Median: $F(x) = 0.5$

Chebyshev Ungleichung
$X: ZV$
$\mu = E[X]$
$\sigma^2: Var[X]$
Es gilt für jedes beliebige $k \in \mathbb{R}^+ \setminus \{0\}$:
$$P(|X - \mu| \ge k \cdot \sigma) \le \frac{1}{k^2}$$

Schwaches Gesetz der großen Zahlen
Mit Chebyshev und
$\mu = E[X_i]\ \ (i = 1, \dots, n) \ \Rightarrow\ \ \mu = E\left[\frac{X_1 + \dots + X_n}{n}\right] = E[\bar{X}]$
folgt:

$X_1, X_2, \dots$ ist eine Folge von unabhängigen, identisch verteilten ZV mit EW $E[X_i] = \mu$ und Varianz $Var[X_i] = \sigma^2$.
Dann gilt für ein beliebig kleines $\epsilon > 0$:
$$P(|\bar{X} - \mu| > \epsilon) \to 0 \quad für\ n \to \infty$$
d.h. der MW $\bar{X}$ konvergiert stochastisch gegen den EW μ.

Diskrete Verteilungen

Bernoulliverteilung $X \sim B_{1,p}$
Indikatorvariable mit den Werten 1 bei Erfolg und 0 bei Misserfolg.
$$P(X = 1) = p, \quad P(X = 0) = 1 - p$$
$$E[X] = p$$

Binomialverteilung $X \sim B_{n,p}$
Anzahl der Erfolge bei n-maligem Ziehen mit zurücklegen.
p = Wahrscheinlichkeit f. Erfolg bei 1mal ziehen.
k = Anz. Erfolge nötig für Gesamterfolg
$P(X = k) = \binom{n}{k} \cdot p^k \cdot (1 - p)^{n-k}, \ \ k \in \{0, 1, \dots, n\}$
$E[x] = np$
$Var[X] = np(1 - p)$
$dbinom(k, n, p) = P(X = k)$
$pbinom(k, n, p) = F(k)$
q-Quantil: $qbinom(q, n, p)$
k binomialverteilte Zufallszahlen: $rbinom(k, n, p)$

Hypergeometrische Verteilung $X \sim H_{M,N,n}$
Anz. d. Erfolge bei n-maligem Ziehen ohne Zurücklegen aus Menge mit M Elementen, die Erfolg bedeuten und N Elementen, die Misserfolg bedeuten.
$$P(X = k) = \frac{\binom{M}{k} \cdot \binom{N}{n-k}}{\binom{M+N}{n}}, k \in \{0, 1, \dots, \min\{n, M\}\}$$
$E[X] = n \cdot \frac{M}{M+N}$
$Var[X] = n \cdot \frac{M}{M+N}\left(1 - \frac{M}{M+N}\right)\frac{M+N-n}{M+N-1}$
$dhyper(k, M, N, n) = P(X = k)$
$phyper(k, M, N, n) = F(k)$

Poisson-Verteilung $X \sim P_\lambda$
Verteilung der seltenen Ereignisse. Häufigkeit punktförmiger Ereignisse in einem Kontinuum. Die durchschnittlich zu erwartende Anzahl der Erfolge λ pro Maßeinheit (i.a. Zeiteinheit) sei bekannt.
$$P(X = k) = \frac{\lambda^k}{k!}e^{-\lambda}, k \in \{0, 1, \dots\} mit \sum_{k=0}^{\infty} P(X = k) = 1$$
$E[X] = \lambda = n * p \ \Leftrightarrow\ 10 = p * 3600 \ \Leftrightarrow\ p = 1/360$
$Var[X] = \lambda$
$dpois(k, \lambda) = P(X = k)$
$ppois(k, \lambda) = F(k)$

Gleichverteilung $X \sim U_{\{x_1, \dots, x_n\}}$
Alle Werte $\{x_1, \dots, x_n\}$ einer ZV x sind gleichwahrsch.
$P(X = x_k) = \frac{1}{n}$
$E[X] = \frac{1}{n}\sum_{k=1}^{n} x_k = \bar{x}$
$Var[x] = \frac{1}{n}\sum_{k=1}^{n} x_k^2 - \bar{x}^2$
$sample(1{:}N, n){:}\ n\ Zufallszahlen\ zw\ 1\ und\ N$

Stetige Verteilungen

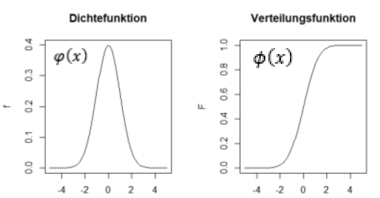
Stetige Gleichverteilung $X \sim U_{[a,b]}$
Zufallszahlen aus einem Intervall [a,b] (Bus-H. bsp)
$f(x) = \frac{1}{b-a}$ für $x \in [a, b]$
$E[X] = \frac{a+b}{2}$
$Var[X] = \frac{(b-a)^2}{12}$
$dunif(x, a, b) = f(x)$
$punif(x, a, b) = F(x)$
$runif(n){:}\ n\ Zufallszahlen\ zw\ 0\ und\ 1$

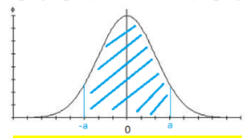
Normalverteilung $X \sim N_{\mu,\sigma^2}$
$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)}$
$E[X] = \mu$
$Var[X] = \sigma^2$
$dnorm(x, \mu, \sigma) = f(x)$
$pnorm(x, \mu, \sigma) = F(x)$
$qnorm(q, \mu, \sigma)$: q-Quantil = $F^{-1}(x)$

Eigenschaften:
- Max von f(x) bei $x = \mu$
- Wendestellen von f(x) bei $x = \mu \pm \sigma$
- $X \sim N_{\mu,\sigma^2} \ \Rightarrow aX + b \sim N_{a\mu+b, a^2\sigma^2}\ und\ \frac{X-\mu}{\sigma} \sim N_{0,1}$
- $X_1 \sim N_{\mu_1,\sigma_1^2}\ und\ X_2 \sim N_{\mu_2,\sigma_2^2} \Rightarrow X_1 + X_2 \sim N_{\mu_1+\mu_2, \sigma_1^2+\sigma_2^2}$

Standardnormalverteilung $X \sim N_{0,1}$

Dichte: $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{\left(-\frac{1}{2}x^2\right)}$
Verteilung: $\phi(x) = \int_{-\infty}^{x} \varphi(t)\,dt$
Quantile Wegen Achsensymmetrie von $\varphi(x)$ gilt:
$$\phi(-x) = 1 - \phi(x) \Rightarrow -x_p = x_{1-p}$$





$P(-a \le X \le a) = 2P(X \le a) - 1$

Stetige Verteilungen

Exponentialverteilung $X \sim Exp_\lambda$ (gedächtnislos)
Modellierung von Lebensdauern, Wartezeiten.
Sei $Y_t \sim P_{\lambda t}$ im Intervall [0,t] von t Zeiteinheiten, dann beschreibt die Exponentialverteilung die Wartezeit X bis Eintreten eines Ereignisses $f(x) = \lambda e^{-\lambda x} \ (x \geq 0)$
λ: Durchschnittliches Eintreten eines Ereignisses pro Zeiteinheit
x: Zeitabstand (höchstens) zw. zwei Ereignissen

$F(x) = 1 - e^{-\lambda x}$
$E[X] = \frac{1}{\lambda}$
$Var[X] = \frac{1}{\lambda^2}$
$dexp(x, \lambda) = f(x)$
$pexp(x, \lambda) = F(x)$

Eigenschaft:
Gedächtnislos:
$P(X > s + t | X > t) = P(X > s)$

Chiquadrat-Verteilung $X \sim \chi_n^2$
$Z_1, \dots, Z_n$ seien unabhängige, **standardnormalverteilte** ZV => $X = Z_1^2 + \cdots + Z_n^2$ hat Chiquadratverteilung mit n Freiheitsgraden.
Summen unabhängiger, standardnormalverteiler ZV

$E[X] = n$
$Var[X] = 2n$
$dchisq(x, n) = f(x)$
$pchisq(x, n) = F(x)$

Eigenschaft:
$X_1 \sim \chi_{n_1}^2 \ und \ X_2 \sim \chi_{n_2}^2 \Rightarrow X_1 + X_2 \sim \chi_{n_1+n_2}^2$

t-Verteilung $Y \sim t_n$
$Z \sim N_{0,1}$ und $X \sim x_n^2 \Rightarrow Y = \frac{Z}{\frac{X}{\sqrt{n}}}$ ist t-verteilt mit n Freiheitsgraden
Schätz- und Testverfahren bei unbekannter Varianz.

$E[Y] = 0 \ für \ n > 1$
$Var[Y] = \frac{n}{n-2} \ für \ n > 2$
$dt(y, n) = f(y)$
$pt(y, n) = F(y)$

Eigenschaften:
Für $n \to \infty$: $t_n \to N_{0,1}$
Achsensymmetrie der Dichtefunktion $\Rightarrow n = x_{1-p}$

ZGWS

Wahrscheinlichkeitsaussagen über X_i, wenn Erwartungswert μ und Varianz σ^2 bekannt sind, nicht aber die Verteilung für n unabhängige identisch verteilte (i.i.d.) ZV X_i.

Für hinreichend große n gilt dann näherungsweise:

$$\sum_{i=1}^{n} X_i \sim N_{n\mu, n\sigma^2} \ und \ \frac{\sum X_1 - n\mu}{\sqrt{n}\sigma} \sim N_{0,1}$$

$$\bar{X} \sim N_{\mu, \frac{\sigma^2}{n}} \ und \ \frac{\bar{X} - \mu}{\sigma} \sqrt{n} \sim N_{0,1}$$

Fausregel für Größe von n:
n > 30: Verteilung ist schief aber ohne markante Ausreißer (Exponentialverteilung)
n > 15: Verteilung annähering symmetrisch (Binomialverteilung)
n <= 15: Verteilung annähernd normalverteilt

Ja nachdem, was gegeben ist, wählen:
1. $\frac{\bar{X} - \mu}{\sigma} \sqrt{n} \sim N_{0,1}$
2. $\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$
3. $\frac{\bar{X} - \mu}{S} \sqrt{n} \sim t_{n-1}$

Parameterschätzung

Unbekannter Parameter (z.B. Erwartungswert μ) der Verteilung der Grundgesamtheit soll basierend auf i.i.d Zufallsvariablen geschätzt werden.

Punktschätzer:
für Erwartungswert: Stichprobenmittel $\bar{X}$
für Varianz: Stickprobenvarianz S^2

Keine Aussage über Unsicherheit der Schätzung.

Intervallschätzer:
Parameter wird mit vorgegebener Sicherheit (Konfidenzniveau $1 - \alpha$) überdeckt.
α: Irrtungswahrscheinlichkeit (meist 1-5%)

Konfidenzintervall f. unbekannten EW μ bei bekannter Varianz σ^2

$$I =] \bar{X} - \phi^{-1}\left(1 - \frac{\alpha}{2}\right) \frac{\sigma}{\sqrt{n}} ; \bar{X} + \phi^{-1}\left(1 - \frac{\alpha}{2}\right) \frac{\sigma}{\sqrt{n}} [$$

$\phi^{-1}(x) = qnorm(x, 0, 1)$
Länge Konfidenzintervall:

$$L = 2\phi^{-1}\left(1 - \frac{\alpha}{2}\right) \frac{\sigma}{\sqrt{n}}$$

Gesucht: **Stichprobenumfang n**

$$\sqrt{n} > 2\phi^{-1}\left(1 - \frac{\alpha}{2}\right) \frac{\sigma}{L}$$

Gesucht: **Konfidenzniveau $1 - \alpha$**

$$1 - \frac{\alpha}{2} = \phi\left(\frac{L}{2} \frac{\sqrt{n}}{\sigma}\right)$$

Konfidenzintervall f. unbekannten EW μ bei unkekannter Varianz σ^2

$$I =] \bar{X} - t_{n-1}^{-1}\left(1 - \frac{\alpha}{2}\right) \frac{S}{\sqrt{n}} ; \bar{X} + t_{n-1}^{-1}\left(1 - \frac{\alpha}{2}\right) \frac{S}{\sqrt{n}} [$$

$t_{n-1}^{-1} = qt(x, n-1)$

Hypothesentests

Entscheidung treffen, ob eine Hypothese für unbekannten Parameter einer Verteilung gültig ist, oder nicht. Bei n i.i.d. Zufallsvariablen.
Nullhypothese H_0: Angezweifelte Aussage, der nicht widersprochen werden kann, wenn Stichprobe keinen Gegenbeweis liefert. z.B. $\mu = \mu_0$
Gegenhypothese H_1: Gegenteil von H_0 z.B. $\mu \neq \mu_0$

TG: Testgröße. z.B. Mittelwert
Kritischer Bereich C:
- Werte von TG, die für H_1 sprechen
- Wenn H_0 gültig ist, treten diese Werte mit Wahrscheinlichkeit $\leq \alpha$ auf (**Signifikanzniveau**).
- **α**: Wahrscheinlichkeit, dass H0 verworfen wird, obwohl richtig => Fehler 1. Art

Fehler 1. Art: H_0 wird verworfen, obwohl richtig
Fehler 2. Art: H_0 wird angenommen, obwohl falsch

Realität	Testentscheidung: H_0 wird angenommen.	H_0 wird abgelehnt.
H_0 ist wahr.	richtig	Fehler 1. Art
H_0 ist falsch.	Fehler 2. Art	richtig

- Durch das vorgegebene Signifikanzniveau wird versucht Fehler 1. Art klein zu halten
- Beide Fehler können nicht gleichzeitig kontrolliert werden
- Fehler 2. Art nicht kontrollierbar, wenn man μ_0 nicht kennt

Wahrscheinlichkeit für Fehler 1. Art mit standartisierter Testgröße TG^*:

$$P(TG \in \bar{C}) \geq 1 - \alpha \Leftrightarrow TG^* \in [\phi^{-1}\left(\frac{\alpha}{2}\right), \phi^{-1}\left(1 - \frac{\alpha}{2}\right)]$$

Wird H_0 verworfen => **signifikante Schlußfolgerung**
Wird H_0 nicht verworfen => es lässt sich keine Schlußfolgerung über Fehler 2. Art treffen, eine **schwache Schlußfolgerung**.

Hypothesentests

Gauß-Test
Test für Erwartungswert einer Normalverteilung bei **bekannter Varianz σ_0^2**

H_0	H_1	Testgröße TG	H_0 ablehnen, falls	p-Wert
$\mu = \mu_0$	$\mu \neq \mu_0$	$\frac{\lvert\bar{X} - \mu_0\rvert}{\sigma_0} \sqrt{n}$	$tg > \Phi^{-1}(1 - \frac{\alpha}{2})$	$2(1 - \Phi(tg))$
$\mu \leq \mu_0$	$\mu > \mu_0$	$\frac{\bar{X} - \mu_0}{\sigma_0} \sqrt{n}$	$tg > \Phi^{-1}(1 - \alpha)$	$1 - \Phi(tg)$
$\mu \geq \mu_0$	$\mu < \mu_0$	$\frac{\bar{X} - \mu_0}{\sigma_0} \sqrt{n}$	$tg < \Phi^{-1}(\alpha)$	$\Phi(tg)$

t-Test
Test für Erwartungswert einer Normalverteilung bei unbekannter Varianz

H_0	H_1	Testgröße TG	H_0 ablehnen, falls	p-Wert
$\mu = \mu_0$	$\mu \neq \mu_0$	$\frac{\lvert\bar{X} - \mu_0\rvert}{S} \sqrt{n}$	$tg > t_{n-1}^{-1}(1 - \frac{\alpha}{2})$	$2(1 - t_{n-1}(tg))$
$\mu \leq \mu_0$	$\mu > \mu_0$	$\frac{\bar{X} - \mu_0}{S} \sqrt{n}$	$tg > t_{n-1}^{-1}(1 - \alpha)$	$1 - t_{n-1}(tg)$
$\mu \geq \mu_0$	$\mu < \mu_0$	$\frac{\bar{X} - \mu_0}{S} \sqrt{n}$	$tg < t_{n-1}^{-1}(\alpha)$	$t_{n-1}(tg)$

p-Wert
„beobachtetes Signifikanzniveau."
- Wahrscheinlichkeit, bei Zutreffen von H_0 den beobachteten Wert tg der Testgröße oder einen noch stärker von μ_0 abweichenden Wert zu bekommen.
- Größter Wert von α, für den H_0 nicht abgelehnt wird
- Wenn H_0 abgelehnt wird, ist das Ergebnis immer signifikant!
- p-Wert <0,01: sehr hohe Signifikanz
- p-Wert <0,05: hohe Signifikanz
- p-Wert <0,1: schwache Signifikanz
- p-Wert >0,1: keine Signifikanz

Zusammenhang Konfidenzintervall – Hypothesentest
Konfidenzintervall zum Konfidenzniveau $1 - \alpha$ ist Annahmebereich für H_0
Test:
H0 wird abgelehnt, falls $\mu_0 \notin I$
H0 wird angenommen, falls $\mu_0 \in I$

Numerik/Fehleranalyse

Fehlerquellen:
- Rundungsfehler
- Fehler aufgrund vn Gleitpunktarithmetik
- Diskretisierungsfehler

Rundungsfehler:
- absolut: | gerundetes Ergbnis – tatsächliches |
- relativ: $\frac{|r \ d(x) - x|}{|x|}$ => größenordnungsbereinigt

Gleitpunktarithmetik
Berechnungsreihenfolge spielt eine Rolle!
Bei Addition in aufsteigender Reihenfolge.

$$\sum_{k=1}^{10} \frac{1}{k} \Rightarrow \sum_{k=0}^{10-1} \frac{1}{10 - k}$$

Auslöschung
Entsteht bei der Subtraktion von zwei fast gleich großen Zahlen.
$3.97403 * 10^2 - 3.97276 * 10^2 = 1.27000 * 10^{-1}$
Erhöhung der Signifikanz weniger signifikanter Stellen

$$x^2 + 200x - 0.000015$$

$$(1) \ x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \quad (2) x_{1,2} = \frac{2c}{-b \mp \sqrt{b^2 - 4ac}}$$

x_1 mit (2) und x_2 mit (1) um Auslöschung zu vermeiden. Weil b > 0

Trick, Erweitern mit 3. binomischen Formel:

$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$
$$= \frac{(-b \mp \sqrt{b^2 - 4ac})(-b \pm \sqrt{b^2 - 4ac})}{(-b \mp \sqrt{b^2 - 4ac}) \cdot 2a}$$
$$= \frac{b^2 - (b^2 - 4ac)}{(-b \mp \sqrt{b^2 - 4ac}) \cdot 2a} = \frac{2c}{-b \mp \sqrt{b^2 - 4ac}}$$

TODO: vielleicht noch Kondition je nach Prüfungsrelevanz

Interpolation

Polynom durch bestimmte Punkte bilden.

Vandermonde Ansatz

Beispiel:

x	-2	3	1
y	-15	-5	3

Ansatz: Polynom 2. Grades
$p_2(x) = a_0 + a_1 x + a_2 x^2$

$$\left.\begin{array}{ll} (I) & a_0 - 2a_1 + 4a_2 = -15 \\ II & a_0 + 3a_1 + 9a_2 = -5 \\ III & a_0 + a_1 + a_2 = 3 \end{array}\right\} \Leftrightarrow \begin{pmatrix} 1 & -2 & 4 \\ 1 & 3 & 9 \\ 1 & 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} -15 \\ -5 \\ 3 \end{pmatrix}$$

Lösung mit Gauß-Algorithmus: $a_0 = 1, \ a_1 = 4, \ a_2 = -2$

$$\Rightarrow p_2(x) = \underline{1 + 4x - 2x^2}$$

Lagrange Ansatz

Ansatz: $p_2(x) = y_0 \cdot L_0(x) + y_1 \cdot L_1(x) + y_2 \cdot L_2(x)$
Polynome 2. Grades gebildet aus den Stützstellen

Beispiel:

x	-2	3	1
y	-15	-5	3

$$L_0(x) = \prod_{j=1}^{2} \frac{x - x_j}{x_0 - x_j} = \frac{\overbrace{(x-3)}^{j=1}\overbrace{(x-1)}^{j=2}}{(-2-3)(-2-1)} = \frac{1}{15}(x-3)(x-1)$$

$$L_1(x) = \prod_{\substack{j=0 \\ j \neq 1}}^{2} \frac{x - x_j}{x_1 - x_j} = \frac{(x+2)(x-1)}{(3+2)(3-1)} = \frac{1}{10}(x+2)(x-1)$$

$$L_2(x) = \prod_{j=0}^{1} \frac{x - x_j}{x_2 - x_j} = \frac{(x+2)(x-3)}{(1+2)(1-3)} = -\frac{1}{6}(x+2)(x-3)$$

$$p_2(x) = -(x-3)(x-1) - \frac{1}{2}(x+2)(x-1) - \frac{1}{2}(x+2)(x-3) = \dots = \underline{1 + 4x - 2x^2}$$

Newton Ansatz
$p_3(x) = c_0 + c_1(x - x_0) + c_2(x - x_0)(x - x_1) + c_3(x - x_0)(x - x_1)(x - x_2)$

Interpolationsbedingungen:

$y_0 = c_0$
$y_1 = c_0 + c_1(x_1 - x_0)$
$y_2 = c_0 + c_1(x_2 - x_0) + c_2(x_2 - x_0)(x_2 - x_1)$
$y_3 = c_0 + c_1(x_3 - x_0) + c_2(x_3 - x_0)(x_3 - x_1) + c_3(x_3 - x_0)(x_3 - x_1)(x_3 - x_2)$

...mit dividierten Differenzen

x	y			
x_0	$\underline{y_0}$ c_0			
		c_1: $\underline{[y_0, y_1]} = \frac{y_1 - y_0}{x_1 - x_0}$		
x_1	y_1		c_2: $\underline{[y_0, y_1, y_2]} = \frac{[y_1, y_2] - [y_0, y_1]}{x_2 - x_0}$	
		$[y_1, y_2] = \frac{y_2 - y_1}{x_2 - x_1}$		c_3: $\underline{[y_0, y_1, y_2, y_3]}$
x_2	y_2		$[y_1, y_2, y_3] = \frac{[y_2, y_3] - [y_1, y_2]}{x_3 - x_1}$	
		$[y_2, y_3] = \frac{y_3 - y_2}{x_3 - x_2}$		
x_3	y_3			

Horner Schema
$p_3(x) = a_3 x^3 + a_2 x^2 + a_1 x + a_0 = ((a_3 x + a_2)x + a_1)x + a_0$

Chebyshev-Punkte
Nicht-äquidistante Stützstellen, die an den Intervallgrenzen dichter sind. So wird Konvergenz erreicht.

Spline-Interpolation
Ansatz um Oszillationen zu vermeiden.
Hinreichend glatte ((k-1)-mal stetig differenzierbare), stückweise zusammengesetzte Polynome, dog. Splines vom Grad k.

Ein kubischer Spline ist eine Funktion $S : [x_0, x_n] \to \mathbb{R}$, die auf den Teilintervallen $[x_i, x_{i+1}]$ zweimal stetig differenzierbar und stückweise aus kubischen Polynomen $S_i(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3$ zusammengesetzt ist.

TODO: SPLINE Beispiel S.7

Numerische Integration

Nicht integrierbarer Integrand wird durch integrierbaren Integranden ersetzt.

Trapezregel

$$\int_a^b f(x) \, dx \approx \frac{b - a}{2}(f(a) + f(b))$$

Simpson Regel

$$\int_a^b f(x) \, dx \approx \frac{b - a}{6}\left(f(a) + 4f\left(\frac{a + b}{2}\right) + f(b)\right)$$

Newton Cotes Regeln
Basieren auf äquidistanten Knoten $t_j = \frac{j}{k}$
Eine Integrationsregel hat Ordnung p, wenn sie für Polynome vom Grad <= p-1 exakte Werte liefert.

k	α_i					Methode	Ordnung p
1	$\frac{1}{2}$	$\frac{1}{2}$				Trapez	2
2	$\frac{1}{6}$	$\frac{4}{6}$	$\frac{1}{6}$			Simpson	4
3	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$		$\frac{3}{8}$-Rule	4
4	$\frac{7}{90}$	$\frac{32}{90}$	$\frac{12}{90}$	$\frac{32}{90}$	$\frac{7}{90}$	Milne	6

Zusammengesetzte Trapezregel
Zusammengesetzte Simpson Regel
Grenzen der Newton Cotes Regeln
Gauß Quadratur
Unstetige Funktionen dürfen nicht numerisch integriert werden

TODO:

- Winkelfunktionen
- Ableitung und Integral Regeln
- Viele Bilder für Verteilungen usw.
- Mitternachtsformel

Beispielaufgaben:

Ü6: 2.
Ü6: 3. b) (Quantile) und c)
Ü7 komplett
Ü5 Lösungsansatz Verschiebungssatz
Beispiel 3.3.4 im Skript
Beispiele im Skript Kapitel Verteilungen