Statistik und Wahrscheinlichkeitstheorie

Ausarbeitung möglicher Theoriefragen

Prof. Dutter / TU Wien

verfasst von Michael Jaros

basierend auf den Fragen von Oliver Zendel und Patrick Kastner

letzte Änderung: 25.06.2006 12:22

Alle Angaben ohne Gewähr. Inhaltliche Quellen: Skriptum von Prof. Dutter, Wikipedia, diverse Ausarbeitungen (zu geringem Teil). Verbesserungsvorschläge etc. bitte an: e0225858@student.tuwien.ac.at

Viel Glück!

Wahrscheinlichkeitstheorie

Erkläre (Elementar-)Ereignisse, Ereignisalgebra, Ereignisraum, Operationen, Borel-Mengen.

Führt man einen Versuch (zB Würfeln mit 2 Würfeln) durch, kann dieser verschiedene **Versuchsausgänge** haben ((1,1), (1,2), … , (6,6)). Die **Menge aller möglichen Versuchsausgänge** heißt Ω, und jede Teilmenge (zB 2 gleiche = {(1,1),(2,2), …, (6,6)} heißt **Ereignis**, einpunktige Teilmengen (zB (1,1)) heißen **Elementarereignisse**.

Auf Ereignisse kann man folgende **Operationen** anwenden: Durchschnittsbildung (A und B), Vereinigung (A oder B), Komplementbildung (nicht A). Das **unmögliche Ereignis Ø** und das **sichere Ereignis Ω** sind Sonderfälle.

Ereignisse können **disjunkt** sein (ihr Durchschnitt ist leer). Ein Ereignis A kann ein anderes B **implizieren (A c B)**, dh A ist in B enthalten. Eine Reihe von Ereignissen ist genau dann **Zerlegung** eines anderen Ereignisses, wenn ihre Vereinigung das andere Ereignis ergibt und sie alle disjunkt sind. **deMorgan-Regeln** gelten: $A \cap B = (A^c \cup B^c)^c$ und $A \cup B = (A^c \cap B^c)^c$

Die Menge aller betrachteten Ereignisse dh die Menge aller Teilmengen von Ω heißt **Ereignisraum** $\mathcal{A}$ (zB {{(1,1)},{1,1}u{1,2}, …}). Zusammen mit den Operationen und ΩØ heißt der Ereignisraum **Ereignisalgebra**. Ist diese bezüglich Vereinigung und Komplementbildung abgeschlossen, heißt sie **Ereignis-σ-Algebra**.

Ist die Menge aller möglichen Versuchsausgänge R, so betrachtet man nicht einfach P(R), sondern nimmt die Menge aller links halboffenen Intervalle sowie Vereinigungen und Komplemente dieser. Die kleinste σ-Algebra, die all diese Mengen enthält, heißt **borelsche σ-Algebra** $\mathcal{B}$, Teilmengen davon **Borel-Mengen**.

Welche Arten von Wahrscheinlichkeiten gibt es?

Während die **a-priori-Wahrscheinlichkeit** theoretischer Natur ist und im Vorhinein angegeben wird (Münze hat 2 gleich wahrscheinliche Seiten, daher P(W) = ½), bestimmt man die **a-posteriori-Wahrscheinlichkeit** empirisch (57 von 100 Münzwürfen haben Wappen gezeigt => P(W)=57/100).

Was ist ein Wahrscheinlichkeitsmaß?

Ein **Maß** μ weist jedem Elementarereignis aus Ω und damit jedem Ereignis aus $\mathcal{A}$ ein **Gewicht** zu, ist also eine Funktion vom Ereignisraum $\mathcal{A}$ in $\mathbb{R}^+$. Für Maße gilt natürlich σ-Additivität, dh die Summe aller Maße einer Zerlegung eines Ereignisses ergibt das Maß dieses Ereignisses. Gilt auch $\mu(\Omega) = 1$, hat man ein **Wahrscheinlichkeitsmaß**, und man schreibt fortan P statt μ. Die Wahrscheinlichkeit eines Ereignisses ist dann die Summe der Wahrscheinlichkeiten seiner Elementarereignisse. Sind alle Gewichte gleich, ist die Wahrscheinlichkeit eines Ereignisses gleich der Anzahl der günstigen dividiert durch die Anzahl der möglichen Elementarereignisse.

Rechenregeln: $P(A) = 1 - P(A^c)$ und $P(A - B) = P(A) - P(A \cap B)$ und

Was ist ein Wahrscheinlichkeitsraum?

Die Menge aller möglichen Versuchsausgänge zusammen mit der darauf definierten σ-Algebra $\mathcal{A}$ und einem Wahrscheinlichkeitsmaß P ($\mathcal{A} \rightarrow [0,\infty]$) bilden zusammen einen Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$.

Was ist eine bedingte Wahrscheinlichkeit? Was bedeutet Unabhängigkeit von Ereignissen?

Durch das Vorwissen wird der **Ereignisraum eingeschränkt**, sodass die Wahrscheinlichkeiten sich ändern (zB 1 Würfel, A…“Zahl≤3“ B…“Zahl=1“, so ist $P(B) = 1/6$, aber $P(B|A)$ nur mehr 1/3). Es gilt: $P(A|B) = P(A \cap B) / P(B)$, also einfach die Wahrscheinlichkeit für *A und B*, normiert auf den durch den Eintritt von B bereits eingeschränkten Ereignisraum. Wenn der Eintritt von B keinen Einfluss mehr hat, dh $P(A|B) = P(A)$, nennt man die Ereignisse A und B **unabhängig**. Aus obiger Formel erhält man ganz einfach: $P(A \cap B) = P(A|B) * P(B)$. Für unabhängige Ereignisse gilt dann wegen $P(A|B) = P(A)$ ganz einfach: $P(A \cap B) = P(A) * P(B)$.

Was ist eine Zufallsvariable? Wann ist sie diskret/kontinuierlich? Was ist eine Dichtefunktion? Was ist eine Verteilungsfunktion? Transformationen von Zufallsvariablen?

Eine **Zufallsvariable** ist formal eine Abbildung von einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ in einen einfacheren $(\mathbb{R}, \mathcal{B}, P_1)$. Jede Menge B aus $\mathcal{B}$ muss ein Urbild $X^{-1}(B)$ als Element aus $\mathcal{A}$ besitzen.

Eine **diskrete** Zufallsvariable kann höchstens abzählbar viele verschiedene Werte annehmen. $p_i = P(X=x_i)$ gibt eine Punktwahrscheinlichkeit an und heißt **Wahrscheinlichkeitsfunktion**, und die Summe über alle p_i ergibt 1. $F(x) = P(X \leq x)$ ist gleich der Summe über p_i bis x und heißt **Verteilungsfunktion**.

Eine Zufallsvariable ist **stetig**, wenn ihre Verteilungsfunktion **absolut stetig** ist, dh sie ist für alle x aus R als Integral über f darstellbar. Die **Verteilungsfunktion** $F(x) = P(X \leq x)$ ist gleich dem Integral über f bis x. f heißt **Dichtefunktion** und ist die Ableitung der Verteilungsfunktion.

Transformationen von Zufallsvariablen: Addition wirkt sich nur auf den Mittelwert aus (verschiebt sich um addierten Betrag), Multiplikation wirkt sich auf die Varianz aus (diese wird dividiert). Zieht man den Mittelwert einer normalverteilten Zufallsvariablen ab und dividiert durch die Standardabweichung, erhält man eine N(0,1), also standard-normalverteilte Zufallsvariable.

Was sind die mathematische Erwartung und die Varianz einer Zufallsvariablen?

Die **Erwartung** (oder der Mittelwert) einer Zufallsvariablen X ist $E(X) = \int x \cdot f(x) \partial x$ im stetigen Fall sowie $E(X) = \sum x_i p_i$ im diskreten Fall. Die **Varianz** einer Zufallsvariablen X ist $\mathrm{Var}(X) = E(X)^2 - (EX)^2$.

Was ist eine mehrdimensionale (multivariate) Zufallsvariable? Was ist eine Randverteilung? Wann sind zwei Zufallsvariablen voneinander unabhängig?

In der Praxis wird selten eine Größe alleine untersucht, da die Zusammenhänge zwischen verschiedenen Merkmalen wichtig sind. Man kann nun einen **p-dimensionalen Zufallsvektor** $(X_1, \ldots, X_p)$ definieren. Verteilungsfunktion und Wahrscheinlichkeitsdichte können ähnlich wie im eindimensionalen Fall definiert werden.

Die Verteilung der einen Zufallsvariablen für einen bestimmten Wert der anderen heißt **Randverteilung**.

Zwei Zufallsvariablen X und Y sind genau dann voneinander **unabhängig**, wenn die Verteilungsfunktion des Zufallsvektors (X, Y) gleich dem Produkt der beiden Randverteilungsfunktionen F_X*F_Y ist: $F(x,y) = F_X(x)*F_Y(y)$ für alle x aus $\mathbb{R}$.

Beschreibende Statistik

Welche Momente einer Verteilung gibt es?

1. Momente sind die **Lageparameter**: **Mittelwert** (Erwartungswert) , **Median** (Zentralwert), **Modalwert** (Modus), **Quantile** (Perzentile)…

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

Der Median ist der mittlere Wert einer geordneten Stichprobe bzw. das Mittel der beiden mittleren Werte bzw. er ist das 50%-Quantil, siehe Quantil. Der Modus ist der häufigste Wert einer Verteilung. Ein p%-Quantil wird so gewählt, dass p % der Verteilung kleiner als dieses Quantil sind. Quartile sind 25 % bzw. 75 %-Quantile. Die Verteilungsfunktion $F(Q_p)$ ergibt also für das p-Quantil genau $F(Q_p) = p$.

2. Momente sind **Streuungsparameter**: **Varianz**, **Standardabweichung** (Streuung – grobe Faustformel: (max – min) / 3)

Varianz: $s^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$ Standardabweichung: $s = \sqrt{s^2}$

Die Varianz ist also das mittlere Abstandsquadrat vom Mittelwert. Siehe Varianz.

3. Moment ist die **Schiefe** (>0 → mehr positive als negative Abweichungen vom Mittelwert):

$$\frac{\sum_{i=1}^{n}(x_i - \overline{x})^3}{n \cdot s^3}$$

4. Moment ist die **Kurtosis** (>0 → ‚heavy tails'):

$$\frac{\sum_{i=1}^{n}(x_i - \overline{x})^4}{n \cdot s^4} - 3$$

Bei Schiefe und Kurtosis hat es sich eingebürgert, einfach durch n zu dividieren. Durch das Subtrahieren von 3 hat die N(0,1)-Verteilung eine Kurtosis von 0.

Was ist das Messniveau? Welche Skalen gibt es, was charakterisiert sie, welche Momente sind dort verwendbar?

Das **Messniveau** gibt eine Unterteilung in verschiedene Datentypen / Skalen vor:

Nominalskala (zB Geschlecht): keine Ordnung:
Modalwert
Ordinalskala (zB Schulnoten): Ordnung, aber keine Abstände:
Modalwert, Median, Quantile
Intervallskala (zB Temparatur °C): Ordnung, Abstände, kein abs. Nullpunkt:
Modalwert, Median, Quantile, Mittel
Verhältnisskala (zB Temparatur K): Ordnung, Abstände, absoluter Nullpunkt:
Modalwert, Median, Quantile, Mittel

Was ist ein MedMed?

Die Verteilung wird am Median nach rechts „umgeklappt“ (dadurch zB bei Normalverteilung doppelt so hoch). Von der Resultierenden Verteilung wird der Median berechnet.

Was ist die Varianz? Warum n-1 Freiheitsgrade?

Die Varianz berechnet sich folgendermaßen: $s^2 = \frac{\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2}{n-1}$

Man geht von n unabhängigen Stichprobenwerten aus, also n Freiheitsgrade. Da man jedoch die Differenz vom bereits bekannten Mittelwert bildet und die Summe aller Abweichungen 0 ist, kennt man bei n-1 Abweichungen bereits auch die letzte Abweichung, daher n-1 Freiheitsgrade. Ein nicht unerwünschter Nebeneffekt ist, das die Varianz nun für n=1 nicht definiert ist statt 0.

Was ist die Standardabweichung?

Die Standardabweichung ist die mittlere Abweichung vom arithmetischen Mittel einer Verteilung. Sie berechnet sich als Quadratwurzel der Varianz und ist ein Streuungsmaß, gibt also die Breite der Verteilung an. Eine sehr grobe Schätzung für die Standardabweichung lautet: (Maximalwert – Minimalwert) / 3.

Was ist ein Wahrscheinlichkeitsnetz?

Im Wahrscheinlichkeitsnetz trägt man die Werte einer geordneten Stichprobe sowie $^{(i-0.5)}/_{n}$ auf. Bei Normalverteilung ergibt sich durch die übliche Anordnung der y-Achse annähernd eine Gerade. Aus der Ausgleichsgeraden lassen sich die Parameter der Verteilung grafisch schätzen.

Verteilungen

Welche Verteilungen kennst du?

Siehe nachfolgende Fragen.

Normalverteilung

$N(\mu, \sigma^2)$, kontinuierlich, symmetrisch.

Die Normalverteilung gilt zumindest annähernd für viele natürliche Prozesse. Als dem Zentralen Grenzwertsatz folgt, dass aus vielen einzelnen unabhängigen Einflüssen entstehende Größen normalverteilt sind. Die Normalverteilung ist bei großem n eine gute Näherung für die Binomialverteilung.

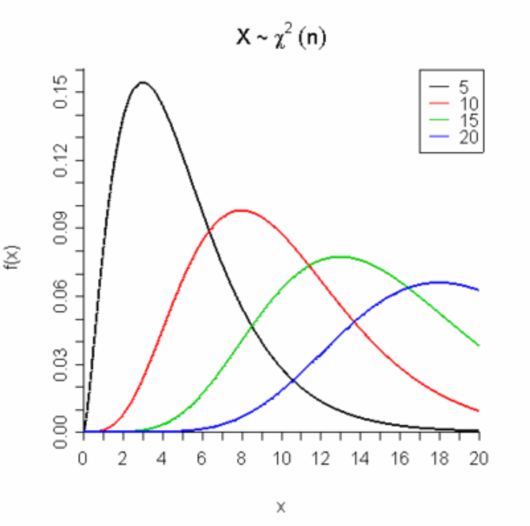
Dichte: Gaußsche Glockenkurve mit arithmetischem Mittel der Verteilung als Höhepunkt, Mittel ± Standardabweichung als Wendepunkte. Standardnormalverteilung ist N(0,1). Verteilungsfunktion: Anfang und Ende sehr flach und um den Median sehr steil.

Chi-Quadrat-Verteilung

χ^2 (n), kontinuierlich, asymmetrisch.

n nennt man die Freiheitsgrade, der Erwartungswert ist n, die Varianz 2n. Meist verwendet man die zentrale χ^2-Verteilung mit nur einem Parameter n.

Die Summe quadrierter Zufallsvariablen (normalverteilt) ist χ^2-verteilt. Man verwendet diese Verteilung zB zur Schätzung der Varianz. Die χ^2-Verteilung ist sozusagen die quadrierte Normalverteilung.



Binomialverteilung

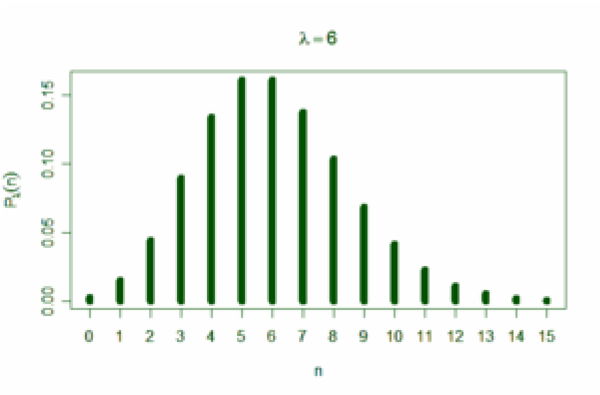
B (n, p), diskret, asymmetrisch. Für p=0,5 symmetrisch. Erwartungswert np, Varianz np(1-p).

Die Binomialverteilung kann zur Beschreibung von n Versuchen, die mit der Wahrscheinlichkeit p erfolgreich sind, verwendet werden. Für große n kann die Binomialverteilung gut durch die Normalverteilung angenähert werden. B (1,p) mit nur 1 Versuch heißt auch **Bernoulliverteilung**.

Poissonverteilung

P (λ), diskret, asymmetrisch. λ ist zugleich Erwartungswert, Varianz und Schiefe.

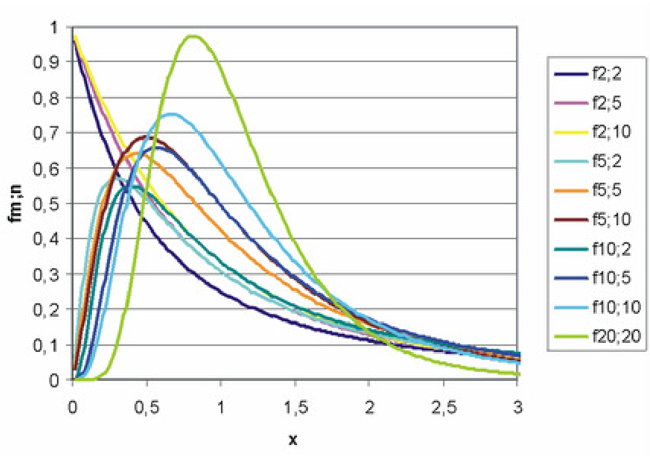
Für große λ lässt sich die Poissonverteilung durch die Normalverteilung annähern. Sie ist die Grenzverteilung der Binomialverteilung (für p $\rightarrow$ 0, n $\rightarrow$ ∞). Sie wird typischerweise für die Zahl von Phänomenen innerhalb einer Zeiteinheit verwendet.



F-Verteilung

F(m,n), kontinuierlich, asymmetrisch. m und n sind Freiheitsgrade.

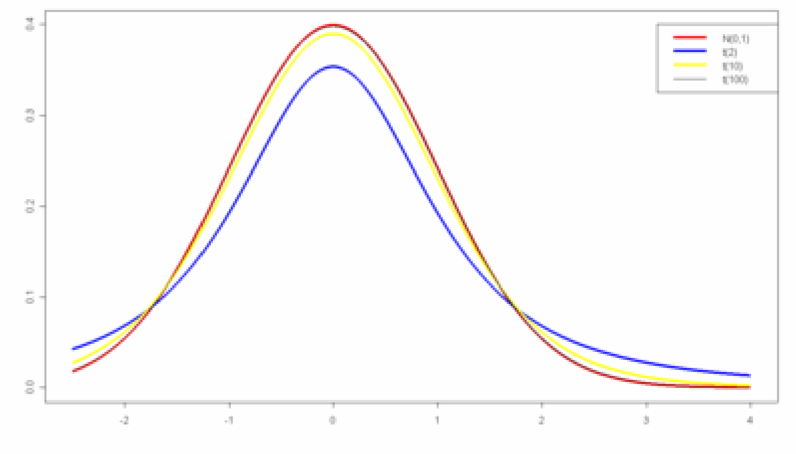
Die F-Verteilung wird in der Varianzanalyse verwendet, um festzustellen, ob die Populationen zweier Stichproben die gleiche Varianz haben.



t-Verteilung

t(m), kontinuierlich, symmetrisch. m sind Freiheitsgrade. Erwartungswert: 0.

Die t-Verteilung wird zur Schätzung des Erwartungswertes bei unbekannter Varianz verwendet (statt der Normalverteilung bei bekannter Varianz). Für große m (>30) kann die t-Verteilung durch die N(0,1)-Verteilung angenähert werden.



Rechtecksverteilung

R (a, b), diskret, symmetrisch. Erwartungswert: ½(a+b), Varianz: ½$(b-a)^2$

Exponentialverteilung

X ~ R (0,1), Y = -ln (X) ➔ Y folgt Exponentialverteilung

Lognormalverteilung

X ~ N (μ, σ^2), Y = e^X ➔ Y folgt Lognormalverteilung

Wie lässt sich die Binomialverteilung durch die Normalverteilung annähern?

Für große Werte von n lässt sich die Binomialverteilung durch die Normalverteilung mit dem Mittelwert μ = np und der Varianz σ^2 = np(1-p) gut annähern. Für Werte von p nahe 0 oder 1 ist die Verteilung allerdings recht schief, für p nahe 0.5 ist die Annäherung recht gut.

Analytische Statistik

Was ist eine Stichprobe?

Eine Untermenge einer Population heißt **Stichprobe**. Mathematisch gesehen stellt sie einen n-dimensionalen Zufallsvektor (X_1, …, X_n) mit unabhängig und identisch verteilten Elementen X_i dar. Die **Stichprobenwerte** (x_1, …, x_n) sind eine Realisation dieses Zufallsvektors. Damit mit einfachen Mitteln Aussagen über die Verteilung oder ihre Parameter gemacht werden können, müssen die Stichprobenwerte zufällig aus der Population gewählt werden.

Was ist ein Schätzer? Wann ist er erwartungstreu, konsistent, oder effizient?

Ein **Schätzer** (eine **Schätzfunktion**) t berechnet einen Parameter θ einer Verteilung näherungsweise aus Stichprobenwerten: $\hat{\theta} = t(x_1, \ldots, x_n)$. Eine Funktion der Stichprobe wird allgemein als **Statistik** bezeichnet (und ist auch eine Zufallsvariable). Eine Realisation eines Schätzers heißt **Schätzwert** oder **Schätzung**.

Eine Schätzfunktion heißt **erwartungstreu**, wenn der Erwartungswert der Schätzfunktion den geschätzten Parameter ergibt.

Eine Schätzfunktion heißt **konsistent**, wenn sie sich mit wachsendem n (größerer Stichprobe) immer mehr dem geschätzten Parameter nähert, d. h. ihre Varianz kleiner wird.

Ein Schätzer ist dann **effizient**, wenn er die kleinstmögliche Varianz aufweist.

Was ist ein Konfidenzintervall?

Ein (1-α)-Konfidenzintervall für einen Parameter ist ein Intervall um den *geschätzten* Parameter, in dem der *tatsächliche* Parameter mit der **Überdeckungswahrscheinlichkeit** 1-α liegt. α heißt dabei **Konfidenzzahl**.

[Ergänzen: Zusammenhang Quantile und Konfidenzintervall]

Was ist eine Hypothese? Was ist die Nullhypothese, was die Alternative? Fehler 1./2. Art?

Eine **Hypothese** ist eine Annahme (zB über die Verteilung einer Zufallsvariablen oder den Wert eines Parameters), die getestet werden soll. Führt man einen statistischen Test durch, so nimmt man die so genannte **Nullhypothese** H_0 an. Gleichzeitig gibt es immer eine Gegenhypothese oder **Alternative** H_1. Die Teststatistik ist eine nach einer bestimmten Vorschrift berechnete Zufallsvariable. Fällt der Wert der Teststatistik in den so genannten **kritischen Bereich**, wird die Nullhypothese H_0 verworfen und H_1 angenommen. Das **Signifikanzniveau** α heißt auch **Fehlerwahrscheinlichkeit 1. Art** und gibt die Wahrscheinlichkeit an, dass die richtige Hypothese abgelehnt wird. Mit der **Fehlerwahrscheinlichkeit 2. Art** β wird die falsche Hypothese angenommen. Die

Fehlerwahrscheinlichkeiten können durch Vergrößern des Stichprobenumfangs verkleinert werden. 1-β heißt **Macht** oder **Schärfe** des Tests.

[Ergänzen: Grafik]

Was ist Regression? Was ist das Regressionsproblem? Regressionsgerade? Residuen? Wie testet man auf Abhängigkeit einer Variablen x?

Das **Regressionsproblem** behandelt die Verteilung einer Variablen Y, wenn (mindestens) eine andere Variable x bestimmte, nicht zufällige Werte annimmt. x heißt **unabhängig** und ist keine Zufallsvariable. Y ist Zufallsvariable und **abhängig** von x, sofern Regression vorliegt. Für jeden gewählten Wert von x gibt es eine Verteilung von Y mit einem Mittelwert $\mu_{y.x}$ und einer Varianz $\sigma^2_{y.x}$.

Liegt einfache, **lineare Regression** vor, so kann die Abhängigkeit der Mittelwerte $\mu_{Y.X}$ von Y durch die **Regressionsgerade** angegeben werden:

$$\mu_{y.x} = a + b \cdot (x - \bar{x})$$

Die **Parameter der Regressionsgeraden** a und b werden aus den Stichprobenwerten geschätzt, wobei man für a den Wert $\bar{x}$ und für b den Quotienten $\frac{s_{XY}}{s_X^2}$ verwendet. s_{XY} ist übrigens die **empirische Kovarianz**, die sich aus der Summe aller Produkte von Mittelabweichungen beider Variablen dividiert durch n-1 Freiheitsgrade ergibt. Auch die Varianz der Beobachtungen s^2 kann geschätzt werden durch:

$$s^2 = \frac{\sum (y_i - \hat{y}_i)^2}{n-2}$$

Die Differenzen zwischen gemessenen und geschätzten Werten $y_i - \hat{y}_i$ nennt man auch **Residuen**. Die Gerade wird so gewählt, dass die Summe der quadrierten Residuen minimal wird.

Nimmt man die Verteilung von Y für jedes x als normal an, kann man **Konfidenzintervalle** für die Parameter a, b, σ^2 und $\mu_{y.x}$ angeben (mit Hilfe der Parameterschätzungen und der t-Verteilung).

Test auf Regression: Die Nullhypothese H_0: b = 0 besagt, dass alle Mittelwerte von Y gleich sind (Alternative: $b \neq 0$) und daher keine Regression vorliegt. Der kritische Bereich ist $|T| > t_{n-2;1-\alpha/2}$, und als Teststatistik verwendet man:

$$T = \frac{b \cdot s_X \cdot \sqrt{n-1}}{S}.$$

Was ist Korrelation? Was ist das Korrelationsproblem? Was sind Kovarianz, empirische Kovarianz und Korrelationskoeffizient? Wie testet man auf Unkorreliertheit?

Das **Korrelationsproblem** behandelt die Frage, ob **Korrelation** vorliegt. Dies ist der Fall, wenn es einen Zusammenhang zwischen den Verteilungen zweier Zufallsvariablen X und Y gibt. Es wird also die gemeinsame Verteilung von X und Y betrachtet, ohne eine Variable zu fixieren. Oft geht man in solchen Problemen von einer **bivariaten Normalverteilung** aus. Für jedes X gibt es eine Verteilung von Y und umgekehrt.

Die **Korrelation** ρ zwischen X und Y ergibt sich aus der **Kovarianz** σ_{XY} dividiert durch das Produkt der beiden Standardabweichungen $\sigma_X\ \sigma_Y$ und ist eine dimensionslose Größe im Intervall $(-1,1)$; bei $\rho = 0$ sind X und Y unabhängig. Als Schätzung verwendet man den **empirischen Korrelationskoeffizienten**, welcher sich wiederum aus der **empirischen Kovarianz** s_{XY}, dividiert durch das Produkt der empirischen Standardabweichungen $s_X\ s_Y$ berechnet:

$$r_{XY} = \frac{1}{s_X s_Y} \frac{1}{n-1} \sum (x_i - \bar{x})(y_i - \bar{y}).$$

Für den **Test auf Unkorreliertheit** sind H_0: $\rho = 0$ und H_1: $\rho \neq 0$. Der kritische Bereich ist $|T| > t_{n-2;1-\alpha/2}$, und als Teststatistik verwendet man:

$$T = R \cdot \sqrt{\frac{n-2}{1-R^2}}.$$

Wozu dienen 1-Stichproben-t-Test und 2-Stichproben-t-Test?

Man vergleicht damit die Mittel zweier Populationen untereinander (2-Stichproben-t-Test) oder man vergleicht das Mittel einer Population mit einem vorgegebenen Wert (1-Stichproben-t-Test, nur bei unbekannter Standardabweichung).

Wozu dient der χ^2-Anpassungstest?

Mit Hilfe dieses Tests kann man eine Hypothese über die Form einer Verteilung prüfen. Dazu teilt man die Stichprobe in k Klassen ein und berechnet

$$T = \sum_{i=1}^{k} \frac{(h_i - e_i)^2}{e_i},$$

wobei h_i die Häufigkeiten und e_i die theoretischen Wahrscheinlichkeiten (aufgrund der mit H_0 angenommenen Verteilung) p_i multipliziert mit n sind. Als kritischen Bereich nimmt man:

$T > \chi^2_{k-1;1-\alpha}$.

Wozu dient der Kolgorov-Smirnov-Test?

Man testet damit, ob eine hypothetische Verteilung F_0 zugrunde liegt, indem man die absolute Differenz zwischen empirischer und hypothetischer Verteilungsfunktion betrachtet.

Was geschieht bei der Varianzanalyse? Was steht in der Varianzanalyse-Tafel?

Eine betrachtete Größe weist oft eine **Variation** auf, die sich aus zufälliger Variation sowie Variation durch einen bestimmten Einfluss zusammensetzt. In der Varianzanalyse trennt man diese beiden Variationen (**Varianzzerlegung**). Dazu wird die **Quadratsumme** (Summe der Abweichungen vom Stichprobenmittel) zerlegt.

Mit Hilfe der Varianzanalyse kann man die Mittelwerte von k Normalverteilungen vergleichen. Man nimmt H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$ gegen H_1: $\mu_r \neq \mu_s$ (für mindestens ein $r \neq s$) an. Danach berechnet man die Quadratsumme innerhalb jeder Stichprobe q_i und die Quadratsumme zwischen den Stichproben q_z:

$$q_I = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (x_{ij} - \overline{x}_j)^2 \qquad q_Z = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (\overline{x}_j - \overline{x})^2 = \sum_{j=1}^{k} n \cdot (\overline{x}_j - \overline{x})$$

Die Varianzanalyse-Tabelle sieht so aus:

Variation	**FG**	**q**	**s^2**	**F**
Zwischen Gruppen	k-1	q_Z	$s_Z^2 = q_Z / (k-1)$	s_Z^2 / s_I^2
Innerhalb d. Gruppen	n-k	q_I	$s_I^2 = q_I / (n-k)$	
Gesamt	n-1	q		

Der Wert der Teststatistik F wird mit dem kritischen Bereich $F > F_{k-1,n-k;1-\alpha}$ getestet.

Was ist das Klassifizierungsproblem? Was ist eine Kontingenztafel?

Oft möchte man Dinge zählen, die in bestimmte Kategorien fallen. Es gibt das **einfache Klassifizierungsproblem** (zB 4 Klassen für Aussehen von Erbsen nach Kreuzungsversuch) und das **zweifache Klassifizierungsproblem** (zB Haarfarbe und Augenfarbe). Man hat jeweils vorgegebene theoretische Häufigkeiten und untersucht, ob die Abweichungen der tatsächlichen Häufigkeiten von ersteren nur zufälliger Natur sind. Eine Tabelle, die die absoluten Häufigkeiten der Merkmale gemeinsam darstellt, heißt **Kontingenztafel**. Als Teststatistik dient bei einem kritischen Bereich von $T > \chi^2_{r-1,c-1;1-\alpha}$:

$$T = \sum_{j=1}^{k} \frac{(h_j - e_j)^2}{e_j}$$

Was macht die likelihood-Funktion? Was ist die Maximum-Likelihood-Methode?

Die **Maximum-Likelihood-Methode** soll einen brauchbaren Schätzer für Parameter einer Verteilung finden, indem sie jenen Wert des Parameters wählt, der die Stichprobe als wahrscheinlichstes Resultat erscheinen lässt. Dazu der Parameter θ so gewählt, dass die **Likelihood-Funktion**

$$l(\theta; x_1, \ldots, x_n) = f_\theta(x_1) \cdot f_\theta(x_2) \cdot \ldots \cdot f_\theta(x_n)$$

ein Maximum annimmt (f ist die Dichte der Verteilung).

Was besagt der zentrale Grenzwertsatz?

Besitzt die Verteilung der Grundgesamtheit eine endliche Varianz, was meist der Fall ist, so ist die Verteilung der Mittelwerte für genügend große Stichproben annähernd normal.

Nicht behandelte Themen

Was ist der Variationskoeffizient?

σ / μ

Was gilt für das Konfidenzintervall der Varianz?

Eine Riesenformel.