STS Summary (12 Pages)

06. Oktober, 2023; rev. 19. Januar 2024

Linda Riesen (rieselin)

1 Vorlesung 01/02 (Skript 1)

Stichprobe, Stichprobengröße, Grundgesamtheit:

- Stichprobe: Teilmenge einer Population zur Untersuchung.
- Stichprobengröße: Anzahl der Elemente in einer Stichprobe.
- Grundgesamtheit: Gesamtheit aller möglichen Elemente in einer Studie.

Merkmal, Merkmalstyp:

- Merkmal: Eine beobachtbare Eigenschaft eines Elements.
- Merkmalstypen: Kategoriell, metrisch, nominal, ordinal, diskret, stetig.

Merkmalsarten:

- Kategoriell: Daten in Kategorien, z.B., Farben (Rot, Blau).
- Metrisch: Messbare Werte, z.B., Gewicht in Kilogramm.
- Nominal: Kategorien ohne Reihenfolge, z.B., Geschlecht.
- Ordinal: Kategorien mit Reihenfolge, z.B., Bildungsniveau.
- Diskret: Ganzzahlige Werte, z.B., Anzahl der Kinder.
- Stetig: Unendlich viele mögliche Werte, z.B., Größe in Zentimetern.

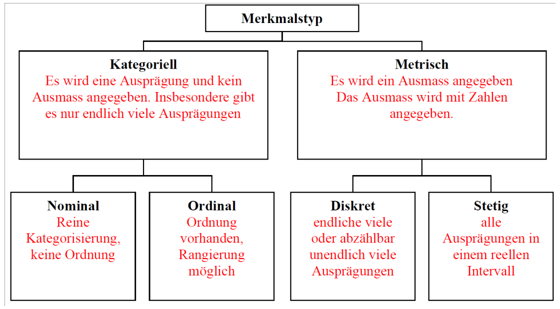


Abbildung 1: Übersicht Diagramm Merkmalsarten

Absolute und relative Häufigkeiten:

- Absolute Häufigkeit: Anzahl des Auftretens eines Werts.
- Relative Häufigkeit: Anteil der absoluten Häufigkeit an der Stichprobengröße.

Klassierte und nichtklassierte Daten:

- Klassierte Daten: In Intervalle (Klassen) gruppierte Werte, z.B., Altersgruppen.
- Nichtklassierte Daten: Einzelne, nicht gruppierte Werte, z.B., individuelle Altersangaben.

Faustregeln für die Klassierung von Daten:

- Mindestens 5-10 Klassen für klassierte Daten.
- Klassenbreite sollte ähnlich sein.
- Klasse ohne Daten sollte vermieden werden.

Dichtefunktion = Häufigkeitsfunktion (PDF/PMF): diskrete Merkmale: PMF (probability mass function) stetige Merkmale: PDF (probability density function) $h = \frac{h_i}{d_i}$ und $f = \frac{f_i}{d_i}$ [mit h = Höhe des Rechtecks = absolute Häufigkeits-Dichte, f = relative Häufigketsdichte, d = Klassenbreite]

- PDF für stetige Daten, PMF für diskrete Daten.

- PDF/PMF beschreibt die Wahrscheinlichkeitsverteilung.
- Summe/Integral über PDF/PMF ist 1.
- PDF gibt die Wahrscheinlichkeit eines bestimmten Werts an.

- Bei nichtklassierten/klassierten Daten ergibt sich die relative Häufigkeit eines Wertes grafisch aus der Höhe des Stabes/der Säule (PMF) / Histogramm (= Continuous Säulendiagram) (PDF) im Diagramm der PMF/PDF
- Berechnung und Eigenschaften der CDF für Stichproben:
  - CDF: $F(x) = H(x)/n$
  - $F(x) = \sum_{r \leq x} f(r)$, mit der relativen Häufigkeitsfunktion (PMF) = f(r)/n = (Anzahl aller Stichprobenwerte $x_i = r$)/n
  - Für alle reellen Zahlen x gilt $0 \leq F(x) \leq 1$.
  - F(x) ist monoton wachsend
  - Der Graph von F(x) ist eine rechtsseitig stetige Treppenfunktion (nicht klassiert) / Polygonzug = Linien mit Punktengraph (klassiert).
  - Es gibt eine reelle Zahl x mit F(x)= 0.
  - Es gibt eine reelle Zahl y mit F(y) = 1.
  - Der Anteil aller Stichprobenwerte xi im Bereich $a < xi \leq b$ berechnet sich als F(b)-F(a)
- Die Lagemasse (Mittelwert) gibt an, wo die Daten zentral tendieren.
- Die Streuungsmasse (Varianz oder Standardabweichung) zeigt, wie stark die Daten um den Mittelwert herum streuen.
- Berechnung von p-Quantilen (= p*100tes Perzentil) :
  1. Daten sortieren (aufsteigend)
  2. Falls n*p eine ganze Zahl: $R = 1/2(x_{n*q} + x_{n*q+1})$, falls nicht ganze Zahl wird aufgerundet
  3. Falls zwei Werte den gleichen Rang haben, wird zufällig einer zugeordnet

- Der **Median** ist der mittlerste Wert in einer geordneten Datenliste. 0.5 Quantil
- Perzentile sind Werte, unter denen ein bestimmter Prozentsatz der Daten liegt.
- Der Interquartilsabstand (IQR) ist die Streuung der mittleren 50% der Daten und wird als Differenz zwischen Q3 und Q1 berechnet.
- Die obere/untere Antenne erstreckt sich bis zum maximalen/minimalen Datenwert innerhalb von 1,5-facher Interquartilsabstand von der Box.
- Im Boxplot werden die Kenngrössen Median (2. Quartil) Q2, erstes Quartil Q1, drittes Quartil Q3, das Minimum und das Maximum der Stichprobe graphisch dargestellt:
- Stichprobenmittelwert (Sample Mean):

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

- Stichprobenvarianz (Sample Variance):

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

- Stichprobenstandardabweichung (Sample Standard Deviation):

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

- Korrigierte Stichprobenvarianz (Bessel-Korrektur):

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

- Korrigierte Stichprobenstandardabweichung:

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

- Modus gibt häufigsten Stichprobenwert an: $Modus(x1, ...xn) = xmod$

2 Vorlesung 03 + 04

2.1 Multivariante / Bivariante Datenmenge (2 vs mehr Datensets)

Zusammenhänge zw. Verschiedenen Datenträgern (bsp Grösse/Gewicht)

2.1.1 Mögliche Darstellungen

- Kontingenztabellen: (für beides eine Achse in Tabelle, anz.mit diesen beiden Merkmalen jeweils reinsortieren)
- Mosaikplot: Felder der Kontingenztabelle in Block-Diagramm
- Boxplots, Histogramm, Stripchart: Kennwerte ausrechnen, nebeneinander in Diagram aufreihen
- Streudiagramm (Scatterplot): Punkte verteilen (geht schlecht b. grossen Datenmengen) dabei können Form, Richtung und Stärke erkannt werden

2.2 Korrelation

Zusammenhang **linear** zw 2 metrischen Merkmalen (v.a. bei grösseren Stichproben) => Korrelationskoeffizienten

2.2.1 Pearson-Korrelationskoeffizient

Empirische Standardabweichung [mit $\bar{x}$ = arithm. Mittel]:

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

Kovarianz

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

Korrelationskoeffizient (Pearson)

$$r_{xy} = \frac{s_{xy}}{s_x * s_y}$$

Interpretation Korrelationskoeffizient nahe 1: stark positiver Zsmenhang, xWerte nehmen mit yWerten zu
Korrelationskoeffizient nahe -1: stark negativer Zsmenhang, xWerte nehmen mit yWerten ab Korrelationskoeffizient nahe 0: kein **linearer** Zsmenhang zwischen den beiden Merkmalen

2.2.2 Spearman-Rangkorrelationskoeffizient

Streuungsmass um eine beliebige monotone Kurve (muss nicht linear sein): Nichtparametrisches Korrelationsmass
mit [r = Rang der sortierten Eigenschaft: da isch da 1,2,3 etc. wo du jdm wert en rang zueordnisch (nachdems sortiert hesch eif de reihe na ufezelle u wenn 2 gliichigi hesch denn bi beidne de durchschnitt aneschriebe als rang)]

$$r_{sp} = \frac{s_{r_{xy}}}{s_{r_x} * s_{r_y}}$$

Interpretation gleich wie bei Pearson, aber Zsmhang monoton muss nicht linear sein

2.3 Probleme mit Koeffizienten

Falsch Positiv/Negativ: kann noch v. drittem Merkmal abhängen, Daten müssen Visualisiert werden da s durch Ausreisser etc. beeinflusst wird

3 Vorlesung 05

3.1 Definitionen

- Startwert 0! = 1
- n! = n*(n-1)!
- Binominalkoeffizient $\binom{n}{k} = \frac{n!}{k!(n-k)!}$

3.2 4 Grundlegende Abzählverfahren d Kombinatorik

3.2.1 Zahlenschlossproblem

6 Zahlenkränze mit jw 10 Ziffern, Zahlen können mehrfach vorkommen, Reihenfolge ist wichtig (entspricht auch Bitproblem (Wieviele Zahlenkombinationen können mit 64 Bits (= 2 Zahlen) dargestellt werden))

$$P = 10^6$$

3.2.2 Schwimmwettkampf

Platzierungen von 10 Schwimmern auf 3 Plätzen, Reihenfolge ist wichtig, Plätze können nicht mehrfach vergeben werden

$$P = 10 * 9 * 8 = \frac{10!}{(10-3)!}$$

3.2.3 Lotto

Ziehung von 6 aus 49, Reihenfolge spielt keine Rolle, Zahlen können nicht mehrfach vorkommen
Möglichkeiten $= 49 * 48 * 47 * 46 * 45 * 44 = \frac{49!}{43!}$
bei $6!$ Verschiedenen Arten diese Anzuordnen

$$P = \binom{49}{6} = \frac{49!}{(49-6)! * 6!}$$

3.2.4 Zahnarztproblem

5 Schüsseln mit Verschiedenen Objekten, 3 Objekte dürfen ausgesucht werden: Lösung durch Ersatzproblem: 3 Objekte = XXX und 5 Schalen IIII als Abtrennungen dazwischen => Verteilung von XXXIIII = Lottoproblem

$$P = \binom{7}{3} = \binom{7}{4} = \frac{7!}{3! * 4!}$$

3.2.5 Fussballmanschaft wählen mit xx Frauen und xx Männern

Multiplikation von jeweiligen Wahrscheinlichkeiten

3.2.6 Teilmengen

Wieviele Teilmengen kann man bilden = Potenzmenge = Zahlenschlossproblem

3.3 Systematik: Einteilung in 4 Klassen

Variation: mit Reihenfolge
Kombination: ohne Reihenfolge

- **Variation** von k aus n Objekten **mit** Wiederholung (Bsp: Zahlenschloss): $P = n^k$
- **Variation** von k aus n Objekten **ohne** Wiederholung (Bsp: Schwimmwettkampf): $P = \frac{n!}{(n-k)!}$
- **Kombination** von k aus n Objekten **ohne** Wiederholung (Bsp: Lotto): $P = \binom{n}{k}$
- **Kombination** von k aus n Objekten **mit** Wiederholung (Bsp: Zahnarztproblem): $P = \binom{n+k-1}{k}$

3.4 Binominalkoeffizient Eigenschaften

- $\binom{n}{0} = 1$
- $\binom{n}{k} = \binom{n}{n-k}$
- Reihe von $\binom{n}{0}$ bis $\binom{n}{n}$ ist symmetrisch
- Rekursionsformel: $\binom{n+1}{k+1} = \binom{n}{k} + \binom{n}{k+1}$

Pascals'sches Dreieck kann zur Berechnung d Binominalkoeffizienten verwendet werden

4 Vorlesung 06

4.1 Begriffe

- Das Mengenkalkül beschäftigt sich mit der Theorie von Mengen, die als Sammlungen von Objekten betrachtet werden. In der Mengenlehre sind grundlegende Konzepte wie Vereinigung, Schnittmenge, Differenz, Komplement und Kartesisches Produkt wichtige Bestandteile
- Das Zahlenkalkül bezieht sich auf die mathematischen Operationen und Eigenschaften von Zahlen, insbesondere ganzen Zahlen, rationalen Zahlen, reellen Zahlen und komplexen Zahlen.
- Ergebnisraum: Menge aller möglichen, sich gegenseitig ausschliessenden Ergebnisse eines Zufallsexperiments (Ergebnisraum ist diskret falls endlich/abzählbar unendlich)
- Zähldichte rho: Funktion $rho => [0, 1]$ mit Summe = 1, gibt Wahrscheinlichkeit für möglichen Ergebnisse des Zufallsexperiments an
- Ereignis A: Zusammenfassung v. gewissen Ergebnissen zu einer Menge (Teilmenge von Ω mit [{} = unmögliches Ergebnis, Ω = sicheres Ergebnis]
- Ereignisraum 2^Ω: Menge aller möglichen Ereignisse (= Menge aller Teilmengen von Ω)
- Wahrschenlichkeitsmass P: Funktion $P : 2^\Omega => [0, 1]$ mit [P(M) = Summe von $\rho(\omega)$] Wahrscheinlichkeit d Ereignis welches durch Menge M beschrieben eintritt
- Laplace Raum: Ein diskreter Ergebnisraum Ω in Kombination mit einem Wahrscheinlichkeitsmass P wird als diskreter Wahrscheinlichkeitsraum (Ω, P) bezeichnet. Wenn die Zähldichte ρ für jedes mögliche Ergebnis denselben Wert hat, wenn also alle Ergebnisse gleich wahrscheinlich sind, dann nennt man (Ω, P) einen Laplace-Raum.
- Zufallsvariable X: eine Variable, die Werte aufgrund von Zufallsexperimenten annimmt, und diese Werte können verschiedene Wahrscheinlichkeiten haben. Mit PMF(X) und CDF(X) lassen sich Wahrscheinlichkeiten von Aussagen systematisch berechnen

Diskreter Wahrscheinlichkeitsraum
Jeder diskrete Wahrscheinlichkeitsraum (Ω, P) hat die folgenden Eigenschaften:

(A1) (Unmögliches Ereignis) $P(\{\}) = 0$
(A2) (Sicheres Ereignis) $P(\Omega) = 1$
(A3) (Komplementäres Ereignis) $P(\Omega \backslash A) = 1 - P(A)$
(A4) (Vereinigung) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
(A5) (Sigma-Additivität) $P(A_1 \cup A_2 \cup A_3 \cup \ldots) = P(A_1) + P(A_2) + P(A_3) + \ldots$ falls die Ereignisse $A_1, A_2, A_3, \ldots$ paarweise disjunkt sind.

Ist (Ω, P) ein Laplace-Raum, so gilt: $P(M) = \frac{|M|}{|\Omega|}$.

Abbildung 2: Diskreter Wahrscheinlichkeitsraum Rechenregeln

Satz (Zufallsvariable, *PMF*, *CDF*)
Gegeben sind ein diskreter Wahrscheinlichkeitsraum (Ω, P) sowie eine Zufallsvariable $X : \Omega \to \mathbb{R}$. Dann haben *PMF* und *CDF* von X folgende Eigenschaften:
(1) $\sum_{x=-\infty}^{\infty} f(x) = 1$ und $F(z) = \sum_{x=-\infty}^{z} f(x)$
(2) $\lim_{x \to \infty} F(x) = 1$ und $\lim_{x \to -\infty} F(x) = 0$
(3) Monotonie: Aus $x \leq y$ folgt $F(x) \leq F(y)$
(4) $f(x) = F(x) - \lim_{y \to x^-} F(y)$
(5) $P(a < X \leq b) = F(b) - F(a)$ und $P(a \leq X \leq b) = F(b) - \lim_{x \to a^-} F(x)$
(6) $P(X > b) = 1 - F(b)$ und $P(X \geq b) = 1 - \lim_{x \to b^-} F(x)$

Abbildung 3: Berechnung CDF(Kumulative Verteilsungsfunktion) PDF (Dichtefunktion)

5 Vorlesung 07: Kenngrössen

Bsp: Erwartungswert, Varianz, Standardabweichung: Werden eingeteilt in Lagemasse und Streumasse
Statistik: Merkmale innehalb Stichprobe, Stochastik: Verteilung Zufallsvariablen

5.1 Definitionen

- Erwartungswert: $\sum_x P(X = x) * x$ (Lagemass d Verteilung v X)
- Varianz: $\sum_x P(X = x) * (x - E(X))^2$ (Streumass d Verteilung v X)

- Standardabweichung: $\sqrt{V(X)}$ (Streumass d Verteilung v X)
- Tschebyscheff'sche Ungleichung [k > 0]:

$$P(|X - E(X)| < k) \geq 1 - \frac{S(X)^2}{k^2}$$

Satz (Kenngrössen)
Gegeben sind ein diskreter Wahrscheinlichkeitsraum (Ω, P) sowie eine Zufallsvariable $X: \Omega \to \mathbb{R}$. Dann haben die Kenngrössen von X folgende Eigenschaften:

(1) **Linearität des Erwartungswertes**:
$E(X+Y) = E(X) + E(Y)$ und $E(\alpha X) = \alpha E(X)$, mit $\alpha \in \mathbb{R}$.

(2) **Verschiebungssatz für die Varianz**
$V(X) = E(X^2) - (E(X))^2 = (\sum_{x \in \mathbb{R}} P(X = x) \cdot x^2) - E(X)^2$

(3) $V(\alpha X + \beta) = \alpha^2 \cdot V(X)$ mit $\alpha, \beta \in \mathbb{R}$.

Abbildung 4: Kenngrössen Zufallsvariablen

6 08: Bedingte Wahrscheinlichkeiten/Stochastische Unabhängigkeit

6.1 Bedingte Wahrscheinlichkeiten

dh. 2 Wahrscheinlichkeiten die miteinander verknüpft sind

- $P(R|F) = \frac{P(F \cap R)}{P(F)}$ R unter Voraussetzung dass F (Bedingte Wahrscheinlichkeit)
- $P(A \cap Z) = P(A) * P(Z|A) = P(Z) * P(A|Z)$ Pfadwahrscheinlichkeit (Wkeitsbaum)
- $P(B) = P(A) * P(B|A) + P(\bar{A}) * P(B|\bar{A})$ Satz d Totalen Wahrscheinlichkeit
- $P(A|B) = \frac{P(A) * P(B|A)}{P(B)}$ Satz von Bayes

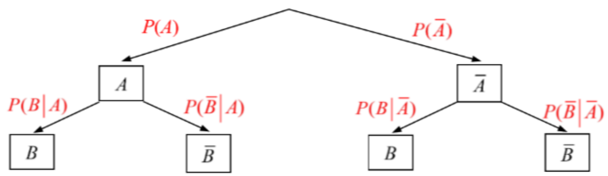


Abbildung 5: Wahrscheinlichkeitsbaum

6.2 Stochastische Unabhängigkeit

wenn Eintreten eines Ereignis keinen Einfluss auf das Eintreten des anderen hat

Definition

- es gilt: $P(B) \neq 0 => P(A|B) = P(A)$ und $P(A) \neq 0 => P(B|A) = P(B)$
- es gilt: $P(A \cap B) = P(A) * P(B)$
- f Zufallsvariabeln: $P(X = x \wedge Y = y) = P(X = x) * P(Y = y)$
- Die Zufallsvariablen X und Y sind dann stochastisch unabhängig, wenn jede Wahrscheinlichkeit in den inneren Feldern der Tabelle gleich dem Produkt der entsprechenden Randwahrscheinlichkeiten ist.

Satz (Stochastische Unabhängigkeit – Ereignisse)
Wir betrachten einen diskreten Wahrscheinlichkeitsraum (Ω, P) und zwei Ereignisse A und $B \subseteq \Omega$.

Folgende Eigenschaften sind äquivalent:

(i) A und B sind stochastisch unabhängig.
(ii) A und $\Omega \backslash B$ sind stochastisch unabhängig.
(iii) $\Omega \backslash A$ und $\Omega \backslash B$ sind stochastisch unabhängig.

Abbildung 6: Satz Stochastische Unabhängigkeit, Gleichbedeutend

Satz (Stochastische Unabhängigkeit – Zufallsvariablen)
Wir betrachten einen diskreten Wahrscheinlichkeitsraum (Ω, P) sowie zwei stochastisch unabhängige Zufallsvariablen $X, Y: \Omega \to \mathbb{R}$.

- Falls die Erwartungswerte $E(X)$ und $E(Y)$ existieren, so gilt:
$$E(X \cdot Y) = E(X) \cdot E(Y)$$
- Falls die Varianzen $V(X)$ und $V(Y)$ existieren, so gilt:
$$V(X+Y) = V(X) + V(Y)$$

Abbildung 7: Satz Stochastische Unabhängigkeit, Zufallsvariablen

7 09: Spezielle Verteilungen

7.1 Stetige oder Diskrete Verteilung

- diskret: es gibt Lücken, stetig: es gibt keine Lücken zw den Werten
- Definition v. stetig über PMF geht nicht da dann für P(X=x) = 0, Definition über CDF
- PDF (stetig): $f(x) \geq 0$ [für alle x in R] und $F(x) = P(X \leq x) = \int_{-\infty}^{\infty} f(u)du = 1$
- $\mu = E(X) = \int_{-\infty}^{\infty} f(x) * xdx$
- $\sigma^2 = V(X) = \int_{-\infty}^{\infty} f(x) * (x - E(X))^2 dx = x^2 * f(x)$
- Stetig mit Grenzen:
  - $E(X) = \frac{a+b}{2}$
  - $V(X) = \frac{(b-a)^2}{12}$

	diskrete Zufallsvariablen	**stetige Zufallsvariablen**
Dichtefunktion / PMF bzw. PDF	$f(x) = P(X = x)$	$f(x) = F'(x) \neq P(X = x)$!!!
Kumulative Verteilungsfunktion/ CDF	$F(x) = P(X \leq x) = \sum_{x \leq X} f(x)$	$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)du$
Wahrscheinlichkeiten	$P(a \leq X \leq b) = \sum_{a \leq x \leq b} f(x)$	$P(a \leq X \leq b) = \int_{a}^{b} f(x)dx$
Graphische Darstellung von f	Stabdiagramm	Graph
Erwartungswert	$E(X) = \sum_{x \in \mathbb{R}} f(x) \cdot x$	$E(X) = \int_{-\infty}^{\infty} f(x) \cdot x\, dx$
Varianz	$V(X) = \sum_{x \in \mathbb{R}} f(x) \cdot (x - E(X))^2$	$V(X) = \int_{-\infty}^{\infty} f(x) \cdot (x - E(X))^2\, dx$

Für diskrete und stetige Zufallsvariablen X und Y gelten die folgenden Regeln:

(1) **Linearität des Erwartungswertes**:
$E(X+Y) = E(X) + E(Y)$ und $E(\alpha X) = \alpha E(X)$ mit $\alpha \in \mathbb{R}$.

(2) **Verschiebungssatz für die Varianz**: $V(X) = E(X^2) - (E(X))^2$
mit $E(X^2) = \int_{-\infty}^{\infty} f(x) \cdot x^2\, dx$

(3) $V(\alpha X + \beta) = \alpha^2 \cdot V(X)$ mit $\alpha, \beta \in \mathbb{R}$.

(4) Sind X und Y unkorreliert ($COV(X,Y) = 0$), so gilt: $V(X+Y) = V(X) + V(Y)$.

Abbildung 8: Stetig vs Diskret Gegenüberstellung

7.2 Diskrete Verteilungen

	Hypergeometrische Vert.	Binomialverteilung	Poissonverteilung
Dichtefunktion für $x = 0,1,2,\ldots$ sonst 0	$P(X = x) = \frac{\binom{M}{x} \cdot \binom{N-M}{n-x}}{\binom{N}{n}}$	$P(X = x) = \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x}$	$P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda}$
Erwartungswert	$n \cdot \frac{M}{N}$	$n \cdot p$	λ
Varianz	$n \cdot \frac{M}{N} \cdot (1 - \frac{M}{N}) \cdot \frac{N-n}{N-1}$	$n \cdot p \cdot (1-p)$	λ
Standardabweichung	$\sqrt{n \cdot \frac{M}{N} \cdot (1 - \frac{M}{N}) \cdot \frac{N-n}{N-1}}$	$\sqrt{n \cdot p \cdot (1-p)}$	$\sqrt{\lambda}$
Parameter	n, N, M	n, p	λ

Abbildung 9: Diskrete Verteilungen Übersicht

7.2.1 Hypergeometrische Verteilung

Urne mit [N Objekten, davon M Objekte einer bestimmten Sorte (Merkmalsträger), Stichprobe n (**ohne Zurücklegen**)]
Geschrieben: $X\tilde{H}(N, M, n)$

$$P(X = x) = \frac{\binom{M}{x} * \binom{M-N}{n-x}}{\binom{N}{n}}$$

- $\mu = E(X) = n * \frac{M}{N}$
- $\sigma^2 = V(X) = n * \frac{M}{N} * (1 - \frac{M}{N}) \frac{N-n}{N-1}$

7.2.2 Bernoulliverteilung

Bernoulli Experimente sind Zufallsexperminente mit nur 2 möglichen Ergebnissen (0, 1)
$P(X=1)=p$ und $P(X=0)=1-p=q$

- $E(X)=p$
- $E(X^2)=p$
- $V(X)=pq$

7.2.3 Binominalverteilung

Für Binominalverteilung wird dasselbe Bernoulli-Expermient n-Mal hintereinander durchgeführt , Wahrscheinlichkeiten sind stochastisch unabhängig (Kugel wird zurückgelegt (oder N ist sehr gross, sodass Vernachlässigt werden kann: $n \lesssim \frac{N}{20}$)): $p=\frac{M}{N}$ und Schreibweise: $X\tilde{B}(n,p)$

$$P(X=x)\binom{n}{x} * p^x * qn-x$$

- Bsp: $P(X \leq 2) = P(X=0) + P(X=1) + P(X=2)$
- $E(X) = n * p$
- $V(X) = n * p * q$

Näherung durch Poisson Für grosses n und kleines p ($n \gtrsim 50$ und $p \lesssim 0.1$): $B(n,p) \approx Poi(n * p)$

7.2.4 Poissonverteilung

Wahrscheinlichkeit Anzahl gleichartiger Ergebnisse in **gegebenem Bereich**, Schreibweise: $X\tilde{P}oi(\lambda)$ mit [$\lambda > 0$ und λ = Durchschnitt im Bereich]

$$P(X=x) = \frac{\lambda^x}{x!} * e^{-\lambda}$$

- $\mu = E(X) = \lambda$
- $\sigma^2 = V(X) = \lambda$

8 10: Stetige Verteilungen

8.1 Gauss-Verteilung/Normalverteilung

$E(X)=\mu$, $V(X)=\sigma^2$

$$\text{Normalverteilung Dichtefunktion: } \varphi_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi} * \sigma} * e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$

$$\text{Standardnormalverteilung mit } [\mu=0, \sigma=1]\text{: } \varphi(x) = \frac{1}{\sqrt{2\pi}} * e^{-\frac{1}{2}x^2}$$

- symmetrisch bezüglich der Geraden: $x=\mu$
- Wendepunkte an Stellen: $\mu-\sigma$ und $\mu+\sigma$
- Ist Normiert: dh. Fläche darunter = 1
- Änderung von μ: Verschiebung in x-Richtung
- Änderung von σ: je grösser σ desto breiter und niedriger die Glockenkurve

8.1.1 Verteilungsfunktion der Normalverteilung

$$\phi_{\mu,\sigma}(x) = P(X \leq x) = \int_{-\infty}^{x} \varphi_{\mu,\sigma}(t)dt = \frac{1}{\sqrt{2\pi} * \sigma} * \int_{-\infty}^{x} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dt$$

8.1.2 Werte Eigenschaften = Stetigkeitskorrektur

- Für **standardnormalverteilte** Zufallsvariable X Tabelle: Rechts: erste Zahl, Oben: 2. Nachkommastelle
- Falls mehr als 2 Nachkommastellen muss linear interpoliert werden:

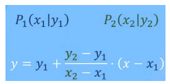


Abbildung 10: Lineare Interpolation, P1 = abrunden, P2 = aufrunden

- Falls beliebige Normalverteilung muss **standartisiert** werden: (X = Gewünschte Grenze), für U dann ablesen

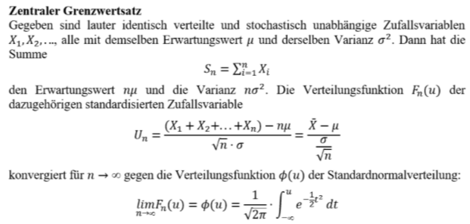
$$U = \frac{X - \mu}{\sigma}$$

- Für symmetrisch zu μ gelegenes Intervall $[\mu - e; \mu + e]$:

$$2 * \phi(\frac{e}{\sigma}) - 1$$

8.1.3 Zentraler Grenzsatz

- Folge von $X_1, X_2, ..., X_n, ...$ Zufallsvariablen:
- Hat n-te Summe: $S_n = X_1 + X_2 + ... + X_n$
- Arithmetisches Mittel: $\frac{1}{n}[X_1 + X_2 + ... + X_n]$
- $E(S_n = E(X_1) + ... + E(X_n)$
- $E(\tilde{X}_n) = \frac{1}{n}[E(X_1) + ... + E(X_n)]$
- Wenn **stochastisch Unabhängig**:
  - $V(S_n) = V(X_1) + ... + V(X_n)$
  - $V(\tilde{X}_n) = \frac{1}{n^2}[V(X_1) + ... + V(X_n)]$
- Wenn alle Zufallsvariabeln denselben Erwartungswert (μ) und dieselbe Varianz (σ^2) haben:
  - $E(S_n) = n * \mu$
  - $V(S_n) = n * \sigma^2$
  - $E(\tilde{X}_n) = \mu$
  - $V(\tilde{X}_n) = \frac{\sigma^2}{n}$
- Sind die Zufallsvariabeln alle identisch $N(\mu, \sigma)$ verteilt
  - $S_n = N(n * \mu, \sqrt{n} * \sigma)$
  - Arithmetisches Mittel: $\tilde{X}_n = N(\mu, \frac{\sigma}{\sqrt{n}}$



Zentraler Grenzwertsatz
Gegeben sind lauter identisch verteilte und stochastisch unabhängige Zufallsvariablen $X_1, X_2, ...$, alle mit demselben Erwartungswert μ und derselben Varianz σ^2. Dann hat die Summe

$$S_n = \sum_{i=1}^{n} X_i$$

den Erwartungswert $n\mu$ und die Varianz $n\sigma^2$. Die Verteilungsfunktion $F_n(u)$ der dazugehörigen standardisierten Zufallsvariable

$$U_n = \frac{(X_1 + X_2 + \ldots + X_n) - n\mu}{\sqrt{n} \cdot \sigma} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

konvergiert für $n \to \infty$ gegen die Verteilungsfunktion $\phi(u)$ der Standardnormalverteilung:

$$\lim_{n \to \infty} F_n(u) = \phi(u) = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{u} e^{-\frac{1}{2}t^2}\, dt$$

Abbildung 11: Zentraler Grenzwertsatz

8.1.4 Approximation Binominalverteilung

Zentraler Grenzwertsatz kann angewendet werden: kann Verwendet werden wenn $npq > 9$
mit $[\mu = np, \sigma^2 = npq]$

$$P(a \leq X \leq b) = \sum_{x=a}^{b} P(X = x) \approx \phi_{\mu,\sigma}(b + \frac{1}{2}) - \phi_{\mu,\sigma}(\alpha - \frac{1}{2})$$

8.1.5 Approximation Poissonverteilung

Zentraler Grenzwertsatz kann angewendet werden: kann Verwendet werden wenn $\lambda > 9$
mit $[\mu = \lambda, \sigma^2 = \lambda]$

$$P(a \leq X \leq b) = \sum_{x=a}^{b} P(X = x) \approx \phi_{\mu,\sigma}(b + \frac{1}{2}) - \phi_{\mu,\sigma}(\alpha - \frac{1}{2})$$

9 11: Die Methode der kleinsten Quadrate

Optimierungsmethode zur Modellierung mathematischer Zusammenhänge in grossen Datenmengen

9.1 Lineare Regression

Bei der linearen Regression wird ein linearer Zusammenhang zwischen den Daten vermutet und versucht, eine optimale Gerade in die Datenmenge einzupassen.

Die öptimale"Gerade (Ausgleichs- / Regressionsgerade), sei diejenige Gerade für die die Summe der quadrieten Residuen ($\sum_{i=1}^{n} \epsilon_i^2$) am kleinsten ist. = Minimum der Summe (Methode der kleinsten Quadrate (KQM)): Erklärte Varianz

$$\sum_{i=1}^{n}(y_i - g(x_i))^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

mit [y_i = beobachtete y - Werte, $\hat{y}_i$ = prognostizierte / erklärte y-Werte]

9.1.1 Vorgehensweise

Minimum der Residualvarianz (Residuenvarianz) => Gleichung der Regressionsgeraden

$$\tilde{s}_\epsilon^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - mx_i - d)^2$$

Daraus folgt: $0 = \bar{y} - m\bar{x} - d$: Greade verläuft durch den Schwerpunkt der Punktmenge: $d = \bar{y} - m\bar{x}$
und es folgt: $m = \frac{\bar{xy} - \bar{x}\bar{y}}{\bar{x^2} - \bar{x}^2} = \frac{\tilde{s}_{xy}}{\tilde{s}_x^2}$ (Kovarianz / Varianz d. x Werte)

Zusammenfassung Regressionsgerade:
Die *Regressionsgerade* $g(x) = mx + d$ mit den Parametern m und d ist die Gerade, für die die *Residualvarianz* $\tilde{s}_\epsilon{}^2$ minimal ist.

Die Regressionsgerade hat die Steigung

$$m = \frac{\tilde{s}_{xy}}{\tilde{s}_x{}^2}$$

und den y-Achsenabschnitt

$$d = \bar{y} - m\bar{x}$$

Für die zugehörige (minimale) Residualvarianz gilt:

$$\tilde{s}_\epsilon{}^2 = \tilde{s}_y{}^2 - \frac{\tilde{s}_{xy}{}^2}{\tilde{s}_x{}^2}$$

mit:
Varianz der x_i-Werte

$$\tilde{s}_x{}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \left(\frac{1}{n}\sum_{i=1}^{n} x_i{}^2\right) - \bar{x}^2$$

Varianz der y_i-Werte

$$\tilde{s}_y{}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 = \left(\frac{1}{n}\sum_{i=1}^{n} y_i{}^2\right) - \bar{y}^2$$

Kovarianz

$$\tilde{s}_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \left(\frac{1}{n}\sum_{i=1}^{n} x_i y_i\right) - \bar{x}\cdot\bar{y}$$

arithmetische Mittelwerte

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \text{ und } \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

Hinweis: Die Berechnung kann auch mit den korrigierten Werten $s_x{}^2$, $s_y{}^2$, s_{xy} erfolgen

Abbildung 12: Zusammenfassung Regressionsgerade

9.1.2 Bestimmtheitsmass

Überprüfung der Güte des Modells

Zusammenfassung Bestimmtheitsmass:
Die Totale Varianz setzt sich zusammen aus der Residualvarianz und der Varianz der prognostizierten Werte:

$$\tilde{s}_y{}^2 = \tilde{s}_\epsilon{}^2 + \tilde{s}_{\hat{y}}{}^2 \text{ bzw. } s_y^2 = s_\epsilon^2 + s_{\hat{y}}^2$$

Das *Bestimmtheitsmass* R^2 beurteilt die globale Anpassungsgüte einer Regression über den Anteil der prognostizierten (erklärten) Varianz $s_{\hat{y}}^2$ an der totalen Varianz s_y^2:

$$R^2 = \frac{\tilde{s}_{\hat{y}}{}^2}{\tilde{s}_y{}^2} \text{ bzw. } R^2 = \frac{s_{\hat{y}}^2}{s_y^2}$$

Das *Bestimmtheitsmass* R^2 stimmt überein mit dem Quadrat *des Korrelationskoeffizienten (nach Bravais-Pearson)*

$$R^2 = \frac{\tilde{s}_{xy}{}^2}{\tilde{s}_x{}^2\tilde{s}_y{}^2} = r_{xy}{}^2 \text{ bzw. } R^2 = \frac{s_{xy}{}^2}{s_x{}^2 s_y{}^2} = r_{xy}{}^2$$

Abbildung 13: Zusammenfassung Bestimmtheitsmass (Totale Varianz)

Bsp: $R^2 = 0.75$ = 75% der gesamten Varianz kann durch Regression erklärt werden, der Rest ist Zufallsstreuung

9.1.3 Residuenbetrachtung

Auch wenn Bestimmtheitsmass ≈ 1 oder sehr klein, kann linearer Trend nicht umbedingt/trotzdem gefolgert werden aus graphischer Darstellung (Residuenplot) / Betrachtung des Streudiagramms
Die Residuen werden dabei bezogen auf die prognostizierten y-Werte $\bar{y}$ dargestellt.
Auf der horizontalen Achse werden die prognostizierten y-Werte $\bar{y}$ und auf der vertikalen Achse die Residuen angetragen.
Die Residuen sollten unsystematisch (d.h. zufällig) und überall etwa gleich um die horizontale Achse streuen, betragsmässig kleine Residuen sollten häufiger sein als grosse.

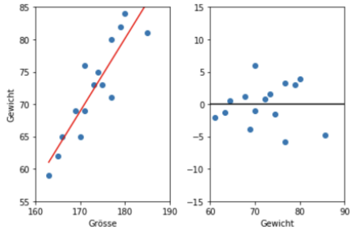


Abbildung 14: Residuen-Plot Beispiel

9.2 Nichtlineares Verhalten

Auch für nichtlineares Verhalten können die im Ansatz enthaltenen Modellparameter mithilfe der Gauss'schen Methode der kleinsten Quadrate bestimmt werden. Aber nicht immer sind die Schätzungen explizit darstellbar. Dies macht den Einsatz von numerischen Verfahren zur Bestimmung der Schätzer erforderlich. In manchen Fällen besteht jedoch die Möglichkeit, durch eine geschickte Transformation ein nichtlineares Regressionsmodell auf ein lineares Modell zurückzuführen.
Für verschiedene Funktionen ist eine Linearisierung möglich, wie z.B. für:

Ausgangsfunktion	Transformation
$y = q \cdot x^m$	$\log(y) = \log(q) + m \cdot \log(x)$
$y = q \cdot m^x$	$\log(y) = \log(q) + \log(m) \cdot x$
$y = q \cdot e^{m \cdot x}$	$\ln(y) = \ln(q) + m \cdot x$
$y = \frac{1}{q + m \cdot x}$	$V = q + m \cdot x; \quad V = \frac{1}{y}$
$y = q + m \cdot \ln(x)$	$y = q + m \cdot U; \quad U = \ln(x)$
$y = \frac{1}{q \cdot m^x}$	$\log\left(\frac{1}{y}\right) = \log(q) + \log(m) \cdot x$

Abbildung 15: Linearisierung / Transformation von nicht linearen Regression

9.3 Allgemeines Vorgehen

In manchen Fällen soll die Abhängigkeit nicht allein von einer, sondern von mehreren Variablen betrachtet werden. Hier ist das Vorgehen vergleichbar und wir können dies als ein Problem der linearen Algebra betrachten.

Im Folgenden ist das Vorgehen für einen linearen Zusammenhang beschrieben.
Allgemein kann man dabei die KQM in der Matrixdarstellung formulieren.
Gegeben sind n Gleichungen und wir suchen k Regressionsparameter: $p_1, p_2, \dots, p_k$.

Dafür ergibt sich das folgende lineare Gleichungssystem für $(k-1)$ Merkmale:

$$(*) \begin{cases} y_1 = \underbrace{p_1 x_{11} + p_2 x_{12} + \cdots + p_{k-1} x_{1(k-1)} + p_k}_{=\hat{y}_1} + \epsilon_1 \\ \dots \\ y_n = \underbrace{p_1 x_{n1} + p_2 x_{n2} + \cdots + p_{k-1} x_{n(k-1)} + p_k}_{=\hat{y}_n} + \epsilon_n \end{cases} \qquad \text{mit den Residuen } \epsilon_i$$

Mit den Vektoren der Parameter $p = \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_k \end{pmatrix}$, der Messwerte $y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$, der Residuen $\epsilon = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$

und der Matrix der Eingangswerte $X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1(k-1)} & 1 \\ x_{21} & x_{22} & \dots & x_{2(k-1)} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{n(k-1)} & 1 \end{pmatrix}$,

lautet das Gleichungssystem (*): $y = Xp + \epsilon.$

Die KQM verlangt dann, den Vektor p so zu bestimmen, dass die Summe der Fehlerquadrate $\sum_{i=1}^{n} \epsilon_i^2$ minimal ist.

Die Lösung ist $p = (X^T X)^{-1} X^T y$
falls $(X^T X)$ invertierbar.

Abbildung 16: Allgemeine Lösung Regression

Wegen der orthogonalen Zerlegung von y in die verktoriellen Komponenten X*p und ϵ gilt auch hier die Quadratsummezerlegung

$$|y|^2 = |X * p|^2 + |\epsilon|^2$$

10 12: Schliessende Statistik (Parameter- und Intervallschätzung

Aussagen über Grundgesamtheit aus Stichprobe erstellen. Dafür müssen alle Objekte der Grundgesamtheit die gleiche Chance haben, gewählt zu werden (stochastisch unabhängig).

10.1 Schätzfunktionen

Der Parameter θ (z.B. der Mittelwert oder die Standardabweichung) charakterisiert die Verteilung der Grundgesamtheit. Der Wert von θ soll geschätzt werden.

- θ ist selber eine Zufallsvariable (wahrer Wert)
- Schätzfunktionen für μ = arithmetisches Mittel / Median
- Schätzfunktionen für σ = empirische (korrigierte) Standardabweichung
- Schätzfunktionen für p = relative Häufigkeit

10.1.1 Optimale Schätzungsfunktion Kriterien

Definition
Eine Schätzfunktion Θ eines Parameters θ heisst *erwartungstreu*, wenn gilt:

$$E(\Theta) = \theta$$

Gegeben sind zwei erwartungstreue Schätzfunktionen Θ_1 und Θ_2 desselben Parameters θ. Man nennt Θ_1 *effizienter* als Θ_2, falls gilt:

$$V(\Theta_1) < V(\Theta_2)$$

Eine Schätzfunktion Θ heisst *konsistent*, wenn gilt:

$$E(\Theta) \to \theta \text{ und } V(\Theta) \to 0 \text{ für } n \to \infty$$

Abbildung 17: Erwartungstreu, Erwartungsgetreu, Effizient, Konsistent

Schätzfunktionen können entweder getestet (durch Anschaulichkeit/Simulation) oder konstruiert (Methode d. k. Quadrate / Maximum Likelihood)

10.1.2 Maximum-Likelihood-Schätzung

Siehe Tabelle, Logarithmieren damit Ableitung einfacher, dann Ableitung = 0

10.2 Vertrauensintervalle

Gesucht wird Intervall zudem man Wahrscheinlichkeit angeben kann mit der es darin liegt.

$$P(|U| \leq \frac{e}{\sigma/\sqrt{n}}) = y$$

- gesucht c mit $c = \frac{e}{\sigma/\sqrt{n}}$
- y [ist obere gewünschte Prozent] $c = \frac{1+y}{2}$ (Ablesen aus Tabelle Normalverteilung)
- $e = x * \frac{\sigma}{\sqrt{n}}$

	(1) Verteilung der Grundgesamtheit	(2) Param.	(3) Schätzfunktionen	(4) zugehörige standardisierte Zufallsvariable	(5) Verteilung und benötigte Quantile	(6) Zufallsvariablen für Intervallgrenzen
1	Normalverteilung (Varianz σ^2 bekannt)	μ	$\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i$	$U = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$	**Standardnormalverteilung** (Tabelle 2) $c = u_p$ mit $p = \frac{1+\gamma}{2}$	$\Theta_u = \bar{X} - c \cdot \frac{\sigma}{\sqrt{n}}$ $\Theta_o = \bar{X} + c \cdot \frac{\sigma}{\sqrt{n}}$
2	Normalverteilung (Varianz σ^2 unbekannt und $n \leq 30$; sonst Fall 1 mit s als Schätzwert für σ)	μ	$\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i$ $S^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (X_i - \bar{X})^2$	$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$	***t*-Verteilung** (Tabelle 4) mit $f = n - 1$ $c = t_{(p;f)}$ mit $p = \frac{1+\gamma}{2}$	$\Theta_u = \bar{X} - c \cdot \frac{S}{\sqrt{n}}$ $\Theta_o = \bar{X} + c \cdot \frac{S}{\sqrt{n}}$
3	Normalverteilung	σ^2	$\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i$ $S^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (X_i - \bar{X})^2$	$Z = (n-1)\frac{S^2}{\sigma^2}$	**Chi-Quadrat-Verteilung** (Tabelle 3) mit $f = n - 1$ $c_1 = z_{(p_1;f)}$ mit $p_1 = \frac{1-\gamma}{2}$ $c_2 = z_{(p_2;f)}$ mit $p_2 = \frac{1+\gamma}{2}$	$\Theta_u = \frac{(n-1) \cdot S^2}{c_2}$ $\Theta_o = \frac{(n-1) \cdot S^2}{c_1}$
4	Bernoulli-Verteilung Anteilsschätzung (mit $n\hat{p}(1-\hat{p}) > 9$)	p	$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ X_i 0/1-wertig mit $P(X_i = 1) = p$	$U = \frac{\bar{X} - p}{\sqrt{p(1-p)/n}}$	**Standardnormalverteilung** (näherungsweise), Tabelle 2 $c = u_q$ mit $q = \frac{1+\gamma}{2}$	$\Theta_u = \bar{X} - c \cdot \sqrt{\frac{\bar{X} \cdot (1-\bar{X})}{n}}$ $\Theta_o = \bar{X} + c \cdot \sqrt{\frac{\bar{X} \cdot (1-\bar{X})}{n}}$
5	beliebig mit $n > 30$	μ, σ^2	wie im Fall 1 (gegebenenfalls mit s als Schätzwert für σ) bzw. im Fall 3			

Abbildung 18: Übersicht Vertrauensinteralle, Schätzfunktionen