

Einführung Begriffe

Kardinal/Metrisch

Ordinal

Nominal

Qualitativ

Quantitativ

- intervall: → Abstände interpretierbar, kein sinnvoller Nullpkt. (Temp.)
- verhältnis: → Abstände interpretierbar, sinnvoller Nullpkt. (Gewicht)
- Ordnung mögl., aber Abstände nicht interpretierbar (Noten)
- Ausprägungen sind Namen, keine Ordnung mögl. (Haarfarbe)
- endlich viele Ausprägungen (höchstens ordinal!)
- Ausprägungen geben Intensität wieder (metrisch!)

Wachstum

Modell	Formel	Absolut. Wachstum $x_t - x_{t-1}$	Wachstumsrate $\frac{x_t - x_{t-1}}{x_{t-1}}$	Steigung $\frac{\partial x_t}{\partial t}$
Konstant. Lineares Wachstum	$x_t = x_0 + a \cdot t$	a	$\frac{a}{x_0 + a(t-1)}$	a
Konstant. %-Wachstum	$x_t = x_0(1+i)^t$	$i \cdot x_0(1+i)^{t-1}$	i	$x_0 \cdot \ln(1+i) \cdot (1+i)^t$
Konstantes expon. W.	$x_t = x_0 \cdot e^{b \cdot t}$	$x_0 \cdot e^{b(t-1)} \cdot (e^b - 1)$	$e^b - 1$	$x_0 \cdot b \cdot e^{b \cdot t}$

Verteilungen

Absol. kumm. Häufigkeitsverteilung $H(x)$ =Anzahl d. Werte x_i mit $x_i \leq x$:

$$\rightarrow H(x) = h(a_1) + \dots + h(a_j) = \sum_{i: a_i \leq x} h_i$$

Relative kumm. Häufigkeitsverteilung $F(x) = \frac{H(x)}{n}$ =Anteil d. Werte x_i mit $x_i \leq x$:

$$\rightarrow F(x) = f(a_1) + \dots + f(a_j) = \sum_{i: a_i \leq x} f_i$$

Lagemasse

Arithm. Mittel ($\bar{x}$) → nicht robust; grösser als geom. Mittel; für additive Verknüpfungen

$$1) \bar{x} = \frac{1}{n}(x_1 + \dots + x_n)$$

$$2) \bar{x} = (q_1 f_1 + \dots + q_n f_n) \text{ mit } 0 \leq q \leq 1$$

$$3) \bar{x} = \frac{1}{n}(n_1 \bar{x}_1 + \dots + n_n \bar{x}_n)$$

Median → 50% Quantil; resistent; Lage: Datenmitte; für metrische, ordinalskal. Merkmale minimiert die Summe der absoluten Abweichungen

$$\rightarrow n \text{ gerade: } \frac{1}{2}(x_{\frac{n}{2}} + x_{\frac{n}{2}+1})$$

$$\rightarrow n \text{ ungerade: } x(\frac{n+1}{2})$$

Modus → Ausprägung mit grössten Häufigkeit
ist eindeutig, falls Häufig.verteilung ein eindeutiges Max. hat

Geom. Mittel

$$\rightarrow \bar{x}_{\text{geom}} = (x_1 \cdot \dots \cdot x_n)^{\frac{1}{n}} = \text{mittlerer Wachstumsfaktor}$$

Harmon. Mittel

$$\rightarrow x_{\text{har}} = \frac{1}{\frac{1}{n} \sum_{i=1}^n \frac{1}{x_i}}$$

Getrimmtes Mittel

→ wie $\bar{x}$, aber Ausreisser weglassen (→ =robust)

Lage-regeln

symmetrisch:

linkssteil:

rechtssteil:

$$\bar{x} \approx x_{\text{med}} \approx x_{\text{mod}}$$

$$\bar{x} > x_{\text{med}} > x_{\text{mod}}$$

$$\bar{x} < x_{\text{med}} < x_{\text{mod}}$$

Übersicht

Nom. Sk. al

Modus

Streuungsmaße

Boxplot

Standardabweichung

$$\rightarrow \tilde{s} = +\sqrt{\tilde{s}^2}$$

→ =Streuung um AM; Var. klein → Werte nahe am AM
→ skalenunabhängig; für metrische M. sinnvoll

Varianz

$$\rightarrow \tilde{s}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \rightarrow \text{Streuung um AM} = \text{mittlere quadr. Abweichung v. Mittelwert}$$

→ wenn klein → Werte nahe am AM; für metrische M. sinnvoll
→ verletzt Prinzip d. Skalenunabhängigkeit

Gesamtvarianz

$$\rightarrow \tilde{s}^2 = \sum_{j=1}^n \tilde{s}_j^2 \left(\frac{n_j}{n} \right) + \sum_{j=1}^n (\bar{x}_j - \bar{x})^2 \left(\frac{n_j}{n} \right) \rightarrow \text{innere} + \text{äussere Varianz}$$

	$\bar{x}_j$	$\tilde{s}_j^2$	n_j	
Schule	10.884	0.277	244	$\sum_{j=1}^n \tilde{s}_j^2 \left(\frac{n_j}{n} \right) = 0.227 + \sum_{j=1}^n (\bar{x}_j - \bar{x})^2 \left(\frac{n_j}{n} \right) = 0.025$
Berufsausbildung	11.216	0.200	1751	
Universität	11.536	0.342	301	
Gesamt	11.222	0.252	2296	

Kovarianz

$$\rightarrow \tilde{s}_{xy} = \sum_{i=1}^k \sum_{j=1}^m (a_i - \bar{x})(b_j - \bar{y}) f_{ij}$$

→ Mass für gemeinsame Variabilität zweier Variablen x u. y
→ =Arithm. Mittel der Abweichungsprodukte d. einzelnen Punktepaare

Stichprobenvarianz

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

→ Division durch "n-Freiheitsgrade"; je grösser n → $s^2 = \tilde{s}^2$

Variationskoeffizient

$$\rightarrow v = \frac{\tilde{s}}{\bar{x}} > 0 \rightarrow \text{Dimensionslose Grösse; für Vgl. d. Streuung zw. Merkmalen}$$

Interquartilsabstand

$$\rightarrow d_q = x_{0.75} - x_{0.25}$$

→ resistent gegen Ausreisser

Spannweite

$$\rightarrow x_{\text{max}} - x_{\text{min}}$$

Optimalität

$$\rightarrow Q(z, x_1, \dots, x_n) = \sum_{i=1}^n (x_i - z)^2$$

→ je weiter d. Wert vom Mittelwert wegliegt,
desto stärker d. Gewichtung (Quadrierung)

68-95-99%-Regel

(bei Normalverteilung)

→ in $\bar{x} \pm \tilde{s}$ liegen 68% aller Daten

→ in $\bar{x} \pm 2\tilde{s}$ liegen 95% aller Daten

→ in $\bar{x} \pm 3\tilde{s}$ liegen 99% aller Daten

Schiefe/Wölbung

Quartilkoeff. der Schiefe

$$\rightarrow g_p = \frac{(x_{1-p} - x_{\text{med}}) - (x_{\text{med}} - x_p)}{x_{1-p} - x_p} \text{ es gilt: } -1 \leq g_p \leq 1 \rightarrow \text{resistent gegen Ausreisser!}$$

→ für $p=0.25$ → Quartilkoeff.
- $g_m, g_p=0$: symmetr. Verteilung
- $g_m, g_p>0$: linkssteile Verteilung
- $g_m, g_p<0$: rechtssteile Verteilung

Momentenkoeff. d. Schiefe

$$\rightarrow g_m = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\tilde{s}^3} \rightarrow \text{nicht resistent; massstabsunabh.}$$

Wölbungsmass v. Fisher

$$\rightarrow \gamma = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{\tilde{s}^4} - 3$$

→ $\gamma = 0$: Normalverteilung

→ $\gamma > 0$: Spitzere Verteilung

→ $\gamma < 0$: Flachere Verteilung

Konzentrationsmasse

Lorenzkurve → Aussagen bez. rel. Konzentration; immer monoton wachsend; Konvexe Linkskurve

Relative Konz.:

L Gini-Koeff.

$$\rightarrow G = \frac{\text{Fläche zw. Diagonale und Lorenzkurve}}{\text{Fläche zw. Diagonale und } x - \text{Achse}}$$

-für geordnete Urliste

$$\rightarrow G = \frac{2 \sum_{i=1}^n i x_i}{n \sum_{i=1}^n x_i} - \frac{n+1}{n}$$

-für Häufigkeitsdaten

$$\rightarrow G = \frac{\sum_{i=1}^k (u_{i-1} + u_i) h_i a_i}{\sum_{i=1}^k h_i a_i} - 1$$

wobei: u_i (= kumm. Anteil Merkmalsträger) = $\sum_{j=1}^i \frac{h_j}{n}$; v_i (= kumm. rel. Merkmalssumme) = $\frac{\sum_{j=1}^i h_j a_j}{\sum_{j=1}^k h_j a_j}$

→ $G_{\min}=0$ bei Nullkonz. (=45Grad Gerade) → $G_{\max}=\frac{n-1}{n}$ bei max. Konz.
→ WB: $0 \leq G \leq 1$

Absol. Konz.:

L. Konzentrationsrate

→ $CR_g = \sum_{i=g+1}^n p_i$, wobei $p_i = \frac{x_i}{\sum_{j=1}^n x_j}$ den Merkmalsanteil der i-ten EH bezeichnet

→ Mit $x_1 \leq \dots \leq x_n$ (Daten werden geordnet)=man nimmt die grössten Anteile zuerst; wenn $CR_g \uparrow$ → Konz. $\uparrow$; rechnen in % → Resultat in %

→ WB: $\frac{g}{n} \leq CR_g \leq 1$ (=max. Konz.)

L. Herfindahl-Index

→ $H = \sum_{i=1}^n p_i^2$, wobei $p_i = \frac{x_i}{\sum_{j=1}^n x_j}$ d. Merkmalsanteil der i-ten EH bezeichnet

→ wenn $H \uparrow$ → Konz. $\uparrow$
→ $H_{\min}=(1/n)$ → min. Konz.=vollk. Konk.; $H_{\max}=1$ → max. Konz.=Monopol

→ WB: $(1/n) \leq H \leq 1$; "Rechnen mit 0,...-Anteilen" → Resultat auch 0,..."

Axiomatische Überlegungen

- 1) **Transferprinzip**: Transfer v. reicherer zu ärmerer Person soll Kennzahl der Ungleichheit verkleinern
- 2) **Skalenunabhängigkeit**: proport. d. aller Einkommen darf Ungleichheit nicht beeinflussen
- 3) **Replikationsprinzip**: eine Zusammenlegung mehrerer identischer Populationen sollte die Ungleichheit nicht beeinflussen
- 4) **Zerlegbarkeit**: es sollte ein systemat. Zusammenh. zw. d. Gesamtungleichheit u. d. Ungleichheit in Teilpopulationen bestehen

Multivariate Daten

Kontingenztabellen

→ mögl. mit absol. od. rel. Häufigkeiten

Bedingte Häufigkeit

→ man schaut eine Zeile/Spalte als ganzes (100%) an
→ 1.) bilde bedingte rel. Häufigkeiten 2.) prüfe ob Verteilung v. Bedingung abhängt
3.) wenn ja: Zusammenh. vorhanden 4.) wenn nein: die beiden Merkm. sind unabh.

Kreuzproduktverhältnis, "Odds-Ratio" (rel. Chance) $\gamma = \frac{h_{11}/h_{12}}{h_{21}/h_{22}}$ → wenn $\gamma = 1$: "gleiche Chancen"

→ macht Aussagen über Ausmass d. relativ. Chance einer Population zur anderen u. um wie viel besser diese Chancen sind

Emp. Unabhängigkeit

→ $\frac{h_{ij}}{h_{i\cdot}} = \frac{h_{\cdot j}}{n}$ [Bsp.: $\frac{\text{Frauen oec.}}{\text{Studis oec.}} = \frac{\text{Frauen}}{\text{Studis}}$] → =Grundprinzip d. Unabhängigkeit

χ^2 -Koeffiz.

→ $\chi^2 = \sum_{i=1}^k \sum_{j=1}^m \frac{(h_{ij} - \frac{h_{i\cdot} h_{\cdot j}}{n})^2}{\frac{h_{i\cdot} h_{\cdot j}}{n}}$ → WB: $\chi^2 \in [0, +\infty)$

→ wenn χ^2 gross: X u. Y abh. / wenn χ^2 klein: X u. Y unabh.
→ Nachteil: Wert hängt v. Dimension der Tabelle ab (nicht normiert!)

L. Kontingenzkoeff.

→ $K = \sqrt{\frac{\chi^2}{n + \chi^2}}$ → $K \in [0, \sqrt{\frac{M-1}{M}}]$ → WB: $[0, 1]$

→ macht Aussagen ob 2 Merkm. unabh./abh. sind

L. Korrig. Kontingenzkoeff.

→ $K^* = K / \sqrt{\frac{M-1}{M}}$ → WB: $[0, 1]$ (K: Ausprägung Klassen / M: Auspräg. Geschlecht)

Eigenschaften v. χ^2 , K, K^*

- 1) nur Stärke des Zusammenhangs, nicht Richtung
- 2) Vergleichendes Mass zw. Prop. X und Prop. Y
- 3) Abhängigkeit vom Stichprobenumfang
- 4) Invarianz

Korrelation (symmetrisch)

Def.

- Kennzahl für Stärke eines linearen Zusammenh. bei qualit. Merkmalen
- Richtung/kausale Zusammenh. werden nicht durch Korr.koeffizienten erfasst
- misst die Streuung v. Pkt. um Gerade herum; ist unabh. v. Lage u. Skalierung

Emp. Korrelationskoeff.

(Bravais-Pearson)

→ $r = \frac{\sum_{i=1}^n x_i \cdot y_i - n \bar{x} \bar{y}}{\sqrt{(\sum_{i=1}^n x_i^2 - n \bar{x}^2)(\sum_{i=1}^n y_i^2 - n \bar{y}^2)}} = \frac{\bar{s}_{xy}}{\bar{s}_x \bar{s}_y} = \frac{\text{Kovarianz } X, Y}{\text{Standardabw. } X \cdot \text{Standardabw. } Y}$

- $r > 0$: +Korrelation, gleichsinn. linearer Zusammenh., +Steigung
- $r < 0$: -Korrelation, gleichsinn. linearer Zusammenh., -Steigung
- $r = 0$: keine Korrelation, kein linearer Zusammenh., Steigung=0
- $r < 0.5$: schwach / $0.5 < r < 0.8$: mittel / $r > 0.8$: stark
- für metrische Merkm. → Bereich der Korrelation: $-1 \leq r \leq 1$
- Mass für Stärke d. linearen Z'hangs zw. 2 Variablen (bez. Stand.abweichung!)

Rangkorrelationskoeff.

(Spearman)

→ $r_{sp} = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{\sqrt{(\sum x_i^2 - n \bar{x}^2)(\sum y_i^2 - n \bar{y}^2)}}$ oder $r_{sp} = 1 - \frac{6 \cdot \sum d_i^2}{(n^2 - 1) \cdot n}$

→ WB: $-1 \leq r_{sp} \leq 1$ (r u. r_{sp} : sind gegenüber linearen Transformationen invariant!)

- di=Differenz zw. den Rängen ($di=x_i-y_i$)
- ACHTUNG: Rangsumme muss gleich gross sein-unbedingt Ränge richtig verteilen!
- bei gleichen Rängen werden Durchschnittsränge vergeben!
- allgemeiner als Bravais-Pearson Koeff. (nicht d. Daten sondern d. Ränge sind entscheidend!)
- "nicht parametrische" Kennzahl: für ordinale/metrische Merkmale

- $r_{sp} > 0$: gleichsinn. monotoner Zusammenh.
- $r_{sp} < 0$: gegensinn. monotoner Zusammenh.
- $r_{sp} = 0$: kein monotoner Zusammenh.
- $r_{sp} = 1$: funktionaler monotoner Zusammenhang vorhanden

Scheinkorrelat.

Verdeckte Korr.

→ drittes Merkmal welches mit d. gemessenen 2 Merkmalen hochkorreliert, wird vergessen
→ eine Population zerfällt hinsichtlich d. interessierenden Zusammenh. in Teilpopulationen

Regression (nicht symmetrisch)

Allg.

→ $y_i = a + b x_i + e_i$ (y=Regressand, abhäng. / x=Regressor, unabh. (kann binär sein) / e_i =Residuen)

Ausgleichsgerade

→ $\hat{y} = \hat{a} + \hat{b}x$

Steigung/Schwerpkt.

→ $b = \frac{\sum_{i=1}^n y_i x_i - n \bar{y} \bar{x}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\text{cov}(x, y)}{s_x^2} = \frac{\text{Kovarianz}}{\text{Varianz}}$, $a = y - b \bar{x}$

Y-Achsenabschnitt

Expon. Regress.gleichung

→ 1) Modell linearisieren 2) Zurückrechnen
→ $\hat{y}_i = a \cdot b^{x_i}$ → Linearisierung → $\ln \hat{y}_i = \ln a + \ln b \cdot x_i = \hat{a} + \hat{b} \cdot x_i$

L. Komponenten: $-\ln b = \hat{b} = \frac{\sum x_i \ln(y_i) - n \bar{x} \cdot \ln(\bar{y})}{\sum x_i^2 - n \bar{x}^2}$
 $-\ln a = \hat{a} = \ln(\bar{y}) - \hat{b} \bar{x}$

L. Für Rückrechnung: $\left. \begin{array}{l} -b = e^{\hat{b}} \\ -a = e^{\hat{a}} \end{array} \right\} \Rightarrow \hat{y}_i = a \cdot b^{x_i}$

Streuungszerlegung

→ $SQT = SQE + SQR = \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2$
→ SQT =Gesamte Streuung / SQE =Erklärte St. / SQR =Residual St.

Bestimmtheitsmass

→ $R^2 = \frac{SQE}{SQT} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = 1 - \frac{SQR}{SQT}$

→ WB: $[0, 1]$, je näher R^2 bei 1 ist, desto genauer ist d. Annahme ($R^2=1$: alle Pkt. liegen auf d. Geraden)

→ d. Quadrat d. Korrelationskoeffizienten; =d. Anteil d. Gesamtvariation

→ R^2 =erklärte Varianz / $1-R^2$ =unerklärte Varianz

→ macht Aussagen über d. Güte d. Anpassung (Ausgleichsgeraden)

→ $R \cdot 100$ =Anteil d. Gesamtvariation d. abhäng. Variablen in %

Residuen (Abweich. v. Gerade)

→ $e_i = y_i - \hat{y}_i$

wobei: y_i =effektiver Ausprägungspkt. / $\hat{y}_i$ =Pkt. auf Ausgleichsgeraden

Wahrscheinlichkeit

Allg.

→ Zufallsvorgang= mögl. Ergebnisse sind bekannt, wobei d. Ergebnis v. Zufall abh. ist / Zufallsexp.=Zufallsvorgang, Ablauf unter kontroll. Bedingungen, beliebig. wiederholbar / Ereignisraum (Ω)=Menge aller Elementarereignisse / Elementarereignis (ω)=Ergebnis eines Zufallsexp.(einelement.) / Zufallereignis (A)=Teilmenge v. Ereignisraum / Subj. W.Begriff=individ. Einschätzung / Obj. W.Begriff=basiert auf Häufigkeitsinterpretation / Axiomatisch. W.Begriff=Wahrsch. als Funktion ($0 < P < 1$)

→ sicheres Ereignis: $A = \Omega$ / unmögl. E.: $A = \{ \}$ / ausschöpfendes E.: $A \cup B = \Omega$ / disjunktes E.: $A \cap B = \{ \}$ (disjunkt=entw. A od. B)

Eigenschaften

- 1) $0 \leq P(A) \leq 1$ 2) $P(\{ \}) = 0$ 3) $P(A) \leq P(B)$ falls $A \subset B$ 4) $P(\bar{A}) = 1 - P(A)$
- 5) $P(A) = P(A \cap B) + P(A \cap \bar{B})$ 6) $P(B) = P(A \cap B) + P(\bar{A} \cap B)$ 7) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Additionssatz

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$ Würfelbsp.: P(gerade od. grösser gleich 3) = $\frac{3}{6} + \frac{4}{6} - \frac{2}{6} = \frac{5}{6}$
→ d. Wahrsch. d. Vereinigung zweier Mengen A und B ist gleich d. Summe d. Wahrsch. beider Mengen abzüglich d. Wahrsch. der Schnittmenge

Laplace (Gleichwahrsh.)

→ $P(A) = \frac{\text{Anz. der für A günstigen Ergebnisse}}{\text{Anz. aller mögl. Ergebnisse}}$
→ für einfache Zufallsstichprobe (Bsp.: Würfeln)
→ Voraussetzung ist Gleichwahrsh. der Ergebnisse

L Bedingte Wahrsch.

→ $P(A/B) = \frac{\text{Anz. der für A u. B günstigen Ergebnisse}}{\text{Anz. der für B günstigen Ergebnisse}} = \frac{P(A \cap B)}{P(B)}$

→ eine neue Info kommt hinzu und beeinflusst!

L Produktregel

→ $P(A \cap B) = P(A/B) \cdot P(B) = P(B/A) \cdot P(A) \triangleq$ Gemeinsame Wahrsch. = Produkt d. bedingten W. u. d. Randw.

Allg. Schema

	ohne Zurückleg.	Mit Zurückleg.
mit Berücksicht. der Reihenfolge	$\frac{N!}{(N-n)!}$	N^n
ohne Berücksicht. der Reihenfolge	$\binom{N}{n}$	$\binom{N+n-1}{n}$

Permutation

→ $N!$ → z.B.: aus Urne alle Lose ziehen u. notieren d. Nr. (Modell ohne Zurückleg.)
→ es gibt $N!$ viele versch. Mögl.

Binomialkoeffizient

→ $\binom{N}{n} = \frac{N!}{n!(N-n)!}$ → =Modell: "ohne ZL+ohne Berücksichtigung d. Reihenfolge"

Unabhängigkeit (1)

→ 2 Ereignisse sind voneinander unabh. wenn gilt: $P(A \cap B) = P(A) \cdot P(B)$

→ Disjunkte Ergebnisse = nicht unabh. (wenn A eingetreten, kann B nicht eintreten!)

Unabhängigkeit (2)

→ $P(A/B) = P(A)$ bzw. $P(B/A) = P(B)$ → d. Eintreten d. einen Ereignisses hat

keinen Einfluss auf Eintreten d. Anderen → Bei Unabh. gilt: $P(A \cap B) = P(A) \cdot P(B)$

→ Unabh. Ergebnisse: "Ziehen mit Zurücklegen"

→ Abh. Ergebnisse: "Ziehen ohne Zurückl."

Satz v. Bayes

→ $P(A \cap B) = P(A/B) \cdot P(B) \Leftrightarrow P(A \cap B) = P(B/A) \cdot P(A) \Rightarrow P(A/B) \cdot P(B) = P(B/A) \cdot P(A) \dots$

...bzw. $P(A/B) = \frac{P(B/A) \cdot P(A)}{P(B)}$

Totale Wahrscheinlich.

→ gibt an wie man v. Daten neuer Infos lernt: = "Regel für induktives Lernen"

Sei $A_1, \dots, A_k$ eine disjunkte Zerlegung von Ω . Dann gilt für

$$B \subset \Omega : P(B) = \sum_{i=1}^k P(B/A_i) \cdot P(A_i)$$

Axiome Kolmogorof

	Normalfall	Für unendliche GG
1	$P(A) \geq 0$	$P(A) \geq 0$
2	$P(\Omega) = 1$	$P(\Omega) = 1$
3	falls $A \cap B = \{\}$ gilt: $P(A \cup B) = P(A) + P(B)$	Seien $A_1, \dots, A_k \subset \Omega$ paarweise disjunkt: $P(A_1 \cup \dots \cup A_k \cup \dots) = \sum_{i=1}^{\infty} P(A_i)$

a-priori-Wahrsch.

→ Wahrsch., subj. Wahrsch., d. vor d. Ankunft neuer Infos besteht, beruht i.d.R.

a-posteriori-Wahrsch.

→ auf verganginem Lernen früherer Infos

→ aktualisierte Wahrsch. → die neue Info wird berücksichtigt → $P(A/B)$

Diskrete Zufallsvariable (Skalierung: nominal-, ordinal-, kardinal-, ist möglich!)

Def. ZV

→ eine Variable od. ein Merkmal X , dessen Werte od. Ausprägungen d. Ergebnisse eines Zufallvorgangs sind heisst ZV X

→ d. Zahl $x \in \mathbb{R}$, d. X bei einer Durchführung d. Zufallvorgangs annimmt, heisst

Realisierung od. Wert v. X

→ ZV X = Abbildung, d. jedem $\omega \in \Omega$ eine reelle Zahl $X(\omega) = x$ zuordnet → $X: \Omega \rightarrow \mathbb{R}$

Def. diskrete ZV

→ ZV X = diskret, falls sie nur endlich od. abzählbar unendlich viele Werte annehm. kann

Wahrscheinlichkeitsfkt.

→ $f(x) = \begin{cases} P(X = x_i) = p_i, & x = x_i \in \{x_1, x_2, \dots, x_k, \dots\} \\ 0, & \text{sonst} \end{cases}$

Wahrsch.keitsverteilung

→ d. ZV X ist d. wahrscheinl.keitstheoret. Analogon zur relativ. Häufigkeitsverteil.

Verteilungsfunktion

→ $F(x) = P(X \leq x) = \sum f(x_i)$ → =Kumulieren d. Wahrscheinlichkeitsverteilung

Standardabweichung

→ $\sigma = \sqrt{\text{Var}(X)}$

Varianz

→ $\text{Var}(X) = \sigma^2 = \sum_{i=1}^k (x_i - \mu)^2 \cdot f(x_i)$

L Verschiebungsregeln

1) $\text{Var}(X) = E(X^2) - (E(X))^2 = E(X^2) - \mu^2 = E(X - \mu)^2$ 2) $\text{Var}(X) = E((X - c)^2) - (\mu - c)^2$

L Transformationsregel

→ Sei $g(x)$ eine reelle Funktion, dann gilt für $Y = g(X)$:

→ $Y = g(X): E(Y) = E(g(X)) = \sum_{i=1}^k g(x_i) \cdot f(x_i)$

→ Bei Linearität gilt: $E(Y) = \sum_{i=1}^k (a + b \cdot x_i) \cdot f(x_i) = a \sum_{i=1}^k f(x_i) + b \sum_{i=1}^k x_i \cdot f(x_i) = a + b E(X)$

L Additionsregel

→ $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$

→ =d. zu erwartende Abweichung d. ZV v. ihrem Erwartungswert

→ Transformationsregeln v. Varianz/Standardabw. = Regeln im deskriptiven Bereich

Erwartungswert (Allg.)

→ $E(X) = \sum x_i p_i = \mu_x$ → charakterisiert d. Verhalten eines Zufallsexperimentes

→ (ACHTUNG: $E(X) \neq AM$) → Transformationsregeln = analog wie bei AM

L Additionsregel

→ $E(X + Y) = E(X) + E(Y)$

L Produktregel

→ $E(X \cdot Y) = E(X) \cdot E(Y)$ (gilt für unabhängige ZV's)

Unabhängigk. disk. ZV's

→ für beliebige Werte aus Wertebereich muss gelten:

→ $P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = P(X_1 = x_1) \cdot P(X_2 = x_2) \cdot \dots \cdot P(X_n = x_n)$

→ Fall: "Augenzahl beim Würfeln"

Diskr. Gleichverteilung

→ $P(X = x_i) = \frac{1}{k}$ → Wahrscheinlichkeitshistogramm:

Geometrische Verteil.

→ Fall: "Anzahl d. Versuche bis zum ersten Mal A eintritt" (gleich bleibende Wahr.!)

→ $P(X = x_i) = (1 - \pi)^{x-1} \cdot \pi$

→ Wahrscheinlichkeitshistogramm:

→ Verteilungsfunktion:

L Erwartungswert

→ $E(X) = \frac{1}{\pi}$

L Varianz

→ $\text{Var}(X) = \frac{1 - \pi}{\pi^2}$

Poissonverteilung

→ $f(x) = P(X = x) = \begin{cases} \frac{\lambda^x}{x!} \cdot e^{-\lambda}, & x \in \{0, 1, \dots\} \\ 0, & \text{sonst} \end{cases}$

λ = Erwartungswert

→ Fall: "Anzahl d. Ereignisse, d. innerhalb d. Intervalls $[0,1]$ eintreten"

(Bsp.: Schadensmeldungen bei Versicherung innerhalb eines Jahres)

→ gegeben wenn folgendes erfüllt ist:

1) 2 Ereignisse können nicht genau gleichzeitig auftreten

2) wenn Δt klein wird, wird d. W. d. Eintreffens d. Ereignisses ebenfalls kl

("Verteilung d. seltenen Ereignisse!")

λ = Intensitätsrate

3) d. W. d. Eintretens hängt nicht v. Lage d. Teilintervalles ab, sondern v. Länge

→ $E(X) = \lambda$

→ $\text{Var}(X) = \lambda$

L Erwartungswert

L Varianz

Binomialverteilung

→ $f(x) = \begin{cases} \binom{n}{x} \cdot \pi^x (1 - \pi)^{n-x}, & x = 0, 1, 2, \dots, n \\ 0, & \text{sonst} \end{cases}$

n = unabh. Wiederholungen

π = Erfolgswahrscheinlichkeit

→ Fall: "Ziehen mit Zurücklegen" (Wahrsch. d. interessierenden Ereignisse = π)

→ Fall: "Anzahl Versuche bei denen A eintritt"

→ $E(X) = n \cdot \pi$

→ $\text{Var}(X) = n \cdot \pi (1 - \pi)$

→ Sind $X \sim B(n, \pi)$ und $Y \sim B(m, \pi)$ unabh., so ist $X + Y \sim B(n + m, \pi)$

→ Sei $X \sim B(n, \pi)$ und $Y = n - X$. Dann gilt: $Y \sim B(n, 1 - \pi)$

L Erwartungswert

L Varianz

L Additionsregel

L Symmetrieeigenschaft.

Hypergeom. Verteilung

→ $f(x) = \begin{cases} \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}}, & x \in T \\ 0, & \text{sonst} \end{cases}$

WB: → $x_{\max} = \min(n, M)$

→ $x_{\min} = \max(0, n - (N - M))$

→ Fall: "Ziehen ohne Zurücklegen / Anz. d. gezogenen Objekte mit Eigenschaft A"

→ Wahrscheinlichkeit ändert mit jedem Zug!

→ N : Umfang d. Grundgesamtheit / M : Anzahl EH, d. Eigenschaft A besitzt /

n : Grösse d. Stichprobe

L Erwartungswert

→ $E(X) = n \cdot \frac{M}{N}$

L Varianz

$$\rightarrow \text{Var}(X) = n \cdot \frac{M}{N} \left(1 - \frac{M}{N} \right) \cdot \frac{N-n}{N-1}$$

Multinomialverteilung

$$\rightarrow f(j) = \begin{cases} \pi_1^{d_1} \cdot \pi_2^{d_2} \cdot \dots \cdot \pi_j^{d_j}, j=1,2,\dots,J \\ 0, \text{sonst} \end{cases} \quad \text{wobei: } \pi_j = 1 - \pi_1 - \dots - \pi_{j-1}$$

→ Bsp: folgende Parteistärke ist gegeben: "SP=25% / SVP=25% / FDP=25% / CVP=12.5% / Ub=12.5%"
Wie gross ist d. W., dass sich unter 8 zufällig ausgewählten Wählern d. genau gleiche Verteilung einstellt?

$$\dots P(2,2,2,1,1) = \frac{8!}{2!2!2!1!1!} \cdot 0.25^2 \cdot 0.25^2 \cdot 0.25^2 \cdot 0.125^1 \cdot 0.125^1 = 0.0192 = 2\%$$

Übersicht Verteilungen

X - G(π) → geom. Verteilung
X - B(n, π) → Binomialverteilung
X - H(n, M, N) → hypergeom. Verteilung
X - Po(λ) → Poisson Verteilung

Lageregeln

– Symmetrische Verteilung: $x_{\text{mod}} = x_{\text{med}} = E(X)$
– Linkssteile Verteilung: $x_{\text{mod}} < x_{\text{med}} < E(X)$
– Rechtssteile Verteilung: $x_{\text{mod}} > x_{\text{med}} > E(X)$

Zeichenerklärung

→ M=Anzahl / n=Grösse d. Stichprobe / N=Grundgesamtheit / λ =Erwartungswert / π =Erfolgswahrscheinlichkeit

Stetige Zufallsvariable (Skalierung: metrisch-, ordinal-, ist möglich!)

Def. → ZV welche jeden Wert im Intervall $[a, b]$ annehmen kann

→ Werte sind nicht mehr abzählbar, sondern überabzählbar

→ die Fkt. $f(x)$ heisst (Wahrscheinlichkeits-) Dichte von X

→ ZV X=stetig, wenn es eine Fkt. $f(x) \geq 0$, sodass für jedes Intervall $[a, b]$ gilt: $P(a \leq X \leq b) = \int_a^b f(x) dx$

Wahrscheinlichkeit

→ $P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b)$ und $P(X = x) = 0$ für jedes $x \in \mathbb{R}$ (ist d. Fall wenn: $a = b = x$)

L Normierungseigenschaft → $\int_{-\infty}^{+\infty} f(x) dx = 1 \Rightarrow$ gesamte Fläche = 1

Dichtefunktion

→ gegeben falls erfüllt:
1.) $f(x) \geq 0$ (da es keine negativen Wahrsch. gibt!)
2.) $\int_{-\infty}^{+\infty} f(x) dx = 1$ (= sichere Ereignis)

Verteilungsfunktion

→ $F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt$

1) $F(x)$ = stetig und monoton wachsend (Mit Werten im Intervall $[0, 1]$)
2) Für Grenzen gilt: $\rightarrow F(-\infty) = \lim_{x \rightarrow -\infty} F(x) = 0$ (unmögl. Ereignis)
 $\rightarrow F(+\infty) = \lim_{x \rightarrow +\infty} F(x) = 1$ (sicheres Ereignis)
3) $F'(x) = \frac{dF(x)}{dx} = f(x) \rightarrow$ Dichte = Ableitung d. Verteilungsfkt.!
4) Für Intervalle gilt: $\rightarrow P(a \leq X \leq b) = F(b) - F(a)$
 $\rightarrow P(X \geq a) = 1 - F(a)$

L Eigenschaften

Erwartungswert

→ $\mu = E(X) = \int_{-\infty}^{+\infty} x f(x) dx$

1) Additionsregel: $E(X+Y) = E(X) + E(Y)$
2) Lin. Transformation: für $Y = aX + b$ gilt: $E(Y) = E(aX + b) = aE(X) + b$
3) Symmetr. Verteilung: ist die Dichte $f(x)$ symmetr. um Pkt. c , dh $f(c-x) = f(c+x)$ für alle x , so gilt: $E(X) = c$
4) Produktregel bei Unabhängigkeit

L Eigenschaften

Varianz

→ $\sigma^2 = \text{Var}(X) = \int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx = E(X^2) - [E(X)]^2$

Standardisierung

$$\rightarrow Z = \frac{X - \mu}{\sigma}$$

Unabhängigkeit

→ $P(X \leq x, Y \leq y) = P(X \leq x) \cdot P(Y \leq y) = F_X(x) \cdot F_Y(y)$
oder allgemeiner: $P(X \leq x_1, \dots, X_n \leq x_n) = P(X_1 \leq x_1, \dots, P(X_n \leq x_n)$

Stetige Gleichverteilung

→ Fall: "Stelle des Glücksrades auf die der Pfeil zeigt"

$$\rightarrow f(x) = \begin{cases} \frac{1}{b-a} & \text{für } a \leq x \leq b \\ 0 & \text{sonst} \end{cases} \quad \boxed{a, b = \text{Intervallgrenzen}}$$

→ eine auf Intervall $[0, 1]$ gleichverteilte ZV nennt man auch Standardgleichverteilt

L Verteilungsfunktion

→ ergibt sich aus Integration: $F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & x > b \end{cases}$

L Erwartungswert

$$\rightarrow E(X) = \frac{b+a}{2}$$

L Varianz

$$\rightarrow \text{Var}(X) = \frac{(b-a)^2}{12}$$

Exponentialverteilung

→ =Pendant zur geometr. Verteilung (bei diskreten ZV's)
→ =gedächtnislos=vorgängige Aktionen haben keinen Einfluss auf Zukunft

$$\rightarrow f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{für } x \geq 0 \\ 0 & \text{für } x < 0 \end{cases} \quad \boxed{\lambda = \text{Anz. Aktionen pro Zeiteinheit}}$$

→ je grösser λ , desto schneller geht d. Exp.fkt. für $x \rightarrow \infty$ gegen 0 u. d. Verteilungsfkt. gegen 1 (Fläche unter $f(x)$ ist immer=1)

L Verteilungsfunktion

→ ergibt sich aus Integration: $F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{für } x \geq 0 \\ 0 & \text{für } x < 0 \end{cases}$

L Erwartungswert

$$\rightarrow E(X) = \int_0^{\infty} x \lambda e^{-\lambda x} dx = \frac{1}{\lambda}$$

L Varianz

$$\rightarrow \text{Var}(X) = \frac{1}{\lambda^2}$$

L Poissonverteilung

$$\rightarrow f(x) = P(X = x) = \begin{cases} \frac{\lambda^x}{x!} e^{-\lambda}, & x \in \{0, 1, \dots\} \\ 0 & \text{sonst} \end{cases} \quad \boxed{E(X) = \lambda} \quad \boxed{\text{Var}(X) = \lambda}$$

→ steht im engen Zusammenhang mit Exponentialverteilung:
d. Anzahl v. Ereignissen in einem Zeitintervall ist genau dann Po(λ) verteilt, wenn d. Zeitdauer zw. aufeinander folgenden Ereignissen unabhängig und exponentialverteilt mit Parameter λ sind

Normalverteilung

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

L Erwartungswert

$$\rightarrow E(X) = \mu$$

L Varianz

$$\rightarrow \text{Var}(X) = \sigma^2$$

L Transformationen

$$\rightarrow \text{Sei } X \sim N(\mu, \sigma^2) \text{ und } Y = a + bX \Rightarrow X \sim N(a + b\mu, b^2\sigma^2)$$

→ =symmetrisch um Erwartungswert; Max. liegt bei μ ; Wendepunkte bei $\mu \pm \sigma$
→ ist d. gleiche Verteilung wie Normalverteilung ausser dass:

$$\rightarrow E(X) = \mu = 0 \text{ und } \sigma^2 = 1 \text{ und } \sigma = 1$$

$$\rightarrow \text{Dichtefunktion} = \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

L Transformationen

$$\rightarrow \text{Sei } X \sim N(\mu, \sigma^2) \text{ und } Y = a + bX \Rightarrow X \sim N(a + b\mu, b^2\sigma^2)$$

Log. Normalverteilt.

→ für Verteilung v. nicht-negativen ZV's wie Lebensdauern usw.; =Linkssteil

→ d. Parameter μ und σ^2 haben für d. Lognormalverteilung nicht d. Bedeutung v. Erwartungswert u. Varianz!

$$\rightarrow X \sim \text{LN}(\mu, \sigma^2) \Leftrightarrow Y = \ln X \sim N(\mu, \sigma^2)$$

L Erwartungswert

$$\rightarrow E(X) = e^{\mu + \sigma^2/2}$$

L Varianz

$$\rightarrow \text{Var}(X) = e^{2\mu + \sigma^2} (e^{\sigma^2} - 1)$$

Logistische Verteilung

→ ähnlich zur Normalverteilung (mit $\sigma = 1,6$)

X²-Verteilung

L Erwartungswert

L Varianz

t-Verteilung

L Erwartungswert

L Varianz

Fisher-Verteilung

L Erwartungswert

L Varianz

68-95-99,7% Regel

- 1) → ca. 68% aller Beobachtungen liegen innerhalb einer Standardabweichung um d. Mittelwert (zw. $\mu - \sigma$ und $\mu + \sigma$)
- 2) → ca. 95% aller Beobachtungen liegen innerhalb zweier Standardabweichung um d. Mittelwert (zw. $\mu - 2\sigma$ und $\mu + 2\sigma$)
- 3) → ca. 99,7% aller Beobachtungen liegen innerhalb dreier Standardabweichung um d. Mittelwert (zw. $\mu - 3\sigma$ und $\mu + 3\sigma$)
- Beobachtungen jenseits der $\pm 3\sigma$ sind also extrem unwahrscheinlich!

Grafik (Grafik auf Seite 303)

Standardisierung

$$Z = \frac{X - \mu}{\sigma} \quad \rightarrow 1) \sigma \text{ und } \mu \text{ einsetzen } 2) \text{ nach } X \text{ auflösen } 3) \text{ Intervallgrenzen setzen } \rightarrow \text{ Tabelle ablesen!}$$

Erwartungswert

L Eigenschaften

$$\mu = E(X) = \int_{-\infty}^{+\infty} x \cdot f(x) dx \quad (\text{einer stetigen ZV } X \text{ mit Dichte } f(x))$$

- 1.) Transformationen: $g(X)$ = reelle Funktion. Dann gilt für $Y = g(X)$: $E(Y) = E(g(X)) = \sum_{-\infty}^{+\infty} g(x) \cdot f(x) dx$
- 2.) Lineare Transformation: für $Y = a \cdot X + b$ ist: $E(Y) = E(a \cdot X + b) = a \cdot E(X) + b$
- 3.) Symmetrische Verteilungen: Ist d. Dichte $f(x)$ symmetrisch um den Pkt. c , d.h. $f(c - x) = f(c + x)$ für alle x , so gilt: $E(X) = c$
- 4.) Additivität: für 2 ZV's X und Y ist: $E(X + Y) = E(X) + E(Y)$
- Allgemeiner: $E(a_1 X_1 + \dots + a_n X_n) = a_1 E(X_1) + \dots + a_n E(X_n)$

Modus

- bezieht sich auf d. Maxima der Dichtefunktion → x_{mod}
- falls Maxima eindeutig ist, u. $f(x)$ keine weitere besitzt, heisst $f(x)$ unimodal
- falls $f(x)$ unimodal und symmetrisch um c ist → $c = x_{\text{mod}} = E(X)$

Median

$$\rightarrow x_{\text{med}} = 50\% \text{ Quantil: } F(x_{\text{med}}) = 1 - F(x_{\text{med}}) = \frac{1}{2}$$
$$\rightarrow \text{bei stetiger Gleichverteilung: } F(x_{\text{med}}) = \frac{a+b}{2}$$

Grafik 13

$$\rightarrow \text{bei Exponentialverteil.: } \int_0^{x_{\text{med}}} \lambda \cdot e^{-\lambda x} dx = \frac{1}{2} \quad \rightarrow \text{Achtung: richtig auflösen (mit logarithmieren)}$$

Quantil

$$\rightarrow F(x_p) = P(X \leq x_p) = p \quad \text{bzw.} \quad x_p = F^{-1}(p)$$

→ für $0 < p < 1$ ist das p -Quantil x_p die Zahl auf der x -Achse!

Symmetrie

$$\rightarrow E(X) = x_{\text{mod}} = x_{\text{med}} \quad (\dots \text{bez. einem Punkt } c)$$

Standardabweichung

Varianz

$$\rightarrow \sigma = \sqrt{\text{Var}(X)}$$

→ einer stetigen ZV X mit Dichte $f(x)$:

$$\text{L Normalfall: } \sigma^2 = \text{Var}(X) = \int_{-\infty}^{+\infty} (x - \mu)^2 \cdot f(x) dx = E[(X - E(X))^2], \text{ mit } \mu = E(X)$$

$$\text{L Exponentialverteilung: } \text{Var}(X) = \frac{1}{\lambda^2}$$

- 1.) $\text{Var}(X) = E((X - \mu)^2)$
- 2.) Verschiebungsregel: $\rightarrow \text{Var}(X) = E(X^2) - (E(X))^2 = E(X^2) - \mu^2$
 $\rightarrow \text{Var}(X) = E((X - c)^2) - (\mu - c)^2$
- 3.) Lineare Transformation: für $Y = a \cdot X + b$:
 $\rightarrow \text{Var}(Y) = \text{Var}(a \cdot X + b) = a^2 \cdot \text{Var}(X)$, $\sigma_Y = |a| \sigma_X$
- 4.) Varianz der $\sum$ v. unabhäng. ZV's: $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$
- allgemeiner: $\text{Var}(a_1 X_1 + \dots + a_n X_n) = a_1^2 \text{Var}(X_1) + \dots + a_n^2 \text{Var}(X_n)$

L Eigenschaften

Symmetrie / Schiefe

- Symmetrisch → wenn: $x_{\text{med}} - x_p = x_{1-p} - x_{\text{med}}$
- Linkssteil → wenn: $x_{\text{med}} - x_p \leq x_{1-p} - x_{\text{med}}$
- Rechtssteil → wenn: $x_{\text{med}} - x_p \geq x_{1-p} - x_{\text{med}}$

Lageregeln

- Symmetrische Verteilung: $x_{\text{mod}} = x_{\text{med}} = E(X)$
- Linkssteile Verteilung: $x_{\text{mod}} < x_{\text{med}} < E(X)$
- Rechtssteile Verteilung: $x_{\text{mod}} > x_{\text{med}} > E(X)$

Summen von ZV's

L Erwartungswert

$$\rightarrow Y = X_1 + \dots + X_n \text{ bzw. in allg. Form: } Y = a_1 X_1 + \dots + a_n X_n$$
$$\rightarrow E(Y) = E(X_1) + \dots + E(X_n) \text{ bzw. } E(Y) = a_1 E(X_1) + \dots + a_n E(X_n)$$

L Erwartungsw. v. Mittelwert

$$\rightarrow E(\bar{X}) = E\left(\frac{1}{n}(X_1 + \dots + X_n)\right) = \mu$$

L Varianz

$$\rightarrow \text{Var}(Y) = \text{Var}(X_1) + \dots + \text{Var}(X_n) \text{ bzw. } \text{Var}(Y) = a_1^2 \text{Var}(X_1) + \dots + a_n^2 \text{Var}(X_n)$$

L Varianz v. Mittelwert

$$\rightarrow \text{Var}(\bar{X}) = \text{Var}\left(\frac{1}{n}(X_1 + \dots + X_n)\right) = \frac{\sigma^2}{n}$$

→ ist kleiner als die Varianz der Komponenten

→ je grösser n , desto kleiner wird d. Varianz d. Mittelwertes

Gesetz d. grossen Zahlen

Zentraler Grenzwertsatz

- $P(|\bar{X}_n - \mu| \leq c) \rightarrow 1$ für $n \rightarrow \infty$
- für Binomial- u. Poissonapproximationen
- Allg. gilt: D. Verteil. einer Summe v. ZV konverg. für $n \rightarrow \infty$ gegen Normalverteil.

$$\rightarrow Z_n = \frac{X_1 + \dots + X_n - n\mu}{\sqrt{n}\sigma} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{X_i - \mu}{\sigma} \rightarrow \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \rightarrow Z_n \sim N(0,1)$$

L Binomialapproximation

$$\rightarrow Z_n = \frac{X_1 + \dots + X_n - n\pi}{\sqrt{n}\sigma} = \frac{X - n\pi}{\sqrt{n}\sqrt{\pi(1-\pi)}} \rightarrow X \sim N(n\pi, n\pi(1-\pi))$$

→ diese Approximation gilt für: $n\pi \geq 5$ und $n(1-\pi) \geq 5$

L Poissonapproximation

$$\rightarrow Z_n = \frac{X_1 + \dots + X_n - n\lambda}{\sqrt{n}\sigma} = \frac{X - n\lambda}{\sqrt{n\lambda}} \rightarrow X \sim N(n\lambda, n\lambda)$$

→ diese Approximation gilt für: $n\lambda \geq 10$

Punktschätzung

Def. Schätzfunktion

- Rückschluss aus ZSP auf Parameter d. Verteilung in GG
- ergibt für jede Stichprobe einen Schätzwert → Schätzfkt. = selber eine ZV

Standardfehler

$$\rightarrow \text{Schätzfkt.} = ZV \rightarrow \text{Wurzel aus d. Varianz dieser ZV} = \text{Standardfehler} = \frac{\sigma}{\sqrt{n}}$$

L Eigenschaften der Schätzfunktion:

1.) Erwartungstreue

- Schätzstatistik sollte tendenziell d. richtigen Wert liefert, d.h. weder systematisch über- noch unterschätzen!
- wenn gilt: $E_g(T) = \theta \Leftrightarrow \text{Bias} = 0$ (Bias = Verzerrung = system. Über- / Unterschätzen)
- $\text{Bias}_g(T) = E_g(T) - \theta$ (θ = exakte Wert)

- → $E(\bar{x}) = E(x) \Rightarrow$ also ist $\bar{x}$ eine unverzerrte Schätzfunktion für:

- $E(x)$ beliebige Verteilung
- μ der Normalverteilung
- π der Bernoulli-Verteilung

- → $S^2 = \frac{n}{n-1} \tilde{S}^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2 \Rightarrow$ Stichprobenvarianz ist unverzerrt für σ^2

– Standardfehler v. erwartungstreuer Schätzstat. gibt Auskunft ü. Genauigk. d. Schätzverfahrens, da $\hat{\sigma}_x = \frac{\sqrt{S^2}}{\sqrt{n}}$

- asymptotisch erwartungstreu, falls gilt: $\lim_{n \rightarrow \infty} E_{\theta}(T) = \theta$
- d. Genauigkeit des Schätzverfahrens wird durch d. Varianz d. Schätzstatistik bestimmt
- d. Standardabweichung d. Schätzstatistik = Standardfehler: $\sigma_{\hat{\theta}} = \sqrt{\text{Var } g(X_1, \dots, X_n)}$
- 2.) **Wirksamkeit (Effizienz)**
- je kl. Varianz (u. damit SP-Fehler!), desto besser; mit steigender Stichprobengröße (Info's) nimmt Varianz ab
- MSE-Wirksamkeit v. Schätzstatistiken: Von zwei Schätzstatistiken T_1 und T_2 heisst T_1 MSE-wirksamer, wenn gilt: $MSE(T_1) \leq MSE(T_2)$
- Wirksamkeit v. erwartungstreuen Schätzstatistiken: Von zwei erwartungstreuen Schätzstatistiken T_1 und T_2 heisst T_1 wirksamer oder effizienter, wenn gilt: $\text{Var}(T_1) \leq \text{Var}(T_2)$

$$-MSE = E[(T - \theta)^2] = \text{Var}(T) + \underbrace{[E(T) - \theta]^2}_{=0} \Rightarrow \text{erwart. Abwertung zw. Schätzfkt. u. wahren } \theta$$

-eine effiziente Schätzfkt. minimiert die MSE!

3.) Konsistenz

- Konsistenz=Verzerrung=0 $\rightarrow$ Varianz=0
- =eine Eigenschaft die d. Verhalten bei grossen Stichprobenumfängen reflektiert!

$$-MSE = E[(T - \theta)^2] = \text{Var}(T) + [E(T) - \theta]^2 \quad \rightarrow (\text{MSE} = \text{mittlere quad. Abweich.})$$

-MSE-Konsistenz falls: $MSE \xrightarrow{n \rightarrow \infty} 0 \quad \rightarrow$ "konsistent im quadratischen Mittel"

-ACHTUNG: Schätzfunktionen können konsistent sein auch wenn sie nicht erwartungstreu sind!!!

-schwache Konsistenz: $\lim_{n \rightarrow \infty} P(|T - \theta| < \varepsilon) = 1$ bzw. $\lim_{n \rightarrow \infty} P(|T - \theta| \geq \varepsilon) = 0$

- Prinzipiell gilt: jede Schätzstatistik die im quadr. Mittel konsistent ist, ist auch schwach konsistent
- jede erwartungstreu Schätzstatistik ist schwach konsistent, wenn $\text{Var}(T) = 0$ für $n \rightarrow \infty$

Konstruktion von Schätzfunktionen

1.) Maximum Likelihood-Schätzung

-Max. Likelihood Fkt.=Produkt aus n identischen Dichte- bzw. Wahrsch.fktionen

$$- \theta = \text{unbekannt, soll als Fkt. für Realisation von } X_1, \dots, X_n \text{ aufgebaut werden: } L(\hat{\theta}) = f(x_1, \dots, x_n | \theta)$$

$\rightarrow$ diese Fkt. ist zu maximieren! $\rightarrow$ meist verwendet man $\ln: \ln L(\theta) = \sum_{i=1}^n \ln f(x_i | \theta)$

-ML-Schätzer sind asymptotisch normalverteilt $\rightarrow$ konsistent und asympt. effizient!

$$- \text{Max. Likelihood-Schätzer für Poissonverteilung: } \hat{\lambda} = \frac{\sum_{i=1}^n x_i}{n} = \bar{x}$$

$$- \text{Max. Likelihood-Schätzer für Normalverteilung: } \hat{\sigma} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

2.) Kleinste Quadrate Schätzung (OLS)

-d. Parameterschätzung besteht darin, d. aufsummierten quadr. Abweichungen zw. Beobachtungswert u.

$$\text{geschätztem Wert zu minimieren: } \sum_{i=1}^n (X_i - \mu)^2 \rightarrow \min$$

$\rightarrow$ daraus resultiert nach einfacher Ableitung als Schätzer das AM $\bar{X}$

Intervallschätzung

Def.

$\rightarrow$ gibt Auskunft über d. Mass d. Unsicherheit d. Stichprobe

$\rightarrow \alpha$ = Irrtumswahrscheinlichkeit / $1 - \alpha$ = Gegenwahrscheinlichkeit (Konfidenzniveau)

$\rightarrow$ liefern den wahren Wert θ

1.) $(1 - \alpha)$ -Konfidenzintervall (bei normalverteiltem Merkmal)

$$a) \sigma^2 \text{ bekannt: } \left[\bar{X} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right], \text{ wobei } Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

$$b) \sigma^2 \text{ unbekannt: } \left[\bar{X} - t_{1-\frac{\alpha}{2}}(n-1) \frac{S}{\sqrt{n}}, \bar{X} + t_{1-\frac{\alpha}{2}}(n-1) \frac{S}{\sqrt{n}} \right] \text{ mit } S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2} \text{ und } T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

2.) $(1 - \alpha)$ -Konfidenzintervall (bei beliebiger Verteilung, $n > 30$)

$$a) \sigma^2 \text{ bekannt: } \left[\bar{X} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right], \text{ wobei } Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

$$b) \sigma^2 \text{ unbekannt: } \left[\bar{X} - z_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}, \bar{X} + z_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}} \right] \text{ mit } S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2} \text{ und } T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

3.) $(1 - \alpha)$ -Konfidenzintervall für den Anteilswert π

$$\left[\hat{\pi} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}, \hat{\pi} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}} \right], \text{ wobei } \hat{\pi} = \bar{X} \text{ die relative Häufigkeit bezeichnet und } Z = \frac{\hat{\pi} - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}}$$

Stichprobenfehler

$\rightarrow$ halbe Breite d. KI = Standardfehler + $\frac{1}{2}$ Wert des $1 - \frac{\alpha}{2}$ Quantils (hat bei Anteilswerten

Interpretation KI

$\rightarrow$ bei wiederholten SP-Ziehungen deckt d. KI mit seinen zufälligen Intervallsgrenzen d. wahren Parameter in $(1 - \alpha)\%$ aller Fälle ab

Breite des KI

$\rightarrow$ ist abhängig von: n, α, σ^2 in der GG (Bsp.: $2 \cdot z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$)

$\rightarrow$ ist grösser, je mehr Anteilswert in d. Mitte d. WB liegen! $\rightarrow$ max für: $\hat{\pi} = 0,5$

Hypothesentest

Def.

$\rightarrow$ Hypothese bezügl. eines Parameters in der GG, prüft anhand SP, ob d. empir. Befund für od. gegen d. Hypothese spricht

$\rightarrow H_0$ = "Wie gehabte Annahmen"

$\rightarrow H_1$ = Forschungshypothese = die zu untersuchende Hypothese

Allg. Vorgehen

L 1) Hypothese

a) $H_0: \pi = \pi_0; H_1: \pi \neq \pi_0 \rightarrow$ zweiseitig

b) $H_0: \pi = \pi_0$ oder $\pi \geq \pi_0; H_1: \pi < \pi_0 \rightarrow$ einseitig

Allg. gilt: $\pi = \mu$

c) $H_0: \pi = \pi_0$ oder $\pi \leq \pi_0; H_1: \pi > \pi_0 \rightarrow$ rechtsseitig

L 2) Signifik. Niveau

$\rightarrow \alpha$ = Wahrsch. dass eine Teststatistik bei Gültigkeit H_0 in Ablehnungsbereich fällt

$\rightarrow$ wenn Teststat. in Abl.bereich, ist es unwahrsch., dass SP aus Verteilung unter H_0 generiert wurde

L 3) Annahme-/Abl.bereich

$\rightarrow$ Bedingungen d. wahr od. unwahr sind, wenn H_0

Methode 1: verwerfen H_0 falls: a) $|z| > z_{1-\frac{\alpha}{2}}$ b) $z < z_{1-\alpha}$ c) $z > z_{1-\alpha}$

Methode 2: verwerfen H_0 falls: $P \leq \alpha \rightarrow$ signifikant

Methode 3: verwerfen H_0 falls Prüfgrösse ausserhalb des $(1 - \alpha)$ -KI liegt

L 4) Prüfgrösse

L $\rightarrow$

$\rightarrow$ falls: σ^2 bekannt: $\bar{X} - \mu \sim N(\mu, \sigma/\sqrt{n})$, $\bar{X} \sim N(0,1)$ od. beliebig verteilt mit $n \geq 30$

$\rightarrow$ Gauss (Z-Test) $\rightarrow$ st.normalvert. Tab.

$$\rightarrow Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$$

L $\rightarrow$

$\rightarrow$ falls: σ^2 unbekannt: $n < 30$ (für $n \geq 30$ ersetze $t_{(n-1)}$ durch $N(0,1)$)

$\rightarrow$ t-Test (od. approx. t-Test für $n \geq 30$) $\rightarrow$ t-Tab.

$$\rightarrow T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t_{(n-1)} \rightarrow \text{Varianz muss geschätzt werden: } S^2 = \sum (X_i - \bar{X})^2 / (n-1)$$

L $\rightarrow$

$\rightarrow$ Anteilswerte: ZSP 0/1, $\pi = P(X=1)$, $X \sim B(n, \pi)$, $X \sim N(n\pi, n\pi(1-\pi))$

$\rightarrow$ appr. Bin-Test (Z-Test) $\rightarrow$ t-Tab. falls: $n \cdot \pi > 5, n(1-\pi) > 5$

$$\rightarrow Z = \frac{X - n\pi}{\sqrt{n \cdot \pi(1-\pi)}} = \frac{\hat{\pi} - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} \sim N(0,1)$$

L $\rightarrow$

$\rightarrow$ Exakter Binomialtest (Berechnung d. Abl.bereich durch exakte Wahrsch. d. Binomialverteilung um H_0) $\rightarrow$ Bin.vert. Tab

$\rightarrow$ falls H_0 verworfen wird, heisst das: 1.) Unterbruch zw. geschätztem Wert aus d. SP u. d. H_0 -Parameter d. GG ist nicht nur wegen SP-Fehler

2.) Geschätz. Wert ist stat. signifikant von μ_0 unterschiedlich

$\rightarrow$ "Wir können nicht beweisen, dass eine Hypo. richtig ist. Wir können nur zeigen, dass sie (wahrscheinlich) falsch ist (Falsifizieren!)"

Fehler

Zutreffendes Ereignis	Testentscheidung	
	H_0 wird abgelehnt = Fehler 1. Art = α - Fehler	H_0 wird angenommen Richtiger Entscheid
H_0 ist korrekt		
H_0 ist falsch	Richtiger Entscheid	= Fehler 2. Art = β - Fehler

Gütefunktion=Macht

$$\rightarrow g(\mu) = 1 - \beta - \text{Fehler} = 1 - \Phi\left(z_{1-\frac{\alpha}{2}} - \frac{\mu - \mu_0}{\sigma} \sqrt{n}\right)$$

$\rightarrow$ =Wahrsch. die falsche H_0 zu verwerfen. "Konsistent" wenn Macht für $n \rightarrow \infty = 1$

$\rightarrow$ ein Test heisst "optimal", wenn er die grösste Macht besitzt (in einer Klasse!)

$\rightarrow$ Macht $\uparrow$: 1.) $n \uparrow$ 2.) je weiter wahres μ von H_0 abweicht 3.) $\alpha \uparrow$ (= je grösser α der Fehler 1. Art)

p-Wert

$$\rightarrow p = P_{H_0}(Z > z) = 1 - P_{H_0}(Z \leq z)$$

$\rightarrow$ lässt sich bei allen Tests anwenden

$\rightarrow$ bei p-Wert Methode ist entscheid. welche Prüfgrössenwerte noch stärker gegen H_0 sprechen $\rightarrow z$

$\rightarrow$ die Nullhypothese kann nicht abgelehnt werden wenn d. p-Wert grösser als der α -Wert ist

Signifikanztest

$\rightarrow$ gegeben falls: $P(H_1 \text{ annehmen} | H_0 \text{ wahr}) \leq \alpha$ d.h. $P(\text{Fehler 1. Art}) \leq \alpha$

$\rightarrow$ es wird dasjenige α (Signifikanzniveau) ausgewählt welches d. kleinste Wahrscheinlichkeit für d. Fehler 2. Art mit sich bringt!

Zweistichprobenfall von Hypothesentests

Allg.

→ H_0 : Homogenität / H_1 : Heterogenität

Vgl. v. Anteilswerten

$$\rightarrow Z = \frac{\hat{\pi}_1 - \hat{\pi}_2 - (\pi_1 - \pi_2)}{\sqrt{\hat{\pi} \cdot (1 - \hat{\pi}) \left(\frac{1}{n} + \frac{1}{m} \right)}} \sim N(0, 1) \text{ wobei } \hat{\pi} = \frac{n_1 \hat{\pi}_1 + n_2 \hat{\pi}_2}{n_1 + n_2}$$

($1 - \alpha$)-KI im ZSP Fall

Verteilungsfreie Tests

└ 1) χ^2 Anpassungstest

$$\rightarrow \left[\bar{X} - \bar{Y} \pm Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \right]$$

→ Vgl. zw. emp. u. hypothet. Verteilungen / Verteilungsmodell
→ H_0 : ZSP aus spez. Verteilung generiert / H_1 : ZSP nicht aus spez. Verteilung generiert

$$\rightarrow H_0: P(X=i) = \pi_i, \quad i=1, \dots, k$$

$$\rightarrow H_1: P(X=i) \neq \pi_i, \quad \text{für ein } i$$

$$\rightarrow \chi^2 = \frac{\sum (h_i - n\pi_i)^2}{n\pi_i} = \frac{\sum (\text{Durchschnitt} - \text{rel. Häufigkeit})^2}{\text{Durchschnitt}}$$

→ Bei voll spezif. Verteilung: approx. $\chi^2(k-1)$ -verteilt

Bei l geschätzten Parametern: approx. $\chi^2(k-1-l)$ -verteilt

└ 2) χ^2 Unabh. test

$$\rightarrow \chi^2 = \sum_{i=1}^k \sum_{j=1}^m \frac{(h_{ij} - \tilde{h}_{ij})^2}{\tilde{h}_{ij}}, \text{ wobei } \tilde{h}_{ij} = \frac{h_{i.} h_{.j}}{n}$$

$$\rightarrow H_0: \underbrace{P(x=i, y=j) = P(x=i) \cdot P(y=j)}_{\text{Unabhängigkeit}}, H_1: P(x=i, y=j) \neq P(x=i) \cdot P(y=j)$$

→ Verteilung unter H_0 : approx. $\chi^2((k-1)(m-1))$

Ablehnungsbereich: $\chi^2 > \chi^2_{1-\alpha}((k-1)(m-1))$

└ 3.) Vorzeichentest

$$\rightarrow H_0: x_{\text{med}} = \bar{d}_0, H_1: x_{\text{med}} \neq \bar{d}_0$$

→ A = Anzahl der X_i mit einem Wert $< \bar{d}_0 \rightarrow$ Unter H_0 gilt: $P(X_i < \bar{d}_0) = 0,5$

$$A \sim B(n, 0.5) \rightarrow H_0 \text{ wird verworfen wenn } A \text{ nahe bei } 0 \text{ od. } 1$$

Freiheitsgrade

$$\begin{aligned} \rightarrow \text{Gleichverteilung: } & k-1 \\ \rightarrow \text{Poissonverteilung: } & k-2 \\ \rightarrow \text{Binomialverteilung: } & k-2 \\ \rightarrow \text{Normalverteilung: } & k-3 \end{aligned}$$

Multivariate Statistik (Mehrdimensionale ZV's)

Def.

→ Einem Ergebnis eines Zufallsvorganges werden simultan mehrere Ergebnisse (reelle Zahlen) zugeordnet → mehrdimensionale Abbildungen!

Gemeinsame W.fkt.

$$\rightarrow f(x, y) = \begin{cases} P(X=x, Y=y) & \text{für } (x, y) \in \{(x_1, y_1), (x_2, y_2), \dots\} \\ 0 & \text{sonst} \end{cases}$$

→ =(gemeinsame) disk. Dichte / Verteilung

Randverteilungen

$$\left. \begin{aligned} -\text{von } x: \quad P(X=x_i) &= \pi_{i.} = \sum_{j=1}^m \pi_{ij} \\ -\text{von } y: \quad P(Y=y_j) &= \pi_{.j} = \sum_{i=1}^k \pi_{ij} \end{aligned} \right\}$$

Bedingte W.fkt.

$$-\text{von } X: \quad f_x(x|y) = \frac{f(x, y)}{f_y(y)} \quad \text{für festen Wert } y$$

$$-\text{von } Y: \quad f_y(y|x) = \frac{f(x, y)}{f_x(x)} \quad \text{für festen Wert } x$$

Gemeinsame Verteilungsfkt.

$$\rightarrow F(x, y) = P(X \leq x, Y \leq y) = \sum_{x_i \leq x} \sum_{y_j \leq y} f(x_i, y_j)$$

Gemeinsame stetige Verteilung und Dichte

$$\rightarrow P(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \int_c^d f(x, y) dy dx$$

Randdichte

$$-\text{von } X: \quad f_x(x) = \int_{-\infty}^{\infty} f(x, y) dy$$

$$-\text{von } Y: \quad f_y(y) = \int_{-\infty}^{\infty} f(x, y) dx$$

$$-\text{von } X|Y=y: \quad f_x(x|y) = \frac{f(x, y)}{f_y(y)} \quad \text{für festes } y \text{ mit } f_y(y) \neq 0$$

Bedingte Dichte

$$-\text{von } Y|X=x: \quad f_y(y|x) = \frac{f(x, y)}{f_x(x)} \quad \text{für festes } x \text{ mit } f_x(x) \neq 0$$

Gemeinsame Verteilungsfkt.

$$\rightarrow P(X \leq a, Y \leq b) = \int_{-\infty}^a \int_{-\infty}^b f(x, y) dy dx$$

Erwartungswert

$$\rightarrow E[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y) dx dy$$

Bedingter Erwart.wert

$$\rightarrow E(Y|x) = \int_{-\infty}^{\infty} y f(y|x) dy$$

Unabhängigkeit (v. 2 ZV's)

→ d. ZV's X u. Y heissen unabhängig, wenn für alle x und y gilt:

$$\rightarrow f(x, y) = f_x(x) \cdot f_y(y) \Leftrightarrow \pi_{ij} = \pi_{i.} \pi_{.j} \quad (\dots \text{ansonsten heissen } X \text{ und } Y \text{ abhängig!})$$

Unabhängigkeit (v. ZV's)

→ die ZV $X_1, \dots, X_n$ heissen unabhängig, wenn für alle $x_1, \dots, x_n$ gilt:

$$\rightarrow P(X_1 \leq x_1, \dots, X_n \leq x_n) = P(X_1 \leq x_1) \cdot \dots \cdot P(X_n \leq x_n)$$

→ äquivalent ist die Produktbedingung:

$$\rightarrow f(x_1, \dots, x_n) = f_{x_1}(x_1) \cdot \dots \cdot f_{x_n}(x_n)$$

Kovarianz

→ ist massstabsunabhängig = Erwartungswert d. ZV die selbst ein Prod. von ZV ist!

→ macht Aussagen ü. Z' häng. d. beiden ZV's (wenn=0 → kein Z' häng./Unabh.)

→ positiv: gleichsinn. Z' häng (pos. Steigung)/negativ: gegensinn. Z' häng (neg. St.)

└ Normalfall

$$\rightarrow \text{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))] = \frac{1}{n} \cdot \sum (x_i - \bar{x})(y_i - \bar{y})$$

└ bei stetiger ZV

$$\rightarrow \text{Cov}(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - E(X))(y - E(Y)) f(x, y) dx dy$$

└ Verschiebungssatz

$$\rightarrow \text{Cov}(X, Y) = E(XY) - E(X)E(Y)$$

└ Symmetrie

$$\rightarrow \text{Cov}(X, Y) = \text{Cov}(Y, X)$$

└ Lineare Transform.

Die Kovarianz der transformierten ZV's $\tilde{X} = a_x X + b_x$, $\tilde{Y} = a_y Y + b_y$ ist bestimmt durch :

$$\rightarrow \text{Cov}(\tilde{X}, \tilde{Y}) = a_x a_y \text{Cov}(X, Y) \rightarrow \text{Bsp.: Cov}(10x, y) = 10 \cdot \text{Cov}(x, y)$$

└ Eigenschaften

$$\begin{aligned} & \bullet \text{Cov}(X, Y) = \text{Cov}(Y, X) \\ & \bullet \text{Cov}(X, Y) = E(XY) - E(X)E(Y) \\ & \bullet \text{Cov}(aX + b, cY + d) = ac \text{Cov}(X, Y) \\ & \bullet \text{Cov}(X, X) = \text{Var}(X) \\ & \bullet \text{Cov}(X, Y) = 0 \Rightarrow \text{Unabhängigkeit!!!} \end{aligned}$$

E(X, Y)

$$-\text{bei stetig: } \iint x \cdot y \cdot f(x, y) dy dx$$

Korrelationskoeffizient

$$-\text{bei diskret: } \sum_i \sum_j x_i y_j f(x_i, y_j)$$

→ besser interpretierbar als Kovarianz (=geeignete Normierung!)

→ positiv: gleichsinn. Z' häng (pos. Steigung)/negativ: gegensinn. Z' häng (neg. St.)

$$\rightarrow \rho_{xy} = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)} \sqrt{\text{Var}(Y)}} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{\text{Cov}(X, Y)}{\sigma_x \sigma_y} \quad \text{wobei WB: } -1 \leq \rho(X, Y) \leq 1$$

└ Unkorreliertheit

$$\rightarrow \rho(X, Y) = 0 \rightarrow \text{Sind 2 ZV's unabhängig, so sind sie auch unkorreliert!}$$

└ Korreliertheit

$$\rightarrow \rho(X, Y) \neq 0$$

Varianz d. Summe zweier Zufallszahlen → $\text{Var}(X_1 + X_2) = \text{Var}(X_1) + \text{Var}(X_2) + 2\text{Cov}(X_1, X_2)$

Gewichtete Summe

$$\rightarrow X = a_1 X_1 + \dots + a_n X_n$$

└ Erwartungswert

$$\rightarrow E(X) = a_1 E(X_1) + \dots + a_n E(X_n)$$

└ Varianz

$$\rightarrow \text{Var}(aX_1 + bX_2) = a^2 \text{Var}(X_1) + b^2 \text{Var}(X_2) + 2ab \text{Cov}(X_1, X_2)$$

Bivariat. Dreieckverteil.

$$\rightarrow f(x, y) = x + y \quad \text{für } 0 \leq x \leq 1, 0 \leq y \leq 1 \rightarrow (\text{=normale Fläche im Raum!})$$

2-dim. Normalverteil.

$$\rightarrow \text{für ZV's } X \text{ und } Y \text{ gilt: } f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\mu_x}{\sigma_x} \right)^2 - 2\rho \left(\frac{x-\mu_x}{\sigma_x} \right) \left(\frac{y-\mu_y}{\sigma_y} \right) + \left(\frac{y-\mu_y}{\sigma_y} \right)^2 \right]}$$

→ Korrelation ρ : bestimmt d. Ausrichtung/Breite d. Bergrücken auf d. (x,y)-Ebene!

($|\rho|$ um 0: breiter Rücken / $|\rho|$ um 1: schmaler Rücken) → falls $|\rho| = 1$: Verteilung entartet!
...d.h. Bergrücken wird beliebig schmal!

→ Varianzen σ_x^2, σ_y^2 : bestimmen d. Höhe d. Bergrücken ($\sigma_x^2, \sigma_y^2 \uparrow \Rightarrow$ desto flacher!)

→ positiv: gleichsinn. Z' häng (pos. Steigung)/negativ: gegensinn. Z' häng (neg. St.)

→ Sind 2 ZV's unabhängig, so sind sie auch unkorreliert

Regressionsanalyse

Allg.

→ Deskript. Stat.-Fall = Beschreibung d. Z' hängs zweier Variablen für bestimmten Datensatz
→ Inferenzstat.-Fall = Punktschätzung für zwei Variablen der GG

Modell d. linearen Einfachregression

$$\rightarrow \text{Emp. Beziehung: } y_i = \alpha + \beta x_i + \epsilon_i, \quad i=1, \dots, n \quad (\text{wobei } \epsilon_i = \text{Fehler/Störvariabel})$$

→ ϵ_i wird nun neu als Realisierung von ZV aufgefasst (wobei vereinfachend angenommen wird): $\rightarrow E(\epsilon_i) = 0, i = 1, \dots, n$

L Stochast. Grundmodell
L Eigensch. d. Zielvariabel

→ $y_i = \alpha + \beta x_i + \epsilon_i, E(\epsilon_i) = 0, i = 1, \dots, n$
- $E(Y_i) = E(\alpha + \beta x_i + \epsilon_i) = \alpha + \beta x_i$
- $\text{Var}(Y_i) = \text{Var}(\alpha + \beta x_i + \epsilon_i) = \sigma^2$

Homoskedastizität
Heteroskedastizität
Normalverteil.annahme
Schätzen

→ =d. Eigenschaft gleicher Varianz σ^2 d. Fehlervariablen ϵ_i
→ es wird zugelassen, dass d. Varianzen ungleich sind
→ $\epsilon_i \sim N(0, \sigma^2)$ bzw. $Y_i \sim N(\alpha + \beta x_i, \sigma^2), i = 1, \dots, n$

L Ziel v. KQ-Prinzip

→ bestimme d. Schätzer $\hat{\alpha}$ und $\hat{\beta}$ so, dass: $\sum_{i=1}^n (Y_i - \alpha - \beta x_i)^2 \rightarrow \min_{\alpha, \beta}$ also die Summe d. quadr. Abweichungen durch $\hat{\alpha}, \hat{\beta}$ minimiert wird.

L Kleinste Quadr. Schätzer

→ $\hat{\beta} = \frac{\sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y})}{\sum_{i=1}^n (x_i - \bar{x})^2}, \hat{\alpha} = \bar{Y} - \hat{\beta} \bar{x}, \hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n \hat{\epsilon}_i^2 = \frac{1}{n-2} \sum_{i=1}^n (Y_i - \hat{\alpha} - \hat{\beta} x_i)^2, \dots$
mit den Residuen $\hat{\epsilon}_i = Y_i - \hat{Y}_i$ und den gefitteten Werten $\hat{Y}_i = \hat{\alpha} + \hat{\beta} x_i$

→ $\hat{\alpha}, \hat{\beta}, \hat{\sigma}^2$ sind erwartungstreue Schätzer: $E(\hat{\alpha}) = \alpha, E(\hat{\beta}) = \beta, E(\hat{\sigma}^2) = \sigma^2$,

$$\text{Var}(\hat{\alpha}) = \sigma_a^2 = \sigma^2 \frac{\sum x_i^2}{n \sum (x_i - \bar{x})^2} = \sigma^2 \frac{\sum x_i^2}{n \sum x_i^2 - n \bar{x}^2}, \text{Var}(\hat{\beta}) = \sigma_\beta^2 = \frac{\sigma^2}{\sum (x_i - \bar{x})^2} = \frac{\sigma^2}{\sum x_i^2 - n \bar{x}^2}$$

→ falls für $n \rightarrow \infty$ gilt: $\sum_{i=1}^n (x_i - \bar{x})^2 \rightarrow \infty \rightarrow \hat{\alpha}, \hat{\beta}, \hat{\sigma}^2$ sind auch noch konsistent!

→ Unter Normalverteilung gilt: $\frac{\hat{\alpha} - \alpha}{\sigma_a} \sim t(n-2), \frac{\hat{\beta} - \beta}{\sigma_\beta} \sim t(n-2)$ mit...

$$\hat{\sigma}_a = \hat{\sigma} \frac{\sqrt{\sum_{i=1}^n x_i^2}}{\sqrt{n \sum_{i=1}^n (x_i - \bar{x})^2}}, \hat{\sigma}_\beta = \frac{\hat{\sigma}}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

→ Konfidenzintervall: $\hat{\alpha} \pm \hat{\sigma}_a t_{1-\frac{\alpha}{2}}(n-2), \hat{\beta} \pm \hat{\sigma}_\beta t_{1-\frac{\alpha}{2}}(n-2)$

...für $n > 30$: t-Quantile d. $t(n-2)$ -Verteil. durch Quantile d. $N(0,1)$ -Verteil. ersetzen

Testen

→ $H_0: \beta = 0$ bedeutet, dass $Y_i = \alpha + \epsilon_i, i = 1, \dots, n$ gilt, u. somit X keinen Erklärungswert für Y besitzt. Ablehnung v. H_0 zugunsten v. H_1 bedeutet, dass es sinnvoll ist, X in Form d. systematisch. Komponente $\alpha + \beta x$ zur Erklärung v. Y einzusetzen
→ Teststatistiken und Ablehnungsbereiche:

L →

→ $T_{\alpha 0} = \frac{\hat{\alpha} - \alpha_0}{\sigma_a}$ bzw. $T_{\beta 0} = \frac{\hat{\beta} - \beta_0}{\sigma_\beta}$

→ $|T_{\alpha 0}| > t_{1-\frac{\alpha}{2}}(n-2)$ bzw. $|T_{\beta 0}| > t_{1-\frac{\alpha}{2}}(n-2)$

→ für $n > 30$: t-Quantile der $t(n-2)$ -Verteilung durch Quantile der $N(0,1)$ -Verteilung ersetzen

L →

→ F-Wert (zur Prüfung ob: $H_0: \beta = 0$):

$$F = \frac{R^2}{1-R^2}(n-2), F = T_{\beta 0}^2, F \sim F(1, n-2)$$

Prognose

→ $Y_0 = \alpha + \beta x_0 + \epsilon_0 \Rightarrow$ weil $E(\epsilon_0) = 0 \Rightarrow \hat{Y}_0 = \hat{\alpha} + \hat{\beta} x_0$

L →

→ Prognose für Konfidenzintervall von $\hat{Y}_0 = \hat{\alpha} + \hat{\beta} x_0$:

L →

→ $\hat{Y}_0 \pm t_{1-\frac{\alpha}{2}}(n-2) \hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n x_i^2 - n \bar{x}^2}}$

→ Konfidenzintervall für d. "wahre" Regressionsgerade $\alpha + \beta x$ (ohne Fehler ϵ_0):

→ $\hat{\alpha} + \hat{\beta} x \pm t_{1-\frac{\alpha}{2}}(n-2) \hat{\sigma} \sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum_{i=1}^n x_i^2 - n \bar{x}^2}}$

Multiple Lineare Regression

Def.

→ Mehrere Variablen ($X_1, \dots, X_p$ =sog. Regressoren) beeinflussen Y
→ Zielvariable=metrisch / Regressoren=metrisch, binär, mehrkategorial
→ $\text{Cov}(\mu_i, \mu_j) = 0 \rightarrow$ Störterme beeinflussen sich nicht
→ Fehlerterm (μ_i)=unbestimmbare Abstände zu Reg.gerade in GG (Autokorrelation: Störterm sind korreliert!)
→ Residuen (ϵ_i)=bestimmbare Abstände d. Stichprobenwerte in d. Stichprobe
→ keine lin. Abhängigkeit zw. d. Regressoren
→ exakter lin. Zusammenhang zw. d. Regressoren
→ $Y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \epsilon_i, E(\epsilon_i) = 0, i = 1, \dots, n$
→ $E(Y_i) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}$
→ $\text{Var}(Y_i) = \sigma^2$

keine Multikollinearität
Multikollinearität
Modell

L Erwartungswert

L Varianz

Schätzen

L Ziel v. KQ-Prinzip

→ bestimme d. Schätzer $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p$ so, dass: $\sum_{i=1}^n (Y_i - \beta_0 - \beta_1 x_{i1} - \dots - \beta_p x_{ip})^2 \rightarrow \min_{\beta_0, \beta_1, \dots, \beta_p}$
also d. Summe d. quadr. Abweichungen durch $\beta_0, \beta_1, \dots, \beta_p$ minimiert wird.
(Voraussetzungen: 1.) $n \geq p + 1$ / 2.) keine Multikollinearität)

L KQ- Schätzer

→ $\hat{\sigma}^2 = \frac{1}{n-p-1} \sum_{i=1}^n \hat{\epsilon}_i^2 = \frac{1}{n-p-1} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$

mit den Residuen $\hat{\epsilon}_i = Y_i - \hat{Y}_i$ und den gefitteten Werten $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots + \hat{\beta}_p x_{ip}$

L Varianz

→ $\sigma_j^2 = \text{Var}(\hat{\beta}_j), j = 0, \dots, p$

L Geschätzt. Standardabw.

→ $\hat{\sigma}_j = \sqrt{\text{Var}(\hat{\beta}_j)}, j = 0, \dots, p$

L Streuungszerlegung

→ $\sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^n (\hat{Y}_i - Y_i)^2$
SQT SQR SQE

L Bestimmtheitsmass

→ $R^2 = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2} = \frac{\text{SQR}}{\text{SQT}} = 1 - \frac{\text{SQE}}{\text{SQT}},$ wobei WB: $0 \leq R^2 \leq 1$
→ je näher bei 1 desto, besser wird d. Zielvariable durch d. Regression erklärt!

L Verteilung d. standard. Schätzer

→ $\frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}_j} \sim t(n-p-1), j = 0, \dots, p$

L Konfidenzintervall für β_j : → $\hat{\beta}_j \pm \hat{\sigma}_j t_{1-\frac{\alpha}{2}}(n-p-1), j = 0, \dots, p$

Testen

L Hypothesen

(a) $H_0: \beta_j = \beta_{0j}, H_1: \beta_j \neq \beta_{0j}$
(b) $H_0: \beta_j \geq \beta_{0j}, H_1: \beta_j < \beta_{0j}$
(c) $H_0: \beta_j \leq \beta_{0j}, H_1: \beta_j > \beta_{0j}$

→ falls: $H_0: \beta_j = 0, H_1: \beta_j \neq 0 \rightarrow$ wenn H_0 zutrifft, hat d. Regressor X_j keinen signifikanten Erklärungswert für Y u. kann aus dem Regressionssatz entfernt werden!

L Teststatistiken

→ $T_j = \frac{\hat{\beta}_j - \beta_{0j}}{\hat{\sigma}_j}, j = 0, \dots, p \rightarrow H_0$ ablehnen falls: (b) $T_j < -t_{1-\alpha}(n-p-1)$
(c) $T_j > t_{1-\alpha}(n-p-1)$

L Overall F-Test

→ Frage: "Tragen X überhaupt zur Erklärung von Y bei?"
- Hypothesen: $H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0, H_1: \beta_j \neq 0$ für mind. ein j

- Teststatistik: $F = \frac{R^2}{1-R^2} \cdot \frac{n-p-1}{p} = \frac{\text{SQE}}{\text{SQR}} \cdot \frac{n-p-1}{p}$

- Unter H_0 gilt: $F \sim F(p, n-p-1)$

- H_0 ablehnen, falls gilt: $F > F_{1-\alpha}(p, n-p-1)$

Prognose

L Prognosewert

→ $\hat{Y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_{01} + \dots + \hat{\beta}_p x_{0p} \rightarrow$ für Schätzfehler: $Y_0 - \hat{Y}_0$ gilt: $E(Y_0 - \hat{Y}_0) = 0$

L Varianz

→ $\text{Var}(Y_0 - \hat{Y}_0) = \hat{\sigma}_{v_0} = \sqrt{\text{Var}(Y_0 - \hat{Y}_0)}$

L $(1 - \alpha)$ -Prognoseintervall für Y_0 gegeben $x_0 \rightarrow \hat{Y}_0 \pm \hat{\sigma}_{v_0} t_{1-\frac{\alpha}{2}}(n-p-1)$