

1. Deskriptive Statistik

Bezeichnung	kurze Beschreibung	§	Literatur
α -Quantil	$q_\alpha = \{ x_{\text{int}(\alpha^*n)+1} \text{ für } \alpha^*n \notin \mathbb{N} \text{ (nat. Zahlen)}$ $\{ 0.5 (x_{\alpha^*n} + x_{\alpha(n+1)}) \text{ für } \alpha^*n \in \mathbb{N} \text{ (nat. Zahlen)}$ bei klassierten Daten: $q_\alpha = x_n + (x_0 - x_n) [(\alpha - F(x_n)) / (F(x_0) - F(x_n))]$	5	S.14, Bol 66, A9,16
χ^2 - Distanz	Abstand zur Unabhängigkeit der Merkmale: $\chi^2 = d^2(M_1, M_2) = n \left[\sum_{j=1}^r \sum_{k=1}^s (F_{jk} - f_j f^k)^2 / (f_j f^k) \right] = \sum_{j=1}^r \sum_{k=1}^s \frac{(n_{jk} \frac{n_j n^k}{n})^2}{\frac{n_j n^k}{n}}$ $\chi^2 = 0$ für Unabhängigkeit, sonst $\chi^2 > 0$	8	S.27, Bol 125 A21
absolute Häufigkeit	abs. Anz. best. Merkmalsausprägungen in einer Urliste: n_k	2	S.7
arithmetisches Mittel	=emp. Mittelwert, $\bar{x} = 1/n \sum x_i$; Ausreißer empfindlich bei klassierten Daten: $\bar{x} = 1/n \sum \bar{x}_k n_k = \sum \bar{x}_k f_k$ (siehe A9d)	5	S.13, Bol 71, A8,10
Balkendiagramm	=Säulendiagramm=Balkendiagramm	4	Bol 45
bedingte Verteilungen	Wahrscheinlichkeit von a_1 unter der Nebenbedingung b_1 : $f(a_1/b_1) = f'_{jk} = f_{jk} / f_j$	8	A21
Bestandsmasse	Masse zu Zeitpunkt t_0 , \leftrightarrow Ereignismasse	1	Bol 13
Bestimmtheitsmaß	(siehe Korrelationskoeffizient) $r^2 = \text{corr}(x,y) / (\text{var}(x) * \text{var}(y)) = \text{var}(\$) / \text{var}(y)$ $r^2 = \cos^2 \alpha = \text{corr}^2(x,y) \quad \epsilon [0,1]$	9	A26c
Bestimmtheitsmaß	$r^2 = a^2 \text{var}(x) / \text{var}(y) = \text{var}(\$) / \text{var}(y)$ (siehe Korrelationskoeffizient $r = \sqrt{r^2}$)	9	A26c
bimodale Verteilung	hat zwei Gipfel, Modalwerte, \leftrightarrow unimodal, \leftrightarrow plurimodal	1	S.15
Boxplot	Box von $q_1 - q_3$, Strich bei $x_m = q_2$, Fühler bis kleinstem/größtem Wert von $q_{1/3} \pm 1.5 (q_3 - q_1)$, alle Außenpunkte mit x einzeichnen.	4	S. 19, Bol 88, A15-17
Codierung	=Skalierung, qual/quant. Merkmale z.B. verh=1, ledig=2	2	Bol 23
dichotom	binärkodiert	2	A 3
disjunktiv	Merkmal ist nicht häufbar, Individuum kann nur eine Ausprägung annehmen	2	S.2
diskret	(nur quant. Merkmale) $m \subseteq \mathbb{Q}$, aber auch Noten 1.3;1.7;2.0;..., \leftrightarrow stetig	2	S.5,A 1, Bol 22

Distanz	$\chi^2 = d^2 (M_1, M_2) = n [\sum_j \sum_k (F_{jk} - f_j f^k)^2 / (f_j f^k)]$ $\chi^2 = 0$ für Unabhängigkeit, sonst $\chi^2 > 0$ (siehe χ^2 - Distanz !)	8	S.27, Bol 125 A21																				
Empirische Verteilungsfunktion	$F(x) = 1/n \{ i \in N \mid x_i \leq x \} $ Eigenschaften: $F(n)=1, F(0)=0$, monoton steigende Treppenfkt., Treppenhöhe=1/n bei einfachen Beobachtungen, rechtsseitig halbstetig	4	S.9, Bol 40, A8																				
empirischer Mittelwert	$\bar{x} = 1/n \sum x_i$ (→ siehe arith. Mittel)	5	S.13, Bol 71, A8,10																				
Ereignismasse	Masse in Zeitraum (z.B. Zugang, Abgang), \leftrightarrow Bestandsmasse	1	Bol 13																				
exhaustiv	Merkmal ist erschöpfend, alle Merkmalsausprägungen werden erfaßt (sonstige)	2	S.2, A 2																				
Flächendiagramm	rel./abs.	4	Bol 43																				
Gini - Koeffizient	$G = F_{\text{Lorenzkurve}} / F_{\text{Dreieck}} = (2 * (\sum i * x_i) - (n+1) (\sum x_i)) / (n (\sum x_i))$	7	22,A18f,B 101																				
häufbares Merkmal	Mehrfachnennungen sind möglich	2	Bol 17																				
Häufigkeitsdichte	$n'_k = n_k / l$	2	Bol 55																				
Häufigkeitspolygon	Verbindung der Mittelpunkte im Histogramm	4	Bol 56																				
Histogramm	Höhe d. Balkens: abs: $n'_k = n_k / l$ (l = Klassenbreite); rel: $f'_k = f_k / l$ sonst Stabdiagramm, Vorsicht: Intervallgrenzen, nur für quant. Merkmal	4	Bol 53, A5,7,9																				
Histogramm (2D)	Volumen = Häufigkeit, Höhe = Häufigkeitsdichte = $n'_{jk} = n_{jk} / l_j l^k$ (Klassenbreiten) 3dimensionale Darstellung, nur für quant. Merkmale	8	A24																				
Kartogramm		4	Bol 49																				
Klassenmitte	$z_i = (\alpha_i + \beta_i) / 2$	1	S.6																				
Klassierung	Zusammenfassung quant. Merkmalsausprägungen zu Klassen/Intervallen	1	S.5, Bol 32																				
Komponentenstabdiagramm	nur EIN Balken	4	S. 8																				
Kontingenzkoeffizienten	<u>Pearson</u> : $\Phi^2 = \chi^2 / n$ <u>Cramer</u> : $C^2 = \chi^2 / (n + \chi^2)$ <u>korrigierter Cramer</u> : $C^* = C \sqrt{k / (k-1)}$	8	S.28, Bol 125, A21, 22																				
Kontingenztafel	<table border="1" style="display: inline-table; border-collapse: collapse;"> <tr> <td></td> <td>b_1</td> <td>b_2</td> <td>...</td> <td></td> </tr> <tr> <td>a_1</td> <td>$n_{b1,a1}$</td> <td>...</td> <td>...</td> <td>n^{a1}</td> </tr> <tr> <td>a_2</td> <td>...</td> <td>...</td> <td>...</td> <td>n^{a2}</td> </tr> <tr> <td>...</td> <td>n_{b1}</td> <td>n_{b2}</td> <td>...</td> <td>n</td> </tr> </table> <p style="text-align: center;">nur für qual. Merkmale, sonst. Korrelationstabelle 1.Merkmal: $j=\{1...r\}$ (Zeile); 2.Merkmal: $k=\{1...s\}$ (Spalte)</p>		b_1	b_2	...		a_1	$n_{b1,a1}$	n^{a1}	a_2	n^{a2}	...	n_{b1}	n_{b2}	...	n	8	Bol 115,110, A21
	b_1	b_2	...																				
a_1	$n_{b1,a1}$	n^{a1}																			
a_2	n^{a2}																			
...	n_{b1}	n_{b2}	...	n																			

Korrelationskoeffizient	$r = \text{corr}(x,y) = \text{cov}(x,y) / (s_x s_y) = a s_x / s_y$, siehe auch Rangkorrelationskoeffizient alle Punkte auf steigender [fallender] Geraden $\Leftrightarrow \text{corr} = 1$ [-1]	9	A25,29
Korrelationstabelle	Kontingenztafel für quantitative Merkmale	8	A24
korreliert	2 Merkmale sind vollständig korreliert, wenn $\text{corr}(x,y) = \cos \alpha = \pm 1$, 2 Merkmale sind unkorreliert, wenn $\text{corr}(x,y) = \cos \alpha = 0$	9	S.32
Kovarianz	$\text{cov}(x,y) = 1/n (\sum (x_i - \bar{x})(y_i - \bar{y})) = 1/n (\sum x_i y_i) - \bar{x}\bar{y}$	8	S. 31, Bol 137
Kreisdiagramm	= Kreissektorendiagramm, nur rel. Hfgkten	4	Bol 48, A4
Lineare Regression	Lösung von $\sum u_i^2$ mit $u_i = \text{Residuen}$; $\hat{y} = a x + b$ $a = \text{cov}(x,y) / \text{var}(x) = [1/n (\sum x_i y_i) - \bar{x}\bar{y}] / [1/n (\sum x_i^2) - \bar{x}^2]$, $b = \bar{y} - a \bar{x}$	9	S.36, A26, Bol 132
Liniendiagramm	rel/abs. Hfgkten. nur Linien	4	Bol 44
Lorenzkurve	$L(k) = (\sum^k x_i) / (\sum^n x_i)$ aus klassierten Daten: $\sum^k \tilde{x}_k / \sum^n \tilde{x}_k$ mit \tilde{x} = Repräsentant der Klasse	7	S. 20,A 18,19
Median	$x_m = \{$ für $n = 2s+1$ (ungerade): $x_m = x_{s+1}$ $\{$ für $n = 2s$ (gerade): $x_m = 0.5 (x_s + x_{s+1})$ bei klassierten Daten: aus Summenhäufigkeitsfunktion (\rightarrow siehe α - Quantile)	5	S.13, Bol 65, A10
Mengenindizes	Paasche: $MP_t = (\sum q_i^t p_i^t) / (\sum q_i^0 p_i^t)$ Laspeyres: $ML_t = (\sum q_i^t p_i^0) / (\sum q_i^0 p_i^0)$ Marshall-Edgeworth: $MME_t = (\sum (p_i^0 + p_i^t) q_i^t) / (\sum (p_i^0 + p_i^t) q_i^0)$ Fisher's idealer Mengenindex: $MF_t = \sqrt{(ML_t * MP_t)}$ (geometr. Mittel aus Lasp./Paasche)	a	S.43, Bol 171
Merkmal	Beschr. der stat. Einheit (z.B. Geschlecht); \rightarrow Rang-, quantitative oder qualitative Merkmale	2	S1,Bol 15, A1
Merkmalsausprägung	Eigenschaft des Merkmals (z.B. männlich, weiblich)	2	S.1, Bol 16
mittlerer abs. Abstand zu \bar{x}	$d_m = 1/n \sum x_i - \bar{x} $	6	S.16, Bol 77
mittlerer abs. Abstand zu x_m	$d_m = 1/n \sum x_i - x_m $	6	S. 16, Bol 77
Modus, Modalwert	$x_{\text{mod}} = \max \{ x_i \}$, häufigster Beobachtungswert, mehrdeutig	5	S.15, Bol 64, A8, A10
Nichtlineare Regression	Linearisierung durch Transformation: Logarithmisieren (\rightarrow S. 42)	9	S. 42
offene Randklassen	Repräsentant der Klasse frei (aber sinnvoll) wählbar	1	Bol 73
Paretokurve	Gegenstück zur Lorenzkurve	7	S. 20
Piktogramm	nur abs. Hfgkten, mit Bildern(Ikonen)	4	Bol 47, A4
plurimodale Verteilung	hat mehr als zwei Gipfel, Modalwerte, \leftrightarrow unimodal, \leftrightarrow bimodal	1	S.15

Population	=stat. Masse, Summe d. stat. Einheiten	1	S.1, Bol 10
Preisindizes	Paasche: $PP_t = (\sum q_i^t p_i^t) / (\sum q_i^t p_i^0)$ Laspeyres: $PL_t = (\sum q_i^0 p_i^t) / (\sum q_i^0 p_i^0)$ Marshall-Edgeworth: $PME_t = (\sum (q_i^0 + q_i^t) p_i^t) / (\sum (q_i^0 + q_i^t) p_i^0)$ Fisher's idealer Preisindex: $PF_t = \sqrt{(PL_t * PP_t)}$ (geometr. Mittel aus Lasp./Paasche)	a	S..43, Bol 169 A33,34
qualitatives Merkmal	m verbal ohne Wertung (z.B. Haarfarbe, Religion), \leftrightarrow quantitatives Merkmal, Rangmerkmal evtl pseudoquantitatives Merkmal (z.B. PLZ)	2	Bol 21, A1c.
Quantil	$q_\alpha = \{ x_{\text{int}(\alpha*n)+1} \text{ für } \alpha*n \notin \mathbb{N} \text{ (nat.Zahlen)} \rightarrow \text{siehe } \alpha \text{- Quantil}$ $\{ 0.5 (x_{\alpha*n} + x_{\alpha(n+1)}) \text{ für } \alpha*n \in \mathbb{N} \text{ (nat. Zahlen)}$	5	S.14, Bol 66, A9,16
quantitatives Merkmal	$m \subseteq \mathbb{R}$, (z.B. Alter, Temperatur), \leftrightarrow qualitatives Merkmal, Rangmerkmal	2	Bol 22, A1c.
Quartilsabstand	QA= $q_3 - q_1$	6	S.16, Bol 76, A10
Randhäufigkeit	$n_j = \sum^s n_{jk}$; $n^k = \sum^r n_{jk}$	8	A21
Rangkorrelationskoeffizient	(nach Spearman): $\text{corr} = 1 - (\sum (r_i - s_i)) / (n(n^2 - 1))$ nur bei zwei Merkmalen; r_i und s_i sind Spaltenvektoren der Ränge der beiden Merkmale, Werte müssen nach Rängen geordnet werden	8	S. 34,A25
Rangmerkmal	Einteilung in Qualitätsstufen, Unterschied nicht quantifizierbar, \leftrightarrow quantitatives/quant. Merkm.	2	Bol 21, A1c.
relative Häufigkeit	abs.Hfgkt durch n: $f_k = n_k/n$	2	S. 8, Bol 30
relative prozentuale Häufigkeit	rel. Hfgkt mal 100: $f_k * 100\%$	2	S.8, Bol 30
Residuen	Abstände bei der lin. Regression, zw. Gerade und tats. Werten, parallel y-Achse es gilt immer $\sum^n u_i = 0$	9	A28
Resthäufigkeit	$R = n - F(x)$ [emp.Verteilungsfunktion] = $n - \int_{-\infty}^x n'(z) dz$	4	
Schiefe	$S = [x_m - \frac{1}{2}(q_1 + q_3)] / [\frac{1}{2}(q_3 - q_1)]$ $S = -1$ oder $x_m < \bar{x} \Leftrightarrow$ größte Rechtsschiefe (Gipfel links); $S=0$ sym.	A	Klaus.7/94 3d
Skalen	Nominalskala: für qual. Merkmale Ordinär/Rangskala: Rangmerkmale Kardinal/metr. Skala: quant. Merkmale Intervallskala: Nullpkt. willkürlich (z.B. °Celsius, °Fahrenheit) Verhältnisskala: Null fest, Abstand fest (z.B. Alter) Absolutskala: Null fest, Abstand fest (z.B. Semesterzahl)	1	S.5, Bol 23

Skalierung	= Codierung, qual/quant. Merkmale z.B. verh=1, ledig=2	2	Bol 23
Spannweite	Spannweite = $\max \{ x_i \} - \min \{ x_i \}$	6	S.16 Bol 75
Stabdiagramm	=Balkendiagramm, Flächen, rel./abs., siehe auch Histogramm	4	Bol 45
Stabdiagramm (2D)	wie Strichdiagramm nur Balken/Volumen	8	A21
Standardabweichung	$s_x = \sqrt{(\text{var}(x))}$	6	S.16, Bol 79, A12,14
Standardisierung	$z_i = (x_i - \bar{x}) / s_x$ (siehe Standardabweichung)	2	S.17
stetig	(nur quant. Merkmale) auch kontinuierlich, $m \subseteq \mathbb{R}$ oder Teilintervall z.B. Größe	2	S.5, A 1, Bol 22
Streudiagramm	1.Achse: Merkmal 1; 2.Achse: Merkmal 2; Punktwolke	8	S.29, Bol 114
Strichdiagramm (2D)	wie eindim. Strichdiagramm nur 2 Merkmale; 3D Zeichnung	8	Bol 113
Summenhäufigkeitsfunktion	=kumuliertes Häufigkeitspolynom, aus klass. Daten, stetig zeichnen abs.: $N^*(x) = \int_{-\infty}^x n^*(z) dz$; rel.: $F^*(x) = \int_{-\infty}^x f^*(z)$	4	S.12, Bol 59, A9
Trajektorie	Vektor aus p Beob.zeitpunkten bzgl. eines Merkmals	1	A 2
Transformation	siehe auch Standardisierung	1	A13
Unabhängigkeit von Merkmalen	für alle i,j muß gelten: $f_{jk} = f_j * f^k$ oder Bol: $p(a,b) = p(a) * p(b)$ (rel.) für alle i,j muß gelten: $n_{jk} = n_j * n^k / n$ $\chi^2 = 0$ (siehe Distanz) oder $\text{cov}(x,y) = 0$ (siehe Kovarianz)	8	S. 27, Bol 120
unimodale Verteilung	hat nur einen Gipfel bzw. Modalwert, \leftrightarrow bimodal, \leftrightarrow plurimodal	1	S.15
unkorreliert	2 Merkmale sind vollständig korreliert, wenn $\text{corr}(x,y) = \cos \alpha = \pm 1$, 2 Merkmale sind unkorreliert, wenn $\text{corr}(x,y) = \cos \alpha = 0$	9	S.32
Urliste	=stat. Reihe, enthält nur Merkmalswerte	1	Bol 27
Varianz	$\text{var}(x) = s_x^2 = 1/n (\sum x_i^2) - \bar{x}^2$ Umformung: $\text{var}(ax-b) = a^2 \text{var}(x)$ bei klassierten Daten: $\text{var}(x) = \sum^k (f_k \bar{x}_k^2) - \bar{x}^2$ (\rightarrow A11)	6	S.16, Bol 79,84, A12,14
Varianzverhältnis	Varianzverhältnis aus Bestimmtheitsmaß: $r^2 / (1 - r^2) = \cotan^2 \alpha = \text{var}(\$) / \text{var}(u)$	9	S.41
Variationskoeffizient	$\text{varkof} = s_x / \bar{x}$	6	S.16

Hinweise: **S. xx** bezeichnet die Seite in meiner Mitschrift zu Statistik I (SS 95), **Bol xx** die Seite in Bol, Deskriptive Statistik, 2.Aufl., Oldenbourg Verlag, **A xx** die Aufgabe zu Statistik I (SS 95), die Spalte § bezeichnet das Thema des Stichwortes (eigene Einteilung)

2. Wahrscheinlichkeitstheorie

Bezeichnung	kurze Beschreibung	Literatur
σ - Algebra	→ Grundgesamtheit mit folgenden Eigenschaften: 1. $\Omega \in \mathcal{A}(\Omega)$ 2. $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}(\Omega)$ 3. $A_i \in \mathcal{A}(\Omega) \Rightarrow \cup_i A_i \in \mathcal{A}(\Omega)$	S45, A36, bol 14
χ_n^2 - Verteilung (mit n Freiheitsgraden)	<u>Dichte</u> : $\varphi_n(x) = \frac{1}{2^{n/2} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}$ mit $\Gamma(n) = (n-1)!$ und $\Gamma(n/2) = 1/\sqrt{\pi}$ <u>Momente</u> : $E(x) = n$; $\text{Var}(x) = 2n$ <u>Annäherung durch $N(n, 2n)$</u> : $\Leftrightarrow n > 30$	a48
χ_1^2 - Verteilung	<u>Dichte</u> : $\varphi_1(x) = 1/\sqrt{2\pi} x^{-1/2} e^{-x/2}$ <u>Momente</u> : $\mu = 1$, $\sigma^2 = 2$	a48
bedingte Dichten	$\varphi(y x) = \varphi(x,y) / \varphi_1(x)$, $\varphi(x y) = \varphi(x,y) / \varphi_2(y)$	a15, a17, a18, bol 133
bedingte Wahrscheinlichkeiten	$P(A B) = P(A \cap B) / P(B) [= P_B(A \cap B)]$; A,B unabhängig $\Leftrightarrow P(A \cap B) = P(A) P(B)$ Satz der totalen Wkt.: $P(B) = \sum_i P(B \cap A_i) = \sum P(B A_i) P(A_i)$ (nur disjunkte Teilmengen!) Satz von Bayes: $P(A_i B) = [P(B A_i) P(A_i)] / [\sum_i P(B A_i) P(A_i)]$ (disj. Teilm.!)	bol 100/108, a7, a8, a9, a10
Binomialverteilung: Bin(n,p)	<u>Masse</u> : $\alpha_k = \binom{n}{k} p^k (1-p)^{n-k}$ ist Ziehung von n Kugeln <u>mit Zurücklegen</u> (vgl. Multinomialvtlg.) <u>Momente</u> : $\mu = np$; $\sigma^2 = np(1-p)$ <u>Faltung</u> : $\text{Bin}(m,p) * \text{Bin}(n,p) = \text{Bin}(n+m,p)$ <u>Transformation</u> : $f: k \rightarrow k/n$; $\text{Bin}_0(n,p)$ auf $[0,1]$ mit $\text{Bin}_0(n,p) \xrightarrow{n \rightarrow \infty} N(p, p(1-p)/n)$ Momente von Bin_0 : $\mu = p$; $\sigma^2 = p(1-p)/n$ (Stetigkeitskorr.!) <u>Annäherung durch $N(np, np(1-p))$</u> $\Leftrightarrow p \approx 0.5$ oder $np(1-p) > 9$ (→ ZGS) (Stetigkeitskorr.!!!)	s2, bol 41, a2, a23 mom.: a29,a37 faltung: a39 transf.: a22 Münze: a47 ZGS: bol 192
Borel'sche Algebra B	$\Omega \equiv \mathbb{R}$; B ist durch S (= Menge aller offenen Intervalle) erzeugte σ - Algebra	S47, A37, bol 17

Dichtefunktion	Eigenschaften: $\int_{-\infty}^{\infty} \varphi(x) dx = 1$ und $\varphi(x) \geq 0 \forall x$	s3, a6
Dirac - Maß	$\varpi \in \Omega$ fest, $\{\varpi\} \in \mathcal{A}$ mit Gesamtmasse im Punkt ϖ	S51
Dreiecksverteilung	$\varphi(x) = (a + x) / a^2$ für $-a \leq x \leq a$ \wedge 0 sonst	a3, a14, bol 52
Erwartungswert: E(x)	kontin.: $E(x) = \mu = \int x \varphi(x) dx$; nach Trafo mit $f(x)$: $E(x) = \int f(x) \varphi(x) dx$ diskret: $E(x) = \mu = \sum x_k \alpha_k$; nach Trafo mit $f(x)$: $E(x) = \sum f(x_k) \alpha_k$ $E(\alpha x + \beta) = \alpha E(x) + \beta$; $E(x+y) = E(x) + E(y)$	a29, trafo:bol 92
Erlang-Verteilung	$\varphi_n(x) = (\lambda^n / (n-1)!) x^{n-1} e^{-\lambda x}$ für $x > 0$, <u>Momente</u> : $\mu = n / \lambda$; $\sigma^2 = n / \lambda^2$ ($\rightarrow \text{Exp}(\lambda)$)	ind 81, a41
Erzeuger	$\mathbf{S} \subset \wp(\Omega)$; Die größte σ -Algebra, die \mathbf{S} enthält, heißt von \mathbf{S} erzeugte σ -Algebra $\mathbf{S} \subset \mathcal{A}_s$	S46f
Exponentialverteilung: $\text{Exp}(\lambda)$	<u>Dichte/Vtlg.</u> : $\varphi(x) = \lambda e^{-\lambda x} \forall x > 0$, $F(a) = 1 - e^{-\lambda a} \forall a > 0$ Verteilung ohne Gedächtnis, $F(x)$ ist Wahrscheinlichkeit, daß es bis x „kaputt geht“ <u>Momente</u> : $E(x) = 1 / \lambda$; $\text{Var}(x) = 1 / \lambda^2$ <u>Faltung</u> : $\text{Exp}(\lambda_1) * \text{Exp}(\lambda_2) \neq \text{Exp}(\lambda_1 + \lambda_2)$ $\text{Exp}(\lambda) * \text{Exp}(\lambda)$ ist @ Erlang Verteilung mit $\mu = 2 / \lambda$ und $\sigma^2 = 2 / \lambda^2$	S54, bol 54, s3, a5, a11, a15, a16, a28, trafo: a25, faltung: a41
Faltung	kontin.: $\psi(z) = \int_{x=0..z} \varphi_A(x) \varphi_B(z-x) dx$ diskret: $\gamma_l = \sum_{j=0}^l \alpha_j \beta_{l-j}$ es gilt: $E(x_1+x_2) = E(x_1) + E(x_2)$ $\text{Var}(x_1+x_2) = \text{Var}(x_1) + \text{Var}(x_2) + 2 \text{Cov}(x_1, x_2)$ mit $\text{Cov}(x_1, x_2) = 0$ bei Unabhängigkeit $\text{Cov}(x_1 + x_2 + \dots + x_n, y) = \text{Cov}(x_1, y) + \text{Cov}(x_2, y) + \dots + \text{Cov}(x_n, y)$ (s. Kovarianz)	a39-42 E(x): bol 158 Var: bol 160
Geometrische Vtlg.	<u>Masse</u> : $\alpha_k = (1-p) p^k$; Wkt. für k-mal Funktion und Fehlfunktion in k+1; p ist Anteil der Guten <u>Momente</u> : $\mu = 1 / (1-p)$; $\sigma^2 = p / (1-p)^2$; x_{Mod} bei $k=0$ für $p \rightarrow 0$: Entartung zum Dirac-Maß: $\alpha_k = 1$ für $k=0$, sonst $\alpha_k = 0$	a31
gestutzte Verteilung auf [a,b]	$\varphi_0(x) = \varphi(x) / \int_a^b \varphi(x) dx$	a11, 12
Gleichverteilung: Gl (a, b)	diskret: $\alpha_k = \text{const} = 1/n$ (Laplace'sche Gl.vtlg); $P(A) = (\#A) / (\#\Omega)$ $E(x) = 0.5 (m + 1)$; $\text{Var}(x) = (1/12) (n^2 - 1)$	S51, bol 21/51 s3, a3 a20, a30

	<p>stetig: $\varphi(x) = \frac{1}{b-a} \mathbb{1}_{[a,b]}(x)$; $E(x) = (b-a)/2$; $\text{Var}(x) = (1/12)(b-a)^2$</p> <p><u>Faltung</u>: $\text{Gl}(0,1) * \text{Gl}(0,1)$ ist Dreiecksverteilung (0,2)</p>	mehrdim: bol 119
Grundgesamtheit	$\Omega = \text{Grundgesamtheit, Population} \neq \emptyset$ (\rightarrow Wahrscheinlichkeitsraum)	S45
Halbkreisverteilung		bol 54
Hypergeometrische Vtlg.: $H(N,M,n)$	<p><u>Masse</u>: $\alpha_k = P(X=k) = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}$ (Ziehen ohne Zurücklegen); $p = M/N$</p> <p>N: Zahl aller Kugel, M: Zahl der guten, n: Stichprobenumfang, k: gezogenen guten Kugeln</p> <p><u>Momente</u>: $\mu_H = n p = n M/N$; $\sigma_H^2 = n p (1-p) (N-n)/(N-1) \xrightarrow{N \rightarrow \infty} \sigma^2_{\text{Bin}}$</p> <p><u>Annäherung durch</u>: $\text{Bin}(n,p) \Leftrightarrow n \leq 0,05 N$ $N(\mu_H, \sigma_H^2) \Leftrightarrow \sigma_H^2 > 9$ oder $n > 30$ (Vorsicht: Stetigkeitskorrektur) $\text{Poi}(n p) \Leftrightarrow n > 50$ und $p < 0,1$</p>	bol 38, a29, näherung: a50
Korrelationskoeffizient	$\rho(x,y) = \text{Cov}(x,y) / (\sigma_x \sigma_y) \in [-1; 1]$ invariant gegen lineare Transformation mit $\alpha > 0$	bol 150
Kovarianz	<p>$\text{Cov}(x,y) = E(x y) - E(x) E(y)$; $\text{Cov}(\alpha x + \beta, \gamma y + \delta) = \alpha \gamma \text{Cov}(x,y)$</p> <p><u>Faltung</u>: $\text{Cov}(x_1 + x_2 + \dots + x_n, y) = \text{Cov}(x_1, y) + \text{Cov}(x_2, y) + \dots + \text{Cov}(x_n, y)$</p> <p>$x,y$ unabh. $\Rightarrow \text{Cov}(x,y) = 0$ ($\text{Cov}(x,y) = 0 \nRightarrow x,y$ unabh.) [s. Unabhängigkeit]</p>	bol 149
Lognormalvertlg.: $\text{Log } N(\mu, \sigma^2)$	<p>Erzeugung: $N(\mu, \sigma^2) \xrightarrow{f: x \mapsto e^x} \text{Log } N(\mu, \sigma^2)$</p> <p><u>Dichte</u>: $\varphi^f(y) = df^{-1}(y)/dy \varphi(\ln y) = \frac{1}{y} \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(\ln(y)-\mu)^2}{2\sigma^2}}$ für $y > 0$</p> <p><u>Momente</u>: $E(y) = \exp(\mu + 0.5 \sigma^2)$; $\text{Var}(y) = \exp(2\mu + \sigma^2)(\exp(\sigma^2)-1)$; $x_M = \exp(\mu)$; $x_{\text{Mod}} = \exp(\mu - \sigma^2)$</p>	a27, a34
Mathe Tricks	geom. Reihen: $\sum^n k = 0.5 n (n+1)$; $\sum^n k^2 = 1/6 n (n+1) (2n+1)$; $\sum^\infty p^k = 1 / (1-p)$	a31
Median	$P(X \geq x_M) \geq 1/2 \Leftrightarrow F(x_M) = 1/2$	bol 61
Mengensystem	$\mathcal{A}(\Omega) = \text{Mengensystem von Teilmengen der Grundgesamtheit}$ ($\rightarrow \sigma$ - Algebra)	S45
Meßbarkeit	<p>$f: \Omega \rightarrow E$ heißt A/B meßbar $\Leftrightarrow f^{-1}(B) \in \mathcal{A} \quad \forall B \subset E$</p> <p>$\wp(\Omega)$ und alle stetigen Funktionen sind immer meßbar</p>	S48, A39, bol 26f
Mittelwertverteilung	$P^* \approx 1/n \sum x_i$ verteilt; $E(P^*) = \mu$; $\text{Var}(P^*) = \sigma^2 / n$	a43
Modalwert	diskret: $P(X = x_{\text{Mod}}) \geq P(X = x) \quad \forall x$; kontin.: $\varphi(x_{\text{Mod}}) = \max \varphi(x)$	bol 60

<p>Multinomialverteilung: Bin (n, p₁, ..., p_n)</p>	<p><u>Masse</u>: $\alpha_{k_1 k_2 \dots k_n} = \binom{n}{k_1 k_2 \dots k_n} p_1^{k_1} p_2^{k_2} \dots p_m^{k_m} = \frac{n!}{\prod_{i=1}^m k_i!} \prod_{i=1}^m p_i^{k_i}$ n unabh. Zieh. <u>mit</u> Zurücklegen mit n=Stichprobenumfang, m=Anzahl der Ausprägungen, p_i ∈ S_m: Einheitssimplex des Rⁿ <u>Momente</u>: μ_j = n p_j; μ = (μ₁, μ₂, ..., μ_n); σ_j² = n p_j (1-p_j)</p>	<p>a2 momente: a32</p>
<p>Normalverteilung: N (μ , σ²)</p>	<p><u>Dichte</u>: $\varphi(x) = (1 / \sigma \sqrt{2 \pi}) e^{-x^2 / 2 \sigma^2}$ <u>Momente</u>: E(x)= μ ; Var(x) = σ² <u>Rücktransformation auf N(0,1)</u>: x → (x - μ) / σ , also P(X ≤ x) = Φ ($\frac{x-\mu}{\sigma}$) P([μ - c , μ + c]) = 2 Φ (c / σ) - 1 <u>Faltung</u>: N(μ₁, σ₁²) * N(μ₂, σ₂²) = N(μ₁+μ₂ , σ₁² + σ₂²)</p>	<p>bol 56, a43, transf.:bol 92,a26 faltung: a42 μ, σ²: a45 mehrdim: bol 119</p>
<p>Poissonverteilung: Poi(a)</p>	<p>Verteilung der seltenen Ereignisse <u>Masse</u>: α_k = (a^k / k!) e^{-a} mit a ≥ 0 ; <u>Momente</u>: μ = a ; σ² = a <u>Faltung</u>: Poi(a) * Poi(b) = Poi (a+b); Poi(a) * Poi(a) * ... * Poi(a) = Poi(na) <u>Annäherung</u> von Poi (n a) durch N(n a , n a) ⇔ n a > 20 von Poi (n p) durch Bin (n,p) ⇔ n > 50 und p < 0.1</p>	<p>S52, bol 43, a28, a37, a44 faltung: a40</p>
<p>Potenzmenge</p>	<p>σ (Ω) ist feinste, diskrete σ - Algebra; { ∅, Ω } ist größte σ - Algebra</p>	<p>S46</p>
<p>Randdichten</p>	<p>$\varphi_1(x) = \int_{-\infty}^{\infty} \varphi(x, y) dy$, $\varphi_2(y) = \int_{-\infty}^{\infty} \varphi(x, y) dx$</p>	<p>a15, a17, a18, a19, a20, bol 126</p>
<p>Simpson's Paradoxon</p>		<p>a13</p>
<p>Standardnormalvertlg.: N (0,1):</p>	<p>$\varphi(x) = (1 / \sqrt{2 \pi}) e^{-x^2 / 2}$; Φ (a) = ∫_{-∞}^a φ (x) dx nur numerisch (→ Tabelle) ; μ = 0; σ = 1 Eigenschaften: Φ (a) = 1 - Φ (-a) , P ([a,b]) = Φ (b) - Φ (a)</p>	<p>S55, bol 56 a4, a12</p>
<p>Träger von P</p>	<p>Tr (P) = { ω_k α_k > 0 } zumeist Tr(P) ≡ N, N₀</p>	<p>S52, bol 42</p>
<p>Transformation (str.monoton)</p>	<p><u>kontin.</u>: j^f (y) = df⁻¹ (y) / dy j (f⁻¹ (y)) $P^f([a,b]) = \int_a^b \varphi^f(y) dy = \int_a^b \varphi(f^{-1}(y)) (f^{-1}(y))' dy = \int_{f^{-1}(a)}^{f^{-1}(b)} \varphi(x) dx$; '- ' für f fallend <u>diskret</u>: P(f(X) = y) = ∑_{g(x_i)=y} P (X = x_i) <u>affin. linear</u>: x → α x + β ⇒ φ^f = (1 / α) φ ((y - β) / α) ⇒ E(α x + β) = α E(x) + β ; Var(α x + β) = α² Var(x)</p>	<p>bol 87, a37, a25, a27 E,Var: bol 94 kontin: ÜbBlatt 7</p>

Tschebyscheff'sche Ungleichung	$P(E(X) - d, E(X) + d ^C) \leq (1/d^2) \text{Var}(X)$; sinnvoll für $d \geq \sigma$, aber sehr ungenau diskret: $P(x - \mu < d) \leq 1 - \text{Var}(x) / d^2$ (Vorsicht diskreter Bereich!)	a38, a43, a49, bol 180
Unabhängigkeit der Merkmale	$\Leftrightarrow \varphi(x,y) = \varphi_1(x) \varphi_2(y) \Leftrightarrow \varphi(y x) = \varphi_2(y) \Leftrightarrow \varphi(x y) = \varphi_1(x)$ $\Rightarrow \text{Cov}(x,y) = 0$ (aber: $\text{Cov}(x,y) = 0 \not\Rightarrow x,y$ unabhängig) x,y unabh. $\Rightarrow x,y$ unkorreliert; (Umkehrschluß falsch, aber x,y korreliert $\Rightarrow x,y$ abhängig)	a15, a17, a18 , a19, a35, (bol 134)
Varianz	kontin: $\text{Var}(x) = \sigma_x^2 = E(x^2) - [E(x)]^2 = [\int x^2 \varphi(x) dx] - \mu^2$ 2-dimensional: $\sigma_{xy} = E(xy) - \mu_x \mu_y \Rightarrow$ symm. (Ko-)Varianzmatrix: $\mathbf{r} = \begin{pmatrix} s_x^2 & s_{xy} \\ s_{xy} & s_y^2 \end{pmatrix}$ diskret: $\text{Var}(x) = \sigma^2 = E(x^2) - [E(x)]^2 = [\sum x_k^2 \alpha_k] - \mu^2$ $\text{Var}(\alpha x + \beta) = \alpha^2 \text{Var}(x)$ Hinweis: $\text{Var}(x+y) = \text{Var}(x) + \text{Var}(y) + 2 \text{Cov}(x,y)$ (siehe Unabhängigkeit)	a29; a36
Verteilungsfunktion von P	$F(a) = P((-\infty; a])$; mon.st., $F(-\infty) = 0, F(\infty) = 1$, <u>kontin.</u> : $F(a) = \int_{-\infty}^a f(x) dx$; <u>diskret</u> : $F(a) = \sum_{0^a} P(X=x_i)$ ist rechts.halbsteig, $P(x_i) = F(x_i) - F(x_i - 1)$ Hinweis: mehrdimensional: Vorsicht bei Bereichen !!!	S53f, bol 33/45
Wahrscheinlichkeitsmaß	$P: A \rightarrow [0,1]$ mit folgenden Bedingungen: 1. $P(\Omega) = 1$ 2. $P(A^c) = 1 - P(A)$ 3. $\{A_n\}_{n \in \mathbb{N}} \subset A$ mit $\forall A_m \cap A_n = \emptyset \Rightarrow P(\cup A_n) = \sum P(A_n)$	S50, A36f, bol 19
Wahrscheinlichkeitsraum	$(\Omega, \mathcal{A}(\Omega), P)$ Ω = Grundgesamtheit $\neq \emptyset$ $\mathcal{A}(\Omega)$ = Mengensystem von Teilmengen der Grundgesamtheit ($\rightarrow \sigma$ -Algebra) P = Wahrscheinlichkeitsmaß (siehe dort) $P: A \rightarrow [0,1]$	S50; A36
Zentraler Grenzwertsatz	$z_n = (1/\sqrt{n}\sigma) \sum (x_i - \mu) \Rightarrow \lim_{n \rightarrow \infty} F_{z_n}(\alpha) = \Phi(\alpha)$	bol 192

*Hinweise: s. **xx** bezeichnet die Seite in meiner Mitschrift zu Statistik II (WS 95/96), **bol xx** die Seite in Bol, Wahrscheinlichkeitstheorie, Oldenbourg Verlag, **a xx** die Aufgabe zu Statistik II, sonstige Abkürzungen siehe Deskriptive Statistik*

3. Schätz- und Testtheorie

Erwartungstreue	$\forall \vartheta \in \Theta: E_{\vartheta}(s) = \mu \Leftrightarrow$ Summe der Gewichte = 1	a55, ind 70
Exponentialfamilie	<u>kontin.:</u> eindim: $\varphi_{\vartheta}(x) = c(\vartheta) \cdot h(x) \cdot \exp(\sum_1^k \xi_j(\vartheta) T_j(x))$ zweidim: $\varphi_{(\vartheta_1, \vartheta_2)}(x) = [c(\vartheta)]^2 [h(x_1) \cdot h(x_2)] \exp(\sum_1^k \xi_j(\vartheta) (T_j(x_1) + T_j(x_2)))$ $\Rightarrow \tau$ suff. für ϑ bei einfacher Stichprobe ; z.b. $N(\mu, \sigma^2), \text{Bin}(n,p)$	a54, ind 55
Gütefunktion, Macht, Power	$\beta(\mu) = P_{H_0}(\bar{x} \in F_1) = 1 - \Phi((\bar{x} - \mu) \sqrt{n} / \bar{s})$ mit F_1 : kritischer Bereich (\rightarrow Tabelle Testen)	ind 176, s.90
Maximum-Likelihood-Methode (ML-Schätzfunktion)	$L_x(\vartheta) = \prod_{i=1}^n P_{\vartheta}(x_i) = \prod_{i=1}^n P(X_i = x_i) = \varphi_{\vartheta}(x)$; $l_x(\vartheta) = \ln L_x(\vartheta)$; bestimme: $(\frac{d}{d\vartheta} L_x / \varphi_{\vartheta}) = 0$ oder $(\frac{d}{d\vartheta} l_x / \varphi_{\vartheta}) = 0$, prüfe, ob $(\frac{d^2}{d\vartheta^2} .. / \varphi_{\vartheta}^2) < 0$ (Maximum!) <u>Bin(1,p):</u> $\hat{p} = \sum x_i / n$ (GBE) <u>Poi(a):</u> $\hat{a} = (\sum x_i) / n = \bar{x}$ (GBE) <u>Exp(λ):</u> $\hat{\lambda} = n / \sum x_i = 1 / \bar{x}$ (nicht erwartungstreu) <u>Erlang:</u> $\hat{\lambda} = m n / \sum x_i$ (nicht erwartungstreu) <u>N(μ, σ^2):</u> $\hat{\mu} = \frac{1}{n} \sum x_i$ (GBE) μ bekannt: $\hat{\sigma}^2 = \frac{1}{n} \sum (x_i - \mu)^2$ (GBE) μ unbekannt: $\hat{\sigma}^2 = \frac{1}{n} \sum (x_i - \bar{x})^2$ (nicht erwartungstreu)	Bin: a57, ind 87 Exp: a58, ind 92 N: a59 Geo: s.80 Poi: ind 95 {6.1} Weibull: {6.4} Gamma: {6.5} Erlang: a59
Schätzer	Poi(a): $\tau(x) = \bar{x} = \frac{1}{n} \sum x_i$ ist GBE für a Bin(m,p): $\tau(x) = \bar{x} / m = (1/nm) \sum x_i$ ist GBE für mp (\rightarrow Max. Likelihood) $\tau(x) = \bar{x}$ ist erwartungstreu, aber nicht Varianzminimal Exp(λ): $\tau(x) = (n-1) / \sum x_i$ ist GBE für λ N(μ, σ^2): $\tau(x) = \bar{x}$ ist GBE für μ mit μ bekannt: $\tau(x) = s^2 = \frac{1}{n} \sum (x_i - \mu)^2$ ist GBE für σ^2 mit μ unbekannt: $\tau(x) = \bar{s}^2 = 1 / (n-1) \sum (x_i - \bar{x})^2$ ist GBE für σ^2 mit GBE (Gleichmäßig Beste Erwartungstreue Schätzfunktion) = UMVU = GMVE	ind 81-83
Statistik	Stichprobenmittelwert: $\tau(x_1, \dots, x_n) = 1/n \sum x_i$ Stichprobenvarianz: $\tau(x_1, \dots, x_n) = s^2(x) = 1/n \sum (x_i - \bar{x})^2$ korrigierte Stichprobenvarianz: $\tau(x_1, \dots, x_n) = s^{*2}(x) = 1 / (n-1) \sum (x_i - \bar{x})^2$	ind 45

Suffizienz (Fisher-Neyman)	$\Leftrightarrow \exists g_{\vartheta}, h: P_{\vartheta}(x_1, \dots, x_n) = g_{\vartheta}(\tau(x_1, \dots, x_n)) h(x_1, \dots, x_n)$ $\tau(x) = \sum x_i$ suff. für: p [Bin (m,p)], a [Poi (a)], λ [Exp (λ)] , μ [N(μ, σ^2) bei fester Var.] $\tau(x) = \bar{x} = 1/n \sum x_i$ ist suffizient für: mp [Bin (m,p)], a[Poi (a)] und, μ [N(μ, σ^2)]	a52, a53, ind 50, ind 81
Tests zum Niveau α	bei $X_i \sim N(\mu, \sigma^2)$: \rightarrow Tabelle Testen, \rightarrow Tabelle ind 198 bei $X_i \sim \text{Bin}(n,p)$: \rightarrow a62	ind 123/130 a62, 63, 64,65,66
t_n - Verteilung	Quantile \rightarrow Tabelle t_n -Verteilung Annäherung durch $N(0, 1) \Leftrightarrow n > 30$	ZGS: a63

Konfidenzintervalle

	Verteilungs- annahme der X_i	ges.	(un)bekannte Parameter, Bedingungen	Berechnung von c	KI (Konfidenzintervall)
1.1	keine Verteilung, aber $n > 30$	μ	σ bekannt	$c = \Phi^{-1} \left(1 - \frac{\alpha}{2} \right) \frac{s}{\sqrt{n}}$	$[\bar{x} - c, \bar{x} + c]$
1.2			σ unbekannt	$c = \Phi^{-1} \left(1 - \frac{\alpha}{2} \right) \frac{\bar{s}}{\sqrt{n}}$	$[\bar{x} - c, \bar{x} + c]$
2.1	$N(\mu, \sigma^2)$ -verteilt	μ	σ bekannt	$c = \Phi^{-1} \left(1 - \frac{\alpha}{2} \right) \frac{s}{\sqrt{n}}$	$[\bar{x} - c, \bar{x} + c]$
2.2.1			σ unbekannt $n \leq 30$	$c = t^{-1}_{(n-1, 1 - \frac{\alpha}{2})} \frac{\bar{s}}{\sqrt{n}}$	$[\bar{x} - c, \bar{x} + c]$
2.2.2			σ unbekannt $n > 30$	$c = \Phi^{-1} \left(1 - \frac{\alpha}{2} \right) \frac{\bar{s}}{\sqrt{n}}$	$[\bar{x} - c, \bar{x} + c]$
2.3		σ	μ bekannt	$c_1 = \chi^2_{(n, \frac{\alpha}{2})}, \quad c_2 = \chi^2_{(n, 1 - \frac{\alpha}{2})}$	$[n s^2/c_1, n s^2/c_2]$
2.4		σ	μ unbekannt	$c_1 = \chi^2_{(n-1, \frac{\alpha}{2})}, \quad c_2 = \chi^2_{(n-1, 1 - \frac{\alpha}{2})}$	$[(n-1) \bar{s}^2/c_1, (n-1) \bar{s}^2/c_2]$
3.1	Bin (1,p) - verteilt	μ	$5 \leq \sum x_i \leq n-5$	$c = \frac{\sqrt{\bar{x}(1-\bar{x})}}{\sqrt{n}} \Phi^{-1} \left(1 - \frac{\alpha}{2} \right)$	$[\bar{x} - c, \bar{x} + c]$

mit $\bar{s}^2 = (1 / (n-1)) \sum (x_i - \bar{x})^2$, $s^2 = \frac{1}{n} \sum (x_i - \mu)^2$, $\bar{x} = \frac{1}{n} \sum x_i$