

Formelsammlung Statistik

3. November 2009

Das ist eine Formelsammlung für Statistik. Die Formelsammlung enthält alle Formeln aus dem Skript des Wintersemesters 2009/2010. Außerdem ein paar Sachen die mir sinnvoll erschienen und für die Klausur notwendig sein könnten, sowie Formblätter zum schnellen Ausfüllen während der Klausur.

Quellen sind (1) Statistikscript Prof. Dr. Müller, HS Wismar und (2) Taschenbuch der Wirtschaftsmathematik, Wolfgang Eichholz und Eberhard Vilkner.

Statistische Masse	Umfang der Einheiten einer statistischen Untersuchung
Statistische Einheit	Untersuchungsobjekt einer statistischen Untersuchung. Träger der interessanten Informationen.
Merkmal	Zu betrachtendes Attribut einer Einheit. Etwa Einkommen, Altern, ...
Merkmalstypen	<p>diskrete Merkmalstypen bestehen aus einer überschaubare, endliche Menge (etwa Geschlecht),</p> <p>stetige Merkmalstypen können in einem bestimmten Bereich jeden reellen Wert annehmen,</p> <p>quasi-stetige Merkmalstypen sind eigentlich diskret, enthalten aber sehr grosse Menge von möglichen Merkmalen</p>
Merkmalsausprägung	<p>Gruppierung Sortierung, gleiche Merkmalsausprägung</p> <p>Klassifizierung benachbarte Ausprägungen werden zu einer Klasse zusammengefasst. Übliche Schreibweise $[200; 400)$ mit der Bedeutung $200 \leq x < 400$.</p>
Skalenniveau	<p>nominal qualitativ (also keine Zahlen), etwa Geschlecht oder Studiengang. Darstellung als gruppierter Wert.</p> <p>ordinal Merkmalsausprägung mit objektiver Rangordnung, Abstände sind aber nicht bezifferbar (etwa Noten). Darstellung als gruppierter Wert.</p> <p>metrisch <i>Interval quantitativ</i>: reelle Zahlen, natürliche Rangfolge, eindeutige Abstände, etwa Sparsumme, <i>Verhältnis quantitativ</i>: reelle Zahlen, natürliche Rangfolge, eindeutige Abstände, absoluter Bezugspunkt (etwa Nullpunkt). Beispiel: Alter. Darstellung als klassierter Wert.</p>

Tabelle 0.1: Begriffe

1 Univariante Datenanalyse

Auto ¹	x_i	h_i	H_i	f_i	F_i	$x_i h_i$	$x_i \cdot f_i$	$(x_i)^2 \cdot f_i$
BMW	342	1	1	0,167	0,167	342	57,00	19494
Mercedes	549	1	2	0,167	0,333	549	91,50	50234
VW/Audi	1501	1	5	0,167	0,833	1501	250,17	3,76E5
Sonstige	1713	1	6	0,167	1,000	1713	285,50	4,89E5

Tabelle 1.1: Beispiel gruppierter, nominaler Werte

x_i	h_i	H_i	f_i	F_i	$x_i h_i$	$x_i \cdot f_i$	$(x_i)^2 \cdot f_i$
280	1	1	0,1	0,1	280
340	2	3	0,2	0,3	680
740	1	9	0,1	0,9	740
1180	1	10	0,1	1,0	1180

Tabelle 1.2: Beispiel gruppierter, ordinaler Werte

x_i	h_i	H_i	f_i	F_i	Δx_i	f_i^*	h_i^*
[200;400)	21	21	0,21	0,21	200	0,00105	0,1050
[700;1000)	19	96	0,19	0,96	300	0,00063	0,0633
[1000;1500)	2	98	0,02	0,98	500	0,00004	0,0040
[1500;2000)	2	100	0,02	1,00	500	0,00004	0,0040

Tabelle 1.3: Beispiel klassierter, metrischer Werte

Math	TR	Formel	Erläuterung
h_i	hi		abs. Häufigkeit
H_i	shi	$h_1 + \dots + h_i = \sum_{j=1}^i h_j$	abs. Summenhäufigkeit cumsum(hi)
f_i	fi	$\frac{h_i}{N}$ mit $\sum_{i=1}^k f_i$	relative Häufigkeit
			gruppiert Stabdiagramm siehe Abbildung 1.3 auf Seite 10
			klassiert Histogramm, siehe Abbildung 1.4 auf Seite 11
F_i	sfi	$f_1 + \dots + f_i = \sum_{j=1}^i f_j$	abs. Summenhäufigkeit
N	n	$\sum_{i=1}^k h_i$	Stat Masse
h_i^*	his	$\frac{h_i}{\Delta x_i}$	abs Häufigkeitsdichte
f_i^*	fis	$\frac{f_i}{\Delta x_i}$	rel Häufigkeitsdichte
$F(x)$	f(x)		Verteilungsfunktion, Funktion der relativen Summenhäufigkeit Als Beispiel für gruppierte Daten: $F(500)=0,30 \rightarrow$ Es wird nicht gerechnet, sondern aus dem Diagramm abgelesen, da es sich um gruppierte Werte handelt! Als grafische Lösung (Treppendiagramm, keine Zwischenwerte!) siehe Abbildung 1.1 auf Seite 10
	Gruppe	$F(x) = \begin{cases} 0 & x < x_1 \\ F_i & x_i \leq x < x_{i+1} \\ 1 & x \geq x_k \end{cases}$	Bei klassierten Daten: Klasse aus Diagramm ablesen (H_i), untere und obere Grenzen der Klasse herauslesen, in Formel einsetzen: $F(500) = 0,21 + \frac{0,56}{300}(500 - 400) = 0,397 = 39,7\%$ als grafische Lösung siehe Funktionsdiagramm 1.2 auf Seite 10
	Klasse	$F(x) = \begin{cases} 0 & x < x_1^u \\ F(x_i^u) + \frac{f_i}{\Delta x_i}(x - x_i^u) & x_i^u \leq x < x_i^o \\ 1 & x \geq x_k^o \end{cases}$	

Tabelle 1.4: Überblick Häufigkeiten

Name	Math	TR	nominal	ordinal	metrisch	Vor- und Nachteile
Modal	x_D	xd	ja	ja	ja	Ist die Merkmalsausprägung, die am häufigsten vorkommt. Es kann mehrere Modalwerte geben.
Median	x_z	xz	?	ja	ja	Mitte aller Merkmalsträger, bzw. welcher Merkmalswert wird von der Hälfte aller Merkmalsträger nicht überschritten. Vorteil: Robust gegen Ausreißer. Mittelwert für ordinale Daten
Quantil	x_p	xp	?	?	?	ein Teil aller Merkmalsträger (etwa 0,25x oder 0,75x) bzw. welcher Merkmalswert wird von einem Teil aller Merkmalsträger nicht überschritten. Dabe ist das $x_p = x_{0.5} = x_z$
Arith. Mittelw.	\bar{x}	xs	nein	nein	ja	Der Durchschnitt oder Mittelwert aller Merkmale Mittelwert für metrische Daten
Geom Mittelw.	x_G	xg	?	?	ja	Mittelwert für Produkte, etwa bei Verhältnissen oder Wachstumswerten. Nur für Zahlen > 0 sinnvoll. Mittelwert für Produkte

Tabelle 1.5: Überblick Lageparameter

Math	TR	Formel	Erläuterung
x_D	xd	Gruppe da x_i wo f_i am größten ist Klasse Mitte der modalen Klasse $x_D = \frac{x_i^u + x_i^o}{2} = x_i'$	Ist die Merkmalsausprägung, die am häufigsten vorkommt. Es kann mehrere Modalwerte geben.
	xz	Gruppe $x_z = 0.5N$ Klasse $x_z = x_i^u + \frac{0.5 - F(x_i^u)}{f_i} * \Delta x_i$	Median bzw. Zentralwert ist der Wert, der in der Mitte der Variantsreihe liegt. Ist N gerade, wird der Mittelwert der zwei mittelsten Werte ermittelt. Beispiel: Zuerst Klasse bestimmen und dann $400 + \frac{0.5 - 0.21}{0.56} * 300 = 555.36$ Mittelwert für ordinale Daten
x_p	xp	Gruppe $x_p = p \cdot N$ Klasse $x_p = x_i^u + \frac{p - F(x_i^u)}{f_i} * \Delta x_i$	Wird eine Variationsreihe in gleich große Teile zerlegt, entstehen Quantile. Typisch sind 0,25, 05, 0,75. Der Quantilabstand ist $Q = x_{0.75} - x_{0.25}$. Das 0,5-Quantil ist gleich dem Median. Quantil ist gewissermaßen das Gegenüber der Verteilungsfunktion!
	xa	Gruppe $\bar{x} = \sum_{i=1}^k x_i f_i$ Klasse $\bar{x} = \sum_{i=1}^k x_i' f_i$	Arithmetischer Mittelwert bzw. Durchschnitt. Durchschnitte werden mit dieser Formel addiert: $\bar{x} = \frac{\sum_k N_m * x_m}{\sum_{m=1}^k N_m}$
x_G	xg		Mittelwert für metrische Daten Der Geometrische Mittelwert wird bei der Mittelung von Wachstumsraten oder multiplikativ verknüpften Daten angewendet.
		$x_G = \sqrt[N]{\prod_{i=1}^k x_i}$	

Tabelle 1.6: Lageparameter

Math	TR	Formel	Erläuterung
R	r	$R = x_{max} - x_{min}$	Die Spannweite ist die Differenz zwischen größtem und kleinstem Merkmalswert.
		K $R = x_k^o - x_1^u$	
Q	q	$Q = x_{0.75} - x_{0.25}$	Der Quantilsabstand ist der Abstand zwischen oberem und unterem Quantil.
s_x^2	s2x	$s_x^2 = \left\{ \sum_{i=1}^k [(x_i)^2 \cdot f_i] \right\} - \bar{x}$ $s_x^2 = \left\{ \sum_{i=1}^k [(x'_i)^2 \cdot f_i] \right\} - \bar{x}^2$	Die Varianz ist die mittlere quadratische Abweichung aller Merkmalsausprägungen vom arith. Mittelwert. Alternative Zeichen der Varianz sind $s_x^2 = s^2 = \sigma^2$
		K G/K	Varianz der Grundgesamtheit. Gleichungsbeispiel bei der Annahme, dass es zwei Teilmengen gibt und die jeweils die Varianzen und Mittelwerte bekannt sind.
		$s^2 = \frac{N_1 s_1^2 + N_2 s_2^2}{N_1 + N_2} + \frac{N_1 (\bar{x}_1 - \bar{x})^2 + N_2 (\bar{x}_2 - \bar{x})^2}{N_1 + N_2}$	
s_x	sx	$s_x = \sqrt{s_x^2}$	Standardabweichung mittlere Abweichung vom Mittelwert. Nachteil: Bei großen Merkmalsmengen nimmt die Schwankungsbreite zu. Ein Vergleich zwischen Messreihen mit großen und kleinen Verteilungen ist daher ggf. nicht mehr sinnvoll. Statt dessen: Variationskoeffizient. Hinweis: Bei $s_x = 0$ gibt es einen eindeutigen Hinweis auf Konzentration. Ansonsten nicht.
v	v	$v = \frac{s_x}{\bar{x}}$	Variationskoeffizient = Auf den Mittelwert bezogenes relatives Streuungsmaß, sofern nur positive Werte auftreten.

Tabelle 1.7: Streuungsparameter

Math	TR	Formel	Erläuterung
p_I	pi		Konzentrationskoeffizient berechnet den Anteil eines Merkalwertes an der Merkmalssumme, bzw. zeigt den prozentualen Anteil an der Gesamtsumme.
		Gruppe $p_i = \frac{x_i \cdot h_i}{N \cdot \bar{x}}$	
		Klasse $p_i = \frac{x'_i \cdot h_i}{N \cdot \bar{x}}$	
P_i	spi	$P_i = \sum_{j=1}^i p_j$	Das Konzentrationsmaß beschreibt die relative Merkmalssumme. Die Lorenzkurve veranschaulicht das Konzentrationsmaß grafisch.
$L(F_i)$	-	-	Die Fläche zwischen Gleichverteilung und Lorenzkurve wird als Lorenzfläche bezeichnet und ist ein weiteres Konzentrationsmaß. Je größer die Fläche, desto größer die Konzentration. Beispiel zur Lorenzkurve siehe 1.5 auf Seite 11
			Lorenzkurve: Welchen Anteil haben Merkmalsträger an Merkmalen. Etwa 0.5 = 50% der Autohersteller(F_i) haben Anteil von 0.25 = 25% P_i der Produktion
G	g	$G = \frac{0.5 - A(L)}{0.5}$ im Bereich $0 \leq G \leq 1$	Stat. Maß zur Darstellung der Ungleichverteilung. Der Gini-Koeffizient misst die Höhe der relativen Konzentration über das Verhältnis der Lorenzfläche zur Fläche bei maximaler Konzentration. Es kann unterschiedliche Lorenzflächen bei identischem G geben. <i>Eigenschaft:</i> Werden alle x_i um denselben Prozentsatz erhöht oder gesenkt, dann bleibt der Gini-Koeffizient unverändert. Werden alle x_i um einen additiven Zuschlag erhöht, dann wird der Gini-Koeffizient kleiner. Wird ein x_i -teibetrag von einem größeren zu einem kleineren x_i transferiert, so wird der Gini-Koeffizient kleiner.
			<i>Beispiel:</i> Der Ginikoeffizient für die Einkommensverteilung liegt in Deutschland bei 0,274 (2003), in Frankreich bei 0,327 (1995), in Großbritannien bei 0,360 (1999), in Japan bei 0,249 (1993) und in den USA bei 0,408
$A(L)$	al	$A(L) = \sum_{i=1}^k \frac{P_{i-1} + P_i}{2} \cdot f_i$ mit $P_0 = 0$	Fläche unter der Trapezkurve Hinweis: Sind großgeschriebene P's also spi.

Tabelle 1.8: Relative Konzentration

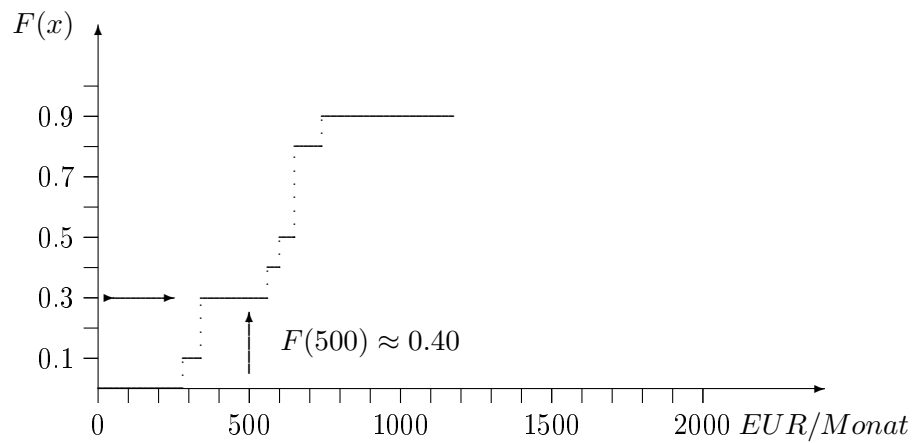


Abbildung 1.1: Funktion relativer Sumenhäufigkeit $F(x)$ bei gruppierten Daten

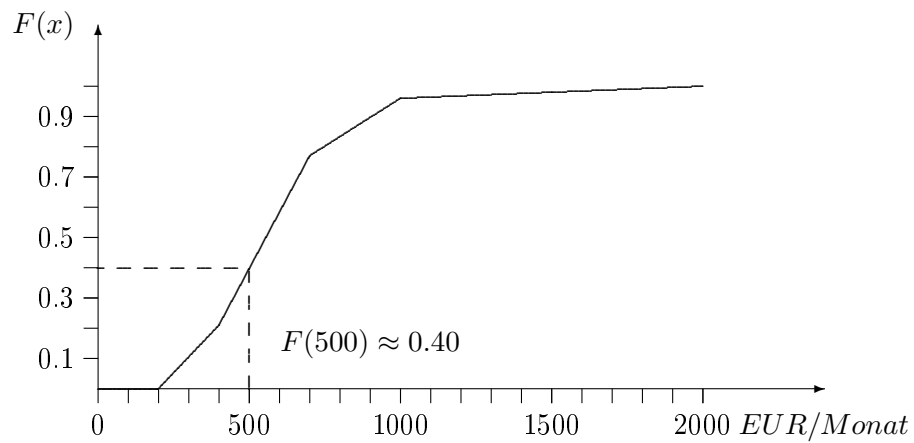


Abbildung 1.2: Funktion $F(x)$ relativer Summenhäufigkeit bei klass. Daten

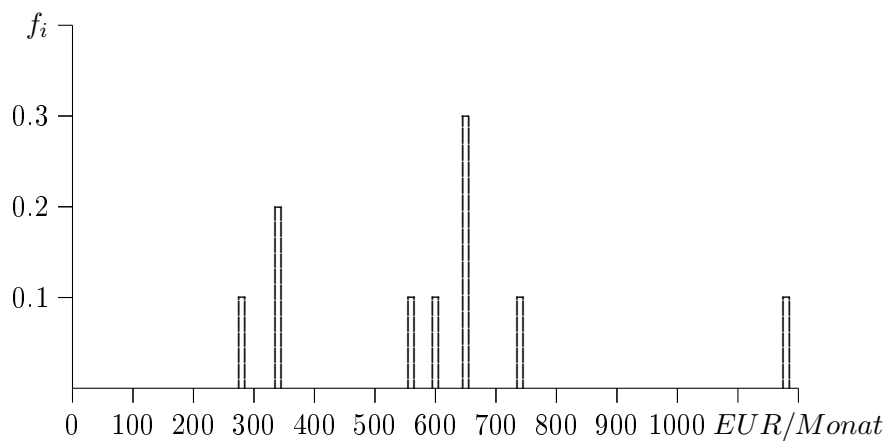


Abbildung 1.3: Darstellung rel Häufigkeit von Gruppen: Stabdiagramm

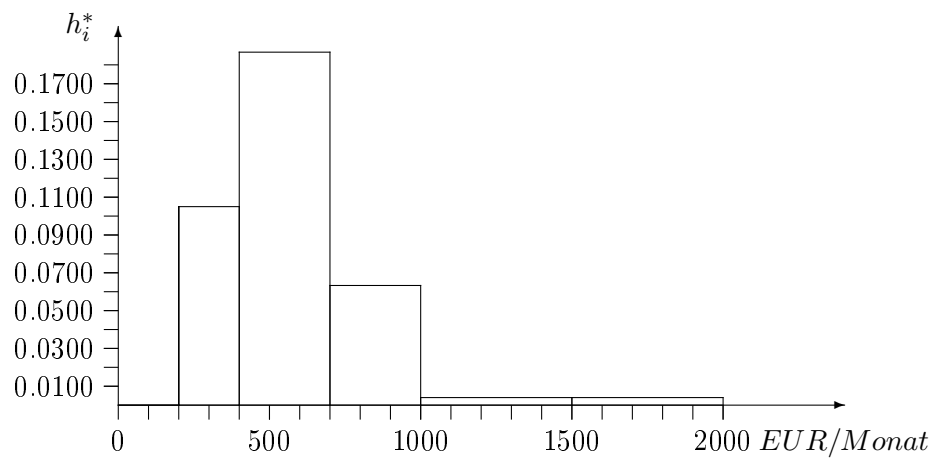


Abbildung 1.4: Darstellung rel. Häufigkeit von Klassen: Histogramm

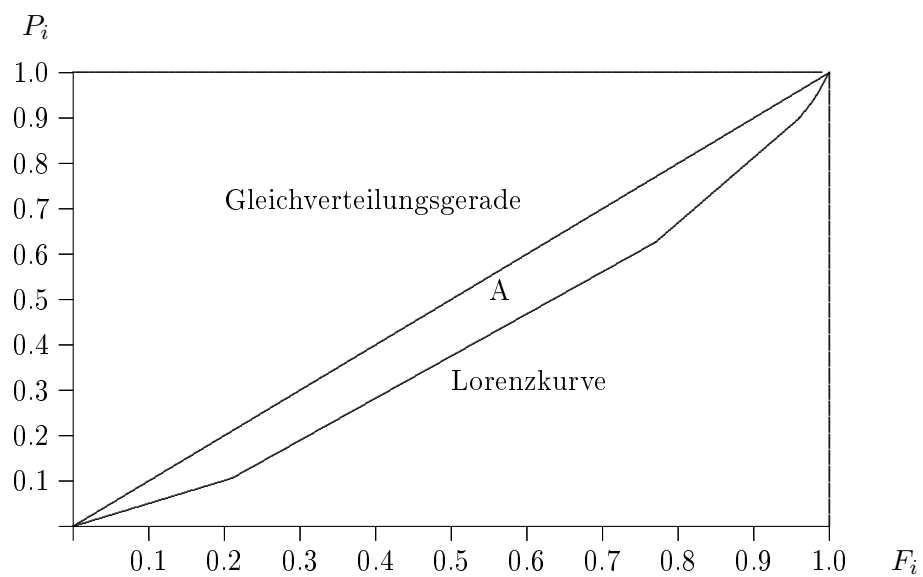


Abbildung 1.5: Lorenzkurve

2 Multivariate Datenanalyse

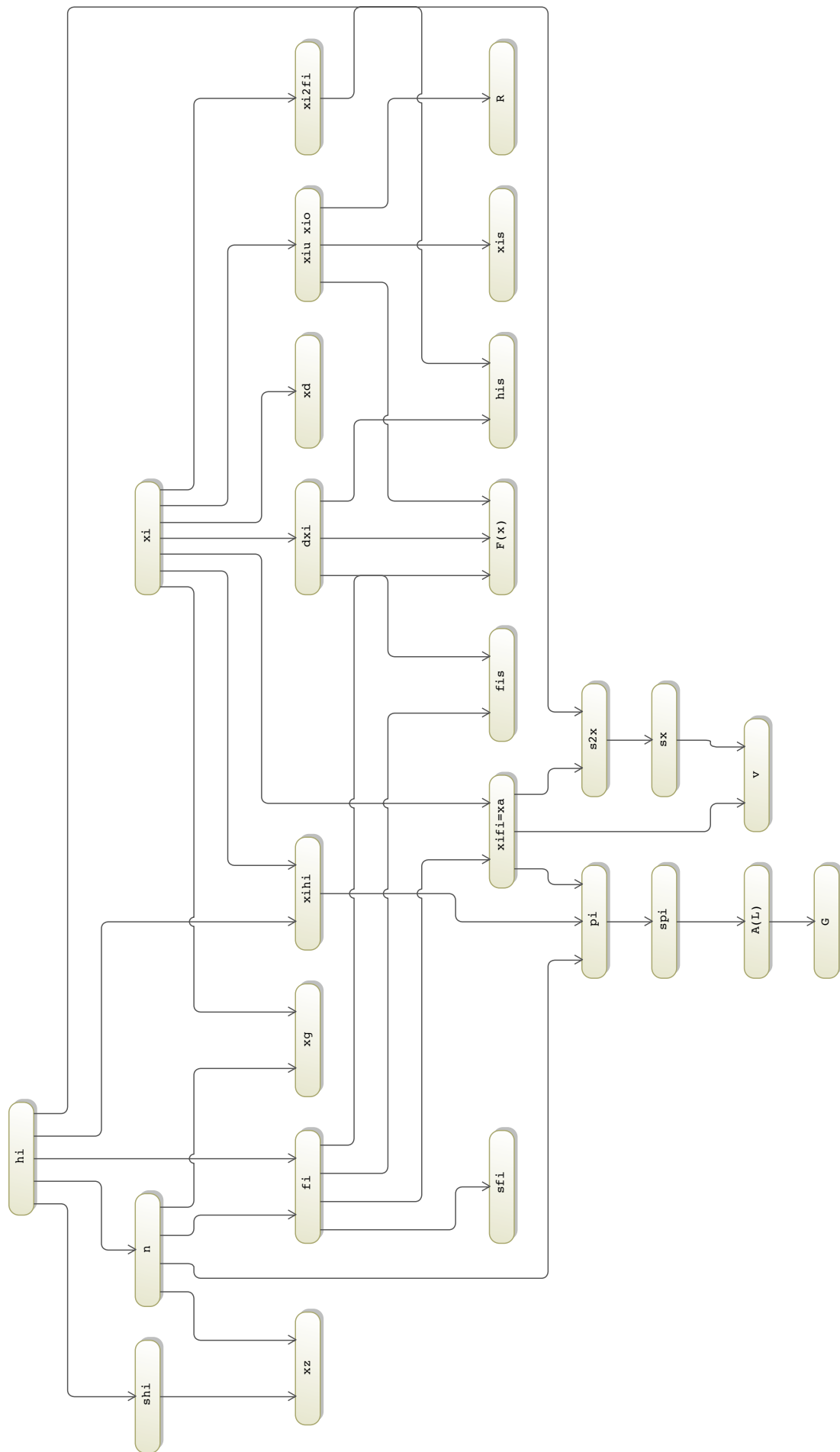
	Studiengang		
Geschlecht	BWL	WI	Summe
weiblich	57	2	59
männlich	43	8	51
Summe	100	10	100

Tabelle 2.1: Beispiel Abs. Häufigkeitsverteilung

3 Anhang

x	x	Klasse oder Gruppe einer statistischen Zählung. Variable kann Zeichen haben wie 1, i , k die für das 1-te, i -te oder letzte Gruppe/Klasse stehen.
x_d	xd	Modalwert, der Wert mit der häufigsten Merkmalsausprägung
x_z	xz	Median, Mitte aller Merkmalsausprägungen, d.h. nach oben und unten gleich viele Merkmalsausprägungen
x_p	xp	Quantile überschreiten einen gewissen Anteil von Merkmalsausprägungen <i>nicht</i>
x'_i		Klassenmitte der i -ten Klasse
x_i^u x_i^o		untere bzw. obere Grenze der i -ten Klasse
h	h	Anzahl von Einheiten innerhalb einer Gruppe oder Klasse. Tiefgestellte Zeichen gleiche Bedeutung wie bei x Die Summe aller h ist die statistische Masse
H_i	shi	absolute Summenhäufigkeit, wie h_i aber aufsteigend addiert. Der größte Wert= N
f_i	fi	relative Häufigkeit. Summe aller $f_i = 1$ Entspricht dem prozentualen Anteil an der statistischen Masse.
F_i	sfi	relative Summenhäufigkeit. Wie f_i aber aufsummiert. Der größte Wert = 1
Δx_i	dx _i	Klassenbreite der i -ten Klasse
s_i	si	relative Summenhäufigkeit einer Klasse
N	n	Statistische Masse, also die Menge aller Merkmalsausprägungen.

Table 3.1: Überblick Variablen



- ◇1 Umschalten Statistikmodus auf Gruppe
- ◇2 Umschalten Statistikmodus auf Klasse
- ◇3
- ◇4 Berechne unteres Quantil
- ◇5 Berechne Median
- ◇6 Berechne oberes Quantil
- ◇7 Berechne Modus
- ◇8
- ◇9 Alles neu berechnen

Table 3.2: Überblick Tastaturabkürzungen

4 Multivariate Datenanalyse

Gruppe/Klassenmitte Merkmal x		Struktur y		
		y_1	y_2	Zeilensumme
1	$x_{i=1}$	h_{11}	h_{12}	$h_{1.}$
2	$x_{i=2}$	h_{21}	h_{22}	$h_{2.}$
Spaltensumme		$h_{.1}$	$h_{.2}$	$N =$

Tabelle 4.1: Zweidimensionale Häufigkeitsverteilung

Gruppe/Klassenmitte Geschlecht		Studiengang		
		BWL	WI	Zeilensumme
1	weiblich	57	2	59
2	männlich	43	8	51
Spaltensumme		100	10	110

Tabelle 4.2: Beispiel Abs. Häufigkeitsverteilung bei nominalen Daten

Gruppe/Klassenmitte Geschlecht		Studiengang		
		BWL	WI	Zeilensumme
1	weiblich	0.518	0.018	0.536
2	männlich	0.391	0.073	0.464
Spaltensumme		0.909	0.091	1.000

Tabelle 4.3: Beispiel Rel. Häufigkeitsverteilung bei nominalen Daten

Studierende No	1	2	3	4	5	6	7	8	9	10
Alter	22	24	26	24	21	28	21	21	23	21
Einkommen	280	650	650	740	340	1180	340	600	650	560

Tabelle 4.4: Beispiel 2D-Häufigkeitstabelle (Wertepaare)

Gruppe/Klassenmitte		Einkommen							Zeilensumme	
Alter		280	340	560	600	650	740	1180		
1	21	2		1	1				4	
2	22	1							1	
3	23						1			1
4	24						1	1	2	
5	26	1							1	
6	28							1	1	
Spaltensumme		1	2	1	1	3	1	1	10	

Tabelle 4.5: Beispiel 2D-Häufigkeitstabelle (Korrelationstabelle)

Math	TR	Formel	Erläuterung
$h_{i.}$	hip	$\sum_{k=1}^n h_{ik} \quad (i = 1, 2, \dots, m)$	abs. Häufigkeit von x bzw. Randhäufigkeit von x oder auch schlicht "Zeilensumme".
$h_{.k}$	hpk	$\sum_{i=1}^m h_{ik} \quad (k = 1, 2, \dots, n)$	abs. Häufigkeit von y bzw. Randhäufigkeit von y oder auch schlicht "Spaltensumme".
h_{ik}	n	$N = \sum_{i=1}^m \sum_{k=1}^n h_{ik} = \sum_{i=1}^m h_{i.} = \sum_{k=1}^n h_{.k}$	abs. Häufigkeit bzw. Anzahl Merkmalsträger oder auch schlicht: Die Summe aller Spalten bzw. Summe aller Zeilen.
f_{ik}	fik	$f_{ik} = \frac{h_{ik}}{N}$ bzw. $f_{i.} = \frac{h_{i.}}{N}$ bzw. $f_{.k} = \frac{h_{.k}}{N}$	rel. Häufigkeit
	sfik	$1 = \sum_{i=1}^m \sum_{k=1}^n f_{ik} = \sum_{i=1}^m f_{i.} = \sum_{k=1}^n f_{.k}$	Summe aller relativen Häufigkeiten
H			abs Häufigkeitstabelle von Matrix mit m_abs hae
F			rel Häufigkeitstabelle von Matrix mit m_rel hae

Tabelle 4.6: Überblick Häufigkeit

Math	TR	Formel	Erläuterung
$f(y_k x_i)$		$f(y_k x_i) = \frac{h_{ik}}{h_{i.}}$	Bedingte relative Häufigkeit. Absolute Häufigkeit durch Zeilensumme. Beispiel: $f(y_1, x_1) = \frac{h_{11}}{h_{1.}} = \frac{57}{59} = 0.966$ m_bedhae(m_abshae())
		$f(y_k x_i) = f(y_k x_2) = \dots = f(y_k x_m) = f_{.k}$ für $k = 1, 2, \dots, m$	Statistische Unabhängigkeit: Wenn alle bedingten, relativen Häufigkeiten gleich sind, sind die Merkmale statistisch unabhängig.
χ^2	kap	$N \cdot \left[\left(\sum_{i=1}^m \sum_{k=1}^n \frac{h_{ik}^2}{h_{i.} \cdot h_{.k}} \right) - 1 \right]$	Chi-Quadrat. Beispiel: $110 \left[\left(\frac{57^2}{59 \cdot 100} + \frac{2^2}{59 \cdot 10} + \frac{43^2}{51 \cdot 100} + \frac{8^2}{51 \cdot 10} \right) - 1 \right] = 5.0047$ =m_chiquad(m_abshae())
		$N \cdot \left[\left(\sum \sum \frac{\text{Zeilensumme}_h^2 \cdot \text{Spaltensumme}_h}{\text{Zeilensumme}_h \cdot \text{Spaltensumme}_h} \right) - 1 \right]$	
C	c	$C = \sqrt{\frac{1}{N} * \frac{\chi^2}{\min((m-1);(n-1))}}$	Maß von Cramer. Im Bereich 0..1. Bei vollständiger Abhängigkeit=1. $C = \sqrt{\frac{1}{110} * \frac{5.0047}{\min((2-1);(2-1))}} = 0.213$ =m_cramer(m_abshae())

Tabelle 4.7: Zusammenhang nominaler Merkmale

Math	TR	Formel	Erläuterung
			<p>Darstellung als</p> <ul style="list-style-type: none"> - Wertpaar - Korrelationstabelle - Streudiagramm (mit Verlauf des Funktionsgraphen) linear, polynom, kein Zusammenhang) <p>Kovarianz für Wertepaare Bei klassierten Daten Klassenmitten benutzen. Bei =0 kein <i>linearer</i> Zusammenhang. Einheit von $cov(x, y)$ ist das Produkt von x-Einheit und y-Einheit (etwa EuroJahr) Basis: Arbeitstabelle Wertepaare m_cov_w(wertematrix mit zwei zeilen) m_cov_wf(wertematrix mit zwei zeilen)</p> <p>Kovarianz für zweidim Häufigkeitsverteilung Bei klassierten Daten Klassenmitten benutzen. Bei =0 kein <i>linearer</i> Zusammenhang. Einheit von $cov(x, y)$ ist das Produkt von x-Einheit und y-Einheit (etwa EuroJahr) Basis: Arbeitstabelle Regressionsanalyse Korrelationskoeffizient nach Pearson/Bravis = Maß für die Höhe des linearen Zusammenhangs Wertebereich im Intervall -1 (=vollständig negative Abhängigkeit) und +1 (vollständig positiver Abhängigkeit) 0=kein <i>linearer</i> Zusammenhang m_s2x_w(wertetabelle) m_s2y_w(wertetabelle) m_rxy(wertematrix mit 2 Zeilen) Mittelwert der x-Zeile für Korrelationstabelle Mittelwert der y-Zeile für Korrelationstabelle</p>
$Cov(x, y)$	$cov(x, y)$	$= \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})$ $= \frac{1}{N} \sum_{i=1}^N x_i y_i - \bar{x}\bar{y}$	
$Cov(x, y)$	$cov(x, y)$	$= \frac{1}{N} \sum_{i=1}^m \sum_{k=1}^n (x_i - \bar{x})(y_k - \bar{y}) h_{ik}$ $= \frac{1}{N} \sum_{i=1}^m \sum_{k=1}^n x_i y_i f_{ik} - \bar{x}\bar{y}$	
r_{xy}	$rx y$	$= \frac{Cov(x, y)}{s_x s_y}$	
\bar{x}	xa	$= m_xa(tabelle, zeilen[])$ oder $= m_xaf(tabelle, zeilen[])$ (nur Wert)	
\bar{y}	ya	$= m_ya(tabelle, spalten[])$ oder $= m_yaf(tabelle, spalten[])$ (nur Wert)	

Tabelle 4.8: Zusammenhang metrischer Merkmale

Math	TR	Formel	Erläuterung
			<p><i>linear</i> $\hat{y} = a_1 x + a_0$</p> <p><i>polynom</i> $\hat{y} = a_2 x^2 + a_1 x + a_0$</p> <p>Geschätzte Regressionsgleichung nur sinnvoll, wenn eingesetzte Werte im Bereich der Messwerte liegen.</p> <p>Einheit x/y z.B. $\frac{Euro}{Jahr}$</p> <p>m_a1_w(wertetabelle)</p> <p>Einheit x z.B. <i>Euro</i></p> <p>m_a0_w(wertetabelle)</p> <p>Bewertung der Schätzqualität: desto besser desto mehr die Schwankungen der y-Werte durch die Gleichung erklärt werden können bzw. desto geringer die Residuen.</p> <p>Bestimmtheitsmaß=Kriterium der Schätzqualität einer Regressionsfunktion</p> <p>B^2 gibt den Erklärungsgrad der Regressionsfunktion im Intervall von 0 bis 1 an. Er bringt zum Ausdruck, wie der Zusammenhang zwischen den beiden Werten durch die Reg-Funktion beschrieben wird. 1=alle Werte liegen auf der Funktion.</p> <p>m_b2_w(wertetabelle)</p>
a_1	a1	$= \frac{Cov(x,y)}{s_x^2} = r_{xy} \frac{s_y}{s_x}$	
a_0	a0	$= \bar{y} - a_1 \bar{x}$	
B^2	b2	$= \frac{s_y^2}{s_y^2} = \frac{a_1^2 s_x^2}{s_y^2} = 1 - \frac{s_e^2}{s_y^2}$	

Tabelle 4.9: Regressionsrechnung

Math	TR	Formel	Erläuterung
R	r	$R = 1 - \frac{6 \cdot \sum_{i=1}^N d_i^2}{N(N^2-1)}$ mit $-1 \leq R \leq 1$	<p>Rangkorrelationskoeffizient nach Spearman als Maß der Ausgeprägtheit des Zusammenhangs zweier Merkmale.</p> <p>Für ordinale Werte (klare Rangordnung, aber Abstand nicht interpretierbar)</p> <p>RZ=Fortlaufend, aber Mittelwert für identische Werte</p> <p>l_rang(list) Rangzahlen einer Liste aufsteigend</p> <p>m_rang(list, list) -> Rangzahlen Arbeitstabelle</p> <p>m_rangko(m_rang) -> Rangzahlenkoeffizient</p>

Tabelle 4.10: Zusammenhang von ordinalen Merkmalen