Formelsammlung Statistik

Andrey Behrens

August 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.

Teil I.

Vorspann

1. Begriffe

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

2. Variablen

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 *nicht*
$x_i^{'}$		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	dxi	Klassenbreite der i-ten Klasse
s_i	si	relative Summenhäufigkeit einer Klasse
N	n	Statistische Masse, also die Menge aller Merkmalsausprägungen.

3. Eindimensionale Häufigkeitsverteilung

3.1. Beispiele

Gruppiert: Für nominale und ordinale Werte

x_i	h_i	H_i	f_i	F_i	$\triangle x_i$	f_i^*	h_i^*
280	1	1	0,1	0,1	-	-	-
340	2	3	0,2	0,3	-	-	-
560	1	4	0,1	0,4	-	-	-
600	1	5	0,1	0,5	-	-	-
650	3	8	0,3	0,8	-	-	-
740	1	9	0,1	0,9	-	-	-
1180	1	10	0,1	1,0	-	-	-

Klassiert: Für metrische Werte

x_i	h_i	H_i	f_i	F_i	$\triangle x_i$	f_i^*	h_i^*
[200;400)	21	21	0,21	0,21	200	0,00105	0,1050
[400;700)	56	77	0,56	0,77	300	0,00187	0,1867
[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

3.2. Formeln:

Name	Math		Formel	TR
abs. Häufigkeit	h_i	hi	-	-
abs. Summenhäufigkeit	H_i	shi	$h_1 + ... + h_i = \sum_{j=1}^{i} h_j$	cusum(hi)
relative Häufigkeit	f_i	fi	$\frac{h_i}{N}$ mit $\sum_{i=1}^{k} f_i$	$relhfg(hi)$
abs. Summenhäufigkeit	F_i	sfi	$f_1 + ... + f_i = \sum_{j=1}^{i} f_j$	$cumsum(relhfg(hi))$
Stat Masse	N	n	$\sum_{i=1}^{k} h_i$	$sum(hi)$
abs Häufigkeitsdichte	h_i^*	his	$\frac{h_i}{\Delta x_i}$	his
rel Häufigkeitsdichte	f_i^*	fis	$\frac{f_i}{\Delta x_i}$	fis

3.3. Funktion der relatitiven Summenhäufigkeit[1]

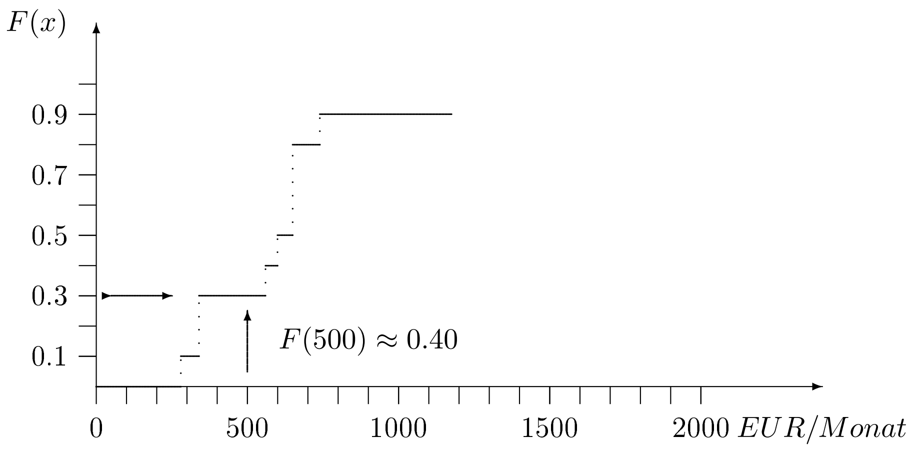
3.3.1. Bei gruppierte Daten

$$F(x) = \begin{cases} 0 & x < x_1 \\ F_i & x_i \leq x < x_{i+1} \\ 1 & x \geq x_k \end{cases}$$

Als Rechenbeispiel:

F(500)=0,30 -> Es wird nicht gerechnet, sondern aus dem Diagramm abgelesen, da es sich um gruppierte Werte handelt!

Als grafische Lösung (Treppendiagramm, keine Zwischenwerte!)



3.3.2. Bei klassierten Daten

$$F(x) = \begin{cases} 0 & x < x_1^u \\ F(x_i^u) + \frac{f_i}{\Delta x_i} * (x - x_i^u) & x_i^u \leq x < x_i^o \\ 1 & x \geq x_k^o \end{cases}$$

als Rechenbeispiel:

1. Klasse aus Diagramm ablesen (H_i), untere und obere Grenzen der Klasse herauslesen.

2. In Formel einsetzen: $F(500) = 0.21 + \frac{0{,}56}{300}(500 - 400) = 0,397 = 39,7\%$

als grafische Lösung siehe Diagramm Verteilungsfunktion bei klass Daten

3.3.3. Stabdiagramm

=relative Häufigkeitsfunktion für klassierte Daten. Numerische Lösung ist f_i die grafische Lösung ein Histogramm über h_i^*

[1] Auch als Verteilungsfunktion bezeichnet

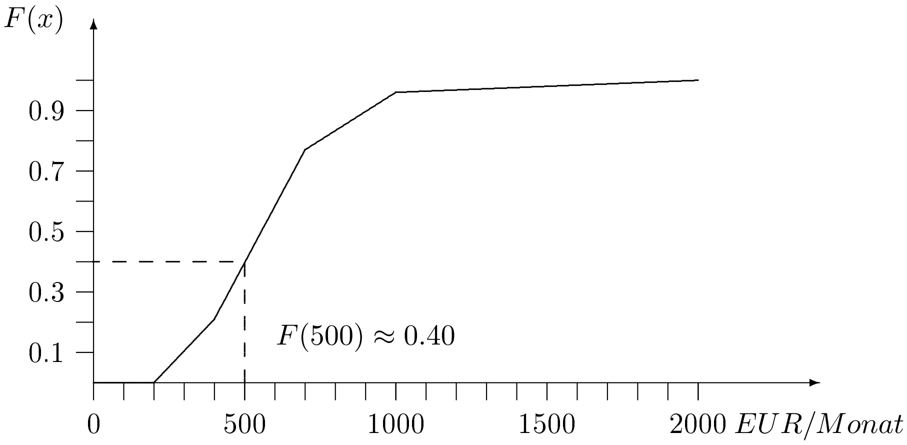


Abbildung 3.1.: Verteilungsfunktion bei klass. Daten

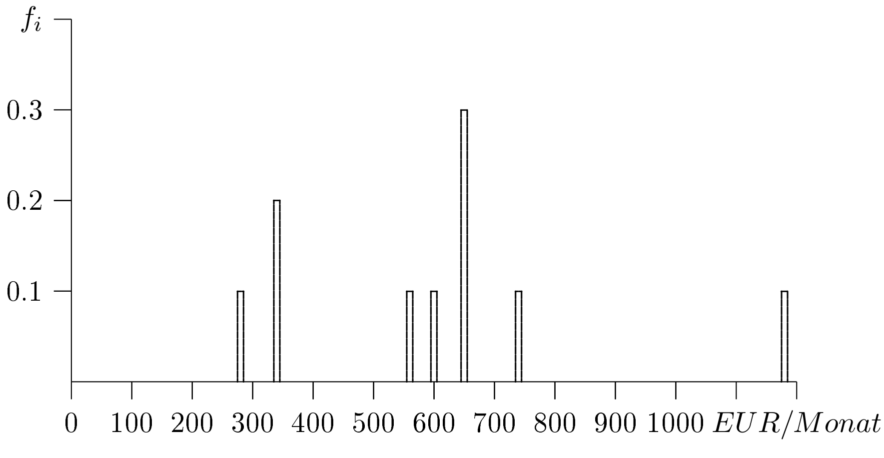


Abbildung 3.2.: Stabdiagramm

3.3.4. Histogramm

=relative Häufigkeitsfunktion für klassierte Daten. Numerische Lösung ist f_i die grafische Lösung ein Histogramm über f_i^*

Zur Darstellung der Klassenhäufigkeit (f_i^* oder h_i^*) einer Klasse

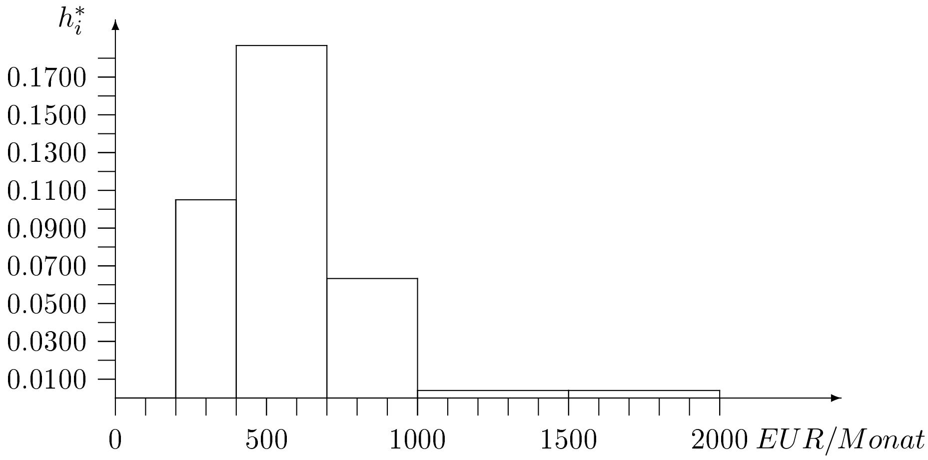


Abbildung 3.3.: Histogramm

3.4. Lageparameter

Name	Math	TR	Beschreibung	
Modal	x_D	xd		
			Gruppiert	$x_D = x_i$ mit $f_i \to max$
			Klassiert	$x_D = \frac{x_i^u + x_i^o}{2} = x_i'$ mit $h_i^* \to max$
Median	x_z	xz		
			Gruppiert	$x_z = 0,5N$
Quantil	x_p	xp		
Arith. Mittelw.	$\bar{x}$	xs		
Geom Mittelw.	x_G	xg		

3.4.1. Modalwert (Modus)

= die Stelle häufigste Merkmalsausprägung

Gruppen da x_i wo f_i am größsten ist: $x_D = x_i$ mit $f_i \to max$

Klassen Mitte der modalen Klasse: $x_D = \frac{x_i^u + x_i^o}{2} = x_i'$ mit $h_i^* \to max$

3.4.2. Median (Zentralwert)

= Mitte aller Merkmalsträger, bzw. welcher Merkmalswert wird von der Hälfte aller Merkmalsträger nicht überschritten.

Gruppe $x_z = 0.5N$ aber: wenn N gerade, dann Mittelwerte von aktueller Gruppe und nächster Gruppe (im Beispiel: 625).

$$x_z = \begin{cases} x_{(k)} & p*N \notin Z \, mit \, k = p*N < k < p*N+1 \, und \, k \in Z \\ \frac{x_{(k)}+x_{(k+1)}}{2} & p*N \in Z \, mit \, k = p*N \end{cases}$$

Klasse $x_z = x_i^u + \frac{0.5-F(x_i^u)}{f_i} * \Delta x_i$ Beispiel: Zuerst Klasse bestimmen und dann

$400 + \frac{0.5-0.21}{0.56} * 300 = 555.36 \, EUR$

3.4.3. Quantile

= ein Teil aller Merkmalsträger (etwa 0,25x oder 0,75x) bzw. welcher Merkmalswert wird von einem Teil aller Merkmalsträger nicht überschritten. Dabe ist das $x_p = x_{0.5} = x_z$

Gruppe $x_p = p * N$ Wobei p das Quantil ist, etwa 0,5, 0,75 oder 0,25. aber: wenn N gerade, dann Mittelwerte von aktueller Gruppe und nächster Gruppe (im Beispiel: 625).

$$x_p = \begin{cases} x_{(k)} & p*N \notin Z \, mit \, k = p*N < k < p*N+1 \, und \, k \in Z \\ \frac{x_{(k)}+x_{(k+1)}}{2} & p*N \in Z \, mit \, k = p*N \end{cases}$$

Klasse $x_p = x_i^u + \frac{p-F(x_i^u)}{f_i} * \Delta x_i$ Beispiel: Zuerst Klasse bestimmen und dann

$400 + \frac{0.5-0.21}{0.56} * 300 = 555.36 \, EUR$

3.4.4. Arithmetischer Mittelwert

= Durchschnitt

Nur für metrische Werte geeignet

Gruppe $\bar{x} = \frac{\sum_{i=1}^{k} h_i * x_i}{N} = \sum_{i=1}^{k} x_i f_i$

Klasse $\bar{x} = \frac{\sum_{i=1}^{k} h_i * x_i'}{N} = \sum_{i=1}^{k} x_i' f_i$

Addition $\bar{x} = \frac{\sum_{m=1}^{k} N_m * \bar{x_m}}{\sum_{m=1}^{k} N_m}$ wobei i i-te Variante der zu addierenden Durchschnitte ist

3.4.5. Geometrischer Mittelwert

=Mittelwert für Wachstumsfaktoren

Gruppe $x_G = \sqrt[N]{\prod_{i=1}^{k} x}$

3.5. Streuungsparameter

3.5.1. Spannweite

Abstand zw. größter und kleinster Merkmalsausprägung

Gruppiert $R = x_{max} - x_{min}$

Klassiert $R = x_k^o - x_1^u$

3.5.2. Quartilsabstand

Abstand zwischen oberem und unterem Quartil $Q = x_{0.75} - x_{0.25}$

3.5.3. Varianz

mittlere quadratische Abweichung aller Merkmalsausprägungen vom arith. Mittelwert

Gruppiert $s_x^2 = \frac{1}{N} \sum_{i=1}^{k} \left[(x_i - \overline{x})^2 \cdot h_i \right] = \sum_{i=1}^{k} \left[x_i^2 \cdot f_i \right] - \overline{x}$

Klassiert $s_x^2 = \frac{1}{N} \sum_{i=1}^{k} \left[(x_i^{'} - \overline{x})^2 \cdot h_i \right] = \sum_{i=1}^{k} \left[(x_i^{'})^2 \cdot f_i \right] - \overline{x}^2$

3.5.4. Standardabweichung

=mittlere Abweichung vom Mittelwert

$s_x = \sqrt{s_x^2}$

3.5.5. Variationskoeffizient

$v = \frac{s_x}{\overline{x}}$

3.6. Relative Konzentration

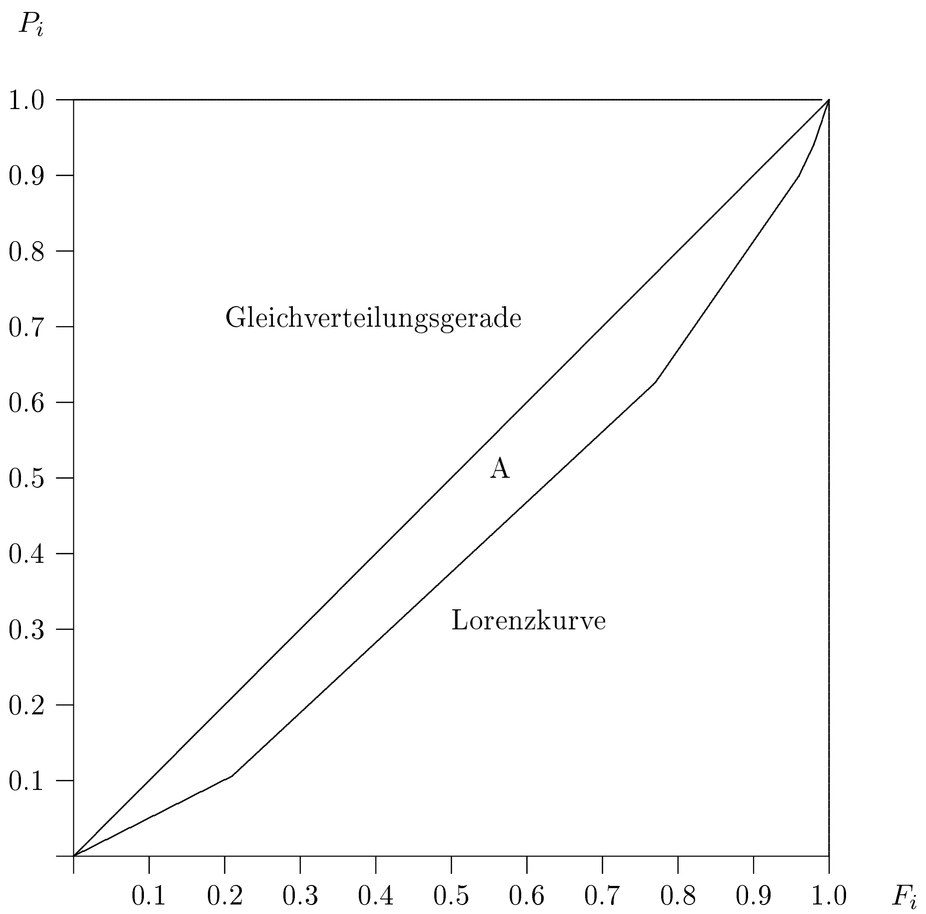
3.6.1. Berechnung

=konzentrieren sich Merkmalssumme auf wenige Merkmalsträger?

Konzentrationskoeffizient $p_i = \frac{x_i \cdot hi}{N \cdot \bar{x}}$

Konzentrationsmaß $P_i = \sum_{j=1}^{i} p_j$

3.6.2. Lorenzkurve



Part II.

Anhang

4. Quellen

(1) Statistikscript Prof. Dr. Müller, HS Wismar

(2) Taschenbuch der Wirtschaftsmathematik, Wolfgang Eichholz und Eberhard Vilkner

Part III.

Formblätter

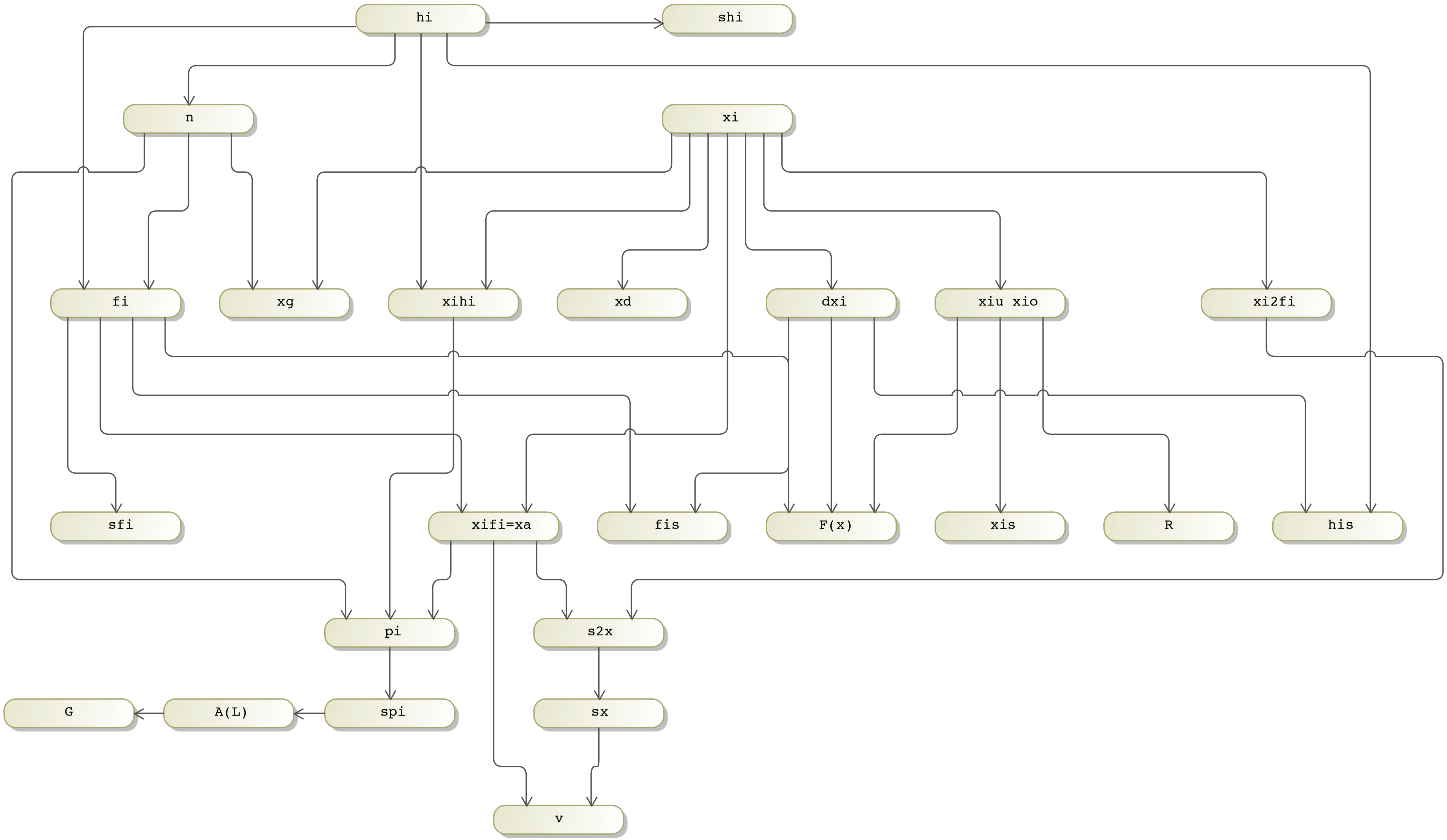


Figure 4.1.: Abhängigkeiten Statistik

Gruppe x_i	abs. Häufig h_i	abs. Summenhäufig H_i	rel. Häufig f_i	rel. Summenhäufig F_i	$x_i f_i$	$x_i \cdot h_i$	$x_i^2 \cdot h_i$	Konz- koeff. p_i	Konz- maß P_i	Fläche unter Lorenzkurve $A(L)$
$\sum$	$N =$	-	$= 1$	-	$\bar{x}=$	$=$	$=$	-	-	$=$

Gruppe x_i	$x_i^{'}$	$\triangle x_i$	abs. Häufig h_i	abs. Summenhäufig H_i	rel. Häufig f_i	rel. Summenhäufig F_i	$x_i f_i$	$x_i \cdot h_i$	$x_i^2 \cdot h_i$	h_i^*	f_i^*	Konz-koeff. p_i	Konz-maß P_i	Fläche unter Lorenzkurve $A(L)$
$\sum$			$N =$	-	$= 1$	-	$\bar{x}=$	$=$	$=$			-	-	$=$