

utb.

Peter Schmidt

Statistik schrittweise verstehen

Lehr- und Arbeitsbuch





Aufgaben

Ü 2-1 Geben Sie für die folgenden Merkmale die Merkmalsträger sowie mögliche Ausprägungen an: a) Hunderasse b) Schulden c) Klausurnoten d) Längen e) Sportart f) Mitarbeiterzahl. [3 Min.]

Ü 2-2 Auf welcher Skala sind die folgenden Merkmale gemessen: a) Tachostand b) Einkommen c) Schulnote d) Uhrzeit e) Gewicht f) Ligatabelle g) Religion h) Alter i) Familienstand j) Getreidesorte k) Handygebühren l) Güteklassen m) Temperatur? [3 Min.]

Ü 2-3 Welche Merkmale sind diskret, welche stetig: a) Wasserstandsanzeige b) Anzahl Vereinsmitglieder c) Kraftstoffverbrauch (in Liter) d) Ausschusstücke pro Tag e) Weinkonsum (in Liter) f) Weinkonsum (in Fl.) g) Einwohner h) Tage bis zur Prüfung? [3 Min.]

K 2-4 Ordnen Sie die folgenden Merkmale den unterschiedlichen Skalierungen und Merkmalstypen zu. Einkommen - Haarfarbe - Alter - soziale Stellung - Körpergröße - Abweichungen von der Norm bei Fertigungsprozessen - Geschlecht - Beruf - Schultypen - Anzahl von Kindern in Schulklassen - Raucher/Nichtraucher - Altersklassen.

Bei Einordnung in mehrere Klassen bitte Erläuterung bzw. Beispiel. [6 Min.]

↓ Merkmale Skalen →	nominal	ordinal	metrisch
Typen diskret			
stetig			

Lernschritt C – Häufigkeiten und Konzentrationsmessung

Fallbeispiel | **StudierBar**

Chris hat sich die Zahlen noch einmal angesehen. Die einzelnen Angaben, wie viel Kaffee die Befragten trinken, sind schon einmal ganz hilfreich, aber wie verteilen sich die Antworten? Sie würde gerne wissen, wie viele Studierende höchstens 5 Tassen in der Woche trinken, wie viele bis zu 15. Welches Maß ist dafür sinnvoll?

Und sie interessiert sich für die Verteilung: wie viel Prozent der Studis trinken wie viel Prozent des Kaffees? Damit möchte sie Gelegenheitstrinker von Stammkunden unterscheiden (was im Marketing gerade besprochen wurde).

2.1.3 Jetzt mal zusammengerechnet: Summenhäufigkeiten

Neben den bisher kennengerlernten Häufigkeiten betrachten wir nun **Summenhäufigkeiten**, auch als „**kumulierte**“ Häufigkeiten bezeichnet. Es geht dabei um die Frage: Wie viele der Antworten weisen eine Merkmalsausprägung *bis zu* einem bestimmten Wert auf? Zum Beispiel: Wie viele der Befragten trinken bis zu (höchstens) 5 Kaffee in der Woche, wie viele bis zu 15 usw.?

Summenhäufigkeiten werden mit den Großbuchstaben H für absolute und F für relative Summenhäufigkeit bezeichnet.

Absolute Summenhäufigkeit: $H(x_j) = \sum_{i=1}^j h(x_i) = \sum_{i=1}^j n_i$ (2-6)

Relative Summenhäufigkeit: $F(x_j) = \sum_{x_i \leq x_j} f(x_i) = \sum_{i=1}^j f(x_i)$ (2-7)

Zur Ermittlung der unterschiedlichen Häufigkeitskonzepte betrachten wir hier den Kaffee Konsum von 32 Gästen. Zur einfacheren Bearbeitung sind die Arbeitstabelle und die Rahmen der Grafiken vorgegeben:

Tassen Kaffee pro Woche							
21	8	7	0	10	10	15	13
0	4	7	9	12	6	18	11
12	0	22	0	12	16	18	0
9	17	21	0	12	8	6	5

Klasse		Häufigkeiten				
Unter- grenze	Ober- grenze	x_i^*	h_i oder n_i	H_i	f_i %	F_i
		0			0	0
0	5					
6	10					
11	15					
16	20					
21	25					
		Summe:		--		--

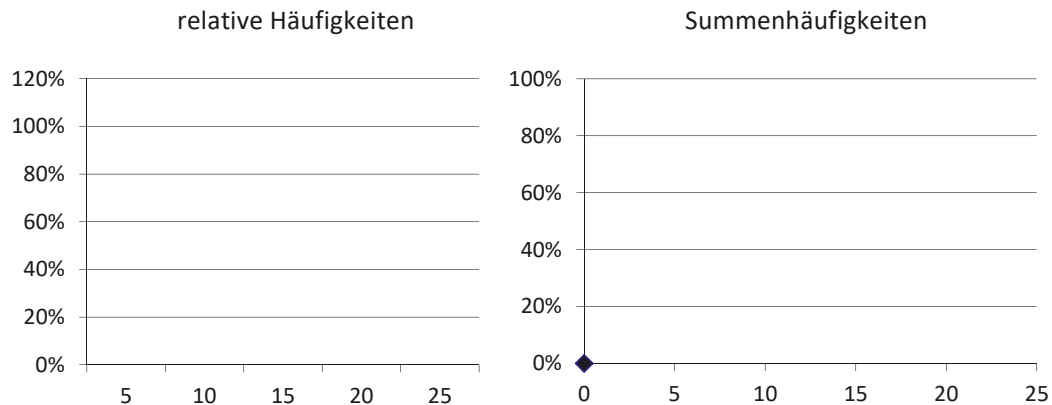


Tabelle 4: Absolute und relative (Summen-)Häufigkeiten für klassierte Daten

2.1.4 Wer hat wie viel?

Konzentrationsmessung (Lorenzkurve / Lorenz'sches Konzentrationsmaß)

Wenn es um „Verteilung“ geht, lautet die Frage oft, wie gleich (= gerecht?) ein bestimmtes Merkmal (z.B. Einkommen, Vermögen ...) auf die Träger dieses Merkmals (Arbeitnehmer, Steuerzahler ...) aufgeteilt ist. Hier hilft die **Lorenzkurve**:

→ Wie viel Prozent der Merkmalsträger verfügen über wie viel Prozent der Merkmalsumme?



Beispielaufgabe⁴

Eine Befragung von 70 Studierenden ergab, dass diese insgesamt gut 800 Tassen Kaffee pro Monat in der *StudierBar* trinken wollen. Chris bildet Trinker-Klassen und zählt, wie viele Studis in die jeweiligen Klassen fallen und wie hoch der Durchschnittsverbrauch in den Klassen ist.

Klasse Anzahl Kaffeetassen pro Monat von ... bis unter ...	n_i Anzahl Studierende (die in diese Klasse fallen)	x_i^* Durchschnittsverbrauch je Studierenden (Klassenmitte)
unter 1	17	0,5
1 – 5	25	3
5 – 10	14	7,5
10 – 50	9	30
50 – 90	5	70

Um *Konzentration* zu messen (konzentriert sich der Umsatz auf wenige KundInnen – oder ist er relativ gleich verteilt?), wird gemessen, wie viel Prozent der Merkmalsträger über wie viel Prozent der Merkmalsumme verfügen – indem erstere auf der waagrechten und letztere auf der senkrechten Achse eines Koordinatensystems abgetragen werden.

⁴ Die typischen Anwendungen einer Lorenz-Analyse sind die Verteilung von Einkommen, Vermögen, Umsatz in einem Markt. Wir gehen hier den nächsten Schritt jedoch weiter mit gemütlichem Kaffee.

Die Formeln für die Lorenzkurve wirken nicht auf den ersten Blick „einleuchtend“. Wenn sie aber *Schritt für Schritt* angewandt werden, erschließt sich das Vorgehen:

Merkmalssumme der *einzelnen* Merkmalsausprägungen: $m_j = x_j \cdot n_j$ (2-8)

Merkmalssumme *aller* Merkmalsausprägungen: $m = \sum_{j=1}^k m_j$ (2-9)

Relative Merkmalssumme: $g_j = \frac{m_j}{m}$ (2-10)

Kumulierte relative Merkmalssumme: $G_j = \sum_{i=1}^j g_i$ (2-11)

Einzelfläche unter der Lorenzkurve: $Fl_j = (f_j \cdot G_{j-1}) + \frac{f_j \cdot g_j}{2}$ (2-12)

Gesamtfläche unter der Lorenzkurve: $Fl = \sum Fl_j$ (2-13)

**Lorenz'sches Konzentrationsmaß (LKM)
oder auch „Gini-Koeffizient“**

$$\text{LKM} = 1 - \frac{Fl}{5000} \quad (2-14)$$

In der folgenden Arbeitstabelle zeichnen wir die Konzentrationskurve für den Kaffeekonsum und errechnen das LORENZ'sche Konzentrationsmaß (= Gini-Koeffizient).

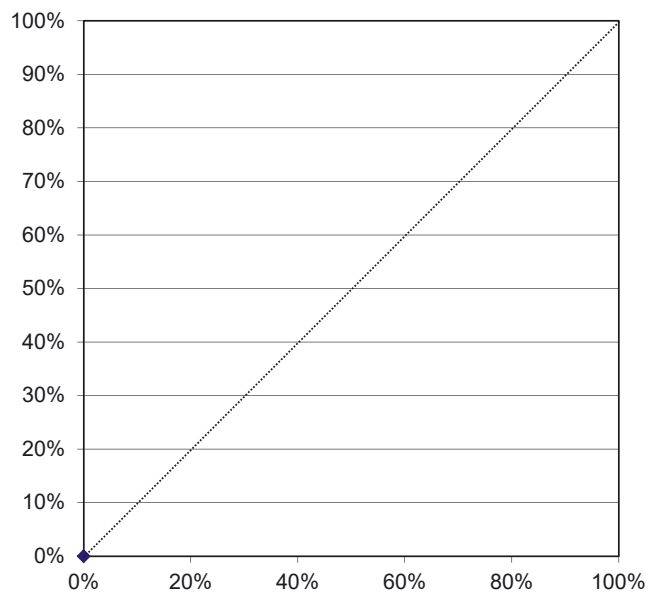
Arbeitstabelle LKM und Lorenzkurve zum Kaffeeverbrauch/Monat:

Klassen	i	n _i	f _i %	F _i	Klassenmitte x _i [*]	vorgegebene Werte		Fläche lt. Formel			
						Summe aller Ausprägungen in x _i [*] · n _i → m _i	Anteil Umsatz g _i %	G _i	f _i · G _{i-1}	$\frac{f_i \cdot g_i}{2}$	Fl _i
				0				0			
unter 1	1	17			0,5						
1 bis 5	2	25			3						
5 bis 10	3	14			7,5						
10 bis 50	4	9			30						
50 bis 90	5	5			70						
Summen:		= n					= m			Fl =	
										5000 - Fl =	

Wie viel Prozent der Merkmalsträger verfügen über wie viel Prozent der Merkmalsumme?

LKM	
=	

Lorenzkurve:



oder auch GINI-Koeffizient genannt

Tabelle 5: Ermittlung der Lorenzkurve und des LKM (Gini-Koeffizient)