2.3 Maßzahlen für klass. Daten

2.3.1. Das arithm. Mittel

Sei X ein diskretes Merkmal:

$$X = \{a_1, \ldots, a_k\}$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i =$$

$$= \frac{1}{n} \left(x_{a_{1_1}} + x_{a_{1_2}} + x_{a_{1_3}} + \right.$$

$$\left. + \dots + x_{a_{1_{i_1}}} \right) +$$

$$+ \left(x_{a_{2_1}} + x_{a_{2_2}} + \dots + x_{a_{2_{i_2}}} \right) +$$

$$+ \dots +$$

$$+ \left(x_{a_{k_1}} + x_{a_{k_2}} + \dots + x_{a_{k_{i_k}}} \right) (=)$$

Wobei:

$x_{a_{j_l}}$ die l-te Beobachtung vom Wert a_j in der Stichprobe ist

$$= \frac{1}{n}\Big[\underbrace{(a_1 + a_1 + \dots + a_1)}_{i_1 - \text{mal}} +$$

$$i_1 = H_n(a_1)$$

$$+ \underbrace{(a_2 + a_2 + \dots + a_2)}_{i_2 - \text{mal}} +$$

$$i_2 = H_n(a_2)$$

$$+ \dots + a_k \cdot \underbrace{i_k}_{H_n(a_k)} \Big] =$$

$$= \frac{1}{n}\left[a_1 \cdot H_n(a_1) + \dots + a_k \cdot H_n(a_k)\right]$$

$$= \frac{1}{n}\sum_{\ell=1}^{k} a_\ell \, H_n(a_\ell) =$$

$$= \sum_{\ell=1}^{k} a_\ell \, h_n(a_\ell)$$

Bsp: Studenten & Noten.

MN	gender	Lungen-
x_1	♂ 1	
x_2		♀ 2
x_3		♀ 2
x_4	♂ 1	
x_5	1	
x_6	1	
x_7		2
x_8		2

$$\frac{1}{8}\sum_{k=1}^{8} x_k = \frac{1}{8}\left(x_1 + x_2 + \dots + x_8\right) =$$

$$= \frac{1}{8}\left(x_1^{(1)} + x_2^{(2)} + x_3^{(2)} + x_4^{(1)} + x_5^{(1)} + x_6^{(1)} + x_7^{(2)} + x_8^{(2)}\right) =$$

$$= \frac{1}{8}\left(\left[x_1^{(1)} + x_4^{(1)} + x_5^{(1)} + x_6^{(1)}\right] + \left[x_2^{(2)} + x_3^{(2)} + x_7^{(2)} + x_8^{(2)}\right]\right)$$

Ist X ein stetiges Merkmal, liegen die Daten sowohl als Urliste als auch in klassierter Form vor, dann gilt:

$$\bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j = \frac{1}{n}\sum_{i=1}^{k} \bar{x}_i H_n(k_i),$$

wobei $\bar{x}_i = \frac{1}{H_n(k_i)} \sum_{x_j \in K_i} x_j$

der Mittelwert der Klasse i ist.

Liegt die Urliste nicht mehr vor, so ersetzt man die Klassenmittelwerte durch Klassenmitte x_i' als Approximation:

$$\bar{x}' = \frac{1}{n} \sum_{i=1}^{k} x_i' \cdot H_n(K_i)$$

Approximation des arithm. Mittels

Approximation des arithmetischen Mittels

Beispiel : Wie groß ist die mittlere Lebensdauer von 1000 Maschinen, deren Lebensdauerverteilung durch folgende Tabelle gegeben ist:

i		K_i	x_i'	$H_n(K_i)$
1	[0,2]	bis 2 Jahre	1	33
2		(2-4] Jahre	3	276
3		(4-6] Jahre	5	404
4		(6-8] Jahre	7	237
5		(8-10] Jahre	9	50

Tab. 2.1

4^* $(6-10]^*$ 8 $237+50 = 287$

1) 5 Klassen:

$$\bar{x}_5' = \frac{33\cdot 1 + 276\cdot 3 + 5\cdot 404 + 7\cdot 237 + 9\cdot 50}{1000} = 4{,}99 \approx$$

≈ 5 Jahre

2) 4 Klassen!

$$\bar{x}_4' = \frac{33 + \overbrace{276\cdot 3}^{828} + \overbrace{5\cdot 404}^{2020} + \overbrace{287\cdot 8}^{2296}}{1000} = 5{,}17 \approx$$

$\approx 5{,}2$ Jahre

2.3.2. Die Streuung

Für diskrete X und

$X = \{a_1, \dots, a_k\}$

gilt:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{k} (a_i - \bar{x})^2 H_n(a_i)$$

Betrachten wir nun

klassifizierte Daten:

Klasse $K_i \longleftrightarrow$ Klassenmittelwert $\overline{x_i}$.

Dann:

$$s_0^2 = \frac{1}{n-1} \sum_{i=1}^{k} \left(\overline{x_i} - \overline{x}\right)^2 \cdot H_i(k)$$

sind die Klassenmittelwerte nicht bekannt ersetzt man sie durch die

Klassenmitten

$$a_i' = \frac{a_i^u + a_i^o}{2},$$

und $\bar{x}$ wird durch $\bar{x}'$ approximiert. Dann erhält man folgende Approximation für die Streuung:

$$s_*^2 = \frac{1}{n-1} \sum_{i=1}^{k} (a_i' - \bar{x}')^2 H_n(k_i)$$

Aus Untersuchungen:

S_*^2 überschätzt die Streuung δ^2, wenn die Klassenzahl k klein im Vergleich zum Stichprobenumfang n ist.

Es hilft uns

Sheppard'sche Korrektorformel:

Sheppardsche Korrektur:

Haben alle Klassen die gleiche Klassenbreite B, so läßt sich der Wert s_*^2 korrigieren:

$$s_{**}^2 = s_*^2 - \frac{B^2}{12},$$

wobei:

$$s_*^2 = \frac{1}{n-1} \sum_{i=1}^{k} \left(a_i' - \bar{x}'\right)^2 \cdot H_n(K_i)$$

Bem: Bei Verwendung

der klass. Daten berechnet

man s_*^2 leichter:

$$s_*^2 = \frac{1}{n-1}\left[\sum_{i=1}^{k} (a_i)^2 H_n(k_i) - n\bar{x}'^2\right]$$

Beispiel: Wie groß sind die mittlere Lebensdauer und die Streuung von 30 Glühbirnen?

i	Klasse K_i	$H_n(K_j)$	a_i'
1	[600, 700)	6	~~500~~ 650
2	[700, 800)	7	750
3	[800, 900)	6	850
4	[900, 1000)	6	950
5	[1000, 1100)	5	~~1250~~ 1050
	Tab. 2.2	30	

$$\bar{x}' = \frac{1}{30} \sum_{i=1}^{5} a_i' \cdot H_n(k_i) =$$

$$= \frac{1}{30} \left(650 \cdot 6 + 750 \cdot 7 + 850 \cdot 6 + 850 \cdot 6 + 1050 \cdot 5 \right) = 840$$

$$S_*^2 = \frac{1}{n-1} \left[\sum_{i=1}^{k} (a_i')^2 H_n(k_i) - (\bar{x}')^2 n \right] =$$

$$= \frac{1}{29} \left[650^2 \cdot 6 + 750^2 \cdot 7 + 850^2 \cdot 6 + 850^2 \cdot 6 + 1050^2 \cdot 5 - 840^2 \cdot 30 \right] = 19551; \quad 51$$

$$S_* = 139{,}8 \simeq 140$$

$$k = \frac{S_*^2}{} = \frac{B^2}{12} = \frac{10^4}{12} \simeq 833$$

$$S_{**}^2 = S_*^2 - k = 19515 - 833 = 18682;$$

$$S_{**} = 137;$$

$$S_*^2 = \frac{1}{n-1} \sum_{i=1}^{k} \left(a_i' - \bar{x}' \right)^2 H_n(k_i);$$

$$S_*^2 = \frac{1}{29} \left((650 - 840)^2 \cdot 6 + (750 - 840)^2 \cdot 7 + (850 - 840)^2 \cdot 6 + \right.$$

$$\left. + (850 - 840)^2 \cdot 6 + (1050 - 840)^2 \cdot 5 \right) = 19551$$

Die Quantile

Wenn die Urliste nicht vorliegt, oder nur die Klassenhäufigkeitstabelle, verwenden wir die Klassensummenhäufigkeitsfunktion

$\tilde{F}_n(x)$,

um die Approximation $\tilde{x}_\alpha$ an α-Quantil x_α zu berechnen.

$F_n(x)$ → Anteil aller Daten x_i, $i = 1, \dots, n$, für die gilt:

$x_i \leq x$.

$\tilde{F}_n(x)$ näherungsweise $F_n(x)$ gleich ist.

x_α :

$F_n(x_\alpha) \approx \alpha$

$\tilde{x}_\alpha$ als Lösung von

$\tilde{F}_n(\tilde{x}_\alpha) = \alpha$.

1) Man bestimme die Klasse mit der Nummer i, in der das α-Quantil liegt:

$$h(i-1) < \alpha \leq h(i)$$

2)

$$\tilde{F}_h(\tilde{x}_\alpha) = h(i-1) + h_h(k_i) \cdot \frac{(\tilde{x}_\alpha - q_i^u)}{(q_i^o - q_i^u)} =$$

$$= \alpha$$

stelle wir nach $\tilde{x}_\alpha$ um:

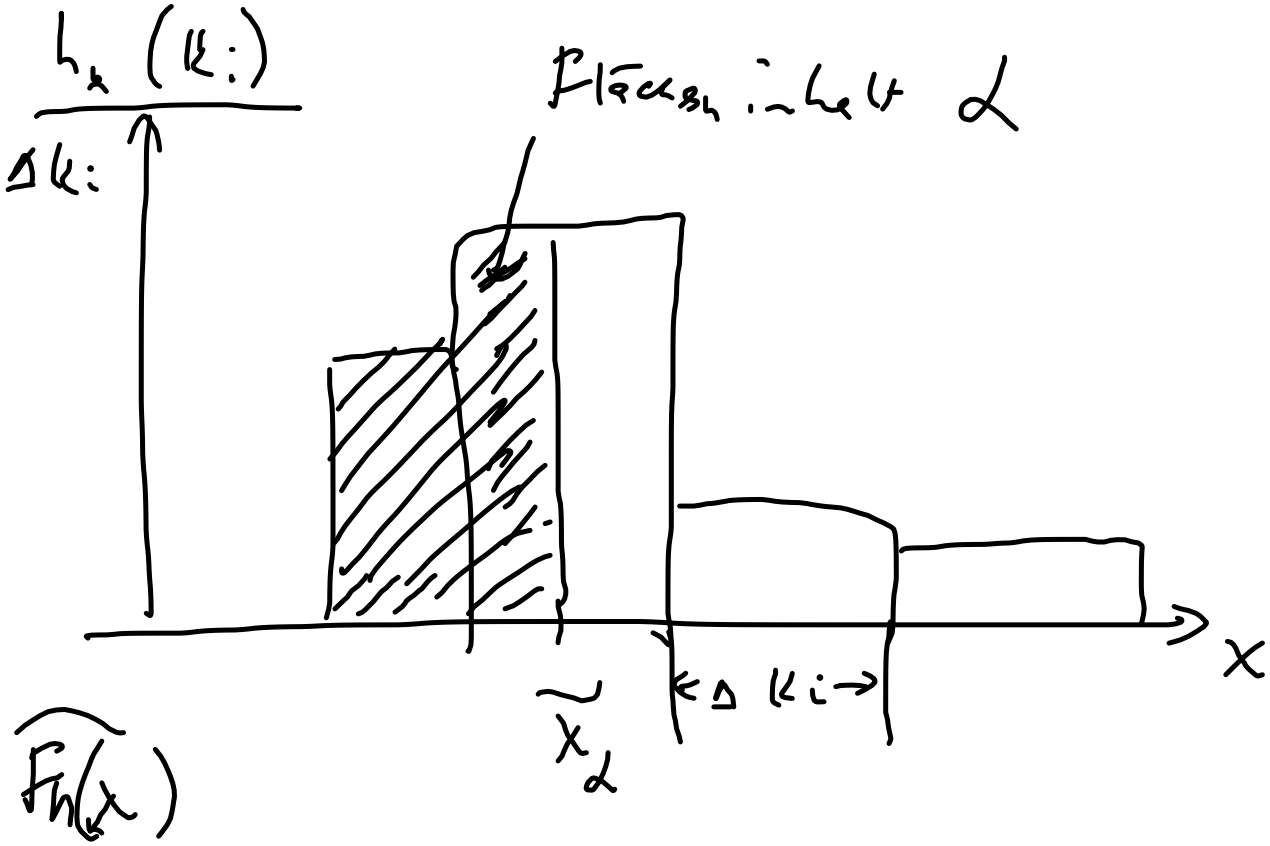
$$\tilde{x}_\alpha = a_i^u + \frac{\alpha - h(i-1)}{h_i(k_i)}\left(a_i^u - a_i^o\right)$$

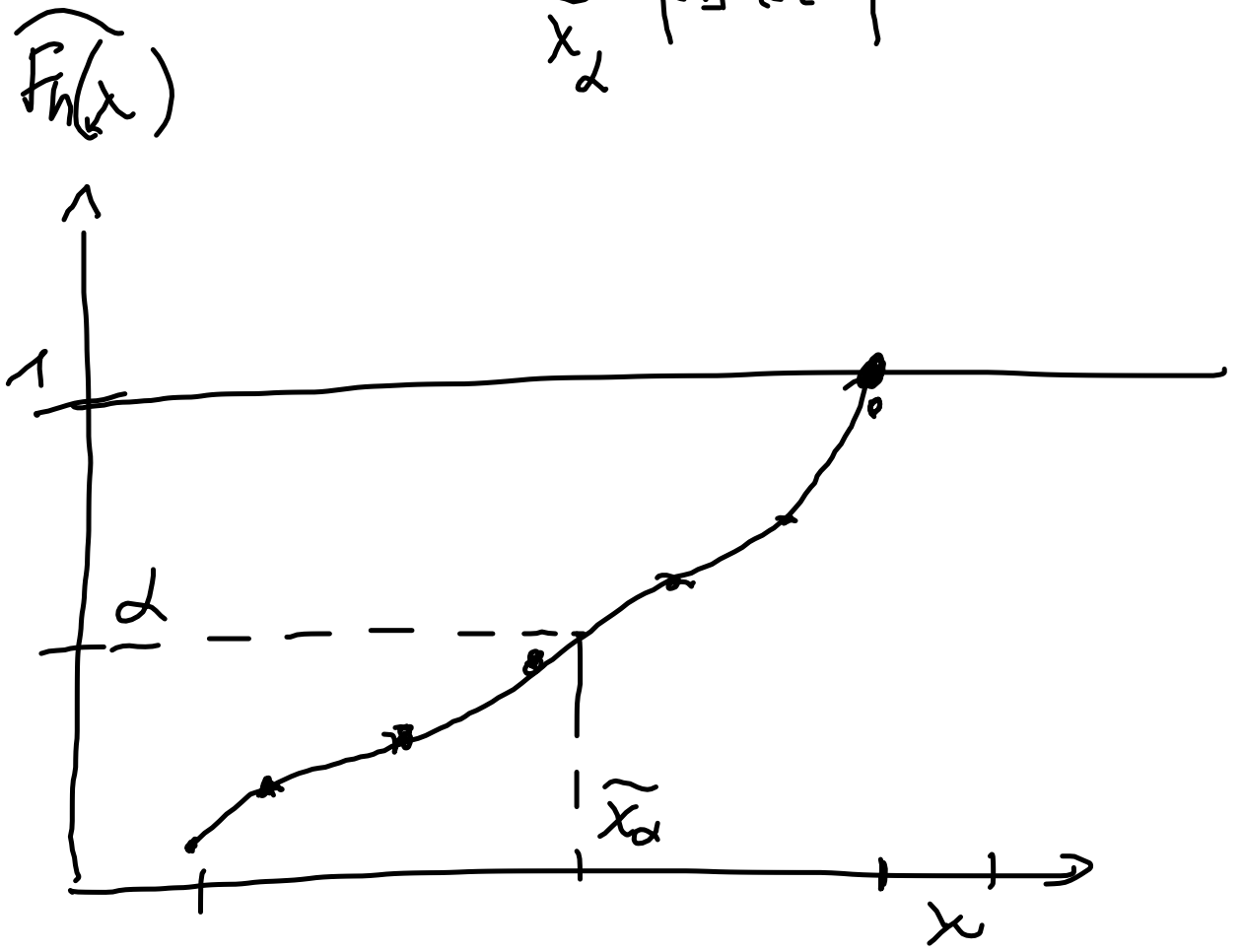
wobei

a_i^u untere und

a_i^o obere Grenze der

Klasse i ist.


$h_n(k_i)$
Flächeninhalt α
Δk_i
$\tilde{x}_\alpha$
Δk_i
x


$F_n(x)$
1
α
$\tilde{x}_\alpha$
x

Klasse Nr. i	Bruttomonatsverdienst in EURO	Beschäftigte		Summenhäufigkeiten Beschäftigte	
		Anzahl $H_n(K_i)$	Anteil $h_n(K_i)$	**Kumulierte Anzahl H(i)**	**Kumulierter Anteil h(i)**
1	500 bis 800	6	0,024	6	0,024
2	über 800 bis 1100	13	0,052	19	0,076
3	über 1100 bis 1400	22	0,088	41	0,164
4	über 1400 bis 1700	32	0,128	73	0,292
5	über 1700 bis 2000	40	0,160	113	0,452
6	über 2000 bis 2300	42	0,168	155	0,620
7	über 2300 bis 2600	39	0,156	194	0,776
8	über 2600 bis 2900	31	0,124	225	0,900
9	über 2900 bis 3200	20	0,080	245	0,980
10	über 3200 bis 3500	5	0,020	250	1,000
Σ		250	1,000		

Maße	Skala			
	Nominalskala	Ordinalskala	Intervallskala	Verhältnis-skala
Lagemaße				
Modus	X	X	X	X
Median α-Quantile		X	X	X
Arithm. Mittel		(X)	X	X
Streuungsmaße				
Spannweite Quantilabstände		X	X	X
Mittlere absolute Abweichung vom Median			X	X
Streuung Standardabweichung			X	X
Variationskoeffizient				X

Stab-, Blakendiagramm

Klasseneinteilung Histogramm