1. EINDIMENSIONALES DATENMATERIAL (1 MERKMAL)

1.1. Notation

n	Zahl der Merkmalsträger (Beobachtungswerte oder Messwerte)
k	Zahl der Merkmalsausprägungen bzw. Klassen (Cave: Abweichende Belegung bei Lorenzkurve)
r	Zahl disjunkter statistischer Massen
x_i	i-ter Beobachtungs- oder Messwert (i = 1;...; n)
a_j	j-te Merkmalsausprägung bzw. j-te Klasse der n Beobachtungs-/Messwerte (j = 1;...; k)
$h(a_j)$	= h_j = Absolute Häufigkeit der j-ten Merkmalsausprägung bzw. der j-ten Klasse (j = 1;...; k)
$f(a_j)$	= f_j = Relative Häufigkeit der j-ten Merkmalsausprägung bzw. der j-ten Klasse (j = 1;...; k)
H(x)	Kumulierte absolute Häufigkeit
F(x)	Kumulierte relative Häufigkeit (empirische Verteilungsfunktion)

1.2. Häufigkeiten

Absolute Häufigkeit der Merkmalsausprägung j: $h\left(a_j\right)$ $\quad j=1;....;k$

Relative Häufigkeit der Merkmalsausprägung j: $f(a_j)=\frac{1}{n}\cdot h\left(a_j\right)$ $\quad j=1;....;k$

Es gilt zwangsläufig: $\sum_{j=1}^{k} h\left(a_j\right)=n$ und $\sum_{j=1}^{k} f\left(a_j\right)=1$

Absolute kumulierte Häufigkeit: $H\left(x\right)=\sum_{a_j\le x} h\left(a_j\right)$

Relative kumulierte Häufigkeit: $F\left(x\right)=\sum_{a_j\le x} f\left(a_j\right)=\frac{1}{n}\cdot H\left(x\right)$

1.3. Lageparameter

Allgemein liegt das p-Quantil an der Stelle: $x_{(n-1)p+1}$ $\quad$ [z.B. $p=0{,}25$ oder $p=0{,}5$]

Median:

$$x_{Med}=\begin{cases} x_{\frac{n+1}{2}} & \text{; falls n ungerade} \\ \frac{1}{2}\cdot\left(x_{\frac{n}{2}}+x_{\frac{n}{2}+1}\right) & \text{; falls n gerade} \end{cases}$$

Arithmetisches Mittel (falls Urliste gegeben):

$$\bar{x}=\frac{x_1+x_2+...+x_n}{n}=\frac{1}{n}\cdot\sum_{i=1}^{n} x_i$$

Arithmetisches Mittel (falls Häufigkeitsverteilung gegeben / keine Klassierung):

$$\bar{x}=\frac{1}{n}\cdot\sum_{j=1}^{k} a_j\cdot h\left(a_j\right)=\sum_{j=1}^{k} a_j\cdot f\left(a_j\right)$$

Arithmetisches Mittel (falls klassierte Daten):

$$\bar{x}^* = \frac{1}{n} \cdot \sum_{j=1}^{k} a_j^* \cdot h(a_j) = \sum_{j=1}^{k} a_j^* \cdot f(a_j)$$

$$\left[a_j^* = \text{Klassenmitte der j-ten Klasse} \right]$$

Gesamtmittel (= Arithmetisches Mittel einer Gesamtmasse):

$$\bar{x}_{Ges} = \frac{1}{n} \cdot \sum_{j=1}^{r} n_j \cdot \bar{x}_j$$

Harmonisches Mittel:

$$\bar{x}_H = \frac{\sum_{j=1}^{k} h(x_j)}{\sum_{j=1}^{k} \frac{h(x_j)}{x_j}}$$

Geometrisches Mittel:

$$\bar{x}_{Geom} = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n}$$

Falls c_1, ..., c_n Zuwachsraten, so gilt für die durchschnittliche Zuwachsrate $\bar{c}$:

$$\bar{c} = \sqrt[n]{(1+c_1) \cdot (1+c_2) \cdot \ldots \cdot (1+c_n)} - 1$$

1.4. Streuungsparameter

Spannweite:

$$SP = \max_i x_i - \min_i x_i \qquad i = 1; \ldots; n$$

Durchschnittliche Abweichung von λ:

$$\bar{s}_\lambda = \frac{1}{n} \cdot \sum_{i=1}^{n} | x_i - \lambda | = \sum_{j=1}^{k} | a_j - \lambda | \cdot f(a_j)$$

Zumeist verwendet man $\lambda = x_{Med}$ oder $\lambda = \bar{x}$

Varianz (falls Urliste gegeben):

$$s^2 = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2 = \underbrace{\left(\frac{1}{n} \cdot \sum_{i=1}^{n} x_i^2 \right) - \bar{x}^2}_{\text{Verschiebungssatz}}$$

Varianz (falls Häufigkeitsverteilung gegeben):

$$s^2 = \frac{1}{n} \cdot \sum_{j=1}^{k} (a_j - \bar{x})^2 \cdot h(a_j) = \sum_{j=1}^{k} (a_j - \bar{x})^2 \cdot f(a_j)$$

Varianz (falls klassierte Daten):

$$s^2 = \frac{1}{n} \cdot \sum_{j=1}^{k} (a_j^* - \bar{x})^2 \cdot h(a_j) = \sum_{j=1}^{k} (a_j^* - \bar{x})^2 \cdot f(a_j)$$

$$\left[a_j^* = \text{Klassenmitte der j-ten Klasse} \right]$$

Varianzzerlegungssatz:

$$s_{Ges}^2 = \underbrace{\frac{1}{n} \cdot \sum_{i=1}^{k} n_i \cdot s_i^2}_{\text{interne Varianz}} + \underbrace{\frac{1}{n} \cdot \sum_{i=1}^{k} n_i \cdot (\bar{x}_i - \bar{x}_{Ges})^2}_{\text{externe Varianz}}$$

Standardabweichung:

$$s = \sqrt{s^2}$$

Variationskoeffizient:

$$V = \frac{s}{\bar{x}}$$

1.5. Konzentrationsmessung

Lorenzkurve: Nicht-Klassierte Daten

Anteil an der gesamten Merkmalssumme, den die k „kleinsten“ Merkmalsträger auf sich vereinigen:

$$\nu_k = \frac{\sum_{i=1}^{k} x_i}{\sum_{i=1}^{n} x_i}$$

Anteil der k „kleinsten“ Merkmalsträger an der Gesamtzahl der Merkmalsträger:

$$\upsilon_k = \frac{k}{n}$$

Lorenzkurve: Klassierte Daten [Cave: Gesamtzahl der Klassen = l]

Anteil an der gesamten Merkmalssumme, den die k „kleinsten“ Klassen auf sich vereinigen:

$$\nu_k = \frac{\sum_{j=1}^{k} h(a_j) \cdot \text{Klassenmitte}_j}{\sum_{j=1}^{l} h(a_j) \cdot \text{Klassenmitte}_j}$$

Anteil der Merkmalsträger in den k „kleinsten“ Klassen an der Gesamtzahl der Merkmalsträger:

$$\upsilon_k = \frac{\sum_{j=1}^{k} h(a_j)}{n}$$

Gini-Koeffizient

Ansatz:

$$G = \frac{\text{Konzentrationsfläche}}{0{,}5}$$

Berechnung:

$$G = \frac{2 \cdot \sum_{i=1}^{n} i \cdot x_i - (n+1) \cdot \sum_{i=1}^{n} x_i}{n \cdot \sum_{i=1}^{n} x_i} = \frac{2 \cdot \sum_{i=1}^{n} i \cdot p_i - (n+1)}{n}$$

mit: $p_i = \frac{x_i}{\sum_{i=1}^{n} x_i}$ $[= \text{relative Merkmalswerte}]$

Normierter Gini-Koeffizient:

$$G^* = \frac{n}{n-1} \cdot G \quad \text{mit } G^* \in [0;1]$$

Weitere Konzentrationsmaße

Konzentrationsindex (Anteil der Merkmalssumme, den die g größten Merkmalsträger auf sich vereinen):

$$CR_g = 1 - L\left(\frac{n-g}{n}\right) \qquad [\text{absolute Werte gegeben}]$$

$$CR_g = 100 - L\left(100 \cdot \frac{n-g}{n}\right) [\text{Prozentwerte gegeben}]$$

Herfindahl-Index: $H = \sum_{i=1}^{n} p_i^2 \quad$ mit: $H \in \left[\frac{1}{n};1\right]$

2. ZWEIDIMENSIONALES DATENMATERIAL (2 MERKMALE)

2.1. Notation

n	Zahl der Merkmalsträger (Beobachtungswerte oder Messwerte)
k	Zahl der Merkmalsausprägungen bzw. Klassen der Variable X
l	Zahl der Merkmalsausprägungen bzw. Klassen der Variable Y
a_i	i-te Merkmalsausprägung bzw. i-te Klasse des Merkmals X (i = 1;...; k)
b_j	j-te Merkmalsausprägung bzw. j-te Klasse des Merkmals Y (j = 1;...; l)

2.2. Häufigkeiten

Gemeinsame Häufigkeiten: $h_{ij} = h\left(a_i ; b_j\right)$

Randhäufigkeiten des ersten Merkmals: $h_{i\bullet} = \sum_{j=1}^{l} h_{ij}$

Randhäufigkeiten des zweiten Merkmals: $h_{\bullet j} = \sum_{i=1}^{k} h_{ij}$

Bedingte relative Häufigkeit der Ausprägung a_i unter der Voraussetzung, dass b_j vorliegt: $f_1\left(a_i \mid b_j\right) = \frac{h_{ij}}{h_{\bullet j}}$

Bedingte relative Häufigkeit der Ausprägung b_j unter der Voraussetzung, dass a_i vorliegt: $f_2\left(b_j \mid a_i\right) = \frac{h_{ij}}{h_{i\bullet}}$

Falls X und Y unabhängig, so gilt:

$$f_1\left(a_i \mid b_j\right) = \frac{h_{ij}}{h_{\bullet j}} = \frac{h_{i\bullet}}{n} = \text{relative Häufigkeit von } a_i$$

$$f_2\left(b_j \mid a_i\right) = \frac{h_{ij}}{h_{i\bullet}} = \frac{h_{\bullet j}}{n} = \text{relative Häufigkeit von } b_j$$

sowie: $h_{ij} = \frac{h_{i\bullet} \cdot h_{\bullet j}}{n}$

2.3. Kovarianz

Kovarianz für n Merkmalswerte:

$$COV(X;Y) = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \bar{x}) \cdot (y_i - \bar{y})$$
$$= \frac{1}{n} \cdot \left(\sum_{i=1}^{n} x_i \cdot y_i \right) - \bar{x} \cdot \bar{y}$$

2.4. Korrelationskoeffizienten

(Bravais-Pearson-) Korrelationskoeffizient:

$$r_{BP} = \frac{\sum_{i=1}^{n} (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \cdot \sum_{i=1}^{n} (y_i - \bar{y})^2}} = \frac{COV(X;Y)}{s_x \cdot s_y}$$

Es gilt stets: $-1 \le r_{BP} \le 1$

Rangkorrelationskoeffizient von Spearman:

$$r_{Sp} = 1 - \frac{6 \cdot \sum_{i=1}^{n} \left(R_i - R_i^{'}\right)^2}{(n-1) \cdot n \cdot (n+1)}$$

wobei $R_i =$ Rangnummern der $x_i -$ Werte

$R_i^{'} =$ Rangnummern der $y_i -$ Werte

Es gilt stets: $-1 \le r_{Sp} \le 1.$

(Korrigierter) Kontingenzkoeffizient:

$$K_{Korr} = \frac{K}{K_{\max}} = \frac{\sqrt{\frac{\mathrm{X}^2}{n + \mathrm{X}^2}}}{\sqrt{\frac{M-1}{M}}}$$

Mit: $M = \min[k;l]$ $\quad \mathrm{X}^2 = \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{\left(h_{ij} - \tilde{h}_{ij}\right)^2}{\tilde{h}_{ij}}$

$$\tilde{h}_{ij} = \frac{h_{i\bullet} \cdot h_{\bullet j}}{n}$$

2.5. Lineare Regression

Regressionsgerade: $y = \hat{a} + \hat{b} \cdot x$

…mit den Parametern: $\hat{a} = \bar{y} - \hat{b} \cdot \bar{x}$

$$\hat{b} = \frac{\sum_{i=1}^{n} (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \frac{COV(X;Y)}{s_x^2}$$

3. INDEXZAHLEN

3.1. Notation

$p_0(1),...,p_0(n)$	Preise der Güter 1,..., n in der Basisperiode 0
$p_t(1),...,p_t(n)$	Preise der Güter 1,..., n in der Berichtsperiode t
$q_0(1),...,q_0(n)$	Mengen der Güter 1,..., n in der Basisperiode 0
$q_t(1),...,q_t(n)$	Mengen der Güter 1,..., n in der Berichtsperiode t

3.2. Indices

LASPEYRES-Preisindex:

$$P_{0t}^{L} = \frac{\sum_{i=1}^{n} p_t(i) \cdot q_0(i)}{\sum_{i=1}^{n} p_0(i) \cdot q_0(i)}$$

PAASCHE-Preisindex:

$$P_{0t}^{P} = \frac{\sum_{i=1}^{n} p_t(i) \cdot q_t(i)}{\sum_{i=1}^{n} p_0(i) \cdot q_t(i)}$$

LASPEYRES-Mengenindex:

$$Q_{0t}^{L} = \frac{\sum_{i=1}^{n} p_0(i) \cdot q_t(i)}{\sum_{i=1}^{n} p_0(i) \cdot q_0(i)}$$

PAASCHE-Mengenindex:

$$Q_{0t}^{P} = \frac{\sum_{i=1}^{n} p_t(i) \cdot q_t(i)}{\sum_{i=1}^{n} p_t(i) \cdot q_0(i)}$$