

## Formelsammlung zur Veranstaltung Statistik

### 1. Eindimensionale empirische Verteilungen

1.1 absolute Häufigkeit:  $h(a_j)$  ungruppiert bzw.  $h_j$  gruppiert

1.2 relative Häufigkeit:  $f(a_j)$  ungruppiert bzw.  $f_j$  gruppiert

1.3 absolute kumulierte Häufigkeiten:  $H(x) = \sum_{a_j \leq x} h(a_j)$

1.4 relative kumulierte Häufigkeiten:  $F(x) = \sum_{a_j \leq x} f(a_j)$

### 2. Zweidimensionale empirische Verteilungen

2.1 X mit realisierten Merkmalsausprägungen:  $a_i (i = 1, 2, \dots, k)$

Y mit realisierten Merkmalsausprägungen:  $b_j (j = 1, 2, \dots, l)$

2.2 gemeinsame absolute Häufigkeit:  $h_{ij} = h(a_i, b_j)$

2.3 Randhäufigkeiten  $h_{i.} = \sum_{j=1}^l h_{ij}$  bzw.  $h_{.j} = \sum_{i=1}^k h_{ij}$

2.4 bedingte relative Häufigkeiten:  $f_X(a_i / b_j) = \frac{h_{ij}}{h_{.j}}$  bzw.  $f_Y(b_j / a_i) = \frac{h_{ij}}{h_{i.}}$

### 3. Lage- und Streuungsparameter

	arithmetisches Mittel $\bar{x}$	empirische Varianz $s^2$	durchschnittliche Abweichung $\bar{s}$
Urliste	$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$	$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$  bzw. $s^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2$	$\bar{s} = \frac{1}{n} \sum_{i=1}^n  x_i - x_{Med}  h(a_j)$
Häufigkeitsfunktion	$\bar{x} = \frac{1}{n} \sum_j a_j h(a_j)$	$s^2 = \frac{1}{n} \sum_j (a_j - \bar{x})^2 h(a_j)$  bzw. $s^2 = \frac{1}{n} \sum_j a_j^2 h(a_j) - \bar{x}^2$	$\bar{s} = \frac{1}{n} \sum_j  a_j - x_{Med}  h(a_j)$
klassierte/ gruppierte Daten	$\bar{x} = \frac{1}{n} \sum_j m_j h_j$	$s^2 = \frac{1}{n} \sum_j (m_j - \bar{x})^2 h_j$  bzw. $s^2 = \frac{1}{n} \sum_j m_j^2 h_j - \bar{x}^2$	$\bar{s} = \frac{1}{n} \sum_j  m_j - x_{Med}  h_j$

Median  $x_{Med}$ : Median ist derjenige Wert  $x$ , für den gilt:

Urliste: Gegeben seien  $n$  geordnete Beobachtungswerte:

falls  $n$  eine ungerade Zahl, so hat der mittlere Wert die

Ordnungsnummer  $x_{(n+1)/2}$  und es gilt:  $x_{Med} = x_{(n+1)/2}$

falls  $n$  eine gerade Zahl, so kommen alle Werte

zwischen  $x_{n/2}$  und  $x_{(n/2+1)}$  als Median in Frage.

Klassiertes Datenmaterial:  $F(x) = 0,5$

Modus  $x_{Mod}$ : Modus ist derjenige Wert  $x$ , der am häufigsten realisiert ist.

## 4. Regressions- und Korrelationsrechnung

### 4.1 Regression

#### 4.1.1 Kleinste – Quadrate - Schätzer

$$\hat{b} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2} = \frac{\sum_i x_i y_i - n\bar{x}\bar{y}}{\sum_i x_i^2 - n\bar{x}^2}$$

$$\hat{a} = \bar{y} - \hat{b}\bar{x}$$

#### 4.1.2 Bestimmtheitsmaß

$$R^2 = \frac{\sum_i (\hat{y}_i - \bar{y})^2}{\sum_i (y_i - \bar{y})^2} = \frac{\sum_i \hat{y}_i^2 - n\bar{y}^2}{\sum_i y_i^2 - n\bar{y}^2}$$

## 4.2 Korrelation

### 4.2.1 Korrelationskoeffizient nach Bravais-Pearson

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}} = \frac{\sum_i x_i y_i - n\bar{x}\bar{y}}{\sqrt{\left(\sum_i x_i^2 - n\bar{x}^2\right) \left(\sum_i y_i^2 - n\bar{y}^2\right)}}$$

### 4.2.2 Korrelationskoeffizient nach Spearman

#### 4.2.2.1 ohne Bindungskorrektur

$$r_{Sp} = 1 - \frac{6 \cdot \sum_{i=1}^n (R_{x_i} - R_{y_i})^2}{n \cdot (n^2 - 1)},$$

wobei  $R_{x_i}$  und  $R_{y_i}$  Rangziffern von X und Y darstellen.

#### 4.2.2.2 mit Bindungskorrektur

$$r_{Sp}^* = 1 - \frac{6 \cdot \sum_{i=1}^n (R_{x_i} - R_{y_i})^2}{n \cdot (n^2 - 1) - \frac{1}{2} \cdot (G_S + H_S)},$$

wobei  $R_{x_i}$  und  $R_{y_i}$  Rangziffern von X und Y darstellen;

$G_S = \sum (g_i^3 - g_i)$  und  $H_S = \sum (h_i^3 - h_i)$ , wobei  $g_i$  die Länge der i-ten Bindung des Merkmals X und  $h_i$  die Länge der i-ten Bindung des Merkmals Y.

#### 4.2.3 Kontingenzkoeffizient

nicht normiert

$$C = \sqrt{\frac{X^2}{X^2 + n}}$$

und in korrigiert-normierter Form:

$$C_{korr} = C \cdot \sqrt{\frac{C_0}{C_0 - 1}}$$

$$C_0 = \min(k, l),$$

wobei:  $k$ : Anzahl der Ausprägungen bei Merkmal X

$l$ : Anzahl der Ausprägungen bei Merkmal Y

## 5. Messzahlen und Indizes sowie Konzentrationsrechnung

### 5.1 Laspeyres - Indizes

**5.1.1 Laspeyres - Mengenindex:** 
$$L_M = \frac{\sum_i m_1^i \cdot p_0^i}{\sum_i m_0^i \cdot p_0^i} \cdot 100\%$$

**5.1.2 Laspeyres - Preisindex:** 
$$L_P = \frac{\sum_i m_0^i \cdot p_1^i}{\sum_i m_0^i \cdot p_0^i} \cdot 100\%$$

### 5.2 Paasche - Indizes

**5.2.1 Paasche - Mengenindex:** 
$$P_M = \frac{\sum_i m_1^i \cdot p_1^i}{\sum_i m_0^i \cdot p_1^i} \cdot 100\%$$

**5.2.2 Paasche - Preisindex:** 
$$P_P = \frac{\sum_i m_1^i \cdot p_1^i}{\sum_i m_1^i \cdot p_0^i} \cdot 100\%$$

**5.3 Umsatzindex (Wertindex):** 
$$U = \frac{\sum_i m_1^i \cdot p_1^i}{\sum_i m_0^i \cdot p_0^i} \cdot 100\%$$

### 5.4 Herfindahl-Index

Für N Merkmalsanteile  $p_i$  ( $i = 1, 2, \dots, N$ ) ergibt sich der Herfindahl-Index  $H$  zu:

$$H = \sum_{i=1}^N p_i^2$$

Fusionieren zwei Unternehmen mit den Anteilen  $p_1$  und  $p_2$ , so erhöht sich der Wert von  $H$ , d. h. die Konzentration nimmt zu; bleiben zudem die Anteile der anderen Marktteilnehmer / Unternehmen unverändert, so verändert sich der Herfindahl-Index um  $\Delta H = 2p_1 p_2$ .