Name: **Datum:**

Grundbegriffe zur Auswertung bivariater Datenmengen

Gegeben sei eine statistische Erhebung mit

- einer Grundgesamtheit mit dem Erhebungsumfang n,
- zwei quantitativen Merkmalen X und Y,
- der durch die Erhebung gewonnen Urliste mit den Messwertepaaren $(x_1 \mid y_1)$,..., $(x_n \mid y_n)$,
- den arithmetischen Mitteln $\bar{x}$ und $\bar{y}$,
- den mittlere quadratische Abweichungen oder Varianzen V_X und V_Y sowie
- den Standardabweichungen s_X und s_Y.

Als **(empirische) Kovarianz c_{XY}** oder **s_{XY}** (der Merkmale X und Y) bezeichnet man die Zahl

$$c_{XY} = \frac{(x_1 - \bar{x}) \cdot (y_1 - \bar{y}) + .. + (x_n - \bar{x}) \cdot (y_n - \bar{y})}{n}$$

Als **(empirischen) Korrelationskoeffizient r** (der Merkmale X und Y) bezeichnet man die Zahl

$$r = \frac{c_{XY}}{s_X \cdot s_Y}$$

Weitere Erklärungen und Beispiele zu diesen Begriffen finden sich

- *unter der Internetadresse www.selbstlernmaterial.de\m\st\lr\lrindex.html*
- *im Selbstlernprogramm ‚Beschreibende Statistik und explorative Datenanalyse' in den Abschnitten 6.1 bis 6.3.*

Als **Regressionsgerade** oder **Trendgerade (bezüglich y)** bezeichnet man die Gerade mit dem Funktionsterm $y(x) = a \cdot x + b$, für die die Summe der Quadratischen Abweichungen der Messwerte y_i von den Funktionswerten $y(x_i)$ minimal ist.

Die Regressionsgerade hat den Funktionsterm

$$y(x) = \left(\frac{c_{XY}}{V_X} \right) \cdot x + \left(\bar{y} - \frac{c_{XY}}{V_X} \cdot \bar{x} \right)$$

Die Summe der Quadratischen Abweichungen der Messwerte y_i von den Funktionswerten $y(x_i)$ der Regressionsgerade hat den Wert

$$Q(a; b) = n \cdot V_Y - \frac{n \cdot {c_{XY}}^2}{V_X} = n \cdot V_Y \cdot \left(1 - \frac{{c_{XY}}^2}{V_X \cdot V_Y} \right) = n \cdot V_Y \cdot (1 - r^2)$$

Weitere Erklärungen und Beispiele zu diesen Begriffen finden sich

- *unter der Internetadresse www.selbstlernmaterial.de\m\st\lr\lrindex.html*
- *im Selbstlernprogramm ‚Beschreibende Statistik und explorative Datenanalyse' in den Abschnitten 7.1 bis 7.4.*