

Appendix D: Statistische Grundbegriffe

Ein Satz von Daten, z.B. Messungen des Körpergewichts an einer Anzahl von Individuen einer Population, sei : 4,3,7,2,9 (z.B. mg Larvalgewicht von *Drosophila*).

- **Mittelwert** : Der Mittelwert $\bar{x}$ berechnet sich als:

$$\bar{x} = \frac{4+3+7+2+9}{5} = 5,0, \text{ wobei Stichprobengrösse } (n) = 5$$

Nebst diesem arithmetischen Mittelwert, definiert man auch andere Kennwerte für das "Zentrum" von Verteilungen:

$$\text{Geometrisches Mittel: } \bar{x}_g = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n} = 4.32$$

Median = derjenige Wert, welcher die Verteilung halbiert, d.h. wo 50% der beobachteten Werte kleiner bzw. grösser sind: $\tilde{x} = 4$

- **Varianz**: Die Varianz ist ein Mass für die Streuung der Daten um den Mittelwert $\bar{x}$:

$$\text{Var}(x) = \frac{1}{n-1} \sum_i (x_i - \bar{x})^2 = \frac{1}{4} \sum (1+4+4+9+16) = 8.50$$

- Folgende Kennzahlen sind ebenfalls ein Mass für die Streuung der Daten:

Standardabweichung $SD_x = \sqrt{\text{Var}(x)} = 2.915$, charakterisiert die Streuung der Daten in der Population. Bei Normal-verteilten Daten befinden sich 68.3 % aller Werte innerhalb ± 1 SD vom Mittelwert.

Standardfehler $SE_x = \frac{SD_x}{\sqrt{n}} = \frac{2.915}{2.236} = 1.304$, charakterisiert die Streuung des Mittelwerts, d.h. der Abweichung des geschätzten Mittelwerts $\bar{x}$ vom wahren Wert μ .

- **Kovarianz**: Es seien zwei Werte pro Individuum gemessen worden, z.B. Körpergewicht (X) und Körpergrösse (Y). Die Kovarianz zwischen beiden Grössen gibt an, in welchem Mass ein Individuum mit hohem X auch ein hohes Y aufweist (falls beide Grössen positiv korrelieren). Es sei in der Reihenfolge der x -Werte: $y_i = 1.8, 1.2, 3.1, 0.8, 3.5$ mit $\bar{y} = 2.08$. Die Kovarianz ist definiert als:

$$\text{Cov}_{x,y} = \frac{1}{n-1} \sum_i (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{4} ([-1] \cdot [-0.28] + [-2.0] \cdot [-0.88] + \dots + 4 \cdot 1.42) = 3.40$$

- **Regression**: Der Regressionskoeffizient ist eine oft gebrauchte Masszahl und gibt den statistischen Zusammenhang zwischen X und Y an:

$$b_{x,y} = \frac{\text{Cov}_{x,y}}{\text{Var}(x)} = \frac{3.40}{8.50} = 0.40$$

Der Regressionskoeffizient beschreibt die Steigung der Regressionslinie, falls Y gegen X aufgetragen wird.