

Deskriptive Statistik für Soziologinnen und Soziologen (Mariana Nold)

Thema: Bivariate Exploration von quantitativen und qualitativen Merkmalen: Korrelation

Abgabe: keine Abgabe, wird in der Übung besprochen

15. Standardisierung

Die folgende Grafik ?? ist Ihnen aus der Vorlesung bekannt. Sie zeigt die Verteilung der Lese-Punkte der 5001 deutschen Schülerinnen und Schüler aus der Pisa-Studie von 2012. In grün ist eine mögliche Approximation aus der Familie der Normalverteilungen eingezeichnet.

Die Lese-Leistung der Person i ($i \in \{1, \dots, 5001\}$) wird mit y_i bezeichnet. Der Mittelwert \bar{y} ergibt sich zu 507.465 Punkten. Die Standardabweichung $\hat{\sigma}_Y$ beträgt 91.263 Punkte.

- (a) Wie können Sie beruhend auf diesen Daten eine Vermutung über die Werte der Parameter μ und σ der Normalverteilung gewinnen?
- (b) Erklären Sie jeweils, wie sich die Form der Normalverteilung ändert, wenn man
 - für festes σ den Erwartungswert μ variiert. Interpretieren Sie die inhaltliche Bedeutung.
 - für festes μ den Erwartungswert σ variiert. Interpretieren Sie die inhaltliche Bedeutung.
- (c) Nutzen Sie die 68 – 95 – 99.7-Regel um beruhend auf ihrer Modellverteilung ein Intervall zu schätzen, indem die Leseleistung von etwa 99.7% der Schülerinnen und Schüler liegt. Dieses Intervall soll symmetrisch um den vermuteten Erwartungswert μ liegen. Interpretieren Sie dieses Intervall.
- (d) Berechnen Sie jetzt ein Intervall, so dass es die Leistungen der mittleren 68% enthält und interpretieren Sie auch dieses Intervall.
- (e) Die Lese-Punkte Y sollen nun standardisiert werden. Lesen Sie S. 80 in dem Buch „Statistik- Eine Einführung für Sozialwissenschaftler “ von Ludwig-Mayerhofer, Liebeskind und Geißler (auf dt-workspace) und erklären Sie die Bedeutung der Standardisierung. Geben Sie die Berechnungsformel für die standardisierten Lese-Punkte Z an.
- (f) Die erste Person im Datensatz hat $y_1 = 475.001$ Lese-Punkte erreicht. Berechnen Sie ihre standardisierte Lese-Punktzahl z_1 .
- (g) Nehmen Sie an, es gibt noch andere Tests als die in der Pisa-Studie verwendeten Tests um die Lese-Leistung von 15-jährigen zu beurteilen. Diese Tests haben eine andere maximale Punktzahl. Wie kann Ihnen das Prinzip der Standardisierung helfen, zu vergleichen, ob die unterschiedlichen Test inhaltlich zu einer vergleichbaren Bewertung kommen?

16. Die durch die Ausgleichsgerade erklärte Streuung

In der Vorlesung hatten wir über den (vermuteten) linearen Zusammenhang der Mathe- und Lese-Punkte gesprochen. Die Grafik ?? zeigt für die ersten 300 Schülerinnen und Schüler den Zusammenhang in einem Streudiagramm.

Die Mathe-Leistung der Person i ($i \in \{1, \dots, 300\}$) wird mit x_i bezeichnet. Der Mittelwert dieser Personengruppe \bar{x} ergibt sich zu 538.92 Punkten. Die Standardabweichung dieser Personengruppe $\hat{\sigma}_X$ beträgt 101.636 Punkte.

Die Lese-Leistung der Person i ($i \in \{1, \dots, 300\}$) wird wieder mit y_i bezeichnet. Der Mittelwert dieser Personengruppe \bar{y} ergibt sich zu 523.431 Punkten. Die Standardabweichung dieser Personengruppe $\hat{\sigma}_Y$ beträgt 94.457 Punkte. Der Korrelationskoeffizient nach Pearson hat den Wert 0.889.

- (a) Berechnen Sie beruhend auf dem Zusammenhang

$$\hat{\beta} = \hat{\rho} \cdot \left(\frac{\hat{\sigma}_Y}{\hat{\sigma}_X} \right),$$

die Steigung der in der Grafik ?? eingezeichneten Geraden und interpretieren Sie diese Steigung.

- (b) Um die Gerade festzulegen, muss neben der Steigung noch der Achsenabschnitt $\hat{\alpha}$ berechnet werden. Die verwendete Formel ist

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}.$$

Berechnen und interpretieren Sie den Achsenabschnitt.

- (c) Die erste Person im Datensatz hat $x_i = 475.001$ Lesepunkte und $y_i = 443.53$ Mathe-Punkte. Berechnen Sie den durch die Gerade vorhergesagten Wert für diese Person. Dieser Wert wird mit \hat{y}_i bezeichnet.
- (d) Interpretieren Sie den Abstand $y_i - \hat{y}_i$ sowohl grafisch als auch inhaltlich.
- (e) Welcher Spezialfall ergibt sich, wenn für eine Erhebung $\sum_i^{300} |y_i - \hat{y}_i| = 0$ gilt.

17. Monotone Funktionen

- (a) Finden Sie je ein Beispiel für eine streng monoton wachsende, eine schwach monoton wachsende, eine streng monoton fallende und eine schwach monoton fallenden mathematische Funktion.
- (b) Überlegen Sie jeweils zu welchem Merkmalszusammenhang zwischen zwei Merkmalen X und Y (Streudiagramm) diese Funktion passen könnte.