

Lösungen zu den Übungsaufgaben in Kapitel 8

- (1) Tabelle 8.2 enthält die Mittelwerte der 3×25 Stichproben, die aus der Population von 150 Schülern mit den Messwerten in Tabelle 6.8 zufällig gezogen wurden. In den Online-Materialien zu diesem Buch sind die dazugehörigen Standardabweichungen der Messwerte in diesen Zufallsstichproben in der Datei mit dem Namen »daten81.txt« abgelegt. In der ersten Spalte stehen die Standardabweichungen der Messwerte in den 25 Stichproben, die $n = 10$ Schüler umfassen. In der zweiten Spalte stehen die Standardabweichungen der 25 Stichproben mit $n = 20$. In der dritten Spalte stehen die Standardabweichungen der 25 Stichproben mit $n = 30$. Lesen Sie die Datei in ein Tabellenkalkulationsprogramm (z. B. Excel) oder ein Statistikprogramm (z. B. R, SPSS) ein und berechnen Sie den Mittelwert und die Standardabweichungen (empirisch bestimmte Standardfehler) der Standardabweichungen für jede der drei Stichprobengrößen! Diskutieren Sie das Ergebnis im Vergleich zur Standardabweichung der Messwerte in der Population.

Der empirische Mittelwert der 25 Standardabweichungen aus Stichproben der Größe $n = 10$ beträgt 15,606. Der empirische Mittelwert der 25 Standardabweichungen aus Stichproben der Größe $n = 20$ beträgt 15,678. Der empirische Mittelwert der 25 Standardabweichungen aus Stichproben der Größe $n = 30$ beträgt 15,881. Wie zu erwarten war, liegt der zuletzt berechnete Wert am nächsten an der tatsächlichen Populationsstandardabweichung ($\sigma_X = 15,86$), weil er aus der größten Stichprobe stammt und die tatsächliche Populationsstandardabweichung daher präziser schätzt als die anderen. Die geschätzten Standardfehler der drei Standardabweichungen betragen 3,31 (für $n = 10$), 2,21 (für $n = 20$) und 1,63 (für $n = 30$). Hier sieht man, dass der Standardfehler der Standardabweichung mit wachsender Stichprobengröße zunehmend kleiner wird. Das muss so sein, denn mit zunehmender Stichprobengröße schwankt die Schätzung einer Standardabweichung weniger stark um die tatsächliche Populationsstandardabweichung.

- (2) In der PISA-Studie 2003 hat sich gezeigt, dass deutsche Schüler ($n = 5500$) im Fähigkeitsbereich »Problemlösen« im Durchschnitt $\bar{x} = 513$ Punkte erzielten. Die Werte sind bei PISA so normiert, dass ein Wert von $\mu = 500$ den Populationsdurchschnitt und ein Wert von $\sigma_x = 100$ die Populationsstandardabweichung darstellt.

- (a) Testen Sie mit Hilfe eines Nullhypothesentests ($\alpha = 1\%$, einseitiger Test), ob deutsche Schüler signifikant besser sind als der entsprechende Normwert.

Die gerichtete H_0 lautet in diesem Beispiel: $\mu_0 \leq 500$. Da die Populationsstandardabweichung bekannt und die Stichprobe sehr groß ist, kann hier ein z -Test verwendet werden. Es wird ein Signifikanzniveau von $\alpha = 0,05$ festgelegt. Der Standardfehler des Mittelwerts kann nach Formel F 8.4 berechnet werden. Er beträgt in unserem Beispiel $\sigma_{\bar{x}} = \frac{\sigma_X}{\sqrt{n}} = \frac{100}{74,16} = 1,35$. Der Wert

der Prüfgröße Z kann nach Formel F 8.6 berechnet werden. Er beträgt in unserem Beispiel: $z_{\bar{x}} = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{513 - 500}{1,35} = 9,64$. Die Wahrscheinlichkeit für diesen oder jeden größeren Wert unter der Nullhypothese ist extrem klein, $p < 0,00001$. Der kritische Wert bei einem 1%-Niveau wäre $z_{(0,99)} = 2,33$ (siehe Tabelle A.2 im Anhang A). Unser Wert liegt deutlich darüber. Deutsche Schüler schneiden demzufolge signifikant besser im PISA-Untertest »Problemlösen« ab als der normierte Durchschnitt.

- (b) Berechnen Sie den empirischen Effektstärkenschätzer (Cohen's d) für dieses Ergebnis inklusive des entsprechenden einseitigen 99 %-Konfidenzintervalls für die Populationseffektgröße δ . Bewerten Sie diese Effektstärke im Hinblick auf ihre praktische Bedeutsamkeit.**

Der Wert für Cohen's d kann nach Formel F 8.9 berechnet werden. Er beträgt in unserem Beispiel $d = \frac{\bar{x} - \mu_0}{\sigma_x} = \frac{513 - 500}{100} = 0,13$. Hier handelt es sich also um einen kleinen Effekt nach der

Taxonomie von Cohen (1988). Die Untergrenze des einseitigen 99%-Konfidenzintervalls für δ kann in Anlehnung an Formel F 8.39 bestimmt werden: $\delta_u = d - z_{(1-\alpha)} \cdot 1/\sqrt{n}$. In unserem Beispiel beträgt $z_{(1-\alpha)} = z_{(0,99)} = 2,33$. Eingesetzt in die o. g. Gleichung ergibt sich ein Wert von

$$\delta_u = 0,13 - \frac{2,33}{\sqrt{5500}} = 0,10. \text{ Die Obergrenze beträgt } \delta_o = +\infty.$$

- (c) Angenommen, die Kultusministerkonferenz habe als wünschenswertes Ziel für den PISA-Ländervergleich festgelegt, dass deutsche Schüler um mindestens 10 Punkte besser sein sollen als der Durchschnitt. Wie groß ist die Wahrscheinlichkeit, mit einer Stichprobengröße von $n = 5500$ auf dem 5 %-Niveau einen solchen Effekt zu finden, falls er wirklich existiert?**

Unter der zentralen z -Verteilung lautet der kritische Wert beim einseitigen Test und $\alpha = 5\%$ $z_{\text{krit}} = 1,65$. Rechnet man diesen Wert zurück in einen Stichprobenmittelwert, ergibt sich ein kritischer Wert von $\bar{x}_{\text{krit}} = z_{\text{krit}} \cdot \sigma_{\bar{x}} + \mu_0 = 1,65 \cdot 1,35 + 500 = 502,22$. Unter einer nonzentralen Stichprobenkennwerteverteilung mit dem Mittelwert $\mu_1 = 510$ und der Standardabweichung $\sigma_{\bar{x}} = 1,35$ schneidet ein Wert von 502,22 einen Flächenanteil von 99,99996 % unter der Kurve nach rechts hin ab. Dies ist die Teststärke. Es ist also nahezu sicher, mit einer Stichprobengröße von $n = 5500$ einen Effekt von $\varepsilon = 10$ auf dem 5 %-Niveau zu finden, falls dieser Effekt wirklich existiert.

- (3) In der Südwest-Realschule der Stadt F. hat sich die Schulleiterin eine eigene Testbatterie ausgedacht, um die Problemlösefähigkeiten ihrer Schülerinnen und Schüler zu testen. Zwar ist der Test so konstruiert, dass man im Durchschnitt $\mu = 100$ Punkte erzielt, allerdings ist die Populationsstandardabweichung dieses Verfahrens unbekannt. In Klasse 8a ($n = 28$) wird nun ein Mittelwert von $\bar{x} = 116$ und eine empirische Standardabweichung von $s_x = 60$ ermittelt. Das Merkmal ist in der Population normalverteilt.**

- (a) Wie groß ist der empirische Effekt, angegeben anhand der Effektgröße d_2 ?**

Um die empirische Effektstärke d_2 nach Formel F 8.40 berechnen zu können, müssen wir zunächst die Stichprobenstandardabweichung $\hat{\sigma}_x$ nach Formel F 8.11 bzw. F 8.12 berechnen. In

unserem Beispiel beträgt sie $\hat{\sigma}_x = \sqrt{s_x^2 \cdot \frac{n}{n-1}} = \sqrt{3600 \cdot \frac{28}{27}} = 61,1$. Setzen wir diesen Wert in

$$\text{Formel F 8.40 ein, erhalten wir einen Wert von } d_2 = \frac{\bar{x} - \mu_0}{\hat{\sigma}_x} = \frac{116 - 100}{61,1} = 0,26.$$

- (b) Bestimmen Sie mit Hilfe des Computerprogramms NDC oder einem vergleichbaren Programm die obere und die untere Grenze des zweiseitigen 95 %-Konfidenzintervalls für die Effektgröße δ .**

Zunächst müssen wir den Wert der Effektgröße δ mit Hilfe von Formel F 8.41b in einen t -Wert umrechnen. Dieser beträgt in unserem Beispiel $t = d_2 \cdot \sqrt{n} = 0,26 \cdot 5,3 = 1,378$. Dieser t -Wert hat $df = n - 1 = 27$ Freiheitsgrade. Nun benötigen wir die Nonzentralitätsparameter zweier non-

zentraler t -Verteilungen, unter denen ein Wert von $t = 1,378$ einen Flächenanteil von 2,5 % unter der Verteilung nach rechts bzw. nach links abschneidet. Mit Hilfe des Computerprogramms NDC ermitteln wir für λ_u einen Wert von $-0,628$ und für λ_o einen Wert von $3,36$. Diese Werte können wir nun unter Zuhilfenahme von Formel F 8.43b in δ -Werte umrechnen. Diese Werte sind dann identisch mit den Grenzen des 95%-Konfidenzintervalls. Für die untere Grenze ergibt sich ein Wert von $\delta_u = \lambda_u / \sqrt{n} = (-0,628) / 5,29 = -0,119$; für die obere Grenze ergibt sich ein Wert von $\delta_o = \lambda_o / \sqrt{n} = 3,36 / 5,29 = 0,635$. Mit einer Wahrscheinlichkeit von 95 % überdeckt das Intervall $[-0,119; 0,635]$ den wahren Populationseffekt δ .

- (c) Bestimmen Sie mit Hilfe des Programms G*Power oder einem vergleichbaren Programm, wie groß die Teststärke bei $n = 28$, $\alpha = 5\%$ und einem Effekt von $\hat{\varepsilon} = 25$ bei einem zweiseitigen Test ist.**

Ein Effekt von $\hat{\varepsilon} = 25$ entspricht einem Wert von $\hat{\delta} = \hat{\varepsilon} / \hat{\sigma}_x = 25 / 61,1 = 0,41$. Setzen wir die Werte $n = 28$, $\alpha = 0,05$ und $\hat{\delta} = 0,41$ in G*Power ein (Testfamilie: t -Tests; Statistischer Test: Mittelwerte, Abweichung von einer Konstante [Einstichprobenfall]; Poweranalysetyp: Post hoc; zweiseitiger Test) und lassen die Teststärke berechnen, erhalten wir einen Wert von $1 - \beta = 0,55$. Die Wahrscheinlichkeit, mit der ein empirischer Mittelwert signifikant von einem fixen Wert abweicht, wenn der Test zweiseitig und auf einem Signifikanzniveau von $\alpha = 0,05$ durchgeführt wird, beträgt bei einer Effektgröße von $\hat{\varepsilon} = 25$ und einer Stichprobengröße von $n = 28$ Personen also 55 %.

- (d) Bestimmen Sie mit Hilfe des Programms G*Power oder einem vergleichbaren Programm, wie viele Testpersonen nötig gewesen wären, um einen Effekt der Größe $\varepsilon_1 = 25$ mit einer Wahrscheinlichkeit von $1 - \beta = 95\%$ auf dem 5 %-Niveau (zweiseitiger Test) zu finden, falls dieser Effekt tatsächlich existiert.**

Um einen Effekt der Größe $\varepsilon_1 = 25$ (bzw. $\delta_1 = 0,41$) auf einem α -Niveau von 5 % (zweiseitiger Test) mit einer Wahrscheinlichkeit von $1 - \beta = 95\%$ zu finden, falls er wirklich existiert, benötigt man laut G*Power $n = 80$ Personen.

- (4) Ein statistisches Hypothesenpaar laute:**

$$H_0: \mu = \mu_0$$

$$H_1: \mu = \mu_1$$

Es gelte weiterhin: $\mu_1 < \mu_0$. Die folgende Abbildung zeigt zwei Stichprobenkennwerteverteilungen, von denen eine die Nullhypothese und die andere die Alternativhypothese beschreibt.

- (a) Welche Stichprobenkennwerteverteilung beschreibt die Null-, welche die Alternativhypothese?
 (b) Was wird durch den grau gefärbten Bereich markiert?
 (c) Was wird durch den blau gefärbten Bereich markiert?
 (d) Wo befindet sich der Ablehnungsbereich unter der H_0 ?
 (e) Wo befindet sich die Teststärke?

Wichtig ist zu bemerken, dass es sich hier um eine gerichtete Null- bzw. Alternativhypothese handelt; der Test wird also einseitig durchgeführt. Der Effekt ist hier negativ, d. h. die H_1 liegt unterhalb (links von) der H_0 . Der α -Fehler ist in der Abbildung grau gefärbt. Der β -Fehler ist in der Abbildung blau gefärbt. Die Teststärke ist die Gegenwahrscheinlichkeit zu β unter der H_1 . Der Ablehnungsbereich ist der Bereich links vom kritischen Wert unter der H_0 .

