Lösungen zu den Übungsaufgaben in Kapitel 13

(1) In unseren Online-Materialien finden Sie die Datei »daten_kap13.txt«, die die Rohdaten eines Vier-Gruppen-Experiments enthält. In diesem Experiment wurde die Hypothese untersucht, dass die Konzentrationsleistung von Personen unter dem Einfluss klassischer Instrumentalmusik (Bedingung 1) höher ist als unter dem Einfluss von Rockmusik (Bedingung 2), von einer Opernarie (Bedingung 3) oder von deutscher Schlagermusik (Bedingung 4). Die Versuchspersonen wurden randomisiert einer der vier experimentellen Bedingungen zugeteilt und sollten sich entsprechende Musikstücke für ca. 15 Minuten anhören. Gleichzeitig sollten sie einen Konzentrationstest bearbeiten. Die erreichte Punktzahl in diesem Test fungiert als abhängige Variable.
Lesen Sie diese Datendatei in ein Statistikprogramm (z. B. R, SPSS) oder ein Tabellenkalkulationsprogramm (z. B. Excel) ein. Beantworten Sie die folgenden Fragen:

(a) Wie lauten die Mittelwerte, die Stichprobenstandardabweichungen und die Standardfehler innerhalb der vier Gruppen?

	n_j	Mittelwert	Standard-abweichung	Standardfehler
Klass. Instrumentalmusik	20	23,25	4,66	1,04
Rockmusik	20	15,55	3,76	0,84
Opernarie	20	14,20	4,60	1,03
Deutscher Schlager	20	9,30	4,05	0,91

(b) Prüfen Sie, ob die Voraussetzungen für die Anwendung des *F*-Tests der einfaktoriellen Varianzanalyse erfüllt sind.

Unabhängige Messwerte. Ob die Unabhängigkeitsannahme wirklich erfüllt ist, lässt sich statistisch nicht ohne weiteres prüfen. Da die Versuchspersonen den vier Bedingungen aber randomisiert zugewiesen wurden, gehen wir (weitgehend ungeprüft) davon aus, dass die Unabhängigkeitsannahme nicht verletzt ist.

Normalverteilung. Ob die Variable „Punkte im Konzentrationstest" normalverteilt ist bzw. ob sie signifikant von einer Normalverteilung abweicht, kann mit Hilfe des KS-Anpassungstests überprüft werden. In unserem Fall ergibt sich für diesen Test:

Kolmogorov-Smirnov-Anpassungstest

		Ergebnis im Konzentrationstest
n		80
Parameter der Normalverteilung (aus den Daten ermittelt)	Mittelwert	15,58
	Standardabweichung	6,56
Extremste Differenzen	Absolut	0,107
	Positiv	0,107
	Negativ	–0,049
Kolmogorov-Smirnov-Statistik (z):		0,959
Asymptotische Signifikanz (2-seitig)		0,317

Da die KS-Statistik unterhalb des kritischen z-Wertes von 1,65 liegt (bei $\alpha = 5$ %), muss die Nullhypothese, der zufolge die Verteilung der Testwerte einer Normalverteilung folgt, nicht verworfen werden.

Varianzhomogenität. Ob die Annahme, dass die Varianzen in den vier Gruppen homogen sind, erfüllt ist bzw. ob die vier Varianzen signifikant voneinander abweichen, kann mit Hilfe eines Levene-Tests überprüft werden. In unserem Fall ergibt sich für die Levene-Statistik ein Wert von 0,906. Dieser Wert liegt bei 3 Zähler- und 76 Nennerfreiheitsgraden auf einem Signifikanzniveau von $\alpha = 5$ % unterhalb des kritischen Wertes von $F_{(0,95;3;76)} = 2{,}725$. Die Nullhypothese, der zufolge die Varianzen in den vier Gruppen gleich sind, muss also nicht verworfen werden.

(c) Prüfen Sie mit Hilfe einer einfaktoriellen Varianzanalyse, ob sich die vier Bedingungsmittelwerte signifikant voneinander unterscheiden. Der F-Test wird auf einem Signifikanzniveau von $\alpha = 5$ % durchgeführt.

Tafel der Varianzanalyse

	QS	*df*	*MQS*	*F*	*p*-Wert
Zwischen den Gruppen	2003,450	3	667,817	36,406	<0,001
Innerhalb der Gruppen	1394,100	76	18,343		
Gesamt	3397,550	79			

Der empirische Wert für den F-Quotienten ist größer als der kritische Wert von $F_{(0,95;3;76)} = 2{,}725$. Die Nullhypothese, der zufolge die Mittelwerte in den vier Gruppen gleich sind, kann verworfen werden.

(d) Wie lautet die empirische Effektstärke $\hat{\eta}^2$? Bewerten Sie die Größe von $\hat{\eta}^2$ anhand der Taxonomie von Cohen. Wo liegen die Grenzen des 90 %-Konfidenzintervalls für die Effektgröße η^2?

Gemäß Formel F 13.10 berechnen wir den Wert $\hat{\eta}^2 = \frac{QS_{\text{zw}}}{QS_{\text{tot}}} = \frac{2003,45}{3397,55} = 0,60$. Zum gleichen Ergebnis wären wir auch mit Formel F 13.49 gekommen:

$\hat{\eta}^2 = \frac{F \cdot df_{\text{zw}}}{F \cdot df_{\text{zw}} + df_{\text{inn}}} = \frac{36,406 \cdot 3}{36,406 \cdot 3 + 76} = 0,60.$ Dieser Wert spricht nach Cohen für einen sehr großen Effekt.

Das Konfidenzintervall für η^2 ermitteln wir, indem wir zunächst die Ober- und die Untergrenze des zweiseitigen 95 %-Konfidenzintervalls für den Nonzentralitätsparameter λ bestimmen. Das tun wir mit Hilfe des Programms NDC (oder eines anderen Programms bzw. Internet-Tools). Wir ermitteln die Werte $\lambda_{\text{u}} = 59,98$ und $\lambda_{\text{o}} = 166,81$. Diese Werte rechnen wir nun mit Hilfe von Formel F13.60a in die Effektstärke η^2 zurück und erhalten $\eta^2_{\text{u}} = \frac{\lambda_{\text{u}}}{\lambda_{\text{u}} + n} = \frac{59,98}{59,98 + 80} = 0,43$ und

$\eta^2_{\text{o}} = \frac{\lambda_{\text{o}}}{\lambda_{\text{o}} + n} = \frac{166,81}{166,81 + 80} = 0,68$. Das Intervall [0,43;0,68] überdeckt also mit einer Wahrscheinlichkeit von 95 % den wahren Populationseffekt η^2.

(e) Wie groß war die Power dieses Tests, wenn man die empirische Effektstärke $\hat{\eta}^2$ als Schätzer für den »wahren« Populationseffekt zugrunde legt?

Wenn wir die Teststärke mit Hilfe des Programms G*Power berechnen wollen, müssen wir die Effektstärke $\hat{\eta}^2$ zunächst in die Effektstärke f umrechnen. Das tun wir mit Hilfe von Formel F 13.57: $f = \sqrt{\frac{\hat{\eta}^2}{1 - \hat{\eta}^2}} = \sqrt{\frac{0,60}{1 - 0,60}} = 1,225.$ Diesen Wert können wir nun in G*Power als Schätzer für den postulierten Populationseffekt ϕ_1 einsetzen. Wir wählen folgende Einstellungen:

- ► TEST FAMILY = F tests
- ► STATISTICAL TEST = ANOVA, omnibus, one-way
- ► TYPE OF POWER ANALYSIS: Post hoc
- ► INPUT PARAMETERS: Effect size f: ϕ_1 = 1,225; α err prob: α = 0,05; Total sample size: n = 80; Number of groups: p = 4.

Dann ergibt sich eine Power von nahezu 1. Die Wahrscheinlichkeit, einen solch großen Effekt unter diesen Bedingungen zu finden, ist also sehr groß.

(f) Prüfen Sie mit Hilfe der Bonferroni-Holm-Methode, welche Bedingungsmittelwerte sich signifikant voneinander unterscheiden.

Die Bonferroni-Holm-Methode wird in vier Schritten durchgeführt, die wir hier nacheinander durchgehen wollen.

(1) Festlegung von α_{kum}. Wir legen fest: α_{kum} = 5 %.

(2) Durchführung der s Paarvergleiche und Ermittlung der jeweiligen p_r-Werte. Wenn wir für alle sechs möglichen Paarvergleiche einen eigenen t-Test für unabhängige Stichproben durchführen, erhalten wir:

Gruppe 1	Gruppe 2	Differenz	Standardfehler	t	p
Klass. Instr.-musik	Rockmusik	7,70	1,35	5,69	<0,001
Klass. Instr.-musik	Opernarie	9,05	1,35	6,68	<0,001
Klass. Instr.-musik	Deutscher Schlager	13,95	1,35	10,30	<0,001
Rockmusik	Opernarie	1,35	1,35	1,00	0,325
Rockmusik	Deutscher Schlager	6,25	1,35	4,61	<0,001
Opernarie	Deutscher Schlager	4,90	1,35	3,62	0,001

(3) Sortieren der Paarvergleiche nach den p_r-Werten in aufsteigender Reihenfolge, so dass P_1 der Mittelwertsunterschied mit dem »signifikantesten« Effekt, P_2 der Mittelwertsunterschied mit dem »zweitsignifikantesten« Effekt etc. und P_s der Mittelwertsunterschied mit dem am wenigsten signifikanten Effekt ist

(4) Bestimmung des adjustierten spezifischen Signifikanzniveaus auf der Basis der Gleichung $\alpha_r = \frac{\alpha_{\text{kum}}}{s-(r-1)}$ (siehe Formel F 13.68).

Da die p-Werte ohnehin so klein sind, können wir die Sortierung anhand der t-Werte (in absteigender Reihenfolge) vornehmen. Die adjustierten spezifischen Signifikanzniveaus sind in der folgenden Tabelle in der vorletzten Spalte eingetragen. In der letzten Spalte ist eingetragen, ob der empirische p-Wert für einen Mittelwertsvergleich kleiner ist als das jeweilige adjustierte spezifische Signifikanzniveau. Wir sehen, dass diese für alle Mittelwertsvergleiche mit Ausnahme des letzten (Vergleich zwischen Rockmusik und Opernarie) der Fall ist.

Gruppe 1	Gruppe 2	t	p	α_r	$p < \alpha_r$?
Klass. Instr.-musik	Deutscher Schlager	10,30	<0,001	0,008	ja
Klass. Instr.-musik	Opernarie	6,68	<0,001	0,010	ja
Klass. Instr.-musik	Rockmusik	5,69	<0,001	0,013	ja
Rockmusik	Deutscher Schlager	4,61	<0,001	0,017	ja
Opernarie	Deutscher Schlager	3,62	0,001	0,025	ja
Rockmusik	Opernarie	1,00	0,325	0,050	nein

(g) Wie lautet die kritische Differenz zwischen zwei Bedingungsmittelwerten in Anlehnung an die Methode von Tukey? Welche Mittelwertsdifferenzen überschreiten in diesem Datenbeispiel die kritische Differenz?

Beim Tukey-Test muss eine empirische Mittelwertsdifferenz P_r den folgenden kritischen Wert P_{krit} überschreiten, damit die Nullhypothese abgelehnt werden kann: $P_{\text{krit}} = q_{(\alpha,\, p,\, df_{\text{inn}})} \cdot \sqrt{\frac{p \cdot \hat{\sigma}_\varepsilon^2}{n}}$.

Um mit Hilfe eines Tukey-Tests zu überprüfen, welche der sechs Mittelwertsunterschiede signifikant von 0 abweichen, muss zunächst der kritische q-Wert ermittelt werden. In unseren Online-Materialien verweisen wir auf Tabellen, in denen die kritischen Werte für eine bestimmte Anzahl von Paarvergleichen, ein bestimmtes α-Niveau und für bestimmte Freiheitsgrade abgetragen sind.

Bei einem Signifikanzniveau von α_{kum} = 5 % beträgt dieser Wert bei s = 6 Paarvergleichen und df_{inn} = 12 Nennerfreiheitsgraden q_{krit} = 4,2. Eingesetzt in die o. g. Formel ergibt sich dann $P_{\text{krit}} = 4,2 \cdot \sqrt{\frac{18,34}{20}} = 4,02$. Auch hier gilt, dass alle paarweisen Mittelwertsdifferenzen bis auf einen (Rockmusik vs. Opernarie) signifikant von 0 verschieden sind.

(h) Prüfen Sie das folgende Set von Kontrasthypothesen. Zeigen Sie anhand der Additivität der drei Quadratsummen, dass es sich hier um ein Set orthogonaler Kontraste handelt.

$$\frac{\mu_1 + \mu_3}{2} < \frac{\mu_2 + \mu_4}{2}$$
$$\frac{\mu_1 + \mu_2}{2} < \frac{\mu_3 + \mu_4}{2}$$
$$\frac{\mu_1 + \mu_4}{2} < \frac{\mu_2 + \mu_3}{2}$$

Zunächst formulieren wir für jeden dieser drei Kontrasthypothesen (KH1, KH2 und KH3) die entsprechenden Kontrastkoeffizienten. Sie lauten:

KH1: $K_1 = -1$ $K_2 = 1$ $K_3 = -1$ $K_4 = 1$
KH2: $K_1 = -1$ $K_2 = -1$ $K_3 = 1$ $K_4 = 1$
KH3: $K_1 = -1$ $K_2 = 1$ $K_3 = 1$ $K_4 = -1$

Nun bilden wir die drei Kontrast-Quadratsummen gemäß Formel F13.74 und erhalten für die erste Kontrasthypothese KH1: QS_{KH1} = 793,80; für die zweite Kontrasthypothese QS_{KH2} = 1170,45 und für die dritte Kontrasthypothese QS_{KH3} = 39,20. Nebenbemerkung: Berechnet man die Summe aus diesen drei Kontrast-Quadratsummen, ergibt sich der Wert für die QS_{zwischen} in der einfaktoriellen Varianzanalyse, nämlich 2003,45. Das ist ein Beleg dafür, dass es sich hier wirklich um orthogonale Kontraste handelt.

Als nächstes bilden wir gemäß Formel F 13.73b den F-Quotienten für jede dieser Kontrasthypothesen und erhalten für die erste Kontrasthypothese F_{KH1} = 0,57; für die zweite Kontrasthypothese F_{KH2} = 0,84 und für die dritte Kontrasthypothese F_{KH3} = 0,03. Alle diese F-Werte liegen unter 1, d. h. ein Vergleich mit dem kritischen Wert erübrigt sich. Keiner der drei Kontraste ist signifikant! Das erkennt man auch schon an dem Mittelwertsmuster: Die Konzentrationsleistung nimmt über die vier Bedingungen (Klassische Instrumentalmusik, Rockmusik, Opernarie, deutscher Schlager) hinweg immer weiter ab und folgt insofern eher einem „linearen“ Trend als einem der oben genannten Kontraste.

(2) Führen Sie mit dem Datenbeispiel aus der Datei »daten_kap13.txt« eine Rangvarianzanalyse (*H*-Test) durch. Zu welchem Ergebnis kommen Sie?

Zunächst berechnen wir die bedingungsspezifischen Rangsummen. Sie lauten RS_1 = 1325; RS_2 = 836; RS_3 = 732 und RS_4 = 347. Als nächstes berechnen wir mit Hilfe dieser Rangsummen den Wert der Kruskal-Wallis-Statistik (H) nach Formel F13.147 und erhalten:

$$\begin{aligned} H &= \frac{12}{n \cdot (n+1)} \cdot \sum_{j=1}^{p} \frac{RS_j^2}{n_j} - 3 \cdot (n+1) \\ &= \frac{12}{80 \cdot (80+1)} \cdot 155537,7 - 3 \cdot (80+1) \\ &= 45,03 \end{aligned}$$

Der kritische χ^2-Wert bei $\alpha = 5\,\%$ und $df = 3$ Freiheitsgraden lautet $\chi^2_{(0,95;3)} = 7{,}81$. Unser empirischer H-Wert liegt deutlich darüber. Die Nullhypothese, der zufolge die Mediane der vier Gruppen auf Populationsebene identisch sind, kann verworfen werden.

(3) In Tabelle 13.5 sind die Ergebnisse des Experiments von Pihl et al. (1981) angegeben. Dort wurden 48 männliche Versuchspersonen einer von vier experimentellen Bedingungen zugewiesen. Jede Bedingung umfasst 12 Versuchspersonen; es wurde variiert, ob (a) den Versuchspersonen ein alkoholisches oder ein nicht-alkoholisches Getränk gegeben wurde und (b) ob ihnen *gesagt* wurde, dass es sich um ein alkoholisches oder um ein nicht-alkoholisches Getränk handele. Wir haben in Abschnitt 13.1.13 so getan, als hätten wir es mit einem vierstufigen Faktor zu tun. Man kann das Experiment aber auch als zweifaktorielles Design mit den Faktoren *A* (»Getränk« mit den Stufen »alkoholisch« vs. »nicht-alkoholisch«) und *B* (»Ankündigung« mit den Stufen »alkoholisch« vs. »nicht-alkoholisch«) auffassen. In einem solchen Auswertungsdesign könnten neben den Haupteffekten des Faktors *A* und den Haupteffekten des Faktors *B* auch noch die Interaktionseffekte *A* × *B* getestet werden.

(a) Testen Sie mit Hilfe einer zweifaktoriellen Varianzanalyse, ob es Haupteffekte des Faktors *A*, Haupteffekte des Faktors *B* und Interaktionseffekte *A* × *B* gibt. Wie lassen sich die Ergebnisse der drei *F*-Tests inhaltlich interpretieren?

Formt man Tabelle 13.5 so um, dass sich ein zweifaktorielles Auswertungsdesign ergibt, erhält man die folgende tabellarische Darstellung:

Mittelwerte ($\bar{x}_{jk}$) und Standardabweichungen ($\hat{\sigma}_{jk}$)		Faktor *B* (Ankündigung): alkoholisch ($k = 1$)	Faktor *B* (Ankündigung): nicht-alkoholisch ($k = 2$)	Zeilenmittelwerte
Faktor *A* (Getränk)	alkoholisch ($j = 1$)	$\bar{x}_{11} = 70$ $\hat{\sigma}_{11} = 9{,}04$	$\bar{x}_{12} = 69{,}33$ $\hat{\sigma}_{12} = 10{,}57$	$\bar{x}_{1\bullet} = 69{,}67$
	nicht-alkoholisch ($j = 2$)	$\bar{x}_{21} = 73{,}25$ $\hat{\sigma}_{21} = 7{,}64$	$\bar{x}_{22} = 61$ $\hat{\sigma}_{22} = 14{,}69$	$\bar{x}_{2\bullet} = 67{,}13$
Spaltenmittelwerte		$\bar{x}_{\bullet 1} = 71{,}63$	$\bar{x}_{\bullet 2} = 65{,}17$	$\bar{x} = 68{,}40$

Zunächst berechnen wir die mittleren Quadratsummen für die drei zu testenden Effekte mit Hilfe der Formeln F 13.118 bis F 13.120. Wir erhalten:

$$
\begin{aligned}
MQS_A &= \frac{q \cdot n_{jk} \cdot \sum_{j=1}^{p} \left(\bar{x}_{j\bullet} - \bar{x}\right)^2}{p-1} \\
&= \frac{2 \cdot 12 \cdot \left((69{,}67 - 68{,}4)^2 + (67{,}13 - 68{,}4)^2\right)}{2-1} \\
&= 77{,}42
\end{aligned}
$$

$$MQS_B = \frac{p \cdot n_{jk} \cdot \sum_{k=1}^{q} (\bar{x}_{\bullet k} - \bar{x})^2}{q-1}$$
$$= \frac{2 \cdot 12 \cdot \left((71{,}63 - 68{,}4)^2 + (65{,}17 - 68{,}4)^2\right)}{2-1}$$
$$= 500{,}78$$

$$MQS_{A \times B} = \frac{n_{jk} \cdot \sum_{k=1}^{q} \sum_{j=1}^{p} \left(\bar{x}_{jk} - \bar{x}_{j\bullet} - \bar{x}_{\bullet k} + \bar{x}\right)^2}{(p-1) \cdot (q-1)}$$
$$= \frac{12 \cdot (8{,}38 + 8{,}38 + 8{,}38 + 8{,}38)}{(2-1) \cdot (2-1)}$$
$$= 402{,}29$$

Die $MQS_{\text{innerhalb}}$ ergibt sich gemäß Formel F 13.121 aus dem Durchschnitt der MQS_{jk}. Diese lassen sich berechnen, indem man die Zell-Standardabweichungen $(\hat{\sigma}_{jk})$ quadriert. Da die Stichprobengröße in allen Zellen identisch ist, brauchen wir nur den Mittelwert der MQS_{jk} zu berechnen. Wir erhalten den Wert $MQS_{\text{inn}} = 116{,}90$.
Es ergibt sich die folgende Tafel der Varianzanalyse:

Quelle der Variation	*QS*	*df*	*MQS*	*F*	*p*
Haupteffekt *A*	77,42	1	77,42	0,66	0,42
Haupteffekt *B*	500,78	1	500,78	4,28	0,04
Wechselwirkung $A \times B$	402,29	1	402,29	3,44	0,07
Residuum	5143,73	44	116,90		
Gesamt	6124,22	47	130,30		

Die Haupteffekte des Faktors *A* (Getränk) sind nicht signifikant, d. h. die Nullhypothese, dass es keinen Einfluss der »echten« Alkoholhaltigkeit des Getränks gibt, kann nicht verworfen werden. Die Haupteffekte des Faktors *B* (Ankündigung) sind sehr wohl signifikant, d. h. ob man den Versuchspersonen sagt, sie tränken ein alkoholisches oder ein nicht-alkoholisches Getränk, macht einen Unterschied. Die Interaktionseffekte $A \times B$ sind nicht signifikant, allerdings unterschreitet der empirische *F*-Wert nur knapp den kritischen *F*-Wert. In solchen Fällen hat sich die Bezeichnung eingebürgert, ein Effekt sei »marginal signifikant«.

(b) Berechnen Sie die Effektgrößen dieser drei Effekte für die Effektgrößenmaße $\hat{\eta}^2$ (nicht-partiell), $\hat{\eta}_p^2$ (partiell), ϕ^2 und ϕ. Bewerten Sie die Größe der Effekte.

Quelle der Variation	$\hat{\eta}^2$	$\hat{\eta}_p^2$	f^2	f
Haupteffekt *A*	0,013	0,015	0,015	0,123
Haupteffekt *B*	0,082	0,089	0,097	0,312
Wechselwirkung $A \times B$	0,066	0,073	0,078	0,280

Nach der Taxonomie von Cohen (1988) handelt es sich bei dem Haupteffekt A um einen kleinen Effekt, bei dem Haupteffekt B und dem Interaktionseffekt $A \times B$ handelt es sich um mittlere Effekte.

(c) Wie viele Personen wären nötig gewesen, um die Nullhypothese, dass alle Interaktionseffekte $A \times B$ in der Population gleich 0 sind, mit einer Wahrscheinlichkeit von 90 % abzulehnen, wenn der Test auf einem Signifikanzniveau von $\alpha = 5$ % durchgeführt wird und der Effekt in der Population $\eta^2_{(p\_A \times B)1} = 0{,}05$ beträgt?

Wenn wir die optimale Stichprobengröße mit Hilfe des Programms G*Power berechnen wollen, müssen wir die Effektstärke $\eta^2_{(\mathrm{p}\_A \times B)1}$ zunächst in die Effektstärke ϕ_1 umrechnen. Das tun wir mit Hilfe von Formel F13.57: $\phi_1 = \sqrt{\frac{0{,}05}{1 - 0{,}05}} = 0{,}23$. Diesen Wert können wir nun in G*Power einsetzen. Wir wählen folgende Einstellungen:

- ► TEST FAMILY = F tests
- ► STATISTICAL TEST = ANOVA, fixed effect, spezial, main effects and interactions
- ► TYPE OF POWER ANALYSIS: A priori
- ► INPUT PARAMETERS: Effect size f: $\phi_1 = 0{,}23$; α err prob: $\alpha = 0{,}05$; Power ($1 - \beta$ err prob): $1 - \beta = 0{,}90$; Numerator df: $df_{A \times B} = (p - 1) \cdot (q - 1) = 1$; Number of groups: $p \cdot q = 4$.

G*Power gibt in diesem Fall an, dass die optimale Stichprobengröße $n = 201$ betragen müsste. Es hätten also 50 Personen pro Zelle untersucht werden müssen, um die Nullhypothese für den Interaktionseffekt $A \times B$ mit einer Wahrscheinlichkeit von 90 % abzulehnen, wenn der Test auf einem Signifikanzniveau von $\alpha = 5$ % durchgeführt wird und der Effekt in der Population $\eta^2_{(\mathrm{p}\_A \times B)1} = 0{,}05$ beträgt.