1. Aufgabe (10 Punkte)

Auf Basis eines Depressionsinventars und der Werte eines IQ-Tests soll der spätere Berufserfolg (gemessen an dem Monatsgehalt) vorhergesagt werden. Hierbei ergeben sich folgende Werte:

Berufserfolg (y)	Depressionsinventar (x1)	IQ-Test (x2)
4	5	5
6	9	8
9	10	8
3	1	4
5	9	6

Sie wollen nun mittels einer multiplen Regressionsanalyse ermitteln, ob die vorhandenen Zusammenhänge bestehen.

a) Welches Minimierungsverfahren können Sie verwenden um die Einflussgewichte b_0, b_1 und b_2 zu berechnen? Erläutern Sie dieses kurz.

b) Ergänzen Sie die folgende Matrix:

$$\mathbf{X'X} = \begin{pmatrix} 5 & 34 & 31 \\ 34 & & 235 \\ & & 205 \end{pmatrix}$$

c) Berechnen Sie die Determinante der Matrix **(X'X).**

d) Gegeben sei die Inverse der Matrix **X'X.** Berechnen Sie die Einflussgewichte b_0, b_1 und b_2 (kürzen Sie die Werte bis auf die zweite Nachkommastelle).

$$\mathbf{(X'X)^{-1}} = \begin{pmatrix} 3815 & 315 & -938 \\ 315 & 64 & -121 \\ -938 & -121 & 284 \end{pmatrix} \div \mathbf{|X'X|}$$

e) Welchen Berufserfolg ($\hat{y}$) würden Sie für eine Person vorhersagen, die im Depressionsinventar einen Wert von 15 und im IQ-Test einen Wert von 7 hat? Tragen Sie die Werte in die Regressionsgleichung ein.

f) In welchem Bereich würde mit 95%iger Wahrscheinlichkeit der Berufserfolg y für eine Person liegen, die im Depressionsinventar einen Wert von $x_1 = 15$ und im IQ Test einen Wert von $x_2 = 7$ hat? Benutzen Sie hierfür eine Formel aus dem Formelblatt.

g) Angenommen, man habe bei der vorliegenden Untersuchung ein b_1 von 0,07 gefunden. In welchem Bereich liegt b_1 mit einer Irrtumswahrscheinlichkeit von 5%?

2. Aufgabe (10 Punkte)

Der Berufserfolg (y) soll durch die Abiturnote (x_1) und durch den IQ (x_2) vorhergesagt werden. Sie haben bereits eine multiple Regression mit folgenden Teilergebnissen berechnet:

$r_{y,x1} = 0{,}7$ $\qquad Q_d = 310{,}79$

$r_{y,x2} = 0{,}8$ $\qquad Q_e = 46{,}44$

$r_{x1,x2} = 0{,}3$

a) Wie viel Varianz des Kriteriums wird prozentual durch die beiden Prädiktoren gemeinsam aufgeklärt?

b) Berechnen Sie prozentual das Inkrement der Abiturnote, wenn zuvor der IQ bereits als Prädiktor gewählt wurde.

c) Was bedeutet das Inkrement inhaltlich?

d) Welches Inkrement hätte der IQ, wenn die Korrelation zwischen IQ und Abiturnote $r_{x1,x2} = 0$ wäre? Begründen Sie Ihre Antwort kurz.

e) Wie hoch ist die Partialkorrelation zwischen IQ und Berufserfolg, wenn die Abiturnote als Drittvariable heraus gerechnet wird?

f) Bei einer Einfach-Regression, bei der der Berufserfolg aus der Abiturnote vorhergesagt wurde, wie groß wäre der $\hat{\beta}$-Koeffizient?

3. Aufgabe (10 Punkte)

Eine Untersuchung zur intrinsischen Motivation bei Schülern wurde durchgeführt. Dazu wurden jeweils 7 Schüler aus unterschiedlichen Klassen für vier Klassenstufen untersucht. Erhoben wurde die intrinsische Motivation anhand eines Fragebogens (höhere Werte = höhere intrinsische Motivation). $QS_{inn} = 348$ sei gegeben.

Vpn	1.Klasse	2.Klasse	3.Klasse	4.Klasse	
1	20	13	14	4	
2	18	8	11	12	
3	12	18	7	9	
4	16	11	4	6	
5	8	15	12	10	
6	14	9	8	5	
7	11	12	17	10	
	$\bar{y}_1 =$	$\bar{y}_2 = 12{,}29$	$\bar{y}_3 = 10{,}43$	$\bar{y}_4 = 8$	$\bar{\bar{y}} = 11{,}215$

a) Vervollständigen Sie die Tabelle.

b) Berechnen Sie QS_{zw} und QS_{total} .

c) Berechnen Sie MQS_{zw} und MQS_{inn}. ($df_{zw} = 3$; $df_{inn} = 24$)

d) Erstellen Sie die Nullhypothese sowohl formal als auch inhaltlich.

e) Führen Sie den Signifikanztest durch (α = 5 %). Was bedeutet das Ergebnis inhaltlich?

f) Was wären die negativen Folgen, wenn man statt einer ANOVA mehrere t-Tests verwenden würde?

g) Was würde bezogen auf das obige Beispiel ein F-Wert nahe 1 für die Varianz der α_k bedeuten? Begründen Sie kurz ihre Antwort unter Berücksichtigung der Formel für den F-Wert.

4. Aufgabe (6 Punkte)

Ein Pharmakonzern hat ein neues Medikament zur Behandlung von Depression entwickelt. Er möchte nun überprüfen, ob dieses Medikament in Kombination mit drei unterschiedlichen Therapieformen einen Einfluss auf den Ausprägungsgrad der sozialen Phobie hat. Hohe Werte sprechen für eine hohe Ausprägung der sozialen Phobie.
Für eine zweifaktorielle ANOVA wurden folgende Mittelwerte gefunden (n=6 pro Zelle).

		Faktor A mit J=3 Stufen „Therapieformen“			
		j = 1	j = 2	j = 3	
Faktor B mit K=2 Stufen „Dosis“	k = 1; Dosis = niedrig	$\bar{y}_{11} = 16$	$\bar{y}_{21} = 25$	$\bar{y}_{31} = 19$	$\bar{y}_{.1} =$
	k = 2; Dosis = hoch	$\bar{y}_{12} = 4$	$\bar{y}_{22} = 15$	$\bar{y}_{32} = 11$	$\bar{y}_{.2} =$
		$\bar{y}_{1.} =$	$\bar{y}_{2.} =$	$\bar{y}_{3.} =$	$\bar{\bar{y}} = 15$

a) Vervollständigen Sie die Tabelle und berechnen Sie die Haupteffekte α_j, β_k (verwenden Sie hierfür die Mittelwerte aus der Tabelle).

b) Wie lautet die Modellgleichung für die zweifaktorielle ANOVA? Gebt auf Grundlage dieser Modellgleichung die Nullhypothese für einen Haupteffekt des Faktors B an. Formulieren Sie die Nullhypothesen sowohl formal als auch inhaltlich.

c) QS_{inn}= 240 sei gegeben. Berechne QS_B , MQS_B und MQS_{inn} . Gibt es einen signifikanten Haupteffekt von Faktor B bei einem α–Niveau von 5%?

d) Wie viel Varianz wird durch den Faktor B erklärt? Benutzen Sie hierfür eine Formel auf dem Formelblatt.

5. Aufgabe (4 Punkte)

Eine einfaktorielle ANOVA mit K=3 Gruppen wurde durchgeführt (balanciertes Design mit n=21, MQS_{inn} = 1.67). Folgende Mittelwerte wurden für die 3 Gruppen berechnet:

Gruppe 1	*Gruppe 2*	*Gruppe 3*
$\bar{y}_1 = 7.1$	$\bar{y}_2 = 5.40$	$\bar{y}_3 = 3.50$

a) Formulieren Sie einen Kontrast ψ , um folgende H0 zu testen:

"Die Mittelwerte verändern sich linear von Gruppe 1 über Gruppe 2 zu Gruppe 3 ".

Geben Sie die Kontrastkoeffizienten an und formulieren Sie die H0 damit.

b) Berechnen Sie $\hat{\psi}$ und den Standardfehler $s_{\hat{\psi}}$ für den Kontrast. Berechnen Sie den

t-Wert.

Lösung Aufgabe 1: Regression

Formel: $\mathbf{b} = (\mathbf{X'X})^{-1}\mathbf{X'y}$

a) Least Squares / Kleinste Quadrate

b)

$$\mathbf{X'X} = \begin{matrix} 5 & 34 & 31 \\ 34 & 288 & 235 \\ 31 & 235 & 205 \end{matrix}$$

c) 5 * 3815 + 34 * 315 + 31 * -938

|X'X| = 707

d)

$$\mathbf{K'} = \begin{matrix} 3815 & 315 & -938 \\ 315 & 64 & -121 \\ -938 & -121 & 284 \end{matrix}$$

$$\mathbf{X'y} = \begin{matrix} 27 \\ \mathbf{212} \end{matrix}$$

182

$$\mathbf{b} = \begin{array}{r} -1{,}3168 \\ 0{,}0721 \\ 1{,}0042 \end{array}$$

e)$\hat{y}$ = −1,3168 + 0,0721*15 + 1,0042*7 = 6,7941

$$\mathbf{(X'X)^{-1}} = \begin{array}{rrr} 5{,}4 & 0{,}44 & -1{,}33 \\ 0{,}44 & 0{,}09 & -0{,}17 \\ -1{,}33 & -0{,}17 & 0{,}40 \end{array}$$

f)

df_e = n – 3 = 2 (5 Probanden)

Q_e = sollte laut Besprechung bereits zuvor gegeben sein, **ist es aber nicht!!!** → In einsetzen, wobei t = 4,3027

Und Y_p = 6,7941

g)

t = 4,3027 & b_1= 0,07

einsetzen in

Lösung Aufgabe 2: Varianzaufklärung/Inkrement

a) $Q_t = Q_d + Q_e$

Q_t = 310,79 + 46,44 = 357,23

Determinationskoeffzient: $R^2 = \frac{Qd}{Qt}$ = 310,79/357,23 = 0,87

R^2 * 100 = 87 %

87 % der Kriteriumsvarianz wird durch die beiden Prädiktoren gemeinsam aufgeklärt.

b) $R_I^2 = R_U^2 + R_E^2$

R_U^2 = 0,87

R_E^2 : Entspricht der aufgeklärten Varianz, die durch den bereits enthalten Prädiktor (IQ) alleine aufgeklärt wird.

$R_E^2 = 0{,}8^2 = 0{,}64$

$R_I^2 = 0{,}87 - 0{,}64 = 0{,}23$

$R^2 * 100 = 23\ \%$

23 % der Kriteriumsvarianz wird durch die Abiturnote zusätzlich aufgeklärt. $\hat{\sigma}_{\varepsilon} = \sqrt{\frac{Q_e}{df_e}}$

$$\hat{y}_p - t_{(1-\alpha/2;df_e)} \cdot \hat{\sigma}_{\varepsilon} \leq y_p \leq \hat{y}_p + t_{(1-\alpha/2;df_e)} \cdot \hat{\sigma}_{\varepsilon}$$

c) Das Inkrement gibt den Zuwachs an erklärter Varianz an, der durch die Hinzunahme eines Prädiktors in ein bestimmtes Modell erreicht wird.

d) Der IQ hätte ein Inkrement von $R^2 = 0{,}8^2 = 0{,}64$.

$$b_j - t_{(1-\alpha/2;df_e)} \cdot \hat{\sigma}_{\beta_j} \leq \beta_j \leq b_j + t_{(1-\alpha/2;df_e)} \cdot \hat{\sigma}_{\beta_j}$$

$$\hat{\sigma}_{\beta_j} = \hat{\sigma}_{\varepsilon} \cdot \sqrt{a_{jj}}$$

Wenn die beiden Prädiktoren nicht miteinander korrelieren, dann klären sie völlig unabhängig voneinander Varianz des Kriteriums auf. Damit entspricht der Zuwachs an erklärter Varianz bei Hinzunahme eines Prädiktors der Varianz, die insgesamt alleine durch den Prädiktor aufgeklärt wird. Deshalb kann hier die Einfachkorrelation zwischen IQ und Berufserfolg direkt quadriert werden um das Inkrement zu erhalten.

e) Formel: (siehe Formelsammlung)

$$r_{x2,y\bullet x3} = \frac{0{,}8 - 0{,}3 * 0{,}7}{\sqrt{1-0{,}3^2} * \sqrt{1-0{,}7^2}} = 0{,}866$$

f) Bei der Einfach-Regression entspricht der $\hat{\beta}$-Koeffizient der Korrelation. In diesem Fall wäre der $\hat{\beta}$-Koeffizient also = 0,7

Lösung Aufgabe 3: Einfaktorielle ANOVA

a) $\bar{y}_1 = 14,14$

b) $QS_{zw} = n_k * \sum_{k=1}^{k}(\overline{y_k} - \bar{\bar{y}})^2$

$QS_{zw} = 7 * ((14,14-11,215)^2 + ...) = 144,646$

$QS_{tot} = QS_{zw} + QS_{inn} = 144,646 + 348 = 492.646$

c) $MQS_{zw} = \frac{QSzw}{dfzw}$ $MQS_{inn} = \frac{QSinn}{dfinn}$

$MQS_{zw} = 144.646/3 = 48,215$

$MQS_{inn} = 348/24 = 14,5$

d) $H_o : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$ oder $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$

H_0: Es gibt keine signifikanten Unterschiede hinsichtlich der intrinsischen Motivation zwischen den Gruppenmittelwerten der Klassenstufen.

e) $F_{emp} = \frac{MQSzw}{MQSinn} = 48,215/14,5 = 3,325$

$F_{krit(3;\ 24;\ 0,05)} = 3,01$

$F_{krit} < F_{emp} \rightarrow$ significant

Die H_o wird abgelehnt.

Inhaltlich: Es gibt einen signifikanten Unterschied hinsichtlich der intrinsischen Motivation zwischen den Klassenstufen.

f) Es würde zu einer α-Fehler Inflation kommen, d.h. der tatsächliche α – Fehler würde sich durch die mehrfache Anwendung der t-Tests deutlich erhöhen. Bei der Anova dagegen wird durch die simultane Testung der Mittelwerte automatisch der vom Anwender gewählte a-Fehler korrekt eingehalten.

g) Es würde bedeuten, dass die var (α_k) nahe 0 wäre: Der empirischen F-Wert wird durch die Formel $F_{emp} = \frac{MQSzw}{MQSinn}$ berechnet. MQS_{zw} entspricht ungefähr Var(ε) + n Var(α) während MQS_{inn} ungefähr Var(ε) entspricht. Ist der empirische F-Wert nahe 1 so bedeutet dies, dass im Zähler fast ausschließlich Fehlervarianz vorherrscht. Der F-Bruch besteht damit nahezu aus Var(ε) / Var(ε) und ist somit nahe 1. Daher gibt es keine Effektvarianz und var (α_k) wäre nahe 0.

Lösung Aufgabe 4: Zweifaktorielle ANOVA

Datensatz:

	j= 1	j= 2	j= 3	
k = 1	12 12 20 20 15 17	20 30 24 26 22 28	21 21 17 17 21 17	
	16	25	19	**20**
k = 2	2 6 6 2 2 6	12 18 14 18 16 12	13 13 9 9 10 12	
	4	15	11	**10**
	10	**20**	**15**	**15**

a)

			= 20
			= 10
= 10	= 20	= 15	

$\alpha 1= -5$ $\alpha 2= 5$ $\alpha 3= 0$ $\beta 1= 5$ $\beta 2= -5$

b)

$$y_{ijk} = \mu_{jk} + \varepsilon_{ijk} = \mu + \alpha_j + \beta_k + \gamma_{jk} + \varepsilon_{ijk}$$

H0B: $\beta 1= \beta 2= 0$
Die Dosis des Medikamentes hat keine Auswirkungen auf den Ausprägungsgrad der sozialen Phobie.

c)
$QS_{TOT}= 1800$

$QS_A = 600$
$QS_B = 900$, $QS_{AxB} = 60$, $QS_{inn} = 240$, $df_B = 1$, $df_{inn} = 30$

$F_B = MQS_B / MQS_{inn} = 900 / 8 = 112{,}5$
$F_{krit} = 7{,}56$
$F_B > F_{krit}$ → signifikant!

Der Faktor B hat einen signifikanten Einfluss in der Population. Dies Dosis des Medikamentes hat somit eine signifikante Auswirkung auf den Ausprägungsgrad der sozialen Phobie.