Aufgabe 1 — 16 Punkte

Ein renommiertes Ernährungsinstitut hat ein neues Diät-Programm entwickelt, das nicht nur auf die pure Gewichtsabnahme abzielt, sondern insbesondere auf die Reduktion von Körperfett. Um die Wirksamkeit des Programms zu überprüfen, wird bei jedem Kunden das Körpergewicht und das Körperfett vor und nach 12 Wochen Diät bestimmt. Die Ergebnisse von $n = 3$ Kunden sind in den folgenden Datenmatrizen zusammengefasst:

$$\text{Vorher: } \mathbf{X}^{(1)} = \begin{pmatrix} 82.1 & 28.7 \\ 75.3 & 22.6 \\ 85.6 & 28.2 \end{pmatrix} \qquad \text{Nachher: } \mathbf{X}^{(2)} = \begin{pmatrix} 78.0 & 27.7 \\ 72.2 & 19.6 \\ 80.2 & 25.0 \end{pmatrix}.$$

Die erste Spalte enthält die jeweils gemessenen Werte für das Körpergewicht, die zweite Spalte für das Körperfett, beides in [kg]. Zusätzlich liegen bereits die individuellen Differenzenvektoren $\mathbf{d}_i = \mathbf{x}_i^{(1)} - \mathbf{x}_i^{(2)}$ für $i = 1, \ldots, 3$

$$\bar{\mathbf{d}} = \begin{pmatrix} 4.2 \\ 2.4 \end{pmatrix}$$

und die Kovarianzmatrix der Differenzenvektoren vor:

$$\mathbf{S}_d = \begin{pmatrix} 1.33 & 0.22 \\ 0.22 & 1.48 \end{pmatrix}$$

Testen Sie mit einem geeigneten Test zum Signifikanzniveau von 5% und unter der Annahme von normalverteilten Variablen zu beiden Zeitpunkten, ob sich das erwartete Körpergewicht und Körperfett nach der Diät signifikant verändert haben.

Aufgabe 2 — 8 Punkte

Multivariate Regressionsmodelle:

Gegeben sei eine stetige Response-Variable der Dimension $q \geq 2$ sowie p Kovariablen. Erklären Sie, bei welcher Art von Fragestellung es sinnvoll ist, ein multivariates Regressionsmodell anstelle von q separaten univariaten Regressionsmodellen zu schätzen.

Aufgabe 3	15 Punkte

In einer Bank werden Kreditnehmer, die einen Kredit in Höhe von 100 000 Euro aufnehmen, in die drei Gruppen gut ($=G$), mittel ($=M$) und schlecht ($=S$) eingeteilt. Die Kategorie schlecht (S) wird als Referenzkategorie festgelegt. Um die Wahrscheinlichkeit der Gruppenzugehörigkeit zu bestimmen, wird ein multinomiales Logit-Modell mit folgenden Einflussgrößen geschätzt:

X_1	Alter in Jahren
X_2	Jahreseinkommen in 1000 Euro
X_3	Geschlecht (männlich=1, weiblich=0)

Die Schätzung der Parameter ergibt:

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_G &= (\hat{\beta}_{G0}, \hat{\beta}_{G1}, \hat{\beta}_{G2}, \hat{\beta}_{G3})^\top &&= (-3.00,\, 0.10,\, 0.10,\, 0.50)^\top \\
\hat{\boldsymbol{\beta}}_M &= (\hat{\beta}_{M0}, \hat{\beta}_{M1}, \hat{\beta}_{M2}, \hat{\beta}_{M3})^\top &&= (0.00,\, 0.05,\, 0.05,\, 0.50)^\top
\end{aligned}$$

(a) Formulieren Sie kurz das zugehörige multinomiale Logit-Modell.

(b) Interpretieren Sie den geschätzten Parameter $\hat{\beta}_{M0}$ und $\hat{\beta}_{M2}$.

Aufgabe 4	23 Punkte

Eine Privatbank beurteilt die Kreditwürdigkeit ihrer Kunden anhand von drei Gehaltsstufen. Die Klassifikation zur Beurteilung der Kreditwürdigkeit von Kreditnehmern erfolgt in die Klassen kreditwürdig (Klasse 1) und nicht kreditwürdig (Klasse 2). Die Verteilung des trichotomen Merkmals X in den beiden Klassen und die apriori Wahrscheinlichkeiten für die jeweiligen Klassenzugehörigkeiten seien durch folgende Tabelle bestimmt:

	X			
	Gehalt niedrig	Gehalt mittel	Gehalt hoch	apriori-Wahrscheinlichkeit
Klasse1	0.1	0.2	0.7	p
Klasse2	0.8	0.1	0.1	1-p

(a) Bestimmen Sie die ML-Zuordnung für die drei Gehaltsstufen. Berechnen Sie anschließend für diese Zuordnung die individuellen Fehlerraten ϵ_{21}

(b) Berechnen Sie für die Gehaltsstufe “hoch” die Bayes-Zuordnung in Abhängigkeit von p.

(c) Es werden nun Fehlklassifikationskosten berücksichtigt. Bestimmen Sie die kostenoptimale Bayes-Zuordnung in Abhängigkeit vom Parameter p für die Gehaltsstufe “niedrig” und den Kosten aus der folgenden Tabelle:

c_{ij}	j=1	j=2
i=1	0	1
i=2	2	0