General Instruction: Answer all questions. Print your answer, your name, number the pages and number the problems on provided exam. Write on one side only. Please use only black pens or pencils.

Notation of Multiple Linear Regression Model (1): $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ where

$$\mathbf{Y} = \begin{pmatrix} \mathbf{y_1} \\ \mathbf{y_2} \\ \mathbf{:} \\ \mathbf{:} \\ \mathbf{y_n} \end{pmatrix} \quad \mathbf{X} = \begin{pmatrix} \mathbf{1} & \mathbf{x_{11}} & \mathbf{x_{12}} & \mathbf{...} & \mathbf{x_{1k}} \\ \mathbf{1} & \mathbf{x_{21}} & \mathbf{x_{22}} & & \mathbf{x_{2k}} \\ & \mathbf{...} & & & \\ & \mathbf{...} & & & \\ \mathbf{1} & \mathbf{x_{n1}} & \mathbf{x_{n2}} & & \mathbf{x_{nk}} \end{pmatrix} \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \mathbf{:} \\ \beta_{\mathbf{k}} \end{pmatrix} \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \mathbf{:} \\ \mathbf{:} \\ \varepsilon_{\mathbf{n}} \end{pmatrix} \quad \mathrm{E}(\boldsymbol{\varepsilon}) = 0, \ \ \mathrm{Cov}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}$$

1. The Model (1)
   a) Give all parameters of the model that are to be estimated.
   b) Give the list square estimates of these all parameters in matrix notation.
   c) Are all the list square estimates also the maximum likelihood estimates? If yes, give the ones that are different.
   d) Estimable is defined as: $\ell'\boldsymbol{\beta}$ is estimable if there exists a vector $\mathbf{c} \in \mathbf{R}_{(\mathbf{n})}$ such that $\mathbf{E}(\mathbf{c}'\mathbf{y}) = \ell'\boldsymbol{\beta}$. Show that $\ell'\boldsymbol{\beta}$ is estimable if and only if $\ell' \in \mathbf{R}(\mathbf{X}')$ and $\ell = \mathbf{X}'\mathbf{c}$, where $\mathbf{c} \in \mathbf{R}_{(\mathbf{n})}$, where ℓ is a $(\mathrm{k}+1)\times 1$ real vector.
   e) State the Gauss-Markov Theorem and give the proof of the theorem.
   f) Interpret $\hat{\beta}_3$

2. The Inferences for Model (1)
   a) Assume that $\boldsymbol{\varepsilon} \sim \mathbf{N}(\mathbf{0}, \sigma^2\mathbf{I})$ give the test statistic for testing $\mathbf{H}_o: \mathbf{C}\boldsymbol{\beta}=\boldsymbol{\gamma}$ vs $\mathbf{H}_a: \mathbf{C}\boldsymbol{\beta} \neq \boldsymbol{\gamma}$ where m is the rank of **C,** a m×(k+1) matrix.
   b) Show that $\mathbf{y}'\left[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right]\mathbf{y}/\sigma^2$ follows $\chi^2_{(\mathbf{n}-\mathbf{k}-\mathbf{1})}$ with degree of freedom n-k-1.
   c) Can the test be a Maximum Likelihood Ratio Test (MLRT)? Give the merit for the test to be a maximum likelihood ratio test.
   d) Give the **C** and $\boldsymbol{\gamma}$ in $\mathbf{H}_o: \mathbf{C}\boldsymbol{\beta}=\boldsymbol{\gamma}$ vs $\mathbf{H}_a: \mathbf{C}\boldsymbol{\beta} \neq \boldsymbol{\gamma}$ for testing $\boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \mathbf{:} \\ \beta_{\mathbf{k}} \end{pmatrix} = \mathbf{0}$
   e) Give the $100(1-\alpha)\%$ confidence interval in matrix notation for $\ell'\boldsymbol{\beta}$ where ℓ is a $(\mathrm{k}+1)\times 1$ real vector.
   f) Give the $100(1-\alpha)\%$ confidence region in matrix notation for $\boldsymbol{\beta}$.

3. The Diagnostics

For Model (2) $\mathbf{Y}_{(\mathbf{i})} = \mathbf{X}_{(\mathbf{i})}\boldsymbol{\beta} + \boldsymbol{\varepsilon}_{(\mathbf{i})}$ where $\mathbf{Y}_{(\mathbf{i})}$ $\mathbf{X}_{(\mathbf{i})}$ $\boldsymbol{\varepsilon}_{(\mathbf{i})}$ are the $\mathbf{Y}$ $\mathbf{X}$ $\boldsymbol{\varepsilon}$ with the i[th] row removed.

$\mathbf{e}_{(\mathbf{i})} = \left(\mathbf{I}_{(\mathbf{i})} - \mathbf{H}_{(\mathbf{i})}\right)\mathbf{y}_{(\mathbf{i})}$ $\mathbf{H}_{(\mathbf{i})} = \mathbf{X}_{(\mathbf{i})}\left(\mathbf{X}'_{(\mathbf{i})}\mathbf{X}_{(\mathbf{i})}\right)^{-1}\mathbf{X}'_{(\mathbf{i})}$

   a) Defined the studentized residuals e_i^*.

b) Show that $\frac{e_i}{\sigma\sqrt{1-h_{ii}}}$ and $\frac{1}{\sigma^2}\mathbf{e}_{(i)}'\mathbf{e}_{(i)}$ are independent.

c) List the five type of diagnostics for regression models with the treatments. Comment om the importance of these diagnostics.

4. The Example

$\mathbf{Y_p} =$ (Time)

219
264
226
242
220
229
253
233
260
235
247

$\mathbf{X_p} =$ (SC DC MR TR)

3	3	77	29
3	3	95	27
6	2	68	24
6	5	80	25
7	1	70	19
3	1	66	30
7	2	81	24
3	2	86	27
4	6	85	25
8	2	72	21
8	0	82	26

$\mathbf{Y_I} =$ (Time)

316
251
216
303
280
285
268
269
307
204
283
233
266

$\mathbf{Y_I} =$ (SC DC MR TR)

3	4	163	30
2	1	141	16
4	2	135	16
13	3	135	16
4	3	138	18
6	2	141	22
2	4	139	25
2	3	152	18
5	3	143	16
5	0	135	17
4	4	151	16
3	2	126	25
4	1	148	20

are the time taken by professional dietitians and interns for four patient contact activities.

a) If model (1) $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ is used, where $\mathbf{Y} = \begin{pmatrix} \mathbf{Y_I} \\ \mathbf{Y_P} \end{pmatrix}$, $\mathbf{X} = \begin{pmatrix} \mathbf{X_I} & \mathbf{0} \\ \mathbf{0} & \mathbf{X_P} \end{pmatrix}$, $\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta_I} \\ \boldsymbol{\beta_P} \end{pmatrix}$, $\boldsymbol{\varepsilon} = \begin{pmatrix} \boldsymbol{\varepsilon_I} \\ \boldsymbol{\varepsilon_P} \end{pmatrix}$ what inference the researcher intend to make?

b) Give the $\mathbf{C}$ and γ in $\mathbf{H}_o: \mathbf{C}\boldsymbol{\beta}=\gamma$ vs $\mathbf{H}_a: \mathbf{C}\boldsymbol{\beta} \neq \gamma$ for testing $\boldsymbol{\beta_I} \neq \boldsymbol{\beta_P}$

c) Give the test statistics in matrix notation and the critical value for 0.05 the level of significance in b).

d) Suppose model (2) is given as $\mathbf{Y} = \mathbf{X_2}\boldsymbol{\beta_2} + \boldsymbol{\varepsilon}$ where $\mathbf{X_2} = \begin{pmatrix} \mathbf{X_I} \\ \mathbf{X_P} \end{pmatrix}$ and $\boldsymbol{\beta}_2' = \begin{pmatrix} \boldsymbol{\beta_{SC}} & \boldsymbol{\beta_{DC}} & \boldsymbol{\beta_{MR}} & \boldsymbol{\beta_{TR}} \end{pmatrix}$. Can this model be used to test the whether the time taken by the professional dietitians and the interns have same regression models?

e) If the hypothesis test in part b) is not rejected, should model (2) be used? Why? Explain.

Use the attached Minitab output for the following questions:

f) Identify cases that don't belong to the model using the e_i^* with $\alpha = 0.05$.

g) Identify influential cases using the recommended method in the textbook.

h) Is there any multicollinearity exists why or why not?

i) If yes, how much the inflation on the $\mathrm{Var}(b_{PMR})$ because of the multicollinearity?

Minitab Output

Time	SC	DC	MR	TR	SC_P	DC_P	MR_P	TR_P	SRES1	HI1
219	3	3	77	29	0	0	0	0	-1.03606	0.237120
264	3	3	95	27	0	0	0	0	1.26512	0.162509
226	6	2	68	24	0	0	0	0	-0.05914	0.122687
242	6	5	80	25	0	0	0	0	-1.01506	0.342927
220	7	1	70	19	0	0	0	0	0.57086	0.205412
229	3	1	66	30	0	0	0	0	0.66019	0.469420
253	7	2	81	24	0	0	0	0	0.34266	0.149624
233	3	2	86	27	0	0	0	0	0.19370	0.174565
260	4	6	85	25	0	0	0	0	0.84121	0.501580
235	8	2	72	21	0	0	0	0	0.00508	0.273816
247	8	0	82	26	0	0	0	0	-0.57894	0.370809
233	3	2	126	25	0	0	0	0	-1.51961	0.302139
266	4	1	148	20	0	0	0	0	0.82187	0.687392
316	0	0	0	0	3	4	163	30	-0.08713	0.517764
251	0	0	0	0	2	1	141	16	1.43973	0.356777
216	0	0	0	0	4	2	135	16	-1.75458	0.117877
303	0	0	0	0	13	3	135	16	-0.46875	0.844607
280	0	0	0	0	4	3	138	18	0.79521	0.105106
285	0	0	0	0	6	2	141	22	1.08289	0.262019
268	0	0	0	0	2	4	139	25	-0.55534	0.392880
269	0	0	0	0	2	3	152	18	-0.41569	0.252576
307	0	0	0	0	5	3	143	16	1.79761	0.180551
204	0	0	0	0	5	0	135	17	-1.48577	0.542610
283	0	0	0	0	4	4	151	16	-1.07442	0.427233

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
SC	8.73	2.17	4.02	0.001	5.26
DC	5.05	3.28	1.54	0.143	3.25
MR	0.881	0.183	4.82	0.000	10.53
TR	4.359	0.837	5.21	0.000	16.86
SC_P	5.05	1.85	2.73	0.015	3.29
DC_P	14.70	5.08	2.89	0.011	7.14
MR_P	1.393	0.220	6.32	0.000	32.58
TR_P	0.54	1.53	0.35	0.730	29.43

Regression Equation

Time = 8.73 SC + 5.05 DC + 0.881 MR + 4.359 TR + 5.05 SC_P + 14.70 DC_P + 1.393 MR_P
+ 0.54 TR_P