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

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

1. The Model (1)
   a) Give the list square estimates of all the parameters in matrix notation.
   b) Define estimable of $\boldsymbol{\ell}'\boldsymbol{\beta}$, a linear function of $\boldsymbol{\beta}$.
   c) State the Gauss-Markov Theorem and prove the theorem.
   d) Interpret $\hat{\beta}_3$

2. The Inferences for Model (1)
   a) Assume that $\boldsymbol{\varepsilon} \sim \mathbf{N}\left(\mathbf{0}, \sigma^2 \mathbf{I}\right)$, give the test statistic for testing H_o: $\mathbf{C}\boldsymbol{\beta} = \boldsymbol{\gamma}$ vs H_a: $\mathbf{C}\boldsymbol{\beta} \neq \boldsymbol{\gamma}$ where m is the rank of **C,** a m×(k+1) matrix.
   b) Prove the test statistics in part a) has the F distribution with degree freedom of m and n-k-1.

3. The Diagnostics

   For Model (2) $\mathbf{Y_{(i)}} = \mathbf{X_{(i)}}\boldsymbol{\beta} + \boldsymbol{\varepsilon}_{\mathbf{(i)}}$ where $\mathbf{Y}_{(i)}$ $\mathbf{X}_{(i)}$ $\boldsymbol{\varepsilon}_{(i)}$ are the $\mathbf{Y}$ $\mathbf{X}$ $\boldsymbol{\varepsilon}$ with the i^{th} row removed;

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

   a) Defined the studentized residuals e_i^*.
   b) Show that $\mathbf{b}_{(\mathbf{i})}$ and $\mathbf{e}_{(\mathrm{i})}$ are independent.
   c) Prove that e_i^* has t distribution with degrees of freedom n-k-2.
   d) Explain how to use e_i^* in regression diagnosis.

4. The Example

$$\mathbf{Y_p} = \begin{array}{c} \texttt{Time} \\ \begin{pmatrix} 219 \\ 264 \\ 226 \\ 242 \\ 220 \\ 229 \\ 253 \\ 233 \\ 260 \\ 235 \\ 247 \end{pmatrix} \end{array} \quad \mathbf{X_p} = \begin{array}{c} \texttt{SC DC MR TR} \\ \begin{pmatrix} 3 & 3 & 77 & 29 \\ 3 & 3 & 95 & 27 \\ 6 & 2 & 68 & 24 \\ 6 & 5 & 80 & 25 \\ 7 & 1 & 70 & 19 \\ 3 & 1 & 66 & 30 \\ 7 & 2 & 81 & 24 \\ 3 & 2 & 86 & 27 \\ 4 & 6 & 85 & 25 \\ 8 & 2 & 72 & 21 \\ 8 & 0 & 82 & 26 \end{pmatrix} \end{array} \quad \mathbf{Y_I} = \begin{array}{c} \texttt{Time} \\ \begin{pmatrix} 316 \\ 251 \\ 216 \\ 303 \\ 280 \\ 285 \\ 268 \\ 269 \\ 307 \\ 204 \\ 283 \\ 233 \\ 266 \end{pmatrix} \end{array} \quad \mathbf{Y_I} = \begin{array}{c} \texttt{SC DC MR TR} \\ \begin{pmatrix} 3 & 4 & 163 & 30 \\ 2 & 1 & 141 & 16 \\ 4 & 2 & 135 & 16 \\ 13 & 3 & 135 & 16 \\ 4 & 3 & 138 & 18 \\ 6 & 2 & 141 & 22 \\ 2 & 4 & 139 & 25 \\ 2 & 3 & 152 & 18 \\ 5 & 3 & 143 & 16 \\ 5 & 0 & 135 & 17 \\ 4 & 4 & 151 & 16 \\ 3 & 2 & 126 & 25 \\ 4 & 1 & 148 & 20 \end{pmatrix} \end{array}$$

Suppose the researcher wanted to test whether the time taken by professional dietitians and interns for the four patient contact activities are different (alternative).

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

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

b) Show that $\mathbf{C}\beta$ is estimable.

c) Give the test statistic in matrix notation for part a).

d) Show that the test statistic in part c) has F distribution. Give the degrees of freedoms of the distribution.

e) Suppose model (2) is given as $\mathbf{Y} = \mathbf{X_2\beta_2} + \boldsymbol{\varepsilon}$ where $\mathbf{X_2} = \begin{pmatrix} \mathbf{X_I} \\ \mathbf{X_P} \end{pmatrix}$ and $\boldsymbol{\beta}_2' = (\beta_{SC} \quad \beta_{DC} \quad \beta_{MR} \quad \beta_{TR})$, can this model be used to test the hypothesis in part a)? Explain.