1. (40 points) Assume the model is

$$y = X\beta + \varepsilon, \tag{1}$$

where $y = (y_1, \cdots, y_n)'$ is the vector of the observations; $X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}$ is the design matrix; $\beta = (\beta_1, \cdots, \beta_p)'$ is the vector of the unknown parameters, and $\varepsilon = (\varepsilon_1, \cdots, \varepsilon_n)'$ is a vector of errors with mean vector, $\mathbf{0}$, and covariance matrix, $\sigma^2 I_n$.

a. Find the least-squares estimate of β, $\hat{\beta}$.

b. Show $\hat{\beta}$ is a unbiased estimator for β and find its covariance matrix.

c. Give a geometrical interpretation of least-squares.

d. Write down SS_T; SS_R and SS_{Res} in terms of y, X and $\hat{\beta}$.

e. Give the corresponding ANOVA table.

2. (10 points) Suppose we want to find the least-squares estimator of β in the model $y = X\beta + \varepsilon$ subject to a set of equality constraints on β, say $T\beta = c$. Show that the estimator is

$$\tilde{\beta} = \hat{\beta} + (X'X)^{-1}T'[T(X'X)^{-1}T]^{-1}(c - T\hat{\beta}), \tag{2}$$

where $\hat{\beta} = (X'X)^{-1}X'y$.

3. (10 points) Give 25 observations, $y_1, \cdots, y_{25}$, and assume the model is $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$. Please fill the following ANOVA table up.

Source Variation	Sum of Square	Degree of Freedom	Mean Square	F_0
Regression	5550.8166			
Residual				
Total	5784.5426			

4. (10 point) In partial regression plot, define that

$$e(y|X) = y - \hat{y} = y - X\hat{\beta}.$$

Assume the model is

$$y = X\beta + \varepsilon = X(j)\beta(j) + \beta_j x_j + \varepsilon, \tag{3}$$

where x_j is the j^{th} variable; $X(j)$ is the original X matrix with the j^{th} regressor (x_j) removed, and $\beta(j) = (\beta_0, \cdots, \beta_{j-1}, \beta_{j+1}, \cdots, \beta_k)'$. Find the relationship between $e(y|X(j))$ and $e(x_j|X(j))$.

5. (20 points) Consider the following analysis of variance table:

Terms added sequentially (from first to least)

	Df	Sum of square	Mean Square	F-value	Pr(F)
x_1	1	5783.780	5783.780	89.49636	0.0000000
x_2	1	811.715	811.715	12.56022	0.0017316
x_3	1	1181.505	1181.505	18.28223	0.0002832
x_4	1	5303.029	5303.029	82.05737	0.0000000
x_5	1	115.504	115.504	1.78727	0.1943335
Residuals	23	1486.395	64.626		

a. Fill out the missing values in the following ANOVA table:

Source Variation	Sum of Square	Degree of Freedom	Mean Square	F_0
Regression	?	?	?	?
Residual	?	?	?	
Total	?	?		

b. Find $SS(\beta_3, \beta_4 | \beta_0, \beta_1, \beta_2)$.

c. Which one is larger $SS(\beta_3, \beta_4 | \beta_0, \beta_1, \beta_2)$ or $SS(\beta_3, \beta_4)$? Why?

6. (10 points) Write out the five major assumptions in the study of regression analysis.