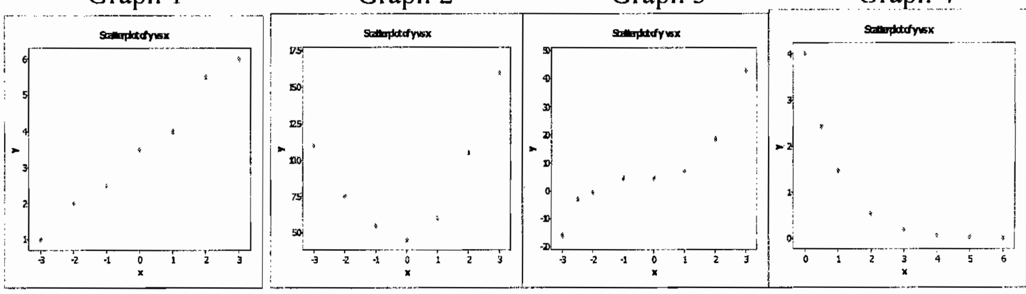
Question 1. (5.5 + 5.5 = 11 marks)

(A) The graphs below show scatter plots of some data on two variables y and x

Graph 1 | Graph 2 | Graph 3 | Graph 4



i) State the kind of relationship likely to exist between the variables y and x in each case (2.0)

ii) Is it possible to find a transformation that might make the relationship of Graph 4 linear? (1.0)

iii) Consider the straight line relationship $y = \beta_0 + \beta_1 x$ between two variables y and x.

(a) Why are β_0 and β_1 called the "intercept" and the "slope", respectively? (1.0)

(b) Obtain the inverse relationship (i.e. write x in terms of y). How much would be the change in x if y is changed by a unit amount? (1.5)

(B) Consider the simple linear regression model $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \ (i = 1, 2, ..., n)$.

i) State the second-order distributional assumptions in the model (1.0)

ii) State the least squares estimators (l.s.e.) of the parameters (which are obtained by minimizing the Error Sum of Squares $Q = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 x_i)^2$ with respect to the parameters). (1.5)

iii) Derive $\text{var}(\hat{\beta}_1)$. (2.0)

iv) State $\text{var}(\hat{\beta}_0)$ (1.0)

Question 2. (6 marks)

Consider the data on the response variable Y and predictor variable x given in the following table.

x	20	20	20	20	20	25	25	25	25	25	30	30	30	30	30	35	35
y	7	7	15	11	9	12	17	12	18	18	14	18	18	19	19	19	25
x	35	35	35	40	40	40	40	40									
y	22	19	23	7	10	11	15	11									

Use MINITAB to do a simple linear regression analysis answering the following:

(a) What is the regression line equation? (0.5)

(b) What are the values of the least squares estimates $\hat{\beta}_0$ and $\hat{\beta}_1$? (0.5)

(c) Obtain a 95% confidence interval for each of the two parameters. (1.0)

(d) State the Analysis of Variance (ANOVA) Table. (1.0)

(e) State the hypotheses of interest and your conclusions about them (1.0)

(d) How good is the fit of the regression line? (0.5)

(e) Use residual plots to decide if the model assumptions are justified. (1.5)

Question 3. (8 marks)

Consider the simple linear regression model $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i\ (i = 1, 2, \ldots, n)$. This model may be rewritten in matrix notation as $\mathbf{Y} = \mathbf{X\beta} + \boldsymbol{\varepsilon}$.

i) Write in full the terms $\mathbf{Y}$, $\mathbf{X}$, $\boldsymbol{\beta}$ and $\boldsymbol{\varepsilon}$ (0.5)

ii) State the second-order distributional assumptions in matrix notation and then the normal theory assumptions using matrix notations (0.5)

iii) State the elements of $\mathbf{X'X}$ and $\mathbf{X'Y}$ (1.0)

iv) Show that the error sum of squares $Q = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 x_i)^2$ is $(\mathbf{Y} - \mathbf{X\beta})'(\mathbf{Y} - \mathbf{X\beta})$ (1.0)

v) State the normal equations (least squares equations) in matrix form. (0.5)

vi) Assuming $\mathbf{X}$ to be full-rank show that the l.s.e.of $\boldsymbol{\beta}$ is given by $\hat{\boldsymbol{\beta}} = (\mathbf{X'X})^{-1}\mathbf{X'Y}$ (0.5)

vii) Show that $\hat{\boldsymbol{\beta}}$ is an unbiased estimator of $\boldsymbol{\beta}$. (1.0)

viii) Derive $\mathbf{Cov}(\hat{\boldsymbol{\beta}})$, the variance-covariance matrix of $\hat{\boldsymbol{\beta}}$. (1.5)

ix) What is the Maximum Likelihood Estimate of $\boldsymbol{\beta}$ when normality assumptions are made? (0.5)

x) What is the Maximum Likelihood Estimate of error variance σ^2 under normality? (1.0)

Question 4. (7 marks)

Consider the multiple linear regression model $Y_i = \beta_0 + \beta_1 x_{i1} + ... + \beta_k x_{ik} + \varepsilon_i \, (i = 1, 2, ..., n)$.

i) State the model and the assumptions in matrix notation. (0.5)

ii) Show that the vector of fitted values $\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{HY}$ where $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. (0.5)

iii) Show that the matrix $\mathbf{H}$ is a symmetric idempotent matrix. (1.0)

iv) Show that trace $\mathbf{H}$ is $k + 1$. (0.5)

v) Show that the residual vector may be written as $\mathbf{e} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$. (0.5)

vi) Show that the $(\mathbf{I} - \mathbf{H})$ is symmetric, idempotent and trace of $(\mathbf{I} - \mathbf{H})$ is $n - k + 1$. (1.0)

vii) Show that the *Residual Sum of Squares* SSE may be written as $\text{SSE} = \mathbf{Y}'(\mathbf{I} - \mathbf{H})\mathbf{Y}$. (1.0)

viii) The *Total (corrected) Sum of Squares* SST is defined as $\text{SST} = \sum_{i=1}^{n} Y_i^2 - [(\sum_{i=1}^{n} Y_i)^2 / n]$. Show that $\text{SST} = \mathbf{Y}'[\mathbf{I} - (\frac{1}{n})\mathbf{J}]\mathbf{Y}$ where $\mathbf{J}$ is the matrix with all elements equal to 1. (1.0)

ix) Show that the *Regression Sum of Squares* SSR may be written as $\text{SSR} = \mathbf{Y}'[\mathbf{H} - (\frac{1}{n})\mathbf{J}]\mathbf{Y}$ (1.0)

Question 5. (9 marks)

(A) In a small-scale regression study, the following data were obtained:

i	1	2	3	4	5	6
X_{1i}	9	6	18	6	23	10
X_{2i}	31	39	5	47	3	29
Y_i	41	32	74	29	90	54

Assume that the multiple linear regression model $Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i \; (i = 1, 2, ..., 6)$ is appropriate for above data where the errors ε_i are assumed to be i.i.d. $N(0, \sigma^2)$ r.v.s. Let $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$. Use MINITAB or any other program for the following.

i) Obtain $\mathbf{X'X}$ and $\mathbf{X'Y}$ (1.0)

ii) Obtain the l.s.e. $\hat{\boldsymbol{\beta}}$ and write it's transpose (only) (1.0)

iii) Compute the vector of fitted values $\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ write it's transpose (only) (1.0)

iv) Compute the matrix $\mathbf{H}$ (1.0)

v) Compute the residual vector $\mathbf{e}$ and write it's transpose only (1.0)

vi) Compute the regression sum of squares SSR (1.0)

vii) Compute the estimated $\{\mathbf{Cov}(\hat{\boldsymbol{\beta}})\}$ (1.0)

viii) Compute the predicted value $\hat{Y}_h$ at $X_{h1} = 30$, $X_{h2} = 40$ (1.0)

ix) Obtain the estimate of $\mathrm{Var}\{\hat{Y}_h\}$ when $X_{h1} = 30$, $X_{h2} = 40$ (1.0)

Question 6 (9 marks)

(A) Whe gasoline is pumped into the tank of a car, vapors are vented into the atmosphere. An experiment was performed to determine whether Y, the amount of vapor, can be predicted using the following variables X_1= tank temperature (°F), X_2= gasoline temperature, X_3 = vapor pressure in tank (psi), and X_4 = vapor pressure of gasoline (psi). The data obtained are as in the following table.

y	29	24	26	22	27	21	33	34	32
x_1	**33**	**31**	**33**	**37**	**36**	**35**	**59**	**60**	**59**
x_2	**53**	**36**	**51**	**51**	**545**	**35**	**56**	**60**	**60**
x_3	**3.32**	**3.10**	**3.18**	**3.39**	**3.20**	**3.03**	**4.78**	**4.72**	**4.60**
x_4	**3.42**	**3.26**	**3.18**	**3.08**	**3.41**	**3.03**	**4.57**	**4.72**	**4.41**

Use MINITAB's multiple regression program to obtain

(i) SSR $(X_2|X_1)$, SSR $(X_3|X_1,X_2)$, SSR $(X_4|X_1,X_2,X_3)$ (only state the answers) (3.0)

(ii) An analysis of the data and make your conclusions (3.0)

(B)

i) State the d-th degree polynomial regression model in one variable (0.5)

ii) State the second-order (degree 2) polynomial regression model in 2 variables and explain the terms in the model (1.0)

iii) Explain briefly how you can do regression with qualitative predictor variables. (1.5)