Tutorial Worksheet 5 - Advanced Regression Analysis

Problem 1.

A study was made on the effect of temperature on the yield of a chemical process, the following data were collected:

X	-5	-4	-3	-2	-1	0	1	2	3	4	5
Y	1	5	4	7	10	8	9	13	14	13	18

(a) Assuming a model, $Y = \beta_0 + \beta_1 X + \epsilon$, what are the least square estimates of β_0 and β_1? What is the fitted equation?

(b) Construct the ANOVA table and test the hypothesis $H_0 : \beta_1 = 0$ with $\alpha = 0.05$.

(c) What are the confidence limits for β_1 at $\alpha = 0.05$?

(d) What are the confidence limits for the true mean value of Y when $X = 3$ at $\alpha = 0.05$?

(e) What are the confidence limits at $\alpha = 0.05$ level of significance for the difference between the true mean value of Y when $X_1 = 3$ and the true mean value of Y when $X_2 = -2$?

Problem 2.

There are a few occasions where it makes sense to fit a model without an intercept β_0. If there were an occasion to fit the model $Y = \beta X + \epsilon$ to a set of data $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, the least square estimate of β would be

$$\hat{\beta} = b = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$$

Suppose you have a programmed calculator that will fit only the intercept model $Y = \beta_0 + \beta_1 X + \epsilon$, but you want to fit a non-intercept model. By adding one more fake data point $(m\bar{x}, m\bar{y})$ to the data above, where $m = \frac{n}{(n+1)^{1/2}-1} = \frac{n}{a}$, say, and letting the calculator fit $Y = \beta_0 + \beta_1 X + \epsilon$, can you estimate β by using b?

Problem 3.

Fit the model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$ for the data given below. Provide an ANOVA table and perform the partial F-tests to test $H_0 : \beta_i = 0$ vs $H_1 : \beta_i \neq 0$ for $i = 1, 2$; given the other variable is already in the model. Comment on the relative contributions of the variables X_1 and X_2, depending on whether they enter the model first or second. Find the regression equation.

X_1	-5	-4	-1	2	2	3	3
X_2	5	4	1	-3	-2	-2	-3
Y	11	11	8	2	5	5	4

Problem 4.

Given a 2-variables linear regression problem $Y = \beta_1 + \beta_2 X_1 + \beta_3 X_2 + \epsilon$, yield the following

$$X^T X = \begin{bmatrix} 33 & 0 & 0 \\ 0 & 40 & 20 \\ 0 & 20 & 60 \end{bmatrix}, \quad X^T Y = \begin{bmatrix} 132 \\ 24 \\ 92 \end{bmatrix} \quad \text{and} \quad \sum (Y - \bar{Y})^2 = 150.$$

(a) What is the sample size?

(b) Write the normal equations and solve for the regression coefficients.

(c) Estimate the standard error of β_2 and test the hypothesis that $\beta_2 = 0$

(d) Compute R^2 and interpret it. Also, interpret the value of regression coefficients.

(e) Predict the value of y given $x_1 = -4$ and $x_2 = 2$

(f) Comment on the possibilities of any regressors being a dummy variable.

X_1	-4	3	1	4	-3	-1
X_2	1	2	3	4	5	6
X_3	3	-5	-4	-8	-2	-5
Y	7.4	14.7	13.9	18.2	12.1	14.8

Problem 5.
Can we use the data in Table to get a unique fit to the model $Y = \beta_0 X_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$

Problem 6.
Show that in linear regression with a β_0 term in the model:

(a) The correlation between the vector e and Y is $(1 - R^2)^{1/2}$. This result implies that it is a mistake to find defective regressions by a plot of residuals e_i versus observations Y_i as this always shows a slope.

(b) Show that the slope is $1 - R^2$.

(c) Show further that the correlation between e and $\hat{Y}$ is zero.

Problem 7.
Prove that the multiple coefficients R^2 is equal to the square of the correlation between Y and $\hat{Y}$.

Problem 8.
A new born baby was weighted weekly. Twenty such weights are shown below, recorded in ounces. Fit to the data, using orthogonal polynomials, a polynomial model of degree justified by the accuracy of the figures, that is, test as you go along for the significance of the linear, quadratic and so fourth, terms.

No. of weeks	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Weights	141	144	148	150	158	161	166	170	175	181	189	194	196	206	218	229	234	242	247	257

Problem 9.
If you are asked to fit a straight line to the data $(X, Y) = (1, 3), (2, 2.5), (2, 1.2), (3, 1),$ and $(6, 4.5)$. What would you do about it?

Problem 10.
Your friend says he/she has fitted a plane to $n = 33$ observations on (X_1, X_2, Y) and his/ her overall regression (given β_0) is just significant at the $\alpha = 0.05$ level.You ask him/ her for R^2 value but s/he does not know. You work it out for him/ her based on what s/he has told you.

Problem 11.
You are given a regression printout that shows a planar to fit X_1, X_2, X_3, X_4, X_5 plus an intercept term obtained from a set of 50 observations. The overall F for regression is ten times as high as the 5% upper-tail F percentage point. How big is R^2 ?

Problem 12.
Consider the simple linear regression model: $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, where the variance of ϵ_i is proportional to x_i^2, i.e., $\mathrm{V}(\epsilon_i) = \sigma^2 x_i^2$ (assumption of constant variance is NoT satisfied).

(a) Suppose that we use these transformation $y' = \frac{y}{x}$ and $x' = \frac{1}{x}$. Is this a variance-stabilizing transformation?

(b) What are the relationships between the parameters in the original and the transformed model?

(c) Suppose we use the method of weighted least squares with $w_i = \frac{1}{x_i^2}$. Is this equivalent to the transformation introduced in part (a).

Problem 13.
Consider the simple linear regression model $y_t = \beta_0 + \beta_1 x_t + \epsilon_t$ where the errors are generated by second-order autoregressive process

$$\epsilon_t = \rho_1 \epsilon_{t-1} + \rho_2 \epsilon_{t-2} + z_t.$$

z_t is an NID $\left(0, \sigma_z^2\right)$ random variable, and ρ_1 and ρ_2 are auto-correlation parameters. Discuss how the Cochrane-Orcutt iterative procedure could be used in this situation. What transformations would be appropriate on the variables y_t and x_t ? How would you estimate the parameters ρ_1 and ρ_2?

Problem 14.

The following 24 residuals from a straight line fit are equally spaced and are given in time sequential order. Is there any evidence of lag-1 serial correlation?

$$8, -5, 7, 1, -3, -6, 1, -2, 10, 1, -1, 8, -6, 1, -6, -8, 10, -6, 9, -3, 3, -5, 1, -9$$

Use a two-sided test at level $\alpha = 0.05$

Problem 15.

Estimate the parameters α & β in the non-linear model $Y = \alpha + (0.49 - \alpha)e^{-\beta(X-8)}$ from the following observations:

X	8	10	12	14	16	18	20	22	24	26	28	30	32	34	36	38	40	42
Y	0.490	0.475	0.450	0.433	0.45	0.423	0.407	0.407	0.407	0.405	0.393	0.405	0.400	0.395	0.400	0.390	0.407	0.390

Problem 16.

Look at these data. I don't know whether to fit two straight lines, one straight line or what. How to solve this dilemma?

X	8	0	12	2
Y	5.3	0.9	7.1	2.4

(a) Set A

X	9	7	8	6
Y	5.1	4.4	5.2	3.8

(b) Set B

Problem 17.

Let $\underset{\sim}{x}$ be a vector of p random variables and α_k is a vector of p constants and we write $\alpha_k' \underset{\sim}{x} = \sum_{j=1}^{p} \alpha_{kj} x_j$. Also, let S be the (known) sample covariance matrix for the random variable $\underset{\sim}{x}$. For $k = 1, 2$, show that the k^{th} principal component is given by $z_k = \alpha_k' \underset{\sim}{x}$ where α_k is an eigenvector of S corresponding to its k^{th} largest eigenvalue λ_k. [Principal component Regression]

Problem 18.

Find the maximum and minimum value of $\underset{\sim}{x}' A \underset{\sim}{x}$ subject to $\underset{\sim}{x}' \underset{\sim}{x} = 1$.

Problem 19.

Show that $\|\hat{\beta}_{\text{Ridge}}\|$ increases as its tuning parameter $\lambda \to 0$. Does the same property hold for the LASSO regression?

Problem 20.

Consider a two-class logistic regression problem with $x \in \mathbb{R}$. Characterize the maximum-likelihood estimates of the slope and intercept parameter if the sample x_i for the two classes are separated by a point $x_0 \in \mathbb{R}$. Generalize this result to (a) $x \in \mathbb{R}^p$ and (b) more than two classes.