Ve401 Probabilistic Methods in Engineering

Spring 2020 — Assigment 8

Date Due: 11:00 PM, Thursday, the 30th of April 2020

This assignment has a total of **(55 Marks)**.

Exercise 8.1
Prove that in simple linear regression

$$\text{SSE} = S_{yy} - b_1 S_{xy}$$

(2 Marks)

Exercise 8.2
In the experiment "Simple Harmonic Motion: Oscillations in Mechanical Systems" of the course Vp141 Physics Lab I, the spring coefficient is measured by using a Jolly balance. A spring is attached to the Jolly balance and weights are added to extend the spring. The extension L of the Jolly balance (not the actual spring extension) is recorded. For one spring the data (rounded) was obtained by two groups:

Group 1		Group 2	
L[cm]	m[g]	L[cm]	m[g]
4.88	0	4.95	0
6.92	4.7	7.00	4.7
8.99	9.5	9.10	9.5
11.09	14.3	11.20	14.3
13.18	19.1	13.30	19.1
15.26	23.9	15.41	24.0
17.39	28.7	17.51	28.7

Use Mathematica to do the following exercises:

i) For the given data, perform a simple linear regression for the random variable L as a function of the (non-random) parameter m. Plot the regression line.
(2 Marks)

ii) Calculate the value of R^2 and check for significance of regression.
(2 Marks)

iii) Perform a test for lack of fit. Is the linear model appropriate?
(2 Marks)

(Many thanks to Li Yingyu, Teachng Assistant for Vp241, for providing the data and advice on the experiment.)

Exercise 8.3
An article in the Journal of the American Statistical Association [Markov Chain Monte Carlo Methods for Computing Bayes Factors: A Comparative Review (2001, Vol. 96, pp. 11221132)] analyzed the tabulated data on compressive strength parallel to the grain versus resin-adjusted density for specimens of radiata pine.

Compressive Strength	Density
3040	29.2
3840	30.7
2470	24.7
3610	32.3
3480	31.3
3810	31.5
2330	24.5
1800	19.9
3110	27.3
3160	27.1
2310	24.0
4360	33.8
1880	21.5
3670	32.2

Compressive Strength	Density
1740	22.5
2250	27.5
2650	25.6
4970	34.5
2620	26.2
2900	26.7
1670	21.1
2540	24.1
3800	32.7
4600	32.6
1900	22.1
2530	25.3
2920	30.8
4990	38.9

Compressive Strength	Density
1670	22.1
3310	29.2
3450	30.1
3600	31.4
2850	26.7
1590	22.1
3770	30.3
3850	32.0
2480	23.2
3570	30.3
2620	29.9
1890	20.8
3030	33.2
3030	28.2

i) Fit a linear regression model for the dependence of the compressive strength $Y \mid x$ on the density x.
(1 Mark)

ii) Estimate σ^2 for this model.
(1 Mark)

iii) Find 90% confidence intervals for the slope and the intercept.
(2 Marks)

iv) Test for significance of regression.
(1 Mark)

v) Calculate R^2 for this model. Provide an interpretation of this quantity.
(2 Marks)

vi) Plot the residuals e_i versus the density x_i. Does the assumption of constant variance seem to be satisfied?
(2 Marks)

Exercise 8.4
Recall that

$$P = \frac{1}{n}\begin{pmatrix} 1 & 1 & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}, \qquad H = X(X^T X)^{-1} X^T$$

where X is the model specification matrix for multiple linear regression.

i) Show that $PH = HP = P$. Conclude that $H - P$ is an orthogonal projection and that

$$\mathrm{SSR} = \langle (H-P)Y, (H-P)Y \rangle.$$

(2 Marks)

ii) Show that $\operatorname{tr} P = 1$ and conclude $\operatorname{tr}(H - P) = p$.
(1 Mark)

iii) Follow the steps in the lecture slides to show that if $\beta = (\beta_0, 0, \dots, 0)$ (i.e., if $\beta_1 = \cdots = \beta_p = 0$), then SSR/σ^2 follows a chi-squared distribution with p degrees of freedom.
(3 Marks)

iv) Show that $(\mathbb{1} - H)(P - H) = (P - H)(\mathbb{1} - H) = 0$. Deduce that

$$\operatorname{ran}(P - H) \subset \ker(\mathbb{1} - H) \qquad \text{and} \qquad \operatorname{ran}(\mathbb{1} - H) \subset \ker(P - H).$$

Explain why this means that the eigenvectors of $H - P$ for the eigenvalue 1 are also eigenvectors of $\mathbb{1} - H$ for the eigenvalue 0 and vice-versa. Construct a matrix U which diagonalizes both $P - H$ and $\mathbb{1} - H$. Use U to show that SSR and SSE are the sums of squares of independent standard normal variables. Deduce that SSR and SSE are independent.
(5 Marks)

Exercise 8.5
In simple linear regression, the significance of regression is equivalent to testing $H_0\colon \beta_1 = 0$. The test can be performed using the statistic

$$T_{n-2} = \frac{B_1}{S/\sqrt{S_{xx}}}.$$

On the other hand, we might test $H_0\colon \beta_1 = 0$ using the statistic

$$F_{1,n-2} = \frac{\mathrm{SSR}}{\mathrm{SSE}/(n-2)} = \frac{\mathrm{SSR}}{S^2}.$$

Prove that both tests are mathematically equivalent.
(3 Marks)

Exercise 8.6

An article entitled A Method for Improving the Accuracy of Polynomial Regression Analysis in the Journal of Quality Technology (1971, pp. 149155) reported the following data for the dependence of the ultimate shear strength of a rubber compound (y, in psi) on the cure temperature (x, °F).

y	770	800	840	810	735	640	590	560
x	280	284	292	295	298	305	308	315

You are encouraged to use Mathematica to help with the calculations in the following exercises. Include a printout of your calculations with your submitted answers.

i) Fit the quadratic model $Y = \beta_0 + \beta_1 x + \beta_2 x^2 + E$ to these data.
(2 Marks)

ii) Test for significance of regression. What is the P-value you obtain?
(2 Marks)

iii) Test the hypothesis that $\beta_2 = 0$. What is the P-value you obtain?
(2 Marks)

iv) Plot the residuals and comment on model adequacy.
(2 Marks)

v) Give confidence intervals for β_0, β_1 and β_2.
(3 Marks)

vi) Give a prediction interval for Y when $x = 285$°F.
(1 Mark)

When fitting polynomial regression models, we often subtract $\overline{x}$ from each x value to produce a "standardized" regressor $x' := (x - \overline{x})/s_x$, where s_x is the standard deviation of x. This reduces the effects of dependencies among the model terms and often leads to more accurate estimates of the regression coefficients.

vii) Fit the standardized model $Y = \beta_0^* + \beta_1^* x' + \beta_2^* (x')^2 + E$.
(2 Marks)

viii) Use the standardized model to give confidence intervals for β_0, β_1 and β_2.
(3 Marks)

ix) Use the standardized model to give a prediction interval for Y when $x = 285$°F.
(1 Mark)

x) What can you say about the relationship between SSE and R^2 for the standardized and unstandardized models?
(3 Marks)

xi) Suppose that $y' = (y - \overline{y})/s_y$ is used in the model along with x'. Fit the model and comment on the relationship between SSE and R^2 in the standardized and unstandardized model.
(3 Marks)