ST102 Exercise 19

In this exercise you will practise least squares estimation and the simple linear regression model. Question 1 involves deriving least squares estimators of two parameters – this requires the same technique used to obtain the least squares estimators in the simple linear regression model. Question 2 concerns point estimation for the simple linear regression model. Question 3 requires you to derive least squares and maximum likelihood estimators for a regression model (which is still linear in the *parameter*). Finally, Question 4 involves estimating the best-fitting line and conducting a hypothesis test of the β_1 parameter.

Your answers to this problem set should be submitted as a pdf file upload to Moodle, *as directed by your class teacher*. It will be covered by your class teacher in your nineteenth class, which will take place in the week commencing Monday 20 March 2023.

1.* Suppose that you are given observations y_1 and y_2 such that:

$$y_1 = \alpha + \beta + \varepsilon_1$$

$$y_2 = -\alpha + \beta + \varepsilon_2.$$

The random variables ε_i, for $i = 1, 2$, are independent and normally distributed with mean 0 and variance σ^2.

(a) Find the least squares estimators of the parameters α and β, and verify that they are unbiased estimators.
Hint: obtain the minimum of the sum of the ε_i^2s using the least squares technique.

(b) Calculate the variance of the estimator of α.

2. An investigation, conducted by a mail-order company, into the relationship between the sales revenues (y_i, in millions of dollars) and the price per gallon of gasoline (x_i, in cents) over a period of 12 months yields:

$$\sum_{i=1}^{12} y_i = 632, \quad \sum_{i=1}^{12} x_i = 6{,}148, \quad \sum_{i=1}^{12} x_i^2 = 5{,}062{,}914 \quad \text{and} \quad \sum_{i=1}^{12} x_i y_i = 287{,}962.$$

Estimate the parameters β_0 and β_1 in the regression model $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, where the ε_is are uncorrelated with a mean of zero and a common variance of σ^2 for $i = 1, 2, \ldots, 12$. Interpret the estimated regression line.

3.* Suppose the observations $\{y_1, y_2, \ldots, y_n\}$ are described by the relationship:

$$y_i = \theta x_i^2 + \varepsilon_i$$

where $\{x_1, x_2, \ldots, x_n\}$ are fixed constants, and $\varepsilon_i \sim N(0, \sigma^2)$ for $i = 1, 2, \ldots, n$, such that the ε_is are independent and identically distributed.

(a) Find the least squares estimator of θ.

(b) Find the maximum likelihood estimator of θ.

4. The table below lists the USA social security costs for 7 specific years between 1965 and 1992.

Year	1965	1970	1975	1980	1985	1990	1992
$x =$ Year $-$ 1960	5	10	15	20	25	30	32
$y =$ social security cost (\$ billion)	17.1	29.6	63.6	117.1	186.4	246.5	285.1

(a) Plot the data using y against x (a hand-drawn graph is acceptable).

(b) Compute $\sum_i x_i$, $\sum_i y_i$, $\sum_i x_i^2$, $\sum_i y_i^2$ and $\sum_i x_i y_i$. Use these figures to fit the data with the simple linear regression model $y = \beta_0 + \beta_1 x + \varepsilon$.

(c) Test the hypothesis $\mathrm{H}_0 : \beta_1 = 0$ vs. $\mathrm{H}_1 : \beta_1 > 0$ at the 5% significance level. What can be concluded about social security costs from this test?

(d) Plot the residuals against x (a hand-drawn graph is acceptable). Are you happy with the fitted model? If not, discuss what you might try to achieve a better fit.