The University of British Columbian —Okanagan

DATA 310

Assignment 1

Due at 11:59 pm on Oct 2. Submit it on Canvas.

Question 1

(3 points) Write out a linear regression model which could be used to predict the prime interest rate P from the core inflation rate C. Make sure to include an error term and the assumptions about it.

Question 2

In the simple linear regression model $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ where $E[\varepsilon_i] = 0$ and $E[\varepsilon_i^2] = \sigma^2$ and where the ε's are independent, the ith fitted value is denoted by $\widehat{y}_i$ and the ith residual is denoted by e_i ($i = 1, 2, \ldots, n$). Show that

(a) (1 point) $\sum_{i=1}^n e_i = 0$

(b) (1 point) $\sum_{i=1}^n x_i e_i = 0$

(c) (1 point) $\sum_{i=1}^n \widehat{y}_i e_i = 0$

Question 3

Consider the regression model

$$y = \beta_0 + \beta_1 x + \varepsilon$$

where the ε_i are i.i.d. N(0,σ^2) random variables, for $i = 1, 2, \ldots, n$.

(a) (4 points) Show $\hat{\beta}_1$ is normally distributed with mean β_1 and variance $\frac{\sigma^2}{S_{XX}}$.

(b) (8 points) Let $SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2$. Show $E[SSE] = (n-2)\sigma^2$

Question 4

Suppose $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ constitute a sample of independent observations. Consider the model

$$y_i = \beta_1 x_i + \beta_2 i + \varepsilon_i$$

where the ε_i are i.i.d. N(0,σ^2) random variables, for $i = 1, 2, \ldots, n$.

(a) (5 points) Derive the least-squares estimators for β_1 and β_2. Under what condition on the predictor (x_i) are these estimators not well-defined?

(b) (2 points) For the case where the coefficient estimators are well-defined, write down an unbiased estimator for σ^2.

Question 5

(4 points) Suppose $y_1, y_2, \ldots, y_n$ constitute a sample of observations. Consider the model

$$y_i = \alpha + \varepsilon_i$$

where the ε_i are i.i.d. N(0,σ^2) random variables, for $i = 1, 2, \ldots, n$. Define

$$\text{MSE} = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2$$

Prove that MSE is an unbiased estimator for σ^2.

Question 6

The rate of spread R (in m/s) of a wildfire is related to wind speed W (in km/h). Suppose data have been collected on 5 fires:

```
R     W
30    35
32    40
18    20
35    50
12    15
```

(a) (1 point) Write out a linear regression model relating rate of spread to wind speed, including an error term. (Do not estimate the parameters β_0 and β_1.)

(b) (2 points) What assumptions are usually made about the error term?

(c) (2 points) What is the expected value of the rate of spread, given that the wind speed is x?

(d) (2 points) What is the variance of the rate of spread given that the wind speed is x?

(e) Suppose W is a random variable with mean μ and variance σ_W^2.

i. (2 points) Find the unconditional expected rate of spread. Contrast this answer with the answer to (c).

ii. (2 points) What is the unconditional variance of the rate of spread? Contrast this answer with the answer to (d).

(f) (2 points) How would you change the model if rate of spread is linearly related to the square root of wind speed?

Question 7

(4 points) Refer to the previous question. Mannually estimate the model parameters from the given data. Write down the fitted model and the estimate of σ^2

(a) (2 points) Calculate $\hat{\beta}_1$

(b) (2 points) Calculate $\hat{\beta}_0$

(c) (2 points) Calculate $\hat{\sigma^2}$

(d) (1 point) Using R language, make a data frame called `speedwind` including `R` and `W`.

(e) (1 point) Using `lm` function to fit the data and using `summary` function to read the result from `lm` function. Compare the result with your calculation.