

Home assignment 1

Deadline: 2024-09-16, 21:00.

1. Each day Elon Musk solves econometrics problems and creates posts in X. Let y_i be the number of solved problems and x_i be the number of posts in X. You have 3 observations: $x_1 = 2, y_1 = 5, x_2 = 1, y_2 = 10, x_3 = 3, y_3 = 4$.

- (a) Find $\hat{\beta}$ if fitted values are given by $y_i = \hat{\beta}x_i$.
- (b) Find $\hat{\beta}_0$ and $\hat{\beta}_1$ if fitted values are given by $y_i = \hat{\beta}_0 + \hat{\beta}_1x_i$.
- (c) Find $\hat{\beta}_0, \hat{\beta}_1$ and $\hat{\beta}_2$ if fitted values are given by $y_i = \hat{\beta}_0 + \hat{\beta}_1x_i + \hat{\beta}_2x_i^2$.

Note: you can use any programming language to calculate the 3×3 matrix inverse but you should provide the code :)

2. Simplify as much as possible the following expressions:

$$A = \sum_{i=1}^n (x_i - \bar{x})\bar{x}, \quad B = \sum_{i=1}^n (x_i - \bar{x})\bar{y}, \quad C = \sum_{i=1}^n (x_i - \bar{x})^2 + n\bar{x}^2.$$

3. Consider simple regression model with $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1x_i$. You have n observations $(x_1, y_1), \dots, (x_n, y_n)$ and you estimate $\hat{\beta}_0$ and $\hat{\beta}_1$ using OLS.

What will happen with $\hat{\beta}_0$ and $\hat{\beta}_1$ in each of the following cases?

- (a) You copy every observation from the original dataset twice.
- (b) You add one new observation $(y_{n+1} = \bar{y}, x_{n+1} = \bar{x})$ to the original dataset.
- (c) You add n more observations given by $(x_{n+i} = -x_i, y_{n+i} = y_i)$ for $i = 1, 2, \dots, n$ to the original dataset.

Hint: you may start by guessing the answer with an experiment, but the proof is required :)

Home assignment 2

Deadline: 2024-09-23, 21:00.

1. Each day Elon Musk solves econometrics problems and creates posts in X. Let y_i be the number of solved problems and x_i be the number of posts in X. You have 3 observations: $x_1 = 2, y_1 = 5, x_2 = 1, y_2 = 10, x_3 = 3, y_3 = 4$.

- (a) Calculate SST, SSE, SSR and R^2 if we regress y on x with constant, ie $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1x_i$.
- (b) Calculate SST, SSE, SSR and R^2 if we regress x on y with constant, ie $\hat{x}_i = \hat{\gamma}_0 + \hat{\gamma}_1y_i$.
- (c) Calculate the hat-matrix H if we regress y on x with constant.

Note: this exercises uses toy dataset from the previous HA, you may reuse old results provided that you state them explicitly.

2. Kamala Harris removes one observation from the initial set of n observations and reestimates the model $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ using OLS.

- (a) Prove that the total sum of squares SST can't increase.
- (b) Provide an example of a dataset where explained sum of squares SSE will decrease and a second example where it will increase.

3. Consider the dataset of diamond prices,

<https://github.com/vincentarelbundock/Rdatasets/raw/master/csv/ggplot2/diamonds.csv>.

Here `price` is the price of diamond in 1000\$ and `carat` is the weight of a diamond in carats. Let y_i be the log of diamond price in 1000\$ and x_i be the log of diamond weight in carats.

- (a) Estimate the model $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{\beta}_2 x_i^2$ using `LinearRegression` from `sklearn.linear_model`.
- (b) Estimate the model $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{\beta}_2 x_i^2$ using `ols` from `statsmodels.formula.api`.
- (c) What is your point forecast of a price of a diamond with 2 carats weight?

Note: the first approach is faster and more stable while the second one gives you much more statistical information.
