STATS 2107
Statistical Modelling and Inference II

Workshop 2: Bias, MSE, and BLUE

Matt Ryan

School of Mathematical Sciences, University of Adelaide

Semester 2 2022

Bias and MSE of Simple Linear Regression Estimates

The model

For data $(x_1, Y_1), (x_2, Y_2), \ldots, (x_n, Y_n)$, $x_i, Y_i \in \mathbb{R}$, consider the simple linear regression model

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \, ,$$

where $\varepsilon_i \sim N(0, \sigma^2)$

The model estimates

Recall that the estimates for β_0 and β_1 are given by

$$\hat{\beta}_1 = \frac{S_{XY}}{S_{XX}}$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x}$$

where

$$S_{XY} = \sum_{i=1}^{n} (x_i - \bar{x})(Y_i - \bar{Y})$$

$$S_{XX} = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

Question

- What is the bias and MSE of $\hat{\beta}_1$?

First note that $\hat{\beta}_1$ is a *linear estimator* of β_1.

$\hat{\beta}_1$ is linear

You can write:

$$\hat{\beta}_1 = \sum_{i=1}^{n} a_i Y_i \, ,$$

where $a_i = \dfrac{(x_i - \bar{x})}{S_{XX}}$.

PROOF:

Expected value and bias of $\hat{\beta}_1$

- $\mathrm{E}\left[\hat{\beta}_1\right] = \beta_1$
- Hence $b_{\hat{\beta}_1}(\beta_1) = 0$

i.e. $\hat{\beta}_1$ is an unbiased estimator of β_1.

PROOF:

The MSE of $\hat{\beta}_1$

Recall that:

$$\mathrm{MSE}_{\hat{\beta}_1}(\beta_1) = \mathrm{Var}(\hat{\beta}_1) + b_{\hat{\beta}_1}(\beta_1)^2 = \mathrm{Var}(\hat{\beta}_1)$$

so

$$\mathrm{MSE}_{\hat{\beta}_1}(\beta_1) = \frac{\sigma^2}{S_{XX}}$$

Your turn

What to do

1. Show that $\mathrm{Var}(\hat{\beta}_1) = \frac{\sigma^2}{S_{XX}}$.
2. Show that $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x}$ is a linear estimator, that is, You can write $\hat{\beta}_0 = \sum_{i=1}^{n} b_i Y_i$ for some constants b_i.
3. Derive the bias and MSE of $\hat{\beta}_0$.

BLUE

A Theorem

Consider the simple linear regression model

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \,,$$

where $\varepsilon_i \stackrel{iid}{\sim} N(0, \sigma^2)$. Then $\hat{\beta}_1$ is the BLUE for β_1.

What does this mean?

Recall what BLUE stands for:

- **B**est
- **L**inear
- **U**nbiased
- **E**stimator

What does this mean?

- We have already shown that $\hat{\beta}_1$ is a linear, unbiased estimator of β_1.
- By "Best", we mean that for ANY other linear unbiased estimator $\tilde{\beta}_1$ of β_1, we must have

$$\mathrm{Var}\left(\tilde{\beta}_1\right) \geq \mathrm{Var}\left(\hat{\beta}_1\right)$$

How do we show this:

We break the proof into 3 parts:

1. Use the fact that $\tilde{\beta}_1$ is linear and unbiased to derive some properties.
2. Add 0.
3. Show the cross term (covariance) is 0.

$\tilde{\beta}_1$ is unbiased

Let $\tilde{\beta}_1 = \sum_{i=1}^{n} c_i Y_i$. Then:

$$\sum_{i=1}^{n} c_i = 0\,,$$
$$\sum_{i=1}^{n} c_i x_i = 1\,.$$

PROOF:

Add 0

Let's look at the variance of $\tilde{\beta}_1$:

$$\begin{aligned}\mathrm{Var}\left(\tilde{\beta}_1\right) &= \mathrm{Var}\left(\tilde{\beta}_1 - \hat{\beta}_1 + \hat{\beta}_1\right) \\ &= \mathrm{Var}\left(\hat{\beta}_1\right) + \mathrm{Var}\left(\tilde{\beta}_1 - \hat{\beta}_1\right) + 2\mathrm{cov}\left(\tilde{\beta}_1 - \hat{\beta}_1, \hat{\beta}_1\right) .\end{aligned}$$

The cross term is 0

We can show that

$$\operatorname{cov}\left(\tilde{\beta}_1 - \hat{\beta}_1, \hat{\beta}_1\right) = 0\,.$$

PROOF:

Putting it all together

Using these results, we have that

$$\mathrm{Var}\left(\tilde{\beta}_1\right) = \mathrm{Var}\left(\hat{\beta}_1\right) + \mathrm{Var}\left(\tilde{\beta}_1 - \hat{\beta}_1\right) \geq \mathrm{Var}\left(\hat{\beta}_1\right) .$$

Your turn

What to do

1. Show that $\hat{\beta}_0$ is the BLUE for β_0.