8 Lecture 8: Feb 5

Last time

- Properties of the LS estimators

Today

- Inference of SLR model
- Lab 1

Statistical inference of the SLR model

Now we have the distribution of $\hat{\beta}_0$ and $\hat{\beta}_1$

$$\begin{aligned}\hat{\beta}_0 &\sim N(\beta_0, \frac{\sigma_\epsilon^2 \sum x_i^2}{n\sum(x_i-\bar{x})^2})\\ \hat{\beta}_1 &\sim N(\beta_1, \frac{\sigma_\epsilon^2}{\sum(x_i-\bar{x})^2}).\end{aligned}$$

However, σ_ϵ is never known in practice. Instead, an *unbiased* estimator of σ_ϵ^2 is given by

$$\hat{\sigma_\epsilon}^2 = MS[E] = \frac{SS[E]}{n-2}.$$

Proof:

Confidence intervals

Now we substitute $\hat{\sigma}_\epsilon^2$ into the distribution of $\hat{\beta}_0$ and $\hat{\beta}_1$

$$\begin{aligned}\hat{\beta}_1 &\sim N(\beta_1, \frac{\sigma_\epsilon^2}{\sum(x_i-\bar{x})^2})\\ \hat{\beta}_0 &\sim N(\beta_0, \frac{\sigma_\epsilon^2 \sum x_i^2}{n\sum(x_i-\bar{x})^2})\end{aligned}$$

to get the estimated standard errors:

$$\begin{aligned}\widehat{SE}(\hat{\beta}_1) &= \sqrt{\frac{MS[E]}{\sum(x_i-\bar{x})^2}}\\ \widehat{SE}(\hat{\beta}_0) &= \sqrt{MS[E]\left(\frac{1}{n}+\frac{\bar{x}^2}{\sum(x_i-\bar{x})^2}\right)}\end{aligned}$$

And the $100(1-\alpha)\%$ confidence intervals for β_1 and β_0 are given by

$$\hat{\beta}_1 \pm t(n-2, \alpha/2)\sqrt{\frac{MS[E]}{S_{xx}}}$$

$$\hat{\beta}_0 \pm t(n-2, \alpha/2)\sqrt{MS[E]\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)}$$

where $S_{xx} = \sum(x_i - \bar{x})^2$

Confidence interval for $\mathbf{E}(Y|X = x_0)$

The conditional mean $\mathbf{E}(Y|X = x_0)$ can be estimated by evaluating the regression function $\mu(x_0)$ at the estimates $\hat{\beta}_0$, $\hat{\beta}_1$. The conditional variance of the expression isn't too difficult (already shown):

$$\mathrm{Var}(\hat{\beta}_0 + \hat{\beta}_1 \mathrm{x}_0 | \mathrm{X} = \mathrm{x}_0) = \sigma^2(\frac{1}{\mathrm{n}} + \frac{(\mathrm{x}_0 - \bar{\mathrm{x}})^2}{\mathrm{S}_{\mathrm{xx}}})$$

This leads to a confidence interval of the form

$$\hat{\beta}_0 + \hat{\beta}_1 x_0 \pm t(n-2, \alpha/2)\sqrt{MS[E]\left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right)}$$

Prediction interval

Often, prediction of the response variable Y for a given value, say x_0, of the independent variable of interest. In order to make statements about future values of Y, we need to take into account

- the sampling distribution of $\hat{\beta}_0$ and $\hat{\beta}_1$
- the randomness of a future value Y.

We have seen the predicted value of Y based on the linear regression is given by $\hat{Y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$.

The 95% prediction interval has the form

$$\hat{Y}_0 \pm t(n-2, \alpha/2)\sqrt{MS[E]\left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right)}.$$

Hypothesis test

To test the hypothesis $\boxed{H_0 : \beta_1 = \beta_{slope_0}}$ that the population slope is equal to a specific value β_{slope_0} (most commonly, the null hypothesis has $\beta_{slope_0} = 0$), we calculate the test statistic (T-statistics) with $df = n - 2$

$$t_0 = \frac{\hat{\beta}_1 - \beta_{slope_0}}{\widehat{SE}(\hat{\beta}_1)} \sim t_{n-2}$$