Lesson 14. Confidence and Prediction Intervals for SLR Response – Part 1

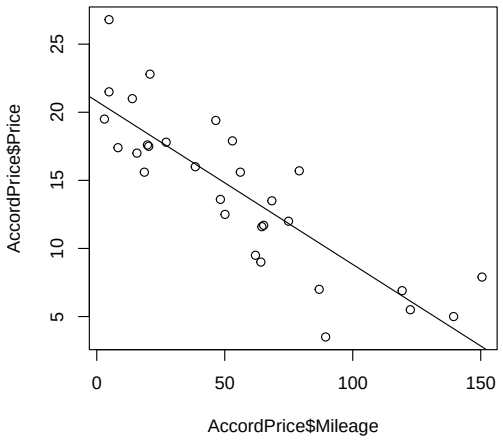
Note. In Part 2 of this lesson, you can run the R code that generates the outputs here in Part 1.

1 Overview

- Recall the simple linear regression model:

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad \text{where} \quad \varepsilon \sim \text{iid } N(0, \sigma_\varepsilon^2)$$

- There are two different things we might want to predict with this model:
  - The mean response for a particular predictor value
  - A future individual response for a particular predictor value



2 Confidence interval for simple linear regression response

- To estimate the mean of Y when X has the value x^*, we use a $100(1-\alpha)$**% confidence interval for** μ_Y:

$$\text{where } SE_{\hat{\mu}} = \hat{\sigma}_\varepsilon \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

- Interpretation:

  We are $100(1-\alpha)$% confident that the average response for all observational units with predictor value x^* is between lower endpoint of CI and upper endpoint of CI units.

  - This means that, with repeated construction and use, the procedure for forming a CI will capture the true μ_Y for the predictor value x^* $100(1-\alpha)$% of the time

Example 1. Let's look once again at the `AccordPrice` data.

First, we run the regression, with *Price* as the response variable and *Mileage* as the predictor:

```
fit <- lm(Price ~ Mileage, data = AccordPrice)
```

Using R, we can construct a 95% confidence interval for the average price of a used Accord with 50,000 miles on it, like this:

```
predict(fit, newdata = data.frame(Mileage = 50), interval="confidence", level=0.95)
```

The output looks like this:

A matrix: 1 × 3 of type dbl

	fit	lwr	upr
1	14.81902	13.65787	15.98016

3 Prediction interval for simple linear regression response

- To estimate a future individual response y when X has the value x^*, we use a $100(1-\alpha)$**% prediction interval for** $\hat{y}$

$$\text{where } SE_{\hat{y}} = \hat{\sigma}_\varepsilon \sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

- Interpretation:

  We are $100(1-\alpha)\%$ confident that the response of a particular observational unit with predictor value x^* is between lower endpoint of CI and upper endpoint of CI units.

  - This means that, with repeated construction and use, the procedure for forming a PI will capture the actual y for the predictor value x^* $100(1-\alpha)\%$ of the time

Example 2. Continuing with the `AccordPrice` data...

Using R, we can construct a 95% prediction interval for the price of an individual used Accord with 50,000 miles on it, like this:

```
predict(fit, newdata = data.frame(Mileage = 50), interval="prediction", level=0.95)
```

The output looks like this:

A matrix: 1 × 3 of type dbl

	fit	lwr	upr
1	14.81902	8.393807	21.24422