

Inference for Prediction in SLR

Warm-up questions

Is education level associated with income? Researchers collected education level (in years) and income (in thousands of dollars) for a random sample of 32 employees working for the city of Riverview.

The researchers fit a simple linear regression of income on education and obtained the following output:

term	estimate	std.error	statistic	p.value
(Intercept)	11.321	6.123	1.849	0.074
education	2.651	0.370	7.173	<0.001

Q1. Write the fitted regression equation.

Q2. Interpret the slope coefficient in the context of the problem. (Don't forget to specify units.)

Q3. Interpret the intercept in the context of the problem. (Don't forget to specify units.)

A new researcher joined the team and decided that education should be standardized (to have mean 0 and SD 1) before fitting the regression model. The output from this regression is shown below:

term	estimate	std.error	statistic	p.value
(Intercept)	53.742	1.587	33.861	<0.001
scale(education)	11.567	1.613	7.173	<0.001

Q4. Write the fitted regression equation for this new model.

Q5. Interpret the slope coefficient in the context of the problem. (Don't forget to specify units.)

Q6. Interpret the intercept in the context of the problem. (Don't forget to specify units.)

Q7. Compare the two models. What's the same? What's different?

Prediction and confidence intervals

Tip

The R code for the following questions is found at
<https://aloy.github.io/stat230-materials/activity/04-slr-prediction>.
The URL is also posted on Moodle.

For this activity you will consider predicting the price of a used car (it's Kelly Blue Book value) based on its mileage. The columns of interest in the **Cars** data set are **Price** and **Mileage**.

Q1. Use the `lm()` command to fit the simple linear regression model where Mileage is used to predict Price. Report the fitted regression equation.

Q2. The first car in the data set is a Buick Century with 8221 miles. Calculate the expected price of this car using the fitted regression equation.

Q3. If we want to predict the price of **this** car, should we use a confidence interval or a prediction interval?

Q4. Use R to construct the appropriate 89% interval for the price of this car. Record this interval below.

Q5. Interpret the interval in context.

Q6. Run the code to produce a scatterplot, regression line, and both types of intervals. Which is the prediction interval and which is the confidence interval? How can you tell?