Lecture 16: Line of Best Fit

For each question below, write the R code you would use and record the result when appropriate. Answer in complete sentences when asked to explain.

Part 1: Understanding the Variables

We will use the built-in `cars` dataset in R. This dataset contains information about the speed of cars and the distance it took them to stop.

```
data(cars)
head(cars)
str(cars)
```

1. Look at the `cars` dataset.

   (a) What does one row of the dataset represent?

   (b) What are the two variables in the dataset?

   (c) Which variable should be the independent variable? Explain briefly.

   (d) Which variable should be the dependent variable? Explain briefly.

   (e) How do you think the two variables are related to each other?

Part 2: Creating a Scatterplot

2. Create a scatterplot with `speed` on the x-axis and `dist` on the y-axis.

   (a) Describe the relationship between speed and stopping distance.

   (b) Does the relationship look positive, negative, or neither? Explain briefly.

   (c) Does a straight line seem like a reasonable way to summarize this relationship? Explain briefly.

Part 3: Creating the Linear Model

3. Use `lm()` to create a line of best fit.

```
linear_model <- lm(cars$dist ~ cars$speed)
linear_model
```

   (a) Record the intercept.

   (b) Record the slope.

   (c) Write the equation of the line of best fit using the output from R.

4. In this model, the equation has the form:

$$\text{predicted stopping distance} = \text{slope} \times \text{speed} + \text{intercept}$$

   (a) In context, interpret the slope.

   (b) In context, what would the intercept represent?

   (c) Does the intercept make sense to interpret literally in this problem? Explain briefly.

Part 4: Adding the Line of Best Fit

5. Add the line of best fit to your scatterplot using the `abline()` function.

   (a) Does the line appear to fit the data reasonably well? Explain briefly.

   (b) Are all of the points exactly on the line? Should we expect them to be? Explain briefly.

(c) What does it mean for the line to be a "best fit" line?

Part 5: Making Predictions

6. Use the model to predict the stopping distance for a car going 20 mph.

   (a) Record the predicted stopping distance.

   (b) Interpret this prediction in context.

Part 6: Extrapolation

7. Look at the range of speeds in the original dataset. What is the smallest and largest speed in the dataset?

8. Suppose someone wants to use this model to predict the stopping distance for a car going 60 mph.

   (a) Would this be interpolation or extrapolation?

   (b) Why should we be cautious about making this prediction?