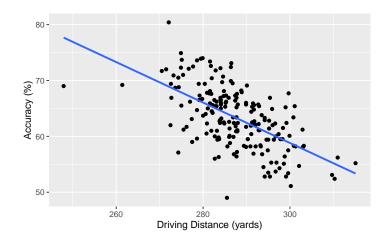# Homework 9: Golf Driving Distances
Simple Linear Regression

In golf the goal is to complete a hole with as few strokes as possible. A long driving distance to start a hole can help minimize the strokes necessary to complete the hole, as long as that drive stays on the fairway. Data were collecting on 197 PGA and LGPA male players in 2008. For each player, the average driving distance (yards), fairway accuracy (percentage), and sex was measured.

```
golf_data <- read_csv("data/golf.csv") |>
  filter(Sex == "male")
head(golf_data)
```

```
# A tibble: 6 x 3
  Driving_Distance Percent_Accuracy Sex
             <dbl>            <dbl> <chr>
1             290.             59.5 male
2             302.             54.7 male
3             287.             62.4 male
4             283.             65.4 male
5             299.             52.8 male
6             300.             51.1 male
```

Research Question: Does the accuracy of a professional male golfer change when they hit the ball farther?

```
ggplot(data = golf_data,
       mapping = aes(x = Driving_Distance,
                     y = Percent_Accuracy)
       ) +
  geom_point() +
  geom_smooth(method = "lm", se = F) +
  labs(x = "Driving Distance (yards)",
```

```
y = "Accuracy (%)")
```


1. Identify the following in context of the problem:

- Observation:

- Explanatory Variable (and data type):

- Response Variable (and data type):

- Population of Interest:

- Sample (and Sample Size):

2. Based on the scatterplot above, does it appear that there is a relationship between driving distance and percent accuracy? Explain.

3. Set up the null and alternative hypotheses for testing for an association using the slope:

- In words.

- In symbols

4. Fill in the code below as if you were running a linear regression model.

```
golf_model <- lm(_____ ~ _____,
                 data = golf_data)

golf_model |>
  tidy(conf.int = TRUE,
       conf.level = TRUE)
```

```
# A tibble: 2 x 7
  term            estimate std.error statistic p.value conf.low conf.high
  <chr>              <dbl>     <dbl>     <dbl> <chr>      <dbl>     <dbl>
1 (Intercept)       167.       10.1      16.6 <0.001     148.      187.
2 Driving_Distance   -0.362     0.035   -10.4 <0.001      -0.431    -0.293
```

5. Check the LINE (Linearity, Independent Observations, Normality, and Equal Variance) conditions for using simple linear regression on our data:
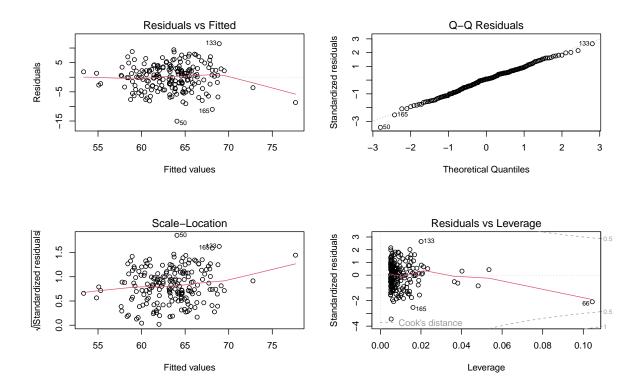
```
par(mfrow = c(2,2))
plot(golf_model)
```

```r
par(mfrow = c(1,1))
```

6. Using the output from the code above, write the equation of the regression line in the context of the problem.

7. Interpret the estimated slope in context of the problem.

8. State your conclusion in context of the research question. Cite evidence to support your conclusion.

9. Interpret the 95% confidence interval for the slope.

10. Suppose you have a golfer with a driving distance of 280 yards. What would you expect their accuracy to be?

11. Assume that same golfer has an actual accuracy of 71%, what is the residual for this golfer?

12. Instead, assume that same golfer has an actual accuracy of 55%, what is the residual for this golfer?

> 🔥 Canvas Quiz (Optional)
>
> Make sure to complete the Homework Quiz on Canvas.