

MA206, Lesson 20 - Multiple Linear Regression

Review: What is the form of our linear regression model with 1 explanatory variable?

If our model is valid, how do we expect our residuals to be distributed?

What are the validity conditions for our regression line?

How would we interpret the affects of variable X_1 on y , given the equation $\hat{y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2$?

1) Use the [ACFT2.csv](#) dataset to examine more inferences we can make with regards to IOCT time.

a) We want to assess if the Sprint Drag Carry may have an association with IOCT time. Generate a scatterplot between *SDC_Raw* and *IOCT_time* and comment on the Form, Direction, Strength, and any unusual observations.

b) Report the correlation between the Sprint Drag Carry and IOCT time. Does this coincide with your previous assessment?

c) Generate a linear model using the Sprint Drag Carry as your explanatory variable and the IOCT time as the response. Write down the resulting equation. How do you interpret the slope?

d) Report the R^2 and explain what it means, in words.

e) Validate the validity conditions and report on your findings.

2) Now we want to assess if the Maximum Dead Lift is a potential confounding variable.

a) Generate two scatterplots comparing *MDL_Raw* with *SDC_Raw* as well as comparing *MDL_Raw* with *IOCT_time*. Comment on the Form, Direction, and Strength for both plots.

b) Based on the previous graphs, might Maximum Deadlift be a confounder? Explain.

c) Report the correlation between the three pairs of variables. Do these results coincide with your previous assessments of Form, Direction, and Strength?

d) Generate a linear model using Sprint Drag Carry and Maximum Deadlift to explain IOCT Time. Write the resulting equation and interpret the slope for Sprint Drag Carry.

e) Did the coefficient for Sprint Drag Carry change between your models? Discuss what that means.

f) Report the R^2 and explain what it means.

g) What are the p-values associated with each coefficient (β_1 and β_2) and what do they mean?

h) Validate the validity conditions and report on your findings.

3) We suspect the categorical variable *Sex* is confounding, so we want to include that in our analysis as well.

a) Generate a linear model using Sprint Drag Carry, Maximum Deadlift, and Sex to explain IOCT Time.

Write the resulting equation and interpret the slope for Sprint Drag Carry.

b) List the p-values for each of our three variable coefficients ($\beta_1, \beta_2, \beta_3$) and discuss the implications.

4) The dataset found below lists different sets from Lego along with the price, number of reviews, number of pieces, and if it is licensed from a Disney, Marvel, DC, or Star Wars franchise. We want to determine if the number of pieces can accurately predict the price of Lego sets.

```
Legos <- read_csv("https://raw.githubusercontent.com/rslasater82/MA206Datasets/main/legos.csv")
```

a) What is the correlation between pairs for *pieces*, *reviews*, and *price*?

b) Create a linear regression model using pieces and reviews to calculate price. Report your equation.

c) Create a new model including Licence as a variable and report your equation.

d) Based on the above equations, would you consider *Licensed* a confounding variable? Explain why.