Cheat Sheet: Multiple Linear Regression

Essential Concepts

- A linear regression model with two or more explanatory variables is called a multiple linear regression model. Since there is more than one explanatory variable, the model is no longer a line. In fact, we can include p explanatory variables in our model. The equation for the estimated model that uses p variables is $\hat{y} = a + b_1 \cdot x_1 + b_2 \cdot x_2 + ... + b_p \cdot x_p$where $b_1, b_2, ..., b_p$ are the regression coefficients for explanatory variables $x_1, x_2, ..., x_p$, respectively. In multiple linear regression, $b_1, b_2, ..., b_p$ are called partial slopes.
- The coefficient of determination, R^2, is used to determine the percentage of variability in the response variable that is accounted for by the explanatory variables.
- In multiple linear regression, the y-axis has the residual values and the x-axis has the explanatory variables and/or the fitted values. For a multiple linear regression model, you create a residual plot for each continuous explanatory variable, as well as the fitted value.
- An indicator variable is a binary variable with only two values: 0 and 1. When creating an indicator variable, we assign the value of 1 for a certain category, and the value of 0 is used for all other categories.
- A reference group is the value of the categorical variable that is not represented explicitly by the indicator variable (which is why we only require $k - 1$ indicator variables to define our regression model).
- An interaction occurs when an explanatory variable has a different effect on the response variable, depending on the values of another explanatory variable. An interaction term is a variable that represents an interaction between two variables.

Key Equations

multiple linear regression model

$$\hat{y} = a + b_1 \cdot x_1 + b_2 \cdot x_2 + ... + b_p \cdot x_p$$

partial slopes

$$b_1, b_2, ..., b_p$$

Glossary

indicator variable

a binary variable with only two values: 0 and 1

interaction

an explanatory variable that has a different effect on the response variable, depending on the values of another explanatory variable

interaction term

a variable that represents an interaction between two variables

multiple linear regression model

a linear regression model with two or more explanatory variables