Week 4 Reading Guide: Basic Regression

# [Modern Dive Chapter 5](https://moderndive.com/5-regression.html)

In Statistics we like to call our variable the “response” variable. What are some other names can be called?

In Statistics we like to call our variable the “explanatory” variable. What are some other names for ?

What is the difference between “explanatory modeling” and “predictive modeling”?

In “basic” / simple linear regression, how many explanatory variables are included in the model?

In linear regression, what type of variable must we **always** have as our response?

## 5.1 – One Numerical Explanatory Variable

### 5.1.1 – Exploratory data analysis

What is the response variable for the evaluation study? What is the explanatory variable?

What is “EDA”? Why should you do EDA *before* you analyze your data?

In the evals\_ch5 dataset, can an instructor appear more than once?

What statistics does the skim() function output?

What does a *correlation coefficient* measure? Specifically, what type of relationship?

For correlation coefficients:

* values near -1 indicate:
* values near +1 indicate:
* values near 0 indicate:

**Quick! – Which correlation is larger, -0.75 or +0.45?**

Why did the authors jitter the points in the scatterplot?

What function adds a regression line to a scatterplot?

What method do you need to specify to get a straight line?

Based on the scatterplot, how would you describe the relationship between instructor’s beauty score and their evaluation score?

### 5.1.2 – Simple Linear Regression

In your math class you’ve seen the equation of a line presented as . How does this equation change in a statistics class?

* becomes:
* becomes:
* becomes:
* becomes:

What does the hat on top of represent?

The intercept () is the value of when is:



Image by Allison Horst

The slope () is the expected change in when is increased by:

What function do we use (in R) to “fit” a linear regression?

What function do we use (in R) to get the table of regression estimates?

Where can you find the value of the intercept on the regression table?

Where can you find the value of the slope on the regression table?

### 5.1.3 – Observed / Fitted Values and Residuals



Image by Allison Horst

How do you calculate the *fitted value* of an observation?

How do you calculate the *residual* of an observation?

What function (in R) can you use to obtain **every** residual for a regression model?

## 5.2 – One Categorical Explanatory Variable

### 5.2.1. – Exploratory Data Analysis

What are the “levels” of the continent variable?

How would you describe the distribution of life expectancy?

Which continent has the largest variation in life expectancy?

### 5.2.2 – Linear Regression

What is an *offset*?

How does R choose what group is the “baseline”?

When the explanatory variable is categorical the numbers in the regression table associated with each group is the:

What is an “indicator function”?

The estimated regression equation for the life expectancy with continent as an explanatory variable is:

When finding the mean life expectancy for Europe, what value would you plug in for , , and ?

For a categorical explanatory variable with levels, how many offsets will you have?

### 5.2.3 – Observed / Fitted Values and Residuals

True or False – For a linear regression with one categorical explanatory variable, **every** observation associated with the same group (e.g., every country in Asia) will have the **same** fitted value

## 5.3 – Related Topics

### 5.3.1. – Correlation is not necessarily causation

Most people have heard the saying “correlation does not imply causation,” but *sometimes* correlation **does** imply causation. What does a study need to have in order for cause and effect statements to be used?

When this (what you stated above) is not present in a study, you have to worry about *confounding* / *lurking* variables. What is a *confounding* variable?