SIPPS Coding Workshop: basic track

06: Introduction to linear models

Hannah & Camille

July 13, 2021

Plan for today

- what is a linear regression model? (quick review)
- prepping your data for analysis
- using the `lm()` function & interpreting the output

What is a linear regression model?

▶ used to quantify the relationship(s) between an **outcome variable** and one (or more!) **predictors**

What is a linear regression model?

▶ used to quantify the relationship(s) between an **outcome variable** and one (or more!) **predictors**

▶ what this analysis does, more specifically, is **fit a line** that best minimizes the error between that line and your data points

What is a linear regression model?

▶ used to quantify the relationship(s) between an **outcome variable** and one (or more!) **predictors**

▶ what this analysis does, more specifically, is **fit a line** that best minimizes the error between that line and your data points

▶ when we use R to fit this line, what we get as output of our analysis is an **intercept** and at least one **slope**

What is a linear regression model?

▶ used to quantify the relationship(s) between an **outcome variable** and one (or more!) **predictors**

▶ what this analysis does, more specifically, is **fit a line** that best minimizes the error between that line and your data points

▶ when we use R to fit this line, what we get as output of our analysis is an **intercept** and at least one **slope**

▶ usually, these **slope** values (also called **betas**) are what we're most interested in: if a slope associated with a given X variable is significantly different than zero, we can conclude that the value of X is meaningfully related to the value of Y
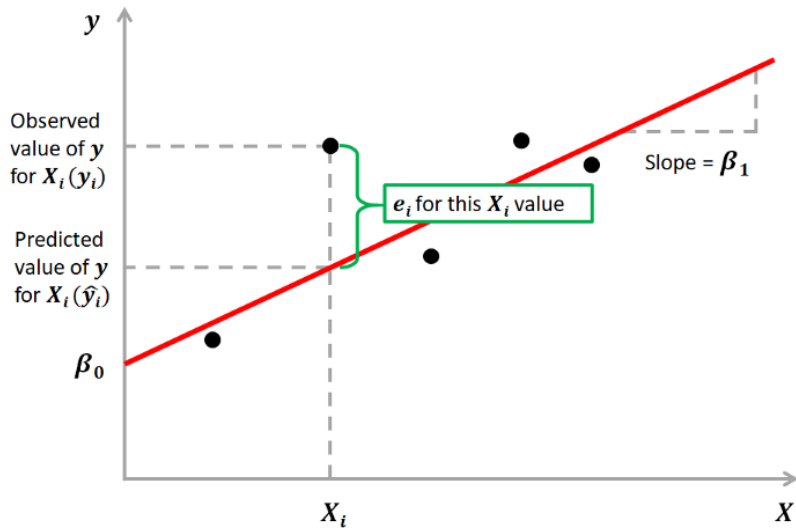
What is a linear regression model?

Figure 1: Illustration of linear regression

What is a linear regression model?

▶ for example, let's say you are interested in which variables affect how much you'll enjoy a particular take-out restaurant

What is a linear regression model?

- ▶ for example, let's say you are interested in which variables affect how much you'll enjoy a particular take-out restaurant

- ▶ you could hypothesize that your enjoyment will depend on (at least) two things:

What is a linear regression model?

▶ for example, let's say you are interested in which variables affect how much you'll enjoy a particular take-out restaurant

▶ you could hypothesize that your enjoyment will depend on (at least) two things:
  ▶ the average Yelp rating

What is a linear regression model?

- ▶ for example, let's say you are interested in which variables affect how much you'll enjoy a particular take-out restaurant

- ▶ you could hypothesize that your enjoyment will depend on (at least) two things:
  - ▶ the average Yelp rating
  - ▶ whether the restaurant is a pizzeria or a taqueria

What is a linear regression model?

- ▶ for example, let's say you are interested in which variables affect how much you'll enjoy a particular take-out restaurant

- ▶ you could hypothesize that your enjoyment will depend on (at least) two things:
  - ▶ the average Yelp rating
  - ▶ whether the restaurant is a pizzeria or a taqueria

- ▶ `enjoyment ~ (yelp rating) + (type of restaurant)`

What is a linear regression model?

- for example, let's say you are interested in which variables affect how much you'll enjoy a particular take-out restaurant

- you could hypothesize that your enjoyment will depend on (at least) two things:
  - the average Yelp rating
  - whether the restaurant is a pizzeria or a taqueria

- `enjoyment ~ (yelp rating) + (type of restaurant)`

- if both slopes are significant, we can say that both the yelp rating and the type of restaurant are significantly associated with food enjoyment

Types of variables

- regression models are inherently flexible, and allow you to quantify many different kinds of variables & relationships

Types of variables

- regression models are inherently flexible, and allow you to quantify many different kinds of variables & relationships

- in our toy model, for example, we are looking at two different types of moderators/predictors: **continuous** (the yelp rating) & **categorical** (the type of restaurant)

Types of variables

▶ example results:

categorical

continuous
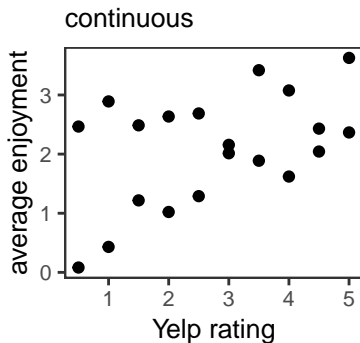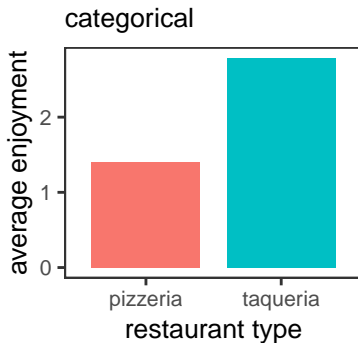
Introducing the lm() function

- lm() is a function that allows you to run linear models

Introducing the `lm()` function

- `lm()` is a function that allows you to run linear models

- using it requires two main arguments: (1) the dataframe you want to work with, and (2) the equation of the model you want to run

Introducing the `lm()` function

- ▶ `lm()` is a function that allows you to run linear models

- ▶ using it requires two main arguments: (1) the dataframe you want to work with, and (2) the equation of the model you want to run

- ▶ equations follow the format: `Y ~ X1 [+ X2 + X3 + ...]`

Introducing the `lm()` function

▶ `lm()` is a function that allows you to run linear models

▶ using it requires two main arguments: (1) the dataframe you want to work with, and (2) the equation of the model you want to run

▶ equations follow the format: `Y ~ X1 [+ X2 + X3 + ...]`

▶ so we might run: `lm(data = mydata, enjoyment ~ yelp_rating + restaurant_type)`

Plan for today

- ~~what is a linear regression model?~~ any questions?
- prepping your data for analysis
- using the `lm()` package & interpreting the output
- visualizing your results