3.1 Linear Equations

Dr. Lauren Perry

Goals

1. Review linear equations.
2. Motivate regression.
3. Interpret a slope and intercept in context.
4. Use a regression line to predict values of a dependent variable.

Linear Equations

- Should already have seen linear equations like

$$y = mx + b$$

- In statistics, we write these as

$$y = b_0 + b_1 x$$

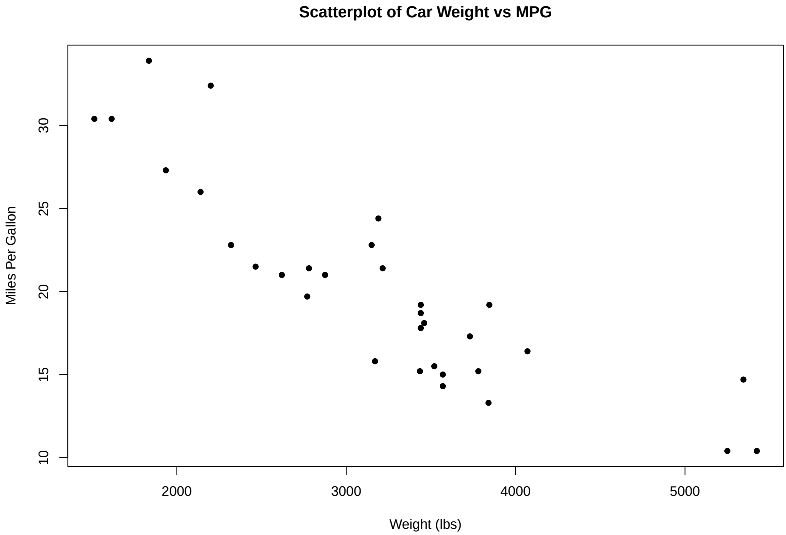
  - b_0 and b_1 are constants.
  - x is the **independent variable**.
  - y is the **dependent variable**

Slope and Intercept

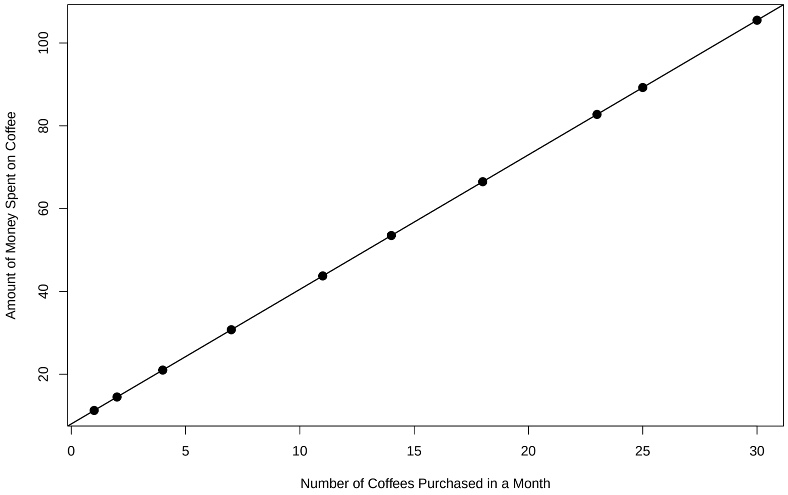
$$y = b_0 + b_1 x$$

- The **y-intercept** is b_0, the value of y when $x = 0$.
- The **slope** is b_1, the change in y for a 1-unit change in x.

A **scatterplot** shows the relationship between two (numeric) variables.



We call this type of data **bivariate data**.



This relationship can be modeled perfectly with a straight line:

$$y = 8 + 3.25x$$

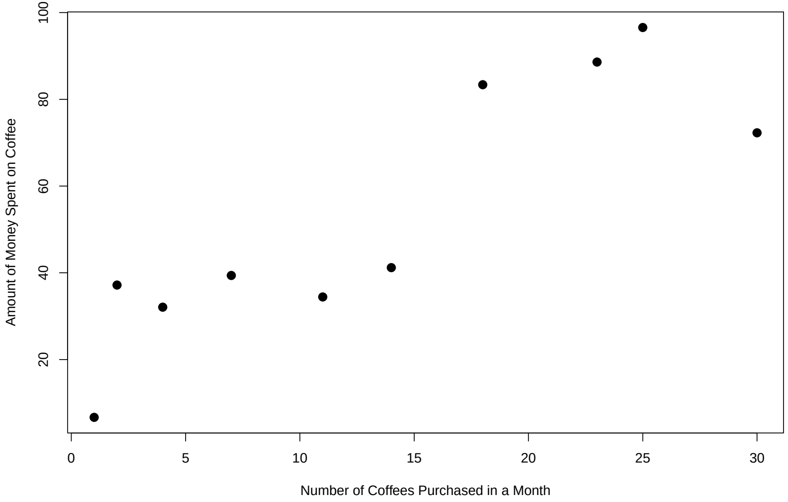
Example: Interpret Slope and Intercept

$$y = 8 + 3.25x$$

where x is the number of coffees purchased in a month and y is the amount of money spent on coffee.

- The *intercept* is 8, which is the dollar amount of money spent on coffee (the value of y) when 0 coffees are purchased in that month (when $x = 0$).
- The *slope* is 3.25, which is the increase in amount of money spent on coffee (increase in y) for each additional coffee purchased (a one-unit increase in x).

- But what if that pound of coffee didn't always cost \$8?
- Or the coffee drinks didn't always cost \$3.25?



The linear regression line looks like

$$y = \beta_0 + \beta_1 x + \epsilon$$

- β is the Greek letter "beta".
- β_0 and β_1 are constants.
- Error (the fact that the points don't all line up perfectly) is represented by ϵ.

We estimate β_0 and β_1 using data and denote the estimated line by

$$\hat{y} = b_0 + b_1 x$$

- $\hat{y}$, "y-hat", is the estimated value of y.
- b_0 is the estimate for β_0.
- b_1 is the estimate for β_1.
- ... and 0 is the estimate for ϵ (so we ignore it).

We use a regression line to make predictions about y using values of x.

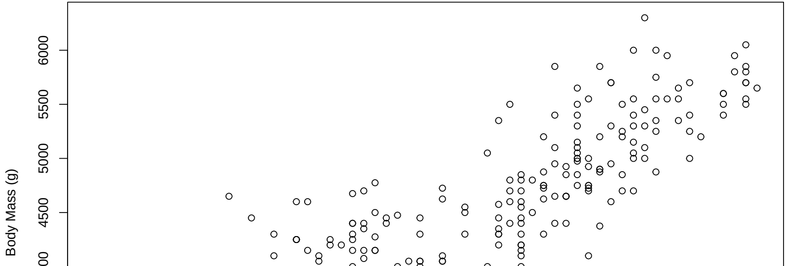
Think of this as the 2-dimensional version of a point estimate!

- y is the **response variable**.
- x is the **predictor variable**.

Example

Example: Researchers took a variety of measurements on 344 adult penguins in Antarctica.

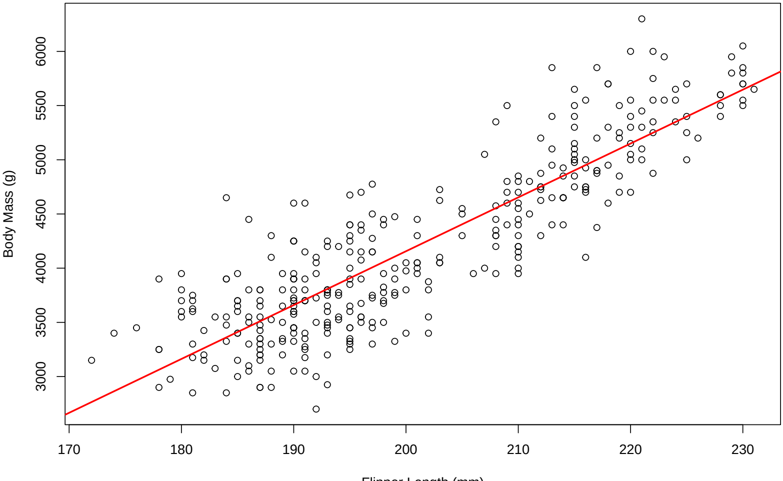
```
##
## Attaching package: 'palmerpenguins'

## The following objects are masked from 'package:datasets
##
##     penguins, penguins_raw
```



Example

The regression model for these data is

$$\hat{y} = -5780.83 + 49.69x$$



Example

To predict the body mass for a penguin with a flipper length of 180mm, we just need to plug in 180 for flipper length (x):

$$\hat{y} = -5780.83 + 49.69 \times 180 = 3163.37\text{g}.$$

- Note that the regression line automatically deals with units.