

Multiple Linear Regression

Grinnell College

September 25, 2024

Practice

We laid out sandwiches and counted the number of ants that appeared after 15 minutes. Our sandwiches had either multigrain, white, or rye bread, and our filling was either peanut butter or ham and pickles

```
1 > lm(Ants ~ Bread + Filling, sandwich)
2
3 Coefficients:
4 (Intercept)      BreadRye      BreadWhite      FillingPeanut_Butter
5      53.250         3.833         0.917         -20.167
```

- ▶ What is the formula for this regression line?
- ▶ What is my reference variable?
- ▶ Interpret intercept?
- ▶ How many ants would we predict to be on a peanut butter sandwich on white bread?

Practice

This dataset gives the speed of cars at the time of braking (mpg) and the distance it continued to travel (ft) before coming to a complete stop

```
1 > lm(dist ~ speed, cars)
2
3 Coefficients:
4 (Intercept)      speed
5      -17.58         3.93
```

- ▶ What is the formula for this regression line?
- ▶ Interpret intercept?
- ▶ How many feet would a car continue to travel if it were going 10mph?

Simple linear regression (Quantitative):

- ▶ $y = \beta_0 + \beta_1 X$
- ▶ Slope and intercept interpretation

Simple linear regression (Categorical):

- ▶ $y = \beta_0 + \mathbb{1}_B \beta_1 + \mathbb{1}_C \beta_2$
- ▶ Indicator variables
- ▶ Reference variables
- ▶ Coefficients represent group differences from reference var

Categorical and Quantitative Predictor

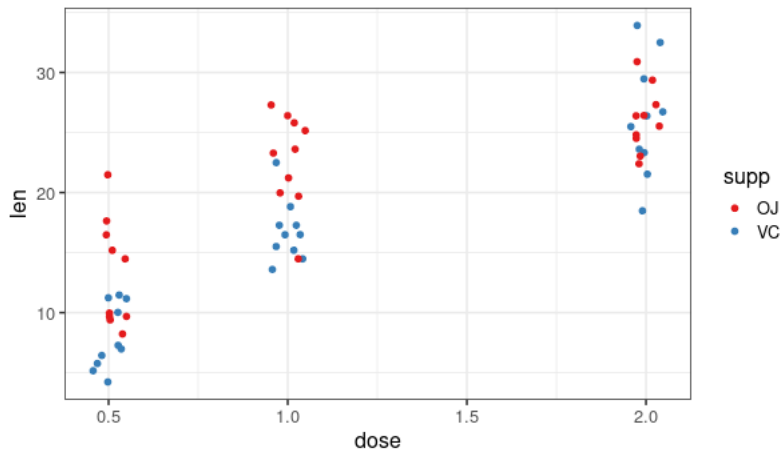
Often times where we have multiple predictors in linear regression model

Consider an example where we investigate odontoblasts measured in 60 guinea pigs, each receiving three doses of vitamin C a day (0.5, 1, 2mg/day) by one of two methods (orange juice (OJ) or ascorbic acid (VC))

We are interested in finding out two things:

1. Is more vitamin C associated with greater odontoblast growth?
2. Does the ROA of vitamin C influence odontoblast growth?

Categorical and Quantitative Predictors



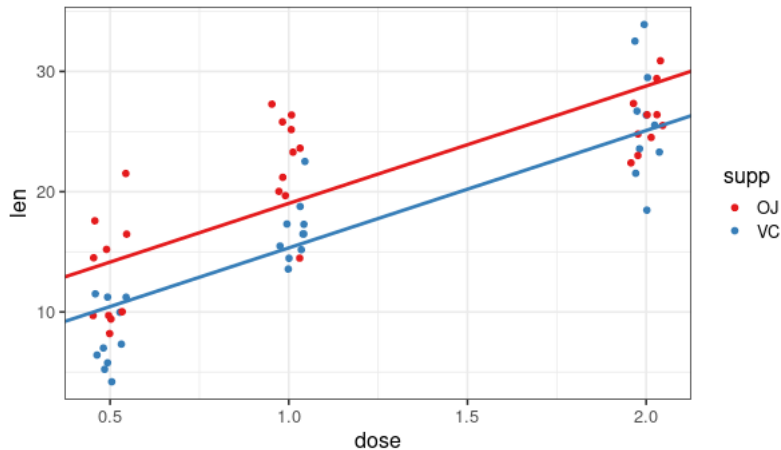
Building the model

```
1 > lm(len ~ supp + dose, ToothGrowth)
2
3 Coefficients:
4 (Intercept)      suppVC          dose
5           9.27         -3.70          9.76
```

- ▶ What is the formula for this regression line?
- ▶ What is my reference variable?
- ▶ Interpret intercept?
- ▶ How does ascorbic acid compare with orange juice?
- ▶ Is more vitamin C associated with greater tooth growth?

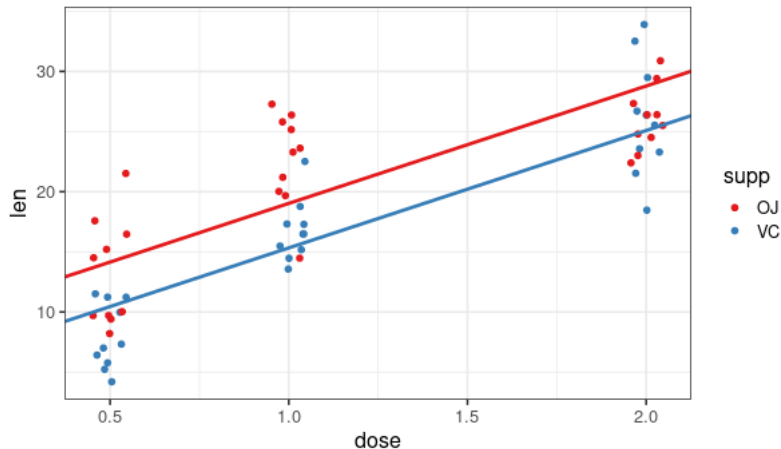
Model Interpretation

$$\widehat{\text{Tooth Growth}} = 9.27 - 3.70 \times \mathbb{1}_{VC} + 9.76 \times \text{Dose}$$



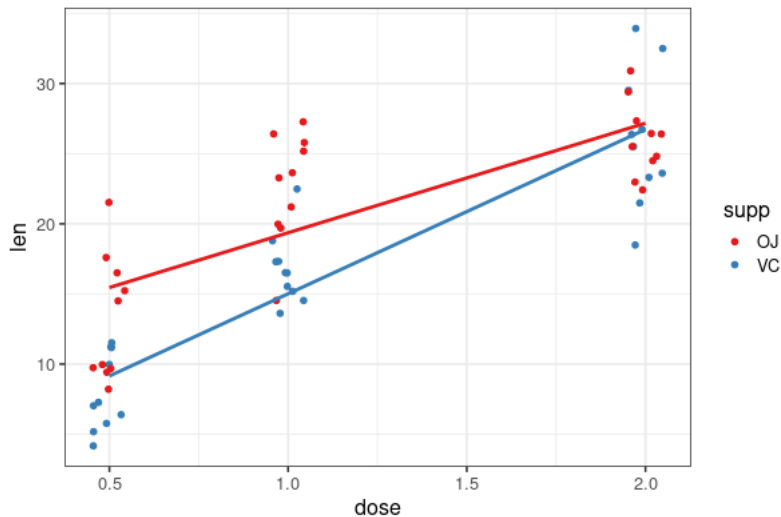
Model Interpretation

$$\widehat{\text{Tooth Growth}} = \begin{cases} 9.27 + 9.76 \times \text{Dose} & \text{if OJ} \\ 5.57 + 9.76 \times \text{Dose} & \text{if VC} \end{cases}$$



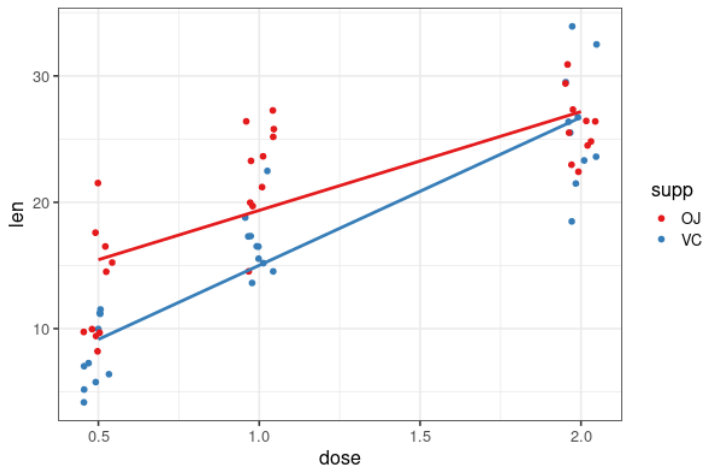
Stratified Models

Contrast this with a case in which we use *stratified models*, fitting two separate regression lines to different subsets of our data



Stratified Models

$$\widehat{\text{Tooth Growth}} = \begin{cases} 3.29 + 11.72 \times \text{Dose} & \text{if VC} \\ 11.55 + 7.81 \times \text{Dose} & \text{if OJ} \end{cases}$$



Stratified Models

Standard Linear Models

Pros:

- ▶ Uses all of the data (more observations = more conclusive statements)
- ▶ Captures overall trends
- ▶ Allows for direct comparisons of groups

Cons:

- ▶ Requires groups to share same slope for quantitative values
- ▶ Can be more difficult to interpret

Stratified Linear Models

Pros:

- ▶ Allows each group to have own slope term
- ▶ Simpler to interpret

Cons:

- ▶ Fewer observations per model
- ▶ Lose general trend information

Generally, we will only be using standard linear models for this course

Review

- ▶ Regression can contain any number of quantitative or categorical predictors
- ▶ Quantitative variables are associated with *slope*, i.e., unit changes in X that are associated with β changes in y
- ▶ Categorical variables use *indicator variables* that vertically shift a regression line up or down depending on groups
- ▶ Interpretation for intercept includes both *reference variables* as well as value of y when $X = 0$
- ▶ Difference between stratified and unstratified models