Multiple linear regression

PSTAT 120C

Outline

- **Matrices and matrix algebra**
  - Simple linear regression in matrix notation
- Multiple linear regression
  - OLS equations
  - Properties
  - Inferences
- Model-level inferences
- Prediction intervals
- Example
- Recap

Why matrices?

- Language for statistical models with multiple parameters

- Used in statistical theory **often**
  - Helpful for other PSTAT courses

Definitions

- **Matrix**: Rectangular array of real numbers.
- **Vector**: Matrix with only one row or column.
- **Elements**: Numbers in the matrix.
- **Size**: Number of rows and columns in the matrix, $m\ x\ n$
- **Dimensions**: m and n
- **Scalar**: A single element/number

$$\underset{3\ x\ 2}{\boldsymbol{X}} = \begin{bmatrix} 6.2 & 0.45 \\ -1.7 & 4 \\ 2 & 7.3 \end{bmatrix}$$

$$\underset{4\ x\ 1}{\boldsymbol{b}} = \begin{bmatrix} 1 \\ -3 \\ 0 \\ 7 \end{bmatrix}$$

Notation

$$\boldsymbol{X} = \begin{bmatrix} 20 & 5 & 2 \\ 17 & 11 & 6 \\ 19 & 20 & 18 \end{bmatrix} \qquad \boldsymbol{X} = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix}$$

$$\boldsymbol{a} = [43 \quad 2 \quad 36 \quad 22] \quad \boldsymbol{a} = [a_{11} \quad a_{12} \quad a_{13} \quad a_{14}]$$

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

Matrix algebra: Addition/subtraction

- Two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ can *only* be added if they are the same size.

$$\underset{\mathbf{3\,x\,3}}{\boldsymbol{A}} = \begin{bmatrix} 6 & 0 & 10 \\ 30 & 12 & 25 \\ 11 & 51 & 40 \end{bmatrix} \qquad \underset{\mathbf{3\,x\,3}}{\boldsymbol{B}} = \begin{bmatrix} 20 & 5 & 2 \\ 17 & 11 & 6 \\ 19 & 20 & 18 \end{bmatrix}$$

- Corresponding elements are added; the result is a matrix of the same dimensions.

$$\underset{\mathbf{3\,x\,3}}{\boldsymbol{A}+\boldsymbol{B}} = \begin{bmatrix} (20+6) & (5+0) & (2+10) \\ (17+30) & (11+12) & (6+25) \\ (19+11) & (20+51) & (18+40) \end{bmatrix} \qquad \underset{\mathbf{3\,x\,3}}{\boldsymbol{A}+\boldsymbol{B}} = \begin{bmatrix} 26 & 5 & 12 \\ 47 & 22 & 31 \\ 30 & 71 & 58 \end{bmatrix}$$

Matrix algebra: Scalar multiplication

$$\boldsymbol{D} = \begin{bmatrix} 6.2 & 0.45 \\ -1.7 & 4 \\ 2 & 7.3 \end{bmatrix}$$

- Consider $3\boldsymbol{D}$. To conform with addition rule, $3\boldsymbol{D} = \boldsymbol{D} + \boldsymbol{D} + \boldsymbol{D}$

$$3\boldsymbol{D} = \begin{bmatrix} 6.2(3) & 0.45(3) \\ -1.7(3) & 4(3) \\ 2(3) & 7.3(3) \end{bmatrix} \quad 3\boldsymbol{D} = \begin{bmatrix} 18.6 & 1.35 \\ -5.1 & 12 \\ 6 & 21.9 \end{bmatrix}$$

Matrix algebra: Multiplication

- To be multiplied, the **inner order** of the matrices must be the same:

$$\underset{\mathbf{2\ x\ 2}}{\boldsymbol{A}} = \begin{bmatrix} 2 & 0 \\ 1 & 4 \end{bmatrix} \qquad \underset{\mathbf{2\ x\ 2}}{\boldsymbol{B}} = \begin{bmatrix} 5 & 2 \\ -1 & 3 \end{bmatrix} \qquad \underset{\mathbf{2\ x\ 2}}{\boldsymbol{AB}} = \begin{bmatrix} 10 & 4 \\ 1 & 14 \end{bmatrix}$$

- $\boldsymbol{A}$ and $\boldsymbol{B}$ are **conformable**. Multiplication proceeds:

$$\boldsymbol{AB} = \underset{\substack{\mathbf{2\ x\ 2} \\ \text{pre-multiplier}}}{\begin{bmatrix} 2 & 0 \\ 1 & 4 \end{bmatrix}} \underset{\substack{\mathbf{2\ x\ 2} \\ \text{post-multiplier}}}{\begin{bmatrix} 5 & 2 \\ -1 & 3 \end{bmatrix}} \qquad \boldsymbol{AB} = \begin{bmatrix} 2(5) + 0(-1) & 2(2) + 0(3) \\ 1(5) + 4(-1) & 1(2) + 4(3) \end{bmatrix}$$

$$\boldsymbol{AB} = \begin{bmatrix} 10 & 4 \\ 1 & 14 \end{bmatrix}$$

More details of matrices

- Square matrices, like $\boldsymbol{A}$, have $m = n$
- Identity matrices are usually denoted $\boldsymbol{I}$
- Upper ($\boldsymbol{C}$) and lower triangular matrices
- Diagonal matrices (like $\boldsymbol{I}$)

$$\boldsymbol{A} = \begin{bmatrix} 6 & 0 & 10 \\ 30 & 12 & 25 \\ 11 & 51 & 40 \end{bmatrix}$$

$$\boldsymbol{C} = \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ 0 & c_{22} & c_{23} \\ 0 & 0 & c_{33} \end{bmatrix}$$

$$\boldsymbol{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Sum of squares in matrix notation

$$\underset{1\,x\,4}{\boldsymbol{y}} = [0.03 \quad 1.15 \quad 1.53 \quad -0.26]$$

$$\underset{4\,x\,1}{\boldsymbol{y}'} = \begin{bmatrix} 0.03 \\ 1.15 \\ 1.53 \\ -0.26 \end{bmatrix}$$

$$\sum_{i=1}^{n} {Y_i}^2 = \cancel{\boldsymbol{yy}} = \boldsymbol{yy}'$$

$$\boldsymbol{yy}' = [0.03(0.03) + 1.15(1.15) + 1.53(1.53) + (-0.26)(-0.26)]$$

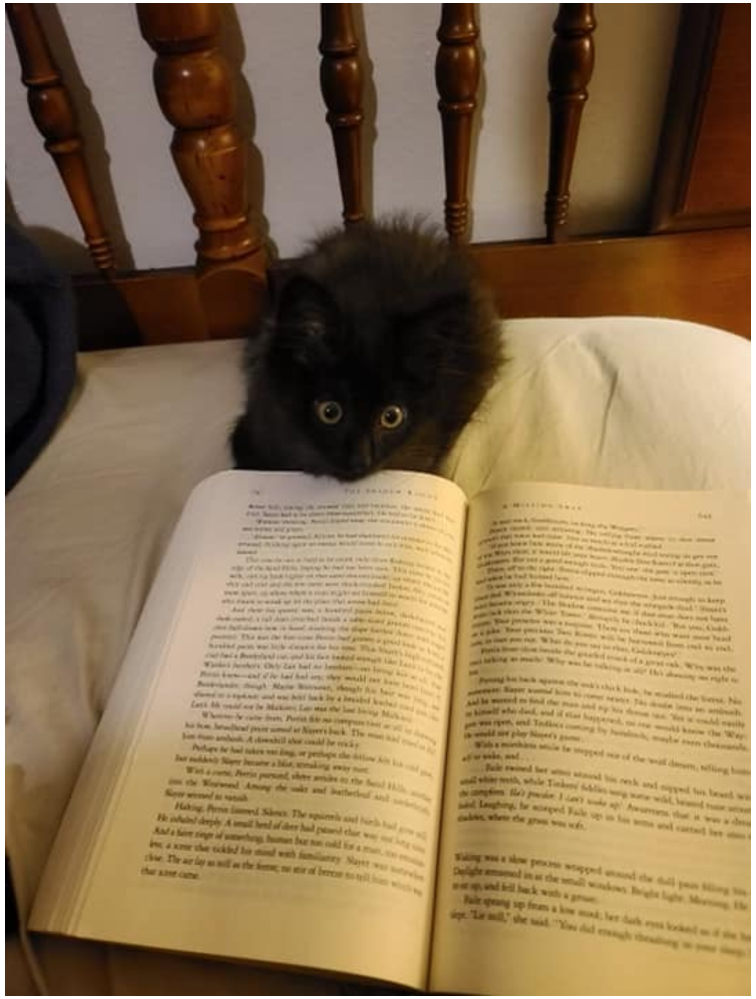
$$\boldsymbol{yy}' = [3.7319]$$

Matrix inversion

$$\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}$$

$$\boldsymbol{A} = \begin{bmatrix} 2 & 0 \\ 1 & -1 \end{bmatrix} \qquad \boldsymbol{A}^{-1} =$$

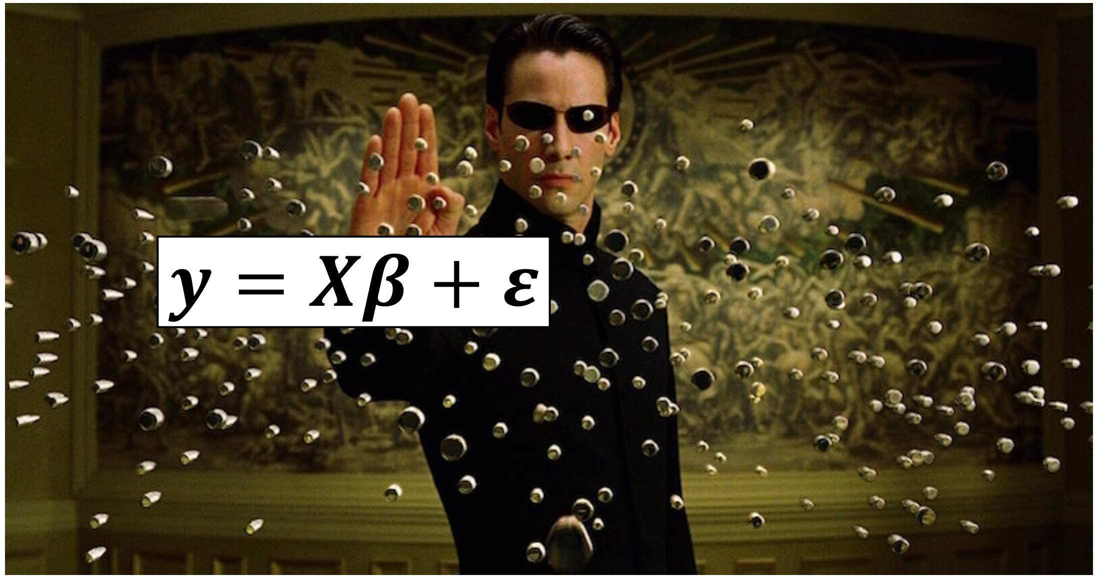
$$\boldsymbol{B} = \begin{bmatrix} \color{red}{2} & \color{red}{0} \\ \color{red}{4} & \color{red}{0} \end{bmatrix}$$



Outline

- Matrices and matrix algebra
  - **Simple linear regression in matrix notation**
- Multiple linear regression
  - OLS equations
  - Properties
  - Inferences
- Model-level inferences
- Prediction intervals
- Recap

Simple linear regression: Matrix style



- Assume that we have $Y_1, \ldots, Y_n$ independent random variables of the form $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

$$\underset{\boldsymbol{n\ x\ 1}}{\boldsymbol{y}} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \underset{\boldsymbol{2\ x\ 1}}{\boldsymbol{\beta}} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \quad \underset{\boldsymbol{n\ x\ 2}}{\boldsymbol{X}} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$$

$$\underset{\boldsymbol{n\ x\ 1}}{\boldsymbol{\varepsilon}} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$\underset{\boldsymbol{n\ x\ 1}}{\boldsymbol{X\beta}} = \begin{bmatrix} \beta_0 + \beta_1 x_1 \\ \beta_0 + \beta_1 x_n \\ \vdots \\ \beta_0 + \beta_1 x_n \end{bmatrix}$$

Simple linear regression, cont.

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$$

$$\hat{\beta}_0 \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i$$

$$\boldsymbol{X}'\boldsymbol{X} = \underset{\boldsymbol{2\ x\ n}}{\begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{bmatrix}} \underset{\boldsymbol{n\ x\ 2}}{\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}} = \underset{\boldsymbol{2\ x\ 2}}{\begin{bmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{bmatrix}}$$

$$\boldsymbol{X'X} = \begin{bmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{bmatrix}$$

2 x 2

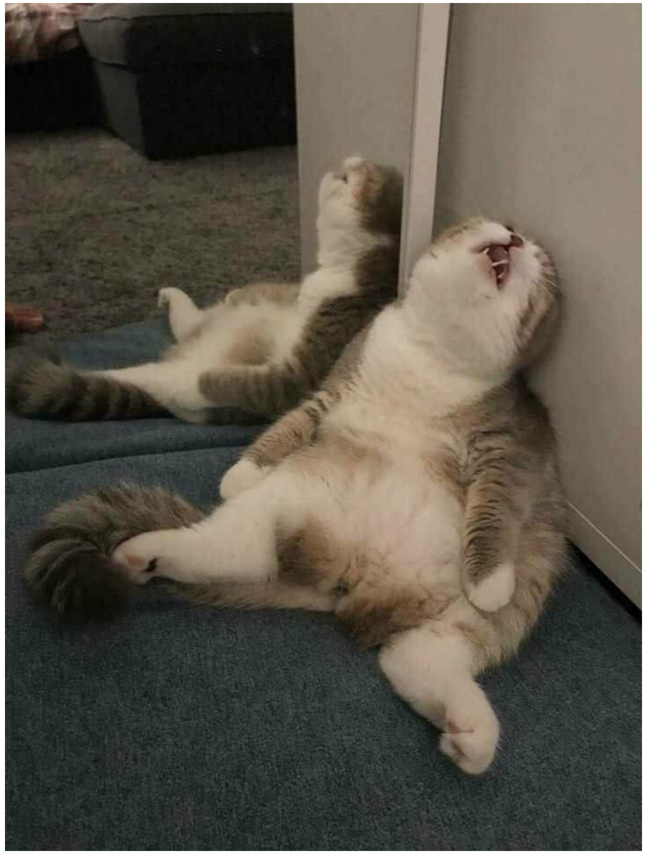
$$n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$$

$$\hat{\beta}_0 \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i$$

$$(\boldsymbol{X'X})\widehat{\boldsymbol{\beta}} = \boldsymbol{X'Y}$$

$$\boldsymbol{X'Y} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \end{bmatrix}$$

2 x n **n x 1** **2 x 1**



$$\underset{\mathbf{2\ x\ 1}}{\widehat{\boldsymbol{\beta}}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix}$$

$$\boldsymbol{X'X} = \begin{bmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{bmatrix}$$

2 x 2

$$\boldsymbol{X'Y} = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \end{bmatrix}$$

2 x 1

$$\underset{2\,x\,1}{\widehat{\boldsymbol{\beta}}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix}$$

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$$

$$\hat{\beta}_0 \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i$$

$$\begin{bmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \end{bmatrix}$$

$$(\boldsymbol{X'X})\widehat{\boldsymbol{\beta}} = \boldsymbol{X'Y}$$

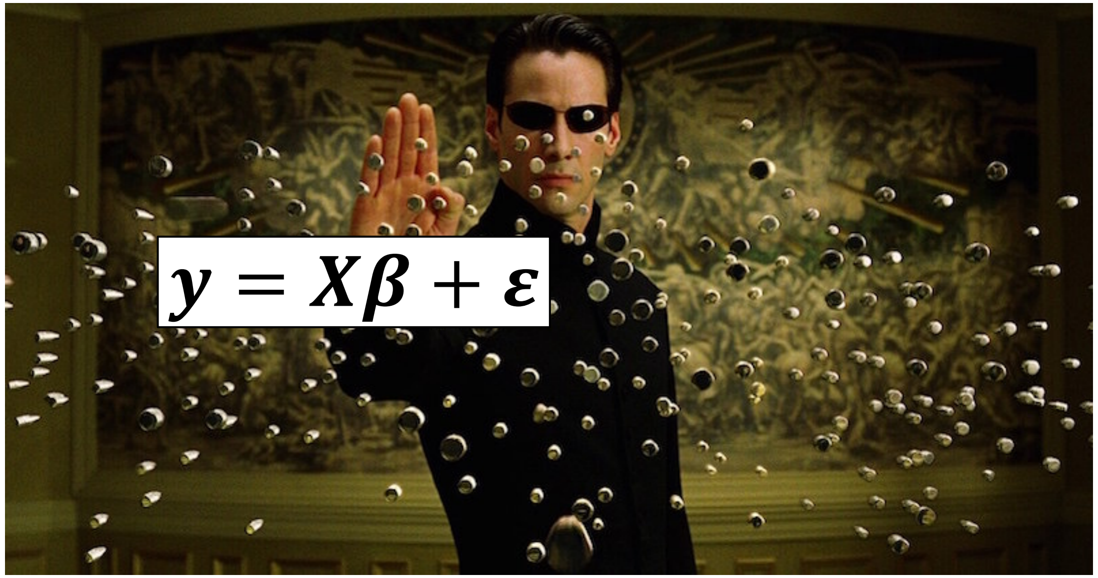
Outline

- **Matrices and matrix algebra**
  - **Simple linear regression in matrix notation**
- **Multiple linear regression**
  - OLS equations
  - Properties
  - Inferences
- Model-level inferences
- Prediction intervals
- Example
- Recap

Multiple linear regression

$$Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon$$

$$\underset{n\,x\,1}{\boldsymbol{y}} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \underset{p\,x\,1}{\boldsymbol{\beta}} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} \quad \underset{n\,x\,1}{\boldsymbol{\varepsilon}} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$



$$\underset{n\,x\,p}{\boldsymbol{X}} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ 1 & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n1} & \cdots & x_{np} \end{bmatrix}$$

$$\underset{n\,x\,1}{\boldsymbol{X\beta}} = \begin{bmatrix} \beta_0 + \beta_1 x_{11} + \cdots + \beta_p x_{1p} \\ \beta_0 + \beta_1 x_{21} + \cdots + \beta_p x_{2p} \\ \vdots \\ \beta_0 + \beta_1 x_{n1} + \cdots + \beta_p x_{np} \end{bmatrix}$$

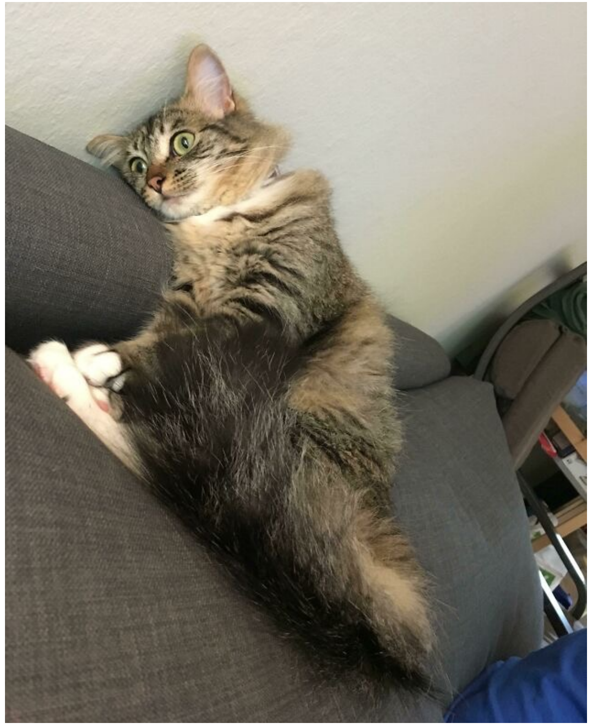
Outline

- Matrices and matrix algebra
  - Simple linear regression in matrix notation
- **Multiple linear regression**
  - **OLS equations**
  - Properties
  - Inferences
- Model-level inferences
- Prediction intervals
- Example
- Recap

Multiple linear regression

$$(\boldsymbol{X'X})\boldsymbol{\widehat{\beta}} = \boldsymbol{X'Y}$$

$$\boldsymbol{X'X} = \underset{\boldsymbol{p\ x\ n}}{\begin{bmatrix} 1 & 1 & 1 & 1 \\ x_{11} & x_{21} & \cdots & x_{n1} \\ x_{12} & x_{22} & \cdots & x_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1p} & x_{2p} & \cdots & x_{np} \end{bmatrix}} \underset{\boldsymbol{n\ x\ p}}{\begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ 1 & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}}$$



$$\boldsymbol{X'X} = \underset{\boldsymbol{p\ x\ p}}{\begin{bmatrix} n & \sum_{i=1}^{n} x_{i1} & \sum_{i=1}^{n} x_{i2} & \cdots & \sum_{i=1}^{n} x_{ip} \\ \sum_{i=1}^{n} x_{i1} & \sum_{i=1}^{n} x_{i1}^2 & \sum_{i=1}^{n} x_{i1}x_{i2} & \cdots & \sum_{i=1}^{n} x_{i1}x_{ip} \\ \sum_{i=1}^{n} x_{i2} & \sum_{i=1}^{n} x_{i2}x_{i1} & \sum_{i=1}^{n} x_{i2}^2 & \cdots & \sum_{i=1}^{n} x_{i2}x_{ip} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{n} x_{ip} & \sum_{i=1}^{n} x_{ip}x_{i1} & \sum_{i=1}^{n} x_{ip}x_{i2} & \cdots & \sum_{i=1}^{n} x_{ip}^2 \end{bmatrix}}$$

Multiple linear regression

$$(\boldsymbol{X'X})\widehat{\boldsymbol{\beta}} = \boldsymbol{X'Y}$$

$$\boldsymbol{X'Y} = \underset{\boldsymbol{p\ x\ n}}{\begin{bmatrix} 1 & 1 & & 1 & 1 \\ x_{11} & x_{21} & \cdots & x_{n1} \\ x_{12} & x_{22} & \cdots & x_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1p} & x_{2p} & \cdots & x_{np} \end{bmatrix}} \underset{\boldsymbol{n\ x\ 1}}{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}} = \underset{\boldsymbol{p\ x\ 1}}{\begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_{n1} y_i \\ \vdots \\ \sum_{i=1}^{n} x_{np} y_i \end{bmatrix}}$$

Multiple linear regression

$$(X'X)\widehat{\boldsymbol{\beta}} = X'Y$$

$$\widehat{\boldsymbol{\beta}} = (X'X)^{-1}X'Y$$

$$\begin{bmatrix} n & \sum_{i=1}^{n} x_{i1} & \sum_{i=1}^{n} x_{i2} & \cdots & \sum_{i=1}^{n} x_{ip} \\ \sum_{i=1}^{n} x_{i1} & \sum_{i=1}^{n} x_{i1}^2 & \sum_{i=1}^{n} x_{i1}x_{i2} & \cdots & \sum_{i=1}^{n} x_{i1}x_{ip} \\ \sum_{i=1}^{n} x_{i2} & \sum_{i=1}^{n} x_{i2}x_{i1} & \sum_{i=1}^{n} x_{i2}^2 & \cdots & \sum_{i=1}^{n} x_{i2}x_{ip} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{n} x_{ip} & \sum_{i=1}^{n} x_{ip}x_{i1} & \sum_{i=1}^{n} x_{ip}x_{i2} & \cdots & \sum_{i=1}^{n} x_{ip}^2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_p \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_{n1}y_i \\ \vdots \\ \sum_{i=1}^{n} x_{np}y_i \end{bmatrix}$$

$\boldsymbol{p \; x \; p}$ $\qquad$ $\boldsymbol{p \; x \; 1}$ $\qquad$ $\boldsymbol{p \; x \; 1}$

Outline

- **Matrices and matrix algebra**
  - **Simple linear regression in matrix notation**
- **Multiple linear regression**
  - **OLS equations**
  - **Properties**
  - Inferences
- Model-level inferences
- Prediction intervals
- Example
- Recap

Multiple linear regression

Properties of the Least-Squares Estimators: Multiple Linear Regression

1. $E(\hat{\beta}_i) = \beta_i, i = 0, 1, \ldots, k$.
2. $V(\hat{\beta}_i) = c_{ii}\sigma^2$, where c_{ii} is the element in row i and column i of $(\mathbf{X}'\mathbf{X})^{-1}$. (Recall that this matrix has a row and column numbered 0.)
3. $\text{Cov}(\hat{\beta}_i, \hat{\beta}_j) = c_{ij}\sigma^2$, where c_{ij} is the element in row i and column j of $(\mathbf{X}'\mathbf{X})^{-1}$.
4. An unbiased estimator of σ^2 is $S^2 = \text{SSE}/[n - (k + 1)]$, where $\text{SSE} = \mathbf{Y}'\mathbf{Y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y}$. (Notice that there are $k + 1$ unknown β_i values in the model.)

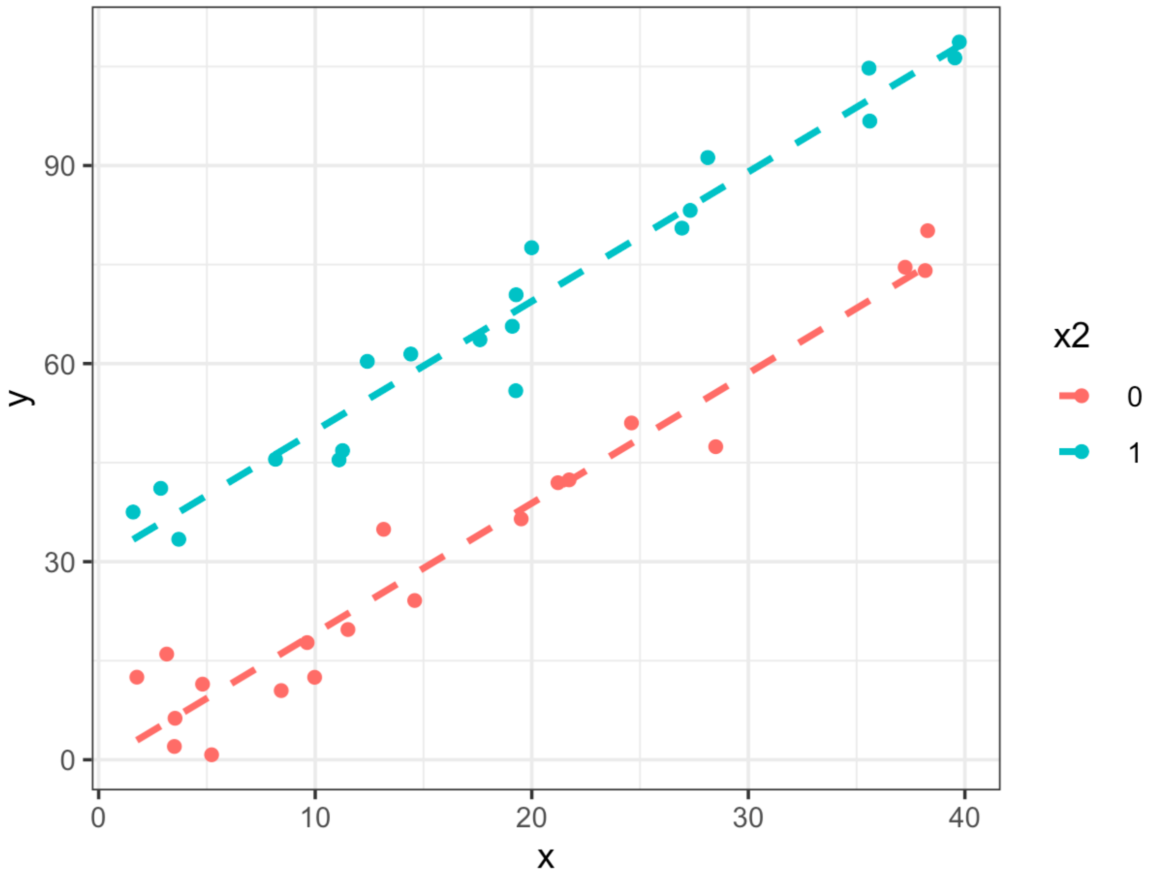
If, in addition, the ε_i, for $i = 1, 2, \ldots, n$ are normally distributed,

5. Each $\hat{\beta}_i$ is normally distributed.
6. The random variable
$$\frac{[n - (k + 1)]S^2}{\sigma^2}$$
has a χ^2 distribution with $n - (k + 1)$ df.
7. The statistic S^2 and $\hat{\beta}_i$ are independent for each $i = 0, 1, 2, \ldots, k$.

Multiple linear regression

1. Linearity in parameters
2. $E[\varepsilon] = 0$
3. $E[\varepsilon|X] = 0$
4. Uncorrelated errors
5. Homoscedasticity
6. No multicollinearity
7. Normally distributed errors

Visualizing multiple linear regression



$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

$$Y = 0 + 2x_1 + 30x_2$$

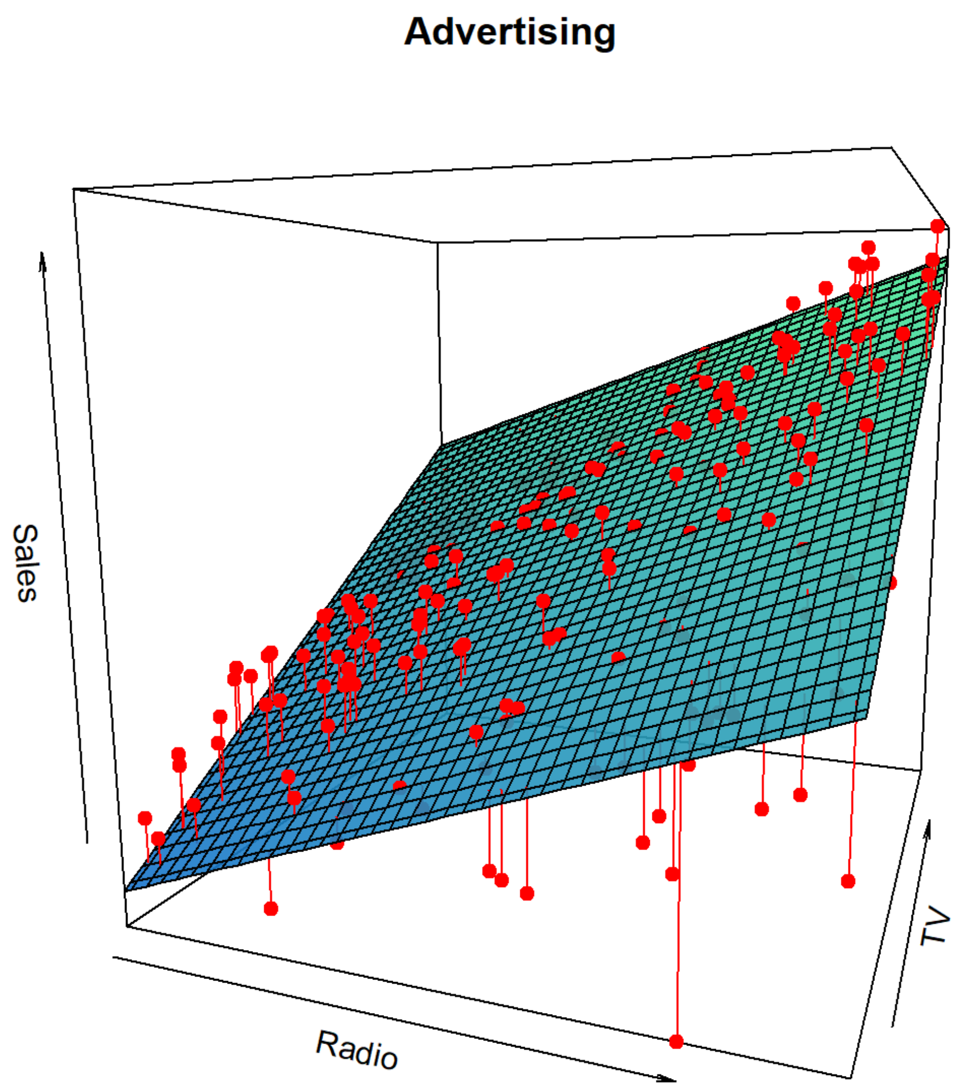
For values of Y where x_2 is 0:

$$Y = \beta_0 + \beta_1 x_1 + \cancel{\beta_2(0)}$$

For values of Y where x_2 is 1:

$$Y = \beta_0 + \beta_1 x_1 + 30(1)$$

Visualizing multiple linear regression



$$Sales = \beta_0 + \beta_1 Radio + \beta_2 TV$$

Interpreting results

```
> warpbreaks
   breaks wool tension
1      26    A       L
2      30    A       L
3      54    A       L
4      25    A       L
5      70    A       L
6      52    A       L
7      51    A       L
8      26    A       L
9      67    A       L
10     18    A       M
11     21    A       M
12     29    A       M
13     17    A       M
14     12    A       M
15     18    A       M
```

```
> lm(data = warpbreaks, formula = breaks ~ wool + tension) %>%
+ summary()

Call:
lm(formula = breaks ~ wool + tension, data = warpbreaks)

Residuals:
    Min      1Q  Median      3Q     Max 
-19.500  -8.083  -2.139   6.472  30.722 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   39.278      3.162  12.423  < 2e-16 ***
woolB         -5.778      3.162  -1.827 0.073614 .  
tensionM     -10.000      3.872  -2.582 0.012787 *  
tensionH     -14.722      3.872  -3.802 0.000391 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11.62 on 50 degrees of freedom
Multiple R-squared:  0.2691,	Adjusted R-squared:  0.2253 
F-statistic: 6.138 on 3 and 50 DF,  p-value: 0.00123
```

Outline

- **Matrices and matrix algebra**
  - **Simple linear regression in matrix notation**
- **Multiple linear regression**
  - **OLS equations**
  - **Properties**
  - **Inferences**
- Model-level inferences
- Prediction intervals
- Example
- Recap

Tests of linear functions of the parameters

$$E[Y] = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$$

$$\theta = \boldsymbol{\beta_0 a_0 + \beta_1 a_1 + \cdots + \beta_p a_p}$$

$$\underset{p\,x\,1}{\boldsymbol{a}} = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} \quad \underset{p\,x\,1}{\boldsymbol{\beta}} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} \quad \underset{1\,x\,1}{\boldsymbol{a'\beta}} = [\boldsymbol{\beta_0 a_0 + \beta_1 a_1 + \cdots + \beta_p a_p}]$$

Thrm. 5.12

Let $Y_1, Y_2, \ldots, Y_n$ and $X_1, X_2, \ldots, X_m$ be random variables with $E(Y_i) = \mu_i$ and $E(X_j) = \xi_j$. Define

$$U_1 = \sum_{i=1}^{n} a_i Y_i \quad \text{and} \quad U_2 = \sum_{j=1}^{m} b_j X_j$$

for constants $a_1, a_2, \ldots, a_n$ and $b_1, b_2, \ldots, b_m$. Then the following hold:

a $E(U_1) = \sum_{i=1}^{n} a_i \mu_i$.

b $V(U_1) = \sum_{i=1}^{n} a_i^2 V(Y_i) + 2 \sum\sum_{1 \le i < j \le n} a_i a_j \text{Cov}(Y_i, Y_j)$, where the double sum is over all pairs (i, j) with $i < j$.

c $\text{Cov}(U_1, U_2) = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j \text{Cov}(Y_i, X_j)$.

$$\widehat{\boldsymbol{a'\beta}} = \boldsymbol{a_0\beta_0 + a_1\beta_1 + \cdots + a_p\beta_p = a'\beta}$$

$$E[\boldsymbol{a'\beta}] = ?$$

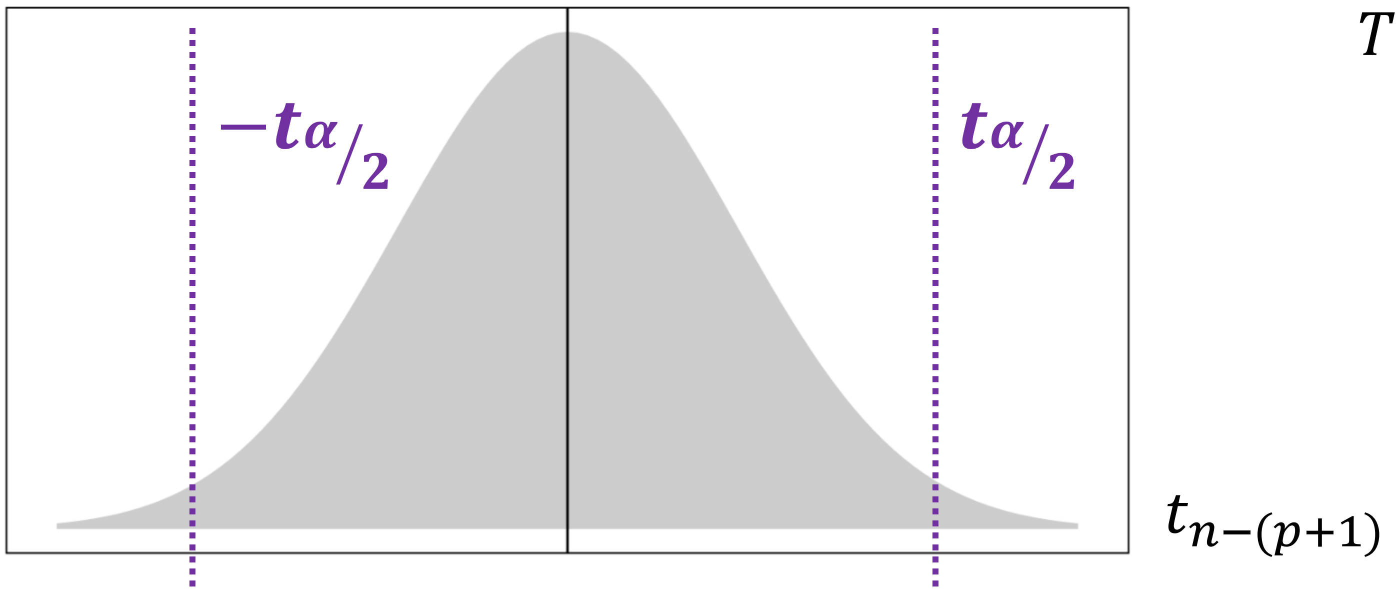
$$V[\boldsymbol{a'\beta}] = ?$$

Tests of linear functions of the parameters

$$\boldsymbol{a}'\boldsymbol{\beta} = \boldsymbol{\beta}_0\boldsymbol{a}_0 + \boldsymbol{\beta}_1\boldsymbol{a}_1 + \cdots + \boldsymbol{\beta}_p\boldsymbol{a}_p$$

$$H_0\colon \boldsymbol{a}'\boldsymbol{\beta} = (\boldsymbol{a}'\boldsymbol{\beta})_0$$
$$H_a\colon \boldsymbol{a}'\boldsymbol{\beta} \neq (\boldsymbol{a}'\boldsymbol{\beta})_0$$

Sampling distribution of $\boldsymbol{a}'\boldsymbol{\beta}$ **under the null**



$$T = \frac{\boldsymbol{a}'\boldsymbol{\beta} - (\boldsymbol{a}'\boldsymbol{\beta})_0}{S\sqrt{(\boldsymbol{a}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{a})}} \sim t_{n-(p+1)}$$

A $100(1-\alpha)\%$ CI:

$$\widehat{\boldsymbol{a}'\boldsymbol{\beta}} \pm t_{\alpha/2} S\sqrt{\boldsymbol{a}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{a}}$$

But what about single parameters?

$$\widehat{\boldsymbol{a'\beta}} = \boldsymbol{a_0\beta_0} + \boldsymbol{a_1\beta_1} + \cdots + \boldsymbol{a_p\beta_p} = \boldsymbol{a'\beta}$$

- If we only want to test β_1, we can let:

$$a_j = \begin{cases} 0, & if\ j \neq 1 \\ 1, & if\ j = 1 \end{cases}$$

- The other equations don't change, and this can be modified to test any combination of parameters.

Back to results

$$T = \frac{\boldsymbol{a}'\boldsymbol{\beta} - (\boldsymbol{a}'\boldsymbol{\beta})_0}{S\sqrt{(\boldsymbol{a}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{a})}} \sim t_{n-(p+1)}$$

```
> lm(data = warpbreaks, formula = breaks ~ wool + tension) %>%
+ summary()

Call:
lm(formula = breaks ~ wool + tension, data = warpbreaks)

Residuals:
    Min      1Q  Median      3Q     Max 
-19.500  -8.083  -2.139   6.472  30.722 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   39.278      3.162  12.423  < 2e-16 ***
woolB         -5.778      3.162  -1.827 0.073614 .  
tensionM     -10.000      3.872  -2.582 0.012787 *  
tensionH     -14.722      3.872  -3.802 0.000391 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11.62 on 50 degrees of freedom
Multiple R-squared:  0.2691,	Adjusted R-squared:  0.2253 
F-statistic: 6.138 on 3 and 50 DF,  p-value: 0.00123
```

Outline

- **Matrices and matrix algebra**
  - **Simple linear regression in matrix notation**
- **Multiple linear regression**
  - **OLS equations**
  - **Properties**
  - **Inferences**
- **Model-level inferences**
- Prediction intervals
- Example
- Recap

Comparing models

Model 1:

$$Y = \beta_0 + \beta_1 x_1 + \beta_p x_p + \varepsilon$$

$$\boldsymbol{SSE}_R$$

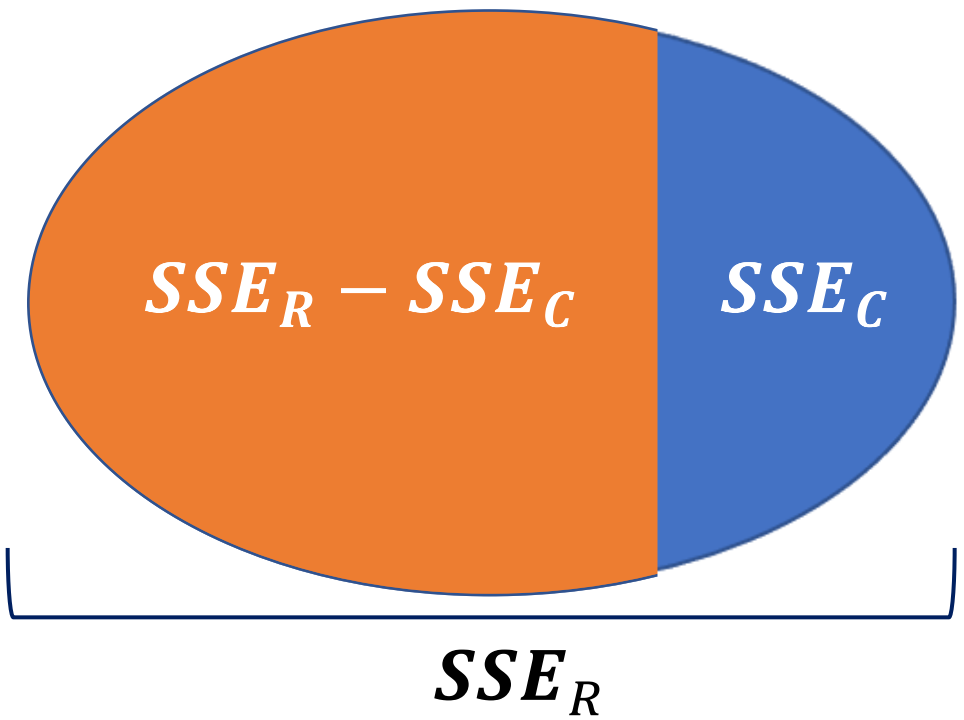
Model 2:

$$Y = \beta_0 + \beta_1 x_1 + \beta_p x_p + \beta_p x_p + \varepsilon$$

$$\boldsymbol{SSE}_{\boldsymbol{C}}$$

- A better model will explain more of the variance in Y and therefore will have a smaller SSE

- $SSE_C \leq SSE_R$



Model 1:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

$$\boldsymbol{SSE_R}$$

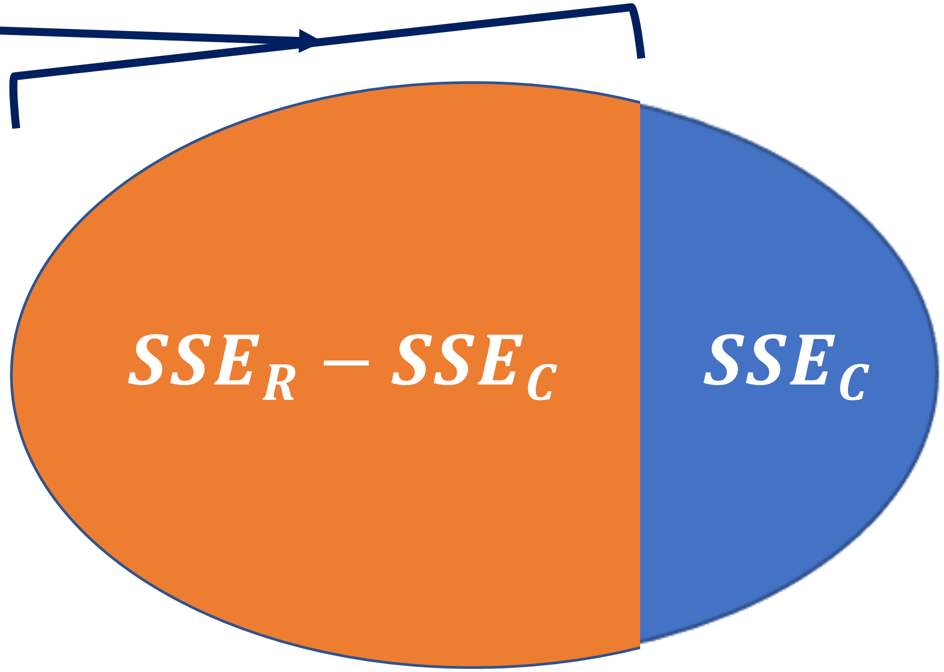
Model 2:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

$$\boldsymbol{SSE_C}$$

- Difference in sum of squares associated with x_3, adjusted for x_1 and x_2

$$H_0: \beta_3 = 0$$



Model 1:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

$$\boldsymbol{SSE_R}$$

$$g = 2$$

Model 2:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

$$\boldsymbol{SSE_C}$$

$$k = 3$$

- If H_0 is true:

$$\frac{SSE_R}{\sigma^2} \sim \chi^2_{(n-(g+1))} \qquad \frac{SSE_C}{\sigma^2} \sim \chi^2_{(n-(k+1))} \qquad \frac{SSE_R - SSE_C}{\sigma^2} \sim \chi^2_{(k-g)}$$

DEFINITION **7.3**

Let W_1 and W_2 be *independent* χ^2-distributed random variables with ν_1 and ν_2 df, respectively. Then

$$F = \frac{W_1/\nu_1}{W_2/\nu_2}$$

is said to have an F distribution with ν_1 numerator degrees of freedom and ν_2 denominator degrees of freedom.

Model 1:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

$$\boldsymbol{SSE}_R$$

$$g = 2$$

Model 2:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

$$\boldsymbol{SSE}_C$$

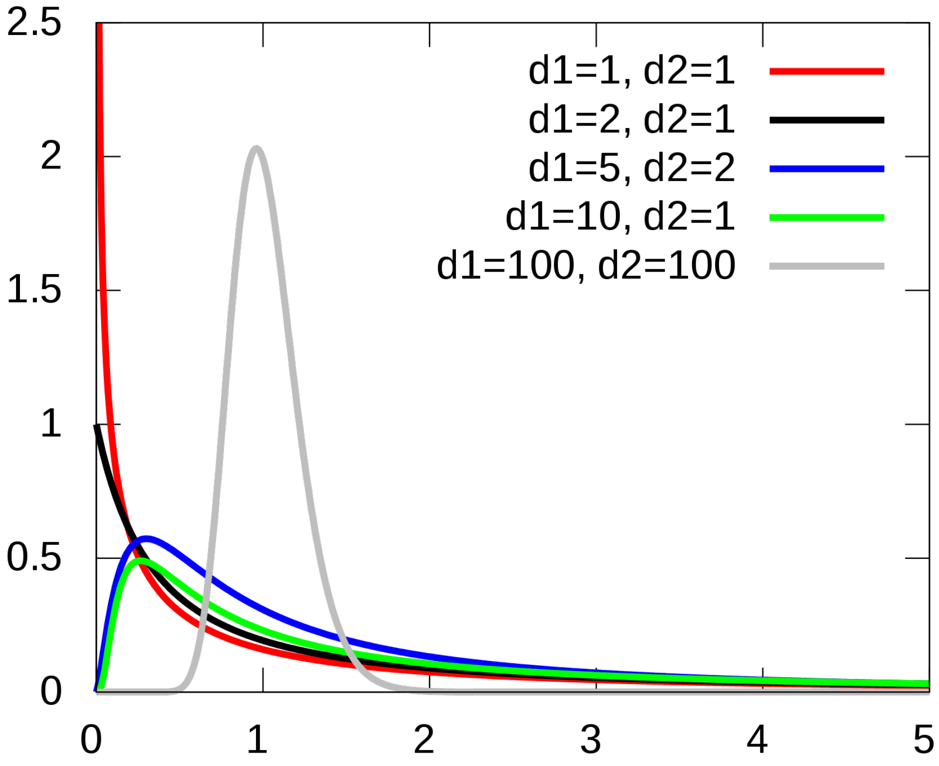
$$k = 3$$

$$\boldsymbol{H_0 : \beta_{g+1} = \beta_{g+2} = \cdots = \beta_k = 0}$$

- So:

$$F = \frac{\chi^2_{(k-g)} \Big/ (k-g)}{\chi^2_{(n-(k+1))} \Big/ n-(k+1)}$$

$$= \frac{(SSE_R - SSE_C)/(k-g)}{SSE_C/(n-(k+1))}$$



Model 1:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

$$\boldsymbol{SSE_R}$$

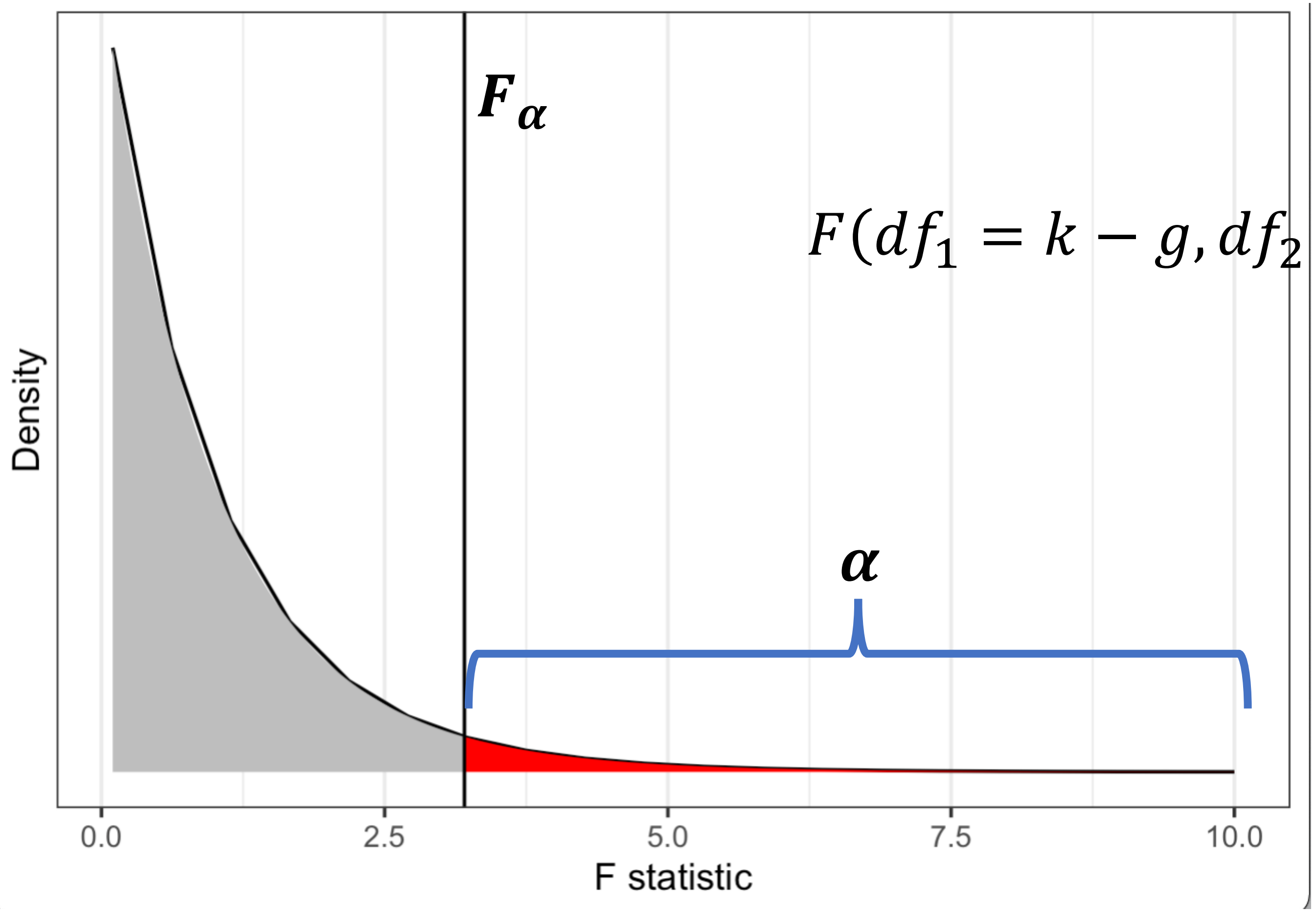
$$\boldsymbol{H_0: \beta_{g+1} = \beta_{g+2} = \cdots = \beta_k = 0}$$

Model 2:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

$$\boldsymbol{SSE_C}$$

$$\boldsymbol{H_A: \beta_{g+1} \neq \beta_{g+2} \neq \cdots \neq \beta_k \neq 0}$$



$$F(df_1 = k - g, df_2 = n - (k+1))$$

Outline

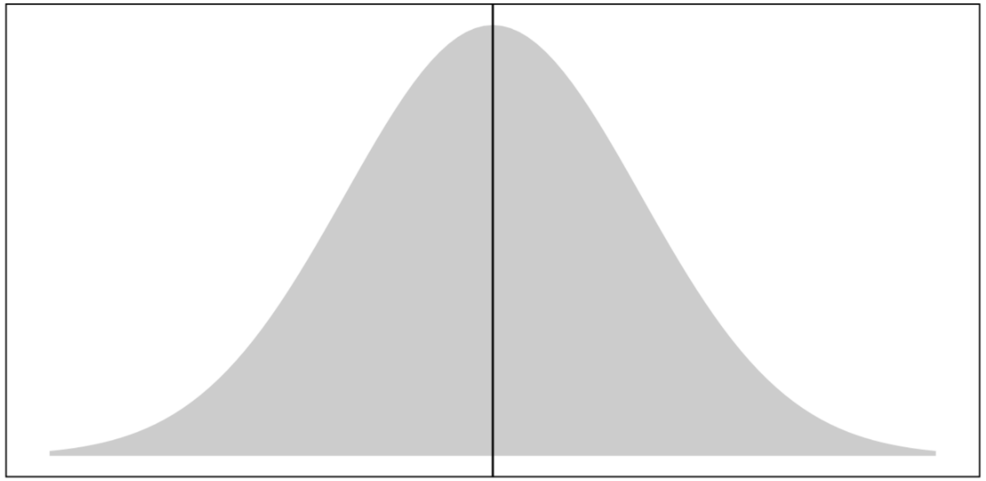
- **Matrices and matrix algebra**
  - **Simple linear regression in matrix notation**
- **Multiple linear regression**
  - **OLS equations**
  - **Properties**
  - **Inferences**
- **Model-level inferences**
- **Prediction intervals**
- Example
- Recap

Prediction intervals

- Again, we want the **exact** value of Y, but not just when $x_1 = x^*$
  - $x_1 = x_1^*, x_2 = x_2^*, \ldots, x_p = x_p^*$

- Same method as before:
- $\widehat{Y^*} = \hat{\beta}_0 + \hat{\beta}_1 x_1^* + \hat{\beta}_2 x_2^* + \cdots + \hat{\beta}_p x_p^*$
- $error = Y^* - \widehat{Y^*}$

$error$



$E[error] = 0$

$V[error] = \sigma^2\left[1 + \boldsymbol{a}'(\boldsymbol{X}'\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{a}\right]$

$$T = \frac{Y^* - \widehat{Y^*}}{S\sqrt{1 + \boldsymbol{a}'(\boldsymbol{X}'\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{a}}}$$

A $\mathbf{100(1-\boldsymbol{\alpha})}$% CI for Y when $x_1 = x_1^*, \ldots, x_p = x_p^*$ **:**

$$\boldsymbol{a}'\widehat{\boldsymbol{\beta}} \pm t_{\alpha/2} S\sqrt{1 + \boldsymbol{a}'(\boldsymbol{X}'\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{a}}$$

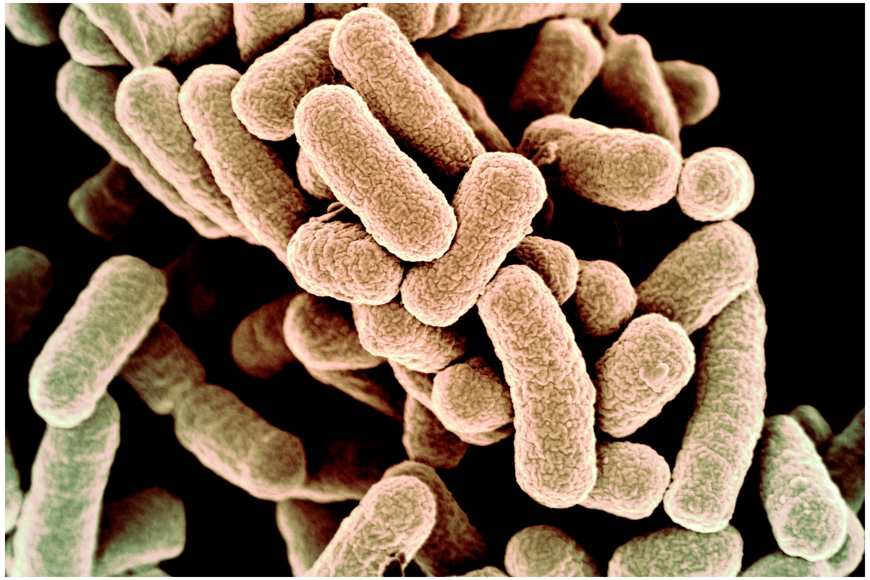
Outline

- Matrices and matrix algebra
  - Simple linear regression in matrix notation
- Multiple linear regression
  - OLS equations
  - Properties
  - Inferences
- Model-level inferences
- Prediction intervals
- **Example**
- Recap

Example: #11.101

The data in the accompanying table come from the comparison of the growth rates for bacteria types A and B. The growth Y recorded at five equally spaced (and coded) points of time is shown in the table.

	Time				
Bacteria Type	−2	−1	0	1	2
A	8.0	9.0	9.1	10.2	10.4
B	10.0	10.3	12.2	12.6	13.9



a Fit the linear model

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \varepsilon$$

to the $n = 10$ data points. Let $x_1 = 1$ if the point refers to bacteria type B and let $x_1 = 0$ if the point refers to type A. Let $x_2 =$ coded time.

b Plot the data points and graph the two growth lines. Notice that β_3 is the difference between the slopes of the two lines and represents time–bacteria interaction.

c Predict the growth of type A at time $x_2 = 0$ and compare the answer with the graph. Repeat the process for type B.

d Do the data present sufficient evidence to indicate a difference in the rates of growth for the two types of bacteria?

e Find a 90% confidence interval for the expected growth for type B at time $x_2 = 1$.

f Find a 90% prediction interval for the growth Y of type B at time $x_2 = 1$.

Outline

- Matrices and matrix algebra
  - Simple linear regression in matrix notation
- Multiple linear regression
  - OLS equations
  - Properties
  - Inferences
- Model-level inferences
- Prediction intervals
- Example
- **Recap**

Recap

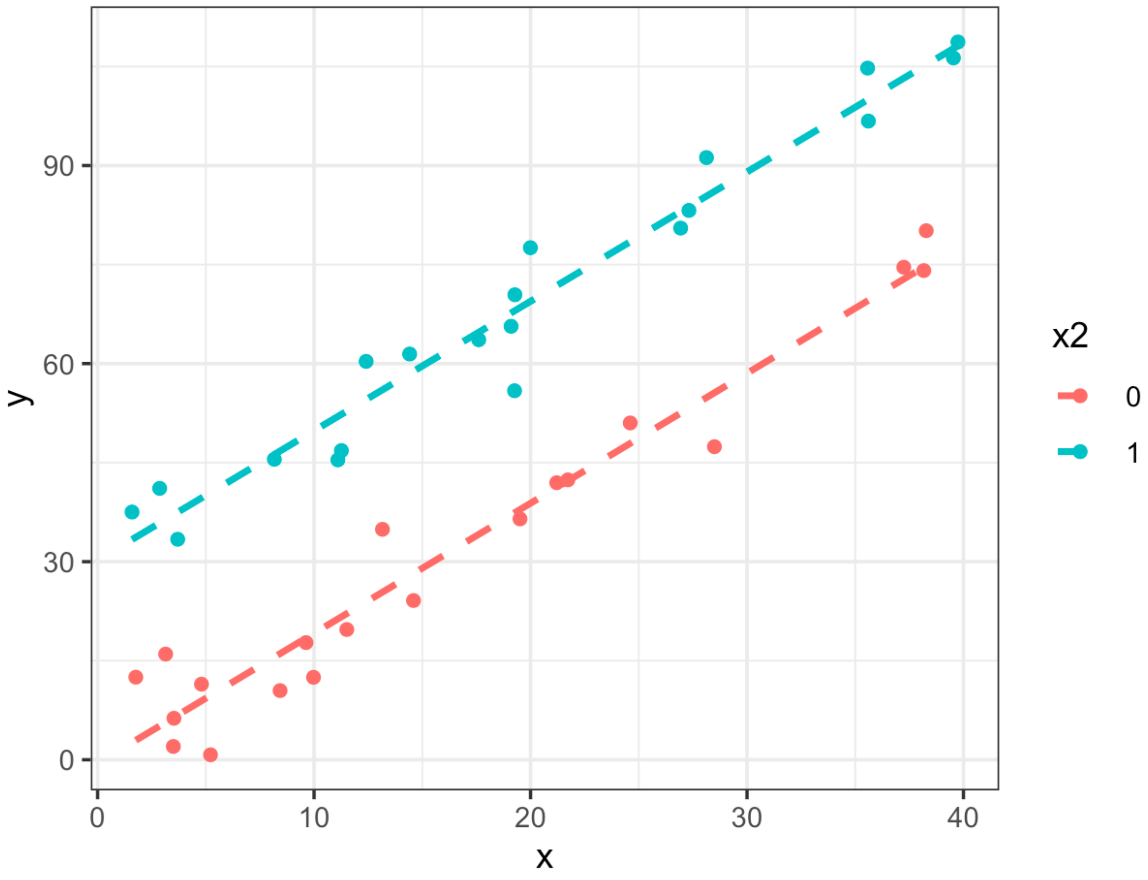
- A probabilistic model for Y as a linear function of multiple independent variables
  - $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \varepsilon$
  - $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$

- Estimates obtained using least squares:
  - $\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X}'\boldsymbol{Y}$

Recap

- Assumptions
  - Linearity
  - $E[\varepsilon] = 0$
  - $E[\varepsilon|X] = 0$
  - Uncorrelated errors
  - Homoscedasticity
  - No multicollinearity
  - Normally distributed errors

- Visualization
- Interpretation



Recap

- Inferences:
  - Single parameters: $H_0: \beta_1 = 0$
  - Multiple parameters: $H_0: \beta_1 + \beta_2 + \cdots + \beta_p = 0$
  - Reduced vs. complete models: $H_0: \beta_1 = \beta_2 = \cdots = \beta_p = 0$

- Prediction intervals:
  - $\boldsymbol{a}'\widehat{\boldsymbol{\beta}} \pm t_{\alpha/2} S\sqrt{1 + \boldsymbol{a}'(\boldsymbol{X}'\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{a}}$

Resources

- Multiple Regression, Clearly Explained
- Confounding Variables and Omitted Variable Bias
- Linear Regression Assumptions
- The Main Idea of Fitting a Line to Data
- Linear Regression, Clearly Explained
- Linear Regression in R, Step by Step
- Multiple Regression in R, Step by Step
- R-squared, Clearly Explained