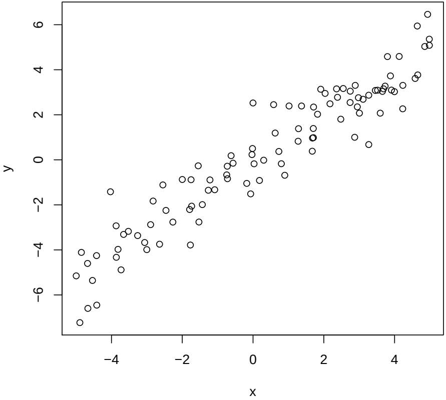
College of Media, Communication and Information
UNIVERSITY OF COLORADO BOULDER

Linear Regression

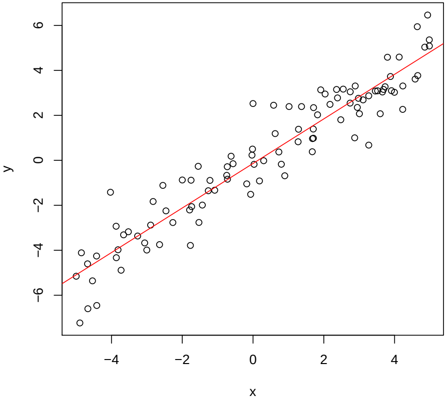
INFO-2301: Quantitative Reasoning 2
Michael Paul and Jordan Boyd-Graber
SLIDES ADAPTED FROM LAUREN HANNAH

Linear Regression



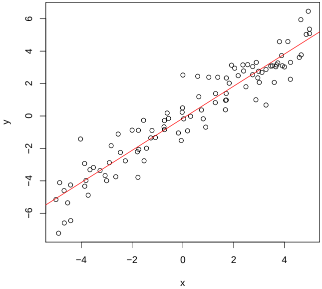
Data are the set of inputs and outputs, $\mathscr{D} = \left\{(x_i, y_i)\right\}_{i=1}^{n}$

Linear Regression



In *linear regression*, the goal is to predict y from x using a linear function

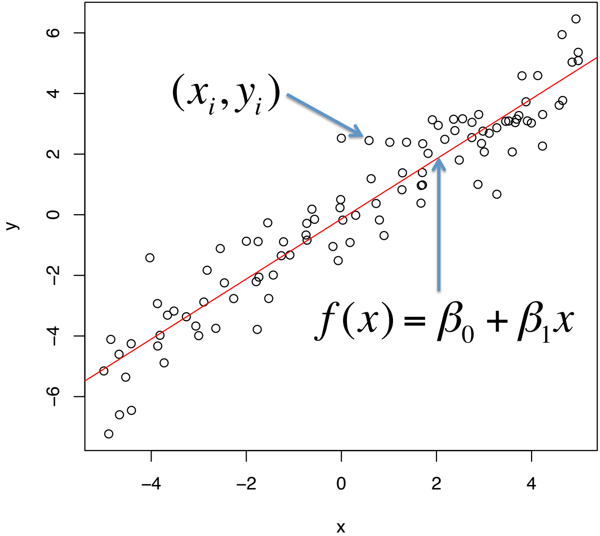
Linear Regression



Examples of linear regression:

- given a child's age and gender, what is his/her height?
- given unemployment, inflation, number of wars, and economic growth, what will the president's approval rating be?
- given a browsing history, how long will a user stay on a page?

Linear Regression



Multiple Covariates

Often, we have a vector of inputs where each represents a different *feature* of the data

$$\mathbf{x} = (x_1, \ldots, x_p)$$

The function fitted to the response is a linear combination of the covariates

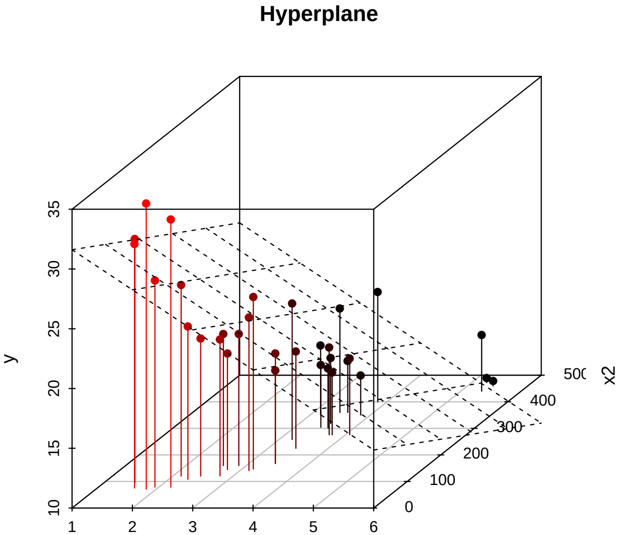
$$f(\mathbf{x}) = \beta_0 + \sum_{j=1}^{p} \beta_j x_j$$

Multiple Covariates

- Often, it is convenient to represent **x** as $(1, x_1, \dots, x_p)$
- In this case **x** is a vector, and so is β (we'll represent them in bold face)
- This is the dot product between these two vectors
- This then becomes a sum

$$\beta \mathbf{x} = \sum_{j=1}^{p} \beta_j x_j$$

Hyperplanes: Linear Functions in Multiple Dimensions



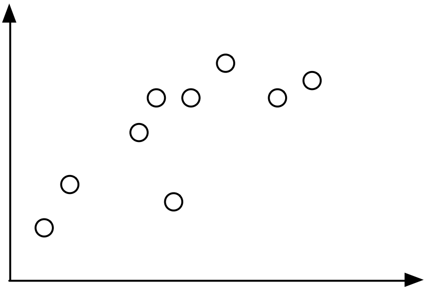
Covariates

- Do not need to be raw value of $x_1, x_2, \ldots$
- Can be any feature or function of the data:
  - Transformations like $x_2 = \log(x_1)$ or $x_2 = \cos(x_1)$
  - Basis expansions like $x_2 = x_1^2$, $x_3 = x_1^3$, $x_4 = x_1^4$, etc
  - Indicators of events like $x_2 = 1_{\{-1 \leq x_1 \leq 1\}}$
  - Interactions between variables like $x_3 = x_1 x_2$
- Because of its simplicity and flexibility, it is one of the most widely implemented regression techniques

Prediction

- After finding $\hat{\beta}$, we would like to predict an output value for a new set of covariates
- We just find the point on the line that corresponds to the new input:

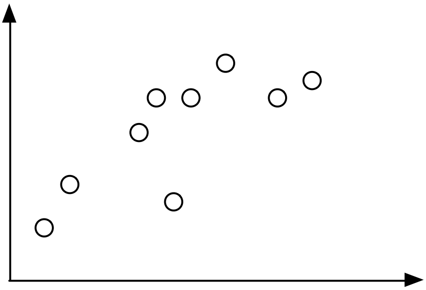
$$\hat{y} = \beta_0 + \beta_1 x \tag{1}$$



Prediction

- After finding $\hat{\beta}$, we would like to predict an output value for a new set of covariates
- We just find the point on the line that corresponds to the new input:

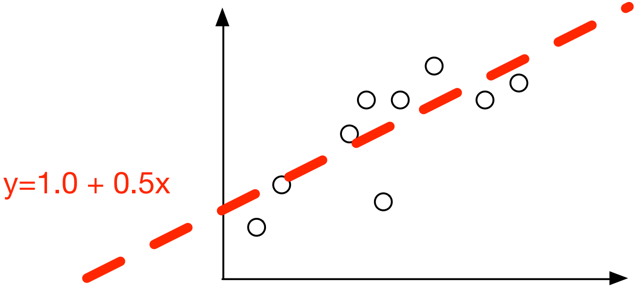
$$\hat{y} = \beta_0 + \beta_1 x \tag{1}$$



Prediction

- After finding $\hat{\beta}$, we would like to predict an output value for a new set of covariates
- We just find the point on the line that corresponds to the new input:

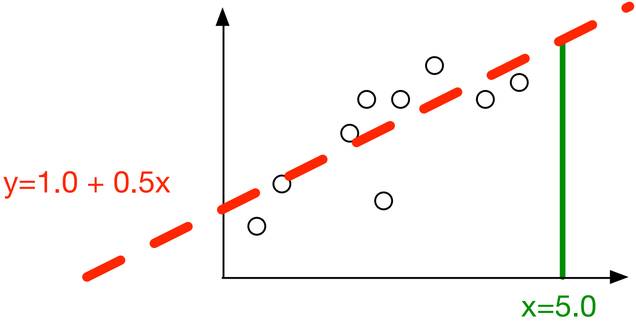
$$\hat{y} = 1.0 + 0.5x \tag{1}$$



Prediction

- After finding $\hat{\beta}$, we would like to predict an output value for a new set of covariates
- We just find the point on the line that corresponds to the new input:

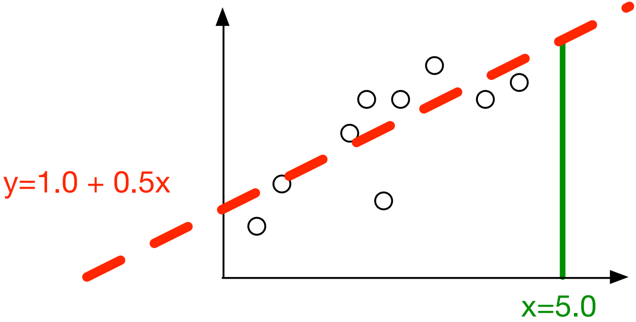
$$\hat{y} = 1.0 + 0.5 * 5 \tag{1}$$



Prediction

- After finding $\hat{\beta}$, we would like to predict an output value for a new set of covariates
- We just find the point on the line that corresponds to the new input:

$$\hat{y} = 3.5 \tag{1}$$



College of Media, Communication and Information
UNIVERSITY OF COLORADO BOULDER

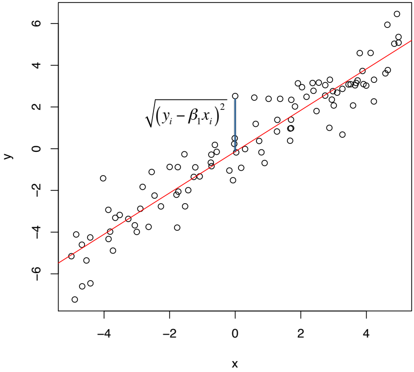
Linear Regression

INFO-2301: Quantitative Reasoning 2

Michael Paul and Jordan Boyd-Graber

SLIDES ADAPTED FROM LAUREN HANNAH

Fitting a Linear Regression



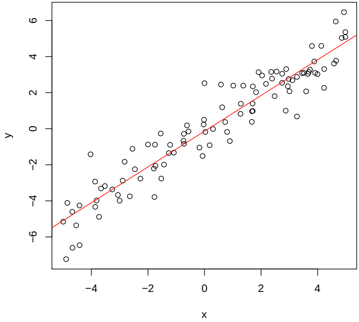
Idea: minimize the Euclidean distance between data and fitted line

$$RSS(\beta) = \frac{1}{2}\sum_{i=1}^{n}(y_i - \beta \mathbf{x}_i)^2$$

How to Find β

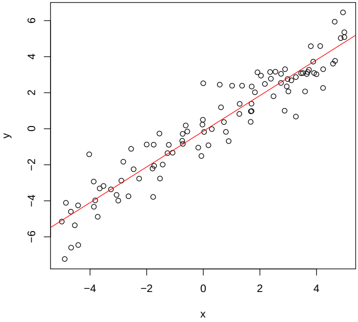
- Use calculus to find the value of β that minimizes the RSS
- The optimal value is

$$\hat{\beta} = \frac{\sum_{i=1}^{n} y_i x_i}{\sum_{i=1}^{n} x_i^2}$$



Probabilistic Interpretation

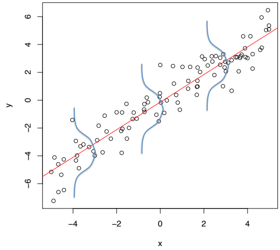
- Our analysis so far has not included any probabilities
- Linear regression does have a *probabilisitc* (probability model-based) interpretation



Probabilistic Interpretation

- Linear regression assumes that response values have a Gaussian distribution around the linear mean function,

$$Y_i \,|\, \mathbf{x}_i, \beta \sim N(\mathbf{x}_i \beta, \sigma^2)$$



- Minimizing RSS is equivalent to maximizing conditional likelihood

College of Media, Communication and Information
UNIVERSITY OF COLORADO BOULDER

Linear Regression

INFO-2301: Quantitative Reasoning 2
Michael Paul and Jordan Boyd-Graber
SLIDES ADAPTED FROM FEDERICO

Deriving Least Squares Regression

- Common theme in data science:
  - Build model
  - Write error model
  - Derive how to minimize error

Model and Objective

Model

$$y_i = b_0 + b_1 x_i + e_i \tag{1}$$

Error

$$e_i = y_i - b_1 x_i - b_0 = e_i \tag{2}$$

Objective

$$\ell \equiv \sum_i e_i^2 \tag{3}$$

Partial Derivatives

Intercept

$$\frac{\partial \ell}{\partial b_0} = \frac{\partial \sum_i (y_i - b_0 - b_1 x_i)^2}{\partial b_0} =$$

Partial Derivatives

Intercept

$$\frac{\partial \ell}{\partial b_0} = \frac{\partial \sum_i (y_i - b_0 - b_1 x_i)^2}{\partial b_0} = -2 \sum_i (y_i - b_0 - b_1 x_i) \tag{4}$$

Partial Derivatives

Intercept

$$\frac{\partial \ell}{\partial b_0} = \frac{\partial \sum_i (y_i - b_0 - b_1 x_i)^2}{\partial b_0} = -2 \sum_i (y_i - b_0 - b_1 x_i) \tag{4}$$

Slope

$$\frac{\partial \ell}{\partial b_1} = \frac{\partial \sum_i (y_i - b_0 - b_1 x_i)^2}{\partial b_1} =$$

Partial Derivatives

Intercept

$$\frac{\partial \ell}{\partial b_0} = \frac{\partial \sum_i (y_i - b_0 - b_1 x_i)^2}{\partial b_0} = -2\sum_i (y_i - b_0 - b_1 x_i) \tag{4}$$

Slope

$$\frac{\partial \ell}{\partial b_1} = \frac{\partial \sum_i (y_i - b_0 - b_1 x_i)^2}{\partial b_1} = -2\sum_i x_i (y_i - b_0 - b_1 x_i) \tag{5}$$

System of Equations with Two Unknowns

Solve for Intercept

(6)

System of Equations with Two Unknowns

Solve for Intercept

$$0 = -2\sum_i (y_i - b_0 - b_1 x_i) \tag{6}$$

$$\tag{7}$$

System of Equations with Two Unknowns

Solve for Intercept

$$0 = -2\sum_i (y_i - b_0 - b_1 x_i) \tag{6}$$

$$0 = \sum_i y_i - \sum_i b_0 - b_i \sum_i x_i \tag{7}$$

$$\tag{8}$$

Multiply by $-\frac{1}{2}$, distribute sum

System of Equations with Two Unknowns

Solve for Intercept

$$0 = -2\sum_i (y_i - b_0 - b_1 x_i) \tag{6}$$

$$0 = \sum_i y_i - \sum_i b_0 - b_i \sum_i x_i \tag{7}$$

$$Nb_0 = \sum_i y_i - b_i \sum_i x_i \tag{8}$$

$$\tag{9}$$

b_0 is constant, so $\sum_i b_0 = Nb_0$, move to LHS

System of Equations with Two Unknowns

Solve for Intercept

$$0 = -2\sum_i (y_i - b_0 - b_1 x_i) \quad (6)$$

$$0 = \sum_i y_i - \sum_i b_0 - b_i \sum_i x_i \quad (7)$$

$$Nb_0 = \sum_i y_i - b_i \sum_i x_i \quad (8)$$

$$b_0 = \left(\frac{\sum_i y_i}{N}\right) - b_1 \left(\frac{\sum_i x_i}{N}\right) \quad (9)$$

$$(10)$$

Divide by N

System of Equations with Two Unknowns

Solve for Intercept

$$0 = -2\sum_i (y_i - b_0 - b_1 x_i) \tag{6}$$

$$0 = \sum_i y_i - \sum_i b_0 - b_i \sum_i x_i \tag{7}$$

$$Nb_0 = \sum_i y_i - b_i \sum_i x_i \tag{8}$$

$$b_0 = \left(\frac{\sum_i y_i}{N}\right) - b_1 \left(\frac{\sum_i x_i}{N}\right) \tag{9}$$

$$b_0 = \bar{y} - b_1 \bar{x} \tag{10}$$

System of Equations with Two Unknowns

Solve for Intercept

$$b_0 = \bar{y} - b_1 \bar{x} \tag{6}$$

Solve for Slope

(7)

System of Equations with Two Unknowns

Solve for Intercept

$$b_0 = \bar{y} - b_1 \bar{x} \tag{6}$$

Solve for Slope

$$0 = -2 \sum_i x_i (y_i - b_0 - b_1 x_i) \tag{7}$$

$$\tag{8}$$

System of Equations with Two Unknowns

Solve for Intercept

$$b_0 = \bar{y} - b_1 \bar{x} \tag{6}$$

Solve for Slope

$$0 = -2 \sum_i x_i (y_i - b_0 - b_1 x_i) \tag{7}$$

$$0 = \sum_i x_i y_i - b_0 \sum_i x_i - \sum_i b_1 x_i^2 \tag{8}$$

$$\tag{9}$$

Multiply by $-\frac{1}{2}$, distribute sum and x_i

System of Equations with Two Unknowns

Solve for Intercept

$$b_0 = \bar{y} - b_1 \bar{x} \tag{6}$$

Solve for Slope

$$0 = -2 \sum_i x_i (y_i - b_0 - b_1 x_i) \tag{7}$$

$$0 = \sum_i x_i y_i - b_0 \sum_i x_i - \sum_i b_1 x_i^2 \tag{8}$$

$$b_1 \sum_i x_i^2 = \sum_i x_i y_i - b_0 \sum_i x_i \tag{9}$$

$$\tag{10}$$

Move last term to RHS

System of Equations with Two Unknowns

Solve for Intercept

$$b_0 = \bar{y} - b_1 \bar{x} \tag{6}$$

Solve for Slope

$$0 = -2 \sum_i x_i (y_i - b_0 - b_1 x_i) \tag{7}$$

$$0 = \sum_i x_i y_i - b_0 \sum_i x_i - \sum_i b_1 x_i^2 \tag{8}$$

$$b_1 \sum_i x_i^2 = \sum_i x_i y_i - b_0 \sum_i x_i \tag{9}$$

$$b_1 \sum_i x_i^2 = \sum_i x_i y_i - \left[\left(\frac{\sum_i y_i}{N} \right) - b_1 \left(\frac{\sum_i x_i}{N} \right) \right] \sum_i x_i \tag{10}$$

Solve for Slope (continued)

$$b_1 \sum_i x_i^2 = \sum_i x_i y_i - \left[\left(\frac{\sum_i y_i}{N} \right) - b_1 \left(\frac{\sum_i x_i}{N} \right) \right] \sum_i x_i$$

Solve for Slope (continued)

$$b_1 \sum_i x_i^2 = \sum_i x_i y_i - \left[\left(\frac{\sum_i y_i}{N} \right) - b_1 \left(\frac{\sum_i x_i}{N} \right) \right] \sum_i x_i$$

$$b_1 \sum_i x_i^2 = \sum_i x_i y_i - \left(\frac{\sum_i y_i \sum_i x_i}{N} \right) - b_1 \left(\frac{(\sum_i x_i)^2}{N} \right)$$

Multiplying out the last term

Solve for Slope (continued)

$$b_1 \sum_i x_i^2 = \sum_i x_i y_i - \left[\left(\frac{\sum_i y_i}{N} \right) - b_1 \left(\frac{\sum_i x_i}{N} \right) \right] \sum_i x_i$$

$$b_1 \sum_i x_i^2 = \sum_i x_i y_i - \left(\frac{\sum_i y_i \sum_i x_i}{N} \right) - b_1 \left(\frac{(\sum_i x_i)^2}{N} \right)$$

$$b_1 \sum_i x_i^2 + b_1 \left(\frac{(\sum_i x_i)^2}{N} \right) = \sum_i x_i y_i - \left(\frac{\sum_i y_i \sum_i x_i}{N} \right)$$

Move last term to LHS

Solve for Slope (continued)

$$b_1 \sum_i x_i^2 = \sum_i x_i y_i - \left[\left(\frac{\sum_i y_i}{N} \right) - b_1 \left(\frac{\sum_i x_i}{N} \right) \right] \sum_i x_i$$

$$b_1 \sum_i x_i^2 = \sum_i x_i y_i - \left(\frac{\sum_i y_i \sum_i x_i}{N} \right) - b_1 \left(\frac{(\sum_i x_i)^2}{N} \right)$$

$$b_1 \sum_i x_i^2 + b_1 \left(\frac{(\sum_i x_i)^2}{N} \right) = \sum_i x_i y_i - \left(\frac{\sum_i y_i \sum_i x_i}{N} \right)$$

$$b_1 \left[\sum_i x_i^2 + \left(\frac{(\sum_i x_i)^2}{N} \right) \right] = \sum_i x_i y_i - \left(\frac{\sum_i y_i \sum_i x_i}{N} \right)$$

Factor out b_1

Solve for Slope (continued)

$$b_1 \sum_i x_i^2 = \sum_i x_i y_i - \left[\left(\frac{\sum_i y_i}{N} \right) - b_1 \left(\frac{\sum_i x_i}{N} \right) \right] \sum_i x_i$$

$$b_1 \sum_i x_i^2 = \sum_i x_i y_i - \left(\frac{\sum_i y_i \sum_i x_i}{N} \right) - b_1 \left(\frac{(\sum_i x_i)^2}{N} \right)$$

$$b_1 \sum_i x_i^2 + b_1 \left(\frac{(\sum_i x_i)^2}{N} \right) = \sum_i x_i y_i - \left(\frac{\sum_i y_i \sum_i x_i}{N} \right)$$

$$b_1 \left[\sum_i x_i^2 + \left(\frac{(\sum_i x_i)^2}{N} \right) \right] = \sum_i x_i y_i - \left(\frac{\sum_i y_i \sum_i x_i}{N} \right)$$

$$b_1 = \frac{\sum_i x_i y_i - \left(\frac{\sum_i y_i \sum_i x_i}{N} \right)}{\sum_i x_i^2 + \left(\frac{(\sum_i x_i)^2}{N} \right)}$$

Solve for Slope (continued)

$$b_1 = \frac{\sum_i x_i y_i - \left(\frac{\sum_i y_i \sum_i x_i}{N} \right)}{\sum_i x_i^2 + \left(\frac{(\sum_i x_i)^2}{N} \right)}$$

Ratio of the sum of the crossproducts of *x* and *y* over the sum of squares for *x*

College of Media, Communication and Information
UNIVERSITY OF COLORADO BOULDER

Linear Regression

INFO-2301: Quantitative Reasoning 2
Michael Paul and Jordan Boyd-Graber
SLIDES ADAPTED FROM DAVID NEAL

Correlation Coefficient

True Correlation

$$\rho = \frac{\mu_{XY} - \mu_X \mu_Y}{\sigma_X \sigma_Y} \tag{1}$$

Sample Correlation

$$r = \frac{\bar{xy} - (\bar{x})(\bar{y})}{\sqrt{\bar{x^2} - (\bar{x})^2}\sqrt{\bar{y^2} - (\bar{y})^2}} \tag{2}$$

- If x and y are independent, then correlation is 0.
- Great if $\rho = \pm 1$
- Can we test how good the regression is?

Statistical Test for Regression

- Null hypothesis $H_0 : \rho = 0$
- Test statistic

$$T = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \tag{3}$$

- Follows a t-distribution with $n-2$ degress of freedom (estimating two parameters)
- Can do either two-tailed or one-tailed test

Wrapup

- Regression: powerful tool for explaining data
- Allows you to tell stories
- Allows you to predict the future
- Foundation for more complicated models