

# STAT 111

## Recitation 7

Mo Huang

Email: [mohuang@wharton.upenn.edu](mailto:mohuang@wharton.upenn.edu)

Office Hours: Wednesdays 3:00 - 4:00 pm, JMHH F96

Slides (adapted from Gemma Moran): [github.com/mohuangx/STAT111-Fall2018](https://github.com/mohuangx/STAT111-Fall2018)

October 26, 2018

# Estimation Summary

- ▶ Binomial parameter  $\theta$ :

Estimate:  $p$

95% confidence interval:  $p \pm \sqrt{1/n}$

- ▶ Mean  $\mu$ :

Estimate:  $\bar{x}$

95% confidence interval:  $\bar{x} \pm 2 \frac{s}{\sqrt{n}}$

- ▶ Difference between proportions  $\theta_1 - \theta_2$ :

Estimate:  $p_1 - p_2$

95% confidence interval:  $p_1 - p_2 \pm \sqrt{\frac{1}{n} + \frac{1}{m}}$

- ▶ Difference between means  $\mu_1 - \mu_2$ :

Estimate:  $\bar{x}_1 - \bar{x}_2$

95% confidence interval:  $\bar{x}_1 - \bar{x}_2 \pm 2 \sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{m}}$

Note:

$$s = \sqrt{\frac{x_1^2 + x_2^2 + \cdots + x_n^2 - n(\bar{x}^2)}{n-1}}.$$

# Linear Regression

- ▶ How does the growth height of a plant in a greenhouse depend on the amount of water that we give it?
- ▶ Let  $x$  be the amount of water we plan to give the plant (fixed).
- ▶ Let  $Y$  be the (future) growth height of a plant (random variable).
- ▶ **Linear Regression Model:** for the  $i$ th observation,
  - ▶ Mean of  $Y_i = \alpha + \beta x_i$ ,      Variance of  $Y_i = \sigma^2$ .

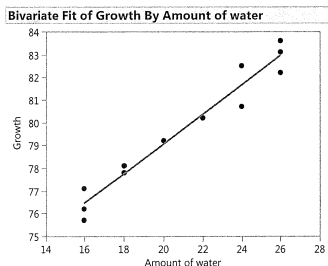


Figure: Data:  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$

# Linear Regression

- Mean of  $Y_i = \alpha + \beta x_i$ , Variance of  $Y_i = \sigma^2$ . How do we estimate  $\alpha$ ,  $\beta$ ?

$$\text{Calculate: } s_{xx} = x_1^2 + x_2^2 + \cdots + x_n^2 - n(\bar{x}^2) = \sum_{i=1}^n x_i^2 - n(\bar{x}^2)$$

$$s_{yy} = y_1^2 + y_2^2 + \cdots + y_n^2 - n(\bar{y}^2) = \sum_{i=1}^n y_i^2 - n(\bar{y}^2)$$

$$s_{xy} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n - n(\bar{x}\bar{y}) = \sum_{i=1}^n x_i y_i - n(\bar{x}\bar{y}).$$

- Estimate  $\beta$  by  $b$  and  $\alpha$  by  $a$ :

$$b = s_{xy}/s_{xx}$$

$$a = \bar{y} - b\bar{x}$$

- Estimate  $\sigma^2$  by  $s_r^2$ :

$$s_r^2 = (s_{yy} - b^2 s_{xx})/(n - 2).$$

- How accurate is the estimate  $b$  of  $\beta$ ?

$$\text{Standard deviation of } b : \frac{s_r}{\sqrt{s_{xx}}} \Rightarrow 95\% \text{ C.I. } b \pm 2 \frac{s_r}{\sqrt{s_{xx}}}$$

## Example

Plant number	1	2	3	4	5	6	7	8	9	10	11	12
Amount of water	16	16	16	18	18	20	22	24	24	26	26	26
Growth height	76.2	77.1	75.7	78.1	77.8	79.2	80.2	82.5	80.7	83.1	82.2	83.6

►  $\bar{x} = 21, \bar{y} = 79.7, s_{xx} = 188, s_{yy} = 83.54, s_{xy} = 122.4$

► Estimate  $\alpha, \beta$

$$b = s_{xy}/s_{xx} = 0.6511$$

$$a = \bar{y} - \beta\bar{x} = 79.7 - (0.6511)(21) = 66.03$$

► Estimate  $\sigma^2$

$$s_r^2 = \frac{83.54 - (0.6511)^2(188)}{10} = 0.3850$$

► Find a 95% C.I. for  $\beta$ .

$$95\% \text{ C.I. for } b : 0.6511 \pm 2 \frac{\sqrt{0.3850}}{\sqrt{188}} \Rightarrow (0.56, 0.74)$$