

# Stat 135 Lab 13

GSI: Yutong Wang  
May 1, 2020

# Questions?



# To-do Today

1. Interpretation of linear model output in R  
[See the [Rmd demo](#)]
2. SE of the regression line
3. Prediction Interval
4. Bayesian statistics
5. Practice problems



# SE of the regression line

**Note:**  $\text{Cov}(aX + bY, cW + dV) = ac\text{Cov}(X, W) + ad\text{Cov}(X, V) + bc\text{Cov}(Y, W) + bd\text{Cov}(Y, V)$

**Lemma:**  $\text{Cov}(\bar{y}, \hat{\beta}_1) = \text{Cov}(\hat{\beta}_0 + \hat{\beta}_1 \bar{x}, \hat{\beta}_1) = \text{Cov}(\hat{\beta}_0, \hat{\beta}_1) + \bar{x}\text{Var}(\hat{\beta}_1)$

$$= -\frac{\sigma^2 \bar{x}}{n\text{Var}(x)} + \frac{\sigma^2 \bar{x}}{n\text{Var}(x)} = 0$$

$$\hat{y}_i = \bar{y} + \hat{\beta}_1(x_i - \bar{x})$$

$$\begin{aligned}\Rightarrow \text{Var}(\hat{y}_i) &= \text{Var}(\bar{y} + \hat{\beta}_1(x_i - \bar{x})) \\ &= \text{Var}(\bar{y}) + (x_i - \bar{x})^2 \text{Var}(\hat{\beta}_1) + 2(x_i - \bar{x})\text{Cov}(\bar{y}, \hat{\beta}_1) \\ &= \text{Var}(\bar{y}) + (x_i - \bar{x})^2 \text{Var}(\hat{\beta}_1) = \sigma^2 \left( \frac{1}{n} + \frac{(x_i - \bar{x})^2}{n\text{Var}(x)} \right)\end{aligned}$$

$$\Rightarrow \hat{y} \sim N(\beta_0 + \beta_1 x, \sigma^2 \left( \frac{1}{n} + \frac{(x - \bar{x})^2}{n\text{Var}(x)} \right))$$

**Standard error:**  $s_{\hat{y}} = \hat{\sigma}^2 \left( \frac{1}{n} + \frac{(x - \bar{x})^2}{n\text{Var}(x)} \right)$

# Prediction interval

$$\hat{y}_{n+1} = \hat{\beta}_0 + \hat{\beta}_1 x_{n+1}$$

$$\text{Var}(y_{n+1} - \hat{y}_{n+1}) = \text{Var}(y_{n+1}) + \text{Var}(\hat{y}_{n+1})$$

$$= \sigma^2 + \sigma^2 \left( \frac{1}{n} + \frac{(x - \bar{x})^2}{n \text{Var}(x)} \right)$$

$$= \sigma^2 \left( \textcolor{red}{1} + \frac{1}{n} + \frac{(x - \bar{x})^2}{n \text{Var}(x)} \right)$$

$$\implies y_{n+1} - \hat{y}_{n+1} \sim N\left(0, \sigma^2 \left( \textcolor{red}{1} + \frac{1}{n} + \frac{(x - \bar{x})^2}{n \text{Var}(x)} \right) \right)$$

$$\implies \frac{y_{n+1} - \hat{y}_{n+1} - 0}{\hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{n \text{Var}(x)}}} \sim t_{n-2}$$

$$\implies \mathbb{P} \left[ y_{n+1} \in \hat{y}_{n+1} \pm t_{n-2}(0.025) \hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{n \text{Var}(x)}} \right] = 0.95$$

# Bayesian statistics

Bayes Rule:

$$f_{Y|X}(y | x) = \frac{f_{X,Y}(x, y)}{f_X(x)} = \frac{f_{X|Y}(x | y) \cdot f_Y(y)}{f_X(x)}$$

Posterior Probability:

$$f_{\Theta|X}(\theta | x) = \frac{f_{X|\Theta}(x | \theta) f_{\Theta}(\theta)}{f_X(x)}$$

where  $f_{\Theta|X}(\theta | x)$  is the **posterior probability**,  $f_{\Theta}(\theta)$  is the **prior probability**,  $f_{X|\Theta}(x | \theta)$  is the **likelihood density**, and  $f_X(x)$  is a normalizing constant.

$$f_{\Theta|X}(\theta | x) \propto f_{X|\Theta}(x | \theta) f_{\Theta}(\theta)$$

- We treat the parameter  $\theta$  as a random variable in Bayesian statistics, when the frequentists treat the parameter as some fixed constant.
- An estimate of  $\theta$  is called the **posterior mean**  $\hat{\theta} = \mathbb{E}(\Theta | X)$ , which is a function of  $X$ .
- When the prior and posterior distributions belong to the same distribution family, we say that the **prior and likelihood are conjugate**.

# Bayesian statistics

**Example** : Beta distribution  $X \sim \text{Beta}(\gamma, s)$

$$f(x) = \frac{\Gamma(\gamma + s)}{\Gamma(\gamma)\Gamma(s)} x^{\gamma-1} (1-x)^{s-1}, \mathbb{E}(x) = \frac{\gamma}{\gamma + s}$$

Denote the prior as  $p \sim \text{Beta}(\gamma, s)$ , and the likelihood density is  $X | p \sim \text{Binomial}(n, p)$ . Then, the posterior is

$$p | X = x \sim \text{Beta}(x + \gamma, n - x + s)$$

and the posterior mean is

$$\hat{p} = \mathbb{E}(p | X = x) = \frac{x + \gamma}{n + \gamma + s} = \frac{n}{n + \gamma + s} \left( \frac{x}{n} \right) + \frac{\gamma + s}{n + \gamma + s} \left( \frac{\gamma}{\gamma + s} \right)$$

where  $\frac{x}{n}$  is MLE. and  $\frac{\gamma}{\gamma+s}$  is the prior mean.

# Problem 1: Bayesian Statistics

Consider a biased coin with probability of landing heads equal to  $\theta$ . Also, let  $X$  be a random variable that is equal to 1 when the coin lands heads, and is otherwise equal to 0. In other words, we assume that  $P(X = 1|\Theta = \theta) = \theta$ , while  $P(X = 0|\Theta = \theta) = 1 - \theta$ . We consider the following prior for  $\Theta$  :

$$f_{\Theta}(\theta) = \begin{cases} \theta e^{\theta} & \text{for } 0 \leq \theta \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

- (a) Find the probability that a coin toss results in heads: i.e., compute  $P(X = 1)$ .
- (b) Given that a coin toss results in heads, find the posterior density for  $\Theta$ .
- (c) Given that the first toss resulted in heads, find the conditional probability of heads on the second toss. (*Hint*: this is tricky. Use the posterior from part (b) as the new prior!)



# Problem 2: Posterior Mean

Let's assume the same setup from problem 2 above but turn to the Bayesian perspective with the flat prior  $\Theta \sim \text{Unif}(0, 1)$ .

Find the posterior distribution of  $\Theta|X_1, \dots, X_n$ .

Show that the posterior mean  $E(\Theta|X_1, \dots, X_n)$  is a weighted average of  $\hat{\Theta}_{ML}$  from problem 2 and the prior mean.

# Problem 3: MLE, MoM

Suppose that we have  $n$  i.i.d random variables  $X_1, X_2, \dots, X_n$ , each with the following probability function (as usual,  $0 \leq \theta \leq 1$  is unknown):

$$P(X = 0|\theta) = \frac{2}{3}\theta, \quad P(X = 1|\theta) = \frac{1}{3}\theta, \quad P(X = 2|\theta) = \frac{2}{3}(1-\theta), \quad P(X = 3|\theta) = \frac{1}{3}(1-\theta)$$

- (a) Find  $\hat{\theta}_{\text{ML}}$ , i.e. the max likelihood estimate of  $\theta$ . Also find  $E(\hat{\theta}_{\text{ML}})$  and  $\text{Var}(\hat{\theta}_{\text{ML}})$ .
- (b) Find  $\hat{\theta}_{\text{MM}}$ , i.e. the method of moments estimate of  $\theta$  (it will depend only on  $\hat{\mu}_1 = \bar{X}$ )
- (c) Compute the standard error of  $\hat{\theta}_{\text{MM}}$ . Note that it is a function of  $\theta$ .
- (d) How does the standard error of  $\sigma_{\hat{\theta}_{\text{MM}}}$  compare to the standard error of  $\sigma_{\hat{\theta}_{\text{ML}}}$  for different values of  $\theta$ ? *Hint*: It is in fact a bit simpler to compare their squares.
- (e) Finally, assume that we have a sample  $x = (3, 0, 2, 1, 3, 2, 1, 0, 2, 1)$ , compute (i): the MLE  $\hat{\theta}_{\text{ML}}$  (it is a number!) and its *estimated* standard error; and (ii): the MOM  $\hat{\theta}_{\text{MM}}$  and its estimated standard error.