Monday: Project
Wed: HW11 due

15: INFERENCE FOR SLR

Larsen & Marx 11.3

Prof Amanda Luby

Additional Assumptions:

1. $f_{Y|x} \sim N(\mu, \sigma^2) \quad \forall x$
2. σ is constant for all x
3. $\mu = E(Y|x) = \beta_0 + \beta_1 x$
4. All conditional distributions are independent
   $E(Y|4000) \perp E(Y|5000)$

1 Properties of MLEs for Simple Linear Regression

1. $\hat{\beta}_0$ and $\hat{\beta}_1$ are normal RV's
2. $\hat{\beta}_0$ and $\hat{\beta}_1$ are unbiased
3. $V(\hat{\beta}_1) = \frac{\sigma^2}{\sum(x_i - \bar{X})^2}$
4. $V(\hat{\beta}_0) = \frac{\sigma^2 \sum x_i^2}{n\sum(x_i - \bar{X})^2} = \sigma^2[\frac{1}{n} + \frac{\bar{x}^2}{\sum(x_i - \bar{x})^2}]$
5. $\hat{\beta}_1$, $\bar{Y}$ and $\hat{\sigma}^2$ are mutually independent
6. $\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-2}$
7. $S^2 = \frac{n}{n-2}\hat{\sigma}^2$ is an unbiased estimator for σ^2

Proof: ($\hat{\beta}_1$ is a normal RV)

Everything but the Y's are constant

$Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$

$$\hat{\beta}_1 = \frac{n\sum x_i Y_i - (\sum x_i)(\sum Y_i)}{n(\sum x_i^2) - (\sum x_i)^2}$$

$$= \frac{\sum x_i Y_i - \frac{1}{n}\sum x_i (\sum Y_i)}{(\sum x_i^2) - \frac{1}{n}(\sum x_i)^2}$$

$$= \frac{\sum x_i Y_i - \bar{x}\sum Y_i}{(\sum x_i^2) - n\bar{x}^2}$$

a, b_i: constants

$$= \frac{\sum (x_i - \bar{x}) Y_i}{(\sum x_i^2) - n\bar{x}^2} \rightarrow \frac{1}{a} \cdot \sum b_i Y_i \rightarrow \text{linear combination of a bunch of Normal RV's}$$

$\rightarrow \hat{\beta}_1$ is also normally distributed!

Proof: ($V(\hat{\beta}_1)$) $= \frac{\sigma^2}{\sum (x_i - \bar{x})^2}$

$(\sum x_i^2) - n\bar{x}^2 = \sum (x_i - \bar{x})^2$

$\sum var[(x_i - \bar{x}) Y_i]$

since all x_i's, $\bar{x}$ constant

$$\hat{\beta}_1 = \frac{\sum (x_i - \bar{x}) Y_i}{\sum (x_i - \bar{x})^2} \rightarrow var(\hat{\beta}_1) = var\left(\frac{\sum (x_i - \bar{x}) Y_i}{\sum (x_i - \bar{x})^2}\right) = \frac{1}{(\sum (x_i - \bar{x})^2)^2} \sum (x_i - \bar{x})^2 Var(Y_i)$$

$$= \frac{\sum (x_i - \bar{x})^2 \sigma^2}{(\sum (x_i - \bar{x})^2)^2}$$

$$= \frac{\sigma^2}{\sum (x_i - \bar{x})^2}$$

2 Inference for Simple Linear Regression

$H_0: \beta_1 = 0$ ← no relationship between x's + y's

2.1 Inference for β_1

$H_1: \beta_1 \neq 0$

Test statistic for β_1

Let $(x_1, Y_1), (x_2, Y_2), ..., (x_n, Y_n)$ be a set of points satisfying $E(Y|X = x) = \beta_0 + \beta_1 x$ and let $S^2 = \frac{1}{n-2} \sum (Y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2$. Then,

$T = \frac{\hat{\beta}_1 - \beta_1}{S/\sqrt{\sum (x_i - \bar{x})^2}} \sim T_{n-2}$

unbiased estimator of σ →

Proof: $E(\hat{\beta}_1) = \beta_1$ $\quad$ $\text{var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum (x_i - \bar{x})^2}$ $\quad$ $\hat{\beta}_1$ is a normal RV

$$\text{let } Z = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\text{var}(\hat{\beta}_1)}} \sim N(0,1) = \frac{\hat{\beta}_1 - \beta_1}{\sigma / \sqrt{\sum (x_i - \bar{x})^2}}$$

independent

$$\frac{n\hat{\sigma}^2}{\sigma^2} = \frac{(n-2)S^2}{\sigma^2} \sim \chi^2_{n-2}$$

$$\frac{Z}{\sqrt{\chi^2_v / v}} = \frac{Z}{\sqrt{\frac{(n-2)S^2}{\sigma^2} / (n-2)}} = \frac{\frac{\hat{\beta}_1 - \beta_1}{\sigma / \sqrt{\sum (x_i - \bar{x})^2}}}{\sqrt{S^2/\sigma^2}} = \frac{\sigma}{S} \cdot \frac{\hat{\beta}_1 - \beta_1}{\sigma / \sqrt{\sum (x_i - \bar{x})^2}} = \frac{\hat{\beta}_1 - \beta_1}{S / \sqrt{\sum (x_i - \bar{x})^2}} \sim T_{n-2}$$

Note: Hypothesis tests based on T are GLRTs!

↳ for β_0 and β_1

2.2 Inference for σ^2

Recall: $Y_i \sim N(\beta_0 + \beta_1 x, \sigma^2)$

$$\frac{(n-2)s^2}{\sigma^2} \sim \chi^2_{n-2} \implies P\left(\chi^2_{\alpha/2,n-2} \le \frac{(n-2)s^2}{\sigma^2} \le \chi^2_{1-\alpha/2,n-2}\right) = 1-\alpha$$

CI for σ^2:

- reciprocal / flip sign
- multiply by $(n-2)s^2$

$$P\left(\frac{(n-2)s^2}{\chi^2_{1-\frac{\alpha}{2},n-2}} \le \sigma^2 \le \frac{(n-2)s^2}{\chi^2_{\alpha/2,n-2}}\right) = 1-\alpha$$

$$\left[\frac{(n-2)s^2}{\chi^2_{1-\frac{\alpha}{2},n-2}}, \frac{(n-2)s^2}{\chi^2_{\alpha/2,n-2}}\right]$$ is a $1-\alpha$ % CI for σ^2

Hypothesis test:

$H_0: \sigma^2 = \sigma_0^2$

$$\chi^2 = \frac{(n-2)s^2}{\sigma_0^2} \sim \chi^2_{n-2}$$

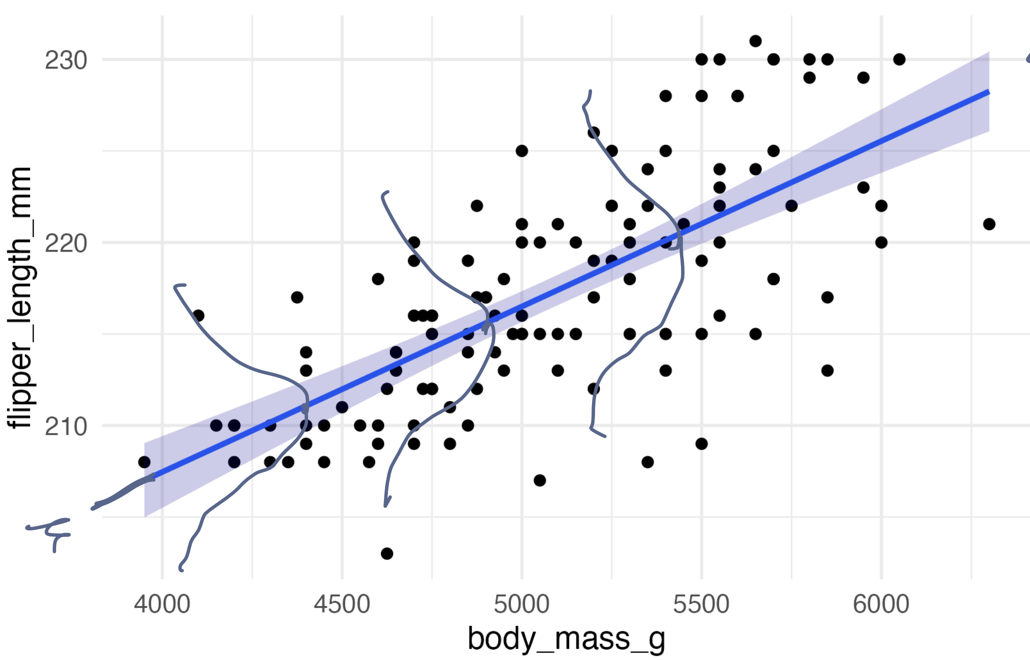
$$Y_i \sim N(\underbrace{\beta_0 + \beta_1 x_i}_{E(Y|x_i)}, \sigma^2)$$

2.3 Inference for $E(Y|x)$

Point estimate: $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ want to find sampling dist. of $\hat{y}_i$ (normal distribution)

$$E(\hat{y}_i) = E(\hat{\beta}_0 + \hat{\beta}_1 x_i) = E(\hat{\beta}_0) + E(\hat{\beta}_1)x_i = \beta_0 + \beta_1 x_i \rightarrow$$ unbiased estimator for $E(Y|x_i)$

$$\begin{aligned} Var(\hat{y}_i) = var(\hat{\beta}_0 + \hat{\beta}_1 x_i) &= var(\bar{Y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 x_i) \\ &= var(\bar{Y} + \hat{\beta}_1 (x_i - \bar{x})) \\ &= var(\bar{Y}) + (x_i - \bar{x})^2 var(\hat{\beta}_1) \\ &= \frac{1}{n}\sigma^2 + (x_i - \bar{x})^2 \cdot \frac{\sigma^2}{\sum (x_i - \bar{x})^2} \\ &= \sigma^2 \left[\frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right] \end{aligned}$$



CI for $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$

2.4 Inference for new Y_i's

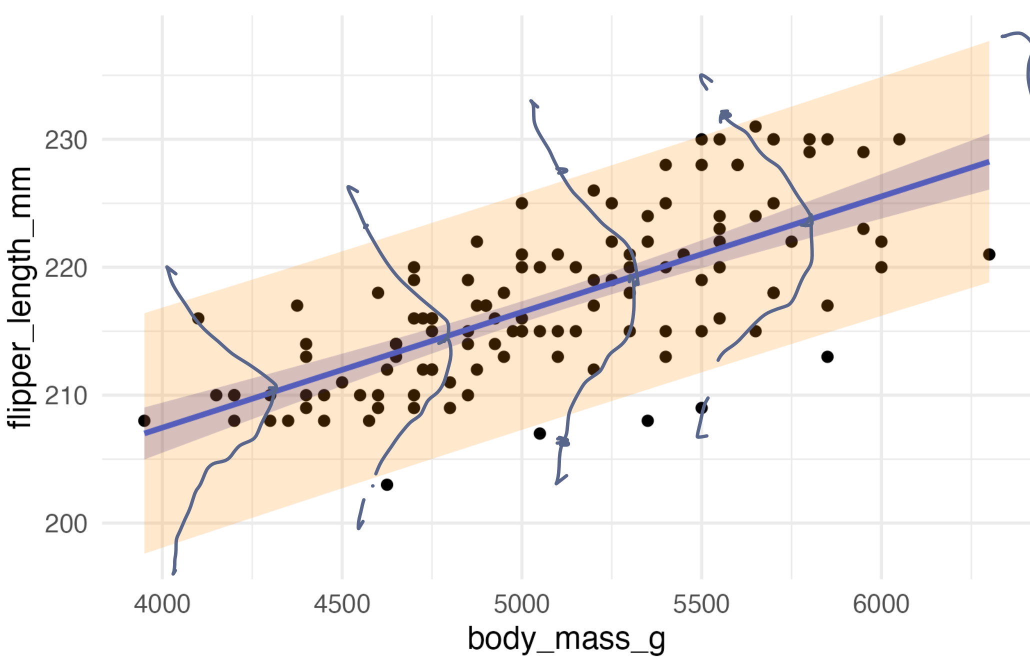
Let (x^*, y^*) be a hypothetical new observation (y^* is independent of Y_i's)

Point estimate for y^* : $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x^*$

let $E = \hat{y}^* - y^*$ ← how far away is prediction from the new observation?

$$E(\hat{y}^* - y^*) = E(\hat{y}^*) - E(y^*) = \beta_0 + \beta_1 x^* - (\beta_0 + \beta_1 x^*) = 0 \leftarrow \text{unbiased}$$

$$\begin{aligned} \mathrm{var}(\hat{y}^* - y^*) &= \mathrm{var}(\hat{y}^*) + \mathrm{var}(y^*) \\ &= \sigma^2 \left[\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right] + \sigma^2 \\ &= \sigma^2 \left[1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right] \end{aligned}$$



95% prediction interval for data points