Jacob Snyder

ACM157 Final

1. $\bar{X} = \frac{1+5}{2} = 3$ $\bar{Y} = 6$ $\rightarrow s_{obs}(X,Y) = |6-3| = 3$

Permute the data: $\{1,5,6\}$

$X = \{1,5\}$ $Y = \{6\}$ or $X = \{5,1\}$ $Y = \{6\}$

$\bar{X} = 3, \bar{Y} = 6, s_1 = s_2 = 3$

$X = \{1,6\}$ or $\{6,1\}$, $Y = \{5\}$

$\bar{X} = 3.5$ $\bar{Y} = 5$ $s_3 = s_4 = 1.5$

$X = \{5,6\}$ or $\{6,5\}$, $Y = \{1\}$

$\bar{X} = 5.5$ $\bar{Y} = 1$ $s_5 = s_6 = 4.5$

p-value $= \frac{1}{6}\sum_{i=1}^{6} I(s_i > s_{obs}) = \frac{1}{6}(2) = \boxed{\frac{1}{3}}$

2. $\hat{p}_A = \frac{X}{n}$ $\hat{p}_B = \frac{Y}{m}$ $\Delta p = p_A - p_B$ $\Delta\hat{p} = \hat{p}_A - \hat{p}_B$

Let us redefine the hypotheses:

$H_0: \Delta p = 0$ vs $H_1: \Delta p \neq 0$

$\Delta\hat{p} = \frac{X}{n} - \frac{Y}{m}$

$se^2[\Delta\hat{p}] = se^2[\hat{p}_A] + se^2[\hat{p}_B] = se^2\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] + se^2\left[\frac{1}{m}\sum_{i=1}^{m} y_i\right]$

$= \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}$

$\Rightarrow \hat{se}^2(\Delta\hat{p}) = \frac{s_1^2}{n} + \frac{s_2^2}{m}$

$\rightarrow$ By the notes (Lecture 12, (15)), the size α Wald test rejects H_0 when

$$\left|\frac{\frac{X}{n} - \frac{Y}{m}}{\sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{m}}}\right| > z_{1-\alpha/2}$$

Jacob Snyder

ACM157 final

3. a) $\mathbb{E}[\hat{\beta} \mid \{X_i\}] = \mathbb{E}\left[\frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2} \middle| \{X_i\}\right]$

$= \frac{\sum_{i=1}^{n} X_i \mathbb{E}[Y_i \mid \{X_i\}]}{\sum_{i=1}^{n} X_i^2} = \frac{\sum_{i=1}^{n} X_i (\beta X_i)}{\sum_{i=1}^{n} X_i^2} = \frac{\beta \sum_{i=1}^{n} (X_i)^2}{\sum_{i=1}^{n} (X_i)^2} = \beta$

therefore $\hat{\beta}$ is unbiased

b) $\mathbb{V}[\hat{\beta} \mid \{X_i\}] = \frac{\sum_{i=1}^{n} (X_i)^2 \mathbb{V}[Y_i \mid \{X_i\}]}{\left(\sum_{i=1}^{n} X_i^2\right)^2} = \frac{\mathbb{V}[Y_i \mid \{X_i\}]}{\sum_{i=1}^{n} X_i^2}$

$= \frac{\sigma^2}{\sum_{i=1}^{n} X_i^2}$

Therefore $se[\hat{\beta}] = \frac{\sigma}{\left[\sum_{i=1}^{n} X_i^2\right]^{1/2}}$

c) Following the notes (Lecture 15b, (14) and on):

$\hat{\hat{\beta}} = \sum_{i=1}^{n} \alpha_i Y_i$

$\beta = \mathbb{E}[\hat{\hat{\beta}} \mid \{X_i\}] = \sum_{i=1}^{n} \alpha_i \mathbb{E}[Y_i \mid \{X_i\}] = \beta \sum_{i=1}^{n} \alpha_i \rightarrow \sum_{i=1}^{n} \alpha_i = 1$

$\mathbb{V}[\hat{\hat{\beta}} \mid \{X_i\}] = \sigma^2 \sum_{i=1}^{n} \alpha_i^2$

$\Rightarrow$ Minimize $\sum_{i=1}^{n} \alpha_i^2$, subject to $\sum_{i=1}^{n} \alpha_i = 1 \rightarrow \boxed{\alpha_i = \frac{1}{n}}$ for all i

$\Longrightarrow \hat{\hat{\beta}} = \frac{1}{n} \sum_{i=1}^{n} Y_i$

4. Let us redefine the hypotheses: $H_0: (\beta_1 - 2019\beta_0) = 0$ vs $H_1: (\beta_1 - 2019\beta_0) \neq 0$

We will use a α size t-test.

Define: $\hat{\beta}_1 = \sum_{i=1}^{n} \frac{X_i - \bar{X}}{S_{XX}} Y_i \quad \mathbb{V}[\hat{\beta}_1] = \frac{\sigma^2}{S_{XX}} \quad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \quad \mathbb{V}[\hat{\beta}_0] = \sigma^2\left(\frac{1}{n} + \frac{X^2}{S_{XX}}\right)$

As defined in the notes, lecture 16a. Thus:

$\mathbb{V}[\hat{\beta}_1 - 2019\hat{\beta}_0] = \mathbb{V}[\hat{\beta}_1] + 2019^2 \mathbb{V}[\hat{\beta}_0] \Rightarrow \hat{se}[\hat{\beta}_1 - 2019\hat{\beta}_0] = \left[\frac{\hat{\sigma}^2}{S_{XX}} + 2019^2 \hat{\sigma}^2 \left(\frac{1}{n} + \frac{X^2}{S_{XX}}\right)\right]^{1/2}$

Therefore, we reject H_0 when

$$\left| \frac{\hat{\beta}_1 - 2019\hat{\beta}_0}{\left[\frac{\hat{\sigma}^2}{S_{XX}} + 2019\hat{\sigma}^2\left(\frac{1}{n} + \frac{X^2}{S_{XX}}\right)\right]^{1/2}} \right| > t_{n-2, 1-\frac{\alpha}{2}}$$

Jacob Snyder

ACM157 final

5. a) Because more information is being used in the creation of the confidence interval when using the prediction $r(X^*)$ as opposed to just Y^*. Specifically, the confidence interval for $r(X^*)$ can take into account the input X^* and consider that the output is likely to closely match the input, whereas the confidence interval for Y^* must consider all possible outputs and thus uses a larger interval.

b) $\mathbb{V}[\hat\beta_1 \mid \{X_i\}] = \dfrac{\sigma^2}{\sum_{i=1}^{n}(X_i-\bar X)^2}$

Therefore → $\underbrace{\text{minimize } \sigma^2}_{\text{All values of } Y \text{ are the same}}$, $\underbrace{\text{maximize } \sum_{i=1}^{n}(X_i-\bar X)^2}_{\text{equally distribute points between } X=-1 \text{ and } X=1\text{, none in between}}$

As an example, you could put 50 points at $(-1,0)$ and 50 points at $(1,0)$

6. $\hat\beta_1 = \sum_{i=1}^{n} \frac{X_i - \bar X}{S_{XX}} Y_i \qquad \hat\beta_0 = \bar Y - \hat\beta_1 \bar X$

→ $\hat\beta_0 + \hat\beta_1 X_i + e_{i1} = a + bX_i + cX_i^2 + e_{i2}$

→ $e_{i1} = (a-\hat\beta_0) + (b-\hat\beta_1)X_i + cX_i^2 + e_{i2}$

→ $\mathbb{E}[e_{i1}] = a - \mathbb{E}[\hat\beta_0] + b\bar X - \mathbb{E}[\hat\beta_1 X_i] + c\,\mathbb{E}[X_i^2]$

$= a - \bar Y + b\bar X + c\,\mathbb{E}[X_i^2]$

$= \mathbb{E}[a + bX_i + cX_i^2 + e_{i2}] - \bar Y$

$= \mathbb{E}[Y_i] - \bar Y$

$= \bar Y - \bar Y$

$= \boxed{0}$

Scanned with CamScanner