Kyle McGraw

ACM 157 Final

1.

$X_1 = 2 \quad X_2 = 5 \quad Y_1 = 7 \qquad H_0: \mu_X = \mu_Y \quad H_1: \mu_X \neq \mu_Y$

$$\text{p-value} = \mathbb{P}_0(s > s_{obs}) = \frac{1}{N!}\sum_{i=1}^{N!} I(s_i > s_{obs})$$

$$s_{obs} = s(X_1, X_2; Y_1) = |\bar{X}_n - \bar{Y}_m| = |3.5 - 7| = 3.5$$

$$N = n + m = 2 + 1 = 3$$

$$z_\pi = (2, 5; 7) \qquad s_\pi = |3.5 - 7| = 3.5$$

$$z_\pi = (2, 7; 5) \qquad s_\pi = |4.5 - 5| = 0.5$$

$$z_\pi = (5, 2; 7) \qquad s_\pi = |3.5 - 7| = 3.5$$

$$z_\pi = (5, 7; 2) \qquad s_\pi = |6 - 2| = 4$$

$$z_\pi = (7, 2; 5) \qquad s_\pi = |4.5 - 5| = 0.5$$

$$z_\pi = (7, 5; 2) \qquad s_\pi = |6 - 2| = 4$$

$$\frac{1}{N!}\sum_{i=1}^{N!} I(s_i > s_{obs}) = \frac{1}{6}\sum_{i=1}^{6} I(s_i > s_{obs}) = \frac{1}{6}\big(I(3.5 > 3.5) + I(0.5 > 3.5) + I(3.5 > 3.5) + I(4 > 3.5) + I(0.5 > 3.5) + I(4 > 3.5)\big)$$

$$= \frac{1}{6}\cdot(0 + 0 + 0 + 1 + 0 + 1)$$

$$= 1/3$$

2.

$A \; n \; X \qquad B \; m \; Y$

$X \sim Bin(n, p_A) \quad Y \sim Bin(n, p_B) \qquad H_0: p_A = p_B \quad H_1: p_A \neq p_B \qquad \alpha$

$$W = \left|\frac{\hat{\theta} - \theta_0}{\hat{se}}\right| > z_{1-\frac{\alpha}{2}}$$

$$\hat{p}_A = \frac{X}{n} \qquad \hat{p}_B = \frac{Y}{n}$$

$$\hat{\sigma}_A^2 = \frac{1}{n}\hat{p}_A(1 - \hat{p}_A) = \frac{X}{n^2}\left(1 - \frac{X}{n}\right) \qquad \hat{\sigma}_B^2 = \frac{1}{m}\hat{p}_B(1 - \hat{p}_B) = \frac{Y}{m^2}\left(1 - \frac{Y}{m}\right)$$

$$\hat{se}^2 = \hat{se}^2(p_A) + \hat{se}^2(p_B) = \frac{\hat{\sigma}_A^2}{n} + \frac{\hat{\sigma}_B^2}{m} = \frac{X}{n^3}\left(1 - \frac{X}{n}\right) + \frac{Y}{m^3}\left(1 - \frac{Y}{m}\right)$$

$$w = \left| \frac{\hat{P}_A - \hat{P}_B}{\sqrt{\frac{X}{n^3}(1-\frac{X}{n}) + \frac{Y}{m^3}(1-\frac{Y}{m})}} \right| > z_{1-\frac{\alpha}{2}}$$

$$= \left| \frac{\frac{X}{n} - \frac{Y}{m}}{\sqrt{\frac{X}{n^3}(1-\frac{X}{n}) + \frac{Y}{m^3}(1-\frac{Y}{m})}} \right| > z_{1-\frac{\alpha}{2}}$$

3.

$Y_i = \beta X_i + e_i$ $\qquad \hat{\beta} = \frac{\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i^2}$

a)

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

$\mathbb{E}[e_i] = 0$

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

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

$$\text{bias}[\hat{\beta}] = \mathbb{E}(\hat{\beta}) - \beta = \beta - \beta = 0$$

unbiased

b)

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

$$= \frac{1}{(\sum_{i=1}^n X_i^2)^2} \sum_{i=1}^n \mathbb{V}[X_i Y_i] = \frac{1}{(\sum_{i=1}^n X_i^2)^2} \sum_{i=1}^n X_i^2 \sigma^2 = \sigma^2 \frac{\sum_{i=1}^n X_i^2}{(\sum_{i=1}^n X_i^2)^2} = \frac{\sigma^2}{\sum_{i=1}^n X_i^2}$$

$$se[\hat{\beta}] = \sqrt{\frac{\mathbb{V}[\hat{\beta}]}{n}} = \sqrt{\frac{\sigma^2}{n\sum_{i=1}^n X_i^2}}$$

c)

$\hat{\beta}$ is a consistent estimate of β if as we get more and more data $\hat{\beta}$ becomes more accurate and goes to β

Since $\hat{\beta}$ is unbiased, we just need:

$$\lim_{n\to\infty} \mathbb{V}[\hat{\beta}] = 0$$

$$\lim_{n\to\infty} \mathbb{V}[\hat{\beta}] = \lim_{n\to\infty} \frac{\sigma^2}{\sum_{i=1}^n X_i^2} = 0$$

This is true as long as there are only a finite number of $x_i = 0$.

4.

$(x_1, Y_1), \ldots, (X_n, Y_n)$ $\quad H_0: \beta_1 = 2022\beta_0 \quad H_1: \beta_1 \neq 2022\beta_0$

$$\left| \frac{\hat{\beta}_1 - 2022\beta_0}{\hat{\sigma}/\sqrt{S_{xx}}} \right| > t_{n-2, 1-\frac{\alpha}{2}} \qquad \hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$$

$$Y = \beta_0 + \beta_1 x + e$$
$$Y = \beta_0 + 2022\beta_0 x + e$$
$$\beta_0 = \frac{Y - e}{1 + 2022x}$$

$$\hat{\sigma}^2 = V[\hat{\beta}_1] = 2022^2 V[\beta_0]$$
$$= 2022^2 \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)$$

$$\left| \frac{\frac{S_{xy}}{S_{xx}} - 2022\frac{Y-e}{1+2022}}{2022\sigma\sqrt{\frac{1}{nS_{xx}} + \frac{\bar{x}^2}{S_{xx}^2}}} \right| > t_{n-2, 1-\frac{\alpha}{2}}$$

5.

a)

$Y = \beta_0 + \beta_1 X + e \qquad 100(1-\alpha)\% \qquad r(X^*) = \beta_0 + \beta_1 X^*$

The confidence interval of the optimal prediction for the response Y^* to new input X^* is smaller than the prediction interval for Y^* because of the additional source of uncertainty from the random statistical error e^*.

$$Y^* = r(X^*) + e^*$$

b)

$n = 200 \quad X_1, \ldots, X_n \quad [-1, 1] \quad \hat{\beta}_1$

$$V[\hat{\beta}_1] = \frac{\sigma^2}{S_{xx}} \qquad S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

To minimize $V[\hat{\beta}_1]$, we want to maximize S_{xx}. To maximize S_{xx}, we want to maximize $|x_i - \bar{x}|$ or put all the points as far as possible from the mean. This means we put 100 points at -1 and 100 points at 1.

6.

$(X_1, Y_1), \ldots, (X_n, Y_n) \quad Y_i = a + bX_i + cX_i^2 + e_i \qquad e_i \quad \mu = 0 \quad \sigma^2$

$$\mathbb{E}[e_i] = \mathbb{E}[Y_i - \hat{Y}_i] = \mathbb{E}[a + bX_i + cX_i^2 - \hat{\beta}_0 - \hat{\beta}_1 X_i]$$

$$= a + bx_i + cx_i^2 - \mathbb{E}[\hat{\beta}_0] - \mathbb{E}[\hat{\beta}_1] x_i$$

$$= a + bx_i + cx_i^2 - \mathbb{E}[\bar{Y}] + \mathbb{E}\left[\frac{S_{xy}}{S_{xx}} \bar{x} - \frac{S_{xy}}{S_{xx}} x_i\right]$$

$$= (a + bx_i + cx_i^2) - (a + bx_i + cx_i^2) + \frac{S_{xy}}{S_{xx}} \mathbb{E}[\bar{x} - x_i]$$

$$= 0$$

7.

Done