# Homework 1

Cindy Lu

February 9, 2021

## 1 1

### 1.1 a

$n = 100, \bar{Y} = 58, s_Y = 8$. Hence the variance of sample average (using that $Y_i$ are iid):

$$Var(\bar{Y}) = Var\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(Y_i) = \frac{1}{100}Var(Y_i) = \frac{1}{100}8^2$$

Thus $\sigma_{\bar{Y}} = \sqrt{Var(\bar{Y})} = 0.8$, and

$$95\%\text{CI:}(58 - 1.96 \times 0.8, 58 + 1.96 \times 0.8) = (56.43, 59.57)$$

### 1.2 b

$\bar{Y_1} - \bar{Y_2} = 58 - 62 = -4, S_{Y_2} = 11$. Need the standard deviation of $\bar{Y_1} - \bar{Y_2}$ (according to random sampling):

$$Var(\bar{Y_1} - \bar{Y_2}) = Var(\bar{Y_1}) + Var(\bar{Y_2}) = \frac{8^2}{100} + \frac{1}{200^2}(200)Var(Y_{i,2}) = \frac{8^2}{100} + \frac{11^2}{200}$$

$$\implies \sigma_{\bar{Y_1}-\bar{Y_2}} = \sqrt{\frac{8^2}{100} + \frac{11^2}{200}} = 1.1158$$

Construct 90% interval: $(-4 \pm 1.64\sigma_{\bar{Y_1}-\bar{Y_2}}) = (-5.83, -2.17)$.

### 1.3 c

Compute the t-statistic and decide to reject if based on 10, 5, and 1% significance levels.

$$t = \frac{-4 - 0}{1.1158} = -3.5849$$

$$|t| = 3.5849 > 2.58$$

Comparing the absolute value of the t-statistic with critical value at 1% (which is 2.58), we see that it is greater in magnitude. Hence reject $H_0 : \bar{Y_1} - \bar{Y_2} = 0$ at all significance levels (1, 5, and 10%). Next, the p-value will lend great evidence to the alternate hypothesis, $H_A = |\bar{Y_1} - \bar{Y_2}| \neq 0$.

$$2\Phi\left(-\frac{4 - 0}{1.1158}\right) = 2\Phi(-|t|) = 2\Phi(-3.5849) \approx 0.00034$$

Since $0.00034 < 0.01$, it is strong evidence against the null, and it is rejected at 1% significance level. Hence the population means are statistically significantly different from zero, concluded with high confidence.

# 2 2

$H_0 : \bar{Y_1} - \bar{Y_2} = 0$, and the alternate $H_A = |\bar{Y_1} - \bar{Y_2}| = 0$. Require standard deviation of difference:

$$Var(\bar{Y_1} - \bar{Y_2}) = Var(\bar{Y_1}) + Var(\bar{Y_2}) = \frac{1}{100^2}(100)Var(Y_{i1}) + \frac{1}{64^2}Var(Y_{i2}) = \frac{1}{100}(200^2) + \frac{1}{64}320^2$$

$$\implies SE_{\bar{Y_1}-\bar{Y_2}} = \sqrt{\frac{200^2}{100} + \frac{320^2}{64}} \approx 44.721$$

Construct t-statistic:

$$t = \frac{(3100 - 2900) - 0}{44.721} \approx 4.4722 > 2.58$$

This t-statistic was positive and it's easily seen that it is greater than 2.58, the critical value at 1% significance level. Next compute the p-value:

$$2\Phi(-|t|) = 2\Phi(-4.4722) \approx 2(3.8744 \times 10^{-6} \approx 7.7488 \times 10^{-6}$$

## 2.1 b

There is overwhelming evidence that the firm has a large gender-wage discrimination. The p-value was far smaller than 0.01, and the t-statistic was much larger than 2.58. Therefore, by these statistical tests, it strongly supports the alternative hypothesis against the null: there is a statistically significant difference in the mean wages of men and women, and the difference between the means is significantly different from zero.

# 3 3

## 3.1 a

Need standard deviation of the mean:

$$Var(\bar{Y}) = \frac{1}{420^2}(420)Var(Y_i) = \frac{1}{420}s_Y^2 = \frac{1}{420}19.1^2$$

$$\implies s_{\bar{Y}} = \frac{19.1}{\sqrt{420}} \approx 0.9319845$$

Confidence interval: $(654.2 - 1.96 * 0.9319845, 654.2 + 1.96 * 0.9319845) = (652.3733, 656.0266)$.

## 3.2 b

Just as in part 2c, state null and alternative hypotheses, compute t-statistic and p-value associated with difference in means.

$H_0 : \bar{Y_1} - \bar{Y_2} = 0$.

$H_A : |\bar{Y_1} - \bar{Y_2}| \neq 0$.

Require the standard deviation / standard error of difference in means:

$$Var(\bar{Y_1} - \bar{Y_2}) = \frac{1}{238^2}(238)Var(Y_{i1}) + \frac{1}{182^2}(182)Var(Y_{i2}) = \frac{1}{238}(19.4^2) + \frac{1}{182}(17.9^2)$$

$$\implies s_{\bar{Y_1}-\bar{Y_2}} = \sqrt{\frac{1}{238}(19.4^2) + \frac{1}{182}(17.9^2)}$$

The t-statistics and p-values are as follows:

$$t = \frac{(657.4 - 650) - 0}{\sqrt{\frac{19.4^2}{238} + \frac{17.9^2}{182}}} \approx 4.0479 > 2.58$$

$$2\Phi(-|t|) = 2\Phi(-4.0479) \approx 2 \times 0.00002584 = 0.00005168 < 0.01$$

There is statistically significant evidence that districts with smaller classes have higher average test scores. Based on the t and p-values, there is evidence against the null and supporting the alternate hypothesis. The t-statistic was much greater in magnitude than the critical value of 2.58 at 1% significance level, while the p-value was much smaller than the significance level of 1% and is highly significant. Hence the difference between the two means is statistically significantly different from zero.