Jasmeen Mann
500653379

Assignment 2

3.3)a) $\hat{p} = \frac{215}{400} = 0.5375$

b) Estimated variance of $\hat{p}$ is

$$\text{Var}(\hat{p}) = \frac{\hat{p}(1-\hat{p})}{n} = \frac{0.5375 \times (1-0.5375)}{400} = 6.2148 \times 10^{-04}$$

Standard error $\Rightarrow SE(\hat{p}) = (\text{Var}(\hat{p}))^{\frac{1}{2}} = (6.2148 \times 10^{-04})^{\frac{1}{2}}$

$= 0.0249$

c) t-statistic

$$t^{act} = \frac{\hat{p} - \mu_{p,0}}{SE(\hat{p})} = \frac{0.5375 - 0.5}{0.0249} = 1.506$$

p-value for the test $\Rightarrow H_0 : p = 0.5$

$H_1 : p \neq 0.5$

$p\text{-value} = 2\,\Phi(-|t^{act}|) = 2\,\Phi(-1.506) = 2 \times 0.066 = 0.132$

d) P-value for the test $\Rightarrow H_0 : p = 0.5$

$H_1 : p > 0.5$

$p\text{-value} = 1 - \Phi(t^{act}) = 1 - \Phi(1.506) = 1 - 0.934 = 0.066$

e) The 2 results differ because part c is a 2 sided test and the p-value is the area in the tails of the standard normal distribution. While part d is a one sided test and the p-value is the area under the standard normal distribution to the right of the calculated t-statistic.

F) The survey did not contain statistically significant evidence that the incumbent was ahead of the challenges at the time of the survey. This is because from the second test $\Rightarrow H_0 : p = 0.5$ and $H_1 : p > 0.5$ the null hypothesis at the 5% significance level cannot be rejected. The p-value 0.066 is larger than 0.05. Also, the t-statistic 1.506 is less than the critical value 1.64 for a one-sided test with 5% significance level.

3.4)a) 95% confidence interval:

$\hat{p} \pm 1.96\ SE(\hat{p}) = 0.5375 \pm 1.96 \times 0.0249 = (0.4887, 0.5863)$

b) 99% confidence interval:

$\hat{p} \pm 2.57\ SE(\hat{p}) = 0.5375 \pm 2.57 \times 0.0249 = (0.4735, 0.6015)$

c) The interval for 'b' is wider than the interval in "a" because

d) The null hypothesis at a 5% significance level cannot be rejected because 0.5 lies inside the 95% confidence interval for p.. of a higher critical value due to a lower significance level

3.12) Standard error $\Rightarrow \bar{Y}_1 - \bar{Y}_2 \Rightarrow SE(\bar{Y}_1 - \bar{Y}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n}}$

$$= \sqrt{\frac{(200)^2}{100} + \frac{(320)^2}{64}} = 44.721$$

a) Hypothesis test $\Rightarrow$ difference in mean monthly salaries

$H_0 : \mu_1 - \mu_2 = 0$; $H_1 : \mu_1 - \mu_2 \neq 0$

- t-Statistic $\Rightarrow$

$$t^{act} = \frac{\bar{Y}_1 - \bar{Y}_2}{SE(\bar{Y}_1 - \bar{Y}_2)} = \frac{3100 - 2900}{44.721} = 4.4722$$

- p-value

$$= 2F(-|t^{act}|) = 2F(-4.4722) = 2 \times (3.8744 \times 10^{-06}) = 7.7488 \times 10^{-06}$$

↳ Since the p-value's are extremely low this suggests that the difference in the monthly salaries for men and women is statistically significant The null hypothesis can be rejected with a high degree of confidence.

b) Part 'a' suggests a large difference between the mean earnings for males and females. TO see if this is gender discrimination or not we have to take the one sided alternative test:

$H_0 : \mu_1 - \mu_2 = 0$

$H_1 : \mu_1 - \mu_2 > 0$

- t-Statistic $t^{act} = 4.4722$
- P-value $= 1 - F(t^{act}) = 1 - F(4.4722) = 1 - 0.999996126 = 3.874 \times 10^{-06}$

-> Since the p-value is small the null hypothesis can be rejected. Even though there is a difference in the mean earnings of both genders it does not suggest gender discrimination by the firm. Gender discrimination means that 2 workers that are the same in every way except gender are paid different wages. This study may not have controlled characteristics of the workers that could affect their productivity such as; education experience etc. Therefore, we cant say there is discrimination for wages since the characteristics are not controlled for the statistical analysis.

3.16) a) $1013 \pm 1.96\,SE$ $SE = \frac{108}{\sqrt{453}} = 5.07$ Interval $\Rightarrow (1003.06, 1022.94)$

b) Since the national average 1000 is not included in the 95% confidence for Florida. We can reject at the 5% confidence level the hypothesis that the Florida students perform the same as the other students in the United States.

c) i) The 95% confidence interval:

$$\bar{Y}_{prep} - \bar{Y}_{nonprep} \pm 1.96\,SE(\bar{Y}_{prep} - \bar{Y}_{nonprep}) \quad \text{where:}$$

$$SE(\bar{Y}_{prep} - \bar{Y}_{nonprep}) = \sqrt{\frac{95^2}{503} + \frac{108^2}{453}} = 6.61$$

95% confidence interval $\Rightarrow (1019 - 1013) \pm 12.95$ or 6 ± 12.96

ii) There is not any statistically significant evidence that suggests that the prep course helped. The 95% confidence interval includes $\mu_{prep} - \mu_{nonprep} = 0$

d) i) Let X denote the change in the test scores.

95% Confidence interval μ_X is $\Rightarrow \bar{x} \pm 1.96\,SE(\bar{x})$

where $SE(\bar{x}) = \frac{60}{\sqrt{453}} = 2.82$

∴ the confidence interval is 9 ± 5.52

ii) Yes there is statistically significant evidence that students will perform better on their second attempt after taking the prep course. The 95% confidence interval does not include $\mu_X = 0$

iii) They should randomly select students that have only take the test 1 time. Randomly select half of these students and have them take a prep course. Make them all take the test again and compare the performance of the 2 sets of test takers

4.1) a) $\widehat{testscore} = 520.4 - 5.82 \times 22 = 392.36$

b) $\widehat{\Delta testscore} = (-5.82 \times 19) - (-5.82 \times 23) = 23.28$

c) $\widehat{Testscore} = \hat{B}_0 + \hat{B}_1 \times \bar{CS} = 520.4 - 5.82 \times 21.4 = 395.85$

d) $SSR = (n-2)SER^2 = (100-2) \times (11.5)^2 = 12961$

$$TSS = \frac{SSR}{1-R^2} = \frac{12961}{1-(0.08)^2} = 13044$$

$$s^2_y = \frac{TSS}{n-1} = \frac{13044}{99} = 131.8 \Rightarrow sd \Rightarrow s_y = \sqrt{s^2_y} = \sqrt{131.8} = 11.48$$

4.6) $E(\mu_i | x_i) = 0$

$E(Y_i | x_i) = E(\beta_0 + \beta_1 x_i + \mu_i | x_i) = \beta_0 + \beta_1 E(x_i | x_i) + E(\mu_i | x_i) = \beta_0 + \beta_1 x_i$

4.7) $E(\hat{\beta}_0) = E(\bar{Y} - \hat{\beta}_1 \bar{x}) = E\left[\left(\beta_0 + \beta_1 \bar{x} + \frac{1}{n}\sum_{i=1}^{n} \mu_i\right) - \hat{\beta}_1 \bar{x}\right]$

$= \beta_0 + E(\beta_1 - \hat{\beta}_1)\bar{x} + \frac{1}{n}\sum_{i=1}^{n} E(\mu_i | x_i) = \beta_0$

4.8) mean of $\hat{\beta}_0 \Rightarrow \beta_0 + 2$

$Y_i = (\beta_0 + 2) + \beta_1 x_i + (\mu_i - 2)$

new regression error $\Rightarrow (\mu_i - 2)$ new intercept $(\beta_0 + 2)$

4.9a) $\hat{\beta}_1 = 0$, $\hat{\beta}_0 = \bar{Y}$ and $\bar{Y}_i = \hat{\beta}_0 = \bar{Y}$ thus $ESS = 0$ and $R^2 = 0$

b) If $R^2 = 0$ then $ESS = 0 \Rightarrow \hat{Y}_i = \bar{Y}$ for all i

But $\Rightarrow \hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ so that $\hat{Y}_i = \bar{Y}$ for all i, which implies that $\hat{\beta}_1 = 0$ or x is constant for all i.

If x is constant $\Rightarrow \sum_{i=1}^{n} (x_i - \bar{x})^2 = 0$ and $\hat{\beta}_1$ is undefined

4.11a) Least squares objective function $\Rightarrow \sum_{i=1}^{n} (Y_i - b_1 x_i)^2$

$\frac{\partial \sum_{i=1}^{n} (Y_i - b_1 x_i)^2}{\partial b_1} = -2 \sum_{i=1}^{n} x_i (Y_i - b_1 x_i)$

$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i Y_i}{\sum_{i=1}^{n} x_i^2}$

b) $\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i (Y_i - \mu)}{\sum_{i=1}^{n} Y_i^2}$