Stat330 HW6 Neha Maddali

1a) $4.2 \pm 1.645 \sqrt{49}/\sqrt{30}$

4.2 ± 2.10234

90% CI: (2.09766, 6.30234)

2b) $4.2 \pm 1.96 \sqrt{49}/\sqrt{30}$

4.2 ± 2.50492

95% CI: (1.69508, 6.70492)

1c) as you increase the confidence interval, the width of the interval also increases

1d) $4.2 \pm 1.645 \sqrt{49}/\sqrt{100}$

4.2 ± 1.1515

90% CI: (3.0485, 5.3515)

1e) as you increase the sample size the width of the confidence interval decreases

2a) mean: 69.75

$\sigma^2: (50-69.75)^2 + (68-69.75)^2 + (85-69.75)^2 + (76-69.75)^2 = 166.188$

$69.75 \pm 1.64 \sqrt{166.188}/\sqrt{4}$

69.75 ± 10.5709

90% CI: (59.1791, 80.3209)

2b) $SD = \sqrt{166.188} = 12.8914$

$69.75 \pm 2.35 \cdot 12.8914/\sqrt{4}$

69.75 ± 15.1474

90% CI: (54.6026, 84.8974)

2c) The t-distribution CI is more conservative (wider) than the normal distribution CI

2d) The data does not provide significant evidence against the hypothesis that the average salary of computer engineers equals \$80,000. So the average salary is different than \$80,000

3a) $37.7 \pm 1.64 \cdot 9.2/\sqrt{100}$

37.7 ± 1.5088

90% CI: (36.19, 39.21)

3b) $H_0: \mu \leq 35$

$H_a: \mu > 35$ $\quad z_{0.99} = 2.33$

$\alpha = 0.01$ $\quad z = \frac{37.7 - 35}{9.2/10} = 2.93$

The value $2.93 > 2.33$ so the null hypothesis is rejected. This means that the population mean of users that are connected at the same time is greater than 35

4a) $n = 200$

$\hat{p}$ = sample prop of defective items = $24/200 = 0.12$

$0.12 \pm 1.96 \cdot \sqrt{0.12(1-0.12)/200}$

0.12 ± 0.045037

95% CI: $(0.074963, 1.65037)$

4b) No, if only 10% of the items were defective, that means there were 20 defective items not 24

4c) $\hat{p} \leq 0.02$

$z = 1.96$ $\quad n \geq \left(0.5 \cdot 1.96 / 0.02\right)^2$

$\Delta = 0.02$ $\quad$ = sample size for random inspection should be 2401 items

5) $H_0: p \leq 0.02$ $\quad n = 200$ $\quad$ 10 failed / 200 tested

$H_a: p > 0.02$ $\quad x = 10$ $\quad = 0.05$

$$z = \frac{0.05 - 0.02}{\sqrt{\frac{0.02(1-0.02)}{200}}} = 3.03046$$

p-value: 0.001223

The result is significant at $p < 0.01$. This verifies the engineers claim for the quality assurance specialist

6a) $(30-20) \pm 2.576 \sqrt{1^2/30 + 2^2/20}$

$10 \pm 2.576 \sqrt{1/30 + 4/30}$

10 ± 1.24433

99% CI: $(8.75567, 11.2443)$

6b) We are 99% confident that the difference between the average number of days required for delivery between website B and A (B-A) is between 8.75 (≈9) and 11.24 (≈11) days

6c) No, not confident therefore need to conduct hypothesis test to check the significance of p-value

6d) $H_0: \mu_B - \mu_A \Rightarrow \mu_B - \mu_A = 0$

$H_a: \mu_B \neq \mu_A$

$$z = \frac{(5.2-4.5)-0}{\sqrt{1/20 + 4/20}} = 1.44914$$

p-value = 0.073655 ⟹ not significant p-value at $p \le 0.01$

So there is not a significant difference between the delivery speeds of the 2 websites

7) $H_0: P_B = P_A$

$H_a: P_B > P_A$

state A: 210 respondents approve

State B: 225 respondents approve

$$\hat{p}_c = \frac{225 + 210}{500 + 600} = 0.395$$

$$z = \frac{225/500 - 210/600}{\sqrt{\frac{0.395 \cdot 0.605}{500} + \frac{0.395 \cdot 0.605}{600}}}$$

$= 3.37826$

p-value = 0.000365 ⟹ significant at $p \le 0.01$

The approval rates differ in the 2 states as the p-value result is significant

8a) $0.4239/0.0368 = 11.519$ $\quad 4969.631/70 - 11.519(123.886/70) = 50.6084$

using x_2: $\hat{y} = 50.6084 + 11.519 x_2$

8b) plug x_1 and x_2 into their models

RMSE for first model: 1.1181

RMSE for second model: 8.2772

First model appears to be better at predicting y as the RMSE is a lower value than for the second model

9a) $\frac{87.51}{84.39} = 1.037$

$65.22 - (1.037)(68.35) = -5.65895$

$\hat{y} = 1.037 - 5.65895x$

9b) $H_0 : B_1 = 0$

$H_a : B_1 \neq 0$

$z = \frac{68.35 - 0}{\sqrt{0(1-0)/78}} =$