STAT/MATH 415 HW#6

October 27, 2017

EXERCISES

7.6.3 For the data given in 6.5.3, with the usual assumptions,

a) Find a 95% confidence interval for $\mu(x)$ when $x = 68, 75$ and 82

Answer: The linear regression line is $\hat{y} = 86.8 + 1.0157(x - 74.5)$ and the 95% CI for $\mu(x)$ is

$$\hat{\alpha} + \hat{\beta}(x - \bar{x}) \pm t_{\alpha/2(n-2)}\sqrt{\frac{n}{n-2}\sigma^2}\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{\sum(x_i - \bar{x})^2}}$$

$$= 86.8 + 1.0157(x - 74.5) \pm t_{0.025,8}\sqrt{\frac{10}{8}18}\sqrt{\frac{1}{10} + \frac{(x-74.5)^2}{414.5}}$$

$$= 1.0157x + 11.13 \pm 2.306 \times 4.74 \times \sqrt{\frac{1}{10} + \frac{(x-74.5)^2}{414.5}}$$

Since $x_1 = 68,\ x_2 = 75,\ x_3 = 82$, we can have their 95% CI as

$$\mu(x_1) = 80.20 \pm 10.93 \times 0.4494 = 80.20 \pm 4.91 = [75.29, 85.11]$$

$$\mu(x_2) = 87.31 \pm 10.93 \times 0.3172 = 87.31 \pm 3.47 = [83.84, 90.78]$$

$$\mu(x_3) = 94.41 \pm 10.93 \times 0.4855 = 94.41 \pm 5.31 = [89.10, 99.72]$$

b) Find a 95% prediction interval for Y when $x = 68, 75$ and 82

Answer: The 95% CI for Y_i is

$$\hat{\alpha} + \hat{\beta}(x - \bar{x}) \pm t_{\alpha/2(n-2)}\sqrt{\frac{n}{n-2}\sigma^2}\sqrt{1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{\sum(x_i - \bar{x})^2}}$$

$$= 86.8 + 1.0157(x - 74.5) \pm t_{0.025,8}\sqrt{\frac{10}{8}18}\sqrt{\frac{11}{10} + \frac{(x-74.5)^2}{414.5}}$$

$$= 1.0157x + 11.13 \pm 2.306 \times 4.74 \times \sqrt{\frac{11}{10} + \frac{(x-74.5)^2}{414.5}}$$

Since $x_1 = 68,\ x_2 = 75,\ x_3 = 82$, we can have their 95% CI as

$$y_1 = 80.20 \pm 10.93 \times 1.0963 = 80.20 \pm 11.98 = [68.22, 92.18]$$

$$y_2 = 87.31 \pm 10.93 \times 1.0491 = 87.31 \pm 11.47 = [75.84, 98.78]$$

$$y_3 = 94.41 \pm 10.93 \times 1.1116 = 94.41 \pm 12.15 = [82.26, 106.56]$$

7.6.7 For the ACT scores in 6.5.6, with the usual assumptions,

a) Find a 95% confidence interval for $\mu(x)$ when $x = 17, 20, 23, 26$ and 29

Answer: First we need to find point estimates for α, β, σ^2

$$\hat{\alpha} = \bar{y} = 26.33 \qquad \hat{\beta} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum(x_i - \bar{x})^2} = \frac{9292 - 9111.33}{356.93} = 0.5062$$

Thus the regression line is $\hat{y} = 26.33 + 0.5062(x - 23.07)$

$$\hat{\sigma^2} = \frac{1}{15}\sum \varepsilon_i^2 = \frac{\sum(y_i - \hat{y}_i)^2}{15} = \frac{211.886}{15} = 14.126$$

The 95% CI for $\mu(x)$ is

$$\hat{\alpha} + \hat{\beta}(x - \bar{x}) \pm t_{\alpha/2(n-2)}\sqrt{\frac{n}{n-2}\sigma^2}\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{\sum(x_i - \bar{x})^2}}$$

$$= 26.33 + 0.5062(x - 23.07) \pm t_{0.025,13}\sqrt{\frac{15}{13}14.126}\sqrt{\frac{1}{15} + \frac{(x - 23.07)^2}{356.93}}$$

$$= 0.5062x + 14.66 \pm 2.160 \times 4.037 \times \sqrt{\frac{1}{15} + \frac{(x - 23.07)^2}{356.93}}$$

For $x_1 = 17,\ x_2 = 20,\ x_3 = 23,\ x_4 = 26,\ x_5 = 29$ their $\mu(x)$ CIs are

$$\mu(x_1) = 23.26 \pm 8.720 \times 0.412 = 23.26 \pm 3.59 = [19.67, 26.85]$$

$$\mu(x_2) = 24.78 \pm 8.720 \times 0.305 = 24.78 \pm 2.66 = [22.12, 27.44]$$

$$\mu(x_3) = 26.30 \pm 8.720 \times 0.258 = 26.30 \pm 2.25 = [24.05, 28.55]$$

$$\mu(x_4) = 27.82 \pm 8.720 \times 0.301 = 27.82 \pm 2.63 = [25.19, 30.45]$$

$$\mu(x_5) = 29.34 \pm 8.720 \times 0.407 = 29.34 \pm 3.55 = [25.79, 32.89]$$

b) Determine a 90% prediction interval for Y when $x = 17, 20, 23, 26$ and 29

Answer: The 90% CI for Y_i is

$$\hat{\alpha} + \hat{\beta}(x - \bar{x}) \pm t_{\alpha/2(n-2)}\sqrt{\frac{n}{n-2}\sigma^2}\sqrt{1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{\sum(x_i - \bar{x})^2}}$$

$$= 26.33 + 0.5062(x - 23.07) \pm t_{0.05,13}\sqrt{\frac{15}{13}14.126}\sqrt{1 + \frac{1}{15} + \frac{(x - 23.07)^2}{356.93}}$$

$$= 0.5062x + 14.66 \pm 1.771 \times 4.037 \times \sqrt{\frac{16}{15} + \frac{(x - 23.07)^2}{356.93}}$$

For $x_1 = 17,\ x_2 = 20,\ x_3 = 23,\ x_4 = 26,\ x_5 = 29$ their Y_i CIs are

$$y_1 = 23.26 \pm 7.150 \times 1.082 = 23.26 \pm 7.73 = [15.53, 30.99]$$

$$y_2 = 24.78 \pm 7.150 \times 1.045 = 24.78 \pm 7.48 = [17.30, 32.26]$$

$$y_3 = 26.30 \pm 7.150 \times 1.033 = 26.30 \pm 7.38 = [18.92, 33.68]$$

$$y_4 = 27.82 \pm 7.150 \times 1.044 = 27.82 \pm 7.47 = [20.35, 35.29]$$

$$y_5 = 29.34 \pm 7.150 \times 1.079 = 29.34 \pm 7.72 = [21.62, 37.06]$$

8.3.1 Let Y be Binom(100, p). To test $H_0 : p = 0.08$ against $H_1 : p < 0.08$, we reject H_0 and accept H_1 if and only if $Y \le 6$.

a) Determine the significance level α of the test.

Answer: $\alpha = P(\text{reject } H_0 \mid H_0 \text{ is true}) = P(Y \le 6 \mid p = 0.08)$

$$\alpha = P(Y \le 6 \mid p = 0.08) = P(Y = 0, 1, ..., 6 \mid p = 0.08) = \sum_{y=0}^{6} \binom{100}{y} 0.08^y 0.92^{100-y} = 0.3032$$

b) Find the probability of the Type II error if, in fact, p=0.04.

Answer: $\beta = P(\text{accept } H_0 \mid H_0 \text{ is false}) = P(Y > 6 \mid p = 0.04)$

$$\beta = P(Y > 6 \mid p = 0.04) = 1 - P(Y = 0, 1, ..., 6 \mid p = 0.04) = 1 - \sum_{y=0}^{6} \binom{100}{y} 0.04^y 0.96^{100-y} = 0.1064$$

8.3.3 Let Y be Binom(192, p). We reject $H_0 : p = 0.75$ and accept $H_1 : p > 0.75$ if and only if $Y \ge 152$. Use the normal approximation to determine

a) $\alpha = P(Y \ge 152;\ p = 0.75)$

Answer: Since $n = 192,\ p = 0.75,\ np = 144 \ge 5,\ n(1-p) = 48 \ge 5$, thus $Y \sim N(144, 36)$ approximately.

$$\alpha = P(Y \ge 152) = P(Y > 151.5) = P(Z > \frac{151.5 - 144}{6}) = P(Z > 1.25) = 0.1056$$

b) $\beta = P(Y < 152)$ when $p = 0.80$

Answer: Similarly, $np = 153.6 \ge 5,\ n(1-p) = 38.4 \ge 5$ thus $Y \sim N(153.6, 30.72)$

$$\beta = P(Y < 152) = P(Y \le 151.5) = P(Z \le \frac{151.5 - 153.6}{\sqrt{30.72}}) = P(Z \le -0.38) = 0.3520$$

8.3.7 The management of the Tigers baseball team decided to sell only low-alcohol beer in their ballpark to help combat rowdy fan conduct. They claimed that more than 40% of the fans would approve of this decision. Let p equal the proportion of Tigers fans on opening day who approved of the decision. We shall test the null hypothesis $H_0 : p = 0.40$ against the alternative hypothesis: $H_1 : p > 0.40$

a) Define a critical region that has an $\alpha = 0.05$ significance level.

Answer: Since this is one-tail, the critical value should be $Z_{0.05} = 1.645$, so the critical region is $[1.645, +\infty)$

b) If, out of a random sample of $n = 1278$ fans, $y = 550$ said that they approved of the new policy, what is your conclusion?

Answer:

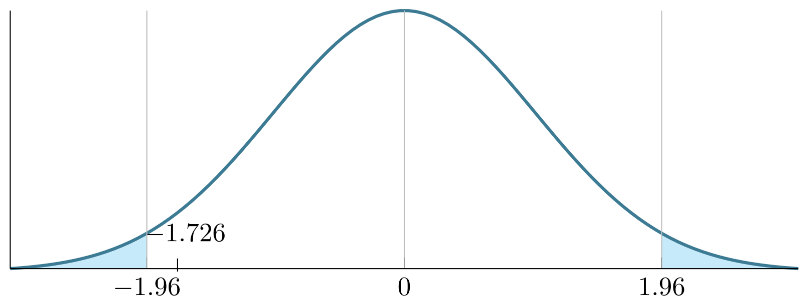
$$\hat{p} = \frac{y}{n} = 0.43036$$
$$Z = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} = \frac{0.43036 - 0.40}{\sqrt{0.4 \times 0.6/1278}} = 2.215$$

Since $2.215 > 1.645$, Z-statistic is in critical region, so reject H_0, that is $p > 0.40$

8.3.11 A machine shop that manufactures toggle levers has both a day and a night shift. A toggle lever is defective if a standard nut cannot be screwed onto the threads. Let p_1 and p_2 be the proportion of defective levers among those manufactured by the day and night shifts repectively. We shall test the null hypothesis $H_0 : p_1 = p_2$ against a two-sided alternative hypothesis based on two random samples, each of 1000 levers taken from the production of the respective shifts.

a) Define the test statistic and a critical region that has an $\alpha = 0.05$ significance level. Sketch a standard normal pdf illustrating this critical region.

Answer: Since it's two-sided, the critical value is $Z_{0.025} = 1.96$, the critical region is $(-\infty, -1.96] \cup [1.96, +\infty)$ as shaded in the picture.



b) If $y_1 = 37$ and $y_2 = 53$ defectives were observed for the day and hight shifts respectively, calculate the value of the test statistic. Locate the calculated test statistic on your figure in (a) and state your conclusion.

Answer:

$$\hat{p_1} = \frac{y_1}{n_1} = 0.037 \qquad \hat{p_2} = \frac{y_2}{n_2} = 0.053 \qquad \hat{p_0} = \frac{y_1 + y_2}{n_1 + n_2} = 0.045$$

$$Z = \frac{\hat{p_1} - \hat{p_2}}{\sqrt{p_0(1-p_0)(\frac{1}{n_1} + \frac{1}{n_2})}} = \frac{-0.016}{\sqrt{0.045 \times 0.955 \times 0.002}} = -1.726$$

The test statistic is not in the critical region, so we do not reject H_0.