STA 570 — Spring 2012 — Dr. Charnigo

Written Assignment 4 Solutions

1a. We have $\hat{p}_1 = 48/127 = 0.3780$, $\hat{p}_2 = 26/73 = 0.3562$, and $\hat{p} = 74/200 = 0.3700$. The test statistic is

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(1/127 + 1/73)}} = \frac{0.0218}{0.0709} = 0.31,$$

which does not exceed 1.96 in absolute value. So, we accept H_0 at level 0.05. [The p-value is 0.76.]

1b. We have $E_{11} = 127 \times 74/200 = 46.99$, $E_{21} = 27.01$, $E_{12} = 80.01$, and $E_{22} = 45.99$. The test statistic is

$$\chi^2 = \frac{(48 - 46.99)^2}{46.99} + \frac{(26 - 27.01)^2}{27.01} + \frac{(79 - 80.01)^2}{80.01} + \frac{(47 - 45.99)^2}{45.99} = 0.09,$$

which does not exceed $3.84 = \chi^2_{1,0.95}$. So, we accept H_0 at level 0.05. [The p-value is 0.76.]

2a. [Exhibit the box plots.] The distribution of perceived pain reduction is somewhat negatively skewed among those receiving the new medication: the bottom whisker of the box plot is noticeably longer than the top whisker. The distribution of perceived pain reduction is even more negatively skewed among those receiving the existing medication: the bottom whisker of the box plot is noticeably longer than the top whisker, the bottom half of the box is noticeably longer than the top half, and there are two outlying observations on the low side. In addition, although not perceptible from the box plots, there is some digit preference within both groups: multiples of 5 or 10 appear disproportionately often.

2b. There are $n = 99$ nonzero difference scores. Using Equation 9.1 we determine that more than $59.8 = 99/2 + 1/2 + 1.96\sqrt{99/4}$ or fewer than $39.2 = 99/2 - 1/2 - 1.96\sqrt{99/4}$ positive difference scores would permit the conclusion of a nonzero median difference in perceived pain reduction between the two groups. Since 45 of the nonzero difference scores are positive, we are not entitled to that conclusion. [The p-value is 0.42 per SAS, which uses a slightly different approach than Equation 9.1.]

2c. The sum of the ranks for the positive difference scores is 2754.5. If the median difference in perceived pain reduction between the two groups were zero, then we would have expected a rank sum of $2475 = 99(100)/4$ for the positive difference scores. Using Equation 9.5, and noting the presence of 14 two-way ties, 4 three-way ties, 1 four-way tie, 3 five-way ties, 1 six-way tie, and 1 seven-way tie, we obtain

$$z = \frac{|2754.5 - 2475| - 1/2}{\sqrt{99(100)(199)/24 - 1146/48}} = 0.97.$$

Since 0.97 does not exceed 1.96 in absolute value, we may not conclude that the median difference in perceived pain reduction between the two groups is nonzero. [The p-value is 0.33 per SAS, which uses a slightly different approach than Equation 9.5.]

2d. The test statistic is

$$t = \frac{-3.27}{35.89/\sqrt{100}} = -0.91.$$

Since -0.91 does not exceed $1.984 = t_{99,0.975}$ in absolute value, we may not conclude that the mean difference in perceived pain reduction between the two groups is nonzero. [The p-value is 0.36.]

2e. Although the distribution of perceived pain reduction is somewhat negatively skewed among those receiving the new medication as well as among those receiving the existing medication, a box plot suggests that the distribution of difference scores is close to normal – apart from the digit preference noted in part a. So, conducting a paired t-test as in part d is permissible.

3a. The sum of ranks is 4052 for the smokers (aged 12 and older) and 11173 for the non-smokers (aged 12 and older). If the median FEV for smokers were equal to that for non-smokers, then we would have expected a rank sum of $50(50 + 124 + 1)/2 = 4375$ for the smokers and a rank sum of $124(50 + 124 + 1)/2 = 10850$ for the non-smokers. The absolute value of the difference between 4052 and 4375, or between 11173 and 10850, is 323. Using Equation 9.7, and noting the presence of several ties (1 four-way, 4 two-way), we obtain

$$z = \frac{323 - 1/2}{\sqrt{\left(\frac{50(124)}{12}\right)\left[175 - \frac{84}{174(173)}\right]}} = 1.07.$$

Since the test statistic does not exceed 1.96 in absolute value, we do not conclude that median FEV differs between the two groups. [The p-value is 0.28.]

3b. After logarithmic transformation the sample means are 1.236 for the non-smokers and 1.184 for the smokers. The sample variances are 0.0494 and 0.0528, yielding a pooled variance estimate of

$$\frac{0.0494 \times 123 + 0.0528 \times 49}{172} = 0.0504.$$

So, the test statistic is

$$t = \frac{1.236 - 1.184}{\sqrt{0.0504(1/124 + 1/50)}} = 1.38.$$

Since the test statistic does not exceed $1.974 = t_{172,0.975}$ in absolute value, we do not conclude that mean log-transformed FEV differs between the two groups. [The p-value is 0.17.]

3c. There should be 202 smokers (aged 12 and older) and 604 – or 606, if we want to have exactly 3×202 – non-smokers (aged 12 and older) since

$$\frac{(0.0528 + 0.0494/3)(1.960 + 0.842)^2}{(1.184 - 1.236)^2} = 201.1$$

and

$$3\frac{(0.0528 + 0.0494/3)(1.960 + 0.842)^2}{(1.184 - 1.236)^2} = 603.4.$$