STA 580 — Spring 2011 — Dr. Charnigo

Written Assignment 4 Solutions

1a. We have $\hat{p}_1 = 13/72 = 0.1806$, $\hat{p}_2 = 7/48 = 0.1458$, and $\hat{p} = 20/120 = 0.1667$. The test statistic is

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(1/72 + 1/48)}} = \frac{0.0348}{0.0695} = 0.50,$$

which does not exceed 1.96 in absolute value. So, we accept H_0 at level 0.05. [The p-value is 0.617.]

1b. We have $E_{11} = 72 \times 20/120 = 12$, $E_{21} = 8$, $E_{12} = 60$, and $E_{22} = 40$. The test statistic is

$$\chi^2 = \frac{(13-12)^2}{12} + \frac{(59-60)^2}{60} + \frac{(7-8)^2}{8} + \frac{(41-40)^2}{40} = 0.25,$$

which does not exceed $3.84 = \chi^2_{1,0.95}$. So, we accept H_0 at level 0.05. [The p-value is 0.617.]

2a. [Exhibit the box plots.] The distribution of effusion durations is somewhat positively skewed among breast fed babies. The sample mean is greater than the sample median, the top whisker of the box plot is longer than the bottom whisker, and there are two outlying observations on the high side. The distribution of effusion durations is very positively skewed among bottle fed babies. The sample mean is greater than not only the sample median but also the sample third quartile, the top whisker of the box plot is considerably longer than the bottom whisker, and there are three (extreme) outlying observations on the high side.

2b. There are $n = 23$ nonzero difference scores. Using Equation 9.1 we determine that more than $16.7 = 23/2 + 1/2 + 1.96\sqrt{23/4}$ or fewer than $6.3 = 23/2 - 1/2 - 1.96\sqrt{23/4}$ positive difference scores would permit the conclusion of a nonzero median difference between effusion duration among breast fed babies and effusion duration among bottle fed babies. Since 7 of the nonzero difference scores are positive, we are not entitled to that conclusion. [The p-value is 0.093 per SAS, which uses a slightly different approach than Equation 9.1.]

2c. The sum of the ranks for the positive difference scores is 61. If the median difference between effusion duration among breast fed babies and effusion duration among bottle fed babies were zero, then we would have expected a rank sum of $138 = 23(24)/4$ for the positive difference scores. Using Equation 9.5, and noting the presence of two two-way ties, one three-way tie, and one four-way tie, we obtain

$$z = \frac{|61 - 138| - 1/2}{\sqrt{23(24)(47)/24 - 96/48}} = 2.33.$$

Since 2.33 exceeds 1.96, we may conclude that the median difference between effusion duration among breast fed babies and effusion duration among bottle fed babies is nonzero. [The p-value is 0.015 per SAS, which uses a slightly different approach than Equation 9.5.]

2d. The test statistic is

$$t = \frac{-26.04}{55.09/\sqrt{24}} = -2.32.$$

Since -2.32 is less than $-2.07 = -t_{23,0.975}$, we may conclude that the mean difference between effusion duration among breast fed babies and effusion duration among bottle fed babies is nonzero. [The p-value is 0.030.]

2e. The sign test, which uses the least amount of information from the data, does not distinguish between the breast fed and bottle fed babies. The paired t-test, which uses the greatest amount of information from the data, does distinguish between the breast fed and bottle fed babies. The signed rank test, which uses an intermediate amount of information from the data, also distinguishes between the breast fed and bottle fed babies.

The difference scores do not appear close to normally distributed, so we are not comfortable using the paired t-test. As such, the signed rank test is the best choice since it uses more information than the sign test and hence gives us a better chance at rejecting a false null hypothesis.

3a. The sum of ranks is 515.5 for the breast fed babies and 660.5 for the bottle fed babies. If the median effusion duration for breast fed babies were equal to that for bottle fed babies, then we would have expected a rank sum of $24(24+24+1)/2 = 588$ for the breast fed babies and a rank sum of $24(24+24+1)/2 = 588$ for the bottle fed babies. The absolute value of the difference between 515.5 and 588, or between 660.5 and 588, is 72.5. Using Equation 9.7, and noting the presence of several ties (one six-way, two three-way, five two-way), we obtain

$$z = \frac{72.5 - 1/2}{\sqrt{\left(\frac{24(24)}{12}\right)\left[49 - \frac{288}{48(47)}\right]}} = 1.49.$$

Since the test statistic does not exceed 1.96 in absolute value, we do not conclude that median effusion duration differs between the two groups of babies. [The p-value is 0.144.]

3b. After logarithmic transformation the sample means are 2.834 for the breast fed babies and 3.306 for the bottle fed babies. The sample variances are 0.588 and 1.098, yielding a pooled variance estimate of

$$\frac{0.588 \times 23 + 1.098 \times 23}{46} = 0.843.$$

So, the test statistic is

$$t = \frac{2.834 - 3.306}{\sqrt{0.843(1/24 + 1/24)}} = -1.78.$$

Since the test statistic does not exceed $2.01 = t_{46,0.975}$ in absolute value, we do not conclude that mean log-transformed effusion duration differs between the two groups of babies. [The p-value is 0.082.]

3c. Each sample should have size 80 since

$$\frac{(0.588 + 1.098)(1.960 + 1.282)^2}{(2.834 - 3.306)^2} = 79.5$$

rounds up to 80.