STA 580 — Spring 2009 — Dr. Charnigo

Written Assignment 3 Solutions

1a. Using notation as in Lecture 7, the test statistic is

$$\chi^2 = \frac{(n_1-1)s_x^2}{225} = \frac{(73)137.55}{225} = 44.63.$$

The critical value is

$$\chi^2_{n_1-1,\alpha} = \chi^2_{73,0.05} = 54.33.$$

Since $44.63 < 54.33$ we reject $H_0 : \sigma_1^2 = 225$ in favor of $H_1 : \sigma_1^2 < 225$.

1b. We have

$$s^2 = \frac{(n_1-1)s_x^2 + (n_2-1)s_y^2}{n_1+n_2-2} = \frac{(73)137.55 + (125)120.49}{198} = 126.78.$$

The test statistic is

$$t = \frac{\bar{x}-\bar{y}}{\sqrt{s^2(1/n_1+1/n_2)}} = \frac{74.35-69.44}{\sqrt{126.78(1/74+1/126)}} = \frac{4.91}{1.649} = 2.978.$$

The (upper) critical value is

$$t_{n_1+n_2-2,1-\alpha/2} = t_{198,0.975} = 1.972.$$

Since $2.978 > 1.972$ we reject $H_0 : \mu_1 = \mu_2$ in favor of $H_1 : \mu_1 \neq \mu_2$.

1c. We have

$$df = \frac{(s_x^2/n_1 + s_y^2/n_2)^2}{(s_x^2/n_1)^2/(n_1-1) + (s_y^2/n_2)^2/(n_2-1)} = \frac{(137.55/74 + 120.49/126)^2}{(137.55/74)^2/(73) + (120.49/126)^2/(125)} = \frac{7.925}{0.0546} \approx 145.$$

The test statistic is

$$t = \frac{\bar{x}-\bar{y}}{\sqrt{s_x^2/n_1 + s_y^2/n_2}} = \frac{74.35-69.44}{\sqrt{137.55/74 + 120.49/126}} = \frac{4.91}{1.678} = 2.926.$$

The (upper) critical value is

$$t_{df,1-\alpha/2} \approx t_{145,0.975} = 1.976.$$

Since $2.926 > 1.976$ we reject $H_0 : \mu_1 = \mu_2$ in favor of $H_1 : \mu_1 \neq \mu_2$.

1d. The test statistic is

$$f = \frac{s_x^2}{s_y^2} = \frac{137.55}{120.49} = 1.142.$$

The upper critical value is

$$f_{n_1-1,n_2-1,1-\alpha/2} = f_{73,125,0.975} = 1.491,$$

while the lower critical value is

$$f_{n_1-1,n_2-1,\alpha/2} = f_{73,125,0.025} = 0.656.$$

(This lower critical value can also be written as $1/1.525 = 1/f_{125,73,0.975} = 1/f_{n_2-1,n_1-1,1-\alpha/2}$.) Since $0.656 < 1.142 < 1.491$ we accept $H_0 : \sigma_1^2 = \sigma_2^2$. Given that we have accepted $H_0 : \sigma_1^2 = \sigma_2^2$, most people would say that the test in part b is more appropriate than the test in part c.

1e. For now our best data analysis option is to perform the large sample ("z") test on page 5 of Lecture 7. (In fact, unless the distribution of diastolic blood pressure is extremely right skewed among either diabetics or non-diabetics, this is a better data analysis option than any to be suggested in Lectures 8 and 9.) The test statistic is

$$z = \frac{\bar{x} - \bar{y}}{\sqrt{s_x^2/n_1 + s_y^2/n_2}} = \frac{74.35 - 69.44}{\sqrt{137.55/74 + 120.49/126}} = \frac{4.91}{1.678} = 2.926.$$

The (upper) critical value is

$$z_{1-\alpha/2} = 1.960.$$

Since $2.926 > 1.960$ we reject $H_0 : \mu_1 = \mu_2$ in favor of $H_1 : \mu_1 \neq \mu_2$.

2a. Using notation as in Lecture 6, the test statistic is

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} = \frac{22/74 - 0.25}{\sqrt{0.25(0.75)/74}} = \frac{0.0473}{0.0503} = 0.940.$$

The critical value is

$$z_{1-\alpha} = z_{0.95} = 1.645.$$

Since $0.940 < 1.645$ we cannot reject $H_0 : p = 0.25$ in favor of $H_1 : p > 0.25$.

2b. The power with a sample size of 150 diabetics is

$$\Phi\left[\sqrt{\frac{p_0(1-p_0)}{p_1(1-p_1)}}\left(-z_{1-\alpha} + \frac{\sqrt{n}|p_0 - p_1|}{\sqrt{p_0(1-p_0)}}\right)\right] = \Phi\left[\sqrt{\frac{0.25(0.75)}{22/74(52/74)}}\left(-1.645 + \frac{\sqrt{150}|0.25 - 22/74|}{\sqrt{0.25(0.75)}}\right)\right] =$$

$$\Phi\left[0.9474(-0.3072)\right] = \Phi[-0.2910] = 0.386.$$

2c. For 80% power the sample size (number of diabetics) is

$$\frac{p_0(1-p_0)\left(z_{1-\alpha} + z_{1-\beta}\sqrt{\frac{p_1(1-p_1)}{p_0(1-p_0)}}\right)^2}{(p_1 - p_0)^2} = \frac{0.25(0.75)\left(1.645 + 0.842\sqrt{\frac{22/74(52/74)}{0.25(0.75)}}\right)^2}{(22/74 - 0.25)^2} \approx 539.$$

For 90% power the sample size (number of diabetics) is

$$\frac{0.25(0.75)\left(1.645 + 1.282\sqrt{\frac{22/74(52/74)}{0.25(0.75)}}\right)^2}{(22/74 - 0.25)^2} \approx 754.$$

2d. There were 74 diabetics in the original sample whose overall size was 200. This is 37%. If we increased the overall sample size by 100, we would expect $100 \times 37\% = 37$ of these to be diabetics.

2e. For 80% power we would need $539 - 74 = 465$ more diabetics. If we increased the overall sample size by 1257, we would expect $1257 \times 37\% \approx 465$ to be diabetics. (One obtains the 1257 not by trial and error but rather by computing $465/0.37$.) For 90% power we would need $754 - 74 = 680$ more diabetics. If we increased the overall sample size by 1838, we would expect $1838 \times 37\% \approx 680$ to be diabetics.