Math 324 Statistics Assignment 3

Jonathan Pearce, 260672004

February 24, 2017

Problem 8a. Let X_1 and X_2 be binomial distributions that represent the 15 patients in stage 1 and stage 2 respectively.

$$\alpha = P(\text{stage 1, reject } H_0 | H_0 \text{ true}) + P(\text{stage 2, reject } H_0 | H_0 \text{ true})$$

$$= P(X_1 \geq 4) + P(X_1 + X_2 \geq 6, X_1 \leq 3) = P(X_1 \geq 4) + \sum_{i=0}^{3} P(X_1 + X_2 \geq 6, X_1 = i)$$

$$= P(X_1 \geq 4) + \sum_{i=0}^{3} P(X_2 \geq 6 - i, X_1 = i) = P(X_1 \geq 4) + \sum_{i=0}^{3} P(X_2 \geq 6 - i) P(X_1 = i)$$

Utilizing Table 1 in Appendix 3 for the following calculations ($p = 0.1$).

$$P(X_1 \geq 4) = [1 - P(X_1 \leq 3)\Big] = 0.0560$$

$$\sum_{i=0}^{3} P(X_2 \geq 6 - i) P(X_1 = i) = \Big[1 - P(X_2 \leq 5)\Big] P(X_1 = 0) + \Big[1 - P(X_2 \leq 4)\Big] P(X_1 = 1) +$$

$$\Big[1 - P(X_2 \leq 3)\Big] P(X_1 = 2) + \Big[1 - P(X_2 \leq 2)\Big] P(X_1 = 3)$$

$$= 0.000412 + 0.00446 + 0.0150 + 0.0236 = 0.0434$$

Therefore,

$$\alpha = 0.0560 + 0.0434 = 0.0994$$

Problem 8b. Carrying out the same calculations as in part a but with $p = 0.3$

$$P(X_1 \geq 4) = 0.703$$

$$\sum_{i=0}^{3} P(X_2 \geq 6 - i) P(X_1 = i) = \Big[1 - P(X_2 \leq 5)\Big] P(X_1 = 0) + \Big[1 - P(X_2 \leq 4)\Big] P(X_1 = 1) +$$

$$\Big[1 - P(X_2 \leq 3)\Big] P(X_1 = 2) + \Big[1 - P(X_2 \leq 2)\Big] P(X_1 = 3)$$

$$= 0.00139 + 0.0146 + 0.0647 + 0.148 = 0.229$$

Therefore,

$$\alpha = 0.703 + 0.229 = 0.932$$

Problem 8c.

$$\beta = P(\text{do not reject } H_0 | p = 0.3)$$

$$= P(X_1 \leq 3, X_1 + X_2 \leq 5) = \sum_{i=0}^{3} P(X_1 = i, X_1 + X_2 \leq 5) = \sum_{i=0}^{3} P(X_1 = i) P(X_2 \leq 5 - i)$$

$$= P(X_1 = 0)P(X_2 \leq 5) + P(X_1 = 1)P(X_2 \leq 4) + P(X_1 = 2)P(X_2 \leq 3) + P(X_1 = 3)P(X_2 \leq 2)$$

$$= 0.00361 + 0.01545 + 0.02732 + 0.02159$$

$$\beta = 0.0670$$

Problem 20.

$$H_0 : \mu = 64$$

$$H_a : \mu < 64$$

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \approx \frac{62 - 64}{\frac{8}{\sqrt{50}}} = -1.768$$

Consider $z_{0.01} = 2.33$, it follows

$$z > -z_{0.01}$$

Therefore we do not reject H_0, there is not enough evidence to refute the manufacturers claim.

Problem 21.

$$H_0 : \mu_1 - \mu_2 = 0$$

$$H_a : \mu_1 - \mu_2 \neq 0$$

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \approx \frac{(1.65 - 1.43) - 0}{\sqrt{\frac{0.26^2}{30} + \frac{0.22^2}{35}}} = 3.648$$

Consider $z_{0.005} = 2.575$, it follows

$$| \ z \ | > z_{0.005}$$

Therefore we reject H_0, there is sufficient evidence that the shear strengths of the two types of soils is different.

Problem 23a.

$$H_0 : \mu_1 - \mu_2 = 0$$

$$H_a : \mu_1 - \mu_2 \neq 0$$

Problem 23b. This is a two tailed test because it considers both cases that $\mu_1 > \mu_2$ or $\mu_1 < \mu_2$.

Problem 23c.

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \approx \frac{(2980 - 3205) - 0}{\sqrt{\frac{1140^2}{40} + \frac{963^2}{40}}} = -0.954$$

Consider $z_{0.05} = 1.645$, it follows

$$| z | < z_{0.005}$$

Therefore we do not reject H_0, there is not sufficient evidence that the mean distances for the two geographical locations is different.

Problem 49. Let u be a 95% upper confidence bound for the average voltage reading

$$u = \bar{x} + z_{0.05}\frac{\sigma}{\sqrt{n}} = 128.6 + 1.645\frac{2.1}{\sqrt{40}} = 129.146$$

Problem 49a. $\mu = 130$ is greater than this upper bound

Problem 49b. Yes, since 130 is greater than the upper bound we reject H_0 and accept the alternative hypothesis.

Problem 49c. No, the result is the same as in question 10.19.

Problem 50.

$$H_0 : \mu = 0.6$$

$$H_a : \mu < 0.6$$

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \approx \frac{0.58 - 0.60}{\frac{.11}{\sqrt{120}}} = -1.992$$

$$\implies \text{p-value} = 0.0233$$

It follows,

$$\text{p-value} < \alpha = 0.10$$

Therefore we reject H_0, there is enough evidence that the flight is unprofitable.

Problem 53a.

$$H_0 : \mu_1 = 3.8$$

$$H_a : \mu_1 < 3.8$$

$$z = \frac{\bar{x} - \mu_1}{\frac{\sigma}{\sqrt{n}}} \approx \frac{3.6 - 3.8}{\frac{1.1}{\sqrt{30}}} = -0.996$$

$$\implies \text{p-value} = 0.1599$$

Problem 53b.

$$0.1599 > \alpha = 0.05$$

Therefore we do not reject H_0.

Problem 53c.

$$H_0 : \mu_2 = 3.1$$

$$H_a : \mu_2 < 3.1$$

$$z = \frac{\bar{x} - \mu_2}{\frac{\sigma}{\sqrt{n}}} \approx \frac{2.7 - 3.1}{\frac{1.2}{\sqrt{30}}} = -1.826$$

$$\implies \text{p-value} = 0.0339$$

Problem 53d.

$$0.0339 < \alpha = 0.05$$

Therefore we reject H_0.

Problem 83a.

$$H_a : \sigma_1^2 = \sigma_2^2$$

The manager of the dairy plant has no preconceived notion of which machine is better, therefore they simply want to see if there is a difference between the two. This is why the manager would want to use a two tail test

Problem 83b.

$$H_a : \sigma_1^2 < \sigma_2^2$$

Salesperson A wants this one sided test so that if H_0 is rejected the manager will buy their machine because it has a lower variance.

Problem 83c.

$$H_a : \sigma_1^2 > \sigma_2^2$$

Salesperson B wants this one sided test so that if H_0 is rejected the manager will buy their machine because it has a lower variance.

Problem 95a.

$$L(\theta) = f(y_1|\theta)...f(y_4|\theta) = \frac{1}{2\theta^3} y_1^2 e^{-\frac{y_1}{\theta}} \cdot ... \cdot \frac{1}{2\theta^3} y_4^2 e^{-\frac{y_4}{\theta}} = \left(\frac{1}{2\theta^3}\right)^4 (y_1 y_2 y_3 y_4)^2 e^{-\frac{1}{\theta}\sum_{i=1}^4 y_i}$$

Therefore

$$\frac{L(\theta_0)}{L(\theta_a)} = \frac{\left(\frac{1}{2\theta_0^3}\right)^4 (y_1 y_2 y_3 y_4)^2 e^{-\frac{1}{\theta_0}\sum_{i=1}^4 y_i}}{\left(\frac{1}{2\theta_a^3}\right)^4 (y_1 y_2 y_3 y_4)^2 e^{-\frac{1}{\theta_a}\sum_{i=1}^4 y_i}} = \left(\frac{\theta_a}{\theta_0}\right)^{12} e^{-\left(\frac{1}{\theta_0} - \frac{1}{\theta_a}\right)\sum_{i=1}^4 y_i} < k$$

$$e^{-\left(\frac{1}{\theta_0} - \frac{1}{\theta_a}\right)\sum_{i=1}^4 y_i} < k\left(\frac{\theta_0}{\theta_a}\right)^{12} \Leftrightarrow -\left(\frac{1}{\theta_0} - \frac{1}{\theta_a}\right)\sum_{i=1}^4 y_i < ln\left[k\left(\frac{\theta_0}{\theta_a}\right)^{12}\right]$$

$$\sum_{i=1}^4 y_i > ln\left[k\left(\frac{\theta_0}{\theta_a}\right)^{12}\right]\left(\frac{1}{\theta_0} - \frac{1}{\theta_a}\right)^{-1}$$

It follows that $\frac{2\sum_{i=1}^{4} y_i}{\theta_0}$ is a chi-square distribution with 24 degrees of freedom. Therefore,

$$\frac{2\sum_{i=1}^{4} y_i}{\theta_0} > \frac{2ln\left[k\left(\frac{\theta_0}{\theta_a}\right)^{12}\right](\frac{1}{\theta_0}-\frac{1}{\theta_a})^{-1}}{\theta_0}$$

Let $k' := \frac{2ln\left[k\left(\frac{\theta_0}{\theta_a}\right)^{12}\right](\frac{1}{\theta_0}-\frac{1}{\theta_a})^{-1}}{\theta_0}$. k' can be selected from the chi square distribution such that the test has significance level of α. The rejection region will be,

$$\left\{\frac{2\sum_{i=1}^{4} y_i}{\theta_0} > k'\right\}$$

Problem 95b. Yes it is the uniformly most powerful test.

Problem 97a.

$$\theta^2 + 2\theta(1-\theta) + (1-\theta)^2 = 1$$

$$N_1 + N_2 + N_3 = n$$

Therefore $L(\theta)$ is the probability mass function of multinomial with 3 classes,

$$L(\theta) = \frac{n!}{n_1!\, n_2!\, n_3!}\theta^{2n_1}[2\theta(1-\theta)]^{n_2}(1-\theta)^{2n_3}$$

Problem 97b.

$$\frac{L(\theta_0)}{L(\theta_a)} = \frac{\frac{n!}{n_1!n_2!n_3!}\theta_0^{2n_1}[2\theta_0(1-\theta_0)]^{n_2}(1-\theta_0)^{2n_3}}{\frac{n!}{n_1!n_2!n_3!}\theta_a^{2n_1}[2\theta_a(1-\theta_a)]^{n_2}(1-\theta_a)^{2n_3}} = \frac{\theta_0^{2n_1}2^{n_2}\theta_0^{n_2}(1-\theta_0)^{n_2}(1-\theta_0)^{2n_3}}{\theta_a^{2n_1}2^{n_2}\theta_a^{n_2}(1-\theta_a)^{n_2}(1-\theta_a)^{2n_3}}$$

$$= \left[\frac{\theta_0}{\theta_a}\right]^{2n_1+n_2}\left[\frac{1-\theta_0}{1-\theta_a}\right]^{n_2+2n_3} < k$$

$$\Leftrightarrow (2n_1+n_2)ln\left[\frac{\theta_0}{\theta_a}\right] + (n_2+2n_3)ln\left[\frac{1-\theta_0}{1-\theta_a}\right] < lnk$$

$$\Leftrightarrow (2n_1+n_2)ln\left[\frac{\theta_0}{\theta_a}\right] + (2n-(2n_1+n_2))ln\left[\frac{1-\theta_0}{1-\theta_a}\right] < lnk$$

$$\Leftrightarrow (2n_1+n_2) > \frac{lnk - 2nln\left[\frac{1-\theta_0}{1-\theta_a}\right]}{-\left[ln\left(\frac{1-\theta_0}{1-\theta_a}\right) - ln\left(\frac{\theta_0}{\theta_a}\right)\right]}$$

Consider $k' := \frac{lnk-2nln\left[\frac{1-\theta_0}{1-\theta_a}\right]}{-\left[ln\left(\frac{1-\theta_0}{1-\theta_a}\right)-ln\left(\frac{\theta_0}{\theta_a}\right)\right]}$. It follows that the rejection region for this test is,

$$\{2N_1 + N_2 > k'\}$$

Problem 97c. Given the trinomial distribution, you can find a critical value k' such that $P(2N_1 + N_2 > k') = \alpha$.

Problem 97d. Yes it is the uniformly most powerful test.

Problem 106.

Problem 111a. With H_0 the likelihood is maximized at θ_0 therefore $L(\hat{\Omega}_0) = L(\theta_0)$. Similarly with H_a the likelihood is maximized at either θ_a or θ_0 therefore $L(\hat{\Omega}) = max\{L(\theta_a), L(\theta_0)\}$. It follows

$$\lambda = \frac{L(\hat{\Omega}_0)}{L(\hat{\Omega})} = \frac{L(\theta_0)}{max\{L(\theta_a), L(\theta_0)\}} = \frac{1}{max\left\{\frac{L(\theta_a)}{L(\theta_0)}, 1\right\}}$$

Problem 111b. define k as follows

$$k := \begin{cases} 2 & \text{if } \lambda = 1 \\ k' & \text{if } \lambda = \frac{L(\theta_0)}{L(\theta_a)} \end{cases}$$

From this it is clear that $\lambda < k$.

Problem 111c. When both the null and alternative hypotheses are simple the likelihood ratio tests agree with the Neyman Pearson Lemma.