Stat 330 Assignment 4 Solutions

Mandy Yao

November 2020

4.2.22

We want to find a_L and a_U such that

$$P(a_L < p_1 - p_2 < a_U) = 90\%.$$

From equation 4.2.14, we have that

$$\hat{p_1} - \hat{p_2} \pm z_{\alpha/2}\sqrt{\frac{\hat{p_1}(1-\hat{p_1})}{n_1} + \frac{\hat{p_2}(1-\hat{p_2})}{n_1}},$$

where $\hat{p_1} = \frac{50}{100}$, $\hat{p_2} = \frac{40}{100}$, $z_{\alpha/2} = 1.645$. Then

$$0.5 - 0.4 \pm 1.645\sqrt{\frac{0.5(1-0.5)}{100} + \frac{0.4(1-0.4)}{100}} = 0.1 \pm 0.11515.$$

Therefore an approximate 90% interval for $p_1 - p_2$ is (-0.015, 0.215).

4.2.27

a)

Note that $W_1 = \frac{(m-1)s_2^2}{\sigma_2^2} \sim \chi^2_{(m-1)}$ and $W_2 = \frac{(n-1)s_1^2}{\sigma_1^2} \sim \chi^2_{(n-1)}$, and they are independent since $X_1, ..., X_n$ and $Y_1, ..., Y_m$ are independent sequences. Therefore

$$F = \frac{W_1/(m-1)}{W_2/(n-1)} = \frac{\frac{(m-1)s_2^2}{\sigma_2^2}/(m-1)}{\frac{(n-1)s_1^2}{\sigma_1^2}/(n-1)} = \frac{s_2^2/\sigma_2^2}{s_1^2/\sigma_1^2} \sim F_{(m-1,n-1)}.$$

b)

$P(F < b) = 0.975$ and $P(a < F < b) = 0.95$, $P(F < a) = 0.025$. We have thus that $a = F_{(m-1,n-1,0.025)}$ and $b = F_{(m-1,n-1,0.975)}$.

c)

Since $F = \frac{s_2^2/\sigma_2^2}{s_1^2/\sigma_1^2} \sim F(m-1, n-1)$,

$$0.95 = P\left(a < \frac{s_2^2/\sigma_2^2}{s_1^2/\sigma_1^2} < b\right) = P\left(F_{(m-1,n-1,0.025)} < \frac{s_2^2/\sigma_2^2}{s_1^2/\sigma_1^2} < F_{(m-1,n-1,0.975)}\right)$$

$$= P\left(\frac{s_1^2}{s_2^2}F_{(m-1,n-1,0.025)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2}{s_2^2}F_{(m-1,n-1,0.975)}\right).$$

4.4.28

a)

Let $Y_1 = min(X_1, X_2)$ and $Y_2 = max(X_1, X_2)$.

First, we know that

$$F_{Y_1}(y_1) = P(min(X_1, X_2) \leq y_1) = 1 - P((X_1, X_2) > y_1) = 1 - P(X_1 > y_1)P(X_2 > y_2)$$

$$= 1 - (1 - F_X(y_1))^2 \quad for \quad -\infty < y_1 < \infty$$

$$F_{Y_2}(y_2) = P(max(X_1, X_2) \leq y_2) = P(X_1 \leq y_2)P(X_2 \leq y_2) = F_X(y_2)^2 \quad for \quad -\infty < y_2 < \infty.$$

Since

$$F_X(\mu) = P(X \leq \mu) = P\left(\frac{X - \mu}{\sigma} \leq 0\right) = \Phi(0) = \frac{1}{2},$$

$$P(Y_1 < \mu < Y_2) = P(\mu < Y_2) - P(\mu < Y_1) = (1 - F_{Y_2}(\mu)) - (1 - F_{Y_1}(\mu))$$

$$= (1 - F_X(\mu)^2) - (1 - \{1 - (1 - F_X(\mu))^2\}) = (1 - (\frac{1}{2})^2) - (1 - \{1 - (1 - (\frac{1}{2}))^2\}) = \frac{1}{2}.$$

Therefore $P(Y_1 < \mu < Y_2) = \frac{1}{2}$.

Now we need to find $E(Y_2 - Y_1) = E(|X_2 - X_1|) = E[|Z|]$ where $X_1 \sim N(\mu, \sigma^2)$, $X_2 \sim N(\mu, \sigma^2)$ and $Z = X_1 - X_2 \sim N(0, 2\sigma^2)$. Thus

$$E(Y_2 - Y_1) = E[|Z|] = \int_{-\infty}^{\infty} \frac{|z|}{\sqrt{2\pi 2\sigma^2}} e^{\frac{-1}{2(2\sigma^2)} z^2} dz = 2\int_0^{\infty} \frac{z}{\sqrt{4\pi\sigma^2}} e^{\frac{-1}{4\sigma^2} z^2} dz$$

$$= \int_0^{\infty} \frac{1}{\sqrt{4\pi\sigma^2}} e^{\frac{-1}{4\sigma^2} u} du = \frac{2\sigma}{\sqrt{\pi}} = 1.13\sigma$$

b)

We have that $c\sigma = z_{\alpha/2}\sigma/\sqrt{n}$, n=2, and $\alpha = 0.5$. Then

$$c = z_{\alpha/2}/\sqrt{n} = 0.674/\sqrt{2} = 0.477.$$

Therefore the length is $2c\sigma = 0.95\sigma$. This can be compared with the expected value of the length in part a, which is 1.13σ.

4.5.5

First, we have

$$z = \frac{f(x_1; 2)f(x_2; 2)}{f(x_1; 1)f(x_2; 1)} \leq \frac{1}{2}.$$

Then

$$z = \frac{1}{2}e^{\frac{-x_1}{2}} \frac{1}{2}e^{\frac{-x_2}{2}} / e^{-x_1 - x_2} = \frac{1}{4}exp\{\frac{x_1 + x_2}{2}\} \leq \frac{1}{2} \Rightarrow \frac{x_1 + x_2}{2} \leq log(2).$$

Therefore the critical region is $x_1 + x_2 \leq 2log(2)$. Then, since $X_1 + X_2 \sim \Gamma(2, \theta)$ is the sum of two $\Gamma(1, 2)$ random variables,

$$\gamma(\theta) = P(X_1 + X_2 \leq 2log(2)) = \int_0^{2log(2)} \frac{1}{\theta^2} xe^{\frac{-x}{\theta}} = -\frac{1}{\theta} 2 \times 4^{\frac{-1}{\theta}} log(2) - 4^{\frac{-1}{\theta}} + 1.$$

Therefore the significance level is $\gamma(2) \approx 0.15$ and the power is $\gamma(1) \approx 0.4$.

4.5.11

a)

First we know that

$$F_{Y_4}(y_4) = F_X(y_4)^4 = (\frac{y_4}{\theta})^4$$

and

$$F_X(x) = \int_0^x \frac{1}{\theta} dx = \frac{x}{\theta}, 0 < x < \theta.$$

Then

$$\gamma(\theta) = P(Y_4 \geq c) = 1 - P(Y_4 < c) = 1 - (\frac{c}{\theta})^4.$$

and the significance level is 0.05 when

$$\gamma(1) = 1 - c^4 = 0.05 \Rightarrow c = (0.95)^{\frac{1}{4}}$$

b)

From part a,

$$\gamma(\theta) = 1 - (\frac{0.95^{\frac{1}{4}}}{\theta})^4 = 1 - \frac{0.95}{\theta^4} \; for \; \alpha = 0.05.$$

4.6.7

a)

The test statistic is

$$t = \frac{\overline{X} - \overline{Y}}{S_p\sqrt{(1/n_1) + (1/n_2)}}.$$

The critical value can be found using $1 - \alpha = 0.95$ and

$$df = n + m - 2 = 13 + 16 - 2 = 27.$$

Using a Student t-distribution table, we get $t = 1.703$. Since the test is left-tailed, the critical region contains all values below -1.703.

b)

$$S_p = \sqrt{\frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}} = \sqrt{\frac{(13-1)25.6^2 + (16-1)28.3^2}{13+16-2}} \approx 27.133.$$

The test statistic is

$$t = \frac{\overline{X} - \overline{Y}}{S_p\sqrt{(1/n_1) + (1/n_2)}} = \frac{72.9 - 81.7}{27.133\sqrt{\frac{1}{13} + \frac{1}{16}}} \approx -0.869.$$

We fail to reject the null hypothesis at level 0.05 since $-0.869 > -1.703$.

4.7.8

The expected value of each cell is $\frac{60\times45}{180} = 15$. The critical value can be found using $1-\alpha = 0.95$ and

$$df = (a-1)(b-1) = (3-1)(4-1) = 6.$$

Using a chi-square distribution table, we get $\chi^2_{0.95} = 12.6$. We have that the random variable

$$W = \sum_{j=1}^{b}\sum_{i=1}^{a} \frac{(X_{ij} - n(X_{i.}/n)(X_{.j}/n))^2}{n(X_{i.}/n)(X_{.j}/n)} \overset{H_0}{\sim} \chi^2(6), approximately.$$

$$W_{obs} = 2\frac{(15-3k-15)^2}{15} + 2\frac{(15-k-15)^2}{15} + 2\frac{(15+k-15)^2}{15} + 2\frac{(15+3k-15)^2}{15} + 4\frac{(15-15)^2}{15}$$
$$= \frac{8k^2}{3}.$$

The rejection of independence occurs when

$$\frac{8k^2}{3} \geq 12.6 \Rightarrow k \geq \sqrt{(3/8)(12.6)} = 2.16.$$

Therefore the smallest value of k that leads to the rejection of independence is 3.

4.8.4

First, we get

$$z = \frac{x-a}{b}.$$

The target pdf is

$$f_X(x) = \frac{1}{b} f_Z((x-a)/b).$$

From this we would randomly generate observations from $f_Z(z)$ and save them. Then we would divide each of the observations by b to obtain the observations from $f_X(x)$.

4.8.21

First, we know that

$$F_{U,V|W\leq1}(u, v|w \leq 1) = \frac{P(U \leq u, V \leq v, W \leq 1)}{P(W \leq 1)}.$$

Now, $P(W \leq 1) = \frac{\pi}{4}$ since the probability of falling within a circle with radius 1 ($W = U^2 + V^2 = 1$) within a square of area 4 with sides with range of U and V from -1 to 1, is $\frac{\pi}{4}$. Therefore

$$F_{U,V|W\leq1}(u, v|w \leq 1) = \frac{\frac{u}{2} \times \frac{v}{2}}{\frac{\pi}{4}} = \frac{uv}{\pi},$$

so

$$f_{U,V|W\leq1}(u, v|w \leq 1) = \frac{\delta F}{\delta u \delta v} = \frac{1}{\pi},\ u^2 + v^2 \leq 1$$

Transforming to polar coordinates, we have $u = r sin\theta$, $v = r cos\theta$, radius r where $0 < r < 1$, $0 < \theta < 2\pi$. Then the Jacobian is

$$|\mathbf{J}| = \begin{vmatrix} \frac{\delta u}{\delta r} & \frac{\delta u}{\delta \theta} \\ \frac{\delta v}{\delta r} & \frac{\delta v}{\delta \theta} \end{vmatrix} = \begin{vmatrix} sin\theta & r cos\theta \\ cos\theta & -r sin\theta \end{vmatrix} = r.$$

Then

$$f_{R,\Theta}(r,\theta|w \leq 1) = rf_{U,V}(u,v) = \frac{r}{\pi},\ 0 < \theta < 2\pi,\ 0 < r < 1.$$

We have that

$$x_1 = uz = rsin\theta\sqrt{\frac{-2log(r^2)}{r^2}} = rsin\theta\sqrt{\frac{-4log(r)}{r^2}},\ -\infty < x_1 < \infty.$$

Similarly,

$$x_2 = rcos\theta\sqrt{\frac{-4log(r)}{r^2}},\ -\infty < x_2 < \infty,$$

so

$$x_1^2 + x_2^2 = r^2\left(\frac{-4log(r)}{r^2}\right)(sin^2\theta + cos^2\theta) = -4log(r) \Rightarrow r = e^{-\frac{1}{4}(x_1^2+x_2^2)}.$$

Also,

$$\frac{x_1}{x_2} = tan\theta \Rightarrow \theta = arctan\left(\frac{x_1}{x_2}\right),$$

so the Jacobian is

$$|\mathbf{J}| = \begin{vmatrix} \frac{\delta\theta}{\delta x_1} & \frac{\delta\theta}{\delta x_2} \\ \frac{\delta r}{\delta x_1} & \frac{\delta r}{\delta x_2} \end{vmatrix} = \begin{vmatrix} \frac{x_2}{x_1^2+x_2^2} & \frac{-x_1}{x_1^2+x_2^2} \\ -\frac{1}{2}x_1 r & -\frac{1}{2}x_2 r \end{vmatrix} = \frac{r}{2}.$$

Finally

$$f_{X_1,X_2}(x_1,x_2) = f_{R,\Theta}(e^{\frac{-1}{4}(x_1^2+x_2^2)}, arctan(\frac{x_1}{x_2}))|J| = \frac{1}{\pi}e^{\frac{-1}{4}(x_1^2+x_2^2)} \times \frac{1}{2}e^{\frac{-1}{4}(x_1^2+x_2^2)}$$

$$= \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x_1^2}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x_2^2},\ -\infty < x_1 < \infty,\ -\infty < x_2 < \infty.$$

Therefore we have shown that the random variables X_1 and X_2 are iid with a common N(0,1) distribution.

4.9.2

a)

1. The x_i^*'s are drawn with replacement, so they are independent.

2. Since they are drawn with replacement, they have equal probability of being drawn, so $P(X_i^* = x_m) = P(X_j^* = x_m)$ for any $m = 1, ..., n$ and $i \neq j$. Thus they are identically distributed.

3. $$\hat{F}_n(t) = P(X_i^* \leq t) = \frac{\#\{x_j : x_j \leq t\}}{n} = \frac{1}{n}\sum_{j=1}^{n} 1(x_j : x_j \leq t),$$

   which is the empirical cdf of $x_1, ..., x_n$.

b)

$$E(X_i^*) = \sum_{j=1}^{n} x_j P(X_i^* = x_j) = \sum_{j=1}^{n} x_j \frac{1}{n} = \bar{x}$$

c)

Note that

$$P(X_i^* < x_{((n+1)/2)}) = \frac{1}{n} \times \left(\frac{(n+1)}{2} - 1\right) = \frac{n-1}{2n}, \ \ and$$

$$P(X_i^* > x_{((n+1)/2)}) = 1 - \frac{1}{n} \times \left(\frac{(n+1)}{2}\right) = \frac{n-1}{2n}.$$

Since $P(X_i^* < x_{((n+1)/2)}) = P(X_i^* > x_{((n+1)/2)})$. Thus $x_{((n+1)/2)}$ is the median.

d)

$$Var(X_i^*) = \sum_{i=1}^{n} (x_i - E(X_i^*))^2 P(X_i^* = x_i) = \sum_{i=1}^{n} (x_i - \bar{x})^2 \frac{1}{n}$$