Solutions for Stat 512 — Take home exam IV

1. In a study of the relationship between birth order and college success, an investigator found that 126 in a sample of 180 college graduates were firstborn or only children; in a sample of 100 nongraduates of comparable age and socioeconomic background, the number of firstborn or only children was 54. Estimate the difference in proportions of firstborn or only children for the two populations from which these samples were drawn using 90% confidence coefficient. Is there sufficient evidence to conclude the proportion of firstborn or children for college graduates is higher than that of nongraduates? (10 pts)

Solution:

Since $n_1 = 180$ and $n_2 = 100$, the large sample size indicates that it is appropriate to say $\hat{p}_1$ and $\hat{p}_2$ have approximately normal distribution. Now,

$$\hat{p}_1 = \frac{126}{180} = 0.7$$
$$\hat{p}_2 = \frac{54}{100} = 0.54$$

Hence, the 90% CI for $p_1 - p_2$ is:

$$\hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$
$$= 0.7 - 0.54 \pm 1.645 * \sqrt{\frac{0.7 * 0.3}{180} + \frac{0.54 * 0.46}{100}}$$
$$= (0.06, 0.26)$$

Yes there is sufficient evidence to conclude that the proportion of firstborn or children for college graduates is higher than that of nongraduates since both upper and lower limits are greater than 0.

****

2. A small amount of the trace element selenium, from 50 to 200 micrograms per day, is considered essential to good health. Suppose that independent random samples of $n_1 = n_2 = 30$ adults were selected from two regions of the United States, and a days intake of selenium, from both liquids and solids, was recorded for each person. The mean and standard deviation of the selenium daily intakes for the 30 adults from region 1 were $\bar{Y}_1 = 165.1$ and $S_1 = 23$, respectively. The corresponding statistics for the 30 adults from region 2 were $\bar{Y}_2 = 138$ and $S_2 = 22$. Find a 90% confidence interval for the difference in the mean selenium intake for the two groups. (10 pts)

Solution:

The 90% CI for $\mu_1 - \mu_2$ is:

$$\bar{Y}_1 - \bar{Y}_2 \pm z_{\alpha/2} \cdot \sqrt{\frac{S_1^2}{n_1} + \frac{S_1^2}{n_2}}$$
$$= 165.1 - 138 \pm 1.645\sqrt{\frac{23^2}{30} + \frac{22^2}{30}}$$
$$= (17.543, 36.657)$$

****

3. Organic chemists often purify organic compounds by a method known as factional crystallization. An experimenter wanted to prepare and purify 4.85g of aniline. Ten 4.85-gram specimens of aniline were prepared and purified to produce acetanilide. The following dry yields were obtained:

3.85, 3.88, 3.90, 3.62, 3.72, 3.80, 3.85, 3.36, 4.01, 3.82

Construct a 95% CI for the mean number of grams of acetanilide that can be recovered from 4.85 grams of aniline. (10 pts) (Hint: 1. you need to use software to calculate sample mean and variance. 2. Be careful about the formula you choose. Is it a small sample or large sample situation?)

Solution:

In R, do the following:

```
> data<-c(3.85,3.88,3.90,3.62,3.72,3.80,3.85,3.36,4.01,3.82)
> mean(data)
[1] 3.781
> var(data)
[1] 0.03274333
> qt(0.025,9)
[1] -2.262157
```

The CI therefore is :

$$3.781 \pm t_{9,0.025}\frac{\sqrt{0.03}}{\sqrt{10}} = 3.781 \pm 2.26\frac{\sqrt{0.03}}{\sqrt{10}} = (3.66, 3.90)$$

****

4. Let Y has pdf:

$$f_Y(y) = \frac{2(\theta - y)}{\theta^2}, \qquad y \in (0, \theta)$$

a. Find out the CDF of Y. (10 pts)

b. Show that $\frac{Y}{\theta}$ is a pivotal quantity. (10 pts)

c. Use the pivotal quantity from part (b) to find a 90% upper confidence limit for θ. (10 pts)

Solution:

For part (a),

$$\begin{aligned} F_Y(y) &= \int_0^y \frac{2(\theta - y)}{\theta^2} dy \\ &= \left(\frac{2y}{\theta} - \frac{y^2}{\theta^2} \right) \Bigg|_{y=0}^{y} \\ &= \frac{2y}{\theta} - \frac{y^2}{\theta^2}, \qquad y \in (0, \theta) \end{aligned}$$

For part (b), let $U = \frac{Y}{\theta}$, then

$$f_U(u) = \frac{2(\theta - u\theta)}{\theta^2} \cdot \theta = 2(1 - u), \qquad u \in (0, 1)$$

Hence U is a function of sample data and unknown target parameter θ, whose distribution does not depend on θ, it is a pivotal quantity.

For part (c),

$$\begin{aligned} & P(U > a) = 0.9 \\ \Longrightarrow\ & P(\theta \leq \frac{Y}{a}) = 0.9 \end{aligned}$$

Determine a? Since $P(U > a) = 0.9$,

$$\begin{aligned} & \int_a^1 2(1 - u) du = 0.9 \\ \Longrightarrow\ & \left(2u - u^2\right) \Big|_a^1 = 0.9 \\ \Longrightarrow\ & a = \pm\sqrt{0.9} + 1 \end{aligned}$$

Since a can only be from 0 to 1, $a = 1 - \sqrt{0.9} = 0.05$, which implies the 90% upper confidence limit for θ is $\frac{Y}{0.05}$.

****

5. The distribution function for a power family distribution is given by:

$$F_Y(y) = \left(\frac{y}{\theta}\right)^{\alpha}, \qquad y \in [0, \theta]$$

where α, $\theta > 0$. Assume that a sample of size n is taken from a population with a power family distribution and that $\alpha = c$ where $c > 0$ is known.

a. Show that the distribution function of $Y_{(n)} = max\{Y_1, \ldots, Y_n\}$ is given by: (10 pts)

$$F_{Y_{(n)}}(y) = \left(\frac{y}{\theta}\right)^{nc}, \qquad y \in [0, \theta]$$

b. Show that $\frac{Y_{(n)}}{\theta}$ is a pivotal quantity and that for $0 < k < 1$ (10 pts)

$$P\left(k < \frac{Y_{(n)}}{\theta} \leq 1\right) = 1 - k^{cn}$$

c. Suppose that $n = 5$ and $\alpha = c = 2.4$, then use the result from part (b) to find k such that (10 pts)

$$P(k < \frac{Y_{(5)}}{\theta} \leq 1) = 0.95$$

and give a 95% CI for θ.

Solution:

For part (a),

$$\begin{aligned} F_{Y_{(n)}}(y) &= P(Y_{(n)} \leq y) \\ &= P(Y_1 \leq y, \ldots, Y_n \leq y) \\ &= [F_Y(y)]^n \\ &= \left(\frac{y}{\theta}\right)^{nc}, \qquad y \in [0, \theta] \end{aligned}$$

For part (b),

$$P(\frac{Y_{(n)}}{\theta} \leq y) = P(Y_{(n)} \leq y\theta)$$
$$= \left(\frac{y\theta}{\theta}\right)^{nc}$$
$$= y^{cn}, \qquad y \in [0, 1]$$

Hence, $\frac{Y_{(n)}}{\theta}$ is a pivotal quantity. Furthermore,

$$P\left(k < \frac{Y_{(n)}}{\theta} \leq 1\right) = F_{Y_{(n)}}(\theta) - F_{Y_{(n)}}(\theta k)$$
$$= \left(\frac{\theta}{\theta}\right)^{nc} - \left(\frac{\theta k}{\theta}\right)^{nc}$$
$$= 1 - k^{cn}$$

For part (c), it can be easily seen from part (b) that $1 - k^{cn} = 1 - k^{2.4*5} = 0.95$ which implies $k = 0.779$. Hence $P\left(0.779 < \frac{Y_{(5)}}{\theta} \leq 1\right) = 0.95$, and a 95% CI for θ is $\left(Y_{(5)}, \frac{Y_{(5)}}{0.779}\right)$.

****

6. Let $X_1, \ldots, X_{10}$ are iid with pdf $f(x|\mu) = e^{-(x-\mu)}$, where $x > \mu$ (This is called a shifted exponential distribution). Find a $100(1-\alpha)\%$ CI for μ using the pivotal method. (Hint: Consider a function of $Y_{(1)} = min(X_1, \ldots, X_{10})$). (10 pts)

Solution:

Firstly let's find out the distribution of $Y_{(1)}$:

$$f_{Y_{(1)}}(y) = ne^{-(y-\mu)}\left[1 - \int_{\mu}^{y} e^{-(x-\mu)}dx\right]^{n-1}$$
$$= ne^{-n(y-\mu)}, \qquad y \in (\mu, \infty)$$

Consider $U = Y_{(1)} - \mu$, one can easily prove that $U \sim exp(\frac{1}{n})$, which implies $U = Y_{(1)} - \mu$ is a pivotal quantity. Hence, we choose a and b such that

$$P(a < U < b) = 1 - \alpha$$

So as long as $b \geq a > 0$, and

$$\int_a^b f_U(u)du = \int_a^b ne^{-nu}du = e^{-an} - e^{-bn} = 1 - \alpha$$

Then $P(a < Y_{(1)} - \mu < b) = P(Y_{(1)} - b < \mu < Y_{(1)} - a) = 1 - \alpha$. Hence, a $100(1-\alpha)\%$ CI for μ is $(Y_{(1)} - b, Y_{(1)} - a)$ where $b \geq a > 0$ and $e^{-an} - e^{-bn} = 1 - \alpha$.