Math 326

Probability and Statistics II

Homework 03 - (8.5 - 8.9)

Matthew Wilder

Due: ~~02/09/22~~ 02/11/22 at 23:59

Contents

1

Let Y have probability density function $f_Y(y) = \dfrac{2(\theta - y)}{\theta^2}$ on support $0 < y < \theta$.

1.a

Determine the cumulative distribution function for Y and explain why it is not pivotal.

Solution: We compute the CDF of Y by integration.

$$\begin{aligned}
P(Y \leq y) &= \int_0^y f_Y(y)dy \\
&= \int_0^y \frac{2(\theta - y)}{\theta^2}\,dy \\
&= \frac{2}{\theta^2}\int_0^y \theta - y\,dy \\
&= \frac{2}{\theta^2}\left[y\theta - \frac{y^2}{2}\right]_{y=0}^{y=y} \\
&= \frac{2}{\theta^2}\left(y\theta - \frac{y^2}{2}\right) \\
&= \frac{2y}{\theta} - \frac{y^2}{\theta^2}.
\end{aligned}$$

Therefore the CDF for Y is

$$F_Y(y) = \begin{cases} 0 & y \leq 0 \\ \frac{2y}{\theta} - \frac{y^2}{\theta^2} & y \in (0, \theta) \\ 1 & y \geq \theta \end{cases}.$$

Which is <u>not pivotal</u> because $F_Y(y)$ depends on some θ.

1.b

Show that the change of variables $U = Y/\theta$ is a pivotal quantity.

Solution: Compute the CDF of U,

$$\begin{aligned} P(U \leq u) &= P(Y/\theta \leq u) \\ &= P(Y \leq \theta u) \\ &= F_Y(\theta u) \\ &= \frac{2\theta u}{\theta} - \frac{(\theta u)^2}{\theta^2} \\ &= 2u - u^2 \\ &= u(2-u). \end{aligned}$$

Therefore the CDF for U is

$$F_U(u) = \begin{cases} 0 & u \leq 0 \\ u(2-u) & u \in (0,1) \\ 1 & u \geq 1 \end{cases}.$$

Which is pivotal because $F_U(u)$ no longer depends on θ.

1.c

For a single observation of Y, construct a 90% lower confidence interval for θ.

Solution: We want

$$P(\hat{\theta}_L \leq \theta) = 0.90.$$

That is, $F_U(u) = 0.90$ and thus

$$\begin{aligned} & u(2-u) = 0.90 \\ \iff \quad & -u^2 + 2u - 0.90 = 0 \\ \iff \quad & u = \frac{-2 \pm \sqrt{2^2 - 4(-1)(-0.90)}}{2(-1)} \\ \iff \quad & u = \frac{2 \pm \sqrt{0.4}}{2} \\ \iff \quad & u = \frac{10 \pm \sqrt{10}}{10} \\ \iff \quad & u \approx 0.6838 \quad \text{or} \quad u \approx 1.3162 \end{aligned}$$

But u is only defined for $u \in [0,1]$, therefore the 90% lower confidence interval is $(0,\, 0.6838)$.

2

AAA reports a study in which 171 out of 880 randomly selected drivers admitted running a red light in the recent past. Assuming those polled answered honestly, find a 90% confidence interval for the percentage of all drivers who have run a red light in the recent past.

Solution: Let $\hat{p} := \{\text{percentage that ran red light}\}$, then $\hat{p} = \dfrac{171}{880} = 0.1943\overline{18}$. Then the error, E, is

$$E = Z_{\alpha/2}\sqrt{\frac{pq}{n}} = 1.645\sqrt{\frac{0.1943\overline{18}(1-0.1943\overline{18})}{880}} \approx 0.021941335937902515694874355 07$$

Thus the interval 90% confidence interval is $(0.1943\overline{18} \pm E)$, also known to be approximately

$$(0.17237684588027930248694382675024 93,\ 0.21625951775608433387669253688611 43)$$

3

A study reported by Steele and Torrie wanted to measure the strength of hydrogen sulfide produced by sewage over 42 hours in warm conditions. As the actual contents of the sewage varied from run to run, $n = 9$ different observations were collected; these observations had mean $\bar{X} = 218$ ppm and $S = 11.1$ ppm. We want to estimate μ, the mean amount of hydrogen sulfide produced over the population of all sewage situations.

3.a

Determine the standard error of the mean.

Solution: The standard error, $\sigma_{\bar{X}}$ is

$$\sigma_{\bar{X}} = \sqrt{\frac{S^2}{n}} = \sqrt{\frac{11.1^2}{9}} = \frac{11.1}{3} = 3.7.$$

3.b

Find a 99% confidence interval for μ.

Solution: There are $n - 1 = 8$ df, and

$$\begin{aligned}
\mu &= \bar{X} \pm t_{\alpha/2}\left(\frac{S}{\sqrt{n}}\right) \\
&= 218 \pm t_{0.005}\left(\frac{11.1}{\sqrt{9}}\right) \\
&= 218 \pm 3.355 \cdot 3.7 \\
&= 218 \pm 3.355 \cdot 3.7 \\
&= 218 \pm 12.4135
\end{aligned}$$

Therefore the 99% C.I. for μ is (205.5865 ppm, 230.4135 ppm).

4

A phone poll survey is undertaken to see if men and women have a difference of opinions on a specific topic.

4.a

If 1000 men and 1000 women are to be interviewed, how accurately could you estimate the difference in the proportions? Find an error bound on the estimation.

Solution: Using

$$\hat{p}_1 - \hat{p}_2 \pm \underbrace{Z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}_{E \text{ error bound}}.$$

With $n_1 = 1000$ men, $n_2 = 1000$ women, and a total size of $n_1 + n_2 = 2000$. Then $\hat{p}_1$, the proportion of men, is $\frac{1000}{2000} = 0.5$, and similarly $\hat{p}_2$, the proportion of women, is also 0.5. Using a 2σ confidence interval, $Z_{0.025} = 1.960$. Thus we can compute the error,

$$\begin{aligned}
E &= Z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \\
&= 1.960\sqrt{\frac{0.5(1-0.5)}{1000} + \frac{0.5(1-0.5)}{1000}} \\
&= 1.960\sqrt{2 \cdot \frac{0.25}{1000}} \\
&\approx 0.0438269323589958780496198039071330142146361198483859041957.
\end{aligned}$$

4.b

Suppose that you were designing the survey and wished to estimate the difference in a pair of proportions, correct to within 2%, with probability 90%. How many interviewees should be included in each sample.

Solution: We want $E < 0.02$, For a 90% C.I., $Z_{0.05} = 1.645$. Assuming that $n_1 = n_2$, we get

$$
\begin{aligned}
0.02 &\geq Z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \\
&= 1.645\sqrt{\frac{0.25}{n_1} + \frac{0.25}{n_2}} \\
&= 1.645\sqrt{\frac{0.5}{n_1}} \\
\iff \quad 0.012158 &\geq \sqrt{\frac{0.5}{n_1}} \\
\iff \quad 0.000147818294 &\geq \frac{0.5}{n_1} \\
\iff \quad n_1 &\geq \frac{0.5}{0.000147818294} \\
&= 3382.53125828
\end{aligned}
$$

Thus we would need at least 3383 interviewees from each group (men/women) to be correct within 2%.

5

Last semester 7 of the 19 MTH 325 students filled out the end of the semester course evaluations. When asked how much time (in hours) they spent outside of class per week on the course, they answered

$$15, 6, 6, 7, 4, 10, \text{and } 8.$$

5.a

Find a 98% confidence interval for the average time spent outside of class on the course per week for all student.

Solution: Let $\bar{X}$ denote the sample mean. Then $\bar{X} = 8$. The sample variance, S^2 is

$$\begin{aligned} S^2 &= \frac{\sum_{i=1}^{7}(X_i - \bar{X})^2}{n-1} \\ &= \frac{1}{6}\left[(15-8)^2 + (6-8)^2 + (6-8)^2 + (7-8)^2 + (4-8)^2 + (10-8)^2 + (8-8)^2\right] \\ &= \frac{1}{6}[49 + 4 + 4 + 1 + 16 + 4 + 0] \\ &= \frac{78}{6} \\ &= 13 \end{aligned}$$

There are 6 degrees of freedom, and by table 5, the 98% C.I. $t_{0.010} = 3.143$. Thus,

$$\begin{aligned} \text{C.I.} &= \bar{X} \pm t_{0.010} \cdot \sqrt{\frac{S^2}{n}} \\ &= 8 \pm 3.143\sqrt{\frac{13}{7}} \\ &\approx 8 \pm 4.28318701436208596469527092124428716574325917937495043504 76 \end{aligned}$$

Thus, the 98% CI is approximately
(3.7168129856379140353047290787557128, 12.2831870143620859646952709212442872).

5.b

Find a 95% confidence interval for the true variance in the time students spent outside of class on the course per week.

Solution: Using table 6 (with 6 degrees of freedom) to obtain values for the 95% C.I., we have that $\chi^2_{0.025} = 14.4494$ and $\chi^2_{0.975} = 1.237347$. Then the confidence interval, I is

$$\begin{aligned} I &= \left(\frac{(n-1)S^2}{14.4494}, \frac{(n-1)S^2}{1.237347} \right) \\ &= \left(\frac{6 \cdot 13}{14.4494}, \frac{6 \cdot 13}{1.237347} \right) \\ &\approx (5.398148019986989079131313410937,\ 63.038096831365817349538973303365 99) \end{aligned}$$