Homework 5

Winter 2024

KEY

2024-02-21

Exercises

1.

a. Before taking off, a pilot is supposed to set the aircraft altimeter to the elevation of the airport. A pilot leaves Denver (altitude 5,280 feet) with her altimeter set to 2,500 feet, so the subsequent altitude measurements are off from the true altitude of the plane.

   This is an example of bias because the altitude measurements are going to be systematically below the true altitude.

b. This is an example of variability since the actual voltage will vary randomly around what it expected.

c. This is an example of bias since we will be systematically underreporting the percentage of alcohol related fatalities due to the flaw in the design.

d. This is an example of bias since again the temperature measurements will be systematically higher than the actual temperatures we would get if the design did a better job of representing various locations.

2. Let X denote the lifetime (in days) of a randomly selected component. Then we are told that we have $X_1, X_2, \ldots, X_{40} \sim Exp\left(\lambda_0 = \frac{1}{45}\right)$. We state the following results for an exponential distribution which will come in handy later:

$$\mu_0 = E\left[X\right] = \frac{1}{\lambda_0}$$
$$\sigma_0^2 = Var\left[X\right] = \frac{1}{\lambda_0^2}.$$

   Let $S = X_1 + X_2 + \cdots + X_{40}$ denote the total number of days that the 40 components will last. We want to calculate $P(S \geq 5 \times 365.25)$.

   By the Central Limit Theorem (assuming n is large and also that the lifetimes of the bulbs are independent, meaning how long one lasts has no bearing on how long another lasts), we can say

$$S \approx Norm\left(40 \times \frac{1}{\lambda_0}, \frac{\sqrt{40}}{\lambda_0}\right).$$

The desired probability is calculated below.

```
prob_s <- pnorm(q = 5*365.25, mean = 40*45, sd = sqrt(40)*45, lower.tail=F )
```

As we can see the probability that the 40 components together will be sufficient for 5 years is 0.4633.

3.

a.

```
sample_summary <- sample_summary %>%
  mutate( lower = sample_mean - qnorm(p=0.95)*sample_sd/sqrt(sampsize),
          upper = sample_mean +qnorm(p=0.95)*sample_sd/sqrt(sampsize))
```

b.

```
ggplot(data=sample_summary)+
  geom_segment(mapping = aes(x = lower,
                             xend = upper,
                             y = 1:nsamp,
                             yend = 1:nsamp)) +
  labs(x = "mean PTSG",
       y = "Confidence Interval",
       title = "100 90% CI's for Mean PTSG (MIAA05 Dataset)") +
  theme(axis.text.y=element_blank(),
        axis.ticks.y=element_blank())
```

c.

```
sample_summary <- sample_summary %>%
  mutate(
  cover = ifelse(lower < mean(MIAA05$PTSG) & mean(MIAA05$PTSG) < upper, 1,0) )
sample_summary %>% summarize("Coverage Proportion" = mean(cover))
```

```
##   Coverage Proportion
## 1                0.91
```

d. We would **expect** the same proportion of coverage (90%), since we should expect 90% of intervals to cover the true mean regardless of sample size. The confidence intervals are constructed in such a way that on average 90% of the intervals will cover the true mean, so even if the intervals will shrink if n increases to 25, the coverage proportion should remain constant.

4.

a. The CDF of Y is shown below.

$$
\begin{aligned}
F_Y(y) &= P(Y \le y) \\
&= P(\frac{X}{\theta_0} \le y), \\
&= F_X\left(\theta_0\, y\right).
\end{aligned}
$$

Therefore the PDF of Y is

$$\begin{aligned} f_Y(y) &= \frac{d}{dy} F_Y(y), \\ &= \frac{d}{dy} F_X\left(\theta_0\, y\right), \\ &= f_X\left(\theta_0\, y\right) \theta_0, && \text{(chain rule of differentiation)} \\ &= \frac{1}{\theta_0} \cdot \theta_0 \qquad 0 \le \theta_0\, y < \theta_0, \\ &= 1 \qquad\qquad 0 \le y < 1 \end{aligned}$$

Hence $Y \sim Unif(0,1)$ and its distribution does not depend on θ_0.

b. We could consider $a = 0.05$ and $b = 0.95$ since $P(0.05 \le Y \le 0.95) = 0.9$.

c. We start with the probability statement in part b as follows:

$$\begin{aligned} P(0.05 \le Y \le 0.95) &= 0.9, \\ P(0.05 \le \frac{X}{\theta_0} \le 0.95) &= 0.9, \\ P\left(\frac{X}{0.95} \le \theta_0 \le \frac{X}{0.05}\right) &= 0.9, \end{aligned}$$

which gives the 90% confidence interval estimator $[\frac{X}{0.95}, \frac{X}{0.05}]$ for θ_0.

d. The 90% confidence interval estimate for θ_0 is $[2.21, 42]$. This says that with 90% confidence, the smallest θ_0 can be is 2.21 and the largest it can be is 42.