Stat135: Homework 2

Deadline: Feb. 9th at 11:55 pm.

1 Write below your solution for ex. 2 ch. 2 (Stat Labs)

- There are in total $\binom{5}{3}$ possible random sample of size 3, in which there are 4 $\{1, 2, 4\}$ triple, 2 $\{2, 2, 4\}$ triple, 1 $\{1, 2, 2\}$ triple, 2 $\{2, 4, 4\}$triple and 1 $\{1, 4, 4\}$triple. Denote the median random variable as Y. Calculate Y of each triple, we got 2 for the first seven triples and 4 for the rest three. So the distribution of median is:

$$\begin{aligned} P(median = 2) &= \frac{7}{10} \\ P(median = 4) &= \frac{3}{10} \end{aligned}$$

- Expectation:

$$\mathbb{E}Y = 2P(Y = 2) + 4P(Y = 4) = 13/5$$

SD:

$$\sqrt{\mathbb{E}(Y - \mathbb{E}Y)^2} = \sqrt{(2 - \frac{13}{5})^2 P(Y = 2) + (4 - \frac{13}{5})^2 P(Y = 4)} = \sqrt{21}/5$$

2 Write below your solution for ex. 10 ch. 2 (Stat Labs)

By CLT, for n large, the probability distribution of:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

is approximately standard normal. So the 95% confidence for μ is: $(\bar{x} - 2\sigma/\sqrt{n}, \bar{x} + 2\sigma/\sqrt{n})$ whose width is $4\sigma/\sqrt{n}$. If the width is less than 4%, then:

$$n = 10^4\sigma^2.$$

Ignore the finite population correction factor. Use the sample percentage $\bar{x}$ to estimate the population variance: $\hat{\sigma}^2 = \bar{x} - \bar{x}^2$. Plug in, then we get

$$n = 10^4(\bar{x} - \bar{x}^2),$$

By the sample data, $\bar{x} = \frac{67}{91}$, then $n = 1941$. If $\bar{x} = 1/2$, $n = 10^4(\bar{x} - \bar{x}^2)$ reaches its maximal, then $n = 2500$.

3 Write below your solution for ex. 11 ch. 2 (Stat Labs)

Use the sample mean to estimate the population mean. Ignore the finite population corrector, the estimate of SE equals to $\hat{SE} = \sqrt{\bar{x}(1-\bar{x})}/\sqrt{n-1}$.

a. Let $\hat{SE}$ be less than 1%. We got roughly $n = 10^4\bar{x}(1-\bar{x})$. For $\bar{x}$ equals roughly to 50%, $n \approx 2500$, for $\bar{x}$ equals roughly to 10%, $n \approx 900$.

b. Let $\hat{SE}$ be less roughly than 10% of the population parameter. For $\bar{x}$ equals roughly to 50%, $n = 400\bar{x}(1-\bar{x})$, then $n \approx 100$. For $\bar{x}$ equals roughly to 10%, $n = 10^4\bar{x}(1-\bar{x})$ then $n \approx 900$.

4 Write below your solution for ex. 13 ch. 2 (Stat Labs)

a. Let $N = 314$ be the number of whole population and $N_w = 131$. The population fraction is:

$$\begin{aligned} \pi &= \frac{\sum_{i=1}^{N} \mathbf{1}\{x_i \text{ is woman and played video games}\}}{N_w} \\ &= \frac{\sum_{i=1}^{N} \mathbf{1}\{x_i \text{ is woman and played video games}\}}{N} \frac{N}{N_w} \end{aligned}$$

The expected value of $\hat{\pi}$ is:

$$\mathbb{E}\hat{\pi} = \mathbb{E}\bar{v}\frac{N}{N_w}$$

where,

$$\begin{aligned} \mathbb{E}\bar{v} &= \frac{\sum_{i=1}^{n} \mathbf{1}\{x_i \text{ is woman and played video games}\}}{n} \\ &= P(x_i \text{ is woman and played video games}) \\ &= \frac{\sum_{i=1}^{N} \mathbf{1}\{x_i \text{ is woman and played video games}\}}{N} \end{aligned}$$

Then we have $\mathbb{E}\hat{\pi} = \pi$, which imply that $\hat{\pi}$ is an unbiased estimator.

b. The SE of $\hat{\pi}$:

$$\hat{SE}(\hat{\pi}) = \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n-1}}\sqrt{\frac{N-n}{N}}$$

where $N = 314,\ n = 91$. Plug in, we get $\hat{SE}(\hat{\pi}) = \sqrt{0.0189\hat{p}(1 - 314\hat{p}/131)}$.

5 Write below your solution for ex. 15 ch. 2 (Stat Labs)

Let a random variable Y to be the number of female students in the sample and random variable $X_i = \mathbf{1}\{x_i \text{ is a woman}\}$. Then we could write Y as:

$$Y = \frac{\sum_{i=1}^{n} X_i}{n}.$$

As $\mathbb{E}X_i = 131/314 = 0.4172$ and $\mathrm{Var}X_i = \frac{131}{314}(1 - \frac{131}{314})\frac{223}{314} = 0.1732$. By CLT, we have the distribution of $\frac{(Y - \mathbb{E}X_i)}{\sigma/\sqrt{n}}$ goes to $\mathcal{N}(0,1)$ as $n \to \infty$. Using the normal approximation to compute the probability of having 38 female:

$$P(37.5 \leq Y < 38.5) \quad = \quad \Phi(0.1347) - \Phi(-0.117) = 0.10014,$$

whereas the exact calculation in the text is 0.1003.

The two answers are nearly the same and it indicates that CLT is a good approximation when sample size $n = 91$.

6 Write below your solution for ex. 17 ch. 2 (Stat Labs)

a. False. A confidence interval does not predict that the true value of the parameter has a particular probability of being in the confidence interval given the data actually obtained.

b. False. What we could say is that: the sample percentage has 0.95 probability within two standard error of μ.

c. False. The object we consider for the interval is the population percentage not the students.

d. True. The 95% interval runs in $[\bar{x} - 2SD, \bar{x} + 2SD]$ which is $[0.26, 0.42]$. This is exactly what confidence interval means.

7 Write below your solution for ex. 19 ch. 2 (Stat Labs)

- Write all the estimators in ex.18 as spacial case of $\bar{x}_w$:

$$\begin{aligned}
\hat{x} &= 1 * x_{I(1)} + \sum_{i=2}^{n} 0 * x_{I(i)} \\
\tilde{x} &= 2 * x_{I(1)} + (-1) * x_{I(2)} + \sum_{i=3}^{n} 0 * x_{I(i)} \\
x^* &= \sum_{i=1}^{n} \frac{2}{n} * x_{I(i)}
\end{aligned}$$

- We know that:

$$\begin{aligned}
\mathbb{E}\bar{x}_w = \mathbb{E}\sum_{i}^{n} w_i x_{I(i)} &= \sum_{i}^{n} w_i \mathbb{E} x_{I(i)} \\
\text{where, } \mathbb{E} x_{I(i)} &= \frac{1}{N}\sum_{j=1}^{N} x_j.
\end{aligned}$$

Then the expectation of $\bar{x}_w$ is:

$$\sum_{i}^{n} w_i * \frac{1}{N}\sum_{j=1}^{N} x_j.$$

If the estimator is unbiased then $\mathbb{E}\bar{x}_w = \frac{1}{N}\sum_{j=1}^{N} x_j$ which indicates $\sum_{i}^{n} w_i = 1$.