Homework assignment #10: Solutions

Milica Cudina

2021-10-24

Problem #1 (10 points)

You are rolling a fair dodecahedral ($20-$sided) die with numbers $1 - 20$ noted its sides. You win if the number rolled is prime and you lose if it is not. What is the probability that you win at most 200 times in 500 rolls? Calculate your answer **both**:

(i) (4 points) using **R** without an approximation, and

(ii) (6 points) using the normal approximation.

Solution: The prime numbers smaller than or equal to 20 are: $2, 3, 5, 7, 11, 13, 17, 19$. So, the probability of the number rolled being prime in a single roll is $8/20 = 2/5 = 0.4$. Hence, the number X of wins in 500 rolls has the following distribution:

$$X \sim Binomial(n = 500, p = 0.4)$$

In `R`, we use the command

```
pbinom(200, size = 500, prob = 0.4)
## [1] 0.5194108
```

The mean and the variance of X are

$$\mathbb{E}[X] = np = 500(0.4) = 200 \quad \text{and} \quad Var[X] = np(1-p) = 120.$$

So, the standard deviation of X is $SD[X] = 2\sqrt{30}$. Using the normal approximation to the binomial (with the continuity correction), we get

$$\mathbb{P}[X \leq 200] = \mathbb{P}[X < 200.5] = \mathbb{P}\left[\frac{X - 200}{2\sqrt{30}} < \frac{200.5 - 200}{2\sqrt{30}}\right] = \mathbb{P}[Z < 0.0456] = \Phi(0.0456)$$

where $Z \sim N(0, 1)$. To obtain the final probability we use

```
pnorm(0.0456)
## [1] 0.5181855
```

Problem #2 (10 points)

A biased coin (probability of {*heads} is 0.7) is tossed 1000 times. Write down the exact expression for the probability that more than 750 {*heads} have been observed.

(i) (4 points) Use **R** to calculate this probability.

(ii) (6 points) Use the normal approximation to approximate this probability.

Solution: The random variable X which equals to the number of heads is binomial with probability $p = 0.7$ and $n = 1000$. We are interested in the probability $\mathbb{P}[X > 750]$. If we split this probability among the elementary outcomes which are greater than 750, we get

$$\mathbb{P}[X > 750] = \sum_{i=751}^{1000} \mathbb{P}[X = i] = \sum_{i=751}^{1000} \binom{1000}{i} (0.7)^i (0.3)^{1000-i}.$$

In `R`, we use the command

```
1 - pbinom(750, size = 1000, prob = 0.7)
## [1] 0.0001985473
```

According to the normal approximation to the binomial distribution, the random variable

$$X' = \frac{X - \mathbb{E}[X]}{\sqrt{Var[X]}} = \frac{X - 700}{\sqrt{1000 \cdot 0.7 \cdot 0.3}},$$

is approximately normally distributed with mean 0 and standard variation 1. Therefore (note the *continuity correction*),

$$\mathbb{P}[X > 750] = \mathbb{P}[X' \geq \frac{750.5 - 700}{\sqrt{210}}] \approx \mathbb{P}[Z \geq 3.48483],$$

where $Z \sim N(0, 1)$ is normally distributed with mean 0 and standard variation 1. In `R`, we have

```
1 - pnorm(3.48483)
## [1] 0.0002462249
```

Problem #3 (2+2+4+1+1=10 points)

Solve **Exercise 6.8** from the textbook.

Solution:

```
knitr::include_graphics("oc-p6-8.png")
```

(a) The population parameter of interest is the proportion of all Greeks who would rate their lives poorly enough to be considered "suffering", p. The point estimate for this parameter is the proportion of Greeks in this sample who would rate their lives as such, $\hat{p} = 0.25$.

(b) 1. Independence: The sample is random, and $1,000 < 10\%$ of all Greeks, therefore the life rating of one Greek in this sample is independent of another.

   2. Success-failure: $1,000 \times 0.25 = 250 > 10$ and $1,000 \times 0.75 = 750 > 10$.

   Since the observations are independent and the success-failure condition is met, $\hat{p}$ is expected to be approximately normal.

(c) A 95% confidence interval can be calculated as follows:

$$\begin{aligned}
\hat{p} \pm z^{\star}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} &= 0.25 \pm 1.96 \times \sqrt{\frac{0.25 \times 0.75}{1000}} \\
&= 0.25 \pm 1.96 \times 0.0137 \\
&= 0.25 \pm 0.0269 \\
&= (0.2231, 0.2769)
\end{aligned}$$

   We are 95% confident that approximately 22% to 28% of Greeks would rate their lives poorly enough to be considered "suffering".

(d) Increasing the confidence level would increase the margin of error hence widen the interval.

(e) Increasing the sample size would decrease the margin of error hence make the interval narrower.

Problem #4 ($4 + 4 = 8$ points)

Solve **Exercise 6.14** from the textbook.

Solution:

```
knitr::include_graphics("oc-p6-14.png")
```

(a) We have previously confirmed that the independence condition is satisfied. We need to recheck the success-failure condition using the sample proportion: $331 \times 0.48 = 158.88 > 10$ and $331 \times 0.52 = 172.12 > 10$. An 80% confidence interval can be calculated as follows:

$$\begin{aligned}\hat{p} \pm z^{\star}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} &= 0.48 \pm 1.65 \times \sqrt{\frac{0.48 \times 0.52}{331}} \\ &= 0.48 \pm 1.65 \times 0.0275 \\ &= 0.48 \pm 0.045 \\ &= (0.435, 0.525)\end{aligned}$$

We are 90% confident that the 43.5% to 52.5% of all Americans who decide not to go to college do so because they cannot afford it. This agrees with the conclusion of the earlier hypothesis test since the interval includes 50%.

(b) We are asked to solve for the sample size required to achieve a 1.5% margin of error for a 90% confidence interval and the point estimate is $\hat{p} = 0.48$.

$$\begin{aligned}ME = z^{\star}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \;\rightarrow\; 0.01 &\geq 1.65\sqrt{\frac{0.48 \times 0.52}{n}} \\ 0.015^2 &\geq 1.65^2 \frac{0.48 \times 0.52}{n} \\ n &\geq \frac{1.65^2 \times 0.48 \times 0.52}{0.015^2} \\ n &\geq 3020.16 \approx 3121\end{aligned}$$

The sample size n should be at least 3,121.

Problem #5 (4 points)

Solve **Exercise 6.16** from the textbook.

Solution:

```r
knitr::include_graphics("oc-p6-16.png")
```

6.16 We are asked to solve for the sample size required to achieve a 2% margin of error for a 95% confidence interval and the point estimate is $\hat{p} = 0.61$.

$$\begin{aligned}ME = z^{\star}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \;\rightarrow\; 0.02 &\geq 1.96\sqrt{\frac{0.61 \times 0.39}{n}} \\ 0.02^2 &\geq 1.96^2 \frac{0.61 \times 0.39}{n} \\ n &\geq \frac{1.96^2 \times 0.61 \times 0.39}{0.02^2} \\ n &\geq 2284.792 \\ n &\geq 2285\end{aligned}$$

The sample size n should be at least 2,285.

Problem #6 (4 points)

In a random sample of 1000 small children, it was found that 880 of them observe Halloween. Find the 80%-confidence interval for the population proportion of childred who observe Halloween.

Solution: Let p denote the population parameter denoting the probability that a randomly chosen child from the population observes Halloween. The point estimate for p based on our data is $\hat{p} = \frac{880}{1000} = 0.88$.

The critical value z^* associated with the 80% confidence level is

$$z^* = \Phi^{-1}(0.90) = 1.28.$$

In our usual notation, the standard error is

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.88(0.12)}{1000}} = 0.01028.$$

Hence, the margin of error is $1.28(0.01028) = 0.01315$. The 80%-confidence interval is

$$p = 0.88 \pm 0.01315.$$

Problem #7 (4 points)

Let p denote the population proportion. How large should the sample size be so that one is at least 95% confident that the true parameter p is within a 0.02 margin of error from the point estimate?

Solution: Let m denote the upper bound on the margin of error. Then,

$$n \geq \left(\frac{z^*}{2m}\right)^2 = \left(\frac{1.96}{2(0.02)}\right)^2 = (49)^2 = 2401.$$