Practice 7

Elena Tuzhilina

February 28, 2023

Question 1

Alice took a test with 100 Yes/No questions (a very long one!).

1. Let Y be the random variable representing the *proportion* of questions that Alice correctly answered. If p is the probability for Alice to answer one question correctly, what is the expectation and variance of Y?

Note that $Y = \frac{X_1 + \ldots + X_{100}}{100}$ where $X_i \sim Bernoulli(p)$. Thus $E(Y) = p$ and $Var(Y) = \frac{p(1-p)}{n}$.

2. What is the approximate distribution of Y?

From the CLT we get that $Y \sim Normal(p, \frac{p(1-p)}{n})$.

3. Alice received the tests results and she got 65 questions correctly. Use this information to find the estimates for $E(Y)$ and $Var(Y)$.

$E(Y) \approx \frac{65}{100} = 0.65$ and $Var(Y) \approx \frac{0.65(1-0.65)}{100} = 0.002275$

4. The Professor wants to understand if Alice was randomly guessing the answers on the test. To do so, the Professor decided to compute the 95% confidence interval for p. Find this interval.

$[0.65 - 1.96 \cdot \sqrt{0.002275}, 0.65 + 1.96 \cdot \sqrt{0.002275}] = [0.557, 0.743]$

5. If Alice was randomly guessing the answers on the test, what would be the value of p?

$p = 0.5$.

6. Use 4 and 5 to answer the following question: is it likely that Alice was randomly guessing the answers on the test?

As it [0.557,0.743] does not cover 0.5, it is unlikely that Alice randomly guessed the answers.

7. Now compute the 99% confidence interval for p. Is it wider that the 90% confidence interval?

The 99% confidence interval is:

$[0.65 - 2.58 \cdot \sqrt{0.002275}, 0.65 + 2.58 \cdot \sqrt{0.002275}] = [0.527, 0.773]$.

It is wider than the 95% confidence interval.

8. Are we still sure that Alice did not randomly guessed the answers?

Yes we are.

Question 2

Elon Musk wants to estimate the average salary in startups in the Silicone Valley.

1. An insider shared that the salary in a Silicone Valley startup ranges between 150K and 450K and follows uniform distribution. Use the fact that $E(X) = \frac{a+b}{2}$ and $Var(X) = \frac{(b-a)^2}{12}$ for $X \sim Uniform(a, b)$ to compute the expectation and variance of the salary in a Silicone Valley startup.

Since $X \sim Uniform(150, 400)$ then $E(X) = \frac{150+450}{2} = 300$ and $Var(X) = \frac{(450-150)^2}{12} = 7500$.

2. Let Y be the random variable representing the average salary in 30 randomly chosen startups. What would be the distribution of Y?

From the CLT $Y \sim Normal(300, \frac{7500}{30}) = Normal(300, 250)$.

3. Elon Musk knows the salaries in 30 startups in the Silicon Valley.

```
salary
```

```
##  [1] 229.7 261.6 321.9 422.5 210.5 419.5 433.4 348.2 338.7 168.5 211.8 203.0
## [13] 356.1 265.2 381.0 299.3 365.3 447.6 264.0 383.2 430.4 213.6 345.5 187.7
## [25] 230.2 265.8 154.0 264.7 410.9 252.1
```

He computed the mean of this sample:

```
mean(salary)
```

```
## [1] 302.8633
```

In addition, an insider shared that the variance of the salary in a Silicone Valley startup is 7500.

Use this information to help Elon Musk computing the 80% confidence interval for the average salary in startups in the Silicone Valley.

Since the theoretical variance is known, we use Normal distribution table and get

$[302.86 - 1.28 \cdot \sqrt{\frac{7500}{30}}, 302.86 + 1.28 \cdot \sqrt{\frac{7500}{30}}] = [282.62, 323.1]$

4. Elon Musk just bought a startup and he wants to attract more talented engineers to the company. Use answer in the previous question to recommend Elon Musk the good salary for the startup.

More than 323.1.

5. Elon Musk doesn't trust the insider anymore, so he computed the variance of his sample.

```
var(salary)
```

```
## [1] 7847.569
```

Use this information to help Elon Musk computing the 80% confidence interval for the average salary in startups in the Silicone Valley. Did we get wider confidence interval?

Since the theoretical variance is unknown, we use t-distribution table and get

$[302.86 - 1.31 \cdot \sqrt{\frac{7847.57}{30}}, 302.86 + 1.31 \cdot \sqrt{\frac{7847.57}{30}}] = [282.15, 323.57]$.

Yes, it is slightly wider.