14.310x: Data Analysis for Social Scientists
Special Distributions, the Sample Mean, and the Central Limit Theorem

Welcome to your sixth homework assignment! You will have about one week to work through the assignment. We encourage you to get an early start, particularly if you still feel you need more experience using R. We have provided this PDF copy of the assignment so that you can print and work through the assignment offline. You can also go online directly to complete the assignment. If you choose to work on the assignment using this PDF, please go back to the online platform to submit your answers based on the output produced. Some of the questions we are asking are not easily solvable using math so we recommend you to use your R knowledge and the content of previous homework assignments to find numeric solutions.

Good luck!

Suppose that $X_i \; i.i.d. \; U[0, \theta]$. You want to build a 90% confidence interval for θ. To do so, you will need an estimator for θ and you will need to know the estimator's distribution. Let's consider $\hat{\theta} = \frac{n+1}{n} X_{(n)}$. (Remember that $X_{(n)}$ is the nth order statistic). This estimator is a variant on the MLE. We have used the n^{th} order statistic, which is the MLE, but multiplied it by $\frac{n+1}{n}$ to remove its bias. Its PDF is $\frac{n^{n+1}}{(n+1)^n} \frac{x^{n-1}}{\theta^n}$ for $x \in [0, \frac{n+1}{n}\theta]$ and 0 otherwise.

Question 1
Let *a* be a function of n and θ such that 5% of the distribution of $\hat{\theta}$ is to the left of *a*.

a is then given by

$$a = \sqrt[n]{A} \frac{B}{C} \theta$$

What is the value of A?

- θ
- n
- $n + 1$
- 0.05

What is the value of B?

- θ
- n
- $n + 1$
- 0.05

What is the value of C?

- θ
- n
- $n + 1$
- 0.05

Question 2

Let b be a function of n and θ such that 5% of the distribution of $\hat{\theta}$ is to the right of b.

b is then given by

$$b = \sqrt[n]{A}\frac{B}{C}\theta$$

What is the value of A?

- θ
- n
- $n+1$
- 0.95

What is the value of B?

- θ
- n
- $n+1$
- 0.95

What is the value of C?

- θ
- n
- $n+1$
- 0.95

Question 3

Using those values that you found above, what is a probability statement of the form $P(a < \hat{\theta} < b) = .90$ as a function of n and θ?

- $P\left(\sqrt[n]{0.95}\,\frac{n+1}{n}\theta \le \hat{\theta} \le \sqrt[n]{0.05}\,\frac{n+1}{n}\theta\right) = .90$
- $P\left(\sqrt[n]{0.95}\,\frac{n}{n+1}\theta \le \hat{\theta} \le \sqrt[n]{0.05}\,\frac{n+1}{n}\theta\right) = .90$
- $P\left(\sqrt[n]{0.05}\,\frac{n}{n+1}\theta \le \hat{\theta} \le \sqrt[n]{0.95}\,\frac{n+1}{n}\theta\right) = .90$
- $P\left(\sqrt[n]{0.05}\,\frac{n+1}{n}\theta \le \hat{\theta} \le \sqrt[n]{0.95}\,\frac{n+1}{n}\theta\right) = .90$

Question 4

If you rearrange the quantities in the probability statement so that θ is alone in the middle, bracketed by functions of the random sample and known quantities, what would be this probability statement?

- $P\left(\frac{\hat{\theta}}{\sqrt[n]{0.95}}\frac{n}{n+1} \le \theta \le \frac{\hat{\theta}}{\sqrt[n]{0.05}}\frac{n+1}{n+n}\right) = .90$
- $P\left(\frac{\hat{\theta}}{\sqrt[n]{0.95}}\frac{n}{n+1} \le \theta \le \frac{\hat{\theta}}{\sqrt[n]{0.05}}\frac{n}{n+1}\right) = .90$
- $P\left(\frac{\hat{\theta}}{\sqrt[n]{0.95}}\frac{n}{n+1} \le \theta \le \frac{\hat{\theta}}{\sqrt[n]{0.05}}\frac{n+1}{n}\right) = .90$

- $P\left(\frac{\hat{\theta}}{\sqrt[n]{0.05}}\frac{n}{n+1} \leq \theta \leq \frac{\hat{\theta}}{\sqrt[n]{0.95}}\frac{n}{n+1}\right) = .90$

We are going to show this in R. We have provided you with this code that demonstrates this.

Question 5
In the code, there are three symbols standing in for specific values: XXX, YYY, and ZZZ. Which of the following values correspond to XXX, YYY, and ZZZ?

- $XXX = \theta, YYY = 1, ZZZ = \&$
- $XXX = \theta, YYY = 1, ZZZ = |$
- $XXX = \theta, YYY = 2, ZZZ = \&$
- $XXX = n, YYY = 2, ZZZ = \&$
- $XXX = n, YYY = 2, ZZZ = |$

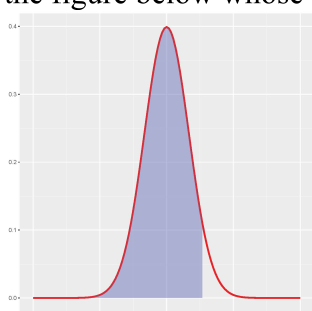
We invite you to run this code to see that it is true that in 90% of the simulated samples this confidence interval (CI) contains the real value of θ. You can play with the code, changing both the value of θ and the sample size.

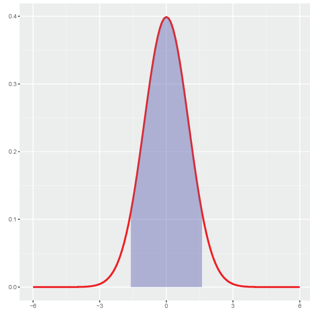
Now suppose X_i $i.i.d. N(\mu, 4)$ and $n = 25$. We want to test $H_0: \mu = 0$ $vs. H_a: \mu \neq 0$.

Question 6
Given the information above, what test statistic should you use?

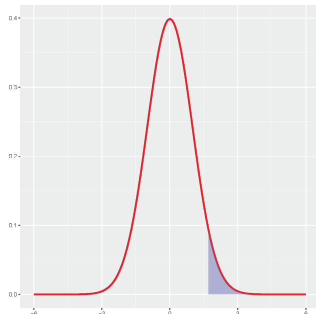
- The minimum
- The sample variance
- The maximum
- The sample mean

Question 7
Choose the figure below whose shaded region corresponds to the critical region.

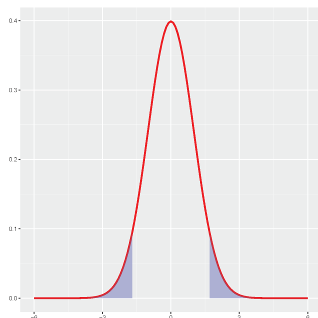
- 



○



○



○

Question 8

Compute α (the probability that we reject the null hypothesis when the null is true) as a function of the critical value(s).

Note: A critical value is the boundary between the critical region and the rest of the sample space. In the example in class, we denoted the critical value k.

$\alpha = \left(1 - \Phi\left(\frac{2k}{5}\right)\right)$

$\alpha = 2\left(1 - \Phi\left(\frac{5k}{2}\right)\right)$

$\alpha = \left(1 - \Phi\left(\frac{5k}{2}\right)\right)$

$\alpha = 2\left(1 - \Phi\left(\frac{2k}{5}\right)\right)$

Question 9

Is there a connection between conducting a hypothesis test for $\mu = \mu_0$ at a significance level α and finding a $(1 - \alpha)$ confidence interval for a parameter μ?

Yes
No

Homework due Oct 19, 2020 19:30 EDT **Past Due**

Suppose that X_i i.i.d. $U[0,\theta]$. You want to build a 90% confidence interval for θ. To do so, you will need an estimator for θ and you will need to know the estimator's distribution. Let's consider $\hat{\theta} = \frac{n+1}{n} X_{(n)}$. (Remember that $X_{(n)}$ is the nth order statistic.) This estimator is a variant on the MLE. We have used the n^{th} order statistic, which is the MLE, but multiplied it by $\frac{n+1}{n}$ to remove its bias. Its PDF is $\frac{n^{n+1}}{(n+1)^n} \frac{x^{n-1}}{\theta^n}$ for $x \in [0, \frac{n+1}{n}\theta]$ and 0 otherwise.

Question 1

0.0/1.0 point (graded)

Let a be a function of n and θ such that 5% of the distribution of $\hat{\theta}$ is to the left of a.

a is then given by

$$a = \sqrt[n]{A}\frac{B}{C}\theta$$

What is the value of A?

- ◯ θ
- ◯ n
- ◯ $n+1$

- 0.05 ✔

What is the value of B?

- $n + 1$ ✔
- 0.05
- θ
- n

What is the value of C?

- n ✔
- 0.05
- θ

○ $n+1$

Explanation

$$Pr\left(\hat{\theta} \leq a\right) = 0.05$$

$$\int_0^a \frac{n^{n+1}}{(n+1)^n}\frac{x^{n-1}}{\theta^n}dx = 0.05$$

$$\frac{n^{n+1}}{(n+1)^n}\frac{a^n}{n\theta^n} = 0.05$$

$$\left(\frac{an}{\theta(n+1)}\right)^n = 0.05$$

$$\frac{an}{\theta(n+1)} = (0.05)^{\frac{1}{n}}$$

$$a\left(\theta, n\right) = \sqrt[n]{0.05}\frac{n+1}{n}\theta$$

Show answer

Question 2

0.0/1.0 point (graded)

Let b be a function of n and θ such that 5% of the distribution of $\hat{\theta}$ is to the right of b.

b is then given by

$$b = \sqrt[n]{A}\frac{B}{C}\theta$$

What is the value of A?

- ○ θ
- ○ n
- ○ $n+1$
- ○ 0.95 ✔

What is the value of B?

- ○ $n+1$ ✔

- 0.95
- θ
- n

What is the value of C?

- n ✔
- 0.95
- θ
- $n+1$

Explanation

We find the value of b such that:

$$Pr\left(\hat{\theta} \geq b\right) = 0.05$$

$$\int_b^{\frac{n+1}{n}\theta} \frac{n^{n+1}}{(n+1)^n} \frac{x^{n-1}}{\theta^n} dx = 0.05$$

$$\frac{n^{n+1}}{(n+1)^n} \left(\frac{(n+1)^n \theta^n}{n^{n+1} \theta^n} - \frac{b^n}{n\theta^n} \right) = 0.05$$

$$b(\theta, n) = \sqrt[n]{0.95} \frac{n+1}{n} \theta$$

Show answer

Submit You have used 0 of 2 attempts

Answers are displayed within the problem

Question 3

0.0/1.0 point (graded)

Using those values that you found above, what is a probability statement of the form $P\left(a < \hat{\theta} < b\right) = .90$ as a function of n and θ?

- $P\left(\sqrt[n]{0.95}\frac{n+1}{n}\theta \leq \hat{\theta} \leq \sqrt[n]{0.05}\frac{n+1}{n}\theta\right) = 0.9$

- $P\left(\sqrt[n]{0.95}\frac{n}{n+1}\theta \leq \hat{\theta} \leq \sqrt[n]{0.05}\frac{n+1}{n}\theta\right) = 0.9$
- $P\left(\sqrt[n]{0.05}\frac{n}{n+1}\theta \leq \hat{\theta} \leq \sqrt[n]{0.95}\frac{n+1}{n}\theta\right) = 0.9$
- $P\left(\sqrt[n]{0.05}\frac{n+1}{n}\theta \leq \hat{\theta} \leq \sqrt[n]{0.95}\frac{n+1}{n}\theta\right) = 0.9$ ✔

Explanation

We know that $P\left(a \leq \hat{\theta} \leq b\right) = 0.9$. So we plug in the values for $a\left(\theta, n\right)$ and $b\left(\theta, n\right)$ that we found above.

Show answer

Submit You have used 0 of 2 attempts

Answers are displayed within the problem

Question 4

0.0/1.0 point (graded)

If you rearrange the quantities in the probability statement so that θ is alone in the middle, bracketed by functions of the random sample and known quantities, what would be this probability statement?

- ○ $P\left(\frac{\hat{\theta}}{\sqrt[n]{0.95}}\frac{n}{n+1} \leq \theta \leq \frac{\hat{\theta}}{\sqrt[n]{0.05}}\frac{n+1}{n+n}\right) = 0.9$
- ○ $P\left(\frac{\hat{\theta}}{\sqrt[n]{0.95}}\frac{n}{n+1} \leq \theta \leq \frac{\hat{\theta}}{\sqrt[n]{0.05}}\frac{n}{n+1}\right) = 0.9$ ✔
- ○ $P\left(\frac{\hat{\theta}}{\sqrt[n]{0.95}}\frac{n}{n+1} \leq \theta \leq \frac{\hat{\theta}}{\sqrt[n]{0.05}}\frac{n+1}{n}\right) = 0.9$
- ○ $P\left(\frac{\hat{\theta}}{\sqrt[n]{0.05}}\frac{n}{n+1} \leq \theta \leq \frac{\hat{\theta}}{\sqrt[n]{0.95}}\frac{n}{n+1}\right) = 0.9$

Explanation

The calculated interval from question 3 is $\sqrt[n]{0.05}\frac{n+1}{n}\theta \leq \hat{\theta} \leq \sqrt[n]{0.95}\frac{n+1}{n}\theta$. To turn this into a 90% confidence interval for θ, we need to isolate θ from the two inequalities: $\sqrt[n]{0.05}\frac{n+1}{n}\theta \leq \hat{\theta} \implies \theta \leq \frac{\hat{\theta}}{\sqrt[n]{0.05}}\frac{n}{n+1}$, and $\hat{\theta} \leq \sqrt[n]{0.95}\frac{n+1}{n}\theta \implies \frac{\hat{\theta}}{\sqrt[n]{0.95}}\frac{n}{n+1} \leq \theta$. Therefore $P\left(\sqrt[n]{0.05}\frac{n+1}{n}\theta \leq \hat{\theta} \leq \sqrt[n]{0.95}\frac{n+1}{n}\theta\right) = 0.9 \implies P\left(\frac{\hat{\theta}}{\sqrt[n]{0.95}}\frac{n}{n+1} \leq \theta \leq \frac{\hat{\theta}}{\sqrt[n]{0.05}}\frac{n}{n+1}\right) = 0.9$ This is your 90% confidence interval for θ. (If you were to draw 100 different samples and construct 100 such confidence intervals, you would expect 90 of them to contain the true value of θ.)

Question 5

0.0/1.0 point (graded)

In the code, there are three symbols standing in for specific values: XXX, YYY, and ZZZ. Which of the following values correspond to XXX, YYY, and ZZZ?

- ○ XXX$=\theta$, YYY$=1$, ZZZ = &
- ○ XXX$=\theta$, YYY$=1$, ZZZ$=|$
- ○ XXX$=\theta$, YYY$=2$, ZZZ = & ✔
- ○ XXX$=n$, YYY$=2$, ZZZ = &
- ○ XXX$=n$, YYY$=2$, ZZZ$=|$

Explanation

XXX corresponds to the maximum value of the uniform distribution we are simulating, which should be θ. We are calculating the maximum value through the columns since each column represents a different sample size of size n. Thus, we should use the apply function over the columns. This implies that YYY$=2$. Finally, our confidence interval contains the real value of θ if $\theta \geq \frac{\hat{\theta}}{\sqrt[n]{0.95}\frac{n}{n+1}}$ and $\theta \leq \frac{\hat{\theta}}{\sqrt[n]{0.05}\frac{n}{n+1}}$. Thus, ZZZ should be equal to &.

$\alpha = \left(1 - \Phi\left(\frac{5k}{2}\right)\right)$

$\alpha = 2\left(1 - \Phi\left(\frac{2k}{5}\right)\right)$

Explanation

α is the probability that we reject the null hypothesis when the null is true:

$$\alpha = P\left(\overline{X} \le -k|\mu = 0\right) + P\left(\overline{X} \ge k|\mu = 0\right).$$

Under the null hypothesis $\overline{X} \sim N\left(0, \frac{4}{25}\right)$. Hence $\alpha = \Phi\left(-\frac{5k}{2}\right) + 1 - \Phi\left(\frac{5k}{2}\right) = 2\left(1 - \Phi\left(\frac{5k}{2}\right)\right)$.

Show answer

Submit You have used 0 of 2 attempts

Answers are displayed within the problem

Question 9

Question 9

0.0/1.0 point (graded)

Is there a connection between conducting a hypothesis test for $\mu = \mu_0$ at significance level α and finding a $(1-\alpha)$ confidence interval for a parameter μ ?

- ◯ Yes ✔
- ◯ No

Explanation

Yes, there is a connection.

A $1-\alpha$ confidence interval for μ from a normal population of known variance is such that $P\left(\mu_0 \in CI_{1-\alpha}\right) = 1-\alpha$. The critical values for a test of size α are the boundaries for the confidence interval centered at $\mu = 0$.

Thus, if a confidence interval realization doesn't contain a parameter μ_0 we can reject the hypothesis $\mu = \mu_0$ with probability α of having committed a Type 1 error.

Show answer

Submit You have used 0 of 1 attempt

Answers are displayed within the problem