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.

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?
- o θ
- o n
- o $n+1$
- o 0.05

What is the value of B?
- o θ
- o n
- o $n+1$
- o 0.05

What is the value of C?
- o θ
- o n
- o $n+1$
- o 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?

- o θ
- o n
- o $n + 1$
- o 0.95

What is the value of B?

- o θ
- o n
- o $n + 1$
- o 0.95

What is the value of C?

- o θ
- o n
- o $n + 1$
- o 0.95

Question 3

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?

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

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?

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

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

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?

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

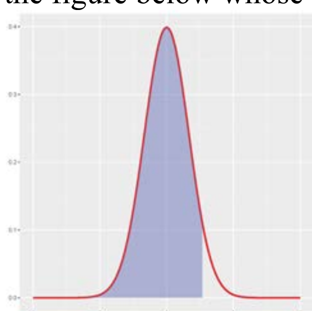
We invite you to run this code to see that it is true that in 90% of the simulated samples this confidence interval (CI) contains the real value of θ. You can play with the code, changing both the value of θ and the sample size.

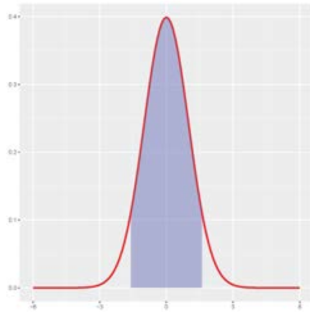
Now suppose $X_i\ i.i.d.N(\mu, 4)$ and $n = 25$. We want to test $H_0: \mu = 0\ vs. H_a: \mu \neq 0$.

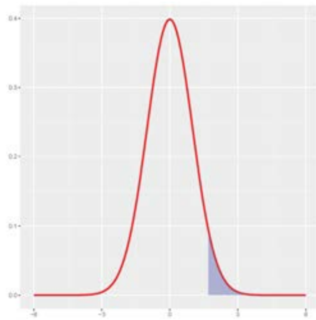
Question 6
Given the information above, what test statistic should you use?

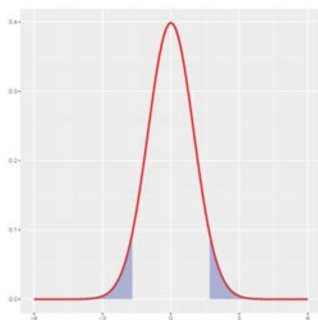
- The minimum
- The sample variance
- The maximum
- The sample mean

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

- 

o 

o 

o 

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
