**14.310x: Data Analysis for Social Scientists - Homework 7**

Welcome to your seventh 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 homeworks 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 $\theta$. To do so, you will need an estimator for $\theta$ 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 nth 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.

1. What is the value $a$ such that 5% of the distribution of $\hat{\theta}$ is to the left of $a$? (It will be a function of $n$ and $\theta$.)

   (a) It is given by $\sqrt[n]{0.1}\frac{n}{n+1}\theta$
   (b) It is given by $\sqrt[n]{0.05}\frac{n}{n+1}\theta$
   (c) It is given by $\sqrt[n]{0.1}\frac{n+1}{n}\theta$
   (d) It is given by $\sqrt[n]{0.05}\frac{n+1}{n}\theta$

2. What is the value $b$ such that 5% of the distribution of is to the right of $b$?

   (a) It is given by $\sqrt[n]{0.9}\frac{n+1}{n}\theta$
   (b) It is given by $\sqrt[n]{0.9}\frac{n}{n+1}\theta$
   (c) It is given by $\sqrt[n]{0.95}\frac{n+1}{n}\theta$
   (d) It is given by $\sqrt[n]{0.95}\frac{n}{n+1}\theta$

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 $\theta$.

   (a) It is $P\left(\sqrt[n]{0.05}\frac{n}{n+1}\theta \leq \hat{\theta} \leq \sqrt[n]{0.95}\frac{n}{n+1}\theta\right) = 0.9$
   (b) It is $P\left(\sqrt[n]{0.1}\frac{n}{n+1}\theta \leq \hat{\theta} \leq \sqrt[n]{0.9}\frac{n}{n+1}\theta\right) = 0.9$
   (c) It is $P\left(\sqrt[n]{0.05}\frac{n}{n+1}\theta \leq \hat{\theta} \leq \sqrt[n]{0.9}\frac{n}{n+1}\theta\right) = 0.9$
   (d) It is $P\left(\sqrt[n]{0.1}\frac{n}{n+1}\theta \leq \hat{\theta} \leq \sqrt[n]{0.95}\frac{n}{n+1}\theta\right) = 0.9$

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

(a) It is $P\left(\frac{\hat{\theta}}{\sqrt[n]{0.05}\frac{n}{n+1}} \le \theta \le \frac{\hat{\theta}}{\sqrt[n]{0.95}\frac{n}{n+1}}\right) = 0.9$

(b) It is $P\left(\frac{\hat{\theta}}{\sqrt[n]{0.95}\frac{n}{n+1}} \le \theta \le \frac{\hat{\theta}}{\sqrt[n]{0.05}\frac{n}{n+1}}\right) = 0.9$

(c) It is $P\left(\frac{\hat{\theta}}{\sqrt[n]{0.9}\frac{n}{n+1}} \le \theta \le \frac{\hat{\theta}}{\sqrt[n]{0.1}\frac{n}{n+1}}\right) = 0.9$

(d) It is $P\left(\frac{\hat{\theta}}{\sqrt[n]{0.95}\frac{n}{n+1}} \le \theta \le \frac{\hat{\theta}}{\sqrt[n]{0.1}\frac{n}{n+1}}\right) = 0.9$

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

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?

   (a) XXX= $n$, YYY= 2, ZZZ= |
   (b) XXX= $\theta$, YYY= 1, ZZZ=&
   (c) XXX= $\theta$, YYY= 2, ZZZ=&
   (d) XXX= $n$, YYY= 2, ZZZ=&
   (e) XXX= $\theta$, YYY= 1, 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 $\theta$. You can play with the code, changing both the value of $\theta$ and the sample size.

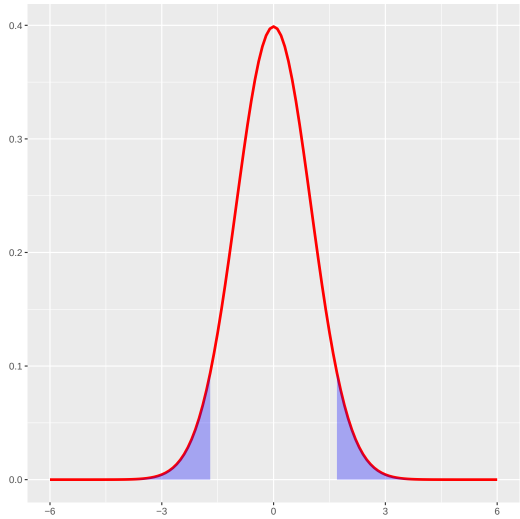
Now suppose $X_i$ i.i.d. $N(\mu, 4)$ and $n = 25$. We want to test $H_0 : \mu = 0$ vs. $H_a : \mu \neq 0$.

6. What test statistic would you propose using?

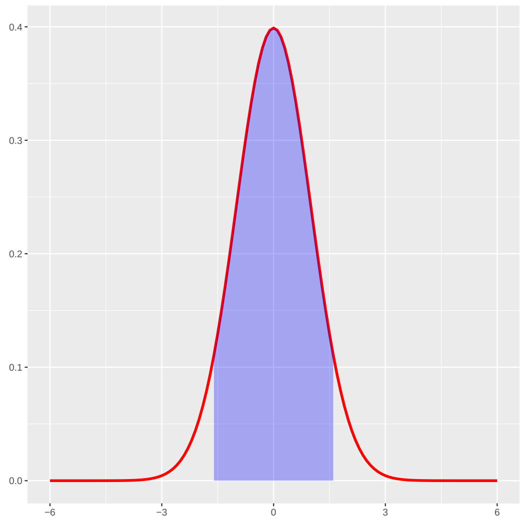
   (a) The sample mean
   (b) The sample variance
   (c) The maximum
   (d) The minimum

7. What shape will the critical region have? In other words, for what values of the test statistic would we want to reject the null in favor of the alternative? [shaded area corresponds to rejection]

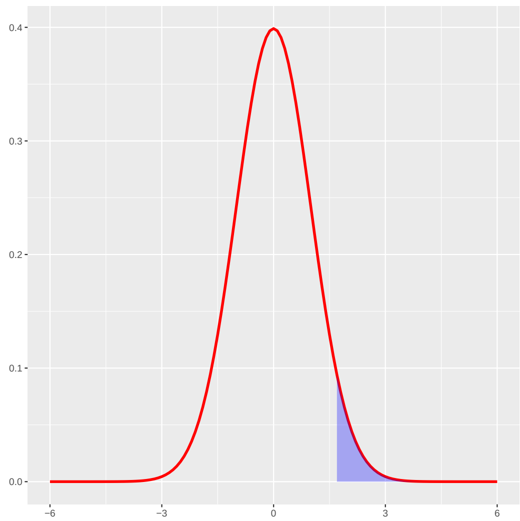
   (a) It is given by:



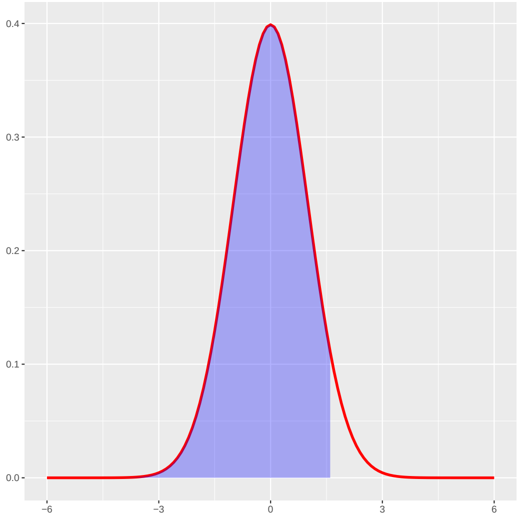

(b) It is given by:




(c) It is given by:



(d) It is given by:




8. Compute $\alpha$ (the probability that we reject the null hypothesis when the null is true) as a function of the critical value(s). (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 $k$.)

   (a) It is given by $2\left(1-\Phi\left(\frac{2k}{5}\right)\right)$

   (b) It is given by $2\left(1-\Phi\left(\frac{5k}{2}\right)\right)$

   (c) It is given by $\left(1-\Phi\left(\frac{5k}{2}\right)\right)$

   (d) It is given by $\left(1-\Phi\left(\frac{2k}{5}\right)\right)$

9. Is there a connection between the procedure for finding $\alpha$ and the construction of a confidence interval for $\mu$?

   (a) Yes

   (b) No