

DDS For Live Session

UNIT 1

Philosophy

Due at 1pm CST the day of live session.

- On the next slide you will see activities and the estimated / expected time that the student should spend on that activity.
- It is important to note that the goal of the activities is to become familiar with the methods, ideas and implementation involved in that activity so that we can efficiently iron out all the details in live session.
- Analogy: You are building the pieces of puzzle in the For Live Session Activity and we are putting them together to see the big picture in live session.
- It is **not** expected that the student have all the correct answers. The expectation is that each student spend the allotted time (indicated next to the activity) on each activity so that we can discuss the details in live session.
- If you max out the indicated time without finishing the activity and you don't have more time to finish, simply write up what you have learned by that time and record any questions you might have and we will address those in live session!
- We want to develop the questions before live session so that we can use the live session time to effectively answer them!

For Live Session: Unit 1

Due at 1pm CST the day of live session.

For each number, provide at least 1 PowerPoint slide to provide a clear presentation of your response. You will present most of these slides in a break out and are given a score based on their completeness and clarity. Most importantly, they will maximize your live session experience!

1. **Data Science Profile:** Make a bar plot for your data science profile. (1-2 hours) The Data Science Profile is described in the Cathy O'Neil reading for Unit 1 from *Doing Data Science*.
2. **CLT:** Using the central limit code from the asynch material (CLT2.Rmd), adapt the code and provide the following analysis: (2-3 hours)
 1. Adapt the code to generate a population of 10,000,000 from a chi-square distribution with 2 degrees of freedom. This is a heavily right skewed distribution. (Hint: `rchisq()`). You will have to read up on this function and probably do some trial and error. Being able to learn new functions and methods is a key skill.
 2. Provide a histogram of this population... display the right skewness.
 3. Record the mean and standard deviation of this population.
 4. According to the central limit theorem, what should be the approximate distribution of sample means of size 50 from this right skewed population? What should be the mean and standard error of the mean (standard deviation of the distribution of sample means)?
 5. Now let's check this: Adapt the CLT code to draw 10,000 means each of size 50 from this population and provide the sampling distribution of this sample mean. Provide a histogram of these 10,000 sample means.
 6. What is the mean and standard deviation of these 10,000 sample means?
3. **Ask ChatGPT (you pick the version) to try and address the previous question.**
4. **T-Test** (2-3 hours)

Research in R how to conduct a T-test (`t.test`) and conduct a six step hypothesis test to answer the question on the next slide. If you would like to brush up on hypothesis tests, please see the *Bridge Course for Statistics*.
5. **Takeaways and/or Questions:** What were your key – takeaways from from this unit and what were any question or comments you would like to make? Your professor will use these to customize live session. (1 hour)

1 Hypothesis Test



The following are ages of 7 randomly chosen patrons seen leaving the Beach Comber in South Mission Beach at 7pm! We assume that the data come from a normal distribution and would like to test the claim that the mean age of the distribution of Comber patrons is different than 21. Conduct a 6 step hypothesis test to test this claim.

25, 19, 37, 29, 40, 28, 31