

# EPSY 5261 : Introductory Statistical Methods

**Day 9**

**Introduction to Hypothesis Testing Using Simulation**

# Learning Goals

- At the end of this lesson, you should be able to...
  - List the steps of a hypothesis test
  - Describe the purpose of a hypothesis test
  - Describe a parametric approach to hypothesis testing for a single mean
  - List the assumptions for using the t-distribution to test a single mean



# Hypothesis Testing

Purpose: to test a claim about a population parameter



# Steps of Hypothesis Testing

1. Formulate a **research question**
2. Write your **hypotheses**
3. Find **Distribution** of the Null Hypothesis
4. **Compare** Sample to the Distribution of Null Hypothesis
5. Get a **p-value**
6. Make a **decision** to reject or fail to reject the p-value
7. Communicate your **conclusion** in context



# Steps of Hypothesis Testing

1. Formulate a **research question**
2. Write your **hypotheses**
3. Find **Distribution** Considering the Null Hypothesis
4. **Compare** Sample to the Distribution of Null Hypothesis
5. Get a **p-value**
6. Make a **decision** to reject or fail to reject the p-value
7. Communicate your **conclusion** in context



# Estimating a Distribution

- Goal: get an estimate for the sampling variability expected given this sample
- Simulation (resampling methods)
- Traditional Parametric Methods (a mathematical function)



# Estimating a Distribution

- Simulation (resampling methods)
- Traditional Parametric Methods (a mathematical function)



# Theoretical Distribution

- Up until now, we have looked at *approximations* of the sampling distribution, with simulation

Now, we will look at theoretical distributions of sample statistics

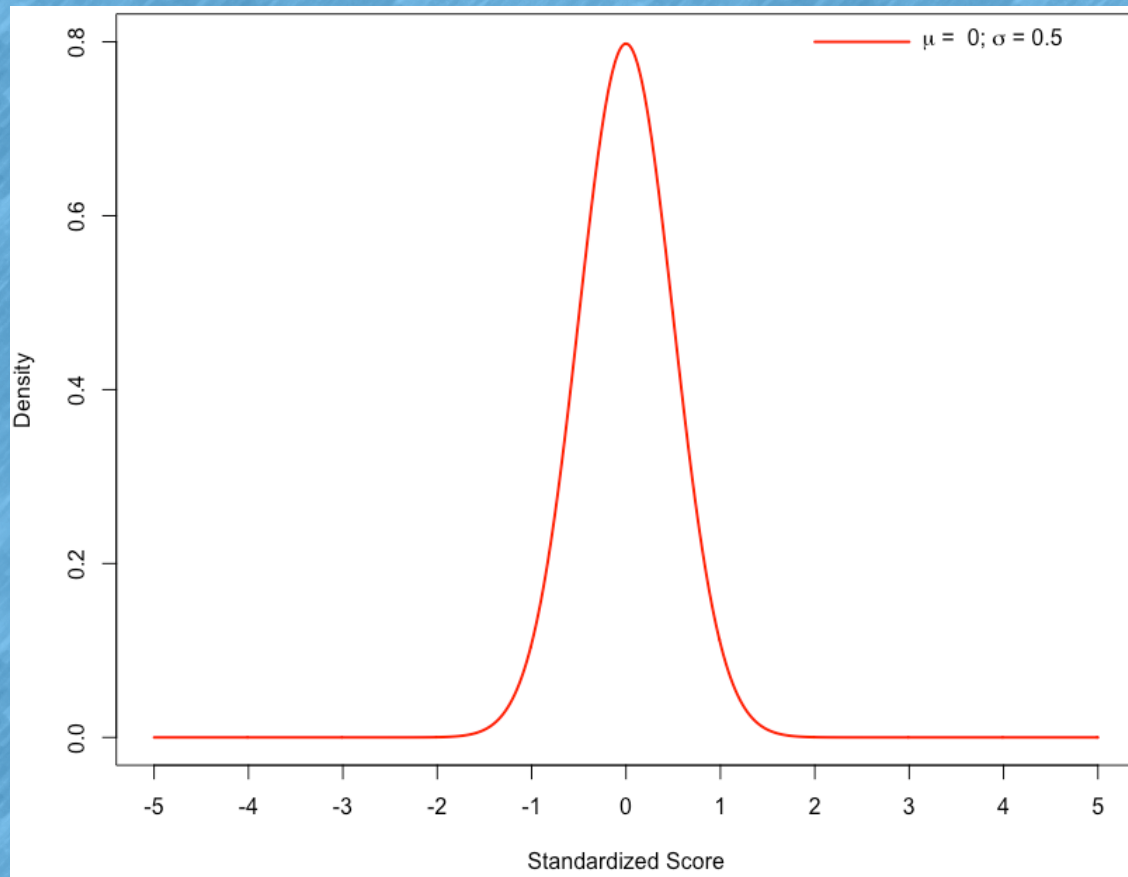


# Normal Distributions

- Normal distributions are bell shaped & symmetric distributions characterized by:
  - Mean (center)
  - Standard deviation (estimate of variability)

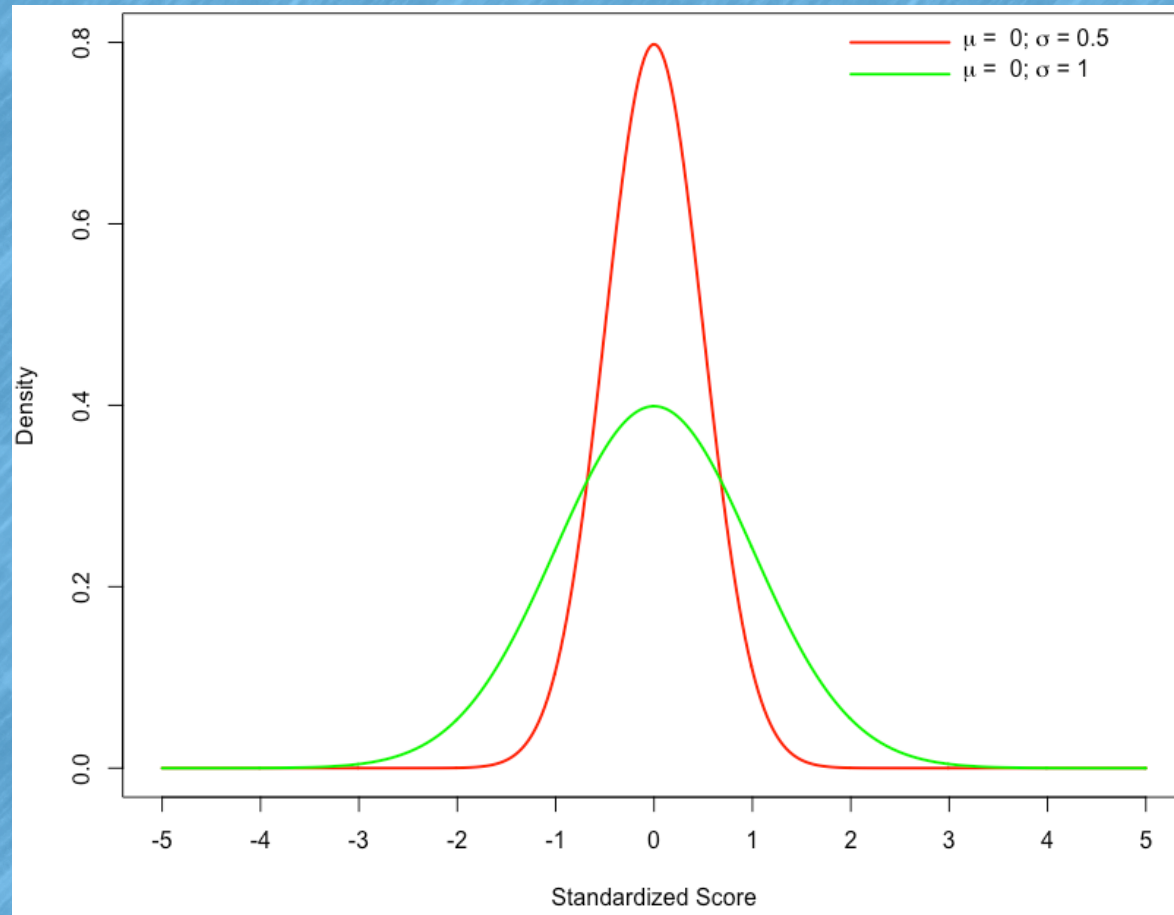


# Normal Distributions



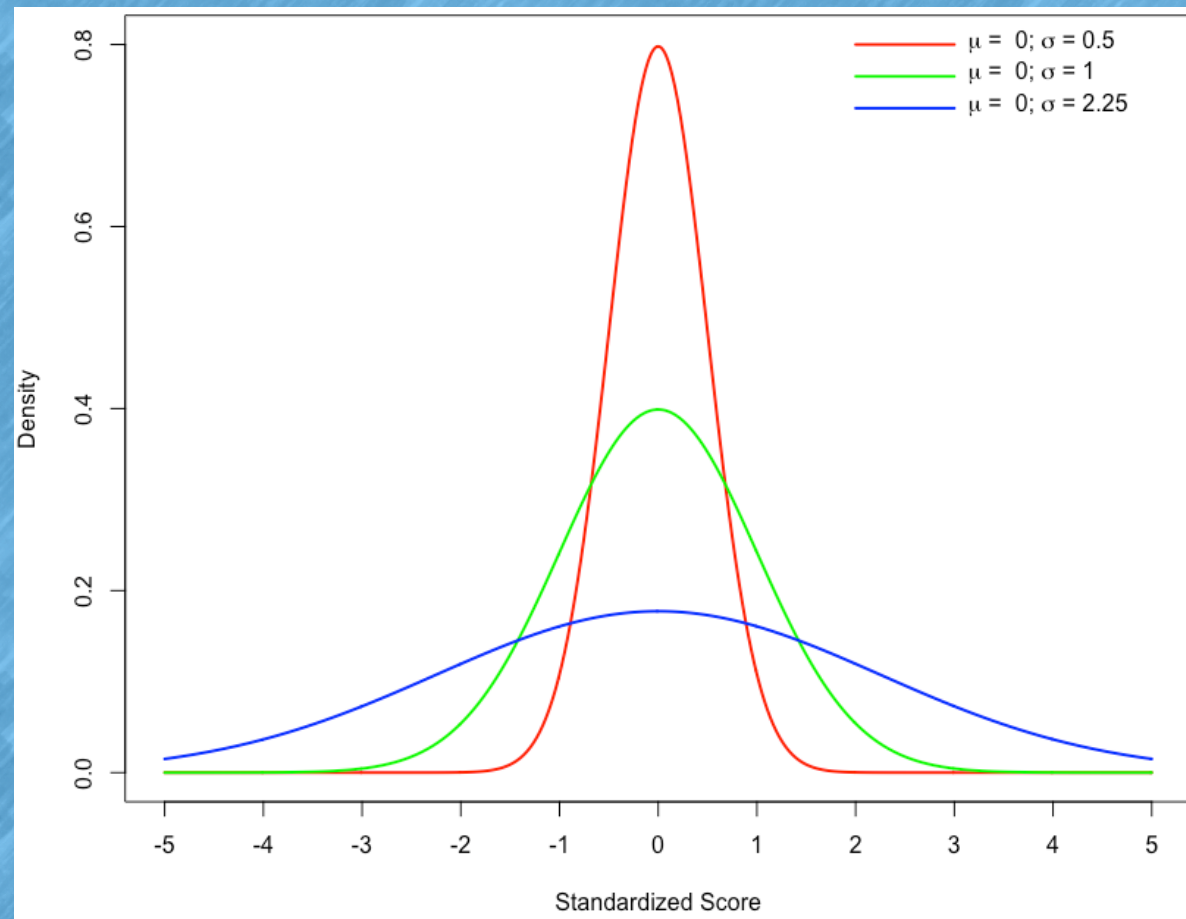


# Normal Distributions



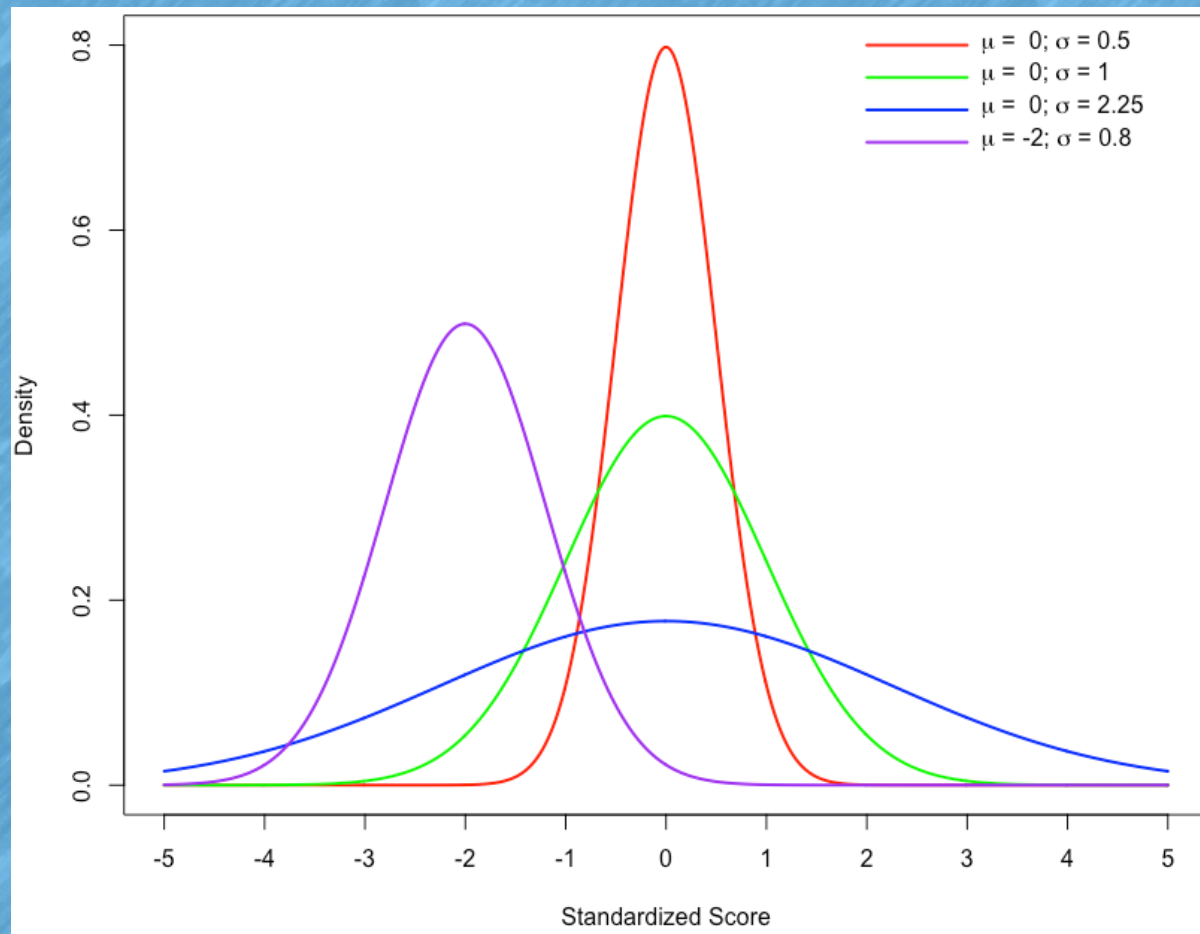


# Normal Distributions



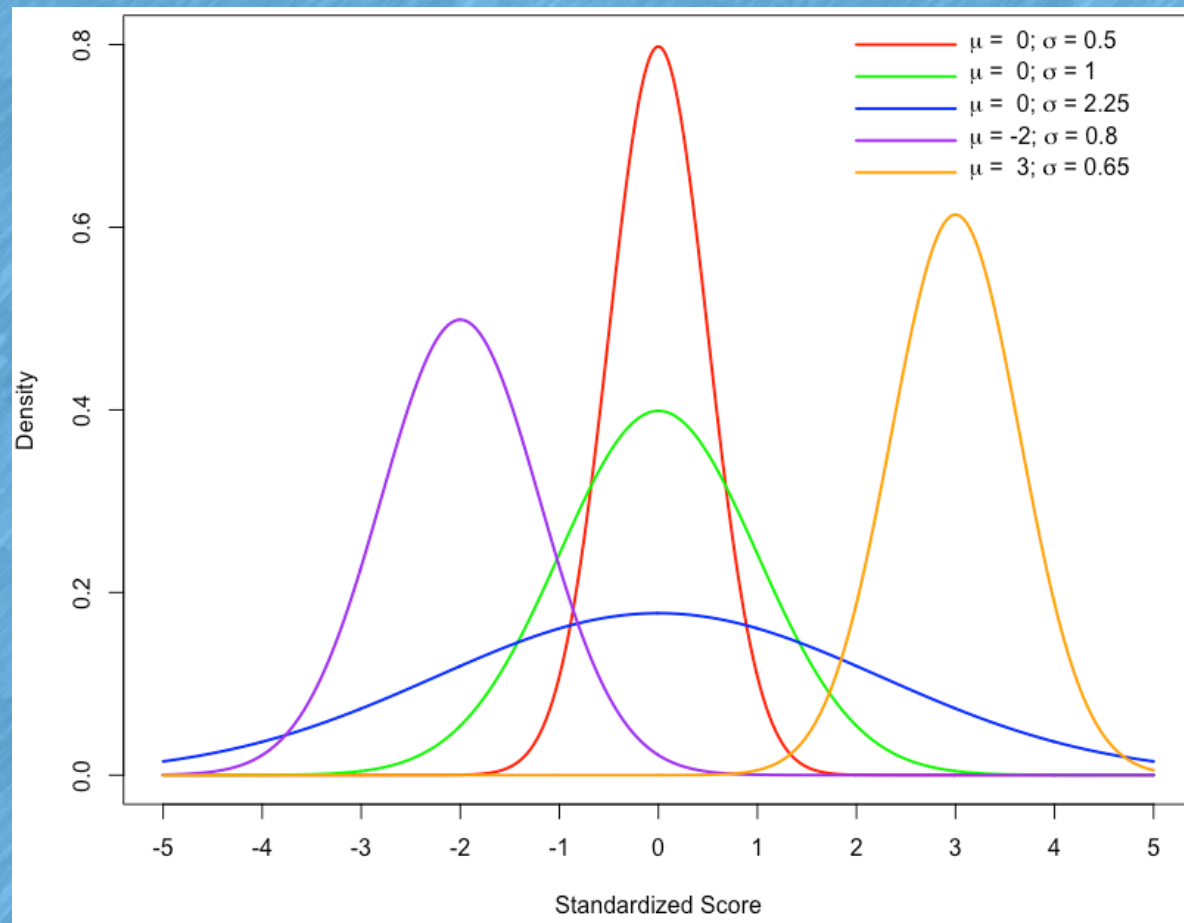


# Normal Distributions





# Normal Distributions





# Central Limit Theorem

- The CLT:

For random samples with a sufficiently large sample size, the distribution of sample statistics for a mean or a proportion is approximately normally distributed and centered at the value of the population parameter.



# T-distribution

- Our mathematical approximation for our sampling distribution
- T-distribution is very similar to the normal distribution, but with slightly thicker tails
- Because we are looking for an estimate for the variability in our sample we will be able to calculate that to be
  - $SE = s/\sqrt{n}$

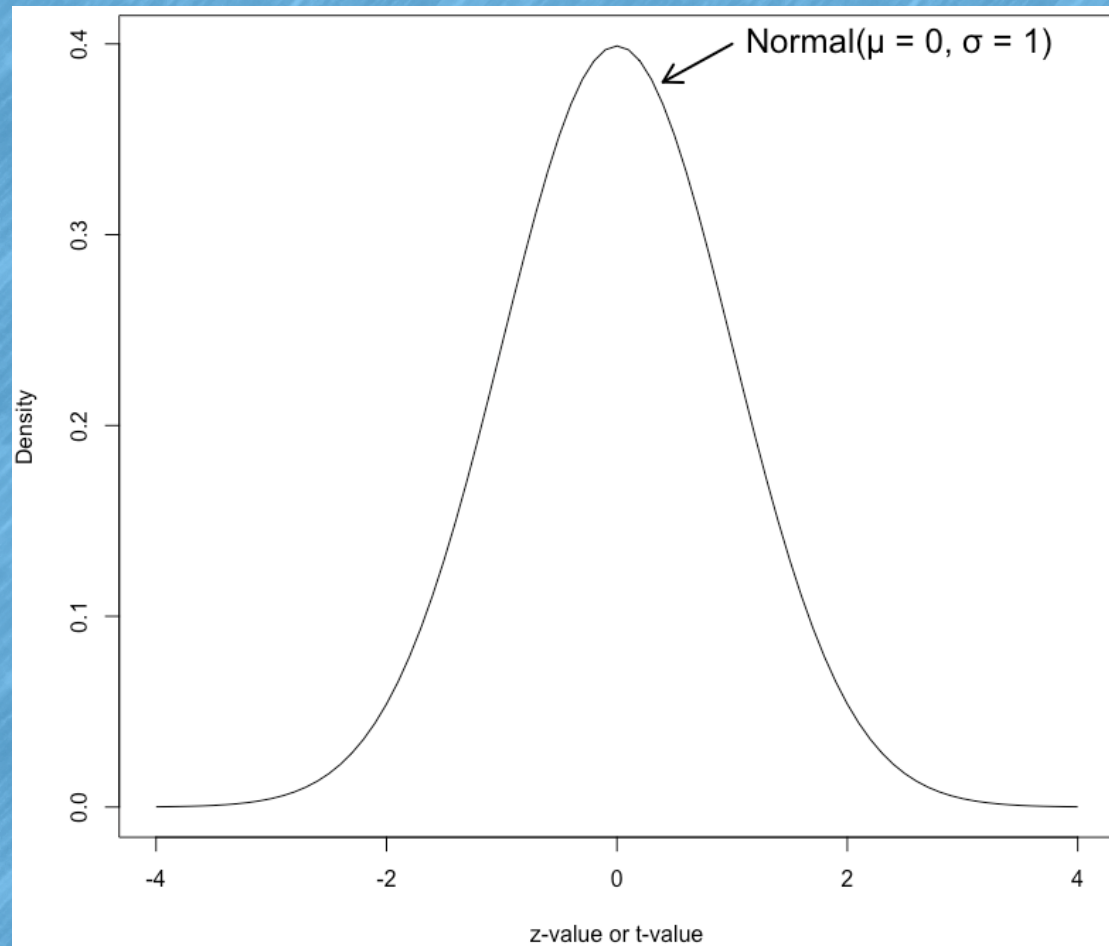


# Degrees of Freedom

- The t-distribution is characterized by degrees of freedom (d.f.)
- Calculated based on sample size
- The higher the d.f., the closer the t-distribution gets to the normal distribution

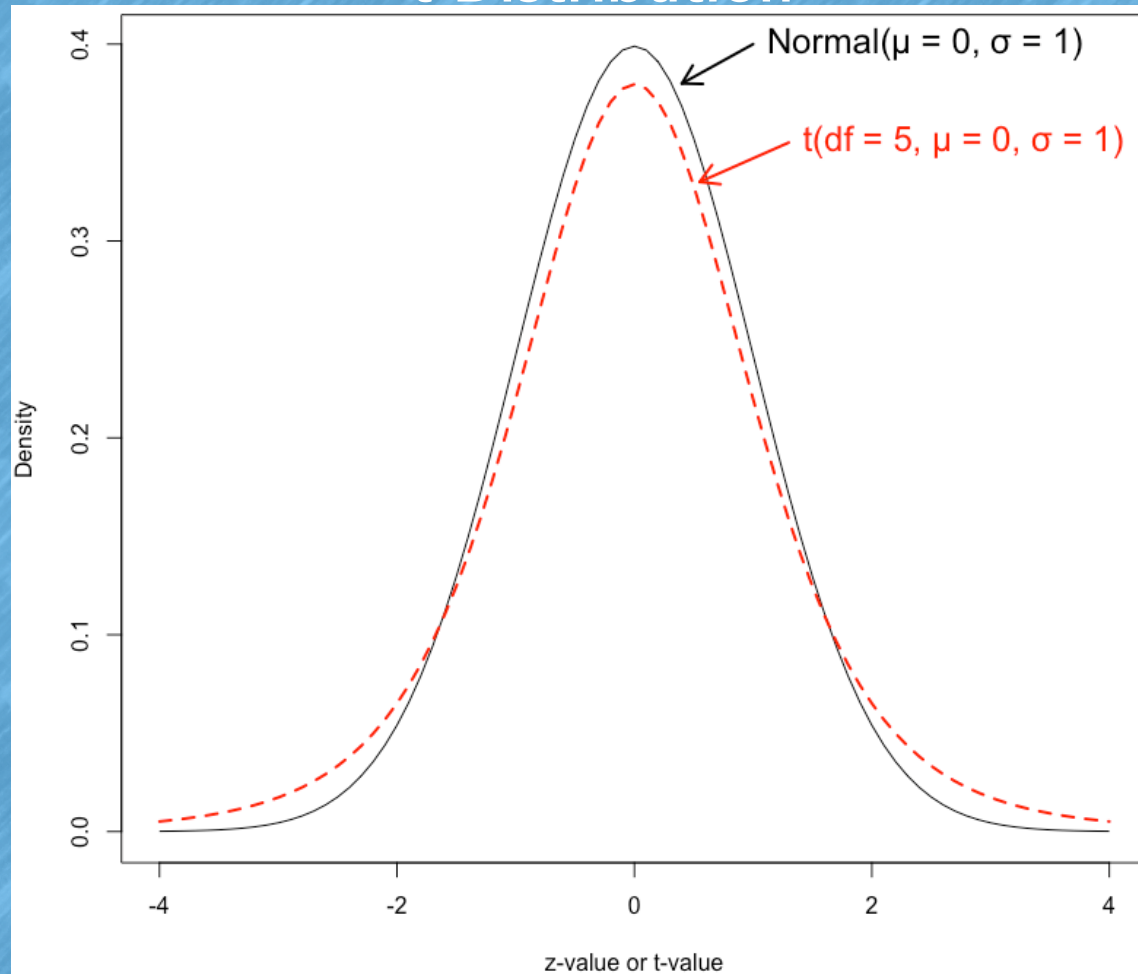


# t-Distribution



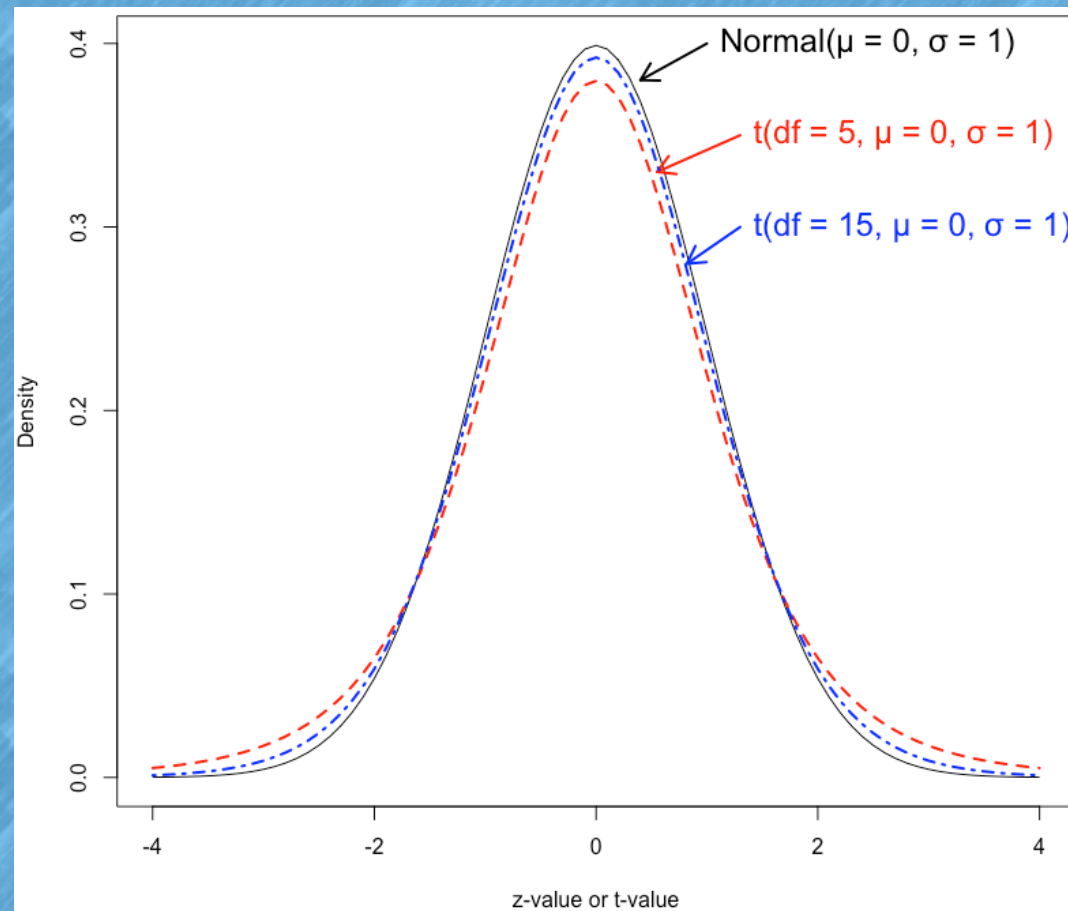


## t-Distribution





# t-Distribution





# Assumptions

- $n \geq 30$  OR sample distribution looks reasonably normal
- If this is not met, better to use a randomization test



# Use R Studio

- Instead of doing a simulation we can use the t-distribution to help us get our estimate for the variability
- Use functions in R Studio to also give us our p-value
- We will explore the entire hypothesis test process in today's activity!



# T-test Parametric Methods Hypothesis Testing Activity



# Summary

- There are many steps to the hypothesis test (overview on slide 9)
- Hypothesis tests help us test a claim while taking into account sampling variability
- They provide one form of evidence to help answer a research question
- We can use a t-distribution to help us conduct our test as an alternative method to simulation