CONFIDENCE INTERVALS

MAT 152 – Statistical Methods I
Lecture 2
Instructor: Dustin Roten
Fall 2020

Student's t-Distribution

- Previously, we discussed how to construct an interval estimate for an unknown population mean. This required us to know the population's standard deviation.

- In practice, we rarely know the population standard deviation!

- If we are studying a population ($\mu_X = ???$ and $\sigma_X = ???$) that we ASSUME has a normally distributed parameter, we can no longer use z-scores. You can think of z-scores as "linking" distributions to the population.

- Now we must rely on a t-score and t-distribution!

Population
$\mu_X = ???$
$\sigma_X = ???$

Sample
$\bar{x}$ is known
s is known

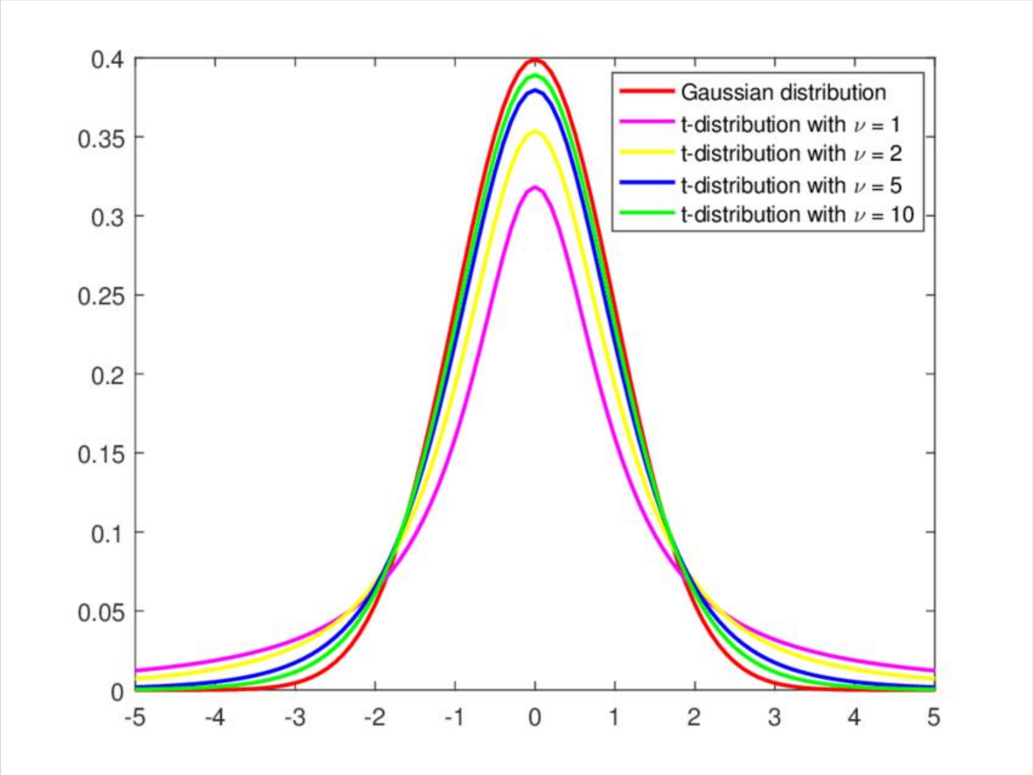
Since we don't have any information about the population, we can't use a normal distribution!

Student's t-Distribution

- William Gosset (1876 – 1937) was an English statistician employed by the Guinness brewing company. In his work, he wanted to make meaningful measurements without using large sample sizes. (Large sample sizes required using up A LOT of good beer!)
- He developed the t-distribution which allowed him to use small sample sizes to estimate parameters for large populations.
- Guinness wouldn't let him publish his new research methods using his own name. Therefore he used the pen name "Student". This name stuck: Student's t-Distribution.

Properties of Student's t-Distribution

- Student's t-Distribution is like a normal distribution.
- Student's t-Distribution is centered around zero.
- Student's t-Distribution has "fat tails". This makes outlier values more likely. (More probability is in the tails.)
- The shape of Student's t-Distribution depends on the sample size.
  - Degrees of freedom = df = ν = n-1.
  - For large df, the t-Distribution starts looking normal.
- The original population is assumed to be normally distributed.



Properties of Student’s t-Distribution

If you draw a simple random sample of size *n* from a population that has an approximately normal distribution with mean μ and unknown population standard deviation σ, a t-score can be calculated.

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

A t-score works just like a z-score: it measures how far the sample mean is from the population mean.

The reason $\nu = n - 1 \ldots$

Suppose we have a sample of four values.

We know three of the values: 3, 6, 9.

If we knew that the mean was 7 then we can determine the fourth value to be 10.

Knowing the mean and n-1 values gives us all the values of the sample.

The notation for the Student’s t-distribution is $T \sim t_{df}$

Example

Suppose you do a study of acupuncture to determine how effective it is in relieving pain. You measure sensory rates for 15 subjects with the results given. Use the sample data to construct a 95% confidence interval for the mean sensory rate for the population (assumed normal) from which you took the data.

8.6; 9.4; 7.9; 6.8; 8.3; 7.3; 9.2; 9.6; 8.7; 11.4; 10.3; 5.4; 8.1; 5.5; 6.9

5-step process:

1. Calculate the sample mean and standard deviation
2. Find the t-score that corresponds to the sample size
3. Calculate the error bound for the mean (EBM)
4. Construct the confidence interval
5. Write a sentence that interprets the estimate in context.

Example (Step 1)

Calculate the **sample** mean and **standard deviation**

8.6; 9.4; 7.9; 6.8; 8.3; 7.3; 9.2; 9.6; 8.7; 11.4; 10.3; 5.4; 8.1; 5.5; 6.9

$n = 15$

$\bar{x} = 8.2267$

$s = 1.6722$

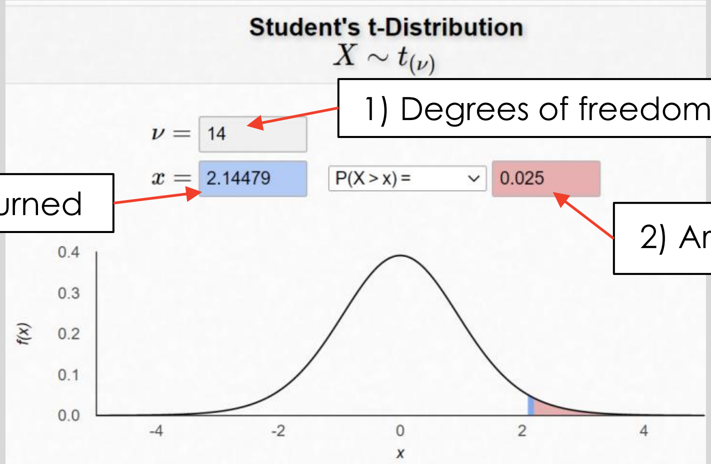
$df = 15 - 1 = 14$

Example (Step 2)

Find the t-score that corresponds to the sample size

$CL = 95\%$ or 0.95
$\alpha = 0.05$

$t_{\frac{\alpha}{2}} = t_{0.025} = ???$

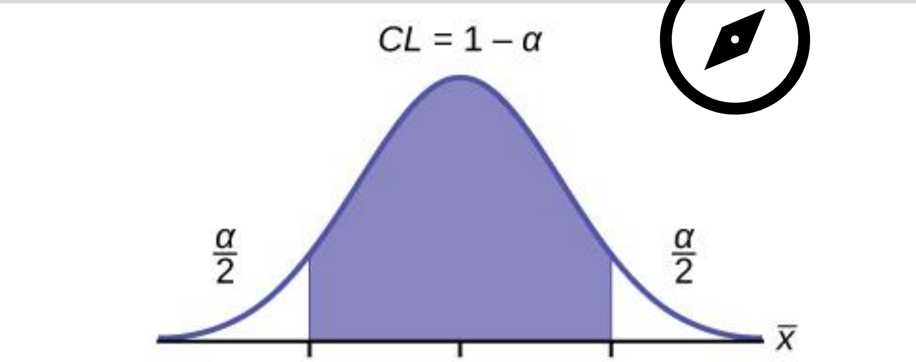


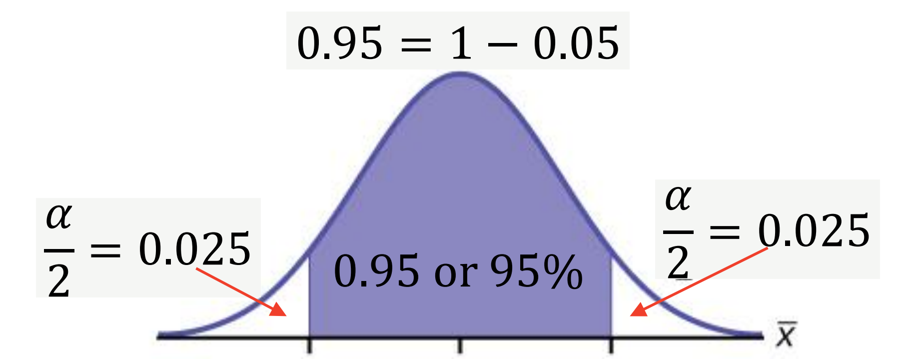
1) Degrees of freedom

2) Area ($\frac{\alpha}{2}$)

3) $t_{0.025}$ is returned

$t_{0.025} \approx 2.14$





Example (Step 3 and Step 4)

Calculate the error bound for the mean (EBM) and construct the confidence interval

$t_{0.025} = 2.14$

$s = 1.6722$

$n = 15$

$$EBM = (t_{0.025})\left(\frac{s}{\sqrt{n}}\right) = (2.14)\left(\frac{1.6722}{\sqrt{15}}\right) = 0.924$$

Lower bound: $\bar{x} - EBM = 8.2267 - 0.924 = 7.3$

Upper bound: $\bar{x} + EBM = 8.2267 + 0.924 = 9.15$

The 95% confidence interval is (7.3, 9.15)

Example (Step 5)

Suppose you do a study of acupuncture to determine how effective it is in relieving pain. You measure sensory rates for 15 subjects with the results given. Use the sample data to construct a 95% confidence interval for the mean sensory rate for the population (assumed normal) from which you took the data.

8.6; 9.4; 7.9; 6.8; 8.3; 7.3; 9.2; 9.6; 8.7; 11.4; 10.3; 5.4; 8.1; 5.5; 6.9

5-step process:

1. Calculate the sample mean and standard deviation $\bar{x} = 8.2267; s = 1.6722$
2. Find the t-score that corresponds to the sample size $t_{0.025} \approx 2.14$
3. Calculate the error bound for the mean (EBM) $EBM = 0.924$
4. Construct the confidence interval (7.3, 9.15)
5. Write a sentence that interprets the estimate in context.

We estimate with 95% confidence that the true population mean sensory rate is between 7.30 and 9.15

A Quick Review

- Student's t-Distribution is used when the population parameters are unknown.
- It is assumed that the population is approximately normally distributed.
- Student's t-Distribution depends on the sample size.
- Student's t-Distribution approximates a normal distribution for large sample sizes.