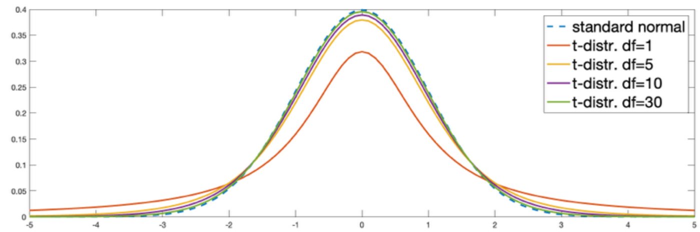
MAT1372, Classwork23, Fall2025

7.1 One-sample means with t-distribution



1. The (Student) t-distribution.

   How to describe it?

   It has just one parameter called degree of freedom (df)

   The difference from standard normal distribution:

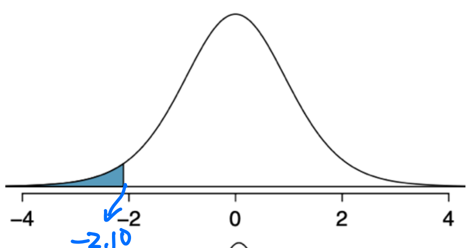
   Its tails are thicker than normal distribution's, meaning observations are more likely to fall beyond two standard deviation from the mean than under the normal distribution. It's nearly normal as $df=30$.

2. The feature of t-distribution.

   Distribution's center: It is always centered at 0.

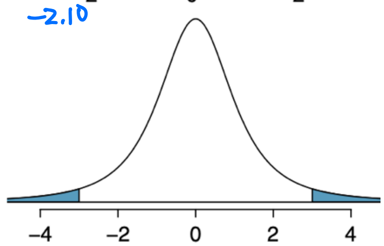
   Shape: The distribution is symmetric, bell-shaped with a thicker tail.

3. Use a t-table to find tail area of t-distribution.



(a) What proportion of the t-distribution with 18 degrees of freedom falls below −2.10?

If we use the t-table, we would only be able to find the value for tail area above 2.10. Since it is symmetrical, then the area below −2.10 is 0.025



(b) A t-distribution with 2 degrees of freedom is shown in the left. Estimate the proportion of the distribution falling more than 3 units from the mean.

If we use a table, we would only be able to find the values for tail area above 2.92 to 4.30

The tail area is between 0.05 and 0.1

4. Sample proportion and Sample mean.

   What is **a sample proportion** $\hat{p}$? The sampling is done to estimate the proportion of population

   Example: The bar exam passing rate is p and the passing rate of sampling n people from all the test takers is

   What is **a sample mean** $\bar{x}$? The average of a sample from the sampling of a population

   Example: The average height of 1000 American adults; The average weight of 200 New Yorkers.

5. **Central Limit Theorem** For the **Sample Mean**

   When we collect a sufficiently large sample of n independent observations from a population with mean μ and standard deviation σ, the sampling distribution of a sample mean $\bar{x}$ will be nearly normal with

   Mean = μ Standard Error (SE) = $\frac{\sigma}{\sqrt{n}}$

6. Why t-distribution?

   In practice, we can't directly calculate the standard Error for $\bar{x}$ since we don't know the population standard deviation σ. We'll employ a similar strategy for computing the standard error of $\bar{x}$, using the sample standard deviation s in place of σ

   Standard Error (SE) = $\frac{\sigma}{\sqrt{n}} \approx \frac{s}{\sqrt{n}}$

   Because of this replacement, the sample mean now follows t-distribution not normal distribution.

7. The conditions required to apply the Central Limit Theorem for a sample mean $\bar{x}$.

Independence. The sample observations must be independent.

Normality. $n<30$: If the sample size n is less than 30 and there are no clear outliers in the data, then we typically assume the data come from a nearly normal distribution to satisfy the condition.

$n \geq 30$: If the sample size n is at least 30 and there are no particularly extreme outliers, then we typically assume sampling distribution of $\bar{x}$ is nearly normal, even if the underlying distribution of individual observations is not.

8. How to find the degree of freedom (df) of t-distribution for a sample mean.

The degree of freedom describes the shape of the t-distribution. The larger the degrees of freedom, the more closely the distribution approximates the normal distribution

When modeling $\bar{x}$ using the t-distribution with sample size n, use $df = n-1$

9. The t-Confidence Interval for the Mean.

Based on a sample of n independent and nearly normal observations, a confidence interval for the population mean is

$$\text{point estimate} \pm t^{\star}_{df} \times \text{SE} \rightarrow \bar{x} \pm t^{*}_{df} \times \frac{s}{\sqrt{n}}$$

where $\bar{x}$ is the sample mean, $t^{\star}_{df}$ corresponds to the confidence level and degrees of freedom df,

t-score (t-value)

and SE is the standard error as estimated by the sample.

10. Example to build a t-Confidence Interval.

n	$\bar{x}$	s	minimum	maximum
19	4.4	2.3	1.7	9.2

Given a summary of mercury content in the muscle of 19 Risso's dolphins from the Taiji area in the table. Measurements are in micrograms of mercury per wet gram of muscle (μg/wet g)

(a) Are the independence and normality conditions satisfied for this data set?

a simple random sample

The summary statistics do not suggest any clear outliers, since all observations are within 2.5 s

above 2.5 s
$4.4 + 2.5 \cdot 2.3 = 10.15$
below -2.5 s
$4.4 - 2.5 \cdot 2.3 = -1.35$

(b) Compute the standard error for the average mercury content in the n = 19 dolphins.

$$SE = \frac{s}{\sqrt{n}} = \frac{2.3}{\sqrt{19}} = 0.528$$

(c) What is the appropriate degrees of freedom (df)? Find $t^{\star}_{df}$ for this df and the 95% confidence level

$df = n-1 = 18$, $95\% \rightarrow$ two tails 5%, one tails 2.5%. Area above t^{*}_{18} is 2.5% $\Longrightarrow t^{*}_{18} = 2.10$

(d) Compute and interpret the 95% confidence interval for the average mercury content in Risso's dolphins.

$$\bar{x} \pm 2.10 \times SE \rightarrow 4.4 \pm 2.10 \cdot (0.528) \rightarrow (3.29, 5.51)$$

11. **Confidence Interval for a Single Mean:** Once you've determined a one-mean confidence interval would be helpful for an application, there are four steps to constructing the interval:

Prepare. Identify $\bar{x}$, s, n and determine what confidence level you wish to use

Check. Verify the conditions to ensure $\bar{x}$ is nearly normal.

Calculate. If the conditions hold, compute SE, find t^{*}_{df}, and construct the interval.

Conclude. Interpret the confidence interval in the context of the problem.