ONLINE MASTERS IN **DATA SCIENCE**

DSC 215 - PROBABILITY AND STATISTICS FOR DATA SCIENCE

ONE SAMPLE T-CONFIDENCE INTERVALS

UC San Diego

COMPUTER SCIENCE & ENGINEERING

HALICIOĞLU DATA SCIENCE INSTITUTE

One Sample t-Confidence Intervals

- **Goal:** use the t-distribution to identify the confidence interval for the average height of 18 year olds (in inches) in a certain population, using a random sample of size 25.

n	Sample mean	s	Sample min	Sample max
25	67.73	2.00	63.48	71.80

- **Question**: Are the independence and normality conditions satisfied?
- **Answer:**
- Independence is satisfied since the sample is random.
- $n < 30$, but we do not see any clear outliers (rule of thumb: all observations within 2.5 standard deviations of the mean), so it is reasonable to conclude that the normality condition is satisfied.

One Sample t-Confidence Intervals

n	Sample mean	s	Sample min	Sample max
25	67.73	2.00	63.48	71.80

- **Question**: What is the standard error SE for the average height in our sample?
- **Answer:** $$SE = \frac{s}{\sqrt{n}} = \frac{2.0}{\sqrt{25}} = 0.4$$
- **Question:** What is the appropriate degrees of freedom df in this example?
- **Answer:** $$df = n - 1 = 25 - 1 = 24$$

One Sample t-Confidence Intervals

- We now know that $\bar{x} = 67.73$, $SE = 0.4$, and $df = 24$. We'd like to now construct a 95% confidence interval around $\bar{x}$.

Categorical Data

$$I = \left(\hat{p} - z^\star \times SE, \quad \hat{p} + z^\star \times SE\right)$$

Found $z^\star$ as the number for which $\mathbb{P}(|x| \leq z^\star) = 0.95$ under a standard normal distribution

Numerical Data

$$I = \left(\bar{x} - t^\star_{df} \times SE, \quad \bar{x} + t^\star_{df} \times SE\right)$$

$t^\star_{df}$ is the number for which $\mathbb{P}(|x| \leq t^\star_{df}) = 0.95$ under a t-distribution with df degrees of

- Using software, or tables, we find that $t^\star_{df} = 2.1$

One Sample t-Confidence Intervals

- **Question:** Having found $\bar{x} = 67.73$, $SE = 0.4$, and $df = 24$, compute and interpret the 95% confidence interval for the average height in our sample of 18 year olds.
- **Answer:** We can construct the confidence interval as

$$\bar{x} \pm t^{\star}_{24} \times SE = 67.73 \pm 2.10 \times 0.4$$

- So

$$I = (66.89,\ 68.57)$$

- We are 95% confident that the average height of 18 year olds in a population that resembles our sample is between 66.89 and 68.57 inches.

One Sample t-Confidence Intervals

- There are 4 steps to constructing a confidence interval for a one sample mean:

  - **Prepare:** Identify or calculate $\bar{x}$, s, n, and determine the confidence level to be used.

  - **Check:** Verify the conditions that $\bar{x}$ is nearly normal.

  - **Calculate:** If $\bar{x}$ is nearly normal, calculate $SE = \dfrac{s}{\sqrt{n}}$, and identify the value of $t^{\star}_{df}$ to use. This should depend on $df = n - 1$, and on the confidence level.

  - **Conclude:** Interpret the confidence interval in the context of the problem.