- So far we have assumed that σ was known for the population

  * CLT → Sampling Distribution for $\bar{X}$ is approximately $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$

  * $z_s = \dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ has a ***N(0,1)*** distribution

- In reality, we will rarely (if ever) know σ

  * we estimate σ with $s = \sqrt{\dfrac{\sum(x_i-\bar{x})^2}{n-1}}$

  * $t_s = \dfrac{\bar{X}-\mu}{s/\sqrt{n}}$ has a *Student's T distribution* with *n-1* degrees of freedom

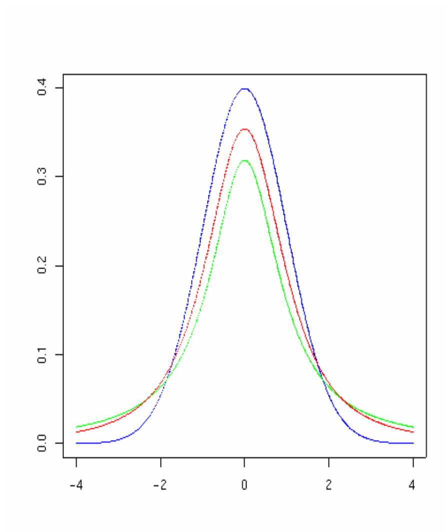
  * Standard Deviation (of the mean) vs. Standard Error (of the mean)

  $SD_{\bar{X}} = \dfrac{\sigma}{\sqrt{n}}$ $\qquad$ $SE_{\bar{X}} = \dfrac{s}{\sqrt{n}}$

Properties of the *T-distribution*

- Symmetric & bell shaped with mean = 0
- Larger spread than the ***N(0,1)*** distribution
- As the d.f. increase, the *T-distribution* approaches the *standard normal curve*
- Table 3/C gives *upper tail probabilities*
- Developed by William S. Gosset

  http://www.uvm.edu/~rsingle/stats/Gosset.html



Normal Distribution in blue
T Distribution df = 5 in green
T Distribution df = 10 in red

Ex: (*d.f.* = 6)

1. Find t^* such that there is 2% area to the right
2. Find t^* such that there is 10% area to the left

100(1-α)% Confidence Interval for μ

$$\bar{x} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}$$ where $t_{\alpha/2}$ is the upper tail critical value with *n-1* d.f.

Ex: A study on cholesterol levels in adult males eating fast food >3 times/week

A SRS of $n = 30$ gives $\bar{X}$ = 180.52 *mg/dL*

s = 41.23 *mg/dL*

1. Find a 95% CI for μ and state an interpretation of the interval
2. Based on this CI, test … H_o: μ=170 vs. H_a: μ≠170 at the α=.05 *l.o.s.*

Suppose we want to find a p-value for the following set of hypotheses

H_o: μ=170 *mg/dL*
H_a: μ>170

$$H_o : \mu = \mu_0 \qquad t_s = \frac{\overline{X} - \mu_0}{s/\sqrt{n}}$$

	Rejection Region	p-value
1) $H_a : \mu > \mu_0$	$t_s \geq t_\alpha$	$p-value = P(T \geq t_s)$
2) $H_a : \mu < \mu_0$	$t_s \leq -t_\alpha$	$p-value = P(T \leq t_s)$
3) $H_a : \mu \neq \mu_0$	$\lvert t_s \rvert \geq t_{\alpha/2}$	$p-value = 2 \cdot P(T \geq \lvert t_s \rvert)$

H_o: μ=165 *mg/dL*

H_a: μ>165

Sample Results: A SRS of $n = 30$ gives $\overline{X} = 180.52$

$s = 41.23$

a) Find the sample *t*-score (t_s).

b) Bracket the p-value.

c) State a conclusion for the test at the α = .05 level.

1) Write the rejection rule (RR) for H_o in terms of *t*-scores.

2) Write the rejection rule (RR) for H_o in terms of $\overline{X}$.

Assumptions, Robustness, and Conditions for Valid CIs & T-Tests

- T-tests and CIs are based on the assumption that the population values being studied have a Normal distribution.
- In reality, populations may be anywhere from slightly non-normal to very non-normal.

Robustness of the T-procedures

- The T-test and CI are called robust to the assumption of normality because p-values and confidence levels are not greatly affected by violations of this assumption of normally distributed populations, especially if sample sizes are large enough.

Conditions for a Valid 1-Sample CI and T-Test

- The data are a SRS from the population.
- The population must be large enough (at least 10 times larger than the sample size).
- Conditions for the sample:
  - If n is small (n < 15), the data should not be grossly non-normal or contain outliers.
  - If n is "medium" (15 ≤ n < 40), the data should not have strong skewness or outliers.
  - If n is large (n ≥ 40), the *T*-procedures are robust to non-normality.

Checking if the conditions are met in your sample

- Always make a plot of the data to check for skewness and outliers before relying on T-procedures in small samples.