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).

- The population is normally distributed.

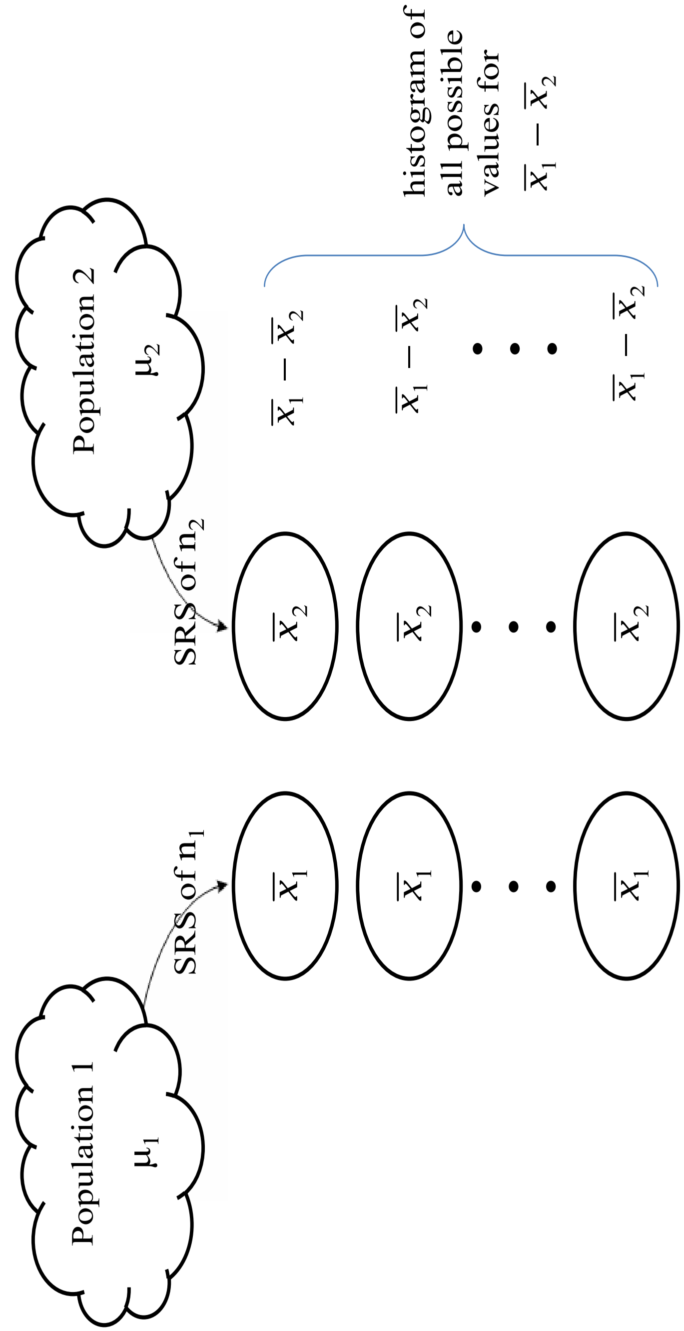
    - If n is small ($n < 15$), the data should not be grossly non-normal or contain outliers.

    - If n is "medium" ($15 \leq n < 40$), the data should not have strong skewness or outliers.

    - If n is large ($n \geq 40$), 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.



Comparing Two Independent Samples

The *sampling distribution* for …

$\bar{X}_1$ is approximately Normal with $Mean = \mu_1$, $SD = \frac{\sigma_1}{\sqrt{n_1}}$

$\bar{X}_2$ is approximately Normal with $Mean = \mu_2$, $SD = \frac{\sigma_2}{\sqrt{n_2}}$

$\bar{X}_1 - \bar{X}_2$ is approx Normal with $Mean = \mu_1 - \mu_2$, $SD = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

$$z_s = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{SD_{\bar{X}_1 - \bar{X}_2}} \rightarrow t_s = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{SE_{\bar{X}_1 - \bar{X}_2}}$$

Example: 30 mice given one of two diets for 21 days.

Reduction in cholesterol in *mg/dl* is measured on day 22

Diet	N	Mean	StDev
Bean	15	26.46	5.90
Oat	15	32.23	9.56

$H_o: \mu_1 - \mu_2 = D_0$

$$t_s = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{SE_{\bar{x}_1 - \bar{x}_2}}$$

$\sigma_1 = \sigma_2 \rightarrow$ $$t_s = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$ with $df = n_1 + n_2 - 2$ (Equal Variance *T*-test)

Rejection Region at the α level of significance:

Reject Ho in favor of

$H_a: \mu_1 - \mu_2 > D_0$ if $t_s \geq t_\alpha$

$H_a: \mu_1 - \mu_2 < D_0$ if $t_s \leq -t_\alpha$

$H_a: \mu_1 - \mu_2 \neq D_0$ if $|t_s| \geq t_{\alpha/2}$

$\sigma_1 \neq \sigma_2 \rightarrow$ $$t_s' = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$ (Unpooled or Unequal Variance *T*-test)

Satterthwaite's approximation for the degrees of freedom:

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{(n_1 - 1)} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{(n_2 - 1)}}$$