- **CLT** (Central Limit Theorem): Draw a random sample of size n from any population with mean μ and finite standard deviation σ. When n is large, the sampling distribution of the sample mean is approximately $N(\mu, \frac{\sigma^2}{n})$.
- Confidence interval : estimate $\pm$ margin of error. A confidence interval for μ is $\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$
- The confidence interval will have a specified **margin of error** E when the sample size is: $n = \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2$
- A **test statistic** calculated from the sample data measures how far the data diverge from the null hypothesis H_0. Define the test statistic $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$

  Decision Rule: Given fixed $\alpha = P(\text{Type-I Error})$.

  $H_a: \mu > \mu_0$, Reject H_0 if $z > z_\alpha$

  $H_a: \mu < \mu_0$, Reject H_0 if $z < -z_\alpha$

  $H_a: \mu \neq \mu_0$, Reject H_0 if $|z| > z_{\alpha/2}$
- If we reject H_0 when in fact H_0 is true, this is a **Type I error**.
- If we fail to reject H_0 when in fact H_a is true, this is a **Type II error**, also called β.

  For one sided alternative test ($H_a: \mu > \mu_0$ or $H_a: \mu < \mu_0$): $\beta = P\Big(Z \le z_\alpha - \frac{|\mu_0 - \mu_a|}{\sigma/\sqrt{n}}\Big)$

  For two sided alternative test ($H_a: \mu \neq \mu_0$): $\beta = P\Big(Z \le z_{\alpha/2} - \frac{|\mu_0 - \mu_a|}{\sigma/\sqrt{n}}\Big)$
- **Power analysis** (Sample size determination), where power $= 1 - \beta$:

  For one sided alternative test ($H_a: \mu > \mu_0$ or $H_a: \mu < \mu_0$): $n = \sigma^2\frac{(z_\alpha + z_\beta)^2}{\Delta^2}$

  For two sided alternative test ($H_a: \mu \neq \mu_0$): $n = \sigma^2\frac{(z_{\alpha/2} + z_\beta)^2}{\Delta^2}$
- The **one-sample t statistic**: $t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$ has the **t distribution** with $n-1$ degrees of freedom.

  The p_value for a test of $H_0: \mu = \mu_0$ vs

  $H_a: \mu > \mu_0$ is p-value $= P(T \ge t)$

  $H_a: \mu < \mu_0$ is p-value $= P(T \le t)$

  $H_a: \mu \neq \mu_0$ is p-value $= 2P(|T| \ge |t|)$
- Confidence interval for μ is:

  $\bar{x} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}$ with $df = n - 1$
- **Power analysis** (2 Independent Samples):

  for ($H_a: \mu_1 > \mu_2$ or $H_a: \mu_1 < \mu_2$):

  $$n = \frac{2\sigma^2(z_\alpha + z_\beta)^2}{\Delta^2}$$

  for ($H_a: \mu_1 \neq \mu_2$):

  $$n = \frac{2\sigma^2(z_{\alpha/2} + z_\beta)^2}{\Delta^2}$$
- **Power analysis** (Paired Samples):

  for ($H_a: \mu_1 > \mu_2$ or $H_a: \mu_1 < \mu_2$):

  $$n = \frac{\sigma_d^2(z_\alpha + z_\beta)^2}{\Delta^2}$$

  for ($H_a: \mu_1 \neq \mu_2$):

  $$n = \frac{\sigma_d^2(z_{\alpha/2} + z_\beta)^2}{\Delta^2}$$

Case 2: Two Numerical Variables – Population Standard Deviations unknown

	Independent Samples		
	$\boldsymbol{\sigma_1 = \sigma_2}$	$\boldsymbol{\sigma_1 \neq \sigma_2}$	**Paired Samples**
Parameters of Interest:	Mean Difference, $\mu_1 - \mu_2$	Mean Difference, $\mu_1 - \mu_2$	Mean Difference, $\mu_1 - \mu_2$
Confidence Interval Formula:	$\bar{x}_1 - \bar{x}_2 \pm t_{\alpha/2} s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ $df = n_1 + n_2 - 2$, where $s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$	$\bar{x}_1 - \bar{x}_2 \pm t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ $df = \frac{(n_1-1)(n_2-1)}{(1-c)^2(n_1-1) + c^2(n_2-1)}$, where $c = \frac{\frac{s_1^2}{n_1}}{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$	$\bar{d} \pm t_{\alpha/2}\frac{s_d}{\sqrt{n}}$ with df = $n - 1$ where the subscript "d" denotes the calculation is performed on the differences "Sample1" – "Sample2"
Name of Hypothesis Test, H_0	Two-sample T Test, $H_0: \mu_1 = \mu_2$ or $H_0: \mu_1 - \mu_2 = 0$	Two-sample T Test, $H_0: \mu_1 = \mu_2$ or $H_0: \mu_1 - \mu_2 = 0$	Paired T Test, $H_0: \mu_1 = \mu_2$ or $H_0: \mu_1 - \mu_2 = 0$
Test Statistic Formula:	$t = \frac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$	$t' = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$	$t = \frac{\bar{d}}{s_d/\sqrt{n}}$ with df = $n - 1$
P-value:	$H_a: \mu_1 \neq \mu_2$, p-value = $2P(T \ge \lvert t \rvert)$ $H_a: \mu_1 > \mu_2$, p-value = $P(T \ge t)$ $H_a: \mu_1 < \mu_2$, p-value = $P(T \le t)$		$H_a: \mu_1 \neq \mu_2$, p-value = $2P(T \ge \lvert t \rvert)$ $H_a: \mu_1 > \mu_2$, p-value = $P(T \ge t)$ $H_a: \mu_1 < \mu_2$, p-value = $P(T \le t)$

- **NON-PARAMETRIC TESTS:** For the population median: use SIGN TEST. For TWO Independent Samples: use WILCOXON RANK-SUM TEST. For TWO Dependent Samples: use WILCOXON SIGNED-RANK TEST