Test 3 Formulas

- z-score in R:

$$z_{\alpha/2} = \text{qnorm}(1-\alpha/2)$$
$$z_{\alpha} = \text{qnorm}(1-\alpha)$$

- t-score in R:

$$t_{\alpha/2} = \text{qt}(1-\alpha/2, \text{df})$$
$$t_{\alpha} = \text{qt}(1-\alpha, \text{df})$$

- The confidence interval for **population mean** with known σ:

$$\overline{X} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = \left[\overline{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}},\ \ \overline{X} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right]$$

- The confidence interval for the difference between **two population means**:

$$(\overline{X} - \overline{Y}) \pm z_{\alpha/2} \cdot \sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}$$

- Sample size needed for a given precision for **population mean**:

In order to attain a margin of error Δ for estimating a **population mean** with a confidence level $(1-\alpha)$, a sample size is required:

$$n \geq \left(\frac{z_{\alpha/2} \cdot \sigma}{\Delta}\right)^2$$

- Sample size needed for a given precision for **population proportion**:

In order to attain a margin of error Δ for estimating a **population proportion** with a confidence level $(1-\alpha)$, a sample size is required:

$$n \geq 0.25 \left(\frac{z_{\alpha/2}}{\Delta}\right)^2$$

Alternatively, we can generate to any proportion if the proportion p is given:

$$n \geq p(1-p) \left(\frac{z_{\alpha/2}}{\Delta}\right)^2$$

- The confidence interval for **population proportion**:

$$\widehat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

- The confidence interval for the difference of **population proportions**:

$$(\widehat{p}_1 - \widehat{p}_2) \pm z_{\alpha/2} \cdot \sqrt{\frac{\widehat{p}_1(1-\widehat{p}_1)}{n_1} + \frac{\widehat{p}_2(1-\widehat{p}_2)}{n_2}}$$

* The confidence interval for the **population Mean** using t-value:

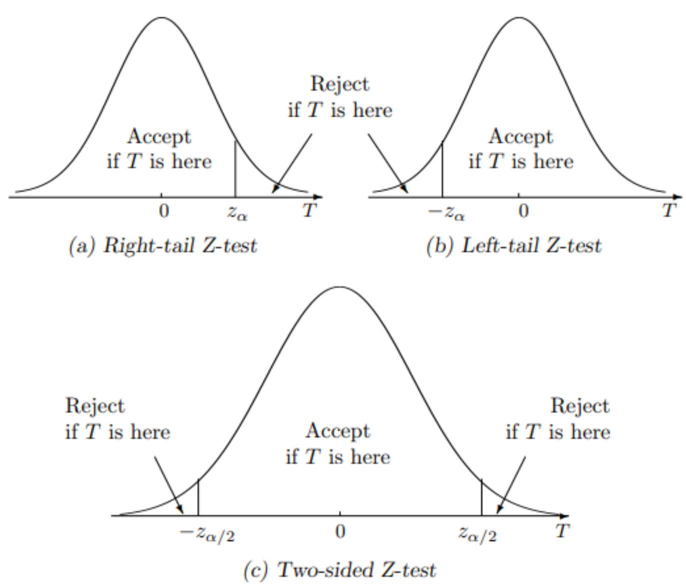
$$\overline{X} \pm t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

where s is the sample standard deviation and $t_{\alpha/2}$ is a critical value from t-distribution with $n-1$ degrees of freedom.

* Possible Errors:

	Result of the test	
	Reject H_0	**Accept H_0**
H_0 **is true**	*Type I error*	correct
H_0 **is false**	correct	*Type II error*

* Rejection Regions:



(a) Right-tail Z-test

(b) Left-tail Z-test

(c) Two-sided Z-test

* Z-tests are summarized in the table below.

Null hypothesis H_0	Parameter, estimator θ, $\widehat{\theta}$	If H_0 is true: $\mathbf{E}(\widehat{\theta})$	If H_0 is true: $\mathrm{Var}(\widehat{\theta})$	Test statistic $Z = \frac{\widehat{\theta} - \theta_0}{\sqrt{\mathrm{Var}(\widehat{\theta})}}$
One-sample Z-tests for means and proportions, based on a sample of size n				
$\mu = \mu_0$	μ, $\overline{X}$	μ_0	$\frac{\sigma^2}{n}$	$\frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$
$p = p_0$	p, $\widehat{p}$	p_0	$\frac{p_0(1-p_0)}{n}$	$\frac{\widehat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$
Two-sample Z-tests comparing means and proportions of two populations, based on independent samples of size n and m				
$\mu_X - \mu_Y = D$	$\mu_X - \mu_Y$, $\overline{X} - \overline{Y}$	D	$\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}$	$\frac{\overline{X} - \overline{Y} - D}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}}$
$p_1 - p_2 = D$	$p_1 - p_2$, $\widehat{p}_1 - \widehat{p}_2$	D	$\frac{p_1(1-p_1)}{n} + \frac{p_2(1-p_2)}{m}$	$\frac{\widehat{p}_1 - \widehat{p}_2 - D}{\sqrt{\frac{\widehat{p}_1(1-\widehat{p}_1)}{n} + \frac{\widehat{p}_2(1-\widehat{p}_2)}{m}}}$
$p_1 = p_2$	$p_1 - p_2$, $\widehat{p}_1 - \widehat{p}_2$	0	$p(1-p)\left(\frac{1}{n} + \frac{1}{m}\right)$, where $p = p_1 = p_2$	$\frac{\widehat{p}_1 - \widehat{p}_2}{\sqrt{\widehat{p}(1-\widehat{p})\left(\frac{1}{n} + \frac{1}{m}\right)}}$ where $\widehat{p} = \frac{n\widehat{p}_1 + m\widehat{p}_2}{n+m}$

* T-tests are summarized in the table below.

Hypothesis H_0	Conditions	Test statistic t	Degrees of freedom
$\mu = \mu_0$	Sample size n; unknown σ	$t = \frac{\overline{X} - \mu_0}{s/\sqrt{n}}$	$n - 1$
$\mu_X - \mu_Y = D$	Sample sizes n, m; unknown but equal standard deviations, $\sigma_X = \sigma_Y$	$t = \frac{\overline{X} - \overline{Y} - D}{s_p\sqrt{\frac{1}{n} + \frac{1}{m}}}$	$n + m - 2$
$\mu_X - \mu_Y = D$	Sample sizes n, m; unknown, unequal standard deviations, $\sigma_X \neq \sigma_Y$	$t = \frac{\overline{X} - \overline{Y} - D}{\sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}}}$	Satterthwaite approximation, formula (9.12)

* R code for 95% Confidence Intervals for one Proportion:

  prop.test(c(x,0), c(n,n), conf.level= 0.95, correct = F)

* R code for 95% Confidence Intervals for two Proportions:

  prop.test(c(x1,x2), c(n1,n2), conf.level= 0.95, correct = F)

* R code to find the critical $t_{\alpha/2}$:

$$t_{\alpha/2} = \mathbf{qt}(1 - \tfrac{\alpha}{2},\ df)$$

* R code for 95% Confidence Intervals when sample standard deviation (s) is given:

  t.test(x, conf.level = 0.95)

  where, x=c(x1, x2, x3, ..., xn).

* R code to perform a hypothesis test for one proportion:

  prop.test(x, n, p, alternative = "greater" (or "less", or "two.sided"), correct = F)

* R code to perform a hypothesis test for two proportions:

  prop.test(c(x1,x2), c(n1,n2), alternative = "greater" (or "less", or "two.sided"), correct = F)

* R code to perform a hypothesis test for one mean if we have the full data set:

  t.test(x, mu = ??, alternative = "greater" (or "less", or "two.sided"))

* R code to find the probability using t-test:

$$\mathbf{P}\{t < a\} = \text{pt(a, df)}$$

$$\mathbf{P}\{t > b\} = \text{1 - pt(b, df)}$$