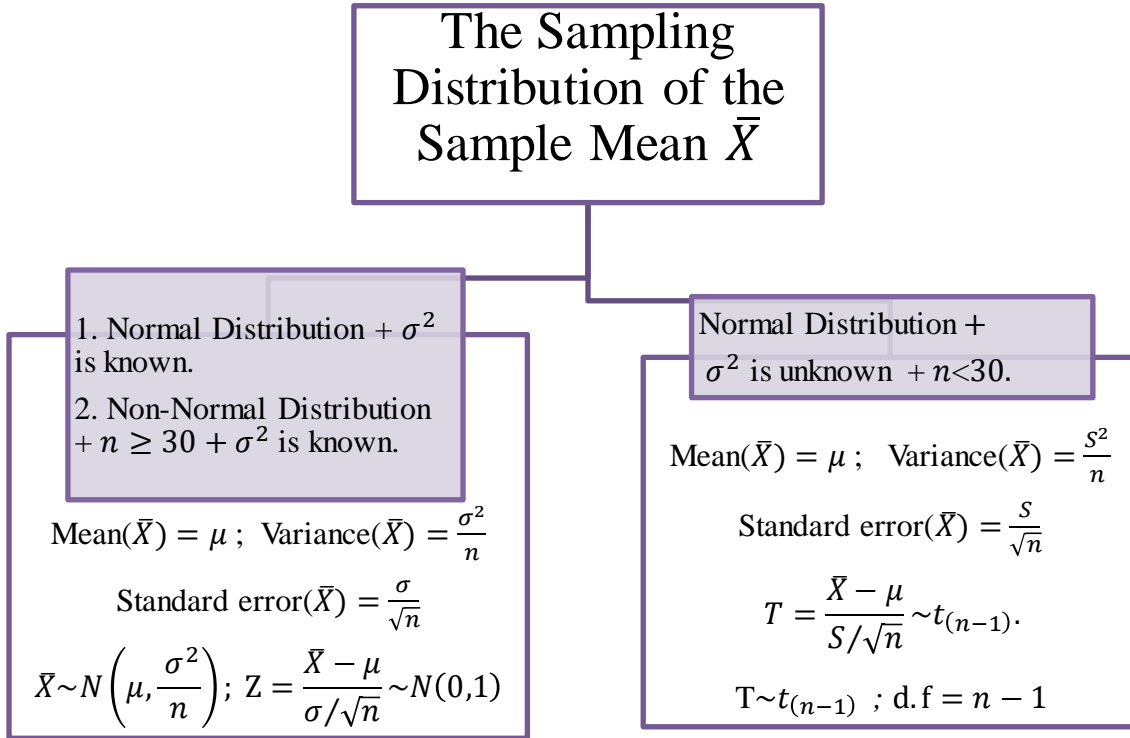


# The Sampling Distributions of Sample Statistics:

## Case 1: The Sampling Distribution of the Sample Mean $\bar{X}$



## Case 2: The Sampling Distribution of the Difference between two Sample Means $\bar{X}_1 - \bar{X}_2$ :

When  $n \geq 30$  +  $\sigma_1^2$  and  $\sigma_2^2$  are Known + either Normal or Non-Normal distribution.

Then

$$\text{Mean}(\bar{X}_1 - \bar{X}_2) = \mu_{\bar{X}_1 - \bar{X}_2} = \mu_1 - \mu_2$$

$$\text{Variance}(\bar{X}_1 - \bar{X}_2) = \sigma_{\bar{X}_1 - \bar{X}_2}^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

$$\text{Standard Error}(\bar{X}_1 - \bar{X}_2) = \sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right) ; \quad Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

### **Case 3: The Sampling Distribution of the Sample Proportion $\hat{p}$ :**

When  $n \geq 30, np > 5, nq > 5$  and  $\hat{p} = \frac{x}{n}$ .

Then

$$\text{Mean } (\hat{p}) = \mu_{\hat{p}} = p$$

$$\text{Variance } (\hat{p}) = \sigma_{\hat{p}}^2 = \frac{pq}{n}$$

$$\text{Standard error}(\hat{p}) = \sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$$

$$\hat{p} \sim N\left(p, \frac{pq}{n}\right)$$

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} \sim N(0,1)$$

### **Case 4: The Sampling Distribution of Difference between two Sample Proportions $\hat{p}_1 - \hat{p}_2$ :**

When  $n_1 \geq 30, n_2 \geq 30, n_1p_1 > 5, n_1q_1 > 5, n_2p_2 > 5, n_2q_2 > 5$  and  $\hat{p}_1 = \frac{x_1}{n_1}, \hat{p}_2 = \frac{x_2}{n_2}$ ;

$$\hat{p}_1 - \hat{p}_2 = \frac{x_1}{n_1} - \frac{x_2}{n_2}.$$

Then

$$\text{Mean } (\hat{p}_1 - \hat{p}_2) = \mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2$$

$$\text{Variance } (\hat{p}_1 - \hat{p}_2) = \sigma_{\hat{p}_1 - \hat{p}_2}^2 = \frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}$$

$$\text{Standard deviation } (\hat{p}_1 - \hat{p}_2) = \sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}$$

$$\hat{p}_1 - \hat{p}_2 \sim N\left(p_1 - p_2, \frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}\right)$$

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}} \sim N(0,1)$$

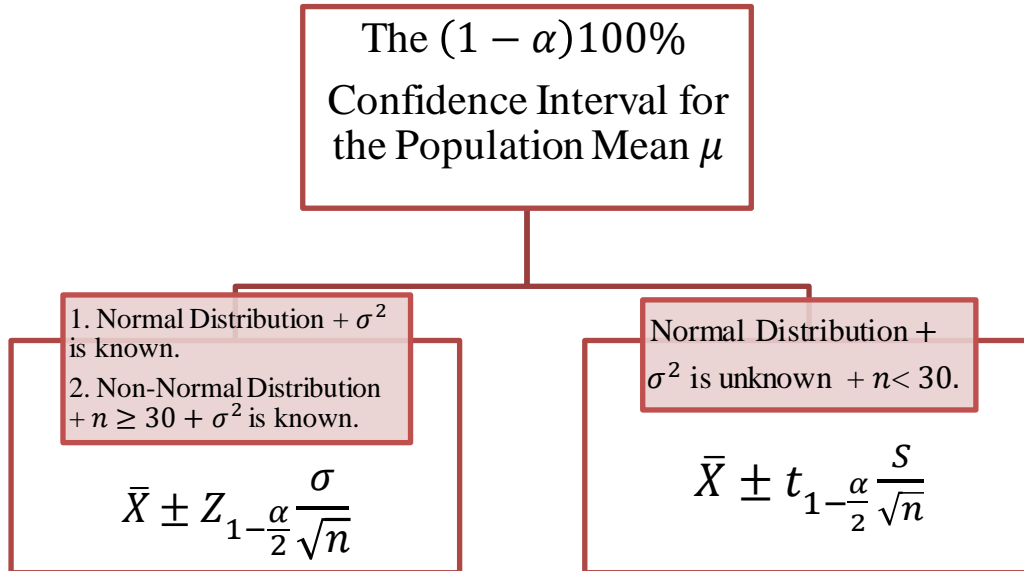
## **Estimations:**

### **1. Point Estimation for the Population Parameters:**

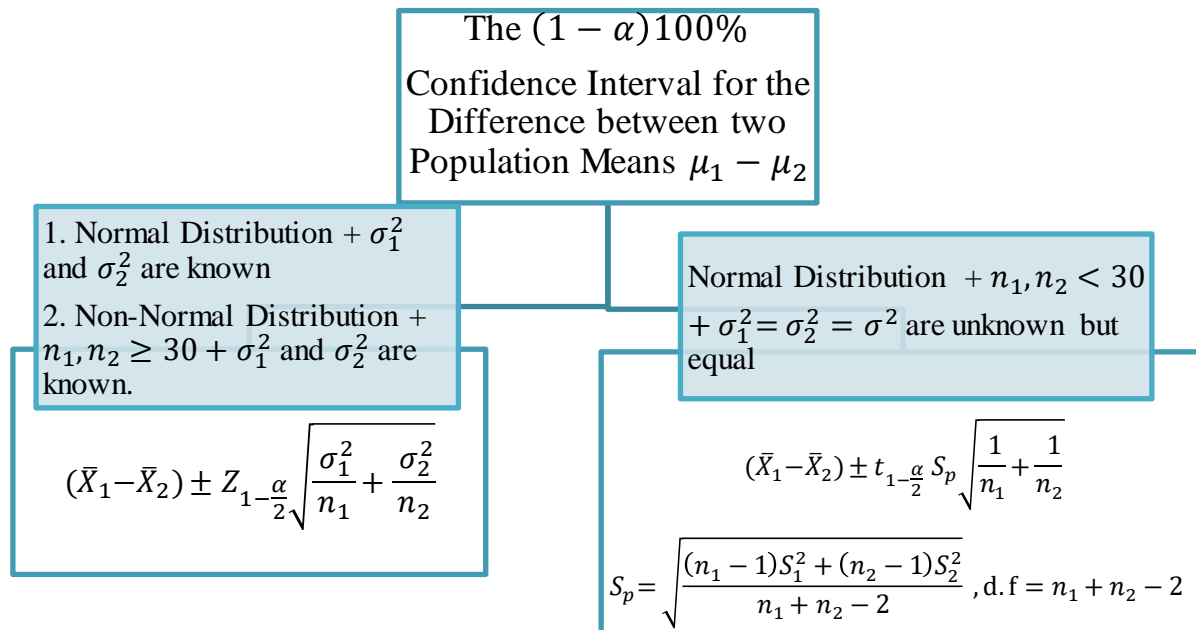
	Population Parameter	Point Estimator (Sample Statistic)
Mean	$\mu$	$\bar{X}$
Variance	$\sigma^2$	$S^2$
Standard Deviation	$\sigma$	$S$
Proportion	$p$	$\hat{p}$
The Difference between two Means	$\mu_1 - \mu_2$	$\bar{X}_1 - \bar{X}_2$
The Difference between two Proportions	$p_1 - p_2$	$\hat{p}_1 - \hat{p}_2$

## 2. Confidence Intervals for the Population Parameters:

### Case 1: The Confidence Interval for the Population Mean $\mu$ :



### Case 2: The Confidence Interval for the Difference between two Population Means $\mu_1 - \mu_2$ :



### **Case 3: The Confidence Interval for the Population Proportion $p$ :**

When  $n \geq 30, np > 5, nq > 5$  and  $\hat{p} = \frac{X}{n}$ .

Then the  $(1 - \alpha)100\%$  confidence interval for  $p$  is

$$\hat{p} \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

### **Case 4: The Confidence Interval for the Difference between two Population Proportions $p_1 - p_2$ :**

When  $n_1 \geq 30, n_2 \geq 30, n_1p_1 > 5, n_1q_1 > 5, n_2p_2 > 5, n_2q_2 > 5$  and  $\hat{p}_1 = \frac{X_1}{n_1}, \hat{p}_2 = \frac{X_2}{n_2}$ ,

$$\hat{p}_1 - \hat{p}_2 = \frac{X_1}{n_1} - \frac{X_2}{n_2}$$

Then the  $(1 - \alpha)100\%$  confidence interval for  $p_1 - p_2$  is

$$(\hat{p}_1 - \hat{p}_2) \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

**The General Formulas:**  $Z = \frac{\text{value} - \text{mean}}{\text{standard error}}$

estimator  $\pm$  (reliability coefficient  $\times$  standard error)

Or estimator  $\pm$  margin of error    Or estimator  $\pm$  precision of the estimate

where,

Standard error = standard deviation

Estimator = Point estimate

Reliability coefficient = table value =  $Z_{1-\frac{\alpha}{2}}$  or  $t_{1-\frac{\alpha}{2}}$

## Hypotheses Testing:

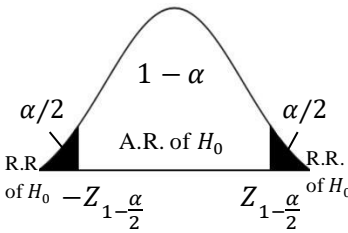
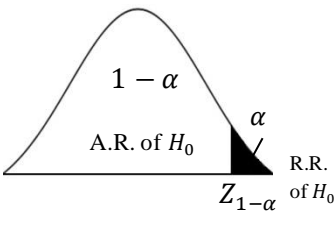
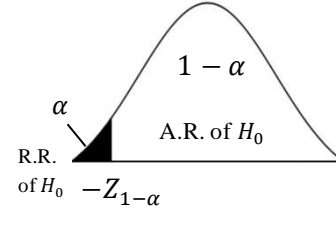
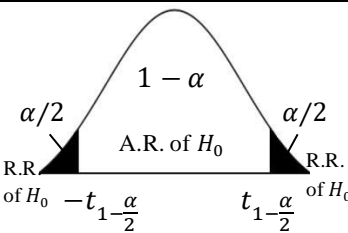
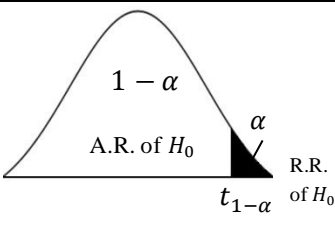
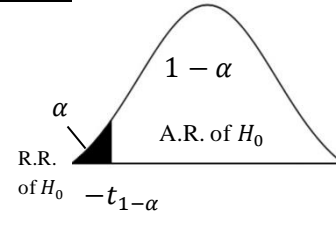
- A hypothesis is a statement about one or more populations.
- A statistical hypothesis is a conjecture (or a statement) concerning the population which can be evaluated by appropriate statistical technique.
- We usually test the null hypothesis ( $H_0$ ) against the alternative (or the research) hypothesis ( $H_A$  or  $H_1$ ) by choosing one of the following situations:
  1. Two-sided hypothesis:  
 $H_0: \theta = \theta_0$  against  $H_A: \theta \neq \theta_0$
  2. One-sided hypothesis:
    - (i)  $H_0: \theta \geq \theta_0$  against  $H_A: \theta < \theta_0$
    - (ii)  $H_0: \theta \leq \theta_0$  against  $H_A: \theta > \theta_0$
- There are 4 possible situations in testing a statistical hypothesis:

		Condition of Null Hypothesis $H_0$ (Nature/Reality)	
		$H_0$ is true	$H_0$ is false
Possible Action (Decision)	Accepting $H_0$	Correct Decision	Type II error ( $\beta$ )
	Rejecting $H_0$	Type I error ( $\alpha$ )	Correct Decision

- There are two types of errors:
  - (i) Type I error = Rejecting  $H_0$  when  $H_0$  is true  
 $P(\text{Type I error}) = P(\text{Rejecting } H_0 \mid H_0 \text{ is true}) = \alpha$   
 Which is called the significance level of the test.
  - (ii) Type II error = Accepting  $H_0$  when  $H_0$  is false  
 $P(\text{Type II error}) = P(\text{Accepting } H_0 \mid H_0 \text{ is false}) = \beta$
- The test statistic has the following form:

$$\text{Test Statistic} = \frac{\text{estimate} - \text{hypothesized parameter}}{\text{standard error of the estimate}}$$

**1. Hypotheses Testing for the population Mean ( $\mu$ ):**

Hypotheses	$H_0: \mu = \mu_0$ vs $H_A: \mu \neq \mu_0$	$H_0: \mu \leq \mu_0$ vs $H_A: \mu > \mu_0$	$H_0: \mu \geq \mu_0$ vs $H_A: \mu < \mu_0$
Assumptions:	First Case: $\sigma^2$ is known + Normal or Non-Normal Distribution but $n \geq 30$		
Test Statistic (T.S.)	$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$		
Rejection Region (R.R.) & Acceptance Region (A.R.) of $H_0$			
Critical Value	$-Z_{1-\frac{\alpha}{2}}$ and $Z_{1-\frac{\alpha}{2}}$	$Z_{1-\alpha}$	$-Z_{1-\alpha}$
Decision	We reject $H_0$ (and accept $H_A$ ) at the significance level $\alpha$ if:		
	$Z < -Z_{1-\frac{\alpha}{2}}$ or $Z > Z_{1-\frac{\alpha}{2}}$	$Z > Z_{1-\alpha}$	$Z < -Z_{1-\alpha}$
Assumptions:	Second Case: $\sigma^2$ is unknown + Normal Distribution + $n < 30$		
Test Statistic (T.S.)	$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t_{(n-1)}; \quad \text{d.f} = v = n - 1$		
Rejection Region (R.R.) & Acceptance Region (A.R.) of $H_0$			
Critical Value	$-t_{1-\frac{\alpha}{2}}$ and $t_{1-\frac{\alpha}{2}}$	$t_{1-\alpha}$	$-t_{1-\alpha}$
Decision	We reject $H_0$ (and accept $H_A$ ) at the significance level $\alpha$ if:		
	$T < -t_{1-\frac{\alpha}{2}}$ or $T > t_{1-\frac{\alpha}{2}}$	$T > t_{1-\alpha}$	$T < -t_{1-\alpha}$
Assumptions:	Third Case: $\sigma^2$ is unknown + Non-Normal Distribution + $n > 30$		
Test Statistic (T.S.)	$Z = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim N(0,1)$		
(R.R.) & (A.R.) of $H_0$	Use the same R.R. & A.R. as in First Case		

## 2. Hypotheses Testing for the Difference Between Two Population Means ( $\mu_1 - \mu_2$ )(Independent Populations):

Hypotheses	$H_0: \mu_1 - \mu_2 = \mu_0$ vs $H_A: \mu_1 - \mu_2 \neq \mu_0$	$H_0: \mu_1 - \mu_2 \leq \mu_0$ vs $H_A: \mu_1 - \mu_2 > \mu_0$	$H_0: \mu_1 - \mu_2 \geq \mu_0$ vs $H_A: \mu_1 - \mu_2 < \mu_0$
Assumptions:	First Case: $\sigma_1^2$ and $\sigma_2^2$ are known, Normal or non-Normal with $n \geq 30$		
Test Statistic (T.S.)	$Z = \frac{\bar{X}_1 - \bar{X}_2 - \mu_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$		
Rejection Region (R.R.) & Acceptance Region (A.R.) of $H_0$			
Critical Value	$-Z_{1-\frac{\alpha}{2}}$ and $Z_{1-\frac{\alpha}{2}}$	$Z_{1-\alpha}$	$-Z_{1-\alpha}$
Decision	We reject $H_0$ (and accept $H_A$ ) at the significance level $\alpha$ if:		
Decision	$Z < -Z_{1-\frac{\alpha}{2}}$ or $Z > Z_{1-\frac{\alpha}{2}}$	$Z > Z_{1-\alpha}$	$Z < -Z_{1-\alpha}$
Assumptions:	Second Case: $\sigma_1^2$ and $\sigma_2^2$ are unknown but equal ( $\sigma_1^2 = \sigma_2^2 = \sigma^2$ ), Normal		
Test Statistic (T.S.)	$T = \frac{\bar{X}_1 - \bar{X}_2 - \mu_0}{\sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}} \sim t_{(n_1+n_2-2)}, \quad df = v = n_1 + n_2 - 2$ $S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$		
Rejection Region (R.R.) & Acceptance Region (A.R.) of $H_0$			
Critical Value	$-t_{1-\frac{\alpha}{2}}$ and $t_{1-\frac{\alpha}{2}}$	$t_{1-\alpha}$	$-t_{1-\alpha}$
Decision	We reject $H_0$ (and accept $H_A$ ) at the significance level $\alpha$ if:		
Decision	$T < -t_{1-\frac{\alpha}{2}}$ or $T > t_{1-\frac{\alpha}{2}}$	$T > t_{1-\alpha}$	$T < -t_{1-\alpha}$



**3. Confidence Interval and Hypotheses Testing for the Difference Between Two Population Means ( $\mu_1 - \mu_2 = \mu_D$ ) for Dependent (Related) Populations: Paired t-Test:**

Calculate the Quantities	<ul style="list-style-type: none"> <li>The differences (D-observations): <math>D_i = X_i - Y_i, i = 1, 2, \dots, n</math></li> <li>Sample Mean of the D-observations: <math>\bar{D} = \frac{\sum_{i=1}^n D_i}{n}</math></li> <li>Sample Variance of the D-observations: <math>S_D^2 = \frac{\sum_{i=1}^n (D_i - \bar{D})^2}{n-1}</math></li> <li>Sample Standard Deviation of the D-observations: <math>S_D = \sqrt{S_D^2}</math></li> </ul>		
Confidence Interval for $\mu_D = \mu_1 - \mu_2$			
100(1 - $\alpha$ )% Confidence Interval for $\mu_D$	$\bar{D} \pm t_{1-\frac{\alpha}{2}} \frac{S_D}{\sqrt{n}}, \quad df = v = n - 1$		
Hypotheses Testing for $\mu_D = \mu_1 - \mu_2$			
Hypotheses	$H_0: \mu_1 = \mu_2 \text{ vs}$ $H_A: \mu_1 \neq \mu_2$ or $H_0: \mu_D = \mu_o \text{ vs}$ $H_A: \mu_D \neq \mu_o$	$H_0: \mu_1 \leq \mu_2 \text{ vs}$ $H_A: \mu_1 > \mu_2$ or $H_0: \mu_D \leq \mu_o \text{ vs}$ $H_A: \mu_D > \mu_o$	$H_0: \mu_1 \geq \mu_2 \text{ vs}$ $H_A: \mu_1 < \mu_2$ or $H_0: \mu_D \geq \mu_o \text{ vs}$ $H_A: \mu_D < \mu_o$
Test Statistic (T.S.)	$T = \frac{\bar{D} - \mu_o}{S_D/\sqrt{n}} \sim t_{(n-1)}, \quad df = v = n - 1$		
Rejection Region (R.R.) & Acceptance Region (A.R.) of $H_0$			
Critical Value	$-t_{1-\frac{\alpha}{2}}$ and $t_{1-\frac{\alpha}{2}}$	$t_{1-\alpha}$	$-t_{1-\alpha}$
Decision	We reject $H_0$ (and accept $H_A$ ) at the significance level $\alpha$ if:		
	$T < -t_{1-\frac{\alpha}{2}}$ or $T > t_{1-\frac{\alpha}{2}}$	$T > t_{1-\alpha}$	$T < -t_{1-\alpha}$

**4. Hypotheses Testing for the Population Proportion ( $p$ ):**

Hypotheses	$H_0: p = p_0$ vs $H_A: p \neq p_0$	$H_0: p \leq p_0$ vs $H_A: p > p_0$	$H_0: p \geq p_0$ vs $H_A: p < p_0$
Test Statistic (T.S.)	$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim N(0,1), \quad \hat{p} = \frac{X}{n}$		
Rejection Region (R.R.) & Acceptance Region (A.R.) of $H_0$			
Critical Value	$-Z_{1-\frac{\alpha}{2}}$ and $Z_{1-\frac{\alpha}{2}}$	$Z_{1-\alpha}$	$-Z_{1-\alpha}$
Decision	We reject $H_0$ (and accept $H_A$ ) at the significance level $\alpha$ if:		
	$Z < -Z_{1-\frac{\alpha}{2}}$ or $Z > Z_{1-\frac{\alpha}{2}}$	$Z > Z_{1-\alpha}$	$Z < -Z_{1-\alpha}$

### 5. Hypotheses Testing for the Difference Between Two Population Proportions ( $p_1 - p_2$ ):

Hypotheses	$H_0: p_1 = p_2$ vs $H_A: p_1 \neq p_2$	$H_0: p_1 \leq p_2$ vs $H_A: p_1 > p_2$	$H_0: p_1 \geq p_2$ vs $H_A: p_1 < p_2$
Test Statistic (T.S.)	$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\bar{p}(1-\bar{p})}{n_1} + \frac{\bar{p}(1-\bar{p})}{n_2}}} \sim N(0,1)$ $\hat{p}_1 = \frac{X_1}{n_1}, \hat{p}_2 = \frac{X_2}{n_2}, \bar{p} = \frac{X_1 + X_2}{n_1 + n_2}$		
Rejection Region (R.R.) & Acceptance Region (A.R.) of $H_0$			
Critical Value	$-Z_{1-\frac{\alpha}{2}}$ and $Z_{1-\frac{\alpha}{2}}$	$Z_{1-\alpha}$	$-Z_{1-\alpha}$
Decision	We reject $H_0$ (and accept $H_A$ ) at the significance level $\alpha$ if:		
	$Z < -Z_{1-\frac{\alpha}{2}}$ or $Z > Z_{1-\frac{\alpha}{2}}$	$Z > Z_{1-\alpha}$	$Z < -Z_{1-\alpha}$