The Sampling Distributions of Sample Statistics:

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

The Sampling Distribution of the Sample Mean $\bar{X}$

σ^2 is known & Normal or Non-Normal Distribution

$$\text{Mean}(\bar{X}) = \mu\ ;\ \text{Variance}(\bar{X}) = \frac{\sigma^2}{n}$$

$$\text{Standard error}(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right);\ Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

σ^2 is unknown & Normal Distribution

$$\text{Mean}(\bar{X}) = \mu\ ;\ \text{Variance}(\bar{X}) = \frac{S^2}{n}$$

$$\text{Standard error}(\bar{X}) = \frac{S}{\sqrt{n}}$$

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{(n-1)};\ \text{d.f} = n - 1$$

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

When $n \geq 30$ & σ_1^2 and σ_2^2 are Known & Normal or Non-Normal distribution.

Then

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

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

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

$$\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

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

<u>Case 3:</u> The Sampling Distribution of the Sample Proportion $\hat{p}$:

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

Then

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

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

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

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

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

<u>Case 4:</u> 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

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

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

Standard deviation $(\hat{p}_1 - \hat{p}_2) = \sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$

$$\hat{p}_1 - \hat{p}_2 \sim N\left(p_1 - p_2, \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}\right)$$

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}} \sim N(0,1)$$

Estimations:

1. Point Estimation for the Population Parameters:

	Population Parameter	Point Estimator (Sample Statistic)
Mean	μ	$\bar{X}$
Variance	σ^2	S^2
Standard Deviation	σ	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 μ:

The $(1-\alpha)100\%$ Confidence Interval for the Population Mean μ

σ^2 is known + Normal or Non-Normal Distribution

$$\bar{X} \pm Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

σ^2 is unknown + Normal Distribution

$$\bar{X} \pm t_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}$$

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

The $(1-\alpha)100\%$ Confidence Interval for the Difference between two Population Means $\mu_1 - \mu_2$

σ_1^2 and σ_2^2 are known

$$(\bar{X}_1 - \bar{X}_2) \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$\sigma_1^2 = \sigma_2^2 = \sigma$ are unknown but equal

$$(\bar{X}_1 - \bar{X}_2) \pm t_{1-\frac{\alpha}{2}} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$S_p = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}, \text{d.f} = n_1 + n_2 - 2$$

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 $n_1 \geq 30, n_2 \geq 30, n_1 p_1 > 5, n_1 q_1 > 5, n_2 p_2 > 5, n_2 q_2 > 5$ 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: $Z = \frac{value - mean}{standard\ error}$

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

Or estimator $\pm$ margin of error

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:

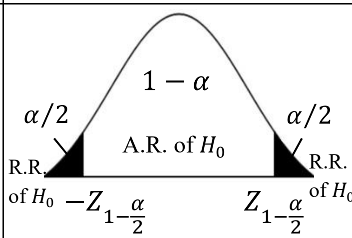
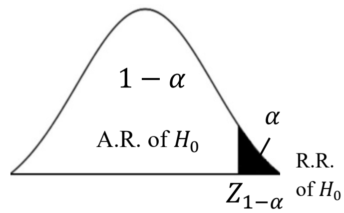
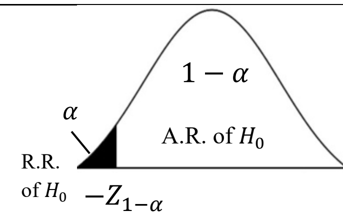
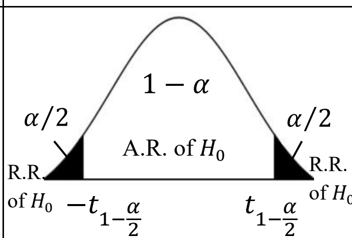
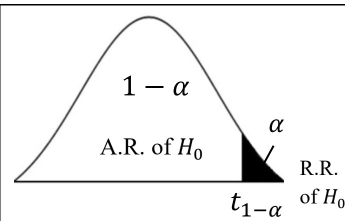
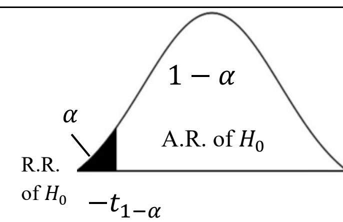
- A hypothesis is a statement about one or more populations.
- A statistical hypothesis is a conjecture (or a statement) concerning the population which can be evaluated by appropriate statistical technique.
- We usually test the null hypothesis (H_0) against the alternative (or the research) hypothesis (H_A or H_1) by choosing one of the following situations:
  Two-sided hypothesis:
  H_0: $\theta = \theta_0$ against H_A: $\theta \neq \theta_0$
  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:

<table>
<tr><td></td><td></td><td colspan="2">Condition of Null Hypothesis H_0 (Nature/Reality)</td></tr>
<tr><td></td><td></td><td>H_0 is true</td><td>H_0 is false</td></tr>
<tr><td rowspan="2">Possible Action (Decision)</td><td>Accepting H_0</td><td>Correct Decision</td><td>Type II error (β)</td></tr>
<tr><td>Rejecting H_0</td><td>Type I error (α)</td><td>Correct Decision</td></tr>
</table>

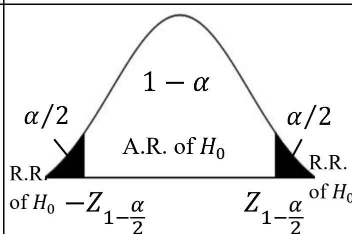
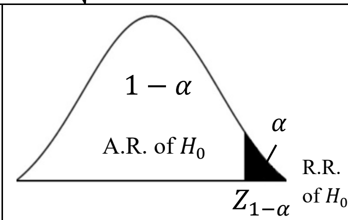
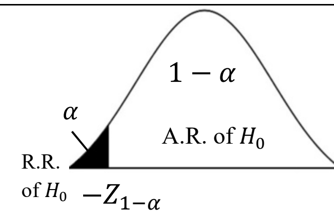
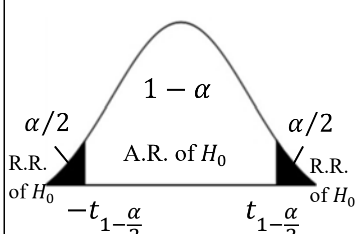
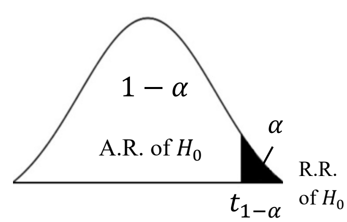
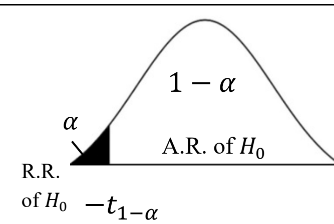
- There are two types of errors:
  (i) Type I error = Rejecting H_0 when H_0 is true
  P(Type I error) = P(Rejecting H_0 | H_0 is true) = α
  Which is called the significance level of the test.
  (ii) Type II error = Accepting H_0 when H_0 is false
  P(Type II error) = P(Accepting H_0 | H_0 is false) = β
- The test statistic has the following form:

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

1. Hypotheses Testing for the population Mean (μ):

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: σ^2 is known; Normal or Non-Normal Distribution		
Test Statistic (T.S.)	$Z = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$		
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 α 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: σ^2 is unknown; Normal Distribution		
Test Statistic (T.S.)	$T = \frac{\bar{X}-\mu_0}{S/\sqrt{n}} \sim t_{(n-1)}$; $\quad$ $\mathrm{d.f} = v = n - 1$		
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 α if:		
	$T < -t_{1-\frac{\alpha}{2}}$ or $T > t_{1-\frac{\alpha}{2}}$	$T > t_{1-\alpha}$	$T < -t_{1-\alpha}$

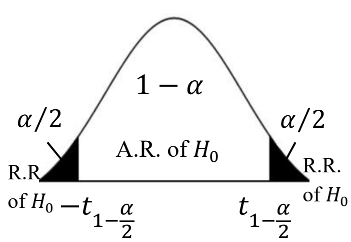
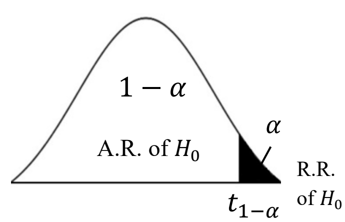
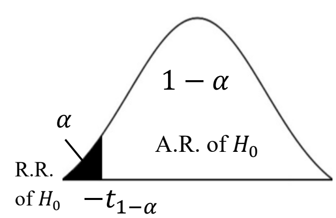
2. Hypotheses Testing for the Difference Between Two Population Means $(\mu_1 - \mu_2)$(Independent Populations):

Hypotheses	$H_0: \mu_1 = \mu_2$ vs $H_A: \mu_1 \neq \mu_2$	$H_0: \mu_1 \leq \mu_2$ vs $H_A: \mu_1 > \mu_2$	$H_0: \mu_1 \geq \mu_2$ vs $H_A: \mu_1 < \mu_2$
Assumptions:	First Case: σ_1^2 and σ_2^2 are known		
Test Statistic (T.S.)	$Z = \dfrac{\bar{X}_1 - \bar{X}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \sim N(0,1)$		
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 α 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: σ_1^2 and σ_2^2 are unknown but equal $(\sigma_1^2 = \sigma_2^2 = \sigma^2)$		
Test Statistic (T.S.)	$T = \dfrac{\bar{X}_1 - \bar{X}_2}{\sqrt{\dfrac{S_p^2}{n_1} + \dfrac{S_p^2}{n_2}}} \sim t(n_1 + n_2 - 2),\ df = v = n_1 + n_2 - 2$ $S_p^2 = \dfrac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$		
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 α if:		
	$T < -t_{1-\frac{\alpha}{2}}$ or $T > t_{1-\frac{\alpha}{2}}$	$T > t_{1-\alpha}$	$T < -t_{1-\alpha}$

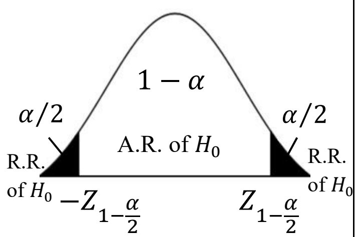
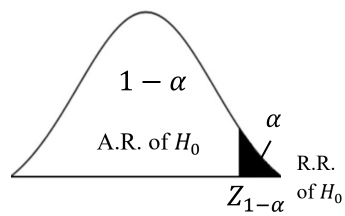
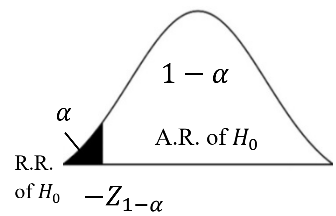
3. Confidence Interval and Hypotheses Testing for the Difference Between Two Population Means $(\mu_1 - \mu_2 = \mu_D)$ for Dependent (Related) Populations: Paired t-Test:

<table>
<tr><td>Calculate the Quantities</td><td colspan="3">

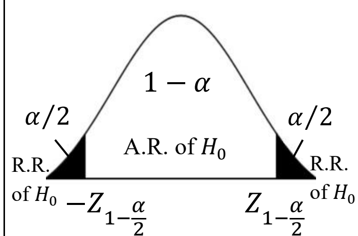
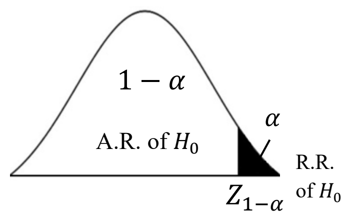
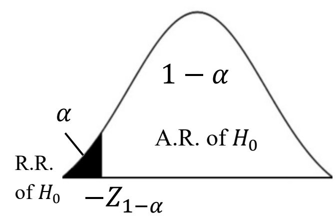
- The differences (D-observations): $D_i = X_i - Y_i$, $i = 1,2,\dots,n$
- Sample Mean of the D-observations: $\overline{D} = \frac{\sum_{i=1}^{n} D_i}{n}$
- Sample Variance of the D-observations: $S_D^2 = \frac{\sum_{i=1}^{n}(D_i-\overline{D})^2}{n-1}$
- Sample Standard Deviation of the D-observations: $S_D = \sqrt{S_D^2}$

</td></tr>
<tr><td colspan="4">Confidence Interval for $\mu_D = \mu_1 - \mu_2$</td></tr>
<tr><td>$100(1-\alpha)\%$ Confidence Interval for μ_D</td><td colspan="3">$\overline{D} \pm t_{1-\frac{\alpha}{2}} \frac{S_D}{\sqrt{n}}$, $df = v = n - 1$</td></tr>
<tr><td colspan="4">Hypotheses Testing for $\mu_D = \mu_1 - \mu_2$</td></tr>
<tr><td>Hypotheses</td><td>$H_0: \mu_1 = \mu_2$ vs
$H_A: \mu_1 \neq \mu_2$
or
$H_0: \mu_D = 0$ vs
$H_A: \mu_D \neq 0$</td><td>$H_0: \mu_1 \leq \mu_2$ vs
$H_A: \mu_1 > \mu_2$
or
$H_0: \mu_D \leq 0$ vs
$H_A: \mu_D > 0$</td><td>$H_0: \mu_1 \geq \mu_2$ vs
$H_A: \mu_1 < \mu_2$
or
$H_0: \mu_D \geq 0$ vs
$H_A: \mu_D < 0$</td></tr>
<tr><td>Test Statistic (T.S.)</td><td colspan="3">$T = \frac{\overline{D}}{S_D/\sqrt{n}} \sim t(n-1)$, $df = v = n - 1$</td></tr>
<tr><td>Rejection Region (R.R.) & Acceptance Region (A.R.) of H_0</td><td>
</td><td>
</td><td>
</td></tr>
<tr><td>Critical Value</td><td>$-t_{1-\frac{\alpha}{2}}$ and $t_{1-\frac{\alpha}{2}}$</td><td>$t_{1-\alpha}$</td><td>$-t_{1-\alpha}$</td></tr>
<tr><td rowspan="2">Decision</td><td colspan="3">We reject H_0 (and accept H_A) at the significance level α if:</td></tr>
<tr><td>$T < -t_{1-\frac{\alpha}{2}}$
or $T > t_{1-\frac{\alpha}{2}}$</td><td>$T > t_{1-\alpha}$</td><td>$T < -t_{1-\alpha}$</td></tr>
</table>

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

Hypotheses	$H_0: p = p_0 \ \ vs$ $H_A: p \neq p_0$	$H_0: p \leq p_0 \ \ vs$ $H_A: p > p$	$H_0: p \geq p_0 \ \ vs$ $H_A: p < p_0$
Test Statistic (T.S.)	$Z = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} \sim N(0,1)\ , \qquad \hat{p} = \dfrac{X}{n}$		
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 α 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.)	$Z = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\dfrac{\bar{p}(1-\bar{p})}{n_1} + \dfrac{\bar{p}(1-\bar{p})}{n_2}}} \sim N(0,1)$ $\hat{p}_1 = \dfrac{X_1}{n_1}, \;\; \hat{p}_2 = \dfrac{X_2}{n_2}, \;\; \bar{p} = \dfrac{X_1 + X_2}{n_1 + n_2}$		
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 α if:		
	$Z < -Z_{1-\frac{\alpha}{2}}$ or $Z > Z_{1-\frac{\alpha}{2}}$	$Z > Z_{1-\alpha}$	$Z < -Z_{1-\alpha}$