# AUA CS 108, Statistics, Fall 2019

## Lecture 40

Michael Poghosyan
YSU, AUA
michael@ysu.am, mpoghosyan@aua.am

25 Nov 2019

# Contents

# Last Lecture ReCap

- Describe the Asymptotic Test by MLE.

# Wald Test

**Model:** $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{F}_\theta$;

# Wald Test

**Model:** $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{F}_\theta$; We take any Estimator $\hat{\theta}$ which is Asymptotically Normal, when $\theta = \theta_0$:

$$W = \frac{\hat{\theta} - \theta_0}{\widehat{SE}(\hat{\theta})} \xrightarrow{D} \mathcal{N}(0, 1)$$

# Wald Test

**Model:** $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{F}_\theta$; We take any Estimator $\hat{\theta}$ which is Asymptotically Normal, when $\theta = \theta_0$:

$$W = \frac{\hat{\theta} - \theta_0}{\widehat{SE}(\hat{\theta})} \xrightarrow{D} \mathcal{N}(0,1)$$

**Null Hypothesis:** $\mathcal{H}_0 : \ \theta = \theta_0$

# Wald Test

**Model:** $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{F}_\theta$; We take any Estimator $\hat{\theta}$ which is Asymptotically Normal, when $\theta = \theta_0$:

$$W = \frac{\hat{\theta} - \theta_0}{\widehat{SE}(\hat{\theta})} \xrightarrow{D} \mathcal{N}(0, 1)$$

**Null Hypothesis:** $\mathcal{H}_0 : \ \theta = \theta_0$

**Asymptotic Significance Level:** $\alpha \in (0, 1)$;

# Wald Test

**Model:** $X_1, X_2, ..., X_n \stackrel{IID}{\sim} \mathcal{F}_\theta$; We take any Estimator $\hat{\theta}$ which is Asymptotically Normal, when $\theta = \theta_0$:

$$W = \frac{\hat{\theta} - \theta_0}{\widehat{SE}(\hat{\theta})} \xrightarrow{D} \mathcal{N}(0, 1)$$

**Null Hypothesis:** $\mathcal{H}_0: \ \theta = \theta_0$

**Asymptotic Significance Level:** $\alpha \in (0, 1)$;

**Test Statistics:** $W = \dfrac{\hat{\theta} - \theta_0}{\widehat{SE}(\hat{\theta})}$ or $W = \dfrac{\hat{\theta} - \theta_0}{SE(\theta_0)}$

# Wald Test

**Model:** $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{F}_\theta$; We take any Estimator $\hat{\theta}$ which is Asymptotically Normal, when $\theta = \theta_0$:

$$W = \frac{\hat{\theta} - \theta_0}{\widehat{SE}(\hat{\theta})} \xrightarrow{D} \mathcal{N}(0, 1)$$

**Null Hypothesis:** $\mathcal{H}_0 : \ \theta = \theta_0$

**Asymptotic Significance Level:** $\alpha \in (0, 1)$;

**Test Statistics:** $W = \dfrac{\hat{\theta} - \theta_0}{\widehat{SE}(\hat{\theta})}$ or $W = \dfrac{\hat{\theta} - \theta_0}{SE(\theta_0)}$

**Asymptotic Distrib of the TS Under $\mathcal{H}_0$:** $W \xrightarrow{D}$

# Wald Test

**Model:** $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{F}_\theta$; We take any Estimator $\hat{\theta}$ which is Asymptotically Normal, when $\theta = \theta_0$:

$$W = \frac{\hat{\theta} - \theta_0}{\widehat{SE}(\hat{\theta})} \xrightarrow{D} \mathcal{N}(0,1)$$

**Null Hypothesis:** $\mathcal{H}_0: \ \theta = \theta_0$

**Asymptotic Significance Level:** $\alpha \in (0,1)$;

**Test Statistics:** $W = \dfrac{\hat{\theta} - \theta_0}{\widehat{SE}(\hat{\theta})}$ or $W = \dfrac{\hat{\theta} - \theta_0}{SE(\theta_0)}$

**Asymptotic Distrib of the TS Under $\mathcal{H}_0$:** $W \xrightarrow{D} \mathcal{N}(0,1)$;

# Wald Test

**Model:** $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{F}_\theta$; We take any Estimator $\hat{\theta}$ which is Asymptotically Normal, when $\theta = \theta_0$:

$$W = \frac{\hat{\theta} - \theta_0}{\widehat{SE}(\hat{\theta})} \xrightarrow{D} \mathcal{N}(0, 1)$$

**Null Hypothesis:** $\mathcal{H}_0 : \ \theta = \theta_0$

**Asymptotic Significance Level:** $\alpha \in (0, 1)$;

**Test Statistics:** $W = \dfrac{\hat{\theta} - \theta_0}{\widehat{SE}(\hat{\theta})}$ or $W = \dfrac{\hat{\theta} - \theta_0}{SE(\theta_0)}$

**Asymptotic Distrib of the TS Under $\mathcal{H}_0$:** $W \xrightarrow{D} \mathcal{N}(0, 1)$;

| $\mathcal{H}_1$ is | RR is |
|---|---|
| $\theta \neq \theta_0$ | $\lvert W \rvert > z_{1-\frac{\alpha}{2}}$ |
| $\theta > \theta_0$ | $W > z_{1-\alpha}$ |
| $\theta < \theta_0$ | $W < z_\alpha$ |

# Note

**Note:** In all above Asymptotic Tests, one can replace the quantiles $z_p$ of the Standard Normal by the quantiles $t_{n-1,p}$ of $t(n-1)$, since, for large $n$,

$$t_{n-1,p} \approx z_p.$$

# Note

**Note:** In all above Asymptotic Tests, one can replace the quantiles $z_p$ of the Standard Normal by the quantiles $t_{n-1,p}$ of $t(n-1)$, since, for large $n$,

$$t_{n-1,p} \approx z_p.$$

```
a <- 0.05
qnorm(1-a/2)
```

```
## [1] 1.959964
```

```
n <- seq(50,250,50)
qt(1-a/2, df = n)
```

```
## [1] 2.008559 1.983972 1.975905 1.971896 1.969498
```

# Note

**Note:** In all above Asymptotic Tests, one can replace the quantiles $z_p$ of the Standard Normal by the quantiles $t_{n-1,p}$ of $t(n-1)$, since, for large $n$,

$$t_{n-1,p} \approx z_p.$$

```r
a <- 0.05
qnorm(1-a/2)
```

```
## [1] 1.959964
```

```r
n <- seq(50,250,50)
qt(1-a/2, df = n)
```

```
## [1] 2.008559 1.983972 1.975905 1.971896 1.969498
```

Another point is that, since $|t_{n-1,p}| > |z_p|$, $p \neq 0.5$, it is safer to have a little bit smaller Rejection Region:

# Note

**Note:** In all above Asymptotic Tests, one can replace the quantiles $z_p$ of the Standard Normal by the quantiles $t_{n-1,p}$ of $t(n-1)$, since, for large $n$,

$$t_{n-1,p} \approx z_p.$$

```
a <- 0.05
qnorm(1-a/2)
```

```
## [1] 1.959964
```

```
n <- seq(50,250,50)
qt(1-a/2, df = n)
```

```
## [1] 2.008559 1.983972 1.975905 1.971896 1.969498
```

Another point is that, since $|t_{n-1,p}| > |z_p|$, $p \neq 0.5$, it is safer to have a little bit smaller Rejection Region: say, for the Two-Sided Tests, if $|W| > t_{n-1,1-\alpha/2}$, then, for sure, also $|W| > z_{1-\alpha/2}$.

# Two Sample Tests

## $Z$-Test for the Difference of two Normals Means

**Model:** $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$,

# $Z$-Test for the Difference of two Normals Means

**Model:** $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$, $Y_1, Y_2, ..., Y_m \overset{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$,

# $Z$-Test for the Difference of two Normals Means

**Model:** $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$, $Y_1, Y_2, ..., Y_m \overset{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$, $\sigma_X, \sigma_Y$ **are known**,

# $Z$-Test for the Difference of two Normals Means

**Model:** $X_1, X_2, \ldots, X_n \overset{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$, $Y_1, Y_2, \ldots, Y_m \overset{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$, $\sigma_X, \sigma_Y$ **are known**, and $X_k$-s and $Y_j$-s are all Independent.

## $Z$-Test for the Difference of two Normals Means

**Model:** $X_1, X_2, ..., X_n \stackrel{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$, $Y_1, Y_2, ..., Y_m \stackrel{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$, $\sigma_X, \sigma_Y$ **are known**, and $X_k$-s and $Y_j$-s are all Independent. The Parameter of interest is $\mu_X - \mu_Y$;

## $Z$-Test for the Difference of two Normals Means

**Model:** $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$, $Y_1, Y_2, ..., Y_m \overset{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$, $\sigma_X, \sigma_Y$ **are known**, and $X_k$-s and $Y_j$-s are all Independent. The Parameter of interest is $\mu_X - \mu_Y$;

**Null Hypothesis:** $\mathcal{H}_0: \ \mu_X - \mu_Y = \mu_0$

# $Z$-Test for the Difference of two Normals Means

**Model:** $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$, $Y_1, Y_2, ..., Y_m \overset{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$, $\sigma_X, \sigma_Y$ **are known**, and $X_k$-s and $Y_j$-s are all Independent. The Parameter of interest is $\mu_X - \mu_Y$;

**Null Hypothesis:** $\mathcal{H}_0: \ \mu_X - \mu_Y = \mu_0$

**Significance Level:** $\alpha \in (0, 1)$;

# $Z$-Test for the Difference of two Normals Means

**Model:** $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$, $Y_1, Y_2, ..., Y_m \overset{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$, $\sigma_X, \sigma_Y$ **are known**, and $X_k$-s and $Y_j$-s are all Independent. The Parameter of interest is $\mu_X - \mu_Y$;

**Null Hypothesis:** $\mathcal{H}_0: \ \mu_X - \mu_Y = \mu_0$

**Significance Level:** $\alpha \in (0, 1)$;

**Test Statistics:** $Z =$

# $Z$-Test for the Difference of two Normals Means

**Model:** $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$, $Y_1, Y_2, ..., Y_m \overset{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$, $\sigma_X, \sigma_Y$ **are known**, and $X_k$-s and $Y_j$-s are all Independent. The Parameter of interest is $\mu_X - \mu_Y$;

**Null Hypothesis:** $\mathcal{H}_0 : \ \mu_X - \mu_Y = \mu_0$

**Significance Level:** $\alpha \in (0, 1)$;

**Test Statistics:** $Z = \dfrac{(\overline{X} - \overline{Y}) - \mu_0}{\phantom{x}}$

# $Z$-Test for the Difference of two Normals Means

**Model:** $X_1, X_2, ..., X_n \stackrel{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$, $Y_1, Y_2, ..., Y_m \stackrel{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$, $\sigma_X, \sigma_Y$ **are known**, and $X_k$-s and $Y_j$-s are all Independent. The Parameter of interest is $\mu_X - \mu_Y$;

**Null Hypothesis:** $\mathcal{H}_0 : \ \mu_X - \mu_Y = \mu_0$

**Significance Level:** $\alpha \in (0, 1)$;

**Test Statistics:** $Z = \dfrac{(\overline{X} - \overline{Y}) - \mu_0}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}}$

# $Z$-Test for the Difference of two Normals Means

**Model:** $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$, $Y_1, Y_2, ..., Y_m \overset{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$, $\sigma_X, \sigma_Y$ **are known**, and $X_k$-s and $Y_j$-s are all Independent. The Parameter of interest is $\mu_X - \mu_Y$;

**Null Hypothesis:** $\mathcal{H}_0: \ \mu_X - \mu_Y = \mu_0$

**Significance Level:** $\alpha \in (0, 1)$;

**Test Statistics:** $Z = \dfrac{(\overline{X} - \overline{Y}) - \mu_0}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}}$

**Distrib of the Test-Statistics Under $\mathcal{H}_0$:** $Z \sim$

# $Z$-Test for the Difference of two Normals Means

**Model:** $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$, $Y_1, Y_2, ..., Y_m \overset{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$, $\sigma_X, \sigma_Y$ **are known**, and $X_k$-s and $Y_j$-s are all Independent. The Parameter of interest is $\mu_X - \mu_Y$;

**Null Hypothesis:** $\mathcal{H}_0 : \ \mu_X - \mu_Y = \mu_0$

**Significance Level:** $\alpha \in (0, 1)$;

**Test Statistics:** $Z = \dfrac{(\overline{X} - \overline{Y}) - \mu_0}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}}$

**Distrib of the Test-Statistics Under $\mathcal{H}_0$:** $Z \sim \mathcal{N}(0, 1)$;

# $Z$-Test for the Difference of two Normals Means

**Model:** $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$, $Y_1, Y_2, ..., Y_m \overset{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$, $\sigma_X, \sigma_Y$ **are known**, and $X_k$-s and $Y_j$-s are all Independent. The Parameter of interest is $\mu_X - \mu_Y$;

**Null Hypothesis:** $\mathcal{H}_0: \ \mu_X - \mu_Y = \mu_0$

**Significance Level:** $\alpha \in (0, 1)$;

**Test Statistics:** $Z = \dfrac{(\overline{X} - \overline{Y}) - \mu_0}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}}$

**Distrib of the Test-Statistics Under** $\mathcal{H}_0$**:** $Z \sim \mathcal{N}(0, 1)$;

| $\mathcal{H}_1$ is | RR is |
|---|---|
| $\mu_X - \mu_Y \neq \mu_0$ | $\lvert Z \rvert > z_{1-\frac{\alpha}{2}}$ |
| $\mu_X - \mu_Y > \mu_0$ | $Z > z_{1-\alpha}$ |
| $\mu_X - \mu_Y < \mu_0$ | $Z < z_{\alpha}$ |

## $t$-Test for the Difference of two Normals Means

**Model:** $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$,

# $t$-Test for the Difference of two Normals Means

**Model:** $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2),\ Y_1, Y_2, ..., Y_m \overset{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2),$

# $t$-Test for the Difference of two Normals Means

**Model:** $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$, $Y_1, Y_2, ..., Y_m \overset{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$, $\sigma_X, \sigma_Y$ **are unknown**,

# $t$-Test for the Difference of two Normals Means

**Model:** $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$, $Y_1, Y_2, ..., Y_m \overset{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$, $\sigma_X, \sigma_Y$ **are unknown**, and $X_k$-s and $Y_j$-s are all Independent.

## $t$-Test for the Difference of two Normals Means

**Model:** $X_1, X_2, ..., X_n \stackrel{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$, $Y_1, Y_2, ..., Y_m \stackrel{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$, $\sigma_X, \sigma_Y$ **are unknown**, and $X_k$-s and $Y_j$-s are all Independent. The Parameter of interest is $\mu_X - \mu_Y$;

# $t$-Test for the Difference of two Normals Means

**Model:** $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$, $Y_1, Y_2, ..., Y_m \overset{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$, $\sigma_X, \sigma_Y$ **are unknown**, and $X_k$-s and $Y_j$-s are all Independent. The Parameter of interest is $\mu_X - \mu_Y$;

**Additional Assumption:** $\sigma_X = \sigma_Y$

## $t$-Test for the Difference of two Normals Means

**Model:** $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$, $Y_1, Y_2, ..., Y_m \overset{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$, $\sigma_X, \sigma_Y$ **are unknown**, and $X_k$-s and $Y_j$-s are all Independent. The Parameter of interest is $\mu_X - \mu_Y$;

**Additional Assumption:** $\sigma_X = \sigma_Y$ (can be Tested separately, by $F$-Test)

# $t$-Test for the Difference of two Normals Means

**Model:** $X_1, X_2, ..., X_n \stackrel{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$, $Y_1, Y_2, ..., Y_m \stackrel{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$, $\sigma_X, \sigma_Y$ **are unknown**, and $X_k$-s and $Y_j$-s are all Independent. The Parameter of interest is $\mu_X - \mu_Y$;

**Additional Assumption:** $\sigma_X = \sigma_Y$ (can be Tested separately, by $F$-Test)

**Null Hypothesis:** $\mathcal{H}_0: \ \mu_X - \mu_Y = \mu_0$

# $t$-Test for the Difference of two Normals Means

**Model:** $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$, $Y_1, Y_2, ..., Y_m \overset{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$, $\sigma_X, \sigma_Y$ **are unknown**, and $X_k$-s and $Y_j$-s are all Independent. The Parameter of interest is $\mu_X - \mu_Y$;

**Additional Assumption:** $\sigma_X = \sigma_Y$ (can be Tested separately, by $F$-Test)

**Null Hypothesis:** $\mathcal{H}_0: \ \mu_X - \mu_Y = \mu_0$

**Significance Level:** $\alpha \in (0, 1)$;

# $t$-Test for the Difference of two Normals Means

**Model:** $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$, $Y_1, Y_2, ..., Y_m \overset{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$, $\sigma_X, \sigma_Y$ **are unknown**, and $X_k$-s and $Y_j$-s are all Independent. The Parameter of interest is $\mu_X - \mu_Y$;

**Additional Assumption:** $\sigma_X = \sigma_Y$ (can be Tested separately, by $F$-Test)

**Null Hypothesis:** $\mathcal{H}_0: \ \mu_X - \mu_Y = \mu_0$

**Significance Level:** $\alpha \in (0, 1)$;

**Test Statistics:** $t = \dfrac{(\overline{X} - \overline{Y}) - \mu_0}{S_p \cdot \sqrt{\frac{1}{n} + \frac{1}{m}}}$, where $S_p$ is the **Pooled Sample Deviation**:

## $t$-Test for the Difference of two Normals Means

**Model:** $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$, $Y_1, Y_2, ..., Y_m \overset{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$, $\sigma_X, \sigma_Y$ **are unknown**, and $X_k$-s and $Y_j$-s are all Independent. The Parameter of interest is $\mu_X - \mu_Y$;

**Additional Assumption:** $\sigma_X = \sigma_Y$ (can be Tested separately, by $F$-Test)

**Null Hypothesis:** $\mathcal{H}_0: \ \mu_X - \mu_Y = \mu_0$

**Significance Level:** $\alpha \in (0, 1)$;

**Test Statistics:** $t = \dfrac{(\overline{X} - \overline{Y}) - \mu_0}{S_p \cdot \sqrt{\frac{1}{n} + \frac{1}{m}}}$, where $S_p$ is the **Pooled Sample Deviation**:

$$S_P^2 = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2} = \frac{\sum_{k=1}^n (X_k - \overline{X})^2 + \sum_{k=1}^m (Y_k - \overline{Y})^2}{n+m-2}.$$

# $t$-Test for the Difference of two Normals Means, Cont'd

**Distrib of the Test-Statistics Under $\mathcal{H}_0$:** $t \sim$

# $t$-Test for the Difference of two Normals Means, Cont'd

**Distrib of the Test-Statistics Under $\mathcal{H}_0$:** $t \sim t(n+m-2)$;

# $t$-Test for the Difference of two Normals Means, Cont'd

**Distrib of the Test-Statistics Under $\mathcal{H}_0$:** $t \sim t(n+m-2)$;

**Rejection Region:**

| $\mathcal{H}_1$ is | RR is |
|---|---|
| $\mu_X - \mu_Y \neq \mu_0$ | $\lvert t \rvert > t_{n+m-2,1-\frac{\alpha}{2}}$ |
| $\mu_X - \mu_Y > \mu_0$ | $t > t_{n+m-2,1-\alpha}$ |
| $\mu_X - \mu_Y < \mu_0$ | $t < t_{n+m-2,\alpha}$ |

# $t$-Test for the Difference of two Normals Means, Cont'd

**Distrib of the Test-Statistics Under $\mathcal{H}_0$:** $t \sim t(n+m-2)$;

**Rejection Region:**

| $\mathcal{H}_1$ is | RR is |
|---|---|
| $\mu_X - \mu_Y \neq \mu_0$ | $\lvert t \rvert > t_{n+m-2,1-\frac{\alpha}{2}}$ |
| $\mu_X - \mu_Y > \mu_0$ | $t > t_{n+m-2,1-\alpha}$ |
| $\mu_X - \mu_Y < \mu_0$ | $t < t_{n+m-2,\alpha}$ |

**Note:** This Test is called the **Pooled $t$-Test**

# $t$-Test for the Difference of two Normals Means

**Model:** $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$,

# $t$-Test for the Difference of two Normals Means

**Model:** $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2),\ Y_1, Y_2, ..., Y_m \overset{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2),$

# $t$-Test for the Difference of two Normals Means

**Model:** $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$, $Y_1, Y_2, ..., Y_m \overset{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$, $\sigma_X, \sigma_Y$ **are unknown**,

# $t$-Test for the Difference of two Normals Means

**Model:** $X_1, X_2, ..., X_n \stackrel{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$, $Y_1, Y_2, ..., Y_m \stackrel{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$, $\sigma_X, \sigma_Y$ **are unknown**, and $X_k$-s and $Y_j$-s are all Independent.

# $t$-Test for the Difference of two Normals Means

**Model:** $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$, $Y_1, Y_2, ..., Y_m \overset{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$, $\sigma_X, \sigma_Y$ **are unknown**, and $X_k$-s and $Y_j$-s are all Independent. The Parameter of interest is $\mu_X - \mu_Y$;

# $t$-Test for the Difference of two Normals Means

**Model:** $X_1, X_2, ..., X_n \stackrel{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$, $Y_1, Y_2, ..., Y_m \stackrel{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$, $\sigma_X, \sigma_Y$ **are unknown**, and $X_k$-s and $Y_j$-s are all Independent. The Parameter of interest is $\mu_X - \mu_Y$;

**Additional Assumption:** $\sigma_X \neq \sigma_Y$ (can be Tested separately, by $F$-Test)

# $t$-Test for the Difference of two Normals Means

**Model:** $X_1, X_2, ..., X_n \stackrel{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$, $Y_1, Y_2, ..., Y_m \stackrel{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$, $\sigma_X, \sigma_Y$ **are unknown**, and $X_k$-s and $Y_j$-s are all Independent. The Parameter of interest is $\mu_X - \mu_Y$;

**Additional Assumption:** $\sigma_X \neq \sigma_Y$ (can be Tested separately, by $F$-Test)

**Null Hypothesis:** $\mathcal{H}_0: \ \mu_X - \mu_Y = \mu_0$

# $t$-Test for the Difference of two Normals Means

**Model:** $X_1, X_2, ..., X_n \stackrel{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$, $Y_1, Y_2, ..., Y_m \stackrel{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$, $\sigma_X, \sigma_Y$ **are unknown**, and $X_k$-s and $Y_j$-s are all Independent. The Parameter of interest is $\mu_X - \mu_Y$;

**Additional Assumption:** $\sigma_X \neq \sigma_Y$ (can be Tested separately, by $F$-Test)

**Null Hypothesis:** $\mathcal{H}_0: \ \mu_X - \mu_Y = \mu_0$

**Significance Level:** $\alpha \in (0, 1)$;

# $t$-Test for the Difference of two Normals Means

**Model:** $X_1, X_2, ..., X_n \stackrel{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$, $Y_1, Y_2, ..., Y_m \stackrel{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$, $\sigma_X, \sigma_Y$ **are unknown**, and $X_k$-s and $Y_j$-s are all Independent. The Parameter of interest is $\mu_X - \mu_Y$;

**Additional Assumption:** $\sigma_X \neq \sigma_Y$ (can be Tested separately, by $F$-Test)

**Null Hypothesis:** $\mathcal{H}_0: \ \mu_X - \mu_Y = \mu_0$

**Significance Level:** $\alpha \in (0, 1)$;

**Test Statistics:**

$$t = \frac{(\overline{X} - \overline{Y}) - \mu_0}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}},$$

# $t$-Test for the Difference of two Normals Means

**Model:** $X_1, X_2, ..., X_n \stackrel{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$, $Y_1, Y_2, ..., Y_m \stackrel{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$, $\sigma_X, \sigma_Y$ **are unknown**, and $X_k$-s and $Y_j$-s are all Independent. The Parameter of interest is $\mu_X - \mu_Y$;

**Additional Assumption:** $\sigma_X \neq \sigma_Y$ (can be Tested separately, by $F$-Test)

**Null Hypothesis:** $\mathcal{H}_0: \ \mu_X - \mu_Y = \mu_0$

**Significance Level:** $\alpha \in (0, 1)$;

**Test Statistics:**

$$t = \frac{(\overline{X} - \overline{Y}) - \mu_0}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}},$$

where $S_X$ and $S_Y$ are the Sample SDs for $X$ and $Y$, respectively.

# $t$-Test for the Difference of two Normals Means, Cont'd

**Distrib of the Test-Statistics Under $\mathcal{H}_0$:** Approximately,

$t \approx$

# $t$-Test for the Difference of two Normals Means, Cont'd

**Distrib of the Test-Statistics Under $\mathcal{H}_0$:** Approximately, $t \approx t(\nu)$, where $\nu$ is

# $t$-Test for the Difference of two Normals Means, Cont'd

**Distrib of the Test-Statistics Under $\mathcal{H}_0$:** Approximately, $t \approx t(\nu)$, where $\nu$ is some scary thing...

# $t$-Test for the Difference of two Normals Means, Cont'd

**Distrib of the Test-Statistics Under $\mathcal{H}_0$:** Approximately, $t \approx t(\nu)$, where $\nu$ is some scary thing... given by

$$\nu = \left\lfloor \frac{\left(\frac{S_X^2}{n} + \frac{S_Y^2}{m}\right)^2}{\frac{\left(S_X^2/n\right)^2}{n-1} + \frac{\left(S_Y^2/m\right)^2}{m-1}} \right\rfloor$$

# $t$-Test for the Difference of two Normals Means, Cont'd

**Distrib of the Test-Statistics Under $\mathcal{H}_0$:** Approximately, $t \approx t(\nu)$, where $\nu$ is some scary thing... given by

$$\nu = \left\lfloor \frac{\left(\frac{S_X^2}{n} + \frac{S_Y^2}{m}\right)^2}{\frac{\left(S_X^2/n\right)^2}{n-1} + \frac{\left(S_Y^2/m\right)^2}{m-1}} \right\rfloor$$

**Rejection Region:**

| $\mathcal{H}_1$ is | RR is |
|---|---|
| $\mu_X - \mu_Y \neq \mu_0$ | $\lvert t \rvert > t_{\nu,1-\frac{\alpha}{2}}$ |
| $\mu_X - \mu_Y > \mu_0$ | $t > t_{\nu,1-\alpha}$ |
| $\mu_X - \mu_Y < \mu_0$ | $t < t_{\nu,\alpha}$ |

## $t$-Test for the Difference of two Normals Means, Cont'd

**Distrib of the Test-Statistics Under $\mathcal{H}_0$:** Approximately, $t \approx t(\nu)$, where $\nu$ is some scary thing... given by

$$\nu = \left\lfloor \frac{\left(\frac{S_X^2}{n} + \frac{S_Y^2}{m}\right)^2}{\frac{\left(S_X^2/n\right)^2}{n-1} + \frac{\left(S_Y^2/m\right)^2}{m-1}} \right\rfloor$$

**Rejection Region:**

| $\mathcal{H}_1$ is | RR is |
|---|---|
| $\mu_X - \mu_Y \neq \mu_0$ | $\lvert t \rvert > t_{\nu,1-\frac{\alpha}{2}}$ |
| $\mu_X - \mu_Y > \mu_0$ | $t > t_{\nu,1-\alpha}$ |
| $\mu_X - \mu_Y < \mu_0$ | $t < t_{\nu,\alpha}$ |

**Note:** The formula above for the DF $\nu$ is called **Welch – Satterthwaite Equation**, and the Tests is called the **Welch Test**.

## Paired $t$-Test for the Difference of two Normals Means

**Model:** $X_1, X_2, ..., X_n \stackrel{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$,

# Paired $t$-Test for the Difference of two Normals Means

**Model:** $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$, $Y_1, Y_2, ..., Y_n \overset{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$,

# Paired $t$-Test for the Difference of two Normals Means

**Model:** $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$, $Y_1, Y_2, ..., Y_n \overset{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$, $\sigma_X, \sigma_Y$ **are unknown**.

## Paired $t$-Test for the Difference of two Normals Means

**Model:** $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$, $Y_1, Y_2, ..., Y_n \overset{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$, $\sigma_X, \sigma_Y$ **are unknown**. The Parameter of interest is $\mu_X - \mu_Y$;

# Paired $t$-Test for the Difference of two Normals Means

**Model:** $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$, $Y_1, Y_2, ..., Y_n \overset{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$, $\sigma_X, \sigma_Y$ **are unknown**. The Parameter of interest is $\mu_X - \mu_Y$;

**Note:** Here we have the same number of $X_k$ and $Y_k$;

# Paired $t$-Test for the Difference of two Normals Means

**Model:** $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$, $Y_1, Y_2, ..., Y_n \overset{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$, $\sigma_X, \sigma_Y$ **are unknown**. The Parameter of interest is $\mu_X - \mu_Y$;

**Note:** Here we have the same number of $X_k$ and $Y_k$; also, importantly, $X_k$ and $Y_k$ can be dependent!

# Paired $t$-Test for the Difference of two Normals Means

**Model:** $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$, $Y_1, Y_2, ..., Y_n \overset{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$, $\sigma_X, \sigma_Y$ **are unknown**. The Parameter of interest is $\mu_X - \mu_Y$;

**Note:** Here we have the same number of $X_k$ and $Y_k$; also, importantly, $X_k$ and $Y_k$ can be dependent!

**Notation:** $D_k = X_k - Y_k$;

# Paired $t$-Test for the Difference of two Normals Means

**Model:** $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$, $Y_1, Y_2, ..., Y_n \overset{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$, $\sigma_X, \sigma_Y$ **are unknown**. The Parameter of interest is $\mu_X - \mu_Y$;

**Note:** Here we have the same number of $X_k$ and $Y_k$; also, importantly, $X_k$ and $Y_k$ can be dependent!

**Notation:** $D_k = X_k - Y_k$; clearly,

$$\mathbb{E}(D_k) = \mu_X - \mu_Y.$$

# Paired $t$-Test for the Difference of two Normals Means

**Model:** $X_1, X_2, \ldots, X_n \overset{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$, $Y_1, Y_2, \ldots, Y_n \overset{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$, $\sigma_X, \sigma_Y$ **are unknown**. The Parameter of interest is $\mu_X - \mu_Y$;

**Note:** Here we have the same number of $X_k$ and $Y_k$; also, importantly, $X_k$ and $Y_k$ can be dependent!

**Notation:** $D_k = X_k - Y_k$; clearly,

$$\mathbb{E}(D_k) = \mu_X - \mu_Y.$$

The Variance of $D_k$, although the same, $\sigma_D^2 = Var(X_k - Y_k)$, cannot be calculated, since $X_k$ and $Y_k$ can be dependent. But that's OK, we do not need it.

[1] The Test will work also in the case when the Differences are nor Normally Distributed, but the Sample Size $n$ is large. We jut need to use the CLT.

# Paired $t$-Test for the Difference of two Normals Means

**Model:** $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$, $Y_1, Y_2, ..., Y_n \overset{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$, $\sigma_X, \sigma_Y$ **are unknown**. The Parameter of interest is $\mu_X - \mu_Y$;

**Note:** Here we have the same number of $X_k$ and $Y_k$; also, importantly, $X_k$ and $Y_k$ can be dependent!

**Notation:** $D_k = X_k - Y_k$; clearly,

$$\mathbb{E}(D_k) = \mu_X - \mu_Y.$$

The Variance of $D_k$, although the same, $\sigma_D^2 = Var(X_k - Y_k)$, cannot be calculated, since $X_k$ and $Y_k$ can be dependent. But that's OK, we do not need it.

**Additional Assumption:** We will assume that the Differences $D_k$ are Normally Distributed[1].

---

[1] The Test will work also in the case when the Differences are nor Normally Distributed, but the Sample Size $n$ is large. We jut need to use the CLT.

# Paired $t$-Test for the Difference of two Normals Means

**Model:** $X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$, $Y_1, Y_2, ..., Y_n \overset{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$, $\sigma_X, \sigma_Y$ **are unknown**. The Parameter of interest is $\mu_X - \mu_Y$;

**Note:** Here we have the same number of $X_k$ and $Y_k$; also, importantly, $X_k$ and $Y_k$ can be dependent!

**Notation:** $D_k = X_k - Y_k$; clearly,

$$\mathbb{E}(D_k) = \mu_X - \mu_Y.$$

The Variance of $D_k$, although the same, $\sigma_D^2 = Var(X_k - Y_k)$, cannot be calculated, since $X_k$ and $Y_k$ can be dependent. But that's OK, we do not need it.

**Additional Assumption:** We will assume that the Differences $D_k$ are Normally Distributed[1].

**Null Hypothesis:** $\mathcal{H}_0: \ \mu_X - \mu_Y = \mu_0$

---

[1]The Test will work also in the case when the Differences are nor Normally Distributed, but the Sample Size $n$ is large. We jut need to use the CLT.

# Paired $t$-Test for the Difference of two Normals Means

**Model:** $X_1, X_2, \dots, X_n \overset{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$, $Y_1, Y_2, \dots, Y_n \overset{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$, $\sigma_X, \sigma_Y$ **are unknown**. The Parameter of interest is $\mu_X - \mu_Y$;

**Note:** Here we have the same number of $X_k$ and $Y_k$; also, importantly, $X_k$ and $Y_k$ can be dependent!

**Notation:** $D_k = X_k - Y_k$; clearly,

$$\mathbb{E}(D_k) = \mu_X - \mu_Y.$$

The Variance of $D_k$, although the same, $\sigma_D^2 = Var(X_k - Y_k)$, cannot be calculated, since $X_k$ and $Y_k$ can be dependent. But that's OK, we do not need it.

**Additional Assumption:** We will assume that the Differences $D_k$ are Normally Distributed[1].

**Null Hypothesis:** $\mathcal{H}_0: \ \mu_X - \mu_Y = \mu_0$

**Asymptotic Significance Level:** $\alpha \in (0, 1)$;

[1] The Test will work also in the case when the Differences are nor Normally Distributed, but the Sample Size $n$ is large. We jut need to use the CLT.

# Paired $t$-Test for the Difference of two Normals Means, Cont'd

**Test Statistics:** $t = \dfrac{\overline{D} - \mu_0}{S_D/\sqrt{n}}$, where $S_D$ is the Sample Deviation of $D$.

[2] Or, Asymptotically, $t \approx t(n-1)$ or $t \approx \mathcal{N}(0,1)$, if $D_k$-s are not Normal, but $n$ is large.

# Paired $t$-Test for the Difference of two Normals Means, Cont'd

**Test Statistics:** $t = \dfrac{\overline{D} - \mu_0}{S_D/\sqrt{n}}$, where $S_D$ is the Sample Deviation of $D$.

**Distrib of the Test-Statistics Under $\mathcal{H}_0$:**[2] $t \sim t(n-1)$;

[2] Or, Asymptotically, $t \approx t(n-1)$ or $t \approx \mathcal{N}(0,1)$, if $D_k$-s are not Normal, but $n$ is large.

# Paired $t$-Test for the Difference of two Normals Means, Cont'd

**Test Statistics:** $t = \dfrac{\overline{D} - \mu_0}{S_D/\sqrt{n}}$, where $S_D$ is the Sample Deviation of $D$.

**Distrib of the Test-Statistics Under $\mathcal{H}_0$:**[2] $t \sim t(n-1)$;

**Rejection Region:**

| $\mathcal{H}_1$ is | RR is |
|---|---|
| $\mu_X - \mu_Y \neq \mu_0$ | $\lvert t \rvert > t_{n-1,1-\frac{\alpha}{2}}$ |
| $\mu_X - \mu_Y > \mu_0$ | $t > t_{n-1,1-\alpha}$ |
| $\mu_X - \mu_Y < \mu_0$ | $t < t_{n-1,\alpha}$ |

[2] Or, Asymptotically, $t \approx t(n-1)$ or $t \approx \mathcal{N}(0,1)$, if $D_k$-s are not Normal, but $n$ is large.

# Paired $t$-Test for the Difference of two Normals Means, Cont'd

**Test Statistics:** $t = \dfrac{\overline{D} - \mu_0}{S_D/\sqrt{n}}$, where $S_D$ is the Sample Deviation of $D$.

**Distrib of the Test-Statistics Under $\mathcal{H}_0$:**[2] $t \sim t(n-1)$;

**Rejection Region:**

| $\mathcal{H}_1$ is | RR is |
|---|---|
| $\mu_X - \mu_Y \neq \mu_0$ | $\lvert t\rvert > t_{n-1,1-\frac{\alpha}{2}}$ |
| $\mu_X - \mu_Y > \mu_0$ | $t > t_{n-1,1-\alpha}$ |
| $\mu_X - \mu_Y < \mu_0$ | $t < t_{n-1,\alpha}$ |

**Note:** This Test is called the **Paired $t$-Test**

[2] Or, Asymptotically, $t \approx t(n-1)$ or $t \approx \mathcal{N}(0,1)$, if $D_k$-s are not Normal, but $n$ is large.

# Two Sample test for Proportions

**Model:** $X_1, X_2, ..., X_n \overset{IID}{\sim} Bernoulli(p_X)$,

# Two Sample test for Proportions

**Model:** $X_1, X_2, ..., X_n \overset{IID}{\sim} Bernoulli(p_X)$,
$Y_1, Y_2, ..., Y_m \overset{IID}{\sim} Bernoulli(p_Y)$,

# Two Sample test for Proportions

**Model:** $X_1, X_2, ..., X_n \overset{IID}{\sim} Bernoulli(p_X)$,
$Y_1, Y_2, ..., Y_m \overset{IID}{\sim} Bernoulli(p_Y)$, and $X_k$-s and $Y_j$-s are all Independent.

# Two Sample test for Proportions

**Model:** $X_1, X_2, ..., X_n \overset{IID}{\sim} Bernoulli(p_X)$,
$Y_1, Y_2, ..., Y_m \overset{IID}{\sim} Bernoulli(p_Y)$, and $X_k$-s and $Y_j$-s are all Independent. The Parameter of interest is $p_X - p_Y$;

# Two Sample test for Proportions

**Model:** $X_1, X_2, ..., X_n \overset{IID}{\sim} Bernoulli(p_X)$, $Y_1, Y_2, ..., Y_m \overset{IID}{\sim} Bernoulli(p_Y)$, and $X_k$-s and $Y_j$-s are all Independent. The Parameter of interest is $p_X - p_Y$;

**Null Hypothesis:** $\mathcal{H}_0: \ p_X - p_Y = p_0$

# Two Sample test for Proportions

**Model:** $X_1, X_2, ..., X_n \stackrel{IID}{\sim} Bernoulli(p_X)$,
$Y_1, Y_2, ..., Y_m \stackrel{IID}{\sim} Bernoulli(p_Y)$, and $X_k$-s and $Y_j$-s are all Independent. The Parameter of interest is $p_X - p_Y$;

**Null Hypothesis:** $\mathcal{H}_0 : \; p_X - p_Y = p_0$

**Significance Level:** $\alpha \in (0, 1)$;

# Two Sample test for Proportions

**Model:** $X_1, X_2, ..., X_n \overset{IID}{\sim} Bernoulli(p_X)$, $Y_1, Y_2, ..., Y_m \overset{IID}{\sim} Bernoulli(p_Y)$, and $X_k$-s and $Y_j$-s are all Independent. The Parameter of interest is $p_X - p_Y$;

**Null Hypothesis:** $\mathcal{H}_0 : \; p_X - p_Y = p_0$

**Significance Level:** $\alpha \in (0, 1)$;

**Test Statistics:**

$$Z = \frac{(\overline{X} - \overline{Y}) - p_0}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n} + \frac{1}{m}\right)}} \qquad or \qquad Z = \frac{(\overline{X} - \overline{Y}) - p_0}{\sqrt{\frac{\overline{X}(1-\overline{X})}{n} + \frac{\overline{Y}(1-\overline{Y})}{m}}}$$

# Two Sample test for Proportions

**Model:** $X_1, X_2, ..., X_n \overset{IID}{\sim} Bernoulli(p_X)$,
$Y_1, Y_2, ..., Y_m \overset{IID}{\sim} Bernoulli(p_Y)$, and $X_k$-s and $Y_j$-s are all Independent. The Parameter of interest is $p_X - p_Y$;

**Null Hypothesis:** $\mathcal{H}_0 : \; p_X - p_Y = p_0$

**Significance Level:** $\alpha \in (0, 1)$;

**Test Statistics:**

$$Z = \frac{(\overline{X} - \overline{Y}) - p_0}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n} + \frac{1}{m}\right)}} \qquad or \qquad Z = \frac{(\overline{X} - \overline{Y}) - p_0}{\sqrt{\frac{\overline{X}(1-\overline{X})}{n} + \frac{\overline{Y}(1-\overline{Y})}{m}}}$$

where $\hat{p}$ is the **Pooled Sample Proportion:**

$$\hat{p} = \frac{n}{n+m} \cdot \overline{X} + \frac{m}{n+m} \cdot \overline{Y}$$

# Two Sample test for Proportions

**Model:** $X_1, X_2, ..., X_n \overset{IID}{\sim} Bernoulli(p_X)$,
$Y_1, Y_2, ..., Y_m \overset{IID}{\sim} Bernoulli(p_Y)$, and $X_k$-s and $Y_j$-s are all Independent. The Parameter of interest is $p_X - p_Y$;

**Null Hypothesis:** $\mathcal{H}_0: \ p_X - p_Y = p_0$

**Significance Level:** $\alpha \in (0, 1)$;

**Test Statistics:**

$$Z = \frac{(\overline{X} - \overline{Y}) - p_0}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n} + \frac{1}{m}\right)}} \qquad or \qquad Z = \frac{(\overline{X} - \overline{Y}) - p_0}{\sqrt{\frac{\overline{X}(1-\overline{X})}{n} + \frac{\overline{Y}(1-\overline{Y})}{m}}}$$

where $\hat{p}$ is the **Pooled Sample Proportion:**

$$\hat{p} = \frac{n}{n+m} \cdot \overline{X} + \frac{m}{n+m} \cdot \overline{Y} = \frac{X_1 + ... + X_n + Y_1 + ... + Y_m}{n+m}.$$

# Two Sample test for Proportions, Cont'd

**Distrib of the Test-Statistics Under $\mathcal{H}_0$:** Approximately,
$Z \approx$

# Two Sample test for Proportions, Cont'd

**Distrib of the Test-Statistics Under $\mathcal{H}_0$:** Approximately,
$Z \approx \mathcal{N}(0, 1)$

# Two Sample test for Proportions, Cont'd

**Distrib of the Test-Statistics Under $\mathcal{H}_0$:** Approximately, $Z \approx \mathcal{N}(0,1)$

**Rejection Region:**

| $\mathcal{H}_1$ is | RR is |
|---|---|
| $p_X - p_Y \neq p_0$ | $\lvert Z \rvert > z_{1-\frac{\alpha}{2}}$ |
| $p_X - p_Y > p_0$ | $Z > z_{1-\alpha}$ |
| $p_X - p_Y < p_0$ | $Z < z_{\alpha}$ |