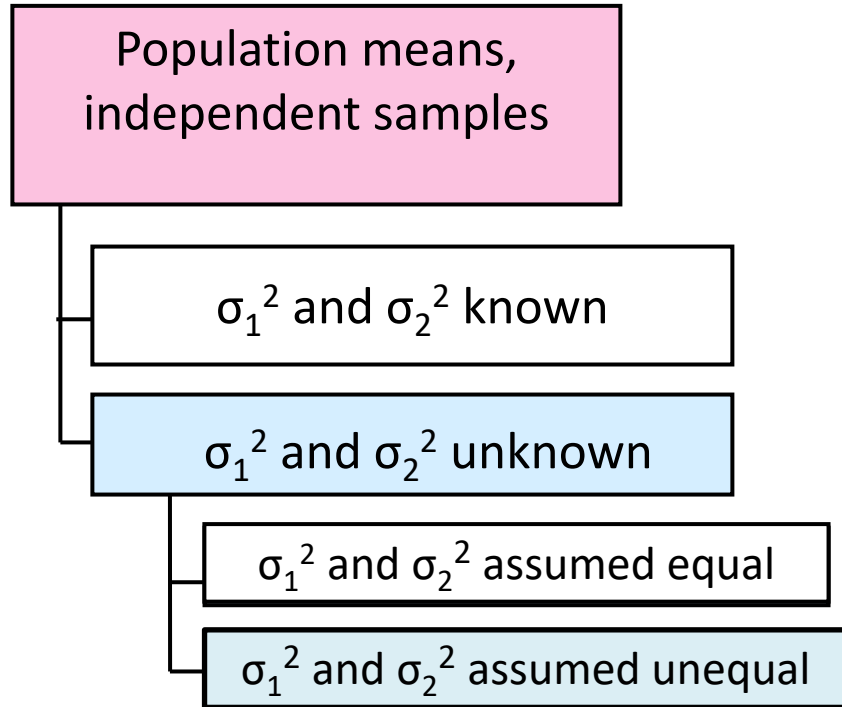# Hypothesis Testing: Two sample test

**Dr. A. Ramesh**
DEPARTMENT OF MANAGEMENT
IIT ROORKEE

# $\sigma_1^2$ and $\sigma_2^2$ Unknown, Assumed Unequal

Population means, independent samples

$\sigma_1^2$ and $\sigma_2^2$ known

$\sigma_1^2$ and $\sigma_2^2$ unknown
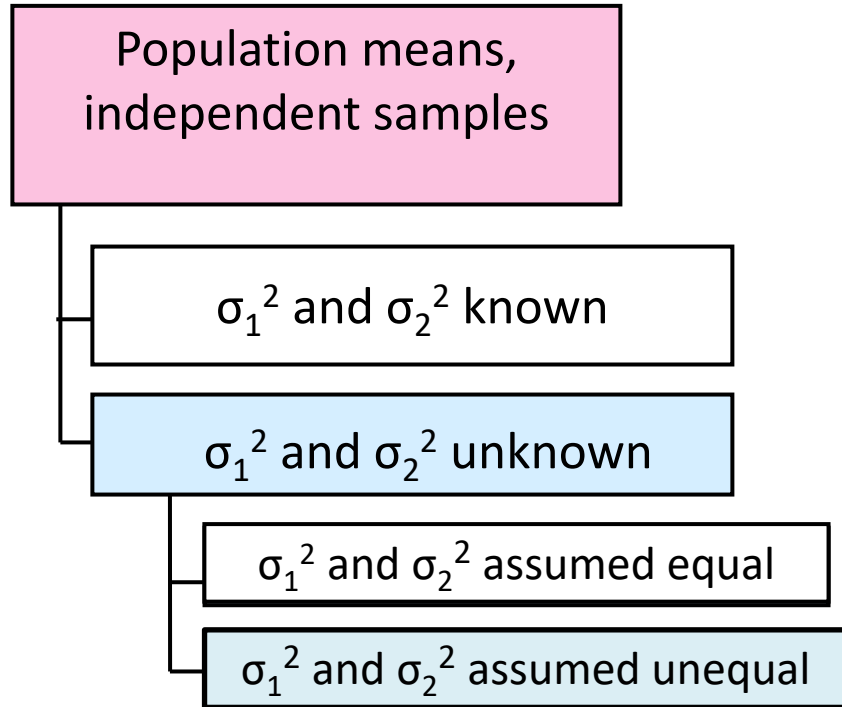
$\sigma_1^2$ and $\sigma_2^2$ assumed equal

$\sigma_1^2$ and $\sigma_2^2$ assumed unequal

**Assumptions:**

- Samples are randomly and independently drawn

- Populations are normally distributed

- Population variances are unknown and assumed unequal

# $\sigma_1^2$ and $\sigma_2^2$ Unknown: Assumed Unequal

Population means, independent samples

$\sigma_1^2$ and $\sigma_2^2$ known

$\sigma_1^2$ and $\sigma_2^2$ unknown

$\sigma_1^2$ and $\sigma_2^2$ assumed equal

$\sigma_1^2$ and $\sigma_2^2$ assumed unequal

**Forming interval estimates:**

- The population variances are assumed unequal, so a pooled variance is not appropriate
- Use a $t$ value with $\nu$ degrees of freedom, where

$$\nu = \frac{\left[ \left(\frac{s_1^2}{n_1}\right) + \left(\frac{s_2^2}{n_2}\right) \right]^2}{\left(\frac{s_1^2}{n_1}\right)^2 / (n_1 - 1) + \left(\frac{s_2^2}{n_2}\right)^2 / (n_2 - 1)}$$

# Test Statistic: $\sigma_1^2$ and $\sigma_2^2$ Unknown, Unequal

$\sigma_1^2$ and $\sigma_2^2$ unknown

$\sigma_1^2$ and $\sigma_2^2$ assumed equal

$\sigma_1^2$ and $\sigma_2^2$ assumed unequal

The test statistic for

$\mu_1 - \mu_2$   is:

$$t = \frac{(\overline{x}_1 - \overline{x}_2) - D_0}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$$

Where t has $\nu$ degrees of freedom:

$$\nu = \frac{\left[ (\dfrac{s_1^2}{n_1}) + (\dfrac{s_2^2}{n_2}) \right]^2}{\left(\dfrac{s_1^2}{n_1}\right)^2 /(n_1 - 1) + \left(\dfrac{s_2^2}{n_2}\right)^2 /(n_2 - 1)}$$

# Problem:Test Statistic: $\sigma_1^2$ and $\sigma_2^2$ Unknown, Unequal

- Arsenic concentration in public drinking water supplies is a potential health risk.

- An article in the Arizona Republic (Sunday, May 27, 2001) reported drinking water arsenic concentrations in parts per billion (ppb) for 10 metropolitan Phoenix communities and 10 communities in rural Arizona.

- The data as shown:

| Metro Phoenix | | Rural Arizona | |
|---|---|---|---|
| Phoenix, | 3 | Rimrock, | 48 |
| Chandler, | 7 | Goodyear, | 44 |
| Gilbert, | 25 | New River, | 40 |
| Glendale, | 10 | Apachie Junction, | 38 |
| Mesa, | 15 | Buckeye, | 33 |
| Paradise Valley, | 6 | Nogales, | 21 |
| Peoria, | 12 | Black Canyon City, | 20 |
| Scottsdale, | 25 | Sedona, | 12 |
| Tempe, | 15 | Payson, | 1 |
| Sun City, | 7 | Casa Grande, | 18 |

$$\bar{x}_1 = 12.5 \qquad \bar{x}_1 = 27.5$$
$$s_1 = 7.63 \qquad s_2 = 15.3$$

# Problem:Test Statistic: $\sigma_1^2$ and $\sigma_2^2$ Unknown, Unequal

- We wish to determine it there is any difference in mean arsenic concentrations between metropolitan Phoenix communities and communities in rural Arizona.

# Problem:Test Statistic: $\sigma_1^2$ and $\sigma_2^2$ Unknown, Unequal

1. The parameters of interest are the mean arsenic concentrations for the two geographic regions, say, $\mu_1$ and $\mu_2$, and we are interested in determining whether $\mu_1 - \mu_2 = 0$.

2. $H_0: \mu_1 - \mu_2 = 0$, or $H_0: \mu_1 = \mu_2$

3. $H_1: \mu_1 \neq \mu_2$

4. $\alpha = 0.05$ (say)

5. The test statistic is

$$t_0^* = \frac{\bar{x}_1 - \bar{x}_2 - 0}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$$

# Problem:Test Statistic: $\sigma_1^2$ and $\sigma_2^2$ Unknown, Unequal

6. The degrees of freedom
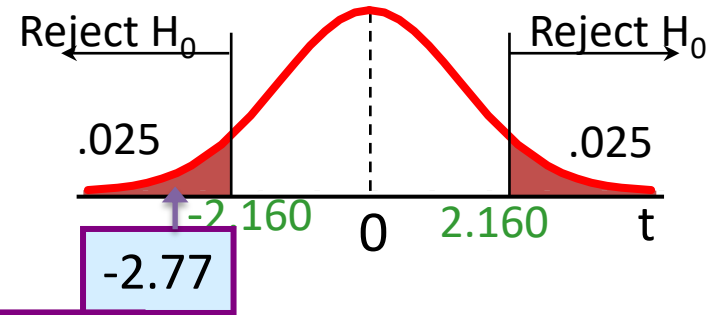
$$v = \frac{\left[(\frac{s_1^2}{n_1}) + (\frac{s_2^2}{n_2})\right]^2}{\left(\frac{s_1^2}{n_1}\right)^2 /(n_1 - 1) + \left(\frac{s_2^2}{n_2}\right)^2 /(n_2 - 1)} = \frac{\left[(\frac{7.63^2}{10}) + (\frac{15.3^2}{10})\right]^2}{\left(\frac{7.63^2}{10}\right)^2 /(10 - 1) + \left(\frac{15.3^2}{10}\right)^2 /(10 - 1)} = 13.2 = 13$$

Therefore, using $\alpha = 0.05$, we would reject $H_0$: $\mu_1 = \mu_2$ if $t_0^* > t_{0.025,13} = 2.160$ or if $t_0^* < -t_{0.025,13} = -2.160$

# Problem:Test Statistic: $\sigma_1^2$ and $\sigma_2^2$ Unknown, Unequal

7. Computations:

$$t = \frac{(12.5 - 27.5) - 0}{\sqrt{\left(\dfrac{7.63^2}{10} + \dfrac{15.3^2}{10}\right)}} = -2.77$$

-2.77



Reject H$_0$

Reject H$_0$

.025

.025

-2.160    0    2.160    t

-2.77

**Decision:**

Reject H$_0$ at $\alpha$ = 0.05

**Conclusion:**
There is evidence of a difference in means.

# Problem:Test Statistic: $\sigma_1^2$ and $\sigma_2^2$ Unknown, Unequal

8. Conclusions: Because $t_0^* = -2.77 < t_{0.025,13} = -2.160$,

- Reject the null hypothesis.
- There is evidence to conclude that mean arsenic concentration in the drinking water in rural Arizona is different from the mean arsenic concentration in metropolitan Phoenix drinking water.

# Problem:Test Statistic: $\sigma_1^2$ and $\sigma_2^2$ Unknown, Unequal

```
In [17]: stats.t.ppf(0.025,13) #critical  t value

Out[17]: -2.160368656461013
```

```
In [18]: metro = [3,7,25,10,15,6,12,25,15,7]
         rural = [48,44,40,38,33,21,20,12,1,18]
```

```
In [20]: stats.ttest_ind(metro,rural, equal_var = False)

Out[20]: Ttest_indResult(statistic=-2.7669395785560558, pvalue=0.015827284816100885)
```

# Dependent Samples

Tests Means of 2 Related Populations

– Paired or matched samples

– Repeated measures (before/after)

– Use difference between paired values:

$$d_i = x_i - y_i$$

- Assumptions:

  – Both Populations Are Normally Distributed

# Test Statistic: Dependent Samples

The test statistic for the mean difference is a t value, with n – 1 degrees of freedom:

$$t = \frac{\bar{d} - D_o}{\dfrac{s_d}{\sqrt{n}}}$$

$$\bar{d} = \frac{\sum d_i}{n} = \bar{x} - \bar{y}$$

$D_0$ = hypothesized mean difference

$s_d$ = sample standard dev. of differences

$n$ = the sample size (number of pairs)

# Decision Rules: Dependent Samples

Lower-tail test:

$H_0: \mu_1 - \mu_2 \geq 0$

$H_1: \mu_1 - \mu_2 < 0$
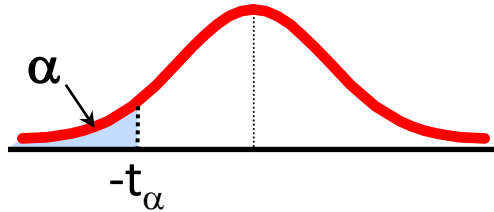
Upper-tail test:

$H_0: \mu_1 - \mu_2 \leq 0$
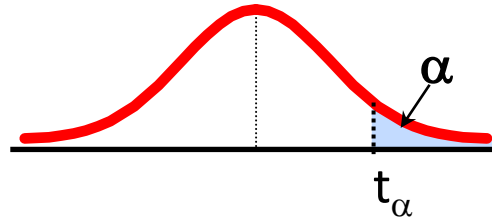
$H_1: \mu_1 - \mu_2 > 0$

Two-tail test:

$H_0: \mu_1 - \mu_2 = 0$

$H_1: \mu_1 - \mu_2 \neq 0$

# Decision Rules: Dependent Samples
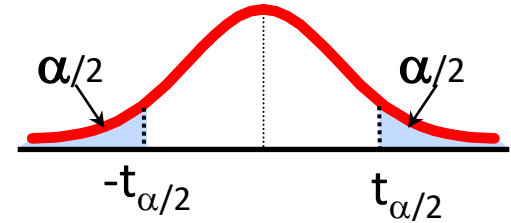


$\alpha$

$-t_\alpha$

Reject $H_0$ if $t < -t_{n-1, \alpha}$

$\alpha$

$t_\alpha$

Reject $H_0$ if $t > t_{n-1, \alpha}$

$\alpha/2$    $\alpha/2$

$-t_{\alpha/2}$    $t_{\alpha/2}$

Reject $H_0$ if $t < -t_{n-1, \alpha/2}$
or $t > t_{n-1, \alpha/2}$

Where $\quad t = \dfrac{\bar{d} - D_o}{\dfrac{s_d}{\sqrt{n}}} \quad$ has n - 1 d.f.

# Dependent Samples: Example

- An article in the Journal of Strain Analysis (1983, Vol. 18, No. 2) compares several methods for predicting the shear strength for steel plate girders.

- Data for two of these methods, the Karlsruhe and Lehigh procedures, when applied to nine specific girders, are shown in Table .

- We wish to determine whether there is any difference (on the average) between the two methods.

## Table : Strength Predictions for Nine Steel Plate Girders
### (Predicted Load/Observed Load)

| Girder | Karlsruhe Method | Lehigh Method | Difference dj |
|--------|------------------|---------------|---------------|
| S11 | 1.186 | 1.061 | 0.119 |
| S21 | 1.151 | 0.992 | 0.159 |
| S31 | 1.322 | 1.063 | 0.259 |
| S41 | 1.339 | 1.062 | 0.277 |
| S51 | 1.200 | 1.065 | 0.138 |
| S21 | 1.402 | 1.178 | 0.224 |
| S22 | 1.365 | 1.037 | 0.328 |
| S23 | 1.537 | 1.086 | 0.451 |
| S24 | 1.559 | 1.052 | 0.507 |

# Inferences About the Difference Between Two Population Means:  Matched Samples

1. The parameter of interest is the difference in mean shear strength between the two methods, say, $\mu_D = \mu_1 - \mu_2 = 0$.

2. $H_0$: $\mu_D = 0$

3. $H_1$: $\mu_D \neq 0$

4. $\alpha = 0.05$

5. The test statistic is

$$t_0 = \frac{\bar{d}}{s_D/\sqrt{n}}$$

# Inferences About the Difference Between Two Population Means: Matched Samples

6. Reject $H_0$ if $t_0 > t_{0.025,8} = 2.306$ or if $t_0 < -t_{0.025,8} = -2.306$.

7. Computations: The sample average and standard deviation of the differences $d_j$ are $\overline{d} = 0.2736$ and $s_D = 0.1356$, so the test statistic is

$$t_0 = \frac{\overline{d}}{s_D/\sqrt{n}} = \frac{0.2736}{0.1356/\sqrt{9}} = 6.05$$

**8. Conclusions:** Since $t_0 = 6.05 > 2.306$,

we conclude that the strength prediction methods yield different results.

The $P$-value for $t_0 = 6.05$ is $P = 0.0002$.

```
In [37]:  KARL= [1.186,1.151,1.322,1.339,1.200,1.402,1.365,1.537,1.559]
          LEH= [1.061,0.992,1.063,1.062,1.065,1.178,1.037,1.086,1.052]
```

```
In [38]:  stats.ttest_rel(KARL,LEH)
```

```
Out[38]:  Ttest_relResult(statistic=6.0819394375848255, pvalue=0.00029529546278604066)
```

# Inferences About the Difference Between Two Population Proportions

- Inferences About the Difference Between Two Population Proportion

# Inferences About the Difference Between Two Population Proportions

- Interval Estimation of $p_1 - p_2$

- Hypothesis Tests About $p_1 - p_2$

# Sampling Distribution of $\bar{p}_1 - \bar{p}_2$

- Expected Value

$$E(\bar{p}_1 - \bar{p}_2) = p_1 - p_2$$

- Standard Deviation (Standard Error)

$$\sigma_{\bar{p}_1 - \bar{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

where: n1 = size of sample taken from population 1

n2 = size of sample taken from population 2

# Sampling Distribution of $\bar{p}_1 - \bar{p}_2$

- If the sample sizes are large, the sampling distribution of $\bar{p}_1 - \bar{p}_2$ can be approximated by a normal probability distribution.

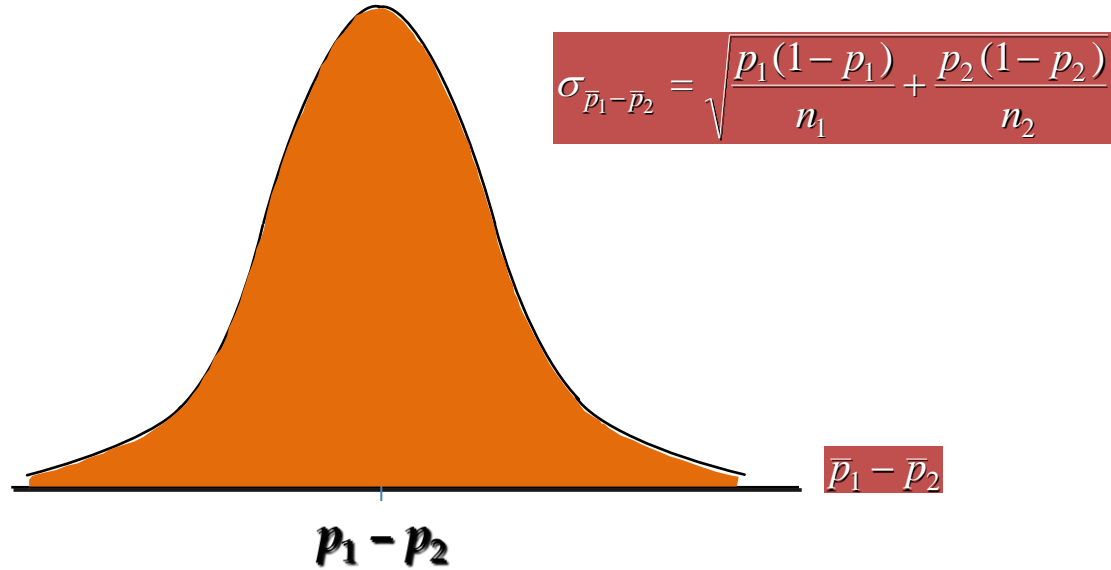- The sample sizes are sufficiently large if all of these conditions are met:

$$n_1 p_1 \geq 5 \qquad n_1(1 - p_1) \geq 5$$

$$n_2 p_2 \geq 5 \qquad n_2(1 - p_2) \geq 5$$

# Sampling Distribution of $\bar{p}_1 - \bar{p}_2$



$$\sigma_{\bar{p}_1 - \bar{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

$\bar{p}_1 - \bar{p}_2$

$p_1 - p_2$

# Interval Estimation of $p_1 - p_2$

- Interval Estimation

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

# Point Estimator of the Difference Between Two Population Proportions

- p1 = proportion of the population of households "aware" of the product after the new campaign

- p2 = proportion of the population of households "aware" of the product before the new campaign

- $\bar{p}_1$ = sample proportion of households "aware" of the product after the new campaign

- $\bar{p}_2$ = sample proportion of households "aware" of the product before the new campaign

$$\bar{p}_1 - \bar{p}_2 = \frac{120}{250} - \frac{60}{150} = .48 - .40 = .08$$

# Hypothesis Tests about $p_1$ - $p_2$

- Hypothesis

We focus on tests involving no difference between the two population proportions (i.e. $p_1 = p_2$)

| $H_0: p_1 - p_2 \geq 0$ | $H_0: p_1 - p_2 \leq 0$ | $H_0: p_1 - p_2 = 0$ |
|---|---|---|
| $H_a: p_1 - p_2 < 0$ | $H_a: p_1 - p_2 > 0$ | $H_a: p_1 - p_2 \neq 0$ |
| **Left-tailed** | **Right-tailed** | **Two-tailed** |

# Hypothesis Tests about $p_1 - p_2$

- Standard Error of $\bar{p}_1 - \bar{p}_2$ when $p_1 = p_2 = p$

$$\sigma_{\bar{p}_1 - \bar{p}_2} = \sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

- Pooled Estimator of p when $p_1 = p_2 = p$

$$\bar{p} = \frac{n_1\bar{p}_1 + n_2\bar{p}_2}{n_1 + n_2}$$

# Hypothesis Tests about $p_1 - p_2$

- Test Statistic

$$z = \frac{(\bar{p}_1 - \bar{p}_2)}{\sqrt{\bar{p}(1-\bar{p})\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$$

# Problem: Hypothesis Tests about $p_1$ - $p_2$

- Extracts of St. John's Wort are widely used to treat depression.

- An article in the April 18, 2001 issue of the *Journal of the American Medical Association* ("Effectiveness of St. John's Worton Major Depression: A Randomized Controlled Trial") compared the efficacy of a standard extract of St. John's Wort with a placebo in 200 outpatients diagnosed with major depression.

- Patients were randomly assigned to two groups; one group received the St. John's Wort, and the other received the placebo.

- After eight weeks, 19 of the placebo-treated patients showed improvement, whereas 27 of those treated with St. John's Wort improved.

- Is there any reason to believe that St. John's Wort is effective in treating major depression? Use  0.05.

# Problem: Hypothesis Tests about $p_1$ - $p_2$

1. The parameters of interest are $p_1$ and $p_2$, the proportion of patients who improve following treatment with St. John's Wort ($p_1$) or the placebo ($p_2$).

2. $H_0$: $p_1 = p_2$

3. $H_1$: $p_1 \neq p_2$

4. $\alpha = 0.05$

5. The test statistic is

$$z_0 = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1 - \hat{p})\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$$

where $\hat{p}_1 = 27/100 = 0.27, \hat{p}_2 = 19/100 = 0.19, n_1 = n_2 = 100,$ and

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{19 + 27}{100 + 100} = 0.23$$

6.  Reject $H_0$: $p_1 = p_2$ if $z_0 > z_{0.025} = 1.96$ or if $z_0 < -z_{0.025} = -1.96$.

7.  Computations: The value of the test statistic is

$$z_0 = \frac{0.27 - 0.19}{\sqrt{0.23(0.77)\left(\frac{1}{100} + \frac{1}{100}\right)}} = 1.35$$

**8.** Conclusions: Since $z_0$ 1.35 does not exceed $z$ 0.025, we cannot reject the null hypothesis.

The *P*-value is $P \cong 0.177$. There is insufficient evidence to support the claim that St. John's Wort is effective in treating major depression.

```python
In [29]: import math
         def two_samp_proportion(p1,p2,n1,n2):
             p_pool = ((p1*n2)+(p2*n1))/(n1+n2)
             x = (p_pool*(1- p_pool)*((1/n1)+(1/n2)))
             s = math.sqrt(x)
             z = (p1- p2)/s
             if (z < 0):
                 p_val = stats.norm.cdf(z)
             else:
                 p_val = 1 - stats.norm.cdf(z)
             return z, p_val*2
```

```python
In [30]: two_samp_proportion(0.27,0.19,100,100)
```

Out[30]: (1.3442056254198995, 0.17888190308175567)

```python
In [27]: stats.norm.cdf(1.3442056254198995)
```

Out[27]: 0.9105590484591222

# Thank You