

Recall : Hypothesis testing

$X_1, \dots, X_n$ iid sample $X \sim \text{Normal}(\mu, \sigma^2)$

Z-test

$$H_0: \mu = c$$

$$H_1: \mu > c$$

$$Z \sim \text{Normal}(0, 1)$$

$$\left(P \left(Z \geq \frac{\sqrt{n} (\bar{X} - c)}{\sigma} \right) \right) > \alpha$$

where α - specified earlier

then we would not reject H_0

$\bar{X}$ - given

fixed known

known

$$H_0: \mu = c$$

$$H_1: \mu \neq c$$

$$P \left(|Z| > \frac{\sqrt{n} |\bar{X} - c|}{\sigma} \right)$$

$> \alpha$

Then we do not reject H_0

t-test . Test for sample mean when variance is not known

$X_1, X_2, \dots, X_n$ be an i.i.d sample from $\text{Normal}(\mu, \sigma^2)$

& σ - unknown.

Fact :-

$$\bar{Y} = \frac{Y_1 + \dots + Y_n}{n}$$

$$Y_i \sim \text{Normal}(\mu, \sigma^2)$$

$$S_y^2 = \frac{1}{n-1} \sum_{i=1}^{n-1} (Y_i - \bar{Y})^2$$

$$\frac{\sqrt{n}(\bar{Y} - \mu)}{S_y} \sim t_{n-1}$$

(null) $H_0: \mu = c$

(Alternative) $H_1: \mu > c$

$X_1, \dots, X_n$ - known from data

$Y_1, \dots, Y_n$ mimic the sample and
are i.i.d X

$$\begin{aligned}
 P(\bar{Y} > \bar{X}) & \stackrel{\text{under null}}{=} P\left(\frac{\sqrt{n}(\bar{Y} - c)}{S} > \frac{\sqrt{n}(\bar{X} - c)}{S}\right) \\
 & = P\left(T > \frac{\sqrt{n}(\bar{X} - c)}{S}\right)
 \end{aligned}$$

$T \sim t_{n-1}$ use t -tables to decide on the test.

Fix $\alpha \in (0, 1)$ If $P\left(T > \frac{\sqrt{n}(\bar{X} - c)}{S}\right) < \alpha$

then we reject the null hypothesis H_0

Example 6

$$n = 16$$

$$\bar{X} = 10.2$$

$$S = 3$$

$$\text{Fix } \alpha = 0.05$$

$$H_0: \mu = 9.5$$

$$H_1: \mu > 9.5$$

Calculate:

$$T \sim t_{15}$$

$$P\left(T > \frac{\sqrt{n}(\bar{X} - c)}{S}\right) = P\left(T > \frac{4(10.2 - 9.5)}{3}\right)$$

$$= 0.183$$

As $0.183 > 0.05$, the null hypothesis is not rejected

approach

(or the relevant statistic)

↑

such that if $\bar{X}$ falls in this region then we reject the null hypothesis

$$H_0: \mu = c$$

Given α

$\rightarrow X_1, \dots, X_n$ i.i.d sample Normal

$$H_1: \mu > c$$

$Y_1, \dots, Y_n$ mimic sample $\sim \text{Normal}$

calculate

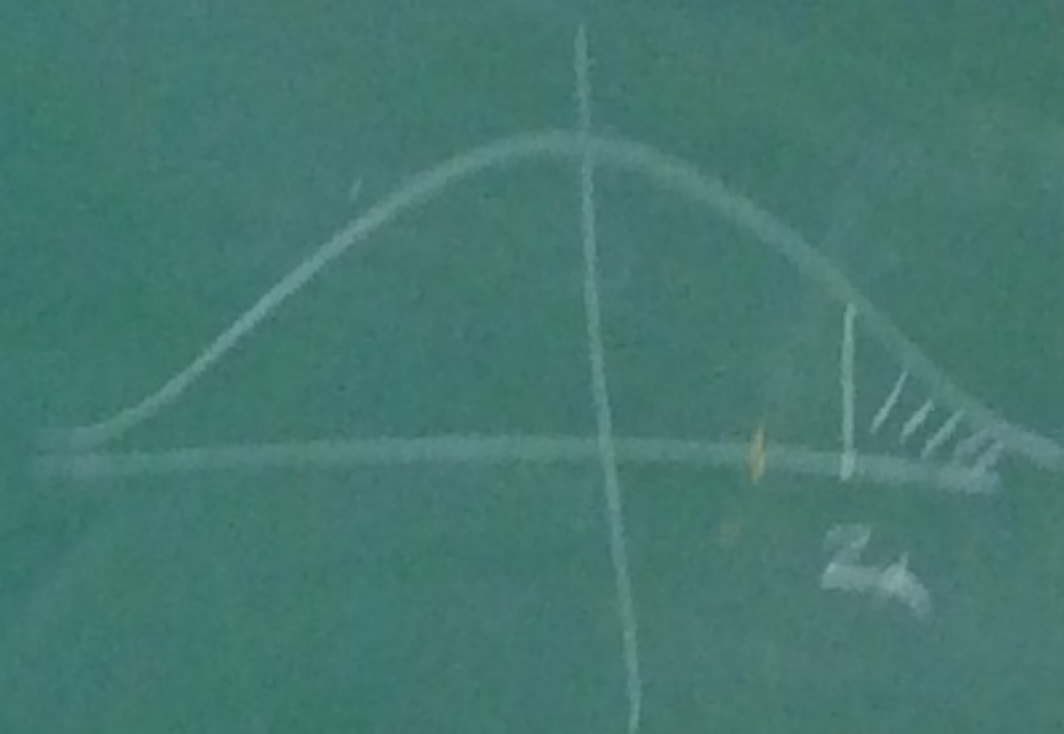
$$P(\bar{Y} > \bar{X}) = P\left(\frac{\sqrt{n}(\bar{Y} - c)}{\sigma} > \frac{\sqrt{n}(\bar{X} - c)}{\sigma}\right)$$

$$Z = \frac{\sqrt{n}(\bar{Y} - c)}{\sigma}$$

$$Z \sim \text{Normal}(0, 1)$$

Find z_α :

$$P(Z > z_\alpha) = \alpha$$



Check: $\frac{\sqrt{n}(\bar{X} - c)}{\sigma} = z_\alpha$

$$(\Rightarrow) \sqrt{n}(\bar{X} - c) = \sigma z_\alpha (\Rightarrow) \bar{X} - c = \frac{\sigma}{\sqrt{n}} z_\alpha (\Rightarrow) \bar{X} = \frac{\sigma}{\sqrt{n}} z_\alpha + c$$

If $\bar{X} = \frac{\sigma}{\sqrt{n}} z_\alpha + c$ then $P(\bar{Y} > \bar{X}) = \alpha$

$\Rightarrow$ we reject the null hypothesis H_0 if $\bar{X} > \frac{\sigma}{\sqrt{n}} z_\alpha + c$

So the critical region at level α is

$$\left(\sigma \frac{Z_{\alpha}}{\sqrt{n}} + c, \infty \right)$$

Example 4a $n=16, \bar{X}=10.2, \sigma=3, \alpha=0.05, c=9.5$

From normal table $Z_{\alpha} = 1.65$

$$\sigma \frac{Z_{\alpha}}{\sqrt{n}} + c = \frac{3}{4}(1.65) + 9.5 = 10.73$$

Do not reject H_0

$$\begin{aligned} \bar{X} &= 10.2 \\ &< 10.73 \\ &= \sigma \frac{Z_{\alpha}}{\sqrt{n}} + c \end{aligned}$$

χ^2 test for sample variance

$X_1, \dots, X_n$ i.i.d sample $X \sim \text{Normal}(\mu, \sigma^2)$

σ - unknown

$$H_0: \sigma = \sigma_0$$

$$H_1: \sigma > \sigma_0$$

Mimic $Y_1, \dots, Y_n$ i.i.d $\sim \text{Normal}(\mu, \sigma^2)$

$$S_Y^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2, \quad \bar{Y} = \frac{\sum_{i=1}^n Y_i}{n}$$

$$(a) \frac{(n-1)S_Y^2}{\sigma^2} \sim \chi_{n-1}^2 \quad (b) \bar{Y} \sim \text{Normal}(\mu, \frac{\sigma^2}{n})$$

(c) S_Y^2 and $\bar{Y}$ are independent

$$P(S_y \geq S_x)$$

$$= P\left(\frac{(n-1)S_y^2}{\sigma_0^2} \geq \frac{(n-1)S_x^2}{\sigma_0^2}\right)$$

$$\chi \sim \chi_{n-1}^2$$

$$= P\left(\chi \geq \frac{(n-1)S_x^2}{\sigma_0^2}\right)$$

Given: α we can suitably decide on H_0 , using χ_{n-1}^2 tables.

Example 5 (a) $X_1, X_2, \dots, X_{16}$ be iid sample from $\text{Normal}(\mu, \sigma^2)$

$$S_x = 3.5$$

$$\alpha = 0.05$$

$$H_0: \sigma = 3$$

$$H_1: \sigma > 3$$

check. $P\left(\chi > \frac{(n-1) S_x^2}{\sigma_0^2}\right)$

$$= P\left(\chi > \frac{15(3.5)^2}{9}\right) \approx 0.157$$

As $0.157 > 0.05$
we do not reject
the null hypothesis
 H_0 .

Example 5(b) $X_1, \dots, X_{16}$ i.i.d. sample from $\text{Normal}(\mu, \sigma^2)$

$$\bar{X} = 3.5$$

$$\alpha = 0.05$$

$$H_0: \sigma = 3$$

$$H_1: \sigma \neq 3$$

Note. unlike $Z \sim \text{Normal}(0,1)$ which is symmetric

χ^2_{n-1} is not

Take a critical region approach for this test

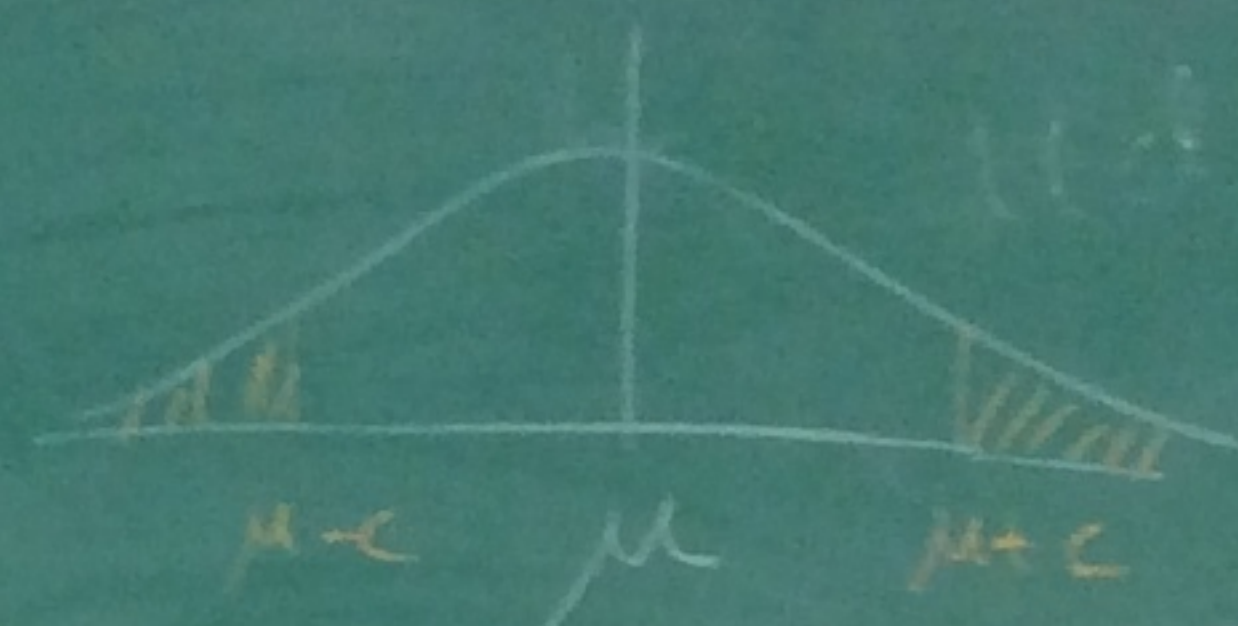
Imitate normal

(μ, σ^2)

$$Z \stackrel{d}{=} \text{Normal}(\mu, \sigma^2)$$

$$P(Z \geq \mu + c)$$

$$= P(Z \leq \mu - c)$$



$$H_0: \mu = c$$

$$H_1: \mu \neq c$$

$$\alpha = P(|Z - \mu| > c)$$

$$= 2 P(Z > \mu + c)$$

As χ^2_{n-1} is not symmetric we cannot take two approach
for a two-tailed test



From χ^2_{15} table: $P(\chi^2_{15} < 6.26) = 0.025$ & $P(\chi^2_{15} > 27.49) = 0.025$

So if $\frac{(n-1)S_x^2}{\sigma_0^2} = \frac{15 S_x^2}{9}$ lies outside

the interval $(6.26, 27.49)$

then we reject H_0 .

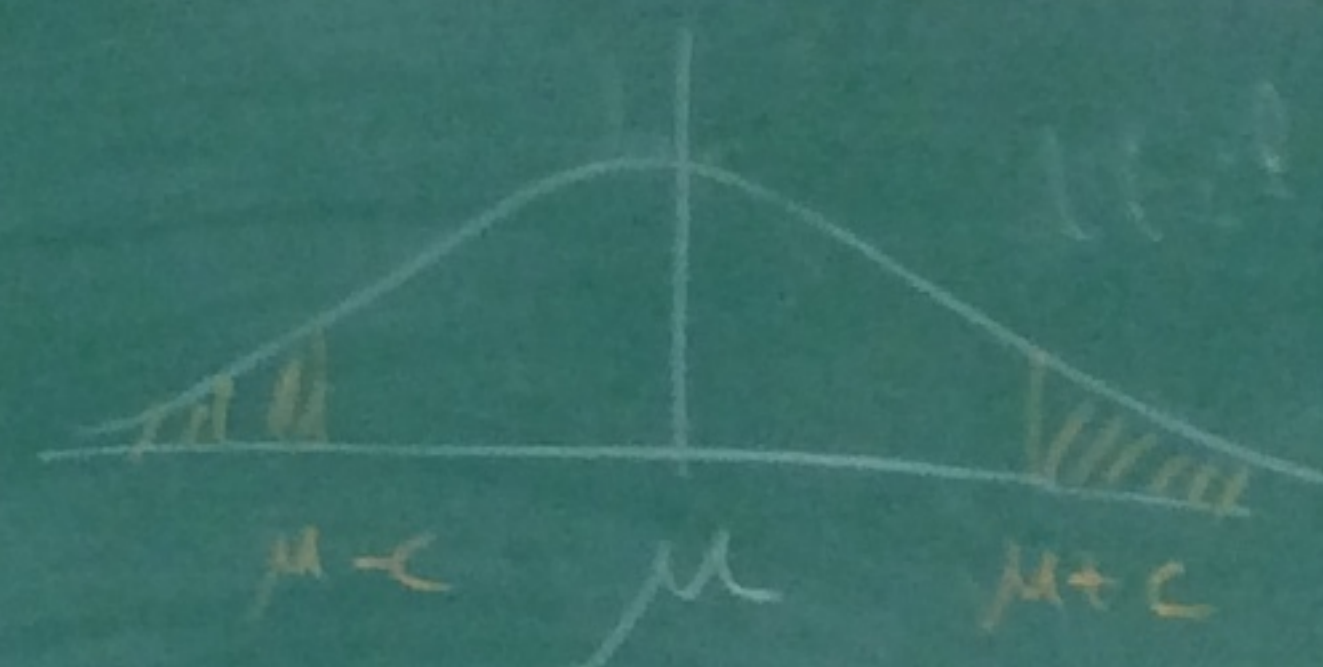
In our case $S_x = 3.5$, $\frac{15 S_x^2}{9} \approx 20.42$

, we do not reject H_0 given the data at level 0.05.

$$Z \stackrel{d}{=} \text{Normal}(\mu, \sigma^2)$$

$$P(Z \geq \mu + c)$$

$$= P(Z \leq \mu - c)$$



$$H_0: \mu = c$$

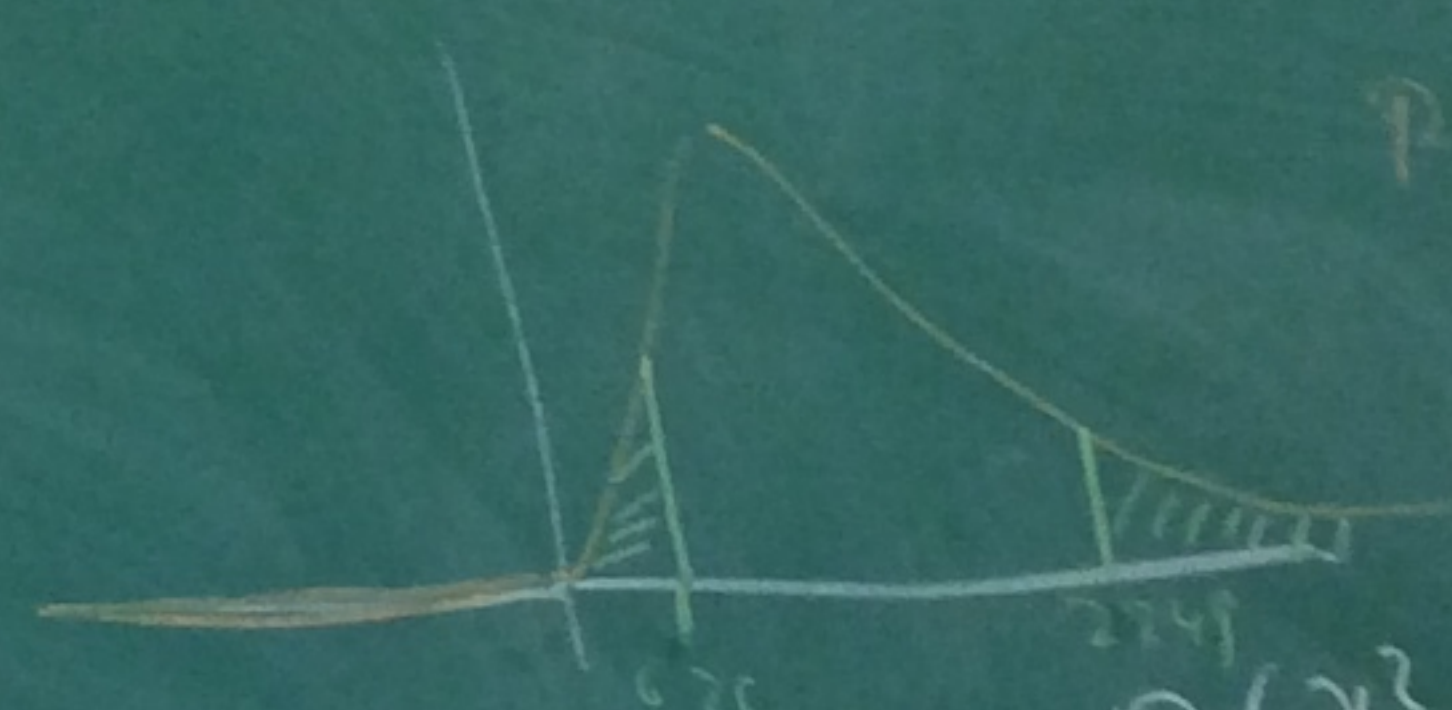
$$H_1: \mu \neq c$$

$$\alpha = P(|Z - \mu| > c)$$

$$= 2 P(Z > \mu + c)$$

As χ^2_{n-1} is not symmetric
for a two-tailed test

we cannot take this approach



pdf of χ^2_n

From χ^2_{15} table: $P(\chi^2_{15} < 6.26) = 0.025$

$$\in P(\chi^2_{15} > 27.49) = 0.025$$

Two sample z-test

$X_1, \dots, X_{n_1}$ — iid sample
 $Y_1, \dots, Y_{n_2}$ — iid sample

$$X \sim \text{Normal}(\mu_1, \sigma_1^2)$$

$$Y \sim \text{Normal}(\mu_2, \sigma_2^2)$$

Assume σ_1 - known & σ_2 - known.

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

Motivation

we may need to compare two
Populations, e.g. Scores of MU score minimum
Students in a test vs Score of Bangalore university
on the same test

Let n_1, n_2 mins $x_1, \dots, x_{n_1}$
 $w_1, \dots, w_{n_2}$ $y_1, \dots, y_{n_2}$

$$\bar{V} \sim \text{Normal}(\mu_1, \frac{\sigma_1^2}{n_1})$$

$$\bar{W} \sim \text{Normal}(\mu_2, \frac{\sigma_2^2}{n_2})$$

$$\bar{V} - \bar{W} \sim \text{Normal}(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$$

$$P(|\bar{V} - \bar{W}| > |\bar{X} - \bar{Y}|)$$

under null

$$= P(|Z| > \frac{|\bar{X} - \bar{Y}|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}})$$

$$= 2 P(Z > \frac{|\bar{X} - \bar{Y}|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}})$$

where $Z \sim \text{Normal}(0, 1)$

Given α . We check if $P(|Z| > \frac{|\bar{X} - \bar{Y}|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}) > \alpha$

$X_1, \dots, X_{n_1}$

$Y_1, \dots, Y_{n_2}$

σ_1^2, σ_2^2

then we do not reject H_0 .

As $H_1: \mu_1 \neq \mu_2$, this implies we accept H_0 .

Example 6. A biologist wants to know if the average weight of squirrels in Bandipur forest is the same as that of squirrels in Banaghat.

Historical data suggests that $\sigma = 10$

30 samples are taken from each forest

Found: Average height in Bandipur is 122.4

Average height in Bannerghata is 127.6

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

(Assume normality)

$$n_1 = n_2 = 30$$

$$\sigma_1^2 = \sigma_2^2 = 100$$

$$\bar{X} = 122.4$$

$$\bar{Y} = 127.6$$

$$\text{Conclude } \frac{|\bar{X} - \bar{Y}|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{5.2}{10} \frac{\sqrt{30}}{\sqrt{2}}$$

$$P(|Z| > (0.52)\sqrt{15}) = 0.044$$

If $\alpha = 0.05$ then as $0.044 < 0.05$ we reject the null hypothesis & conclude that mean weight of squirrels are different in each forest.

F-test to compare two sample variances

$X_1, \dots, X_{n_1}$ are iid sample $\sim \text{Normal}(\mu_1, \sigma_1^2)$

$Y_1, \dots, Y_{n_2}$ are iid sample $\sim \text{Normal}(\mu_2, \sigma_2^2)$

$$H_0: \sigma_1 = \sigma_2$$

$$H_1: \sigma_1 \neq \sigma_2$$

Fact:

$V_1, \dots, V_{n_1} \sim \text{iid Normal}(\mu_1, \sigma_1^2)$

$W_1, \dots, W_{n_2} \sim \text{iid Normal}(\mu_2, \sigma_2^2)$

$$\frac{(n_1-1)S_v^2}{\sigma_1^2} \sim \chi_{n_1-1}^2 \quad \& \quad \frac{(n_2-1)S_w^2}{\sigma_2^2} \sim \chi_{n_2-1}^2$$

$$\Rightarrow \frac{\sigma_2^2 S_v^2}{\sigma_1^2 S_w^2} \sim F(n_1-1, n_2-1)$$

F - not symmetric

$$H_0: \sigma_1 = \sigma_2, \quad H_1: \sigma_1 \neq \sigma_2$$

Example 7

$$n_1 = 30, \quad n_2 = 25,$$

$$s_x^2 = 11.4, \quad s_y^2 = 5.1$$

(under null)

$$R = \frac{s_v^2}{s_w^2} \sim F(29, 24)$$

F-tables

$$P(R < 0.464) = 0.025$$

$$P(R > 2.22) = 0.025$$

Reject H_0 if $\frac{s_x^2}{s_y^2}$ lies outside $(0.464, 2.22)$

Here

$$\frac{s_x^2}{s_y^2} = \frac{11.4}{5.1} = 2.24$$

$\Rightarrow H_0$ is rejected