UNIT - 3

Sampling-Distributions, Parameter-Estimations, Hypothesis-Testing, Two-population, Linear Regression, Tests, Regression and Correlation, Block Chain

R. K. Sharma, CSED, TIET Patiala

Parameter-Estimations, Hypothesis-Testing

Parameter Estimation using Method of Moments

- **Estimating the Moments about origin and about population mean (μ)**

The k-th **population moment** is defined as

$$\mu_k = \mathbf{E}(X^k).$$

The k-th **sample moment**

$$m_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

estimates μ_k from a sample $(X_1, \ldots, X_n)$.

The first sample moment is the sample mean $\bar{X}$.

For $k \geq 2$, the k-th **population central moment** is defined as

$$\mu'_k = \mathbf{E}(X - \mu_1)^k.$$

The k-th **sample central moment**

$$m'_k = \frac{1}{n}\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^k$$

estimates μ_k from a sample $(X_1, \ldots, X_n)$.

Example

Suppose that we take a random sample from a rectangular distribution, i.e., a uniform distribution over [a, b]. The 2 parameters are a and b here. We have to estimate these 2 parameters given a sample of size *n*.

Example

Suppose that we take a random sample from a rectangular distribution followed by *X*, i.e., a uniform distribution over [a, b]. The 2 parameters are a and b here. We have to estimate these 2 parameters given a sample of size *n*.

- Let $(X_1, X_2, \dots, X_n)$ be a random sample of size *n* taken from the variable *X*.
- E(*X*) = (a + b)/2 = $(x_1, + x_2 + \cdots + x_n)/n$ = m (say);
  - This gives us a + b = 2m (i)
- $E(X^2)$ = (a^2 + b^2 + ab)/3 = (x_1 ^2+ x_2 ^2 +… + x_n ^2)/n = p (say);
  - This gives us (a^2 + b^2 + ab) = 3p (ii)
- Solving (i) and (ii) for a and b gives the estimates using method of moments.
- a = m - sqrt(3*p - 3*m*m)
- b = 2*m – a = m + sqrt(3*p - 3*m*m)

```
#DEMO
n<-1000
a<-10
b<-15
x<-runif(n, a, b)
m<-sum(x)/n
p<-sum(x^2)/n
aestimated <- m-
sqrt(3*p - 3*m*m)
bestimated <- 2*m - a
print(abs(a-aestimated))
print(abs(b-bestimated))
```

Maximum Likelihood Estimate

Maximum likelihood estimator is the parameter value that maximizes the likelihood of the observed sample. For a discrete distribution, we maximize the joint pmf of data $P(X_1, \ldots, X_n)$. For a continuous distribution, we maximize the joint density $f(X_1, \ldots, X_n)$.

Estimating the parameter of Exponential distribution

- Let us take a sample of size *n* from an exponential population having the parameter λ. So, *X* is exponentially distributed with parameter λ.
- A sample of size *n* is taken: $(X_1, X_2, \ldots, X_n)$.
- Exponential density is: $f(x) = \lambda . exp(-\lambda x), \lambda > 0, \ x \geq 0$
- What is the density function for X_1? $\lambda . exp(-\lambda x_1)$; So the density of X_i is $\lambda . exp(-\lambda xi)$, *i* =1, 2, …, *n*.
- Joint density of $(X_1, X_2, \ldots, X_n)$, L(x1, x2, …., xn) = $\prod_{i=1}^{n} \lambda . exp(-\lambda x_i)$
- ln(L(x1, x2, …., xn)) = ln(λ^n * exp($-\lambda x1 + -\lambda x2 + \cdots + -\lambda xn$))
- = n*ln(λ) + ($-\lambda x1 + -\lambda x2 + \cdots + -\lambda xn$) = n*ln($\lambda$) – $\lambda(x1 + x2 + .. + xn)$
- Differentiate this equation with respect to λ, and equate to zero, this will give us the estimator that is obtained using this method of maximum likelihood estimation.
  - $\lambda = 1/sample(mean)$

Maximum Likelihood Estimate: Normal Distribution

Let $(X_1, X_2, \dots, X_n)$ be a random sample of size n taken from a Normal Population with parameters: mean θ_1 and variance θ_2. Find the Maximum Likelihood Estimates of these two parameters.

We have a random sample $x_1, x_2, \dots, x_n$, where $X_i \sim N(\theta_1, \theta_2)$.

θ_1 = Mean of X_i, and

θ_2 = Variance of X_i.

So, $f_{X_i}(x_i; \theta_1, \theta_2) = \frac{1}{\sqrt{2\pi\theta_2}} \cdot e^{-(x_i-\theta_1)^2/2\theta_2}$

Likelihood function is:

$$L(x_1, x_2, \dots, x_n; \theta_1, \theta_2) = \frac{1}{(2\pi)^{n/2} \cdot \theta_2^{n/2}} \exp\left(-\frac{1}{2\theta_2} \cdot \sum_{i=1}^{n} (x_i - \theta_1)^2\right)$$

So, $\ln L(x_1, x_2, \dots, x_n; \theta_1, \theta_2)$

$$= -\frac{n}{2} \cdot \ln(2\pi) - \frac{n}{2} \cdot \ln\theta_2 - \frac{1}{2\theta_2} \sum_{i=1}^{n} (x_i - \theta_1)^2$$

Now $\frac{\partial L}{\partial \theta_1} = 0$ and $\frac{\partial L}{\partial \theta_2} = 0$ gives

$$-\frac{1}{2\theta_2} \cdot \sum_{i=1}^{n} 2(x_i - \theta_1)(-1) = 0 \text{ or } \frac{1}{\theta_2} \sum_{i=1}^{n} (x_i - \theta_1) = 0 \quad \text{—(1)}$$

and

$$-\frac{n}{2\theta_2} + \frac{1}{2\theta_2^2} \cdot \sum_{i=1}^{n} (x_i - \theta_1)^2 = 0 \quad \text{—(2)}$$

(1) gives, $\hat{\theta}_1 = \frac{1}{n} \sum_{i=1}^{n} x_i$, and

(2) gives, $\hat{\theta}_2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \theta_1)^2$.

$\hat{\theta}_1$ and $\hat{\theta}_2$ are the MLEs of θ_1 and θ_2.

Hypothesis Testing

HYPOTHESIS TESTING

A statistical hypothesis is an assumption about a population parameter. This assumption about the parameter may or may not be true.

- *Null Hypothesis*
  - The null hypothesis, denoted by H_0, is the hypothesis that sample observations result simply from options. For example, H_0: $\mu = 5$.
- *Alternative Hypothesis*
  - The alternative hypothesis, denoted by H_1 or H_a, is the hypothesis that sample observations are prone to some non-random reasons. There are three types of alternative hypotheses. These are: (i) H_1: $\mu \neq 5$, (ii) H_1: $\mu > 5$ and (iii) H_1: $\mu < 5$. We have to consider one of these three hypotheses in a problem. This consideration is based on the facts stated in the problem.

HYPOTHESIS TESTING

- The two hypotheses are complementary to each other. We accept (or reject) the null hypothesis; this is equivalent to rejecting (or accepting) the alternative hypothesis. In this process, we make two types of decision errors, namely, Type I error and Type II error.
- *Type I Error*
  - We say that we have committed a Type I error when we reject a null hypothesis when it is true. The probability of committing a Type I error is called the significance level denoted by α.
- *Type II Error*
  - We say that we have committed a Type II error when we accept a null hypothesis when it is false. The probability of committing a Type II error is denoted by β. The probability of not committing a Type II error is called the Power of the test (= 1 - β).

	Result of the test	
	Reject H_0	Accept H_0
H_0 is true	*Type I error*	correct
H_0 is false	correct	*Type II error*

HYPOTHESIS TESTING

Two-Tailed Test

- A test of hypothesis, in which the region of rejection is on both sides of the distribution used for test statistic, is called a two-tailed test. For example, if we hypothesize that population mean is 12, i.e., null hypothesis is $\mu = 12$ and we consider that the alternative hypothesis is; mean is less than 12 or mean is greater than 12, then we have to employ a two-tailed test. In such a situation, alternative hypothesis shall be $\mu \neq 12$.

One-Tailed Tests

- A test of hypothesis, in which the region of rejection is only on one side of the distribution used for test statistic, is called a one-tailed test. For example, if we hypothesize that population mean is 12, i.e., null hypothesis is $\mu = 12$ and we consider that the alternative hypothesis is; mean is less than 12, and then we have to employ a one-tailed test. In this situation, alternative hypothesis shall be $\mu < 12$. One can see that in some other situation, we may have to consider that $\mu > 12$.

HYPOTHESIS TESTING

In every test of hypothesis, we have to follow 4 basic steps. These are:

- State the null and alternative hypotheses
- Decide the sample statistic
- Analyze sample data, and calculate the value of Sample Statistic
- accept or reject the null hypothesis based on the significance level, and interpret the same.

Z-tests

Null hypothesis H_0	Parameter, estimator $\theta, \hat\theta$	If H_0 is true: $\mathbf{E}(\hat\theta)$	If H_0 is true: $\mathrm{Var}(\hat\theta)$	Test statistic $Z = \frac{\hat\theta - \theta_0}{\sqrt{\mathrm{Var}(\hat\theta)}}$
One-sample Z-tests for means and proportions, based on a sample of size n				
$\mu = \mu_0$	$\mu, \bar X$	μ_0	$\frac{\sigma^2}{n}$	$\frac{\bar X - \mu_0}{\sigma/\sqrt{n}}$
$p = p_0$	$p, \hat p$	p_0	$\frac{p_0(1-p_0)}{n}$	$\frac{\hat p - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$
Two-sample Z-tests comparing means and proportions of two populations, based on independent samples of size n and m				
$\mu_X - \mu_Y = D$	$\mu_X - \mu_Y$, $\bar X - \bar Y$	D	$\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}$	$\frac{\bar X - \bar Y - D}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}}$
$p_1 - p_2 = D$	$p_1 - p_2$, $\hat p_1 - \hat p_2$	D	$\frac{p_1(1-p_1)}{n} + \frac{p_2(1-p_2)}{m}$	$\frac{\hat p_1 - \hat p_2 - D}{\sqrt{\frac{\hat p_1(1-\hat p_1)}{n} + \frac{\hat p_2(1-\hat p_2)}{m}}}$
$p_1 = p_2$	$p_1 - p_2$, $\hat p_1 - \hat p_2$	0	$p(1-p)\left(\frac{1}{n} + \frac{1}{m}\right)$, where $p = p_1 = p_2$	$\frac{\hat p_1 - \hat p_2}{\sqrt{\hat p(1-\hat p)\left(\frac{1}{n}+\frac{1}{m}\right)}}$ where $\hat p = \frac{n\hat p_1 + m\hat p_2}{n+m}$

Bernoulli Trial / Experiment

Y	P(Y = y_i)	
1	p	E(Y) = p
0	1 - p	V(Y) = p(1 - p)

Binomial Variable (Sum of n independent variables following the distribution of Y)

X	P(X = x_i)	
0	${}^n c_0 p^0 q^n$	E(X) = np
1	${}^n c_1 p^1 q^{n-1}$	V(X) = np(1 - p)
…		
n	${}^n c_n p^n q^0$	

Distribution of proportion ($Z = Y / n$)

$E(Z) = E(X) / n = p$

$V(Z) = V(X / n) = np(1-p)/(n^2) = p(1-p) / n$

Z-tests

Null hypothesis H_0	Parameter, estimator $\theta,\ \hat\theta$	If H_0 is true: $\mathbf{E}(\hat\theta)$	$\mathrm{Var}(\hat\theta)$	Test statistic $Z=\dfrac{\hat\theta-\theta_0}{\sqrt{\mathrm{Var}(\hat\theta)}}$
One-sample Z-tests for means and proportions, based on a sample of size n				
$\mu=\mu_0$	$\mu,\ \bar X$	μ_0	$\dfrac{\sigma^2}{n}$	$\dfrac{\bar X-\mu_0}{\sigma/\sqrt{n}}$
$p=p_0$	$p,\ \hat p$	p_0	$\dfrac{p_0(1-p_0)}{n}$	$\dfrac{\hat p-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$
Two-sample Z-tests comparing means and proportions of two populations, based on independent samples of size n and m				
$\mu_X-\mu_Y=D$	$\mu_X-\mu_Y,$ $\bar X-\bar Y$	D	$\dfrac{\sigma_X^2}{n}+\dfrac{\sigma_Y^2}{m}$	$\dfrac{\bar X-\bar Y-D}{\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}}$
$p_1-p_2=D$	$p_1-p_2,$ $\hat p_1-\hat p_2$	D	$\dfrac{p_1(1-p_1)}{n}+\dfrac{p_2(1-p_2)}{m}$	$\dfrac{\hat p_1-\hat p_2-D}{\sqrt{\frac{\hat p_1(1-\hat p_1)}{n}+\frac{\hat p_2(1-\hat p_2)}{m}}}$
$p_1=p_2$	$p_1-p_2,$ $\hat p_1-\hat p_2$	0	$p(1-p)\left(\dfrac{1}{n}+\dfrac{1}{m}\right),$ where $p=p_1=p_2$	$\dfrac{\hat p_1-\hat p_2}{\sqrt{\hat p(1-\hat p)\left(\frac{1}{n}+\frac{1}{m}\right)}}$ where $\hat p=\dfrac{n\hat p_1+m\hat p_2}{n+m}$

Suppose that X has mean mu.X and variance v.X.

Y has mean mu.Y and variance v.Y.

Let us take a sample of size n from X and a sample of size m from Y.

The mean of sample from X is Xbar, and mean of sample from Y is ybar.

E(Xbar – Ybar)
= E(Xbar) – E(Ybar) = mu.X – mu.Y

v(Xbar – Ybar)
= v.Xbar + v.Ybar
= v.X + v.Y
= v.X/n + v.y/m

[(Xbar – Ybar) – (mu.X – mu.Y)] /
[sqrt(v.X/n + v.y/m]
shall have standard normal distribution

t-tests

Hypothesis H_0	Conditions	Test statistic t	Degrees of freedom
$\mu = \mu_0$	Sample size n; unknown σ	$t = \dfrac{\bar{X} - \mu_0}{s/\sqrt{n}}$	$n - 1$
$\mu_X - \mu_Y = D$	Sample sizes n, m; unknown but equal standard deviations, $\sigma_X = \sigma_Y$	$t = \dfrac{\bar{X} - \bar{Y} - D}{s_p\sqrt{\frac{1}{n} + \frac{1}{m}}}$	$n + m - 2$
$\mu_X - \mu_Y = D$	Sample sizes n, m; unknown, unequal standard deviations, $\sigma_X \neq \sigma_Y$	$t = \dfrac{\bar{X} - \bar{Y} - D}{\sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}}}$	Satterthwaite approximation, formula (9.12)

Welch Two-Sample Test

Pooled Sample Variance:

$$s_p^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2 + \sum_{i=1}^{m}(Y_i - \bar{Y})^2}{n + m - 2} = \frac{(n-1)s_X^2 + (m-1)s_Y^2}{n + m - 2}.$$

Satterthwaite Approximation for Degrees of Freedom

$$\nu = \frac{\left(\frac{s_X^2}{n} + \frac{s_Y^2}{m}\right)^2}{\frac{s_X^4}{n^2(n-1)} + \frac{s_Y^4}{m^2(m-1)}}. \qquad \textbf{(9.12)}$$

Example: One sample t-test

- Let us consider the daily energy intake in kJ for 11 women: 5260, 5470, 5640, 6180, 6390, 6515, 6805, 7515, 7515, 8230, and 8770. We might wish to investigate whether the women's energy intake deviates systematically from a recommended value of 7725 kJ at 5% level of significance.
- 4 Basic Steps: State the null and alternative hypotheses, Decide the sample statistic, Analyze sample data, and calculate the value of Sample Statistic, Accept or reject the null hypothesis based on the significance level, and interpret the same.
- H0: mu = 7725, H1: mu != 7725
- We have to decide for the mean, population sd is not know known, so we will use t-statistic.
- t = $\frac{\bar{X}-\mu}{S/\sqrt{n}}$ = -2.820754 with parameter (degree of freedom) 10 (= sample size -1)
- Rejection region is defined by:

  (-∞, qt(.025, 10)) ∪ (qt(.975, 10), ∞) = (-∞, -2.2281) ∪ (2.2281, ∞).
- The calculated value of t lies in rejection region, so we will reject the null hypothesis OR we will accept the alternative hypothesis. It means that mean energy intake is not 7725 kJ.

Example: One sample t-test in R

- Let us consider the daily energy intake in kJ for 11 women: 5260, 5470, 5640, 6180, 6390, 6515, 6805, 7515, 7515, 8230, and 8770. We might wish to investigate whether the women's energy intake deviates systematically from a recommended value of 7725 kJ at 5% level of significance.
- daily.intake <- c(5260,5470,5640,6180,6390,6515,6805,7515,7515,8230,8770)
- mean(daily.intake)
- sd(daily.intake)
- quantile(daily.intake)
- #t-test
  - t.test(daily.intake,mu=7725)
- #Explanation
  - tvalue <- (mean(daily.intake) - 7725)/(sd(daily.intake)/sqrt(11))
  - print(p_value <- 2*pt(-abs(tvalue), (length(daily.intake) -1)))
- #Confidence Interval
  - print (lowerlimit <- mean(daily.intake)-qt(0.975,10)*sd(daily.intake)/sqrt(length(daily.intake)))
  - print (upperlimit <- mean(daily.intake)+qt(0.975,10)*sd(daily.intake)/sqrt(length(daily.intake)))

Wilcoxon signed-rank test

- The t tests are fairly robust against departures from the normal distribution especially in larger samples, but sometimes we wish to avoid making the normality assumption. We use the distribution-free methods here.
- For the one-sample Wilcoxon test:
  - the procedure is to subtract the theoretical $\mu 0$ and rank the differences according to their numerical value, ignoring the sign, and then calculate the sum of the positive or negative ranks.
  - The point is that, assuming only that the distribution is symmetric around $\mu 0$, the test statistic corresponds to selecting each number from 1 to n with probability 1/2 and calculating the sum. The distribution of the test statistic can be calculated exactly, at least in principle.
  - It becomes computationally excessive in large samples, but the distribution is then very well approximated by a normal distribution.
- The test statistic V is the sum of the positive ranks

Wilcoxon signed-rank test in R

- daily.intake <- c(5260,5470,5640,6180,6390,6515,6805,7515,7515,8230,8770)
- wilcox.test(daily.intake, mu=7725)

- x<-runif(1000)
- wilcox.test(x, mu=0.5)

Two sample t-test Var(X) = Var(Y)

- x <- runif(100)
- y <- runif(100, 0, 1.5)
- t.test(x, y, var.equal = T)

Two Sample t-test

data: x and y

t = -3.1173, df = 198, p-value = 0.002097

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-0.26638341 -0.05994594

sample estimates:

mean of x mean of y

0.5371837 0.7003484

- In this test, Population Variances are not known.
  - There are 2 cases: v.X = v.Y, and v.X != v.Y
- When v.X = v.Y, we calculate:

- sp2<- ((n-1)*var(x) +(m-1)*var(y))/(n+m-2)
- tvalue<-(mean(x)-mean(y))/(sqrt(sp2)*sqrt(1/n + 1/m))
- tvalue

- pvalue<-2*pt(-abs(tvalue),n+m-2)
- pvalue

Two-sample t-test (Var(X) != Var(Y))

- data <- read.csv("C:/Users/RKS/OneDrive/marks for two sample t-test.csv")
- t.test(data$marks~data$year)

```
Welch Two Sample t-test

data:  data$marks by data$year
t = 10.033, df = 147.57, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 17.91445 26.70257
sample estimates:
mean in group Y1 mean in group Y2
        68.15386         45.84535
```

  - H0: mu1-mu2 =0
  - H1: mu1-mu2 != 0

- We have to decide for the mean, populations' sds are not know known, so we will use t-statistic.
- Since the p-value < .05, we reject the null hypothesis and conclude that the mean of marks of the 2 classes are not equal.

Two-sample t-test (Var(X) != Var(Y))

- `data <- read.csv("C:/Users/RKS/OneDrive/marks for two sample t-test.csv")`
- `t.test(data$marks~data$year)`

  - H0: mu1-mu2 =0
  - H1: mu1-mu2 != 0

```
Welch Two Sample t-test

data:  data$marks by data$year
t = 10.033, df = 147.57, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 17.91445 26.70257
sample estimates:
mean in group Y1 mean in group Y2
        68.15386         45.84535
```

- `x<- subset(data$marks, data$year == "Y1")`
- `y<- subset(data$marks, data$year == "Y2")`
- `n<-length(x)`
- `m<-length(y)`
- `tvalue<-(mean(x)-mean(y))/sqrt(var(x)/n + var(y)/m)`
- `tvalue`
- `dof<-((var(x)/n + var(y)/m)^2) / (var(x)*var(x)/(n*n*(n-1)) + var(y)*var(y)/(m*m*(m-1)))`
- `dof`

Chi-square test for the population variance

Null Hypothesis	Alternative Hypothesis	Test statistic	Rejection region	P-value
$\sigma^2 = \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$\frac{(n-1)s^2}{\sigma_0^2}$	$\chi^2_{\text{obs}} > \chi^2_\alpha$	$P\left\{\chi^2 \geq \chi^2_{\text{obs}}\right\}$
	$\sigma^2 < \sigma_0^2$		$\chi^2_{\text{obs}} < \chi^2_\alpha$	$P\left\{\chi^2 \leq \chi^2_{\text{obs}}\right\}$
	$\sigma^2 \neq \sigma_0^2$		$\chi^2_{\text{obs}} \geq \chi^2_{\alpha/2}$ or $\chi^2_{\text{obs}} \leq \chi^2_{1-\alpha/2}$	$2\min\left(P\left\{\chi^2 \geq \chi^2_{\text{obs}}\right\}, P\left\{\chi^2 \leq \chi^2_{\text{obs}}\right\}\right)$

F-test for the ratio of Population Variances

Null Hypothesis $H_0 : \dfrac{\sigma_X^2}{\sigma_Y^2} = \theta_0$		Test statistic $F_{\text{obs}} = \dfrac{s_X^2}{s_Y^2} / \theta_0$
Alternative Hypothesis	Rejection region	P-value Use F$(n-1, m-1)$ distribution
$\dfrac{\sigma_X^2}{\sigma_Y^2} > \theta_0$	$F_{\text{obs}} \geq F_\alpha(n-1, m-1)$	$\boldsymbol{P}\{F \geq F_{\text{obs}}\}$
$\dfrac{\sigma_X^2}{\sigma_Y^2} < \theta_0$	$F_{\text{obs}} \leq F_\alpha(n-1, m-1)$	$\boldsymbol{P}\{F \leq F_{\text{obs}}\}$
$\dfrac{\sigma_X^2}{\sigma_Y^2} \neq \theta_0$	$F_{\text{obs}} \geq F_{\alpha/2}(n-1, m-1)$ or $F_{\text{obs}} < 1/F_{\alpha/2}(m-1, n-1)$	$2\min\left(\boldsymbol{P}\{F \geq F_{\text{obs}}\}, \boldsymbol{P}\{F \leq F_{\text{obs}}\}\right)$

Comparison of variances

- data <- read.csv("C:/Users/RKS/OneDrive/marks for two sample t-test.csv")
- x<- subset(data$marks, data$year == "Y1")
- y<- subset(data$marks, data$year == "Y2")
- var.test(x, y)

```
 F test to compare two variances

data:  x and y
F = 0.93407, num df = 100, denom df = 70, p-
value = 0.7473
alternative hypothesis: true ratio of variances is
not equal to 1
95 percent confidence interval:
 0.5994874 1.4310026
sample estimates:
ratio of variances
         0.9340661
```

- H0: Var(X) – Var(Y) =0
- H1: Var(X) – Var(Y) != 0

- print(F <- var(x) / var(y))
- print(num_df <- length(x) - 1)
- print(denom_df <- length(y)-1)
- print(p_value <- 2*pf(F, num_df, denom_df))

```
pp <- pf(F, num_df, denom_df)
pp
if (pp < 0.5) print(p_value <- 2*pp)
if (pp >= 0.5) print(p_value <- 2*(1-pp))
```

Thank You !