**KUMARAGURU COLLEGE OF TECHNOLOGY**

**LABORATORY MANUAL**

**Experiment Number: 5**

**Lab Code : U18MAI4201**

**Lab : Probability and Statistics**

**Course / Branch : B.E-CSE,ISE, B.Tech-IT**

**Title of the Experiment : Applications of Student t-test**

**STEP 1: INTRODUCTION**

**OBJECTIVES OF THE EXPERIMENT**

1. To apply t-test to test hypothesis about population mean

2. To apply t-test to test hypothesis about two means

3. To apply paired t-test to test hypotheses about means of two dependent samples

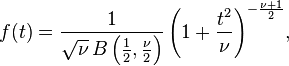
**STEP 2: ACQUISITION**

**Student’s t – distribution**

Student's **t-distribution** has the [probability density function](https://en.wikipedia.org/wiki/Probability_density_function) given by

f(t) = \frac{\Gamma(\frac{\nu+1}{2})} {\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})} \left(1+\frac{t^2}{\nu} \right)^{-\frac{\nu+1}{2}},\!

where \nu is the number of [degrees of freedom](https://en.wikipedia.org/wiki/Degrees_of_freedom_(statistics)) and \Gamma is the [gamma function](https://en.wikipedia.org/wiki/Gamma_function). This may also be written as



Note: (a) The values of can be got from the t – table

(b) gives the critical value of t for a single tail test at LOS and d.f

For eg, t8(0.05) for single tailed test = t8(10) for two-tailed test = 1.86

**Test of Hypothesis about the Population Mean**

Test statistic follows t – distribution with n-1 degrees of freedom.

where and

Null hypothesis H0 : There is no significant difference between the sample mean and the population mean .

**If tabulated t, then H0 is accepted and the difference between and is not considered significant.**

**Assumptions for t – test for population mean**

1. The parent population from which the sample is drawn is normal.
2. The sample observations are independent
3. The population standard deviation is unknown.

**Test of Hypothesis about the difference between two means**

To test a hypothesis concerning the difference between the means of two normally distributed populations, when the population variances are unknown, t – test is used.

H0: The samples have been drawn from populations with same means, ie,

Test statistic is ~

where

or **,** where

(Note : S2 is an unbiased estimate of the population variance )

The test statistic follows t-distribution with n1 +n2 -2 degrees of freedom.

**If tabulated t, then H0 is accepted and the difference between and is not considered significant.**

**Paired t-test for difference of Means**

If the two given samples are dependent, ie, each observation in one sample is associated with a particular observation in the second sample, then we use paired t – test to test whether the means differ significantly or not. Here , both the samples will have same number of units.

The test statistic is

follows t – distribution with n-1 d.f. Here n is the number of pairs in the sample

**Using R for testing of hypothesis**

The R function t.test() can be used to perform both one and two sample t-tests on vectors of data.

The function contains a variety of options and can be called as follows:

t.test(x, y = NULL, alternative = c("two.sided", "less", "greater"), mu = 0, paired = FALSE, var.equal = FALSE, conf.level = 0.95)

Here x is a numeric vector of data values and y is an optional numeric vector of data values. If y is excluded, the function performs a one-sample t-test on the data contained in x, if it is included it performs a two-sample t-tests using both x and y.

The option mu provides a number indicating the true value of the mean (or difference in means if you are performing a two sample test) under the null hypothesis. The option alternative is a character string specifying the alternative hypothesis, and must be one of the following: "two.sided" (which is the default), "greater" or "less" depending on whether the alternative hypothesis is that the mean is different than, greater than or less than mu, respectively.

**Procedure for doing the Experiment:**

|  |  |
| --- | --- |
| **1.** | To test hypothesis about population mean:  (a)For a two-tailed test  x = c()  t.test(x,alternative="two.sided",mu=)  (b) For a one-tailed test  x = c()  t.test(x,alternative="less"/"greater",mu=) |
| **2.** | To test hypothesis about two means  A= c()  B = c()  t.test(A,B,alternative="two.sided"/"less"/"greater",, var.equal=TRUE) |
| **3.** | To use paired t-test  A= c()  B = c()  t.test(A,B,alternative="greater"/"less"/"two.sided",paired=TRUE) |

**EXAMPLE – Single mean**

**Eleven articles produced by a factory were chosen at random and their weights were found to be (in kgs) 63,63,66,67,68,69,70,70,71,71,71 respectively. In the light of the above data, can we assume that the mean weight of the articles produced by the factory is 66 kgs? (Given: the critical value of for 10 degrees of freedom at 5% LOS is 2.28).**

**R-code**

x = c(63,63,66,67,68,69,70,70,71,71,71)

t.test(x,alternative="two.sided",mu=66)

**Output:**

One Sample t-test

data: x

t = 2.3, df = 10, p-value = 0.04425

alternative hypothesis: true mean is not equal to 66

95 percent confidence interval:

66.06533 70.11649

sample estimates:

mean of x

68.09091

**Conclusion**: -value = 2.3 > 2.228. Hence we reject and we may conclude that the mean

weight of the articles produced by the factory is not 66

**Task 1**

**Tests made on the breaking strength of 10 pieces of a metal gave the followingresults. 578, 572, 570, 568, 572, 570, 570, 572, 596 and 584 kg.**

**Test if the mean breaking strength of the wire can be assumed as 577kg.**

**Null hypothesis:**

**Alternate hypothesis:**

**R-code**

x = c(578,572,570,568,572,570,570,572,596,584)

t.test(x,alternative="two.sided",mu=577)

**Output:**

One Sample t-test

data: x

t = -0.65408, df = 9, p-value = 0.5294

alternative hypothesis: true mean is not equal to 577

95 percent confidence interval:

568.9746 581.4254

sample estimates:

mean of x

575.2

**Conclusion:**

-value = 0.65408 < 2.262. Hence we accept and we may conclude that the mean breaking strength of the wire can be assumed as 577kg.

**Task 2**

**The heights of 10 men in a given locality are found to be 70, 67, 62, 68, 61, 68, 70, 64, 64, 66 inches. Is it reasonable to believe that the average height is greater than 64 inches?**

**Null hypothesis**

**Alternate hypothesis: :**

**R-code:**

x = c(70, 67, 62, 68, 61, 68, 70, 64, 64, 66)

t.test(x,alternative="greater",mu=64)

**Output :**

One Sample t-test

data: x

t = 2, df = 9, p-value = 0.03828

alternative hypothesis: true mean is greater than 64

95 percent confidence interval:

64.16689 Inf

sample estimates:

mean of x

66

**Conclusion:**

-value = 2 > 1.833 . Hence we reject and we may conclude that the mean height is greater than 64 inches.

**Example 2: Two means**

**6 subjects were given a drug (treatment group) and an additional 6 subjects a placebo (control group). Their reaction time to a stimulus was measured (in ms).**

**Placebo group: 91, 87, 99, 77, 88, 91**

**Treatment group : 101, 110, 103, 93, 99, 104**

**Can we conclude that the reaction time of the placebo group is less than that of the treatment group? (Required table value of t = 1.1812)**

**Null hypothesis** : , ie. the reaction times of the two groups are equal.

**Alternate hypothesis**: ie, the reaction time of the placebo group is less than that of the treatment group

**R-code:**

Control = c(91, 87, 99, 77, 88, 91)

Treat = c(101, 110, 103, 93, 99, 104)

t.test(Control,Treat,alternative="less", var.equal=TRUE)

**Output:**

Two Sample t-test

data: Control and Treat t = -3.4456, df = 10, p-value = 0.003136 alternative hypothesis: true difference in means is less than 0

**Conclusion**: -value =-3.4456 , =3.4456 > 1.1812. Hence we may conclude that the reaction time of placebo group is less than that of treatment group.

**Task 3**

**Two independent samples are chosen from two schools A and B and common test is given in a subject. The scores of the students are as follows:**

**School A: 76 68 70 43 94 68 33**

**School B: 40 48 92 85 70 76 68 22.**

**Can we conclude that students of school A performed better than students of school B.**

**Null hypothesis**: , ie, Students of both schools performed equally well.

**Alternate hypothesis**: ie, Students of school A performed better than students of school B.

**R-code:**

A = c(76,68,70,43,94,68,33)

B = c(40,48,92,85,70,76,68,22)

t.test(A,B,alternative="greater",var.equal=TRUE)

**Output:**

Two Sample t-test

data: A and B

t = 0.16802, df = 13, p-value = 0.4346

alternative hypothesis: true difference in means is greater than 0

95 percent confidence interval:

-18.56956 Inf

sample estimates:

mean of x mean of y

64.57143 62.62500

**Conclusion:**

-value = 0.16802 < 1.771. Hence we accept ,we may conclude that there is no significant difference in the performance of the students of the two schools.

**Task 4**

**Two independent samples of sizes 8 and 7 contained the following values.**

**Sample 1:19 17 15 21 16 18 16 14**

**Sample 2:15 14 15 19 15 18 16**

**Is the difference between the sample means significant?**

**Null hypothesis :**  , ie, There is no significant difference between the means of the two samples.

**Alternate hypothesis** : ie, There is a significant difference between the means of the two samples.

**R-code:**

Samp1 = c(19,17,15,21,16,18,16,14)

Samp2 = c(15,14,15,19,15,18,16)

t.test(Samp1,Samp2,alternative="two.sided",var.equal=TRUE)

**Output:**

Two Sample t-test

data: Samp1 and Samp2

t = 0.93095, df = 13, p-value = 0.3688

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

95 percent confidence interval:

-1.320608 3.320608

sample estimates:

mean of x mean of y

17 16

**Conclusion:**

-value = 0.93095 < 2.160. Hence we accept ,we may conclude that there is no significant difference in the means of the two samples.

**Example 3: Paired t-test**

**A study was performed to test whether cars get better mileage on premium gas than on regular gas. Each of 10 cars was first filled with either regular or premium gas, decided by a coin toss, and the mileage for that tank was recorded. The mileage was recorded again for the same cars using the other kind of gasoline. The relevant mileages : Regular: 16, 20, 21, 22, 23, 22, 27, 25, 27, 28 Premium :19, 22, 24, 24, 25, 25, 26, 26, 28, 32 . Use a paired t test to determine whether cars get significantly better mileage with premium gas.**

**Null Hypothesis H0 :** , ie, the two types of bulbs are identical regarding length of life.

**Alternative Hypothesis: H1** :

reg=c(16,20,21,22,23,22,27,25,27,28)

prem=c(19,22,24,24,25,25,26,26,28,32)

t.test(prem,reg,alternative="greater",paired=TRUE)

Paired t-test

data: prem and reg

t = 4.4721, df = 9, p-value = 0.0007749

alternative hypothesis: true difference in means is greater than 0

95 percent confidence interval:

1.180207 Inf

sample estimates:

mean of the differences

2

Conclusion: p-value = 0.0007749 < 0.05 Hence we reject and we may conclude that cars get significantly better mileage with premium gas.

**Task 5**

**The weight gain in pounds under two systems of feeding of calves of 10 pairs of identical twins is given below.**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Twin pair** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
| **Weight gain under System A** | **43** | **39** | **39** | **42** | **46** | **43** | **38** | **44** | **51** | **43** |
| **Weight gain under System B** | **37** | **35** | **34** | **41** | **39** | **37** | **37** | **40** | **48** | **36** |

**Discuss whether the difference between the two systems of feeding is significant.**

**Null Hypothesis :**  , ie, There is no significant difference between the two systems of feeding.

**Alternate hypothesis** : ie, There is a significant difference between the two systems of feeding.

**R-code:**

SysA=c(43,39,39,42,46,43,38,44,51,43)

SysB=c(37,35,34,41,39,37,37,40,48,36)

t.test(SysA,SysB,alternative="two.sided",paired=TRUE)

**Output:**

Paired t-test

data: SysA and SysB

t = 6.2644, df = 9, p-value = 0.0001471

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

95 percent confidence interval:

2.811113 5.988887

sample estimates:

mean of the differences

4.4

**Conclusion:**

-value = 6.2644 > 2.262. Hence we reject ,we may conclude that there is a significant difference between the two systems of feeding.

**Task 6**

**Ten persons were appointed in the officer cadre in an office. Their performance was noted by giving a test and the marks were recorded out of 100.**

**Employee A B C D E F G H I J**

**Before training 80 76 92 60 70 56 74 56 70 56**

**After training 84 70 96 80 70 52 84 72 72 50**

**By applying t test, can it be concluded that the employees have been benefited by the training?**

**Null hypothesis: :**  , ie, The employees have not been benefitted by the training.

**Alternate hypothesis** : ie, The employees have been benefitted by the training.

**R-code:**

Before=c(80,76,92,60,70,56,74,56,70,56)

After=c(84,70,96,80,70,52,84,72,72,50)

t.test(Before,After,alternative="less",paired=TRUE)

**Output:**

Paired t-test

data: Before and After

t = -1.4142, df = 9, p-value = 0.09547

alternative hypothesis: true difference in means is less than 0

95 percent confidence interval:

-Inf 1.184826

sample estimates:

mean of the differences

-4

**Conclusion:**

-value = 1.4142 < 1.83. Hence we accept ,we may conclude that the employees have not been benefitted by the training.

**STEP 3: PRACTICE/TESTING**

1. **Write the test statistic for testing hypothesis about a population mean.**

T=(x-y)/(s/sqrt(n))

1. **Write the test statistic for testing of hypothesis about the difference between two means .**

T=(x-y)/s(sqrt(1/n1+1/n2))

1. **Write the test statistic for testing of hypothesis about the difference between means of two dependent samples. (paired t-test)**

T= d/(s/sqrt(n))

1. **Define level of significance.**

The significance level of an event is the probability that the event could have occurred by chance

**KUMARAGURU COLLEGE OF TECHNOLOGY**

**LABORATORY MANUAL**

**Experiment Number: 6**

**Lab Code : U18MAI4201**

**Lab : Probability and Statistics**

**Course / Branch : B.E-CSE,ISE, B.Tech-IT**

**Title of the Experiment/experiment :Applications of F test**

**STEP 1: INTRODUCTION**

**OBJECTIVES OF THE EXPERIMENT**

To apply F-test to compare the variances of two samples from normal populations.

**STEP 2: ACQUISITION**

The null hypothesis is that the ratio of the variances of the populations from which x and y were drawn, or in the data to which the linear models x and y were fitted, is equal to ratio.

**Procedure for doing the Experiment:**

|  |  |
| --- | --- |
|  | R-Code for F-test:  var.test(x, y, ratio = 1,alternative = c("two.sided", "less", "greater"),conf.level = 0.95, ...) |

**Note:**

|  |  |
| --- | --- |
| x, y | * numeric vectors of data values, or fitted linear model objects (inheriting from class "lm"). |
| Ratio | * the hypothesized ratio of the population variances of x and y. |
| Alternative | * a character string specifying the alternative hypothesis, must be one of "two.sided" (default), "greater" or "less". You can specify just the initial letter. |
| conf.level | * confidence level for the returned confidence interval. |

In the test statistic, the greater of the two variances and is to be taken in the numerator and corresponds to the greater variance.

**Example:**

**Two samples of 6 and 7 items respectively have the following values for a variable**

**Sample 1 39 41 42 42 44 40**

**Sample 2 40 42 39 45 38 39 40**

**Do the sample variances differ significantly?**

**Null Hypothesis: There is no significant difference in sample variances.**

**Alternative Hypothesis: There is a significant difference in sample variances.**

**Code:**

x=c(40,42,39,45,38,39,40)

y=c(39,41,42,42,44,40)

var.test(x, y, ratio = 1,

alternative = c("two.sided"),

conf.level = 0.95)

**Output:**

F test to compare two variances

data: x and y

F = 1.8323, numdf = 6, denomdf = 5, p-value = 0.523

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

0.2625934 10.9710044

sample estimates:

ratio of variances

1.832298

Critical value of for (6, 5) d.f. is

**Conclusion:Since**  , **we accept the null hypothesis and we may conclude that**

**there is no significant difference in the sample variances.**

**Task 1:**

**Two random samples drawn from two normal populations are**

**Sample 1: 20 16 26 27 23 22 18 24 25 19**

**Sample 2: 27 33 42 35 32 34 38 28 41 43 30 37**

**Test whether the populations have the same variances.**

**Null Hypothesis: The population have the same variances.**

**Alternative Hypothesis: : The population have different variances.**

**R Code:**

Samp1=c(20,16,26,27,23,22,18,24,25,19)

Samp2=c(27,33,42,35,32,34,38,28,41,43,30,37)

var.test(Samp1,Samp2,ratio = 1,alternative = c("two.sided"),conf.level = 0.95)

**Output:**

F test to compare two variances

data: Samp1 and Samp2

F = 0.46709, num df = 9, denom df = 11, p-value = 0.2629

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

0.1301852 1.8272959

sample estimates:

ratio of variances

0.4670913

Critical value of for (9,11) d.f. is

**Conclusion:**

Since , we accept the null hypothesis and we may conclude that the population have the same variances.

**Task 2:**

**The nicotine content in 2 random samples of tobacco are given below:**

**Sample 1: 21 24 25 26 27**

**Sample 2: 22 27 28 30 31 36**

**Test whether the populations have the same variances.**

**Null Hypothesis: The population have the same variances.**

**Alternative Hypothesis: The population have different variances.**

**R Code:**

Samp1=c(21,24,25,26,27)

Samp2=c(22,27,28,30,31,36)

var.test(Samp1,Samp2,ratio = 1,alternative = c("two.sided"),conf.level = 0.95)

**Output:**

F test to compare two variances

data: Samp1 and Samp2

F = 0.24537, num df = 4, denom df = 5, p-value = 0.1981

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

0.03321253 2.29776367

sample estimates:

ratio of variances

0.2453704

Critical value of for (4,5) d.f. is

**Conclusion:**

Since , we accept the null hypothesis and we may conclude that the population have the same variances.

**Task 3:**

**2 independent samples of 8 and 7 items have the following values.**

**Sample 1: 9 11 13 11 15 9 12 14**

**Sample 2: 10 12 10 14 9 8 10**

**Can we conclude that the two samples have drawn from the same normal population**.

To test whether the samples come from the same normal population, we have to test for

1. Equality of population means
2. Equality of population variances.

Equality of means is tested using t-test and equality of variances is tested using F-test.

Since t-test assumes , we first apply *F*-test and then t-test.

***F*-test:**

**Null Hypothesis:**

**Alternative Hypothesis:**

**R Code:**

Sample1=c(9,11,13,11,15,9,12,14)

Sample2=c(10,12,10,14,9,8,10)

var.test(Sample1,Sample2,ratio = 1,alternative = c("two.sided"),conf.level = 0.95)

**Output:**

F test to compare two variances

data: Sample1 and Sample2

F = 1.2108, num df = 7, denom df = 6, p-value = 0.8315

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

0.2125976 6.1978188

sample estimates:

ratio of variances

1.210843

Critical value of for (7,6) d.f. is

**Conclusion:**

Since , we accept the null hypothesis and we may conclude that the population have the same variances.

***t*-test:**

**Null Hypothesis: :**  ,

**Alternate hypothesis** :

**R Code:**

Sample1=c(9,11,13,11,15,9,12,14)

Sample2=c(10,12,10,14,9,8,10)

t.test(Sample1,Sample2,alternative="two.sided",var.equal=TRUE)

**Output:**

Two Sample t-test

data: Sample1 and Sample2

t = 1.2171, df = 13, p-value = 0.2452

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

95 percent confidence interval:

-1.024204 3.667061

sample estimates:

mean of x mean of y

11.75000 10.42857

**Conclusion:**

-value = 1.2171 < 2.160 . Hence we accept ,we may conclude that the population means are same.

**Final conclusion:**

Since both the null hypothesis are accepted, we may conclude that the given samples have drawn from the same normal population.

**Task 4:**

**Two horses A and B were tested according to the time(in seconds) to run a particular track with the following results:**

**Horse A: 28 30 32 33 33 29 34**

**Horse B: 29 30 30 24 27 29**

**Test whether the two horses have the same running capacity in terms of average and variance of time taken.**

**F** – **Test:**

**Null Hypothesis:**  The two horses have the same running capacity in terms of variance of time taken.

**Alternative Hypothesis:**  The two horses does not have the same running capacity in terms of variance of time taken.

**R Code:**

HorseA=c(28,30,32,33,33,29,34)

HorseB=c(29,30,30,24,27,29)

var.test(HorseA,HorseB,ratio = 1,alternative = c("two.sided"),conf.level = 0.95)

**Output:**

F test to compare two variances

data: HorseA and HorseB

F = 0.97604, num df = 6, denom df = 5, p-value = 0.9573

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

0.1398802 5.8441186

sample estimates:

ratio of variances

0.9760426

Critical value of for (6,5) d.f. is

**Conclusion:**

Since , we accept the null hypothesis and we may conclude that the two horses have the same running capacity in terms of variance of time taken.

**T - Test:**

**Null Hypothesis: :**  , The two horses have the same running capacity in terms of average of time taken.

**Alternate hypothesis** : The two horses does not have the same running capacity in terms of average of time taken.

**R Code:**

HorseA=c(28,30,32,33,33,29,34)

HorseB=c(29,30,30,24,27,29)

t.test(HorseA,HorseB,alternative="two.sided",var.equal=TRUE)

**Output:**

Two Sample t-test

data: HorseA and HorseB

t = 2.436, df = 11, p-value = 0.03306

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

95 percent confidence interval:

0.3009233 5.9371719

sample estimates:

mean of x mean of y

31.28571 28.16667

**Conclusion:**

-value = 2.436 > 2.201 . Hence we reject ,we may conclude that the two horses does not have the same running capacity in terms of average of time taken.

**Final Conclusion:**

In the t-test, the null hypothesis is rejected. So the two horses have the same running capacity only in terms of variance and not in terms of average of time taken.

**Task 5:**

**Two samples are drawn from two normal populations. From the following data test whether the two samples have the same variance at 5% level:**

**Sample 1: 60 65 71 74 76 82 85 87**

**Sample 2: 61 66 67 85 78 63 85 86 88 91.**

**Null Hypothesis:**  the two samples have the same variance.

**Alternative Hypothesis:**  the two samples have different variance.

**R Code:**

Samp1=c(60,65,71,74,76,82,85,87)

Samp2=c(61,66,67,85,78,63,85,86,88,91)

var.test(Samp1,Samp2,ratio = 1,alternative = c("two.sided"),conf.level = 0.95)

**Output:**

F test to compare two variances

data: Samp1 and Samp2

F = 0.68143, num df = 7, denom df = 9, p-value = 0.6271

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

0.1623591 3.2866779

sample estimates:

ratio of variances

0.6814286

Critical value of for (7,9) d.f. is

**Conclusion:**

Since , we accept the null hypothesis and we may conclude that the two samples have the same variance at 5 % level.

**STEP 3: PRACTICE/TESTING**

**1. What is the use of *F*-distribution?**

The main use of F distribution is to chech whether two independent samples have been drawn for the same variance or if two independent estimates of the population variance are homogeneous or not, since it is often desirable to compare two variance rather than two averages.

1. **State the important properties of *F*-distribution.**

1)F- distribution is positively skewed.

2)Value of F lies between 0 and

1. **What is the difference between *F*-test and *t*-test?**

t-test is used to test if two sample have the same mean. The assumptions are that they are samples from normal distribution. F-test is used to test if two sample have the same variance.