

# MVLU COLLEGE

## 7 Performing one-way ANOVA using aov() (R).

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
practical no.7 module 2.R x practical no.8 module 2.R x practical no.9 module 2.R x
29:1 (Top Level)
R Script
R 4.5.2 ~ /
14 14 14 2.0 99 00
15 15 16 2.3 79 69
16 16 18 4.2 97 91
17 17 17 1.8 72 58
18 18 16 3.6 90 84
19 19 18 2.1 77 66
20 20 17 3.9 96 90
>
> colnames(data1)[colnames(data1) == "score (%)"] <- "score"
>
> data1$age <- as.factor(data1$age)
>
> group_means <- tapply(data1$score, data1$age, mean)
>
> grand_mean <- mean(data1$score)
>
> group_sizes <- table(data1$age)
>
> SSB <- sum(group_sizes * (group_means - grand_mean)^2)
>
> SSW <- sum((data1$score - group_means[data1$age])^2)
>
> df_between <- length(group_means) - 1
> df_within <- nrow(data1) - length(group_means)
>
> MSB <- SSB / df_between
> MSW <- SSW / df_within
>
> F_value <- MSB / MSW
>
> F_value
[1] 0.3095695
> |
```

Environment History Connections  
R Global Environment  
data3 499 obs. of 5 v...  
values  
df\_bet... 2  
df\_wit... 171  
F\_value 0.30956948774359  
grand... 74.25  
group... num [1:3(1d)] 70...  
group... 'table' int [1:3(1d)]

Files Plots Packages Help Vi  
Home  
Name  
RData  
Rhistory  
Custom Office Templates  
desktop.ini  
Downloads - ShortcutLink  
FeedbackHub  
MU0341120240122781\_Bachelor\_of\_S  
My Music  
My Pictures  
My Videos  
NetBeansProjects  
policy details - ShortcutLink  
policy details.txt

## 8 Performing two-way ANOVA using aov() (R).

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
Source
R 4.5.2 ~ /
> data2 <- read.csv("C:/Users/Admin/Desktop/ANKITA TIWARI (data science R)/datasets/titanic.csv")
>
> data2$pclass <- as.factor(data2$pclass)
> data2$sex <- as.factor(data2$sex)
>
> grand_mean <- mean(data2$fare)
>
> tm_means <- tapply(data2$fare, data2$pclass, mean)
> tm_sizes <- table(data2$pclass)
> SS_TM <- sum(tm_sizes * (tm_means - grand_mean)^2)
>
> gender_means <- tapply(data2$fare, data2$sex, mean)
> gender_sizes <- table(data2$sex)
> SS_G <- sum(gender_sizes * (gender_means - grand_mean)^2)
>
> SS_total <- sum((data2$fare - grand_mean)^2)
> SS_error <- SS_total - SS_TM - SS_G
>
> df_TM <- length(tm_means) - 1
> df_G <- length(gender_means) - 1
> df_error <- nrow(data2) - length(tm_means) - length(gender_means) + 1
>
> MS_TM <- SS_TM / df_TM
> MS_G <- SS_G / df_G
> MS_error <- SS_error / df_error
>
> F_TM <- MS_TM / MS_error
> F_G <- MS_G / MS_error
>
> F_TM
[1] 255.1855
> F_G
[1] 48.05352
> |
```

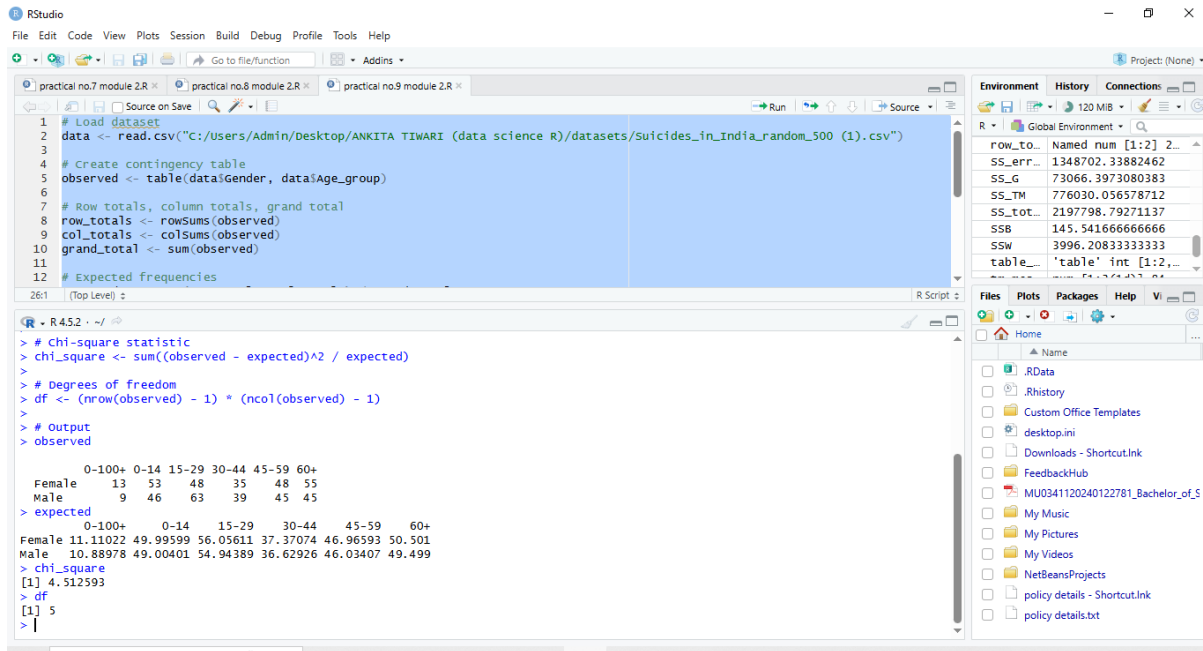
Environment History Connections  
R Global Environment  
SS\_G 73066.3973080383  
SS\_TM 776030.056578712  
SS\_tot... 2197798.79271137  
SSB 145.54166666666666  
SSW 3996.20833333333  
table... 'table' int [1:2,...  
tm\_mea... num [1:3(1d)] 84...  
tm\_siz... 'table' int [1:3(1d)]

Files Plots Packages Help Vi  
Home  
Name  
RData  
Rhistory  
Custom Office Templates  
desktop.ini  
Downloads - ShortcutLink  
FeedbackHub  
MU0341120240122781\_Bachelor\_of\_S  
My Music  
My Pictures  
My Videos  
NetBeansProjects  
policy details - ShortcutLink  
policy details.txt

# MVLU COLLEGE

## 9 Conducting Chi-square tests using chisq.test() (R)

### Chi-Square Test of Independence



```
1 # Load dataset
2 data <- read.csv("C:/Users/Admin/Desktop/ANKITA TIWARI (data science R)/datasets/Suicides_in_India_random_500 (1).csv")
3
4 # Create contingency table
5 observed <- table(data$Gender, data$Age_group)
6
7 # Row totals, column totals, grand total
8 row_totals <- rowSums(observed)
9 col_totals <- colSums(observed)
10 grand_total <- sum(observed)
11
12 # Expected frequencies
13 expected <- (row_totals * col_totals) / grand_total
14
15 # Chi-square statistic
16 chi_square <- sum((observed - expected)^2 / expected)
17
18 # Degrees of freedom
19 df <- (nrow(observed) - 1) * (ncol(observed) - 1)
20
21 # Output
22 observed
23 expected
24 chi_square
25 df
```

Output:

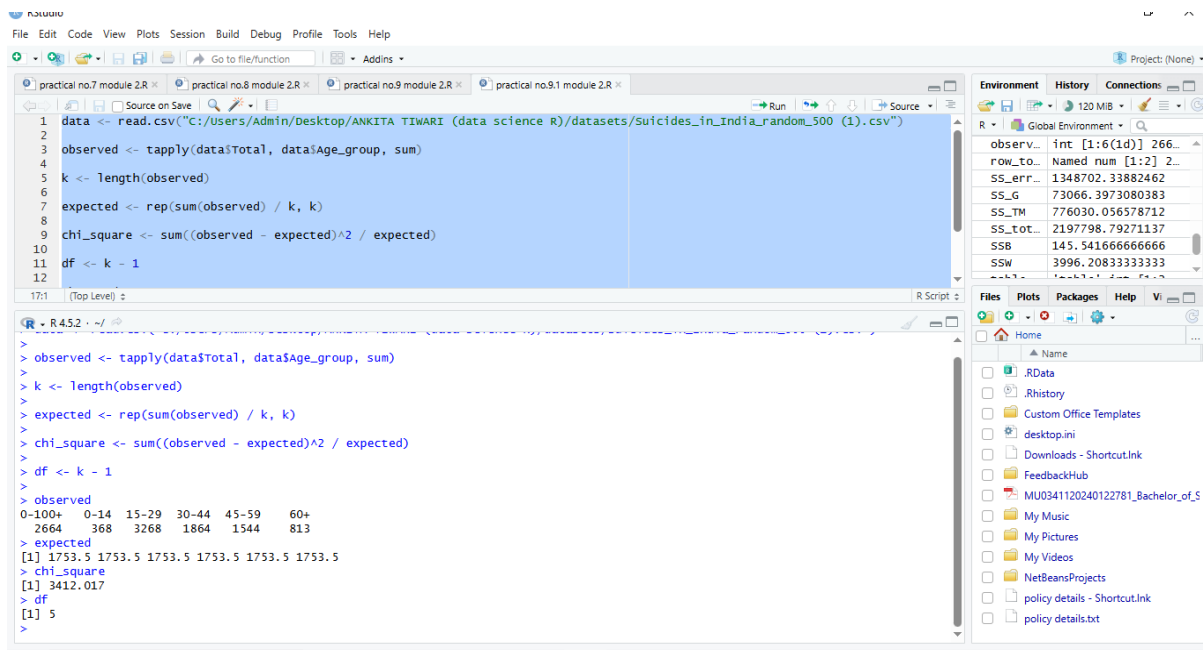
```
observed
  Female 0-100+ 0-14 15-29 30-44 45-59 60+
  Male   9    46    63    39    45    55

expected
  Female 0-100+ 0-14 15-29 30-44 45-59 60+
  Male   11.11022 49.99599 56.05611 37.37074 46.96593 50.501
  Male   10.88978 49.00401 54.94389 36.62926 46.03407 49.499

chi_square
[1] 4.512593

df
[1] 5
```

### Chi-Square Test of Goodness of Fit



```
1 data <- read.csv("C:/Users/Admin/Desktop/ANKITA TIWARI (data science R)/datasets/Suicides_in_India_random_500 (1).csv")
2
3 observed <- tapply(data$Total, data$Age_group, sum)
4
5 k <- length(observed)
6
7 expected <- rep(sum(observed) / k, k)
8
9 chi_square <- sum((observed - expected)^2 / expected)
10
11 df <- k - 1
12
```

Output:

```
observed
0-100+ 0-14 15-29 30-44 45-59 60+
2664 368 3268 1864 1544 813

expected
[1] 1753.5 1753.5 1753.5 1753.5 1753.5 1753.5

chi_square
[1] 3412.017

df
[1] 5
```