DATE:- 13-10-2023 VAMSI MANUOTHU-192011429

DAY 3 LAB MANUAL

ITA0443-STATISTICS WITH R PROGRAMMING

GITHUB LINK:- https://github.com/Vamsim29/ITA0443-STATISTICS-WITH-R-PROGRAMMING

Exercise:
I. ARITHMETIC MEAN
a) Write suitable R code to compute the average of the following values.
12,7,3,4.2,18,2,54,-21,8,-5
CODE:

```
x <- c(12,7,3,4.2,18,2,54,-21,8,-5)
result.mean <- mean(x)
print(result.mean)
```

OUTPUT:

```
> x <- c(12,7,3,4.2,18,2,54,-21,8,-5)
> result.mean <- mean(x)
> print(result.mean)
[1] 8.22
```

b) Compute the mean after applying the trim option and removing 3 values from each
end.
CODE:

```
x <- c(12,7,3,4.2,18,2,54,-21,8,-5)
result.mean <- mean(x,trim = 0.3)
print(result.mean)
```

OUTPUT:

```
> x <- c(12,7,3,4.2,18,2,54,-21,8,-5)
> result.mean <- mean(x,trim = 0.3)
> print(result.mean)
[1] 5.55
```

c) Compute the mean of the following vector .
(12,7,3,4.2,18,2,54,-21,8,-5,NA)
#If there are missing values, then the mean function returns NA.
Find mean dropping NA values.
#To drop the missing values from the calculation use na.rm = TRUE
CODE:

```
x <- c(12,7,3,4.2,18,2,54,-21,8,-5,NA)
result.mean <- mean(x)
print(result.mean)
result.mean <- mean(x,na.rm = TRUE)
print(result.mean)
```

OUTPUT:

```
> x <- c(12,7,3,4.2,18,2,54,-21,8,-5,NA)
> result.mean <- mean(x)
> print(result.mean)
[1] NA
> result.mean <- mean(x,na.rm = TRUE)
> print(result.mean)
[1] 8.22
```

II.MEDIAN

Write suitable R code to compute the median of the following values.

12,7,3,4.2,18,2,54,-21,8,-5

CODE:

```
x <- c(12,7,3,4.2,18,2,54,-21,8,-5)
median.result <- median(x)
print(median.result)
```

OUTPUT:

```
> x <- c(12,7,3,4.2,18,2,54,-21,8,-5)
> median.result <- median(x)
> print(median.result)
[1] 5.6
```

III. MODE

Calculate the mode for the following numeric as well as character data set in R.

(2,1,2,3,1,2,3,4,1,5,5,3,2,3) ,

("o","it","the","it","it")

CODE:

```
getmode <- function(v) {
uniqv <- unique(v)
uniqv[which.max(tabulate(match(v, uniqv)))]
}
v <- c(2,1,2,3,1,2,3,4,1,5,5,3,2,3)
result <- getmode(v)
print(result)
charv <- c("o","it","the","it","it")
result <- getmode(charv)
print(result)
```

OUTPUT:

```
> getmode <- function(v) {
+ uniqv <- unique(v)
```

```
+ uniqv[which.max(tabulate(match(v, uniqv)))]
+ }
> v <- c(2,1,2,3,1,2,3,4,1,5,5,3,2,3)
> result <- getmode(v)
> print(result)
[1] 2
> charv <- c("o","it","the","it","it")
> result <- getmode(charv)
> print(result)
[1] "it"
```

UNIVARIATE ANALYSIS IN R - MEASURES OF DISPERSION

Exercise: 1

Download mpg dataset which contains Fuel economy data from 1999 and 2008 for 38 popular models of car from the URL given below.

https://vincentarelbundock.github.io/Rdatasets/datasets.html

Answer the following queries

i) Find the car which gives maximum city miles per gallon

CODE:

```
require(ggplot2)
str(mpg)
nrow(mpg)
ncol(mpg)
summary(mpg$cty)
```

OUTPUT:

```
> require(ggplot2)
> str(mpg)
tibble [234 × 11] (S3: tbl_df/tbl/data.frame)
$ manufacturer: chr [1:234] "audi" "audi" "audi" "audi" ...
$ model : chr [1:234] "a4" "a4" "a4" "a4" ...
$ displ : num [1:234] 1.8 1.8 2 2 2.8 2.8 3.1 1.8 1.8 2 ...
$ year : int [1:234] 1999 1999 2008 2008 1999 1999 2008 1999 1999 2008 ...
$ cyl : int [1:234] 4 4 4 4 6 6 6 4 4 4 ...
$ trans : chr [1:234] "auto(l5)" "manual(m5)" "manual(m6)" "auto(av)" ...
$ drv : chr [1:234] "f" "f" "f" "f" ...
$ cty : int [1:234] 18 21 20 21 16 18 18 18 16 20 ...
$ hwy : int [1:234] 29 29 31 30 26 26 27 26 25 28 ...
$ fl : chr [1:234] "p" "p" "p" "p" ...
$ class : chr [1:234] "compact" "compact" "compact" "compact" ...
> nrow(mpg)
[1] 234
> ncol(mpg)
[1] 11
```

> summary(mpg$cty)
Min. 1st Qu. Median Mean 3rd Qu. Max.
9.00 14.00 17.00 16.86 19.00 35.00
ii) Find the cars which gives minimum disp in compact and subcompact class
CODE:
```
require(ggplot2)
str(mpg)
nrow(mpg)
ncol(mpg)
summary(mpg$cty)
table(mpg$class)
```
OUTPUT:
> summary(mpg$cty)
Min. 1st Qu. Median Mean 3rd Qu. Max.
9.00 14.00 17.00 16.86 19.00 35.00
> table(mpg$class)
2seater compact midsize minivan pickup subcompact suv
5 47 41 11 33 35 62
Exercise: 2
Use the same dataset as used in Exercise 1 and perform the following queries
i) Find the standard deviation of city milles per gallon
CODE:
```
str(mpg)
nrow(mpg)
ncol(mpg)
summary(mpg$cty)
table(mpg$class)
sd(mpg)
```
OUTPUT:
> str(mpg)
tibble [234 × 11] (S3: tbl_df/tbl/data.frame)
$ manufacturer: chr [1:234] "audi" "audi" "audi" "audi" ...
$ model : chr [1:234] "a4" "a4" "a4" "a4" ...
$ displ : num [1:234] 1.8 1.8 2 2 2.8 2.8 3.1 1.8 1.8 2 ...
$ year : int [1:234] 1999 1999 2008 2008 1999 1999 2008 1999 1999 2008 ...
$ cyl : int [1:234] 4 4 4 4 6 6 6 4 4 4 ...
$ trans : chr [1:234] "auto(l5)" "manual(m5)" "manual(m6)" "auto(av)" ...
$ drv : chr [1:234] "f" "f" "f" "f" ...
$ cty : int [1:234] 18 21 20 21 16 18 18 18 16 20 ...
$ hwy : int [1:234] 29 29 31 30 26 26 27 26 25 28 ...
$ fl : chr [1:234] "p" "p" "p" "p" ...
$ class : chr [1:234] "compact" "compact" "compact" "compact" ...
> nrow(mpg)

[1] 234
> ncol(mpg)
[1] 11
> summary(mpg$cty)
Min. 1st Qu. Median Mean 3rd Qu. Max.
9.00 14.00 17.00 16.86 19.00 35.00
> table(mpg$class)
2seater compact midsize minivan pickup subcompact suv
5 47 41 11 33 35 62
> sd(mpg)
ii) Find the variance of highway milles per gallon
Exercise 3
Use the same dataset and perform the following queries
i) Find the range of the disp in the data set mpg
CODE:
str(mpg)
nrow(mpg)
ncol(mpg)
summary(mpg$cty)
table(mpg$class)
range(mpg)
OUTPUT:
> str(mpg)
tibble [234 × 11] (S3: tbl_df/tbl/data.frame)
$ manufacturer: chr [1:234] "audi" "audi" "audi" "audi" ...
$ model : chr [1:234] "a4" "a4" "a4" "a4" ...
$ displ : num [1:234] 1.8 1.8 2 2 2.8 2.8 3.1 1.8 1.8 2 ...
$ year : int [1:234] 1999 1999 2008 2008 1999 1999 2008 1999 1999 2008 ...
$ cyl : int [1:234] 4 4 4 4 6 6 6 4 4 4 ...
$ trans : chr [1:234] "auto(l5)" "manual(m5)" "manual(m6)" "auto(av)" ...
$ drv : chr [1:234] "f" "f" "f" "f" ...
$ cty : int [1:234] 18 21 20 21 16 18 18 18 16 20 ...
$ hwy : int [1:234] 29 29 31 30 26 26 27 26 25 28 ...
$ fl : chr [1:234] "p" "p" "p" "p" ...
$ class : chr [1:234] "compact" "compact" "compact" "compact" ...
> nrow(mpg)
[1] 234
> ncol(mpg)
[1] 11
> summary(mpg$cty)
Min. 1st Qu. Median Mean 3rd Qu. Max.
9.00 14.00 17.00 16.86 19.00 35.00
> table(mpg$class)

2seater compact midsize minivan pickup subcompact suv
5 47 41 11 33 35 62
> range(mpg)
ii) Find the Quartile of the disp in the data set mpg
CODE:
str(mpg)
nrow(mpg)
ncol(mpg)
summary(mpg$cty)
table(mpg$class)
quantile(mpg)
OUTPUT:
> str(mpg)
tibble [234 × 11] (S3: tbl_df/tbl/data.frame)
$ manufacturer: chr [1:234] "audi" "audi" "audi" "audi" ...
$ model : chr [1:234] "a4" "a4" "a4" "a4" ...
$ displ : num [1:234] 1.8 1.8 2 2 2.8 2.8 3.1 1.8 1.8 2 ...
$ year : int [1:234] 1999 1999 2008 2008 1999 1999 2008 1999 1999 2008 ...
$ cyl : int [1:234] 4 4 4 4 6 6 6 4 4 4 ...
$ trans : chr [1:234] "auto(l5)" "manual(m5)" "manual(m6)" "auto(av)" ...
$ drv : chr [1:234] "f" "f" "f" "f" ...
$ cty : int [1:234] 18 21 20 21 16 18 18 18 16 20 ...
$ hwy : int [1:234] 29 29 31 30 26 26 27 26 25 28 ...
$ fl : chr [1:234] "p" "p" "p" "p" ...
$ class : chr [1:234] "compact" "compact" "compact" "compact" ...
> nrow(mpg)
[1] 234
> ncol(mpg)
[1] 11
> summary(mpg$cty)
Min. 1st Qu. Median Mean 3rd Qu. Max.
9.00 14.00 17.00 16.86 19.00 35.00
> table(mpg$class)
2seater compact midsize minivan pickup subcompact suv
5 47 41 11 33 35 62
> quantile(mpg
iii) Find the IQR of the disp column in the data set mpg
CODE:
OUTPUT:
Exercise 4
#Install Library
library(e1071)
a. Find the skewness of city miles per mileage in the data set mpg ?

Use qplot function and display the graph for the city miles per mileage column
CODE:

```
head(mpg)
ggplot(data = mpg, aes (x = displ, y = hwy)) +
geom_point() +
ggtitle("Engine displacement (x-axis) vs Mileage (y-axis)") +
theme(plot.title = element_text(hjust = 0.5)) +
geom_smooth(method = 'lm', se = F)
```

OUTPUT:

```
> head(mpg)
# A tibble: 6 × 11
manufacturer model displ year cyl trans drv cty hwy fl class
<chr> <chr> <dbl> <int> <int> <chr> <chr> <int> <int> <chr> <chr>
1 audi a4 1.8 1999 4 auto(l5) f 18 29 p compact
2 audi a4 1.8 1999 4 manual(m5) f 21 29 p compact
3 audi a4 2 2008 4 manual(m6) f 20 31 p compact
4 audi a4 2 2008 4 auto(av) f 21 30 p compact
5 audi a4 2.8 1999 6 auto(l5) f 16 26 p compact
6 audi a4 2.8 1999 6 manual(m5) f 18 26 p compact
> ggplot(data = mpg, aes (x = displ, y = hwy)) +
+ geom_point() +
+ ggtitle("Engine displacement (x-axis) vs Mileage (y-axis)") +
+ theme(plot.title = element_text(hjust = 0.5)) +
+ geom_smooth(method = 'lm', se = F)
`geom_smooth()` using formula = 'y ~ x'
```

b. Find the kurtosis of city miles per mileage in the data set mpg
Use qplot function and display the graph for the city miles per mileage column
CODE:

```
head(mpg)
ggplot(data = mpg) +
geom_point(mapping = aes (x=displ, y=hwy)) +
ggtitle("Engine displacement (x-axis) vs Mileage (y-axis)") +
theme(plot.title = element_text(hjust = 0.5))
```

OUTPUT:

```
head(mpg)
# A tibble: 6 × 11
manufacturer model displ year cyl trans drv cty hwy fl class
<chr> <chr> <dbl> <int> <int> <chr> <chr> <int> <int> <chr> <chr>
1 audi a4 1.8 1999 4 auto(l5) f 18 29 p compact
2 audi a4 1.8 1999 4 manual(m5) f 21 29 p compact
3 audi a4 2 2008 4 manual(m6) f 20 31 p compact
4 audi a4 2 2008 4 auto(av) f 21 30 p compact
5 audi a4 2.8 1999 6 auto(l5) f 16 26 p compact
```

6 audi a4 2.8 1999 6 manual(m5) f 18 26 p compact
> ggplot(data = mpg) +
+ geom_point(mapping = aes (x=displ, y=hwy)) +
+ ggtitle("Engine displacement (x-axis) vs Mileage (y-axis)") +
+ theme(plot.title = element_text(hjust = 0.5))
BIVARIATEANALYSIS IN R -COVARIANCE,CORRELATION,CROSSTAB
Exercise: 1
Reference Status Gender TestNewOrFollowUp
1 KRXH Accepted Female Test1 New
2 KRPT Accepted Male Test1 New
3 FHRA Rejected Male Test2 New
4 CZKK Accepted Female Test3 New
5 CQTN Rejected Female Test1 New
6 PZXW Accepted Female Test4 Follow-up
7 SZRZ Rejected Male Test4 New
8 RMZE Rejected Female Test2 New
9 STNX Accepted Female Test3 New
10 TMDW Accepted Female Test1 New
i) Load the dataset and Create a data frame and name it as dataframe1
ii) Load the function for crosstab
xtabs(~colname , data=Data frame name)
CODE:
Reference = c("KRXH", "KRPT", "FHRA", "CZKK", "CQTN", "PZXW", "SZRZ", "RMZE", "STNX", "TMDW")
Status = c("Accepted", "Accepted", "Rejected", "Accepted", "Rejected", "Accepted", "Rejected", "Rejected", "Accepted", "Accepted")
Gender = c("Female", "Male", "Male", "Female", "Female", "Female", "Male", "Female", "Female", "Female")
Test = c("Test1", "Test1", "Test2", "Test3", "Test1", "Test4", "Test4", "Test2", "Test3", "Test1")
NewOrFollowUp = c("New", "New", "New", "New", "New", "Follow-up", "New", "New", "New", "New")
data1<-data.frame(Reference,Status,Test,NewOrFollowUp)
print(data1)
CODE:
Reference = c("KRXH", "KRPT", "FHRA", "CZKK", "CQTN", "PZXW", "SZRZ", "RMZE", "STNX", "TMDW")
Status = c("Accepted", "Accepted", "Rejected", "Accepted", "Rejected", "Accepted", "Rejected", "Rejected", "Accepted", "Accepted")
Gender = c("Female", "Male", "Male", "Female", "Female", "Female", "Male", "Female", "Female", "Female")
Test = c("Test1", "Test1", "Test2", "Test3", "Test1", "Test4", "Test4", "Test2", "Test3", "Test1")

```
NewOrFollowUp = c("New", "New", "New", "New", "New", "Follow-up", "New",
"New", "New", "New")
data1<-data.frame(Reference,Status,Test,NewOrFollowUp )
print(data1)
xtab(data(Reference,Status,Test,NewOrFollowUp))
```

OUTPUT:

```
> Reference = c("KRXH", "KRPT", "FHRA", "CZKK", "CQTN", "PZXW", "SZRZ",
"RMZE", "STNX", "TMDW")
> Status = c("Accepted", "Accepted", "Rejected", "Accepted", "Rejected",
"Accepted", "Rejected", "Rejected", "Accepted", "Accepted")
> Gender = c("Female", "Male", "Male", "Female", "Female", "Female", "Male",
"Female", "Female", "Female")
> Test = c("Test1", "Test1", "Test2", "Test3", "Test1", "Test4", "Test4", "Test2",
"Test3", "Test1")
> NewOrFollowUp = c("New", "New", "New", "New", "New", "Follow-up", "New",
"New", "New", "New")
> data1<-data.frame(Reference,Status,Test,NewOrFollowUp )
> print(data1)
Reference Status Test NewOrFollowUp
1 KRXH Accepted Test1 New
2 KRPT Accepted Test1 New
3 FHRA Rejected Test2 New
4 CZKK Accepted Test3 New
5 CQTN Rejected Test1 New
6 PZXW Accepted Test4 Follow-up
7 SZRZ Rejected Test4 New
8 RMZE Rejected Test2 New
9 STNX Accepted Test3 New
10 TMDW Accepted Test1 New
> xtab(data(Reference,Status,Test,NewOrFollowUp))
```

Exercise: 2

i) Use Two Categorical Variables and Discover the relationships within a dataset

ii) Next, using the xtabs() function, apply two variables from “dataframe1 “, to create a table delineating the relationship between the “Reference” category, and the “Status” category.

iii) Save the file in the name of dataframe2

Exercise: 3

Use the same data frame using three Categorical Variables create a Multi-Dimensional Table

Apply three variables from “dataframe1” to create a Multi-Dimensional Cross-Tabulation of

“Status“, “Gender“, and “Test“.

Exercise: 4
Row Percentages
The R package “tigerstats” is required for the next two exercises.
1) Create an xtabs() formula that cross-tabulates “Status“, and “Test“.
2) Enclose the xtabs() formula in the tigerstats function, “rowPerc()” to display row percentages for “Status” by “Test“.
Exercise 5
Column Percentages
1) Create an xtabs() formula that cross-tabulates “Status“, and “Test“.
2) Enclose the xtabs() formula in the tigerstats function, “colPerc()” to display row percentages for “Status” by “Test“.
Exercise 6
Covariance
i) For the Dataframe1 created from exercise 2 calculate the covariance between Refrence column and Status column
ii) Display the covariance matrix
Exercise 7
Correlation
Find the Correlation between gender and status. what kind of correlation does exist between
the two?
VISUALIZATION IN R
1. Write a program for creating a pie-chart in R using the input vector(21,62,10,53).
Provide labels for the chart as ‘London’, ‘New York’, ‘Singapore’, ‘Mumbai’. Add a title to the chart as ‘city pie-chart’ and add a legend at the top right corner of the chart.
CODE:

```
x <- c(21, 62, 10, 53)
labels <- c("London", "New York", "Singapore", "Mumbai")
png(file = "city_title_colours.jpg")
pie(x, labels, main = "City pie chart", col = rainbow(length(x)))
dev.off()
```

OUTPUT:
2. Create a 3D Pie Chart for the dataset “political Knowledge” with suitable labels,colours and a legend at the top right corner of the chart.
CODE:
OUTPUT:
3. Write a program for creating a bar chart using the vectors H=c(7,12,28,3,41) and M=c(“mar”, “apr”, “may”, “jun”, “jul”). Add a title to the chart as “Revenue chart”.
CODE:

```
H <- c(7,12,28,3,41)
M<-c(“mar”, “apr”, “may”, “jun”, “jul”).
```

```
png(file = "barchart.png")
barplot(H)
dev.off()
```
4. Make a histogram for the “AirPassengers“dataset, start at 100 on the x-axis, and from
values 200 to 700, make the bins 200 wide
CODE:
```
hist(AirPassengers,
main="Histogram for Air Passengers",
xlab="Passengers",
border="blue",
col="green",
xlim=c(100,700),
las=1,
breaks=5)
```
OUTPUT:
5. Create a Boxplot graph for the relation between "mpg"(miles per galloon) and
"cyl"(number of Cylinders) for the dataset "mtcars" available in R Environment.
CODE:
```
input <- mtcars[,c('mpg','cyl')]
print(head(input))
png(file = "boxplot.png")
boxplot(mpg ~ cyl, data = mtcars, xlab = "Number of Cylinders", ylab = "Miles Per Gallon", main = "Mileage Data")
dev.off()
```
OUTPUT:
```
> input <- mtcars[,c('mpg','cyl')]
> print(head(input))
mpg cyl
Mazda RX4 21.0 6
Mazda RX4 Wag 21.0 6
Datsun 710 22.8 4
Hornet 4 Drive 21.4 6
Hornet Sportabout 18.7 8
Valiant 18.1 6
> png(file = "boxplot.png")
> boxplot(mpg ~ cyl, data = mtcars, xlab = "Number of Cylinders", ylab = "Miles Per Gallon", main = "Mileage Data")
> dev.off()
png
2
```