EX.1

```
print(R.version.string)
rm(list = ls())
print(2 + 3)
print(10 - 4)
print(6 * 5)
print(20 / 4)
print(2^5)
v <- c(1, 2, 3, 4, 5)
print(v)
print(getwd())
print(ls())
rm(v)
print(ls())
help(mean)
?mean
library(ggplot2)
q()
```

EX.2

```
library(dplyr)
data <- data.frame(x1 = 1:6, x2 = c(1, 2, 2, 3, 1, 2), x3 = c("F", "B", "C", "E", "A", "D"))
print(data)
arranged_data <- arrange(data, x2)
print(arranged_data)
filtered_data <- filter(data, x2 == 2)
print(filtered_data)
mutated_data <- mutate(data, x4 = x2 * 2)
print(mutated_data)
x3_values <- pull(data, x3)
print(x3_values)
renamed_data <- rename(data, ID = x1)
print(renamed_data)
sampled_data <- sample_n(data, 3)
print(sampled_data)
selected_data <- select(data, x1, x3)
print(selected_data)
```

EX.3

```
data("swiss")
selected_rows <- c(1,2,3,10,11,12,13)
selected_columns <- c("Examination","Education","Infant.Mortality")
swiss_subset <- swiss[selected_rows, selected_columns]
swiss_subset[4, "Infant.Mortality"] <- NA
total_row <- colSums(swiss_subset, na.rm = TRUE)
swiss_subset <- rbind(swiss_subset, Total = total_row)
swiss_subset$Examination_Proportion <- round(swiss_subset$Examination / total_row["Examination"]
print(swiss_subset)
```

EX.4

```
library(dplyr)
hotel_data <- data.frame(Booking_ID = paste0("BKG", 101:110), Booking_Date = as.Date(c("2025-07-
print(hotel_data)
filtered_data <- filter(hotel_data, Children > 1)
print(filtered_data)
mutated_data <- mutate(hotel_data, Total_Guests = Children + Adults)
print(mutated_data)
stay_lengths <- pull(hotel_data, Length_of_Stay)
print(stay_lengths)
renamed_data <- rename(hotel_data, Parking_Spots = Parking_Spaces)
print(renamed_data)
```

EX.5

```
dates <- as.Date(c("2023-Mar-01","2023-Apr-15","2023-May-10","2023-Jun-25","2023-Jul-05"), forma
print(dates)
names <- c("Alice","Bob","Charlie","Diana","Eve")
scores <- c(85,92,78,88,95)
combined1 <- paste(names, "visited on", dates)
combined2 <- paste(names, "scored", scores, "on", dates)
print(combined1)
print(combined2)
```

EX.6

```
students <- read.csv("C:\\Users\\system10\\Documents\\R Studio.program\\students.csv", stringsAs
cat("Rows:", nrow(students), "Columns:", ncol(students), "\n")
print(sapply(students, class))
print(colnames(students))
print(rownames(students))
print(head(students,3))
print(tail(students,2))
print(students[2,3])
print(students[1,])
print(students$Marks)
students$Result <- ifelse(students$Marks >= 75, "Pass", "Fail")
students$Result <- as.factor(students$Result)
students$Grade <- ifelse(students$Marks >= 85, "Distinction", ifelse(students$Marks >= 75, "Firs
students$Grade <- as.factor(students$Grade)
new_student <- data.frame(Name = "Frank", Age = 22, Gender = "Male", Department = "Maths", Marks
students <- rbind(students, new_student)
students$Gender <- as.factor(students$Gender)
students$Department <- as.factor(students$Department)
print(levels(students$Gender))
print(levels(students$Department))
print(levels(students$Grade))
print(summary(students))
print(table(students$Gender))
print(table(students$Department))
print(table(students$Grade))
print(table(students$Result))
```

EX.7

```
x <- c(3,1,4,1,5,9)
print(x[1])
print(x[c(2,4,6)])
print(x[-1])
print(x[-c(2,5)])
print(x[0])
print(x[c(TRUE,FALSE,TRUE,FALSE,TRUE,FALSE)])
print(x[])
print(x[c(1,3,5)])
print(x[c(TRUE,TRUE,FALSE,TRUE,FALSE,TRUE)])
```

EX.8

```
raw_names <- c(" alice smith ", "BOB JOHNSON", "charlie Brown", " diana LEE", "EVAN kelly ")
clean_names <- trimws(raw_names)
lower_names <- tolower(clean_names)
formatted_names <- tools::toTitleCase(lower_names)
split_names <- strsplit(formatted_names, " ")
first_names <- sapply(split_names, `[`, 1)
last_names <- sapply(split_names, `[`, 2)
students_df <- data.frame(FullName = formatted_names, FirstName = first_names, LastName = last_n
print(students_df)
```

EX.9

```
student_names <- c("Anita","Bala","Charan","Divya","Elango","Farah","Ganesh","Hema","Imran","Jay
marks <- c(78,85,67,90,76,85,92,88,70,80)
avg <- mean(marks)
count <- length(marks)
first <- marks[1]
last <- marks[length(marks)]
maximum <- max(marks)
med <- median(marks)
minimum <- min(marks)
get_mode <- function(x){ ux <- unique(x); ux[which.max(tabulate(match(x, ux)))] }
mode_val <- get_mode(marks)
rms <- sqrt(mean(marks^2))
stddev <- sd(marks)
total <- sum(marks)
print(avg)
print(count)
print(first)
print(last)
print(maximum)
print(med)
print(minimum)
print(mode_val)
print(rms)
print(stddev)
print(total)
```

EX.10

```
library(reshape2)
names <- c("Anita","Bala","Charan","Divya","Elango","Farah","Ganesh","Hema","Imran","Jaya")
addresses <- c("Chennai","Madurai","Coimbatore","Trichy","Salem","Erode","Tirunelveli","Vellore"
states <- rep("Tamil Nadu", 10)
df_cbind <- cbind(Name = names, Address = addresses, State = states)
print(df_cbind)
row1 <- c("Anita","Chennai","Tamil Nadu")
row2 <- c("Bala","Madurai","Tamil Nadu")
row3 <- c("Charan","Coimbatore","Tamil Nadu")
df_rbind <- rbind(row1, row2, row3)
colnames(df_rbind) <- c("Name","Address","State")
print(df_rbind)
student_df <- data.frame(ID = 1:10, Name = names, Address = addresses, State = states, stringsAs
melted_df <- melt(student_df, id.vars = "ID")
print(melted_df)
casted_df <- dcast(melted_df, ID ~ variable)
print(casted_df)
```

EX.11

```
customers <- c(45,52,48,60,55,49,62,58,53,47,51,50,65,59,54,56,61,57,63,46)
print(mean(customers))
print(sd(customers))
print(pnorm(55, mean = mean(customers), sd = sd(customers)))
print(pnorm(60, mean = mean(customers), sd = sd(customers)) - pnorm(50, mean = mean(customers),
hist(customers, probability = TRUE, main = "Customer Arrivals Distribution", xlab = "Number of C
curve(dnorm(x, mean = mean(customers), sd = sd(customers)), lwd = 2, add = TRUE)
print(dbinom(7, size = 10, prob = 0.6))
print(pbinom(7, size = 10, prob = 0.6))
print(dpois(5, lambda = 3))
print(ppois(5, lambda = 3))
```

EX.12

```
library(pwr)
pwr.t.test(d = 2.228027, sig.level = 0.05, power = 0.8, type = "two.sample", alternative = "two.
pwr.t.test(d = -2.86972, sig.level = 0.05, power = 0.8, type = "two.sample", alternative = "less
```

EX.13

```
X <- c(10,20,30,40,50,60,70,80,90,100)
Y <- c(15,22,29,35,45,52,60,72,81,95)
correlation <- cor(X, Y)
print(correlation)
reg_model <- lm(Y ~ X)
print(summary(reg_model))
plot(X, Y, main = "Regression and Correlation", xlab = "X", ylab = "Y", pch = 16)
abline(reg_model, col = "red", lwd = 2)
legend("topleft", legend = c("Data Points", "Regression Line"), col = c("black","red"), pch = c(
```

EX.14

```
group1 <- c(12,15,14,16,13)
group2 <- c(14,17,15,16,18)
t.test(group1, group2, var.equal = TRUE)
before <- c(5,6,7,8,5)
after  <- c(6,7,8,9,6)
t.test(before, after, paired = TRUE)
```

EX.15

```
traditional <- c(23,18,30,15,27)
discussion <- c(21,25,20,28,19)
online <- c(30,26,33,24,29)
scores <- c(traditional, discussion, online)
method <- factor(c(rep("Traditional Lecture", length(traditional)), rep("Group Discussion", leng
data <- data.frame(scores, method)
anova_result <- aov(scores ~ method, data = data)
print(summary(anova_result))
kruskal_result <- kruskal.test(scores ~ method, data = data)
print(kruskal_result)
boxplot(scores ~ method, data = data, main = "Exam Scores by Teaching Method", xlab = "Teaching
```

EX.16

```
students <- read.csv("C:\\Users\\system10\\Documents\\R Studio.program\\students.csv")
print(students[2,3])
print(students$Name)
students$Result <- ifelse(students$Score >= 80, "Pass", "Fail")
students$Result <- as.factor(students$Result)
print(students)
new_row <- data.frame(ID = 6, Name = "Fiona", Age = 22, Gender = "F", Score = 90, Result = "Pass
students <- rbind(students, new_row)
print(students)
students$Gender <- as.factor(students$Gender)
print(levels(students$Gender))
print(summary(students))
```

EX.17

```
group <- factor(c("A","A","B","B","A","B","A","B"))
outcome <- factor(c("Yes","No","Yes","No","Yes","Yes","No","No"))
tbl <- table(group, outcome)
print(tbl)
chisq.test(tbl)
```

EX.18

```
employees <- read.csv("employees.csv", stringsAsFactors = FALSE)
high_paid_IT <- subset(employees, Department == "IT" & Salary > 70000)
print(high_paid_IT)
```

```
reduced <- employees[, c("Name","Department","Salary")]
print(reduced)
employees$AnnualBonus <- employees$Salary * 0.10
print(head(employees))
```

EX.19

```
cars_df <- data.frame(mpg = c(21,22,19,24,18,30,27,23,20,25), wt = c(2.8,3.0,3.5,2.6,3.8,2.1,2.4
model <- lm(mpg ~ wt + disp + hp, data = cars_df)
print(summary(model))
new_car <- data.frame(wt = 3.0, disp = 180, hp = 120)
predicted_mpg <- predict(model, new_car)
print(predicted_mpg)
```

EX.20

```
library(tidyverse)
library(broom)
library(car)
library(lmtest)
library(sandwich)
insurance <- read.csv("insurance.csv", stringsAsFactors = FALSE)
insurance <- insurance %>% rename(gender = sex) %>% mutate(gender = factor(gender), smoker = fac
set.seed(123)
train_index <- sample(seq_len(nrow(insurance)), size = 0.8 * nrow(insurance))
train <- insurance[train_index, ]
test  <- insurance[-train_index, ]
M1 <- lm(charges ~ age, data = train)
M2 <- lm(charges ~ age + gender + bmi + smoker + region, data = train)
M3 <- lm(charges ~ age + gender + bmi * smoker + age * smoker + region, data = train)
M4 <- lm(log(charges) ~ age + gender + bmi * smoker + age * smoker + region, data = train)
summary(M1)
summary(M2)
summary(M3)
summary(M4)
coeftest(M2, vcov = sandwich::vcovHC(M2, type = "HC3"))
vif(M2)
bptest(M2)
```