Exercise 6.1: Simple Linear Regression Analysis in R (Height vs. Weight)

Aim:
To perform a simple linear regression analysis in R, predicting weight based on height.

Procedure:

- **Create Dataset:** Use a dataset with two variables: height (independent variable) and weight (dependent variable).
- **Fit Linear Model:** Fit a linear model using the lm() function.
- **View Model Summary:** Analyze the summary of the regression model to understand the relationship between height and weight.
- **Plot Regression Line:** Visualize the regression line on a scatter plot.
- Predict Weight for a height of 172.5 cm

Program:

```
# create dataset
height <- c(150, 152, 160, 165, 170, 175, 180, 185, 190, 195)
weight <- c(40, 45, 52, 55, 60, 65, 67, 70, 73, 75)
# Linear regression
model <- lm(weight ~ height)
cat("The summary of Linear regression is \n")
Print(summary(model))
# plotting
Plot(height, weight, main = "height vs weight", xlab = "height(cm)"
     ylab = "weight(kg)", Pch = 16, col = "red")
abline (model, col = "green")
# Prediction
Predicted_output <- Predict(model, data.frame(height = 172.5))
cat ("Predicted output for height 172.5cm is", Predicted_
output")
```

Output:

The Summary of linear regression is

call:

lm (formula=weight ~ height, data=data)

Residuals:

Min	1Q	Median	3Q	Max
-3.3571	-0.5063	0.1940	1.0125	2.6757

coefficients:

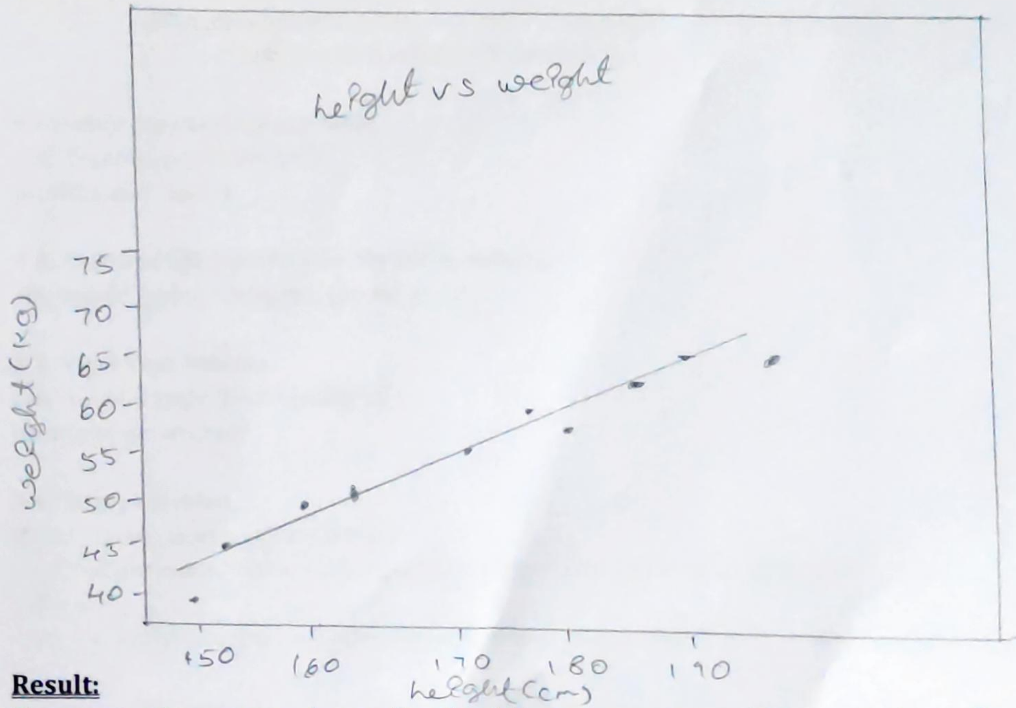
	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	-70.44595	7.09251	-9.932	8.93e-06
height	0.75869	0.04104	18.488	7.55e-08

Residual standard error: 1.914 on 8 degrees of freedom

Multiple R-Squared: 0.9771, Adjusted R-Squared: 0.9743.

F-Statistic: 341.8 on 1 and 8 DF, P-value: 7.55e-08

Predicted output for height 172.5cm is 60.42761



Result:

The exercise demonstrates how to fit a simple linear regression model, interpret the results, and make predictions based on the model.

Exercise 7.1: Chi-Square Test for Independence in R

Aim:

To perform a chi-square test on a contingency table to determine if two categorical variables are independent.

Procedure:

- Create Contingency Table: A survey of preference for different types of cuisine (Italian, Chinese) by age group (Young, Middle-aged)

	Cuisine Preference	
Age Group	Italian	Chinese
Young	25	30
Middle-aged	20	35

- Display the contingency table
- Perform Chi-Square Test for Independence
- View the Chi-Square Test Results
- Interpret the result

Programe:

```
# create contigency table
cuisine_age <- matrix(c(25,30,20,35), nrow=2, byrow=TRUE,
                      dimnames=list("age group"=c("Young",
                      "middle-aged"), "cuisine"=
cat("contigency table is\n")          c("Italian", "chinese")))
print(cuisine_age)
# chi-square test
chi_square_test <- chisq.test(cuisine_age)
cat("The result of chi_square test is\n")
print(chi_square_test)
if(chi_square_test$p.value < 0.05){
  cat("conclusion: The variables are not independent
                 (reject the null hypothesis), \n")
} else {
```

cat ("conclusion : The variables are independent (fail to reject the null hypothesis)\n")

Output:

The contingency table is

	cuisine	
age group	Italian	chinese
young	25	30
middle-aged	20	35

The result of chi square test is

Pearson's chi-squared test with Yates' continuity correction.

data : cuisine_age

x-squared = 0.60171 df = 1 P-value = 0.4379

conclusion : The variables are independent (fail to reject the null hypothesis).

Result:

The chi-square test for independence has been performed, and the result shows whether there is a significant association between gender and sport preference.

Exercise 8.1: One-Way ANOVA in R Using Weight and Diets

Aim:

To perform a one-way ANOVA on a dataset comparing the average weight of individuals on three different diets.

Procedure:

- **Create a Dataset:** Example: Weights of individuals on three different diets (Diet A, Diet B, and Diet C)
  weights <- c(68, 72, 65, 70, 74, 60, 63, 67, 69, 64, 76, 78, 71, 73, 75)
  diets <- factor(c(rep("Diet A", 5), rep("Diet B", 5), rep("Diet C", 5)))
- Combine into a data frame
- Display the dataset
- Perform One-Way ANOVA
- View the ANOVA Summary
- Interpret the result

Program:

```
#creating a dataset
weights<-c(68,72,65,70,74,60,63,67,69,64,76,78,71,73,75)
diets <-factor(c(rep("Diet A",5), rep("Diet B",5), rep("Diet C",5)))
# data frame
data <-data.frame(weights,diets)
Print(data)
# anova_result
anova_result<-aov(weights~diets, data=data)
Print(summary(anova_result))

# significance
P_value <-Summary(anova_result)[[1]][["Pr(>F)"]]
[1]
```

```
# test significance
if (p.value < 0.05){
  cat("conclusion : There is a significant difference between
      the group means(reject the null hypothesis)\n")
}else{
   cat("conclusion : There is no significant difference between
  the group means(fail to reject the null hypothesis)\n")
  }
```

Output

	weight	diet
1	68	diet A
2	72	diet A
3	65	diet A
4	70	diet A
5	74	diet A
6	60	diet B
7	63	diet B
8	67	diet B
9	69	diet B
10	64	diet B
11	76	diet c
12	78	diet c
13	71	diet c
14	73	diet c
15	75	diet c

The one way anova result is

	Df	Sumsq	Meansq	Fvalue	Pr(>F)
diets	2	250.1	125.1	11.8	0.0047
Residuals	12	127.2	10.6		

conclusion : There is significance difference between two groups (reject the null hypothesis)

Result:

This program demonstrates how to perform a one-way ANOVA in R, allowing you to analyze whether the means of different groups are statistically different.

Exercise 9.1: Two-Sample t-Test in R

Aim:

To perform a two-sample t-test on a dataset comparing the means of two independent samples.

Procedure:

- **Create a Dataset**: Define the dataset with scores from two independent groups.
- **Perform the Two-Sample t-Test**: Use the t.test() function to conduct the test.
- **Interpret the Results**: Analyze the p-value to determine if there is a significant difference between the two groups.

Program:

```
# create a Dataset
 group1 <- c(85,90,88,92,87)
 group2 <- c(78,82,80,84,79)
 cat("group1 score\n")
 Print(group1)
 cat("group 2 score\n")
 Print(group2)
 t_test<-t.test(group1,group2)
 cat("two sample t_test result\n")
 print(t_test)
# significant testing
 P_value <- t_test$p.value
 if (P_value < 0.05){
   cat("conclusion: There is significant difference between
            two groups (reject null hypothesis)\n")
 } else{
   cat("conclusion: There is no significant difference
   between two groups (fail to reject null hypothesis)\n")
 }
```

Output:

group 1 Score
[1] 85 90 88 92 87
group-2 score
[2] 78 82 80 84 79
Two Sample t-test result
Welch two Sample t-Test
data : group1 and group2
t = 4.8189, df = 7.8965, P-value = 0.001373
alternative hypothesis: true difference in means is not equal to 0 95 percent confidence interval:
4.058871 11.541129
Sample estimates:
mean of x mean of y
88.4 80.6
conclusion: there is significant difference between two groups (reject null hypothesis).

Result:

This program demonstrates how to perform a two-sample t-test in R to compare the means of two independent samples. The analysis reveals whether the means of the two groups are significantly different based on the given data.

Exercise 10.1: Plotting Gamma Distribution in R

Aim:

To demonstrate the plotting of a Gamma distribution, which is used to model the time until the occurrence of an event, for multiple shape parameters.

Procedure:

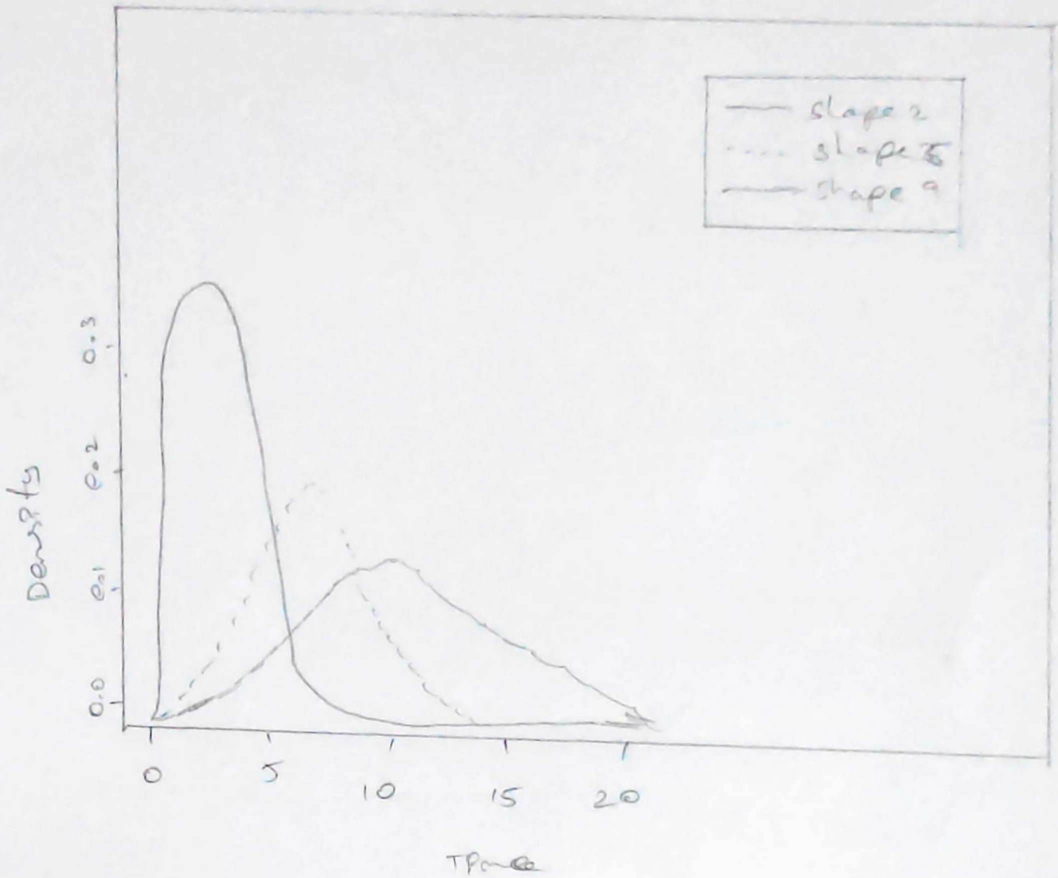
- **Create Data for Gamma Distribution**: Generate data using the dgamma() function for different shape parameters.
- **Plot Gamma Distribution**: Visualize the distribution using the plot() function.

Program:

```
#create data
 x <- seq(0,20, length=100)
gamma_shape1 <-dgamma(x, shape=2, scale=1)
gamma_shape2 <- dgamma(x, shape=5, scale=1)
gamma_shape3 <- dgamma(x, shape=9, scale=1)
plot(x,gamma_shape1, type="l", col="blue", lwd=2,
     main="Gamma Distribution with Different Shape
          Paraments",
          xlab="Time",
          ylab="Density")
lines(x,gamma_shape2, col="red", lwd=2)
lines(x,gamma_shape3, col="green", lwd=2)
legend("topright", legend=c("shape=2", "shape=5",
                            "shape=9"),
       col=c("blue", "red", "green"), lwd=2)
```

Output:

Gamma Distribution with different shape Parameters



Result:

This program successfully demonstrates how to plot the Gamma distribution for different shape parameters, showing how the shape affects the spread and peak of the distribution.