MMU®

MULTIMEDIA UNIVERSITY

TDS2201

Assignment 2

Group 5

Name	StudentID
Haziq Zairul	1221303388
Tan Xin Thong	1211104274
Chin Wei Ling	1211104398
Lee Jia Ying	1221303972

Introduction:

In the pursuit of understanding the factors influencing human reaction times, a comprehensive study was conducted in a town with a population of approximately 18,000. This research aimed to explore the impact of age, gender, hand dominance, and physical activity frequency on the reaction times of individuals. A sample of 100 residents was meticulously selected to represent the town's demographic diversity. The ensuing analysis, utilizing R for statistical computations and graphical representations, seeks to shed light on the intricate relationships between these variables and reaction time. This report presents the findings from the sample data, offering insights into the average reaction time of the citizens, the differences in reaction times between left-handed and right-handed individuals, and the construction of a regression model to predict reaction times based on the identified factors.

Data Description:

- **Study Objective**: Investigate the influence of **age, gender, hand dominance**, and **physical activity frequency** on reaction times.

- **Data Collection**: Sample of **100 individuals** from a town of approximately **18,000 people**.

- **Variables**: Includes **age** (years), **gender** (male/female), **hand dominance** (left/right), **physical activity frequency** (daily/weekly/occasionally/rarely), and **reaction time** (milliseconds).

- **Analysis Tools**: Utilization of **R** for statistical analysis and plotting, with emphasis on clear labeling.

Analysis & Results:

Confidence Interval Analysis

- **95% CI for Average Reaction Time**: Calculated using the sample mean and standard deviation, resulting in an interval that may or may not support the researcher's belief of 0.28 seconds.
- **Sample Size Calculation**: Determined for 90% confidence that the population mean estimate is within 0.05 of the true mean, using the formula for margin of error.

1) The researcher believes that the average reaction time of all citizen in this town is 0.28 seconds.

a) Calculate a 95% confidence interval for the average reaction time.

```
# Assigning variables from the dataset
> Physical_activity_frequency <- data$Physical.Activity.Frequency
> Age <- data$Age
> Gender <- data$Gender
> Hand_dominance<-data$Hand.Dominance
> Reaction_times <- data$Reaction.Time..ms.
>
> # Calculate mean and standard deviation
> mean_rt <- mean(Reaction_times)
> sd_rt <- sd(Reaction_times)
> n <- length(Reaction_times)
>
> # Margin of error for a 95% confidence interval
> alpha <- 0.05
> t_value <- qt(1 - alpha/2, df = n - 1)
> margin_of_error <- t_value * (sd_rt / sqrt(n))
> margin_of_error
[1] 7.891038
>
> # Confidence interval
> lower_bound <- mean_rt - margin_of_error
> upper_bound <- mean_rt + margin_of_error
> conf_interval <- c(lower_bound, upper_bound)
[1] 285.309 301.091
```

b) Do you agree with the researcher? Justify your answer based on your confidence interval in part (a).

To determine if we agree with the researcher's belief that the average reaction time of all citizens in the town is 0.28 (280 ms) seconds, we need to check if this value falls within our calculated 95% confidence interval of 285.309 to 301.091 ms. Since the researcher's belief of 0.28 seconds is not within the confidence interval, this suggests that it is unlikely to be true average reaction time based on our sample data. Therefore, we might not fully agree with the researcher. However, it's essential to note that the confidence interval provides an estimate of where the true population mean is likely to be, and it's possible that the researcher's belief is close to the true value.

c) Calculate the sample size if we want to be 90% confident that the estimate of population mean is off by at most 0.05.

```
> # Calculate the sample size for a 90% confidence level
> confidence_level <- 0.90
> margin_of_error <- 0.05
> z_value <- qnorm((1 + confidence_level) / 2)
> sample_size <- (z_value * sd_rt / margin_of_error)^2
> sample_size
[1] 1711609
```

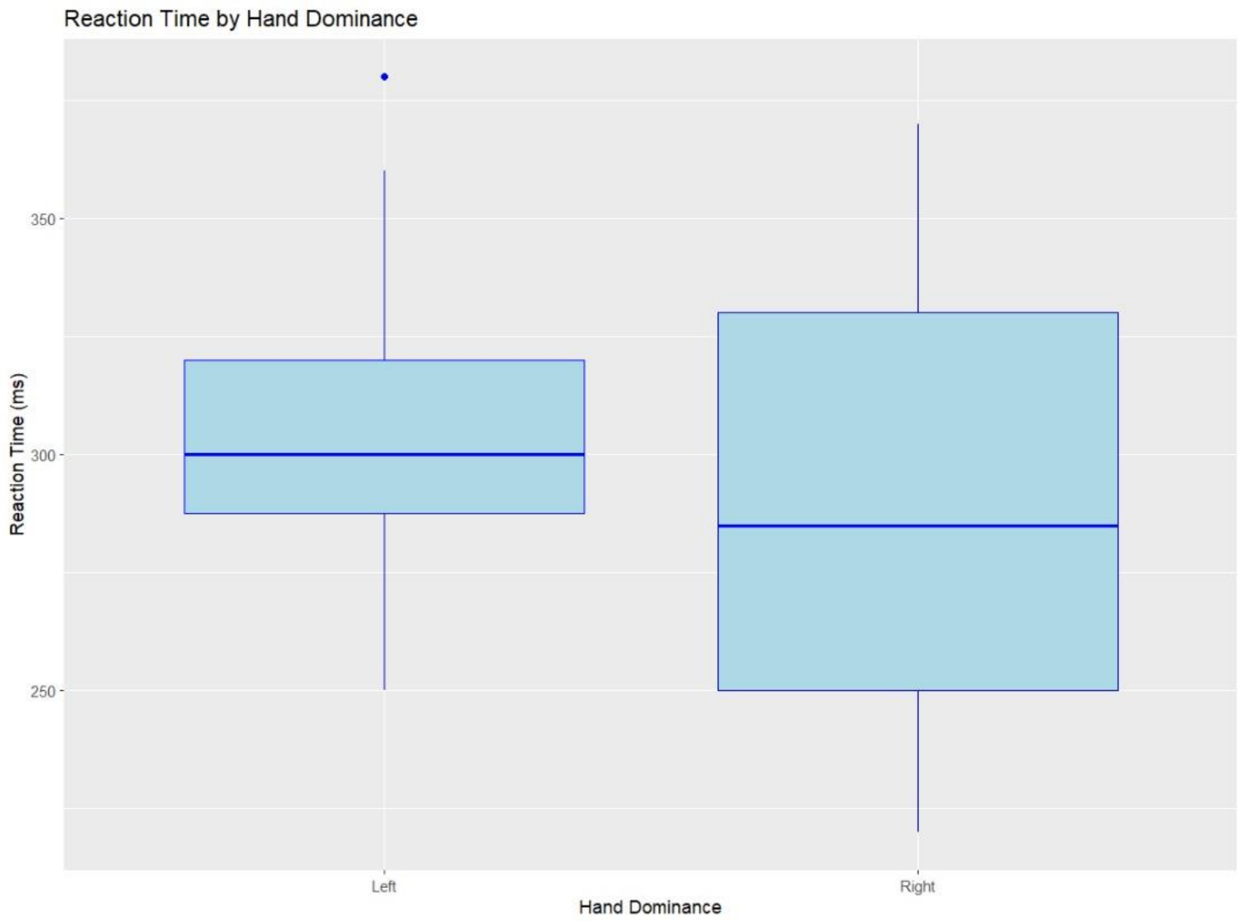
Hand Dominance Comparison

- **Data Exploration**: Visualized with a box plot to compare reaction times between left-handed and right-handed citizens, revealing any significant differences.
- **Confidence Interval for Variability**: A 97% confidence interval calculated to assess if there's a difference in variability of reaction times between the two groups.
- **Hypothesis Testing**: Performed to determine if there's a statistically significant difference in average reaction times, using a 3% significance level.

2) Is there a different in the reaction time for left-handed and right-handed citizen on average? You should

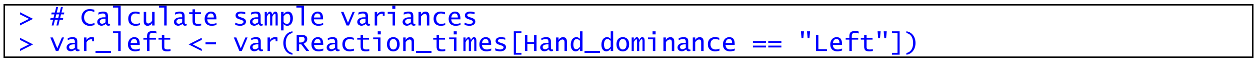
a) Explore the data with appropriate plot. Comment on your plot and answer the question.

```
> # Load ggplot2 package
> library(ggplot2)

> # Plot to explore reaction time by hand dominance
> ggplot(data, aes(x = Hand_dominance, y = Reaction_times)) +
+    geom_boxplot(fill = "lightblue", color = "blue") +
+    labs(title = "Reaction Time by Hand Dominance",
+         x = "Hand Dominance",
+         y = "Reaction Time (ms)")
```



- The minimum reaction time(ms) for right-handed citizens is smaller than left-handed citizens. This difference suggests that, on average, right-handed citizens have a quicker minimum reaction time compared to left-handed citizens.
- The medium value of left-handed citizens is higher than right-handed citizens, which indicates that left-handed citizens have slower reaction times on average compared to right-handed citizens.
- The length of right-handed citizen box is wider than left-handed citizen. This indicates that right-handed citizens have greater variability in reaction times.
- The higher maximum time of right-handed citizens compares to left-handed citizens means that, on average, some right-handed citizens may experience longer reaction times than left-handed citizens in extreme situations.
- There is an outlier in left-handed citizen's boxplot suggesting there is at least one left-handed citizen with a reaction time significantly higher than most left-handed citizens.

b) Calculate an appropriate confidence interval to determine if left-handed and right-handed citizen have different variability in their reaction time. Use a 97% confidence level. You must write all the steps clearly



```
> # Calculate sample variances
> var_left <- var(Reaction_times[Hand_dominance == "Left"])
> var_right <- var(Reaction_times[Hand_dominance == "Right"])
>
> # Calculate the F-statistic
> F_statistic <- var_left / var_right
> F_statistic
[1] 0.462781

> # Define degrees of freedom
> df_left <- length(Reaction_times[Hand_dominance == "Left"]) - 1
> df_right <- length(Reaction_times[Hand_dominance == "Right"]) - 1

> # Calculate critical values
> alpha <- 0.03
> F_critical_lower <- qf(alpha / 2, df_left, df_right)
F_critical_lower
[1] 0.4879511
> F_critical_upper <- qf(1 - alpha / 2, df_left, df_right)
F_critical_upper
[1] 1.892772


> # Calculate confidence interval
> confidence_interval<- c((F_statistic/ F_critical_upper), F_statistic/
F_critical_lower)
> confidence_interval
[1] 0.2444990 0.9484167
```

c) Construct an appropriate hypothesis test to determine if there a different in the reaction time for left-handed and right-handed citizen on average. Use a 3% level of significance. You must write all the steps clearly.

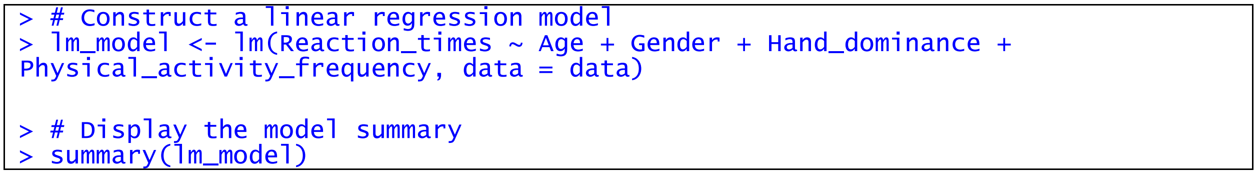
```
> t_test <- t.test(Reaction_times ~ Hand_dominance, data = data)
> t_test


      Welch Two Sample t-test


data:  Reaction_Time by Hand_Dominance
t = 1.9134, df = 85.294, p-value = 0.05906
alternative hypothesis: true difference in means between group Left and group
Right is not equal to 0
95 percent confidence interval:
 -0.5526695 28.8247283
sample estimates:
 mean in group Left mean in group Right
           302.8125            288.6765

> t_stat <- t_test$statistic
> t_stat
       t
1.913363
> p_value <- t_test$p.value
> p_value
[1] 0.05905625
```

Regression Model Construction

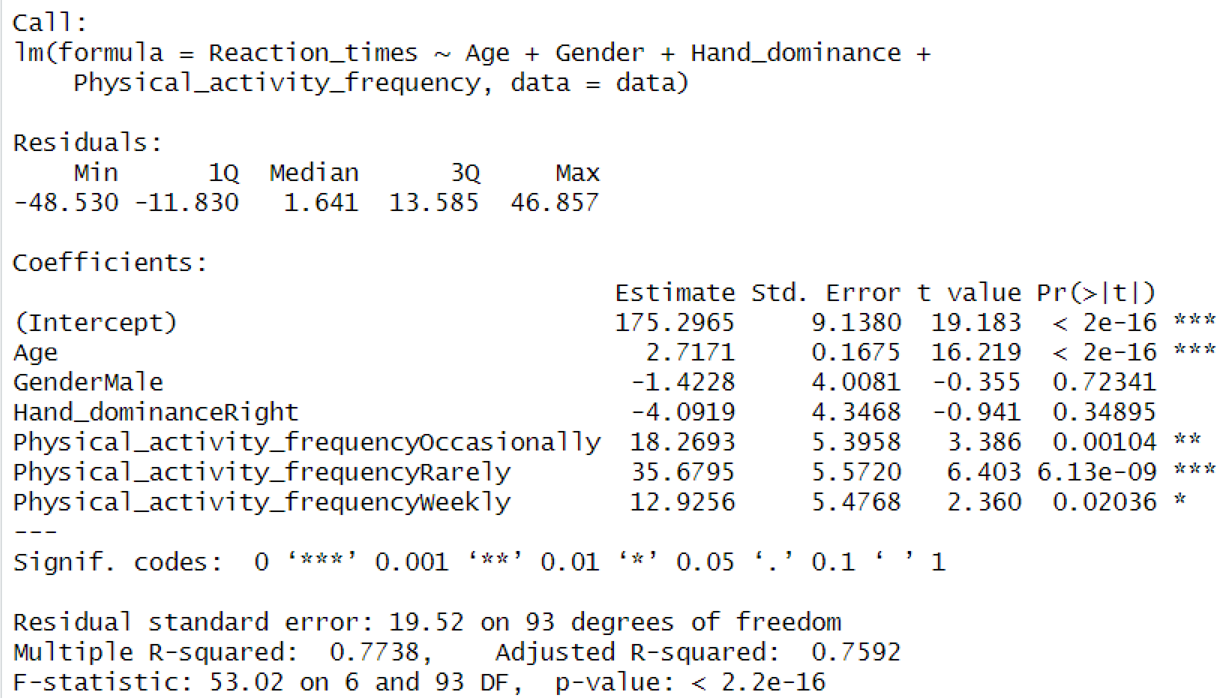
- **Linear Regression Model**: Developed to predict reaction time based on age, gender, hand dominance, and physical activity frequency, with coefficients representing the relationship strength.
- **Age Impact on Reaction Time**: Analyzed how reaction time changes with an additional 10 years in age, using the regression coefficient.
- **ANOVA Test**: Conducted to test the overall significance of the regression model.
- **Model Appropriateness**: Evaluated using residual plots to check for any patterns that might indicate model inadequacies.

3) The researcher wants to build a regression model for respond time, with the other 4 variables in the dataset as explanatory variables.

a) Construct a linear regression model for respond time on the 4 variables and write your model.

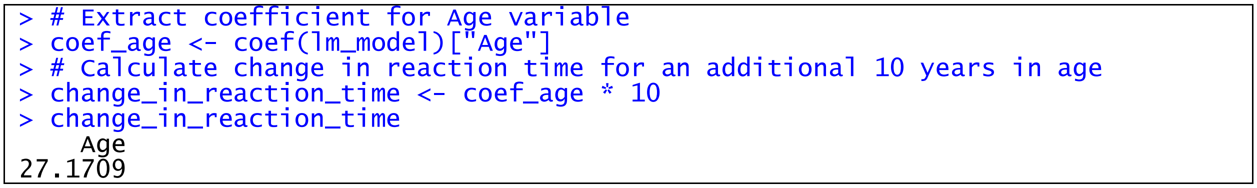


```
> # Construct a linear regression model
> lm_model <- lm(Reaction_times ~ Age + Gender + Hand_dominance +
Physical_activity_frequency, data = data)

> # Display the model summary
> summary(lm_model)
```



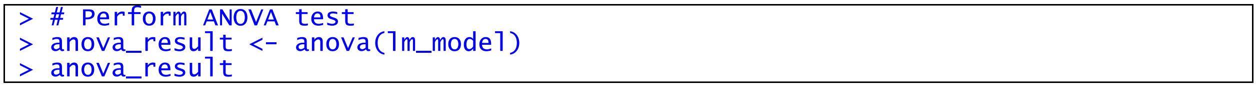
```
Call:
lm(formula = Reaction_times ~ Age + Gender + Hand_dominance + 
    Physical_activity_frequency, data = data)

Residuals:
    Min      1Q  Median      3Q     Max 
-48.530 -11.830   1.641  13.585  46.857 

Coefficients:
                                       Estimate Std. Error t value Pr(>|t|)    
(Intercept)                            175.2965     9.1380  19.183  < 2e-16 ***
Age                                      2.7171     0.1675  16.219  < 2e-16 ***
GenderMale                              -1.4228     4.0081  -0.355  0.72341    
Hand_dominanceRight                     -4.0919     4.3468  -0.941  0.34895    
Physical_activity_frequencyOccasionally 18.2693     5.3958   3.386  0.00104 ** 
Physical_activity_frequencyRarely       35.6795     5.5720   6.403 6.13e-09 ***
Physical_activity_frequencyWeekly       12.9256     5.4768   2.360  0.02036 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 19.52 on 93 degrees of freedom
Multiple R-squared:  0.7738,    Adjusted R-squared:  0.7592 
F-statistic: 53.02 on 6 and 93 DF,  p-value: < 2.2e-16
```

b) How is the reaction time change for an additional of 10 years in age? Show all the steps of your calculation.

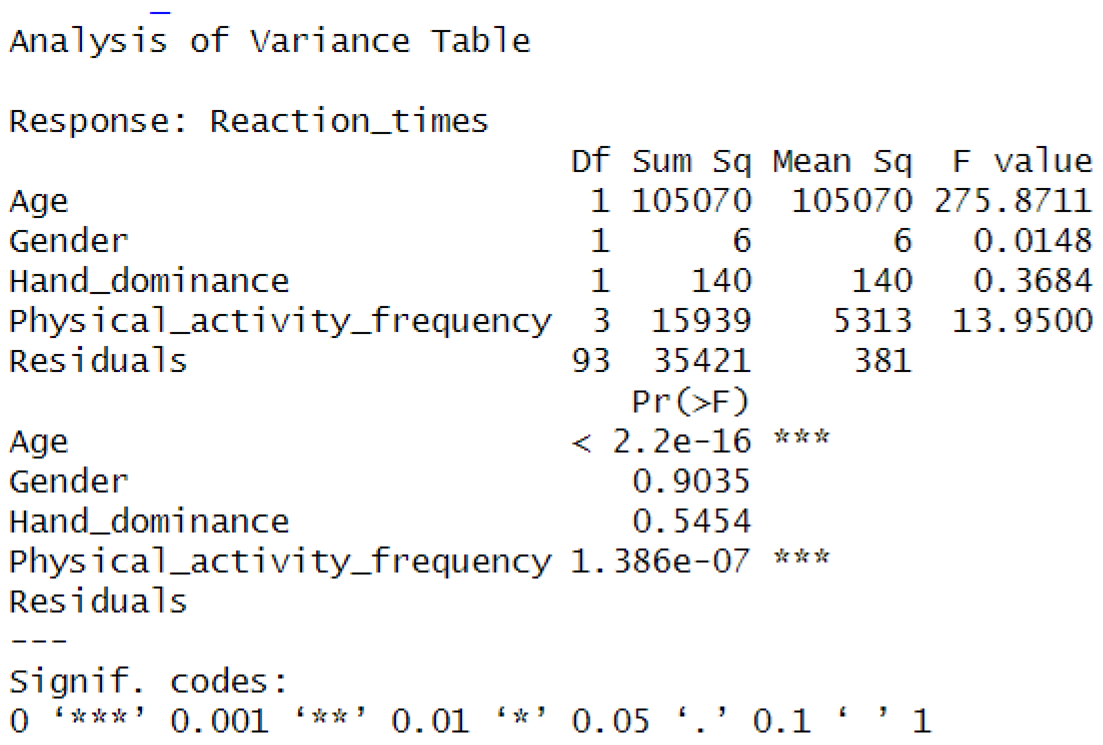


```
> # Extract coefficient for Age variable
> coef_age <- coef(lm_model)["Age"]
> # Calculate change in reaction time for an additional 10 years in age
> change_in_reaction_time <- coef_age * 10
> change_in_reaction_time
    Age 
27.1709 
```

c) Test the significance of your model with the ANOVA approach. You must write all the steps clearly.



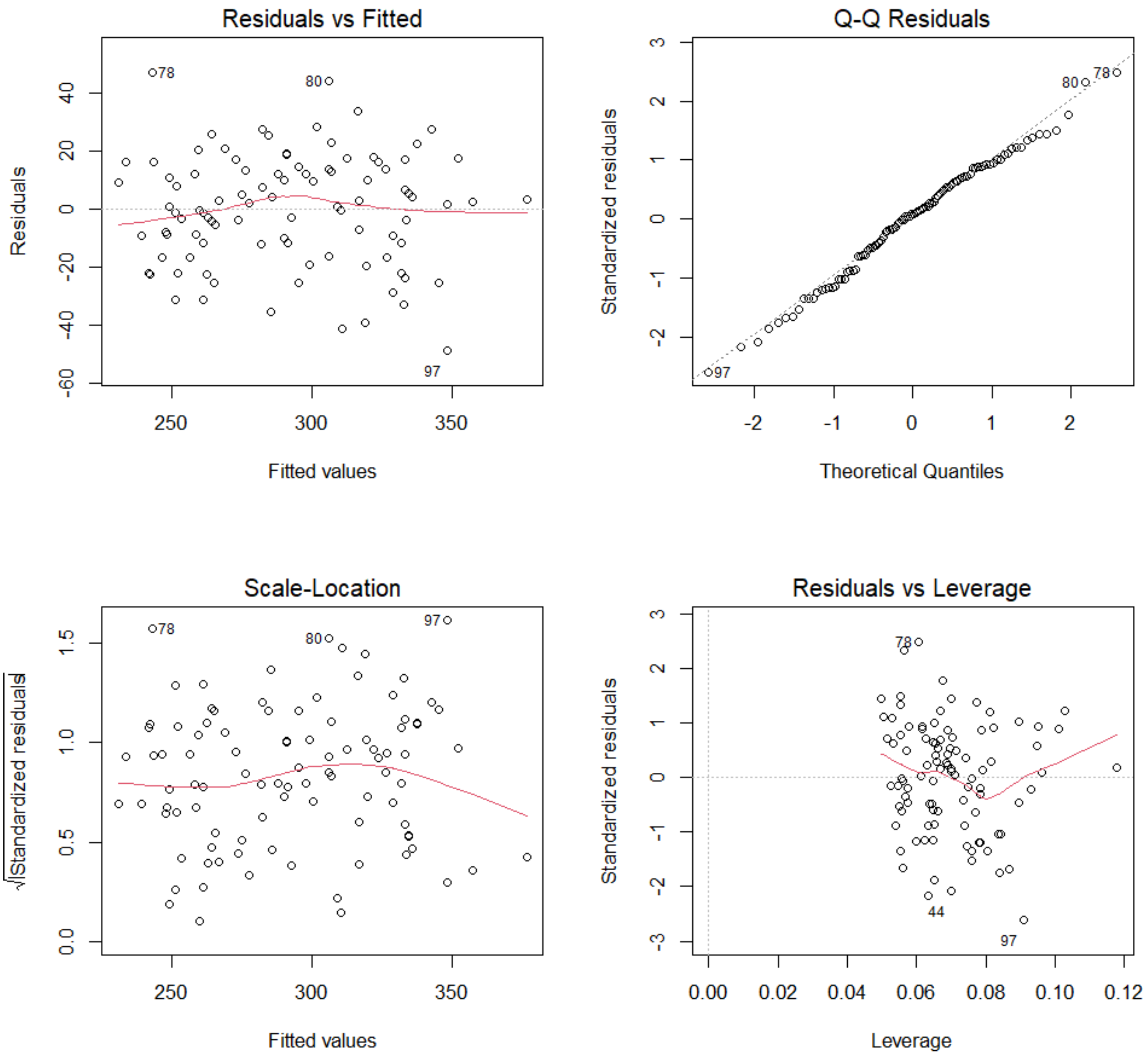
```
> # Perform ANOVA test
> anova_result <- anova(lm_model)
> anova_result
```



```
Analysis of Variance Table

Response: Reaction_times
                           Df Sum Sq Mean Sq  F value
Age                         1 105070  105070 275.8711
Gender                      1      6       6   0.0148
Hand_dominance              1    140     140   0.3684
Physical_activity_frequency 3  15939    5313  13.9500
Residuals                  93  35421     381
                               Pr(>F)
Age                         < 2.2e-16 ***
Gender                         0.9035
Hand_dominance                 0.5454
Physical_activity_frequency 1.386e-07 ***
Residuals
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

d) Verify if the linear regression model is appropriate for your dataset. You must justify your answer based on appropriate plot.

```
> # Plot diagnostic plots
> par(mfrow = c(2, 2))
> plot(lm_model)
```



- **Residual vs Fitted plot**: The plot shows residuals spread randomly around the horizontal line (residual = 0), indicating the relationship between the predictors (age, gender, hand dominance and physical activity frequency) and the response variable (reaction time) is approximately near. There is no clear pattern, which suggests that the assumption of linearity is reasonable. However, some points (e.g., observation 78 and 97) may be potential outliers, suggesting they have large residuals compared to others.
- **Q-Q Residuals**: The residuals mostly fall along the 45-degree reference line, indicating that the residuals are approximately normally distributed. This is a good sign for the normality assumption. Some deviations at the tails suggest minor deviations from normality, but they are not severe enough to invalidate the assumption.
- **Scale-Location Plot:** The plot shows residuals scattered randomly without a clear pattern. The red line that represents the mean of squared residual is relatively flat, indicating that the

variance of residuals is approximately constant across different levels of fitted values. This supports the assumption of homoscedasticity. The homoscedasticity supports the assumption that variability of residuals is uniform across the range of predicted values.

- **Residuals vs Leverage Plot**: Most residuals are within the Cook's distance lines, indicating that there are no highly influential points. Observations like 78, 80, and 97 might have slightly higher leverage, suggesting that they have more influence on the outcomes, but they do not significantly affect the model performance.

Conclusion:

- **Study Objective**: Investigate the influence of age, gender, hand dominance, and physical activity frequency on reaction times.

- **Data Analysis**: Utilize R for statistical analysis, including confidence intervals, hypothesis testing, and regression modeling.
- **Key Tasks**:
  - Calculate a 95% confidence interval for average reaction time and discuss agreement with the researcher’s belief.
  - Explore differences in reaction times between left-handed and right-handed individuals using plots and confidence intervals.
  - Construct and test a linear regression model for reaction time with the given variables and assess the model’s appropriateness.