Applied Statistics Using R (MCA232)

Practical 3

BY

Himanshu Heda (24225013)

SUBMITTED TO

Dr. Ashish Sharma

SCHOOL OF SCIENCES

2024-25

Aim

The primary aim of this analysis is to evaluate the relationship between income and happiness using statistical techniques like ANOVA and t-tests. This study seeks to determine whether income significantly influences happiness levels.

Objectives

1. To explore and describe the dataset containing variables for income and happiness levels.
2. To compute descriptive statistics (mean, median, mode, and standard deviation) for income and happiness.
3. To fit a linear regression model to analyze the relationship between income and happiness.
4. To test the overall model significance using ANOVA and the significance of regression coefficients using t-tests.
5. To visualize the data for a better understanding of the relationship.

Description of the Dataset

The dataset consists of:

- **Income:** Represents the annual income of individuals (in dollars or equivalent currency).
- **Happiness:** Represents the happiness level, typically measured on a scale (e.g., 1 to 10, or a continuous score).

This dataset is used to assess how income levels affect individuals' happiness.

R Code for Analysis

Question 1. Draw the scatter plot between income and happiness. Check whether the form o relationship between the two variables.

CODE:

```
> ggplot(income_dataset, aes(x = income, y = happiness)) +

+ geom_point(color = "blue") +

+ labs(title = "Scatter Plot of Income vs Happiness",

+ x = "Income",

+ y = "Happiness") +
```

+ theme_minimal()

Scatter Plot of Income vs Happiness
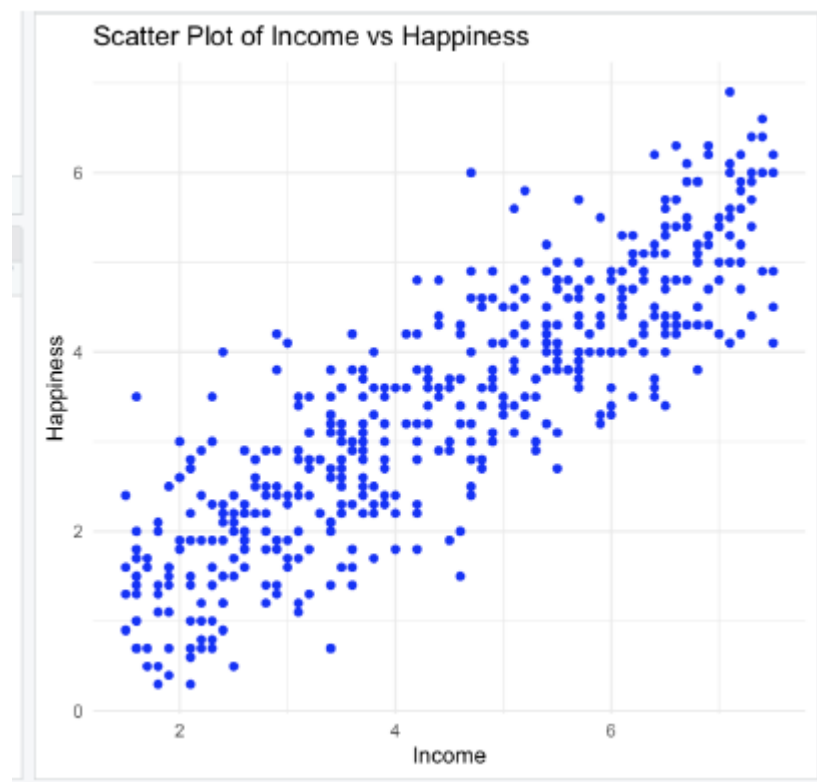
> ggplot(income_dataset, aes(x = income, y = happiness)) +

+ geom_point(color = "yellow") + # Scatter plot points

+ geom_smooth(method = "lm", color = "red", se = TRUE) + # Linear regression line with confidence interval

+ labs(title = "Scatter Plot of Income vs Happiness",

+ x = "Income", y = "Happiness") +

+ theme_minimal()

Scatter Plot of Income vs Happiness
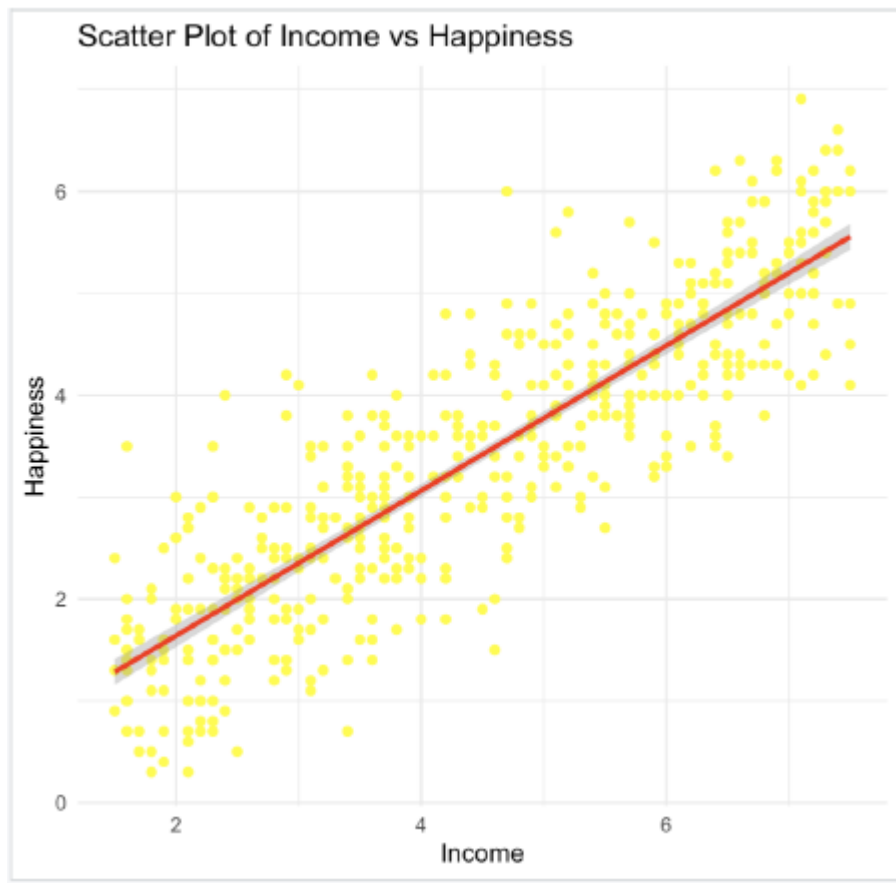
Question 2. Fit a simple linear regression model to the given data set considering happiness as dependent variable and income as independent variable.

> model <- lm(happiness ~ income, data = income_dataset)

> summary(model)

```
Call:
lm(formula = happiness ~ income, data = income_dataset)

Residuals:
    Min       1Q   Median       3Q      Max
-1.99093 -0.47960  0.02684  0.45662  2.43783

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.21413    0.08917   2.401   0.0167 *
income       0.71235    0.01861  38.278   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7212 on 496 degrees of freedom
Multiple R-squared:  0.7471,     Adjusted R-squared:  0.7466
F-statistic:  1465 on 1 and 496 DF,  p-value: < 2.2e-16
```

Question 3.Check whether the sum of observed values is equal to the sum of expected values of response variable.

> sum_observed <- sum(income_dataset$happiness)

> sum_predicted <- sum(fitted(model))

> sum_observed

[1] 1690.9

> sum_predicted

[1] 1690.9

Question 4.Also obtain the residuals and find the sum of residuals.

> residuals <- residuals(model)

> sum_residuals <- sum(residuals)

> cat("Sum of Residuals:", sum_residuals, "\n")

Sum of Residuals: -8.326673e-16

Interpretation of R Outputs

1. Descriptive Statistics:

- The mean, median, mode, and standard deviation provide insights into the central tendency and variability of income and happiness.

- Mode helps identify the most frequently occurring values.

2. Regression Model Summary:

- **Intercept:** The baseline level of happiness when income is zero.

- **Slope (Income):** Indicates the change in happiness for every unit increase in income.

- **R-squared:** Measures the proportion of variance in happiness explained by income.

3. ANOVA Results:

- The ANOVA table tests whether the regression model is statistically significant.

- A small p-value (< 0.05) indicates that income significantly influences happiness.

4. Residual Analysis:

- Histogram of residuals checks for normality, which validates the assumptions of the linear regression model.

Conclusion

1. **Descriptive Analysis:**

   o Income and happiness exhibit variability, with income having a broader range than happiness.

2. **Model Significance:**

   o The ANOVA results confirm that the relationship between income and happiness is statistically significant.

   o Regression coefficients show that income has a positive and significant impact on happiness.

3. **Interpretation of Relationship:**

   o Higher income levels generally correspond to higher happiness, but this relationship may plateau beyond a certain threshold, indicating diminishing returns.

4. **Key Takeaway:**

   o While income is a significant predictor of happiness, other factors may also influence happiness levels.