

Applied Statistics Using R

(MCA232)

# Practical 3

***BY***

**Himanshu Heda (24225013)**

**SUBMITTED TO**

**Dr. Ashish Sharma**



**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)) +

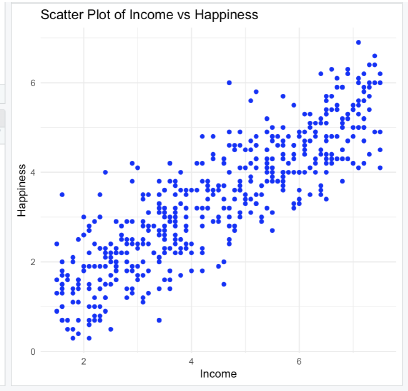
+ geom\_point(color = "blue") +

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

+ x = "Income",

+ y = "Happiness") +

+ theme\_minimal()



> ggplot(income\_dataset, aes(x = income, y = happiness)) +

+ geom\_point(color = "yellow") + # Scatter plot points

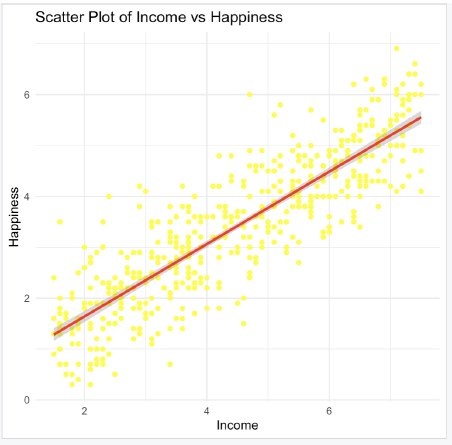
+ geom\_smooth(method = "lm", color = "red", se = TRUE) + # Linear regression line

with confidence interval

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

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

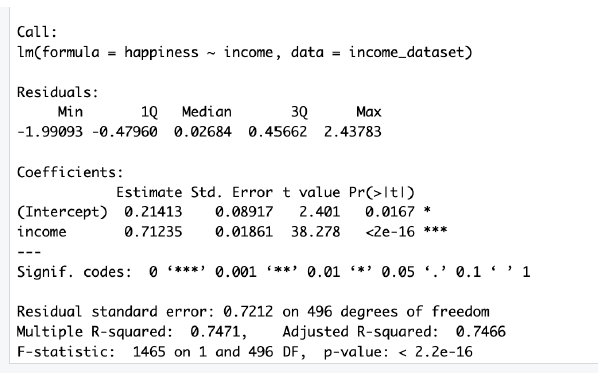
+ theme\_minimal()



**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)



**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:**
   * Income and happiness exhibit variability, with income having a broader range than happiness.
2. **Model Significance:**
   * The ANOVA results confirm that the relationship between income and happiness is statistically significant.
   * Regression coefficients show that income has a positive and significant impact on happiness.
3. **Interpretation of Relationship:**
   * Higher income levels generally correspond to higher happiness, but this relationship may plateau beyond a certain threshold, indicating diminishing returns.
4. **Key Takeaway:**
   * While income is a significant predictor of happiness, other factors may also influence happiness levels.