

Week 9 Statistical Analysis Activity

Shimranbanu

October 13, 2023

Week 9: Statistical Analysis

In this activity, we explore various statistical tests using R. Each scenario is analyzed with appropriate tests and visualizations.

Scenario 1: Independent t-test

Objective: Determine if there is a significant difference in test scores between students who attended a revision class and those who did not.

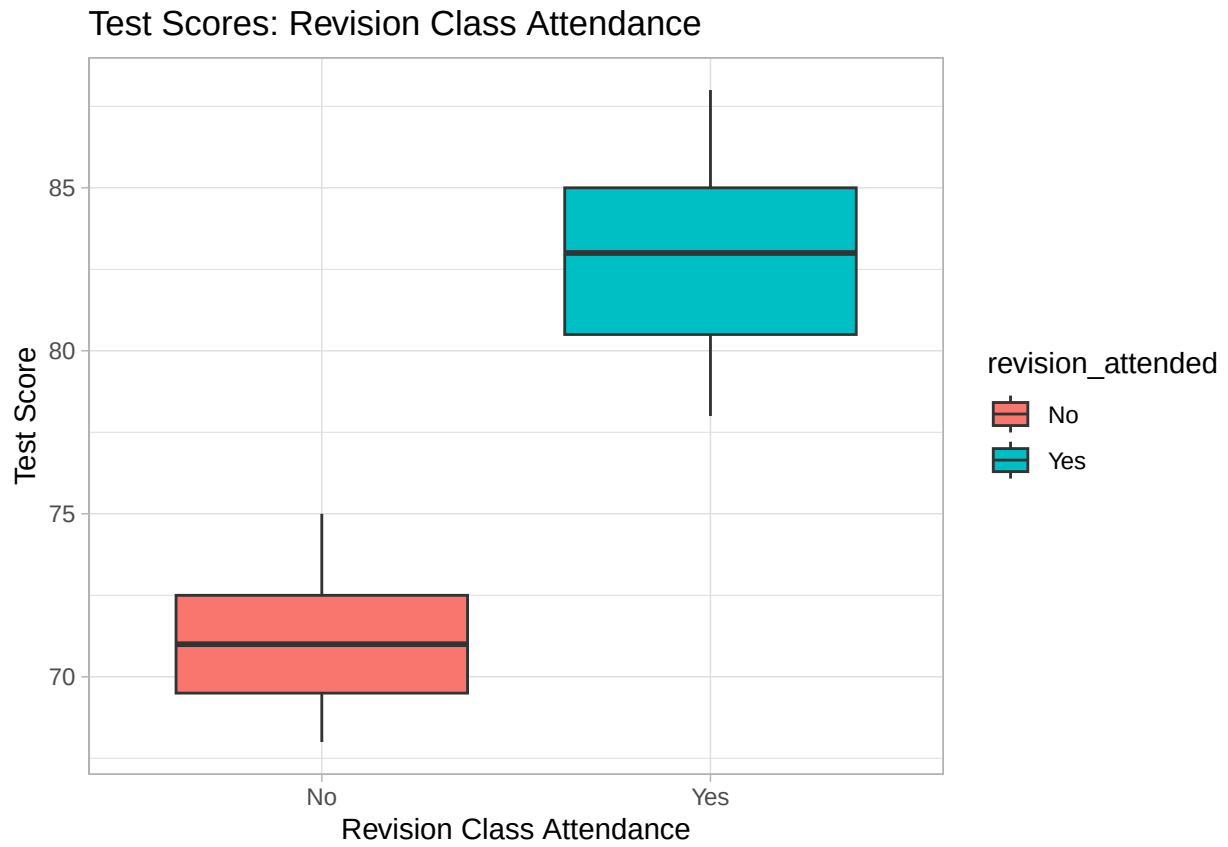
```
# Load necessary library
library(ggplot2)

# Data: Test scores of students
scores_data <- data.frame(
  revision_attended = rep(c("Yes", "No"), each = 15),
  test_score = c(85, 80, 78, 83, 81, 79, 84, 82, 87, 83, 85, 88, 80, 86, 84,
                  72, 70, 69, 73, 75, 68, 71, 74, 70, 69, 72, 71, 68, 73, 70)
)

# Independent t-test
test_result <- t.test(test_score ~ revision_attended, data = scores_data)
test_result

##
## Welch Two Sample t-test
##
## data: test_score by revision_attended
## t = -12.683, df = 25.412, p-value = 1.709e-12
## alternative hypothesis: true difference in means between group No and group Yes is not equal to 0
## 95 percent confidence interval:
## -13.94707 -10.05293
## sample estimates:
## mean in group No mean in group Yes
## 71 83

# Visualization: Boxplot
ggplot(scores_data, aes(x = revision_attended, y = test_score, fill = revision_attended)) +
  geom_boxplot() +
  labs(title = "Test Scores: Revision Class Attendance",
       x = "Revision Class Attendance",
       y = "Test Score") +
  theme_light()
```



Scenario 2: Paired t-test

Objective: Assess changes in blood pressure levels after a 2-week diet plan.

```
# Data: Blood pressure before and after diet
bp_data <- data.frame(
  bp_before = c(140, 145, 138, 142, 141, 139, 144, 143, 140, 137,
                141, 142, 139, 143, 138, 140, 144, 146, 139, 141),
  bp_after = c(130, 132, 129, 134, 133, 130, 131, 132, 129, 128,
               131, 132, 130, 133, 129, 130, 134, 135, 130, 132)
)

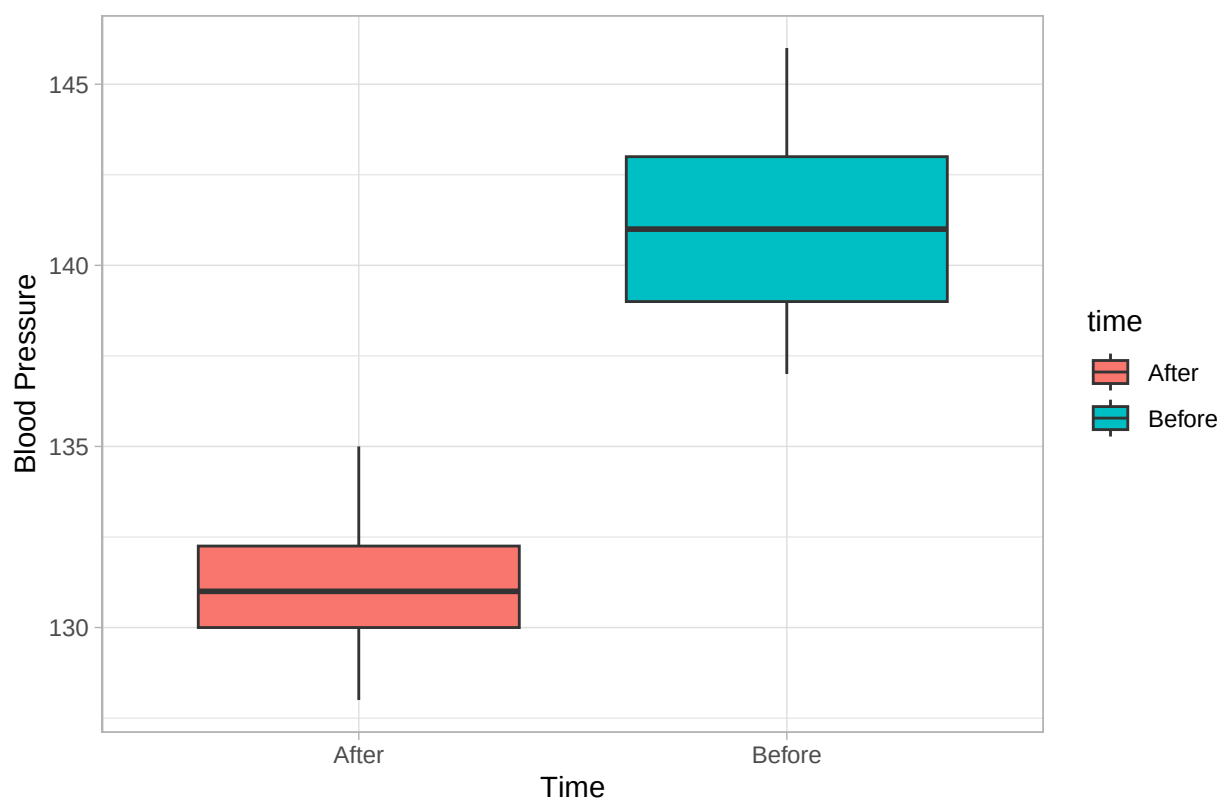
# Paired t-test
paired_test_result <- t.test(bp_data$bp_before, bp_data$bp_after, paired = TRUE)
paired_test_result

##
## Paired t-test
##
## data: bp_data$bp_before and bp_data$bp_after
## t = 32.254, df = 19, p-value < 2.2e-16
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  9.257573 10.542427
## sample estimates:
## mean difference
##          9.9

# Visualization: Boxplot
bp_long_data <- data.frame(
  time = rep(c("Before", "After"), each = 20),
  bp = c(bp_data$bp_before, bp_data$bp_after)
)

ggplot(bp_long_data, aes(x = time, y = bp, fill = time)) +
  geom_boxplot() +
  labs(title = "Blood Pressure: Before vs. After Diet",
       x = "Time",
       y = "Blood Pressure") +
  theme_light()
```

Blood Pressure: Before vs. After Diet



##Scenario 3: Independent t-test Objective: Compare resting heart rate between two supplement groups.

Data: Resting heart rate for two groups

```
heart_rate_data <- data.frame(  
  supplement_group = rep(c("Group A", "Group B"), each = 20),  
  heart_rate = c(72, 70, 71, 73, 74, 72, 70, 72, 73, 71,  
                 70, 72, 74, 72, 71, 73, 72, 72, 73, 71, 72,  
                 72, 73, 71, 72, 70, 72, 74, 73, 71, 70,  
                 71, 72, 73, 72, 70, 72, 71, 73, 72)  
)
```

Independent t-test

```
heart_test_result <- t.test(heart_rate ~ supplement_group, data = heart_rate_data)  
heart_test_result
```

##

Welch Two Sample t-test

##

data: heart_rate by supplement_group

t = 0.27296, df = 37.693, p-value = 0.7864

alternative hypothesis: true difference in means between group Group A and group Group B is not equal to 0

95 percent confidence interval:

-0.64183 0.84183

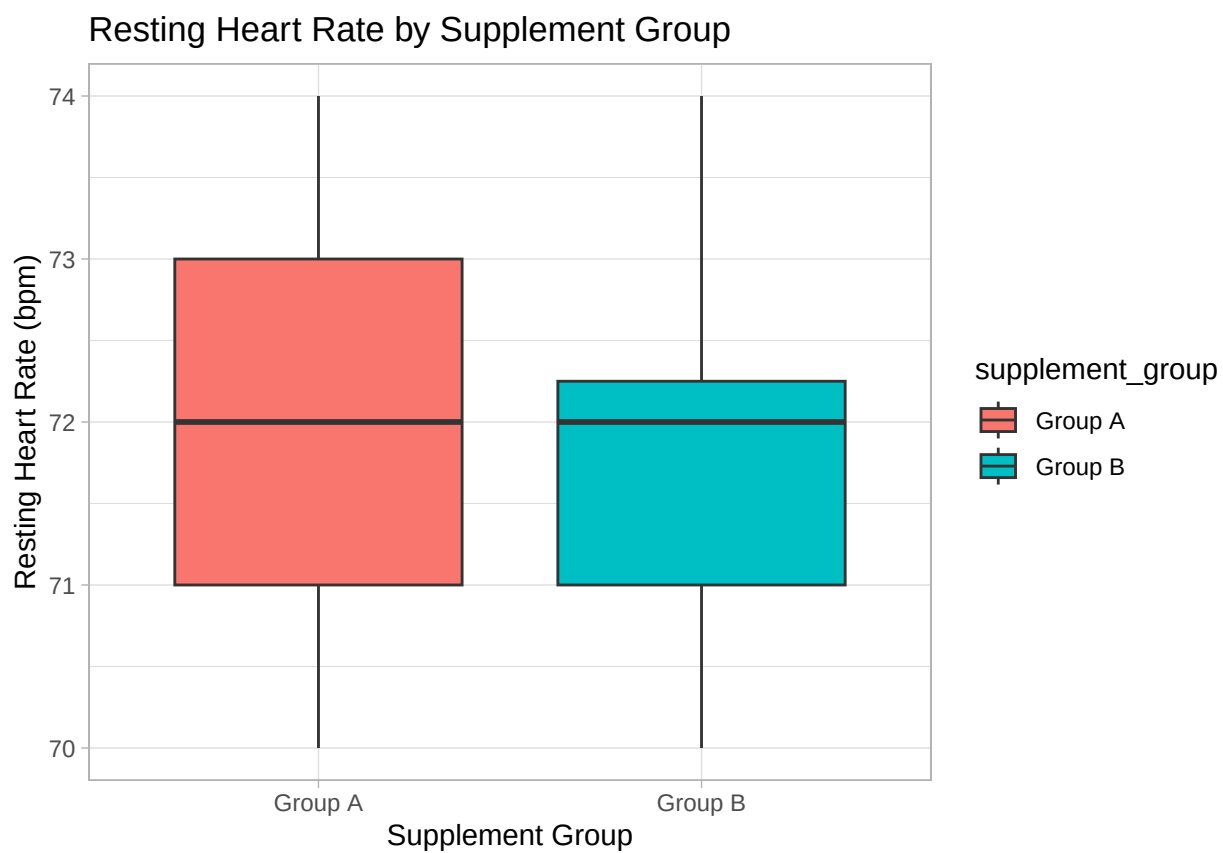
sample estimates:

mean in group Group A mean in group Group B

71.9 71.8

Visualization: Boxplot

```
ggplot(heart_rate_data, aes(x = supplement_group, y = heart_rate, fill = supplement_group)) +  
  geom_boxplot() +  
  labs(title = "Resting Heart Rate by Supplement Group",  
        x = "Supplement Group",  
        y = "Resting Heart Rate (bpm)") +  
  theme_light()
```



Scenario 4: One-way ANOVA

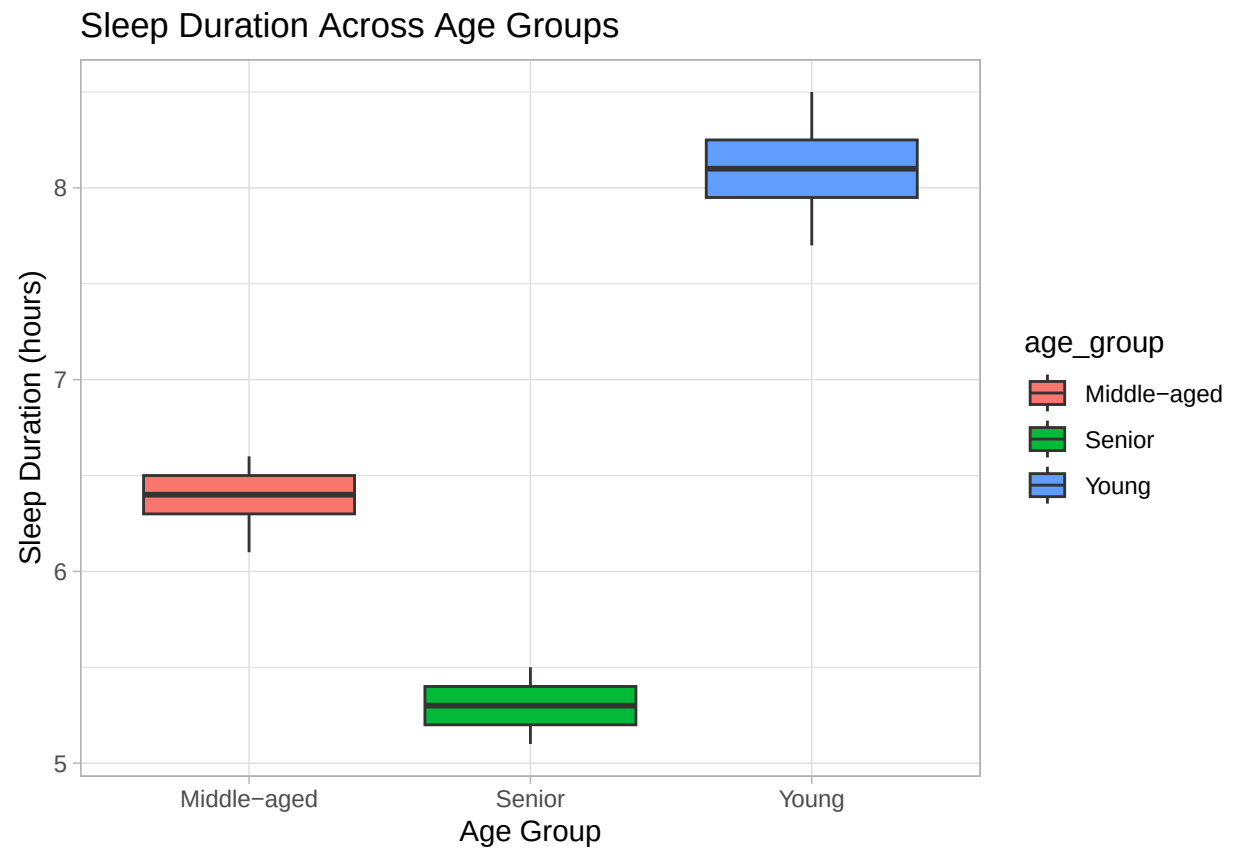
Objective: Examine differences in sleep duration across age groups.

```
# Data: Sleep duration across age groups
sleep_duration_data <- data.frame(
  age_group = rep(c("Young", "Middle-aged", "Senior"), each = 15),
  sleep_hours = c(8.5, 8.2, 8.1, 7.8, 8.0, 8.3, 7.9, 8.4, 8.2, 8.1, 8.0, 7.7, 7.9, 8.3, 8.0,
                  6.5, 6.3, 6.6, 6.4, 6.2, 6.5, 6.6, 6.4, 6.3, 6.1, 6.4, 6.2, 6.3, 6.5, 6.4,
                  5.4, 5.2, 5.5, 5.3, 5.1, 5.3, 5.4, 5.2, 5.3, 5.5, 5.2, 5.3, 5.4, 5.1, 5.3)
)

# One-way ANOVA
anova_model <- aov(sleep_hours ~ age_group, data = sleep_duration_data)
anova_result <- summary(anova_model)
anova_result

##           Df Sum Sq Mean Sq F value Pr(>F)
## age_group    2  59.52   29.762    1014 <2e-16 ***
## Residuals   42   1.23    0.029
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Visualization: Boxplot
ggplot(sleep_duration_data, aes(x = age_group, y = sleep_hours, fill = age_group)) +
  geom_boxplot() +
  labs(title = "Sleep Duration Across Age Groups",
       x = "Age Group",
       y = "Sleep Duration (hours)") +
  theme_light()
```



Scenario 5: Correlation

Objective: Investigate the relationship between study hours and exam scores.

```
# Data: Study hours and exam scores
study_data <- data.frame(
  study_hours = c(2, 3, 4, 5, 6, 7, 8, 9, 10, 11),
  exam_scores = c(55, 60, 65, 70, 74, 78, 82, 85, 88, 90)
)

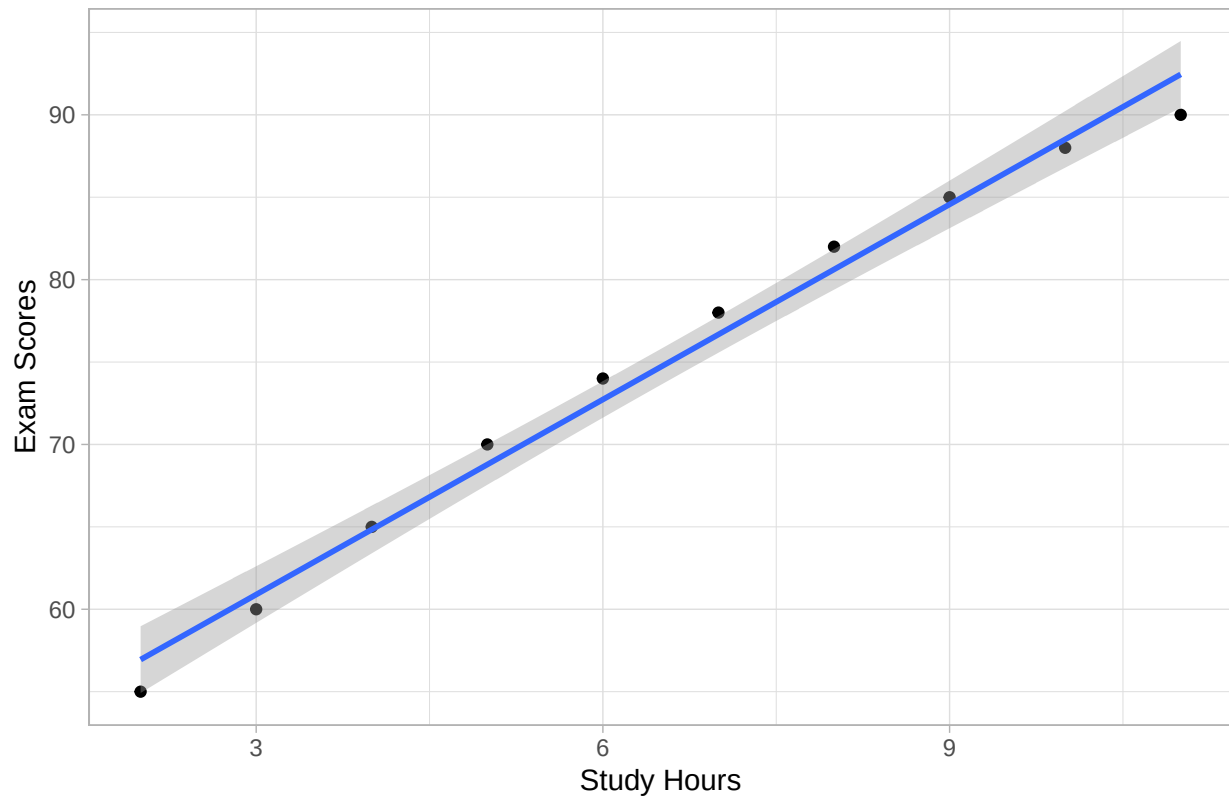
# Correlation test
correlation_result <- cor.test(study_data$study_hours, study_data$exam_scores, method = "pearson")
correlation_result

##
## Pearson's product-moment correlation
##
## data: study_data$study_hours and study_data$exam_scores
## t = 23.988, df = 8, p-value = 9.72e-09
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## 0.9700795 0.9984322
## sample estimates:
## cor
## 0.9931203

# Visualization: Scatter plot
ggplot(study_data, aes(x = study_hours, y = exam_scores)) +
  geom_point() +
  geom_smooth(method = "lm", se = TRUE) +
  labs(title = "Study Hours vs. Exam Scores",
       x = "Study Hours",
       y = "Exam Scores") +
  theme_light()

## `geom_smooth()` using formula = 'y ~ x'
```

Study Hours vs. Exam Scores



Scenario 6: Chi-Square Test

Objective: Check for association between gender and remote work preference.

```
# Data: Gender and work preference
gender_data <- c(rep("Male", 20), rep("Female", 20))
preference_data <- c("Remote", "Office", "Remote", "Remote", "Remote", "Office", "Remote", "Remote", "O",
                     "Office", "Remote", "Remote", "Remote", "Remote", "Office", "Remote", "Office", "R",
                     "Office", "Office", "Office", "Remote", "Office", "Remote", "Office", "Office", "R",
                     "Remote", "Remote", "Office", "Remote", "Remote", "Office", "Office", "Remote", "R")

# Chi-Square test
preference_table <- table(gender_data, preference_data)
chi_square_result <- chisq.test(preference_table)
chi_square_result
```

```
##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data: preference_table
## X-squared = 1.6368, df = 1, p-value = 0.2008
```

Scenario 7: One-way ANOVA

Objective: Analyze differences in work hours across departments.

```
# Data: Work hours across departments
work_hours_data <- data.frame(
  department = rep(c("HR", "Finance", "IT"), each = 15),
  work_hours = c(
    38, 39, 40, 41, 39, 40, 38, 40, 39, 41, 40, 39, 38, 40, 39,
    39, 40, 41, 40, 39, 40, 38, 41, 40, 39, 40, 40, 39, 40, 39,
    40, 41, 39, 40, 40, 38, 39, 40, 41, 40, 39, 39, 40, 40, 40)
)

# One-way ANOVA
work_anova_model <- aov(work_hours ~ department, data = work_hours_data)
work_anova_result <- summary(work_anova_model)
work_anova_result
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## department    2   0.93  0.4667   0.615  0.545
## Residuals   42  31.87  0.7587
```

```
# Visualization: Boxplot
ggplot(work_hours_data, aes(x = department, y = work_hours, fill = department)) +
  geom_boxplot() +
  labs(title = "Work Hours by Department",
       x = "Department",
       y = "Work Hours") +
  theme_light()
```

