

Week-9

Anas

2025-07-09

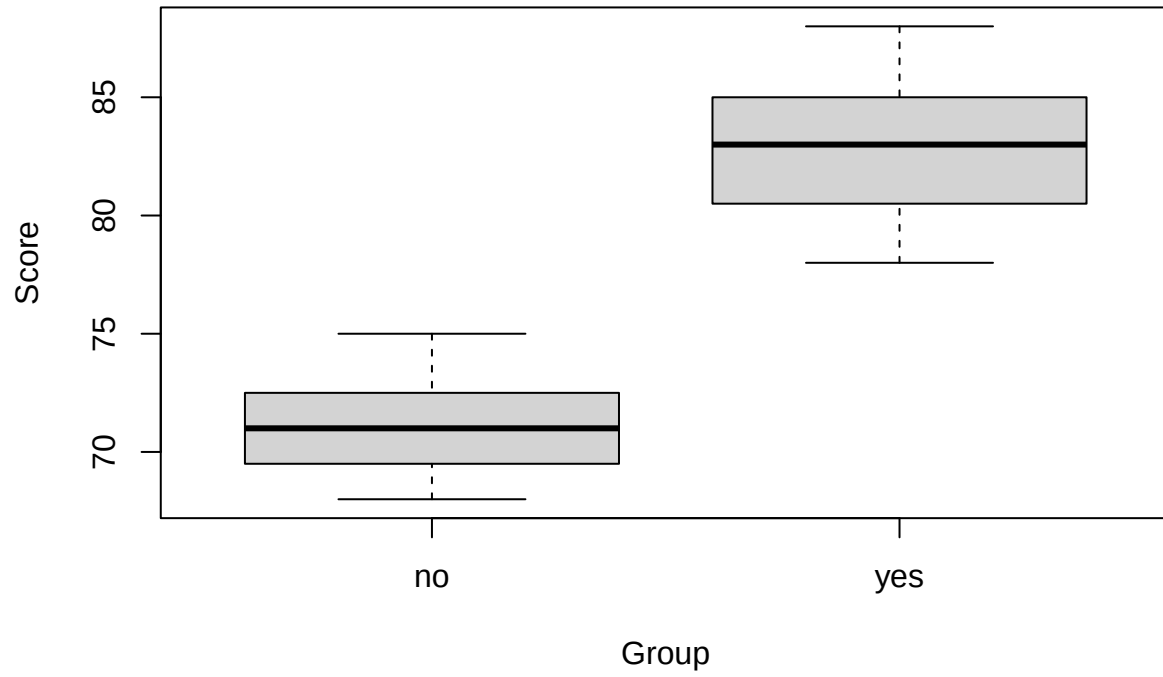
Scenario 1: Independent t-test

Question:

Is there a significant difference in test scores between students who attended a revision class and those who did not?

```
scores <- data.frame(  
  group = rep(c("yes", "no"), each = 15),  
  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)  
)  
  
t.test(score ~ group, data = scores, var.equal = TRUE)  
  
##  
## Two Sample t-test  
##  
## data: score by group  
## t = -12.683, df = 28, p-value = 3.994e-13  
## alternative hypothesis: true difference in means between group no and group yes is not equal to 0  
## 95 percent confidence interval:  
## -13.93814 -10.06186  
## sample estimates:  
## mean in group no mean in group yes  
## 71 83  
  
boxplot(score ~ group, data = scores,  
  main = "Test Scores by Revision Class Attendance",  
  xlab = "Group", ylab = "Score", col = )
```

Test Scores by Revision Class Attendance



Scenario 2: Paired t-test

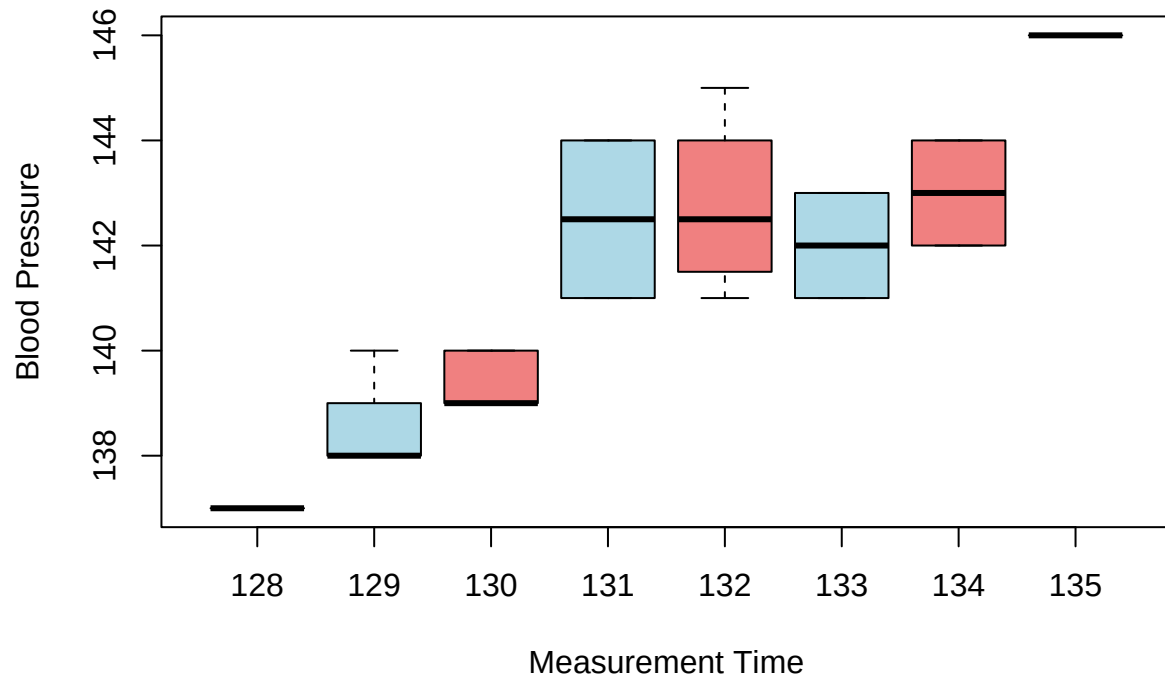
Question:

Do blood pressure levels change after a 2-week diet plan?

```
bp <- data.frame(
  before = c(140, 145, 138, 142, 141, 139, 144, 143, 140, 137,
             141, 142, 139, 143, 138, 140, 144, 146, 139, 141),
  after  = c(130, 132, 129, 134, 133, 130, 131, 132, 129, 128,
             131, 132, 130, 133, 129, 130, 134, 135, 130, 132)
)
t.test(bp$before, bp$after, paired = TRUE)

##
## Paired t-test
##
## data: bp$before and bp$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
# Create the boxplot
boxplot(before ~ after, data = bp,
        main = "Blood Pressure Before and After Treatment",
        xlab = "Measurement Time", ylab = "Blood Pressure",
        col = c("lightcoral", "lightblue"))
```

Blood Pressure Before and After Treatment



Scenario 3: Independent t-test

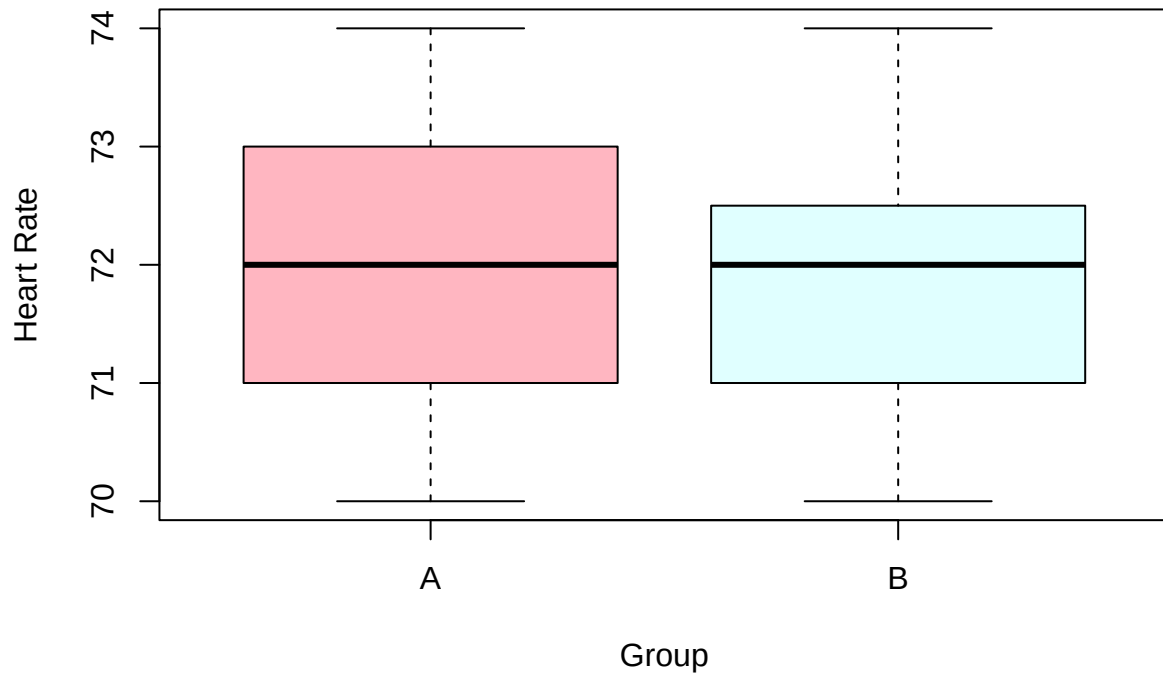
Question:

Is there a significant difference in resting heart rate between two groups who take two different types of mild supplements?

Group A and B have similar means and small variance

```
heart_data <- data.frame(  
  group = rep(c("A", "B"), each = 20),  
  bpm = 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)  
)  
t.test(bpm ~ group, data = heart_data)  
  
##  
## Welch Two Sample t-test  
##  
## data: bpm by group  
## t = 0.27296, df = 37.693, p-value = 0.7864  
## alternative hypothesis: true difference in means between group A and group B is not equal to 0  
## 95 percent confidence interval:  
## -0.64183 0.84183  
## sample estimates:  
## mean in group A mean in group B  
## 71.9 71.8  
  
boxplot(bpm ~ group, data = heart_data,  
  main = "Heart Rate Comparison Between Groups A and B",  
  xlab = "Group", ylab = "Heart Rate",  
  col = c("lightpink", "lightcyan"))
```

Heart Rate Comparison Between Groups A and B



Scenario 4: One-way ANOVA

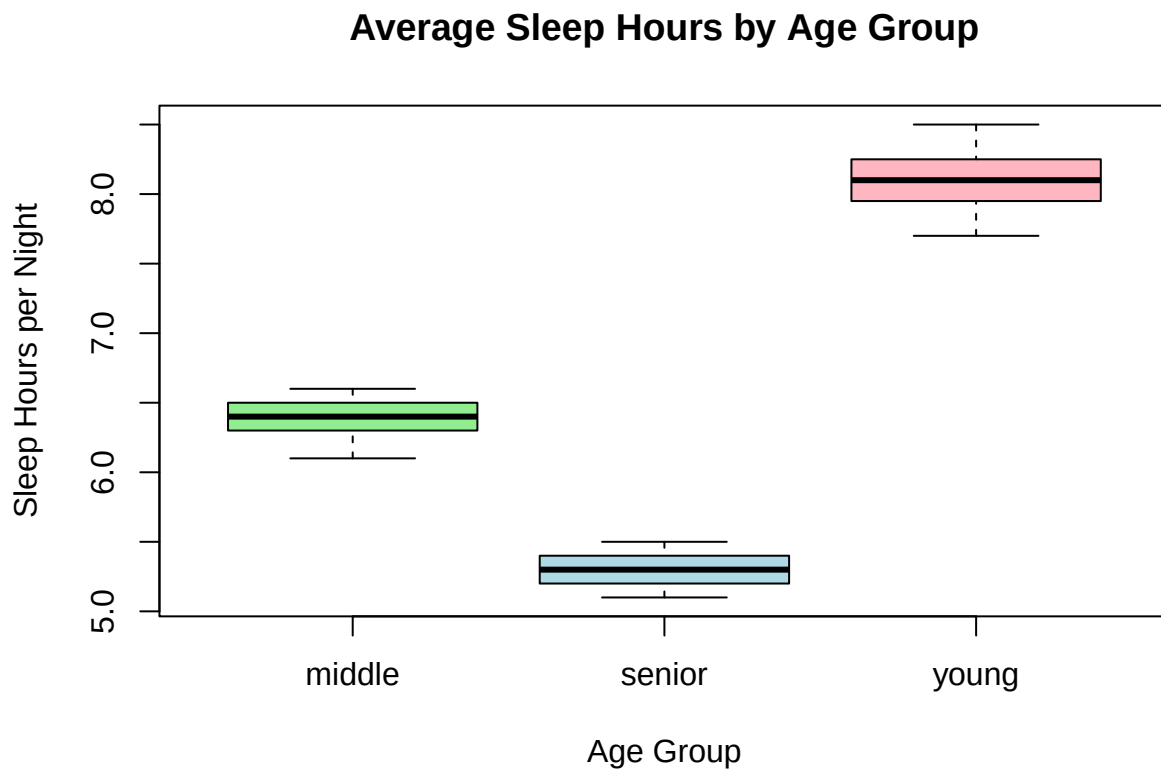
Question:

Does average sleep duration differ across age groups?

```
sleep_data <- data.frame(  
  age_group = rep(c("young", "middle", "senior"), each = 15),  
  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)  
)  
anova_model <- aov(hours ~ age_group, data = sleep_data)  
summary(anova_model)
```

```
##           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
```

```
boxplot(hours ~ age_group, data = sleep_data,  
  main = "Average Sleep Hours by Age Group",  
  xlab = "Age Group", ylab = "Sleep Hours per Night",  
  col = c("lightgreen", "lightblue", "lightpink"))
```



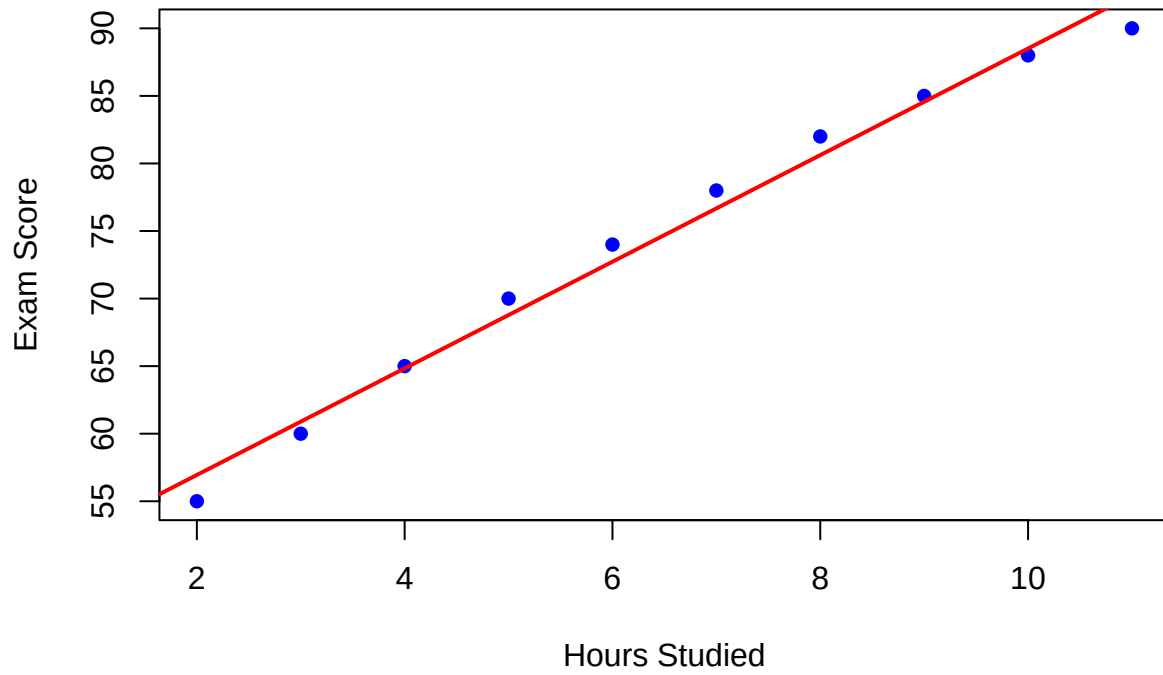
Scenario 5: Correlation

Question:

Is there a relationship between study hours and final exam scores?

```
study <- data.frame(  
  hours = c(2, 3, 4, 5, 6, 7, 8, 9, 10, 11),  
  score = c(55, 60, 65, 70, 74, 78, 82, 85, 88, 90)  
)  
cor.test(study$hours, study$score, method = "pearson")  
  
##  
## Pearson's product-moment correlation  
##  
## data: study$hours and study$score  
## 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  
  
plot(score ~ hours, data = study,  
  main = "Relationship Between Study Hours and Exam Score",  
  xlab = "Hours Studied", ylab = "Exam Score",  
  col = "blue", pch = 16)  
  
# Add regression line  
abline(lm(score ~ hours, data = study), col = "red", lwd = 2)
```

Relationship Between Study Hours and Exam Score



Scenario 6: Chi-square Test

Question:

Is there an association between gender and preference for remote work?

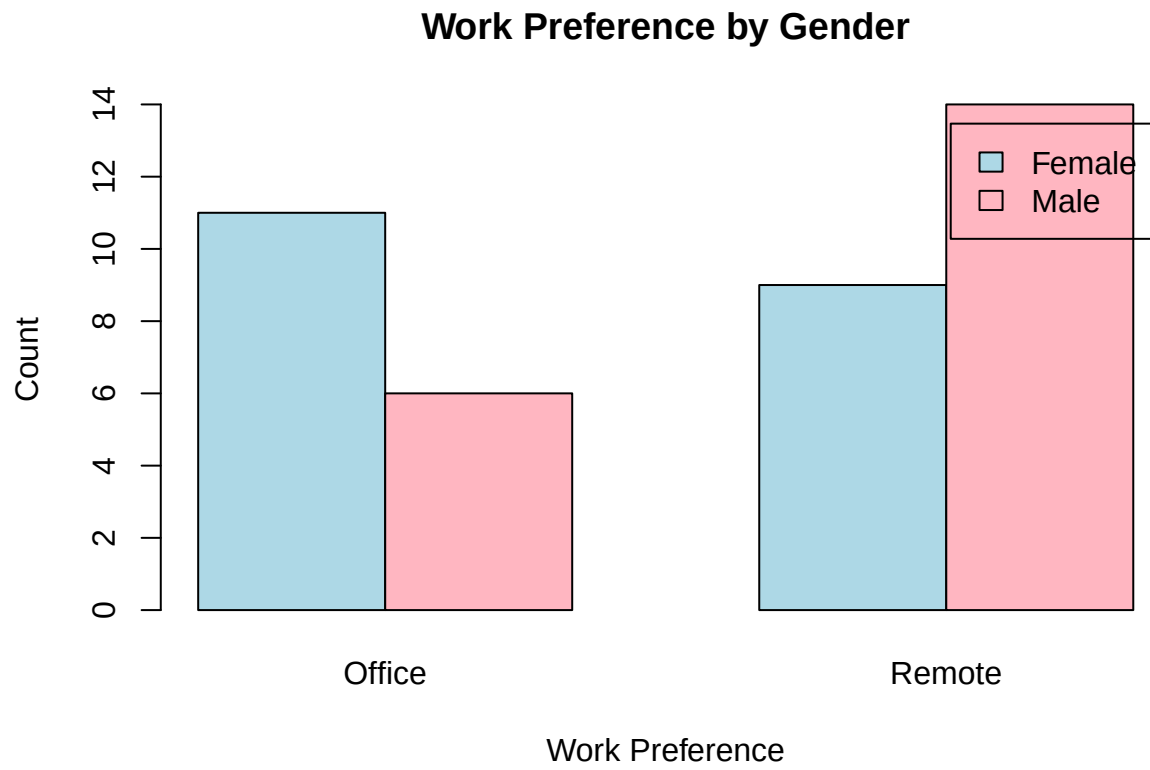
```
gender <- c(rep("Male", 20), rep("Female", 20))
preference <- c("Remote", "Office", "Remote", "Remote", "Remote",
               "Office", "Remote", "Remote", "Office", "Remote",
               "Office", "Remote", "Remote", "Remote", "Remote",
               "Office", "Remote", "Office", "Remote", "Remote",
               "Office", "Office", "Office", "Remote", "Office",
               "Remote", "Office", "Office", "Remote", "Office",
               "Remote", "Remote", "Office", "Remote", "Remote",
               "Office", "Office", "Remote", "Remote", "Office")

table_data <- table(gender, preference)
chisq.test(table_data)

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

# Create contingency table
table_data <- table(gender, preference)

# Create the barplot
barplot(table_data,
        beside = TRUE,
        col = c("lightblue", "lightpink"),
        main = "Work Preference by Gender",
        xlab = "Work Preference",
        ylab = "Count",
        legend = rownames(table_data))
```



Scenario 7: One-way ANOVA

Question:

Does the number of work hours differ significantly across three departments?

```
work_hours <- data.frame(  
  department = rep(c("HR", "Finance", "IT"), each = 15),  
  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)  
)  
anova_model2 <- aov(hours ~ department, data = work_hours)  
summary(anova_model2)
```

```
##           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
```

```
boxplot(hours ~ department, data = work_hours,  
  main = "Work Hours by Department",  
  xlab = "Department", ylab = "Weekly Work Hours",  
  col = c("lightblue", "lightgreen", "lightpink"))
```

