

Data Processing

Team 2

2025-11-03

```
data <- read.csv("00_ProstateCancer_Data.csv", header=T)

cols1 <- c("Tani_Tarihi", "Tedavi_Tarihi", "PSA_Takip_3ay", "PSA_Takip_6ay", "PSA_Takip_12ay", "BCR_Tar")
data <- data[, !(names(data) %in% cols1)]

cols2 <- c("Klinik_Evre", "Biyopsi_Gleason", "Risk_Grubu", "Komorbidite_Skor", "Tedavi_Tipi", "ADT_Tipi")
data[cols2] <- lapply(data[cols2], as.factor)

names(data) <- c("Patient ID", "Age", "PSA_before", "CTstage", "GleasonScore_before", "RiskClass", "Alb")
```

Data Analysis 1: Chi-squared Test, Treatment vs BCR

Hypotheses

H0: Treatment Type and Biochemical Recurrence (BCR) Status are independent. (The recurrence rate is the same across all treatment groups)

Ha: Treatment Type and Biochemical Recurrence (BCR) Status are not independent. (The recurrence rate is significantly different for at least one treatment group)

```
bcr_table <- table(data$Treatment, data$BCR)
cat("Frequency Table (Counts):\n")
```

```
## Frequency Table (Counts):
```

```
print(bcr_table)
```

```
##
##      False True
##  1    189   65
##  2    123   59
##  3     40   31
##  4     64   29
```

```
bcr_test <- chisq.test(bcr_table)
cat("\nChi-squared Test Results:\n")
```

```
##
## Chi-squared Test Results:

print(bcr_test)

##
## Pearson's Chi-squared test
##
## data:  bcr_table
## X-squared = 8.9915, df = 3, p-value = 0.0294

cat("\nBCR Rate (%) within each Treatment Group:\n")

##
## BCR Rate (%) within each Treatment Group:

round(prop.table(bcr_table, margin = 1) * 100, 1)

##
##      False True
##  1  74.4 25.6
##  2  67.6 32.4
##  3  56.3 43.7
##  4  68.8 31.2
```

Conclusion: Since the p-value (0.0294) is less than the significance level ($\alpha=0.05$), we reject the null hypothesis (H_0).

Interpretation: There is a statistically significant association between the type of treatment a patient receives and the likelihood of experiencing Biochemical Recurrence (BCR).

Highest Recurrence Rate: Treatment Type 3 (Hormone Therapy (ADT) Monotherapy) showed the highest biochemical recurrence rate at 43.7%.

Lowest Recurrence Rate: Treatment Type 1 (Radical Prostatectomy) showed the lowest recurrence rate at 25.6%.

```
bcr_prop_table <- prop.table(bcr_table, margin = 1)

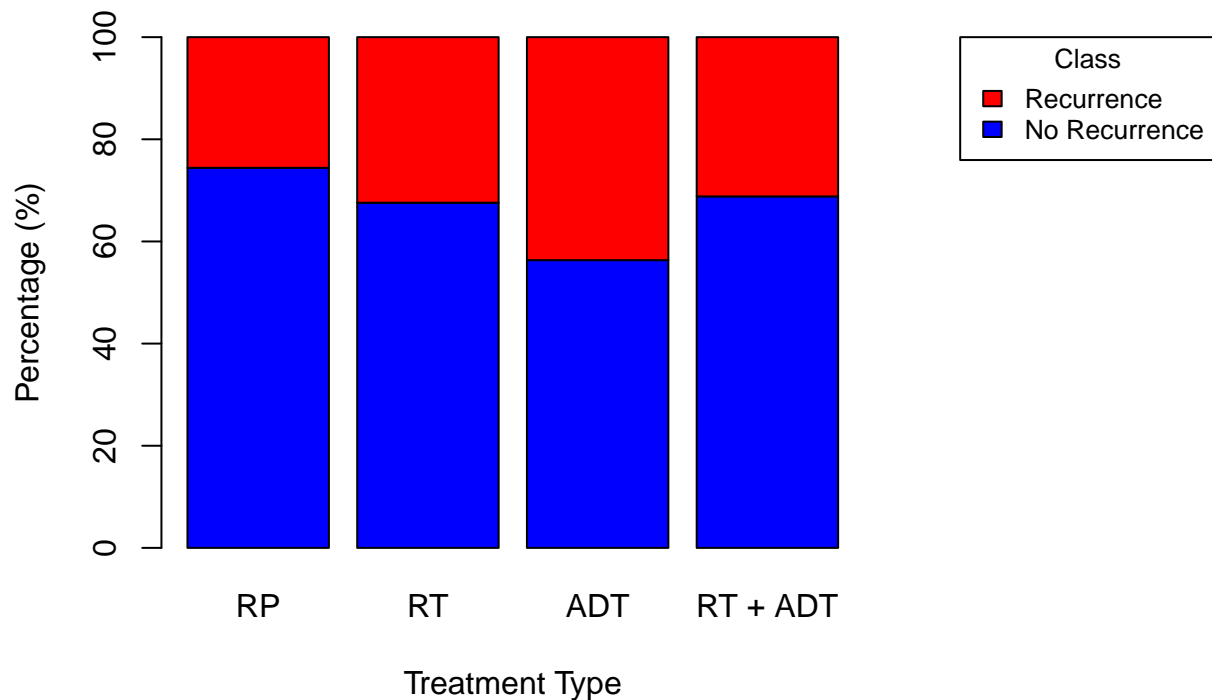
par(mar = c(5, 4, 4, 10) + 0.1)
barplot(
  t(bcr_prop_table) * 100,
  main = "Biochemical Recurrence (BCR) Rate by Treatment Type",
  xlab = "Treatment Type",
  ylab = "Percentage (%)",
  col = c("blue", "red"),
  legend.text = c("No Recurrence", "Recurrence"),
  args.legend = list(
    title = "Class",
    x = "topright",
    inset = c(-0.5, 0),
    xpd = TRUE,
    cex = 0.8
  )
)
```

```

),
ylim = c(0, 100),
names.arg = c("RP", "RT", "ADT", "RT + ADT")
)

```

Biochemical Recurrence (BCR) Rate by Treatment Type



Data Analysis 2: Chi-squared Test, Treatment vs Survival

Hypotheses

H0: Treatment Type and Survival Status are independent. (The survival rate is the same across all treatment groups)

Ha: Treatment Type and Survival Status are not independent. (The survival rate is significantly different for at least one treatment group)

```

survival_table <- table(data$Treatment, data$Survival)
cat("Frequency Table (Counts):\n")

```

```
## Frequency Table (Counts):
```

```
print(survival_table)
```

```
##
```

```
##      0    1
##    1 35 219
##    2 24 158
##    3  6  65
##    4  6  87
```

```
survival_test <- chisq.test(survival_table)
cat("\nChi-squared Test Results:\n")
```

```
##
## Chi-squared Test Results:
```

```
print(survival_test)
```

```
##
## Pearson's Chi-squared test
##
## data: survival_table
## X-squared = 4.6021, df = 3, p-value = 0.2034
```

```
cat("\nSurvival Rate (%) within each Treatment Group (Survival = Alive):\n")
```

```
##
## Survival Rate (%) within each Treatment Group (Survival = Alive):
```

```
round(prop.table(survival_table, margin = 1) * 100, 1)
```

```
##
##      0    1
##    1 13.8 86.2
##    2 13.2 86.8
##    3  8.5 91.5
##    4  6.5 93.5
```

Conclusion: Since the p-value (0.2034) is greater than the significance level ($\alpha=0.05$), we fail to reject the null hypothesis (H_0).

Interpretation: There is no statistically significant association between the type of treatment a patient receives and the likelihood of their Survival Status (Alive vs Deceased).

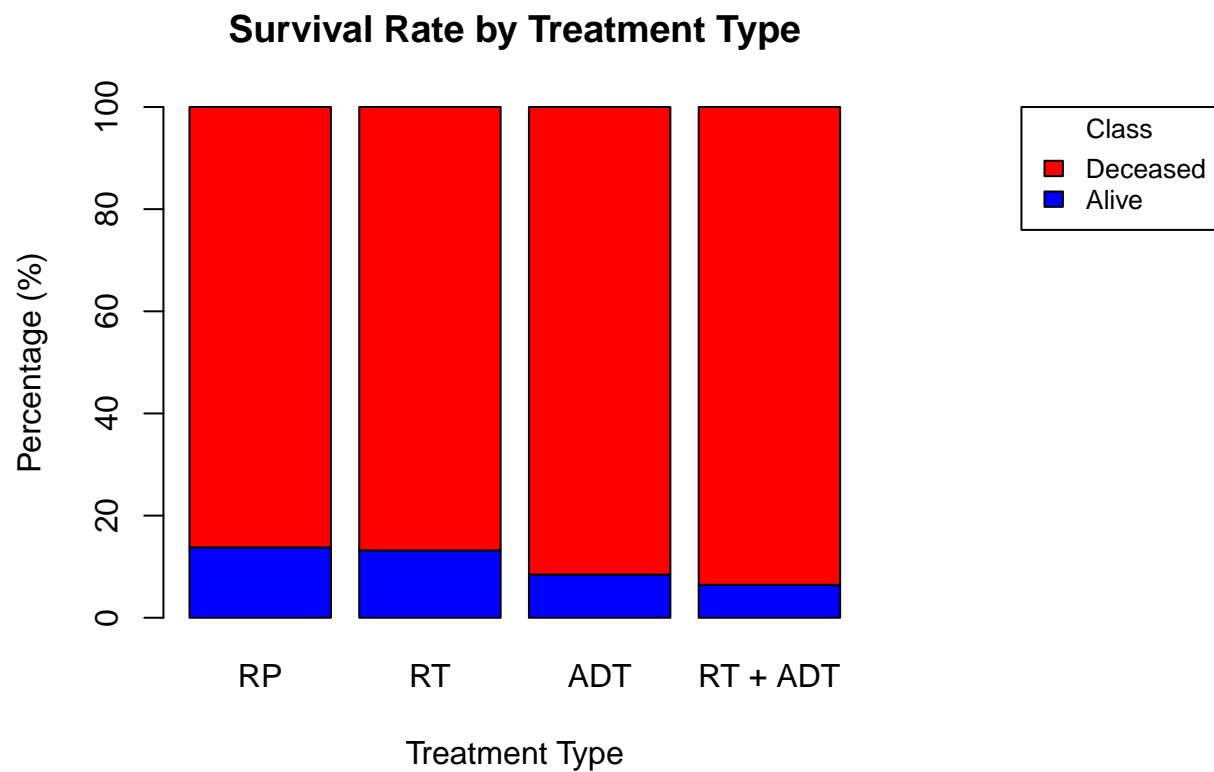
```
survival_prop_table <- prop.table(survival_table, margin = 1)
```

```
par(mar = c(5, 4, 4, 10) + 0.1)
barplot(
  t(survival_prop_table) * 100,
  main = "Survival Rate by Treatment Type",
  xlab = "Treatment Type",
  ylab = "Percentage (%)",
  col = c("blue", "red"),
  legend.text = c("Alive", "Deceased"),
  args.legend = list(
```

```

title = "Class",
x = "topright",
inset = c(-0.5, 0),
xpd = TRUE,
cex = 0.8
),
ylim = c(0, 100),
names.arg = c("RP", "RT", "ADT", "RT + ADT")
)

```



Data Analysis 3: ANOVA Test, Treatment vs PSA Difference

Hypotheses

H0: The mean change in PSA (Delta PSA) is the same across all four Treatment Type groups

Ha: The mean change in PSA (Delta PSA) is significantly different for at least one treatment group

```

data$Delta_PSA <- data$PSA_before - data$PSA_after

anova_result <- aov(Delta_PSA ~ Treatment, data = data)
summary(anova_result)

```

```
##           Df  Sum Sq Mean Sq F value    Pr(>F)
```

```
## Treatment      3    60755    20252    9.774 2.66e-06 ***
## Residuals     596 1234905     2072
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
TukeyHSD(anova_result)
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = Delta_PSA ~ Treatment, data = data)
##
## $Treatment
##      diff      lwr      upr    p adj
## 2-1  14.59788   3.209034 25.986728 0.0055972
## 3-1  29.75452  14.011616 45.497414 0.0000085
## 4-1  16.52282   2.309503 30.736139 0.0151264
## 3-2  15.15663  -1.252454 31.565723 0.0821099
## 4-2   1.92494 -13.022893 16.872773 0.9873986
## 4-3 -13.23169 -31.713328  5.249938 0.2535812
```

Conclusion: Since the p-value (2.66e-06) is less than the significance level ($\alpha=0.05$), we reject the null hypothesis (H_0).

Interpretation: Interpretation: There is a statistically significant difference in the mean change in PSA (Delta PSA) among the different treatment groups.

Tukey's HSD Interpretation:

The mean PSA change (Delta PSA, where a larger value indicates a greater reduction) in the Radical Prostatectomy (Treatment 1) group is significantly smaller than the mean change in all other treatment groups (2, 3, and 4).

Largest Mean Difference: Treatment 3 (Hormone Therapy Monotherapy) showed the largest positive mean difference (Delta PSA = 29.755) compared to Treatment 1, indicating the greatest average PSA reduction after treatment.

No Significant Difference: There is no significant difference in the mean PSA Difference among the three non-surgical groups (Treatment 2, 3, and 4).

```
par(mar = c(5, 4, 4, 2) + 0.1)
boxplot(Delta_PSA ~ Treatment,
        data = data,
        main = "Delta PSA Distribution by Treatment Type",
        xlab = "Treatment Type",
        ylab = "Delta PSA (PSA at diagnosis - PSA after treatment)",
        col = rainbow(4),
        names = c("RP", "RT", "ADT", "RT + ADT"))

means <- tapply(data$Delta_PSA, data$Treatment, mean, na.rm=TRUE)
points(means, pch = 8, col = "black", cex = 1.5)
```

