

Sheth L.U.J. & Sir M.V. College

Aim :- Performing two-way ANOVA using aov() (R)____(Practical8).

Two-way ANOVA Practical Report – Cholesterol by Sex and Chest Pain Type

Dataset: <https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction>

1. Objective

To test whether the mean cholesterol level differs by **Sex** and **Chest Pain Type**, and to examine if there is an interaction effect between these two factors.

2. Hypotheses

Main Effects

Sex

- **Null Hypothesis (H_0):** Mean cholesterol is the same for males and females.
- **Alternative Hypothesis (H_1):** Mean cholesterol differs between males and females.

Chest Pain Type

- **Null Hypothesis (H_0):** Mean cholesterol is the same across all chest pain types.
- **Alternative Hypothesis (H_1):** At least one chest pain type has a different mean.

Interaction Effect

- **Null Hypothesis (H_0):** No interaction effect between Sex and Chest Pain Type.
 - **Alternative Hypothesis (H_1):** There is an interaction effect.
-

3. Assumption Checks

3.1 Homogeneity of Variances (Levene's Test)

| Test | F-value | p-value | Decision |
|--------|---------|------------------------|-------------------------|
| Levene | 8.8289 | 1.54×10^{-10} | Variances are not equal |

Interpretation: The assumption of equal variances is violated. Results should be interpreted with caution.

Name :- Priya Gupta

Roll no. :- S081

Sheth L.U.J. & Sir M.V. College

3.2 Normality of Residuals (Shapiro-Wilk Test)

| Test | W | p-value | Decision |
|--------------|---------|-------------------------|----------------------------------|
| Shapiro-Wilk | 0.91817 | $< 2.2 \times 10^{-16}$ | Residuals deviate from normality |

Interpretation: Residuals are not perfectly normal. ANOVA is still robust for large sample sizes ($n = 918$).

4. ANOVA Results

| Source | Df | Sum Sq | Mean Sq | F value | Pr(>F) | Significance |
|-------------------|-----|----------|---------|---------|----------|--------------|
| Sex | 1 | 439277 | 439277 | 38.776 | 7.25e-10 | *** |
| ChestPainType | 3 | 172305 | 57435 | 5.070 | 0.00174 | ** |
| Sex:ChestPainType | 3 | 51263 | 17088 | 1.508 | 0.21084 | |
| Residuals | 910 | 10308961 | 11329 | | | |

Interpretation:

- Both **Sex** and **Chest Pain Type** significantly affect cholesterol.
 - Interaction effect is **not significant**.
-

5. Post-hoc Analysis (Tukey HSD) – Chest Pain Type

| Comparison | Mean Difference | 95% CI | p-value | Significance |
|------------|-----------------|---------------|---------|--------------|
| ATA - ASY | 35.36 | 11.17, 59.55 | 0.0010 | Yes |
| NAP - ASY | 4.35 | -18.47, 27.18 | 0.9611 | No |
| TA - ASY | 16.33 | -25.90, 58.55 | 0.7522 | No |

Name :- Priya Gupta

Roll no. :- S081

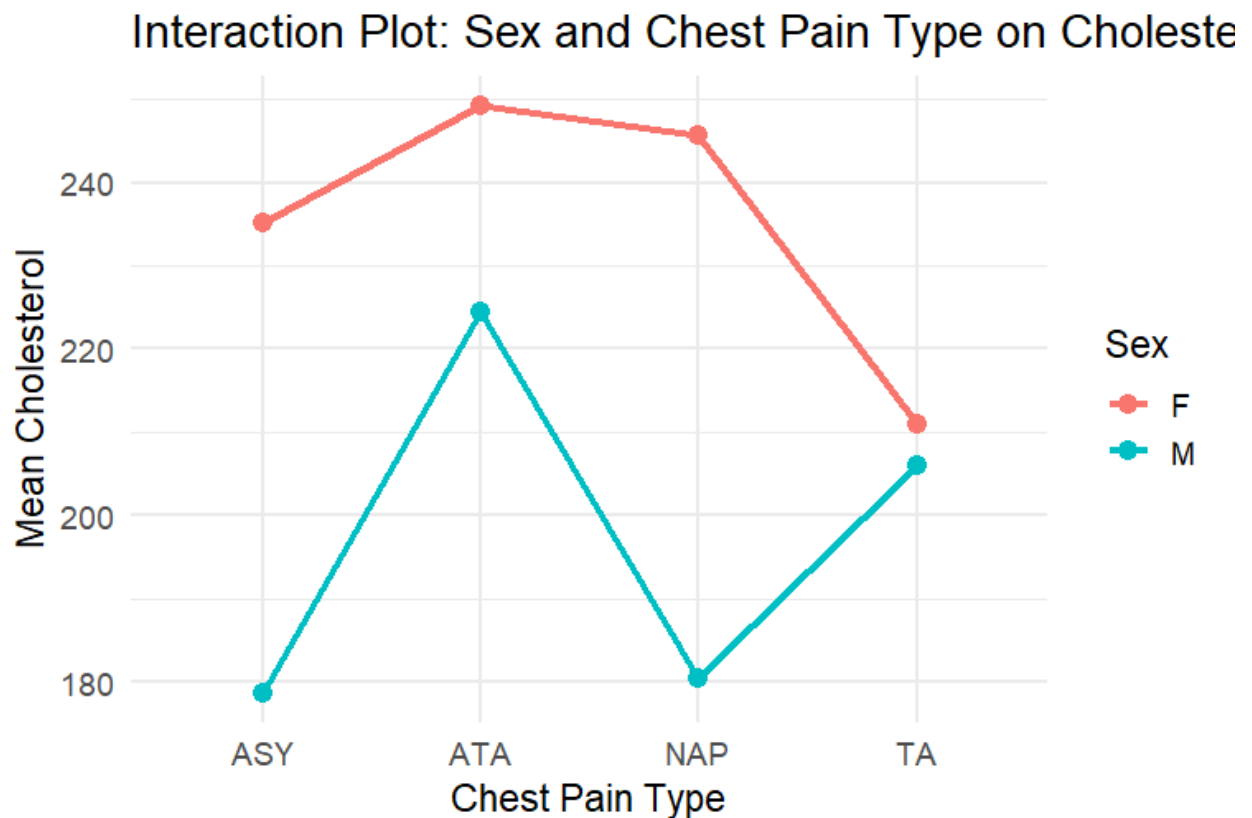
Sheth L.U.J. & Sir M.V. College

| Comparison | Mean Difference | 95% CI | p-value | Significance |
|------------|-----------------|---------------|---------|--------------|
| NAP - ATA | -31.01 | -59.35, -2.66 | 0.0256 | Yes |
| TA - ATA | -19.03 | -64.48, 26.41 | 0.7031 | No |
| TA - NAP | 11.97 | -32.76, 56.71 | 0.9014 | No |

Interpretation: Significant differences observed mainly for **ATA vs ASY** and **NAP vs ATA** chest pain types.

6. Visualization

Interaction Plot:

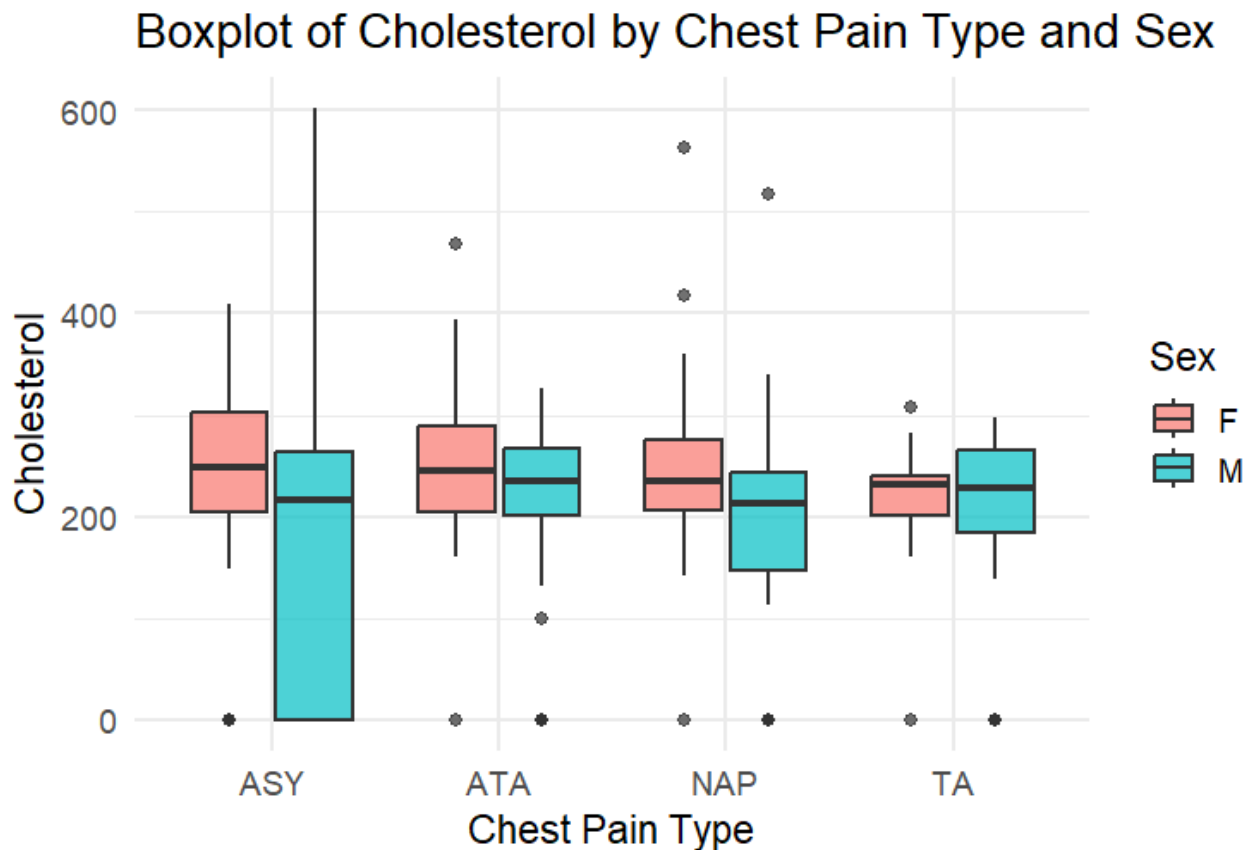


Name :- Priya Gupta

Roll no. :- S081

Sheth L.U.J. & Sir M.V. College

Boxplot:



Observation:

- Males generally have higher cholesterol than females.
- ATA chest pain type shows higher mean cholesterol compared to ASY and NAP.
- Lines in interaction plot are mostly parallel → interaction effect is weak.

7. Conclusion

- **Sex** and **Chest Pain Type** significantly affect cholesterol levels.
 - Interaction effect between Sex and Chest Pain Type is **not significant**.
 - Tukey HSD shows significant differences mainly between **ATA vs ASY** and **NAP vs ATA**.
 - Despite minor violations of ANOVA assumptions, results are reliable for the large sample size (n = 918).
-

Name :- Priya Gupta

Roll no. :- S081

Sheth L.U.J. & Sir M.V. College

Screenshots

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins Project (None)

Source
Console Terminal Background Jobs

R 4.5.2 ~ /
> # Load dataset
> df <- read.csv("C:/Users/Priya Gupta/Downloads/heart.csv")
>
> # Convert categorical variables
> df$Sex <- as.factor(df$Sex)
> df$ChestPainType <- as.factor(df$ChestPainType)
>
> # Fit Two-way ANOVA
> anova2_model <- aov(Cholesterol ~ Sex * ChestPainType, data = df)
> summary(anova2_model)
          Df Sum Sq Mean Sq F value    Pr(>F)
Sex          1  439277   439277   38.776 7.25e-10 ***
ChestPainType 3   172305    57435    5.070 0.00174 **
Sex:ChestPainType 3    51263    17088    1.508 0.21084
Residuals    910 10308961    11329
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> # Check residual normality
> res2 <- residuals(anova2_model)
> shapiro.test(res2)

      Shapiro-Wilk normality test

data:  res2
W = 0.91817, p-value < 2.2e-16

> # Check homogeneity of variance
> library(car)
> leveneTest(Cholesterol ~ Sex * ChestPainType, data = df)
Levene's Test for Homogeneity of Variance (center = median)
          Df F value    Pr(>F)
group       7  8.8289 1.54e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> # Post-hoc test
> TukeyHSD(anova2_model, "ChestPainType")
Tukey multiple comparisons of means
```

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins Project (None)

Source
Console Terminal Background Jobs

R 4.5.2 ~ /
> # Post-hoc test
> TukeyHSD(anova2_model, "ChestPainType")
Tukey multiple comparisons of means
95% Family-wise confidence level

Fit: aov(formula = Cholesterol ~ Sex * ChestPainType, data = df)

$ChestPainType
          diff      lwr      upr      p adj
ATA-ASY  35.358768  11.17032  59.547220 0.0010267
NAP-ASY   4.353639 -18.47125  27.178527 0.9611207
TA-ASY   16.326008 -25.89599  58.548009 0.7521936
NAP-ATA -31.005129 -59.35048 -2.659782 0.0256020
TA-ATA   -19.032759 -64.47702  26.411498 0.7031151
TA-NAP   11.972369 -32.76100  56.705743 0.9013730

>
> # Load required libraries
> library(ggplot2)
> library(dplyr)

Attaching package: 'dplyr'

The following object is masked from 'package:car':
  recode

The following objects are masked from 'package:stats':
  filter, lag

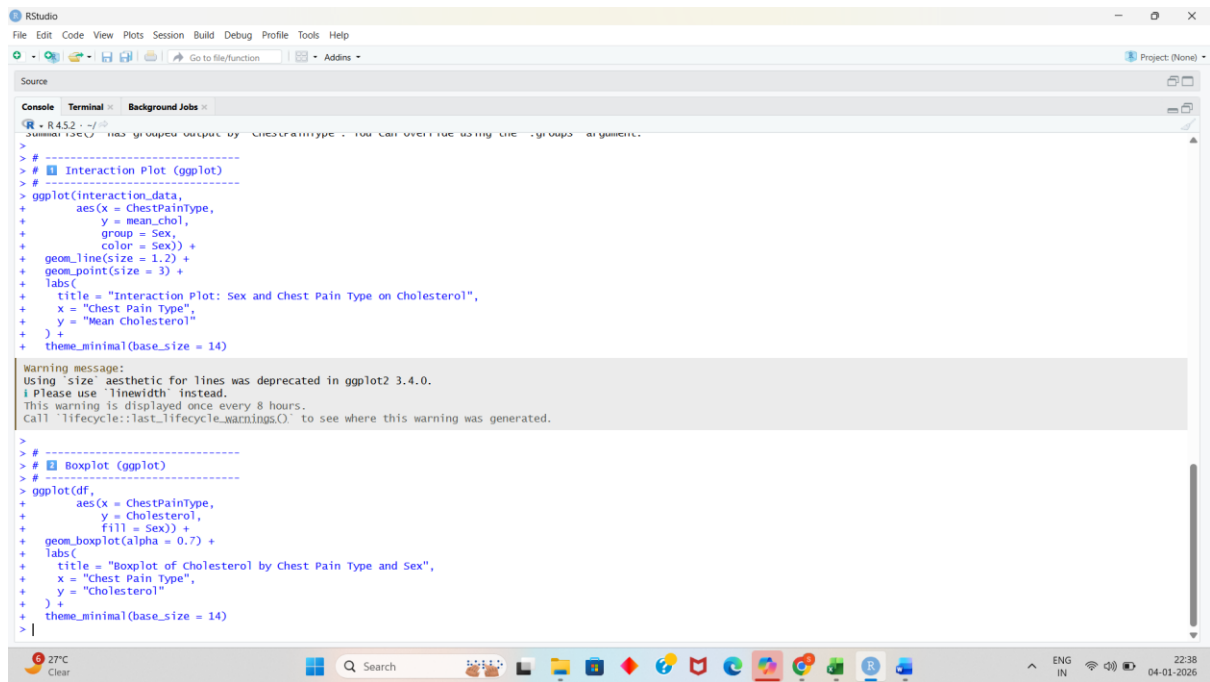
The following objects are masked from 'package:base':
  intersect, setdiff, setequal, union

>
> # Interaction plot data (mean cholesterol)
> interaction_data <- df %>%
+   group_by(ChestPainType, Sex) %>%
+   summarise(mean_cho = mean(Cholesterol, na.rm = TRUE))
```

Name :- Priya Gupta

Roll no. :- S081

Sheth L.U.J. & Sir M.V. College



```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins Project (None)
Source
Console Terminal Background Jobs
R 4.5.2 ~
Warning message:
'geom_line()' has grouped output by 'ChestPainType', but can't overruling the 'group' argument.
>
> # Interaction Plot (ggplot)
> # -----
> ggplot(interaction_data,
+       aes(x = ChestPainType,
+           y = mean_chol,
+           group = Sex,
+           color = Sex)) +
+   geom_line(size = 1.2) +
+   geom_point(size = 3) +
+   labs(
+     title = "Interaction Plot: Sex and Chest Pain Type on Cholesterol",
+     x = "Chest Pain Type",
+     y = "Mean Cholesterol"
+   ) +
+   theme_minimal(base_size = 14)
Warning message:
Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
! Please use 'linewidth' instead.
This warning is displayed once every 8 hours.
Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was generated.
>
> # Boxplot (ggplot)
> # -----
> ggplot(df,
+       aes(x = ChestPainType,
+           y = Cholesterol,
+           fill = Sex)) +
+   geom_boxplot(alpha = 0.7) +
+   labs(
+     title = "Boxplot of Cholesterol by Chest Pain Type and Sex",
+     x = "Chest Pain Type",
+     y = "Cholesterol"
+   ) +
+   theme_minimal(base_size = 14)
> |
```

Name :- Priya Gupta

Roll no. :- S081