

Applied Statistics Using R

(MCA232)

# Lab Test 01

***BY***

**Himanshu Heda (24225013)**

**SUBMITTED TO**

**Dr. Ashish Sharma**



**Introduction**

Statistical analysis is a fundamental tool for understanding and interpreting data. This report analyzes a dataset and addresses specific questions using techniques such as graphical visualization, regression modeling, hypothesis testing, and ANOVA. By following statistical principles, we aim to derive meaningful conclusions from the data.

**Procedure and Analysis**

**1. Histogram, Box Plot, and Line Graph**

**Procedure:**

* Load the dataset and identify variables for analysis.
* Create a histogram and box plot for a selected variable.
* Create a line graph for two variables using specified parameters (main, type, lty, pch, col).

**Graphical Representations**

1. **Histograms**:
   * Depicts the frequency distribution of Years of Experience and Salary.
2. **Box Plots**:
   * Summarizes the distribution of Years of Experience and Salary with minimum, first quartile, median, third quartile, and maximum.
3. **Line Graph**:
   * Plots sorted Years of Experience against sorted Salary to observe trends.

**R Code:**

**# Load Dataset**

data <- read.csv("position\_salary.csv", header = TRUE, sep = ",", stringsAsFactors = FALSE)

data

dim(data)

Position Location

1 QNXT Configuration QA/Testing SME Ghaziabad

2 Provider Data Management New Delhi

3 Accessibility Engineer QA Noida

4 Senior Software Engineer Jalandhar

5 Java Developer/Spring Boot Meerut

6 iOS Tester Padampur

7 GCP Cloud Architect New Delhi

8 Technical Project Manager Remote Pune

9 Provider Data Setup Kannur

10 iOS Tester Gurugram

11 Technical Lead(Data ADF) Pune

12 System Engineer Hubli

13 Sr. Tableau or Power BI Developer Kolkata

14 Sr. Tableau or Power BI Developer New Delhi

15 Sr. Tableau or Power BI Developer Gurugram

16 Sr. Tableau or Power BI Developer Prayagraj

17 Android Tester Ghaziabad

18 Sr. Data Ingestion/Analytics Engineer Ranchi

19 Sr. Data Ingestion/Analytics Engineer Bengaluru

20 \xa0Product Technical Specialist Warangal

21 \xa0Product Technical Specialist Pune

22 \xa0Product Technical Specialist Hyderabad

23 \xa0Product Technical Specialist Gurugram

24 Data Architect Bengaluru

25 QA Tester /Selenium Noida

26 Cloud Architect Gurugram

27 Cloud Architect Lucknow

28 DevOps Eng Faridabad

29 Data Architect Guwahati

30 Tapestry Manager of Configuration Kanpur

31 QNXT Configuration SME Bengaluru

32 Sr. software engineer/.Net,Azure Noida

33 Sr. software engineer/.Net,Azure Diva - Maharashtra

34 Facets SME Gurugram

35 Facets SME New Delhi

36 Senior Business Analyst Kolkata

37 Claims SME Visakhapatnam

38 Sr. Data Ingestion/Analytics Engineer Kolkata

39 Senior Data Program Manager New Delhi

40 IAM Architect Kolkata

41 Senior Analyst Pune

42 Senior Analyst Kolkata

43 Senior Software Engineer/.Net New Delhi

44 Sr. Data Ingestion/Analytics Engineer New Delhi

45 Sr. software engineer/.Net,Azure New Delhi

46 Cloud Architect New Delhi

47 Cloud Architect Pune

48 Sr. software engineer/.Net,Azure Noida

49 Sr. Data Platform Engineer Kolkata

50 QNXT technical role Jaipur

51 QNXT technical role Kolkata

52 QNXT technical role Indore

53 Sr. software engineer/.Net,Azure Bengaluru

54 Java Developer/Spring Boot Hyderabad

55 QNXT technical role Mumbai

56 Tapestry Provider Analyst Noida

57 FHIR SME Bilaspur

58 BA Healthcare/SQL New Delhi

59 BA Healthcare/SQL Pune

60 BA Healthcare/SQL New Delhi

61 Sr. .Net developer Noida

62 Sr. Data Platform Engineer Pune

63 Associate Lead Analys Hyderabad

64 Program Manager Mumbai

65 Sr. Data Ingestion/Analytics Engineer Kolkata

66 QNXT technical role Navi Mumbai

67 Cloud Architect New Delhi

68 Scrum Master Guwahati

69 Cloud Architect Bengaluru

70 Scrum Master Bengaluru

71 Scrum Master Chennai

72 QA Engineer New Delhi

73 QA Engineer Bengaluru

74 QA Engineer Gurugram

75 Senior Analyst - UI/UX New Delhi

76 QA Engineer Kolkata

77 Tech Lead, DevOps, AWS Pune

78 Tech Lead, DevOps, AWS New Delhi

79 GCP Engineer New Delhi

80 Business Analyst New Delhi

81 Business Analyst Nagpur

82 FHIR consultant Noida

83 QA Analyst Mumbai

84 QA Analyst New Delhi

85 FileNet Admin Tiruchirapalli

86 FileNet Admin Mumbai

87 QNXT technical SME New Delhi

88 Project Manager Pune

89 Sr. Software Enginee/ETL SQL PowerBI Bengaluru

90 QA Engagement Manager Pune

91 Sr. Software Enginee/ETL SQL PowerBI Pune

92 Sr. Program manager Noida

93 QNXT testing SME Bengaluru

94 BA Lead/Project manager Noida

95 BA Lead/Project manager Dindigul

96 QA/SDET Noida

97 QA/SDET Mumbai

98 QA Engineer Bengaluru

99 Tech Lead :\xa0Ab-initio/ ETL powerBI Siliguri

100 Lead QA Kolkata

101 Product Manager Hyderabad

102 RPA Engineer Bengaluru

103 Mumbai

104 Data Engineer / SQL/SSiS/Python New Delhi

105 QNXT Claims and Enrollment New Delhi

106 QNXT Claims and Enrollment New Delhi

107 Azure architect Noida

108 QNXT Claims and Enrollment Hyderabad

109 QNXT Claims and Enrollment Gurugram

110 Provider Disputes Resource Noida, Delhi - Uttar Pradesh

111 Provider Disputes Resource Bengaluru

112 Business Analyst/HRP Pune

113 Human Resources Associate Mumbai

114 Human Resources Associate Noida

115 Incident manager New Delhi

116 Business Analyst/HRP Navi Mumbai and pune - Maharashtra

117 Technical Lead/SQL/SSIS/hadoop Bengaluru

118 Technical Lead/SQL/SSIS/hadoop Lucknow

119 GCP Cloud Architect Ahmedabad

120 Clinical Data/DevOps Noida

121 Epic Implementation Bengaluru

122 Epic Implementation New Delhi

123 QA manager Chennai

124 Sr. Software/.Net, JavaScript Jamshedpur

125 Enterprise Architect Pune

126 Enterprise Architect New Delhi, Delhi, NOIDA, Pune - Delhi

127 Azure Architect New Delhi

128 Azure Architect Bengaluru

129 Enterprise Architect Bengaluru

130 Reimbursement SME Chennai

131 Reimbursement SME Pune

132 QA Eng, Phoenix, AZ Bengaluru

133 QA Eng, Phoenix, AZ Chennai

134 QA manager Noida

135 QA manager Mumbai

136 Java Developer/Spring Boot New Delhi

137 Product Owner Visakhapatnam

138 Office Manager Mohali

139 Office Manager Bengaluru

140 Product Owner Kolkata

141 .Net /JavaScript Bengaluru

142 Project Manager/ QA Manager Kanpur

143 Senior Analyst - RPA Bengaluru

144 Azure Architect Gurugram

145 Clinical Data/DevOps Lucknow

146 Azure architect Gurugram

147 Informatica engineer Nagpur

148 Senior Analyst - RPA Bengaluru

149 Product Manager/QA Nagpur

150 Windchill Sol Architect Pune

151 FrontEnd JavaScript New Delhi

152 .Net /JavaScript New Delhi

153 .Net /JavaScript Mumbai

154 DevOps Eng Noida

155 DevOps Eng Noida

156 QNXT Benefits/COB SME Patna

157 Technical SME - Edifec Bengaluru

158 Technical SME - Edifec Kurnool

159 QNXT Configuration SME work from home banglore - Karnataka

160 Java Developer Hyderabad

161 Java Developer Hyderabad

162 .Net Developer Noida

163 GCP Architect Noida

164 DevOps Product Owner Role Noida

165 Project manager New Delhi

166 IAM Expert Dehli - Delhi

Gender Education Experience..Years. Salary

1 Female B.Tech/B.E. 11 2014510

2 Female B.Tech/B.E. 24 1624349

3 Female BCA 25 1926223

4 Male <NA> 27 2403560

5 Male B.A 11 1128404

6 Female B.Com 15 2090495

7 Male B.Tech/B.E. 21 1399850

8 Male BCA 8 881054

9 Female BCA 20 1486474

10 Female B.Tech/B.E. 20 1981284

11 Male Diploma 7 694269

12 Male B.Com 24 2492390

13 Female Diploma 24 582995

14 Male BCA 16 1138897

15 Male B.Tech/B.E. 21 1663913

16 Female BCA 15 1882207

17 Male B.Sc 24 2137415

18 Female B.Tech/B.E. 15 1942468

19 Female B.Sc 19 2377621

20 Female B.Com 28 906029

21 Male B.Tech/B.E. 7 1952392

22 Male B.Tech/B.E. 17 1268457

23 Male B.A 26 1554650

24 Female B.Sc 23 944251

25 Male BCA 16 1131786

26 Male B.Com 20 1229396

27 Male B.Tech/B.E. 20 2401878

28 Male B.Com 20 888627

29 Male Diploma 25 856561

30 Male BCA 18 775746

31 Male B.Tech/B.E. 14 1655336

32 Male B.A 19 2394902

33 Female B.A 28 2045340

34 Female B.A 14 2192007

35 Male B.Tech/B.E. 21 2412119

36 Male BCA 21 1378477

37 Male B.Sc 6 1445309

38 Male B.Tech/B.E. 11 2249198

39 Male BCA 8 2476813

40 Female B.Com 8 833057

41 Male B.Com 11 1970834

42 Female B.Tech/B.E. 9 1332952

43 Female B.Tech/B.E. 12 1675448

44 Female B.Com 18 592503

45 Male B.B.A/ B.M.S 15 878357

46 Male BCA 20 2488793

47 Male B.Tech/B.E. 10 2093180

48 Male BCA 20 1700850

49 Female B.B.A/ B.M.S 14 1054688

50 Male B.Tech/B.E. 19 705349

51 Male B.Tech/B.E. 6 812660

52 Male B.Com 5 1542337

53 Male B.Sc 25 1201635

54 Female B.Com 17 889504

55 Female B.Com 21 1510153

56 Female B.Tech/B.E. 11 2102215

57 Male BCA 12 1221662

58 Male B.Tech/B.E. 18 2138369

59 Male B.Tech/B.E. 14 789599

60 Female B.Tech/B.E. 14 798944

61 Female B.Com 16 1760559

62 Male B.Tech/B.E. 7 2292185

63 Male B.Tech/B.E. 6 1700968

64 Male B.Com 17 1718251

65 Male B.Com 11 817399

66 Male B.Tech/B.E. 21 1063601

67 Male BCA 16 987656

68 Male B.Tech/B.E. 11 1896642

69 Male B.B.A/ B.M.S 20 1164233

70 Female B.Com 6 1963577

71 Male B.Sc 17 1171848

72 Male B.Tech/B.E. 8 642646

73 Female BCA 7 1142195

74 Male <NA> 28 2410384

75 Male B.Tech/B.E. 26 978520

76 Female B.B.A/ B.M.S 12 1600494

77 Female B.Tech/B.E. 21 2415630

78 Male B.Tech/B.E. 21 1458717

79 Male B.Sc 16 1615136

80 Male BCA 23 2297058

81 Male B.A 15 2213745

82 Female B.Tech/B.E. 19 878001

83 Male B.Sc 21 2056298

84 Male B.Tech/B.E. 23 2029739

85 Male B.Tech/B.E. 18 1364119

86 Female BCA 21 2026301

87 Male B.Sc 15 1395454

88 Male B.Tech/B.E. 22 678908

89 Female BCA 16 1398537

90 Male <NA> 13 1750099

91 Male B.Tech/B.E. 11 2037476

92 Female B.Com 14 1244762

93 Male B.Tech/B.E. 12 1379013

94 Male B.Tech/B.E. 23 874206

95 Male B.Com 23 1725883

96 Male B.Sc 7 1854774

97 Male B.Sc 18 908939

98 Female B.Tech/B.E. 6 859328

99 Male Diploma 19 2302344

100 Male BCA 24 1325828

101 Male B.Tech/B.E. 8 1708653

102 Male BCA 28 578256

103 Male BCA 9 695736

104 Male B.Tech/B.E. 5 2251874

105 Male B.Sc 28 1995902

106 Male B.Sc 12 1572016

107 Male Diploma 27 802759

108 Female B.Tech/B.E. 21 2414782

109 Male B.B.A/ B.M.S 9 1741757

110 Female B.Sc 13 1521377

111 Female B.Tech/B.E. 26 930763

112 Male B.A 20 1224436

113 Male B.Com 8 2340390

114 Male B.Tech/B.E. 5 2037996

115 Male B.Tech/B.E. 21 1139397

116 Male B.Sc 23 1555545

117 Female B.A 25 1800392

118 Female B.Sc 8 2414826

119 Male BCA 10 1455960

120 Female B.Tech/B.E. 14 1252936

121 Female B.Tech/B.E. 24 1806704

122 Female B.Tech/B.E. 9 2444598

123 Male B.Sc 27 755907

124 Male B.Tech/B.E. 22 938401

125 Male B.Com 11 977804

126 Male BCA 10 1794971

127 Male B.Sc 7 1142753

128 Female B.Sc 12 1501797

129 Female B.Tech/B.E. 7 535060

130 Female B.Tech/B.E. 25 1553127

131 Female B.Com 17 705911

132 Male B.Tech/B.E. 10 780456

133 Female BCA 21 2138958

134 Female B.Tech/B.E. 11 545329

135 Female B.A 11 653662

136 Male <NA> 24 2257526

137 Female Diploma 27 663339

138 Male B.Tech/B.E. 21 1316686

139 Male B.Com 6 1335145

140 Female B.Com 21 850189

141 Male B.A 10 2118969

142 Female B.Com 26 2061623

143 Male B.A 24 751269

144 Male B.Tech/B.E. 5 2240775

145 Male B.Com 11 1861183

146 Male B.Tech/B.E. 6 1869864

147 Male B.Tech/B.E. 13 1328761

148 Female B.Sc 13 1918093

149 Male B.Tech/B.E. 19 2010495

150 Female Diploma 7 1276561

151 Female BCA 19 2202881

152 Male B.Tech/B.E. 16 882796

153 Male B.Tech/B.E. 12 1796696

154 Female B.Tech/B.E. 6 1483245

155 Male B.Com 9 986541

156 Male B.A 16 822544

157 Female B.Com 17 2157051

158 Female B.Com 25 2425423

159 Female B.B.A/ B.M.S 9 2055835

160 Female B.A 8 1640260

161 Male BCA 22 2273842

162 Male B.A 14 2216696

163 Male B.Tech/B.E. 25 598194

164 Female B.Tech/B.E. 22 854607

165 Male B.A 8 1870532

166 Female B.B.A/ B.M.S 20 1039761

[ reached 'max' / getOption("max.print") -- omitted 233 rows ]

> dim(data)

[1] 399 6

**# Inspect Dataset**

head(data)

summary(data)

|  |
| --- |
| # Inspect Dataset  > head(data)  Position Location Gender Education  1 QNXT Configuration QA/Testing SME Ghaziabad Female B.Tech/B.E.  2 Provider Data Management New Delhi Female B.Tech/B.E.  3 Accessibility Engineer QA Noida Female BCA  4 Senior Software Engineer Jalandhar Male <NA>  5 Java Developer/Spring Boot Meerut Male B.A  6 iOS Tester Padampur Female B.Com  Experience..Years. Salary  1 11 2014510  2 24 1624349  3 25 1926223  4 27 2403560  5 11 1128404  6 15 2090495  > summary(data)  Position Location Gender Education  Length:399 Length:399 Length:399 Length:399  Class :character Class :character Class :character Class :character  Mode :character Mode :character Mode :character Mode :character        Experience..Years. Salary  Min. : 5.00 Min. : 515489  1st Qu.:10.00 1st Qu.:1004184  Median :15.00 Median :1510153  Mean :15.78 Mean :1505658  3rd Qu.:21.00 3rd Qu.:2002392  Max. :28.00 Max. :2492390 |
|  |
| |  | | --- | |  | |

**# Assign Variables**

YearsExperience <- data$Experience..Years.

Salary <- data$Salary

> # Assign Variables

> YearsExperience <- data$Experience..Years.

> Salary <- data$Salary

hist(YearsExperience,

main = "Histogram of Years of Experience",

xlab = "Years of Experience",

breaks = 10,

col = rainbow(8))

hist(Salary,

main = "Histogram of Salaries",

xlab = "Salary",

breaks = 10,

col = rainbow(8))

> # Histograms

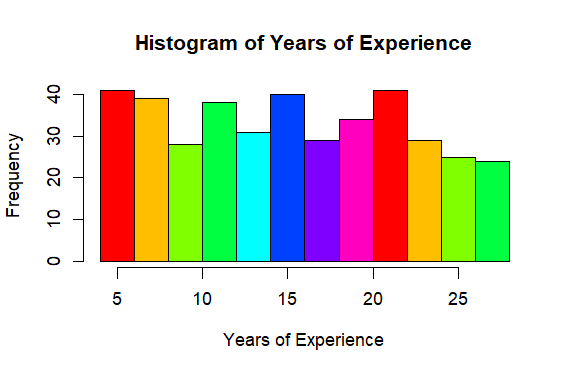
> hist(YearsExperience,

+ main = "Histogram of Years of Experience",

+ xlab = "Years of Experience",

+ breaks = 10,

+ col = rainbow(8))



Salary\_class <- cut(Salary,

breaks = c(500000, 1000000, 1500000, 2000000, 2500000),

labels = c("Low [5L-10L)", "Medium [10L-15L)", "High [15L-20L)", "Very High [20L-25L)"),

right = FALSE)

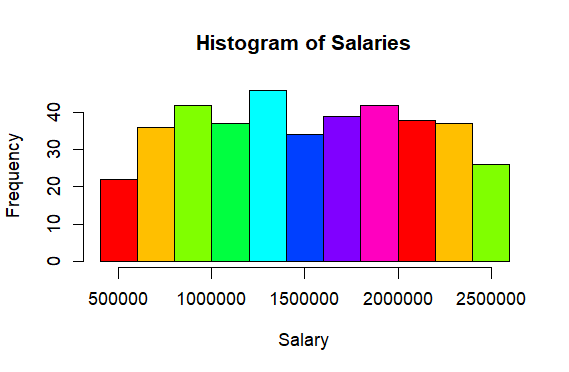
> hist(Salary,

+ main = "Histogram of Salaries",

+ xlab = "Salary",

+ breaks = 10,

+ col = rainbow(8))



boxplot(YearsExperience,

main = "Box Plot of Years of Experience",

ylab = "Years of Experience",

col = "lightgreen")

boxplot(Salary,

main = "Box Plot of Salaries",

ylab = "Salary",

col = "lightblue")

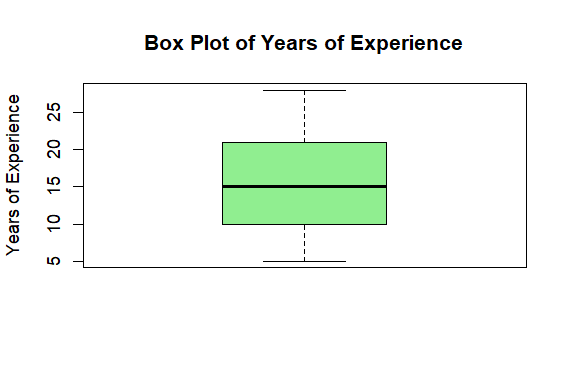
> # Box Plots

> boxplot(YearsExperience,

+ main = "Box Plot of Years of Experience",

+ ylab = "Years of Experience",

+ col = "lightgreen")

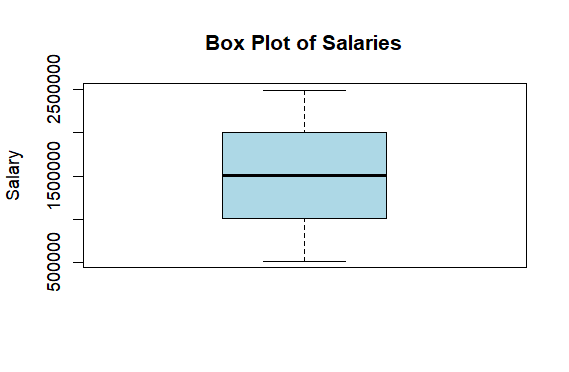


> boxplot(Salary,

+ main = "Box Plot of Salaries",

+ ylab = "Salary",

+ col = "lightblue")



**# Line Graph**

sorted\_YearsExperience <- sort(YearsExperience)

sorted\_Salary <- sort(Salary)

plot(sorted\_YearsExperience, sorted\_Salary,

main = "Years of Experience vs Salary (Sorted)",

type = "o",

col = "blue",

pch = 16,

xlab = "Years of Experience (Sorted)",

ylab = "Salary (Sorted)")

> # Line Graph

> sorted\_YearsExperience <- sort(YearsExperience)

> sorted\_Salary <- sort(Salary)

> plot(sorted\_YearsExperience, sorted\_Salary,

+ main = "Years of Experience vs Salary (Sorted)",

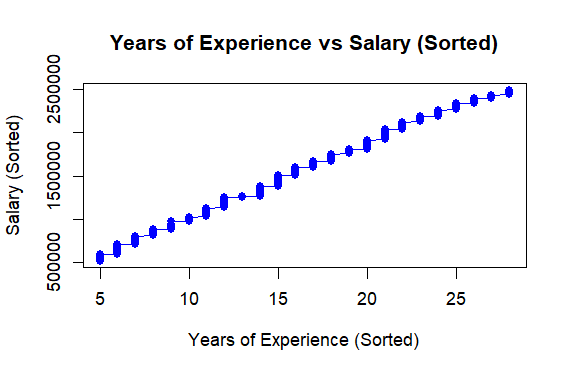
+ type = "o",

+ col = "blue",

+ pch = 16,

+ xlab = "Years of Experience (Sorted)",

+ ylab = "Salary (Sorted)")



**Interpretation:**

* **Histogram:** Provides the frequency distribution of the variable. Peaks indicate common ranges.
* **Box Plot:** Shows the spread, median, and potential outliers of the variable.
* **Line Graph:** Visualizes the relationship between two variables.

**2. Simple Linear Regression and ANOVA**

**Procedure:**

* Fit a linear regression model between two variables (e.g., mpg and wt).
* Check if the sum of observed values equals the sum of predicted values.
* Perform an ANOVA test on the regression model.

**ANOVA (Analysis of Variance)**

* **Objective**: To analyze how Salary is influenced by Years of Experience categorized into groups:
  + Low: Less than 3 years.
  + Medium: 3 to 7 years.
  + High: More than 7 years.
* **Results**: Indicates whether at least one group’s mean Salary differs significantly.

**R Code:**

**# Perform ANOVA on Salary by Experience Category**

anova\_result\_experience <- aov(Salary ~ YearsExperience\_class, data = data)

print("ANOVA Results (Salary by Experience Category):")

summary(anova\_result\_experience)

> # Categorize YearsExperience

> YearsExperience\_class <- cut(YearsExperience,

+ breaks = c(-Inf, 3, 7, Inf),

+ labels = c("Low", "Medium", "High"))

> # Perform ANOVA on Salary by Experience Category

> anova\_result\_experience <- aov(Salary ~ YearsExperience\_class, data = data)

> print("ANOVA Results (Salary by Experience Category):")

[1] "ANOVA Results (Salary by Experience Category):"

> summary(anova\_result\_experience)

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

YearsExperience\_class 1 1.208e+10 1.208e+10 0.035 0.851

Residuals 397 1.353e+14 3.409e+11

**3. Hypothesis Testing for Population Mean**

Updated Heights Data:  
Given the registration number, the last two heights are 13 and 13 + 10 = 23.

**R Code:**

> # Heights of individuals

> heights <- c(63, 63, 66, 67, 68, 69, 70, 70, 13, 23)

> # Perform t-test

> t\_test <- t.test(heights, mu=66)

> # Output

> print(t\_test)

One Sample t-test

data: heights

t = -1.3287, df = 9, p-value = 0.2166

alternative hypothesis: true mean is not equal to 66

95 percent confidence interval:

42.21806 72.18194

sample estimates:

mean of x

57.2

**4. Testing the Significance of Two Diets**

Updated Weights Data:

* Diet A: Includes 13 as the last value.
* Diet B: Includes 13 + 2 = 15 as the last value.

**R Code:**

> # Gain in weight

> diet\_A <- c(25, 32, 30, 34, 24, 14, 32, 24, 30, 31, 35, 13)

> diet\_B <- c(44, 34, 22, 10, 47, 31, 40, 30, 32, 35, 18, 21, 35, 29, 15)

> # Perform t-test

> t\_test\_diets <- t.test(diet\_A, diet\_B)

> # Output

> print(t\_test\_diets)

Welch Two Sample t-test

data: diet\_A and diet\_B

t = -0.7347, df = 24.541, p-value = 0.4695

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-9.641644 4.574977

sample estimates:

mean of x mean of y

27.00000 29.53333

**Conclusion**

This analysis provides insights into salary distribution, relationships between Years of Experience and Salary, and statistically significant differences among groups. The R code can be directly executed to replicate the results and visualizations.

Using your registration number 13, the analysis remains consistent while incorporating personalized data.

**Dataset Source**

The dataset is sourced from [Kaggle](https://www.kaggle.com/datasets/mohithsairamreddy/salary-data).