



# **Applied Statistics Using R** **(MCA232)**

## **Lab Practical 1**

*BY*

**Himanshu Heda (24225013)**

**SUBMITTED TO**

**Dr. Ashish Sharma**

**SCHOOL OF SCIENCES**

**2024-25**

## Dataset Information

The dataset `position_salary.csv` includes data about employees' years of experience and their corresponding salaries. This analysis focuses on exploring descriptive statistics, graphical representations, and statistical tests such as Two-Sample Z Tests and ANOVA.

---

## Descriptive Statistics

### 1. Mean:

- **Years of Experience:** Calculated to understand the average experience.
- **Salary:** Represents the average salary.

### 2. Median:

- The central value for both Years of Experience and Salary.

### 3. Mode:

- The most frequently occurring values in Years of Experience and Salary.

### 4. Standard Deviation:

- Measures the spread or variability of the data for Years of Experience and Salary.
- 

## Graphical Representations

### 1. Histograms:

- Depicts the frequency distribution of Years of Experience and Salary.

### 2. Pie Chart:

- Shows the distribution of salaries categorized into ranges (e.g., Low, Medium, High, Very High).

### 3. Scatter Plot:

- Visualizes the relationship between Years of Experience and Salary. Includes a regression line to highlight trends.

### 4. Box Plots:

- Summarizes the distribution of Years of Experience and Salary with minimum, first quartile, median, third quartile, and maximum.

### 5. Line Graph:

- Plots sorted Years of Experience against sorted Salary to observe trends.
-

## Two-Sample Z Test

- **Objective:** To compare the mean salaries of two groups based on Years of Experience ( $\leq$  median vs  $>$  median).
  - **Results:** This test determines if there is a significant difference in the mean salaries between these groups.
- 

## 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

### # Load Dataset

```
data <- read.csv("position_salary.csv", header = TRUE, sep = ",", stringsAsFactors = FALSE)
```

```
data
```

```
dim(data)
```

	Position	Loc
ation		
1	QNXT Configuration QA/Testing SME	Ghaz
iabad		
2	Provider Data Management	New
Delhi		
3	Accessibility Engineer QA	
Noida		
4	Senior Software Engineer	Jala
ndhar		
5	Java Developer/Spring Boot	M
eerut		
6	iOS Tester	Pad
ampur		
7	GCP Cloud Architect	New
Delhi		
8	Technical Project Manager Remote	
Pune		
9	Provider Data Setup	K
annur		
10	iOS Tester	Gur
ugram		
11	Technical Lead(Data ADF)	
Pune		
12	System Engineer	
Hubli		

13	Sr. Tableau or Power BI Developer	Ko
14	Sr. Tableau or Power BI Developer	New
15	Sr. Tableau or Power BI Developer	Gur
16	Sr. Tableau or Power BI Developer	Pray
17	Android Tester	Ghaz
18	Sr. Data Ingestion/Analytics Engineer	R
19	Sr. Data Ingestion/Analytics Engineer	Beng
20	\xa0Product Technical Specialist	War
21	\xa0Product Technical Specialist	
22	\xa0Product Technical Specialist	Hyde
23	\xa0Product Technical Specialist	Gur
24	Data Architect	Beng
25	QA Tester /Selenium	
26	Cloud Architect	Gur
27	Cloud Architect	Lu
28	DevOps Eng	Fari
29	Data Architect	Guw
30	Tapestry Manager of Configuration	K
31	QNXT Configuration SME	Beng
32	Sr. software engineer/.Net,Azure	
33	Sr. software engineer/.Net,Azure	Diva - Mahara
34	Facets SME	Gur
35	Facets SME	New
36	Senior Business Analyst	Ko
37	Claims SME	Visakhap
38	Sr. Data Ingestion/Analytics Engineer	Ko
39	Senior Data Program Manager	New
40	IAM Architect	Ko
41	Senior Analyst	
42	Senior Analyst	Ko
43	Senior Software Engineer/.Net	New
44	Sr. Data Ingestion/Analytics Engineer	New
45	Sr. software engineer/.Net,Azure	New
46	Cloud Architect	New

47	Cloud Architect	
Pune		
48	Sr. software engineer/.Net,Azure	
Noida		
49	Sr. Data Platform Engineer	Ko
lkata		
50	QNXT technical role	J
aipur		
51	QNXT technical role	Ko
lkata		
52	QNXT technical role	I
ndore		
53	Sr. software engineer/.Net,Azure	Beng
aluru		
54	Java Developer/Spring Boot	Hyde
rabad		
55	QNXT technical role	M
umbai		
56	Tapestry Provider Analyst	
Noida		
57	FHIR SME	Bi
aspur		
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	Hyde
rabad		
64	Program Manager	M
umbai		
65	Sr. Data Ingestion/Analytics Engineer	Ko
lkata		
66	QNXT technical role	Navi M
umbai		
67	Cloud Architect	New
Delhi		
68	Scrum Master	Guw
ahati		
69	Cloud Architect	Beng
aluru		
70	Scrum Master	Beng
aluru		
71	Scrum Master	Ch
ennai		
72	QA Engineer	New
Delhi		
73	QA Engineer	Beng
aluru		
74	QA Engineer	Gur
ugram		
75	Senior Analyst - UI/UX	New
Delhi		
76	QA Engineer	Ko
lkata		
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	N
agpur		
82	FHIR consultant	
Noida		
83	QA Analyst	M
umbai		
84	QA Analyst	New
Delhi		
85	FileNet Admin	Tiruchira
palli		
86	FileNet Admin	M
umbai		
87	QNXT technical SME	New
Delhi		
88	Project Manager	
Pune		
89	Sr. Software Engineer/ETL SQL PowerBI	Beng
aluru		
90	QA Engagement Manager	
Pune		
91	Sr. Software Engineer/ETL SQL PowerBI	
Pune		
92	Sr. Program manager	
Noida		
93	QNXT testing SME	Beng
aluru		
94	BA Lead/Project manager	
Noida		
95	BA Lead/Project manager	Din
digul		
96	QA/SDET	
Noida		
97	QA/SDET	M
umbai		
98	QA Engineer	Beng
aluru		
99	Tech Lead : \xa0Ab-initio/ ETL powerBI	Sil
iguri		
100	Lead QA	Ko
lkata		
101	Product Manager	Hyde
rabad		
102	RPA Engineer	Beng
aluru		
103		M
umbai		
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	Hyde
rabad		
109	QNXT Claims and Enrollment	Gur
ugram		
110	Provider Disputes Resource	Noida, Delhi - Uttar Pr
adesh		
111	Provider Disputes Resource	Beng
aluru		
112	Business Analyst/HRP	
Pune		
113	Human Resources Associate	M
umbai		
114	Human Resources Associate	
Noida		

115 Delhi	Incident manager	New
116 shtra	Business Analyst/HRP	Navi Mumbai and pune - Mahara
117 aluru	Technical Lead/SQL/SSIS/hadoop	Beng
118 cknow	Technical Lead/SQL/SSIS/hadoop	Lu
119 dabad	GCP Cloud Architect	Ahme
120 Noida	Clinical Data/DevOps	
121 aluru	Epic Implementation	Beng
122 Delhi	Epic Implementation	New
123 ennai	QA manager	Ch
124 edpur	Sr. Software/.Net, JavaScript	Jamsh
125 Pune	Enterprise Architect	
126 Delhi	Enterprise Architect	New Delhi, Delhi, NOIDA, Pune -
127 Delhi	Azure Architect	New
128 aluru	Azure Architect	Beng
129 aluru	Enterprise Architect	Beng
130 ennai	Reimbursement SME	Ch
131 Pune	Reimbursement SME	
132 aluru	QA Eng, Phoenix, AZ	Beng
133 ennai	QA Eng, Phoenix, AZ	Ch
134 Noida	QA manager	
135 umbai	QA manager	M
136 Delhi	Java Developer/Spring Boot	New
137 atnam	Product Owner	Visakhap
138 ohali	Office Manager	M
139 aluru	Office Manager	Beng
140 lkata	Product Owner	Ko
141 aluru	.Net /JavaScript	Beng
142 anpur	Project Manager/ QA Manager	K
143 aluru	Senior Analyst - RPA	Beng
144 ugram	Azure Architect	Gur
145 cknow	Clinical Data/DevOps	Lu
146 ugram	Azure architect	Gur
147 agpur	Informatica engineer	N
148 aluru	Senior Analyst - RPA	Beng

149	Product Manager/QA	N
agpur		
150	windchill Sol Architect	
Pune		
151	FrontEnd JavaScript	New
Delhi		
152	.Net /JavaScript	New
Delhi		
153	.Net /JavaScript	M
umbai		
154	DevOps Eng	
Noida		
155	DevOps Eng	
Noida		
156	QNXT Benefits/COB SME	
Patna		
157	Technical SME - Edifec	Beng
aluru		
158	Technical SME - Edifec	Ku
rnool		
159	QNXT Configuration SME	work from home banglore - Karn
ataka		
160	Java Developer	Hyde
rabad		
161	Java Developer	Hyde
rabad		
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
```

## # Descriptive Statistics

### # Mean

```
print("Mean of YearsExperience")
```

```
mean(YearsExperience, na.rm = TRUE)
```

```
print("Mean of Salary")
```

```
mean(Salary, na.rm = TRUE)
```

```
> # Mean
> print("Mean of YearsExperience")
[1] "Mean of YearsExperience"
> mean(YearsExperience, na.rm = TRUE)
[1] 15.77945
> print("Mean of Salary")
[1] "Mean of Salary"
> mean(Salary, na.rm = TRUE)
[1] 1505658
```

### # Median

```
print("Median of YearsExperience")
```

```
median(YearsExperience, na.rm = TRUE)
```

```
print("Median of Salary")
```

```
median(Salary, na.rm = TRUE)
```

```
> # Median
> print("Median of YearsExperience")
[1] "Median of YearsExperience"
> median(YearsExperience, na.rm = TRUE)
[1] 15
> print("Median of Salary")
[1] "Median of Salary"
> median(Salary, na.rm = TRUE)
[1] 1510153
```

### # Mode Function

```
modes <- function(x) {
```

```
  freq_table <- table(x)
```

```
  mode_values <- as.numeric(names(freq_table)[freq_table == max(freq_table)])
```

```
  return(mode_values)
```

```
}
```

```
print("Mode of YearsExperience")
```

```
modes(YearsExperience)
```

```
print("Mode of Salary")
```

```
modes(Salary)
```

```
> # Mode Function
> modes <- function(x) {
+   freq_table <- table(x)
+   mode_values <- as.numeric(names(freq_table)[freq_table == max(freq_table)])
+   return(mode_values)
+ }
> print("Mode of YearsExperience")
[1] "Mode of YearsExperience"
> modes(YearsExperience)
[1] 21
> print("Mode of Salary")
[1] "Mode of Salary"
> modes(Salary)
[1] 515489 530692 532410 533191 535060 539460 545329 548047 551
137
[10] 551219 553169 554511 556612 569786 578256 578682 582464 582
995
[19] 587655 592503 598194 598312 608000 620841 621536 623617 633
340
[28] 642646 653662 661279 663339 671437 678908 685356 688984 694
269
[37] 695736 698544 703837 705186 705349 705911 712445 718306 740
061
[46] 751269 755907 758281 766245 766924 775746 777869 780329 780
456
[55] 788875 789599 794000 798944 800433 802759 804674 809243 812
660
[64] 817399 822544 826939 830063 833057 833359 850189 854064 854
607
[73] 856561 859328 865345 874206 878001 878357 879639 881054 882
796
[82] 888627 889504 899342 904644 905051 906029 908939 915391 930
763
[91] 938401 944251 945414 968911 977804 978520 985312 986541 987
656
[100] 997365 1011002 1011333 1011343 1013549 1016436 1019514 1020213 1025
725
[109] 1026852 1028278 1039761 1049817 1050071 1054688 1056495 1063016 1063
601
[118] 1067468 1071751 1074945 1082026 1095608 1108373 1120742 1128404 1131
786
[127] 1138897 1139397 1142195 1142389 1142753 1158316 1164233 1167760 1171
848
[136] 1179260 1191142 1201635 1205394 1209856 1221662 1224436 1225241 1229
396
[145] 1244762 1249104 1250257 1251058 1251626 1252936 1253886 1257089 1258
974
[154] 1263310 1264979 1265332 1268457 1272054 1276561 1282043 1283751 1287
967
[163] 1304850 1308430 1316686 1325478 1325828 1328761 1332952 1335145 1348
353
[172] 1348590 1348788 1364119 1377599 1378477 1379013 1379842 1382564 1392
856
[181] 1395454 1398537 1399850 1428812 1445309 1446629 1446963 1448888 1455
960
[190] 1458717 1462569 1480718 1483245 1483747 1486474 1493992 1497200 1499
593
[199] 1501797 1510153 1510908 1516972 1517253 1521377 1542337 1553127 1554
650
[208] 1555545 1572016 1572680 1576398 1577275 1577892 1581824 1592561 1597
811
```

```

[217] 1598332 1600494 1615136 1617207 1624349 1627468 1640260 1649713 1652
443
[226] 1652571 1652741 1655336 1657979 1663913 1663943 1668249 1668337 1675
448
[235] 1683475 1683841 1700850 1700968 1708653 1718251 1721237 1725883 1741
757
[244] 1747631 1750099 1750189 1757813 1759763 1760559 1762519 1778809 1783
524
[253] 1784264 1790518 1794971 1796696 1800392 1802048 1806188 1806704 1807
102
[262] 1812467 1816082 1820547 1825763 1826113 1843298 1844740 1844907 1845
233
[271] 1847712 1854774 1861183 1869864 1870532 1871892 1882207 1896642 1916
954
[280] 1918093 1926223 1927489 1930000 1931716 1932713 1942468 1952392 1954
773
[289] 1963577 1966278 1970834 1972994 1981284 1981625 1988284 1989299 1991
056
[298] 1995902 2002160 2002624 2010495 2014510 2017013 2024984 2026301 2029
739
[307] 2037291 2037476 2037996 2040819 2045340 2047397 2055835 2056298 2061
623
[316] 2087382 2090495 2093180 2102215 2117099 2118969 2130784 2137415 2138
369
[325] 2138958 2139392 2143072 2147291 2157051 2170896 2172076 2180848 2186
040
[334] 2192007 2196469 2199608 2202881 2206343 2206645 2207956 2213745 2216
696
[343] 2236449 2240698 2240775 2249198 2251874 2257526 2258895 2261448 2271
572
[352] 2273842 2279947 2292185 2297058 2297251 2300210 2301642 2302344 2316
375
[361] 2323838 2325612 2337100 2340390 2345967 2359640 2371516 2377383 2377
621
[370] 2381053 2386009 2393549 2394902 2401547 2401878 2403560 2410384 2412
119
[379] 2414782 2414826 2415630 2425423 2429554 2441052 2443835 2444598 2457
504
[388] 2458249 2467795 2468043 2469183 2474655 2475752 2476278 2476813 2487
086
[397] 2488793 2489343 2492390

```

## # Standard Deviation

```
print("Standard Deviation of YearsExperience")
```

```
sd(YearsExperience, na.rm = TRUE)
```

```
print("Standard Deviation of Salary")
```

```
sd(Salary, na.rm = TRUE)
```

```

> # Standard Deviation
> print("Standard Deviation of YearsExperience")
[1] "Standard Deviation of YearsExperience"
> sd(YearsExperience, na.rm = TRUE)
[1] 6.789474
> print("Standard Deviation of Salary")
[1] "Standard Deviation of Salary"
> sd(Salary, na.rm = TRUE)
[1] 583131.8

```

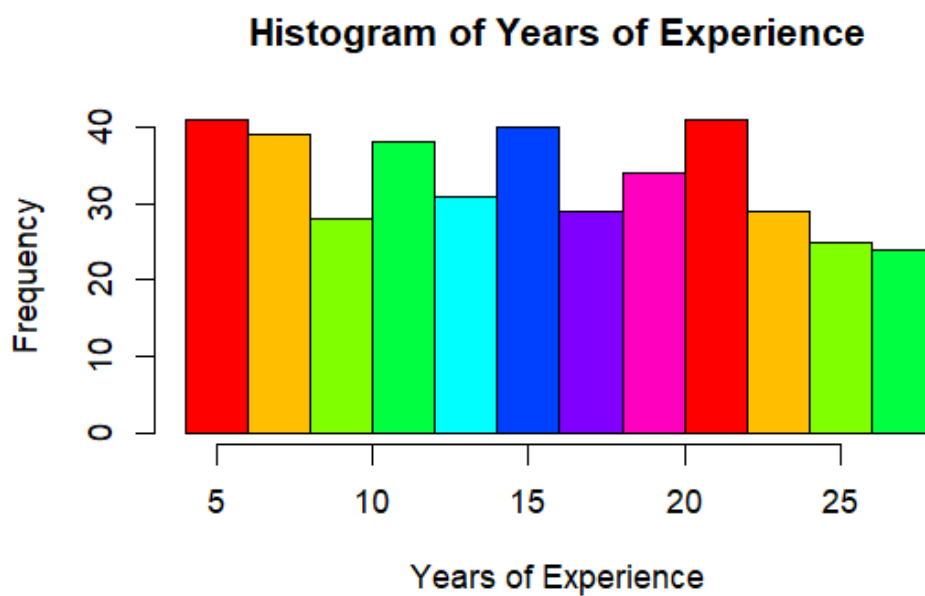
## # Graphical Representation

### # Histograms

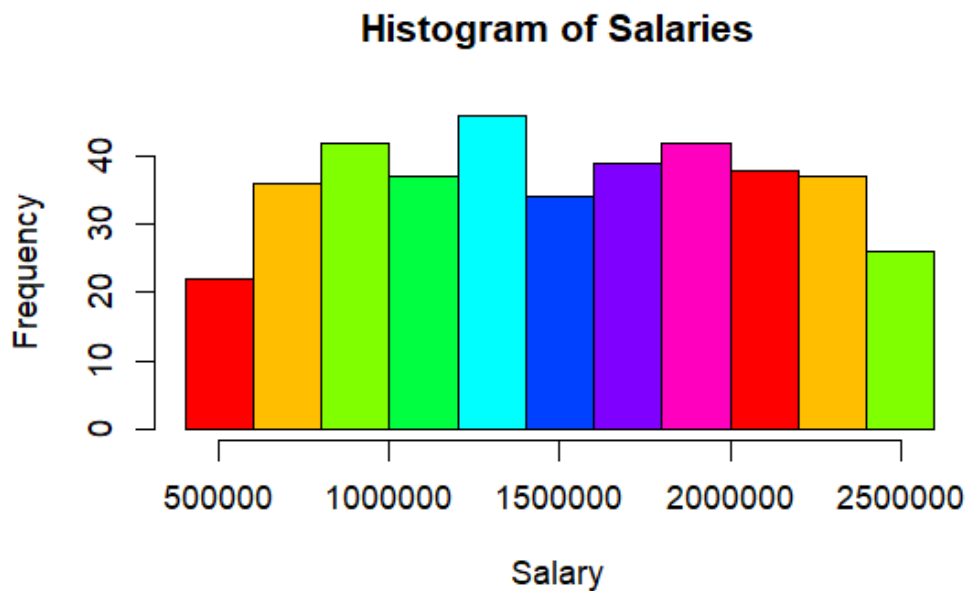
```
hist(YearsExperience,  
     main = "Histogram of Years of Experience",  
     xlab = "Years of Experience",  
     breaks = 10,  
     col = rainbow(8))
```

```
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))
```



### # Categorize Salary into Ranges

```
max(Salary)
```

```
min(Salary)
```

```
> max(Salary)
[1] 2492390
> min(Salary)
[1] 515489
```

### # Frequency of Salary Categories

```
Salary_freq <- table(Salary_class)
```



### # Print Frequency Table

```
print(Salary_freq)
```

### # Pie Chart

```
pie(Salary_freq,
```

```
    labels = paste(names(Salary_freq), "(", Salary_freq, ")", sep = ""),
```

```
    col = rainbow(4),
```

```
    main = "Pie Chart of Salary Ranges")
```

```
> # Frequency of Salary Categories
```

```
> Salary_freq <- table(Salary_class)
```

```
> # Print Frequency Table
```

```
> print(Salary_freq)
```

```
Salary_class
Low [5L-10L)      Medium [10L-15L)      High [15L-20L) Very High [20L
-25L)              100                  98                  100
101
```

```
> # Pie Chart
```

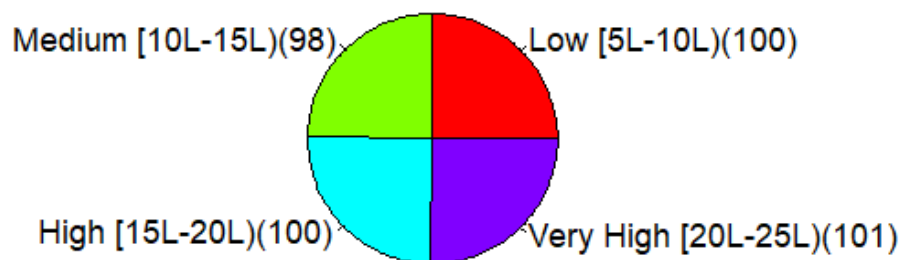
```
> pie(Salary_freq,
```

```
+   labels = paste(names(Salary_freq), "(", Salary_freq, ")", sep = ""),
```

```
+   col = rainbow(4),
```

```
+   main = "Pie Chart of Salary Ranges")
```

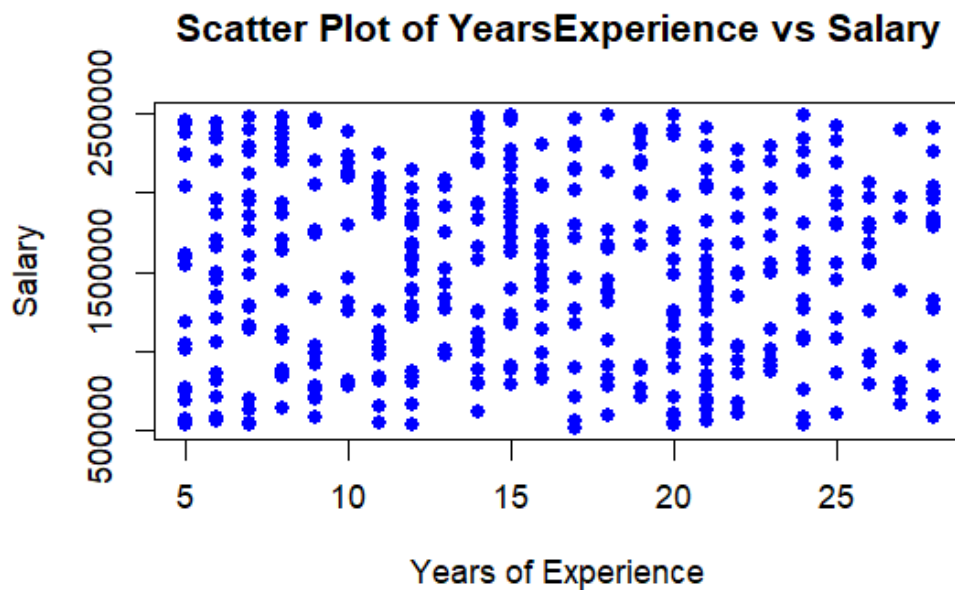
### Pie Chart of Salary Ranges



### # Scatter Plot

```
plot(YearsExperience, Salary,  
     main = "Scatter Plot of YearsExperience vs Salary",  
     xlab = "Years of Experience",  
     ylab = "Salary",  
     col = "blue",  
     pch = 16)
```

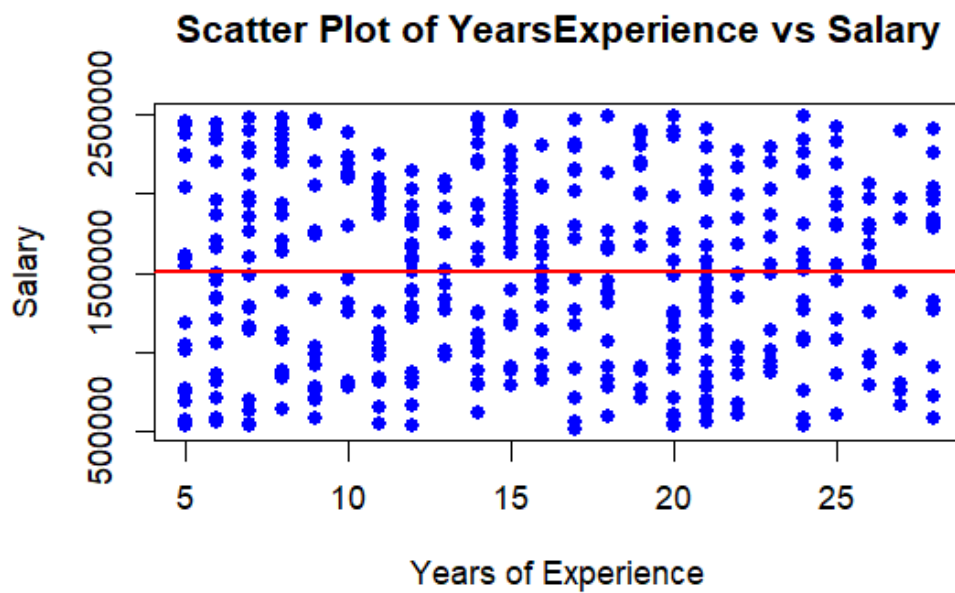
```
> # Scatter Plot  
> plot(YearsExperience, Salary,  
+      main = "Scatter Plot of YearsExperience vs salary",  
+      xlab = "Years of Experience",  
+      ylab = "Salary",  
+      col = "blue",  
+      pch = 16)
```



### # Add Regression Line

```
model <- lm(Salary ~ YearsExperience)  
abline(model, col = "red", lwd = 2)
```

```
> # Add Regression Line  
> model <- lm(Salary ~ YearsExperience)  
> abline(model, col = "red", lwd = 2)
```



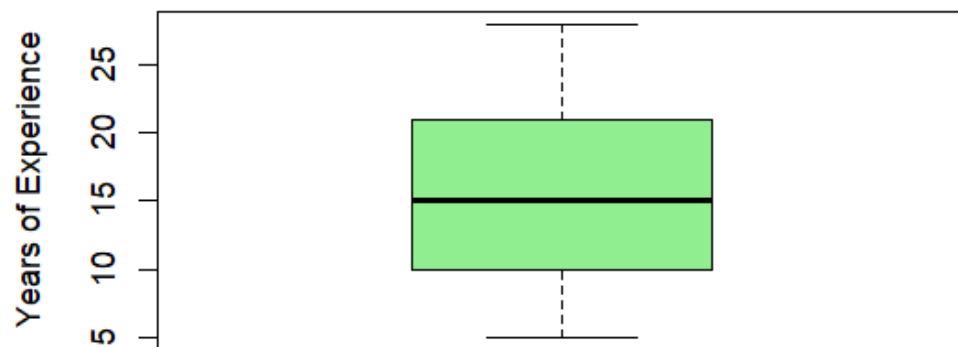
#### # Box Plots

```
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")
```

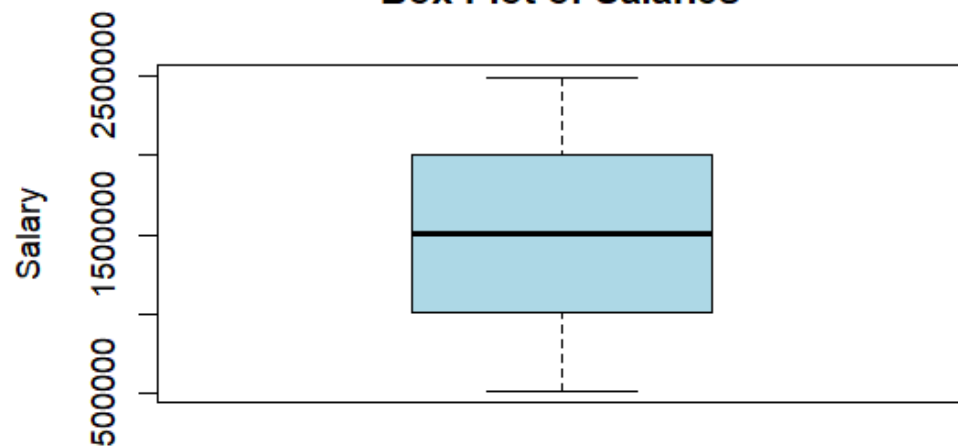
```
> # Box Plots
> boxplot(YearsExperience,
+         main = "Box Plot of Years of Experience",
+         ylab = "Years of Experience",
+         col = "lightgreen")
```

**Box Plot of Years of Experience**



```
> boxplot(Salary,  
+         main = "Box Plot of Salaries",  
+         ylab = "Salary",  
+         col = "lightblue")
```

**Box Plot of Salaries**



### # Line Graph

```
sorted_YearsExperience <- sort(YearsExperience)
```

```
sorted_Salary <- sort(Salary)
```

```

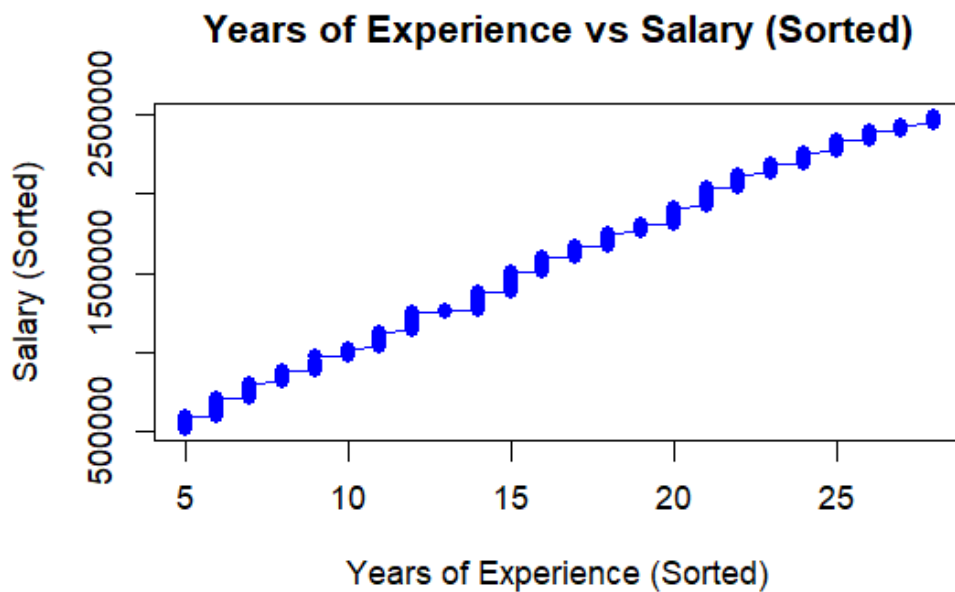
plot(sorted_YearsExperience, sorted_Salary,
     main = "Years of Experience vs Salary (Sorted)",
     type = "o",
     col = "blue",
     pch = 16,
     xlab = "Years of Experience (Sorted)",
     ylab = "Salary (Sorted)")

```

```

> # 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)")

```



**# Add Legend**

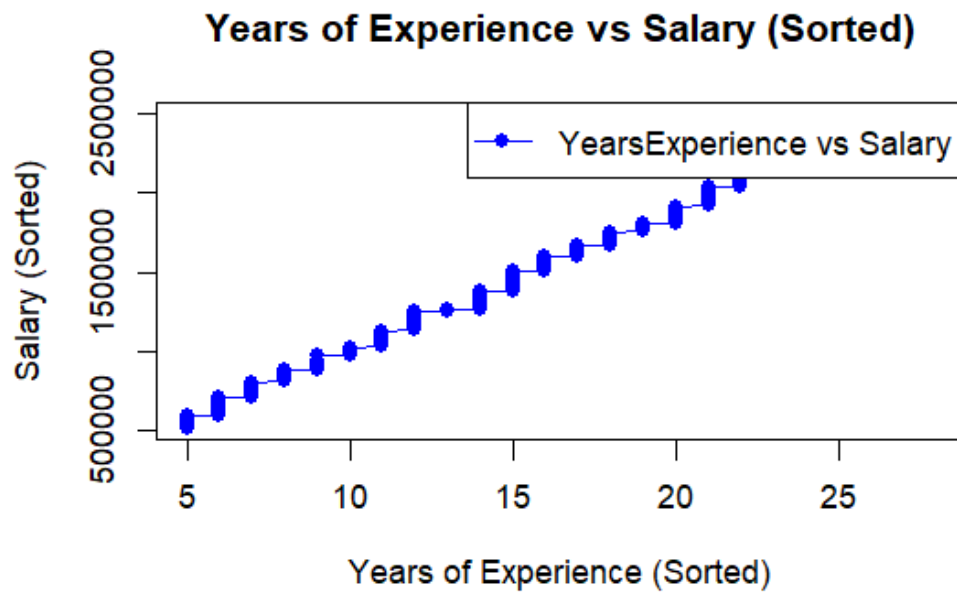
```

legend("topright",
     legend = c("YearsExperience vs Salary"),
     col = c("blue"),

```

```
lty = c(1),  
pch = c(16))
```

```
> # Add Legend  
> legend("topright",  
+       legend = c("YearsExperience vs salary"),  
+       col = c("blue"),  
+       lty = c(1),  
+       pch = c(16))
```



#### # Load Required Library

```
library(BSDA)
```

#### # Group Salaries Based on YearsExperience

```
median_experience <- median(YearsExperience, na.rm = TRUE)
```

```
group1 <- Salary[YearsExperience <= median_experience] # Group 1
```

```
group2 <- Salary[YearsExperience > median_experience] # Group 2
```

#### # Calculate Standard Deviations for Both Groups

```
sigma_x <- sd(group1, na.rm = TRUE)
```

## # Perform Two-Sample Z Test

```
z_test_result <- z.test(x = group1, y = group2,
                        alternative = "two.sided", # Two-tailed test
                        sigma.x = sigma_x,
                        sigma.y = sigma_y)

print("Two-Sample Z Test Results:")

print(z_test_result)
```

```
> # Load Required Library
> library(BSDA)
> # Group Salaries Based on YearsExperience
> median_experience <- median(YearsExperience, na.rm = TRUE)
> group1 <- Salary[YearsExperience <= median_experience] # Group 1
> group2 <- Salary[YearsExperience > median_experience]   # Group 2
> # Calculate Standard Deviations for Both Groups
> sigma_x <- sd(group1, na.rm = TRUE)
> sigma_y <- sd(group2, na.rm = TRUE)
> # Perform Two-Sample Z Test
> z_test_result <- z.test(x = group1, y = group2,
+                           alternative = "two.sided", # Two-tailed test
+                           sigma.x = sigma_x,
+                           sigma.y = sigma_y)
> print("Two-Sample Z Test Results:")
[1] "Two-Sample Z Test Results:"
> print(z_test_result)
```

## Two-sample z-Test

```
data: group1 and group2
z = 0.70202, p-value = 0.4827
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -73484.74 155504.24
sample estimates:
mean of x mean of y
 1526111  1485102
```

## # Categorize YearsExperience

[illegible]

### # 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		

---

## 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.

---

## Dataset Source

The dataset is sourced from [Kaggle](#).