

# MVLU COLLEGE.

PRACTICAL NO :- 07

AIM :- Performing one-way ANOVA using aov() (R).

CODE :-

```
# Load dataset
```

```
salary_data <- read.csv("C:/Users/Arvind/Downloads/Salary_Data.csv")
```

```
str(salary_data)
```

```
# Convert Years.of.Experience to numeric
```

```
salary_data$Years.of.Experience <- as.numeric(salary_data$Years.of.Experience)
```

```
# Create Experience Groups
```

```
salary_data$Exp_Group <- cut(  
  salary_data$Years.of.Experience,  
  breaks = c(0, 3, 6, 10, 50),  
  labels = c("Low", "Medium", "High", "Very High"),  
  include.lowest = TRUE  
)
```

```
# One-Way ANOVA
```

```
anova_one <- aov(Salary ~ Exp_Group, data = salary_data)
```

```
summary(anova_one)
```

The screenshot shows the RStudio interface with the following components:

- Source Editor:** Contains the R code for loading the dataset, converting experience to numeric, creating experience groups, and performing a one-way ANOVA.
- Console:** Displays the output of the code, including the structure of the salary\_data dataset and the results of the ANOVA test.
- Environment:** Shows the objects created in the R session, including anova\_one, data, gender\_marks, jobs, lang, lang\_clean, phones, salary, salary\_data, and t\_test\_result.

**Console Output:**

```
> colnames(salary_data)
[1] "Age" "Gender" "Education.Level" "Job.Title" "Years.of.Experience"
[6] "Salary"
> # Check structure (optional but good practice)
> str(salary_data)
'data.frame': 6704 obs. of 6 variables:
 $ Age      : int  32 28 45 36 52 29 42 31 26 38 ...
 $ Gender   : chr  "Male" "Female" "Male" "Female" ...
 $ Education.Level : chr  "Bachelor's" "Master's" "PhD" "Bachelor's" ...
 $ Job.Title : chr  "Software Engineer" "Data Analyst" "Senior Manager" "Sales Associate" ...
 $ Years.of.Experience : num  5 3 15 7 20 2 12 4 1 10 ...
 $ Salary   : int  90000 65000 150000 60000 200000 55000 120000 80000 45000 110000 ...
> # Convert Years.of.Experience to numeric
> salary_data$Years.of.Experience <- as.numeric(salary_data$Years.of.Experience)
> # Create Experience Groups
> salary_data$Exp_Group <- cut(
+   salary_data$Years.of.Experience,
+   breaks = c(0, 3, 6, 10, 50),
+   labels = c("Low", "Medium", "High", "Very High"),
+   include.lowest = TRUE
+ )
> # One-Way ANOVA
> anova_one <- aov(Salary ~ Exp_Group, data = salary_data)
> summary(anova_one)
              Df Sum Sq Mean Sq F value Pr(>F)
Exp_Group      3 1.288e+13 4.294e+12  4974 <2e-16 ***
Residuals    6695 5.781e+12 8.634e+08
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
5 observations deleted due to missingness
>
```

**Environment:**

Object	Class	Size
anova_one	Large aov	(14 elements, 610.4 kb)
data	data.frame	9994 obs. of 21 variables
gender_marks	data.frame	6 obs. of 2 variables
jobs	data.frame	1602 obs. of 8 variables
lang	data.frame	262 obs. of 4 variables
lang_clean	data.frame	262 obs. of 2 variables
phones	data.frame	980 obs. of 26 variables
salary	data.frame	6704 obs. of 6 variables
salary_data	data.frame	6704 obs. of 7 variables
t_test_result	List	of 10

NANDINI PANDIT S100

DATA ANALYSIS WITH SAS/SPSS/R.

# MVLU COLLEGE.

PRACTICAL NO :- 08

AIM :- Performing two-way ANOVA using aov() (R).

CODE :-

# Load dataset

```
phone <- read.csv("C:/Users/Arvind/Downloads/Smartphones_cleaned_dataset.csv")
colnames(phone)
```

# Convert factors

```
phone$brand_name <- as.factor(phone$brand_name)
phone$os <- as.factor(phone$os)
```

# Two-Way ANOVA

```
anova_two <- aov(price ~ brand_name * os, data = phone)
```

# Display result

```
summary(anova_two)
```

The screenshot displays the RStudio interface with the following components:

- Source Editor:** Contains the R script for loading the dataset, converting factors, and performing a two-way ANOVA.
- Console:** Shows the execution of the script, including an error message: "Error in '\$<-data.frame'(\*tmp\*, brand, value = integer(0)) : replacement has 0 rows, data has 980".
- Environment:** Lists the objects in the global environment, including 'anova\_one', 'anova\_two', 'Data', 'gender\_marks', 'jobs', 'jobs\_clean', 'lang', 'lang\_clean', 'lang\_data', 'lang\_table', 'phone', 'phones', 'salary', 'salary\_data', 't\_test\_result', and 'values'.
- Terminal:** Displays the output of the ANOVA test, showing the summary of the model.

```
> # Load dataset
> phone <- read.csv("C:/Users/Arvind/Downloads/Smartphones_cleaned_dataset.csv")
> # Convert variables to factors
> phone$brand_name <- as.factor(phone$brand_name)

Error in '$<-data.frame'(*tmp*, brand, value = integer(0)) :
  replacement has 0 rows, data has 980

> colnames(phone)
[1] "brand_name" "rating" "has_ir_blaster" "processor_speed" "fast_charging" "screen_size" "num_front_cameras" "primary_camera_front" "resolution_width"
[10] "model" "has_5g" "processor_brand" "battery_capacity" "ram_capacity" "refresh_rate" "os" "extended_memory_available" "resolution_height"
[19] "price" "has_nfc" "num_cores" "fast_charging_available" "internal_memory" "num_rear_cameras" "primary_camera_rear" "extended_upto"

> # Convert factors
> phone$brand_name <- as.factor(phone$brand_name)
> phone$os <- as.factor(phone$os)
> # Two-Way ANOVA
> anova_two <- aov(price ~ brand_name * os, data = phone)
> # Display result
> summary(anova_two)

brand_name Df Sum Sq Mean Sq F value Pr(>F)
os 2 2.495e+11 1.248e+11 20.714 < 2e-16 ***
brand_name:os 8 1.230e+10 1.537e+09 1.912 0.055 .
Residuals 924 7.430e+11 8.041e+08
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

NANDINI PANDIT S100

DATA ANALYSIS WITH SAS/SPSS/R.

# MVLU COLLEGE.

PRACTICAL NO :- 09

AIM :- Conducting Chi-square tests using chisq.test() (R)

CODE :-

# Load dataset

```
lang_data <- read.csv("C:/Users/Arvind/Downloads/programming language trend over time.csv")
```

# Check column names

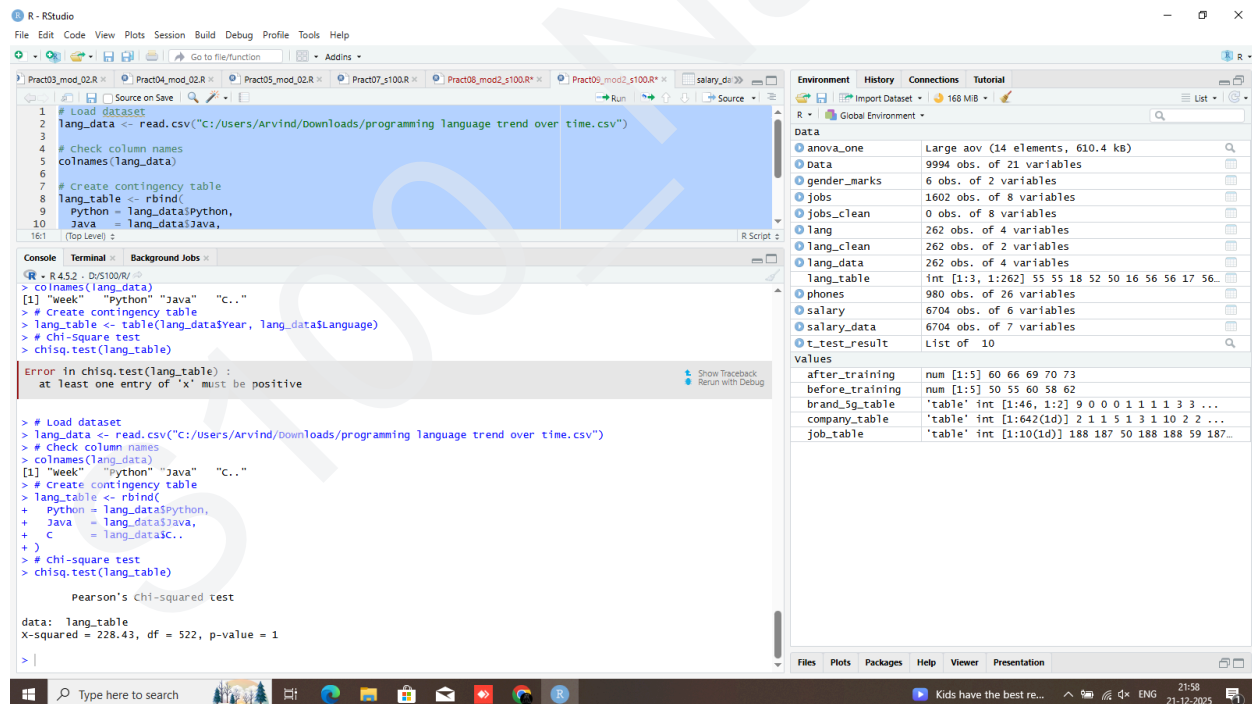
```
colnames(lang_data)
```

# Create contingency table

```
lang_table <- rbind(  
  Python = lang_data$Python,  
  Java   = lang_data$Java,  
  C      = lang_data$C..  
)
```

# Chi-square test

```
chisq.test(lang_table)
```



NANDINI PANDIT S100

DATA ANALYSIS WITH SAS/SPSS/R.