

## EXPERIMENT-17

### CORRELATION ANALYSIS:

#### AIM:

To write the program for correlation analysis using R-tool.

#### PROGRAM:

```
library(readxl) # Load readxl package

# Read the Excel file
file_path <- "C:/Users/harik/Downloads/Data Excel sheet for study on DM.xlsx"
diabetest1 <- read_excel(file_path)

# Clean column names (remove extra spaces)
colnames(diabetest1) <- trimws(colnames(diabetest1))

# Function to compute frequency distribution and Chi-Square Goodness-of-Fit Test
analyze_age_distribution <- function(data, column) {
  if (column %in% colnames(data)) {
    # Convert column to factor (categorical)
    data[[column]] <- as.factor(data[[column]])

    # Frequency distribution of AGE
    age_table <- table(data[[column]])
    print("Frequency Distribution of AGE:")
    print(age_table)

    # Expected distribution (assuming uniform distribution)
    expected_counts <- rep(sum(age_table) / length(age_table), length(age_table))

    # Perform Chi-Square Goodness-of-Fit Test
    chi_test <- chisq.test(age_table, p = expected_counts / sum(expected_counts))
```

```

# Print Chi-Square test results

print("Chi-Square Goodness-of-Fit Test for AGE Distribution:")

print(chi_test)

} else {

    print(paste("Column", column, "not found in the dataset."))

}

}

```

# Example: Analyze AGE column

```
analyze_age_distribution(diabetest1, "AGE")
```

OUTPUT:

```

> # Example: Analyze AGE column
> analyze_age_distribution(diabetest1, "AGE")
[1] "Frequency Distribution of AGE:"

35 39 40 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62
 1  2  3  2  3  4  3  3  2  3  4  4  2  2  1  1  1  1  5  1  2  2  3
63 64 65 66 69
 4  2  2  6  1
[1] "Chi-Square Goodness-of-Fit Test for AGE Distribution:"

      Chi-squared test for given probabilities

data:  age_table
X-squared = 18.8, df = 27, p-value = 0.8772

```

RESULT:

Thus the correlation analysis was executed successfully.