EXPT NO : 11	STATISTICAL ANALYSIS WITH R
DATE : 09.04.2025	

AIM:

To Implement statistical analysis in R programming.

1.Create a dataset containing loan application details such as Application_ID, Applicant_Age, Income,

Loan_Amount, Loan_Term, Credit_Score

Tasks:

1. Use summary() to analyze the dataset.

2. Compute the mean and median loan amount.

3. Find the variance and standard deviation of Credit_Score.

4. Identify the age group that takes the highest loan amounts.

5. Determine the percentage of applicants with credit scores above 750.

CODE:

```
loan_data <- data.frame(
  Application_ID = 1:15, Applicant_Age = c(25, 34, 45, 29, 52, 40, 31, 47, 38, 26, 55, 43, 36, 30, 48),
  Income = c(35000, 48000, 55000, 40000, 62000, 50000, 45000, 58000, 47000, 39000, 63000, 52000, 49000, 41000, 60000),
  Loan_Amount = c(150000, 180000, 200000, 160000, 220000, 190000, 170000, 210000, 175000, 155000, 225000, 195000, 185000, 165000, 205000),
  Loan_Term = c(10, 15, 20, 10, 25, 15, 10, 20, 15, 10, 25, 20, 15, 10, 20),
  Credit_Score = c(700, 720, 680, 750, 790, 710, 730, 770, 690, 740, 800, 760, 720, 710, 780))
write.csv(loan_data,"loan.csv",row.names=FALSE)
summary(loan_data)
mean_loan <- mean(loan_data$Loan_Amount)
median_loan <- median(loan_data$Loan_Amount)
mean_loan
median_loan
variance_score <- var(loan_data$Credit_Score)
stddev_score <- sd(loan_data$Credit_Score)
variance_score
stddev_score
```

```
loan_data$Age_Group <- cut(
  loan_data$Applicant_Age,
  breaks = c(20, 30, 40, 50, 60),
  labels = c("20-30", "31-40", "41-50", "51-60")
)
aggregate(Loan_Amount ~ Age_Group, data = loan_data, mean)
high_score_count <- sum(loan_data$Credit_Score > 750)
total_applicants <- nrow(loan_data)
percentage_above_750 <- (high_score_count / total_applicants) * 100
percentage_above_750
```

OUTPUT:

```
> summary(loan_data)
 Application_ID Applicant_Age      Income        Loan_Amount       Loan_Term    Credit_Score
 Min.   : 1.0   Min.   :25.0   Min.   :35000   Min.   :150000   Min.   :10   Min.   :680.0
 1st Qu.: 4.5   1st Qu.:30.5   1st Qu.:43000   1st Qu.:167500   1st Qu.:10   1st Qu.:710.0
 Median : 8.0   Median :38.0   Median :49000   Median :185000   Median :15   Median :730.0
 Mean   : 8.0   Mean   :38.6   Mean   :49600   Mean   :185667   Mean   :16   Mean   :736.7
 3rd Qu.:11.5   3rd Qu.:46.0   3rd Qu.:56500   3rd Qu.:202500   3rd Qu.:20   3rd Qu.:765.0
 Max.   :15.0   Max.   :55.0   Max.   :63000   Max.   :225000   Max.   :25   Max.   :800.0
> mean_loan <- mean(loan_data$Loan_Amount)
> median_loan <- median(loan_data$Loan_Amount)
>
> mean_loan
[1] 185666.7
> median_loan
[1] 185000
> variance_score <- var(loan_data$Credit_Score)
> stddev_score <- sd(loan_data$Credit_Score)
>
> variance_score
[1] 1380.952
> stddev_score
[1] 37.16117
```

```
> aggregate(Loan_Amount ~ Age_Group, data = loan_data, mean)
  Age_Group Loan_Amount
1     20-30      157500
2     31-40      180000
3     41-50      202500
4     51-60      222500
> high_score_count <- sum(loan_data$Credit_Score > 750)
> total_applicants <- nrow(loan_data)
> percentage_above_750 <- (high_score_count / total_applicants) * 100
>
> percentage_above_750
[1] 33.33333
```

RESULT:

Thus , the R program is implemented and output is verified successfully.

2303717624322023