# Lab-1

The information of 20 persons is given in the following table:

| SN | Sex | Religion | Level of Education | SN | Sex | Religion | Level of Education |
|---|---|---|---|---|---|---|---|
| 1 | Male | Muslim | Primary | 11 | Female | Hindu | Primary |
| 2 | Female | Hindu | Graduate | 12 | Male | Christian | Graduate |
| 3 | Male | Muslim | Illiterate | 13 | Male | Others | Secondary |
| 4 | Male | Hindu | Graduate | 14 | Female | Muslim | Secondary |
| 5 | Female | Muslim | Primary | 15 | Male | Hindu | Higher Secondary |
| 6 | Female | Christian | Graduate | 16 | Male | Christian | Others |
| 7 | Male | Muslim | Illiterate | 17 | Female | Muslim | Primary |
| 8 | Male | Hindu | Primary | 18 | Male | Others | Illiterate |
| 9 | Female | Muslim | Others | 19 | Female | Muslim | Higher Secondary |
| 10 | Male | Others | Higher Secondary | 20 | Male | Others | Secondary |

I. Construct the frequency distribution for variables “Religion” and “Level of education”.

II. Draw pie diagram for the variable “Religion ” and comment.

III. Draw bar diagram for the variable “Level of Education ” and comment.

# Solve:

```
sex<c("Male","Female","Male","Male","Female","Female","Male","Male","Female","Male","Fe
male","Male","Male","Female","Male","Male","Female","Male","Female","Male")

religion<c("Muslim","Hindu","Muslim","Hindu","Muslim","Christian","Muslim","Hindu","Musli
m","Others","Hindu","Christian","Others","Muslim","Hindu","Christian","Muslim","Others","Mu
slim","Others")

edu<c("BTech","BEng","MSc","BSc","Hons","Primary","Primary","Primary","HSC","HSC","HSC","
HSC","SSC","SSC","SSC","SSC","SSC","SSC","SSC","SSC")

blood<c("A+","AB+","A+","A+","AB-","AB-","B+","B+","B+","AB+","AB-","AB-
","B+","AB+","O+","O+","O+","AB-","AB+","AB-")


sexF<-table(sex)

religionF<-table(religion)

eduF<-table(edu)

bloodF<-table(blood)
```

```
dataTable<- data.frame(sex,religion,edu,blood)

dataTable


sexF

religionF

eduF

bloodF


pie(sexF,main = "Pie chart for Sex",col = c("red","green"))

pie(religionF, main = "Pie chart for religions",col = c("red","green","yellow"))


barplot(eduF,main = "Father's level of Education", xlab = "Degree",ylab = "Number of Fathers",col = c("red","green","yellow","black","gray","green","red","yellow"))

barplot(bloodF,main = "Blood Distribution of Students",xlab = "Blood Groups",ylab = "Number of Students", col = c("red","green","yellow","black"))
```

# LAB-2

Suppose that 80 students are enrolled in a statistics class and the following are the test scores received by them:

77 44 49 33 38 33 76 55 68 39 44 59 36 55 47 61 53 32 65 51 29 41 32 45 83 58 73 47 40 26 59 43 66 44 41 25 39 72

37 55 34 47 66 53 55 58 49 45 61 41 55 92 83 77 45 62 45 36 78 48 54 50 51 66 80 73 57 61 56 50 45 82 71 48 46 69

38 72 56 64

I. Compute Mean, Median, Mode, Variance and Standard deviation of the above raw data and comment on

your results.

II. Find the five number summaries

III. Select an appropriate class interval and organize data set into a frequency distribution.

IV. Using frequency distribution obtained in question (iii) construct a histogram and an ogive. Also approximate

the median and mode with the help of ogive and histogram respectively.

V. Find the mean, median and mode using the frequency distribution obtained in question (iii).

I. Show the data stem and leaf display. Find the mode and median from stem leaf plot.

## Solve:

```
# Raw data

scores <- c(77, 44, 49, 33, 38, 33, 76, 55, 68, 39, 44, 59, 36, 55, 47, 61, 53, 32, 65, 51, 29, 41, 32, 45, 83, 58, 73, 47, 40, 26,
59, 43, 66, 44, 41, 25, 39, 72, 37, 55, 34, 47, 66, 53, 55, 58, 49, 45, 61, 41, 55, 92, 83, 77, 45, 62, 45, 36, 78, 48, 54, 50, 51,
66, 80, 73, 57, 61, 56, 50, 45, 82, 71, 48, 46, 69, 38, 72, 56, 64)


# I. Mean, Median, Mode, Variance, and Standard Deviation

mean_scores <- mean(scores)

median_scores <- median(scores)

mode_scores <- as.numeric(names(table(scores))[table(scores) == max(table(scores))])

var_scores <- var(scores)

sd_scores <- sd(scores)


# II. Five-number summary

five_number_summary <- summary(scores)


# III. Frequency Distribution

class_interval <- 10

# Choose an appropriate class interval, you can modify it if needed

freq_table <- table(cut(scores, breaks = seq(min(scores), max(scores) + class_interval, by = class_interval)))

# Calculate cumulative frequencies

cumulative_freq <- cumsum(freq_table)

# IV. Histogram and Ogive

hist(scores, breaks = seq(min(scores), max(scores) + class_interval, by = class_interval), main = "Histogram", xlab =
"Scores",

   ylab = "Frequency")


plot(cumulative_freq, type = "o", main = "Ogive", xlab = "Scores", ylab = "Cumulative Frequency")

lines(cumulative_freq, type = "o", col = "blue")


#VI. Stem-and-leaf plot

stem<-stem(scores)
```

# LAB-3

The following data represents the ages of the 50 richest people in the world in 2009.

89,89,87,86,85,83,83,82,81,80,78,78,77,76,73,73,73,72,69,69,68,67,66,66,65,65,64,63,61,61,60,59,58,57,56,54,54,53,53,51,51,49,47,46,44,43,42,36,2000.

II. Find the mean, median and mode of the ages of the 50 richest people. Which measures of central tendency best describes a typical entry of this data set.

III. Replace 35 instead of 2000 from the data set then rework (i). Compare these measures of central tendency with those found in (i).

IV. Compute trimmed mean extracting 2% observation.

V. Construct a frequency distribution using the above data after replacing 35 instead of 2000

VI. Construct a relative frequency histogram.

VII. Compute skewness and kurtosis. Comment on the shape of the distribution.

VIII. Find standard deviation, mean deviation about median, coefficient of variation.

**#Solve:ii**

ages <- c(89, 89, 87, 86, 85, 83, 83, 82, 81, 80, 78, 78, 77, 76, 73, 73, 73, 72, 69, 69, 68, 67, 66, 66, 65, 65, 64, 63, 61, 61, 60, 59, 58, 57, 56, 54, 54, 53, 53, 51, 51, 49, 47, 46, 44, 43, 42, 51, 36, 2000)

# Calculating mean

mean_age <- mean(ages)

# Calculating median

median_age <- median(ages)

# Calculating mode

mode_age <- as.numeric(names(table(ages))[table(ages) == max(table(ages))])

#The measure of central tendency that best describes a typical entry in this dataset would be the median (66), as it is less affected by the extreme value (2000) compared to the mean (104.9388).

## #Solve:iii

ages[ages == 2000] <- 35

# Calculating mean

mean_age_2 <- mean(ages)

# Calculating median

median_age_2 <- median(ages)

# Calculating mode

mode_age_2 <- as.numeric(names(table(ages))[table(ages) == max(table(ages))])

**#Solve:iv**

# Sorted ages

sorted_ages <- sort(ages)

# Trimming 2% from both ends

trim_pct <- 0.02

trim_length <- round(length(sorted_ages) * trim_pct)

trimmed_mean_age <- mean(sorted_ages[(trim_length + 1):(length(sorted_ages) - trim_length)])

# #Solve:v

# Frequency table

freq_table_age <- table(ages)

# Creating relative frequency table

rel_freq_table_age <- prop.table(freq_table_age)

## #Solve:vi

# Creating histogram

hist(ages, breaks = 10, freq = FALSE, main = "Relative Frequency Histogram of Ages",

ylim = c(0, 0.25), xlab = "Age", ylab = "Relative Frequency")

# Adding relative frequencies as bars

rel_freq_table_age <- prop.table(freq_table_age)

barplot(rel_freq_table_age, horiz = TRUE, col = "lightblue", border = "white", las = 1, xlim = c(0, 0.25))

# **# Solve: Vii**

# Calculating skewness

skewness_age <- skewness(ages)

# Calculating kurtosis

kurtosis_age <- kurtosis(ages)

## **# Solve: Viii**

# Calculating standard deviation

sd_age <- sd(ages)

# Calculating mean deviation about median

mean_dev_median <- median(abs(ages - median(ages)))

# Calculating coefficient of variation

cv_age <- sd_age / mean_age

## LAB 4

The grade point average (GPA) in different semesters of two students are shown below:

| Student | GPA in Semister | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| A | 2.5 | 2.5 | 3.0 | 3.5 | 3.5 | 4.0 | 3.5 | 3.5 |
| B | 2.5 | 3.0 | 4.0 | 4.0 | 4.0 | 2.0 | 2.5 | 4.0 |

Which students would you consider better throughout the courses of studies?

**#Solve:**

# Define the GPA data for each student

student_A <- c(2.5, 2.5, 3.0, 3.5, 3.5, 4.0, 3.5, 3.5)

student_B <- c(2.5, 3.0, 4.0, 4.0, 4.0, 2.0, 2.5, 4.0)

# Calculate the average GPAs for each student

avg_GPA_A <- mean(student_A)

avg_GPA_B <- mean(student_B)

## Lab-5

Generate a random sample of size 1000 from binomial (5, 0.3), i.e. n=5, p=0.3. Compare observed distribution with true binomial distribution.

# Solve:

# Set the seed for reproducibility

set.seed(1234)

# Generate a random sample from the binomial distribution

sample_size <- 1000

n <- 5

p <- 0.3

random_sample <- rbinom(sample_size, n, p)

```
# Calculate the observed frequencies in the sample
observed_freq <- table(random_sample)

# Calculate the probabilities in the true binomial distribution
binom_probs <- dbinom(0:n, n, p)

# Create a data frame to store the observed and expected frequencies
comparison <- data.frame(Observed = observed_freq, Expected = binom_probs)

# Print the comparison
print(comparison)
```

# LAB-06

If Z~N(0,1). For the following values of Z

-4.0, -3.9, -3.8, -3.7,…………………………..,3.8, 3.9, 4.0.

I. Create pdf of Z . Draw standard normal curve and comment the sharp characteristics of the distribution.

II. Create pdf and cdf of X~N(1000,250000).

III. Find

a)P(X=850)

b)P(X>1200)

a)P(1000<X<2000)

Construct normal density curve and normal cumulative distribution curve. Comment on your results.

## Solve:

```r
#Number I
# Generate Z values from -4.0 to 4.0
z_values <- seq(-4.0, 4.0, by = 0.1)

# Calculate PDF of Z
pdf_z <- dnorm(z_values, mean = 0, sd = 1)

# Plot standard normal curve
plot(z_values, pdf_z, type = "l", xlab = "Z", ylab = "PDF", main = "Standard Normal Distribution")

#Number II
# Define the parameters
mean_X <- 1000
variance_X <- 250000
sd_X <- sqrt(variance_X)

# Generate X values
x_values <- seq(-5000, 3000, by = 100)

# Calculate PDF of X
pdf_X <- dnorm(x_values, mean = mean_X, sd = sd_X)

# Calculate CDF of X
cdf_X <- pnorm(x_values, mean = mean_X, sd = sd_X)

# Plot normal density curve
plot(x_values, pdf_X, type = "l", xlab = "X", ylab = "PDF", main = "Normal Distribution: X ~ N(1000,
250000)")
```

```
# Plot normal cumulative distribution curve

plot(x_values, cdf_X, type = "l", xlab = "X", ylab = "CDF", main = "Normal Distribution: X ~ N(1000, 250000)")


#Number III

# Find P(X = 850)

prob_X_850 <- pnorm(850, mean = mean_X, sd = sd_X, lower.tail = TRUE)

print(prob_X_850)


# Find P(X > 1200)

prob_X_gt_1200 <- 1 - pnorm(1200, mean = mean_X, sd = sd_X, lower.tail = TRUE)

print(prob_X_gt_1200)


# Find P(1000 < X < 2000)

prob_X_between <- pnorm(2000, mean = mean_X, sd = sd_X, lower.tail = TRUE) - pnorm(1000, mean = mean_X, sd = sd_X, lower.tail = TRUE)

print(prob_X_between)
```