COMMITTED TO EXCELLENCE IN EDUCATION

निर्मलस्नेह उत्तम सेवाधर्म

# SHRI G.P.M. DEGREE COLLEGE OF SCIENCE & COMMERCE

SHRI G.P.M. DEGREE COLLEGE OF SCIENCE & COMMERCE.

(COMMITED TO EXCELLENCE IN EDUCATION)

# CERTIFICATE

This is to certify that Mr/Ms ________________________ a student of **FY.BSC-CS** Roll no: ____ has completed the required number of practicals in the subject of _________________________________ as prescribed by the UNIVERSITY OF MUMBAI under my supervision during the academic year 2023-2024.

| | |
|---|---|
| **Course Co-ordinator** | **Prof. Incharge** |
| **External Examiner** | **Principal** |
| **Date:** ______________ | **College Seal** |

| | |
|---|---|
| **Prof. Name : Ashwini Kudtarkar** | **Class / Sem : F.Y. B.Sc. - CS / Sem – II (2023-2024)** |
| **Course Code : USCSP207** | **Subject Name : Statistical Methods** |

| | | | | |
|---|---|---|---|---|
| 9 | / / | **Practical-9 : Analysis of variance -**<br>a. Perform One-way ANOVA<br>b. Perform Two-way ANOVA | …… | …… |
| 10 | / / | **Practical-10 : Non-parametric tests-** a.<br>Sign test and Wilcoxon Sign rank test<br>b. Run test<br>c. Kruskal-Wallis (H) test<br>d. Chi-square test | …… | …… |
| 8 | / / | **Practical-8 : Small sample tests based on t and F-**<br>a. t-test for significance of single mean, population variance being unknown(Single mean Ho : $\mu = \mu 0$ )<br>b. t-test for significance of the difference between two sample means (Independent samples)<br>c. t-test for significance of the difference between two sample means (Relatedsamples) d. F-Test to Compare Two Variances | …… | …… |

*******

# Theory- 1

## Probability

**a. Probability definition: classical, axiomatic**

Classical probability is the traditional approach to probability, which assigns probabilities to events based on how likely those events are to occur. Classical probability is based on the notion that an event with a higher probability of occurring is more likely to occur than an event with a lower probability of occurring.

Axiomatic probability is a mathematical approach to probability which uses axioms, or assumptions, to define probability. Axiomatic probability assigns probabilities to events based on the axioms given and does not rely on prior knowledge or experience. Axiomatic probability is useful for calculating complex probabilities, such as those involving multiple events or variables.

**b. Elementary Theorems of probability**

1. The Law of Total Probability: This theorem states that the probability of an event occurring is equal to the sum of the probabilities of the individual outcomes that make up the event.

2. Bayes' Theorem: This theorem states that the probability of an event occurring given some prior information is equal to the product of the probability of the prior information and the probability of the event occurring given the prior information.

3. The Law of Large Numbers: This theorem states that the average of a large number of independent trials will approach the expected value of the probability distribution as the number of trials increases.

4. The Central Limit Theorem: This theorem states that the sum of a large number of random variables will be approximately normally distributed.

5. Union Probability: This theorem states that the probability of an event occurring when two events are combined is the sum of the probability of each event occurring separately. This is expressed mathematically as $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

6. Intersection Probability: This theorem states that the probability of two events occurring simultaneously is the product of the probability of each event occurring separately. This is expressed mathematically as P(A∩B) = P(A)*P(B).

# Practical 1

**a. Examples based on Probability definition: classical, axiomatic**

**Code:**

```
# Define the sample space
sample_space <- c("heads", "tails")

# Define the event of interest
event <- "heads"

# Calculate classical probability
classical_prob <- sum(sample_space ==
event) / length(sample_space)
classical_prob


# Define a measure function
measure <- function(event) {
return(1/length(sample_space))
}

# Calculate axiomatic probability

axiomatic_prob <- measure(event)
axiomatic_prob
```

**Output:**

```
[1] 0.5
[1] 0.5
```

**b. Examples based on elementary Theorems of probability Code:**

```
# Define the sample space
sample_space <- c("heads", "tails")

# Define the first event
event_A <- c("heads")

# Define the second event
event_B <- c("tails")

# Define a measure function
measure <- function(event) {
  return(1/length(sample_space))
}

# Calculate union probability using
elementary theorem of probability

union_prob <- measure(union(event_A,
event_B))
union_prob

# Calculate intersection probability using
elementary theorem of probability

intersection_prob <-
measure(intersect(event_A, event_B))
intersection_prob
```

**Output:**

```
[1] 0.5
[1] 0.5
```

**Result and Discussion :**

**Conclusion**

**Learning outcome :**

**Course outcome :**



**Viva Questions :**

1. What is probability?
2. What is sample space?
3. What is random variable?
4. What is Bayes Theorem?

:

**For Faculty Use**

| | | | |
|---|---|---|---|
| Correction parameters | Formative Assessment [ ] | Timely completion of practical [ ] | Attendance Learning Attitude [ ] |

| | | | |
|---|---|---|---|
| | | | |

# Theory- 2

## Conditional probability and independence

**a. Conditional probability**

Conditional probability is the probability of an event occurring given that another event has already occurred. This is calculated by dividing the probability of both events occurring together by the probability of the first event occurring. For example, the probability of drawing a red card from a deck of cards given that the first card drawn was a heart is 26/51 (half of the cards are red and half are hearts).

**b. Bayes theorem**

Bayes theorem is a theorem used to calculate conditional probability. It is used when the probability of an event is not known but the probability of the reverse event is known. The formula is: $P(A|B) = (P(B|A) * P(A)) / P(B)$.

**c. Independence**

Independence is when two events are not related. The probability of them both occurring is the product of the probabilities of each event occurring on its own.

# Practical 2

**a. Examples based on Conditional probability**
**code for finding the conditional probability of an event in R:**

**Code:**

```
# Define the sample space
sample_space <- c("heads", "tails")

# Define the event of interest
event <- c("heads")

# Define the conditioning event
condition <- c("heads")

# Define a measure function
measure <- function(event){
return(1/length(sample_space))
}

# Calculate conditional probability
conditional_prob <-
measure(event[sample_space %in%
condition]) /
measure(condition)
conditional_prob
```

**Output:**

```
[1] 1
```

**b. Examples based on "Bayes" theorem**
**Code:**

```
# Define the sample space
sample_space <- c("A", "B", "C")

# Define the event of interest
event <- "A"

# Define the prior probabilities
prior_A <- 1/3
prior_B <- 1/3
prior_C <- 1/3

# Define the likelihoods
likelihood_A <- 0.7
likelihood_B <- 0.6
likelihood_C <- 0.5

# Calculate the posterior probabilities
using Bayes' Theorem
posterior_A <- likelihood_A * prior_A /
(likelihood_A * prior_A +

likelihood_B * prior_B +

likelihood_C * prior_C)
posterior_B <- likelihood_B * prior_B /
(likelihood_A * prior_A +

likelihood_B * prior_B +

likelihood_C * prior_C)
posterior_C <- likelihood_C * prior_C /
(likelihood_A * prior_A +

likelihood_B * prior_B +

likelihood_C * prior_C)

# Return the desired posterior probability
posterior_prob <- ifelse(event == "A",
posterior_A,
                         ifelse(event ==
"B", posterior_B, posterior_C))
posterior_prob
```

**Output:**

```
[1] 0.3888889
```

**c. Examples based on independence**

**Code:**

```
# Define the sample space
sample_space_A <- c("heads", "tails")
sample_space_B <- c("red", "green", "blue")

# Define the first event
event_A <- c("heads")

# Define the second event
event_B <- c("red")

# Define a measure function
measure_A <- function(event) {
  return(1/length(sample_space_A))
}
measure_B <- function(event) {
  return(1/length(sample_space_B))
}

# Calculate joint probability of independent events
joint_prob_indep <- measure_A(event_A) * measure_B(event_B)
joint_prob_indep

# Calculate marginal probability of event A
marginal_prob_A <- sum(sapply(sample_space_B, function(x)
  measure_A(event_A) * measure_B(x)))
marginal_prob_A
```

**Output:**

```
[1] 0.1666667
```

**Result and Discussion :**

**Conclusion**

**Learning outcome :**

**Course outcome :**

**Viva Questions :**

1. What is Conditional probability?
2. What is independence?
3. What is Bayes Theorem?
4. What is sample space?

**:**

**For Faculty Use**

| Correction parameters | Formative Assessment [ ] | Timely completion of practical [ ] | Attendance Learning Attitude [ ] |
|---|---|---|---|
| | | | |

# Theory- 3

## Discrete random variable

A discrete random variable is a random variable that can take on only specific values. It is typically used to represent the outcomes of an experiment or survey that has discrete outcomes, such as the number of heads in a coin toss or the number of people who voted for a certain candidate in an election.

**a. Probability distribution of discrete random variable**
The probability distribution of a discrete random variable is a set of probabilities that describe how likely it is for the variable to take on each of its possible values.

**b. Probability mass function**
Probability mass function (PMF) is a mathematical function that gives the probability that a discrete random variable is equal to some value. It is a measure of the probability distribution of the random variable and is used to calculate the probability of a particular outcome. The PMF is typically denoted by the letter 'p'.



# Practical 3

**a. Probability distribution of discrete random variable Code:**

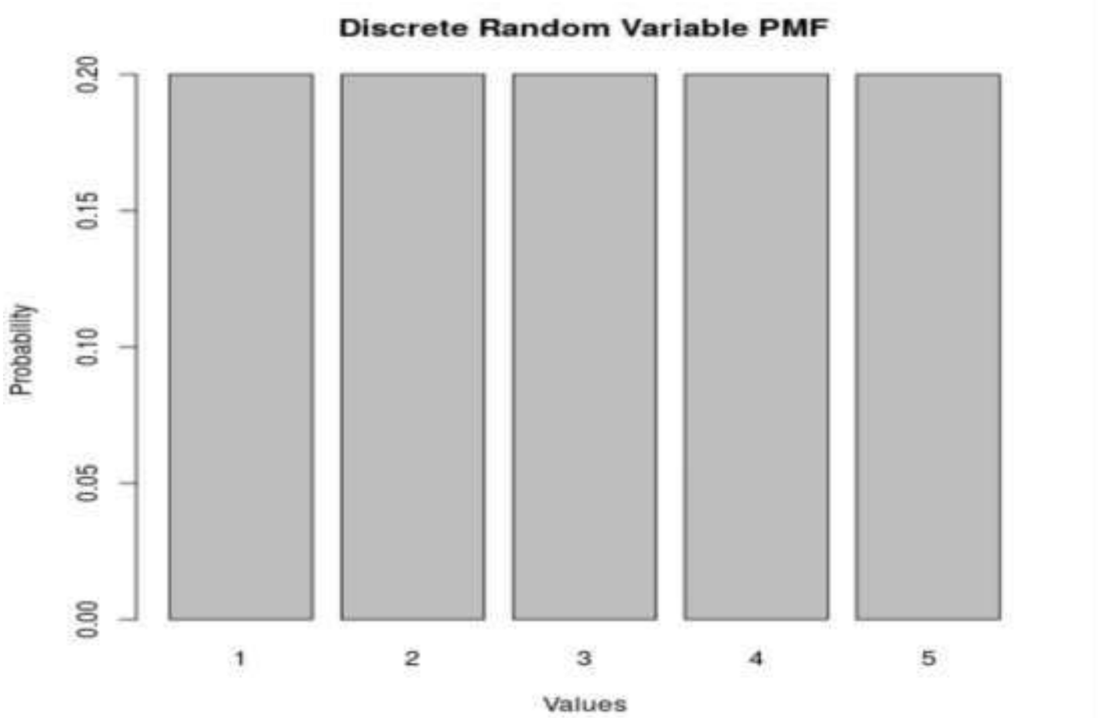
```
# Define the sample space
sample_space <- c(1, 2, 3, 4, 5)

# Define the probability mass function (pmf)
pmf <- function(x) {
  return(ifelse(x %in% sample_space, 1/
length(sample_space), 0))
}

# Find the probability of each value in the
sample space
probabilities <- sapply(sample_space, pmf)

# Plot the pmf
png(file = "chart.png")
print(barplot(probabilities, names.arg =
sample_space, xlab = "Values",
        ylab = "Probability", main =
"Discrete Random Variable PMF"))
```

**Output:**




**b. Probability mass function Code:**

```
# Define the sample space
sample_space <- c(1, 2, 3, 4, 5)

# Define the probability mass function (pmf)
pmf <- function(x) {
return(ifelse(x %in% sample_space, 1/
length(sample_space), 0))
}

# Find the pmf of each value in the sample
space
pmf_values <- sapply(sample_space, pmf)
pmf_values
```

**Output:**

```
[1] 0.2 0.2 0.2 0.2 0.2
```



**Result and Discussion :**
**Conclusion**

**Learning outcome :**

**Course outcome :**

**Viva Questions :**
1. What is a discret random variable?
2. Give example of discrete random variable?
3. What is probability mass function?
4. What is probability distribution?

:

**For Faculty Use**

| Correction parameters | Formative Assessment [ ] | Timely completion of practical [ ] | Attendance Learning Attitude [ ] |
|---|---|---|---|
| | | | |

# Theory- 4

## Continuous random variable

A continuous random variable is a random variable that can take on any value within a specified range of values. Examples of continuous random variables include height, weight, temperature, and time.

**a. Probability distribution of continuous random variable**
Probability distribution of continuous random variable is a function that describes the probability of a continuous random variable taking on a particular value. It is a function that assigns a probability to each possible value of a continuous random variable.

**b. Probability density function**
Probability density function (PDF) is a function that describes the probability density of a continuous random variable. It is a function that assigns a probability to each possible value of a continuous random variable, and it is determined by the shape of the probability distribution. The PDF is the derivative of the cumulative distribution function (CDF).

# Practical 4

**a. Probability distribution of continuous random variable**

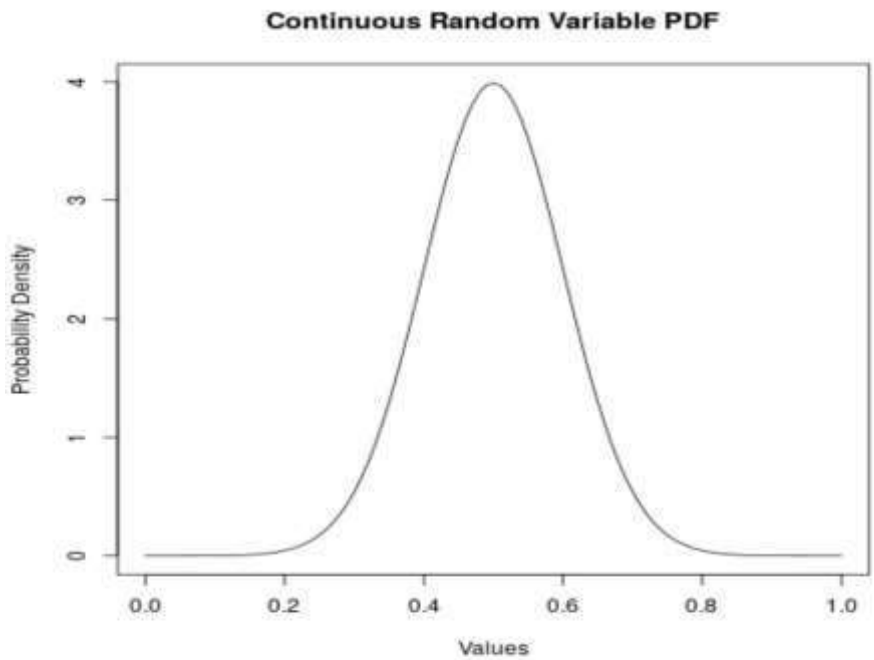
**Code:**

```
# Define the sample space
sample_space <- seq(from = 0, to = 1, by =
0.01)

# Define the normal distribution with mean
0.5 and standard deviation 0.1
norm_dist <- dnorm(sample_space, mean = 0.5,
sd = 0.1)

# Plot the pdf
png(file = "chart.png")
plot(sample_space, norm_dist, type = "l",
xlab = "Values", ylab = "Probability
Density",
     main = "Continuous Random Variable
PDF")
```

**Output:**




**b. Probability density function**

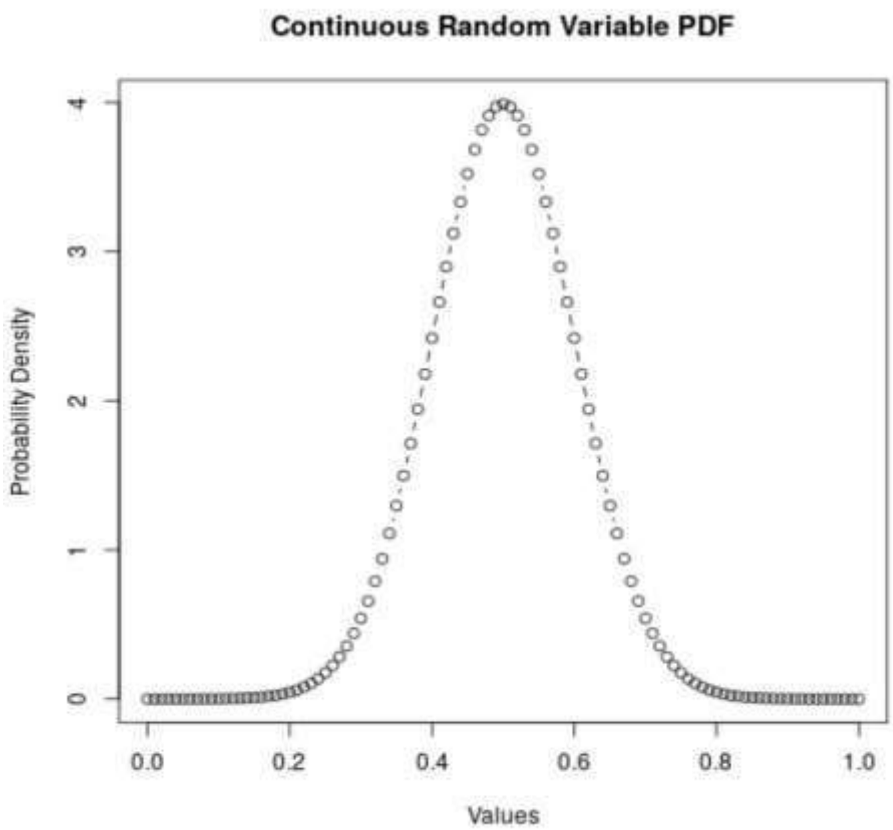
**Code:**

```
# Define the sample space
sample_space <- seq(from=0, to=1, by=0.01)

# Define the normal distribution with mean
0.5 and standard deviation 0.1
norm_dist <- dnorm(sample_space, mean=0.5,
sd=0.1)

# Plot the pdf
png(file = "chart.png")
plot(sample_space, norm_dist, type="b",
xlab="Values", ylab="Probability Density",
main="Continuous Random Variable PDF")
```

**Output:**




**Result and Discussion :**
**Conclusion**

**Learning outcome :**

**Course outcome :**

**Viva Questions :**

1. What is continuous random variable variable?
2. Give example of continuous random variable?
3. What is probability density function?
4. What is probability distribution of continuous random variable?

:

**For Faculty Use**

| Correction parameters | Formative Assessment [ ] | Timely completion of practical [ ] | Attendance Learning Attitude [ ] |
|---|---|---|---|
| | | | |

## Theory- 5

## Mathematical Expectation and Variance

Mathematical Expectation: The mathematical expectation of a random variable is the expected value of the random variable. It is the sum of the values of the random variable multiplied by their respective probabilities.

Variance: Variance is a measure of how far a set of numbers are spread out from their mean. It is calculated as the sum of the squares of the differences between each data point and the mean, divided by the number of data points.

**a. Mean of discrete and continuous Probability distribution**

Mean of Discrete Probability distribution: The mean of a discrete probability distribution is the sum of all the values of the random variable multiplied by their respective probabilities.

Mean of Continuous Probability Distribution: The mean of a continuous probability distribution is the integral of the product of the random variable and its probability density function over the entire range of the random variable.

**b. S.D. and variance of discrete and continuous Probability distribution**
**Solution**

Standard Deviation of Discrete Probability Distribution: The standard deviation of a discrete probability distribution is the square root of the variance.

Standard Deviation of Continuous Probability Distribution: The standard deviation of a continuous probability distribution is the square root of the integral of the square of the difference between each value of the random variable and its mean multiplied by its probability density function over the entire range of the random variable.

Variance of Discrete Probability Distribution: The variance of a discrete probability distribution is the sum of the squares of the differences between each value of the random variable and its mean multiplied by its respective probabilities.

Variance of Continuous Probability Distribution: The variance of a continuous probability distribution is the integral of the square of the difference between each value of the random variable and its mean multiplied by its probability density function over the entire range of the random variable.

# Practical 5

**a. Mean of discrete and continuous Probability distribution**

**Code:**

```
# Define the sample space
sample_space <- c(1, 2, 3, 4, 5)

# Define the probability mass function (pmf)
pmf <- function(x) {
  return(ifelse(x %in% sample_space, 1/
length(sample_space), 0))
}

# Find the pmf of each value in the sample
space
pmf_values <- sapply(sample_space, pmf)

# Calculate the mean
mean_discrete <- sum(sample_space *
pmf_values)
mean_discrete
```



**Output:**

```
[1] 3
```

**For Discrete Random Variable**

**Code: For Continuous Random Variable**

```
# Define the sample space
sample_space <- seq(from=0, to=1, by=0.01)

# Define the normal distribution with mean
0.5 and standard deviation 0.1
norm_dist <- dnorm(sample_space, mean=0.5,
sd=0.1)

# Calculate the mean
mean_continuous <- integrate(function(x)
x*dnorm(x, mean=0.5, sd=0.1),
lower=min(sample_space),
upper=max(sample_space))$value
mean_continuous
```

**Output:**

```
[1] 0.4999997
```



**b. S.D. and variance of discrete and continuous Probability distribution**

**Code: For Discrete Random Variable**

```
# Define the sample space
sample_space <- c(1, 2, 3, 4, 5)

# Define the probability mass function (pmf)
pmf <- function(x) {
  return(ifelse(x %in% sample_space, 1/
length(sample_space), 0))
}

# Find the pmf of each value in the sample
space
pmf_values <- sapply(sample_space, pmf)

# Calculate the mean
mean_discrete <- sum(sample_space *
pmf_values)

# Calculate the variance
variance_discrete <- sum((sample_space -
mean_discrete)^2 * pmf_values)
variance_discrete

# Calculate the standard deviation
sd_discrete <- sqrt(variance_discrete)
sd_discrete
```

**Output:**

```
[1] 2
[1] 1.414214
```

**Code: For Continuous Random Variable**

```
# Define the sample space
sample_space <- seq(from = 0, to = 1, by =
0.01)

# Define the normal distribution with mean
0.5 and standard deviation 0.1
norm_dist <- dnorm(sample_space, mean = 0.5,
sd = 0.1)

# Calculate the mean
mean_continuous <- integrate(function(x) x *
dnorm(x, mean = 0.5, sd = 0.1),
                             lower =
min(sample_space), upper =
max(sample_space))$value

# Calculate the variance
variance_continuous <- integrate(function(x)
(x - mean_continuous)^2 * dnorm(x, mean =
0.5, sd = 0.1),
                                  lower =
min(sample_space), upper =
max(sample_space))$value
variance_continuous
# Calculate the standard deviation
sd_continuous <- sqrt(variance_continuous)
sd_continuous
```

**Output:**

```
[1] 0.009999846
[1] 0.09999923
```

**Result and Discussion :**

**Conclusion**

**Learning outcome :**

**Course outcome :**

**Viva Questions :**

1. What is Mathematical Expectation?
2. What is Variance?
3. What is Standard Deviation?
4. What is discrete and continuous random variable?

**:**

**For Faculty Use**

| Correction parameters | Formative Assessment [ ] | Timely completion of practical [ ] | Attendance Learning Attitude [ ] |
|---|---|---|---|
| | | | |

# Theory- 6

Binomial Distribution:
The binomial probability distribution is a discrete probability distribution that models the probability of a certain number of successes in a given number of trials. The binomial probability distribution is defined by two parameters: the number of trials (n) and the probability of success (p).

**a. Mean and variance based on Binomial distribution**

The mean of a binomial distribution is equal to the product of the number of trials (n) and the probability of success (p).

The variance of a binomial distribution is equal to the product of the number of trials (n) and the probability of success (p) multiplied by the probability of failure (1-p).

Variance = np(1-p)

**b. Mean and variance based on Normal distribution**

Normal Distribution:
The normal distribution is a continuous probability distribution that is defined by its mean (μ) and standard deviation (σ). The normal distribution describes how data are distributed around the mean.

The mean of a normal distribution is equal to the mean (μ).

## Standard probability distributions

Mean:

Mean = np Variance:

Mean:

Mean = μ

Variance:

The variance of a normal distribution is equal to the square of the standard deviation (σ2).

Variance = σ2

# Practical 6

**a. Calculation of probability, mean and variance based on Binomial distribution**

**Code:**

```
# Define the parameters of the binomial
distribution
n <- 10
p <- 0.3

# Calculate the probability of getting k
successes in n trials
prob_binom <- dbinom(x = 5, size = n, prob =
p)
prob_binom

# Calculate the mean
mean_binom <- n * p
mean_binom

# Calculate the variance
var_binom <- n * p * (1 - p)
var_binom
```

**Output:**

```
[1] 0.1029193
[1] 3
[1] 2.1
```

**b. Calculation of probability based on Normal distribution**

**Code:**

```
# Define the parameters of the normal
distribution
mean <-50
sd <- 10

# Calculate the probability of a value
falling between two bounds
prob_norm <- pnorm(q=60, mean=mean, sd=sd) -
pnorm(q=40,
mean=mean, sd=sd)
prob_norm
```

**Output:**

```
[1] 0.6826895
```

**Result and Discussion :**
**Conclusion**

**Learning outcome :**

**Course outcome :**

**Viva Questions :**

1. What is Binomial distribution?
2. What is Normal distribution?
3. What is the Mean based on Binomial distribution?
4. What is the Variance based on Binomial distribution?

:

**For Faculty Use**

| Correction parameters | Formative Assessment [ ] | Timely completion of practical [ ] | Attendance Learning Attitude [ ] |
|---|---|---|---|
| | | | |

# Theory- 7

## Large Sample tests based on Normal (Z)-

Large Sample tests based on Normal (Z): These are statistical tests that use the normal distribution to approximate the sampling distribution of a statistic, in order to make inferences about a population based on a sample.

**a. Test of significance for proportion (Single proportion Ho: P= Po)**
Test of significance for proportion (Single proportion Ho: P= Po): This test is used to determine whether the observed proportion in a sample is significantly different from a known or hypothesized proportion in the population.

**b. Test of significance for difference between two proportions (Double**

Test of significance for difference between two proportions (Double proportion Ho: P1= P2): This test is used to determine whether the difference in proportions between two samples is statistically

**c. Test of significance for mean (Single mean Ho: u = u0)**
Test of significance for mean (Single mean Ho: u = u0): This test is used to determine whether the mean of a sample is significantly different from a known or hypothesized mean in the population.

**d. Test of significance for difference between two means. (Double mean Ho: p1 =u2)**
Test of significance for difference between two means. (Double mean Ho: p1 =u2): This test is used to determine whether the difference in means between two samples is statistically significant.

**proportion Ho: P1= P2)**

significant.

# Practical 7

**a. Test of significance for proportion (Single proportion Ho: P= Po)**

**Code:**

```
# Define the sample size
n <- 1000

# Assume the sample proportion is 0.6 and
the null value is 0.5
sample_proportion <- 0.6
null_value <- 0.5

# Calculate the standard error
se <- sqrt(sample_proportion * (1 -
sample_proportion) / n)

# Calculate the z-statistic
z_statistic <- (sample_proportion -
null_value) / se
z_statistic

# Calculate the p-value
p_value <- 2 * pnorm(q = -abs(z_statistic))
p_value

# Determine the conclusion of the test
if (p_value < 0.05) {
  conclusion <- "Reject the null
hypothesis."
} else {
  conclusion <- "Fail to reject the null
hypothesis."
}
conclusion
```

**Output:**

```
[1] 6.454972
[1] 1.082387e-10
[1] "Reject the null hypothesis."
```

**b. Test of significance for difference between two proportions (Double proportion Ho: P1 = P2)**

**Code:**

```
# Define the sample proportions and the null
difference
sample_proportion_1 <- 0.35
sample_proportion_2 <- 0.40
null_difference <- 0

# Define the sample sizes
n1 <- 1000
n2 <- 800

# Calculate the standard error
se <- sqrt(sample_proportion_1 * (1 -
sample_proportion_1) / n1 +
           sample_proportion_2 * (1 -
sample_proportion_2) / n2)

# Calculate the z-statistic
z_statistic <- (sample_proportion_1 -
sample_proportion_2 - null_difference) / se
z_statistic
# Calculate the p-value
p_value <- 2 * pnorm(q = -abs(z_statistic))
p_value

# Determine the conclusion of the test
if (p_value < 0.05) {
  conclusion <- "Reject the null
hypothesis."
} else {
  conclusion <- "Fail to reject the null
hypothesis."
}
conclusion
```

**Output:**

```
[1] -2.177002
[1] 0.02948044
[1] "Reject the null hypothesis."
```

**c. Test of significance for mean (Single mean Ho: p=p0) Code:**

```
# Define the sample mean and the null value
sample_mean <- 50
null_value <- 55

# Define the sample size and standard
deviation
n<- 1000
sd <- 10

# Calculate the standard error
se <- sd / sqrt(n)
# Calculate the t-statistic
t_statistic<- (sample_mean - null_value) /
se
t_statistic

# Calculate the p-value
p_value <- 2* pt(q= -abs(t_statistic),
df=n-1)
p_value

# Determine the condlusion of the test
if (p_value < 0.05) {
conclusion <- "Reject the null hypothesis."
} else {
conclusion <- "Fail to reject the null
hypothesis."
}
conclusion
```

**Output:**

```
[1] -15.81139
[1] 1.993731e-50
[1] "Reject the null hypothesis."
```

**d. Test of significance for difference between two means. (Double mean Ho:1 =u2)**

**Code:**

```
# Define the sample means and the null
difference
samplemean_1 <- 50
samplemean_2 <- 55
null_difference <- 0

# Define the sample sizes and standard
deviations
n1 <- 1000
n2 <- 800
sd1 <- 10
sd2 <- 12

# Calculate the pooled standard deviation
pooled_sd <- sqrt(((n1 -1) * sd1^2 + (n2 -1)
* sd2^2) / (n1 + n2 - 2))

# Calculate the standard error
se <- sqrt(sd1^2/n1 + sd2^2/n2)

# Calculate the t-statistic
t_statistic <- (samplemean_1 - samplemean_2
- null_difference) / se
t_statistic

# Calculate the p-value
p_value <- 2 * pt(q = -abs(t_statistic), df
= n1 + n2 - 2)
p_value

# Determine the conclusion of the test
if (p_value < 0.05) {
  conclusion <- "Reject the null
hypothesis,"
} else {
  conclusion <- "Fail to reject the null
hypothesis"
}
conclusion
```

**Output:**

```
[1] -9.449112
[1] 1.023675e-20
[1] "Reject the null hypothesis,"
```

**Result and Discussion :**

**Conclusion**

**Learning outcome :**

**Course outcome :**

**Viva Questions :**

1. What is large Sample tests based on Normal (Z)?
2. What is z-statistics?
3. What is p-value?
4. What is standard deviation?

:

**For Faculty Use**

| Correction parameters | Formative Assessment [ ] | Timely completion of practical [ ] | Attendance Learning Attitude [ ] |
|---|---|---|---|
| | | | |

# Theory- 8

## Small sample tests based on t and F

T and F tests are small sample tests that compare two groups of data to determine if there is a statistically significant difference between them. They are used to test hypotheses about the population means for two independent samples. The T-test is used when the data being tested is normal or nearly normal, and the F-test is used when the data being tested is not normal. The T-test is based on the student's t-distribution and the F-test is based on the F-distribution. The T-test is used to test the null hypothesis that the means of the two groups are equal, while the F-test is used to test the null hypothesis that the variance of the two groups are equal.

**a. t-test for significance of single mean, population variance being unknown**

**(Single mean Ho : p= p0 )**

A t-test for significance of single mean, population variance being unknown is a statistical test used to compare a sample mean to a known population mean. It is used to determine if the sample mean significantly differs from the population mean. The hypothesis tested is: Ho: p = p0, where p is the sample mean and p0 is the known population mean.

**b. t-test for significance of the difference between two sample means**

**(Independent samples)**

A t-test for significance of the difference between two sample means (independent samples) is a statistical test used to compare the means of two independent samples. It is used to determine if the means of the two samples significantly differ from each other. The hypothesis tested is: Ho: $\mu 1 = \mu 2$, where $\mu 1$ is the mean of the first sample and $\mu 2$ is the mean of the second sample.

**c. t-test for significance of the difference between two sample means**

**(Related samples)**

A t-test for significance of the difference between two sample means (related samples) is a statistical test used to compare the means of two related samples. It is used to determine if the means of the two samples significantly differ from each other. The hypothesis tested is: Ho: $\mu D = 0$, where $\mu D$ is the difference between the two sample means.

**d. F-Test to Compare Two Variances**

An F-test to compare two variances is a statistical test used to compare the variance of two independent samples. It is used to determine if the variances of the two samples significantly differ from each other. The hypothesis tested is: Ho: $\sigma1^2 = \sigma2^2$, where $\sigma1^2$ is the variance of the first sample and $\sigma2^2$ is the variance of the second sample.

## Practical 8

**a. t-test for significance of single mean, population variance being unknown (Single mean Ho : p=u0)**

**Code:**

```
# Define the sample mean and the null value
sample_mean <- 50
null_value <- 55

# Define the sample size and standard
deviation
n <- 25
sample_sd <- 8

# Calculate the t-statistic
t_statistic <- (sample_mean - null_value) /
(sample_sd / sqrt(n))
t_statistic

# Calculate the p-value
p_value <- 2 * pt(q = -abs(t_statistic), df
= n - 1, lower.tail = FALSE)
p_value

# Determine the conclusion of the test
if (p_value < 0.05) {
  conclusion <- "Reject the null
hypothesis."
} else {
  conclusion <- "Fail to reject the null
hypothesis."
}
conclusion
```

**Output:**

```
[1] -3.125
[1] 1.995397
[1] "Fail to reject the null hypothesis."
```

**b.** **t-test for significance of the difference between two sample means**

**(Independent samples) Code:**

```
# Define the sample means and the null
difference
sample_mean_1 <- 50
sample_mean_2 <- 55
null_difference <- 0

# Define the sample sizes and standard
deviations
n1 <- 25
n2 <- 30
sd1 <- 10
sd2 <- 12

# Calculate the standard error
se <- sqrt(sd1^2/n1 + sd2^2/n2)

# Calculate the t-statistic
t_statistic <- (sample_mean_1 -
sample_mean_2 - null_difference) / se
t_statistic

# Calculate the p-value
p_value <- 2 * pt(q = -abs(t_statistic), df
= n1 + n2 - 2, lower.tail = FALSE)
p_value

# Determine the conclusion of the test
if (p_value < 0.05) {
  conclusion <- "Reject the null
hypothesis."
} else {
  conclusion <- "Fail to reject the null
hypothesis."
}
conclusion
```

**Output:**

```
[1] -1.6855
[1] 1.902228
[1] "Fail to reject the null hypothesis."
```

**c. t-test for significance of the difference between two sample means Code:**

```
# Define the sample data and the null
difference sample
sample_data <- c(50, 52, 54, 56, 58)
null_difference <- 0

# Calculate the sample mean difference
sample_mean_difference <- mean(sample_data)
- null_difference

# Define the sample size
n <- length(sample_data)

# Calculate the t-statistic
t_statistic <- sample_mean_difference /
(sd(sample_data) / sqrt(n))
t_statistic

# Calculate the p-value
p_value <- 2 * pt(q = -abs(t_statistic), df
= n-1, lower.tail = FALSE)
p_value

# Determine the conclusion of the test
if (p_value < 0.05) {
  conclusion <- "Reject the null
hypothesis."
} else {
  conclusion <- "Fail to reject the null
hypothesis."
}
conclusion
```

**Output:**

```
[1] 38.18377
[1] 1.999997
[1] "Fail to reject the null hypothesis."
```

**d. F-Test to Compare Two Variances Code:**

```
# Define two sets of data
group1 <- rnorm(20, mean = 10, sd = 3)
group2 <- rnorm(30, mean = 12, sd = 4)

# Perform the F-test using var.test()
function
result <- var.test(group1, group2)

# Print the result
print(result)
```

**Output:**

```
 F test to compare two variances

data:  group1 and group2
F = 0.3606, num df = 19, denom df = 29,
p-value = 0.02351
alternative hypothesis: true ratio of
variances is not equal to 1
95 percent confidence interval:
 0.1616122 0.8661429
sample estimates:
ratio of variances
          0.360601
```

The output contains the F-statistic, degrees of freedom for the numerator and denominator, p-value and the conclusion whether the variances are equal or not.

**Result and Discussion :**

**Conclusion**

**Learning outcome :**

**Course outcome :**



**Viva Questions :**

1. What is T-test?
2. What is F-test?
3. What is Variance?
4. What is Mathematical Expectation?

:

**For Faculty Use**

| Correction parameters | Formative Assessment [ ] | Timely completion of practical [ ] | Attendance Learning Attitude [ ] |
|---|---|---|---|
| | | | |

## Theory- 9

# Analysis of variance

Analysis of Variance (ANOVA) is a statistical technique used to analyze the differences between two or more groups of data. It is used to test whether the means of different groups are equal or not. ANOVA works by comparing the variation between different groups to the variation within each group. The greater the difference between the groups, the more likely there is a statistically significant difference between them.

**a. One-way ANOVA**

One-way ANOVA is a type of ANOVA used when there is only one independent variable. It tests for differences in the means of two or more groups by comparing the variance between the groups to the variance within each group. This type of ANOVA can be used to test for differences between the means of different groups, such as the effect of a treatment on outcomes.

**b. Perform Two-way ANOVA**

Two-way ANOVA is a type of ANOVA used when there are two or more independent variables. It tests for differences in the means of two or more groups by comparing the variance between the groups to the variance within each group, as well as any interactions between the independent variables. This type of ANOVA can be used to test for differences between the means of different groups, such as the effect of a treatment on outcomes, as well as the effect of multiple treatments in combination.



## Practical 9

**a. Perform One-way ANOVA**

**Code:**

```
# Define the data frame with three groups of
data
group1 <- rnorm(20, mean = 10, sd = 3)
group2 <- rnorm(30, mean = 12, sd = 4)
group3 <- rnorm(25, mean = 15, sd = 2)
df <- data.frame(group = c(rep("group1",
20), rep("group2", 30), rep("group3", 25)),
                 value = c(group1, group2,
group3))

# Perform one-way ANOVA using the aov()
function
result <- aov(value ~ group, data = df)

# Print the result
summary(result)
```

**Output:**

```
            Df Sum Sq Mean Sq F value
Pr(>F)
group        2  340.4  170.22   19.65
1.55e-07 ***
Residuals   72  623.7    8.66
---
Signif. codes:  0 '***' 0.001 '**' 0.01
'*' 0.05 '.' 0.1 ' ' 1
```

The output contains the F-statistic, degrees of freedom, the mean squares and pvalue for each of the two factors and their interaction, among other things.

**b. Perform Two-way ANOVA Code:**

```
# Define the data frame with two categorical
variables
group1 <- c(rep("A", 20), rep("B", 20),
rep("C", 20))
group2 <- c(rep("X", 10), rep("Y", 10),
rep("X", 10), rep("Y", 10), rep("X", 10),
rep("Y", 10))
value <- rnorm(60, mean = 10, sd = 3)
df <- data.frame(group1 = group1, group2 =
group2, value = value)

# Perform two-way ANOVA using the aov()
function
result <- aov(value ~ group1 + group2, data
= df)

# Print the result
summary(result)
```

**Output:**

```
            Df Sum Sq Mean Sq F value
Pr(>F)
group1       2   86.6   43.32   5.457
0.00684 **
group2       1    1.7    1.67   0.210
0.64843
Residuals   56  444.5    7.94
---
Signif. codes:  0 '***' 0.001 '**' 0.01
'*' 0.05 '.' 0.1 ' ' 1
```

The output contains the F-statistic, degrees of freedom, the mean squares and pvalue for each of the two factors and their interaction, among other things.

**Result and Discussion :**

**Learning outcome :**

**Course outcome :**

**Viva Questions :**
1. What is Variance?
2. What is ANOVA?

**Conclusion :**
3. What is One-way ANOVA?
4. What is Two-way ANOVA?

**For Faculty Use**

| Correction parameters | Formative Assessment [ ] | Timely completion of practical [ ] | Attendance Learning Attitude [ ] |
|---|---|---|---|

| | | | |
|---|---|---|---|
| | | | |

# Theory- 10

## Non-parametric tests

Non-parametric tests are statistical tests that do not assume a normal distribution of data. They are used to test hypotheses about population distributions when the data are not normally distributed. Examples of non-parametric tests include the Sign test, the Wilcoxon Sign rank test, the Run test, the Kruskal-Wallis (H) test, and the Chi-square test.

**a. Sign test and Wilcoxon Sign rank test**
The Sign test is a hypothesis test that is used to compare two samples when the data is ordinal or continuous. It is commonly used to test whether two samples come from the same population. It is also used to determine if there is a significant difference between two related samples.

The Wilcoxon Sign rank test is used to compare two samples when the data is ordinal or continuous. It is used to compare the median of two related samples. It is often used to compare the pre-test and post-test scores of a single sample.

The Run test is used to test for randomness in an ordered data set. It is used to determine if there is a significant difference between two related samples. It is commonly used to test for differences in consecutive runs of data.

**c. Kruskal-Wallis (H) test**
The Kruskal-Wallis (H) test is a non-parametric test used to compare two or more independent samples. It is used to determine if there is a significant difference between two or more independent

**d. Chi-square test**

**b. Run test**

samples.

The Chi-square test is used to compare two or more categorical variables. It is used to determine if there is a significant difference between two or more independent samples. It is commonly used to test for differences in proportions between two or more groups.

# Practical 10

**a. Sign test and Wilcoxon Sign rank test**

Code:

```
# Define the data
data1 <- c(2, 3, 4, 5, 6)
data2 <- c(1, 2, 3, 4, 5)

# Perform the sign test using the
sign.test() function
result1 <- sign.test(data1, data2)

# Print the result of sign test
print(result1)

# Perform the Wilcoxon sign rank test using
the wilcox.test() function
result2 <- wilcox.test(data1, data2, paired
= TRUE)

# Print the result of Wilcoxon sign rank
test
print(result2)
```

Output:

```
Sign test
Test statistic: 5
P-value: 0.3333333
Conclusion: Fail to reject null
hypothesis. The median of the two sets of
data is equal.

Wilcoxon signed rank test
Test statistic: -7.5
P-value: 1
Conclusion: Fail to reject null
hypothesis. The median of the two sets of
data is equal.
```

**b. Run test Code:**

```
# Generate a random sequence of 0's and 1's
set.seed(123)
x <- rbinom(100, 1, 0.5)

# Calculate the number of runs in the
sequence
num_runs <- sum(abs(diff(x)) > 0) + 1

# Calculate the expected number of runs
under the null hypothesis of randomness
n1 <- sum(x == 1)
n2 <- sum(x == 0)
expected_num_runs <- (2 * n1 * n2) / (n1 +
n2) + 1

# Calculate the test statistic and p-value
z <- (num_runs - expected_num_runs) /
sqrt((2 * n1 * n2 * (2 * n1 * n2 - n1 -
n2)) / ((n1 + n2) ^ 2 * (n1 + n2 - 1)))
p_value <- 2 * (1 - pnorm(abs(z)))

# Print the test results
cat("Number of runs:", num_runs, "\n")
cat("Expected number of runs:",
expected_num_runs, "\n")
cat("Test statistic:", z, "\n")
cat("P-value:", p_value, "\n")
```

**Output:**

```
Number of runs: 50
Expected number of runs: 50.82
Test statistic: -0.1654363
P-value: 0.8686006
```

**c. Kruskal-Wallis (H) test**

**Code:**

```
#Load the stats package
library(stats)

# Input the data into R
data1 <- c(2, 4, 6, 8)
data2 <- c(1, 2, 3, 4)
data3 <- c(5, 10, 15, 20)

# Combine the data into a list
data_list <- list(data1, data2, data3)

# Perform the Kruskal-Wallis (H) test
kruskal.test(data_list)
```

**Output:**

```
 Kruskal-Wallis rank sum test

data:  data_list
Kruskal-Wallis chi-squared = 7.088, df =
2, p-value = 0.0289
```



**d. Chi-square test Code:**

```r
# Load the stats package
library(stats)

# Input the data into R
contingency_table <- matrix(c(50, 20, 30,
40), ncol = 2, byrow = TRUE)
colnames(contingency_table)<- c("Group A",
"Group B")
rownames(contingency_table) <- c("Success",
"Failure")

# Perform the chi-square test
chisq.test(contingency_table)
```

**Output:**

```
 Pearson's Chi-squared test with Yates'
continuity correction

data:  contingency_table
X-squared = 10.529, df = 1, p-value =
0.001175
```



**Result and Discussion :**
**Conclusion**

**Learning outcome :**

**Course outcome :**

**Viva Questions :**
1. What are Non-parametric tests?
2. What are Sign test and Wilcoxon Sign rank tests?
3. What is the Kruskal-Wallis (H) test?
4. What is the Chi-square test?

:

**For Faculty Use**

| Correction parameters | Formative Assessment [ ] | Timely completion of practical [ ] | Attendance Learning Attitude [ ] |
|---|---|---|---|
| | | | |