



A COMPREHENSIVE STUDY ON STATISTICAL TECHNIQUES AND R CODE IN RESEARCH ANALYSIS

Deepika Prabhu

Research Scholar, Department of Statistics, JAIN (Deemed-to-be University), Bengaluru, India.

Dr. Suma A P

Research Coordinator, Centre for Researcher Training and Administration, JAIN (Deemed-to-be University), Bengaluru, India.

ABSTRACT

Statistics is a scientific discipline focused on the collection, presentation, examination, and interpretation of numerical data. Statistics has a very important role in research design. A good grasp of statistics is essential for systematic study. This requires an appropriate selection of samples, techniques, and tools for analysis. Improper statistical techniques may lead to erroneous conclusions. R is a statistical computing language, and many programs are built into it for analysis. This article discusses various research types and the suitable statistical techniques to collect the sample and R codes for various analyses of the data and draw meaningful inferences.

KEYWORDS : Research Types, Statistical Techniques, Fundamental Research, Applied Research, Qualitative Research, Quantitative Inquiry, Parametric Test, Non-Parametric Test, R Code.

INTRODUCTION

Research involves the meticulous examination of a specific issue or challenge through the application of scientific methodologies. Currently, there is a significant amount of discussion surrounding research, occurring frequently in both academic settings and beyond. Numerous research studies are undertaken and accomplished year after year. A research study is said to be scientific and systematic if it has a good research design. In many instances, insufficient emphasis is placed on a crucial aspect of research, which is the consideration of research methodology and statistical techniques. Understanding the significance of fundamental concepts in statistical methods is essential for researchers. Sampling or Statistical design describes the techniques and procedures adopted to select the sample. The importance of the research is derived from its caliber, not solely its quantity. Thus, those engaged in research must diligently focus on devising and adhering to suitable methodologies in order to enhance the researcher's quality. This article helps a researcher choose the appropriate statistical technique for different kinds of research. Various types of research methodologies encompass Basic research, Applied research, Qualitative research, Quantitative research, Exploratory, Experimental, Longitudinal, Clinical research, and more.

Basic research:

It is a type of research in which the main aim is to understand and formulate scientific theories and to predict the nature of a phenomenon. This research is also called pure research or fundamental research or theoretical research. Fundamental research is primary research which revolves around simplification and the establishment of theories. Examples of fundamental research encompass investigations into natural phenomena or the realm of pure mathematics. Data collection is very essential in any research design. The following are the important methods of collecting data in basic research – Interview: In basic research, an interview represents a prevalent approach for gathering data, encompassing direct one-to-one interactions with individuals to acquire insights into a particular phenomenon. Interviews, as part of the research process, can take on different forms: structured, semi-structured, or unstructured. A structured interview, often referred to as a patterned interview, strategic interview, or standardized interview, is a quantitative approach that employs a predetermined set of questions to collect data about a research topic. Observation is another data-gathering method involving the careful observation of a phenomenon over a specific duration to obtain relevant information about its behaviors. Experiments, on the other hand, are quantitative data-gathering techniques used to either confirm or refute a hypothesis or test existing theories. A questionnaire, comprising a series of series of questions answered by research participants, serves as a cost-effective tool collecting extensive data from group members

simultaneously.

Examples of basic research in psychology include:

Can elevated stress levels lead to increased aggression in individuals?

How significantly do gender stereotypes contribute to the onset of depression?

Examples of basic research in Behavioral science include:

How does the level of caffeine intake impact concentration in the classroom?

Research on variances in behavior between children raised in family settings and those raised by single parents.

To study the association between literacy and smoking habit.

Example of Experiment method (Basic research):

A researcher wants to know whether the attribute 'smoking' and 'literacy' are independent. level of significance is 5%.

	Smokers	Nonsmokers
Literates	40	60
Illiterates	42	58

Here, H_0 : Smoking and literacy are not dependent on each other.

H_1 : Smoking and literacy are dependent attributes.

R-code for Chi-square test:

```
data <- matrix(c(40, 60, 42, 58), ncol=2, byrow=TRUE)
> colnames(data) <- c("smokers", "nonsmokers")
> rownames(data) <- c("literates", "illiterates")
> data <- as.table(data)
> #Perform Chi-Square Test of Independence
> chisq.test(data)
```

Output:

Chi-square test statistic=0.02067, p-value= 0.8857

Conclusion: Since the p-value is more than 0.05, we fail to reject the null hypothesis. Therefore the two attributes smoking and literacy are independent.

Applied Research:

Scientists in this field endeavor to discover prompt resolutions for challenges encountered by societal, industrial, or corporate entities. This form of research is also referred to as action research. It represents an unsystematic approach to

addressing particular problems or concerns, which may pertain to individual, group, or societal levels. It is commonly described as a scientific method since it leverages existing scientific methodologies to actively seek solutions. Hence, the primary objective of applied research is to identify a resolution for a practical problem. This research has three types. Evaluation research, a subtype of applied research, involves the analysis of existing data related to a research topic in order to derive unbiased research findings or make well informed decisions. This form of research is predominantly utilized in the business sector. Research and development: Research and development (R&D) represents a form of applied research with a specific emphasis on creating novel products and services tailored to the demands of target markets. Its primary focus revolves around gathering data concerning market requirements and devising strategies to enhance existing products or develop new ones that meet those identified needs. Action research employs data collection methods, which may include quantitative, qualitative, or a combination of both approaches. It is a subtype of applied research primarily geared toward delivering practical solutions to business-related challenges. Action research is employed when a researcher encounters a specific educational issue that requires resolution. Applied research is a crucial approach to research as it assists organizations in identifying practical solutions to a particular problems, simultaneously enhancing their efficiency and productivity. This stands in contrast to basic research, which concentrates on elucidating evidence without necessarily seeking solutions. This form of research is mainly applied in Psychology, Education, Health, Science, Business, etc. The various statistical techniques used for data collection are : Descriptive Analysis: Descriptive statistical analysis allows for organizing and summarizing large data into graphs and tables. The descriptive analysis involves various processes such as tabulation, a measure of central tendency, a measure of dispersion or variance, skewness, etc. Inferential Analysis: The inferential statistical analysis allows to extrapolate the data acquired from small sample size to the complete population. This type of analysis helps draw conclusions and make decisions about the whole population on the basis of sample data.

Predictive Analysis:

Predictive analysis is used to make a prediction of forthcoming measures. This analysis is used by marketing companies, insurance organizations, data-driven marketing, and financial corporations. Exploratory Data Analysis: EDA is generally the initial step of the data analysis process that is conducted before performing any other statistical analysis technique. It completely focuses on analyzing patterns in the data to identify possible relationships. Causal Analysis: In this type of analysis we study the cause-and-effect relationship between the variables. The purpose of causal analysis is to find the root cause of a problem instead of finding symptoms. We use a Cross lagged model, a latent growth model (LGM), and a synthesis of the two, an autoregressive latent trajectories model(ALT) as statistical methods for causal analysis.

Examples of applied research in psychology include:
Investigating treatment and management options for anxiety and panic attacks.

Investigating factors that improve workers' productivity.
Examples of applied research in Business include:
Applied research to improve an organization's hiring process.
Applied research to improve employee's efficiency and organizational policies.

Examples of applied research in Health include:
A study to identify the healing properties of a specific herb.

A study to identify the side effects of using a particular drug.

Example for descriptive analysis (Applied research):

A researcher wants to know whether advertisement helps increase sales. Advertisement cost and Sales are given. A suitable tool for checking related variation between variables is Spearman's rank correlation coefficient.

Advertisement Cost(in'000 Rs)	39	65	62	90	82	75	25	98	36	78
Sales(in lakh Rs)	47	54	58	86	62	68	60	91	51	84

```
Solution:
R-code:>
Advertisement Cost=c(39,65,62,90,82,75,25,98,36,78)
> Sales=c(47,54,58,86,62,68,60,91,51,84)
> #Obtain correlation coefficient
> cor.test(x,y,method = "spearman")
```

Output is:
rho= 0.8333, p-value=0.01538
Conclusion: Since the p-value is more than 0.05, we fail to reject the null hypothesis. That is, there is a high positive correlation between advertisement cost and sales.

Qualitative research:

It is concerned with qualitative phenomena, i.e., phenomena concerning to or involving quality or kind. It is a type of research that aims to gather and analyze descriptive data in order to gain an understanding of individuals' social reality, including understanding their attitudes, beliefs, and motivation. Qualitative research is often used to discover composite phenomena or to gain insight into people's experiences and perspectives on a particular topic. This type of research can also be used in the business world if we want to know what a specific group thinks about new ideas for products or services or simply want to try something since it allows you to access content-rich information about perceptions, emotions, and interactions between people. This form of research is mainly applied in Psychology, Social and Behavioral Science, Business, etc. Qualitative research is particularly important in the behavioral sciences where the aim is to discover the underlying objects of human behavior. Through such research, we can analyze the various factors which inspire people to behave in a particular manner or which make people like or dislike a particular thing.

- Examples of Qualitative research in Business include:
- Impact of social media reviews on brand perception.
 - Consumer's attitude towards food products advertising.

- Examples of Qualitative research in Education include:
- Teacher perceptions of professional learning communities.
 - Course scores and preservation comparing online and face-to-face classes.

- Example of Qualitative research in Social and Behavioral Science include:
- Level of social media addiction and mental health research.

Example for Qualitative research:

A researcher interested in studying whether intensive coaching helps in improving marks. The suitable method is paired t-test. The hypothesis is,
H0: The marks before and after coaching remains the same.
H1: The marks after coaching shows an improvement.

```
R-code :
> before <- c(50,42,51,26,35,42,60,41,70,55,62,38)
> after <- c(62,40,61,35,30,52,68,51,84,63,72,50)
```

```
> data<-data.frame(subject = rep(c(1:12), 2), time =
rep(c("before", "after"), each = 12),
score = c(before, after))
> t.test (before, after, paired=TRUE, alternative = "greater",
mu = 0, conf.level = 0.95)
```

Output:

t = -4.8851, p-value = 0.9998

Conclusion: Since the p-value is more than 0.05, we fail to reject the null hypothesis. That is intensive coaching does not help in improving marks.

Quantitative research:

It is the process of collecting, presenting, and analyzing numerical data. It is constructed on the measurement of quantity or amount. It is a research methodology used by researchers to test theories and hypothesis about the attitudes and behaviors of their customers on the basis of numerical and statistical evidence. Quantitative research is helpful when you collect data from a large group of diverse populations. This type of research is widely used in the Natural and Social sciences: Chemistry, Biology, Economics, Marketing, Psychology, Market research, etc. It helps in collecting reliable and accurate data. There are different types of quantitative research methods, namely: Descriptive Research Design, Observational Research Design, Survey Research, Correlation Research Design, Experimental Research Design, and Quasi-experimental Research Design.

Descriptive research is used to understand a phenomenon, situation, or population. This type of research helps us to understand the current status or to measure or to summarize the variables. This is an appropriate choice when the research aim is to identify characteristics, trends, frequencies, and categories. The research design should be cautiously developed to guarantee that the results are valid and reliable. The main characteristic of this method is that the researcher has no control over the variables; he can only report what has happened and or what is happening. Observations: Observations allow us to gather data on behaviors and phenomena without having to rely on the honesty and accuracy of respondents. This method is often used by psychological, social, business and market researchers to understand how people act in real-life situations. Survey research is described as the practice of gathering information through surveys distributed by researchers to individuals chosen as survey respondents. The main goal of a survey is to widely gather and describe the characteristics of a target population or customers. This allows us to gather large volumes of data that can be analyzed for frequencies, averages, and patterns. Common uses of surveys include: Describing the demographics of a country or region and evaluating satisfaction with a company's products or an organization's services. In correlation research, we investigate relationships between the variables. It is a type of non-experimental research method that studies the relationship between two variables with the help of statistical analysis. Correlation can be positive, negative, or zero. In experimental research, we systematically examine whether there is causation between variables. Here we follow scientific research to test the hypothesis. Experimental research can be Pre-experimental, True experimental, or Quasi-experimental research. Quasi-experimental research aims to identify a cause-and-effect relationship between two variables; independent and dependent variables. It involves subjects being assigned to groups based on non-random criteria. Within the realm of quantitative research methodologies, this approach is frequently utilized when it becomes impractical or ethically problematic to conduct true experiments.

Examples of Quantitative research in Biology:

It is used to make accurate measurements to test a hypothesis

in order to compare experimental data with predictions generated by theoretical models.

Examples of Quantitative research in Pharmacology:

By using quantitative research, we can calculate the right amount of dosage of a medicine that is compatible and safe for the patients.

Examples of Quantitative research in Business:

Quantitative analysis has found wide applications in project management. These methods are employed to enhance the efficient distribution of human resources, machinery, materials, finances, and time.

Quantitative methods offer guidance on how much raw material to purchase, levels of inventory to keep and various costs involved in shipping and storing finished products.

Examples of Quantitative research in Agriculture:

Operations research techniques have long been used by farmers. They rely on decision trees and make weather forecast-based assumptions to determine the choice of crops to cultivate.

The statistical techniques used for data collection and analysis are- descriptive statistical method and inferential statistical method. Under the inferential statistical method, we have parametric and non-parametric tests.

Descriptive Statistics: Descriptive Statistics serves the purpose of summarizing or providing a description of the attributes or features within a dataset. The main purpose of descriptive statistics is to provide information about the population. Descriptive statistics can be divided into three main types. i.e., Frequency distribution, Central tendency, and Dispersion. Frequency distribution is the distribution of variable values and their frequencies. It simplifies complex data. It reveals the trends and tendencies of the data. Central tendency is the property of concentration of the values around a central value in data. The main objective is to get a single value to represent the characteristics of the entire mass of data. The different measures of central tendency are : Arithmetic mean, Median, Mode, Geometric mean, Harmonic mean. These measures are also called positional averages.

Dispersion: The study of scatter or variation of values in data from any measure of central tendency is called dispersion. In other words, measures that measure the lack of uniformity are called measures of dispersion or variation. The main objectives of dispersion are to judge the reliability of an average, to compare two or more series with respect to its variability, and to enable the use of other statistical measures. Measures of dispersion can be categorized as either absolute or relative in nature. Types of absolute measures of variation: Range, Quartile Deviation (Q.D), Mean Deviation (M.D), and Standard Deviation. The corresponding relative measures of variation are, Coefficient of Range, Coefficient of Quartile Deviation, Coefficient of Mean Deviation, Coefficient of Standard Deviation.

Inferential Statistics: Statistical Inference is the theory of making decisions about population parameters using sampling and the concept of probability. It has two branches namely, Estimation and Testing of Hypothesis. Estimation: The process of obtaining the most likely value of the population parameter using statistic is called estimation. Estimation can be point- estimation or interval estimation. If a single value is proposed as an estimate of the unknown parameter, then it is a point estimation. If an interval is projected as an estimate of the unknown parameter, then it is interval estimation. Testing of Hypothesis: A statistical hypothesis is a statement regarding the parameters of the population. While testing a

statistical hypothesis, we come across two-tailed and one-tailed tests. A two-tailed test is a test of a statistical hypothesis, where the rejection region is located at both the tails of the probability curve of the null distribution of the test statistic. A one-tailed test is a test of a statistical hypothesis, where the rejection region will be located at only one tail of the probability curve of the null distribution of the test statistic. Testing of a hypothesis generally involves testing the null and the alternative hypothesis. A null hypothesis is a hypothesis that is being tested for possible rejection, under the assumption that it is true. On the other hand, the alternative hypothesis is the hypothesis which is accepted when null hypothesis is rejected. There are significant tests, like t-test, z-test, f-test, chi-square test, etc. that are mentioned to as quantitative techniques in quantitative analysis. The errors that occur in the quantitative analysis are basically categorized into two categories, namely, Type I error and Type II error.

Parametric test:

Parametric tests are those that assume that the sample data comes from a population that follows a normal distribution with a fixed set of parameters. Common parametric tests are attentive on analyzing and comparing the mean or variance of data.

Various types of parametric tests are:

Student's t-tests: This is a small sample test. The important applications of t-tests are,

- To test the significance of the mean of a population using small sample.
- To test the difference between the means of the two populations using two small samples. (independent samples).
- To test the difference between the means of the two populations using paired observations. (dependent samples).

Z- test: Here tests are based on large samples. Sampling distributions of large samples($n \geq 30$) approach a normal distribution. The important applications of Z-tests are,

- Test for a population mean.
- Evaluating the equality of means of two populations.
- Test for a population proportion.
- Evaluating the equality of proportions of two populations.

ANOVA tests: The analysis of variance frequently referred to by the contraction ANOVA is a statistical technique specially designed to test whether the means of more than two quantitative populations are equal. Analysis of variance can be a one-way classification or a two-way classification. Using ANOVA we can calculate the variance between the samples, variance within the samples, and F-distribution measures, the ratio of the variance between groups to the variance within groups.

Non-parametric tests: When the data is skewed type, non-parametric tests are used to evaluate the data. Non-parametric models are distribution-free models. It does not mean that these models do not have any parameters, but the number of parameters are not predefined. Various types of non-parametric tests are,

Kruskal Wallis Test: This test is used when there are more than two independent samples.

Wilcoxon's Signed Rank Test: When the given data is dependent and ordinal, we use Wilcoxon's Signed Rank test.

Sign Test: This test is used when the samples are paired.

Mann Whitney U test: When two independent samples are

given and it is ordinal in nature, to compare the outcomes we use Mann Whitney U test.

Example for Quantitative research using descriptive statistics: A researcher wants to know whether there is a correlation between marks obtained by students in two different subjects.

Marks in Statistics	25	43	27	35	54	61	37	45
Marks in Mathematics	35	47	20	37	63	54	28	40

Solution:

H0: There is no correlation between marks scored in Statistics and Mathematics.

H1: There is a correlation between marks scored in Statistics and Mathematics.

R-code:

```
> x=c(25,43,27,35,54,61,37,45)
> y=c(35,47,20,37,63,54,28,40)
> #Obtain correlation coefficient
> cor.test(x,y,method="pearson")
```

Output is:

Correlation coefficient = 0.8933, p-value= 0.009159

Conclusion:

Since the p-value is less than 0.05, we reject the null hypothesis.

There is a highly positive correlation between marks scored in Mathematics and Statistics.

Example for Quantitative research using Parametric test Two independent sample t-test

	Group I	Group-II
Sample size	7	5
Mean	50kg	48kg
Variance	5kg ²	3kg ²

A researcher wants to know whether two different types of drugs are effective in increasing weight.

Solution:

H0: Two drugs do not differ significantly in increasing weight.

H1: Two drugs do differ significantly in increasing weight.

R-code:

```
# weight of persons using drug A
> weight_before <- rnorm(7, mean = 50, sd = 2.236)
> # weight of persons using drug B
> weight_after <- rnorm(5, mean = 48, sd = 1.732)
> # draw the distribution
> t.test(weight_before, weight_after, alternative = "greater", var.equal = TRUE)
```

Output is:

t-test statistic = 3.0392, p-value= 0.00624

Conclusion:

Since p-value= 0.00624 which is less than 0.05, we reject the null hypothesis. Therefore Two drugs do differ significantly in increasing weight.

Example for Quantitative research using Fisher's F test

The performance of two sets of groups of students is known. The researcher wants to know whether both group performs same.

H0: The performance of two sets of groups of students are same.

H1: The performance of two sets of groups of students are not same.

R-code is:

```
x <- rnorm(48, mean = 50)
> y <- rnorm(78, mean = 30)
> # var test in R
```

```
> var.test(x, y, alternative = "two.sided")
```

Output:
 F = 1.0586, p-value = 0.8117,
 95% confidence interval: [0.6408 1.8070]
 The p-value of the F-test is = 0.8117 > 0.05, we fail to reject the null hypothesis.

Conclusion: There is no significant difference between the marks obtained by two samples.

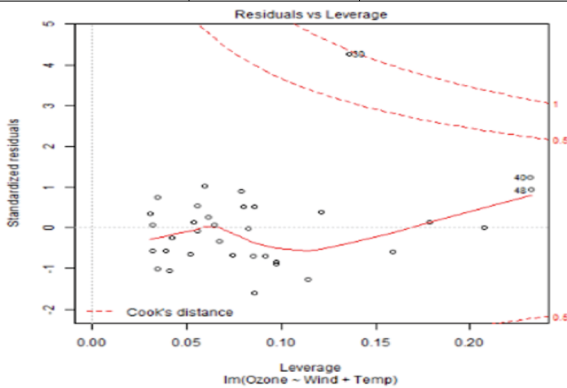
Example for Quantitative research for Developing Multiple Linear Regression model

A researcher is interested in developing a multiple regression model to check air quality.

```
R-code is:
# Using air quality dataset
> input <- airquality[1:40,c("Ozone", "Wind", "Temp")]
# To create regression model
> model <- lm(Ozone~Wind + Temp, data = input)
# Print the regression model
> cat("Regression model:\n")
> print(model)
> plot(model)
```

Output is:
 Coefficients are given by,

Intercept	Wind	Temp
-73.397	-1.124	1.641



Conclusion: The developed regression model is given by,
 $Y = -73.397 - 1.124 X_1 + 1.6141 X_2$

Example for Quantitative research using One-way ANOVA

A researcher is interested in studying if the average weights of the plants in the three experimental situations vary significantly.

H0: the average weights of the plants in the three experimental situations does not vary significantly.

H1: the average weights of the plants in the three experimental situations vary significantly.

```
R-code is:
> library("ggpubr")
dplyr::sample_n(data, 10)
> ggboxplot(data, x = "group", y = "weight",
color = "group", palette = c("#00AFBB", "#E7B800",
"#FC4E07"),
order = c("ctrl", "trl1", "trl2", ylab = "Weight", xlab = "Treatment"))
```

	1	2	3	4	5	6	7	8	9	10
Weight	5.87	4.32	3.59	5.18	5.14	4.89	5.12	4.81	4.50	4.69
group	trl1	trl1	trl1	ctrl	ctrl	trl1	trl2	trl1	ctrl	trl1

```
> res.aov <- aov(weight ~ group, data = data)
```

```
> summary(res.aov)
```

Output is:

Group	d f	Sum sq	Mean sq	F value	p-value
Residuals	2	3.766	1.8832	4.846	0.0159

Conclusion:
 Since p-value = 0.0159 < 0.05, we reject H0. There-fore the average weights of the plants in the three experimental situations vary significantly.

Some other types of Research:

All other types of research are distinctions of one or more of the above-specified methods, based on either the purpose of research, the time required to achieve research, on the environment in which research is done, or on the basis of some other similar factor. In the former case the research is confined to a single time period, whereas in the latter case, the research is carried on over several time periods. Research can be simulation research or field-setting research or laboratory research, depending upon the environment in which it is to be carried out. Research follows case-study methods or in-depth approaches to reach the basic causal relations. The research may be formalized or it may be exploratory. The objective of exploratory research is the development of a hypothesis instead of testing. Formalized research studies are those with significant structure and with detailed hypotheses to be tested. Example of Longitudinal research

Longitudinal data analysis - Longitudinal data can be observed as a special case of multilevel data where time is nested within individual participants. All longitudinal data share the following structures: (1) the same objects are repeatedly observed over time; (2) the same measurements are used; and (3) the timing for each measurement is identified.

Examples:
 Blood pressure in patients was measured every week for 5 weeks.
 Student enrolment in extracurriculars each semester.

Example of biomedical research:

Calculating Odds Ratio for a Cohort or Case-Control Study
 The odds ratio is commonly calculated in a case-control study. For example, the odds ratio for association between the exposure and the disease is the ratio of (i) the probability of having the disease for people who were exposed to the treatment or environment factor, divided by the probability of not having the disease for people who were exposed, and (ii). The probability of people who were affected and not exposed to the treatment or environmental factor, divided by the probability of not having the disease for people who were not exposed. For example,

	Affected	Neutral
Exposed	155	9420
Unexposed	1530	14796

```
R-code is:
mymatrix< matrix(c(155, 9420,1530, 14796), nrow=2, byrow=
TRUE)
> colnames(mymatrix) <- c("Affected","Neutral")
> rownames(mymatrix) <- c("Exposed","Unexposed")
> print(mymatrix)
      Affected  Neutral
Exposed    155    9420
Unexposed 1530   14796
>
> calcOddsRatio <-
function(mymatrix,alpha=0.05,referencerow=2,quiet=FALS
E)
{ numrow <- nrow(mymatrix)
  myrownames <- rownames(mymatrix)
```

```

for (i in 1:numrow)
{ rowname <- myrownames[i]
  AffectedUnexposed <- mymatrix[referencerow,1]
  NeutralUnexposed <- mymatrix[referencerow,2]
  if (i != referencerow)
  { AffectedExposed <- mymatrix[i,1]
    NeutralExposed <- mymatrix[i,2]
    totExposed <- AffectedExposed + NeutralExposed
    totUnexposed <- AffectedUnexposed + NeutralUnexposed
    probDiseaseGivenExposed <- AffectedExposed/totExposed
    probDiseaseGivenUnexposed <- AffectedUnexposed/totUnexposed
    probControlGivenExposed <- NeutralExposed/totExposed
    probControlGivenUnexposed <- NeutralUnexposed/totUnexposed
    # calculate the odds ratio
    oddsRatio <- (probDiseaseGivenExposed * probControlGivenUnexposed) /
    (probControlGivenExposed * probDiseaseGivenUnexposed)
    if (quiet == FALSE)
    { print(paste("category =", rowname, ", odds ratio = ", oddsRatio))
    }
    # calculate a confidence interval
    confidenceLevel <- (1 - alpha) * 100
    sigma <- sqrt((1/AffectedExposed) + (1/NeutralExposed) +
    (1/AffectedUnexposed) + (1/NeutralUnexposed))
    # sigma is the standard error of our estimate of the log of the odds ratio
    z <- qnorm(1-(alpha/2))
    lowervalue <- oddsRatio * exp(-z * sigma)
    uppervalue <- oddsRatio * exp(z * sigma)
    if (quiet == FALSE)
    { print(paste("category =", rowname, ", confidenceLevel, % confidence interval = [", lowervalue, ", ", uppervalue, "]"))
    }
  }
}

if (quiet == TRUE && numrow == 2) # If there are just two treatments (exposed/nonexposed)
{ return(oddsRatio)
}
}

```

> calcOddsRatio(mymatrix,alpha=0.05)
Output:
Odds ratio= 0.15912
95% Confidence Interval = [0.1346 0.1880]
Conclusion:

If the odds ratio =0.15912 (< 1), there is evidence of a negative association. Therefore the odds ratio can be used for estimation of either a cohort study or a case-control study.

Example of biomedical research:

Calculating Relative Risks for a Cohort Study

A very common type of data set in biomedical statistics is a cohort study, where we have information on people who were exposed to some treatment or environment (for example, people who took a certain drug, or people who smoke) and also on whether the same people have a particular disease or not. Suppose the following data regarding people who are having disease and are given exposure.

	Disease	Control
Exposed	155	9420
Unexposed	1530	14796

```

R-code is:
calcRelativeRisk <- function (mymatrix, alpha=0.05, referencerow=2)
{ numrow <- nrow(mymatrix)
myrownames <- rownames(mymatrix)
for (i in 1:numrow)

```

```

{ rowname <- myrownames[i]
DiseaseUnexposed <- mymatrix[referencerow,1]
ControlUnexposed <- mymatrix[referencerow,2]
if (i != referencerow)
{ DiseaseExposed <- mymatrix[i,1]
ControlExposed <- mymatrix[i,2]
totExposed <- DiseaseExposed + ControlExposed
totUnexposed <- DiseaseUnexposed + ControlUnexposed
probDiseaseGivenExposed <- DiseaseExposed/totExposed
probDiseaseGivenUnexposed <- DiseaseUnexposed/totUnexposed
# calculate the relative risk
relativeRisk <- probDiseaseGivenExposed / probDiseaseGivenUnexposed
print(paste("category =", rowname, ", relative risk = ", relativeRisk))
# calculate a confidence interval
confidenceLevel <- (1 - alpha) * 100
sigma <- sqrt((1/DiseaseExposed) - (1/totExposed) +
(1/DiseaseUnexposed) - (1/totUnexposed))
# sigma is the standard error of estimate of log of relative risk
z <- qnorm(1-(alpha/2))
lowervalue <- relativeRisk * exp(-z * sigma)
uppervalue <- relativeRisk * exp(z * sigma)
print(paste("category =", rowname, ", confidenceLevel, % confidence interval = [", lowervalue, ", ", uppervalue, "]"))
}}
> calcRelativeRisk(mymatrix,alpha=0.01)

```

Output :
Relative risk= 0.1727
Confidence Interval= [0.1393 0.2141]

Conclusion:
This tells us that the estimate of the relative risk is about 0.1727, and that a 99% confidence interval is [0.1393 0.2141]. A relative risk of 0.1727 means that the risk of disease in people who are exposed (to the treatment or environmental factor etc. that we are examining) is 0.1727 times the risk of disease of people who are not exposed.

CONCLUSION:

Statistical analysis is used in almost all fields to study the vast data that are available. It is important that a researcher knows the concepts of the basic statistical methods used for the conduct of a research study. This type of well-designed study leads to efficient and reliable results. This article also gives R code for various statistical tests. Many examples of descriptive analysis, parametric tests, non-parametric tests, clinical trials, etc. have been presented with R code and output for the same.

REFERENCES:

1. Discover Ph.D., Types of Research-Explained with Examples, October 2, 2020.
2. Jacqueline Bloomfield and Murray J Fisher, Quantitative Research Design, August 01, 2019, Journal of the Australasian Rehabilitation Nurses Association.
3. Jin Zhang et.al, Application of Inferential Statistical methods in Library and Information Science, August 2018.
4. Indeed Editorial Team, 7 Types of Statistical Analysis Techniques (And Process Steps), June 2, 2021 career guide.
5. Mara Calvello, Statistical Analysis Methods That Take data to the Next Level, March 6, 2020.
6. Shrutika Sirisilla, Effective Use of Statistics in Research- Methods and tools for Data Analysis Dec 19, 2022 enago academy.
7. Watson, Roger, Quantitative Research, London Vol 29, April 1, 2015.
8. Zulfiqar Ali and S Bala Bhaskar, Basic Statistical Tools in Research and data analysis, 2016 Sep; 60(9):662-669, Indian Journal of Anaesthesia.
9. <https://influentialpoints.com/notes/nrcmcl1a.htm>
10. <https://www.educba.com/types-of-research-methodology/>
11. <https://www.formpl.us/blog/applied-research>
12. <https://www.scribbr.com/methodology/types-of-research/>
13. <https://www.questionpro.com/blog/qualitative-research-methods/>
14. <https://www.voxco.com/blog/applied-research/>
15. <https://influentialpoints.com/notes/nrcmcl1a.htm>
16. <https://www.geeksforgoeks.org/t-test-approach-in-r-programming/>
17. <https://www.youtube.com/watch?v=RIhnNbPZCOA>
18. <https://www.guru99.com/t-t-test-one-sample.html>
19. <http://www.sthda.com/english/wiki/paired-samples-t-test-in-r>

20. <https://www.educba.com/statistical-analysis-with-r/>
21. <https://statisticsbyjim.com/basics/introduction-statistics-r-programming-language/>
22. https://www.researchgate.net/publication/356782229_Use_of_Statistics_in_Research