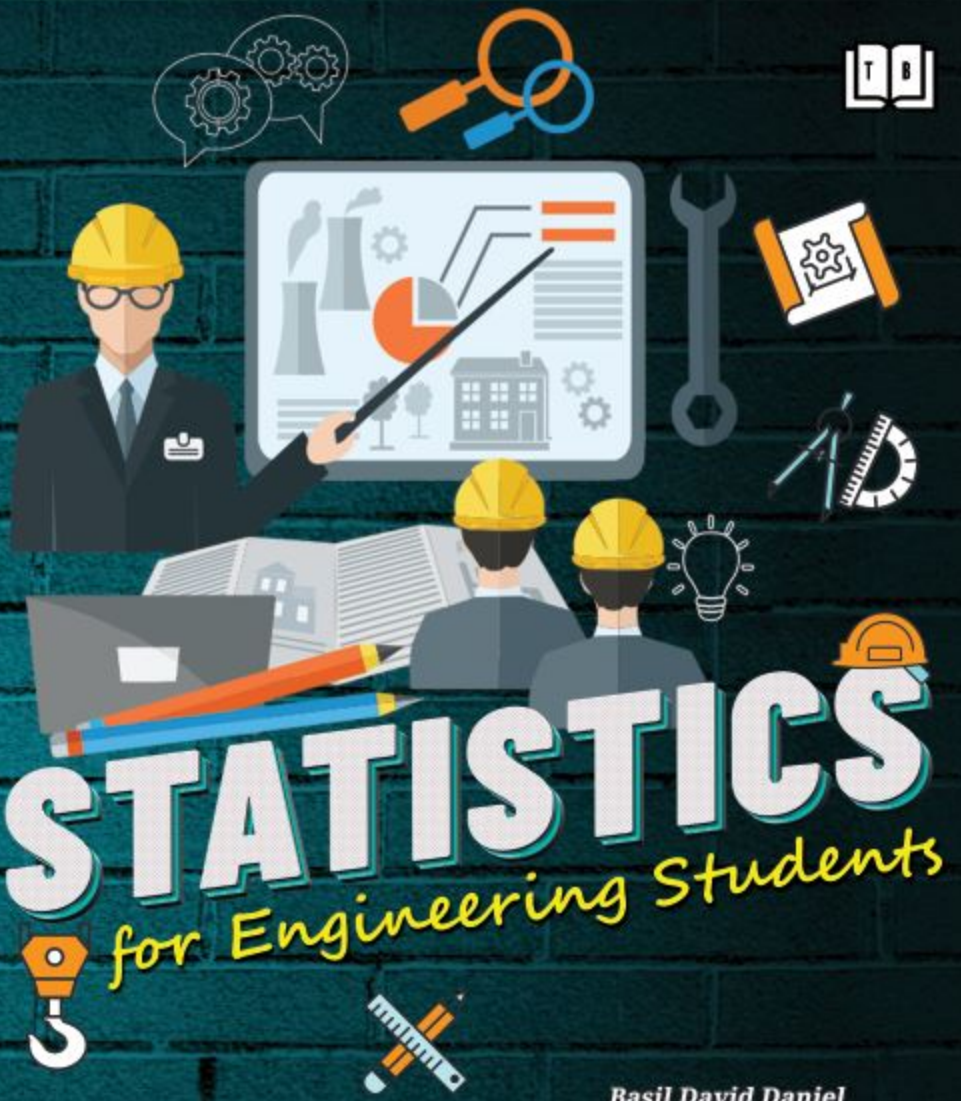
STATISTIC FOR ENGINEERING STUDENTS

Authors:
Basil David Daniel

email:
basil@uthm.edu.my

Abstract:
Engineers have long valued statistics because they may apply them to understand the data gathered from field work, experiments, testing, assessments, research and other operations. Numerous aspects of engineering largely rely on statistics, including the use of statistical models for problem-solving and making predictions, statistical tests for testing hypotheses, and probability-based decision-making. Statistics For Engineering Students was specifically written with pre-university and university students in mind, and with the aim of providing a better understanding on statistics and how it may be used in the engineering world. This book explores common statistical tests and analysis that are presented in the simplest manner as possible to aid students' comprehension of the topics covered, particularly descriptive and inferential statistics. Examples and exercises dealing with parametric and non-parametric tests have been included in this book as well. This is not only to ensure students thoroughly understand the need for conducting statistical analysis, but are able to relate statistics with engineering, and apply them if required in their respective job scopes when they embark on a career in the engineering industry.

Keywords: statistical models, statistical tests and analysis, tatistics with engineering



STATISTICS

for Engineering Students

Basil David Daniel

Penerbit UTHM

STATISTICS
for Engineering Students

Basil David Daniel

Penerbit UTHM
2024

Published 2024

Author:
Basil David Daniel

Published by:
Penerbit UTHM
Universiti Tun Hussein Onn Malaysia
86400 Parit Raja, Batu Pahat, Johor
Tel: 07-453 8698/8529
Fax: 07-453 6145

Website: http://penerbit.uthm.edu.my
E-mail: pt@uthm.edu.my
http://e-bookstore.uthm.edu.my
Penerbit UTHM is a member of
Majlis Penerbitan Ilmiah Malaysia
(MAPIM)

Printed by:
Mihas Grafik Sdn. Bhd.
No.8, 10, Jalan SR 4/18,
Taman Serdang Raya,
43300 Seri Kembangan, Selangor

Cataloguing-in-Publication Data

Perpustakaan Negara Malaysia

A catalogue record for this book is available from the National Library of Malaysia

ISBN 978-629-490-076-9

Table of Contents

Preface

Engineers have long valued statistics because they may apply them to understand the data gathered from field work, experiments, testing, assessments, research and other operations. Numerous aspects of engineering largely rely on statistics, including the use of statistical models for problem-solving and making predictions, statistical tests for testing hypotheses, and probability-based decision-making. *Statistics For Engineering Students* was specifically written with pre-university and university students in mind, and with the aim of providing a better understanding on statistics and how it may be used in the engineering world. This book explores common statistical tests and analysis that are presented in the simplest manner as possible to aid students' comprehension of the topics covered, particularly descriptive and inferential statistics. Examples and exercises dealing with parametric and non-parametric tests have been included in this book as well. This is not only to ensure students thoroughly understand the need for conducting statistical analysis, but are able to relate statistics with engineering, and apply them if required in their respective job scopes when they embark on a career in the engineering industry.

Basil David Daniel

CHAPTER 1

Data and Variables

1.1 Data

A set of data is a collection of observations, measurements or information obtained for a study. Data can be classified as either **quantitative data** or **qualitative data**.

Quantitative Data

- Data that can be measured numerically or counted.
- Can be continuous or discrete.
- e.g. length, time, mass, temperature

Qualitative Data

- Data that are not in numerical form but instead assigned as attributes.
- e.g. race, age, gender, marital status

Qualitative data is sometimes called **categorical data**, because the data is grouped into categories. Quantitative data can also be called **numerical data** because the data is expressed in numbers. Quantitative data can be categorised as **discrete data** or **continuous data**.

CHAPTER 2

Understanding Statistics

2.1 Descriptive and Inferential Statistics

Statistics deals with collecting, classifying, presenting, describing, analysing and interpreting data that allows us to make conclusions and reasonable decisions. Statistics can be grouped into two categories i.e., **descriptive statistics** and **inferential statistics**.

Descriptive Statistics

- Deals with collecting, classifying, presenting and describing quantitative data.
- Methods for organising (e.g. frequency table), representing (e.g. graphs) and summarising data (e.g. variability and central tendency).

Inferential Statistics

- Deals with techniques and methods of interpretation of results from the descriptive statistics.

Usually ,we begin with a statistical **population** to be studied when applying statistics to a industrial, social or scientific problem. When it is not feasible to collect data from a large population, a subset of the population, called a **sample**, is studied.

CHAPTER 3

Probability Distributions of Random Variables

3.1 Discrete and Continuous Random Variables

A random variable is a quantity resulting from an experiment that, by chance, can take on different values.

For instance, we flip two coins and count the number of heads. The result could be zero, one or two heads. Therefore, the number of heads occurring is the random variable. If X is the number of heads appearing, then $X = 0$, 1 or 2.

Discrete Random Variable	Continuous Random Variable
• Takes on only certain clearly separated values. • Measures discrete data (only integer values). • For example, number of concrete cube samples, number of accidents, etc.	• Takes on infinitely many values, with some limitations. • Measures continuous data (values include decimals and fractions). • For example, compressive strength of concrete, accident rates, etc.

3.2 Probability Distributions

A probability distribution is a mathematical function that explains the probability (or chance) of a set of values for a variable. Generally, graphs or probability tables are used to represent probability distributions.

CHAPTER 4

Special Probability Distributions

4.1 Binomial Probability Distribution

In a binomial probability distribution, there is only two possible outcomes on a particular trial of an experiment. For example, the statement in a true/false question is either true or false. It cannot be both true and false at the same time.

Characteristics of a Binomial Probability Distribution
• An outcome on each trial of an experiment is classified into one of two mutually exclusive categories, i.e. a success or a failure. • The random variable counts the number of successes in a fixed number of trials. • The probability of a success stays the same for each trial, so does the probability of failure. • The trials are independent, meaning that the outcome of one trial does not affect the outcome of any other trial.

To construct a particular binomial probability distribution, we use

- the number of trials, n
- the probability of success on each trial, p

Given that X is the random variable whereby X is the number of successes in n trials, X can be written as:

$X \sim B(n, p)$

The symbol $\sim$ (called "tilde") is used to indicate that X "is distributed as" or "follows the distribution of".

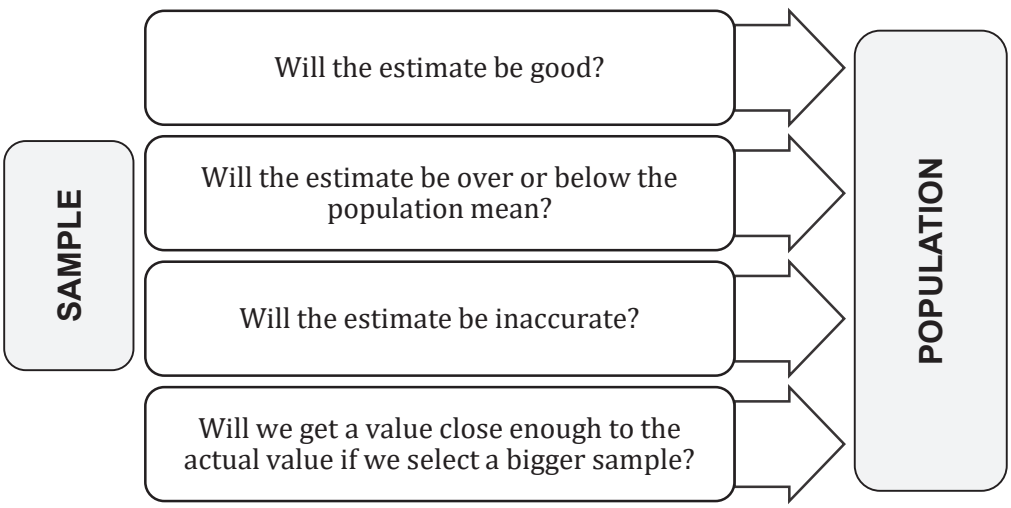
CHAPTER 5

Sampling and Estimation

5.1 Sampling

We have understood from Chapter 2 that a sample is a subset of a population and we study a sample to make conclusions about the population. If we are studying a large population, say to find the height of people in that population, it may be difficult to look at every individual. We therefore take a random sample. We might sample 100 people and take the average height of the sample.

Some questions we should consider when sampling:



In many cases, sampling is the only way to determine something about the population. Some of the reasons for sampling are:

- The physical impossibility of checking all items in the population
- To reduce the cost of studying the entire population
- To save time because contacting the entire population would often be time consuming

CHAPTER 6

Significance Testing

6.1 Testing a Hypothesis

A statistical **hypothesis** is a statement about a population parameter developed for the purpose of testing. In statistical analysis we make a claim, that is, state a hypothesis, then follow up with tests to verify the assertion or to determine that it is untrue.

Since populations are mostly so large that it is not feasible to study all the items or people in the population, we take a sample from the population. We can, therefore, test a statement to determine whether the sample does or does not support the statement concerning the population.

Hypothesis testing (or significance testing) is a procedure based on sample evidence and probability theory to determine whether the hypothesis is a reasonable statement.

5-Step Procedure For Testing A Hypothesis

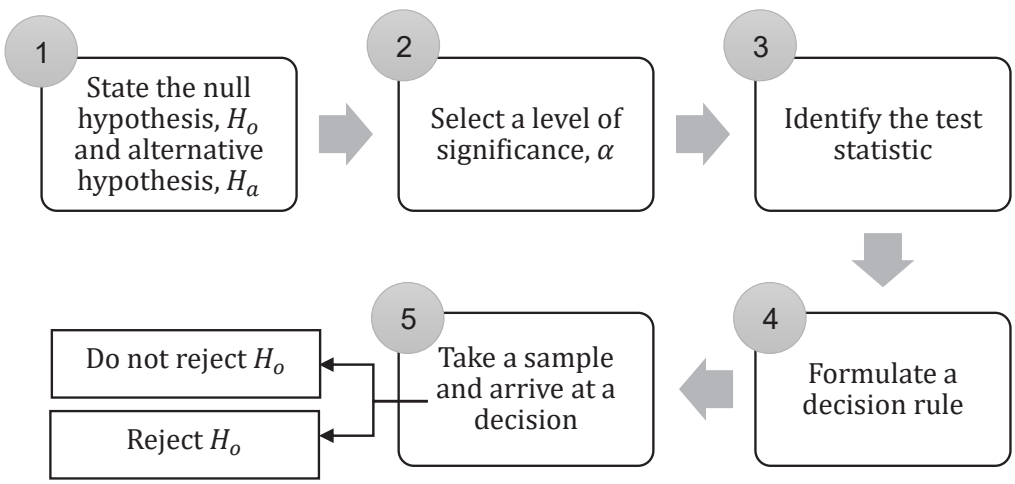


Figure 6.1: Significance testing procedure

CHAPTER 7

Correlation and Regression

7.1 Correlation Analysis

Correlation analysis is the study of the relationship between variables. The basic idea of correlation analysis is to report the strength of the relationship between two variables, i.e., the dependent variable and the independent variable.

- The **dependent variable** (Y) is the variable that is predicted or estimated. It is also called the response variable.
- The **independent variable** (X) is the variable that provides the basis for estimation. It is also known as the predictor or explanatory variable.

A basic way of studying correlation is to plot the data in a scatter diagram or scatter plot, which is a chart that portrays the relationship between two variables. There are several statistical tests that we can use to assess the correlation between variables. The most common ones are using Pearson's correlation coefficient.

The strength of the relationship between the dependent and independent variables is expressed using the coefficient of correlation.

7.2 Coefficient Of Correlation

Introduced by Karl Pearson, the coefficient of correlation denoted by the symbol r, is used to describe the strength of the relationship between two sets of variables. This coefficient is also referred to as Pearson's correlation or Pearson's r. It can take any value from –1 to +1.

If the value of one variable increases linearly with an increase in the other variable, it is a positive correlation ($r > 0$). If the value of one variable decreases with an increase in the other variable, it is a negative correlation ($r < 0$).

CHAPTER 8

Analysis of Variance

8.1 ANOVA Test

Analysis of variance (ANOVA) is a hypothesis testing technique used to test the equality of two or more population means by examining the variances of samples that are taken.

It is a parametric test. Parametric tests are those that make assumptions about the parameters of the population distribution from which the sample is drawn. It is used to determine whether:

- the differences between the samples are simply due to random error.
- there are systematic treatment effects that causes the mean in one group to differ from the mean in another.

This is achieved by calculating the *F*-ratio.

ANOVA is based on comparing the variance (or variation) between the data samples to variation within each particular sample.

- If the "between" variation is much larger than the "within" variation, the means of different samples will not be equal.
- If the "between" and "within" variations are approximately the same size, then there will be no significant difference between sample means.

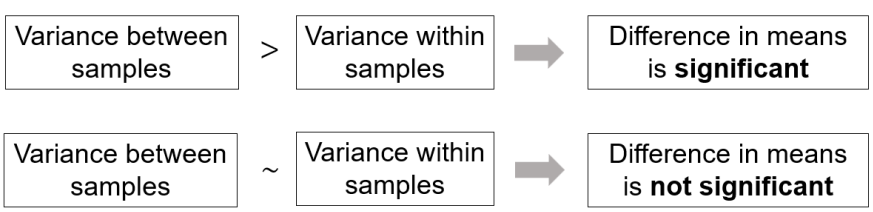


Figure 8.1: Significance of differences in variance

CHAPTER 9

Chi-Squared Test

9.1 Chi-Squared Distribution

The chi-squared, χ^2 distribution is a continuous probability distribution that is widely used in statistical inference. It is related to the standard normal distribution. If a random variable Z has the standard normal distribution, then Z^2 has the C distribution with one degree of freedom.

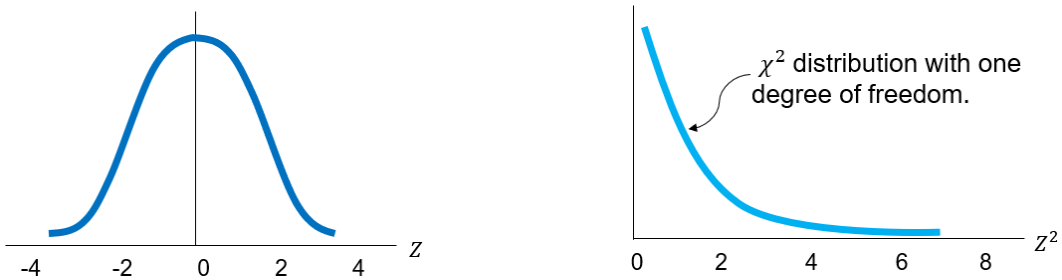


Figure 9.1: Deriving the c² Distribution

If $Z_1, Z_2, \ldots, Z_k$ are independent standard normal variables, then ${Z_1}^2 + {Z_2}^2 + \ldots + {Z_k}^2$ has a χ^2 distribution with k degrees of freedom.

The probability distribution function (PDF) of the χ^2 distribution with degrees of freedom is given by:

$$f(x) = \frac{x^{\frac{k}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{k}{2}} \cdot \Gamma\left(\frac{k}{2}\right)} \quad for\ x \geq 0$$

The mean of the χ^2 distribution, $\mu = k$ while the variance, $\sigma^2 = 2k$. The mode of the χ^2 distribution is equal to $k - 2$.

CHAPTER 10

Analysis of Ranked Data

10.1 The Sign Test

The sign test is based on the sign of a difference between two related observations. We usually designate a plus (+) sign for a positive difference and a minus (-) sign for a negative difference.

The sign test has many applications, one is for before/after experiments. For example, a tune-up program for cars is evaluated whereby the mileage per gallon of fuel before and after the tune-up are recorded. A "+" sign is assigned to an increase in mileage, and a "-" sign to a decrease in mileage.

A hypothesis testing is conducted to see if the intervention is effective. The procedure is explained in the following example.

Example 10.1

15 engineers were randomly selected to assess their level of competence in using software for design and analysis. The engineers underwent a software training program and were rated Outstanding, Excellent, Good, Fair or Poor before and after the training. Can it be concluded that the engineers were more competent after the training?

Engineer	Before	After
Adnan	Good	Outstanding
Felix	Fair	Excellent
Firdaus	Excellent	Good
Khaidir	Poor	Good
Adib	Excellent	Excellent

Bibliography

Berenson, M. L., Levine, D.M., Szabat, K.A. and Stephan, D.F. (2020). *Basic Business Statistics: Concepts and Applications, 14th Edition*. Pearson.

Black, K., Asafu-Adjaye, J., Burke, P., Perera, N., Sherwood, C. and Wasimi, S. (2021). *Business Analytics and Statistics, 1st Edition*. Wiley.

Blatchley, B. (2018). *Statistics in Context*. Oxford University Press.

Bruce, N., Pope, D. and Stanistreet, D. (2008). *Probability Distributions, Hypothesis Testing and Bayesian Methods*. Wiley.

Chernick, M.R. and Friis, R.H. (2003). *Introductory Biostatistics for the Health Sciences*. Wiley.

Fenton, G.A. and Griffiths, D.V. (2008). *Risk Assessment in Geotechnical Engineering*. Wiley.

Hubbard, M.R. (2003). *Statistical Quality Control for the Food Industry*. Springer Link.

Illowsky, B. and Dean, S. (2018). *Introductory Statistics*. OpenStax.

Kenett, R.S., Zacks, S. and Amberti, D. (2013). *Modern Industrial Statistics*. Wiley.

Lewis-Beck, M. S., Bryman, A. and Liao, T. F. (2004). *The SAGE Encyclopedia of Social Science Research Methods*. Sage Publications, Inc.

Lind, D. Marchal, W. and Wathen, S. (2011). *Statistical Techniques in Business and Economics, 15th Edition*. McGraw Hill/Irwin.

Lind, D. Marchal, W. and Wathen, S. (2012). *Basic Statistics for Business and Economics, 8th Edition*. McGraw-Hill.

Maity, R. (2022). *Statistical Methods in Hydrology and Hydrocilmatology*. Springer Science and Business Media LLC.

May, G.S. and Spanos, C.J. (2006). *Fundamentals of Semiconductor Manufacturing and Process Control*. Wiley

Mendenhall, W. and Sincich, T. (2011). *Second Course in Statistics: Regression Analysis*. Pearson.

Mendenhall, W., Beaver, R.J. and Beaver, B.M. (2012). *Introduction to Probability and Statistics*. Cengage Learning.

Mendes, E. (2008). Understanding Project Data in *Cost Estimation Techniques for Web Projects*. IGI Global. pp. 303-374.

Miah, A.Q. (2016). *Applied Statistics for Social and Management Sciences*. Springer Link.

Mitra, A. (2008). *Fundamentals of Quality Control and Improvement*. Wiley.

Moore, D., Notz, W.I. and Fligner, M.A. (2011). *The Basic Practice of Statistics, 6th Edition*. W.H. Freeman and Company.

Rohatgi, V.K. and Ehsanes Saleh, A.K.MD. (2000). *An Introduction to Probability and Statistics*. Wiley.

Schiller, J., Srinivasan, R. and Spiegel, M. (2012). *Schaum's Outline of Probability and Statistics, 4th Edition*. McGraw Hill.

Simplilearn (2023). *Descriptive vs. Inferential Statistics: Key Differences and Measurement Techniques*. Retrieved on July 13, 2023 from https://www.simplilearn.com/difference-between-descriptive-inferential-statistics-article.

Thonemann, U.W. and Brandeau, M.L. (2000). Optimal Commonality in Component Design. *Operations Research 48(1)*, pp 1-19.

TRB (2016). *Highway Capacity Manual*. Washington, D.C.: Transportation Research Board.

Utts, J.M. and Heckard, R.F. (2011). *Mind on Statistics*, *4th Edition*. Cengage Learning.

Van Belle, G., Fisher, L.D., Heagerty, P.J. and Lumley, T. (2004). *Biostatistics*. Wiley.

Walpole, R.E., Myers, R.H., Myers, S.L. and Ye, K.E. (2011). *Probability and Statistics for Engineers and Scientists, 9th Edition*. Pearson.

Weathington, B.L., Cunningham, C.J.L. and Pittenger, D.J. (2012). *The Foundations of Research*. Wiley.

Weiss, N. (2015). *Elementary Statistics, 9th Edition*. Pearson.

Author Biography

Basil David Daniel is a senior lecturer at the Faculty of Civil Engineering and Built Environment, Universiti Tun Hussein Onn Malaysia, specialising in Transportation Engineering. With 20 years experience in the academia, he obtained a Doctor of Philosophy from the University of Canterbury, New Zealand in 2013 and a Master of Science from Universiti Putra Malaysia in 2003. Both his PhD research and Master's dissertation dealt with various statistical techniques applied in his area of expertise. He has been teaching Statistics for Civil Engineering over the last 10 years and hopes graduates develop a deep appreciation for Statistics, which may come in handy in when they embark on a professional career.

Index

STATISTICS
for Engineering Students

Engineers have long valued statistics because they may apply them to understand the data gathered from field work, experiments, testing, assessments, research and other operations. Numerous aspects of engineering largely rely on statistics, including the use of statistical models for problem-solving and making predictions, statistical tests for testing hypotheses, and probability-based decision-making. Statistics For Engineering Students was specifically written with pre-university and university students in mind, and with the aim of providing a better understanding on statistics and how it may be used in the engineering world. This book explores common statistical tests and analysis that are presented in the simplest manner as possible to aid students' comprehension of the topics covered, particularly descriptive and inferential statistics. Examples and exercises dealing with parametric and non-parametric tests have been included in this book as well. This is not only to ensure students thoroughly understand the need for conducting statistical analysis, but are able to relate statistics with engineering, and apply them if required in their respective job scopes when they embark on a career in the engineering industry.

Penerbit UTHM

ISBN 978-629-490-076-9

9 786294 900769 04000

For more information, please scan the code