DR. SITI ZANARIAH BINTI SATARI

Siti Zanariah Satari has been in the academic profession for more than ten years. She has experience in teaching, developing syllabus, and module writing for various mathematics subjects such as elementary statistics, applied statistics, fundamental discrete structure, discrete structure & applications, basic mathematics, and calculus for diploma and degree level students in Universiti Malaysia Pahang. She has a PhD in Statistics from Universiti of Malaya, and Master's degree in Mathematics (Statistics) from Universiti Teknologi Malaysia (UTM), and Bachelor's degree in Industrial Mathematics from Universiti Teknologi Malaysia (UTM). Her general interest is in Statistics and Data Analysis. Her main contributions have been in the areas of (i) circular statistics, (ii) circular regression model, (iii) functional relationship model for circular data, (iv) outlier detection for circular regression model and multiple linear regression model, (v) clustering algorithm for detecting multiple outliers, (vi) multiple regression method, (vii) wind direction, and (viii) rain modelling. She also involved in research collaboration with different fields of studies such as social science and computing, as data analyst, mathematical modeler, and co supervisor. She has published three books under Penerbit UMP (UMP Publisher). She is the main author for Applied Statistics Module (2015) and Calculus for Science & Engineering (2009), and contributed as an author in Research Methodology for Social Science (2007). Currently she teaches Applied Statistics and Mathematics courses at the Faculty of Industrial Sciences & Technology, FIST, UMP.

INSTITUTE OF POSTGRADUATE STUDIES
UNIVERSITI MALAYSIA PAHANG
Lebuhraya Tun Razak,
26300 Gambang, Kuantan,
Pahang, Malaysia.
Tel: +609-549 3197/3198/3199
Fax: +609-549 3190
Email: ips@ump.edu my
Website: ips.ump.edu.my

ISBN 978-967-2054-05-4

PUBLISHER
UNIVERSITI MALAYSIA PAHANG

DATA COLLECTION AND ANALYSIS (ENGINEERING)

SITI ZANARIAH SATARI

FOR INTERNAL CIRCULATION ONLY

DATA COLLECTION AND ANALYSIS (ENGINEERING)

DATA COLLECTION AND ANALYSIS (ENGINEERING)

SITI ZANARIAH SATARI

Penerbit
Universiti Malaysia Pahang
Kuantan
2017

Copyright © Universiti Malaysia Pahang, 2017

First Published, 2017

All right reserved.
Apart from fair dealing for the purpose of study, research,
criticism or review, as permitted under the Copyright Act, no part of this
book may be reproduced, strored in a retrieval system, or transmited, in any form or
by any means, electronic, mechanical, photocopying, recording or otherwise
without the prior written permission of the publisher. Enquiries to be made
to the author and the publisher Penerbit Universiti Malaysia Pahang,
Lebuhraya Tun Razak, 26300 Gambang, Kuantan, Pahang Darul Makmur.
Negotiation is subject to royalty arrangement or honorarium.

Perpustakaan Negara Malaysia Cataloguing-in-Publication Data
Siti Zanariah Satari
DATA COLLECTION AND ANALYSIS : (ENGINEERING)
REV.4 / Siti Zanariah Satari
ISBN 978-967-2054-05-4
1. Engineering. I. Title
620

Published By:
Publisher
Universiti Malaysia Pahang
Lebuhraya Tun Razak, 26300 Gambang
Kuantan, Pahang Darul Makmur
Tel: 09-549 3273 Fax: 09-549 3281

Printing:
Syarikat Percetakan Inderapura Sdn. Bhd
Jalan Tanjong Api Off Jalan Telok Sisek
25200 Kuantan, Pahang Darul Makmur
Tel: 09-5177225/5177031 Fax: 095139434

PREFACE

The rebranding of 8 modules has been approved by the UMP Senate in February 2016. The 8 modules comprise:

i. Philosophy of Science and Ethics in Research
ii. Research Overview
iii. Literature Review
iv. Methodology
v. Data Collection and Analysis
vi. Scientific Writing
vii. Preparation for viva-voce
viii. Conceptual Paper

8 modules are one of the Institute of Postgraduate Studies (IPS) initiatives in helping students to understand ways and means to carry out research at Master's and PhD level. There are 2 sets of 8 modules provided by IPS in accordance with the technology and engineering cluster. These 8 modules are intended to help students in conducting systematic research owing to its utmost importance in determining the effectiveness and efficiency of the research process. The 8 modules shall help students to understand the research problem and identify the areas of research. The modules also assist students on how to write a literature review in order to understand how other researchers approach, define or manage the problem to keep the research pertinent to what is current in the field. In addition, 8 modules also facilitate students to understand the methods of collecting data in an organized and controlled way to achieve valid results, analyze the data according to the considered problem and make conclusions. Therefore, IPS hopes that these modules will be able to provide some insights and benefit students in conducting research at Master's and PhD level.

INSTITUTE OF POSTGRADUATE STUDIES
UNIVERSITI MALAYSIA PAHANG

TABLE OF CONTENTS

Module Description vii
Module Outcome vii
Module Subtopic vii

Topic 1 : Statistical Problem-Solving Methodology 1
1.1 Overview 1
1.2 Identify The Problem or Opportunity 1
1.3 Data Collection 3
1.4 Sampling Techniques 4
1.5 Data Classification 9
1.6 Presenting and Analysing The Data 11
1.7 Decision Making 11

Topic 2 : Summary Statistics Using *Microsoft Excel* 13
2.1 Role of Computer in Statistics 13
2.2 Data Organization 14
2.3 Summarising Quantitative Data 16
2.3.1 Descriptive Statistics 16
2.3.2 Accuracy and Precision (Validity and Reliability) 20
2.3.3 Graphical Summary 21
2.4 Summarising Qualitative Data 32
2.5 Normality Assumption and Normal Probability Plot 33

Topic 3 : Statistical Analysis for One Population Using *Microsoft Excel* 35
3.1 Hypothesis Testing and its Procedure 35
3.2 Statistical Analysis for One Population Mean 39
3.3 Statistical Analysis for One Population Variance 42
3.4 Statistical Analysis for One Population Proportion 45

Topic 4 : Statistical Analysis for Two Populations Using *Microsoft Excel* 49
4.1 Statistical Analysis for Two Population Variances 49
4.2 Statistical Analysis for Two Population Means 52
4.2.1 Independent Samples 52
4.2.2 Dependent Samples 56
4.3 Statistical Analysis for Two Population Proportions 58

Topic 5 : Statistical Analysis for Three or More Population Means Using *Microsoft Excel* 63
5.1 Introduction to Design of Experiments and Analysis of Variance 63
5.2 One-Way Analysis of Variance 64
5.2.1 The Scheffe Test and Tukey Test 68
5.3 Two-Way Analysis of Variance with Replications 71

Topic 6 : Statistical Analysis for Categorical Data Using *Microsoft Excel* 79
6.1 Goodness of Fit Test for Categorical Data 79
6.2 Contingency Table for Categorical Data 82
6.2.1 Independent Test 82
6.2.2 Test of Homogeneity Proportions 84

Topic 7 : Linear Regression and Correlation Using *Microsoft Excel* 87
7.1 Scatter Plot and Correlation 87
7.2 Simple Linear Regression Analysis 90
7.3 Multiple Linear Regression Analysis and Correlation 95
7.4 Model Selection 100

Conclusion 109
References 111
Further Readings 111

MODULE DESCRIPTION

This module is specially tailored to assist postgraduate students of science, engineering, and technology in data collection and analysis. The contents in this module are primarily focused on the applications of various statistical techniques for quantitative and categorical data. Through this course, students are able to learn and directly apply the statistical knowledge in the research or problems being studied. This module provides examples and exercises to show connections between theory and application in scientific research. The material in this module also does integrate well with computer software packages, especially *Microsoft Excel*.

There are seven chapters in this module. Topic 1 covers statistical problem-solving methodology. Topic 2 discusses on summarising quantitative and qualitative data using descriptive and graphical summary. Topic 3 presents the terms and procedure of hypothesis testing for one population parameter. The parameters involve are population mean, population variance, and population proportion. Topic 4 is on statistical analysis for two populations' parameters that are discussed in Topic 3.

Topic 5 covers some of the commonly used methods of experimental designs that are a one-way and two-way analysis of variance which is covered extensively via computer software. Topic 6 discusses the concepts of the goodness of fit test and contingency table in dealing with categorical or frequency data. Finally, Topic 7 covers linear regression and correlation, for both single and multiple cases.

We hope that postgraduate students will be able to fully utilise this Data Collection and Analysis Module in assisting them in the process of research and thesis writing.

MODULE OUTCOME

By the end of this module, you should be able to:

1. Acquire fundamental principle of statistics.
2. Organise real life data to solve related problems in various disciplines using appropriate statistical methodology.
3. Conduct statistical analysis using appropriate software tools.
4. Recommend a conclusion or suggestion based on the fundamental principle of applied and advanced statistical techniques.

MODULE SUBTOPIC

- ✓ Topic 1 Statistical Problem Solving Methodology
- ✓ Topic 2 Summary Statistics Using *Microsoft Excel*
- ✓ Topic 3 Statistical Analysis for One Population using *Microsoft Excel*
- ✓ Topic 4 Statistical Analysis for Two Populations using *Microsoft Excel*
- ✓ Topic 5 Statistical Analysis for Three or More Population Means using *Microsoft Excel*
- ✓ Topic 6 Statistical Analysis for Categorical Data using *Microsoft Excel*
- ✓ Topic 7 Linear Regression and Correlation using *Microsoft Excel*
- ✓ Conclusion
- ✓ References and Further Readings