STATISTICAL ANALYSIS OF FACTORS AFFECTING MONOCLONAL ANTIBODY PRODUCTION BY USING PRINCIPAL COMPONENT ANALYSIS : PHYSIOLOGICAL CHARACTERISTICS OF CELL LINE

WONG SHIN YIE

UNIVERSITI MALAYSIA PAHANG

SUPERVISOR'S DECLARATION

I hereby declare that I have checked this thesis and in my opinion, this thesis is adequate in terms of scope and quality for the award of the degree of Bachelor in Manufacturing Engineering Technology (Pharmaceutical) with Hons.

_____

(Supervisor's Signature)

Full Name : DR RAIHANA ZAHIRAH BINTI EDROS

Position : HEAD OF PROGRAMME

Date : 11 DECEMBER 2017

STUDENT'S DECLARATION

I hereby declare that the work in this thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at Universiti Malaysia Pahang or any other institutions.

_____

(Student's Signature)

Full Name : WONG SHIN YIE

ID Number : TD14015

Date : 11 DECEMBER 2017

STATISTICAL ANALYSIS OF FACTORS AFFECTING MONOCLONAL ANTIBODY PRODUCTION BY USING PRINCIPAL COMPONENT ANALYSIS : PHYSIOLOGICAL CHARACTERISTICS OF CELL LINE

WONG SHIN YIE

Thesis submitted in fulfillment of the requirements
for the award of the degree of
Bachelor of Manufacturing Engineering Technology (Pharmaceutical)

Faculty of Engineering Technology

UNIVERSITI MALAYSIA PAHANG

January 2018

ACKNOWLEDGEMENTS

Foremost, I would like to express my sincere gratitude to my advisor Dr Raihana Zahirah Edros for the continuous support throughout my research; for her patience, motivation, and immense knowledge. I would not be able to complete this research and thesis without her guidance.

Besides, I would like to thank my teammates, Wong Li Ung and Gan Zun Jiat, for their encouragement, tolerance, and help throughout this research. Their guidance, invaluably constructive criticism and friendly advice are highly appreciated. It is my pleasure to have them sharing their truthful and instructive opinions on a number of issues related to the research.

Finally, my highest appreciation to my family members; my parents, Wong Kok Kooi and Lim Soo Har and my brother Wong Lean Chun whose love and guidance are with me in whatever I pursue. I am grateful for their spiritual support throughout the completion of this research and thesis.

TABLE OF CONTENT

DECLARATION

TITLE PAGE

ACKNOWLEDGEMENTS **ii**

ABSTRAK **iii**

ABSTRACT **iv**

TABLE OF CONTENT **v**

LIST OF TABLES **ix**

LIST OF FIGURES **x**

LIST OF SYMBOLS **xi**

LIST OF ABBREVIATIONS **xii**

CHAPTER 1 INTRODUCTION **2**

1.1 Background 2

1.2 Problem Statement 4

1.3 Research Objectives 5

1.4 Scope of the Study 5

CHAPTER 2 LITERATURE REVIEW **7**

2.1 Antibody 7

2.1.1 Structure of Antibody 8

2.1.1.1 Heavy Chain 9

2.1.1.2 Light Chain 9

2.1.2 Classes of Antibody 10

2.2 Mammalian Cell Culture 12

2.2.1 Chinese Hamster Ovary (CHO) Cells 12

2.3 Physiological Characteristics Affecting Productivity of CHO Cell Lines 13

2.4 Data Mining 15

2.5 Multivariate Analysis 16

2.6 Principal Component Analysis 17

2.6.1 Data Preprocessing 19

2.6.1.1 Methods of data Preprocessing 20

2.6.2 Principal Components 22

2.6.2.1 Mathematical Model of Principal Components 22

2.6.2.2 Limit and Significance of Principal Components 23

2.6.3 Cross-Validation 24

2.7 Partial Least Square 25

CHAPTER 3 METHODOLOGY **28**

3.1 Data Acquisition 28

3.2 Data Preprocessing 28

3.3 Partial Least Square (PLS) 29

3.4 Principal Component Analysis (PCA) 29

CHAPTER 4 RESULTS AND DISCUSSION **31**

4.1 Data Acquisition 31

4.1.1 Growth, Viability and Antibody Concentration Profile 31

4.1.2 Summary of Data 34

4.2 Data Preprocessing 36

4.2.1 Arrangement of Data 36

4.2.2 Standardization of Data 36

4.3 Partial Least Square (PLS) 37

4.4 Principal Component Analysis (PCA) 42

4.4.1 Arrangement of Data by All the Six Cell Lines for Five Consecutive Days 42

4.4.2 Arrangement of Data by Individual Cell Line for Five Consecutive Days 46

4.4.2.1 Scatter Plot for Data Arranged by Individual Cell Line for Five Consecutive Days 46

4.4.2.2 Standardized Biplot and Loading Scatter Plot for Data Arranged by Individual Cell Line for Five Consecutive Days 49

4.4.3 Arrangement of Data by Individual Day for All the Six Cell Lines 50

4.4.3.1 Scatter Plot for Data Arranged by Individual Day for All the Six Cell Lines 50

4.4.3.2 Standardized Biplot and Loading Scatter Plot for Data Arranged by Individual Day for All the Six Cell Lines 55

CHAPTER 5 CONCLUSION **58**

REFERENCES **59**

APPENDIX A RAW DATA OBTAINED FROM Edros et. al. 2013 **65**

APPENDIX B PREPROCESSED DATA **67**

APPENDIX C STANDARDIZED DATA **68**

APPENDIX D SCATTER PLOT FOR DATA ARRANGED BY INDIVIDUAL CELL LINE FOR FIVE CONSECUTIVE DAYS **69**

APPENDIX E SCATTER PLOT FOR DATA ARRANGED BY INDIVIDUAL DAY FOR ALL THE SIX CELL LINES **70**

APPENDIX F STANDARDIZED BIPLOT AND LOADING SCATTER PLOT FOR DATA ARRANGED BY INDIVIDUAL DAY FOR ALL THE SIX CELL LINES **71**

LIST OF TABLES

Table 4.1	Growth characteristics and cell productivity of GS-CHO cell lines	33
Table 4.2	Preprocessed data of the six GS-CHO cell lines	36
Table 4.3	Standardized data using STATISTICA	37

LIST OF FIGURES

Figure 2.1	Structure of human antibody molecule	8
Figure 2.2	Structure of different classes of antibody	10
Figure 2.3	Schematic diagram for four human IgG subclasses	11
Figure 2.4	Data mean centering in two dimensions	20
Figure 2.5	Algebraic representation of a PC model	23
Figure 2.6	A graphical representation of X block and Y block	26
Figure 3.1	Summary of research	30
Figure 4.1	Growth and viability curve of GS-CHO cell lines	31
Figure 4.2	Antibody concentration profile of GS-CHO cell lines	32
Figure 4.3	Data distribution collected throughout exponential phase of triplicate batch cultures in the six GS-CHO cell lines	35
Figure 4.4	PLS scatter plot for day 1 (A), day 2 (B) and day 3 (C) respectively	39
Figure 4.5	PLS scatter plot for cell line 38 (A), cell line 76 (B) and cell line 150 (C) respectively	42
Figure 4.6	Scatter plot for PCA analysis on all the six cell lines for five consecutive days	43
Figure 4.7	Standardized biplot (A) and loading scatter plot (B) generated from PCA analysis on all the six cell lines for five consecutive days	44
Figure 4.8	Scatter plot for PCA analysis on cell line 38 (A), cell line 47 (B), cell line 160 (C) and cell line 164 (D) for five consecutive days	47
Figure 4.9	Standardized biplot and loading scatter plot for cell line 38 (A), cell line 47 (B), cell line 76 (C), cell line 150 (D), cell line 160 (E) and cell line 164 (F) respectively	51
Figure 4.10	Scatter plot for PCA analysis on day 1 (A), day 2 (B), day 4 (C) and day 5 (D) for all the six cell lines.	53
Figure 4.11	Standardized biplot and loading scatter plot for day 1 (A), day 2 (B), day 3 (C) and day 5 (D) respectively	56

LIST OF SYMBOLS

q_p	Specific productivity
α	Alpha
β	Beta
γ	Gamma
κ	Kappa
λ	Lambda
%	Percentage
ºC	Degree Celcius
μ_{net}	Specific growth
t_1	Principal component 1
t_2	Principal component 2

LIST OF ABBREVIATIONS

µM	Micromolar
ANOVA	Analysis of variance
BHK	Baby hamster kidney
cells/ml	Cells per millilitre
CHO	Chinese Hamster Ovary
ekf	Element wise *k* fold method
ELISA	Enzyme-linked immunosorbent assay
ER	Endoplasmic reticulum
g/L/cell	Gram per litre per cell
GS-CHO	Glutamine synthetase Chinese Hamster Ovary
HEK	Human embryonic kidney
IgA	Immunoglobulin A
IgD	Immunoglobulin D
IgE	Immunoglobulin E
IgG	Immunoglobulin G
IgG1	Immunoglobulin G 1
IgG2	Immunoglobulin G 2
IgG3	Immunoglobulin G 3
IgG4	Immunoglobulin G 4
IgM	Immunoglobulin M
KDD	Knowledge Discovery in Database
LV	Latent variable
mAb	Monoclonal antibody
ml	millilitre
MSX	Methionine sulphoxamine
MVS	Multivariate statistics
NSO	Mouse myeloma
PC	Principal component
PCA	Principal Component Analysis
pg/cell/day	Pictogram per cell per day
PLS	Partial Least Square

rkf	Row wise *k* fold method
rpm	Revolutions per minute