

**STATISTICAL ANALYSIS OF FACTORS
AFFECTING MONOCLONAL ANTIBODY
PRODUCTION BY USING PRINCIPAL
COMPONENT ANALYSIS: MOLECULAR
MARKERS**

GAN ZUN JIAT

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 of Manufacturing Engineering Technology (Pharmaceutical) with Hons.

(Supervisor's Signature)

Full Name : Raihana Zahirah Binti Edros

Position : Senior Lecturer

Date : 10 January 2018



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 : GAN ZUN JIAT

ID Number : TD 14012

Date : 10 January 2018

STATISTICAL ANALYSIS OF FACTORS AFFECTING MONOELONAL
ANTIBODY PRODUCTION BY USING PRINCIPAL COMPONENT ANALYSIS:
MOLECULAR MARKERS

GAN ZUN JIAT

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 supervisor Dr. Raihana Zahirah Edros for the continuous support of my Degree study and research, for her patience, motivation, enthusiasm, and immense knowledge. Her guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my Degree study.

Next, I would like to thank my family: my parents Gan Teah Ann and Mok Chui Lai for giving birth to me at the first place and supporting me spiritually throughout my life. I would like to thank my siblings also, Gan Zun Xiang and Gan Zun Han, for the company throughout my life

Besides my supervisor, I would like to thank the rest of my colleagues, Wong Li Ung and Wong Shin Yir, for their encouragement, insightful comments, and hard questions.

Other than that, I would like to appreciate my close friends,

Last but not the least, I would like to appreciate my close friends, they have been giving me support all the time.

TABLE OF CONTENT

DECLARATION	
TITLE PAGE	
ACKNOWLEDGEMENTS	ii
ABSTRAK	iii
ABSTRACT	iv
TABLE OF CONTENT	v
LIST OF TABLES	viii
LIST OF FIGURES	ix
LIST OF SYMBOLS	x
LIST OF ABBREVIATIONS	xi
CHAPTER 1 INTRODUCTION	1
1.1 Background Study	1
1.2 Problem Statement	3
1.3 Objective	4
1.4 Scope of Study	4
CHAPTER 2 LITERATURE REVIEW	6
2.1 Monoclonal Antibodies (IgG1)	6
2.2 Monoclonal Antibody Production	7
2.2.1 Expression System	7
2.2.2 Chinese Hamster Ovary (CHO)	9
2.2.3 Glutamine Synthetase CHO (GS-CHO) Cell Line	10

2.3	Molecular Markers	11
2.3.1	Endoplasmic Reticulum (ER)	12
2.3.2	Intracellular Heavy Chain (HC) and Light Chain (LC) Polypeptides	13
2.3.3	Golgi Apparatus	13
2.4	Principal Component Analysis (PCA)	14
2.4.1	Eigenvectors and Eigenvalues	15
2.4.2	Preprocessing	17
2.4.3	Data Standardization	17
2.4.4	Mean Centering and Unit Variance	18
2.4.5	The Mathematical Model	19
2.4.6	Limit and Significance of a Principal Component	20
2.4.7	The Method of Cross Validation	21
2.5	Partial Least Square (PLS)	21
CHAPTER 3 METHODOLOGY		23
3.1	Introduction	23
3.2	Data Acquisition	24
3.3	Data Selection	24
3.4	Data Treatment	24
3.5	Partial Least Square (PLS)	25
3.6	Removing Outliers	26
3.7	Principal Component Analysis (PCA)	26
CHAPTER 4 RESULTS AND DISCUSSION		27
4.1	Data Acquisition	27

4.1.1	Growth, Viability and Specific Productivity	27
4.1.2	Summary of Data	29
4.2	Data Preprocessing	32
4.3	PLS Analysis	33
4.4	PCA Analysis	35
4.4.1	PCA for All Cell Lines in Five Consecutive Days	35
4.4.2	PCA for Each Cell Line in Five Consecutive Days	37
4.4.3	PCA for All Cell Lines in Five Respective Days	41
CHAPTER 5 CONCLUSION		44
REFERENCES		45
APPENDIX A		50
APPENDIX B		49
APPENDIX C		56
APPENDIX D		59
APPENDIX E		63

LIST OF TABLES

Table 4.1	Growth characteristics and cell productivity of GS-CHO cell lines determined from day 1 to day 5 of batch culture	29
Table 4.2	Data arranged by cell line throughout the days. The shaded data are the out-of-range data which are going to be removed from the data set.	32
Table 4.3	Standardized data for intracellular HC and intracellular LC.	33

LIST OF FIGURES

Figure 1.1	Alpha backbone structure of human IgG showing functional regions; light chain (blue) and heavy chain (red).	2
Figure 2.1	Structure of immunoglobulin molecule	7
Figure 2.2	Secretory pathway of protein synthesis	11
Figure 2.3	Structure endoplasmic reticulum which consists of SER and RER .	12
Figure 2.4	The three dimensional structure of Golgi apparatus	14
Figure 2.5	Determination of the principal component. (A) An example set of data. (B) A vertical straight line is drawn on the plane of data. (C) A horizontal line is drawn on the plane of the data.	15
Figure 2.6	An example to determine eigenvectors. (A) A two dimensional set of data. (B) Principal component of the data set. (C) Second eigenvector of the data set. (D) New dimension of the data.	16
Figure 2.7	Two dimensional data mean centering. The mean of the multivariate data becomes the point of origin for the mean centered variables.	19
Figure 2.8	An algebraic representation of a PC model.	20
Figure 3.1	The flow chart of the methodology in this study.	23
Figure 4.1	Growth and viability curves of GS-CHO cell lines	28
Figure 4.2	Antibody concentration profiles of GS-CHO cell lines	28
Figure 4.3	Data distribution collected throughout exponential phase of triplicate batch cultures in the six GS-CHO cell lines.	31
Figure 4.4	PLS plot with outliers detected in Day 4 for all six cell lines.	34
Figure 4.5	PLS plot with outliers recoded in Day 4 for all six cell lines.	34
Figure 4.6	PCA scatter plot for all six cell lines in five consecutive days.	36
Figure 4.7	PCA biplot for all six cell lines in five consecutive days.	36
Figure 4.8	PCA loading scatterplot for all six cell lines in five consecutive days.	37
Figure 4.9	PCA score scatterplot (A); Standardized biplot (B); Loading scatterplot (C) of cell line 47 in five consecutive days.	39
Figure 4.10	PCA score scatterplot (A); Standardized biplot (B); Loading scatterplot (C) of cell line 76 in five consecutive days.	40
Figure 4.11	PCA score scatterplot (A); Standardized biplot (B); Loading scatterplot (C) of all cell lines on day 2.	42
Figure 4.12	PCA score scatterplot (A); Standardized biplot (B); Loading scatterplot (C) of all cell lines on day 4.	43

LIST OF SYMBOLS

pg	Pictogram
mL	Millilitre
°C	Degree Celsius
rpm	Revolutions Per Minute
q_p	Specific Productivity
μM	Micro Molar
%	Percentage

LIST OF ABBREVIATIONS

CHO	Chinese Hamster Ovary
HEK	Human Embryonic Kidney
BHK	Baby Hamster Kidney
tPA	Tissue Plaminogen Activator
EPO	Erythropoietin
PCA	Principal Component Analysis
PLS	Partial Least Square
mAb	Monoclonal Antibody
DHFR	Dihydrofolate Reductase
GS	Glutamine Synthetase
ER	Endoplasmic Reticulum
HC	Heavy Chain
LC	Light Chain
ATP	Adenosine Triphosphate
ADP	Adenosine Diphosphate
MSX	Methionine Sulphoximine
CL	Cell Line
IQR	Interquartile Range
ANOVA	Analysis Of Variance