UNIVERSITI MALAYA

Institute of Mathematical Sciences

PROGRAMME HANDBOOK

MASTER OF SCIENCE IN STATISTICS BY COURSEWORK

SESSION 2021/2022

https://ism.um.edu.my/postgraduate-info

Master of Science in Statistics Session 2021/2022 (42 CREDITS)		
1. Programme Core Courses (26 CREDITS)		
Course Code	**Course Name**	**Credits**
SQB7001	Research Methodology for Statistics	3
SQB7002	Research Project for Statistics	10
SQB7003	Statistical Inference	4
SQB7004	Probability Theory	4
SQB7005	Statistical Laboratory	2
SQB7006	Statistical Consultancy and Data Analysis	3
2. Programme Elective Courses (16 CREDITS)		
Course Code	**Course Name**	**Credits**
SQB7007	Multivariate Analysis	4
SQB7008	Stochastic Models	4
SQB7009	Bayesian Statistics	4
SQB7010	Decision Statistics	4
SQB7011	Generalized Linear Models	4
SQB7012	Experimental Design and Quality Engineering	4
SQB7013	Statistical Time Series	4
SQB7014	Risk Theory	4
SQB7015	Stochastic Processes in Finance	4
SQB7016	Computer Intensive Methods	4
SQB7017	Robust Statistics	4
SQB7018	Statistical Methods in Bioinformatics	4
SQB7019	Data Mining	4
SQB7020	Survival Data Analysis	4
SQB7021	Epidemiology Modelling	4

PROGRAMME GOAL

To produce master graduates with strong statistical knowledge, able to analyse and solve problems as well as to think critically; capable of adapting to diverse environments and contributing meaningfully towards professions in different fields.

PROGRAMME LEARNING OUTCOMES

At the end of the programme, graduates of Master of Science in Statistics are expected to demonstrate the following qualities:

1. Comprehend advanced statistical theories with regards to the statistical and mathematical arguments, proofs and abstract concepts.
2. Apply technical and practical skills related to advanced statistics in different fields.
3. Perform professional work with good social skills, behaving responsibly to the community.
4. Practice professionally and ethically in one's profession.
5. Work collaboratively, and articulate statistical ideas clearly, accurately, and effectively.
6. Solve statistical problems using logical, analytical, and critical thinking.
7. Habitually use statistical thinking to understand and manage information.

LIST OF COURSES ACCORDING TO SEMESTER
(PLANNING OF COURSES)

COMPONENT	YEAR 1				TOTAL CREDIT HOURS
	SEMESTER 1		SEMESTER 2		
	COURSES	CREDIT HOUR	COURSES	CREDIT HOUR	
Core Courses	SQB7001 Research Methodology for Statistics	3	SQB7003 Statistical Inference	4	**26**
	SQB7004 Probability Theory	4	SQB7006 Statistical Consultancy and Data Analysis	3	
	SQB7005 Statistical Laboratory	2	SQB7002 Research Project for Statistics	5	
	SQB7002 Research Project for Statistics	5			
Elective Courses	*SQB70***	8	*SQB70***	8	**16**
TOTAL CREDIT HOURS		**22**		**20**	**42**

INSTITUTE OF MATHEMATICAL SCIENCES

The Institute of Mathematical Sciences (ISM) was established as a department in the Faculty of Science when the University of Malaya was founded in Kuala Lumpur in 1959. It has grown into three branches: Pure Mathematics, Applied Mathematics and Statistics. The Institute offers 4 first degree programmes, one master's by coursework programme as well as master's by research and doctoral programmes.

Since the 2017/2018 academic session, ISM has been conducting the Master of Science in Statistics coursework programme. This programme provides opportunities to first degree graduates to acquire advanced statistical knowledge and skills.

STAFF

ISM has a group of experienced lecturers in teaching. They are active in research and have published their work in international as well as local journals. Our research activities encompass a broad spectrum, from findings and knowledge which are abstract in nature, to those with direct applications in the industry. ISM strives to establish and forge close relationships with the industry and other research institutions. This strengthens the quality of teaching and supervision of projects/theses for students at the Bachelors, Masters and Doctoral levels.

HEAD:
Assoc. Prof. Dr. Zailan Siri, *BSc(UM), MSc(UPM), PhD(UKM)*

DEPUTY HEADS:
Assoc. Prof. Dr. Ng Kok Haur, *BSc, MSc(UPM), PhD(UM)*
Assoc. Prof. Dr. Wong Kok Bin, *BSc, MSc, PhD(UM)*
Dr. Noor Fadiya Mohd Noor, *BSc, MSc(UTM), PhD(UKM)*
Dr. Rossita Mohamad Yunus, *BSc, MSc(UM), PhD(USQ)*
Dr. Siti Suzlin Supadi, *BSc, MSc, PhD(UM)*

MATHEMATICS UNIT

COORDINATORS (B.Sc. in MATHEMATICS):
Dr. Kwa Kiam Heong, *BSc, MSc(UM), PhD(Ohio)*
Dr. Loo Tee How, *BSc, MSc, PhD(UM)*

PROFESSOR:
Dr. Angelina Chin Yan Mui, *BSc, MSc(UM), PhD(Queensland)*

ASSOCIATE PROFESSORS:
Dr. Chooi Wai Leong, *BSc, MSc, PhD(UM)*
Dr. Deng Chai Ling, *BSc, MSc, PhD(UM)*
Dr. Wan Ainun Mior Othman, *BSc(UNCC), MSc(N Carolina State), PhD(USM)*
Dr. Wong Kok Bin, *BSc, MSc, PhD(UM)*
Dr. Zailan Siri, *BSc(UM), MSc(UPM), PhD(UKM)*

LECTURERS:
Dr. Amizah Malip, *BSc(IIUM), MSc, PhD(London)*
Dr. Elayaraja Aruchunan, *BSc, MSc(UMS), PhD(Curtin)*
Dr. Kwa Kiam Heong, *BSc, MSc(UM), PhD(Ohio)*
Dr. Loo Tee How, *BSc, MSc, PhD(UM)*
Mr. Mohamad Bakri Zubir, *BSc(Bristol), MSc(Exeter)*
Dr. Muhamad Hifzhudin Noor Aziz, *BSc(UM), MSc, PhD(Glasgow)*
Dr. Noor Fadiya Mohd Noor, *BSc, MSc(UTM), PhD(UKM)*
Dr. Ong Siew Hui, *BSc, MSc, PhD(UM)*
Dr. Oon Shea Ming, *BSc, MSc, PhD(UHP)*
Dr. Siti Suzlin Supadi, *BSc, MSc, PhD(UM)*
Dr. Tan Ta Sheng, *BA, MA, PhD(Cambridge)*

STATISTICS AND ACTUARIAL SCIENCE UNIT

COORDINATOR (B.Sc. in STATISTICS):
Dr. Ng Choung Min, *BSc(UTM), MSc, PhD(UM)*

COORDINATOR (B. ACTUARIAL SCIENCE):
Dr. Koh You Beng, *BSc(UMS), MSc(UM), PhD(HKU)*

COORDINATOR (M.Sc. in STATISTICS):
Dr. Khang Tsung Fei, *BSc, MSc(UM), PhD(NUS)*

PROFESSORS:
Dr. Ibrahim Mohamed, *BSc(Bristol), MSc(Reading), PhD(UiTM)*
Dr. **Nor Aishah Hamzah**, *BSc(Southampton), MSc(Leeds), PhD(Bristol), DipEd(UKM), MIS(UK)*

ASSOCIATE PROFESSORS:
Dr. Khang Tsung Fei, *BSc, MSc(UM), PhD(NUS)*
Dr. Ng Kok Haur, *BSc, MSc(UPM), PhD(UM)*
Mr. Raveendran A/L V G K Menon, *BEng(UM), MActuarialSc(Georgia State), ASA, AIAA*

LECTURERS:
Dr. Adriana Irawati Nur Ibrahim, *BSc(USM), MSc(UM), PhD(Bath)*
Dr. Dharini A/P Pathmanathan, *BSc, MSc, PhD(UM)*
Dr. Koh You Beng, *BSc(UMS), MSc(UM), PhD(HKU)*
Dr. Lim Sok Li, *BEd, MSc, PhD(USM)*
Dr. Mohd Azmi Haron, *BBA, MBA, PhD(UPM)*
Mdm. Nadiah Zabri, *BBA(Wisconsin-Madison), MSc(Kent)*, *AIA*
Dr. Ng Choung Min, *BSc(UTM), MSc, PhD(UM)*
Dr. Nur Anisah Mohamed @ A. Rahman, *BSc, MSc(UM), PhD(Newcastle)*
Dr. Rossita Mohamad Yunus, *BSc, MSc(UM), PhD(USQ)*
Dr. Shaiful Anuar Abu Bakar, *DipActuarialSc, BSc(UiTM), MSc(Heriot-Watt), PhD(Manchester)*

COORDINATOR (B.Sc. Ed. Mathematics):
Dr. Dharini A/P Pathmanathan, *BSc, MSc, PhD(UM)*

RESEARCH AREAS

Research areas at ISM include the following fields: differential geometry, group theory, ring theory, linear preserver problems, functional identities, linear and multilinear algebra, matrix theory, combinatorial graph theory, graph theory, social network analysis, supply chain management, operations research, numerical analysis, computational statistics, robust statistics, probability distribution theory, nonlinear time series, image processing, regression analysis, statistical quality control, mathematical biology, and data science.

COMPUTER FACILITIES

Currently, ISM has a computer lab equipped with 10 laptops, 17 workstations, 121 desktops, 3 laser printers, 1 colour printer, and 4 heavy-duty dot matrix printers, all of which are interconnected through a network system. The lab is also equipped with 4 LCD projectors, 2 visualizers, and 3 scanners. The lab utilises state-of-the-art software such as MATLAB (with various toolboxes), SPSS, Wolfram Mathematica, MathType, Minitab, Microsoft Visual C++, Dev-C++, and S-PLUS. In addition, three of the lecture halls and tutorial rooms are each equipped with an LCD projector and a visualiser.

JOB OPPORTUNITIES

The learning of mathematics increases one's skills in problem solving and analysis. It trains one's mind to process information, form sophisticated ideas, and discern complex arguments. The training to think quantitatively, logically, and analytically is invaluable to one's career.

Since the use of mathematics is all encompassing in human endeavours, a graduate's career opportunity is numerous and not confined to only teaching and research. Our graduates are well-represented in diverse sectors such as finance, business, industry, and information science.

SYNOPSIS OF COURSES

SQB7001 RESEARCH METHODOLOGY FOR STATISTICS

This course is designed to give knowledge and skills to students related to suitable methodologies in statistical research. It includes searching and critically evaluating journal articles on statistical problems. The course will introduce fundamental statistical concepts and techniques useful for research in statistics. The students will be guided in writing research proposal in the area of statistics.

Assessment
Continuous Assessment: 100%

References:

1. Bordens, K.S., & Bruce, B. (2011). *Research and design methods: A process approach* (8th ed.). McGraw-Hill.
2. Flick, U. (2011). *Introducing research methodology: A beginner's guide to doing a research project*. Sage Publication.
3. Hogg, R.V., McKean, J.W., & Craig, A.T. (2013). *Introduction to mathematical statistics* (7th ed.). Pearson.
4. Agresti, A. (2015). *Foundations of linear and generalized linear models*. John Wiley & Sons.

SQB7002 RESEARCH PROJECT FOR STATISTICS

Refer to lecturers concerned for project synopsis and reference texts.

Assessment
Continuous Assessment: 100%

References:
Refer to lecturers.

SQB7003 STATISTICAL INFERENCE

Principles of data reduction. Sufficient statistic. Factorization theorem. Minimal sufficient statistic. Lehman-Scheffe Theorem. Ancillary and complete statistics. Basu's theorem. Exponential class of distributions. Likelihood ratio test. Union-intersection and intersection-union tests. Neyman-Pearson Lemma and its generalization. Most powerful test. Unbiased test. Locally most powerful test. Asymptotic distribution of the likelihood ratio. Sequential probability ratio test.

Assessment:
Final Examination: 50%
Continuous Assessment: 50%

References:

1. Casella, G., & Berger, R.L. (2008). *Statistical inference*. Wadsworth & Brooks/Cole.
2. Devore, J.L., & Berk, K.N. (2012). *Modern mathematical statistics with applications*. Cengage Learning.
3. Hogg, R.V., Craig, A.T., & Mckean, J.W. (2004). *Introduction to mathematical statistics* (6th ed.). Collier MacMillan.
4. Mukhopadhyay, N. (2020). *Probability of statistical inference*. CRC Press.

SQB7004 PROBABILITY THEORY

Introduction to basic concepts, probability measure and space, sigma-fields. Random variables, measurability. Distribution functions. Generating functions, characteristic functions. Various modes of convergence of sequences of random variables. Classical limit theorems. Examples of applications of basic results.

Assessment:
Final Examination: 50%
Continuous Assessment: 50%

References:

1. Bhattacharya, R., & Waymire, E.C. (2016). *A basic course on probability theory*. Springer.
2. Borovkov, A. A. (2013). *Probability theory*. Springer.
3. Karr, A. F. (1993). *Probability*. Springer-Verlag.
4. Ross, S. M. (2019). *Introduction to probability models* (12th ed.). Academic Press.
5. Rotar, V. I. (1997). *Probability theory*. World Scientific.

SQB7005 STATISTICAL LABORATORY

Introduce statistical packages such as S-PLUS, Minitab, SPSS and/or R. Use of functions and commands in statistical packages for exploratory data analysis, modelling and statistical inferences. Coding in a statistical programming language.

Assessment
Final Examination: 50%
Continuous Assessment: 50%

References:

1. Chambers, J.M., Cleveland, W.S., Kleiner, B., & Tukey, P.A. (1983). *Graphical methods for data analysis*. Chapman and Hall.
2. Cornillon, P.-A. et al. (2012). *R for statistics*. CRC Press.
3. Esbensen, K.H., Guyot, D., Westad, F., & Houmoller, L.P. (2009). *Multivariate data analysis: In practice*. CAMO.
4. Lander, J.P. (2017). *R for everyone: Advanced analytics and graphics* (2nd ed.). Addison-Wesley.
5. Venables, W.N., & Ripley, B.D. (2002). *Modern applied statistics with S*. Springer-Verlag.

SQB7006 STATISTICAL CONSULTANCY AND DATA ANALYSIS

Introduction to consulting methods. Related problems and issues. Exposure to the use of secondary data from various sources. Application of suitable statistical methods such as analysis of multivariate data, regression and time series in the analysis of real data. Producing reports and presenting the findings that are suitable for the needs of the statistical practitioners. Introduction to consultancy activities.

Assessment
Continuous Assessment: 100%

References:

1. Boen, J.R., & Zahn, D.A. (1982). *The human side of statistical consulting*. Lifetime Learning Publications.
2. Hand, D.J., & Everitt, B.S. (1987). *The statistical consultant in action*. Cambridge University Press.

3. Tufte, E.R. (2001). *The visual display of quantitative information*. (2nd ed.). Graphics Press.
4. Wickham, H., & Grolemund, G. (2017). *R for data science: Import, tidy, transform, visualize, and model data*. O'Reilly Media.

SQB7007 MULTIVARIATE ANALYSIS

The use/application of multivariate analysis. Managing and handling multivariate data. Matrix theory. Random vectors and matrices. The multivariate normal and Wishart distributions. Selected topics from graphical methods, Regression analysis, correlation, principal components, factor analysis, discriminant analysis and clustering methods.

Assessment:
Final Examination: 50%
Continuous Assessment: 50%

References:
1. Anderson, T. W. (1984). *An introduction to multivariate statistical analysis* (2nd ed.). John Wiley.
2. Arnold, S. F. (1981). *The theory of linear models and multivariate statistical analysis*. John Wiley.
3. Chatfield, C., & Collins, A. J. (1980). *Introduction to multivariate analysis*. Chapman & Hall.
4. Konishi, S. (2014). *Introduction to multivariate analysis - Linear and nonlinear modellin*g. Chapman & Hall.
5. Johnson R.A., & Wichern, D.W. (2007). *Applied multivariate statistical analysis* (6th ed.). Pearson International Education.

SQB7008 STOCHASTIC MODELS

Poisson processes, backward and forward Kolmogorov equations, birth and death processes and examples. Definition and concepts in renewal processes, distribution for the number of renewal, renewal function and theorems for renewal processes. Backward and forward renewal times. Examples for various types of renewal processes. Examples of applications of the theory in renewal processes.

Assessment:
Final Examination: 50%
Continuous Assessment: 50%

References:
1. Cox, D. R. (1962). *Renewal theory*. Methuen and Company, Ltd.
2. Cox, D. R., & Miller, H. D. (1977). *The theory of stochastic processes*. Chapman & Hall.
3. Durrett, R. (2016). *Essentials of stochastic processes*. Springer.
4. Taylor, H. M., & Karlin, S. (1998). *An introduction to stochastic modelling*. Academic Press.

SQB7009 BAYESIAN STATISTICS

Different functions relevant to Bayesian statistics, calculation of $E(x)$ and Var(x). Hypothesis testing of proportion, mean for posterior distribution, choice of sample size. Sufficient statistics and efficiency. Bayesian estimators and properties of estimators. Loss function, Bayesian risk. Decision theory on x^2 , subjective information compared to objective information. Bayesian decision criterion. Expected opportunity loss (EOL). Bayesian inference - beta-binomial, uniform prior, beta prior, conjugate family, Jeffrey's prior. Choosing the prior beta-binomial - with vague prior, with conjugate prior information, choosing prior when you have real prior knowledge, constructing a general continuous prior, effect of prior. Bayes' theorem for normal mean with discrete and continuous prior. Flat prior density (Jeffrey's prior) for normal mean.

Assessment:
Final Examination: 50%
Continuous Assessment: 50%

References:
1. Berger, J. O. (2010). *Statistical decision theory and Bayesian analysis* (2nd ed.). Springer-Verlag.
2. Cover, T.M., & Thomas, J.A. (2006). *Elements of information theory* (2nd ed.). John Wiley.
3. Cowles, M. K. (2013). *Applied Bayesian statistics with R and Open Bugs example*. Springer.
4. Gelman, A. (1995). *Bayesian data analysis*. Chapman & Hall.
5. Lee, P.M. (1997). *Bayesian statistics: An introduction*. Arnold.
6. Polson, N.G., & Tiao, G.C. (1995). *Bayesian inference*. Edward Elgar Pub.

SQB7010 DECISION STATISTICS

Introduction to decision theory, risk function, non-randomized rules, randomized rules, completeness. Bayes rules and minimax. Comparison of decision theory and classical theory.

Assessment:
Final Examination: 50%
Continuous Assessment: 50%

References:
1. Blackwell, D., & Girshick, M. A. (1954). *Theory of games and statistical decision*. John Wiley.
2. Ferguson, T. S. (1967). *Mathematical statistics: A decision theory approach*. Academic Press.
3. French, S., & Insua, D. R. (2000). *Statistical decision theory*. Arnold.
4. Hoggar, S. G. (2006). *Mathematics of digital images*. Cambridge University Press.
5. Longford, N.T. (2013). *Statistical decision theory*. Springer.

SQB7011 GENERALIZED LINEAR MODELS

Generalized linear models based on exponential family. Regression models like binomial (logistic) and Poisson in detailed. Statistical software for data analysis.

Assessment:
Final Examination: 50%
Continuous Assessment: 50%

References:
1. Agresti, A. (2013). *Categorical data analysis* (3rd ed.). Wiley-Interscience.
2. Dobson A.J., & Barnett, A.G. (2008). *An introduction to generalized linear models* (3rd ed.). Chapman & Hall/CRC.
3. McCullagh, P., & Nelder J.A. (1989). *Generalized linear models* (2nd ed.). Chapman & Hall.
4. Myers, R.H., Montgomery, D.C., Vining, G.G., & Robinson, T.J. (2010). *Generalized linear models: With applications in engineering and the sciences* (2nd ed.). John Wiley & Sons.
5. Smithson, M., & Merkle, E.C. (2014). *Generalized linear models for categorical and continuous limited dependent variables*. Chapman & Hall/CRC.

SQB7012 EXPERIMENTAL DESIGN AND QUALITY ENGINEERING

Factorial designs. Blocking and confounding. Fractional factorial designs and its resolution. Supersaturated design. Response surface method and design. Robust parameter design.

Assessment:
Final Examination: 50%
Continuous Assessment: 50%

References:

1. Box, G.E.P., Hunter, J.S., & Hunter, W.G. (2005). *Statistics for experimenters*. John Wiley.
2. Fowlkes, W.Y., & Creveling, C.M. (1995). *Engineering methods for robust product design: Using Taguchi methods in technology and product development*. Addison-Wesley.
3. Montgomery, D.C. (2012). *Design and analysis of experiments* (8th ed.). John Wiley.
4. Myers, R.H., Montgomery, D.C., & Anderson-Cook, C. M. (2009). *Response surface methodology: Process and product using designed experiments* (3rd ed.). John Wiley.
5. Wu, C.F.J., & Hanada, M.S. (2009). *Experiments: Planning, analysis and optimization* (2nd ed.). John Wiley.

SQB7013 STATISTICAL TIME SERIES

Basic concept of model-building. Stationary model linear. Autoregressive and moving average model (ARMA). ARMA process: includes identification, estimation, fitting and prediction. Non-stationary and seasonal model. GARCH model. Multivariate time series. State-space model.

Assessment:
Final Examination: 50%
Continuous Assessment: 50%

References:

1. Box, G.E.P., Jenkins, G.W., & Reinsel, G. (2008) *Time series analysis, forecasting and control* (4th ed.). Prentice Hall.
2. Brockwell, P. J., & Davis, R. A. (2010). *Introduction to time series and forecasting* (2nd ed.). Springer.
3. Cryer, J.D. (2008). *Time series analysis: With application in R.* Springer.
4. Shumway, R. H., & Stoffer, D. S. (2011). *Time series analysis and its applications*: *With R examples* (2nd ed.). Springer.

SQB7014 RISK THEORY

This course will emphasize on the applications of probabilistic models in the risky business, especially in insurance using the theory in stochastic processes. The topics include individual risk model, computation of net premium, security loading and reinsurance, utility theory and its applications in reinsurance, collective risk model, number of claims distribution, aggregate claims distribution, surplus process and ruin probabilities.

Assessment:
Final Examination: 50%
Continuous Assessment: 50%

References:

1. Bower, N.L., Gerber, H.U., Hickman, J.C., Jone, D.A., & Nesbitt, C.J. (1997*). Actuarial mathematics* (2nd ed.). The Society of Actuaries.
2. Buhlmann, H. (1996). *Mathematical methods in risk theory*. Springer.
3. Kellison, S.G., & London, R.L. (2011). *Risk models and their estimation*. ACTEX.
4. Klugman S.A., Panjer, H.H., & Willmot, G.E. (2008). *Loss models: From data to decisions* (3rd ed.). John Wiley & Sons.
5. Tse, Y. (2009). *Nonlife actuarial models: Theory, methods and evaluation.* Cambridge University Press.

SQB7015 STOCHASTIC PROCESSES IN FINANCE

Brownian motion and Ito's lemma. Evaluation of option and future prices using Martingale and risk-neutral probabilities. Black-Scholes formula, Stochastic interest rate and volatility.

Assessment:
Final Examination: 50%
Continuous Assessment: 50%

References:

1. Hirsa, A., and Neftci, S.N. (2014). *An introduction to the mathematics of financial derivatives* (3rd ed.). Academic Press.
2. Klebaner, F.C. (2005). *Introduction to stochastic calculus with application* (2nd ed.). Imperial College Press.
3. Privault. N. (2014). *Stochastic finance: An introduction with market examples*. CRC Press.
4. Shreve, S.E. (2004). *Stochastic calculus for finance I: The binomial asset pricing model*. Springer.
5. Shreve, S.E. (2004). *Stochastic calculus for finance II: Continuous-time models*. Springer.

SQB7016 COMPUTER INTENSIVE METHODS

Error in floating point calculations. Probability function and distribution function approximations. Generating random numbers, including evaluating the quality of the generator and calculation methods in linear algebra: Gaussian elimination, sweep operators. Calculation methods for multiple regression, (not constrained) nonlinear regression and model fitting other than the least squares, bootstrap and Markov chain Monte Carlo.

Assessment:
Final Examination: 50%
Continuous Assessment: 50%

References:

1. Chernick, M.R. (2011). *Bootstrap methods: A guide for practitioners and researchers* (2nd ed.). Wiley-Interscience.
2. Gentle, J.E., Hardle, W.K., & Mori, Y. (2012). *Handbook of computational statistics: Concept and methods*. Springer.
3. Hjorth, J.S. (1994). *Computer intensive statistical methods: Validation, model selection and bootstrap*. Chapman & Hall.
4. Kennedy, W.J., & Gentle, J.E. (1980). *Statistical computing*. Marcel Dekker.
5. Peng, R.D. (2016). *R programming for data science*. Leanpub.

SQB7017 ROBUST STATISTICS

Foundation for robust methods: tools for judging robustness, measures of location and inference functions, measures of scales, M-estimation of location. Some outlier detection methods. Inference statistics: Construction of confidence intervals. Robust regression estimation M, GM, MM estimation dan inference functions.

Assessment:
Final Examination: 50%
Continuous Assessment: 50%

References:

1. Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., & Stahel, W. A. (1986). *Robust statistics: The approach based on influence functions*. Wiley.
2. Huber, P. J., & Ronchetti, E.M. (2009). *Robust statistics* (2nd ed.). John Wiley & Sons.

3. Maronna, R.A., Martin, D.R., & Yohai, V.J. (2006). *Robust statistics: Theory and methods*. John Wiley & Sons.
4. Rousseeuw, P.J., & Leroy, A. M. (2003). *Robust regression and outlier detection*. Wiley-Interscience.

SQB7018 STATISTICAL METHODS IN BIOINFORMATICS

Statistical modelling of DNA/protein sequences: Assessing statistical significance in BLAST using the Gumbel distribution; DNA substitution models; Poisson and negative binomial models for gene counts; Hidden Markov Model. Algorithms for sequence analysis and tree construction: Dynamic programming for sequence alignment and Viterbi decoding; neighbour-joining, UPGMA, parsimony and maximum likelihood tree-building methods. Analysis of high-dimensional microarray / RNA-Seq gene expression data: Statistical tests for detecting differential expression, feature selection, visualisation, and phenotype classification.

Assessment:
Final Examination: 50%
Continuous Assessment: 50%

References:
1. Buffalo, V. (2015). *Bioinformatics data skills*. O' Reilly Media.
2. Durbin, R., Eddy, S., Krogh, A., and Mitchison, G. (1998). *Biological sequence analysis: Probabilistic models of proteins and nucleic acids*. Cambridge University Press.
3. Ewens, W.J., and Grant, G.R. (2005). *Statistical methods in bioinformatics: An introduction* (2nd ed.). Springer.
4. Jones, N.C., and Pevzner, P.A. (2004). *An Introduction to bioinformatics algorithms*. MIT Press.
5. Pevsner, J. (2009). *Bioinformatics and functional genomics* (2nd ed.). Wiley-Blackwell.

SQB7019 DATA MINING

Introduction to statistical methods and tools for analysing very large data sets and search for interesting and unexpected relationships in data. Data measurement: types of measurements, distance measure, data quality. Data reduction: Data organisation and display; Principal component, multidimensional scaling. Data analysis and uncertainty: Handling uncertainty; statistical inference; sampling Data mining Algorithms: Classification and clustering – CART; artificial neural network; support vector machine; mining ordered dependence. Modelling: model structure; curse of dimensionality; score function; optimisation methods; descriptive modelling and prediction. Data organisation.

Assessment:
Final Examination: 50%
Continuous Assessment: 50%

References:
1. Bramer, M. (2013). *Principles of data mining* (2nd ed.). Springer-Verlag.
2. Flach, P. (2012). *Machine learning: The art and science of algorithms that make sense of data*. Cambridge University Press.
3. Hastie, T., Tibshirani, R., & Friedman, J. (2017). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.
4. Wickham, H., & Grolemund, G. (2017). *R for data science: Import, tidy, transform, visualize, and model data*. O'Reilly Media.

SQB7020 SURVIVAL DATA ANALYSIS

Basic concepts such as survival and hazard functions. Survival data analysis including life table, Kaplan-Maier; log-rank and Wilcoxon tests. Survival regression modeling including the Cox regression model, several parametric models and the accelerated lifetime model and risk model. Diagnostic checking of the models. Application to real datasets.

Assessment
Final Examination: 50%
Continuous Assessment: 50%

References:
1. Collett, D. (2015). *Modelling survival data in medical research* (3rd ed.). Chapman and Hall.
2. Hosmer, D.W., Lemershow, S., & May, S. (2011). *Applied survival analysis: Regression modeling of time to event data*. John Wiley.
3. Lee, E.T. (2013). *Statistical methods for survival data analysis* (2nd ed.). John Wiley & Sons.
4. Miller, R.G. (2011). *Survival analysis* (2nd ed.). John Wiley & Sons.

SQB7021 EPIDEMIOLOGY MODELLING

Measures of disease frequency and risk. Assess strengths and limitation of different sources of epidemiology data. Principles of study design: cross-sectional, cohort, case-control, and intervention studies. Interpretation of epidemiology studies; causality, random errors, bias, confounding. Regression methods for case-control studies. Unconditional and conditional logistic regression. Regression methods for cohort studies. Issues in the case-control and cohort studies. Evaluating published papers on epidemiology studies.

Assessment:
Final Examination: 50%
Continuous Examination: 50%

References:
1. Hebel, J.R., & McCarter, R.J. (2006). *Study guide to epidemiology and biostatistics*. Jones and Bartlett Publisher.
2. Merrill, R. M. (2010). *Introduction to epidemiology* (5th ed.). Jones and Bartlett Publisher.
3. Rothman, K.J. (2012). *Epidemiology: An introduction* (2nd ed.). Oxford University Press.
4. Susan, C. (2007). *An introduction to public health and epidemiology*. Open University Press.