Short Title:	Statistics APPROVED
Full Title:	Statistics

Module Code:	ADSA H6019

ECTS credits:	10

NFQ Level:	9

Module Delivered in	1 programme(s)

Module Contributor:	Damian Cox

Module Description:	The purpose of this module is to provide the postgraduate student with the concepts, tools and techniques needed to undertake standard statistical analysis and to use these concepts to underpin their adoption of data mining techniques.

Learning Outcomes:

On successful completion of this module the learner will be able to

1. Summarize large sets of data, including grouped data, using the standard measures of central tendency and dispersion and their definitions and properties, and represent it graphically, by following an agreed set of conventions.
2. Apply the laws of probability to questions involving random variables and events, and move on to the concept of a random variable and its distribution, the meaning of expected values, and the properties of common distributions such as the normal, binomial, Poisson and exponential distributions.
3. Interpret the concept of a statistic as a random variable arising from sample data, with the central limit theorem determining the behaviour of such statistics and thereby underpinning many statistical tests.
4. Frame and use an appropriate test for a statistical problem, based on their knowledge of hypothesis testing, the central limit theorem and those distributions used in a range of common statistical tests. This will include multivariate analyses – Manova, Mancova.
5. Design or explain the chosen structure of an experiment and the meaning of any data analysis produced for that experiment, based on the students understanding of the properties of Analysis of Variance and Analysis of Covariance and other statistical tests.
6. Apply their knowledge of techniques derived from linear algebra to the matrix formulation of the general linear model, including eigenvector decompositions of the covariance matrix and their application to Principal Component Analysis.

Module Content & Assessment

Indicative Content

Linear Algebra
The definition of a matrix. Matrix algebra, including the addition and multiplication of matrices and multiplication by scalars. The representation of vectors as matrices. The definition of the determinant and the inverse for square matrices. Methods for calculating the inverse of a matrix, including the cofactor method and Gauss-Jordan elimination. Solving systems of linear equations using the inverse. Eigenvalues and eigenvectors.

Review of Descriptive Statistics
Calculation of mean, mode, median and standard deviation. Grouped data, calculation of class intervals, calculation of mean, mode, median and standard deviation for grouped data. Data representation and types of charts. Linear regression and correlation as geometric ideas and as data analysis techniques.

Probability
The definition of the fundamental ideas of events, experiments and probability. Independent events, conditional probabilities and the addition and multiplication laws. Permutations and combinations. The concepts of a random variable and its distribution, the definition of population parameters in terms of the probability distribution function and the cumulative probability distribution. Discrete and continuous probability distributions, including the exponential, normal, binomial and Poisson distributions. Examples of the role of these distributions in reliability prediction, component failures and designing for reliability.

Fundamentals of Hypothesis Testing
The concept of a Hypothesis test. The concept of a statistic. The common population parameters as statistics. The Central Limit Theorem and the concept of standard error. The role of the normal distribution arising from the Central Limit theorem. The representation of the results of a test; critical values and confidence intervals. The concept and limitations of a Hypothesis test, including type I and II errors and their probabilities.

Standard Hypothesis Tests
Distributions including the 'Student t', the chi-square and the F distributions. The F distribution as a ratio of chi-square distributions. Standard tests, including tests on means and variances, paired sample and unpaired tests on comparisons of means. Categorical tests using the chi-square distribution, such as goodness-of-fit tests to a distribution and tests for independence. Linear regression and correlation as statistical tests. The power of a statistical test. Effects sizes and the calculation of sample sizes. Reporting the results of an experiment and Hypothesis test, communicating the meaning of a test to peers and colleagues from non-technical backgrounds, interpreting existing reports and academic papers.

Multivariate Statistics
The design of experiments and the comparison of group means by one- and two-way analysis of variance (ANOVA). Relating an experiment to the form of the data collected. The type and nature of response variables and the concept of an attribute. Multiple regression and the General Linear Model. Easing of assumptions on the errors for generalised linear models. The General linear model as the foundation for Analysis of Variance and Analysis of Covariance, including Multivariate models (MANOVA, ANCOVA, MANCOVA)

Principal Component Analysis
Rotations and Orthogonal transformations. Eigenvalue decomposition. The eigenvectors of the covariance matrix, rotations and orthogonal transformations of the variables. Eigenvalue decomposition of the covariance matrix. The role of transformations in investigating attributes.

Bayesian inference
The Bayesian concept and method. The nature of priors. Bayesian testing. Comparison of Bayesian methods with Null Hypothesis based statistical testing. Large sample properties of Bayesian inference.

Parameter estimation
Parametric inference and the Maximum likelihood estimate. The maximum likelihood estimator and its properties, including asymptotic normality. The method of moments for parametric inference.

Indicative Assessment Breakdown	%
Course Work Assessment %	100.00%

Course Work Assessment %				
Assessment Type	Assessment Description	Outcome addressed	% of total	Assessment Date
Practical/Skills Evaluation	Hypothesis testing I: The student will be given an assignment on Hypothesis testing, implementing a range of the statistical tests covered in the module, including tests on means and variances, tests on group means, correlation and regression, and tests for goodness-of-fit and independence. The student will be assessed on their ability to establish the conceptual framework of any test, the Null and alternative Hypothesis, identify the parameters of a given test and draw the correct conclusions and the meaning of type I and II errors. The students will be given chaotically generated, recoverable, data sets for this assignment, so that they may collaborate up to a point.	1,3,4	20.00	Week 4
Practical/Skills Evaluation	Hypothesis testing II: The student will be given an assignment on Analysis of Variance, where they will identify a range of experimental designs testing scientific Hypotheses, the corresponding test and the required partitions of sums of squares for the analysis of variance layout. The student will be assessed on their ability to establish the conceptual framework of the tests and drawing the correct conclusions. The students will be given chaotically generated, recoverable, data sets for this assignment, so that they may collaborate up to a point.	4,5	25.00	Week 6
Open-book Examination	Probability: The student will be set a number of questions on the theoretical, probability element of the module, including its application to problems such as reliability and quality control, the fundamental definitions of probability, the Central limit theorem and its implications, the properties and definitions of common distributions and the theory of the general linear model.	2,6	30.00	Week 10
Case study	Interpreting the results of an analysis of an existing or historical data set, writing up a report at an appropriate academic standard on these results, and interpreting them for peers and non-technical colleagues.	1,4,5	25.00	n/a

No Final Exam Assessment %

Indicative Reassessment Requirement
Repeat the module *The assessment of this module is inextricably linked to the delivery. The student must reattend the module in its entirety in order to be reassessed.*

ITB reserves the right to alter the nature and timings of assessment

Indicative Module Workload & Resources

Indicative Workload: Part Time	
Frequency	**Indicative Average Weekly Learner Workload**
Every Week	52.00
Every Week	148.00

Resources

Recommended Book Resources

Chris Chatfield 1983, *Statistics for technology*, Chapman & Hall London [ISBN: 0412253402]

Larry Wasserman, *All of Statistics*, Springer New York [ISBN: 1441923225]

James E. Gentle, *Matrix Algebra*, Springer New York [ISBN: 1441924248]

Michael Baron., *Probability and statistics for computer scientists*, ; Chapman and Hall/CRC [ISBN: 1439875901]

Henk Tijms, *Understanding Probability*, Cambridge University Press [ISBN: 110765856X]

John Fox., *Applied regression analysis and generalized linear models*, Thousand Oaks, Calif; Sage [ISBN: 1452205663]

Supplementary Book Resources

David A. Freedman, *Statistical models*, Cambridge ; Cambridge University Press, 2009. [ISBN: 0521743850]

Leonard Mlodinow, *The Drunkard's Walk: How Randomness Rules Our Lives*, Vintage [ISBN: 9780307275172]

This module does not have any article/paper resources

This module does not have any other resources

Module Delivered in

Programme Code	Programme	Semester	Delivery
BN_KADSA_R	Master of Science in Computing in Applied Data Science & Analytics	3	Elective