The University of Western Australia
DEPARTMENT OF MATHEMATICS AND STATISTICS

STAT2402 Analysis of Observations

Semester 2 2023

UNIT OUTLINE

Co-ordinator: R. Nazim Khan Room 207, Mathematics Building **Email**: nazim.khan@uwa.edu.au

1 Prerequisites

Students should preferably have taken STAT1400 Statistics for Science or STAT1520 Economic and Business Statistics prior to takin this unit.

2 Unit Resource Material

Lecture slides are available from the unit LMS. Laboratory sheets are also available online.

1. **Recommended text** Faraway, J. J. (2016). Extending the Linear Model with `R`: Generalized Linear, Mixed Effects and Nonparametric Regression Models, Second Edition. Chapman and Hall/CRC.

   Recommended Reading

2. Faraway, J. J. (2014). Linear Model with `R`, Second Edition. Chapman and Hall/CRC.

3. Ramsey, F. L., and Schafer, D. W. (2013). The statistical sleuth: A course in methods of data analysis. Third edition. Boston: Brooks/Cole, Cengage Learning. Available online in the UWA library.

3 Teaching and Learning Regime

Teaching and learning will be based on a weekly two-hour lecture session and a weekly two-hour laboratory session. Lecture notes will be available on LMS prior to the class. This unit is delivered in *flipped mode*. ALL the lectures will be pre-recorded and put on LMS before the lecture session. Each recorded lecture will be around ten to fifteen minutes long, and will cover the core content. The lectures will follow the slides in the Unit Reader. In the weekly lecture session the lecture material will be reinforced and extended, student questions answered and misunderstandings addressed. Real studies will be discussed, and students will get the chance to work on these in groups. Practice and demonstrations with `R` will also be covered.

The lectures will be supported by a weekly two-hour laboratory class in which students will analyse data using the `R` statistical environment. This is a crucial part of the teaching and learning and if students do not complete the laboratory work each week then they will quickly fall behind. Note that the laboratory work will be assessed, and some of the sessions will be used for short lab-based tests. Solutions to the laboratory exercises will be posted on LMS weekly in time with the coverage of the material.

THE LABORATORIES BEGIN IN WEEK 1 OF SEMESTER.

Warning The material moves quite fast, and it is very easy to fall behind. You should stay up to date with lectures and laboratory material. This will give you the best chance to do your best.

4 LEARNING OUTCOMES

Overall Learning Outcome: Students should be able to

1. demonstrate their knowledge of fundamental concepts in probability and statistics;
2. apply statistical models to real-world problems for discrete data, in particular count data;
3. use computer package(s) for fitting such models to data;
4. communicate the results of these analyses effectively to non-statisticians.

Detailed Learning Outcomes

1. Fit a linear model to data. Perform model diagnostics, and interpret the results.
2. Understand the structure of discrete data.
3. Perform inference for a single proportion, including hypothesis tests and obtaining confidence intervals.
4. Perform hypothesis tests for two proportions from independent samples, and obtain confidence intervals for a difference of two proportions.
5. Fit a log-linear model for a proportion.
6. Fir a log-linear model to binomial data.
7. Fit a log-linear model to count data based on a Poisson distribution.
8. Understand the principle of over-dispersion, and fit a negative binomial distribution to data where appropriate.
9. Understand the principle of likelihood,and write likelihood equations for models in this unit.
10. Understand the principle of maximum likelihood, and obtain maximum likelihood estimates using software.
11. Understand the principles of the likelihood ratio test.
12. Interpret the results of statistical analyses and communicate the findings to a non-statistician.

5 UNIT CONTENT

Proposed lecture schedule for each week is indicated. Some variation from this may occur. Please note the assessment dates.

Monday	Week	Topics	Important dates
24 Jul	1	**Part I: Introduction** Aims and Motivation. Unit outline. Introduction to `R, RStudio, R Markdown`. *1. Revision of Linear Statistical model.* Simple linear regression, Model equation and assumptions, model fitting in `R`, model diagnostics. `R libraries, ggplot`.	
31 Jul	2	Interpretation and reporting results. Verifying model assumptions. Models with categorical variables—Linear statistical model.	
7 Aug	3	**Part II: Population Proportion** *2. Probability* Probability rules, Conditional probability, Independent events. *3. Discrete distributions.* Discrete random variables, mean and variance, Bernoulli distribution, Binomial distribution, Poisson distribution.	
14 Aug	4	*4. Inference for population proportion.* Inference for population proportion based on a single sample, distribution of sample proportion, confidence interval for single proportion.	
21 Aug	5	*5. Comparing proportions and odds.* Comparing two proportions based on normal approximation, confidence intervals of difference of two proportions. Two-by-two contingency tables, Odds ratios.	
28 Aug	6	*6. Analysis of categorical or count data.* Chi-squared test for homogeneity and independence. Fisher's exact test.	
4 Sep		**NON-TEACHING WEEK**.	Assignment 1 Due 11 pm, Sunday 10 September
11 Sep	7	**Part III: The Generalised Linear Model (GLM)** *7. Binary logistic regression model.* Model equation, fitting the model in `R`. Model selection—Wald test. Plotting predicted probabilities. Model interpretation. Confidence intervals for parameters. Model selection—decrease is deviance test.	
18 Sep	8	*8. Logistic regression model for binomial data.* Logistic regression for binomial counts, probability model, model fitting in `R`, model selection. Model interpretation. Model assessment, Pearson residuals, Deviance goodness of fit test. Common limitations of logistic regression. The dispersion parameter. Conditional logistic regression.	
25 Sep	9	*9. Poisson regression.* The Poisson probability model, hypothesis test for Poisson mean. The Poisson regression model, interpretation of model coefficients. Model fitting in `R`, model selection, model interpretation, model diagnostics. Prediction.	
2 Oct	10	*10. Negative binomial regression.* Dispersion parameter, overdispersion. Negative binomial probability model. Fitting negative binomial regression in `R`, model interpretation. Prediction.	
9 Oct	11	*11. Zero-inflated Poisson regression.* Examples, model specification, model fitting. Steps to fitting Poisson model.	Assignment 2 Due 11 pm Sunday 15 October.
16 Oct	12	**Part IV: Maximum Likelihood** *12. Likelihood function.* Joint distributions, independence. The likelihood function. Examples in discrete cases. Likelihood for binary data, binomial data. Evaluating the likelihood in `R`, plotting the likelihood. Likelihood for Poisson data. Likelihood for two or more parameter models. *13. Maximum likelihood estimation.* Maximum likelihood estimates. Two-parameter model. Score function. Numerical solution in `R`. Maximum likelihood for Poisson regression parameters. *13. Information and Sampling distribution of MLEs* Observed information, Expected or Fisher Information, asymptotic variance of the MLE. Properties of MLEs. Invariance principle of MLE. *14. Likelihood based inference.* Confidence intervals, hypothesis tests, Wald test for a single parameter. Likelihood ratio tests.	

6 Assessments

6.1 Assessment components

Your final mark in STAT2402 will be made up as follows:

12 weekly quizzes @ 3%	36%
Assignment 1 @ 10%	10%
Assignment 2 @ 14%	14%
Final Exam	40%
Total	100%

All assessments, including the final examination, will be open book. Please note the following.

1. **Given the timings of the semester assessments no opportunity for an alternative sitting of the assessments is possible.**
2. Any student who is unable to take an assessment needs to submit a special consideration to their faculty BEFORE the assessment. Students with approved special consideration who do not take an assessment will have the weighting transferred to the final exam.

6.2 Lab workbook

Students should keep a record of their weekly laboratory work, either as a hard copy or a soft copy. These will be a very useful resource in the tests and final examination.

6.3 Lab-based tests

The lab-based tests will use `R` but may also contain some written questions.

6.4 Final Examination

The final examination will be due in the Study Break or the first week of examinations. It will be an open book, take home examination. More details will be provided a few weeks before the due date.

6.5 Assessment details

Test	Due Date	Topics	Learning Outcomes
Weekly online quizzes	11 pm Sunday each week	Weekly lab material	All
Assignment 1	11 pm Sunday 10 September	Up to Section 4: Analysis of Categorical or Count data.	All
Assignment 2	11 pm Sunday 15 October	Up to Section 6: Logistic Regression for Binomial data.	All

7 University and Faculty Policies

It is your responsibility to be familiar with **ALL** University and Faculty policies that apply to you as a student.

R. Nazim Khan
MATH1721 Co-ordinator
July 2023