SM-1402: Basic Statistics

Dr. Haziq Jamil
Assistant Professor in Statistics
haziq.jamil@ubd.edu.bn
https://haziqj.ml

Ms. Indira Puteri Kinasih
Teaching Assistant
21h8203@ubd.edu.bn

Semester II 2022/23
document last updated: 2022-12-27

Module Description

Welcome to SM-1402! This is a Level 1 Breadth module offered by FOS weighing 2 MCs. As stated on the module description, Basic Statistics *"is designed to provide the students the fundamental knowledge of statistics, its application and the basic concepts of random variables and sampling"*. Being a breadth module, there will be less emphasis on theory, and more on statistical concepts and applications.

The goal is to teach you, at the introductory level, the toolbox required for statistical analysis of data. In 14 weeks, we aim to cover the following topics:

1. Data collection *[self-learnt]*
2. Presentation of data
3. Descriptive statistics
4. The normal distribution
5. Analysis of qualitative data
6. Linear relationships
7. Comparing grouped data

Module Contents and Learning Outcomes

Visualisation and descriptive statistics

- Know which best visualisation tool to use for a particular data type.
- Construct and draw a histogram for continuous data.
- Summarise data in the form of tables for concisely presenting information.
- Calculate the various measures of location (average, median, mode) and know the differences between them.
- Calculate the various measures of spread (range, IQR, SD) and know the differences between them.
- Understand that different data types require different measures of location and spread.

The normal distribution

- Characterize the normal distribution by its two parameters, and understand the effect of changing these two values.
- Understand the role of the normal distribution function and its relation to areas and probabilities.

- Use the standard normal table to calculate normal probabilities and also use it in reverse to find quantiles.
- Be able to simply evaluate the assumption of normality for data.
- Caculate the point estimate for the mean as well as the 95% confidence interval for it.

Analysis of qualitative data

- Be familiar with the binomial distribution and its relationship to the normal distribution when n is large.
- Calculate sample proportions and construct a confidence interval to quantify the statistical uncertainty.
- Perform a statistical test for testing hypothesis regarding proportions.
- Perform a Chi-squared test of independence for frequency tables.

Linear relationships

- Understand and quantify the linear association between two variables by calculating the correlation coefficient between them.
- Be familiar with the concept of least squares regression and know how to compute the least squares estimates for the intercept and slope for a simple linear regression model.
- Diagnose a linear model by analysing the residuals.
- Conduct a test of significance for the slope parameter of a linear regression model.
- Use the fitted model for statistical inference (interpreting the coefficients) and forecasting.

Comparing grouped data

- Perform the paired t-test to compare group means in a before/after situation.
- Perform the two sample t-test (unequal variances) to compare means of two groups.
- Perform the ANOVA test to compare means of two or more groups.
- Understand the assumptions and limitations of the t-tests/ANOVA, and know when to apply each depending on the situation.

Reading list

The main textbook for this module will be

- Madsen, B. S. (2016). *Statistics for Non-Statisticians* (2nd ed.). Springer Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-49349-6

Each topic presented during lectures correspond to at least one chapter from the Madsen book. You are advised to read the accompanying chapter prior to attending lectures, and definitely prior to attending problem classes.

Supplementary reading list

- Ross, S. M. (2019). *A first course in probability.* Pearson Boston
- Navarro, D. (2013). *Learning statistics with R.* https://learningstatisticswithr.com

Class Format

Classes will alternate between lectures and problem classes. Both kinds of classes are timetabled once a week:

- Mondays, 2.10pm–4.00pm @ Mini Theatre 1, UTH (ICTC Building)

Lecture slides are provided, and you are welcome to supplement these with your own notes. It is recommended to bring the following items to class (lecture or tutorial)

- A copy of the slides or exercise sheet
- Blank paper
- Pens and pencils
- Calculators
- Laptops or tablets that run MS Excel (optional)

Statistics is all about applying math to real world data. As we learn statistical concepts, we will do some 'manual' calculations on pen and paper. The book chapter contains instructions to doing the calculations using MS Excel. You are encouraged to run through those examples yourself, and you might need them for your assignment.

Assessment

Formative assessment

- (Biweekly) Exercise sheets

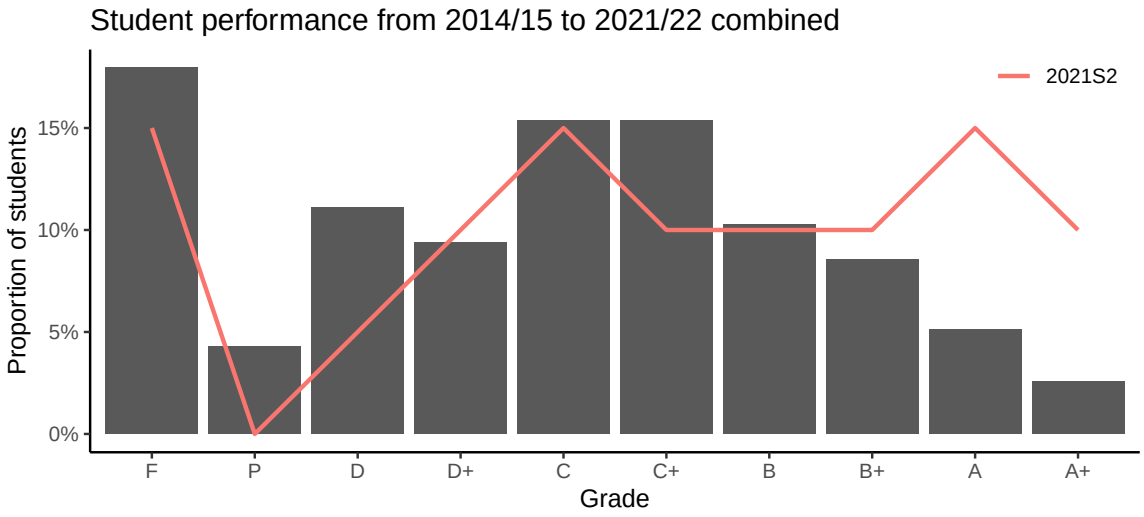
Summative assessment

- **60% examination**: The exam will contain the 'usual' questions (similar to the exercise sheets). Expect to do calculations. A formula sheet will be provided.

  The scheduled date for the exam is **Monday, 15th May 2023 @ Chancellor's Hall**.
- **10% quizzes**: 2 × Canvas quizzes
- **10% graded discussions**: 2 × Canvas discussions
- **20% assignment**: You will be tackling a *data analysis*-type question, and you are expected to apply any or all techniques that you have learnt in the course. Submission will be a PDF report, and you may use any statistical software to aid your analysis.

Key Data

- Past class sizes: 2014S2 = 9, 2015S2 = 23, 2016S2 = 19, 2019S2 = 16, 2020S2 = 30, 2021S2 = 20 (average: 19.5)
- SFE grade average: 4.22 / 5.00



Class Schedule

Week 01 (09/01/23)

- **No classes scheduled**
- Reading: Chapter 1 (Madsen, 2016)
- Canvas discussion

Week 02 (16/01/23)

- Lecture: Presentation of data
- Reading: Chapter 2 (Madsen, 2016)

Week 03 (23/01/23)

- **No classes** *due to Chinese New Year holiday*
- Canvas discussion

Week 04 (30/01/23)

- Lecture: Descriptive statistics
- Reading: Chapter 3 (Madsen, 2016)

Week 05 (06/02/23)

- Tutorial: Exercises 1 & 2

Week 06 (13/02/23)

- Lecture: The normal distribution
- Reading: Chapter 4 (4.1–4.8 only) (Madsen, 2016)

Week 07 (20/02/23)

- **Replacement class schedule TBC** *due to Isra Mikraj holiday*
- Tutorial: Exercise 3
- Canvas quiz

27/02/23 – 05/03: Mid-semester Break

No classes. Take a break!

Week 08 (06/03/23)

- Lecture: Analysis of qualitative data
- Reading: Chapter 5 (Madsen, 2016)

Week 09 (13/03/23)

- Tutorial: Exercise 4

Week 10 (20/03/23)

- Lecture: Linear relationships
- Reading: Chapter 7 (Madsen, 2016)

Week 11 (27/03/23)

- *Ramadhan month*
- Tutorial: Exercise 5
- Canvas discussion

Week 12 (03/04/23)

- *Ramadhan month*
- Lecture: Comparing grouped data
- Reading: Chapter 8 (Madsen, 2016)

Week 13 (10/04/23)

- *Ramadhan month*
- Tutorial: Exercise 6
- Canvas quiz

Week 14 (17/04/23)

- *Ramadhan month*
- **No classes scheduled**
- Assignment due

24/04/23 – 30/04: Revision Week

Revision week & Hari Raya Aidilfitri holiday

Monday 15/05/2023

- 2.10pm–4.00pm SM-1402 Exam @ Chancellor's Hall

Frequently Asked Questions (FAQ)

1. *Are there any prerequisites? What level of mathematics is expected?*

   There are no prerequisites for this module. The course is introductory level with less emphasis on statistical and probability theory. Therefore, only basic arithmetic is required. We will encounter some (possibly) new mathematical symbols and jargon, but these will be explained. A-level statistics is a bonus but not required. No calculus required.

2. *Why should I take this module?*

   No matter what your discipline, I'm certain that you will encounter a situation where you will have to deal with a data set. Perhaps you're running experiment(s) or conducting surveys. There's a reason why most programmes make it mandatory to take statistics courses (AFAIK: Biology, Economics, Business Studies, Health Sciences, Computer Science). In this digital day and age, we cannot help but be inundiated with data. So embrace it, and let's learn together!

3. *Is this an easy course?*

   This is a level-1 module with a promise of "non-heavy" mathematics, so perhaps you can make that judgement yourself. My goal is to shatter the perception that statistics and number crunching is hard, so my hope is that the module design achieves that goal.

   On the other hand, do not conflate 'easy' with 'simple'. While the statistical concepts taught are indeed simple as they are catered to non-statisticians and non-mathematicians, the course does expect you to work for your grade (as much as a 2MC module requires, which is a minimum of 6 hours of study per week). You should participate in discussions, study well for the quizzes, read the required chapters, and attend all classes.

4. *But the topics seem really easy, and I looked at the first few slides...*

   I think this module (as with every single module out there) starts very easy by drawing plots, calculating averages, etc. The turning point will be after the mid-semester break,

where you will learn many kinds of hypothesis tests and linear regression. The cool thing is that these techniques form the basis of many advanced statistical work, and you will be learning them in this module.

My advice: Don't get swayed by the "easy"ness of the first 2 chapters. Stay on track, do the work, and you'll be well prepared for the later chapters.

5. *Do I have to read the book chapters?*

   Yes.

6. *Will you teach software in class?*

   Short answer–no. The reason is that the module is (still) assessed by a 60% exam component, so it would be time better spent to prepare you well for the exam.

   Having said that, the assignment (20%) is all about applying the techniques you have learnt to an actual data set. Sure, you can use pen and paper to carry out the calculations and plots, but it would highly make sense to use software to do this.

   So the middle road is for you to self-learn the computational part using MS Excel. I think everyone has experience using MS Excel, and the book is a great resource. Just follow along the examples given.

7. *Do I have to memorise any formulae?*

   A formula sheet will be provided for the exam, along with a copy of statistical tables (χ^2, t and F distributions).