Part III — Statistics

Based on lectures by Brian

Notes taken by Dexter Chua

Lent 2017-2018

These notes are not endorsed by the lecturers, and I have modified them (often significantly) after lectures. They are nowhere near accurate representations of what was actually lectured, and in particular, all errors are almost surely mine.

Contents

1 Representation and summary of data - location

1.1 Basic Concepts of Variable

Definition (Quantitative variables and Qualitative variables)**.** Quantitative variable associated with numerical observation. Qualitative variables associated with non-numerical observations.

Definition (Continuous variable and discrete variable)**.** Continuous variable can take ant value in given range. Discrete can take only specific values in a given range.

1.2 Grouped data

Definition (Grouped data)**.** The groups are more commonly known as classes.

- class boundaries.
- mid-point of a class.
- class width.

Example. Example 5-6

Definition (Frequency and cumulative frequency)**.** Number of anything; example is how many sheeps. It is sometimes helpful to add a column to the table showing the running total of the frequencies. This is called the cumulative frequency

Definition (Ungrouped data)**.** Show all data

1.3 Mean , mode and median

Definition (Mode)**.** The mode is the value that occurs most often

Definition (Median)**.** n/2 term or 1 term above

Definition (Mean)**.**

$$\bar{x} = \frac{\sum_i^n x_i}{n}$$

1.4 Linear interpolation

Example. Example 14-15

1.5 Coding

Example. pick 1 example

2 Representation and summary of data - measures of dispersion

2.1 Range and interquartile range

The list of formula:

- Range = Upper value − Lowest value

Example. example 3

2.2 Percentiles split the data into 100 parts

Example. example 4

2.3 Range and Interquartile range

Example (Linear Interpolation)**.**

2.4 Variance and standard deviation

Definition (Variance)**.** Let f stand for the frequency, then $n = \sum f$ and

$$\text{Variance} = \frac{\sum f(x-\bar{x})^2}{\sum f} \text{ or } \frac{\sum fx^2}{\sum f} - (\frac{\sum fx^2}{\sum f})$$

2.5 Variance and standard deviation for grouped data

Definition.

Example. example 7-8

2.6 Coding

Example. example 9-11

3 Representation of data

3.1 Stem and Leaf diagrams

3.2 Outlier

Definition.

Example.

3.3 Box plot

3.4 Histogram

Definition (Frequency density)**.**

3.5 Skewness (Shape)

3.6 What!?

4 Probability

4.1 Classical Probability

4.2 Venn diagram and their rules

Definition (Complementary Probability)**.**

4.3 Conditional Probabilites

4.3.1 Vann diagram

4.3.2 Tree diagram

4.4 Special Events of Probabilites

Definition (Mutually exclusive)**.**

Definition (Independent events)**.**

5 Correlation

5.1 Correlation

5.2 Bivariate data

Definition (Co-Variance)**.**

5.3 Product moment Correlation coefficient r

5.4 Coding

6 Regression

6.1 Linear

6.2 Coding

7 Discrete random variables

7.1 Probability distribution

Definition (Variable)**.**

Definition (Expected value)**.**

7.2 Coding

8 The normal distribution

9 Binomial distribution

10 Poisson distribution

11 Continuous random variables

12 Continuous uniform distribution

13 Normal approximation

14 Population and samples

15 Hypothesis testing

16 Combination of random variables

17 Sampling

18 Estimation , confidence intervals and tests

19 Goodness of fit and contingency tables

20 Regression and correlation

21 Quality of tests and estimators

22 One-sample procedures

23 Two-sample procedures