

# 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

<b>1</b>	<b>Representation and summary of data - location</b>	<b>3</b>
1.1	Basic Concepts of Variable . . . . .	3
1.2	Grouped data . . . . .	3
1.3	Mean , mode and median . . . . .	3
1.4	Linear interpolation . . . . .	3
1.5	Coding . . . . .	3
<b>2</b>	<b>Representation and summary of data - measures of dispersion</b>	<b>4</b>
2.1	Range and interquartile range . . . . .	4
2.2	Percentiles split the data into 100 parts . . . . .	4
2.3	Range and Interquartile range . . . . .	4
2.4	Variance and standard deviation . . . . .	4
2.5	Variance and standard deviation for grouped data . . . . .	4
2.6	Coding . . . . .	4
<b>3</b>	<b>Representation of data</b>	<b>5</b>
3.1	Stem and Leaf diagrams . . . . .	5
3.2	Outlier . . . . .	5
3.3	Box plot . . . . .	5
3.4	Histogram . . . . .	5
3.5	Skewness (Shape) . . . . .	5
3.6	What!/? . . . . .	5
<b>4</b>	<b>Probability</b>	<b>6</b>
4.1	Classical Probability . . . . .	6
4.2	Venn diagram and their rules . . . . .	6
4.3	Conditional Probabilites . . . . .	6
4.3.1	Vann diagram . . . . .	6
4.3.2	Tree diagram . . . . .	6
4.4	Special Events of Probabilites . . . . .	6
<b>5</b>	<b>Correlation</b>	<b>7</b>
5.1	Correlation . . . . .	7
5.2	Bivariate data . . . . .	7
5.3	Product moment Correlation coefficient $r$ . . . . .	7
5.4	Coding . . . . .	7
<b>6</b>	<b>Regression</b>	<b>8</b>
6.1	Linear . . . . .	8
6.2	Coding . . . . .	8
<b>7</b>	<b>Discrete random variables</b>	<b>9</b>
7.1	Probability distribution . . . . .	9
7.2	Coding . . . . .	9
<b>8</b>	<b>The normal distribution</b>	<b>10</b>
<b>9</b>	<b>Binomial distribution</b>	<b>11</b>

---

10 Poisson distribution	12
11 Continuous random variables	13
12 Continuous uniform distribution	14
13 Normal approximation	15
14 Population and samples	16
15 Hypothesis testing	17
16 Combination of random variables	18
17 Sampling	19
18 Estimation , confidence intervals and tests	20
19 Goodness of fit and contingency tables	21
20 Regression and correlation	22
21 Quality of tests and estimators	23
22 One-sample procedures	24
23 Two-sample procedures	25

# 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 any 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 sheep. 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:

$$\text{Range} = \text{Upper value} - \text{Lowest value}$$

**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

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

### 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