

# 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 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
<b>2 Representation and summary of data - measures of dispersion</b>	<b>4</b>
2.1 Range and Interquartile range . . . . .	4
2.2 Variance and standard deviation . . . . .	4
<b>3 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 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 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 Regression</b>	<b>8</b>
6.1 Linear . . . . .	8
6.2 Coding . . . . .	8
<b>7 Discrete random variables</b>	<b>9</b>
7.1 Probability distribution . . . . .	9
7.2 Coding . . . . .	9
<b>8 The normal distribution</b>	<b>10</b>
<b>9 Binomial distribution</b>	<b>11</b>
<b>10 Poisson distribution</b>	<b>12</b>
<b>11 Continuous random variables</b>	<b>13</b>
<b>12 Continuous uniform distribution</b>	<b>14</b>
<b>13 Normal approximation</b>	<b>15</b>

<b>14 Population and samples</b>	<b>16</b>
<b>15 Hypothesis testing</b>	<b>17</b>
<b>16 Combination of random variables</b>	<b>18</b>
<b>17 Sampling</b>	<b>19</b>
<b>18 Estimation , confidence intervals and tests</b>	<b>20</b>
<b>19 Goodness of fit and contingency tables</b>	<b>21</b>
<b>20 Regression and correlation</b>	<b>22</b>
<b>21 Quality of tests and estimators</b>	<b>23</b>
<b>22 One-sample procedures</b>	<b>24</b>
<b>23 Two-sample procedures</b>	<b>25</b>

# 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.

**Example.** – Example of :

– Example of :

**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.

**Example.** – Example of :

– Example of :

## 1.2 Grouped data

**Definition** (Grouped data).

**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

**Example.**

**Definition** (Ungrouped data).

## 1.3 Mean , mode and median

## **2 Representation and summary of data - measures of dispersion**

### **2.1 Range and Interquartile range**

**Example** (Linear Interpolation).

### **2.2 Variance and standard deviation**

## **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