# Mathematics

*Senior 2 Part II*

MELVIN CHIA

Started on 1 January 2023

Finished on ...

# Contents

# Chapter 18

# Statistics

## 18.1 Basic Concepts

Statistics mainly study how to collect, organize, summarize, and interpret data. It is a branch of mathematics that deals with the collection, analysis, interpretation, and presentation of data. It is used to answer questions about the data and to make decisions based on the data.

### Population and Sample

In statistics, a population is the entire group of individuals that we are studying, and the units that form a population are called individuals or elements. A sample is a subset of the population. The number of elements in a sample is called the sample size. For example: select 20 of the 4,000 senior high school mathematics UEC exam papers and record their scores:

| | | | | |
|---|---|---|---|---|
| 72 | 80 | 96 | 20 | 42 |
| 75 | 60 | 92 | 18 | 53 |
| 82 | 77 | 53 | 29 | 34 |
| 57 | 79 | 82 | 90 | 41 |

Here, the population is the 4,000 scores, each of which is an element of the population. The sample is the 20 scores, the sample size is 20.

### Census and Sample Survey

The way of surveying can be divided into two types: census and sample survey. A census is a survey in which every element of the population is included in the sample. For example: national census. The data collected in a census is more accurate and reliable, but it is very expensive and time-consuming.

A sample survey is a survey in which only a part of the population is included in the sample. Researchers can use a sample survey to estimate the characteristics of the population. For example: a light bulb manufacturer produces a lot of light bulbs, thus it is impossible to test every single light bulb. The manufacturer can randomly select a sample of light bulbs and test them.

## 18.2 Data Processing

Data that are collected must be processed before they can be analyzed.

### Frequency Distribution

When the possible values of a dataset are not too many, we can use a frequency distribution table to organize the data. The frequency distribution table is a table that shows the frequency of each value in a dataset. The frequency of a value is the number of times that value appears in the dataset.

When there are too many possible values, we must group the values into classes. Before grouping the values, we must first determine the range of the values, aka the difference between the largest and smallest values, then determine the number of classes. The number of classes should be determined according to the purpose of the study and the identity of the data. After classifying the data, the range of each group is called the class interval. Typically, the class interval is the same for all classes, and must be greater than the number of classes divided by the range of the data. After the number and interval of the classes are determined, we can arrange the frequency of each class in a frequency distribution table.

Take 100 sample from a population of some kind of component, their weight (in $g$), are as below:

| | | | | | |
|---|---|---|---|---|---|
| 1.36 | 1.49 | 1.43 | 1.41 | 1.37 | 1.40 |
| 1.32 | 1.42 | 1.47 | 1.39 | 1.41 | 1.36 |
| 1.40 | 1.34 | 1.42 | 1.42 | 1.45 | 1.35 |
| 1.42 | 1.39 | 1.44 | 1.42 | 1.39 | 1.42 |
| 1.42 | 1.30 | 1.34 | 1.42 | 1.37 | 1.36 |
| 1.37 | 1.34 | 1.37 | 1.37 | 1.44 | 1.45 |
| 1.32 | 1.48 | 1.40 | 1.45 | 1.39 | 1.46 |
| 1.39 | 1.53 | 1.36 | 1.48 | 1.40 | 1.39 |
| 1.38 | 1.40 | 1.36 | 1.45 | 1.50 | 1.43 |
| 1.38 | 1.43 | 1.41 | 1.48 | 1.39 | 1.45 |

| | | | | | |
|---|---|---|---|---|---|
| 1.37 | 1.37 | 1.39 | 1.45 | 1.31 | 1.41 |
| 1.44 | 1.44 | 1.42 | 1.47 | 1.35 | 1.36 |
| 1.39 | 1.40 | 1.38 | 1.35 | 1.38 | 1.43 |
| 1.42 | 1.42 | 1.42 | 1.40 | 1.41 | 1.37 |
| 1.46 | 1.36 | 1.37 | 1.27 | 1.37 | 1.38 |
| 1.42 | 1.34 | 1.43 | 1.42 | 1.41 | 1.41 |
| 1.44 | 1.48 | 1.55 | 1.39 | | |

In the dataset above, the minimum value is 1.27 and the maximum value is 1.55.

$\therefore$ The range of the data is $1.55 - 1.27 = 0.28$.

If we classify the data into 10 classes, then the class interval must be greater than $\frac{0.28}{10} = 0.028$. Thus, we can use a class interval of 0.03.

Let the lower limit of the first class be 1.27, then the lower limit of the second class is $1.27 + 0.03 = 1.30$.

Since all the values in the dataset are of 2 decimal places, the upper limit of the first class is should be 1.29. By the same logic, we can get all the classes: $1.27 - 1.29$, $1.30 - 1.32$, $\cdots$, $1.54 - 1.56$.

Now we can arrange the data into the frequency distribution table:

| Weight $m(g)$ | Frequency |
|---|---|
| $1.27 - 1.29$ | 1 |
| $1.30 - 1.32$ | 4 |
| $1.33 - 1.35$ | 7 |
| $1.36 - 1.38$ | 22 |
| $1.39 - 1.41$ | 24 |
| $1.42 - 1.44$ | 24 |
| $1.45 - 1.47$ | 10 |
| $1.48 - 1.50$ | 6 |
| $1.51 - 1.53$ | 1 |
| $1.54 - 1.56$ | 1 |

In the example above, we assume that the weight of the components is accurate to 2 decimal places. Hence, if a component has a weight of $1.443g$, it is rounded to $1.44g$, thus it belongs to the class $1.42 - 1.44$. Hence, the actual range of the first class $1.27 - 1.29$ is $1.265 \leq m < 1.295$, written as $1.265 - 1.295$, while 1.265 and 1.295 are the boundaries of the first class, 1.265 is the lower boundary and 1.295 is the upper boundary. The mean of the lower boundary and upper boundary of a class is called the class midpoint. For example, the class midpoint of the first class is $\frac{1.265+1.295}{2} = 1.28$.

When we are analyzing the data data that have been classified into classes, the midpoint of each class is used as the representative value of the class. Thus, we should try our best to make the data-intensive place the group midpoint when choosing the class interval and boundaries, so that the data can be analyzed more precisely.

The distribution of frequency can be represented by a histogram or a frequency polygon.

The histogram is a row of continuous bars, the bottom side of each bar on the x-axis. For unclassified data, the bottom side of each bar is marked with the values, while the height of each bar is the frequency of the corresponding value. For classified data, the bottom side of each bar is marked with the boundaries of the corresponding class, while the area of each bar must be proportional to the frequency of the corresponding class. When the class interval of each class is the same, we can use the frequency of each class as the height of the bar.

The frequency polygon is a continuous line graph, the x-axis is the midpoint of each class, and the y-axis is the frequency of each class. To draw a frequency polygon, we plot each point, including the point before the first class and the point after the last class that uses 0 as their frequency, and then connect the points with a continuous line.

### Accumulative Frequency Distribution

## 18.3 Central Tendency

## 18.4 Measures of Dispersion

## 18.5 Coefficient of Variation

## 18.6 Correlation and Correlation Coefficient

## 18.7 Statistical Index

# Chapter 19

# Permutations and Combinations

# Chapter 20

# Bionomial Theorem

## 20.1 Bionomial Theorem when $n$ is a Natural Number

## 20.2 General Form of Bionomial Expansion

# Chapter 21

# Probability