CEG2002: Statistics and Numerical Methods for Civil Engineers (Part I: Statistics)

Dr. Jian Shi
Room 2.29, Level 2, Herschel Building
Email: j.q.shi@ncl.ac.uk
http://www.staff.ncl.ac.uk/j.q.shi/teaching/CEG2002

Semester 2, 2011/2012

- This course includes two parts:
  - Part I (weeks 1-5, Dr Jian Qing Shi): Statistics.
  - Part II (weeks 6-10, Dr David Swailes): Numerical Methods.
  - Revision (Week 11: Numerical Methods; Week 12: Statistics)
- Assessment

Assignment/project	20% (10% for each part).
Exam (2hrs)	80%

Arrangement for part I

- Lectures: Monday 11am-12pm (STB F13), Wednesday 10am-11am (STB. F16)
- Tutorials: Tuesday 5pm-6pm (Weeks 3 and 5, STB F13)
- Computer Practical: Monday 1pm-2pm (Weeks 2, 4 and 6, STB Tree, room 3.04C)
- Office hours: Monday 12pm-1pm and Wednesday 11am-12pm (room 2.29, Herschel Building)
- References
  - Montgomery, D. C. and Runger, G. C. (2006). Applied Statistics and Probability for Engineers (4th edition). Wiley.
  - Ross, S. M. (2004). Introduction to Probability and Statistics for Engineers and Scientists (3rd edition). Academic Press.

What is statistical analysis?

A catastrophic example: Venezuela, 1999

On December 23rd, 1999, a **daily precipitation event of more than 410mm**, almost 3 times the magnitude of the previously recorded maximum, occurred in Caracas, Venezuela.

The flooding which followed caused **50,000 deaths** and an estimated **£5.5 billion worth of damage**.

Flood defences were designed, using probability, to withstand the "**one in one hundred year**" flood event – estimated at a level of 350mm.

Estimated level of 350mm v.s. the observed level of 410mm. If a better statistical model is used, efforts could have been made to strengthen flood defences.

Statistical Analysis

- Define a random variable: e.g X-daily rainfall level;
- To find a distribution (or model), e.g. $X \sim N(\mu, \sigma^2)$;
- To collect a set of data to estimate unknown parameters e.g. (artificial) $\hat{\mu} = 220, \hat{\sigma} = 40$.
- Conclusion/interpretation:

$$P(X > 350) = 0.000577 = 1/1733(days) = 1/5(years)$$

This is not a rare event!

- The old flood defence system was based on the assumption that

$$P(X > 350) = 1/1000(years) = 1/365,000(days) = 2.74 \times 10^{-6},$$

which is very **wrong**!

- Possible reasons:
  - The old model is **wrong** or the data was **biased**;
  - Need a new model due to globe warming.

Statistical Analysis

- Statistical analysis:
  - Problem (using statistical language)
  - Modelling
  - Data collecting
  - Inference (e.g. parameter estimating or hypothesis testing)
  - Conclusion/interpretation
- Some related topics
  - How to collect data?
  - Is the model a good fit for the data?

Contents

Chap 1. Collecting, Presenting and Summarising data

In this chapter, we will focus on **data** and discuss the following topics.

- Definitions and types of data
- Sampling
- Frequency tables
- Graphical summaries
  - Stem–and–leaf plots
  - Bar charts
  - Histograms
  - Time series plots
  - Scatterplots
  - Box–and–whisker plots
- Numerical summaries
  - Measure of location: mean, median and mode
  - Measure of spread: variance (standard deviation), range, inter-quartile range
  - Numerical summaries linked to order statistics: quartiles

Chap 1. Collecting, Presenting and Summarising data

1.1 Introduction

Data are the key to many important engineering decisions:

- Is a sea–wall high enough to withstand extreme sea levels which are expected to occur once every fifty years?
- Or once every five hundred years?
- Is such a defence *strong enough*?
- What are the design requirements for a lift which should be able to carry, on average, 17 people?

These, and many other questions, can be answered with **data**.

We begin this course by looking at some basic methods of **collecting**, **representing** and **describing** data.

Definitions

The quantities measured in a study are called **random variables**.

A particular outcome is called an **observation**.

A collection of observations is the **data**.

The collection of all possible outcomes is the **population**.

Example

If we were interested in the height of people doing Civil Engineering courses at Newcastle,

- our **random variable** would be X: height of people doing Civil Engineering at Newcastle
- an **observation** would be a particular person's height
- if we measured everyone doing CEG2002, those would be our **data**
- these data would form a **sample** from the **population** of all students registered with the School of Civil Engineering.

In practice it is difficult to observe whole populations. We usually observe a subset of the population (i.e. a set of data or a sample); the data will be used to estimate the distribution of the whole population.

Types of data

Variables are usually one of two types: **discrete** or **continuous**.

Discrete random variables

- can only take a sequence of distinct values which are usually integers
- are **countable**
- **Examples:** number of defective pieces in a manufacturing batch, the number of people in a tutorial group, a person's shoe size, ...
- There are other kinds of discrete data. **Ordinal** data (data which are ordered) usually take numeric values, but these values are not really numbers in the usual sense.

Continuous random variables

- can take any value over some continuous scale
- can be measured to differing degrees of accuracy using different equipment; but we can never say *absolutely*, *precisely*, how much someone weighs (for example)
- are often expressed up to a number of significant digits and could appear to be discrete
- **Examples:** height, weight, wind speed, compressive strength, or the fuel consumption of a car

It is the underlying variable which defines their status and not the form in which they are expressed.

1.2 Sampling

We can rarely observe the whole **population**. Instead, we observe some sub–set of this called the **sample**. The difficulty lies in obtaining a **representative** sample.
How to obtain a good sample? Data should be

- random
- unbias (double blined)

For example, if you were to ask the people leaving a gym if they took exercise this would produce a **biased** sample and would not be representative of the population as a whole.

The importance of obtaining a representative sample can not be stressed too highly.

Different sampling methods

- Simple Random Sampling
- Stratified Sampling
- Systematic Sampling
- Cluster Sampling
- Judgemental sampling

1.3 Frequency tables

The following table presents the modes of transport used daily by 30 students to get to and from University.

Student	Mode	Student	Mode	Student	Mode
1	Car	11	Walk	21	Walk
2	Walk	12	Walk	22	Metro
3	Car	13	Metro	23	Car
4	Walk	14	Bus	24	Car
5	Bus	15	Train	25	Car
6	Metro	16	Bike	26	Bus
7	Car	17	Bus	27	Car
8	Bike	18	Bike	28	Walk
9	Walk	19	Bike	29	Car
10	Car	20	Metro	30	Car

The table obviously contains much information. However, it is difficult to see which method of transport is the most widely used.

Idea: count the number of students using each mode of transport!

Mode	Frequency
Car	10
Walk	7
Bike	4
Bus	4
Metro	4
Train	1
Total	**30**

This gives us a much clearer picture of the methods of transport used.

Relative frequency

Also of interest might be the **relative** frequency of each of the modes of transport.

This is simply the frequency expressed as a proportion of the total number of students surveyed. If this is given as a percentage, as here, this is known as the **percentage relative frequency**.

Mode	Frequency	Relative Frequency (%)
Car	10	33.3
Walk	7	23.4
Bike	4	13.3
Bus	4	13.3
Metro	4	13.3
Train	1	3.4
Total	**30**	**100**

continuous data

Example The following data set represents rainfall totals (in mm) for 50 recording stations across the Deep South over a 48 hour period during Hurricane Katrina:

214.8412	220.6484	216.7294	195.1217	211.4795
195.8980	201.1724	185.8529	183.4600	178.8625
196.3321	199.7596	206.7053	203.8093	203.1321
200.8080	201.3215	205.6930	181.6718	201.7461
180.2062	193.3125	188.2127	199.9597	204.7813
198.3838	193.1742	204.0352	197.2206	193.5201
205.5048	217.5945	208.8684	197.7658	212.3491
209.9000	197.6215	204.9101	203.1654	192.9706
208.9901	202.0090	195.0241	192.7098	219.8277
208.8920	200.7965	191.9784	188.8587	206.8912

This is what we do...

1. Divide the range of the variable into smaller ranges called **class intervals**
2. There should be **no gaps** between these intervals
3. The class interval width should be a convenient number (e.g. 5, 10, 100, depending on the data)
4. You should aim for no more than about **ten to fifteen classes**

Continuous data frequency tables

For the Deep South data, we might get:

Class Interval	Tally	Frequency	Relative Frequency %
175–179.9999	|	1	2
180–184.9999	|||	3	6
185–189.9999	|||	3	6
220–224.9999			
Totals		**50**	**100**

1.4 Summarising data– Graphical summaries

So far, we have looked at

- how **data** can differ
- how to **sample** from the **population**
- how to display data in **tabular form**
  - frequency tables
  - *relative* frequency tables

We now look at how to **summarise** data

- **graphically** and
- **numerically**.

1.4 Graphical summaries

We will consider six different methods of summarising data **graphically**:

- Stem–and–leaf plots
- Bar charts
- Histograms
- Time series plots
- Scatterplots
- Box–and–whisker plots

We will look at the first five of these now, and consider the last of these – the **box–and–whisker plot** – towards the end of the chapter.

1.4.1 Stem–and–leaf plots

Stem and leaf plots are a quick and easy way of representing data graphically.

They can be used with both **discrete** and **continuous** data.

The easiest way to describe how such a plot is constructed is via demonstration...

...but first, you need to think about the following...

- You need to decide on a reasonable number of **intervals** which span the range of the data
- These interval widths must be **equal**
- You should use **sensible** values for the interval widths

Construct a stem and leaf plot for the following data:

11 12 9 15 21 25 19 8

- Put the data in ascending order:

8 9 11 12 15 19 21 25

- Need to decide on an **interval width**
  Idea: try 10 – i.e. 0–9, 10–19 and 20–29

0	**8 9**
1	**1 2 5 9**
2	**1 5**

Stem Leaf

$n = 8$ stem unit $= 10$, leaf unit $= 1$.

- Findings:

Example 2: Dam flows

Consider the following data, which are the total monthly flows (October 2006), in cubic metres per second, through the spillways of ten dams in southern India:

17 18 15 14 12 19 20 21 24 15

If you were to choose 10 as the interval width (i.e. go up in 10s), the stem and leaf plot would look like

1	2 4 5 5 7 8 9
2	0 1 4

Stem Leaf

$n = 10$, stem unit $= 10$, leaf unit $= 1$.

- The interval width is **too large**!
- Two intervals is not enough to reveal any **patterns** in the data

Idea: Use an interval width of 5!

1	2 4
1	5 5 7 8 9
2	0 1 4

Stem **Leaf**

$n = 10$, stem unit $= 5$, leaf unit $= 1$.

We can see the pattern more clearly now!

1.4.2 Bar charts

Bar charts are a commonly–used and clear way of presenting categorical data

- As with stem and leaf plots, various computer packages allow you to produce these
- First, let us work through the process of producing these by *hand*.

Example

Recall the example on students' modes of transport:

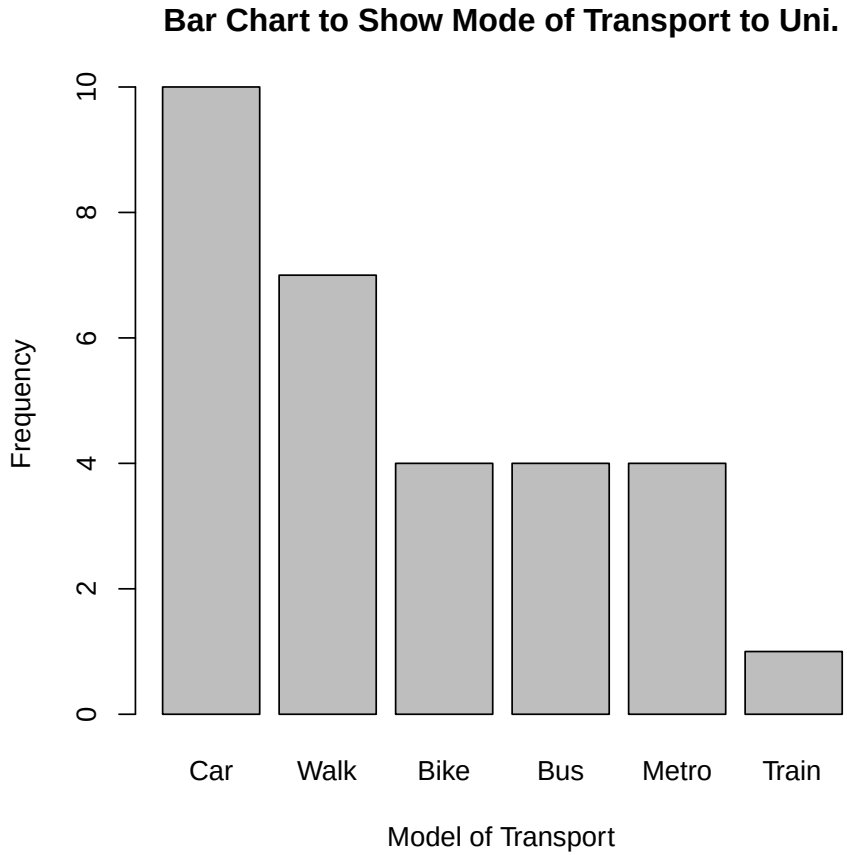
Student	Mode	Student	Mode	Student	Mode
1	Car	11	Walk	21	Walk
2	Walk	12	Walk	22	Metro
3	Car	13	Metro	23	Car
4	Walk	14	Bus	24	Car
5	Bus	15	Train	25	Car
6	Metro	16	Bike	26	Bus
7	Car	17	Bus	27	Car
8	Bike	18	Bike	28	Walk
9	Walk	19	Bike	29	Car
10	Car	20	Metro	30	Car

Draw a bar chart to represent these data

The first logical step is to put these into a frequency table, giving

Mode	Frequency
Car	10
Walk	7
Bike	4
Bus	4
Metro	4
Train	1
Total	30

We can then present this information as a bar char.


Bar Chart to Show Mode of Transport to Uni.
Frequency
0
2
4
6
8
10
Car
Walk
Bike
Bus
Metro
Train
Model of Transport

1.4.3 Histograms

Bar charts have their limitations, one of which is that they cannot be used to present **continuous data**.

When dealing with continuous random variables a different kind of graph is required – one such graph is the **histogram**.

Example

Consider the following data, which show rainfall totals (in mm) for fifty recording stations across the Deep South over a 48 hour period during Hurricane Katrina in August 2005.

214.8412	220.6484	216.7294	195.1217	211.4795
195.8980	201.1724	185.8529	183.4600	178.8625
196.3321	199.7596	206.7053	203.8093	203.1321
200.8080	201.3215	205.6930	181.6718	201.7461
180.2062	193.3125	188.2127	199.9597	204.7813
198.3838	193.1742	204.0352	197.2206	193.5201
205.5048	217.5945	208.8684	197.7658	212.3491
209.9000	197.6215	204.9101	203.1654	192.9706
208.9901	202.0090	195.0241	192.7098	219.8277
208.8920	200.7965	191.9784	188.8587	206.8912

Produce a histogram to display these data graphically

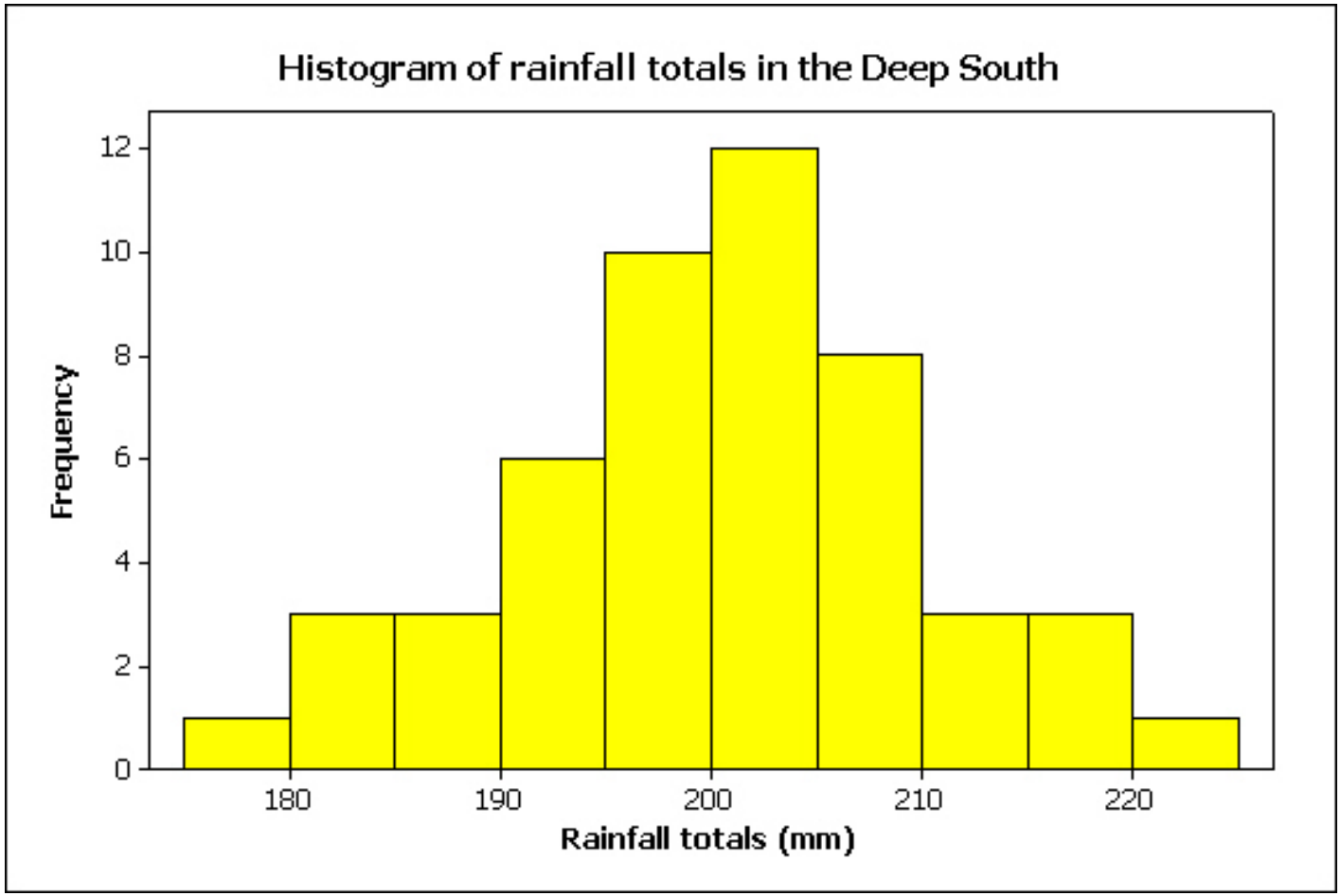
Producing a histogram is much like producing a bar chart.

It is often best to produce a frequency table first which collects all the data together in an ordered format.

Rainfall totals x	**Frequency**
$175 \leq x < 180$	1
$180 \leq x < 185$	3
$185 \leq x < 190$	3
$190 \leq x < 195$	6
$195 \leq x < 200$	10
$200 \leq x < 205$	12
$205 \leq x < 210$	8
$210 \leq x < 215$	3
$215 \leq x < 220$	3
$220 \leq x < 225$	1
Total	**50**

Once we have the frequency table, the process is very similar to before...

- Find the maximum frequency and draw the **vertical** (y–axis) from zero to this value, including a sensible numeric scale
- The range of the **horizontal** (x–axis) needs to include not only the full range of observations but also the full range of the class intervals from the frequency table.
- Draw a bar for each group in your frequency table. These should be the same width and touch each other


Histogram of rainfall totals in the Deep South
Frequency
0
2
4
6
8
10
12
180
190
200
210
220
Rainfall totals (mm)

Findings:

1.4.4 Percentage relative frequency histograms

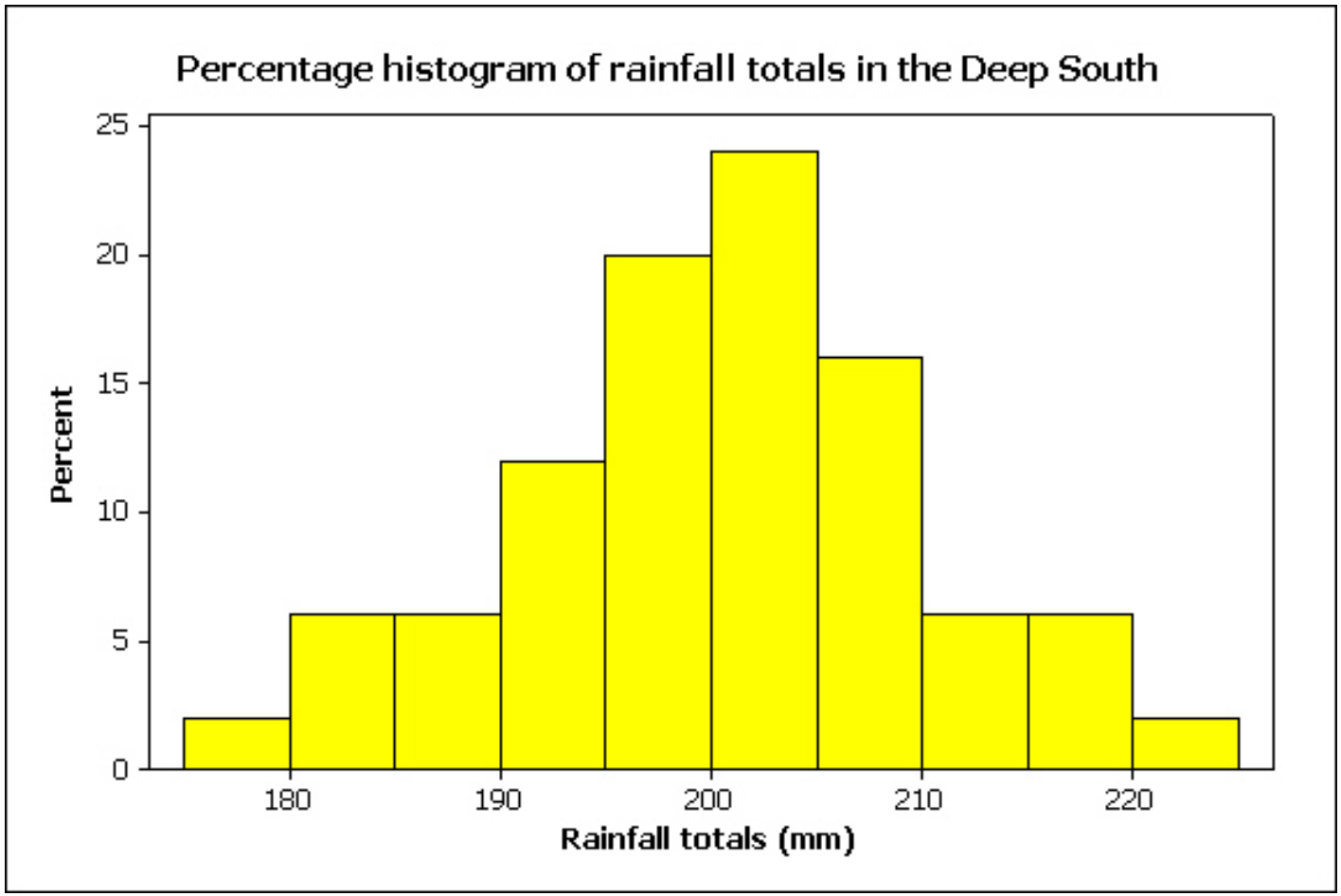
These are an extension to the **percentage relative frequency tables** we have already considered.

To produce a percentage relative frequency histogram, we follow the same procedure as for a regular histogram, but use *percentage frequencies* instead of just *frequencies*.

A percentage relative frequency table for the Deep South data is:

Rainfall totals x	Frequency	Relative Frequency (%)
$175 \leq x < 180$	1	2
$180 \leq x < 185$	3	6
$185 \leq x < 190$	3	6
$190 \leq x < 195$	6	12
$195 \leq x < 200$	10	20
$200 \leq x < 205$	12	24
$205 \leq x < 210$	8	16
$210 \leq x < 215$	3	6
$215 \leq x < 220$	3	6
$220 \leq x < 225$	1	2
Totals	**50**	**100**

Instead of using **frequency** on the vertical axis (y-axis), you *could* use the **percentage relative frequency**.


Percentage histogram of rainfall totals in the Deep South
Percent
25
20
15
10
5
0
180
190
200
210
220
Rainfall totals (mm)

1.4.5 Time Series Plots

So far we have only considered data where we can **ignore the order** in which the data come.

Not all data are like this ...

... one exception is data which have been collected over **time**.

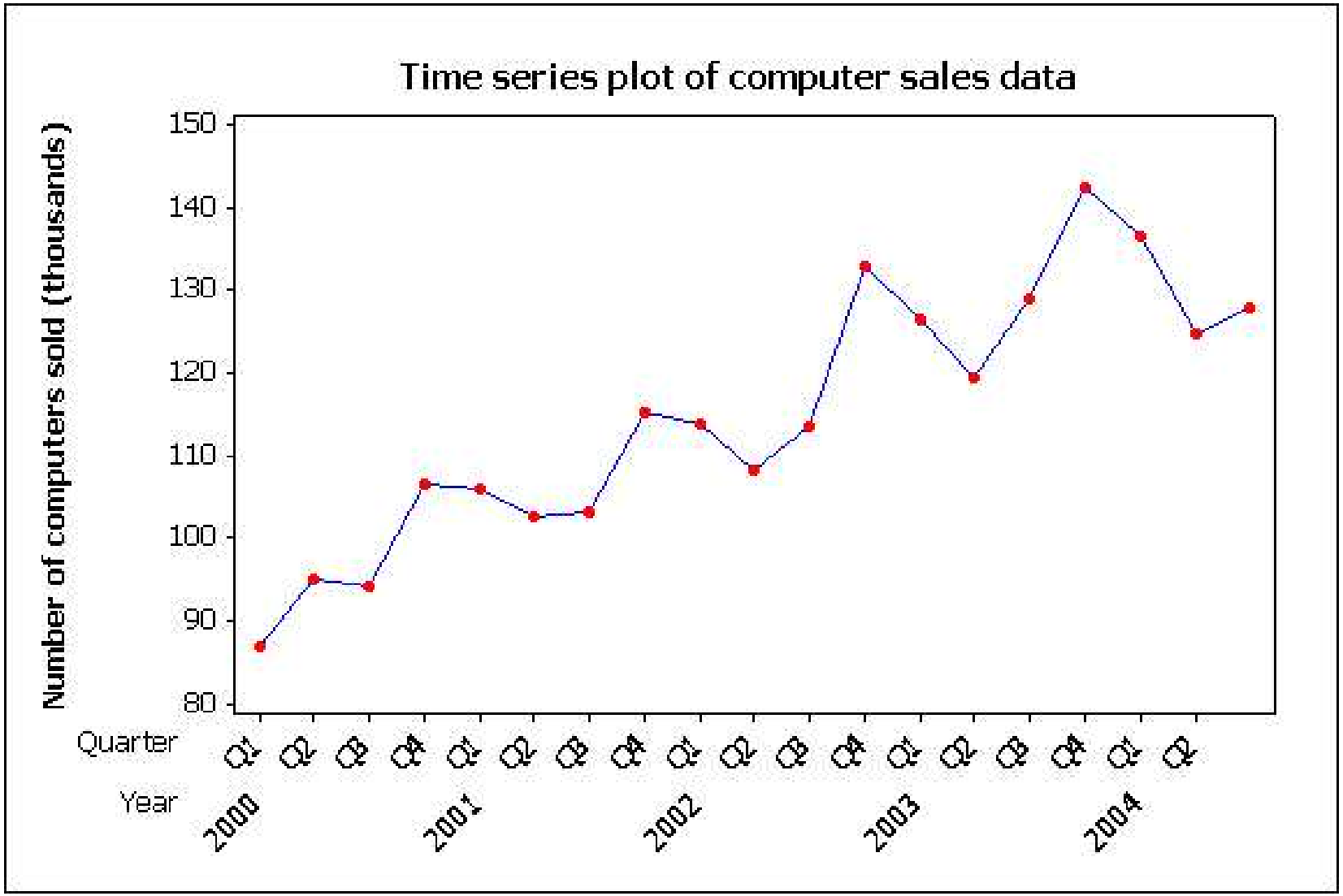
- Monthly sales of a product
- The price of a share at the end of each day
- The air temperature at midday each day

Such data can be plotted simply using **time** as the x-axis.

Example

Consider the following data on the number of computers sold (in thousands) by quarter at a large warehouse outlet.

Quarter	Units Sold	Quarter	Units Sold
Q1 2000	86.7	**Q1** 2002	113.7
Q2 2000	94.9	**Q2** 2002	108.0
Q3 2000	94.2	**Q3** 2002	113.5
Q4 2000	106.5	**Q4** 2002	132.9
Q1 2001	105.9	**Q1** 2003	126.3
Q2 2001	102.4	**Q2** 2003	119.4
Q3 2001	103.1	**Q3** 2003	128.9
Q4 2001	115.2	**Q4** 2003	142.3
		Q1 2004	136.4
		Q2 2004	124.6
		Q3 2004	127.9


Time series plot of computer sales data
Number of computers sold (thousands)
150
140
130
120
110
100
90
80
Quarter
Year
Q1 Q2 Q3 Q4 Q1 Q2 Q3 Q4 Q1 Q2 Q3 Q4 Q1 Q2 Q3 Q4 Q1 Q2
2000
2001
2002
2003
2004

Findings:

1.4.6 Scatter plots

The final type of graph we are going to look at (for now!) is the **scatter plot**.

Such graphs are used to plot two variables which you believe might be **related** – for example:

- height and weight
- advertising expenditure and sales
- age of machinery and maintenance costs

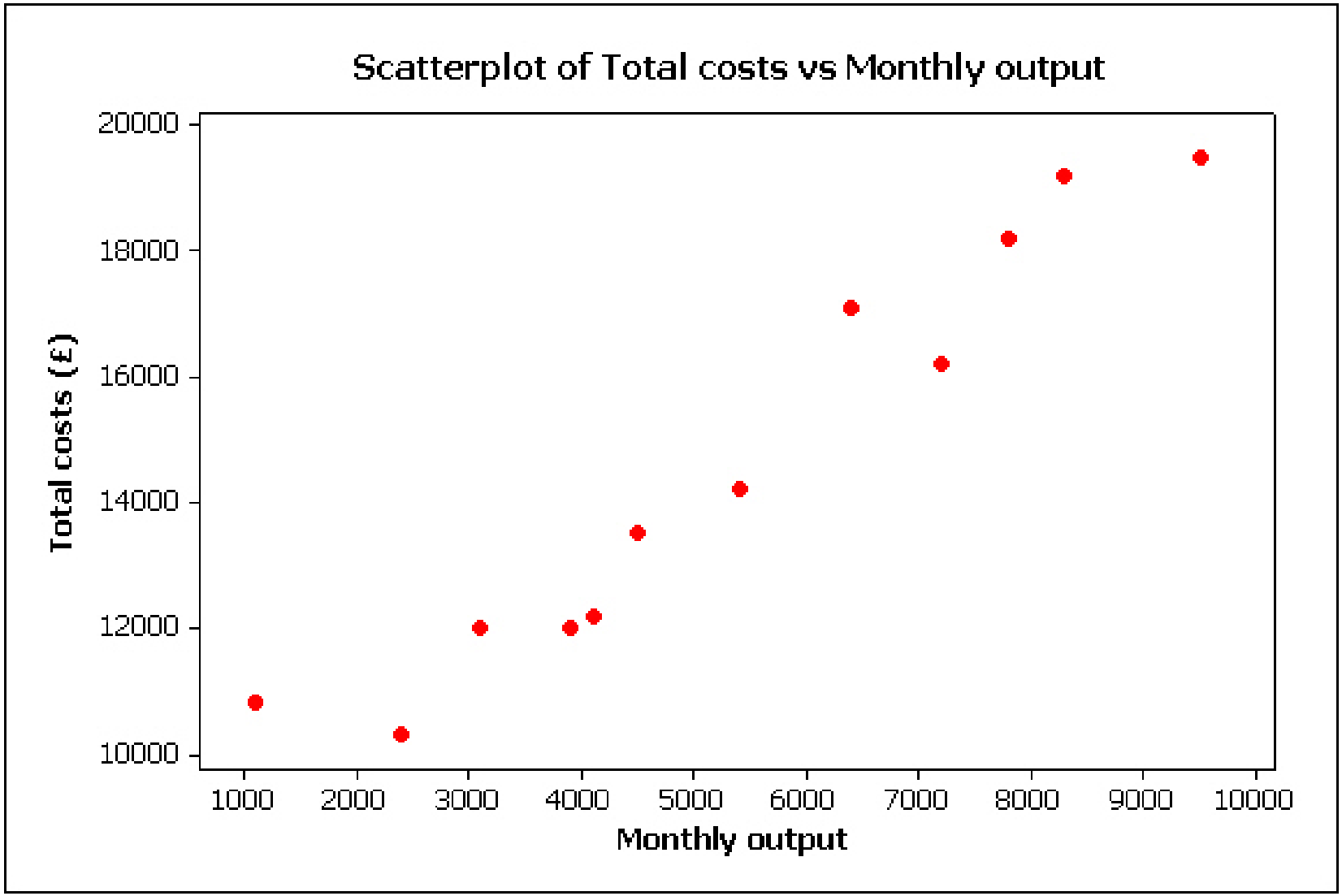
Example

Consider the following data for total costs and monthly output at a factory.

Total costs (£)	Monthly Output
10300	2400
12000	3900
12000	3100
13500	4500
12200	4100
14200	5400
10800	1100
18200	7800
16200	7200
19500	9500
17100	6400
19200	8300

If you were interested in the relationship between the cost of production and the number of units produced you could easily plot this by hand.

1. The "**response**" variable is placed on the y-axis. Here the response variable is "total costs"
2. The variable that is used to try to explain the response variable (here, monthly output) is placed on the x-axis – this is the **explanatory** variable
3. Plot the pairs of points on the graph


Scatterplot of Total costs vs Monthly output
Total costs (£)
20000
18000
16000
14000
12000
10000
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
Monthly output

Findings:

1.5 Summarising data – Numerical summaries

If you summaries data numerically, it is essential that you calculate both

- a measure of **location** (or "average"), and
- a measure of **spread**

There are three measures of **location** which are commonly used:

1. the **mean**,
2. the **median** and
3. the **mode**.

We will consider these in turn.

1.5.1 The (arithmetic) mean

- The **arithmetic mean** is the most commonly used measure of location. We often refer to it as the **average** or just the **mean**.
- So if our data were **10**, **12**, and **14**, then our mean would be

$$\frac{10 + 12 + 14}{3} = \frac{36}{3} = 12.$$

- In general, if we have a set of observations (**sample**) $x_1, \ldots, x_n$, the **sample mean** is calculated using the formula

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}.$$

- For small data sets this is easy to calculate by hand, but is simplified by using the statistics (**SD**) mode on a University approved calculator.

Sometimes we might not have the **raw data**; instead, the data might be available in the form of a **table**. For example:

Cars Sold $(x_{(j)})$	**Frequency** (f_j)
4	1
5	1
6	2
7	2
8	3
9	2
10	2
11	1
Total (n)	14

The sample mean can be calculated from these data as

$$\bar{x} = \frac{4 \times 1 + 5 \times 1 + 6 \times 2 + \ldots + 11 \times 1}{14} = 7.71.$$

1.5.2 The Median

The median is occasionally used instead of the mean, particularly when the data are **asymmetric**.

The median is the **middle value** of the observations when they are listed in ascending order:

$$\text{median} = \left(\frac{n+1}{2}\right)^{th} \text{ smallest observation.}$$

How to find **median**

- For example, the data is $\{3, 5, 3, 2, 9, 7, 6\}$,
- Always put the data in **ascending order** first!

$$2 \quad 3 \quad 3 \quad \mathbf{5} \quad 6 \quad 7 \quad 9$$

If n is an odd number, the median is the

$$\frac{7+1}{2} = 4^{th} \text{ smallest observation},$$

i.e. median $= 5$.

- If the sample size (n) is an **even** number, for example:

$$2 \quad 3 \quad 3 \quad \mathbf{5} \quad 6 \quad 7 \quad 9 \quad 10$$

$$\text{median} \quad = \quad \left(\frac{8+1}{2}\right)^{th} \text{ smallest observations}$$

The median = average of 4th and 5th smallest observations = 5.5.

Comparing mean with median

- For the following sample:

$$2 \quad 3 \quad 3 \quad 5 \quad 6 \quad 7 \quad 9 \quad 10$$

$$\text{mean} = 5.625, \quad \text{median} = 5.5$$

- But, for the sample:

$$2 \quad 3 \quad 3 \quad 5 \quad 6 \quad 7 \quad 9 \quad 100$$

$$\text{mean} = 16.875, \quad \text{median} = 5.5$$

- **Median** is more robust! When there are outliers (or when the distribution of the data is not symmetrical), we usually use median, e.g. income;
- but if we are interested in those extreme values, we should be careful...

1.5.3 The Mode

It is the value of the random variable in the sample which occurs with the **highest frequency**.

It is usually found by **inspection**.

For **discrete** data this is easy. The mode is simply the most common value. On a bar chart, it would be the category with the highest bar.

2.2.4 Measures of spread

A measure of **location** is insufficient in itself to summarise data as it only describes the value of a typical outcome.

For example:

Sample 1	**6**	**22**	**38**	$\bar{x} = 22$	median $= 22$
Sample 2	**21**	**22**	**23**	$\bar{x} = 22$	median $= 22$

Both samples have the same measures of average. But they are clearly very different samples!

The mean or the median does not fully represent the data.

There are three basic **measures of spread** which we will consider:

1. the **sample variance** — average of the squared distance between each observation and the mean,
2. the **range** ($Max - Min$), and
3. the **inter–quartile range** ($Q3 - Q1$).

The Sample Variance and Standard Deviation

- The **sample variance** is the standard measure of spread used in statistics.
- It is usually denoted by s^2 and is the **average of the squared distances** of the observations from the sample mean.
- We use the formula...

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$

- Or equivalently as

$$s^2 = \frac{1}{n-1}\left\{\sum_{i=1}^{n} x_i^2 - n(\bar{x})^2\right\}$$

- s is called the **sample standard deviation**, which is the square root of the sample variance.

Example Consider the following example($n = 20$)

8.4	8.7	9.0	9.0	9.2	9.3	9.3	9.5	9.6	9.6
9.6	9.7	9.7	9.9	10.3	10.4	10.5	10.7	10.8	11.4

The sample mean is

$$\bar{x} = (8.4 + 8.7 + \ldots + 11.4)/20 = 9.73.$$

Now

$$\sum x^2 = 8.4^2 + 8.7^2 + \cdots + 11.4^2 = 1904.38$$
$$n(\bar{x})^2 = 1893.458$$

and so the sample variance is

$$s^2 = \frac{1}{19}(1904.38 - 1893.458) = 0.57484$$

and the sample standard deviation is

$$s = \sqrt{s^2} = \sqrt{0.57484} = 0.75818.$$

Numerical summaries

The followings are **Five** numerical summaries linked to order statistics

- Min and Max.
- The 1st quartile (**lower quartile**, $Q1$) has 25% of the data below it

$$Q1 = \frac{(n+1)}{4}\text{th smallest observation}$$

- The 2nd quartile (**median**, $Q2$) has 50% of the data below it
- The 3rd quartile (**upper quartile**, $Q3$) has 75% of the data below it

$$Q3 = \frac{3(n+1)}{4}\text{th smallest observation}$$

Consider again the data $n = 20$:

8.4	8.7	9.0	9.0	9.2	9.3	9.3	9.5	9.6	9.6
9.6	9.7	9.7	9.9	10.3	10.4	10.5	10.7	10.8	11.4

- the lower quartile is the $(n+1)/4 = 5\,\frac{1}{4}$th smallest observation
  Here the 5th and 6th smallest observations are 9.2 and 9.3.
  Therefore, the **lower quartile** is $Q1 = 9.225$.
- Similarly, the **upper quartile** is the $3(n+1)/4 = 15\,\frac{3}{4}$th smallest observation, that is, three quarters of the way between 10.3 and 10.4; so $Q3 = 10.375$.

Numerical summaries and box plot

Range and inter–quartile range

- Range $= Max - Min = 11.4 - 8.4 = 3$.
- The **inter–quartile range** is simply the difference between the upper and lower quartiles, that is

$$IQR = Q3 - Q1 = 10.375 - 9.225 = 1.15.$$

- **Range** is unduly influenced by extreme observations or outliers.
- The **inter–quartile range** is a robust statistic.
- Box and Whisker Plot.

Box and Whisker Plots

Box and whisker plots are another graphical method for displaying data

They are particularly useful in highlighting differences between groups

These plots use some of the key summary statistics we have looked at earlier, the **quartiles**, as well as the **maximum** and **minimum**.

The plot is constructed as follows:

- Lay out an x–axis for the full range of the data
- Draw a rectangle with ends at the the **upper** and **lower quartiles** (the "box")
- Split the rectangle in two using the **median**
- Draw lines from the "box" to the **minimum** and **maximum** values (the "whiskers")

Draw a Box and Whisker plot for data with the following summaries:

Minimum	$min = 10$
Lower quartile	$Q1 = 40$
Median	$Q2 = 43$
Upper quartile	$Q3 = 45$
Maximum	$max = 50$