



MAS1403

Quantitative Methods for Business Management

Semester 1, 2017–2018

Module leader: Dr. Lee Fawcett

Additional lecturers: Dr. Dave Walshaw and Dr. Ged Cowburn

Tutorials

- This week tutorials are replaced with **computer practicals**
- Days/times:

Group	Day	Time
A	Tues	10–11
B	Wed	10–11
C	Thurs	11–12

- All sessions take place in the Herschel PC cluster
- You will be taught how to use the “NUMBAS” software to access the online **Computer Based Assessments**
- Your first assessed assignment, using NUMBAS, goes live next week – so attendance at this week’s practical session is vital!

Lecture 2

GRAPHICAL METHODS FOR PRESENTING DATA

So far, we have looked at

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

We now look at how to display data **graphically**

2.1 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 (same as for frequency tables last week)
- These interval widths must be **equal**
- You should use **sensible** values for the interval widths

2.1 Stem-and-leaf plots

Construct a stem-and-leaf plot for the following data:

11 12 9 15 21 25 19 8

We need to decide on an **interval width** – **idea: try 10** – i.e.

- 0–9
- 10–19
- 20–29

2.1 Stem-and-leaf plots

This gives a **stem unit** of 10 and a **leaf unit** of 1:

0		8	9		
1		1	2	5	9
2		1	5		

Stem Leaf

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

2.1 Stem-and-leaf plots

Some notes...

- “Stem” units are to the left of the line, “leaves” to the right
- For example, for the first observation – **11** – we put a “1” in the stem (one ten) and a “1” as the first leaf
- Each leaf must be equally spaced along the row
- It’s like a **bar chart** sideways! (see later)
- But better – it contains all the **raw observations**
- Put the data in **ascending order** first!

Example 1

The observations in the table below are the recorded time it takes to get through to an operator at a telephone call centre (in seconds).

54	56	50	67	55	38	49	45	39	50
45	51	47	53	29	42	44	61	51	50
30	39	65	54	44	54	72	65	58	62

2		9											
3		0	8	9	9								
4		2	4	4	5	5	7	9					
5		0	0	0	1	1	3	4	4	4	5	6	8
6		1	2	5	5	7							
7		2											

Stem Leaf

$n = 30$, stem unit = 10, leaf unit = 1

Extra example

If there is more than one significant figure in the data, the extra digits are **cut**, not **rounded** – e.g. 2.97 would be “cut” to 2.9.

Consider the following data on lengths of items on a production line (in cm):

2.97	3.81	2.54	2.01	3.49
3.09	1.99	2.64	2.31	2.22

The stem and leaf plot for this is as follows:

1		9		
2		0	2	3
2		5	6	9
3		0	4	
3		8		

$n = 10$, stem unit = 1 cm, leaf unit = 0.1 cm.

2.2 Bar charts

Bar charts are a commonly-used and clear way of presenting categorical data or any ungrouped discrete 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 2

The following frequency table represents the modes of transport used daily by 30 students to get to 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

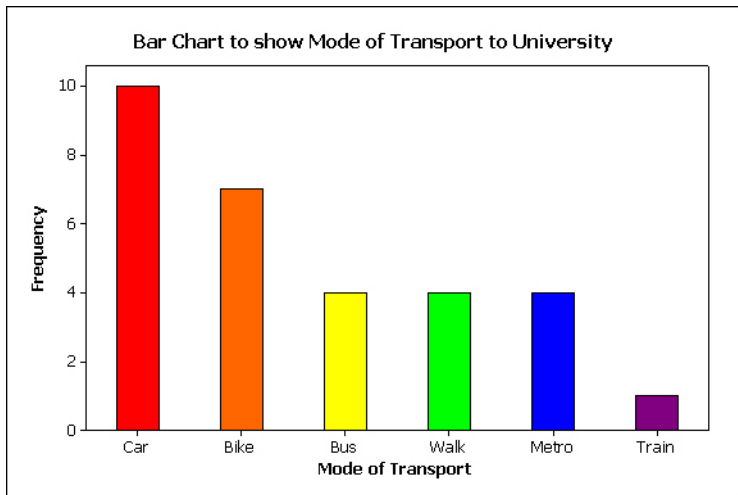
Example 2

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 easily present this information as a bar chart:

Example 2



2.2 Extension: Multiple bar charts

The data below gives the daily sales of CDs (in £) by music type for an independent retailer.

Day	Chart	Dance	Rest	Total
Monday	12000	10000	2700	24700
Tuesday	11000	8000	3000	22000
Wednesday	9000	6000	2000	17000
Thursday	10000	5000	2500	17500
Friday	12000	11000	3000	26000
Saturday	19000	12000	4000	35000
Sunday	10000	8000	2000	20000
Total	83000	60000	19200	162200

2.2 Extension: Multiple bar charts

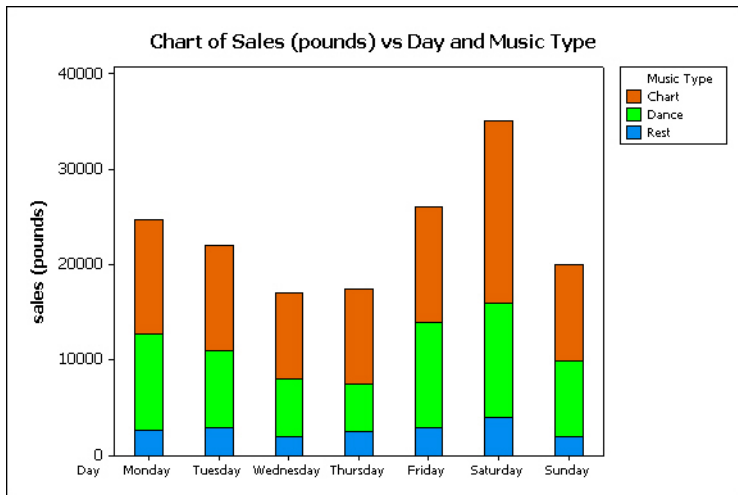
Bar charts could be drawn of **total sales per music type** in the week...

...or of **total daily sales**...

It might be interesting to see **daily sales broken down into music types!**

This can be done using a **multiple bar chart**.

2.2 Extension: Multiple bar charts



2.3 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**.

At first sight these look similar to bar charts. There are, however, some critical differences:

- The horizontal (x -axis) is a *continuous scale*
- As a result there are *no gaps between the bars*
- Strictly speaking, the *area* of the rectangle is proportional to the frequency – not the height...
- ... however, most computer packages will automatically use equal class widths, meaning the resulting histogram would look the same on a “frequency” or “frequency density” scale anyway

Example 3

Consider the following data, which show service times (in seconds) for a telephone call centre.

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

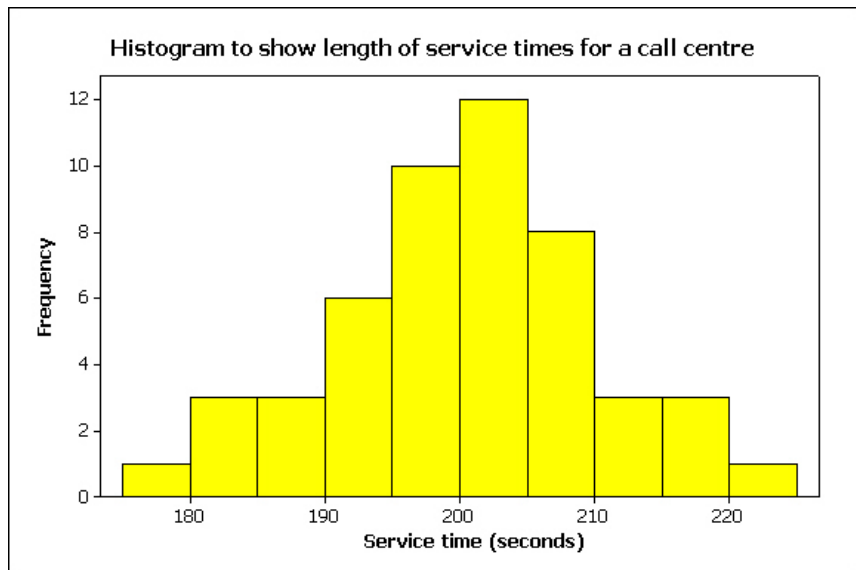
Example 3

Producing a histogram is much like producing a bar chart.

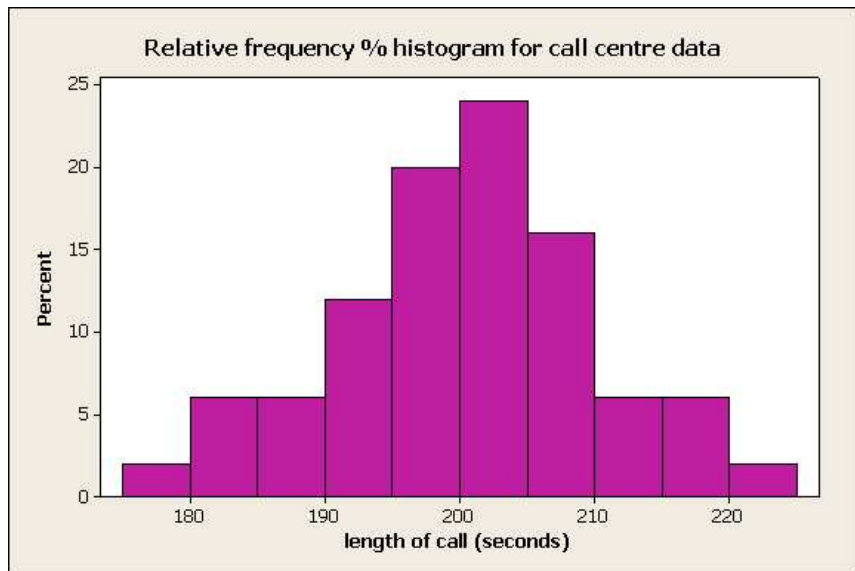
It is often best to produce a frequency table first which collects all the data together in an ordered format.

Service time	Frequency	Rel. Freq. (%)
$175 \leq \text{time} < 180$	1	2
$180 \leq \text{time} < 185$	3	6
$185 \leq \text{time} < 190$	3	6
$190 \leq \text{time} < 195$	6	12
$195 \leq \text{time} < 200$	10	20
$200 \leq \text{time} < 205$	12	24
$205 \leq \text{time} < 210$	8	16
$210 \leq \text{time} < 215$	3	6
$215 \leq \text{time} < 220$	3	6
$220 \leq \text{time} < 225$	1	2
Total	50	100

Example 3



Example 3



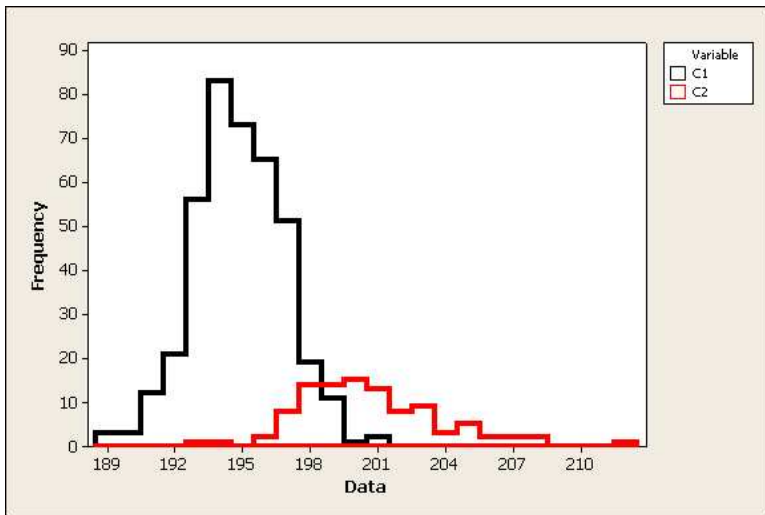
Why use percentage relative frequency?

It is useful for comparing two or more histograms

- If one sample were larger than the other, the standard histograms would look different just because of the different sample sizes
- Looking at percentages “**puts both samples on the same scale**” and removes this difference
- This enables us to look at **relative** differences

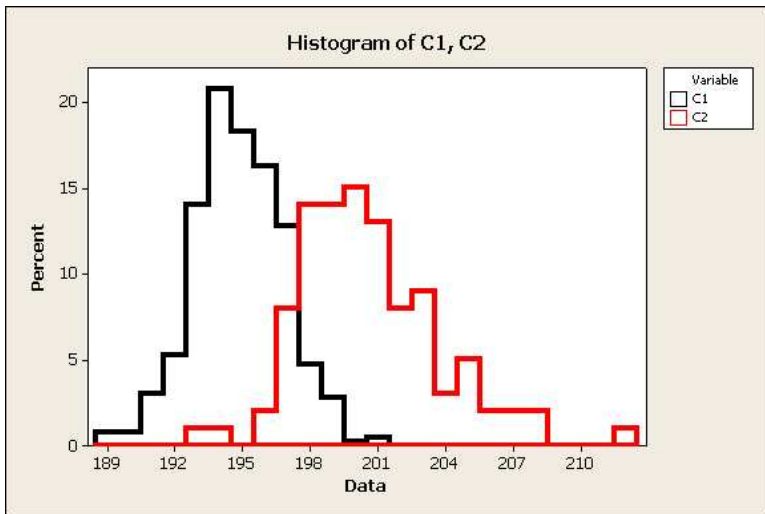
Why use percentage relative frequency?

The following (frequency) histograms were created from two samples, one of size 100 and one of size 400 ...



Why use percentage relative frequency?

... and this is what happens when we use **% relative frequency**:



2.4 Relative frequency polygons

So percentage relative frequency histograms are useful for comparing two groups.

However ...

... can you imagine how **messy** these “superimposed” histograms would get if we had three or more groups?

To get round this **messy problem** we use **relative frequency polygons**.

2.4 Relative frequency polygons

These are a **natural extension** to the relative frequency histogram.

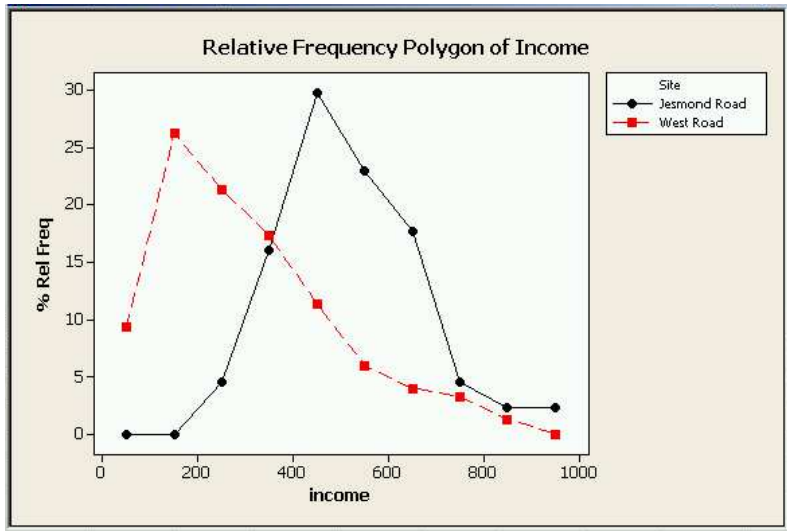
They differ in that, rather than drawing bars, each class is represented by one point (**class interval mid-point**) and these are joined together by straight lines.

Example 4

Consider the following data on gross weekly income (in £) collected from two sites in Newcastle.

Weekly Income (£)	West Road (%)	Jesmond Road (%)
$0 \leq \text{income} < 100$	9.3	0.0
$100 \leq \text{income} < 200$	26.2	0.0
$200 \leq \text{income} < 300$	21.3	4.5
$300 \leq \text{income} < 400$	17.3	16.0
$400 \leq \text{income} < 500$	11.3	29.7
$500 \leq \text{income} < 600$	6.0	22.9
$600 \leq \text{income} < 700$	4.0	17.7
$700 \leq \text{income} < 800$	3.3	4.6
$800 \leq \text{income} < 900$	1.3	2.3
$900 \leq \text{income} < 1000$	0.0	2.3

Example 4



Example 4

Example comments

- Distribution of incomes on West Road is **skewed towards lower values**; those on Jesmond Road are more **symmetric**
- There is evidence of higher incomes in the Jesmond Road area than in the West Road area
- The **spread** of incomes is roughly the same in the two areas
- There are no obvious **outliers**

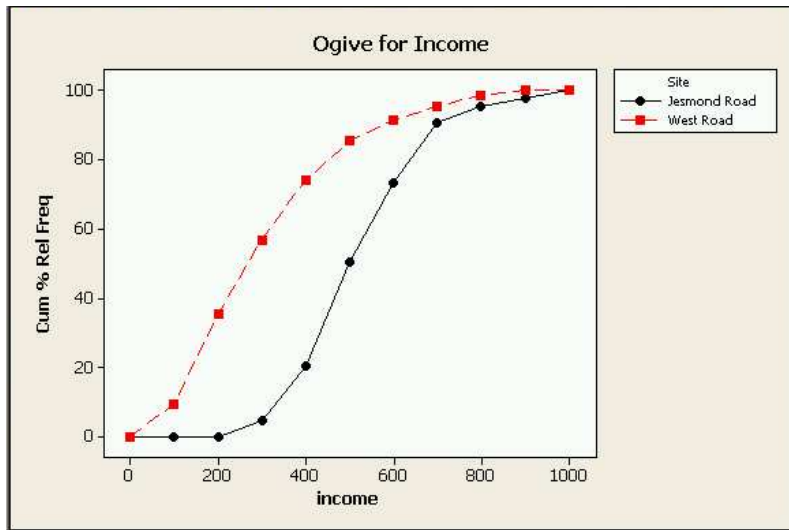
2.5 Cumulative frequency polygons

These are very useful for comparing datasets

- Construct a percentage relative frequency table
- Add a “cumulative” column by adding up the percentages as you go along
- Plot the **upper end-point** of each class interval against the cumulative value

Example 5

For the income data on West Road and Jesmond Road, we get:



2.6 Scatter plots

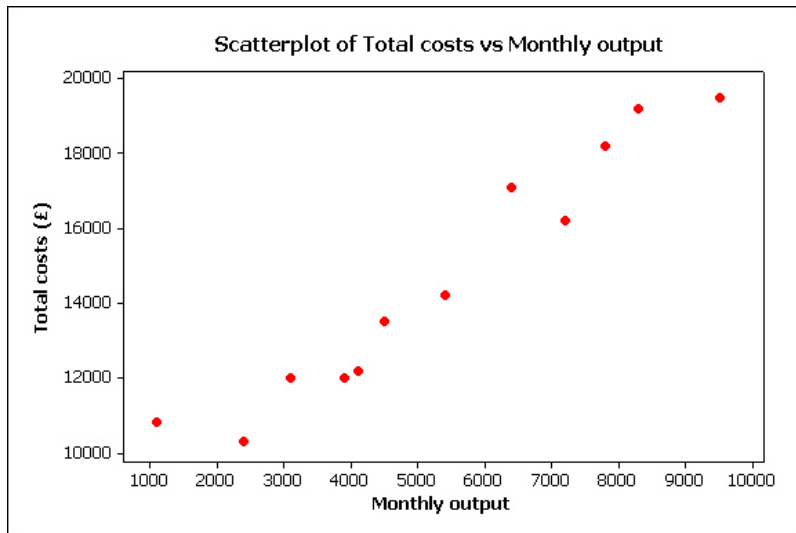
Scatter plots are used to plot two variables which you think might be related – for example: height versus weight; advertising expenditure versus sales...

Example 6

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

Total costs (£)	Monthly Output (units)
10,300	2,400
12,000	3,900
12,000	3,100
13,500	4,500
12,200	4,100
14,200	5,400
10,800	1,100
18,200	7,800
16,200	7,200
19,500	9,500
17,100	6,400
19,200	8,300

Example 6



Example 6

Example comments

- Clear **positive**, roughly **linear**, relationship between the two variables
- Strong, positive **correlation**
- The more units that are made, the higher the total costs

2.7 Time series plots

Data collected over time can be plotted by using a scatter plot, but with **time** as the (horizontal) x -axis, and where the points are connected by lines.

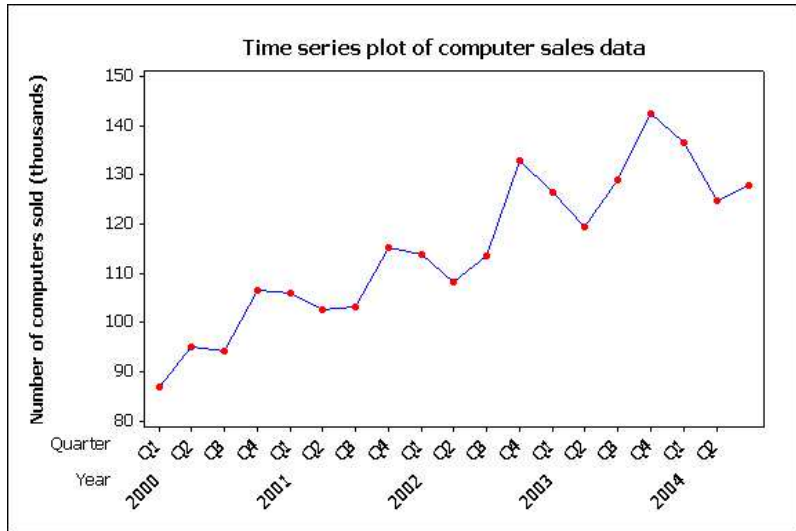
Such a plot is usually called a time series plot (or sometimes simply a “line plot”).

Example 7

Consider the following data on the number of computers sold (in thousands) by quarter (January–March, April–June, July–September, October–December) at a large warehouse outlet.

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

Example 7



Example 7

Example comments

- There is a clear **positive trend** in computer sales through time
- There appear to be **seasonal cycles** (highest sales in Q4 in each year, lowest sales in Q1)
- Seasonal differences appear to be getting more defined?
- No obvious **outliers**