MAS1403

Quantitative Methods for Business Management

Semester 1, 2009–2010

Lecturer: Dr. Lee Fawcett

Additional lecturers: Dr. Andy Golightly and Dr. Tom Nye

Contact details

My name:	Dr. Lee Fawcett
Office:	Room 2.07 Herschel Building
Phone:	0191 222 7228
Email:	`lee.fawcett@ncl.ac.uk`
www:	`www.mas.ncl.ac.uk/~nlf8`

Classes

Weekly **lectures** take place on Mondays, at 11am (or 3pm), in the Curtis Auditorium in the Herschel Building.

Each student should attend one **tutorial** session later on in the week. Due to the size of the class, there are *seven* sessions, A, B, C, D, E, F and G.

Group	Day	Time	Room
A	Wed	9–10	Herschel LT2
B	Wed	12–1	Herschel LT2
C	Thurs	1–2	Herschel LT2
D	Thurs	2–3	Cassie 2.32
E	Thurs	3–4	Agriculture, CSLT
F	Fri	9–10	Stephenson F.13
G	Fri	10–11	Bedson LT3

A register is taken in tutorial sessions!

Classes

Occasionally throughout the year, a **computer practical session** will take the place of the scheduled tutorial session.

All sessions take place in the Herschel Building Computing Cluster on the first floor of the Herschel Building.

Group	Day	Time	Dates
A	Wed	9–10	7 Oct, 11 Nov
B	**Tues**	1–2	6 Oct, 10 Nov
C	Thurs	1–2	8 Oct, 12 Nov
D	Thurs	2–3	8 Oct, 12 Nov
E	Thurs	3–4	8 Oct, 12 Nov
F	Fri	9–10	9 Oct, 13 Nov
G	**Thurs**	10–11	8 Oct, 12 Nov

Office 'hour'

As well as seeing you for two hours a week in the **lecture** and the **tutorial**/**computer practical**, I have also timetabled a 2–hour "drop-in" session each week.

Each Wednesday, from 1pm – 3pm, I will be available in my office to see students needing extra help with the work. This will be on a voluntary "drop–in" basis.

Lecture notes and handouts

Lecture notes, together with solutions to tutorial exercises and any other handouts, will be posted on the course website. You should make the following link one of your "favourites":

www.mas.ncl.ac.uk/∼nlf8

There will be a link to this website from the MAS1403 homepage in Blackboard.

You should print out and read (!) lecture notes before each lecture.

Some handouts will only be available to collect in class!

At the end of each lecture I will usually set you some work to do, from the lecture notes, before the tutorial.

Assessment

The assessment for this module consists of two 1 hour **exams** at the end of each Semester (next January for Semester 1 and May for Semester 2), two **assignments** (one in each Semester) and six **computer based assessments** (three in each Semester).

- **Exam (30% in each Semester)**
  The exams for MAS1403 will be "open–book". This means you can take all your stuff in!

  In previous years, the exam has been a traditional written paper; this year, it will be an automated computerised exam. More info on the exam will be given later in the course.

Assessment

- **Assignments (10% in each Semester)**
  There are two assignments, one in each Semester. An announcement will be made in class when these are available to download. Assignment 1 is due in by *4pm, Wednesday 9th December 2009*.

- **CBAs (10% in each Semester)**
  There will be three computer based assessments in each Semester. You will be given a tutorial on how to use the system in the practical sessions next week.

Other info

- **Announcements**
  Announcements relating to the course will be made via email. You should check your email at least twice a week!

- **Videos**
  We are in the process of shooting some videos which will show detailed calculations relating to the more difficult parts of the course. These will be available to view from the course website.

Other info

- **Wine/beer/book questions**
  Some weeks, in tutorials, I will set "prize questions". You will have until the following lecture to submit solutions to these questions to me. These questions will be quite taxing, and as such will attract a prize (usually alcoholic!) for students submitting the best solutions.

- **Calculator**
  All students should buy a **Casio fX-83** or a **Casio fx-113** calculator. Others might be allowed – you should check the University's regulations (see course website).

Some pitfalls

False sense of security

The course starts *very easy*

- This week, we'll look at how data are collected
- Next week, we'll draw some graphs
- The week after, we'll draw more graphs

That's when attendance usually drops off... students think the work is too easy, so why bother coming to lectures?

Some pitfalls

False sense of security

- You need the basics in order to get a handle on the more difficult stuff
- The course will get more difficult by about week 4 or 5
- By Christmas you will be tackling A–Level standard Statistics problems
- By Easter next year, you will be tackling degree–level Statistics problems!

Solution – regular attendance, regular testing!

Some pitfalls

Maths is too hard. I'll never pass, so what's the point?

- This is not a maths course. It's a Stats course – there is a difference!
- We start with stuff you did in Year 8 and build up gradually!
- Out of 454 students taking MAS1403 last year, only 5 students failed
- Regular lecture/tutorial attendance has a proven track record – even for those who are really scared of the subject! Practice makes perfect!
- Lots of support – and I don't think I'm scary!
- Out of 300 Marketing and Management students taking MAS1403 last year, 219 said this was their favourite module

Some pitfalls

Denial – Maths and Stats is unnecessary for me!

- Employers love graduates who are numerate
- You will need quite a few of the techniques this course provides in other courses and in your final year dissertation
- Real–life Accounting/Finance/Management/Marketing companies use the techniques you will learn in this course to help them make very big decisions – I know this for a fact
- We will use real–life data
- This course will make you a better person, and you will gain a massive sense of achievement when you pass next year!
- "As soon as I graduate I want to be a full–time live at home Mother for the rest of my life"

Main piece of advice

Keep on top of the work!

Enough of that...

... let's start the course!

Lecture 1

COLLECTING AND PRESENTING DATA

Introduction

Data are the key to many important management decisions.

- Is a new product selling well?
- Do potential customers like the new advertising campaign?

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

In this lecture we will look at

- the **different kinds** of data that exist
- how we might **obtain** our data
- how we might **present** our data

Examples

Sizing clothes

- Most clothing now comes in essentially standard sizes...
- but where do these standards come from?
- By sampling from the population as a whole, standards can be set around the most common sizes
- We cannot say that an individual is *exactly* a standard size...
- however, we *can* say that they will probably fall within a range either side of a standard size

Car maintenance

- Just imagine you're buying a new car...
- it might be useful to know how much it will cost to run over the next few years
- Obviously this cannot be predicted exactly –
  - Each car will be slightly different
  - Each user will be slightly different
- Collecting data from people who have bought similar cars will give us some idea!

Some definitions...

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

A particular outcome is called an **observation**

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

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

Example

Suppose we are interested in the height of students on Management courses at Newcastle...

- Our **random variable** is "the height of students on Management courses at Newcastle"
- If Joe Bloggs was a Management student, and we measured his height, then that value would be a single **observation**
- If we measured the height of every first year Management student, we would have a collection of such observations which would be our **data**
- This would be a **sample** from the **population** which consists of all students registered with the School of Management

Ideally, to get a true idea of what is going on, we'd like to observe the whole population (take a **census**). However, this can be difficult:

- If the population is huge, then this would take ages!
- *And* it would be very costly!
- In reality, we usually observe a subset of the population... but how do we choose who to observe?

Types of data

There are two types of variables: **qualitative** and **quantitative**.

- *Qualitative* variables have non–numeric outcomes, and are usually **categorical**.
  - sex of a person (categories: male or female)
  - colour of a car (categories: red, black, silver, blue – could code this to 1, 2, 3 and 4)
  - mode of transport
- *Quantitative* variables have numeric outcomes with a natural ordering
  - people's height
  - number of defective components in a batch

Quantitative variables

We are most interested in **quantitative** variables.

These variables cane be subdivided into two types:

discrete and **continuous**.

Discrete random variables

- are usually integers
- are *countable* – e.g. the number of red cars going past a junction in an hour
- can be **ordinal** – where the outcomes are ordered

Continuous random variables

- can take any value over some continuous scale – e.g. height or weight
- can be measured to a very high degree of accuracy (provided we have the equipment to do so)...
- ...however, we can never say *precisely* how much someone weighs, for example
- might be measured to the nearest whole number – and so could "look" discrete – be careful!

Surveys

surveys or **questionnaires** are often used to gain insight into the impact of many management decisions – for example, customer views on the impact and potential of new technologies.

When preparing a survey, you should consider the following key questions:

- what is the **purpose** of the survey?
- what is the **target population**?
- is there a list of the target population?
- how can **bias** be avoided?
- how **accurate** does the survey have to be?
- what **resources** are available for conducting the survey?
- how are the data to be **collected**?

How can we collect survey information?

There are many ways of collecting survey information. Each has its advantages and disadvantages.

1. *Postal questionnaire* – low cost, low response rate, slow turn around time, low quality information
2. *Telephone interviewing* – moderate cost, moderate response rate, fast turn around time, good quality information (?)
3. *Face–to–face interviewing* – high cost, high response rate, fast turn around time, high quality information

Sampling

We can rarely observe the whole population. Instead we observe some sub–set of this (called the **sample**).

The difficulty is in obtaining a **representative** sample.

For example, if you were to ask 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 cannot be stressed too highly.

Sampling techniques

There are three general forms of sampling techniques:

1. **Random sampling** – where the members of the sample are chosen by some random mechanism
2. **Quasi–random sampling** – where the mechanism for choosing the sample is only partly random
3. **Non–random sampling** – where the sample is specifically selected rather than randomly selected

Simple random sampling...

This method is the **simplest to understand**.

If we had a population of 200 students we could put all their names into a hat and draw out 20 names as our sample.

Each name has an **equally likely chance** of being drawn and so the sample is completely random.

Furthermore each possible sample of 20 has an equal chance of being selected.

In reality the drawing of the names would be done by a computer and the population and samples would be considerably larger.

... and some disadvantages

We often don't have a **complete list** of the population.

For example, if you were surveying the market for some new software, the population would be everybody with a compatible computer!

Not all elements of the population are **equally accessible**.

Purely by chance, you could pick an **unrepresentative sample**!

Stratified sampling...

This is a form of random sample where clearly defined groups, or **strata**, exist within the population.

If we know the overall proportion of the population that falls into each of these groups, we can randomly sample from each of the groups and then adjust the results according to the known proportions.

For example, if we assume that the population is 55% female and 45% male and we wanted a sample of 1000, we would

- decide to have 550 females and 450 males in our sample, and then
- pick the members of our sample from their respective groups randomly

... and some disadvantages

We need clear information on the **size** and **composition** of each stratum which can be difficult to obtain.

We still need a **list of the entire population** to sample from.

Systematic sampling...

This is a form of **quasi–random sampling** which can be used when the population is clearly structured.

For example, if you were interested in obtaining a 10% sample from a batch of components being manufactured, you would

- select the first component at random, and then
- pick every 10th item to come off the production line after that

This scheme is **very easy to implement**!

... and some disadvantages

This method is **not entirely random**!

If there is a pattern in the process, it is very easy to obtain a **biased sample**.

Multi–stage sampling...

This is another form of **quasi–random sampling**, and is used when the population is spread across a wide geographic area.

Example

If we were interested in sampling school children, we might

- take a random (or stratified) sample of education authorities, then
- within each selected authority, take a random (or stratified) sample of schools, then
- within each selected school, take a random (or stratified) sample of children

This is likely to **save time**, and **cost less** than sampling the whole population.

... and some disadvantages

The sample can be **biased** if the stages are not carefully selected.

Indeed the whole scheme needs to be carefully thought through and designed to be **truly representative**.

Cluster sampling...

This is a method of **non–random sampling**.

For example, a geographic area is sub–divided into clusters, and all the members of these clusters are then surveyed.

If necessary, these clusters can be further subdivided into smaller clusters until we can easily sample all members of each cluster.

The main advantage of this approach is that it is **relatively inexpensive** because the sampling takes place in a concentrated area.

... and some disadvantages

Using small clusters to allow the entire cluster to be surveyed introduces the strong possibility of **bias** within the sample.

Example
If you were interested in the take up of organic foods and were sampling via the cluster method you could easily get biased results:

- If you picked an economically deprived area – the proportion of those who ate organically might be very low
- If you picked a middle–class suburb, however, the proportion might be higher than in the overall population

Other methods of sampling

Judgemental sampling

- Here, the person interested in obtaining the data decides whom they are going to ask
- This scheme is **non–random**
- It is also prone to **bias**

Accessibility sampling

- Here, only the **most easily accessible** individuals are sampled
- This is obviously prone to **bias**
- It is, however, **convenient** and **cheap**

Quota sampling

- This method is similar to stratified sampling but uses judgemental sampling within groups
- The population is broken down into groups (could be age groups) and we sample within these groups until our **quota** is reached
- One benefit is the accuracy of the results obtained as a result of specific targeting.

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

Non–categorical data

The following table shows the raw data for car sales at a new car showroom over a two week period in July.

Date	Cars Sold	Date	Cars Sold
01/07/04	9	08/07/04	10
02/07/04	8	09/07/04	5
03/07/04	6	10/07/04	8
04/07/04	7	11/07/04	4
05/07/04	7	12/07/04	6
06/07/04	10	13/07/04	8
07/07/04	11	14/07/04	9

Presenting these data in a relative frequency table by number of days on which numbers of cars were sold, we get the following table:

Cars Sold	Tally	Frequency	Relative Frequency %
1		0	0
2		0	0
3		0	0
4	I	1	7.14
5	I	1	7.14
6	II	2	14.29
7	II	2	14.29
8	III	3	21.43
9	II	2	14.29
10	II	2	14.29
11	I	1	7.14
Totals	14	14	100

Continuous data frequency tables

With discrete data it is easy to count the quantities in the defined categories. With **continuous data** this is not possible.

For example the following data set represents the service time in seconds for callers to a credit card 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

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**

For the call centre 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