MA208 Quantitative Techniques for Business

Lecture 1: Introduction to Statistics

Dr Kirsten Pfeiffer

School of Mathematics, Applied Mathematics and Statistics
NUI Galway

Lecture 1 - Outline

- About this module
- Textbooks
- Introduction to Statistics
- Graphical methods to describe and analyse data sets

About this module

In this module we will cover the following topics:

1. Descriptive Statistics
2. Counting and Probability
3. Mathematics of Finance
4. Linear Systems / Matrices
5. Linear Programming

About this module

Lecturer: Dr Kirsten Pfeiffer,
School of Mathematics, Applied Mathematics and Statistics.
Room ADB-G012, Arás De Brún

Feel free to call to the office anytime for any course questions.

You can also use the **SUMS** facility in Arás de Brún.

This module is taken by

- 2nd Commerce (Business Information Systems)
- 1st Arts (Mathematics, Mathematical Studies)
- Visiting student(s)

About this module

- Lectures:
  - Monday at 12:00 in AC216,
  - Friday at 15:00 in AM150.
- Tutorials to be confirmed.
- Assessment:
  - **Continuous assessment:** four online assignments, worth 30% in total.
  - **Final assessment:** 2 hour exam at the end of the semester, worth 70%.

SUMS

The School of Maths provides a free drop-in centre called **SUMS**: **S**upport for **U**ndergraduate **M**athematics and **S**tatistics.

SUMS opens

- Monday to Friday from 2pm to 5pm and
- Wednesday 7pm-9pm.

For more information, see

http://www.nuigalway.ie/public-sites/s-u-m-s/

Textbooks

The main recommended textbooks are

- A. Barnett, M.R. Ziegler and K. Byleen: *College Mathematics for Business, Economics, Life Sciences & Social Sciences*.
  Available at Main Library Open Access (510BAR).
- J.T. McClave, G. Benson, T. Sincich: *Statistics for Business and Economics*.
  Available at Main Library Open Access (519.5 McC).

Textbooks

Other recommended texts include

- Frank S. Budnick: *Applied mathematics for business, economics, and the social sciences*,
- David R. Anderson, Dennis J. Sweeney and Thomas A. Williams: *Quantitative methods for business*,
- Clare Morris: *Quantitative approaches in business studies*,
- Louise Swift: *Quantitative methods for business, management and finance*.

Section I: Statistics

In this section we will cover

- Some terminology
- Graphical methods to describe and analyse data
  - Dot Plots
  - Bar Charts / Pie Charts
  - Histograms
  - Stem and Leaf Plots
- Numerical methods to describe and analyse data
  - Measures of centre
  - Measures of spread
  - Five Number Summary
  - Box Plots
  - Standard deviation
  - Coefficient of variation

Introduction to Statistics

Statistics is the science that studies the collection, organization, analysis, interpretation and presentation of numerical data.

A common goal of surveys and other data collecting tools is to collect data from a smaller part of a larger group so that something can be learned about the larger group.

In statistics the larger group is the *population* and the smaller group is the *sample*. Populations can be topics as diverse as "all people living in Kerry" or "amount of customers that lodge money in a bank on a given day".

Introduction to Statistics

Statistics can be divided into three parts:

- **Producing data**. Once a sample that is the representation of the population is determined, data is collected from the sample members.
- **Descriptive Statistics: organising and analysing data**. The population data is summarised by describing what was observed in the sample (numerically or graphically). This is the area we will study in this module.
- **Inferential Statistics: drawing conclusions.** Patterns in the sample data are used to draw inferences about the population represented. Opinion polls fall under the heading of inferential statistics. In a poll we survey a sample of the population and then try to make statements about the entire population.

Data

There are two main types of data

- **Quantitative Data:** such as measurements or counts, e.g. measuring age, height, weight etc.
- **Qualitative Data:** such as traits or characteristics, e.g. hair colour, eye colour, favourite food etc.

Note: not all data represented by numbers are numerical data (e.g. 1 = female, 2 = male).

Descriptive statistics summarise or describe the important characteristics of a known set of population data. The important characteristics of data can be described using **numerical** or **graphical** methods.

Data

Having the collection of data just listed may not be useful to analyse these data.

Example

The length in cm of fish caught in a river.

19	20	11	20	18	12	20	14	18	12
12	13	20	13	11	19	15	19	13	20
17	15	20	14	12	18	16	20	12	17
18	19	11	20	17	14	13	21	11	20
14	17	20	15	12	11	20	17	13	14

Dot Plots

There are a few graphical tools which can aid us in our understanding of a given data set. One of these tools is a dot plot. A dot plot is a very simple display. It involves representing each data value as a dot. Such a graphical representation of a list of numbers can often tell us something interesting which might not be immediately apparent from a raw list of numbers.

Exercise

Below are the number of delays per week experienced by an airline.

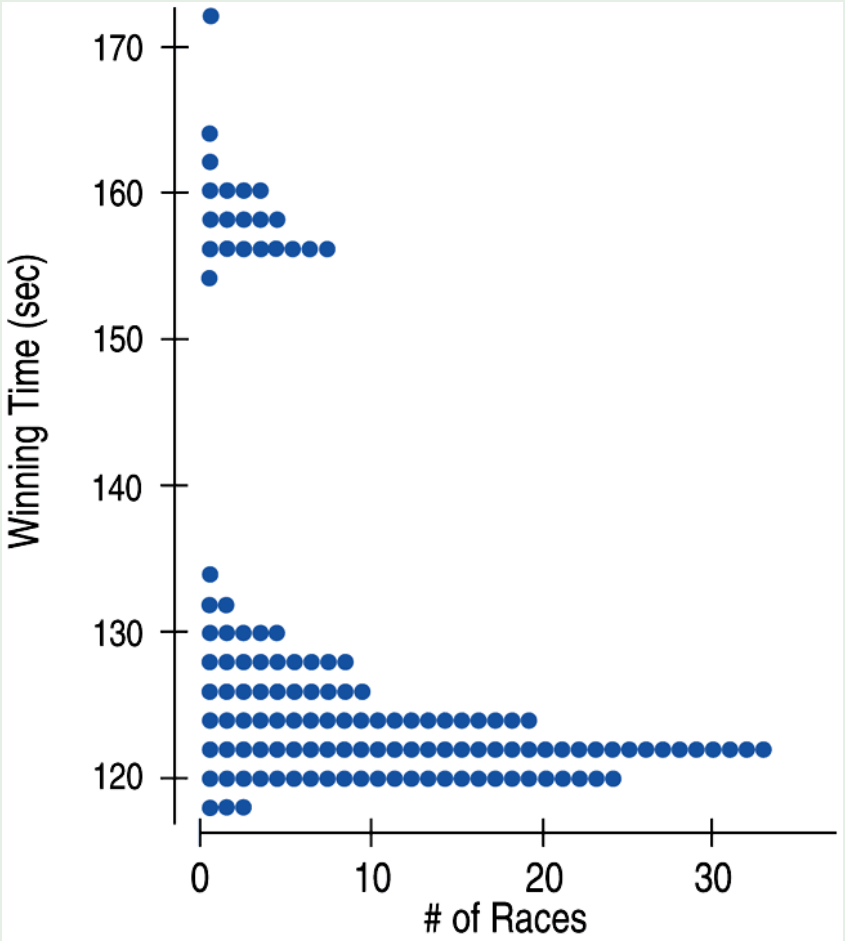
2	1	5	0	1
2	4	1	2	2
0	2	4	1	3

Represent these data in a dot plot diagram.

Dot Plots

Example

The dot plot below shows Kentucky Derby winning times with each race plotted as a dot.



Dot Plots

The dot plot seems to tell us that winning times are separated into two groups. Times around 160 seconds and times around 125 seconds. With a little research we find that the race distance changed in 1995 from 1.5 miles to 1.25 miles. A raw list of numbers would not give us this information. It's often the unusual feature of a data set that tells us something interesting.

Frequency Tables

Example

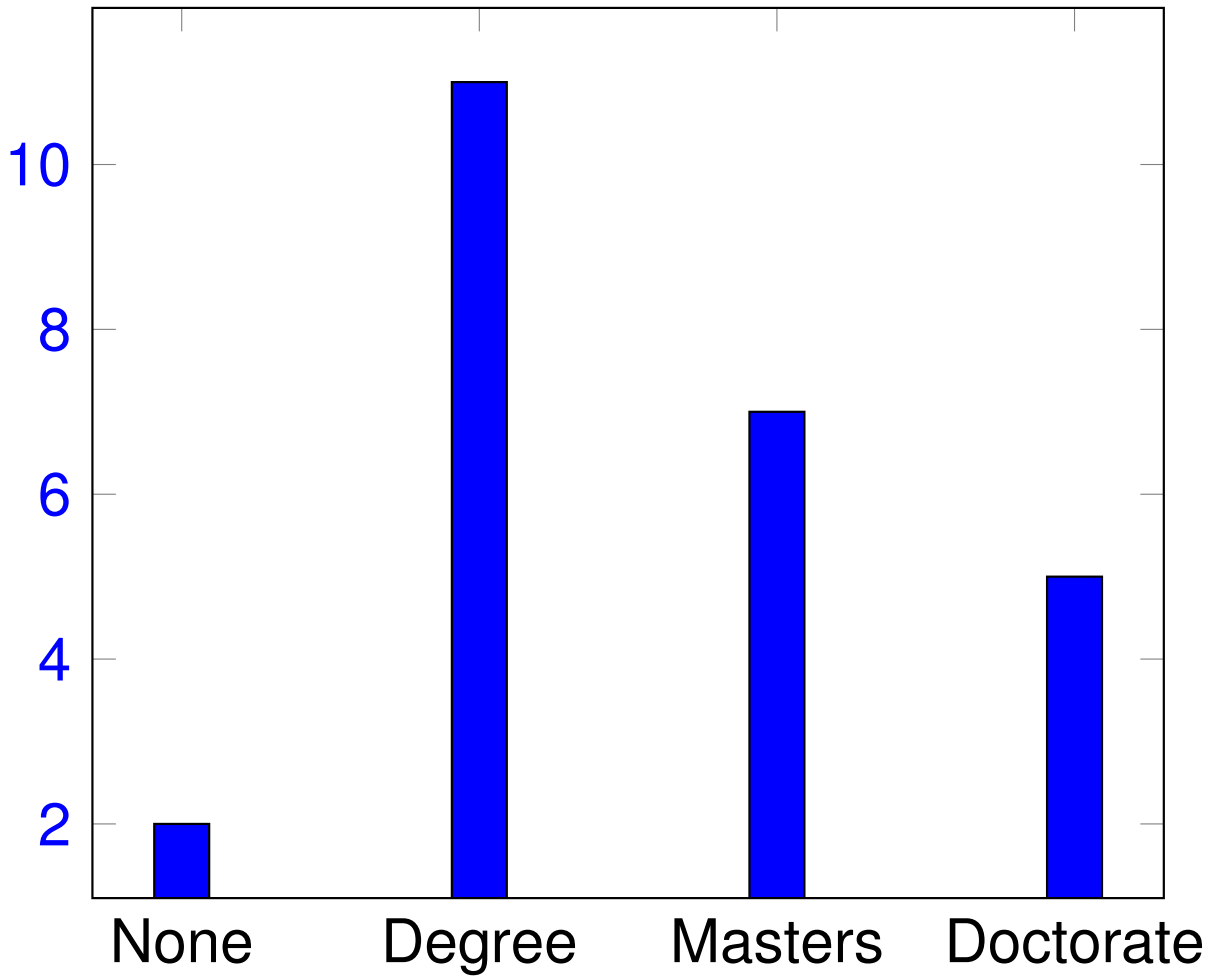
Suppose we survey the level of (3rd level) education of the top twenty five highest paid CEOs in Ireland. We obtain 25 responses which are summarized in the following table.

Education	**Frequency**
None	2
Degree	11
Masters	7
Doctorate	5
Total	25

This is a frequency table. It records the number of responses that fall into each category.

This data can then be summarized graphically using e.g. a bar chart.

Bar Charts



Relative Frequency / Percentage Frequency

Definition

We define the relative frequency of a particular class to be

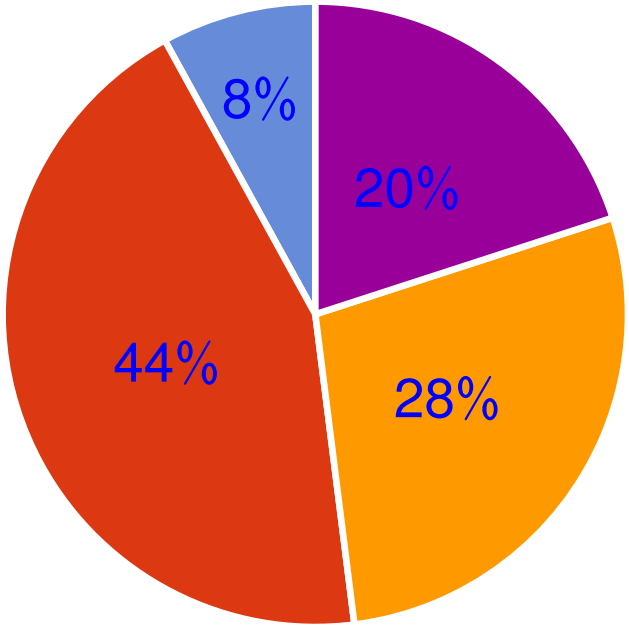
$$\frac{\text{frequency of the class}}{\text{total}}.$$

The relative frequency can then be easily converted to percentage frequency.

Education	Frequency	Rel. Freq.	% Freq.
None	2	0.08	8%
Degree	11	0.44	44%
Masters	7	0.28	28%
Doctorate	5	0.20	20%
Total	**25**	**1.00**	**100%**

Pie Charts

Now we can use the relative (or percentage) frequencies to construct a pie chart.



None
Masters
Degree
Doctorate

Histograms

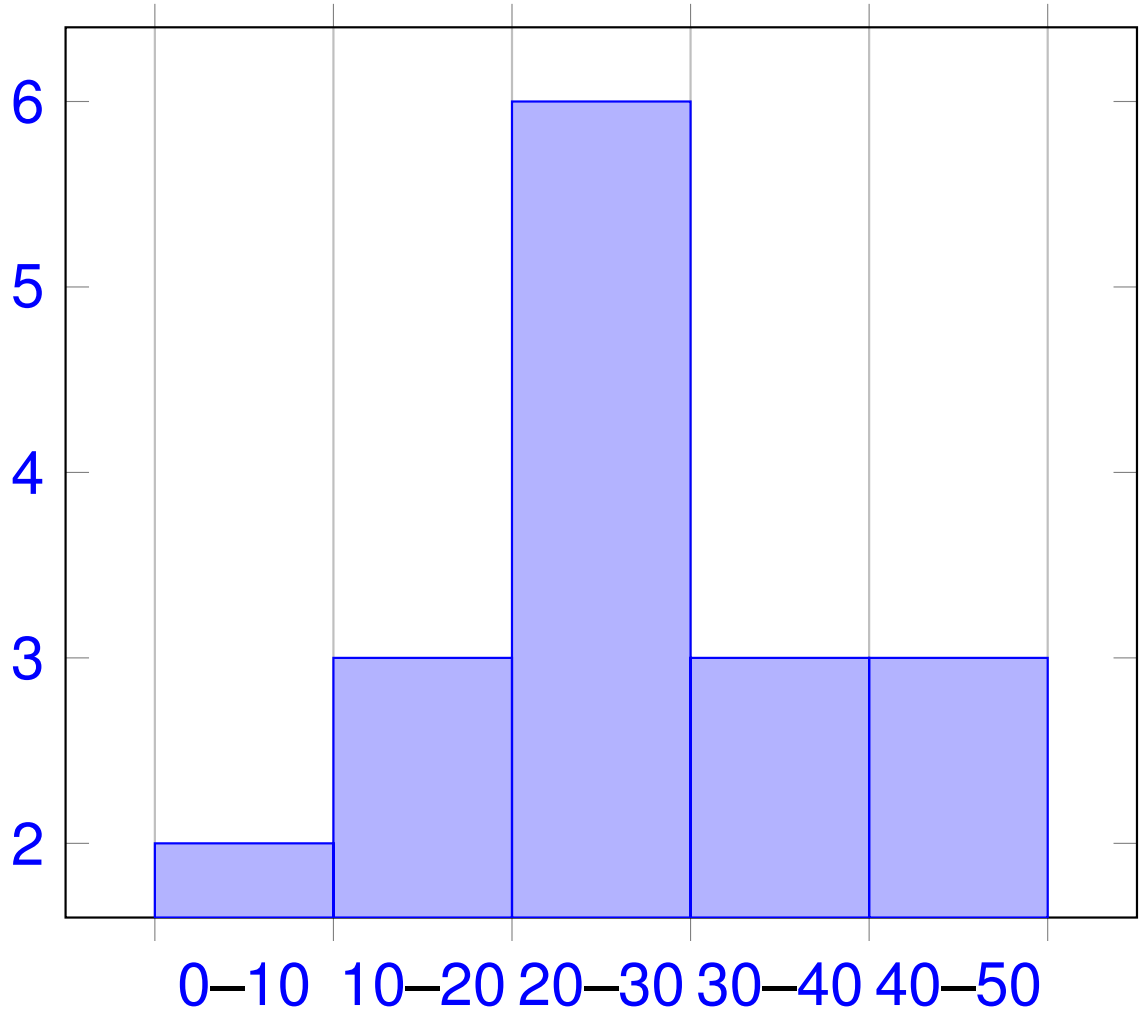
Another useful graphical tool to display quantitative data is a histogram. In a histogram the data values are separated into classes called bins. We then record the number of data values in each class and plot this on a chart similar to a bar chart. Suppose we are given the data set

$$\{0, 3, 11, 12, 19, 22, 23, 24, 25, 27, 29, 35, 36, 37, 45, 49, 49\}$$

Let's construct a histogram for this data set using class (bin) widths of 10. The first thing we need is a frequency table.

Range	Frequency
0 -9	2
10-19	3
20-29	6
30-39	3
40-49	3

Histograms



Histograms

Exercises

(i) The scores of twenty students (out of 100) in a maths exam are given below

25	28	30	40	43	45	48	51	52	52
55	60	61	61	61	63	64	66	67	70

(ii) The scores of 40 students in a second year statistics module at NUIG were as shown below.

29	34	40	61	67	75	33	46	77	41
29	34	44	66	67	75	41	66	78	67
30	35	42	63	70	75	35	71	79	66
33	37	45	64	72	77	37	85	82	97

Construct a frequency table for each with classes of width 10. Draw a histogram and comment on any unusual features (of the histogram).

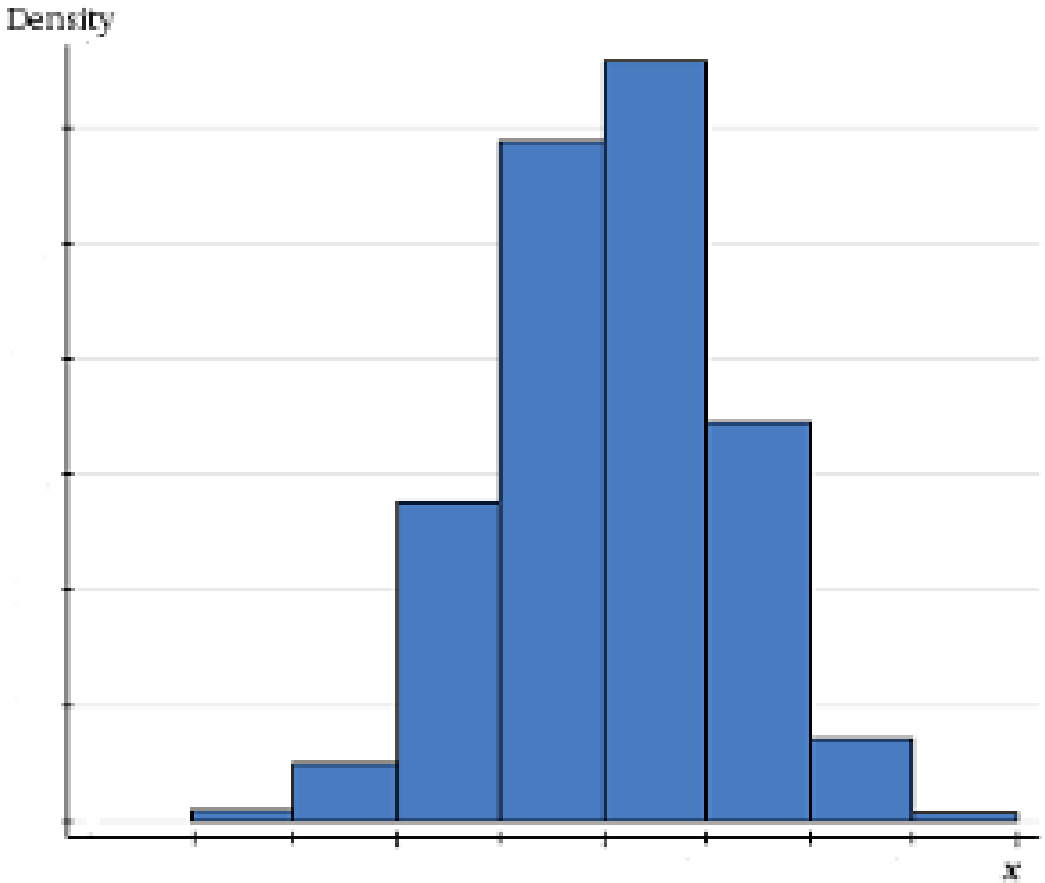
Interpretation of a Histogram

When analysing a histogram we should pay attention to the following:

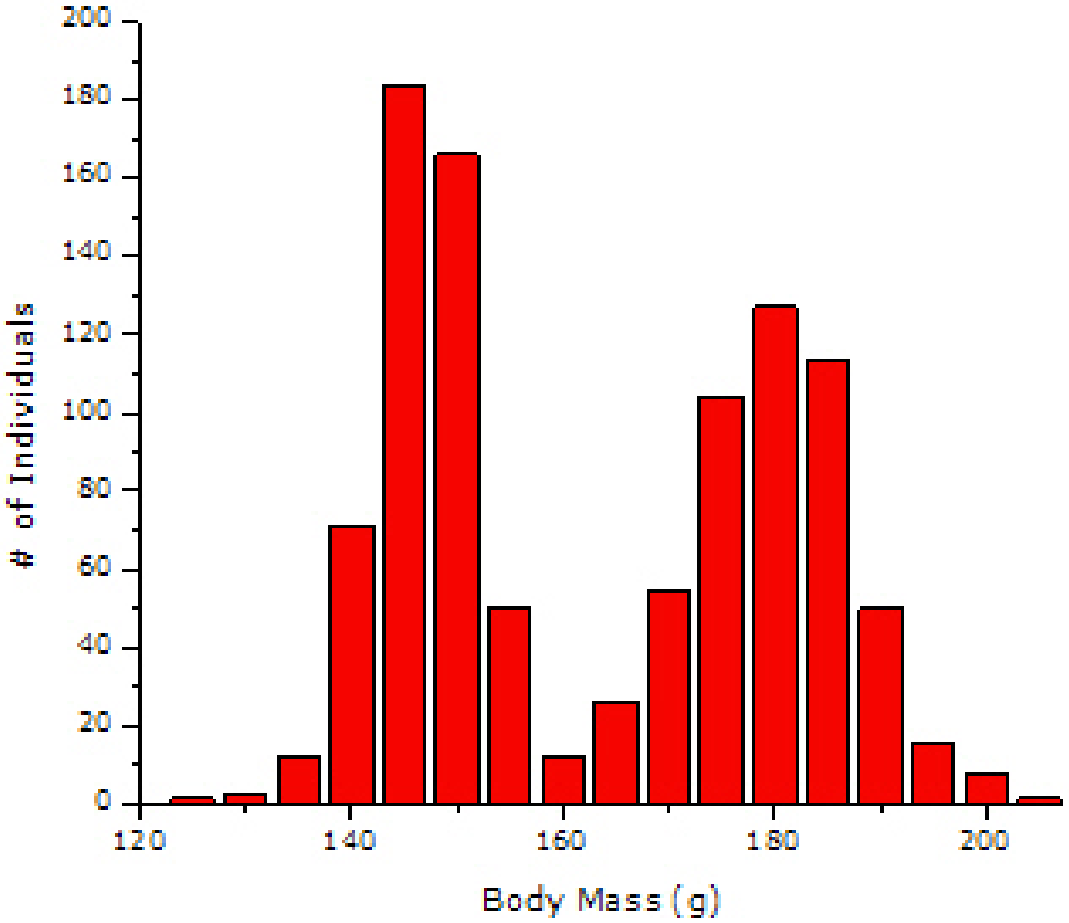
- What is the shape of the histogram? Does it have a single hump or several?
- Where is the histogram centered?
- How are the data values spread out across the histogram? Are they mostly in the centre or towards either ends?

Interpretation of a Histogram

Humps in a histogram are called **modes**. A histogram with one main hump is called unimodal.

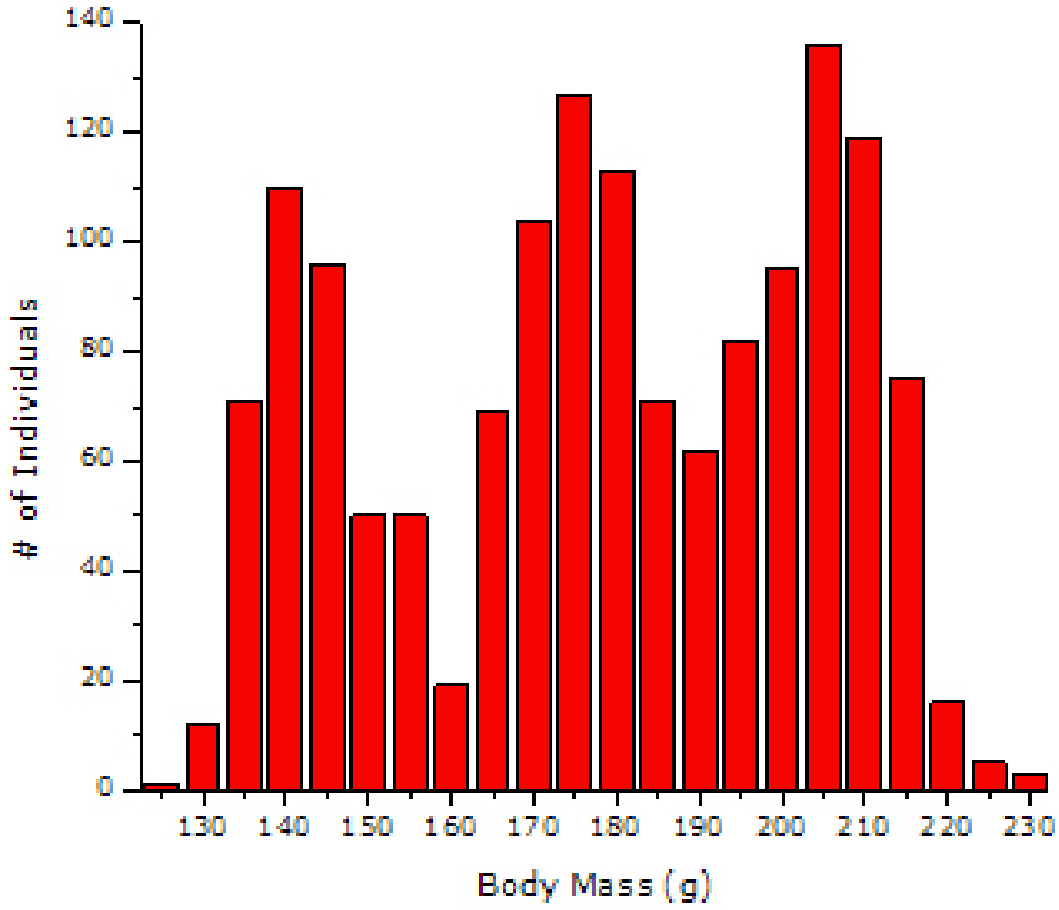


A histogram with two humps is called bimodal.



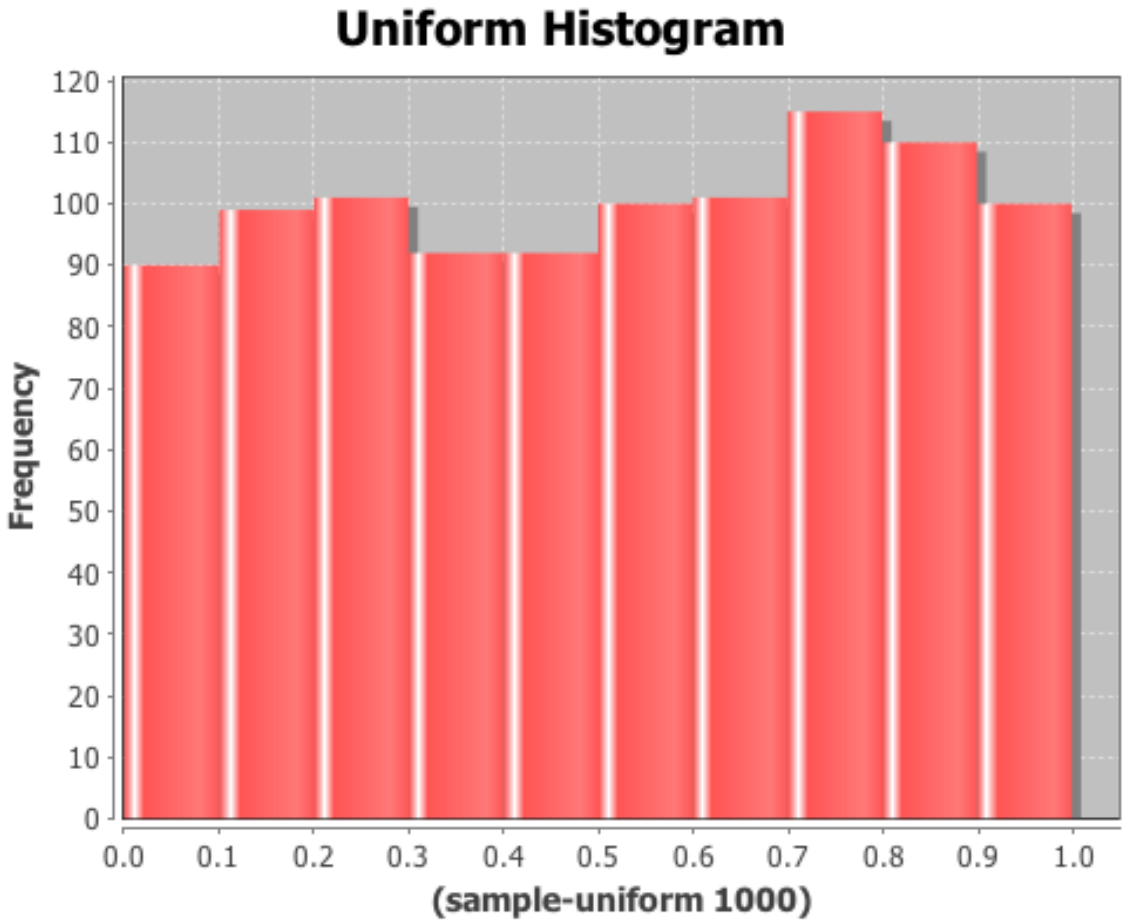
Interpretation of a Histogram

A histogram with three or more humps is called multimodal.



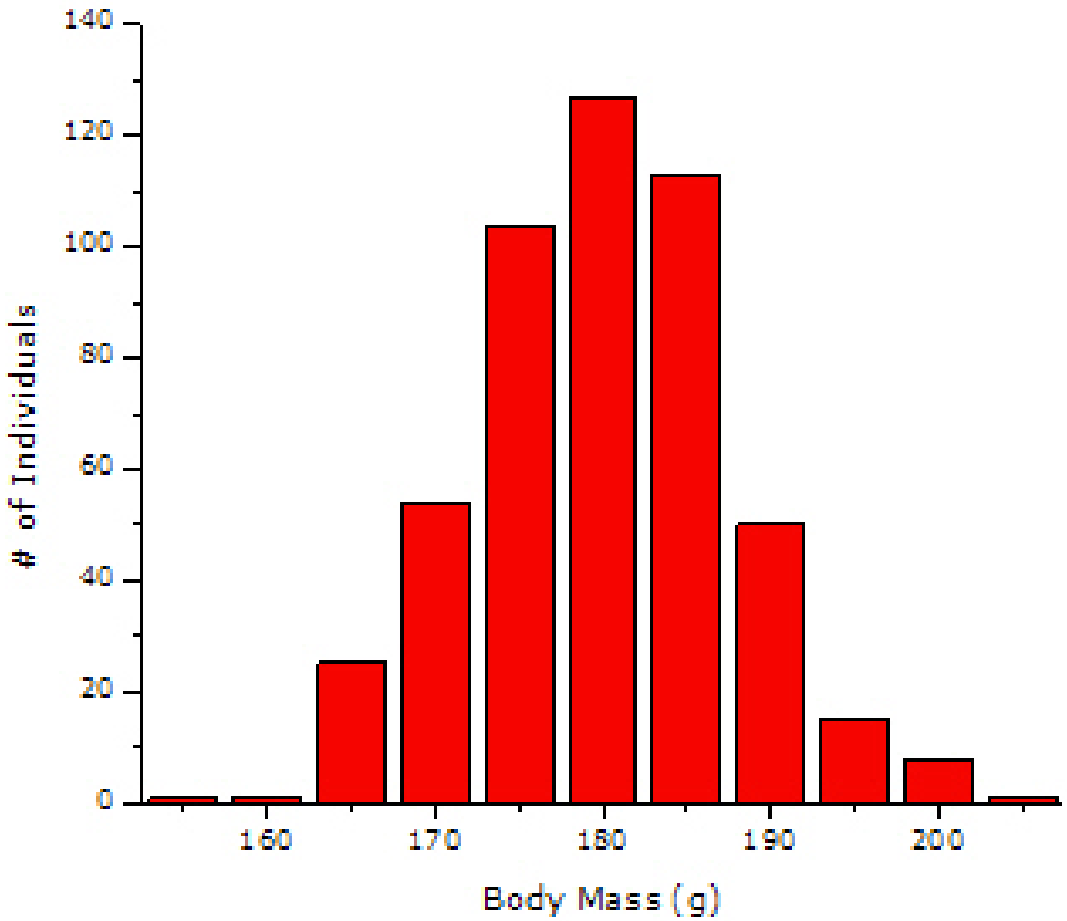
Interpretation of a Histogram

A histogram with no obvious humps is called uniform.

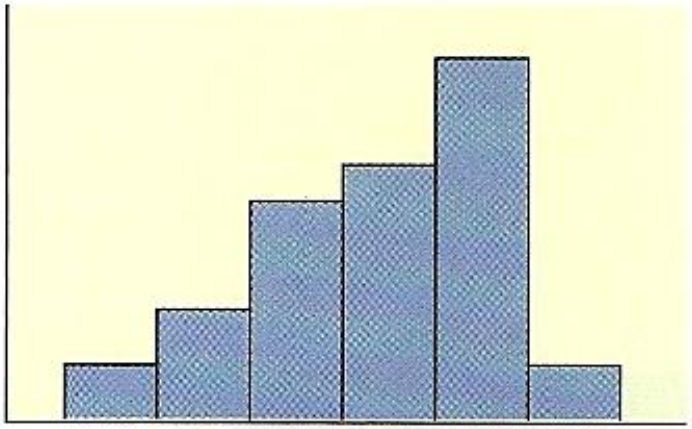


Interpretation of a Histogram

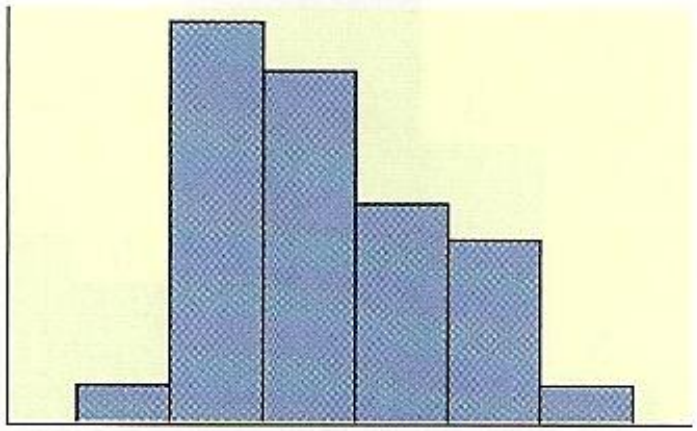
Is the histogram symmetric, or are the data values more to one side than another?



Typical skewed histogram



Skewed left



Skewed right

Interpretation of a Histogram

- Are there any data values far away from the others? These values are known as outliers.
- Are there any gaps? If so we may be looking at data from more than one group.

Graphical tools such as histograms and dot plots tell us quite a lot about a data set. They are easy to draw and easy to talk about.

In our next lecture we will introduce another **graphical** method to describe and analyse data sets and then move on to **numerical** methods.