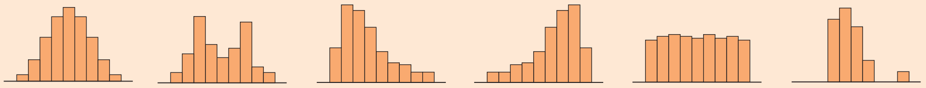
A Describing a Distribution

The shape of a histogram reveals the distribution of the data.

Interesting details of the distribution could be the peak score (mode), gaps, outliers, or clusters

Vocabulary to describe the shape of distributions :



1 The histogram shows weights of checked in luggage on an Air New Zealand flight.

a) Name and describe the shape of the distribution. ……………………………………….....

…………………………………………………………………………………………………

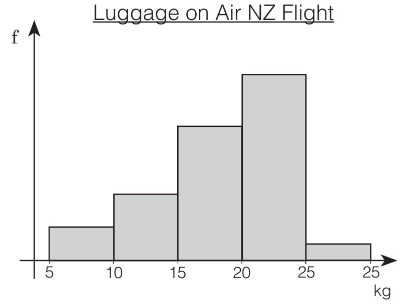
b) Can you think of a reason for the data to have this shape? ………………………………...

…………………………………………………………………………………………………

…………………………………………………………………………………………………

…………………………………………………………………………………………………

Luggage on Air NZ Flight



2 Ages of the audience of Shakespeare's play *'Hamlet'* are displayed in a histogram. Describe the shape and interesting features of the distribution.

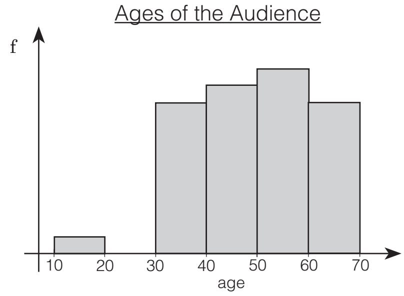
…………………………………………………………………………………………………

…………………………………………………………………………………………………

…………………………………………………………………………………………………

…………………………………………………………………………………………………

Ages of the Audience



3 Students of Year 11 at Bay View College drew the four graphs i) to iv) below. The four graphs have no titles, no labels, no scales. In the grey boxes are the titles, together with descriptions of the graph.

a) For each graph select the correct title - A, B, C or D.

b) Place labels and scales on the horizontal axes.

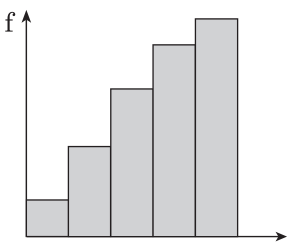
A Travel Time from Home to School
Times are clustered between 0 and 30 minutes with an extreme time of 40-50 minutes.

B Number of Students per Class
A triangular distribution, with class sizes ranging from minimum 5 students to maximum 30 students.

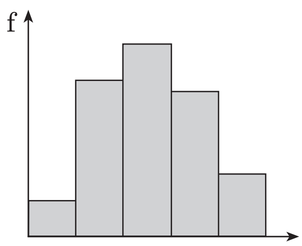
C Time Students Spend in the Library at Lunchtime
A bimodal distribution with most common times 0-5 mins and 15-20 minutes.

D Heights of Year 11 Boys
Almost symmetrical bell shaped distribution centred at 170-180 cm, ranging from 150 cm to 200 cm.

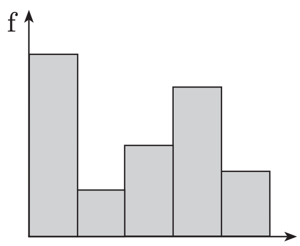
Graph i) ; Title



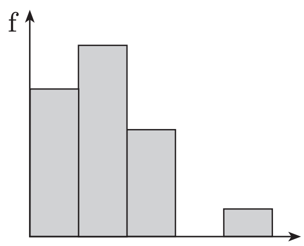
Graph ii) ; Title



Graph iii) ; Title



Graph iv) ; Title



29 Comparing the Samples 1

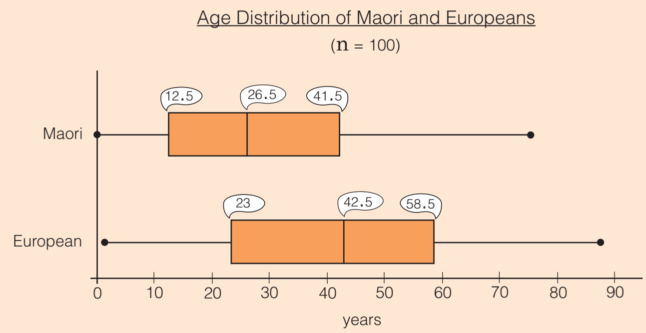
A Comparing Two Samples

Box-and-whisker plots can be used to compare two sets of data.

Method : Draw two box-and-whisker plots on the same scale. Concentrate on the 'box' part of each plot, describe the **spread, shift and overlap** of the middle 50% of the scores.

Back up your observations by referring to measures of centre and spread (median and IQR)

Example : Investigate whether in New Zealand the Maori population is generally younger than people of European background. Two large samples were taken and sample statistics calculated. Compare the two samples shown in the box-and -whisker plot.



Possible comments :

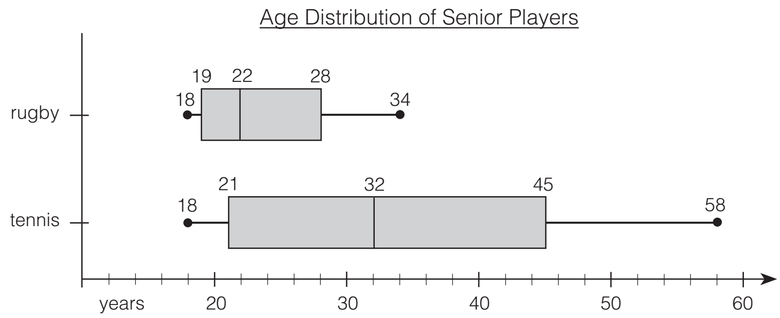
Spread : **I notice that** the bottom half of the plot in the Maori sample is shorter that the top half, while the European sample has a short section in the third quarter. **This means** that the Maori sample is more clustered in the younger age groups (under 23 years) and the European sample is more clustered in middle age (43-59 years).

Shift : **I notice that** the middle 50% of the European sample is shifted 11 to 17 years higher up the scale than the middle 50% of the Maori sample. **This means** that the ages in the sample of Europeans are on average higher than the ages of Maori. Also the median age of Europeans is 42.5 which is 16 years up from the median age of Maori at 26.5 which confirms my observations.

Overlap : **I notice that** there is an overlap between the third quarter of the Maori sample and the second quarter of the European sample but the median of the European sample is outside the middle 50% of the Maori sample. **This means** that, in the samples, half of the Europeans are older than three quarters of the Maori.

1 A random sample of senior players was selected from the tennis clubs and another from the rugby clubs of Christchurch. Box-and-whisker plots have been drawn from this data.

Comment on the **spread**, **shift** and **overlap** of the two age distributions.



....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

  ISBN 978-1-877567-61-2.

37 Statistical Inference 4

A Making the Call - Third Rule of Thumb

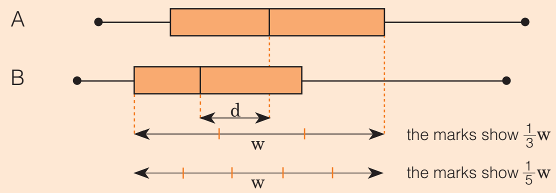
If there seems to be a large gap between two medians, but both medians are in the overlap of the boxes, then the second rule does not work. There is another rule of thumb :

Steps : i) Look at distance between the two medians ($=\mathrm{d}$).

ii) Look at the overall visible spread, that is the total width of the boxes together ($=\mathrm{w}$).

Rule : With a sample size of 30, you can make the call that the scores of population A tend to be larger on average than those of population B, only if $\mathrm{d} > \frac{1}{3}$ of w.

With a sample size of 100, the call can be made if $\mathrm{d} > \frac{1}{5}$ of w.

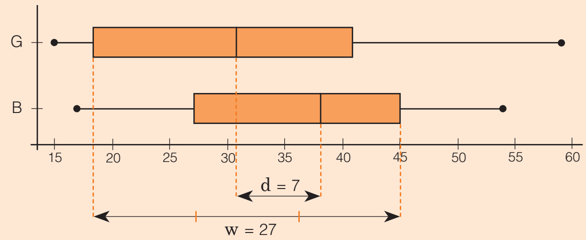


Example :

Gemma and Benjamin recorded how many text messages they sent each day in the previous month ($\mathrm{n} = 30$). The box-and-whisker plot show the results.

Is there sufficient evidence to conclude that in general Benjamin tends to send more text messages per day than Gemma?

Number of Text Messages Sent per Day



Conclusion :

The difference between the medians (d) is about 7 messages, the overall visible spread (w) is about 27 messages.

Since $\frac{1}{3}$ of $27 = 9$, d is not larger than $\frac{1}{3}$ of w, the difference between the medians is not large enough to make the call. The shift shown here could just be due to sampling variation. If we would take repeated samples of 30 days, each would give a different picture and Benjamin's number of text messages may not always be further up the scale than Gemma's. Back in the population of all daily messages sent by Gemma and Benjamin, the pattern shown here is not necessarily happening.

There is *insufficient evidence* to conclude that back in the populations, Benjamin tends to send more text messages per day than Gemma.

1a) Look back at the example. What if Gemma and Benjamin had recorded their scores for 100 days in stead of 30, and the above box-and-whisker graph showed their results over that time. Would the conclusion be different? Show your working.

....

....

....

....

b) What is the reason that there are different criteria for different sample sizes?

....

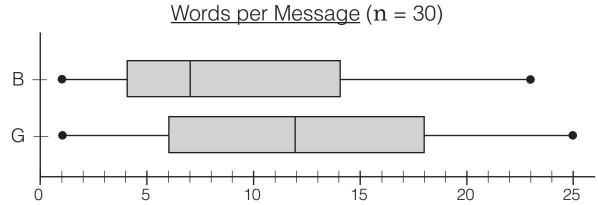
....

....

....

2 Gemma and Benjamin did a survey to see whether Gemma sends longer text messages (more 'words') than Benjamin. They took random samples of 30 text messages on their phones. Find d and w and write a conclusion about who sends longer texts.

Words per Message ($\mathrm{n} = 30$)



....

....

....

....

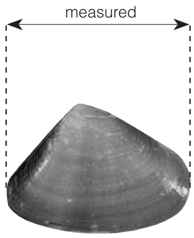
....

....

45 An Investigation 1

A Shell Sizes

1 Tim has two favourite places where he gathers pipis. He reckons site A has larger pipis than site B. Tim investigates his assertion by collecting 30 pipis from each site. He measured them to the nearest millimetre, at the widest part of the shell.



Pipi Shell Sizes (mm)

Site A					Site B				
57	64	65	71	54	53	56	49	68	60
46	74	58	53	35	62	57	47	65	47
48	62	76	57	58	34	59	63	51	43
72	43	44	60	38	40	37	54	59	58
62	59	41	57	47	53	45	39	47	56
67	38	44	55	54	41	51	43	58	44

Write an investigative question for Tim.

....

....

....

2 Draw a back-to-back stem-and-leaf plot for the data. (Use scrap paper first, then order the leaves and copy the result into the space provided.)

Shell Sizes

Site A Site B

key :

3 Calculate summary statistics.

	Site A	Site B
mean		
minimum		
LQ		
median		
UQ		
maximum		
IQR		

4 Draw box-and-whisker plots.

Pipi Shell Sizes

Site A

Site B