

## Descriptions and inferences: comparing two samples (Level 5, Year 10 to Level 7 , Year 12)

### PROBLEM

#### Background

There is a general belief that 13 year-old boys are bigger than 13 year-old girls. I wish to check out the claim that, for 13 year old students in New Zealand, the boys are bigger than the girls.

#### Question

Do 13 year-old New Zealand boys have bigger feet than 13 year-old New Zealand girls?

#### Investigative Question

Do right foot lengths for 13 year-old NZ boys tend to be bigger than right foot lengths for 13 year-old NZ girls?

I expect boys to have bigger feet than girls.

### PLAN

I will get our two random samples using the CensusAtSchool random sampler. Take a random sample of 25 boys from the population of 13 year-old NZ boys in the CensusAtSchool database.

Take a random sample of 25 girls from the population of 13 year-old NZ girls in the CensusAtSchool database. Use the responses on foot length.

### DATA

Managed through CensusAtSchool survey team.

The data came from the student responses to the following 2005 CensusAtSchool survey questions:

1. Are you:

2. How old are you?  years

male

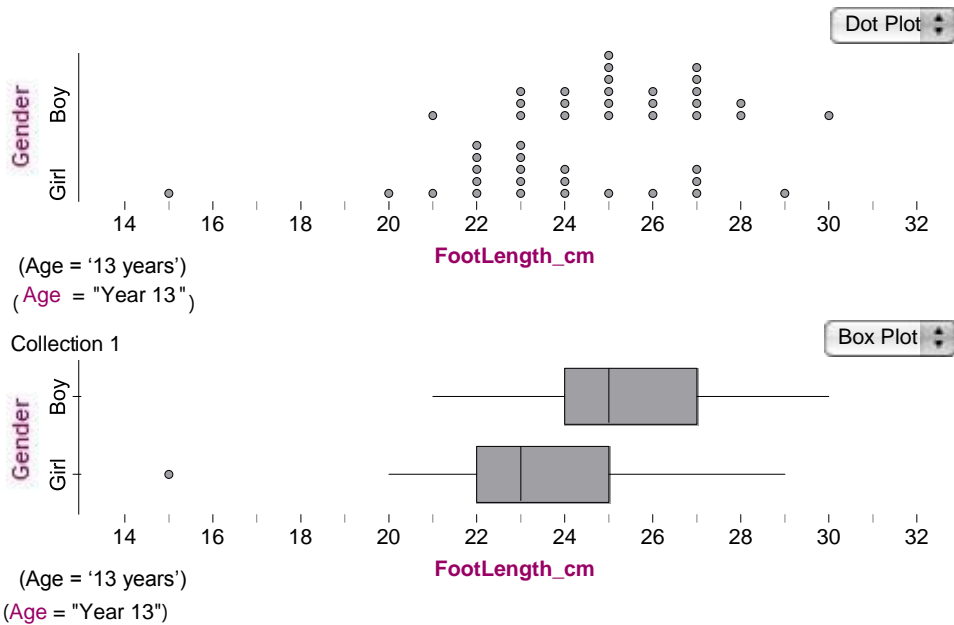
female

3. What is the length of your right foot?  cm

In the random sample of 25 boys, one boy did not record a foot length leaving 24 recorded foot lengths, i.e., there was one missing value.

In the random sample of 25 girls, there were three missing values leaving 22 recorded foot lengths. ***I worry:***

- about the quality of the foot length data since students measured and recorded their own foot lengths
  - Were measurements made with shoes on or shoes off?
  - Would all students have seen 'cm' to the right of the entry box?
  - To what level of precision did the students make their measurement?
  - Why were there missing values?



FootLength_cm	Gender	
	Boy	Girl
S1 = Min	21	15
S2 = LQ	24	22
S3 = Median	25	23
S4 = UQ	27	25
S5 = Max	30	29
S6 = Count	24	22
S7 = Mean	25.5	23.4

Figure 1: Fathom output, 13 year-old boys' & girls' foot lengths

### Shape

***I notice:***

- the sample distribution for the boys' foot lengths is roughly symmetrical with a mound around 24 to 27cm, i.e., unimodal
- the sample distribution for the girls' foot lengths shows a large mound around 22 to 24 cm and a hint of a small mound around 27cm, i.e., a hint of bimodality ***I wonder:***
- if boys' and girls' foot length distributions back in the two populations are roughly symmetric and unimodal. I expect so for a body measurement such as foot length for both girls and boys.

### Individual

***I notice:***

- one of the girls has a foot length (15cm) far smaller than any other girl

***I worry:***

- that this may be a mistake. It could be a measurement or recording mistake or perhaps this girl is much younger than 13 years. I wouldn't expect a 13 year-old girl to have a foot size this small. I need to check her other measurements such as age, height etc. to further investigate this extreme value.

**Gaps/Clusters *I notice:***

- the dots are stacked on whole numbers. This is because the foot lengths are measured to the nearest cm.
- there is a gap in the girls' group at 28cm and gaps in the boys' group at 22 and 29cm ***I wonder:***
- if boys' and girls' foot length distributions back in the two populations would have gaps at these same values. I don't expect so because I don't know any reason for this to happen.

**Overall visual comparisons *I notice:***

- there is a lot of overlap between the boys' and girls' foot lengths
- the boys' foot lengths are shifted further up the scale
- one of the girls has a recorded foot length *far* shorter than any other girl

**Shift and overlap *I notice:***

- the middle 50% of the boys' foot lengths (the box) is shifted *much* further along the scale than the middle 50% of the girls'
- there is *some* overlap for the middle 50% of the boys' right foot lengths and the middle 50% of the girls'

some of the boys have bigger right foot lengths than some of the girls and vice versa

- **that more than half the girls have footlengths shorter than the lengths of  $\frac{3}{4}$  of the boys.**

**Summary (looking at summary statistics) *I notice:***

• the boys' median foot length is bigger than the girls' median by 2cm (boys' median is 25cm, girls' median is 23cm) i.e., there is a difference in the medians of 2cm. Half of the boys have a foot length of at least 25cm

whereas half of the girls have a foot length of at least 23cm.

- the boys' median and the girls' upper quartile are the same, i.e., half of the boys have a foot length at least 25cm whereas only a quarter of the girls do

## Spread

### *I notice:*

- the middle 50% of boys have a right foot measuring between 24cm and 27cm (IQR = 3cm) whereas the middle 50% of the girls are between 22 and 25cm (IQR = 3cm). This means that the foot lengths for these boys vary by about the same amount as these girls' do.
- the boys' foot lengths went from a minimum of 21cm to a maximum of 30cm whereas the girls went from 15 to 29cm.

### *I wonder:*

- if boys' and girls' foot length distributions back in the two populations have similar variability. I expect so.

## Sampling Variation

If a new random sample of 24 13-year-old boys and a new random sample of 22 13-year-old girls were taken I would expect the plots to look different because of sampling variability. With these sample sizes, I would expect each IQR spread to change slightly and that each box would be slightly further down or up the scale. *I wonder:*

- if I repeated this sampling process many times the boys' foot lengths would, just about always, be shifted further up the scale than the girls'
- if boys tend to have a greater foot length than girls back in the two populations
- if the median foot length of boys *really* is greater than that of girls back in the two populations

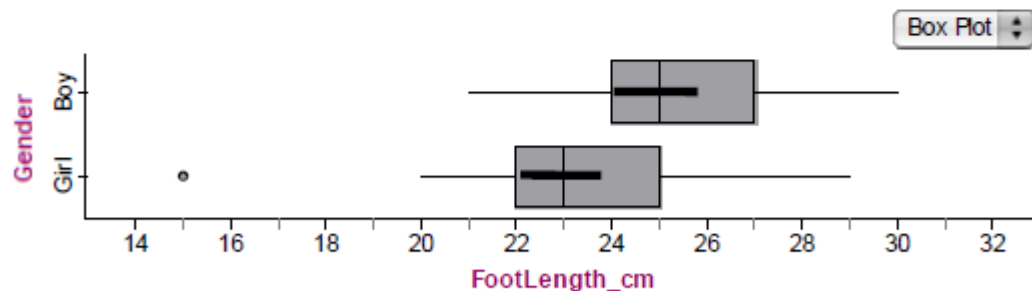
## Descriptions and inferences: comparing two samples (Level 7, Year 12)

### Sampling

If a new random sample of 24 13-year-old boys and a new random sample of 22 13-year-old girls were taken I would expect the plots to look different because of sampling variability. With these sample sizes, I would expect each IQR spread to change slightly and that each box would be slightly further down or up the scale.

### *I wonder:*

- if I repeated this sampling process many times the boys' foot lengths would, just about always, be shifted further up the scale than the girls'
- if boys tend to have a greater foot length than girls back in the two populations
- if the median foot length of boys *really* is greater than that of girls back in the populations



### *I notice:*

- that the informal confidence intervals for the population medians do not overlap

## CONCLUSION

I am going to claim that, on average, the right foot length of 13 year-old New Zealand boys is longer than the right foot length of 13 year-old New Zealand girls back in the two populations. I am prepared to make this call because, from my data, we are reasonably sure that the possible values for the boys' and girls population medians are somewhere within their respective informal confidence intervals. To make this call, with sample sizes of around 30, these informal confidence intervals for the population medians must not overlap. This is true for my data.

I don't believe that the pattern in my data of the boys' median foot length being greater than the girls' has just happened *by chance*. I am prepared to claim that this pattern in the data is *real*, i.e., that population median foot length is greater than the population median foot length for the girls.

## *Explanatory*

I expected that, on average, boys have bigger feet than girls back in the populations and the information I collected (my data) supports this belief.

I can't think of any other factor which can explain the difference in foot size other than gender.