Statistical Investigation Worksheet – IRON DATA

This activity arises out of information gleaned from a video interview with Elaine Ferguson: Is iron deficiency common among NZ infants and toddlers?
http://www.maths.otago.ac.nz/video/statistics/Iron/index.html

Other statistics videos useful for school complete with data sets and activities can also be found on the website: http://www.maths.otago.ac.nz/video/statistics/

In the late 1990s a study was undertaken in the South Island to explore iron levels in babies and toddlers (age 6-24 months). The participants were selected randomly from Christchurch, Dunedin and Invercargill (South Island Urban).

The iron, fibre, calcium and vitamin C intake per day was collected over three non-consecutive days. Haemoglobin, mean cell volume, zinc protoporphyrin and ferritin were all measured. Information such as whether the child was being breastfed, fed with formula milk or cows milk, as well as things like gender, ethnicity, maternal education, income level of household, if there were smoker(s) in the household and marital status of the mother.

From exploring the literature a number of factors were suggested that could have an effect on the levels of iron. Among these were: sex – boys at higher risk; premature babies – lower iron stores; formula fed babies – formula is fortified with iron; and cows milk – low in iron.

This gives three investigative questions to explore.
1. Do the iron levels of South Island urban boys tend to be lower than the iron levels of South Island urban girls?
2. Do the iron levels of South Island urban children who are given formula tend to be higher than the iron levels of South Island urban children who have high cows milk intake (more than 0.5 litre)?
3. Do the iron levels of children who were born prematurely tend to be lower than the iron levels of children who were not born prematurely (for South Island urban children)?

Before you analyse the sample data, predict and draw the population distributions for the ferritin levels in the questions. Show one population distribution relative to the other as suggested by the literature.

Snippet of the data table: downloaded from website.

id	hb	mcv	zpp	ferritin	age	infant	birthwt	bf	premi	girl
258.00	124.00	79.00	54.00	22.80	22.23	0.00	2,870.00	0.00	0.00	1.00
328.00	107.00	80.00	40.00	8.00	24.43	0.00	4,500.00	0.00	0.00	1.00
349.00	110.00	75.00			24.93	0.00	3,020.00	0.00	0.00	1.00
362.00	115.00	81.00	50.00	14.00	21.90	0.00	4,410.00	0.00	0.00	0.00
390.00	110.00	73.00	48.00	6.00	21.37	0.00	4,310.00	1.00	0.00	0.00
444.00	99.00	76.00	33.00	16.60	20.07	0.00	3,665.00	0.00	0.00	0.00
455.00	101.00	81.00	45.00	16.80	14.53	0.00	2,970.00	1.00	1.00	1.00
462.00	111.00	79.00	30.00	8.20	18.23	0.00	3,321.00	0.00	0.00	0.00
496.00	112.00	82.00	33.00	7.70	24.77	0.00	3,490.00	0.00	0.00	0.00
819.00	112.00	78.00	28.00	22.40	16.00	0.00	3,020.00	0.00	1.00	1.00
104.00	112.00	79.00	45.00		12.63	0.00	3,140.00	1.00	0.00	1.00
261.00	122.00	79.00	39.00	17.30	19.90	0.00	2,055.15	0.00	1.00	1.00
381.00	95.00	75.00	62.00	23.70	12.43	0.00	1,160.00	0.00	1.00	0.00
432.00	99.00	77.00	47.00	11.60	17.13	0.00	3,890.00	0.00	0.00	0.00
268.00	121.00	75.00	48.00	11.43	19.63	0.00	4,107.00	0.00	0.00	0.00
304.00	118.00	80.00	36.00	48.20	16.10	0.00	1,450.00	1.00	1.00	1.00

Complete list of variables in the data table:

Variable	Type	Description
ID	Continuous	subject ID number
hb	Continuous	haemoglobin (g/L)
mcv	Continuous	mean cell volume (fL)
zpp	Continuous	zinc protoporphyrin (μmol/mol hb)
ferritin	Continuous	ferritin (μg/l)
iron3	1=iron deficiency anaemic (IDA) (Stage3)	Children with iron deficiency anaemia with ferritin <10, hb <110, mcv < 73, zpp > 70
iron2	1=iron deficient without anaemia (ID) (Stage2)	Children with iron deficiency without anaemia with ferritin <10, hb <110, mcv<73 fL, zpp>70
iron1	1=depleted iron stores (but without IDA or ID) (Stage1)	Children with depleted iron stores (not IDA or ID) with ferritin<10, hb<110, mcv<73, zpp>70
ncrp10	1=elevated C-reactive protein (infection), 0 otherwise	to define children with infection (elevated C-reactive protein)
age	Continuous	age of child
infant	1=infant, 0=toddler	infant = 5-11.9 months of age; toddler=12-24 months of age
birthwt	Continuous	infant birth weight
bf	1=currently breastfeeding 0 otherwise	to define children who were currently breastfeeding
premi	1=born prematurely 0 otherwise	to define children who were born prematurely
curff	1=currently formula feeding, 0 otherwise	to define children who were currently formula feeding
sex	1=girl, 0=boy	sex
caucasia	1=Caucasian, 0 otherwise	ethnicity
tertiary	1=mother has tertiary level education,0 otherwise	maternal education
lowincom	1=low income (<$20,000 in 1998 & 1999)	household income level
hiincome	1=high income (>$70,000 in 1998 & 1999)	household income level
medincom	1=mid income ($20,000 to $70,000 in 1998 & 1999)	household income level
smokers	1=a smoker in the household,0 otherwise	smoker in the household
marital	1=mother in a permanent relationship,0 otherwise	marital status
nkjall	Continuous	the estimated total average energy intake per day (breast milk & food)
nfeall	Continuous	total average iron intake per day from food and breast milk
fibre	Continuous	total average fibre intake per day from food & breast milk
ca	Continuous	total average calcium intake per day from food & breast milk
vtc	Continuous	total average vitamin C intake per day from food & breast milk
milk500	1=more than 500 ml of milk per day,0 otherwise	to define children with a high milk intake (> 0.5 litre)

STATISTICAL INVESTIGATION 1

PROBLEM: Do the iron levels of South Island urban boys tend to be lower than the iron levels of South Island urban girls?

PLAN/DATA: Take a sample of 30 boys and 30 girls from the iron data used in the study. Some of these boys and girls may not have recorded ferritin levels (*μg/l*).

ANALYSIS:

Sample of boys

	ferritin
	18.579259
	2
	11.43
	16.3
	23.6
	53.1
	27

S1 = mean ( )
S2 = min ( )
S3 = Q1 ( )
S4 = median ( )
S5 = Q3 ( )
S6 = max ( )
S7 = count ( )
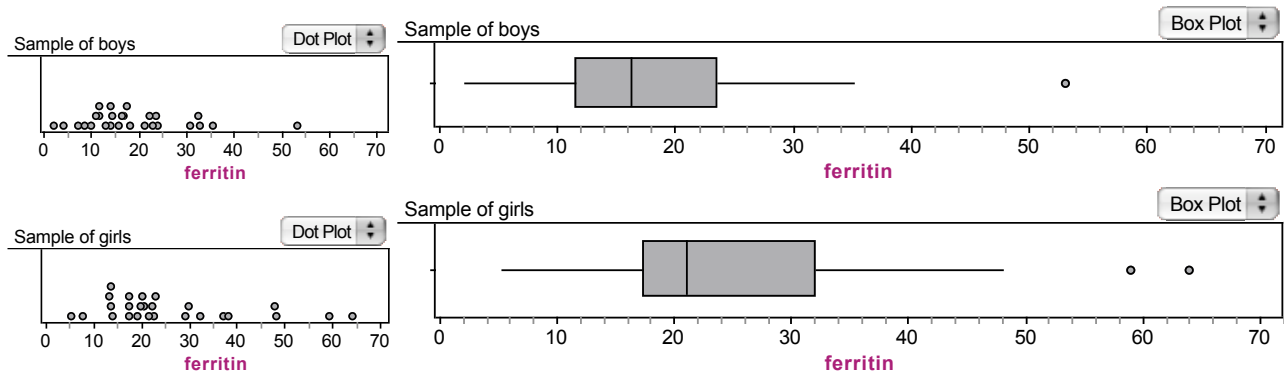
Sample of girls

	ferritin
	25.892308
	5.2
	17.3
	21
	32.1
	64
	26

S1 = mean ( )
S2 = min ( )
S3 = Q1 ( )
S4 = median ( )
S5 = Q3 ( )
S6 = max ( )
S7 = count ( )

Description of the sample distributions.

Middle 50%:
Shift: From the samples I notice…

Overlap: From the samples I notice …

Anything unusual:
From the samples I notice…

I worry or think that …

<u>Shape</u> (Describe the shape of each sample distribution, compare the shapes of the two sample distributions):
From the samples I notice…

Back in the two populations I wonder if …

<u>Spread</u> (Describe the spread of each sample distribution, compare the spreads of the two sample distributions)**:**
From the samples I notice…

Back in the two populations I wonder if …

CONCLUSION
Write a conclusion using the headings below.
Answer the problem:
"Do the iron levels of South Island urban boys tend to be lower than the iron levels of South Island urban girls?"

I would claim that …
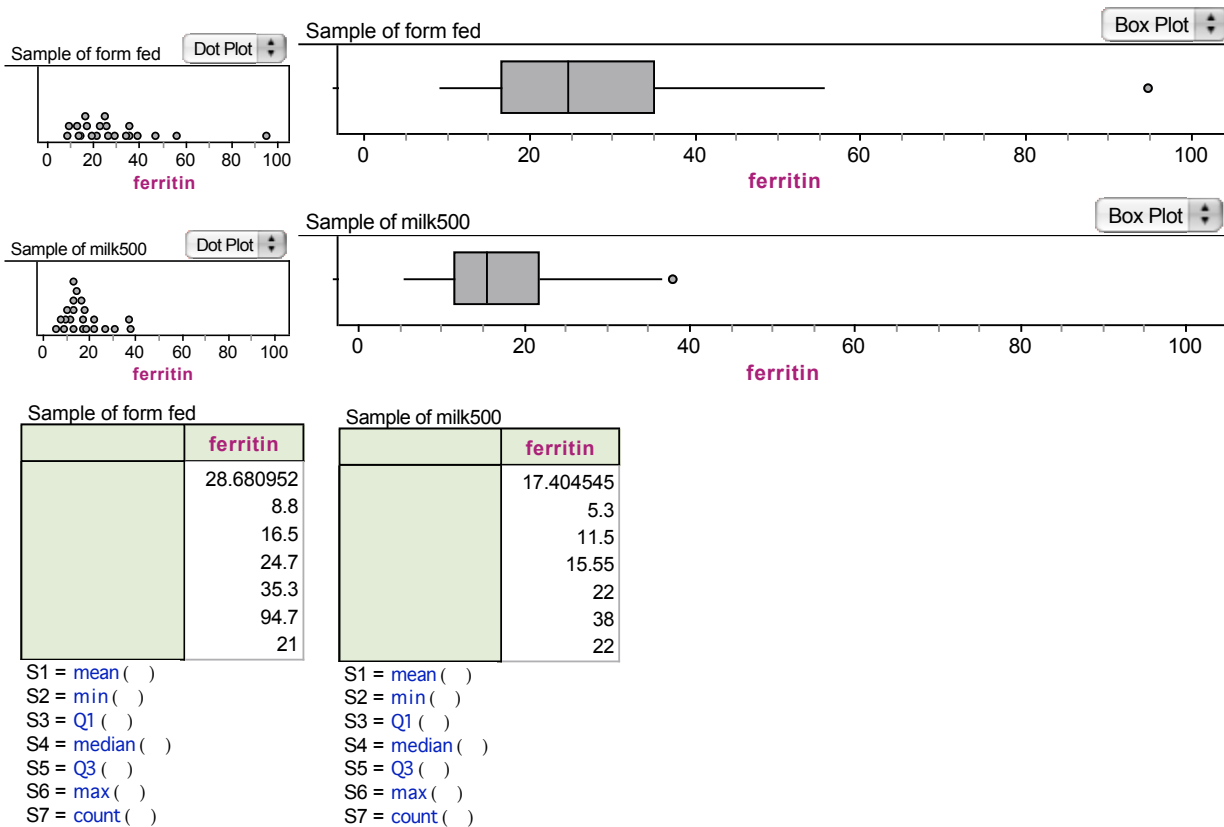
Explain why you have made this conclusion.

Is my conclusion consistent with the literature?

STATISTICAL INVESTIGATION 2

PROBLEM: Do the iron levels of South Island urban children who are given formula tend to be higher than the iron levels of South Island urban children who have high cows milk intake (more than 0.5 litre)?

PLAN/DATA: Take a sample of 30 formula fed (form fed) children and 30 high cows milk intake (milk500) children from the iron data used in the study. Some of these children may not have recorded ferritin levels ($\mu g/l$).

ANALYSIS:

Sample of form fed — Dot Plot / Box Plot (ferritin)

Sample of milk500 — Dot Plot / Box Plot (ferritin)

Sample of form fed

	ferritin
	28.680952
	8.8
	16.5
	24.7
	35.3
	94.7
	21

S1 = mean (   )
S2 = min (   )
S3 = Q1 (   )
S4 = median (   )
S5 = Q3 (   )
S6 = max (   )
S7 = count (   )

Sample of milk500

	ferritin
	17.404545
	5.3
	11.5
	15.55
	22
	38
	22

S1 = mean (   )
S2 = min (   )
S3 = Q1 (   )
S4 = median (   )
S5 = Q3 (   )
S6 = max (   )
S7 = count (   )

Description of the sample distributions.

Middle 50%:
Shift: From the samples I notice…

Overlap: From the samples I notice …

<u>Anything unusual:</u>
From the samples I notice…

I worry or think that …

<u>Shape</u> (Describe the shape of each sample distribution, compare the shapes of the two sample distributions):
From the samples I notice…

Back in the two populations I wonder if …

<u>Spread</u> (Describe the spread of each sample distribution, compare the spreads of the two sample distributions)**:**
From the samples I notice…

Back in the two populations I wonder if …

CONCLUSION
Write a conclusion using the headings below.
Answer the problem:
"Do the iron levels of South Island urban children who are given formula tend to be higher than the iron levels of South Island urban children who have high cows milk intake (more than 0.5 litre)?"

I would claim that …
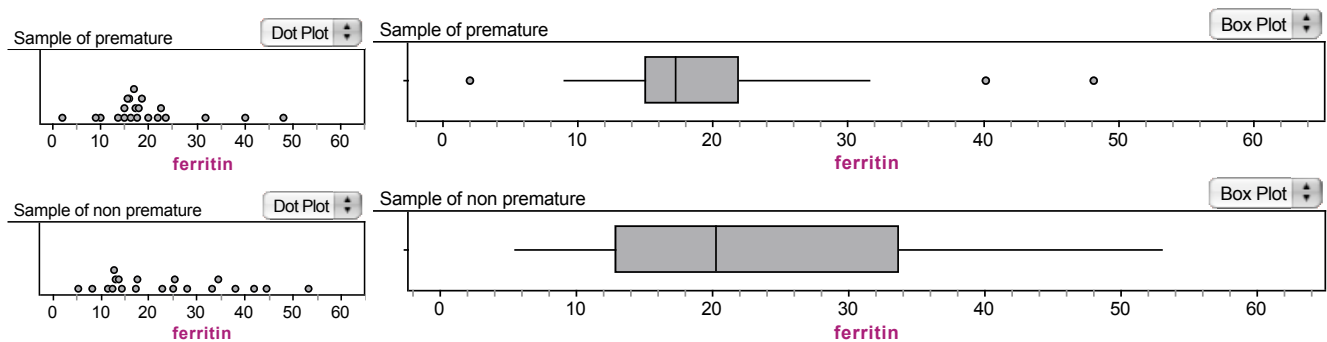
Explain why you have made this conclusion.

Is my conclusion consistent with the literature?

STATISTICAL INVESTIGATION 3

PROBLEM: Do the iron levels of children who were born prematurely tend to be lower than the iron levels of children who were not born prematurely (for South Island urban children)?

PLAN/DATA: Take a sample of 30 children who were premature babies and 30 who were non-premature babies from the iron data used in the study. Some of these children may not have recorded ferritin levels ($\mu g/l$).

ANALYSIS:

Sample of premature	
	ferritin
	19.461905
	2
	15
	17.3
	22
	48.2
	21

S1 = mean (   )
S2 = min (   )
S3 = Q1 (   )
S4 = median (   )
S5 = Q3 (   )
S6 = max (   )
S7 = count (   )

Sample of non premature	
	ferritin
	23.61
	5.3
	12.8
	20.25
	33.75
	53.1
	20

S1 = mean (   )
S2 = min (   )
S3 = Q1 (   )
S4 = median (   )
S5 = Q3 (   )
S6 = max (   )
S7 = count (   )

Description of the sample distributions.

<u>Middle 50%:</u>
Shift: From the samples I notice…

Overlap: From the samples I notice …

<u>Anything unusual</u>:
From the samples I notice …

I worry or think that …

<u>Shape</u> (Describe the shape of each sample distribution, compare the shapes of the two sample distributions):
From the samples I notice…

Back in the two populations I wonder if …

<u>Spread</u> (Describe the spread of each sample distribution, compare the spreads of the two sample distributions)**:**
From the samples I notice…

Back in the two populations I wonder if …

CONCLUSION
Write a conclusion using the headings below.
Answer the problem:
"Do the iron levels of children who were born prematurely tend to be lower thanthe iron levels of children who were not born prematurely (for South Island urban children)?"

I would claim that …

Explain why you have made this conclusion.

Is my conclusion consistent with the literature?