**Laying Foundations for Inference (Year 10)**

**KareKare College:** 616 students (389 females, 227 males)

**Data card variables:**

 Card colour indicates gender

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

**C@S Survey questions for the variables.**

1. Are you: male; female

2. How old are you?
 (Note: year level was by default, this was automatically entered with the class entry.)

4. Which ethnic group do you belong to? Mark the space or spaces which apply to you.

 New Zealand European; Māori; Samoan; Cook Island Maori; Tongan; Niuean; Chinese; Indian;

 Other such as DUTCH, JAPANESE, TOKELAUAN. Please state:

7. How tall are you without your shoes on? Answer to the nearest centimetre.

12. What is your popliteal length? Answer to the nearest centimetre. (The popliteal length is the

 measurement from the underside of the leg right behind the knee when seated, to the floor.

 Taken with shoes off.)

13. What is the length of your index finger to the nearest millimetre?

14. What is the length of your ring finger to the nearest millimetre?

15. What is the main method of transportation that you usually use to get to school?
 (Pick one of the following: walk; car; bus; train; bike; boat; other)

16. How long does it usually take you to get to school? Answer to the nearest minute.

17. What is the weight of your school bag today? Answer to the nearest 100g.
 (Weigh your school bag with all your books and other materials you brought to school today.)

18. How do you usually carry your school bag: (Two straps, one on each shoulder; One strap, over

 one shoulder; One strap, going diagonally across your body; Hold in your hand)

27. How physically fit do you think you are? (Unfit; A little bit fit; Quite fit; Very fit)

**Summary Question:**

 What are typical popliteal lengths for the students at Karekare College?

**Comparison Questions:**

*At Karekare College, do boys tend to be taller than the girls?*

*At Karekare College, who tends to take a longer time to get to school; students who walk or students who travel by bus?*

**Statistical Investigation Exemplar – 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 |

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:





Description of the sample distributions.

**Middle 50%:**

**Shift:** From the samples I notice…

*that the ferritin levels of these girls are shifted further up the scale than the ferritin levels of these boys.*

**Overlap:** From the samples I notice …

*that there is some overlap of ferritin levels between these two groups.*

**Anything unusual:**

From the samples I notice…

 *that one of these boys has an unusually high ferritin level.*

I worry or think that …

 *I worry about the boy who has an unusually high ferritin level about this and should try to check it.*

**Shape** (Describe the shape of each sample distribution, compare the shapes of the two sample distributions):

From the samplesI notice…

*\* that the ferritin levels of both these groups are a mound shape. The mound for these boys is around 15 μg/l and around 20 μg/l for these girls.*

*\* that the ferritin levels of both these groups are slightly skewed to the right.*

Back in the two populations I wonder if …

*the shapes will be like these, I expect so. (this is based on the predicted population distributions)*

**Spread** (Describe the spread of each sample distribution, compare the spreads of the two sample distributions)**:**

From the samplesI notice…

*\* that the middle 50% of ferritin levels for these girls are slightly more spread out than the middle 50% of these boys.*

\* *these girls’ ferritin levels IQR = 32.1 - 17.3 = 14.8 μg/l*

*\* these boys’ ferritin levels IQR = 23.6 – 11.43 = 12.17 μg/l*

Back in the two populations I wonder if …

*the spreads will be like these, that is, similar for boys and girls. I expect so. (this is based on the predicted population distributions)*

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 …

*ferritin levels of girls tend to be higher than ferritin levels of boys.*

Explain why you have made this conclusion.

*This shift is big with only a small amount of overlap. The boys’ median is below the girls’ middle 50%. If I were to take new samples I would expect to make the same claim that is that ferritin levels of girls tends to be higher.*

Is my conclusion consistent with the literature?

*My conclusion confirms / ~~does not confirm~~ that boys tend to have a higher risk of having lower iron levels than girls.*

**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 (*μg/l*).

ANALYSIS:





Description of the sample distributions.

**Middle 50%:**

**Shift:** From the samples I notice…

*that the ferritin levels of these formula fed children are shifted further up the scale than the ferritin levels of these high cows milk intake children.*

**Overlap:** From the samples I notice …

*that there is some overlap of ferritin levels between these two groups.*

**Anything unusual:**

From the samples I notice…

 *One of these formula fed children has unusually high ferritin levels.*

I worry or think that …

*I worry that this may be a measurement mistake.*

**Shape** (Describe the shape of each sample distribution, compare the shapes of the two sample distributions):

From the samplesI notice…

*\* that the ferritin levels of both these groups are a mound shape. The mound for these formula fed children is around 20 μg/l and around 15 μg/l for these high cows milk children.*

*\* that the ferritin levels of both these groups are slightly skewed to the right.*

Back in the two populations I wonder if …

*the shapes will be like these, I expect so. (this is based on the predicted population distributions)*

**Spread** (Describe the spread of each sample distribution, compare the spreads of the two sample distributions)**:**

From the samplesI notice…

*\* that the middle 50% of ferritin levels for these formula fed children are more spread out than the middle 50% of these high cows milk children.*

\* *these formula fed children ferritin levels IQR = 35.3 - 16.5 = 18.8 μg/l*

*\* these high cows milk children ferritin levels IQR = 22 – 11.5 = 10.5 μg/l*

Back in the two populations I wonder if …

*the spreads will be like these, I don’t know. (this is based on the predicted population distributions)*

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 …

*ferritin levels of formula fed children tend to be higher than ferritin levels of children who have high cows milk intake.*

Explain why you have made this conclusion.

*This shift is big with only a small amount of overlap. The median of each of these groups is outside the box of the other group. If I were to take new samples I would expect to make the same claim that is that ferritin levels of formula fed children tend to be higher.*

Is my conclusion consistent with the literature?

*My conclusion confirms / ~~does not confirm~~ that formula fed children tend to have higher ferritin levels than children who have high cows milk intake (as cows milk is known to be low in iron).*

 **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 (*μg/l*).

ANALYSIS:





Description of the sample distributions.

**Middle 50%:**

**Shift:** From the samples I notice…

*that the ferritin levels of these non-premature children are shifted slightly further up the scale than the ferritin levels of these premature children.*

**Overlap:** From the samples I notice …

*that ferritin levels of these non-premature children completely overlaps the ferritin levels of these premature children.*

**Anything unusual:**

From the samples I notice … *Nothing unusual.*

I worry or think that …

**Shape** (Describe the shape of each sample distribution, compare the shapes of the two sample distributions):

From the samplesI notice…

*\* that the ferritin levels of both these groups are a mound shape. The mound for these premature children is around 18 μg/l and around 15 μg/l for these non-premature children.*

*\* that the ferritin levels of both these groups are slightly skewed to the right.*

Back in the two populations I wonder if …

*the shapes will be like these, I expect so. (this is based on the predicted population distributions)*

**Spread** (Describe the spread of each sample distribution, compare the spreads of the two sample distributions)**:**

From the samplesI notice…

*\* that the middle 50% of ferritin levels for these non-premature are a lot more spread out than the middle 50% of these premature.*

\* *these premature ferritin levels IQR = 22 - 15 = 7 μg/l*

*\* these non-premature ferritin levels IQR = 33.75 – 12.8 = 20.95 μg/l*

Back in the two populations I wonder if …

*the spreads will be like these. I don’t know. (this is based on the predicted population distributions)*

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 …

*I am unable to make a call as to which group (premature or non-premature) has the highest iron level back in the population.*

Explain why you have made this conclusion.

*This shift is not big enough; there is a large overlap. Both medians are within the overlap. If I were to take new samples I could easily get the medians the other way around.*

Is my conclusion consistent with the literature?

*My conclusion ~~confirms~~ / does not confirm that children born prematurely have lower iron levels than children who were not born prematurely.*

|  |
| --- |
| **Key Ideas from Workshop** * Appreciating sampling variability
* Considering shift and overlap
* Making a call
* Justify the call

All handouts and classroom materials from Workshop are located on CensusAtSchool:<http://www.censusatschool.org.nz/2009/informal-inference/> |