Statistics Education Early 21st Century New Zealand

Pip Arnold, Team Solutions, The University of Auckland

- **What are the big ideas in Statistics Education in the early 21st Century…**

PROBLEM

- My theory is that there is *question posing* and *question asking*.
- Question posing results in a question being formally structured, whereas question asking is a continual spontaneous interrogative process.

Posing questions

- Question posing arises as a result of having a problem that needs to be addressed using a statistical investigation.
- Investigative question - The question being asked of the data
- Survey question - The question asked to get the data

Asking questions

- The *interrogative* questions – questions that are asked as checks within the cycle (the problem, the plan, the data (given data sets), the analysis, the conclusion);
- The *analysis* questions that are asked about the statistics, graphs and tables to develop a description of and an inference about what is noticed (the analysis).

Types of Investigative Questions

- Summary - A description of the data, usually a single data set
- Comparison - Comparing two (or more) sets of data across a common variable
- Relationship - Interrelationship between two paired variables

Questioning the question

- What was the original question that was used to collect the data (survey question)?
- What type of data is being used?
- What graph or display of the data will be made?
- What hypothesis can be made about the data?
- Who would be interested in the answers to this question?
- Is there enough data available to answer the question (issues around sample size)?
- What background information is available about the data (how it was collected, who it was collected from, when it was collected etc.)?
- Is the variable of interest in the dataset?
- **Is the *investigative* question "right"?**

What makes a "good" question?

- can be answered with the data: sample size, variable(s) available
- population of interest is clear: a, the, nothing
    - sample v target population
    - Indvidual v aggregate
- variable(s) of interest is/are clear
- intent is clear: summary, comparison, relationship; What is typical? is clear
- one that we are interested in the answer

PLAN

- Planning where the data is collected.
  - Happens once the problem is set.
- Planning where the data is given.
  - Happens as part of the problem setting process.
  - Exploring the planning that was done.
  - Understanding the situation.

Reaction times – sleeping sheep

http://aucksecmaths.wikispaces.com/Statistics+Workshop+2008

Draft activity, includes teacher notes, workshop plan and handouts.

- Problem: Do my reaction times tend to be faster than my partners?
- Plan: How many measures will they take?
  - Will they allow a practice run?
  - How will they deal with penalty shots?
  - How do we know everyone has done the same trialling?
  - Will we collect individual results or averages?

Being typical

http://aucksecmaths.wikispaces.com/Statistics+Workshop+2008

The growing scatterplots activity has the cards. Further ideas in data and dots.

- Take a few data cards each.
- Discuss in pairs/groups what you think the variables might be.
  - Are you: male/female
  - What is your wrist circumference in cm?
  - What is your neck circumference in cm?
  - Year level was by default - based on teacher entry.
  - How old are you?
- Describe two people from the cards to your neighbour, **write** the description for one of the cards down!

- This student is a girl. She is in Year 9 and is 15 years of age. Her wrist circumference is 15cm and her neck circumference is 10cm.
  - I suspect there is a problem with one of the measures as the neck circumference should be larger than the wrist circumference.
- This student is a boy. He is in Year 9 and is 14 years of age. His wrist circumference is 15cm and his neck circumference is 30cm.
  - Draw a data card to show this student's data.

- Students need to understand that each card is a person.
- They need to become familiar with the context.
- They need to know what the questions were that were asked to get the data.

DATA

Websites for data

•CensusAtSchool NZ: http://www.censusatschool.org.nz/
•CensusAtSchool International: http://www.censusatschool.ntu.ac.uk/
•Statistics NZ: http://www.stats.govt.nz/schools-corner/default.htm
•Exploring data: http://exploringdata.cqu.edu.au/datasets.htm

CensusAtSchool

•2003: 18,000 participants; 682 schools; 26 variables; Year 5-10
•2005: 33,205 participants; 18,516 girls; 14,689 boys; 32 variables; Year 5-10
•2007: 25,007 participants; 13,578 girls; 11,429 boys; 31 variables; Years 5-13

CensusAtSchool 2009
3 March 2009 until 9 April 2009

Register online: http://www.censusatschool.org.nz/2007/register/
•If you have previously registered, you won't need to register again. You should have already received an email to confirm this.

The more data we get, the more data you have, get your classes involved in 2009.

New for 2009

•Teachers will get their class results back if they choose.
•Early in year so 2009 data can be used for 2009 teaching.

Cleaning Data

•Sort the data you have in your bag.
•Identify any data that you think is unusual.
  – What makes it unusual?
  – Do you think it is "correct" data? Why? Why not?
  – What will you do with the data that is "not correct"?

Sampling variation

•CensusAtSchool and other databases provide an opportunity to give students different samples from the same population.

Sampling

•Can use Fathom to take a sample from a hidden population.
•Household SURF.

Categorical variables: sex, partnered, highest qualification gained, employment status, ethnicity.
Numerical variables: age of respondent from 20 - 49 years, age of partner, (individual) wage or salary, total income (for individual or couple), total debt (for individual or couple), total net worth (for individual or couple).

ANALYSIS
CensusAtSchool

- Simple analysis tools: Data viewer, Summary tables
- New for 2009: PPDAC structured worksheet (online)

Ideas for summary

- Use the starter *I notice…*
- Describe
  - Shape
  - Spread
  - Middle group(s)
  - Anything unusual
- Remember context. If I cover any labels can I still tell what the graphs are showing.

Ideas for comparison

- Use the starter *I notice…*
- Describe
  - Shape
  - Spread
  - Middle 50%
  - Summary statistics
  - Anything unusual

See handout. http://www.censusatschool.org.nz/2008/informal-inference/
Handout 2: A teacher's guide to informal comparative reasoning by Pfannkuch, Wild, Horton and Regan (240 Kb pdf file)

Analysis

- Need to model.
- Need to give structure.
- Need to remember context.
- Need to do in class.
- Need to describe not just answer question.

Analysis

- Inference.
- Chris Wild's talk. http://www.censusatschool.org.nz/2008/informal-inference/
A talk on the subject of informal statistical inference given by Chris Wild to the Auckland Mathematical Association on 14 June 2008.

CONCLUSION

- Some starting points for conclusions.
  - Answer the question.
  - Provide supporting evidence from your analysis.
  - Generalise to the population of interest.

What are the big ideas in Statistics Education in the early 21st Century…

- **Problem** is critical.
  - Variables of interest
  - Population of interest
  - Intention
  - Able to be answered by the data
  - Answer is of interest to someone
- Don't hurry this part. Take time to establish the context - link to planning for given data sets.

- **Plan**ning is an often over looked aspect of the cycle.
- With given data sets it is important to provide background information about how the data was collected; what were the survey questions, who was surveyed.
- Context - students need to be familiar with the context.

- **Data**
- Use real multivariate data sets.
- Give groups different samples from the same population. Compare and contrast outcomes.

- **Analysis** is about describing the data.
- *I notice…*
- What is it that the data tells you?
- Look at the shape, spread, middle group, anything unusual.
- Remember to relate to the context.
- Make inferential statements when you can.

- **Conclusion**
- Answer the problem.
- Give supporting evidence.
- Generalise to the population.

MODEL, MODEL, MODEL