Building Inferential Reasoning in Statistics

Chris Wild and **Maxine Pfannkuch**
Dept of Statistics, University of Auckland
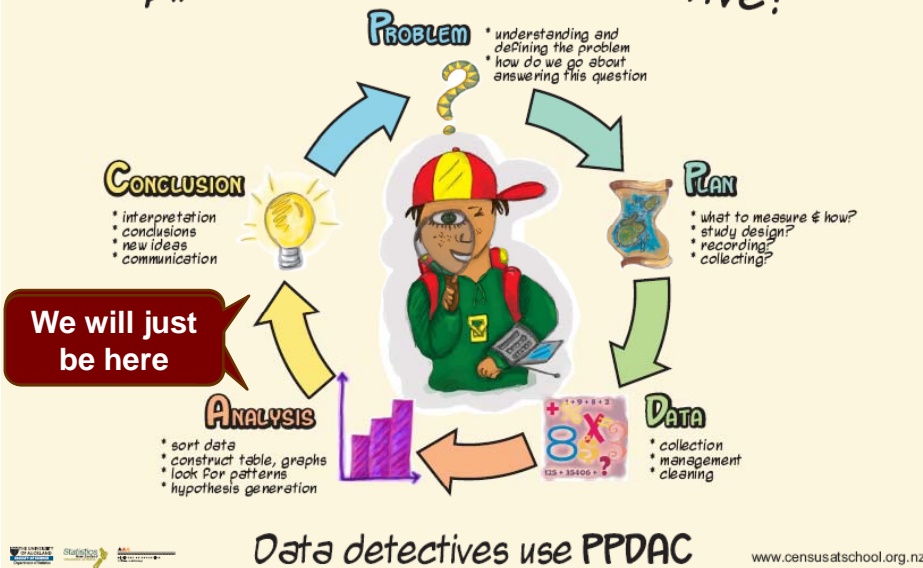New Zealand

---
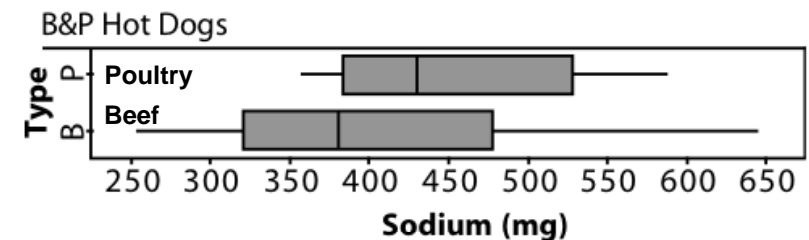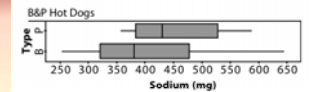
Part 1: Maxine

On Description, Inference
& the Game of Statistics

Are You a Data Detective?

PROBLEM
* understanding and defining the problem
* how do we go about answering this question

PLAN
* what to measure & how?
* study design?
* recording?
* collecting?

DATA
* collection
* management
* cleaning

ANALYSIS
* sort data
* construct table, graphs
* look for patterns
* hypothesis generation

CONCLUSION
* interpretation
* conclusions
* new ideas
* communication

We will just be here

Data detectives use PPDAC

www.censusatschool.org.nz

Discussion - Part 1

B&P Hot Dogs

Type
Poultry (P)
Beef (B)

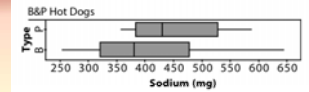Sodium (mg)
250 300 350 400 450 500 550 600 650

- Investigators analysed the sodium content of some beef (B) hot dogs and some poultry (P) hot dogs. This is their graph (Franklin et al., 2007).
- **Discuss in pairs what statements you would make from these plots.**

Slide 1

Discussion - Part 1

B&P Hot Dogs

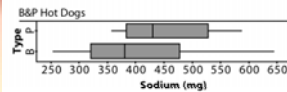Type P: Poultry
Type B: Beef

Sodium (mg)
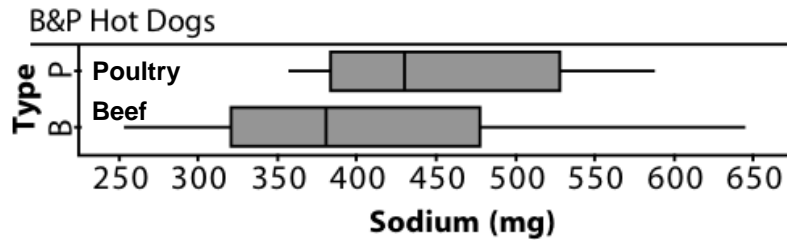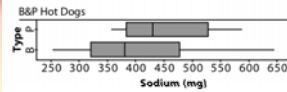
- What sort of statements did you **want** to make?
- What statements **did** you make?
  - Did you **describe** what you can see?
  - Did you **infer** about …?
 - Did you consider **sample sizes**?
    - Does it matter?

Slide 2

Distribute handouts

Slide 3

Discussion - Part 2

Statements from the box plots	Descriptive or Inferential?
The median sodium content for poultry hot dogs is 430 mg, almost 50mg more than the median sodium content for beef hot dogs	
The medians indicate that a typical value for the sodium content of poultry hot dogs is greater than a typical value for beef hot dogs	
The range for the beef hot dogs is 392 mg, versus 231 mg for the poultry hot dogs	
The ranges indicate that, overall, there is more spread (variation) in the sodium content of beef hot dogs than poultry hot dogs	
The IQRs for sodium content are 157.5 mg for beef hot dogs and	

- The investigators made statements from the plots.
  - Note for beef hot dogs n=20, for poultry n=17
- In pairs decide whether each statement is *descriptive* or *inferential* (see handout)
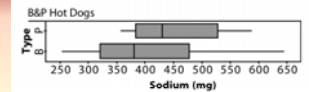
Slide 4

Discussion - Part 2

Descr/Inf

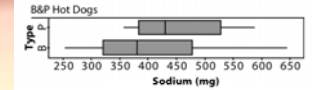|	Descr/Inf
The median sodium content for poultry hot dogs *is* 430 mg, almost 50mg more than the median sodium content for beef hot dogs	D
The medians *indicate* that a typical value for the sodium content of poultry hot dogs is greater than a typical value for beef hot dogs	I
The range for *the* beef hot dogs *is* 392 mg, versus 231 mg for *the* poultry hot dogs	D
The ranges *indicate* that, overall, there is more spread (variation) in the sodium content of beef hot dogs than poultry hot dogs	I

Discussion - Part 2

B&P Hot Dogs
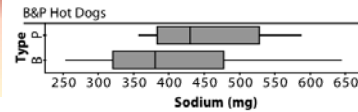
- **Statement**
  - *Considering the degree of variation in the data and the amount of overlap in the box plots, a difference of 50 mg between the medians is not really that large*
- What is this statement about?
  - Practical importance?
  - Is it an intuitive t-test?
  - Considering sample sizes? Sampling variability?
- Can we expect students to make such a statement?
 - Under what circumstances?

Discussion - Part 2

B&P Hot Dogs

- What is the point of the hot dog comparison?
  - To describe?
  - To "infer"?
    - To draw a conclusion, make a decision?
 - To "do a comparison" ?

Discussion - Part 3

B&P Hot Dogs
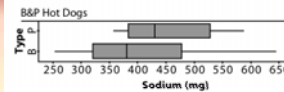
In pairs discuss the following questions (see handout).

- What statistical game is played in the Year 11 NZ classroom (NCEA Level 1 – pose a question …?)
  - Is the question about the data collected or is the purpose of the question to make a decision about some wider universe?
  - If Year 11 NZ students posed a question such as "Do poultry hot dogs tend to have a higher sodium content level than beef hot dogs?", **on what basis would your students make a decision**?
  - Are Year 11 NZ students aware of concepts such as *sample*, *population*, *sample distribution*, *population distribution*, *sampling variability*?

Should students be doing inference?

B&P Hot Dogs
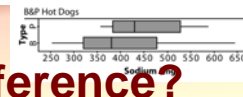
- Research
- Everyday life (political polls etc.)
- If not, all rather pointless
- Therefore must provide a learning pathway
  - to introduce some of the "big ideas" behind inference
  - to allow students to "conceive of samples and sampling
    - in ways that support their developing coherent understandings of why statisticians have such confidence in this practice" (Saldanha & Thompson, 2002, p. 268).

What is the statistical game?
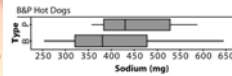
- Statistics is about
  making decisions in the face of uncertainty

- Statistical inference
  "moves beyond the data in hand to draw conclusions about some wider universe, taking into account that variation is everywhere and the conclusions are uncertain" (Moore, 2004)

How can NZ students play the statistical game?
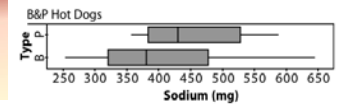
1. **Know what the game** is about and why it is played.
2. **Give structure**
   - "I notice …" ↔ *descriptive thinking*
   - "I wonder …" ↔ *inferential thinking*
     what might be happening back in the populations
3. **Appreciate sampling variability**
   - hands-on, computer-based visual experiences
   - track the variability (imagery)
4. **Develop "decision rules"** to enable them to *make a call* on whether one group tends to give bigger values than the other group
   - base "decision rules" on things like *t*-statistic, confidence intervals
   - introduce underlying concepts - centres, variability, sample sizes, …

Part 2: Chris

On Informal Statistical Inference