The Use and Misuse of Statistics

MAP4C
You have been invited to try out for the debate team. Your debate position is as follows: It would be better for families if women stayed at home rather than joined the workforce.

What would you do to prove this statement true? Where would you collect data and how would you display it to show that this is true?
This data was collected by Statistics Canada. Does this support the statement that follows? As the number of women in the workforce increases, the number of single-parent families increases. If we want to reduce the number of single-parent families in Canada, women should stay home and not go out to work.
Valid Conclusions??

Ask yourself the following questions to ensure that the study was valid and the conclusion has merit.

- Is there any bias in the data collection?—The way the sample was selected, questions, way it was conducted?
- If data has measurements – are they accurate?
- Are any graphs drawn accurately or do they mislead the viewer? – look, scale, etc.
Assessing graphs

Choose the graph that displays the data more accurately. Justify your choice.

Canada’s Population by age according to the 2001 census

i) Ages of Canadians, 2001 Census

- 0 to 14: 19%
- 15 to 24: 20%
- 25 to 44: 24%
- 45 to 64: 20%
- 65 to 84: 11%
- 85 and older: 6%

ii) Ages of Canadians, 2001 Census

- 0 to 14: 19%
- 15 to 24: 20%
- 25 to 44: 24%
- 45 to 64: 20%
- 65 to 84: 11%
- 85 and older: 6%
ii displays the graph more accurately. Generally, using 3–Dimensional graphs makes some pieces appear larger than others.
A company’s profits over a 5 year period.
The first one is the better graph. The scale has been adjusted to make the profits look better in number ii. The first graph has a symbol in the scale to alert the reader that it has been adjusted.
What is wrong with this headline?
Four grade 9 students collected the following data on lunch preferences. Their conclusions were: We asked students to tell us their favourite lunch meals and displayed the results in this bar graph. We conclude that the school cafeteria should serve more pizza since it is clearly the favourite lunch of students.
IS THEIR CONCLUSION VALID??

In groups of 2 discuss the following:

- Sample size
- Was sample representative?
- Was the survey question biased?
- How was the survey conducted?
- Is the graph constructed accurately?
Our Conclusion:

Their conclusion is not valid (it may still be true). You need more information about the sample and size of sample as well as the survey itself.
Cause and Effect

This is referring to two-variable data. Look at the graph below. Students performed a linear regression on data collected from Stat. Canada on the number of seniors and the number of weapons crimes in Canada (see if there is a relationship between the two)

--What is your hypothesis?
There is a strong positive correlation between the two variables. As the number of seniors increases, weapons charges increase. Therefore, criminals in Canada are becoming bolder because of our ageing population. (Is this a valid conclusion given the data?)
Questions to Ask

- Was there bias in the data?
- Is the graph constructed accurately?
- Is the correlation strong?
- Does the analysis support a cause-and-effect relationship? Does one variable CAUSE the other variable to change?
  - What other conclusion could you have made looking at the data?
Types of statistical relationships

- What is a correlation?
  - two variables appear to be related
  - i.e., a change in one variable is associated with a change in the other
  - e.g., salary increases as age increases

- What is a causal relationship?
  - a change in one variable is proven to cause a change in the other
  - usually requires an in-depth study i.e.
  - e.g., incidence of cancer among smokers
  - e.g., hours of study per week with the approach of exams
**Example**

- Ice Cream sales in the United States and the frequency of Shark attacks increase in the month of July.
  - Correlation or Causal?
- The size of a child’s vocabulary increases with the number of cavities the child has while in elementary school.
  - Correlation or Causal?
Homer: Not a bear in sight. The "Bear Patrol" is working like a charm!

Lisa: That's specious* reasoning, Dad.

Homer: [uncomprehendingly] Thanks, honey.

Lisa: By your logic, I could claim that this rock keeps tigers away.

Homer: Hmm. How does it work?

Lisa: It doesn't work; it's just a stupid rock!

Homer: Uh-huh.

Lisa: But I don't see any tigers around, do you?

Homer: (pause) Lisa, I want to buy your rock.

* plausible but false, deceptively pleasing

http://en.wikipedia.org/wiki/Correlation_implies_causation_(logical_fallacy)
The idea that correlation implies causation is a wrong way of thinking that many people fall into. We can never assume that because two things are correlated that one causes the other. This information must be supplied by a theory that we have about the variables.
Why does this happen?

- A third event may be the cause of the correlation
  - My age and the price of Coke
  - Wearing a skirt and breast cancer
  - Cavities and vocabulary
- The direction of the causation might be reversed
- It might be a coincidence (tiger attacks and the rock)
Types of Causal Relationships

- **Cause and Effect Relationship**
  - A change in $x$ leads to a change in $y$.

- **Common Cause Factor**
  - An external variable causes both variables to change.

- **Accidental Relationship**
  - There is no causation – no relationship

- **Presumed Relationship**
  - The relationship does not seem accidental, but it is difficult to show one variable affects the other.
Do females seem more likely to be interested in student government?

Does gender appear to have an effect on interest in student government?

Is this a correlation?

Is it likely that being female causes interest?
A survey as reported in a British newspaper involved questioning a group of teenagers about their behaviour, and establishing whether their parents smoked.

The newspaper reported as fact that children whose parents smoked were more likely to exhibit delinquent behaviour.

The results seemed to show a correlation between the variables, so the paper printed a headline “Parental Smoking Causes Children to Misbehave”

Would you agree with the media report?
Homework:

p.229 # 1,2,3,4,7,8,10,12