

**4.3 The Use and Misuse of Statistics**

Learning Goals: I am learning to...

- Assess data presented in graphical form
- Assess data based on how it was collected
- Assess data based on cause and effect assumptions



A **valid conclusion** is one that is supported by unbiased data that has been interpreted appropriately.

When you read a conclusion, someone has made based on statistics, you must decide whether the conclusion is valid or not. To do this, ask yourself:

- Is there any bias in the data collection, in the way the
  - Sample was selected? (random vs. non-random)
  - Questions were phrased or asked? (leading / loaded)
  - Survey was conducted ? (anonymous or not)
- If the data involved measurements, were they accurate?
- Are any graphs drawn accurate or do they mislead the viewer?

**Part A: Assessing Graphs**

The graphs in each pair show the SAME DATA. Choose the graph that displays the data MORE accurately. Justify your choice.

1. The following graphs both compare men and women's weekly income.

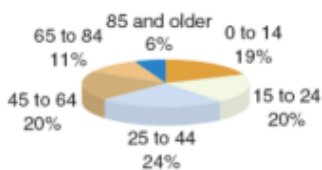


**Justification:**

Graph (a) was better because the scale on the y-axis is started at 0. Graph (b) makes it look like there is a bigger gap between men and women.

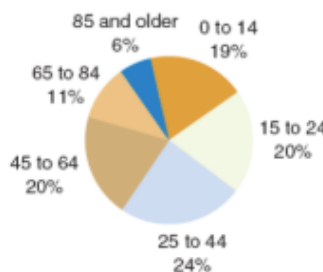
2. The following pie charts show Canada's population by age, according to the 2001 census.

i) Ages of Canadians, 2001 Census



OR

ii) Ages of Canadians, 2001 Census

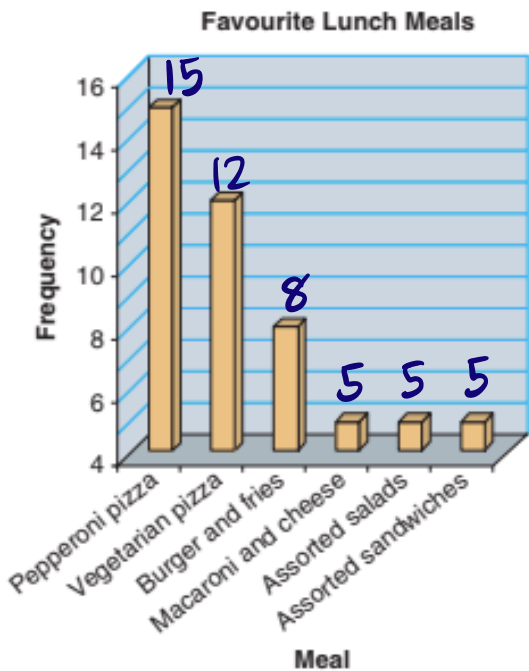


**Justification:**

Graph (ii) is better as it is a 2D representation rather than 3D. The 3D graph makes some sections appear larger than others.

**Part B: Assessing How Data was Collected and Graphed**

Four Grade 9 students collected data on school lunch preferences. Is their conclusion valid?



Their conclusion:

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.

In order to decide if a conclusion is valid, ask yourself the following questions.

- Was the sample size appropriate?
- Was the sample representative?
- Was the survey question biased?
- How was the survey conducted?
- Is the graph constructed accurately?

1. Was the sample size appropriate?

50 students were surveyed. Depending on the size of the school this may not be large enough.

2. Was the sample representative?

We don't know how they chose their sample. We don't have enough information to make an informed conclusion.

3. Was the survey question biased?

The question was unbiased. They just asked students their favourite lunch food.

4. How was the survey conducted?

Most likely was conducted verbally. They could have just asked people they know → bias. People may not have been given many options in their response.

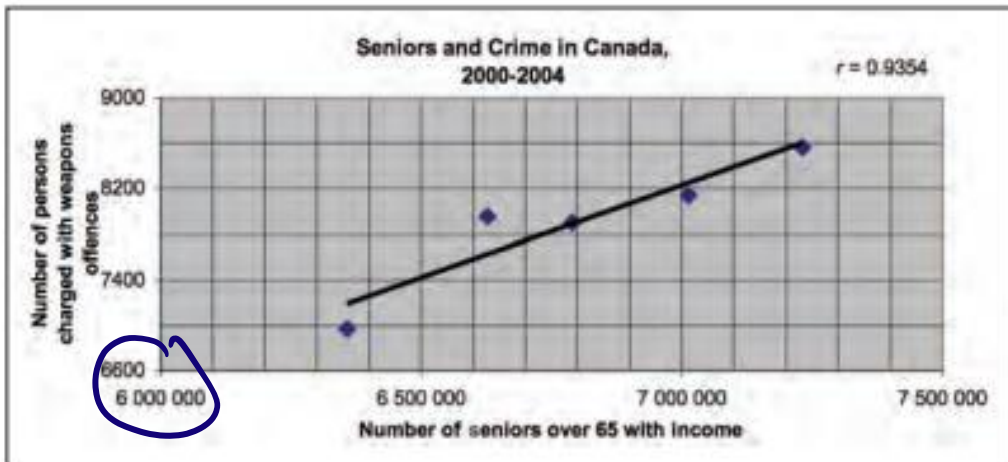
5. Is the graph constructed accurately?

No, the graph is 3D, which makes it hard to read. The y-axis does not start at 0.

Overall the conclusion is not valid. We need to know more information.

**Part C: Assessing Assumptions About Cause and Effect**

A group of Grade 12 students performed a linear regression on data they collected from Statistics Canada about the number of seniors and the number of weapons crimes in Canada. Is their conclusion valid?



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.

Conclusion:  
To assess the validity of the conclusion, ask yourself the following questions:

- Was there bias in the data collection?
- Is the graph constructed accurately?
- Is the correlation strong?
- Does the analysis support a cause-and-effect relationship?

1. Was there bias in the data collection?

The source of data was good, but there were not enough data points.

2. Is the graph constructed accurately?

Both the x and y axis don't start at 0, nor do they include a break.

3. Is the correlation strong?

strong  $\rightarrow$  moderate, but it is hard to conclude on because there are few points.

4. Does the analysis support a cause-and-effect relationship?

Not really, both variables are likely increasing because of an overall increase in population. There could also have been a reason based on the time period (2000-2004)