

Formulating an Argument Using Statistics

by L. Shorser

In order to make an argument that makes use of statistics, one must decide what data to use and how to use it. In this document we will describe the steps towards creating an argument by asking ourselves questions and then answering them. The argument will be intended to prove the following statement:

Employment among teenagers is higher now than at any other time during the twentieth or twenty-first century.

Before we can make this argument, we need to ask ourselves:

1) *What are we counting?*

We are counting the number of teenagers with jobs at various times throughout time since January 1st, 1900.

2) *What raw data do we need to collect in order to make the count?*

We need to know how many Canadian teenagers were alive at the time of the census for each census taken since January 1st, 1900. We would also need to know how many Canadian teenagers were employed at those times.

3) *What limits or restrictions do you need to place on the data?*

First we define *teenager* (does this include 13, 14, 15 year olds, or just legal working age?). Also, the term *employed* must be defined. Does having employment status mean having a job that earns minimum wage or higher, or do we want to include chores around the house that get paid by the job, not by the hour, etc. In other words, we need to define out terms so that we do not count too few or too many people.

4) *What is the most accurate way to compare data?*

We have a choice: do we want to compare the number of employed teenagers in each census year or the percentage of employed teenagers at each census year? If the definition of *employment* and/or *teenager* change from census to census, then it makes sense to look at the raw data. However, if we want to account for the shifting number of teenagers, we would look at the percentages.

5) *To prove your original point, what relationships would you need to find in the data?*

In order for the point to be proven, a greater percentage of teenagers would have to be working now than in each of the previous census years since January 1st, 1900.

6) *How might the information be misleading? How misleading could it be?*

Rarely is survey data completely accurate. In this case, there may have been a spike in employment that was not recorded because it occurred between census years or at the wrong time of a census year. Likewise, there may be forms of employment that are yet to be included in the census data. To determine the accuracy of the percentages, one must look at the other jobs which are not yet counted and determine if they are statistically relevant (i.e., if there are enough otherwise-employed teenagers to significantly change the employment rates).

In summary, an argument based on statistics requires the following questions to be answered:

1. What are we counting?
2. What raw data do we need to collect in order to make the count?
3. What limits or restrictions do you need to place on the data?
4. What is the most accurate way to compare data?
5. To prove your original point, what relationships would you need to find in the data?
6. How might the information be misleading? How misleading could it be?