Unit 5: Organization of Data

Lesson 5.3: Collecting, Interpreting, and Analyzing Data

Learning Goal: Design effective surveys & experiments to collect primary data

There are two different types of studies:

1. **Observational Studies**:

   For example, a researcher might compare two groups of people, one with members who exercise and another with members who do not exercise. Surveys are typically a form of observational study.
2. **Experimental Studies**:

   For example: a researcher might randomly choose two similar groups and have members of one group perform rigorous exercise once a day for 30 days while members of the other group continue with their normal lifestyle.

In experimental studies, you are comparing two groups:

1. **Treatment group**:
   In the example above it is the group that exercises
2. **Control group**:
   In the example above they are the group that does not exercise.

Three things to occur to determine the cause:

1. **Control**: as many aspects of the experiment need to be controlled as possible so that if there is an effect, the researchers know what caused it.
2. **Randomization**: when groups are chosen, subjects need to be randomized so that no biases occur in any of the groups.
3. **Replication**: even though the groups are random, when researchers repeat and experiment they should be similar in make-up so that changes from one group to another are easier to detect.

Bias: Occurs when there is a prejudice for or against an idea or response. Biased samples can result from:

1. The _____ (where the characteristics of the sample do not reflect the population)
   Example: there is an over representation of male to female ratio

2. The _____ (where there is a flaw in the measurement system or wording of questions provokes an emotional response)

Part 2: Interpreting & Analyzing Data

Learning Goal: Analyze data collected from primary and secondary sources

Recall: Types of data

1. **Numerical** data which can be continuous or discrete
2. **Categorical** data which can be **ordinal** or **nominal**

More about Data

Next, let's consider where data comes from or the source of data.

Primary Source Data is data that has been collected _____.
In a survey, an individual set of data about a single respondent is called _____.

Secondary Source Data is data that is used by someone _____.
Often secondary source data has been summarized and reported as _____ where the individual microdata can no longer be determined.

Cross-sectional Data is a data that is collected with _____ and/or _____ variables about subjects (such as people, firms, countries, etc.) at the same point of time.

Analyzing Data

Tables of data are sometimes difficult to interpret, particularly when there is a lot of data or many variables to compare. When data are displayed in graphical form they are often easier to analyze.

Example 1: The following table shows cross-sectional data of average domestic airfares for 10 Canadian cities for 2010 and 2011. The data was collected by Statistics Canada using a stratified random sample.

a) Does this table show microdata or aggregate data?

b) Is this table a primary or secondary source of data?

c) Identify the independent and dependent variables.

City	2010 $	2011 $	% Change 2010 to 2011
Canada	182.5	190.7	4.5
Calgary	165.5	176.2	6.5
Edmonton	160.8	170.0	5.7
Halifax	172.0	179.3	4.2
Montreal	191.1	194.1	1.6
Ottawa	196.0	194.8	-0.6
Regina	168.1	177.8	5.8
Saskatoon	170.2	178.8	5.1
Toronto	205.2	214.9	4.7
Vancouver	199.2	206.7	3.8
Winnipeg	181.0	189.4	4.6

Average Domestic Air Fares for Canada and 10 Major Cities

Source: Table 1 Average domestic air fares for 10 major Canadian cities, *The Daily*, Wednesday, January 9, 2013, Statistics Canada

d) Why do you suppose a stratified random sample was used to collect this data?

Average Domestic Air Fares for 10 Major Canadian Cities

e) Use Excel to reproduce the graph shown.

f) What kind of story do the data in the table and graph tell?

Example 2: Suppose you conducted an experiment, dropping a ball from various heights and measuring the height of the bounce. The table shows the results.

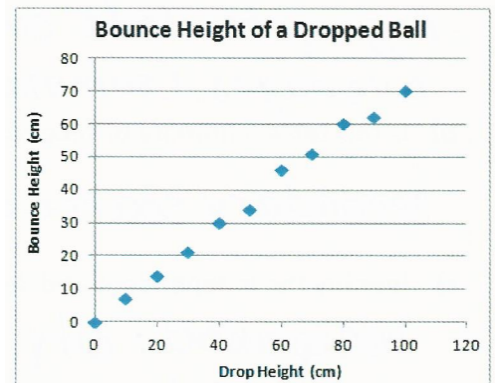a) Is this primary or secondary source of data?

Drop Height (cm)	Bounce Height (cm)
0	0
10	7
20	14
30	21
40	30
50	34
60	46
70	51
80	60
90	62
100	70

b) Identify the independent and dependent variables.

c) Which variable is being controlled in this activity?

d) Use Excel to reproduce the graph shown.

Bounce Height of a Dropped Ball

e) Describe the relationship.

f) If you dropped the ball from 130m, how could you use this information to predict its bounce height?