

Data Analysis—Leading Digits

One-Page Overview

By Robert B. Brown, The Ohio State University

Topics:

Statistics; Probability; Patterns

Levels:

Grades 5 – 12

Problem:

How many of the numbers in a set of data have a leading digit of 1 and how many have a final digit of 1?

Getting Started:

Ask students what they would expect if they did the following experiment:
 Find a large table of data, such as lengths of the world's rivers or populations of countries. Count and record how many of the tabulated numbers have a leading digit of 1 and how many have a final digit of 1. Do the same for all the other digits 2, 3, 4, etc., and record how many times each occurs in the leading and in the final place. Also record how many times 0 occurs in the final place. Make two histograms exhibiting the results, one for how many times each digit occurs as a leading digit, and one for the final digits.

Ohio Academic Content Standards, 2002

5-7		8-10		11-12	
1. Number, Number Sense and Operations		1. Number, Number Sense and Operations		1. Number, Number Sense and Operations	
2. Measurement		2. Measurement		2. Measurement	
3. Geometry and Spatial Sense		3. Geometry and Spatial Sense		3. Geometry and Spatial Sense	
4. Patterns, Functions and Algebra	X	4. Patterns, Functions and Algebra	x	4. Patterns, Functions and Algebra	x
5. Data Analysis and Probability	X	5. Data Analysis and Probability	X	5. Data Analysis and Probability	X
Mathematical Processes Reasoning		Mathematical Processes Reasoning		Mathematical Processes Reasoning	

NCTM Principles and Standards, 2000

6-8		9-12	
1. Number and Operations		1. Number and Operations	
2. Algebra	X	2. Algebra	X
3. Geometry		3. Geometry	
4. Measurement		4. Measurement	
5. Data Analysis and Probability	X	5. Data Analysis and Probability	X
6. Problem Solving		6. Problem Solving	
7. Reasoning and Proof	x	7. Reasoning and Proof	x
8. Communication		8. Communication	
9. Connections		9. Connections	
10. Representation		10. Representation	

Note: Capital X denotes major emphasis; lower case x denotes minor emphasis.

Data Analysis—Leading Digits

By Robert B. Brown, The Ohio State University

<u>Topics:</u> Statistics, Probability, Patterns	
<u>Levels:</u> Grades 5-12	<u>Timing:</u> 1 - 2 class periods
<u>Materials:</u> Sources of tables, such as almanacs or access to the World Wide Web	<u>Prerequisites:</u> None

Problem:

How many of the numbers in a set of data have a leading digit 1 and how many have a final digit 1? Do the same for all the other digits 2, 3, 4, etc.

Goals:

- Make predictions based on an informal probability model
 - Face unexpected results
 - Try to explain unexpected results
 - Create histograms for data analysis
 - Practice cooperative work, as this problem lends itself a division of labor if you can provide each student with a copy of the table that they have chosen for their experiment.
-

Procedure:

1. Break the students into groups of three or four. Describe to the class the experiment that each group will do:

Each group will choose a large table of data that includes about 100 numbers. Almanacs and encyclopedias are good sources. Examples of tables of data are the lengths of the world's rivers or populations of countries.

The groups will then count and record how many of the tabulated numbers have leading digit 1 and how many have final digit 1. Do the same for all the other digits 2, 3, 4, etc., and record how many times each occurs in the leading and in the final place. They should also record how many times 0 occurs in the final place. Make two histograms exhibiting the results, one for how many times each digit occurs as a leading digit, and one for the final digits.

2. Discuss what the students expect will be the shape of the histograms. They will probably expect that each digit will occur with about the same frequency in any given place.
3. Have each group choose a table and carry out the experiment including making the histograms.
4. Discuss whether their resulting histograms are shaped as they expected. If they differ, describe how they differ and have the students propose reasons for the differences. Most likely, the digit 1 will occur as leading digit much more frequently than predicted, and at the expense of the larger digits, such as 7, 8, and 9. Also, the digit 0 may occur more frequently than predicted as the final digit.
5. Aggregate the results from all of the groups and make two histograms that cover the aggregate, one for leading digits, the other for final digits.
6. Discuss whether the aggregate histograms fit the original expectations of the students, and if not, possible reasons for the differences.
7. Discuss whether the aggregate histograms resemble the histograms for the individual experiments. Try to explain any observed differences.

Closure:

Discuss the idea that data analysis is more subtle than it might at first seem.

Extension:

See the following websites for extensions and more in-depth information about Benford's Law:

<http://more.abcnews.go.com/sections/scitech/WhosCounting/whoscounting020301.html>

<http://www.mathpages.com/home/kmath302/kmath302.htm>

The Mathematics:

This activity can be done with students who can informally express the expectation that each digit should occur with about the same frequency as every other digit. Unless there has been some rounding or doctoring of the table, this is usually the outcome of tabulating the final digits. But it is rarely true for the leading digits. Here is a table of the results of amalgamating the experiments in one class of size about 25:

Digit	Leading	Final
1	323	91
2	169	113
3	148	109
4	96	105
5	92	101
6	61	97
7	87	113
8	108	106
9	72	132
0	0	153

The actual table is full of 1's in the leading place. The digits 2 and 3, as well, appear more often than expected. The distribution for the leading digits is not at all a uniform distribution, in which each digit would occur with about the same frequency. The distribution for the final digits looks much more uniform, except the digit 0 occurs a little more often than might be expected. Also, the total count of leading digits is more than the total count of final digits. What might account for that?

Here are the results from the house numbers of 1156 consecutive telephone book entries. (1156 is the same as the total count of leading digits in the first table.)

<u>Digit</u>	<u>Leading</u>	<u>Final</u>
1	291	120
2	170	100
3	143	123
4	126	87
5	161	139
6	101	112
7	62	87
8	64	118
9	38	93
0	0	170

What are the similarities between this table and the preceding one?

Relationships to the Ohio Academic Content Standards, 2002:

Grades 5-7:

Patterns, Functions and Algebra Standard

The student will be able to...

- Describe, extend and determine the rule for patterns and relationships occurring in numeric patterns, computation, geometry, graphs and other applications.

Data Analysis and Probability Standard

The student will be able to...

- Read, create and use line graphs, histograms, circle graphs, box-and-whisker plots, stem-and-leaf plots, and other representations when appropriate.
- Collect, organize, display and interpret data for a specific purpose or need.
- Evaluate conjectures and predictions based upon data presented in tables and graphs, and identify misuses of statistical data and displays.
- Make and justify predictions based on experimental and theoretical probabilities.

Mathematical Processes Standard

The student will be able to...

- Use deductive thinking to construct informal arguments to support reasoning and to justify solutions to problems.
- Use inductive thinking to generalize a pattern of observations for particular cases, make conjectures, and provide supporting arguments for conjectures.

Grades 8-10:

Data Analysis and Probability Standard

The student will be able to...

- Create, interpret and use graphical displays and statistical measures to describe data; e.g., box-and-whisker plots, histograms, scatterplots, measures of center and variability.
- Construct convincing arguments based on analysis of data and interpretation of graphs.
- Make predictions based on theoretical probabilities and experimental results.

Mathematical Processes Standard

The student will be able to...

- Apply reasoning processes and skills to construct logical verifications or counter-examples to test conjectures and to justify and defend algorithms and solutions.

Grades 11-12:

Data Analysis and Probability Standard

The student will be able to...

- Create and analyze tabular and graphical displays of data using appropriate tools, including spreadsheets and graphing calculators.

Mathematical Processes Standard

The student will be able to...

- Present complete and convincing arguments and justifications, using inductive and deductive reasoning, adapted to be effective for various audiences.

Relationships to the NCTM Principles and Standards, 2000:

Grades 6-8 and 9-12:

Algebra Standard

Instructional programs from pre-kindergarten through grade 12 should enable all students to...

- Understand patterns, relations, and functions.
- Use mathematical models to represent and understand quantitative relationships.

Data Analysis and Probability Standard

Instructional programs from pre-kindergarten through grade 12 should enable all students to...

- Formulate questions that can be addressed with data and collect, organize, and display relevant data to answer them.
- Develop and evaluate inferences and predictions that are based on data.
- Understand and apply basic concepts of probability.

Reasoning and Proof Standard

Instructional programs from pre-kindergarten through grade 12 should enable all students to...

- Make and investigate mathematical conjectures.