Statistical Literacy

AN INTRODUCTION
TO A WEB-BASED
STATEMENT VALIDATOR
FOR
RATES AND PERCENTAGES

A Product of the
W. M. Keck Statistical Literacy Project
Augsburg College
Minneapolis, MN

schield@augsburg.edu
www.StatLit.org

This document at
www.StatLit.org/pdf/2006BurnhamSchieldIASSIST.pdf

Program access at
www.StatLit.org/GC

REFERENCES

STATISTICAL LITERACY LEARNING OBJECT:

Burnham, Tom and Milo Schield (2005). An Online Grammar-parsing Program That Decodes Ordinary English Descriptions and Comparisons of Percentages and Rates. *2005 American Statistical Association Proceedings of Section on Statistical Education*, [CD-ROM] 2202-2208. At www.StatLit.org/pdf/2005BurnhamSchieldASA.pdf.

STATISTICAL LITERACY AND QUANTITATIVE LITERACY

Schield, Milo (2004a). Statistical Literacy and Liberal Education at Augsburg College. *Peer Review,* Sept. 2004, 7-14. AACU. At www.StatLit.org/pdf/2004SchieldAACU.pdf.

Best, Joel (2001). *Damned Lies and Statistics.* University of California Press.

Best, Joel (2004). *More Damned Lies and Statistics.* University of California Press.

Steen, Lynn (2005). Achieving Quantitative Literacy. MAA.

Schield, Milo (2004b). Information Literacy, Statistical Literacy and Data Literacy. *IQ (IASSIST Quarterly)*, 7-14. At www.StatLit.org/pdf/2004SchieldIASSIST.pdf.

Schield, Milo (2000). *Statistical Literacy and Mathematical Reasoning*. International Conference on Mathematics Education (ICME-9), Tokyo. See www.StatLit.org/pdf/2000SchieldICME.pdf.

Schield, Milo (2005a). *Statistical Prevarication: Telling Half Truths Using Statistics.* International Assoc. of Statistical Educators (IASE). Invited paper, Sydney. At www.StatLit.org/pdf/2005SchieldIASE.pdf.

STATISTICAL LITERACY AND RATIOS:

Schield, Milo (2006a). *Statistical Literacy Survey Results: Reading Tables and Graphs of Rates and Percentages.* International Association for Social Science Information Service and Technology (IASSIST). www.StatLit.org/pdf/2006SchieldIASSIST.pdf.

Schield, Milo (2006b). *Statistical Literacy Survey Analysis: Reading Tables and Graphs of Rates and Percentages.* Int'l. Conference Teaching Statistics. www.StatLit.org/pdf/2006SchieldICOTS.pdf.

Schield, M. (2001). Statistical Literacy: Reading Tables of Rates and Percentages. *2001 ASA Proceedings of the Section on Statistical Education*. See www.StatLit.org/pdf/2001SchieldASA.pdf.

STATISTICAL LITERACY AND STANDARDIZATION:

Schield, Milo (2006c). *Presenting Confounding and Standardization Graphically.* Draft for STATS Magazine. At www.StatLit.org/pdf/2006SchieldSTATS.pdf.

Schield, Milo (2004d). *Three Graphs to Promote Statistical Literacy.* International Conference on Mathematics Education (ICME-10). At www.StatLit.org/pdf/2004SchieldICME.pdf.

TEACHING STATISTICAL LITERACY:

Isaacson, Marc (2005). Statistical Literacy – Online at Capella University. 2005 *American Statistical Association Proceedings of the Section on Statistical Education*, [CD-ROM] 2244-2252. At www.StatLit.org/pdf/2005IsaacsonASA.pdf.

Schield, Milo (2004c). *Statistical Literacy Curriculum Design.* IASE Curriculum Design Roundtable. At www.StatLit.org/pdf/2004SchieldIASE.pdf.

Schield, Milo (2005b). Statistical Literacy and Chance. *2005 American Statistical Association Proceedings of the Section on Statistical Education.* [CD-ROM] 2302-2310. At www.StatLit.org/pdf/2005SchieldASA.pdf.

Statistical Literacy studies the use of statistics in everyday language. The statistics involved are typically ratios – rates and percentages – that are presented in tables and graphs. Students have considerable difficulty describing and comparing ratios[1] when they are presented in tables and graphs[2]. It takes considerable class time to teach this material to students. Drilling on this is a burden on class time, so a computer-aided drill program is needed.

Technically, part-whole ratios are conditional probabilities. Statisticians have developed an algebraic notation for describing conditional probabilities that is clear, concise and unambiguous. In talking about adults, let M = Men and R = Runner, then P(R|M) is the prevalence of Runners among Men (the probability that a randomly selected man will be found to be a runner). Unfortunately, most students do not speak or even think in terms of algebra. Even if they do so for a course, they are unlikely to retain much if they don't use algebra thereafter. However, a great majority of students use ordinary English to describe and compare conditional probabilities in the forms of rates and percentages.

Ordinary language descriptions of ratios have been organized into four grammatical families: "% of," "percentage," "rate" and "chance." "Percentage" includes "fraction" and "share." "Rate" includes "prevalence" and "incidence." "Chance" includes "risk," "odds,"[3] "likelihood" and "probability."[4] For making comparisons, there is also the "likely/prevalent" grammar.

A statistically-literate student should be able to decode any table of rates or percentages, describe any ratio using any applicable form of grammar and compare any two ratios using any applicable grammar."

---

[1] Schield (2000).

[2] Schield (2001).

[3] The grammar for odds is unique within this family.

[4] The "chance" family also includes confident and accurate. E.g., 95% confident, 90% accurate.

GRAMMARS FOR DESCRIBING AND COMPARING RATIOS.

Table 1: Unemployment Rates in 1997 (% of civilian labor force)
by Educational Attainment, Sex and Race

| SEX and RACE | Total | Highest Grade Achieved | | | |
		< 12	12	13-15	> 15
Total: \2	4.4	10.4	5.1	3.8	2.0
Male:	4.7	9.9	5.6	4.0	2.1
Female:	4.1	11.3	4.5	3.6	2.0
White:	3.9	9.4	4.6	3.4	1.8
Black:	8.1	16.6	8.2	6.1	4.4
Hispanic: \4	7.3	9.6	7.5	5.5	3.0

\2 Includes other races, not shown separately.
\4 Persons of Hispanic origin may be of any race.

In referring to the 1997 US civilian labor force in the above table, the circled 4.7% can be described 11 ways:

"% of" grammar has two forms:
1. 4.7% of males are <u>unemployed</u>
2. Among men, 4.7% are <u>unemployed</u>.

"Percentage" grammar has two forms:
1. The percentage of men who are <u>unemployed</u> is 4.7%
2. Among men, the percentage of <u>unemployed</u> is 4.7%.

"Rate" grammar has seven forms:
1. The rate of <u>unemployment</u> is 47 per 1,000 men.
2. The <u>unemployment</u> rate of men is 47 per 1,000.
3. Men's <u>unemployment</u> rate is 47 per 1,000.
4. Men are <u>unemployed</u> at a rate of 4.7%.
5. <u>Unemployment</u> occurs at a rate of 47 per 1,000 men.
6. The rate at which men are <u>unemployed</u> is 4.7%.
7. The rate at which <u>unemployment</u> occurs is 47 per 1,000 men.

Note the various grammatical devices used to indicated that "unemployment" is the part in a part-whole ratio. This program handles these 11 forms of descriptions but is not yet programmed to handle "chance" grammar.

Each description involving "percentage" or "rate" has a *corresponding comparison*. A *corresponding comparison* uses the same grammar as the descriptions being compared. Consider using percentage grammar to compare the unemployment of women and men. Since the percentage of females who are unemployed is 4.1%, we can form this comparison, "*The percentage who are unemployed is 50% greater among men than among women*."

While there are no comparisons using "% of" grammar, there are two forms involving "likely/prevalent" grammar for a total of 11 kinds of comparisons. Most of these have both a long and a short form, and all of these can involve one of three comparisons.[5]

---

[5] 1. Simple Difference: {Test} is X more than {base}. E.g., 8 is 6 more than 2. 2. Simple Ratio: {Test} is X times as much as {base}. E.g., 8 is 4 times as much as 2. 3. Relative Difference. Percentage Compare: {Test} is X% more than {base}. E.g., 8 is 300% more than 2. Times Difference: {Test} is X times more than {base}. E.g., 8 is 3 times more than 2. The program does not handle #1.

NEED FOR AN ON-LINE DRILL PROGRAM:

For those who did not learn English at an early age, describing and comparing ratios in ordinary English may involve learning the rules for using prepositions and relative clauses. Yet those who have learned English as a second language may do better than native English speakers since the native speakers are not accustomed to using English in such a precise manner. Both groups need time to practice decoding tables and writing descriptions and comparisons of conditional probabilities in ordinary English.

This material has been taught to over a 1,000 students over a period of 10 years. Students need 3 to 6 hours of class time for drill for 80% of the class to achieve a passing level of skill at describing and comparing both percentages and rates in tables. This is a lot of class time. Yet 20% of the students needed still more time. Tutors were provided but busy students have difficulty working in special study sessions. For these reasons an on-line grammar-parsing drill program was needed by both students and teachers.

At first it seemed that asking a computer program to decode ordinary language statements was asking too much. Great minds working in artificial intelligence have realized that analyzing context is not easy for computers. But this situation differs in several ways. First, the semantics of all correct answers is known in advance. Second, the patterns that can be used to describe or compare rates and percentages are fairly limited in their grammatical structures. Third, the semantic domain for answers is limited. These facts make it possible to create a program that is useful to students.

If an on-line program simply says OK or Wrong, this is not helpful to a student. A teacher or tutor would describe the error in such a way as to help the student without giving them a correct answer. The quality of the program will be largely determined by the appropriateness and helpfulness of the user-error messages generated by the program.

Goals

- User Goal: To write a single sentence in ordinary English that describes a single ratio or compares two ratios contained in a table of percentages or rates and that can be quoted out of context without being misunderstood (no grammar errors, no ambiguities and no missing relevant terms).

- Program Goal: To decode the user's sentence describing a ratio or comparing two ratios contained in a table of percentages or rates, to accurately identify user errors relative to the table data and the problem statement, to provide error messages that help the user improve their answer, and to capture student comments for program improvement.

The program shown herein[6] is an expanded version of the one described by Burnham and Schield (2005). In March 2006 the program was expanded to include statements as a data source. In May 2006, the program was expanded to include pie charts as a data source. It is expected that presenting part-whole percentage in pie charts and statements should make it much easier for users to write out descriptions and comparisons before they take on doing so with tables and bar charts.

---

[6] Version 16, 21 May 2006

The program screen has two parts:
- The top part of the screen contains the option buttons that select the kind of problem.
- The rest of the screen involves working with a specific problem.

There are four command buttons on this screen:
- The "NEW PROBLEM" button in the upper-right corner brings up a new problem that matches the "kind of problem" selected by the option button.
- The "ANALYZE ANSWER" button on the right side asks the program to analyze what the user has entered.
- The "SHOW ANSWER" button in the lower-left corner gives the user one of the many answers that are acceptable for this problem.
- The "EXIT" button in the lower-right corner exits the program and enters the user's results into an on-line log.

The top part of the screen involves up to five rows of options. Consider the first two.
The first row allows the user to select the "**Source**" of the problem.
> There are four options: **Pie Chart**, **Statement**, **Table** and **Bar Graph**.
> These are arranged in order of difficulty. Beginners should start with pie chart.
- "Pie chart" presents a percentage in the form of a pie chart as shown above.
- "Statement" presents a written statement of a percentage or rate.
- "Table" presents a table of percentages or rates.
- "Bar graph" presents a bar chart of a percentage or a rate.

The second row allows the user to select the "**Activity**": Describe or Compare.
- "Describe" asks the user to describe a single ratio
- "Compare" asks the user to compare two ratios.

The content of the remaining rows varies with the selections in these first two rows.

To see how the program operates, consider these options: In the first line, the Source is PIE CHART. In the second line, the Activity is DESCRIBE. In the third line, the Answer grammar is "% OF/IS."   In the middle-left, a pie-chart is shown. In the middle right, the problem is stated. The user is asked to type in their answer below the graphic and press the "Analyze Answer" button when they are done with their answer.

Suppose the user types in the answer shown below and presses "Analyze Answer."

The program analyzes the user answer in the textbox immediately below. In this example, the answer is wrong. If the answer is wrong, two command buttons may appear.

- More Detail: Pressing this button, gives more detail on the error message.

- Analyze Ps and Ws (Parts and Wholes): Pressing this button asks the program to analyze the grammatical part-whole status of each of the terms in the user's answer.

In the lower right corner, a new User Comment textbox appears. If the user has a comment or question to be read by those who can change the program, enter it into this Comment text box and press the "Store Comment" command button.

When the user enters a correct answer (there are a number of possibilities),

- The "# Answers" and "# Correct" are incremented by one.
- In the "Answer Analysis" box, the program says, "No Problems Found."

The textbox at the very bottom of the screen (between "Show an Answer" and "Exit") shows the user entry that was the last one analyzed.

Consider selecting "STATEMENT" for Source and "COMPARE" for Activity.
The second row has additional options when a comparison is selected.
- Comparison Type: This selects the type of comparison that is allowed.

The third and fourth rows give these options:
- Answer Grammar: The grammar the user must use in their answer.
- Source Grammar: The grammar used in the statement presented.
  The options for Source grammar depend on the Answer grammar selected.

Source:	○ Pie Chart ● Statement ○ Table ○ Bar Graph		New Problem
Activity:	○ Describe ● Compare	Compare type: ● Any ○ Difference ○ Ratio ○ Relative Difference	
Answer grammar:	○ Percentage ● Likely ○ Prevalent ○ Rate	User name (opt):	User Name
Source grammar:	● # % of/is ○ Percentage ○ Prevalence ○ Rate		

Among college students who are males, 60% are business majors.

Among college students who are females, 20% are business majors.

Generate a comparison from the given description statements using 'likely' grammar; use the smaller as the base.

Answers: 0
Correct: 0

Consider selecting "TABLE" for Source and "COMPARE" for Activity.

Tables are classified by the types of margins involved:
- 100%: These tables have at least one 100% margin value.
- Full: These tables have margins for both rows and columns.
- Missing: These tables have at least one missing margin.
  Again, these options are organized in ascending difficulty. Users should start with the 100% tables.

Source:	○ Pie Chart ○ Statement ● Table ○ Bar Graph	Table Margin Type: ● 100% ○ Full ○ Missing	New Problem
Activity:	○ Describe ● Compare	Compare type: ● Any ○ Difference ○ Ratio ○ Relative Difference	
Answer grammar:	○ Percentage ● Likely ○ Prevalent ○ Rate	User name (opt):	User Name

College Students

MAJOR	Male	Female	TOTAL
Business	(60%)	(20%)	40%
Economics	10%	50%	30%
MIS	30%	30%	30%
TOTAL	100%	100%	100%

Compare the circled 60% to the circled 20% using 'likely' grammar; use the smaller as the base.

Answers: 0
Correct: 0

Consider selecting "BAR GRAPH" for Source and "DESCRIBE" for Activity.

Source:	○ Pie Chart ○ Statement ○ Table ● Bar Graph		New Problem
Activity:	● Describe ○ Compare		
Answer grammar:	● # % of/is ○ Percentage ○ Rate	User name (opt):	

Percentage of Business Majors among College Students

Female	20%
Male	60%

0%   20%   40%   60%   80%

Describe the 60% using 'percent of/is' grammar.

Answers: 0
Correct: 0

CONCLUSION

The W. M. Keck Statistical Literacy Survey found that students, data analysts and college professors have difficulty reading and interpreting rates and percentages when presented in tables and graphs. See Schield (2004a and 2004b).

This program is unique in decoding ordinary English statements and giving immediate feedback on the syntax and semantics involved in describing and comparing rates and percentages as presented in pie charts, statements, tables and bar graphs. It has been used by hundreds of students.

To access this program, go to www.StatLit.org/GC