The Mathematics Education into the 21st Century Project
Proceedings of the International Conference
The Decidable and the Undecidable in Mathematics Education
Brno, Czech Republic, September 2003

Two-Way Tables: Introducing Probability Using Real Data

Gail Burrill
Michigan State University

Not too long ago, statistics courses in the United States consisted of some statistics and lots of probability. The probability of an event was defined as the number of successes out of the total number of possible outcomes. Thus, most students learned probability as a study in how to count and spent time learning about permutations and combinations using cards, dice, and urns. In the United States, applications were essentially about finding the probability that two people would have the same set of initials or the probability that a five-card hand would have four aces. This approach to probability did not seem to help students develop any notion of chance, and various studies have shown that much of probability is not intuitive (Fischbein & Schnarch, 1997; Keeler & Steenhorst, 2001; Kahneman & Tversky, 1972). Furthermore, students in introductory courses rarely experienced applications from the world in which they lived. Drawing balls out of urns became the focus rather than thinking of the use of balls and urns as a model for an actual probability situation. When students finally met "p-values" in statistical inference, most of them seldom had any real notion of what "p" meant.

A different approach to probability can help students develop an understanding of probability and appreciate it as a measure of chance. The use of two way tables and simple practical data can help students relate probability to very real situations (Scheaffer et al, 1999). Students can use data they gather in a survey, observational data, or experimental data. They can gather the data themselves or make use of data reported in the media. This paper will explore three situations and illustrate how the use of two-way tables can lead to the development of simple probability rules, to an understanding of whether or not there seems to be an association between two variables, and to the use of a simulated sampling distribution to determine whether an observed outcome seems likely to have occurred by chance.

Situation 1: Students collect their own data about a topic of interest to them and from this begin to generate some of the traditional probability formulas.

Students in an urban school were curious about how many of their classmates ate breakfast. The teacher posed questions such as: Is there a bigger chance that a student chosen at random eats breakfast on a regular basis than that the student does not eat breakfast? Do boys eat breakfast more often than girls? The class surveyed the student population asking "Do you eat breakfast on a regular basis?" and recorded the results along with gender (Scheaffer et al, 1999).

Their data can be organized in a table (Table 1)   Table 1: Eating breakfast and gender

	Male	Female	Total
Eat breakfast on a regular basis	190	110	300
Do not eat breakfast on a regular basis	130	165	395
Total	320	375	695

Using the results, the students can explore how to use the definition of probability to find the probability of A, of A given B, of A and B, and of A or B. For example, what is the probability that a male student eats breakfast on a regular basis? Of the 320 males, 190 or about 60% of them reported that they ate breakfast on a regular basis. Of the 375 females 74 or about 20% reported eating breakfast regularly. By experimenting with the tables, students learn the importance of the row and column totals and how they are related to the probability; they soon recognize that given the row and column totals, knowing any one of the interior cells will determine the others.

Given a question about whether events are independent, they can refer back to a two-way table. Is eating breakfast on a regular basis independent of gender or is it possible some association exists between the two variables? For example, overall 43% of the students indicated they ate breakfast regularly. Thus, if there were no difference in the breakfast eating habits of boys and girls, about 14% of each would eat breakfast on a regular basis (Figure 1). Figure 2 shows the situation for the school data. There seems to be a difference between boys and girls in terms of eating breakfast on a regular basis.

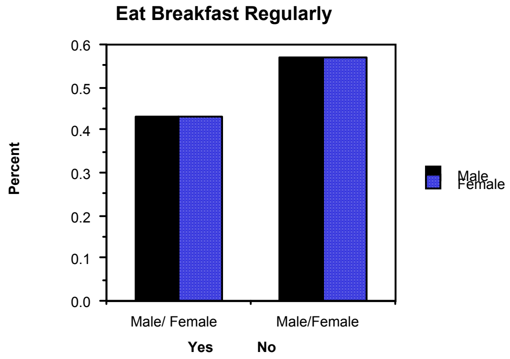

Figure 1

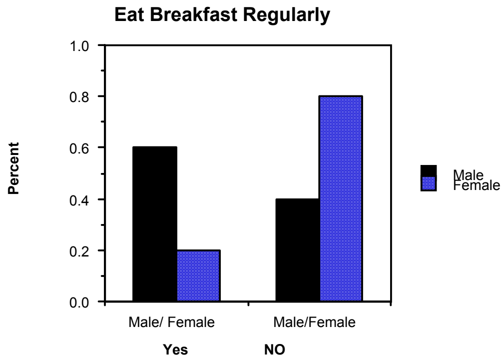

Figure 2

Situation 2. Students take the results of an experiment reported in the media and analyze them in terms of chance and relationships.

Once students have explored probability in situations from their own environment, they can take information about other chance situations and sort out the reasoning. For example, do the data support the claim in the headlines of the following article?

"Antibiotics can worsen E coli complications

Children who may be infected with the bacterial E coli 0157-H7 should not be treated with antibiotics because they raise the risk of a potentially deadly complication called hemolytic uremic syndrome (HUS) according to a report in the New England Journal of Medicine. Researchers at the University of Washington School of Medicine looked at 71 children with E coli 0157-H7 poisoning, nine of whom were treated with antibiotics. Of the nine five developed HUS. Among the remaining 62, five developed HUS. E coli 0157:H7 can be transmitted by water or undercooked foods." (USA TODAY, 2000; Wong, et al., 2000)

Table 2 shows the data in the article

Table 2. Antibiotics and HUS

	Antibiotics	No Antibiotics	
HUS	5	5	10
No HUS	4	57	61
	9	62	71

Again, students can explore chance. What is the chance of getting HUS? What is the chance that someone who was treated with antibiotics contacted HUS? How does this compare to the chance of getting HUS without taking antibiotics? The raw data make it difficult to see the story in the numbers. The chance of getting HUS for those participating in this study was 10 out of 71 or 14%. Of the 9 treated with antibiotics, 5 or 56% contacted HUS. Of the 62 who were not treated with antibiotics, 5 or 8% contacted HUS.

What do the data indicate for those with E-Coli about a connection between taking antibiotics and contacting HUS? Suppose the data had clearly indicated no connection between HUS and taking antibiotics. Because 14% of those in the study contacted HUS, it could be expected that 14% of those taking and of those not taking antibiotics would contact HUS (Table 3). Table 4 shows a clear association between the two with all 9 of those taking antibiotics contacting HUS.

Table 3 No relation between taking antibiotics and HUS

	Antibiotics	No Antibiotics	
HUS	1	9	10
No HUS	8	53	61
	9	62	71

Table 4 Relationship between taking antibiotics and HUS

	Antibiotics	No Antibiotics	
HUS	9	1	10
No HUS	0	61	61
	9	62	71

Again, the numbers alone make the story hard to see. What is the difference here between no association, clear association, and the observed data? Figure 3 shows the observed data, and Figure 4 depicts the situation if there were clearly no connection.

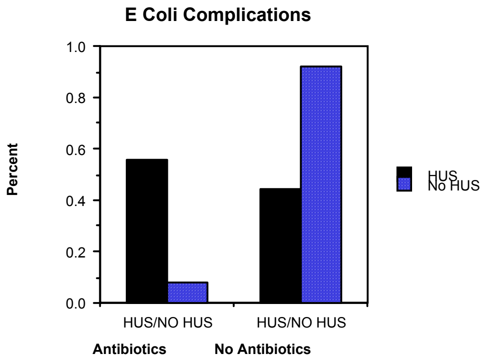

Figure 3

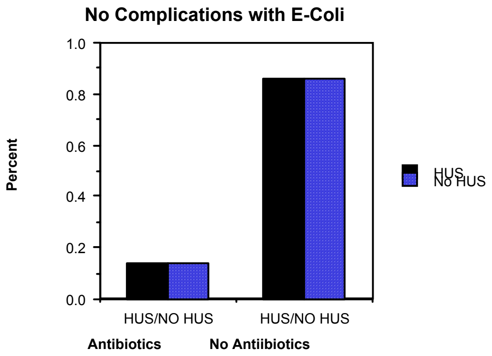

Figure 4

The data seem to suggest a possible association implying that the relationship between contacting HUS and antibiotics might need more investigation.

Situation 3. Students take the results of a survey given in the media and analyze it for a possible association between the variables. Probability is a tool in drawing a conclusion.

A newspaper article reported the results of a survey on the question of whether men and women get the same pay for the same work. The results showed that 2/3 of the women felt this to be a problem while only one half of the men did. Does this show a difference in the way men and women feel about the issue? Table 5 shows the data for the 44 people surveyed (Teachers Teaching with Technology, 2001).

Table 5: Equal pay for equal work

	A problem	Not a problem	Total
Men	10	10	20
Women	16	8	24
Total	26	18	44

Do the data suggest a real difference in the way men and women perceive the situation? How likely is obtaining the number of men and women who feel the situation is a problem just by chance. What is the probability that the data in the table would occur by chance? The situation can be simulated by using a deck of cards, keeping the row and column totals constant. The 20 men can be represented by 20 black cards, and 24 red cards can represent the women. Shuffling and dealing out 18 cards will simulate the 18 responders who feel that the situation is not a problem. Just as in the first situation, knowing one of the cells determines the others. Thus, counting the number of black cards in the18 that were dealt gives the number of men in that set who feel that the equal pay for equal work is not a problem.

Table 6 shows the situation when the data would indicate no difference in the way men and women perceive the problem.

Table 6: No difference between men and women

	A problem	Not a problem	Total
Men	12	8	20
Women	14	10	24
Total	26	18	44

If there were no difference in the way men and women responded, you would expect to have 10 men agreeing that equal pay for equal work is a problem with some variability due to chance. The observed number of men who felt this was a problem was 8. Would 8 be likely to happen as part of the chance variability or could it indicate that some other factors might be involved, such as a real difference between the way men and women perceive the situation?

Figure 5 shows the results of repeating the simulation with the cards 50 times in a sampling distribution for the number of men who feel the problem of equal pay for equal work for men and women is serious. How does the observed value (the number of men who felt the problem to be serious) compare to a distribution generated by chance? From the simulated distribution, the probability that at least 8 men felt this was a problem is 35/50 or about 70%. In this case, the observed value seems just as likely to have occurred by chance rather than indicate a real difference between the way men and women in the study feel about the problem of equal pay for equal work.

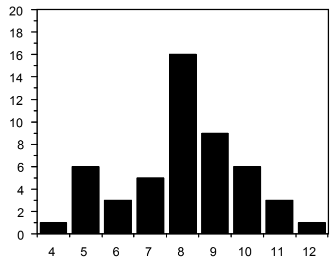

Figure 5

Conclusion

The situations described in this paper are typical of many situations that appear in the media or in reports on surveys and experiments. Using two-way tables to organize data and establish chance events and the informal approach to analyzing whether the data provide evidence to make a decision described above also serves as a preparation for more formal hypothesis testing and grounds the notion a p-value. Both the E-Coli data and the problem of equal pay for equal work can be analyzed using the chi square test or a difference between two proportions z-test.

Two-way tables are a practical and sense making way to help students develop an understanding of probability. The use of real data makes the notion of chance important, one that touches lives in ways other than the lottery or games of chance. To be literate and contributing members of society and in their work, people should understand the role of chance, measured by probability, in analyzing information and in making decisions. Probability can help students think critically about choices they will face as adults: Is hormone therapy a good choice for women as they age? What are the consequences of choosing one of two possible treatments for an illness? Should I take this recommended diet seriously and in fact, change my life style to do so? What are the risks for my company of a given course of action.

References

Fischbein, E., & Schnarch, D. (1997). The evolution with age of probabilistic intuitively based misconceptions. *Journal of Research in Mathematics Education*. 28, 96-105.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representiveness. *Cognitive Psychology*. 3, 430-454.

Keeler, C., & Steenhorst, K. (2001). A new approach to learning probability in the first course in statistics. *Journal of Statistical Education* 9, 3.

Kranondonk, H., Hopfensperger, P. & Scheaffer, R. (1999). *Exploring Probability*. Palo Alto
CA: Dale Seymour Publications.

Teachers Teaching with Technology. (2001). Advanced placement statistics with the TI-83. Dallas, TX: Texas Instruments, Incorporated.

Wong, C. S., Jelacic, S., Habeeb, R. L., Watkins, S. L., & Tarr, P. I. (2000). The risk of the
hemolytic-uremic syndrome after antibiotic treatment of escherichia coli O157:H7 infections. *The New England Journal of Medicine 342*, 1930-6.