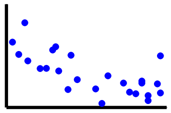


TOPIC D5: RELATIONSHIPS BETWEEN CATEGORICAL VARIABLES

SPOTLIGHT: THE TITANIC: WOMEN AND CHILDREN FIRST?

On April 15, 1912, the ocean liner Titanic collided with an iceberg, sank, and many of its passengers lost their lives. Many books have been written about this catastrophic event. One issue that has been discussed is the role of economic status in determining the mortality status of the passengers on this ship. In the recent movie *Titanic*, it was suggested that many of the lower-class passengers were less likely than the higher-class passengers to survive this accident. Dawson (1995) in the *Journal of Statistics Education* (vol. 3, n. 3) makes available an interesting dataset on the passengers of the Titanic. For each of the 2201 passengers, the dataset records

- the economic status (crew, first-class, second-class, third-class)
- the age (adult or child)
- the gender (male or female)
- the survival status (either yes or no)

By use of these data, we will see if there is a relationship between the economic status and survival of the Titanic passengers.

PREVIEW

In this topic, we discuss ways of exploring relationships when the variables are categorical. We begin with a two-way table of counts and by computing relevant *conditional* proportions from the table, we can discuss how knowledge of one variable is informative about the second variable.

In this topic your learning objectives are to:

- Understand how to construct a two-way table of counts from a data table.

- Understand how to compute conditional row percentages or conditional column percentages from the two-way table.
- Understand how to graph sets of conditional percentages.
- Understand how to use conditional percentages to describe the relationship between two categorical variables.

NCTM Standards

✓ In Grades 6-8, all students should collect data about different characteristics within one population.

✓In Grades 9-12, all students should display and discuss bivariate data where at least one variable is categorical.

A TWO-WAY TABLE OF COUNTS

The following table displays a partial listing of the Titanic data. Each row corresponds to the values of the four categorical variables for a passenger.

CLASS	AGE	SEX	SURVIVED
First	Adult	male	Yes
First	Adult	male	Yes
First	Adult	male	Yes
First	Child	female	No
Crew	Adult	female	Yes
Crew	Adult	female	No
Crew	Adult	female	No

Crew	Adult	female	No

The Titanic accident always will be remembered for the large number of fatalities. Let's first focus on the "Survived" variable in the dataset and construct a frequency table. The table below shows the counts of people who survived and who did not survive; this table also shows the corresponding percentages. From this table, we see that about 2/3 of the passengers did not survive this trip.

.

	Survived		Total
	no	yes	
Count	1490	711	2201
%	67.7	32.3	100

Let's consider the question posed earlier -- were passengers of certain economic classes more or less likely to survive the accident?

To answer this question, we classify the passengers in the below table by the survival status and the economic status (crew, or first, second, or third class). We see, for example, there were 122 people in first class who did not survive, and 203 people in the first class who survived.

		Survived		Total
		no	yes	
Class	Crew	673	212	885
	First	122	203	325
	Second	167	118	285
	Third	528	178	706
	Total	1490	711	2201

Although this table is useful in understanding the numbers of people who died and survived in different classes, it is difficult to compare the distributions since there were many more non-survivors in the dataset. For comparison purposes, it is better to compute the percentage of survivors and non-survivors for each class.

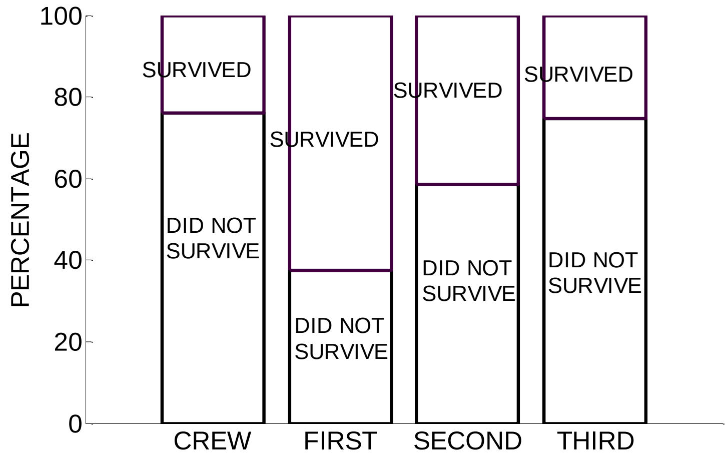
We can get these percentages from the table by dividing each basic count by the corresponding row total. For example, the percentage of first-class people who survived is 203 / 325 = 62%. If we compute these row percentages for all of the table entries,

		Survived		Total
		No	yes	
Class	crew	673/885	212/885	885
	first	122/325	203/325	325
	second	167/285	118/285	285
	third	528/706	178/706	706
	TOTAL	1490/2201	711/2201	2201

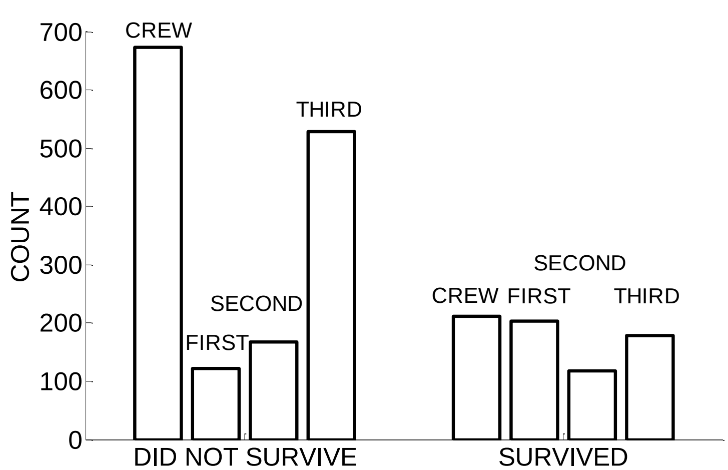
we obtain the row percentages.

		Survived		TOTAL
		no	yes	
Class	Crew	76%	24%	100%
	First	38%	62%	100%
	Second	59%	41%	100%
	Third	75%	25%	100%
	Total	68%	32%	100%

We see from the row percentages that there is a strong association between class and survival. Overall, 32% of the passengers survived. However, a majority (62%) of first-class passengers survived compared to a survival rate of 41% for second-class passengers and a survival rate of 25% for third-class passengers and crew. To display this association graphically we can use a set of stacked bar charts. Each bar corresponds to a passenger class, and the bar is divided by the survival status. This graph clearly shows the high survival rate of the first-class passengers.



In the above analysis, we divided each table count by the row total and found the row percentages. This told us the percentage of each class group that survived and did not survive. But we can look at the association in the table a different way. To compare the class of the non-survivors with the class of the survivors, we can construct side-by-side bar charts of the two sets of column frequencies, as shown below.



This graph is informative, but it is difficult to compare the two sets of bars, since there are different numbers of passengers in the two classes. It is easier to compare the two sets of frequencies if they are converted to percentages. Again we start with our basic count table

		Survived		Total
		no	yes	
Class	Crew	673	212	885
	First	122	203	325
	second	167	118	285
	Third	528	178	706
	Total	1490	711	2201

and divide the counts by the column totals

		Survived		TOTAL
		No	yes	
CLASS	crew	673/1490	212/711	885/2201
	first	122/1490	203/711	325/2201
	second	167/1490	118/711	285/2201
	third	528/1490	178/711	706/2201
	TOTAL	1490	711	2201

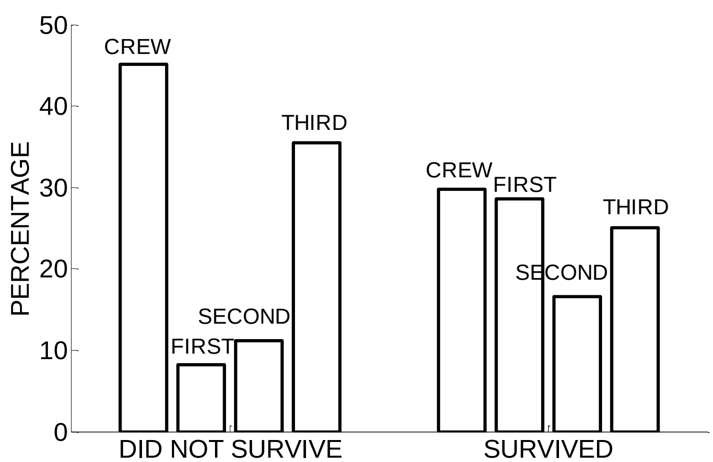
to obtain the column percentages

		Survived		TOTAL
		no	yes	

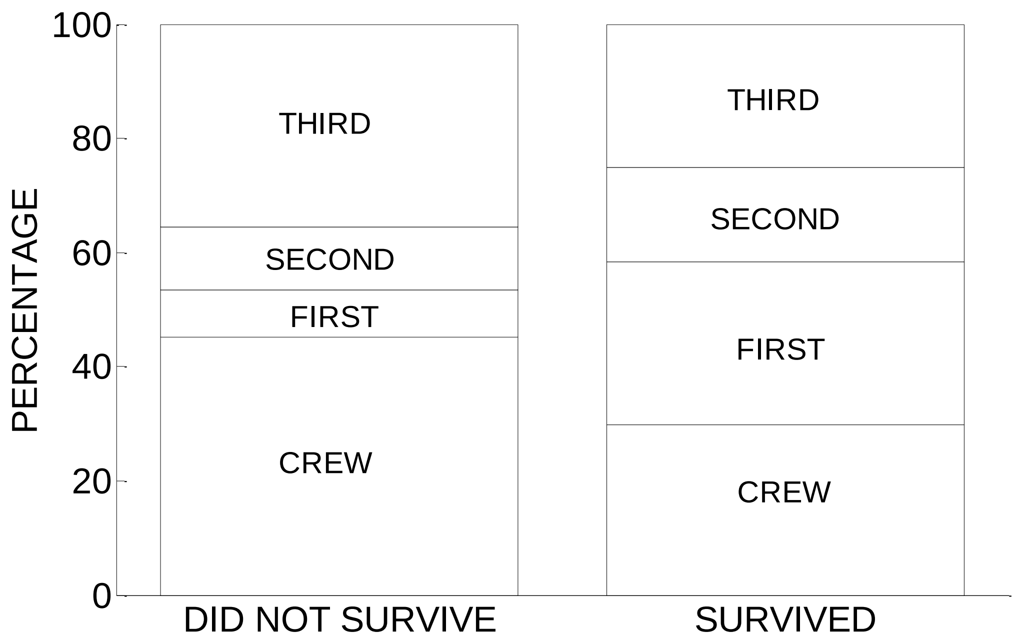
CLASS	Crew	45%	30%	40%
	First	8%	29%	15%
	Second	11%	17%	13%
	Third	35%	25%	32%
	TOTAL	100%	100%	100%

Let's interpret these column percentages. Of the 2201 passengers in the Titanic, 40%, 15%, 13%, and 32% were respectively crew, first-class, second-class, and third-class passengers. However, there was a different composition if you consider only the 711 people who survived the accident. Of this group, there were about equal percentages of crew and first-class passengers, 25% third class, and 17% that were second class. But more importantly, if the passengers who died, almost half of them were crew and only 8% were first-class.

We can display these two sets of column percentages in several ways. We can use side-by-side bar charts, where the percentage (instead of a frequency) is plotted.



Alternately, we can use two stacked bar charts, where each bar chart contains the percentages of class memberships for a survival group.



One can study the association in a two-way table by computing row percentages or column percentages. Which way to go depends on the particular application and ease of interpretation.

In this example, is it easier to talk about

- the survival rate of different classes of passengers

or is it easier to discuss

- the class status of the group of passengers who survived and the status of the passengers who did not survive?

Here the author thinks it makes more sense to talk about how survival depends on the economic status of the passenger, so I would prefer computing row percentages and using a corresponding graph (such as a set of segmented bar charts) to communicate the difference in row percentages across class.

PRACTICE: LOOKING AT THE AGE AND GENDER OF THE TITANIC PASSENGERS

It is also interesting to relate the survival of the Titanic passengers with respect to their age and their gender.

1. The below table classifies the passengers with respect to their survival (no or yes) against their age (child or adult). For each age, compute the percentage of survivors and percentages of nonsurvivors and place your results in the second table. Were children more or less likely to survive than adults? Explain.

		Survived?	
		No	Yes
Age	Child	52	57
	Adult	1438	654

		Survived?	
		No	Yes
Age	Child		
	Adult		

2. Using two stacked bar charts, show how the survival of the passenger depends on age.

3. In a similar fashion, compute relevant percentages to see if the survival status of the passengers depended on their gender. Place your percentages in the second table and describe your conclusions.

		Survived?	
		No	Yes
Gender	Female	126	344
	Male	1364	367

		Survived?	
		No	Yes
Gender	Female		
	Male		

4. Use stacked bar charts to show how the survival depends on gender.

5. We have seen that the survival status of the Titanic passengers was dependent on their economic status and also dependent on their gender. That raises the question: was there any relationship between economic status and gender? By using the below table and computing relevant row or column percentages, answer this question.

		Class			
		Crew	First	Second	Third
Gender	Female	23	145	106	196
	Male	862	180	179	540

6. Does it make sense that children and women were more likely to survive this disaster? Explain.

HANDS-ON ACTIVITY: PREDICTABLE PAIRS

DESCRIPTION: This is a class participation activity, where students will indicate, through a show of hands, if they have seen or not seen different movies. If two movies are similar in some way, then it is possible that one can use information about students seeing one movie to predict whether they have seen (or not seen) a second movie.

1. Find a movie that roughly half of the class has seen. (You can ask the class for possible suggestions of movie titles.) Write down the name of the movie.

2. Now find a second movie that you think is similar in some way to the first movie. Write down the name of this second movie. ______________________

- For the students that ***have seen*** the first movie, ask if they have or have not seen the second movie.
- For the students who ***have not seen*** the first movie, ask if they have or have not seen the second movie.

Put your data in the table below.

	Seen 1st movie?	
Send 2nd movie?	Yes	No
Yes		
No		

3. By the computation of relevant row or column percentages, answer the question: Is there a relationship between seeing the first movie and seeing the second movie? Explain.

4. Repeat the above analysis for a movie that is a romantic comedy.

5. Repeat the above analysis for an action movie with some violence.

SIMPSON'S PARADOX

The following two-way table classifies hypothetical hospital patients according to the hospital that treated them and whether they survived or died:

	Survived	Died	Total
Hospital A	800	200	1000
Hospital B	900	100	1000

Which hospital had the better survival rate? We can answer this question by computing the proportion of hospital A's patients who survived and the proportion of hospital B's patients who survived.

	Survival Rate
Hospital A	800/1000 = 0.8
Hospital B	900/1000=0.9

We see that Hospital B saved the higher percentage of its patients.

Before you decide that Hospital B is the better hospital, suppose that when we further categorize each patient according to whether they were in good condition or poor condition prior to treatment we obtain the following two-way tables:

Good condition:

	Survived	Died	Total
Hospital A	590	10	600
Hospital B	870	30	900

Poor condition:

	Survived	Died	Total
Hospital A	210	190	400
Hospital B	30	70	100

Suppose that we compute the recovery rates for the two hospitals for each condition. Among those who were in good condition, we compute the recovery rates for the two hospitals by dividing the number who survived (590 and 871) by the number of good patients (600 and 900).

Good condition	Survival rate

Hospital A	590/600=0.983
Hospital B	870/890=0.967

Hospital A saved the greater percentage of its patients who had been in good condition.

Let's look next at the patients who were in poor condition. Using the counts in the second table, we compute the recovery rates for poor condition patients.

Poor patients	Survival Rate
Hospital A	210/400=0.525
Hospital B	30/100=0.30

Hospital A also saved a greater proportion of patients in good condition.

We have discovered a surprising result known as Simpson's paradox. Hospital A has the higher recovery rate for each type of patient. But when we aggregate the table, the association goes the other direction – Hospital B has the higher recovery rate for all patients. In generally, Simpson's paradox refers to the fact that aggregate proportions can reverse the direction of the relationship seen in the individual pieces.

Why are we observing Simpson's paradox in this illustration? First, note that good patients tend to survive more often than poor patients. If we look at the original set of tables, we see that 900/1000 = 90% of the patients going to Hospital B are good patients; in contrast only 400/1000 = 40% of the Hospital A patients are good. So the reason why Hospital B has an overall higher survival rate is due primarily to the fact that it is treating better patients.

PRACTICE: SIMPSON'S PARADOX

In the two-year period 1995-1996, the baseball player Derek Jeter had 195 hits in 630 at-bats (opportunities) and David Justice had 149 hits in 551 at-bats.

1. Which player had the higher batting average during this period (batting average is defined as hits divided by at-bats).

2. Suppose we break the above data by year – we get the following table.

	1995		1996	
	Derek Jeter	David Justice	Derek Jeter	David Justice
Hits	12	104	183	45
At-bats	48	411	582	140

Who was the better hitter in 1995? Who was the better in 1996?

3. Explain how this example demonstrates Simpson's paradox.

WRAP-UP

In this topic, we discussed how to summarize relationships between two categorical variables when the data are presented in a two-way table. One first computes row percentages (or column percentages) and then looks for association by comparing the sets of row percentages. If there are differences in the row percentages across rows, this indicates that the outcome of the column variable is dependent on the outcome of the row variable. Side-by-side bar charts or segmented bar charts are useful for displaying the relationship that one finds in a two-way table.

EXERCISES

1. **Movie Ratings**

On the Internet Movie Database (www.imbd.com), people are given the opportunity to rate movies on a scale from 1 (hate the movie) to 10 (love the movie). For 27337 people who rated the movie *Die Hard*, the first table classifies these people by their rating and their gender. The second table classifies the people by their rating and their age.

Rating

		9, 10	6, 7, 8	5 or lower
Gender	Males	9228	13154	1634
	Female	891	1862	468

		Rating		
		9, 10	6, 7, 8	5 or lower
	under 18	423	333	43
Age	18-29	6490	9360	1186
	30-44	2308	3878	620
	45+	518	919	193

a. Explore the relationship between gender and rating in the first table by computing relevant conditional percentages. Use a graph to show this relationship. Is this movie more popular among men or women?

b. Explore the relationship between age and rating in the second table. Is Die Hard more popular among viewers of a particular age?

c. Do you think that the sample of people who rated this movie is representative of all people who watched this movie? Explain.

2. **Movie Ratings**

The following table classifies people who rated the movie *Sleepless in Seattle* by rating and gender. By computing relevant row or column percentages, investigate if the ratings of males is different from the ratings of females. Also use a graph to display the relationship. Are you surprised by the findings?

		Rating		
		8,9,10	5,6,7	1,2,3,4
Gender	Males	2217	3649	754
	Females	1059	835	178

3. Available Vehicles to Households

The following table from U.S. Census Bureau classifies the American households for five years with respect to the number of vehicles (unit = 1000 households). We see from the table that, for example, there were 11,417 (thousand) households in 1960 that did not have a vehicle.

	Year				
Vehicles Available	1960	1970	1980	1990	2000
None	11,417	11,081	10,390	10,602	10,861
1	30,189	30,268	28,565	31,039	36,124
2	10,074	18,600	27,347	34,361	40,462
3 or more	1,342	3,495	14,088	15,945	18,034

a. For each year, compute the percentage of households having different number of available vehicles.
b. By comparing the column percentages, describe how the number of vehicles available to American households has changed over time.

4. Participation in Selected Sports Activities

The following table gives the number (in thousands) of Americans in 2001 that participated in selected sports activities. (These data were taken from *Statistical Abstract of the United States* 2003.)

	Age of Participant							
Sport	7-11 yrs	12-17 yrs	18-24yrs	25-34 yrs	35-44 yrs	45-54 yrs	55-64 yrs	65 yrs and over
Basketball	6356	7818	3955	4397	3616	1278	422	361
Bowling	5330	5893	6806	8597	7205	3649	1265	1558
Exercise	2417	3550	6936	12332	14692	13616	8237	9438

walking								
Golf	1011	2264	3022	5197	5906	4754	2033	2450
Soccer	5867	1811	1312	1115	972	447	146	196
Tennis	728	1963	1487	2256	2157	1217	535	569

By computing relevant conditional percentages, explore the relationship between sport and age of the participant. Classify the sports into those that are enjoyed by "young" people and those that are enjoyed by "old" people.

5. **Cigarette Smoking among Americans**

The following table (from *Statistical Abstract of the United States* 2003) gives the percent of people in each category who smoke at least "some days" for the years 1990, 1995, 2000. By constructing suitable graphs, show how the tendency to smoke depends on gender and age. Also, use graphs to show how the percent of people who smoke has changed over time.

	AGE	1990	1995	2000
Male, total		28.4%	27%	25.7%
	18 to 24 years	26.6%	27.8%	25.7%
	25 to 34 years	31.6%	29.5%	29.0%
	35 to 44 years	34.5%	31.5%	30.2%
	45 to 64 years	29.3%	27.1%	26.4%
	65 years and over	14.6%	14.9%	10.2%
Female, total		22.8%	22.6%	21.0%
	18 to 24 years	22.5%	21.8%	25.1%
	25 to 34 years	28.2%	26.4%	22.5%
	35 to 44 years	24.8%	27.1%	26.2%
	45 to 64 years	24.8%	24.0%	21.6%
	65 years and over	11.5%	11.5 %	9.3%

6. **Spending Money**

How do Americans spend their money? The following table gives the total expenditures of U.S. consumers (in $1000) in several categories for three years. from *Statistical Abstract of the United States* 2003.

	1990	1995	2000
Food	4296	4505	5158
Alcoholic beverages	293	277	372
Housing	8703	10458	12137
Apparel and services	1618	1704	1856
Transportation	5120	6014	7404
Health care	1480	1732	2066
Entertainment	1422	1612	1863
Personal insurance and pensions	2592	2954	3365
TOTAL	25524	29256	34221

a. For each year, compute the percentage of the total expenditure that was spent on the different categories.

b. Construct a graph to compare the percentages you computed in (a).

c. Has the pattern of spending in the different categories changed over time? Explain.

7. **Births in the United States**

When are babies born? All of the births in the United States in the year 1978 were classified by month and day of the week. The unit is thousands of births – so, for example, there were a total of 38 thousand babies born on Sundays in January.

	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	TOTAL
January	38	45	46	36	37	37	32	271
February	31	36	38	37	37	37	33	249
March	31	37	38	46	46	47	32	277
April	37	36	37	35	35	36	39	255

May	30	44	47	46	36	37	32	272
June	31	37	38	38	47	48	32	271
July	42	49	39	39	41	41	44	295
August	34	40	52	50	51	41	35	303
September	34	39	42	42	41	51	44	293
October	41	48	50	39	39	39	34	290
November	32	39	40	47	46	38	33	275
December	40	37	39	39	39	48	42	284
TOTAL	421	487	506	494	495	500	432	3335

a. Compute the proportion of babies that were born on different days of the week. Are there particular days of the week when babies are less likely to be born? Can you explain the reason why these days are not popular?

b. Compute the proportion of babies born in different months of the year. Are there particular seasons where babies are more or less likely to be born? Can you explain why?

8. **Children's Living Arrangements**

The following table classifies the living arrangements of American children (under 18 years of age) in 1993 according to their race/Hispanic origin and which parent(s) they lived with.

	Both	Just mom	Just dad	Neither
White	40,842,340	9,017,140	2,121,680	1,060,840
Black	3,833,640	5,750,460	319,470	745,430
Hispanic	4,974,720	2,176,440	310,920	310,920

Analyze these data to address the issue whether a relationship exists between race/Hispanic origin and parental living arrangements. Write a paragraph reporting your findings, supported by appropriate calculations and visual displays.

9. **Throwing and Batting Side of Baseball Players**

For all of the major league baseball players who where born in 1970 or later, the following table classifies the players by their throwing hand (left or right) and their batting side (left only, right only, or both sides).

		Throwing hand	
		Left	Right
Batting	Both	58	587
side	Left only	1216	610
	Right only	195	4287

a. If a baseball player is a southpaw (that is, throws with his left hand), what is the chance that he will also bat from the left side only?

b. If a player throws right, what is the chance that he will bat from the left side only?

c. A switch-hitter is a player that bats from both sides. Is it more likely for a left-hand thrower or a right-hand thrower to be a switch-hitter? Can you explain why this might be the case? (HINT: In baseball, most pitchers are right-handed and it is easier to get a hit from a right-handed pitcher if you bat from the left side.)

10. Purchases by Gender

In a survey, students in a statistics class specified the number of pairs of shoes they owned and the number of movie DVDs they owned. The following tables classify the responses by gender and number of pairs of shoes, and by gender and number of DVDs owned.

Number of pairs of shoes owned	Female	Male
Eight or less	53	181
Between 9 and 20	225	34
Betweeen 21 and 50	124	6
Over 50	11	0

Number of DVDs owned	Female	Male
None	16	10
Between 1 and 10	140	61
Between 11 and 20	115	57
Between 21 and 50	118	56
Over 50	32	36

a. For each gender, compute the percentage of students who own Eight or less, Between 9 and 20, Between 21 and 50, and Over 50 pairs of shoes. Graph the two sets of percentages, and explain if there is a relationship between gender and number of pairs of shoes owned.

b. For each gender, compute the percentage of students who own no DVDs, between 1 and 10 DVDs, and so on. By comparing the two sets of percentages, is there a relationship between gender and the number of DVDs owned?

11. **The Pop and Soda Controversy**

On the Pop vs. Soda web page at www.**pop**vs**soda**.com, American readers were asked to give their state of their home town and answer the question "What generic word do you use to describe carbonated soft drinks?" The following table gives the answer to the question for respondents from six different states.

State	pop	soda	coke	other
Alabama	30	260	2161	198
California	669	12843	2301	6006
Kansas	1938	483	266	212
Maine	25	993	15	57
New Jersey	96	5830	212	113
Ohio	12317	1730	329	343

a. For each state, compute the percentages of people giving the answers "pop", "soda", "coke" and "other".

b. Graph the six sets of state percentages using parallel bar plots.

c. Some states in the U.S. are regarded as “pop” states, other states are regarded as “soda” states, and other states are “coke” states. Based on your computations, describe what regions of the United States correspond to each type.

d. What is your answer to this question? Do you believe that your answer is consistent with the majority of people from your home town state? (Check the website if your state is not included in the above list.)

12. **The Ohio Graduation Test**

All students in Ohio must currently pass the Ohio Graduation Test (OGT) that measures the level of reading, writing, mathematics, science, and social studies skills of students at the end of the sophomore year. Each school is rated on the basis of their students’ performance in the OGT and a school earns a state indicator if 75% or more students at that school obtain a proficient level on a particular section of the test.

All of the public high schools in Ohio were classified as either earning a state indicator or not earning a state indicator for the math and science sections of the OGT and the following table gives the corresponding two-way table of counts.

		State indicator in science?	
		No	yes
State indicator in math?	no	73	0
	yes	163	373

a. What percentage of high schools earn a state indicator in math?

b. What percentage of high schools earn a state indicator in science?

c. What percentage of high schools earned state indicators in both math and science?

d. Of the schools earning a state indicator in math, what percentage also earned a state indicator in science?

e. Of the schools not earning a state indicator in math, what percentage also earned a state indicator in science?

f. Is there an association between a school’s performance on the OGT math section and the OGT science section? Explain why this association exists.

g. Suppose the high schools are divided into “small” and “large” schools depending on their enrollment. The following table classifies the schools by their size and their performance on the OGT science test. By computing relevant percentages, check if there is an association between a school’s size and its performance on the science OGT test.

		State indicator in science?	
		No	Yes
Size of school	small	108	195
	large	128	178

13. **Salaries for Associate and Bachelor Degrees**

An administrator is interested in comparing the salaries of students who receive Associate Degrees with the salaries of students who receive Bachelor Degrees. For 5000 recent Associate Degree graduates, suppose we classify the graduate by Subject Area (health, business, other) and the Salary (low or high), obtaining the following table.

Associate Degree graduates:

	Salary		
Subject Area	Low	High	TOTAL
Health	250	2250	2500
Business	600	900	1500
Other	700	300	1000
TOTAL	1550	3450	5000

Likewise, suppose 5000 Bachelor Degree graduates are classified with respect to Subject Area and Salary.

Bachelor Degree graduates:

	Salary		
Subject Area	Low	High	TOTAL
Health	25	475	2500

Business	700	1300	1500
Other	1625	875	1000
TOTAL	2350	2650	5000

a. What proportion of Associate Degree graduates earn high salaries? What proportion of Bachelor Degree graduates earn high salaries? Is it accurate to say that a greater proportion of Associate Degree holders get high salaries?

b. Suppose that you are a Business major. What proportion of these majors earn high salaries if you have an Associate Degree. What proportion of these majors earn high salaries if you have a Bachelor Degree? Is it advantageous to get a Bachelor Degree?

c. Suppose that you are a Health major. Are you more likely to get a high salary with an Associate Degree or a Bachelor Degree?

d. Suppose that you are an "Other" major. Are you more likely to get a high salary with an Associate Degree or a Bachelor Degree?

e. Based on your answers to parts b, c, d, which type of degree gives you a higher salary? Why does the comparison using the totals in part a go the other direction?

f. How does this example illustrate Simpson's paradox?

14. **Berkeley Graduate Admissions**

The University of California at Berkeley was charged with having discriminated against women in their graduate admissions process for the fall quarter of 1973. The table below identifies the number of acceptances and denials for both men and women applicants in each of the six largest graduate programs at the institution at that time:

	Men accepted	Men denied	Women accepted	Women denied
Program A	511	314	89	19
Program B	352	208	17	8
Program C	120	205	202	391
Program D	137	270	132	243
Program E	53	138	95	298
Program F	22	351	24	317

TOTAL				

a. Start by ignoring the program distinction, collapsing the data into a two-way table of gender by admission status. To do this, find the total number of men accepted and denied and the total number of women accepted and denied. Construct a table such as the one below:

	Admitted	Denied	TOTAL
Men			
Women			
TOTAL			

b. Consider for the moment just the men applicants. Of the men who applied to one of these programs, what proportion was admitted? Now consider the women applicants; what proportion of them were admitted? Do these proportions seem to support the claim that men were given preferential treatment in admissions decisions?

c. To try to isolate the program or programs responsible for the mistreatment of women applicants, calculate the proportion of men and the proportion of women within each program who were admitted. Record your results in a table such as the one below.

	Proportion of men admitted	Proportion of women admitted
Program A		
Program B		
Program C		
Program D		
Program E		
Program F		

d. Does it seem as if any program is responsible for the large discrepancy between men and women in the overall proportions admitted?

e. Reason from the data given to explain how it happened that men had a much higher rate of admission overall even though women had higher rates in most programs and no program favored men very strongly.