Homework 1
30. More politics. Look once more at the table summarizing the political views of Intro Stats students in Exercise 26. 
a) Produce a graphical display comparing the conditional distributions of males and females among the three categories of politics.
b) Comment briefly on what you see from the display in a.

31. Back to school. Examine the table about ethnicity and acceptance for the Houston Independent School District’s magnet schools program, shown in Exercise 29. Does it appear that the admissions decisions are made independent of the applicant’s ethnicity? Explain.

32. Cars. A survey of autos parked in student and staff lots at a large university classified the brands by country of origin, as seen in the table.

<table>
<thead>
<tr>
<th>Driver</th>
<th>Student</th>
<th>Staff</th>
</tr>
</thead>
<tbody>
<tr>
<td>American</td>
<td>107</td>
<td>105</td>
</tr>
<tr>
<td>European</td>
<td>33</td>
<td>12</td>
</tr>
<tr>
<td>Asian</td>
<td>55</td>
<td>47</td>
</tr>
</tbody>
</table>

a) What percent of all the cars surveyed were foreign? 
b) What percent of the American cars were owned by students? 
c) What percent of the students owned American cars? 
d) What is the marginal distribution of origin? 
e) What are the conditional distributions of origin by driver classification? 
f) Do you think that the origin of the car is independent of the type of driver? Explain.

33. Weather forecasts. Just how accurate are the weather forecasts we hear every day? The following table compares the daily forecast with a city’s actual weather for a year:

<table>
<thead>
<tr>
<th>Actual Weather</th>
<th>Rain</th>
<th>No rain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forecast</td>
<td>27</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>268</td>
</tr>
</tbody>
</table>

a) On what percent of days did it actually rain? 
b) On what percent of days was rain predicted? 
c) What percent of the time was the forecast correct? 
d) Do you see evidence of an association between the type of weather and the ability of forecasters to make an accurate prediction? Write a brief explanation, including an appropriate graph.

34. Twins. In 2000, the Journal of the American Medical Association (JAMA) published a study that examined pregnancies that resulted in the birth of twins. Births were classified as preterm with intervention (induced labor or cesarean), preterm without procedures, or term/post-term. Researchers also classified the pregnancies by the level of prenatal medical care the mother received (inadequate, adequate, or intensive). The data, from the years 1995–1997, are summarized in the table below. Figures are in thousands of births. (JAMA 284 [2000]:335–341)

<table>
<thead>
<tr>
<th>Level of Prenatal Care</th>
<th>Preterm (induced or cesarean)</th>
<th>Preterm (without procedures)</th>
<th>Term or post-term</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intensive</td>
<td>18</td>
<td>15</td>
<td>28</td>
<td>61</td>
</tr>
<tr>
<td>Adequate</td>
<td>46</td>
<td>43</td>
<td>65</td>
<td>154</td>
</tr>
<tr>
<td>Inadequate</td>
<td>12</td>
<td>13</td>
<td>38</td>
<td>63</td>
</tr>
<tr>
<td>Total</td>
<td>76</td>
<td>71</td>
<td>131</td>
<td>278</td>
</tr>
</tbody>
</table>

a) What percent of these mothers received inadequate medical care during their pregnancies? 
b) What percent of all twin births were preterm? 
c) Among the mothers who received inadequate medical care, what percent of the twin births were preterm? 
d) Create an appropriate graph comparing the outcomes of these pregnancies by the level of medical care the mother received. 
e) Write a few sentences describing the association between these two variables.

35. Blood pressure. A company held a blood pressure screening clinic for its employees. The results are summarized in the table below by age group and blood pressure level:

<table>
<thead>
<tr>
<th>Blood Pressure</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Under 30</td>
</tr>
<tr>
<td>Low</td>
<td>27</td>
</tr>
<tr>
<td>Normal</td>
<td>48</td>
</tr>
<tr>
<td>High</td>
<td>23</td>
</tr>
</tbody>
</table>

a) Find the marginal distribution of blood pressure level. 
b) Find the conditional distribution of blood pressure level within each age group. 
c) Compare these distributions with a segmented bar graph. 
d) Write a brief description of the association between age and blood pressure among these employees. 
e) Does this prove that people’s blood pressure increases as they age? Explain.

36. Obesity and exercise. The Centers for Disease Control and Prevention (CDC) has estimated that 19.8% of Americans over 15 years old are obese. The CDC conducts a survey on obesity and various behaviors. Here is a table on self-reported exercise classified by body mass index (BMI):

<table>
<thead>
<tr>
<th>Blood Pressure</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Under 30</td>
</tr>
<tr>
<td>Low</td>
<td>27</td>
</tr>
<tr>
<td>Normal</td>
<td>48</td>
</tr>
<tr>
<td>High</td>
<td>23</td>
</tr>
</tbody>
</table>
1. **Statistics in print.** Find a histogram that shows the distribution of a variable in a newspaper or magazine article.
   a) Does the article identify the W's?
   b) Discuss whether the display is appropriate for the data.
   c) Discuss what the display reveals about the variable and its distribution.
   d) Does the article accurately describe and interpret the data? Explain.

2. **Not a histogram.** Find a graph other than a histogram that shows the distribution of a quantitative variable in a newspaper or magazine article.
   a) Does the article identify the W's?
   b) Discuss whether the display is appropriate for the data.
   c) Discuss what the display reveals about the variable and its distribution.
   d) Does the article accurately describe and interpret the data? Explain.

3. **In the news.** Find an article in a newspaper or a magazine that discusses an "average."
   a) Does the article discuss the W's for the data?
   b) What are the units of the variable?
   c) Is the average used the median or the mean? How can you tell?
   d) Is the choice of median or mean appropriate for the situation? Explain.

4. **In the news II.** Find an article in a newspaper or a magazine that discusses a measure of spread.
   a) Does the article discuss the W's for the data?
   b) What are the units of the variable?
   c) Does the article use the range, IQR, or standard deviation?
   d) Is the choice of measure of spread appropriate for the situation? Explain.

5. **Thinking about shape.** Would you expect distributions of these variables to be uniform, unimodal, or bimodal? Symmetric or skewed? Explain why.
   a) The number of speeding tickets each student in the senior class of a college has ever had.
   b) Players' scores (number of strokes) at the U.S. Open golf tournament in a given year.
   c) Weights of female babies born in a particular hospital over the course of a year.
   d) The length of average hair on the heads of students in a large class.

6. **More shapes.** Would you expect distributions of these variables to be uniform, unimodal, or bimodal? Symmetric or skewed? Explain why.
   a) Ages of people at a Little League game.
   b) Number of siblings of people in your class.
   c) Pulse rates of college-age males.
   d) Number of times each face of a die shows in 100 tosses.

7. **Sugar in cereals.** The histogram displays the sugar content (as a percent of weight) of 49 brands of breakfast cereals.

   a) Describe this distribution.
   b) What do you think might account for this shape?

8. **Singers.** The display shows the heights of some of the singers in a chorus, collected so that the singers could be positioned on stage with shorter ones in front and taller ones in back.

   a) Describe the distribution.
   b) Can you account for the features you see here?

9. **Vineyards.** The histogram shows the sizes (in acres) of 36 vineyards in the Finger Lakes region of New York.

   a) Approximately what percentage of these vineyards are under 60 acres?
   b) Write a brief description of this distribution (shape, center, spread, unusual features).
27. Movie budgets. The histogram shows the budgets (in millions of dollars) of major release movies in 2005.

An industry publication reports that the average movie costs $35 million to make, but a watchdog group concerned with rising ticket prices says that the average cost is $46.8 million. What statistic do you think each group is using? Explain.

28. Sick days. During contract negotiations, a company seeks to change the number of sick days employees may take, saying that the annual "average" is 7 days of absence per employee. The union negotiators counter that the "average" employee misses only 3 days of work each year. Explain how both sides might be correct, identifying the measure of center you think each side is using and why the difference might exist.

29. Standard deviation I. For each lettered part, a through c, examine the two given sets of numbers. Without doing any calculations, decide which set has the larger standard deviation and explain why. Then check by finding the standard deviations by hand.

<table>
<thead>
<tr>
<th>Set 1</th>
<th>Set 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) 3, 5, 6, 7, 9</td>
<td>2, 4, 6, 8, 10</td>
</tr>
<tr>
<td>b) 10, 14, 15, 16, 20</td>
<td>10, 11, 15, 19, 20</td>
</tr>
<tr>
<td>c) 2, 6, 9, 11, 14</td>
<td>82, 86, 89, 91, 94</td>
</tr>
</tbody>
</table>

30. Standard deviation II. For each lettered part, a through c, examine the two given sets of numbers. Without doing any calculations, decide which set has the larger standard deviation and explain why. Then check by finding the standard deviations by hand.

<table>
<thead>
<tr>
<th>Set 1</th>
<th>Set 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) 4, 7, 7, 7, 10</td>
<td>4, 6, 7, 8, 10</td>
</tr>
<tr>
<td>b) 100, 140, 150, 160, 200</td>
<td>10, 50, 60, 70, 110</td>
</tr>
<tr>
<td>c) 10, 16, 18, 20, 22, 28</td>
<td>48, 56, 58, 60, 62, 70</td>
</tr>
</tbody>
</table>

31. Payroll. A small warehouse employs a supervisor at $1200 a week, an inventory manager at $700 a week, six stock boys at $400 a week, and four drivers at $500 a week.

a) Find the mean and median wage.

b) How many employees earn more than the mean wage?

c) Which measure of center best describes a typical wage at this company, the mean or the median?

d) Which measure of spread would best describe the payroll, the range, the IQR, or the standard deviation? Why?

32. Singers. The frequency table shows the heights (in inches) of 130 members of a choir.

<table>
<thead>
<tr>
<th>Height</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>2</td>
</tr>
<tr>
<td>61</td>
<td>6</td>
</tr>
<tr>
<td>62</td>
<td>9</td>
</tr>
<tr>
<td>63</td>
<td>7</td>
</tr>
<tr>
<td>64</td>
<td>5</td>
</tr>
<tr>
<td>65</td>
<td>20</td>
</tr>
<tr>
<td>66</td>
<td>18</td>
</tr>
<tr>
<td>67</td>
<td>7</td>
</tr>
<tr>
<td>68</td>
<td>12</td>
</tr>
</tbody>
</table>

a) Find the median and IQR.

b) Find the mean and standard deviation.

c) Display these data with a histogram.

d) Write a few sentences describing the distribution of heights.

33. Gasoline. In March 2006, 16 gas stations in Grand Junction, Colorado, posted these prices for a gallon of regular gasoline:

<table>
<thead>
<tr>
<th>$2.22</th>
<th>$2.21</th>
<th>$2.45</th>
<th>$2.24</th>
</tr>
</thead>
<tbody>
<tr>
<td>$2.27</td>
<td>$2.28</td>
<td>$2.27</td>
<td>$2.23</td>
</tr>
<tr>
<td>$2.26</td>
<td>$2.46</td>
<td>$2.29</td>
<td>$2.32</td>
</tr>
<tr>
<td>$2.36</td>
<td>$2.38</td>
<td>$2.33</td>
<td>$2.27</td>
</tr>
</tbody>
</table>

a) Make a stem-and-leaf display of these gas prices. Use split stems; for example, use two 2.2 stems, one for prices between $2.2 and $2.24, the other for prices from $2.25 to $2.29.

b) Describe the shape, center, and spread of this distribution.

c) What unusual feature do you see?

34. The Great One. During his 20 seasons in the NHL, Wayne Gretzky scored 50% more points than anyone who ever played professional hockey. He accomplished this amazing feat while playing in 280 fewer games than Gordie Howe, the previous record holder. Here are the number of games Gretzky played during each season:

| 79, 80, 80, 80, 74, 80, 80, 79, 64, 78, 73, 78, 74, 45, 81, 48, 80, 82, 82, 70 |

a) Create a stem-and-leaf display for these data, using split stems.

b) Describe the shape of the distribution.

C) Describe the center and spread of this distribution.

d) What unusual feature do you see? What might explain this?

35. Sip size. Researchers in Cornell University's Food Science department study how people experience foods. One study considered how much liquid people typically take into their mouths in one "sip" (in milliliters). The researchers also recorded the height (meters) and weight (kilograms) of the participants. Here are histograms of three of the variables from that study:
38. Wayne Gretzky. In Exercise 34 you examined the number of games played by hockey great Wayne Gretzky during his 20-year career in the NHL.
   a) Would you use the median or the mean to describe the center of this distribution? Why?
   b) Find the median.
   c) Without actually finding the mean, would you expect it to be higher or lower than the median? Explain.

39. Home runs. The stem-and-leaf display shows the number of home runs hit by Mark McGwire during the 1986–2001 seasons. Describe the distribution, mentioning its shape and any unusual features.

<table>
<thead>
<tr>
<th>Home Runs</th>
</tr>
</thead>
<tbody>
<tr>
<td>70</td>
</tr>
<tr>
<td>65</td>
</tr>
<tr>
<td>528</td>
</tr>
<tr>
<td>429</td>
</tr>
<tr>
<td>322399</td>
</tr>
<tr>
<td>229</td>
</tr>
<tr>
<td>1099</td>
</tr>
<tr>
<td>70 means 70</td>
</tr>
</tbody>
</table>

40. Bird species. The Cornell Lab of Ornithology holds an annual Christmas Bird Count (www.birdsource.org), in which bird watchers at various locations around the country see how many different species of birds they can spot. Here are some of the counts reported from sites in Texas during the 1999 event:

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>228</td>
<td>178</td>
<td>186</td>
<td>162</td>
<td>206</td>
</tr>
<tr>
<td>166</td>
<td>163</td>
<td>183</td>
<td>181</td>
<td>206</td>
</tr>
<tr>
<td>177</td>
<td>175</td>
<td>167</td>
<td>162</td>
<td></td>
</tr>
<tr>
<td>160</td>
<td>157</td>
<td>156</td>
<td>153</td>
<td>153</td>
</tr>
</tbody>
</table>

   a) Create a stem-and-leaf display of these data.
   b) Write a brief description of the distribution. Be sure to discuss the overall shape as well as any unusual features.

41. Hurricanes 2006. The data below give the number of hurricanes classified as major hurricanes in the Atlantic Ocean each year from 1944 through 2006, as reported by NOAA (www.nhc.noaa.gov):

<p>| | | | | | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>32,1,2,4,3,7,2,3,5,5,2,4,2,2,6,2,5,1,3,3,1,0,3,2,1,0,1,2,3,2,1,2,2,3,1,1,1,3,0,1,3,2,1,2,1,1,1,6,1,3,5,3,2,2,3,6,7,2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

   a) Create a dotplot or stem-and-leaf display of these data.
   b) Describe the distribution.

42. Horsepower. Create a stem-and-leaf display for these horsepowers of autos reviewed by Consumer Reports one year, and describe the distribution:

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>155</td>
<td>103</td>
<td>130</td>
<td>80</td>
<td>65</td>
</tr>
<tr>
<td>142</td>
<td>123</td>
<td>129</td>
<td>71</td>
<td>69</td>
</tr>
<tr>
<td>125</td>
<td>115</td>
<td>138</td>
<td>68</td>
<td>78</td>
</tr>
<tr>
<td>150</td>
<td>133</td>
<td>135</td>
<td>90</td>
<td>97</td>
</tr>
<tr>
<td>68</td>
<td>105</td>
<td>88</td>
<td>115</td>
<td>110</td>
</tr>
<tr>
<td>95</td>
<td>85</td>
<td>109</td>
<td>115</td>
<td>71</td>
</tr>
<tr>
<td>97</td>
<td>110</td>
<td>65</td>
<td>90</td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>120</td>
<td>80</td>
<td>70</td>
<td></td>
</tr>
</tbody>
</table>
16. **Camp sites.** Shown below are the histogram and summary statistics for the number of camp sites at public parks in Vermont.

![Histogram of camp sites](image)

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count</td>
<td>46</td>
</tr>
<tr>
<td>Mean</td>
<td>62.8 sites</td>
</tr>
<tr>
<td>Median</td>
<td>43.5</td>
</tr>
<tr>
<td>StdDev</td>
<td>56.2</td>
</tr>
<tr>
<td>Min</td>
<td>0</td>
</tr>
<tr>
<td>Max</td>
<td>275</td>
</tr>
<tr>
<td>Q1</td>
<td>28</td>
</tr>
<tr>
<td>Q3</td>
<td>78</td>
</tr>
</tbody>
</table>

a) Which statistics would you use to identify the center and spread of this distribution? Why?
b) How many parks would you classify as outliers? Explain.
c) Create a boxplot for these data.
d) Write a few sentences describing the distribution.

17. **Hospital stays.** The U.S. National Center for Health Statistics compiles data on the length of stay by patients in short-term hospitals and publishes its findings in *Vital and Health Statistics*. Data from a sample of 39 male patients and 35 female patients on length of stay (in days) are displayed in the histograms below.

![Histograms of hospital stays](image)

a) What would you suggest be changed about these histograms to make them easier to compare?
b) Describe these distributions by writing a few sentences comparing the duration of hospitalization for men and women.
c) Can you suggest a reason for the peak in women’s length of stay?


![Histograms of ages at death](image)

19. **Women’s basketball.** Here are boxplots of the points scored during the first 10 games of the season for both Scyrine and Alexandra.

![Boxplots of points](image)

a) Describe the overall shapes of these distributions.
b) How do the distributions differ?
c) Look carefully at the bar definitions. Where do these plots violate the rules for statistical graphs?

20. **Gas prices.** Here are boxplots of weekly gas prices at a service station in the Midwestern United States (prices in $ per gallon):

![Boxplots of gas prices](image)

a) Summarize the similarities and differences in their performance so far.
b) The coach can take only one player to the state championship. Which one should she take? Why?
21. **Marriage age.** In 1975, did men and women marry at the same age? Here are boxplots of the age at first marriage for a sample of U.S. citizens then. Write a brief report discussing what these data show.

22. **Fuel economy.** Describe what these boxplots tell you about the relationship between the number of cylinders a car's engine has and the car's fuel economy (mpg):

23. **Fuel economy II.** The Environmental Protection Agency provides fuel economy and pollution information on over 2000 car models. Here is a boxplot of Combined Fuel Economy (using an average of driving conditions) in miles per gallon by vehicle Type (car, van, or SUV). Summarize what you see about the fuel economies of the three vehicle types.

24. **Fuel economy III.** The Environmental Protection Agency provides fuel economy and pollution information on over 2000 car models. Here is a boxplot of Combined Fuel Economy (using an average of driving conditions) in miles per gallon by whether the car has two- or four-wheel drive. Summarize what you see about the fuel economies of these two types of vehicles.

25. **Wines.** The boxplots display case prices (in dollars) of varieties of wines produced by vineyards along three of the Finger Lakes:

a) Which lake region produces the most expensive wine?

b) Which lake region produces the cheapest wine?

c) In which region are the wines generally more expensive?

d) Write a few sentences describing these wine prices.

26. **Ozone.** Ozone levels (in parts per billion, ppb) were recorded at sites in New Jersey monthly between 1926
28. **Unemployment.** In May of 2001, the U.S. Bureau of Labor Statistics (BLS) released a news release that said, 

*In April, 223 metropolitan areas recorded unemployment rates below the U.S. average of 4.2 percent (not seasonally adjusted), while 99 areas registered higher rates.*

Sketch what the distribution of unemployment rates for the 322 metropolitan areas reported on by BLS probably looks like.

29. **Test scores.** Three Statistics classes all took the same test. Histograms of the scores for each class are shown below.

![Histograms of test scores](image)

a) Which class had the highest mean score?
b) Which class had the highest median score? 
c) For which class are the mean and median most different? Which is higher? Why?
d) Which class had the smallest standard deviation? 
e) Which class had the smallest IQR?

30. **Test scores.** Look again at the histograms of test scores for the three Statistics classes in Exercise 29.
a) Overall, which class do you think performed best on the test? Why? 
b) How would you describe the shape of each distribution? 
c) Match each class with the corresponding boxplot below.

---

8 IOC president Juan Antonio Samaranch initiated a special program in the 2000 Olympics. As a way of spreading sports around the world, a handful of athletes were allowed to compete even though they didn't meet qualifying standards.

Eric Moussambani of Equatorial Guinea, who trains in a 20-m pool with no lane markers in his African homeland, became an international celebrity after completing his swim in more than twice the time of the next slowest competitor. He carried the national flag in the opening ceremony, leading an 11-person team that also included a female swimmer. He actually won his heat of 3 swimmers when the other 2 failed to qualify.
1. **Standardized test scores.** For his Statistics class experiment, researcher J. Gilbert decided to study how parents' income affects children's performance on standardized tests like the SAT. He proposed to collect information from a random sample of test takers and examine the relationship between parental income and SAT score.
   a) Is this an experiment? If not, what kind of study is it?
   b) If there is relationship between parental income and SAT score, why can't we conclude that differences in score are caused by differences in parental income?

2. **Heart attacks and height.** Researchers who examined health records of thousands of males found that men who died of myocardial infarction (heart attack) tended to be shorter than men who did not.
   a) Is this an experiment? If not, what kind of study is it?
   b) Is it correct to conclude that shorter men are at higher risk for heart attack? Explain.

3. **MS and vitamin D.** Multiple sclerosis (MS) is an autoimmune disease that strikes more often the farther people live from the equator. Could vitamin D—which most people get from the sun's ultraviolet rays—be a factor? Researchers compared vitamin D levels in blood samples from 150 U.S. military personnel who have developed MS with blood samples of nearly 300 who have not. The samples were taken, on average, five years before the disease was diagnosed. Those with the highest blood vitamin D levels had a 62% lower risk of MS than those with the lowest levels. (The link was only in whites, not in blacks or Hispanics.)
   a) What kind of study was this?
   b) Is that an appropriate choice for investigating this problem? Explain.
   c) Who were the subjects?
   d) What were the variables?

4. **Super Bowl commercials.** When spending large amounts to purchase advertising time, companies want to know what audience they'll reach. In January 2007, a poll asked 1008 American adults whether they planned to watch the upcoming Super Bowl. Men and women were asked separately whether they were looking forward more to the football game or to watching the commercials. Among the men, 16% were planning to watch and were looking forward primarily to the commercials. Among women, 30% were looking forward primarily to the commercials.
   a) Was this a stratified sample or a blocked experiment? Explain.
   b) Was the design of the study appropriate for the advertisers' questions?

5. **Menopause.** Researchers studied the herb black cohosh as a treatment for hot flashes caused by menopause. They randomly assigned 351 women aged 45 to 55 who reported at least two hot flashes a day to one of five groups: (1) black cohosh, (2) a multiherb supplement with black cohosh, (3) the multiherb supplement plus advice to consume more soy foods, (4) estrogen replacement therapy, or (5) receive a placebo. After a year, only the women given estrogen replacement therapy had symptoms different from those of the placebo group. [Annals of Internal Medicine 145:12, 869-897]
   a) What kind of study was this?
   b) Is that an appropriate choice for this problem?
   c) Who were the subjects?
   d) Identify the treatment and response variables.

6. **Honesty.** Coffee stations in offices often just ask users to leave money in a tray to pay for their coffee, but man people cheat. Researchers at Newcastle University replaced the picture of flowers on the wall behind the coffee station with a picture of staring eyes. They found that the average contribution increased significantly above the well-established standard when people felt they were being watched, even though the eyes were patentely not re.
   (NY Times 12/10/06)
   a) Was this a survey, an observational study, or an experiment? How can we tell?
   b) Identify the variables.
   c) What does "increased significantly" mean in a statistical sense?

7-26. **Read each brief report of statistical research, and identify**
   a) whether it was an observational study or an experiment
   b) if it was an observational study, identify (if possible)
      c) whether it was retrospective or prospective
      d) the subjects studied and how they were selected
      e) the parameter of interest
      f) the nature and scope of the conclusion the study can reach.

---

7. **Over a 4-month period, among 30 people with bipolar disorder, patients who were given a high dose (10 g/day) of omega-3 fats from fish oil improved more than those given a placebo. (Archives of General Psychiatry 56 [1999]: 407)**
22.2 days ($SD = 8.3$ days). Do you think this difference is statistically significant? Explain.

38. **Diet and blood pressure.** Exercise 12 reports on an experiment that showed that subjects fed the DASH diet were able to lower their blood pressure by an average of 6.7 points compared to a group fed a "control diet." All meals were prepared by dieticians.
   a) Why were the subjects randomly assigned to the diets instead of letting people pick what they wanted to eat?
   b) Why were the meals prepared by dieticians?
   c) Why did the researchers need the control group? If the DASH diet group's blood pressure was lower at the end of the experiment than at the beginning, wouldn't that prove the effectiveness of that diet?
   d) What additional information would you want to know in order to decide whether an average reduction in blood pressure of 6.7 points was statistically significant?

39. **Mozart.** Will listening to a Mozart piano sonata make you smarter? In a 1995 study published in the journal *Psychological Science*, Rauscher, Shaw, and Ky reported that when students were given a spatial reasoning section of a standard IQ test, those who listened to Mozart for 10 minutes improved their scores more than those who simply sat quietly.
   a) These researchers said the differences were statistically significant. Explain what that means in this context.
   b) Steele, Bass, and Crook tried to replicate the original study. In their study, also published in *Psychological Science* (1999), the subjects were 125 college students who participated in the experiment for course credit. Subjects first took the test. Then they were assigned to one of three groups: listening to a Mozart piano sonata, listening to music by Philip Glass, and sitting for 10 minutes in silence. Three days after the treatments, they were retested. Draw a diagram displaying the design of this experiment.
   c) These boxplots show the differences in score before and after treatment for the three groups. Did the Mozart group show improvement?

40. **More Mozart.** The *NY Times* (Sept. 9, 1991) reported that "A sake brewer [near] Tokyo is marketing a Mozart brand of sake, asserting that Mozart is regularly played at the brewery. It seems that the malt grows well especially with Mozart music in the background," the company director [said]." Suppose, just for the sake (as it were) of discussion, you wished to design an experiment to test whether this is true. Assume you have the full cooperation of the sake brewery. Specify how you would design the experiment. Indicate factors and response and how they would be measured, controlled, or randomized.

41. **Frumpies.** The makers of Frumpies, "the breakfast of rug rats," want to improve their marketing, so they consult you:
   a) They first want to know what fraction of children, ages 10 to 13, like their celery-flavored cereal. What kind of study should they perform?
   b) They are thinking of introducing a new flavor, maple-marshmallow Frumpies, and want to know whether children will prefer the new flavor to the old one. Design a completely randomized experiment to investigate this question.
   c) They suspect that children who regularly watch the Saturday morning cartoon show starring Frump, the flying teenage warrior rabbit who eats Frumpies in every episode, respond differently to the new flavor. How would you take that into account in your design?

42. **Full moon.** It's a common belief that people behave strangely when there's a full moon and that as a result police and emergency rooms are busier than usual. Design a way you could find out whether there is any merit to this belief. Will you use an observational study or an experiment? Why?

43. **Wine.** A 2001 Danish study published in the *Archives of Internal Medicine* casts significant doubt on suggestions that adults who drink wine have higher levels of "good" cholesterol and fewer heart attacks. These researchers followed a group of individuals born at a Copenhagen hospital between 1959 and 1961 for 40 years. Their study found that in this group the adults who drank wine were richer and better educated than those who did not.
   a) What kind of study was this?
   b) It is generally true that people with high levels of education and high socioeconomic status are healthier than others. How does this call into question the supposed health benefits of wine?
   c) Can studies such as these prove causation (that wine helps prevent heart attacks, that drinking wine makes one richer, that being rich helps prevent heart attacks, etc.)? Explain.

44. **Swimming.** Recently, a group of adults who swim regularly for exercise were evaluated for depression. It turned out that these swimmers were less likely to be depressed than the general population. The researchers said the difference was statistically significant.
   a) What does "statistically significant" mean in this context?
b) Is this an experiment or an observational study? Explain.

c) News reports claimed this study proved that swimming can prevent depression. Explain why this conclusion is not justified by the study. Include an example of a possible confounding variable.

d) But perhaps it is true. We wonder if exercise can ward off depression, and whether anaerobic exercise (like weight training) is as effective as aerobic exercise (like swimming). We find 120 volunteers not currently engaged in a regular program of exercise. Design an appropriate experiment.

45. Dowsing. A water dowsing claims to be able to sense the presence of water using a forked stick. Suppose we wish to set up an experiment to test his ability. We get 20 identical containers, fill some with water, and ask the dowsing to tell which ones are full and which empty.

a) How will we randomize this procedure?

b) The dowsing correctly identifies the contents of 12 out of 20 containers. Do you think this level of success is statistically significant? Explain.

c) How many correct identifications (out of 20) would the dowsing have to make to convince you that the forked stick trick works? Explain.

46. Healing. A medical researcher suspects that giving postsurgical patients large doses of vitamin E will speed their recovery times by helping their incisions heal more quickly. Design an experiment to test this conjecture. Be sure to identify the factors, levels, treatments, response variable, and the role of randomization.

47. Reading. Some schools teach reading using phonics (the sounds made by letters) and others using whole language (word recognition). Suppose a school district wants to know which method works better. Suggest a design for an appropriate experiment.

48. Gas mileage. Do cars get better gas mileage with premium instead of regular unleaded gasoline? While it might be possible to test some engines in a laboratory setting, we'd rather use real cars and real drivers in real day-to-day driving, so we get 20 volunteers. Design the experiment.

49. Weekend deaths. A study published in the New England Journal of Medicine (Aug. 2001) suggests that it's dangerous to enter a hospital on a weekend. During a 10-year period, researchers tracked over 4 million emergency admissions to hospitals in Ontario, Canada. Their findings revealed that patients admitted on weekends had a much higher risk of death than those who went to the emergency room on weekdays.

a) The researchers said the difference in death rates was "statistically significant." Explain in this context what that means.

b) What kind of study was this? Explain.

c) If you think you're quite ill on a Saturday, should you wait until Monday to seek medical help? Explain.

d) Suggest some possible explanations for this troubling finding.

50. Shingles. A research doctor has discovered a new ointment that she believes will be more effective than the current medication in the treatment of shingles (painful skin rash). Eight patients have volunteered to participate in the initial trials of this ointment. You the statistician hired as a consultant to help design completely randomized experiment.

a) Describe how you will conduct this experiment.

b) Suppose the eight patients' last names start with the letters A to H. Using the random numbers listed, show which patients you will assign to each group. Explain your randomization procedure clearly.

   41098 18329 78458 31685 55259

c) Can you make this experiment double-blind? If so, explain how.

d) The initial experiment revealed that males and females may respond differently to the ointment. Further testing of the drug's effectiveness is now planned, and many patients have volunteered. If you were the statistician, what changes in your first design, if any, would you make for the second stage of testing?

51. Beetles. Hoping to learn how to control crop damage by a certain species of beetle, a researcher plans to analyze data from two different pesticides on small plots of corn. After application of the chemicals, he'll check the number of beetle larvae found on each plant. The researcher wants to know whether either pesticide was more effective and whether there is a significant difference in effectiveness between them. Design an appropriate experiment.

52. SAT Prep. Can special study courses actually help SAT scores? One organization says that the 30 students they tutored achieved an average gain of 60 points when they retook the test.

a) Explain why this does not necessarily prove that special course caused the scores to go up.

b) Propose a design for an experiment that could test effectiveness of the tutoring course.

c) Suppose you suspect that the tutoring course might be more helpful for students whose initial scores were particularly low. How would this affect your proposed design?

53. Safety switch. An industrial machine requires an emergency shutoff switch that must be designed so it can be easily operated with either hand. Design an experiment to find out whether workers will be able to operate the machine as quickly with their left hands as with their right hands. Be sure to explain the role of randomization in your design.

54. Washing clothes. A consumer group wants to test effectiveness of a new "organic" laundry detergent. Make recommendations to customers about how to use the product. They intentionally get grass stains on white T-shirts in order to see how well the detergent can clean them. They want to try the detergent in cold and hot water on both the "regular" and "delicate" wash cycles. Design an appropriate experiment, including the number of factors, levels, and treatments. Explain the role of randomization in your experiment.