2.14 (Schools)
- **Who:** students – each student is an individual case
- **What:** age, race/ethnicity, number of absences, grade level, reading score, math score, disabilities/special needs
- **When:** current
- **Where:** not specified
- **Why:** state requirement to keep this information
- **How:** information is collected and stored as part of school records

Variables:
- Race/ethnicity – categorical
- Number of absences – quantitative, integer count
- Current grade level – could be considered categorical or quantitative/integer
- Standardized reading score – quantitative, units not specified
- Standardized math score – quantitative, units not specified
- Disability/special needs – categorical

2.24 (Walking in Circles)
- **Who:** 32 volunteers – each volunteer is a case
- **What:** sex, height, handedness, distance walked, sideline crossed
- **When/Where:** not specified
- **Why:** to see if people naturally walk in circles
- **How:** collected during blindfolded a test on a football field

Variables:
- Sex – categorical
- Height – quantitative, units not specified (presumably inches)
- Handedness – categorical (right/left)
- Distance – quantitative (yards)
- Sideline crossed – categorical

3.5 (Movie Genres)

a) Yes, display is appropriate – the categories are mutually exclusive. **However, pie charts are not ideal because human perception typically has a more difficult time judging ratios in pie charts.**

b) Thriller/Horror is the least common genre

3.7 (Genres again)

a) Comedy is the most common genre

b) It is easier to tell the difference in height of bars; categories so close in size would have been difficult to confidently judge in a pie chart
3.9 (Magnet Schools)
1755 students applied for admission; 53% were accepted, 17% were waitlisted, and 30% turned away

3.10 (Magnet Schools again)
Of the 1755 students who applied for admission, 29.5% were Hispanic, 16.6% Asian, and 53.9% White

3.15 (Global Warming)
- No title on the chart
- Percentages add up to 92% (should be 100)
- 3-D (ish) display distorts size of the pie pieces

3.25 (Magnet Schools revisited)
a) 16.6% of applicants were Asian
b) 11.8% of students accepted were Asian
c) 37.7% of Asian applicants were accepted
d) 53% of total applicants were accepted; 56.1% of non-Asian applicants were accepted

3.31 (Blood Pressure)
a) Low 20%, Normal 48.9%, High 31%
b) Under 30 30-49 Over 50
<table>
<thead>
<tr>
<th></th>
<th>Under 30</th>
<th>30-49</th>
<th>Over 50</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>27.6%</td>
<td>20.7%</td>
<td>15.7%</td>
</tr>
<tr>
<td>Normal</td>
<td>49.0%</td>
<td>50.8%</td>
<td>47.2%</td>
</tr>
<tr>
<td>High</td>
<td>23.5%</td>
<td>28.5%</td>
<td>37.1%</td>
</tr>
</tbody>
</table>
c) As age increases, the percentage of adults with high blood pressure increases
d) This doesn’t prove that blood pressure rises with age, although it implies that it might. It is possible, for example, that the over-50 group had higher blood pressure when they were younger, because of cohort differences (e.g., diet when they were raised)
3.39 (Graduate Admissions)
   a) 42.6% of applicants were admitted
   b) A higher percentage of males than females were admitted: 47.2% of males, 30.9% of females
   c) Program 1: 47.2% of males, 82.4% of females
      Program 2: 62.9% of males, 68% of females
      Program 3: 33.7% of males, 35.2% of females
      Program 4: 5.9% of males, 7% of females
   d) The comparisons in (c) show that males have a lower admittance rate in every program, even though the overall rate shows males with a higher rate of admittance. This is an example of Simpson’s paradox. It means that, on average, males are more-represented in the higher-admittance programs.

Cocaine Addiction problem (not in textbook)

Note: the Total for the “Relapse” column should be 48, not 38 as listed in the assignment

<table>
<thead>
<tr>
<th>Treatment</th>
<th>N</th>
<th>Relapse</th>
<th>No relapse</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Desipramine</td>
<td>24</td>
<td>41.7%</td>
<td>58.3%</td>
<td>100.0%</td>
</tr>
<tr>
<td>Lithium</td>
<td>24</td>
<td>75.0%</td>
<td>25.0%</td>
<td>100.0%</td>
</tr>
<tr>
<td>Placebo</td>
<td>24</td>
<td>83.3%</td>
<td>16.7%</td>
<td>100.0%</td>
</tr>
<tr>
<td>Total</td>
<td>72</td>
<td>66.7%</td>
<td>33.3%</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

Based on this trial, Desipramine was by far the most successful in preventing relapses of Cocaine usage. 58.3% of the addicts treated with Desipramine had successful treatment (i.e., did not suffer a relapse), whereas the success rates were lower with Lithium (25%) and Placebo (17%). However, it is worth noting that the sample size is quite small – 24 people were tested for each group.

Swiss problem (not in textbook)

a) Histogram of Catholic variable:
This distribution tells you that the majority of districts in Switzerland in 1888 were either highly Catholic or highly non-Catholic (i.e., highly Protestant). The religious groups were likely divided in different regions of the country. Also, you can see that Catholic makes up less than 50% of the country’s total population, as the low-Catholic-percentage side of the graph is slightly better-represented than the high-Catholic side of the graph.

b) Histogram of *Fertility* variable:

![Histogram of Fertility variable](image)

This distribution has a fairly normal shape (just slightly left-skewed), so the mean and standard deviation would be useful to summarily describe the data. Based on the histogram, the mean appears to be between the 60-70 group and the 70-80 group, so we could say “about 70.” *(In fact, the mean is 70.14 and the median is 70.40)*

4.8 (Singers)

a) The distribution appears bimodal, with one mode around 65 inches, the other around 71 inches. Potentially convenient if the chorus is to be arranged in a front and a back row.

b) Best guess is a mix of males and females in the chorus

4.11 (Heart Attack stays)

a) The histogram is strongly right-skewed, so the mean will be larger than the median

b) The distribution is bimodal and right-skewed. The center mode is around 8 days of hospital stay, while the other mode is at 1 day (the shortest possible stay). The 1-day mode is likely so common because it includes patients who died, and/or patients who required little to no treatment. Most patients stay between 1 and 15 days, although there are some outliers who stay many more days.

c) The median and Inter-Quartile Range are useful because the distribution is so skewed. The mean is useful for comparing to the median (as another indicator of skew).
4.16 (Tornados 2008)
   a) mean: 69
   b) median: 55
      Q1: 40
      Q2: 90.75
   c) range: 95
      Inter-Quartile Range: 50.75

4.22 (Neck Size)
The data are normally distributed and unimodal, so the mean and standard deviation (and possible the range, since the distribution appears to be bounded without extreme outliers) are useful summary stats.

4.32 (Singers)
   a) median: 66 inches
      Inter-Quartile Range: 5 inches
   b) mean: 67.1 inches
      standard deviation: 3.8 inches
   c) histogram of singers heights:

   ![Histogram of Singers Heights](image)
   d) The distribution is centered near 67 inches, with median of 66 inches and mean 67.1 inches. The middle 50% of heights are between 65 and 70 inches. The distribution appears to be unimodal, although skewed slightly to the right.
4.35 (States)
   a) Since the data are strongly skewed, the median and IQR are the best statistics to report
   b) The mean will be larger than the median, because the data is skewed toward the large numbers
   c) median: 4 million
      IQR: 4.5 million (Q3: 6 million, Q1: 1.5 million)
   d) The distribution of populations of states (and Washington, D.C.) is unimodal and right-skewed.
      Median is 4 million. One state (California) is an outlier, with a population of 34 million.

4.49 (Math Scores 2005)
   a) median: 239
      IQR: 9
      mean: 237.6
      standard deviation: 5.7
   b) Because the distribution is skewed to the left, probably better to report median and IQR
   c) Skewed to the left, and maybe bimodal. The center is around 239, and the middle 50% of states
      scored between 233 and 242. Alabama, Mississippi, and New Mexico scores were low outliers.

5.9 (Cereals)
   a) The range of sugar of sugar contents in these cereals is about 59 (you can tell exact increments
      from the mix/max indicators on the box plots)
   b) The distribution is bimodal: some cereals cluster around high sugar content, some around low.
   c) see (b)
   d) Yes, children’s cereals are higher in sugar than adult-targeted cereals. In fact, in this sample of
      49 cereals, the distributions do not even overlap.
   e) Although the ranges appear comparable between children’s and adults’ cereals, the IQR is larger
      for adult cereals, indicating more variability in the sugar content of the middle 50% of adult
      cereals.

5.28 (Framingham)
Boxplot with following (approximate) characteristics based on the ogive graph (looking at 25th, 50th, and
75th percentiles)
   • minimum: ~90
   • Q1: ~205
   • median: ~230
   • Q3: ~265
   • Maximum: ~425
The distribution is essentially symmetric, with median just above 225, IQR 60-70 points. The outliers are
slightly right-skewed, just because nature allows more variability at the high end of blood pressure
(whereas too low and death comes more quickly)
5.37 (Assets)
   a) Most of the data are found in the far left of his histogram; the distribution is very right-skewed.
   b) Transforming the data (e.g., by taking the log or possibly square root) might make the
distribution more symmetric and visual-friendly.

5.39 (Assets again)
   a) The logarithm makes the histogram more symmetric. It is easy to see that the center is around
   3.5 log-assets (which shows the difficulty of using log-transformation, as it is difficult to
conceptualize a log-dollar compared to just thinking of a dollar)
   b) The value of 50 square-root assets indicates an actual dollar value around $2500 million ($2.5b)
   c) The value of 3 log-assets indicates an actual dollar value around $1000 million ($1 billion)

Chile problem (not in textbook)
   Parallel box plots of Status Quo by Education categories