Line charts vs. bar charts

1.19 A psychologist measures the length of time it takes a rat to successfully navigate a maze on each of five days.

The times are:

<table>
<thead>
<tr>
<th>Day</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>45</td>
</tr>
<tr>
<td>2</td>
<td>43</td>
</tr>
<tr>
<td>3</td>
<td>46</td>
</tr>
<tr>
<td>4</td>
<td>32</td>
</tr>
<tr>
<td>5</td>
<td>25</td>
</tr>
</tbody>
</table>

(a) Draw a line chart to represent the data.

Note that the corresponding bar chart consists of bars whose heights are the nodes of the line chart.

Fractons, decimals, percents

$\frac{3}{4}$, .75, 75% are three ways of representing the same number. .75 = 75% in the same sense that 12 inches = 1 foot. The units for percents are one hundreths. Thus 75% means $75 \times (0.01)$ which is .75. Instead of being something special, percents are just another unit and another way of writing numbers. Although percentages usually are used for numbers in $[0, 1]$, there is no formal reason for this. For example, it is possible for company to experience 120% growth. If we wanted to, we could write 13.56 as 1356%.

Rounding

Unless otherwise specified, assume that data you are given is accurate to two decimal places. In this case you answers should be rounded to two significant decimal places.

- What is the average of 1.25, 5, 3.40?
  
  average = \frac{(1.25 + 5 + 3.40)}{3} = 9.65/3 = 3.2166666...

  Rounding to two places gives

  Answer: 3.22.

  Why shouldn’t we use a 4-place answer 3.2167 or even 3.216666...?

  The reason is that giving an answer with four or more decimals places gives a dishonest impression of accuracy. If the data is only accurate to 2 places, then one should be correspondingly modest in calculating things like averages.

  However only the final answer should be rounded. Do not round intermediate numbers used to calculate the answer. Doing so will lead to rounding errors.

  It is best to learn to be able to do all calculations on your calculator without having to write down and reenter intermediate numbers. This means learning to use the memory registers of your calculator and learning how to use parentheses.

- The average $A$ of $x, y, z$ is $A = \frac{(x + y + z)}{3}$.

  The standard deviation is $S = \sqrt{\frac{(x-A)^2 + (y-A)^2 + (z-A)^2)}{3}}$

To calculate $S$, don't write down the average $A$ found with the first formula and then reenter it in the formula for $S$. Rather, store $A$ in a memory register and then recall $A$ each time it is needed. This way there will be no rounding errors or errors due to mistyping $A$.

Calculator

For this course you will need a calculator which can do “two variable statistics”. You will need it for the very first exam. You can buy such calculators for a little around $20 at the bookstore. But be sure that it lists “two variable statistics” in its specifications. Calculators which only do one variable statistics won’t do.

Stem-and-leaf plots vs. histograms

For the data 1.2, 1.3, 0.8, 4.0, 3.3, 3.5, 2.8, 0.8, 1.3, with .1 as the leaf unit, the stem-and-leaf plot is:

- 0 | 8, 8
- 1 | 2, 3, 3
- 2 | 8
- 3 | 3, 5
- 4 | 0

The frequency histogram for the same data with interval of length 1 is

Note that rotating the lines of the stem-and-leaf plot 90° counterclockwise gives basically the bars of the histogram.

Stem-and-leaf plots are a convenient way to organize data for an initial analysis. However, when the data is published, histograms are preferred.

Stem-and-leaf plots can also handle alphabetical data. Try making a stem-and-leaf plot for the data set:

{ hand, and, band, hat, hate, hard, bend, bat, barb }

- a | nd
- b | at
- ba | nd, rb
- be | nd
- h | at
- ha | nd, rd, te

leaf = last two letters