DESCRIBING DATA WITH GRAPHS

Data can be described clearly and concisely with the aid of a well-constructed frequency distribution and data can often be described even more vividly by converting frequency distributions into graphs.

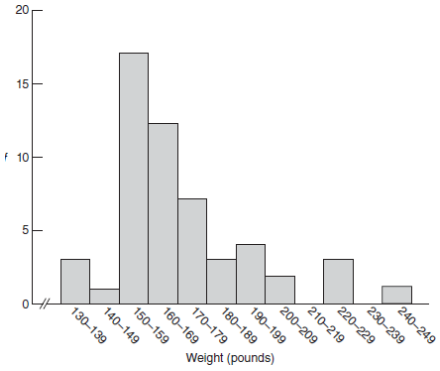
Graphs for Quantitative data

Histograms

A bar-type graph for quantitative data. The common boundaries between adjacent bars emphasize the continuity of the data, as with continuous variables. A histogram is a display of statistical information that uses rectangles to show the frequency of data items in successive numerical intervals of equal size.

Important features of histograms

- Equal units along the horizontal axis (the X axis, or abscissa) reflect the various class intervals of the frequency distribution.
- Equal units along the vertical axis (the Y axis, or ordinate) reflect increases in frequency. (The units along the vertical axis do not have to be the same width as those along the horizontal axis.)
- The intersection of the two axes defines the origin at which both numerical scales equal 0.
- Numerical scales always increase from left to right along the horizontal axis and from bottom to top along the vertical axis
- The body of the histogram consists of a series of bars whose heights reflect the frequencies for the various classes.
- The adjacent bars in histograms have common boundaries that emphasize the continuity of quantitative data for continuous variables.
- The introduction of gaps between adjacent bars would suggest an artificial disruption in the data more appropriate for discrete quantitative variables or for qualitative variables.



Frequency Polygon

An important variation on a histogram is the frequency polygon, or line graph. Frequency polygons may be constructed directly from frequency distributions.

Step-by-step transformation of a histogram into a frequency polygon

- This panel shows the histogram for the weight distribution.
- Place dots at the midpoints of each bar top or, in the absence of bar tops, at midpoints for classes on the horizontal axis, and connect them with straight lines.
- Anchor the frequency polygon to the horizontal axis. First, extend the upper tail to the midpoint of the first unoccupied class on the upper flank of the histogram. Then extend the lower tail to the midpoint of the first unoccupied class on the lower flank of the histogram. Now all of the area under the frequency polygon is enclosed completely.
- Finally, erase all of the histogram bars, leaving only the frequency polygon.

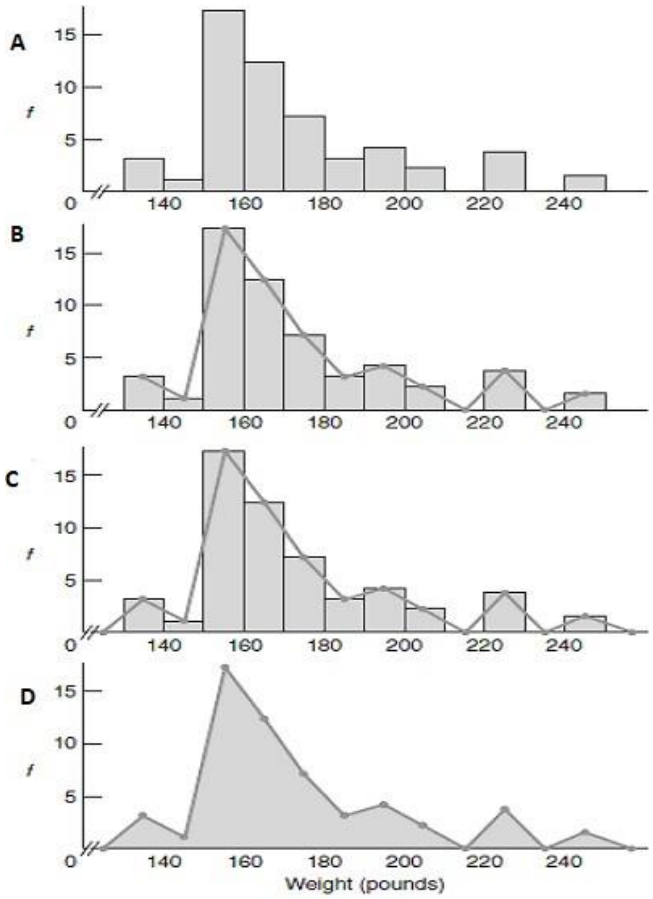


FIGURE
Transition from histogram to frequency polygon.

Stem and Leaf Displays

Another technique for summarizing quantitative data is a stem and leaf display. Stem and leaf displays are ideal for summarizing distributions, such as that for weight data, without destroying the identities of individual observations.

Constructing Stem and Leaf Display

The leftmost panel of table re-creates the weights. To construct the stem and leaf display for the table given below, first note that, when counting by tens, the weights range from the 130s to the 240s. Arrange a column of numbers, the stems, beginning with 13 (representing the 130s) and ending with 24 (representing the 240s). Draw a vertical line to separate the stems, which represent multiples of 10, from the space to be occupied by the leaves, which represent multiples of 1.

For example

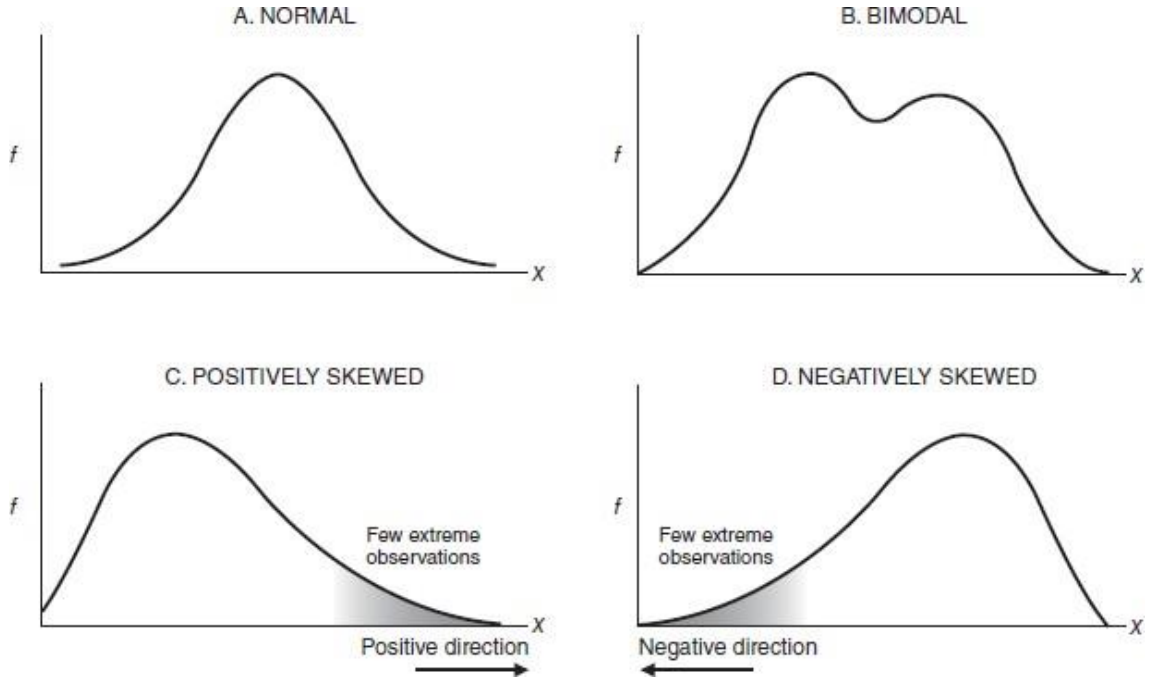
Enter each raw score into the stem and leaf display. As suggested by the shaded coding in Table 2.9, the first raw score of 160 reappears as a leaf of 0 on a stem of 16. The next raw score of 193 reappears as a leaf of 3 on a stem of 19, and the third raw score of 226 reappears as a leaf of 6 on a stem of 22, and so on, until each raw score reappears as a leaf on its appropriate stem.

Table 2.9
CONSTRUCTING STEM AND LEAF DISPLAY FROM WEIGHTS OF MALE STATISTICS STUDENTS

RAW SCORES					STEM AND LEAF DISPLAY
160	165	135	175		
193	168	245	165	13	3 5 5
226	169	170	185	14	5
152	160	156	154	15	2 7 1 7 8 0 2 0 2 6 9 8 2 6 4 7 6
180	170	160	179	16	0 3 5 8 9 0 0 0 6 5 5 5
205	150	225	165	17	2 0 0 0 2 5 9
163	152	190	206	18	0 0 5
157	160	159	165	19	3 0 0 0
151	190	172	157	20	5 6
157	150	190	156	21	
220	133	166	135	22	6 0 5
145	180	158		23	
158	152	152		24	5
172	170	156			

TYPICAL SHAPES

Whether expressed as a histogram, a frequency polygon, or a stem and leaf display, an important characteristic of a frequency distribution is its shape. Below figure shows some of the more typical shapes for smoothed frequency polygons (which ignore the inevitable irregularities of real data).



A GRAPH FOR QUALITATIVE (NOMINAL) DATA

- As with histograms, equal segments along the horizontal axis are allocated to the different words or classes that appear in the frequency distribution for qualitative data. Likewise, equal segments along the vertical axis reflect increases in frequency. The body of the bar graph consists of a series of bars whose heights reflect the frequencies for the various words or classes.
- A person's answer to the question "Do you have a Facebook profile?" is either Yes or No, not some impossible intermediate value, such as 40 percent Yes and 60 percent No.
- Gaps are placed between adjacent bars of bar graphs to emphasize the discontinuous nature of qualitative data.

MISLEADING GRAPHS

Graphs can be constructed in an unscrupulous manner to support a particular point of view. Popular sayings say, including "Numbers don't lie, but statisticians do" and "There are three kinds of lies— lies, damned lies, and statistics."