

Chapter 13.1

Displaying Distributions with Graphs

Homework: 3, 5, 7, 9, 11-23 all, 29-31 all

Sep 30-8:52 AM

Statistical Terminology

- **Qualitative (Categorical) variable** – places an individual into one of several groups or categories. {Gender, Race, Blood Type}
- **Quantitative variable** – takes numerical values for which arithmetic operations such as adding and averaging make sense. {Height, Income, Time, etc.}
 - Discrete Variables** – There is a gap between possible values.
 - Continuous Variables** – Variables that can take on values in an interval.

Sep 30-8:52 AM

More Terminology

- **Distribution** – tells us what values the variable takes and how often it takes these values
- **Frequency Distribution** – a list of the distinct quantitative data values along with a frequency count or the relative frequency percentage.

Sep 30-8:52 AM

Histogram

- A histogram is the most common graph of the distribution of a quantitative variable.
- To create a histogram:
 1. Divide the range of the data into classes of equal size.
 - * Be sure that each individual fall into exactly one class.
 - * Typically use 5 to 12 classes.
 2. Count the number of individuals in each class.
 3. Draw the histogram.
 - * Each bar represents a class.
 - * The base of the bar covers the class.
 - * The bar height is the class count (or percentage).

** Note: A histogram is used for continuous quantitative variables, so there is no gap between the bars.

** If you have categorical data, use a bar graph, which has gaps between the bars, instead of a histogram.

Sep 30-8:52 AM

Entering Freshman at UWL must either take the ACT or SAT. A random sample of 21 student scores on the ACT are listed below. Construct a histogram and describe its shape, center, and spread.

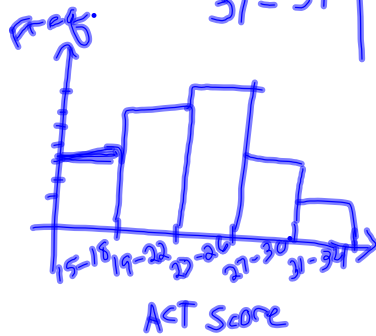
ACT Scores

31	26	23
32	24	25
25	22	30
17	18	19
15	21	20
20	28	24
22	29	23

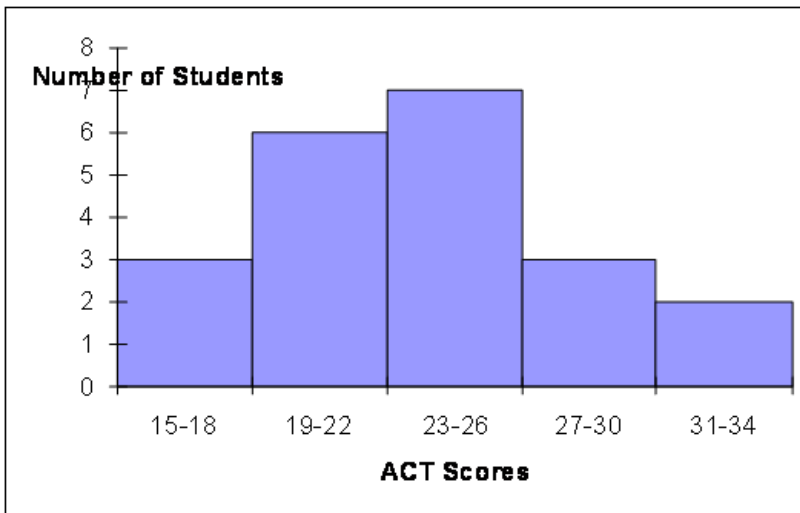
Range: $32 - 15 = 17$

5 classes, ea/ 4 units.

Class	Freq	Rel. Freq
15-18	3	3/21
19-22	6	
23-26	7	
27-30	3	
31-34	2	



Sep 30-8:52 AM



Sep 30-8:52 AM

Occurrence of Diabetes

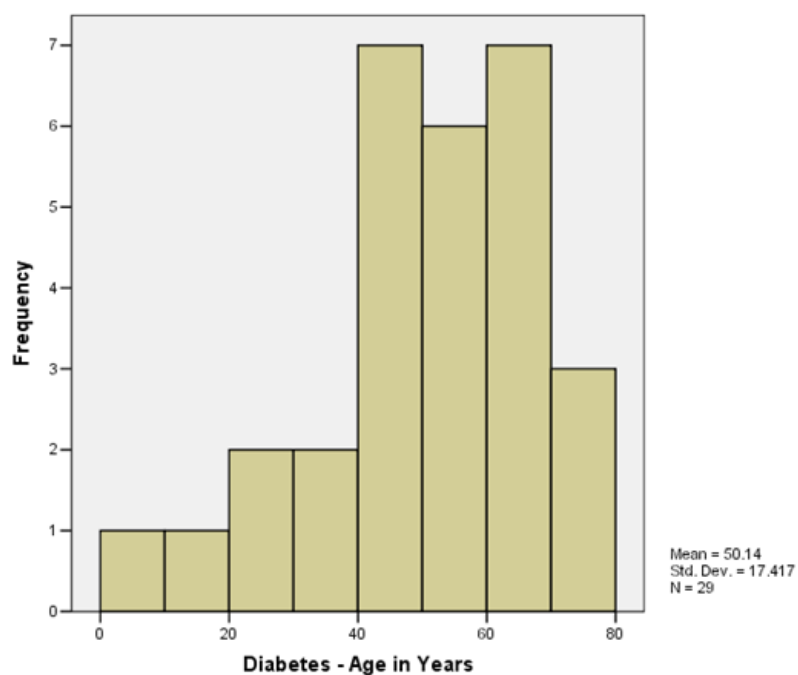
Age of Diagnosis	Frequency
0-9	1
10-19	1
20-29	2
30-39	2
40-49	7
50-59	6
60-69	7
70-79	2

Raw Data

8 12 22 27 34 38 40 41 45
 45 47 48 48 51 51 54 55 55
 58 63 63 65 65 66 68 68 71
 72 74

Sep 30-8:52 AM

Occurrence of Diabetes

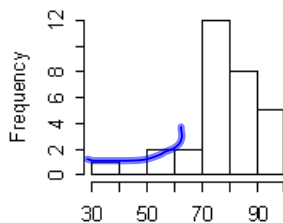


Sep 30-8:52 AM

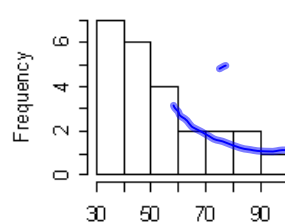
Defining a Common Language about Shape

- **Symmetric:** if draw line through center, picture on one side would be mirror image of picture on other side.
- *Example:* bell-shaped data set.
- **Skewed to the Right:** *higher* values more spread out than lower values
- **Skewed to the Left:** *lower* values more spread out and higher ones tend to be clumped

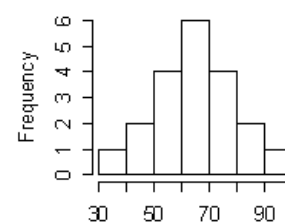
Look for the "tail".



Skewed Left



Skewed Right



Symmetric

Sep 30-8:52 AM

Stemplot

- Only use with a **small data set**.
- To make a stemplot:
 1. separate the numbers into stems and leaves. The stems can have as many numbers as needed, but the leaves can only have one number (the final number).
 2. Write the stems in a vertical column on the left (starting with the lowest value) and draw a line to separate them from the leaves.
 3. Write the leaves ascending to the right of the appropriate stem.

Sep 30-8:52 AM

Creating a Stemplot

Step 1: Create the Stems

Divide range of data into equal units to be used on **stem**. Have 6 – 15 stem values, representing equally spaced intervals.

Step 1: Creating the stem

3| 2
4|
5| 5
6| 012448
7| 35568899
8| 0023458
9| 02358

Ordered Listing of 28 Exam Scores

32, 55, 60, 61, 62, 64, 64, 68, 73, 75, 75, 76, 78, 78, 79, 79, 80, 80, 82, 83, 84, 85, 88, 90, 92, 93, 95, 98

Sep 30-8:52 AM

Creating a Stemplot

Attach a **leaf** to represent each data point. Next digit in number used as leaf; drop remaining digits.

Step 2: Attaching leaves

3|2
4|
5|
6|
7|
8|
9|

Ordered Listing of 28 (Fictional) Exam Scores

32, 55, 60, 61, 62, 64, 64, 68, 73, 75, 75, 76, 78, 78, 79, 79, 80, 80, 82, 83, 84, 85, 88, 90, 92, 93, 95, 98

Sep 30-8:52 AM

ACT Scores

31	26	23
32	24	25
25	22	30
17	18	19
15	21	20
20	28	24
22	29	23

Splitting Stems:

```

(10-14) | 1
(15-19) | 1 5789
etc.    | 2 001223344
        | 2 55689
        | 3 012
        | 3
    
```

Sep 30-8:52 AM

Obtaining Info from the Stemplot

Determine shape, identify outliers, locate center.

Pulse

Rates:

```

5|4
5|789
6|023344
6|55567789
7|00124
7|58
    
```

Exam Scores

```

3|2
4|
5|5
6|024418
7|56598398
8|5430820
9|53208
    
```

Incomes:

```

4|66789
5|11344
5|56666688899999
6|011112334
6|556666789
7|01223
7|
8|0022
    
```

Key

4|6

means

\$46,000

Sep 30-8:52 AM

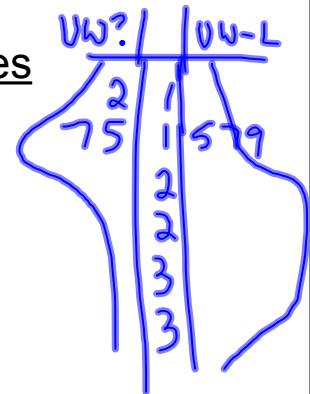
Is there a difference in ACT scores?

UW? ACT Scores

23	22	23
12	23	22
25	22	30
17	28	29
15	21	20
20	24	24
22	26	23

UWL ACT Scores

31	26	23
32	24	25
25	22	30
17	18	19
15	21	20
20	28	24
22	29	23



Sep 30-8:52 AM

Example: larger numbers

- $15765 \approx 15800 \Rightarrow$ stem: 15 leaf: 8
- $6423 \approx 6400 \Rightarrow$ stem: 6 leaf: 4
- $19,333 \approx 19,300 \Rightarrow$ stem: 19 leaf:
3
- $842 \approx 800 \Rightarrow$ stem: 0 leaf:
8
- You can also use decimals as the leaf.
Example: $8.6 \rightarrow$ stem: 8 leaf: 6
- Just make sure you give a key!

Sep 30-8:52 AM

Histogram and Stemplot

- A stemplot looks like a histogram turned on end.
- You can choose the classes in a histogram. The classes of a stemplot are given to you.
- The chief advantage of a stemplot is that it displays the actual values of the observations. Stemplot are also faster to draw than histograms.
- Stemplots do not work well with large data sets, because the stems then have too many leaves.

Sep 30-8:52 AM

Sep 30-9:00 AM