# A Few More Things for the Test and then Measures of Central Tendency and Spread.

## Stem and Leaf Plots

- Displays the distribution of the data while maintaining the actual data values.
- Each data value is split into a stem and a leaf.
  - Example: The height 71” can be split as stem = 7 and leaf = 1

**PROCEDURE**

**How to make a stem-and-leaf display**

1. Divide the digits of each data value into two parts. The leftmost part is called the *stem* while the rightmost part is called the *leaf*.
2. Align all the stems in a vertical column from smallest to largest. Draw a vertical line to the right of all the stems.
3. Place all the leaves with the same stem on the same row as the stem, and arrange the leaves in increasing order.
4. Use a label to indicate the magnitude of the numbers in the display. We include the decimal position in the label rather than with the stems or leaves.

## Stem and Leaf Plot Example

- Below are the points scored by the winning basketball team for the past 35 championship games.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 132 | 118 | 124 | 109 | 104 | 101 | 125 | 83 | 99 |
| 131 | 98 | 125 | 97 | 106 | 112 | 92 | 120 | 103 |
| 111 | 117 | 135 | 143 | 112 | 112 | 116 | 106 | 117 |
| 119 | 110 | 105 | 128 | 112 | 126 | 105 | 102 | |

## Stem and Leaf Plot Example

- We can use the first two digits of each score as the stem.
- We can use the third digit of each score as the leaf.
  - Score = 125 Stem = 12 Leaf = 5
  - Score = 83 Stem = 8 (or 08) Leaf = 3

Winning Scores of Conference Basketball Championship Games

08 | 3 represents 083 or 83 points

| | |
|---|---|
| 08 | 3 |
| 09 | 2 7 8 9 |
| 10 | 1 2 3 4 5 5 6 6 9 |
| 11 | 0 1 2 2 2 2 6 7 7 8 9 |
| 12 | 0 4 5 5 6 8 |
| 13 | 1 2 5 |
| 14 | 3 |

## Notation

The population size is denoted *N*.

The sample size is denoted *n.*

## Mean

$$\text{Sample mean} = \bar{x} = \frac{\Sigma x}{n}$$

- Read "*x-bar*"

$$\text{Population mean} = \mu = \frac{\Sigma x}{N}$$

- Read "*mu*"

## Mean Example

- Seven light bulbs are measured to see how long they last (in hours).
  - {1, 180, 192, 194, 196, 221, 237}

$$\bar{x} = \frac{\sum x}{n} = \frac{1221}{7} = 174.43$$

You should have done this with the sub!

**INVESTIGATION 2** **EFFECTS OF OUTLIERS**

We have seen that an **outlier** or **extreme value** is a value which is much greater than, or much less than, the other values.

Your task is to examine the effect of an outlier on the three measures of central tendency.

**What to do:**

**1** Consider the set of data: 4, 5, 6, 6, 6, 7, 7, 8, 9, 10. Calculate:
   **a** the mean **b** the mode **c** the median.

**2** We now introduce the extreme value 100 to the data, so the data set is now: 4, 5, 6, 6, 6, 7, 7, 8, 9, 10, 100. Calculate:
   **a** the mean **b** the mode **c** the median.

**3** Comment on the effect that the extreme value has on:
   **a** the mean **b** the mode **c** the median.

**4** Which of the three measures of central tendency is most affected by the inclusion of an outlier?

**5** Discuss with your class when it would not be appropriate to use a particular measure of the centre of a data set.

## Resistant Measures of Central Tendency

- A resistant measure will not be affected by extreme values in the data set.
- The mean is not resistant to extreme values.
- The median is resistant to extreme values.

## Weighted Average

- At times, we may need to assign more importance to some of the data values.

$$Weighted \quad Average = \frac{\sum xw}{\sum w}$$

- *x* is a data value.
- *w* is the weight assigned to that value.

## Weighted Average Example

In a pageant, the interview is worth 30% and appearance is worth 70%. Find the weighted average for a contestant with an interview score of 90 and an appearance score of 80.

$$\text{Weighted average} = \frac{0.30(90) + 0.70(80)}{0.30 + 0.70}$$

$$= \frac{27 + 56}{1.00} = 83$$

## Measures of Variation: Range

- Range = Largest value – smallest value
- Only two data values are used in the computation, so much of the information in the data is lost.

## Measures of Variation – Standard Deviation

- The standard deviation and variance measure the spread of the distribution around the mean of the data.
- The formulae for calculating the standard deviation and variance are slightly different if we are dealing with a population or a sample.

Finishing standard deviation from last class.

Measures of Variation

Standard deviation- How much the data varies from the mean.

EX: a sample of 8 grades were taken randomly from a class. They were as follows:
72, 82, 84, 90, 58, 95, 100, 75

$\Sigma$ = sum

$$\bar{x} = \frac{\Sigma x}{n} = \frac{656}{8} = 82$$

8 #'s

| $x$ | $x-\bar{x}$ (Deviations) | $(x-\bar{x})^2$ (Variances) |
|---|---|---|
| 72 | $72-82=-10$ | 100 |
| 82 | 0 | 0 |
| 84 | 2 | 4 |
| 90 | 8 | 64 |
| 58 | $-24$ | 576 |
| 95 | 13 | 169 |
| 100 | 18 | 324 |
| 75 | $-7$ | 49 |

$$\text{Var:} = S_x^2 = \frac{\Sigma(x-\bar{x})^2}{n-1} = \frac{1286}{7}$$

$$S_x = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}$$

## Calcluating a Sample Standard Deviation

1) Calculate the mean of the sample.
2) Subtract each data value from the mean (these are the deviations).
3) Square these terms (each deviation).
4) Add up the squared deviations.
5) Divide the sum by *n-1*; this is the sample variance.
6) Take the square root of the sample variance to obtain the sample standard deviation.

## Another Standard Deviation Example

| Column I $x$ | Column II $x - \bar{x}$ | Column III $(x - \bar{x})^2$ |
|---|---|---|
| 2 | $2 - 6 = -4$ | $(-4)^2 = 16$ |
| 3 | $3 - 6 = -3$ | $(-3)^2 = 9$ |
| 3 | $3 - 6 = -3$ | $(-3)^2 = 9$ |
| 8 | $8 - 6 = 2$ | $(2)^2 = 4$ |
| 10 | $10 - 6 = 4$ | $(4)^2 = 16$ |
| 10 | $10 - 6 = 4$ | $(4)^2 = 16$ |
| $\Sigma x = 36$ | | $\Sigma(x - \bar{x})^2 = 70$ |

Thus, $s^2 = \frac{70}{n-1} = \frac{70}{5} = 14$ And $s = \sqrt{14} = 3.742$

## Population Variance and Standard Deviation

- When dealing with the entire population, rather than a sample, we will divide the sum of squared-deviations by *N*, the entire population size.

## Population Variance and Standard Deviation

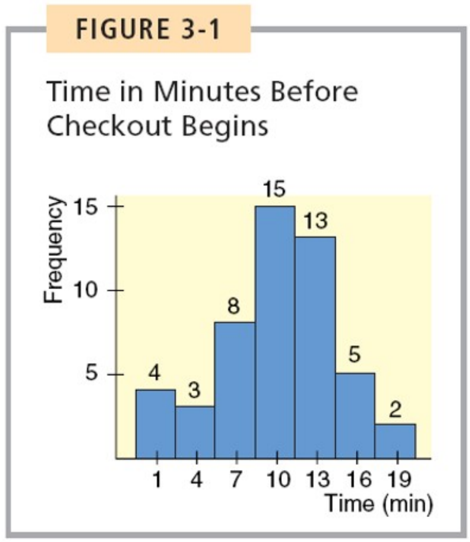
**Population Variance =** $\sigma^2 = \dfrac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}$

**Population Standard Deviation =** $\sigma = \sqrt{\sigma^2}$

## Grouped Data

- Sometimes, we wish to compute the mean and standard deviation of grouped data.

# Grouped Data Example




| I<br>Class | II<br>Frequency<br>*f* | III<br>Midpoint<br>*x* |
|---|---|---|
| 0–2 | 4 | 1 |
| 3–5 | 3 | 4 |
| 6–8 | 8 | 7 |
| 9–11 | 15 | 10 |
| 12–14 | 13 | 13 |
| 15–17 | 5 | 16 |
| 18–20 | 2 | 19 |

# Grouped Data Example

1) Make a frequency table.
2) Compute the midpoint, call it $x$.
3) Count the number of entries in each class, call that $f$.
4) Add the number of entries in each class to find $n$.

# Grouped Data Example

**Sample mean for a frequency distribution**

$$\bar{x} = \frac{\Sigma xf}{n}$$

**Sample standard deviation for a frequency distribution (defining formula)**

$$s = \sqrt{\frac{\Sigma(x - \bar{x})^2 f}{n - 1}}$$

# Grouped Data Example

| I<br>Class | II<br>Frequency<br>$f$ | III<br>Midpoint<br>$x$ | IV<br>$xf$ | V<br>$x - \bar{x}$ | VI<br>$(x - \bar{x})^2$ | VII<br>$(x - \bar{x})^2 f$ |
|---|---|---|---|---|---|---|
| 0–2 | 4 | | | | | |
| 3–5 | 3 | | | | | |
| 6–8 | 8 | | | | | |
| 9–11 | 15 | | | | | |
| 12–14 | 13 | | | | | |
| 15–17 | 5 | | | | | |
| 18–20 | 2 | | | | | |
| | $\Sigma f = 50$ | | | | | |

## Grouped Data Example

$$s = \sqrt{\frac{\Sigma(x - \overline{x})^2 f}{n - 1}}$$

$$= \sqrt{\frac{\text{Sum of Column VII}}{(\text{Sum of Column II}) - 1}}$$

$$= \sqrt{\frac{961.40}{50 - 1}} \approx \sqrt{19.62} \approx 4.43$$

Study for Test.

# Percentiles

- For whole numbers *P*, $1 \le P \le 99$, the $P^{th}$ percentile of a distribution is a value such that *P%* of the data fall below it, and *(100-P)%* of the data fall at or above it
  - Example: Your score on the psychology midterm was at the $73^{rd}$ percentile of your class. Thus, *73%* of the scores were below yours, and *27%* of the scores were at or above yours

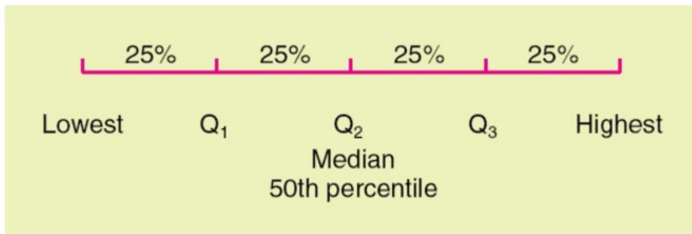
# Special Percentiles: Quartiles

- $Q_1 = 25^{th}$ Percentile
- $Q_2 = 50^{th}$ Percentile = The Median
- $Q_3 = 75^{th}$ Percentile

# Quartiles and Interquartile Range (IQR)

FIGURE 3-7

Quartiles




$$\text{Interquartile range} = Q_3 - Q_1$$

# Computing Quartiles

**PROCEDURE**

**How to compute quartiles**

1. Order the data from smallest to largest,
2. Find the median. This is the 2nd quartile.
3. The first quartile $Q_1$ is then the median of the lower half of the data; that is, it is the median of the data falling *below* the $Q_2$ position (and not including $Q_2$).
4. The third quartile $Q_3$ is the median of the upper half of the data; that is, it is the median of the data falling *above* the $Q_2$ position (and not including $Q_2$).

# Quartiles Example

- The calorie counts of 22 vanilla ice cream bars were measured and ordered:

| | | | | | |
|---|---|---|---|---|---|
| 111 | 131 | 147 | 151 | 151 | 182 |
| 182 | 190 | 197 | 201 | 209 | 234 |
| 286 | 294 | 295 | 310 | 319 | 342 |
| 353 | 377 | 377 | 439 | | |

- Median = (209 + 234) / 2 = 221.5

- $Q_1 = 182$
- $Q_3 = 319$
- IQR = 319 – 182 = 137

## Five Number Summary

- A listing of the following statistics:
  - Minimum, $Q_1$, Median, $Q_3$, Maximum
- For the ice cream calories example:
  - 111, 182, 221.5, 319, 439

# Box-and-Whisker Plot

- We can represent the five-number summary graphically by creating a box-and-whisker plot.

# Box-and-Whisker Plot Construction

**PROCEDURE**

**How to make a box-and-whisker plot**

1. Draw a vertical scale to include the lowest and highest data values.
2. To the right of the scale draw a box from $Q_1$ to $Q_3$.
3. Include a solid line through the box at the median level.
4. Draw solid lines, called *whiskers*, from $Q_1$ to the lowest value and from $Q_3$ to the highest value.

FIGURE 3-8

Box-and-Whisker Plot