Measures of Variation

Section 3-3

Objectives

- Describe data using measures of variation, such as range, variance, and standard deviation

Example: You own a bank and wish to determine which customer waiting line system is best. You sample 10 customer waiting line times (in minutes)

Branch A (Single Waiting Line)

6.5 6.6 6.7 6.8
7.1 7.3 7.4 7.7
7.7 7.7

Branch B (Multiple Waiting Lines)

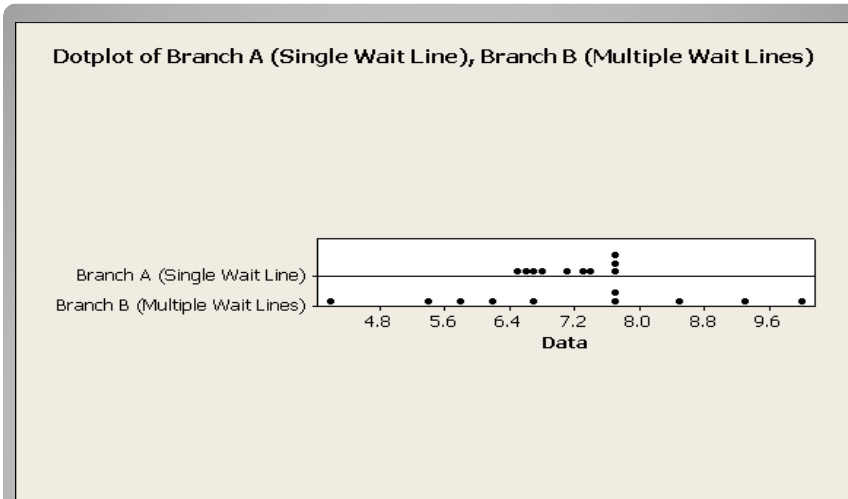
4.2 5.4 5.8 6.2
6.7 7.7 7.7 8.5
9.3 10.0

Find the measures of central tendency and compare the two customer waiting line systems. Which system is best?

	Branch A (Single Wait Line)	Branch B (Multiple Wait Lines)
Mean	7.15	7.15
Median	7.2	7.2
Mode	7.7	7.7
Midrange	7.1	7.1

Does this information help us to decide which is best?
Let's take a look at the distributions of each branch's wait times

Which is best –Branch A or Branch B?



Which is best –Branch A or Branch B?

- Since measures of central tendency are equal, one might conclude that neither customer waiting line system is better.
- But, if examined graphically, a somewhat different conclusion might be drawn. The waiting times for customers at Branch B (multiple lines) vary much more than those at Branch A (single line).

Insights

Measures of Variation

- Range
- Variance
- Standard Deviation

Range

- Range is the simplest of the three measures
- Range is the highest value (maximum) minus the lowest value (minimum)
- Denoted by **R**

R = maximum – minimum

- Not as useful as other two measures since it only depends on maximum and minimum

Measures of Variation

Variance

- Variance is the average of the squares of the distance each value is from the mean.
- Variance is an "unbiased estimator" (the variance for a sample tends to target the variance for a population instead of systematically under/over estimating the population variance)
- Serious disadvantage: the units of variance are different from the units of the raw data (variance = units squared or $(\text{units})^2$

Standard Deviation

- Standard Deviation is the square root of the variance
- Standard Deviation is usually positive
- Standard deviation units are the same as the units of the raw data

---

Variance

- Population variance, σ^2 (lowercase Greek sigma)

$$\sigma^2 = \frac{\sum (x-\mu)^2}{N}$$

where x is a data point

μ = population mean and

N = size of the population

Standard Deviation

- Population standard deviation, σ

$$\sigma = \sqrt{\frac{\sum (x-\mu)^2}{N}}$$

where x is a data point

μ = population mean and

N = size of the population

Variance

- Sample variance, s^2

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

where x is a data point

$\overline{x}$ = sample mean and

n = sample size

Standard Deviation

- Population standard deviation, σ

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

where x is a data point

$\overline{x}$ = sample mean

n = sample size

Example: You own a bank and wish to determine which customer waiting line system is best. You sample 10 customer waiting line times (in minutes)

Branch A (Single Waiting Line)

6.5	6.6	6.7	6.8
7.1	7.3	7.4	7.7
7.7	7.7		

Branch B (Multiple Waiting Lines)

4.2	5.4	5.8	6.2
6.7	7.7	7.7	8.5
9.3	10.0		

Find the range, variance, and standard deviation for each set of waiting times. Which system is best?

- STEP 1: Find the mean for the sample data set
- STEP 2: Subtract the mean from each data value (this helps us see how much "deviation" each data value has from the mean)
- STEP 3: Square each result from step 2 (Guarantees a positive value for the amount of "deviation" or distance from the mean)
- STEP4: Find the sum of the squares from step 3
- STEP 5: Divide the sum by (n-1), the sample size minus 1 (If you stop at this step, you have found the variance)
- STEP 6: Take square root of value from step 5 (This gives you the standard deviation)

Variance & Standard Deviation Calculation Procedure

- Since the formulas are so involved, we will use our calculators or MINITAB to determine the variance or standard deviation and focus our attention on the ***interpretation*** of the variance or standard deviation
- Why did I bother showing you? So you have some sense of what is going on behind the scenes and realize it is not magic, it's MATH

NO WORRIES!!!

- Variances and standard deviations are used to determine the spread of the data.
  - If the variance or standard deviation is large, the data is more dispersed. This information is useful in comparing two or more data sets to determine which is more (most) variable
- The measures of variance and standard deviation are used to determine the consistency of a variable
  - For example, in manufacturing of fittings, such as nuts and bolts, the variation in the diameters must be small, or the parts will not fit together

Uses of the Variance and Standard Deviation

- The variance and standard deviation are used to determine the number of data values that fall within a specified interval in a distribution
  - For example, Chebyshev's theorem shows that, for any distribution, at least 75% of the data values will fall within 2 standard deviations of the mean
- The variance and standard deviation are used quite often in inferential statistics

Uses of the Variance and Standard Deviation

	Branch A (Single Wait Line)	Branch B (Multiple Wait Lines)
Mean	7.15	7.15
Median	7.2	7.2
Mode	7.7	7.7
Midrange	7.1	7.1
Standard Deviation	0.48	1.82

Does this information help us to decide which is best?

Which is best –Branch A or Branch B?

- If you are in a hurry and do not have a calculator to assist with the calculation of the standard deviation, we can use the **Range Rule Of Thumb (RROT)**
- **RROT**

$$\approx \frac{R}{4} = \frac{\max - \min}{4}$$

- This is ONLY an estimate, but it is in the ballpark

In a hurry?

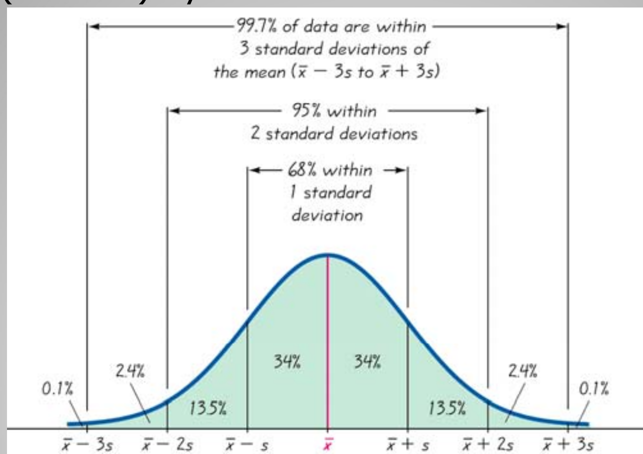
- RROT can also be used to *estimate* the maximum and minimum values of a data set. Most of the data in a dataset will lie within two standard deviations of the mean.
- Minimum "usual" value = $\bar{x} - 2s$
- Maximum "usual" value = $\bar{x} + 2s$

RROT

- Specifies the proportions of the spread in terms of the standard deviation
- Applies to ANY distribution
- The proportion of data values from a data set that will fall with k standard deviations of the mean will be AT LEAST

Chebyshev's Theorem (p.126)

- Only applies to bell-shaped (normal) symmetric distributions



Empirical (Normal) Rule