MATH 3339 - 03 15951
Statistics for the Sciences
Numerical description

Wendy Wang
wwang60@central.uh.edu

Lecture 9 - 3339
Outline

1. The 1.5IQR Rule
2. Understanding Standard Deviation
3. Calculating The Standard Deviation
Popper Set Up

- Fill in all of the proper bubbles.
- Make sure your ID number is correct.
- Make sure the filled in circles are very dark.
- This is popper number 04.
R-code for finding $Q_1$, $Q_2$, & $Q_3$

The values: Minimum, $Q_1$, Median ($Q_2$), $Q_3$, and Maximum are called the **Five Number Summary**

```r
> shoeprice=c(100,110,120,120,140,140,140,150,185,185,215,215,250,250,290)
> fivenum(shoeprice)
[1] 100 130 150 215 290
```
Interquartile range, $\text{IQR}$, is the difference between $Q_3$ and $Q_1$

$$IQR = Q_3 - Q_1$$
Detecting Outliers: 1.5IQR Rule

- An **outlier** is an observation that is "distant" from the rest of the data.

- Outliers can occur by chance or by measurement errors.

- Any point that falls outside the interval calculated by $Q_1 - 1.5(IQR)$ and $Q_3 + 1.5(IQR)$ is considered an outlier.
Outliers for Basketball Shoe Prices?

- Recall: $Q_1 = 130$, $Q_3 = 215$, So $IQR = 215 - 130 = 85$.

- $Q_1 - 1.5(IQR) = 130 - 1.5(85) = 2.5$

- $Q_3 + 1.5(IQR) = 215 + 1.5(85) = 342.5$

- Any price that is below $2.50$ or above $342.50$ is considered an outlier.
Outliers?

The following is information from 91 pairs of basketball shoes:

> fivenum(shoes$Price)
[1] 40 75 90 120 250

The highest four numbers in the dataset is . . . , 170, 225, 250, 250. Are there any prices that are considered an outlier?
A Graph of the Five Number Summary: Boxplot

- A central box spans the quartiles.
- A line inside the box marks the median.
- Lines extend from the box out to the smallest and largest observations.
- Asterisks represents any values that are considered to be outliers.
- Boxplots are most useful for side-by-side comparison of several distributions.

Rcode: `boxplot(dataset name$variable name)`
boxplot(shoes$Price, horizontal = T)
Boxplot of Course Scores by Session

```r
boxplot(grades$Score~grades$Session, horizontal=TRUE)
```
Measuring Spread: The Standard Deviation

- Measures spread by looking at how far the observations are from their mean.
- Most common numerical description for the spread of a distribution.
- A larger standard deviation implies that the values have a wider spread from the mean.
- Denoted $s$ when used with a sample. This is the one we calculate from a list of values.
- Denoted $\sigma$ when used with a population. This is the "idealized" standard deviation.
- The standard deviation has the same units of measurements as the original observations.
The standard deviation is the average distance each observation is from the mean.

- Using this list of values from a sample: 3, 3, 9, 15, 15
- The mean is 9.
- By definition, the average distance each of these values are from the mean is 6. So the standard deviation is 6.
The standard deviation is the average distance each observation is from the mean.

- Using this list of values from a sample: 3, 3, 9, 15, 15

- The mean is 9.

- By definition, the average distance each of these values are from the mean is 6. So the standard deviation is 6.
Definition of the Standard Deviation

The standard deviation is the average distance each observation is from the mean.

- Using this list of values from a sample: 3, 3, 9, 15, 15

- The mean is 9.

- By definition, the average distance each of these values are from the mean is 6. So the standard deviation is 6.
Definition of the Standard Deviation

The standard deviation is the average distance each observation is from the mean.

- Using this list of values from a sample: 3, 3, 9, 15, 15

- The mean is 9.

- By definition, the average distance each of these values are from the mean is 6. So the standard deviation is 6.
Values of the Standard Deviation

- The standard deviation is a value that is greater than or equal to zero.
- It is equal to zero only when all of the observations have the same value.
- By the definition of standard deviation determine $s$ for the following list of values.
  - $2, 2, 2, 2$: standard deviation = 0
  - $125, 125, 125, 125, 125$: standard deviation = 0
The standard deviation is a value that is greater than or equal to zero.

It is equal to zero only when all of the observations have the same value.

By the definition of standard deviation determine $s$ for the following list of values.

- $2, 2, 2, 2$ : standard deviation $= 0$
- $125, 125, 125, 125, 125$ : standard deviation $= 0$
The standard deviation is a value that is greater than or equal to zero.

It is equal to zero only when all of the observations have the same value.

By the definition of standard deviation determine $s$ for the following list of values.

- 2, 2, 2, 2 : standard deviation = 0
- 125, 125, 125, 125, 125: standard deviation = 0
The standard deviation is a value that is greater than or equal to zero.

It is equal to zero only when all of the observations have the same value.

By the definition of standard deviation determine \( s \) for the following list of values.

- 2, 2, 2, 2: standard deviation = 0
- 125, 125, 125, 125, 125: standard deviation = 0
The standard deviation is a value that is greater than or equal to zero.

It is equal to zero only when all of the observations have the same value.

By the definition of standard deviation determine $s$ for the following list of values.

- $2, 2, 2, 2$ : standard deviation = 0
- $125, 125, 125, 125, 125$ : standard deviation = 0
Values of the Standard Deviation

- The standard deviation is a value that is greater than or equal to zero.
- It is equal to zero only when all of the observations have the same value.
- By the definition of standard deviation determine $s$ for the following list of values.
  - $2, 2, 2, 2$: standard deviation $= 0$
  - $125, 125, 125, 125, 125$: standard deviation $= 0$
The standard deviation is a value that is greater than or equal to zero.

It is equal to zero only when all of the observations have the same value.

By the definition of standard deviation determine $s$ for the following list of values.

- $2, 2, 2, 2$: standard deviation $= 0$
- $125, 125, 125, 125, 125$: standard deviation $= 0$
Adding or subtracting the same value to all the original observations does not change the standard deviation of the list.

Using this list of values: 3, 3, 9, 15, 15 mean = 9, standard deviation = 6.

If we add 4 to all the values: 7, 7, 13, 19, 19

mean = 13, standard deviation = 6
Adding or Subtracting a Value to the Observations

- Adding or subtracting the same value to all the original observations does not change the standard deviation of the list.

- Using this list of values: 3, 3, 9, 15, 15 mean = 9, standard deviation = 6.

- If we add 4 to all the values: 7, 7, 13, 19, 19

  mean = 13, standard deviation = 6
Adding or Subtracting a Value to the Observations

- Adding or subtracting the same value to all the original observations does not change the standard deviation of the list.
- Using this list of values: 3, 3, 9, 15, 15 mean = 9, standard deviation = 6.
- If we add 4 to all the values: 7, 7, 13, 19, 19

  mean = 13, standard deviation = 6
Adding or Subtracting a Value to the Observations

- Adding or subtracting the same value to all the original observations does not change the standard deviation of the list.
- Using this list of values: 3, 3, 9, 15, 15 mean = 9, standard deviation = 6.
- If we add 4 to all the values: 7, 7, 13, 19, 19

  mean = 13, standard deviation = 6
Multiplying or dividing the same value to all the original observations will change the standard deviation by that factor.

Using this list of values: 3, 3, 9, 15, 15: mean = 9, standard deviation = 6.

If we double all the values: 6, 6, 18, 30, 30

mean = 18, standard deviation = 12
Multiplying or Dividing a Value to the Observations

- Multiplying or dividing the same value to all the original observations will change the standard deviation by that factor.

- Using this list of values: 3, 3, 9, 15, 15: mean = 9, standard deviation = 6.

- If we double all the values: 6, 6, 18, 30, 30

  mean = 18, standard deviation = 12
Multiplying or dividing the same value to all the original observations will change the standard deviation by that factor.

Using this list of values: 3, 3, 9, 15, 15: mean = 9, standard deviation = 6.

If we double all the values: 6, 6, 18, 30, 30

mean = 18, standard deviation = 12
Multiplying or dividing a Value to the Observations

Multiplying or dividing the same value to all the original observations will change the standard deviation by that factor.

Using this list of values: 3, 3, 9, 15, 15: mean = 9, standard deviation = 6.

If we double all the values: 6, 6, 18, 30, 30

mean = 18, standard deviation = 12
If $N$ is the number of values in a population with mean $\mu$, and $x_i$ represents each individual in the population, the population variance is found by:

$$\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$$

and the population standard deviation is the square root, $\sigma = \sqrt{\sigma^2}$. 
Most of the time we are working with a sample instead of a population. So the **sample variance** is found by:

\[
s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n - 1}
\]

and the **sample standard deviation** is the square root, \( s = \sqrt{s^2} \). Where \( n \) is the number of observations (samples), \( x_i \) is the value for the \( i^{th} \) observation and \( \bar{x} \) is the sample mean.
Calculating the Standard Deviation By Hand

When calculating by hand we will calculate $s$.

1. Find the mean of the observations $\bar{x}$.
2. Calculate the difference between the observations and the mean for each observation $x_i - \bar{x}$. This is called the deviations of the observations.
3. Square the deviations for each observation $(x_i - \bar{x})^2$.
4. Add up the squared deviations together $\sum_{i=1}^{n}(x_i - \bar{x})^2$.
5. Divide the sum of the squared deviations by one less than the number of observations $n - 1$. This is the variance

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n}(x_i - \bar{x})^2$$
Step 6: Standard Deviation

6. Find the square root of the variance. This is the **standard deviation**

\[ s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2} \]
Example: Section A

Determine the sample standard deviation of the test scores for Section A.

<table>
<thead>
<tr>
<th>Section A Scores ($X_i$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>65</td>
</tr>
<tr>
<td>66</td>
</tr>
<tr>
<td>67</td>
</tr>
<tr>
<td>68</td>
</tr>
<tr>
<td>71</td>
</tr>
<tr>
<td>73</td>
</tr>
<tr>
<td>74</td>
</tr>
<tr>
<td>77</td>
</tr>
<tr>
<td>77</td>
</tr>
<tr>
<td>77</td>
</tr>
</tbody>
</table>
Step 1: Calculate the Mean

The sample mean is $\bar{x} = 71.5$. 

### Use Table To Calculate Standard Deviation

<table>
<thead>
<tr>
<th>Variable Score ($X_i$)</th>
<th>Deviations $X_i - \bar{X}$</th>
<th>Deviations Squared $(X_i - \bar{X})^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>66</td>
<td></td>
<td></td>
</tr>
<tr>
<td>67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>68</td>
<td></td>
<td></td>
</tr>
<tr>
<td>71</td>
<td></td>
<td></td>
</tr>
<tr>
<td>73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>77</td>
<td></td>
<td></td>
</tr>
<tr>
<td>77</td>
<td></td>
<td></td>
</tr>
<tr>
<td>77</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sum</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Step 2: Calculate Deviations For All Values

<table>
<thead>
<tr>
<th>Variable Score ((X_i))</th>
<th>Deviations (X_i - \bar{X})</th>
<th>Deviations Squared ((X_i - \bar{X})^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>65</td>
<td>65 − 71.5 = −6.5</td>
<td></td>
</tr>
<tr>
<td>66</td>
<td>66 − 71.5 = −5.5</td>
<td></td>
</tr>
<tr>
<td>67</td>
<td>67 − 71.5 = −4.5</td>
<td></td>
</tr>
<tr>
<td>68</td>
<td>68 − 71.5 = −3.5</td>
<td></td>
</tr>
<tr>
<td>71</td>
<td>71 − 71.5 = −0.5</td>
<td></td>
</tr>
<tr>
<td>73</td>
<td>73 − 71.5 = 1.5</td>
<td></td>
</tr>
<tr>
<td>74</td>
<td>74 − 71.5 = 2.5</td>
<td></td>
</tr>
<tr>
<td>77</td>
<td>77 − 71.5 = 5.5</td>
<td></td>
</tr>
<tr>
<td>77</td>
<td>77 − 71.5 = 5.5</td>
<td></td>
</tr>
<tr>
<td>77</td>
<td>77 − 71.5 = 5.5</td>
<td></td>
</tr>
</tbody>
</table>

sum
### Step 3: Calculate Squared Deviations

<table>
<thead>
<tr>
<th>Variable Score $(X_i)$</th>
<th>Deviations $X_i - \bar{X}$</th>
<th>Deviations Squared $(X_i - \bar{X})^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>65</td>
<td>$65 - 71.5 = -6.5$</td>
<td>$(-6.5)^2 = 42.25$</td>
</tr>
<tr>
<td>66</td>
<td>$66 - 71.5 = -5.5$</td>
<td>$(-5.5)^2 = 30.25$</td>
</tr>
<tr>
<td>67</td>
<td>$67 - 71.5 = -4.5$</td>
<td>$(-4.5)^2 = 20.25$</td>
</tr>
<tr>
<td>68</td>
<td>$68 - 71.5 = -3.5$</td>
<td>$(-3.5)^2 = 12.25$</td>
</tr>
<tr>
<td>71</td>
<td>$71 - 71.5 = -0.5$</td>
<td>$(-0.5)^2 = 0.25$</td>
</tr>
<tr>
<td>73</td>
<td>$73 - 71.5 = 1.5$</td>
<td>$1.5^2 = 2.25$</td>
</tr>
<tr>
<td>74</td>
<td>$74 - 71.5 = 2.5$</td>
<td>$2.5^2 = 6.25$</td>
</tr>
<tr>
<td>77</td>
<td>$77 - 71.5 = 5.5$</td>
<td>$5.5^2 = 30.25$</td>
</tr>
<tr>
<td>77</td>
<td>$77 - 71.5 = 5.5$</td>
<td>$5.5^2 = 30.25$</td>
</tr>
<tr>
<td>77</td>
<td>$77 - 71.5 = 5.5$</td>
<td>$5.5^2 = 30.25$</td>
</tr>
</tbody>
</table>

**sum**
Step 4: Calculate the Sum of the Squared Deviations

<table>
<thead>
<tr>
<th>Variable Score ($X_i$)</th>
<th>Deviations $X_i - \bar{X}$</th>
<th>Deviations Squared $(X_i - \bar{X})^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>65</td>
<td>$65 - 71.5 = -6.5$</td>
<td>$(-6.5)^2 = 42.25$</td>
</tr>
<tr>
<td>66</td>
<td>$66 - 71.5 = -5.5$</td>
<td>$(-5.5)^2 = 30.25$</td>
</tr>
<tr>
<td>67</td>
<td>$67 - 71.5 = -4.5$</td>
<td>$(-4.5)^2 = 20.25$</td>
</tr>
<tr>
<td>68</td>
<td>$68 - 71.5 = -3.5$</td>
<td>$(-3.5)^2 = 12.25$</td>
</tr>
<tr>
<td>71</td>
<td>$71 - 71.5 = -0.5$</td>
<td>$(-0.5)^2 = 0.25$</td>
</tr>
<tr>
<td>73</td>
<td>$73 - 71.5 = 1.5$</td>
<td>$1.5^2 = 2.25$</td>
</tr>
<tr>
<td>74</td>
<td>$74 - 71.5 = 2.5$</td>
<td>$2.5^2 = 6.25$</td>
</tr>
<tr>
<td>77</td>
<td>$77 - 71.5 = 5.5$</td>
<td>$5.5^2 = 30.25$</td>
</tr>
<tr>
<td>77</td>
<td>$77 - 71.5 = 5.5$</td>
<td>$5.5^2 = 30.25$</td>
</tr>
<tr>
<td>77</td>
<td>$77 - 71.5 = 5.5$</td>
<td>$5.5^2 = 30.25$</td>
</tr>
<tr>
<td>sum</td>
<td>$\sum_{i=1}^{n} (X_i - \bar{X})^2 = 204.5$</td>
<td></td>
</tr>
</tbody>
</table>
Step 5: Calculate the Variance

\[
\text{variance} = s^2 \\
= \frac{1}{n - 1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \\
= \frac{204.5}{9} \\
= 22.7222
\]
Step 6: Take the Square Root of the Variance

standard deviation \(= s\)  
\[= \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2}\]  
\[= \sqrt{22.7222}\]  
\[= 4.77\]