Do Now:

1.) Give me your height in inches when I check homework.

2.) A group of numbers has mean of 375, median 280, and a mode of 250. Tell which numbers you would use to represent the data and why.

3.) A class had an average of 69.8 on the first test, with a standard deviation of 6.2. Another class has an average of 70.4, with a standard deviation of 13.4. Compare the performance of the two classes.

Homework Questions?

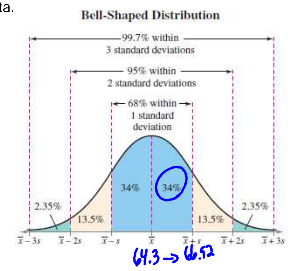
Sep 8-10:25 AM

Empirical Rule:

In a survey conducted by the National Center for Health Statistics, the sample mean height of women in the United States (ages 20-29) was 64.3 inches, with a sample standard deviation of 2.62 inches.

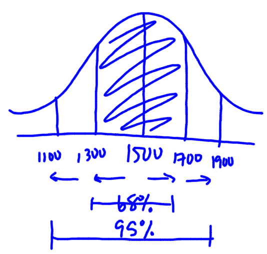
Estimate the percent of women's heights that are between 64.3 and 66.92 inches tall.

a. How many *standard deviations* is 66.92 to the right of 64.3?
b. Use the *Empirical Rule* to estimate the percent of the data
c. *Interpret* the result in the context of the data.



64.3 → 66.92

Jan 11-7:15 AM



.68 x 75 =

.95 x 75 =

Jan 11-7:56 AM

Unit 1 Day 4
Data Description

(3-3) Measures of Position

Mar 29-5:41 PM

Measures of position are used to locate the relative position of a data value in a data set.

For example, when you took the ACT. You received a score and then a percentile. So for instance...

80th Percentile: 80% of the people who took the ACT scored below your score.

It **does not** mean you earned an 80%

Mar 29-5:42 PM

I. z-scores

Standard Scores (a.k.a. **z scores**):

a z-score is used to compare different sets of data.

Who is taller, a man 73 inches tall or a woman 68 inches tall? The obvious answer is that the man is taller. However, men are taller than women on the average. Let's ask the question this way: Who is taller relative to their gender, a man 73 inches tall or a woman 68 inches tall?

The **z-score** of an individual data value tells how many standard deviations that value is from its population mean.
Let x be a value from a population with mean μ and standard deviation σ. The z-score for x is

$$z = \frac{X - \mu}{\sigma}$$

Mar 29-5:43 PM

I. z-scores

$$z = \frac{X - \mu}{\sigma} \rightarrow z\sigma = X - \mu$$

$$X = \mu + z\sigma$$

Practice

1. A National Center for Health Statistics study states that the mean height for adult men in the U.S. is μ = 69.4 inches, with a standard deviation of σ = 3.1 inches. The mean height for adult women is μ = 63.8 inches, with a standard deviation of σ = 2.8 inches. Who is taller relative to their gender, a man 73 inches tall, or a woman 68 inches tall?

man $z = \frac{73 - 69.4}{3.1} \approx 1.16$ woman $z = \frac{68 - 63.8}{2.8} \approx 1.50$

2. Eric proudly tells his brother Bruce that he got 94 points on his last math exam, which had an average of 73 points and a standard deviation of 9. Bruce says that he did even better on his math exam, on which he got 96 points, and this exam had an average of 79 points and a standard deviation of 7. Who did better on his exam relative to their class scores?

Eric $z = 2.33$ Bruce $z = 2.43$

3. Suppose Eric's classmate got a z-score of -1.7 on his math exam. What was his exam score?

$X = 73 + (-1.7)(9)$

$X = 58$ points

Mar 29-5:46 PM

I. z-scores

* The larger the z score the higher the relative position.

* If the z score is **positive** then the score is above the mean.

*If the z score is **negative** then the score is below the mean.

Mar 29-5:48 PM

II. Percentiles

Percentiles are position measures used to indicate the position of an individual in a group. (in educational and health-related fields)

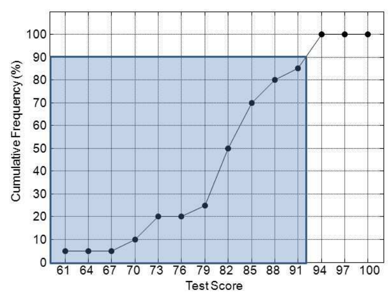
Percentile graphs can be used to estimate a percentile or a corresponding value for a given percentile.

They use the same value as the cumulative frequency, but change them to percents.

Mar 29-5:49 PM

Estimating percentile points and percentile ranks from the cumulative percentage curve

Estimate the percentile point for a percentile rank of 90%



90% of the scores fall below a score of about 92. (P_{90} is about 92)

Feb 8-11:05 AM

II. Percentiles

$$\boxed{\text{Percentile rank of } x = \frac{(\text{Number of values less than } x) + 0.5}{n} \cdot 100}$$

(where x is some number in the data set)

Round the result to the nearest whole number.

ex. *Given the following data set*

~~15~~ ~~9~~ ~~12~~ ~~11~~ ~~7~~ ~~6~~ ~~9~~ ~~10~~ ~~14~~ ~~3~~ ~~6~~ ~~5~~

3 5 6 6 7 9 9 10 11 12 14 15

(a) *Calculate the approximate value of the 55th percentile.*

Find the percentile rank of 7

$\frac{4 + 0.5}{12} \cdot 100 = 37.5$

38^{th}

Mar 29-5:51 PM

II. Percentiles

To find the approximate value of the *p*th percentile:

1. sort the data in increasing order
2. calculate L= $\left(\frac{p}{100} \cdot n\right)$
3. If L is a whole number, the *p*th percentile is the average of the number in position L and the number in position L+1.
   If L is NOT a whole number, the *p*th percentile is the number in the position of the next whole number higher than L.

(b) ~~Find the percentile rank of 7.~~

Calc the approx value of the 55^{th} percentile

$L = \frac{55}{100} \cdot 12 = 6.6$ round up 7

a score of 9 would be in the 55^{th} percentile

Sep 6-3:11 PM

II. Percentiles

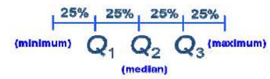
Using the heights of students (in inches) in this classroom.

EX: Find the percentile rank of a height of___

EX: Find the value corresponding to the 75th percentile

Jan 31-7:27 AM

Quartiles Q_1, Q_2, Q_3



5 - number summary

1. Minimum value
2. Q_1
3. The median Q_2
4. Q_3
5. Maximum value

Jan 9-3:28 PM

IQR = Interquartile Range = $Q_3 - Q_1$

The **IQR method** allows us to determine which values are outliers. Outliers are data values that are below $Q_1 - 1.5 \cdot IQR$, the lower outlier boundary, or above $Q_3 + 1.5 \cdot IQR$, the upper outlier boundary.

Find the five number summary and the IQR for the given data sets below. Determine if they have any outliers.

Q_1 average Q_2 Q_3 -> average

ex. 1 3 6 7 9 9 10 14 15

min 1
Q_1 4.5
Q_2 9
Q_3 12
max 15

$IQR = 12 - 4.5 = 7.5$

$4.5 - 1.5 \cdot 7.5 \rightarrow 4.5 - 11.25 = -6.75$ } range
$12 + 1.5 \cdot 7.5 \rightarrow 12 + 11.25 = 23.25$

no outliers

Q_2 average

ex. 3 2 10 8 2 9

2 2 3 8 9 10

min 2
Q_1 2
Q_2 5.5
Q_3 9
max 10

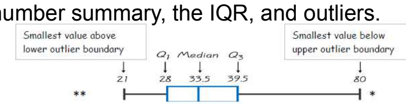
$IQR = 9 - 2 = 7$

$2 - 1.5 \cdot 7 \rightarrow 2 - 10.5 = -8.5$ } range
$9 + 1.5 \cdot 7 \rightarrow 9 + 10.5 = 19.5$

no outliers

Jan 10-12:28 PM

Boxplots- graphical representation of the five number summary, the IQR, and outliers.



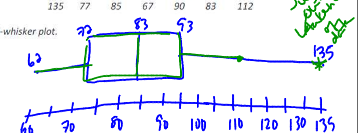
ex. *The time (in minutes) that a student spent in the laundromat in a week, for 15 randomly selected weeks, is as follows:*

72	62	84	73	107	81	93	72
135	77	85	67	90	83	112	

green with outlier taken out of data

(a) *Prepare a box-and-whisker plot.*

min 62
Q_1 72
Q_2 83
Q_3 93
max 135



(b) *Is the data skewed?* skewed right -> positively

(c) *Does the data contain any outlier?*

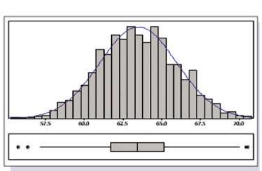
$IQR = 21$

$72 - 21 \cdot 1.5 = 72 - 31.5 = 40.5$ } range
$93 + 21 \cdot 1.5 = 93 + 31.5 = 124.5$

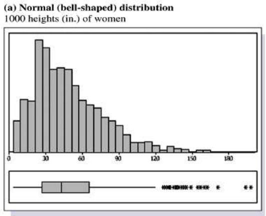
yes 135 is an outlier

Jan 10-12:30 PM

A boxplot can help us better see the distribution of the data, such as the center, spread, skewness, and outliers.



(a) Normal (bell-shaped) distribution
1000 heights (in.) of women



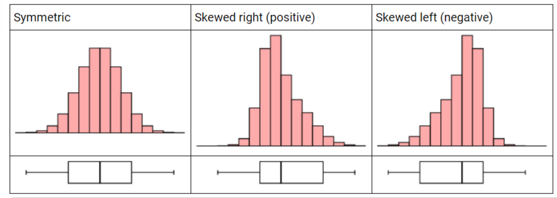
(c) Skewed distribution
Incomes (thousands of dollars) of 1000 statistics professors

Jan 10-2:40 PM

Information Obtained from a Boxplot

1. *a.* If the median is near the center of the box, the distribution is approximately symmetric.
   b. If the median falls to the left of the center of the box, the distribution is positively skewed.
   c. If the median falls to the right of the center, the distribution is negatively skewed.
2. *a.* If the lines are about the same length, the distribution is approximately symmetric.
   b. If the right line is larger than the left line, the distribution is positively skewed.
   c. If the left line is larger than the right line, the distribution is negatively skewed.

Symmetric	Skewed right (positive)	Skewed left (negative)



Jan 9-3:31 PM

Homework:
WS Measures of Position, Box Plots, and outliers

Unit 1 Test

Thursday 1/17

Jan 31-8:45 AM