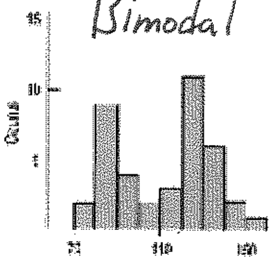
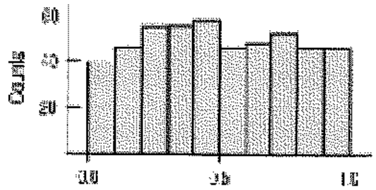
Algebra 2 **Chapter 9 Notes**

9.1 Notes Shapes

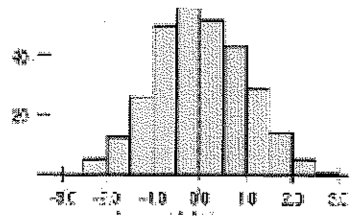
Bimodal



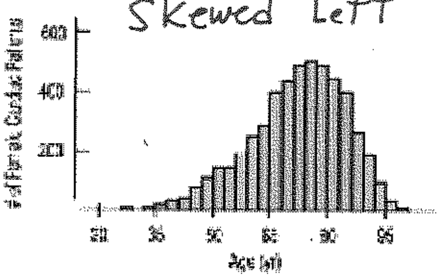
Uniform



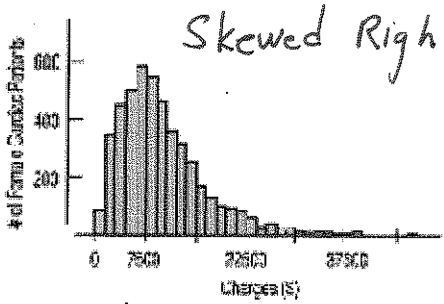
(Bell Curve)
Normal



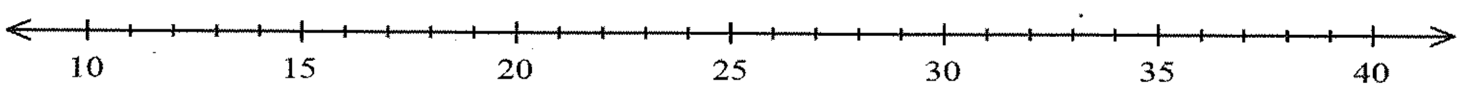
Skewed Left



Skewed Right



Students: come up to the board and create a dot plot of your ages:



Which shape does this most closely resemble?

If your teacher decides to add his or her age, how does it change the shape?

When is data normal?

What shape do you think these data would have?

1. The test scores for an Algebra 2 test:

2. The # of Instagram followers for kids in Alg 2 at Damonte:

3. The # of pushups the boys in the room can do:

4. The # on a 6-sided die if tossed 100 times:

Types of Data:

Categorical:
categories, not values
(not #'s w/ meaning)

Quantitative:
numerical data
(you can measure with)

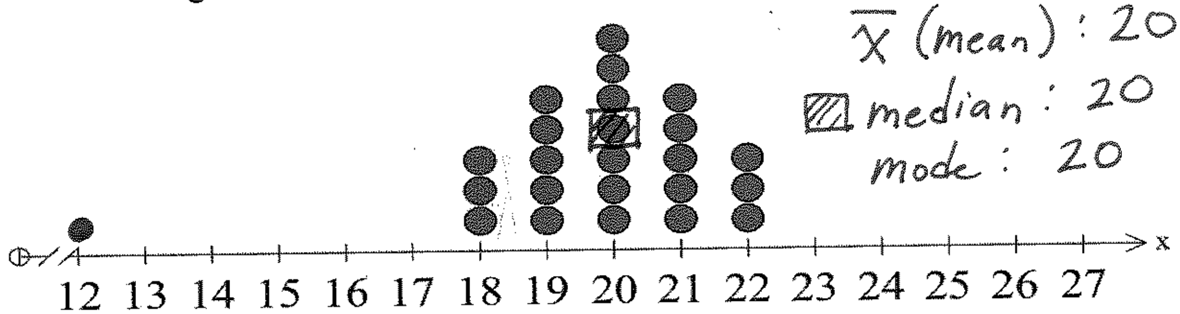
What type of data?

Type of pet C
GPA Q
Gender C
Level of approval C
Age Q
Phone Number C
(These #'s don't have real meaning)

Center

Here are the ages of students in a college course. Mark the median, mean, and mode.

$\bar{x}$ (mean) : 20
median : 20
mode : 20



Let's say a 12-year old child genius is added to the course. Now mark the median, mean, and mode. What has happened to these values?

$\bar{x} = 19.67$
med: 20
mode: 20

- mean decreases
- median move 1/2 a space (still 20)
- mode doesn't change

What is the new shape of the data?

skewed Left

Measures of Center

- Mean: (average) measure of center when data is normal
- Median: middle # when data is in order
  (* use if data is skewed &/or has outliers)
- Mode: # that occurs the most.

5-number summary

Create a 5-number summary from the data on the right, which represents the number of minutes people waited for their doctor when arriving on time for an appointment:

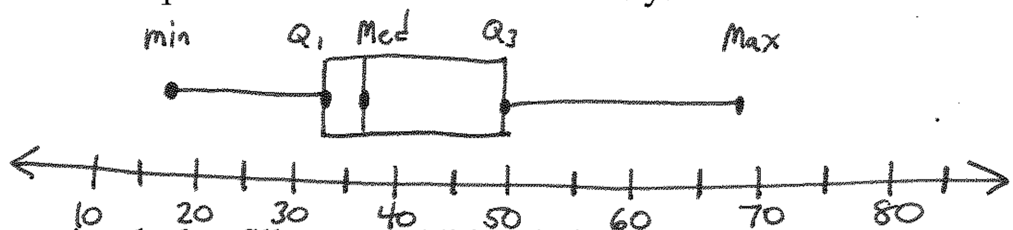
6	28
5	0468
4	0346
3	3344466$\boxed{7}$8
2	289
1	8

Min: 18
Q1: 33
Median: 37
Q3: 50
Max: 68

Find the range: 68 − 18 = 50
(max − min)

Find the Interquartile Range:
50 − 33 = 17
$(Q_3 - Q_1)$

Now create a box plot from the 5 number-summary:



If everyone is asked to fill out an additional piece of paperwork and their wait times all increase by 5 minutes, what happens to the:

IQR: stays the same.　　　Median: increases by 5 min.

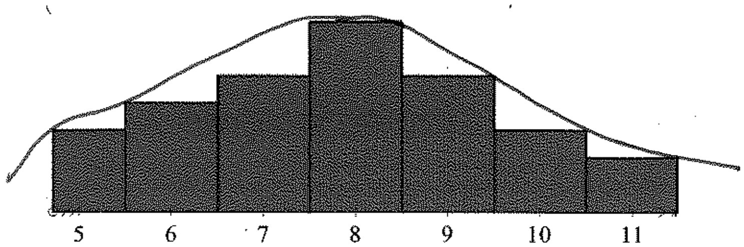
If all of the wait times are doubled, what happens to the:

IQR: doubles　　　Median: doubles.

Spread

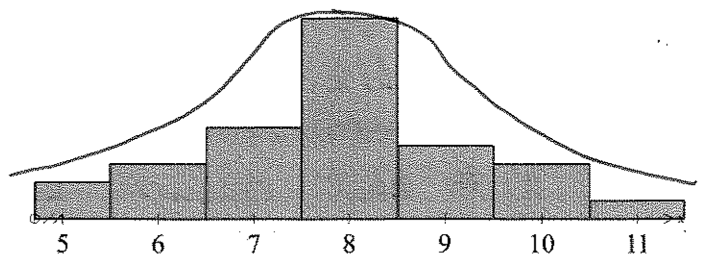
Flatter bell
more variability

Class A Quiz Scores



Higher Peak
less variability

Class B Quiz Scores



Compare the center and the range of the two classes:

Same for both

Which class would have a larger IQR?

class A

Which class shows less variability?

class B, higher peak

Measures of Spread

- Standard Deviation: shows how spread out data is from the mean. Best used w/ a normal distribution
- Range: High − Low
- IQR: $Q_3 - Q_1$

Outliers

If data looks *normal* (symmetric/unimodal): $\underline{mean} \pm \underline{2 \cdot standard\ deviation}.$
(We will focus more on this in 9.3.)

If data doesn't look *normal*:
use the IQR method.
1. Find IQR: $Q_3 - Q_1$
2. Find $1.5(IQR)$
3. Subtract $1.5(IQR)$ from Q_1, add $1.5(IQR)$ to Q_3 these are your "fences"

★ Any value outside of the fence is an outlier

Here are some Algebra 2 test scores. Find the five-number summary, identify any outliers, and create a box plot. *Denote the outliers with asterisks.*

63*, 72, 75, 78, 80, 81, 85, 86, 86, | 86, 87, 87, 88, 90, 95, 96, 98, 108*

min = 63 $\quad Q_1 = 80 \quad$ Med = 86 $\quad Q_3 = 90 \quad$ Max = 108

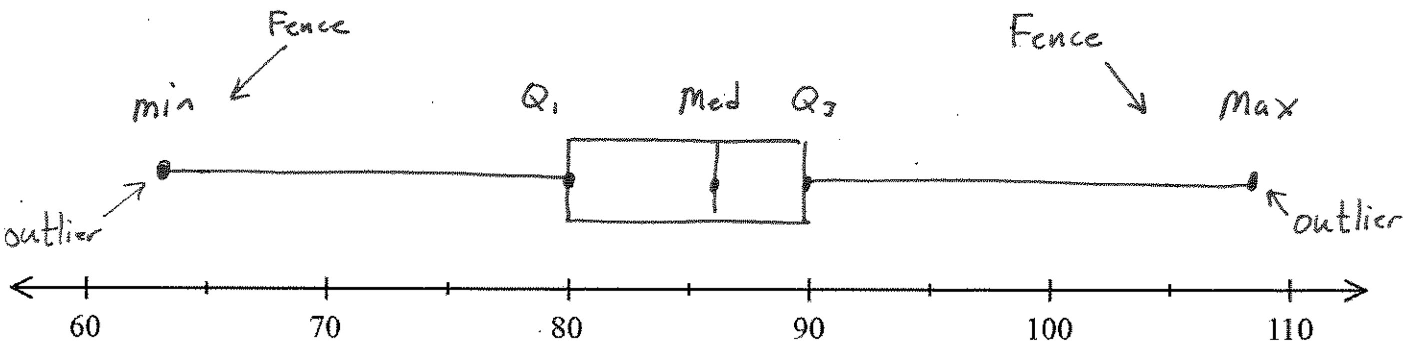
$\bar{x} = 85.61$

IQR = 90 − 80 = 10

1.5(IQR) = 1.5(10) = 15

Fences: 80 − 15 = 65

90 + 15 = 105



What is the shape of this data?

Normal

Would you expect the mean or median to be larger?

Median

What is the IQR?

$Q_3 - Q_1$

$90 - 80 = \boxed{10}$

What is the range?

Max − Min

108 − 63 = 45

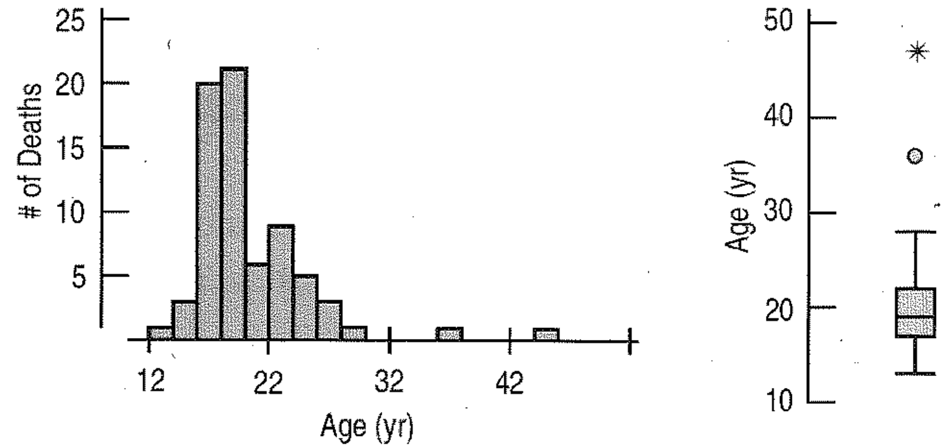
What percentage of the data is between 80 and 108?

75%

Between 63 and 80?

25%

Example: Crowd Management Strategies monitors accidents at rock concerts. In their database, they list the names and other variables of victims whose deaths were attributed to "crowd crush" at rock concerts. Here are the histogram and boxplot of the victims' ages for data from 1999 to 2000:



a) What features of the distribution can you see in both the histogram and the boxplot?

b) What features of the distribution can you see in the histogram that you could not see in the boxplot? Mode

c) What summary statistic would you choose to summarize the center of this distribution? Why?
Median, the data is skewed.

d) What summary statistic would you choose to summarize the spread of this distribution? Why? IQR, b/c data is skewed.

9.2 Notes: Statistical Studies and Sampling Techniques

Population – all the members of a group being studied

Parameter – a measurement of the *population*

Sample – a subset of the population that's being studied

Statistic – a measurement of the *sample*

1. The principal wants to know how many students at her school have jobs. She samples 100 of the students and found that 25% of them were employed.

Population: All students at school

Sample: 100 students

Parameter: The true (unknown) % employed

Statistic: 25% employed.

2. A biology teacher wants to know which dissection his students liked best this semester, so he randomly selects eight students from each of his five classes to participate in a survey.

What is the population?
All students

What is the sample?
40 students, 8 from each class

3. At the end of the month, a restaurant manager used the total food sales for the month and the total number of customers for the month to calculate that the mean amount spent per customer during that month was $12.59.
What's the population? Total food sales.

Is this value a statistic or a parameter?
Parameter

Sample survey – asks every member of sample the same set of questions and records the answers

Observational Study – measures or observes members of a sample in such a way that they are not affected by the study.

Experiment – involving applying a treatment to some group or groups and measures the effects of the treatments

Identify if this is a survey, observational study, or experiment:

4. Lions are watched in the wild and the amount of time they sleep is recorded.

O.S.

5. Ask randomly chosen students in the school their opinion and record the results.

S

6. Doctors track children throughout their childhood to see if cancer is ever diagnosed.

O.S.

7. Split plants into groups randomly and give each group different conditions such as extra water, light, or different fertilizer. Leave one group under the current conditions to see if the new treatments show improvement.

E

8. A gym asks their customers if they would prefer the gym to be open earlier in the morning.

S

9. A gym counts the number of customers who work out before 8am.

O.S.

Bias – a systematic error in a study that results in a sample misrepresenting a population

Variability – spread or range of the data. Within most data there is natural variation.

- ***Do you think there could be bias in these samples and why?***

10. A soft drink company calls 500 people at random and asks, "Isn't it true that our product is better than our rival's product?" and 75% of the people respond, "Yes."

bias, wording of ? may influence people.

11. A PTA has a community meeting and they ask everyone who attends if they agree with the budget and 70% agree.

bias, only people attending vote.

12. A teacher wants to know if kids at her school enjoy her class, so sends 5 randomly selected students from each of her classes to a computer lab to take a survey.

not bias, random w/o teacher influence.

13. A teacher wants to know if his students find it easy to turn in homework, so he gives a survey to all of his students and asks them to turn it in when they're done filling it out. He figures since has 150 students, he will have enough.

bias, only those good at turning in will respond.

14. A scientist wants to know how happy people are in the morning at 9:00. She samples 100 people in a subway car on their way to work and her statistic was 35%. Her colleague sampled 100 people at the dog park and his statistic was 65%. Are their differences due to variability or bias?

bias, population not representative.

15. A radio station wants to know if its listeners liked the last song, so they tell them to go to their facebook page and complete the survey.

bias, only those w/ strong opinion will respond.

Experiment – involving applying a treatment to some group or groups and measures the effects of the treatments.

Treatments – the treatments must be **randomly assigned** to more than one group. If there are no treatments *assigned* and/or they are not *randomly* assigned, there is no experiment.

Control Group – A control group is a group that receives the treatment of *nothing* (can be a placebo). This way the **experimental group** can be compared with a baseline. There doesn't need to be a control group in an experiment, but it is very common.

16. Describe a design for a controlled experiment to test a new drug for treating the flu on mice. How does randomization apply to your design?

subjects $\xrightarrow{\text{random}}$ new drug → compare results.
subjects $\xrightarrow{\text{random}}$ placebo → compare results.

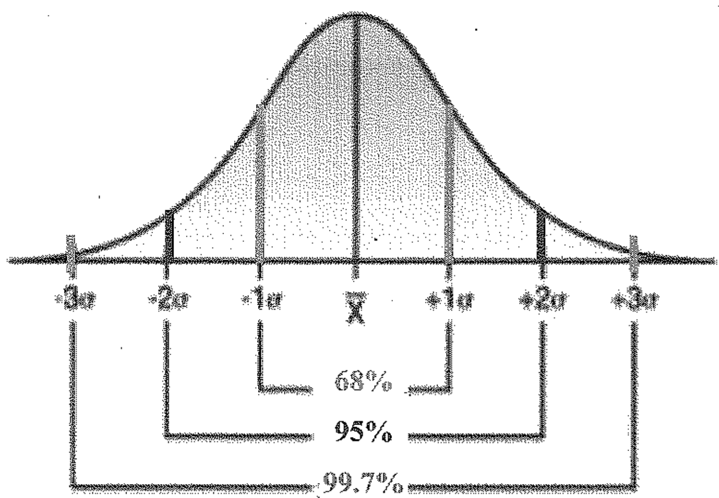
17. Design a controlled experiment to test whether drinking coffee improves memory. How would you choose the experimental and control groups?

subjects → must drink coffee → compare results.
random
subjects → no coffee → compare results.

9.3 Notes: Normal Curves and Histograms

Given a distribution with a normal shape that is made out of a curve, the distribution can be separated into sections by adding and subtracting standard deviations from the mean.

The percentages of the area under the curve for each section will be the same for every normal curve, as shown below.



The percentages under the curve can be remembered by using the

68 - 95 - 99.7 **rule.**

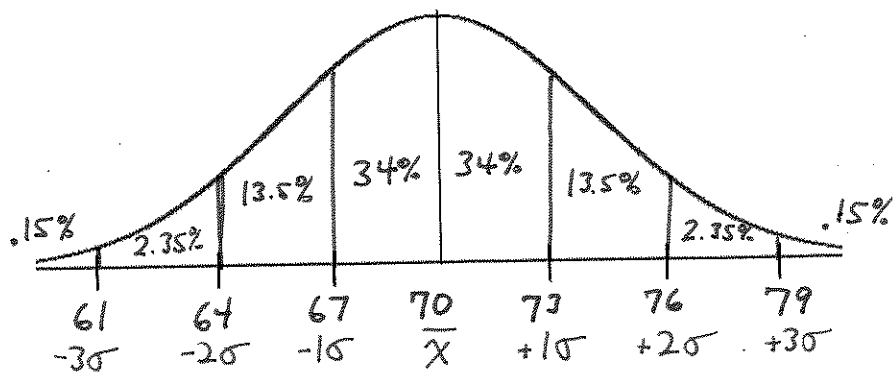
Thus, the area under the curve that is within 1 standard deviation of the mean is 68 % of the total area under the normal curve.

Also, the area under the curve that is within 2 standard deviations of the mean is 95 % of the total area under the normal curve.

Finally, the area under the curve that is within 3 standard deviations of the mean is 99.7 % of the total area under the normal curve.

Ex 1: The height of adult American males is normally distributed with a mean of 70 in and a standard deviation of 3 in.

a) Draw the normal curve to represent the heights of adult American males.



b) What percentage of adult American males are between 64 inches and 76 inches tall?

95%

c) What percentage of adult American males are greater than 70 inches tall?

50%

d) What percentage of adult American males are less than 67 inches tall?

16%

e) What percentage of adult American males are greater than 76 inches tall?

2.5%

f) What percentage of adult American males are between 67 and 79 inches tall?

83.85%

g) True or false? Approximately the same number of adult American males are taller than 79 inches as those who are shorter than 61 inches.

True

h) As of the US Census taken in 2010, there are approximately 113 million adult men in the United States. *How many* of them are between 67 and 73 inches tall?

68% of 113 million

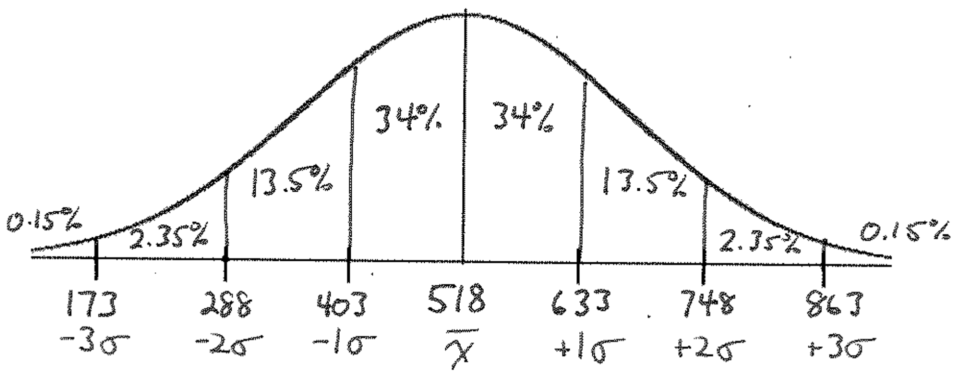
$\boxed{\text{76.84 million}}$

Outliers: A score is an outlier if it is more than ___2___ standard deviations away from the mean.

Percentile ranking: A score has a certain percentile based on the percentage of scores ___less___ than that score.

Example 2: In 2011, SAT scores are have a normal distribution with the mean score of the math portion of the SAT is 518, with a standard deviation of 115.

a) Draw the normal curve.



b) What is the **probability** that a student scored less than 403 on the math portion of the SAT?

16% ~~or~~ .16

c) What is the **probability** that a student scored between 288 and 518 on the math portion of the SAT?

47.5% or .475

d) What is the **probability** that a student scored between 288 and 633 **or** more than 748 on the math portion of the SAT? 81.5% 2.5%

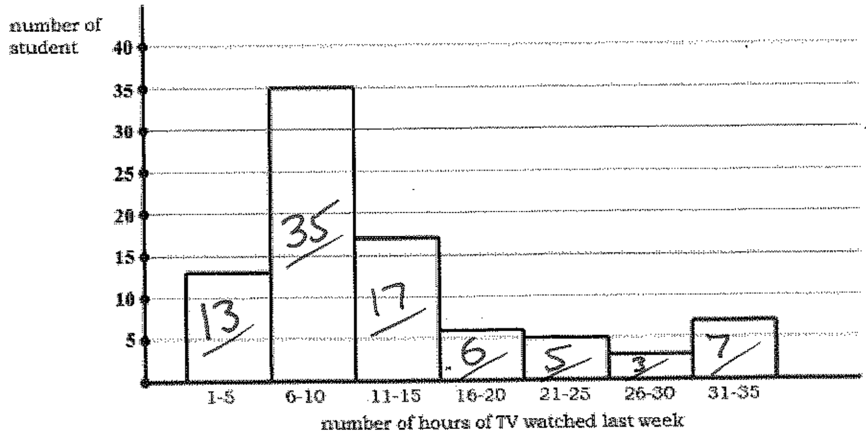
84% or .84

e) Terrance scored 750 on the math portion of the SAT. Is his score an outlier? Justify your conclusion. Yes, 750 is more than 2σ above $\bar{x}$.

f) Sara scored 633 on the math portion of the SAT. The college that she wants to attend requires that students score in the 85th percentile or higher on that math portion of the SAT. Is Sara's score high enough for her to be considered by this college?

No, her score is in the 84th percentile.

Example 3: In a survey, high school students were randomly selected and asked how many hours of television they had watched in the previous week. The histogram to the right displays their answers.



a) Approximately how many students participated in the survey?

86

b) Describe the shape of the distribution.

skewed right

c) Approximately how many students watched 10 hours or less of TV last week?

48

d) Approximately how many students watched between 16 and 30 hours of TV last week?

14

e) In which category is the median of the data? Explain. The median is the average of the 43^{rd} and 44^{th} data which are both in the 6-10 category.

A) 6 – 10 hours B) 11 – 15 hours C) 16 – 20 hours D) 21 – 25 hours

f) Is it possible to calculate the *mean* of the data from a histogram? Explain.

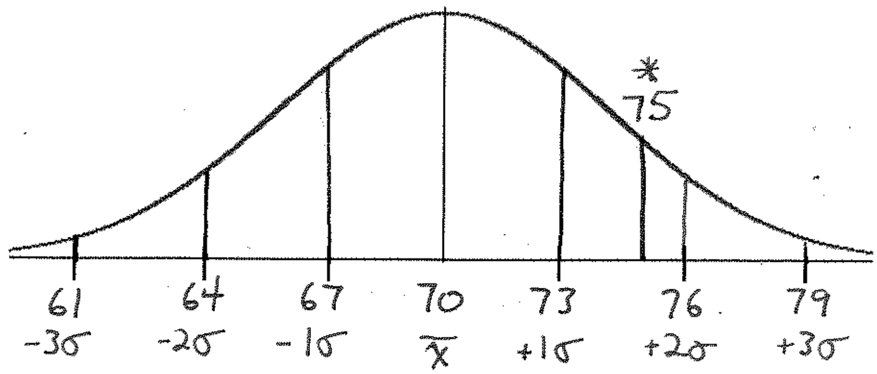
No, we don't know the exact values of the data.

g) Which has a higher value, the median of the data or the mean of the data? Explain.

Mean, b/c data is skewed right.

9.4 Notes: Normal Curves and z-scores

Ex 1: The heights of adult American males is normally distributed with a mean of 70 in and a standard deviation of 3 in. Draw the normal curve to represent the heights of adult American males.



b) Suppose you want to find the probability that an adult American male is less than 75 inches tall. Why is this problem hard to do?

75 is not an exact σ from the $\bar{x}$.

We want to find out *how many* standard deviations 75 is away from our mean. To do this, we can calculate a **z-score.**

z-score formula: $z = \dfrac{\text{value} - \text{mean}}{\text{standard deviation}} = \dfrac{x - \bar{x}}{\sigma}$

A z-score tells us how many **standard deviations we are away from the mean.**

c) Calculate the z-score for 75 inches.

$$Z_{75} = \frac{75-70}{3} = \frac{5}{3} = 1.6\overline{6} \; \boxed{\approx 1.7}$$

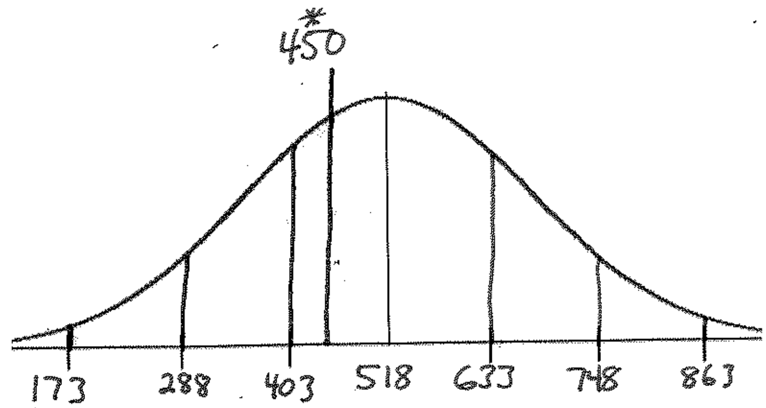
d) We can use a table of values (which will be given to you) in order to find the probability that we randomly choose someone **less than** our given score. Use the table below to find the probability that an adult American male is less than 75 inches tall.

$P(<75) = .9554$

z	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9
−2	.0228	.0179	.0139	.0107	.0082	.0062	.0047	.0035	.0026	.0019
−1	.1587	.1357	.1151	.0968	.0808	.0668	.0548	.0446	.0359	.0287
−0	.5000	.4602	.4207	.3821	.3446	.3085	.2743	.2420	.2119	.1841
0	.5000	.5398	.5793	.6179	.6554	.6915	.7257	.7580	.7881	.8159
1	.8413	.8643	.8849	.9032	.9192	.9332	.9452	.9554	.9641	.9713
2	.9772	.9821	.9861	.9893	.9918	.9938	.9953	.9965	.9974	.9981

Example 2: In 2011, SAT scores have a normal distribution with the mean score of the math portion of the SAT being 518, with a standard deviation of 115. What is the **probability** that a student scored less than 450 on the math portion of the SAT?

Step 1: Draw a sketch of the normal curve.



Step 2: Calculate the z-score.
Why is this a negative value?

$$Z_{450} = \frac{450-518}{115} = \frac{-68}{115} = -0.591 \approx -0.6$$

Step 3: Use the table below to answer the question.

$P(<450) \approx .2743$

z	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9
−2	.0228	.0179	.0139	.0107	.0082	.0062	.0047	.0035	.0026	.0019
−1	.1587	.1357	.1151	.0968	.0808	.0668	.0548	.0446	.0359	.0287
−0	.5000	.4602	.4207	.3821	.3446	.3085	.2743	.2420	.2119	.1841
0	.5000	.5398	.5793	.6179	.6554	.6915	.7257	.7580	.7881	.8159
1	.8413	.8643	.8849	.9032	.9192	.9332	.9452	.9554	.9641	.9713
2	.9772	.9821	.9861	.9893	.9918	.9938	.9953	.9965	.9974	.9981

Example 3: In 2011, the mean score of the math portion of the SAT is 518, with a standard deviation of 115. What is the **probability** that a student scored ***more than*** 450 on the math portion of the SAT?

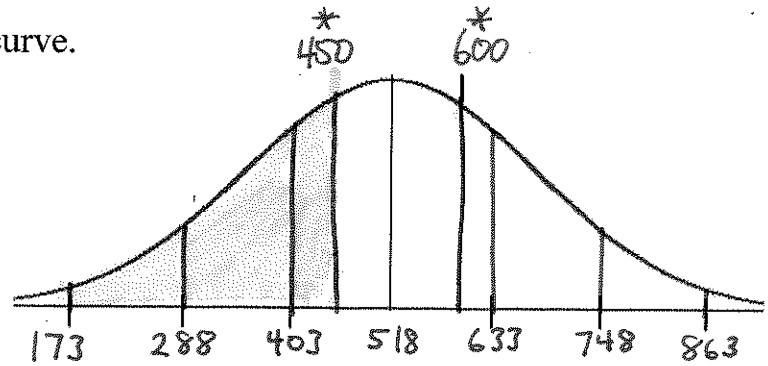
$$P(>450) = 1 - P(<450)$$

$$= 1 - .2743$$

$$\boxed{P(>450) = .7257}$$

Example 4: In 2011, SAT scores have a normal distribution with the mean score of the math portion of the SAT being 518, with a standard deviation of 115. What is the **probability** that a student scored ***between*** 450 and 600 on the math portion of the SAT?

Step 1: Draw a sketch of the normal curve.



Step 2: Find the z-score for 600.

$$z_{600} = \frac{600-518}{115} = \frac{82}{115} = 0.713 \boxed{\approx 0.7}$$

Step 3: Find the probability for **scoring less** than 600. $P(<600) \approx .7580$

z	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9
−2	.0228	.0179	.0139	.0107	.0082	.0062	.0047	.0035	.0026	.0019
−1	.1587	.1357	.1151	.0968	.0808	.0668	.0548	.0446	.0359	.0287
−0	.5000	.4602	.4207	.3821	.3446	.3085	.2743	.2420	.2119	.1841
0	.5000	.5398	.5793	.6179	.6554	.6915	.7257	.7580	.7881	.8159
1	.8413	.8643	.8849	.9032	.9192	.9332	.9452	.9554	.9641	.9713
2	.9772	.9821	.9861	.9893	.9918	.9938	.9953	.9965	.9974	.9981

Step 4: Reminder: You already know the probability for scoring less than 450, which is .2743. Now, you can subtract the probabilities for 600 and 450 to find the probability of scoring **between** those values.

$$P(450-600) = P(<600) - P(<450)$$
$$= .7580 - .2743$$
$$\boxed{\approx .4838}$$

Why does this strategy work?