

Home Run Sluggers

1.) Babe Ruth's 60 home runs in a single season in 1927 was a record that stood for 34 years as a milestone in all of sports history. Roger Maris had a remarkable breakthrough season hitting 61 home runs in 1961. Mark McGwire set a new mark at 70 home runs in 1998. Then in 2001, Barry Bonds hit 73.

Ruth: 54 59 35 41 46 25 47 60 54 46 49 46
41 34 22

Maris: 13 23 26 16 33 61 28 39 14 8

McGwire: 49 32 33 39 22 42 9 9 39 52 58 70
65 32 29

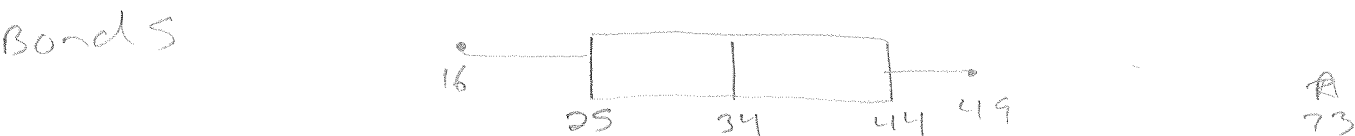
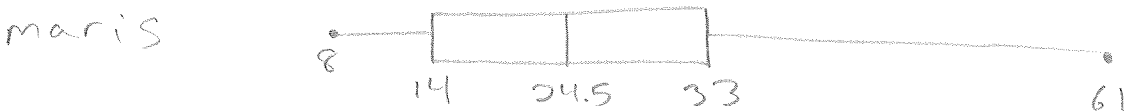
Bonds: 16 25 24 19 33 25 34 46 37 33 42 40
37 34 49 73 46

a.) Write out the five-number summaries:

Ruth:	Maris:	McGwire:	Bonds:
20, 35, 46, 54, 60	8, 14, 24.5, 33, 61	9, 29, 39, 52, 70	16, 25, 34, 44, 73

b.) Create a side-by-side box and whisker plot.

Single Season Home Run Totals



Bonds: $Q_1 = 25, Q_3 = 44, IQR = 19$

c.) Were any of their record-setting performances really unusual (outliers)? Be sure to show the math behind your answer.

outlier if... $\sigma = 73$ is an outlier for Bonds.
 $< 25 - (1.5 \cdot 19)$ or $> 44 + (1.5 \cdot 19)$
 < -3.5 or > 72.5

d.) Calculate the mean, median, mode, standard deviation, range, and IQR for Maris.

$\bar{x} = 26.1, m = 24.5, \text{mode} = \text{None}, \sigma = 14.81$

Range = 53, IQR = 19

e.) What is the 75th percentile for the Ruth graph?

54

Salaries of the Los Angeles Lakers, 2002-2003

Player: Salary	Player: Salary
Shaquille O'Neal: \$23.6 million	Derek Fisher: \$3.0 million
Kobe Bryant: \$12.4 million	Kareem Rush: \$1.0 million
Robert Horry: \$5.3 million	Brain Shaw: \$1.0 million
Rick Fox: \$3.9 million	Mark Madsen: \$0.8 million
Tracy Murray: \$3.7 million	Jannero Pargo: \$0.4 million
Devean George: \$4.6 million	Samaki Walker: \$1.5 million

2.) 2002-2003 Laker's Salaries

a.) Find the mean and median of the salaries for the Lakers players.

$\bar{x} = 5.1$ million, $m = 3.35$ million

b.) Why is the mean so much greater than the median?

The \bar{x} is impacted by Shaq and Kobe's extreme salaries.

c.) Explain a situation in which someone studying the data might want to use the mean to represent the Laker's salaries for their own benefit:

The owner. They could use it to argue they pay their players a lot of money.

d.) Explain a situation in which someone studying the data might want to use the median to represent the Laker's salaries for their own benefit:

The players or player's union. They might want to argue that they aren't paid enough.

← Normally Distributed

3.) **How Spread out are the Laker's Salaries?**

a.) Find the range and interquartile range of the Lakers salaries.

$$\text{Range} = 23.6 - 0.4 = 23.2$$

$$\text{IQR} = 4.95 - 1 = 3.95$$

b.) Are there any outliers in the data set (if so what are they)? Show your reasoning.

outlier if...

$$< 1 - (1.5 \cdot 3.95) \text{ or } > 4.95 + (1.5 \cdot 3.95)$$

$$< -4.925 \text{ or } > 10.875$$

Shaq
A and
Kobe

c.) Find the standard deviation of the Lakers salaries.

$$\sigma = 6.40 \text{ million}$$

d.) What percentage of the data values lie within one standard deviation of the mean?

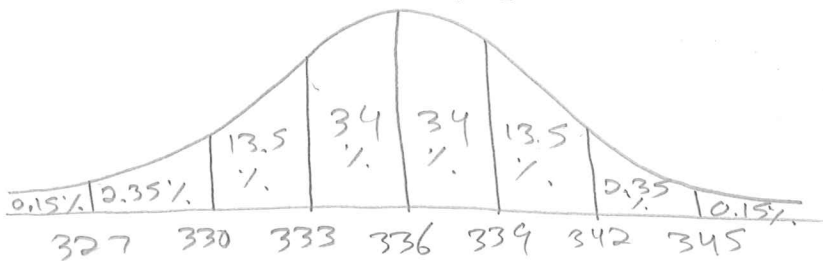
← Normal distribution

68%

4.) **Horse Pregnancies**

Bigger animals tend to carry their young longer before birth. The length of horse pregnancies from conception to birth varies according to a roughly normal distribution with mean 336 days and standard deviation 3 days. Use the 68-95-99.7 rule to answer the following questions.

a.) Almost all (99.7%) horse pregnancies fall within what range of lengths? (Draw the curve 1st)



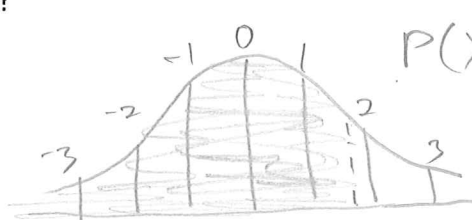
$$327 < X < 345 \text{ days}$$

b.) What percent of horse pregnancies are longer than 339 days?

16%

c.) What is the z-score for a horse pregnancy of 341 days? This pregnancy would last longer than what % of horse pregnancies?

$$z = \frac{(341 - 336)}{3} = 1.67$$

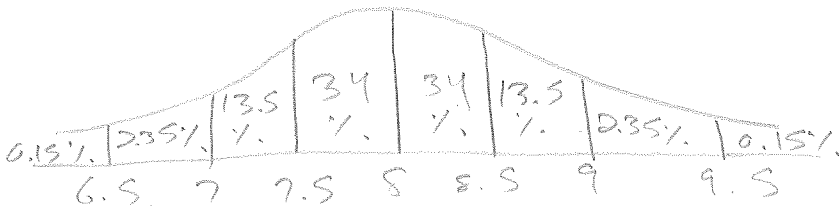


$$P(X < 341) = 0.9525 \text{ or } 95.25\%$$

5.) **Oranges**

Bags of oranges in a shipment averaged 8 pounds with a standard deviation of 0.5 pounds. A histogram of these weights followed a normal distribution quite closely. (Draw the curve 1st)

a.) Sketch the normal curve.



b.) The orange factory has 1,000 bags of oranges to ship out. How many bags weight between 7.5 pounds and 9 pounds?

81.5% # of Bags = 0.815 · 1000 = 815 bags

c.) What percent of the bag weighed more than 8.30 pounds? Be sure to sketch the curve and write the probability statement.

$z = \frac{8.3 - 8}{0.5} = 0.60$

$P(X > 8.30) = 1 - 0.7257 = 0.2743$

or 27.43%

Does this make sense?

d.) What percentile would 9 lb. be?

97.5th

e.) What value is the 16th percentile?

7.5 lbs.

f.) Out of 1,000 bags of oranges, how many would weigh between 6.8 and 8.75 pounds? Be sure to sketch the curve and write the probability statement.

$z = \frac{6.8 - 8}{0.5} = -2.40$

$z = \frac{8.75 - 8}{0.5} = 1.50$

$P(6.8 < X < 8.75) = 0.9332 - 0.0082 = 0.925$

of bags = 0.925 · 1,000 = 925 bags

g.) What would a bag's weight be if it had a z-score of -1.38?

$z = \frac{x - \bar{x}}{\sigma} \rightarrow -1.38 = \frac{x - 8}{0.5} \rightarrow -0.69 = x - 8$

7.31 = x

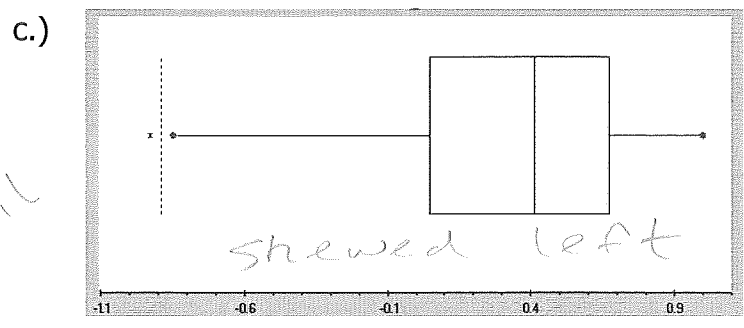
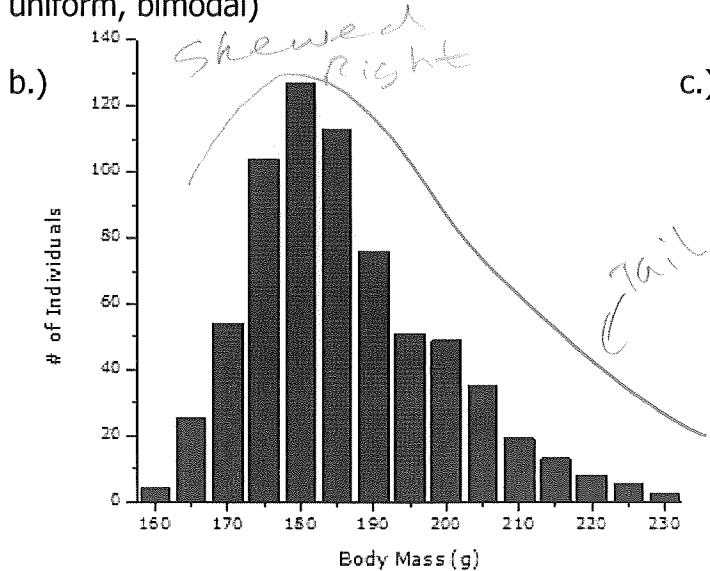
h.) If a bag has a z-score of 2.39 would it be considered an outlier? Why?

YES. Anything more than 2σ or above or below the \bar{x} is an outlier.

6.) **What's the shape?**



a.) We have collected the salaries of employees at the local ice cream parlor. We included the salary of the owner and manager in the data as well (the salaries of these members of the business is significantly higher than the rest). Draw a curve to display the salaries of the employees. What type of distribution does it follow? (symmetrical, skewed right, skewed left, uniform, bimodal)



For Exercises 7-9, state whether each scenario involves an observational study or an experimental study. If it's experimental determine the treatment.

Skip # 7-9

7.) A Tigers fan attends all of the games keeping the stats of all the players. He is trying to know as much as he possibly can about his team to show off and stake claim as the biggest Tiger fan.

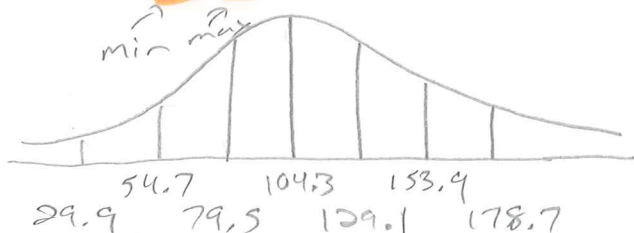
8.) The Michigan Department of Education is trying to find schools that are excelling in particular content areas in order to improve the educational system. They have decided to stop in several high schools, beginning to administer tests to various groups of students.

9.) Post Brand Cereal has decided that they make a superior raisin bran to the Kellogg Corporation raisin bran. Following in Pepsi's footsteps, they hit the road and set up the stands outside local grocery stores having customers taste test the two brands to choose a favorite.

10.) Katie keeps track of the number of phone calls received by her business each day. Within how many standard deviations of the mean do all the values fall?

84, 112, 82, 61, 150, 130, 110, 97, 94, 123

$\bar{x} = 104.3$ $\sigma = 24.8$



Between 2σ below and 2σ above the \bar{x} .

For exercises 11-13, identify what sampling method is being used in each instance. Also, if bias exists identify the type of bias. Be sure to justify your answer:

Sampling Methods: convenience, cluster, random, stratified, systematic, voluntary

Types of Bias: wording, voluntary, convenience, undercoverage. skip Bias

11.) The Romeo Observer is visiting Romeo Senior High to find out whether the students find their course load: easy, challenging, or hard. In order to get an idea of what the entire high school population thinks, they decide to find 50 female and 50 male students to survey. After pulling these students aside, they randomly survey the students.

Stratified
Random

12.) Mr. Feeny wants to survey his students to get information about what they think about the amount of homework they are given. He collects the survey back from his students that choose to fill it out and return it.

voluntary
Response

13.) In order to determine what their customers think of their wireless service, Verizon Wireless calls every 10th person from the database that contains the records of their customers.

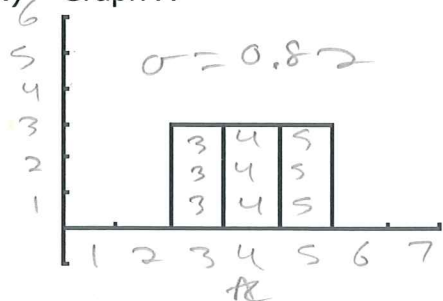
Systematic

For exercises 14-15, order each of the following from smallest to largest standard deviation:

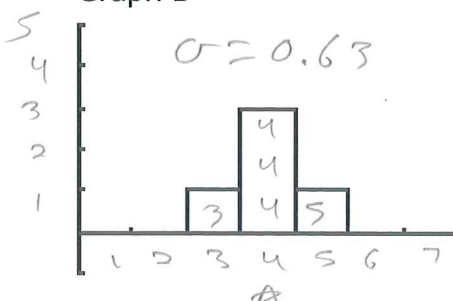
14.) Data Set F: 2, 2, 3, 3, 3, 4, 5, 7 $\sigma = 1.60$
 Date Set G: 4, 4, 5, 6, 6, 7, 9 $\sigma = 1.64$
 Data Set H: 3, 3, 3, 3, 3, 3, 7 $\sigma = 1.40$

H → F → G

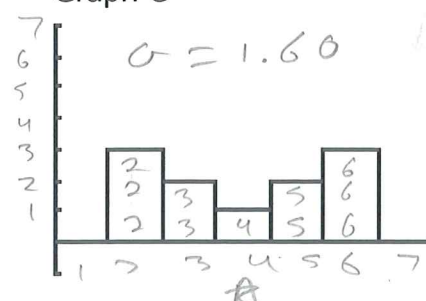
15.) Graph A



Graph B



Graph C



B → A → C

16.) Alex is trying to get information on how his fellow classmates feel about allowing juniors and seniors to attend prom.

Alex's Survey Question:

Should the school allow juniors to attend prom even though it has always been a special day for just the seniors?

a.) Explain why his survey question may be biased

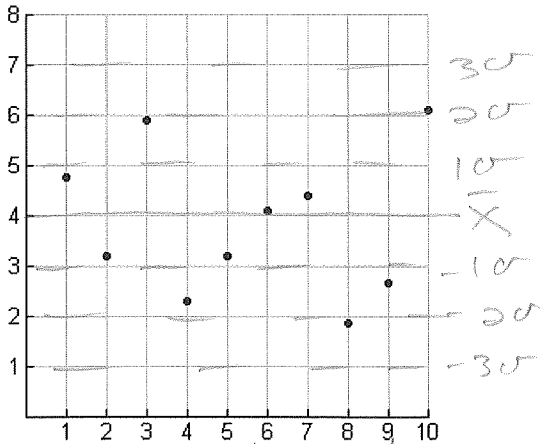
The wording "even though it has always been a special day for just the seniors" is unnecessary.

b.) Rewrite the survey question in an un-biased way.

Should the school allow juniors to attend prom?

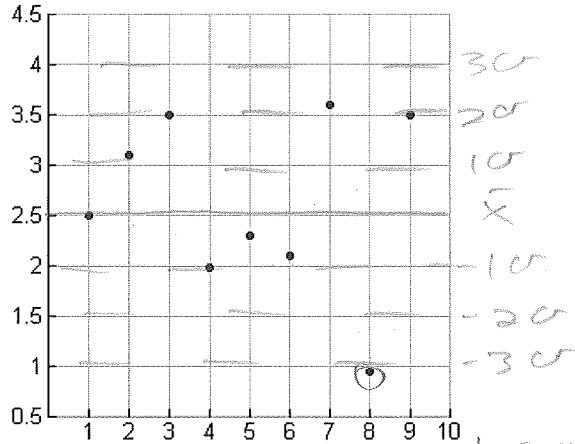
17.) Determine whether each system is under control. If out of control state the rule it breaks. Label the mean and every standard deviation above and below the mean on the control chart.

a.) Mean=4, Standard Deviation=1



In control

b.) Mean=2.5, Standard Deviation=0.5



Out of control

Breaks Rule #1

- Rule #1:**
One point is outside the "control limits".
(Outside of three standard deviations)
- Rule #2:**
Two out of three points in a row are between 2 and 3 standard deviations on the same side of the mean.
- Rule #3:**
Four out of five points in a row are between 1 and 2 standard deviations on the same side of the mean.
- Rule #4:**
Eight in a row are on the same side of the mean.

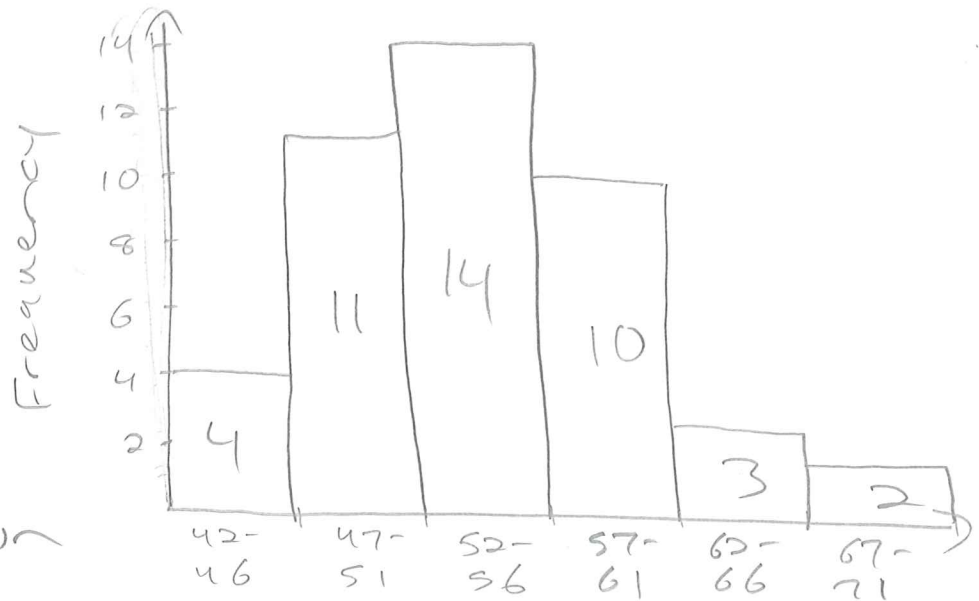
$$\text{Add} = \frac{\text{Range}}{7} = \frac{(69-42)}{7} = \frac{27}{7} = 3.86 \approx 4$$

If you want 7-intervals

18.) The following list contains the ages of Presidents (Washington → Obama) at inauguration:
 Ages: 57, 61, 57, 57, 58, 57, 61, 54, 68, 51, 49, 64, 50, 48, 65, 52, 56, 46, 54, 49, 51, 47, 55, 55, 54, 42, 51, 56, 55, 51, 54, 51, 60, 61, 43, 55, 56, 61, 52, 69, 64, 46, 54, 47
 Create a frequency histogram with at least 5 bars. How would a relative frequency histogram differ?

Intervals	Frequency
42-46	4
47-51	11
52-56	14
57-61	10
62-66	3
67-71	2

Ages of Presidents at Inauguration



relative Frequency has percentages on the y-axis.

19.) Below you are given the total number of points scored by the New England Patriot and Indianapolis Colt football teams during the 2010 football season.

Patriots: 17, 19, 13, 30, 10, 41, 23, 15, 20, 31, 48, 10, 24, 49, 47, 13

Colts: 13, 18, 21, 31, 31, 14, 21, 18, 24, 33, 23, 10, 35, 31, 31, 23

Create an ordered side-by-side stem and leaf plot for the data. **Be sure to include a title and legend!**

skip