MATH 1030: Unit 3 Study Guide Name: _____

Chapter 5 Statistical Reasoning

5A Fundamentals of Statistics

Common Sampling Methods: (page 290)

1. You want to determine the typical dietary habits of students at a college. Which of the following would make the best sample, and why? Also explain why each of the other choices would *not* make a good sample for this study.

a) Students in a single dormitory

b) Students majoring in public health

c) Students who participate in intercollegiate sports

d) Students who enrolled in a required mathematics class

Identify which of these types of sampling is used.

2. A market researcher selects 500 people from each of 10 cities.

A) Convenience B) Systematic C) Stratified D) Random

3. Identify the Sampling Method (simple random sampling, systematic sampling, convenience sampling, or stratified sampling) of the following studies:

a. A taste test for chips and salsa is conducted at the entrance to a supermarket.

b. A computer randomly calls 500 names from a list of all residents of Smallville. Those selected are surveyed to predict who will win the election for city councilman.

c. People magazine chooses its "25 best-dressed celebrities" by looking at responses from readers who voluntarily mail in a survey printed in the magazine.

d. A principal takes an alphabetized list of student names and picks a random starting point. Every 20^{th} student is selected to take a survey about the effects of class size on student grades.

Two types of a Statistical Study: Treatment and Control Groups: Placebo Effect: (page 292-293)

4. Determine whether the following studies are observational studies or experiments. If the study is an experiment, identify the control and treatment groups, and discuss whether single-or double-binding is necessary. If the study is observational, state whether it is a case-control study, and if so, identify the cases and controls.
    a. A National Cancer Institute study of 716 melanoma patients and 1014 cancer-free patients matched by age, sex, and race found that those having a single large mole had twice the risk of melanoma. Having 10 or more moles was associated with a 12 times greater risk of melanoma.

    b. A double-blind drug versus placebo study of 80 patients suffering from tendonitis (the inflammation of a tendon) demonstrated the effectiveness of a non-steroidal anti-inflammatory drug. The non-steroidal treatment improved the condition of all the tendonitis patients.

5B Should you believe a statistical study?

Eight guidelines for evaluating a statistical study: (page 301)

5. Based solely on the information given, do you have reason to question the results of the following hypothetical studies? Explain your reasoning.
    a) A study financed by a major pharmaceutical company is intended to determine whether its new cholesterol drug is more effective than similar drugs of competing companies.

    b) A government study is designed to determine the percentage of taxpayers who understate their income, based on people who had their tax returns audited.

6. Identify at least one potential source of bias in the following studies. Explain why the bias would or would not affect your view of the study.
   a) An article in *Journal of Nutrition* noted that chocolate is rich in flavonoids. This article reports that "regular consumption of foods rich in flavonoids may reduce the risk of coronary heart disease." The study received funding from Mars, Inc. the candy company, and the Chocolate Manufacturers Association.

   b) A study (in the Canadian Medical Association Journal) of 20 nations discovered that Germany has the highest number of average (mean) annual visits to a doctor (8.5), while Finland has the fewest (3.2)

5C Statistical Tables and Graphs

7. Amazon collects rating for their products sold online. The ratings of a dog toy on are listed below.

5	3	3	3	2	1	5	4	3	1
4	2	1	5	3	2	3	4	1	5
1	3	2	2	5	4	3	1	4	3

Construct a frequency table for the ratings:

Grade	Frequency	Relative Frequency (Round 2 decimal places)
5		
4		
3		
2		
1		
Total		

8. The following table shows the average (mean) annual salaries of public school teachers for each state and the District of Columbia. Make a histogram of the data.

Average Teacher Salaries by State	
Avg Annual Salary	Number of States
<$40,000	1
$40,000–$45,000	1
$45,000–$50,000	25
$50,000–$55,000	7
$55,000–$60,000	9
$60,000–$65,000	4
>$65,000	4

9. Construct pie charts for the following data set. The first step is to compute percentage for each category in the data set.

The five leading tourist destinations (in millions of visitors) are shown in the table.

Country	Visitors (millions)
France	76.8
U.S.	59.7
China	55.7
Spain	52.7
Italy	43.6

Determine whether each variable if qualitative (categorical) or quantitative.

10. Write your answer in the blank provided.

a) Score on a math placement exam. _____

b) A student's favorite type of music. _____

c) The zip code that a person lives in. _____

d) Fat content in a bag of potato chips. _____

e) The gas mileage on a car model. _____

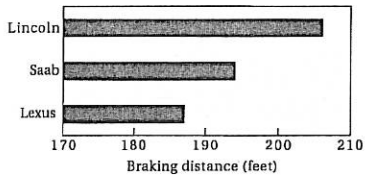
11. The following table shows the revenue (in millions of dollars) of the leading U.S. food and drug retailers in 2011.
    a. State whether the variables are qualitative or quantitative.

    b. Draw a bar graph or a pie chart if the data are qualitative. Draw a histogram or a line chart if the data are quantitative.

Advertiser	Revenue ($ billions)
CVS Caremark	96.4
Kroger	82.2
Walgreens	67.4
Safeway	41.0
Supervalu	40.6
Rite Aid	25.7

5D Graphics in the Media

12. The figure shows the braking distance for three different cars. Discuss the ways in which it might be deceptive. How much greater is the braking distance of a Lincoln than the breaking distance of a Lexus? Draw the distance in a fairer way.



5E Correlation and Causality

13. Consider the following statements about a correlation. In each case, state the correlation clearly (for example, there is a positive correlation between variable A and variable B). Then state whether the correlation is most likely due to coincidence, a common underlying cause, or a direct cause. Explain your answer.

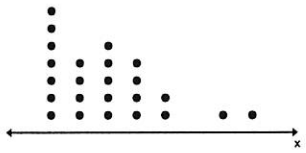
    a. In a large resort city, the crime rate increased as the number of taxi cabs increased.

    b. Automobile gas mileage decreases with tire pressure.

Chapter 6 Putting Statistics to Work
6A Characterizing Data

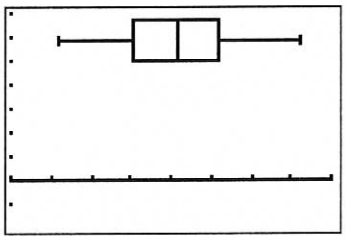
14. For each of the following graphs, Determine the shape of the data (symmetrical, skewed left, or skewed right). Then Determine if the mean is higher, lower, or the same as the median. Determine whether the mean or the median be the best measure of center.

a)



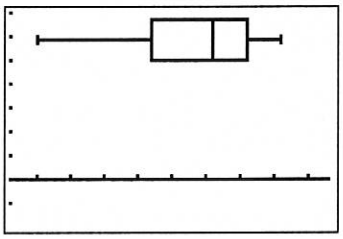
Shape: _____
Mean _____ Median
Best Measure of Center: _____

b)



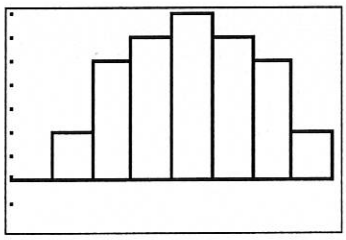
Shape: _____
Mean _____ Median
Best Measure of Center: _____

c)



Shape: _____
Mean _____ Median
Best Measure of Center: _____

d)



Shape: _____
Mean _____ Median
Best Measure of Center: _____

15. Compute the mean, median, and mode of the following data set.

Body temperature (in degrees Fahrenheit) of randomly selected normal and healthy adults.

98.6	98.6	98.0	98.0	99.0
98.4	98.4	98.4	98.4	98.6

16. The following data give the margin of victory in the Super Bowl from 2003-2013

27 3 3 11 12 3 4 14 6 4 3

a. Find the mean and median margin of victory.

b. Identify any outlier(s) in the set. If you eliminate the outlier(s), what are the new mean and median?

6B Measures of Variation

17. A nurse works in the maternity ward at the Healing Hospital. During one shift at the Healing hospital, she records the following weights (in pounds) of newborn babies.

Healing	6.6	7.3	7.6	7.8	7.8	8.6	8.9

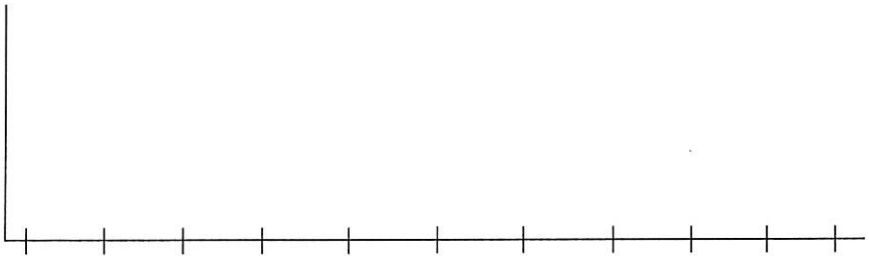
a) Find the mean, mode and range for the newborn babies at the **Healing Hospital**.

Mean = _____ Mode = _____ Range = _____

b) Find the 5-number summary.

Min = _____ Q1 = _____ Q2 = _____ Q3 = _____ Max = _____

c) Draw a box and whiskers plot of the new born babies from both hospitals.



Healing Hospital

18. A small animal veterinarian reviews her records for the day and notes that she has seen eight dogs and eight cats with the following weights (in pounds):

Dogs: 12, 18, 26, 33, 41, 56, 74, 109

Cats: 4, 5, 9, 9, 12, 15, 21, 22

$$\text{standard deviation} = \sqrt{\frac{\text{sum of (deviations from the mean)}^2}{\text{total number of data values} - 1}}$$

$$\text{standard deviation} \approx \frac{\text{range}}{4}$$

a) Compute the mean and median for each set.

b) Find the standard deviation for each of the two data sets

c) Apply the range rule of thumb to estimate the standard deviation of each of the two data sets. How well does the rule work in each case? Briefly discuss why it does or does not work well.

6C The Normal Distribution

Use the 68-95-99.7 Rule to answer the following.

19. The number of minutes per day spent talking on a cell phone by a sample of 412 students is normally distributed with a mean of 41 minutes and a standard deviation of 14 minutes.

a) What percentage of the scores are below 55?

b) What percentage of the scores are above 69?

c) Approximately how many students spend less than 27 minutes talking on their cell phones?

d) Approximately how many students spend more than 69 minutes talking on their cell phones?

e) What is the probability that a randomly selected student will spend less than 13 minutes per day talking on their cell phone?

20. In the United States the average weight of a newborn baby is 7.2 pounds with a standard deviation of 1.1 pounds.

a) Determine what a baby would weigh to be considered in the 90^{th} percentile.

b) Determine what a baby would weigh to be considered in the 15^{th} percentile.

c) Describe what it means for a baby to be in the 10^{th} percentile.

Name:

MATH 1030: Unit 3 Study Guide

Chapter 5 Statistical Reasoning

5A Fundamentals of Statistics

Common Sampling Methods: (page 290)

1. You want to determine the typical dietary habits of students at a college. Which of the following would make the best sample, and why? Also explain why each of the other choices would *not* make a good sample for this study.

a) Students in a single dormitory NO - students in a dorm have diff. eating habits,

b) Students majoring in public health NO - majoring in public health. have more education about dietary habits

c) Students who participate in intercollegiate sports NO - athletes tend to eat better

d) Students who enrolled in a required mathematics class yes not biased towards any sub groups

Identify which of these types of sampling is used.

2. A market researcher selects 500 people from each of 10 cities.

A) Convenience B) Systematic C) Stratified D) Random

3. Identify the Sampling Method (simple random sampling, systematic sampling, convenience sampling, or stratified sampling) of the following studies:

a. A taste test for chips and salsa is conducted at the entrance to a supermarket.

Convenience

b. A computer randomly calls 500 names from a list of all residents of Smallville. Those selected are surveyed to predict who will win the election for city councilman.

Simple random

c. People magazine chooses its "25 best-dressed celebrities" by looking at responses from readers who voluntarily mail in a survey printed in the magazine.

Convenience

d. A principal takes an alphabetized list of student names and picks a random starting point. Every 20th student is selected to take a survey about the effects of class size on student grades.

systematic

Two types of a Statistical Study: Treatment and Control Groups: Placebo Effect: (page 292-293)

4. Determine whether the following studies are observational studies or experiments. If the study is an experiment, identify the control and treatment groups, and discuss whether single-or double-binding is necessary. If the study is observational, state whether it is a case-control study, and if so, identify the cases and controls.
   a. A National Cancer Institute study of 716 melanoma patients and 1014 cancer-free patients matched by age, sex, and race found that those having a single large mole had twice the risk of melanoma. Having 10 or more moles was associated with a 12 times greater risk of melanoma.

observational: case-control
cases are patients with melanoma
controls: are patients who are cancer-free

   b. A double-blind drug versus placebo study of 80 patients suffering from tendonitis (the inflammation of a tendon) demonstrated the effectiveness of a non-steroidal anti-inflammatory drug. The non-steroidal treatment improved the condition of all the tendonitis patients.

Experiment:
treatment group: received the drug
control group: received placebo
double-blind used

5B Should you believe a statistical study?

Eight guidelines for evaluating a statistical study: (page 301)

5. Based solely on the information given, do you have reason to question the results of the following hypothetical studies? Explain your reasoning.
   a) A study financed by a major pharmaceutical company is intended to determine whether its new cholesterol drug is more effective than similar drugs of competing companies.

You should doubt the results because of possible bias, company is bias to their own product.

   b) A government study is designed to determine the percentage of taxpayers who understate their income, based on people who had their tax returns audited.

There is no reason to doubt the study.

6. Identify at least one potential source of bias in the following studies. Explain why the bias would or would not affect your view of the study.
   a) An article in *Journal of Nutrition* noted that chocolate is rich in flavonoids. This article reports that "regular consumption of foods rich in flavonoids may reduce the risk of coronary heart disease." The study received funding from Mars, Inc. the candy company, and the Chocolate Manufacturers Association.

Potential bias because of funding source

   b) A study (in the Canadian Medical Association Journal) of 20 nations discovered that Germany has the highest number of average (mean) annual visits to a doctor (8.5), while Finland has the fewest (3.2)

a well respected journal would have no reason to distort data that is readily available.

5C Statistical Tables and Graphs

7. Amazon collects rating for their products sold online. The ratings of a dog toy on are listed below.

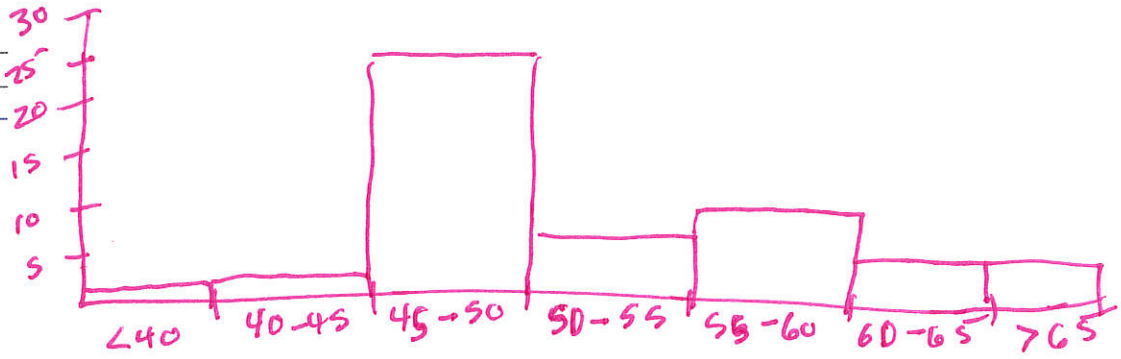
5	3	3	3	2	1	5	4	3	1
4	2	1	5	3	2	3	4	1	5
1	3	2	2	5	4	3	1	4	3

Construct a frequency table for the ratings:

Grade	Frequency	Relative Frequency (Round 2 decimal places)
5	𝍸 5	5/30 ≈ 0.17
4	𝍸 5	5/30 ≈ 0.17
3	𝍸 𝍸 9	9/30 = 0.3
2	𝍸 5	5/30 ≈ 0.17
1	𝍸 1 6	6/30 ≈ 0.2
Total	30	

8. The following table shows the average (mean) annual salaries of public school teachers for each state and the District of Columbia. Make a histogram of the data.

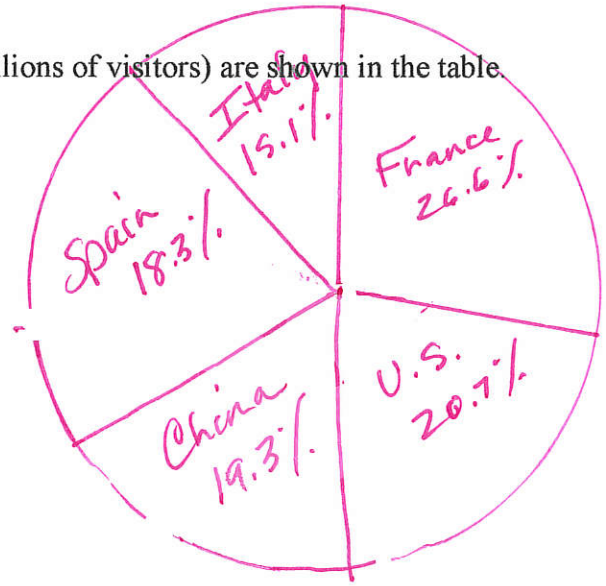
Average Teacher Salaries by State	
Avg Annual Salary	Number of States
<$40,000	1
$40,000–$45,000	1
$45,000–$50,000	25
$50,000–$55,000	7
$55,000–$60,000	9
$60,000–$65,000	4
>$65,000	4



9. Construct pie charts for the following data set. The first step is to compute percentage for each category in the data set.

The five leading tourist destinations (in millions of visitors) are shown in the table.

Country	Visitors (millions)	
France	76.8	26.6%
U.S.	59.7	20.7%
China	55.7	19.3%
Spain	52.7	18.3%
Italy	43.6	15.1%



Determine whether each variable if qualitative (categorical) or quantitative.

10. Write your answer in the blank provided.

a) Score on a math placement exam. quantitative

b) A student's favorite type of music. qualitative

c) The zip code that a person lives in. qualitative

d) Fat content in a bag of potato chips. quantitative

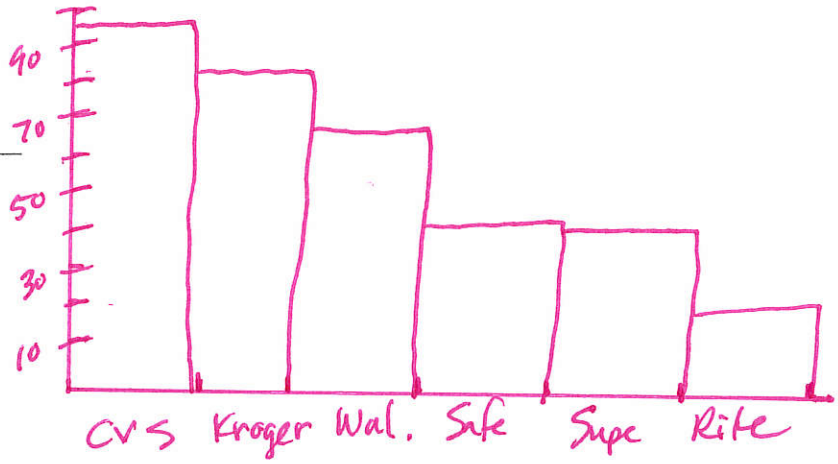
e) The gas mileage on a car model. quantitative

11. The following table shows the revenue (in millions of dollars) of the leading U.S. food and drug retailers in 2011.

a. State whether the variables are qualitative or quantitative.

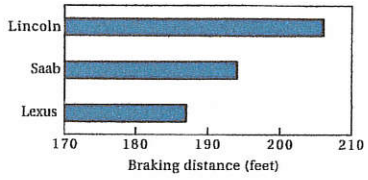
b. Draw a bar graph or a pie chart if the data are qualitative. Draw a histogram or a line chart if the data are quantitative.

Advertiser	Revenue ($ billions)
CVS Caremark	96.4
Kroger	82.2
Walgreens	67.4
Safeway	41.0
Supervalu	40.6
Rite Aid	25.7



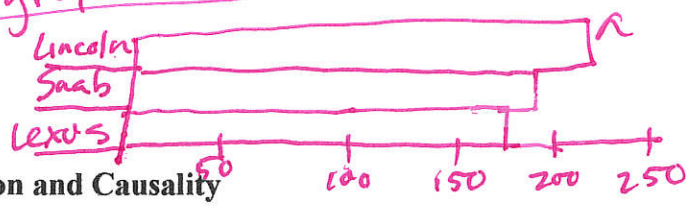
5D Graphics in the Media

12. The figure shows the braking distance for three different cars. Discuss the ways in which it might be deceptive. How much greater is the braking distance of a Lincoln than the breaking distance of a Lexus? Draw the distance in a fairer way.



The breaking distance of a Lexus is about 10% greater than the breaking distance of a Lexus.

start graph at zero



5E Correlation and Causality

13. Consider the following statements about a correlation. In each case, state the correlation clearly (for example, there is a positive correlation between variable A and variable B). Then state whether the correlation is most likely due to coincidence, a common underlying cause, or a direct cause. Explain your answer.

a. In a large resort city, the crime rate increased as the number of taxi cabs increased.

positive correlation between crime rate and # of cabs. Common underlying cause

b. Automobile gas mileage decreases with tire pressure.

gas mileag is a positively correlated to tire pressure

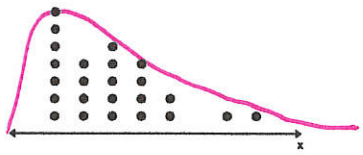
Direct Cause

Chapter 6 Putting Statistics to Work

6A Characterizing Data

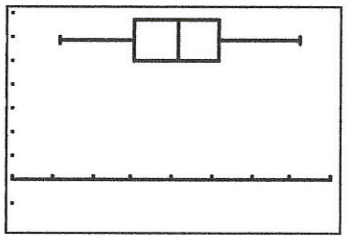
14. For each of the following graphs, Determine the shape of the data (symmetrical, skewed left, or skewed right). Then Determine if the mean is higher, lower, or the same as the median. Determine whether the mean or the median be the best measure of center.

a)



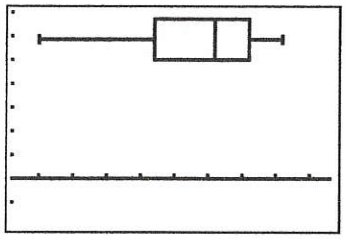
Shape: right-skewed
Mean > Median
Best Measure of Center: median

b)



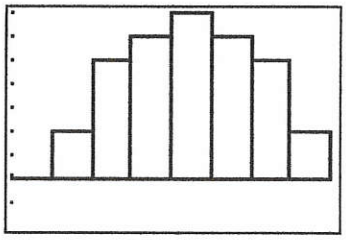
Shape: Symmetric
Mean = Median
Best Measure of Center: mean

c)



Shape: left-skew
Mean < Median
Best Measure of Center: median

d)



Shape: Symmetric
Mean = Median
Best Measure of Center: mean

15. Compute the mean, median, and mode of the following data set.

Body temperature (in degrees Fahrenheit) of randomly selected normal and healthy adults.

98.6	98.6	98.0	98.0	99.0
98.4	98.4	98.4	98.4	98.6

mean = 98.44°
median = 98.4°
Mode = 98.4

16. The following data give the margin of victory in the Super Bowl from 2003-2013

27	3	3	11	12	3	4	14	6	4	3

a. Find the mean and median margin of victory.

mean = 8.18 median = 4

b. Identify any outlier(s) in the set. If you eliminate the outlier(s), what are the new mean and median?

outlier is 27
mean = 6.3 median: 4

6B Measures of Variation

17. A nurse works in the maternity ward at the Healing Hospital. During one shift at the Healing hospital, she records the following weights (in pounds) of newborn babies.

Healing	6.6	7.3	7.6	7.8	7.8	8.6	8.9

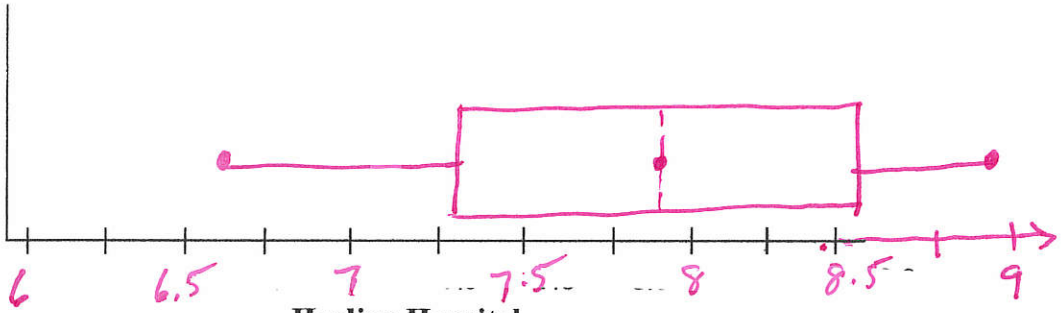
a) Find the mean, mode and range for the newborn babies at the **Healing Hospital**.

Mean = 7.8 Mode = 7.8 Range = 2.3

b) Find the 5-number summary.

Min = 6.6 Q1 = 7.3 Q2 = 7.8 Q3 = 8.6 Max = 8.9

c) Draw a box and whiskers plot of the new born babies from both hospitals.



Healing Hospital

18. A small animal veterinarian reviews her records for the day and notes that she has seen eight dogs and eight cats with the following weights (in pounds):

Dogs: 12, 18, 26, 33, 41, 56, 74, 109

Cats: 4, 5, 9, 9, 12, 15, 21, 22

$$\text{standard deviation} = \sqrt{\frac{\text{sum of (deviations from the mean)}^2}{\text{total number of data values} - 1}}$$

$$\text{standard deviation} \approx \frac{\text{range}}{4}$$

a) Compute the mean and median for each set.

Dogs: mean = 46.1 median = 37 Cats: mean: 12.1 median: 10.5

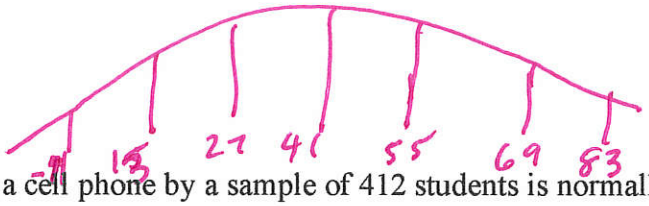
b) Compute the standard deviation of each set.

Dogs SD: 32.5 Cats SD: 6.8

d) Apply the range rule of thumb to estimate the standard deviation of each of the two data sets. How well does the rule work in each case? Briefly discuss why it does or does not work well.

Dogs $\frac{97}{4} = 24.25$ Cats $\frac{18}{4} = 4.5$ Not for dogs because of outlier ok for Cats

6C The Normal Distribution
Use the 68-95-99.7 Rule to answer the following.



19. The number of minutes per day spent talking on a cell phone by a sample of 412 students is normally distributed with a mean of 41 minutes and a standard deviation of 14 minutes.

a) What percentage of the scores are below 55?

84%

b) What percentage of the scores are above 69?

25%

c) Approximately how many students spend less than 27 minutes talking on their cell phones?

$.16(412) = 65.92$ (65)

d) Approximately how many students spend more than 69 minutes talking on their cell phones?

$.025(412) = 10.3$ (10)

e) What is the probability that a randomly selected student will spend less than 13 minutes per day talking on their cell phone?

2.5%

20. In the United States the average weight of a newborn baby is 7.2 pounds with a standard deviation of 1.1 pounds.

a) Determine what a baby would weigh to be considered in the 90th percentile.

$.90 = 1.28$ $\quad 1.28 = \frac{x-7.2}{1.1}$ (8.6 lbs)

b) Determine what a baby would weigh to be considered in the 15th percentile.

$.15 = -1.04$ $\quad -1.04 = \frac{x-7.2}{1.1}$ (6.1 lbs)

c) Describe what it means for a baby to be in the 10th percentile.

only 10% of the babies weigh less