Statistics **Chapter 1 Review**

Determine if each statement is a parameter or statistic.

1. The average credit card debt of 100 families sampled in Ohio is $12,375. Statistic

2. In a study of all 2147 parolees in a city, 392 had violated parole in the last 6 months. parameter

3. The number of home runs hit by all Major League baseball players in the 2005 season was 5017. (Baseball Almanac) parameter

4. In a recent survey of 2000 school-age children, it was found that they spend 6.5 hours per day watching TV, using computers, or enjoying other electronic devices. Statistic

For the questions 5 and 6, identify the (a) sample and the (b) population from each study.

5. A survey of 2850 smokers found that 78% had tried quitting smoking at least once in the last year.

(a) 2850 smokers

(b) All smokers

6. A survey 100 patients in a hospital ER showed the average wait time was 36 minutes.

(a) 100 patients in the ER

(b) All patients in the ER

Determine if the data below is:
A) qualitative or quantitative?
B) at the nominal, ordinal, interval or ratio level of measurement?

7. The rating of a restaurant, on a scale from 1 to 5. (1 being terrible, 5 being great)

A) Qualitative B) Ordinal

8. A list of states you have visited.

A) Qualitative B) Nominal

9. The number of chicken wings John can eat.

A) Quantitative B) Ratio

10. The years in which all of your teachers were born.

A) Quantitative B) Interval

Determine if the data below consists of qualitative or quantitative data.

11. The salaries of football players on a professional football team. Quantitative

12. Favorite breed of dog among 30 second-graders. Qualitative

Identify which data collection method is best for each situation below: *observation, experiment, simulation, or survey.*

13. A study on the effect of eliminating a turn lane on a major road to examine if it increases or decreases the number of car crashes. Simulation

14. A study on how often a child puts toys in their mouth. Observation

15. A study of student ratings about school lunches. Survey

16. A study of the effect of eating oatmeal every day to see if it lowers blood pressure.
Experiment

For #7 – 21, identify which types of sampling are being used below: *random, systematic, convenience, stratified, or cluster.*

17. Arrange all students in school in alphabetical order, and then choose every 50th person to participate in a survey.
Systematic

18. To do a survey on smoking, a sample of 500 people is divided into age groups (10-19, 20-29, 30-39, etc). Then 15 people from each age group are given the survey.
Some from all → stratified

19. A city needs to collect data on services used. Households in the city are numbered from 1 to 45,000. A computer is used to generate random numbers to choose 150 households to participate in this survey.
Random

20. A researcher stands outside Target to survey shoppers about their spending habits.
Convenience

21. The office of the Mayor of New York City wants to know what residents think about the subway system. The city has 176 zip codes. A survey of all the residents in 12 zip codes is conducted.
All from some → Cluster

22. Fifty people who walked in the 3-Day Breast Cancer walk are selected at random and their ages are recorded.

A. Is the data qualitative or quantitative? Quantitative

B. What level (nominal, ordinal, interval, ratio) of measurement are the data? Ratio

C. What is the population? All people who walked in the 3 day walk

23. Give an example of a convenience sample that is biased.

You go to a winery to ask if a person drinks wine.

24. Give an example of an experiment that has a confounding variable.

You stop drinking pop and now walk to school. You lose weight. Which is more effective for weight loss?

25. How is a data value at the interval level different from a data value at the ratio level?

Data at the interval level can be meaningfully added, subtracted, and ordered. Ratio level data can do that, too, but can also be multiplied or divided meaningfully. Ratio level data also has an inherent zero.

26. Use the information "In a sample of 350 college seniors, 318 had applied to college by February of their senior year."

A) What is the descriptive branch?

318 out of 350 seniors applied for college

B) Make an inference about the sample:

Most seniors applied for college by February

27. What is replication and why is it important?

Replication is the ability of an experiment to be repeated many, many times, while still getting the same results. It is important to show that the results are reliable (not a fluke) and the study is valid.

28. What methods are used to ensure that subjects are randomly assigned to a treatment/placebo group?

Blinding, and double blinding can be used to guard against the placebo and Hawthorne effects

p. 117-118 7-12

Review of 2.2

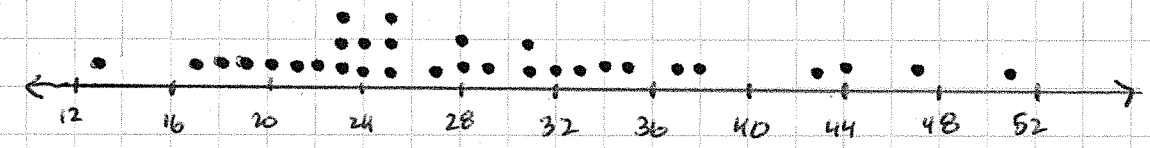
Daily High Temperature, January in Chicago

(7)

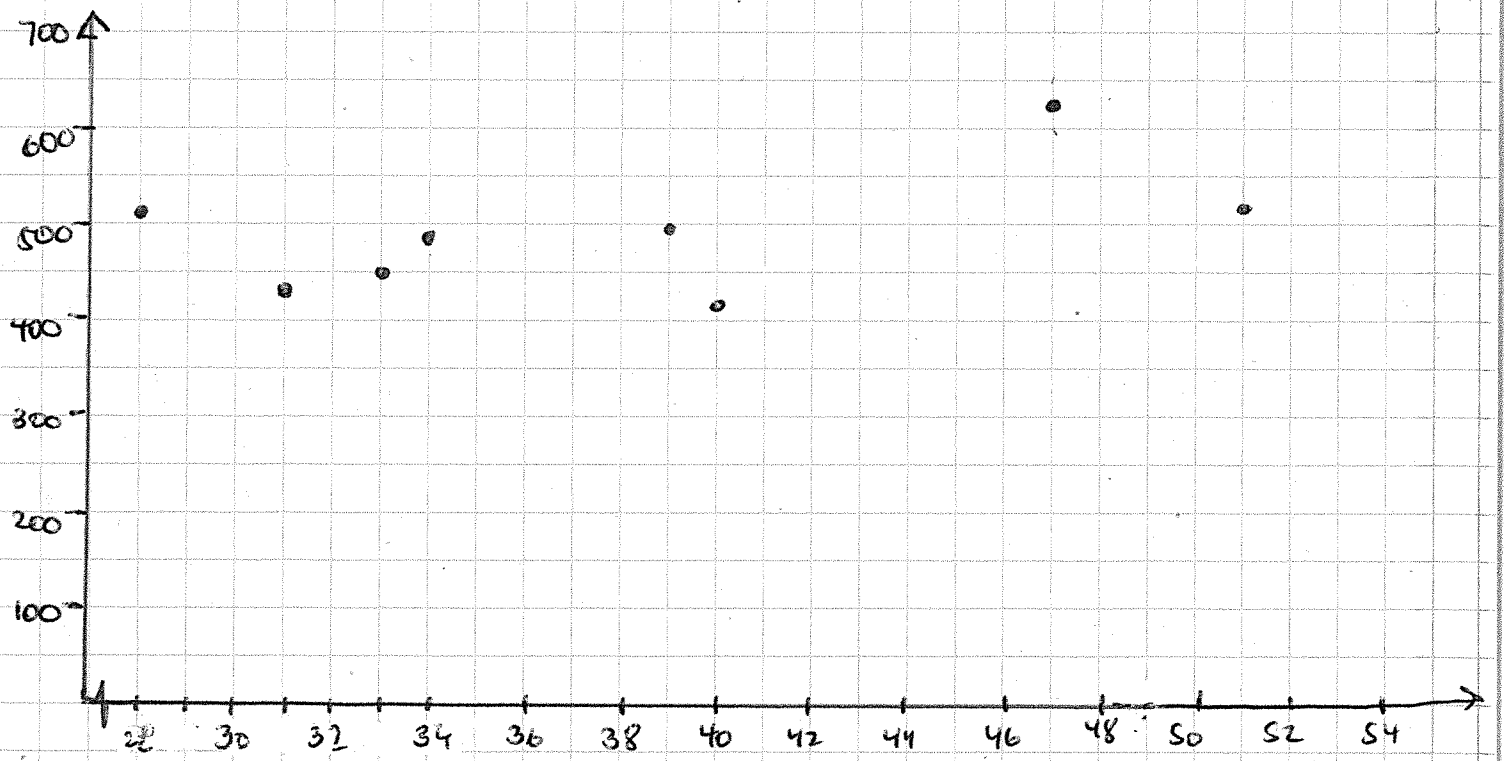
1	3 7 8 9
2	0 1 2 3 3 3 4 4 5 5 5 7 8 8 9
3	1 1 2 3 4 5 7 8
4	3 4 7
5	1

Key 3|1 = 31 °F

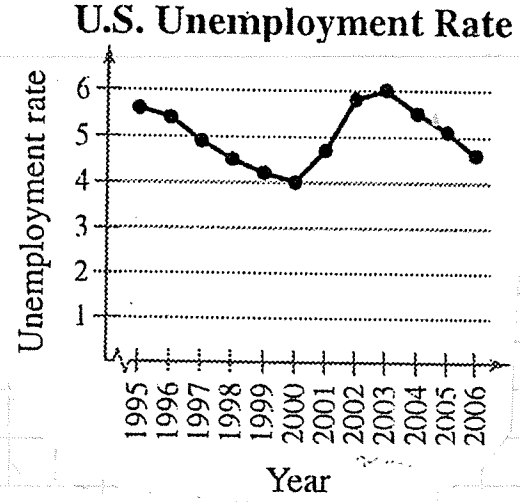
(8) Daily High Temps in January in Chicago



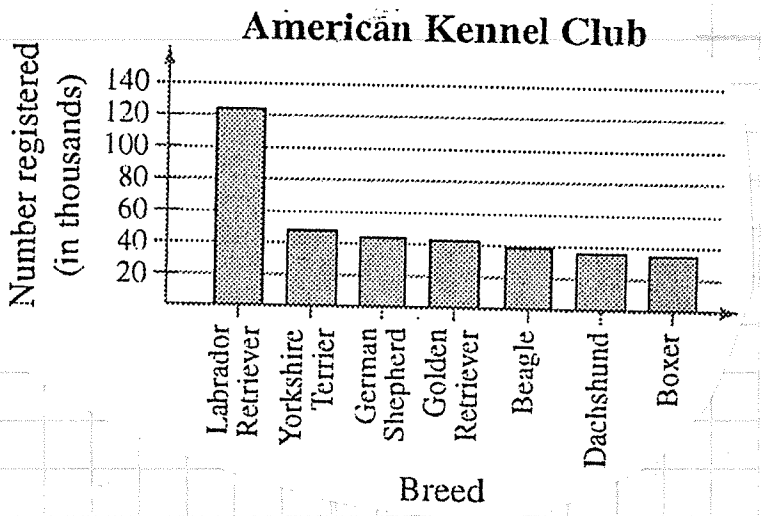
(9)



10.

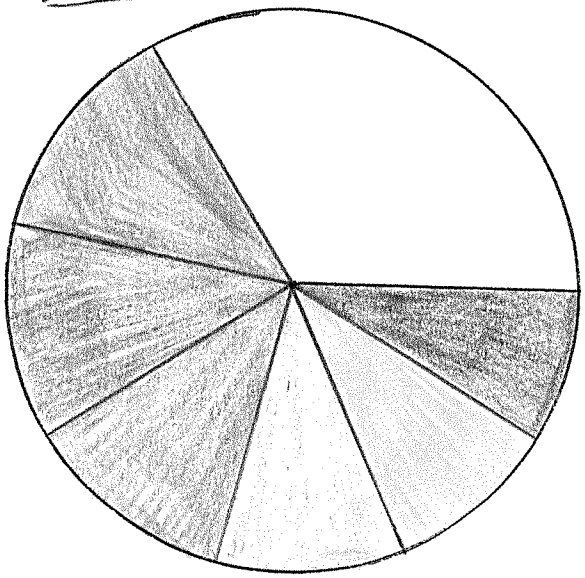

U.S. Unemployment Rate
Unemployment rate
6
5
4
3
2
1
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
Year

11.


American Kennel Club
Number registered (in thousands)
140
120
100
80
60
40
20
Labrador Retriever
Yorkshire Terrier
German Shepherd
Golden Retriever
Beagle
Dachshund
Boxer
Breed

#12 p. 118

2006 American Kennel Club Registrations



Key

- Labrador Retrievers
- Yorkshire Terriers
- German Shepards
- Golden Retrievers
- Beagle
- Dachshunds
- Boxers

Labs $\frac{124}{369}(360°) \approx 121°$

Yorkies $\frac{48}{369}(360°) \approx 47°$

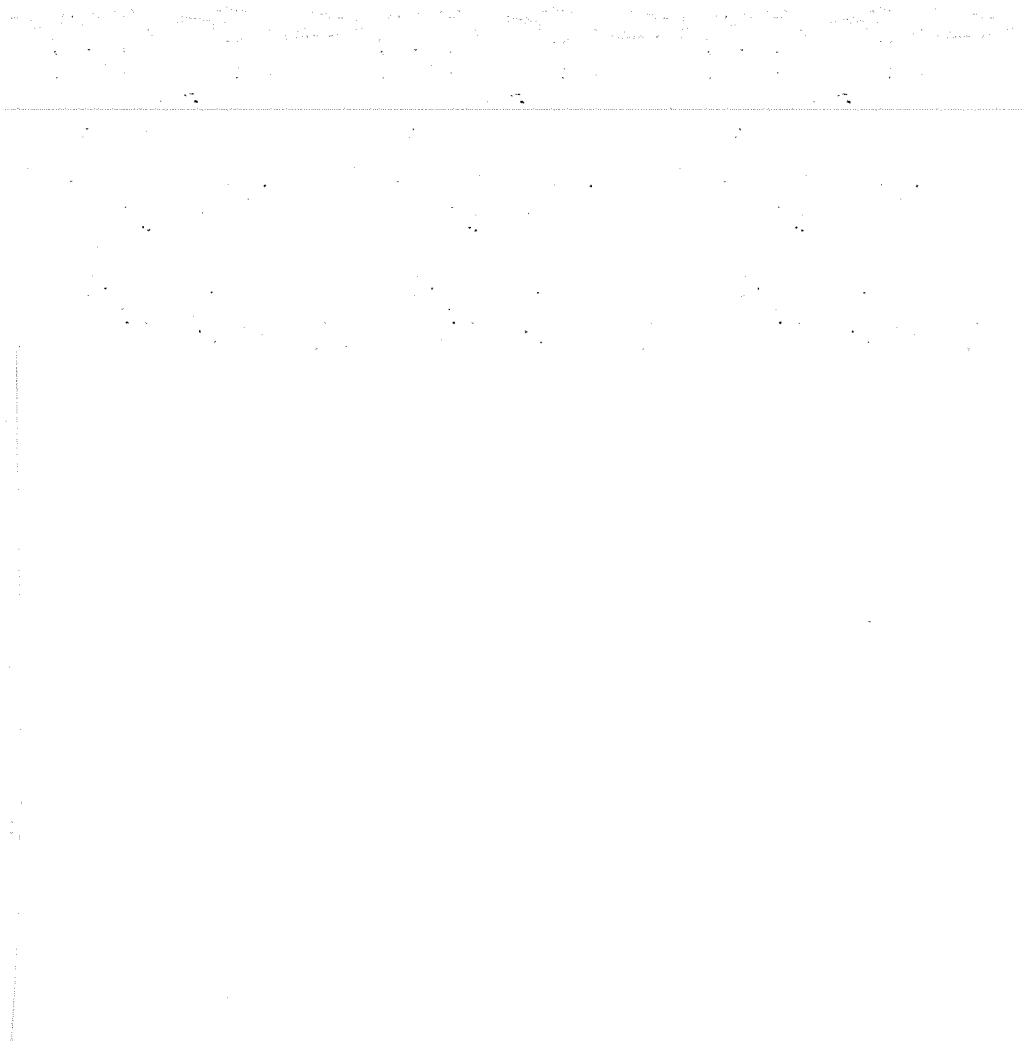
German Shepard $\frac{44}{369}(360°) \approx 43°$

Golden Retrievers $\frac{43}{369}(360°) \approx 42°$

Beagle $\frac{39}{369}(360°) \approx 38°$

Dachshund $\frac{36}{369}(360°) \approx 35°$

Boxer $\frac{35}{369}(360°) \approx 34°$



Statistics 2.1-2.2 Review Name: Key

Section 2-1: Frequency Distributions and Their Graphs

*Construct a frequency distribution including limits, midpoints, relative frequencies, cumulative frequencies and boundaries.

*Answer questions about a frequency distribution or its graph.

*Construct a frequency histogram, frequency polygon, relative frequency histogram, or ogive from a given data set or frequency distribution.

*Answer questions about a graph.

Use the table below to answer questions 1 -9.

Daily Low Temperature ($F°$)	Frequency	Midpoints	Relative Frequency	Cumulative Frequency	Class Boundaries
35-39	1	$35+39=74/2=37$	$\frac{1}{35}\approx .03$	1	34.5-39.5
40-44	3	42	$\frac{3}{35}\approx .09$	4	39.5-44.5
45-49	5	47	$\frac{5}{35}\approx .14$	9	44.5-49.5
50-54	11	52	$\frac{11}{35}\approx .31$	20	49.5-54.5
55-59	7	57	$\frac{7}{35}\approx .20$	27	54.5-59.5
60-64	7	62	$\frac{7}{35}\approx .20$	34	59.5-64.5
65-69	1 $\Sigma f=35$	67	$\frac{1}{35}\approx .03$	35	64.5-69.5

1. Identify the class width. $40-35=5$ or $64-59=5$ or $59-55+1=5$
2. Complete the remaining columns. ✓
3. Create a frequency histogram. Identify the class with the least and greatest frequency. 35-39 + 65-69 (least); 50-54 (greatest)
4. Create a relative frequency histogram. ✓
5. Create a frequency polygon. ✓
6. Create an ogive. Which class has the largest increase in frequency? 50-54
7. What percent of the days had a low temperature of 50° or higher? $\frac{26}{35}\approx 74\%$
8. What did you do to answer #6? Added relative frequencies for 50-69
9. How many days had low temperatures in the 40's? 8 days
10. What did you do to answer #9? Added the 2 frequencies (3+5) in the 40's range (40-44 + 45-49)

*Be able to explain advantages/disadvantages when using the above graphical displays (ex: why choose stem-and-leaf plot over histogram).

Use the stem-and-leaf plot below for questions 11-13.

Key 1|7 = 1.7

0	6
1	1 7 9
2	2 6
3	2 4 7 8
4	1 5 6 9 9
5	3 6 8
6	2 4 4 5 7
7	
~~8~~	~~5 6~~

(12) .6, 1.1, 1.7, 1.9, 2.2, 2.6, 3.2, 3.4, 3.7, 3.8, 4.1, 4.5, 4.6, 4.9, 4.9, 5.3, 5.6, 5.8, 6.2, 6.4, 6.4, 6.5, 6.7

11. What is the minimum and maximum value in the stem-and-leaf plot? .6, 6.7 (If you take out last 2
12. Write the data set represented by the stem-and-leaf plot.
13. Create a dot plot using the data in the stem-and-leaf plot.

Use the data below to answer questions 14 – 16.

The table below shows the responses of a poll of adults on June 16, 2006 about whether America should keep the penny.

Response	# of people
Yes	58
No	148
Not Sure	44

n = 250

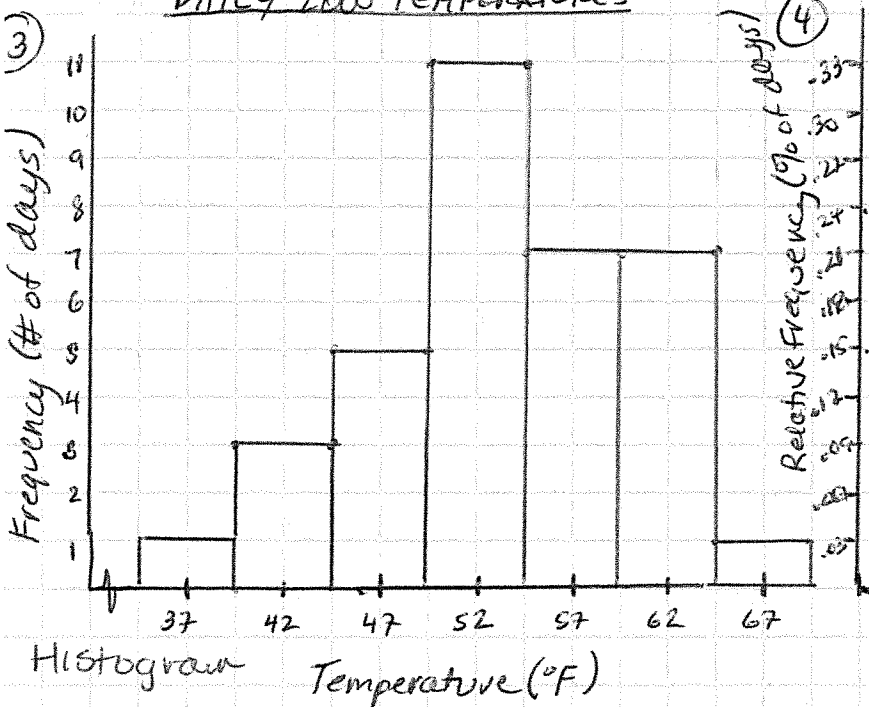
14. What percent said "yes" to keeping the penny? $\frac{58}{250} \approx 23\%$
15 Create a pie chart using the data above.
16 Create a Pareto chart using the responses above.

17 Use a time series chart to display the data below.

Number of Collision Between Wildlife and Aircraft

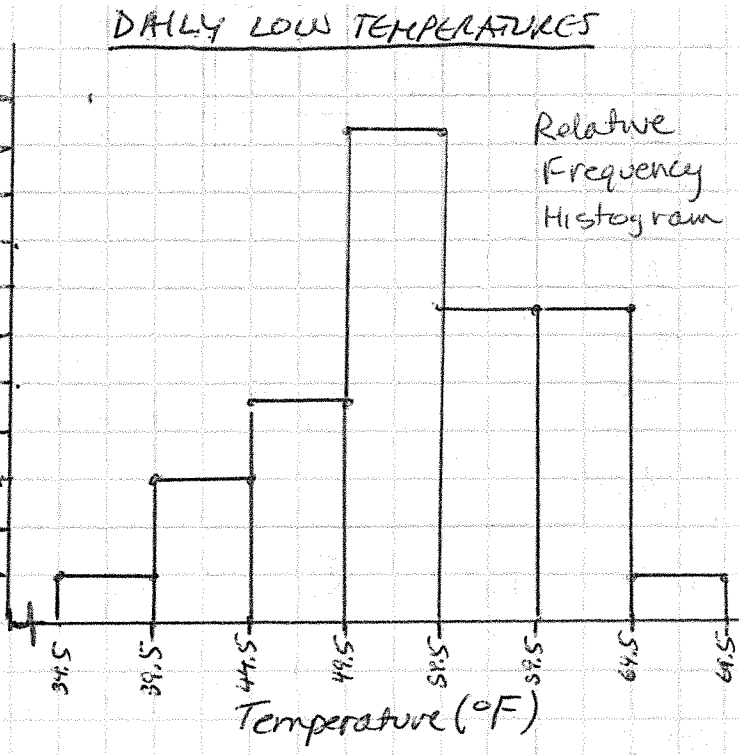
Year	Collisions
1990	1285
1992	1383
1994	2775
1996	6323
1998	5204
2000	4091
2002	5622
2004	4990
2006	5457

(3)

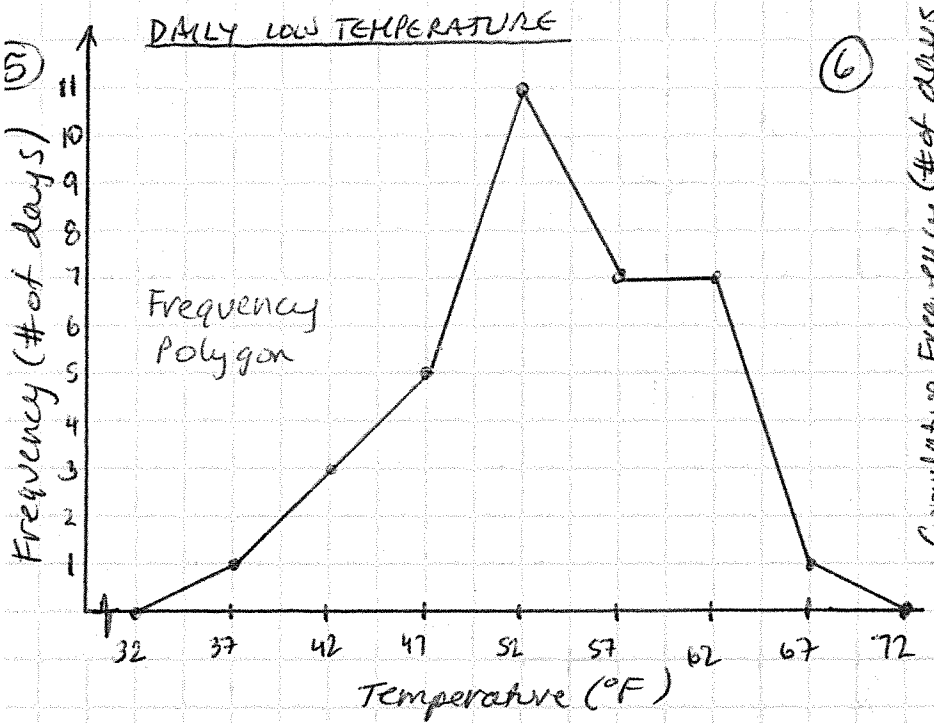


Histogram

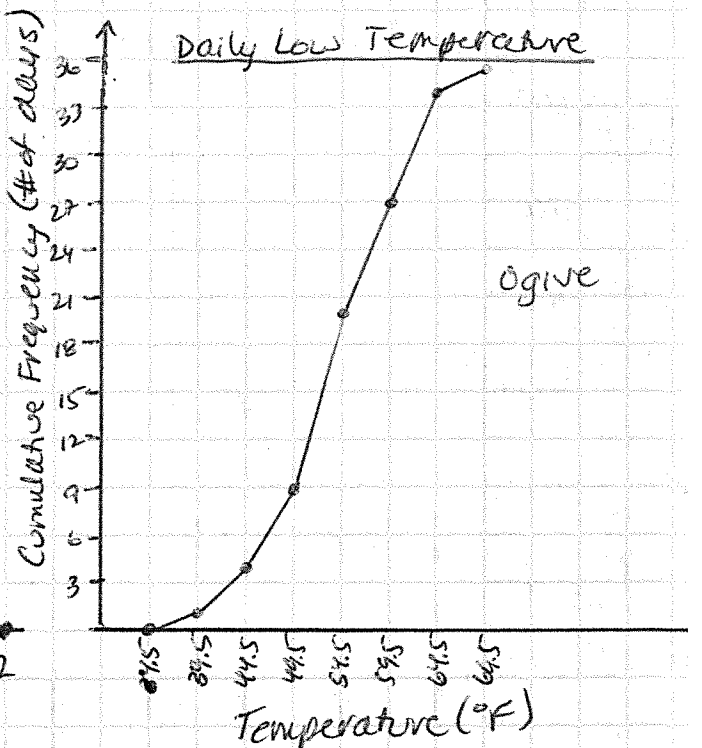
(4)



(5)



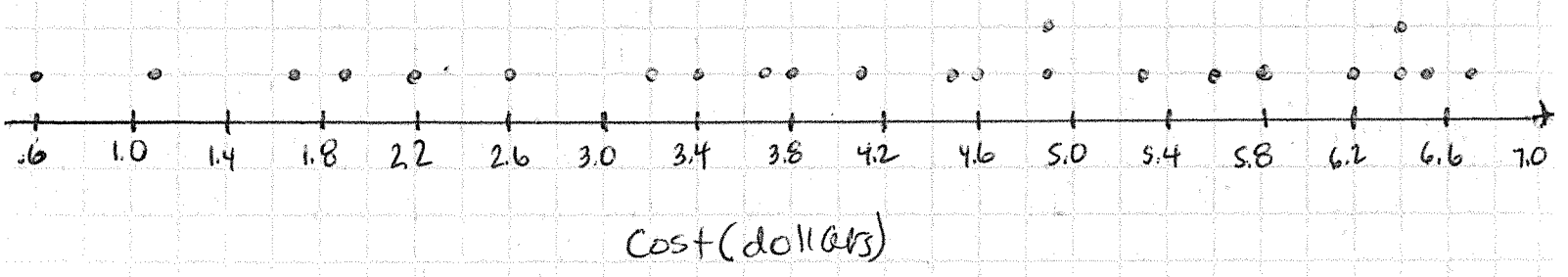
(6)



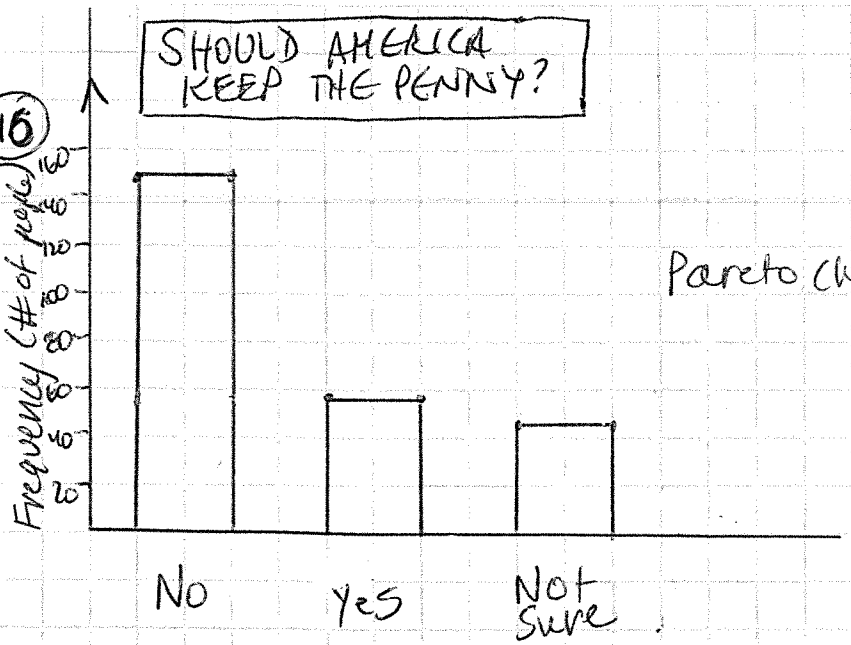
(3)

Prices at a Flea Market

Dot Plot

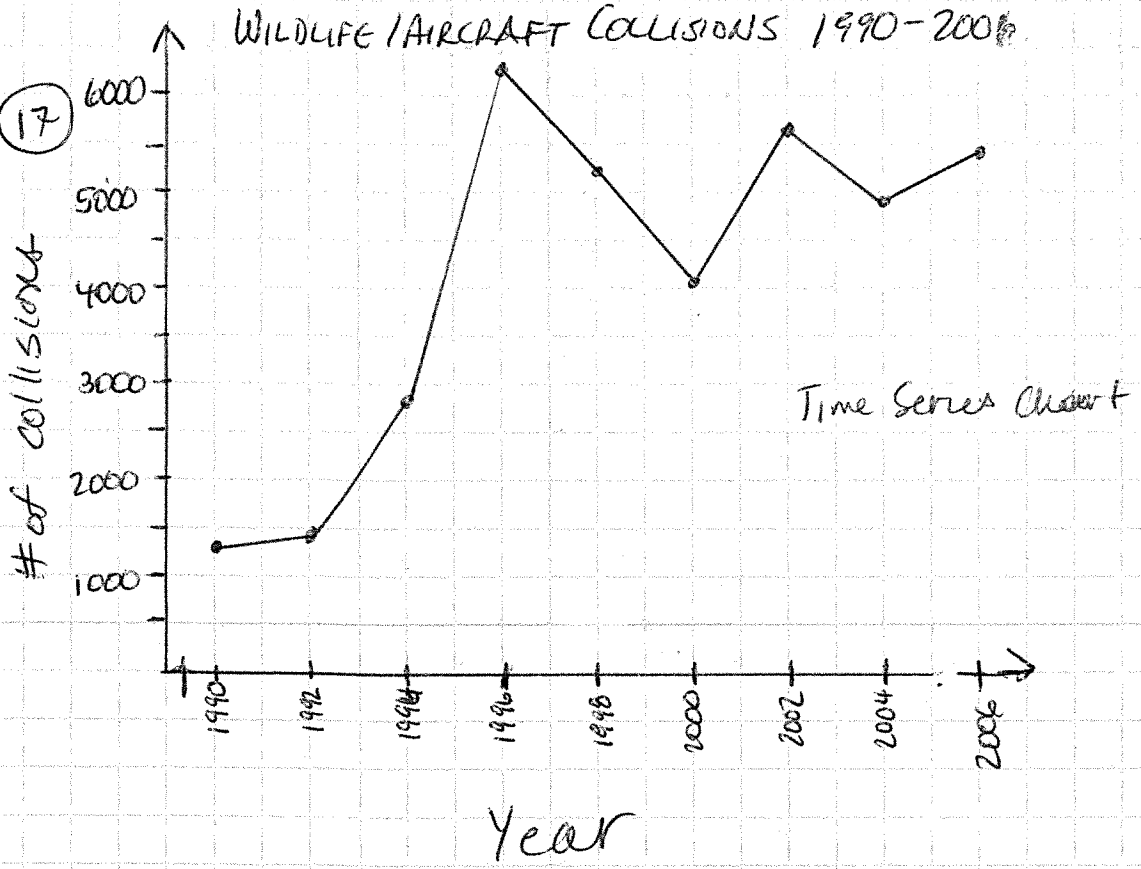


(16)



Pareto chart

(17)



Time Series Chart

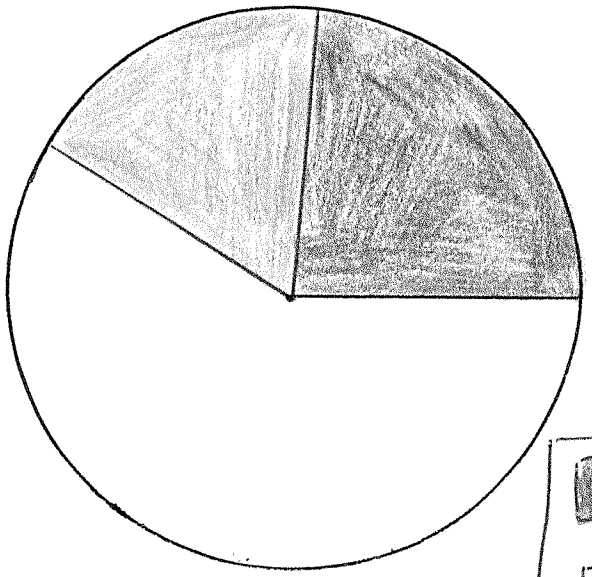
Should America Keep The Penny?

(15) $\frac{58}{250}(360°) \approx 84°$

$\frac{148}{250}(360°) = 213°$

$\frac{44}{250}(360°) = 63°$

* you can avoid having to draw a 213° ∠ by just drawing the other two.



Key

- Yes
- No
- Not Sure

Statistics 2.3-2.5 Name: Key Hour: _____

Use the data below to answer questions 1 and 2.
The costs of 20 textbooks chosen randomly from a college bookstore are listed below.

$125	$98	$132	$112	$86	$100	$98	$120	$90	$38
$115	$130	$95	$100	$89	$93	$105	$98	$122	$130

$ → Round to penny (100ths)

1. Find the following values for the data set. (Be careful with SD—is it sample or pop?)

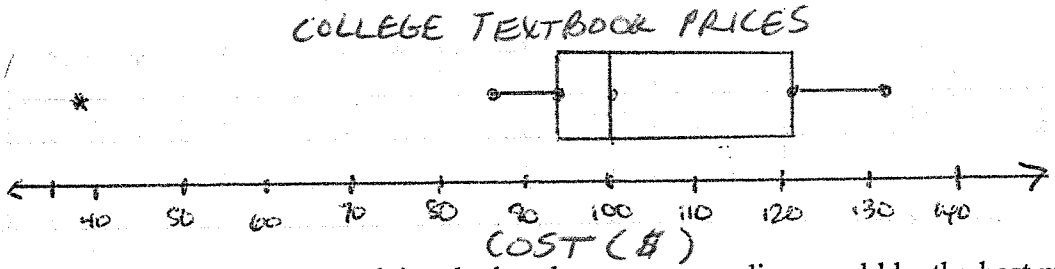
Mean: $\bar{x} = \frac{2076}{20}$ $103.80 Stand Dev: $s = \$21.50$ Min = $38 Max = $132

Q_1 = $94 Median: $100 Q_3 = $121 Mode: $98

IQR = 121 − 94 = 27 27(1.5) = 40.5

Outlier Check
94 − 40.5 = $53.50
121 + 40.5 = $161.50
$38 is an outlier

2. Construct a box and whisker plot of the data below. You can add more tick marks if necessary.



3. Given the following data set, explain whether the mean or median would be the best measure of central tendency. There are no outliers to skew the data, so mean is the best measure of center.

Ages of people in a library on a Tuesday afternoon:

63	36	28	42	39	53	59	45	62	23

min = 23 Q_1 = 36 Q_2 = 43.5 Q_3 = 59 max 63 IQR = 59 − 36 = 23

23 × 1.5 = 34.5

Q_1: 36 − 34.5 = 1.5 Q_3: 59 + 34.5 = 93.5

(1.5, 93.5) non-outliers

4. The ages in years when 12 randomly people moved the first time shown below. Use the data to find the range and the sample standard deviation.

6	9	12	15	12	13	14	21	10	17	11	14

Range = 21 − 6 = 15 yrs Standard deviation: $s \approx 3.8$ yrs

Did you use sample or population standard deviation? sample

5. Grade points are assigned as follows: A = 4, B = 3, C = 2, D = 1 and E = 0. Grades are weighted according to credit hours. If a student receives an B in a 3-credit hour class, an A in a 4-credit hour class, an B in a 4-credit hour class, and a C in a 2-credit hour class, what is the student's grade point average?

Grade (x)	Cr Hrs (w)	xw
3	3	9
4	4	16
3	4	12
2	2	4

$\sum w = 13$ $\sum xw = 41$

$GPA = \frac{\sum xw}{\sum w} = \frac{41}{13} \approx 3.15$

B

6. The data below shows the number of miles that members of the cross country team ran in one week. Find the mean of the frequency distribution.

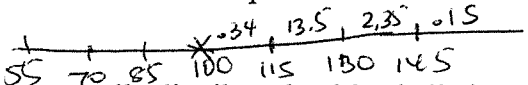
Number of miles run per week

Miles	x (midpts)	f Frequency	xf
1 – 3	2	4	8
4 – 6	5	8	40
7 – 9	8	9	72
10 – 12	11	7	77
13 – 15	12	3	36

$\Sigma f = 31$ $\Sigma xf = 233$

$\mu = \frac{\Sigma xf}{\Sigma f} = \frac{233}{31} \approx 7.5$ miles

7. IQ scores in the U.S. generally have a mean of 100 with a standard deviation of 15. IQ scores follow a bell-shaped distribution. Estimate the percent of Americans with IQ scores above 115.

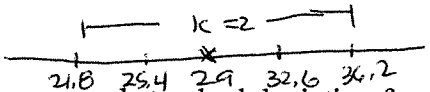


$13.5 + 2.35 + .15 = 16\%$

8. ACT scores in the U.S. are normally distributed, with a bell-shaped curve. It has been found that on the ACT, $\mu = 21$ and $\sigma = 4.7$. Between what two values would 99.7% of scores fall?

$21 \pm 3(4.7) = 21 \pm 14.1$ → 6.9, 35.1 $k=3$ 99.7% of students score between 6.9 + 35.1 on the ACT

9. The ages of new hires at a corporation were collected and found that the mean age at the date of hire was 29 years old with a standard deviation of 3.6 years. If the data is skewed right, about what percentage of the employees are hired between the ages of 21.8 and 36.2? Chebychev's



$k = 2$

At least 75% of new hires are between 21.8 + 36.2 yrs old

10. The mean and standard deviation for a psychology test is $\bar{x} = 83.5$ and $s = 2.7$. If Martha received a score of 96, calculate her z-score. Is Martha's test score usual or unusual? Explain.

z-score: 4.63

$z = \frac{96 - 83.5}{2.7} = \frac{12.5}{2.7} = 4.63$

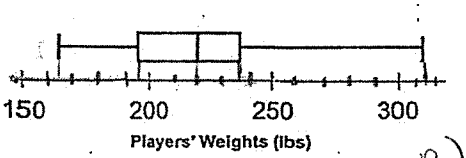
Explanation:

Her score is highly unusual. Any z score > 3 is highly unusual

11. A newborn baby's weight placed her in the 84th percentile. What can you conclude about this baby's weight compared to other newborns?

She weighs more than 84% of babies of the same age.

12. The box and whisker plot graphs the weights of players on a football team. Use it to answer parts [A] – [D].



[A] About 25% of weights are above 236 lbs / 237 lbs

[B] One half of the weights are between _____ and see below

[C] About 75% of weights are below 236/237 lbs

[D] About 75% of weights are above 195 lbs

B) 3 answers possible

165 – 220 (min → Q2)

195 – 236/7 (Q1 → Q3)

236/237 – 315 (Q3 → max)

SLOT STUDY GUIDE

① $\frac{7x}{x-7} - \frac{4}{x} = \frac{28}{x(x-7)}$ LCM: $x(x-7)$

$\frac{x(x-7)}{1} \cdot \frac{7x}{(x-7)} - \frac{x(x-7)}{1} \cdot \frac{4}{x} = \frac{28}{x(x-7)} \cdot \frac{x(x-7)}{1}$

$7x \cdot x - 4(x-7) = 28$

$7x^2 - 4x + 28 = 28$

$7x^2 - 4x = 0$

$x(7x-4) = 0$

extraneous ~~$x=0$~~ $7x-4=0$ $7x=4$ $\boxed{x = 4/7}$

② LCM: $(x+2)(x-2)$ $\frac{(x+2)(x-2)}{1} \cdot \frac{10}{x-2} = \frac{(x+2)(x-2)}{1} \cdot \frac{1}{1} + \frac{12}{x+2} \cdot \frac{(x+2)(x-2)}{1}$

$10(x+2) = (x+2)(x-2) + 12(x-2)$

$10x + 20 = x^2 - 4 + 12x - 24$

$0 = x^2 + 2x - 48$

$(x+8)(x-6) = 0$

$x+8=0$ $x-6=0$

$\boxed{x=-8}$ $\boxed{x=6}$

③ LCM: $4x$

$\frac{4x}{1} \cdot \frac{5-x}{x} + \frac{4x}{1} \cdot \frac{3}{4} = \frac{7}{x} \cdot \frac{4x}{1}$

$4(5-x) + 3x = 7 \cdot 4$

$20 - 4x + 3x = 28$

$-x + 20 = 28$

$-x = 8$

$\boxed{x=-8}$

④ $\frac{4}{x-5} + \frac{9}{x-8} = \frac{7}{(x-5)(x-8)}$

LCM $= (x-5)(x-8)$

$\frac{(x-5)(x-8)}{1} \cdot \frac{4}{x-5} + \frac{9}{x-8} \cdot \frac{(x-5)(x-8)}{1} = \frac{7}{(x-5)(x-8)} \cdot \frac{(x-8)}{1}$

$4(x-8) + 9(x-5) = 7$

$4x - 32 + 9x - 45 = 7$

$13x - 77 = 7$

$13x = 84$

$x = \frac{84}{13}$

(5) $\frac{2x}{x+2} + \frac{5}{x-5} = \frac{8}{(x-5)(x+2)}$ LCM: $(x-5)(x+2)$

$\frac{(x-5)(x+2)}{1} \cdot \frac{2x}{x+2} + \frac{5}{x-5} \cdot \frac{(x-5)(x+2)}{1} = \frac{8}{(x-5)(x+2)} \cdot \frac{(x-5)(x+2)}{1}$

$2x(x-5) + 5(x+2) = 8$

$2x^2 - 10x + 5x + 10 = 8$

$2x^2 - 5x + 2 = 0$

$(2x-1)(x-2) = 0$

$2x-1=0$ $\quad 2x=1$ $\quad \boxed{x = \frac{1}{2}}$

$x-2=0$ $\quad \boxed{x=2}$

(6) $\frac{-1}{x-1} = \frac{1}{1} - \frac{x}{x-1}$ LCM: $x-1$

$\frac{(x-1)}{1} \cdot \frac{-1}{(x-1)} = \frac{(x-1)}{1} \cdot \frac{1}{1} - \frac{(x-1)}{1} \cdot \frac{x}{(x-1)}$

$-1 = x-1-x$

$-1 = -1$ infinitely many solns

(7) LCM: $x-1$

$\frac{x-1}{1} \cdot \frac{x}{1} = \frac{x-1}{1} \cdot \frac{4x-3}{x-1} + \frac{x-1}{1} \cdot \frac{-x}{x-1}$

$x(x-1) = 4x-3-x$

$x^2 - x = 3x - 3$

$x^2 - 4x + 3 = 0$

$(x-3)(x-1) = 0$

$x-3=0$ $\quad \boxed{x=3}$

$x-1=0$ $\quad \boxed{x=1}$ extraneous

(8) $\frac{x+4}{x} + \frac{3}{x-4} = \frac{-16}{x^2-4x}$ $\quad x(x-4)$

LCM: $x(x-4)$

$\frac{x(x-4)}{1} \cdot \frac{x+4}{x} + \frac{x(x-4)}{1} \cdot \frac{3}{x-4} = \frac{-16}{x(x-4)} \cdot \frac{x(x-4)}{1}$

$(x-4)(x+4) + 3x = -16$

$x^2 - 16 + 3x - 16$

$x^2 + 3x = 0$

$x(x+3) = 0$

$\boxed{x=0}$ extraneous

$x+3=0$ $\quad \boxed{x=-3}$

Rational Study Guide

	LCM	SOL'N(S)
(1)	$x(x-7)$	$x = 4/7$ ~~$x = 0$~~
(2)	$(x+2)(x-2)$	$x = -8$ $x = 6$
(3)	$4x$	$x = -8$
(4)	$(x-5)(x-8)$	$x = 84/13$
(5)	$(x+2)(x-5)$	$x = 2$ $x = 1/2$
(6)	$x-1$	Infinitely many sol'ns
(7)	$x-1$	$x = 3$ ~~$x = 1$~~
(8)	$x(x-4)$	$x = -3$ ~~$x = 0$~~