Dulaney, Stewart
MATH 54
Section 4053
SID: 1545566

10/10

HW 1

1.1

1. Define statistics

Statistics is the science of collecting, organizing, summarizing, and analyzing information to draw conclusions or answer questions. In addition, statistics is about providing a measure of confidence in any conclusions.

3. Explain the difference between a population and a sample.

The entire group to be studied is called the population. A sample is a subset of the population that is being studied.

5. A statistic is a numerical summary of a sample.
A parameter is a numerical summary of a population.

In problems 7-14, determine whether the underlined value is a parameter or statistic.

7. State Government: Following the 2014 national midterm election, 18% of the governors of the 50 United States were female.

A: parameter

9. School Bullies: In a national survey of 1300 high school students (grades 9 to 12), 32% of respondents reported that someone had bullied them at school.

A: statistic

11. Batting Average: Ty Cobb is one of Major League Baseball's greatest hitters of all time, with a career batting average of 0.366.

A: statistic

13. Hygiene Habits: A study of 6076 adults in public rest rooms (in Atlanta, Chicago, New York City, and San Francisco) found that 23% did not wash their hands before exiting.

A: statistic

In problems 15-22, classify the variable as qualitative or quantitative

15. Nation of origin

A: qualitative

17. Grams of carbohydrates in a doughnut

A: quantitative

19. Number of unpopped kernels in a bag of microwave popcorn

A: quantitative

21. Phone number

A: qualitative

In problems 23-30, determine whether the quantitative variable is discrete or continuous

23. Goals scored in a season by a soccer player

A: discrete

25. Length (in minutes) of a country song

A: continuous

27. High temperature on a randomly selected day in Memphis, Tennessee

A: continuous

29. Points scored in an NCAA basketball game

A: discrete

In problems 31-38, determine the level of measurement of each variable

31. Nation of origin

A: nominal level of measurement

33. Volume of water used by a household in a day
A: ratio level of measurement

35. Highest degree conferred (high school, bachelor's, and so on)
A: ordinal level of measurement

37. Assessed value of a house
A: ratio level of measurement

In problems 39-44, a research objective is presented. For each, identify the population and sample in the study.

39. The Gallup Organization contacts 1028 teenagers who are 13 to 17 years of age and live in the United States and asks whether or not they have been prescribed medications for any mental disorders, such as depression or anxiety.
- population: all teenagers 13 to 17 years of age
- sample: 1028 teenagers 13 to 17 years of age in the study

41. A farmer interested in the weight of his soybean crop randomly samples 100 plants and weighs the soybeans on each plant.
- population: all plants in the soybean crop
- sample: 100 plants sampled in the study

43. Folate and Hypertension: Researchers want to determine whether or not folate intake is associated with a lower risk of hypertension (high blood pressure) in women (27 to 44 years of age). To make this determination, they look at 7373 cases of hypertension in these women and find that those who consume at least 1000 micrograms per day ($\mu g/d$) of total folate had a decreased risk of hypertension compared with those who consume less than 200 $\mu g/d$.
- population: all women 27 to 44 years of age
- sample: 7373 cases of hypertension in women 27 to 44 years of age in the study

In problems 45 and 46, identify the individuals, variables, and data corresponding to the variables. Determine whether each variable is qualitative, continuous, or discrete.

45. Driver's License Laws

State	Minimum Age for Driver's License (unrestricted)	Mandatory Belt Use Seating Positions	Maximum Allowable Speed Limit (cars on rural interstate), mph
Alabama	17	Front	70
Colorado	17	Front	75
Indiana	18	All	70
North Carolina	18	All	70
Wisconsin	18	All	65

- individuals: driver's license laws for a given state
- variables:
  - Minimum Age For Driver's License (unrestricted) - discrete
  - Mandatory Belt Use Seating Positions - qualitative
  - Maximum Allowable Speed Limit, mph - continuous
- data: the data for each variable are the values in the above table

47. Smoker's IQ: A study was conducted in which 20,211 18-year-old Israeli male military recruits were given an exam to measure IQ. In addition, the recruits were asked to disclose their smoking status. An individual was considered a smoker if he smoked at least one cigarette per day. The goal of the study was to determine whether adolescents aged 18 to 21 who smoke have a lower IQ than nonsmokers. It was found that the average IQ of the smokers was 99, while the average IQ of the nonsmokers was 101. The researchers concluded that lower IQ individuals are more likely to choose to smoke, not that smoking makes people less intelligent.

a) What is the research objective? The goal of the study was to determine whether adolescents aged 18 to 21 who smoke have a lower IQ than nonsmokers.

b) What is the population being studied? What is the sample?
The population is all 18-year-old Israeli male military recruits. The sample is the 20,211 18-year-old Israeli male military recruits in the study.

c) What are the descriptive statistics?
It was found that the average IQ of the smokers was 94, while the average IQ of the nonsmokers was 101.

d) What are the conclusions of the study?
The researchers concluded that lower IQ individuals are more likely to choose to smoke, not that smoking makes people less intelligent.

49. Government Waste: Gallup News Service conducted a survey of 1017 American adults aged 18 years or older, September 4-7, 2014. The respondents were asked, "Of every tax dollar that goes through the federal government in Washington, D.C., how many cents of each dollar would you say are wasted?" Of the 1017 individuals surveyed, 35% indicated that 51 cents or more is wasted. Gallup reported that 35% of all adult Americans 18 years or older believe the federal government wastes at least 51 cents of each dollar spent, with a margin of error of 4% and a 95% level of confidence.

a) What is the research objective? The goal was to find out how many cents out of every tax dollar American adults 18 years or older think are wasted.

b) What is the population? The population is all American adults 18 years or older.

c) What is the sample? The sample is the 1017 American adults in the study.

d) List the descriptive statistics. Of the 1017 individuals surveyed, 35% indicated that 51 cents or more is wasted.

e) What can be inferred from this survey? 35% of all adult Americans 18 years or older believe the federal government wastes at least 51 cents of each dollar spent, with a margin of error of 4% and a 95% level of confidence.

51. What Level of Measurement? It is extremely important for a researcher to clearly define the variables in a study because this helps determine the type of analysis that can be performed on the data. For example, if a researcher wanted to describe baseball players based on jersey number, what level of measurement would the variable jersey number be?

A: nominal level of measurement

Now suppose the researcher felt that certain players who were of lower caliber received higher numbers. Does the level of measurement of the variable change? If so, how?

A: Yes, it becomes ordinal level of measurement because the naming scheme now allows the value of the variable to be arranged in a ranked order.

53. Analyze the Article

a) What is the research question the study addresses?

The study addresses whether your season of birth affects your mood later in life.

b) What is the sample?

The sample is the 400 people in the study.

c) What type of variable is season in which you were born?

Qualitative

d) What can be said (in general) about individuals born in summer? Winter?

A tendency to be excessively positive was significantly higher among those born in summer. Those born in winter were less likely to be irritable.

e) What conclusion was drawn from the study?
It seems that when you are born may increase or decrease your chance of developing certain mood disorders, though they can't say anything about the mechanisms involved.

1.2

1. An explanatory variable is the one that's varying level may have an effect on the response variable. The response variable is the outcome being studied which may be influenced by the explanatory variable.

3. Confounding in a study occurs when the effects of two or more explanatory variables are not separated. A lurking variable is an explanatory variable that was not considered in a study but that affects the value of the response variable in the study. A confounding variable is an explanatory variable that was considered in a study whose effect cannot be distinguished from a second explanatory variable in the study.

9. Observational study

13. Observational study

17. a) This is a cohort study because group of individuals was identified then observed over a long period of time.
b) The response variable is whether or not heart disease was contracted. The explanatory variable is level of happiness.
c) This sentence means that genetics could be a lurking variable.

19. a) This is an observational study because researchers don't intentionally change the value of an explanatory variable. This is a cross-sectional study because information was only collected at a specific point in time.

6) The response variable is BMI and obesity. The explanatory variable is a TV in the bedroom.

c) Lurking variables could include genetics, diet, exercise habits, or social factors.

d) This means that socioeconomic status was not a confounding varinble, meaning its effect could be distinguished from the explanatory variable in the study.

e) We can only conclude that a TV in the bedroom is associated with a higher BMI, because observational studies don't allow a researcher to claim causation, only association.

1.3

1. A frame is a list of all individuals within the population.

3. In a sample without replacement, an individual who is selected is removed from the population and cannot be chosen again.

7. a)

EPR 616, EPR 630	EPR 616, EPR 631	EPR 649, EPR 650
EPR 616, EPR 632	EPR 616, EPR 645	
EPR 616, EPR 649	EPR 616, EPR 650	
EPR 630, EPR 631	EPR 630, EPR 632	
EPR 630, EPR 645	EPR 630, EPR 649	
EPR 630, EPR 650	EPR 631, EPR 632	
EPR 631, EPR 645	EPR 631, EPR 649	
EPR 631, EPR 650	EPR 632, EPR 645	
EPR 632, EPR 649	EPR 632, EPR 650	
EPR 645, EPR 649	EPR 645, EPR 650	

b) Only 1 in 21 possible samples contains EPR 630 and EPR 645, so there is a 1 in 21 chance that the simple random sample will contain these three.

9. a) 83, 67, 84, 38, 22, 24, 36, 58, 34
b) 31, 9, 58, 83, 31, 78, 63, 13, 29

1.4

1. One circumstance in which stratified sampling would be an appropriate sampling method is a poll on company rules with three strata: employees with management background, employees with sales background, and employees with technical background.

3. Convenience samples are ill advised because they yield unreliable results because the individuals participating in the survey are not chosen using random sampling.

5. A stratified sample is obtained by dividing the population into homogeneous groups and randomly selecting individuals from each group.

7. False

9. True

11. Systematic sample

13. Cluster sample

15. Simple random sample

17. Cluster sample

19. Convenience sample

21. Stratified sample

23. 16, 41, 66, 91, 116, 141, 166, 191, 216, 241
266, 291, 316, 341, 366, 391, 416, 441, 466, 491

25. Democrats

1 Bolden	5 Fallenbuchel	9 Motola	13 Ramirez
2 Bolt	6 Haydra	10 Nolan	14 Tate
3 Carter	7 Khouri	11 Opacian	15 Washington
4 Debold	8 Lukens	12 Pawlak	16 Wright

Stratified sample obtained using technology: 14, 16 → Tate, Wright

Republicans

1 Blouin	5 Grajewski	9 Ochs	13 Salihar
2 Cooper	6 Keating	10 Pechtold	14 Thompson
3 De Young	7 May	11 Redmond	15 Trudeau
4 Engler	8 Niemeyer	12 Rice	16 Zenkel

Stratified sample obtained using technology: 1, 4 → Blouin, Engler

27. a) $N = 4502$

$n = 50$

$\frac{N}{n} = \frac{4502}{50} = 90.04$ → rounded down → $k = 90$

b) $p = 17$

17, 107, 197, 287, 377, ..., 4427

29. - Simple random sample: Number the students from 1 to 1280. Randomly select 128 students to survey.
- Stratified sample: Randomly select four students from each section to survey ($4 \times 32 = 128$).
- Cluster: Randomly select four sections and survey all students in the sections ($4 \times 32 = 128$).
- In this situation I think a stratified sample would be best so that all classes are equally represented in the survey and the survey is not influenced by only representing students from four of the professors.

31. A good choice would be stratified sampling with the strata being noncommuters, commuters who currently drive, and commuters who currently take some form of public transportation. This would have the advantage of showing the difference in opinion between those who would and would not be likely to actually use the rail station.

1.5

1. A closed question requires the respondent to choose from a list of predetermined responses. An open question allows the respondent to choose his or her response. An adantage of closed questions is that responses are easy to analyze but a disadvantage is that they do not always include a respondent's desired choice. An advantage of open questions is a respondent can always use their desired response but a disadvantage is the variety of responses may be difficult to analyze.

3. If the results of the sample are not representative of the population, then the sample has bias.
   - 1. Sampling bias, e.g. a political poll with undercoverage of Democrats
   - 2. Nonresponse bias, e.g. an e-mail survey with response rate of only 40%
   - 3. Response bias, e.g. a survey of recent college graduates with self-reported salaries that are inflated

A census could have bias if a question is misunderstood, thereby leading to a response bias in the results.

5. a) Sampling bias due to undercoverage because the first 60 customers might not be representative of the customer population
   b) If a complete frame is not possible, systematic sampling would be used to make the sample more representative of the customer population.

7. a) Response bias due to a poorly worded question.
   b) The survey should begin by stating the currents laws and penalty for selling a gun illegally. The question could be rewritten as

"Do you approve or disapprove of harsher penalties for individuals who sell guns illegally?", with the words "approve" and "disapprove" rotated with each respondent.

9. a) Nonresponse bias; if the survey is only written in English, non-English speaking homes won't complete the survey
b) The survey could be conducted using random digit dialing telephone surveys which typically has response rates around 70%

11. a) Sampling bias due to undercoverage as the readers of the magazine may not be representative of all Australian women. There is also likely response bias due to messaging or images in the magazine that could affect women's view of themselves.
b) A well designed sampling plan wouldn't be conducted through a magazine and could use stratified sampling with the strata being geographic areas in Australia to make the sample more representative of the intended population.

13. a) Response bias due to a poorly worded question
b) The question should be rewritten so as not to imply any opinions of the editors. One option would be "Do you believe that a marriage can be maintained after an extramarital relation?".

15. a) Response bias because the students are not likely to tell the truth to their teachers.
b) The survey should be conducted by an independent third party so the students are more likely to respond truthfully.

17. No, the survey still suffers from sampling bias and undercoverage, nonresponse bias, and potentially response bias.

19. The ordering of the questions will likely affect the survey results. One solution would be to rotate the questions randomly.

21. The company is incentivizing the reader by telling them their participation will make a difference and using a reward in the form of a $5.00 payment.

23. The frame is anyone with a landline or cell phone. Any household without a phone, those on the DNC registry, and homeless individuals would be excluded. This would result in sampling bias due to undercoverage.

25. Yes, especially if households on the do-not-call registry have a characteristic those not on the registry do not have.

27. Nonsampling errors included poorly trained interviewers, interviewer bias that may have caused response bias, nonresponse bias due to surveying too many female voters.

29. The wording of a survey question can have a significant effect on the responses. In this case, the words "along with its allies in Europe" resulted in many more respondents saying they supported the idea.

31. Question 1 - Do you believe the U.S. should have stricter gun laws?
Question 2 - Do you believe the rise in mass school shootings in recent years means obtaining a gun is too easy?

Order of Questions	Answered Yes to Q1	Answered Yes to Q2
1, 2	8/25	14/25
2, 1	15/25	15/25

Yes, the results differ because respondents are more likely to answer yes to Q1 if it is asked after Q2.

33. Sampling bias due to undercoverage of Democrats caused by using an incorrect frame. There was also nonresponse bias, the rather low response rate.

35. a) One option would be a stratified sample with strata by political affiliation (Democrat, Republican, Independent). You could account for registered voters vs. likely voters by having the sample size from each strata be proportional to the party turnout rates in the previous election.
b) Historically Democratic turnout is higher in presidential election cycles, so differences such as this one could be accounted for by the sample size taken from each strata based on the type of election we are studying.
c) A higher percentage of Democrats in FiveThirtyEight polls versus turnout would result in overstating the predicted percentage of Democratic votes.

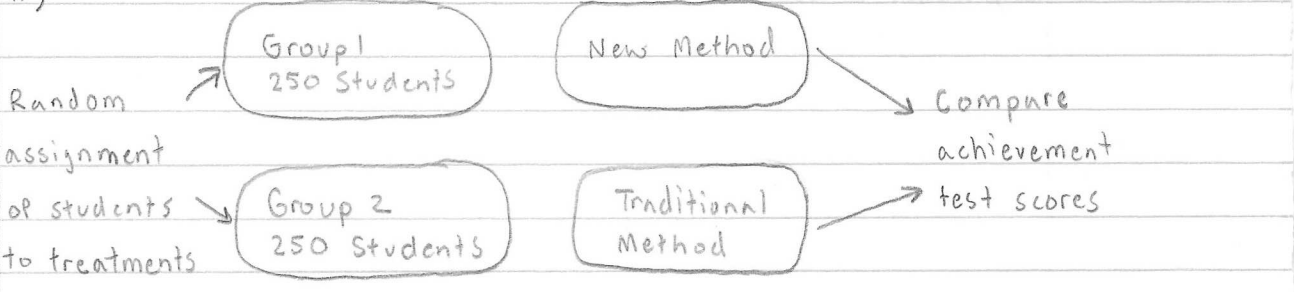
1.6

1. a) In an experiment, the experimental unit is a person, object, or some other well-defined item upon which a treatment is applied.
b) Any combination of the values of the factors, or explanatory variables, is called a treatment.
c) A response variable is a quantitative or qualitative variable that represents the variable of interest.
d) A factor is the variable whose effect on the response variable is to be assessed by the experimenter.
e) A placebo is an innocuous medication, such as a sugar tablet, that looks, tastes, and smells like the experimental medication.
f) Confounding means the effect of two factors (explanatory variables) on the response variable cannot be distinguished.

3. In single-blind experiments, the experimental unit (or subject) does not know which treatment he or she is receiving. In double-blind experiments, neither the experimental unit nor the researcher in contact with the experimental unit knows which treatment the experimental unit is receiving.

7. a) Researchers wanted to determine the association between number of times one chews food and food consumption.
b) The response variable is food intake, which quantitative.
c) The explanatory variable is number of times one chews food, which is quantitative.
d) The experimental units are 45 individuals 18 to 45 years of age.
e) Control is used by determining a baseline number of chews before swallowing. Also, the same type of food and same time of day (lunch) is used in the baseline as in the experiment.
f) Randomization reduces the effect of the order in which number of chews required plays. For example, perhaps the first time through subjects are more diligent about their chewing than the last time through the study.

9. a) The response variable is the score on the achievement test.
b) Factors include method of teaching, grade level, intelligence, school district, teacher. Fixed: grade level, school district, teacher. Set at a predetermined level: method of teaching.
c) There are two treatments, the new teaching method and the traditional method.
d) The factors that are not controlled are dealt with by random assignment to the two groups.
e) Group 2 is the control group.
f) This experiment is a completely randomized design.
g) The subjects are the 500 first-grade students in District 203.
h)



11. a) Matched-pair experimental design

b) Whiteness of teeth

c) Crest Whitestrips Premium in addition to regular brushing and flossing, brushing and flossing alone

d) Other factors that could affect whiteness of teeth include diet, specifically foods and tobacco products that are more likely to stain teeth

e) Using twins helps control for genetic factors like weak teeth that may affect the results of the study.

13. a) Completely randomized design

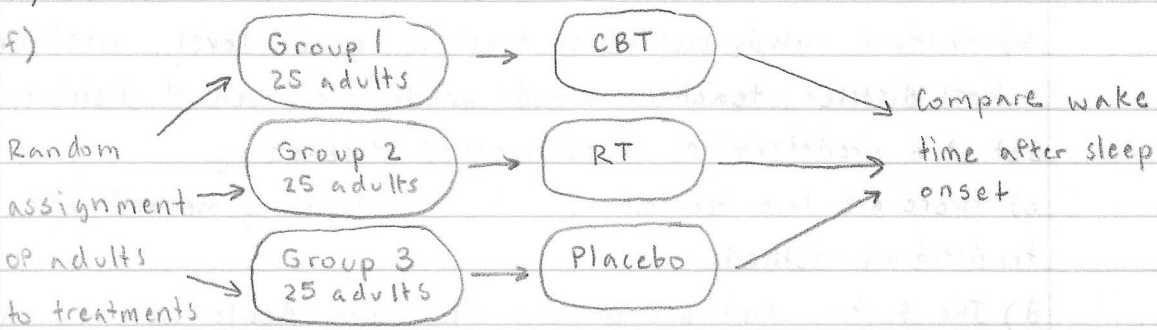
b) Adults with insomnia

c) Wake time after sleep onset (WASO)

d) CBT, RT, placebo

e) 75 adults with insomnia

f)



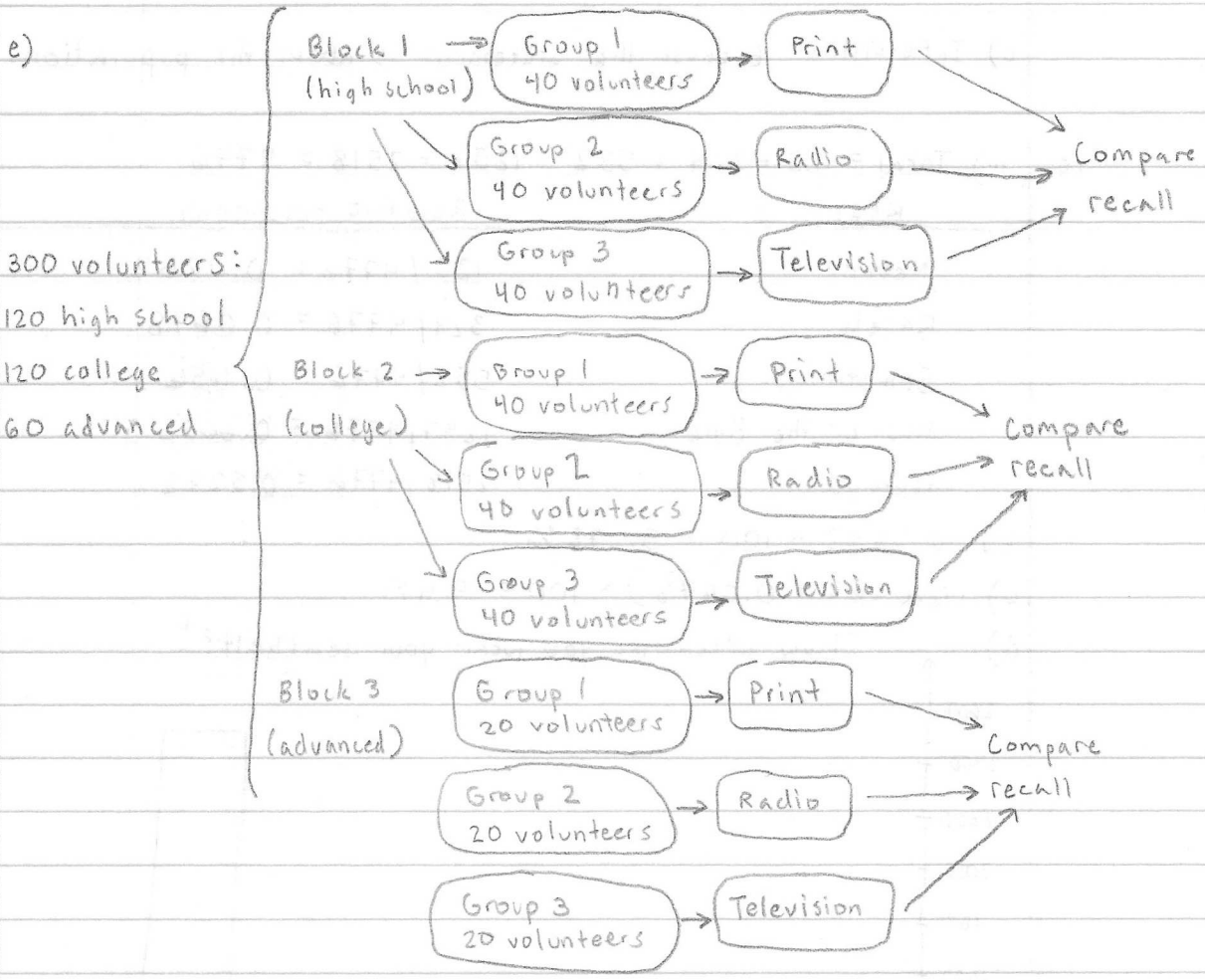
19. a) Randomized block design

b) Number of correct answers on recall exam

c) Print, radio, or television advertising

d) Level of education

e)



2.1 1. Raw data are data that have not yet been organized.

3. One, although rounding may cause the result to vary slightly.

7. a) OF

b) 15

c) 30 - 15 = 15

d) Instead of treating left field, center field, right field as the same position, each position should have the frequency of MVP's shown separately in the bar graph.

9. a) 69 %

b) $23\left(\frac{1}{100}\right)(240,000,000) = 55,200,000$

c) Inferential because the statement is about the population.

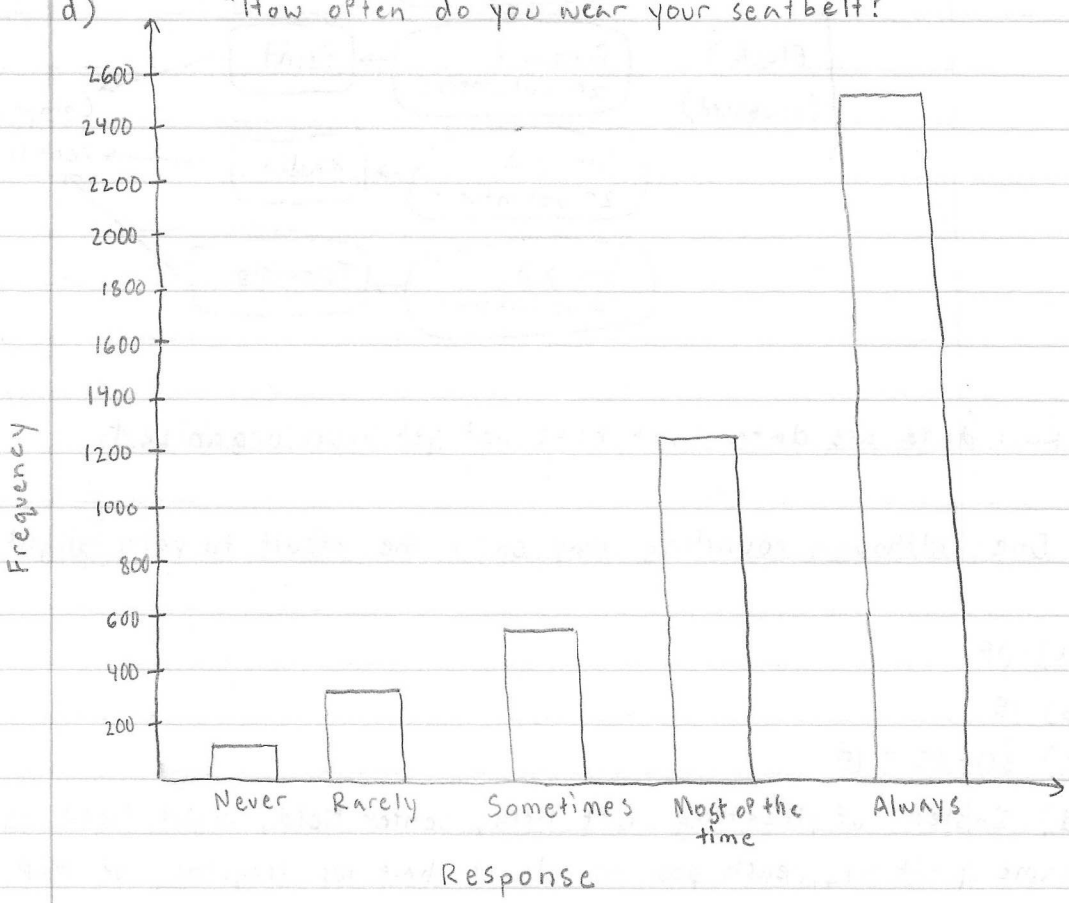
13. a) Total = 125 + 324 + 552 + 1257 + 2518 = 4776

Response	Relative Frequency
Never	125 / 4776 = 0.0262
Rarely	324 / 4776 = 0.0678
Sometimes	552 / 4776 = 0.1156
Most of the time	1257 / 4776 = 0.2632
Always	2518 / 4776 = 0.5272

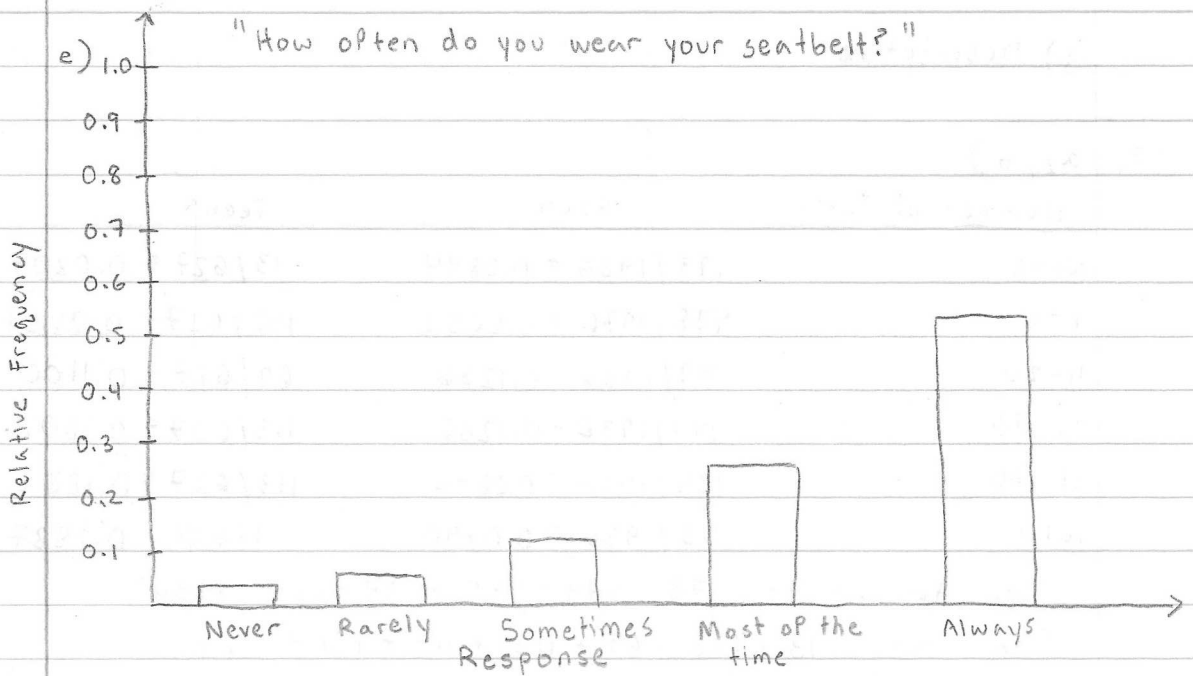
b) $0.5272 \times 100 = 52.72\%$

c) $(0.0262 + 0.0678) \times 100 = 9.4\%$

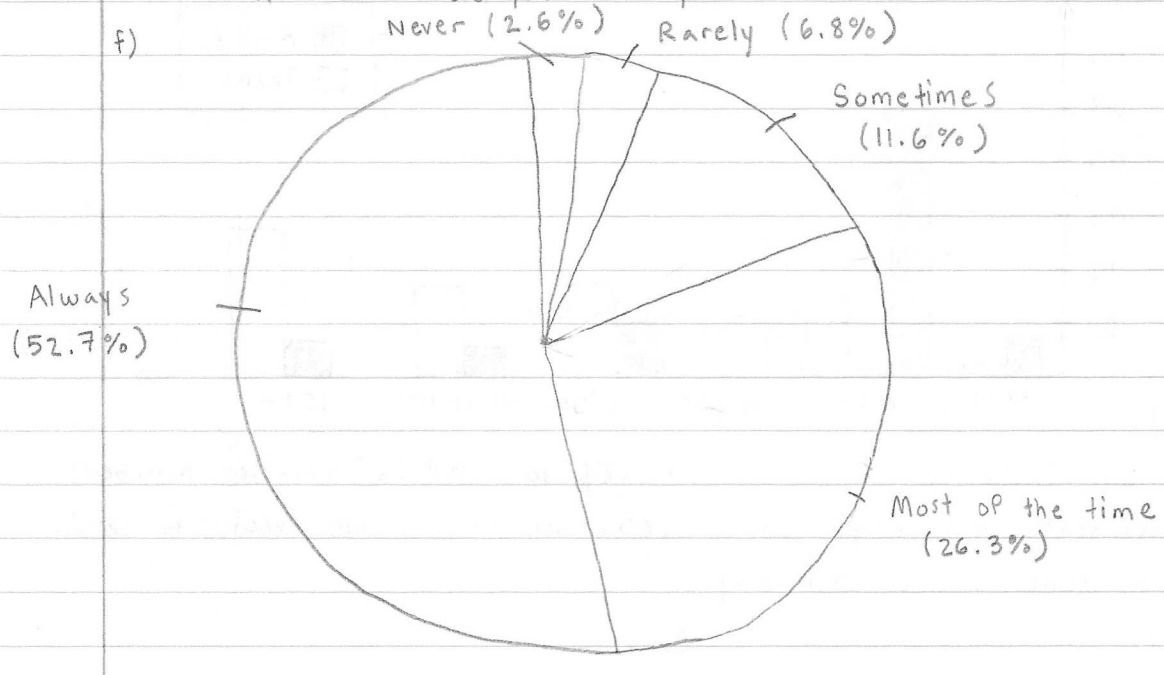
d) "How often do you wear your seatbelt?"



e) "How often do you wear your seatbelt?"



f) "How often do you wear your seatbelt?"



Response	Degree Measure of Each Sector
Never	$(0.0262)360^\circ = 9^\circ$
Rarely	$(0.0678)360^\circ = 24^\circ$
Sometimes	$(0.1156)360^\circ = 42^\circ$
Most of the time	$(0.2632)360^\circ = 95^\circ$
Always	$(0.5272)360^\circ = 190^\circ$

g) Descriptive

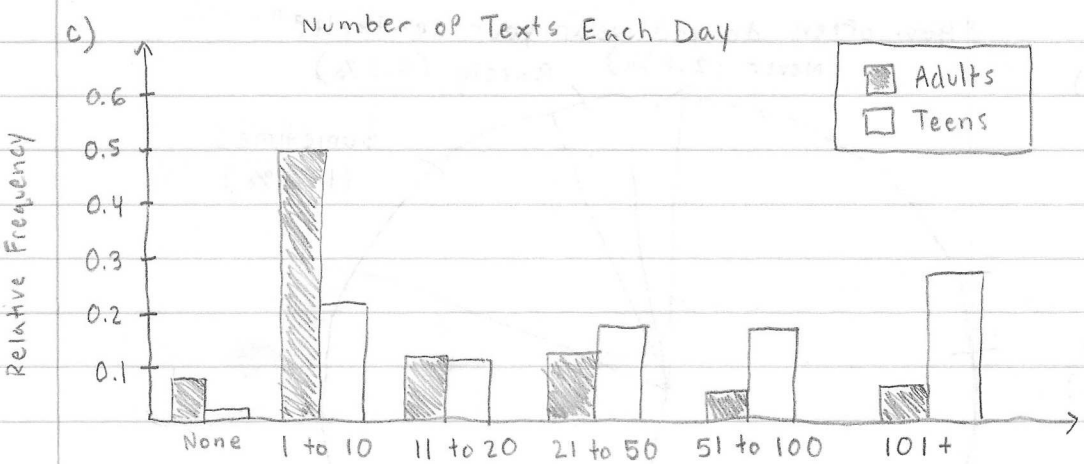
17. a), b)

Number of Texts	Adults	Teens
None	173/1936 = 0.0894	13/627 = 0.0207
1-10	978/1936 = 0.5052	138/627 = 0.2201
11-20	249/1936 = 0.1286	69/627 = 0.1100
21-50	249/1936 = 0.1286	113/627 = 0.1802
51-100	134/1936 = 0.0692	113/627 = 0.1802
101+	153/1936 = 0.0790	181/627 = 0.2887

Total Adults = 173 + 978 + 249 + 249 + 134 + 153 = 1936

Total Teens = 13 + 138 + 69 + 113 + 113 + 181 = 627

c)



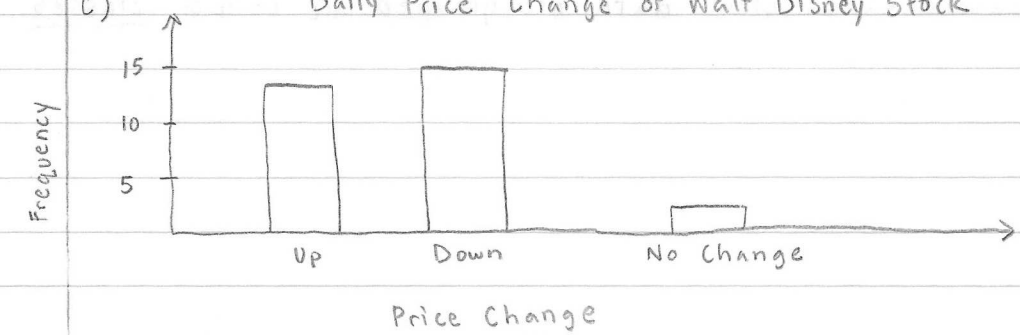
d) Adults are much more likely to send a smaller number of texts each day, which teens are much more likely to send a larger number each day.

21. a), b)

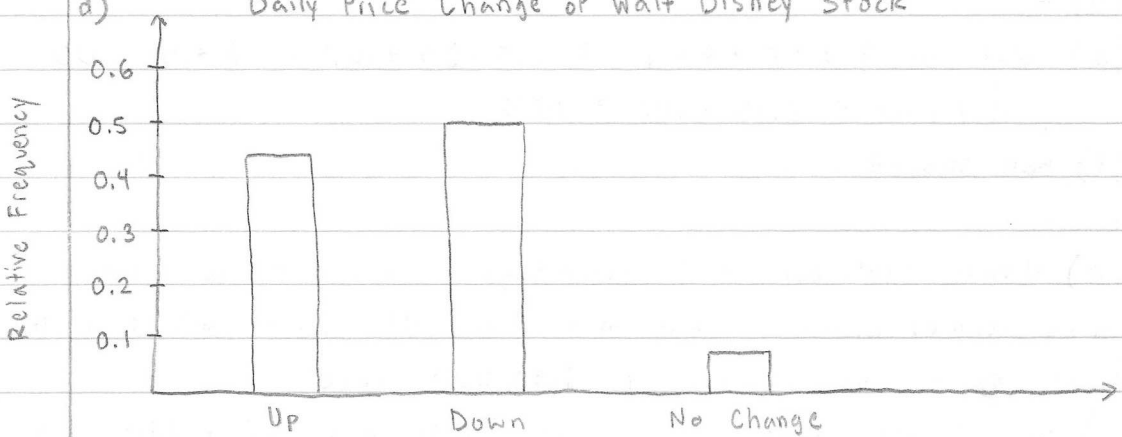
Price Change	Tally	Frequency	Relative Frequency
Up	𝍸 𝍸 | | |	13	13/30 = 0.433
Down	𝍸 𝍸 𝍸 |	15	15/30 = 0.500
No Change	| |	2	2/30 = 0.067

Total price changes = 13 + 15 + 2 = 30

c) Daily Price Change of Walt Disney Stock



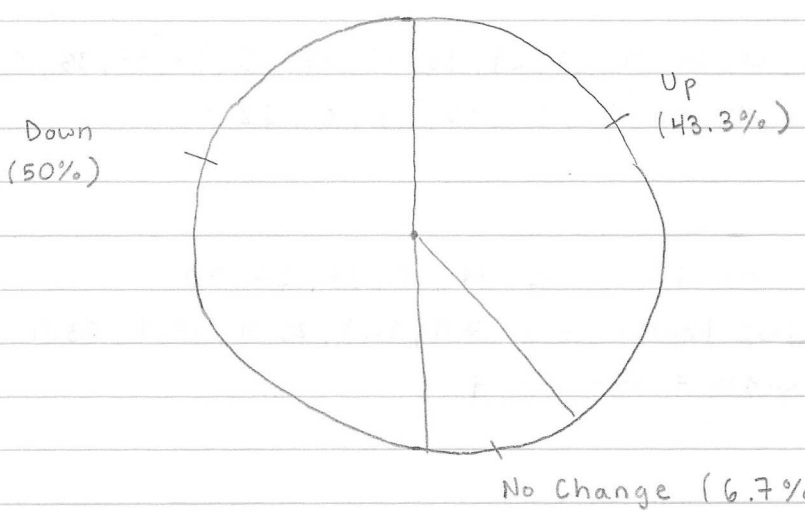
d) Daily Price Change of Walt Disney Stock



e)

Price Change	Degree Measure of Each Sector
Up	$0.433(360°) = 156°$
Down	$0.500(360°) = 180°$
No Change	$0.067(360°) = 24°$

Daily Price Change of Walt Disney Stock



2.2

1. The categories by which data are grouped are called classes.

5. True

9. a) 8
b) 2
c) 15
d) 4
e) Total rolls = $2 + 5 + 7 + 11 + 9 + 15 + 20 + 13 + 7 + 7 + 4 = 100$

$15 / 100 = 0.15 \times 100 = 15\%$

f) bell shaped

13. a) skewed right because the percentage of households with extremely high incomes is small, while most household incomes will be on the left, say in the $50,000 to $150,000 range
b) bell shaped because most scores will be in the middle with scores tapering off evenly to the right and to the left
c) skewed right because most households will have values on the left (1 to 4 people) with fewer households having more than that
d) skewed left because the majority of patients will be on the right (60+ age range) with fewer younger patients being diagnosed with Alzheimer's disease

17. 10, 11, 14, 21, 24, 24, 27, 29, 33, 35, 35, 35, 37, 37, 38, 40, 40, 41, 42, 46, 46, 48, 49, 49, 53, 53, 55, 58, 61, 62

21. a) 6
b) Lower class limits: 10, 14, 18, 22, 26, 30
Upper class limits: 13.9, 17.9, 21.9, 25.9, 29.9, 33.9
c) Class width = $14 - 10 = 4$

25. a) discrete because you obtain them by counting

b), c)

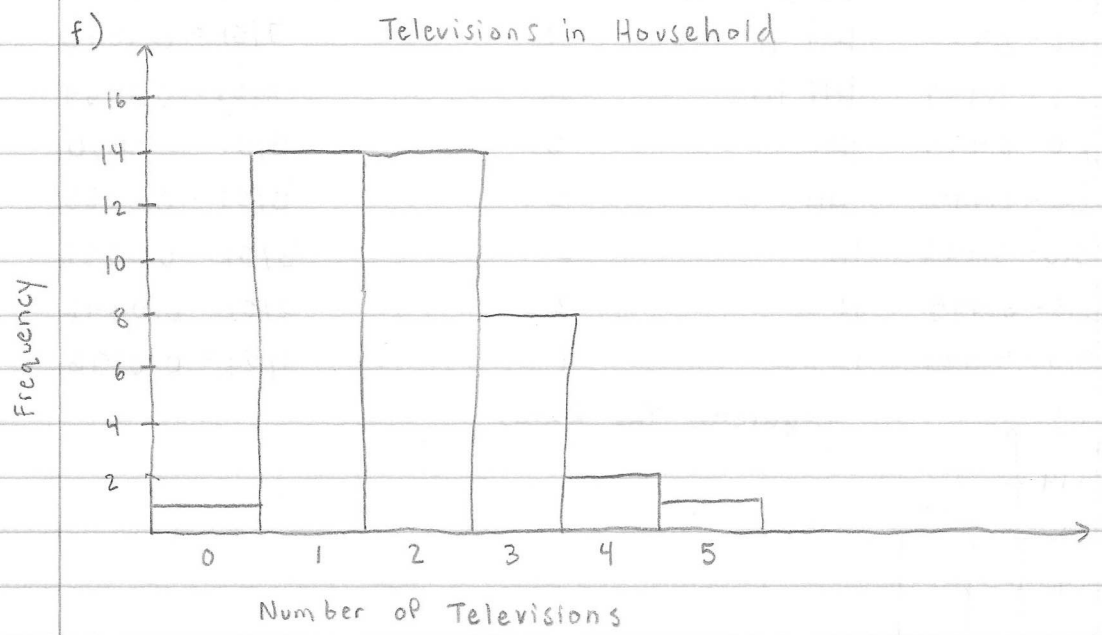
Televisions	Tally	Frequency	Relative Frequency
0	|	1	1/40 = 0.025
1	𝍸 𝍸 ||||	14	14/40 = 0.350
2	𝍸 𝍸 ||||	14	14/40 = 0.350
3	𝍸 |||	8	8/40 = 0.200
4	||	2	2/40 = 0.050
5	|	1	1/40 = 0.025

Total = 1 + 14 + 14 + 8 + 2 + 1 = 40

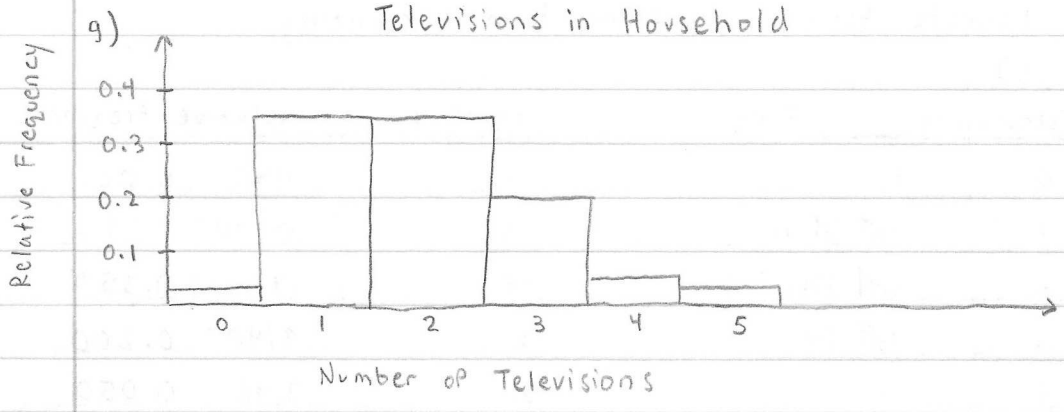
d) $\frac{8}{40} \times 100 = 20\%$

e) $\frac{2+1}{40} \times 100 = 7.5\%$

f)



g)

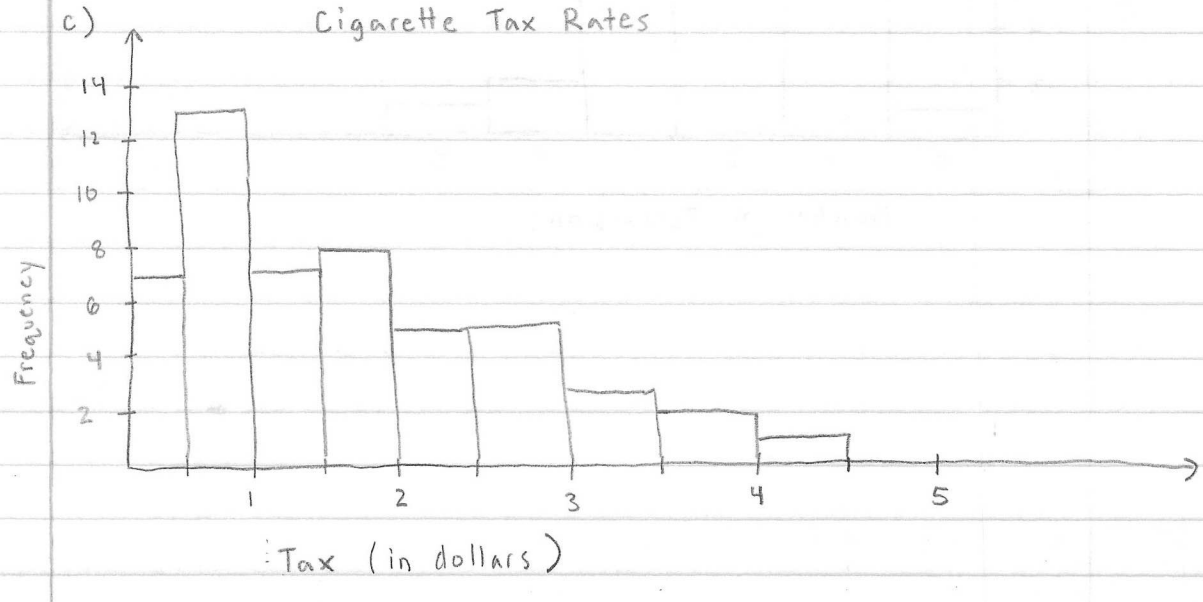


h) skewed right

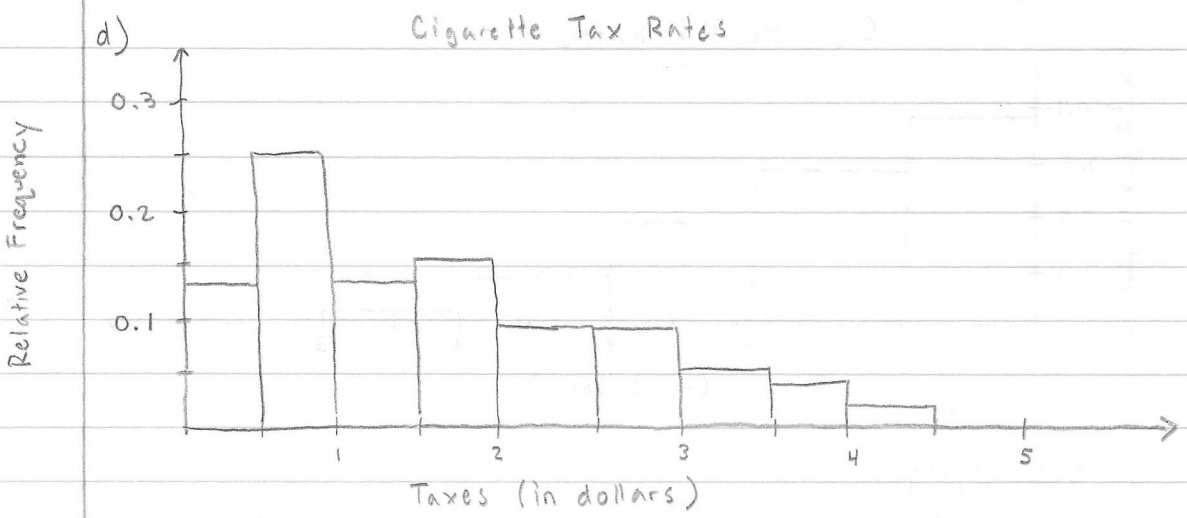
29. a), b)

Tax	Tally	Frequency	Relative Frequency
0.0 - 0.499	𝍸 ||	7	7/51 = 0.1373
0.5 - 0.999	𝍸 𝍸 |||	13	13/51 = 0.2549
1.0 - 1.499	𝍸 ||	7	7/51 = 0.1373
1.5 - 1.999	𝍸 |||	8	8/51 = 0.1569
2.0 - 2.499	𝍸	5	5/51 = 0.0980
2.5 - 2.999	𝍸	5	5/51 = 0.0980
3.0 - 3.499	|||	3	3/51 = 0.0588
3.5 - 3.999	||	2	2/51 = 0.0392
4.0 - 4.499	|	1	1/51 = 0.0196

c)



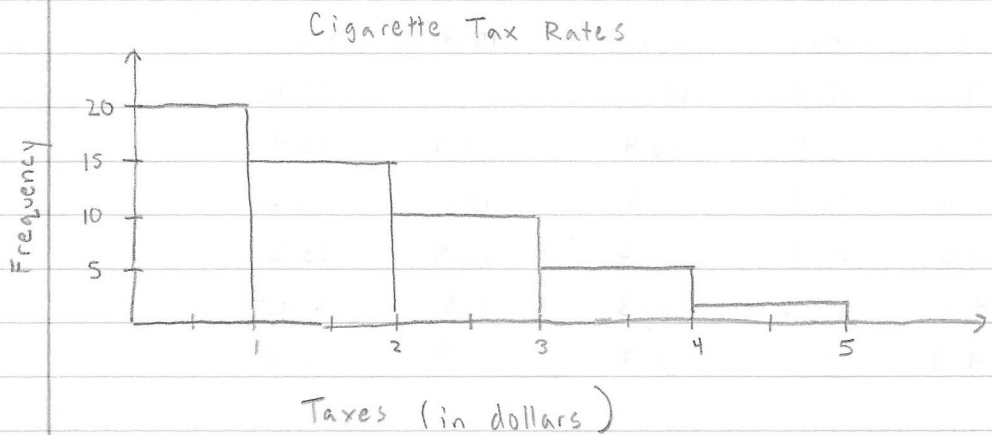
d)

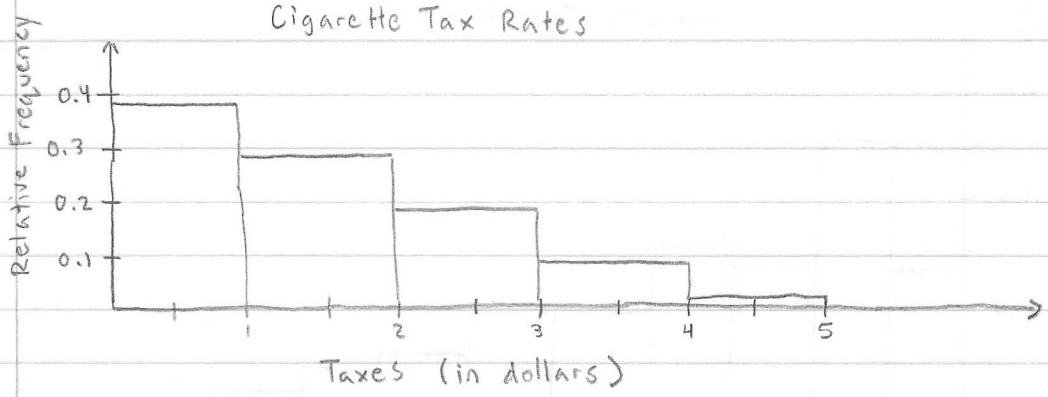


e) skewed right

f)

Tax	Frequency	Relative Frequency
0.0 - 0.999	20	20/51 = 0.3922
1.0 - 1.999	15	15/51 = 0.2941
2.0 - 2.999	10	10/51 = 0.1961
3.0 - 3.999	5	5/51 = 0.0980
4.0 - 4.999	1	1/51 = 0.0196





- skewed right

g) Both do a good job of summarizing the data

33. a)

President Ages at Inauguration

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

Legend: 4 | 2 represents 42 years

b) bell shaped

37. a)

10.9	14.2	12.4	13.6	13.0
10.5	10.3	13.1	15.7	14.9
14.1	12.8	13.3	9.9	15.6
12.3	13.9	13.4	19.4	13.4
12.2	14.8	11.9	10.1	13.6
14.6	14.8	13.5	13.9	13.2
14.0	15.2	8.3	9.0	8.9
14.9	16.0	13.7	13.9	12.8

b)

Five-Year Rate of Return

8	3 7
9	9 0
10	9 5 3 1
11	9
12	3 2 8 4 8
13	9 1 3 4 5 7 6 9 9 0 4 6 2
14	1 6 0 9 2 8 8 9
15	2 7 6
16	0
17	
18	
19	4

Five-Year Rate of Return

8	3 7
9	0 9
10	1 3 5 9
11	9
12	2 3 4 8 8
13	0 1 2 3 4 4 5 6 6 7 9 9 9
14	0 1 2 6 8 8 9 9
15	2 7 6
16	0
17	
18	
19	4

Legend: 8|3 means 8.3%

c) bell shaped

41. a)

McGwire		Bonds
	32	0 0
	33	
1 0	34	7
0 0	35	0
0 9 0 0	36	5 0 1 0 0
0 7 0 0 0	37	5 5 5 0 0 5
5 8 0 0 0 0 5 0	38	0 0 5 0 0 0
8 0 0 0 0	39	6 1 0 0 4
9 0 0	40	0 4 0 0 5 0
0 0 0 0 0	41	0 0 0 7 0 0 1 0 0 0 6 5 7 0 5 0
5 3 0 0 0 0 0	42	0 0 0 0 0 0 9 0 0
0 0 0 0 0 0 0	43	6 0 5 0 0 0 0 5
0 0 0	44	2 0 0 0 0
0 0 8 0 2 0	45	0 4
1 0 0	46	
0 8 0 0	47	
0	48	8
	49	
0	50	
0 0	51	
7	52	
	53	
	54	
0	55	

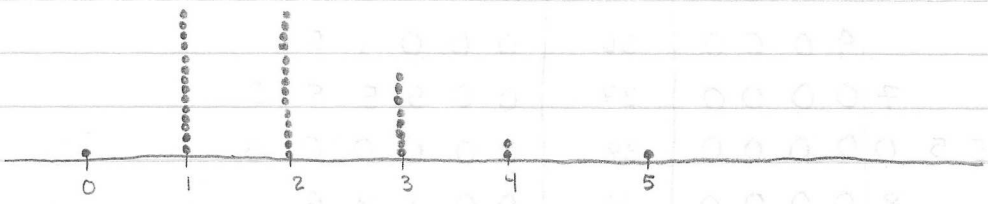
Home Run Distances

McGwire		Bonds
	32	0 0
	33	
1 0	34	7
0 0	35	0
9 0 0 0	36	0 0 0 1 5
7 0 0 0 0 0	37	0 0 5 5 5 5
8 5 5 0 0 0 0 0	38	0 0 0 0 0 5
8 0 0 0 0 0	39	0 0 1 4 6
9 0 0	40	0 0 0 0 4 5
0 0 0 0 0	41	0 0 0 0 0 0 0 0 0 0 1 5 5 6 7 7
5 3 0 0 0 0 0	42	0 0 0 0 0 0 0 0 9
0 0 0 0 0 0 0	43	0 0 0 0 0 5 5 6
0 0 0	44	0 0 0 0 2
8 2 0 0 0 0	45	0 4
1 0 0	46	
8 0 0 0	47	
0	48	8
	49	
0	50	
0 0	51	
7	52	
	53	
	54	
0	55	

Legend: 0 | 34 | 7 means 340 feet for McGwire and 347 feet for Bonds

b) The data for home run distances of McGwire has more spread than that of Bonds. McGwire's longest home run (550 feet) was further than Bonds's (488 feet).

45. Televisions in Household



2.3

1. An ogive is a graph that represents the cumulative frequency or cumulative relative frequency for the class.

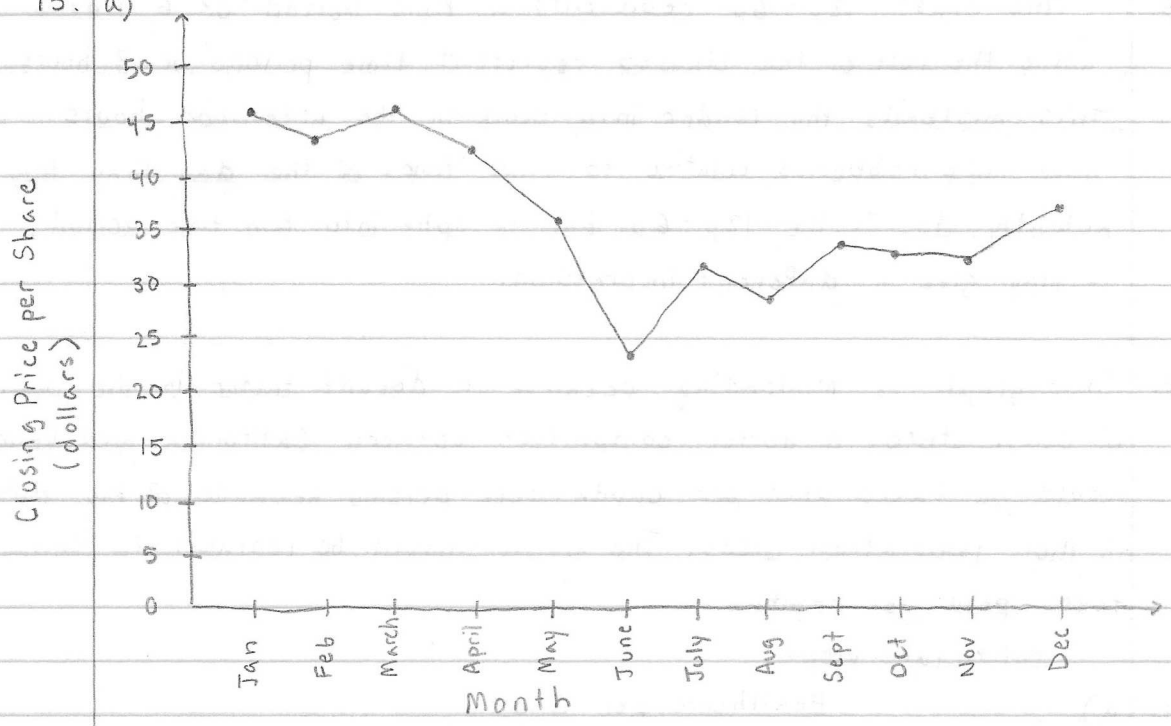
3. True

5. a) class width = 1.25 - 0.75 = 0.50, 6 classes
b) midpoint = 1.25

lower limit = 1.00
upper limit = 1.49
c) midpoint = 3.75
lower limit = 3.50
upper limit = 3.99
d) lower limit = 2.50
upper limit = 2.99
e) lower limit = 1.00
upper limit = 1.49

9. a) 9%
b) 2010
c) 2008
d) 2001, 2009
e) 1999: 1.6 + 4.2 = 5.8, 2014: 1.5 + 6.5 = 8 → ∴ 2014 more miserable
f) misery index has decreased since 2010 due to decreases in unemployment

15. a)

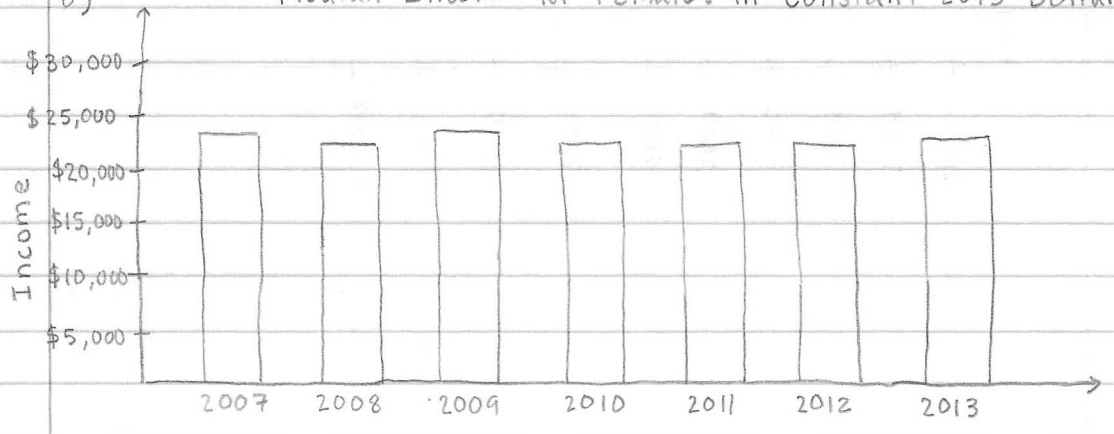


b) Percentage change $= \frac{P_2 - P_1}{P_1} = \frac{24.02 - 35.72}{35.72} = -0.328 = -32.8\%$

2.4

3. a) The vertical axis starts at \$21,500 instead of \$0. This seems to convey that median income for females declined much more than they actually did.

b)



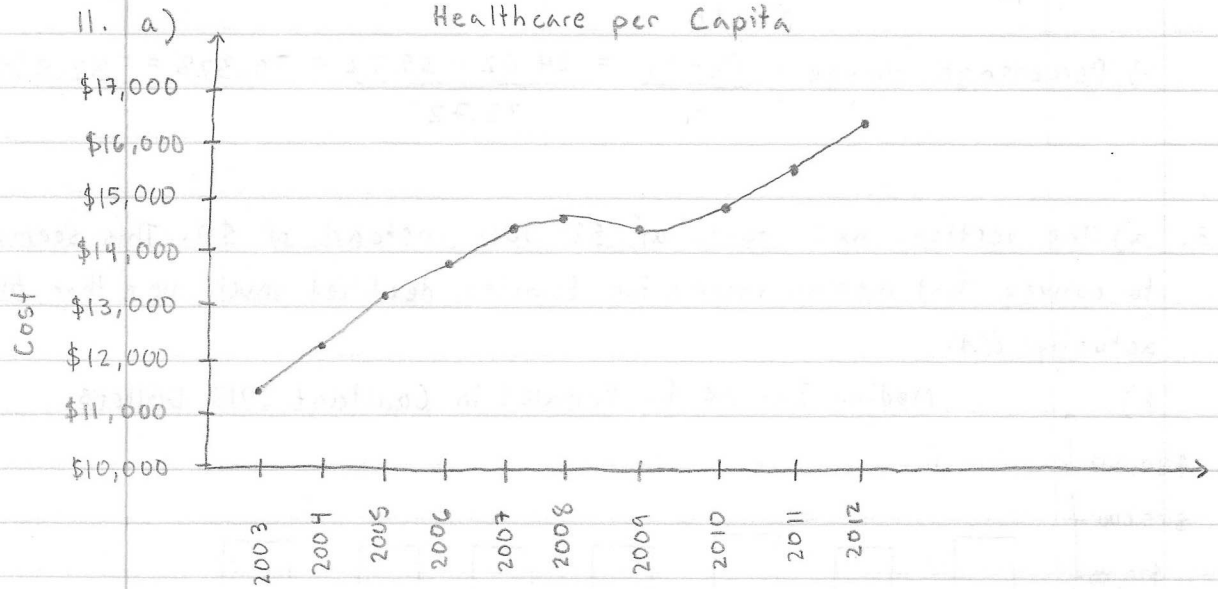
The graph indicates that the median income for females has declined, just not as significantly as the graph from part (a) seems to suggest.

5. The class 12p-6p represents a time period of 6 hours while the rest of the classes represent time periods of 3 hours. This misleads the reader into thinking the afternoon hours have more robberies relative to other times of the day than they actually do. If the 12p-6p bar is split into two time period, it may give a different impression.

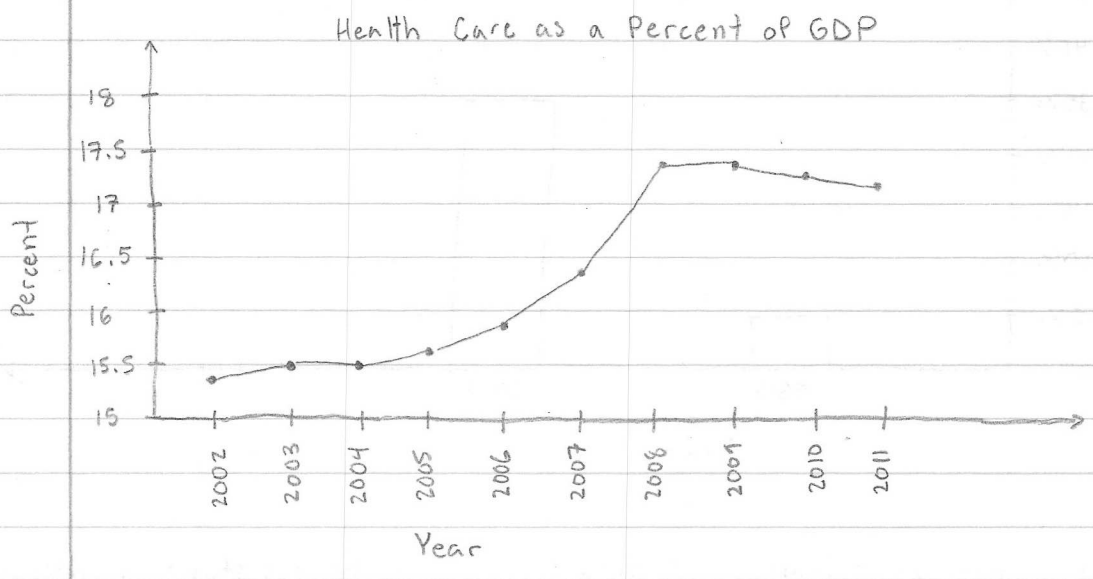
7. This graph is misleading because it doesn't factor in the population of each state. A better comparison between California and Vermont would be taxes paid per capita (per person) because of the difference in their population sizes. The graph should be redrawn to show taxes paid per capita.

Politician's view:

11. a)

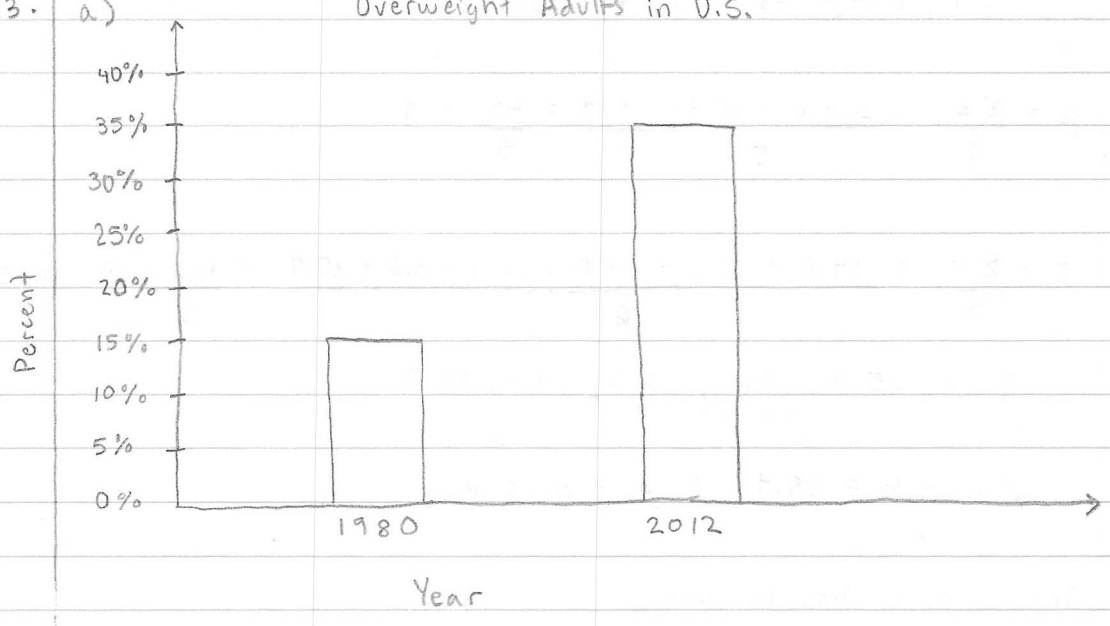


b) Healthcare industry's view:

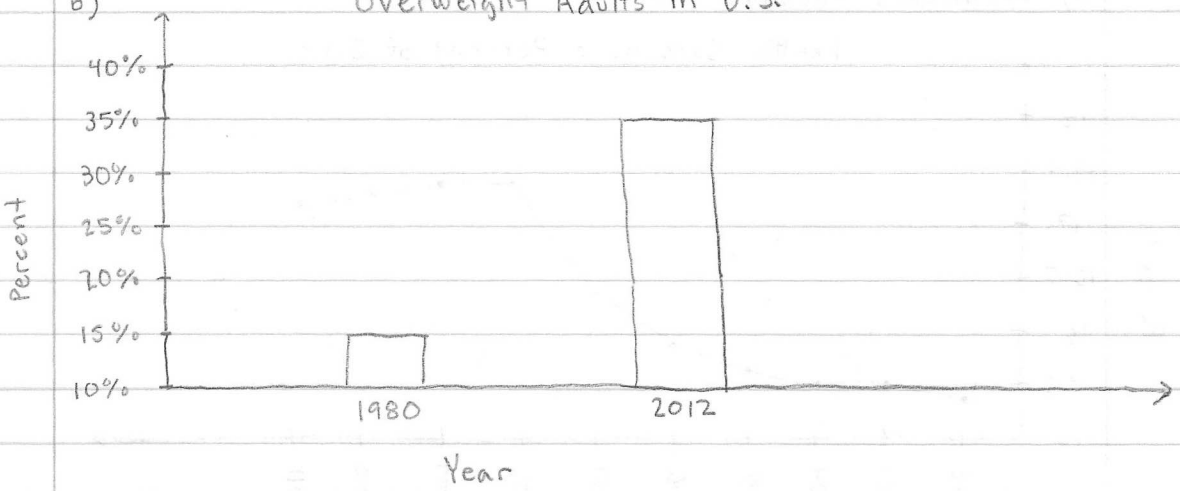


c) The variable used to measure the argument, the scale, and the starting point of the vertical axis all affect the message conveyed by the graphic.

13. a)



b) Overweight Adults in U.S.



3.1

1. A numerical summary of data is said to be resistant if extreme values (very large or small) relative to the data do not affect its value substantially.

5. The median is the mean of the values at position $\frac{n}{2} = \frac{10,000}{2} = 5000$ and $\frac{n}{2} + 1 = \frac{10000}{2} + 1 = 5000 + 1 = 5001$.

9. $\mu = \frac{\Sigma x_i}{N} = \frac{3 + 6 + 10 + 12 + 14}{5} = \frac{45}{5} = 9$

13. $\bar{x} = \frac{\Sigma x_i}{n} = \frac{34.0 + 33.2 + 37.0 + 29.4 + 23.6 + 25.9}{6} = \frac{183.1}{6} = 30.52$ mpg

23.6, 25.9, 29.4, 33.2, 34.0, 37.0

Median = $M = \frac{29.4 + 33.2}{2} = 31.3$ mpg

The sample has no mode.

17. a) Mean > Median because skewed right

b) Mean = Median because symmetric

c) Mean < Median because skewed left

21. a) $\mu = \frac{\Sigma x_i}{N} = \frac{76 + 60 + 60 + 81 + 72 + 80 + 80 + 68 + 73}{9} = \frac{650}{9} = 72.22$ beats per min

b) Sample 1: Kevin, Tammy, Kathy

$\bar{x} = \frac{\Sigma x_i}{n} = \frac{80 + 68 + 73}{3} = \frac{221}{3} = 73.67$ beats per min

Sample 2: Perpectual, Janette, Kathy

$\bar{x} = \frac{\Sigma x_i}{n} = \frac{76 + 80 + 73}{3} = \frac{229}{3} = 76.33$ beats per min

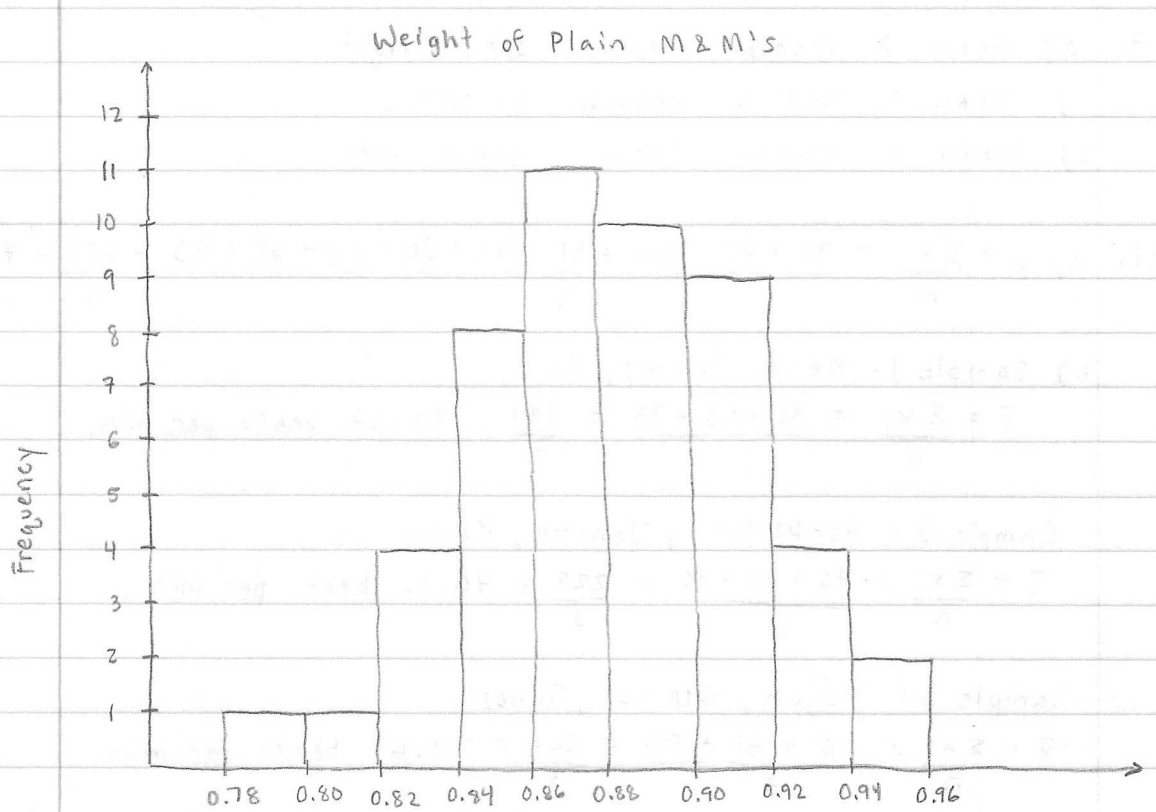
Sample 3: Megan, Clarice, Janette

$\bar{x} = \frac{\Sigma x_i}{n} = \frac{60 + 81 + 80}{3} = \frac{221}{3} = 73.67$ beats per min

c) All three samples result in a sample mean that overestimate the population mean. None underestimate it or are equal to it.

25.

Weight	Tally	Frequency
0.78 - 0.799	|	1
0.80 - 0.819	|	1
0.82 - 0.839	||||	4
0.84 - 0.859	𝍸 |||	8
0.86 - 0.879	𝍸 𝍸 |	11
0.88 - 0.899	𝍸 𝍸	10
0.90 - 0.919	𝍸 ||||	9
0.92 - 0.939	||||	4
0.94 - 0.959	||	2



The distribution is symmetric, so the mean is the better measure of central tendency.

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{0.87 + 0.91 + 0.89 + 0.83 + \ldots + 0.87}{50} = \frac{43.73}{50} = 0.875 \text{ g}$$

Median = M = 0.875 g

(calculated using TI-84 Plus)

29. a)

Response	Tally	Frequency
Conservative	𝍸 III	8
Moderate	𝍸 𝍸 II	12
Liberal	𝍸 𝍸	10

The mode is moderate.

b) Yes to avoid response bias.

33. $\mu = 82 \quad \bar{x} = 84$

$$\frac{\Sigma x_i}{19} = 84$$

$$\Sigma x_i = 1596$$

$$\mu = 82 = \frac{\Sigma x_i}{20} = \frac{1596 + x_{20}}{20}$$

$$82 \times 20 = 1596 + x_{20}$$

$$1640 = 1596 + x_{20}$$

$$1640 - 1596 = 1596 - 1596 + x_{20}$$

$$44 = x_{20}$$

3.2

1. The sum of deviations about the mean always equals zero.

3. True

5. $$s = \sqrt{\frac{\Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n}}{n-1}} = \sqrt{\frac{(20^2+13^2+4^2+8^2+10^2) - \frac{(20+13+4+8+10)^2}{5}}{5-1}}$$

$$= \sqrt{\frac{749 - \frac{3025}{5}}{4}} = \sqrt{\frac{749-605}{4}} = \sqrt{\frac{144}{4}}\sqrt{36} = 6$$

$s^2 = 36$

7. $$\sigma = \sqrt{\frac{\Sigma x_i^2 - \frac{(\Sigma x_i)^2}{N}}{N}} = \sqrt{\frac{(3^2+6^2+10^2+12^2+14^2) - \frac{(3+6+10+12+14)^2}{5}}{5}}$$

$$= \sqrt{\frac{485 - \frac{2025}{5}}{5}} = \sqrt{\frac{485-405}{5}} = \sqrt{\frac{80}{5}} = \sqrt{16} = 4$$

$\sigma^2 = 16$

9. $$s = \sqrt{\frac{\Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n}}{n-1}} = \sqrt{\frac{(6^2+52^2+13^2+49^2+35^2+25^2+31^2+29^2+31^2+29^2) - \frac{(6+52+13+49+35+25+31+29+31+29)^2}{10}}{10-1}}$$

$$= \sqrt{\frac{10764 - \frac{90000}{10}}{9}} = \sqrt{\frac{10764-9000}{9}} = \sqrt{\frac{1764}{9}} = \sqrt{196} = 14$$

$s^2 = 196$

11. Range = 37.0 − 23.6 = 13.4 mpg

$$s = \sqrt{\frac{\Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n}}{n-1}} = \sqrt{\frac{(34.0^2 + 33.2^2 + 37.0^2 + 29.4^2 + 23.6^2 + 25.9^2) - \frac{(34.0+33.2+37.0+29.4+23.6+25.9)^2}{6}}{6-1}}$$

$$= \sqrt{\frac{5719.37 - \frac{33525.61}{6}}{5}} = \sqrt{\frac{5719.37 - 5587.60}{5}} = \sqrt{\frac{131.77}{5}} = \sqrt{26.354} = 5.13 \text{ mpg}$$

$s^2 = 26.35$ mpg

15. (b) has a higher standard deviation because the data go from 30 to 75. In (a) they only go from 40 to 60.

19. a) $\sigma = 7.7$ beats per min
Calculated on TI-84 Plus

b) Sample 1: Perpectual, Kevin, Kathy
$s = 3.5$ beats per min (TI-84)
Sample 2: Jeff, Crystal, Kathy
$s = 7.2$ beats per min (TI-84)
Sample 3: Megan, Crystal, Kevin
$s = 10.0$ beats per min (TI-84)

c) Samples 1 and 2 underestimate. Sample 3 overestimates.

21. a) Ethan: Range = 24 − 5 = 19 fish

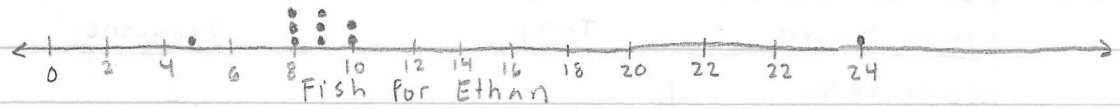
$$\mu = \frac{9+24+8+9+5+8+9+10+8+10}{10} = \frac{100}{10} = 10 \text{ fish}$$

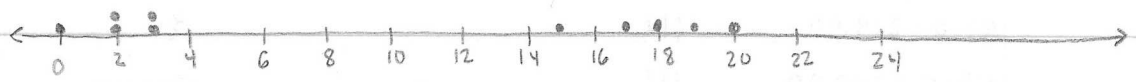
Drew: Range = 20 − 1 = 19 fish

$$\mu = \frac{15+2+3+18+20+1+17+2+19+3}{10} = \frac{100}{10} = 10 \text{ fish}$$

There is no difference based on these measures.

b)





Fish for Drew

Ethan seems more consistent.

c) Ethan: σ = 4.9 fish (TI-84)

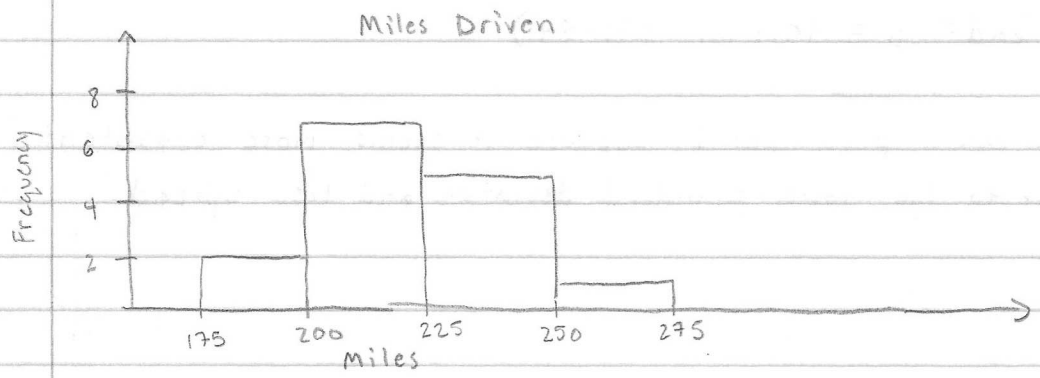
Drew: σ = 7.9 fish (TI-84)

No change, Ethan has the more consistent record.

d) The range is not resistant, and is also calculated using only two values from the data set.

25. Car 1:

Miles Driven	Tally	Frequency
175.0 - 199.99	II	2
200.0 - 224.99	~~IIII~~ II	7
225.0 - 249.99	~~IIII~~	5
250.0 - 274.99	I	1



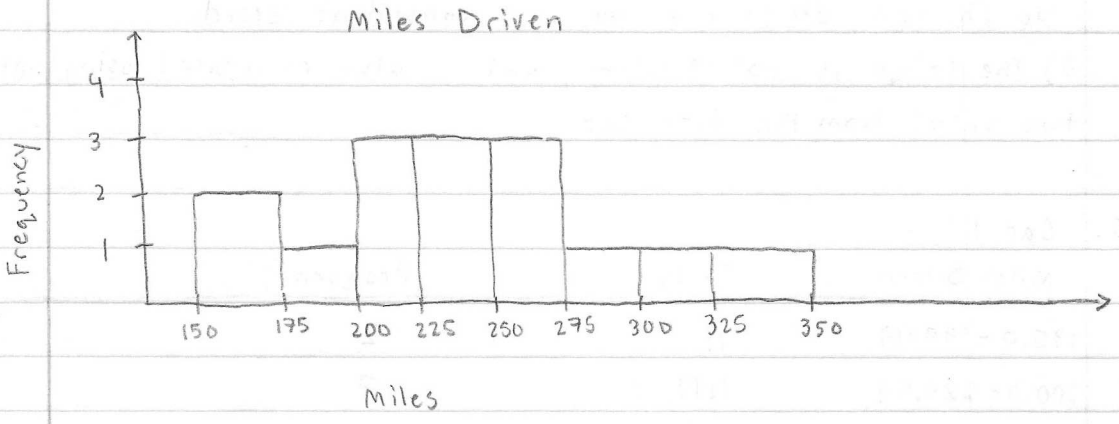
Shape: symmetric

Center: $\bar{x}$ = 223.5 mi (TI-84)

Spread: s = 21.8 mi (TI-84)

Car 2:

Miles Driven	Tally	Frequency
150.0 - 174.99	II	2
175.0 - 199.99	I	1
200.0 - 224.99	III	3
225.0 - 249.99	III	3
250.0 - 274.99	III	3
275.0 - 299.99	I	1
300.0 - 324.99	I	1
325.0 - 349.99	I	1



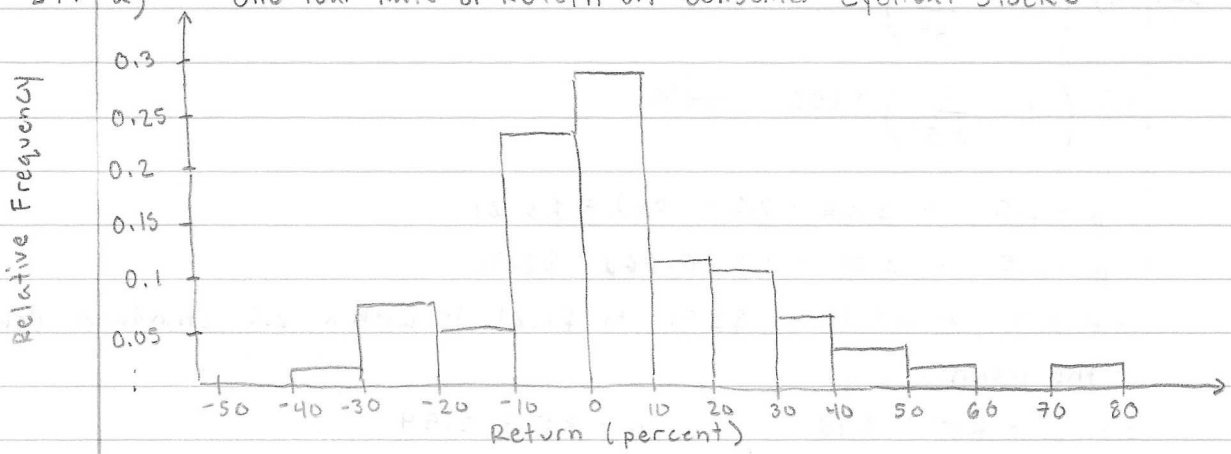
Shape: bell shaped

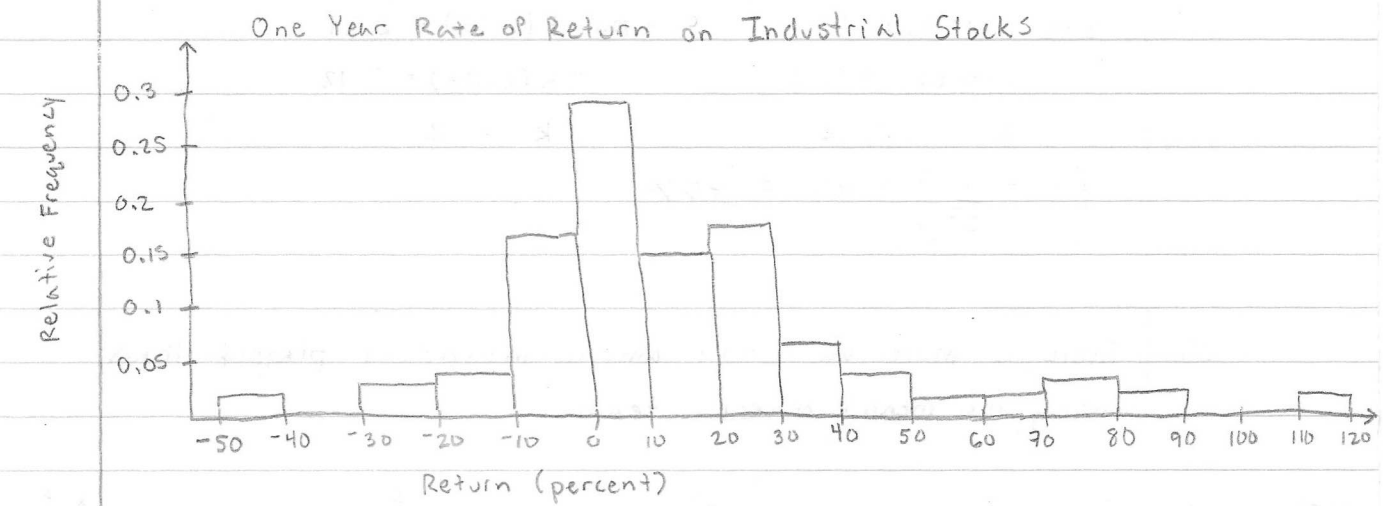
Center: $\bar{x}$ = 237.2 mi (TI-84)

Spread: s = 49.1 mi (TI-84)

I would pick car 1 because it seems more consistent due to the lower standard deviation and less spread.

27. a)





It seems that the industrial stocks have more dispersion.

b) Consumer Cyclical: $\bar{x} = 6.595\%$

$M = 3.915\%$

Industrial: $\bar{x} = 14.425\%$

$M = 9.595\%$

Industrial stocks have the higher mean and median rate of return.

c) Consumer Cyclical: $s = 19.078\%$

Industrial: $s = 23.851\%$

The sector with the higher rate of return, industrial, also has the higher risk. The risk is worth it if you have a high appetite for risk as an investor.

35. a) $\left(1 - \frac{1}{3^2}\right) \times 100 = 88.9\%$

b) $\left(1 - \frac{1}{2.5^2}\right) \times 100 = 84\%$

$\mu + 2.5\sigma = 3.06 + 2.5(0.06) = \$3.21$

$\mu - 2.5\sigma = 3.06 - 2.5(0.06) = \$2.91$

Gasoline priced from \$2.91 to \$3.21 is within 2.5 standard deviations of the mean.

c) $\mu + k\sigma = 3.18$ $\qquad$ $\mu - k\sigma = 2.94$

$3.06 + k(0.06) = 3.18$ $\qquad$ $3.06 - k(0.06) = 2.94$

$k(0.06) = 0.12$ $\qquad$ $-k(0.06) = -.12$

$k = 2$ $\qquad$ $k = 2$

$\left(1 - \frac{1}{2^2}\right) \times 100 = 75\%$

37. There is more variation among individual players than among the means of each team.

3.3 1.

Savings	x_i	f_i	$x_i f_i$
\$0 - \$19,999	\$10,000	344	3,440,000
\$20,000 - \$39,999	\$30,000	98	2,940,000
\$40,000 - \$59,999	\$50,000	52	2,600,000
\$60,000 - \$79,999	\$70,000	19	1,330,000
\$80,000 - \$99,999	\$90,000	13	1,170,000
\$100,000 - \$119,999	\$110,000	6	660,000
\$120,000 - \$139,999	\$130,000	2	260,000
		$\Sigma f_i = 534$	$\Sigma x_i f_i = 12,400,000$

$\bar{x} = \frac{\Sigma x_i f_i}{\Sigma f_i} = \frac{12,400,000}{534} = \$23,221$

* (cont'd)

$\bar{x}$	$x_i - \bar{x}$	$(x_i - \bar{x})^2 f_i$
\$23,221	-13221	60,129,425,300
\$23,221	6779	45,035,744,18
\$23,221	26779	37,289,971,730
\$23,221	46779	41,577,221,980
\$23,221	66779	57,972,652,930
\$23,221	86779	45,183,569,050
\$23,221	106779	22,803,509,680

$\Sigma (x_i - \bar{x})^2 f_i = 269,459,925,100$

$$s = \sqrt{\frac{\Sigma (x_i - \bar{x})^2 f_i}{(\Sigma f_i) - 1}} = \sqrt{\frac{269,459,925,100}{534 - 1}} = \$22,484.50$$

3.

Temperature (°F)	x_i	f_i	$x_i f_i$ *
61-64	63	31	1953
65-67	66.5	67	4455.5
68-69	69	198	13662
70	70	195	13650
71-72	72	120	8640
73-76	75	89	6675
77-80	79	50	3950
		$\Sigma f_i = 750$	$\Sigma x_i f_i = 52985.5$

$$\bar{x} = \frac{\Sigma x_i f_i}{\Sigma f_i} = \frac{52985.5}{750} = 70.6°$$

* (cont'd)

$(x_i - \bar{x})$	$(x_i - \bar{x})^2 f_i$
-7.6	$(-7.6)^2 (31) = 1790.56$
-4.1	$(-4.1)^2 (67) = 1126.27$
-1.6	$(-1.6)^2 (198) = 506.88$
-.6	$(-.6)^2 (195) = 70.2$
1.4	$(1.4)^2 (120) = 235.2$
4.4	$(4.4)^2 (89) = 1723.04$
8.4	$(8.4)^2 (50) = 3528$
	$\Sigma (x_i - \bar{x})^2 f_i = 8980.15$

$$s = \sqrt{\frac{8980.15}{750 - 1}} = 3.5°F$$

5. a)

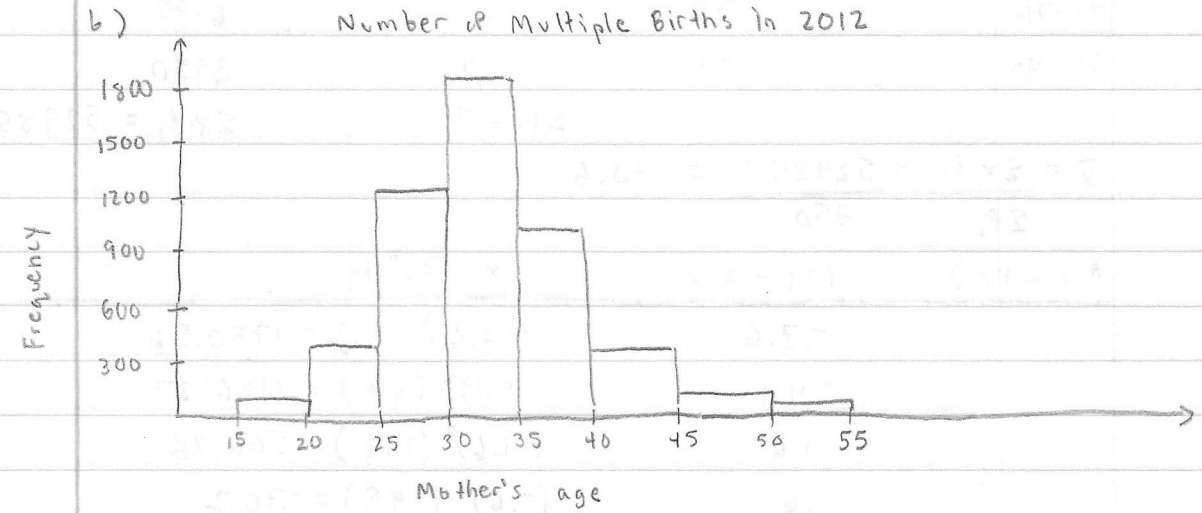
Age	x_i	f_i	$x_i f_i$	$\bar{x}$	$(x_i - \bar{x})$	$(x_i - \bar{x})^2 f_i$
15-19	17.5	44	770	32.2	-14.7	9507.96
20-24	22.5	404	9090	32.2	-9.7	38012.36
25-29	27.5	1204	33110	32.2	-4.7	26596.36
30-34	32.5	1872	60840	32.2	.3	168.48
35-39	37.5	1000	37500	32.2	5.3	28090
40-44	42.5	332	14110	32.2	10.3	35221.88
45-49	47.5	44	2090	32.2	15.3	10299.96
50-54	52.5	19	997.5	32.2	20.3	7829.71

$\Sigma f_i = 4919$ $\Sigma x_i f_i = 158507.5$ $\Sigma (x_i - \bar{x})^2 f_i = 155726.71$

$$\bar{x} = \frac{158507.5}{4919} = 32.2 \text{ years}$$

$$s = \sqrt{\frac{155726.71}{4919 - 1}} = 5.6 \text{ years}$$

b)



c) $\bar{x} + 2s = 32.2 + 2(5.6) = 43.4$

$\bar{x} - 2s = 32.2 - 2(5.6) = 21$

between 21.0 and 43.4 years of age

7.

Tax	x_i	f_i	$x_i f_i$	μ	$(x_i - \mu)$	$(x_i - \mu)^2 f_i$
0.0 - 0.499	0.25	7	1.75	1.583	-1.333	12.438
0.5 - 0.999	0.75	13	9.75	1.583	-.833	9.021
1.0 - 1.499	1.25	7	8.75	1.583	-.333	.776
1.5 - 1.999	1.75	8	14	1.583	.167	.223
2.0 - 2.499	2.25	5	11.25	1.583	.667	2.224
2.5 - 2.999	2.75	5	13.75	1.583	1.167	6.809
3.0 - 3.499	3.25	3	9.75	1.583	1.667	8.337
3.5 - 3.999	3.75	2	7.5	1.583	2.167	9.392
4.0 - 4.499	4.25	1	4.25	1.583	2.667	7.113
		$\Sigma f_i = 51$	$\Sigma x_i f_i = 80.75$			$\Sigma (x_i - \mu)^2 f_i = 56.333$

Grouped data:

$$\mu = \frac{80.75}{51} = \$1.583$$

$$\sigma = \sqrt{\frac{56.333}{51 - 1}} = \$1.061$$

Raw data:

$\mu = \$1.531$ (TI-84)

$\sigma = \$1.003$ (TI-84)

The approximations are close to the actual values.

9. $$GPA = \bar{x}_w = \frac{\Sigma w_i x_i}{\Sigma w_i} = \frac{5(3) + 3(4) + 4(4) + 3(2)}{5 + 3 + 4 + 3} = \frac{49}{15} = 3.27$$

11. $$\text{Cost per pound} = \bar{x}_w = \frac{4(3.50) + 3(2.75) + 2(2.25)}{4 + 3 + 2} = \frac{26.75}{9} = \$2.97$$

13. a) Male:

Age	x_i	f_i	$x_i f_i$	$(x_i - \mu)$	$(x_i - \mu)^2 f_i$
0-9	5	20,700,000	103,500,000	-32.1	21329487000
10-19	15	21,367,000	320,505,000	-22.1	10435856470
20-29	25	21,417,000	535,425,000	-12.1	3135662970
30-39	35	19,455,000	680,925,000	-2.1	85796550
40-49	45	20,839,000	937,755,000	7.9	1300561990
50-59	55	20,785,000	1,143,175,000	17.9	66597 21850
60-69	65	14,739,000	958,035,000	27.9	11472984990
70-79	75	7,641,000	573,075,000	37.9	10975608810
≥ 80	85	4,230,000	359,550,000	47.9	970535400
		$\Sigma f_i = 151,173,000$	$\Sigma x_i f_i = 5,611,945,000$		$\Sigma (x_i - \mu^2) f_i =$ 75101034930

$\mu = \frac{5611945000}{151173000} = 37.1$ years

$\sigma = \sqrt{\frac{75101034930}{151173000}} = 22.3$ years

b) Female:

"	"	f_i	$x_i f_i$	$(x_i - \mu)$	$(x_i - \mu)^2 f_i$
		19,826,000	99130000	-34	22918856000
		20,475,000	307125000	-24	11793600000
		21,355,000	533875000	-14	4185580000
		20,011,000	700385000	-4	320176000
		21,532,000	968940000	6	775152000
		22,058,000	1213190000	16	5646848000
		16,362,000	1063530000	26	11060712000
		9,474,000	710550000	36	12278304000
		6,561,000	557685000	46	13883076000
		$\Sigma f_i = 157654000$	$\Sigma x_i f_i = 6154310000$		$\Sigma (x_i - \mu)^2 f_i =$ 82862304000

$\mu = \frac{6154310000}{157654000} = 39.0$ years

$\sigma = \sqrt{\frac{82862304000}{157654000}} = 22.9$ years

c) Females

d) Females

15.

Class	Freq.	Cum. Freq.
0 - 499	5	5
500 - 999	17	22
1000 - 1499	36	58
1500 - 1999	121	179
2000 - 2499	119	298
2500 - 2999	81	379
3000 - 3499	47	426
3500 - 3999	45	471
4000 - 4499	22	493
4500 - 4999	7	500

The median is in class 2000 - 2499.

$$M = L + \frac{\frac{n}{2} - CF}{f}(i) = 2000 + \frac{\frac{500}{2} - 179}{119}(500)$$

$$= 2000 + \frac{71}{119}(500) = 2000 + 298.3 = 2298.3 \text{ sq ft}$$

17. Modal class: \$0 - \$19,999

3.4 1. z-score

34 week:

5. $z = \frac{x - \mu}{\sigma} = \frac{2400 - 2600}{660} = -.30$

40 week:

$z = \frac{3300 - 3500}{470} = -.43$

The 40 week gestation baby weighs less relative to the gestation period.

9. Kershaw:

$$z = \frac{1.77 - 3.430}{0.721} = -2.30$$

Hernandez:

$$z = \frac{2.14 - 3.598}{0.762} = -1.91$$

Kershaw because he was 2.30 standard deviations below the mean while Hernandez was 1.91 below.

13. $200 + 1.5(26) = 239$

17. a) 25% of the states have a violent crime rate that is 252.4 crimes per 100,000 population or less, and 75% have a rate that is more. 50% of the states have a rate that is 333.8 or less, and 50% have a rate that is more. 75% of the states have a rate that is 454.5 or less, and 25% have a rate that is more.

b) $IQR = Q_3 - Q_1 = 454.5 - 252.4 = 202.1$

The middle 50% have a range of 202.1 crimes per 100,000 population.

c) Upper Fence $= Q_3 + 1.5(IQR) = 454.5 + 1.5(202.1) = 757.65$

Yes

d) Skewed right because the difference between Q_1 and Q_2 is less than Q_2 and Q_3, and there is an outlier on the right.

21. a) $\bar{x} = 38.775$ mpg

$s = 3.416$ mpg

$$z = \frac{36.3 - 38.775}{3.416} = -0.72$$

b) $Q_1 = 36.85$ mpg

$Q_2 = 38.35$ mpg

$Q_3 = 41$ mpg

c) IQR = 41 - 36.85 = 4.15 mpg

The middle 50% have a range of 4.15 mpg

d) LF = 36.85 - 1.5 (4.15) = 30.625

UF = 41 + 1.5(4.15) = 47.225

No outliers

25. $Q_1 = 433$

$Q_2 = 466$

$Q_3 = 489.5$

IQR = 489.5 - 433 = 56.5

UF = 489.5 + 1.5(56.5) = 574.25

Cutoff point is 574 minutes

29.

Student	z-score
Perpectual	0.49
Megan	-1.58
Jeff	-1.58
Clarice	1.14
Crystal	-0.03
Janette	1.01
Kevin	1.01
Tommy	-0.55
Kathy	0.10

The mean of the z-scores is 0 and the standard deviation is 1.

33. Because the percentile of a score is rounded to the nearest integer, it is possible for two different scores to have the same percentile.

37. Comparing z-scores allows a comparison of the number of standard deviations both observations are from the mean that is unitless.

3.5

1. The five-number summary of a set of data consists of the smallest data value, Q_1, the median, Q_3, and the largest data value.

3. a) skewed right

b) 0, 1, 3, 6, 16

5. a) 40

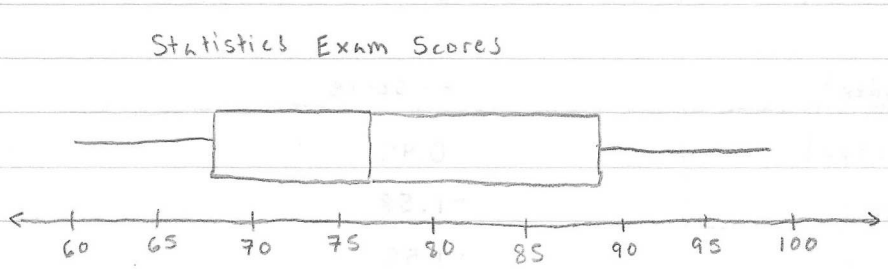
b) 52

c) y because the IQR is larger

d) symmetric because the median is in the middle of Q_1, Q_3

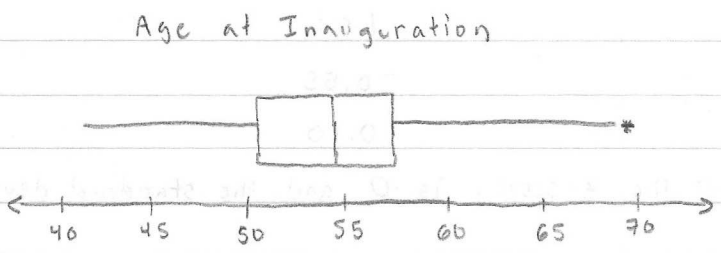
e) skewed right because the median is closer to Q_1 than Q_3

7.



9. a) 42, 50.5, 54.5, 57.5, 69

b)



c) symmetric with an outlier

$IQR = 57.5 - 50.5 = 7$

$LF = 50.5 - 1.5(7) = 40$

$UF = 57.5 + 1.5(7) = 68$

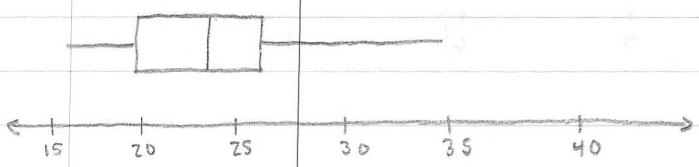
11. a) 16, 20, 23.5, 26, 35

IQR = 26 - 20 = 6

LF = 20 - 1.5(6) = 11

UF = 26 + 1.5(6) = 35

Age of Mother at Birth



b) slightly skewed right

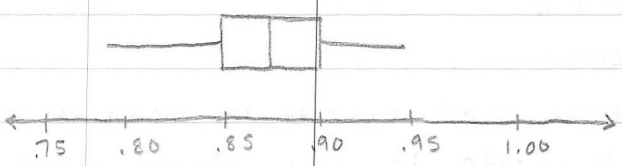
13. .79, .85, .875, .9, .95

IQR = .9 - .85 = .05

LF = .85 - 1.5(.05) = .775

UF = .9 + 1.5(.05) = .975

Weight of M & M's



The distribution is symmetric because the range from MINIMUM to M is roughly the same as M to MAXIMUM, and the range from Q_1 to M is the same as from M to Q_3.

15. a) Centrum: 2.15, 2.6, 3.02, 3.53, 4.33

IQR = 3.53 - 2.6 = .93

LF = 2.6 - 1.5(.93) = 1.205

UF = 3.53 + 1.5(.93) = 4.925

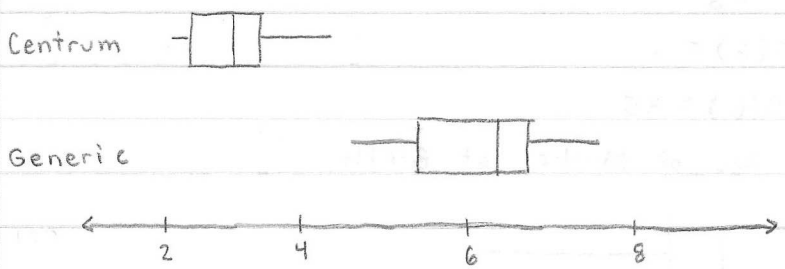
Generic: 4.97, 5.57, 6.315, 6.73, 7.58

IQR = 6.73 - 5.57 = 1.16

LF = 5.57 - 1.5(1.16) = 3.83

UF = 6.73 + 1.5(1.16) = 8.47

Dissolving Time of Vitamins



b) Generic

c) Centrum

17. a) Depth: $\mu = 19.483$ km $Q_1 = 2.99$ km

$M = 7.26$ km $Q_3 = 15.38$ km

Range = 534.05 km

$\sigma = 41.4981$ km

Magnitude: $\mu = 1.6967$ $Q_1 = 0.9$

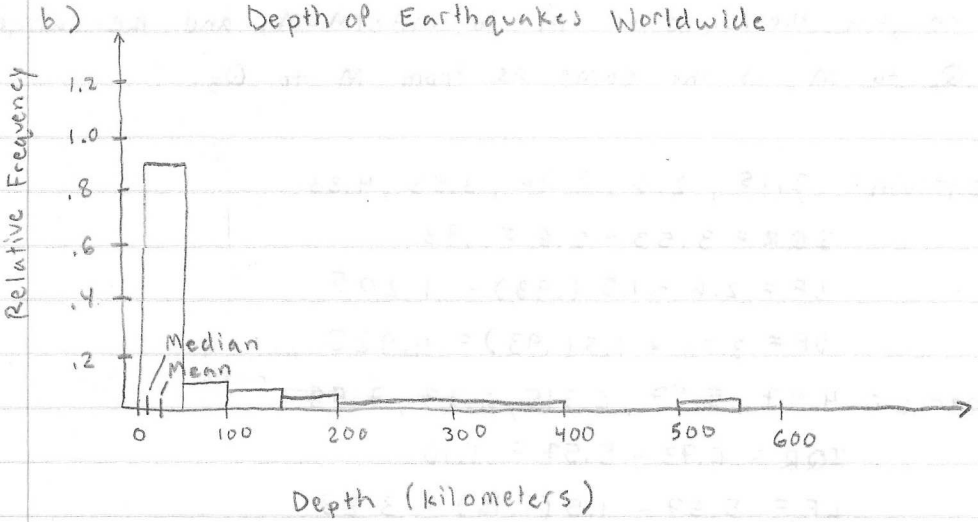
$M = 1.38$ $Q_3 = 2.01$

Range = 6.8

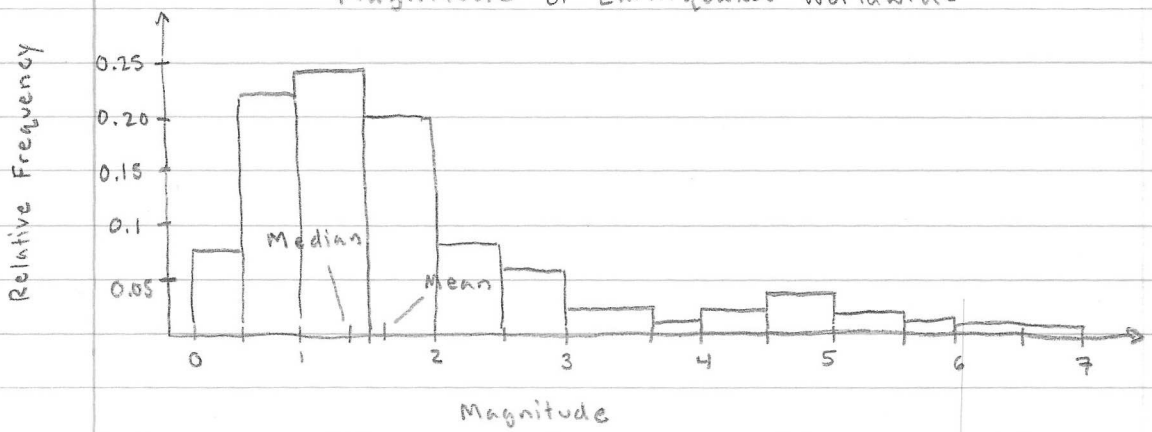
$\sigma = 1.2061$

Depth is likely skewed right because the mean is much larger than the median. Magnitude is likely skewed right because the mean is much larger than the median.

b) Depth of Earthquakes Worldwide

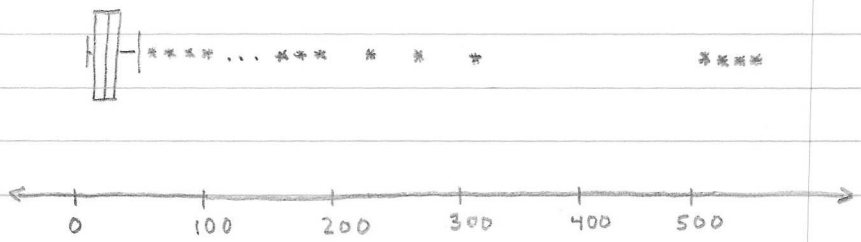


Magnitude of Earthquakes Worldwide

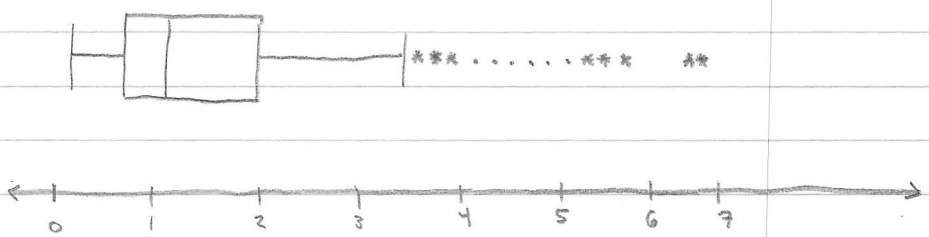


c) Both data sets have many outliers.

Depth of Earthquakes Worldwide



Magnitude of Earthquakes Worldwide



d) Depth: LF = -15.595

UF = 33.965

Magnitude: LF = -0.765

UF = 3.675