1.3 Data from a medical study contain values of many variables for each of the people who were the subjects of the study. Which of the following variables are categorical and which are quantitative?

(a) Gender (female or male) categorical
(b) Age (years) quantitative
(c) Race (Asian, black, white, or other) categorical
(d) Smoker (yes or no) categorical
(e) Systolic blood pressure (millimeters of mercury) quantitative
(f) Level of calcium in the blood (micrograms per milliliter) quantitative

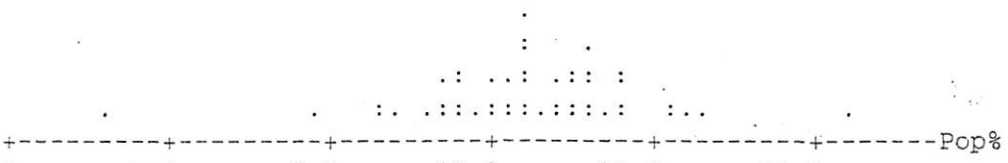
Dotplots and histograms

EXAMPLE 1.2

dotplot

Table 1.1 presents the percent of residents aged 65 years and over in each of the 50 states. One way to quickly visualize a data set is to construct a ***dotplot.*** To make a dotplot, draw a horizontal line to represent the variable and impose a number scale for the values of the variable. Then mark a dot at the appropriate place for each observation. If you enter the data from Table 1.1 into a Minitab worksheet and ask for a dotplot, you would see the following:

```
MTB > Dotplot 'Pop%';
```




```
MTB >
```

range

Notice that you need not begin with zero on the left; simply cover the *range* of the data. If you sort the data—the keystrokes on the TI-83 are STATS / 2:SortA(L_1)—you see that the smallest percent is 4.9 (Alaska) and the largest percent is 18.6 (Florida). We say that the range of the data is from 4.9 to 18.6. If you are constructing a dotplot by hand, you would mark a horizontal scale that extends from about 4 to about 19.

histogram

Sometimes quantitiative variables take so many values that a graph of the distribution is clearer if nearby values are grouped together. A ***histogram*** is the most common graph of distributions with one quantitative variable. To illus-

Dotplots

A dotplot is a simple way to display numerical data when the data set is reasonably small. Each observation is represented by a dot above the location corresponding to its value on a horizontal measurement scale. When a value occurs more than once, there is a dot for each occurrence and these dots are stacked vertically.

Dotplots

When to Use: Small numerical data sets

How to Construct:

1. Draw a horizontal line and mark it with an appropriate measurement scale.
2. Locate each value in the data set along the measurement scale, and represent it by a dot. If there are two or more observations with the same value, stack the dots vertically.

(*continued*)

What to Look For: Dotplots convey information about a representative or typical value in the data set, the extent to which the data values spread out, the nature of the distribution of values along the number line, and the presence of unusual values in the data set.

EXAMPLE 3.7 The accompanying data on gender and birth weight (in kilograms) of foals born to 15 thoroughbred mares appeared in the article "Suckling Behavior Does Not Measure Milk Intake in Horses" (*Animal Behaviour* (1999): 673–678). Figure 3.9 shows a MINITAB dotplot of the weight values.

Foal	1	2	3	4	5	6	7	8	9	10
Gender	F	M	M	F	F	M	F	F	M	F
Weight	129	119	132	123	112	113	95	104	104	93
Foal	11	12	13	14	15					
Gender	M	F	M	F	F					
Weight	108	95	117	128	127					

A typical birth weight is around 113. The 15 observations are quite spread out around this value. A gap separates the three smallest values from the rest of the data. Figure 3.10 shows the dotplots of male and female foal weights. Although the weights for male and female foals seem similar in the main part of the data, it is interesting to note that all three of the foals with low birth weight were female.

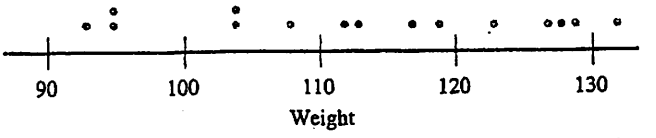


FIGURE 3.9 Dotplot of foal birth weights

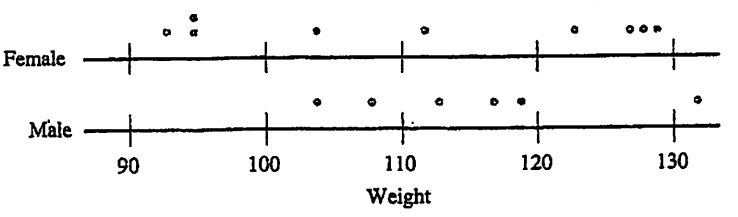


FIGURE 3.10 Dotplot of birth weight for male and female foals

Variables: categorical and quantitative

Some variables, like gender and job title, simply place individuals into categories. Others, like height and annual income, take numerical values for which we can do arithmetic. It makes sense to give an average income for a company's employees, but it does not make sense to give an "average" gender. We can, however, count the numbers of female and male employees and do arithmetic with these counts.

CATEGORICAL AND QUANTITATIVE VARIABLES

A **categorical variable** records which of several groups or categories an individual belongs to.

A **quantitative variable** takes numerical values for which it makes sense to do arithmetic operations like adding and averaging.

The **distribution** of a variable tells us what values the variable takes and how often it takes these values.

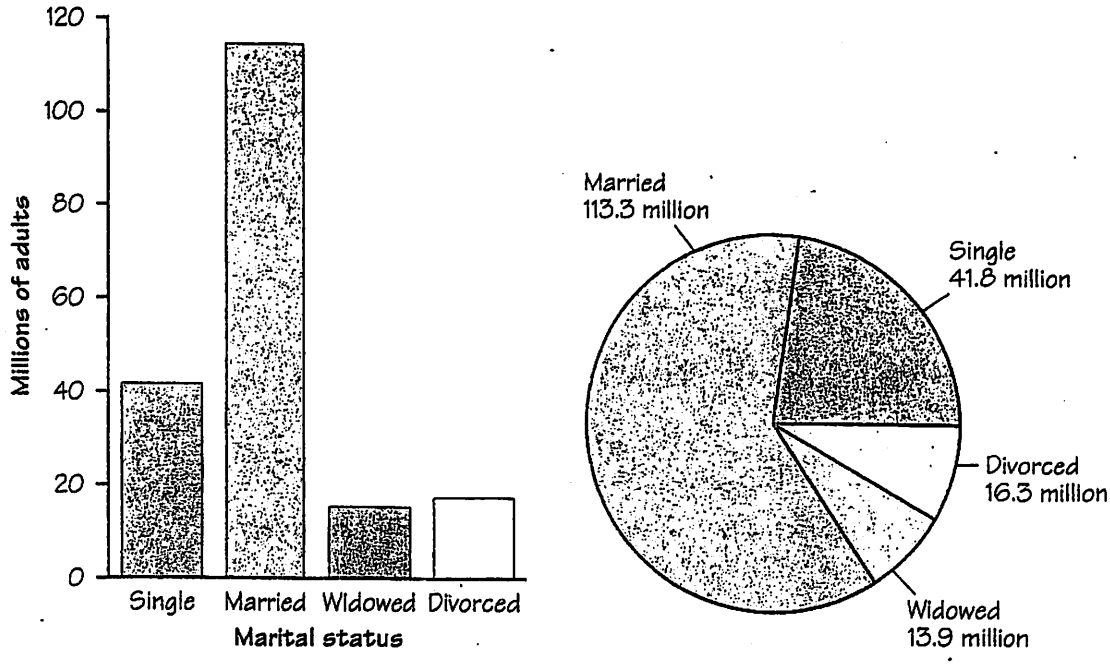
A variable generally takes values that vary. One variable may take values that are very close together while another variable takes values that are quite spread out. We say that the *pattern of variation* of a variable is its distribution.

The values of a categorical variable are just labels for the categories, like "male" and "female." The distribution of a categorical variable lists the categories and gives either the count or the percent of individuals who fall in each category. For example, here is the distribution of marital status for all Americans age 18 and over.

Marital status	Count (millions)	Percent
Single	41.8	22.6
Married	113.3	61.1
Widowed	13.9	7.5
Divorced	16.3	8.8

bar chart

To present such data to an audience, you may wish to use graphs like those in Figure 1.1. The ***bar chart*** in Figure 1.1(a) helps us compare the sizes of the four marital status groups. The heights of the four bars show the counts



(a) (b)

FIGURE 1.1 (*a*) Bar chart of the marital status of U.S. adults. (*b*) Pie chart of the same data.

pie chart

in the four categories. The ***pie chart*** in Figure 1.1(b) helps us see what part of the whole each group forms. For example, the "married" slice makes up 61% of the pie because 61% of adults are married. Bar charts and pie charts help an audience grasp the distribution quickly.

EXERCISES

1.1 In Activity 1, you collected some pulse rate data.

(a) What individuals do the data describe?
(b) How many variables are there? What are they?
(c) In what units is each variable recorded?
(d) Are the variables quantitative or categorical?

1.2 Example 1.1 presents data on the states. The first column identifies the states. Each of the remaining seven columns contains values of a variable. Which of these variables are categorical and which are quantitative?

cathy sun

9+

1-3, 5 or 6

1a) Longevity of cowboys

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

8|4 = 84

Analysis
- center: 15.5 - median
- shape: skewed left
- spread: 70-89, medium
- outliers: 47, 97
- were taken care of, got vitamin D, maybe contributed to long lives

b) yes, most lived long lives and to do that, they needed to be responsible, reserved, and somewhat wealthy.

2) % loss of wetlands

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

3|1 = 33

Analysis
- center: 48.5 - median
- shape: skewed right
- spread: 20-49, medium low
- outliers: 9

Deforestation or oil companies are equally spread, causing most to be concentrated in the middle

3) Average length of stay (days)

5	2 3 5 5 6 7
6	0 2 4 6 6 7 7 8 8 8 8 9 9
7	0 0 0 0 0 0 1 1 1 2 2 2 3 3 3 3 4 4 5 5 6 6 8
8	4 5 7
9	4 6 9
10	0 3
11	1

9|4 = 9.4

Analysis
- center: 71 - median
- shape: skewed right
- spread: 60-78, low
- outliers: 111 days

similar surgeries → similar stays, outliers could be like a chemo treatment

5a) Mins > 2 hours for winning marathon times (1961-1980)

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

0|9: 9 min > 2 hrs

Why not bk-bk?

Analysis
- center: 15.5 > 2hrs - median
- shape: roughly symmetric
- spread: 12-18, medium
- outliers: X

people can only run so fast and the human body has a limit so the times are close

b) Mins > 2 hrs for winning marathon times (1981-2000)

0·	7 7 7 8 8 8 8 9 9 9 9 9 9 9 9 9
1*	0 0 1 1 4

0|8 = 8 mins > 2 hrs

Analysis
- center: 9 min > 2hrs - median
- shape: skewed left?
- spread: 7-11, low-med
- outliers: 14

c) From 1961-1980, only 8 times < 2 hr 15 min but from 1981-2000, 20 times were < 2hr 15 min. This could be because of better clothing and shoe developments. Also, training w/ monitors from tech could help decrease times + maximize performance.

Ch2 Review 2-4,8

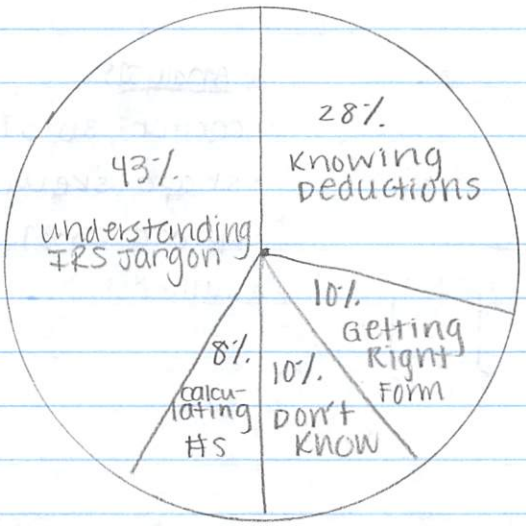
2a) 140,000 prisoners per 100,000 in 1980 and 440,000 per 100,000 in 1997

b) The time plot shows that every year, the number of prisoners per 100,000 people increases by roughly the same amount, resulting in a steady but gradual rise in prisoners.

c) $\frac{266,574,000}{100,000} \times 444 \approx 1183589$ prisoners in 1997

$\frac{323,724,000}{100,000} \times 444 \approx 1437335$ prisoners in 2020

3. Most Difficult Part of Filing a Tax Return



Analysis
- high - 43% - understanding jargon
- low - 8% calculating #s
- why - the most difficult parts also are the most time consuming, such as jargon and deductions require in depth research wheras calculating #s is simply easy busy work.

4. Ages of DUI Arrests

a)

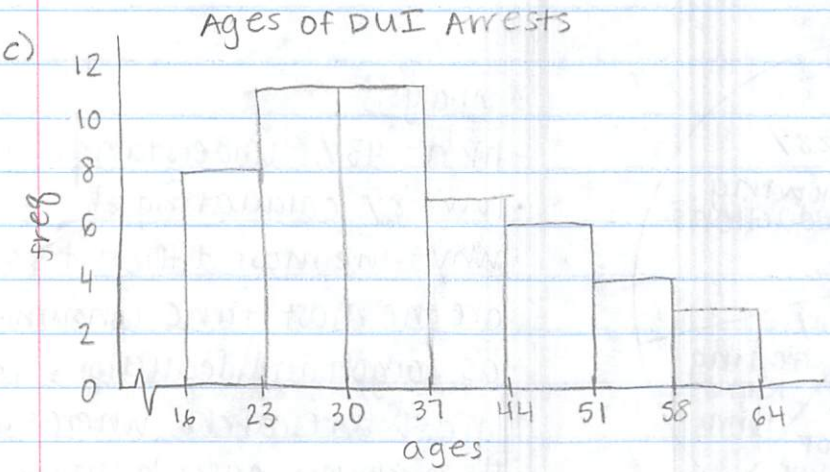
1	6	8																
2	0	1	1	2	2	2	3	4	4	5	6	6	6	7	7	7	9	
3	0	0	1	1	2	3	4	4	5	5	6	7	8	9				
4	0	0	1	3	5	6	7	7	9	9								
5	1	3	5	6	8													
6	3	4																

Analysis
- center: 33-34
  - ↳ 33.5 - median
- shape: skewed right
- spread: 20-49, mod
- outliers: 63, 64

b) $\frac{64-16}{7} = 7$

classes	Tallies	Freq	Rel Freq
16-22	𝍸 III	8	8/50 = 16%
23-29	𝍸 𝍸 I	11	11/50 = 22%
30-36	𝍸 𝍸 I	11	11/50 = 22%
37-43	𝍸 II	7	7/50 = 14%
44-50	𝍸 I	6	6/50 = 12%
51-57	IIII	4	4/50 = 8%
58-64	III	3	3/50 = 6%

c)



Analysis
- center: 30-37
- shape: skewed right
- spread: 23-37, low-mod
- outliers: X

8a) skewed left

b) 0.75-1.25; 1.25-1.75; 1.75-2.25; 2.25-2.75; 2.75-3.25; 3.25-3.75; 3.75-4.25

c) ↓ 3.25 = 1+1+2+8+17 = 29%

↓ 3.75 = 1+1+2+8+17+27 = 56%

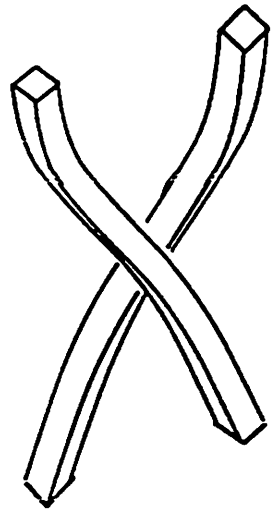
P75 #2

2. skewed left means low freqs are in the lower classes and a peak on the right, and skewed right means there is a peak on the left, and the rest of the #s are scattered in small amounts across the higher classes. Bimodal means there are multiple peaks with gaps between them. Yes, the weights of football players will be concentrated in the higher classes and the weights of cheerleaders will be clustered in the lower classes, resulting in 2 peaks are low freqs on the very edges and in the middle.

Name on zainab's paper :)

French Fry Statistics

Personal Data Sheet



Restaurant:_____

Size:_____

Cost:_____ Number of fries: _____

Length in Inches

Length

Mean:_____

10% Trimmed Mean:

Median:_____

Mode:_____

Range: _____

Standard Deviation: _____

Variance:

CV:

Total Number of Inches: _____

Cost per inches:_____

Cost per french fry: _____

Why is the cost per inch a more reliable comparison that the cost per french fry?

#3 lee Name cathy sun

Why do we divide by $n-1$ instead of n when calculating the sample standard deviation or variance?

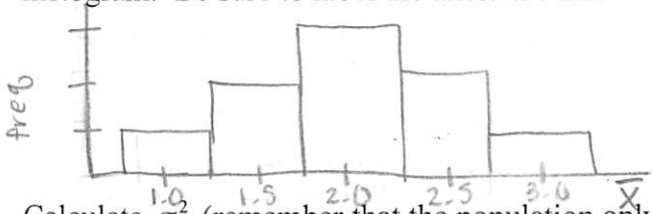
The goal of this activity is to guide you to discover the reason for the above and to practice the many things you have learned so far.

In order for a statistic, for example s^2, to be an unbiased estimator of a parameter, for this example σ^2, then the expected value or average of all s^2's should equal σ^2. Find out which formula for s^2 is the unbiased estimator by filling out the following table and answering the questions. Under the column "Samples of Size 2", list the values for each of the different samples that could come from a population of only three numbers: 1, 2, and 3. There are nine (9) different samples with replacement. Use these values to help fill in the rest of the table. The first two rows have been done for you. Be sure you understand where these numbers came from before proceeding.

	Samples of Size 2	$\bar{x}$	s^2 using $n-1$	s^2 using n
1	1, 1	1	0	0
2	1, 2	1.5	0.5	0.25
3	1, 3	2	2	1
4	2, 1	1.5	0.5	0.25
5	2, 2	2	0	0
6	2, 3	2.5	0.5	0.25
7	3, 1	2	2	1
8	3, 2	2.5	0.5	0.25
9	3, 3	3	0	0
Average	n/a	2	2/3	1/3

The tattoo parlor near campus got busy when the professor required hand calculation of the standard deviation.

1. Draw the distributions of the population and of the sample means. You can use a stemplot or histogram. Be sure to label the axes. Be informative!



Ana
- center = 2.0
- spread = 1-3
- shape = symmetrical
- outliers = X

2. Calculate σ^2 (remember that the population only has three values). Show your work.

$$\sigma^2 = \frac{\sum (x - \mu_x)^2}{n}$$

$$\sigma^2 = \frac{(3-2)^2 + (2-2)^2 + (1-2)^2}{3} = \frac{2}{3}$$

3. Which formula for s^2 is the unbiased estimator? Why?

s^2 using $n-1$, because it is equal to σ^2 of the three values and

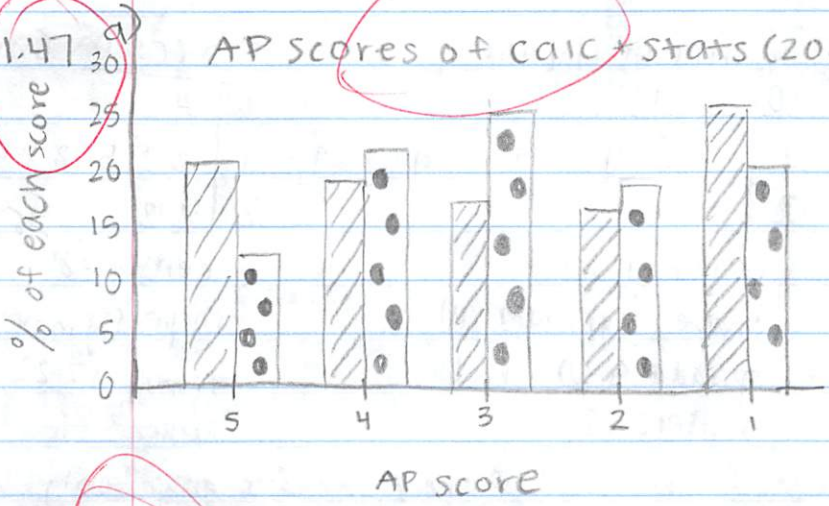
4. How do μ_x (population mean of x's) and $\mu_{\bar{x}}$ (population mean of $\bar{x}$'s, the sample means) compare?

μ_x and $\mu_{\bar{x}}$ are the same

Cathy Sun

10

1.47 a)



Ana
- high: calc 1's, lows: stats 5's
- Trends: look at B.

b) More people get 5s on calc AB than on stats, but the pass rate for stats is about 10% higher. Also, while the most frequent score for calc is 1, the most frequent score for stats is 3.

1.51 a) It is hard to remember the exact # of minutes so people tend to round. I think 0 minutes is suspicious because afterall, the students are in an AP class and they should most likely be studying. 300 and 360 minutes is also suspicious b/c that is 5-6 hours of only stats.

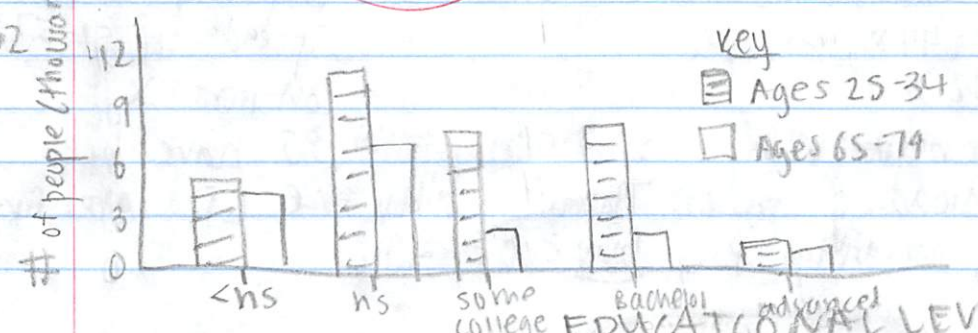
b)

Girls		Boys
90 60	0	0 30 30 30 30 45 60 60 60 75 90 90 90-95
80 80 70 50 50 50 50 20 20 20 20 20 20 20 15	1	20 20 20 20 20 20 20 50 50 80
40 40 40 00	2	00 00 30 40
60	3	00

2|30 = 230 min

Girls midpoint: 170-180
Boys midpoint: 120
Girls claim to study more than boys.

1.52



Ana
- high: 25-34 yrolds w/ hs graduate
- low: 65-74 w/ advanced degree
- Trend: People are pursuing more higher education, maybe college is more affordable

1.53 a) Trees in Group 1

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

center: 23
shape: symmetrical
spread: 20-24, med
outliers: 16, 33

Trees in Group 2

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

center: 14.5
shape: symmetrical
spread: 9-20, med
outliers: 2

Trees in Group 3

0	4
1	2 2 5 8 8 9
2	2 2

center: 8
shape: symmetrical
spread: 12-22, high
outliers: 4

b) Logging has greatly affected the count of trees, the 3rd group only has a median of 8 compared to 23 in the never logged group.

c) $\bar{x}$ to find mean and s to find spread and then you can get CV, which will allow you to compare spread of each group

1.54

Data A

3	10
4	74
5	
6	13
7	26
8	10 14 74 77
9	13 14 26

6|13 = 6.13

Ana
center: 8.14
shape: skewed left
spread: 6.13 - 9.267
outliers: 3.10, 4.74

Data B

5	25 56 76
6	58 89
7	04 71 91
8	47 84
9	
10	
11	
12	50

8|47 = 8.47

Ana
center: 7.04
shape: uniform
spread: 5.25 - 8.84
outliers: 12.50
↳ skews mean

$\bar{x}$ of A: 7.50 s of A: 2.03
$\bar{x}$ of B: 7.50 $\bar{x}$ of B: 2.03

1.55

Sales for 15-34

3	2.3 4.2 5.7 6.0 6.1 9.4
4	2.3 4.7 4.7 4.8

4|2.3 = 42.3

center: 37.75
shape: symmetrical
spread: 34.2 - 44.8, med
outliers: 32.3

Sales for over 35

4	6.5 6.6 7.3 7.9
5	0.4 3.8 4.5 4.5 6.0 7.9

4|6.5 = 46.5

center:
shape: symmetrical
spread: 46.5 - 57.9, high
outliers: X

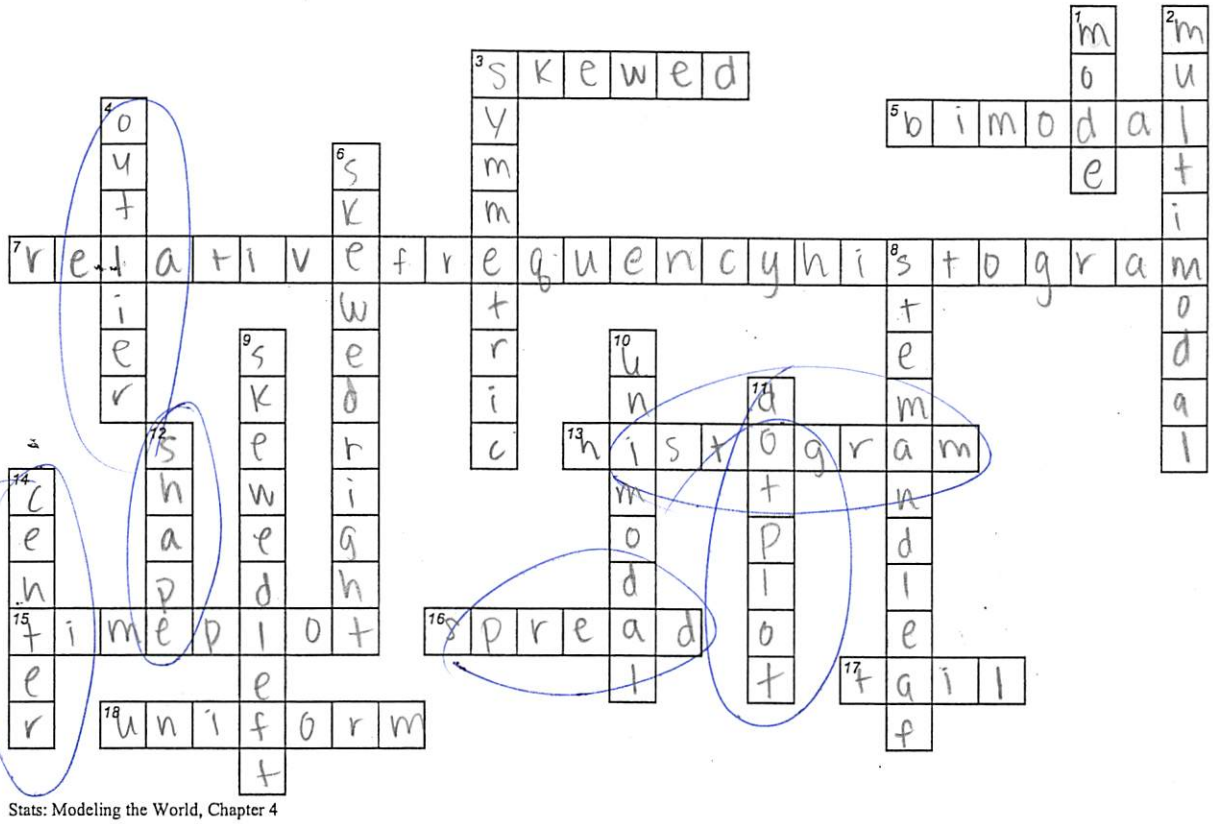
More people between ages 15-34 than over 35 have been illegally downloading music. That's why the CD sales for over 35 are increasing, they buy CDs still.

cathy sun

10

Displaying Quantitative Data

Advanced Placement Statistics



Stats: Modeling the World, Chapter 4

ACROSS

3 when a distribution is not symmetric and one tail stretches out farther than the other
5 distributions with two modes
7 uses adjacent bars to show the distribution of values in a quantitative variable, where each bar represents the proportion of values falling in an interval
13 uses adjacent bars to show the distribution of values in a quantitative variable, where each bar represents the number of values falling in an interval
15 used to display data that change over time
16 a numerical summary of how tightly the values are clustered around the "center"
17 the parts of a distribution that typically trail off on either side
18 a distribution roughly flat in shape

DOWN

1 a hump or high point in the shape of the distribution of a variable
2 distributions with more than two modes
3 shape where the two halves on either side of the center look approximately like mirror images of each other
4 an extreme value that doesn't appear to belong with the rest of the data
6 shape where the longer tail stretches to the right
8 type of display that shows quantitative data values in a way that shows the shape of the distribution in addition to individual data values
9 shape where the longer tail stretches to the left
10 having one mode
11 graphs a dot for each case against a single axis
12 reveals single vs. multiple modes and symmetry vs. skewness
14 a value that summarizes the entire distribution with a single number, a "typical" value

Cathy Sun
10/19/13
P4-Stats

9+

3.3 # 1, 3, 6, 7, 9-11

1. 82% are at or below, 18% are above

3. NO, there may have been many scores in the 90s and the 70th percentile could be a score of 85.

6a) median = 23

Q1 = 16 IQR = 30 - 16 = 14

Q3 = 30

Months of being in ~~staff~~ positions

label?

Ana
- center: 23 (median)
- shape: skewed right
- spread: 16-30, mod
- outliers: 72 ?

b) medians are same, middle half is more spread out in #5; distance from Q1 and Q3 to extreme values is less in #5

7a) IQR = 27 - 22 = 5

% with bachelors degrees

Ana
- center: 24 (median)
- shape: skewed right
- spread: 22-25, low
- outliers: X

b) 26% falls into the 3rd quartile

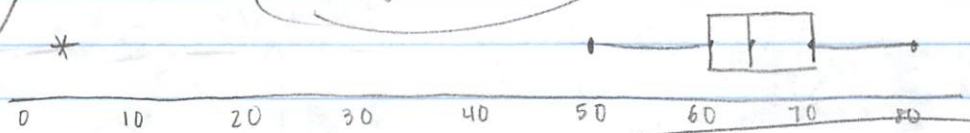
9a) California has lowest, Pennsylvania has highest

~~b)~~ Pennsylvania has highest median premium

c) California has smallest range, Texas has smallest IQR

d) A - Texas, B - Pennsylvania, C - California

10a) Heights (in)



Ana
- center: 65.5 - median
- shape: symmetrical
- spread: 10, low
- outliers: 4

where?

b) ~~IQR~~: 71.5 - 61.5 = 10

c) IQR × 1.5 = 15

lower limit: 46.5
upper limit: 86.5

d) 4 is an outlier, maybe the person recorded his/her height in feet

11a) Assistant had smallest median % salary increase while associate had the highest increase

b) instructor

c) Assistant

d) professor, yes

e) ~~Professor: $1.5 \times (4,975 - 2,100) + 4,975$: 9.29~~

Associate: $1.5 \times (5,075 - 2,350) + 5,075$: 9.16

~~Assistant: $1.5 \times (5,500 - 2,900) + 5,500$: 9.4~~

Instructor: $1.5 \times (5,800 - 2,850) + 5,800$: 10.23

Yes, marked by artesisks

cathy sun
10/16/13

10
Good

3.2 # 4, 6-8

4 a) · Range = 14.1 − 6.8 = 7.3

· Mean = $\frac{14.1+12.4+7.7+6.9+9.0+6.8+6.8}{7}$ = 9.1

· sample variance = $\frac{\Sigma(x-\bar{x})^2}{n-1}$

= $\frac{(14.1-9.1)^2+(12.4-9.1)^2+(7.7-9.1)^2+(6.9-9.1)^2+(9.0-9.1)^2+(6.8-9.1)^2+(6.8-9.1)^2}{6}$

= $\frac{53.28}{6}$

= 8.88

· $s = \sqrt{8.88}$

= 2.98

· CV = $\frac{2.98}{9.1} \times 100$ = 32.75%

b) · Range = 31.0 − 19.1 = 11.9

· Mean = 26.13

· sample variance = $\frac{(29.0-26.13)^2+(24.5-26.13)^2+(31.0-26.13)^2+(29.8-26.13)^2+(21.8-26.13)^2+(27.7-26.13)^2+(19.1-26.13)^2}{6}$

= 19.79

· $s = \sqrt{19.79}$

= 4.45

· CV = $\frac{4.45}{26.13} \times 100$ = 17.03%

c) Since a small CV means low variability, that means that many of the values do not differ too much from the mean. With a large mean and small CV, that means many of the values are clustered around the same high mean, showing that many employees get high profits.

6a) on calculator

b) $1-\frac{1}{2^2} = \frac{3}{4} = 75\%$ (730 ± 2σ)

we conclude that at least 75% of years will have 386 - 1074 tornados

c) $1-\frac{1}{3^2} = \frac{8}{9} = 88.9\%$ (730 ± 3σ)

we conclude that at least 88.9% of years will have 214 - 1246 tornados.

7a) Range = 956 − 219 = 737

mean = 566.86

b) $s^2 = \frac{(851-566.86)^2+(596-566.86)^2+(444-566.86)^2+(956-566.86)^2+(576-566.86)^2+(219-566.86)^2+(326-566.86)^2}{6}$

$= 71202.14$

$s = \sqrt{71202.14} = 266.84$

c) $CV = \frac{266.84}{566.86} \times 100 = 47.07\%$

There is a large spread and the artifact counts in the excavation sites are very different. Some places might have been the location of fights, while other sites were not as interesting and had way fewer artifacts.

d) $1 - \frac{1}{2^2} = \frac{3}{4} = 75\%$ (−2s … 566.86 … 2s)

we conclude that at least 75% of artifact counts fall from 33.18 − 1100.54.

8. $CV = \frac{s}{\bar{x}} \times 100$

$1.5 = \frac{s}{2.2} \times 100$

$\frac{1.5 \times 2.2}{100} = \boxed{0.033}$

cathy sun

ch3 Review #1, 5-7

1a) $\bar{x} = \frac{73+140+78+142+80+140+90+133}{8}$

$= 109.5$

$s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}} = 31.72$

$CV = \frac{31.72}{109.5} \times 100 = 28.97\%$

range $= 142 - 73 = 69$

b) $\bar{x} = \frac{100+120+108+114+105+117+103+114}{8}$

$= 110.13$

$s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}} = 7.16$

$CV = \frac{7.16}{110.13} \times 100 = 6.50\%$

range $= 120 - 100 = 20$

c) means are about equal. The distribution w/ the power surge protector is more compact and tighter. The standard deviation and CV show that the values in b are closely packed together. The range is also much smaller in b.

5a) $\bar{x} = \frac{10.1+6.2+9.8+5.3+9.9+5.7}{6}$

$= 7.83$

$s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}} = 2.32$

$CV = \frac{2.32}{7.83} \times 100 = 29.63\%$

range $= 10.1 - 5.3 = 4.8$

b) $\bar{x} = \frac{10.2+9.7+9.8+10.3+9.6+10.1}{6}$

$= 9.95$

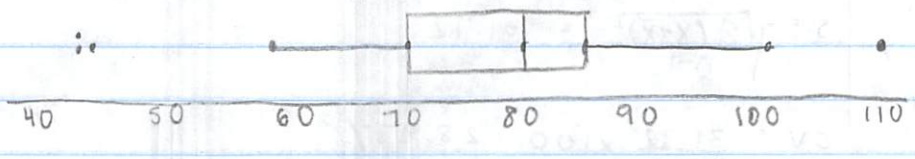
$s = 0.29$

range $= 10.3 - 9.6 = 0.7$

$CV = \frac{0.29}{9.95} \times 100 = 2.91\%$

c) Line B was more consistent, and the standard deviation is very small, proving the values are close together. The CV in b < in a, and the data is very packed, which is proven by the small range as well.

6a) min = 45
Q1 = 71
Median = 80
Q3 = 84
Max = 109
IQR = 13



7. $\frac{2500}{16} = 156.25$ is mean weight of people in the elevator

Ana
- center: 80 - median
- shape: roughly uniform
- spread: 13 (IQR) low
- outliers: 45, 45, 46, 109

Cathy Sun

10 Good

Trade Winds p123

a) Calculator

b) $\bar{x}$ = 33.17

med = 21

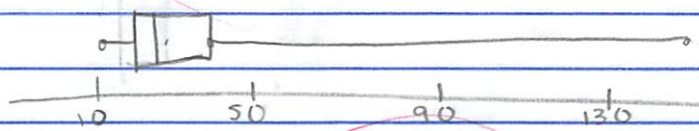
mode = 18

c) range = 127

variance = 845.06

s = 29.07

d) Total Air movement



miles/day

Ana

center - 21, median

shape - skewed high

spread = 18 (IQR), mod

outliers = 72, 100, 105, 113, 138

e) $\bar{x}$ = 25.48

med = 20

mode = 18

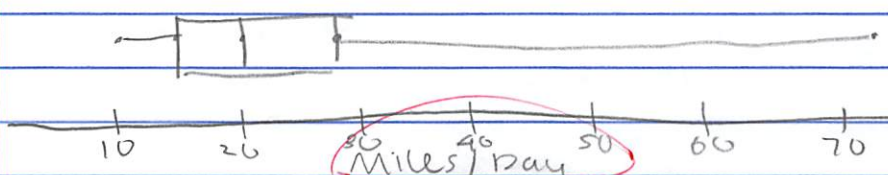
range = 61

variance = 208.22

s = 14.43

Mean is most affected b/c it relies on values. The standard deviation plummets because not outliers make the spread high.

Total Air Movement



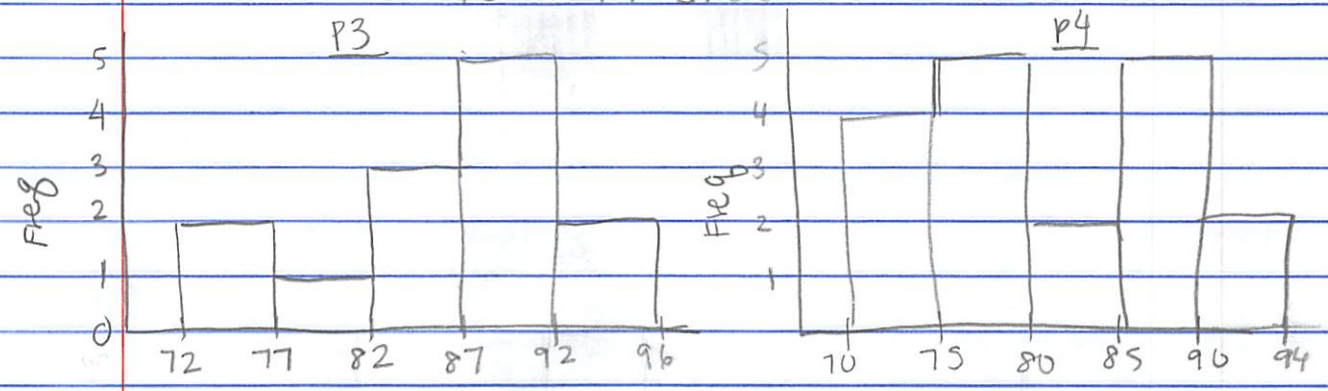
Ana

- center: 20 - med
- shape - skewed high
- spread = 12 (IQR), low
- outliers = 56, 50, 52, 57, 57, 72

- Far less spread, still skewed high

f) From May-Sept in Hawaii, median wind mph is 12. In winter months, the median is 10 or 11 mph.

P3 v P4 Grades



Ana

center: 82-87
shape: skewed low
spread: 77-92, mod
outliers: X

Ana

center: 80-85
shape: bimodal
spread: 75-90, mod
outliers: X

cathy sun

±3σ

P3 vs P4 Grades

P3	P4
87	79
83	88
77	70
72	77
94	89
91	91
73	79
89	84
88	74
88	73
85	89
83	88
93	87
	84
	77
	77
	74
	90

	P3	P4
mode	88	77
mean	84.85	81.67
Min	72	70
Q_1	80	77
Med	87	81.5
Q_3	90	88
MAX	94	91
standard dev	6.81	6.63
variance	46.38	43.96
range	22	21
trimmed mean	85.67	81.86
CV	8.03%	8.12%
IQR	18	18

chebyshev ± 2s

Period 3: we conclude that at least 75% of values will be between 71.23 and 98.47

Period 4: we conclude that at least 75% of values will be between 68.41 and 94.93

Period 3 vs. Period 4

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

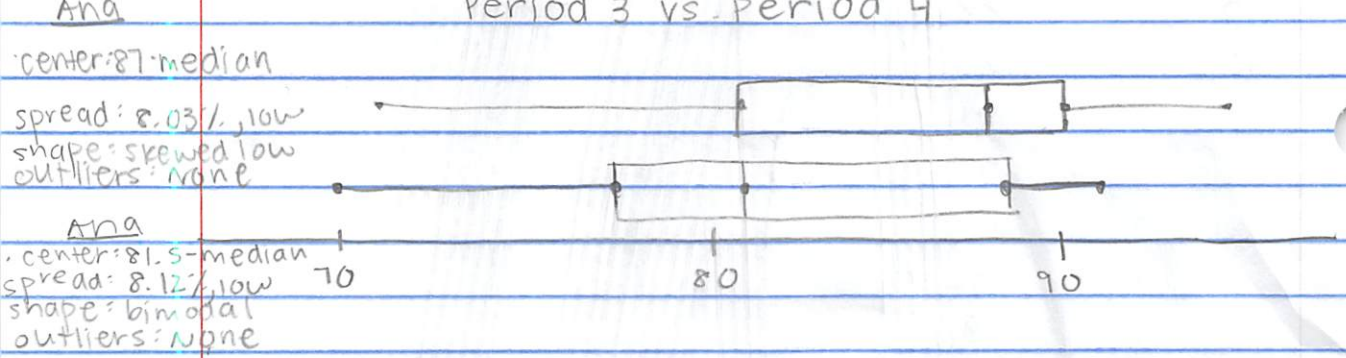
Ana

center: 87-median
spread: 8.03%, low
shape: skewed low
outliers: none

Ana

center: 81.5 median
spread: 8.12%, low
shape: bimodal
outliers: none

Period 3 vs. Period 4



Frequency Table

	Class	Tallies	Freq	Rel Freq
Period 3	72-76	II	2	15.38%
	77-81	I	1	7.69%
	82-86	III	3	23.08%
	87-91	IIII	5	38.46%
	92-96	II	2	15.38%
Period 4	70-74	IIII	4	22.22%
	75-79	IIII	5	27.78%
	80-84	II	2	11.11%
	85-89	IIII	5	27.78%
	90-94	II	2	11.11%