5.1 & 5.2 Frequency Tables, Histograms & Freq Polygons

Frequency Tables, Histograms, & Frequency Polygons [5.1 & 5.2]

Are you ready for some STATS???? Let's start by exploring the similarities & differences between 2 sets of data.

Measured Lifespans of 30 Car Batteries (years)									
Brand X					Brand Y				
5.1	7.3	6.9	4.7	5.0	5.4	6.3	4.8	5.9	5.5
6.2	6.4	5.5	5.7	6.8	4.7	6.0	4.5	6.6	6.0
6.0	4.8	4.1	5.2	8.1	5.0	6.5	5.8	5.4	5.1
6.3	7.5	5.0	5.7	8.2	5.7	6.8	5.6	4.9	6.1
3.3	3.1	4.3	5.9	6.6	4.9	5.7	6.2	7.0	5.8
5.8	6.4	6.1	4.6	5.7	6.8	5.9	5.3	5.6	5.9

Vocab:

Mean	Median	Mode
"the average" add all values and divide by how many values there are.	"the middle" line values up in order and cross off from both sides.	"most common" occurs most often in the data
Brand X (5.1+7.3+6.9+4.7+5.0+6.2+6.4+5.5+5.7+6.8) = 59.6 $\frac{59.6}{10} = 6.0$ X: 6.0 Brand Y: $\frac{55.7}{10} = 5.6$ y: 5.6	X: ~~4.7~~, ~~5.0~~, ~~5.1~~, ~~5.5~~, 5.7, 6.2, ~~6.4~~, ~~6.8~~, ~~6.9~~, ~~7.3~~ $\frac{5.7+6.2}{2} = 5.95$ $= 6.0$ Y: 5.7	X: look at chart no values occur more than once X: none y: 6.0 occurs 2 times y: 6.0

Range: Highest # "minus" Lowest #

X: 7.3 − 4.7 = 2.6 (range of x)

y: 6.6 − 4.5 = 2.1 (range of y)

Example 1

Maximum Water Flow Rates for the Red River, from 1950 to 1999, Measured at Redwood Bridge*									
Year	Flow Rate (m^3/s)	Year	Flow Rate (m^3/s)	Year	Flow Rate (m^3/s)	Year	Flow Rate (m^3/s)	Year	Flow Rate (m^3/s)
1950	3058	1960	1965	1970	2280	1980	881	1990	396
1951	1065	1961	481	1971	1526	1981	159	1991	280
1952	1008	1962	1688	1972	1589	1982	1458	1992	1399
1953	357	1963	660	1973	530	1983	1393	1993	946
1954	524	1964	1002	1974	2718	1984	1048	1994	1121
1955	1521	1965	1809	1975	1671	1985	991	1995	1877
1956	1974	1966	2498	1976	1807	1986	1812	1996	3058
1957	654	1967	1727	1977	187	1987	2339	1997	4587
1958	524	1968	510	1978	1750	1988	564	1998	1557
1959	991	1969	2209	1979	3030	1989	1390	1999	2180

Example 1: Determine the water flow rate that is associated with serious flooding by creating a frequency distribution. (table or graph)

* a frequency distribution should have between 5 and 12 intervals

to determine intervals to use, we look at the range

Highest = 4587
Lowest = 159

Range = 4587 − 159
= 4428 m^3/s

10 intervals

$\frac{4428}{10} = 442.8$

→ 500 is a nicer interval

Frequency Distribution Table

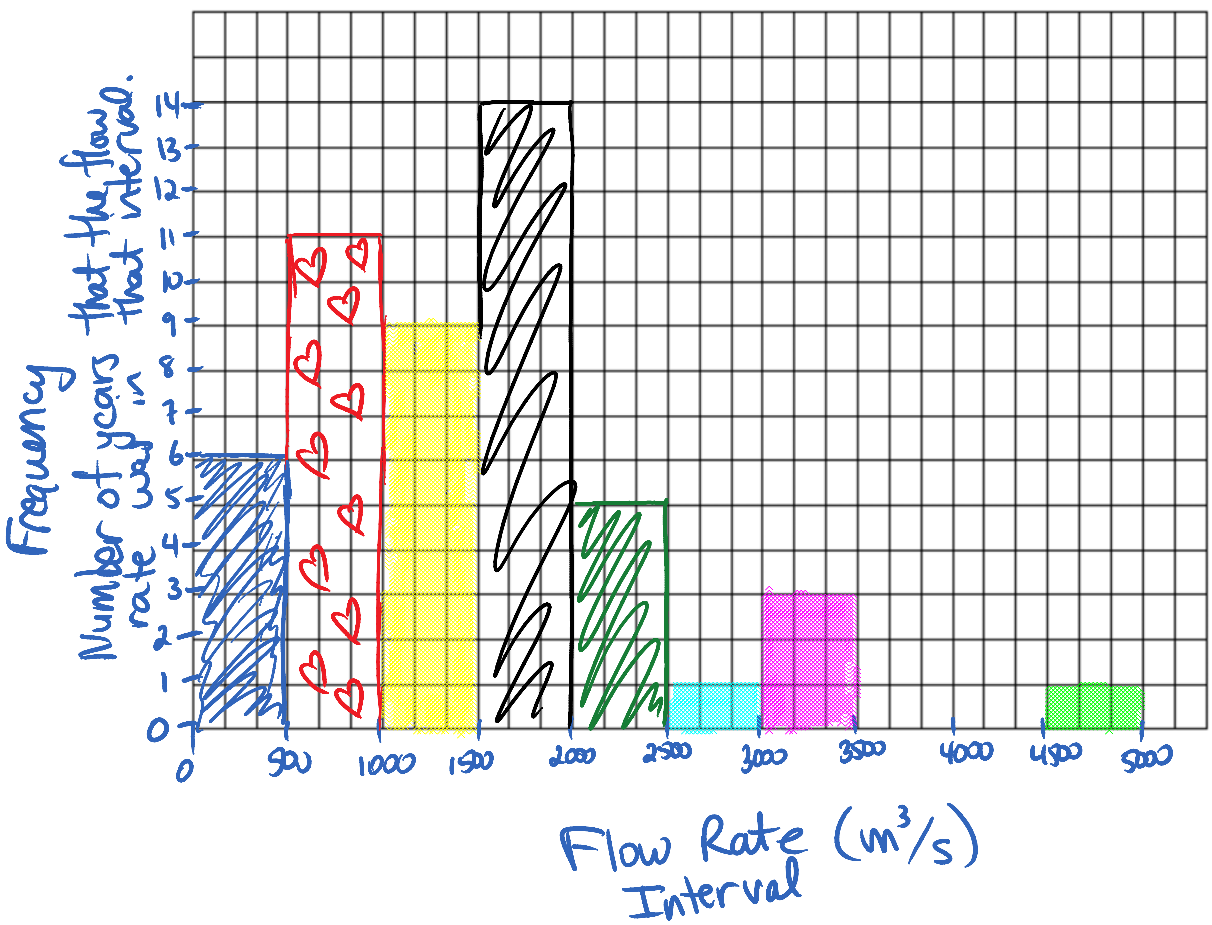
Flow Rate (m^3/s) intervals	Tally	Frequency (# of years)	Midpoint of the interval
0-500	~~||||~~ |	6	250
500-1000	~~||||~~ ~~||||~~ |	11	750
1000-1500		9	1250
1500-2000		14	1750
2000-2500		5	2250
2500-3000		1	2750 $= \frac{2501+3000}{2}$
3000-3500		3	3250
3500-4000		0	3750
4000-4500		0	4250
4500-5000		1	4750

a count of the # of times the max. water flow rates happened

Using the same information from above, create a Histogram. = bar graph

Things to remember when creating Histograms
- graph of a frequency table
- equal intervals on the horizontal (x) axis
- frequencies on the vertical (y) axis
- no space between bars
- Must have title and axes labelled.

Red River Flow Rates
(1950–1999)

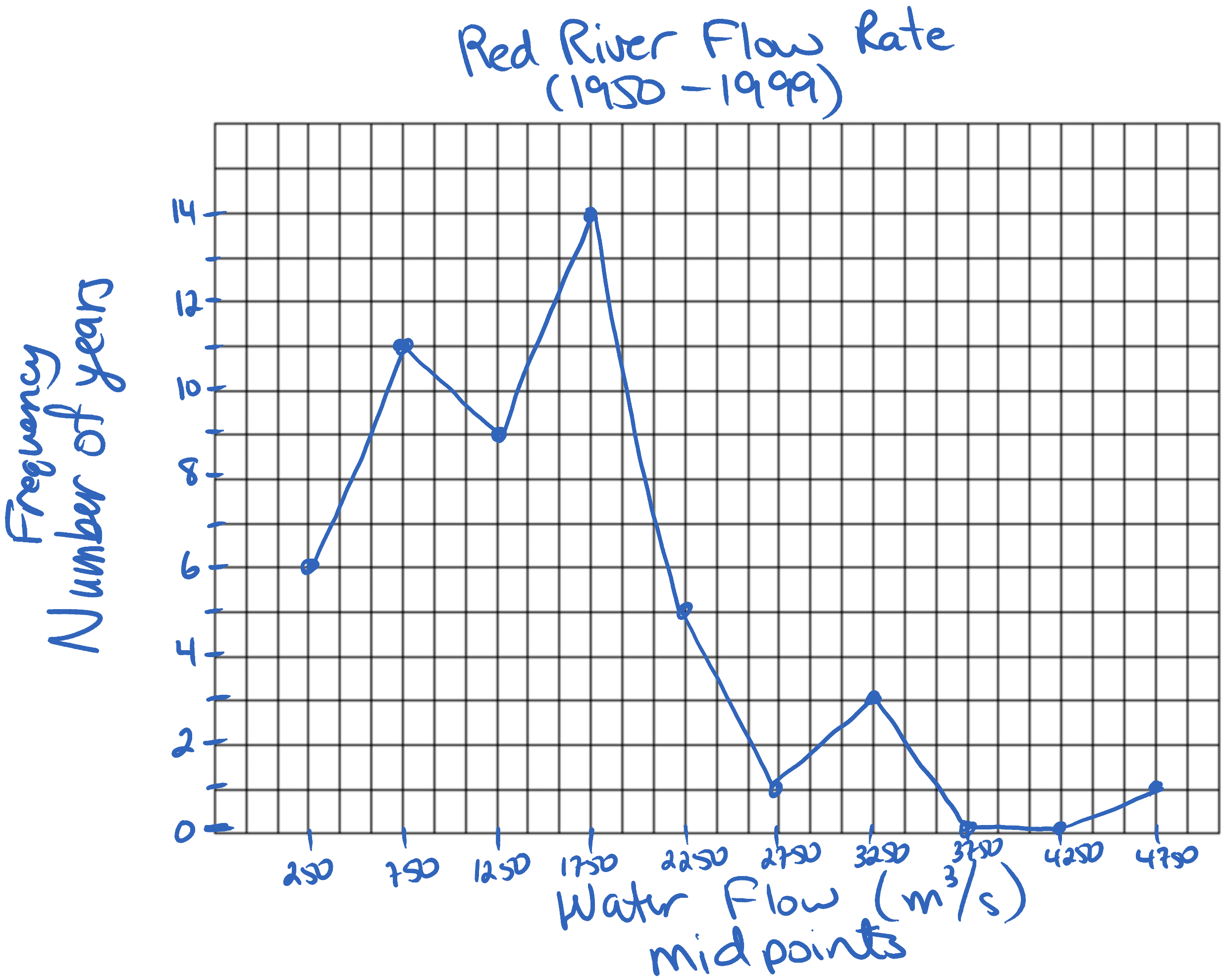


Look what else we can do with the same information… a <u>Frequency Polygon</u>!

water flow midpoints vs # of years └line graph

A Quick How To: the Frequency Polygon

- use the midpoints of the intervals
- connect points using straight lines
- title and labelled axes!!



Practice
pg 211 #1,2
pg 222 #4

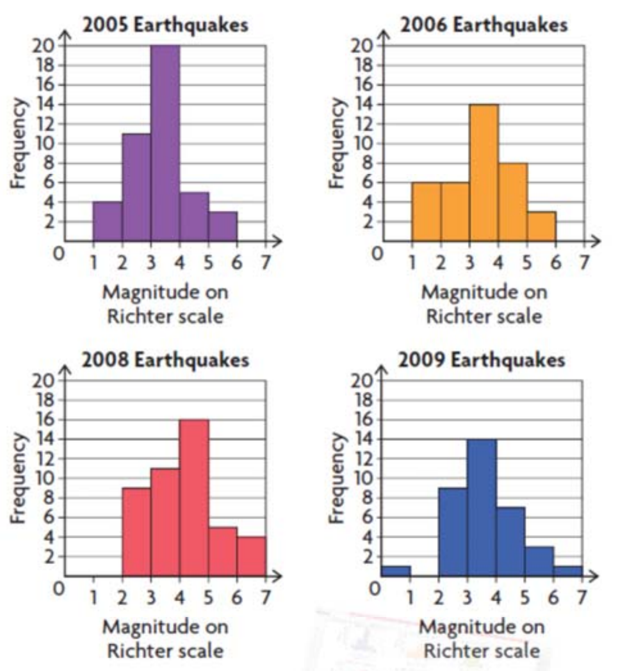
With Freq. Polygon can graph more than one line so can compare data more easily. :)

Example 2:

Which of these years could have had the most damage from earthquakes?

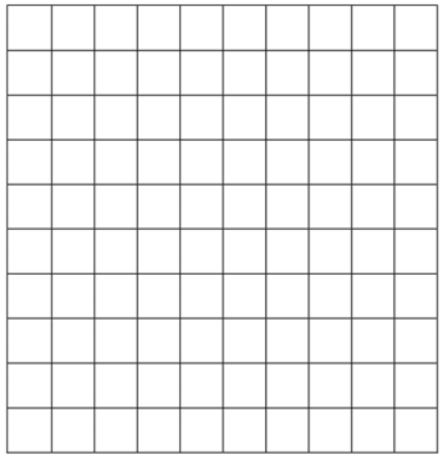
Understanding the Richter Scale*	
Magnitude	**Effects**
less than 3.0	recorded by seismographs; not felt
3.0–3.9	feels like a passing truck; no damage
4.0–4.9	felt by nearly everyone; movement of unstable objects
5.0–5.9	felt by all; considerable damage to weak buildings
6.0–6.9	difficult to stand; partial collapse of ordinary buildings
7.0–7.9	loss of life; destruction of ordinary buildings
more than 7.9	widespread loss of life and destruction

*Every unit increase on the Richter scale represents an earthquake 10 times more powerful. For example, an earthquake measuring 5.6 is 10 times more powerful than an earthquake measuring 4.6.



National Research Council Canada

Frequency Polygon:



pg 222 #8

Holly 22max − 0min = 22km

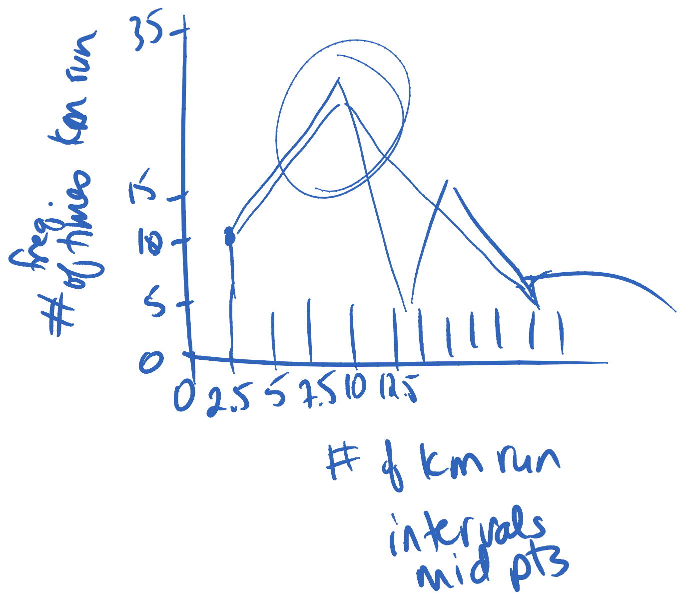
5 intervals $\frac{22}{5} = 4.5 \rightarrow 5$

Jason $\frac{35\text{km}}{7 \text{ intervals}} = 5$

mdpt	(interval) # of km run	# of times this many km's were run
2.5	0−5	11
7.5	6−10	30
12.5	11−15	19
17.5	16−20	8
22.5	21−25	1

Jason

0−5	
6−10	
⋮	
31−35	


35
15
10
5
0
freq.
# of times km run
0
2.5
5
7.5
10
12.5
# of km run
intervals
mid pts

5.3 Standard Deviation Pg 7

Standard Deviation [5.3]

σ = Standard deviation: a measure of the dispersion or scatter of data values in relation to the mean.

$$\boxed{\sigma = \sqrt{\frac{\sum(x-\bar{x})^2}{n}}}$$

σ (read as sigma – lower case): represents the standard deviation of the data

$\sum$ (read as sigma – upper case): summation operator

x: each data value

$\bar{x}$ (read as x bar): represents the mean of the data

average = $\frac{\text{add all \#'s together}}{\text{divide by \# of data values}} = \bar{x}$

n: the number of data values

* the smaller the standard deviation (σ), the more consistent the data.
→ more of the data #'s are close to the average (mean)

Example 1:

Brendan works part-time in the canteen at his local community centre. One of his tasks is to unload delivery trucks. He wondered about the accuracy of the mass measurements given on two cartons that contained sunflower seeds. He decided to measure the masses of the 20 bags in the two cartons. One carton contained 227 g bags, and the other carton contained 454 g bags.

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

Masses of 227 g Bags (g)			
228	220	233	227
230	227	221	229
224	235	224	231
226	232	218	218
229	232	236	223

x

add = 4543

Masses of 454 g Bags (g)			
458	445	457	458
452	457	445	452
463	455	451	460
455	453	456	459
451	455	456	450

How can measures of dispersion be used to determine if the accuracy of measurement is the same for both bag sizes?

1 Start by finding the mean (avg) of the population (add up, ÷ #)

$$\bar{x} = \frac{4543}{20 \text{ bags}} = 227.15\text{g} \qquad n = 20$$

2

x	$(x-\bar{x})^2$
228	0.7225
230	8.1225
224	9.9225
226	1.3225
229	3.4225
220	51.1225
227	0.0225
235	61.6225
232	23.5225
232	23.5225
233	34.2225
221	37.8225
224	9.9225
218	83.7225

$$(228 - 227.15)^2 = (0.85)^2 = 0.7225$$

x	$(x-\bar{x})^2$
236	78.3225
227	0.0225
229	3.4225
231	14.8225
218	83.7225
223	17.2225

3 add all $(x-\bar{x})^2$ together

$$\Sigma(x-\bar{x})^2 = 546.55$$

4

$$\sigma = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}} = \sqrt{\frac{546.55}{20}}$$

$$\sigma_{227g} = 5.22757$$

$$\sigma_{454g \text{ bags}} = 4.4988$$

∴ the 454g bags are more consistently close to 454g. since σ is smaller pg 9

Example 2:

Angèle conducted a survey to determine the number of hours per week that Grade 11 males in her school play video games. She determined that the mean was 12.84 h, with a standard deviation of 2.16 h.

Janessa conducted a similar survey of Grade 11 females in her school. She organized her results in this frequency table. Compare the results of the two surveys.

Gaming Hours per Week for Grade 11 Females		
Hours (Interval)	Frequency	Midpoint
3–5	7 girls	4 hrs
5–7	11 girls	6 hrs
7–9	16 "	8 "
9–11	19	10
11–13	12	12
13–15	5	14

$$\bar{x} = \frac{(4\text{hrs} \cdot 7\text{girls}) + (6\text{hrs} \cdot 11\text{girls}) + (8 \cdot 16) + (10 \cdot 19) + (12 \cdot 12) + (14 \cdot 5)}{70}$$

total # of girls
7+11+16+19+12+5

$\boxed{\bar{x} = 9}$ $\boxed{n = 70}$

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

x	$(x-\bar{x})^2$	$\times f =$
4	25	×7 = 175
6	9	×11 = 99
8	1	×16 = 16
10	1	×19 = 19
12	9	×12 = 108
14	25	×5 = 125

$\Sigma f(x-\bar{x})^2 \rightarrow 542$

$$\sigma = \sqrt{\frac{542}{70}}$$

$\boxed{\sigma = 2.78}$

Comparison

– larger standard deviation (girls) means the

data is more scattered. (more high hours and/or low hours)
- boys play closer to all the same amount - closer to the average/mean

Practice pg 235 # 11, 13

(11)

# of employees that handle the range of daily calls indicated.

Daily Calls	Freq	"X" mid point	freq × mid pt = Total Daily calls	$f \cdot (x-\bar{x})^2$
26-30	2	28	56	$2 \cdot (28-45)^2 = 578$ ✓
31-35	13	33	429	$13 \cdot (33-45)^2 = 1872$ ✓
36-40	42	38	1596	→ 2058 ✓
41-45	53	43	2279	212 ✓
46-50	42	48	2016	378 ✓
51-55	36	53	1908	2304 ✓
56-60	8	58	464	1352 ✓
61-65	4	63	252	1296 ✓
	200 people		Total calls for all employees: 9000 calls	$\Sigma f(x-\bar{x})^2 = 10050$

$$\sigma = \sqrt{\frac{\Sigma f(x-\bar{x})^2}{n}} = \sqrt{\frac{10050}{200}} = 7.1$$

200 employees handled 9000 calls in a day.

$$\text{Average} = \text{Mean} = \frac{9000 \text{ calls}}{200 \text{ people}} = 45 \text{ calls/day/person} = \bar{x}$$

Average = 45 but standard deviation = 7.1

ok

too high

∴ must hire more workers.

13 The mean (average) could be the same:

for ex

Jordana : 7, 7, 7, 7, 7 → avg = 7 = $\bar{x}$

Jane : 5, 6, 7, 8, 9 → avg = 7

Jordana is more consistent, her standard deviation would be zero $\sigma = \sqrt{\frac{\sum(x-\bar{x})^2}{n}} = 0$

Jane's scores have more variation

$$\sigma = \sqrt{\frac{\sum(x-\bar{x})^2}{n}} = \sqrt{\frac{10}{5}} = \sqrt{2}$$

x	$(x-\bar{x})^2$
5	4
6	1
7	0
8	1
9	4
	10

5.3 Homework   pg 233 # 5-7

5) Group 1

mean: $\bar{G}_1 = \dfrac{63+78+79+75+73+72+62+75+63+77+77+65+70+69+80}{15}$

$\bar{G}_1 = \dfrac{1078}{15} = 71.9$ bpm $n = 15$

$(63-71.9)^2$

G_1	$(G_1-\bar{G}_1)^2$
63	79.21
78	37.21
79	50.41
75	9.61
73	1.21
72	0.01
62	98.01
75	9.61
63	79.21
77	26.01
77	26.01
65	47.61
70	3.61
69	8.41
80	65.61

$\Sigma(G_1-\bar{G}_1)^2 = 79.21 + 37.21 + \cdots + 65.61$
$= 541.75$

$\sigma = \sqrt{\dfrac{\Sigma(G_1-\bar{G}_1)^2}{n}} = \sqrt{\dfrac{541.75}{15}} = 6.0$ bpm

standard deviation for group 1

G_1: $\bar{G}_1 = 71.9$ bpm $\sigma_1 = 6.0$ bpm

G_2: $\bar{G}_2 = 71.0$ bpm $\sigma_2 = 4.0$ bpm

G_3: $\bar{G}_3 =$ ______ $\sigma_3 =$ ______

G_4: $\bar{G}_4 = 76.9$ bpm $\sigma_4 = 1.9$ bpm

G3 1 mean → add all the pulses for Group 3 and divide by 15 and divide by 15

average or mean $\rightarrow \bar{G}_3 = 70.4$

② Standard Deviation formula $\sigma = \sqrt{\frac{\Sigma(G_3-\bar{G}_3)^2}{n}}$

$= \sqrt{\frac{\Sigma(G_3-\bar{G}_3)^2}{15}} =$

③ $\Sigma(G_3-\bar{G}_3)^2$

G_3	$(G_3-\bar{G}_3)^2$
68	5.76
75	
78	
⋮	
79	

$\leftarrow (68-70.4)^2 = (-2.4)^2 = 5.76$

$\Sigma \rightarrow$ add all the $(G_3-\bar{G}_3)^2$ #'s

④ Plug that total value into the σ

$\sigma = \sqrt{\frac{\text{total value}}{15}}$ ← n

$= 5.7$

7a) mean $= \frac{\text{add all TD}}{14} = \frac{147}{14} = 10.5 = \bar{T}$ ← average # of touchdowns per season.

$\sigma = \sqrt{\frac{\Sigma(T-10.5)^2}{14}}$ ← $\bar{T}$

T	$(T-\bar{T})^2$
4	42.25

$\leftarrow (4-10.5)^2$

$$\sigma = \sqrt{\frac{\sum (T - 10.5)^2}{14}}$$

n

T	$(T-\bar{T})^2$
4	42.25
6	20.25
14	12.25
7	12.25
7	12.25
15	20.25
14	12.25
23	156.25
15	20.25
7	12.25
17	42.25
7	12.25
8	6.25
3	52.56

$\leftarrow (4-10.5)^2$

$4(12.25)$

$\sum (T-\bar{T})^2$ (add all)

$= 433.81$

$$\sigma = \sqrt{\frac{433.81}{14}}$$

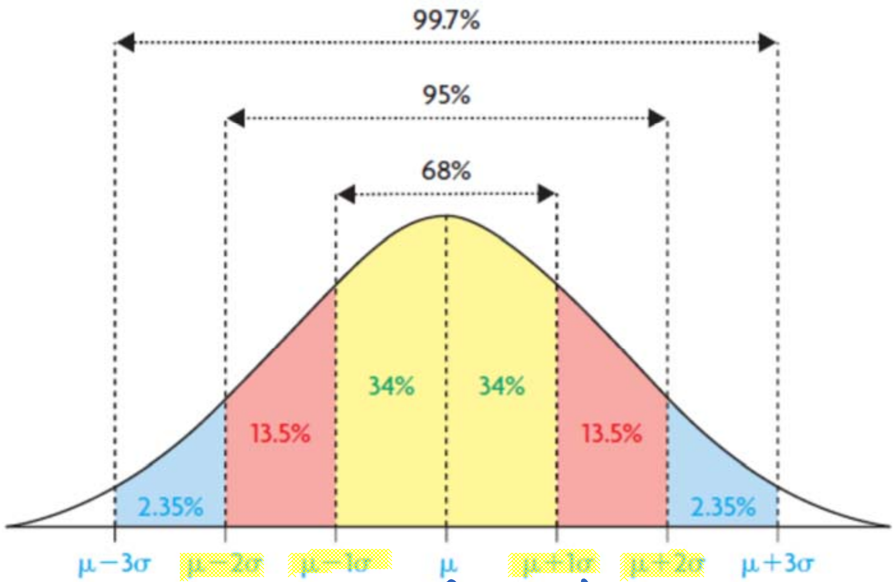
$= 5.6$

5.4 The Normal Distribution

The Normal Distribution [5.4]

Need to Know

- The properties of a normal distribution can be summarized as follows:
  - The graph is symmetrical. The mean, median, and mode are equal (or close) and fall at the line of symmetry.
  - The normal curve is shaped like a bell, peaking in the middle, sloping down toward the sides, and approaching zero at the extremes.
  - About 68% of the data is within one standard deviation of the mean.
  - About 95% of the data is within two standard deviations of the mean.
  - About 99.7% of the data is within three standard deviations of the mean.
  - The area under the curve can be considered as 1 unit, since it represents 100% of the data.



- Generally, measurements of living things (such as mass, height, and length) have a normal distribution.

mean "$\bar{X}$"

σ = standard deviation

Example 1:

Heidi is opening a new snowboard shop near a local ski resort. She knows that the recommended length of a snowboard is related to a person's height. Her research shows that most of the snowboarders who visit this resort are males, 20 to 39 years old. To ensure that she stocks the most popular snowboard lengths, she collects height data for 1000 Canadian men, 20 to 39 years old. How can she use the data to help her stock her store with boards that are the appropriate lengths?

Mid Pt	Height (in.)	Frequency
60.5	61 or shorter	3
61.5	61–62	4
62.5	62–63	10
:	63–64	18
	64–65	30
	65–66	52
	66–67	64
	67–68	116
	68–69	128
	69–70	147 ← 500th person
	70–71	129
	71–72	115
	72–73	63
	73–74	53
	74–75	29
	75–76	20
	76–77	12
77.5	77–78	5
78.5	taller than 78	2

425

1000 people

mean/mode/meadian should be ? close to equal if normal distrib.

mean:
$$\frac{60.5\times3+61.5\times4+62.5\times10\cdots78.5\times2}{1000}$$

$\bar{x} = 69.521$

$(\sigma = 2.987)$

median: 500th person if lined up by height
500th person is 69.5 in tall

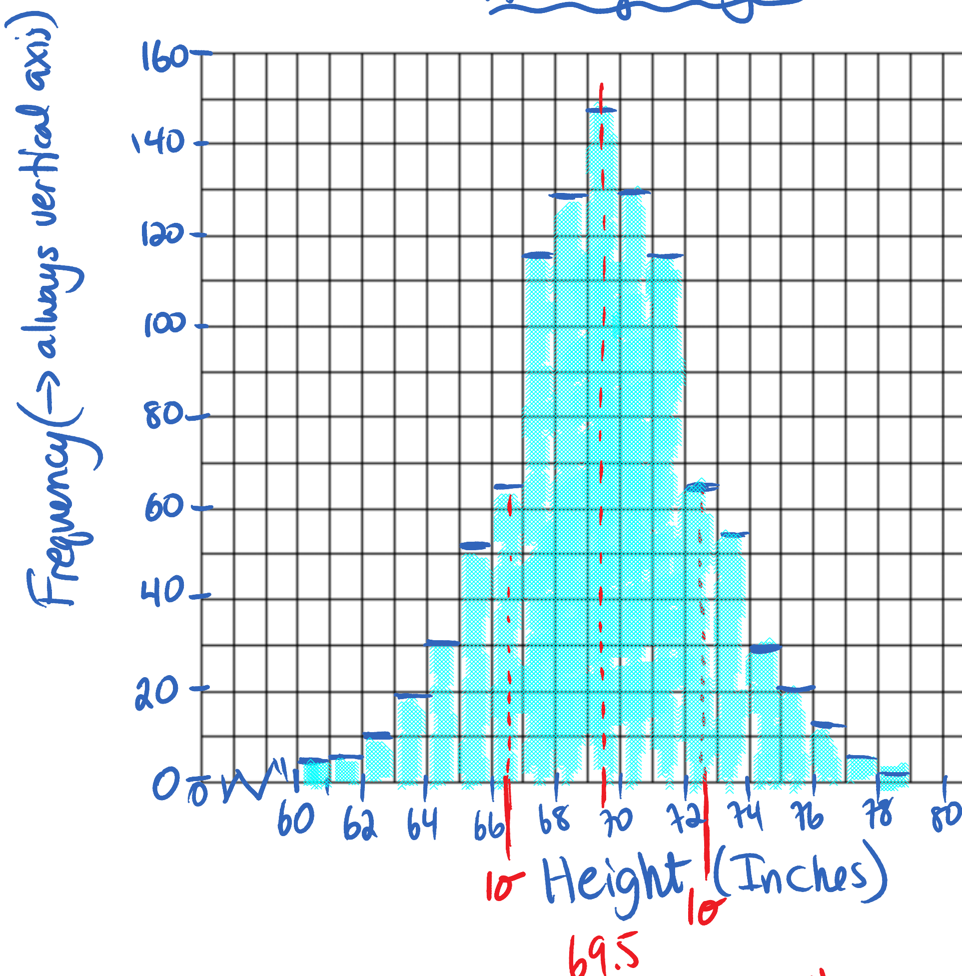
mode: "most common"
147 men between 69 and 70 in
= 69.5 in

The mean, mode and median are close to the same, so this seems to be a normal distribution.
Draw Histogram

Histogram

Histogram

Average Height



notice
- Bell curve
- symmetrical

→ normal distribution

Check: Are 68% of the men within 1σ of the mean (69.5 in)?

$1\sigma = 2.987 \sim 3$ inches

→ are 68% of men between 66.5 in and 72.5 in?

66-67 67-68 72-73

64 men + 116 + 128 + 147 + 129 + 115 + 63 = 762

$\frac{762}{1000} = 76\%$

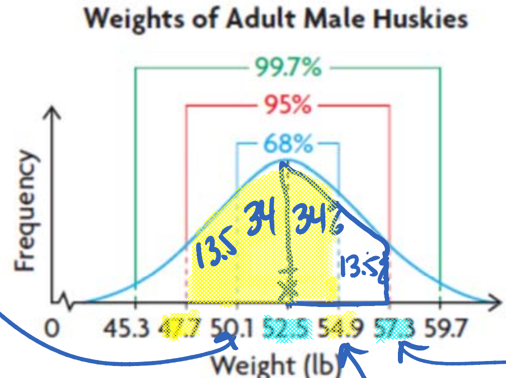
She should have 76% of her boards to fit men btwn 66.5 and 72.5 in

Example 2:

$\bar{x} = 52.5$ lbs
$\sigma = 2.4$ lbs

Jim raises Siberian husky sled dogs at his kennel. He knows, from the data he has collected over the years, that the weights of adult male dogs are normally distributed, with a mean of 52.5 lb and a standard deviation of 2.4 lb. Jim used this information to sketch a normal curve, with

- 68% of the data within one standard deviation of the mean
- 95% of the data within two standard deviations of the mean
- 99.7% of the data within three standard deviations of the mean



$\bar{x} - \sigma$: $52.5 - 2.4 = 50.1$

$\bar{x} + \sigma$: $52.5 + 2.4 = 54.9$

$54.9 + 2.4 = 57.3$

What percent of adult male dogs at Jim's kennel would you expect to have a weight between 47.7 lb and 54.9 lb?

$13.5\% + 34\% + 34\% = 81.5\%$

What % of dogs would be between 52.5 and 57.3 lbs?

$34\% + 13.5\% = 47.5\%$

Pg 251 #6, 7

Example 3:

Two baseball teams flew to the North American Indigenous Games. The members of each team had carry-on luggage for their sports equipment. The masses of the carry-on luggage were normally distributed, with the characteristics shown to the right.

Team	μ (kg)	σ (kg)
Men	6.35	1.04
Women	6.35	0.59

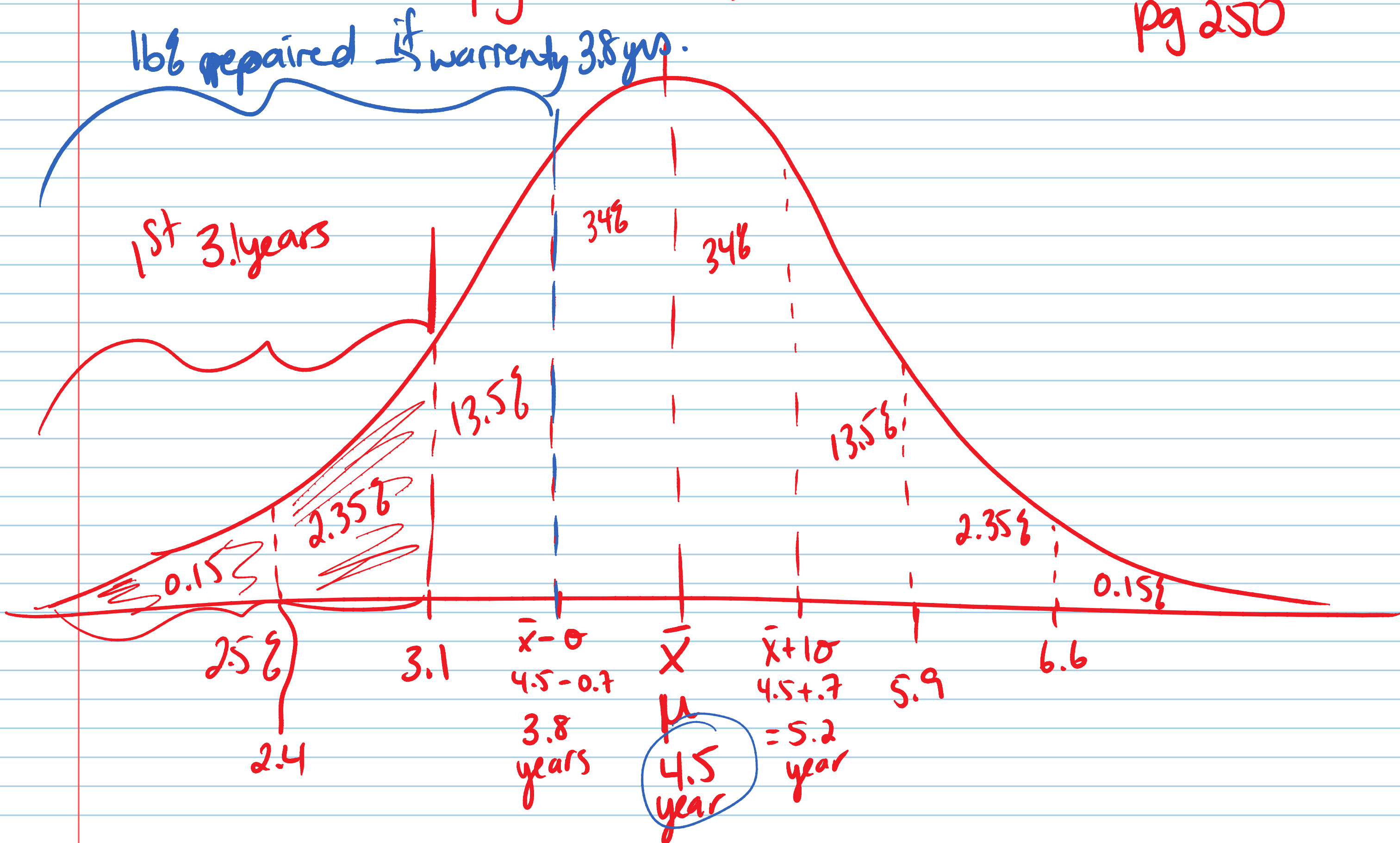
a) Sketch a graph to show the distribution of the masses of the luggage for each team.

b) The women's team won the championship. Each member received a medal and a souvenir baseball, with a combined mass of 1.18 kg, which they packed in their carry-on luggage. Sketch a graph that shows how the distribution of the masses of their carry-on luggage changed for the flight home.

5.4 Homework

pg 251 # 6, 7

pg 250



A warrenty for 3.1 years (3 years) is given and this means any coffee makers that break down after that, won't get repaired.
The company will only have to repair 2.5% (the area under the graph up to 3.1 years).

7

x Sum	f Freq	Total = Sum×freq	$(x-\bar{x})^2 \cdot f$
3	1	3	56.25 ← $(3-10.5)^2 \cdot 1$
4	3	12	126.75 ← $(4-10.5)^2 \cdot 3$
5	6	30	181.5
6	10	60	202.5
7	15	105	183.75
8	21	168	131.25
9	25	225	⋮
10	27 add	270	⋮

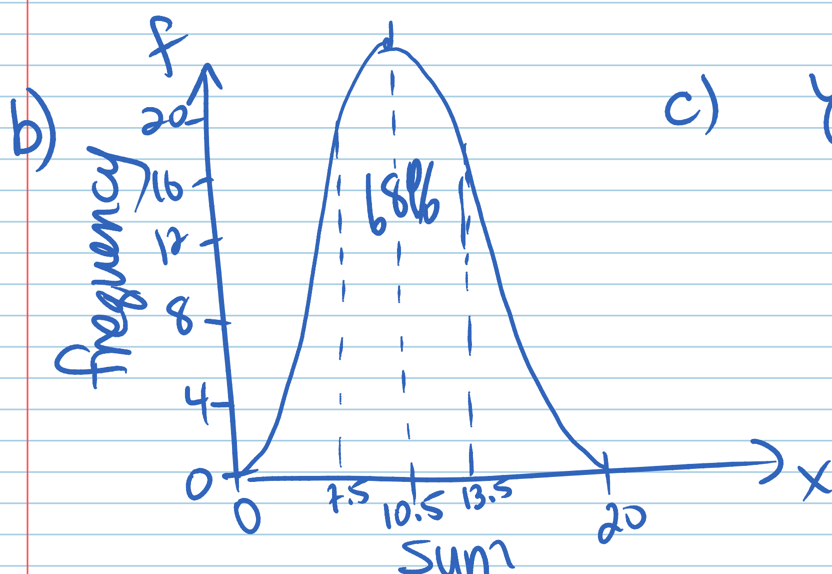
8	21	168	131.25
9	25	225	⋮
10	27	270	
11	27	297	
12	25	300	
13	21	273	
14	15	210	
15	10	150	
16	6	96	
17	3	51	
18	1	18	
	$n = 216$	2268	$\Sigma = 1890$

add

$$\text{mean} = \frac{3\cdot 1 + 4\cdot 3 + \dots}{216} = \frac{2268}{216} = \boxed{10.5 = \bar{x}}$$

$$\sigma = \sqrt{\frac{\Sigma (x-\bar{x})^2 \cdot f}{n}} = \sqrt{\frac{1890}{216}} = \boxed{2.96 = \sigma}$$

b)

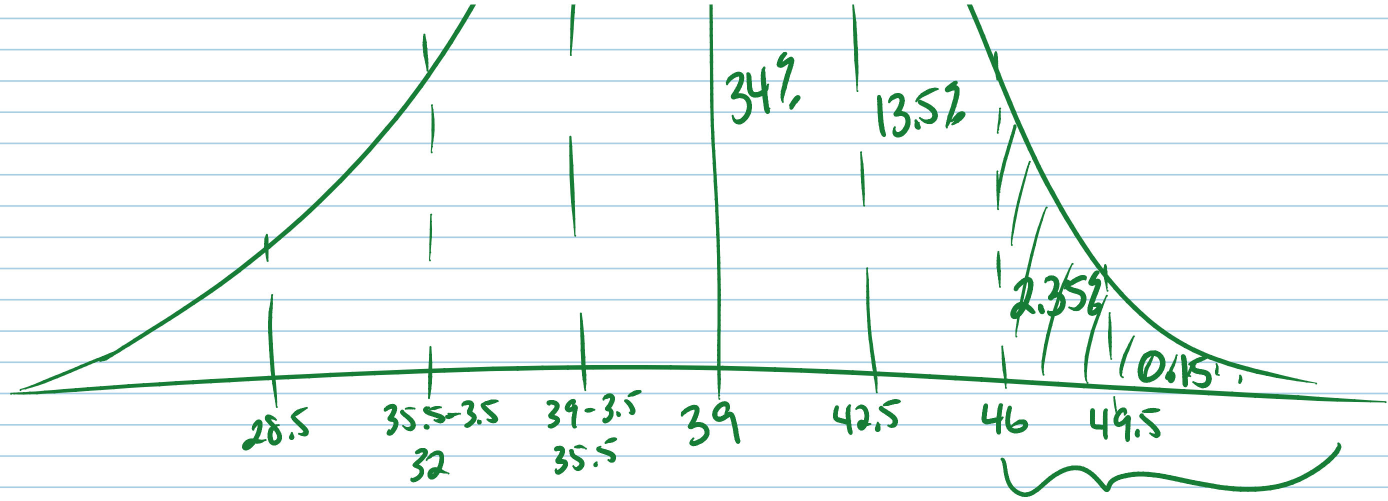


c) yes is normal distrib.

symmetrical

1 SD away from mean has 34% + 34% of the data

10



% that would live beyond 46 years = 2.5%

130 dolphins total

$2.5\% \times 130 = 0.025 \times 130 = 3.25$ dolphins

$\rightarrow$ 3 dolphins