Chapter 5 - Statistical Reasoning

Measures of Central Tendency

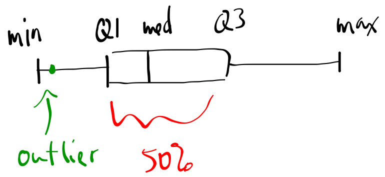
mean (~~average~~) → (sum of all values) ÷ n

median → middle value when sorted ascending or descending.

mode → most frequent value.

range → how spread out the data is (min, max)

outlier → a value that is very different from the other values



dispersion → how spread out the data it
(the range is a measure of the dispersion)

§5-1 Exploring Data

```
L1     L2     L3    1
3.1    6.8
5.2    5.7
5.5    5.9
5.7    4.8
5.7    4.5
5.7    5.8
5.8    5.6
L1(10) =5.1
```

3 occurrences
So mode is 5.7

	Brand X	Brand Y
mode	5.7 yrs	
mean ($\bar{x}$)	5.74 yrs	
median	5.75 yrs	
range	(3.1, 8.2)	

mean

```
1-Var Stats
x̄=5.743333333
Σx=172.3
Σx²=1034.37
Sx=1.242823075
σx=1.221933804
↓n=30
```

```
1-Var Stats
↑n=30
minX=3.1
Q1=5
Med=5.75
Q3=6.4
maxX=8.2
```

median

To Do:
p212/2

§5-2 Frequency Tables, Histograms + Frequency Polygons

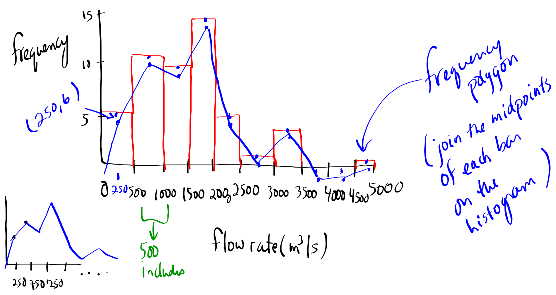
Frequency Table

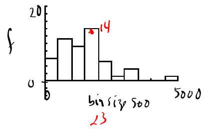
min ~ 159
max ~ 4587
} 0 - 5000, 10 groups

Flow rate (m^3/s)	Tally	Frequency	
0 - 500	HHH I	6	
500 - 1000		11	
includes 1000 → 1000 - 1500 ← doesn't include		9	
1500 - 2000		14	
2000 - 2500		5	
2500 - 3000		1	
3000 - 3500		3	floods
3500 - 4000		0	
4000 - 4500		0	
4500 - 5000		1	floods

Bin size → 500

Histogram (use the frequency table)





0 to 5000 bin size 500

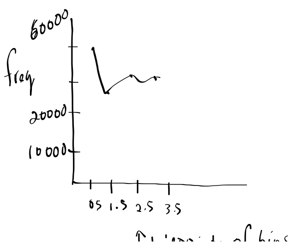
0 to 5000 bin size 1000

100 to 4600 bin size 500

TO DO

CYU (p221) - do by hand

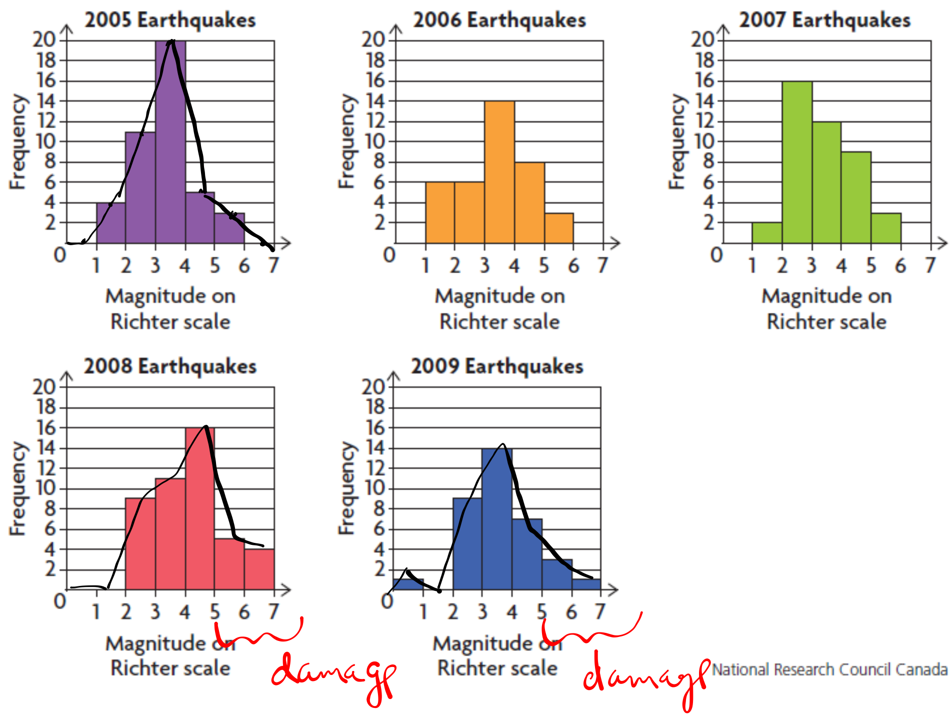
Practise (p222) - do by hand 3, 4 + 5 (check with calc)



← freq polygon.

EXAMPLE 2 Comparing data using histograms

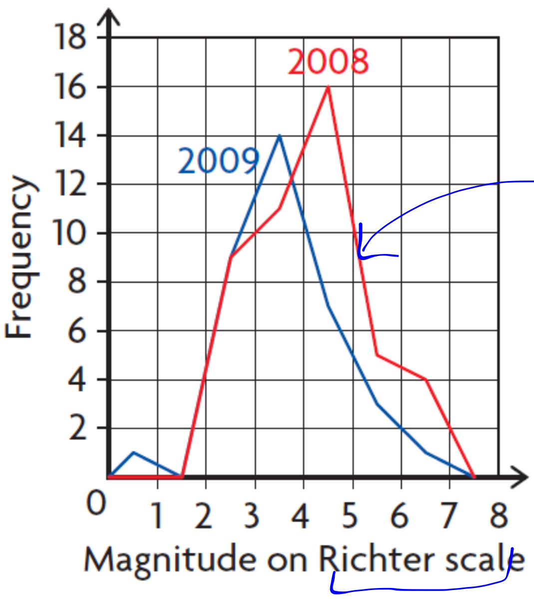
The magnitude of an earthquake is measured using the Richter scale. Examine the histograms for the frequency of earthquake magnitudes in Canada from 2005 to 2009. Which of these years could have had the most damage from earthquakes?



National Research Council Canada

Both 2008 and 2009 had the strongest earthquakes, registering from 6.0 to 6.9 on the Richter scale.

I exam histog than 6 had th registe damag



frequency polygon shows more damage (>5) for 2008

more damage

I decid of hist polygo

The number of earthquakes in the three hi intervals was greater in 2008 than in 2009, so 2 could have had the most damage from earthquakes.

I comp

To Do:
p222/3-5

§5-3 Standard Deviation (p226)

TEST	Class A	Class B
1	94%	84%
2	56%	77%
3	89%	76%
4	67%	81%
5	84%	74%
(mean) $\bar{x}$	78%	78%
range min	56%	74%
max	94%	84%
range ~~spread~~	38%	10%

Class B has more consistent marks since they are closer together and less spread out.

Both classes have about the same mean.

p226

Player	Field Goal Percent in Last 10 Basketball Games									
Anna	36	41	43	39	45	27	40	37	31	28
Patrice	36	39	36	38	35	37	35	36	38	34
Morgan	34	41	38	37	48	19	33	43	21	44
Paige	34	35	33	35	33	34	33	35	34	33
Star	41	33	39	36	38	36	29	34	38	39

$\bar{x} = ?$
33.9

How can the coach use the data to determine which player should be substituted into the game?

A. Which player seems to be the most consistent shooter? Explain.

B. Analyze the data for Paige using a table like the one shown on the next page. Determine the mean of the data, $\bar{x}$, for Paige, and record this value in the first column.

p226

$\bar{x} = 33.9$

Paige's Field Goal (%)	Deviation $(x - \bar{x})$	Square of Deviation $(x-\bar{x})^2$
34	$34 - 33.9 =$ 0.1	0.01
35	1.1	1.21
33	−0.9	0.81
35	1.1	1.21
33	−0.9	0.81
34	0.1	0.01
33	−0.9	0.81
35	1.1	1.21
34	0.1	0.01
33	−0.9	0.81
		6.90

σ

Σ

big sigma (means sum)

little sigma (standard deviation) $\rightarrow$ $\sigma = \sqrt{\dfrac{\Sigma(x-\bar{x})^2}{n}}$

$\sigma = \sqrt{\dfrac{6.90}{10}}$

$\sigma = \sqrt{0.69}$

$\sigma \doteq 0.83$

```
1-Var Stats
 x̄=33.9
 Σx=339
 Σx²=11499
 Sx=.8755950358
 σx=.8306623863
↓n=10
```

mean

stand. dev.

mean:

$\bar{x} = \dfrac{\Sigma x}{n}$

Standard deviation tells you how clustered the data is about the mean. The smaller the value the closer the data is to the mean

Standard deviation:

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

	$\bar{x}$	σ
Anna		
Patrice		
Morgan		
Paige	33.9	0.83
Star		

Example 1 (p228)

05_Chapter 5.pdf - Adobe Reader

File Edit View Window Help

Open 228 (23 of 84) 150% Tools Fill & Sign Comment

Sign Document

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.

L1

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

L2

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?

228 Chapter 5 Statistical Reasoning

	227g bags		454g bags	
min	218 g	18g	445g	18g
max	236g		463g	
mean ($\bar{x}$)	227.15g		454.4g	
St.dev (σ)	5.227 g		4.498g	

more consistent

* Both samples have the range, but the 454g bags were more consistent (less spread out).
* Standard deviation is a better measure of the dispersion of data as opposed to the range.

To Do

1. Finish the BBall example with calculator.
2. CYU (p233) / #1 (do by hand)
3. Practice (p233) / 5 - 8 (with calc)
4. MidChapter Review (p238-240 / 1-5)