Foundations 11

Class notes: Chapter 5

5.1 Exploring Data

mean - average

median - middle number (numbers must be in order least - greatest)

mode - number occuring most often

range - largest number - smallest number

dispersion - increases as data becomes more spread out

* remember: when comparing data look at both similarities + differences

5.2 Frequency Tables, Histograms and Frequency Polygons

To create a Frequency Table

① decide on intervals depending on data.

ex

highest # is 200
lowest # is 4

> intervals could be by 20's

0-20
20-40
40-60
⋮

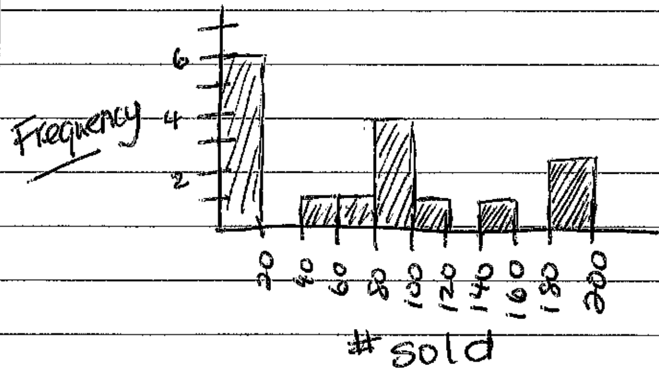
(2) add the data to the table by doing a Tally

Data
4, 7, 12, 15, 200, 199, 18, 50, 75, 82, 93, 101, 19, 155, 11, 92, 87, 193

	Tally	Frequency
0-20	~~||||~~ |	6
20-40		0
40-60	|	1
60-80	|	1
80-100	||||	4
100-120	|	1
120-140		0
140-160	|	1
160-180		0
180-200	|||	3

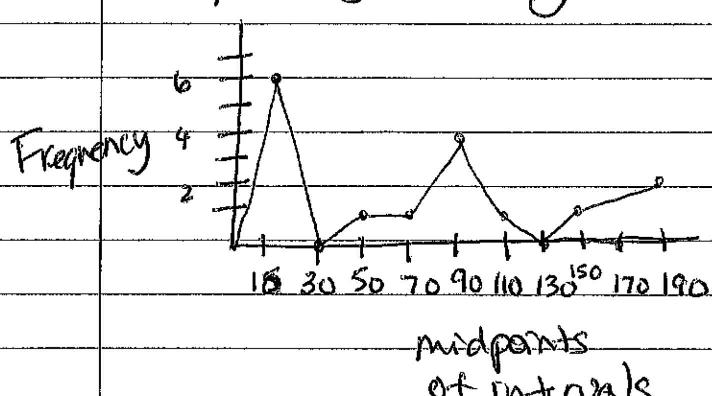
← add up tally marks

Histogram - is a bar graph without spaces



→ the spaces in this graph indicate 0

Frequency Poly gon - a line graph using mid point of interval



← add data

midpoints	Frequency
10	6
30	0
50	1
70	1
90	4
110	1
130	0
150	1
170	0
190	3

Comparing Data

You can use histograms to compare or you can use some of the data & create a frequency table
See page 218 & 219

Note:

A frequency table should have a minimum of 5 intervals & a maximum of 12

Interval width can be determined by dividing the range (largest - smallest) by the number of intervals you want

Frequency polygons are useful because multiple graphs can be graphed on the same grid.

Frequency Tables / Polygons & Histograms give an overview of data showing trends - to find mean, mode or median, you must go back to the data.

5.3 Standard Deviation

deviation – the difference between the data value and mean of the data

Standard deviation – measure of the scatter of the data
- low number = data close to mean
- high number = data scattered from the mean

Symbols used:

$\bar{x}$ = mean (add all numbers then divide by quantity of numbers)

$$\bar{x} = \frac{\sum x}{n}$$

$\sum$ = sum of numbers

n = quantity of numbers

σ = standard deviation

Formula

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

Data

7, 6, 7, 9, 6, 7, 12, 7, 6, 15, 12,

1

	Tally
6	3
7	4
9	1
12	2
15	1

2

$$\bar{x} = \frac{6\times3 + 7\times4 + 9\times1 + 12\times2 + 15\times1}{11} = \frac{94}{11} \Rightarrow \bar{x} = 8.5$$

3

	f	$(x-\bar{x})^2$	$f\cdot(x-\bar{x})^2$
6	3	$(6-8.5)^2 = 6.25$	$3\cdot6.25 = 18.75$
7	4	$(7-8.5)^2 = 2.25$	$4\cdot2.25 = 9$
9	1	$(9-8.5)^2 = 1.12$	1.12
12	2	$(12-8.5)^2 = 12.25$	$2\cdot12.25 = 24.5$
15	1	$(15-8.5)^2 = 42.25$	42.25
	11		add 95.62 $\Leftarrow \sum(x-\bar{x})^2 \Rightarrow$

4

$$\sigma = \sqrt{\frac{95.62}{11}} = 2.9$$

To find standard deviation with intervals in a frequency table

* find midpoint

(1)

	f	midpoint (x)	$f \cdot x$
0-5	4	2.5	10
5-10	2	7.5	15
10-15	8	12.5	100
15-20	1	17.5	17.5
20-25	3	22.5	67.5
	$\boxed{18}$		$\boxed{210}$

add all f $\Rightarrow n$

add $f \cdot x$ to get $\Sigma(f)(x)$

To find $\bar{x} = \dfrac{\Sigma(f \cdot x)}{n}$

$= \dfrac{210}{18} = 11.7$

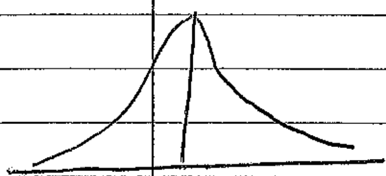
(2)

midpoint	f	$(x-\bar{x})^2$	$f \cdot (x-\bar{x})^2$
2.5	4	$(2.5-11.7)^2 = 84.64$	$4 \cdot 84.64 = 338.56$
7.5	2	$(7.5-11.7)^2 = 17.64$	$2 \cdot 17.64 = 35.28$
12.5	8	$(12.5-11.7)^2 = 0.64$	$8 \times 0.64 = 5.12$
17.5	1	$(17.5-11.7)^2 = 33.64$	$1 \times 33.64 = 33.64$
22.5	3	$(22.5-11.7)^2 = 116.64$	$3 \times 116.64 = 349.92$
	n= 18		add to get $\Sigma f(x-\bar{x})^2$ $\boxed{762.52}$

$$\sigma = \sqrt{\frac{\Sigma f(x-\bar{x})^2}{n}} = \sqrt{\frac{762.52}{18}} = 6.5$$

5.4 The Normal Distribution

Normal distribution – when a set of data is graphed, it results in a symmetrical pattern around the mean.



<u>Normal curve</u> – a symmetrical curve in a graph that shows normal distribution (bell curve)

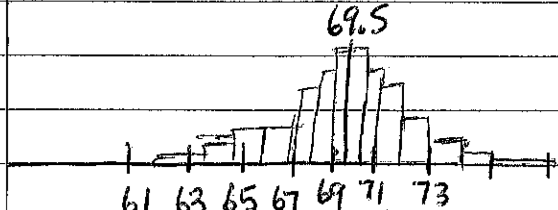
* use the standard deviation & the normal distribution to predict

<u>Data</u>

$\boxed{\mu} = \overline{x} = 69.52$ in $\quad \sigma = 2.987$ in $\quad$ total surveyed $= 1000$

new symbol

range = 61 in or shorter → 78 in or taller



$\boxed{\text{To find h within } 1\sigma}$

$69.5 - 2.987 \sim 66.5$ in

$69.5 + 2.987 \sim 72.5$ in

$\boxed{\text{To find \%}}$

– 100k on page 242

1σ – add frequency between 66.5 and 72.5 (67–73)

↳ $116 + 128 + 147 + 129 + 115 + 63 = 698$ $\Rightarrow \frac{698}{1000} = 70\%$

$\boxed{\text{To find h within } 2\sigma}$

$69.5 - 2(2.987) = 63.5$

$69.5 + 2(2.987) = 75.5$

÷ by total

2σ – add frequency between 64–76

$= 30 + 52 + 64 + 116 + 128 + 147 + 129 + 115 + 63 + 53 + 29 + 20 = 946 \Rightarrow \frac{946}{1000} = 95\%$

Using the Normal Curve

You are given the following data

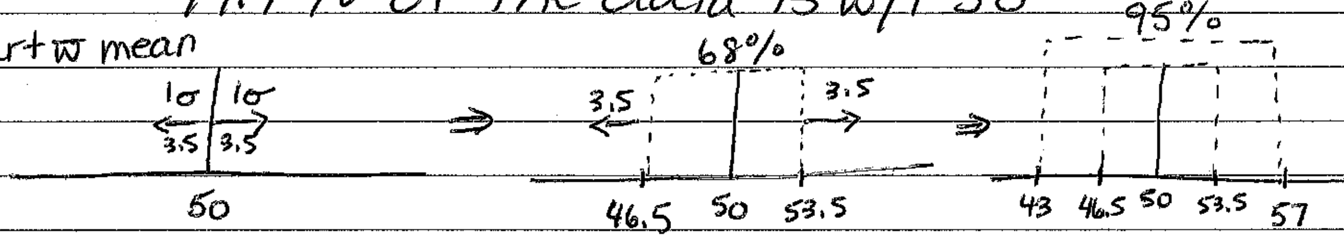
The mean weight of a 6 yr old is 50 lbs with a standard deviation of 3.5 lbs.

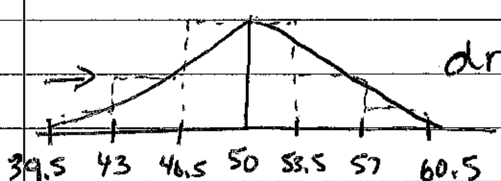
Using the normal curve

68% of the data is w/i 1σ
95% of the data is w/i 2σ

99.7% of the data is w/i 3σ

Start w mean

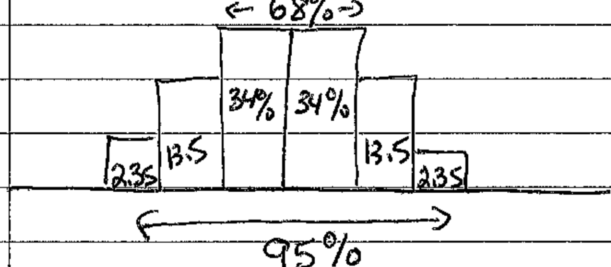


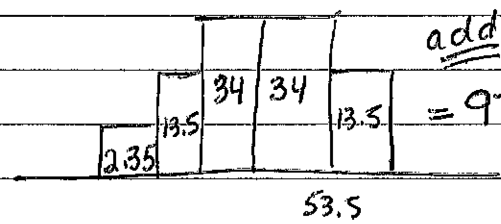


draw normal curve.

How to find % of given value (43 - 57 lb)

*use natural curve





add
= 97.4%

95% - 68% = 27% or 13.5% each side

99.7% - 95% = 4.7% or 2.35% each side

5.5 Z-scores

$$z = \frac{x - \mu}{\sigma}$$

μ = mean
σ = standard deviation

Z score → shows number of standard deviations a value is from the mean
→ − means below; + means above
→ good for comparing 2 values from the data.

	mean	σ	Hailey	Serge
run 1	25.75	0.62	24.95	25.95
run 2	26.03	0.70	25.62	25.23

Compare Hailey's time:

$z = \frac{x - \mu}{\sigma}$ run 1 $= \frac{24.95 - 25.75}{0.62} = -1.29$ * better time

run 2 $= \frac{25.62 - 26.03}{0.70} = -0.59$

Compare Serge to Hailey

run 1 $= \frac{25.95 - 25.75}{0.62} = +0.32$ → Hailey did better on run 1

run 2 $= \frac{25.23 - 26.03}{.70} = -0.8$ → Serge did better on run 2

Compare Serge
− he did better in run 2 by 1.12

Using z scores to determine data values.

	0.09	0.08	0.07	0.06
-0.7	0.2148	0.2177	0.2206	0.2236
-0.6	0.2451	0.2483	0.2514	0.2546
-0.5	0.2776	0.2810	0.2843	0.2877
⋮	⋮	⋮	⋮	⋮

← these numbers represent decimal form of %

Running Shoes loose shock-absorption at about 640 km with standard deviation of 160 km

You want to replace your shoes at 28%

	0.09	0.08
-0.5	0.2776	0.2810

between

add 2 numbers = -0.58 and -0.59

So about -0.585

plug this value into the equation

$$z = \frac{x - \mu}{\sigma} \qquad -0.585 = \frac{x - (640)}{160}$$ (x both sides by 160)

$$-93.6 = x - 640$$ (+640 to both sides)

$$546.4 = x$$

* replace shoes at 546 km

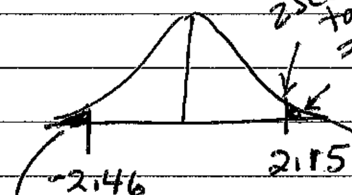
Use z scores to solve a word problem

Bungee chords need to be between 48–42cm

mean = 45.2 cm σ = 1.3cm

$$z_{min} = \frac{42-45.2}{1.3} = -2.46$$

$$z_{max} = \frac{48-45.2}{1.3} = 2.15$$



z score % from table = 0.9842 this means all area to left so subtract from 1 (100%

area = z score % from table = 0.0069

100% − 98.42%
or 1 − 0.9842 = 0.0158

Total # rejected
0.0069 + 0.0158 = 0.0227 or 2.27%

IF the factory makes 20000
then 20000 x 0.0227 or 454
are rejected.

5.6 Confidence Intervals

<u>margin of error</u> - possible difference between the estimate & real value
- it is true for entire population
- expressed as ± %

<u>confidence interval</u> - range of values you are trying to estimate using margin of error
ex 50% ± 2.3% or 47.7 → 52.3

<u>Confidence level</u> - % that the result will lie in confidence interval
ex. accurate 19 times out of 20 ⇒ 95%.

<u>Ex</u> 600 people take a survey

76% of the people aged 18-34 have a social networking account

Results are accurate ± 4%, 19 times out of 20

<u>confidence interval</u>	<u>confidence level</u>
76% − 4% = 72%	$\frac{19}{20} = 95\%$
76% + 4% = 80%	

So if the total # of people betwen 18-34 is 100,000

Confidence = 100000 × 4% = 4000

\# → 100000 × 76% = 76,000 $\boxed{\text{So } 76{,}000 \pm 4000}$ people have social networking account w 95% confidence