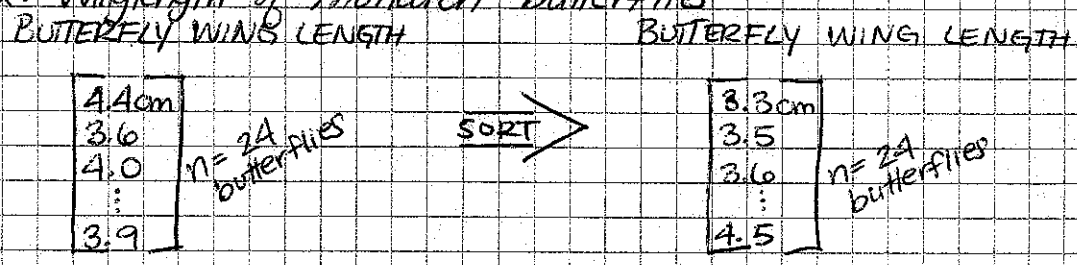


01/13/09

# RAW FREQUENCY DISTRIBUTION

When we collect raw data, it tends to be disordered. First step is to sort the data (ex: smallest  $\rightarrow$  largest)  
 ex: Winglength of monarch butterflies



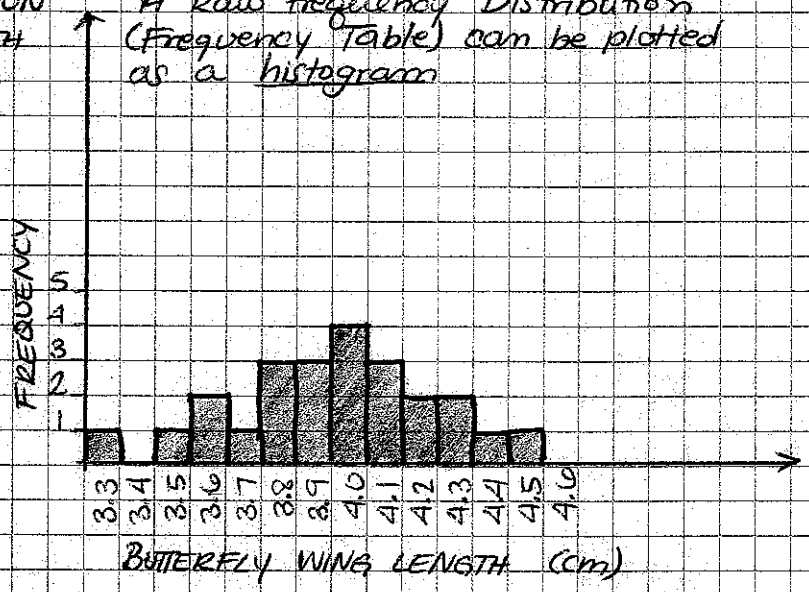
Sorted table is used to make a raw frequency distribution

**RAW FREQUENCY DISTRIBUTION**

BUTTERFLY WING LENGTH Value	Frequency
3.3cm	1
3.4	0
3.5	1
3.6	2
3.7	1
3.8	3
3.9	3
4.0	4
4.1	3
4.2	2
4.3	2
4.4	1
4.5	1

Total:  $n=24$

A Raw Frequency Distribution (Frequency Table) can be plotted as a histogram



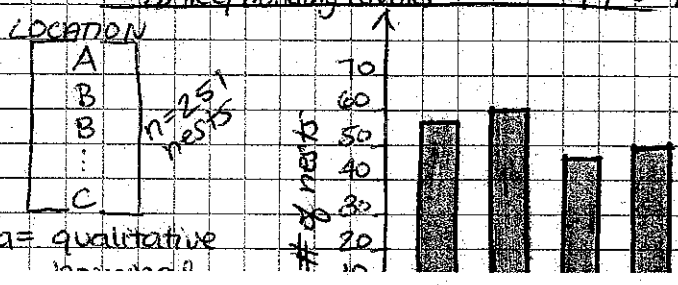
**HISTOGRAM**: a bar graph that helps to visualize a frequency table and see patterns in the value frequencies

- Can only be used for quantitative data
- gap between values on the X-axis (axis that shows the value in question) must always be the same  
 gap from 3.3  $\rightarrow$  3.4 = gap from 4.0 - 4.1 = 0.1cm

Tip for good statistical work: VISUALIZE THE RAW DATA FROM A NUMERICAL SUMMARY

ex: Frequency Table - Location of Sparrow Nests  $\rightarrow$  # of nests observed

- |                          |    |   |
|--------------------------|----|---|
| A) Vines                 | 50 | } may look quantitative because of the numbers involved, but when you reproduce raw data: |
| B) Building eaves        | 60 |   |
| C) tree branches         | 40 |   |
| D) tree/building ravines | 49 |   |



Because data is not quantitative, histogram not possible

Bar graph possible & useful because shows visually where sparrows prefer to nest

Data = qualitative

01/13/09

EX: Amount of pigmentation in sunfish

PIGMENTATION CLASS	AMT OF PIGMENTATION	# OF FISH
0	no black pigmentation	13
1	faintly speckled	68
2	moderately speckled	41
3	heavily speckled	21
4	solid black pigmentation	8

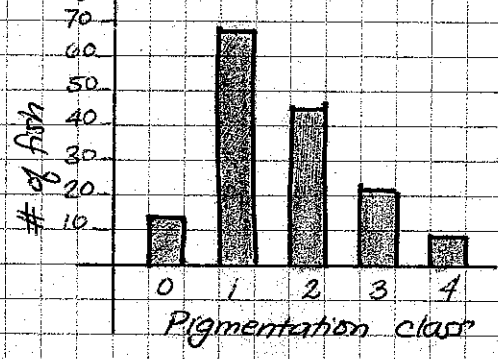
Raw Data Table:

BLACK PIGMENTATION

- faint
- heavy
- none
- :
- faint

n = 154 sunfish

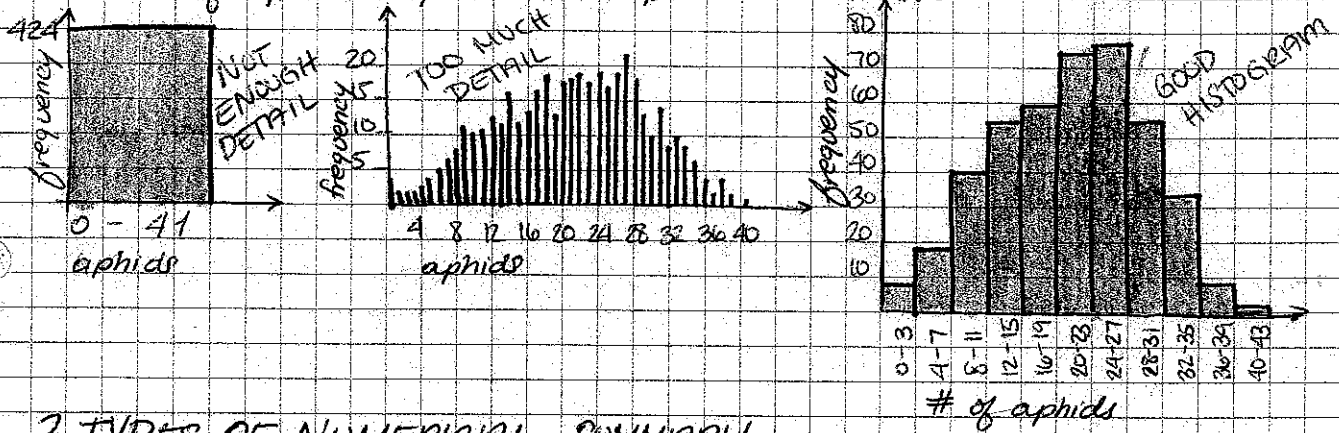
qualitative  
ordinal



The bar graph in the sunfish example is still not a histogram - Data is qualitative, ordinal, meaning that the gaps between the values may not be equal throughout.  
 • Bar graphs of ordinal variables can look a lot like histograms & give a lot of similar information.

Values in a histogram can be organized in different ways, depending on the detail needed to clear up uncertainty.

- EX. of aphids w/ 41 data points on X axis:



2 TYPES OF NUMERICAL SUMMARY

- MEASURES OF CENTER (central tendency)

- mean = average ( $\bar{y}$ )

$$\bar{y} = \frac{y_1 + y_2 + \dots + y_{(n-1)} + y_n}{n}$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

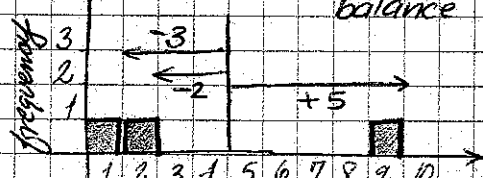
(p. R.26 in reader has list of formulas)

- Graphical interpretation as balance point of data

ex: sample data

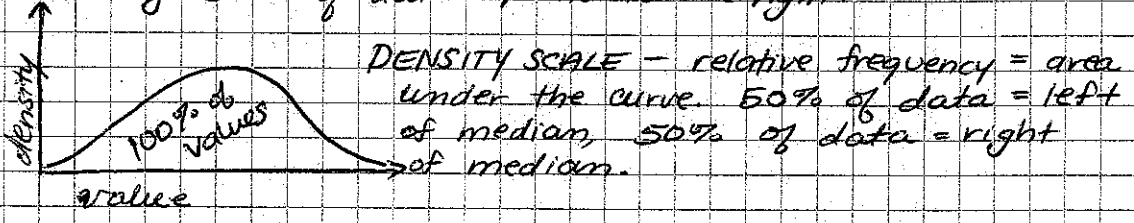
1	n = 3
2	
9	

mean =  $\frac{1+2+9}{3} = 4$



P. 10  
01/13/09

- median ( $\tilde{x}$ ) - the middle value of a list of values that are in order from smallest to largest
  - If there is an odd number of values, median is uniquely defined by the middle value
  - If data set has even # of values, median is the mean of the 2 middle numbers
  - On a histogram, median is where 50% of data points is to the left & 50% of data points to the right.

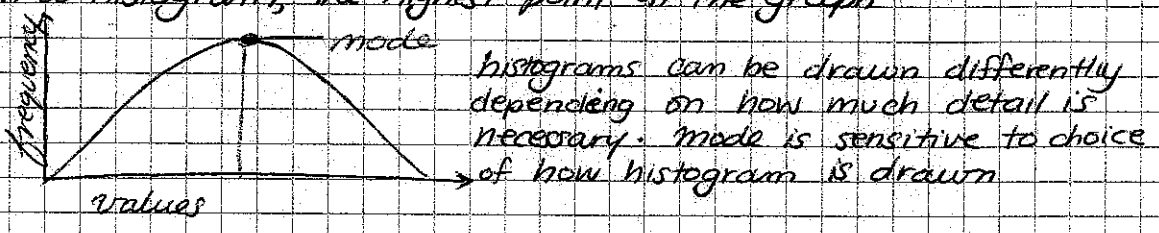


ex: sample data

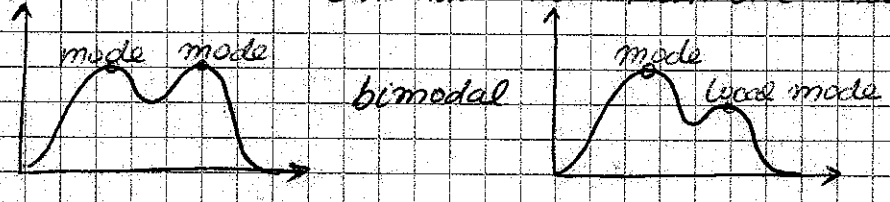
1	n=3
2	
9	

median = 2

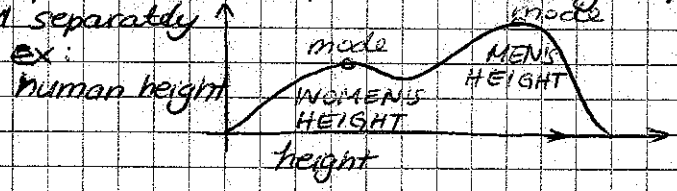
- mode: value w/ the highest frequency, value that recurs most often
  - On a histogram, the highest point on the graph



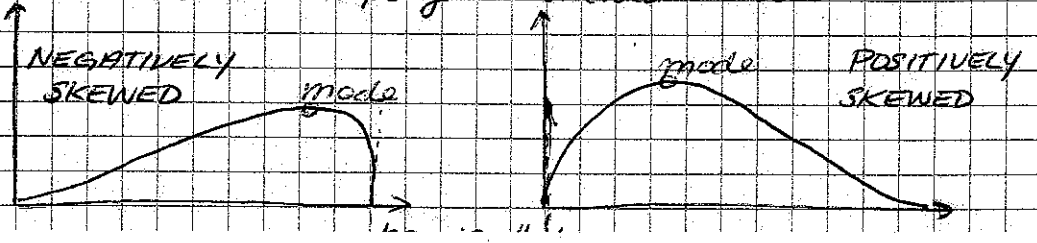
- a set of values can have more than one mode



A multimodal histogram tends to mean that there are multiple subgroups in the sample → the subgroups should be studied separately

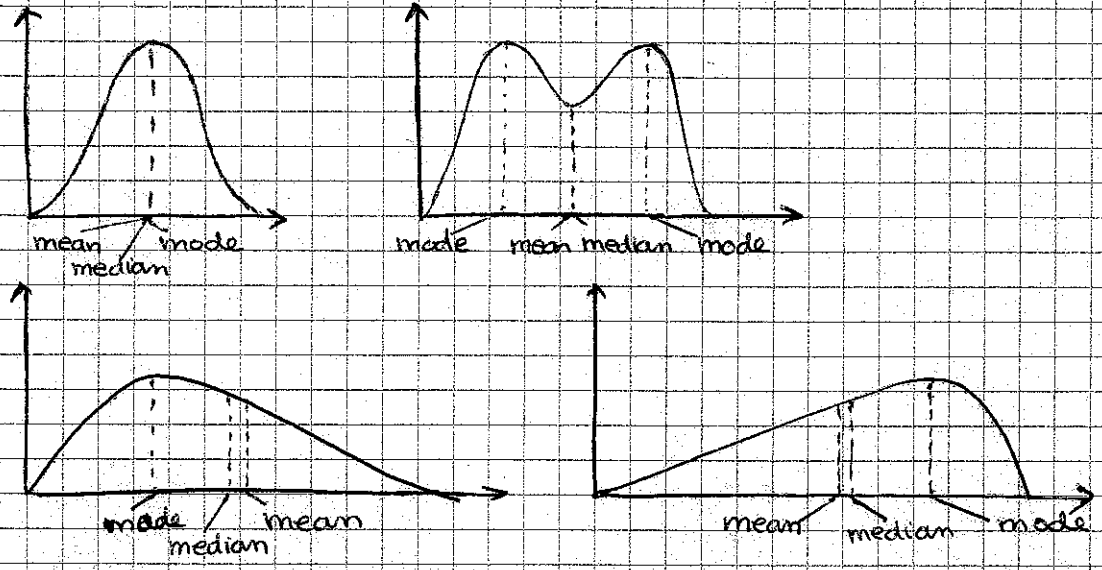


When variables can't go above or below a barrier, graph can have a skewed shape & a skewed mode

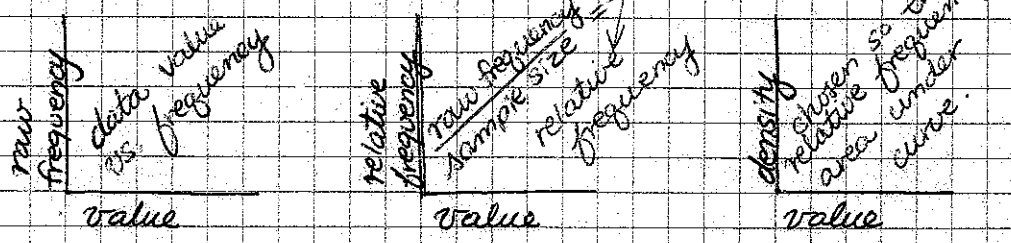


01/15/07

MEAN, MEDIAN, & MODE COMPARISONS IN SAMPLE GRAPHS



WAYS TO DRAW HISTOGRAMS



default sketches of histograms in class are always on density scale

- On Density scale
  - relative frequency = area under the curve
  - 50% of data = left of median, 50% of data = right of median
  - normal curve drawn to density scale

01/15/09

PERCENTILE

The median is a special case of percentile → it is the 50<sup>th</sup> percentile

- 25<sup>th</sup> percentile (1<sup>st</sup> quartile) → 25% of the data is to the left of the first quartile.

