

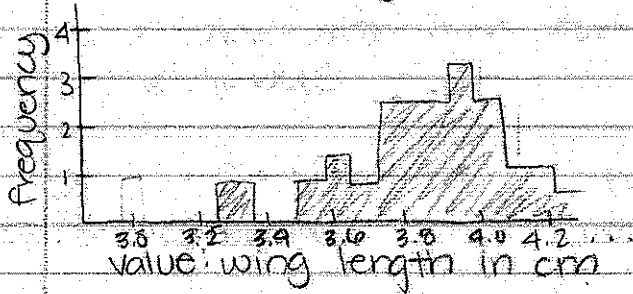
Histograms, measures of center + spread

10/2

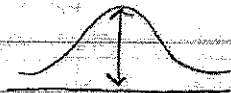
NOTE labs start next week 10/6

Example from butterfly wing length

* Raw frequency histograms

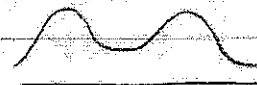


mode = highest freq. or highest pt. of histogram



1 mode = uni-modal

bimodal ex: height of ppl. in class (♀ generally shorter than ♂)



2+ modes = bimodal

variable is question is quantitative (place on # line)

* Ex: where sparrows like to build nests

Nest sites	# of nest observed
vines	56
building eaves	60
tree branches	46
tree/building cavities	49

} raw freq. distribution = SUMMARY

raw data

Nests
B
V
B
:

n = 211

- variable is qualitative + nominal
- Not dichotomous b/c more than 2 variables
 - ↳ can't draw histogram
 - ↳ CAN draw bar graph

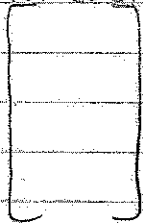
* diff. btwn bar graph + histogram

• histogram: variables have unique place on # line (no choice of where to place them)

- bar graph: no specific place for "vines" or "trees" on # line

*secret of statistical work → visualize raw data as a table with rows for subjects & columns for variables

Ex. Sunfish Pigment
1 row for each sunfish



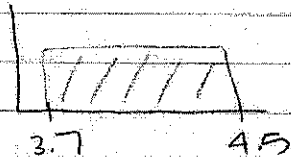
n=154

Pig- class	Amount of Pig	# of fish
0	no black	13
1	faint speckle	68
2	Mod. speckle	44
3	heavy "	21
4	solid black	8

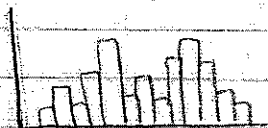
could make these variables diff. #'s & then the data would be located in diff. place on histogram.

variable = qualitative, ordinal, no histogram, only bar graph

*There are many possible histograms for the same data set depending on how wide the bars are where they start.



1 bar = loose shape



too many bars = can't see shape well

Example: litter size

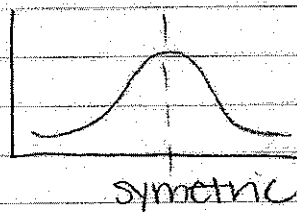
5
3
6
⋮

$n = 64$

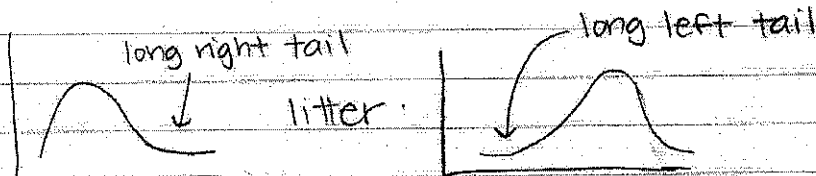
Variable quantitative, discrete, ratio, can make both histogram & bar graph.

1 row for each litter

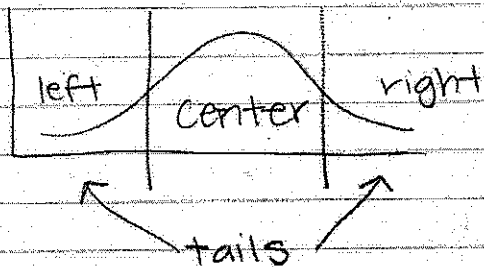
histogram is bar graph



butterfly: mode in middle
* unimodal



asymmetric = skewed



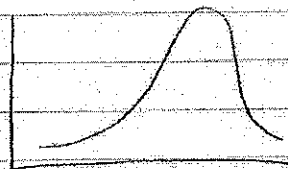
Ex: clover plant & aphids

aphids

24
7
19
⋮

1 row for each plant

Variable quantitative, discrete, ratio, histogram



long left hand tail

ratio = zero is meaningful

Ex: phosphorus in leaves

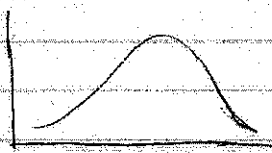
variable: [Phosp.]

8.92
8.64
⋮

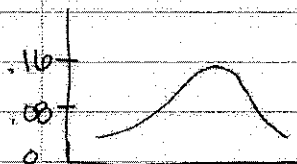
1 row for each leaf

variable quantitative, continuous, ratio (0 = absence of phos.)

raw freq. scale

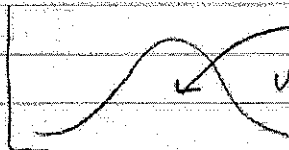


* unimodal, slight left hand tail



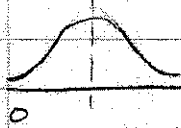
← relative freq. histogram

density scale

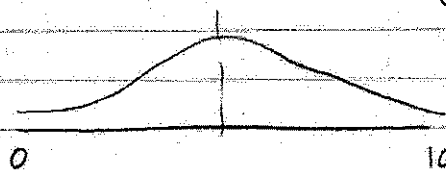


total area under curve which summarizes hist. is 1 = 100% & relative freq. is characterized by area under curve.

mode = 25



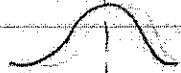
100



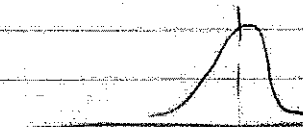
0

100

mode = 75



100



0

100

same shape different center

* same shape, same center, different spread

* Measures of center

1. Mode

2. mean →

3. mode

variable y
 $\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$
 \uparrow
 n
 \downarrow

mean = \bar{y}

$$= \frac{1}{n} (y_1 + \dots + y_n)$$

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

butterfly data mean:

$$\text{mean } \bar{y} = \frac{4.4 + \dots + 3.9}{24}$$

$$= 3.96 \text{ cm}$$

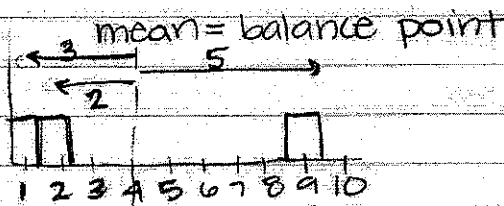
4.4
⋮
3.9

$n = 24$

$$\text{mode} = 4$$

* Data Set (new)

1
2
9

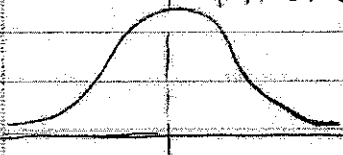


↑ 9 = outlier (unusually high or low)

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

* # line = plywood, histogram = bricks, mean = place of balance

* pt. of symmetry = mode = mean *



symmetric + unimodal

* Median: the middle value in the list in which (y_1, \dots, y_n) has been sorted from smallest to largest.

• halfway pt. in terms of relative raw freq. of data

1
2
9

→ median ←

1
7
9

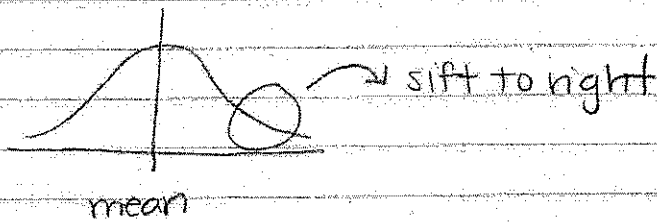
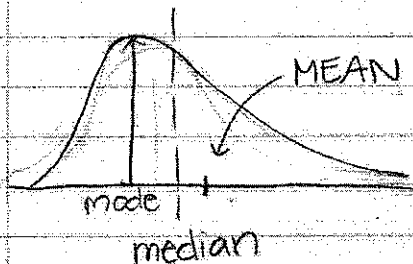
1
2
3
9

$$\frac{2 + 3}{2} = 2.5$$

↑ median

* graphical interpretation

- median = area to the left is 50% of data +
 area to the right is other 50% of data



- ∴ with long right hand tail: 1. mode 2. median 3. mean
- with long left hand tail: 1. mean 2. median 3. mode

* Back to butterfly data:

median = 4.0 cm (1/2 data to right, other 1/2 to left)

