AMS 7
12 Apr
17

this time: numerical measures of center & spread

next time: the normal curve

variable types: (1)

hwk 1
new due date: 11:55pm on wed
19 Apr 17

ex. litter size in foxes (1)-19

pups
4
7
3
4
⋮

← 1 row for each litter

n = lot + 1

hist? yes

quant, discrete, ratio

jumps between possible values represent impossibility

mean? yes

aphids on clover

aphids
14
36
⋮
⋮

← 1 row for each clover plant

n = 424

hist? yes

mean? yes

quant., discrete, ratio

phosphorus conc. (y)

8.42
9.08
⋮

← 1 row for each leaf

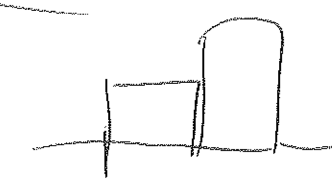
n = 130

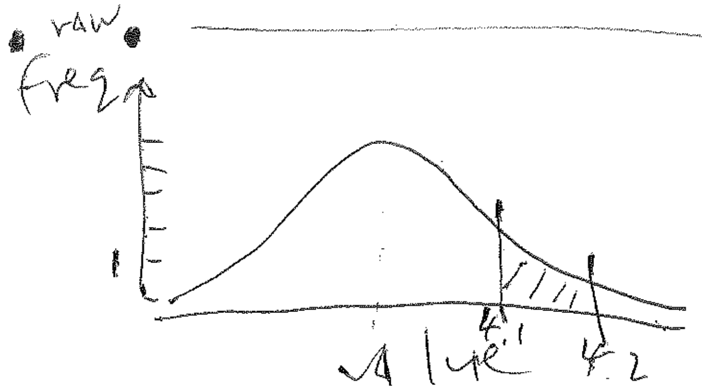
quant. (conceptually) continuous, ratio

mean? yes!

hist? yes

$$\frac{y_{max} - y_{min}}{\# \text{ bars}} = \text{bar width}$$





relative freq. (%)

$\frac{1}{24} \times 100 = 4\%$

density scale

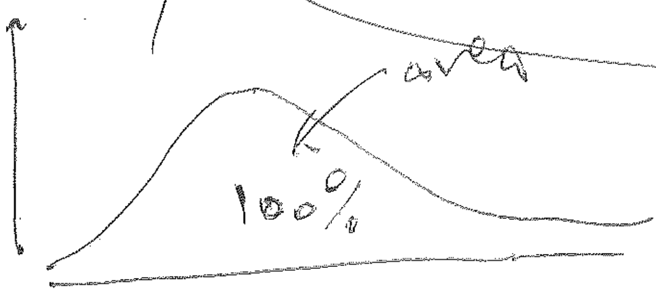
---

convention: unless otherwise noted, all hist. in this class will be drawn on the density scale

density scale chosen so that

relative frequency = area under hist. (curve)

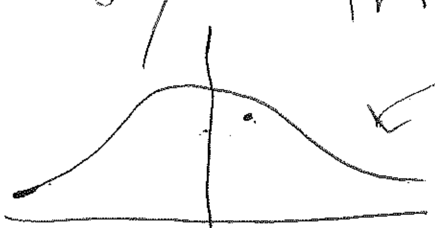
density



with hist. on density scale, total area under curve = 1 = 100%

histogram shapes

symmetric | unimodal | asymmetric = skewed

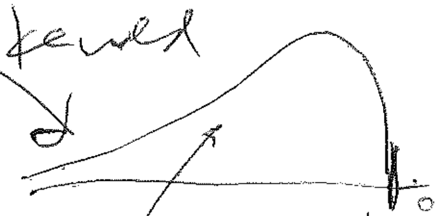


phosphorus

point of symmetry = mode

income

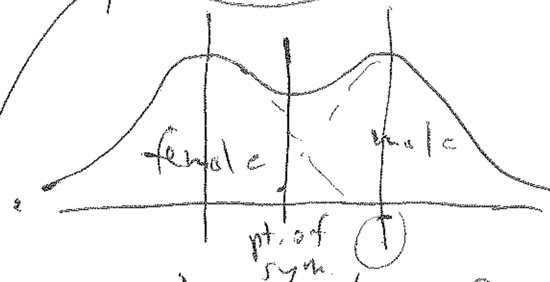
long right-hand tail



midterm scores

long left-hand tail

bimodal



height of everybody in class today

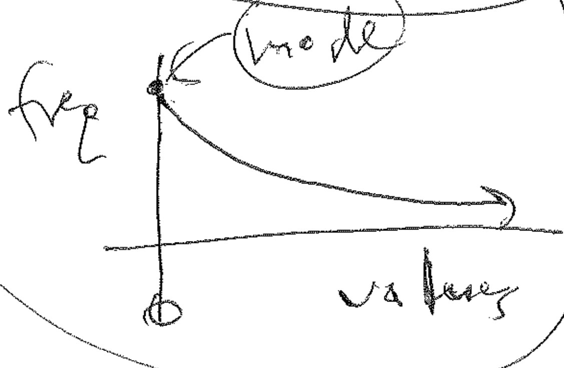
or

multimodal

mode = point of highest frequency

quantitative (numerical) measures of center of a distribution

(1) mode



↑ not necessarily central, but always typical in terms of freq

L-15

(2) mean

(butterflies)

wing length (cm)

$$\begin{pmatrix} 4.4 \\ 3.6 \\ \vdots \\ 3.9 \end{pmatrix} \quad n = 24$$

mean = ?

variable y

generic data set

y_1 — subscript

y_2

⋮

$y_n = y_{24}$

n = sample size = 24

mean $\bar{y} = \dfrac{y_1 + y_2 + \dots + y_n}{n}$

"y-bar"

$= \frac{1}{n}(y_1 + y_2 + \dots + y_n)$

~~capital sigma~~

Summation notation

$$\sum_{i=1}^{n} y_i = y_1 + y_2 + \dots + y_n$$

(expand out the sum)

largest index → n

$i = 1$ ← smallest index (subscript)

index of summation i, j, k

the sum, as i runs from 1 to n, of

summation notation tutorial (web search string)

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$ in this data set

↑ sample mean of variable Y

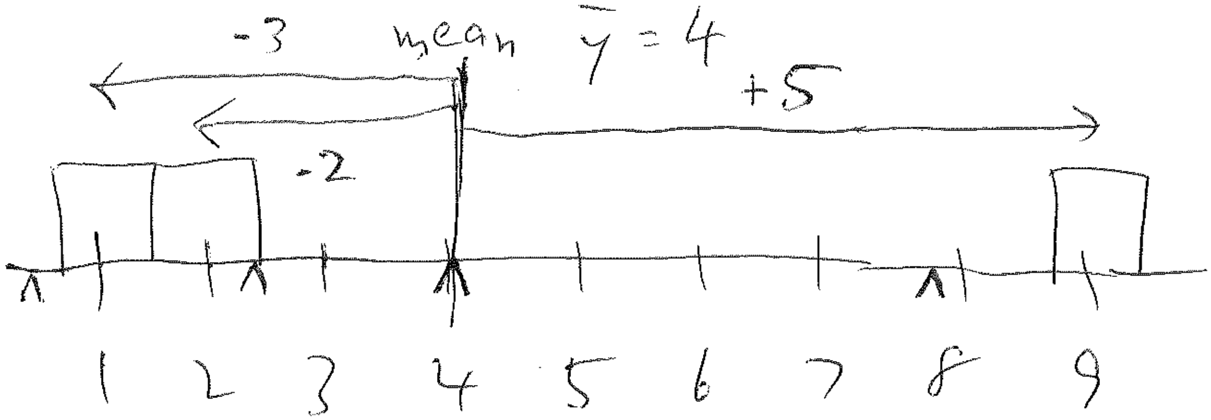
$$\bar{Y} = \frac{95.0}{24} \doteq 3.96 \text{ cm}$$

graphical interpretation of the mean

y

$$\begin{matrix} y_1 \\ y_2 \\ y_n \end{matrix} \begin{bmatrix} 1 \\ 2 \\ 9 \end{bmatrix} \quad n = 3$$

fake data set



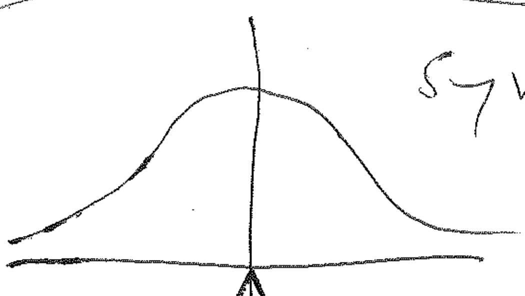
mean = balance point

$$\begin{bmatrix}1\\2\\9\end{bmatrix} = \begin{bmatrix}y_1\\y_2\\\vdots\\y_n\end{bmatrix} \xrightarrow[\bar{y}=4]{\text{subtract}} \begin{pmatrix}1-4=-3\\2-4=-2\\9-4=+5\end{pmatrix}_{n=3}$$

mean 0

$$= \begin{pmatrix}y_1-\bar{y}\\y_2-\bar{y}\\\vdots\\y_n-\bar{y}\end{pmatrix}$$

theorem: let n be any positive integer & let $y_1, \ldots, y_n$ be any real numbers; then $\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y}) = 0$



symmetric & unimodal

point of symmetry = mode = mean =