AMS7
24 June
2016

this time: graphical & numerical descriptive methods; the normal curve

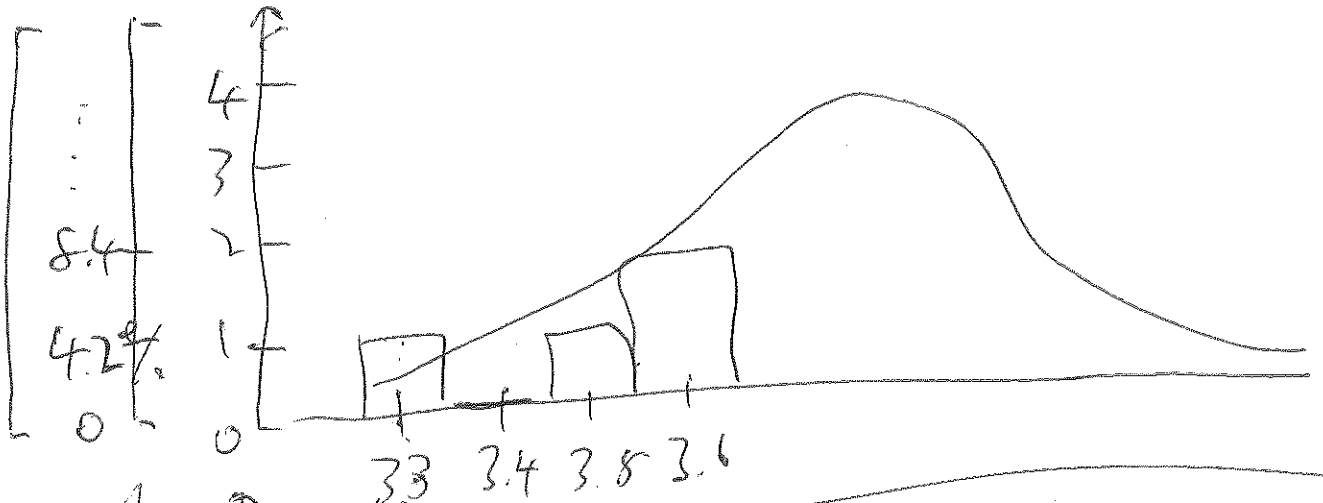
next time: experimental design

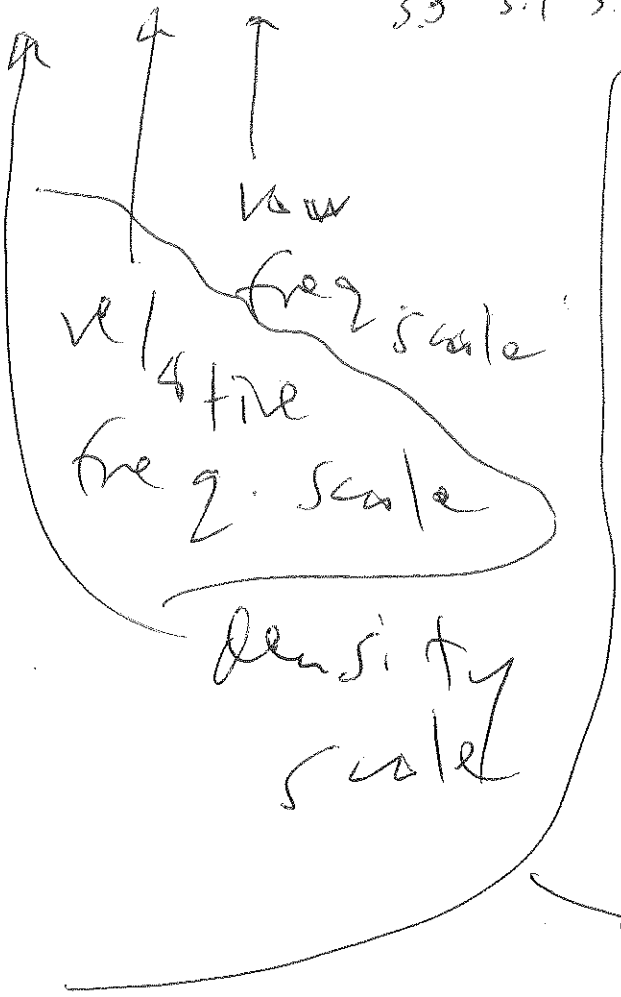
hwk 1 due Wed 29 Jun in class

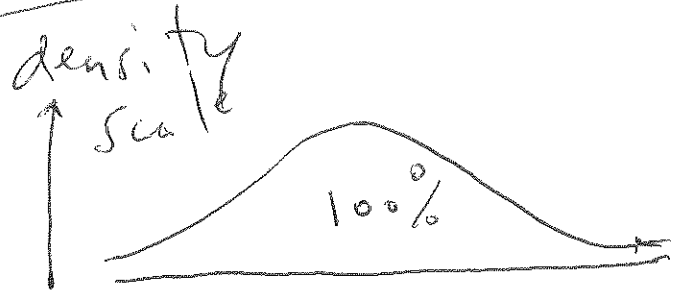
reqd: DD (O) 1-5
DD (N) 1-3
LN pp. 1-94

making class Tue 5 July 5-7.30 pm location ?

value	raw freq.	relative frequency (%)
3.3	1	$(1/24) \cdot 100\% = 4.2\%$
3.4	0	0
3.5	1	4.2
3.6	2	8.4
⋮	⋮	⋮
4.5	1	4.2
	sum $n = 24$	sum 100.1 = roundoff error





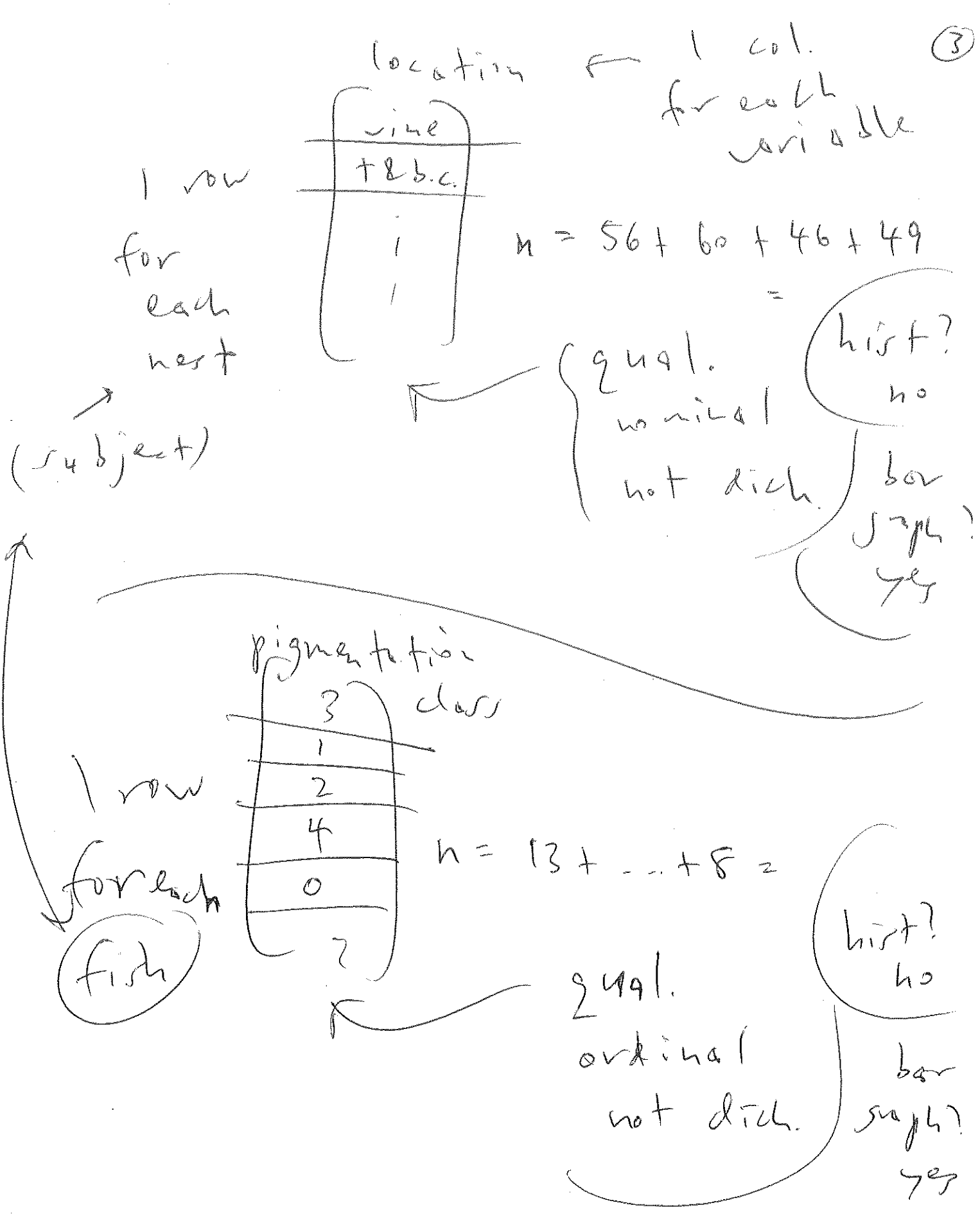


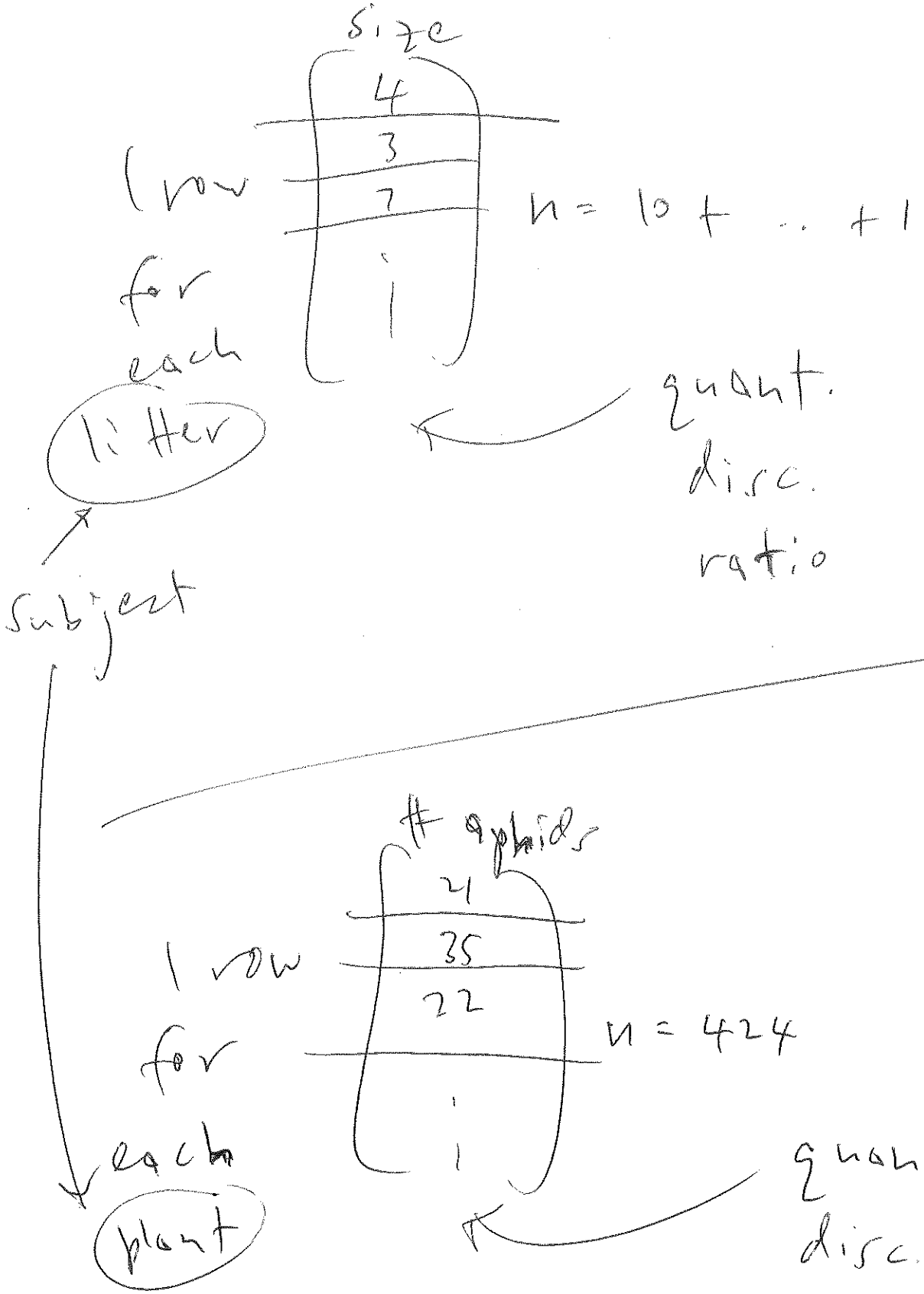
on density scale,

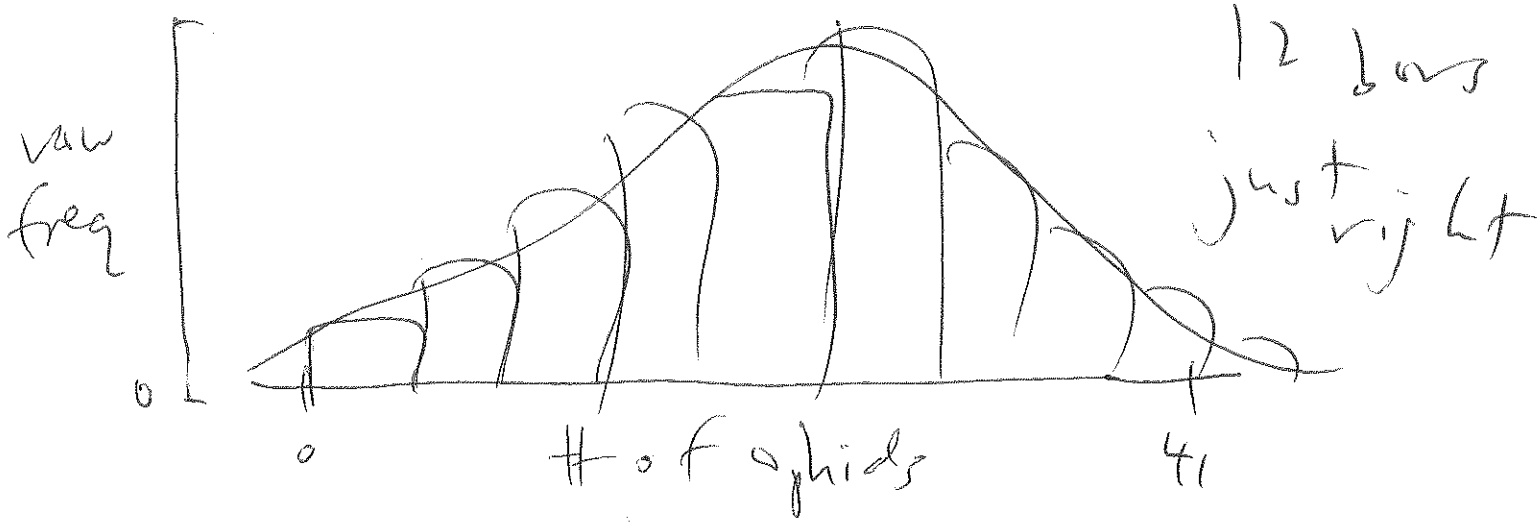
relative frequency = $\frac{\text{area}}{\text{under histogram}}$

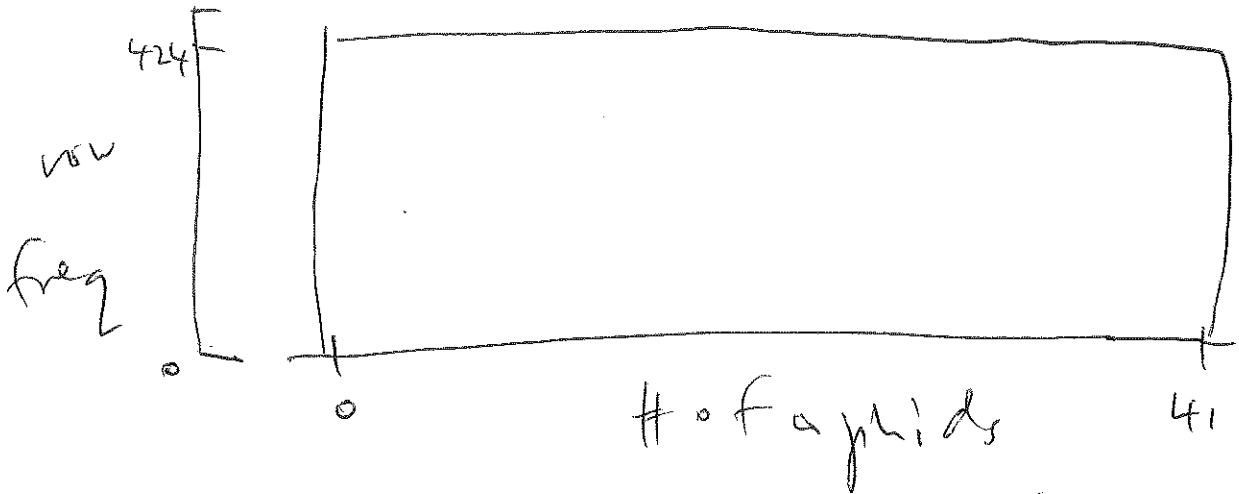
= area under curve

convention: all hist. sketches from now on are on density scale

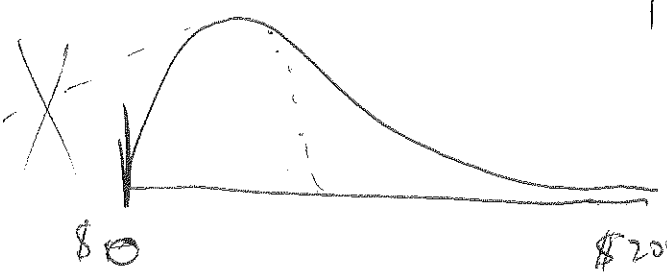

location
1 col. for each variable
vine
t&b.c.
1 row for each next (subject)
1
1
n = 56 + 60 + 46 + 49 =
equal. nominal not dich.
hist? no
bar graph? yes
pigmentation class
3
1
2
4
0
7
1 row foreach fish
n = 13 + ... + 8 =
equal. ordinal not dich.
hist? no
bar graph? yes


size
4
3
2
1
row
for
each
litter
n = 10 + .. + 1
quant.
disc.
ratio
hist?
yes
subject
xplids
4
35
22
1
row
for
each
plant
n = 424
quant.
disc.
ratio
hist?
yes





(can't see shape at all)

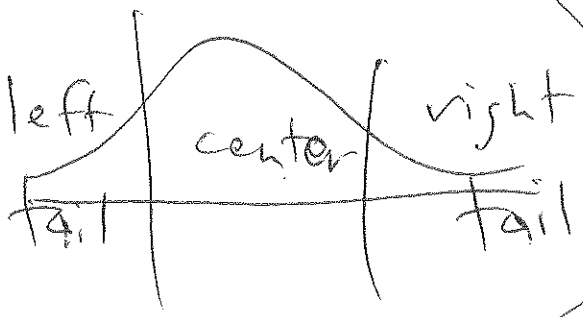


yearly family income in U.S.

right-skewed
(positive skew)

long right-hand tail

shape

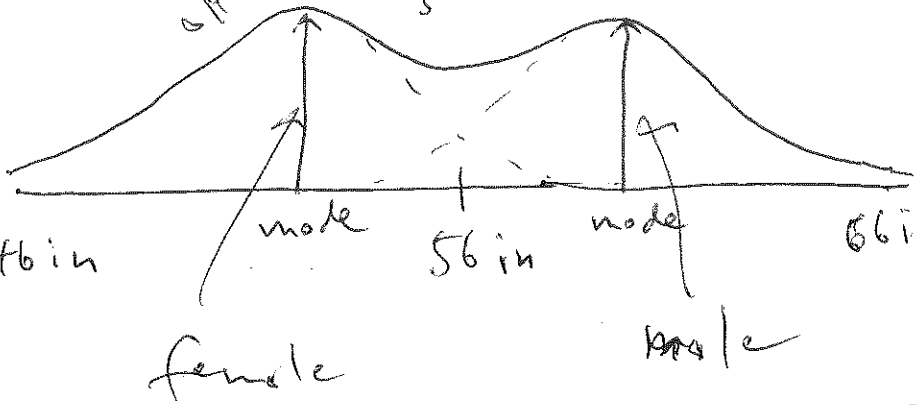


midterm scores

long left hand tail

(negative skew)

height of UCSC students



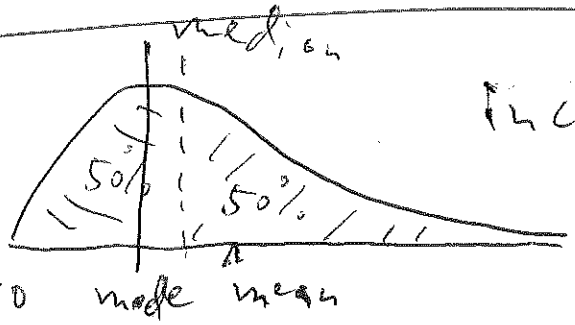
female

male

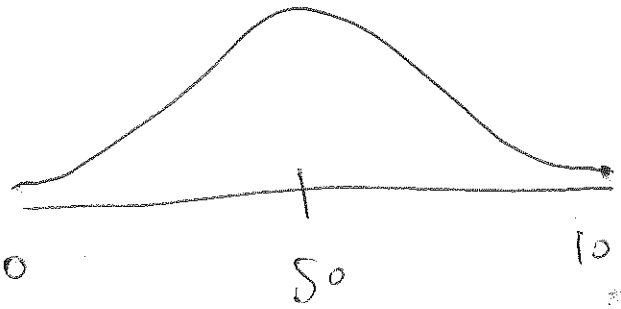
symmetric bimodal

(multi-modal)

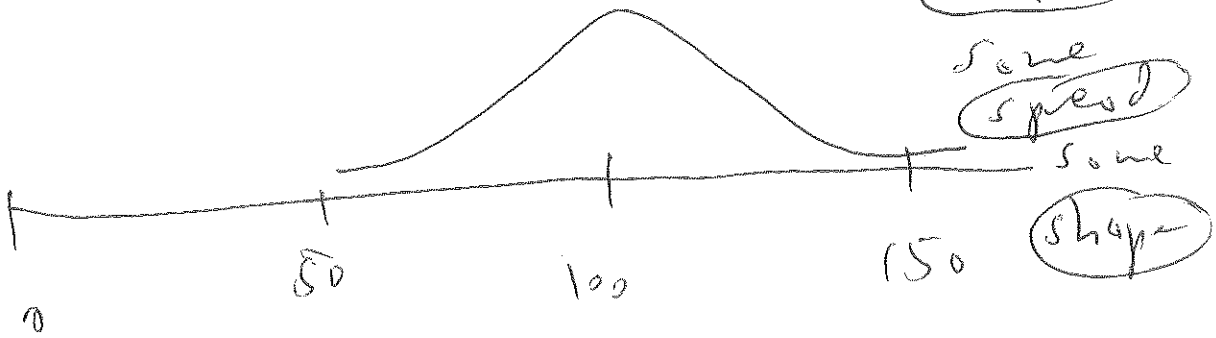
mode = point of highest frequency



income

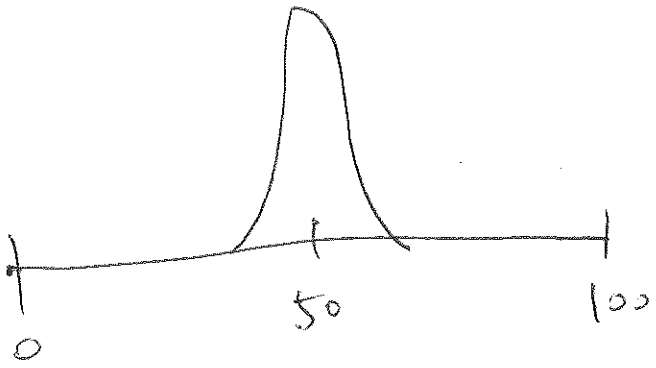


different center

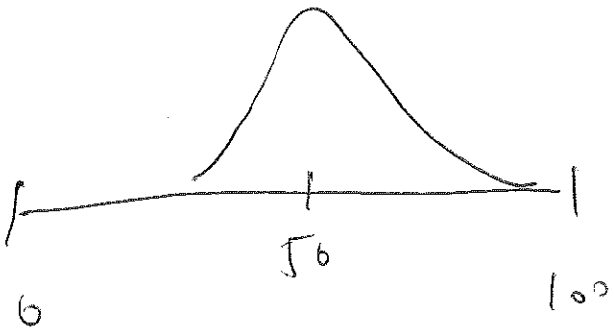


some spread

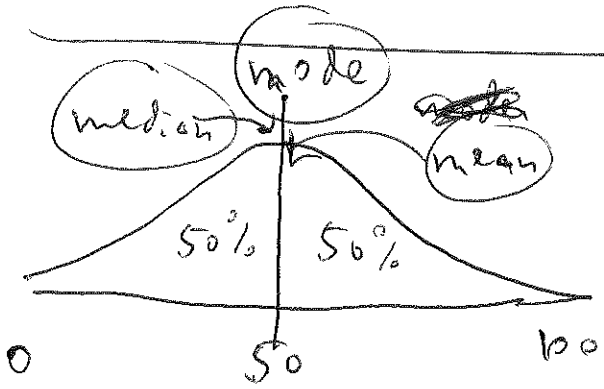
some shape



some center
different spread
some shape



some center
some spread
different shape

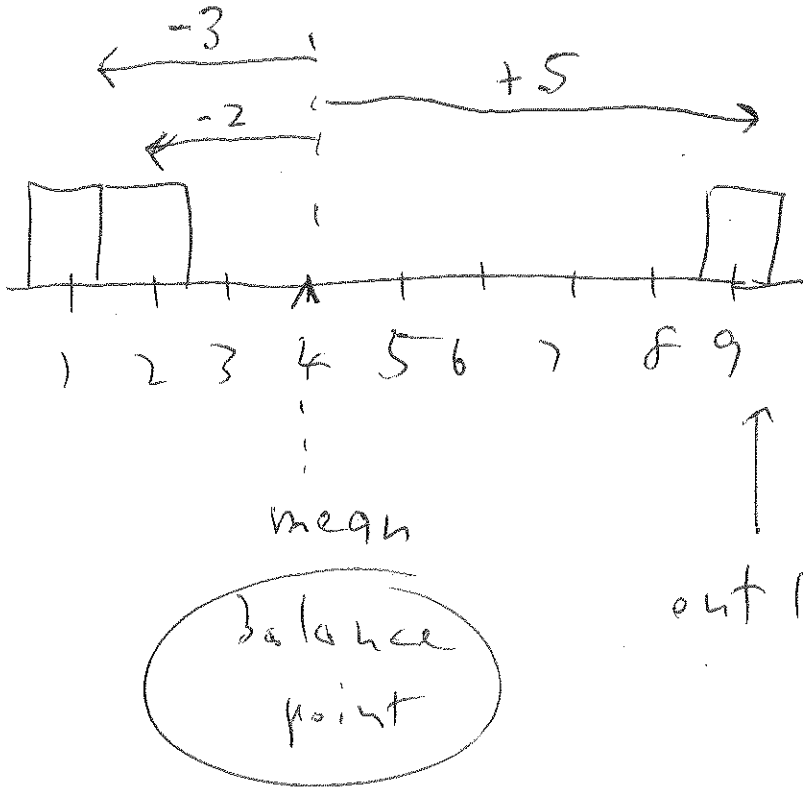


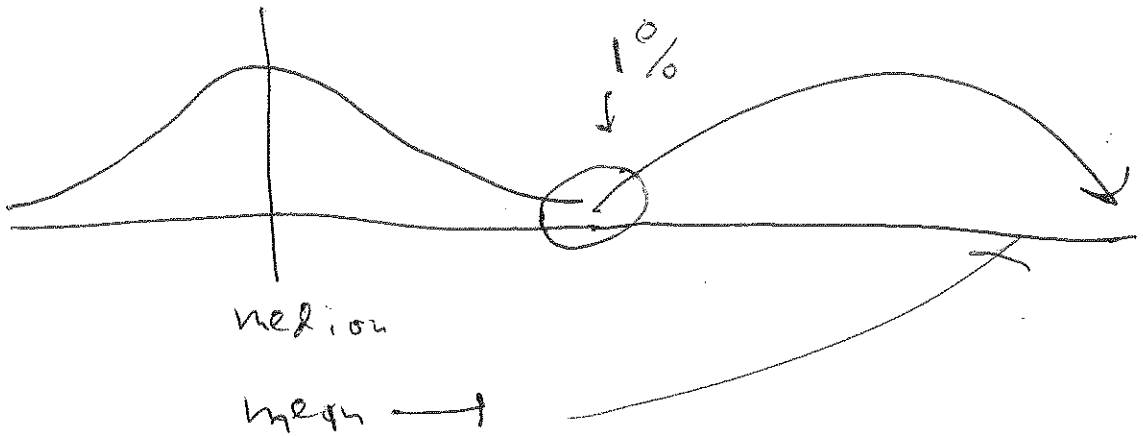
symmetric
unimodal

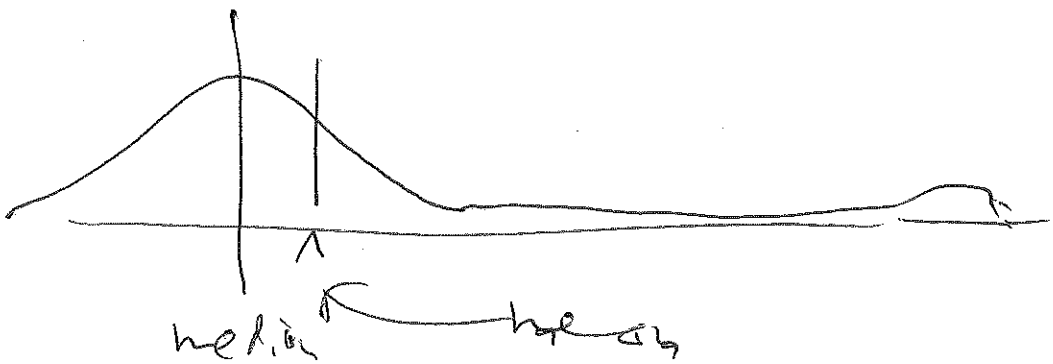
← point of symmetry

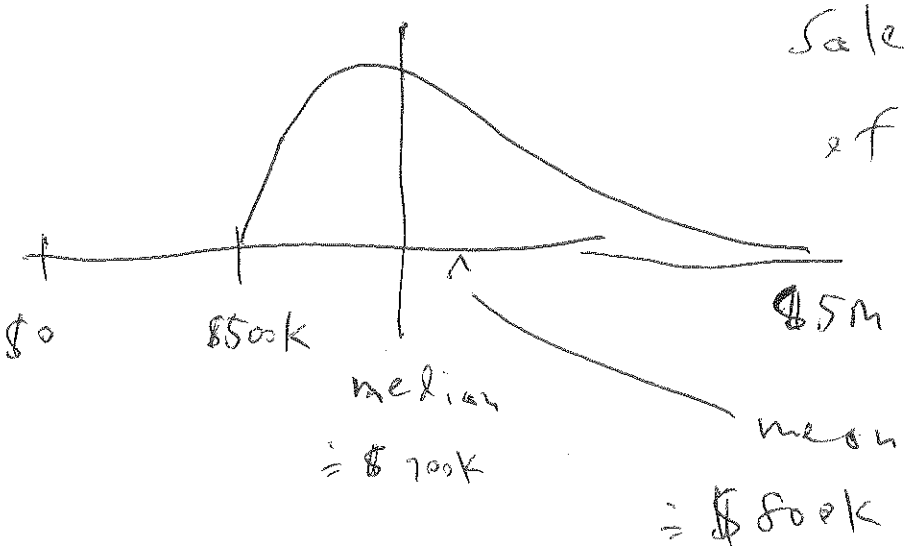
$$\begin{matrix} y_1 \\ y_2 \\ y_n \end{matrix}\begin{pmatrix} 1 \\ 2 \\ 9 \end{pmatrix} \quad n = 3$$

mean $\bar{y} = 4$



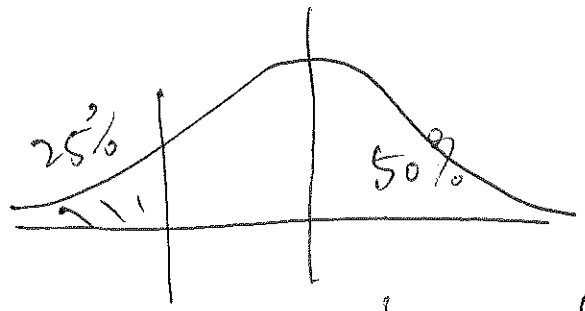






$$\begin{bmatrix} 9 \\ 1 \\ 2 \end{bmatrix} \xrightarrow{\text{sort}} \begin{bmatrix} 1 \\ \textcircled{2} \\ 9 \end{bmatrix} \quad \text{median } \tilde{y}, \; n=3$$

$$\begin{bmatrix} 9 \\ 1 \\ 3 \\ 2 \end{bmatrix} \xrightarrow{\text{sort}} \begin{bmatrix} 1 \\ \textcircled{2} \\ \textcircled{3} \\ 9 \end{bmatrix} \; n=4 \qquad \frac{2+3}{2} = 2.5 = \text{median}$$



median = 50th percentile

25th percentile = 25th quantile

$\begin{bmatrix} 1 \\ 2 \\ 9 \end{bmatrix}$ $n=3$ → subtract mean → $\begin{bmatrix} -3 \\ -2 \\ +5 \end{bmatrix}$ → absolute value → $\begin{bmatrix} 3 \\ 2 \\ 5 \end{bmatrix}$

mean 4 mean 0 mean $\frac{10}{3} = 3.3$

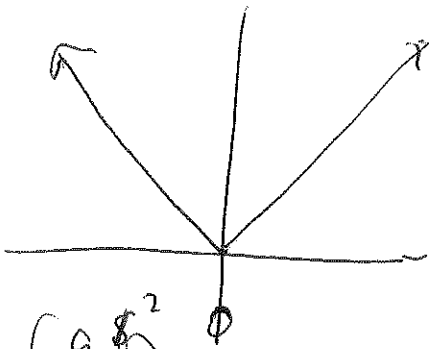
$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$ n → subtract mean → $\begin{bmatrix} y_1 - \bar{y} \\ \vdots \\ y_n - \bar{y} \end{bmatrix}$ → abs. values → $\begin{bmatrix} |y_1 - \bar{y}| \\ \vdots \\ |y_n - \bar{y}| \end{bmatrix}$

mean $\bar{y}$ mean 0 mean

deviations from the mean

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

mean absolute deviation



$\begin{bmatrix} \$_1 \\ \$_2 \\ \$_9 \end{bmatrix}$ → subt. $\bar{y}$ → $\begin{bmatrix} \$_1 - \$_2 \\ \$_2 - \$_5 \\ +\$_5 \end{bmatrix}$ → square → $\begin{bmatrix} 9\ \$^2 \\ 4\ \$^2 \\ 25\ \$^2 \end{bmatrix}$

mean $\frac{38}{3} = 12.7\ \$^2$

$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \rightarrow \begin{bmatrix} y_1 - \bar{y} \\ \vdots \\ y_n - \bar{y} \end{bmatrix} \rightarrow \begin{bmatrix} (y_1 - \bar{y})^2 \\ \vdots \\ (y_n - \bar{y})^2 \end{bmatrix}$ ⊛

mean of squares

$$\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2 = (*)$$

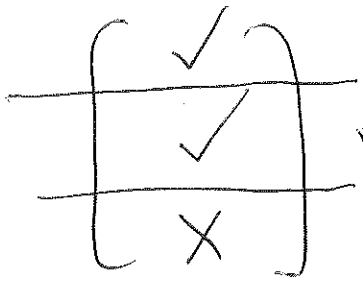
(*) ↓ units wrong

$$\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})^2 = \text{sample variance}$$

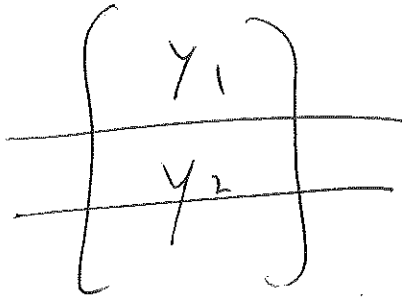
$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})^2} = \text{sample standard deviation}$$

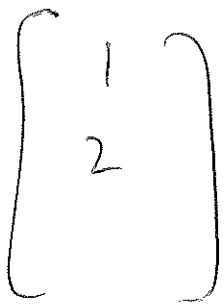
(SD)

here $\sqrt{\frac{9+4+25}{2}} = \sqrt{19} = 4.4$



$n = 3$





mean (4) mean 4 mean 4

$(n-1)$ degrees of freedom for learning about spread from data set with n values