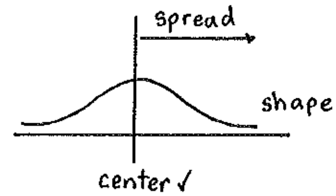
This time: measure of center & spread, normal curve

Next time: experimental design

Read: DD Ⓐ Ch. 1-3 Ⓑ Ch. 1-6 LN pp L-1-85

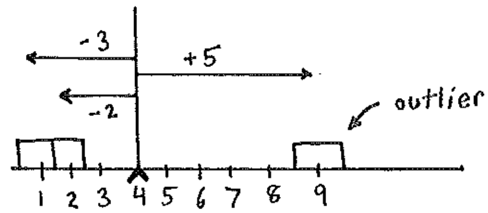
Can turn in homework 1 on Fri. or over weekend in box outside Baskin.



graphical interpretation of mean:

$$\begin{array}{c} y_1 \\ \vdots \\ y_n \end{array} \begin{bmatrix} 1 \\ 2 \\ 9 \end{bmatrix} n=3 \xrightarrow[\;4\;]{\text{subtract}} \begin{bmatrix} -3 \\ -2 \\ +5 \end{bmatrix}$$

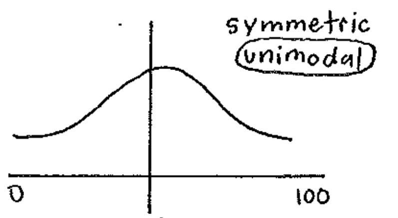
mean $\bar{y} = 4$ mean $\bar{y} = 0$



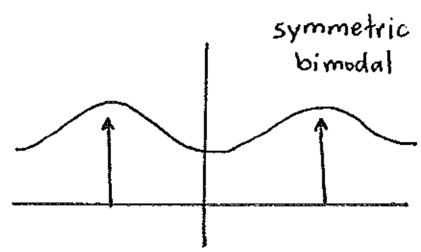
$$\begin{bmatrix} y_1 \\ \cdot \\ \cdot \\ y_n \end{bmatrix} \xrightarrow[\;\bar{y}\;]{\text{subtract}} \begin{bmatrix} y_1 - \bar{y} \\ y_2 - \bar{y} \\ \vdots \\ y_n - \bar{y} \end{bmatrix}$$ — deviations from mean

mean $\bar{y}$ mean 0 ✓

mean = balance point

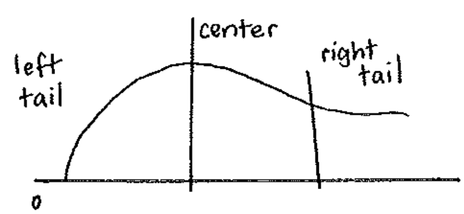


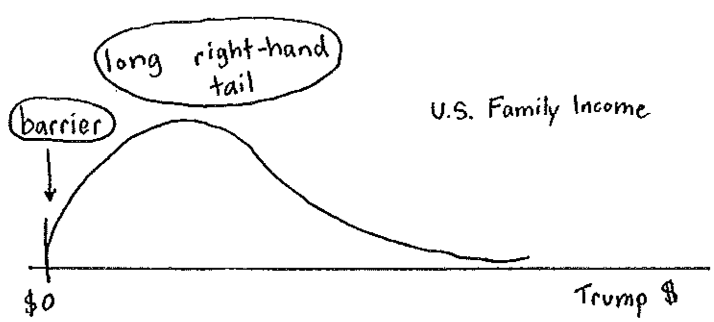
point of symmetry
= mean if hist. is symmetric

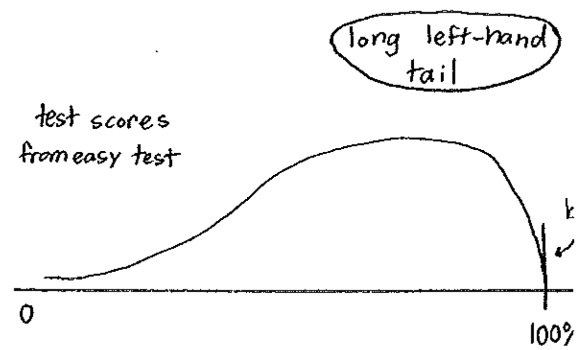


ex of multimodal

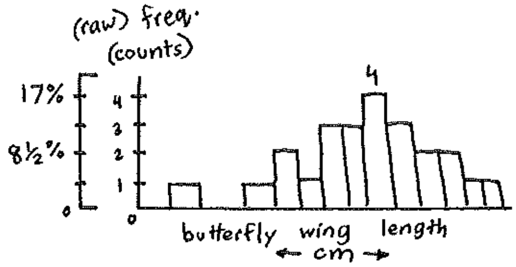
mode = highest poin
on hist.
(in freq. terms







skewed = asymetric

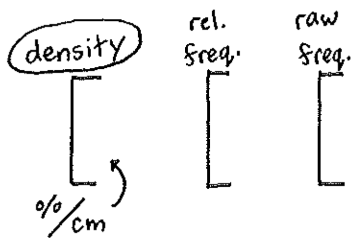


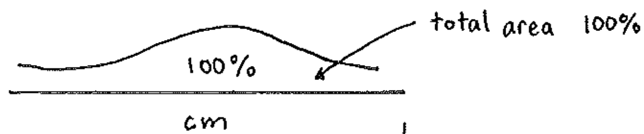
Q: What relative freq. is in highest?

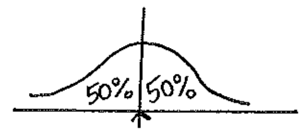
A: $\frac{4}{n=24} = \frac{1}{6} \cdot 100\% = 17\%$ (16.67...)

Rule: when hist. are drawn on density scale, relative frequency (%) corresponds (≡) exactly to area under hist.

Convention: all histograms in this class from now on will be on the density scale







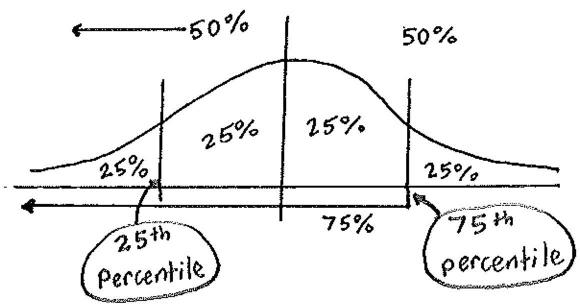
∧ = pt. of sym.
= mean
= mode
= median

median = 50/50 pt. in the histogram in rel. freq. terms.

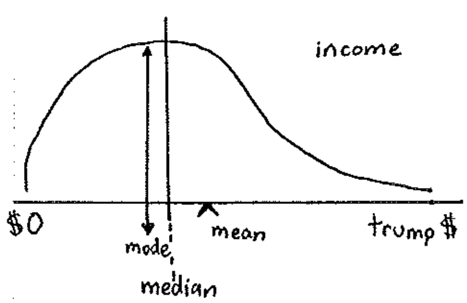
$$\begin{bmatrix} 1 \\ 9 \\ 2 \end{bmatrix} \xrightarrow{\text{sort}} \begin{bmatrix} 1 \\ 2 \\ 9 \end{bmatrix} \text{median}$$

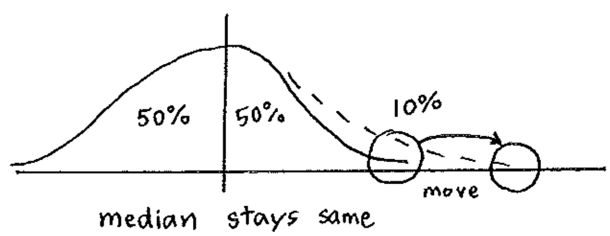
$$\begin{bmatrix} 9 \\ 1 \\ 3 \\ 2 \end{bmatrix} \xrightarrow{\text{sort}} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 9 \end{bmatrix} \text{mean of } \frac{2+3}{2} = 2.5$$

median = 2.5

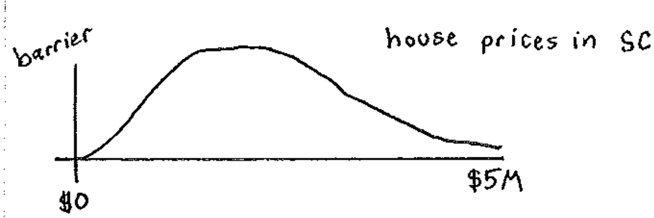


medi..
=
50th
perce...





median stays same
mean goes up



idea 1:

$$\begin{bmatrix}1\\2\\9\end{bmatrix} \xrightarrow[4]{\text{subtract}} \begin{bmatrix}-3\\-2\\+5\end{bmatrix} \xrightarrow[|\cdot|]{\text{take}} \begin{bmatrix}3\\2\\5\end{bmatrix}$$

absolute value

mean $\frac{10}{3} \doteq 3.3$ ← (MAD) not used much

mean absolute deviation (from the mean)

idea 2: square

$$\begin{bmatrix}\$1\\\$2\\\$9\end{bmatrix} \xrightarrow[\$4]{\text{subtract}} \begin{bmatrix}-3\\-2\\+5\end{bmatrix} \xrightarrow{\text{square}} \begin{array}{l}(-3)^2 = +9\,\$^2\\(-2)^2 = +4\,\$^2\\(+5)^2 = +25\,\$^2\end{array}$$

mean $\bar{y} = \$4$

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

Final step: take $\sqrt{}$

$\sqrt{\$^2\, 12.7} = \$\,3.6$

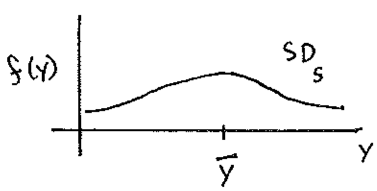
$$\begin{bmatrix}y_1\\\vdots\\y_n\end{bmatrix} \xrightarrow[\bar{y}]{\text{subtract}} \begin{bmatrix}y_1-\bar{y}\\\vdots\\y_n-\bar{y}\end{bmatrix} \xrightarrow{\text{square}} \begin{bmatrix}(y_1-\bar{y})^2\\\vdots\\(y_n-\bar{y})^2\end{bmatrix}$$

mean $= \bar{y}$

$\sqrt{\text{mean}}$

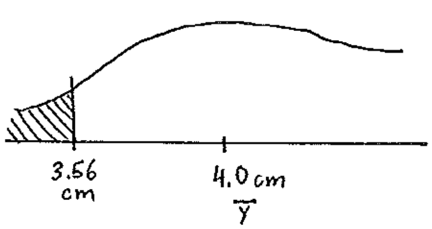
$$\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})^2} \rightarrow \overset{\text{def}}{=} \text{(sample) standard deviation (SD)} = s$$

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

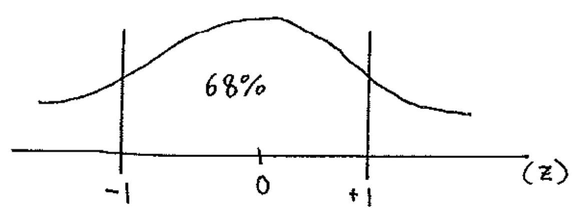


$$f(y) = \frac{1}{s\sqrt{2\pi}} \exp\left[-\frac{(y-\bar{y})^2}{2s^2}\right]$$

SD $s = 0.29$ cm



standard normal curve



SD 1 $\quad 0.1587 = 16\%$

fact: all normal curves satisfy empirical rule <u>exactly</u>