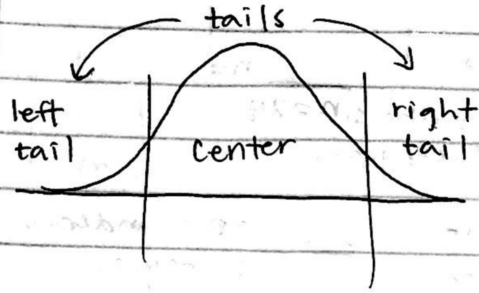
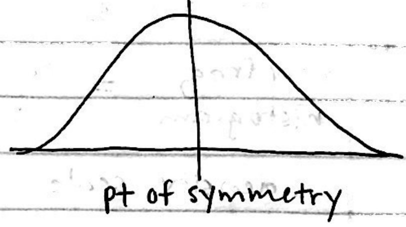
AMS 7
12 Apr 18

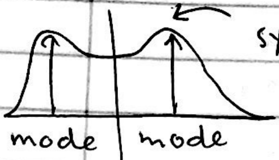
this time: center & spread ; normal curve

next time: experimental design

LN p. 1-94

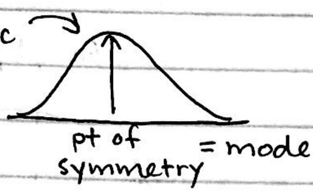






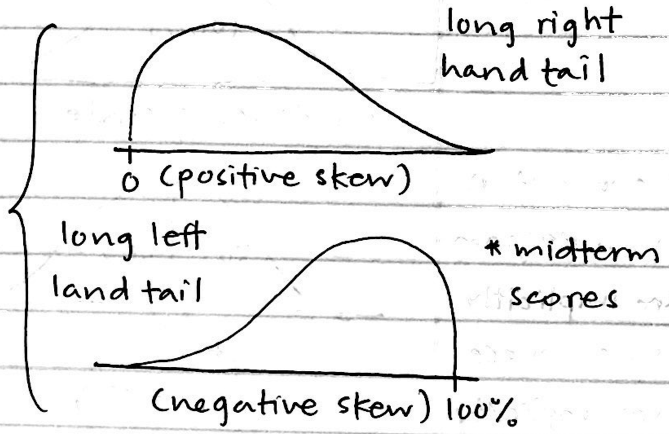
bimodal

multimodal



unimodal

skewed (asymmetric)

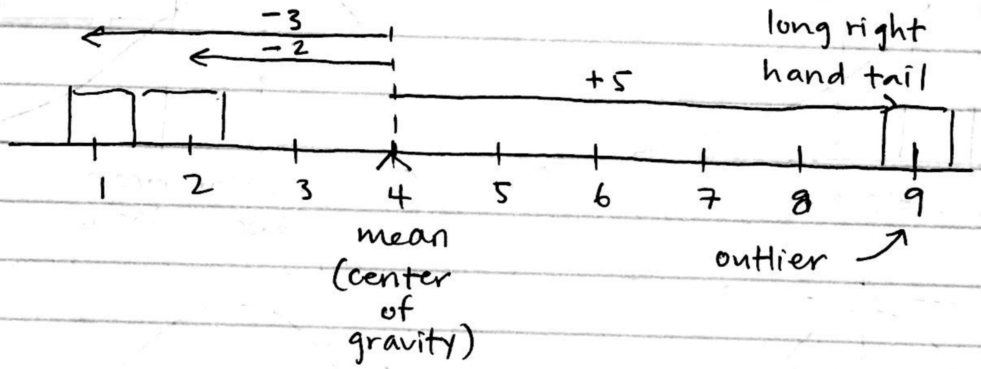


Numerical Measures of Center

1) mode when unimodal

2) mean

$$\bar{Y} = \frac{Y_1 + \dots Y_n}{n} = \frac{1}{n}(Y_1 + Y_2 + \dots + Y_n) = \frac{1}{n}\sum_{i=1}^{n} Y_i$$



$$\begin{array}{c} Y \\ \begin{bmatrix} 1 \\ 2 \\ 9 \end{bmatrix} \end{array} \quad n = 3$$

$\bar{Y} = 4$

$$\begin{bmatrix} 1 \\ 2 \\ 9 \end{bmatrix} \xrightarrow[4]{\text{subtract}} \begin{bmatrix} -3 \\ -2 \\ +5 \end{bmatrix} \qquad \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} \xrightarrow[\bar{Y}]{\text{subtract}} \begin{bmatrix} Y_1 - \bar{Y} \\ Y_2 - \bar{Y} \\ \vdots \\ Y_n - \bar{Y} \end{bmatrix}$$

deviations (from the mean)

3) median

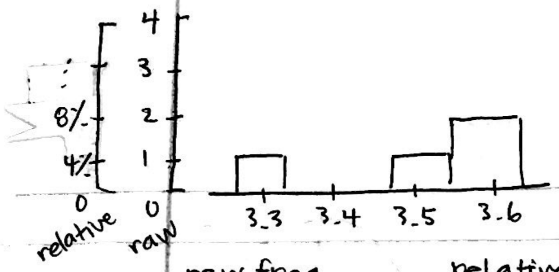
50/50 point in relative freq

$$\xrightarrow{\text{sort}} \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} \qquad \begin{bmatrix} 9 \\ 1 \\ 2 \end{bmatrix} \xrightarrow{\text{sort}} \begin{bmatrix} 1 \\ 2 \\ 9 \end{bmatrix}$$

median = center data value, after sorting from smallest to largest

$$n = 4 \quad \begin{bmatrix} 9 \\ 1 \\ 4 \\ 2 \end{bmatrix} \rightarrow \begin{bmatrix} 1 \\ 2 \\ 4 \\ 9 \end{bmatrix} \rightarrow$$

median = mean of 2 middle #s



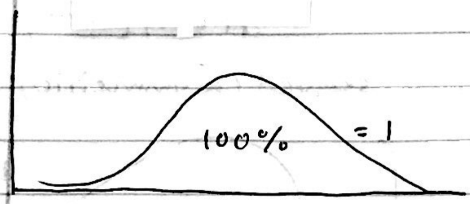
value	raw freq	relative freq
3.3	1	$(1/24) \cdot 100\% = 4\%$
3.4	0	$(0/24) \cdot 100\% = 0\%$
3.5	1	$(1/24) \cdot 100\% = 4\%$
⋮		
4.5	1	

$n = 24$

raw freq histogram = relative freq histogram = density scale histogram

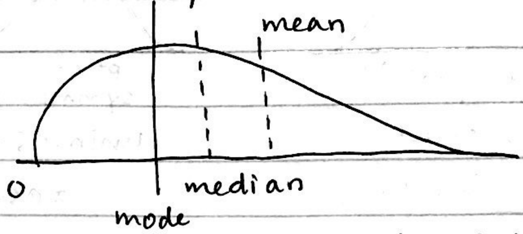
relative freq = area under histogram

density scale



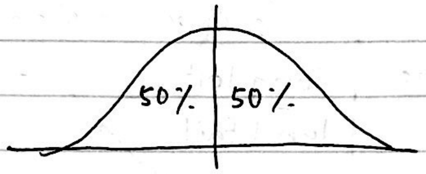
on density scale

family income

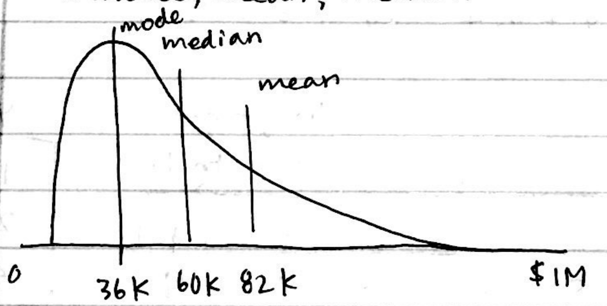


* same median, mean shifted right

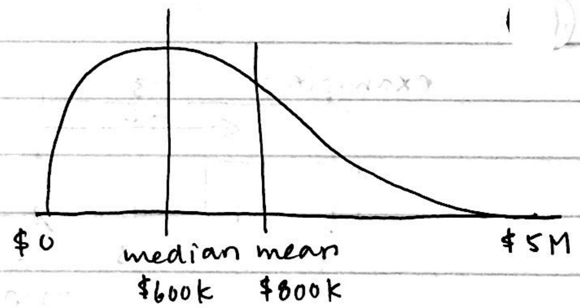
* convention: all histogram sketches in this course are implicitly drawn on density scale

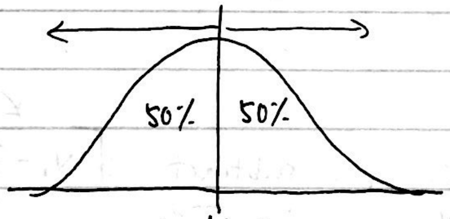


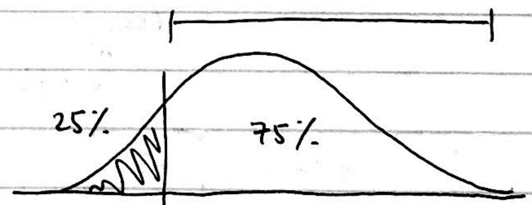
pt of symmetry = mode, mean, median



SC home sale prices

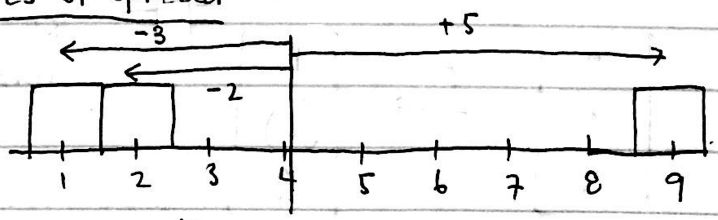






Measures of Spread

$n = 3$ $\begin{bmatrix} 1 \\ 2 \\ 9 \end{bmatrix}$



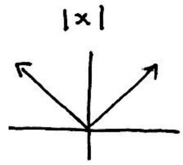
mean = 0

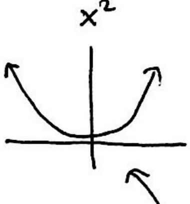
$\begin{bmatrix} 1 \\ 2 \\ 9 \end{bmatrix} \rightarrow \begin{bmatrix} -3 \\ -2 \\ +5 \end{bmatrix}$

$\bar{y} = 4$ mean $= \frac{10}{3} = 3.3$

$\begin{bmatrix} -3 \\ -2 \\ +5 \end{bmatrix}$ $\xrightarrow{\text{take absolute values}}$ $\begin{bmatrix} +3 \\ +2 \\ +5 \end{bmatrix}$ $\begin{bmatrix} y_1 - \bar{y} \\ \vdots \\ y_n - \bar{y} \end{bmatrix} \rightarrow \begin{bmatrix} |y_1 - \bar{y}| \\ \vdots \\ |y_n - \bar{y}| \end{bmatrix}$

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





$\begin{bmatrix}-3\\-2\\+5\end{bmatrix} \xrightarrow{\text{square}} \begin{bmatrix}(-3)^2=9\\(-2)^2=4\\(+5)^2=25\end{bmatrix}$ mean $=\frac{38}{3}=12.7$

$\begin{bmatrix}y_1\\ \vdots\\ y_n\end{bmatrix} \xrightarrow[\bar{y}]{\text{subtract}} \begin{bmatrix}y_1-\bar{y}\\ \vdots\\ y_n-\bar{y}\end{bmatrix} \xrightarrow{(\,)^2} \begin{bmatrix}(y_1-\bar{y})^2\\ \vdots\\ (y_n-\bar{y})^2\end{bmatrix}$ mean $=\frac{1}{n}\sum_{i=1}^{n}(y_i-y)^2$

matches bell curve

intuitive: $\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2$ better: $\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})^2=s^2$ $\begin{bmatrix}+9\\+4\\+25\end{bmatrix}$ sum = 38

variance $=\frac{38}{2}=19$

sample variance

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

sample standard deviation (SD)

$\begin{bmatrix}y_1\\y_2\\y_3\end{bmatrix}$ $\bar{y}=4$ $\begin{bmatrix}1\\2\\9\end{bmatrix} \rightarrow$ $\bar{y}=0$ $\begin{bmatrix}-3\\-2\\+5\end{bmatrix}$ $\bar{y}=4$ $\begin{bmatrix}\checkmark\\ \checkmark\\ \ \end{bmatrix}$ free to vary

mean $\bar{y}$

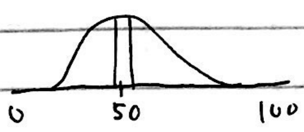
constrain mean to

* a data set w/ n observations actually only has (n-1) independent <u>pieces of information about spread</u>

degrees of freedom

<u>Empirical Rule</u>

↳ for almost all data sets, if you start at the mean and go $\begin{Bmatrix}1\text{ SD}\\2\text{ SD}\\3\text{ SD}\end{Bmatrix}$ either way, you will usually encompass about $\begin{Bmatrix}2/3\\\text{most}\end{Bmatrix}=\begin{Bmatrix}68\%\\95\%\end{Bmatrix}$ of the data



$\boxed{SD=1}$

50 ± 1 ⟷ 49 51

too small

$\boxed{SD=50}$ → too big

$\boxed{SD=10}$ → too small

$\boxed{SD=20}$ → about right

bell curve = normal curve = Gaussion

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