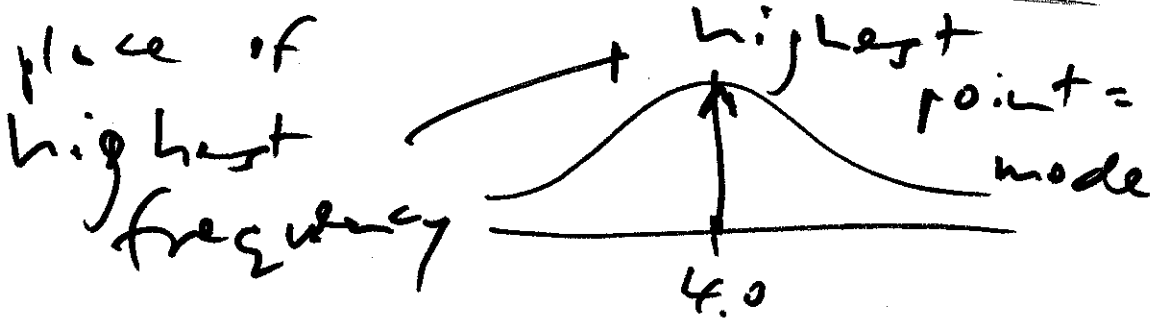
AMS 7
20 Oct 08

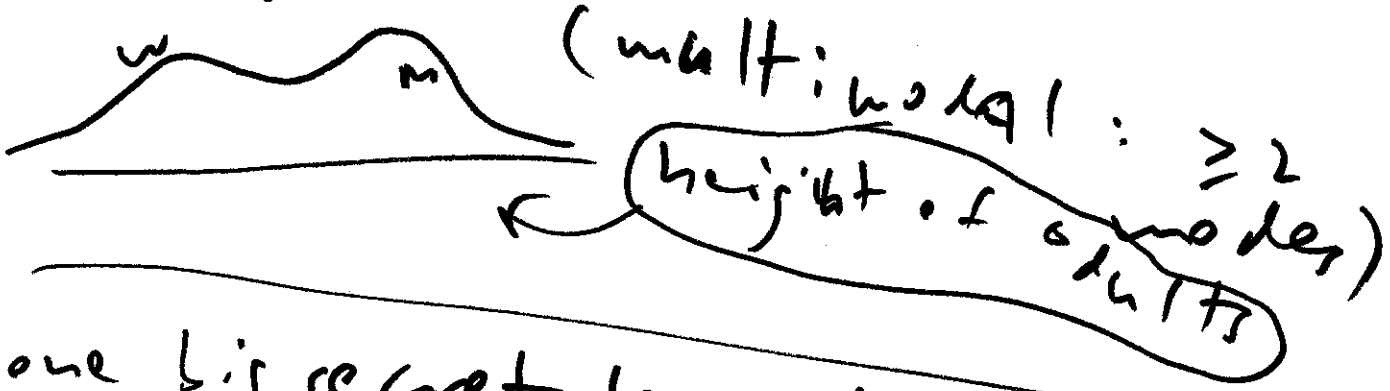
this time: histograms, measures of center & spread;

next time: using normal curve descriptively; controlled experiments

read: T&T ch. 1, 2, 5

labs start next week



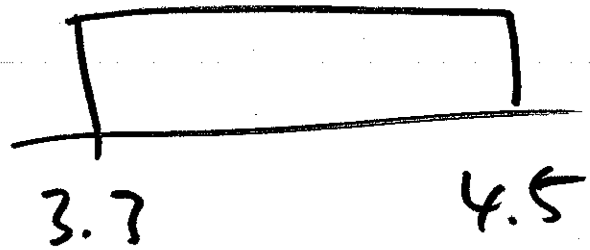


one big secret to good statistical work: visualize the raw data or a table with rows for subjects & columns for variables

many possible histograms for same dataset, depending on how wide bins are & where they start

---

bins? too few - lose info about shape



bins too many: can't see shape well, too noisy

~~symmetric unimodal~~

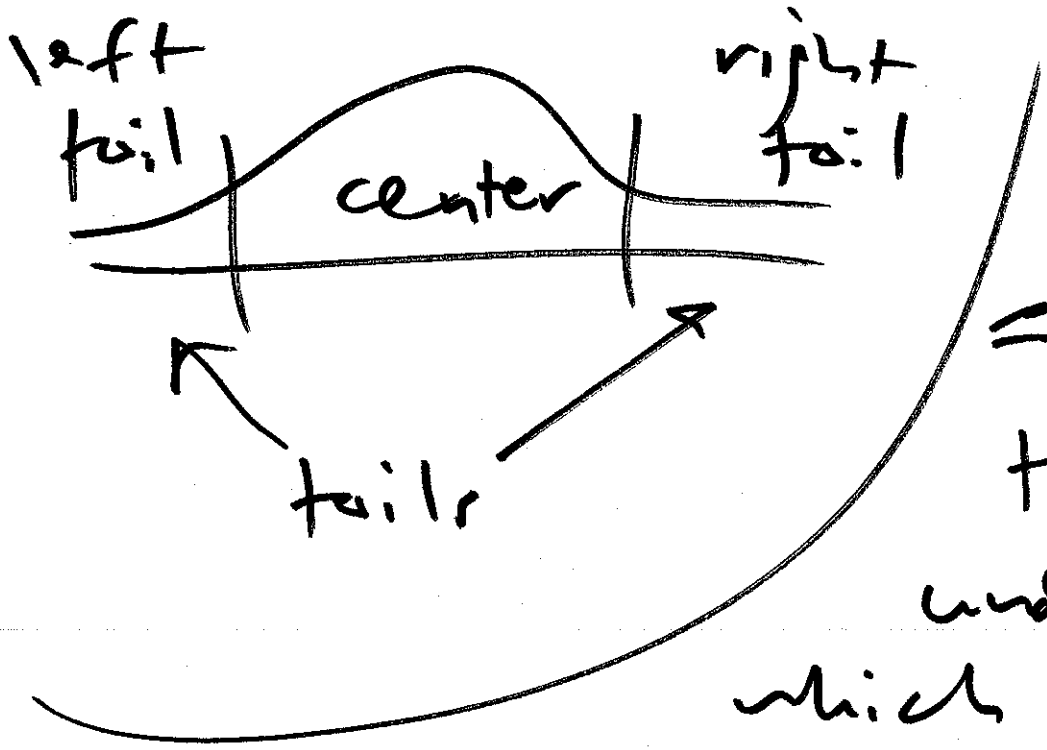
butterfly wing length

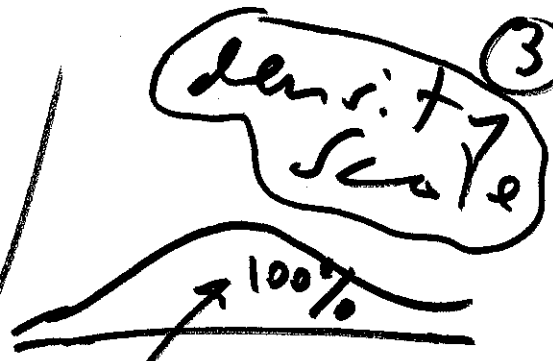
long right hand tail

fox litter size

long left hand tail

asymmetric = skewed

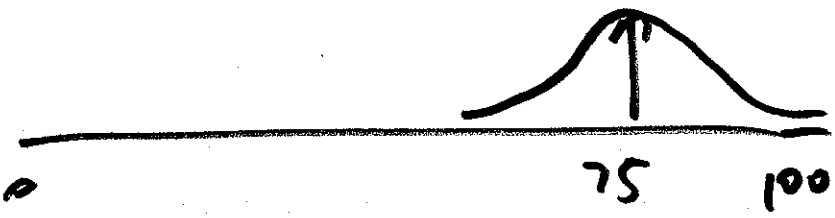


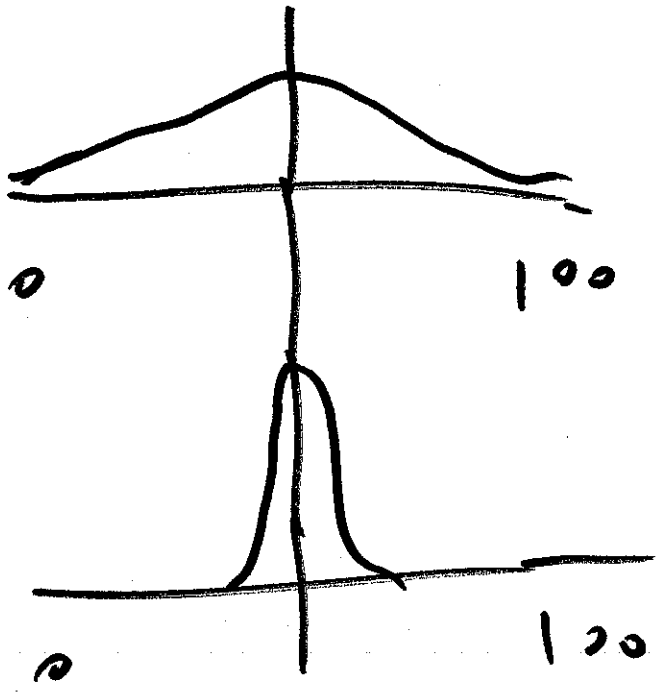


total area under curve which summarizes

hist is 1 = 100%, & relative freq. is characterized by area under curve

same spread same shape different center





same shape
same center
different spread

---

measures of center

① mode

② mean

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

$= \frac{1}{n}(y_1 + \cdots + y_n) =$ mean $\bar{y}$

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

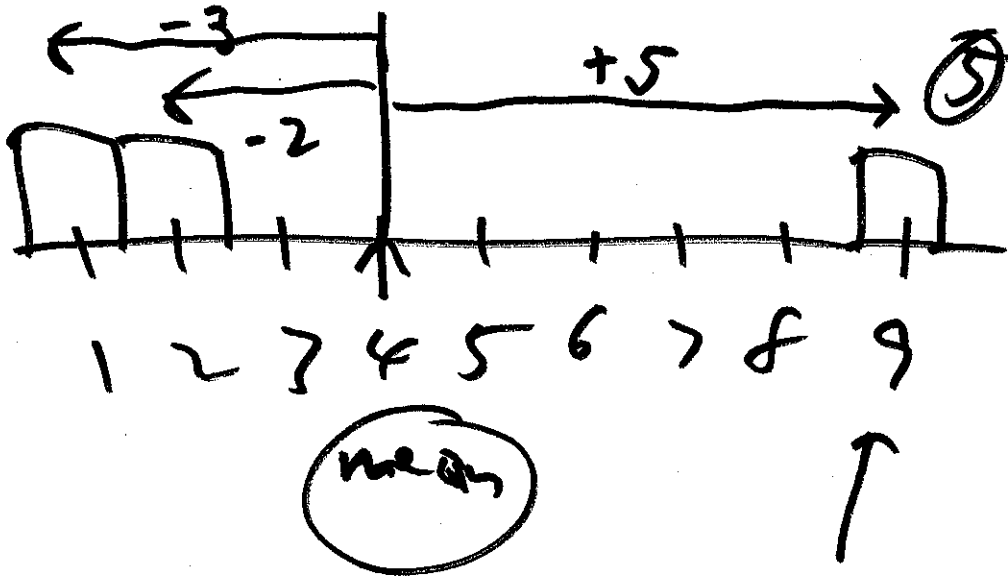
$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad n$$

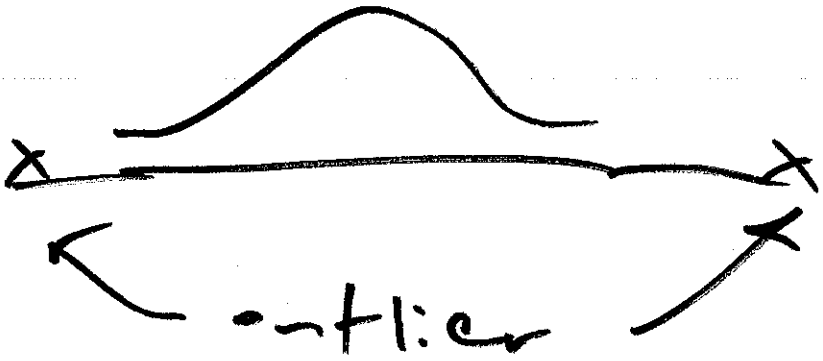
$$\begin{bmatrix} 4.4 \\ \vdots \\ 3.9 \end{bmatrix} \quad n = 24$$

mean $\bar{y} = \dfrac{4.4 + \cdots + 3.9}{24} = 3.96$ cm

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

mean $\bar{y} = 4$

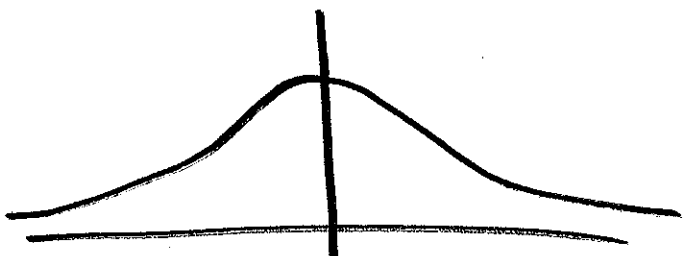




Commonly high or low value

mean = balance point

(number line = plywood, dist = bricks, mean = place where this balances on fulcrum)



pt of symmetry

= mode

= mean = median

median

$\begin{bmatrix} 1 \\ 2 \\ 9 \end{bmatrix}$ median $\quad \begin{bmatrix} 1 \\ 7 \\ 9 \end{bmatrix}$ median

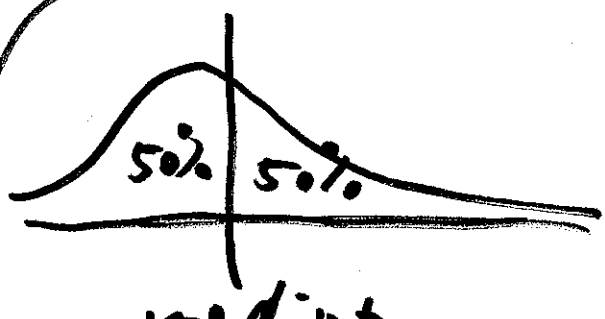
↑

halfway point in terms of relative or raw frequency of data : 50/50 point in frequency

---

median = midpoint of data sorted from smallest to largest

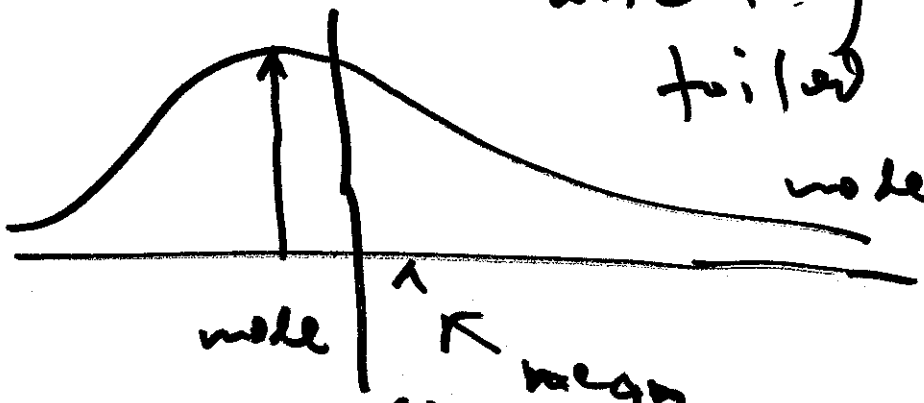
$\begin{bmatrix} 2 \\ 3 \\ 9 \end{bmatrix}$ median 2.5



Skewed right

with long right hand tailed data,

mode < median < mean



* backwards for long left hand tail