this time: spread; normal curve;

next time: experimental design

read: DD ch. 1, 2, 3, 4, 5, 6

AMS 7 15 Jan 08

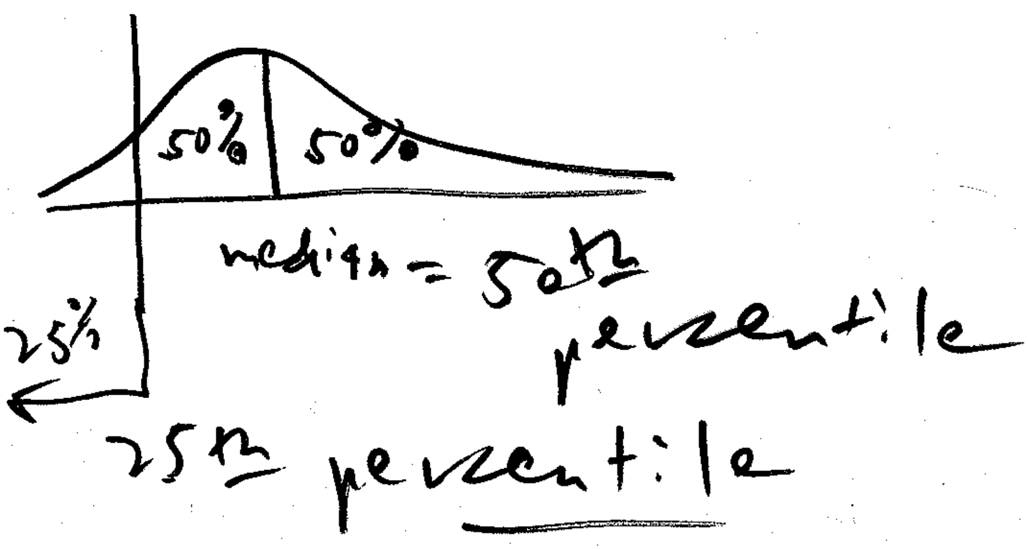
hwk 1 due Tue 20 Jan

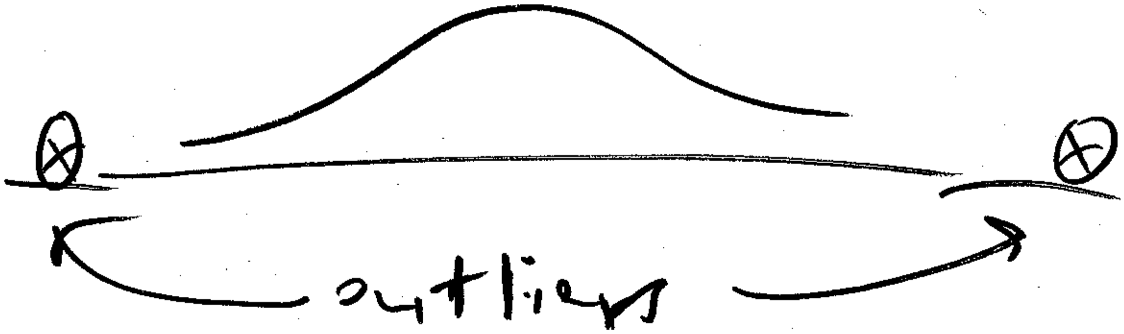
course packet (690 pages; book; lecture notes, reader) available in back of room: $60 cash

labs next week? } yes, except Mon is a holiday

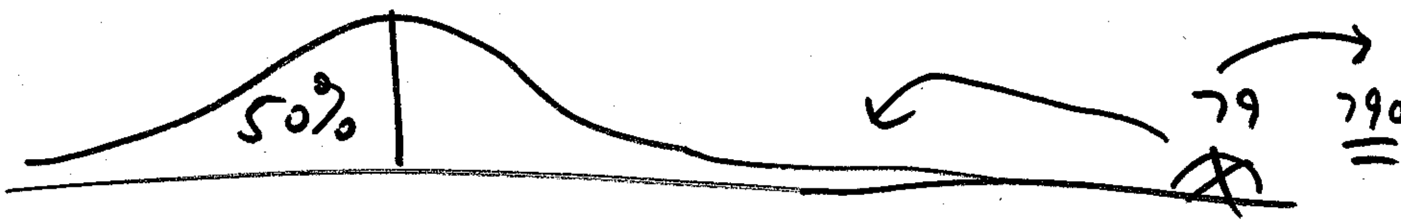
disc. sec next week? }

if you usually go to disc. sec. &/or lab on Mon, next week you have to go to a disc. sec &/or lab Tue - Fri;

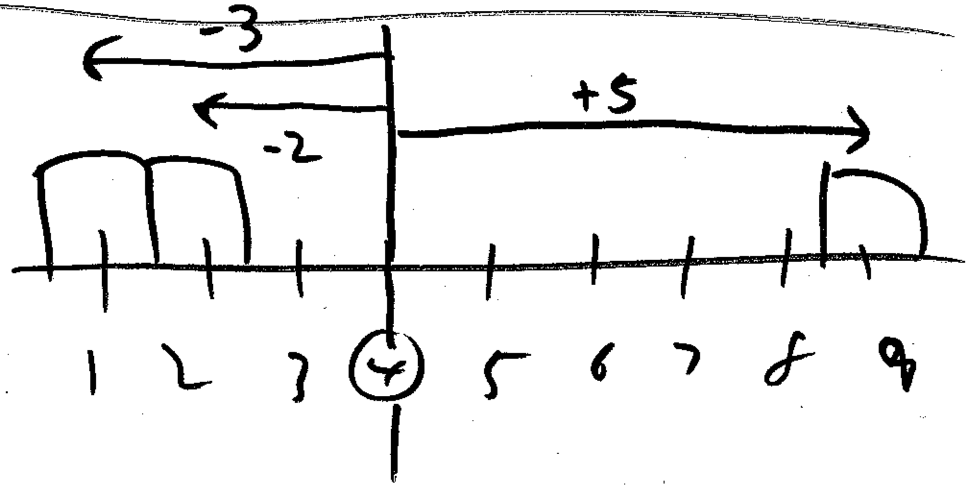




A



$$\begin{pmatrix} \$1 \\ \$2 \\ \$9 \end{pmatrix}$$



mean $\$4$

$$\begin{pmatrix} 1 \\ 2 \\ 9 \end{pmatrix} \xrightarrow[\bar{y}=4]{\text{subtract}} \begin{pmatrix} 1-4=-3 \\ 2-4=-2 \\ 9-4=+5 \end{pmatrix}$$

deviations from mean

mean $\bar{y}=4$

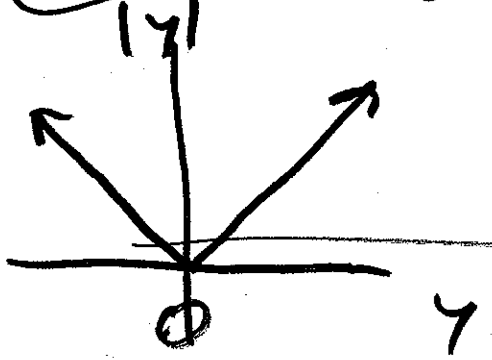
$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \xrightarrow[\bar{y}]{\text{subtract}} \begin{pmatrix} y_1-\bar{y} \\ y_2-\bar{y} \\ \vdots \\ y_n-\bar{y} \end{pmatrix}$$

get rid of cancellation of (+), (−) signs

<u>idea 1</u>: use absolute values

$$\begin{pmatrix} |y_1 - \bar{y}| \\ \vdots \\ |y_n - \bar{y}| \end{pmatrix} = \begin{pmatrix} |-\overset{\$}{3}| = +\overset{\$}{3} \\ |-2| = +\overset{\$}{2} \\ |+5| = +\overset{\$}{5} \end{pmatrix}$$

<u>mean absolute deviation</u>



mean $\frac{10}{3} \doteq \overset{\$}{3.3}$ (MAD)

~~MAD not used much~~

<u>idea 2</u>: square the deviations

$$\begin{pmatrix} (y_1 - \bar{y})^2 \\ (y_2 - \bar{y})^2 \\ \vdots \\ (y_n - \bar{y})^2 \end{pmatrix} = \begin{pmatrix} (\overset{\$}{-3})^2 \\ (\overset{\$}{-2})^2 \\ (\overset{\$}{+5})^2 \end{pmatrix} = \begin{pmatrix} 9\,\$^2 \\ 4\,\$^2 \\ 25\,\$^2 \end{pmatrix}$$

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

$$s = \sqrt{\frac{(y_1-\bar{y})^2 + (y_2-\bar{y})^2 + \dots + (y_n-\bar{y})^2}{n-1}}$$

idea so far

(sample) standard deviation

(SD)

$$s^2 = \frac{(y_1-\bar{y})^2 + \dots + (y_n-\bar{y})^2}{n-1}$$

(sample) variance

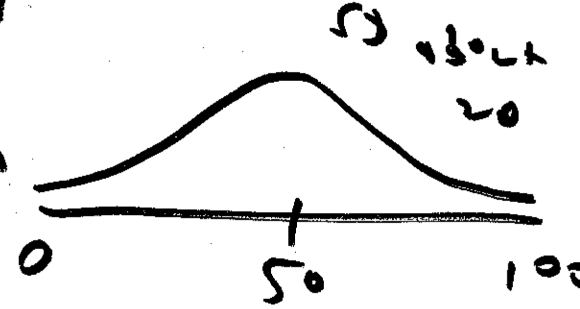
$\begin{bmatrix}1\\2\\9\end{bmatrix}$ mean 4

$\begin{bmatrix}\checkmark\\\checkmark\\X\end{bmatrix}$ mean 4

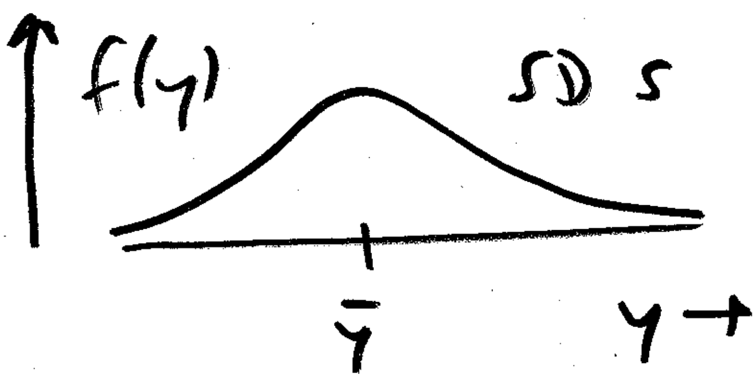
a dataset with n obs. has only (n-1) degrees of freedom for measuring spread

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

each # in this dataset is around 50, give or take around 20

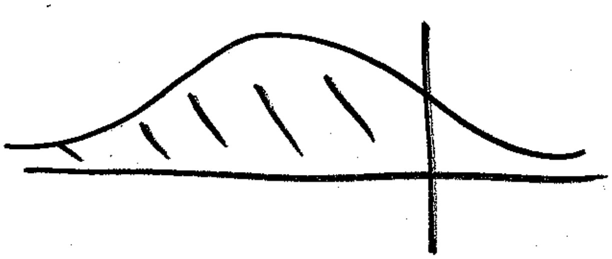


C.F. Gauss (1800) Gaussian dist

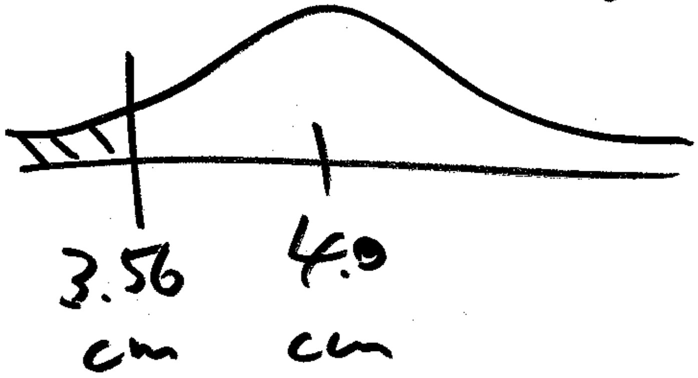


$$f(y) = \frac{1}{s\sqrt{2\pi}} e^{-\frac{(y-\bar{y})^2}{2s^2}}$$

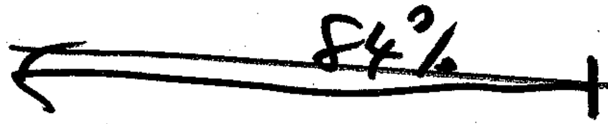
density scale



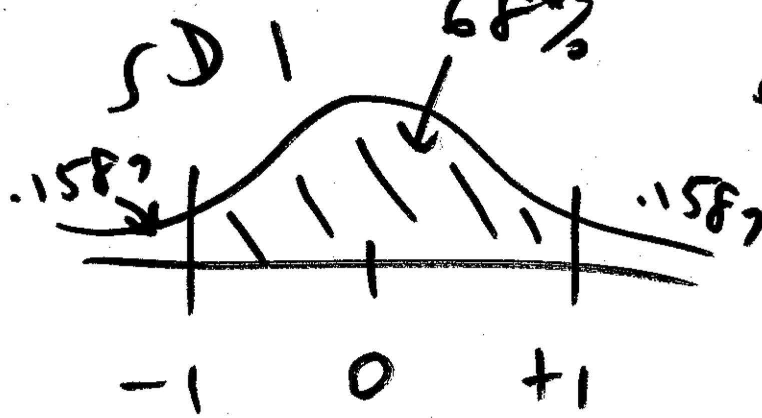
SD 0.29 cm



butterfly wing length



SD 1



standard normal curve

(Z)

$\begin{bmatrix} 5 \\ 5 \\ \vdots \\ 5 \end{bmatrix}$

mean 5

SD 0

$\begin{bmatrix} c \\ \vdots \\ c \end{bmatrix}$

mean c

SD 0

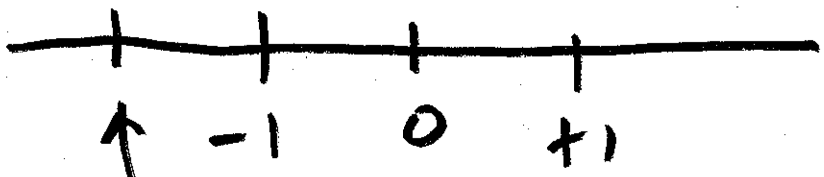
facts about normal curve

① symmetric

② total area under it is 1 (or 100%)

SD 0.29 cm → 0.0643 from table ≈ 6.4%

raw units (cm)



standard units (SU) (pure #s)

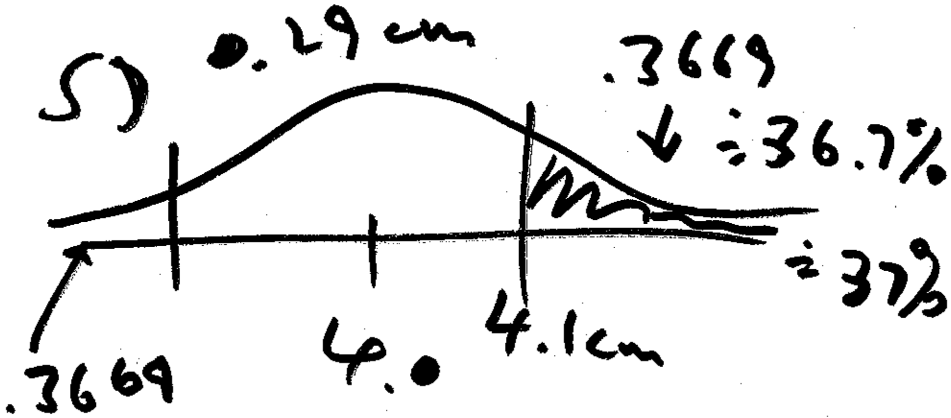
$$\frac{3.56\,\text{cm} - 4.0\,\text{cm}}{0.29\,\text{cm}}$$

converting to standard units

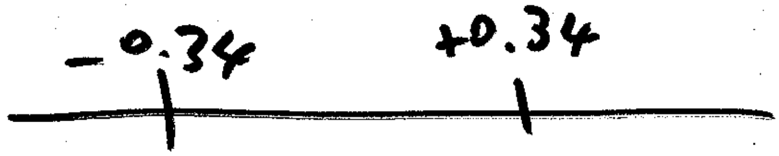
$$= \frac{-0.44\,\cancel{\text{cm}}}{0.29\,\cancel{\text{cm}}} \approx -1.52$$

---

actual answer: 8.3% (normal approximation here: good but not great)



what % of data lie above 4.1 cm?



$$+0.34 = \frac{+0.1}{0.29} = \frac{4.1\cancel{cm} - 4.0\cancel{cm}}{0.29\cancel{cm}}$$

to convert to su,

$$su = \frac{\# - mean}{SD}$$

$$Z = \frac{Y - \overline{Y}}{s}$$

right ans: $\frac{9}{24}$ ≐ 38%