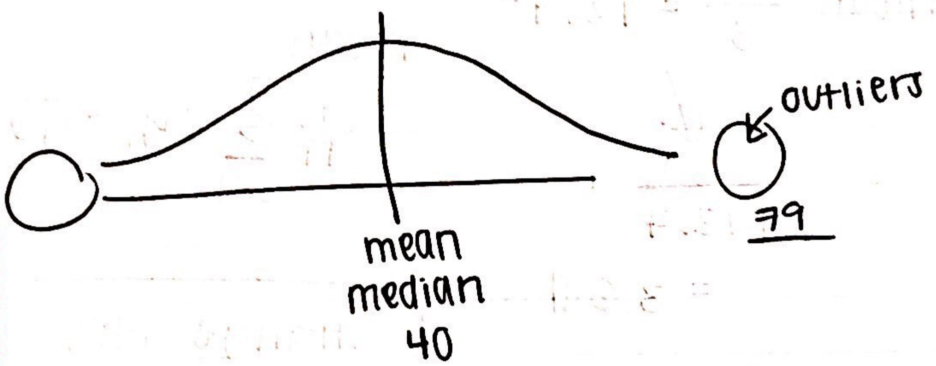
STANDARD DEVIATION, EMPIRICAL RULE, Normal curve (10/15/19)

Read: DD
(A) chpt 1-3
(B) chpt 1-6

Today:
LN pg L-25



mean pulled by the tail

---

absolute value

$$\begin{bmatrix} \$3 \\ \$2 \\ \$5 \end{bmatrix} \qquad \begin{bmatrix} 3 \\ 2 \\ 5 \end{bmatrix}$$

mean 0

mean $\frac{10}{3} = 3.3\$$ ok

Laplace ~1785

MAD

mean absolute deviation → mean

$$\begin{bmatrix} y_1 - \bar{y} \\ \vdots \\ y_n - \bar{y} \end{bmatrix} \updownarrow n$$

$$\begin{bmatrix} |y_1 - \bar{y}| \\ \vdots \\ |y_n - \bar{y}| \end{bmatrix}$$

\$

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

$$\begin{bmatrix} (-3\$)^2 \\ (-2\$)^2 \\ (+5\$)^2 \end{bmatrix} = \begin{bmatrix} 9\$^2 \\ 4\$^2 \\ 25\$^2 \end{bmatrix}$$

(Gauss ~1788)

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

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

mean

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

take $\sqrt{\ }$ at end

$\sqrt{12.7}$

$= 3.6\$$

strange idea at least for now

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

lower case

Fisher (1910) sample (name) variance

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

sample standard deviation (SD)

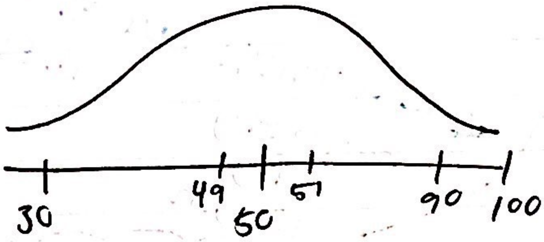
K. Pearson (~1910)

$\boxed{s \geq 0}$

Properties of SD $\longrightarrow$ s can't be negative

Graphical Interpretation of SD

SD = 15



mean

too small

SD 1
10

40

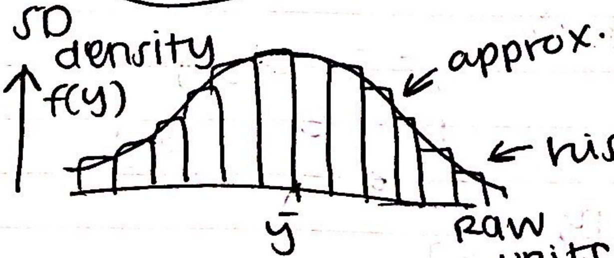
too big

Empirical Rule

↳ start at mean

so $\begin{Bmatrix}1\\2\\3\end{Bmatrix}$ $SD_{(s)}$ either way : you will capture

about $\begin{cases} 2/3 \leftarrow 68\% \\ \text{most} \leftarrow 95\% \\ \text{almost all} \leftarrow 99.7\% \end{cases}$ of data



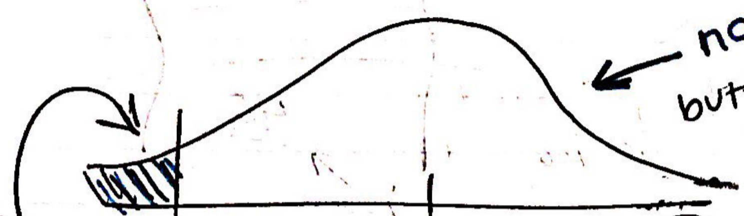
← approx.

← histogram, drawn on density scale

Relative Frequency = area under histogram (curve) de moivre (~1705)

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

SD 0.29cm



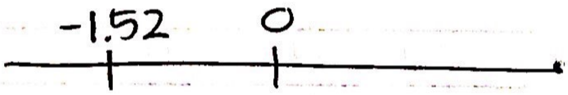
normal curve (K. Pearson ~1900)

butterfly wing length

?
.0643
= 6.4%

(raw units)

Gaussian Distribution (~1790)



(standard units) z

$$\frac{3.56\text{cm} - 4.00\text{cm}}{0.29\text{cm}}$$

Q: what % of butterflies in data set had wing length $\leq$ 3.56cm?

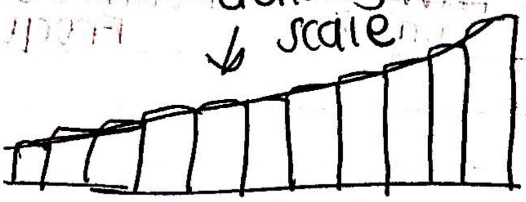
exact: $\frac{2}{24} = \frac{1}{12} = 8.3\%$
A

$$\int_{-\infty}^{c} e^{-\frac{1}{2}y^2}\, dy$$

3.3
3.5
3.56
3.6
3.6
⋮
4.5

n=24

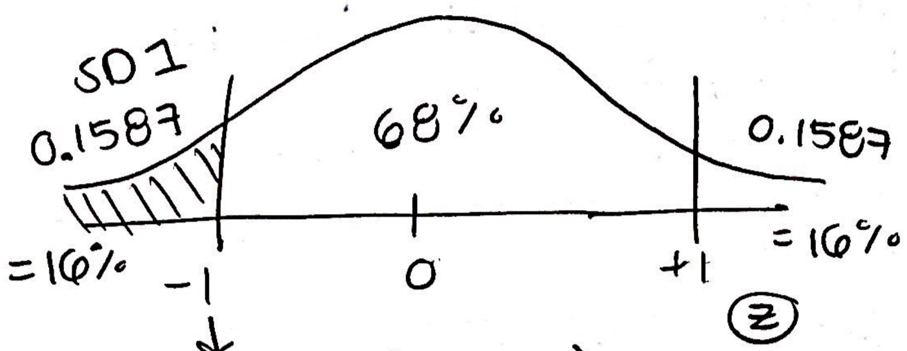
density scale



numerical integration

Standard Normal curve

$\boxed{L-34}$



$-1.00 = -(1+0.00)$

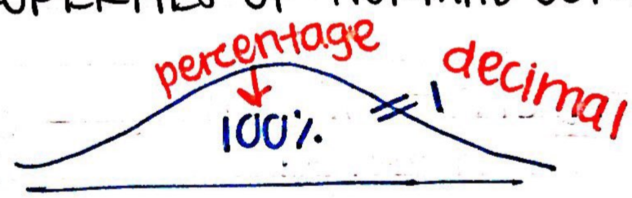
$$\begin{bmatrix} 7 \\ 7 \\ \vdots \\ 7 \end{bmatrix}$$

mean: 7
SD: 0

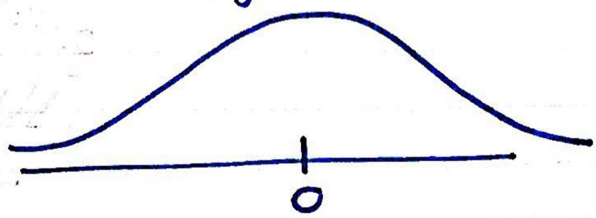
MATH FACT: every normal curve satisfies the Empirical Rule exactly.

2 PROPERTIES OF NORMAL CURVE

1



2 symmetric



Cumulative probability

$P(Z \leq z) \Longleftrightarrow$ area under normal curve to left of z to convert from raw units (y) to standard units (z), ask:

how far is y (#) from $\bar{y}$ (mean), relative to the SD?

$$z = \frac{\# - \text{mean}}{SD} = \frac{y - \bar{y}}{s}$$

pure #'s w/no units

$$\frac{3.56\,\cancel{cm} - 4.0\,\cancel{cm}}{0.29\,\cancel{cm}} = -\frac{0.44}{0.29} = -1.52$$