this time: spread (SD); normal curve

next time: experimental design

read: DD (A) ch. 1-3; (B) ch. 1-6

STAT7 15 Sections

today: 1

LN pp. L-25 →

Quiz 2 due Sat night;

hwk 1 due Fri night

mean pulled by the tail

$$\begin{bmatrix} -3 \\ -2 \\ +5 \end{bmatrix} \qquad \begin{array}{cc} \begin{bmatrix} y_1 - \bar{y} \\ \vdots \\ y_n - \bar{y} \end{bmatrix} & \updownarrow n \end{array}$$

absolute value

mean 0   mean 0

$$\begin{bmatrix} 3 \\ 2 \\ 5 \end{bmatrix} \qquad \begin{bmatrix} |y_1 - \bar{y}| \\ \vdots \\ |y_n - \bar{y}| \end{bmatrix}$$

ok

mean $\frac{10}{3} \doteq 3.3$ \$

Laplace ~1785

MAD

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

8

$$\begin{bmatrix} (-3\$)^2 \\ (-2\$)^2 \\ (+5\$)^2 \end{bmatrix} = \begin{bmatrix} 9\$^2 \\ 4\$^2 \\ 25\$^2 \end{bmatrix} \quad \begin{pmatrix} \text{Gauss} \\ \sim 1785 \end{pmatrix} \begin{bmatrix} (y_1 - \bar{y})^2 \\ \vdots \\ (y_n - \bar{y})^2 \end{bmatrix}$$

mean $\frac{38}{3} \doteq 12.7\$^2$ mean $\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2$

$\downarrow$

$\sqrt{12.7\$^2} \doteq 3.6$

take $\sqrt{\ }$ at end

strange idea, at least for now:

Fisher (1910)

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

(name)

lower-case

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

(SD)

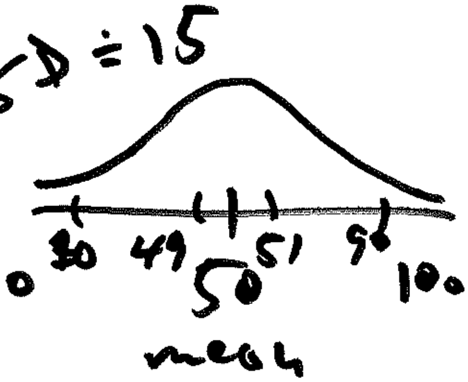
K. Pearson ($\sim$1910)

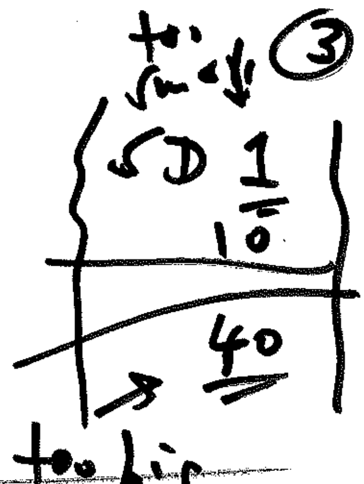
properties of SD

$s \geq 0$

s can't be negative

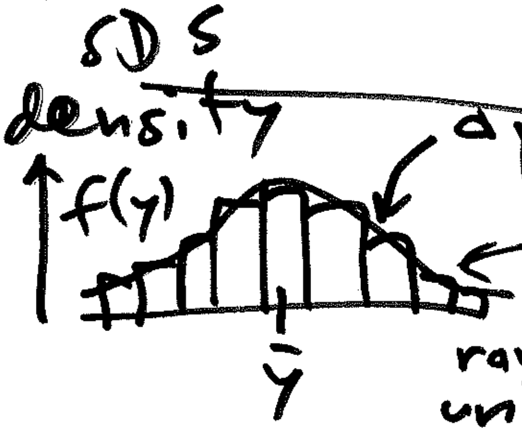
graphical interpretation of SD





Empirical Rule

start at mean, go $\left\{\begin{matrix}1\\2\\3\end{matrix}\right\}$ SD(s) either way: you will capture about $\left\{\begin{matrix}\frac{2}{3} \leftarrow 68\% \\ \text{most} \leftarrow \doteq 95\% \\ \text{almost all} \leftarrow 99.7\%\end{matrix}\right\}$ of data



histogram, drawn on density scale

relative frequency = area under hist. (curve)

deMoivre (~1705)

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

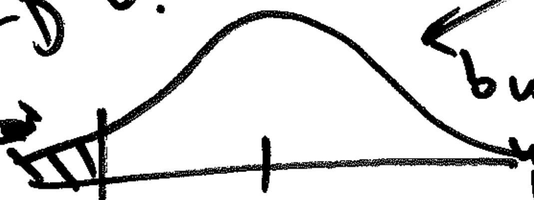
normal curve (4)

(K. Pearson) ~ 1900

Gaussian distribution (~ 1790)

SD 0.29 cm

butterfly wing length



3.56 cm   4.0 cm   y (raw units)

? .0643 $\doteq$ 6.4%

−1.52   0

z (standard units)

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

Q: what % of butterflies in data set had wing length ≤ 3.56cm?

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

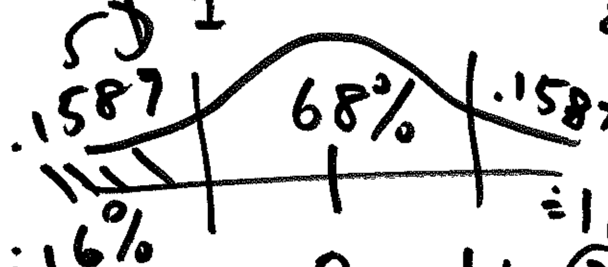
3.3
3.5
──── 3.56
3.6
3.6
⋮
4.5

(2)   n = 24

density scale

numerical integration

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



standard normal curve

L-34

$-1.00 = -(1 + .00)$

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

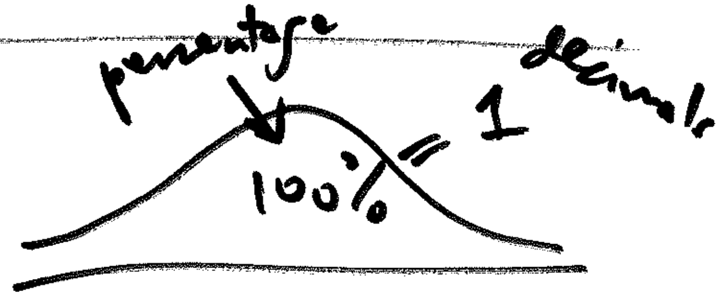
mean 7

SD 0

math fact: every normal curve satisfies the Empirical Rule exactly.

2 properties of normal curve

1



2 symmetric

cumulative probability $P(Z \le z) \leftrightarrow$ 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 s?

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

↑ pure #s, with no units

$$\frac{3.56\ \cancel{gm} - 4.0\ \cancel{gm}}{0.29\ \cancel{gm}} = \frac{-0.44}{0.29}$$

$$\doteq -1.52$$

$(\doteq -1.5)$