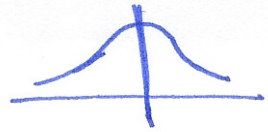
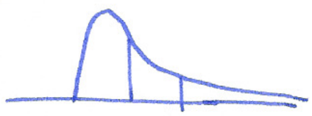
for symmetric distributions, mean close to mode



for skew, mean ≠ mode



mean : sensitive to outliers.

	mean	median
$\underbrace{0,0,\dots 0}_{10}, \underbrace{1,\dots,1}_{10}$. ←	$\frac{1}{2}$	$\frac{1}{2}$.
$\underbrace{0,0\dots 0}_{10}, \underbrace{1,\dots,1}_{10}, 10^6$ ←	10^5	1

spread want to distinguish 

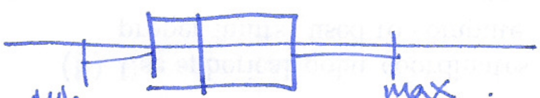
quartiles: 0, 0, 1, 1 | 1, 1, 2, 3 . . | . | 10

Q_1 first quartile — median M — Q_3 ~~second~~ third quartile.

interquartile range : $IQR = Q_3 - Q_1$

five number summary : min, Q_1, M, Q_3, max.

boxplot :



outliers? $1.5 \times IQR$ above Q_3 or below Q_1.

standard deviation "average distance from the mean".

data: ~~wrong w~~ $X = \{0, +1, -1\}$ mean: 0

wrong way: take average of $x_i - \bar{x}$: $-1, 0, +1$ ← average 0 !

better : take average of $(x_i - \bar{x})^2$ $1, 0, 1$ ← average $\frac{2}{3}$.

this is variance $s_x^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$

standard deviation $= s_x = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - x)^2}$.

Q: why not $|x_i - \bar{x}|$?

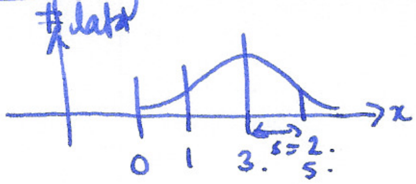
properties : s measures spread about mean (not so good for non-symm)

$s = 0$ iff all x_i the same.

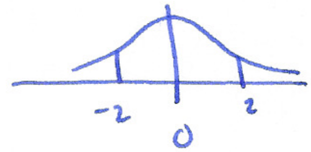
s depends on outliers   s has same units as original variable measures of center and spread

mean	standard deviation	
median	interquartile range.	} may be different. $\boxed{\text{always plot your data!}}$

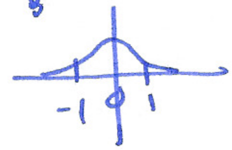
Changing units of measurement



$y_i = x_i - 3$.



$z_i = \frac{x_i - 3}{2}$.



general linear change: $z_i = a + b x_i$

multiply by b: $z_i = b x_i$

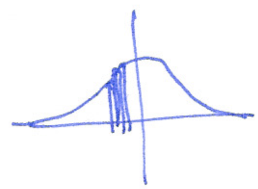
mean (z_i) = b mean (x_i)
median (z_i) = b median (x_i).
s.d. (z_i) = b s.d (x_i)
IQR (z_i) = b IQR (z_i).

add a: $z_i = a + x_i$

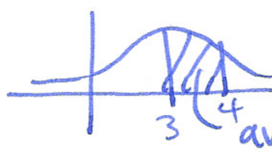
mean (z_i) = a + mean (x_i)
median (z_i) = a + med (x_i).
s.d.
IQR } don't change!

§1.3 Density curves and normal distributions

lots of data: histogram looks like a smooth curve.



density curves: area under graph = 1.

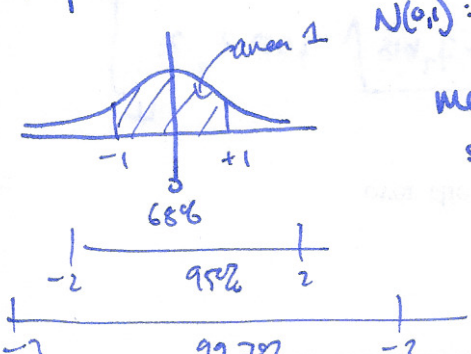


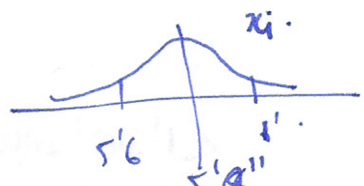
area = proportion of data between 3 and 4.

important distribution: <u>normal distribution</u>

$N(0,1)$: $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2}$

mean $\mu = 0$
s.d. $\sigma = 1$





male height
μ = 5'9".
s.d = 3 in.

$z_i = \frac{x_i - 5'9"}{3"}$ ← gives $N(0,1)$

notation: $N(0,1)$	normal with	$\mu = 0, \sigma = 1$	z_i
$N(4,2)$		$\mu = 4$ $\sigma = 2$	$x_i = 2z_i + 4$
$N(\mu,\sigma)$		μ, σ	$x_i = \sigma z_i + \mu$.