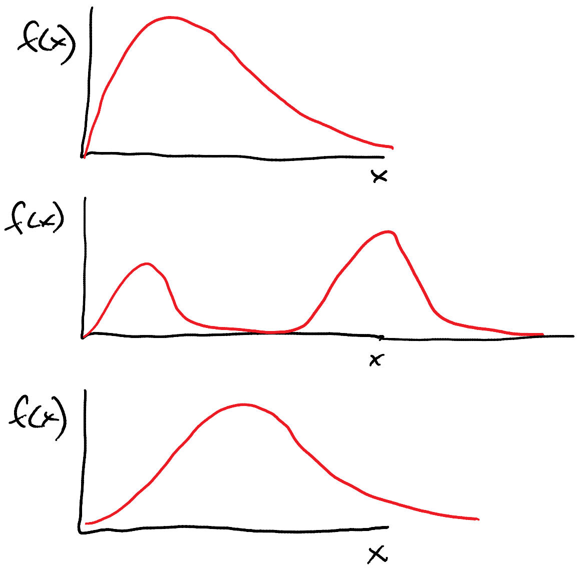
Section 1.5: Tchebysheff and the Empirical Rule

Thursday, May 14, 2015
10:34 AM

distributions



skewed to the right
asymmetrical
unimodal

bimodal
asymmetrical

symmetrical
unimodal
"mound shaped"

questions: how big is a standard deviation on each graph?
what can we say about the proportion of measurements in general?

Tchebysheff's theorem: works for all distributions (symmetrical or not unimodal or multimodal)

- for <u>any</u> set of measurements,
at least $\left(1 - \frac{1}{k^2}\right)$ of the measurements will lie within k standard deviations of the mean for $k \geq 1$.

k	$1 - 1/k^2$			
1	0	so $\geq 0\%$	lie within	$\mu \pm 1\sigma$ (or $\bar{x} \pm 1s$)
1.5	5/9	$\geq 55.\bar{5}\%$		$\mu \pm 1.5\sigma$
2	3/4	$\geq 75\%$		$\mu \pm 2\sigma$
2.5	21/25	$\geq 84\%$		$\mu \pm 2.5\sigma$
3	8/9	$\geq 88.\bar{8}\%$		$\mu \pm 3\sigma$

completely useless statement

example: A sample data set has a mean of 7.2 and a standard deviation of 0.3.

a) What can you say about the proportion of data points between the values 6.6 and 7.8 ?

b) What can you say about the proportion of data points between 6.9 and 7.5?

c) What can you say about the proportion of data points less than 6.6 or greater than 7.8 ?

d) What can you say about the proportion of data points above 7.8?

e) at least 50% of the data points lie within a range of values. What are the upper and lower limits of those values?

a) $\bar{x} = 7.2$ what is $6.6 - 7.8$? $\bar{x} \pm 2s$
$s = 0.3$ 7.2 ± 0.6

$\geq 75\%$ of data fall between 6.6 and 7.8

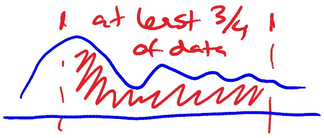
b) $6.9 - 7.5$ $\bar{x} \pm 1s$

nothing. we can't say anything.

(if you insist, $\geq 0\%$ ...)

c) < 6.6 or > 7.8?

$< 25\%$ because $\geq 75\%$ lie within that range



how much

d) 

how much up here?

$< 25\%$ because no info about symmetry and there could be no data below 6.6

e) $\geq 50\%$

$$0.5 = 1 - \frac{1}{k^2}$$

$$\frac{1}{k^2} = \frac{1}{2}$$

$$k = \sqrt{2}$$

$$\bar{x} \pm ks = \boxed{7.2 \pm \sqrt{2}(0.3)}$$

$$= 6.77 - 7.62$$

$$= 6.8 - 7.6$$

The Empirical Rule:

mound-shaped: used to describe roughly symmetrical, unimodal distributions

Empirical Rule says:

if a data set is mound-shaped, then

$\sim 68\%$ of the data lie within $\mu \pm 1\sigma$

approximately

$\sim 95\%$ " " " " " $\mu \pm 2\sigma$

~99.7% " " " " " $\mu \pm 3\sigma$