Section 6.3: Tchebysheff's Theorem and the Empirical Rule

Tuesday, November 07, 2023 3:59 PM

Chebyshev

Tchebysheff's theorem: works for all distributions/ data sets

(symmetrical or skewed, unimodal, bi or multi-modal)

- for any set of measurements

at least $\left(1 - \frac{1}{k^2}\right)$ of the measurements will fall within k standard deviations of the mean for $k > 1$.

what does it mean?

suppose we have a data set with the following graph, where the sample mean is 50 and the std dev is 10.

look at the interval from 30 to 70

30: 2 std dev below the mean

70: 2 std dev above the mean

the shaded area is "within 2 std dev of the mean"

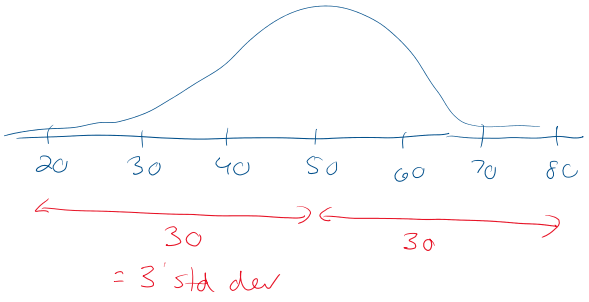
so $k = 2$

so $k = 2$

$$1 - \frac{1}{k^2} = 1 - \frac{1}{2^2} = 1 - \frac{1}{4} = \frac{3}{4} \text{ or } 75\%$$

Tcheby says "at least 75% of the measurements lie within 30 and 70"

what about the number of measurements between 20 and 80?



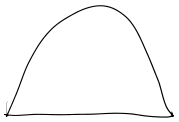
so $k = 3$

and $1 - \frac{1}{k^2} = 1 - \frac{1}{3^2} = \frac{8}{9}$ or $88.\overline{8}\%$

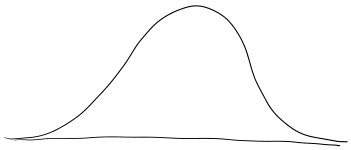
then "at least 89% of measurements fall within the interval 20 to 80"

k	$1 - 1/k^2$		
1.5	5/9	$\geq 55\%$	lie within (mean $\pm$ 1.5 std dev)
2	3/4	$\geq 75\%$	$\pm$ 2
2.5	21/25	$\geq 84\%$	$\pm$ 2.5
3	8/9	$\geq 89\%$	$\pm$ 3

The Empirical Rule: only works for "mound-shaped" or "bell-shaped" curves



mound



bell

- for unimodal and roughly symmetrical

approximately 68% of measurements fall within (mean ± 1 std dev)
" 95% ± 2 std dev
" 99.7% ± 3