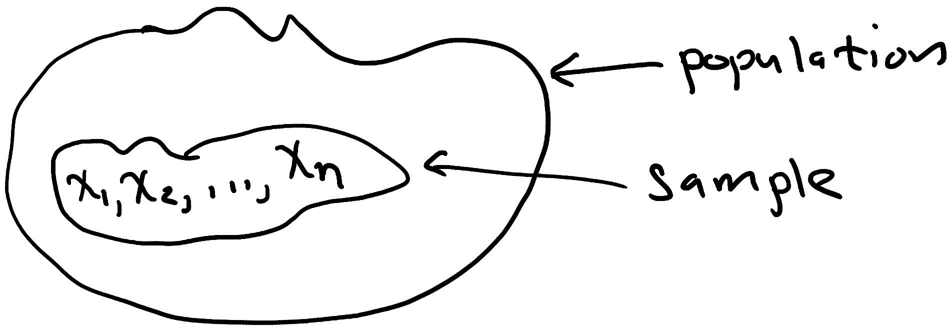
01/25/2023

Wednesday, January 25, 2023 6:23 AM

01/25/2023

Chapter 3 - Describing, Exploring, and Comparing Data



Summarize sample and/or population values using measures of center, variability, and other quantities.

Measure of Center

Def: Sample mean $\bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n} = \dfrac{x_1 + x_2 + \ldots + x_n}{n}$

$\bar{x}$ measures the "balance point" of the sample $x_1, x_2, \ldots, x_n$.

Ex: x = height (inches) of class of students (last time) $n = 22$

$\sum_{i=1}^{22} x_i = x_1 + \ldots + x_{22} = 1542$

$\bar{x} = \dfrac{1542}{22} = 70.\overline{09}$ inches

$$\bar{x} = \frac{1542}{22} = 70.\overline{09} \text{ inches}$$

or about 5 ft 10 inches (5' 10")

Def: Sample median

$$\tilde{x} = \begin{cases} \text{single middle ordered value, if } n \text{ is odd} \\ \text{mean of the two middle ordered values, if } n \text{ is even} \end{cases}$$

Ex: (continued)

$n = 22$

11th 12th

$$\underline{64} \quad \underline{66} \quad \ldots \quad \underline{70} \,\vdots\, \underline{70} \quad \ldots \quad \underline{75} \quad \underline{75}$$

Find $\tilde{x}$

$$\tilde{x} = \frac{70+70}{2} = 70 \text{ inches}$$

If $x_{max} = 75$ is removed, then $\tilde{x} = 70$.

Remark: $\bar{x}$ is highly affected by outliers whereas $\tilde{x}$ is relatively unaffected by outliers.

Ex: (follow on) Remove $x_{max} = 75$ and replace it with 107 inches.

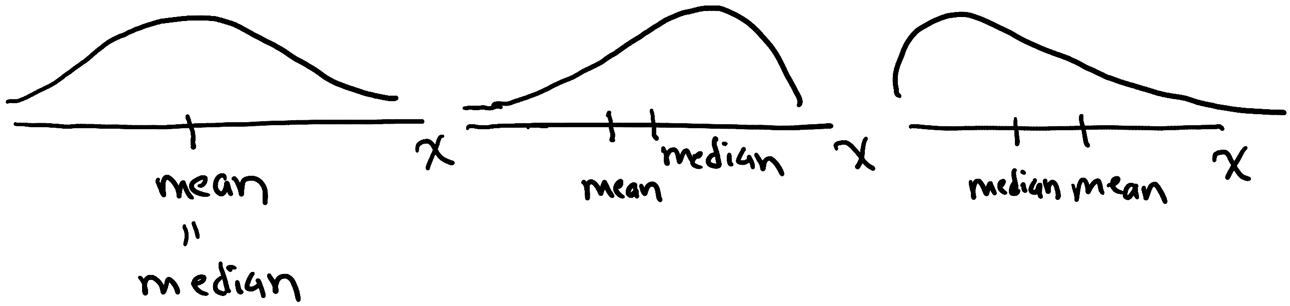
$$\bar{x} = 71.\overline{54}$$

$$\tilde{x} = 70$$

Notation: population mean $= \mu$

<u>Notation:</u> population mean $= \mu$

population median $= \tilde{\mu}$



<u>Measures of Variability</u>

<u>Def</u>: sample Range $= x_{max} - x_{min}$

<u>Def</u>: i^{th} deviation from sample mean is

$$x_i - \bar{x} \quad , \quad i = 1, 2, \ldots, n$$

<u>Remark</u>: $\sum_{i=1}^{n} (x_i - \bar{x}) = 0$ (sum of all deviations is zero)

<u>Def</u>: sample variance $s^2 = \dfrac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$ "Average" squared deviation from the sample mean

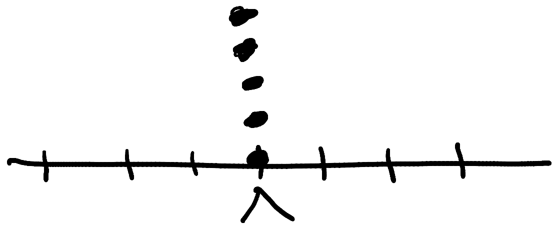
<u>Def</u>: sample standard deviation

$s = \sqrt{s^2}$ "Typical" distance an x-value is from the

$s = \sqrt{s^2}$

"Typical distance an x-value is from the sample mean.

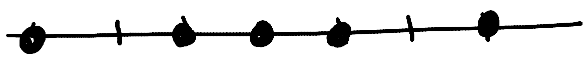
Ex:



$n = 5$
$s = 0$

1.)

2.)



$s_1 \quad s_2$?

Ex: x = height (inches)

$n = 22$ $\quad \bar{x} = 70.\overline{09}$

i	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	x_i^2
1	64	$-6.\overline{09}$	≈ 37.0992	$(64)^2$
⋮	⋮	⋮	⋮	⋮
22	75	$4.\overline{90}$	≈ 24.0992	$(75)^2$
	1542	0	197.8181	108278.

$$s^2 = \frac{197.8181}{22-1} = 9.4199 \text{ inches}^2$$

$$s = \sqrt{s^2} = 3.0692 \text{ inches}$$

Interpretation of s?

Computational Formula for s^2:

$$s^2 = \sum^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2$$

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1}$$

Ex:

$$s^2 = \frac{108278 - (1542)^2/22}{22-1}$$

Notation: Population Variance $= \sigma^2$ "sigma^2"

Population standard deviation $= \sigma$ "sigma"