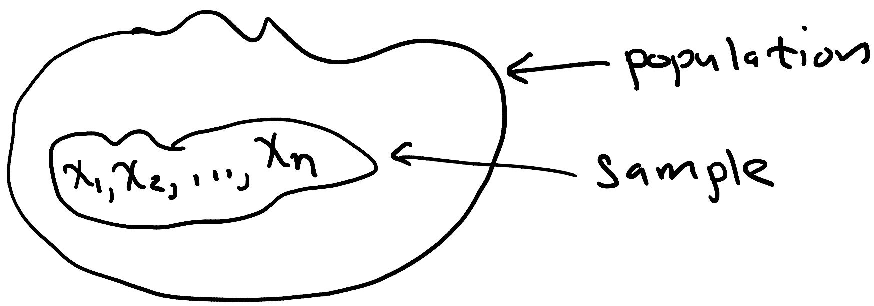
01/25/2023
(Revisited on 01/30/2023)

Chapter 3 – Describing, Exploring, and Comparing Data

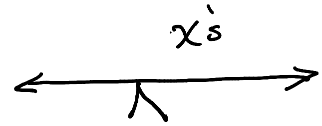


Summarize sample and/or population values using measures of center, variability, and other quantities.

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) $n = 22$

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

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

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)

$n = 22$ 64 66 ... 70 | 70 ... 75 75

11th, 12th ordered position

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. $(n = 22)$

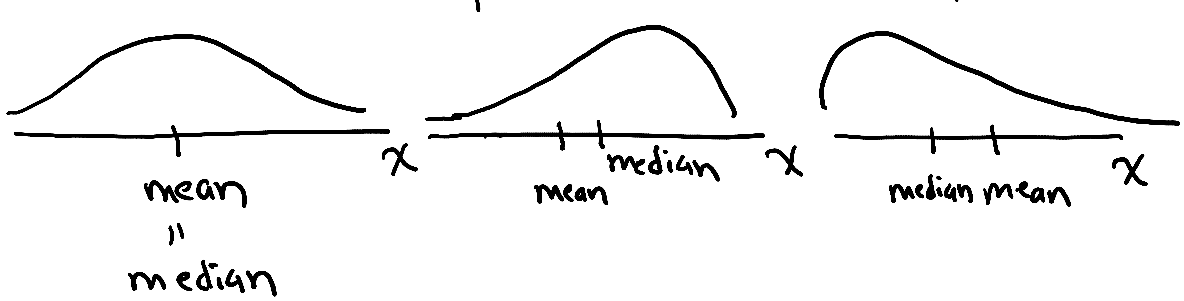
$$\bar{x} = 71.\overline{54}$$

$$\tilde{x} = 70$$

Notation: population mean $= \mu$ ← "mu"

population median $= \tilde{\mu}$

Examples of Histogram Shapes

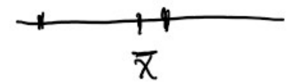


Measures of Variability

Def: sample Range $= x_{max} - x_{min}$

Def: i^{th} deviation from sample mean is

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



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

Def: sample variance

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

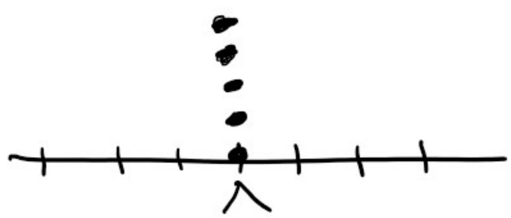
"Average" squared deviation from the sample mean

Def: sample standard deviation

$$s = \sqrt{s^2}$$

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

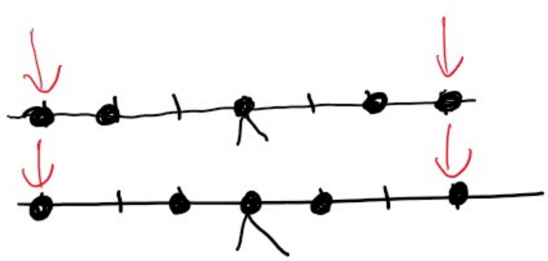
Ex:



$n = 5$

$s = 0$

1.)
2.)



$s_1 > s_2$

Ex: x = height (inches)

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

i	x_i	$x_i-\bar{x}$	$(x_i-\bar{x})^2$	x_i^2
1	64	$-6.\overline{09}$	≈ 37.0992	$(64)^2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
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:

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

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

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

Population standard deviation $= \sigma$ "sigma"

01/30/2023
(NEW)

Sample Quartiles

First, order the x-values from smallest to largest.

Def: lower quartile = median of the lower half of the data of the data

Q_1

upper quartile = median of the upper half of the data

Q_3

Def: Interquartile Range (IQR)

$$IQR = Q_3 - Q_1$$

= upper quartile − lower quartile

= difference between upper and lower quartiles

Ex: x = height (continued)

$n = 22$

$Q_1 = 68$ $Q_2 = 70$ $Q_3 = 71$

64 ... 68 ... 70 : 70 ... 71 ... 75

68 is the 6th, 70 : 70 are the 11th and 12th, 71 is the 17th

$$IQR = Q_3 - Q_1 = 3 \text{ inches}$$
$$= 71 - 68$$

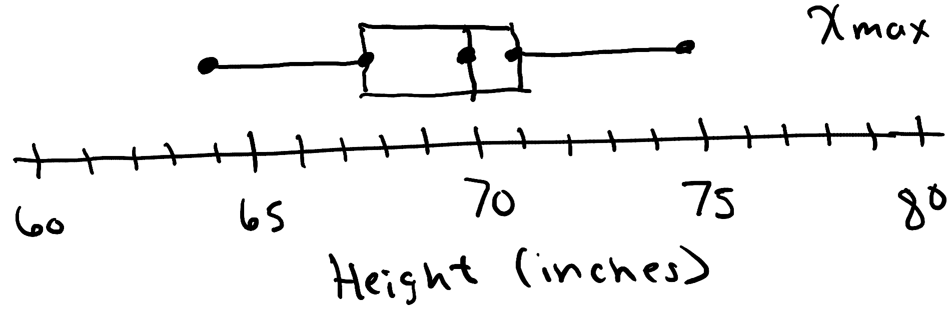
Five number summary of a sample and boxplot are useful summaries of the sample data.

Def: A skeletal boxplot is a visual display of data based on the five numbers below:

min, Q_1, median, Q_3, max

Ex: x = height
$n = 22$

$x_{min} = 64$
$Q_1 = 68$
$\tilde{x}$ = median = 70
$Q_3 = 71$
$x_{max} = 75$



02/01/2023

Def: Percentiles, denoted by $P_1, P_2, \ldots, P_{100}$ divide data into 100 groups with about 1% of the values in each group.