Statistics

Lecture 3

Feb 19-8:47 AM

Basic Computations in Statistics: (Sk 5-8)

$x \to$ data element

$\sum x \to$ Sum of data elements

Summation

$n \to$ Sample Size

$\bar{x} \to$ x-bar $\to$ Sample Mean (Average)

$$\boxed{\bar{x} = \frac{\sum x}{n}}$$

Consider the Sample below

2 3 3 4 8

1) $n = 5$

2) Range $=$ Max $-$ Min $= 8 - 2 = 6$

3) Midrange $= \frac{\text{Max} + \text{Min}}{2} = \frac{8+2}{2} = \frac{10}{2} = 5$

4) Mode $= 3$

5) Median 3

"Data must be Sorted"

6) $\sum x = 2 + 3 + 3 + 4 + 8 = \boxed{20}$

7) $\bar{x} = \frac{\sum x}{n} = \frac{20}{5} = \boxed{4}$

Mar 9-1:48 PM

Consider the Sample below

1 3 $\boxed{3 \quad 5}$ 5 7

1) $n = 6$

2) Range $= 7 - 1 = 6$

3) Midrange $= \dfrac{7+1}{2} = \dfrac{8}{2} = \boxed{4}$

4) Mode $=$ 3 & 5
Bimodal

5) Median $= \dfrac{3+5}{2} = \boxed{4}$
"Data must be Sorted"

6) $\sum x = 1 + 3 + 3 + 5 + 5 + 7 = \boxed{24}$

7) $\bar{x} = \dfrac{\sum x}{n} = \dfrac{24}{6} = \boxed{4}$

Mar 9-1:56 PM

$x \rightarrow$ Data element

$x^2 \rightarrow$ Square of data element

$\sum x \rightarrow$ Sum of data elements

Summation

$\sum x^2 \rightarrow$ Sum of Squares of data elements

$\bar{x} \rightarrow$ x-bar $\rightarrow$ Sample Mean

$S^2 \rightarrow$ Sample Variance

$\boxed{\bar{x} = \dfrac{\sum x}{n}}$, $\boxed{S^2 = \dfrac{\sum (x-\bar{x})^2}{n-1}}$, $\boxed{S^2 = \dfrac{n\sum x^2 - (\sum x)^2}{n(n-1)}}$

Mar 9-2:06 PM

Consider the Sample below

2 4 6 8 10

1) $n=5$

2) Range $= 10-2 = 8$

3) Midrange $= \frac{10+2}{2} = 6$

4) Mode $\boxed{\text{None}}$

5) Median 6

6) $\sum x = 2+4+6+8+10$
$= 30$

7) $\bar{x} = \frac{\sum x}{n} = \frac{30}{5} = \boxed{6}$

8) $\sum x^2 = 2^2+4^2+6^2+8^2+10^2 = \boxed{220}$

9) $S^2 = \frac{n\sum x^2 - (\sum x)^2}{n(n-1)} = \frac{5\cdot 220 - 30^2}{5(5-1)} = \frac{200}{20} = \boxed{10}$

Mar 9-2:11 PM

Consider the Sample below

1 3 3 $\boxed{3 \quad 5}$ 5 5 7

1) $n=8$

2) Range $= 7-1 = 6$

3) Midrange $= \frac{7+1}{2} = 4$

4) Mode $=$ 3 & 5
Bimodal

5) Median $\frac{3+5}{2} = \boxed{4}$

6) $\sum x = 1+3+3+3+5+5+5+7 = \boxed{32}$

7) $\sum x^2 = 1^2+3^2+3^2+3^2+5^2+5^2+5^2+7^2 = \boxed{152}$

8) $\bar{x} = \frac{\sum x}{n} = \frac{32}{8} = \boxed{4}$

9) $S^2 = \frac{n\sum x^2 - (\sum x)^2}{n(n-1)}$
$= \frac{8\cdot 152 - 32^2}{8(8-1)}$
$= \frac{192}{56}$
$\approx \boxed{3.429}$

Mar 9-2:19 PM

$\bar{x} \rightarrow$ Sample Mean

$s^2 \rightarrow$ Sample Variance

$s \rightarrow$ Sample Standard deviation

$$\bar{x} = \frac{\sum x}{n} \qquad s^2 = \frac{\sum (x-\bar{x})^2}{n-1} \qquad s^2 = \frac{n\sum x^2 - (\sum x)^2}{n(n-1)}$$

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

Mar 9-2:27 PM

Consider the Sample below

2 3 3 3 4

1) $n = 5$

2) $\sum x = 15$

3) $\sum x^2 = 47$

4) $\bar{x} = \frac{\sum x}{n} = \frac{15}{5} = \boxed{3}$

5) $s^2 = \frac{n\sum x^2 - (\sum x)^2}{n(n-1)} = \frac{5 \cdot 47 - 15^2}{5(5-1)} = \frac{10}{20} = \boxed{.5}$

6) $s = \sqrt{s^2} = \sqrt{.5} \approx \boxed{.707}$

Mar 9-2:30 PM

Given $n=8$ $\quad \sum x=115$ $\quad \sum x^2=1745$

1) $\bar{x} = \frac{\sum x}{n} = \frac{115}{8} = \boxed{14.375}$

whole	14
1-Dec.	14.4
2-Dec.	14.38

2) $s^2 = \frac{n\sum x^2 - (\sum x)^2}{n(n-1)} = \frac{8 \cdot 1745 - 115^2}{8(8-1)} = \frac{735}{56} = \boxed{13.125}$

whole	13
1-Dec.	13.1
2-Dec.	13.13

3) $s = \sqrt{s^2} = \sqrt{13.125}$
$= \boxed{3.623}$

whole	4
1-Dec.	3.6
2-Dec.	3.62

How to estimate S:

$$\boxed{s \approx \frac{\text{Range}}{4}}$$

The range rule-of-thumb.

Given min = 45, Max = 95, Estimate S.

$s \approx \frac{\text{Range}}{4} = \frac{95-45}{4} = \frac{50}{4} = \boxed{12.5}$

Mar 9-2:39 PM

Given $n=8$ $\quad \sum x=56$ $\quad \sum x^2=392$

1) $\bar{x} = \frac{\sum x}{n} = \frac{56}{8} = \boxed{7}$

2) $s^2 = \frac{n \cdot \sum x^2 - (\sum x)^2}{n(n-1)} = \frac{8 \cdot 392 - 56^2}{8(8-1)} = \frac{0}{56} = \boxed{0}$

3) $s = \sqrt{s^2} = \sqrt{0} = \boxed{0}$

Mar 9-2:50 PM

What is sample standard deviation?

$S \geq 0$

It is a number that indicates how data elements are spread from $\bar{x}$.

If S is small $\rightarrow$ Data elements are close to $\bar{x}$.

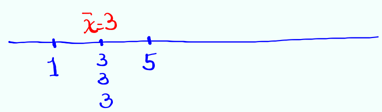
If S is large $\rightarrow$ Data elements are more spread out from $\bar{x}$.

$\boxed{\text{If } S = 0 \rightarrow \text{All data elements are equal to } \bar{x}.}$

Mar 9-2:56 PM

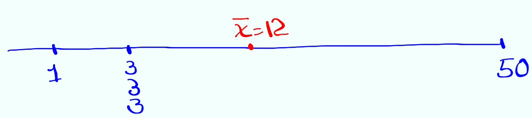
Sample: 1 3 3 3 5

$\bar{x} = 3$, $S = 1.414$



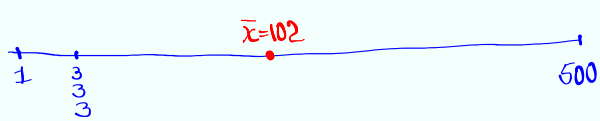
Sample: 1 3 3 3 50

$\bar{x} = 12$ $S = 21.260$



Sample 1 3 3 3 500

$\bar{x} = 102$ $S = 222.490$



Sample 2 3 3 3 4

$\bar{x} = 3$ $S = .707$

Mar 9-3:02 PM

Empirical Rule:

Mean = Mode = Median $\Rightarrow$ Data dist. to be Symmetric

68% Range $\rightarrow \bar{x} \pm S$

$\boxed{95\% \text{ Range} \rightarrow \bar{x} \pm 2S}$ Usual Range

99.7% Range $\rightarrow \bar{x} \pm 3S$

Mar 9-3:18 PM

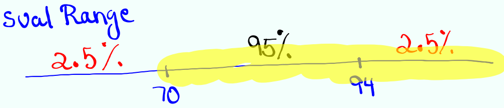
I randomly Selected $\boxed{80 \text{ exams.}}$

$\bar{x} = 82$, $S = 6$

68% Range $\rightarrow \bar{x} \pm S = 82 \pm 6 \rightarrow \boxed{76 \text{ to } 88}$

95% Range $\rightarrow \bar{x} \pm 2S = 82 \pm 2(6) \rightarrow \boxed{70 \text{ to } 94}$

Usual Range



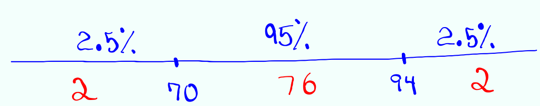
What % of Scores were at least 70?

95% + 2.5% = 97.5%

How many of Scores were at least 70?

97.5% of 80

$.975(80) = \boxed{78}$



Mar 9-3:21 PM

Z- Score

Always round to 3-dec. places.

$$Z = \frac{x - \bar{x}}{S}$$

It is a method to standardize data elements.

It allows us to compare data elements from different Samples.

It tells us how many Standard deviations is the data element from $\bar{x}$.

If $Z > 0 \rightarrow x > \bar{x}$

If $Z < 0 \rightarrow x < \bar{x}$

If $Z = 0 \rightarrow x = \bar{x}$

When $-2 \leq Z \leq 2 \rightarrow$ Data element is usual.

When $Z < -2$ or $Z > 2 \rightarrow$ Data element is unusual.

Mar 9-3:31 PM

John made 90 on exam 1 & 80 on exam 2.

Exam 1 : $\bar{x} = 84$, $S = 8$

$$Z = \frac{x - \bar{x}}{S} = \frac{90 - 84}{8} = \boxed{.75}$$ usual Score

Exam 2 : $\bar{x} = 70$, $S = 4$

$$Z = \frac{x - \bar{x}}{S} = \frac{80 - 70}{4} = \boxed{2.5}$$ unusual Score

unusual | -2 | usual | 2 | unusual

Mar 9-3:39 PM

Salaries of Selected nurses in LA County had $\boxed{\bar{x} = \$6500 \quad , \quad S = \$400}$

1) Maria makes \$7500, find her Z-Score.

$$Z = \frac{x - \bar{x}}{S} = \frac{7500 - 6500}{400} = \boxed{2.5}$$ unusual Salary

2) My Z-Score as a nurse is -1.75. Find my Salary.

$$Z = \frac{x - \bar{x}}{S}$$

$$-1.75 = \frac{x - 6500}{400}$$

Cross-multiply

$$x - 6500 = -1.75(400)$$

$$x = 6500 - 1.75(400)$$

$$\boxed{x = 5800}$$

Mar 9-3:45 PM

5 - Number Summary

Min. $\quad Q_1 \quad$ Median $\quad Q_3 \quad$ Max.

Draw Box Plot

IQR (Inter - Quartile - Range) $= Q_3 - Q_1$

Upper fence $= Q_3 + 1.5(IQR)$

Lower fence $= Q_1 - 1.5(IQR)$

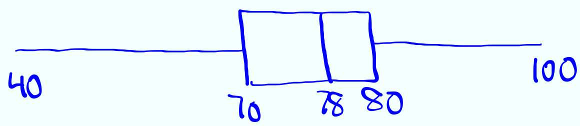
Outliers

Mar 9-3:54 PM

Consider the Following 5-Number Summary.

40	70	78	80	100
Min.	Q_1	Med.	Q_3	Max

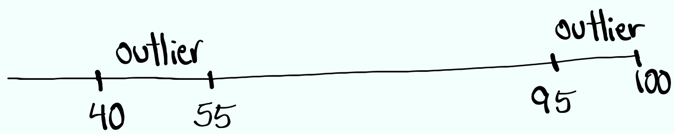
Box Plot



$IQR = Q_3 - Q_1 = 80 - 70 = 10$

Upper Fence $= Q_3 + 1.5(IQR) = 80 + 1.5(10) = \boxed{95}$

Lower Fence $= Q_1 - 1.5(IQR) = 70 - 1.5(10) = \boxed{55}$



Mar 9-3:59 PM

Class QZ 1:

Consider the Sample below

$2 \quad 4 \quad \boxed{4 \quad 6} \quad 6 \quad 7$

1) Median $= \dfrac{4+6}{2} = \boxed{5}$

2) $\sum x = \boxed{29}$

3) $\sum x^2 = \boxed{157}$

4) $\bar{x} = \dfrac{\sum x}{n} = \dfrac{29}{6} = \boxed{4.8\bar{3}}$

Mar 9-4:04 PM