Statistics

Lecture 3

Feb 19-8:47 AM

SG 5

Computations in Statistics:

n → Sample Size

x → Data element

$\sum x$ → Sum of data elements

Summation

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

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

Sep 16-6:53 PM

Consider the Sample below

1 3 3 3 5

1) $n=5$

2) Range $=5-1=4$

3) Midrange $=\frac{5+1}{2}=3$

4) Mode $\boxed{3}$

5) Median $\boxed{3}$

Value in the middle

Data must be Sorted

6) $\sum x=1+3+3+3+5=\boxed{15}$

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

whenever Mean, Mode, and midian are equal, data dist. will be Symmetric.

Sep 16-6:56 PM

Consider the Sample below

1 2 $\boxed{2 \quad 4}$ 4 5

1) $n=\boxed{6}$

2) Range $=5-1=\boxed{4}$

3) Midrange $=\frac{5+1}{2}=\boxed{3}$

4) Mode 2 & 4

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

6) $\sum x=1+2+2+4+4+5=\boxed{18}$

7) $\bar{x}=\frac{\sum x}{n}=\frac{18}{6}=\boxed{3}$

Sep 16-7:01 PM

$x \to$ data element

$n \to$ Sample Size

$\sum x \to$ sum of data elements

$\bar{x} \to$ Sample Mean $\boxed{\bar{x} = \frac{\sum x}{n}}$

$\sum x^2 \to$ square every data element, then add them together

$S^2 \to$ Sample Variance

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

$\boxed{S^2 \geq 0}$ Always

Variance is a value that measures how data elements are spread out from $\bar{x}$.

It is a measure of variation of elements of a data.

Sep 16-7:06 PM

Consider the Sample below

2 3 3 3 4

$n = \boxed{5}$ Range $= 4-2 = \boxed{2}$ Midrange $= \frac{4+2}{2} = \boxed{3}$

Mode $\boxed{3}$ Median $\boxed{3}$

$\sum x = 2 + 3 + 3 + 3 + 4 = \boxed{15}$

$\sum x^2 = 2^2 + 3^2 + 3^2 + 3^2 + 4^2 = \boxed{47}$

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

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

Sep 16-7:13 PM

Given $n=8$ $\sum x=34$ $\sum x^2=186$

Min $=1$ Max $=8$

Range $=8-1=\boxed{7}$ Midrange $=\frac{8+1}{2}=\boxed{4.5}$

$\bar{x}=\frac{\sum x}{n}=\frac{34}{8}=\boxed{4.25}$

whole $\to 4$

1-Dec. $\to 4.3$

Round-up to whole $\to 5$

$$S^2=\frac{n\sum x^2-(\sum x)^2}{n(n-1)}$$

$$=\frac{8\cdot 186-34^2}{8(8-1)}$$

$$=\frac{332}{56}\approx\boxed{5.929}$$

Sep 16-7:20 PM

Given $n=6$ $\sum x=42$ $\sum x^2=294$

1) $\bar{x}=\frac{\sum x}{n}=\frac{42}{6}=\boxed{7}$

2) $S^2=\frac{n\sum x^2-(\sum x)^2}{n(n-1)}=\frac{6\cdot 294-42^2}{6(6-1)}=\frac{0}{30}=\boxed{0}$

When $S^2=0$, all data elements are equal to $\bar{x}$.

Sep 16-7:28 PM

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

$s^2 \rightarrow$ Sample Variance

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

$s \rightarrow$ Sample Standard deviation

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

Standard deviation measures how data elements are spread out from $\bar{x}$.

when s is small $\rightarrow$ data elements are close to $\bar{x}$.

when s is big $\rightarrow$ data elements are more spread out from $\bar{x}$.

when $s = 0 \Rightarrow$ All data elements are the Same as $\bar{x}$.

s has same unit as data elements.

Sep 16-7:34 PM

Consider the Sample below

2 3 3 4 4

1) $n = 5$

2) $\sum x = 16$

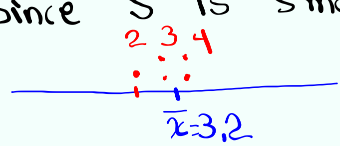
3) $\sum x^2 = 54$

4) $\bar{x} = \dfrac{\sum x}{n} = \dfrac{16}{5} = \boxed{3.2}$

5) $s^2 = \dfrac{n\sum x^2 - (\sum x)^2}{n(n-1)} = \dfrac{5 \cdot 54 - 16^2}{5(5-1)} = \dfrac{14}{20} = \boxed{.7}$

6) $s = \sqrt{s^2} = \sqrt{.7} \approx \boxed{.837}$

Since s is small, data elements are close to $\bar{x} = 3.2$



Sep 16-7:40 PM

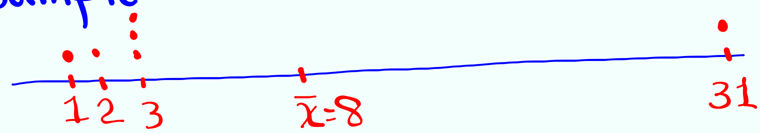
Given $n=5$ $\quad \sum x = 40$ $\quad \sum x^2 = 984$

1) $\bar{x} = \dfrac{\sum x}{n} = \dfrac{40}{5} = \boxed{8}$

2) $S^2 = \dfrac{n\sum x^2 - (\sum x)^2}{n(n-1)} = \dfrac{5 \cdot 984 - 40^2}{5(5-1)} = \dfrac{3320}{20} = \boxed{166}$

3) $S = \sqrt{S^2} = \sqrt{166} \approx \boxed{12.884}$

Sample 1 2 3 3 31



when S is big, data elements are more spread out from $\bar{x}$

Sep 16-7:48 PM

How to estimate S:

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

Always

The range rule-of-thumb

Sep 16-7:55 PM

whenever Mean, Mode, and median are the same, the data dist. will be Symmetric and has bell-shape graph.

Empirical Rule

About 68% of data fall within $\bar{x} \pm S$

$\boxed{\text{About 95\% of data fall within } \bar{x} \pm 2S}$ → Usual Range

About 99.7% of data fall within $\bar{x} \pm 3S$

Exam Scores had a bell-shape dist with $\bar{x} = 82$ & $S = 6$.

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

95% Range → $\bar{x} \pm 2S = 82 \pm 2(6)$

$= 82 \pm 12$ → $\boxed{70 \text{ to } 94}$ Usual Range

2.5% Unusual | 70 | Usual 95% | 94 | 2.5% Unusual

Lisa made 90 Usual

Moe made 95 Unusual

Rahim made 68 Unusual

SG 5 ✓

99.7% Range $\bar{x} \pm 3S = 82 \pm 3(6) = 82 \pm 18$ → $\boxed{64 \text{ to } 100}$

Sep 16-8:09 PM

Z - Score Always round to 3-dec. places

SG 6

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

Z-Score measures how many standard deviations is the data element from $\bar{x}$.

It allows us to standardize data elements from different Samples.

If $-2 \le Z \le 2$ → data element is usual

If $Z < -2$ or $Z > 2$ → data element is unusual.

unusual | -2 | usual | 2 | Unusual

Sep 16-8:21 PM

Lisa got 96 on exam 1.

$\bar{x} = 88$, $S = 5$

$Z = \frac{x - \bar{x}}{S} = \frac{96 - 88}{5} = \frac{8}{5} = \boxed{1.6}$

Since $-2 \le Z \le 2$ → her Score is usual.

My Z-Score was −2.2, what was my exam Score?

$Z < -2$ → unusual test Score

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

$-2.2 = \frac{x - 88}{5}$

Cross-multiply

$x - 88 = 5(-2.2)$

$x = 88 - 11$

$\boxed{x = 77}$

Sep 16-8:26 PM

Maria makes \$5000/mo. as a nurse.

$\bar{x} = 4000$ $S = 400$

Jose 〃 \$6000/mo. as a teacher

$\bar{x} = 5000$ $S = 800$

who performs better?

$Z_{Maria} = \frac{5000 - 4000}{400} = \frac{1000}{400} = \boxed{2.5}$ unusually high

$Z_{Jose} = \frac{6000 - 5000}{800} = \frac{1000}{800} = \boxed{1.25}$ usual

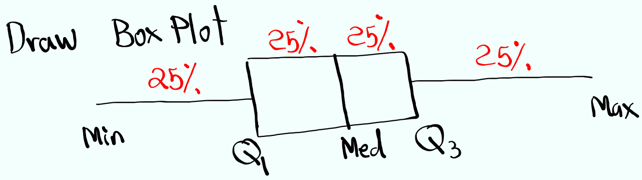
Maria

Sep 16-8:31 PM

5-Number Summary

Min Q1 Med. Q3 Max

Draw Box Plot



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

Upper Fence $= Q_3 + 1.5(IQR)$

Lower Fence $= Q_1 - 1.5(IQR)$

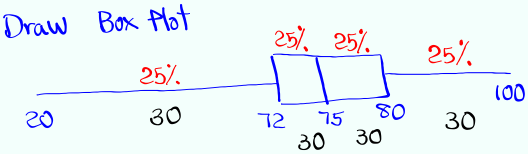
Sep 16-8:36 PM

I randomly Selected (120 exams). The 5-Number Summary of Scores were

$120 \div 4 = 30$

20	72	75	80	100
Min	Q1	Med	Q3	Max

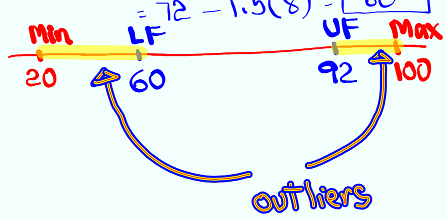
Draw Box Plot



$IQR = Q_3 - Q_1 = 80 - 72 = \boxed{8}$

Upper Fence $= Q_3 + 1.5(IQR)$
$= 80 + 1.5(8) = \boxed{92}$

Lower Fence $= Q_1 - 1.5(IQR)$
$= 72 - 1.5(8) = \boxed{60}$

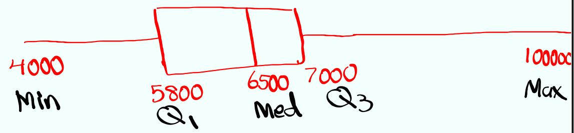


Sep 16-8:42 PM

Salaries of teachers have the following 5 - Number Summary

4000 5800 6500 7000 100000

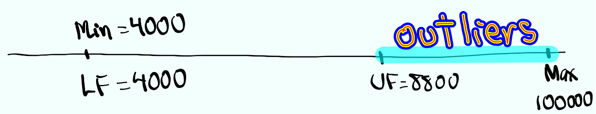
Box Plot



$IQR = Q_3 - Q_1 = 7000 - 5800 = \boxed{1200}$

Upper Fence $= Q_3 + 1.5(IQR)$
$= 7000 + 1.5(1200) = \boxed{8800}$

Lower Fence $= Q_1 - 1.5(IQR)$
$= 5800 - 1.5(1200) = \boxed{4000}$

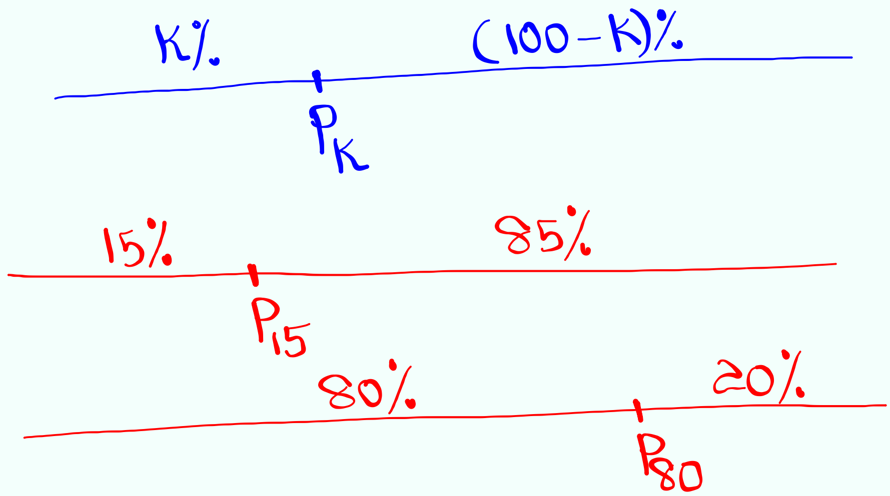


Sep 16-8:51 PM

Percentile & Percentile Ranking

Notation P_K

Data must be Sorted



Sep 16-8:59 PM

How to find P_K

1) find location $L = \frac{K}{100} \cdot n$ ← Sample Size

if L is decimal → Round-up to whole #

P_K = Lth element

if L is a whole # →

$$P_K = \frac{\text{Lth element} + \text{Next element}}{2}$$

Sep 16-9:01 PM

I randomly selected 20 exams. here are the scores

53	58	60	64	64
68	70	72	75	78
78	78	80	85	88
90	92	98	100	100

STEM Plot

```
5 | 3 8
6 | 0 4 4 8
7 | 0 2 5 8 8 8
8 | 0 5 8
9 | 0 2 8
10| 0 0
```

find P_{10}

$L = \frac{10}{100} \cdot 20 = 2$

$P_{10} = \frac{\text{2nd element} + \text{Next one}}{2}$

$= \frac{58 + 60}{2} = \boxed{59}$

10% | 90%

$P_{10} = 59$

find P_{82}

$L = \frac{82}{100} \cdot 20 = 16.4$

$L = 17$

P_{82} = 17th element

$P_{82} = 92$

82% | 18%

$P_{82} = 92$

Sep 16-9:04 PM

STEM Plot

5	3 8
6	0 4 4 8
7	0 2 5 8 8 8
8	0 5 8
9	0 2 8
10	0 0

doing Reverse

Find K when P_K is given.

$$K = \frac{\overset{\text{Below}}{B}}{n} \cdot 100$$

Round to whole %

Find K such that $P_K = 80$

Below

$$K = \frac{B}{n} \cdot 100 = \frac{12}{20} \cdot 100 = \boxed{60}$$

So $P_{60} = 80$

60% | 40%

$P_{60} = 80$

SG 6 ✓

Sep 16-9:13 PM