Statistics

Lecture 3

Feb 19-8:47 AM

SG 5

$x \rightarrow$ Data element

$n \rightarrow$ Sample Size

$\sum x \rightarrow$ Sum of data elements

$\sum$ ↑ Summation

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

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

Consider the Sample below

2 4 4 4 8

$n = 5$

Range $= 8 - 2 = 6$

Midrange $= \frac{8+2}{2} = 5$

Mode 4

Median 4

the value in the middle

(Data must be Sorted)

$\sum x = 2 + 4 + 4 + 4 + 8 = 22$

$\bar{x} = \frac{\sum x}{n} = \frac{22}{5} = 4.4$

Round to whole # $\rightarrow$ 4

Round up to whole # $\rightarrow$ 5

Sep 12-8:03 AM

$n \rightarrow$ Sample Size

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

$x \rightarrow$ Data element

$\Sigma x \rightarrow$ Sum of data elements $\quad \bar{x} = \frac{\Sigma x}{n}$

$\Sigma x^2 \rightarrow$ Square every data element, then add them

$s^2 \rightarrow$ Sample variance

$\rightarrow$ we use variance to find standard deviation

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

Variance is the sum of square of the differences between every data element and the mean then divide be $n-1$.

Sep 12-8:11 AM

Consider the Sample below

1 4 4 6 6 9

$n = 6$ Range $= 9 - 1 = 8$ Midrange $= \frac{9+1}{2} = 5$

Mode: 4 & 6 Median $= \frac{4+6}{2} = 5$

value in the middle

(Data must be Sorted)

$\Sigma x = 1 + 4 + 4 + 6 + 6 + 9 = 30$

$\Sigma x^2 = 1^2 + 4^2 + 4^2 + 6^2 + 6^2 + 9^2 = 186$

$\bar{x} = \frac{\Sigma x}{n} = \frac{30}{6} = 5$ $\qquad s^2 = \frac{n\Sigma x^2 - (\Sigma x)^2}{n(n-1)}$

$= \frac{6 \cdot 186 - 30^2}{6(6-1)} = \frac{216}{30} = 7.2$

$s^2 \geq 0$

It is never negative.

Sep 12-8:18 AM

$\bar{x}$ → Sample Mean $\quad \bar{x} = \frac{\Sigma x}{n}$

s^2 → Sample Variance $\quad s^2 = \frac{\Sigma(x-\bar{x})^2}{n-1}$

$s^2 = \frac{n\Sigma x^2 - (\Sigma x)^2}{n(n-1)}$

s → Sample Standard deviation

$s = \sqrt{s^2}$ $\quad s \geq 0$

Standard deviation is a value that indicates the deviation of data elements from the mean.

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

when s is big → data elements are more spread out from the mean.

when s is zero → all data elements are the same and equal to $\bar{x}$.

Sep 12-8:26 AM

Consider the Sample below

1 2 2 2 3

$n = 5$ $\quad$ Range = 2 $\quad$ Midrange = 2

Mode = 2 $\quad$ Median = 2

$\Sigma x = 10$ $\quad \Sigma x^2 = 22$

$\bar{x} = \frac{\Sigma x}{n} = \frac{10}{5} = \boxed{2}$

$s^2 = \frac{n\Sigma x^2 - (\Sigma x)^2}{n(n-1)}$

$= \frac{5 \cdot 22 - 10^2}{5(5-1)} = \frac{10}{20}$

$= \boxed{.5}$

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

there is not that much deviation between $\bar{x} = 2$ and data elements 1 2 2 2 3

Sep 12-8:35 AM

Consider the sample below

1 2 2 2 30

$\Sigma x = 37$ $\quad \Sigma x^2 = 913$

$\bar{x} = \frac{\Sigma x}{n} = \frac{37}{5} = 7.4$

$s^2 = \frac{n\Sigma x^2 - (\Sigma x)^2}{n(n-1)} = \frac{5 \cdot 913 - 37^2}{5(5-1)} = \frac{3196}{20} = 159.8$

$s = \sqrt{s^2} = \sqrt{159.8} \approx 12.641$

big

Since s is a big #, there is more deviation between data elements and the sample mean $\bar{x}$.

Sep 12-8:41 AM

Given $n = 10$ $\quad \Sigma x = 80$ $\quad \Sigma x^2 = 640$

Find

$\bar{x} = \frac{\Sigma x}{n} = \frac{80}{10} = 8$

$s^2 = \frac{n\Sigma x^2 - (\Sigma x)^2}{n(n-1)} = \frac{10 \cdot 640 - 80^2}{10(10-1)} = \frac{0}{90} = 0$

$s = \sqrt{s^2} = \sqrt{0} = 0$

what do you conlude about data elements?

Since $s = 0$, all data elements are the same and equal to $\bar{x} = 8$

How to estimate Sample Standard deviation:

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

the range rule-of-thumb.

Sep 12-8:48 AM

Given $n=20$ $\sum x = 1583$

$\sum x^2 = 128743$ Min $=52$ Max $=100$

Find

$\bar{x} = \frac{\sum x}{n} = \frac{1583}{20} = 79.15$ $s^2 = \frac{n\sum x^2 - (\sum x)^2}{n(n-1)} = \frac{20 \cdot 128743 - 1583^2}{20(20-1)}$

$s = \sqrt{s^2} = \sqrt{181.503}$ $= \frac{68971}{380} \approx 181.503$

$\approx 13.472 \approx 13$

Estimate s $\approx \frac{\text{Range}}{4} = \frac{100-52}{4} = \frac{48}{4} = 12$

Sep 12-8:57 AM

when mean, mode, and median are the same, data dist. will be symmetric and bell-shape looking.

Empirical Rule 68% – 95% – 99.7% Ranges

About 68% of data elements are within

$\bar{x} \pm s$

About 95% of data elements are within

Usual Range $\bar{x} \pm 2s$

About 99.7% of data elements are within

$\bar{x} \pm 3s$

Sep 12-9:08 AM

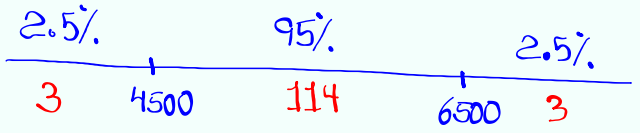
I randomly selected 120 nurses, their salaries had a bell-shape dist with $\bar{x} = \$5500/\text{mo.}$ with $S = \$500/\text{mo.}$

68% Range $\rightarrow \bar{x} \pm S = 5500 \pm 500$

68% of $120 \approx 82$ of them $\Rightarrow$ 5000 to 6000

95% Range $\rightarrow \bar{x} \pm 2S = 5500 \pm 2(500)$

$\Rightarrow$ 4500 to 6500



95% of 120 = 114

Sep 12-9:13 AM

Z-Score Round to 3-dec. places

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

Z-Scores indicates how many standard deviation is the data element from $\bar{x}$.

$-2 \le Z \le 2 \longleftrightarrow$ usual data element

$Z < -2$ or $Z > 2 \longleftrightarrow$ unusual data element

Z-Score allows us to compare data elements from different samples

It is process of standardize data elements.

Sep 12-9:19 AM

Lisa makes $4000 as a nurse while Mike makes $6000 as a Store manager. who is performing better?

Nurse → $\bar{x} = 3500$, $S = 400$

Managers → $\bar{x} = 5000$, $S = 1000$

Lisa $Z = \frac{x - \bar{x}}{S} = \frac{4000 - 3500}{400} = 1.25$ usual

Mike $Z = \frac{x - \bar{x}}{S} = \frac{6000 - 5000}{1000} = 1$ usual

I have a Z-Score of −2.25 as a Store manager, what is my salary?

$Z = \frac{x - \bar{x}}{S}$ $\quad -2.25 = \frac{x - 5000}{1000}$

Cross-Multiply $\quad -2.25(1000) = x - 5000$

Solve for x $\quad -2.25(1000) + 5000 = x$

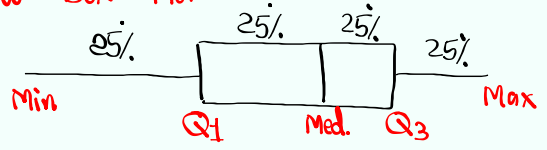
$\boxed{x = 2750}$

Sep 12-9:23 AM

5-Number Summary

Min Q_1 Med. Q_3 Max

Draw Box Plot



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

Upper Fence $= Q_3 + 1.5(IQR)$

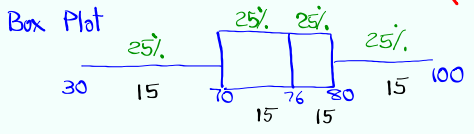
Lower Fence $= Q_1 - 1.5(IQR)$

Sep 12-9:45 AM

I randomly Selected 60 exams and here are the 5-Number Summary of Scores

30 Min | 70 Q_1 | 76 Med. | 80 Q_3 | 100 Max } 60÷4=15

Box Plot



25% below Q_1, 75% above Q_1

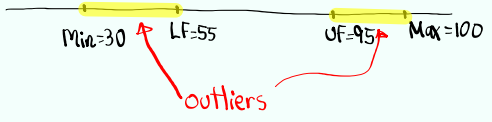
50% = Med., 50% above Med.

75% below Q_3, 25% above Q_3

$IQR = Q_3 - Q_1 = 80 - 70 = \boxed{10}$

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

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



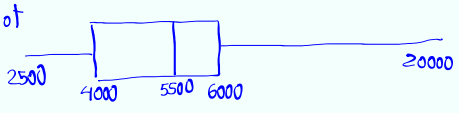
outliers

Sep 12-9:50 AM

5-Number Summary of monthly Salaries of nurses randomly Selected were

2500 4000 5500 6000 20000

Box Plot



$IQR = Q_3 - Q_1 = \boxed{2000}$

Upper Fence $= Q_3 + 1.5(IQR) = 6000 + 1.5(2000)$

$= \boxed{9000}$

Lower Fence $= Q_3 - 1.5(IQR) = 4000 - 1.5(2000)$

$= \boxed{1000}$

Outliers → 9000 to 20,000

SG 5 ✓

You can work on SG 6.

Look for my email.

Sep 12-9:59 AM

TI Instructions:

1) To clear the Screen

[Clear]

2) To quit

[2nd] [MODE]

3) To clear all lists.

[2nd] [+] [4:ClearAll Lists] [Enter]

4) To reset all lists.

[STAT] Edit [Enter]
[5:Set UpEditor]

Sep 12-10:09 AM

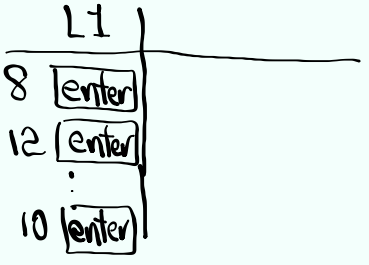
How to store a data Set in a list:

Consider the Sample below

8 12 5 20 15

10 5 14 15 10

[STAT] Edit
[1:Edit]



Let's quit & clear Screen

[2nd] [Mode] [clear]

Sep 12-10:15 AM

How to view a list.

I want to see what is in L1.

2nd 1 Enter

{8 12 5 20 ...}

go right → → →

go left ← ← ←

How to Sort a list:

STAT Edit 2:SortA(2nd 1 Enter

L1

Let's quit & clear Screen

2nd Mode Clear

Sep 12-10:19 AM

Let's view L1:

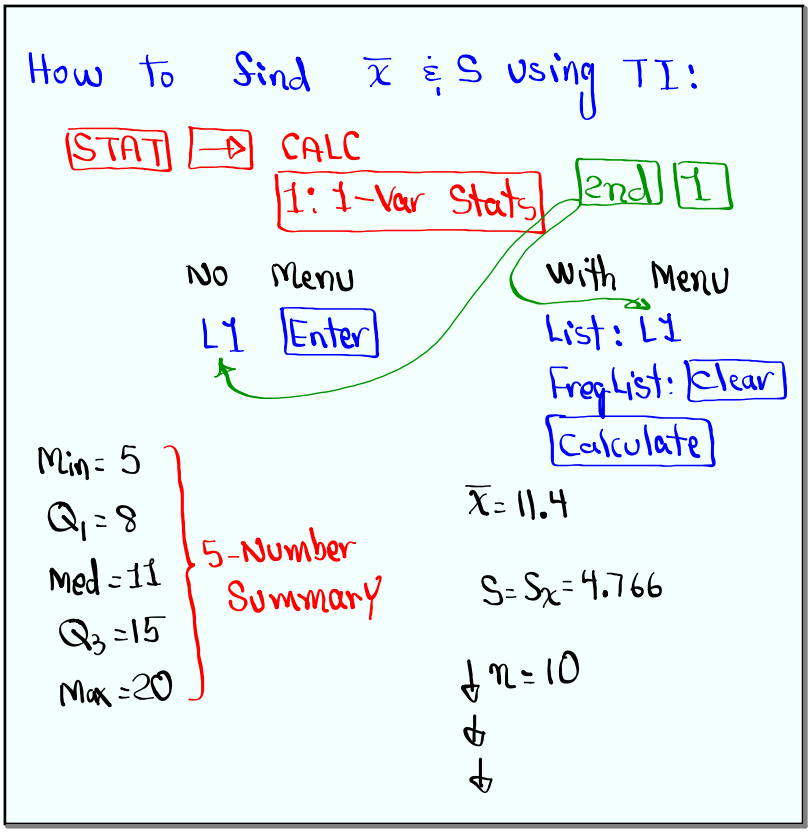
2nd 1 Enter

{5 5 8 10 ...

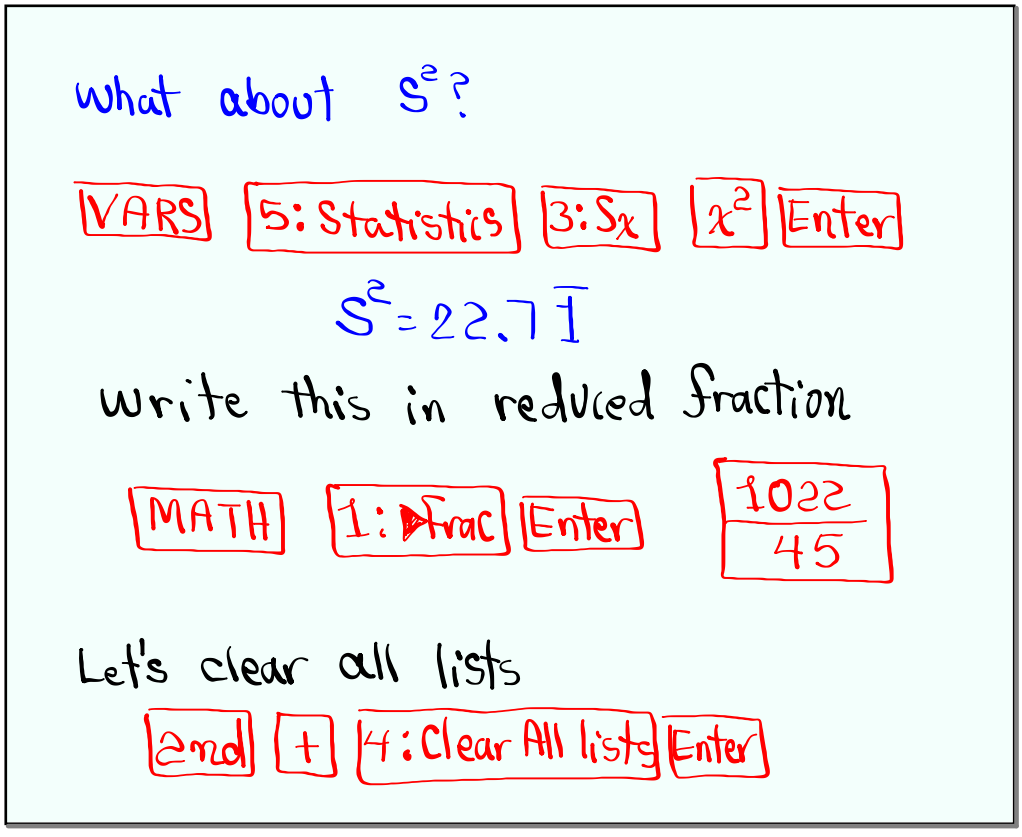
→ → → Right

Left ← ← ←

Sep 12-10:25 AM



Sep 12-10:27 AM



Sep 12-10:33 AM

I randomly selected 20 students, here are their ages:

23	27	18	20	30
35	32	25	19	40
48	50	45	38	19
24	25	50	46	44

1) Store in L1.

STAT Edit

1: Edit

L1
23 Enter
27
18
⋮
44

quit & clear Screen

2nd Mode clear

Sep 12-10:40 AM

Sort L1, view L1, then make STEM plot

STAT Edit

2: SortA(

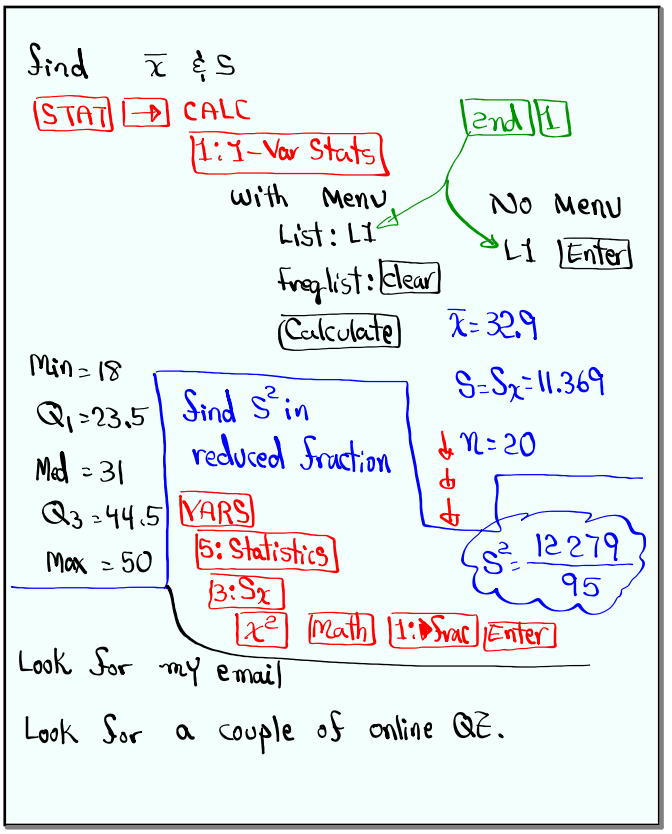
2nd 1 Enter

2nd 1 Enter

{18 19 19 20 → → →

1	8 9 9
2	0 3 4 5 5 7
3	0 2 5 8
4	0 4 5 6 8
5	0 0

Sep 12-10:45 AM



Sep 12-10:50 AM