Statistics

Lecture 3

Feb 19-8:47 AM

Consider the freq. table below

class limits	class BNDRS	class MP	class F	Cum. F	Rel. F	% F
18 - 26	17.5 - 26.5	22	7	7	.175	17.5%
27 - 35	26.5 - 35.5	31	13	20	.325	32.5%
36 - 44	35.5 - 44.5	40	12	32	.300	30.0%
45 - 53	44.5 - 53.5	49	8	40	.200	20.0%

$n = 40$

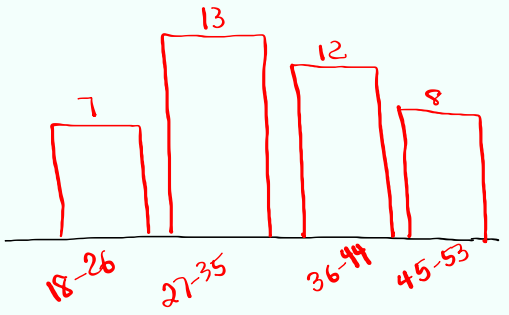
4 Rows → 4 classes

Class width = $27-18 = 36-27 = 45-36 = \boxed{9}$

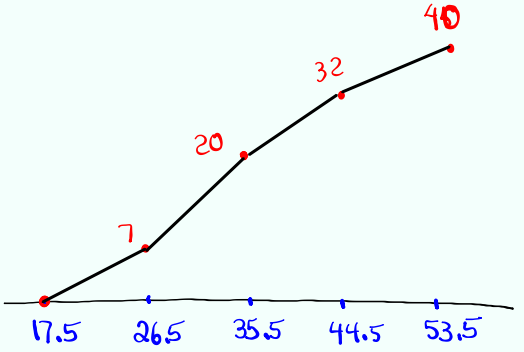
$$\text{Rel. } f = \frac{f}{n} = \frac{f}{40}$$

Jun 26-4:30 PM

Bar Graph
- Class limits
- Class F



Ogive
- Class BNDRS
- Cum. F
- Start at 0 level
- Always increasing



Jun 26-4:40 PM

Consider the Sample below

2 3 3 3 5 5 5 8

1) $n = 8$ 2) Range $= 8 - 2 = 6$

3) Midrange $= \frac{8+2}{2} = 5$ 4) Mode: 3 & 5

5) $\Sigma x = 34$ 6) $\Sigma x^2 = 170$

7) $\bar{x} = \frac{\Sigma x}{n} = \frac{34}{8}$
$= 4.25$
whole $\rightarrow 4$
1-Dec. $\rightarrow 4.3$

8) $s^2 = \frac{n\Sigma x^2 - (\Sigma x)^2}{n(n-1)}$
$= \frac{8 \cdot 170 - 34^2}{8(8-1)} = \frac{204}{56}$
≈ 3.643

Find

9) $s = \sqrt{s^2} = \sqrt{3.643} \approx 1.909$

10) Estimate $s \approx \frac{\text{Range}}{4} = \frac{6}{4} = 1.5$

Jun 26-4:46 PM

What is Variance?

It is the measure of spread of data elements from the mean.

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

Why divide by $n-1$?

It turns out if we divide by $n-1$ instead of n, we get a better estimator for population standard deviation.

Sample standard deviation:

$$S = \sqrt{S^2} \qquad S \geq 0, S^2 \geq 0$$

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

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

If $S=0$, All data elements are the same as $\bar{x}$.

Jun 26-4:57 PM

Consider the Sample below

5 5 5 5 5 5

$n=6$ $\quad \sum x = 30$ $\quad \sum x^2 = 150$

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

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

$$= \frac{6 \cdot 150 - 30^2}{6(6-1)} = \frac{0}{30} = \boxed{0}$$

$$S = \sqrt{0} = \boxed{0}$$

Jun 26-5:04 PM

Consider the Sample below

5 5 5 5 5 50

$n=6$ $\Sigma x=75$ $\Sigma x^2=2625$

$\bar{x}=\frac{\Sigma x}{n}=\frac{75}{6}=12.5$

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

$=\frac{6\cdot 2625-75^2}{6(6-1)}$

$=337.5$

$s=\sqrt{s^2}$

$=\sqrt{337.5}\approx 18.371$

Now the Sample below

5 5 5 5 5 500

$n=6$ $\Sigma x=525$ $\Sigma x^2=250125$

$\bar{x}=87.5$ $s^2=40837.5$ $s=202.083$

Jun 26-5:08 PM

Z - Score

It is a process to standardize data elements.

It tells us how many Standard deviation is the data element from the mean.

Always round to 3-decimal places.

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

why do we need it? we need it to compare data elements from different Samples.

If $-2\leq Z\leq 2$ ⇒ Data element is usual

If $Z<-2$ or $Z>2$ ⇒ unusual

Jun 26-5:15 PM

Robert got 88 on exam 1.

$\bar{x} = 80$, $S = 5$

$$Z = \frac{x - \bar{x}}{S} = \frac{88 - 80}{5} = \frac{8}{5} = 1.6$$

Since $-2 \le Z \le 2$ → usual Score

Robert got 82 on exam 2.

$\bar{x} = 74$ $S = 4$

$$Z = \frac{x - \bar{x}}{S} = \frac{82 - 74}{4} = \frac{8}{4} = 2$$

This score is on the border of being unusual.

Jun 26-5:20 PM

Lisa makes $5000/mo. as a nurse.

Jose makes $10000/mo. as a teacher.

who is doing better in their field?

Nurse → $\bar{x} = 4000$, $S = 400$

Teacher → $\bar{x} = 8000$, $S = 500$

$$Z_{Lisa} = \frac{x - \bar{x}}{S} = \frac{5000 - 4000}{400} = \frac{1000}{400} = 2.5$$

$$Z_{Jose} = \frac{x - \bar{x}}{S} = \frac{10000 - 8000}{500} = \frac{2000}{500} = 4$$

Jose

Jun 26-5:25 PM

When mean, mode, and median are the same, data dist. will be symmetric and takes a shape called bell curve.

Emperical Rule

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

" 95% of " " " $\bar{x} \pm 2S$

" 99.7% of " " " $\bar{x} \pm 3S$

Usual Range $\Rightarrow \bar{x} \pm 2S$

Jun 26-5:33 PM

I took a Survey of 80 students. the age dist. were symmetric with $\bar{x} = 32$ & $S = 6$.

68% Range: $\bar{x} \pm S = 32 \pm 6 \Rightarrow$ 26 to 38

68% of 80 = 54.4 $\approx$ 55

About 55 of them were between 26 & 38 yrs old.

Usual Range → 95% Range

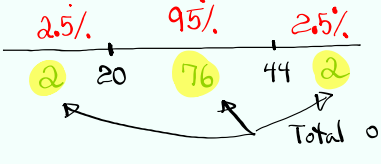
$\bar{x} \pm 2S = 32 \pm 2(6)$

$= 32 \pm 12$

$\Rightarrow$ 20 to 44

95% of 80 = .95(80) = 76

76 students were aged between 20 & 44.



Jun 26-5:37 PM

Exam 1 Scores were bell-Shape with $\bar{x} = 84$ & $S = 7$.

Find its usual Range.

95% Range → $\bar{x} \pm 2S$

$84 \pm 2(7) = 84 \pm 14 \Rightarrow$ **70 to 98**

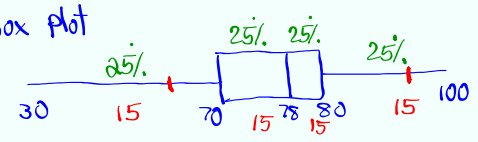
Jun 26-5:45 PM

60 exams had the following 5-Number Summary

$60 \div 4 = 15$

30	70	78	80	100
Min	Q_1	MD	Q_3	Max

Draw Box plot



How many Scored below 80? 45

What % of Score were below 80? 75%

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

Upper Fence $= Q_3 + 1.5(IQR) = 80 + 1.5(10) = 95$

Lower Fence $= Q_1 - 1.5(IQR) = 70 - 1.5(10) = 55$

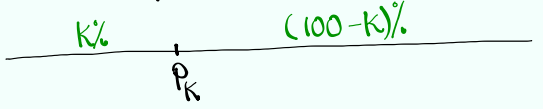
Discuss outliers. **30 to 55** OR **95 to 100**

Jun 26-5:52 PM

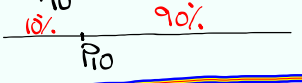
Percentile:

Data must be Sorted (Smallest → Largest)

Notation P_K



ex: P_{10}



P_{80}

How to find P_K:

1) Location $L = \frac{K}{100} \cdot n$ (n: Sample Size)

If L is decimal → Round-up

P_K = Lth element

If L is a whole #

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

Jun 26-6:14 PM

Consider the Sample below

24	25	25	28	30
32	32	35	35	35
39	40	40	43	46
46	46	48	50	52
53	56	60	60	64

1) $n = 25$

2) Range = 64 − 24 = 40

3) Estimate S

$S \approx \frac{\text{Range}}{4} = \frac{40}{4} = 10$

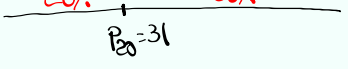
4) STEM Plot

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

5) P_{20}

$L = \frac{20}{100} \cdot 25 = 5$

$P_{20} = \frac{5\text{th } \# + 6\text{th } \#}{2} = \frac{30+32}{2} = 31$



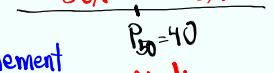
6) P_{50}

$L = \frac{50}{100} \cdot 25 = 12.5$

$L = 13$

P_{50} = 13th element

$P_{50} = 40$



Median

Jun 26-6:21 PM

Doing Reverse

Find K such that $P_K = 50$

Percentile ranking of 50

2	4558
3	0225559
4	0036668
5	0236
6	004

$$K = \frac{B}{n} \cdot 100 \quad \text{(B = \# number)}$$

Round to whole %.

$$= \frac{18}{25} \cdot 100$$

$$= 72$$

$P_{72} = 50$

72% | 28%

$P_{72} = 50$

Jun 26-6:32 PM

Consider the Stem Plot below:

5	58
6	2559
7	02355558
8	24555689
9	00235
10	00

1) $n = 28$

2) Range $= 100 - 55 = 45$

3) Estimate $S \approx \frac{\text{Range}}{4} = \frac{45}{4} = 11.25$

4) P_{30} $\quad L = \frac{30}{100} \cdot 28 = 8.4 \quad L = 9$

P_{30} = 9th element = 73

5) P_{80} $\quad L = \frac{80}{100} \cdot 28 = 22.4 \quad L = 23$

P_{80} = 23rd element = 90

6) Find K such that $P_K = 70$ (Below)

$$K = \frac{B}{n} \cdot 100 = \frac{6}{28} \cdot 100 = 21.42\ldots \approx 21$$

$P_{21} = 70$

21% | 79%

70

Jun 26-6:35 PM

Some TI instructions:

1) To clear Screen — clear

2) To quit — 2nd MODE

3) To clear all lists — 2nd + 4:ClearAllLists Enter

4) To reset all lists — STAT Edit 5:SetupEditor Enter

Jun 26-6:45 PM

How to store data elements in a list:

I wish to store the following in L1

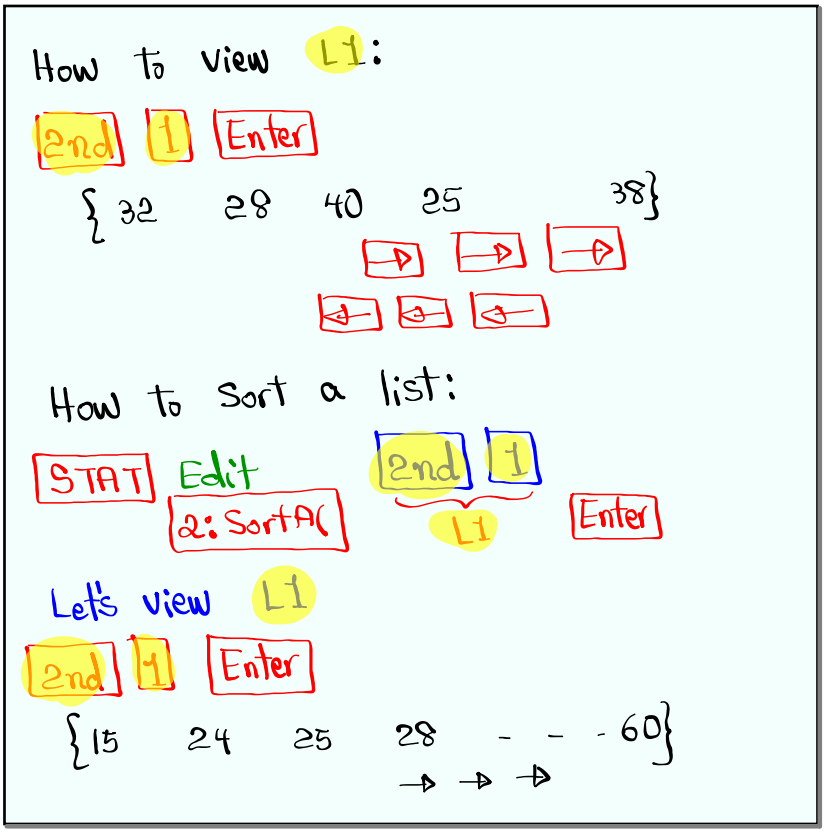
32 28 40 25 55
15 60 24 36 38

STAT Edit
1:Edit

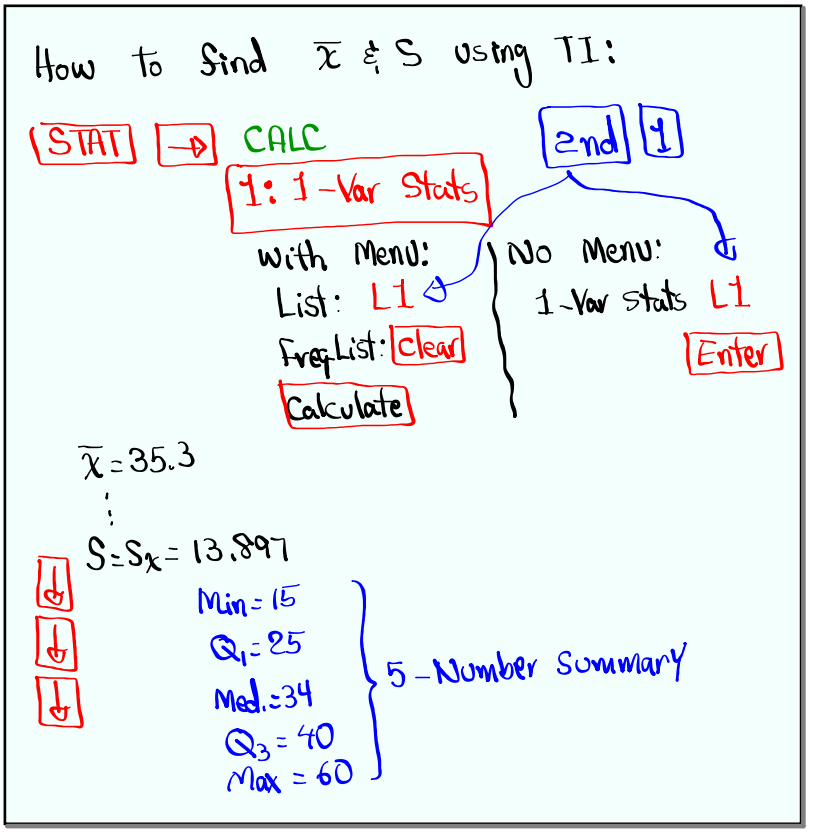
L1
32 Enter
28 "
40 "
⋮ ⋮
38 "

Quit & clear Screen
2nd MODE Clear

Jun 26-6:51 PM



Jun 26-6:55 PM



Jun 26-7:02 PM

How to find S^2:

VARS → VARS → 5: Statistics → $\bar{x}y$ → 3: S_x → x^2 → Enter

$S^2 = 193.1222\ldots$

How to convert to reduced fraction

MATH → Math → 1: ▶Frac → Enter

$\frac{17381}{90}$

Jun 26-7:11 PM