Statistics

Lecture 3

Feb 19-8:47 AM

Consider the sample below given in Stem Plot:

2	0 35
3	0 1 2 3 6 8
4	0 0 2 4 5 5 8 9
5	2 3 5 6 7
6	0 2 5

1) $n = 25$

2) Range $= 65 - 20 = 45$

3) Midrange $= \frac{65+20}{2} = \frac{85}{2} = 42.5$

4) Mode $= 40$ & 45

5) what % of data elements fall below 40?

$\frac{9}{25} \cdot 100 = 36$ (36%)

6) find class width if we wish to have a freq. table with

a) 3 classes

$CW = \frac{\text{Range}}{3} = \frac{45}{3} = 15$

$CW = 16$

b) 4 classes

$CW = \frac{\text{Range}}{4} = \frac{45}{4} = 11.25$

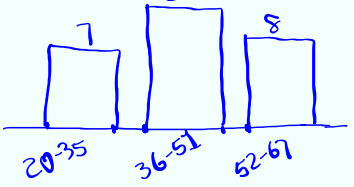
$CW = 12$

Jun 26-4:28 PM

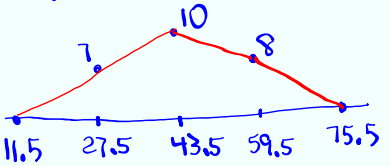
Let's make a Freq. table with 3 classes. CW=16

class limits	class BNDRS	class MP	class F	Cum. F	Rel. F	% F
20-35	19.5-35.5	27.5	7	7	.28	28%
36-51	35.5-51.5	43.5	10	17	.40	40%
52-67	51.5-67.5	59.5	8	25	.32	32%

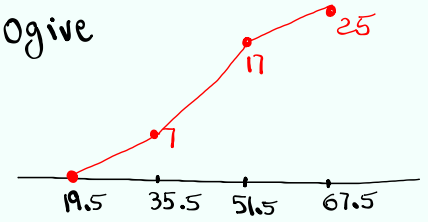
Bar chart



Freq. Polygon



Ogive



SG 3 & 4

Jun 26-4:36 PM

Consider the Sample below

2 3 3 3 5

6 6 8 8 10

1) $n = 10$

2) Range $= 10 - 2 = 8$

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

4) Mode $= 3$

5) $\Sigma x = 54$

6) $\Sigma x^2 = 356$

7) $\bar{x} = \frac{\Sigma x}{n} = \frac{54}{10} = 5.4$ Sample Mean

8) $s^2 = \frac{n\Sigma x^2 - (\Sigma x)^2}{n(n-1)} = \frac{10(356) - 54^2}{10(10-1)}$ Sample Variance

$= \frac{644}{90} = 7.1\overline{5} \approx 7.156$

646 ÷ 90 Math 1:▶Frac Enter

$\frac{322}{45}$

9) $s = \sqrt{s^2} = \sqrt{7.156}$

≈ 2.675

Sample Standard deviation

Standard Deviation $= \sqrt{\text{Variance}}$

Jun 26-4:51 PM

What is Standard deviation?

It is a non-negative numerical value that indicates how data elements are spread out with respect to the mean.

If S is Small $\Rightarrow$ Data elements are close to the mean.

If S is large $\Rightarrow$ Data elements are more spread out from the mean.

If S is Zero $\Rightarrow$ No deviation from the mean. All data elements are equal to $\bar{x}$.

Jun 26-5:01 PM

Consider the Sample below

6 6 6 6 6

$n = 5$, $\sum x = 30$, $\sum x^2 = 180$

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

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

$$= \frac{5 \cdot 180 - 30^2}{5(5-1)} = \frac{0}{20} = \boxed{0}$$

$$s = \sqrt{s^2} = \sqrt{0} = \boxed{0}$$ ← Since $S = 0$, All data elements are equal to $\bar{x} = 6$

Jun 26-5:06 PM

Whenever Mean = Mode = Median, data dist. will be Symmetric & Bell-Shape.

Empirical Rule:

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

About 95% of data elements fall within $\bar{x} \pm 2s$

Usual Range

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

Jun 26-5:11 PM

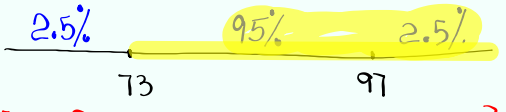
I randomly selected 120 exams. Scores had a bell-shape dist. with $\bar{x} = 85$ and $s = 6$

68% Range $\Rightarrow \bar{x} \pm s = 85 \pm 6$

$\Rightarrow$ 79 to 91

95% Range $\Rightarrow \bar{x} \pm 2s = 85 \pm 2(6)$

Usual Range $\Rightarrow$ 73 to 97



What % of Scores were above 73?

97.5%

How many Scores were above 73?

97.5% • 120 = 117

Jun 26-5:17 PM

How to compare data elements from different samples:

1) we standardize Score

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

2) Compare Z- Scores.

When $-2 \le Z \le 2 \Rightarrow$ usual

$Z < -2$ or $Z > 2 \Rightarrow$ unusual

Jun 26-5:24 PM

John makes $4000/mo. as a Sales person.

Mary makes $6500/mo. as a nurse.

Who is doing better?

Sales $\rightarrow \bar{x} = 3000$, $S = 400$

John $Z = \frac{4000 - 3000}{400} = \frac{1000}{400} = 2.5$

Nurses $\rightarrow \bar{x} = 6000$, $S = 500$

Mary $Z = \frac{6500 - 6000}{500} = \frac{500}{500} = 1$

Jun 26-5:27 PM

Suppose $\bar{x} = 125$, $S = 15$ $\quad Z = \frac{x - \bar{x}}{S}$

1) Find Z-Score for data element 160.

$Z = \frac{160 - 125}{15} = \frac{35}{15} \approx \boxed{2.333}$ unusual

2) Find the x data element if Z-Score is -2.4.

$-2.4 = \frac{x - 125}{15}$ Cross-Multiply

$x - 125 = -2.4(15)$

$x = -36 + 125$

$\boxed{x = 89}$

SG 5 Soon → SG 6

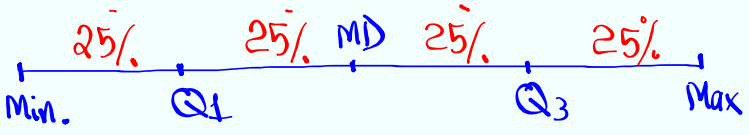
Jun 26-5:32 PM

5-Number Summary

Min Q_1 Median Q_3 Max

Q_1: First Quartile

Q_3: Third Quartile

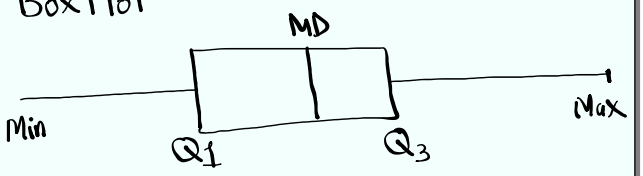


About 25% fall below Q1, and 75% fall above Q1.

About 75% fall below Q_3 and 25% fall above Q_3.

Jun 26-5:53 PM

1) Draw Box Plot



2) Find IQR (Inter-Quartile-Range)

3) Find upper fence & Lower fence

Upper fence $= Q_3 + 1.5(IQR)$

Lower fence $= Q_1 - 1.5(IQR)$

4) Identify outliers.

outliers $>$ upper fence OR

outliers $<$ Lower fence

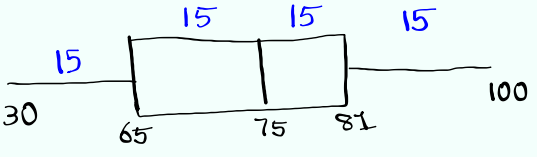
Jun 26-5:56 PM

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

$60 \div 4 = 15$

30	65	75	81	100
Min	Q_1	MD	Q_3	Max

1) Box Plot



2) $IQR = Q_3 - Q_1 = 81 - 65 = \boxed{16}$

3) upper fence $= Q_3 + 1.5(IQR) = 81 + 1.5(16) = \boxed{105}$

4) Lower fence $= Q_1 - 1.5(IQR) = 65 - 1.5(16) = \boxed{41}$

5) Discuss outliers:

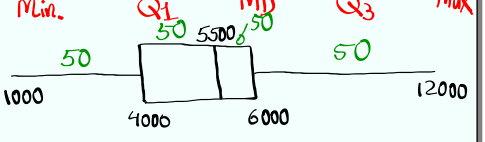
Jun 26-6:01 PM

I randomly Selected 200 nurses. Below are the 5-Number Summary of their monthly Salaries.

1000 Min. | 4000 Q1 | 5500 MD | 6000 Q3 | 12000 Max

200÷4 = 50

Box Plot

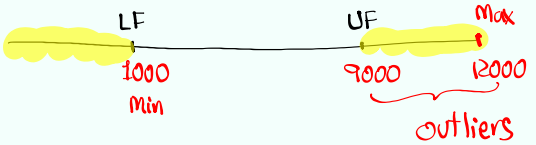


$IQR = Q_3 - Q_1 = 6000 - 4000 = 2000$

Upper Fence $= Q_3 + 1.5(IQR) =$ \$9000

Lower Fence $= Q_1 - 1.5(IQR) =$ \$1000

Discuss outliers \$9000 – \$12000



Jun 26-6:08 PM

I randomly selected 40 exams, and here are the Scores

56	58	60	62	65
68	69	70	73	75
75	75	76	78	79
79	79	80	82	83
85	85	85	88	90
92	93	93	94	94
95	95	95	96	97
98	99	100	100	100

1) $n = 40$

2) Range $= 100 - 56 = 44$

3) Midrange $= \frac{100+56}{2} = 78$

4) Estimate S

$S \approx \frac{\text{Range}}{4} = \frac{44}{4} = 11$

5) STEM Plot

5	68
6	02589
7	03555689999
8	0235558
9	023344555689
10	000

6) what % of data elements are below 70?

$\frac{7}{40} \cdot 100 = 17.5 \approx 18\%$

18% | 82% | 70

7) what % are below 90?

$\frac{24}{40} \cdot 100$ (60%)

60% | 40% | 90

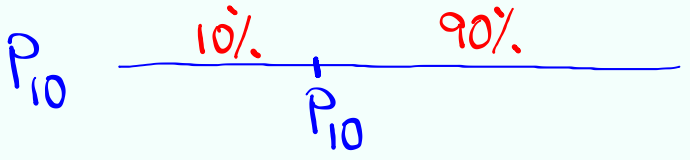
Jun 26-6:19 PM

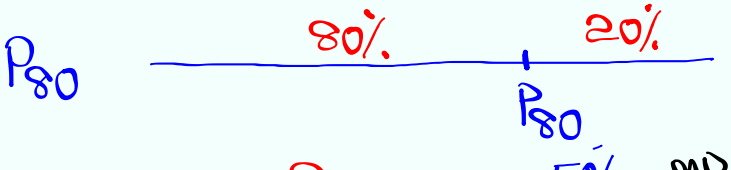
Percentile

Data must be Sorted

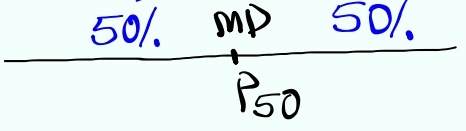
Notation P_K

K% fall below it, (100−K)% fall above it.





Median $= P_{50}$



Jun 26-6:31 PM

Find P_{20}

How to Find P_K

$L = \frac{K}{100} \cdot n$

Location

If L is decimal → Round-up P_K = Lth element

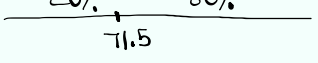
If L is a whole # $P_K = \frac{\text{Lth element} + \text{Next one}}{2}$

5	68
6	02589
7	0355568999
8	0235558
9	023344555 6789
10	000

P_{20}

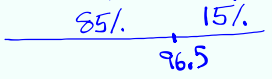
$L = \frac{20}{100} \cdot 40 = 8$

$P_{20} = \frac{8\text{th} + 9\text{th}}{2} = \frac{70+73}{2} = \boxed{71.5}$



Find P_{85}

$L = \frac{85}{100} \cdot 40 = 34$



$P_{85} = \frac{34\text{th element} + \text{Next one}}{2} = \frac{96+97}{2} = \boxed{96.5}$

Jun 26-6:35 PM

Doing Reverse:

Find K when P_K is given.

$$K = PR = \frac{B}{n} \cdot 100$$

B: Below P_K

n: Sample Size

Round to whole %.

Find K such that $P_K = 80$

Below

$$K = \frac{17}{40} \cdot 100 = 42.5 \approx 43$$

5	68
6	02589
7	0355568999
8	0235558
9	0233445556789
10	000

43% | 57%

$P_{43} = 80$

Jun 26-6:42 PM

Consider the Stem Plot below

1	89
2	03588
3	02559
4	23448
5	04

1) $n = 20$

2) Find P_{15}

$$L = \frac{15}{100} \cdot 20 = 3$$

$$P_{15} = \frac{3\text{rd} + 4\text{th}}{2} = \frac{20+23}{2} = 21.5$$

3) Find P_{94}

P_{94} = 19th element

$$L = \frac{94}{100} \cdot 20 = 18.8 \quad L = 19 \quad P_{94} = 50$$

4) Find K such that $P_K = 32$ (Below)

$P_{40} = 32$

40% | 60%

$P_{40} = 32$

$$K = \frac{B}{n} \cdot 100 = \frac{8}{20} \cdot 100 = 40$$

Jun 26-6:48 PM

Class QZ 2

Consider the Sample below

2	3	3	5
8	9	10	12

1) Range = $12 - 2 = \boxed{10}$

2) Midrange = $\frac{12+2}{2} = \boxed{7}$

3) $\sum x = 2+3+3+5+8+9+10+12 = \boxed{52}$

4) $\sum x^2 = 2^2+3^2+3^2+5^2+8^2+9^2+10^2+12^2 = \boxed{436}$

Jun 26-7:02 PM