Statistics

Lecture 2

Feb 19-8:47 AM

Consider the table below

class limits	class BNDRS	class MP	class F	Cum. F	Rel. F	% F
18 – 26	17.5 – 26.5	22	5	5	.156	15.6%
27 – 35	26.5 – 35.5	31	8	13	.250	25.0%
36 – 44	35.5 – 44.5	40	12	25	.375	37.5%
45 – 53	44.5 – 53.5	49	7	32	.219	21.9%

1) 4 classes

2) CW = 27 − 18 = 9
= 36 − 27 = 9
= 45 − 36 = 9

3) class MP = $\frac{\text{+class limits}}{2}$
= $\frac{18+26}{2}$ = 22

4) n = 32

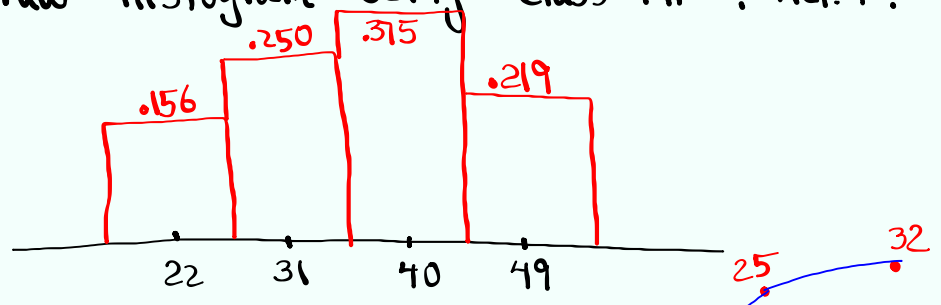
5) Rel.F = $\frac{f}{n} = \frac{f}{32}$

6) what % of data elements are between 27 & 44, inclusive?

25.0% + 37.5%
= 62.5%

Jan 8-4:30 PM

Draw histogram using class MP & Rel. F.

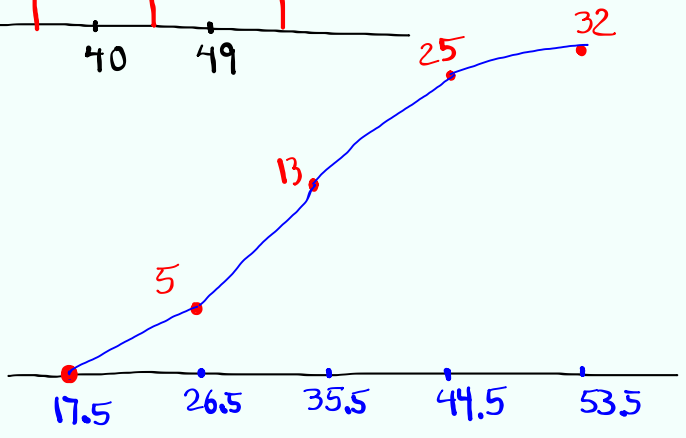


Draw ogive.

Class BNDRS

Cum. F.

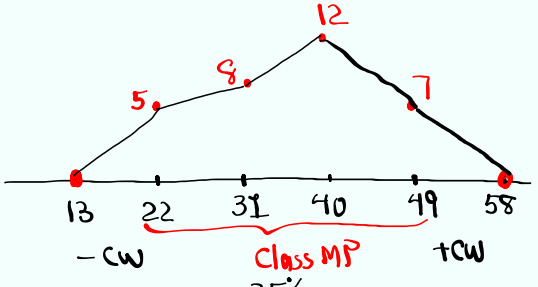
Start at 0.



Jan 8-4:43 PM

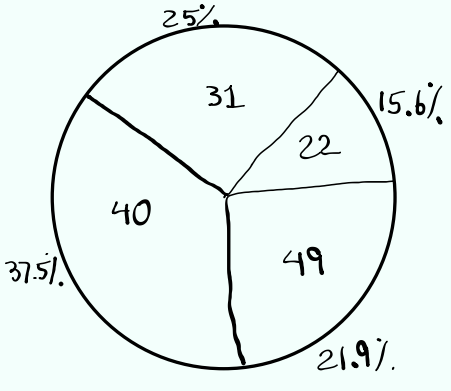
Draw Freq. Polygon.

- Class MP
- Class F
- extra MP, one on each side
- Start & stop at 0 level.



Pie chart

- Circle
- class MP to name each slice
- % F for slice size



Jan 8-4:48 PM

I randomly selected 28 exams, here are the scores:

52 58 60 65 65 68
70 72 75 75 75 78
79 80 83 86 86 86
86 88 90 92 93 95
98 100 100 100

1) $n = 28$
2) Min. = 52, Max = 100
3) Range = Max − Min = 100 − 52 = 48
4) Midrange = $\frac{\text{Max} + \text{Min}}{2} = \frac{100+52}{2} = 76$
5) Mode = 86

Find class width if we wish to have a freq. table with

6) 3 classes: $CW = \frac{\text{Range}}{3} = \frac{48}{3} = 16$, CW = 17 ✓
7) 4 classes: $CW = \frac{\text{Range}}{4} = \frac{48}{4} = 12$, CW = 13
8) 5 classes: $CW = \frac{\text{Range}}{5} = \frac{48}{5} = 9.6$, CW = 10

Make STEM plot (Data must be sorted)

5	28
6	0558
7	0255589
8	0366668
9	02358
10	000

what % of scores are below 90?

$\frac{20}{28} \cdot 100 \approx 71\%$ ($n = 28$)

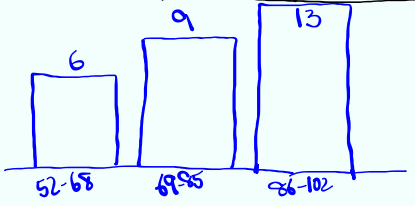
20 scores were below 90

Jan 8-4:54 PM

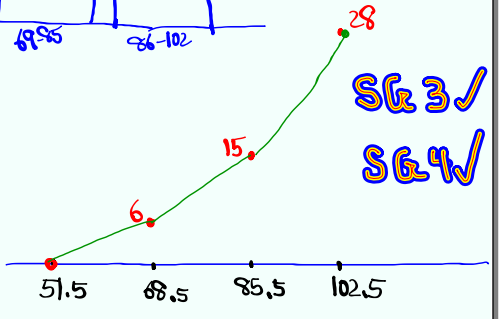
Make a freq. table with 3 classes.

Class limits	Class BNDRS	Class MP	Class F	Cum. F	Rel. F	% F
52 – 68	51.5 – 68.5	60	6	6	.214	21.4%
69 – 85	68.5 – 85.5	77	9	15	.321	32.1%
86 – 102	85.5 – 102.5	94	13	28	.464	46.4%

Bar chart
class limits
class F



Ogive
– class BNDRS
– Cum. F
– Start at 0 level.



SG 3 ✓
SG 4 ✓

Jan 8-5:08 PM

S& 5 – 8

x → Data elements

$\sum x$ → Sum of data elements

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

$\bar{x} = \frac{\sum x}{n}$ ← Sample Size

Consider the Sample below

1 3 3 3 8

Mode = 3

Range = 8 − 1 = 7

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

$n = 5$

$\sum x = 1+3+3+3+8 = 18$

$\bar{x} = \frac{\sum x}{n} = \frac{18}{5} = \boxed{3.6}$

Jan 8-5:21 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) $\sum x = 2+3+3+3+5+5+5+8 = \boxed{34}$

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

Jan 8-5:27 PM

$x \rightarrow$ Data element

$n \rightarrow$ Sample Size

$\bar{x} \rightarrow$ Sample Mean

$s^2 \rightarrow$ Sample Variance

$s^2 \geq 0$

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

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

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

Consider the Sample below

2, 3, 4, 4, 7

$n = 5$

$\sum x = 2+3+4+4+7 = 20$

$\sum x^2 = 2^2+3^2+4^2+4^2+7^2 = 94$

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

$= \frac{(2-4)^2+(3-4)^2+(4-4)^2+(4-4)^2+(7-4)^2}{5-1}$

$= \frac{4+1+9}{4} = \frac{14}{4} = \frac{7}{2} = \boxed{3.5}$

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

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

$= \frac{5 \cdot 94 - 20^2}{5(5-1)}$

$= \frac{70}{20} = \frac{7}{2}$

$= \boxed{3.5}$

Jan 8-5:31 PM

Given $n = 8$, $\sum x = 48$, $\sum x^2 = 288$

Find

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

$= \frac{48}{8}$

$= \boxed{6}$

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

$= \frac{8 \cdot 288 - 48^2}{8(8-1)}$

$= \frac{0}{56} = \boxed{0}$

Sample Standard deviation s

$s = \sqrt{s^2}$

$s \geq 0$

Mean $\bar{x} = 6$

Variance $s^2 = 0$

$s = \sqrt{0} = 0$

Jan 8-5:42 PM

Consider the Sample below

2 3 5 5 6
7 8 8 9 10

1) $n = 10$

2) Range $= 10 - 2 = 8$

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

4) Mode $= 5$ & 8

5) $\Sigma x = 2+3+5+5+6+7+8+8+9+10 = \boxed{63}$

6) $\Sigma x^2 = 2^2+3^2+5^2+5^2+6^2+7^2+8^2+8^2+9^2+10^2 = \boxed{457}$

7) $\bar{x} = \frac{\Sigma x}{n} = \frac{63}{10} = \boxed{6.3}$ (Sample Mean)

8) $s^2 = \frac{n\Sigma x^2 - (\Sigma x)^2}{n(n-1)} = \frac{10 \cdot 457 - 63^2}{10(10-1)} = \frac{601}{90} = 6.6\overline{7} \approx \boxed{6.678}$ (Sample Variance)

9) s Sample Standard deviation $= \sqrt{s^2} = \sqrt{6.678} \approx \boxed{2.584}$

How to estimate Sample Standard deviation:

$s \approx \frac{\text{Range}}{4}$ the range rule-of-thumb.

Jan 8-6:02 PM

Given $n = 10$, $\Sigma x = 78$, $\Sigma x^2 = 660$,

Min $= 4$, Max $= 12$.

1) Range $= 8$

2) Midrange $= 8$

3) $\bar{x} = \frac{\Sigma x}{n} = \frac{78}{10} = 7.8 \approx \boxed{8}$

4) $s^2 = \frac{n\Sigma x^2 - (\Sigma x)^2}{n(n-1)} = \frac{10 \cdot 660 - 78^2}{10(9)} = \frac{516}{90} \approx \boxed{5.733}$

5) $s = \sqrt{s^2} = \sqrt{5.733} \approx 2.394 \approx \boxed{2}$

6) Estimate $s \approx \frac{\text{Range}}{4} = \frac{8}{4} = \boxed{2}$

Jan 8-6:12 PM

What is Standard Deviation?

It is a numerical value that indicates how data elements are spread out from the mean.

If S is Small ⇒ Data elements are close to $\bar{x}$.

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

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

No deviation from $\bar{x}$.

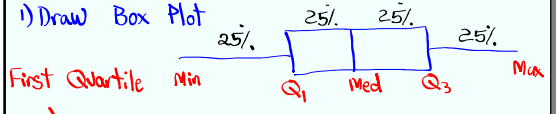
Jan 8-6:19 PM

5 – Number Summary:

Min. Q_1 Med. Q_3 Max

Data is Sorted

1) Draw Box Plot



Q_1 25% below & 75% above

Q_3 75% below & 25% above

Med. 50% below & 50% above

3rd Quartile

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

Upper Fence = $Q_3 + 1.5(IQR)$

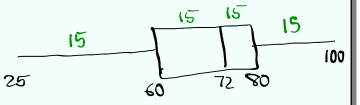
Lower Fence = $Q_1 - 1.5(IQR)$

Jan 8-6:24 PM

I randomly selected 60 exams, here are the 5-number summary of scores: (60÷4=15)

25 (Min.) 60 (Q_1) 72 (Med) 80 (Q_3) 100 (Max)

1) Draw Box Plot

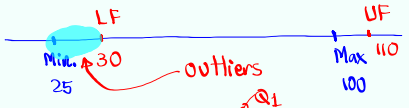


2) $IQR = Q_3 - Q_1 = 80 - 60 = \boxed{20}$

3) Upper Fence $= Q_3 + 1.5(IQR) = 80 + 1.5(20) = \boxed{110}$

4) Lower Fence $= Q_1 - 1.5(IQR) = 60 - 1.5(20) = \boxed{30}$

5) Discuss outliers



6) what % did more than 60? (Q1) 75%

7) what %: did less than 80? (Q_3) 75%

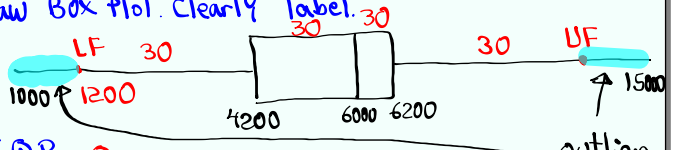
Jan 8-6:33 PM

120÷4=30

120 nurses were randomly selected. 5-Number Summary of their (monthly) Salaries were

1000 4200 6000 6200 15000

1) Draw Box Plot. Clearly label.



2) $IQR = Q_3 - Q_1 = 2000$

3) Upper fence $= Q_3 + 1.5(IQR) = 6200 + 1.5(2000) = \boxed{9200}$

4) Lower fence $= Q_1 - 1.5(IQR) = 4200 - 1.5(2000) = \boxed{1200}$

5) Discuss outliers 1000 - 1200 and 9200 - 15000

Jan 8-6:42 PM

Consider the STEM plot below for ages of randomly selected students

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

1) $n = 27$

2) Range $= 62 - 18 = 44$

3) Midrange $= \frac{62+18}{2} = 40$

4) Mode $= 35$

5) Find class width if we wish to have 4 classes

$CW = \frac{\text{Range}}{4} = \frac{44}{4} = 11$ $\boxed{CW = 12}$

6) Estimate S

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

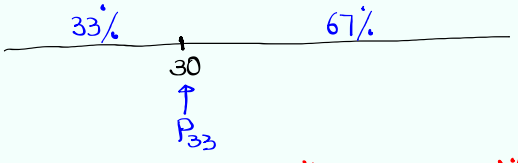
Jan 8-6:52 PM

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

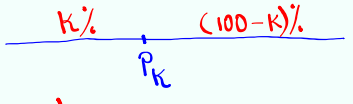
1) How many had an age below 30? 9

2) what % of them were below 30?

$\frac{9}{27} \cdot 100 \approx \boxed{33\%}$



Percentile P_K



Data must be Sorted

Median P_{50}

$P_{25} \approx Q_1$

$P_{75} \approx Q_3$

Jan 8-6:58 PM

How to find P_K

1) Find location

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

If decimal → Round-up

P_K = Lth element

If whole #

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

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

$n = 27$

Find P_{10}

$L = \frac{10}{100} \cdot 27 = 2.7$

$L = 3$

P_{10} = 3rd element

$P_{10} = 19$

10% 90%

19

Find P_{60}

$L = \frac{60}{100} \cdot 27 = 16.2$

$L = 17$

P_{60} = 17th element = 39

60% 40%

39

Jan 8-7:06 PM

$n = 30$

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

Find P_{20}

$L = \frac{20}{100} \cdot 30 = 6$

$P_{20} = \frac{\text{6th element} + \text{Next element}}{2}$

$= \frac{23 + 25}{2} = 24$

20% 80%

24

Find Median = P_{50}

$L = \frac{50}{100} \cdot 30 = 15$

$P_{50} = \frac{\text{15th element} + \text{Next one}}{2}$

$= \frac{35 + 38}{2} = 36.5$

Jan 8-7:13 PM