Statistics

Lecture 3

Feb 19-8:47 AM

Class Quiz 1

Complete the chart below then find

class limits	class MP	class F	Cum. F
12 – 20	16	5	5
21 – 29	25	8	13
30 – 38	34	7	20

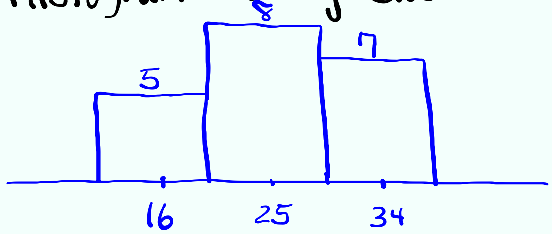
1) class width

$\boxed{9}$ ✓

2) Sample size

$\boxed{20}$

3) Draw histogram using class MP & class F.



Mar 14-8:03 AM

SG 5-8

$x \to$ Data element

$\sum x \to$ Sum of data elements

$n \to$ Sample Size

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

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

Ex: Consider the Sample below

2, 3, 5, 5, 10

1) $n = 5$

Mode = 5

2) $\sum x = 2 + 3 + 5 + 5 + 10 = 25$

Range = Max − Min = $10 - 2 = \boxed{8}$

3) $\bar{x} = \frac{\sum x}{n} = \frac{25}{5} = \boxed{5}$

Midrange = $\frac{\text{Max} + \text{Min}}{2} = \frac{10+2}{2} = \frac{12}{2} = \boxed{6}$

Mar 14-8:20 AM

$x \to$ Data element

$n \to$ Sample Size

$\bar{x} \to$ Sample Mean

$s^2 \to$ Sample Variance $s^2 \geq 0$

Variance is a measure of the Spread between data element x and Sample mean $\bar{x}$.

Variance is the $\boxed{\text{average}}$ of the Squared of differences between data elements x and Sample mean $\bar{x}$

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

n data elements

we divide by $n-1$ instead of n to get a closer value to the Population Variance.

from Last example

2, 3, 5, 5, 10 $\bar{x} = 5$, $n = 5$

$$s^2 = \frac{(2-5)^2 + (3-5)^2 + (5-5)^2 + (5-5)^2 + (10-5)^2}{5-1}$$

$$= \frac{(-3)^2 + (-2)^2 + (0)^2 + (0)^2 + 5^2}{4} = \frac{9 + 4 + 0 + 0 + 25}{4}$$

$$= \frac{38}{4} = \boxed{\frac{19}{2}} = \boxed{9.5}$$

Mar 14-8:24 AM

Consider the Sample below

1 3 3 5 5 13

1) $n = \boxed{6}$

2) $\sum x = 1+3+3+5+5+13 = \boxed{30}$

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

4) Mode $\boxed{3 \,\&\, 5}$

5) Range $= 13 - 1 = \boxed{12}$

6) Midrange $= \frac{13+1}{2} = \boxed{7}$

7) $$S^2 = \frac{\sum(x-\bar{x})^2}{n-1} = \frac{(1-5)^2+(3-5)^2+(3-5)^2+(5-5)^2+(5-5)^2+(13-5)^2}{6-1}$$

$$= \frac{(-4)^2+(-2)^2+(-2)^2+0^2+0^2+8^2}{5} = \frac{16+4+4+0+0+64}{5}$$

$$= \boxed{\frac{88}{5}} = \boxed{17.6}$$

When S^2 is Small → Spread from $\bar{x}$ is Small

" " " Large → " " " large

" " " Zero → there is no Spread from the mean. All data elements are equal to $\bar{x}$.

Mar 14-8:34 AM

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

with Some algebra ⟹ $$\boxed{S^2 = \frac{n\sum x^2 - (\sum x)^2}{n(n-1)}}$$

Consider the Sample below

1 3 3 5 9

$\sum x = 1+3+3+5+9 = \boxed{21}$

$\sum x^2 = 1^2+3^2+3^2+5^2+9^2 = \boxed{125}$

$$\bar{x} = \frac{\sum x}{n} = \frac{21}{5} = \boxed{4.2}$$

$$S^2 = \frac{n\sum x^2 - (\sum x)^2}{n(n-1)} = \frac{5 \cdot 125 - 21^2}{5(5-1)} = \frac{184}{20} = \boxed{9.2}$$

Mar 14-8:45 AM

Sample standard deviation

$\underbrace{\text{Sample standard deviation}}_{S}$

$S = \sqrt{S^2}$

$S \geq 0$

If S is small ⇒ Data elements are close to $\bar{x}$. Less spread from $\bar{x}$.

If S is large ⇒ Data elements are spread out from $\bar{x}$

If S is zero ⇒ No spread between data elements and $\bar{x}$

$\Rightarrow x = \bar{x}$

Mar 14-8:50 AM

Consider the Sample below

2 3 3 5

5 8 8 10

1) $n = \boxed{8}$

2) Mode $= \boxed{3, 5, 8}$

3) Range $= 10 - 2 = \boxed{8}$

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

5) $\sum x = \boxed{44}$

6) $\sum x^2 = \boxed{300}$

7) $\bar{x} = \frac{\sum x}{n} = \frac{44}{8} = \boxed{5.5}$

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

$= \frac{8 \cdot 300 - 44^2}{8(8-1)}$

$= \frac{464}{56} = \frac{58}{7} \approx \boxed{8.286}$

9) $S = \sqrt{S^2}$

$= \sqrt{8.286}$

$\approx \boxed{2.878}$

Mar 14-8:54 AM

How to $\boxed{\text{estimate}}$ $\underbrace{\text{Sample Standard deviation}}_{S}$

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

the range rule-of-thumb

ex: we have a Sample with min. 20 and max. 80.

1) Range = Max − Min = $\boxed{60}$

2) Midrange = $\frac{\text{Max}+\text{Min}}{2}$ = $\boxed{50}$

3) Estimate $S \approx \frac{\text{Range}}{4} = \frac{60}{4} = \boxed{15}$

4) Estimate S^2

$15^2 = \boxed{225}$

Mar 14-9:05 AM

Empirical Rule:

This is best when data distribution is symmetric.

This happens when mean, mode, and median are all the Same or almost equal.

68% Range → $\bar{x} \pm S$

$\boxed{95\% \text{ Range} \rightarrow \bar{x} \pm 2S}$ Usual Range

99.7% Range → $\bar{x} \pm 3S$

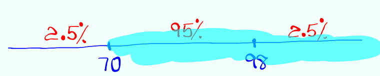
ex: Exam Scores had a symmetric dist. with $\bar{x} = 84$ and $S = 7$.

68% Range ⇒ $\bar{x} \pm S = 84 \pm 7$ ⇒ $\boxed{77 \text{ to } 91}$

95% Range ⇒ $\bar{x} \pm 2S = 84 \pm 2(7) = 84 \pm 14$

⇒ $\boxed{70 \text{ to } 98}$ Usual Range

99.7% Range ⇒ $\bar{x} \pm 3S = 84 \pm 3(7)$ ⇒ $\boxed{63 \text{ to } 105}$



100% − 95% = 5% , 5% ÷ 2 = 2.5%

what % of exam Scores are at least 70?

95% + 2.5% $\boxed{97.5\%}$

Mar 14-9:24 AM

I randomly selected 120 nurses, their Salaries had a symmetric dist with

$\bar{x} = \$7500$ and $S = \$250$.

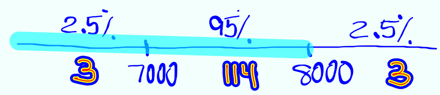
1) Usual Range $\quad \bar{x} \pm 2S = 7500 \pm 2(250)$

95% Range $\quad \Rightarrow \boxed{7000 \text{ to } 8000}$

2) How many had a usual Salary?

95% of 120 = $.95(120) = \boxed{114}$

3) How many had a Salary of at most $8000?



97.5% of 120 = $.975(120) = \boxed{117}$

SG 5 ✓

Mar 14-9:35 AM

5 - Number Summary

Data must be Sorted from Smallest to largest

First Quartile

Min. $\quad Q_1 \quad$ Median $\quad Q_3 \quad$ Max

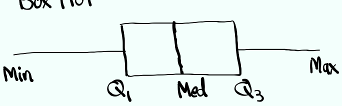
$Q_1 \rightarrow$ 25% below it, 75% above it.

Median $\rightarrow$ 50% below it, 50% above it.

$Q_3 \rightarrow$ 75% below it, 25% above it.

Third Quartile

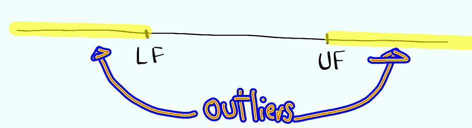
Draw Box Plot



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

Upper Fence $\quad Q_3 + 1.5(IQR)$

Lower Fence $\quad Q_1 - 1.5(IQR)$



Mar 14-9:44 AM

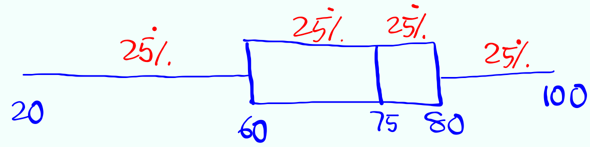
A Sample had the following 5-number Summary

20 60 75 80 100

Min Q_1 Med. Q_3 Max

$100\% \div 4 = 25\%$

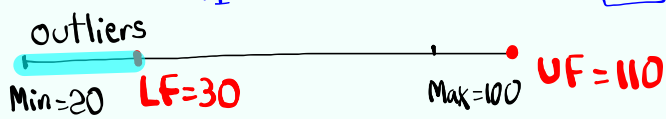
Box Plot



$IQR = Q_3 - Q_1 = 20$

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

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



Mar 14-9:52 AM

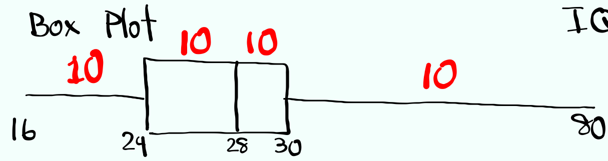
I randomly selected 40 students, the 5-Number Summary of their ages were

$40 \div 4 = 10$

16 24 28 30 80

Min Q_1 Med Q_3 Max

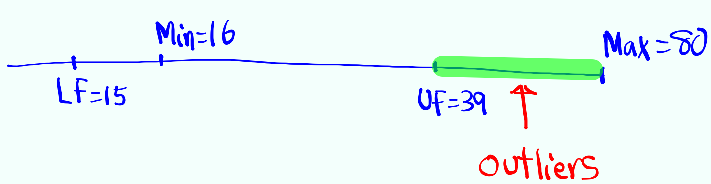
Box Plot



$IQR = Q_3 - Q_1$
$= 6$

Upper fence $Q_3 + 1.5(IQR) = 30 + 1.5(6) = \boxed{39}$

Lower fence $Q_1 - 1.5(IQR) = 24 - 1.5(6) = \boxed{15}$



Mar 14-9:58 AM

Z - Score:

Always round to 3-decimal places

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

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

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

Z-Score allows us to standardize data elements, then make comparison from different samples.

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

Mar 14-10:06 AM

In a sample of students, the mean age was 30 with standard deviation of 5 yrs.

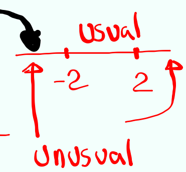
Jose was 35 yrs old, Is that a usual age?

$$Z = \frac{x - \bar{x}}{s} = \frac{35 - 30}{5} = \frac{5}{5} = 1$$

Usual age

Lisa had a Z-Score of -2.4. How old is Lisa?

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



$$-2.4 = \frac{x - 30}{5}$$

Cross-Multiply

$$x - 30 = 5(-2.4)$$

$$x = 30 - 12 = \boxed{18}$$

Mar 14-10:11 AM

STEM Plot

Data must be Sorted

STEM	Leaf						
2	58	25	28				
3	005559	30	30	35	35	35	39
⋮	⋮						
10	028	100	102	108			

Mar 14-10:17 AM

I randomly Selected 25 exams, here are the Scores

58	60	65	69	70
72	73	75	75	75
79	82	84	84	84
88	89	90	92	92
93	95	95	100	100

Data must be Sorted

STEM Plot

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

Mar 14-10:31 AM

Percentile

Data must be Sorted

Notation P_K

$K\%$ | $(100-K)\%$ at P_K

Ex: P_{15}: 15% | 85% at P_{15}

P_{90}: 90% | 10% at P_{90}

How to find P_K

1) Location $L = \frac{K}{100} \cdot n$

Sample Size → n

If L is decimal → Round-up to a whole #

$P_K = L$th element

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

Mar 14-10:35 AM

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

$n = 25$

Find P_{20}

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

$P_{20} = \frac{5\text{th} + 6\text{th}}{2} = \frac{70+72}{2} = \boxed{71}$

20% | 80% at 71

Find P_{50} Median

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

$L = 13$

$P_{50} = 13$th element

$= 84$

50% | 50% at 84

Mar 14-10:41 AM

Doing Reverse

Find K when P_K is given.

Find K such that $P_K = 90$

K is percentile ranking of

How many below

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

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

Round to whole %.

Sample Size

68% | 32%

90

$$K = \frac{17}{25} \cdot 100 = 68$$

$$\boxed{P_{68} = 90}$$

Mar 14-10:45 AM

I randomly selected 30 students, stem plot below is for their ages

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

$n = 30$

Find P_{15}

$L = \frac{15}{100} \cdot 30 = 4.5 = 5$

P_{15} = 5th element

$P_{15} = 22$

Find P_{50} Median

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

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

$= \frac{38 + 40}{2} = \boxed{39}$

Find percentile ranking of 50.

Find K such that $P_K = 50$

below

$$K = \frac{B}{n} \cdot 100 = \frac{21}{30} \cdot 100 = 70$$

70% | 30%

50

$$\boxed{P_{70} = 50}$$

Mar 14-10:51 AM