Statistics

Lecture 2

Feb 19-8:47 AM

Class QZ 1

Complete the chart below

class BNDRS	class F	Cum.F
12.5 – 22.5	2	2
22.5 – 32.5	5	7
32.5 – 42.5	3	10

Find

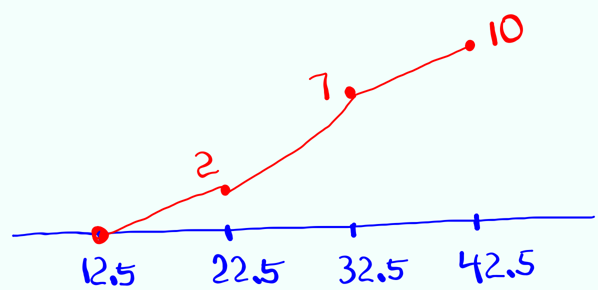
1) Sample Size n

$\boxed{n = 10}$

2) class width $\boxed{10}$

Draw its ogive display.

- class BNDRS
- Cum. F.
- Start @ 0 level.



Mar 6-7:45 AM

Computations in Statistics (SG 5-8)

$x \rightarrow$ Data element

$\sum x \rightarrow$ Sum of data elements

Summation

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

$n \rightarrow$ Sample Size

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

Consider the Sample below

1 3 3 3 5

1) $n = 5$

2) Min. = 1, Max. = 5

3) Range = Max − Min = 5 − 1 = 4

4) Midrange = $\frac{\text{Max} + \text{Min}}{2} = \frac{5+1}{2} = \frac{6}{2} = 3$

5) Mode 3

6) Median 3

"Data must be Sorted"

7) $\sum x = 1 + 3 + 3 + 3 + 5 = \boxed{15}$

8) $\bar{x} = \frac{\sum x}{n} = \frac{15}{5} = \boxed{3}$

Sample Mean

Mar 6-8:12 AM

Consider the Sample below

1 2 $\boxed{2 \quad 4}$ 4 8

1) $n = 6$

2) Range $= 8 - 1 = 7$

3) Midrange $= \frac{8+1}{2} = \frac{9}{2} = 4.5$

4) Mode 2 & 4

Bimodal

5) Median $= \frac{2+4}{2} = \frac{6}{2} = \boxed{3}$

"Data must be Sorted"

6) $\sum x = 1 + 2 + 2 + 4 + 4 + 8 = \boxed{21}$

7) $\bar{x} = \frac{\sum x}{n} = \frac{21}{6} = \frac{7}{2} = \boxed{3.5}$

Mar 6-8:21 AM

$x \to$ Data element

$x^2 \to$ Square each data element

$n \to$ Sample Size

$\sum x \to$ Sum of data elements

$\sum x^2 \to$ Sum of squares of each data element

$\bar{x} \to$ Sample Mean $\qquad \bar{x} = \dfrac{\sum x}{n}$

$s^2 \to$ Sample Variance

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

Mar 6-8:28 AM

Consider the Sample below

2 3 3 3 4

1) $n = 5$

2) Range $= 4 - 2 = 2$

3) Midrange $= \dfrac{4+2}{2} = 3$

4) Mode 3

5) Median 3

6) $\sum x = 2+3+3+3+4 = \boxed{15}$

7) $\sum x^2 = 2^2 + 3^2 + 3^2 + 3^2 + 4^2 = \boxed{47}$

8) $\bar{x} = \dfrac{\sum x}{n} = \dfrac{15}{5} = \boxed{3}$

9) $s^2 = \dfrac{n\sum x^2 - (\sum x)^2}{n(n-1)} = \dfrac{5 \cdot 47 - 15^2}{5(5-1)} = \dfrac{10}{20} = \boxed{.5}$

Mar 6-8:33 AM

Consider the Sample below

1 3 3 5 5 7

1) $n = 6$

2) Range $= 7 - 1 = 6$

3) Midrange $= \frac{7+1}{2} = 4$

4) Mode $3 \ \& \ 5$

5) Median $\frac{3+5}{2} = \boxed{4}$

6) $\sum x = 1+3+3+5+5+7$
$= \boxed{24}$

7) $\sum x^2 = 1^2+3^2+3^2+5^2+5^2+7^2 = \boxed{118}$

8) $\bar{x} = \frac{\sum x}{n} = \frac{24}{6} = \boxed{4}$

9) $S^2 = \frac{n\sum x^2 - (\sum x)^2}{n(n-1)} = \frac{6 \cdot 118 - 24^2}{6(6-1)} = \frac{132}{30} = \boxed{4.4}$

Mar 6-8:40 AM

Given $n = 8$, $\sum x = 72$, $\sum x^2 = 648$

Find

1) $\bar{x} = \frac{\sum x}{n} = \frac{72}{8} = \boxed{9}$

2) $S^2 = \frac{n\sum x^2 - (\sum x)^2}{n(n-1)} = \frac{8 \cdot 648 - 72^2}{8(8-1)} = \frac{0}{56} = \boxed{0}$

Mar 6-8:48 AM

$x \rightarrow$ Data element

$n \rightarrow$ Sample Size

$\bar{x} \rightarrow$ Sample Mean

$s^2 \rightarrow$ Sample Variance

$s \rightarrow$ Sample Standard Deviation

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

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

Mar 6-8:51 AM

Consider the Sample below

3 4 6 8 10

1) $n = 5$

2) Range $= 10 - 3 = 7$

3) Midrange $= \frac{10+3}{2} = 6.5$

4) Mode $\boxed{\text{None}}$

5) Median $\boxed{6}$

6) $\sum x = 3+4+6+8+10 = \boxed{31}$ ✓

7) $\sum x^2 = 3^2+4^2+6^2+8^2+10^2 = \boxed{225}$ ✓

8) $\bar{x} = \frac{\sum x}{n} = \frac{31}{5} = \boxed{6.2}$

9) $s^2 = \frac{n\sum x^2 - (\sum x)^2}{n(n-1)} = \frac{5 \cdot 225 - 31^2}{5(5-1)} = \frac{164}{20} = \boxed{8.2}$

10) $s = \sqrt{s^2} = \sqrt{8.2} \approx \boxed{2.864}$

Standard deviation is a non-negative value that indicates how data elements are spread from the mean.

Mar 6-8:55 AM

Consider the Sample below

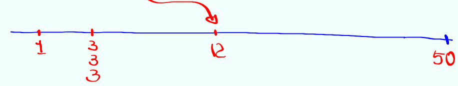
1 3 3 3 5

$n=5$ $\bar{x}=3$ $S=1.414$

NOW

1 3 3 3 50

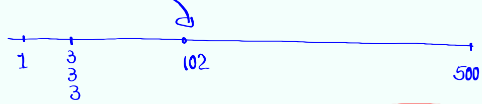
$n=5$ $\bar{x}=12$ $S=21.260$



NOW

1 3 3 3 500

$n=5$ $\bar{x}=102$ $S=222.490$



Small Standard deviation $\rightarrow$ Values are close to $\bar{x}$.

Large Standard deviation $\rightarrow$ Values are more spread out from $\bar{x}$.

Zero Standard deviation $\rightarrow$ All values are equal to $\bar{x}$.

Mar 6-9:04 AM

Consider the Sample below

5 5 5 5 5 5 5 5

1) $n=8$

2) $\sum x = 40$

3) $\sum x^2 = 200$

4) $\bar{x} = \frac{\sum x}{n} = \frac{40}{8} = \boxed{5}$

5) $S^2 = \frac{n\sum x^2 - (\sum x)^2}{n(n-1)} = \frac{8\cdot 200 - 40^2}{8(8-1)} = \frac{0}{56} = \boxed{0}$

6) $S = \sqrt{S^2} = \sqrt{0} = \boxed{0}$

Mar 6-9:13 AM

to Estimate Sample Standard deviation

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

The range rule-of-thumb

ex: A Sample has a min. of 30 and max. of 80.

1) Estimate its Standard deviation

$$S \approx \frac{\text{Range}}{4} = \frac{\text{Max} - \text{Min}}{4} = \frac{80-30}{4} = \frac{50}{4} = \boxed{12.5}$$

2) Estimate its Variance

$$s^2 \approx S^2 \approx 12.5^2 = \boxed{156.25}$$

Mar 6-9:35 AM

Empirical Rule:

It is best when mean, mode, and median are the Same.

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

About 95% of data falls within $\bar{x} \pm 2S$ **Usual Range**

About 99.7% of data falls within $\bar{x} \pm 3S$

When mean, mode, and median are the Same, data distribution will be Symmetric.

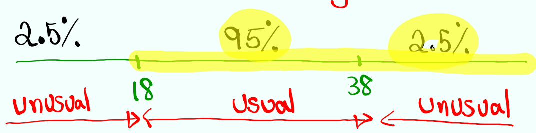
Mar 6-9:40 AM

Suppose ages of 80 students had a Symmetric dist. with $\bar{x} = 28$ and $S = 5$.

68% Range $\rightarrow \bar{x} \pm S = 28 \pm 5 \rightarrow$ $\boxed{23 \text{ to } 33}$

95% Range $\rightarrow \bar{x} \pm 2S = 28 \pm 2(5) = 28 \pm 10$

Usual Range $\rightarrow$ $\boxed{18 \text{ to } 38}$



97.5% of 80 Students were 18 or above.

$.975(80) = 78$ Students were at least 18 Yrs old.

99.7% Range $= \bar{x} \pm 3S = 28 \pm 3(5)$

$= 28 \pm 15$

$\rightarrow$ $\boxed{13 \text{ to } 43}$

Mar 6-9:46 AM

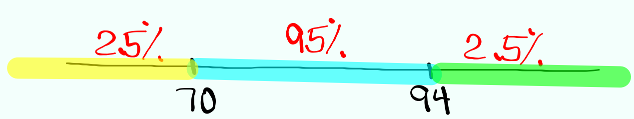
I randomly Selected 40 exams, Scores had a Symmetric dist. with $\bar{x} = 82$ and $S = 6$.

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

1) Find the Usual Range

$82 \pm 2(6) = 82 \pm 12 \rightarrow$ $\boxed{70 \text{ to } 94}$

2) How many exams had a unusual low Score?



what is 2.5% of 40?

$x = .025(40) = \boxed{1}$

Mar 6-9:53 AM

Z – Score

Always Round to 3-dec. places.

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

it tells us how many standard deviations is our data element away from $\bar{x}$.

If $Z > 0$ → above $\bar{x}$

If $Z < 0$ → below $\bar{x}$

If $Z = 0$ → equal to $\bar{x}$

If $-2 \le Z \le 2$ → Usual Value

If $Z < -2$ or $Z > 2$ → Unusual Value

Unusual | -2 | **Usual** | 2 | **Unusual**

we can use Z-Score to compare values from different Samples.

Mar 6-9:59 AM

Sherry got 92 on exam 1 and 80 on exam 2.

Exam 1 : $\bar{x} = 85$, $S = 8$

$$Z = \frac{x - \bar{x}}{S} = \frac{92 - 85}{8} = \frac{7}{8} = \boxed{.875}$$

$-2 \le Z \le 2$ → Score was usual.

Exam 2 : $\bar{x} = 70$, $S = 4$

$$Z = \frac{x - \bar{x}}{S} = \frac{80 - 70}{4} = \frac{10}{4} = \boxed{2.5}$$

$Z > 2$ → Unusual high Score.

Mar 6-10:06 AM

Salaries of 120 randomly selected nurses had a mean of \$6400 with standard deviation of \$500.

1) John makes \$7500/month Find his Z-Score. Usual or unusual.

$$Z = \frac{x - \bar{x}}{s} = \frac{7500 - 6400}{500} = \boxed{2.2}$$

Ususual high Salary

2) Maria's Z-Score is −1.8. Find her Salary.

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

$$-1.8 = \frac{x - 6400}{500}$$

Cross-multiply

$x - 6400 = -1.8(500)$

Isolate x

$x = 6400 - 1.8(500)$

$x = 5500$

\$5500

Mar 6-10:11 AM

5-Number Summary

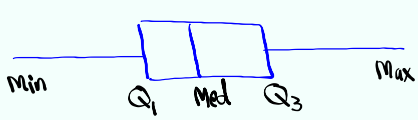
Min Q1 Med Q3 Max

Q1: First Quartile

Med: Median

Q3: third Quartile

1) Draw Box Plot



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

3) Upper Fence $= Q_3 + 1.5(IQR)$

Lower Fence $= Q_1 - 1.5(IQR)$

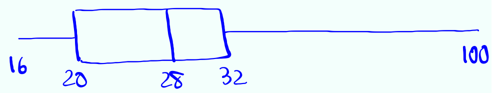
4) Discuss outliers

Mar 6-10:18 AM

Ages of students randomly selected has the following 5-Number summary.

16	20	28	32	100
Min.	Q_1	Med.	Q_3	Max

Box Plot



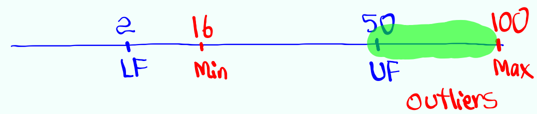
$IQR = Q_3 - Q_1 = 32 - 20 = \boxed{12}$

Upper fence $= Q_3 + 1.5(IQR)$

$= 32 + 1.5(12) = \boxed{50}$

Lower fence $= Q_1 - 1.5(IQR)$

$= 20 - 1.5(12) = \boxed{2}$

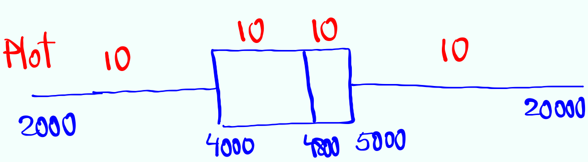


Mar 6-10:25 AM

I randomly selected 40 nurses. The 5-Number Summary of their Salaries are $40 \div 4 = 10$

2000 4000 4800 5000 20,000

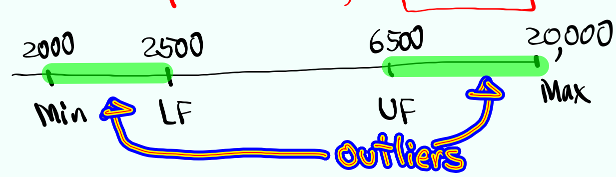
1) Draw Box Plot



2) $IQR = Q_3 - Q_1 = 5000 - 4000 = \boxed{1000}$

3) Upper fence $= Q_3 + 1.5(IQR) = \boxed{\$6500}$

Lower fence $= Q_1 - 1.5(IQR) = \boxed{\$2500}$



Mar 6-10:31 AM

Percentile

"Data must be Sorted"

Notation

$P_K \rightarrow K\%$ fall below it

$(100-K)\%$ fall above it.

$K\%$ $(100-K)\%$

P_K

P_{15} 15% 85%

P_{15}

Mar 6-10:39 AM

How to find P_K:

Sample Size

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

If decimal $\rightarrow$ Round-up , $P_K = L$th element

If whole # $\rightarrow P_K = \dfrac{L\text{th element} + \text{Next element}}{2}$

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

Sorted data

$n = 20$

Find P_{30}

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

$P_{30} = \dfrac{6\text{th element} + \text{Next element}}{2}$

$= \dfrac{65 + 68}{2} = \boxed{66.5}$

30% 70%

$P_{30} = 66.5$

Mar 6-10:42 AM

Use the Stem Plot below to find P_{70}

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

$n = 26$

$L = \frac{70}{100} \cdot 26 = 18.2$

$\boxed{L = 19}$

$P_{70} = 19\text{th element} = \boxed{45}$

Doing Reverse ⇒ Find K such that $P_K = 30$.

$K = \frac{\overset{\text{\# below}}{B}}{n} \cdot 100 = \frac{8}{26} \cdot 100 = 30.769\ldots \approx \boxed{31}$

Round to whole %.

31% 69%

$P_{31} = 30$

Mar 6-10:49 AM

Class QZ 2 (open notes)

Consider the Sample below

1 2 2 2 8

1) $\sum x = \boxed{15}$

2) $\sum x^2 = \boxed{77}$

3) $\bar{x} = \frac{\sum x}{n} = \frac{15}{5} = \boxed{3}$

4) $S^2 = \frac{n\sum x^2 - (\sum x)^2}{n(n-1)} = \frac{5 \cdot 77 - 15^2}{5(5-1)}$

$= \frac{160}{20} = \boxed{8}$

Mar 6-10:57 AM