Statistics

Summer 2023

Lecture 2

Driver's age at the time of accident

Class Limits	Class Boundaries	Class Midpoints	Freq.	Cumulative Frequency	Relative Frequency	Percentage Frequency
16-26	15.5 – 26.5	21	8	8	.100	10.0%
27-37	26.5-37.5	32	12	20	.150	15.0%
38-48	37.5-48.5	43	30	50	.375	37.5%
49-59	48.5-59.5	54	20	70	.250	25.0%
60-70	59.5-70.5	65	10	80	.125	12.5%

1) 5 Rows
5 Classes

2) Sample Size
$n = \Sigma f$
$n = 8 + 12 + 30 + 20 + 10$
$\boxed{n = 80}$

3) class width $= 27 - 16 = 38 - 27 = 49 - 38 = 60 - 49$
$\boxed{CW = 11}$

4) 26.5
26 27

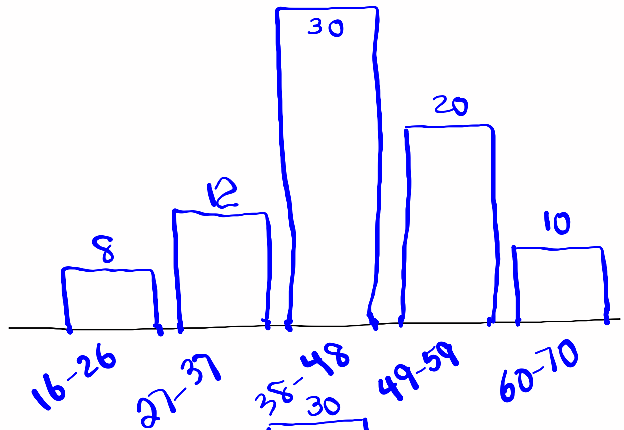
5) Class MP $= \dfrac{+\text{Class limits}}{2}$

6) Rel. f. $= \dfrac{f}{n} = \dfrac{f}{80}$

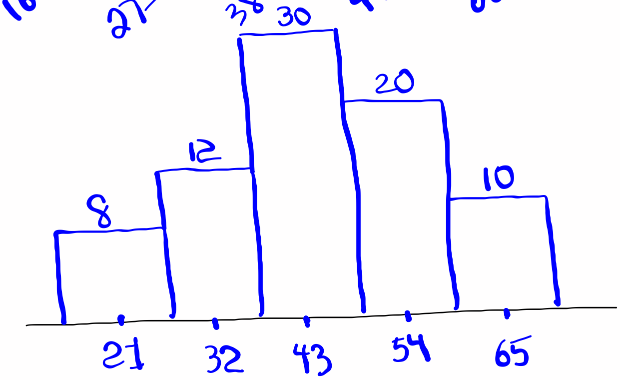
7) what % of drivers were between 27 & 59?
$15\% + 37.5\% + 25\% = \boxed{77.5\%}$

8) what % of drivers were below 60 Yrs?
$10\% + 15\% + 37.5\% + 25\% = \boxed{87.5\%}$

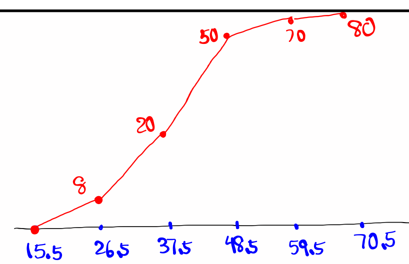
Draw bar chart
- Class limits
- class F



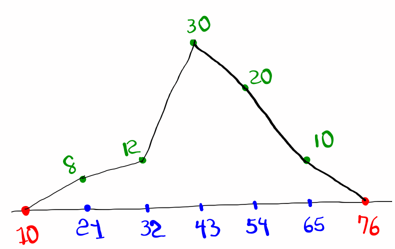
Draw Histogram
- class MP
- class F



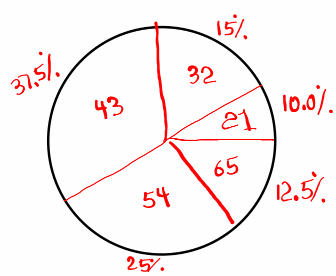
Draw ogive
- Class BNDRS
- Cum. F.
- Start at 0 level



Draw freq. Polygon
- class MP
- Extra MP, one on each side
- class F
- Start & finish at 0 level



Draw Pie chart
- Circle
- class MP
- % F



Consider the STEM plot below:

→ Data must be Sorted.
"Smallest → Largest"

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

1) Sample Size $n = \boxed{25}$

2) Min. $= \boxed{30}$ Max. $= \boxed{80}$

3) Range $=$ Max $-$ Min $= \boxed{50}$

4) Midrange $= \dfrac{\text{Max}+\text{Min}}{2} = \boxed{55}$

5) Mode $= \boxed{55}$

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

a) 3 classes

$CW = \dfrac{\text{Range}}{\#\text{classes}} = \dfrac{50}{3} = 16.\overline{6}$ $\boxed{CW = 17}$

when decimal → Round-up

b) 5 classes $CW = \dfrac{\text{Range}}{5} = \dfrac{50}{5} = 10$ $\boxed{CW = 11}$

when whole #, Add 1

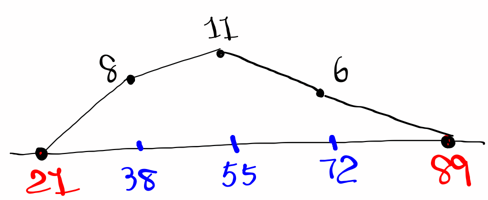
Make a freq. table with 3 classes. CW=17

class limits	class BNDRS	class MP	class F	Cum. F	Rel. F	% F
30 – 46	29.5 – 46.5	38	8	8	.32	32%
47 – 63	46.5 – 63.5	55	11	19	.44	44%
64 – 80	63.5 – 80.5	72	6	25	.24	24%

$\text{Rel. F} = \dfrac{f}{n} = \dfrac{f}{25}$

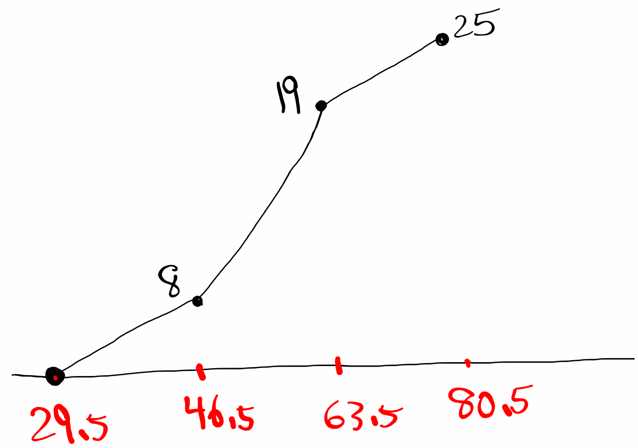
Freq. Polygon

- class MP
- Extra MP, one on each side
- Start & End at 0 level



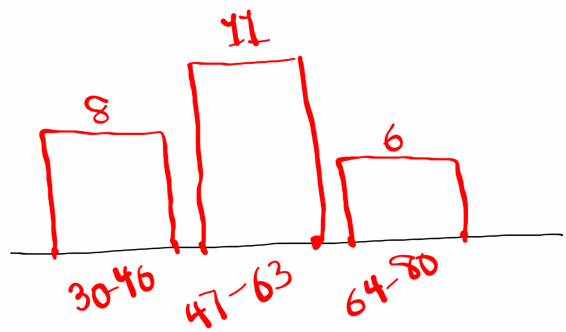
Ogive

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



Bar chart

- class limits
- class F



Class QZ 1

1) Evaluate $\dfrac{28-17}{\frac{10}{\sqrt{25}}} = \dfrac{11}{\frac{10}{5}} = \dfrac{11}{2} = \boxed{5.5}$

2) Find y when $x = -4$ given $y = 2.5x + 10$

$y = 2.5(-4) + 10 = -10 + 10 = \boxed{0}$

3) Determine the data type for # of Students

$\boxed{\text{Discrete}}$ or Continuous.

Countable

Consider the Sample below:

2, 2, 4, 4, 8

1) $n = \boxed{5}$

2) Range = Max − Min $= \boxed{6}$

3) Midrange = $\dfrac{\text{Max}+\text{Min}}{2} = \boxed{5}$

4) Mode $\boxed{2 \,\&\, 4}$ Bimodal

5) $\sum x = 2+2+4+4+8 = \boxed{20}$

6) $\sum x^2 = 2^2+2^2+4^2+4^2+8^2 = \boxed{104}$

7) $\dfrac{\sum x}{n} = \dfrac{20}{5} = \boxed{4}$

8) $\dfrac{n\sum x^2 - (\sum x)^2}{n(n-1)} = \dfrac{5\cdot 104 - 20^2}{5(5-1)} = \dfrac{120}{20} = \boxed{6}$

9) $\sqrt{\text{Last answer}} = \sqrt{6} \approx \boxed{2.449}$ 3-decimal Places

Computations in Statistics

SG 5−8

x → data element

$\sum x$ → Summation of data elements

n → Sample Size

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

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

Ex: Consider the Sample below

1, 3, 3, 3, 7

$n = 5$ Mode $= 3$

Range $= 7 - 1 = 6$ Midrange $= \dfrac{7+1}{2} = 4$

$\sum x = 1+3+3+3+7 = 17$

$\bar{x} = \dfrac{\sum x}{n} = \dfrac{17}{5} = \boxed{3.4}$

Consider the Sample below

1, 2, 2, 2, 4, 4, 4, 9

1) $n = \boxed{8}$

2) Range $= 9 - 1 = \boxed{8}$

3) Midrange $= \frac{9+1}{2} = \boxed{5}$

4) Mode $\boxed{2 \,\&\, 4}$

5) $\sum x = 1+2+2+2+4+4+4+9$
$= \boxed{28}$

6) $\bar{x} = \frac{\sum x}{n} = \frac{28}{8} = \boxed{3.5}$

$x \to$ Data element

$\sum x \to$ Sum of data elements

$x^2 \to$ Data element to the second power

$\sum x^2 \to$ Sum of data element^2.

$S^2 \to$ Sample Variance

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

Ex: $n = 6$, $\sum x = 26$, $\sum x^2 = 134$

$\bar{x} = \frac{\sum x}{n} = \frac{26}{6} = 4.\bar{3} \approx \boxed{4.333}$

$$S^2 = \frac{n\sum x^2 - (\sum x)^2}{n(n-1)} = \frac{6(134) - (26)^2}{6(6-1)} = \frac{128}{30}$$
$= 4.2\bar{6}$
$\approx \boxed{4.267}$

Consider the Sample below:

1 3 5 7

1) $n = \boxed{4}$

2) Range $= \boxed{6}$

3) Midrange $= \boxed{4}$

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

5) $\sum x = \boxed{16}$

6) $\sum x^2 = \boxed{84}$

7) $\bar{x} = \frac{\sum x}{n} = \frac{16}{4} = \boxed{4}$

8) $S^2 = \frac{n\sum x^2 - (\sum x)^2}{n(n-1)} = \frac{4 \cdot 84 - 16^2}{4(4-1)} = \frac{80}{12} \approx \boxed{6.667}$

9) $\sqrt{S^2} = \sqrt{6.667} \approx \boxed{2.582}$

Consider the Sample below

8 8 8 8 8

1) $n = \boxed{5}$

2) $\sum x = \boxed{40}$

3) $\sum x^2 = \boxed{320}$

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

5) $S^2 = \frac{n\sum x^2 - (\sum x)^2}{n(n-1)} = \frac{5 \cdot 320 - 40^2}{5(5-1)} = \frac{0}{20} = \boxed{0}$

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

$\bar{x} \to$ Sample Mean

$s^2 \to$ Sample Variance

$s \to$ Sample Standard Deviation $\boxed{s = \sqrt{s^2}}$

Given $n = 10$, $\sum x = 91$, $\sum x^2 = 903$

$$\bar{x} = \frac{\sum x}{n} = \frac{91}{10} = \boxed{9.1}$$

$$s^2 = \frac{n\sum x^2 - (\sum x)^2}{n(n-1)} = \frac{10 \cdot 903 - 91^2}{10(10-1)} = \frac{749}{90} \approx \boxed{8.322}$$

$$s = \sqrt{s^2} = \sqrt{8.322} \approx \boxed{2.885}$$

To estimate s:

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

Range rule-of-thumb

Given: $n = 10$, $\sum x = 272$, $\sum x^2 = 7600$,
Min. $= 18$, Max $= 36$

1) $\bar{x} = \frac{\sum x}{n} = \frac{272}{10} = \boxed{27.2}$

2) $s^2 = \frac{n\sum x^2 - (\sum x)^2}{n(n-1)} = \frac{10 \cdot 7600 - 272^2}{10(10-1)} = \frac{2016}{90} = \boxed{22.4}$

3) $s = \sqrt{s^2} = \sqrt{22.4} \approx \boxed{4.733}$

The range rule-of-thumb.

4) Estimate s $\to$ $s \approx \frac{\text{Range}}{4} = \frac{\text{Max} - \text{Min}}{4} = \frac{36 - 18}{4} = \frac{18}{4} \approx \boxed{4.5}$

5- Number Summary

Min. Q_1 Median Q_3 Max

$Q_1 \Rightarrow$ 25% below & 75% above

Med. $\Rightarrow$ 50% below & 50% above

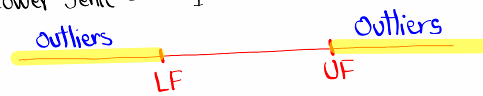
$Q_3 \Rightarrow$ 75% below & 25% above

Q_1 First Quartile $\Rightarrow IQR = Q_3 - Q_1$

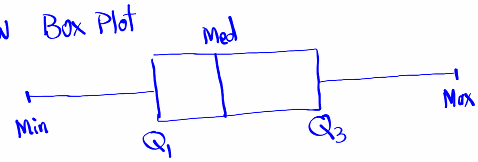
Q_3 Third Quartile inter-Quartile-Range

Upper Fence = $Q_3 + 1.5(IQR)$

Lower Fence = $Q_1 - 1.5(IQR)$



Draw Box Plot

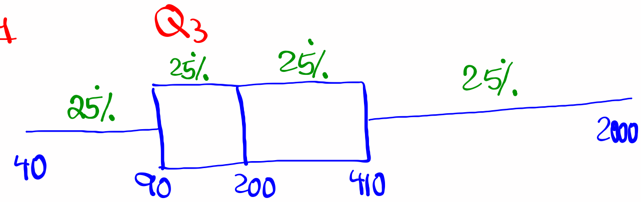


A Sample has the Following 5-Number Summary

40, 90, 200, 410, 2000

Min. Q_1 Med. Q_3 Max.

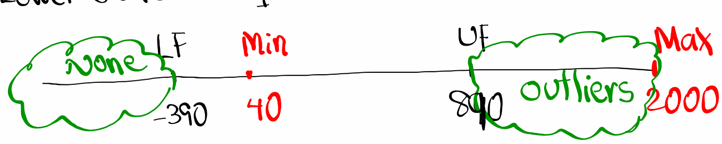
Box Plot



$IQR = Q_3 - Q_1 = 410 - 90 = \boxed{320}$

Upper Fence $= Q_3 + 1.5(IQR) = 410 + 1.5(320) = \boxed{890}$

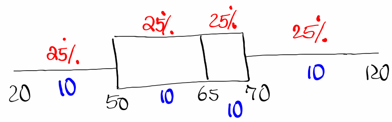
Lower Fence $= Q_1 - 1.5(IQR) = 90 - 1.5(320) = \boxed{-390}$



I randomly selected 40 exams, and here are the 5-number summary of scores: $\rightarrow 40 \div 4 = 10$

20	50	65	70	120
Min.	Q_1	Med	Q_3	Max

Box Plot



about 30 scores were above 50

About 50% of scores were between 50 & 70%

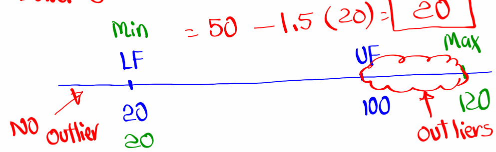
$IQR = Q_3 - Q_1 = 70 - 50 = \boxed{20}$

Upper fence $= Q_3 + 1.5(IQR)$

$= 70 + 1.5(20) = \boxed{100}$

Lower fence $= Q_1 - 1.5(IQR)$

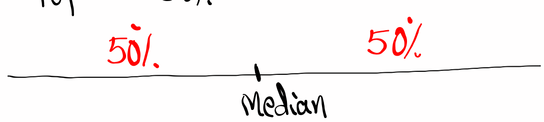
$= 50 - 1.5(20) = \boxed{20}$



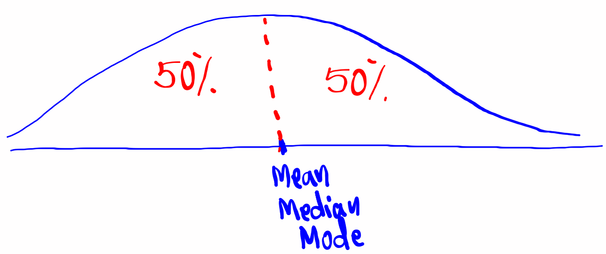
$\bar{x} \rightarrow$ mean

Mode $\rightarrow$ Data element that is repeated the most

Median $\rightarrow$ it is the value that separates the bottom 50% of data from the top 50% of data



When mean = mode = median, data distribution is symmetric and bell-shape.



Empirical Rule:

About 68% of data elements are within

$\bar{x} \pm S$

About 95% of data elements are within

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

About 99.7% of data elements are within

$\bar{x} \pm 3S$

Salaries of 300 nurses had a bell-shape dist with $\bar{x} = \$6400$ and $S = \$500$ per month.

By Empirical Rule

About 68% of them have Salaries within

$\bar{x} \pm S =$

$6400 \pm 500 \Rightarrow \boxed{5900 \text{ to } 6900}$

68% of 300 = $\boxed{204}$

About 95% of them have Salaries within

$\bar{x} \pm 2S =$

$6400 \pm 2(500) \Rightarrow \boxed{5400 \text{ to } 7400}$

Usual Range

95% of 300 = $\boxed{285}$

I graded 40 exams, Scores had a bell-Shape dist with $\bar{x}=80$ & $S=7.5$.

95% Range

1) find its $\boxed{\text{Usual Range.}}$ $\Rightarrow \bar{x} \pm 2S =$

$80 \pm 2(7.5) \rightarrow \boxed{65 \text{ to } 95}$

Unusual | 65 | Usual | 95 | Unusual

1 | | 38 | | 1

2) How many Scores were Usual?

95% of 40 = .95(40) = $\boxed{38}$

3) How many Scores were Unusually low?

$\boxed{1}$

2.5% | 95% Range | 2.5%

65 | 95

what % are above 65? 97.5%

Z - Score

It is a value that indicates how many Standard deviation is the data element above or below the mean.

$Z = \dfrac{x - \bar{x}}{S}$ Always round to 3-decimal places.

Suppose $\bar{x}=75$, $S=5$, and $x=82$

Find Z-Score

$$Z = \frac{x - \bar{x}}{S} = \frac{82-75}{5} = \frac{7}{5} = 1.4$$

1.4 S

$\bar{x}=75$ | 80 | 82 | 85

Suppose $\bar{x}=78$, $S=4$, $x=70$

Find Z-Score

$$Z = \frac{x - \bar{x}}{S} = \frac{70-78}{4} = \boxed{-2}$$

2 S

70 | 74 | $\bar{x}=78$

$Z<-2$ | $-2 \le Z \le 2$ | $Z>2$

Unusual | -2 | -1 | 0 | 1 | 2 | Unusual

Usual

Isabella got 95 on exam 1 and 86 on exam 2.

Exam 1: $\bar{x} = 88$, $S = 8$ $Z = \frac{95-88}{8} = \boxed{.875}$

Since $-2 \leq Z \leq 2 \rightarrow$ 95 is an usual Score

Exam 2: $\bar{x} = 75$, $S = 4$ $Z = \frac{86-75}{4} = \boxed{2.75}$

Since $Z > 2 \rightarrow$ 86 was unusual Score

Comparing these two exams, she did better in exam 2.

Z-Scores allow us to compare different Samples.

Class QZ 2

Complete the chart below

class BNDRS	class F	Cum. F.
18.5 – 26.5	3	3
26.5 – 34.5	7	10
34.5 – 42.5	2	12

Draw (clearly label)

1) Histogram

2) ogive