Feb 10, 25

Ungrouped data:

Measures of skewness
↳ outliers

- Mean:

  μ, $\bar{X}$

- Trimmed Mean:

  μ_{trim}, $\bar{X}_{trim}$

- Median:

  $\tilde{\mu}$, $\tilde{X}$

- Mode:

  $\hat{u}$, $\hat{X}$

Range $= R = X_m - X_0$

absolute
↳ same unit

~~Ab~~ Coeffiecient of despersion

$$= \frac{\text{Range}}{X_m + X_0}$$

$$\text{Ver} = S^2 = \frac{\Sigma(X - \bar{X})^2}{n-1}$$

$$S.d = S = \sqrt{\text{ver}}$$

$$C.V = \left(\frac{S.d}{\text{Mean}} \times 100\right)\%$$

median

Q_1 Q_2 Q_3

↓ ↓ ↓

P_{25} P_{50} P_{75}

$$i = \left(\frac{P}{100}\right) \times \text{total obs.}$$

→ integer : $\frac{i + (i+1) \text{ value}}{2}$

→ not an integer : Round up th observation.

$Q_1 = 0.25 \times$ total no. of observation.

$Q_2 =$ median $= P_{50} = 0.50 \times$ total

$P_{75} = Q_3 = 0.75 \times$

$P_{80} = 0.80 \times$ total..

$IQR = Q_3 - Q_1$

$$\text{Coef. of Q. Dev} = \frac{IQR}{Q_3 + Q_1}$$

Semi $IQR = IQR/2$.

Data:

Range has significant limitations:
- → Sensitive to outliners,
- → Variance & s.d consider all the pts
- → Var. & s.d are more consistent & reliable.

if naturally outliness exists : median / trimmed mean

Question:
data of 12 graduates

3450, 3550, 3650,
3480, 3355, 3310, 3490, 3730
3540, 3925, 3520, 3480.

Sorted data:

3310, 3355, 3450, 3480, 3480, 3490,
3520, 3540, ~~348~~, 3550, 3650,
3730, 3925.

1) **mean** = 3540

2) **Trimmed mean:**
20%

10% 10%

10% of 12 observations = 1.2%
round off to 1
drop 1, 1 values from bothsides.

trimmed mean = $\frac{\sum X_{trim}}{n_{trim}}$ = 3524.5

3) **Median:**

Even: $\frac{\left(\frac{n}{2}\right)th + \left(\frac{n}{2}+1\right)th}{2} = \frac{6^{th} + 7^{th}}{2}$

odd: $\left(\frac{n+1}{2}\right)th$

$= \frac{3490+3520}{2}$

3505

8.9 10.2 11.5 7.8 10.0 12.2 12.5 14.1
10.0 12.2 6.8 9.5 11.5 11.2 14.9 7.5 10.0 6.0
15.8 11.5

4) mode:
Most Frequent = 3480

Range:
$X_m - X_0 = 3925 - 3310$
$= 615$

Coef. of Dispersion:

$= \frac{Range}{X_m + X_0} = \frac{615}{7235} = 0.085$

$P_{25} = Q_1 = 0.25 \times 12 = 3$
$= \frac{3rd + 4th}{2} = 3465 = \left(\frac{3450 + 3480}{2}\right)$

$Q_2 = 0.50 \times 12 = 6$
$= \frac{6th + 7th}{2}$

$Q_3 = 0.75 \times 12 = 9 =$
$= \frac{9th + 10th}{2} = \frac{3550 + 3650}{2}$
$= 3600$

$P_{80} = 0.80 \times 12 = 9.6 = 10th$ value.
3650.

$IQR = Q_3 - Q_1 = 3600 - 3465$
$= 135$

Semi IQR : $\frac{IQR}{2} =$

coef of Q. Dev $= \frac{IQR}{Q_3 + Q_1}$

mean > median > mode
↳ +ively skewed.

mean = med. = mode
↳ symmetrical

mean < med < mod
↳ -ively skewed.

Pearson's coef. of Skewness:
-3 to +3

→ $Sk = \frac{\text{mean} - \text{mode}}{\text{s.d}}$

* mode = 3 median - 2 mean

$Sk = \frac{\text{mean} - 3\,\text{med} + 2\,\text{mean}}{\text{s.d}}$

can be determine Histogram left -3 to +3 → skewed to right.

slight 0.8
mode 1.8
highly —

mean, mode, med.
→ Pearson's = $\frac{3mean - 3med}{s.d.}$
→ Box & whisker plot

→ S.k = $\frac{3(mean - median)}{s.d}$ exact skeness.

$$\frac{3(3540 - 3505)}{165.65} = 0.63$$

$\cancel{s}^2$ =

	X	$(X-\bar{X})^2$
1)	3310	52900.
2)	3355	34225
3)	3450	8100
4)	3480	3600
5	3480	3600
6	3490	2500
7	3520	400
8	3540	0
9	3550	12100
10	3650	36100
11)	3730	148225
12	3925	

$$\frac{301850}{11} = 27440.91$$

Var = 27440.91

$sd = \sqrt{var} = 165.65$

For outliers
↳ Box & W Plot
↳ Z - Scores

Skeness
→ histogram
→ mean / mode / median
→ Pearson's
→ Box & Whisker's Plot.

bigger box → skewed to —
upper → right
lower → left

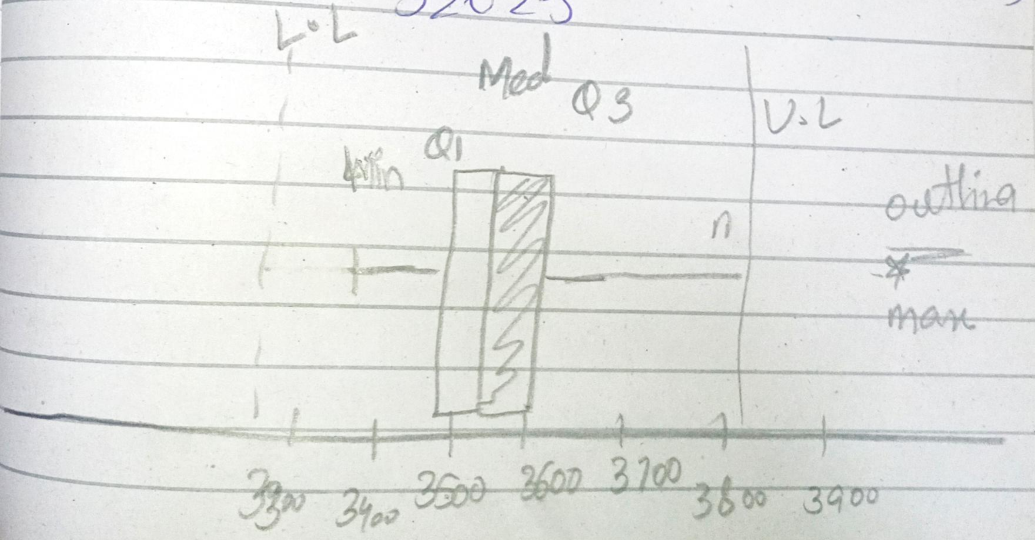
→ Fine Point Summary:

Min	Q_1	Med	Q_3	Max
3310.	3465	3505	3600	3925

Fences:

$U.L = Q_3 + 1.5(IQR)$
$= 3600 + 1.5(135)$
$= 3802.5$

$L.L : Q_1 - 1.5(IQR) = 3465 - 1.5(135)$
$= 3262.5$



Machine learning
↳ anamoly.

Z-scores:

$$Z_i = \frac{X - mean}{s.d}$$

Sample : $\frac{X - \bar{X}}{s}$

pop : $\frac{X - U_x}{s_x}$

if z_i lies b/w -3 to +3, ~~then~~ there is no outlier in data.

Z ~ mean = 0
s.d = 1