Statistics, lecture 5:

Note: the m% to n% inter-percentile range is:

IPR = $P_n - P_m$

e.g) Find the 35% to 70% inter-percentile range for:

3, 5, 5, 6, 8, 8, 9, 10, 11, 12

sol) $P_{35} = \frac{35}{100} \times 10 = 3.5 \rightarrow 4^{th}$ value

$P_{35} = 6$

$P_{70} = \frac{70}{100} \times 10 = \frac{7^{th} + 8^{th}}{2}$

$\frac{9+10}{2} = 9.5$

IPR $= 9.5 - 6$
$= 3.5$

e.g)

x	1	2	3	4	5	sum
f	3	8	7	2	5	25
cf	3 1-3	11 4-11	18 5-18	20 6-20	25 21-25	

Find the 40^{th} percentile:

sol) $P_{40} = \frac{40}{100} \times 25 = \frac{10^{th} + 11^{th}}{2} = \frac{2+2}{2}$

$P_{40} = 2$

* في المثال السابق ، ممكن يعطيني c.f و تكون f مجهولة ولكن يمكن ايجادها ببساطة .

eg) Find the 60^{th} percentile for

I	0-4	5-9	10-14	15-19	Sum
f	3	8	7	2	20
cf	3	11 12	18	20	
URB	4.5	9.5 P60	14.5	19.5	

sol) $\frac{60}{100} \times 20 = 12^{th}$ value

$\frac{12-11}{18-11} = \frac{P60 - 9.5}{14.5 - 9.5}$

11 — 12 — 18

9.5 — P60 — 14.5

$\frac{1}{7} = \frac{P60 - 9.5}{5}$

$P60 = 10.2$

مؤشر صحة الحل: بما انّ 12 اقرب لل 11
اذاً P60 اقرب لل 9.5 ✓

* The standard deviation and variance

i) For raw data and steam and leaf

eg) Find the variance and standard deviation for

i) 2, 7, 5, 11, 5

Sol)

method 1: $\bar{X} = \frac{30}{5} = 6$

$x-\bar{x}$	-4	1	-1	5	-1	0
$(x-\bar{x})^2$	16	1	1	25	1	44

Sum

$\text{variance} = \frac{44}{4} = 11$

$\text{standard deviation} = \sqrt{11} \approx 3.32$

Note: $\Sigma(x-\bar{x}) = 0$

Method 2:

						Sum
x	2	7	5	11	5	30
x^2	4	49	25	121	25	224

$$S^2 = \frac{224}{4} - \frac{(30)^2}{5 \times 4} = 11$$

$\text{standard deviation} = \sqrt{11} \approx 3.32$

****

eg) 1, 2, 3, 4, 5 "Population"

$$\mu = \frac{\Sigma x}{n} = \frac{15}{5} = 3$$

$(x-\mu)$: Deviation of x from μ

تقيس بعد البيانات عن وسطها الحسابي

$x-\mu$: $-2, -1, 0, 1, 2$

ملاحظة: الاشارة السالبة تدل ان الرقم اصغر من الوسط الحسابي والموجبة تدل انه اكبر منه

تذكروا: $\Sigma(x-\mu)=0$ لذلك نلجأ للتربيع للتخلص من الصفر

$(x-\mu)^2$: 4, 1, 0, 1, 4 $\rightarrow \Sigma(x-\mu)^2 = 10$

$$\sigma = \sqrt{\frac{\Sigma(x-\mu)^2}{n}} = \sqrt{\frac{10}{5}} \approx 1.41$$

$\text{variance} = \sigma^2 = 2$

eg) 1, 2, 3, 4, 5 "sample"

$$\bar{x} = \frac{\Sigma x}{n} = \frac{15}{5} = 3$$

$(x-\bar{x})$: $-2, -1, 0, 1, 2 \rightarrow \Sigma(x-\bar{x}) = 0$

$(x-\bar{x})^2$: 4, 1, 0, 1, 4 $\rightarrow \Sigma(x-\bar{x})^2 = 10$

$$S = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}} = \sqrt{\frac{10}{4}}$$

$$= \sqrt{2.5} \approx 1.58$$

$\text{variance} = S^2 = 2.5$

لاحظوا: يوجد اختلاف في النتيجة حسب السؤال هل هو sample أم population.

eg) 1, 2, 3, 4, 5

If the following value is added to the set of data, then what will happen to the mean?

$$\left\{ \begin{array}{l} \bar{x} = 3 \\ S = 1.58 \end{array} \right.$$

تم حسابهم سابقاً

a) $x = 3$

بما أنّ العدد الذي اضفناه يساوي الوسط الحسابي اذاً تبقى قيمة الوسط ثابتة

b) $X=4$

بما ان 4 اقل من الوسط الحسابي اذاً سوف يزداد

c) $X=2$

بما ان 2 اقل من الوسط الحسابي اذاً سيقل

what will happen to S if the following data is added to the set of data?

a) $X=3 \longrightarrow$ لمعرفة الجواب علينا تذكّر ان S لا تتعامل مع المشاهدة وانما متوسط المزوقات فيكون الحل:

$X-\bar{X}=0$

$0<S \quad \therefore$ S will decrease

b) $X=7$

$X-\bar{X}=7-3=4>S \quad \therefore$ S will increase

c) $X=4.58$

$X-\bar{X}=4.58-3=1.58=S$

$\therefore$ S will remain the same

d) $X=0$

$|X-\bar{X}|=|0-3|=|-3|=3>S \quad \therefore$ S will increase

لاحظوا ... لم نهتم بالاشارة السالبة فالمهم هو البعد عن الوسط كقيمة موجبة

آلية الحكم على S: ننظر الى المزمه بغض النظر عن اشارته ثم نحكم ... كما في الامثلة السابقة

e.g) A set of data has a mean $\bar{X}$ and a standard deviation S. Given that $X-\bar{X}>S$, then.

a) $\bar{X}$ will increase and S will decrese
b) $\bar{X}$ will decrease and S will increase
c) $\bar{X}$ will decrease and S will decrease
d) $\bar{X}$ remains the same and S will increase
e) $\bar{X}$ will increase and S will increase

Sol) $X-\bar{X}>S>0$

$X-\bar{X}>0$

$X>\bar{X} \quad \therefore \bar{X}$ will increase

$X-\bar{X}>S \quad \therefore$ S will increase

the answer is: E

Note: $S=0 \Rightarrow$ all the observations are equal

* For frequency distribution and grouped " " :

Population	Sample
$\sigma=\sqrt{\frac{\Sigma f.(x-\mu)^2}{\Sigma f}}$	$S=\sqrt{\frac{\Sigma f(x-\bar{x})^2}{\Sigma f-1}}$
$=\sqrt{\frac{\Sigma f x^2}{\Sigma f}-\mu^2}$	$\sqrt{\frac{\Sigma f x^2}{\Sigma f-1}-\frac{(\Sigma f x)^2}{\Sigma f(\Sigma f-1)}}$
	variance $=S^2$

e.g)

Sample

x	1	2	3	4	5	Sum
f	3	8	7	4	5	27
fx	3	16	21	8	25	81
x^2	1	4	9	16	25	
$f.x^2$	3	32	63	64	125	287

Sol) $\bar{X} = \frac{\sum f.x}{\sum f} = \frac{81}{27} = 3$

$$S = \sqrt{\frac{\sum fx^2}{\sum f - 1} - \frac{(\sum fx)^2}{\sum f(\sum f - 1)}}$$

$$= \sqrt{\frac{287}{27-1} - \frac{(81)^2}{27(26)}}$$

$= 1.30$

variance $= S^2 = 1.69$ Thus $S = 1.30$

e.g)

I	0-4	5-9	10-14	15-19	Sum
F	3	8	7	2	20
x	2	7	12	17	
f.x	6	56	84	34	180
x^2	4	49	144	289	
$f.x^2$	12	392	1008	578	1990

$\bar{X} = \frac{\sum fx}{\sum f} = \frac{180}{20} = 90$

$$S = \sqrt{\frac{\sum fx^2}{\sum f - 1} - \frac{(\sum fx)^2}{\sum f(f-1)}}$$

$$= \sqrt{\frac{1990}{19} - \frac{(180)^2}{20(19)}} \simeq 4.41$$

variance $= S^2 \simeq 19.5$

$S = \sqrt{19.5} = 4.41$

* Z-score (standard score)

e.g) 1, 2, 3, 4, 5 $\to \bar{X} = 3$, $S = 1.58$

$X - \bar{X}$: $-2, -1, 0, 1, 2$

ملاحظة ← $(X-\bar{X})$ جميعها قيم لكنها بدون وحدة لذلك فعلنا التالي .

$\frac{X-\bar{X}}{S}$: $\underbrace{-1.27, -0.633, 0, 0.633, 1.27}$

هذه القيم لها وحدة وهي standard deviation

مثلاً : 1.27 معناها ان الرقم 5 اعلى من الوسط الحسابي بمقدار 1.27 standard deviation اذاً :

$Z = \frac{X - \mu}{\sigma}$ "population"

$Z = \frac{X - S}{\sigma}$ " Sample "

$Z < 2 \Rightarrow$ usual observation

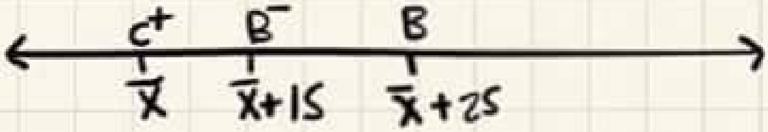
$Z > 2 \Rightarrow$ unusual observation

$Z > 3 \Rightarrow$ very unusual observation e.g)

	statistics	physics
$\bar{X}$	14	17
S	2	3
mark X	15	18

$Z = \frac{X - \bar{X}}{S} = \frac{15-14}{2} = 0.5$ "Statistics"

$Z = \frac{X - \bar{X}}{S} = \frac{18-17}{3} = 0.33$ "Physics"



اذاً الهدف من المثال هو توضيح ان اداء الطالب في اختبار معين لا يعتمد على علامته بل على اداء زملائه

Note: The student did better in Statistics than in physics

*Coefficient of variation (C.V.)

$C.V. = \frac{S}{\bar{X}} \times 100\%$

" يعطينا انطباع ادق للدراسة بحيث نعرف اي المجموعات لها تشتت اقل "

e.g)

	section I	section II
$\bar{X}$	60	70
Standard deviation	4.5	5

C.V. of section I $\rightarrow \frac{4.5}{60} \times 100\% = 7.5\%$

C.V. of section II $\rightarrow \frac{5}{70} \times 100\% = 7.14\%$

* على الرغم من ان standard deviation للمجموعة الثانية اعلى الا ان تشتتها اقل

* the variability of section I is more than section II

نستنتج: عند مقارنة التشتت بين مجموعتين يجب حساب (C.V.) الا في حال كان الوسط الحسابي متساوٍ و يمكن الاعتماد على standard deviation فقط

****

e.g) If the mean mark of 10 students is 15 and the standard deviation is 3. Ahmad with mark 18 joined the class. Find the new mean and the new standard deviation sol) $\left.\begin{matrix} n=10 \\ \bar{x}=15 \end{matrix}\right]$ $\Sigma x = 150$

$$\begin{array}{r} 150 \\ 18\ + \\ \hline 168 \end{array}$$ "new Σx"

new no. = 11

new $\bar{x} = \frac{168}{11} = 15.3$

Notes: $S^2 = \frac{\Sigma x^2}{n-1} - \frac{(\Sigma x)^2}{n(n-1)}$

$3^2 = \frac{\Sigma x^2}{10-1} - \frac{(150)^2}{10(9)}$

$9 = \frac{\Sigma x^2}{9} - 250$

$\Sigma x^2 = 2331$

$18^2 +$

new $\Sigma x^2 = 2655$

new no = 11

new $\Sigma x = 168$

new $S^2 = \frac{2655}{10} - \frac{(168)^2}{11 \times 10}$

$= 8.92$

new $S = \sqrt{8.92} \approx 2.99$

ملاحظة : جربوا حلوا السؤال اذا كان المكتوب

Joined بدلاً من left the class .

e.g) If the mean mark of 10 boys is 12 and the standard deviation is 2 while the mean of 12 girls is 10 and the standard deviation is 3. Find the mean mark and the standard deviation for the students alltogether

sol)

Boys: $n=10$, $\bar{x}=12$, $S_1=2$, $\Sigma x=120$

Girls: $m=12$, $\bar{y}=10$, $S_2=3$, $\Sigma x=120$

$C=22$

$\Sigma x = 240$

$\bar{F} = \frac{240}{22} = 10.9$

$\Sigma x^2 = 2775$

Boys $\rightarrow$ $S^2 = \frac{\Sigma x^2}{n-1} - \frac{(\Sigma x)^2}{n(n-1)}$

$2^2 = \frac{\Sigma x^2}{9} - \frac{120^2}{10 \times 9}$

$\Sigma x^2 = 1476$

Girls $\rightarrow$ $3^2 = \frac{\Sigma x^2}{11} - \frac{(120)^2}{12 \times 11}$

$\Sigma x^2 = 1299$

Girls and boys $\rightarrow$ $\Sigma x^2 = 2775$

alltogether $\rightarrow$

$S^2 = \frac{2775}{21} - \frac{(240)^2}{22 \times 21} = 7.47$

$S = \sqrt{7.47} \approx 2.73$

* Measures of variation are never negative

Note: relative Frequency (r.f.)

$$r.f. = \frac{f}{\Sigma f}$$

e.g)

x	1	2	3	4	5	Sum
f	3	8	7	2	5	25
r.f.	$\frac{3}{25}$ 0.12	$\frac{8}{25}$ 0.32	$\frac{7}{25}$ 0.28	$\frac{2}{25}$ 0.08	$\frac{5}{25}$ 0.20	1

مجموعهم

ملاحظة :- ممكن يكون (r.f.) معلوم ويطلب f او c.f