Statistics, lecture 4:

*Measures of variation:

1) The range

Range = Max - Min

= Max (U.R.B) - min (L.R.B)

eg) Find the range for:

i) 2, 7, 5, 11, 3, 8

sol) 11 - 2 = 9

ii)

S	L		
1	1	2	3
2	1	1	2
3	1	2	2

Key: 1|2 represents 17

sol) Range = 32 - 11 = 21

iii)

X	1	2	3	4	5
f	100	200	13	7	25

sol) Range = 5 - 1 = 4

لاحظوا: التكرارات ليس لها علاقة بالكل.

iv)

I	0-4	5-9	10-14	15-19
f	3	8	7	2

-0.5 ... 19.5 ... 11

Range = 19.5 - (-0.5)

= 20

Note: class length (C.L) or " width (C.L) is:

C.L. = U.R.B - L.R.B

eg)

-0.5 ... 3.5 ... 9.5 ... 11.5 ... 20.5

I	0-3	4-9	10-11	12-20
f	3	7	8	9
CL	4	6	2	9

سؤال: ايهما افضل، ان يكون الوسط الحسابي والـ Range مرتفعين ام منخفضين؟

جواب: حسب المسألة، هنا بالنسبة الوسط الحسابي، كندا بتحليه بالاحتمالات الافضل ان يكون مرتفع، اما كندا بتحليه بمرض معين - لا سمح الله - فالافضل ان يكون منخفض

بالنسبة لـ Range -- لو قلنا ان قيمته بالنسبة للدخل الشهري لاعب المباريات = 2000 فهذا يعتبر مؤشر لخسارة عالية، فارتفاعه هنا كان سيء، وهكذا في السؤال

ملاحظة: كلما قلّ Range ازداد تقارب البيانات

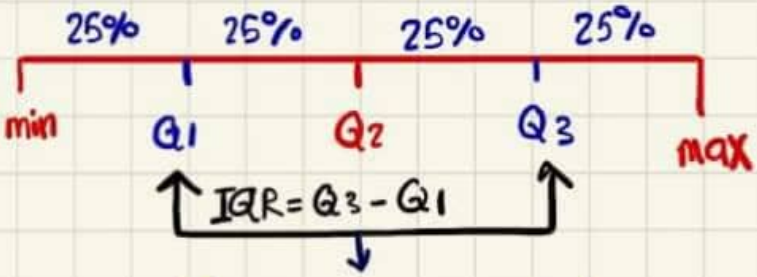
2) The inter-quartile range (IQR):

IQR = Q3 - Q1, where

Q1: Lower quartile (25% of the data lie below Q1)

Q3: Upper quartile (75% of the data lie below Q3)

الممسوحة ضوئيا بـ CamScanner



The middle 50% data lie between Q1 and Q3

One can find Q1 and Q3 in the same way as finding Q2 but we use:

هذه الطريقة فقط للجدول

Q_1: $\left(\frac{n}{4}\right)^{th}$ value

Q_3: $\left(\frac{3n}{4}\right)^{th}$ value

eg) 2, 7, 5, 11, 3, 8, 10

sol) 2, 3, 5, 7, 8, 10, 11

Q1 Q2 Q3

IQR = 10 - 3
= 7

eg) 2, 3, 5, 7, 7, 9, 11, 12

Q1 = 4 Q2 Q3 = 10

IQR = 10 - 4
= 6

eg)

S	L
1	1 2 3
2	1 1 2
3	2 2 3

Key : 1|2 represents 12

$Q_1 = \frac{12+13}{2} = 12.5$

$Q_3 = \frac{32+32}{2} = 32$

IQR = 32 - 12.5 = 19.5

مهم : يوجد اختلاف في طريقة حساب Q1 و Q3 بين اليوتيوب والشرح الجديد ، حيث الطريقة المعتمدة هي الطريقة اعلاه ويجب اكل بها في الاختبار .

* بالنسبة لطريقة اليوتيوب سوف نعتمدها للجداول .

الممسوحة ضوئيا بـ CamScanner

e.g)

x	1	2	3	4	5
f	3	8	7	2	5
cf	3 (1-3)	11 (4-11)	18 (12-18)	20 (19-20)	25 (21-25)

$Q1 \rightarrow \frac{n}{4} \rightarrow \frac{25}{4} = 6.25 \rightarrow 7^{th}$ value

$Q1 = 2$

$Q3 \rightarrow \frac{3n}{4} = \frac{75}{4} = 18.75 \rightarrow 19^{th}$ value

$Q3 = 4$

$IQR = 4 - 2$

2

e.g)

I	0-4	5-9	10-14	15-19
f	5	7	3	5
cf	5	12	15	20
U.R.B	4.5	9.5	14.5	19.5

$Q1 \rightarrow \frac{n}{4} = \frac{20}{4} = 5^{th}$ value $= 4.5$

$Q3 \rightarrow \frac{3n}{4} = \frac{60}{4} = 15^{th}$ value $= 14.5$

$IQR = 14.5 - 4.5$

10

5

e.g)

I	0-4	5-9	10-14	15-19
f	3	8	7	2
cf	3	11	18	20
U.R.B	4.5	9.5	14.5	19.5

$Q1 \rightarrow \frac{n}{4} = \frac{20}{4} = 5^{th}$ value

$\frac{5-3}{11-3} = \frac{Q1-4.5}{9.5-4.5}$

$Q1 = 5.75$

3 — 5 — 11

4.5 — Q1 — 9.5

الخمسة اقرب للـ ٣ بالتالي

منطقيا Q1 اقرب للـ 4.5

$Q3 = \frac{3n}{4} = \frac{60}{4} = 15^{th}$ value

$\frac{15-11}{18-11} = \frac{Q3-9.5}{14.5-9.5}$

$Q3 = 11.4$

11 — 15 — 18

9.5 — Q3 — 14.5

$IQR = Q3 - Q1$

$11.4 - 5.75$

≈ 6.61

6

الممسوحة ضوئيا بـ CamScanner

*Outliers

Any value less than Q1 - 1.5 IQR or greater than Q3 + 1.5 IQR

e.g) Find the outliers (if any) for:-

i) 2, 10, 11, 12, 13, 20

Q1 Q2 Q3

IQR = 13 - 10 = 3

Q1 - 1.5 IQR = 10 - 1.5 x 3 = 5.5 ∴ 2 is an outlier

Q3 + 1.5 IQR = 13 + 1.5 x 3 = 17.5 ∴ 20 is an outlier

ii) 2, 10, 12, 14, 16, 20

Q1 Q2 Q3

IQR = 6

Q1 - 1.5 IQR = 10 - 1.5 x 6 = 1

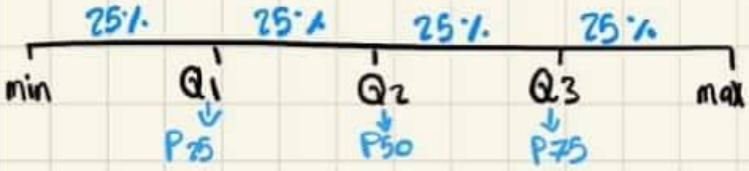
Q3 + 1.5 IQR = 16 + 1.5 x 6 = 20

∴ No outliers

ملاحظة: لو عندي 1,2,3,1000 واضح جداً ان 1000 هي outlier لكن ممكن ما يكون هنا لا بد من الحسابات. 1,2,3,8

*Fractiles

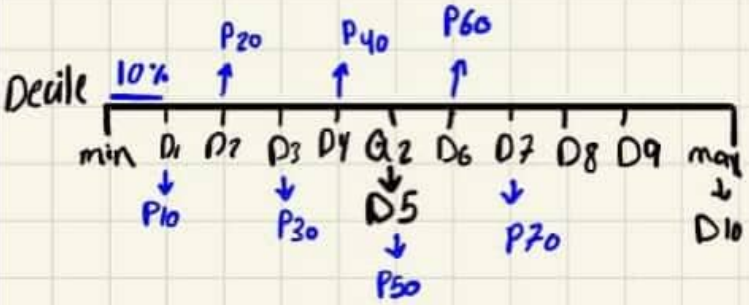
Quartiles (Q1, Q2, Q3) — Percentile P_k — Decile D_k



20% to 80% IPR = P80 - P20

لمعرفة المدى الربعي IQR

نكتب ال Percentile بدل ال Quartile لكن ما يهمنا فيها هو نسبة المشاهدات يسار القيمة



Note: $P_{80} \rightarrow \left(\frac{80}{100} \times n\right)^{th}$ value

$P_k = \left(\frac{k \times n}{100}\right)^{th}$ value, $D_k \left(\frac{k}{10} n\right)^{th}$ value

الممسوحة ضوئيا بـ CamScanner

e.g) Find the 60^{th} Percentile and the 7^{th} decile for:

i) 2, 3, 5, 7, 8, 10, 11 "يجب ان تكون الارقام مرتبة"

sol) $P_{60} \rightarrow \frac{60}{100} \times 7 = 4.2 \rightarrow 5^{th}$ value

$P_{60} = 8$

$D_7 \rightarrow \frac{7}{10} \times 7 = 4.9 \rightarrow 5^{th}$ value

$D_7 = 8$

Percentile of x =

$$\frac{\text{number of data less than x}}{\text{total number of data}} \times 100$$

e.g) 1, 2, 3, 4, 5

Percentile of $4 = \frac{3}{5} \times 100 = 60 \quad \therefore$

$4 = P_{60}$

* Measures of location:
Quartiles, Percentiles, deciles

* Midquartile $= \frac{Q_1 + Q_3}{2}$

3] The standard deviation and variation

i) For raw data and steam and leaf:

Population μ

$\sigma^2 = \frac{\Sigma (x-\mu)^2}{n}$

$= \frac{\Sigma x^2}{n} - \mu^2$

Sample

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

$\frac{\Sigma x^2}{n-1} - \frac{(\Sigma x)^2}{n(n-1)}$

$\sigma \uparrow$ $\uparrow s$

Standard deviation $= \sqrt{\text{variance}}$

Note: $x - \mu$ deviation of x from μ

الممسوحة ضوئيا بـ CamScanner