جامعة 20 أوت 1955 سكيكدة السنة الجامعية : 2022/2021

السنة الأولى MI مقياس : مدخل إلى الإحتمالات و الإحصاء الوصفي

الفصل الأول : مبادئ أولية وبعض المفاهيم الإحصائية الضرورية

I. تعريف علم الإحصاء

الإحصاء هو فرع من فروع الرياضيات يشمل النظريات و الطرق الموجهة نحو جمع البيانات ووصفها و الاستقراء و صنع القرارات.

يصنف الإحصاء كعلم إلى :

- **الإحصاء الوصفي Statistique descriptive:** يتناول طرق تنظيم وتلخيص وعرض البيانات في صورة مبسطة.
- **الإحصاء الإستدلالي أو التطبيقي Statistique inférentielle:** يهتم بطرق الوصول إلى نتائج معينة أو توقعات ما عن المجتمع من خلال دراسة عينة من ذلك المجتمع.

II. المفاهيم الأساسية لعلم الإحصاء

1. **المعطيات أو البيانات(les données):** هي المعلومات التي يتم جمعها و تنظيمها و تحليلها من طرف الإحصائيين.
2. **المجتمع الإحصائي (population) :** هو مجموعة الأفراد الذين تخصّهم دراسة إحصائية معينّة أي المجموعة التي تجرى عليها المشاهدات. مثلا سكان مدينة، طلبة جامعة ما، مجتمع من الأسر ... إلخ.
3. **الفرد أو الوحدة الإحصائية (unité statistique ou individu) :** هو العنصر المكون للمجتمع الإحصائي. مثلا طالب من جامعة ما.
4. **العينّة الإحصائية (échantillon) :** هو مجموعة جزئية من المجتمع الإحصائي، تجرى المشاهدات عليها إذا كان عدد عناصر المجتمع كبيراً.
5. **الميزة الإحصائية (الصفة أو المتغير الإحصائي) (caractère-variable):** هي الخاصية التي يرغب الباحث في دراستها و العنصر المشترك بين أفراد مجتمع ، مثل الطول بالنسبة لمجموعة الطلبة، الوزن،...،الخ. الهيئات(القيم) التي يمكن أن تظهر عليها الميزة تسمى **الكيفيات (Modalités)**. الميزة نوعان:

أ. **ميزة نوعية (كيفية-وصفية) (caractère qualitatif):** هي الميزة التي لا يمكن قياسها (لا تأخذ قيما عددية). مثلا: اللون، الجنس، الجنسية، الحالة العائلية ...الخ. تنقسم إلى:

- **ميزة نوعية خاضعة للترتيب - ordinale:** مثل المستوى التعليمي، الرتب العسكرية، تقديرات النجاح... الخ.
- **ميزة نوعية غير خاضعة للترتيب- nominale :** مثل الجنسية، أنواع السيارات، أنواع الأمراض ...إلخ.

ب. **ميزة كمية (عددية)- caractère quantitatifs:** هي الميزة التي يمكن قياسها (تأخذ قيما عددية). مثلا : العمر، القامة، نقط اختبار، عدد الإخوة . المتغير الكمي نوعان :

- **كمي متقطع (منفصل) -discret:** هي الميزة الكمية التي تأخذ قيما معزولة صحيحة أو قابلة للعد مثلا: العمر، عدد الإخوة، سنة الميلاد، عدد أفراد أسرة،....

- **كمي مستمر(متصل) -continu:** هي الميزة الكمية التي تأخذ قيما في مجال معطى مستمر في R أو غير قابلة للعد. مثلا: القامة، الوزن، المسافة، ....

عندما تكون قيم المتغير الإحصائي مستمرة، تصنف هذه القيم على شكل فئات من R .

6. **الفئة - classe:** هي كل مجال من الشكل $[a\,;b[$ من R غالبا ما تكون الفئات متساوية الطول. طولها هو العدد الموجب $(b-a)$ و مركزها هو العدد $\frac{a+b}{2}$ و المدى هو $E=b-a$.
7. **التكرار المطلق – effectif (fréquence absolue) :** يمثل عدد المرات التي تتكرر فيه نفس القيمة x_i (قيمة المتغير الإحصائي) ونرمز له ب n_i مع $\sum n_i = N$ ، حيث N هو حجم العينة (أو المجتمع الإحصائي).
8. **التكرار النسبي- fréquence (relative) :** هو حاصل قسمة التكرار المطلق لكل قيمة على مجموع التكرارات، أي:

$$f_i=\frac{n_i}{\sum n_i}=\frac{n_i}{N} \text{ و } \sum f_i=1$$

9. **التكرار النسبي المئوي (fréquence) :** هو التكرار النسبي مضروبا في 100:

$$f_{i\%}=f_i\times 100=\frac{n_i}{\sum n_i}\times 100 \text{ و } \sum f_{i\%}=100\%$$

III. العرض الجدولي للبيانات (تبويب البيانات)

هي إحدى وسائل تصنيف أو تبويب البيانات الإحصائية على صورة جدول منتظم يوضح كيفية توزيع القيم التي حصلنا عليها من الظاهرة المدروسة.

جدول التوزيع التكراري

هو جدول بسيط يتكون من عمودين(سطرين) حيث يدل العمود (السطر) الأول على قيم الظاهرة (كيفيات الميزة)، ويدل العمود (السطر) الثاني على التكرار (المطلق) المقابل لهذه القيم.

قيم الميزة الإحصائية (الكيفيات) x_i	x_1	x_2	...	x_k	المجموع
التكرار المطلق n_i	n_1	n_2	...	n_k	N

مثال1: : تمثل البيانات التالية توزيع 20 مرضى حسب نوعية فصيلة الدم:

O A O A AB B O A A B A O O A B B A AB O A

المجتمع الإحصائي: جماعة ال 20 مرضى المستجوبة.

المتغير الإحصائي : نوعية فصيلة الدم. طبيعته: كيفي غير قابل للترتيب.

فصيلة الدم	A	B	O	AB	المجموع
عدد المرضى (التكرارات) n_i	8	4	6	2	20

مثال2: تمثل البيانات التالية عدد الساعات التي قضاها 15 طالب في إستعمال الانترنت في أسبوع:

5 6 9 7 5 4 8 10 7 8 11 10 9 8 9

المجتمع الإحصائي: جماعة ال15 الطلبة المستجوبة.

المتغير الإحصائي : عدد ساعات إستعمال الانترنت في أسبوع من طرف الطلبة المستجوبة. طبيعته: كمي متقطع.

عدد الساعات	4	5	6	7	8	9	10	11	Σ
عدد الطلبة (التكرار) n_i	1	2	1	2	3	3	2	1	15

❖ أما في حالة متغير كمي مستمر يتم بناء جدول إحصائي بإتباع الخطوات الآتية:

طريقة تحديد جدول التوزيع التكراري حالة متغير كمي مستمر

1. نحدد المجال (المدى) الذي تنتشر فيه البيانات، وهو الفرق بين أكبر قيمة للبيانات وأصغر قيمة لها، أي أن:

المدى = أكبر قيمة – أصغر قيمة.

2. نقسم المدى إلى فئات متساوية الطول k بحيث يكون عددها مناسبا (ما بين 5 و 15 فئة) وهناك عدة طرق لحساب عدد الفئات نذكر منها:

✓ معادلة ستيرجس Sturages : التي تنص على أن $k = 1 + 3.322\log(N)$ ، حيث N عدد البيانات.

✓ معادلة يول yule : التي تنص على أن $k = 2.5\sqrt[4]{N}$، حيث N عدد البيانات.

3. نحسب طول الفئة L وهو يساوي المدى مقسوما على عدد الفئات k

$$L = \frac{E}{k}$$

عند تحديد طول الفئة يجب مراعاة المعادلة المرنة:

طول الفئة × عدد الفئات ≥ المدى

$$E \leq k \times L$$

4. حساب مراكز الفئات c_i.

5. يكون الجدول من عمود (سطرين) يحتوي العمود (السطر) الأول على الفئات أما الثاني فيحتوي على مراكز الفئات و العمود (السطر) الثالث على التكرار (المطلق) المقابل لهذه الفئات.

مثال3: البيانات التالية متعلقة بأوزان 50 عاملا في مؤسسة ما .

165 162 150 154 155 162 155 164 158 152 171 163 <u>140</u> 157 162 171 158 164 163 159 153

158 150 160 160 149 158 152 165 156 164 157 159 158 159 153 163 158 <u>174</u> 162 156 151

160 158 162 166 162 164 158 153

- تحديد المدى E الذي تنتشر فيه البيانات :

$$E = 174 - 140 = 34$$

- حساب عدد الفئات

✓ معادلة ستيرجس Sturages : $k = 1 + 3.322\log(N) = 1 + 3.322\log 50 = 6.64 \approx 7$

✓ معادلة يول yule : $k = 2{,}5.\sqrt[4]{N} = 2{,}5.\sqrt[4]{50} = 6.65 \approx 7$

- حساب طول الفئة L وهو

$$L = \frac{E}{k} = \frac{34}{6.64} = 5.12 \approx 5$$

المعادلة المرنة هي: $34 \leq 7 \times 5 = 35$

الأطوال	[140,145[	[145,150[	[150,155[	[155,160[	[160,165[	[165,170[	[170,175[	Σ
مركز الفئة	142.5	147.5	152.5	157.5	162.5	167.5	172.5	
التكرار	1	1	9	17	16	3	3	50
f_i	0.02	0.02	0.18	0.34	0.32	0.06	0.06	1

ملاحظة: إن إختلاف طول الفئة لا يؤثر على الدراسة، لأننا في كل الحالات سواء إختيار الطول أو حسابه، لا نضيع شيئا من المعلومات وهذا هو المهم.

IV. التكرارات المجمعة

عندما نرغب في معرفة عدد المشاهدات التي تقل أو تزيد عن قيمة معينة نلجأ حينها إلى حساب :

1. التكرار المجمع الصاعد – Fréquences (effectifs) Cumulées Ascendantes

التكرار المجمع الصاعد $(N_i^{\uparrow})$ لأي قيمة (فئة) هو تكرار(المطلق أو النسبي) هذه القيمة (الفئة) مضافا إليه مجموع تكرارات القيم (الفئات) السابقة.

$$N_{i+1}^{\uparrow} = N_i^{\uparrow} + n_{i+1} = \sum_{k=1}^{i+1} n_k \ , N_1^{\uparrow} = n_1, N_2^{\uparrow} = n_1 + n_2 \ , \dots, N_r^{\uparrow} = n_1 + n_2 + \cdots + n_r = \sum_{k=1}^{k} n_k = N$$

2. التكرار المجمع النازل - Fréquences (effectifs) Cumulées Descendantes

التكرار المجمع النازل $(N_i^{\downarrow})$لأي قيمة (فئة) هو عبارة عن مجموع التكرارات (N) مطروحا منه تكرارات (المطلقة أو النسبية) القيم (الفئات) السابقة.

$$N_{i+1}^{\downarrow} = N - N_i^{\downarrow}\ , \qquad N_1^{\downarrow} = N, \qquad N_2^{\downarrow} = N - n_1\ , \dots, N_r^{\downarrow} = N - n_1 - n_2 - \cdots - n_{r-1} = n_r$$

مثال: (مثال 3)

Σ	[170,175[	[165,170[	[160,165[	[155,160[	[150,155[	[145,150[	[140,145[	الم الا
	172.5	167.5	162.5	157.5	152.5	147.5	142.5	c_i
50	3	3	16	17	9	1	1	n_i
1	0.06	0.06	0.32	0.34	0.18	0.02	0.02	f_i
	50	47	44	28	11	2	1	$N_i^{\uparrow}$
	1	0.94	0.88	0.56	0.22	0.04	0.02	$F_i^{\uparrow}$
	3	6	22	39	48	49	50	$N_i^{\downarrow}$
	0.06	0.12	0.44	0.78	0.96	0.98	1	$F_i^{\downarrow}$

III. العرض البياني للتكرارات (الأشكال البيانية)

1. حالة متغير نوعي :

أ. القطع الدائرية – Diagramme circulaire (secteurs angulaires)

هي دائرة مقسمة إلى عدة أجزاء، كل جزء يقابل زاوية مركزية Φ_i تتناسب مع التكرارات المقابلة لكل خاصية حيث:

$$\Phi_i = \frac{n_i}{N} \times 360° = f_i \times 360°$$

ب. الأعمدة المستطيلة (الشريط) - Tuyaux d'orgue

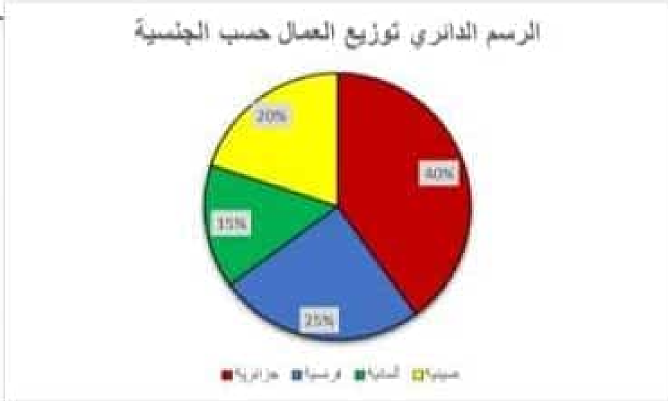
هي مجموعة من الأعمدة المتجاورة ذات قواعد متساوية إلا أن إرتفاعها يتناسب مع تكرار كل خاصية، كما أنها تكون متباعدة بمسافة متساوية.

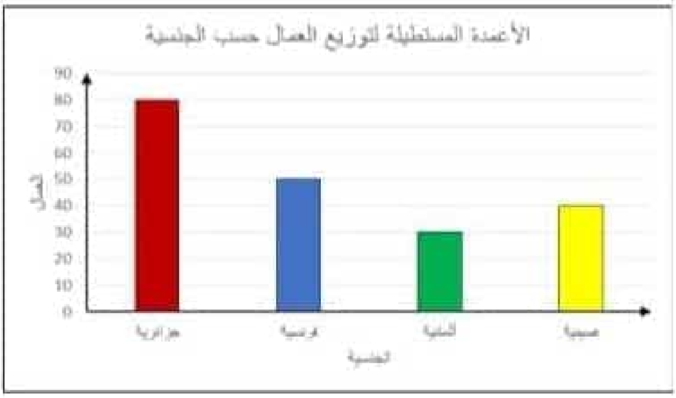
ج. العمود المجزأ – Diagrammes en barres

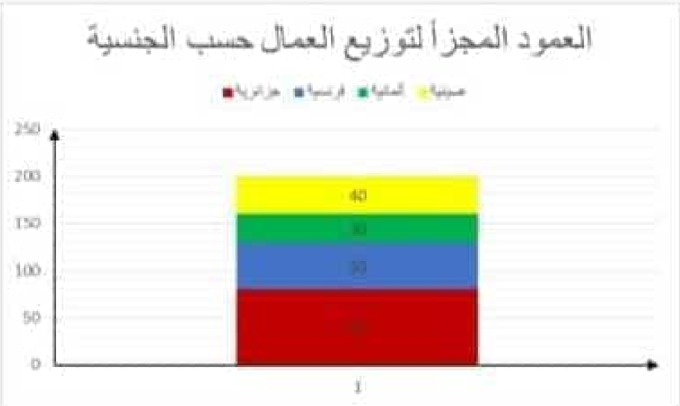
هو مستطيل مقسم إلى أجزاء، كل جزء يقابل تكرار معين لخاصية مدروسة.

مثال: توزيع العمال حسب الجنسية

النسبة المئوية	الزاوية	f_i	العدد n_i	الجنسية
40%	144°	0.40	80	جزائرية
25%	90°	0.25	50	فرنسية
15%	54°	0.15	30	ألمانية
20%	72°	0.20	40	صينية
100%	360°	1.00	200	Σ





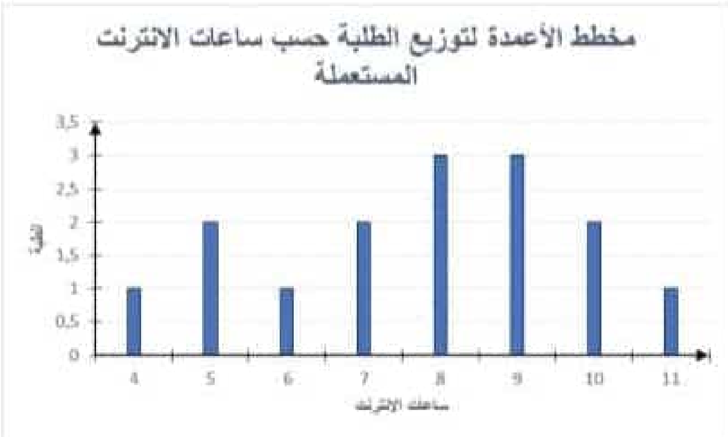


2. حالة متغير كمي متقطع:

الأعمدة البسيطة – Diagrammes en bâtons

مجموعة من الأعمدة البسيطة تتناسب أطوالها مع التكرار المقابل لكل قيمة.

مثال: (مثال 2)



3. حالة متغير كمي مستمر:

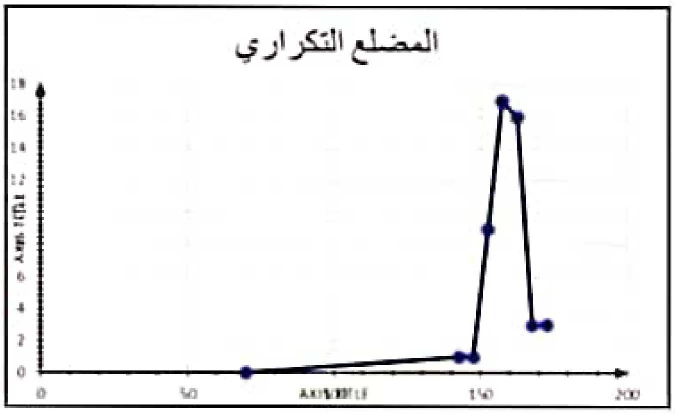
أ. المدرج التكراري – Histogramme

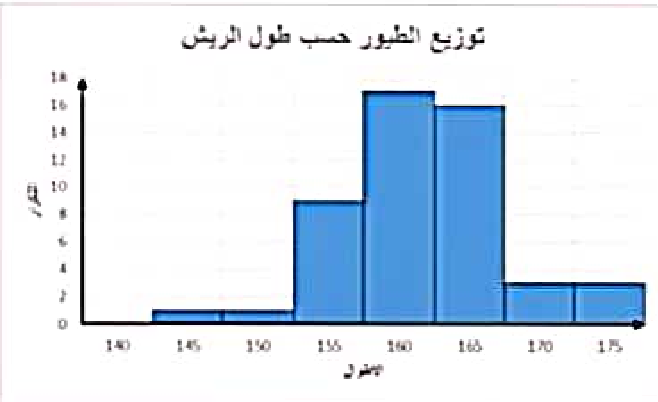
مجموعة من المستطيلات المتلاصقة، تتناسب أطوالها مع التكرار المقابل لكل فئة و قاعدة كل مستطيل تساوي طول الفئة المقابلة.

ب. المضلع التكراري – Polygone de fréquence

مجموع من قطع مستقيمة متصلة ومنكسرة تتحدد بنقاط إحداثياتها مركز الفئة و التكرارات المقابلة (c_i, n_i) .

مثال: (مثال3)



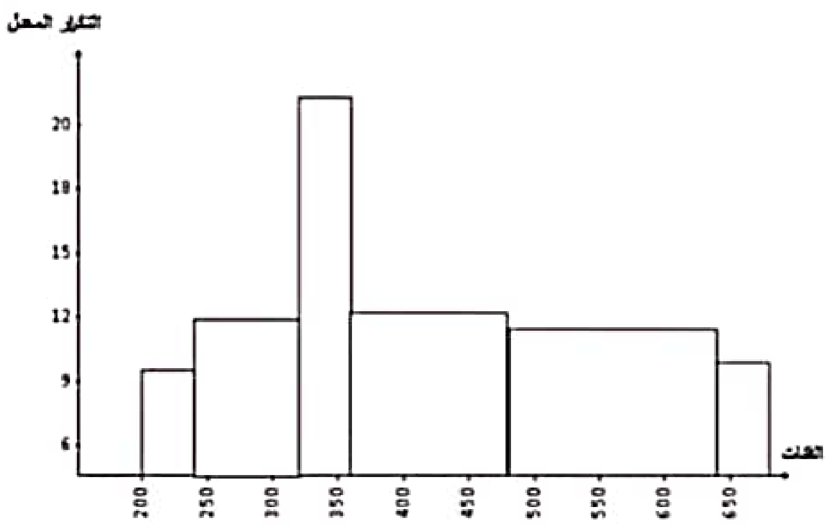


ملاحظة: إذا كانت فئات التوزيع غير متساوية، نقوم بتعديل التكرار بإستعمال المعادلة التالية:

$$\text{التكرار المعدل} = \frac{\text{تكرار الفئة}}{\text{طول الفئة}} \times \text{طول الفئة المختار}$$

مثال : يبين الجدول التالي توزيع عينة من 100 عامل حسب الأجر اليومي

فئة الأجر	عدد العمال	طول الفئة	التكرار المعدل
250 – 200	5	5	5
350 – 250	15	10	7.5
400 – 350	20	5	20
550 – 400	25	15	8.33
750 – 550	30	20	7.5
800 – 750	5	5	5
المجموع	100		



4. منحنى التكرارات المجمعة الصاعدة و النازلة

أ. متغير كمي متقطع

عبارة عن قطع مستقيمة متصاعدة (متنازلة) حسب تصاعد (تنازل) التكرار المجمع سواء كانت مطلقة أو نسبية لكل قيمة من قيم المتغير الإحصائي المدروس.

مثال:

عدد الساعات	4	5	6	7	8	9	10	11	Σ
n_i	1	2	1	2	3	3	2	1	15
$N_i^{\uparrow}$	1	3	4	6	9	12	14	15	
$N_i^{\downarrow}$	15	14	12	11	9	6	3	1	

ب. متغير كمي مستمر

يرسم منحنى التكرار المجمع الصاعد (النازل) عن طريق إيصال مجموعة النقاط ذات الإحداثيات التالية: الحدود العليا(الدنيا) للفئات و التكرار المجمع الصاعد (النازل) المقابل لها: $\{(a_i, n_i)\}$ (b_i, n_i). يرسم المنحنى بإيصال هذه النقاط بالمسطرة.

مثال:

الم الا	[140,145[	[145,150[	[150,155[	[155,160[	[160,165[	[165,170[	[170,175[
$F_i^{\uparrow}$	0.02	0.04	0.22	0.56	0.88	0.94	1
$F_i^{\downarrow}$	1	0.98	0.96	0.78	0.44	0.12	0.06