المحاضرة 5

ثانيا: التوزيع التكراري للمتغير الإحصائي المتصل (المستمر)

1- جدول التوزيع التكراري لمتغير إحصائي مستمر

ثانيا: التوزيع التكراري للمتغير الإحصائي المتصل (المستمر)

1- جدول التوزيع التكراري لمتغير إحصائي مستمر:

إذا كان المتغير الإحصائي من النوع المتصل رأينا أنه يقبل عددا غير متناهي من القيم الممكنة، المحصورة بين أصغر قيمة X_{min} وأكبر قيمة X_{max} وعليه يستحيل أن نمثله بجدول على شكل قيم فردية كما هو الحال في المتغير المنفصل، فنلجأ في هذه الحالة إلى تجميع أو تكثيف البيانات في مجموعات جزئية نسميها "**فئات**". فما هو عدد الفئات التي يمكن تحديدها وكيف؟

ليس هناك قاعدة نظرية لذلك، وإنما يشترط أن لا يكون عدد الفئات كثير جدا (يفوق 15 فئة) فيصبح الجدول ضخما يصعب تحليله وقراءته، أو يكون عدد الفئات قليل جدا (أقل من 5 فئات) فيصبح الجدول مبسط جدا أين يفقد حينها دقة وتفاصيل البيانات.

رغم ذلك فقد اجتهد بعض العلماء في تحديد قاعدة نظرية لإيجاد عدد الفئات، ومنهم العالم ستورجس (*Sturges*) الذي وضع قاعدة تجريبية لحساب طول الفئات، حيث تعتمد هذه القاعدة على مجال الدراسة وحجم المجتمع أو العينة.

أ- تحديد عدد الفئات K:

$$K = 1 + 1{,}322\ ln(n) \quad \text{أو} \quad K = 1 + 3{,}322\ log(n)$$

حيث: K تمثل عدد الفئات و n تمثل حجم العينة أو المجتمع

ب- تحديد أطوال الفئات C:

$$C = \frac{E}{K}$$

C : تمثل طول الفئة.

E: المدى العام وهو الفرق بين أكبر قيمة وأقل قيمة، أي: $E = X_{max} - X_{min}$

وعلى العموم فإن كل فئة تتميز بما يلي:

- كل فئة تتميز بحدين، حد أدنى وحد أقصى، هذه الحدود وبالخصوص الحد الأقصى يمكن أن يكون حدا فعليا أو غير فعلي، كأن نقول الفئة من أ إلى ب أي: $[أ - ب]$ ⇐ ب يعتبر حدا فعليا أي ينتمي للفئة، أو من أ إلى أقل من ب أي: $]أ - ب[$ ⇐ ب يعتبر حدا غير فعلي أي لا ينتمي للفئة.
- يمكن أن نلاحظ في بعض الجداول أن الفئة الأولى في الجدول غير محددة الحد الأدنى، كأن نقول مثلا 100 فما أقل أي (≤ 100)، كما أن الفئة الأخيرة قد تكون غير محددة الحد الأقصى كأن نقول مثلا 10000 فما أكثر (≤ 10000)، وهذا النوع من الفئات يطرح إشكالا في حساب مراكز الفئات.
- كل فئة تتميز بمجال أو طول وهو: $\Delta x_i = Lim_{sup} - Lim_{inf}$
- كل فئة تتميز بمركز يحسب كالتالي: $C_i = \frac{Li_{sup} + Lim_{inf}}{2}$
- إذا كانت أطوال الفئات متساوية فإن الفروق بين المراكز تساوي أطوال الفئات.

ملاحظة هامة: مهما تكن الطريقة المستعملة في تحديد عدد الفئات وأطوالها فإن المهم في كل ذلك هو أن لا يكون هناك تداخل بين الفئات، بحيث كل قيمة في السلسلة لا يمكن وضعها إلا في فئة واحدة.

كما أن جدول التوزيع التكراري لمتغير إحصائي مستمر قد يحتوي على التكرارات التالية:

- التكرار النسبي f_i والتكرار النسبي المئوي $f_i\%$.
- التكرار التجميعي الصاعد المطلق $N_i^{\uparrow}$ والنازل المطلق $N_i^{\downarrow}$
- التكرار التجميعي الصاعد النسبي $F_i^{\uparrow}$، والصاعد النسبي المئوي $F_i^{\uparrow}\%$.
- التكرار التجميعي النازل النسبي $F_i^{\downarrow}$، والنازل النسبي المئوي $F_i^{\downarrow}\%$.

ملاحظة: يتم حساب التكرارات السابقة بنفس الطريقة المذكورة في المتغير الكمي المتقطع.

مثال(2-6): البيانات التالية تمثل أوزان 60 طالبا بالكيلوغرام في أحد أقسام الـــ LMD بكلية العلوم الاقتصادية وعلوم التسيير بجامعة المسيلة:

67	64	68	73	73	54	61	74	60	78
80	74	65	63	60	69	72	66	77	65
74	50	76	69	68	66	78	63	70	55
67	67	64	76	61	72	72	57	65	77
59	71	79	78	58	63	74	66	73	67
61	71	69	68	73	81	64	61	84	55

المطلوب: أنشئ جدول التوزيع التكراري واحسب كلا من: $N_i^{\uparrow}$، $N_i^{\downarrow}$، f_i ، $f_i\%$، $F_i^{\uparrow}\%$، $F_i^{\downarrow}\%$ ، ثم إشرح كلا من:

n_2 ، $f_2\%$ ، $N_2^{\uparrow}$ ، $N_5^{\downarrow}$ ، $F_2^{\uparrow}\%$، $F_5^{\downarrow}\%$

الحل: أول خطوة نقوم بها هي ترتيب البيانات السابقة ترتيبا تصاعديا:

61	60	60	59	58	57	55	55	54	50
65	64	64	64	63	63	63	61	61	61
68	67	67	67	67	66	66	66	65	65
72	72	71	71	70	69	69	69	68	68
76	74	74	74	74	73	73	73	73	72
84	81	80	79	78	78	78	77	77	76

لدينا: $C = \frac{E}{K}$

- **حساب المدى:** $E = X_{max} - X_{min} = 84 - 50 = 34$
- **حساب عدد الفئات:** $K = 1 + 3{,}322\ log(n) = 6{,}9 \approx 7$ ، أي 7 فئات.
- **حساب طول الفئة:** $K = \frac{E}{K} = \frac{34}{7} = 4{,}92 \approx 5$

أي طول كل فئة يساوي 5 كلغ، ومنه تكون الفئات هي: $[50 - 55[$، $[55 - 60[$، ..... $[80 - 85[$.

نقوم بتفريغ البيانات في الفئات المذكورة فنحصل على الجدول الموالي:

الجدول(2-4): توزيع 60 طالبا حسب الوزن بالكيلوغرام في أحد أقسام الـ: LMD بكلية العلوم الاقتصادية - المسيلة

أوزان الطلبة X_i	عدد الطلبة n_i	C_i	f_i	$f_i\%$	$N_i^{\uparrow}$	$N_i^{\downarrow}$	$F_i^{\uparrow}\%$	$F_i^{\downarrow}\%$
$]55-50]$	2	52,5	0,033	3,3	2	60	3,3	100
$]60-55]$	5	57,5	0,083	8,3	7	58	11,6	96,7
$]65-60]$	12	62,5	0,2	20	19	53	31,6	88,4
$]70-65]$	16	67,5	0,268	26,8	35	41	58,4	68,4
$]75-70]$	14	72,5	0,233	23,3	49	25	81,7	41,6
$]80-75]$	8	77,5	0,133	13,3	57	11	95	18,3
$]85-80]$	3	82,5	0,05	5	60	3	100	5
المجموع	$\sum n_i = 60$	/	1	100	/	/	/	/

الشرح:

$n_2 = 5$: هناك 5 طلبة من بين 60 طالبا أوزانهم تتراوح بين 55 و 60 كلغ.

$f_2\% = 8{,}3\%$: هناك 8,3% من الطلبة أوزانهم تتراوح مابين 55 و 60 كلغ.

$N_2^{\uparrow} = 7$: هناك 7 طلبة من بين 60 طالبا أوزانهم أقل من 60 كلغ.

$N_5^{\downarrow} = 25$: هناك 25 طالبا من بين 60 طالبا أوزانهم أكبر أو يساوي 70 كلغ.

$F_2^{\uparrow}\% = 11{,}6\%$: هناك 11,6% من الطلبة أوزانهم أقل من 60 كلغ.

$F_5^{\downarrow}\% = 41{,}6\%$: هناك 41,6% من الطلبة أوزانهم أكبر أو تساوي 70 كلغ.