محاضرة حول: الجداول التكرارية

ترتيب الأرقام الإحصائية:

بعد جمع البيانات الإحصائية والتعبير عنها بشكل رقمي، ومن أجل تحديد درجة الدقة والأهمية فيها يجب علينا ترتيب هذه المعلومات الرقمية لاظهار خواصها وصفاتها.

ومن أبسط أنواع ترتيب الأرقام الإحصائية هو الترتيب التصاعدي، حيث نبدأ بأقل القيم ثم بالتي تليها وهكذا حتى ننتهي بأكبر قيمة بين القيم المعطاة. كما يمكن أن يكون الترتيب تنازليًا، حيث نبدأ بأكبر قيمة بين القيم ثم بالتي تليها حتى نصل إلى أصغر قيمة.

وفائدة الترتيب الصاعد أو النازل (الهابط) أنه يعطينا فكرة سريعة عن المدى الذي تتراوح فيه الأرقام بين أصغر وأكبر قيمة، كما يدلنا بصورة مبدئية على اتجاه التركيز في قيم المعلومات الإحصائية، إلاّ أن هذا الترتيب لا يبين لنا بوضوح مكان تركز القيم، ولا يدل على القيم التي تقيس تركز القيم. لهذا يعمد إلى توزيع البيانات الإحصائية المرتبة تصاعديًا أو تنازليًا على فئات معينة، ثم نحسب تكرار الوحدات الإحصائية في تلك الفئات. فكأننا في الواقع قسمنا المعلومات الإحصائية إلى أقسام متقاربة في الخاصية المدروسة، ثم نحسب البيانات التي تنتمي لكل قسم، وبذلك نحصل على شكل مختصر للبيانات.

إنشاء الجداول التكرارية :

من أجل إنشاء جدول توزيع تكراري لابد من تحديد المصطلحات المستعملة في الجدول بكل وضوح، وهذه المصطلحات هي:

1 . طول أو مدى الفئات.

2 . كيفية كتابة حدود الفئات.

3 . كيفية تفريغ البيانات وحساب التكرارات.

1 ـ تحديد طول الفئة :

نعني بطول أو مدى الفئة العدد الواقع بين الحد الأدنى والحد الأعلى للفئة. وعلى العموم ليس هناك طريقة معينة متفق عليها من طرف علماء الإحصاء لتحديد طول الفئة، إذ أن العملية تعود بالأساس إلى الباحث، فهو الذي يحدد الدقة المطلوبة من بحثه.

فهناك من الباحثين الذين يقومون بتقسيم البيانات التي جمعوها إلى عدد من الفئات، ومن ثم يمكن تحديد طول الفئات وذلك بقسمة المدى العام على عدد الفئات المفترض. ويرى البعض أنه يجب أن لا يتعدى عدد الفئات 15 فئة في جميع الحالات، ذلك أن الجداول التكرارية هي عملية إختصار للبيانات، وإذا زاد عدد الفئات عن هذا الحد فإن الجداول التكرارية تفقد أهميتها.

وهناك طريقة إستخلصها العالم الإحصائي ستورجز H.A.Sturges لإيجاد طول الفئة نلخصها في المعادلة التالية :

$$L = \frac{R}{K} = \frac{X_{Max} - X_{Min}}{1 + 3.322 \, Log \, N}$$

حيث : L يساوي مدى أو طول الفئة .

R المدى العام ، علمًا أن : R = XMax - Xmin .

K عدد الفئات ، علمًا أن : K = 1+3.322 Log N .

Xmax أكبر قيمة في البيانات .

Xmin أصغر قيمة في البيانات .

N عدد البيانات محل الدراسة .

ونشير هنا بأن قيمة L التي نحصل عليها تشمل جزء صحيح وجزء عشري، يستحسن تقريب الجزء العشري إلى وحدة صحيحة، وذلك وفقًا لما رأيناه في عملية التقريب في الفصل الأول.

2 ـ كيفية كتابة حدود الفئات :

بعد أن نحدد مدى الفئة يجب تعيين حدود الفئات بدقة من خلال إظهار حدا الفئة بشكل يمنع الالتباس والتداخل ، فلا يجوز مثلاً أن نكتب الفئات بالشكل التالي:

10 . 20

20 . 30

30 . 40

حيث نلاحظ وجود تداخل بين الفئات، فالوحدة الإحصائية التي قيمتها 20 هل تتبع الفئة الأولى أم الثانية؟ وكذلك بالنسبة للوحدات التي قيمتها 30 أو 40. ويمكننا منع هذا التداخل بطرق مختلفة، يمكن أن تأخذ الأشكال التالية :

10 . 19.9

20 . 29.9

30 . 39.9

أو بطريقة آخر كما هو موضح بالصورة التالية :

10.1 . 20

20.1 . 30

30.1 . 40

إلا أنه يؤخذ على هذا التحديد أنه في المتغيرات المستمرة قد يكون لدينا أجزاء أكثر دقة من المئة أو من أجزاء الألف بعد الفاصلة عكس ما رأينا سابقًا . لذلك فإننا نفضل إستخدام مفهوم المجالات لتوضيح حدود الفئات وذلك كما يلي :

فالمجال [a , b] هو مجال مغلق تنتمي إليه جميع القيم الواقعة بين a و b وكذلك كل من قيمتي a و b . أي $a \leq Xi \leq b$.

أما المجال] a , b [فهو مجال مفتوح من الطرفين تنتمي إليه جميع القيم الواقعة بين a و b ، ولا تنتمي إليه القيمتين a و b أي $a < Xi < b$.

والمجال [a , b] هو مجال نصف مفتوح أي مفتوح من طرف واحد وهو يشمل كافة القيم الواقعة بين a و b وكذلك القيمة a أما القيمة b فلا تنتمي إلى هذا المجال أي $a \leq Xi < b$.

واستنادًا إلى مفهوم المجالات يمكن كتابة حدود الفئات بالشكل التالي :

[10 ـ 20 [

[20 ـ 30 [

[30 ـ 40 [وهكذا ...

كما يمكن كتابة صيغة المجالات السابقة بالشكل التالي :

10 وأقل من 20

20 وأقل من 30

30 وأقل من 40 وهكذا ...

حيث أن الوحدة الإحصائية التي قيمتها 20 تنتمي إلى الفئة الثانية ولا تنتمي إلى الفئة الأولى. أما القيم الأقل من 20 فإنها تنتمي إلى الفئة الأولى. وكذلك القيمة 40 فإنها لا تنتمي إلى الفئة الثالثة وإنما تنتمي إلى الفئة اللاحقة لها.

أما الحد الأدنى للفئة الأولى فيكون عادة أصغر قيمة ضمن البيانات المعطاة، كما يستحسن أن يكون هذا الحد من مضاعفات الخمسة وذلك من أجل تسهيل الحسابات لا غير.

3 ـ كيفية تفريغ البيانات وحساب التكرارات :

بعد تعيين حدود الفئات وتحديد مداها ، نهيئ جدولاً بثلاثة أعمدة نكتب في العمود الأول الفئات وفي العمود الثاني تفريغ البيانات بشكل من الأشكال التي سنراها لاحقًا والعمود الثالث نسجل فيه التكرارات بالأرقام، ثم نقوم بتفريغ البيانات في الجدول، وتتم عملية تفريغ البيانات بعدة أساليب منها تسجيل البيانات في الجدول حسب الفئات التي تنتمي إليها القيم.

مثال: البيانات التالية تمثل أوزان 60 طالبا في أحد الأقسام بكلية العلوم الإقتصادية وعلوم التسيير، والمطلوم تهيئة هذه البيانات في جدول توزيع تكراري:

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

أول خطوة نقوم بها هي ترتيب البيانات السابقة ترتيبًا تصاعديًا كما هو مبين أدناه:

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

نلاحظ من خلال هذا الترتيب أن أصغر قيمة هي 50 كلغ وأكبر قيمة هي 84 كلغ ، ولكن هذا الترتيب لا يكفي لذلك لابد من جعل هذه البيانات في جدول توزيع تكراري . نحسب مدى الفئات وفق معادلة ستورجس كما يلي :

$$L = \frac{R}{K} = \frac{X_{Max} - X_{Min}}{1 + 3{,}322 \text{ Log } N} = \frac{84 - 50}{1 + 3{,}322 \log 60} = \frac{34}{1 + 3{,}322(1{,}778)} = 4{,}922 \approx 5$$

إذًا مدى الفئات يساوي 5 ولدينا أصغر قيمة هي 50 كلغ، فإذا أخذنا الحد الأدنى للفئة الأولى يساوي 50 يمكننا كتابة حدود الفئات كما يلي:

الفئـات	تفريـغ البيانات بالقيـم	التكـرار
50 وأقل من 55	54 ، 50	2
55 ، 60	59 ، 58 ، 57 ، 55 ، 55	5
60 ، 65	64 ، 64 ، 64 ، 63 ، 63 ، 63 ، 61 ، 61 ، 61 ، 60 ، 60	12
65 ، 70	69 ، 69 ، 68 ، 68 ، 68 ، 67 ، 67 ، 67 ، 67 ، 66 ، 66 ، 66 ، 65 ، 65	16
70 ، 75	74 ، 74 ، 74 ، 74 ، 73 ، 73 ، 73 ، 73 ، 72 ، 72 ، 72 ، 71 ، 71 ، 70	14
75 ، 80	79 ، 78 ، 78 ، 78 ، 77 ، 77 ، 76 ، 76	8
80 ، 85	84 ، 81 ، 80	3
المجمـوع		60

كما يمكن كتابة حدود الفئات وتفريغ البيانات بشكل آخر كما هو مبين في الجدول التالي :

الفئات	تفريغ البيانات بالخطوط المائلة	التكرار
[50 - 55 [	//	2
[55 - 60 [	////	5
[60 - 65 [	// //// ////	12
[65 - 70 [	/ //// //// ////	16
[70 - 75 [	//// //// ////	14
[75 - 80 [	/// ////	8
[80 - 85 [	///	3
المجموع		60

نلاحظ أن الفئة الأولى قيمها تتراوح ما بين 50 وأقل من 55 كلغ . تسمى 50 بالحد الأدنى للفئة الأولى و55 تسمى الحد الأعلى للفئة الأولى ، وعليه نسمي كل من 50 ، 55 ، 60 ، 65 ، 70 ، 75 ، 80 بالحدود الدنيا للفئات ونسمي 55 ، 60 ، 65 ، 70 ، 75 ، 80 ، 85 بالحدود العليا للفئات . إن الفرق بين الحد الأعلى والحد الأدنى لكل فئة يعطينا طول الفئة ، على سبيل المثال طول الفئة الأولى = الحد الأعلى – الحد الأدنى = 55 – 50 = 5 .

كما أن حاصل قسمة مجموع الحدين الأعلى والأدنى لكل فئة على 2 يعطينا مركز الفئة ، فمركز الفئة الأولى = (الحد الأعلى + الحد الأدنى) ÷ 2 = (55 + 50) ÷ 2 = 52.5 .

بالإضافة إلى التكرار العادي هناك أنواع أخرى من التكرارات يمكن شرحها فيما يلي :

إذا طرحنا السؤال التالي : ما هو عدد الطلبة الذين وزنهم أقل من 55 كلغ ؟ فالجواب بطبيعة الحال هو 2 . أما عدد الطلبة الذين يقل وزنهم عن 60 كلغ فهو 7 طلبة (2+5=7) وهكذا ... نسمي هذا النوع من الكرارات بالتكرار التجميعي الصاعد (ت. ت. ص) .

أما إذا طرحنا السؤال بشكل آخر وقلنا ما هو عدد الطلبة الذين وزنهم 50 كلغ فأكثر ؟ بما أن اقل وزن هو 50 كلغ ، فإن عدد الطلبة الذين وزنهم 50 كلغ فأكثر هو 60 طالب . أما عدد الطلبة الذين وزنهم 55 كلغ فأكثر فهو 58 طالب (60 – 2 = 58) ، وهكذا ... نسمي هذا النوع من التكرارات بالتكرار التجميعي الهابط (ت . ت . هـ) . والجدول التالي يوضح ذلك .

الفئات	التكرار	مراكز الفئات	ت . ت . ص	ت . ت . هـ
[50 – 55]	2	52.5	2	60
[55 – 60]	5	57.5	7	58
[60 – 65]	12	62.5	19	53
[65 – 70]	16	67.5	35	41
[70 – 75]	14	72.5	49	25
[75 – 80]	8	77.5	57	11
[80 – 85]	3	82.5	60	3
المجموع	60		-	-

تسمى التكرارات السابقة بالتكرارات المطلقة، ويمكن التعبير عنها بشكل نسب مئوية كما هو مبيّن في الجدول التالي:

الفئات	%التكرار	%ت . ت . ص	%ت . ت . هـ
[50 – 55]	3.3	3.3	100
[55 – 60]	8.3	11.6	96.7
[60 – 65]	20	31.6	88.4
[65 – 70]	26.8	58.4	68.4
[70 – 75]	23.3	81.7	41.6
[75 – 80]	13.3	95	18.3
[80 – 85]	5.0	100	5.0
المجموع	100	-	-

من خلال الجدول السابق يمكن القول على سبيل المثال أن نسبة الطلبة الذين تتراوح أوزانهم بين 60 وأقل من 65 كلغ هي 20 % من مجموع الطلبة، أما نسبة الطلبة الذين تقل أوزانهم عن 65 كلغ فهي 31.6 %، ونسبة الطلبة الذين أوزانهم 60 كلغ فأكثر هي 88.4 %