وزارة التعليم العالي والبحث العلمي
جامعة تكريت
كلية الإدارة والاقتصاد
قسم ادارة الاعمال
المرحلة الاولى

محاضرات في مادة الاحصاء

مدرس المادة

أ.م.د هادي مران احمد

1446 هـ **2024 مـ**

الفصل الثاني

التوزيعات التكرارية وأساليب عرض البيانات

تصنيف وتبويب البيانات :

بعد جمع البيانات فأن هذه البيانات التي تم الحصول عليها بخصوص ظاهرة معينة بشكلها الاولي تسمى بيانات خام raw data أو البيانات الأولية elementary data أو البيانات غير المصنفة ، وتكون هذه البيانات في الغالب غير منظمة وبالتالي يتعذر الاعتماد عليها لأغراض التحليل الإحصائي للوصول إلى النتائج المطلوبة ، لذلك فأن أولى الخطوات بعد جمع البيانات هي :

أ- مراجعة البيانات : وهي العملية التي تأتي بعد عملية جمع البيانات حيث يتم مراجعة وتدقيق البيانات.

ب- تصنيف البيانات : تبدأ هذه العملية بتصنيف البيانات على أساس الظواهر التي جمعت عنها فقد يكون للتصنيف على أساس العمر ، الجنس ، الوزن ... الخ.

ج- تبويب البيانات : ويقصد بالتبويب عملية تفريغ البيانات المصنفة في جداول خاصة بحيث أن كل جزء من البيانات المصنفة عن الظاهرة المعينة يعود إلى مستوى معين لتلك الظاهرة.

طبيعة البيانات الإحصائية والرموز الإحصائية :

عند توفر بيانات حول ظاهرة ما فأننا نرمز للظاهرة بالرمز (Y) وكل مفردة أو مشاهدة منها نرمز لها بـ (y) فمثلاً عند دراسة عدد الشركات في القطاع الصناعي فأننا نرمز لصفة القطاع الصناعي بالرمز (Y) وأي شركة تابعة له بالرمز (y_i) وتسمى y_i المشاهدة أو المفردة ، هذا وأن قيمة (y_i) قد تختلف من منشأة إلى أخرى ولهذا نقول أن (Y) متغير.

المتغيرات العشوائية Random variable :

يعرف المتغير العشوائي بأنه دالة ذات قيمة حقيقية real – valued function معرفة على مجال العينة sample space وغالباً ما يرمز للمتغير العشوائي بأحد الاحرف الكبيرة مثل X , Y , Z ... الخ ، ولقيم المتغير عند تنفيذ التجربة بأحد الأحرف الصغيرة مثل x , y , z ... الخ.

<u>وتقسم المتغيرات إلى :</u>

1- المتغيرات النوعية (الوصفية): وهي المتغيرات التي لا يمكن قياسها بوسائل قياس مألوفة كالعدد أو التقييس أنما تشكل صفات لذلك المتغير مثل الحالة الإنتاجية لإحدى الشركات (رديئة ، متوسطة ، جيدة ، جيدة جداً) وغيرها من الأمثلة.

2- المتغيرات الكمية: وهي المتغيرات التي يمكن قياسها بوسائل قياس مثل الموظفين في وزارة معينة وهذا النوع من المتغيرات يكون على نوعين أهمها :

أ- المتغيرات المستمرة:

إذا كانت مجموع القيم الممكنة للمتغير (x) مجموعة غير قابلة للعد سواء كانت مجموعة محدودة أو غير محدودة عندئذ يقال أن (x) متغير عشوائي مستمر ، بمعنى أن المتغير المستمر هو المتغير الذي تأخذ لمشاهدة أو المفردة فيه أي قيمة رقمية في مدى وجيز ، فلو فرضنا أطوال طلبة كلية معينة تتراوح بين 130,5–180,5 فعندئذ يقول بأن $130,5 \leq x \leq 180,5$ أي أن المتغير (x) يمكن ان يأخذ أي قيمة بين 130,5 و 180,5 وهنالك أمثلة أخرى مثل الوزن وكمية المحصول والزمن لأنها ممكن قياسها بأجزاء صغيرة جداً.

ب- المتغيرات المتقطعة أو المنفصلة:

إذا كانت مجموعة القيم الممكنة للمتغير (x) مجموعة قابلة للعد سواء كانت مجموعة محددة أو غير محددة عندئذ يقال أن (x) متغير عشوائي متقطع.

الخطوات العامة في إنشاء جدول التوزيع التكراري :

لتكوين جدول التوزيع التكراري يجب إتباع ما يلي :

1- استخراج المدى :

المدى = أكبر قيمة – أقل قيمة + 1

$$T.R = x_L - x_S + 1$$

2- اختيار وتحديد عدد الفئات : تمثل عدد الفئات عدد المجاميع التي يتألف منها التوزيع التكراري ، وهناك صيغ تقريبية يمكن من خلالها تحديد عدد فئات التوزيع أهمها :

أ- صيغة يول Yule وهي :

$$m = 2.5 \sqrt[4]{n}$$

حيث أن n تمثل عدد المفردات.

ب- صيغة سترجس وهي :

عدد الفئات = m

$$m = 1 + 3 . 322 \log_{10}(n)$$

وعند التطبيق يتم تقريب الناتج لأقرب عدد صحيح.

لكل من الطريقتين مزايا وعيوب ولكن سوف تستخدم في العمل الحسابي عدداً من الفئات لا يقل عن 5 ولا يزيد عن 15.

3- إيجاد طول الفئة :

$$\text{طول الفئة} = \frac{\text{المدى}}{\text{عدد الفئات}}$$

$$L = \frac{T.R}{m}$$

ويستحسن أن تكون أطوال الفئات متساوية.

مثال : البيانات التالية تمثل أعمار 40 موظف في إحدى الشركات التابعة لوزارة التجارة ، المطلوب : تكوين جدول التوزيع التكراري لهذه البيانات :

19 , 31 , 22 , 22 , 27 , 28 , 28 , 28 , 26 , 25 , 26 , 25 , 26 , 26 , 24 , 22 , 25 , 24 , 24 , 22 , 24 , 24 , 24 , 23 , 21, 22 , 22 , 29 , 29 , 24 , 24 , 19 , 20 ,24 , 24 , 24 , 24 , 24 , 24 24 , 24 , 24

الحل :/ الخطوة الأولى للحل نرتب القيم تصاعديا لسهولة الحل :

19 , 19 , 20 , 21 , 22 , 22 , 22 , 22 , 22 , 23 , 24 , 24 ,24 ,24 ,24 ,24 ,24 ,24 , 25 , 25 , 25 , 26 , 26 , 26 , 26 , 27 , 28 , 28 , 24 ,24 ,24 ,24 ,24 ,24 ,24 ,24 ,24 ,24 , 28 , 29 , 29 , 31

2- المدى : $T.R = x_L - x_S + 1$

$= 31 - 19 + 1 = 13$

3- عدد الفئات : $m = 1 + 3.322 \log(n)$

$= 1 + 3.322 \log(40) = 1+3.322 (1.6) = 1 + 5.322 = 6.322 = 6$ بالتقريب

4- طول الفئة :

$$L = \frac{T.R}{m} = \frac{13}{6} = 2.16 = 2$$ بالتقريب

1- الحد الأدنى للفئة الأولى يمثل أصغر قيمة حسب المثال هنا = 19

2- يضاف إلى الحد الأدنى طول الفئة 21 = 2 + 19

3- ويستمر الجدول الحد الأعلى للفئة الأولى يكون الحد الأدنى للفئة الثانية يضاف إليه طول الفئة يكون الحد الأعلى.

عدد الفئات	التكرار	f_i
19 – 21	III	3
21 – 23	IIIIIII	7
23 – 25	IIIIIIIIIIIIIII	16
25 – 27	IIIIIII	7
27 – 29	IIII	4
29 – 31	III	3
المجموع		40

الطرق المختلفة في كتابة حدود الفئات :

نفرض لدينا الجدول التالي :

الفئات	الفئات	الفئات
40 – 49	40 -	40 – 50
50 – 59	50 -	50 – 60
60 – 69	60 – 70	60 – 70

جدول توزيع تكراري متقطع جدول توزيع مستمر جدول توزيع تكراري متصل أو مستمر

جدول التوزيع التكراري حسب نوع المتغير :

1- جدول التوزيع التكراري في حالة المتغيرات العشوائية المستمرة :

في هذه الحالة تكتب حدود الفئات كما هو موضح في المثال التالي :

مثال 2 : البيانات التالية تمثل أوزان (كغم) عينة من طلبة إحدى الكليات قوامها (37) طالب ، المطلوب : كون جدول التوزيع التكراري لهذه البيانات :

100 , 65 , 70 , 83 , 94 , 55 , 46 , 47.8 , 62.3 , 77.2 , 61.3 , 62.9 , 66.5 , 66.9 , 68.3 , 70 , 73.2 , 73.8 , 82.4 , 58.2 , 59.9 , 54.4 , 58.7 , 52.6 , 51.8 , 74.1 , 80.2 , 75.1 , 78.3 , 58.5 , 62 , 66.3 , 65 , 67.1 , 88.2 , 89.1 , 80.1 , 76.3

الحل :

1- نجد المدى $T.R = x_L - x_S + 1$

$= 100 - 46 + 1 = 55$

2- نجد عدد الفئات $m = 2.5 \sqrt[4]{n}$

3- طول الفئه $= 2.5 \sqrt[4]{38} = 6.16 = 6$

$L = \frac{T.R}{m} = \frac{55}{6} = 9$

الفئات	التكرار
46 – 55	5
55 – 64	9
64 – 73	8
73 – 82	9
82 – 91	4
91 – 100	2

2- في حالة المتغيرات العشوائية المتقطعة :

في هذه الحالة تكتب حدود الفئات حسب التوزيع التكراري التالي ، بحيث تضمن أن كل قيمة من قيم البيانات تضمن في فئة واحدة من فئات التوزيع دون أي تكرار قد يحصل في هذه الفئات أو تلك ، بحيث أن طول الفئة L يكون مساوياً للفرق ما بين الحد الأعلى للفئة وحدها الأدنى مضاف للفرق العدد واحد.

مثال 3 : البيانات التالية تمثل عدد الموظفين الفنيين والإداريين والخدميين لـ 60 منشأة تابعة لوزارة الصناعة والتجارة.

المطلوب : تبويب هذه البيانات في جدول تفريغ تكراري :

60 , 76 , 80 , 120 , 132 , 82 , 90 , 65 , 68 , 73 , 150 , 142 , 157 , 164 , 88 , 90 , 98 , 101 , 103 , 110 , 119 , 116 , 120 , 126 , 109 , 114 , 120 , 122 , 111 , 116 , 90 , 78 , 93 , 95 , 98 , 104 , 120 , 113 , 121 , 119 , 125 , 126 , 130 , 131 , 136 , 118 , 142 , 148 , 150 , 154 , 122 , 123 , 139 , 125 , 156 , 154 , 136 , 137 , 110 , 136

1- T.R = 164 – 60 + 1 = 105 المدى

2- m = 2.5 $\sqrt[4]{\boldsymbol{n}}$ = 2.5 $\sqrt[4]{\boldsymbol{60}}$ = 2.5 (2.783) عدد الفئات

m = 6.958 = 7

3- L = $\frac{\boldsymbol{T.R}}{\boldsymbol{m}}$ = $\frac{\boldsymbol{105}}{\boldsymbol{7}}$ = 15 طول الفئة

ترتيب القيم تصاعدياً :

60 , 65 , 68 , 73 , 76 , 78 , 80 , 82 , 88 , 90 , 90 , 90 , 93 , 95 , 98 , 98 , 101 , 103 , 104 , 109 , 110 , 110 , 111 , 113 , 114 , 116 , 116 , 118 , 119 , 119 , 120 , 120 ,120 ,120 , 121 , 122 , 122 , 123 , 125 , 125 , 126 , 126 , 130 , 131 , 132 , 136 ,136,136, 137 , 139 , 142 , 142 ,148,150, 150 , 154 , 154 , 156 , 157 , 164

الفئات	التكرار	التكرار
60 – 74	IIII	4
75 – 89	IIIII	5
90 – 104	IIIIIIIII	10
105 – 119	IIIIIIIIII	11
120 – 134	IIIIIIIIIIIIII	15
135 – 149	IIIIIII	8
150 – 164	IIIIIII	7
		60

ملاحظات على تكوين الجداول التكرارية :

1- عند تحديد عدد الفئات يفضل أن يكون العدد مناسب وعلى هذا يجب أن توازن دائماً عند تحديد الفئات بين الدقة بالنتائج والسهولة بالعمل الحسابي.

2- يفضل جعل فئات الجداول متساوية في الطول.

3- عند كتابة الفئات في الغالب نبدأ بأقل مفردة لدينا كحد أدنى للفئة الأولى ويجوز في بعض الأحيان أن نبدأ بقيمة أقل عن قيمة أقل مفردة كحد أدنى للفئة الأولى ، وينتهي الجدول بقيمة أكبر مفردة كحد أعلى للفئة الأخيرة ويجوز أن ينتهي بقيمة أكبر من هذه القيمة ولا يجوز أن ينتهي بقيمة أقل من قيمة أكبر مفردة.

4- يفضل قدر الإمكان تجنب الجداول المفتوحة حيث يفضل عليها الجداول المغلقة.

5- يسمى الجدول مقفلاً إذا كان الحد الأعلى للفئة الأولى معروفاً وكذلك الحد الأعلى للفئة الأخيرة معروف أيضاً أما الجدول المفتوح فيكون على الحالات التالية :

أ- مفتوحاً من طرفين إذا كان الحد الأدنى للفئة الأولى غير معلوم وكذلك الحد الأعلى للفئة الأخيرة غير معلوم أيضاً.

ب- يكون مفتوحاً من طرفه الادنى فقط إذا كان الحد الأدنى للفئة الأولى غير معلوم فقط.

ج- يكون مفتوحاً من طرفه الاعلى فقط إذا كان الحد الاعلى للفئة الاخيره غير معلوم فقط.

الفئات	الفئات	الفئات	الفئات	الفئات
40 - 50	40 – 50	أقل من 40	أقل من 40	30 – 40
50 – 60	50 – 70	40 – 50	40 – 50	40 – 50
60 – 70	70 – 80	50 – 60	50 – 60	50 – 60
70 – 80	80 – 100	60 – 70	60 – 70	60 فأكثر
جدول منتظم	جدول غير منتظم	70 فأكثر مفتوح من الطرفين	مفتوح من الطرف الأدنى فقط	مفتوح في الطرف الاعلى فقط

التكرار النسبي والمئوي

ان جدول التوزيع التكراري النسبي هو جدول يبين الأهمية النسبية لكل فئة حيث :

التكرار النسبي = $\frac{\text{تكرار تلك الفئة}}{\text{المجموع الكلي التكرارات}}$

التكرار النسبي = $\frac{fi}{\sum_{i=1}^{m} fi}$

اما التكرار المئوي فانه يبين الاهمية المئوية لكل فئة :

التكرار المئوي = $\frac{\text{تكرار تلك الفئة}}{\text{المجموع الكلي التكرارات}} \times 100$

التكرار المئوي = $\frac{fi}{\sum_{i=1}^{m} fi} * \mathbf{100}\%$

وعادة يصبح التكرار النسبي تكرار مئوي وذلك بضربه في 100%

التكرارات المئوي = $\frac{\text{تكرار الفئة}}{\text{مجموع التكرارات}} \times 100$

مثال 5 :

للجدول التالي جد التكرار النسبي والمئوي لكل فئة :

الفئات	Fi	التكرار النسبي	التكرار المئوي
31 – 40	1	$= 0.0125\frac{\mathbf{1}}{\mathbf{80}}$	$= 0.0125 * 100 = 1.25\frac{\mathbf{1}}{\mathbf{80}}$
41 – 50	2	$= 0.0250\frac{\mathbf{2}}{\mathbf{80}}$	$= 0.025 * \mathbf{100} = 2.5\frac{\mathbf{2}}{\mathbf{80}}$
51 – 60	5	$= 0.0625\frac{\mathbf{5}}{\mathbf{80}}$	$0.06\ 25 * \mathbf{100} = 6.25$
61 – 70	15	0.187	18.7
71 – 80	25	0.3125	31.25
81 – 90	20	0.25	25
91 – 100	12	0.15	15
المجموع	80	1	100

التكرار النسبي = $\frac{fi}{\sum_{i=1}^{m} fi}$

ملاحظة الرمز $\sum$ يعني المجموع الكلي.

على سبيل المثال $= \frac{\mathbf{1} \quad \boldsymbol{fi}}{\mathbf{80}\sum_{\boldsymbol{i=1}}^{\boldsymbol{m}} \boldsymbol{fi}}$

التكرار المئوي = $\frac{\text{تكرار كل فئة}}{\text{المجموع الكلي التكرارات}} \times 100\%$

$= \frac{\mathbf{1}}{\mathbf{80}} * 100\% = = 0.0125 * 100\% = 1.25\%$

مادة الإحصاء