الفصـل الثـاني

أساليب عرض البيانات

أولا: العرض الجدولي للبيانات:

1- التوزيع التكراري Frequency Distribution:

وهو عبارة عن تلخيص وترتيب لبيانات المتغير العشوائي التي سبق أن جمعت وصنفت مقسمة إلى عدد من المجاميع كل منها تسمى بـ (الفئة class). هذه الفئات قد تكون مرتبة تصاعدياً أو تنازلياً حسب طبيعة البيانات ويسمى توزيع عدد قيم X حسب الفئات بـ (التوزيع التكراري). وقد تكون فئات التوزيع التكراري متساوية في الطول أو غير متساوية وذلك يعتمد على طبيعة الدراسة ومتطلباتها.

وتقسم البيانات إلى نوعين:

1- البيانات الغير مبوبة: وهي البيانات الأولية أو الأصلية (الخام) التي جمعت ولم تبوب.

2- البيانات المبوبة: وهي البيانات التي بوبت ونظمت في جدول توزيع تكراري.

لشرح مكونات التوزيع التكراري لدينا المثال الافتراضي الآتي:

إذا كانت $X_1, X_2, ..., X_n$ تمثل بيانات المتغير العشوائي X من عينة عشوائية من المفردات قوامها n مفردة. ونرغب في تلخيص هذه البيانات في توزيع تكراري عدد فئاته هو m.

لنفرض أن X_S هو أصغر قيمة و X_L هو أكبر قيمة في مجموعة البيانات هذه عندئذ تعرف مكونات التوزيع على النحو الآتي:

1- المدى الكلي للتوزيع Total range:

وهو الفرق مابين أكبر قيمة وأصغر قيمة في المجموعة مضافاً له العدد واحد. ويرمز له بالرمز $T.R$ وعليه فإن

$$T.R = X_L - X_S + 1$$

2- عدد فئات التوزيع Number of classes:

وهي عدد المجاميع التي يتألف منها التوزيع التكراري. وهنالك صيغ تقريبية لإيجادها أهمها:

أ. صيغة يول Yule:

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

إذا إن n تمثل عدد المفردات (المشاهدات).

ب. صيغة سترجس Sturges:

$$m = 1 + 3.322 \log_{10} n$$

وعند التطبيق يتم تقريب الناتج لأقرب عدد صحيح.

الممسوحة ضوئيا بـ CamScanner

3- طول الفئة Length of a class:

وهو مقدار سعة الفئة، أي مقدار المسافة مابين الحد الأدنى للفئة وحدها الأعلى ويسمى أحياناً المدى الفئوي. ويرمز لطول الفئة بالرمز L وعليه يمكن تحديد قيمة L من خلال الصيغة الآتية:

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

4- الحد الأدنى والحد الأعلى للفئة Lower and upper bound of a class:

لكل فئة من فئات التوزيع التكراري بداية ونهاية. فالبداية هي الحد الأدنى للفئة والنهاية هي الحد الأعلى لها.

- في حالة المتغيرات المستمرة فإن الحدود تكون كالآتي

تسلسل الفئة	الحد الأدنى	الحد الأعلى
1	Xs	Xs+L
2	Xs+L	Xs+2L
3	Xs+2L	Xs+3L
.	.	.
m	Xs+(m-1)L	Xs+mL

- في حالة المتغيرات المتقطعة فالحدود هي

تسلسل الفئة	الحد الأدنى للفئة	الحد الأعلى للفئة
1	Xs	Xs+L-1
2	Xs+L	Xs+2L-1
3	Xs+2L	Xs+3L-1
.	.	.
.	.	.
.	.	.
M	Xs+(m-1)L	Xs+ML-1

5- مركز الفئة Center of a class:

وهو قيمة من قيم المتغير العشوائي X التي تتوسط المسافة بين الحد الأدنى والحد الأعلى للفئة. فإذا رمزنا للحد الأدنى بالرمز $L.L$ والحد الأعلى بالرمز $U.L$ ولمركز الفئة بالرمز X فإن

$$X = \frac{L.L + U.L}{2}$$

الممسوحة ضوئيا بـ CamScanner

6- تكرار الفئة Class frequency:

وهو جزء من مفردات العينة التي تتصف بكونها تقع من حيث القيمة العددية مابين حدي الفئة بحيث أن مجموع هذه الأجزاء يشكل عدد مفردات العينة n. فإذا رمزنا لتكرارات الفئات بالرموز $f_1, f_2, ..., f_m$ فإن

$$f_1+f_2+f_3+...+f_m=n$$

ملاحظة:

التوزيع التكراري قد يكون توزيع مغلق أو مفتوح وذلك يعتمد على طبيعة الدراسة. والتوزيع المغلق هو ذلك التوزيع الذي يمتلك حداً أدنى للفئة الأولى وحداً أعلى للفئة الأخيرة أما التوزيع المفتوح هو التوزيع الذي لا يمتلك حداً أدنى أو حداً أعلى للفئة الأخيرة أو كليهما معاً.

مثال:

البيانات التالية تمثل عدد أشجار البرتقال المملوكة من قبل 60 عائلة فلاحية. المطلوب تبويب هذه البيانات في جدول توزيع تكراري.

60,76,80,120,132,82,90,65,68,72,150,142,157,164,88,90,98,101,103,110,
119,116,120,126,109,114,120,122,111,116,90,78,93,95,98,104,120,113,
121,119,125,126,130,131,136,118,120,142,150,154,122,123,139,125,156
,154,136,137,110,136.

الحل:

1- إن اصغر قيمة في المجموعة هو العدد 60 وان اكبر قيمة فيها هو العدد 164 وبذلك فإن المدى الكلي لهذه المجموعة هو

$$T.R=164-60+1=105$$

2- نستخدم صيغة Yule في تحديد عدد فئات التوزيع وعليه فإن

$$m=2.5\sqrt[4]{n}=2.5\sqrt[4]{60}=2.5(2.783)$$

$$\therefore m=6.958$$

إذ إن عدد الفئات يجب أن يكون عدد صحيح لذا يتم تقريب قيمة m إلى اقرب عدد صحيح وهو 7 وعليه فإن عدد فئات التوزيع التكراري هي 7.

3- طول الفئة يكون وفق الآتي:

$$L=\frac{T.R}{m}=\frac{105}{7}=15$$

الممسوحة ضوئيا بـ CamScanner

تسلسل الفئات	الفئات (عدد الأشجار)	مركز الفئة	التكرار (عدد العوائل)
1	60-74	67	4
2	75-89	82	5
3	90-104	97	10
4	105-119	112	12
5	120-134	127	16
6	135-149	142	7
7	150-164	157	6
المجموع			60

مثال:

البيانات التالية تمثل أوزان (50) طالبا في احدى المدارس والمطلوب إنشاء جدول توزيع تكراري؟

29 47 56 49 37 41 47 45 53 29 57 49 54 19 38 44 24 46 43 57
29 34 37 18 21 39 28 45 42 22 34 49 43 28 45 42 52 51 32 31
28 37 32 27 26 41 39 43 35 23.

الحل:

1- إيجاد المدى الكلي

$$T.R = X_L - X_S + 1 = 57 - 18 + 1 = 40$$

2- تحديد عدد الفئات حسب طريقة سترجس

$$M = 1 + 3.322 \log(n)$$
$$= 1 + 3.322 \log(50)$$
$$= 1 + 3.322(1.69) = 1 + 5.614 = 6.614 \approx 7$$

3- طول الفئة يكون وفق الآتي:

$$L = \frac{T.R}{m} = \frac{40}{7} = 5.7 \simeq 6$$

نحدد الحدود الدنيا والعليا للفئات، وبما أن المتغير من النوع المستمر بذلك سوف نستخدم الجدول الأول لتحديد حدود الفئات:

الحد الأدنى الأول	$X_S = 18$
الحد الأعلى الأول	$X_S + L = 18 + 6 = 24$
الحد الأدنى الثاني	$X_S + L = 18 + 6 = 24$
الحد الأعلى الثاني	$X_S + 2L = 18 + 2*6 = 30$
الحد الأدنى الثالث	$X_S + 2L = 18 + 2*6 = 30$
الحد الأعلى الثالث	$X_S + 3L = 18 + 18 = 36$
الحد الأدنى الرابع	$X_S + 3L = 18 + 18 = 36$

الممسوحة ضوئيا بـ CamScanner

الحد الأعلى الرابع $X_S+4L = 18+24 = 42$

بذلك سيكون الجدول كما يلي:

الفئات	تفريغ التكرارات	f_i	طريقة ثانية للفئات	مركز الفئة
18-24	|||||	5	18-	21
24-30	|||| |||||	9	24-	27
30-36		||||||	6	30-	33
36-42	||| |||||	8	36-	39
42-48	|| |||||| |||||	12	42-	45
48-54		||||||	6	48-	51
54-60	||||	4	54-	57
المجموع		50		

2- التوزيع التكراري النسبي Proportionate frequency distribution:

هو توزيع تكراري اعتيادي فيه التكرارات معبر عنها بنسب مئوية يمكن الحصول عليها من قسمة تكرار كل فئة على مجموع التكرارات الكلية. فإذا كانت التكرارات المقابلة للفئات وحسب تسلسلها $f_1, f_2, ..., f_m$ عندئذ فإن التكرارات النسبية المئوية المقابلة لهذه الفئات ستكون $f^*_1, f^*_2, ..., f^*_m$ إذ إن

$$f^*_i = \frac{f_i}{n} \bullet 100 \qquad i = 1,2,3,....,m$$

وإن n تمثل مجموع التكرارات.

مثال:

لنفس بيانات المثال السابق، المطلوب تكوين التوزيع التكراري النسبي المئوي.

الفئات	التكرار	التكرار النسبي %
60-74	4	6.666%
75-89	5	8.333%
90-104	10	16.666%
105-119	12	20%
120-134	16	26.666%
135-149	7	11.666%
150-164	6	10%
المجموع	60	100%

الممسوحة ضوئيا بـ CamScanner