العرض الجدولي للبيانات الاحصائية :

بعد عملية جمع البيانات الإحصائية وفرزها تصبح لدينا مشاهدات على شكل معطيات أو بيانات منفردة أو غير مبوبة، يكون عددها كبير جدا يصعب علينا استيعابها أو استخلاص النتائج منها ما يجعلنا في أمس الحاجة إلى تنظيمها وتبويبها حتى نتمكن من التعامل معها ودراستها والاستفادة منها، فنستعين بعملية تنظيم البيانات في جداول إحصائية وتصنيفها إلى مجموعات متجانسة وتبويبها في جداول تلخيصية.

يقصد بالعرض الجدولي للبيانات الإحصائية وضع البيانات المتحصل على الظاهرة محل الدراسة في جداول خاصة تسمى بجداول التوزيعات التكرارية. يتكون الجدول التكراري من عمودين أساسيين (سطرين) يوضع في العمود الأول (السطر الأول) قيم الخاصية المدروسة والتي نرمز لها بالرمز X_i ، وهي تعبر عن قيم المتغير الإحصائي المدروس، وتكون هذه القيم على شكل صفات، قيم نقطية أو على شكل مجالات. أما العمود الثاني (السطر الثاني) فيخص عدد الوحدات الإحصائية المقابلة لكل قيمة من قيم المتغير الإحصائي ويطلق على عدد الوحدات الإحصائية بالتكرار ونرمز له بالرمز n_i.

I–جداول التوزيعات التكرارية (الجداول الإحصائية):

تختلف الجداول الإحصائية باختلاف نوع البيانات أي حسب نوع المتغير الإحصائي، بين المتغير الإحصائي الكيفي والكمي وإذا كان كمي بين المتغير المنفصل والمتغير المتصل. إذا لا بد من تحديد أولا نوع المتغير قبل تبويب البيانات.

I –1–الجدول التكراري لمتغير كيفي:

يتكون الجدول التكراري لمتغير كيفي من عمودين (أو سطرين)، يحتوي العمود الأول على رموز كتابية للخاصية المدروسة ويحتوي الثاني على تكرارات كل رمز كتابي.

مثال (1–1): أخذت عينة إحصائية تتكون من 200 شخص مغترب في بلد أوروبي لمعرفة جنسية كل مغترب، فتحصلنا على النتائج المدونة في الجدول الإحصائي التالي:

الجدول رقم (1–1): توزيع المغتربين حسب الجنسية

الجنسية (X_i)	تونسية	مغربية	جزائرية	خليجية	مصرية	برتغالية	نيجيرية	المجموع
التكرار (n_i)	40	30	35	25	20	15	35	200

I–2–الجدول التكراري لمتغير كمي منفصل: (منقطع)

المتغير الكمي المنفصل أو المنقطع هو ذلك المتغير الذي لا يمكن تجزئة قيمه، حيث يأخذ قيم صحيحة.

مثال (1–2): لدراسة متوسط عدد أطفال في العائلة لبلد ما، أخذت عينة من 100 أسرة فكانت النتائج كما يلي:

الجدول رقم (1–2): عدد الأطفال في الأسرة

عدد الأطفال (X_i)	0	1	2	3	4	5	6	مجموع
التكرار (n_i)	10	12	16	18	22	14	8	100

I –3–الجدول التكراري لمتغير كمي متصل (مستمر):

المتغير الكمي المتصل هو ذلك المتغير الذي يمكن تجزئة قيمه ووضعها في مجالات (فئات). حيث أن قيمه توضع على شكل مجالات متساوية أو غير متساوية وتسمى فئات، ولكل فئة حد أدنى وحد أعلى.

مثال (1–3): لدراسة ظاهرة تأخر وصول الطلبة إلى مقاعد الدراسة أخذنا عينة من 150 طالب، فكانت النتائج كما يلي:

الجدول رقم (1–3): ظاهرة تأخر وصول الطلبة إلى مقاعد الدراسة

الفئات	$[0-5[$	$[5-10[$	$[10-15[$	$[15-20[$	$[20-25[$	$[25-30[$	مجموع
التكرار (n_i)	10	20	35	55	20	10	150

ويحدد عدد هذه الفئات حسب حجم العينة وحسب توزيع الوحدات الإحصائية في مجال الدراسة.

1–3–1– تحديد الفئات: لتحديد هذه الفئات أو هذه المجالات وضع الإحصائي ستورجس (staurges) قاعدة تجريبية لتحديد طول الفئات وتعتمد هذه القاعدة على مجال الدراسة (المدى العام) وحجم المجتمع أو العينة. وتعطى بالعبارتين التاليتين وذلك حسب اللوغاريتم المستخدم النيبيري أو العشري.

$$K = \frac{E}{1 + 1.32Ln(n)}$$

أو

$$K = \frac{E}{1 + 3.32Log(n)}$$

حيث :

n: حجم العينة أو المجتمع؛ E: المدى العام وهو الفرق بين أكبر قيمة و أصغر قيمة $E = X_n - X_1$

K: طول الفئة ، والعبارة $1 + 1.32Ln(n)$ تمثل عدد الفئات.

ملاحظة: لا يوجد طول متفق عليه لهذه الفئة كما انه لا يوجد عدد متفق عليه من الفئات لتوزيع إحصائي معين ,ولكن يتحدد عدد الفئات وطولها حسب أهمية الدراسة وطول مجال التعريف وحجم العينة أو المجتمع. وللحفاظ على توازن جدول التكرارات من الأفضل أن يتراوح عدد الفئات من 9 إلى 25 فئة.كما أنه يستحسن أن تكون أطوال الفئات متساوية وذلك حتى يمكن مقارنة الظاهرة من مجال إلى مجال آخر. كما يتم تحديد طول الفئة بالعلاقة التالية:

$$\mathrm{K} = \frac{\mathrm{E}}{\mathrm{L}}$$

حيث L يمثل عدد الفئات، ويجب مراعاة المتباينة التالية في تحديد طول الفئة: $E \leq L \times K$

ويمكن تطبيق معادلة يول (YULE) لحساب عدد الفئات بالعلاقة التالية: $\mathrm{L} = 2.5\sqrt[4]{n}$

1-3-2- تحديد مراكز الفئات: في حالة تكوين جدول توزيع تكراري لمتغير كمي متصل تضيع القيم الإحصائية الأصلية للمشاهدات ونصبح لا نعرف عنها شيء سوى أنها تنتمي إلى فئة معينة محدودة بحدين معلومين، ولتخطي هذه المشكلة نقوم باستخراج مركز الفئة والذي نحصل عليه بحساب الصيغة التالية:

$$x_i = \frac{x_{min} + x_{max}}{2} = \frac{\text{الحد الأعلى للفئة} + \text{الحدى الأدنى للفئة}}{2}$$

1-3-3- تحديد حدود الفئات: تبدأ حدود الفئة الأولى بوضع أقل قيمة في البيانات في الحد الأدنى ثم يضاف إليها طول الفئة لتحديد الحد الأعلى، ونستمر بهذه الطريقة إلى غاية الوصول إلى الفئة الأخيرة يكون فيها الحد الأعلى لهذه الفئة عبارة عن القيمة العظمى في البيانات.

II-جداول التكرارات التجميعية: نميز هنا بين ثلاثة أنواع من التكرارات

II -1-التكرار المطلق: وهو التكرار العادي، أي عدد المشاهدات الخاصة بكل قيمة التي يمكن أن يأخذها المتغير الإحصائي.

II –2–التكرار النسبي: وهو عبارة عن نسب مئوية نتحصل عليها بقسمة التكرار العادي على مجموع التكرارات $f_i = {}^{n_i}/_{\sum n_i}$. مع العلم أن $\sum f_i = 1$. ويمكن تحويل التكرار النسبي إلى تكرار نسبي مئوي وذلك بضرب التكرار النسبي في 100، ويرمز له بـ $f_i\%$.

يفيد التكرار النسبي في تقليص الشكل البياني في حال كان عدد القيم كبيرا، بينما التكرار النسبي المئوي يفيد في إظهار الشكل عندما يكون عدد القيم صغيرا.

II –3–التكرارات التجميعية: يستخدم هذا النوع من الجداول عندما نحتاج إلى معلومات إضافية عن البيانات، فمثلا قد نحتاج إلى معرفة عدد المفردات أو المشاهدات التي تقل قيمتها أو تزيد عن حد معين. فنقوم لذلك بإيجاد التكرارات التجميعية الصاعدة والنازلة. ومنه تنقسم هذه التكرارات إلى قسمين:

II –3–1– التكرارات التجميعية الصاعدة: وهي عبارة عن تكرار فئة مضافا إليه مجموع تكرارات الفئات السابقة. وهو يمثل مجموع الأفراد الذين تقل قيمتهم الإحصائية عن الحد الأعلى للفئة المقابلة.

II –3–2– التكرارات التجميعية النازلة: عبارة عن مجموع التكرارات ($\sum n_i$) مطروح منه تكرارات الفئات السابقة، ويمثل مجموع الأفراد الذين تزيد قيمتهم الإحصائية عن الحد الأدنى للفئة المقابلة.

تمارين محلولة:

التمرين الأول:

إذا كان لدينا بيانات الحالة المدنية لـ20 شخص في بلدية ما كما يلي:

M	M	C	D	D	C	M	C	D	M
C	D	C	C	C	M	C	M	V	V

حيث:

(C) أعزب، (M) متزوج، (D) مطلق، (V) أرمل.

المطلوب:

1–تحديد الخاصية المدروسة، المجتمع المدروس، الوحدة الإحصائية، ونوع المتغير المدروس.

2–أعد ترتيب هذه البيانات في جدول توزيع تكراري.

الحل:

الخاصية المدروسة: الحالة المدنية للأشخاص. المجتمع المدروس: 20 شخص قاطن في بلدية معينة. الوحدة الإحصائية: شخص. المتغير المدروس: متغير كيفي (نوعي) غير قابل للترتيب.

X_i	n_i	f_i	$f_i\%$
C	8	0,40	40
M	6	0,30	30
D	4	0,20	20
V	2	0,10	10
$\sum$	20	1	100

التمرين الثاني:

ليكن لدينا بيانات حول الحالة الاجتماعية المهنية في مؤسسة معينة و المدونة في الجدول التالي:

Cs	C	C	Ab	Cs	Ab	C	Ab	C	Cs
Ab	Ab	Ab	Ap	Ab	Ap	Ab	Ab	Ap	Ab
C	Ap	Cs	Ab	C	S	Cs	C	Ab	C
Ab	Ap	S	C	Ab	Ap	Ab	S	Ab	Ap
Ab	C	Ap	Ab	Cs	S	C	Ab	C	Ab

حيث:

(Cs) إطار سامي، (C) إطار، (Ab) عون مكتب (Ap) عون متعدد الخدمات، (S) أمين مكتب.

المطلوب:

1-تحديد الخاصية المدروسة، المجتمع المدروس، الوحدة الإحصائية، ونوع المتغير المدروس.

2-أعد ترتيب هذه البيانات في جدول توزيع تكراري.

الحل:

الخاصية المدروسة: الحالة الاجتماعية المهنية لعمال مؤسسة. المجتمع المدروس: 50 عامل.

الوحدة الإحصائية: عامل.

المتغير المدروس: متغير كيفي (نوعي) قابل للترتيب.

X_i	n_i	f_i	$f_i\%$	$F^{\uparrow}$
Cs	6	0,12	12	6
C	12	0,24	24	18
Ab	20	0,40	40	38
Ap	8	0,16	16	46
S	4	0,08	8	50
$\sum$	50	1	100	/