

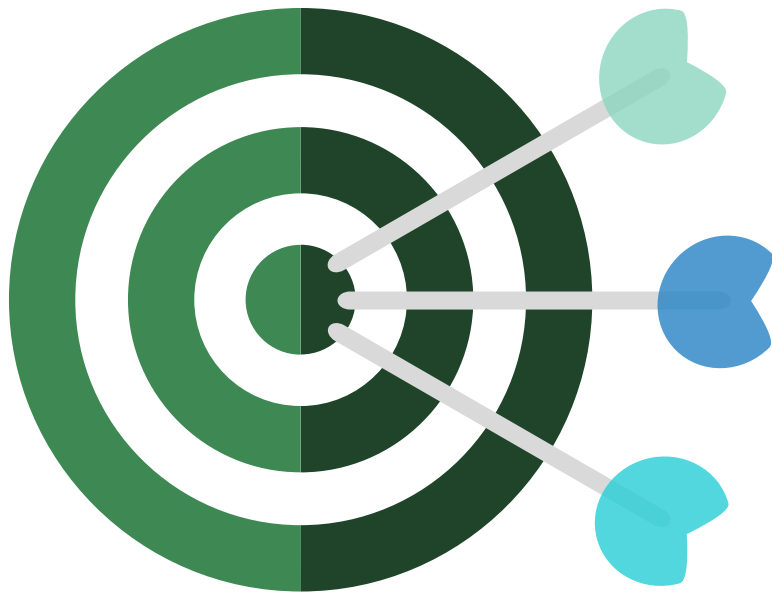
١١.٣ احص مقدمة في الإحصاء المحاضرة الأولى

الفصل الأول: البيانات الإحصائية جمعها وتنظيمها

فصل البيانات الإحصائية جمعها وتنظيمها

✓ تعاريف ومفاهيم أساسية.

✓ تنظيم البيانات الخام وتمثيلها الشرائطي وبالقطاعات الدائرية.



تعريف علم الإحصاء (statistics)

إنَّ علم الإحصاء هو ذلك الفرع من الرياضيات الذي يهتمّ بجمع البيانات، وتنظيمها (في جداول وعروض بيانية مناسبة)، ودراسة خصائصها وتحليلها، واستقرائها، وأخيراً اتخاذ القرارات المناسبة بشأنها.

الإحصاء الوصفي

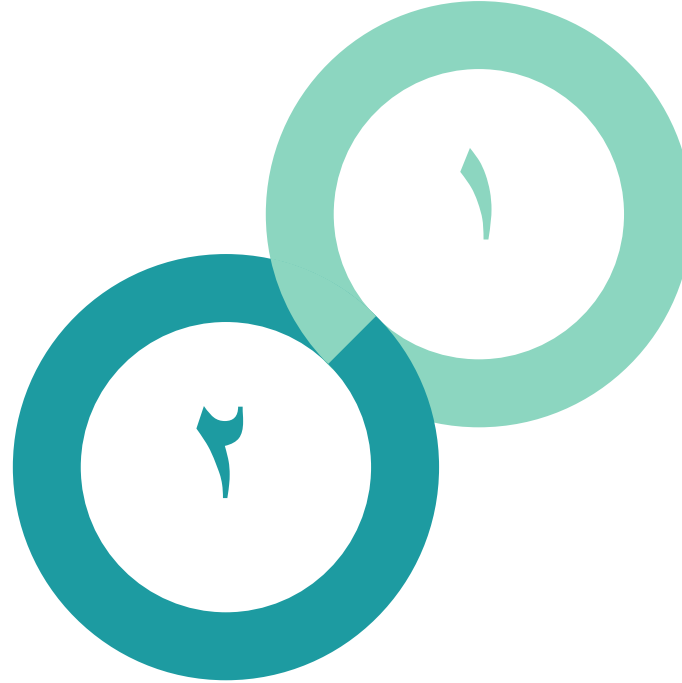
Descriptive statistics

يهتم بجمع البيانات وتنظيمها
ودراسة خصائصها العددية

الإحصاء الاستدلالي

Inferential statistics

يهتم بتحليل البيانات واستقرائها
واتخاذ القرارات المناسبة بشأنها، وفي
مراجع أخرى يذكر باسم الإستدلال
الإحصائي Inference Statistical



➤ الإحصاء الوصفي Descriptive Statistics :

هو ذلك الفرع من الإحصاء الذي يهتم بجمع البيانات وتنظيمها في جداول وعروض بيانية مناسبة، ودراسة خصائصها العددية.

➤ البيانات Data :

هي قياسات Measures أو ملحوظات أو مشاهدات Observations تهدف إلى غرض معين في مجتمع إحصائي مُحدّد، ويتم تدوينها كنتيجة لعملية إنتاجية (مثل كميات القمح الناتج عن الزراعة في عام أو أعوام متتالية في بلد ما) أو لتجربة معملية (مثل معرفة الألوان الناتجة عن تحليل ضوء الشمس) أو لمراقبات (ملاحظات) لكائنات موجودة في الطبيعة (مثل أعداد النجوم في مجرة في الفضاء الكوني) أو

تعريف

➤ الإحصاء الاستدلالي Inferential Statistics:

هو ذلك الفرع من الإحصاء الذي يهتم بتحليل البيانات واستقرائها (تعميم نتائج العينات على المجتمع الإحصائي) ومن ثمّ اتخاذ القرارات المناسبة بشأنها.

➤ المجتمع الإحصائي Population:

هو أي تجمّع لأشياء تجمّع بينها صفة مشتركة واحدة على الأقل لتكون محل دراسة لهدف محدد. تُدعى مكونات المجتمع عناصر أو أفراداً.

تعريف

المجتمع

هو عدد عناصر المجتمع

حجم المجتمع

١- محدود

٢- غير محدود

حجم المجتمع
ممکن أن يكون

المجتمع

المجتمع المحدود

عدد عناصر المجتمع منتهياً ويُرمز له بعدد طبيعي، وقد درجت العادة على استخدام أحرف لاتينية كبيرة من قبيل M, N للدلالة على حجم المجتمع، فعلى سبيل المثال من أجل دراسة إحصائية ما على طلاب جامعة الملك سعود يمكن النظر إلى طلاب هذه الجامعة على أنه مجتمع محدود.

المجتمع

المجتمع غير المحدود

عدد عناصر المجتمع غير منته، ولذلك لا يستخدم رمزاً للدلالة على حجم المجتمع في هذه الحالة، فعلى سبيل المثال من أجل دراسة إحصائية ما على سلوك الأعداد الأولية (مثل عشوائيتها وتوزيعها) فإنه يمكن النظر إلى مجموعة الأعداد الأولية على أنها مجتمع غير محدود .

من الأمور المهمة هنا هي أن ندرك أن المجتمع ليس بالضرورة أن يكون مجتمعاً بشرياً أو حتى مجتمعاً لأحياء، إذ إنه من الممكن أن يكون جماداً أو أي شيء آخر أيضاً، والمثالان الآتيان يوضحان لنا ذلك .

- لدى تحديد نسبة خام النحاس في فلز معدني في منجم معين، حيث يمكن أن تتواجد أنواع عديدة من مركبات النحاس في هذا المنجم ، ولكنها جميعاً تحوي على معدن النحاس، ولذلك فلز النحاس في هذا المنجم يكون مجتمعاً.
- في دراسة لتحديد الحالة الفنية للطائرات السفرية في المملكة العربية السعودية، حيث يوجد أنواع عديدة من الطائرات السفرية، ولكنها جميعاً تتصف بأنها طائرات سفرية ، ولذلك الطائرات السفرية الموجودة في المملكة العربية السعودية تكون مجتمعاً.

العينات Samples

بالطبع هناك عوامل أخرى قد تضطرننا إلى عدم التعامل مع عناصر المجتمع كله لأسباب أخرى منها الاقتصادية والزمنية أيضاً، ولذلك يلجأ المرء في مثل هذه الحالات إلى أخذ جزء من المجتمع لدراسته، وهذا العمل يندرج تحت مفهوم العينة والذي يقدمه لنا التعريف الآتي :

هي جزء من المجتمع يتم اختياره بشكل مناسب بحيث يمثل المجتمع تمثيلاً جيداً.

العينة
Sample

العينات Samples

قد تكون عملية إخضاع جميع عناصر المجتمع للبحث والدراسة شاقة، وأكثر من ذلك قد تكون في كثير من الحالات غير ممكنة أيضاً، فعلى سبيل المثال :

في عملية تحليل الدم لمريض فمن غير المعقول ولا المقبول أخذ كل دم المريض (ومن ثم قتل المريض) لتحليله من أجل الكشف على أسباب مرضه .

ملاحظات



- سنستخدم كلمة **عينة** عوضاً عن كلمة **عينة إحصائية** على سبيل الاختصار والتبسيط .
- نشير إلى أن **عدد عناصر العينة** يجب أن يكون **منتهياً**.
- يطلق على عدد عناصر العينة اسم "**حجم العينة**"، وقد درجت العادة على استخدام أحرف لاتينية صغيرة من قبيل m, n و... للدلالة على **حجم العينة**.

المتغيرات

نوعية

كمية

منتظمة

مستمرة

تعريف المتغير

هو تطبيق (وقد يكون دالة) مجاله أو (مجموعة قيمه) العينة أو المجتمع نفسه (حسب طبيعة الدراسة الإحصائية) وأما مجاله المقابل فهو مجموعة ذات طبيعة ما، فيمكن لها أن تكون أعداداً أو رموزاً أو مسميات، ويستخدم لقياس خاصية معينة لعناصر العينة أو المجتمع .

ملاحظة :

من التعريف السابق يتضح لنا أن القياسات أو المشاهدات التي يمكن أن تنتج عن متغير قد تكون قيماً عددية أو أحرفاً أو رموزاً أو وبناءً على ذلك يمكننا تصنيف المتغيرات في نوعين رئيسيين هما:

● كمية ● نوعية

١- المتغيرات الكمية

هي متغيرات تكون قيمها أعداداً حقيقية.

تعريفها

متغيرات مستمرة
(متصلة)

متغيرات متقطعة

أنواعها

١- المتغيرات الكمية

متغيرات متقطعة

هي تلك المتغيرات الكمية التي مجالها (مجموعة قيمها) منتهٍ أو غير منتهٍ ولكن قابلة للعد،

ومن الأمثلة على ذلك:

- المتغير الذي يرصد عدد السيارات المباعة من معرض ما في يوم معين حيث يكون لمجاله المجموعة $(0,1,2,3,\dots,N)$ مع N عدد طبيعي مثبت.
- المتغير الذي يرصد عدد التجارب التي يجب تنفيذها حتى الحصول على شعار لأول مرة لدى قذف قطعة نقود معدنية، فنجد أن مجاله المجموعة $(1,2,3,\dots)$ وهي مجموعة الأعداد الطبيعية كاملة.

هي تلك المتغيرات الكمية التي مجالها (مجموعة قيمها) غير قابل للعدّ (وبالتالي غير منتهية أيضاً) ومن الأمثلة على ذلك :

المتغير الذي يرصد عمر الإنسان في القرن الأخير 1916-2016 ، فنجد أنّ مجاله (لمجموعة قيمة) الفترة [0,179]، وهي مجموعة غير قابلة للعدّ.

المتغير الذي يرصد الوقت المستغرق من قبل طالب لإنهاء اختبارهِ (بزمن ١٢٠ دقيقة) في مقرر معين ، فنجد أنّ مجاله الفترة [0,120] ، وهي مجموعة غير قابلة للعدّ.

المتغير الذي يرصد طول الطفل عند الولادة في مستشفى للتوليد ، فنجد أنّ مجاله الفترة [20 , 45] على وجه التقريب (أطراف الفترة بالسنتيمتر)، وهي مجموعة غير قابلة للعدّ.

١- المتغيرات الكمية

متغيرات مستمرة
(متصلة)

٢- المتغيرات النوعية

تعريفها

هي متغيرات تكون قيمها عبارة عن رموز أو أسماء أو أرقام دالة على نوع أو اسم أو صفة أو تمييز ، وهذه القيم تنتج عن السؤال ب"ما".

إنّ البيانات التي تنتج عن هذا النوع من المتغيرات تُدعى بيانات نوعية ؛ فعلى سبيل المثال :

- أ- المتغير الذي يرصد ألوان الزهور في حديقة معينة هو متغير نوعي ، والقيم التي تنتج عنه (أحمر ، أصفر ، أبيض و....) ونحصل عليها بالسؤال : ما لون الزهرة ؟
- ب- المتغير الذي يرصد فصيلة الدم لدى البشر تكون مجموعة قيمه رموزاً AB, B, A, O ونحصل على هذه القيم بالسؤال: ما فصيلة دم الشخص X ؟
- ج- المتغير الذي يرصد الرقم الجامعي لطالب في جامعة الملك سعود هو متغير نوعي ، والقيم التي تنتج عنه هي أرقام من قبيل $436....., 437.....$ و..... وهذه الأرقام تميز الطالب ولا تعني مقداراً كمياً له ، ونحصل على هذه القيم بالسؤال: ما رقم الطالب X ؟

البيانات

نوعية

كمية

مستمرة

منتظمة

تنظيم البيانات الخام وتمثيلها

■ إن البيانات التي نحصل عليها قد تكون على أشكال مختلفة، فمنها على شكل قيم عددية مفردة، وبعضها الآخر قد يعرض تغير ظاهرة ما مع مرور الزمن أو مع مسميات كالبلدان أو المدن أو مع كليهما معاً، وبعضها الآخر قد يكون مجمعاً في جداول.

اتجاهات تنظيم البيانات

١
يهتم بتنظيم البيانات المفردة التي تنتج مباشرة عن الدراسة الإحصائية (كمية أو نوعية) في جداول تدعى الجداول التكرارية ومن ثم تمثيل هذه البيانات في عروض بيانية مناسبة.

٢
يهتم بتجميع البيانات المفردة الكمية فقط في جداول من نوع خاص تدعى جداول التوزيع التكرارية، حيث يقال عن البيانات المقدمة بهذه الجداول إنها بيانات مجمعة (أو مبوبة، أو مجدولة)، ومن ثم تمثيل بيانات هذه الجداول في عروض بيانية مناسبة.

البيانات الخام

لدى تنفيذ دراسة إحصائية معيّنة حول ظاهرة ما وجمع البيانات حول هذه الظاهرة تنتج لدينا بيانات مفردة تُدعى **بيانات خام**.

إذا كان عدد البيانات صغير

فإنّه يمكن التعامل مع هذه البيانات بشكل مباشر (مع كل قيمة على انفراد) لدراستها.

إذا كان عدد البيانات كبير نسبياً

فإنه قد يكون من الصعب التعامل معها بشكل مباشر، ولذلك لا بد من تقديم طرائق تسهل التعامل مع هذه البيانات كي يتم الاستفادة منها بأقصى قدر ممكن .

البيانات الخام

١- لدى الاطلاع على تقديرات ٦٠ طالباً في مقرر الإحصاء وجدنا المعطيات الآتية:

C	A	D	B	D	A	F	D	C	A
D	D	D	C	C	B	C	F	F	C
A	B	C	D	D	A	D	A	A	B
D	C	F	D	C	B	C	C	B	C
B	D	A	B	B	C	B	A	D	C
C	C	F	C	B	D	C	D	B	F

فلاحظ أنَّ هذه البيانات هي بيانات خام نوعية

٢- لقد سئل ٣٠ طالباً من كلية X عن عدد الحوادث المرورية التي حصلت معه خلال الفصل الدراسي الأول لهذا العام فكانت الإجابات كما يلي:

0	0	1	3	1	0	1	2	2	3
2	0	1	2	1	1	1	1	2	1
1	0	0	0	3	2	2	1	0	0

فلاحظ أنَّ هذه البيانات هي بيانات خام كمية

البيانات الخام

٣- البيانات الآتية تمثل الطول لخمسین طالباً جامعياً (مقدّرة بالسنتيمتر):

140	155	168	171	158	168	159	149	172	145
155	154	166	169	168	158	149	172	168	166
156	166	149	157	156	159	167	166	169	171
170	159	168	168	167	157	154	166	169	168
158	157	172	155	154	166	168	167	171	168

وهذه البيانات هي بيانات خام كميّة، ولكنها تتبع مجموعة بيانات مستمرة (أو متصلة)، حيث نلاحظ أنّ كمية البيانات الخام

المقدّمة أعلاه لاتعدّ كبيرة نسبياً إلا أنّه يصعب أخذ انطباع سريع عن سلوك هذه القيم بشكل مباشر، فعلى سبيل المثال: هل

كل قيمة في هذه المجموعة تتكرّر بالقدر نفسه؟

التمثيل الجدولي للبيانات الخام

إن تمثيل البيانات الخام جدولياً يعني صبَّ هذه البيانات في جدول بتصميم معيّن، وهذا الجدول يُدعى **الجدول التكراري للبيانات**، فإذا أردنا صب مجموعة بيانات خام في جدول تكراري نقوم بإدراج جدول يحتوي على خمسة أعمدة ، وهذه الأعمدة تخصص على النحو الآتي :

أ- يدوّن في العمود الأول **ممثّل** لكل نوع من الأسماء أو الرموز أو الأعداد حسب طبيعة البيانات التي قيد الدراسة ، فعلى سبيل المثال لدينا:

- الرمز A هو ممثّل لكل الرموز A في مجموعة بيانات المثال (١) .
- الرمز O هو ممثّل لكل الأعداد O في مجموعة بيانات المثال (٢) .

ب - يدوّن في العمود الثاني **التعداد** لكل ممثّل (من الأسماء أو الرموز أو الأعداد) في مجمل البيانات التي قيد الدراسة ، ويتم ذلك برسم خط عمودي عن كل بيان موافق للاسم أو الرمز أو العدد ، وإذا أصبح لدينا أربعة خطوط عمودية فإن الخط الخامس يحزمها على النحو ++++ .

ج - يدون في العمود الثالث **عدد** يُعبّر عن تعداد كل ممثّل (من الأسماء أو الرموز أو الأعداد) في مجمل البيانات التي قيد الدراسة، وهذا العدد يُدعى التكرار.

د - يدون في العمود الرابع **عدد** يعبّر عن نسبة تكرار كل ممثّل (من الأسماء أو الرموز أو الأعداد) إلى العدد الكلي للبيانات التي قيد الدراسة، وهذا العدد يُدعى التكرار النسبي. أي أنّ التكرار النسبي يساوي تكرار النوع مقسوماً على المجموع الكلي للتكرارات.

هـ - يدون في العمود الخامس **عدد** يعبّر عن حاصل ضرب العدد ١٠٠ في التكرار النسبي لكل ممثّل (من الأسماء أو الرموز أو الأعداد) ويقراً كنسبة مئوية، وهذا العدد يدعى التكرار المئوي. أي أنّ التكرار المئوي يساوي التكرار النسبي مضروباً في ١٠٠

التمثيل الجدولي للبيانات الخام

١- لنقم بصبّ البيانات الموجودة في المثال (١) في جدول تكراري، فنجد أنّ لهذا الجدول العرض الآتي:

التقدير	التعداد	التكرار	التكرار النسبي	التكرار المئوي
A		9	$9/60 = 0.15$	$0.15 \times 100 = 15\%$
B		12	$12/60 = 0.20$	$0.20 \times 100 = 20\%$
C		18	$18/60 = 0.30$	$0.30 \times 100 = 30\%$
D		15	$15/60 = 0.25$	$0.25 \times 100 = 25\%$
F		6	$6/60 = 0.10$	$0.10 \times 100 = 10\%$
Total	-----	60	1	100

- الشكل ١ -

التمثيل الجدولي للبيانات الخام

٢- لتكن لدينا البيانات الآتية والناجمة عن فحص فصيلة الدم لـ ٦٠ شخصاً.

B	A	B	A	B	O	A	O	AB	A
A	O	A	AB	O	A	AB	O	A	AB
A	B	B	B	A	AB	O	A	AB	A
B	AB	A	A	AB	A	A	O	B	B
AB	A	B	O	A	B	A	AB	A	AB
A	B	A	A	AB	A	O	A	B	B

فلو قمنا بصب هذه البيانات في جدول تكراري على النحو السابق، فإننا سنجد له العرض الآتي :

رمز فصيلة الدم	التعداد	التكرار	التكرار النسبي	التكرار المئوي
A		24	$24/60 = 0.40$	$0.40 \times 100 = 40\%$
B		15	$15/60 = 0.25$	$0.25 \times 100 = 25\%$
AB		12	$12/60 = 0.20$	$0.20 \times 100 = 20\%$
O		9	$9/60 = 0.15$	$0.15 \times 100 = 15\%$
Total		60	1	100

ملاحظات



■ إن مجموع التكرارات النسبية يجب أن يساوي الواحد تماماً ، ولكن عند تنفيذ بعض الحساب نضطر إلى إجراء عملية تدوير للأرقام ، وفي هذه الحالة قد لا نحصل على مجموع يساوي الواحد تماماً ، فيكون المجموع أكبر أو أصغر من الواحد بقليل.

■ إن مجموع التكرارات المئوية يجب أن يساوي المئة تماماً ، ولكن إذا ما حصلت عملية تدوير للأرقام فإن مجموع التكرارات المئوية قد لا يساوي المئة تماماً ، فيكون لدينا مجموع أكبر أو أصغر من المئة بقليل.

ملاحظات



- بعد الانتهاء من صبّ البيانات يمكن الاستغناء عن عمود التعداد لأنّ عمود التكرار يؤدي الغاية نفسها، وأمّا في حال تقديم البيانات مجمّعة في جدول تكراري فإنّه (وفي معظم الحالات) لا يدرج عمود التعداد معه بسبب عدم وجود مبرر لظهوره، وبذلك يتبقى لدينا جدول بأربعة أعمدة فقط.

التمثيل الشرائطي للبيانات الخام

يعدّ تمثيل البيانات الخام باستخدام الشرائط العمودية (ويعرف باسم التمثيل بالأعمدة أيضاً) أو الشرائط الأفقية من أحد التمثيلات الجيدة للبيانات الخام ، وذلك لأنها تعطي انطباعاً سريعاً حول طبيعة البيانات الخام وسلوكها ، وسبب ذلك أنه من طبيعة الإنسان سرعة الملاحظة عند النظر إلى المشاهد والرسومات.

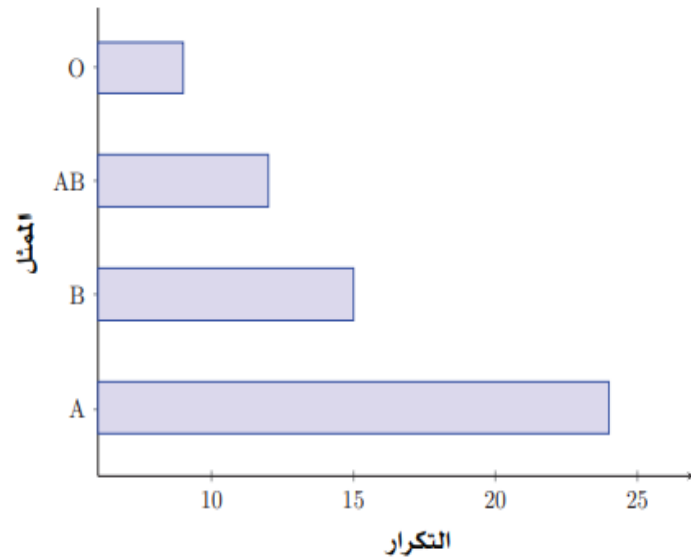
فكيف يتم التمثيل الشرائطي لمجموعة بيانات خام؟

بغض النظر عن نوع البيانات الخام اسمية، رموزاً أو عددية، نقوم برسم محورين متعامدين XoY ، ومن ثم يدون أسفل المحور Ox الممثل لكل صنف في البيانات الخام (فإن كانت أسماء كتبت الأسماء، وإن كانت رموزاً وضعت الرموز وإن كانت أعداداً سجلت الأعداد). وأما المحور Oy فيدون عليه قيم تكرارات البيانات الخام. بعد ذلك يُرسم عمود فوق كل مُمثل بيانات بارتفاع قدره يساوي قيمة تكرار هذا الممثل مع الأخذ بالحسبان أن تكون هذه الأعمدة منفصلة بعضها عن البعض الآخر بتباعد ثابت (وغالباً ما تؤخذ بمقدار وحدة قياس). في هذه الحالة نحصل على التمثيل بالشرائط العمودية أو التمثيل بالأعمدة.

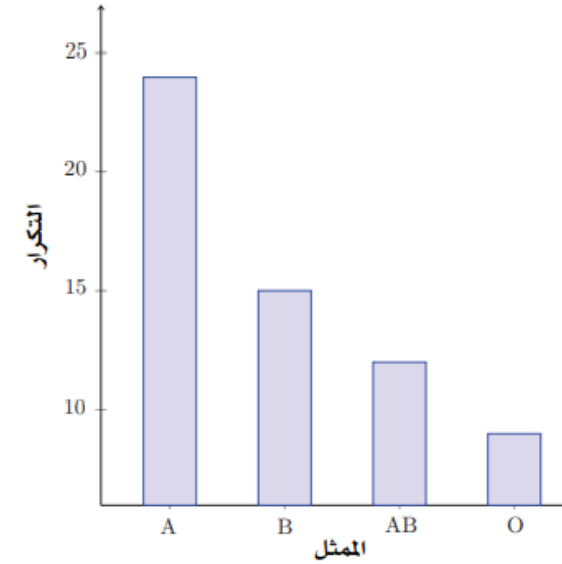
أما إذا أردنا تمثيل البيانات بالشرائط الأفقية فإننا نعكس العمليات التي تمت على المحورين المتعامدين، فيصبح المحور Oy من أجل تدوين الممثل لكل صنف في البيانات الخام في حين يستخدم المحور Ox لتدوين قيم تكرارات البيانات الخام، وفي هذه الحالة لا يُقال عن التمثيل الناتج إنه تمثيل بالأعمدة للبيانات الخام. بعد ذلك يُرسم شريط أفقي إلى جانب كل مُمثل للبيانات بطول قدره يساوي قيمة تكرار هذا الممثل، ومع الأخذ بالحسبان أن تكون هذه الأشرطة منفصلة بعضها عن البعض الآخر بتباعد ثابت أيضاً.

بالعودة إلى المثال (٢) ، فإننا نجد تمثيل البيانات الخام المعطاة باستخدام:

وأما التمثيل باستخدام الشرائط الأفقية له هذا الشكل :

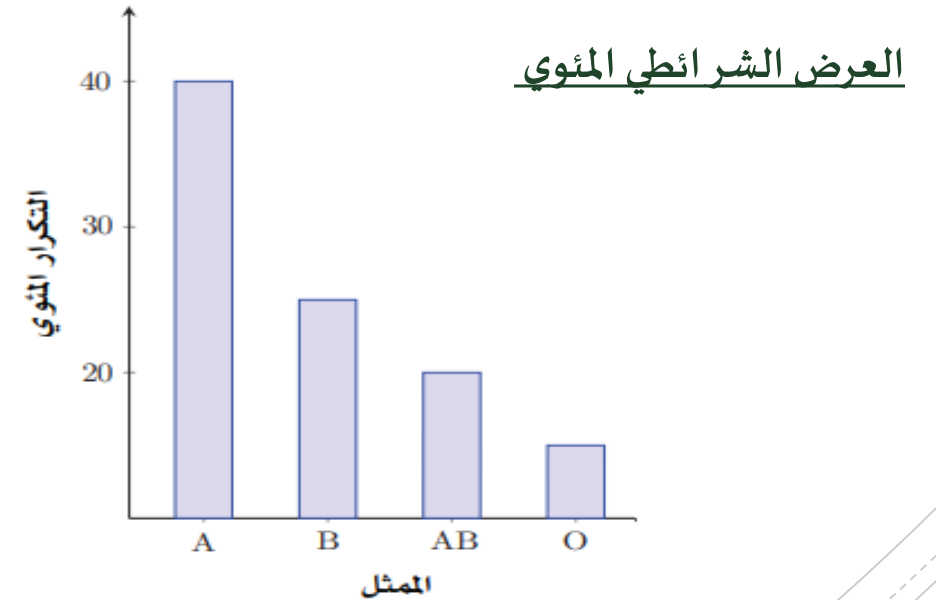
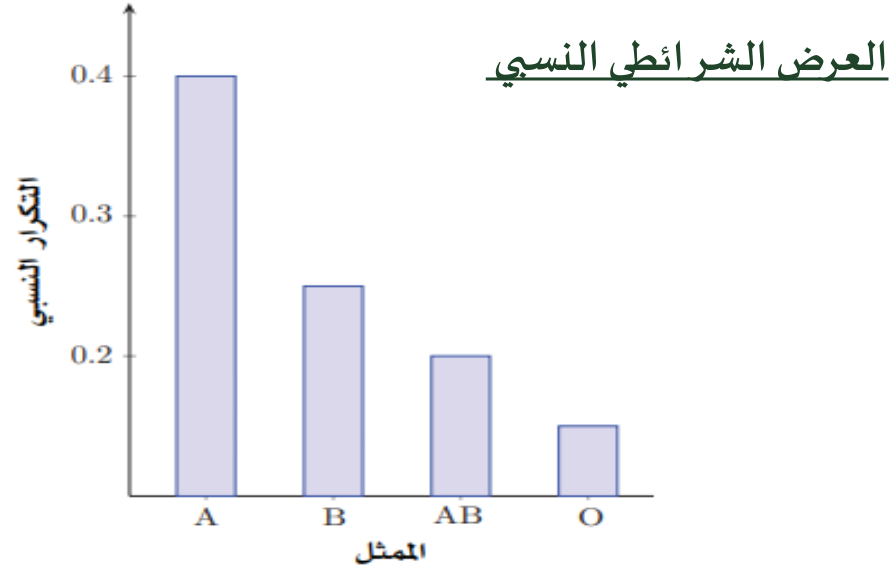


الشرائط العمودية (أو التمثيل بالأعمدة) له هذا الشكل :



ملاحظة:

يمكن استخدام التكرارات النسبية والتكرارات المئوية بدلاً من التكرارات في التمثيل الشرائطي أيضاً ، حيث تستبدل قيم التكرارات بقيم التكرارات النسبية أو التكرارات المئوية، فعلى سبيل المثال نجد من أجل المثال السابق أن العرض الشرائطي النسبي والمئوي لهما الشكلين الآتيين :



التمثيل بالقطاعات الدائرية (أو القرص الدائري)

يُلبأ عادة إلى استعمال هذه الطريقة عندما نكون بحاجة لتقسيم الكل إلى K من الأجزاء ، وأما لرسمها فإننا نقوم أولاً برسم دائرة بنصف قطر مُثبَّت (غالباً يكون عمودياً) يُعدّ مبدأً لقياس الزاوية عنه، ثم تُحسب زوايا القطاعات الدائرية α_i (مقدرة بالدرجات Degrees) وتأخذ إلى يمين العمود السابق باتجاه دوران عقارب الساعة ، وبحيث يكون للممثل (أو الجزء) قطاع دائري زاويته

تُحسب بوساطة العلاقة الآتية:

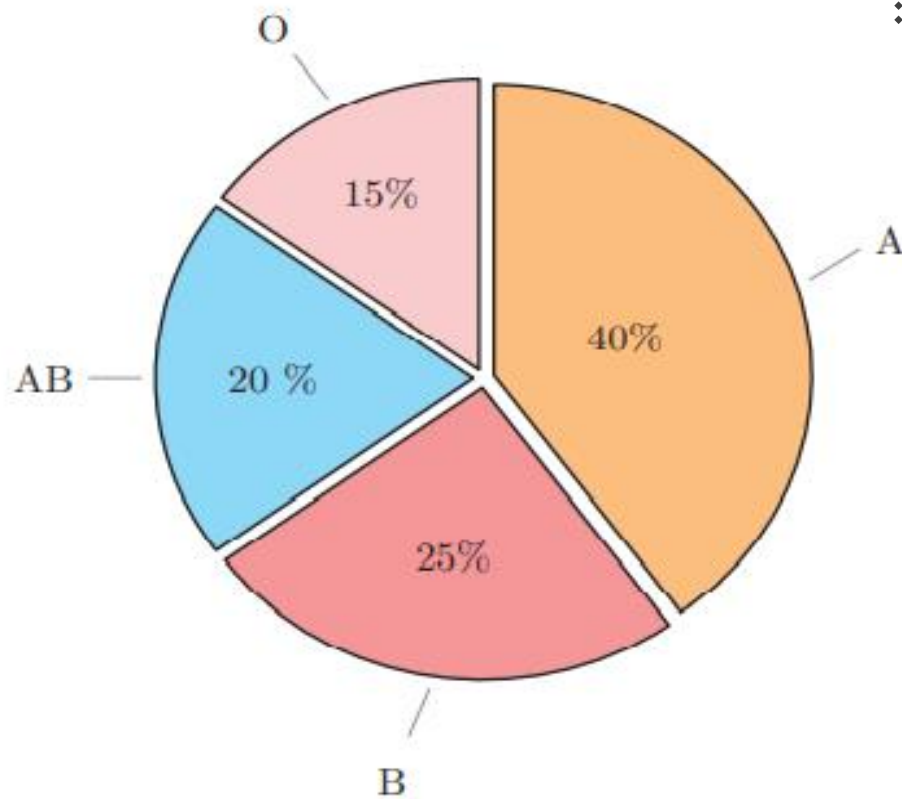
$$\alpha_i := \frac{n_i}{n} \times 360$$

[1-1]

علماً أنّ n هو عدد البيانات الخام المُعطاة و n_i هو عدد العناصر (أو البيانات) التابعة للممثل (أو الجزء) i ، بمعنى آخر ، فإننا نحصل على قيمة الزاوية للقطاع التابع للممثل (أو الجزء) i من خلال ضرب قيمة التكرار النسبي لهذا الممثل في 360 ، والمثالان الآتيان يوضحان لنا ذلك .

التمثيل بالقطاعات الدائرية (أو القرص الدائري)

١- بالرجوع إلى المثال (٢) من الشكل ١-١، وباستخدام العلاقة [١-١] نجد أن:



$$\alpha_A = \frac{24}{60} \times 360 = 144^\circ$$

زاوية القطاع الدائري لممثل البيانات A هي

$$\alpha_B = \frac{15}{60} \times 360 = 90^\circ$$

زاوية القطاع الدائري لممثل البيانات B هي

$$\alpha_{AB} = \frac{12}{60} \times 360 = 72^\circ$$

زاوية القطاع الدائري لممثل البيانات AB هي

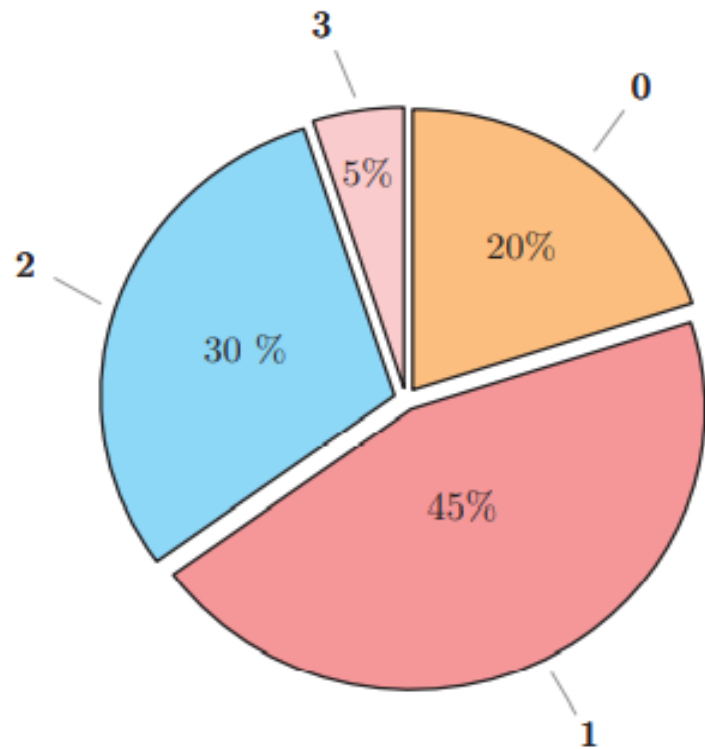
$$\alpha_O = \frac{9}{60} \times 360 = 54^\circ$$

زاوية القطاع الدائري لممثل البيانات O هي

ومن ثم يكون لدينا العرض الآتي للقطاعات الدائرية:

التمثيل بالقطاعات الدائرية (أو القرص الدائري)

١- بالرجوع إلى المثال (٣) من - الشكل ١- ، وباستخدام العلاقة [١-١] نجد أن :



$$\alpha_0 = \frac{8}{40} \times 360 = 72^\circ$$

زاوية القطاع الدائري لممثل البيانات 0 هي

$$\alpha_1 = \frac{18}{40} \times 360 = 162^\circ$$

زاوية القطاع الدائري لممثل البيانات 1 هي

$$\alpha_2 = \frac{12}{40} \times 360 = 108^\circ$$

زاوية القطاع الدائري لممثل البيانات 2 هي

$$\alpha_3 = \frac{2}{40} \times 360 = 18^\circ$$

زاوية القطاع الدائري لممثل البيانات 3 هي

ومن ثم يكون لدينا العرض الآتي للقطاعات الدائرية :