مدخل إلى علم الإحصاء

Introduction to Statistics

علم الإحصاء Statistics: هو أحد فروع الرياضيات الهامّة ذات التطبيقات الواسعة سواءً في العلوم الطبيعية أو الاجتماعية أو غيرها .. بحيث يهتمّ هذا العلم بشكلٍ أساسي بجمع وتنظيم وتلخيص وعرض وتحليل وتفسير ونشر البيانات الإحصائية.

البيانات Data: هي عبارة عن معلومات تكون مرتّبة ومنظّمة ومخزّنة بطريقة سهلة التداول والفهم والتفسير ومدعّمة بالأرقام.

من أين نحصل على البيانات ؟

- من مشاهدة ومراقبة الظواهر الموجودة في الطبيعة.

مثل: كميّة الأمطار الهاطلة في مكان معيّن، أو درجات الحرارة، أو سرعة الرياح..

- من خلال إجراء الاختبارات أو التجارب.

مثل: تجربة كميّات مختلفة من السماد أو كميّات مختلفة من مياه الري للوصول إلى الكميّة المُثلَى التي تعطي أفضل إنتاج من نوع نباتي معيّن.

- من خلال إجراء الدراسات والأبحاث والمقابلات.

مثل: إجراء دراسة ميدانية لجمع بيانات حول الوضع الاجتماعي والاقتصادي للمزارعين في منطقة معيّنة، أو إجراء مقابلات مع العاملين في شركة معيّنة لأخذ بيانات حول ظروف العمل في الشركة ومدى رضاهم عن العمل بما يمكن أن يساهم في تحسين ظروف العمل في الشركة وبالتالي إنتاجية العمال.

تصنيف البيانات: يمكن تصنيف البيانات بشكلٍ عام إلى نوعين رئيسيين:

- البيانات النوعيّة أو الوصفية أو الكيفية Qualitive Data: وتنقسم بدورها إلى نوعين:

– البيانات الإسمية Nominal Data: تكون في صورة غير عددية أي لا يمكن قياسها وإجراء العمليات الحسابية عليها، ولكن بالإمكان عدّ أفراد كل فئة. مثل الجنس (ذكور، إناث) (1، 2) أو لون بتلات الأزهار (أحمر1، وردي2، برتقالي3) أو سؤال إجابته نعم أم لا (1، 0).

- البيانات الترتيبية Ordinal Data: تكون في صورة غير عددية ولا يمكن إجراء العمليات الحسابية عليها والفرق بينها وبين البيانات الإسمية هي عملية المفاضلة والترتيب بين فئات أو رتب المتغير ولكن ليس شرطاً تساوي المسافات بين الرتب المتجاورة مثل المستوى التعليمي (ابتدائي 1، إعدادي 2، ثانوي 3، جامعي 4).

- البيانات الكميّة Quantitative Data:

- البيانات الفترية Interval Data: تكون في صورة عددية ويمكن إجراء العمليات الحسابية عليها مثل المتوسّط الحسابي والانحراف المعياري وغير ذلك.. وتمتاز هذه البيانات بتساوي المسافات بين الرتب المتجاورة كما أنّه في هذه البيانات لا يوجد صفر حقيقي وإنَّما صفر افتراضي لا يعني انعدام الخاصيّة فدرجة طالب تساوي صفر مثلاً لا يعني أنّه لا يعرف شيئاً في المقرر، ومن أمثلة هذه البيانات التواريخ ودرجات الحرارة.
- البيانات النسبية Rational Data: هي بيانات ذات مستوى أعلى من البيانات السابقة حيث تمتاز بالتصنيف (مثل البيانات الاسمية) والترتيب (مثل البيانات الترتيبية) والمسافات المتساوية (مثل البيانات الفترية) وخاصيّة النسبية التي تعني أنّ للصفر خاصيّة انعدام الظاهرة (مثل أنّ سرعة سيارة تساوي الصفر تعني أن السيارة متوقّفة) أي أنّ الصفر هنا صفر حقيقي.

يتمّ الحصول على البيانات الكميّة غالباً من خلال قراءات الأجهزة والأدوات العلمية المستخدمة في الأبحاث، كقياسات الأوزان والأطوال والحجوم والأعمار والزمن والسعة والمسافات ودرجات الحرارة والدخل وغيرها من أنواع البيانات الكمية .. كما يمكن أن يتم ذلك من خلال الاستفتاءات الميدانية والاستبانات والجداول الاحصائية الرسمية وغير الرسمية.

مصادر جمع البيانات الإحصائية:

- المصادر الأولية (المباشرة أو الميدانية): هي البيانات التي يجمعها الباحث بنفسه من عيّنات البحث، وهو أكثر دقّة من المصدر الآخر لكنه يستهلك الكثير من الجهد والوقت والمال.
- المصادر الثانوية (غير المباشرة أو التاريخية): هي البيانات التي يتمّ الحصول عليها بشكلٍ غير مباشر من جهات معيّنة كاستخدام الدراسات السابقة وهو أقلّ دقّة لكنه يوفّر الوقت والجهد والمال.

المجتمع Population: عدد لا نهائي من الأفراد والعناصر التي تتعايش مع بعضها البعض وتتميّز بخصائص ومواصفات تميّزها عن بقيّة المجتمعات حيث أنّه لكل مجتمع خصائص ومميزات تميّزه عن المجتمعات الأخرى.

العيّنة Sample: هي جزء من المجتمع يجب ألَّا يقلّ عدد أفرادها عن 2 – 10 % من عدد أفراد المجتمع، وهي تُؤخَذ بطريقة عشوائية.

أساليب جمع البيانات:

- أسلوب الحصر الشامل: يتمّ في هذا الأسلوب دراسة كل فرد خاضع للبحث من دون استثناءات ممَّا يجعله دقيقاً جداً وواقعي لكنه يحتاج للكثير من الوقت والجهد والمال.
- أسلوب المعاينة: يتمّ في هذا الأسلوب دراسة مجموعة صغيرة مختارة بأسس علمية ثمّ تعميم النتائج على المجتمع ككل ممَّا يجعله غير دقيق ولكنه يوفر الوقت والجهد والمال.

أدوات جمع البيانات:

إذا اختار الباحث طريقة جمع البيانات من المصدر المباشر، يمكنه اختيار أحد الأسلوبين التاليين:

- الأسلوب الميداني المباشر: ويتطلّب هذا الأسلوب الاتصال المباشر للباحث أو من ينوب عنه بعناصر المجتمع أو أفراد العينة، ومن الأدوات التي يستخدمها هذا الأسلوب: الاستبيان، المقابلة، الملاحظة...
- أسلوب المراسلة غير المباشر: يتطلّب هذا الأسلوب إرسال الإستمارة لعناصر المجتمع أو أفراد العينة من خلال البريد أو الهاتف أو الفاكس أو الإنترنت....

وعلى وجه العموم فإنّه يمكن الجمع بين أكثر من طريقة تبعاً لطبيعة الدراسة وأهدافها.

الجدول (1) للاطّلاع: مقارنة بين بعض أدوات جمع البيانات الإحصائية.

الخصائص	توزيع الاستبيانات بشكل مباشر	توزيع الاستبيانات باستخدام البريد	الاستبيانات عن طريق الهاتف	الاستبيانات الالكترونية	المقابلة
السرعة	بطيئة	بطيئة	بطيئة	سريعة	سريعة
التكلفة	مرتفعة	مرتفعة	مرتفعة	منخفضة	مرتفعة
الحجم	العينات الكبيرة	العينات الكبيرة	العينات الكبيرة	العينات الكبيرة	العينات الصغيرة
+ **نقطة إيجابية**	نسبة استرجاع الاستمارات عالية وفي نفس وقت	الحصول على معلومات دقيقة ، لأن الاجابة تكون في	لا تحتاج الى عملية التفريغ، حيث تتم عملية إضافة أي بيانات إلى قاعدة البيانات آلياً		الحصول على توضيحات مباشرة من المفحوص لأجوبته،

	التوزيع	وقت عدم انشغال المفحوص		مما يضمن الحصول على بيانات أدق وأصح
– **نقطة سلبية**	عملية التوزيع و التفريغ متعبة صعوبة توفر الوقت المناسب دائماً	عملية التوزيع والتفريغ متعبة الشخص قد لا يعبئ الاستبيان بنفسه (التزوير)	الشخص قد لا يعبئ الاستبيان بنفسه احتمال التزوير وارد جداً	صعوبة القيام بها في حال العينات الكبيرة

مراجعة وتدقيق البيانات:

عند إتمام عملية جمع البيانات وفق الوسيلة المناسبة لذلك يجب أن تتمّ مراجعة وتدقيق البيانات للتأكد من مطابقتها وتكاملها مع متطلّبات الدراسة فعلى سبيل المثال لو كانت الوسيلة المستخدمة في جمع البيانات هي الاستبيان عندئذٍ يتوجّب مراجعة وتدقيق الإستمارات الإحصائية التي تمّ جمعها واستبعاد ما هو غير متكامل أو واضح أو دقيق وعزل الاستمارات التي يعتقد الباحث أنّها غير مطابقة لما هو مطلوب. ومن الأمور التي يتمّ التركيز عليها في هذه العملية ما يلي:

- عدم وضوح الكتابة، في هذه الحالة يُراجَع المبحوث للاستيضاح وفي حالة التعذّر تُلغَى اجابته.
- مدى توخّي الدقّة من المبحوث في اجابته، ويُلاحَظ ذلك من خلال التعارض في الاجابة، والاجابات النمطية إلخ....
- ترك بعض الأسئلة دون اجابة.

طرائق عرض البيانات الإحصائية

Methods of graphical data view

بعد الانتهاء من جمع البيانات تُستخدَم طرق عديدة لعرضها منها البسيطة ومنها المعقّدة، هذه الطرق سنقوم بذكرها فيما يلي:

1. **الجداول البسيطة:** عند استخدام هذا النوع من الجداول يجب مراعاة ما يلي:

- رقم الجدول: حيث يُخصَّص لكل جدول رقم خاص به يختلف عن الرقم الذي يسبقه وعن الرقم الذي يليه .. مع مراعاة أن تكون أرقام هذه الجداول مرتّبة بشكل تسلسلي تصاعدي.
- عنوان الجدول: حيث يُخصَّص لكل جدول عنوان يوضِّح بشكلٍ مختصر طبيعة المعلومات المُدوَّنة في الجدول.
- وحدة القياس: يجب أن تُوضَع الوحدات التي تُقَاس بها المعلومات المدونة في الجدول سواءً أعلى أو أسفل الجدول.
- المصدر: يتمّ ذكر المصدر الذي أُخِذَت منه المعلومات ويتمّ وضعه أسفل الجدول عادةً.
- تقريب الأرقام: يجب أن تكون الأرقام الموضوعة في الجدول من طرازٍ واحد كأن تكون كلّها صحيحة إن أمكن أو مقرَّبة لرقمين بعد الفاصلة أو ثلاثة..إلخ، وذلك حسب الحاجة ونوع المعلومات.

مثال على الجداول البسيطة:

Table (2): Development of Arable & Cultivated Lands in Syria during (2007 -2014).

الجدول (2): تطوّر الأراضي القابلة للزراعة والمُستثمرَة في سورية خلال الفترة (2007 - 2014).

Area: 1000 Hec **المساحة: ألف هيكتار**

Years السنوات	Arable Lands أراضي قابلة للزراعة	Cultivated Lands أراضي مُستثمرَة
2007	6039	5682
2008	6024	5666
2009	6012	5664
2010	6045	5696
2011	6068	5716
2012	6079	5731
2013	6083	5733
2014	6081	5732

المصدر: المجموعة الإحصائية الزراعية السنوية.

2. **الجداول التكرارية (الفئوية أو المبوَّبة):** هي جداول خاصّة لها أهمية كبيرة في الإحصاء، وتعتبر أفضل من النوع السابق من الجداول حيث تعطينا معلومات أكثر وضوحاً وتعبيراً عن البيانات المدرجة فيها. وهي تتألَّف من عدد من الأسطر والأعمدة. تشمل الأعمدة ما يلي:
   - العمود الأول يمثّل الحدود العليا والدنيا للفئات.
   - العمود الثاني يمثّل مراكز الفئات.
   - العمود الثالث يمثّل تكرارات الفئات.
   - العمود الرابع يمثّل التكرارات المتجمّعة الصاعدة للفئات.
   - العمود الخامس يمثّل التكرارات المتجمّعة الهابطة للفئات.
   - العمود السادس يمثّل أيّة معلومات توضيحية إضافية يراها الباحث مناسبة.

خطوات إنشاء الجدول التكراري: لإنشاء جدول التوزيع التكراري يتوجّب حساب بعض المقادير الهامّة التي سيتمّ عرضها في الجدول، وهي كالتالي:

1. عدد الفئات (K): يتمّ بدايةً تقسيم البيانات الإحصائية إلى عدد محدود من الفئات، ويمكن استخدام أحد ثلاث طرق معروفة لهذا الغرض، وهي التالية:
   - إذا كان n<1000 يتمّ استخدام معادلة يول Joil، وهي المعادلة التالية:

$$\boldsymbol{K = 2.5 * \sqrt[4]{n}}$$

   - إذا كان n≥1000 يتمّ استخدام معادلة ستيورجس Sturges، وهي المعادلة التالية:

$$\boldsymbol{K = \{1 + 3.3 * log(n)\}}$$

   - أو يمكن استخدام جدول خاص مجهّز مسبقاً من قبل التخصّصين في مجال الإحصاء:

n	≤30	45	60	100	>100
K	4	6	7	8	10

n: تمثّل عدد عناصر العيّنة التي تمّ أخذها من المجتمع المدروس.

يجدر بالملاحظة أنّه إذا كان عدد الفئات المحسوب بإحدى الطرق السابقة عدد غير صحيح فإنّه يتوجّب تقريب العدد إلى أقرب عدد صحيح.

2. المدى العام (W): وهو عبارة عن الفرق بين أعلى قيمة وأصغر قيمة من قيم العينة .. أي:

$$\boldsymbol{W = Xmax - Xmin}$$

3. مدى الفئة أو طول الفئة (I): بعد حساب كل من عدد الفئات والمدى العام يمكن بعد ذلك حساب مدى الفئة الواحدة بكل سهولة، من خلال تطبيق العلاقة التالية:

$$I = W/K$$

4. مركز الفئة (C): هو عبارة عن القيمة الواقعة في وسط أو مركز الفئة، وهو حاصل جمع الحد الأعلى إلى الحد الأدنى مقسوماً على اثنين.
5. تكرارات الفئة (R): هو عدد العناصر الموجودة ضمن مجال الفئة، والتي تكون أكبر أو مساوية للحد الأدنى وأصغر تماماً من الحد الأعلى للفئة.
6. التكرار المتجمّع الصاعد: هو عدد العناصر التي تكون أصغر من الحد الأعلى للفئة.
7. التكرار المتجمّع الهابط: هو عدد العناصر التي تكون أكبر أو مساوية للحد الأدنى للفئة.

مثال على إنشاء جدول تكراري: إذا كان لدينا الجدول التالي الذي يعطينا معلومات عن عدد البذور في سنبلة القمح في 13 عيّنة (سنبلة) تمّ عرضها في الجدول التالي:

رقم السنبلة	1	2	3	4	5	6	7	8	9	10	11	12	13
عدد البذور	80	90	40	75	61	30	101	73	82	95	45	65	77

- المطلوب: إنشاء جدول توزيع تكراري للبيانات الواردة في الجدول.

الحل:

1. نحسب عدد الفئات باستخدام معادلة يول:

$$K = 2.5 * \sqrt[4]{n}$$

$$K = 2.5 * \sqrt[4]{13} = 5$$

2. نحسب المدى العام:

$$W = Xmax - Xmin = 101 - 30 = 71$$

3. نحسب مدى الفئة أو طول الفئة:

$$I = W/K = 71/5 = 15$$

4. نحدّد الحدود الدنيا والعليا للفئات، وذلك كما يلي:

- الحد الأدنى للفئة الأولى هو $Xmin$ أي القيمة الدنيا للفئة الأولى، أمّا الأعلى فهو $(Xmax + I)$.

- الحد الأدنى للفئة الثانية هو الحد الأعلى للفئة الأولى، أمّا الأعلى فهو الحد الأدنى لها مضافاً له ***I***.

وهكذا.....

5. تحديد مراكز الفئات: يتمّ تحديد مراكز الفئات كما يلي:

مركز الفئة الأولى = (الحد الأدنى + الحد الأعلى)/2

C1 = (30+45) / 2 = 37.5

وبنفس الطريقة لباقي الفئات..

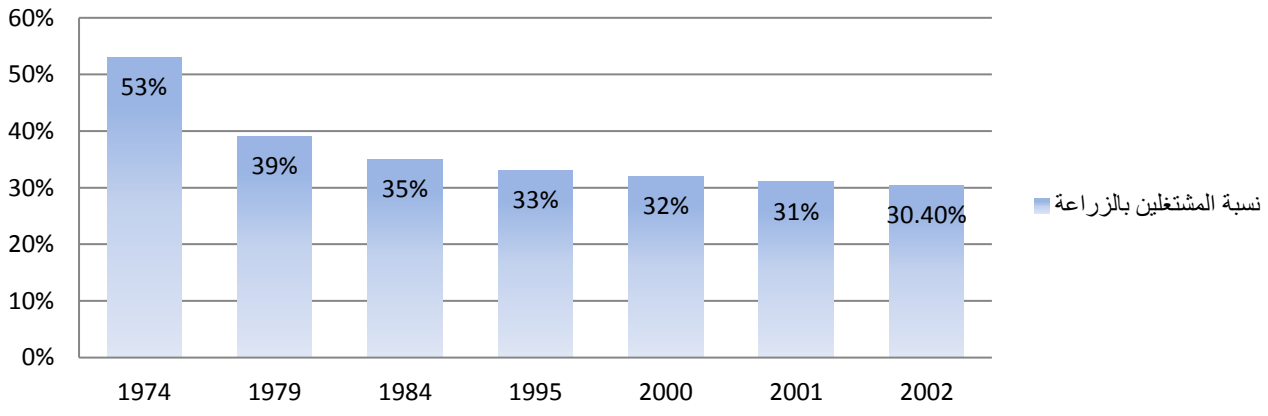
في الختام نحصل على جدول التوزيع التكراري للبيانات التي تمثّل عدد البذور في 13 سنبلة من نبات القمح:

الفئات	حدود الفئات	مراكز الفئات	تكرارات الفئات	التكرار المتجمع الصاعد	التكرار المتجمع الهابط
1	[30, 45[	37.5	2	2	13
2	[45, 60[	52.5	1	3	11
3	[60, 75[	67.5	3	6	10
4	[75, 90[	82.5	4	10	7
5	[90, 105]	97.5	3	13	3

3. الأعمدة (الأشرطة) البيانية: هي عبارة عن أعمدة أفقية أو عمودية ذات عرض واحد وطول يختلف تبعاً للمتغيّر الذي تمثّله، وتنقسم هذه الأشرطة البيانية إلى مجموعتين:

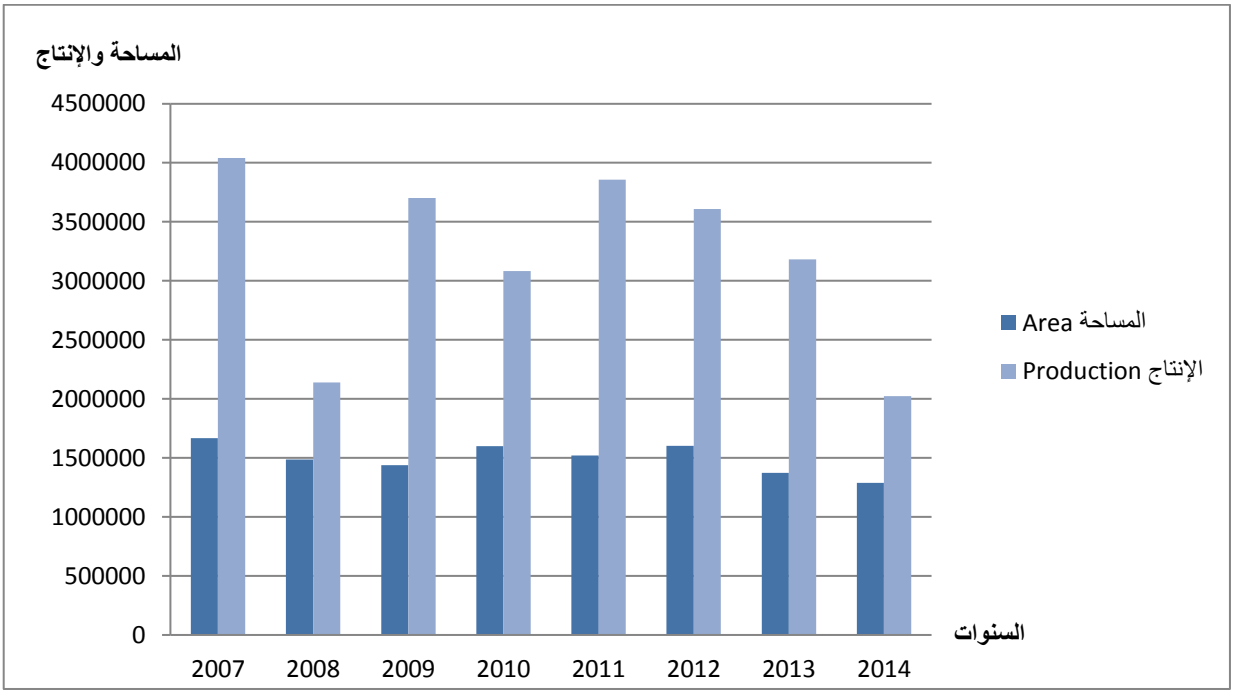
أ. الأشرطة البيانية البسيطة: هذا النوع من الأشكال البيانية يستخدم للتمثيل البياني عندما تحتوي البيانات متغير واحد مع الزمان أو المكان، كما في المثال التالي:

المخطط (1): تغيّر نسبة المشتغلين بالزراعة في سورية بين عامي 1974 - 2002 م.



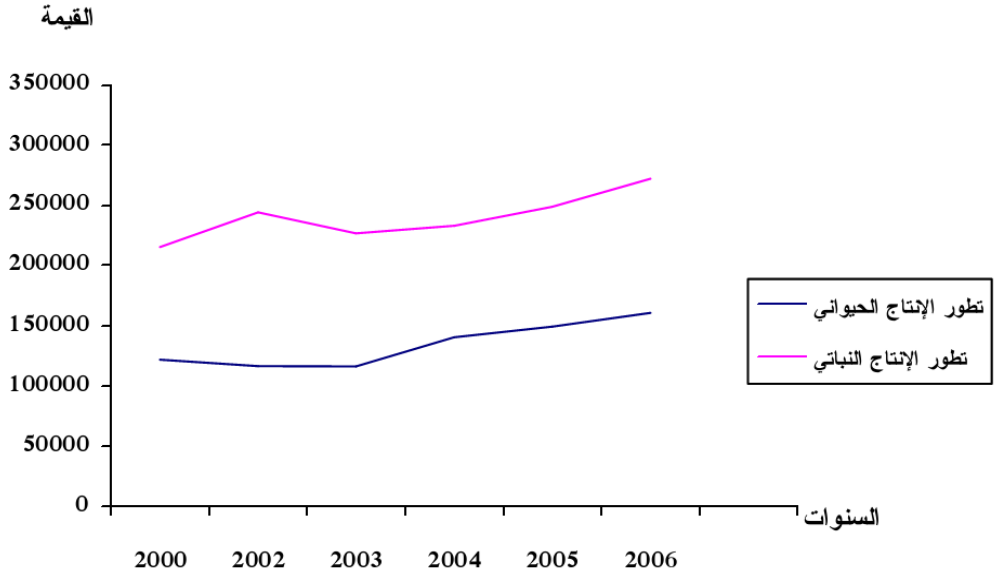
ب. الأشرطة البيانية المركّبة: هذا النوع من الأشكال البيانية يستخدم للتمثيل البياني عندما تحتوي البيانات أكثر من متغير مع الزمان أو المكان، كما في المثال التالي:

المخطط (2): تطوّر كل من إنتاج القمح والمساحات المزروعة به في سورية خلال الفترة (2007 – 2014) م.



4. المنحنيات البيانية (المضلع والمنحني التكراري): هو عبارة عن خط يمثّل العلاقة بين أحد المتغيرات مع متغير آخر مثل الزمن أو المكان. كما في المثال التالي:

المخطط (3): تطوّر قيمة الإنتاج النباتي والحيواني في سورية بأسعار عام 2000 الثابتة/مليون ليرة سوريّة.



5. **الرسوم البيانية القطاعية:** هي إحدى طرق العرض البياني شائعة الاستخدام، حيث يتمّ فيها تمثيل البيانات الإحصائية باستخدام دائرة تمثّل المجموع الكلي للبيانات، ومن ثمّ تُقسَم الدائرة إلى أجزاء أو شطائر تمثّل نسبة كل قيمة من القيم الإحصائية إلى المجموع الكلي للبيانات الإحصائية كنسبة مئوية. وعادةً ما يتمّ تطبيق هذه الطريقة وفق الخطوات التالية:

أ. حساب النسبة المئوية لكل قيمة أو مكون من مكونات العينة.

ب. حساب عدد الدرجات المقابلة كل قيمة أو كل مكون من مكونات العينة.

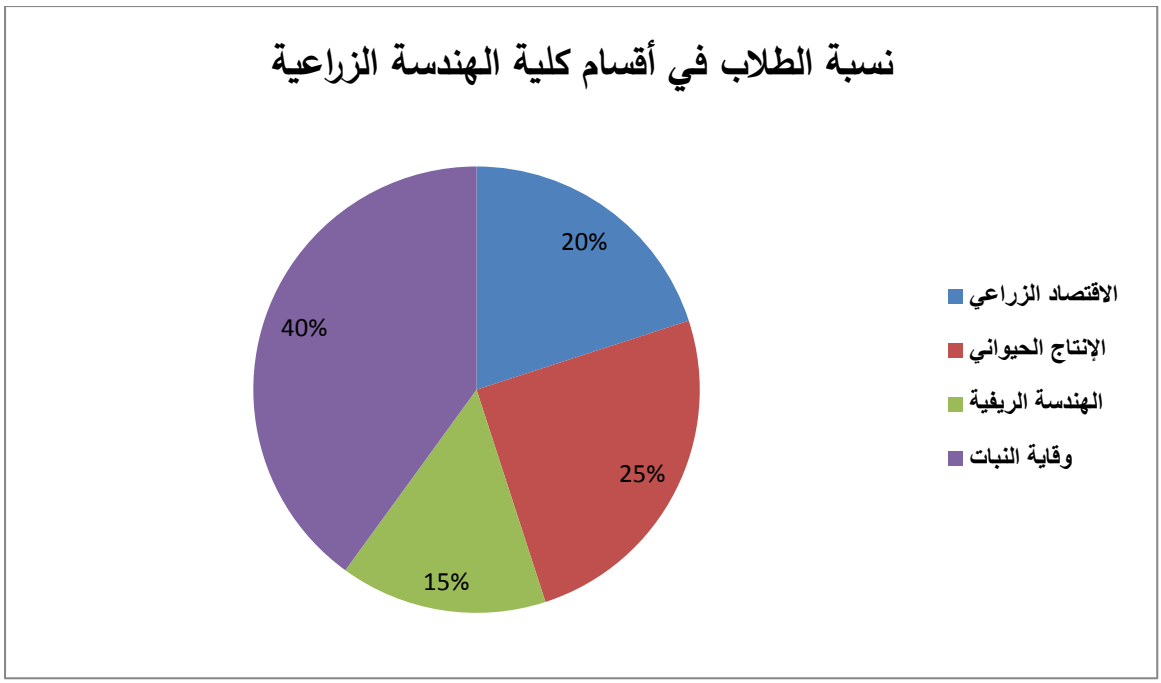
ج. رسم دائرة وتجزئتها إلى أقسام متناسبة أو متطابقة مع عدد الدرجات المقابلة لكل قيمة أو كل مكون من مكونات العينة.

كما في المثال التالي الذي يتضمّن بيانات ناتجة عن دراسة نسبة أعداد الطلاب في الأقسام المختلفة من كلية الهندسة الزراعية إلى العدد الكلي، وذلك في إحدى السنوات:

القسم	الاقتصاد الزراعي	الإنتاج الحيواني	الهندسة الريفية	وقاية النبات	المجموع
عدد الطلاب	2000	2500	1500	4000	10000
النسبة	%20	%25	%15	%40	%100

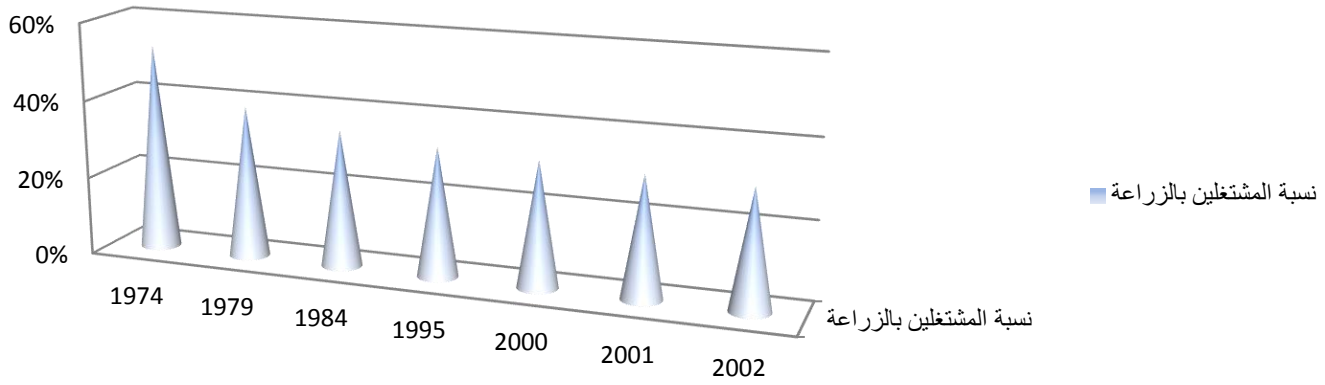
يتمّ تمثيل البيانات الواردة في الجدول بيانياً باستخدام الرسوم البيانية القطاعية وفق الخطوات السابقة لنحصل على المخطط البياني التالي:

المخطط (4): نسبة أعداد الطلاب في الأقسام المختلفة من كلية الهندسة الزراعية في إحدى السنوات إلى العدد الكلي.



6. **الرسوم البيانية المساحية:** هي عبارة عن أشكال هندسية مثل المربع أو المستطيل وغيرهما.. بحيث تستخدم مساحة هذه الأشكال لتمثيل قيم المتغير المدروس، ويتمّ حساب مساحة الشكل الهندسي من خلال حساب طول الضلع، وذلك حسب الشكل الهندسي وفقاً للدستور الرياضي المناسب.
7. **الرسوم البيانية الحجمية:** هي عبارة عن أشكال هندسية مثل المكعّب والاسطوانة والمخروط وغيرها.. بحيث يستخدم حجم هذه الأشكال لتمثيل قيم المتغير المدروس، ويتمّ حساب حجم الشكل الهندسي من خلال حساب طول الضلع، وذلك حسب الشكل الهندسي وفقاً للدستور الرياضي المناسب، كما في المخطط التالي:

المخطط (5): تغيّر نسبة المشتغلين بالزراعة في سورية بين عامي 1974 - 2002 م.



8. **الخرائط البيانية:** وهي عبارة عن خرائط جغرافية من حيث المبدأ تستخدم أشكال أو رموز مختلفة للتعبير عن قيم المتغير.

تمارين:

1. ارسم المنحني البياني للبيانات التالية التي تبيّن تطور إنتاج اللحوم في إحدى مزارع الإنتاج حيواني:

السنة	2013	2014	2015	2016
الإنتاج (طن)	22,2	45,6	34,0	87,5

2. لدينا البيانات التالية التي تبيّن تطور إنتاج أحد المحاصيل النباتية في منطقة سلمية:

السنة	2015	2016	2017	2018
الإنتاج (طن)	60000	70000	40000	20000

والمطلوب:

- مثّل البيانات السابقة في الجدول أعلاه بطريقة الأعمدة البيانية.
- أنشئ جدول التوزيع التكراري.
- ارسم المدرج التكراري.
- ارسم منحني ومضلّع التوزيع التكراري.
- مثّل البيانات أعلاه بطريقة التمثيل المساحي البيانية.
- مثّل البيانات أعلاه بطريقة التمثيل الحجمي البيانية.