الإحصاء والاحتمالات- المحاضرة الأولى

Statistics and probabilities- Lecture 1

Dr Fadi KHALIL

Doctor lecturer in statistics and programing

2022

مقدمة:

أصبح من المعروف في العالم اليوم أنّه لتعلم أو دراسة شيء ما لا بدّ من جمع بيانات تتعلق به. والإحصاء ببساطة هو فن التعلم من هذه البيانات. يهتم بالدرجة الأولى بجمع البيانات، وتوصيفها، ومن ثمّ تحليلها تمهيداً لاستخلاص النتائج منها. وفي الواقع ليس من البديهي تنفيذ هذه المهام خاصة في ظل عدم التأكد والتقلبات الحادة التي أصبحت تتسم بها البيانات هذه الأيام.

من هنا فإنّ علم الهندسة يتناول الإحصاء من ناحية أهميته في قياس دقة المخرجات ودراسة العوامل التي تؤثر في هذه الدقة (variability). فالمهندس معني أكثر من غيره بفهم مجموعة العوامل العشوائية التي تؤثر على خواص وأبعاد ودقة المنتج النهائي الذي يعمل عليه.

على سبيل المثال من مهام المهندس بشكل عام هو توصيف المشكلة، تحديد العوامل المؤثرة بها، اقتراح النموذج والتصميم، وتنفيذ التجربة تمهيداً لاعتماده وتطويره لاحقاً. كل هذه الخطوات تتطلب أحياناً الإعادة بالتالي جمع معلومات، تحليلها وتفسيرها، لفهم كيف يعمل النموذج أو التصميم لحل المشكلة قيد الدراسة.

مما سبق يمكن فهم دور الإحصاء كمساعد في اقتراح تصميم منتج جديد، تحسين تصميم حالي، أو تطوير عملية الإنتاج.

يهدف مقرر الإحصاء والاحتمالات إلى إطلاع طالب الهندسة على المفاهيم الإحصائية التي تساعد في توظيف الإحصاء في علم الهندسة. هذه المفاهيم تتعلق بأنواع البيانات وطرق تمثيلها بيانياً describing data، ومؤشرات الإحصاء الوصفي descriptive statistics، بالإضافة إلى أهم أدوات التحليل الإحصائي. يتضمن المقرر أيضاً عرض مبسّط لمفهوم المتغير العشوائي والتوزيع الإحتمالي probability distribution مع شرح لأهم التوزيعات الإحتمالية المنقطعة والمستمرة والعديد من المواضيع ذات الصلة. علاوةً على ذلك، يتناول المقرر مفهوم بالغ الأهمية في علوم الهندسة وخاصة التطبيقية منها وهو مراقبة الجودة quality control حيث يعرض كيفية الإستفادة من الأدوات والتقنيات الإحصائية في خدمة هذا الغرض وهو ما يعرف بـ statistical quality control.

ما يميّز هذا المقرر أنّه لا يكتفي بعرض للمفهوم الإحصائي فقط بل يقوم بعرض العديد من الأمثلة التي ترتبط بالقضايا التي تصادف المهندس في عمله وأبحاثه مما يساعد على استيعاب المفاهيم الإحصائية وتطبيقها في الواقع.

يعمل المقرر في نهايته على تطبيق المواضيع التي تمّ التطرق إليها على برنامج التحليل الإحصائي SPSS وهذا انطلاقاً من عضوية العلاقة بين علم الإحصاء من جهة والبرمجة والحاسوب من جهة أخرى.

يمكن تلخيص مخطط المقرر بما يأتي:

1. علم الإحصاء، أنواع البيانات وطرق جمعها
2. مفهوم الاستدلال الإحصائي statistical inference

3. طرق الوصف الإحصائي Describing data method
4. مقاييس النزعة المركزية والتشتت Central tendency and dispersion measures
5. مفهوم التوزيع الاحتمالي Probabilvity distribution
    a. التوزيعات الإحتمالية المنقطعة Discrete probability distribution
    b. التوزيعات الاحتمالية المستمرة Continuous probability distribution
6. مفهوم التوقع الرياضي، التباين، والتباين المشترك Expectation, Variance, covariance
7. قياس العلاقة والتوافق بين المتغيرات Correlation and correspondence
8. مجالات الثقة Interval of confodence
9. مفهوم مقايس لورينز ومعامل جيني The Lorenz curve and Gini index
10. اختبار الفرضيات Hypothesis Testing
11. مراقبة الجودة وخرائط مراقبة الجودة Quality control and control charts
12. تطبيق على الحزمة الإحصائية SPSS

في المحاضرة الأولى سيتم التطرق للمواضيع الآتية:

1. علم الإحصاء، أنواع البيانات وطرق جمعها
2. مفهوم الاستدلال الإحصائي statistical inference
3. أنواع المتغيرات، وطرق تمثيلها بيانياً

1- علم الإحصاء:

الإحصاء هو مجموعة من الطرق التي تساهم في تقديم ملخص عن ظاهرة مدروسة. وهو يعبر عن إجمالي البيانات الرقمية وغير الرقمية المتعلقة بهذه الظاهرة. يعبر علم الإحصاء أيضاً عن مجموعة الطرق العلمية التي تستخدم لـ

[1] لجمع ووصف البيانات.

[2] ترتيب وتصنيف البيانات.

[3] عرض وتمثيل البيانات

[4] وتحليل البيانات باستخدام الادوات الإحصائية.

من الضروري إذا في البداية التعرف على أنواع البيانات ووسائل جميع البيانات، قبل التعرف على الأساليب الإحصائية المستخدمة لتحليل هذه البيانات.

2- أنواع البيانات Data measures:

عموماً يوجد أربعة أنواع أو مستويات للقياس مرتبة تصاعدياً من البسيط إلى الأكثر وضوحاً وهى القياس : الإسمى ، والرتبي ، الفترى ، والنسبى ويمكن المقارنة بين الأنواع الأربعة على النحو الآتي:

المستوى	العمليات الرياضية	الخصائص القياسية	أمثلة
الاسمى Nominal	العد	-عدد لا يدل على كم أو مقدار (أعداد منفصلة) - الأرقام تحل محل الأسماء، - لا يمكن ترتيب الأرقام ترتيب تصاعدي او تنازلي. - الأرقام تمثل فئات – وضع الأشخاص فى فئات - لا تمثل الأرقام كميات من الخصائص التي يملكها الفرد. - تميز الأرقام بين المجموعات. - لا يمكن إجراء العمليات الحسابية على الأرقام	المهنة الجنس الحالة الاجتماعية نوع الحاسوب
الرتبى Ordinal	الترتيب العد	- كم لا يشار إليه بعدد (قيم منفصلة) - الأرقام مرتبة ترتيباً تنازلياً أو تصاعدياً. - يهتم بترتيب الأفراد فى الخاصية، ولكن لا يعبر الرقم عن كمية امتلاك الفرد من الخاصية، بل تدل على أن الرتبة تملك من الخاصية مقدار أكبر أو أصغر من رتبة أخرى. مثلاً إذا	المستوى التعليمي، درجة الإصابة بمرض ما، درجة أداء محرك السيارة

		كان لدينا المقياس الآتي 1- حزين، 2- غير حزين 3- لا بأس 4- سعيد، 5- سعيد جداً هنا نعرف أنّ الرمز 2 يدل على مقدار ما من السعادة أكبر من الرمز 1، ولكن لا نعرف إذا كان الفرق بين 3 و 2 هو نفسه الفرق بين 4 و 5 مثلاً. من هنا نرى أنّ المقياس الرتبي لا يعبر بالضرورة عن فروق متساوية حسابياً في الرتب لا يمكن تنفيذ العمليات الحسابية عليه مباشرة.	
الفترى Interval	الجمع الضرب الطرح	- عدد يدل على كم أو مقدار (قيم متصلة). - وضع الأشخاص في مقياس متصل يتكون من مسافات متساوية وله صفر اعتباري وليس مطلق. - الفرق بين الرتب يعبر عن الفرق بين الدرجات. مثلاً الفرق بين 60 و65 درجة حرارة هو قابل للقياس وهو 5 درجات وهو نفسه الفرق بين 40 و 45 درجة. - ليس له مقياس نسبي بمعنى إنّ درجة الحرارة 10 مئوية لا تعني أنّها ضعف درجة الحرارة 20 مئوية، أي إذا النسبة 10/20 ليس لها معنى مع هذا النوع من البيانات لأنّه مثلاً بتحويل هذه الدرجات إلى مقياس فهرنهايت تصبح 50 و 68 على الترتيب وبنسبتهما يظهر رقم أخر 50/68.	عند تقسيم درجات الطلاب إلى مستويات وكل مستوى يأخذ قيمة معينة، ودرجات الحرارة، قوة المحرك بالحصان
النسبي Ratio	جميع العمليات الرياضية	- عدد يدل على كم أو مقدار (قيم متصلة) - وضع الأشخاص في مقياس متصل يتكون من وحدات متساوية وله صفر مطلق - يمكن استخدام النسب لمقارنة الأرقام. يمكن إجراء العمليات الحسابية عليه. - له مقياس نسبي، بمعنى إذا نسبنا دخل موظف والذي يبلغ 20 ألف على دخل موظف أخر والذي يبلغ 10 آلاف أي 10/20=2 فإنّ الرقم الناتج له معنى أي انّ الموظف الأول يملك دخل أكبر من الثاني بمقدار الضعف.	الطول الوزن درجة الأمبير الفولت

- نود الإشارة إلى أن معرفتنا بمستويات قياس المتغيرات التى نستخدمها تسهل علينا اختيار الأداة الإحصائية المناسبة لتحليل البيانات المدروسة.

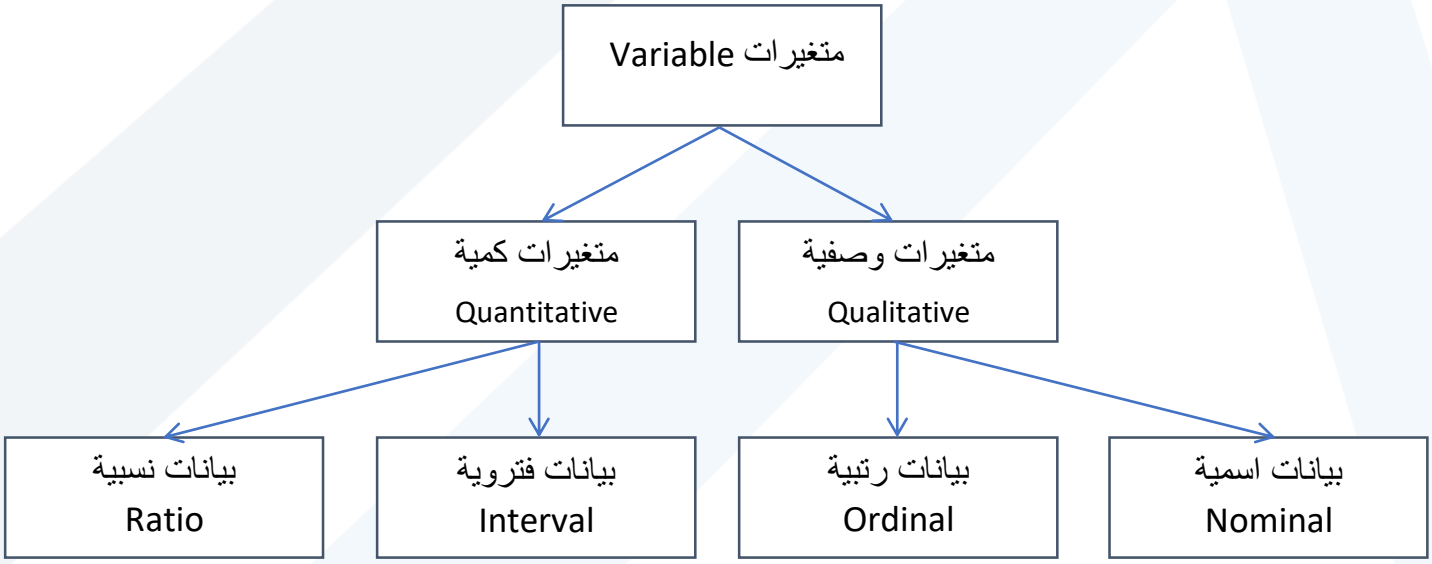
3- أنواع المتغيرات:

يمكن تمييز نوعين رئيسين من المتغيرات :

[1] متغيرات مستمرة تأخذ قيم مستمرة ضمن مجال معين (مقدار ثخانة سلك ملم،).

[2] ومتغيرات منقطعة تأخذ قيم صحيحة فقط او متقطعة (مثل نجاح تجربة أم لا).

ولكن بالإضافة لهذين النوعين الرئيسيين، تلعب أنواع البيانات التي ذكرناها سابقاً دوراً في تحديد انواع أخرى. ويمكن تلخيص أنواع المتغيرات وفقاً للمخطط التدفقي الآتي:



4- مصادر جمع البيانات (Collecting data Sources):

تنقسم مصادر جمع البيانات إلى نوعين: مصادر تاريخية ومصادر ميدانية:

[1] المصادر التاريخية: ويقصد بها السجلات المحفوظة والبيانات التي يتم نشرها من قبل الشركات والمؤسسات أو المنظمات الدولة.

[2] المصادر الميدانية: يتم وفقاً لها جمع البيانات بطريقة مباشرة من قبل الباحث وذلك عن طريق:

a. المقابلة الشخصية (Personal interview): حيث يقوم الباحث بمقابلة أفراد المجتمع او العينة المراد دراستها، ويتم توجيه الأسئلة لكل فرد وتسجيل إجابته. ومن عيوب هذه الطريقة أنّها تستغرق جهداً وتكاليف مادية عالية.

b. الاستمارة الإحصائية (الاستبيان Questionary): حيث يقوم الباحث بتصميم استمارة تشتمل أسئلة تتعلق بالبحث، ويقوم بتوجيه الاستبيان إما عن طريق الحصور الشخصي، البريد، الانترنت، عبر الهاتف، أو مواقع التواصل الاجتماعي.

c. التجربة (Experiment): حيث يقوم الباحث بإجراء تجربة ويسجل نتائجها، وذلك للحصول على معلومات مفيدة. فمثلاً لقياس تأثير بعض العوامل في منتج زراعي نقوم بتصميم التجربة لقياس شدة تأثير نوع السماد ونوع البذار في كمية المحصول الزراعي.

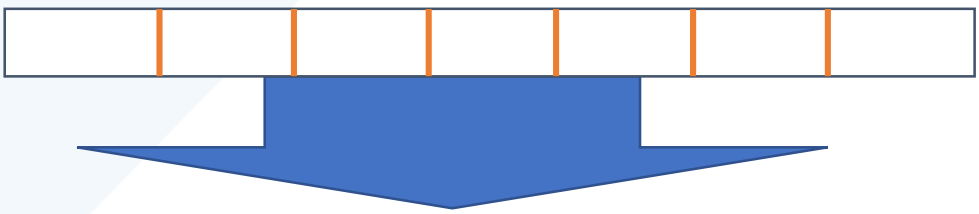
5- أساليب جمع البيانات (Collecting data methods):

يتم جمع البيانات بأحد الأسلوبين التاليين:

[1] الحصر الشامل (Complete Census): حيث يتم جمع البيانات من جميع أفراد مجتمع الدراسة، ويستخدم في المجتمعات الصغيرة (مدرسة، شركة، ..) أو في حالات تنفيذ البحث على المستوى الوطني مثل التعداد السكاني. وتكمن قوة هذا الأسلوب في إعطاء الباحث صورة حقيقية وكاملة عن مجتمع الدراسة، ومن عيوبه تكاليفه الباهظة وطول المدة الزمنية اللازمة لإجرائه.

العينات (Sampling): يستخدم أسلوب العينات عند دراسة مجتمعات كبيرة جداً، ويمكن تعريف العينة على انّها جزء من مجتمع الدراسة يتم اختياره بطريقة مناسبة ويمثل جميع خصائص المجتمع بصدق. ويوجد عدة طرق لسحب العينة:

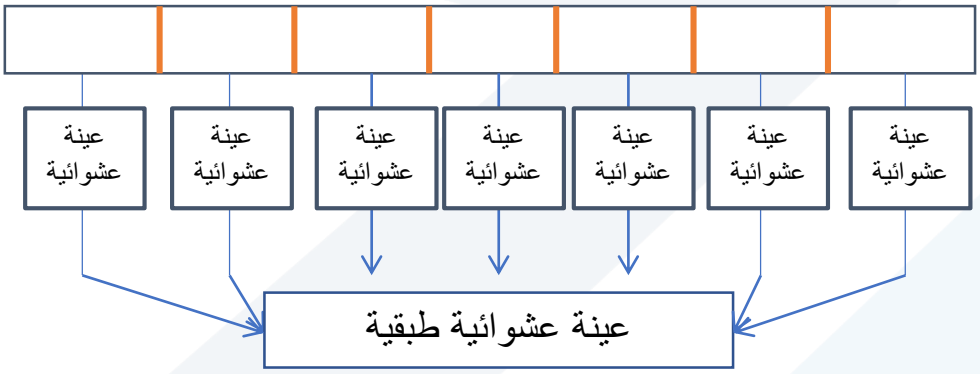
a. العينة العشوائية البسيطة (Simple Random Sample): هنا يتم سحب العينة من المجتمع بشكل مباشر وتستخدم في الحالات التي يكون فيها المجتمع متجانس.



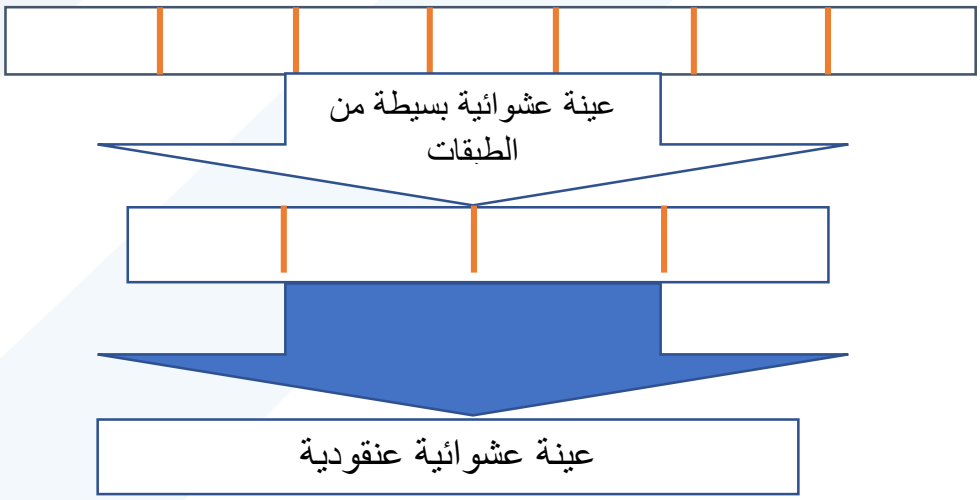
عينة عشوائية بسيطة

b. العينة العشوائية الطبقية (Stratified Random Sample): تستخدم هذه الطريقة عندما يكون المجتمع غير متجانس، حيث يتم تقسيم المجتمع إلى طبقات ومن ثم عينة عشوائية بسيطة من كل

طبقة. مثل تقسيم طلاب الجامعة إلى اختصاصات وسحب عينة عشوائية بسيطة من كل اختصاص، والاختصاص هنا يمثل الطبقة.



c. العينة العشوائية العنقودية (Cluster Random Sample): تستخدم هذه الطريقة عندما يكون المجتمع غير متجانس ويمكن تقسيمة إلى عدد كبير من الطبقات، هنا يتم سحب عينة عشوائية بسيطة من الطبقات، ومن ثم سحب عينة عشوائية بسيطة من كل طبقة. مثلاً، لسحب عينة لدراسة مستوى أداء السيرفرات في إحدى شركات التواصل الإجتماعي الضخمة نعتبر كل فرع عبارة عن طبقة ومن ثم نقوم بسحب عينة عشوائية بسيطة من هذه الطبقات (الفروع) ومن ثم نقوم بسحب عينة عشوائية بسيطة من السيرفرات من كل فرع من هذه الطبقات المختارة.



d. العينة العمدية او القصدية (Purposive sample): حيث يتم اختيار عينة تناسب هدف البحث، بحيث توفر في كل عنصر من العينة شروط محددة يرى الباحث أنّها تساعده في الحصول إلى نتائج أفضل في دراسته (مثال اختيار طلاب إحدى فروع الهندسة لإجراء دراسة عليهم أو اختيار مجموعة من المصابين بمرض معين لتطبيق الدراسة عليهم).

من أسلوب جمع البيانات عن طريق سحب العينة ينشأ مفهوم مهم جداً وهو ما يسمى بالاستدلال الإحصائي (Statistical Inference).

6- مفهوم الاستدلال الإحصائي:

الاستدلال الإحصائي هو أسلوب أو منهج عمل يدل على دراسة خصائص مجتمع ما عن طريق عينة مأخوذة من هذا المجتمع. أو بعبارة أخرى هي تعميم الخصائص المستمدة من عينة ما على المجتمع الذي أخذت منه هذه العينة. وبالتالي قد يكون هذا المفهوم مرادف لمصطلح التعميم الإحصائي.

بالتأكيد قد يكون الاستدلال الإحصائي مترافقاً مع أخطاء تسمى أخطاء القياس والتقدير. ويقاس هذا الخطأ عن طريق أدوات أو صيغ رياضية تسمى إحصائيات.

وبمقدار ما تكون طريقة سحب العينة مثالية وحقيقية بمقدار ما يكون خطأ التقدير صغيراً، والعينة المثالية هي العينة التي تمثل المجتمع الأم أفضل تمثيل. على سبيل المثال، إذا كان لدينا عينة تتألف من مفردات مأخوذة من مدرسة واحدة لطلاب الصف الخامس، فمن المحتمل أن تكون هذه العينة غير قابلة للتعميم على مجتمع الصف الخامس في المدينة بأكملها وبالتالي خطأ التقدير قد يكون كبيراً.

يقول Grosso modo من أجل ضمان أن تكون العينة ممثلة للمجتمع بشكل جيّد، من الضروري أن تسحب مفردات العينة بشكل كافي. حيث أنّه كلّما اقترب عدد مفردات العينة من عدد مفردات المجتمع الأم، يكون خطأ تقدير معالم المجتمع صغيراً.

7- تبويب، عرض وتمثيل البيانات:

بعد جمع البيانات مهما كان مصدرها (تاريخية أو ميدانية) أو أسلوب جمعها (حصر شامل، أو عن طريق عينة)، تكون البيانات خاماً ليست مرتبة أو منظمة. وبالتالي يصعب دراستها أو معرفة خصائصها وتحليلها. وبالتالي لا بدّ من تبويبها وترتيبها، ومن ثم عرضها وتمثيلها بيانياً لكي يسهل دراستها واستخلاص بعض النتائج منها.

7-1- الجداول التكرارية (Frequency table):

بالإضافة لأنواع المتغيرات وبياناتها التي رأيناها سابقاً، يمكن ذكر تصنيف آخر للبيانات يتعلق بالبيانات الوصفية (اسمية، رتبية) و الكمية (فئوية نسبية). حيث تختلف طريقة عرض وتمثيل كل نوع من هذه البيانات كما سنرى:

7-1-1- الجداول التكرارية والتجميعية للبيانات الوصفية :

يتضمن الجدول التكراري للبيانات الوصفية بشكل رئيسي عمودين الاول يتضمن القيم الوصفية والثاني يحتوي على التكرار المقابل لكل صفة. ويمكن أن يتضمن أعمدة أخرى مثل النسب التكرارية أو التكرار التجميعي الصاعد. ويطلق على الجدول التكراري اسم التوزيع التكراري لأنّه يوزع لكل صفة التكرار المقابل لها.

مثال 1:

ليكن لدينا الجدول الآتي الذي يبين مستويات 60 طالب في أحد المقررات:

المستوى	التكرار Frequency	النسبة Percent	النسب التجميعية
ممتاز	14	23,3	23,3
جيد جداً	14	23,3	46,7
جيد	12	20,0	66,7
مقبول	12	20,0	86,7
ضعيف	8	13,3	100,0
Total	60	100,0	

7-1-2- الجداول التكرارية والتجميعية للبيانات الكمية Freauency table :

تختلف طريقة إنشاء الجداول التكرارية للبيانات الكمية عن البيانات الوصفية، لأننا لو اعتبرنا أنّ كل قيمة من قيم المتغير الكمي هي فئة بحد ذاتها لنتج لدينا جدول توزيع تكرار يحوي عدد كبير من الفئات. لذلك يعتبر من الضروري تجميع كل مجموع من القيم ضمن فئة، بحيث يتم تقسيم جميع القيم إلى فئات متساوية الطول وإعطاء كل فئة عدد التكرارات للقيم المتضمنة في هذه الفئة. ويتم إنشاء الفئات باتباع الخطوات الآتية:

[1] ترتيب البيانات ترتيباً تصاعدياً من الأصغر إلى الأكبر.

[2] حساب عدد الفئات بما يتناسب مع البيانات، بحيث يتراوح بين 5 و 15 فئة، أو يمكن استخدام قاعدة ستارجس (Sturges rules) لتحديد عدد الفئات (k) لبيانات عددها (n):

$$k = 1 + 3.322log(n)$$

[3] تحديد الفئات بحساب طول الفئة، وذلك عن طريق حساب المدى وهو الفرق بين أكبر قيمة (Xmax) وأصغر قيمة (Xmin)، $R = X_{max} - X_{min}$، وقسمته على عدد الفئات (K) التي تم حسابها في الخطوة السابقة. وبالتالي يكون طول الفئة (w) مساوياً لـ $w = R/k$.

[4] يتم إنشاء حدود أول فئة $[L_1 = X_{min}, U_1 = L_1 + w[$، ويتم وضع فيها كل القيم بين X_{min} و $L_1 + w$. ومن ثم يتم إنشاء الفئة الثانية، $[L_2 = U_1, U_2 = L_2 + w[$، وهكذا ... إلى أن نصل إلى الفئة الأخيرة، $[L_k = U_{k-1}, X_{max}]$.

مثال 2:

ليكن لدينا البيانات الآتية عن 30 طالب المتعلقة بعلامات أحد المقررات:

45 40 57 43 56 63 39 68 44 49 61 47 58 34 45 59 48 54 75 54 48 37 62 40 50 54 58 48 49 46

المطلوب إنشاء جدول التوزيع التكراري لتمثيل علامات الطلاب في 7 فئات.

نلاحظ أنّ هذه البيانات من النوع الكمي حيث إذا اعتبرنا أنّ لكل قيمة فئة ينتج لدينا الكثير من الفئات، وبالتالي لا بدّ من تقسيم هذه القيم إلى فئات كما يأتي:

[1] طالما أنّ عدد الفئات مُعطى وهو 7 فئات فلا داعي لاستخدام قاعدة ستاردج (sturges).

[2] نقوم بحساب طول الفئة وذلك بقسمة المدى $R = X_{max} - X_{min} = 75 - 34 = 41$ على عدد الفئات k=7 ،

$w = \frac{R}{k} = \frac{41}{7} = 5.7$ أي يساوي تقريباً 6 درجات وهو طول الفئة.

[3] نقوم بإنشاء الفئة الأولى

$$[L_1 = X_{min}, U_1 = L_1 + w[= [34{,}34 + 6[= [34{,}40[$$

الفئة الثانية :

$$[L_2 = U_1, U_2 = L_2 + w[= [40{,}40 + 6[= [40{,}46[$$

الفئة الثالثة:

$$[L_3 = U_2, U_3 = L_3 + w[= [46{,}46 + 6[= [46{,}52[$$

الفئة الرابعة:

$$[L_4 = U_3, U_4 = L_4 + w[= [52{,}52 + 6[= [52{,}58[$$

الفئة الخامسة:

$$[L_5 = U_4, U_5 = L_5 + w[= [58{,}58 + 6[= [58{,}64[$$

الفئة السادسة:

$$[L_6 = U_5, U_6 = L_6 + w[= [64{,}64 + 6[= [64{,}70[$$

الفئة السابعة:

$$[L_7 = U_6, U_7 = X_{max}[= [70{,}75]$$

[4] يتم إنشاء جدول التوزيع التكراري كما يلي:

الفئات	التكرار f_i	النسبة f_i/n	النسبة التجميعية
$[34,40[$	3	$\frac{3}{30} * 100 = 10\%$	%10
$[40,46[$	6	$\frac{6}{30} * 100 = 20\%$	%30
$[46,52[$	8	$\frac{8}{30} * 100 = 27\%$	%57
$[52,58[$	6	$\frac{6}{30} * 100 = 20\%$	%77
$[58,64[$	5	$\frac{5}{30} * 100 = 17\%$	%94
$[64,70[$	1	$\frac{1}{30} * 100 = 3\%$	%97
$[70,75]$	1	$\frac{1}{30} * 100 = 3\%$	%100
المجموع	30	1	

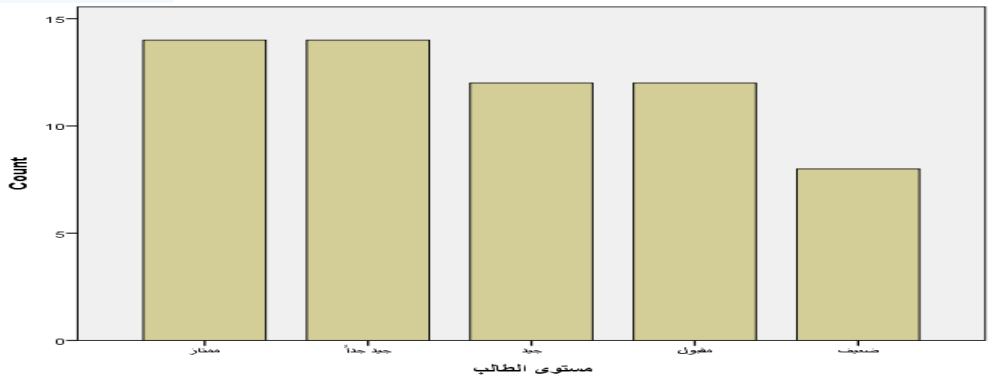
2-7- العرض البياني للجداول التكرارية Graphic representation:

1-2-7- المدرج التكراري (Frequency Histogram):

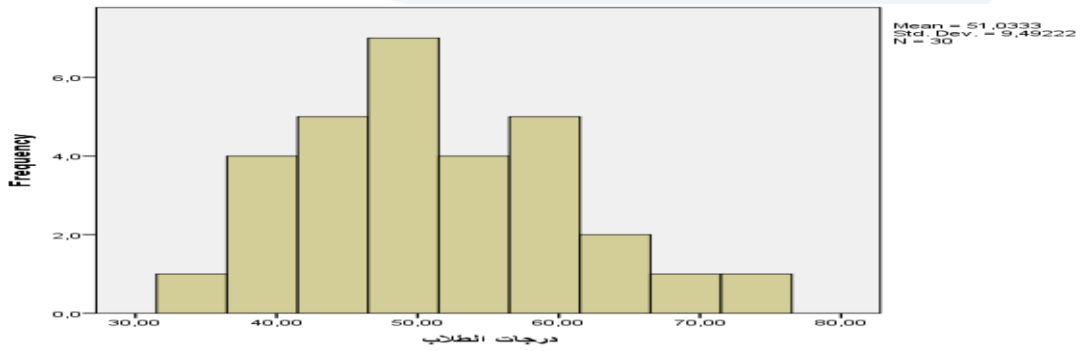
المدرج التكراري يتكون من مجموعة من الأعمدة المتجاورة التي تتوضع على المحور الأفقي (التي تمثل حالات الصفة المدروسة عندما تكن البيانات وصفية وفترات الفئات عندما تكون البيانات كمية) بحيث يساوي ارتفاع كل عمود مقدار التكرار المقابل للفئة.

نعود للمثالين السابقين:

تمثيل بيانات المثال 1:

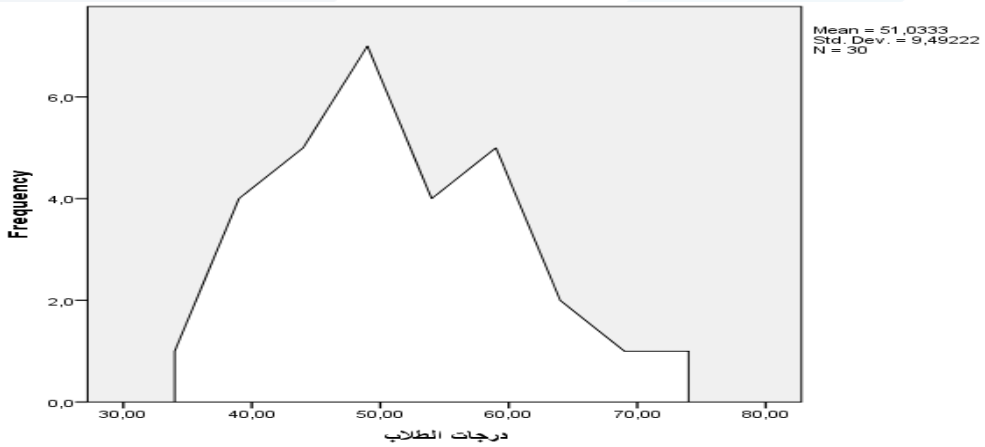


مثال 2:



7-2-2- المضلع التكراري (Frequency Polygon):

هو عبارة عن خط منكسر يصل بقطع مستقيمة النقاط التي مساقطها مراكز الفئات، ويبدأ الخط من الصفر ثم ينتهي بالصفر، للتوضيح أكثر لنرى الشكل البياني الآتي والذي يمثل المضلع التكراري لدرجات الطلاب:

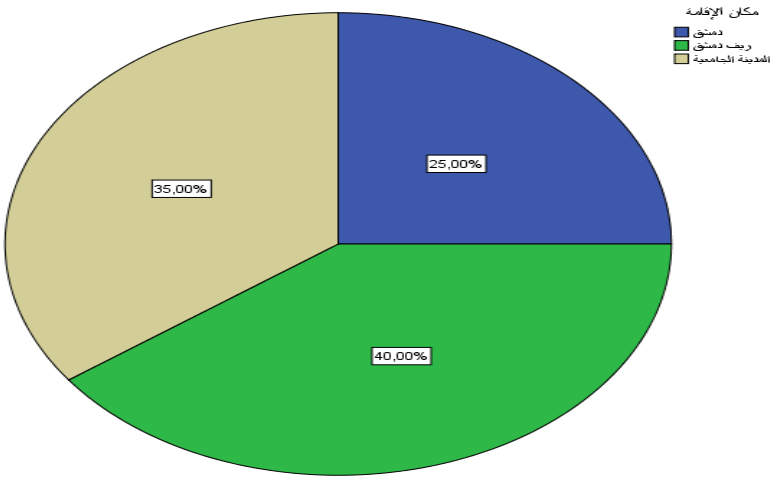


7-2-3- مخطط الدائرة (Pie Charts):

تعتبر هذه الوسيلة من وسائل العرض البياني التي تتيح للمتهم التعرف بنظرة سريعة على الظاهرة المدروسة. حيث يتم تجزئة الدائرة (360) إلى مجموعة من القطاعات التي تتناسب قياسات زواياها مع تكرارات الفئات.

مثال 3: ليكن لدينا البيانات الآتية عن مكان إقامة 100 طالب إحدى الجامعات:

مكان الإقامة	التكرار
المدينة الجامعية	35
دمشق	25
ريف دمشق	40
المجموع	100



حيث يتم حساب زاوية القطاع بضرب 360 بالنسبة المقابلة لكل فئة:

مكان الإقامة	التكرار	الزاوية
المدينة الجامعية	35	360*(35/100)=126
دمشق	25	360*(25/100)=90
ريف دمشق	40	360*(40/100)=144
المجموع	100	360

7-2-4- شكل الانتشار (Scatter diagram):

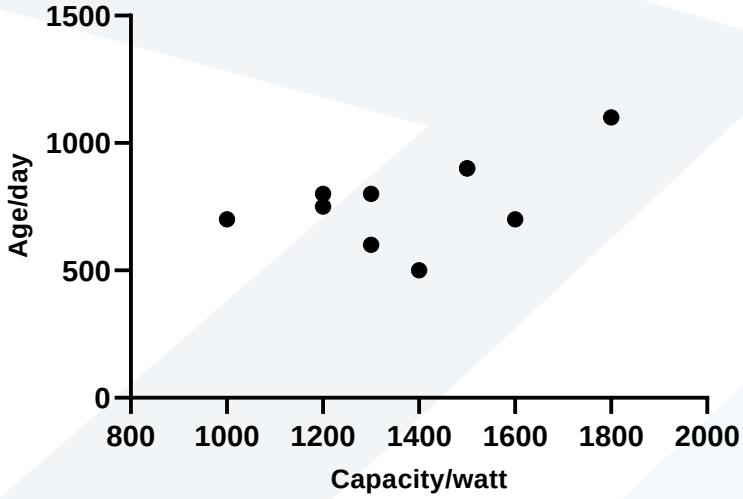
يمثل هذا النوع من الرسوم البيانية توزع قيم متغير (محور العامودي) بالنسبة لمتغير آخر (محور أفقي)، حيث يناسب البيانات الكمية. كذلك يمكن الاستفادة منه في معرفة مدى وجود علاقة بين المتغيرين.

مثال 4:

ليكن لدينا البيانات الآتية عن الإستطاعة والعمر (بالأيام) لـ 10 أجهزة:

watt/ capacity	1000	1500	1300	1200	1800	1500	1600	1400	1300	1200
day/Age	700	900	800	750	1100	900	700	500	600	800

يمكن استخدام شكل الانتشار لتمثيل العلاقة:



يمكن الإستفادة من شكل الانتشار للكشف فيما إذا كان انتشار نقاط احد المتغيرا ت بالنسبة لمتغير آخر يتم بشكل عشوائي أو وفقاً لنموذج ذو معالم، بمعنى آخر يسمح شكل الانتشار بتحديد طبيعة واتجاه العاقة بين المتغيرين المدروسين.

5-2-7- مخطط الساق والورقة Stem and Leaf diagram:

يعتبر هذا الأسلوب مناسباً للبيانات الكمية والمتصلة ، ويعلى الرغم من طريقة بنائه السهلة إلّا انّه وسيلة سريعة لوصف لبيانات كبيرة الحجم.

اسم هذا الأسلوب مشتق من طريقة بنائه، إذ يتم تقسيم الأرقام إلى قسمين ، الأول يدعى الساق Stem ويعبر عن العدد الاول أو الثاني في الرقم ، أمّا القسم الثاني يدعى الورقة Leaf فهو يوضع إلى جانب العدد المتبقي في هذا الرقم.

بعد ذلك يتم وضع قيم الساق Stem تحت بعضها وإلى جانبها قيم الأوراق Leaf.

مثلاً في حاول وجد رقم 85، يتم تمثيله في هذا المخطط بوضع العدد 8 على الساق والعدد 5 على الورقة. أما في حال كانت أغلب البيانات تحتوي أرقام أكبر من 100، مثل 152 في هذه الحالة يتم وضع أول عددين 15 على الساق بينما يتم وضع العدد 2 على الورقة.

للتوضيح اكثر لتكن الأمثلة الآتية.

مثال:

الشكل الآتي يعبر عن مخطط الساق والورقة لـ 25 قيمة إنتاج في مصنع معين:

Stem	Leaf
6	1 3 4 5 5 6
7	0 1 1 3 5 7 8 8 9
8	1 3 4 4 7 8 8
9	2 3 5

حيث يلاحظ من المخطط أنّ قيمة الإنتاج الأكثر تكراراً هو القيمة التي تقارب 70، مثلاً يوجد 70، 71، 73، 75، ... 79.

ولاستنتاج خصائص اكثر لإنتاج هذه المصنع يمنك تفصيل المخطط بشكل أكبر كما يلي:

Stem	Leaf
6L	1 3 4
6U	5 5 6
7L	0 1 1 3
7U	5 7 8 8 9
8L	1 3 4 4
8U	7 8 8
9L	2 3
9U	5

يلاحظ أنّ هذا المخطط يعطي معلومات تفصيلية أكثر دقة لإنتاج المصنع من المخطط السابق.

مثال:

البيانات الآتية تخص قوة الضغط (pound/ps) لـ 80 سبيكة من الألمنيوم:

105	221	183	186	121	181	180	143
97	154	153	174	120	168	167	141
245	228	174	199	181	158	176	110
163	131	154	115	160	208	158	133
207	180	190	193	194	133	156	123
134	178	76	167	184	135	229	146
218	157	101	171	165	172	158	169
199	151	142	163	145	171	148	158
160	175	149	87	160	237	150	135
196	201	200	176	150	170	118	149

بالضبع من الصعوبة استخراج خصائص قوة التحمل لسبائك الألمنيوم من هذا الشكل الخام للبيانات، وبالتالي من الضروري عرض هذه البيانات بشكل مفيد أكثر وذلك باستخدام مخطط الساق والورقة:

Stem	Leaf	Frequency
7	6	1
8	7	1
9	7	1
10	5 1	2
11	5 8 0	3
12	1 0 3	3
13	4 1 3 5 3 5	6
14	2 9 5 8 3 1 6 9	8
15	4 7 1 3 4 0 8 8 6 8 0 8	12
16	3 0 7 3 0 5 0 8 7 9	10
17	8 5 4 4 1 6 2 1 0 6	10
18	0 3 6 1 4 1 0	7
19	9 6 0 9 3 4	6
20	7 1 0 8	4
21	8	1
22	1 8 9	3
23	7	1
24	5	1

بمقارنة أرقام الجدول مع أرقام المخطط يلاحظ مثلاً أن الرقمين 105، و 101 ، تم تمثياهما بوضع 10 على الساق و وبضع ورقتين تمثلان العدد1 والعدد 5 .

من المخطط السابق يمكن استخلاص خاصية أنّ غالبية سبائك الألمنيوم المدورسة تتحمل ضغط يتراوح 100 و 200 باوند في البوصة المربعة. بينما القيم المركزية تتراوح بين 142 و169.

العمود الأخير في هذه المخطط يعطي معلومات عن تكرار كل قيمة مثلاً يوجد قيمة واحد للرقم 76 بينما يوجد 3 قيم للرقم 12 وبمعنى آخر يوجد 3 أوراق للعدد 12. كذلك يوجد 8 أوراق Leaf على الساق رقم 14.

يمكن تلخيص أنواع وسائل العرض البياني السابقة وفقاً لنوع البيانات كما يأتي:

نوع البيانات Data type	طريقة العرض البياني
وصفية Qualitative	المدرج التكراري، المضلع التكراري، الدائرة
كمية Quantitative	المدرج التكراري (فئات)، المضلع التكراري (فئات)، شكل الانتشار، مخطط الساق والورقة