الأهداف التعليمية لهذا الدرس:

من المتوقع مع نهاية هذا الدرس أن تكون قادراً على التعرف على:

- الخطوات المنهجية للتحليل الاحصائي في البحث العلمي.
- أنواع البيانات.
- قياس البيانات.
- أخطاء البيانات.
- عرض وتنظيم البيانات (الجداول التكرارية – التمثيل البياني)
- المقاييس الاحصائية (مقاييس النزعة المركزية)
(الوسط الحسابي – الوسيط – المنوال)

الخطوات المنهجية للتحليل الإحصائي في البحث العلمي

البيانات وأنواعها

- إن من أهم المبادئ الأساسية في التحليل الإحصائي معرفة البيانات المناسبة والتي قد تجيب على أسئلة الباحث حول ظاهرة محل اهتمامه .
- البيانات : هي مجموعة القيم التي يتم جمعها من مفردات المجتمع أو العينة لخاصية (متغير) معينة .

أنواع البيانات :

١- نوعية أو وصفية .
٢- كمية ، وتنقسم إلى :
أ) منفصلة .
ب) متصلة .

انواع البيانات

البيانات النوعية (الوصفية) : Qualitative Data

- هي البيانات التي يمكن حصرها في عدة اوجه وصفية ولا يمكن إجراء عمليات رياضية حسابية عليها كالجمع والطرح .

أمثلة :
نوع الشخص (ذكر – أنثى)
الجنسية (سعودي ، مصري ، فرنسي)
المستوى الاقتصادي للأسر (غنية ، متوسطة ، فقيرة)

البيانات الكمية : Quantitative Data

- هي البيانات التي يتم الحصول عليها على شكل أعداد ويمكن ترتيبها .

أمثلة :
الرواتب الشهرية
أطوال الطلاب
عدد الحجرات في المسكن
درجات الاختبار في مادة الإحصاء

البيانات الكمية المنفصلة : Discrete Data

- هي البيانات التي يمكن عدها وبمعنى آخر هي البيانات التي تكون مفرداتها منفصلة عن بعضها البعض ، ولا تأخذ قيمًا كسرية ، تأخذ فقط قيمًا صحيحة .

أمثلة :
عدد الحجرات في المسكن
عدد أفراد الأسرة
عدد حوادث السيارات اليومية
عدد الأسهم المباعة من كل شركة

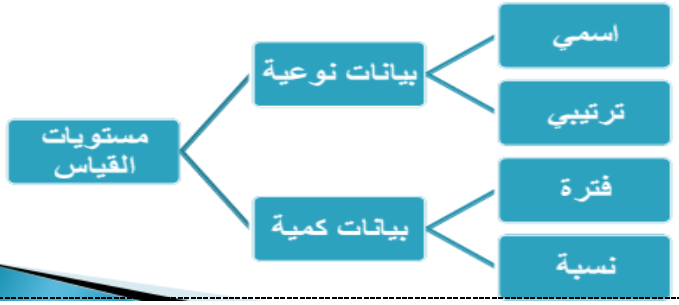
البيانات الكمية المتصلة : Continuous Data

هي البيانات التي لا يمكن عدها إنما يتم الحصول عليها عن طريق القياس ويمكن أن تأخذ أي قيمة داخل مدى معين سواء كانت صحيحة أو كسرية .

أمثلة:
الدخل الشهري
درجات الحرارة
المعدل الدراسي
أوزان الطلاب

قياس البيانات

تقاس بيانات الخاصية والتي تكون محل الاهتمام من المجتمع أو العينة بأحد أربع قياسات :



المقياس الاسمي : Nominal Scale

- مجموعة من الأوجه أو الصفات التي يأخذها المتغير الوصفي ، ويمكن أن تعطى الصفات أرقام تعكس مدلول الصفة ولكن ليس لها معنى رياضي في مفهوم أكبر أو أصغر (مثل فصيلة الدم والجنسية)

المقياس الترتيبي (التفضيلي) : Ordinal Scale

- مجموعة من الأوجه أو الصفات التي يأخذها المتغير الوصفي مع إمكانية ترتيبها ، ويمكن أن تعطى الصفات أرقام تعكس مدلول الصفة ولها معنى رياضي في مفهوم أكبر أو أصغر ولكن لا تعكس معنى حقيقي للفروق (مثل المستوى التعليمي ومدى الموافقة على رأي معين) .

مقياس الفترة : Interval Scale

- مجموعة من الأعداد أو القيم التي يأخذها المتغير الكمي ، وتعطى الصفات أرقام تعكس مدلول الصفة ولها معنى رياضي أكبر أو أصغر وتعكس معنى حقيقي للفروق ولكن ليس للصفر معنى حقيقي أي أن المتغير إذا كان مساويا للصفر فلا يعني هذا انعدام الصفة (مثل درجة الحرارة ودرجة امتحان الذكاء) .

مقياس النسبة : Ratio Scale

- مجموعة من الأعداد أو القيم التي يأخذها المتغير الكمي ، وتعطى الصفات أرقام تعكس مدلول الصفة ولها معنى رياضي أكبر أو أصغر وتعكس معنى حقيقي للفروق ، والصفر له معنى حقيقي أي أن المتغير إذا كان مساويًا للصفر فهذا يعني انعدام الصفة (مثل الوزن ، الطول) .

أنواع الأخطاء التي تتعرض لها البيانات عند جمعها :

١- خطأ التحيز :

- هو الخطأ الذي يحدث عند جمع البيانات ومصدر هذا الخطأ إما من الباحث أو من مفردات المجتمع محل الدراسة . ويمكن أن يحدث هذا النوع من الخطأ عند إجراء الحصر الشامل أو عند استخدام العينة العشوائية .

مثال :

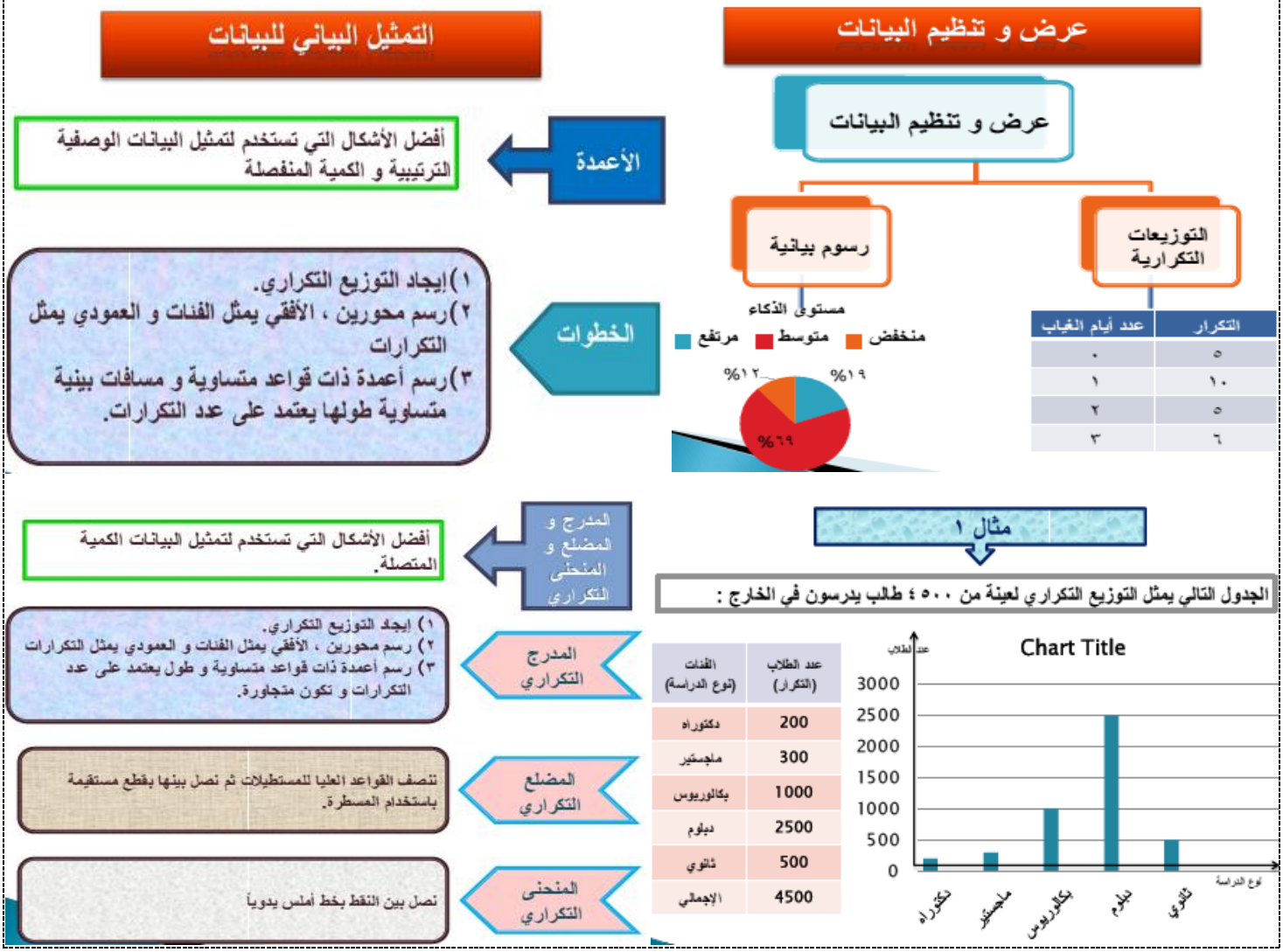
المغالاة في الإجابة من قبل المبحوث كتقليل الدخل وتضخيم النفقات، إهمال مفردات معينة واستبدالها بأخرى ، استخدام أسئلة إيحائية ، أن يقوم جامع البيانات بملء البيانات بنفسه دون مقابلة المبحوث .

٢ - خطأ المعاينة العشوائية :

- هو الخطأ الذي يحدث عند إجراء الدراسة الإحصائية بأسلوب العينة العشوائية ويرجع فقط إلى الصدفة وليس لأخطاء من الباحث والمبحوث .

مثال :

عند إجراء دراسة متوسط درجة الذكاء لدى الأطفال في عمر السادسة بأسلوب العينة العشوائية البسيطة وجد أن متوسط الدرجة في العينة يقل عن الدرجة العلمية النظرية التي حددها علماء علم النفس بمقدار درجتين



السلاسل الزمنية
Time Series

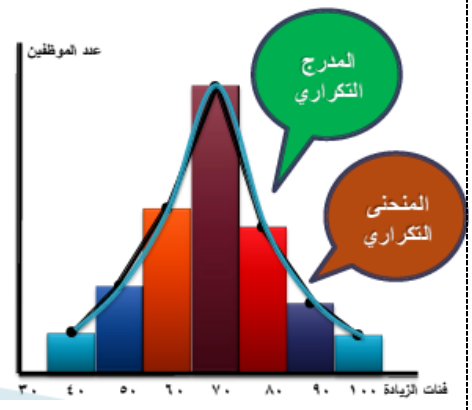
شكل السلاسل الزمنية يمكن أن يحدد :

- **الاتجاه العام** اتجاه تطور السلسلة رغم التذبذبات
  - اتجاه الزيادة.
  - اتجاه النقصان
- **التغيرات الموسمية** التغيرات التي تتكرر بانتظام خلال فترة زمنية أقل من سنة
  - تحدث في المواسم
- **التغيرات الدورية** تحدث في فترات زمنية أكثر من سنة.
  - كل خمس أو عشر سنوات
- **التغيرات العرضية** تحدث نتيجة الحوادث المفاجئة
  - الزلازل - الفيضانات - الحروب

مثال

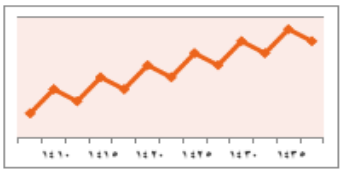
الجدول التالي يبين توزيع عينة من ١٠٠ موظف حسب الزيادة التي حصلوا عليها في الراتب

فئات الزيادة	عدد الموظفين
30-	4
40-	11
50-	20
60-	36
70-	17
80-	8
90-	4
Σ	100



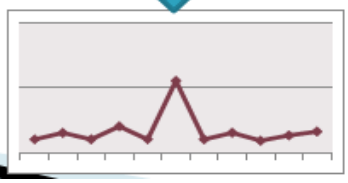
السلاسل الزمنية: الاتجاه العام

تغيرات دورية



تغيرات موسمية

تغيرات عرضية



وصف البيانات
Data Description

أولاً

مقاييس النزعة المركزية

القيم التي تقترب منها أو تتركز حولها أو تتوزع بالقرب منها معظم البيانات

المقاييس الإحصائية الوصفية

تشمل كل من

- مقاييس النزعة المركزية.
- مقاييس التشتت.
- معاملات الالتواء.
- و غيرها.....

منها

- الوسط الحسابي
- الوسيط
- المنوال

الوسط الحسابي (المتوسط)
Average

الوسط الحسابي لمجموعة من البيانات هو حاصل جمع هذه البيانات مقسوماً على عددها.
يرمز بالرمز μ لمتوسط المجتمع.
و يرمز بالرمز $\bar{X}$ لمتوسط العينة.

طريقة حساب الوسط الحسابي

البيانات الغير مبوبة

إذا كانت $X_1, X_2, \dots X_N$ تمثل بيانات مجتمع ما

الوسط الحسابي يعطى بالعلاقة:

$$\mu = \frac{\sum X}{N} = \frac{X_1 + X_2 + \cdots X_N}{N}$$

إذا كانت $x_1, x_2, \dots x_N$ تمثل بيانات عينة من المجتمع

الوسط الحسابي يعطى بالعلاقة:

$$\bar{x} = \frac{\sum x}{N} = \frac{x_1 + x_2 + \cdots x_N}{N}$$

أمثلة

البيانات التالية تمثل عدد أيام الغياب خلال ربع السنة لعينة عشوائية من الموظفين، أوجد الوسط الحسابي

10 2 3 7 5 9 6

$$\bar{x} = \frac{\text{مجموع البيانات}}{\text{عددها}} = \frac{10+2+3+7+5+9+6}{7} = \frac{42}{7} = 6 \text{ أيام}$$

قد يساوي المتوسط إحدى القيم و قد يكون مختلف

شركة لديها 6 مصانع في مناطق مختلفة لإنتاج منتج معين سعتها الإنتاجية كما يلي

1200 2500 1000 2000 3000 1000

أوجد الوسط الحسابي لإنتاج الشركة الكلي

$$\mu = \frac{\text{مجموع البيانات}}{\text{عددها}} = \frac{1200+2500+1000+2000+3000+1000}{6} = \frac{10700}{6} = 1783.3 \text{ وحدة}$$

طريقة حساب الوسط الحسابي

البيانات المبوبة

الفئات	التكرار f_i	مراكز الفئات x_i	$x_i f_i$
a—	f_1	x_1	$x_1 f_1$
b—	f_2	x_2	$x_2 f_2$
c—	f_3	x_3	$x_3 f_3$
f—	f_k	x_k	$x_k f_k$
Σ	$\Sigma f = N$		$\Sigma x_i f_i$

+h +h

نحسب طول الفئة $h = b - a$

نحسب مركز الفئة الأولى $x_1 = \frac{a+b}{2}$

مركز الفئة الثانية= مركز الفئة الأولى +h

نحسب المتوسط $\bar{X} = \frac{\Sigma x_i f_i}{\Sigma f_i}$

مثال

الجدول التالي يوضح الأجر اليومي لعينة عشوائية من ٣٦ عامل بالريال. أوجدي الوسط الحسابي

فئات الأجر	30 –	34 –	38 –	42 –	46 –	50 –	54 – 58
عدد العمال	1	3	7	10	8	4	3

نكمل الجدول

فئات الأجر	عدد العمال f_i	مراكز الفئات x_i	$x_i f_i$
30—	1	32	32
34—	3	36	108
38—	7	40	280
42—	10	44	440
46—	8	48	384
50—	4	52	208
54—58	3	56	168
Σ	36		1620

نحسب المتوسط

$$\bar{X} = \frac{\Sigma x_i f_i}{\Sigma f_i} = \frac{1620}{36} = 45 \text{ ريال}$$

- نوجد طول الفئة = الحد الأدنى للفئة الثانية – الحد الأدنى للفئة الأولى

$$h = 34 - 30 = 4$$

- نوجد مركز الفئة الأولى

$$x_1 = \frac{30+34}{2} = 32$$

- نوجد مراكز الفئات الأخرى بإضافة طول الفئة(4) في كل مرة.

الوسيط Mediator

القيمة العددية التي تقسم البيانات إلى قسمين متساويين بعد ترتيبها تصاعدياً أو تنازلياً.

يرمز للوسيط بالرمز m

مزايا و عيوب الوسط الحسابي

المزايا

- سهولة حسابه و التعامل معه جبرياً
- لا يحتاج لترتيب البيانات.
- تدخل في حسابه جميع القيم.
- يعتبر الأساس في معظم عمليات الإحصاء الاستدلالي.

العيوب

- لا يمكن إيجاده للبيانات الوصفية.
- لا يمكن إيجاده من خلال الرسم.
- يتأثر بالقيم الشاذة.
- قد لا يساوي عدداً صحيحاً أو أي من القيم الداخلة في حسابه.

طريقة حساب الوسيط

البيانات الغير مبوبة

إذا كانت $x_1, x_2, \ldots . x_n$ تمثل بيانات عينة من المجتمع فإن الوسيط يحسب كالتالي:

١. نرتب البيانات تصاعدياً أو تنازلياً.

٢. نوجد موقع الوسيط $\frac{n+1}{2}$.

٣. إذا كان n عدد فردي فإن الناتج يكون عدد صحيح و بالتالي الوسيط هو $x_{\frac{n+1}{2}}$.

٤. إذا كان n عدد زوجي فإن الناتج يكون عدد غير صحيحو بالتالي الوسيط هو الوسط الحسابي للقيمتين اللتين يقع بينهما العنصر $x_{\frac{n+1}{2}}$.

أمثلة

أوجد الوسيط للأجور اليومية بالدولار للبيانات التالية:

العينة الأولى : 80 50 40 70 50

العينة الثانية : 80 30 40 60 70 50

العينة الأولى

نعيد ترتيب البيانات :

80 70 50 50 40

عدد البيانات n = 5

$$\frac{n+1}{2} = \frac{6}{2} = 3$$

$$m = X_3 = 50$$

العينة الثانية

نعيد ترتيب البيانات :

80 70 60 50 40 30

عدد البيانات n = 6

$$\frac{n+1}{2} = \frac{7}{2} = 3.5$$

إذن الوسيط يقع بين X_3 , X_4

$$m = \frac{X_3 + X_4}{2} = \frac{50+60}{2} = 55$$

المنوال
Mode

هو القيمة التي تكررت أكثر من غيرها.
القيمة الأكثر شيوعاً أو تكراراً.
يرمز للمنوال بالرمز D

مزايا و عيوب الوسيط

المزايا

- سهولة حسابه .
- لا يتأثر بالقيم الشاذة.
- يمكن إيجاده عن طريق الرسم.
- يمكن إيجاده للبيانات الوصفية الترتيبية.

العيوب

- يحتاج لترتيب البيانات.
- لا تدخل في حسابه جميع القيم.
- يصعب استخدامه في الإحصاء الاستدلالي لصعوبة التعامل معه جبرياً.

طريقة حساب المنوال

البيانات الغير مبوبة

أمثلة

أوجد المنوال لكل من العينات التالية:

المخالفات المرورية التي ارتكبها كل شخص في عينة من 10 أشخاص:

4	6	4	1	0	3	4	5	4	1

المنوال: $D = 4$

أحادية المنوال

تقديرات عينة من 10 طلاب :

C	C	D	B	D	F	D	A	C	A

مرات3 تكرر D
مرات3 تكرر C

المنوال: D, C

ثنائية المنوال

جنسيات عينة من 10 حجاج آجانب :

لبناني	مصري	لبناني	تونسي	مصري
تونسي	سوداني	كويتي	قطري	أمريكي

كل من المصري التونسي و اللبناني تكرر مرتين

المنوال: تونسي ، لبناني، مصري

ثلاثية المنوال (متعددة المنوال)

عدد أيام الغياب عينة من 10 طلاب خلال شهر :

10	8	7	3	6	5	0	4	2	1

جميع القيم تكررت مرة واحدة

المنوال: غير موجود

عديمة المنوال

طريقة حساب المنوال

البيانات المبوبة

البيانات الوصفية أو الكمية المنفصلة

يحسب المنوال من التعريف مباشرة أي القيمة التي يقابلها أكبر تكرار

أمثلة

الجدول التالي يمثل عدد أجهزة الهاتف النقال المباعة خلال شهر في أحد المحلات و المطلوب إيجاد المنوال

التكرار	عدد الأجهزة المباعة
6	0
13	1
6	2
3	3
2	4

أكبر تكرار

المنوال: 1

أوجدي المنوال لبيانات عينة عشوائية من العمال موزعين حسب الحالة الاجتماعية:

عدد العمال	الحالة الاجتماعية
40	متزوج
25	أعزب
14	مطلق
10	أرمل

أكبر تكرار

المنوال: متزوج

مزايا و عيوب المنوال

المزايا

- سهولة حسابه و إيجاده
- لا يتأثر بالقيم الشاذة.
- يمكن حسابه في الجداول التكرارية المفتوحة.
- يمكن إيجاده لجميع أنواع البيانات الوصفية و الكمية.
- يمكن إيجاده من خلال الرسم
- يعتبر المقياس الوحيد الذي يمكن استخدامه للبيانات الاسمية

العيوب

- لا تدخل في حسابه جميع البيانات.
- قد لا يقع في مركز البيانات بل في طرفها.
- تتغير قيمته باختلاف طريقة اختيار الفئات.
- يصعب التعامل معه في الإحصاء الاستدلالي لأنه قد تكون له أكثر من قيمة.