المحاضرة الرابعة عشر
عرض وتحليل البيانات

الخطوات المنهجية للتحليل الإحصائي في البحث العلمي:

جمع البيانات ← تنظيم وعرض البيانات ← وصف وتلخيص البيانات ← تحليل البيانات واتخاذ القرار

*** 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:**

1- نوعية (وصفية).

2- كمية. **وتقسم إلى:** (مفصلة، متصلة).

- **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 الشهري، 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:

()" **مستويات القياس: 1- بيانات نوعية:** (اسمي، ترتيبي). **2- بيانات كمية:** (فترة، نسبية).

- **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 مسآويًا للصفر فهذا يعني انعدام الصفة (مثل الوزن، الطول).

أنواع الأخطاء التي تتعرض لها البيانات عند جمعها:

1- 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.

مثل: 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.

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

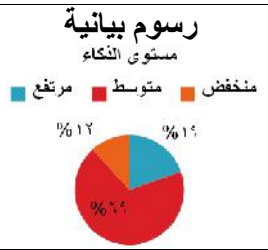
مثل: عند إجراء دراسة متوسط درجة الذكاء لدى الأطفال في عمر السادسة بأسلوب العينة العشوائية البسيطة وجد أن متوسط الدرجة في العينة يقل عن الدرجة العلمية النظرية التي حددها علماء علم النفس بمقدار درجتين.

- **عرض وتنظيم البيانات:**

التوزيعات التكرارية

التكرار	عدد أيام الغياب
٥	٠
١٠	١
٥	٢
٦	٣

رسوم بيانية



التمثيل البياني للبيانات:

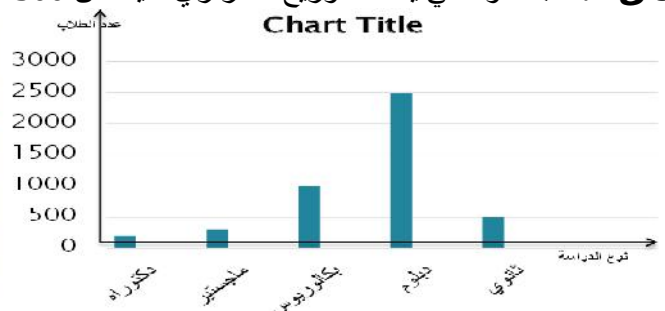
- **الأعمدة:** أفضل الأشكال التي تستخدم لتمثيل البيانات الوصفية الترتيبية والكمية المنفصلة.
- **الخطوات:** 1- إيجاد التوزيع التكراري.

2- رسم محورين، الأفقي يمثل الفئات والعمودي يمثل التكرارات.

3- رسم أعمدة ذات قواعد متساوية و مسافات بينية متساوية طولها يعتمد على عدد التكرارات.

مثل 1: الجدول والتالي يمثل التوزيع التكراري لعينة من 4500 طالب يدرسون في الخارج:

الفئات (نوع الدراسة)	عدد الطلاب (التكرار)
دكتوراه	200
ماجستير	300
بكالوريوس	1000
دبلوم	2500
ثانوي	500
الإجمالي	4500



- **المدرج و المضلع و المنحنى التكراري:** أفضل الأشكال التي تستخدم لتمثيل البيانات الكمية المتصلة.
- **المدرج التكراري:** 1- إيجاد التوزيع التكراري.

2- رسم محورين، الأفقي يمثل الفئات والعمودي يمثل التكرارات.

3- رسم أعمدة ذات قواعد متساوية و طول يعتمد على عدد التكرارات و تكون متجاورة.

- **المضلع التكراري:** ننصف القواعد العليا للمستطيلات ثم نصل بينها بقطع مستقيمة باستخدام المسطرة.
- **المنحنى التكراري:** نصل بين النقط بخط أملس يدوياً.

مثل 2: الجدول التالي يبين توزيع عينة من 100 موظف حسب الزيادة التي حصلوا عليها في الراتب

فئات الزيادة	عدد الموظفين
30-	7
40-	11
50-	20
60-	36
70-	17
80	8
90	4
>	100

السلاسل الزمنية: شكل السلاسل الزمنية يمكن أن يحدد:

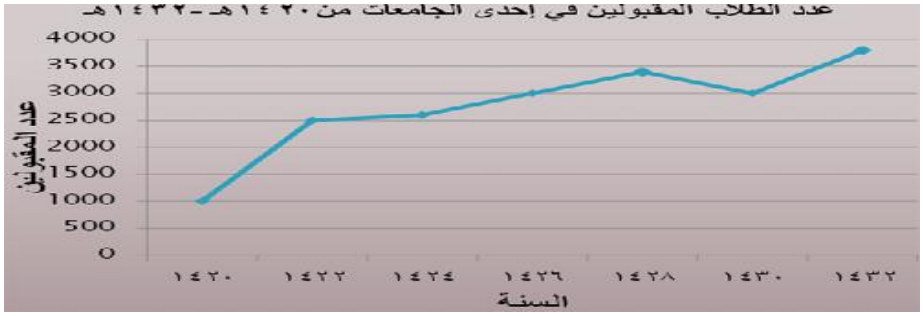
الاتجاه العام: اتجاه تطوير السلسلة رغم التذبذبات: (اتجاه الزيادة، اتجاه النقصان).

التغيرات الموسمية: التغيرات التي تتكرر بانتظام خلال فترة زمنية أقل من سنة: (تحدث في المولد).

التغيرات الدورية: تحدث في فترات زمنية أكثر من سنة: (كل خمس أو عشرسنوات).

التغيرات العرضية: تحدث نتيجة الحوادث المفاجئة: (الزلازل، الفيضانات، الحروب).

السلاسل الزمنية: **الاتجاه العام**



وصف البيانات Data Description : المقاييس الإحصائية الوصفية تشمل كل من:

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

أولاً مقاييس النزعة المركزية: القيم التي تقترب منها أو تتركز حولها أو تتوزع بالقرب منها معظم البيانات:

منها: 1- **الوسط الحسابي.** 2- الوسيط. 3- **المنوال.**

الوسط الحسابي: 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 غير مبوبة:

- **إذا كانت** $x_1, x_2, \dots x_n$ تمثل بيانات مجتمع ما: $\mu = \frac{\sum X}{N} = \frac{X_1+X_2+\dots+X_N}{N}$
- **إذا كانت** $N_1, N_2, \dots N_n$ تمثل بيانات عينة من المجتمع: $\bar{x} = \frac{\sum X}{N} = \frac{X_1+X_2+\dots+X_N}{N}$

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

6 5 9 7 3 2 10

الحل: ايام $\overline{X} = \frac{\text{مجموع البيانات}}{\text{عددها}} = \frac{10+2+3+7+9+5+6}{7} = 6$

*** شركة لديها 6 مصانع في مناطق مختلفة لإنتاج منتج معين سعتها الإنتاجية كما يلي، أوجد الوسط الحسابي لإنتاج الشركة الكلي:**

1000 3000 2000 1000 2500 1200

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

طريقة حساب الوسط الحسابي: (البيانات مبوبة)

الفئات	التكرار	مركز	
A –			
B –			
C –			
F –			

نحسب طول الفئة $h=b+a$

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

مركز الفئة الثانية = مركز الفئة الأولى + h

مثال: الجدول التالي يوضح الأجر اليومي لعينة عشوائية من 36 عامل بالريال. **أوجدي الوسط الحسابي**

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

- نوجد **طول الفئة** = الحد الادنى 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.

نكمل الجدول:

فئات الأجر	عدد العمال f_i	مراكز الفئات x_1	$x_i f_i$
- 30	1	32	32
- 34	3	36	108
- 38	7	40	280
- 42	10	44	440
- 46	8	48	384
- 50	4	52	208
58 – 54	3	56	168
	36		1620

نحسب المتوسط :

مزايا الوسط الحسابي:	عيوب الوسط احسابي:
سهولة حسابه و التعامل معه جبرياً.	**لا** يمكن إيجاده للبيانات الوصفية
لا يحتاج لترتيب البيانات.	لا يمكن إيجاده من خلال الرسم
تدخل في حسابه جميع القيم.	يتأثر بالقيم الشاذة
يعتبر الأساس في معظم عمليات.	قد لا يساوي عدد صحيح أو أي من القيم الداخلة في حسابه.
الإحصاء الاستدلالي.	

الوسيط Mediator : القيمة العددية التي تقسم البيانات إلى قسمين متساويين بعد ترتيبها تصاعدياً أو تنازلياً. يرمز للوسيط بالرمز n 31

طريقة حساب الوسيط: (البيانات الغير مبوبة) **إذا كانت** $x\,1$, $x\,2$,.... $x\,n$ تمثل بيانات عينة من المجتمع فإن الوسيط يحسب **كالتالي:**

1- نرتب البيانات تصاعدياً أو تنازلياً.

2- نوجد موقع الوسيط . $n+1\,2$ / 2

3- **إذا كان** n عدد فردي فإن الناتج يكون عدد صحيح و بالتالي الوسيط هو . $x\,n+1\,2$ /2

4- **إذا كان** n عدد زوجي فإن الناتج يكون عدد غير صحيحو بالتالي الوسيط هو الوسط الحسابي للقيمتين اللتين يقع بينهما العنصر x 2 / $n+1\,2$. 32

أمثلة: أوجد الوسيط للأجور اليومية بالدولار للبيانات التالية

العينة الأولى	50	70	40	50	80		
العينة الثانية	50	70	60	40	30	80	

العينة الأولى: نعيد ترتيب البيانات: 40 50 50 70 80

عدد البيانات = 5

$$M = X_3 = 50$$

العينة الثانية: نعيد ترتيب البيانات: 30 40 50 60 70 80

عدد البيانات = 6

إذا الوسيط يقع بين x_4, x_3

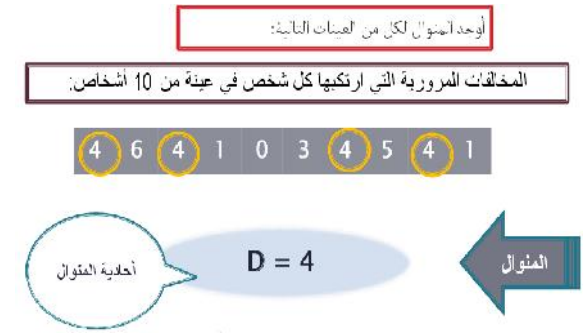
$$M = \frac{x_3+x_4}{2} = \frac{50+60}{2} 55$$

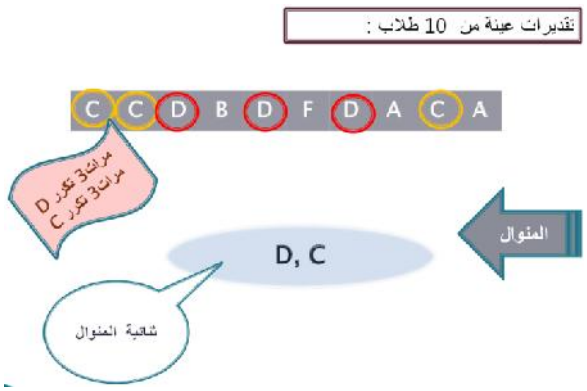
مزايا الوسيط:	عيوب الوسيط:
سهولة حسابه.	يحتاج لترتيب البيانات.
لا يتأثر بالقيم الشاذة.	لا تدخل في حسابه جميع القيم.
يمكن إيجاده عن طريق الرسم.	يصعب استخدامه في الإحصاء.
يمكن إيجاده للبيانات الوصفية الترتيبية.	الاستدلالي لصعوبة التعامل معه جبرياً.

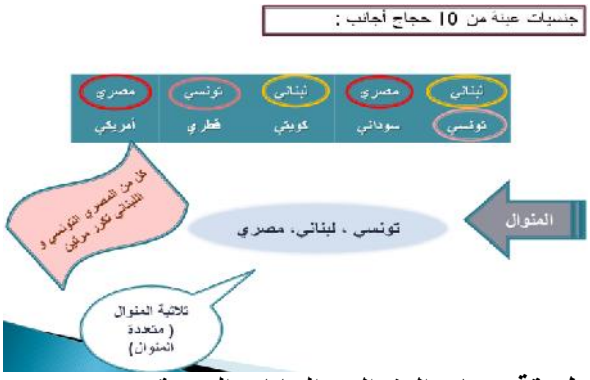
المنوال Mode : هو القيمة التي تكررت أكثر من غيرها. القيمة الأكثر شيوعاً أو تكراراً. يرمز للمنوال بالرمز D

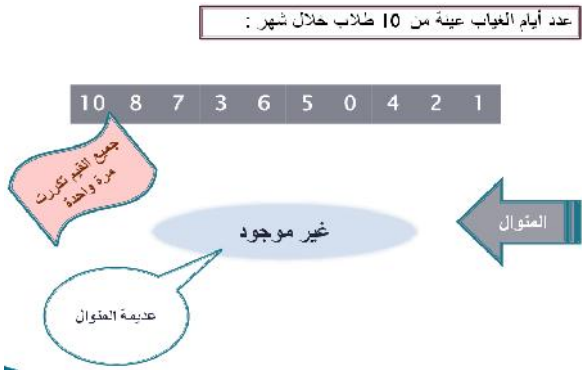
طريقة حساب المنوال: (البيانات غير المبوبة)

أمثلة:









طريقة حساب المنوال: (البيانات المبوبة)

البيانات الوصفية أو الكمية المنفصلة: يحسب المنوال من التعريف مباشرة أي القيمة التي يقابلها أكبر تكرار.

أمثلة:

مزايا المنوال:	عيوب المنوال:
سهولة حسابه وإيجاده. لا يتأثر بالقيم الشاذة. يمكن حسابه في الجداول التكرارية المفتوحة. يمكن إيجاده لجميع أنواع البيانات. الوصفية والكمية. يمكن إيجاده من خلال **الرسم**. يعتبر المقياس الوحيد الذي يمكن. استخدامه للبيانات الاسمية.	**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.