مقياس: المعالجة الآلية وتحليل المعطيات

يهدف هذا المقياس إلى تذكير الطلبة بالمفاهيم الأساسية في علم الإحصاء وطرق تمثيل وعرض ووصف البيانات مع تبيان الهدف من علم الإحصاء باستخدام الإحصاء الاستدلالي .

مقدمة:

تم اقتباس المسمى الإنجليزي لعلم الإحصاء (Statistics) من اللفظ اللاتيني (Status) أي بمعنى الدولة أي كل ما يخص الوصف الرقمي للأوضاع الاقتصادية والسكانية والاجتماعية للدولة. وقد تطور علم الإحصاء ليدخل في معظم مجالات المعرفة سواء كانت طبيعية أو علمية أو إنسانية، في عام 1662 قام العالم " جون جرونت " بنشر معلومات إحصائية حول المواليد والوفيات ، ثم تلي عمل جرونت بكثير من الدراسات حول الوفيات ونسب ومعدلات الأمراض وأحجام السكان والدخول ونسب البطالة .

1. تعريف علم الإحصاء

علم الإحصاء هو ذلك الفرع من المعرفة المكون من مجموعة من الأساليب الطرق الرياضية التي تستخدم لجمع وتبويب وتنظيم وعرض وتلخيص وتحليل المعلومات والبيانات وفق طرق وأساليب علمية للحصول على استدلالات واستنتاجات تستخدم لاتخاذ القرارات "السليمة". وينقسم علم الاحصاء الى قسمين رئيسين هما:

الإحصاء الوصفي: الإحصاء الوصفي هو الفرع المتعلق بطرق جمع وتبويب وتنظيم وعرض وتلخيص المعلومات والبيانات ووصف توزيع البيانات وذلك باستخدام جداول تكرارية أو رسوم بيانية وكذلك إيجاد بعض المقاييس العددية أو الوصفية التي تصف توزيع البيانات.

الإحصاء الإستدلالي: مجموعة الطرق والأساليب التي تستخدم في تعميم **نتائج العينة** على **المجتمع** التي سحبت منه.

2. مصطلح المجتمع والعينة والبيانات:

المجتمع: المجتمع هو المجموعة الكلية من الأشياء ذات خصائص مشتركة ويكون الباحث مهتمًا بعمل بعض الاستدلالات والنتائج حولها.

العينة: العينة هي مجموعة مكونة من عدد من أفراد المجتمع يتم اختيارهم بطريقة مناسبة بحيث تمثل المجتمع تمثيلاً جيدًا وذلك لدراسة خصائص المجتمع.

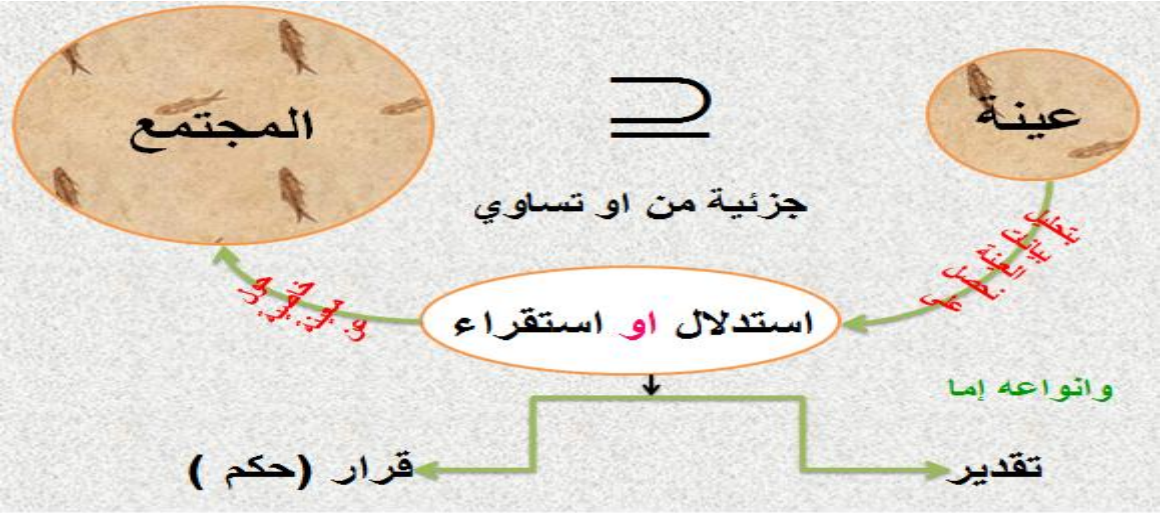
البيانات: البيانات هي مجموعة المشاهدات المأخوذة من الدراسة الإحصائية. أما **مصادر البيانات فهي:**

مصادر تجريبية مثل: التجارب المخبرية

مصادر ميدانية مثل: المسوحات الميدانية

مصادر تاريخية مثل: السجلات المحفوظة

3. الهدف من علم الإحصاء الاستدلالي : هو صناعة الاستدلال او الاستقراء حول المجتمع مستخدماً البيانات الموجودة داخل العينة المأخوذة من هذا المجتمع.



فالهدف من علم الإحصاء هو عملية الاستدلال حول خاصية معينة عن المجتمع وهذه العملية تعتمد على ثلاثة عناصر سنسميها بعناصر المشكلة الإحصائية (خطوات حل المشكلة الإحصائية)

4. طبيعة البيانات (القياسات): وتصنف إلى قسمين:

1.4 قياسات (بيانات) نوعية (طبقية– وصفية): هى التي يمكن تقسيمها أو فصلها إلى طبقات أو أنواع من القياسات يمكن التمييز بينها ببعض الخصائص غير العددية مثل: مستوى الدراسة، الجنس (ذكر – انثى)، نوع المنتج من شركة ما للإلكترونيات (راديو– تلفزيون –....)

2.4 قياسات (بيانات) كمية: تحتوى على مجموعة من الأرقام تمثل مقدار القياس أو الخاصية لشيْ معين مقارنة بمجموعته مثل: عدد أجهزة الراديو التي تنتجها شركة ما للإلكترونيات في أشهر مختلفة. وتصنف القياسات الكمية إلى:

1.2.4 قياسات كمية منفصلة : هى التي تأخذ قيما محددة مثل: عدد المواد الدراسية في الفرقة الأولى، عدد أفراد الأسرة، ترتيب أخ بين أخوانه وهكذا .....

2.2.4 قياسات كمية متصلة : هى التي تأخذ أي قيم داخل مدى معين أو داخل فترة معينة مثل: العمر، الطول، الوزن، درجة الحرارة..

وهناك طريقة أخرى لتصنيف القياسات وهى استخدام المستويات الأربعة للقياس:

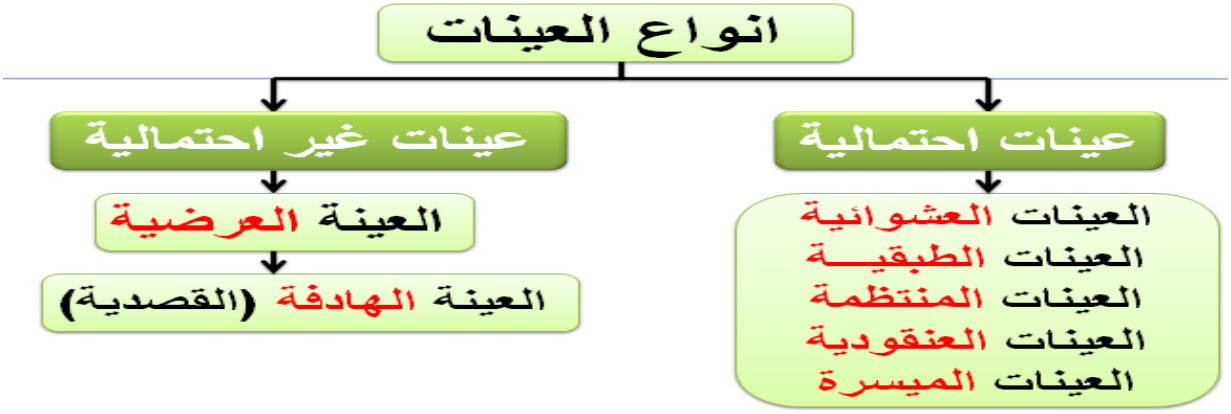
-**المستوى الاسمي (التصنيفي) للقياس:** يتميز هذا النوع بالقياسات التي تحتوى على الأسماء، العناوين، أو الأصناف فقط. وفي هذا المستوى لا يمكن ترتيب القياسات بأي طريقة.

-**المستوى الرتبي (الترتيبي – التفضيلي) للقياس:** يتميز هذا النوع بأنه يحتوى على القياسات التي يمكن إجراء عمليات الترتيب عليها، ولكن الفروق بين القياسات (الرتب) إما أن تكون لا يمكن تعيينها (تحديد قيمتها) أو تكون لا معنى لها.

-**المستوى الفئوي (الفترى – الفترة) للقياس:** يشبه هذا المستوى إلى حد كبير المستوى الرتبى مع تمييزه بخاصية إضافية وهى إمكانية تحديد الفروق بين القياسات ومعرفة دلالتها ولا يوجد لهذا المقياس صفرا أو (نقطة بداية) محددة بل تكون دائما اختيارية أو افتراضية.

-**المستوى النسبي (النسبة) للقياس:** يعتبر هذا المستوى تطويرا للمستوى الفئوي حيث إنه يحتوى على نقطة بداية طبيعية (الصفر المطلق) والذي يعنى غياب الخاصية، كما أن الفروق والنسب بين القياسات في هذا المستوى لها دلالة ومعنى.

5- **أنواع العينات:** وتنقسم إلى نوعين عينات احتمالية وعينات غير احتمالية.



1.1.5 العينات العشوائية: وفيها يتساوى جميع أفراد المجتمع في فرصة اختيارهم داخل العينة. تعتبر العينة العشوائية عينة ممثلة لمجتمع الدراسة في جميع خصائصه ونادرا ما يحدث اختلاف بين (إحصائيات المجتمع) و (إحصائيات العينة)

2.1.5 العينات الطبقية: ويتم في هذا النوع تقسيم المجتمع إلى مجتمعين جزئيين (طبقيين) على الأقل بشرط اشتراكهم فى نفس الخاصية، ثم يتم اخذ العينة من كل مجتمع جزئي (طبقة).

3.1.5 العينات المنتظمة: يتم في هذا النوع اختيار نقطة بداية ثم اختيار عناصر العينة بشكل دوري بداية من هذه النقطة كأن نختار كل خامس طالب إذا كان طول الدورة خمسة أو ثامن كتاب إذا كان طول الدورة ثمانية، وهكذا.

4.1.5 العينات العنقودية: يتم في هذا النوع تقسيم المجتمع إلى قطاعات أو أقسام تسمى عناقيد (ليست ذات خصائص مشتركة كما في حالة العينة الطبقية) ثم يتم الاختيار عشوائيا لعدد من هذه العناقيد ثم اختيار كل أعضاء هذه العناقيد في العينة.

5.1.5 العينات الميسرة: وفيها يتم استخدام النتائج المتاحة من قبل بسهولة و بسرعة. في بعض الحالات تكون العينات الجاهزة جيدة وفى بعض الحالات تحتوى على انحرافات (لصغر حجم العينة).

أما العينات غير الإحتمالية وهى العينات التي يتم أخذها تحت شروط ومواصفات أو معايير يراها الباحث لتحقيق غرض معين في التجربة. وبذلك فإن هذا النوع لا يتبع نظرية الاحتمالات في الاختيار. وتنقسم إلى:

1.2.5 العينة العرضية: ويتم اخذ مثل هذه العينات عن طريق الصدفة . ولا يمكن تعميم نتائجها على مجتمع الدراسة.

2.2.5 العينة الهادفة (القصدية): وفى هذا النوع يقوم الباحث باختيار العينة تحت شروط معينة لتحقيق الهدف أو الغرض من التجربة أو الدراسة.

6. طرق وصف القياسات: لكى نقوم بعمل جملة استدلالية (إستدلال إحصائي) فإننا نحتاج إلى طريقة لوصف هذا المجتمع، من خلال الطرق البيانية (الرسوم) والطرق العددية (الحسابية)

1.6 الطرق البيانية: تستخدم الطريقة البيانية لعرض مجموعة القياسات في صورة أشكال أو رسومات بيانية تعطى الدارس أو الباحث وصفا مرئيا كافيا عن هذه المجموعة من القياسات. ويستخدم الجدول

التكراري لسرد أو تلخيص القياسات داخل فترات (أو فصول) تبعا لعدد مرات (أو تكرار) وقوع البيان داخل كل فترة (أو فصل). والأشكال المختلفة للجدول التكراري نعرضها فيما يلي:

(شكل 1 – الجدول التكراري)

رقم الفترة (الفصل) i	حدود الفترة	التكرارات المطلقة	
		علامات التفريغ	التكرار f_i
1	الحد الادني - الحد الاعلى	|	f_1 =1
2	الحد الادني - الحد الاعلى	||	f_2 =2
3	الحد الادني - الحد الاعلى	|||	f_3 =3
.....	.....	.....	.....
المجموع	.....	.....	$\sum_{i=1}^{k} f_i = n$

f_i تكرار الفترة (الفصل) رقم i أي ان f_1 = تكرار الفترة (الفصل) رقم 1 وهكذا.....

(شكل 2 – الجدول التكراري النسبي)

رقم الفترة (الفصل)	حدود الفترة	التكرارات النسبية $\tilde{f}_i = \frac{f_i}{n}$
1	الحد الادني - الحد الاعلى	التكرار المطلق للفترة / العدد الكلي للبيانات (n)
2	الحد الادني - الحد الاعلى	التكرار المطلق للفترة / العدد الكلي للبيانات (n)
3	الحد الادني - الحد الاعلى	التكرار المطلق للفترة / العدد الكلي للبيانات (n)
.....	.....	.....
.....	.....	.....
المجموع	.....	$\sum_{i=1}^{k} \frac{f_i}{n} = 1$

(شكل 3 – الجدول التكراري التراكمي (التجميعي))

رقم الفترة (الفصل) i	حدود الفترة	التكرارات المطلقة	التكرارات التراكمية (التجميعية)
1	الحد الادني - الحد الاعلى	1	1
2	الحد الادني - الحد الاعلى	2	3
3	الحد الادني - الحد الاعلى	3	6
.....	.....	.....	.....
k	.....	.....	n

وبالمثل يمكن الحصول علي التكرار التراكمي النسبي كما هو موضح بالجدول التالي (شكل 4)

(شكل 4 – الجدول التكراري التراكمي النسبي)

رقم الفترة (الفصل) i	حدود الفترة	التكرارات المطلقة	التكرارات التراكمية	التكرارات التراكمية (النسبية)
1	الحد الادني - الحد الاعلى	1	1	التكرار التراكمي للفترة / العدد الكلي للبيانات(n)
2	الحد الادني - الحد الاعلى	2	3	التكرار التراكمي للفترة / العدد الكلي للبيانات(n)
3	الحد الادني - الحد الاعلى	3	6	التكرار التراكمي للفترة / العدد الكلي للبيانات(n)
.....	.....	.....	.....	.....
.....	.....	.....	.....	.....
k	.....	.....	.....	1

مثال: في اختبار لقياس درجة الذكاء يحتوى على 30 فقرة رصدت درجات 25 طالبا على النحو التالي:

(لاحظ ان حجم البيانات = 25 (عدد الطلاب)) **الإجابة** من خلال الجدول التكراري الكامل:

رقم الفترة (الفصل) i	حدود الفترة	التكرارات المطلقة f_i	التكرارات النسبية $\tilde{f}_i$	التكرارات التراكمية	التكرارات التراكمية النسبية
1	16.5– 14.5	1	25/1	1	25/1
2	18.5– 16.5	2	25/2	3	25/3
3	20.5– 18.5	3	25/3	6	25/6
4	22.5– 20.5	3	25/3	9	25/9
5	24.5– 22.5	7	25/7	16	25/16
6	26.5– 24.5	4	25/4	20	25/20
7	28.5– 26.5	3	25/3	23	25/23
8	30.5– 28.5	2	25/2	25	1 = 25/25
المجموع	------	25	1	------	------

2.6 الطرق العددية لوصف البيانات: من أهم مميزات الطريقة البيانية لوصف القياسات هي عملية التمثيل المرئي للبيانات. في كثير من الأوقات نحتاج إلى تقرير كتابي عن البيانات، وفى هذه الحالة لا يمكن استخدام الطريقة البيانية للحصول على هذا النوع من التقارير.

المعدل (المتوسط)	تعريفه	مدى استخدامه	إيجاده	تأثره بالقيم المتطرفة	تأثره بجميع القيم	مميزاته وعيوبه
الوسط الحسابي	$\overline{x} = \frac{\sum x}{n}$	الأكثر استخداما	دائما	نعم	نعم	يعمل بكفاءة مع جميع الطرق الإحصائية
الوسيط	القيمة التي تتوسط القيم	غالبا	دائما	لا	لا	غالبا ما يستخدم في حالة وجود قيم متطرفة
المنوال	القيم الأكثر تكرارا	أحيانا	أحيانا لا يوجد وأحيانا أكثر من واحد	لا	لا	صالح للبيانات من المستوى الاسمي
نصف المدى	$\frac{\text{الأكبر} + \text{الأصغر}}{2}$	نادرا	دائما	نعم	لا	يتأثر بالقيم المتطرفة