الأستاذة : عظيمي مسعودة

البريد الالكتروني: lamia072009@yahoo.fr

المحور الأول: الطريقة الإحصائية في البحث التربوي

# المحاضـــــــــــــــرة الثالثة: التحليل الإحصائي للبيانات:

# I- عرض البيانات وتلخيصها.

بعدما ينتهي الباحث من تحديد المصادر التي سيعتمد عليها في الحصول على بياناته، وتحديد أساليب وتقنيات جمعها، يأتي إلى مرحلة أخرى تعتبر ذات أهمية في تسهيل عملية قراءة هذه البيانات والتعامل معها، وهي عملية عرض وتلخيص البيانات، ذلك أن هذه الأخيرة تكون عبارة عن مجموعة من القيم المجمعة، لذا فإن تنظيمها وتلخيصها لاستنتاج مميزاتها، يشكل خطوة أساسية لتحليلها واتخاذ قرارات بشأنها، ويعتمد الباحث في ذلك على ما يعرف بأساليب الإحصاء الوصفي الذي سبق وأن أشرنا إلى أن وظيفته الأساسية في البحوث العلمية تتمثل في: عرض البيانات جدوليا وبيانيا وتلخيصها باستخدام مختلف أساليب (مقاييس) الإحصاء الوصفي، من مقاييس النزعة المركزية، مقاييس التشتت،...، ويمكننا تقديم عرض موجز لأهم هذه الأساليب، على أساس أن الطالب قد سبق له التعامل معها في المقياس الإحصاء الوصفي في السنة الأولى من دراسته الجامعية.

## I- طرق عرض البيانات وتلخيصها:

### I-1 العرض الجدولي للبيانات:

**الجدول التكراري:** عبارة عن جدول يتألف من عمودين، يضم العمود الأول البيانات أو العلامات الخام مرتبة تنازليا أو تصاعديا، في حين يضم العمود الثاني عدد مرات ظهور أو تكرار كل علامة، وإذا كان عدد القيم أو العلامات كبيرا فإنه من الأفضل تجميعها في فئات، مع الأخذ بعين الاعتبار بعض المفاهيم المتعلقة بمثل هكذا بيانات وهي: مدى الفئة $R$، وهو الفرق بين أعلى وأقل قيمة، سعة أو طول الفئة $L = R/N_C$ حيث $N_C$ هو عدد الفئات.

عدد الفئات هذه، يمكن إيجادها عن طريق تطبيق أحد القانونين الآتيين:

- قانون ستروج *Struges*: $N_C = 1+3.33Log_n$
- قانون بروكس *Brooks*: $N_C = 5Log_n$

### I-2 العرض (التمثيل) البياني:

**I-2-1 الدوائر البيانية:** أداة بسيطة من أدوات العرض، تستخدم عادة لتمثيل البيانات الكيفية، حيث يمثل كل نمط من أنماط المتغير الكيفي بزاوية قياسها يتناسب مع التكرار، ويتم تمييز كل نمط بلون مختلف عن غيره.

**I-2-2 المدرج التكراري ( الأعمدة التكرارية):** يستخدم المدرج التكراري والأعمدة التكرارية لتوضيح بيانات جدول تكراري معين ونستعمل عادة المحور الأفقي لتمثيل الفئات والرأسي لتمثيل التكرارات، ففي حالة ما إذا رسمنا المستطيلات متصلة فيما بينها فإن الشكل الناتج هو المدرج التكراري (يستخدم عادة مع المتغيرات المتصلة بعد تجميع القيم في فئات)، أما إذا رسمنا الأعمدة منفصلة فإن الشكل الناتج هو أعمدة تكرارية (تستخدم في حالة البيانات الكيفية والكمية المنفصلة).

**I-2-3 المضلع التكراري:** يختلف المضلع التكراري عن المدرج التكراري في كون هذا الأخير يفترض أن التكرارات الخاصة بكل فئة موزعة بانتظام على جميع قيم هذه الفئة، أما المضلع التكراري فيفترض أن جميع قيم الفئة تمثلهم قيمة واحدة هي مركز الفئة، وعملية الإيصال بين النقاط المتتالية(مراكز الفئات) بمستقيمات، ينتج لنا المضلع.

**I-2-4 المنحنى التكراري:** يختلف عن المضلع التكراري في استخدام الخطوط المنحنية بدلا من الخطوط المستقيمة المنكسرة، ويستعمل عادة لإعطاء فكرة عامة عن التوزيع وذلك برسم منحنى يمر بأكبر عدد من النقط المعبرة عن التكرار الحقيقي للفئات، ويستخدم هذا التمثيل عادة، إذا كان المتغير المدروس متصلا.

كما تشمل عملية عرض البيانات وتلخيصها، استخدام بعض مقاييس الإحصاء الوصفي والمتمثلة أساسا في مقاييس النزعة المركزية ومقاييس التشتت.

**I-3 مقاييس النزعة المركزية:** سميت كذلك، على أساس أن الأرقام التي نتعامل معها لها نزعة نحو التمركز في وسط التوزيع، يمكن تعيينها باستخدام عدة أساليب رياضية، ويلجأ إليها الباحث لإعطاء تفاصيل أكثر عن خواص البيانات التي قام بجمعها، فهو يحتاج أن يعبر عن درجات العينة التي شملها البحث بقيمة واحدة تمثلها، تعد بمثابة القيمة المعيارية أو المركزية التي يمكن أن تستخدم من أجل التوضيح والمقارنة، فإذا نظرنا إلى مجموعة من الدرجات لعينة ما في اختبار أدائي موضحة في جدول، فإننا سوف نبحث عن شيء عام يربط هذه الأرقام معا، أي أننا نبحث عن مركزة هذه الأرقم جميعا في رقم متوسط، وأبسط خطوات هذا البحث هي حساب الوسط الحسابي Mean، أو الوسيط Median، أو حساب المنوال Mode، ويمكن عرضها بإيجاز كالتالي:

**I-3-1 المتوسط الحسابي *Mean*:** المتوسط الحسابي لقيم متغير ما هو مجموع قيم ذلك المتغير مقسوما على عدد هذه القيم (المشاهدات)، ويمثل هذا القياس الأكثر هرة وأهمية، حيث يمثل القيمة التي تتمركز حولها قيم المتغير الكمي، يرمز له بالرمز $\bar{x}$ ويعطى بالعلاقة التالية: $\bar{x} = \frac{\sum x_i}{n}$

**I-3-2 الوسيط *Median*:** ثاني أهم مقياس للنزعة المركزية، يقسم الوسيط توزيع مجموعة من الدرجات إلى قسمين متساويين، بحيث يقع 50٪ من القيم أدنى الوسيط، و50٪ أعلاه، وهذا بعد ترتيب القيم تصاعديا أو تنازليا. إذا كان عدد البيانات فرديا فإن قيمة الوسيط هي القيمة التي تتوسط هذه البيانات، أما إذا كان عددها زوجيا، فإن قيمة الوسيط، هي الوسط الحسابي للقيمتين اللتين تتوسطان هذا التوزيع أي. ويمكن تعيين قيمة الوسيط من خلال حساب رتبته والتي تعطى بالعلاقة التالية:

$$Me = \frac{n+1}{2}$$

**I-3- 2 المنوال *Mode*:** هو القيمة الأكثر تكرارا لمجموعة من القيم، وقد يكون للقيم أكثر من منوال، وهو وسيلة بسيطة تدلنا بصورة أولية على مركز التجمع للقيم أو العلامات أو لفئات، ويعتبر أقل دقة مقارنة بالوسط الحسابي والوسيط.

**I-4 مقاييس التشتت:** تعرفنا على الهدف العام من استخدام مقاييس النزعة المركزية والمتمثل أساسا في محاولة التعرف على ملمح أو شيء مشترك بين أفراد العينة، بحيث تسهل معه عملية الوصف والمقارنة، غير أن هذه المقاييس لا تكفي لوحدها لمعرفة الصفات الإحصائية اللازمة لوصف الظواهر، لأن الفروق بين قيم الظواهر تزداد أو تنقص رغم تساوي المتوسطات لهذه الظواهر، لذلك وجب الاستعانة بمقاييس أخرى لقياس مدى تباعد أو تقارب البيانات من بعضها البعض أو من متوسطها الحسابي، تسمى

بمقاييس التشتت. إذن فماهية مقاييس التشتت تنبع من حقيقة أنه يمكن تساوي المتوسطات الحسابية أو الوسيط لعدة مجموعات، رغم كون هذه المجموعات مختلفة كثيرا وغير متجانسة، وعلية فمن الخطأ القول بأن هذه المجموعات متساوية لمجرد تساوي متوسطاتها، وهذا كله يفسر بوجود نزعة للأرقام نحو التشتت أو التباعد.

يقصد بالتشتت، مقدار أو درجة التفاوت: (التباعد)، (الاختلاف)، (التباين) بين مفردات هذه الظاهرة، وتعتبر بيانات الظاهرة متجانسة عندما تكون قيمها قريبة من بعضها البعض، ونقول في هذه الحالة بأن تشتتها ضعيف (غير مشتتة)، أما إذا كانت بيانات الظاهرة متباعدة وغير متجانسة (الاختلاف بين القيم كبير)، فنقول أن مفردات الظاهرة مشتتة وغير مركزة، ويمكن الإشارة فيما يلي إلى بعض مقاييس التشتت الشائعة الاستخدام:

**I-4-1 المدى *Range*:** هو أبسط مقياس لقياس التشتت، والمدى لمجموعة من البيانات هو الفرق بين أكبر قيمة وأصغر قيمة: $R = X_{max} - X_{min}$ . يعتبر هذا المقياس قليل الدقة في التعبير عن مدى تشتت القيم عن متوسطها، وذلك لاعتماده على النقطتين المتطرفتين في المجموعة التي ينتميان إليها $X_{max}$ و $X_{min}$، فيبرز تشتتا قد يكون وهميا، مثلا لدينا القيم التالية:2، 10، 15، 18، 23، 29، 52، فالمدى هنا لا يعبر عن التشتت الحقيقي للقيم، وبالتالي فإن هذا المقياس لايصلح إلا إذا أراد الباحث أن يأخذ فكرة سريعة عن التشتت الموجود، دون الاهتمام بالدقة في المقياس.

**I-4-2 التباين *Variance*:**

هو أحد مقاييس التشتت، يمكن التفكير فيه كمقياس للمسافة، حيث تقاس المسافة ببعد القيمة عن المتوسط الحسابي، وكلما كانت قيمة التباين كبيرة، كان التوزيع أكثر تبعثرا وأقل تجانسا، أما عن صيغته الرياضية فهي عبارة عن المتوسط لمجموع مربعات الفروق بين قيم الإحصائي ومتوسطها الحسابي، ويرمز له بالرمز $S^2$ :

$$s^2 = \frac{\sum(x - \bar{x})^2}{n}$$

أو نستخدم الصيغة المختصرة التالية:

$$s^2 = \frac{\sum(x)^2 - \frac{(\sum x)^2}{n}}{n}$$

**I-4-3 الانحراف المعياري *Standard Deviation*:** هو الجذر التربيعي للانحراف المعياري، يفضل استخدامه لأن وحدة القياس فيه مساوية لوحدة القياس للبيانات الأصلية، وبالتالي يمكن التفكير فيه كمتوسط للمسافات بين القيم والمتوسط الحسابي، يرمز له بالرمز $S_d$:

$$S_d = \sqrt{\frac{\sum(x - \bar{x})^2}{n}}$$