AL MUSTAQBAL UNIVERSITY

جامعة المستقبل

1/1/2024

وزارة التعليم العالي والبحث العلمي

جامعة المستقبل

كلية العلوم الإدارية

قسم إدارة الاعمال

مادة مبادئ الاحصاء

منهج مادة مبادئ الاحصاء

المرحلة الاولى

تدريسي المادة

المدرس الدكتور

رياض مالك محسن

مبادئ علم الاحصاء

المصدر الاساسي :مبادئ الاحصاء (تأليف: د. محمد حسن ، أمير حنا هرمز) ومصادر اخرى

الفصل الاول

المحاضرة الاولى

مقدمة عن علم الاحصاء

نشوء وتطور علم الاحصاء :-

يعتبر علم الاحصاء علم قديم كقدم المجتمع البشري نفسه ، حيث ان اصل الاحصاء يمكن ان ينسب الى الازمنة السابقة عندما كان يعامل كنظام للعد والترقيم لا نشطة الدولة المختلفة كما هو مدون في مسلة حمورابي واثار الحضارات البابلية والاشورية واليونانية ووادي النيل ، فمظاهر البيع والشراء وحساب الجند والحيوانات تعتبر كلها مظاهر لعمليات احصائية برغم بساطتها.

بعد ظهور الاسلام وانتشاره حتى الصين وصل الصين شرقا والمحيط الاطلسي وجنوب واوربا غربا احتك العرب بأقوام تلك البلدان واطلعوا على ما لديهم من المعرفة وذلك عن طريق الترجمة والدراسة كما قام العرب بإنجاز اضافات هامة في مختلف المجالات كانت حصيلتها ازدهار علمي اثار الاعجاب والدهشة .

ان ارتباط الاحصاء بأسس ومفاهيم رياضية يقودنا للقول بان التطور الذي حصل في الرياضيات ساهم كثيرا في تطور الاعمال الاحصائية حيث ان غالبية المعالجات الاحصائية تستند على بيانات كمية (ارقام) ، لذلك فلابد من ذكر ماساهم به العرب فيما يتعلق الامر بالأرقام ، لقد انظمة الترقيم كثيرة ومتنوعة باختلاف الامم والحضارات ، ففي مصر كانت انظمة الترقيم بالحروف القبطية وفي سوريا كانت تستخدم الارقام اليونانية المستندة على حساب الجمل وان عددها كان بعدد حروف الهجاء اما الارقام اليونانية فتحتاج الى اشكال عددية للدلالة على بعض الاعداد وقد طور العرب نظاما جديدا يكتب بأشكال معينة وقد استخدموا نوعين من هذه الاشكال احدهما هو (1,2,3,4,5,6,7,8,9) ولأخر هو (١,٢, ٣ ٤,٥, ٦ ، ٧، ٨ ، ٩) الشكل الاول اخذ عما كان متداول في وادي الرافدين في حضارة بابل حيث كان هذا النظام يعتمد الزوايا في تحديد قيمة الارقام واشكالها فعلى سبيل المثال كان الرقم واحد يتضمن زاوية واحدة ويكتب (1) وان الرقم اثنان يتضمن زاويتان ويكتب (2) وهكذا وقد انتشر استخدام هذه الارقام في بلاد المغرب والاندلس وعن طريقهم دخل نظام الترقيم هذا الى واربا وعرف فيها باسم الارقام العربية التي استخدمت في واربا في اواخر القرن السادس عشر للميلاد ويقال ان اخذ واربا بالأرقام العربية قد اسرع في نهضتها وازدهارها .

اما نظام الترقيم الثاني الذي طوره العرب فقد اخذ من الهنود عندما اطلع العرب على ما لدى الهنود من الرياضيات وهو النظام الذي تستعمله غالبية البلدان العربية والاسلامية ولم يقتصر العرب على تطوير وتهذيب هاتين السلسلتين بل وجدوا نظاما اكثر تطورا وهو النظام الصفري . ان لهاتين السلسلتين مزايا كثيرة منها :

١- اقتصارها على عشرة اشكال بما فيها الصفر .
٢- امكانية تركيب اي عدد من هذه الاشكال العشرة مهما كان كبيرا بخلاف الارقام الرومانية التي تحتاج الى اشكال جديدة للدلالة على بعض الارقام .
٣- تقوم على اساس النظام العشري وعلى اساس القيم الوصفية فللرقم قيمتان هما قيمة شكله وقيمته بالنسبة الى المنزلة التي يقع فيها .

٤- ادخال الصفر في الترقيم واستعماله في المنازل الخالية من الارقام ويعتبر هذا التطور احد الاختراعات الاساسية ذات الفوائد الرائعة التي توصل اليها العقل البشري فلم تنحصر مزاياه في تسهيل عملية الترقيم وحده بل تعدته الى تسير جميع العمليات الحسابية بما فيها الضرب والقسمة التي قد تحتاج الى طرق معقدة لأجرائها في غير النظام العربي .

ويبدو ان كلمة الاحصاء (Statistics) مشتقة من الكلمة اللاتينية (Status) أو من الكلمة الايطالية (Statista) أو من الكلمة الالمانية (Statistik) وجميعها تعني فيما تعنيه حقائق ومعلومات عن الدولة (Political State) حيث أستخدم هذا المفهوم لجميع المعلومات الخاصة بأفراد المجتمع لأغراض تكوين فكرة عن قوة العمل حينذاك وتكوين قاعدة معلومات من خلالها يمكن للدول فرض ضرائب لتعزيز وضعها المالي .

اما في العصور الحديثة فان اول ملامح تطور علم الاحصاء كان خلال القرن السابع عشر الذي شهد ولادة الاحصاء الحيوي (Vital Statistics) من قبل الانكليزي (Jhon Grant) (١٦٢٠-١٦٧٤) الذي كان اول من درس احصاء الولادات والوفيات وحساب توقعات البقاء عند اعمار مختلفة ونشر نتائج ابحاثه عام ١٦٦٢ في كتاب ، وحول نفـ س الموضوع قام صديق له يدعى (William Petty) عام ١٦٨٧ بمحاولة لتقدير عدد نفوس لندن ومفترضا نسبة وفاة جارية قدرها ١ الى ٣٠ وقد استنتج ان عدد سكان لندن عام ١٦٨٥ هو حوالي ٦٩٦ الف نسمة وهكذا استكملت الدراسات بشكل اوسع ..... الخ .

تعريف علم الاحصاء :-

هنالك تعاريف عديدة لعلم الاحصاء اختلفت وتباينت من حيث المضمون والشمول باختلاف مراحل تطور هذا العلم والفوائد المتوخاة منه وبشكل عام يلاحظ ان هذه التعاريف يمكن اجمالها في نوعين رئيسيين الاول منها اعتبر الاحصاء بأنه جمع لبيانات احصائية اي جمع لجمل عددية للحقائق والظواهر في حين ان الثاني اعتبر الاحصاء بانه جمع لطرق احصائية اي جمع متكامل لمبادئ واساليب تستخدم في تجميع وتحليل البيانات والمعلومات الاحصائية وفيما يلي بعض هذه التعاريف :-

تعريف Webster للإحصاء :

الاحصاء عبارة عن حقائق مصنفة تمثل معلومات عن الفرد في الدولة وخصوصا تلك الحقائق التي يمكن وصفها بأعداد او اية وسيلة اخرى للتبويب او التصنيف .

تعريف Boddington للإحصاء :

الاحصاء هو علم التقديرات والاحتمالات .

تعريف King للإحصاء :

الاحصاء هو الطريقة التي تختص بجمع المعلومات عن الظواهر الطبيعية او الاجتماعية من النتائج المتحصل عليها من تحليل او تعداد او تجميع التقديرات .

مما تقدم يمكن تعريف علم الاحصاء على النحو الاتي :

علم الاحصاء : هو الطريقة العلمية التي تختص بجمع البيانات والحقائق عن ظاهر او فرضية معينة (ظواهر او فرضيات) معينة وتنظيم وتبويب هذه البيانات والحقائق بالشكل الذي يسهل عملية تحليلها وتفسيرها ومن ثم استخلاص النتائج واتخاذ القرار على ضوء ذلك .

وبشكل عام فان علم الاحصاء وبسبب تطوره السريع وكثرة فروعه التطبيقية في مجالات الحياة كافة فإن معظم الإحصائيين يميلون للنظر الى هذا العلم على انه جمع لفرعين رئيسين هما :

أ- الاحصاء الوصفي Descriptive Statistics .

ويتضمن هذا الفرع الطرق والاساليب المستخدمة في جمع البيانات والمعلومات عن ظاهرة معينة او مجموعة ظواهر وكيفية تنظيم وتصنيف وتبويب هذه البيانات والمعلومات مع امكانية عرضها في جداول ورسومات بيانية وحساب بعض المؤشرات الاحصائية منها والتي سيأتي ذكرها في الفصول اللاحقة .

ب- الاحصاء الاستدلالي Inferential Statistics .

وهو الشطر الثاني من علم الاحصاء الذي يهتم عادة بموضوعي التخمين Estimation واختبار الفرضيات Testing Hypotheses .

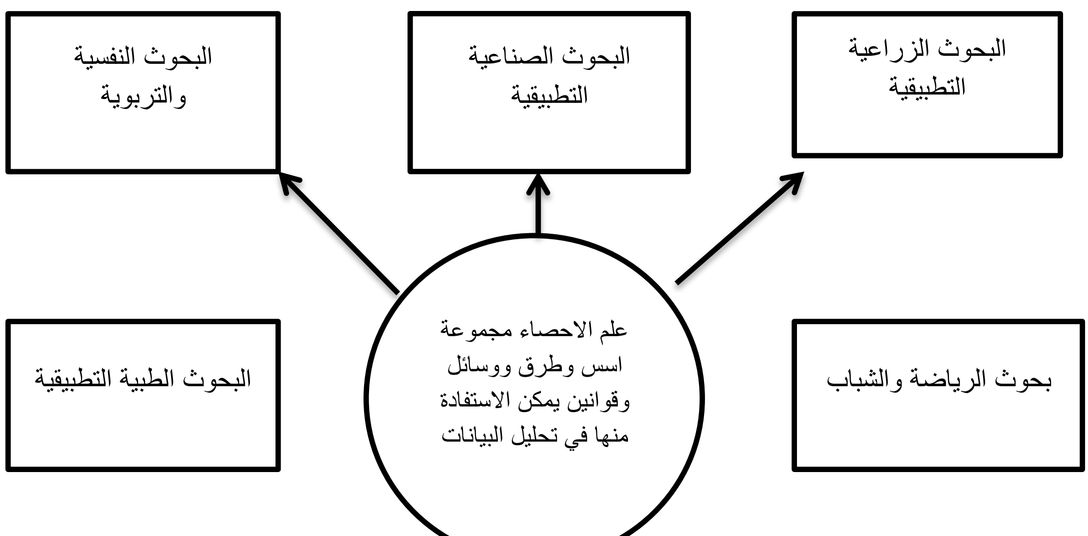
الفصل الاول

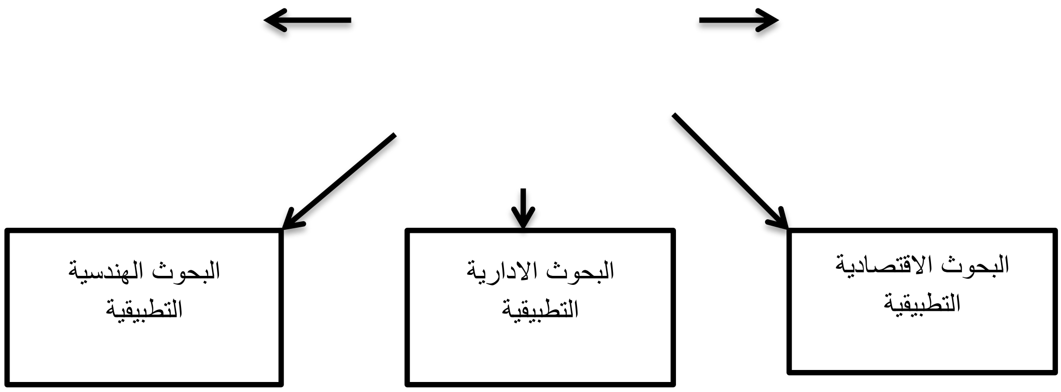
المحاضرة الثانية

مقدمة عن علم الاحصاء

أهمية علم الاحصاء ومجالات تطبيقاته :

يعتبر علم الاحصاء احد الوسائل المهمة في البحث العلمي من خلال استخدام قواعده وقوانينه وطرقه في جمع البيانات والمعلومات اللازمة للبحث العلمي وتحليل البيانات والمعلومات للوصول الى النتائج التي يهدف لها البحث كما وان للإحصاء دورا بارزا في وضع الخطط المستقبلية عن طريق التنبؤ بالنتائج ولكافة القطاعات سواء كانت انتاجية ام خدمية وحيث ان الاحصاء بحد ذاته يعتبر وسيلة وليس غاية فذلك يعني امكانية استخدامه أينما وجد البحث العلمي ، وذلك يعني ان مجالات تطبيق علم الاحصاء ممكنة سواء كان ذلك في مجال العلوم الصرفة او العلوم الانسانية وغيرها ويمكن تمثيل مجالات تطبيق علم الاحصاء حسب ما هو موضح بالشكل التالي :





شكل (١-١) مجالات تطبيق علم الاحصاء

الطريقة الاحصائية في البحث العلمي:

ان استخدام الاسلوب الاحصائي في البحث العلمي يعني توفير بيانات ومعلومات عن الظاهرة او الظواهر المطلوب دراستها في ذلك البحث وهذا يعني ان امكانية تطبيق الطريقة الاحصائية مرهون بإمكانية التعبير عن هذه الظاهرة او تلك تعبيرا كميا ، فمثلا كان من الصعوبة استخدام الطريقة الاحصائية في البحوث الاقتصادية بسبب العجز في أمكانية قياس الكثير من الظواهر الاقتصادية كميا وعليه فان البحث الاقتصادي حينذاك كان يعتمد اسلوب التحليل الاستنباطي والتحليل الوصفي الا انه وبمرور الزمن وتقدم القياس لبعض الظواهر اصبح بالإمكان استخدام الطريقة الاحصائية في تحليل تلك الظواهر ويلاحظ ان الكثير من البحوث التطبيقية وفي العديد من مجالات المعرفة قد استخدمت الطريقة الاحصائية وذلك بسبب ان هذه الطريقة تمتاز بكونها تهيئ اسلوبا علميا موضوعيا محايدا دون ان يكون للباحث اي تدخل او انحياز تجاه نتائج البحث هذه الطريقة جعلت الطريقة الاحصائية تلاقي اهتمام الكثير الباحثين في مجالات المعرفة المختلفة .

اهم مراحل الطريقة الاحصائية في البحث العلمي :

١- تحديد مشكلة او فرضية البحث او الدراسة .
٢- جمع البيانات والمعلومات عن الظاهرة او الظواهر ذات العلاقة بالبحث او الدراسة .
٣- تصنيف البيانات وتبويبها وعرضها .
٤- حساب المؤشرات الاحصائية كتقديرات لمعالم مجتمع البحث او الدراسة .
٥- تحليل معطيات الدراسة والتوصل للنتائج على ضوء فرضية او فرضيات البحث او الدراسة .
٦- تفسير النتائج وعملية اتخاذ القرار بشأن فرضيات البحث .

اسلوب تصميم البحوث :-

هنالك اعتبارات كثيرة يتوقف عليها تصميم البحث ففي كل تصميم يتوجب على الباحث الاخذ بنظر الاعتبار مسالة الحصول على البيانات والمعلومات بأقصر وقت واقل جهد واوطأ كلفة وهذا يعني انه يجب مراعاة ما يلي عند تصميم البحث : -

ا- تحديد الغرض من البحث .

يجب ان يكون الهدف من البحث محدد بشكل واضح ودقيق بحيث يمكن التعرف على اوجه الاستفادة من نتائجه . فمثلا لوكان البحث منصب في دراسة نمط الاستهلاك من اللحوم الحمراء فذلك وبالتأكيد يعني ان الهدف من هذا البحث هو التعرف وعلى نحو مركز نمط استهلاك الفرد من هذه السلعة .

ت- تحديد امكانية التنفيذ الفعلي للبحث : -

من الضروري جدا تحديد المتطلبات التي تستلزمها عملية تنفيذ البحث وبشكل واضح ودقيق كالموارد المالية المطلوبة عند التنفيذ والامكانيات البشرية المتاحة المطلوبة في تحقيق بعض فقرات البحث كذلك التأكد من مدى توفر البيانات والمعلومات الدقيقة عن مشكلة البحث .

ث- تحديد اطار البحث Frame .

ان احد الامور قبل البدء بتنفيذ البحث هو تحديد نوع وطبيعة مجال ذلك البحث اي ما نعنيه تحديد المجتمع الاحصائي على نحو واضح ودقيق والمجتمع الاحصائي عبارة عن مجموعة من الوحدات او المفردات التي تشترك بصفة او صفات معينة والتي غالبا ما يتم الحصول منها على البيانات والمعلومات المطلوبة .

فمثلا لو كان البحث هو التعرف على نسبة عدد الافراد المصابين بأحد الامراض المتوطنة في قرية معينة فأن المجتمع الاحصائي هنا هو كافة الافراد المقيمين في هذه القرية والمفردة الاحصائية هنا تمثل الفرد المقيم في هذه القرية الممكن إخضاعه لعملية الفحص الطبي .

اما أذا كان البحث منصب على دراسة انتاجية الدونم الواحد من الحنطة في المحافظات الشمالية من العراق فأن المجتمع الاحصائي هنا هو كافة المساحات المخصصة لزراعة الحنطة في المحافظات الشمالية من العراق وان المفردة الاحصائية هنا قد تكون العائلة الفلاحية المتخصصة في زراعة الحنطة .

والمجتمع الاحصائي قد يكون مجتمع محدد او غير محدد ، فالمجتمع الاحصائي المحدد هو ذلك المجتمع الذي يمكن مواجهة او ملاحظة كل مفردة من مفرداته اي ما نعنيه انه يمكن حصر كافة المفردات التي تنتمي له مثال على ذلك طلبة جامعة بغداد ، سكان بلد معين ، المرضى الراقدين في مستشفى معين ، أنتاج مصنع للمصابيح الكهربائية خلال يوم محدد ، اما المجتمع الاحصائي غير المحدد فهو ذلك المجتمع الذي لا يمكن مواجهة او ملاحظة كل مفرداته اي لا يمكن حصر كافة مفرداته مثال على ذلك كريات الدم الحمراء في جسم الانسان ، الاسماك في بحيرة دوكان .

وعموما يمكن القول بان تحديد المجتمع الاحصائي والمفردة الاحصائية امر مهم ويتوقف ذلك على الهدف او الاهداف المتوخاة من البحث.

ج- تحديد اسلوب جمع البيانات والمعلومات .

بغية الحصول على البيانات والمعلومات التي يتطلبها البحث او الدراسة هنالك اسلوبان يمكن من خلالها جمع البيانات والمعلومات كل منها له ميزاته وعيوبه هذان الاسلوبان هما اسلوب التسجيل الشامل عن كافة مفردات المجتمع الاحصائي واسلوب التسجيل عن طريق العينات .

الفصل الثاني
مقاييس النزعة المركزية
Central Tendency Measurement

مقاييس النزعة المركزية
Central Tendency Measurement

بعد أن استعرضنا في محاضراتنا السابقة أساليب جمع وتصنيف وتبويب البيانات وكيفية تمثيلها في جداول ورسوم هندسية وبيانية الهدف منها إعطاء صورة واضحة وسريعة تبين ماهية هذه البيانات . أما ألان فسوف نتطرق الى كيفية تمثيل مجموعة من البيانات بقيمة واحدة فقط من خلال مقياس يدعى مقياس النزعة مركزية أو مقياس متوسط هذه القيمة وهذا العدد يميل لان يقع في وسط تلك المجموعة من البيانات في حال ترتيبها حسب صغرها أو كبرها . أي أن هذا العدد يؤول لان يتمركز في وسط المجموعة التي احتسب منها . وأهم مقاييس التوسط هي :

1 – الوسط الحسابي The Arithmetic Mean
2 – الوسيط The Median Mean
3 – المنوال The Mode
4 – الوسط الهندسي The Geometric Mean
5 – الوسط التوافقي The Harmonic Mean

هذه المقاييس جميعا تبحث عن قيمة تتمركز حولها أغلبية هذه البيانات برقم واحد يعبر عن أو يمثل جميع بيانات تلك المجموعة .وسنشرح كيفية حساب كل مقياس من المقاييس أعلاه في حالتين :

أ – بيانات مبوبة .
ب – بيانات غير مبوبة .

1.2 – الوسط الحسابي *The Arithmetic Mean*

وهو أهم مقاييس النزعة المركزية لما يمتاز به من خصائص وسهولة حسابه كما انه من أكثر مقاييس النزعة المركزية استخداما وهو عبارة عن مجموع القيم مقسوما على عددها وعادة ما يرمز له ($\bar{X}, \bar{Y}, \bar{Z} \ldots$)
طرق حسابه :-

- حساب الوسط الحسابي للبيانات غير المبوبة :

إذا كان المتغير X_i يأخذ القيم $X_1, X_2, X_3, \ldots . X_n$ فان الوسط الحسابي لهذه القيم يرمز له بالرمز $\bar{X}$ ويعبر عنه على انه :

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

مثال (1.2)

أوجد الوسط الحسابي لأجور 10 عمال كانت أجورهم الشهرية (بآلاف الدنانير) كما يلي :

الحل :

600 ، 750 ، 450 ، 380 ، 360 ، 520 ، 440 ، 525 ، 425 ، 550

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{5000}{10} = 500$$ ألف دينار

- حساب الوسط الحسابي للبيانات المبوبة :

يمكن حساب الوسط الحسابي للبيانات المبوبة باستخدام القانون التالي :

$$\bar{X} = \frac{\sum_{i=1}^{n} f_i X_i}{\sum_{i=1}^{n} f_i}$$

حيث أن f_i تمثل التكرارات وان X_i تمثل مراكز الفئات .

ويتم الحل على النحو التالي :

1 – نحسب مراكز الفئات .

2 - نحسب $f_i X_i$ لكل فئة (حاصل ضرب مراكز الفئات بالتكرارات المقابلة لها) .

3 – نطبق القانون .

مثال (2.2):

من جدول التوزيع التكراري التالي اوجد الوسط الحسابي :

الفئات	التكرارات f_i	مراكز الفئات	$x\ f_i$
03 – 07	8	5	40
08 – 12	3	10	30
13 – 17	2	15	30
18 – 22	3	20	60
23 - 27	4	25	100
	20		260

الحل

الفئات	التكرارات f_i	مراكز الفئات X_i	$f_i\ x_i$
03 – 07	8	5	40
08 – 12	3	10	30
13 – 17	2	15	30
18 – 22	3	20	60
23 - 27	4	25	100
	20		**260**

$$\bar{X} = \frac{\sum f_i X_i}{\sum f_i} = \frac{260}{20} = 13$$

1.1.2 - الوسط الحسابي المرجح :

بعض الظواهر تختلف قيمتها من حيث الأهمية النسبية تسمى هذه الأهمية النسبية وزن أو ترجيح لكل مفردة وعند حساب الوسط الحسابي وعند حساب الوسط الحسابي لتلك القيم يسمى بالوسط الحسابي المرجح أو الموزون والذي يرمز له بالرمز $\bar{X}_W$ حيث W تمثل الأوزان الترجيحية وتكون صيغته كما يلي :

$$\bar{X}_W = \frac{\sum_{i=1}^{n} X_i W_i}{\sum_{i=1}^{n} W_i}$$

مثال (3.2) :

الجدول التالي يبين درجات أحد طلبة قسم العلوم المالية والمصرفية في 9 مواد دراسية وعدد الوحدات الأسبوعية لتلك المواد والمطلوب حساب معدل الطالب في تلك المواد مرجحا بعدد الوحدات :

اسم المادة	عدد الوحدات W_i	الدرجة X_i
مبادئ إحصاء	3	70
مبادئ إدارة	3	65
مبادئ اقتصاد	3	50
مبادئ محاسبة	3	70
ثقافة مصرفية	3	82
نصوص E	2	63
حاسوب	2	72
حقوق إنسان	2	77
لغة عربية	2	79
	23	$\sum$

الحل :

X_iW_i	عدد الوحدات W_i	الدرجة X_i
210	3	70
195	3	65
150	3	50
210	3	70
246	3	82
126	2	63
144	2	72
154	2	77
158	2	79
1593	**23**	$\sum$

$$\bar{X}_W = \frac{\sum_{i=1}^{n} X_i W_i}{\sum_{i=1}^{n} W_i} = \frac{1593}{23} = 69.2$$

2 .1.2 - مزايا وعيوب الوسط الحسابي :

- مزايا الوسط الحسابي :

1. هو نقطة اتزان المشاهدات ويمتاز بسهولة حسابه وبساطة فكرته .
2. خضوعه للعمليات الحسابية .
3. إن حسابه يستند الى كافة البيانات المتاحة .

- عيوب الوسط الحسابي

1. يتأثر بالقيم المتطرفة والقيم الشاذة لذا لا يصلح للتوزيعات التكرارية الملتوية .
2. لا يصلح في حالة الفئات المفتوحة (لعدم وجود مركز فئة) .
3. لا يمكن حسابه في المتغيرات النوعية إلا إذا أمكن تحويل التغير النوعي الى كمي .
4. لا يمكن حسابه في حالة فقدان قيمة أو أكثر لعدم إمكانية حساب المجموع الكلي للقيم لأنه يعتمد على جميع القيم في حسابه .

خواص الوسط الحسابي :[1]

أ – مجموع انحرافات القيم عن وسطها الحسابي تساوي صفر أي :

للبيانات غير المبوبة: $\sum (X_i - \bar{X}) = 0$

ويمكن برهنته كما يلي :

$$\sum (X_i - \bar{X}) = \sum X_i - n\bar{X}$$

$$= \sum X_i - n\left(\frac{\sum X_i}{n}\right)$$

$$= \sum X_i - \sum X_i = 0$$

ولتوضيح ذلك في البيانات غير المبوبة نأخذ المثال التالي :

مثال (4.2) :

$X_i = 8 ، 12 ، 6 ، 9 ، 5$

$$\bar{X} = \frac{40}{5} = 8$$

فلو تم طرح الوسط الحسابي من كل قيمة من القيم الأصلية :

X_i	$X_i - \bar{X}$
8	0
12	4
6	– 2
9	1
5	– 3
40	0

للبيانات المبوبة :

2 – انظر الى :
- د.خاشع الراوي ، مبادئ الإحصاء ، مطابع جامعة الموصل ، ص ص 68 – 74 .
- د. محمود حسن المشهداني وأمين حنا هرمز ، الإحصاء ، بغداد ، ص ص162 – 166.
- د. ثائر فيصل شاهر ،الإحصاء في العلوم الإدارية والمالية ، دار الحامد للنشر والتوزيع ، عمان 2010 ، ص ص 84 -92 .

$$\sum f_i(X_i - \bar{X}) = 0$$

$$\sum f_i(X_i - \bar{X}) = \sum f_i X_i - \bar{X}\sum f_i$$

$$= \sum f_i X_i - \left(\frac{\sum f_i X_i}{\sum f_i}\right)\sum f_i$$

$$= \sum f_i X_i - \sum f_i X_i = 0$$

مثال (5.2) :

للتحقق من تلك الخاصية في البيانات المبوبة نأخذ البيانات التالية التي وسطها الحسابي 13:

$f_i(X_i - \bar{X})$	$X_i - \bar{X}$	مراكز الفئات X_i	التكرارات f_i	الفئات
-64	-8	05	8	03 –07
- 9	-3	10	3	08 – 12
4	2	15	2	13 – 17
21	7	20	3	18 – 22
48	12	25	4	23 - 27
0			20	

ب – عند إضافة عدد ثابت (k) الى كل قيمة من قيم المشاهدات فأن الوسط الحسابي للقيم الجديدة = الوسط الحسابي للقيم الأصلية + العدد الثابت (k) :

$$Y_i = X_i + k$$

$$\bar{Y} = \bar{X} + k$$

البرهان :

$$Y_i = X_i + k$$

$$\sum Y_i = \sum (X_i + \bar{X})$$

$$\sum Y_i = \sum X_i + nk$$

$$\frac{\sum Y_i}{n} = \frac{\sum X_i}{n} + \frac{nk}{n}$$

$$\bar{Y} = \bar{X} + k$$

مثال (6.2) : باستخدام بيانات المثال (4.2) والتي وسطها الحسابي = 8 وللتحقق من تلك الخاصية نضيف العدد 3 الى كل مفردات العينة وكما يلي :

X_i	$X_i + 3$
8	11
12	15
6	9
9	12
5	8
$\sum X_i = 40$ $\bar{X} = 8$	$\sum X_i = 55$

$$\bar{X} = \frac{55}{5} = 11$$

والذي يعني الوسط الحسابي القديم + 3 (أي 8 + 3 = 11)

ج – عند طرح عدد ثابت (k) من كل قيمة من قيم المشاهدات فأن الوسط الحسابي للقيم الجديدة = الوسط الحسابي للقيم الأصلية - العدد الثابت (k) :

$$Y_i = X_i - k$$

$$\bar{Y} = \bar{X} - k$$

البرهان :

تطبق نفس الخطوات السابقة مع تغيير الإشارة من موجب الى سالب .

مثال (7.2) :

باستخدام بيانات المثال (4.2) والتي وسطها الحسابي = 8 وبطرح العدد 2 من كل مفردات العينة :

X_i	$X_i - 2$
8 12 6 9 5	6 10 4 7 3
$\sum X_i = 40$ $\bar{X} = 8$	$\sum X_i = 30$

$$\bar{X} = \frac{30}{5} = 6$$

أي أن $\bar{X} - 2 = 8 - 2 = 6$

د – إذا ضربت كل قيمة من قيم المشاهدات في قيمة ثابتة (k) فأن الوسط الحسابي للقيم الجديدة يساوي = الوسط الحسابي للقيم الأصلية × العدد الثابت k أي أن :

$$Y_i = kX_i$$

$$\bar{Y} = k\,\bar{X}$$

البرهان :

$$Y_i = kX_i$$

$$\sum Y_i = k\sum X_i$$

$$\frac{\sum Y_i}{n} = \frac{k\sum X_i}{n}$$

$$\bar{Y} = k\,\bar{X}$$

مثال (8.2) : باستخدام بيانات المثال (4.2) والتي وسطها الحسابي = 6 وبضرب العينة

بالعدد 4 :

X_i	X_i
8 12 6 9 5	32 48 24 36 20
$\sum X_i = 40$ $\bar{X} = 8$	$\sum X_i = 160$

$$\bar{X} = \frac{160}{5} = 32$$

ويلاحظ أن الـوسط الحسابي الجديد = 32 هو عبارة عن 8 ×4 وهذا ما يحقق تلك الخاصية .

هـ - الوسط الحسابي لمجموع قيم متغيرين = مجموع الوسطين الحسابيين للمتغيرين :

$$Z_i = X_i + Y_i$$

$$\bar{Z} = \bar{X} + \bar{Y}$$

البرهان :

$$Z_i = X_i + Y_i$$

$$\sum Z_i = \sum (X_i + Y_i)$$

$$\sum Z_i = \sum X_i + \sum Y_i$$

$$\frac{\sum Z_i}{n} = \frac{\sum X_i}{n} + \frac{\sum Y_i}{n}$$

$$\bar{Z} = \bar{X} + \bar{Y}$$

مثال (9.2) :

باستخدام نفس العينة في المثال (2.4) والتي وسطها الحسابي = 8 وعينة أخرى كما في الجدول أدناه ووسطها الحسابي = 7 يمكن برهنة أعلاه رقميا كما يلي :

X_i	Y_i	$Z_i = X_i + Y_i$
8	7	15
12	11	23
6	5	11
9	8	17
5	4	9
$\sum X_i = 40$ $\bar{X} = 8$	$\sum X_i = 35$ $\bar{X} = 7$	$\sum Z_i = 75$

$$\bar{Z} = \frac{\sum Z_i}{n} = \frac{75}{5} = 13$$

الوسيط: *The Median*

أحد مقاييس النزعة المركزية وهو القيمة التي تتوسط مجموعة القيم بعد ترتيبها تصاعديا أو تنازليا ويرمز له بالرمز *(Me)* .

أ – حساب الوسيط للبيانات غير المبوبة :

يمكن حساب الوسيط للبيانات غير المبوبة بالخطوات الآتية :

- يتم ترتيب القيم تصاعديا أو تنازليا .
- إذا كان عدد القيم فرديا فأن الوسيط هو القيمة التي ترتيبها $\frac{n+1}{2}$.
- أذا كان عدد القيم زوجيا فأن الوسيط عبارة عن الوسط الحسابي للقيمتين الواقعتين في الترتيبين $\frac{n}{2}$ و $\frac{n}{2}+1$.

مثال (10.2) :

الأتي أوزان 11 فرادا والمطلوب حساب متوسط وزنهم باستخدام الوسيط .

73 ، 69 ، 82 ، 66 ، 85 ، 67 ، 59 ، 77 ، 80 ، 65 ، 71

- نرتب القيم تصاعديا :

59 ، 65 ، 66 ، 67 ، 69 ، 71 ، 73 ، 77 ، 80 ، 82 ، 85

- نجد ترتيب الوسيط AMe : $AMe = \frac{11+1}{2} = 6$

$$Me = 71$$

مثال (11.2) :

ألأتي درجات 8 طلاب استخدم الوسيط لتقدير وزن الطالب .

76 ، 68 ، 90 ، 45 ، 57 ، 78 ، 89 ، 44

الحل :

نرتب القيم تصاعديا :

44 ، 45 ، 57 ، 68 ، 76 ، 78 ، 89 ، 90

ترتيب الوسيط AMe : $\frac{n}{2} = \frac{8}{2} = 4$ $\quad \frac{n}{2}+1 = \frac{8}{2}+1 = 5$

$$Me = \frac{68+76}{2} = 72$$

ب – حساب الوسيط للبيانات المبوبة :

يعتمد حساب الوسيط للبيانات المبوبة على التوزيع التكراري المتجمع الصاعد أو النازل ويتم استخراجه بتطبيق القانون التالي :

$$Me = L_1 + \left[\frac{\sum f_i / 2 - f^*}{f_m} \right] W$$

حيث أن :

L_1 = الحد الأدنى الحقيقي لفئة الوسيط .

$\sum f_i$ = مجموع التكرارات .

f^* = التكرار المتجمع الصاعد السابق لفئة الوسيط (عند بداية فئة الوسيط) .

f_m = تكرار فئة الوسيط (أي التكرار المتجمع الصاعد عند نهاية فئة الوسيط - التكرار المتجمع الصاعد السابق لفئة الوسيط).

W: طول الفئة

ويمكن تلخيص خطوات الحل كما يلي :-

- تكوين جدول توزيع تكراري تجميعي تصاعدي أو تنازلي .
- إيجاد ترتيب الوسيط (موقع الوسيط) AMe باستخدام القانون : $AMe = \frac{\sum f_i}{2}$.
- البحث عن القيمة $\frac{\sum f_i}{2}$ (ما يساويها أو اكبر منها) .
- نحدد فئة الوسيط وهي الفئة التي تقع تلك القيمة بين حديها وذلك عن طريق إيجاد قيمتين متتاليتين في التكرار التجميعي التصاعدي يقع بينهما ترتيب الوسيط ، يقابل هاتين القيمتين حدا فئة الوسيط الأدنى والأعلى (ويستحسن أخذ الحدود الحقيقية لهذه الفئة) ، فإذا كانت قيمة الوسيط مساوية لأي تكرار متجمع صاعد فأن فئة ذلك التكرار ستكون هي الفئة الوسيطية ، أما إذا وقعت بين تكرارين متجمعين فإن الفئة اللاحقة لقيمة ترتيب الوسيط ستكون هي فئة الوسيط [1]أي الفئة التي تقابل التكرار المتجمع الصاعد الأكبر من بينهما [2].

---

[1] - د عبد الحميد عبد المجيد البلداوي ، أساليب الإحصاء ، مصدر سابق ، ص 75 .

[2] - د.ثائر فيصل شاهر ، مصدر سابق ،ص 97 .

مثال (12.2) :
أوجد الوسيط للتوزيع التكراري التالي :

الفئات	التكرارات f_i
03 – 07	8
08 – 12	3
13 – 17	2
18 – 22	3
23 - 27	4
	20

الحل :
- نستخرج الحدود الحقيقية للفئات والتكرار التجميعي التصاعدي:

الفئات	التكرارات f_i	الحدود الحقيقية للفئات	التكرار التجميعي التصاعدي
03 – 07	8	02.5 – 07.5	8
08 – 12	3	07.5 – 12.5	11
13 – 17	2	12.5 – 17.5	13
18 – 22	3	17.5 – 22.5	16
23 - 27	4	22.5 – 27.5	20
	20		

نستخرج ترتيب الوسيط : $AMe = \frac{\sum f_i}{2} = \frac{20}{2} = 10$

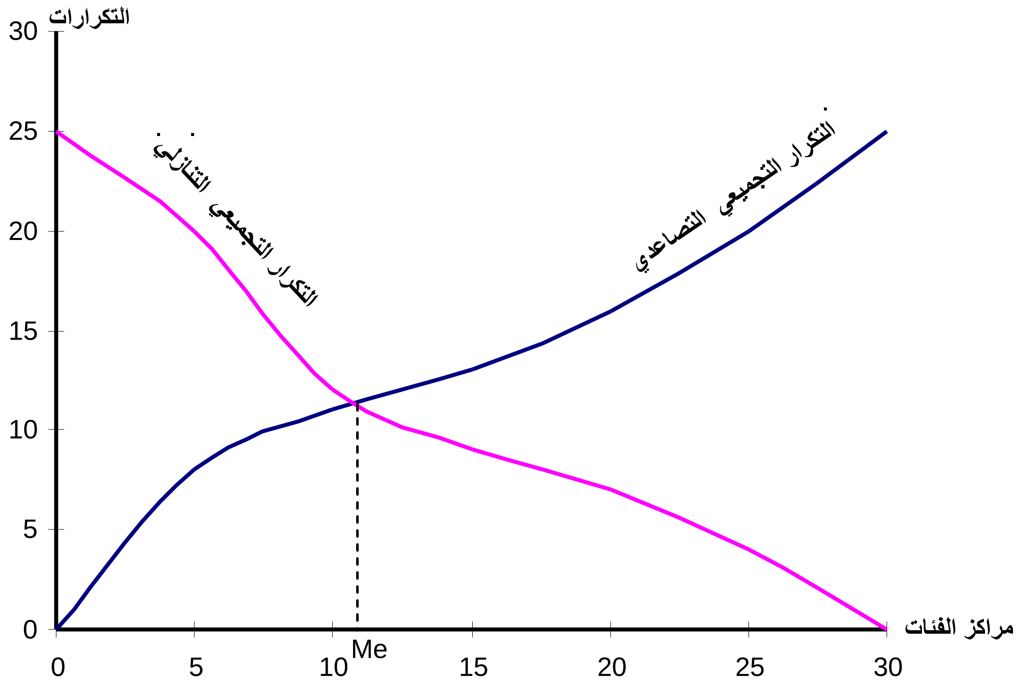
- إن الوسيط يقع بين التكرارين التصاعديين 8 و 11 .
- f_m تكرار فئة الوسيط (3= 8- 11) .
- $f^* =$ التكرار المتجمع الصاعد السابق لفئة الوسيط (عند بداية فئة الوسيط) = 8
- L_1 الحد الأدنى الحقيقي لفئة الوسيط $L_1 = 07.5$.
- نحسب قيمة الوسيط وفق الصيغة التالية :

$$Me = L_1 + \left[\frac{\sum f_i / 2 - f^*}{f_m} \right] W$$

$$Me = 7.5 + \left[\frac{20/2 - 8}{3} \right](5) = 10.8$$

وبنفس بيانات المثال السابق يمكن أيضا إيجاد ترتيب قيمة الوسيط هندسيا للمنحنيين التصاعدي والتنازلي وذلك بإنزال عمود من نقطة تقاطعهما الى المحور الأفقي ليقطعه في نقطة هي قيمة الوسيط كما مبين في الشكل البياني التالي :

شكل (2 – 1) استخراج قيمة الوسيط هندسيا



ويلاحظ بأن قيمة الوسيط المستخرجة بيانيا هي ذات القيمة المستخرجة بموجب المعادلة السابقة .

كما يمكن استخراج الوسيط من الجداول التكرارية المفتوحة والمثال التالي يوضح ذلك .

مثال (13.2) :

أوجد الوسيط من جدول التكرار التالي الذي يمثل القروض الممنوحة من احد المصارف التجارية في بغداد خلال عام 2011 لعدد من المزارعين بملايين الدنانير:

عدد المزارعين	مبالغ القروض
50	5 - أقل من 1
33	06 – 10
65	11 – 15
29	16 – 20
31	21 - 25
16	فأكثر 30 - 26
224	

الحل :

بتطبيق نفس خطوات حل المثال السابق نصل على :

$$Me = L_1 + \left[\frac{\sum f_i / 2 - f^*}{f_m} \right] W$$

التكرار التجميعي الصاعد	الحدود الحقيقية للفئات	عدد المزارعين	مبالغ القروض
50	4.5 – أقل من 0.5	50	5 - أقل من 1
83	5.5 – 10.5	33	06 – 10
148	10.5 – 15.5	65	11 – 15
177	15.5 – 20.5	29	16 – 20
208	20.5 – 25.5	31	21 - 25
224	فأكثر 29.5 -25.5	16	فأكثر 30 - 26
		224	

$$Me = 10.5 + \left[\frac{112 - 83}{65} \right] 5 = 12.73$$

2.2. 1- مميزات وعيوب الوسيط :

- مميزات الوسيط :

1. سهولة حسابه سوءا كانت البيانات مبوبة أو غير مبوبة .
2. يمكن حسابه من جداول التوزيع التكراري المفتوحة.

3. يمثل الوسيط البيانات تمثيلا سليما حينما تكون القيم للمفردات المرتبة متقاربة .
4. يمكن تقدير الوسيط في حالة الصفات الوصفية التي لا تقاس بأعداد مباشرة شريطة أن تكون قابلة للترتيب .
5. مجموع الانحرافات المطلقة لأفراد العينة عن قيمة الوسيط أقل من أي قيمة أخرى .
6. يمكن تعينه بيانيا .

- **عيوب الوسيط** :
1. ليس الوسيط شائع الاستعمال كالوسط الحسابي بالرغم من سهولة فهمه فهو اقل استعمالا منه .
2. لا يمكن الاستفادة من الوسيط حسابيا ومثال ذلك انه لا يمكن حساب وسيط عام لعدد من العينات المعروف قيمة وسيط كل منها كما هو الحال في الوسط الحسابي .
3. ليس الوسيط حساسا للتغيرات التي تحدث في قيم المفردات الداخلة في التوزيع ، فقد نتمكن من تغيير هذه القيم دون أن تتأثر قيمة الوسيط طالما أن هذا التغيير لا ينقله الى الجهة الأخرى .
4. لا تستعمل جميع قيم المتغير في حسابه بل يعتمد على جزء منها .

المنوال: *The Mode*

يعرف المنوال لمجموعة من القيم بأنه القيمة الأكثر شيوعاً بينها ، وقد يوجد أكثر من منوال لمجموعة القيم ، كما إن هذا المقياس يعد من المقاييس المستخدمة في الحالة العملية بكثرة فمثل هذه القيمة نجدها في مقاسات الملابس إذ توجد بعض المقاسات تتكرر لأغلبية الناس أو في مقاسات الأحذية .

أ ـ حساب المنوال للبيانات غير المبوبة :

إذا كان لدينا n من المشاهدات فان المنوال لهذه المشاهدات هو المشاهدة او القيمة الأكثر تكرارا بين هذه المشاهدات ويرمز له بالرمز Mo .

ويجب ملاحظة ما يلي :

1 - في حال وجود أكثر من منوال (متجاورة) فمتوسط هذه المنوالات يعتبر منوال التوزيع حال وجود تلك القيمة في التوزيع.

2 - يسمى التوزيع أحادي المنوال إذا وجد منوال واحد، وثنائي المنوال إذا وجد منوالان ، وهكذا.

3 - إذا وجد أكثر من منوال فالأكثر تكرار بينهم يدعى المنوال الرئيسي *(Mode Major)* والأخرى تعرف بالمنوال الفرعي أو الثانوي *(Mode Minor)* .

1 - قد لا يكون هناك منوال أو لا يوجد منوال للمشاهدات .

مثال (14.2) :

مجموعة القيم : 10 ، 13، 8 ، 9 ، 10 ، 10، 23 لها منوال واحد هو القيمة 10 .

مجموعة القيم : 12 ، 13، 7 ، 7 ، 12 ، 12، 23 يوجد أكثر من منوال 12، 7 .

مجموعة القيم : 12 ، 5 ، 4 ، 18 ،10، 16، 8 لا يوجد منوال لها .

ب ـ حساب المنوال للبيانات المبوبة :

يستخرج المنوال للبيانات المبوبة بتطبيق القانون التالي :

$$Mo = L_1 + \left[\frac{d_1}{d_1 + d_2} \right] W$$

حيث أن :

فئة المنوال تلك الفئة التي تمتلك أكبر التكرارات .

وان : L_1 = الحد الأدنى الحقيقي لفئة المنوال .

d_1 = الفرق بين تكرار فئة المنوال والفئة السابقة لها .

d_2 = الفرق بين تكرار فئة المنوال والفئة اللاحقة لها .

W = طول الفئة .

مثال (15.2) :

أوجد المنوال لجدول التوزيع التكراري التالي :

الفئات	التكرارات f_i
06 – 10	7
11 – 15	10
16 – 20	18
21 – 25	9
26 – 30	13
	57

الحل :

الفئات	التكرارات f_i
05.5 – 10.5	7
10.5 – 15.5	10
15.5 – 20.5	18
20.5 – 25.5	9
25.5 – 30.5	13
	57

- فئة المنوال هي 20.5 – 15.5 لأنها تقابل اكبر التكرارات وعليه فأن $15.5 = L_1$

- $8 = 10 - 18 = d_1$

- $9 = 9 - 18 = d_2$

- $5 = W$

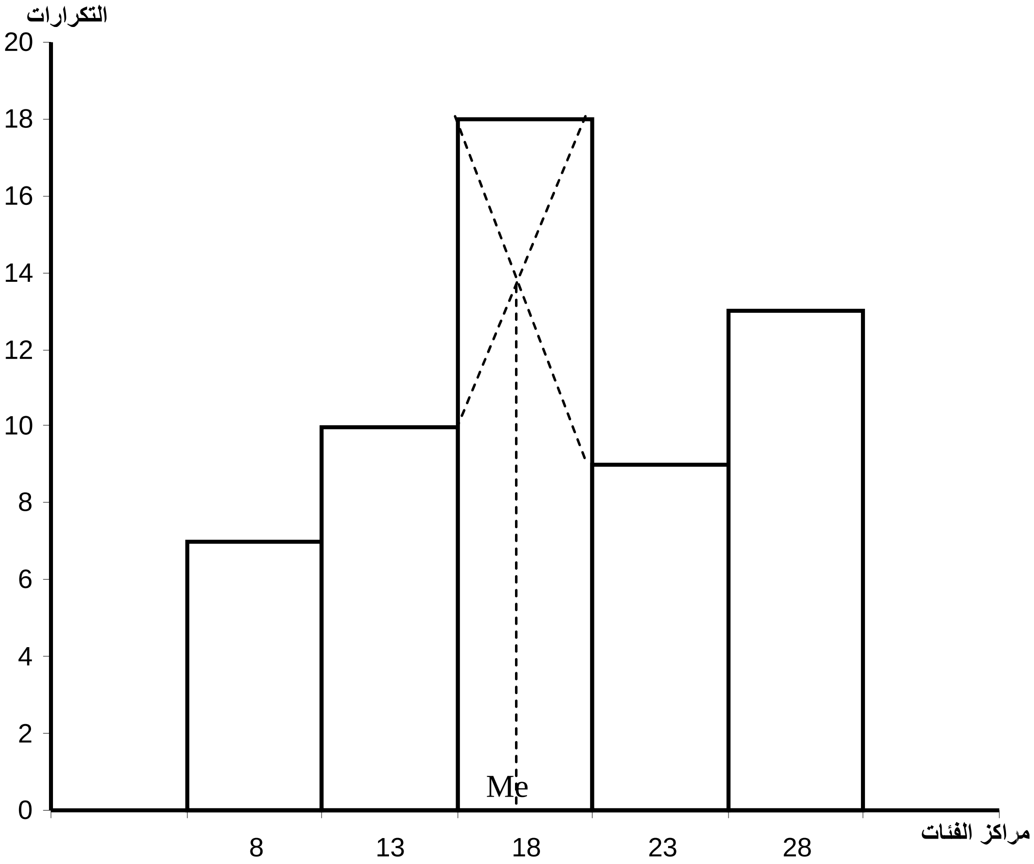
ثم:

$$Mo = L_1 + \left[\frac{d_1}{d_1 + d_2} \right] W$$

$$Mo = 15.5 + \left[\frac{8}{8+9} \right] 5 = 17.85$$

ويمكن كذلك تقدير قيمة المنوال هندسيا ، فمن خلال بيانات المثال السابق يتم رسم المدرج التكراري واستعمال مستطيل الفئة المنوالية باعتباره يمثل أكبر التكرارات والمستطيلان المجاوران له وكما يلي :

شكل (2 -2) استخراج قيمة المنوال بيانيا



ويمكن استخراج المنوال من بيانات الجداول التكرارية المفتوحة والمثال التالي يوضح ذلك :

مثال (16.2) :

من بيانات المثال (13.2) اوجد قيمة المنوال :

الحل :

- نستخرج الحدود الحقيقية للفئات :

الحدود الحقيقية للفئات	عدد المزارعين	مبالغ القروض
4.5 – أقل من 0.5	50	5 - أقل من 1
5.5 – 10.5	33	06 – 10
10.5 – 15.5	65	11 – 15
15.5 – 20.5	29	16 – 20
20.5 – 25.5	31	21 - 25
25.5- 29.5 فأكثر	16	26 - 30 فأكثر

	224	

- ثم نحتسب قيمة المنوال بنفس الخطوات السابقة :

$$Mo = L_1 + \left[\frac{d_1}{d_1 + d_2} \right] W$$

$$Mo = 10.5 + \left[\frac{32}{32+36} \right] 5 = 12.8$$

3.2. 1 - مميزات وعيوب المنوال :

- مميزات المنوال :

1. في حسابه لا تستخدم جميع المفردات في التوزيع .
2. المنوال ليس مقياسا يمكن ضربه في عدد المفردات في المجموعة لينتج المجموع الكلي.
3. لا يتأثر المنوال بالقيم الشاذة أو المتطرفة كبيرة كانت أم صغيرة .
4. يمكن حساب المنوال في حالة الجداول المفتوحة .
5. يسهل تقدير المنوال بمجرد النظر إذا كان عدد المفردات قليلا .
6. يمكن تعينه بيانيا .

- عيوب المنوال :

1. قيمة المنوال تقريبية غير معتمد عليها وتتأثر قيمته بشدة من عينة الى أخرى من نفس المجتمع .
2. لا تتغير قيمة المنوال بأحداث تغيرات في القيم الأخرى ما دام تكرار القيمة لم يتغير .
3. أقل المقاييس للمتوسط في دقة حسابه .
4. تتغير قيمة المنوال عند تقديرها في الجدول التكراري إذا تغير عدد الفئات وأطوالها لنفس البيانات .

الارتبــاط

١.١٠ مقدمة :

تناولنا في الفصول السابقة دراسة متغير واحد أو ظاهرة واحدة من حيث قياس وحساب متوسط هذه الظاهرة وكذلك حساب مقياس لتشتتها. ولكن في الحياة العملية كثيراً ما يحتاج الباحث لدراسة العلاقة بين ظاهرتين (أو متغيرين) لمعرفة مدى الارتباط بينهما ونوع هذا الارتباط. فقد يريد الباحث معرفة ما إذا كان هناك علاقة بين التدخين والإصابة بمرض في الرئة، أو بين درجة تعليم الشخص ومستوى دخله. أو بين الحالة التعليمية والحالة الاجتماعية للناخب. وكما نرى فإنه يمكن أن نذكر الكثير بين الأمثلة في مختلف المجالات بل قد يرغب الباحث في دراسة العلاقة بين أكثر من متغيرين في وقت واحد. فمثلاً قد يريد الباحث معرفة تأثير درجة التعليم ومستوى الدخل وحجم الأسرة على درجة الوعي السياسي للشخص. في هذا المثال يريد الباحث معرفة تأثير ثلاثة متغيرات على متغير رابع. وفي هذا الكتاب سوف نركز على دراسة العلاقة بين **متغيرين اثنين فقط** وهو ما يعرف **بالارتباط " البسيط "** Simple Correlation. بينما الحالات التي تتناول الدراسة فيها **أكثر من متغيرين** تعرف **بالارتباط المتعدد** Multiple Correlation وهي – كما ذكرنا – خارج نطاق هذا الكتاب.

٢.١٠ أنواع العلاقة بين المتغيرين :

إذا كان المتغيران يتغيران معاً في الاتجاه نفسه بمعنى أنه إذا زاد أو نقص أحدهما، زاد أو نقص الأخر، فإن العلاقة بينهما تكون طردية والارتباط بينهما يكون موجباً. مثال ذلك العلاقة بين زيادة حجم الطبقة الوسطى في المجتمع وزيادة الاستقرار السياسي.

وإذا كان المتغيران يتغيران معاً ولكن في عكس الاتجاه بمعنى أنه إذا زاد أحدهما نقص الأخر، أو إذا نقص أحدهما زاد الأخر، فإن العلاقة بينهما تكون عكسية والارتباط بينهما يكون سالباً. مثال العلاقة بين تدني مستوى الفرد التعليمي ودرجة الوعي الاجتماعي.

وتختلف العلاقات بين الظواهر من حيث القوة. فقد تكون العلاقة قوية جداً (أو حتى تامة)، وقد تكون متوسطة، أو ضعيفة، أو منعدمة تماماً. وفي هذا الفصل سوف نتناول بالتفصيل كيفية حساب الارتباط بين متغيرين سواء كان المتغيران كميين أو وصفيين (ترتيبيين أو اسميين)، أو أحدهما كمياً والأخر وصفياً.

٣.١٠ شكل الانتشار Scatter Diagram :

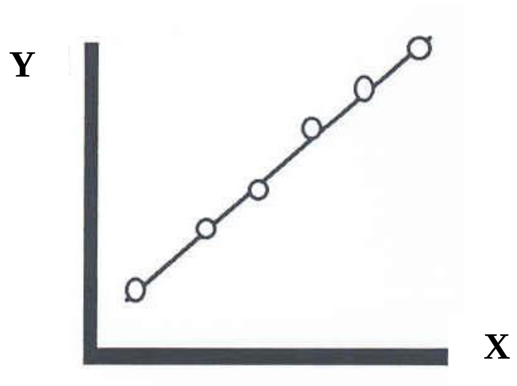
هناك وسيلة مبدئية يعرف الباحث من خلالها نوع الارتباط بين المتغيرين وما إذا كان الارتباط قوياً وضعيفاً أو منعدماً، وما إذا كانت العلاقة خطية أو غير خطية، موجبة أو سالبة. هذه الوسيلة هي " شكل الانتشار " والتي تصلح إذا كان المتغيران كميين. وجدير بالذكر أن هذه وسيلة مبدئية تساعد فقط في معرفة نوع الارتباط ولا تعتبر بديلاً عن الطرق الإحصائية التي سوف نتناولها بالتفصيل في هذا الفصل.

والمقصود بشكل الانتشار هو تمثيل قيم الظاهرتين بيانياً على المحورين، المتغير الأول X على المحور الأفقي، والمتغير الثاني Y على المحور الرأسي، حيث يتم تمثيل كل زوج Pair من القيم بنقطة، فنحصل على شكل يمثل كيفية انتشار القيم على المستوى، وهو الذي يسمى شكل الانتشار. وطريقة انتشار القيم تدل على وجود أو عدم وجود علاقة بين المتغيرين ومدى قوتها ونوعها. فإذا كانت تتوزع بشكل منتظم دل ذلك على وجود علاقة (يمكن استنتاجها)، أما إذا كانت النقط مبعثرة ولا تنتشر حسب نظام معين دل ذلك على عدم وجود علاقة بين المتغيرين أو أن العلاقة بينهما ضعيفة. والأشكال التالية تظهر بعض أشكال الانتشار المعروفة :

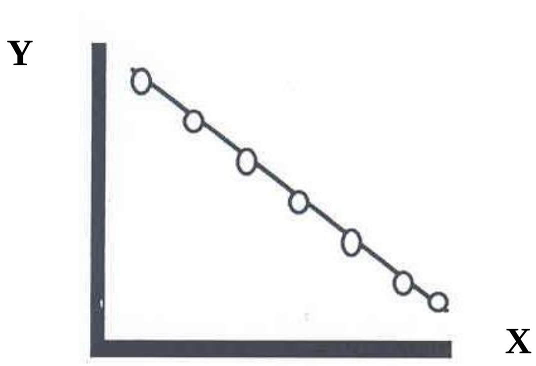
<u>الشكل الأول :</u>

إذا وقعت جميع النقاط على خط مستقيم، دل ذلك على أن العلاقة بينهما خطية وأنها ثابتة أو تامة. وهذه تمثل أقوى أنواع الارتباط بين المتغيرين " ارتباط تام ". فإذا كانت العلاقة طردية فإن " الارتباط طردي تام " كما في الشكل الأول (أ). ومثاله العلاقة بين الكمية المشتراة من سلعة والمبلغ المدفوع لشراء هذه الكمية.

أما إذا كانت العلاقة عكسية (وجميع النقاط تقع على خط مستقيم واحد) فإن " الارتباط عكسي تام " كما في الشكل الأول (ب). ومثال على ذلك العلاقة بين السرعة والزمن.



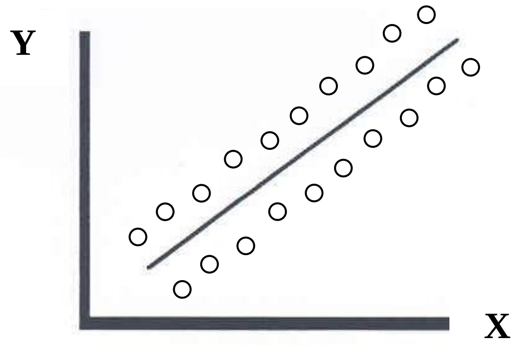
الشكل الأول (أ)
ارتباط طردي تام
(موجب)



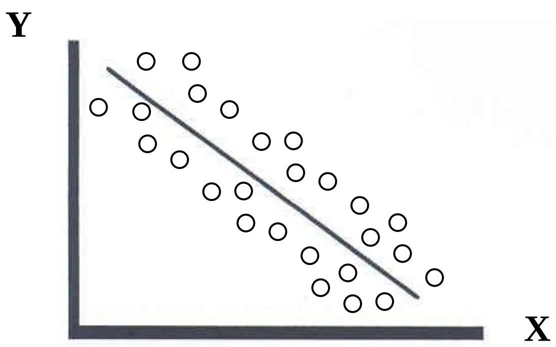
الشكل الأول (ب)
ارتباط عكسي تام
(سالب)

الشكل الثاني :

أما إذا كانت النقاط تأخذ شكل خط مستقيم ولكن لا تقع جميعها على الخط قيل أن العلاقة خطية (موجبة أو سالبة) كما في الشكل الثاني أ، ب.



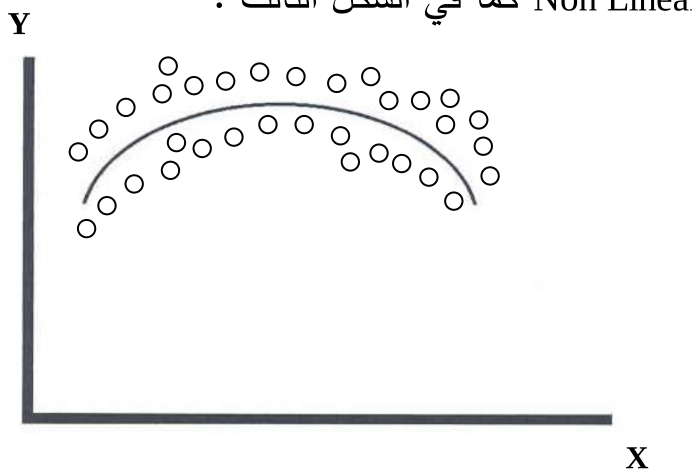
الشكل الثاني (أ)
ارتباط موجب قوي
(ارتباط خطي طردي)



الشكل الثاني (ب)
ارتباط سالب قوي
(ارتباط خطي عكسي)

الشكل الثالث :

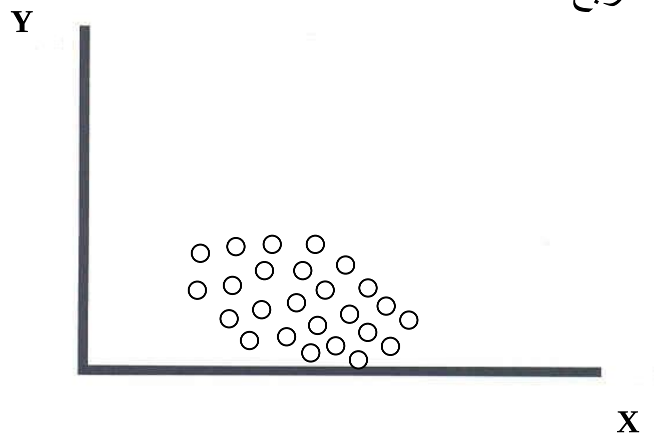
وإذا كانت العلاقة تأخذ شكل منحنى فإن الارتباط لا يكون خطياً "ارتباط غير خطي Non Linear Correlation " كما في الشكل الثالث :



الشكل الثالث
(ارتباط غير خطي)

الشكل الرابع :

أما إذا كانت النقاط تتبعثر بدون نظام معين فإن ذلك يدل على عدم وجود علاقة بين المتغيرين (أو أن العلاقة بينهما ضعيفة جداً) كالعلاقة مثلاً بين دخل الشخص وطوله كما في الشكل الرابع :



الشكل الرابع
(لا توجد علاقة)

٤.١٠ معامل الارتباط Correlation Coefficient :

يقاس الارتباط بين متغرين بمقياس إحصائي يسمى " معامل الارتباط " ويعكس هذا المقياس درجة أو قوة العلاقة بين المتغيرين واتجاه هذه العلاقة. وتنحصر قيمة معامل الارتباط بين + ١، – ١. فإذا كانت قيمة معامل الارتباط تساوي + ١ فمعنى ذلك أن الارتباط بين المتغيرين طردي تام، وهو أقوى أنواع الارتباط الطردي بين متغيرين. وإذا كانت قيمة معامل الارتباط تساوي – ١ فمعنى ذلك أن الارتباط بين المتغيرين عكسي تام، وهو أقوى أنواع الارتباط العكسي بين متغيرين. وإذا كانت قيمة معامل الارتباط تساوي صفر، فمعنى ذلك أنه لا يوجد ارتباط بين المتغيرين. وكلما اقتربت قيمة معامل الارتباط من + ١ أو – ١ كلما كان الارتباط قوياً، وكلما اقترب من الصفر كلما كان الارتباط ضعيفاً.

والخلاصة :

أنه كلما كانت العلاقة قوية بين المتغيرين كلما اقترب معامل الارتباط من + ١ أو – ١ فإذا وصلت قيمة المعامل إلى + ١ أو – ١ كان الارتباط تاماً بين المتغيرين. وأنه كلما كانت العلاقة ضعيفة بين المتغيرين كلما اقترب معامل الارتباط من الصفر، فإذا وصلت قيمة المعامل إلى الصفر كان الارتباط منعدماً بين المتغيرين. ومعنى ذلك أيضاً أنه لا يوجد ارتباط بين متغيرين تكون قيمة المعامل فيه أكبر من + ١ ولا أصغر من – ١. وسنبدأ بقياس العلاقة بين متغيرين كميين، ثم متغيرين ترتيبين، وأخيراً متغيرين أسميين.

٥.١٠ معامل بيرسون للارتباط الخطي البسيط :

يفترض بيرسون Pearson أن المتغيرين كميان، وأن العلاقة بينهما خطية (أي تأخذ شكل خط مستقيم، أنظر الشكل الثاني من أشكال الانتشار).

ويرى بيرسون أن أفضل مقياس للارتباط بين متغيرين قد يختلفان في وحدات القياس و / أو في مستواهما العام (مثل الارتباط بين العمر والدخل) حيث يقاس العمر بالسنوات ويقاس الدخل بالعملة، بالريال أو الدولار.. كما أن المستوى العام للعمر – أي متوسط العمر – قد يساوي **أربعين** عاماً. فبينما المستوى العام – أي متوسط – الدخل الشهري قد يكون **خمسة آلاف** ريال مثلاً).

وبالتالي فإن أفضل مقياس للارتباط بين مثل هذين المتغيرين – حسب رأي بيرسون – هو عن طريق حساب انحرافات كل من المتغيرين عن وسطه الحسابي وقسمة هذه الانحرافات على الانحراف المعياري لكل منهما، فنحصل على ما يسمى بالوحدات المعيارية لكل متغير. ويكون معامل ارتباط بيرسون هو " متوسط حاصل ضرب هذه الوحدات المعيارية ". ومعامل الارتباط يكون بدون تمييز.

وبالرموز، إذا فرضنا أن المتغيرين هما Y , X وأن لدينا عدد n من أزواج القيم هي :

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$$

وأن الوسط الحسابي للمتغير X هو $\overline{X}$ وللمتغير Y هو $\overline{Y}$ وأن الانحراف المعياري للمتغير X هو S_x وللمتغير Y هو S_y فإن معامل بيرسون للارتباط الخطي والذي يرمز له بالرمز r هو :

الصيغة التعريفية لمعامل الارتباط

$$\boxed{r = \frac{1}{n} \sum \left(\frac{x - \bar{x}}{S_x} \right) \left(\frac{y - \bar{y}}{S_y} \right)}$$

ونلاحظ من تعريف معامل بيرسون للارتباط الخطي البسيط أنه يجب أولاً حساب كل من $\overline{X}, \overline{Y}, S_x, S_y$، ثم حساب $\frac{x - \bar{x}}{S_x}$ لكل قيمة من قيم X، وحساب $\frac{y - \bar{y}}{S_y}$ لكل قيمة من قيم y ثم ضرب $\frac{x - \bar{x}}{S_x}$ في $\frac{y - \bar{y}}{S_y}$ لكل زوج من القيم وأخذ مجموع حاصل الضرب ثم القسمة على n. إن هذه العملية كما نرى تستغرق وقتاً طويلاً ونحتاج عمليات حسابية معقدة، لذلك فإنه عادة لا تستخدم الصيغة السابقة في حساب معامل الارتباط وتستخدم بدلاً منه **<u>الصيغة المختصرة التالية</u>** والتي تعطي بطبيعة الحال النتائج نفسها : –

(٢) الصيغة المختصرة لمعامل الارتباط

$$\boxed{r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{n \sum x^2 - (\sum x)^2}\ \sqrt{n \sum y^2 - (\sum y)^2}}}$$

وكل ما نحتاجه لحساب معامل الارتباط الخطي لبيرسون بالصيغة المختصرة رقم (2) هو حساب : $\sum xy , \sum x^2 , \sum y^2$ أي مجموع مربعات قيم x ومجموع مربعات قيم y ومجموع حاصل ضربهما بعد معرفة $\sum X$، $\sum Y$ ، n (حيث n هي عدد أزواج القيم).

<u>مثال (١) :</u>

البيانات التالية تمثل أعمار ثمانية من الناخبين ودخولهم اليومية بالدولار، والمطلوب حساب معامل بيرسون للارتباط الخطي بين الأعمار والدخول.

الأعمار x :	25	32	29	43	38	51	47	35
الدخول y :	10	18	15	35	40	62	100	50

الحل :

لحساب معامل بيرسون للارتباط الخطي يلزم حساب المجاميع:

$\sum x, \sum y, \sum xy, \sum x^2, \sum y^2$ لذلك يتم تنظيم حساب هذه المجاميع كما في الجدول التالي:

الأعمار x	الدخول y	xy	x^2	y^2
25	10	250	625	100
32	18	576	1024	324
29	15	435	841	225
43	35	1505	1849	1225
38	40	1520	1444	1600
51	62	3162	2601	3844
47	100	4700	2209	10000
35	50	1750	1225	2500
300	330	13898	11818	19818

ثم نطبق في الصيغة المختصرة رقم (2) لمعامل الارتباط حيث n = 8 :

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{n\sum x^2 - (\sum x)^2}\ \sqrt{n\sum y^2 - (\sum y)^2}}$$

$$= \frac{8(13898) - (300)(330)}{\sqrt{8(11818) - (300)^2}\ \sqrt{8(19818) - (330)^2}}$$

$$= \frac{111184 - 99000}{\sqrt{94544 - 9000}\ \sqrt{158544 - 108900}}$$

$$= \frac{12184}{\sqrt{4544}\ \sqrt{49644}}$$

$$= \frac{12184}{15019.6}$$

$$r = 0.81$$

أي أن معامل بيرسون للارتباط الخطي بين أعمار الناخبين ودخولهم اليومية يساوي 0.81 وهو ارتباط طردي (لأن إشارته موجبة) وقوى (لأنه قريب من الواحد الصحيح). بمعنى أخر، إن هناك علاقة طردية قوية بين عمر الناخب ودخله مقدارها 81 %. فمع زيادة عمر الناخب يزيد دخله، والعكس صحيح.

ملاحظة مهمة :

من خواص معامل بيرسون للارتباط الخطي أنه لا يتأثر بالعمليات الحسابية التي تجري على المتغيرين x , y. بمعنى أنه لا يتأثر بالطرح (أو الجمع)، ولا بالقسمة (أو الضرب). أي إذا طرحنا (أو جمعنا) قيمة معينة من كل قيم x وقيمة أخرى من كل قيم y، أو قسمنا (أو ضربنا) قيم x على قيمة معينة وكل قيم y على قيمة أخرى فإن قيمة معامل الارتباط لا تتغير أي نحصل على القيمة نفسها.