تحليل البيانات الصحفية

د. جخدم موسى
جامعة عمارثليجي بالأغواط
كلية العلوم الإنسانية والإسلامية والحضارة – قسم علوم الإعلام والاتصال

2023/2022

مطبوعة موجهة لطلبة سنة ثانية إعلام واتصال

1. معلومات عن المقياس:

اسم المادة
تحليل البيانات الصحفية

المهارات المطلوبة للنجاح في المادة	المهارات التحليلية، التفكير الكمّي، الإحساس المنطقي جميعها مهارات مطلوبة و مهمةَ للنَجاح في هذا المقياس
الهدف العام للمادة العلمية	يهدف هذا المقياس إلى تعريف الطالب بالمفاهيم الأساسية في علم الإحصاء وطرق تمثيل وعرض ووصف البيانات مع إعطاء نماذج تطبيقية .

2. الأهداف التعليمية والمهارات المكتسبة:

أ- **معرفية** :

بعد الانتهاء من دراسة هذا المقياس يتوقع من الطالب أن تستوعب المفاهيم والمصطلحات الأساسية في علم الإحصاء وأن يكون قادر على وصف البيانات باستخدام المقاييس الإحصائية مع إدراك مدلولاتها وأن يتعرف على العلاقات بين الظواهر من خلال بياناتها بالإضافة إلى أخذ فكرة عن تطبيقات الإحصاء في مجالات العلوم الاجتماعية والإنسانية.

ب- **مهارات علمية** :

يتوقع من الطالب تطبيق المنهج الإحصائي في بيانات الظواهر الاجتماعية والانسانية التي يحتاج لدراستها وبحثها لحل المشاكل وذلك بأن يميز الطالب بين أنواع البيانات الإحصائية والتحليل الإحصائي المناسب لتعميم النتائج.

ج- **مهارات شخصية وتحمل المسؤولية** :

يهدف هذا المقياس إلى تعزيز مهارة الطالب في مواجهة حل بعض المشاكل باستخدام الطرق الإحصائية والالتزام ببعض السياسات التي تنمي في الطالب تحمل المسئولية مثل:

- حضور الأعمال الموجهة في الزمن المحدد.
- أداء الواجبات المطلوبة منه في الموعد المحدد.
- أداء الاختبارات في الموعد المحدد.
- الالتزام بقواعد الحوار والمناقشة.

د- ***مهارات التحليل والاتصال***

يتوقع من الطالب أن يكون لديه فكرة عامة عن تحليل وتفسير البيانات باستخدام التطبيقات الإحصائية والقدرة على قراءة مخرجات التطبيقات الإحصائية .

مقدمة

تحليل البيانات أو مايعرف بالإحصاء هو علم يهتم بالمعلومات، ويهدف إلى تجميعها وتبويبها وتنظيمها وتحليلها واستخلاص النتائج منها،واستخدامها في اتخاذ القرارات ، وأدى التقدم المذهل في تكنولوجيا المعلومات إلى مساعدة الدارسين والباحثين ومتخذي القرارات فى الوصول إلى درجات عالية ومستويات متقدمة من التحليل ووصف الواقع ومتابعته ثم إلى التنبؤ بالمستقبل .

ولم تعد البحوث في شتى مجالاتها (الإنسانية،الإجتماعية،........إلخ، تكتفي بمجرد عرض المشاكل و دراسة الظواهر و تحديد الأسباب و استخلاص النتائج و اتخاذ القرارات بطريقة سطحية مجردة عن أسلوب الإقناع و التقدير والقياس .

ولقد أصبح الاتجاه العام في مثل هذه البحوث و الدراسات هو استخدام طرق القياس الكمية ووسائل الإقناع الإحصائية و ذلك لتحديد الخصائص وإبراز الاتجاهات العامة في الظواهر الاجتماعية والإدارية ، و تحليل العلاقات المتشابكة و المتبادلة بين الظواهر علي أساس موضوع غير متميز .

وتحليل البيانات (الإحصاء) يعطي للباحثين في مجال العلوم الإنسانية و الاجتماعية والإدارية ، العديد من الطرق والأساليب اللازمة لضرورة القيام بالدراسات والبحوث وتستخدم كلمة الإحصاء لتشير إلي عملية جمع البيانات الكمية و الأساليب المستعملة في معالجة تلك البيانات ، و قد نعني بهذه الكلمة أيضا عملية استخلاص بعض الاستنتاجات من دراسة عينة صغيرة لصياغة تعميمات يمكن تطبيقها علي مجتمعات اكبر حجما .

ومن هنا يتضح أن الإحصاء لا غنى عنه لأى باحث فى شتى المجالات المختلفة إذ اعتمد فى بحثه على الأسلوب العلمى. أي أن الإحصاء هوبمثابة عصى تقود الباحث إلى الطريق الصحيح الباحث وهى الأداة التى تساعده على تفسير الظواهر التى يدرسها وتوضيح النتائج التى يحصل عليها ودلالات البيانات والأرقام التى يحصل عليها .

د. جخدم موسى

الفصل الأول

تحليل البيانات (الإحصاء)

أولا : تعريف تحليل البيانات (الإحصاء) .

هو فرع من فروع الرياضيات يشمل النظريات و الطرق الموجهة نحو جمع البيانات ووصف البيانات و الاستقراء و صنع القرارات .

و عندما نتكلم عن الإحصاء لا نعنى بذلك البيانات الإحصائية وإنما نقصد حينئذ الطريقة الإحصائية . وهى الطريقة التى تمكننا من جميع الحقائق عن الظواهر المختلفة فى صورة قياسية رقمية وعرضها بيانيا ووضعها فى جداول تلخيصية بطريقة تسهل تحليلها بهدف معرفة اتجاهات هذه الظواهر وعلاقات بعضها ببعض .ولقد كان الهدف الرئيسى من الإحصاء قديما هو عد أو حصر الأشياء المراد توفير بيانات إحصائية عنها ، وكانت الجهة التى تقوم بإعداد الإحصاءات تعرف بمصلحة التعداد ولذلك كان التعريف القديم للإحصاء بأنه علم العد ، أي العلم الذي يشتمل على أساليب جمع البيانات الكمية عن المتغيرات والظواهر موضوع الدراسة .

ولكن مع تطور المجتمعات وتشابه جوانب الحياة المختلفة ، لم يعد مجرد توفير البيانات الكمية عن المتغيرات والظواهر موضوع الدراسة يفى بحاجات متخذى القرارات على وجه الخصوص إلى تكوين صورة متكاملة الجوانب عن مجتمعهم والمجتمعات المحيطة بهم . فقام العديد من العلماء بتحديث نظريات علم الإحصاء وأساليبه وأدواته لكى يعين الباحثين وغيرهم استخلاص استنتاجات معينة من البيانات الكمية التى أمكن لهم جمعها عن طريق العد .

ثانيا: أهمية تحليل البيانات (الإحصاء) .

لقد أصبح لعلم الإحصاء أهميه بالغة فى الحياة المعاصرة فصارت الإحصاءات مألوفة لدينا وتمثل جانبا مهما من المعلومات التى نطالعها كل يوم مثل نتائج الإنتخابات ،نتائح ومؤشرات البورصة،وغيرها

إن دراسة الإحصاء أمر له فوائد عديدة بالنسبة للدراسين في حقل العلوم الإنسانية وخاصة بعد أن تفتحت أمامهم مجالات عمل كثيرة فى تنظيمات وهياكل متعددة ومنوعة مثل: الشرطة والعلاقات العامة بالشركات ومراكز البحوث وغير ذلك من مجالات العمل المختلفة . بل إن المعرفة بالإحصاء قد تفيد الإنسان على المستوى الشخصي فتكسبه مهارة التخطيط لحياته الاقتصادية الخاصة خاصة في ظل الثورة الرقمية المتسارعة والسريعة التي يشهدها مجتمعنا اليوم.

ولكن ينبغي أن نشير إلى أن النتائج التى تسفر عن تطبيق أداة إحصائية أو أكثر ليست نتائج

قطعيه أو غير قابله للتمحيص والمراجعة . فإذا كانت الأدوات الإحصائية تستطيع أن تعين المرء على وصف البيانات وتصميم التجارب .وبعبارة أخرى ، فأن دور الأدوات الإحصائية لايقتصر على توفير المؤشرات المبدئية التى تساعد الباحث على رفض أو قبول الفروض التى يقوم بدراستها .والإحصاء أيضا أداه لا تستخدم إلا في العثور على إجابات عن أسئلة تتصل ببيانات يمكن التعبير عنها بصيغ كمية . وهناك فى مجال العلوم الانسانية موضوعات لا حصر لها لا يمكن صياغة البيانات الخاصة بها فى صورة كميه على نحو دقيق ، ومن ثم لا يستطيع الباحث استخدام التحليل الإحصائي فى دراستها .

ومما يعكس أهميه علم الإحصاء أنها يستخدم فى توجيه عمليه جمع البيانات وفى تفسير العلاقات التى تعكسها تلك البيانات . ومن ابرز المجالات التى تستخدم فيها المعالجات الإحصائية إجراء المقارنة بين عديد من الأشياء فى كثير من المناسبات . ويمكننا القول أن الحياة الإنسانية سلسلة من المواقف التى يتخذ فيها الفرد قراره بناء على ما تسفر عنه المقارنة التى يجريها بين عديد من الاحتمالات وهذه المقارنة فى جوهرها عمليه إحصائية تقترن بالقياس والتقييم والتقدير . فنجاح الإنسان فى حياته يتحدد وفق مقياس معين فى ذهنه يقدر به هذا النجاح ، وحرية الفرد فى مجتمعه تقاس أيضا وفق معايير يتعارف عليها الأفراد فى مجتمعهم .

ثالثا : تطور تحليل البيانات (الإحصاء) .

تطور علم الإحصاء وتطبيقاته عبر سنوات طويلة ، وتم ذلك بجهود كثيرة من العلماء من دول مختلفة وكان التطور بطيئا إلى أن جاء القرن العشرين ليشهد معدلا هائلا للتطور فى النظريات الإحصائية فى مجالات كثيرة .

ويرجع الاهتمام بالإحصاء إلى عصور قديمة ، ويبدو أن كلمه إحصاء (statistics) قد ظهرت لأول مره عام **1749** وهى مشتقة من الكلمة اللاتينية (status) أو الايطالية (statista) وتعني كلاهما الدولة السياسية . ومن الطبيعي أن تكون الدولة أول من اهتم بجمع البيانات وذلك لإدارة شؤون البلاد خاصة عن السكان لأغراض حربية وضريبية ، وامتدت بعد ذلك لتشمل إحصاءات حجم السكان والمواليد والوفيات والإنتاج.

ولقد تطور علم الإحصاء من مجرد فكره الحصر والعد إلى أن أصبح الآن علما له قواعده ونظرياته ويرجع الفضل فى ذلك إلى كثير من العلماء من أمثال عائله برونلي Bernoulli وفردريك جاوس F.gauss وكيتليه Quetlet وجولتونF.galton وأخيرا كارل بيرسون Karl.pearson وبولي A.bowley وبول U.yule فيشر L.fisher و........الخ .

وجاء التطور في علم الإحصاء بصفه عامه ملازما وموازيا للتطور في نظرية الاحتمالات . فقد نشأت نظريه الاحتمالات على أساس رياضى فى (1494) بواسطة باسيولي Lucapacidi . ومن الدراسات الفلكية لكل من كبلر (1517-1630) Keplr وجاليليو (1564-1642) Galilio قاما بتطوير نماذج الاحتمالات . غير أن التاريخ الحقيقي لنظريه الاحتمالات بدء فى القرن السابع عشر حيث وضعت أسسها في عام 1654 بواسطة كلا من العالمين : باسكان .Pascal,B (1623 – 1662) عالم الرياضيات والفيزياء والفيلسوف الفرنسى – وكذا العالم فرمات Fermat (1608 – 1665) .

وبعد ذلك بثلاث سنوات قام هينجينز Huygens (1629 – 1695) بنشر كتيب صغير في موضوع المعالجة الرياضية لفرص الفوز في مباريات ورق اللعب وزهرة النرد .

وفى نفس الوقت تقريبا قام جرونت grunt (1620 – 1674) بنشر ملاحظاته عن معالجة البيانات المتعلقة بالحكومة خاصة في النواحي الطبيعية والسياسية والتجارية والنمو والوفيات والأمراض.

وقد كان العمل الذى قام به هيجيتر دافعا للكثيرين لدراسة النظريات والمشاكل المتعلقة بمباريات الصدفة ومنهم برنوللى (1654 – 1705) ودي موافر De Moivre (1667 – 1754) واربوثنوت Arbuthnott ولابلاس laplace (1749 – 1827) وجاوس Gauss (1777 – 1855). [9] ويعد العالم البلجيكى كتيليه (1796 – 1874) أول من وضع قواعد محددة لعلم الإحصاء ، وكلمة إحصاء في الوقت الحاضر ذات معان متعددة فمنها يفهم جمع المعلومات التى تبين الحالة في الدولة مثل عدد المواليد والوفيات وبيانات عن المحاصيل والتجارة الخارجية ...... الخ ويسمى نشر الأجهزة الحكومية لمثل هذه المعلومات في شكل كتب وتقارير " **بالإحصاء الرسمي** ".

وكلمة إحصاء (Statistics) لها ثلاث معانى :

(1) الإحصاءات أو البيانات : مثال ذلك إحصاءات السكان والمواليد والوفيات والإنتاج – الصادرات – الاستهلاك .

(2) المؤشرات المحسوبة من عينة (العينة هى مجموعة جزئية من الوحدات محل الدراسة)

(3) علم الإحصاء : وهو فرع من فروع الرياضيات يشمل النظريات والطرق الموجهة نحو جميع البيانات ووصف البيانات والاستقراء وصنع القرارات .

ولقد تطور علم الإحصاء وتنوعت طرائقه ، وأصبح له من القواعد ما يمكنه من القيام كعلم مستقل يمكن الاستعانة به في رسم وتحديد السياسات الاجتماعية التى ينتهجها المجتمع. كما برز دور

الإحصاء – بما يقدمه من بيانات وإحصاءات – فى عمليات التخطيط والتنمية التي تمر بها مجتمعاتنا اليوم

ويمكن القول أن الإحصاء لا يكاد يخلو ميدان من ميادين البحث العلمى إلا وطرقته الإحصاء وساهمت فيه مساهمة فعالة . وقد أثار روبرت بارسوز فى مستهل كتابه " التحليل الإحصائي " أن كلمة إحصاء لها أكثر من استخدام إلا أن أكثر الاستخدامات شيوعاً هو ذلك الذى يرى أن كلمة إحصاء تشير إلى تلك الأساليب والإجراءات التحليلية المستخدمة في معالجة البيانات الرقمية.

بمعنى أنه للحصول علي معلومات ذات قيمة من تلك البيانات الرقمية فإنها يجب أن تخضع للتحليل الاحصائىStatistical Analysis بمساعدة تلك الأساليب والإجراءات والأدوات التي توفرها لنا الإحصاء.

ويذهب كل من Whittaker, Startup إلى وجود ثلاثة استخدامات لكلمة إحصاء :

أ- للإشارة إلى الحقائق الرقمية التى جمعت بطريقة منتظمة من الواقع الاجتماعي.

ب- تشير إلى الأساليب المستخدمة فى جمع ، وتصنيف وتحليل البيانات الرقمية.

ج- للإشارة إلى صفة أو خاصية للعينة تحت الدراسة.

الفصل الثاني

عرض البيانات ا لإحصائية ووصفها

أولا :عرض البيانات :

عادة ما تكون البيانات التي يتم جمعها غير منتظمة عددياً **(بيانات خام)**. ولذا يجب تنظيمها قبل بدء عملية التحليل الإحصائي. وتوجد ثلاثة أنواع مختلفة من هذه التوزيعات ويمكن استخدام إحداها حسب قيم المشاهدات المختلفة وتوزيعا التكراري إن وجد.

1- البيانات أو التوزيعات غير التكرارية أو غير المبوبة:

هي عبارة عن **قيم** (**أعداد**) **تتكون من فئة قليلة ومحدودة من المشاهدات وغالباً لا تتكرر**.

طريقة عرض هذه البيانات لا تحتاج إلى علاج إحصائي معين.

مثال توضيحي:

وجدنا القيم التالية ونريد التعامل معها إحصائيا

7 8 5 3 0 3 4 5 10

الحل:

يمكن التعامل معها كالآتي:

x	10	5	4	3	0	3	5	8	7

عادة الذي يستخدم في هذا النوع من البيانات هو أخذها كما هي، والتعامل معها .

2- البيانات أو التوزيعات التكرارية أو المبوبة :

عبارة عن ***قيم (أعداد) كثيرة، متكررة ومداها قصير أو صغير***. في هذه الحالة نستخدم عمودين: أحداهما لقيم المشاهدات (ترتب هذه القيم إما تصاعدياً أو تنازلياً) والعمود الآخر للتكرار.

مثال:

العلامات التالية تمثل طلبة الفوج :07 في مادة تحليل البيانات، تخصص اتصال وعلاقات عامة، خلال الفصل الدراسي 2022-2023:

11	12	20	12	00	19	19	18	18	10	18
18	10	11	10	10	18	11	10	18	16	11
18	10	18	18	12	10	12	18	11	14	12

اعرض هذه النتائج في جدول إحصائي؟

الحل:

بعد تفحص البيانات نرى أنها من النوع الثاني (كثيرة، متكررة، مداها قصير). نقوم

ببناء هذا التوزيع في جدول، ونرتب الدرجات ترتيباً تصاعدياً في العمود الأول (X)، أما عناصر العمود الثاني فتمثل عدد المرات التي تكررت بها كل درجة (f)، أما الدرجة التي لم تظهر في البيانات فتكرارها صفر.

الدرجات X	18	10	19	00	12	20	11	16	14
التكرار f	10	7	02	01	05	01	05	01	01

N= 36

3- البيانات أو التوزيعات التكرارية أو المبوبة:

إذا كان ***عدد البيانات كثيراً (متعددة ومتكررة) ومداها طويلاً أو كبيراً***، فإنه يجدر بنا في هذه الحالة أن نقسم قيم البيانات إلى فئات يتراوح عددها من 5 إلى 15 فئة حسب كون عدد البيانات صغيراً أو كبيراً. ولكن من الأصح يمكننا إيجاد عدد الفئات باستخدام الجذر التربيعي لعدد البيانات بحيث يقرب الناتج إلى اقرب رقم صحيح في حالة وجود كسور.

لبناء التوزيع التكراري نتبع الخطوات الآتية:

1.- نحسب مدى البيانات (R) بإحدى الطريقتين التاليتين:

الطريقة الأولى:

المدى = اكبر البيانات – اصغر البيانات + وحدة قياس

(وحدة القياس قيمتها = 1 في الأعداد الصحيحة، بينما في الكسور قيمتها تعتمد على الكسر)

$$R = P> - P< + a$$

الطريقة الثانية:

المدى = الحد الحقيقي العلوي لأكبر البيانات – الحد الحقيقي السفلي لأصغر البيانات

$$R = L+1 - L-1$$

الحدود الحقيقية أو الفعلية:

إذا نظرنا إلى تعريف المتغير المتصل، نلاحظ انه توجد قيم غير محدودة بين قيمتين. على سبيل المثال: إذا كانت أوزان طلبة قسم علوم الإعلام والاتصال بكلية العلوم

الإنسانية والإسلامية والحضارة بجامعة عمار ثليجي بالأغواط–الجزائر. سجلت إلى اقرب Kg فإن فترة (مدى) الفئة 70-60 تتضمن من الناحية النظرية كل القياسات من 59,5Kg – 70,5Kg

هذه الأرقام تسمى بالحدود الحقيقية أو الفعلية للفئة. الرقم الأصغر 59,5 هو الحد الأدنى الحقيقي أو الفعلي للفئة والرقم 70,5 هو الحد الأعلى الحقيقي أو الفعلي للفئة.

من الناحية العملية فإن الحدود الحقيقية للفئة يمكن الحصول عليها بجمع الحد الأعلى (النظري أو الظاهري) لفترة فئة والحد الأدنى (النظري أو الظاهري) لفترة الفئة التالية لها والقسمة على 2 .

يعني

60-70

50-59

40-49

للحصول على الحد الحقيقي الأعلى للفئة الأولى نقوم بجمع 50+49 = 99 ونقسمه على 2 فنحصل على الناتج 49,5 ونفس القيمة تكون الحد الحقيقي الأدنى للفئة التي تليها.

2.- نحدد عدد الفئات باستخدام الجذر التربيعي لمجموع التكرار، ويتم تقريب الناتج في حالة وجود كسور إلى اقرب عدد صحيح *(تقريب رقم مثل 44,8 إلى اقرب رقم عشري هو 45 حيث أن 44,8 اقرب إلى 45* منها إلى 44 . كذلك فإن تقريب الرقم 44,9146 إلى اقرب رقم مئوي أو إلى رقمين عشريين هو 44,91 حيث أن 44,9146 اقرب إلى 44,91 منها إلى 44,92 . في تقريب رقم مثل 44,365 إلى اقرب رقم مئوي تصادفنا صعوبة حيث أن الرقم 44,365 في نفس درجة البعد عن الرقمين 44,36 ، 44,37 وقد اصطلح من الناحية العملية أن يتم في هذه الحالات التقريب إلى الرقم الزوجي السابق على 5 . مثال ذلك 44,365 تقرب إلى 44,36 ، 44,375 تقرب إلى 44,38) .

3.- نحسب طول الفئة (i) بقسمة المدى (R) على عدد الفئات ثم التقريب دائماً إلى أعلى في حالة وجود كسور.

طول الفئة (i) = المدى ÷ عدد الفئات

4.- نتجه للجدول ونضع كحد أدنى للفئة الأولى اصغر قيمة في البيانات، ثم نضيف طول الفئة عليها لكي نحصل على الحد الأعلى إلى هذه الفئة، ثم ننتقل إلى الفئة التالية ونضيف واحد للحد الأدنى وبعد ذلك طول الفئة، ......الخ. ونكرر هذه العملية حتى نحصل على جميع الفئات.

5.- نسجل تكرارات كل فئة أمامها في عمود التكرارات (f) .

6.- نجمع التكرارات لجميع الفئات (N) .

7.- نحسب التوزيع التكراري المتجمع (fa) وهو عبارة عن جمع التكرارات المقابلة لجميع القيم الأقل من الحد الأعلى الحقيقي لفئة معينة ويتضمن تكرارها أيضا. وهناك نوعان من التوزيع التكراري المتجمع هما: التصاعدي والتنازلي.

8.- نحسب التكرار النسبي (p) لكل فئة وهو عبارة عن نسبة تكرار تلك الفئة إلى مجموع التكرارات.

9.- لحساب التوزيع التكراري النسبي المتجمع (pa) نستخدم نفس الطريقة رقم 7 . ونلاحظ أن مجموع التكرارت النسبية يجب أن يساوي 1 .

10.- نحسب التوزيع التكراري المئوي (P) وهو عبارة عن التكرار النسب مضروباً في 100 .

$$P = p \cdot 100 \qquad \text{or} \qquad P = (f / N) \cdot 100$$

11.- لحساب التوزيع التكراري المئوي المتجمع (Pa) نستخدم نفس الطريقة رقم 7 . ونلاحظ أن مجموع التكرارت المئوية يجب أن يساوي 100 .

12.- نجد مركز كل فئة (Xm) بقسمة مجموع حديها على 2 .

$$Xm = \frac{L+1 \quad + \quad L-1}{2}$$

هناك طريقة أخرى هي:

$$Xm = L-1 \; + \; \frac{i}{2}$$

13.- نحسب الحدود الحقيقية لكل الفئات وذلك بطرح نصف وحدة من الحد الأدنى النظري (الظاهري) للفئة الأولى وبذلك نحصل على الحد الأدنى الحقيقي أو الفعلي للفئة الأولى، وللحصول على

الحد الأعلى الحقيقي أو الفعلي لنفس الفئة نقوم بإضافة نصف وحدة. ونكرر هذا مع باقي الفئات.

مثال:

البيانات التالية تمثل أوزان مختلفة لفوجين من طلبة قسم علوم الإعلام والاتصال بكلية العلوم الإنسانية والإسلامية والحضارة بجامعة عمار ثليجي بالأغواط-الجزائر. على النحو التالي:

77 96 78 87 60 90 91 90 90 91 96 87 80 80 76 80 90 91 97 75 90 60
81 79 91 71 87 78 93 91 75 96 60 100 93 90 75 91 80 70 80 87 95 96
90 85 85 70 60 93 85 80 70 85 80 95 60 70 60
92 90 80 74 87 86 80 95 92 97 75 100 93 67 90 90 78 92 96 90 85 84

- حدد نوع التوزيع ولماذا؟

- إذا كان هذا التوزيع من النوع الثالث اجب عن الأسئلة الآتية؟

1.- اكبر درجة 2.- اقل درجة 3.- المدى 4.- عدد الفئات 5.- طول الفئات

6.- تكوين الفئات بالحدود الظاهرية أو النظرية 7.- تكرار كل فئة

8.- مجموع التكرار الكلي 9.- التكرار المتجمع التصاعدي والتنازلي

10.- التكرار النسبي 12.- التكرار النسبي المتجمع 13.- التكرار المئوي

14.- التكرار المئوي المتجمع 15.- مراكز الفئات 16.- الحدود النظرية

17.- الحدود الحقيقية أو الفعلية

الحل:

نوع التوزيع تكراري من النوع الثالث لان عدد الدرجات كثيرة ومتكررة ومداها كبيراً.

1.- اكبر درجة هي 100

2.- اقل درجة هي 60

3.- المدى = 100 − 60 + 1 = 41 أو المدى = 100,5 − 59,5 = 41

4.- عدد الفئات = الجذر التربيعي للمجموع = 81 = 9

5.- طول الفئات = 9 / 41 = 4,55 = 5 (دائماً في حالة العمل مع أعداد صحيحة نقوم بالتقريب إلى اكبر عدد صحيح).

X	f	fa	p	pa	P	Pa	Xm	L_{-1} + L_{+1}
100 -104	2	**81**	0,03	**1**	3	**100**	102	99,5-104,5
95 - 99	10	79	0,12	0,97	12	97	97	94,5-99,5
90 - 94	24	69	0,30	0,85	30	85	92	89,5-94,5
85- 89	11	45	0,14	0,55	14	55	87	84,5-89,5
80 - 84	11	34	0,14	0,41	14	41	82	79,5-84,5
75 - 79	10	23	0,12	0,27	12	27	77	74,5-79,5
70 - 74	6	13	0,07	0,15	7	15	72	69,5-74,5
65 - 69	1	7	0,01	0,08	1	8	67	64,5-69,5
60 - 64	6	6	0,07	0,07	7	7	62	59,5-64,5

N = 81

البيانات التي تحتوي على كسور:

الحدود الحقيقية أو الفعلية:

إذا كانت البيانات تحتوي على كسور ونريد أن نجد الحدود الحقيقية لها. يعني نجد بعد الفاصلة كسر واحد فقط، مثلاً: 7,8

هذه القيمة تحتوي على كسر واحد فقط ولذلك يجب إضافة 0,05 إلى 7,8 لكي نحصل على الحد الحقيقي العلوي، وللحصول على الحد الأدنى الحقيقي يجب طرح 0,05 من 7,8 فتصبح الحدود كالآتي:

الحد الحقيقي أو الفعلي الأدنى		الحد الحقيقي أو الفعلي العلوي
7,75	**7,8**	**7,85**

طول الفئة:

إذا وجدنا أكثر من كسر، فالإضافة أو الطرح حسب عدد الكسور، مثلاً: 12,3479

للحصول على الحد الحقيقي العلوي يجب إضافة 0,00005 إلى 12,3479 ونطرح 0,00005 من هذه القيمة، فتصبح الحدود الفعلية أو الحقيقية كالآتي:

الحد الحقيقي أو الفعلي الأدنى		الحد الحقيقي أو الفعلي العلوي
12,34785	**12,3479**	**12,34795**

التمثيل البياني :

يوجد تمثيل بياني لكل مستوى قياس. فمثلاً عندما تكون المتغيرات مقاسه بمقياس التصنيف والترتيب نستخدم لتمثيلها بيانياً الأعمدة البيانية (**BAR CHARTS**) والرسوم الدائرية (**PIE CHARTS**)، بينما إذا كانت مقاسه بمقياس الفئات أو النسبة فيتم استخدام المضلع التكراري (**FREQUENCY POLYGON**) والمدرج التكراري (**FREQUENCY HISTOGRAM**) حسب شروط معينة.

في دراستنا للعرض (التمثيل) البياني سنتناول أنواع الرسوم البيانية في حالة القيم غير المبوبة ثم نتناول الرسوم البيانية في حالة القيم المبوبة (التوزيعات التكرارية).

1.- الرسوم البيانية في حالة القيم الغير مبوبة:

هذه الرسوم يمكن تمثيلها بيانياً إذا كانت مقاسه بمقياس التصنيف أو الترتيب. من أهم أنواعها:

1.- الأعمدة البيانية:

أكثر الطرق انتشاراً وسهلة الرسم والفهم. ووفقاً لهذه الطريقة نبدأ برسم محورين متعامدين، وتمثل المتغيرات بقواعد متساوية والمسافات بينها يجب أن تكون متساوية في المحور الأفقي أو الاحداثي السيني (ABSCISSA) بينما نمثل التكرارات على المحور الرأسي أو الاحداثي الصادي (ORDINATE) وتقسم تقسيماً ملائماً.

2.- الرسوم الدائرية:

عبارة عن تمثيل أو رسم دائري. نقسم الدائرة إلى قطاعات أو عناصر جزئية بعدد مكونات الظاهرة ويستخدم المنطق التالي: حيث أن الزاوية المركزية تساوي 360 درجة وهي تمثل 100% من مساحة الدائرة، إذن فأن 1% من مساحة الدائرة يمثل عن طريق زاوية مقدارها 3,6 درجة، وعلى هذا الأساس يمكن رسم الزوايا المختلفة داخل الدائرة.

مثال:

تم عمل استقصاء في جامعة عمار ثليجي بالأغواط–الجزائر– لمعرفة جنسية الوافدين للجامعة، ووجدنا النتائج الآتية:

الصحراء الغربية = 50 موريتانيا = 15 اليمن = 25

فلسطينيين = 23 آخرين = 17

نريد تمثيل هذه البيانات بيانياً.

الحل:

يمكننا استخدام الأعمدة البيانية أو الرسوم الدائرية لأن هذه البيانات مقاسه بمقياس التصنيف.

II.– الرسوم البيانية في حالة القيم المبوبة أو تمثيل التوزيعات التكرارية بيانياً:

لكي نستطيع تمثيلها بيانياً يجب أن تكون مقاسه إما في مستوى الفئات أو مقياس النسبة. هناك طريقتان للتعبير عن هذه التوزيعات:

1.- المضلع التكراري:

هو خط بياني لتكرار الفئة المقابلة لمركز الفئة. ويمكن رسمه بوضع قيمة مركز كل فئة (Xm) في المحور الأفقي أو الاحداثي السيني (ABSCISSA) وفي المحور الرأسي أو الاحداثي الصادي (ORDINATE) تكرار كل فئة (f) .

2.- المدرج التكراري:

لرسمه بيانياً نضع الحدود الحقيقية لكل الفئات في الاحداثي السيني وفي الاحداثي الصادي تكرار كل فئة (f) .

إذا كانت كل الفئات متساوية من حيث الطول فانه من المعتاد أن نأخذ الارتفاعات مساوية لتكرار الفئات. أما إذا كانت الفئات غير متساوية الطول فان هذه الأطوال يجب أن تعدل. وذلك باستخدام حساب الارتفاعات (المساحات) ويمكن حسابها بالطريقة الآتية:

مساحة كل فترة فئة (h) تساوي تكرار الفئة مقسوماً على طول هذه الفئة (i) .

$$h = \frac{f}{i}$$

وفي ختام هذا الفصل هذا هو جدول الرموز العلمية الأكثر استخداماً في علم الإحصاء والمتعلقة بكل ماتم ذكره.

المصطلح بالعربية	الرمز المستخدم	المصطلح بالإنجليزية
فئات قيم المشاهدات أو القياسات	X	INTERVAL
التكرارات المقابلة لهذه الفئات أو القياسات	f	FREQUENCY
التوزيع التكراري المتجمع	fa	CUMULATIVE FREQUENCY DISTRIBUTION
مركز الفئة	Xm	CLASS MIDPOINT (MARK)
التكرارات النسبية	p	RELATIVE FREQUENCIES
التكرارات النسبية المتجمعة	pa	CUMULATIVE FREQUENCIES RELATIVE
التكرارات المئوية	P	PORCENTAGE FREQUENCY
التكرارات المئوية المتجمعة	Pa	CUMULATIVE FREQUENCY PORCENTAGE
مجموع التكرارات أو التكرار الكلي	N	TOTAL
عدد البيانات	n	NUMBER OF DATA
الحد الحقيقي أو الفعلي الأدنى أو الأسفل	L-1	LOWER CLASS BOUNDRAY
الحد الحقيقي أو الفعلي العلوي أو الأكبر	L+1	UPPER CLASS BOUNDRAY
مدى أو طول كل فئة	i	CLASS SIZE (WIDTH)
المدى	R	RANGE

TYPE (I)	Type I	البيانات أو التوزيعات غير التكرارية (النوع الأول)
TYPE (II)	Type II	البيانات أو التوزيعات التكرارية أو المبوبة (النوع الثاني)
TYPE (III)	Type III	البيانات أو التوزيعات التكرارية أو المبوبة (النوع الثالث)
HIGH POINT	P>	اكبر قيمة أو عدد
DOWN POINT	P<	اصغر قيمة أو عدد
	a	وحدة قياس

الفصل الثالث

مقاييس النزعة المركزية

أول هذه المقاييس التي نبدأ بها هي مقاييس النزعة المركزية: المنوال (MODE) والوسيط (MEDIAN) والوسط (MEAN).

1.- **المنوال (MODE):** هو عبارة عن القيمة التي تتكرر أكثر من غيرها أو الأكثر شيوعاً. ونرمز لها Mo . المنوال من السهل حسابه، لكنه أيضا غير دقيق ومن هنا الأقل استخداماً.

كيفية حساب قيمة المنوال غير المصحح أو غير المعدل *البيانات غير المبوبة:*

أمثلة:

1.- احسب المنوال للبيانات التالية:

11	**1**	**2**	**3**	**4**	**5**	**7**	**10**	**9**	**2**

نلاحظ أن القيمة الأكثر تكراراً هي 0 إذن هي المنوال، ولذلك نقول:

Mo = 2 (Unimodal)

وحيد المنوال

2.- احسب المنوال للبيانات التالية:

10	**11**	**7**	**10**	**1**	**2**	**3**	**2**	**5**	**8**

نلاحظ أن القيم الأكثر تكراراً هي 2 و 10 إذن هاتان القيمتان هما المنوال، ولذلك نقول:

Mo = 2 & Mo= 10 (Bimodal)

مزدوج المنوال

3.- احسب المنوال للبيانات التالية:

3	**1**	**3**	**7**	**5**	**6**	**7**	**1**	**11**	**8**

نلاحظ أن القيم الأكثر تكراراً هي 3 و 1 و 7 إذن هذه القيم هي المنوال، ولذلك نقول:

Mo = 3 & Mo= 1 & Mo = 7 (Trimodal)

ثلاثي المنوال

البيانات المبوبة (النوع الثاني) : المنوال هو القيمة التي يقابلها اكبر تكرار

مثال:

إذا كانت هذه الأعداد تمثل علامات طلبة قسم علوم الإعلام والاتصال في مقياس تحليل البيانات: (انظر الجدول).

المنوال هو القيمة التي يقابلها اكبر تكرار، يعني العلامة 10 هي المنوال لأن تكرارها اكبر التكرارات.

نقول أن هذه العلامات وحيدة المنوال = 13Mo (Unimodal)

X	14	10	20	18	5	7	15	12	0
f	7	13	1	6	4	8	11	9	2

ملاحظة: يمكننا أن نجد نفس الحالات كما هو الحال في البيانات غير المبوبة

البيانات المبوبة (النوع الثالث): المنوال هو مركز الفئة (Xm) التي تمتلك اكبر تكرار

مثال:

X	20-10	30-20	40-30	50-40	60-50	70-60	80-70	90-80	100-
f	7	13	1	6	4	18	11	19	21

الفئة التي تمتلك اكبر تكرار هي 80 – 90 فنقوم بإيجاد مركزها ويكون الناتج هو المنوال.

هكذا:

$$Mo = \frac{80 + 90}{2} = 85 \quad \text{(Unimodal)}$$

كيفية حساب قيمة المنوال المصحح أو المعدل:

1.– البيانات المبوبة من النوع الثالث التي تحتوي على فئات متساوية.

X	60-64	65-69	70-74	75-79	80-84	85-89	90-94	95-99	104
f	6	1	6	10	11	11	24	10	2

N= 102

الحل:

1.– نقوم بتحديد الفئة التي يقابلها اكبر تكرار وهذه الفئة تسمى الفئة المنوالية.

2.- نستخدم الصيغة الآتية:

$$Mo = L\text{-}1 + i \quad \frac{f+1}{(f+1) + (f\text{-}1)}$$

حيث:

التكرار الذي يلي الفئة المنوالية $f+1$

التكرار الذي يسبق الفئة المنوالية $f-1$

3.– نقوم بالتعويض في الصيغة السابقة لكي نحصل على قيمة المنوال المصحح أو المعدل:

$$Mo = 89,5 + 5 \frac{10}{10 + 11} = 91,88$$

2.– *البيانات المبوبة من النوع الثالث التي لا تحتوي على فئات متساوية.*

14–16	17–20	21–24	25–30	31–40	X
5	7	1	10	2	f
1.67	1.75	0.25	1.67	0.20	H

N=25

الحل:

1.– نقوم بحساب الارتفاعات (h) لكل الفئات بالطريقة الآتية:

$$h = \frac{f}{i}$$

2.– نقوم بتحديد الفئة التي يقابلها اكبر ارتفاع وهذه الفئة تسمى الفئة المنوالية.

3.- نستخدم الصيغة الآتية:

$$Mo = L\text{-}1 + i \frac{h+1}{(h+1) + (h-1)}$$

حيث:

الارتفاع الذي يلي الفئة المنوالية $h+1$

الارتفاع الذي يسبق الفئة المنوالية $h-1$

4.– نقوم بالتعويض في الصيغة السابقة لكي نحصل على قيمة المنوال المصحح أو المعدل:

$$Mo = 16,5 + 4 \frac{0,25}{0,25 + 1,67} = 17,02$$

2.– الوسيط (Median): هو القيمة التي يسبقها نصف البيانات ويليها النصف الآخر. أو يمكن تعريفه بأنه القيمة التي تقسم القيم إلى جزئيين بحيث يكون عدد القيم التي أقل منها يساوي عدد القيم التي أكبر منها. فالوسيط إذن هو متوسط ترتيب القيم. ونرمز له **Mdn**

البيانات الغير مبوبة (النوع الأول):

.– إذا كان عدد الحالات فردياً: عند معرفة مكانة الوسيط، نجد هذه المكانة قيمتها لا تتكرر.

مثال: إذا كانت الأعداد الآتية تمثل أعمار سبعة طلاب قسم علوم الإعلام والاتصال تخصص اتصال وعلاقات عامة للعام الدراسي 2023/2022:

18, 19, 21, 20, 23, 22, 24

فالطريقة المستخدمة لحساب وسيطها هي كالتالي:

1.– نرتب الأعداد إما تصاعدياً أو تنازلياً.

18 19 20 21 22 23 24

2.– نحدد مكانة أو ترتيب الوسيط.

$$Pmdn = \frac{n+1}{2} = \frac{7+1}{2} = 4$$

3.– نبحث عن هذه المكانة بين الأعداد.

18 19 20 21 22 23 24

4.– بعد ذلك نطرح السؤالين الآتيين:

– هل هذه القيمة تتكرر ؟. في هذا المثال لا تتكرر

– هل يوجد نفس الوحدات على يمين هذه القيمة وعلى يسارها؟. بالطبع في مثالنا نعم.

5.– إذن الوسيط هو القيمة 21

2.– عند معرفة مكانة الوسيط، نجد هذه المكانة قيمتها تتكرر.

مثال:

إذا كانت الأعداد الآتية تمثل أعمار سبعة طلاب قسم علوم الإعلام والاتصال تخصص اتصال وعلاقات عامة للعام الدراسي 2023/2022:

18, 19, 21, 20, 23, 22, 21

فالطريقة المستخدمة لحساب وسيطها هي كالتالي:

1.– نرتب الأعداد إما تصاعدياً أو تنازلياً.

18 19 20 21 21 22 23

2.– نحدد مكانة أو ترتيب الوسيط.

$$Pmdn = \frac{n+1}{2} = \frac{7+1}{2} = 4$$

3.– نبحث عن هذه المكانة بين الأعداد.

18 19 20 21 21 22 23

4.– بعد ذلك نطرح السؤالين الآتيين:

– هل هذه القيمة تتكرر ؟. في هذا المثال نعم تتكرر

– هل يوجد نفس الوحدات على يمين هذه القيمة وعلى يسارها؟. بالطبع في مثالنا نعم.

5.–كما نعرف أن الحدود الحقيقية للقيمة 21 هي 21,5 – 20,5 وان طول هذه الفئة هو 1.

كما نرى أن عدد تكرار هذه المكانة هو 2 ولذلك يجب تقسيم طول الفئة بين هاتين القيمتين كالآتي:

18 19 20 21 21 22 23

6.– الوسيط في هذه الحالة هو

$$Mdn = \frac{20,5 + 21}{} = 20,75$$

إذا كان عدد الحالات زوجياً:

1.- عند معرفة مكانة الوسيط، نجد هذه المكانة قيمتها تقع بين قيمتين مختلفتين فيما بينهما.

مثال:

إذا كانت الأعداد الآتية تمثل أعمار سبعة طلاب قسم علوم الإعلام والاتصال تخصص اتصال وعلاقات عامة للعام الدراسي 2023/2022:

18, 19, 21, 20, 23, 22, 24, 25

فالطريقة المستخدمة لحساب وسيطها هي كالتالي:

1.- نرتب الأعداد إما تصاعدياً أو تنازلياً.

18 19 20 21 22 23 24 25

2.- نحدد مكانة أو ترتيب الوسيط.

$$Pmdn = \frac{n+1}{2} = \frac{8+1}{2} = 4,5$$

3.- نبحث عن هذه المكانة بين الأعداد.

18 19 20 21 22 23 24 25

4.- بعد ذلك نطرح السؤالين الآتيين:

- هل هذه القيمة تتكرر ؟. في هذا المثال لا تتكرر
- هل يوجد نفس الوحدات على يمين هذه القيمة وعلى يسارها؟. بالطبع في مثالنا نعم.

5.- إذن الوسيط هو الوسط الحسابي بين هاتين القيمتين

$$Mdn = \frac{21+22}{2} = 21,5$$

.- عند معرفة مكانة الوسيط، نجد هذه المكانة قيمتها تقع بين قيمتين متساويتين فيما بينهما.

مثال:

إذا كانت الأعداد الآتية تمثل أعمار سبعة طلاب قسم علوم الإعلام والاتصال تخصص اتصال وعلاقات عامة للعام الدراسي 2023/2022:

18, 19, 22, 20, 23, 22, 24, 25

فالطريقة المستخدمة لحساب وسيطها هي كالتالي:

1.- نرتب الأعداد إما تصاعدياً أو تنازلياً.

18 19 20 22 22 23 24 25

2.- نحدد مكانة أو ترتيب الوسيط.

$$Pmdn = \frac{n+1}{2} = \frac{8+1}{2} = 4,5$$

3.- نبحث عن هذه المكانة بين الأعداد.

18 19 20 22 22 23 24 25

4.- بعد ذلك نطرح السؤالين الآتيين:

- هل هذه القيمة تتكرر ؟. في هذا المثال تتكرر

- هل يوجد نفس الوحدات على يمين هذه القيمة وعلى يسارها؟. بالطبع في مثالنا نعم.

5.- إذن الوسيط هو هذه القيمة

$$Mdn = \frac{22+22}{2} = 22$$

.- عند معرفة مكانة الوسيط، نجد هذه المكانة قيمتها تقع بين قيمتين متساويتين فيما بينهما وكذلك متساوين مع عدد آخر أو أعداد أخرى.

مثال:

إذا كانت الأعداد الآتية تمثل أعمار سبعة طلاب قسم علوم الإعلام والاتصال تخصص اتصال وعلاقات عامة للعام الدراسي 2023/2022:

21 ،18, 19, 20, 20, 23, 22, 20

فالطريقة المستخدمة لحساب وسيطها هي كالتالي:

1.- نرتب الأعداد إما تصاعدياً أو تنازلياً.

18 19 20 20 20 21 22 23

2.- نحدد مكانة أو ترتيب الوسيط.

$$Pmdn = \frac{n+1}{2} = \frac{8+1}{2} = 4,5$$

3.- نبحث عن هذه المكانة بين الأعداد.

18 19 20 20 20 21 22 23

4.– بعد ذلك نطرح السؤالين الآتيين:

– هل هذه القيمة تتكرر ؟. في هذا المثال نعم تتكرر

– هل يوجد نفس الوحدات على يمين هذه القيمة وعلى يسارها؟. بالطبع في مثالنا نعم.

5.–كما نعرف أن الحدود الحقيقية للقيمة 20 هي 20,5 – 19,5 وان طول هذه الفئة هو 1.

كما نرى أن عدد تكرار هذه المكانة هو 3 ولذلك يجب تقسيم طول الفئة بين القيم الثلاثة كالآتي:

1/3 = 0,33

18 19 20 20 20 21 22 23

6.– الوسيط في هذه الحالة هو

Mdn = 20,16

البيانات مبوبة (النوع الثاني) :

X	f	fa
100	2	36
99	0	34
98	0	34
97	2	34
96	5	32
95	3	27
94	0	24
93	4	24
92	3	20
91	6	17
90	11	11

N=36

1.– نقوم بإيجاد التكرار المتجمع التصاعدي.

2.– نحدد مكانة الوسيط كالمعتاد

$$Pmdn = \frac{N}{2} = \frac{36}{2} = 18$$

3.– نحدد مكانة الوسيط في التكرار المتجمع.

4.– الوسيط هو الفئة التي تقابل مكانة الوسيط. في هذه الحالة قيمة الوسيط هي 92

البيانات مبوبة او التوزيعات تكرارية (النوع الثالث) :

X	f	fa
100-104	2	81
95-99	10	79
90-94	24	69
85-89	11	45
80-84	11	34
75-79	10	23
70-74	6	13
65-69	1	7
60-64	6	6

1.– نقوم بإيجاد التكرار المتجمع التصاعدي.

2.– نحدد مكانة الوسيط كالمعتاد

$$Pmdn = \frac{N}{2} = \frac{81}{2} = 40,5$$

3.– نحدد مكانة الوسيط في التكرار المتجمع.

4.– نحسب قيمة الوسيط بالصيغة الآتية:

$$Mdn = L_{-1} + \left(i \; \frac{N/2 - fai}{fb} \right)$$

$$Mdn = 84,5 + 5 \; \frac{40,5 - 34}{11} = 87,45$$

حالات خاصة لحساب الوسيط (Median):

.- نجد مكانة الوسيط تطابق قيمة التكرار المتجمع. في هذه الحالة قيمة الوسيط هي الحد الحقيقي العلوي للفئة الموجود بها التكرار المتجمع. لنرى مثالاً:

X	f	Fa
100-104	11	90
95-99	10	79
90-94	24	69
85-89	11	45
80-84	11	34
75-79	10	23
70-74	6	13
65-69	1	7
60-64	6	6

1.- نقوم بإيجاد التكرار المتجمع التصاعدي.

2.- نحدد مكانة الوسيط كالمعتاد

$$Pmdn = \frac{N}{2} = \frac{90}{2} = 45$$

3.- نحدد مكانة الوسيط في التكرار المتجمع، ونجد أن هذه المكانة تتفق (تساوي بالضبط) مع التكرار المتجمع.

4.- في هذه الحالة قيمة الوسيط هي الحد الحقيقي العلوي للفئة الموجود بها التكرار المتجمع، يعني 89,5

توضيح للإجابة: واضح أن القيمة التي تترك 50% من الحالات هي الحد الحقيقي العلوي للفئة التي يوجد بها التكرار المتجمع، في هذه الحالة الوسيط = 89,5

.- نجد مكانة الوسيط تطابق قيمة التكرار المتجمع و إضافة على ذلك أن الفئة التالية تكرارها صفر. في هذه الحالة قيمة الوسيط هي مركز الفئة التي تكرارها صفر.

X	f	fa
100-	11	90
104	34	79
95-99	0	45
90-94	11	45
85-89	11	34
80-84	10	23
75-79	6	13
70-74	1	7
65-69	6	6
60-64		

1.- نقوم بإيجاد التكرار المتجمع التصاعدي.

2.- نحدد مكانة الوسيط كالمعتاد

$$Pmdn = \frac{N}{2} = \frac{90}{2} = 45$$

3.- نحدد مكانة الوسيط في التكرار المتجمع، ونجد أن هذه المكانة تتفق (تساوي بالضبط) مع قيمتين في التكرار المتجمع.

4.- في هذه الحالة قيمة الوسيط هي مركز الفئة التي تكرارها صفر، يعني 92

توضيح للإجابة: بإمكاننا أن نفكر أن الحد العلوي الحقيقي 94,5 والحد العلوي الحقيقي 89,5 يقسموا القيم إلى قسمين أو بمعنى آخر يتركوا 50% إلى أعلى والى أسفل، وحتى كل قيمة مشمولة بين هاتين القيمتين تكمل تعريف الوسيط.

ملاحظة هامة:

يمكن حل السؤال السابق أيضا بإحدى الطرق التالية:

1.- نحسب الوسيط لكلا القيمتين ثم بعد ذلك نأخذ متوسطه، ثم نضيف إليه نصف طول الفئة التي تكرارها يساوي صفر.

2.- نأخذ الحد العلوي الحقيقي للفئة الأولى التي بها مكانة الوسيط تتفق مع التكرار المتجمع، وكذلك

الحد الحقيقي العلوي للفئة الثانية والتي بها مكانة الوسيط تتفق أيضا مع التكرار المتجمع، وبعد ذلك نحسب المتوسط، ويكون الناتج هو قيمة الوسيط.

3.– الوسط الحسابي (ARITHMETIC MEAN): هو الإحصائي الذي يأخذ في الاعتبار جميع البيانات. أكثر مقاييس النزعة المركزية استخداماً وأهمية. ونرمز له **X**

البيانات غير مبوبة (النوع الأول)

مثال:

احسب الوسط الحسابي لما يلي:

2 1 8 5

لحساب الوسط الحسابي نقوم بجمع جميع البيانات ونقسمها على عددها. للاختصار في الكتابة نستعمل الحرف اليوناني Σ ونعني به جمع الحدود التي في داخله.

$$X = \frac{\Sigma X}{n} = \frac{2+1+8+5}{4} = 4$$

البيانات مبوبة (النوع الثاني)

X	f	f. Xm
100	2	200
99	0	0
98	0	0
97	2	194
96	5	480
95	3	285
94	0	0
93	4	372
92	3	276
91	6	546
90	11	990

N=36 Σ=3343

نقوم بحساب الوسط الحسابي مستخدمين الصيغة الآتية:

$$X = \frac{\sum f . Xm}{N} = \frac{3343}{36} = 92{,}86$$

البيانات مبوبة أو التوزيعات التكرارية (النوع الثالث)

حساب الوسط الحسابي بالطريقة المباشرة:

$$\mathbf{X = \frac{\sum f . Xm}{N}}$$

X	f	Xm	f . Xm
100-104	2	102	204
95-99	10	97	970
90-94	24	92	2208
85-89	11	87	957
80-84	11	82	902
75-79	10	77	770
70-74	6	72	432
65-69	1	67	67
60-64	6	62	372
	N= 81		Σ = 6882

$$X = \frac{6882}{81} = 84{,}96$$

أوساط حسابية خاصة:

1.– الوسط الحسابي للأوساط أو الوسط الحسابي الكلي: عندما يكون عندنا مجموعة عينات ويهمنا معرفة الوسط الحسابي الكلي لها في اختبار معين. يمكننا أن نجد احد الاحتمالين:

1.- ***عدد الأشخاص في كل مجموعة متساوٍ***. للحصول على الوسط الحسابي الكلي نقوم بجمع كل الأوساط الحسابية للعينات ونقسمها على عددها.

2.- ***عدد الأشخاص في كل مجموعة غيرمتساوٍ***. للحصول على الوسط الحسابي الكلي نقوم بضرب كل وسط حسابي في عينته الخاصة به ثم نجمعهم وبعد ذلك نقسم المجموع على مجموع العينات، كالآتي:

$$Xt = \frac{N1.X1 + N2.X2 + \ldots\ldots\ldots + Nn.Xn}{N1 + N2 + \ldots\ldots\ldots + Nn}$$

2.– الوسط الحسابي المئوي: طريقة الحل نفس الطريقة السابقة والفرق أننا نجد البيانات مستخدمة النسبة المئوية.

3.– الوسط الحسابي المرجح:

نستخدم هذا الوسط عندما نقرن بيانات الشخص $X1, X2, X3, \ldots, Xn$ بمعاملات أو أوزان ترجيحية $W1, W2, W3, \ldots, Wn$ وهذة الترجيحات تعتمد على الدلالة أو الأهمية المرتبطة بهذه البيانات.

$$XW = \frac{W1.X1 + W2.X2 + \ldots + Wn.Xn}{W1 + W2 + \ldots + Wn}$$

خصائص واعتبارات مقاييس النزعة المركزية

1.- عند حساب الوسط الحسابي (X) نأخذ في الحسبان كل البيانات، بينما الوسيط (Mdn) يهتم فقط بالأماكن المركزية، وعند المنوال (Mo) نركز على القيم الأكثر تكراراً.

2.- عند وجود البيانات غير المبوبة من النوع الأول ، الوسط الحسابي (X) هو الأكثر ملائمة عندما تكون كل البيانات قريبة من بعضها، لكن إذا حصل العكس (وجود احد البيانات أو أكثر بعيداً نسبياً عن الباقي) فإن الوسيط (Mdn) هو الأكثر ملائمة.

3.- عند وجود بيانات مبوبة من النوع الثاني أو الثالث ، فالوسط الحسابي (X) هو الملائم إذا توفرت 3 شروط مجتمعة:

- التوزيع مغلق (نعرف كل حدود الفئات).
- كل الفئات تحتوي على نفس الطول.
- التوزيعات متماثلة.

إذا لم يتوفر احد هذه الشروط أو أكثر فالوسيط (Mdn) هو الملائم.

4.- إذا كان التوزيع مفتوح فقط نستطيع حساب الوسيط (Mdn) . لكن إذا وقع مكان الوسيط في فئة مفتوحة لا يمكن أن نحسب الوسيط (Mdn) .

5.- إذا كل الفئات تحتوي على نفس الطول يمكننا حساب الوسط الحسابي (X) بالطريقة المباشرة والطريقة المختصرة، لكن إذا وجدنا إحدى الفئات بطول مختلف يمكن حساب الوسط الحسابي (X) فقط بواسطة الطريقة المباشرة.

6.- لا يمكن أن نحسب المنوال المصحح (Mo) إذا كانت إحدى فئات البيانات المبوبة مفتوحة. لكن يمكن أن نحسب المنوال غير المصحح.

7.- إذا جمعنا أو طرحنا أو ضربنا أو قسمنا كمية محددة للبيانات سواء كانت من النوع الأول أو الثاني أو الثالث ، قيم المنوال (Mo) والوسيط (Mdn) والوسط الحسابي (X) تتأثر بالجمع أو الطرح أو الضرب أو القسمة بنفس الكمية.

كالعادة نقوم بإدراج جدول الرموز العلمية الأكثر استخداماً في هذا الفصل.

المصطلح بالعربية	الرمز المستخدم	المصطلح بالإنجليزية
المنوال	Mo	MODE
لا منوال له		AMODAL
وحيد المنوال		UNIMODAL
ذو منوالين		BIMODAL
ثلاثي المنوال		TRIMODAL
التكرار الذي يلي الفئة المنوالية	f+1	
التكرار الذي يسبق الفئة المنوالية	f-1	
الارتفاعات	H	
الارتفاع الذي يلي الفئة المنوالية	h+1	
الارتفاع الذي يلي الفئة المنوالية	h-1	
الوسيط	Mdn	MEDIAN
مكانة أو ترتيب الوسيط	Pmdn	
تكرار الفئة الوسيطية	Fb	
التكرار المتجمع للفئة التي تسبق الفئة الوسيطية مباشرة	fai or fainf	
الوسط الحسابي	X	ARITHMETIC MEAN
مركز الفئة المفترضة أو المقدرة أو المعدلة	Xs	
المصحح أو المعدل	X	CORRECTIVE
الوسط الحسابي للأوساط أو الوسط الحسابي الكلي	Xt	
الوسط الحسابي المئوي	XP	
الوسط الحسابي المرجح	XW	WEIGHTED ARITHMETIC MEAN

الفصل الرابع
الربيعيات

إذا رتبنا مجموعة من الأرقام حسب قيمها فإن القيمة التي في المنتصف والتي تقسم مجموعة القيم إلى مجموعتين متساويتين في العدد هي الوسيط. وبتعميم هذه الفكرة يمكن أن نفكر في القيم التي تقسم المجموعة إلى أربعة أجزاء متساوية. هذه القيم هي:

الربيع الأول **(Q1)** أو الأدنى: هو القيمة التي يسبقها ربع البيانات ويليها ثلاثة أرباع البيانات.

الربيع الثاني **(Q2)** أو الوسيط: هو القيمة التي يسبقها نصف البيانات ويليها النصف الآخر.

الربيع الثالث **(Q3)** أو الأعلى: هو القيمة التي يسبقها ثلاثة أرباع البيانات ويليها أربع الآخر.

كذلك فإن القيم التي تقسم المجموعة إلى عشرة أجزاء متساوية تسمى بالعشيرات وهي:

العشير الأول **(D1)** : هو القيمة التي يسبقها 10% من البيانات ويليها 90% .

العشير الثاني **(D2)** : هو القيمة التي يسبقها 20% من البيانات ويليها 80% .

العشير الثالث **(D3)** : هو القيمة التي يسبقها 30% من البيانات ويليها 70% .

العشير الرابع **(D4)** : هو القيمة التي يسبقها 40% من البيانات ويليها 60% .

العشير الخامس **(D5)** أو الوسيط : هو القيمة التي يسبقها 50% من البيانات ويليها 50% .

العشير السادس **(D6)** : هو القيمة التي يسبقها 60% من البيانات ويليها 40% .

العشير السابع **(D7)** : هو القيمة التي يسبقها 70% من البيانات ويليها 30% .

العشير الثامن **(D8)** : هو القيمة التي يسبقها 80% من البيانات ويليها 20% .

العشير التاسع **(D9)** : هو القيمة التي يسبقها 90% من البيانات ويليها 10% .

مما سبق نستنتج الآتي:

1.- عدد الربيعات هو **3** من اليسار إلى اليمين، بينما عدد العشيرات هو **9** من اليسار إلى اليمين، وعدد المئينات هو **99** من اليسار إلى اليمين.

2.- يوجد تساوي بين:

$$\mathbf{Q_2 = D_5 = C_{50} = Mdn}$$
$$\mathbf{C_{25} = Q_1}$$
$$\mathbf{C_{75} = Q_3}$$

كيفية حساب قيمة الربيعات

فقط سنقوم بالتركيز على حساب الربيعات في النوع الثاني والثالث من التوزيعات التكرارية وذلك لكثرة استخدامها في هذان النوعان.

البيانات مبوبة (النوع الثاني)

مثال:

احسب قيمة Q1

X	f	fa
100	2	36
99	0	34
98	0	34
97	2	34
96	5	32
95	3	27
94	0	24
93	4	24
92	3	20
91	6	17
90	11	11

N=36

1.- نقوم بإيجاد التكرار المتجمع التصاعدي.

2.- نحدد مكانة أو ترتيب الفئة الربيعية كالآتي:

$$PQ = \frac{N.n}{4} = \frac{36.1}{4} = 9$$

3.- نبحث عن هذه المكانة في التكرار المتجمع، ونحددها.

4.- الربيع الأول هو الفئة التي تقابل مكانة الفئة الربيعية. في هذه الحالة قيمة الربيع هي 90

مثال:

احسب قيمة الربيع الثاني والثالث في المثال السابق؟

البيانات مبوبة أو التوزيعات تكرارية (النوع الثالث):

مثال:

احسب قيمة الربيع الثالث؟

X	f	Fa
100-104	2	81
95-99	10	79
90-94	24	69
85-89	11	45
80-84	11	34
75-79	10	23
70-74	6	13
65-69	1	7
60-64	6	6

1.– نقوم بإيجاد التكرار المتجمع التصاعدي.

2.– نحدد مكانة الربيع الثالث كالمعتاد

$$PQ = \frac{N.n}{4} = \frac{81.3}{4} = 60,75$$

3.– نحدد مكانة الربيع الثالث في التكرار المتجمع.

4.– نحسب قيمة الربيع بالصيغة الآتية:

$$XQ_3 = L_{-1} + \left(i \; \frac{N.n/4 - fai}{fb} \right)$$

$$XQ_3 = 89,5 + 5 \; \frac{60,75 - 45}{24} = 92,78$$

مثال:

احسب قيمة الربيع الأول والثاني في المثال السابق؟

كيفية حساب قيمة العشيرات

فقط سنقوم بالتركيز على حساب العشيرات في النوع الثاني والثالث من التوزيعات التكرارية وذلك لكثرة استخدامها في هذان النوعان.

البيانات مبوبة (النوع الثاني):

مثال:

احسب قيمة D3

X	f	fa
100	2	36
99	0	34
98	0	34
97	2	34
96	5	32
95	3	27
94	0	24
93	4	24
92	3	20
91	6	17
90	11	11

N=36

1.- نقوم بإيجاد التكرار المتجمع التصاعدي.

2.- نحدد مكانة أو ترتيب الفئة العشيرية كالآتي:

$$PD = \frac{N.n}{10} = \frac{36.3}{10} = 10,8$$

3.- نبحث عن هذه المكانة في التكرار المتجمع، ونحددها.

4.- العشير الثالث هو الفئة التي تقابل مكانة هذا العشير. في هذه الحالة قيمتة هي 90

مثال:

احسب قيمة باقي العشيرات في المثال السابق؟

البيانات مبوبة أو التوزيعات تكرارية (النوع الثالث(:

مثال:

احسب قيمة العشير السابع؟

X	f	Fa
100-104	2	81
95-99	10	79
90-94	24	69
85-89	11	45
80-84	11	34
75-79	10	23
70-74	6	13
65-69	1	7
60-64	6	6

1.– نقوم بإيجاد التكرار المتجمع التصاعدي.

2.– نحدد مكانة العشير السابع كالمعتاد

$$PD = \frac{N.n}{10} = \frac{81.7}{10} = 56,7$$

3.– نحدد مكانة العشير السابع في التكرار المتجمع.

4.– نحسب قيمة بالصيغة الآتية:

$$XD_7 = L_{-1} + \left(i \ \frac{N.n/10 - fai}{fb} \right)$$

$$XD_7 = 89,5 + 5 \ \frac{56,7 - 45}{24} = 91,94$$

كالعادة نقوم بإدراج جدول الرموز العلمية الأكثر استخداماً في هذا الفصل.

المصطلح بالعربية	الرمز المستخدم	المصطلح بالإنجليزية
الربيعات	Q	QUARTILES
العشيرات	D	DECILES
قيمة الربيع	XQ	
قيمة العشير	XD	
قيمة المئين	XC or XP	
مكانة أو ترتيب الربيع	$PQ = \frac{N.n}{4}$	
مكانة أو ترتيب العشير	$PD = \frac{N.n}{10}$	
مكانة أو ترتيب المئين	$PC = \frac{N.n}{100}$	
تكرار الفئة الربيعية أو العشيرية أو المئينية (حسب الحالة)	Fb	
التكرار المتجمع	fai	
مجموع التكرار	N	
عدد البيانات	N	

الفصل الخامس

مقاييس التشتت

سبق أن درسنا عرض البيانات الإحصائية ووصفها، ثم مقاييس النزعة المركزية التي تصف هذه البيانات عددياً، إلا أن هذه المقاييس غير كافية لتحديد صفات التوزيعات التكرارية والبيانات الإحصائية، فربما يكون لدينا ظاهرتان متساويتان في مقاييس النزعة المركزية كالوسيط والوسط الحسابي إلا أنهما مختلفتان، فمثلاً: إذا كانت درجات الحرارة في مدينتي الأغواط وغزة خلال أشهر العام الماضي 2022 هي:

الأغواط	**18**	**18**	**17**	**17**	**15**	**20**	**24**	**30**	**28**	**30**	**28**	**25**
الجلفة	**18**	**14**	**20**	**21**	**23**	**27**	**30**	**30**	**28**	**25**	**16**	**18**

هاتان المدينتان لهما نفس الوسط الحسابي $(X = 22,5)$ ونفس الوسيط $(Mdn = 22)$ وهذا يعني أن معدل درجات الحرارة في المدينتين متساوٍ، إلا أنه بالنظر إلى البيانات في كل من المدينتين نجد اختلافاً بينهما، وهذا يعني أن الوسط الحسابي لا يكفي لوصف البيانات أو الحكم على تشابهها. إذن لابد من استعمال مقاييس أخرى تبين لنا مدى اختلاف البيانات فيما بينها ومدى التفاوت والتغير بين مفرداتها. من هذه المقاييس: المدى، الانحراف المتوسط، التباين، الانحراف المعياري، معامل التغير، معامل التغير الربيعي، العزوم.

التشتت أو التغير: يقصد به الدرجة التي تتجه بها البيانات الرقمية للانتشار حول قيمة وسطى تسمى تشتت أو تغير البيانات، أو بمعنى آخر، التشتت يقصد به درجة التفاوت أو الاختلاف بين قيم مجموعة فإذا كانت هذه القيم متقاربة من بعضها البعض يكون التشتت صغيراً وإذا كانت متباعدة عن بعضها البعض أي متباينة يكون التشتت كبيراً.

1.- المدى (RANGE): هو عبارة عن إحصائي فعّال لبناء التوزيعات التكرارية من النوع الثالث، لكنه كمقياس للتباين أهميته ضئيلة. ولكي نقوم بحساب قيمته نستخدم إحدى الصيغتين الآتيتين (سبق استخدامهما):

المدى = اكبر قيمه – اصغر قيمه + وحدة قياس

المدى = الحد الحقيقي العلوي لأكبر قيمه – الحد الحقيقي السفلي لأصغر قيمه

2.- نصف المدى الربيعي أو الانحراف الربيعي (SEMI-INTERQUARTILE AMPLITUDE):

هو عبارة عن متوسط الفرق بين الربيع الثالث والربيع الأول. لكي نحسب قيمته نستخدم الصيغة الآتية:

$$Q = \frac{XQ_3 - XQ_1}{2}$$

حيث XQ_1 هو الربيع الأول و XQ_3 هو الربيع الثالث.

<u>البيانات مبوبة أو التوزيعات تكرارية (النوع الثالث) :</u>

مثال:

اوجد قيمة نصف المدى الربيعي (Q) للتوزيع الآتي:

X	f	Fa
90-98	27	72
81-89	24	45
72-80	11	21
63-71	4	10
54-62	4	6
45-53	0	2
36-44	1	2
27-35	1	1
	72	

الطريقة:

1.- نقوم بإيجاد التكرار المتجمع التصاعدي.

2.- نحدد مكانة الربيع الأول كالمعتاد

$$PQ_1 = \frac{N.n}{4} = \frac{72.1}{4} = 18$$

3.- نبحث عن مكانة الربيع الأول في التكرار المتجمع.

4.- نحسب قيمة الربيع بالصيغة الآتية:

$$XQ_1 = L_{-1} + \left(i \, \frac{N.n/4 - fai}{fb} \right)$$

$$XQ_1 = 71{,}5 + 9 \, \frac{18 - 10}{11} = 78{,}04$$

وبنفس الطريقة نحسب قيمة الربيع الثالث.

- نحدد مكانة الربيع الثالث كالمعتاد

$$PQ = \frac{N.n}{4} = \frac{72.3}{4} = 54$$

- نبحث عن مكانة الربيع الثالث في التكرار المتجمع.
- نحسب قيمة الربيع بالصيغة الآتية:

$$XQ_3 = L_{-1} + \left(i \, \frac{N.n/4 - fai}{fb} \right)$$

$$XQ_3 = 89{,}5 + \left(9 \, \frac{54 - 45}{27} \right) = 92{,}5$$

5.- بعد إيجاد قيمة كل من XQ_1 و XQ_3 نقوم بالتعويض في الصيغة الآتية للحصول على قيمة نصف المدى الربيعي:

$$Q = \frac{XQ_3 - XQ_1}{2} = \frac{92{,}5 - 78{,}04}{2} = 7{,}23$$

3.- الانحراف المتوسط (THE MEAN DEVIATION):

هو عبارة عن مجموع انحرافات البيانات عن وسطها الحسابي ولكن هذا المجموع يساوي صفراً دائماً لأن مجموع الانحرافات الموجبة عن الوسط الحسابي يساوي مجموع الانحرافات السالبة، ولذلك لابد من حذف الإشارة السالبة لنحصل على مقياس ذي معنى. وللتخلص من الإشارة السالبة نأخذ القيمة المطلقة (ABSOLUTE VALUE) . ونرمز له Dm

البيانات غير مبوبة (النوع الأول) :

$$Dm = \frac{\Sigma \left(|X - X| \right)}{n}$$

لكي نقوم بحساب الانحراف المتوسط نستعين بالمثال الآتي:

احسب الانحراف المتوسط للبيانات التالية:5 8 1 2

X	|X-X|
2	-2
1	-3
8	4
5	1
	|10|

1.– نحسب الوسط الحسابي.

$$X = \frac{\Sigma X}{n} = \frac{2+1+8+5}{4} = 4$$

2.– نحسب الفرق بين بين كل واحدة من القيم والوسط الحسابي.

3.– نجمع كل الفروق السابقة بدون النظر للإشارة (قيم مطلقة).

4.– نعوض في الصيغة السابقة

$$Dm = \frac{\Sigma (|X-X|)}{n} = \frac{10}{4} = 2,5$$

البيانات مبوبة (النوع الثاني):

$$Dm = \frac{\Sigma f (|Xm-X|)}{N}$$

مثال:

X	f	f. Xm	Xm-X	f (|Xm-X|)
7	2	14	3,48	6,96
6	1	6	2,48	2,48
5	9	45	1,48	13,32
4	5	20	0,48	2,4
3	6	18	-0,52	-3,12
2	3	6	-1,52	-4,56
1	7	7	-2,52	-17,64
	33	116		50,48

الطريقة:

1.- نحسب الوسط الحسابي.

$$\bar{X} = \frac{\Sigma f \cdot Xm}{N} = \frac{116}{33} = 3,52$$

2.- نحسب الفرق بين بين كل واحدة من القيم والوسط الحسابي.

3.- نضرب كل فرق من هذه الفروق في تكراره.

4.- نجمع كل الفروق السابقة بدون النظر للإشارة (قيم مطلقة).

5.- نعوض في الصيغة السابقة

$$Dm = \frac{\Sigma f(|Xm-X|)}{N} = \frac{50,48}{33} = 1,53$$

البيانات مبوبة أو التوزيعات التكرارية (النوع الثالث)"

$$Dm = \frac{\Sigma f(|Xm-X|)}{N}$$

لكي نقوم بحساب الانحراف المتوسط نستعين بالمثال الآتي:

X	f	Xm	f . Xm	Xm – X	f (|Xm-X|)
35-40	7	37,5	262,5	14,86	104,02
30-35	1	32,5	32,5	9,86	9,86
25-30	6	27,5	165	4,86	29,16
20-25	9	22,5	202,5	-,14	-1,26
15-20	5	17,5	87,5	-5,14	-25,7
10-15	1	12,5	12,5	-10,14	-10,14
5-10	7	7,5	52,5	-15,14	-105,98
	36		815		286,12

الطريقة:

1.- نحسب الوسط الحسابي.

$$\bar{X} = \frac{\Sigma f \cdot Xm}{N} = \frac{815}{36} = 22,64$$

2.- نحسب الفرق بين بين كل واحدة من القيم والوسط الحسابي.

3.- نضرب كل فرق من هذه الفروق في تكراره.

4.– نجمع كل الفروق السابقة بدون النظر للإشارة (قيم مطلقة).

5.– نعوض في الصيغة السابقة

$$Dm = \frac{\Sigma f\,(|Xm-X|\,)}{N} = \frac{286,12}{36} = 7,95$$

4.– **التباين** **(THE VARIANCE)**:

هو عبارة عن متوسط مربعات انحرافات القيم عن وسطها الحسابي. ونرمز له بالرمز Sx

البيانات غير مبوبة (النوع الأول):

$$Sx = \frac{\Sigma\,(X-X)}{n}$$

لكي نقوم بحساب التباين نستعين بالمثال الآتي:

مثال:

احسب التباين للبيانات التالية:

2 1 8 5

X	X-X	(X-X)
2	-2	4
1	-3	9
8	4	16
5	1	1
		30

الطريقة:

1.– نحسب الوسط الحسابي.

$$X = \frac{\Sigma X}{n} = \frac{2+1+8+5}{4} = 4$$

2.– نحسب الفرق بين بين كل واحدة من القيم والوسط الحسابي.

3.– نحسب مربعات هذه الانحرافات، ثم نجمعها.

4.– نعوض في الصيغة السابقة

$$Sx = \frac{\Sigma\,(X-X)}{n} = \frac{30}{4} = 7,5$$

البيانات مبوبة (النوع الثاني) :

$$Sx = \frac{\Sigma f\,(Xm-X\,)}{N}$$

X	f	f. Xm	Xm-X	(Xm-X)	f (Xm-X)
7	2	14	3,48	12,11	24,22
6	1	6	2,48	6,15	6,15
5	9	45	1,48	2,19	19,71
4	5	20	0,48	0,23	1,15
3	6	18	-0,52	0,27	1,62
2	3	6	-1,52	2,31	6,93
1	7	7	-2,52	6,35	44,45
	33	116			104,23

الطريقة:

1.- نحسب الوسط الحسابي.

$$X = \frac{\Sigma\, f\,.\,Xm}{N} = \frac{116}{33} = 3{,}52$$

2.- نحسب انحراف كل واحدة من القيم عن الوسط الحسابي.

3.- نحسب مربعات هذه الانحرافات.

4.- نضرب كل تكرار في مربع الانحرافات المقابل له، ثم نجمعهم.

6.- نعوض في الصيغة السابقة

$$Sx = \frac{\Sigma f\,(Xm-X)}{N} = \frac{104{,}23}{33} = 3{,}16$$

البيانات مبوبة أو التوزيعات التكرارية (النوع الثالث)"

حساب التباين باستخدام الطريقة المباشرة:

$$Sx = \frac{\Sigma f\,(Xm-X\,)}{N}$$

لكي نقوم بحساب التباين نستعين بالمثال الآتي:

X	f	Xm	f . Xm	Xm – X	(Xm – X)	f (Xm-X)
35-40	7	37,5	262,5	14,86	220,82	1545,74
30-35	1	32,5	32,5	9,86	97,22	97,22
25-30	6	27,5	165	4,86	23,62	141,72
20-25	9	22,5	202,5	-,14	0,02	0,18
15-20	5	17,5	87,5	-5,14	26,42	132,10
10-15	1	12,5	12,5	-10,14	102,82	102,82
5-10	7	7,5	52,5	-15,14	229,22	1604,54
	36		815			3624,32

الطريقة:

1.– نحسب الوسط الحسابي.

$$X = \frac{\Sigma\, f\,.\,Xm}{N} = \frac{815}{36} = 22,64$$

2.– نحسب انحراف مركز البيانات عن الوسط الحسابي.

3.– نحسب مربعات هذه الانحرافات.

4.– نضرب كل تكرار في مربع الانحرافات المقابل له، ثم نجمعهم.

5.– نعوض في الصيغة السابقة

$$Sx = \frac{\Sigma f\,(Xm-X\,)}{N} = \frac{3624,32}{36} = 100,675$$

5.– الانحراف المعياري (STANDARD DEVIATION):

هو عبارة عن الجذر التربيعي للتباين. أهم مقاييس التشتت وأكثرها استخداما وذلك لدخوله في حساب كثير من المقاييس الإحصائية الأخرى. ونرمز له بالرمز Sx

البيانات غير مبوبة (النوع الأول) :

$$Sx = \frac{\Sigma (X-X)}{n}$$

مثال:

$Sx = 7,5$ $Sx = 2,74$

البيانات مبوبة (النوع الثاني):

$$Sx = \frac{\Sigma f (Xm-X)}{N}$$

$Sx = 3,16$ $Sx = 1,78$

البيانات مبوبة أو التوزيعات التكرارية (النوع الثالث) Type (III):

$$Sx = \frac{\Sigma f (Xm-X)}{N}$$

$Sx =100,675$ $Sx = 10,03$

الانحراف المعياري الكلي (Sxt):

لحساب الانحراف المعياري الكلي لعينتين أو أكثر نستخدم الصيغة الآتية:

$$Sxt = \frac{N_1 (Sx_1 + d_1) + N_2 (Sx_2 + d_2) + \ldots + Nn (Sxn + dn)}{N_1 + N_2 + \ldots + Nn}$$

حيث:

$d = X – Xt$

مثال:

احسب الانحراف المعياري الكلي للبيانات الآتية:

$N_1 = 100$	$N_2 = 150$	$N_3 = 148$
$X_1 = 57$	$X_2 = 71$	$X_3 = 60$
$Sx_1 = 8,16$	$Sx_2 = 10,92$	$Sx_3 = 12,26$

الحل:

1.– نقوم بحساب الوسط الحسابي الكلي.

$$Xt = \frac{N_1.X_1 + N_2.X_2 + \ldots\ldots + Nn.Xn}{N_1 + N_2 + \ldots\ldots + Nn} = \frac{100.57 + 150.71 + 148.60}{100 + 150 + 148} = 63{,}39$$

2.– نحسب الفروق بين كل وسط حسابي والوسط الحسابي الكلي:

$d = X - Xt$

$d_1 = X_1 - Xt = 57 - 63{,}39 = -6{,}39$

$d_2 = X_2 - Xt = 71 - 63{,}39 = 7{,}61$

$d_3 = X_3 - Xt = 60 - 63{,}39 = -3{,}39$

$$Sxt = \frac{N_1 (Sx_1 + d_1) + N_2 (Sx_2 + d_2) + \ldots + Nn (Sxn + dn)}{N_1 + N_2 + \ldots\ldots + Nn}$$

$$= \frac{100[8{,}16 + (-6{,}39)] + 150\,(10{,}92 + 7{,}61) + 148\,[12{,}26 + (-3{,}39)]}{100 + 150 + 148} = 12{,}407$$

6.– معامل الاختلاف (COEFFICIENT OF VARIATION) أو التشتت (DISPERSION) :

التغير الفعلي أو التشتت كما نحصل عليه من الانحراف المعياري أو غيره من مقاييس التشتت يسمى بالتشتت المطلق. ولكن تغير أو تشتت 1 متر عند قياس مسافة 1000 متر يختلف في تأثيره عن نفس تغير 1 متر في مسافة 20 متر. ومقياس لهذا التأثير تحصل عليه بالتشتت النسبي ويعرف بما يلي:

$$\text{التشتت النسبي} = \frac{\text{التشتت المطلق}}{\text{المتوسط}}$$

إذا كان التشتت المطلق هو الانحراف المعياري Sx والمتوسط هو الوسط الحسابي X فإن التشتت النسبي يسمى بمعامل الاختلاف أو معامل التشتت ويعرف كالآتي:

$$CV = \frac{SX}{X} \cdot 100$$

وبشكل عام يعبر عنه كنسبة. يفيد عند مقارنة توزيعات ذات وحدات مختلفة (الوزن والطول) أو توزيعات ذات وحدات متساوية لكن وسطها الحسابي مختلف.

أحد عيوب معامل الاختلاف هو أنه يصبح عديم الفائدة عندما تكون X قريبة من الصفر.

مثال:

مصنع لإطارات السيارات ينتج نموذجين من الإطارات A و B . النموذج A متوسط عمره 10.000 KM ومتوسط انحرافه 2.000 KM . النموذج B متوسط عمره 11.000 KM ومتوسط انحرافه 1.000 KM . نود معرفة أي النموذجين أحسن؟

الحل:

نستخدم معامل التشتت وبعد ذلك نقارن:

النموذج A :

$$CV = \frac{SX}{X} .100 = \frac{2.000}{10.000} .100 = 20\%$$

النموذج B :

$$CV = \frac{SX}{X} .100 = \frac{1.000}{11.000} .100 = 9\%$$

معامل التشتت للنموذج B اقل من النموذج A ، وبالتالي النموذج B أحسن من النموذج A .

7.– نسبة الاختلاف (COCIENT OF VARIATION) :

يستخدم فقط في المتغيرات النوعية (الكيفية)، وقيمته تتراوح بين 0 و 1. ولحسابه نستخدم الصيغة الآتية:

$$\boxed{V = \frac{N - f\,Mo}{N}}$$

حيث:

N هو المجموع الكلي للبيانات.

f Mo هو تكرار المنوال.

خصائص واعتبارات مقاييس التشتت

1.- المدى (R) والانحراف الربيعي (Q) والانحراف المتوسط (DM) والانحراف المعياري (Sx) قيمتهم تكون بنفس الوحدات المستخدمة في السؤال. التباين يكون بنفس الوحدات لكن قيمته مربعة. **معامل الاختلاف (CV) يشير إلى ما هي نسبة الانحراف المعياري الموجودة في الوسط الحسابي.**

2.- عند حساب الانحراف المتوسط (DM) والتباين (Sx) والانحراف المعياري (Sx) نأخذ في الحسبان كل البيانات. بينما الانحراف الربيعي (Q) يهتم فقط بقيمة الربيع الثالث (XQ_3) والأول (XQ_1) ، ولحساب المدى (R) نركز فقط على أطراف القيم (أكبر وأصغر قيمة).

3.- الانحراف المتوسط (DM) والانحراف الربيعي (Q) دائماً قيمتهم اصغر من الانحراف المعياري (Sx) (لنفس البيانات أو التوزيع). إذا كان التوزيع طبيعي (سيتم شرحة فيما بعد) فالعلاقة بين هذه المقاييس الثلاثة ثابتة:

$$Sx = \frac{Q}{0,6745}$$

$$Sx = \frac{DM}{0,7948}$$

4.- عندما يكون الوسط الحسابي (X) هو الأكثر ملائمة في مقاييس النزعة المركزية، فالانحراف المعياري كمقياس تشتت هو الملائم. وعندما يكون الوسيط هو الملائم، فالانحراف الربيعي كمقياس تشتت هو الملائم.

$$\mathbf{X \rightarrow Sx}$$
$$\mathbf{Mdn \rightarrow Q}$$

5.- لا يمكن أن نحسب الانحراف الربيعي إذا كانت الفئة الحرجة للربيع الثالث (Q3) أو الأول (Q1) مفتوحة.

6.- إذا كل الفئات تحتوي على نفس الطول يمكننا حساب التباين (Sx) أو الانحراف المعياري (Sx) بالطريقة المباشرة والطريقة المختصرة، لكن إذا وجدنا إحدى الفئات بطول مختلف يمكن حساب التباين

(Sx) أو الانحراف المعياري (Sx) فقط بواسطة الطريقة المباشرة.

7.- إذا جمعنا أو طرحنا كمية محددة للبيانات سواء كانت من النوع الأول أو الثاني أو الثالث ، فإن قيم المدى (R) والانحراف الربيعي (Q) والانحراف المتوسط (DM) والتباين (Sx) والانحراف المعياري (Sx) لا تتغير.

8.- إذا ضربنا أو قسمنا كمية محددة للبيانات سواء كانت من النوع الأول (Type I) أو الثاني أو الثالث، فإن قيم المدى (R) والانحراف الربيعي (Q) والانحراف المتوسط (DM) والتباين (Sx) والانحراف المعياري (Sx) تتغير بنفس هذه الكمية.

9.- إذا جمعنا كمية محددة للبيانات سواء كانت من النوع الأول أو الثاني أو الثالث ، فإن قيمة معامل الاختلاف (CV) تنقص ولكن بكمية مختلفة.

10.- إذا طرحنا كمية محددة من البيانات سواء كانت من النوع الأول أو الثاني أو الثالث ، فإن قيمة معامل الاختلاف (CV) تزيد ولكن بكمية مختلفة.

9.- إذا ضربنا أو قسمنا كمية محددة للبيانات سواء كانت من النوع الأول أو الثاني أو الثالث ، فإن قيمة معامل الاختلاف (CV) لا تتغير.

كالعادة نقوم بإدراج جدول الرموز العلمية الأكثر استخداماً في هذا الفصل.

المصطلح بالعربية	الرمز المستخدم	المصطلح بالإنجليزية
المدى	R	RANGE
نصف المدى الربيعي أو الانحراف الربيعي	Q	SEMI-INTERQUARTILE AMPLITUDE
الانحراف المتوسط	DM	MEAN DEVIATION
التباين	Sx	THE VARIANCE
الانحراف المعياري	Sx	STANDARD DEVIATION
معامل التغير	CV	COEFFICIENT OF VARIATION
معامل التغير الربيعي	CQV	COEFFICIENT OF QUARTILE VARIATION
تصحيح شبرد	Sxc	SHEPPARDS CORRECTION
الانحراف المعياري الكلي	Sxt	TOTAL STANDARD DEVIATION
نسبة الاختلاف	V	COCIENT OF VARIATION

الفصل السادس

معاملات الارتباط

لقد بحثنا في الفصول السابقة بعض المسائل المتعلقة بقياسات ومشاهدات متعلقة بمتغير واحد، وفي هذا الفصل نبدأ بدراسة بعض الأمور التي تتعلق بمتغيرين اثنين.

إذا كان لكل قيمة من قيم المتغير X يوجد قيمة مقابلة لمتغير آخر Y فإن الأزواج المرتبة من هذه القيم تسمّى مجتمعاً ذا بعدين ويسمّى الزوج المرتب $(X, \ Y)$ متغيراً عشوائياً ذا بعدين. إن الأمثلة على المجتمعات ذات البعدين كثيرة وهي ذات أهمية كبيرة في التربية وعلم النفس وعلم الاجتماع والإدارة والاقتصاد والزراعة والصناعة. فمثلاً ربما يكون لدينا بيانات عن طلبة الجامعة تمثل علامة الطالب في شهادة الدراسة الثانوية ومعدله عند التخرج من الجامعة. يمكن دراسة هذه البيانات للإجابة على السؤالين:

1.- هل هناك علاقة بين المتغيرين؟

2.- إذا كان هناك علاقة بين المتغيرين فكيف نعبر عنها بمعادلة؟

سنعتمد في دراستنا لمسائل الارتباط والانحدار إلى تمثيل مجموعة الأزواج المرتبة من المشاهدات بيانياً وهذا ما يسمّى بلوحة الانتشار.

الارتباط الخطي (Linear Correlation)

بالنظر إلى لوحة انتشار أي متغير عشوائي ذي بعدين (X, Y) نلاحظ فيما إذا كان هناك ارتباط أو علاقة بين المتغيرين X و Y . وغرض دراسة الارتباط هو قياس قوة الارتباط الخطي بين متغيرين. أما معامل الارتباط r_{xy} فهو مقياس قوة العلاقة الخطية بين المتغيرين ويقيس مدى التغير والتأثير الذي يطرأ على Y عندما يزداد X مقداراً معيناً. إنه يعطينا فكرة فيما إذا كانت Y تزداد كلما ازدادت X أو أنها تنقص كلما ازدادت X أو أنها لا تتأثر. في الحالة الأولى يكون هناك ارتباط موجب وفي الثانية ارتباط سالب أما في الثالثة فلا يوجد ارتباط. الشكل الآتي يوضح لنا لوحات انتشار نلاحظ فيها وجود الارتباط أو عدمه.

خواص معامل الارتباط:

1.- من الممكن رياضياً إثبات أن قيمة r_{xy} تتراوح بين ±1 ، فالقيمة $r_{xy}=1$ تعني ارتباطا كاملاً موجباً بين المتغيرين فإذا ازداد أحدهما أزداد الآخر، أما $r_{xy}=-1$ فتعني وجود ارتباط كامل سالب بين المتغيرين فإذا ازداد احدهما نقص الآخر، أما $r_{xy}=0$ فتعني عدم وجود ارتباط خطي بين المتغيرين.

إن قيم r_{xy} الموجبة التي لا تساوي 1 وقيم r_{xy} السالبة التي لا تساوي -1 تدل على وجود ارتباط ولكن نحتاج إلى اختبار دلالة هذه القيم وذلك بإيجاد قيمتين احدهما سالبة والأخرى موجبة فإذا كان r_{xy} اكبر من القيمة الموجبة أو اصغر من القيمة السالبة فهناك ارتباط بين المتغيرين أما إذا وقع r_{xy} بين القيمتين فليس هناك ارتباط خطي بين المتغيرين.

ويجب التأكيد أن وجود ارتباط خطي بين متغيرين لا يعني السببية أي انه لا يعني أن احد المتغيرين سبب في وجود المتغير الآخر.

2.- إذا وقعت جميع نقاط لوحة الانتشار على خط مستقيم فقيمة r_{xy} تكون ±1 .

3.- إذا لم يوجد علاقة خطية بين المتغيرين يكون $r_{xy}=0$.

مثال:

إذا كان معامل الارتباط في إحدى الدراسات $r_{xy}=0{,}8$ ، وفي دراسة أخرى و $r_{xy}=0{,}4$. فهل من الصحيح القول بأن معامل الارتباط الأول وهو 0,8 أقوى مرتين من معامل الارتباط الثاني وهو 0,4 ؟ الإجابة بالنفي، وذلك لان معامل الارتباط يساوي 0,8 ومعامل التحديد Coefficient of determination (r^2_{xy}) والذي يساوي مربع معامل الارتباط = $(0{,}8)^2 = 0{,}64$ ويبلغ معامل التحديد للعلاقة الثانية = $(0{,}4)^2 = 0{,}16$ ، وهذا يعني انه في العلاقة الأولى يفسر المتغير المستقل حوالي 64% من التغيرات في المتغير التابع، بينما تبلغ هذه النسبة فقط 16% في العلاقة الثانية، وعلى هذا فأنه يمكن القول بأن العلاقة ذات معامل الارتباط الذي يبلغ 0,8 تبلغ في قوتها أربعة أضعاف أو أمثال العلاقة ذات معامل الارتباط = 0,4 .

معامل ارتباط بيرسون (Pearson's Correlation Coefficient):

عند استخدامه يجب أن يكون كلا المتغيرين متصلين.

البيانات غير مبوبة (النوع الأول)

هناك أربعة طرق لحساب معامل ارتباط بيرسن وتعطي كلها نفس النتيجة وهي:

أولا: باستخدام متوسط الانحرافات

$$r_{xy} = \frac{\sum x.y}{n.SX.Sy}$$

ثانياً: باستخدام الدرجات الخام مباشرة

$$r_{xy} = \frac{n \ \sum X.Y - \sum X \ . \sum Y}{\left[n \sum X - (\sum X)\right] \left[n \sum Y \quad (\sum Y)\right]}$$

ثالثاً: باستخدام الدرجات المعيارية

$$r_{xy} = \frac{\sum Zx.Zy}{n}$$

رابعاً: باستخدام الفروق بين الدرجات الخام

$$r_{xy} = \frac{\sum x.y}{\sum X^2 . \sum y^2}$$

مثال: أوجد معامل ارتباط بيرسون بين المتغيرين X و Y المبينين في الجدول الآتي:

X	Y	x	y	x^2	y^2	Zx	Zy	ZxZy	xy	X^2	Y^2	XY
15	8	7	2	49	4	1,61	0,57	0,917	14	225	64	120
13	12	5	6	25	36	1,15	1,71	1,967	30	169	144	156
11	6	3	0	9	0	0,69	0	0	0	121	36	66
10	10	2	4	4	16	0,46	1,14	0,524	8	100	100	100
9	8	1	2	1	4	0,23	0,57	0,131	2	81	64	72
7	6	-1	0	1	0	-0,23	0	0	0	49	36	42
7	5	-1	-1	1	1	-0,23	-0,29	0,067	1	49	25	35
5	3	-3	-3	9	9	-0,69	-0,86	0,593	9	25	9	15
3	2	-5	-4	25	16	-1,15	-1,14	1,311	20	9	4	6
0	0	-8	-6	64	36	-1,84	-1,71	3,146	48	0	0	0
80	**60**			**188**	**122**			**8,656**	**132**	**828**	**482**	**612**

$$X = \frac{80}{10} = 8 \qquad Sx = 4{,}34$$

$$Y = \frac{60}{10} = 6 \qquad Sy = 3{,}5$$

أولا: باستخدام متوسط الانحرافات

$$r_{xy} = \frac{\Sigma x.y}{n.SX.Sy} = \frac{132}{10 \,.\, 4{,}34 \,.\, 3{,}5} = 0{,}87$$

ثانياً: باستخدام الدرجات الخام مباشرة

$$r_{xy} = \frac{n\ \Sigma X.Y - \Sigma X\ .\ \Sigma Y}{\left[n\,\Sigma X - (\Sigma X)\right]\ \left[n\,\Sigma Y \quad (\Sigma Y)\right]}$$

$$= \frac{10 \,.\, 612 - 80 \,.\, 60}{10 \,.828 - 80^2 \quad 10.482 - 60^2} = 0{,}87$$

ثالثاً: باستخدام الدرجات المعيارية

$$r_{xy} = \frac{\Sigma Zx.Zy}{n} = \frac{8{,}656}{10} = 0{,}87$$

رابعاً: باستخدام الفروق بين الدرجات الخام

$$r_{xy} = \frac{\Sigma x.y}{\Sigma X^2 . \Sigma y^2} = \frac{132}{188 . 122} = 0,87$$

التفسير: توجد علاقة بين المتغيرين موجبة وغير تامة ومرتفعة

معامل التحديد **Coefficient of determination** (r^2_{xy}):

$$r^2_{xy} = (0,87)^2 = 0,7569$$

وهذا يعني أن المتغير المستقل يفسر حوالي 75,69% من التغيرات في المتغير التابع.

معامل ارتباط سبيرمان **(Spearman's coefficient of rank correlation)**:

عند استخدامه يجب أن يكون كلا المتغيرين شبه كمي.

بدلاً من استخدام قيم محددة للمتغيرات، أو عندما لا يكون مثل هذا التحديد متاح، فإنه يمكن ترتيب البيانات حسب ترتيب حجمها، أهميتها،... وغير ذلك باستخدام الأرقام $1, 2, 3, 4, \ldots\ldots\ldots n$. إذا رتبنا متغيرين X و Y بهذه الطريقة فإن معامل ارتباط سبيرمان لارتباط الرتب كما يلي:

$$\rho\ (\text{rho}) = 1 - \frac{6\ \Sigma\ d^2}{n(n^2 - 1)}$$

حيث :

d = الفرق بين رتب القيم المتقابلة في X ، Y

n = عدد أزواج القيم (X ، Y) في البيانات.

مثال:

احسب معامل سبيرمان للارتباط بالرتب بين المعدلات التالية لعشرة طلاب في امتحان الصف الثالث الثانوي وامتحان شهادة الدراسة الثانوية:

معدل الطالب في الصف الثالث الثانوي *X*	معدل الطالب في شهادة الدراسة الثانوية *Y*	رتب x	رتب y	d	d^2
75	*69*	6	7	-1	1
82	*85*	4	4	0	0
65	*55*	8	9	-1	1
90	*90*	2	1	1	1
77	*80*	5	5	0	0
60	*50*	9	10	-1	1
55	*57*	10	8	2	4
87	*88*	3	3	0	0
91	*89*	1	2	-1	1
73	*71*	7	6	1	1
					10

الخطوات:

1.- نقوم بترتيب المتغير (X) داخلياً من الأقل إلى الأكبر أو من الأكبر إلى الأقل.

2.- نقوم بترتيب المتغير (Y) بنفس الطريقة.

3.- نحسب (d) أي الفرق بين رتب القيم المتقابلة في X ، Y

4.- نحسب d^2 ثم نقوم بعد ذلك بتجميعها.

5.- نقوم بتطبيق القانون.

$$\rho_{\text{(rho)}} = 1 - \frac{6 \sum d^2}{n(n^2 - 1)} = 1 - \frac{6 \,.\, 10}{10(100-1)} = 0{,}94$$

مثال:

احسب معامل سبيرمان للارتباط بالرتب بين عدد ساعات الدراسة والمستويات

في الامتحانات لعشرة من الطلاب:

عدد ساعات الدراسة *X*	المستوى في الامتحان *Y*	رتب x	رتب y	d	d^2
9	*56*	5	4	1	1
5	*44*	2,5	2	0,5	0,25
11	*79*	7	8	-1	1
13	*72*	8	7	1	1
10	*70*	6	6	0	0
5	*54*	2,5	3	-0,5	0,25
18	*94*	10	10	0	0
15	*85*	9	9	0	0
2	*33*	1	1	0	0
8	*65*	4	5	-1	1
					4,25

الخطوات:

1.– نقوم بترتيب المتغير (X) داخلياً من الأقل إلى الأكبر أو من الأكبر إلى الأقل. عندما نجد أكثر من قيمة لها نفس القيمة فأنه يمكن عمل ترتيب واحد (أو مشترك لهما). فمثلاً القيمتين الثانية والثالثة (في الترتيب من الأقل للأعلى) لهم نفس القيمة 5 فأنه يمكن عمل ترتيب واحد أو مشترك كالآتي:

$$\frac{2+3}{2} = 2,5$$

2.– نقوم بترتيب المتغير (Y) بنفس الطريقة.

3.– نحسب (d) أي الفرق بين رتب القيم المتقابلة في X ، Y

4.– نحسب d^2 ثم نقوم بعد ذلك بتجميعها.

5.– نقوم بتطبيق القانون.

$$\rho_{\text{(rho)}} = 1 - \frac{6\,\Sigma\, d^2}{n(n^2-1)} = 1 - \frac{6\,.\,4,25}{10(100-1)} = 0,97$$

الفهرس

المراجع

المراجع باللغة العربية:

- أحمد عامر، الإحصاء I أو الإحصاء II – دار الغرب للنشر والتوزيع، الجزائر، سنة 2020.
- جيلالي جلاطو، " الإحصاء مع تمارين ومسائل محلولة "، الطبعة الثامنة، ديوان المطبوعات الجامعية، ، الجزائر ، سنة 2010 .
- شرف الدين خليل، الإحصاء الوصفي " – شبكة الأبحاث والدراسات الاقتصادية، من الموقع الإلكتروني: www.rr4ee.net ، 2012/05/02.
- عبد الحفيظ مصطفى، " نظرية الاحتمالات مبادئ وتطبيقات "، ديوان المطبوعات الجامعية، الجزائر، سنة 2001.
- عبد الرزاق عزوز، " الكامل في الإحصاء، – دروس مفصلة، تمارين ومسائل مع الحلول"– جزئيين، ديوان المطبوعات الجامعية، ، الجزائر ، سنة 2018 .
- محمد حسين محمد رشيد، "مبادئ الإحصاء والاحتمالات ومعالجتها– باستخدام برنامج SPSS "، دار الصفاء للنشر والتوزيع– عمان، سنة 2012.
- محمد شفيق ياسين أنور اللحام، " مبادئ الإحصاء والاحتمال "، ديوان المطبوعات الجامعية، الجزائر، سنة 2020.
- موسى محمد أماني، " التحليل الإحصائي للبيانات" – مركز تطوير الدراسات العليا والبحوث كلية الهندسة – جامعة القاهرة، من الموقع الإلكتروني:

www.Pathways.cu.edu.eg , www.capscu.com ، 2020/08/01.

المراجع باللغة الأجنبية

- Grais.B « Statistique descriptive », Dumod, 2éme édition, Année :2022.
- Anderson-Sweeney-Williams « Statistique pour l'économie et la gestion », Traduction de la 5e édition américaine par Claire Borenberge, édition de boeck université, France, 2019.
- Bernard Verlant et Geneviéve Saint-Pierre « Statistique et Probabilités –Manuel de cours » , LMD,DUEA, BTS, Editions Berti, Année : 2010.
- Célyne Laliberté , « Probabilités et statistique », Distribué Par ERPI,-Québec.Année :2009.
- G . SAPORTA , Probabilités Analyse de Données Et Statistique – France,Année:2013
- Gilbert.N, « Staistique », édition H.R.W, etéé Montréal, Année :2018.
- Khaldi Khaled , « FAC Méthodes Statistiques et Probabilités » Editions CASBAH, Année : 2000.
- walder masiéri, « Cours Eco-Gestion , Statistique et Calcul des Probabilités », édition campus dalloz, France, Année : 2022.