مبادئ الاحصاء

المدرس المساعد

زين العابدين عبود كاظم الحسيني

الفصل الأول

علم الإحصاء تعريفه و أهميته

أولا : تعريف علم 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 علم 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 علم الإحصاء قديما هو عد أو حصر الأشياء المراد توفير بيانات إحصائية عنها ، وكانت الجهة التى تقوم بإعداد الإحصاءات على مستوى الدولة تعرف بمصلحة التعداد ولذلك كان التعريف القديم لعلم الإحصاء أنه علم العد ، أي العلم الذي يشتمل على أساليب جمع البيانات الكمية عن المتغيرات والظواهر موضوع الدراسة .

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

من ذلك على سبيل المثال ، أن نظرية العينات ساعدت الباحثين على استخلاص استنتاجات عديدة من دراسة عدد صغير من الأفراد أو الأشياء –*العينة* - وتعميم تلك الاستنتاجات على المجتمع الذى سحبت منه العينة بأسره ولذلك يعرف علم الإحصاء حديثاً بأنه : (علم متكامل يتضمن الأسلوب العلمى الضرورى لتقصى حقائق الظواهر واستخلاص النتائج عنها ، كما يتضمن أيضاً أيضا النظرية اللازمة للقياس واتخاذ القرار فى كافة الميادين الاقتصادية والاجتماعية والسياسية والعسكرية)

ثانيا: أهميه علم الإحصاء

لقد أصبح لعلم الإحصاء أهميه بالغه فى حياتنا الحديثة فصارت الإحصاءات مألوفة لدينا وتمثل جانبا مهما من المعلومات التى نطالعها كل يوم مثل جداول النقاط التى تحرزها أنديه كره القدم وتنشر فى الصحف والمجلات والتقديرات الخاصة بالتنبؤات الجوية ومؤشرات البورصة وانجازات الحكومة فى مجال الإسكان والتعمير والتغيرات التي تطرأ على أسعار العملات وأثمان السلع . وربما يتساءل المرء عن أهميه الإحصاء بالنسبة لدارس علم الاجتماع أو علم النفس معتقداً أن الإحصاء موضوع يدخل فى صميم تخصص التجاريين والاقتصاديين والواقع أن الباحث الاجتماعي والمتخصص فى العلوم الاجتماعية بوجه عام يحتاج فى كثير من الأحيان إلى استخدام الأرقام لكي يلخص ويعرض بها مجموعه من المشاهدات التى تتعلق بظاهرة يهتم بدراستها ، فقد يطلب منه أن يقدم تقريراً عن مدي التطور الذى حققه برنامج معين لمحو الأمية بين نزلاء المؤسسة التي يعمل بها ، وقد يكلف بدراسة الأسباب التى تجعل الذكور أكثر تقدما وحرصا على التعليم من الإناث فى المدرسة التى يشتغل فيها .

ففي كل مناسبة من هذه المناسبات سيحتاج الباحث أو الدارس إلى أداة من الأدوات الإحصائية لكي يستخدمها فى تلخيص أفكاره والتعبير عنها بصورة محدده ومؤثرة ، فالعبارة التى مؤداها " لقد نجحنا فى محو أميه 90%من العاملين الأميين بالمصنع " أقوى وأشد من العبارة التى مفادها : " لقد نجحنا فى محو أمية عدد كبير من العاملين الأميين بالمصنع " : [(4)] يحتل الإحصاء (أو الأساليب الإحصائية) أهميه خاصة فى الأبحاث العلمية الحديثة ، إذ لا تخلو أى دراسة أو بحث من دراسة تحليليه إحصائية تتعرض لأصل الظاهرة أو الظاهرات المدروسة فتصور واقعها فى قالب رقمى ، وتنتهي إلى ابرز اتجاهاتها وعلاقاتها بالظاهرات الأخرى .

إن دراسة الإحصاء أمر له فوائد كثيرة بالنسبة لدارسي العلوم الاجتماعية وخاصة بعد أن تفتحت أمامهم مجالات عمل كثيرة فى تنظيمات 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 .

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 وعلى اختبار العلاقات بين الأشياء والوقائع التى يهتم بها إلا أن ذلك لا يلغى بصيرته السوسيولوجية وخبرته المهنية .

وبعبارة أخري ، يقتصر دور الأدوات الإحصائية على توفير المؤشرات المبدئية التى تساعد الباحث على رفض أو قبول الفروض التى يقوم بدراستها في حدود درجه معينه من الثقة . والإحصاء أيضا أداه لا تستخدم إلا في العثور على إجابات عن أسئلة تتصل ببيانات يمكن التعبير عنها بصيغ كميه . وهناك فى مجال العلوم الاجتماعية 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 .

ثالثا : تطور علم 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 قديمه ، وان تعداد السكان عند القدماء المصرين وفي الصين أمثلة توضح اهتمام الحكومات منذ القدم بالمعلومات الاجتماعية وذلك لأغراض التنظيم والتخطيط فى أحوال السلم والحرب .

ويبدو أن كلمه إحصاء (statistics) قد ظهرت لأول مره عام1749 وهى مشتقه 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 والاستهلاك والثروة 00000الخ . وهكذا بداء العلم وتطوره باعتباره علم الدولة أو علم الملوك .

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 ".

وأخيرا يفهم بالإحصاء فرع من العلم له نظريته الخاصة . وعلم الإحصاء ، شأنه في ذلك شأن أى فرع آخر من فروع العلم له أسلوبه وموضوعات البحث الخاص به

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

رابعا : علاقة علم الإحصاء بالعلوم الاجتماعية

تأثرت العلوم الاجتماعية وخاصة علم الاجتماع وعلم النفس وعلم السياسة بالتطورات . التي حققها علم الإحصاء ، واستعان العلماء الاجتماعيون بمنهج جديد فى دراساتهم . وهو المنهج الاحصائى الذى ينطوي علي نفس خطوات المنهج العلمي في البحث ، حيث يقدم علي عمليتين منطقيتين هما القياس و الاستنتاج ، وإذن يقوم العالم بملاحظة الحقائق في البداية ثم يجري تجاربه ويرصد عددا من النتائج التي يستخلصها من تلك التجارب بنمط أو إطار عام للظاهرة. وبعد أن يقوم بصياغة نظريته علي ذلك النحو ، ينتقل إلي عملية الاستنتاج التي تعينه علي التنبؤ بسلسلة من النتائج الأخرى .

الفصل الثاني

مبادئ نظرية العينات

1-1 مقدمة

تهتم نظرية العينات بدراسة العلاقة بين المجتمع والعينات المسحوبة منه فيما يسمى بالاستدلال الإحصائي ، **statistical inference**. هناك عدة طرق لأخذ العينات من المجتمع لاستخدامها في الاستدلال الإحصائي ومن أشهر هذه الطرق هي العينة العشوائية وهي العينة التي تكون لكل مفرده من مفردات المجتمع نفس فرصة الاختيار في العيينة. فمثلاً نستعين بعينه مسحوبة من المجتمع لتقدير معالم هذا المجتمع مثل متوسطه أو تباينة أو غير ذلك. أو أعطاء عينه من المرضى بارتفاع الضغط، مثلاً دواء معين ثم قياس ضغطهم قبل وبعد تناولهم لهذا الدواء لمعرفة ما إذا كان هذا الدواء مفيد في خفض الضغط أم لا.

أي مجموعات من المفردات تشترك في صفه أصفات وتكون موضوع دراسة أو بحث فإن هذه المجموعة يطلق عليها إحصائيا مجتمع الدراسة (أو اختصاراً المجتمع **Population**). والمجتمع قد يكون مجموعة ما من البشر أو أشجار أنواع معينه من الفاكهة أو الحيوانات الزراعية أو إنتاج دولة ما لسلع معينه خلال فترة زمنية محدده...الخ.

والمجتمع قد يكون محدوداً إذا كان يمكن حصر عدد أفراده مثل سكان مدينة ما أو طلاب مرحلة تعليمية معينة ، وقد يكون المجتمع غير محدود (لانهائي) إذا كان لا يمكن حصر عدد أفراده مثل النجوم والكواكب أو الكائنات الحية بمياه المحيطات والأنهار وعند دراسة صفة ما أو صفات معينه لمجتمع ما ، فإن البيانات الإحصائية عن تلك الصفة أو الصفات تجمع بأحد أسلوبين

أولاً: أسلوب الحصر الشامل (**census**) وفيه تجمع البيانات عن كل مفرده من مفردات المجتمع، وهذا الأسلوب يتطلب وفرة في الوقت والمال والمجهود الفني وتزداد هذه المتطلبات وتتضاعف كلما ازداد حجم المجتمع (عدد أفراد المجتمع). وهذا الأسلوب لا يتبع عادة إلا في حالة التعدادات التي تجريها الدول وتدعمها بإمكانيات ضخمه مثل تعدادات السكان والتعدادات الصناعية والتعدادات الزراعية.

الثاني: أسلوب المعاينة (**Sampling method**) وفيه يتم جمع البيانات عن جزء من مفردات المجتمع يختار بطريقة أو بأخرى ويطلق عليه عينه (**Sample**) ثم بعد ذلك يتم تعميم نتائج الدراسة على المجتمع بأكمله. أي أن أسلوب المعاينة يقصد به دراسة خصائص المجتمع من خلال دراسة عينه مسحوبة منه ، ونجاح هذا الأسلوب يعتمد على أن تحمل العينة أقصى درجة من دقة التمثيل للمجتمع المسحوبة منه.

من الأفضل في بعض الحالات الحصول على معلومات دقيقة عن طريق التعداد التام أو الحصر الشامل لجميع عناصر المجتمع، لكن لاستخدام أسلوب المعاينة فوائد جمة مقارنة بالتعداد الشامل يرد بيانها في الفقرة التالية.

2-1 بعض مزايا أسلوب المعاينة

يتميز أسلوب المعاينة عن أسلوب الحصر الشامل بمزايا عديده منها:

1- يؤدي استخدام العينات العشوائية إلى خفض تكاليف الدراسات الميدانية بسبب صغر حجم العينة بالنسبة إلى حجم المجتمع وهو ما يؤدي إلى تخفيض الأعباء الإدارية والفنية التي تتطلبها أي دراسة ميدانية.

2- يتحقق وفر واضح في الوقت الذي ينفق في دراسة ميدانية على أساس عينة بدلاً من الحصر الشامل وتتضح أهمية الوقت عندما نقوم بدراسة ظاهرة تتغير بمرور الوقت ، وعينة قد يترتب على دراسة تلك الظاهرة في المجتمع كله بجمع البيانات من جميع مفردات المجتمع أن يمر وقت بديل فتكون البيانات والنتائج وقت ظهورها غير مطابقة لواقع المجتمع وتصبح النتائج ذات قيمه محدودة بعد أن فقدت عنصر المطابقة مع واقع الظاهرة وتوزيعها الحالي لمجتمع ، والتعدادات الدورية للسكان وبسبب ضخامة حجم العمل بها تستغرق وقتاً طويلاً حتى تصبح نتائجها جاهزة ومنشورة وقد يطول هذا الوقت إلى أكثر من ثلاث أو أربع سنوات حتى مع استخدام أحدث أجهزة الحاسبات الأليه الضخمة ، ويكون على الباحثين مستخدمي هذه النتائج مراعاة الوقت الذي ينقض بين تاريخ إجراء التعداد وتاريخ نشر نتائجه وتعديل هذه النتائج في حدود ذلك .. وهذا دفع الكثير من الدول إلى تعزيز نتائج التعدادات الدورية للسكان بنتائج تعدادات تجري بين كل تعدادين متتاليين على أساس العينة.

3- في المجتمعات غير المحدودة (اللانهائية) مثل مجتمع الكائنات الحية في البحار والمحيطات لا يمكن أن تتم الدراسة على أساس الحصر الشامل ولكن لابد وأن تتم الدراسة بأسلوب المعاينة.

4- أيضاً هناك بعض الاختبارات لابد وأن تتم بأسلوب المعاينة لأن إجراء مثل هذه الاختبارات على أساس الحصر الشامل يؤدي إلى تلف المادة المختبرة أو هلاكها.. فاختبار صلاحية شحنه من المفرقعات مثلاً لابد وأن يتم على أساس العينة وبالمثل تحليل دم المرضى يتم على أساس عينه.

3-1 أقسام العينات

تنقسم العينات عادة إلى قسمين رئيسين وهما عينات عشوائية وعينات غير عشوائية، وفيما يلي تفصيل لكل قسم منها.

1-3-1 العينات العشوائية

هي تلك العينات التي يتم اختيار مفرداتها حسب خطه إحصائية لا يكون فيها للباحث أو لمفردات العينة دخل في اختيار أي مفرده فيها ، حيث يتم الاختيار باستخدام أساليب معينة تلعب الصدفة خلالها الدور الأول في اختيار المفردة ولكن بشرط أن يتحقق لجميع المفردات احتمال ثابت ومحدد للاختيار. والعينات العشوائية إذا ما تم اختيارها بالطريقة العلمية السليمة والمناسبة يمكن أن تكفل درجه عالية من دقة التمثيل للمجتمعات المسحوبة منها لذلك فهي الوسيلة الأساسية في حالة البحوث العلمية الدقيقة .. من أهم أنواع العينات العشوائية مايلي.

(أ) العينة العشوائية البسيطة: **Simple random sample**

ويلجأ إليها الباحث في حالة ما إذا كان مجتمع الدراسة ليس كبيراً ويحمل قدراً من التجانس بين المفردات للصفة أو الصفات موضع الدراسة. والعينة العشوائية البسيطة تستغل فرص متكافئة لمفردات المجتمع للدخول في العينة ولكن المفردات التي تدخل في العينة تكون عن طريق الصدفة البحتة. والاختيار العشوائي يتم يدوياً عن طريق بطاقات متماثلة في الحجم واللون أو عن طريق جداول الأعداد العشوائية أو عن طريق الحاسب الآلي. ولكي يتحقق ذلك فإن الأمر يتطلب تحديد مفردات المجتمع تحديداً كاملاً ويكون هذا التحديد على شكل قائمة (أو خريطة) تضم كل مفردات المجتمع وهذه القائمة تسمى الإطار(Frame) ولا يجوز الاختيار العشوائي إلا من المفردات التي يضمها الإطار.

(ب) العينة المنتظمة: Systematic sample

اختيار هذه العينة يتطلب وجود إطار للمجتمع كما في حالة العينة العشوائية البسيطة بحيث يعطى لكل مفرده من مفردات المجتمع رقماً متسلسلاً داخل الإطار ، ثم نختار مفردات العينة من الإطار بحيث يكون الرقم المتسلسل لكل مفردة يبعد بعداً ثابتاً منتظماً عن رقم المفردة السابقة لها وكذلك رقم المفردة اللاحقة لها. فمثلاً إذا كان لدينا مجتمعاً حجمه **2000** مفرده ونريد اختيار عينه منتظمة حجمها **100** مفرده فإننا نقسم الإطار إلى فترات منتظمة طول كل فترة = $\frac{2000}{100}=20$ مفردة ومن داخل مفردات الفترة الأولى **(1 – 20)** يختار مفرده واحدة عشوائياً ولتكن رقم **14** مثلاً وبناء على رقم تلك المفردة يتحدد باقي مفردات العينة المنتظمة فتكون هي المفردات ذات الأرقام **34**، **54**، **....**، **1974**، **1994.**

والعينة المنتظمة كثيرة الاستعمال في التطبيقات العملية لقلة تكاليفها وقلة الأخطاء التي ترتكب في اختيار مفردات العينة فضلاً عن سهولة إجرائها. ولكن أهم عيوب المعاينة المنتظمة هو عدم صلاحيتها إذا ما وجدت علاقة دورية مع ترتيب العناصر في القائمة وكان طول الفترة بين عناصر العينة مساوياً لطول الدورة أو إحدى مضاعفاتها.

(جـ) العينة العشوائية الطبقية: Stratified random sample

ويلجأ إليها الباحث في حالة ما إذا كان مجتمع الدراسة واضحاً به فئات (طبقات) بحيث أن التجانس أو التقارب داخل كل طبقة من طبقات مجتمع الدراسة أكبر من التجانس داخل المجتمع ككل (أي أن التشتت داخل المجتمع ككل أكبر من التشتت داخل كل فئة من فئاته على حده). في هذه الحالة يجب على الباحث مراعاة أن الطبقة داخل العينة بنفس نسبة وجودها داخل المجتمع (وأحياناً يوضع في الاعتبار عناصر أخرى مثل التشتت داخل الطبقة أو عنصر التكلفة لجمع البيانات عن الطبقة). بعد أن يتم تحديد عدد المفردات التي يجب سحبها من كل طبقة للدخول في العينة فإن هذه المفردات يتم سحبها عشوائياً من داخل الطبقة ومجموع هذه المفردات تكون العينة الطبقية العشوائية.

(د) العينة متعددة المراحل أو العنقودية: clustered sample

يلجأ إليها الباحث عندما يكون مجتمع الدراسة كبير جداً ومتناثراً على مساحات شاسعة تكلف الكثير من الوقت والجهد في التنقل بينها عند جمع البيانات، أيضاً في حالة عدم وجود إطار يضم جميع مفردات المجتمع فيستحيل الاختيار العشوائي مباشر من المجتمع. لهذا يلجأ الباحث إلى أخذ العينة على مراحل متعددة متتالية. في المرحلة الأولى يتم تقسيم المجتمع إلى عدد محدد من وحدات المعاينة

الكبيرة الحجم ومنها يختار بعضها عشوائياً ثم يتلو ذلك كمرحلة ثانية تقسيم الوحدات المختارة عشوائياً من المرحلة الأولى إلى وحدات أقل منها في الحجم ثم يختار بعضها عشوائياً.. وهكذا تتابع مراحل التقسيم والاختيار العشوائي ، وعدد هذه المراحل ليس ثابت بل يتوقف على طبيعة مجتمع الدراسة وإمكانيات الباحث .. في المرحلة الأخيرة يصل الباحث إلى وحدات المعاينة التي سيجمع عنها بيانات البحث ويطلق عليها وحدات المعاينة الأولية.

2-3-1 العينات غير العشوائية:

هي تلك العينات التي لا تكفل لجميع مفردات المجتمع احتمال ثابت ومحدد للاختيار، وغالباً يتدخل الباحث في عملية الاختيار بصورة أو بأخرى ... ومن أهم أنواع العينات غير العشوائية:

(أ) العينة العمدية أو المقصودة: Purposive sample

يلجأ الباحث إلى هذه الطريقة فيما إذا كان مجتمع الدراسة كبير جداً وكانت إمكانياته لا تسمح له إلا بدراسة عينة حجمها صغير جداً بالنسبة لمجتمع الدراسة، في هذه الحالة يتعمد الباحث اختيار مفردات معينة كعينة لمجتمع الدراسة يرى بخبرته السابقة أن هذه العينة يمكن أن تعطي تمثيلاً مقبولاً لمجتمع الدراسة.

مثلاً إذا أراد باحث دراسة خصائص اقتصادية أو اجتماعية معينه عن ريف دولة ما ، وكانت إمكانياته المالية والإدارية لا تسمح له بعينة سوى سكان قرية واحدة ، فإنه في هذه الحالة إذا ما تم اختيار القرية عشوائياً من بين آلاف القرى بتلك الدولة فإن الصدفة قد تأتي بقرية بعيدة في خصائصها (من حيث الظاهرة موضوع الدراسة) عن خصائص معظم قرى تلك الدولة ... كأن تأتي

بالصدفة قرية ساحلية معظم سكانها من الصيادين أو قرية قريبة من مشروع صناعي ضخم يستوعب في قواه العاملة معظم سكانها .. هذه القرية أو تلك قد يأخذ النمط المعيشي لسكناها طابعاً خاصاً _

نابعاً عن ظروفها الخاصة _ بعيداً عن النمط المعيشي المعتاد لبقية القرى، لذلك فأي منها لا يمكن أن

يعطي تمثيلاً مقبولا لريف تلك الدولة . لهذا فإن الباحث وعلى ضوء خبراته السابقة يتعمد اختيار قرية

معينة يرى أنها _ من وجهة نظره الشخصية_ يمكن أن تمثل الريف. وهذه الطريقة غير علميه وغالباً يتم اللجوء إليها في حالة البحوث التمهيدية.

(ب) العينة الحصصيه: Quota sample

وهي نوع خاص من العينات غير العشوائية وتستخدم كثيراً في معاينة الرأي العام (على سبيل المثال عمليات استطلاعية الرأي العام التي يقوم بها معهد جالوب قبل إجراء انتخابات الرئاسة في الولايات المتحدة الأمريكية)... في هذه الطريقة يقسم المجتمع موضوع الدراسة إلى طبقات بالنسبة إلى صفات أو خصائص معينة ويتم العمل على تمثيل كل طبقة منها في العينة بنسبة وجودها في المجتمع الأصلي (وعلى سبيل المثال في حالة دراسة الدخل لمنطقة ما ورؤى أن يكون حجم العينة المطلوبة 100 فرد مثلاً عندما يريد الباحث أن يقوم جامعوا البيانات بالحصول على البيانات من 20 موظفاً، 45 من العمال الحرفيين ، 35 من ذوي الأعمال الحرة .. وتترك الحرية لجامعي البيانات في اختيار الأفراد المطلوبة فيها حدود المواصفات الموضوعة لكل طبقة من الطبقات المذكورة.
واضح أنه رغماً من أن هذه الطريقة في ظاهرها مماثلة للعينة الطبقية العشوائية.. إلا أنه في الحالة الأخيرة (العينة الطبقية العشوائية) يكون اختيار المفردات عشوائياً من داخل كل طبقة ولا يترك لجامع البيانات حرية اختيار المفردات من كل طبقة والذي قد يترتب عليه تميزاً كبيراً.
عموماً.. يلجأ الباحث إلى العينة الحصصيه إذا كان من المرغوب فيها اظهار النتائج في وقت قصير مع التغاضي عن توافر درجة دقة عالية بتلك النتائج.

4-1 أخطاء البيانات الإحصائية

تتعرض البيانات الإحصائية التي يتم جمعها إلى نوعين من الأخطاء:

1-4-1 خطأ التميز

وهو ينتج عن مصادر متعددة، منها أخطاء في تصميم البحث أو التجربة أو أخطاء فنيه أثناء جمع البيانات أو خلال العمليات الحسابية التي تتم على البيانات المتجمعة.. أخطاء التميز تزداد بازدياد

الفروق بين الإمكانيات (المادية والفنية) اللازم توافرها لضمان أقصى درجة دقة ممكنه وبين الإمكانيات الفعلية المتاحة للباحث.

أخطاء التميز قد توجد في البيانات التي يتم جمعها بأسلوب الحصر الشامل وقد توجد أيضاً في البيانات التي يتم جمعها بأسلوب المعاينة، ولكنها إن وجدت فهي غالباً أكبر في الحالة الأولى (الحصر الشامل) مما هي عليه في الحالة الثانية (المعاينة) باعتبار أن حجم العمل في تلك الحالة يكون أقل وبالتالي قد يسهل توفير الإمكانيات اللازمة وتجنب الأخطاء الفنية.

2-4-1 خطأ المعاينة العشوائية أو خطأ الصدفة

وهو الخطأ الناتج عن فروق الصدفة بين مفردات المجتمع التي دخلت العينة وبين تلك المفردات التي لم تشأ الصدفة أن تدخل العينة.. وخطأ الصدفة يمكن تقليل قيمته إذا ما تم اختيار العينة بالطريقة المناسبة وإذا ما كان حجم العينة مناسباً لحجم المجتمع وخصائصه.

5-1 المعالم والإحصاءات

المقاييس الإحصائية التي تحسب من بيانات مجتمع الدراسة بأكمله يطلق عليها معالم المجتمع **(Parameters of population)**، أما المقاييس الإحصائية التي تحسب من بيانات عينه مسحوبة من مجتمع الدراسة فيطلق عليها إحصاءات **(Statistics)** ويعتبر كل إحصاء منها بمثابة تقدير أو قيمه تقديرية لمعلمة المجتمع المناظر، فيكون المتوسط الحسابي المحسوب من بيانات العينة تقدير لمعلمة المجتمع المناظرة وهي المتوسط الحسابي المسحوب منه هذه العينة وهكذا.. ويجب ألا يغيب عن الأذهان بأن حساب قيمة المتوسط الحسابي من بيانات العينة ليس هدفاً في حد ذاته ولكن وسيلة للتعرف على المتوسط الحسابي للمجتمع موضوع الدراسة.. وهكذا بالحال بالنسبة لباقي المقاييس الإحصائية التي تحسب من العينة.

للتفرقة بين المعالم والإحصاءات يجب أن نرمز لكل منها برموز تختلف عن رموز الأخرى، على سبيل المثال يرمز للمتوسط الحسابي للمجتمع بالرمز μ بينما يرمز للمتوسط الحسابي للعينة بالرمز $\bar{x}$ ، أيضاً للانحراف المعياري للمجتمع بالرمز σ بينما يرمز للانحراف المعياري للعينة بالرمز S وهكذا.

1-5-1 توزيعات المعاينة: Sampling Distributions

نفرض أننا أخذنا عينه حجمها n من مجتمع ما ، ثم سحبنا منها بعض المقاييس الإحصائية مثل المتوسط الحسابي ، التباين ، ... فإن كل مقياس من هذه المقاييس يعتبر متغير عشوائي في ذاته يختلف من عينه إلى أخرى – هذا المتغير العشوائي يخضع لتوزيع معين – هذا التوزيع يسمى بتوزيع العينة. فمثلاً نقول أن توزيع المعاينة للمتوسط الحسابي وهو عبارة عن توزيع جميع المتوسطات الحسابية للعينات المأخوذة من نفس هذا المجتمع ذات الحجم n ، وكذلك فإن توزيع المعاينة للتباين هو توزيع جميع التباينات المحسوبة من عينات لها نفس الحجم n ومأخوذة من نفس المجتمع ، وهكذا ... .

2-5-1 توزيعات المعاينة للأوساط: Sampling Distributions of Means

نفرض أننا سحبنا عينه حجمها n من مجتمع لانهائي ، القيمة المتوقعة له تساوي μ والانحراف المعياري هو σ فإن المتوسط الحسابي $\overline{X}$ يخضع لتوزيع ما ، متوسط هذا التوزيع وانحرافه المعياري هما

$$\mu_{\overline{X}} = \mu \quad ، \quad \sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} \qquad (4\text{-}1)$$

وفي الحالة التي يكون فيها المجتمع الأصلي المسحوبة منه العينة مجتمع طبيعي (ويرمز له بالرمز $N(\mu , \sigma^2)$ فإن توزيع المعاينة للمتوسط الحسابي $\overline{X}$ يكون في هذه الحالة توزيع طبيعي أيضاً له نفس المتوسط الأصلي μ ولكن انحرافه المعياري يساوي $\sigma / \sqrt{n}$، أي بمعنى أن

$$X \sim N(\mu,\sigma^2) \quad \Rightarrow \quad \overline{X} \sim N(\mu,\frac{\sigma^2}{n}) \qquad (4\text{-}2)$$

ومن ثم يكون

$$z = \frac{\sqrt{n}(\overline{x} - \mu)}{\sigma} \sim N(0,1) \qquad (4\text{-}3)$$

أما إذا كان المجتمع غير طبيعي فإن $\overline{X}$ لا تخضع للتوزيع الطبيعي ولكنها تتوزع توزيع يكون قريباً من التوزيع الطبيعي لقيم n الكبيرة $(n \geq 30)$ حيث أن

$$z = \frac{\sqrt{n}(\overline{x} - \mu)}{\sigma} \xrightarrow[n \to \infty]{as} N(0,1) \qquad (4\text{-}4)$$

وتعتبر النتيجة السابقة الهامة جداً في الإحصاء وخاصة في التطبيقات العلمية وتسمى نظرية النهاية المركزية Central Limit Theorem والتي تنص على أنه في حالة العينات الكبيرة الحجم فإن المتوسط الحسابي $\bar{X}$ يخضع للتوزيع الطبيعي بالمعاملات μ و $\frac{\sigma^2}{n}$ ، حيث أن μ, σ^2 هما متوسط وتباين المجتمع الأصلي بغض النظر عن شكل توزيع المجتمع الأصلي. ومن ثم فإنه لقيم n الكبيرة تتحقق العلاقة (4-3) بصرف النظر عن توزيع المجتمع الأصلي.

كذلك فإنه إذا كان $\bar{X}_1$ هو المتوسط الحسابي لعينه عشوائية مسحوبة من مجتمع لانهائي متوسطه هو μ_1 وانحرافه المعياري هو σ_1، وكان $\bar{X}_2$ هو المتوسط الحسابي لعينة عشوائية مسحوبة من مجتمع لانهائي آخر متوسط μ_2 وانحرافه المعياري σ_2 وكانت العينتين مستقلتين فإن المجموع الجبري لمتوسط العينتين يخضع لتوزيع المعاينة بالمعاملات

$$\mu_{(\bar{x}\pm\bar{x}_2)} = \mu_1 \pm \mu_2 \quad \text{and} \quad \sigma^2_{(\bar{x}_1\pm\bar{x}_2)} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \qquad (4\text{-}5)$$

حيث n_1, n_2 هما حجم العينة الأولى والثانية.

وإذا كان المجتمعين الأصليين طبيعيين فإن $(\bar{x}_1 \pm \bar{x}_2)$ يخضع لتوزيع طبيعي أيضاً بالبارامترات المعطاة في (4-5) وعليه فإنه في هذه الحالة

$$z = \frac{(\bar{x}_1 \pm \bar{x}_2) - (\mu_1 \pm \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1) \qquad (4\text{-}6)$$

أما إذا كان أحد المجتمعين أو كليهما لا يتوزع توزيعاً طبيعياً فإن $(\bar{x}_1 \pm \bar{x}_2)$ لا يتوزع توزيعاً طبيعياً كذلك ، ولكن لقيم n_1, n_2 الكبيرة فإنه طبقاً لنظرية النهاية المركزية السابقة فإن $(\bar{x}_1 \pm \bar{x}_2)$ يتوزع توزيعاً قريباً من التوزيع الطبيعي وبذلك يمكننا استخدام نفس العلاقة (4-6)في حالة العينات الكبيرة.

3-5-1 توزيع المعاينة للتباين: Sampling Distribution of The Variance

إذا كان $s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$ هو تباين عينه عشوائية حجمها n مأخوذة من مجتمع متوسطه μ وتباينه σ^2 وعزمه الرابع حول المتوسط هو μ_4 فإن

$$\mu_{s^2} = \sigma^2 \quad \text{and} \quad \sigma^2{}_{s^2} = \frac{\mu_4 - \sigma^2}{n-1} \qquad (4\text{-}7)$$

وإذا كان المجتمع طبيعي فإن $\mu_4 = 3\sigma^2$ وبالتالي فإن

$$\sigma^2_{s^2} = \left(\frac{2}{n-1}\right)\sigma^2 \qquad (4\text{-}8)$$

نلاحظ هنا أن s^2 لا تتوزع طبيعي حتى ولو كان المجتمع طبيعي ، ولكنه يتوزع توزيع قريب من التوزيع الطبيعي وذلك لقيم nالكبيرة $(n\geq 100)$.   أما إن كان المجتمع الأصلي يخضع للتوزيع الطبيعي فإن المتغير $(n-1)s^2/\sigma^2$ يخضع لتوزيع يسمى توزيع مربع كاي χ^2 بعدد درجات حرية يساوي $n-1$. أي أن

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1) \qquad (4\text{-}9)$$

ويعتبر توزيع مربع كاي من التوزيعات الهامة في الإحصاء التطبيقي ودالة كثافته هي

$$f(y) = y^{\frac{v-1}{2}} e^{-y/2} \quad , \quad y > 0 \qquad (4\text{-}10)$$

حيث v هي عدد درجات الحريه للتوزيع وتعتبر هي المعامل الوحيد له ويتضح من شكل الدالة أنها داله متصلة وتقع بأكملها فوق النصف الموجب لمحور السينات ، منحنى هذه الداله غير متماثل ويعتبر من المنحنيات موجبة الالتواء ويقل التواؤه (وبالتالي يقترب من التماثل) كما زادت درجات الحريه v. وتكون القيمة المتوقعة لهذا التوزيع هي v و تباينه هو $2v$ أي بمعنى أن

$$E(y) = \mu_y = v$$

$$V(y) = \sigma^2 = 2v \qquad (4\text{-}11)$$

فإذا كان s_1^2 هو تباين عينه عشوائية حجمها n_1 مسحوبة من مجتمع طبيعي $N(\mu_1, \sigma_1^2)$ ، وكان s_2^2 هو تباين عينه عشوائية أخرى حجمها n_2 ومسحوبة من مجتمع طبيعي آخر $N(\mu_2, \sigma_2^2)$ وكانت العينتان مستقلتان فإن المتغير:

$$\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \sim F(n_1 - 1, n_2 - 1) \qquad (4\text{-}12)$$

حيث أن $F(n_1-1, n_2-1)$ تسمى بتوزيع F بدرجتي الحريه n_1-1 و n_2-1 و دالة الكثافة الإحتماليه للمتغير y الذي يخضع لتوزيع F بدرجتي الحريه v_1, v_2 تعطى بالصورة:

$$f(y) = \frac{y^{\frac{v_1}{2}-1}}{(v_1 y + v_2)^{\frac{v_1+v_2}{2}}} \quad , \quad y > 0 \qquad (4\text{-}13)$$

وكما يتضح من الداله في (4-13) أن المنحنى يقع بالكامل في النصف الموجب لمحور السينات كما في حالة توزيع χ^2 ، وهو أيضاً غير متماثل وموجب الالتواء ولكن يقترب من التماثل كلما زادت درجات الحريه v_1, v_2.

ذكرنا سابقاً أنه إذا كان $\bar{X}$ هو المتوسط الحسابي لعينه حجمها n مأخوذة من مجتمع طبيعي بالمعاملات μ, σ^2 فإن

$$z = \frac{\sqrt{n}\,(\bar{x}-\mu)}{\sigma} \quad \sim \quad N(0,1)$$

هذا إذا كانت σ معلومة ، ولكن في حالة ما إذا كانت قيمة σ غير معلومة فإننا نستخدم بدلا منها الانحراف المعياري للعينة S ولكن في هذه الحالة يصبح المتغير $\frac{\sqrt{n}\,(\bar{x}-\mu)}{S}$ يخضع لتوزيع يعرف بتوزيع t ستيودنت $t-$student بدرجات حريه $n-1$ ، أي أن

$$t = \frac{\sqrt{n}\,(\bar{x}-\mu)}{s} \quad \sim \quad t(n-1) \qquad (4\text{-}14)$$

دالة الكثافة لتوزيع t بدرجات حريه v تعطي بالصورة:

$$f(t) = \left(1+\frac{t^2}{v}\right)^{-\frac{v+1}{2}} \qquad (4\text{-}15)$$

وهو توزيع متماثل حول محور y وهو يشبه في ذلك المنحنى الطبيعي القياسي $N(0,1)$ ولكنه أقل تحدباً من التوزيع الطبيعي القياسي ولكنه يقترب من التوزيع الطبيعي كلما زادت درجات الحريه.

وإذا كان $\bar{X}_2$ و S_1^2 هما المتوسط الحسابي والتباين لعينه حجمها n_1 مأخوذة من مجتمع طبيعي متوسط هو μ_1 وكان $\bar{X}_2$ و S_2^2 هما المتوسط الحسابي والتباين لعينه أخرى حجمها n_2 ومأخوذة من مجتمع طبيعي آخر له المتوسط μ_2 وكانت العينتان مستقلتان فإن المتغير

$$t = \frac{(\bar{X}_2 \pm \bar{X}_2) - (\mu_1 \pm \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2) \qquad (4\text{-}16)$$

حيث أن $S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$ يسمى بالتباين المشترك للعينتين The Pooled Variance.

المتغيرات Variables :

تشير كلمة المتغيرات إلى الخصائص التى تشترك فيها أفراد المجتمع الاحصائى ولكنها تختلف من فرد إلى فرد آخر فالعمر ، درجة الذكاء ، وطول القامة ، واللياقة البدنية والقدرة على القراءة ، والدخول التى يحصل عليها الأفراد أمثلة للمتغيرات وتتميز هذه المتغيرات بأنها قابلة للقياس الكمى وبإمكانية تحديد قيمة معينة لها .

والمتغيرات عبارة عن ظاهرات أو صفات تختلف قيمها باختلاف الحالات . ومن أمثلتها : درجة الحرارة فى مناطق مختلفة أو فى فترات مختلفة لمكان واحد ، كميات الإنتاج الزراعى أو الصناعى .

والمتغيرات التى تقاس كميا تنقسم من حيث قيمتها العددية إلى نوعين:

1 – المتغير المتصل Continuous Variable .

لما كان التعريف العام للمتغير Variable هو ظاهرة أو صفات تختلف قيمها باختلاف الحالات فان المتغير يكون متصلا عندما يأخذ أى قيمة متدرجة على المقياس المستخدم . مثال ذلك قياس درجات الحرارة باستخدام الترمومتر فالمتغير يأخذ أى قيمة بين رقمين صحيحين ، بمعنى أن المتغير يمكن أن يأخذ أى قيمة بين 36 درجة ، 37 درجة (36.1 ، 36.2 الخ) .

2 – المتغير المتقطع Discrete Variable

عندما يأخذ المتغير قيما محددة يطلق عليه متغيرا متقطعاً أو بمعنى آخر ، المتغير المتقطع هو الذى يحتوى مداه على عدد محدود من القيم أو يحتوى عدد لانهائي من القيم ولكن لكل منها قيمة

محددة يمكن عدها أو ترتيبها فى نهاية الأمر تعدد الأولاد أو الأفراد فى الأسرة لابد أن يكون أعدادا صحيحة غير حقيقة مثل 1 ، 2 ، 3 ، 4 وهكذا ومن أمثال المتغيرات المتقطعة ، النوع ، الحالة الزواجية Martial Status ، عدد أيام الإنتاج فى احد المصانع ، عدد حوادث السيارات وهكذا .

كما يمكن تصنيف المتغيرات إلى عدد من التصنيفات بحسب الغاية من كل تصنيف وذلك على النحو التالى : –

1 – المتغيرات الكمية والمتغيرات الكيفية :

يمكن تصنيف المتغيرات من حيث طريقة التعبير عنها إلى فئتين هما : المتغيرات الكمية Quantitative Variables وهى التى يمكن أن نصفها عدديا بأنها اكبر من أو أقل من قيمة معينة ويعتبر العمر وعدد سنوات التعليم أمثلة لهذه المتغيرات . والفئة الثانية من المتغيرات هى المتغيرات الكيفية Qualitative Variables وهى التى تصف الأشياء بصفاتها مثل متغير النوع الذى ينقسم إلى قسمين : ذكور وإناث . والحالة العملية للفرد حيث تكون إما مزارع أو عامل غير ماهر ، أو عامل ماهر أو موظف أو تاجر وما إلى ذلك من صفات ، وهذه المتغيرات الكيفية يتعذر معالجتها إحصائياً ما لم يميزها عن بعضها بعضاً باستخدام الأرقام فنرمز لمتغير الإناث برقم 1 و لمتغير الذكور برقم 2 أو العكس ، والرقم فى هذه الحالة لا يعنى أكثر من أنه أداه للتميز بين المتغيرات الكيفية لتسهيل تفريغ البيانات التى جمعت عنها من ميدان الدراسة تمهيداً لمعالجتها إحصائياً ولا تكون لها قيمة عددية فى حد ذاته .

2- المتغيرات التابعة والمستقلة والضابطة :

ويمكن تصنيف المتغيرات تصنيفاً أخر بحسب دورها فى حدوث الظاهرة محل الدراسة وذلك إلى :

(أ) متغيرات تابعة Dependent Variables

وهى تلك المتغيرات التى نحاول تفسيرها ومعرفة أسباب حدوثها وتحديد مدى إمكان التنبؤ بها .

(ب) متغيرات مستقلة Independent Variables

وهى التى لعبت دوراً مباشر فى حدوث المتغيرات التابعة ونستخدمها فى تأييد تفسيرنا وفهمنا لما طرأ على هذه المتغيرات من تغيير ، وفى التنبؤ بالحالة التى ستؤول إليها بعد ذلك .

(ج) متغيرات وسيطة Intermediate Variables

وهى تلك المتغيرات التى يمر من خلالها تأثير المتغيرات المستقلة إلى المتغيرات التابعة والمتغيرات الوسيطة بالغة الأهمية فى تفسير حدوث الظواهر الاجتماعية إذ قد يغفل عنها الباحثون أو قد ينظرون إليها على أنها متغيرات مستقلة لارتباطها المباشر بالمتغيرات التابعة .

الفصل الثالث

تبويب وعرض البيانات

أولاً : العرض الجدولى للبيانات الإحصائية .

- تبويب البيانات الخام فى جدول تكرارى بسيط .
- تبويب البيانات فى جدول تكرارى ذو فئات .
- تبويب البيانات فى الجدول التكرارى المتجمع الصاعد .
- تبويب البيانات فى الجدول التكرارى المتجمع الهابط .
- الجدول المزدوج .

ثانياً : العرض البيانى للبيانات الإحصائية .

- العرض البيانى للبيانات الغير مبوبة .
  1. طريقة الأعمدة البيانية البسيطة .
  2. طريقة المنحنى البيانى البسيط .
  3. طريقة الخط البيانى المنكسر .
  4. طريقة الدائرة البيانية .
  5. طريقة الأعمدة البيانية المتلاصقة .
  6. طريقة الأعمدة البيانية المجزأة .
- العرض البيانى للبيانات الغير مبوبة .
  1. المدرج التكرارى .
  2. المضلع التكرارى .
  3. المنحنى التكرارى .

تبويب البيانات :

يقصد بتبويب البيانات عرض هذه البيانات (البيانات الخام) فى جداول مناسبة وذلك حتى يمكن تلخيصها وفهمها واستيعابها واستنتاج النتائج منها ومقارنتها بغيرها من البيانات ، كما يسهل الرجوع إليها فى صورة جداول دون الاطلاع على الاستمارات الأصلية التى قد تحمل أسماء أصحابها مما يخل بمبدأ سرية البيانات الإحصائية .

كما يعتبر عرض وتبويب البيانات الإحصائية الخطوة الثانية (بعد تجميع هذه البيانات الخام) في مفهوم التحليل الإحصائي، ويلجأ الباحث إلى حصر وتصنيف هذه البيانات وعرضها بطريقة مختصرة تساعد على فهمها وتحليلها إحصائياً للتعرف عليها ووصفها ومقارنتها بغيرها من الظواهر ، والخروج ببعض المدلولات الإحصائية عن مجتمع الدراسة .

عرض البيانات :

تتوقف طريقة عرض البيانات على نوع هذه البيانات وعلى الحقائق المطلوب إبرازها. وهناك طريقتان أساسيتان لعرض وتبويب البيانات الإحصائية وهما :

أولاً : العرض الجدولى للبيانات الإحصائية :

بعد عملية تبويب وتعيين الصفات التى تميز المفردات ، ترصد النتائج فى جداول مناسبة توضح الشكل النهائى للمجموعات المميزة وتسمى هذه العملية التى يتم تجميع البيانات فى مجموعات مميزة ومتجانسة بعملية التصنيف وتصنف البيانات الإحصائية بوجه عام وفقاً لإحدى القواعد التالية :

1- تصنيف جغرافى

2- تصنيف تاريخى أو زمنى .

3- تصنيف نوعى أو وصفى .

4- تصنيف كمى .

ويمكن التمييز بين مجموعة أشكال من الجداول الإحصائية نذكرها فيما يلى :

تبويب البيانات الخام فى جدول تكرارى بسيط :

والمقصود بالجدول البسيط هو ذلك الجدول الذى يتم وضع قيم الدرجات فيه مرتبة ترتيباً تصاعدياً فى عموده الأول أما العمود الثانى فيسمى بعمود التكرار ويرصد فيه عدد مرات تكرار كل درجة أو حدث .

مثال :

البيانات التالية هى درجات حصل عليها عشرون طالباً فى مادة الإحصاء الاجتماعي بالفرقة الأولى قسم الاجتماع فى امتحان نهاية العام :

10	12	13	15	10	12	14	15	11	12
15	12	10	12	13	15	12	13	10	14

والمطلوب تبويب هذه البيانات فى جدول توزيع تكرارى بسيط ؟

الحل :

يتم ترتيب البيانات دون تكرار تصاعديا ثم وضع هذه البيانات فى العمود الأول من الجدول وتسمى (x) ثم وضع عدد مرات التكرار باستخدام العلامات فى العمود الثانى أما العمود الثالث فيمثل التكرار ويرمز له بالرمز (f) .

x	العلامات	f
10	////	4
11	/	1
12	~~////~~ /	6
13	///	3
14	//	2
15	////	4
مج		20

<u>مثال :</u>

البيانات التالية هى تقديرات 20 طالباً فى مادة الإحصاء بالفرقة الأولى لقسم الاجتماع فى العام الجامعى 2006/2005 والمطلوب هو وضع هذه البيانات فى جدول بسيط ؟

جيد	مقبول	جيد	جيد	مقبول	جيد	جيد جداً	مقبول	جيد	جيد جداً
ممتاز	جيد	ممتاز	جيد جداً	مقبول	جيد	ممتاز	جيد	جيد	مقبول

<u>الحل :</u>

التقدير	التكرار
مقبول	5
جيد	9
جيد جداً	3
ممتاز	3
المجموع	20

<u>تبويب البيانات فى جدول تكرارى ذو فئات :</u>

قبل التعرض إلى إعداد هذا الجدول سنقوم أولاً بالتعرف على معنى الفئات وطرق كتابتها .

<u>المقصود بالفئات :</u>

الفئة هى مجموعة من البيانات متشابهة إلى حد كبير جداً فى الصفات ، وفى حالة زيادة عدد البيانات الخام التى يتم الحصول عليها من الاستبيان لا يمكن استخدام الجداول البسيطة في التعبير عن هذه

الحـالات وإلا سـنحتاج إلـى مئـات الصـفحات ، وإنمـا يـتم تقسـيم البيانـات إلـى مجموعـات متقاربـة ومتشابهة فى الصفات تسمى فئات .

f	x
5	20-10
20	30-20
50	40-30
25	50-40

وتنطق الفئـة الأولـى مثلاً (مـن 20 إلـى 30) ولـيس (20 شـرطة 30) وهـذه الطريقـة معيبـة لأن نـهاية الفئة الأولى هى نفسها بداية الفئة الثانية وهكذا وفى هذه الحالة لا نعرف إلى أى فئـة ينتمـى هذا الرقم .

<u>الطريقة الثانية :</u>

نذكر كلا من الحد الأدنـى والحد الأعلـى للفئـة ولكـن نقـوم بتـرك فاصـل مقـداره الواحـد الصـحيح بـين نـهاية الفئة الأولى وبداية الفئة الثانية وهكذا كما بالجدول التالى .

f	x
5	19-10
20	29-20
50	39-30
25	49-40

ويعاب على هذه الطريقة أنها لا تصلح فى حالة البيانات التى تحتوى على كسور .

الطريقة الثالثة :

نذكر الحد الأدنى فقط للفئة ونضع بعده شرطة وتنطق الفئة الأولى مثلاً (10 إلى أقل من 20) وهذه الطريقة تصلح لكافة الظواهر.

ف	ك
-10	5
-20	20
-30	50
-40	25

الطريقة الرابعة :

نذكر الحد الأعلى فقط للفئة ونضع قبله شرطة وتنطق الفئة الأولى مثلاً (أكثر من صفر الى 20) وهذه الطريقة تصلح لكافة الظواهر أيضاً ولكنها أقل شيوعاً .

ف	ك
20-	5
30-	20
40-	50
50-	25

خطوات بناء جدول التوزيع التكراري ذو الفئات :

1- حساب المدى = أكبر قيمة – أصغر قيمة

2- حساب عدد الفئات = $n * 3.3$

3- حساب طول الفئة = المدى / عدد الفئات

4- اختيار بداية الفئة الأولى أى الحد الأدنى لها مساوى لأقل قيمة موجودة بالبيانات أو أقل بقليل منها فمثلاً تكون من الأرقام الصفرية لتسهيل الحسابات بعد ذلك .

5- بناء الجدول ووضع العلامات التى تمثل التكرار .

<u>مثال :</u>

قام باحث بجمع بيانات تمثل درجات اختبار مادة الحاسب الآلى لخمسين طالباً من طلاب المرحلة الثانية كما فى الجدول التالى :

57	42	51	55	70
53	63	47	60	45
55	82	39	65	33
42	65	61	58	64
55	45	53	52	50
39	63	59	36	25
64	54	49	45	65
78	52	41	42	75
26	48	25	35	30
88	46	55	40	20

والمطلوب هو إعداد جدول توزيع تكرارى ذو فئات للجدول السابق؟

<u>الحل :</u>

- المدى = أكبر قيمة – أصغر قيمة = 88 – 20 = 68
- نختار عدد الفئات = 7
- طول الفئة = المدى / عدد الفئات = 68 / 7 = 9.7
- نقرب طول الفئة لأقرب رقم صحيح فتصبح

  طول الفئة = 10
- نختار بداية الفئة الأولى وهو أصغر رقم = 20
- نبدأ فى بناء الجدول كالتالى :

التكرار f	العلامات	الفئات x
4	~~////~~	-20
6	/ ~~////~~	-30

-40	~~////~~ ~~////~~ //	12
-50	~~////~~ ~~////~~ ////	14
-60	~~////~~ ////	9
-70	///	3
90-80	//	2
المجموع		50

<u>تبويب البيانات فى الجدول التكراري المتجمع الصاعد :</u>

ويقصد بالتكرار المتجمع الصاعد هو تجميع تكرار كل فئة على جميع التكرارات السابقة لها بحيث يكون مجموع التكرار التصاعدى للفئة الأخيرة مساوى لمجموع التكرارات .

<u>مثال :</u>

من نفس بيانات المثال السابق كون جدول التكرار المتجمع الصاعد.

<u>الحل :</u>

بنفس الخطوات السابقة نكون جدول التوزيع التكرارى ذو الفئات ومنه نكون جدول التوزيع التكرارى المتجمع الصاعد كالتالى :

حدود الفئات	التكرارات	التكرار المتجمع الصاعد (ك.م.ص)
أقل من 20	4	صفر
أقل من 30	6	4
أقل من 40	12	10
أقل من 50	14	22
أقل من 60	9	36
أقل من 70	3	45
أقل من 80	2	48
أقل من 90		50

تبويب البيانات فى الجدول التكرارى المتجمع الهابط :

ويقصد بالتكرار المتجمع الهابط هو تجميع تكرار كل فئة على جميع التكرارات التالية لها بحيث يكون مجموع التكرار التنازلى للفئة الأولى مساوى لمجموع التكرارات .

مثال :

من نفس بيانات المثال السابق كون جدول التكرار المتجمع الهابط

الحل :

بنفس الخطوات السابقة نكون جدول التوزيع التكرارى ذو الفئات ومنه نكون جدول التوزيع التكرارى المتجمع الصاعد كالتالى :

حدود الفئات	التكرارات	التكرار المتجمع الهابط (ك.م.هـ)
20 فأكثر	4	50
30 فأكثر	6	46
40 فأكثر	12	40
50 فأكثر	14	28
60 فأكثر	9	14
70 فأكثر	3	5
80 فأكثر	2	2
90 فأكثر		صفر

الجدول المزدوج

وهو الجدول الذى يربط بين متغيرين فى نفس الوقت وكل متغير منهم له فئاته فيتم بناؤه بإتباع عدة خطوات هى :

1- تحديد المتغيرين

2- تحديد المتغير المستقل والمتغير التابع

3- تحديد فئات كل من المتغيرين

4- تكوين الجدول بحيث يحتل المتغير المستقل أعلى الجدول أى يكون أفقياً أما المتغير التابع فيحتل الجزء الأسفل أى يكون عمودياً.

5- وضع العلامات التى تمثل التكرار.

6- إعادة كتابة الجدول بالأرقام .

مثال :

الجدول التالى يوضح البيانات التى حصل باحث فى دراسة بين النوع و مشاهدة البرامج التعليمية لمجموعة من طلاب الصف الثالث متوسط على النحو التالى :

النوع	مشاهدة البرامج	النوع	مشاهدة البرامج
ذكر	يشاهد	ذكر	لا يشاهد
ذكر	يشاهد	أنثى	لا يشاهد
أنثى	يشاهد	أنثى	لا يشاهد
ذكر	لا يشاهد	أنثى	يشاهد
أنثى	يشاهد	ذكر	يشاهد
أنثى	لا يشاهد	ذكر	يشاهد
أنثى	لا يشاهد	ذكر	لا يشاهد
ذكر	لا يشاهد	ذكر	لا يشاهد
ذكر	يشاهد	أنثى	يشاهد
أنثى	لا يشاهد	أنثى	لا يشاهد

والمطلوب تكوين الجدول المزدوج للعلاقة بين المتغيرين (النوع ومشاهدة البرامج التعليمية) ؟

الحل :

1- المتغيرين (النوع – مشاهدة البرامج التعليمية)

2- المتغير المستقل هو النوع والمتغير التابع هو مشاهدة البرامج التعليمية .

3- فئات المتغير النوع هى (ذكور – إناث)

فئات المتغير مشاهدة البرامج التعليمية (يشاهد – لا يشاهد)

4- تكوين الجدول بحيث يحتل المتغير المستقل أعلى الجدول أى يكون أفقياً أما المتغير التابع فيحتل الجزء الأسفل اى يكون عمودياً .

5- وضع العلامات .

النوع / مشاهدة البرامج التعليمية	ذكور	إناث
يشاهد	~~////~~	////
لا يشاهد	~~////~~	~~////~~ /

6- إعادة كتابة الجدول بالأرقام .

النوع / مشاهدة البرامج التعليمية	ذكور	إناث	مج
يشاهد	5	4	9
لا يشاهد	5	6	11
مج	10	10	20

<u>ثانياً : العرض البياني للبيانات الإحصائية</u>

يعتبر العرض البيانى للبيانات الإحصائية بمثابة تلخيص للبيانات الإحصائية فى شكل يسهل منه استيعاب خصائص موضوع بحث الدراسة ، وتختلف طرق عرض البيانات المبوبة عن البيانات الغير مبوبة ، وسنتعرض لكل منها بالتفصيل فيما يلى :-

<u>أولاً : العرض البيانى للبيانات الغير مبوبة :</u>

والمقصود بالبيانات الغير مبوبة تلك البيانات المفردة أى لا يوجد بها فئات وهناك عدة طرق لعرض البيانات الغير مبوبة .

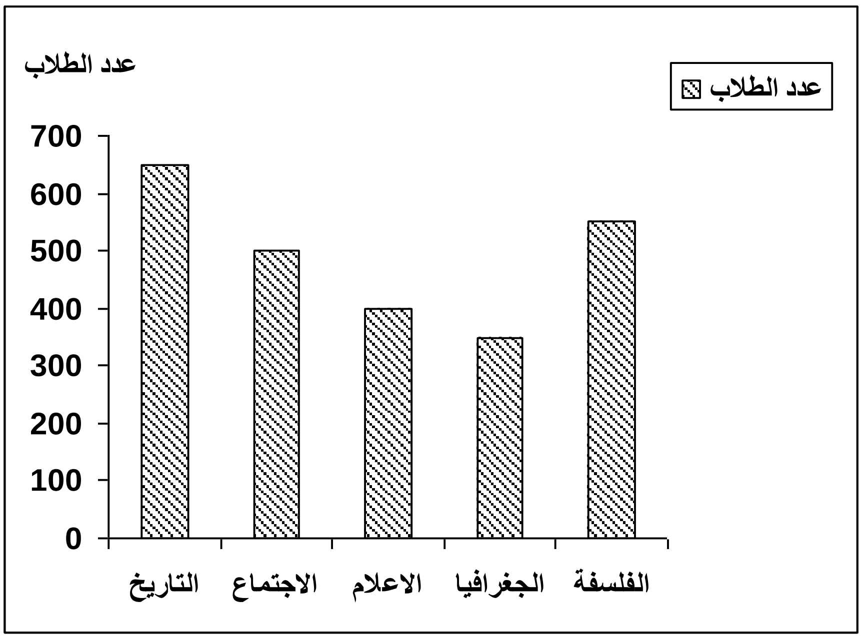
<u>(1) طريقة الأعمدة البيانية البسيطة :</u>

وفى هذه الطريقة يمثل محور السينات قيم المتغير أما محور الصادات يمثل القيمة المقابلة لقيمة المتغير ويتم رسم عمود حول المتغير وارتفاعه يمثل قيمة المتغير .

<u>مثال :</u>

الجدول التالى يوضح أعداد الطلاب ببعض أقسام كلية التربية والمطلوب عرض هذه البيانات باستخدام طريقة الأعمدة البيانية البسيطة ؟

القسم	التاريخ	الاجتماع	الإعلام	الجغرافيا	الفلسفة
عدد الطلاب	650	500	400	350	550



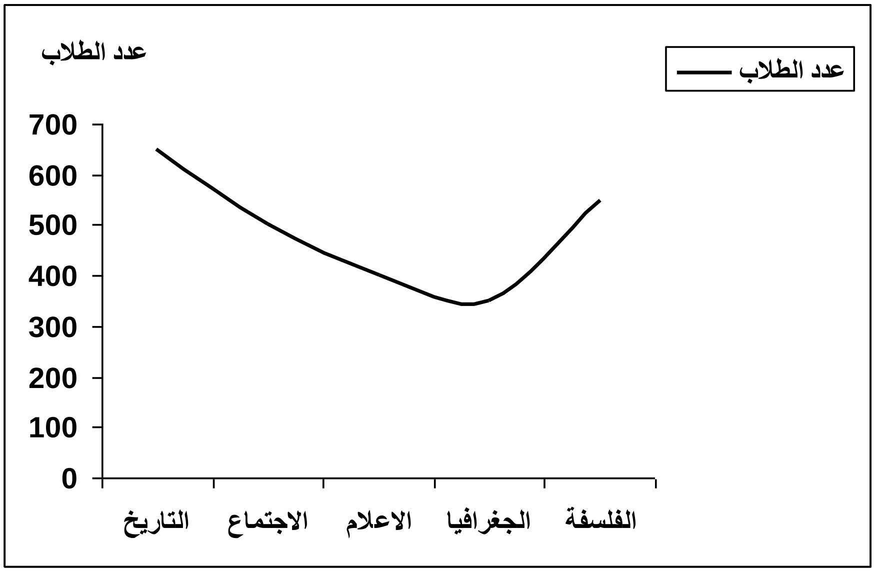
<u>(2) طريقة المنحنى البيانى البسيط :</u>

وفى هذه الطريقة يمثل محور السينات المتغير أما محور الصادات يمثل قيمة المتغير ويتم توقيع نقاط بين كل قيمة من قيم المتغير على محور السينات والقيمة المقابلة على محور الصادات ثم يتم توصيل تلك النقاط بخط منحنى باليد .

<u>مثال :</u>

الجدول التالى يوضح أعداد الطلاب ببعض أقسام كلية الآداب والمطلوب عرض هذه البيانات باستخدام طريقة المنحنى البياني البسيطة؟

القسم	التاريخ	الاجتماع	الإعلام	الجغرافيا	الفلسفة
عدد الطلاب	650	500	400	350	550



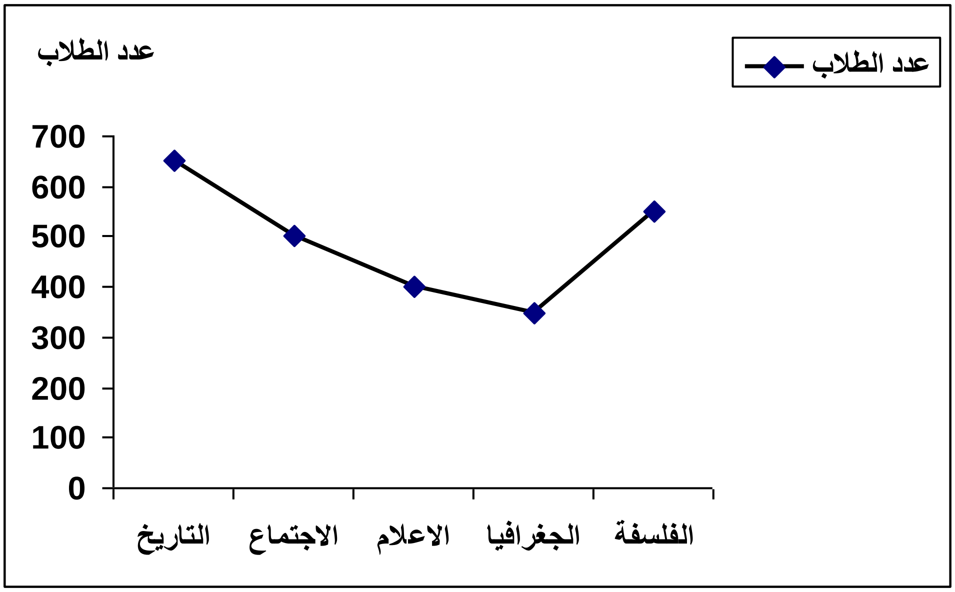
<u>(3) طريقة الخط البياني المنكسر :</u>

وفى هذه الطريقة يمثل محور السينات المتغير أما محور الصادات يمثل قيمة المتغير ويتم توقيع نقاط بين كل قيمة من قيم المتغير على محور السينات والقيمة المقابلة على محور الصادات ثم يتم توصيل تلك النقاط بخط منكسر باستخدام المسطرة .

<u>مثال :</u>

الجدول التالى يوضح أعداد الطلاب ببعض أقسام كلية العلوم والمطلوب عرض هذه البيانات باستخدام طريقة الخط البياني المنكسر؟

القسم	التاريخ	الاجتماع	الإعلام	الجغرافيا	الفلسفة
عدد الطلاب	650	500	400	350	550



(4) طريقة الدائرة البيانية :

وفي هذه الطريقة يتم رسم دائرة ثم نحسب زاوية قطاع كل قيمة على حدة ونقوم برسم تلك الزاوية داخل الدائرة حتى تنتهي الدائرة.

ونحسب زاوية قطاع الجزء من العلاقة :

$$\text{زاوية قطاع الجزء} = \frac{\text{التكرار الفعلي للجزء}}{\text{مجموع التكرارات}} \times 360$$

مثال :

الجدول التالي يوضح أعداد الطلاب ببعض أقسام كلية الآداب والمطلوب عرض هذه البيانات باستخدام طريقة الدائرة البيانية ؟

القسم	التاريخ	الاجتماع	الإعلام	الجغرافيا	الفلسفة
عدد الطلاب	650	500	400	350	550

الحل :

نحسب مجموع التكرارات = 550+350+400+500+650

مجموع التكرارات = 2450

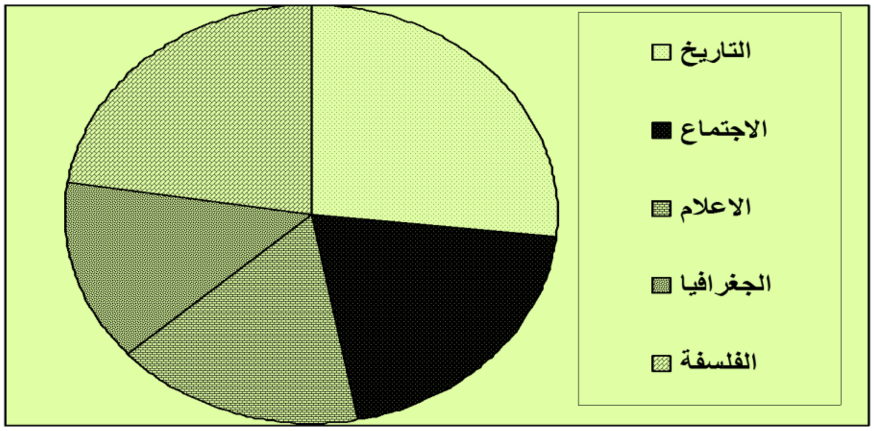
$$\text{زاوية قطاع التاريخ} = \frac{650}{2450} \times 360 = 95.5$$

$$\text{زاوية قطاع الاجتماع} = \frac{500}{2450} \times 360 = 73.5$$

$$\text{زاوية قطاع الإعلام} = \frac{400}{2450} \times 360 = 58.7$$

$$\text{زاوية قطاع الجغرافيا} = \frac{350}{2450} \times 360 = 51.4$$

$$\text{زاوية قطاع الفلسفة} = \frac{550}{2450} \times 360 = 80.8$$



(5) طريقة الأعمدة البيانية المتلاصقة :

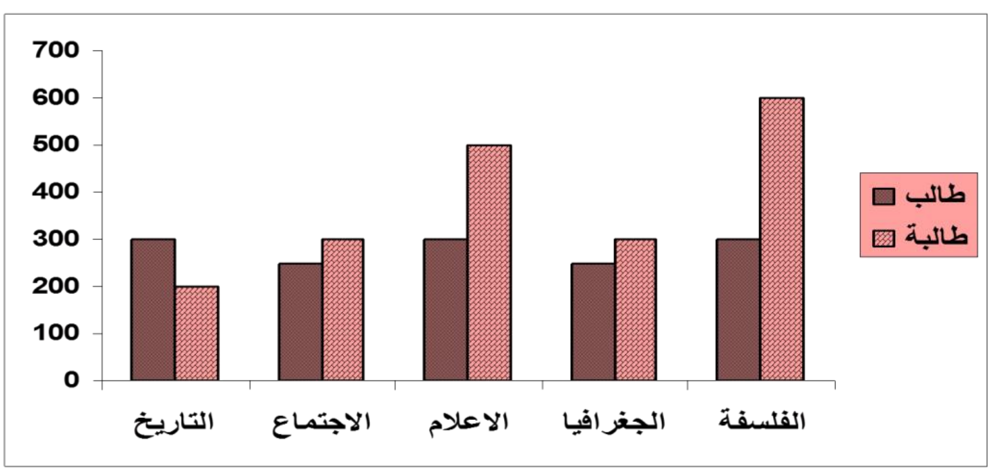
تسمى هذه الطريقة أيضا بطريقة الأعمدة البيانية المتجاورة وهى تشبه طريقة الاعمدة البيانية البسيطة ولكن يتم رسم عدد من الأعمدة متلاصقة يمثل كل منهم احد قيم المتغير .

مثال :

الجدول التالى يوضح أعداد الطلاب والمطلوب عرض هذه البيانات باستخدام طريقة الأعمدة البيانية المتلاصقة ؟

القسم	التاريخ	الاجتماع	الإعلام	الجغرافيا	الفلسفة
طالب	300	250	300	250	300
طالبة	200	300	500	300	600

الحل :



ثانياً : العرض البياني للبيانات المبوبة :

والمقصود بالبيانات المبوبة تلك البيانات المقسمة إلى فئات وهناك عدة طرق لعرض البيانات المبوبة

(1) المدرج التكرارى :

أحد طرق عرض البيانات المبوبة حيث يتم تخصيص عمود لكل فئة وتكرارها ، بحيث يكون طول الفئة هى قاعدة العمود والتكرار هو ارتفاع العمود ، ويفضل ترك فراغ كاف قبل الفئة الأولى وفراغ آخر بعد الفئة الأخيرة ، أما بالنسبة لمنتصف العمود فيكون هو مركز الفئة .

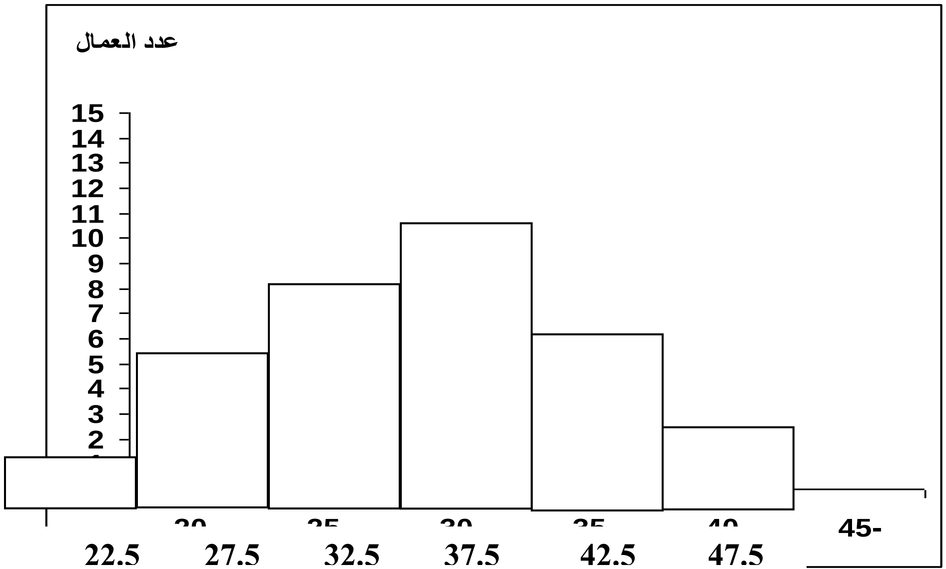
مثال :

اعرض لهذا الجدول بيانياً باستخدام المدرج التكرارى ؟

فئات العمر	20-	25-	30-	35-	40-	45-
عدد العمال	2	6	9	11	7	3

الحل :

ف	ك	مركز الفئة
20-	2	22.5
25-	6	27.5
30-	9	32.5
35-	11	37.5
40-	7	42.5
45-	3	47.5



(2) المضلع التكراري :

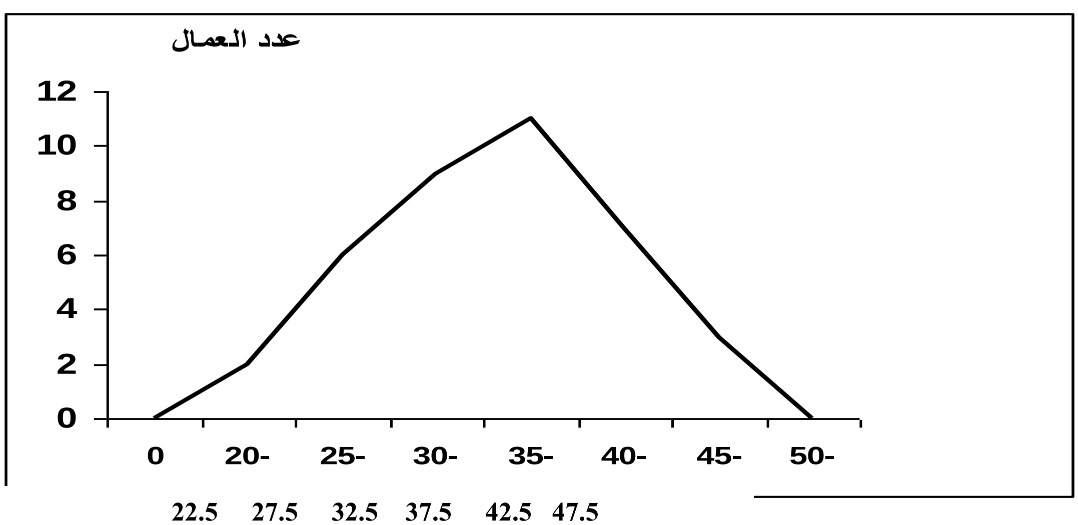
تخصص لكل فئة وتكرارها نقطة ، بحيث يكون الاحداثي السيني لها هو مركز الفئة بينما الاحداثي الصادي لها هو التكرار ، نفترض فئة سابقة للفئة الأولى وفئة لاحقة للفئة الأخيرة وتكرار كل منهما صفر ، ثم نوصل كل نقطتين متتاليتين بخط مستقيم بالمسطرة .

مثال :

اعرض لهذا الجدول بيانياً باستخدام المضلع التكراري ؟

فئات العمر	-20	-25	-30	-35	-40	-45
عدد العمال	2	6	9	11	7	3

الحل :



(3) المنحنى التكرارى :

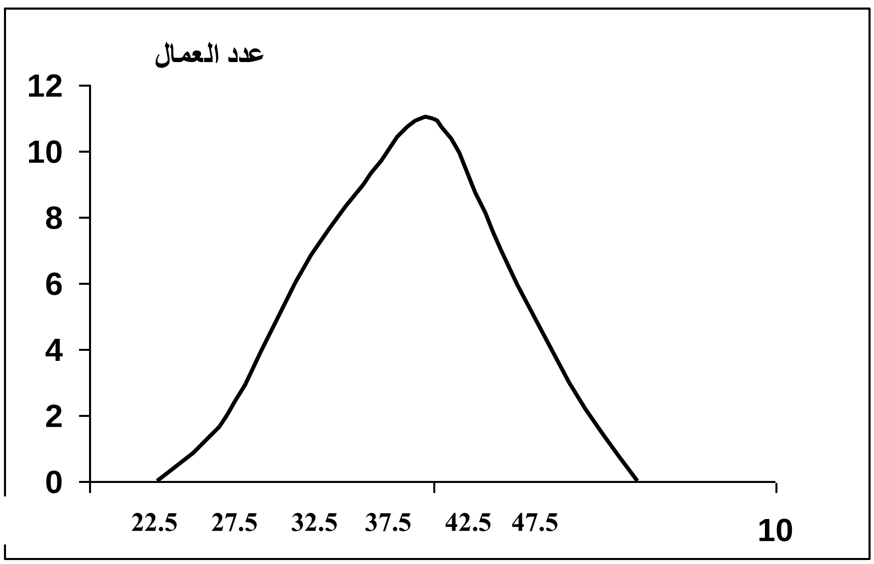
بعد رصد النقاط كما فى الطريقة السابقة نوصل كل نقطتين متتاليتين بمنحنى باليد .

مثال :

اعرض لهذا الجدول بيانياً باستخدام المنحنى التكرارى ؟

فئات العمر	20-	25-	30-	35-	40-	45-
عدد العمال	2	6	9	11	7	3

الحل :



تمارين

1- حصل عدد من الطلاب في مادة الإحصاء على الدرجات التالية :

5	4	4	5	3	4	2	3	1	2
3	7	4	1	6	3	2	5	3	4
7	3	2	6	5	3	4	2	4	1

المطلوب : تكوين جدول تكراري بسيط لهذه الدرجات.

2- تمثل البيانات التالية تقديرات عشرون طالبا في مادة علم النفس والمطلوب وضعها في جدول تكراري بسيط لتلك التقديرات .

ممتاز	مقبول	جيد جدا	مقبول	جيد
جيد جدا	جيد	ضعيف	جيد	مقبول
جيد	ممتاز	مقبول	ضعيف	جيد
جيد جدا	جيد	مقبول	جيد	مقبول

3- هذه درجات 50 طالبا في اختبار ذكاء ، والمطلوب وضع هذه الدرجات في جدول تكراري للفئات .

28	39	33	40	27	55	37	35	37	25
29	28	51	29	51	22	36	44	29	34
32	47	38	25	20	41	36	15	42	33
14	18	34	16	10	46	33	27	27	15
16	27	21	24	17	19	36	19	21	46

4- الدرجات التالية تمثل درجات 50 طالبا فى أحد الاختبارات:

5	6	5	7	5	6	6	4	5	4
6	6	5	6	6	7	9	8	7	5
5	3	3	5	4	9	7	8	6	7
5	8	8	6	7	7	6	7	7	6
4	6	6	7	6	4	7	7	8	5

والمطلوب : وضع هذه الدرجات فى جدول تكرارى للفئات .

5- حصل 80 طالبا فى اختبار ذكاء على الدرجات التالية:

46	38	30	20	11	46	23	46	45	18
47	39	33	25	29	49	28	13	36	25
50	43	32	21	19	51	25	15	48	16
49	41	35	27	13	37	29	27	55	37
51	45	21	23	18	50	27	17	12	48
52	42	37	26	14	38	26	14	28	50
53	44	34	22	28	47	30	16	26	36
48	40	31	29	12	35	24	22	20	19

والمطلوب :

- وضع هذه الدرجات فى جدول تكرارى للفئات بحيث يكون عدد الفئات .
- تكوين جدول التكرار المتجمع الصاعد .
- تكوين جدول التكرار المتجمع الهابط .

6- الجدول التالى يمثل أعداد الكتب بمكتبة الكلية فى مجموعة من التخصصات :

التخصص	علم الاجتماع	علم النفس	الرياضيات	اللغة العربية	بحوث العمليات
عدد الكتب	550	350	400	600	300

والمطلوب عرض هذه الجدول بيانياً باستخدام الطرق التالية :

- الأعمدة البيانية البسيطة .
- الخط البياني .
- الخط المنكسر .
- الدائرة البيانية .

7- الجدول التالى يمثل أعداد الذكور والإناث ببعض إدارات أحد الهيئات الحكومية .

الإدارة	الشئون الإدارية	الصيانة	الإحصاء	المعاشات
عدد الذكور	10	20	30	10
عدد الإناث	20	5	60	50

والمطلوب عرض هذه الجدول بيانياً باستخدام الطرق التالية :

- الأعمدة البيانية المتلاصقة .
- الأعمدة البيانية المجزأة .

8- الجدول التالى يمثل فئات درجات مجموعة من الطلاب فى اختبار للتحصيل وتكراراتهم :

الفئات	-5	-10	-15	-20	-25	-30	-35	-40
التكرار	10	13	8	9	12	5	6	7

والمطلوب هو عرض هذا الجدول بيانياً باستخدام الطرق التالية :

- المدرج التكرارى .
- المضلع التكرارى .
- المنحنى التكرارى .

الفصل الرابع

مقاييس النزعة المركزية_مقاييس النزعة المركزية (مقاييس الموضع)

Measures of Central Tendency (Location Measures)

مقدمة:

مقاييس النزعة المركزية هي مقاييس عددية تستخدم لقياس موضع تركز أو تجمع البيانات. إذ أن بيانات أي ظاهر تنزع في الغالب إلى التركز والتجمع حول قيم معينة. هذه القيم هي ما تسمى بمقاييس النزعة المركزية. ومقاييس النزعة المركزية تستخدم لتلخيص البيانات عدديًا إذ أنها تعتبر قيم نموذجية أو مثالية للبيانات. كما أن هذه المقاييس تستخدم لوصف مجموعة البيانات وكذلك لمقارنة مجموعات البيانات المختلفة. ومن أهم هذه المقاييس نذكر: الوسط الحسابي (أو المتوسط)، الوسط الموزون (أو المرجح)، الوسيط، والمنوال.

تعريف (1): رمز التجميع:

إذا كان عدد البيانات هو n وكانت البيانات هي $x_1, x_2, \ldots, x_n$ فإن مجموع هذه البيانات هو:

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n$$

خواص رمز التجميع:

يتمتع التجميع بالخواص التالية:

1. $\sum_{i=1}^{n}(x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i$
2. $\sum_{i=1}^{n} c x_i = c \sum_{i=1}^{n} x_i$
3. $\sum_{i=1}^{n} c = n\, c$

وبعد أن عرفنا رمز التجميع وخواصه نشرع في التطرق لمقاييس النزعة المركزية مبتدئين بالوسط الحسابي.

الوسط (المتوسط) الحسابي : Arithmetic Mean

يعتبر المتوسط من أهم وأفضل مقاييس النزعة المركزية ومن أكثرها شيوعًا واستخدامًا في التحليل الإحصائي وذلك لما يتمتع به من خصائص وصفات إحصائية جيدة. ولإيجاد المتوسط للبيانات فإننا لابد أن نفرق بين البيانات المفردة (غير مبوبة في جدول تكراري) والبيانات المبوبة (الملخصة في جدول تكراري).

أولاً: الوسط الحسابي للبيانات غير المبوبة :

إذا كان عدد البيانات (حجم العينة) هو n وكانت قيم أو مشاهدات العينة هي $x_1, x_2, \ldots, x_n$ فإن الوسط الحسابي يرمز له بالرمز $\bar{x}$ ويعرف بالصيغة التالية:

$$\text{الوسط الحسابي} = \frac{\text{مجموع البيانات}}{\text{عدد البيانات}}$$

أي أن:

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

مثال (1):

أوجد الوسط الحسابي للمشاهدات التالية والتي هي عبارة عن أوزان (بالكيلوغرام) لمجموعة مكونة من سبعة أشخاص:

25, 30, 40, 45, 35, 55, 50

الحل:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + \cdots + x_7}{7}$$

$$\bar{x} = \frac{25+30+40+45+35+55+50}{7} = \frac{280}{7} = 40$$

مثال(2) :

اذا كانت درجات مادة الرياضيات لمجموعة من الطلاب هي كالاتي

9,4,5,7,6,8 جد الوسط الحسابي ؟

الحل /

$$\bar{X} = \frac{\sum x_i}{n}$$

$$= \frac{9+4+5+7+6+8}{6} = \frac{39}{6}$$

$$= 6.5$$

مثال(3) / اذا كان مستوى هيموغلوبين الدم لدى ثمانية اشخاص كالاتي

13,10,12,13,11,10,13,10

جد الوسط الحسابي ؟

الحل /

$$\overline{X} = \frac{\sum x_i}{n}$$

$$= \frac{10+13+10+12+13+11+10+13}{8}$$

$$= \frac{92}{8} = 11.5$$

مثال(4) :

اذا كان متوسط مستوى الهرمون المحفز لنمو الحويصلات يساوي 18 حيث كان مستوى الهرمون المحفز في انثى الارنب الاولى = 18 وفي الثانية = 19 والثالثة = 17 اما في الرابعة = 19 جد مستوى الهرمون في انثى الارنب الخامسة ؟

الحل /

$$\overline{X} = \frac{\sum x_i}{n}$$

$$18 = \frac{18+19+17+19+X_5}{5}$$

$$90 = 73 + X_5 \quad ، X_5 = 17$$

ثانياً: الوسط الحسابي للبيانات المبوبة:

ينبغي علينا ملاحظة ما يلي في حالة البيانات الملخصة في جدول توزيع تكراري:

1. البيانات الأصلية غير معروفة.
2. عدد البيانات في كل فترة (تكرار الفترة) معروف.
3. يستخدم مركز الفترة كقيمة تقريبية لجميع البيانات في الفترة.

إذا كان لدينا بيانات عددها n وكانت هذه البيانات ملخصة في جدول تكراري بحيث أن:

1. عدد الفترات هو k.
2. مراكز الفترات هي $x_1, x_2, \ldots, x_k$.

3. تكرارات الفترات هي $f_1, f_2, \ldots, f_k$.

أي أن البيانات قد تم تلخيصها في جدول التوزيع التكراري التالي:

الفترة	مركز الفترة x	التكرار f
الفترة رقم 1	x_1	f_1
الفترة رقم 2	x_2	f_2
: :	: :	: :
الفترة رقم k	x_k	f_k
المجموع		$\sum f = n$

ولحساب المتوسط بالطريقة الحسابية فإنه يلزمنا فقط معرفة ما يلي:

1. حجم العينة = عدد البيانات = $n = \sum f$
2. مجموع البيانات = $\sum xf$

ولذلك فإن الوسط الحسابي للتوزيع التكراري المبوب يمكن حسابه بشكل تقريبي بالصيغة التالية:

$$\bar{x} = \frac{\sum xf}{\sum f} = \frac{\sum xf}{n} = \frac{x_1 f_1 + x_2 f_2 + \cdots + x_k f_k}{f_1 + f_2 + \cdots + f_k}$$

ويكمن تلخيص عملية إيجاد المتوسط باستخدام الجدول التالي:

الفترة	مركز الفترة x	التكرار f	$x\,f$
الفترة رقم 1	x_1	f_1	$x_1\,f_1$
الفترة رقم 2	x_2	f_2	$x_2\,f_2$
:	:	:	:
:	:	:	:
الفترة رقم k	x_k	f_k	$x_k\,f_k$
المجموع		$= n\sum f$	$\sum xf$

مثال (5):

أوجد الوسط الحسابي لمستوى الهيموجلوبين في الدم لعينة مكونة من خمسين شخصا

مستوى الهيموجلوبين	12.95 – 13.95	13.9 5– 14.95	14. 95– 15.95	15. 95– 16.95	16. 95– 17.95	17. 95– 18.95
التكرار	**3**	**5**	**15**	**16**	**10**	**1**

الحل:

مستوى الهيموجلوبين	مركز الفترة x	التكرار f	xf
12.95 – 13.95	13.45	3	40.35
13.9 5– 14.95	14.45	5	72.25
14. 95– 15.95	15.45	15	231.75
15. 95– 16.95	16.45	16	263.20
16. 95– 17.95	17.45	10	174.50
17. 95– 18.95	18.45	1	18.45
المجموع		$n = \sum f = 50$	$= 800.5 \sum xf$

الوسط الحسابي هو:

$$\bar{x} = \frac{\sum xf}{\sum f} = \frac{800.5}{50} = 16.01$$

بعض خصائص الوسط الحسابي :

1. المجموع الجبري لانحرافات القيم عن الوسط الحسابي $\bar{x}$ يساوي الصفر. أي أن:

$$\sum_{i=1}^{n}(x_i - \bar{x}) = (x_1 - \bar{x}) + (x_2 - \bar{x}) + ... + (x_n - \bar{x}) = 0$$

حيث أن $(x_i - \bar{x})$ هو انحراف القيمة x_i عن وسطها الحسابي $\bar{x}$.

2. مجموع مربعات انحرافات القيم عن الوسط الحسابي $\bar{x}$ أصغر من أو يساوي مجموع مربعات انحرافات القيم عن أي قيمة حقيقية c، أي أن:

$$\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2 \le \sum_{i=1}^{n}\left(x_i-c\right)^2$$

وبشكل مكافئ فإن:

$$\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2 = \min_{-\infty<c<\infty} \sum_{i=1}^{n}\left(x_i-c\right)^2$$

3. الوسط الحسابي يخضع للعمليات الجبرية بسهولة. فإذا كان $\bar{x}$ هو متوسط البيانات $x_1,x_2,\dots,x_n$ وكان a و b مقدارين ثابتين، فإن:

أ- متوسط البيانات $x_1 \pm b, x_2 \pm b, \cdots, x_n \pm b$ هو $\overline{(x \pm b)} = \bar{x} \pm b$.

ب- متوسط البيانات $ax_1, ax_2, \dots, ax_n$ هو $\overline{ax} = a\bar{x}$.

ج- متوسط البيانات $ax_1 \pm b, ax_2 \pm b, \cdots, ax_n \pm b$ هو $\overline{(ax \pm b)} = a\bar{x} \pm b$.

ويمكن تلخيص هذه الخاصية في الجدول التالي:

المشاهدات	الوسط الحسابي (المتوسط)
$x_1,x_2,\dots,x_n$	$\bar{x}$
$x_1 \pm b, x_2 \pm b, \cdots, x_n \pm b$	$\overline{(x \pm b)} = \bar{x} \pm b$
$ax_1, ax_2, \dots, ax_n$	$\overline{ax} = a\bar{x}$
$ax_1 \pm b, ax_2 \pm b, \cdots, ax_n \pm b$	$\overline{(ax \pm b)} = a\bar{x} \pm b$

مثال (6):

	المشاهدات	الوسط الحسابي
$: x$	2, 6, 4, 3, 5	$\bar{x} = 4$
$: x+5$	7, 11, 9, 8, 10	$\bar{x} + 5 = 9$
$:3x$	6, 18, 12, 9, 15	$3\bar{x} = 12$
$: 3x +5$	11, 23, 17, 14, 20	$3\bar{x} + 5 = 17$

مثال (7):

إذا كان الوسط الحسابي للمشاهدات $x_1, x_2, \ldots, x_n$ هو 15 فإن الوسط الحسابي للمشاهدات $\frac{x_1 - 10}{2}, \frac{x_2 - 10}{2}, \ldots, \frac{x_n - 10}{2}$

هو $\frac{\bar{x} - 10}{2} = \frac{15 - 10}{2} = 2.5$.

4. إذا كان لدينا مجموعتان من البيانات بحيث أن عدد بيانات المجموعة الأولى هو n_1 و وسطها هو $\bar{x}_1$ وكان عدد بيانات المجموعة الثانية هو n_2 و وسطها هو $\bar{x}_2$ فإن وسط المجموعة الكلية $(\bar{x})$ المكونة من دمج هاتين المجموعتين يمكن حسابه بالصيغة التالية:

$$\bar{X} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2}$$

مثال (8):

إذا كان لدينا مجموعتان من البيانات بحيث أن عدد بيانات المجموعة الأولى هو 10 و وسطها الحسابي هو 5 وكان عدد بيانات المجموعة الثانية هو 20 و وسطها الحسابي هو 2، فأوجد الوسط الحسابي للمجموعة الكلية المكونة من دمج هاتين المجموعتين.

الحل:

$$n_1 = 10, \bar{X}_1 = 5$$
$$n_2 = 20, \bar{X}_2 = 2$$

$$\bar{X} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2} = \frac{10 \times 5 + 20 \times 2}{10 + 20} = \frac{90}{30} = 3$$

مثال(9) :

في امتحان مادة الرياضيات لمئة طالب وكان الامتحان من عشرة درجات وتم الحصول على البيانات الاتية:

x_i	10	9	8	7	6	5	4
f_i	5	16	21	35	13	8	2

جد الوسط الحسابي ؟

الحل/

$$\bar{X} = \frac{\sum x_i f_i}{\sum f_i}$$

x_i	f_i	$x_i f_i$
10	5	50
9	16	144
8	21	168
7	35	245
6	13	78
5	8	40
4	2	8
المجموع	100	733

$$\bar{X} = \frac{733}{100} = 7.33$$

1- في حالة وجود فئات يجب استخراج مركز الفئة عن طريق العلاقة الاتية :

$$مركز\ الفئة = \frac{\text{الحد الاعلى للفئة + الحد الادنى للفئة}}{2}$$

مثال(10)/ اوجد الوسط الحسابي لمجموعة البيانات الاتية :

الفئات	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59
التكرارات	12	8	6	2	27	16	14	8	5	2

الحل/

$$\overline{X} = \frac{\sum x_i f_i}{\sum f_i}$$

الفئات	التكرارات(f)	مركز الفئة (x)	$X*f$
10-14	12	12	144
15-19	8	17	136
20-24	6	22	132
25-29	2	27	54
30-34	27	32	864
35-39	16	37	592
40-44	14	42	588
45-49	8	47	376
50-54	5	52	260
55-59	2	57	114
المجموع	*100*		*3260*

$$\overline{X} = \frac{\sum x_i f_i}{\sum f_i}$$

$$= \frac{3260}{100} = 32.6$$

بعض مميزات وعيوب الوسط الحسابي :

لكل مقياس من مقاييس النزعة المركزية مميزاته وعيوبه. ونورد فيما يلي بعض مميزات وعيوب المتوسط.

(أ) مميزات الوسط الحسابي:

يعتبر الوسط الحسابي من أفضل مقاييس النزعة المركزية ومن أكثرها شيوعًا وذلك لما يتمتع به من صفات جيدة. ومن مميزات الوسط الحسابي نذكر ما يلي:

1. الوسط الحسابي سهل التعريف والحساب ويخضع للعمليات الجبرية بسهولة.
2. الوسط الحسابي وحيد لمجموعة البيانات الواحدة.
3. يأخذ الوسط الحسابي في الاعتبار جميع البيانات.

(ب) **عيوب** الوسط الحسابي:

بالرغم من أن الوسط الحسابي يعتبر من أفضل مقاييس النزعة المركزية إلا أن له بعض العيوب نذكر منها ما يلي:

1. يتأثر الوسط الحسابي بالقيم الشاذة أو المتطرفة.
2. الوسط الحسابي غير معرف للبيانات الوصفية (النوعية) إذ يمكن حسابه للبيانات الكمية فقط.

(3−3) الوسط المرجح (الموزون): Weighted Mean

في بعض الأحيان تكون المشاهدات $x_1, x_2, \ldots, x_n$ مقرونة بالأوزان $w_1, w_2, \ldots, w_n$ على التوالي.

وفي هذه الحالة نعرف الوسط المرجح كما يلي:

$$\bar{x}_w = \frac{\sum x\ w}{\sum w} = \frac{x_1 w_1 + x_2 w_2 + \cdots + x_n w_n}{w_1 + w_2 + \cdots + w_n} \qquad (3)$$

مثال (1):

أوجد الوسط المرجح لدرجات أحد الطلاب في مقررات الإحصاء والفيزياء والرياضيات باعتبار أن الوزن هو عدد الساعات للمقرر فيما يلي:

المقرر	عدد الساعات (w)	الدرجة (x)
إحصاء	2	40
فيزياء	4	65
رياضيات	3	70

الحل:

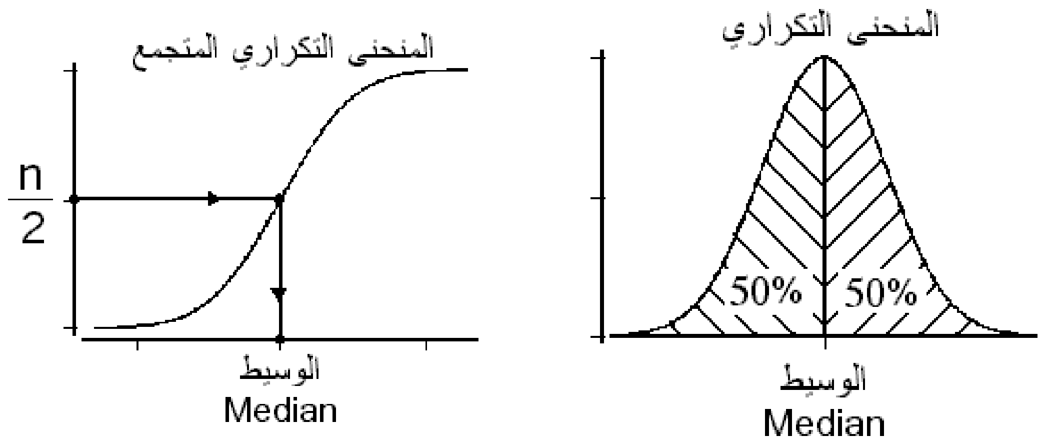
$$\bar{x}_w = \frac{\sum x\ w}{\sum w} = \frac{x_1 w_1 + x_2 w_2 + x_3 w_3}{w_1 + w_2 + w_3}$$

$$= \frac{40 \times 2 + 65 \times 4 + 70 \times 3}{2 + 4 + 3} = \frac{550}{9} = 61.11$$

وبعد أن تطرقنا الوسط الحسابي كمقياس من مقاييس النزعة المركزية، نورد فيما يلي أحد المقاييس الأخرى شائعة الاستخدام ألا وهو الوسيط. ويتميز الوسيط بأنه أقل تأثرًا بالقيم الشاذة والمتطرفة من المتوسط ولكن يعاب عليه بأنه لا يأخذ في الاعتبار جميع القيم.

الوسيط: Median

الوسيط هو أحد مقاييس النزعة المركزية المشهورة. ويعرف الوسيط لمجموعة من البيانات على أنه تلك القيمة التي تتوسط البيانات عندما يتم ترتيبها تصاعديًا (أو تنازليًا) أي أنه تلك القيمة التي تقسم البيانات بعد ترتيبها إلى جزأين متساويين فتكون البيانات في الجزء الأول تقل عن أو تساوى الوسيط والبيانات في الجزء الثاني تزيد عن أو تساوى الوسيط. أي أن 50% من البيانات تساوي أو تقل عن الوسيط و 50% من البيانات تساوي أو تزيد عن الوسيط. يرمز للوسيط بالرمز (Med). والشكل (1) يوضح لنا فكرة الوسيط.



شكل (1): الوسيط

طرق حساب الوسيط:

أولاً: الوسيط للبيانات غير مبوبة:

إذا كانت قيم العينة هي $x_1,x_2,\ldots,x_n$ وحجم العينة هو n فإن الوسيط يعرف كما يلي:

أولاً: إذا كان حجم العينة n عددًا فرديًا فإن الوسيط هو:

الوسيط = القيمة التي في منتصف البيانات بعد ترتيبها وهي القيمة المرتبة ذات الترتيب $\frac{n+1}{2}$.

والجدول التالي يبين أن الوسيط هو القيمة التي تتوسط البيانات بعد ترتيبها في هذه الحالة، أي أن الوسيط = القيمة في المنتصف = $X_{\left(\frac{n+1}{2}\right)}$

البيانات مرتبة	$X_{(1)}$	$X_{(2)}$	$\ldots$	$X_{\left(\frac{n+1}{2}\right)}$	$\ldots$	$X_{(n)}$
الترتيب	1	2	$\ldots$	$\frac{n+1}{2}$	$\ldots$	n

ثانيًا: إذا كان حجم العينة n عددًا زوجيًا فإن الوسيط يعرف كما يلي:

الوسيط = متوسط القيمتين في منتصف البيانات بعد ترتيبها وهما القيمتان المرتبتان ذاتا الترتيب $\frac{n}{2}$ و $\frac{n}{2}+1$.

والجدول التالي يبين أن الوسيط هو متوسط القيمتين اللتين تتوسط البيانات بعد ترتيبها في هذه الحالة، أي أن الوسيط = القيمة في المنتصف = $\frac{X_{\left(\frac{n}{2}\right)}+X_{\left(\frac{n}{2}+1\right)}}{2}$،

حيث أن القيمتين في المنتصف هما: $X_{\left(\frac{n}{2}\right)}$ و $X_{\left(\frac{n}{2}+1\right)}$.

البيانات مرتبة	$X_{(1)}$	$X_{(2)}$	…	$X_{\left(\frac{n}{2}\right)}$	$X_{\left(\frac{n}{2}+1\right)}$	…	$X_{(n)}$
الترتيب	1	2	…	$\frac{n}{2}$	$\frac{n}{2}+1$	…	$n.$

مثال (1):

أوجد الوسيط لمجموعة الأوزان (بالكيلوغرام) التالية: 8.3 ,5.4 ,2.5 ,2.5 ,7.1.

الحل:

بما أن $n = 5$ عدد فردي فإن الوسيط هو القيمة التي في المنتصف بعد ترتيب البيانات وهي القيمة ذات الترتيب $3 = \frac{5+1}{2} = \frac{n+1}{2}$. نرتب البيانات في الجدول التالي:

البيانات مرتبة	2.5	2.5	5.4	7.1	8.3
الترتيب	1	2	3	4	5

نجد أن الوسيط هو القيمة ذات الترتيب 3 لذلك فإن: الوسيط = 5.4 كيلوغراما

مثال (2):

أوجد الوسيط لمجموعة الأوزان (بالكيلوغرام) التالية: 8.3 ,9.2 ,5.4 ,2.5 ,2.5 ,7.1.

الحل:

بما أن $n = 6$ عدد زوجي فإن الوسيط هو متوسط القيمتين اللتين في المنتصف بعد ترتيب البيانات

وهما القيمتان ذاتا الترتيب $3 = \frac{6}{2} = \frac{n}{2}$ و $\frac{n}{2}+1=4$. نرتب البيانات في الجدول التالي:

البيانات مرتبة	2.5	2.5	5.4	7.1	8.3	9.2
الترتيب	1	2	3	4	5	6

نجد أن القيمتين في المنتصف هما 5.4 و 7.1 ولذلك فإن:

الوسيط = $\frac{5.4+7.1}{2}$ = 6.25 كيلوغراما.

مثال (3) / اذا كانت لديك البيانات الاتية

5,3,2,6,7,8,10

اوجد الوسيط ؟

الحل /

نرتب البيانات بشكل تصاعدي

2,3,5,6,7,8,10

بما ان عدد افراد العينة هو عدد فردي فان

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

اذن قيمة الوسيط M=6

مثال(4) / اذا كان مستوى الهيموغلوبين لدى 9 رجال هو كالاتي

15,12,13,12,13,11,14,13,14

اوجد الوسيط ؟

الحل/

نرتب البيانات بشكل تصاعدي

11,11,12,12,13,13,13,14,14

بما ان عدد افراد العينة هو عدد فردي فان

ترتيب الوسيط $= \frac{n+1}{2} = \frac{9+1}{2} = 5$

اذن الوسيط $M=13$

مثال(5)/ اذا كانت اوزان مجموعة من الاشخاص هي كالاتي

75,80,50,64,65,58,66,60,78,84

اوجد الوسيط ؟

الحل /

نرتب البيانات بشكل تصاعدي

50,58,60,64,65,66,75,78,80,85

بما عدد افراد العينة زوجي فان

ترتيب الوسيط الاول $= \frac{n}{2} = \frac{10}{2} = 5$

ترتيب الوسيط الثاني $= \frac{n}{2} + 1 = 6$

$M = \frac{65+66}{2} = 65.5$

مثال(6) / اذا كانت البيانات التالية تمثل مستوى الهيموغلوبين لدى 10 رجال

10,11,12,11,13,12,13,11,14,10

اوجد الوسيط ؟

الحل /

نرتب البيانات بشكل تصاعدي

10,10,11,11,11,12,12,13,13,14

بما عدد افراد العينة زوجي فان

ترتيب الوسيط الاول $= \frac{n}{2} = \frac{10}{2} = 5$

ترتيب الوسيط الثاني $= \frac{n}{2} + 1 = 6$

$M = \frac{12+11}{2} = 11.5$

H.W اذا كانت لديك البيانات الاتية فجد قيمة الوسيط ؟

1- 5,10,20,15,16,36,40,25

2- 16,14,12,20,23,12,10,25,15

3- 9,12,8,6,7,10,5,4

4- 15,18,12,13,11,14,19

ثانيًا: الوسيط للبيانات المبوبة:

يمكن حساب الوسيط للبيانات الملخصة في جدول تكراري بطرقتين هما: (أ) الطريقة الحسابية و(ب) الطريقة البيانية. ويستخدم الجدول التكراري المتجمع الصاعد لإيجاد الوسيط حسابيًا بينما يستخدم المضلع التكراري المتجمع الصاعد لإيجاد الوسيط بيانيًا.

إيجاد الوسيط حسابيًا:

لأيجاد الوسيط في حالة البيانات المبوبة نتبع الخطوات الاتية

1- ايجاد التكرار المتجمع الصاعد
2- ايجاد ترتيب الوسيط = $\frac{\sum fi}{2}$
3- ايجاد الحد الادنى الفعلي للفئة الوسيطة = الحد الادنى للفئه – 0.5
4- نطبق قانون الوسيط الاتي

$$\mathbf{M = L + \left(\frac{\frac{\sum fi}{2} - c}{fi}\right) * I}$$

حيث ان :

M: الوسيط

L: الحد الادنى الفعلي للفئة الوسيطيه

I: طول الفئة الوسيطية

$\frac{\sum fi}{2}$: موقع الوسيط (رتبة الوسيط)

C: التكرار المتجمع الصاعد السابق لفئة الوسيط

F: تكرار الفئة الوسيطية

مثال(7) : البيانات التالية تبين توزيع 100 طالب من طلبة كلية الهندسة حسب صفة الطول

الفئات	التكرارات
60-69	6
70-79	12
80-89	47
90-99	25
100-109	10

فجد الوسيط ؟

الحل/

$$\mathbf{M= L + \left(\frac{\frac{\sum fi}{2}-c}{fi}\right) * I}$$

الفئات	التكرارات	التكرار المتجمع الصاعد
60-69	6	6
70-79	12	18
80-89	47	65
90-99	25	90
100-109	10	100
المجموع	100	

نجد ترتيب الوسيط

$\frac{\sum fi}{2} = \frac{100}{2} = 50$

طول الفئه = الحد الاعلى – الحد الادنى +1

$= 69 - 60 + 1 = 10$

الحد الادنى الفعلي للفئة الوسطية (L)

$80 - 0.5 = 79.5$

تكرار الفئة الوسطية (fi)

47

=

$$M = L + \left(\frac{\frac{\sum fi}{2} - c}{fi}\right) * I = 79.5 + \frac{50-18}{47} * 10$$

$= 86.30$

مثال(8) : جد الوسيط للبيانات الاتية

الفئات	التكرارات
60-62	5
63-65	15
66-68	45
69-71	27
72-74	8

الحل /

الفئات	التكرارات	التكرار المتجمع الصاعد
60-62	5	5
63-65	15	20
66-68	**45**	**65**
69-71	27	92
72-74	8	100
المجموع	**100**	

نجد ترتيب الوسيط

$\frac{\sum fi}{2}$ $50 = \frac{100}{2} =$

طول الفئه = الحد الاعلى – الحد الادنى +1

$= 68 - 66 + 1 = 3$

الحد الادنى الفعلي للفئة الوسطية (L) $= 66 - 0.5 = 65.5$

تكرار الفئة الوسطية (fi)

$= 45$

$$M = L + \left(\frac{\frac{\sum fi}{2} - c}{fi}\right) * I = 65.5 + \frac{50-20}{45} * 3 = 67.51$$

<u>H.W</u> اذا كانت البيانات التالية تمثل اجور 100 عاملا فجد قيمة الوسيط

الفئات	التكرارات
80-89	8
90-99	22
100-109	41
110-119	19
120-129	10

بعض مميزات وعيوب الوسيط:

أ- مميزات الوسيط:

يعتبر الوسيط من مقاييس النزعة المركزية الشائعة وذلك لما يتمتع به من بعض الصفات الجيدة. ومن مميزات الوسيط نذكر ما يلي:

1. الوسيط سهل التعريف والحساب.
2. الوسيط وحيد لمجموعة البيانات الواحدة.
3. الوسيط أقل تأثرًا من المتوسط بالقيم الشاذة أو المتطرفة.

ب- عيوب الوسيط:

بالرغم من أن الوسيط يعتبر من مقاييس النزعة المركزية الجيدة إلا أن له بعض العيوب نذكر منها ما يلي:

1. لا يأخذ الوسيط في الاعتبار جميع البيانات إذا أنه يعتمد فقط على القيم التي في المنتصف وعلى ترتيب البيانات بغض النظر عن قيمها.
2. لا يمكن حساب الوسيط للبيانات الوصفية (النوعية) بشكل عام. ولكن يمكن حسابه لبعض حالات البيانات الوصفية التي يكون للترتيب معنى فيها.

المنوال: Mode

المنوال هو أحد مقاييس النزعة المركزية شائعة الاستخدام ولاسيما في حالة البيانات الوصفية (النوعية). ويعرف المنوال لمجموعة من البيانات على أنه تلك القيمة التي تتكرر أكثر من غيرها أي أنها القيمة ذات التكرار الأكبر (إن وجدت). يرمز للمنوال بالرمز (Mod). ومن تعريف المنوال تتضح لنا عدة أنواع من البيانات:

1. بيانات ليس لها منوال وتسمى عديمة المنوال.
2. بيانات لها منوال واحد وتسمى وحيدة المنوال.
3. بيانات لها أكثر من منوال وتسمى متعددة المنوال.

أولاً: المنوال للبيانات غير مبوبة :

المنوال = المشاهدة الأكثر تكرارًا (إن وجدت).

مثال (1):

الجدول أدناه يتضمن بيانات إحدى الدراسات التي طبقت على خمسة أشخاص لقياس العمر (بالسنة) والوزن (بالكيلوغرام) والطول (بالسم) وفصيلة الدم. أوجد منوال للبيانات المختلفة.

رقم الشخص	1	2	3	4	5
العمر	25	20	25	30	35
الوزن	70	55	65	70	65

رقم الشخص	1	2	3	4	5
الطول	164	162	155	165	158
فصيلة الدم	O	A	B	A	AB

الحل:

نلخص حل هذا المثال بالجدول التالي:

البيانات	المنوال	نوع البيانات بالنسبة للمنوال
العمر	25	وحيدة المنوال
الوزن	المنوال الأول = 65 المنوال الثاني = 70	متعددة المنوال (ثنائية المنوال)
الطول	لا يوجد	عديمة المنوال
فصيلة الدم	A	وحيدة المنوال

مثال (2) : جد المنوال 10,7,18,7,12,10,18,12

:

لا يوجد منوال بيانات عديمة المنوال

مثال (3) : جد المنوال

6- 12,18,10,12,7,18,7

المنوال = 7,12,18 بيانات متعددة المنوال

ثانياً: المنوال للبيانات المبوبة:

نعرف الفترة المنوالية بأنها الفترة ذات التكرار الأكبر وهي الفترة التي يقع فيها منوال. وفي الجدوال التكراري قد يكون هناك فترة منوالية واحدة أو عدة فترات منوالية أو قد لا يوجد فترة منوالية. ويمكن حساب المنوال للبيانات الملخصة في جدول تكراري بطرقتين هما (أ) الطريقة الحسابية و(ب) الطريقة البيانية. ويستخدم الجدول التكراري لإيجاد المنوال حسابيًا بينما يستخدم المدرج التكراري لإيجاد المنوال بيانيًا.

إيجاد المنوال حسابيًا:

هناك طريقتان حسابيتان لإيجاد المنوال للبيانات المبوبة. فأما الطريقة الأولى فهي طريقة تقريبية وتعطى بالصيغة التالية:

المنوال = مركز الفترة المنوالية.

وأما الطريقة الثانية وهي الأكثر دقة فتعطى بالصيغة التالية:

$$\mathbf{Mo = Li + \frac{d1}{d1+ d2} * I}$$

نستخدم الطريقة الجبرية لأيجاد المنوال التي تسمى طريقة الفروق لبيرسون وحسب الخطوات المبينة ادناه .

1- نحدد الفئة المنوالية (الفئة التي تقابل اكثر تكرار)

2- نجد الفرق بين تكرار الفئة المنوالية وتكرار الفئة السابقة لها وليكن d_1

3- نجد الفرق بين تكرار الفئة المنوالية وتكرار الفئة اللاحقة لها وليكن d_2

4- نجد الحد الادنى الفعلي للفئة المنوالية وليكن L

5- نجد طول الفئة .

مثال (4) : جد المنوال للبيانات الاتية

الفئات	التكرارات
11-13	1
14-16	3
17-19	9
20-22	13
23-25	11
26-28	3

الحل /

$$\mathbf{Mo = Li + \frac{d1}{d1+d2} * I}$$

$d_1 = 13-9=4$

$d_2 = 13 - 11 =2$

$\mathbf{Li = 20 - 0.5 =19.5}$

$\mathbf{I = 13-11 +1 = 3}$

$$\mathbf{Mo = 19.5 + \frac{4}{4+2} * 3}$$

$= 21.5$

مثال (5) : جد المنوال للبيانات التالية التي تبين صفة الوزن لدى مجموعة من الاشخاص

الفئات	التكرارات
60-62	5
63-65	15
66-68	45
69-71	27
72-74	8

الحل /

$$\mathbf{Mo = Li + \frac{d1}{d1+d2} * I}$$

$d_1 = 45\text{-}15 = 30$

$d_2 = 45\text{-}27 = 18$

Li = 66– 0.5 =65.5

I = 68-66 +1 = 3

$$\mathbf{Mo = 65.5 + \frac{30}{30+18} * 3}$$

= 67.38

H.W **اذا كانت** البيانات التالية تمثل فئات الاوزان ل (100) **طالب جد كل من**

1- الوسط الحسابي

2- الوسيط

3- المنوال

التكرارات	الفئات
8	40-44
18	45-49
44	50-54
20	55-59
10	60-65

بعض مميزات وعيوب المنوال:

(أ) مميزات المنوال:

يعتبر المنوال من مقاييس النزعة المركزية الشائعة ومن مميزاته نذكر ما يلي:

1. المنوال سهل التعريف والحساب.
2. المنوال أقل تأثرًا من المتوسط بالقيم الشاذة أو المتطرفة.
3. يمكن حساب المنوال للبيانات الكمية والوصفية (النوعية).

(ب) عيوب المنوال:

بالرغم من أن المنوال يعتبر من مقاييس النزعة المركزية الشائعة إلا أن له بعض العيوب نذكر منها ما يلي:

1. لا يأخذ المنوال في الاعتبار جميع البيانات إذا أنه يعتمد فقط على البيانات ذات التكرار الأكثر.
2. قد لا يوجد منوال لمجموعة من البيانات أو قد يكون هناك أكثر من منوال.

العلاقة بين الوسط الحسابي والوسيط والمنوال

1- اذا كان التوزيع متماثل فان العلاقة التي تربط المقاييس الثلاثة هي كالاتي :

$$Mo = M = \bar{X}$$

2- اذا كان التوزيع غير متماثل فان العلاقة التي تربط المقاييس الثلاثة هي كالاتي :

$$\bar{X} - Mo = 3(\bar{X} - M)$$

H.W اذا كان الوسط الحسابي = 10 فجد قيمة الوسيط عندما يكون المنوال = 4 ، ثم بين نوع التوزيع ؟

الربيعات والعشيرات والمئينات:

مقاييس النزعة التي ذكرناها سابقًا عبارة عن مقاييس عددية تصف مجموعة البيانات ككل إذ أنها تستخدم لتلخيص البيانات جماعيًا بقيمة عددية واحدة. فمتوسط درجات الطلاب، مثلاً، يستخدم لوصف مستوى تحصيل مجموعة جميع الطلاب. وفي كثير من الحالات نرغب في معرفة وضع قيمة معينة مقارنة بجميع البيانات. فعلى سبيل المثال، عندما يعلم الطالب بدرجته في الاختبار فإنه لا شك سيكون راغبًا في معرفة وضعه بالنسبة لزملائه من خلال معرفة نسبة الطلاب الذين تقل درجاتهم عن درجته ونسبة الطلاب الذين تزيد درجاتهم عن درجته.

إن الربيعات والعشيرات والمئينات هي مقاييس عددية تستخدم لوصف وضع مشاهدة محددة مقارنة بمجموعة المشاهدات الكلية. وقد رأينا سابقًا أن الوسيط هو القيمة التي تجزئ البيانات بعد ترتيبها إلى جزأين متساويين، والشكل التالي يوضح هذا المعنى.

$$\frac{n}{2}=50\% \qquad\qquad \frac{n}{2}=50\%$$

$$\text{– – – – – – – – – – – – – –}\mid\text{– – – – – – – – – – – – –}$$

$$Med$$

والآن، لو جزأنا البيانات بعد ترتيبها إلى أربعة أجزاء متساوية كما في الشكل التالي:

$$\frac{n}{4}=25\% \qquad \frac{n}{4}=25\% \qquad \frac{n}{4}=25\% \qquad \frac{n}{4}=25\%$$

$$\text{– – – – – – – –}\underset{Q_1}{\mid}\text{– – – – – – – – – –}\underset{Q_2}{\mid}\text{– – – – – – – – – –}\underset{Q_3}{\mid}\text{– – – – – – –}$$

فإن نقاط التقسيم هي:

Q_1 = الربيع الأول

= القيمة التي يسبقها $\frac{1}{4}$ البيانات أو 25% من البيانات وتكون رتبتها = $\frac{n}{4}$.

Q_2 = الربيع الثاني

= القيمة التي يسبقها $\frac{2}{4}$ البيانات أو 50% من البيانات وتكون رتبتها = $\frac{2n}{4}$.

Q_3 = الربيع الثالث

= القيمة التي يسبقها $\frac{3}{4}$ البيانات أو 75% من البيانات وتكون رتبتها = $\frac{3n}{4}$.

أما إذا جزأنا البيانات بعد ترتيبها إلى عشرة أجزاء متساوية كما في الشكل التالي:

$$\begin{array}{cccccccccc}
\frac{n}{10} & \frac{n}{10} & \frac{n}{10} & \frac{n}{10} & \frac{n}{10} & \frac{n}{10} & \frac{n}{10} & \frac{n}{10} & \frac{n}{10} & \frac{n}{10} \\
---| & ---| & ---| & ---| & ---| & ---| & ---| & ---| & ---| & ---- \\
D_1 & D_2 & D_3 & D_4 & D_5 & D_6 & D_7 & D_8 & D_9 &
\end{array}$$

فإن نقاط التقسيم هي:

D_1 = العشير الأول

= القيمة التي يسبقها $\frac{1}{10}$ البيانات أو 10% من البيانات ويكون ترتيبها = $\frac{n}{10}$.

D_2 = العشير الثاني

= القيمة التي يسبقها $\frac{2}{10}$ البيانات أو 20% من البيانات ويكون ترتيبها = $\frac{2n}{10}$.

D_3 = العشير الثالث

= القيمة التي يسبقها $\frac{3}{10}$ البيانات أو 30% من البيانات ويكون ترتيبها = $\frac{3n}{10}$.

وهكذا ...

D_9 = العشير التاسع

= القيمة التي يسبقها $\frac{9}{10}$ البيانات أو 90% من البيانات ويكون ترتيبها = $\frac{9n}{10}$.

أما إذا جزأنا البيانات بعد ترتيبها إلى مائة جزء متساوي:

$$\begin{array}{ccccccccccccc} & \frac{n}{100} & & \frac{n}{100} & & \frac{n}{100} & & \cdots & & \frac{n}{100} & & \frac{n}{100} & \\ ---- & | & --- & | & --- & | & ----\cdots---- & & & | & --- & | & ---- \\ & P_1 & & P_2 & & P_3 & & \cdots & & P_{98} & & P_{99} & \end{array}$$

فإن نقاط التقسيم هي:

P_1 = المئين الأول

= القيمة التي يسبقها $\frac{1}{100}$ البيانات أو 1% من البيانات ويكون ترتيبها = $\frac{n}{100}$.

P_2 = المئين الثاني

= القيمة التي يسبقها $\frac{2}{100}$ البيانات أو 2% من البيانات ويكون ترتيبها = $\frac{2n}{100}$.

P_3 = المئين الثالث

= القيمة التي يسبقها $\frac{3}{100}$ البيانات أو 3% من البيانات ويكون ترتيبها = $\frac{3n}{100}$.

وهكذا ...

P_{99} = المئين التاسع والتسعون

= القيمة التي يسبقها $\frac{99}{100}$ البيانات أو 99% من البيانات ويكون ترتيبها= $\frac{99n}{100}$.

ملاحظة (5):

1. $Med = Q_2 = D_5 = P_{50}$
2. $Q_1 = P_{25}$
3. $Q_3 = P_{75}$
4. $D_1 = P_{10}, \quad D_2 = P_{20}, \quad D_3 = P_{30}, \ldots, D_9 = P_{90}$

<u>إيجاد الربيعات والعشيرات والمئينات للبيانات المبوبة:</u>

إن طريقة حساب الربيعات والعشيرات والمئينات سواءً كانت حسابية أم بيانية مشابهه لطريقة حساب الوسيط التي ذكرناها آنفًا إذ نستخدم الجدول التكراري المتجمع الصاعد والمضلع التكراري المتجمع الصاعد مع ملاحظة تحديد الرتبة المناسبة للمقياس المطلوب (R) بدلاً من رتبة الوسيط $\frac{n}{2}$.

أولاً: إيجاد الربيعات والعشيرات والمئينات حسابيًا:

نستخدم الجدول التكراري المتجمع الصاعد لحساب هذه المقاييس (الربيعات والعشيرات والمئينات).

بعد تحديد رتبة المقياس المطلوب، ولتكن R ، نستخدم الصيغة التالية:

$$\text{المقياس} = A \left(\frac{R - F_1}{F_2 - F_1} \right) \times L$$

حيث أن:

R = رتبة المقياس.

A = بداية فترة المقياس.

L = طول فترة المقياس.

F_1 = التكرار المتجمع الصاعد السابق لرتبة المقياس R.

F_2 = التكرار المتجمع الصاعد اللاحق لرتبة المقياس R.

ويمكن تلخيص رتب وقوانين هذه المقاييس (الربيعات والعشيرات والمئينات) في الجدول التالي:

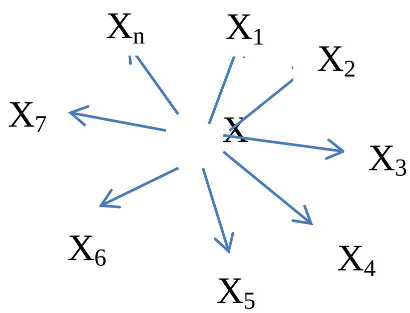
المقياس	الرمز	الرتبة (R)	القانون
الربيع رقم k	Q_k	$\frac{k\ n}{4}$	$Q_k = A + \left(\frac{\frac{kn}{4} - F_1}{F_2 - F_1} \right) \times L$
العشير رقم k	D_k	$\frac{k\ n}{10}$	$D_k = A + \left(\frac{\frac{kn}{10} - F_1}{F_2 - F_1} \right) \times L$
المئين رقم k	P_k	$\frac{k\ n}{100}$	$P_k = A + \left(\frac{\frac{kn}{100} - F_1}{F_2 - F_1} \right) \times L$

الفصل الخامس

مقاييس التشتت Measures of Variation

التشتت : هو مقدار التباعد او التقارب بين القيم والنقطة المركزية .اي هو عكس مقاييس النزعة المركزية .

لو كان لدينا X يمثل الوسط الحسابي والقيم الاخرة والمتمثلة ب $(X_1,X_2......X_n)$ تمثل التشتت حول القيمة المركزية :



- و من اهم مقاييس التشتت :

1- التباين Variance (S^2)
2- الانحراف المعياري Standard Deviation (S)
3- المدى Range (R)
4- الانحراف المتوسط The Mean Deviation (M.D)

1- التباين Variance (S^2)

-a- التباين في حالة البيانات غير المبوبة :

يعرف التباين على انه مجموع مربعات انحرافات القيم عن وسطها الحسابي مقسوما على حجم العينة اي ان :

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

حيث ان :

S^2: التباين

خطوات حساب التباين

1- حساب الوسط الحسابي

2- حساب انحراف القيم عن وسطها الحسابي

3- نجد مربعات انحرافات القيم عن وسطها الحسابي

4- استخدام قانون التباين

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

اما اذا كان حجم العينة $n>30$ فيتم استخدام الصيغة الاتية :

$$S^2 = \frac{\sum(x-\bar{x})^2}{n-1}$$

2- الانحراف المعياري (S) Standard Deviation

هو الجذر التربيعي للتباين اي ان :

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

$$S = \sqrt{s^2}$$

3- المدى Range(R)

يعرف على انه ابسط مقاييس التشتت ويتم حسابه عن طريق

المدى = اكبر قيمة – اصغر قيمة

مثال (1) : اذا كانت لديك البيانات الاتية

5,14,11,7,3

جد كل من التباين والانحراف المعياري والمدى ؟

الحل /

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

$$\bar{x} = \frac{\sum x}{n}$$

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

$(x-\bar{x})^2$	$(x-\bar{x})$	$\bar{x}$	X
9	-3	8	5
36	6	8	14
9	3	8	11
1	-1	8	7
25	-5	8	3
80			**المجموع**

$$S^2 = \frac{80}{5} \qquad = 16$$

$$S = \sqrt{s^2} \qquad = \sqrt{16} \quad = 4$$

المدى = اكبر قيمة – اصغر قيمة

R = 14 – 3 = 11

B- التباين في حالة البيانات المبوبة

لايجاد التباين في حالة البيانات المبوبة نتبع الخطوات الاتية :

1- نجد مركز الفئات X

2- نجد مربع مركز الفئات X^2

3- نجد حاصل ضرب التكرارات * مركز الفئات X*f

4- نجد حاصل ضرب التكرارات * مربع مركز الفئات X^2*f

5- تطبيق قانون التباين الاتي

$$S^2 = \frac{n\sum x^2 * f - \sum(xf)^2}{n(n-1)}$$

وكذلك الانحراف المعياري

$$S = \sqrt{s^2}$$

و كذلك المدى

R = الحد الاعلى للفئة الاخيرة – الحد الادنى للفئة الاول

مثال (2): جد التباين والانحراف المعياري للبيانات الاتية

الفئات	**التكرارات**
0-2	2
3-5	4
6-8	6
9-11	5
12-14	1

الحل /

الفئات	**التكرارات f**	**مركز الفئة X**	**X^2**	**fx**	**$f X^2$**
0-2	2	1	1	2	2
3-5	4	4	16	16	64
6-8	6	7	49	42	294
9-11	5	10	100	50	500
12-14	1	13	169	13	169
المجموع	**18**			**123**	**1029**

$$S^2 = \frac{n\sum x^2 * f - \sum(xf)^2}{n(n-1)}$$

$$S^2 = \frac{18(1029)-(123)^2}{18(18-1)}$$

$$S^2 = \frac{18522-15129}{306} \qquad = \quad 11.8$$

$$S = \sqrt{s^2}$$

$$= \sqrt{11.8}$$

$$= 3.32$$

H.W / جد الانحراف المعياري للبيانات الاتية :

التكرارات	الفئات
5	5-10
8	10-15
12	15-20
10	20-25
6	25-30
4	30-35
3	35-40
7	40-45
10	45-50

4- الانحراف المتوسط The Mean Deviation (M.D)

يعرف الانحراف المتوسط بانه المجموع المطلق لانحرافات القيم عن وسطها الحسابي مقسوما على حجم العينة اي ان

$$M.D = \frac{\sum|X-\bar{x}|}{N}$$

مثال (3): اذا كانت لديك البيانات الاتية

6,7,10,8,5,4,9,7

جد كل من الانحراف المتوسط ، التباين ، الانحراف المعياري ، المدى

الحل /

$M.D = \frac{\sum|X-\overline{x}|}{N}$

$\overline{x} = \frac{\sum X}{N} = \frac{56}{8} = 7$

x	$\overline{x}$	$X-\overline{x}$	$\mid X-\overline{x}\mid$	$(X-\overline{x})^2$
6	**7**	-1	1	1
7	**7**	0	0	0
10	**7**	3	3	9
8	**7**	1	1	1
5	**7**	-2	2	4
4	**7**	-3	3	9
9	**7**	2	2	4
7	**7**	0	0	0
المجموع			12	28

$M.D = \frac{12}{8} = 1.5$

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

$= \frac{28}{8} = 3.5$

$S = \sqrt{s^2}$

$S = \sqrt{3.5} = 1.87$

المدى = اكبر قيمة – اصغر قيمة

R = 10 – 4 = 6

H.W / احسب الانحراف المتوسط للتوزيع التكراري الاتي :

x	F
3	2
7	5
11	9
15	7
19	2

H.W / اذا كانت لديك مجموعة البيانات الاتية :

التكرارات	الفئات
8	30-32
10	33-35
12	36-38
5	39-41
3	43-45
7	46-48

جد كل من :

1- الوسط الحسابي
2- الوسيط
3- المنوال

4- الانحراف المعياري

5- المدى

6 – الانحراف المتوسط

- خواص التباين والانحراف المعياري

1- التباين والانحراف المعياري لا يتأثران بالاضافة والطرح .

2- التباين والانحراف المعياري يتأثران عند ضرب القيم باي عدد ثابت حيث ان :

- التباين الجديد = التباين السابق * (مقدار الثابت)^2
- الانحراف المعياري الجديد = الانحراف المعياري السابق * |مقدار الثابت|

3- يتأثر كل من التباين والانحراف المعياري بالقسمة على اي عدد ثابت حيث ان :

- التباين الجديد = $\frac{\text{التباين السابق}}{(\text{مقدار الثابت})^2}$
- الانحراف المعياري الجديد = $\frac{\text{الانحراف المعياري السابق}}{|\text{الثابت مقدار}|}$

معامل الاختلاف

يستخدم معامل الاختلاف للمقارنة بين المجموعات لمعرفة اي مجموعة اكثر في البيانات تشتت

ويرمز له بالرمز C حيث ان :

$$C = 100\% * \frac{\text{المعياري الانحراف}}{\text{الوسط الحسابي}}$$

$$C = 100\% * \frac{S}{\bar{x}}$$

مثال (4): اذا كانت لديك مجموعه من البيانات بوسط حسابي مقداره 17 وبتباين 36 جد معامل الاختلاف ؟

الحل /

$$C = \frac{S}{\bar{x}} *\%100$$

$$S = \sqrt{s^2}$$

$$= \sqrt{36} = 6$$

$$C = \frac{6}{36} \quad * 100\%$$

$$= 35.29\ \%$$

مثال (5): اذا كانت لديك المعلومات الاتية :

متوسط ارباح احدى شركات القطاع الخاص = 50 و بانحراف معياري = 10 ، ومتوسط ارباح احدى شركات القطاع العام = 30 وبانحراف معياري = 9 . بين اي الاجور اكثر تشتت بين الشركتين ؟

الحل/

$$C = \frac{S}{\bar{x}} \qquad C = \frac{S}{\bar{x}} * 100\%$$

القطاع الخاص

$$C = \frac{10}{50} * 100\%$$

$$= 20\%$$

القطاع العام

$$C = \frac{9}{30} \quad *100\%$$

$$= 30\%.$$

فتكون شركات القطاع العام اكثر تشتت من القطاع الخاص

الفصل السادس

الارتبــاط

(دراسة العلاقة بين متغيرين)

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 .

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.

شكل الانتشار Scatter Diagram :

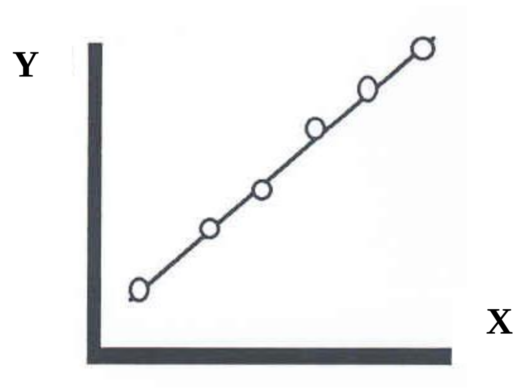
هناك وسيلة مبدئية يعرف الباحث من خلالها نوع الارتباط بين المتغيرين وما إذا كان الارتباط قوياً وضعيفاً أو منعدماً، وما إذا كانت العلاقة خطية أو غير خطية، موجبة أو سالبة. هذه الوسيلة هي " شكل الانتشار " والتي تصلح إذا كان المتغيران كميين. وجدير بالذكر أن هذه وسيلة مبدئية تساعد فقط في معرفة نوع الارتباط 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 :

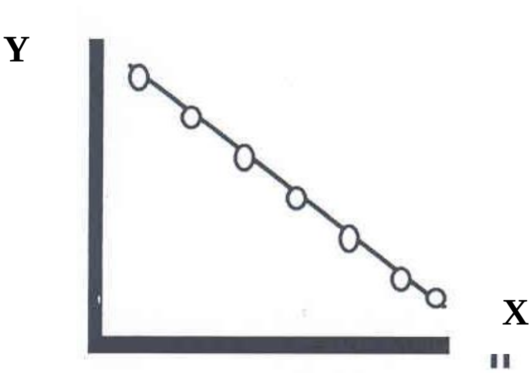
<u>**الشكل الأول :**</u>

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.

أمــا إذا كانــت العلاقــة عكســية (وجميــع النقــاط تقــع علــى خــط مســتقيم واحــد) فــإن " الارتباط عكسي تام " كما في الشكل الأول (ب). ومثال على ذلك العلاقة بين السرعة والزمن.



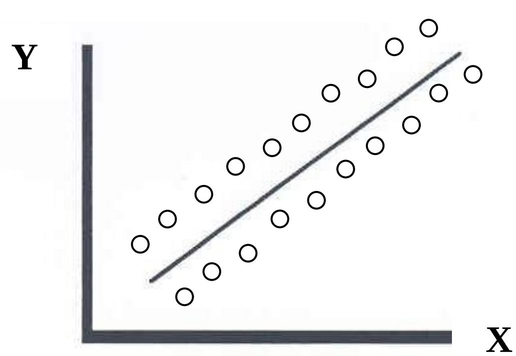
الشكل الأول (أ)
ارتباط طردي تام
(موجب)



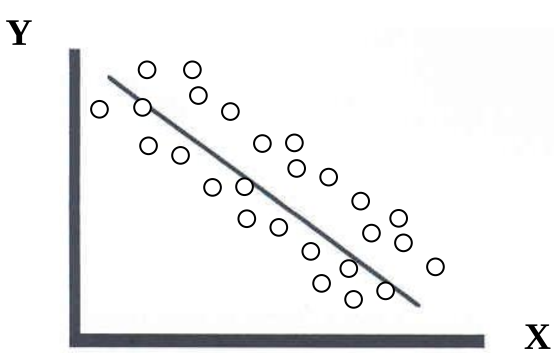
الشكل الأول (ب)
ارتباط عكسي تام
(سالب)

<u>الشكل الثاني :</u>

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



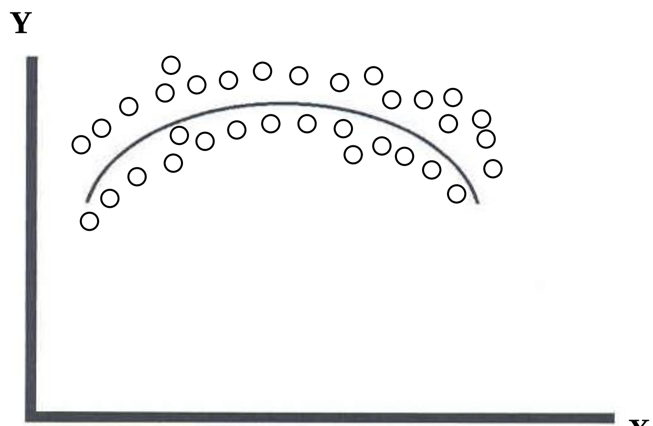
الشكل الثاني (ب)
ارتباط سالب قوي

<u>الشكل الثالث :</u>

(ارتباط خطي طردي)

(ارتباط خطي عكسي)

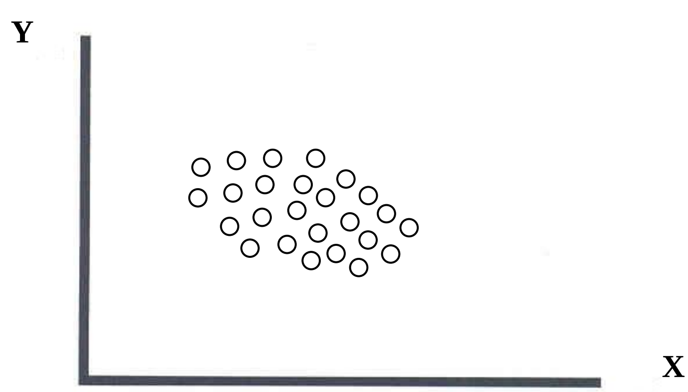
وإذا كانت العلاقة تأخذ شكل منحنى فإن الارتباط لا يكون خطياً "ارتباط غير خطي " Non Linear Correlation كما في الشكل الثالث :



الشكل الثالث
(ارتباط غير خطي)

<u>الشكل الرابع :</u>

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)

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 + 1، - 1. فإذا كانت قيمة معامل الارتباط تساوي + 1 فمعنى ذلك أن الارتباط بين المتغيرين طردي تام، وهو أقوى أنواع الارتباط الطردي بين متغيرين. وإذا كانت قيمة معامل الارتباط تساوي – 1 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 + 1 أو – 1 كلما كان الارتباط قوياً، وكلما اقترب من الصفر كلما كان الارتباط ضعيفاً.

<u>والخلاصة :</u>

أنه كلما كانت العلاقة قوية بين المتغيرين كلما اقترب معامل الارتباط من + 1 أو – 1 فإذا وصلت قيمة المعامل إلى + 1 أو – 1 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 + 1 ولا أصغر من – 1. وسنبدأ بقياس العلاقة بين متغيرين كميين، ثم متغيرين ترتيبين، وأخيراً متغيرين أسميين.

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، بالدينار أو الدولار.. كما أن المستوى العام للعمر – أي متوسط العمر – قد يساوي **<u>أربعين</u>** عاماً. فبينما المستوى العام – أي متوسط – الدخل الشهري قد يكون **<u>خمسة آلاف</u>** دولار 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 X , Y 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 : -

(2) الصيغة المختصرة لمعامل الارتباط

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

او الصيغة

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

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).

مثال (1) :

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.

معامل ارتباط الرتب Rank Correlation :

قد يرغب الباحث في حساب معامل الارتباط بين رتب المتغيرين وليس بين القيم ذاتها، فقد يكون المتغيران **وصفيين ترتيبين** Ordinal **أو أن يكون أحد المتغيرين كمياً بينما الأخر وصفياً ترتيبياً**، أو أن يكون **المتغيران كميين**، ويكون اهتمام الباحث منصباً على الرتب أكثر من القيم.

فإذا كانت رتب المتغيرين تسير في الاتجاه نفسه : بمعنى أن الرتب الأعلى للمتغير الأول تناظرها رتب أعلى للمتغير الثاني كانت العلاقة طردية بينهما. وإذا كانت الرتب الأعلى للمتغير الأول تناظرها رتب أدنى للمتغير الثاني كانت العلاقة بينهما عكسية. ففي مثالنا السابق عن العلاقة بين دخل الناخب وعمره، كان الناخب الأكبر عمراً (بصفة عامة) هو الأعلى دخلاً، فمن الواضح أن العلاقة بينهما طردية، أما إذا كان الناخب الأكبر عمراً (بصفة عامة) هو الأقل مشاركة في العمل السياسي فإننا في هذه الحالة نكون أمام علاقة عكسية.

ولحساب معامل ارتباط الرتب هناك طرق مختلفة أهمها معاملي سبيرمان وكيندال.

معامل سبيرمان لارتباط الرتب :

Spearman rank Correlation Coefficient

لحساب معامل سبيرمان لارتباط الرتب يقوم بترتيب كل من المتغيرين ترتيباً تصاعدياً أو تنازلياً (أما تصاعدياً لكلا المتغيرين أو تنازلياً لكليهما). وفي حالة الترتيب التصاعدي تأخذ أقل قيمة من قيم المتغير الرتبة رقم 1، والقيمة الأعلى منها مباشرة الرتبة رقم 2 وهكذا (بالنسبة لكل من المتغيرين). أما في حالة الترتيب التنازلي تأخذ أكبر قيمة من قيم المتغير الرتبة رقم 1، والقيمة الأقل منها مباشرة الرتبة رقم 2 وهكذا (بالنسبة لكل من المتغيرين). وعند تساوي قيمتين (أو أكثر) من قيم المتغير نعطي كل قيمة رتبة مختلفة (كما لو كانت القيم غير متساوية) ثم نحسب متوسط هذه الرتب، ويعطى هذا المتوسط لكل من هذه القيم المتساوية.

وبعد ترتيب المتغيرين نحسب الفروق بين رتب كل من المتغيرين (ونرمز للفروق بالرمز d) ثم نقوم بتربيع هذه الفروق ونحصل على مجموعها أي نحصل على $\sum d^2$ ثم نعوض في معامل سبيرمان لارتباط الرتب والذي يأخذ الشكل التالي :

$$\boxed{r_s = 1 - \frac{6(\sum d^2)}{n(n^2-1)}}$$

حيث : $\sum d^2$ هو مجموع مربعات الفروق بين رتب المتغيرين، n هي عدد أزواج القيم.

مما سبق نستطيع إجمال بعضاً من الملاحظات فيما يلي :

1 – مجموع الفروق بين الرتب يساوي صفر.

2 – أن قيمة معامل ارتباط الرتب تنحصر بين – 1، + 1 فإذا كانت الرتبة رقم 1 للمتغير الأول تناظرها الرتبة 1 للمتغير الثاني، والرتبة 2 للمتغير الأول تناظرها الرتبة رقم 2 للمتغير الثاني، وهكذا.. فإن معامل ارتباط الرتب يساوي + 1 (ارتباط طردي تام بين

الرتب). وإذا كانت الرتبة رقم 1 (أقل رتبة) للمتغير الأول تناظرها أعلى رتبة للمتغير الثاني وهكذا.. فإن معامل ارتباط الرتب يساوي – 1 (ارتباط عكسي تام بين الرتب).

3 – كذلك نلاحظ أن مجموع الرتب لكل من المتغيرين تساوي $\frac{n(n+1)}{2}$.

مثال (2):

البيانات التالية تمثل إجابات عينة من سبعة أشخاص حول برامج الضمان الاجتماعي، ومدى ملاءمتها لحاجات الناس.

السؤال الأول	جيد	مقبول	ممتاز	جيد	جيد جداً	مقبول	جيد
السؤال الثاني	جيد جداً	مقبول	جيد جداً	جيد	جيد	جيد	ممتاز

والمطلوب حساب معامل سبيرمان لارتباط الرتب بين هذين السؤالين ؟

الحل :

تنظم الحل في الجدول التالي مع ملاحظة ما يلي :

1 – بالنسبة للسؤال الأول، فإن التقدير الأعلى سيحصل على الرتبة رقم 1 والأقل منه مباشرة سيحصل على الرتبة رقم 2 وهكذا.. أي أن الترتيب تنازلي. ونكرر العمل نفسه مع السؤال الثاني.

2 – عند حصول إجابتين أو أكثر على التقدير نفسه نعطي لكل إجابة مبدئياً رتبة **كما لو كانوا مختلفين** ثم نحسب متوسط هذه الرتب، وهذا المتوسط هو الذي يعطى لكل إجابة.

3 – ثم نحسب الفروق بين رتب السؤالين ونرمز لها بالرمز d ثم نربع هذه الفروق فنحصل على d^2 ونعوض في القانون عن $\sum d^2$ مع ملاحظة أن $n = 7$.

السؤال الأول X	**السؤال الثاني Y**	**رتب X**	**رتب Y**	**d الفرق بين الرتب**	**d^2 مربعات الفرق**
جيد	جيد جداً	4	2.5	1.5	2.25
مقبول	مقبول	6.5	7	- 0.05	0.25
ممتاز	جيد جداً	1	2.5	- 1.5	2.25

جيد	جيد	4	5	- 1	1
جيد جداً	جيد	2	5	- 3	9
مقبول	جيد	6.5	5	1.5	2.25
جيد	ممتاز	4	1	3	9
المجموع				**Zero**	**26**

$$r_s = 1 - \frac{6\,(Sd^2)}{n\,(n^2-1)} = 1 - \frac{6\,(\sum d^2)}{n\,(n^2-1)} = 1 - \frac{6\,(26)}{7\,(49-1)}$$

$$= 1 - \frac{156}{336} = 1 - 0.46$$

$$r_s = 0.54$$

وهذا يعني أن الارتباط بين إجابات المبحوثين بالنسبة للسؤالين هو ارتباط طردي متوسط. وبالتالي فليس بالضرورة أن يكون رأي المجيبين في برامج الضمان الاجتماعي تعني ملاءمتها لحاجات الناس.

مثال (3)

البيانات التالية تمثل أعداد الساعات التي يدرس خلالها عشرة طلاب والدرجات التي حصلوا عليها في امتحان أحد المقررات :

عدد الساعات X	10	6	12	14	11	6	19	16	3	9
الدرجات y	60	48	83	76	74	58	98	89	37	69

أحسب معامل سبيرمان لارتباط الرتب.

الحل :

كما في المثال السابق ننظم الحل في الجدول التالي مع ملاحظة أن n = 10

عدد الساعات X	الدرجات Y	رتب X	رتب Y	الفروق d	d^2
10	60	6	7	- 1	1
6	48	8.5	9	- 0.5	0.05

1	1	3	4	83	12
1	- 1	4	3	76	14
0	0	5	5	74	11
0.25	0.5	8	8.5	58	6
0	0	1	1	98	19
0	0	2	2	89	16
0	0	10	10	37	3
1	1	6	7	69	9
4.50	**Zero**				**المجموع**

وبالتعويض في القانون حيث $\sum d^2 = 4.5$ ، $n = 10$ نحصل على :

$$r_s = 1 - \frac{6(\sum d^2)}{n(n^2-1)} = 1 - \frac{6(4.5)}{10(1-99)} = 1 - \frac{27}{990} = 1 - 0.027$$

$$r_s = 0.973$$

مما يعني أننا أمام علاقة طردية قوية بين المتغيرين. فكلما زادت عدد الساعات التي يدرسها الطالب في هذا المثال، كلما زادت درجاته في الامتحانات قوة وذلك بما نسبته 97 %.

امثلة تطبيقية على الارتباط

مثال(4) :اوجد معامل الارتباط بين اطوال و أوزان مجموعة من طلبة احدى الجامعات من البيانات التالية-:

الطول	164	152	184	164	176	156	168	164
الوزن	52	40	60	52	60	42	50	52

الحل-:

الطول X	الوزن y	XY	X^2	Y^2
164	52	8528	26896	2704
152	40	6080	23104	1600
184	60	11040	33856	3600
164	52	8528	26896	2704
176	60	10560	30976	3600
156	42	6552	24336	1764
168	50	8400	28224	2500
164	52	8528	26896	2704

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

$$r=\frac{68216-\frac{(1328)(408)}{8}}{\sqrt{\left(221184-\frac{(1328)^2}{8}\right)\left(21176-\frac{(408)^2}{8}\right)}}$$

$$r=\frac{488}{\sqrt{(736)(368)}}$$

$$r=\frac{488}{520.43}$$

$$r=0.937$$

النتيجة ان معامل الارتباط 0.937 يدل على وجود علاقة قوية جدا وطردية بين الطول والوزن لان الاشارة موجب

مثال : (5) الجدول التالي يعرض الدخل الشهري والاستهلاك بعشرات الدنانير لمجموعة من الافراد

الدخل X	الاستهلاك Y
465	367
682	495
396	373
837	612
784	687
922	764
850	621
482	370

المطلوب :اوجد بطريقة الرتب معامل الارتباط بين الدخل والاستهلاك؟

الحل :نتبع الخطوات التالية

1- نرتب البيانات تصاعديا او تنازليا

2- نرصد رتب (x) ونرصد رتب(y) مثلا نعطي اصغر قيمة رقم 1 والقيمة التي تليها 2 لكل من x و y

3- نحصل على الفرق d بين رتب x ورتب y

4- نربع الفرق d^2

5- نطبق القانون

ملاحظة :اذا تكررت القيم ناخذ الوسط الحسابي للرتب

X الدخل	Y الاستهلاك	رتب الدخل X	رتب الاستهلاك Y	d=x-y	d^2
465	367	2	1	1	1
682	495	4	4	0	0
396	373	1	3	-2	4
837	612	6	5	1	1
784	687	5	7	-2	4
922	764	8	8	0	0
850	621	7	6	1	1
482	370	3	2	1	-1
					$\sum d^2$=12

$\sum d^2$=**12**

n=8

$n^2 = 64$

$$r_s=1-\frac{6\sum d^2}{n(n^2-1)}$$

$$r_s=1-\frac{6(12)}{8(64-1)}$$

$$r_s=1-\frac{72}{504}$$

$r_s = 1 - 0.14 \qquad r_s = 0.86$

وهذا يدل على ان وجود ارتباط طردي قوي جدا بين الدخل والاستهلاك .

مثال(6) :الجدول التالي يعرض تقديرات عشرة من الطلاب في ماديتي الاحصاء و الاقتصاد

الاحصاء X	الاقتصاد Y
A	A+
C+	D
D	C
D+	C
B+	A
C+	B
A+	B+
B	B
B+	C
B+	B

المطلوب :اوجد معامل ارتباط الرتب بين اداء الطلاب في مادة الاحصاء والاقتصاد

الحل : نرتب التقديرات تصاعديا او تنازليا للمادتين كالتالي

<table>
<tr><th>رتب X</th><th>الاحصاء X</th><th>رتب Y</th><th>الاقتصاد Y</th></tr>
<tr><td>1</td><td>A+</td><td>1</td><td>A+</td></tr>
<tr><td>2</td><td>A</td><td>2</td><td>A</td></tr>
<tr><td rowspan="3">(3+4+5)/3=4</td><td>B+</td><td>3</td><td>B+</td></tr>
<tr><td>B+</td><td rowspan="3">(4+5+6)/3 = 5</td><td>B</td></tr>
<tr><td>B+</td><td>B</td></tr>
<tr><td>6</td><td>B</td><td>B</td></tr>
<tr><td rowspan="2">(7+8)/2 = 7.5</td><td>C+</td><td rowspan="3">(7+8+9)/3 = 8</td><td>C</td></tr>
<tr><td>C+</td><td>C</td></tr>
<tr><td>9</td><td>D+</td><td>C</td></tr>
<tr><td>10</td><td>D</td><td>10</td><td>D</td></tr>
</table>

الاقتصاد Y	الاحصاء X	رتب Y	رتب X	d=x-y	d^2
A+	**A**	1	2	1	1
D	**C+**	10	7.5	-2.5	6.25
C	**D**	8	10	-2	4
C	**D+**	8	9	1	1
A	**B+**	2	4	-2	4
B	**C+**	5	7.5	-2.5	6.25
B+	**A+**	3	1	2	4
B	**B**	5	6	-1	1
C	**B+**	8	4	4	16
B	**B+**	5	4	1	1
					$\sum d^2=44.5$

$\sum d^2=44.5$

n=10

$n^2 = 100$

$$r_s=1-\frac{6\sum d^2}{n(n^2-1)}$$

$$r_s=1-\frac{6(44.5)}{10(100-1)}$$

$$r_s=1-\frac{267}{990}$$

$$r_s=1-0.26$$

$$r_s=0.73$$

هناك ارتباط طردي قوي بين اداء الطالب في مادة الاقتصاد ومادة الاحصاء

الفصل السابع

الإنحدار Regression

الانحدار أو يسمى التنبؤ Prediction وهو تقدير القيمة المستقبلية لمتغير واحد بناءً على معرفة قيم متغير أو أكثر، وهنالك عدة أنواع من معامل الإنحدار:

1) الإنحدار الخطي البسيط Simple Linear Regression تشير تسمية هذا المعامل "بسيط" الى أنه يتضمن متغير تابع y يعتمد على متغير واحد مستقل x وكلمة خطي تشير الى أن العلاقة ين المتغيرين y و x هي علاقة خطية.

2) الإنحدار المتعدد Multiple Linear Regression هذا النوع من الإنحدار يتضمن إعتماد المتغير y على أكثر من متغير مستقل مثل x1 و x2 ... الخ.

3) الإنحدار غير الخطي Non-Linear Regression إذا كانت العلاقة بين المتغير y والمتغيرات المستقلة غير خطية مثل علاقة أسية أو لوغاريتمية أو تربيعية ... الخ. وهنالك أنواع أخرى مثل الإنحدار الهرمي Hierarchical Regression والإنحدار التدريجي Stepwise Regression وغيرها.

الإنحدار الخطي البسيط Simple Linear Regression

الانحدار الخطي البسيط هو أداة إحصائية تستعمل لبيان العلاقة بين متغيرين كميين بحيث يمكن توقع قيمة المتغير التابع (y) Dependent variable من المتغير المستقل (x) Independent variable . على سبيل المثال، إذا كان الباحث يعرف العلاقة بين النسبة

المئوية لتراكم المادة الجافة وإنتاجية الحنطة فانه يمكنه التنبؤ بالإنتاجية عن طريق الإنحدار الخطي البسيط بمجرد تحديد مستوى تراكم المادة الجافة، بصورة عامة يستعمل الإنحدار للأغراض الآتية:

1) تعد هذه الطريقة تقنية لنمذجة وتحليل البيانات العددية.

2) إستغلال العلاقة بين متغيرين للتنبوء بقيم أحد المتغيرات من خلال قيم المتغير الأخر.

3) التنبؤ وتقدير وإختبار فرضية ونمذجة العلاقات السببية.

هناك العديد من نماذج الانحدار ولكن النموذج الأكثر أهمية والأكثر شيوعاً في الاستعمال هو نموذج الانحدار الخطي بسيط، في هذا النموذج يوجد لدينا المتغير التابع y المعروف أيضاً باسم متغير الإستجابة والمتغير المستقل x المعروف أيضاً باسم المتغير المتنبئ، ويمكن ذكر النموذج على النحو التالي:

$$y = a + bx + e$$

إذ أن:

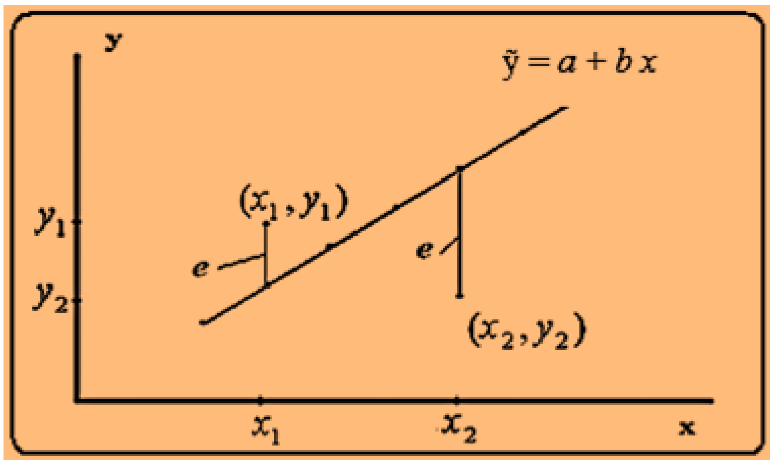
y : المتغير التابع.

x : المتغير المستقل.

a : ثابت الإنحدار وهو الجزء المقطوع من المحور العمودي y الذي يعكس قيمة المتغير التابع y في حالة عدم وجود قيمة للمتغير المستقل x ، بمعنى أخر (x = 0).

b : معامل الإنحدار (الميل) وهو مقدار التغيير في y إذا تغيرت x وحدة واحدة، ويساوي منحدر الخط المستقيم ($a + b\,x$).

e : الخطأ العشوائي الذي يشير إلى الفرق بين القيمة الفعلية للمتغير التابع y والقيمة المقدرة التي يرمز لها $\tilde{y} = a + b\,x$ ، وهذا يعني أن الخطأ العشوائي يساوي $e = y - (a + b\,x)$ ، ويمكن توضيح هذا الخطأ في الشكل البياني أدناه:



الفرق بين نموذجي الإنحدار الخطي البسيط النظري والفعلي

النموذج النظري (المفترض): نموذج يفترض أن العلاقة بين المتغيرين (x و y) تحكمها المعادلة الآتية:

$$y = a + bx + e$$

النموذج الفعلي (المقدر): نموذج عملي يقدره الباحث بجمع البيانات عن المتغيرين (x و y) ومن ثم حساب معاملات الانحدار (a و b) (الثابت والميل) ويعبر عن هذا النموذج بالمعادلة الآتية:

$$\hat{y} = \hat{a} + \hat{b}\,x + e$$

بعد تقدير النموذج يمكن حساب الخطأ العشوائي لكل عينة كمايلي:

$$\hat{e} = y - \hat{y} = y - (\hat{a} + \hat{b}x)$$

تقدير نموذج الانحدار الخطي البسيط

هنالك عدة طرق لتقدير أو حساب نموذج الانحدار الخطي البسيط وكل الطرق تعتمد على حساب قيم معاملات الانحدار (a و b)، وتعد طريقة المربعات الصغرى (OLS) Least Squares Method من أفضل الطرق لأنها تجعل مجموع مربعات الأخطاء العشوائية أقل ما يمكن، ولحساب القيمة المقدرة لمعامل الإنحدار البسيط للمتغير التابع y بدلالة المتغير المستقل x تطبق المعادلة الآتية:

$$\hat{y} = \hat{a} + \hat{b}\,x + e$$

ولحساب قيمة $\hat{b}$ كمايلي:

$$\hat{b} = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - \left(\sum x\right)^2}$$

وتحسب قيمة $\hat{a}$ كمايلي:

$$\hat{a} = \bar{y} - \hat{b}\,\bar{x}$$

مثال(1):

لدراسة علاقة الاستهلاك المحلي لمادة الإسفلت (بالمليون برميل) خلال عدة سنوات اخذنا عشر قراءات تقريبية كما يلي:

y	6	8	9	8	7	6	5	6	5	5
x	10	13	15	14	9	7	6	6	5	5

أوجد معادلة الانحدار الخطي البسيط وتوقع قيمة الاستهلاك عندما يصل إنتاج 16 مليون برميل؟

الحل :

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

$$\hat{a} = \bar{y} - \hat{b}\,\bar{x}$$

x	y	xy	x^2
10	6	60	**100**
13	8	104	**169**
15	9	135	**225**
14	8	112	**196**
9	7	63	**81**
7	6	42	**49**
6	5	30	**36**
6	6	36	**36**
5	5	25	**25**
5	5	25	**25**
90	65	632	**942**

$$\hat{b} = \frac{10(632)-(90)(65)}{10(942)-(90)^2} = 0.36$$

$\bar{y} = 6.5$

$\bar{x} = 9$

$\hat{a} = 6.5 - (0.36)(9) = 3.26$

$\hat{y} = 3.26 + 0.36\ x$

ولتوقع قيمة الاستهلاك نعوض في النموذج اعلاه x=16

$\hat{y} = 3.26 + 0.36\ (16)$

$= 9.02$

أي أن الاستهلاك قد يصل إلى 9.02 مليون برميل، أي ما يعادل 9020000 برميل خلال السنة

مثال(2):

فيما يلي بيانات إضافة السماد المركب لنباتات البطاطا (x)، ومقدار الزيادة في تركيز النشاء في درنات النبات (y)، عند قياس عينة من 10 نباتات لكل معاملة سماد.

Fertilizer (x)	10	11	14	15	20	25	46	50	59	70
Starch (y)	10	10	12	12	13	13	19	15	16	20

المطلوب :

1) ارسم نقاط الانتشار؟

2) قدر معادلة انحدار تركيز النشاء في الدرنات (y) على كمية السماد المضاف (x)؟

3) فسر معادلة الانحدار؟

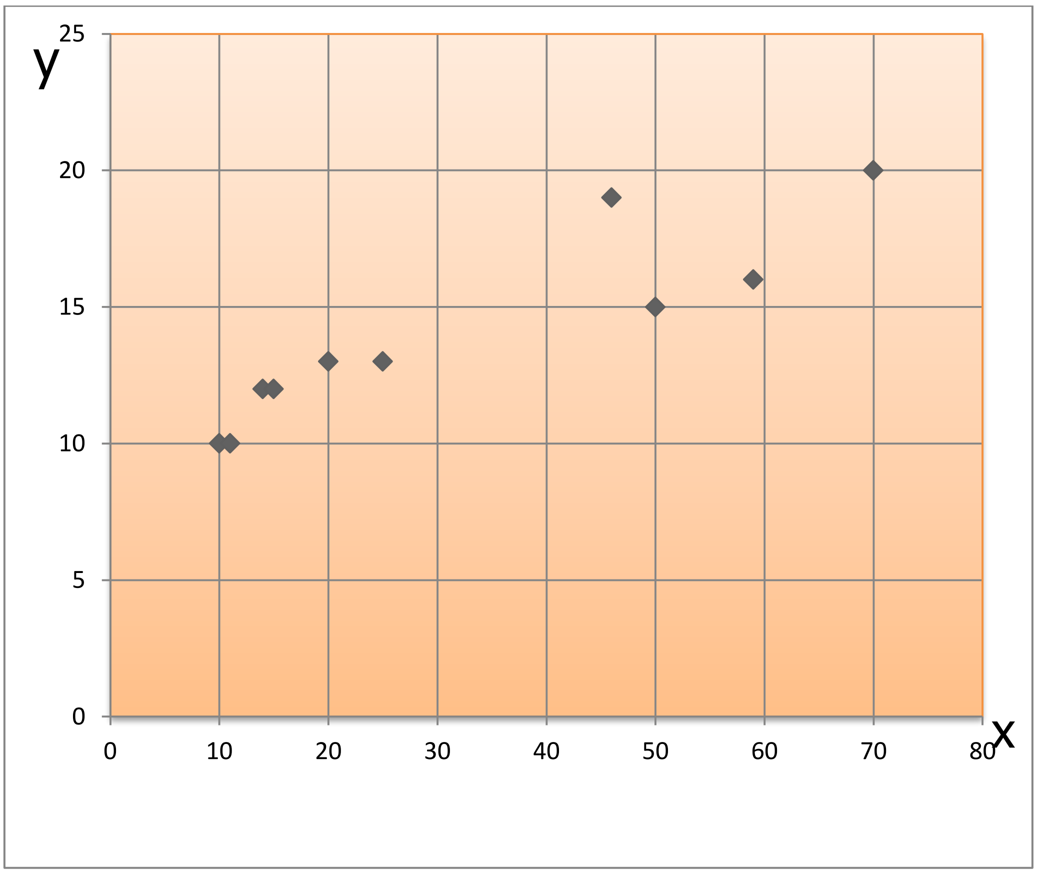
4) ما هو مقدار الزيادة في تركيز النشاء عند التسميد بمقدار 50 كغم.هـ$^{-1}$ (x = 50)؟

5) ما هي مقدار الخطأ العشوائي؟

6) ارسم معادلة الانحدار على نقط الانتشار، وما هي توقعاتك لشكل العلاقة؟

الحل:

1) رسم نقاط الانتشار



2) حساب نموذج الإنحدار المقدر بتطبيق المعادلة $\hat{y} = \hat{a} + \hat{b}\,x + e$ تتبع الخطوات الآتية:

– حساب قيم العلاقات بين x و y كما في الجدول الآتي:

x	y	xy	x^2	y^2
10	10	100	100	100
11	10	110	121	100
14	12	168	196	144
15	12	180	225	144
20	13	260	400	169
25	13	325	625	169
46	19	874	2116	361
50	15	750	2500	225
59	16	944	3481	256
70	20	1400	4900	400
$\sum \boldsymbol{x} = \mathbf{320}$	$\sum y = 140$	$\sum xy = 5111$	$\sum x^2 = 14664$	$\sum y^2 = 2068$

$\sum x^2 =$ 14664 $\quad n = 10 \quad \sum y^2 = 2068$

$\sum x = 320 \quad \bar{y} = \frac{140}{10} = 14$

$\sum y = 140 \quad \bar{x} = \frac{320}{10} = 32$

$$\hat{b} = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - \left(\sum x\right)^2}$$

$$= \frac{(10)(5111) - (320)(140)}{(10)(14664) - (320)^2} = \frac{6310}{44240} = 0.1426$$

الان يمكن حساب قيمة $\hat{a}$ بتطبيق المعادلة كمايلي:

$$\hat{a} = \bar{y} - \hat{b}\,\bar{x} = 14 - (0.1426)(32) = 9.4368$$

اذن حساب قيمة النموذج الفعلي:

$$\hat{y} = \hat{a} - \hat{b}\,x = 9.4368 + 0.1426\,x$$

3) تفسير معادلة الإنحدار:

1) معامل الإنحدار الثابت $\hat{a} = 9.4368$ يدل على أنه في حالة عدم إستعمال التسميد (x = 0) فان تركيز النشاء يزداد بمقدار 9.4368 ملغم..

2) معامل الإنحدار (الميل) $\hat{b} = 0.143$ يدل على أنه كلما زادت كمية السماد كغم. واحد فان هنالك زيادة تحدث في تركيز النشاء بمقدار 0.1426 ملغم..

4) ما هو مقدار الزيادة في تركيز النشاء عند التسميد بمقدار 50 كغم. (x = 50)؟

$$\hat{y} = 9.4368 + 0.1426\,x$$
$$\hat{y} = 9.4368 + 0.1426\,(50) = 16.59$$

5) لتقدير قيمة الخطأ العشوائي عند (x = 50) تستخرج قيمة تقاطع $y_{x=50}$ من المخطط البياني أعلاه والتي تساوي 15 ثم تطبق المعادلة كمايلي:

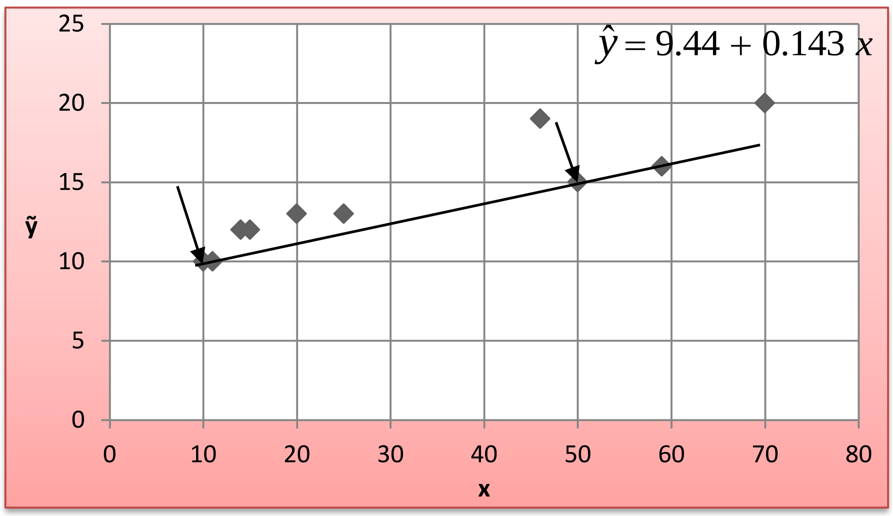
$$\hat{e}_{x=50} = y_{x=50} - \hat{y}_{x=50} = 15 - 16.59 = -1.59$$

6) رسم معادلة الانحدار على نقاط الانتشار:

ملاحظة: من البديهي يمكن رسم خط مستقيم إذا علم أي نقطتين على ذلك الخط، فاذا كانت:

x	50	10
$\hat{y}$	16.59	10.87

لذا يمكن رسم الخط المستقيم كمايلي:



يستنتج من ذلك وجود علاقة الخط المستقيم بين المتغيرين

الإنحدار الخطي المتعدد *Multiple Linear Regression*

يعد الإنحدار الخطي المتعدد من الأساليب الإحصائية المتقدمة والتي تضمن دقة الإستدلال من أجل تحسين نتائج البحث عن طريق الإستخدام الأمثل للبيانات في إيجاد علاقات سببية بين الظواهر موضوع البحث .

والإنحدار الخطي المتعدد هو عبارة عن إيجاد معادلة رياضية تعبر عن العلاقة بين متغيرين وتستعمل لتقدير قيم سابقة ولتنبؤ قيم مستقبلية ، وهو عبارة أيضاً عن إنحدار للمتغير التابع (**Y**) على العديد من المتغيرات المستقلة $X_1 , X_2 , ...X_K$ لذا فهو يستخدم في التنبؤ بتغيرات المتغير التابع الذي يؤثر فيه عدة متغيرات مستقلة أي تعتمد فكرته على العلاقات الدلالية التي تستخدم ما يعرف بشكل التشتت أو االانتشار ، فبإمكاننا التنبؤ بالمستوى الرقمي في فعالية رمي المطرقة على سبيل المثال إعتماداً على دراسة حالات أخرى للرامي كالعمر الزمني والعمر التدريبي والمهارة والمواصفات الجسمية وغيرها .

إن الإنحدار الخطي المتعدد ليس مجرد أسلوب واحد وإنما مجموعة من الأساليب التي يمكن أستخدامها لمعرفة العلاقة بين متغير تابع مستمر وعدد من المتغيرات المستقلة التي عادةً ما تكون مستمرة)

والمعادلة الخطية في الإنحدار الخطي المتعدد هي :

$$Y = a + b_1X_1 + b_2X_2 + \dots\dots\dots + e$$

حيث أن **Y** = المتغير التابع

a = قيمة ثابتة ***Constant*** أو ***Intercept***

b$_1$ = ميل ألإنحدار ***y*** على المتغير المستقل الأول

b$_2$ = ميل ألإنحدار ***y*** على المتغير المستقل الثاني

X$_1$ = المتغير المستقل الأول

X$_2$ = المتغير المستقل الثاني

ويمكن استخدام الإنحدار الخطي المتعدد في حالة توافر الشروط التالية :

1. أن تكون العلاقة خطية بين المتغيرات المستقلة والمتغير التابع .

2. أن تكون البيانات موزعة توزيعاً طبيعياً للمتغيرات المستقلة والمتغير التابع .
3. يجب أن تكون قيم المتغير التابع من المستوى الترتيبي على الأقل .

بعد الحصول على نتائج معادلة الانحدار يجب علينا أن نبين هل أن هذه المعاملات مقبولة من الناحية الإحصائية أي معنوية احصائياً مع التنويه بأن المعنوية تكون لكل معامل على حدة .

ولكي نحكم على معنوية معاملات الإنحدار نستعين باختبار T ومستوى الاحتمالية المقابل له وبالطبع فإن برنامج ***SPSS*** سيقوم تلقائياً باستخراج اختبار T ومستوى الاحتمالية المقابل له .

كما سيتم الحصول على إحصائيات تستخدم لمعرفة المعنوية الإجمالية للنموذج ومنها (R) ، (R^2) ، (R^{2-}) .

فالأول R هو معامل الأرتباط البسيط والذي يقيس قوة العلاقة بين متغيرين أو أكثر ، أما R^2 فهو يسمى بمعامل التحديد والذي يستخدم لمعرفة القوة التفسيرية للنموذج المقدر (المعادلة المقدرة) في حالة الإنحدار الخطي البسيط (متغير مستقل واحد مع متغير معتمد واحد) ، أما R^{2-} فهو يستخدم لتفسير القوة التفسيرية لنموذج الانحدار الخطي المتعدد (لأنه يأخذ بنظر الاعتبار عدد المتغيرات المستقلة ولذلك يسمى بالمصحح لأنه بالأصل مشتق من R^2) .

كما نستخدم أيضاً أحصائية F للحكم على معنوية النموذج المقدر ككل عند مستوى معنوية معين .

معامل التحديد Determination Coefficient

هو مقياس لتقدير دقة معامل الانحدار ويرمز له R^2 وذلك لأنه يساوي مربع معامل الإرتباط البسيط ويأخذ هذا المعامل قيم بين 0 الى 1 أي أنه $0 \leq r^2 \leq 1$ وكلما إقتربت قيمة معامل التحديد من 1 فان ذلك يدل على قلة قيمة الخطأ العشوائي، مثال ذلك لو كانت قيمة معامل التحديد لمتغيرين تساوي 0.87 فهذا يفسر (يدل) على أن معادلة الإنحدار تفسر 87% من التغير الحاصل في المتغير التابع y حدثت بسبب التغير الحاصل في المتغير المستقل x والباقي من التغير البالغ 13% حدث بسبب عوامل أخرى غير المتغير المستقل x، ويتم حساب R^2 على النحو التالي:

$$SS_e = \sum y^2 - \hat{a}\sum y - \hat{b}\sum xy$$
$$= 2068 - 9.44(140) - 0.1426(5111) = 17.5714$$
$$SS_T = \sum y^2 - n(\bar{y})^2 = 2068 - 10(14)^2 = 108$$
$$SS_R = SS_T - SS_e = 108 - 17.5714 = 90.4286$$
$$R^2 = \frac{SS_R}{SS_T} = \frac{90.4286}{108} = 0.83$$

SS_e : مجموع مربعات الخطأ التجريبي.

SS_T : مجموع المربعات الكلية.

SS_R : مجموع مربعات الإنحدار.

مثال (2) :

أراد أحد الباحثين التنبؤ بالزيادة الحاصلة في تركيز المادة القلويدية الفعالة (y) في كالس نبات الداتورة عند إضافة تراكيز من Ascorbic acid (x) الى الوسط الغذائي للكالس، وعند تقدير 13 عينة من الكالس حصل الباحث على البيانات الآتية:

Ascorbic acid (x)	**0.5**	**0.75**	**1.0**	**1.25**	**1.5**	**1.75**	**2.0**	**2.25**	**2.5**	**2.75**	**3.0**	**3.25**	**3.5**
Alkaloids (y)	10	8	12	12	14	12	16	18	17	20	18	20	21

المطلوب :

1) ارسم نقاط الانتشار؟

2) قدر معادلة انحدار تركيز المادة القلويدية الفعالة (y) بدلالة تراكيز من Ascorbic acid (x)؟

3) فسر معادلة الانحدار؟

4) ما هو مقدار الزيادة في تركيز المادة القلويدية الفعالة عند إضافة تركيز 3.0 من Ascorbic acid (x = 3.0)؟

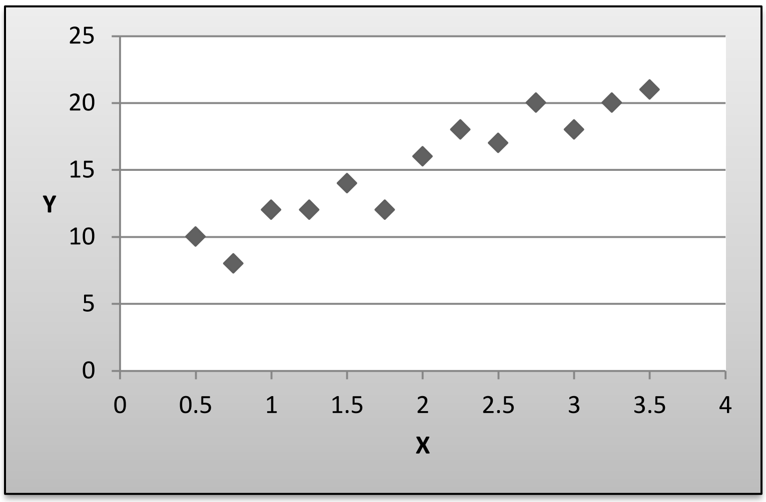
5) ما هي مقدار الخطأ العشوائي؟

6) ما هي قيمة معامل التحديد وكيف يفسر؟

7) ارسم معادلة الانحدار على نقط الانتشار، وما هي توقعاتك لشكل العلاقة؟

الحل:

1) رسم نقاط الانتشار؟



2) حساب نموذج الإنحدار المقدر بتطبيق المعادلة $\hat{y} = \hat{a} + \hat{b}\,x + e$ تتبع الخطوات الآتية:

– حساب قيم العلاقات بين x و y كما في الجدول الآتي:

n	x	y	x^2	y^2	xy
1	0.5	10	0.25	100	5
2	0.75	8	0.5625	64	6
3	1	12	1	144	12
4	1.25	12	1.5625	144	15
5	1.5	14	2.25	196	21
6	1.75	12	3.0625	144	21
7	2	16	4	256	32

8	2.25	18	5.0625	324	40.5
9	2.5	17	6.25	289	42.5
10	2.75	20	7.5625	400	55
11	3	18	9	324	54
12	3.25	20	10.5625	400	65
13	3.5	21	12.25	441	73.5
$\boldsymbol{n = 13}$	$\sum x = 26$	$\sum y = 198$	$\sum x^2 = 63.375$	$\sum y^2 = 3226$	$\sum xy = 442.54$

$n = 13 \qquad \sum xy = 442.54 \qquad \sum x = 26 \qquad \sum y = 198$

$\sum x^2 = 63.375 \qquad \sum y^2 = 3226 \qquad \bar{y} = \frac{198}{13} = 15.23 \qquad \bar{x} = \frac{26}{13} = 2$

– حساب قيمة $\hat{b}$ بتطبيق المعادلة كمايلي:

$$\hat{b} = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - \left(\sum x\right)^2} = \frac{(13)(442.54) - (26)(198)}{(13)(63.375) - (26)^2}$$

$$= \frac{605.02}{147.875} = 4.09$$

اذن قيمة $\hat{a}$ بتطبيق المعادلة كمايلي:

$$\hat{a} = \bar{y} - \hat{b}\,\bar{x} = 15.23 - (4.09)(2) = 7.05$$

النموذج الفعلي:

$$\hat{y} = \hat{a} - \hat{b}\,x = 7.05 + 4.09\,x$$

3) تفسير معادلة الإنحدار:

1) معامل الإنحدار الثابت $\hat{a} = 7.05$ يدل على أنه في حالة عدم إستعمال Ascorbic acid (x = 0) فان تركيز المادة القلويدية الفعالة تزداد بمقدار 7.05 ملغم.100غم$^{-1}$.

2) معامل الإنحدار $\hat{b} = 4.09$ يدل على أنه كلما زاد تركيز Ascorbic acid وحدة واحد فان هنالك زيادة تحدث في تركيز المادة القلويدية الفعالة بمقدار 4.09 ملغم.100غم$^{-1}$.

4) ما هو مقدار الزيادة في تركيز المادة القلويدية الفعالة عند تركيز الحامض 3.0 ملغم.لتر$^{-1}$ (x = 3.0)؟

$$\hat{y} = 7.05 + 4.09\,x$$
$$\hat{y} = 7.05 + 4.09\,(3.0) = 19.32$$

5) لتقدير قيمة الخطأ العشوائي عند (x=3.0) تستخرج قيمة تقاطع $y_{x=3.0}$ من المخطط البياني أعلاه والتي تساوي 18 ثم تطبق المعادلة كمايلي:

$$\hat{e}_{x=3.0} = y_{x=3.0} - \hat{y}_{x=3.0} = 18 - 19.32 = -1.32$$

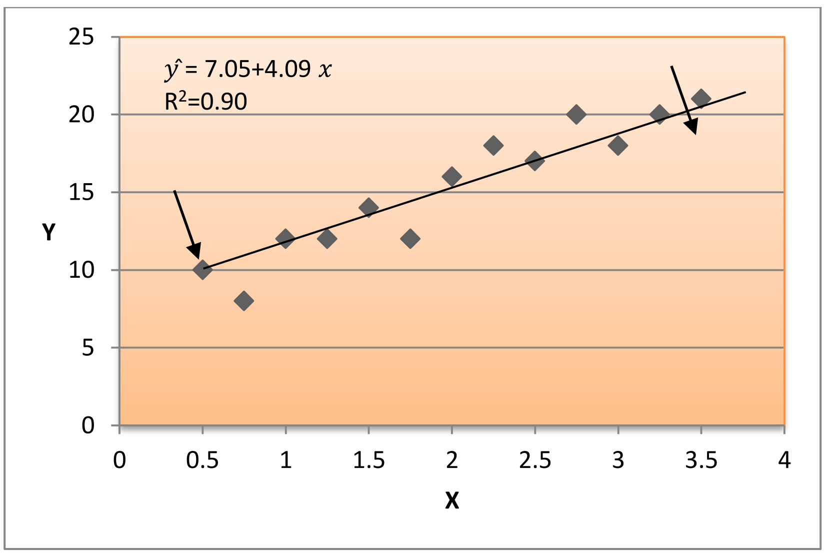
6) حساب قيمة معامل التحديد R^2 وكيف يفسر؟

$$SS_e = \sum y^2 - \hat{a}\sum y - \hat{b}\sum xy$$
$$= 3226 - 7.05\,(198) - 4.09\,(442.54) = 20.1$$
$$SS_T = \sum y^2 - n(\bar{y})^2 = 3226 - 13(15.23)^2 = 210.6$$
$$SS_R = SS_T - SS_e = 210.6123 - 20.1114 = 190.5$$
$$R^2 = \frac{SS_R}{SS_T} = \frac{190.5}{210.6} = 0.90$$

7) رسم معادلة الانحدار على نقاط الانتشار:

ملاحظة: من البديهي يمكن رسم خط مستقيم إذا علم أي نقطتين على ذلك الخط، فاذا كانت:

x	3.0	0.5
$\hat{y}$	19.32	10.0



يستنتج من ذلك وجود علاقة الخط المستقيم بين المتغيرين

مثال (3):

البيانات الآتية تمثل المتغير y والمتغير x لتجربة ما كمايلي:

$n = 7 \qquad \sum xy = 2093 \qquad \sum x = 103 \qquad \sum y = 144$

$\sum x^2 = 1531 \qquad \sum y^2 = 3012 \qquad \bar{y} = \frac{144}{7} = 20.5 \qquad \bar{x} = \frac{103}{7} = 14.7$

المطلوب:

1- تقدير معادلة انحدار المتغير y بدلالة المتغير x .

2- تقدير المتغير y عندما يكون المتغير x يساوي 10.

3- حساب معامل التحديد وتفسير معناه.

الحل:

1) تقدير معادلة انحدار المتغير y بدلالة المتغير x .

- حساب قيمة $\hat{b}$ بتطبيق المعادلة كمايلي:

$$\hat{b} = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - \left(\sum x\right)^2} = \frac{(7)(2093) - (103)(144)}{(7)(1531) - (103)^2}$$

$$= \frac{-181}{108} = -1.68$$

- حساب قيمة $\hat{a}$ بتطبيق المعادلة كمايلي:

$$\hat{a} = \bar{y} - \hat{b}\,\bar{x} = 20.5 - (-1.68)(14.7) = 45.29$$

النموذج الفعلي:

$$\hat{y} = \hat{a} - \hat{b}\,x = 45.29 - 1.68\,x$$

2) تقدير المتغير y عندما يكون المتغير x يساوي 10.

$$\hat{y} = 45.29 - 1.68\,x$$

$$\hat{y} = 45.29 - 1.68\,(10) = 28.49$$

3) حساب قيمة معامل التحديد R^2 وكيف يفسر؟

$$SS_e = \sum y^2 - \hat{a}\sum y - \hat{b}\sum xy$$

$$= 3012 - 45.29\,(144) - (-1.68)(2093) = 6.48$$

$$SS_T = \sum y^2 - n(\bar{y})^2 = 3012 - 7(20.5)^2 = 70.25$$

$$SS_R = SS_T - SS_e = 70.25 - 6.48 = 63.77$$

$$R^2 = \frac{SS_R}{SS_T} = \frac{63.77}{70.25} = 0.90$$

قيمة معامل التحديد لمتغيرين تساوي 0.90 وهذا يفسر (يدل) على أن معادلة الإنحدار تفسر 90% من التغير الحاصل في المتغير التابع y حدثت بسبب التغير الحاصل في المتغير المستقل x والباقي من التغير البالغ 10% حدث بسبب عوامل أخرى غير المتغير المستقل x.

الفصل الثامن

اختبارات الفروض الإحصائية

مقدمة:

يتعرض الإنسان في كثير من الحالات وفي مجالات العمل الى مواقف معينة تتطلب منه اتخاذ قرار بناء على معلومات محسوبة من عينه ، وعليه يجب اتخاذ هذا القرار بأقل قدر ممكن من الخطأ.

مثال 1:نفرض أن باحثا اجتماعيا ادعى ان متوسط اعمار طلاب الجامعة لايختلف عن متوسط أعمار الطالبات. للتأكد من ذلك فإن الشئ الطبيعي أن نقوم بحصر اعمار الطلاب والطالبات ومنها نحسب المتوسط لكل منهما ثم نقرر من منهما اكبر..ولكن عملية الحصر صعبة ومجهدة لذلك نضطر الى اختيار عينة عشوائية من بين الطلاب وعينة عشوائية من بين الطالبات ونحسب متوسط العمر في كل عينة منهما،،،،فإذا كان متوسط عمر الطالب هو 24 وكان متوسط عمر الطالبة هو 22 فهل يعني ذلك ان متوسط عمر الطالب اكبر من متوسط عمر الطالبة ؟؟ هل الفرق راجع لمجرد الصدفة ؟؟ متى يكون الفرق نتيجة للصدفة؟؟ متى يكون الفرق دالا على وجود اختلاف حقيقي أو جوهري بين متوسطي المجتمعين الأصليين ..

مفاهيم مهمة :

هناك بعض المفاهيم المتعلقة باختبارات الفروض لابد من معرفتها:

- الفرض الاحصائي statistical hypothesis

هو عبارة عن إدعاء او تخمين معين حول معلمة من معالم المجتمع ويكون المطلوب اختبار صحة هذا الادعاء أو التخمين ...هناك نوعين من الفروض :

- فرض العدم (null hypothesis) ويرمز له بالرمز H_o ويصاغ في صورة عدم وجود فرق أو عدم وجود علاقة أو عدم وجود تغير – مثال : في مثال أعمار الطلاب وطالبات الجامعة فإن فرض العدم هو

H_o : نفترض عدم وجود اختلاف بين متوسطي اعمار الطلاب والطالبات

- الفرض البديل (alternative hypothesis) ويرمز له بالرمز H_1 وهو الفرض الذي يجب أن يكون صحيحا اذا كان فرض العدم غير صحيح – مثال : في مثال أعمار الطلاب وطالبات الجامعة فإن الفرض البديل هو

H_1: يوجد اختلاف حقيقي وليس ظاهري بين متوسط اعمار الطلاب والطالبات .

مستوى المعنوية α ودرجة الثقة (1 – α):

- إن القرار الذي سوف نتخذه بناء على الاختبار الإحصائي لا يمكن اعتباره صحيح 100 % فهناك مقدار من الخطأ لأن المعلومات التي نتخذ قرارنا بناء عليها بيانات مأخوذة من عينة وليس من المجتمع الأصلي
- في اختبار فرض معين، فإن مقدار ثقتنا في القرار المتخذ بالرفض أو القبول يسمى بدرجة الثقة ويرمز له بالرمز (1 – α) كما وأن مقدار عدم الثقة أو مقدار الخطأ يسمى بمستوى المعنوية ويرمز له بالرمز α

وعادة يحدد الباحث مستوى المعنوية أو درجة الثقة قبل البدء في عملية الاختبار.

عند اختبار فرض العدم H_o ضد الفرض البديل H_1 نجد أننا امام احدى الحالات الاربع الاتية :

	H0 صحيح	H0 خطأ
قبول H0	قرار سليم	خطأ من النوع الثاني
رفض H0	خطأ من النوع الاول	قرار سليم

1) أن يكون فرض العدم صحيحا ويكون القرار بقبوله ....وهذا قرار سليم
2) أن يكون فرض العدم صحيحا ويكون القرار برفضه .... وهذا قرار خاطئ

(الخطأ من النوع الأول : رفض H_o عندما يكون H_o صحيحا ويرمز لحجم هذا الخطأ بالرمز α)

3) أن يكون فرض العدم خطأ ويكون القرار برفضه وهذا قرار سليم

4) أن يكون فرض العدم خطأ ويكون القرار بقبوله ..وهذا قرار خاطئ (الخطأ من النوع الثاني : قبول H_o عندما يكون H_o خاطئ ويرمز لحجم هذا الخطأ بالرمز β)

- احتمال الوقوع في الخطأ من النوع الأول يسمى مستوى المعنوية ويرمزله بالرمز α أي ان α = احتمال رفض فرض العدم H_o عندما يكون صحيح = مستوى المعنوية

- احتمال الوقوع في خطأ من النوع الثاني يرمز له بالرمز β أي أن

β = احتمال قبول فرض العدم H_o عندما يكون خطأ

خطوات اختبار الفرض الإحصائي حول متوسط المجتمع لعينة كبيرة

لإجراء الاختبار الإحصائي فإننا نتبع الخطوات التالية :

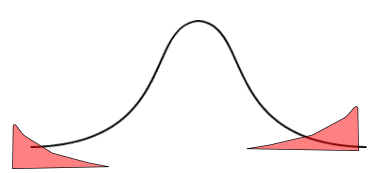
1- صياغة فرض العدم Ho

$Ho : \mu = \mu 0$

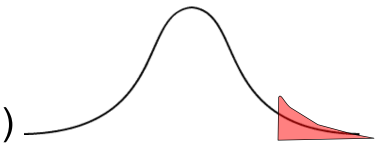
والفرض البديل هو احد الحالات التالية :

1- $H1: \mu \neq \mu 0$

(اختبار من طرفين)

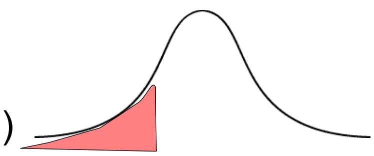


2- $H1: \mu > \mu 0$



(اختبار من طرف واحد ،الجهة اليمنى)

3- H1:μ < μ0



(اختبار من طرف واحد ،الجهة اليسرى)

2- تحديد قيمة احصاءة الاختبار(قيمة Z المحسوبة):

حيث أن هذا الاحصاءة يتبع تقريبا توزيعا طبيعيا قياسياً

$$Z_C = \frac{\overline{X} - \mu_O}{\frac{\sigma}{\sqrt{n}}}$$

3- تحديد القيمة الجدولية و تحدد على حسب نوع الاختبار وقيمة α :

نوع الاختبار	مستوى المعنوية α	درجه الثقة $(1-\alpha)$	الدرجة المعيارية
اختبار من طرفين	$5\% = 0.05$	$95\% = 0.95$	$= Z_{\alpha/2} \pm 1.96$
	$1\% = 0.01$	$99\% = 0.99$	$= Z_{\alpha/2} \pm 2.58$

$95\% = 0.95$

نوع الاختبار	مستوى المعنوية α	درجه الثقة $(1-\alpha)$	القيمة الجدولية (القيمة الحرجة)
اختبار من طرف واحد	$5\% = 0.05$	$95\% = 0.95$	$Z\alpha$ =1.64
(الجهة اليمنى)	$1\% = 0.01$	$99\% = 0.99$	$Z\alpha$ =2.33
اختبار من طرف	$5\% = 0.05$	$95\% = 0.95$	$Z\alpha$ = -1.64
واحد(الجهةاليسرى)	$1\% = 0.01$	$99\% = 0.99$	$Z\alpha$ = -2.33

4- اتخاذ القرار:

نتخذ القرار بناءاً على قيمة احصاءة الاختبار

نرفض H0 إذا وقعت قيمة احصاءة الاختبار في منطقة الرفض

لا نرفض H0 إذا وقعت قيمة احصاءة الاختبار في منطقة القبول

إذا كان الاختبار من طرفين : نقبل فرض العدم إذا تحققت المعادلة التالية :

$$-Z_{\alpha/2} < Z_C < Z_{\alpha/2}$$

نرفض فرض العدم إذا تحققت إحدى المعادلتين : $Z_C > Z_{\alpha/2}$ ، $Z_C < -Z_{\alpha/2}$

إذا كان الاختبار من طرف واحد الجهة اليمنى :

نقبل فرض العدم إذا تحققت المعادلة : $Z_C < Z_\alpha$

نرفض فرض العدم إذا تحققت المعادلة : $Z_C > Z_\alpha$

إذا كان الاختبار من طرف واحد الجهة اليسرى :

نقبل فرض العدم إذا تحققت المعادلة : $Z_C > -Z_\alpha$

نرفض فرض العدم إذا تحققت المعادلة : $Z_C < -Z_\alpha$

مثال (1):

شركة متخصصة في صناعة لعب الأطفال تعاقدت لشراء نوع جديد من الخيوط الصناعية، يدعي صانع هذه الخيوط أن متوسط قوة تحمل الخيط 15 كغم بانحراف معياري نصف كغم .
ولاختبار صحة ادعاء الصانع أخذت عينة عشوائية من 50 خيطاً وتم اختبارها فوجد أن متوسط قوة التحمل في العينة 14.8كغم . فهل يمكننا تأييد ادعاء المدير؟ استخدمي مستوى معنوية 5 % .

الحل:

$$n=50 \qquad \mu_o=15\ kg$$
$$\overline{X}=14.8\ kg \qquad \sigma=0.5\ kg$$

1- صياغة الفرض الإحصائي:

الفرض العدم : عدم وجود فرق معنوي بين متوسط الخيوط القديمة والحديثة $H_o:\mu=15$

ضـــــــــد

الفرض البديل : وجود فرق معنوي بين متوسط الخيوط القديمة والحديثة $H_1:\mu\neq 15$

حيث μ هي متوسط قوة تحمل الخيط .

2- إيجاد قيمة إحصاء الاختبار:

لان σ مجهولة والعينة كبيرة فانه يمكن استخدام S بدلاً من σ :

$$Z_C=\frac{\overline{X}-\mu_O}{\frac{\sigma}{\sqrt{n}}}=\frac{14.8-15}{\frac{0.5}{\sqrt{50}}}=-2.83$$

3- تحديد القيمة الجدولية :

$$Z_{\alpha/2}=\pm 1.96$$

4- اتخاذ القرار:

بما أن القيمة المحسوبة وقعت في منطقة الرفض، فإن القرار هو: رفض فرض العدم أي أن الادعاء غير صحيح وأن هناك فرق معنوي بين المتوسط الحقيقي والمتوسط المدعى.

مثال(2) :

إذا كان من المعروف أن جسم الإنسان البالغ يحتاج يوميا في المتوسط 800 ميللغرام من الكالسيوم لكي يقوم بوظائفه خير قام.ويعتقد احد علماء التغذية أن الأفراد ذوي الدخل المنخفض لا يستطيعون تحقيق هذا المتوسط ، ولاختبار ذلك تم اختيار عينة من 50 شخصا بالغا من بين ذوي الدخل المنخفض فكان متوسط ما يتناوله من كالسيوم يوميا هو 755.3 ميللغرام والانحراف المعياري هو 239.3 ميللغرام .فهل تدل هذه النتائج على أن متوسط ما يتناوله الأشخاص البالغون من ذوي الدخل المنخفض من كالسيوم يقل عن 800 ميللغرام؟ استخدمي مستوى معنوية 0.05

الحل :

1- صياغة الفرض الإحصائي:

فرض العدم هو

$$Ho\mu \ : \ = 800$$

والفرض البديل

$$H1 : \mu \ < 800$$

حيث μ هي متوسط ما يتناوله الإنسان البالغ ذوي الدخل المنخفض من الكالسيوم .

2- إيجاد قيمة إحصاء الاختبار:

لان σ مجهولة والعينة كبيرة فانه يمكن استخدام $s=239.3$ بدلا منها .وبالتعويض نجد ان قيمة احصاءة الاختبار هي

$$Z= \frac{\bar{x}-\mu_0}{s/\sqrt{n}} = \frac{755.3-800}{239.3/\sqrt{50}} = -1.32$$

3- تحديد القيمة الجدولية :

ونلاحظ هنا أن الاختبار ذو جانب (طرف) واحد هو الجانب (الطرف) الأيسر وحيث أن مستوى المعنوية α=0.05 فانه من جدول التوزيع الطبيعي القياسي نجد أن القيمة الحرجة هي :

$Z\,\alpha$ = -1.64

4- اتخاذ القرار:

بما أن قيمة الاحصاءة 1.32− اكبر من القيمة الحرجة 1.64- وهي تقع في منطقة القبول وبالتالي فإننا لانرفض فرض العدم H_0 وهو أن متوسط ما يتناوله الإنسان البالغ ذو الدخل المنخفض من الكالسيوم يساوي 800 ميللجرام .

مثال (3) :

في عينة عشوائية مكونة من تسجيل 100 حالة وفاة في قرية معينة تبين أن متوسط العمر في العينة 67.5 عاما والانحراف المعياري 8 أعوام. فهل هذا يوضح أن متوسط العمر في هذه القرية اكبر من 65 عاما ؟

تحت مستوى معنوية 5 %.

الحل:

نفرض أن μ متوسط العمر في هذه القرية .

1- صياغة الفرض الإحصائي:

$$H_0 : \mu = 65$$

$$H_1 : \mu > 65$$

2- إيجاد قيمة إحصاء الاختبار:

$$Z = \frac{\bar{x}-\mu_0}{s/\sqrt{n}} = \frac{67.5-65}{8/\sqrt{100}} = 3.125$$

3- تحديد القيمة الجدولية :

ونلاحظ هنا أن الاختبار ذو جانب (طرف) واحد هو الجانب (الطرف) الأيمن وحيث أن مستوى المعنوية $\alpha=0.05$ فانه من جدول التوزيع الطبيعي القياسي نجد أن القيمة الحرجة هي :

$Z_\alpha = 1.64$

4- اتخاذ القرار:

نجد أن قيمة Z المحسوبة 3.125 اكبر من القيمة الجدولية 1.64 لهذا فان Z المحسوبة تقع في منطقة الرفض لهذا فان القرار هو رفض H_0

ونستنتج من ذلك أن متوسط العمر في هذه القرية اكبر من 65 عاما .

الفصل التاسع

جزء من التوزيعات

التوزيع الطبيعي

يعتبر التوزيع الطبيعي من أهم التوزيعات الاحتماليه في علم الاحصاء لأنه يمثل كثيراً من الظواهر التي تقابلنا في الحياة العمليه مثل الأطوال و الأوزان و الأعماروردجات الحراره والدخول الشهريه ... و غيرها من الظواهر المتصلة .

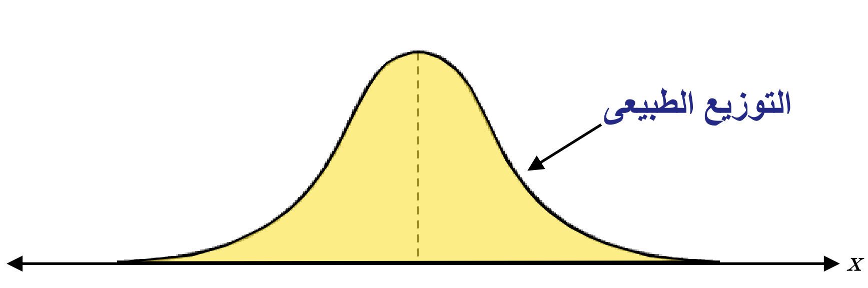
ولرسم التوزيع الطبيعى نستخدم المنحنى الاحتمالي للتوزيع الطبيعي

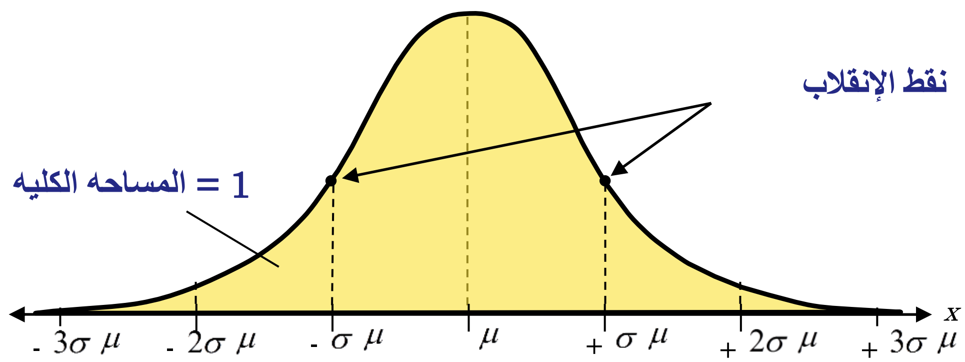
إذا كانت x متغيراً عشوائياً متصلاً يتبع توزيعاً طبيعياً متوسطه μ و تباينه σ^2 يمكن رسم المنحنى الاحتمالي للتوزيع الطبيعي بإستخدام المعادله

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \; e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad ; \; -\infty \le x \le \infty$$

فإنه يقال إن x تتبع توزيعاً طبيعياً متوسطه μ و تباينه σ^2. و تكتب اختصاراً

$$X \sim N(\mu , \sigma^2)$$





خصائص المنحنى الإحتمالى للتوزيع الطبيعى

1. طرفا التوزيع تمتد من $-\infty$ إلى ∞
2. متماثل حول العمود الذى يمر بقمته أى عند $X=\mu$ لذلك فإن هذا العمود يجزئ المنحنى الطبيعى القياسى إلى قسمين متماثلين فى الشكل والمساحه.
3. يشبه الناقوس من حيث الشكل.
4. المنحنى له نقطتا إنقلاب عند $X=\mu\pm\sigma$
5. المساحه الكليه تحت المنحنى تساوى الواحد الصحيح .
6. نظرا لتطابق جانبيه وتوسط تفرطحه فإن مقاييس النزعه المركزيه الثلاثه (الوسط الحسابى والوسيط والمنوال) تلتقى فى نقطه واحده (فى منتصف التوزيع) أى أن الوسط الحسابى = الوسيط = المنوال
7. نلاحظ أن 99.7% من قيم المتغير تقع تقريبا فى الفتره ($\mu-3\sigma\ ,\ \mu+3\sigma$) أى أنه نادرا ما نجد قيمه من قيم X خارج هذه الفتره. كما أن 95.5% من قيم المتغير X تقع فى الفتره ($\mu-2\sigma\ ,\ \mu+2\sigma$) وأن% 68 منها تقع فى الفتره($\mu-\sigma\ ,\ \mu+\sigma$)

ويمكن حساب إحتمال وقوع X فى أى فتره تريدها وذلك بإيجاد قيمه تكامل داله الكثافه الإحتماليه فى هذه الفتره أى بإيجاد المساحه المحصوره تحت منحنى هذه الداله داخل هذه الفتره . فمثلا الإحتمال $P(a \le X \le b)$ يعطى بالمساحه المظلله فى الشكل الآتى:-

مثال

إذا كان متوسط أطوال المجندين فى وحده عسكريه ما 70 بوصه بإنحراف معيارى قدره 2 بوصه صف هذه البيانات مستخدما القوانين العمليه فى الفقره 7.

الحل:

1. 68% من الحالات فى تلك الوحده يتراوح أطوالهم بين $\mu-\sigma$ ، $\mu+\sigma$

$=(70 - 2\times 1)$،$(70 + 1\times 2)$

$=70 - 2$، $70 + 2$

$=68$ ، 72

أى أن 68 % من المجندين فى تلك الوحده يتراوح أطوالهم بين 68 بوصه و 72 بوصه .

2. 9505% من الحالات فى تلك الوحده يتراوح أطوالهم بين $\mu-2\sigma$ ، $\mu+2\sigma$

$=(70 - 2\times 2)$،$(70 + 2\times 2)$

$=(70 - 4)$،$(70 + 4)$

$=66$ ، 74

أى أن 95.5 % من المجندين فى تلك الوحده يتراوح أطوالهم بين 66 بوصه و 74 بوصه .

3. 99.7% من الحالات فى تلك الوحده يتراوح أطوالهم بين $\mu-3\sigma$ ، $\mu+3\sigma$

$=(70 - 3\times 2)$،$(70 + 3\times 2)$

$=(70 - 6)$،$(70 + 6)$

76 ، 64=

أى أن 99.7 % من المجندين فى تلك الوحده يتراوح أطوالهم بين 64 بوصه و 76 بوصه .

و تختلف التوزيعات الطبيعية باختلاف كل من المتوسط و التباين و لتسهيل حساب الاحتمالات في حالة التوزيع الطبيعي فإننا لابد وأن نتعرض للتوزيع الطبيعي القياسي.

التوزيع الطبيعى القياسى

التوزيع الطبيعى القياسى هو التوزيع الطبيعى بمتوسط $\mu = 0$ وتباين $\sigma^2 = 1$.أي أن أي قيمه من قيم المتغير X يمكن تحويلها إلي المتغير Z الذي يتبع التوزيع الطبيعى القياسى بإستخدام التحويله

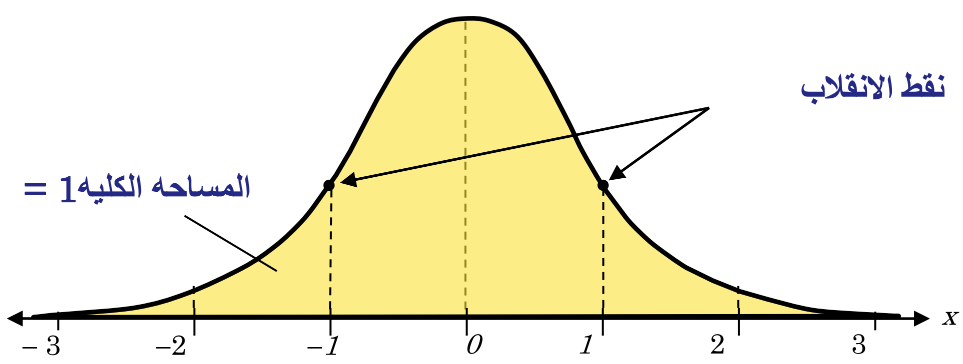
$$Z = \frac{X - \mu}{\sigma}$$

ويمكن رسم المنحنى الإحتمالى للتوزيع الطبيعى القياسى بإستخدام المعادله

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \ -\infty \leq z \leq \infty.$$

والشكل الأتى يوضح المنحنى الإحتمالى للتوزيع الطبيعى القياسى

المنحنى الإحتمالى للتوزيع الطبيعى القياسى



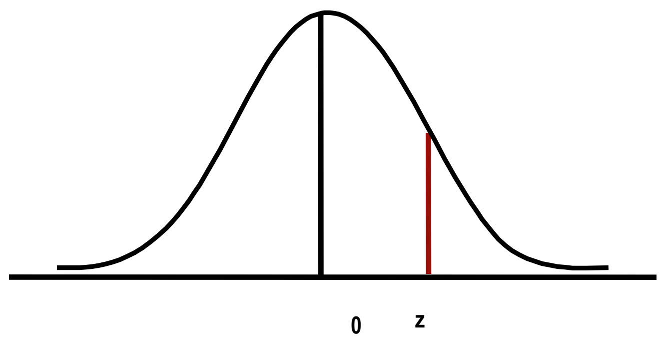
خصائص المنحنى الإحتمالى للتوزيع الطبيعى القياسى

1. طرفا التوزيع تمتد من $-\infty$ إلى ∞
2. متماثل حول العمود الذى يمر بقمته أى عند $X=\mu=0$ لذلك فإن هذا العمود يجزئ المنحنى الطبيعى القياسى إلى قسمين متماثلين فى الشكل والمساحه.
3. يشبه الناقوس من حيث الشكل.
4. المنحنى له نقطتا إنقلاب عند $X=\pm 1$
5. المساحه الكليه تحت المنحنى تساوى الواحد الصحيح .
6. نظرا لتطابق جانبيه وتوسط تفرطحه فإن مقاييس النزعه المركزيه الثلاثه (الوسط الحسابى والوسيط والمنوال) تلتقى فى نقطه واحده (فى منتصف التوزيع) أى أن الوسط الحسابى = الوسيط = المنوال = 0
7. نلاحظ أن 99.7% من قيم المتغير تقع تقريبا فى الفتره (3 ,3-) أى أنه نادرا ما نجد قيمه من قيم X خارج هذه الفتره. كما أن 95.5% من قيم المتغير X تقع فى الفتره (2, 2-) وأن% 68 منها تقع فى الفتره(1,1-)

ويمكن حساب إحتمال وقوع Z فى أى فتره تريدها وذلك بإيجاد قيمه تكامل داله الكثافه الإحتماليه فى هذه الفتره أى بإيجاد المساحه المحصوره تحت منحنى هذه الداله داخل هذه الفتره

ولان متوسط التوزيع الطبيعى القياسى دائما يساوى صفر وتباينه يساوى واحد فقد تم عمل جدول يعطى إحتمالات وقوع المتغير Z فى فتره معينه .

فمثلا يمكن بواسطه الجدول رقم (1) حساب الإحتمال $p(1 \le Z \le 2)$. والجدول يعطى إحتمال وقوع Z بين الصفر وأى قيمه موجبه Z أى يعطى $p(0 \le Z \le z)$ وهى المساحه المظلله فى الشكل الآتى:-



ملاحظات:-

1. $P(-\infty \le Z \le \infty) = 1$
2. $P(Z \le 0) = P(Z \ge 0) = 0.5$

مثال:-

إذا كان Z متغير عشوائيا متصلا يتبع التوزيع الطبيعى القياسى فأوجد

1. $P(Z \le 1.54)$
2. $P(-1.8 \le Z \le 0)$
3. $P(1 \le Z \le 2)$

الحل

1. $P(Z \leq 1.54) = 0.9382$

z	0.00	0.01	0.02	0.03	0.04	0.05
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505

2. $P(-1.8 \leq Z \leq 0)$

$= P(Z \leq 0) - P(Z \leq -1.8)$

$= 0.5 - 0.359$

$= 0.4641$

z	0.00	0.01	0.02
0.0	0.5000	0.5040	0.5080
0.1	0.5398	0.5438	0.5478
0.2	0.5793	0.5832	0.5871
0.3	0.6179	0.6217	0.6255
0.4	0.6554	0.6591	0.6628
0.5	0.6915	0.6950	0.6985

z	0.00	0.01
−3.4	0.0003	0.0003
−3.3	0.0005	0.0005
−3.2	0.0007	0.0007
−3.1	0.0010	0.0009
−3.0	0.0013	0.0013
−2.9	0.0019	0.0018
−2.8	0.0026	0.0025
−2.7	0.0035	0.0034
−2.6	0.0047	0.0045
−2.5	0.0062	0.0060
−2.4	0.0082	0.0080
−2.3	0.0107	0.0104
−2.2	0.0139	0.0136
−2.1	0.0179	0.0174
−2.0	0.0228	0.0222
−1.9	0.0287	0.0281
−1.8	0.0359	0.0351
−1.7	0.0446	0.0436

3. $\mathbf{P(1 \le Z \le 2)}$

$$= P(Z \le 2) - P(Z \le 1)$$

$$P(1 \le Z \le 2) =$$

$$= P(Z \le 2) - P(Z \le 1)$$

$$= 0.1359$$

z	0.00	0.01	0.02	0.03	0.04
0.0	0.5000	0.5040	0.5080	0.5120	0.5160
0.1	0.5398	0.5438	0.5478	0.5517	0.5557
0.2	0.5793	0.5832	0.5871	0.5910	0.5948
0.3	0.6179	0.6217	0.6255	0.6293	0.6331
0.4	0.6554	0.6591	0.6628	0.6664	0.6700
0.5	0.6915	0.6950	0.6985	0.7019	0.7054
0.6	0.7257	0.7291	0.7324	0.7357	0.7389
0.7	0.7580	0.7611	0.7642	0.7673	0.7704
0.8	0.7881	0.7910	0.7939	0.7967	0.7995
0.9	0.8159	0.8186	0.8212	0.8238	0.8264
1.0	0.8413	0.8438	0.8461	0.8485	0.8508
1.1	0.8643	0.8665	0.8686	0.8708	0.8729
1.2	0.8849	0.8869	0.8888	0.8907	0.8925
1.3	0.9032	0.9049	0.9066	0.9082	0.9099
1.4	0.9192	0.9207	0.9222	0.9236	0.9251
1.5	0.9332	0.9345	0.9357	0.9370	0.9382
1.6	0.9452	0.9463	0.9474	0.9484	0.9495
1.7	0.9554	0.9564	0.9573	0.9582	0.9591
1.8	0.9641	0.9649	0.9656	0.9664	0.9671
1.9	0.9713	0.9719	0.9726	0.9732	0.9738
2.0	0.9772	0.9778	0.9783	0.9788	0.9793
2.1	0.9821	0.9826	0.9830	0.9834	0.9838
2.2	0.9861	0.9864	0.9868	0.9871	0.9875
2.3	0.9893	0.9896	0.9898	0.9901	0.9904
2.4	0.9918	0.9920	0.9922	0.9925	0.9927

حساب الاحتمالات في حالة التوزيع الطبيعي العادي

إذا كانت X تتبع توزيعاً طبيعياً متوسطه μ و تباينه σ^2 و أردنا حساب أي احتمال حول المتغير X فإننا نحوله أولاً إلى متغير يتبع التوزيع الطبيعي القياسي

إذا كانت $X \sim N(\mu, \sigma^2)$ فإن $Z = \frac{x-\mu}{\sigma} \sim N(0, 1)$

و على ذلك يمكن حساب الاحتمال $P(a < X < b)$ كما يلي :

$$P(a \le X \le b) = P\left(\frac{a-\mu}{\sigma} \le \frac{x-\mu}{\sigma} \le \frac{b-\mu}{\sigma}\right)$$
$$= P\left(\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma}\right)$$

و هذا الاحتمال الأخير يمكن الحصول عليه من جدول التوزيع الطبيعي القياسي

إذا كانت أطوال مجموعه من النباتات تتبع توزيعاً طبيعياً متوسطه 168 سم و انحرافه المعياري 6 سم. أخذنا عشوائياً أحد النباتات. ما هو احتمال أن يكون طوله:

1. أقل من 159 سم؟
2. أكبر من 180 سم؟
3. واقعاً في الفترة(174 , 165) ؟

الحل

1. أقل من 159 سم

إذا جعلنا X ترمز لأطوال النباتات، فإن X تكون متغيراً عشوائياً يتبع توزيعاً طبيعياً متوسطه 168 سم
وانحرافه المعياري 6 سم.

$$
\begin{aligned}
P(X \le 159) &= P\left(\frac{x-\mu}{\sigma} \le \frac{159-168}{6}\right) \\
&= P(Z \le -1.5) \\
&= 0.0668
\end{aligned}
$$

z	0.00	0.01	0.02
−3.4	0.0003	0.0003	0.0003
−3.3	0.0005	0.0005	0.0005
−3.2	0.0007	0.0007	0.0006
−3.1	0.0010	0.0009	0.0009
−3.0	0.0013	0.0013	0.0013
−2.9	0.0019	0.0018	0.0018
−2.8	0.0026	0.0025	0.0024
−2.7	0.0035	0.0034	0.0033
−2.6	0.0047	0.0045	0.0044
−2.5	0.0062	0.0060	0.0059
−2.4	0.0082	0.0080	0.0078
−2.3	0.0107	0.0104	0.0102
−2.2	0.0139	0.0136	0.0132
−2.1	0.0179	0.0174	0.0170
−2.0	0.0228	0.0222	0.0217
−1.9	0.0287	0.0281	0.0274
−1.8	0.0359	0.0351	0.0344
−1.7	0.0446	0.0436	0.0427
−1.6	0.0548	0.0537	0.0526
−1.5	0.0668	0.0655	0.0643
−1.4	0.0808	0.0793	0.0778

2. أكبر من 180 سم؟

$$P(X \geq 180) = P\left(\frac{x-\mu}{\sigma} \geq \frac{180-168}{6}\right)$$
$$= P(Z \geq 2)$$
$$= 1 - P(Z \leq 2)$$
$$= 1 - 0.9772$$
$$= 0.0228$$

z	0.00	0.01
0.0	0.5000	0.5040
0.1	0.5398	0.5438
0.2	0.5793	0.5832
0.3	0.6179	0.6217
0.4	0.6554	0.6591
0.5	0.6915	0.6950
0.6	0.7257	0.7291
0.7	0.7580	0.7611
0.8	0.7881	0.7910
0.9	0.8159	0.8186
1.0	0.8413	0.8438
1.1	0.8643	0.8665
1.2	0.8849	0.8869
1.3	0.9032	0.9049
1.4	0.9192	0.9207
1.5	0.9332	0.9345
1.6	0.9452	0.9463
1.7	0.9554	0.9564
1.8	0.9641	0.9649
1.9	0.9713	0.9719
2.0	0.9772	0.9778
2.1	0.9821	0.9826
2.2	0.9861	0.9864

3. واقعاً في الفترة(174 , 165)؟

$$P(165 \le X \le 174)$$
$$= P\left(\frac{165-168}{6} \le \frac{x-\mu}{\sigma} \le \frac{174-168}{6}\right)$$
$$= P(-0.5 \le Z \le 1)$$
$$= P(Z \le 1) - P(Z \le -0.5)$$
$$= 0.8413 - 0.3085 = 0.5328$$

z	0.00	0.01
−3.4	0.0003	0.0003
−3.3	0.0005	0.0005
−3.2	0.0007	0.0007
−3.1	0.0010	0.0009
−3.0	0.0013	0.0013
−2.9	0.0019	0.0018
−2.8	0.0026	0.0025
−2.7	0.0035	0.0034
−2.6	0.0047	0.0045
−2.5	0.0062	0.0060
−2.4	0.0082	0.0080
−2.3	0.0107	0.0104
−2.2	0.0139	0.0136
−2.1	0.0179	0.0174
−2.0	0.0228	0.0222
−1.9	0.0287	0.0281
−1.8	0.0359	0.0351
−1.7	0.0446	0.0436
−1.6	0.0548	0.0537
−1.5	0.0668	0.0655
−1.4	0.0808	0.0793
−1.3	0.0968	0.0951
−1.2	0.1151	0.1131
−1.1	0.1357	0.1335
−1.0	0.1587	0.1562
−0.9	0.1841	0.1814
−0.8	0.2119	0.2090
−0.7	0.2420	0.2389
−0.6	0.2743	0.2709
−0.5	0.3085	0.3050
−0.4	0.3446	0.3409

z	0.00	0.01
0.0	0.5000	0.5040
0.1	0.5398	0.5438
0.2	0.5793	0.5832
0.3	0.6179	0.6217
0.4	0.6554	0.6591
0.5	0.6915	0.6950
0.6	0.7257	0.7291
0.7	0.7580	0.7611
0.8	0.7881	0.7910
0.9	0.8159	0.8186
1.0	0.8413	0.8438
1.1	0.8643	0.8665
1.2	0.8849	0.8869

مثال

وجد أستاذ مادة الإحصاء الإجتماعى أن متوسط الوقت الكافى الذى يحتاجه الطلاب لإكمال إمتحانهم النهائى = 150 دقيقه بإنحراف معيارى قدره 30 دقيقه.

أوجدى الآتى:

1. ما إحتمال أن يكمل الطلاب إمتحانهم بين 125و 150 دقيقه.

2. ما إحتمال أن يكمل الطلاب إختبارهم فى 185 دقيقه أو أقل

3. ما إحتمال أن يكمل الطلاب إختبارهم فى وقت يزيد على 195 دقيقه

4. إذا كان عدد الطلاب 1000 طالب . أوجد عدد الطلاب الذين اكملوا امتحانهم في وقت يزيد على 185 دقيقة ؟

<u>الحل</u>

1. إحتمال أن يكمل الطلاب إمتحانهم بين 125و 150 دقيقه

$$P(125 \le X \le 150)$$

$$\begin{aligned}
&P(125 \le X \le 150) \\
&= P\left(\frac{125-150}{30} \le \frac{x-\mu}{\sigma} \le \frac{150-150}{30}\right) \\
&= P(-0.83 \le Z \le 0) \\
&= P(Z \le 0) - P(Z \le -0.83) \\
&= 0.5000 \;-\; 0.2033 \;=\; 0.2967
\end{aligned}$$

وعليه نقرر أن إحتمال إكمال الطلاب لإمتحانهم فى فتره زمنيه تتراوح بين 125 و 150 دقيقه =29.7 %.

2. إحتمال أن يكمل الطلاب إختبارهم فى 185 دقيقه أو أقل

$$\begin{aligned}
&P(X \le 185) \\
&= P\left(\frac{x-\mu}{\sigma} \le \frac{185-150}{30}\right) \\
&= P(Z \le 1.17) \;=\; 0.8790
\end{aligned}$$

وعليه نقرر أن إحتمال إكمال الطلاب لإمتحانهم فى فتره زمنيه 185 دقيقه أو أقل = 87.9 %

3. ما إحتمال أن يكمل الطلاب إختبارهم فى وقت يزيد على 195 دقيقه .

الحل :

أولاً: تحويل المسافه بين المتوسط 150 والدرجه الخام 195 إلى درجه معياريه بإستخدام طريقة تحويل (z).

$$
\begin{aligned}
&P(X \ge 185)\\
&= P\left(\frac{x-\mu}{\sigma} \ge \frac{195-150}{30}\right)\\
&= P(Z \ge 1.5) \;\; = 1 - P(Z \le 1.5)\\
&= 1 - 0.9332 = .0668
\end{aligned}
$$

أى أن إحتمال أن يكمل الطلاب إختيارهم فى وقت يزيد على 195 دقيقه =6.7 %تقريباً.

4. عدد الطلاب الذين اكملوا امتحانهم في وقت يزيد على 185 دقيقة :

$P(X \ge 185) \times$ عدد الطلاب الكلي

$1000 \times 0.0668 = 66.8 \approx 66$ طالب

توزيع ذي الحدين :- The Binomial Distribution.

إذا كانت هناك نجربه عشوائية لها نتيجتان فقط هما ظهور حدث معين نجاح او عدم ظهور فشل مثل نجاح

الطالب او فشلة، المصباح الكهربائي جيد او تالف، وصول طائرة فى موعدها او عدم وصولها، ظهور الصورة

عند القاء قطعة نقود او عدم ظهورها.

ـــ و إذا أجريت هذه التجربة n من المرات

و اذا افترضنا أيضا أن احتمال نجاح هذه التجربة هو p . و احتمال فشلها هو q =1-p حيث p + q = 1

ـــ نفرض أن x هو التغير العشوائي المعرف على هذه التجربة و يرمز الى عدد مرات النجاح لهذه التجربة.

تعطى دالة كثافة الاحتمال للمتغير العشوائي X و التي نرمز لها بالرمز f (x) بالمعادلة

$$p\left(X=x\right)= f(x)=\begin{cases}\binom{n}{x}P^{x}(1-p)^{n-x} & x=0,1,2,....,n\\ 0 & else\ where.\end{cases}$$

حيث n عدد صحيح موجب، $0 \langle p \langle 1$

نحت هذه الشروط واضح أن $f(x) \geq 0$

مفكوك ذي الحدين.

$$(a+b)^n = \sum_{x=o}^{n} \binom{n}{x} b^x \; a^{n-x}$$

ــ أذا كانت f(x) دالة كثافة احتمال للتغير العشوائي x يجب أن تحقق

$$1)\; f(x) \geq o$$

$$2) \sum_{x=0}^{n} f(x) = 1$$

$$1) \binom{n}{x} = \frac{n!}{x!(n-x)!} \rangle\; o \qquad f(x) \geq 0$$

$$x = 0,1,2,\ldots\ldots n$$

وايضا p^x , $q^{n-x} = (1-p)^{n-x}$ موجبتان لجميع قيم x حيث $x = 0,1,2,\ldots., n$ و منها نستنتج أن f(x) موجبة.

2ـ ثانيا

$$\sum_{x=o}^{n} \binom{n}{x} p^x (1-p)^{n-x} = [(1-p) + p]^n = 1$$

إي أن f (x) نحقق شروط دالة كثافة الاحتمال للمتغير العشوائي X المنفصل و يطلق عليها دالة كثافة احتمال

ذى الحديين binomial p.d.f و يقال فى هذة الحالة أن المتغير العشوائي X يتبع توزيع ذو الحديين بعدد مرات n.p و يرمز لذلك بالرمز $X \sim b(n, p)$

و اذا قلنا مثلاً أن في هذه الحالة $X \sim b\left(5, \frac{1}{3}\right)$ تكون دالة كثافة الاحتمال

$$f(x)=\binom{5}{x}\left(\frac{1}{3}\right)^{x}\left(\frac{2}{3}\right)^{5-x} \qquad x=0,1,2,.....,5$$

ثانيا :ـ الدالة المولدة للعزوم لتوزيع ذى الحديين.

$$M(t) = \sum_{x=0}^{n} e^{tx}\ f(x)$$

$$= \sum_{x=0}^{n} e^{tx} \binom{n}{x} P^{x}\ (1-p)^{n-x}$$

$$= \sum_{x=0}^{n} \binom{n}{x} \left(p e^{t}\right)^{x}\ (1-p)^{n-x}$$

لجميع قيم t الحقيقية.

$$M(t) = \left[(1-p) + p e^{t}\right]^{n}$$