جامعة الكوت
مركز البحوث والدراسات والنشر

الإحصاء الحيوي الوصفي
باستعمال البرنامج الاحصائي
(SPSS)

Descriptive Biostatistics
Using Statistical Software
(SPSS)

تأليف
الاستاذ الدكتور
جاسم ناصر حسين
جامعة الكوت

الطبعة الاولى
٢٠٢٥م

منشورات

مركز البحوث والدراسات والنشر

جامعة الكوت

٥١٩ / ٥٣

ح ٥٩٩ حسين، جاسم ناصر.

الاحصاء الحيوي الوصفي باستعمال البرنامج الاحصائي (SPSS)

/ جاسم ناصر حسين.- ط١. بغداد. مطبعة جامعة الكوت، ٢٠٢٥.

٢٦٦ ص ؛ ٢٤ سم

١. الاحصاء الوصفي (برنامج SPSS)   أ. العنوان

رقم الايداع

٢٠٢٥ / ٥١١٤

المكتبة الوطنية/الفهرسة اثناء النشر

رقم الايداع في دار الكتب والوثائق ببغداد

٥١١٤ لسنة ٢٠٢٥ م

الرقم الدولي: ISBN: 978-9922-726-58-8

ملاحظة

مركز البحوث والدراسات والنشر في جامعة الكوت
غير مسؤول عن الافكار والرؤى التي يتضمنها الكتاب
والمسؤول عن ذلك الكاتب او الباحث فقط.

مطبعة جامعة الكوت

University of Kut Printing Press

الشكر والتقدير

الحمد لله والشكر أولا في إتمام هذا الجهد المتواضع، ومن ثم الشكر لسيد الكونين أبي القاسم محمد صل الله عليه وعلى أله ثانيا وبعد كما يقال من شكر المخلوق فقد شكر الخالق.

واجب الشكر والعرفان ان أتوجه بالشكر لكل من: الأستاذ الدكتور باسم شليبه مسلم – قسم الإحصاء كلية الإدارة والاقتصاد – جامعة كربلاء، والأستاذ الدكتور سيف الدين هاشم قمر – كلية الإدارة والاقتصاد – الجامعة العراقية لجهودهما في مراجعة وتقويم الكتاب، وقد كان لملاحظاتهما الأثر الكبير في اظهار الكتاب بصورته النهائية. وكذلك أتوجه بالشكر لكل من الأستاذ الدكتور ثامر ياسر البكري المقوم العلمي والأستاذ الدكتور كريم عبيد هليل الوائلي المقوم اللغوي لجهودهما في تقويم الكتاب.

كما أتوجه بالشكر الى مجلس جامعة الكوت لموافقتهم على طبع الكتاب في مطبعة الجامعة، كذلك أتوجه بالشكر الجزيل لكل العاملين في مطبعة جامعة الكوت لجهودهم في اخراج الكتاب في صورته النهائية.

المؤلف

كانون الثاني/ 2025

بسم الله الرحمن الرحيم

تحتوي المكتبة العربية العديد من المؤلفات العلمية العربية لمختلف التخصصات ومنها تخصص الاحصاء خصوصاً المتعلق بأساليب التحليل الاحصائي للبيانات باستعمال برنامج spss في المجالات الطبية والتي تشترك جميعها بنقاط اساسية من حيث العرض للامثلة وآليات التعامل مع البيانات عند ادخالها في البرنامج لغرض البدء بعملية التحليل ولكن تبقى صفتي الوضوح والبساطة في عرض الخطوات وآليات التحليل باستعمال البرنامج هي المهمة عند القراء والمهتمين خصوصاً من غير المتخصصين في علم الاحصاء عند انجاز تحليلهم الاحصائي للبيانات والحصول على نتائج دقيقة بشكل بسيط دون تعقيد او غموض، ولذا تعد عملية عرض خطوات التعامل في برنامج SPSS فضلاً عن اساليب الاحصاء هدفاً مهماً جداً لاغلب المؤلفين عند تأليف هكذا نوعية من الكتب الاحصائية.

عند مراجعتي وتقييمي لمفردات كتاب الاحصاء الوصفي باستعمال البرنامج الاحصائي "Spss" والمؤلف من قبل الاستاذ الدكتور جاسم ناصر حسين المحترم وما تضمنه من عرضاً علمياً لعملية التحليل الاحصائي لبيانات امثلة تطبيقية تتعلق بظواهر طبية وجدته كتاباً مهماً خصوصاً انه يعرض الخطوات التطبيقية لعملية التحليل الاحصائي (الاحصاء الوصفي) والمتضمن مقاييس النزعة المركزية، مقاييس التشتت، تكوين جداول التوزيع التكراري ... الخ والتي تعد المهمة لدى الباحثين ضمن التخصص الطبي. وايضاً تميز هذا الكتاب بلغته البسيطة الواضحة تدل على مستوى راقٍ في عملية التأليف ليكون مصدراً مهماً في المكتبة الاحصائية العربية.

مبارك للدكتور انجازه العلمي هذا ونشد على يديه الكريمة لانجاز الجزء الثاني من التحليل الاحصائي (الاحصاء الاستدلالي) ليكون مع الاول حلقة متكاملة لعملية التحليل الاحصائي باستعمال برنامج spss.

واخيراً ما كان لله ينمو.

الاستاذ الدكتور

باسم شليبه مسلم العقيلي/[illegible]

كانون الثاني/ ٢٠٢٥

بسم الله الرحمن الرحيم

لطالما شكّل الإحصاء الوصفي حجر الأساس في تحليل البيانات وفهمها، حيث قدمت المؤلفات السابقة في هذا المجال أسساً راسخة للمفاهيم النظرية والتطبيقات العملية في مختلف المجالات. ومع ذلك، تظل هناك حاجة ملحّة إلى مؤلفات تُركز على ربط الإحصاء الوصفي بتطبيقات محددة تُعالج التحديات الواقعية في القطاعات المتخصصة، وتُركز على استخدام البرمجيات المتقدمة لتحليل البيانات، خاصة في المجال الطبي والحيوي.

في هذا الإطار، يأتي كتاب **"الإحصاء الحيوي الوصفي باستعمال البرنامج الإحصائي SPSS"**، للأستاذ الدكتور جاسم ناصر حسين، ليجمع بين أسس الإحصاء الحيوي الوصفي والتطبيق العملي باستعمال برنامج SPSS وهو أحد أشهر الأدوات الإحصائية وأكثرها استخداماً في الأبحاث الطبية، ليشكل إضافة متميزة تُثري المكتبة العلمية. فهو لا يكتفي بتناول المفاهيم الأساسية للإحصاء الوصفي، بل يُعززها بمناقشة معمّقة لتطبيقات التحليل الإحصائي ضمن سياقات طبية متنوعة ودقيقة.

يتفرد هذا الكتاب بتقديم أمثلة واقعية مستوحاة من التطبيقات العملية المرتبطة بالمجال الطبي والصحي تُظهر كيفية توظيف أدوات التحليل الإحصائي باستعمال SPSS لفهم الأنماط وتقييم المخاطر ودعم القرارات السريرية ليسُد فجوة معرفية هامة، مما يجعله مؤلف لا غنى عنه للطلبة والباحثين والعاملين في المجال الطبي والحيوي. كما يُرشد الباحثين إلى كيفية صياغة الاستنتاجات المستندة إلى الأدلة العلمية بطريقة دقيقة ومُحكمة، مما يُسهم في دعم القرارات الطبية وتحسين جودة الأبحاث العلمية.

يُعد هذا الكتاب، بما يحتويه من مادة علمية مُتخصصة وأمثلة تطبيقية من الواقع الطبي مرجعاً أساسياً للطلبة والباحثين والعاملين في المجالات الطبية والصحية. ويُظهر الجهد الكبير الذي بذله المؤلف لتقديم محتوى علمي رصين يُساعد على بناء قدرات القارئ في التعامل مع الإحصاء الوصفي بشكل عملي وفعّال.

إننا على يقين بأن هذا العمل سيُسهم في تعزيز المعرفة وتطوير المهارات التحليلية لدى المختصين في المجال الطبي، ويُعد خطوة هامة نحو تحسين جودة البحث العلمي واتخاذ القرارات السليمة.

مع أطيب التمنيات بالتوفيق للمؤلف...

الأستاذ الدكتور

سيف الدين هاشم قمر

الجامعة العراقية/ كلية الإدارية والاقتصاد

٢٠٢٥/١/١١

جدول المحتويات

المقدمة

الحمد لله رب العالمين على عونه، فمنه نستمد العون وبه نستعين، والصلاة والسلام على نبينا محمد واله الطيبين الطاهرين، اولي المكارم والجود وصحبه الغر الميامين، ومن اتبعهم بإحسان الى يوم الدين. اكتسب علم الإحصاء أهميته من امكانية تطبيق نظرياته ومبادئه وأساليبه في العلوم كلها، اذ بالإمكان التعبير عن الظواهر المختلفة ببيانات يمكن تجميعها بأحد الأساليب الاحصائية. إذ أصبح بالإمكان استعمال الأساليب الاحصائية وتطبيقها في العلوم ومن بينها العلوم الطبية المتنوعة ومن ثم اخذ أسماء بحسب الفروع العلمية كالإحصاء الحيوي والاحصاء الطبي والاحصاء في الصيدلة وغيرها من المسميات، على الرغم من انها تشترك في تطبيق الأساليب الإحصائية في مختلف العلوم.

وبشكل عام فان الاحصاء يتكون من فرعين هما: الاحصاء الوصفي، والاحصاء الاستدلالي، ولكل منهما اساليبه في تحليل ومعالجتها البيانات حيث يعنى الفرع الاول بدراسة اساليب وصف، البيانات والفرع الثاني يعنى بدراسة اساليب استدلال النتائج واستنتاجها من البيانات التي جمعت من العينة واعمام النتائج على المجتمع ويمثل هذا الكتاب الجزء الاول والذي يتضمن الاساليب الوصفية في الاحصاء وتطبيقها في مجال العلوم الطبية، لذلك سمي بالإحصاء الحيوي الوصفي، وان شاء الله سوف يتضمن الجزء الثاني منه اساليب الاستدلال الاحصائي في مجال العلوم الطبية الذي سيكون اسمه الاحصاء الحيوي الاستدلالي مع تطبيق هذه الاساليب باستعمال البرنامج الاحصائي (SPSS).

كان الباحثون قبل انتشار الحاسبات الالكترونية والبرامج الاحصائية يقومون بتحليل البيانات وحساب النتائج يدويا اي باستعمال الحاسبات الصغيرة، ومن ثم كانت كتب الاحصاء مليئة بالمعادلات والاشتقاقات الرياضية لهذه المعادلات فضلا عن الجداول التي تساعد على حل القوانين الإحصائية وفهمها. ولهذا كان الحصول على النتائج في حالة العينات الصغيرة ممكنا، ولكن في حالة العينات الكبيرة يعد استهلاكا للوقت،

فضلا عن امكانية حصول الخطأ تكون كبيرة، اما في الوقت الحالي مع انتشار الحاسبات الالكترونية والبرامج الاحصائية أصبحت مهمة الباحثين في مجال التحليل الاحصائي تنحصر في كيفية التعامل مع هذه البرمجيات، واختيار الانسب منها ومن ثم الحصول على النتائج المطلوبة بحسب هدف الدراسة ومن ثم تفسير هذه النتائج.

وفي الجانب الأخر يواجه طلبة الدراسات الأولية في العلوم الطبية وطلبة الدراسات العليا والباحثون والعاملون في مجال العلوم الطبية والصيدلة الصناعية دائما في حياتهم العملية كميات كبيرة من النتائج والارقام التي تحتاج الى تقويم بعض الكورسات وتفسيرها في مجال الدراسة يحتاجون لاجتيازها، ولتحقيق ذلك فان فهم المفاهيم والطرق الإحصائية ودراستها سوف يحقق هذا الهدف. لهذا فان هدفنا في هذا الكتاب هو تزويد الطلبة والباحثين والقراء بالمفاهيم والطرائق الاحصائية التي يمكن استعمالها في مجال العلوم الطبيبة المختلفة، بطريقة سهلة القراءة ومكثفة وملائمة، دون الدخول في الاشتقاقات الرياضية وتفاصيل النظريات الاحصائية، فضلا عن كمية من التطبيقات الطبية العملية، وباستعمال البرامج الاحصائية.

نعتقد بان هذا الكتاب بفصوله المتنوعة يمكن ان يكون كتابا منهجيا في مجال الدراسات الاولية للعلوم الطبية، وكذلك يمكن ان يكون مرجعا للباحثين في هذا المجال، لأنه سيعزز مفهوم من المفاهيم الإحصائية كلها بأمثله تطبيقية لتوضيح ذلك المفهوم فضلا عن خطوات استعمال البرامج الاحصائية في حل التطبيقات العملية والامثلة التوضيحية ولقد تم استعمال البرنامج الاحصائي الاكثر تطبيقا في مجال العلوم الطبية والاجتماعية من بين البرامج الاحصائية (SPSS) لإمكانية توافره ولسهولة استعماله وحل التطبيقات بواسطة هذا البرنامج لاختصار الوقت والجهد فضلا عن دقة النتائج ووضوح الاشكال البيانية التي تمثل الحالات التطبيقية.

الفصل الأول

مفاهيم عامة في الاحصاء

General Concepts in Statistics

1-1 ما الإحصاء؟ What is Statistics?

تؤكد المعلومات التاريخية ان تاريخ بدء التعامل بالإحصاء يرجع الى الفين سنة قبل الميلاد، حين بدأ المصريون الاوائل بإحصاء السكان والشعوب في ذلك الوقت، وحينما بدء الصينيون بدراسة الارقام الخاصة بمنتوجهم الزراعي والسكان، وحصل التطور الكبير في الاحصاء عند اكتشاف نظرية الاحتمالات مع بداية القرن التاسع عشر التي تعد المساعد الكبير لعلم الاحصاء، ومنذ ذلك الوقت أصبح الاحصاء علما له فوائده في مختلف المجالات العلمية، ومنها مجال العلوم الطبية، ولهذا فان الباحث او الدارس في مجال العلوم الطبية يحتاج الى ان يكون ملما بشكل مبسط بالأساسيات لعلم الاحصاء والحاسوب والعلوم الطبية، كل بحسب اختصاصه، كي يتمكن من انجاز بحثه او دراسة مقرراته الدراسية وانجاز تجاربه العلمية وتفسير النتائج التي يحصل عليها عبر هذه التجارب. لذلك كان الهدف من هذا الكتاب هو تقديم بعض المفاهيم الأساسية لعلم الإحصاء دون الدخول في الاشتقاقات الرياضية، بل يكون التركيز على الجانب التطبيقي باستعمال الحاسوب والبرنامج الاحصائي (SPSS)، لتسهيل مهمة الباحث في كيفية التعرف على الاساليب الاحصائية الملائمة لدراسته، او بحثه، ومن ثم كيفية تحليل بياناته باستعمال البرامج الاحصائية وكيفية تفسير نتائجه التي يحصل عليها. وبوصفه مدخلا لهذه الاساسيات ينبغي الاجابة عن السؤالين الآتيين: ما الاحصاء؟ ولماذا نحتاج الاحصاء في مجال العلوم الطبية؟

تطور مفهوم علم الاحصاء مع تطور نظرياته خلال مدة زمنية تمتد لأكثر من قرنين من الان، فكانت بداية الاحصاء التعامل مع التعدادات والارقام وتطور هذا

المفهوم مع تطور نظريات الإحصاء، وأصبح علما يهتم بجمع البيانات وتنظيمها وتحليلها وتفسير نتائجها من اجل اتخاذ القرارات بشأن الظاهرة محل الدراسة. ومن ثم بحسب هذا المفهوم تمر الطريقة الاحصائية بعدة مراحل اولها تعريف البيانات وتجميعها، والمرحلة الثانية: تنظيم البيانات وترتيبها، ومن ثم وصف هذه البيانات، وهذه المراحل تشكل ما يسمى بالإحصاء الوصفي الذي يمثل كل ما يخص جمع البيانات وتحليلها وتفسيرها، كما أنه يتضمن تمثيل البيانات؛ كحساب متوسط الوزن لمجموعة من المرضى او مجموعة من الممرضات العاملات في احد المستشفيات، أو حساب نسب الطلاق والزواج في أحدى الدول، أو عمل استبانة لتبيّن رأي المجتمع حول انتشار مرض معين، ولهذا يستخدم الإحصاء الوصفي لوصف البيانات وتحويلها إلى أرقام لعرضها بالصورة المناسبة سوآءا كان ذلك باستعمال الخرائط، أم الجداول الإحصائية، أم الرسومات والمنحنيات البيانية التي توضّح الظواهر بشكل أفضل من أي أسلوب آخر، كما يتضمن حساب بعض المؤشرات الإحصائية؛ بوصفها مقاييس النزعة المركزية التي تتضمن، المنوال، والوسط، والوسيط،...وغيرها، ومقاييس التشتت التي تتضمن الانحراف المعياري، والتباين، والمدى،... وغيرها. وبعد هذه المراحل تأتي مرحلة التحليل واستخدام النظريات الاحصائية لاستنباط القرارات المناسبة اعتمادا على البيانات، وهذا ما يسمى بالإحصاء الاستدلالي الذي يُطلق عليه أيضاً الإحصاء التحليلي، وهو يهتم بوضع القرارات المناسبة بناء على النتائج التي تم استنتجت من البيانات، التي جمعت، ويتم ذلك بأسلوبين هما، اولا: التقدير ويعني تقدير معلمات المجتمع المطلوب دراسته اعتمادا على البيانات التي جمعت من العينة المختارة من المجتمع. وثانيهما اختبار الفرضيات: الذي يعني استعمال البيانات التي جمعت من العينة، والمؤشرات الإحصائية، بهدف الوصول إلى قرار نحو الفرضيات التي تبنتها الدراسة، وبناءً عليه تقبل الفرضية أو ترفض. وفيما يخص استعمالات علم الإحصاء فهي كثيرة؛ كاستعماله في العلوم الطبية، وعلم الاجتماع، والاقتصاد، والصناعة، والكيمياء، والرياضة، والإدارة، وغيرها في عدة من المجالات. ويمكن توضيح الحاجة

لعلم الاحصاء عبر استعراض الوظائف التي يقوم بها الاحصاء للعلوم الاخرى، إذ يساعد علم الإحصاء على التخطيط لجمع البيانات وتحديد الطرق المناسبة لجمع البيانات، وماهي البيانات المناسبة عن كل ظاهرة وبدرجة عالية من الكفاءة. كما يساعد على عرض وتمثيل هذه البيانات حول الظاهرة المدروسة بصورة مناسبة وواضحة، كذلك يساهم في تلخيص البيانات حول الظاهرة عبر استعمال المقاييس المناسبة التي تكون ذات معنى، ومن ثم يساهم في تحليل البيانات واختيار الطرق المناسبة للتحليل حسب نوع البيانات حول الظاهرة وكذلك اختبار صحة الفرضيات البحثية التي تتبناها الدراسة في بداية كل بحث علمي، ومن ثم صياغة القرارات المناسبة والتنبؤ بما سيحدث في المستقبل بالنسبة للظاهرة محل الدراسة.

2-1 أهمية علم الإحصاء للعلوم الاخرى Importance of Statistics for other Sciences

اكتسب علم الإحصاء أهميته من امكانية تطبيق نظرياته ومبادئه وأساليبه في العلوم كلها، كما يمكن التعبير عن الظواهر المختلفة ببيانات يمكن تجميعها بأحد الأساليب الاحصائية. فقد أصبح بالإمكان استعمال الأساليب الاحصائية وتطبيقها في العلوم المختلفة ومنها العلوم الطبية المختلفة ومن ثم أخذ أسماء حسب الفروع العلمية كالإحصاء الحيوي، والاحصاء الطبي، والاحصاء في الصيدلة، وغيرها من المسميات، على الرغم من انها تشترك في تطبيق الأساليب الإحصائية نفسها في مختلف العلوم.

لقد أصبح لعلم الإحصاء أهمية بالغة في حياتنا الحديثة، فصارت الإحصاءات مألوفة لدينا وتمثل جانبا مهما من المعلومات التي نطالعها كل يوم، وعلى سبيل المثال: الامراض التي يمكن ان تنتشر وعدد الإصابات في مرض معين، وعدد حالات الشفاء من الإصابات في مرض آخر وتنشر في الصحف، وأيضا المجلات والتقديرات الخاصة بالتنبؤات الجوية ومؤشرات البورصة وانجازات الحكومة في مجال الإسكان، والتعمير والتغيرات التي تطرأ على أسعار العملات وأثمان السلع. وربما يتساءل المرء عن أهمية

الإحصاء بالنسبة لدارس العلوم الطبية معتقداً أن الإحصاء موضوع يدخل في صميم تخصص التجاريين والاقتصاديين فقط، والواقع أن الباحث المتخصص في العلوم الطبية بشكل عام يحتاج في كثير من الأحيان إلى استعمال الأرقام لكي يلخص ويعرض بها مجموعة من المشاهدات التي تتعلق بظاهرة يعنى بدراستها، فقد يطلب منه أن يقدم تقريراً عن مدى التطور الذي حققه برنامج معين لمكافحة مرض معين، وقد يكلف بدراسة الأسباب التي تجعل الذكور أكثر تعرضا للإصابة بمرض معين من الإناث في مجتمع معين.

ففي كل حالة من هذه الحالات يحتاج الباحث أو الدارس في المجال الطبي إلى أداة من الأدوات الإحصائية لكي يستعملها في تلخيص أفكاره والتعبير عنها بصورة محدده ومؤثرة، فالعبارة التي مؤداها " لقد نجحنا في معالجة 90% من مرضى مرض معين بين الأطفال في مجتمع معين" أقوى وأشد من العبارة التي مفادها: " لقد نجحنا في معالجة عدد كبير من مرضى مرض معين من الأطفال في ذلك المجتمع ". ويحتل الإحصاء أهمية خاصة في الأبحاث العلمية الحديثة، إذ لا تخلو أية دراسة أو بحث من دراسة تحليليه إحصائية تتعرض لأصل الظاهرة أو الظواهر المختلفة المدروسة، فتعرض على شكل أرقام ومؤشرات إحصائية تنتهي إلى اظهار اتجاهاتها وعلاقاتها مع الظواهر الأخرى. وفي ضوء ما تقدم يمكن تثبيت بعض الفوائد الكثيرة لدارسي العلوم الطبية وكالاتي:

1. تبسيط البيانات الإحصائية بعرضها في الجداول، أو الرسومات البيانية؛ وذلك لتسهيل فهمها وتحليلها.
2. التعبير عن الحقائق بصورة عددية واضحة ودقيقة بدلاً من عرضها بطريقة السرد الانشائي.
3. مقارنة المجموعات المختلفة وايجاد العلاقات القائمة بينها.
4. يساعد في عملية التخطيط من خلال التنبؤ لبيانات مستقبلية.

5. استخلاص النتائج واتخاذ القرارات المناسبة بقدر كبير من الدقة، وذلك بعد قيام الباحث في اي فرع من فروع العلوم الطبية المختلفة بتحليل البيانات المتوافرة لديه.

ولكن ينبغي أن نشير إلى أن النتائج التي تسفر عن تطبيق أداة إحصائية أو أكثر ليست نتائج قطعية أو غير قابلة للتدقيق والمراجعة. فإذا كانت الأدوات الإحصائية تستطيع أن تساعد الباحث في وصف البيانات وتصميم التجارب واختبار العلاقات بين الظواهر والوقائع الطبية التي يهتم بها فأن ذلك لا يلغي خبرته المهنية. وبعبارة أخرى، يقتصر دور الأدوات الإحصائية على توفير المؤشرات المبدئية التي تساعد الباحث على رفض أو قبول الفروض التي يقوم بدراستها في حدود درجة معينة من الثقة.

3-1 المجتمع والعينة Population and Sample

تتنوع أساليب جمع البيانات والمعلومات تبعا لتنوع اهداف الدارسات الإحصائية ولمختلف المجالات. وبما ان الإحصاء يهتم بدراسة ظواهر تتعلق بمجتمعات، لذلك فان مفهوم المجتمع في علم الإحصاء يشمل جميع الوحدات او المفردات التي تشترك بصفة الدراسة، فعلى سبيل المثال عندما يكون الهدف معرفة تطور خدمة التمريض في مدينة معينة يصبح مجتمع الدراسة جميع الممرضات اللاتي يعملن في تلك المدينة، وعندما يكون الهدف دراسة حوادث السيارات في مدينة معينة يصبح المجتمع جميع الحوادث التي حصلت في تلك المدينة خلال مدة زمنية معينة سنة مثلا. ويمكن تقسيم المجتمعات حسب عدة معايير، فهناك المجتمعات المحدودة والمجتمعات غير المحدودة، اذ يعد المجتمع محدودا إذا كان بالإمكان الوصول الى جميع مفرداته، مثال ذلك مجتمع المرضى في مستشفى معين في حين ان مجتمع الطيور في مدينة معينة يمثل مجتمعا غير محدود، لأنه لا يمكن الوصول الى جميع مفرداته. كذلك يمكن تقسيم المجتمع حسب الحجم الى مجتمعات كبيرة ومجتمعات صغيرة. بالإضافة لهذه العوامل وطبيعة المجتمع تحدد أحيانا الأسلوب المستعمل في جمع البيانات فعلى سبيل المثال

تفرض طبيعة المجتمع أحيانا استعمال أسلوب العينات لان الدراسة الشاملة تدمر مجتمع الدراسة وتؤدي الى خسارة وحدات المجتمع مثال ذلك عند اجراء تحليل الدم لاحد المرضى لأنه لا يمكن استعمال الدراسة الشاملة، بل نعتمد على العينات. والمبحث الاتي يوضح تعريف المعاينة وأنواع العينات في الإحصاء.

4-1 تعريف البيانات Data Definition

البيانات هي مجموعة من الحقائق والأرقام غير المنظمة من مصادر مختلفة، يمكن أن تختلف مصادر البيانات اعتمادًا على ما يحتاجه البحث، ويعتمد تحليل البيانات وتفسيرها فقط على جمع أنواع مختلفة من البيانات من مصادرها. كذلك تعرف البيانات على انها مجموعة من الحقائق الاولية التي تجمع بطرائق متعددة حول ظاهرة ما لدراسة خصائص تلك الظاهرة وتحليل التغييرات التي من المحتمل ان تطرأ عليها، إن اغلب الباحثين يركزون على مشكلة ما لدراستها لجمع البيانات لاحقا عن تلك المشكلة، او ما يطلق عليها الظاهرة المدروسة، وهناك نوعان من البيانات هي (النوعية والكمية) وتختلف هذه البيانات بطرق جمعها وتبويبها.

وبشكـــــل عـــــام فالبيـــانـــات هــــي مجمـــوعـــة مـــن الحــروف أو الكلمات أو الأرقام أو الرموز أو الصور (الخام) المتعلقة بموضوع معين. مثال على ذلك: بيانات الموظفين (الأسماء – والأرقام الوظيفية – والمهن – والصور) بدون ترتيب، وينتج عن هذه البيانات بعد المعالجة ما يطلق عليه مصطلح معلومات. إن نوع البيانات والخصائص للأفراد والمشاهدات، هو العامل الرئيس في توجيه البحث نحو استعمال النوع المناسب من الأساليب الإحصائية في التحليل. وبالتالي تم التعارف بين الاحصائيين على الأنواع التالية من البيانات (مع الاختصار):

1-4-1 البيانات الوصفية (النوعية) Qualitative Data

وهي الصفات والاسماء لمفردات المجتمع التي بدورها تصنف الى نوعين هما:

1- **البيانات الاسمية (Nominal data)**: التي تمثل البيانات التي تأخذ أسماء، مثل النوع: (ذكور- إناث)، المهنة (عمال- موظفين) واللون (احمر-اصفر-...) ويمكن الاستعاضة عن الأسماء بالأرقام فعلى سبيل المثال يعطى للتصنيف (الذكور) رقم (1) و (الاناث) رقم (2) وهنا يعطى لكل صنف رقماً لا يعبر الا عن صنفه ، إذ لا يمكن اجراء العمليات الحسابية الأربع عليه، وهنا لا يمكن ان تجرى عملية الجمع فنقول (1+2) لأنها أرقام معبرة عن اصناف وليست اعدادا، والتعامل مع هذا النوع من المقاييس بالطرق الوصفية عبر الجداول التكرارية والنسب المئوية للتكرارات لكل صنف.

2-**البيانات الرتبية (Ordinal data)**: وهذا النوع افضل من سابقه إحصائيا، إذ ترمز البيانات على وفق رتب لتبين المواقع النسبية للأشياء، ومن ثم يمكن الترتيب من أعلى الى ادنى، او العكس من الأدنى الى الأعلى، مثل الحجم (كبير، متوسط وصغير)، او درجات الطلبة في مقرر دراسي معين (امتياز، جيد جدا، جيد، متوسط، مقبول، راسب) ، أو الوزن لمادة معينة (ثقيل جدا، ثقيل، متوسط ، خفيف، خفيف جدا) اذ ترتب المشاهدات تصاعديا أو تنازليا، ولا تمثل الأرقام فيها كميات بل هي اوزان تمثل التدرج في حالة الصفة، وأغلب استعمالاتها في استمارات الاستبيان عند استعمال التصنيف الثنائي (نعم او كلا) فيكون الوزن 1 يقابل نعم و0 يقابل كلا او عند استعمال التصنيف الخماسي (موافق بشدة، موافق، محايد، غير موافق، وغير موافق بشدة) اذ تعطى الدرجة (5) لـ (موافق بشدة) والدرجة (4) لـ (موافق) ...وهكذا بقية التصنيفات.

1-4-2 البيانات الكمية Quantitative Data

وهي الصفات التي يمكن قياسها او عدها من مفردات المجتمع التي تقسم بدورها الى نوعين هما:

1- **البيانات المتقطعة (Discrete Data)**: وهي البيانات التي نحصل عليها من عملية العد لمختلف الصفات في المجتمع مثل عدد السكان في مدن العراق، او عدد الطلبة في المراحل الدراسية في قسم علمي معين او عدد الممرضات في

مستشفيات مدينة معينة او عدد العيادات الخاصة للأطباء في مدينة معينة او عدد المرضى الذين تماثلوا للشفاء في اثناء جائحة كورنا في أحد المستشفيات وغيرها من الظواهر.

2- **البيانات المستمرة (Continuous Data):** وهي البيانات التي نحصل عليها من عملية القياس لمختلف الصفات لوحدات المجتمع مثل وزن المرضى في أحدى ردهات المستشفى، الوزن للولادات الحديثة في احدى مستشفيات الولادة، او الطول لهذه الولادات او قياسات ضغط الدم العالي او الواطئ لمجموعة من المرضى او قياسات درجة الحرارة لاحد المرضى خلال فترة رقوده في المستشفى وغيرها من الصفات القابلة للقياس.

1-5 مصادر البيانات الإحصائية Resources of Data

تعرف مصادر البيانات بالوحدات او المشاهدات التي تجمع البيانات منها، ويمكن الحصول على البيانات من مصدرين اساسين هما:

1-5-1 مصادر الميدان (المباشرة) Direct Resources

تجمع البيانات بشكل مباشر عن طريق جمع الحقائق والتحري عنها حول دراسة معينة من قبل الباحث نفسه؛ باستعمال أحد الأدوات كالاستبانة، او المقابلة الشخصية، او التجارب المختبرية، او المعملية.

1-5-2 المصادر التاريخية (الرسمية) Historical Resources

تتولّى المؤسسات المختصة المختلفة تسجيل البيانات الإحصائية وجمعها عن الظواهر المختلفة، كل بحسب اختصاصه وتسجيلها في سجلات خاصة لخدمة أغراض تلك المؤسسات؛ فعلى سبيل المثال تقوم المؤسسات الصحية بتسجيل البيانات عن الظواهر الصحية، وتقوم المؤسسات التعليمية بتسجيل البيانات عن الظواهر التعليمية، والمؤسسات الاقتصادية تقوم بتسجيل البيانات عن الظواهر الاقتصادية، وهذه السجلات

تستعمل لأخذ البيانات منها، لذلك سميت بالمصادر التاريخية، لان البيانات سجلت منذ فترات زمنية سابقة للدراسة.

6-1 اساليب جمع البيانات Data Collection Methods

إن استعمال الأساليب الإحصائية في البحوث العلمية، يتطلب بيانات عن الظواهر المدروسة ونحصل عليها باحد اسلوبين هما:

1-6-1 أسلوب التعداد الشامل (Census Method)

يشمل هذا الأسلوب وحدات المجتمع في الدراسة جميعها، وهذا يعني جمع البيانات من المفردات التي تؤلف المجتمع الإحصائي للظاهرة قيد البحث كافة، كالتعدادات السكانية والزراعية وغيرها. ويشترط في هذا الاسلوب ان يكون المجتمع محدداً وان يتوافر الوقت والإمكانات المطلوبة لجمع البيانات، وبشكل عام يعتمد هذا الاسلوب في حالة كون الباحث جاهلاً تماماً بمفردات المجتمع ولا تتوافر لديه اية معلومات سابقة عن المجتمع، وان اهم ما يميز استخدام التعداد الشامل هو، دقة البيانات المستخرجة من المسح نظراً لجمع البيانات من كل فرد من المجتمع وتجنب أخطاء التعميم، التي تنتج من استعمال بيانات مأخوذة عن عينة محددة من المجتمع. ويتم في هذه الطريقة تجنب الأخطاء الشائعة في غيرها من الطرائق وخصوصا خطأ التحيز، أو الصدفة. إما عيوبها فإنها باهظة التكاليف، وتحتا ج إلى إمكانات كبيرة، وتستغرق وقتاً طويلاً، وتبذل فيها جهوداً كبيرة في جمع البيانات وتصنيفها، وكذلك تحتاج إلى فريق عمل ضخم ومدرب للقيام بهذا العمل، كما هو الحال في التعدادات السكانية، لذلك نلجأ الى النوع الثاني.

2-6-1 أسلوب العينات Samples method

يستعمل هذا الأسلوب عندما يمتلك الباحث بعض المعلومات عن المجتمع تساعد على اختيار العينة. ويعتمد هذا الأسلوب إذا كان مجتمع الدراسة غير محدد، أو عدم

إمكان استعمال أسلوب التعداد الشامل لعدم توافر الإمكانات اللازمة له، او عدم توفر الوقت او التكاليف. فعلى سبيل المثال عند دراسة الاستجابة لدواء معين لمرض ما يتم اختيار عدد من المرضى المصابين بذلك المرض، ويعالجون بهذا الدواء والحصول على النتائج التي تضم الاستنتاجات الخاصة بالدواء. ان من مميزات اسلوب العينات إنه يقلل الوقت، والجهد والتكاليف المطلوبة لإنجاز العمل منها. ومن ثم فان المفردات المختارة من المجتمع تسمى بالعينة في حين الاسلوب الذي يستعمل في اختيار هذه العينة يسمى بالمعاينة. ومن ثم فان الأساليب السابقة جميعها تعتمد على نوع الدراسة والتصميم المستعمل في كل دراسة والمبحث الاتي يوضح أنواع العينات.

7-1 أنواع العينات Type of Samples

المعاينة :هي العملية التي تمكننا من اختيار عدد من الافراد للدراسة بطريقة تجعل هؤلاء الافراد يمثلون المجتمع تمثيلا جيدا. وكذلك تعرف المعاينة على انها عملية تستعمل لتسهيل البحث العلمي وتعطي نتائج، وتجيب على معظم أسئلة البحث، وهي اختيار جزء من الكل، والعينة تمثل عددا معينا من أفراد المجتمع الأصلي، ثم اعمام نتائج الدراسة على مجتمع الدراسة كله، ووحدات العينة قد تكون أحياء سكنية، أو شوارع، أو مدنا، أو غير ذلك. وهناك عدد من المبررات لاستعمال العينات في البحوث العلمية منها:

1- اختصار الوقت والجهد والتكاليف.
2- يمكن الحصول على النتائج بسرعة وبسهولة وبصورة كاملة لأن العينة أصغر حجما من المجتمع.
3- يمكن الحصول على بيانات أكثر بواسطة العينة مما نستطيع الحصول عليه من افراد المجتمع كله.
4- هناك بعض الحالات لا يمكن فيها الحصر الشامل للمجتمع، لذا نلجأ الى اسلوب العينة.

5- ليس هناك طريقة لمعرفة مدى الدقة التي تنتج عن الحصر الشامل في حين طريقة العينات فان هناك طرائق لتحديد مدى الدقة للنتائج الحاصلة والثقة لتمثيلها للمجتمع .

وبشكل عام هناك نوعان من العينات حسب طريقة الاختيار إذ يفرض نوع المشكلة وخصائص المجتمع الطريقة المناسبة للاختيار هما.

1-7-1 العينات غير العشوائية Non-Random Samples

وهي العينات التي تعتمد في اختيارها على رغبة الباحث واحكامه الشخصية، ولا يدخل الاحتمال أساسا في عملية الاختيار، ولذا تسمى بالعينات غير الاحتمالية. فقد يختار الباحث عناصر العينة من الذين يقابلهم بشكل عرضي، او لأنه يعرف مسبقا انهم لم يرفضوا طلبة كأن يكونوا من معارفه الذين يتيسر الوصول إليهم ويشار إليهم عادة بالعينة المتيسرة، وقد ينتقي عناصر العينة لأنه يعرف مسبقاً بأنهم الاقدر على تقديم معلومات عن مشكلة معينة أكثر من غيرهم، لأنهم عاشوا المشكلة او عاصروها بمعنى أن عينة من هذا النوع عينة مقصودة. وغالبا ما يستعمل هذا النوع من العينات في حالة المجتمعات غير المحدودة والتي لا يمكن الوصول لجميع مفردات المجتمع وهناك الأنواع الاتية:

1- العينة العمدية أو الغرضية Purposive Sample

يكون الاختيار في هذا النوع من العينات على أساس حر من الباحث، وحسب طبيعة بحثه إذ يحقق هذا الاختيار هدف الدراسة أو أهداف الدراسة المطلوبة مثال ذلك اختيار الطلبة الذين تكون معدلاتهم في الامتحان النهائي جيده جدا فما فوق فقط، لأن هدف الدراسة هو معرفة العوامل التي تؤدي إلى التفوق عند هذا النوع من الطلبة.

2- العينة الحصصية Quota Sample

و هي مهمة في استطلاع الرأي العام لأنها تتم بسرعة أكبر و بتكاليف أقل، و يتم اختيار العينات من الفئات ذات الخصائص المعينة حسب الحجم العددي للجماعات، و الباحث ملزم بتعليمات معطاة مسبقا، و تشابه هذه العملية ما يتم في العينة الطبقية من ضمن العينات العشوائية لكن العينة الطبقية تستعمل العشوائية في اختيار مفردات العينة من الطبقة في حين يترك الاختيار للباحث في العينة الحصصية كي يحصل على الحصة المطلوبة من كل طبقة أو فئة مما يؤدي إلى بعـ ض التحيز في اختيار مفردات العينة من الطبقات حسب رغبة الباحث.

3- عينة كرة الثلج Snowball Sample

تقوم هذه الطريقة على اختيار فرد معين، وبناء على ما يقدمه هذا الفرد من معلومات تهم موضوع دراسة الباحث يقرر الباحث من هو الشخص الثاني الذي سيقوم باختياره لاستكمال المعلومات والمشاهدات المطلوبة، لذلك سميت بعينة الكرة الثلجية إذ يعتبر الفرد الأول النقطة التي سيبدأ حولها التكثيف لاكتمال الكرة، أي اكتمال العينة .

4- العينة العرضية Accidental Sample

يلجأ إلى هذا الصنف من العينة، حين لا يتوافر للباحث أي اختيار لسحب عينة، اذ يعتمد على العناصر التي تقع في يده، إذ يلعب هنا عامل الحظ بالمعنى العامي دورا مهما في الحصول على هذا النوع من العينة، إذ يقوم الباحث بالتقاء أشخاص مارين في طريق معين أو داخلين إلى محل معين لأخذ البيانات منهم بالمقابلة ويظهر هذا النوع من العينة بعض الصعوبات تتعلق أساسا بالتمثيل للمجتمع وبالتالي تعميم النتائج .

5- العينة النمطية

يتم التركيز في هذا الصنف من العينات على بعض الصفات النمطية لمجتمع الدراسة فعلى سبيل المثال اذا كان الهدف من الدراسة هو تصورات الطلبة حول الازمة الاقتصادية فإن هذا يوجه الانتباه الى اختيار عينة الدراسة من طلبة العلوم الاقتصادية، ليكونوا عينة دراستنا، انطلاقا من الاعتقاد بان هؤلاء الطلبة لديهم اهتمام ومعرفة أكثر من غيرهم بالمسائل المتعلقة بالأزمة الاقتصادية.

2-7-1 العينات العشوائية Random Samples

وهي العينات التي يتم اختيارها بطرق علمية محددة ويكون الاحتمال أساسا في عملية الاختيار ولذا تسمى العينات الاحتمالية، مثل العينة العشوائية البسيطة والعينة الطبقية المنتظمة والعينة العنقودية. لذلك فان **العينات الاحتمالية هي تلك العينات التي يكون لكل فرد في المجتمع فرصة محددة لاختياره، او يكون له الفرصة نفسها لاختياره ضمن الفئة الواحدة المتجانسة من الفئات المكونة للمجتمع الاحصائي**، وتشترك الطرائق الاحصائية في اختيار العينات في خطوة اساسية وهي تحديد مجتمع الدراسة واعداد قائمة بعناصره ثم اختيار عينة بحجم يكفي لتمثيل خصائص المجتمع، وأهم هذه الطرائق هي:

1-العينة العشوائية البسيطة Simple Random Sample

وهي من ابسط أنواع العينات وتستعمل عندما يكون مجتمع الدراسة متجانساً، فيما يتعلق بالعوامل التي تؤثر في الدارسة. وهي العينة التي يتم اختيارها بطريقة يكون لكل عنصر في المجتمع فرصة الاختيار نفسها وان اختيار اي عنصر لا يرتبط باختيار اي عنصر اخر. وتستعمل عند توافر شرطين هما اولا أن يكون جميع أفراد مجتمع البحث معروفين، وثانيا أن يكون مجتمع الدراسة متجانسا أي تجانس بين افراده. وتختار العينة العشوائية البسيطة على وفق أحد اسلوبين، هما أسلوب القرعة او جدول الأرقام

العشوائية. ويتلخص أسلوب القرعة في أن يكتب الاعداد على قصاصات من الورق، وتوضع في صندوق ثم تسحب الأوراق عشوائياً الى ان يتم اختيار العدد من الأوراق الذي يساوي حجم العينة. اما طريقة جدول الأرقام العشوائية فتتلخص هذه الطريقة في إعداد قائمة بأفراد المجتمع ويعطي كل فرد رقماً يُعرِّفه لتشكيل ما يسمى بإطار العينة، فاذا كان عدد الأفراد في الإطار مثلا يصل الى 785 فان الارقام في القائمة تتسلسل كما يأتي 1، 2، 3، ....، 784، 785 وبعد ان يحدد الباحث حجم العينة المطلوب يختار افرادها باستخدام اي جدول من جداول الارقام العشوائية التي تتكون عادة باستخدام الحاسب الالكتروني.

2- العينة العشوائية المنتظمة Systematic Random Sample

تستعمل هذه الطريقة في حالة توافر قائمة بأفراد المجتمع، أي إذا كان ترتيب افراد المجتمع في القائمة المشار اليها في العينة العشوائية البسيطة عشوائياً فانه يمكن اختيار عناصر العينة بشكل دوري، إذ يقوم الباحث بتحديد طول الفترة عن طريق قسمة حجم المجتمع على حجم العينة. فعلى سبيل المثال إذا كان حجم المجتمع هو (5000) وان حجم العينة المطلوب هو (400) فان طول الفترة هو (10) أي اننا نختار فرداً واحداً من كل (10) أفراد، على ان يختار الفرد الأول الذي يحمل الرقم من (1-10) وان لا يتجاوز هذا الرقم، فعلى سبيل المثال: إذا اختار الفرد رقم (4) عشوائياً فان الفرد الثاني هو الذي يحمل الرقم (14) والفرد الثالث الذي يحمل الرقم (24) والفرد الرابع يحمل الرقم (34) وهكذا. وتختلف هذه الطريقة عن العينة العشوائية البسيطة في ان اختيار نقطة البداية يحدد رتب الأفراد الباقين بمعنى انها غير مستقلة.

3- العينة العشوائية الطبقية Stratified Random Sample

عندما يكون مجتمع الدراسة غير متجانس بالنسبة لبعض العوامل المؤثرة في الدراسة لا يوفر الاختيار العشوائي عينة ممثلة لخصائص المجتمع اذ لا يوجد ما يضمن ان تكون خصائص المجتمع ممثلة في العينة بالنسب نفسها الواردة في المجتمع أي ان

أفراد مجتمع الدراسة متباينون في الخصائص إذ تشكل كل منهم طبقة مثل (طبقة المعلمين، وطبقة العمّال، وطبقة المهندسين ...). وتقسمهم لطبقات بحسب صفة الدراسة وهدفها ويتم اختيار عينة الدراسة عشوائيا بأخذ نسبة محددة من كل طبقة من هذه الطبقات. وهناك طريقتان لتحديد حصة كل طبقة من حجم العينة هما: اما التوزيع المتساوي بغض النظر عن حجم الطبقة ومن ثم تكون حصة كل طبقة مساوية لباقي الطبقات أي حجم العينة يقسم على عدد الطبقات لمعرفة حصة كل طبقة، او التوزيع المتناسب وهو اخذ حجم كل طبقة بنظر الاعتبار ومن ثم تصبح حصة كل طبقة من العينة تحدد على وفق الصيغة الاتية:

حصة الطبقة من العينة = حجم العينة المطلوب × (حجم الطبقة /حجم المجتمع)

وبعد تحديد حصة كل طبقة من العينة تصبح كل طبقة كأنها مجتمع متجانس مستقل تسحب عينة عشوائية بسيطة منه تمثل حصة العينة من تلك الطبقة، ومن ثم جمع هذه العينات لتشكل العينة الكلية من المجتمع وبالتالي نضمن تمثيل طبقات المجتمع كلها في العينة والدراسة.

4-العينة العشوائية متعددة المراحل(العنقودية) Cluster Random Sample

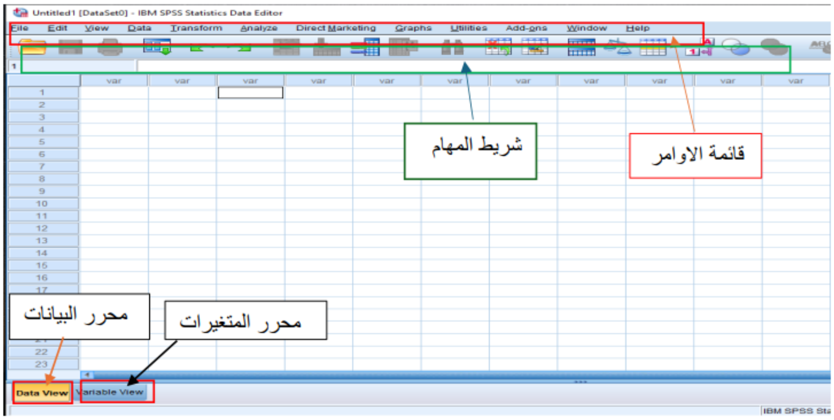
العينة العنقودية هي النوع الرابع من أنواع العينات العشوائية، و فيها يلجأ الباحث إلى تحديد العينة أو اختيارها ضمن مراحل عدة، ففي المرحلة الأولى يقسم مجتمع الدراسة الأصلي إلى مجتمعات اصغر أو مجموعات بحسب معيار معين، ومن ثم اختيار مجموعة أو أكثر بطريقة عشوائية، و بالنسبة للمجموعات التي لم يخترها في هذه المرحلة فانه يستبعدها من العينة نهائيا، وفي المرحلة الثانية تقسم المجموعات التي وقع عليها الاختيار في المرحلة السابقة إلى مجموعات جزئية أخرى ثم نختار مجموعة جزئية أو أكثر منها بطريقة عشوائية أيضا، وهكذا يستمر الباحث حتى يتم الوصول إلى المجتمع النهائي الذي يمثل المجموعة الجزئية النهائية التي تمثل مجتمع الدراسة، يقوم الباحث باختيار منها و بشكل عشوائي عدد مفردات العينة المطلوبة،

وعليه فان تحديد مفردات العينة النهائية يتم على مراحل. مثال ذلك دراسة انتشار مرض الانفلونزا في العراق أولاً تقسيم العراق الى محافظات، وتسحب عينة من المحافظات حسب توافر الإمكانات المادية والبشرية للباحث، ثانياً تقسيم المحافظات التي اختيرت في المرحلة الأولى الى اقضية، واختيار عينة من الاقضية، وفي المرحلة الثالثة تقسم الاقضية الى وحدات إدارية أصغر واختيار العينة النهائية من هذه الوحدات. ومن ثم فان اية طريقة استعملت في جمع البيانات فان الحصيلة ستكون مجموعة من الحقائق تسمى البيانات.

8-1 نبذة عن البرنامج (SPSS)

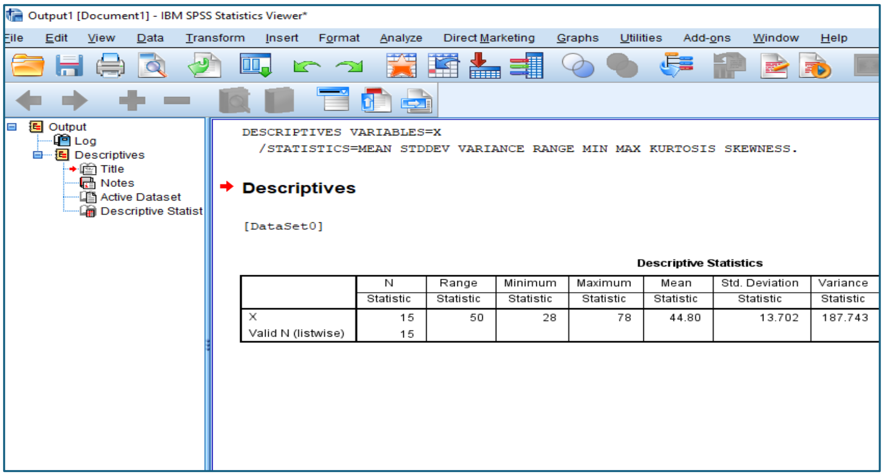
يعد البرنامج (SPSS) هو اختصاراُ للاسم الكامل Statistical Package for) the Social Sciences) الذي يعني (الحزمة الاحصائية للعلوم الاجتماعية) وهو برنامج حاسوب متخصص في التحليل الإحصائي وإدارة البيانات. وقد طورته شركة IBMويستخدم على نطاق واسع في البحوث الاجتماعية والعلوم السلوكية والعلوم الطبية والتجارة والتسويق ومجالات أخرى. وهو أحد البرامج الرائدة في مجال التحليل الإحصائي ومعالجة البيانات، ويعد البرنامج SPSS أداة قوية وشاملة تستعمل في البحوث الاجتماعية والعلوم الطبية والإدارة وعدد من العلوم الاخرى. يتميز برنامج SPSSبواجهة استخدام سهلة تسمح للمستخدمين بإجراء تحليلات إحصائية متقدمة دون الحاجة إلى خبرة مسبقة في البرمجة. كما يوفر البرنامج عدد من الأدوات والوظائف لتنظيم البيانات، وتحليلها، واظهار النتائج.

نفترض بان القارئ لديه فكرة وخبرة بسيطة في تشغيل البرنامج، ومن لم تتوافر لديه هذه الخبرة يمكن الرجوع الى المصادر الخاصة بالبرنامج او مشاهدة الفيديوهات الخاصة بكيفية تشغيل البرنامج عبر الانترنت. وبشكل مختصر يشتمل البرنامج عند تشغيله في البداية على صفحتين، تظهر في البداية عند التشغيل صفحة محرر البيانات (Data view) ويمكن الانتقال الى الصفحة الثانية التي تمثل محرر المتغيرات (Variable view) وعبر النقر على المربع في أسفل صفحة محرر البيانات. كما موضح في الشكل (1-1).



الشكل (1-1) محرر البيانات في برنامج (SPSS)

ويعرض البرنامج النتائج في صفحة ثالثة تسمى محرر النتائج (output viewer) وهذه الصفحة لا تظهر عند التشغيل، ولكنها تظهر بعد اجراء اي عملية لتحليل البيانات، او عملية أخرى ويكون شكلها كما في الشكل (1−2)



الشكل (1−2) صفحة النتائج بعد اجراء عملية حسابية

يعد برنامج (SPSS) أحد البرامج التي تعمل تحت بيئة نظام التشغيل ويندوز، لذلك يتضمن مجموعة من قوائم الاوامر التي من خلالها يمكن القيام بعمليات التحليل الاحصائي المطلوبة في البرنامج جميعها، وتكون هذه القوائم عادة في أعلى صفحة

محرر البيانات كما في الشكل (1−1). كذلك يوجد شريط أسفل قوائم الأوامر يسمى شريط المهام لإنجاز مهام محددة كما في الشكل (1−1) وسوف يستعمل هذا البرنامج في الأماكن التي تحتاج الى تطبيق مثل هذه البرامج وعادة ما تكون الامثلة او التمارين التي تستعمل لتوضيح القواعد العامة في احجام عينات صغيرة لذلك يكون استعمال الحاسبة او البرامج الاحصائية فقط للتطبيق ويبقى الهدف تعليمياً على كيفية استعمال هذا البرنامج في حالة توفر بيانات بأحجام كبيرة تحتاج لمثل هذه البرامج.

تمارين الفصل الأول

1− لماذا نحتاج الاحصاء؟

2− حدد مفهوم الاحصاء في العلوم الطبية.

3− ماهي اهم المجالات التي يطبق فيها الاحصاء في العلوم الطبية؟

4− عرف المصطلحات الاتية: أ− المجتمع ب− العينة ت− العشوائية ج− البيانات.

5− وضح ماذا نقصد بالإحصاء الوصفي؟ وماذا نقصد بالإحصاء الاستدلالي؟ وما هو الفرق بينهما؟

6− وضح الفرق بين: أ− البيانات والمعلومات ب− المجتمع والعينة ت− العينات العشوائية وغير العشوائية ج− البيانات الوصفية والبيانات الكمية.

7− طلب منك اختيار عينة من بين 600 موظف في المستشفى 200 طبيب، و150 ممرضين، 50 مشرفين على المستشفى، 200 عامل في مختلف المجالات الأخرى. ما هو نوع العينة المناسب وكيف يتم سحب هذه العينة؟

8− عدد أنواع العينات، مع تحديد طبيعة المجتمع التي تناسب كل عينة.

9− لماذا تستعمل العينات بدلا عن التعداد الشامل في جمع البيانات لمختلف البحوث والدراسات؟ ومن هي الأكثر دقة؟

الفصل الثاني

تصميم الدراسة في العلوم الطبية

Study Design in Medical Sciences

1-2 تمهيد

يعد تصميم الدراسة في العلوم الطبية أكثر اهمية من تحليل البيانات، لان التحليل غير الجيد للبيانات يمكن اعادته، ولكن التصميم غير الجيد للدراسة تصعب اعادته، ومن هنا يعد تصميم الدراسة في تجارب العلوم الطبية المفتاح لجمع البيانات وتحليلها. والدراسات بشكل عام يمكن ان تصنف الى: دراسات تعتمد على المشاهدة، ودراسات تعتمد على التجربة. إن دراسات المشاهدة هي الدراسات التي يقوم الباحثون فيها بمشاهدة وحدات الدراسة وتصنيفها حسب واحد او أكثر من مستويات المتغيرات المستقلة و/او مستويات المتغير المعتمد، اما الدراسات التجريبية فهي الدراسات التي يتدخل فيها الباحث عبر اعطاء وحدات الدراسة معالجة معينة، وتتابع هذه الوحدات خلال فترة من الزمن وتسجيل الاستجابة المطلوبة. ومن ثم هناك العديد من الطرق التي تستعمل في تصميم الدراسات الحيوية في كلا النوعين، ومنها:

2-2 الدراسات التي تعتمد على المشاهدة Observational Studies

بحسب هذا النوع من الدراسة على الباحث ان يشاهد وحدات الدراسة كما تحدث في الطبيعية، ويصنفها بحسب مستويات المتغيرات المعتمدة او المتغيرات المستقلة. وهذه الدراسات تكون على نوعين: اما دراسات مستقبلية (prospective Studies) وفيها تحدد مستويات المتغيرات المستقلة لوحدات الدراسة اولا ثم مشاهدة المخرجات او الاستجابات المطلوبة لهذه الوحدات. او دراسات استرجاعية (Studies Retrospective) وهي الدراسات التي تعتمد او تنظر للماضي وفيها تحدد مستويات متغيرات الاستجابة أولا، وبعد ذلك تجمع المعلومات عن المتغيرات المستقلة. فعلى

سبيل المثال اذا رغبنا في دراسة العلاقة بين التدخين ومرض سرطان الرئة، فأما يصنف الاشخاص الى مدخنين وغير مدخنين (المتغير التوضيحي) ومتابعتهم عبر مدة من الزمن، لمعرفة فيما اصيبوا بمرض سرطان الرئة ام لم تحدث الاصابة (المتغير المعتمد)، وهذا ما يسمى بالدراسات المستقبلية. أو أخذ مجموعة من المرضى ويصنفون الى مصابين بالسرطان وغير مصابين (وهذا يمثل المتغير المعتمد) ونرجع الى تاريخهم الطبي لمعرفة ما إذا كانوا مدخنين او غير مدخنين (المتغير التوضيحي) وهذا ما يعرف بالدراسات التاريخية او الاسترجاعية، وبشكل عام فان دراسات المشاهدة تتكون من ثلاثة انواع هي:

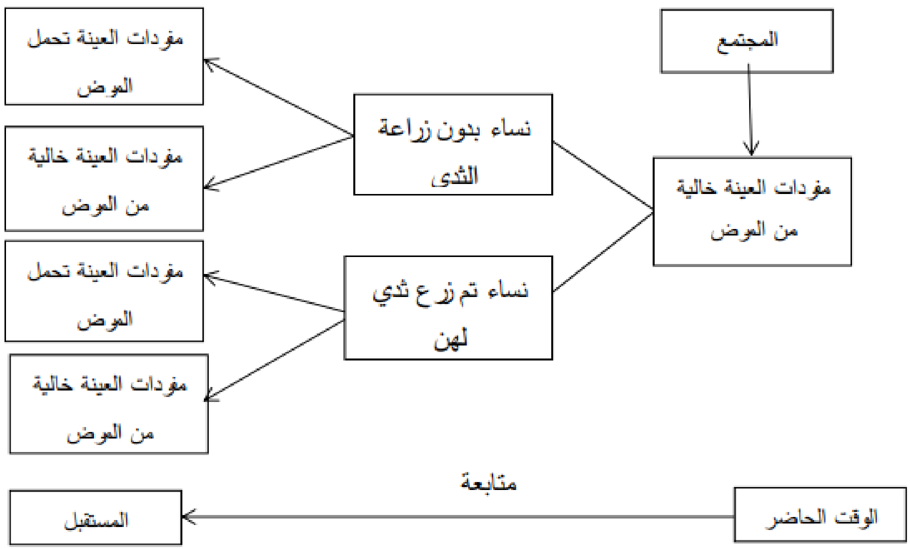
1-2-2 تصميم الدراسات المقطعية Cross-Sectional studies design

يستعمل هذا التصميم في دراسة العلاقة بين المرض والعوامل المسببة له عند فترة معينة من الزمن في المجتمع، ويتضمن دراسة عينة من المجتمع مختارة عشوائيا، ومن ثم تحديد مستويات المتغيرات التوضيحية ومستويات متغيرات الاستجابة. وهذه الدراسات تجرى تاريخيا بالاعتماد على قاعدة بيانات في المنظمات الصحية، او البلدان او على المستوى العالمي. وفي مثل هذه الحالات هناك عدد من المفردات، ولكل مفردة تاريخا طبيا، إذ تصنف المفردات وتدرس العلاقات بين المتغيرات. فعلى سبيل المثال إذا كان الهدف من الدراسة معرفة العلاقة بين البدانة (السمنة) والعوامل الديموغرافية المسببة لها لعينة من الناس مثل العمر، والنوع، والوزن، والطول، وضغط الدم، ومستوى الكولسترول في الدم وغيرها من العوامل. فالبيانات التي تجمع على وفق هذا الخليط من العوامل تتضمن بيانات مستمرة وبيانات متقطعة مرتبة وبيانات اسمية. وفي تحليل البيانات في مثل هذه الدراسات ربما يكون هدف الباحث وصف الخصائص لعينة الدراسة عبر المقاييس الوصفية وحسب نوع البيانات، او هدفه تحديد العلاقات بين هذه الخصائص في مدة من الزمن، او يكون هدفه بناء نموذج يضم هذه الخصائص للتنبؤ المستقبلي بأحدها عب توافر البيانات عن الخصائص الاخرى.

2-2-2 تصميم دراسات المتابعة Cohort Studies Design

يعد هذا النوع من الدراسات أحد انواع الدراسات المستقبلية (Prospective Studies)، وفيه تحدد المفردات اعتمادا على مستويات المتغيرات التوضيحية ومتابعة مفردات الدراسة (العينة) من نقطة محددة في الزمن الحاضر، وعبر فترة من الزمن وصولا الى المستقبل للحصول على قياسات الاستجابة. فعلى سبيل المثال هناك عدد من الدراسات التي اعدت للمقارنة بين تطور مرض السرطان عند النساء بسبب زراعة الثدي، أذ تصنف النساء الى صنفين (نساء زرع الثدي لهن ونساء بدون زراعة الثدي) وهذا يمثل مستويات المتغير التوضيحي وتتابع حالتهن مدة من الزمن لملاحظة ما إذا ظهر لديهن سرطان الثدي، وهذا يمثل متغير الاستجابة (المعتمد) وكما موضح في الشكل رقم (2-1). هذا النوع من الدراسات يكون شائعا ويكون احيانا غير اخلاقيا عند تخصيص عامل معين مثل التدخين او زراعة الثدي لمفردات العينة، ولكن احيانا نجد مجتمعاً يتضمن مثل هذه المفردات.

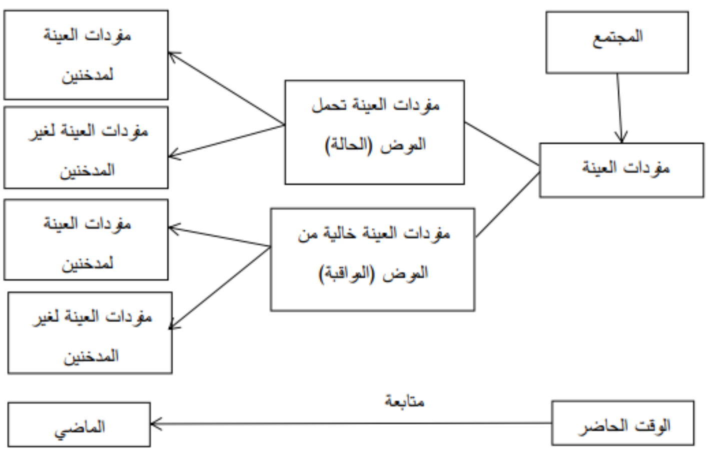
عادة هذا النوع من الدراسة يحتاج الى عينات بحجم كبير عندما يكون عامل الاستجابة في المجتمع نادراً.



الشكل (2-1) خطوات دراسة المتابعة

3-2-2 تصميم دراسة الحالة Case- Control study design

يعد هذا النوع من الدراسات أحد انواع الدراسات التاريخية أذ تحدد مفردات العينة اعتمادا على مستويات متغيرات الاستجابة، ومن ثم قياس مستويات المتغيرات التوضيحية (وغالبا ما تكون قد تعرضت مفردات المجتمع لمستويات من الخطر). وتدرس عينة من المجتمع غالبا ما يحمل افرادها مرضاً معيناً عند نقطة محددة من الزمن تسمى (عينة الحالة) وتقارن مع افراد عينة اخرى لا يحملون المرض (عينة المراقبة). ويستعمل هذا التصميم في اغلب الاحيان عندما يكون متغير الاستجابة نادرا جدا في المجتمع. قد يكون هذا النوع من الدراسات عرضة للتحيز بسبب ان مفردات عينة الحالة تتذكر عوامل الخطر التي سببت لها المرض مثلا، وعلى العكس من مفردات عينة المراقبة لا يتذكرون مثل هذه العوامل. فعلى سبيل المثال عينة من الاطفال الذين ولدوا مع اعاقة معينة، تمثل عينة الحالة وعينة من الاطفال الذين ولدوا بدون اعاقة معينة تمثل عينة المراقبة، فإن امهات الاطفال المعاقين ربما يتذكرن بأنهن استخدمن بعض الادوية في اثناء الحمل مما سبب الاعاقة، ولكن امهات الاطفال بدون الاعاقة لا يتذكرن ذلك. وبشكل عام يمكن القول إن هذا النوع من الدراسات هو الاضعف في تحديد العلاقة السببية بين المرض وعناصر الخطر التي تسبب هذا المرض الذي تعرض لها افراد المجتمع. وهذا التصميم يوفر أرخص وأسرع طريقة لدراسة عناصر الخطر للأمراض ويمكن توضيح هذا التصميم بالشكل رقم (2-2).



الشكل (2-2) خطوات تصميم دراسة الحالة

2-3 تصاميم الدراسات التجريبية Experimental Studies designs

هي الدراسات التي يتدخل الباحثون في اجرائها عبر اقامة بعض التجارب، او تخصيص بعض المعالجات لمفردات الدراسة لمعالجة بعض الامراض او الظواهر المطلوب دراستها، وتتضمن انواع عدة من التجارب اهمها ما يأتي:

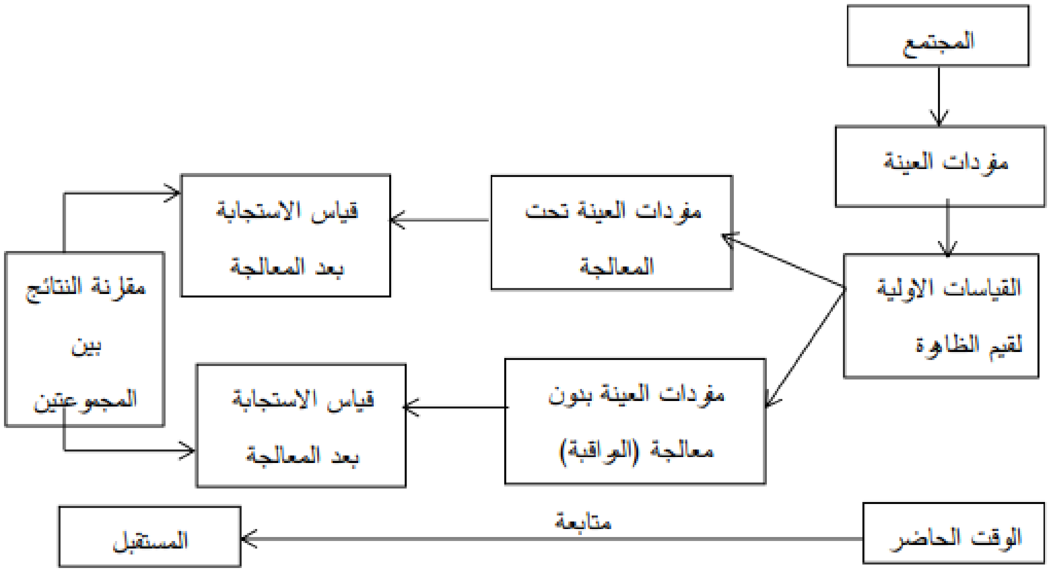
2-3-1 التجارب السريرية العشوائية Randomized Clinical Trials

وهي تجارب تحت السيطرة حيث تختار بعض المفردات من المجتمع التي تطابق بعض المعايير الفيزيائية، وتصنف عشوائيا في المجاميع، ومن ثم فان مستويات المتغيرات التوضيحية تخصص عشوائيا. هذه التجارب اصبحت مشهورة في اواخر الاربعينات من القرن الماضي، وحاليا تعد القاعدة الذهبية للتجارب الطبية من حيث تحديد السبب والتأثير للمعالجة. يعتبر الانكليزي السير برادفورد الرائد في بذل الجهود للبدء باستخدام هذا النوع من التجارب (Winner, 2004)، وعندما لا يكون مهماً تخصيص المريض لاي معالجة فان التجربة تعرف بالتجربة العمياء (Blind trials) وعندما لا يكون مهماً للباحث الطبي والمريض ان تعمل التجربة بأية معالجة، فأنها تسمى بالتجربة العمياء المزدوجة (Double-Blind trials). ويمكن اجراء هذه التجارب بإحدى طريقتين:

أ- دراسات المجاميع المتوازية Parallel Groups Studies

ويسمى هذا التصميم بالمتوازي لان التغيرات بين مفردات الدراسة تحدد عبر الزمن. وهي دراسات وتجارب مستقبلية التي يتاح فيها الحصول على البيانات الاولية في بيئة طبية مسيطر عليها. إذ تخصص المفردات المتشابهة في القياسات الأولية بطريقة عشوائية الى مجموعتين أو أكثر من مجاميع المعالجة، وبعد مدة متابعة كافية يتم مقارنة المخرجات بين المجموعات. وبمعنى آخر ان المفردات تخصص بطريقة عشوائية لواحدة من مجموعتين أو أكثر، من ثم تعد واحدة من المجاميع بوصفها مجموعة مراقبة. لذا فان مفردات مجموعة المعالجة تختلف كليا عن المفردات في

مجموعة المراقبة. فعلى سبيل المثال نرغب في دراسة كفاءة دواء جديد في معالجة مستوى الكولسترول في الدم، نختار مجموعة من الاشخاص ونسجل المستويات الاولية للكولسترول عندهم، ثم بعد ذلك نوزعهم عشوائيا في مجموعتين، إذ ان المجموعة الاولى تمثل مجموعة المعالجة بالدواء الجديد والمجموعة الثانية تمثل مجموعة المراقبة كما مبين في الشكل رقم (2-3).



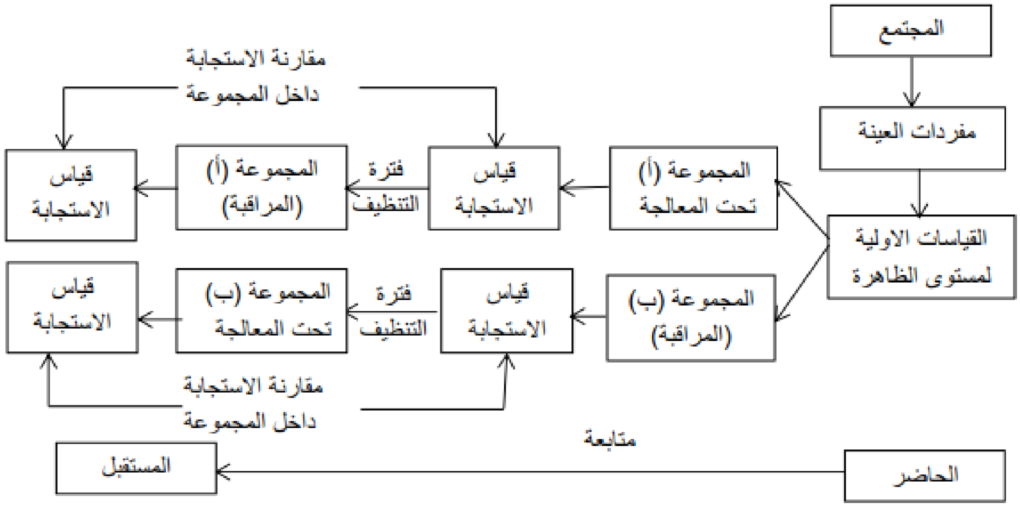
الشكل (2-3) تصميم دراسة المجاميع المتوازية

ويقاس مستوى الكولسترول بعد مدة من الزمن، على سبيل المثال أربعة أسابيع، ومن ثم مقارنة النتائج بين المجموعتين لمعرفة كفاءة الدواء الجديد. ولتقويم كفاءة الدواء هناك عدة طرق يمكن استخدامها مثل حساب الفروقات بين القيم الاولية للكولسترول قبل بدء التجربة والقيم النهائية للكولسترول بعد نهاية التجربة، ومقارنة متوسطات الفروق في المجموعتين. وعلى الرغم من اننا نتوقع انه من الطبيعي حصول انخفاض في مستوى الكولسترول في مفردات مجموعة المعالجة، ولكن هذا قد يحصل في

مجموعة المراقبة أيضا، ويعزى هذا الى التأثير الوهمي في المجموعة. وإن حدث مثل ذلك فان المقارنة بين كمية التغير في المجموعتين خلال الفترة الزمنية غير ذي معنى.

ب- **دراسات المجاميع المتقاطعة (Cross-over Groups Studies):**

وهي دراسات تتكون من ثلاث مراحل تخصص فيها المفردات بطريقة عشوائية لواحدة من مجموعتين مثل (أ و ب)، وبعد تسجيل البيانات الاولية لعناصر العينة كلها حول موضوع الدراسة، وتختبر على نوع واحد من المعالجات او نوعين، وتوضع في المرحلة الاولى من التجربة مفردات المجموعة (أ) تحت المعالجة ومفردات المجموعة (ب) تبقى بدون معالجة ومن ثم تعد مجموعة مراقبة، وبعد مدة معالجة كافية يتم قياس الاستجابة بعد المعالجة، بعدها تبدأ المرحلة الثانية، وتسمى مرحلة التنظيف وذلك لتنظيف مفردات المجموعتين من تأثير المعالجة. ومن ثم تبدأ المرحلة الثالثة حيث تغير المعالجة بين المجموعتين فتبقى عناصر المجموعة (أ) بدون معالجة وتعد مجموعة مراقبة، وتوضع عناصر المجموعة (ب) تحت المعالجة ولنفس المدة وفي نهاية المدة الثالثة تقاس الاستجابة مرة اخرى في المجموعتين، وفي نهاية التجربة تقارن قياسات الاستجابة في نهاية المرحلة الاولى مع قياسات الاستجابة في نهاية المرحلة الثالثة، ولكل مجموعة على حده وكما موضح في الشكل رقم (4.2). وبهذه الطريقة فان كل مفردة من مفردات العينة أخذت كل المعالجات، ومن ثم أصبحت كل مفردة هي مفردة مراقبة على نفسها، ولا يمكن استعمال هذا النوع من الدراسات في الحالات التي تترك فيها المعالجات اثار دائمة ومن ثم فان عملية التنظيف غير ممكنة.



الشكل (2-4) مراحل دراسة المجاميع المتقاطعة

2-3-2 تصاميم الدراسات المتتابعة Sequential Studies Designs

المشكلة الاخلاقية التي تواجه القائمين على التجارب الطبية والمتمثلة بتعريض مفردات الدراسة الى جرعات من الأدوية أو أنواع من المعالجات ولفترات طويلة او محددة التي قد تسبب لهم امراضاً او تنهي حياتهم، دعت الباحثين الى استعمال تصاميم الدراسات المتتابعة التي تسمح للباحثين بإيقاف الدراسة مبكرا، وعند مشاهدة اول تأثير واضح للمعالجة دون التأثير في الخصائص الاحصائية للتجربة. في مثل هذه التجارب في حال مشاهدة اي مستوى من النجاح لمعالجة معينة دون المعالجات الأخرى توقف التجربة واعطاء مفردات العينة علاجات متفوقة لتقليل تأثير العلاجات الأولى، ونلاحظ مؤخرا بان هناك عدداً من المعايير التي تستعمل لتحديد الفروق بين المجموعات بضمنها احجام العينات الثابتة وأخطار الاخطاء في الاستنتاج في التجارب، ولكن تصاميم المتتابعة يمكن اجراؤها عند مستويات مقبولة من المعنوية، ولا توجد خطة مسبقة لتحديد حجم العينة عندما يظهر أول تأثير للمعالجة.

3-3-2 تصاميم الدراسات المتكررة Repeated Studies Designs

تركز تصاميم الدراسات السابقة على مقارنة تأثير المعالجة مع مجموعة المراقبة. ولكن احيانا توجد بعض الحالات التي تكون فيها مفردات مجتمع الدراسة متشابهة الى درجة كبيرة. ولدراسة تأثير المعالجة في مثل هذه الحالات يكون من المقبول استعمال تجارب القياسات المتكررة على المفردات في حال اختيارها ويتضمن هذا النوع من الدراسات الطرائق الاتية:

أ- ابسط انواع هذه الدراسات هي طريقة العينات المزدوجة، وهي من التجارب المستقبلية وفيها نختار المفردات عشوائياً وخصوصاً التي تمتلك المعايير التي تؤهلها الى الدخول في الدراسة، وبعد تسجيل البيانات الاولية تخضع للمعالجة، وبعد مدة علاج كافية تقارن المخرجات مع البيانات الاولية. وهذا ما يعرف أيضا بطريقة اختبار قبل وبعد (Pre-Post test).

ب- تجارب القياسات المتكررة في أكثر من نقطتين من الزمن، وفي مثل هذه الدراسات تكرر القياسات لأكثر من مرتين ضمن مدة زمنية محددة. فعلى سبيل المثال في تجارب فحص سكر الدم، احيانا نرغب في تقدير الزمن عندما يكون مستوى السكر في الدم في القمة، او تقدير الوقت اللازم كي ينخفض مستوى سكر الدم الى القيم الاولية الطبيعية، وأحيانا نرغب في رسم مستويات تركيز سكر الدم عبر مدة من الزمن للتعرف على الزمن الذي يكون فيه السكر في القمة، وأحيانا يستعمل هذا المنحنى للمقارنة بين أكثر من مجموعة او أكثر من شخص.

ت- تجارب تحليل التباين المتكرر (ANOVA) وهي تجارب تجرى فيها القياسات المتكررة في أكثر من نقطتين من الزمن ولأنواع مختلفة من المعالجات. فعلى سبيل المثال يرغب باحث في مقارنة تأثير مكملات الكالسيوم والحليب في كثافة نخاع العظم عند النساء اللواتي في سن اليأس، لذلك نختار النساء في سن اليأس اللاتي يحققن الشروط المحددة مسبقا للدخول في الدراسة يقسمن عشوائيا الى مجموعتين. تعطى للمجموعة الاولى مكملات الكالسيوم بينما تعطى للمجموعة

الثانية الحليب، وتسجل قياسات كثافة نخاع العظم كل ستة أشهر ولمدة عامين في المجموعتين، وبعد فحص التغيرات التي تحصل عبر هذه المدة في هذا التصميم والمقارنة بين المعالجتين يمكن ان تتحقق كل ستة أشهر أو في نهاية فترة الدراسة، ويمكن ان توضع المقارنات مع الاخذ بنظر الاعتبار متغيرات اخرى كالعمر والوزن وغيرها، وهذا من جانب ومن جانب اخر طالما الدراسة تتطلب مدة طويلة لذلك نتوقع خروج بعض المفردات عن الدراسة. وعادة ما تكون حصيلة هذه التجارب مجموعة من الحقائق او البيانات لمجموعة من الصفات، وتسمى هذه الصفات بالمتغيرات العشوائية (Random Variables) لأنها متغيرة حسب طريقة جمعها من مجموعة من المشاهدات والمبحث الاتي يوضح أنواع المتغيرات العشوائية.

2-4 المتغير العشوائي Random Variable

يختلف مفهوم المتغير العشوائي وهذا الاختلاف يعود الى اختلاف وجهة نظر العلوم المختلفة الى المتغير العشوائي فعلى سبيل المثال، من وجهة نظر الرياضيات، وبالتحديد في الاحتمالات والإحصاء، **المتغير العشوائي** (Random Variable) متغير ذو قيمة متغيرة طبقًا للصدفة (أي أنه يحقق مفهوم العشوائية)، فلا يكون ثابتا على قيمة معينة محددة، بل يساوي قيمة من القيم الممكنة المختلفة التي يكون لكل واحدة منهن احتمال ما. لذلك فالمتغير العشوائي هو دالة رياضية تعين لكلِ مفردة من فَضَاء العينة Ω، عددا مناسبا من فضاء الأعداد الحقيقية $\mathbb{R}$. اما من وجهة نظر العلوم الاجتماعية ومنها العلوم الطبية فإنها تعرف المتغير العشوائي على انه الصفة المشتركة لكل الوحدات الإحصائية التي تشكل المجتمع الإحصائي، التي تمثل موضوع الدراسة، مثل: الطول، والعمر، ومستوى التأهيل العلمي، ونسبة انتشار مرض معين وغيرها. وقبل البدء في اختيار أي طريقة إحصائية لتحليل البيانات، لابد من معرفة مستويات القياس الخاصة بالبيانات (Data Scales)، والقياس على وجه العموم هو

تحديد السمات والخصائص للمتغيرات حسب قواعد معينة، ومستويات القياس تشير الى العلاقة بين القيم التي تخصص لمفردات المتغير. ومن ثم فان طرق التحليل الاحصائي تختلف حسب مستويات قياس المتغيرات. ويجرى القياس الإحصائي للمتغيرات حسب طبيعة تلك المتغيرات من منظور الإحصاء التطبيقي، إذ يتم تصنف المتغيرات إلى:

1-4-2 المتغيرات العشوائية الاسمية Nominal Random Variables

هي الخاصية أو النوعية التي تستعمل لمجرد تصنيف مفردات الدراسة (مثل: ذكور مقابل إناث، وكذلك: ريف مقابل حضر)، وبشكل مبسط، يدل هذا النوع من المتغيرات على التصنيفات أو الأنواع، فعلى سبيل المثال، الأرقام الموجودة في ملفات المرضى في العيادات الطبية او المستشفيات، هي فقط لتصنيف المرضى، ولا يمكن بأي حال من الأحوال أن نقول إن المريض رقم 10 أكبر مرتين من المريض رقم 5. ومن الأمثلة الأخرى على هذا النوع من البيانات، الفئة العمرية، ونوع المدرسة، والموضوع الذي يدرس، وغيرها. وأحيانا تستعمل الأرقام في الترميز الإحصائي لتلك المتغيرات، فإن الأرقام يقتصر دورها على مجرد الترميز دون أي دلالة كمية، على سبيل المثال قد نرمز إلى الذكور بالرقم (1) ونرمز إلى الإناث بالرقم (2)، فالرقم (2) لم يستعمل بمعنى أنه أكبر من الرقم (1)، كما أن الرقم (1) لم يستعمل بمعنى أنه أصغر من الرقم (2) وإنما استعمل كلاهما لتصنيف الأفراد إلى ذكور وإناث، والتصنيفات في هذه الحالة مختلفة وغير متكررة وليس لها أي دلالة رقمية، فالأرقام في هذه الحالة لم توضع إلا لتبسيط التعامل مع الأصناف. فالبيانات (الأرقام) في هذه الحالة فقط تصنّف البيانات ولا تعطي لها أي ترتيب، ولا تجرى عليها العمليات الحسابية (كالضرب والجمع والطرح والقسمة). والقياس الاسمي هو أدنى المستويات الأربعة وهو ليس كمياً، بل مجرد تصنيف الأشياء مثل الذكر والأنثى وهذا الترميز ضروري احيانا للتعامل الإحصائي مع البيانات.

2-4-2 المتغيرات العشوائية الرتبية Ordinal Random Variables

وهي الصفة التي تستعمل لتصنيف الأفراد أو الموضوعات ومن ثم فان هذه الصفة تكون قابلة للترتيب عبر ترتيب الافراد تصاعدياً أو تنازلياً، وبغض النظر عن الفروق بين مستويات التصنيف، وتستعمل الأعداد في ترتيب الأشياء (تنازليا أو تصاعديا) وتجرى على المتغيرات عملية المقارنة أكبر من (>) أو أصغر من (<) فعلى سبيل المثال 2 < 3، 3 < 5 فإن 2 < 5. ويهتم هذا المقياس بترتيب الصفة كالأول على طلبة المدرسة والثاني والثالث ...الخ، وتصنيف الترتيب كميا فالترتيب يكون 1، 2، 3، ... وليس 1.5، ولا معنى هنا للفرق بين 2 و3. وهذا يعني عدم قبول هذا المقياس للعمليات الحسابية، ولكننا يمكن إجراء حساب التكرار لكل جزء ومن ثم يمكن حساب هذه الأعداد (التكرارات) لحساب بعض مقاييس الإحصاء كمعامل سبيرمان للرتب وغير ذلك، ومثال اخر هو ترتيب أفراد البحث تصاعدياً حسب الفئة العمرية، فالشخص الأكبر عمراً يحتل ترتيباً أكبر من الشخص الأوسط عمراً، وهذا بدوره يحتل ترتيباً أكبر من الشخص الأصغر عمراً، والشخص الذي يليه في العمر يحتل ترتيباً أقل .... وهكذا. فإذا كانت الأرقام الترتيبية هي: 1، 2، 3، 4 فإن الرقم 4 يرمز إلى الترتيب الأكبر، بينما الرقم 1 يرمز إلى الترتيب الأصغر، لكن ذلك لا يعني تساوي الفروق في الأعمار بين المفردات محل الدراسة.

2-4-3 المتغيرات العشوائية الفئوية Interval Random Variables

يستعمل هذا المقياس الأعداد للمقارنة بين المفردات وما شابهها أي بمقارنة المدى بين قياسين، فعلى سبيل المثال درجة سعد تفوق درجة أحمد بمقدار 10 درجات، وفيه تستعمل عمليتا الجمع والطرح فقط، لذلك فان درجة الحرارة 40 لا تعني بأنها ضعف الحرارة عندما تكون الدرجة 20، ويستعمل هذا المقياس بكثرة في المقاييس التربوية والنفسية. وتتضمن هذه المتغيرات معنى المتغيرات الرتبية ولكنها فضلا عن ذلك تتضمن تساوي الفروق بين الفئات في الصفة المقاسة، وهذا يعنى أن لهذا النوع من

المتغيرات وحدة قياس إلا أن الصفر أو نقطة الأصل أو البداية لا تعنى غياب الظاهرة أو الخاصية المقاسة، فحصول بعض الطلاب على (صفر) في مادة دراسية معينة لا يعنى أنهم لا يعرفون شيئاً في هذه المادة الدراسية، وإنما هم (بالتأكيد) لديهم بعض المعلومات عنها، فالصفر هنا صفر نسبى أي غير مطلق. ويمكن حساب المتوسطات والانحراف المعياري وغيرها لهذا النوع من المتغيرات الفئوية.

4-4-2 المتغيرات العشوائية النسبية Ratio Random Variables

وهي المتغيرات التي تتوافر فيها معاني المتغيرات الفئوية بجانب كونها تتضمن الصفر المطلق، والصفر هنا يعني غياب الصفة المقاسة ويعني وجود العدد السالب كدرجات الحرارة فوق وتحت الصفر، والمتغيرات من هذا النوع توجد في العلوم الطبيعية، بالإضافة الى بعض الظواهر في العلوم الاجتماعية والانسانية. وفي هذا النوع من المقاييس تستعمل الأعداد لوضع علاقة بين الأشياء وتجرى عليها العمليات الحسابية، فعلى سبيل المثال وزن أحمد ثلاثة أمثال وزن علي (احمد 90 كغم، علي 30 كغم). ويعتبر هذا المقياس من أعلى مستويات القياس لقبوله العمليات الحسابية الأربع المعروفة وطرق الإحصاء المعلمي. وخلاصة ما تقدم تتمثل العلاقة بين الإحصاء في العلوم الطبية ومستويات القياس من خلال ان القياس يقصد به تعيين أرقام أو مستويات مختلفة للصفات المقاسة باختلاف الأفراد، أما الإحصاء فهو يستعمل هذه الأرقام أو المستويات، ويتعامل معها بأساليب معينة تناسب مشكلة الدراسة أو تساؤلاتها. والمهم هو اختيار الأسلوب الإحصائي المناسب للبيانات شريطة أن يكون لذلك معنى مفهوم وواضح، فعلى سبيل المثال يمكن حساب المتوسط لعدد الأبناء في عينة ما ولذلك معنى مفهوم وواضح بينما متوسط الجنس لا معنى له، والجدول رقم (2-1) يوضح بعض الأساليب الإحصائية التي يمكن استعمالها حسب نوع المتغير ومستوى القياس له.

جدول رقم (2–1) العلاقة بين الاحصاء الحيوي ومستويات القياس

الإحصاء	المتغير الاسمي	المتغير الترتيبي	المتغير الفئوي	المتغير النسبي
الوصفي	التكرارات النسبة المئوية الأعمدة البيانية المنوال	التكرارات النسبة المئوية الأعمدة البيانية الوسيط نصف المدى الربيعي ارتباط سبيرمان	التكرارات النسبة المئوية المدرج / المضلع المنوال الوسيط المتوسط التباين الانحراف المعياري ارتباط بيرسون	التكرارات النسبة المئوية المدرج / المضلع المنوال الوسيط المتوسط التباين الانحراف المعياري ارتباط بيرسون
الاستدلالي	Chi-Square Test	Mann-Whitney Test. Friedman Test. Wilcoxon Tests. Kruskal-Wallis Test.	T- Test ANOVA	T- Test ANOVA

تمارين الفصل الثاني

1- ناقش العبارة الاتية (يعد تصميم الدراسة في الدراسات الطبية أكثر اهمية من تحليل البيانات) موضحاً أهم التصاميم للدراسات الاحصائية.

2- وضح بالرسم اهم تصاميم الدراسات في مجال العلوم الطبية.

3- فكر في دراسة طبية ترغب في اجرائها وحدد العوامل الاتية:

أ- حدد الهدف من الدراسة. ب- اشرح التصميم المناسب للدراسة. ت- وضح كيف يتم اختيار مفردات الدراسة. ث- وضح نوع البيانات التي سوف يتم جمعها.

4- عرف المتغير العشوائي، وحدد انواعه.

5- حدد أي من المتغيرات الآتية متقطع (منفصل) وأيها متصل (مستمر):

أ- عدد المكالمات التي يستقبلها المنزل خلال ساعة معينة.

ب- درجات الطلاب في امتحان الإحصاء الحيوي.

ت- عدد الزبائن الذين يدخلون أحد المستشفيات في اليوم.

ث- درجات الحرارة خلال شهر يناير في محافظة بابل.

ج- مساحة أراضي محافظة كربلاء المزروعة بالحمضيات.

ح- الدخل السنوي لموظفي أحد الدوائر.

6- حدد نوع البيانات الاتية التي هي عبارة عن أوزان (بالكيلوغرام) مجموعة مكونة من سبعة أشخاص: 50 ,55 ,35 ,45 ,40 ,30 ,25.

7- يعتقد أحد الباحثين أن استعمال اسيتامينوفين في اثناء فترة الثلث الأول من الحمل للنساء يمكن ان يسبب تلف الانبوب العصبي. واذ تم تقدير خطر تلف الانبوب العصبي في عموم المجتمع بنسبة 1: 1000. ما هو أفضل تصميم للدراسة من بين التصاميم الاتية:

أ- دراسة المتابعة (Cohort study).

ب-دراسة الحالة (Case-Control study).

ت-التجربة الطبية (Clinical study).

ث-الدراسة المقطعية (Cross-sectional study).

8- يدعي المسؤولين في أحد المستشفيات زيادة في حدوث سرطان الدم اللمفاوي الحاد (ALL) بين الأطفال بعمر 5-12 سنة. لقد أشاروا الى ان عدد من الأمهات في المدينة قد تعرضوا الى نفايات كيمياوية من مصنع قريب. ويعتقدون أن النفايات الكيمياوية تسبب السرطان. إذا صممت دراسة لتقييم ادعاء المسؤولين في المستشفى، أي من المواضيع التالية يكون الأقرب ليضمن ضمن مجموعة السيطرة؟

أ- الأطفال الذين تعرضوا للنفايات الكيمياوية، ولكن لا يعانون من السرطان.

ب- الأطفال الذين لم يتعرضوا للنفايات الكيمياوية والذين لا يعانون من السرطان.

ت-الأطفال الذين يراجعون العيادات الخارجية للمستشفى ولا يعانون من السرطان.

ث-الأطفال الذين لم يتعرضوا للنفايات الكيمياوية ويعانون من السرطان.

ج- الأطفال الذين يعانون من السرطان ويأخذون العلاج له.

9- مجموعة من النساء اللاتي يراجعن للفحص الروتيني قد تم سؤالهن عن نوع اللحم الذي يستهلكنه. 20% من النساء كانت اجابتهن بأنهن نباتيات. وخلال 5 سنوات مدة التدقيق، كان 5 نساء من النباتيات و43 امرأة من غير النباتيات ظهر لديهن سرطان القولون والمستقيم. أيا من التصاميم الاتية يعد التصميم المناسب للدراسة؟

أ- دراسة المتابعة (Cohort study).

ب-دراسة الحالة (Case-control study).

ت-الدراسة المقطعية (Cross-sectional study).

ث-التجربة الطبية العشوائية (Randomized clinical trial).

10- مجموعة من مرضى سرطان الرئة انضموا الى مجموعة من غير المرضى، وقورنت عادات التدخين لهم خلال حياتهم. عل ضوء هذه المعلومات حسب الباحثون نسبة المدخنين من ذوي الإصابة بسرطان الرئة مقارنة بنسبتهم من المدخنين من غير المصابين بالسرطان. هذا يعد مثالاً لاي من التصاميم:

أ- دراسة الحالة (Case-Control study).

ب-دراسة المتابعة (Cohort study).

ت-الدراسة المقطعية (Cross-sectional study).

ث-الدراسة العشوائية المسيطرة (Randomized controlled study).

11- في أحدى الدراسات تعرضت مجموعة من الأشخاص لنوع من السموم الطبيعية ولم يعالجوا، بدلا من ذلك قاموا بمتابعتهم عبر مدة من الزمن عبر مجموعة من المقاييس المعيارية للتأكد من التأثير المحتمل للسم. هذا النوع من تصميم الدراسة يسمى:

أ- التجربة الطبية (Clinical trial).

ب-مزدوج التعمية (Double-blind).

ت-تصميم دراسات المتابعة (Prospective cohort).

ث-الفوج بأثر رجعي (Retrospective cohort).

12- نوع من العلاج اظهر مختبريا نشاطاً ضد مرض الايدز واختبر على مجتمع من المرضى المؤكد اصابتهم بالإيدز، ومن بين 200 شخص من مجتمع المرضى اختير بالقرعة 100 شخص لأخذ العلاج. هذا العلاج غير المفحوص مزج بكوب من عصير البرتقال، المرضى الاخرين اعطوهم عصير البرتقال فقط. لا أحد في

المستشفى من الممرضات او الأطباء او المرضى يعرفون من المرضى الذين اعطوهم العلاج. وفي نهاية الدراسة حدد عدد الخلايا (CD4+T). هذا يمثل أي نوع من التصاميم؟

أ- دراسة الحالة (Case-control study).

ب-تقرير الحالة (Case report).

ت-دراية الفوج (المتابعة) (Cohort study).

ث-الدراسة المقطعية (Cross-sectional study).

ج- تجربة سريرية عشوائية مزدوجة التعمية (Double-blind randomizes clinical trial).

13- الغرض من تصميم دراسة التعمية المزدوجة (Double-Blind) هو:

أ- الوصول الى المقارنة بين العناصر المعالجة وغير المعالجة.

ب-تقليل تأثير التغيير في المعاينة.

ت- تجاوز تحيز المشاهدة والعناصر.

ث- تجاوز تحيز المشاهدة وتغير المعاينة.

الفصل الثالث

عرض البيانات باستعمال الجداول التكرارية

Frequency Tables Presentation

3-1 تمهيد

تعرف عملية وصف المتغيرات على انها عملية تلخيص المتغيرات بحيث يحصل على جداول تكرارية او اشكال بيانية او قيم عددية تعبر عن أو تصف الحالات في البيانات . ومن ثم فان نوع المتغير ومستوى قياسه يحدد الوسيلة التي يمكن استعمالها لوصف هذا المتغير كما موضح بالجدول (2-1) في الفصل السابق. وبشكل عام هناك ثلاث طرائق لوصف المتغيرات العشوائية تتمثل بالجداول التكرارية (Frequency Tables) والاشكال البيانية (Graphical Charts) والمقاييس العددية (Numerical Measurements). وسوف يخصص هذا الفصل للعرض الجدولي للبيانات الذي يعد الخطوة الأولى في تلخيص ووصف البيانات لاختصار الاعداد الكبيرة لها ووضعها في جداول توضح الصورة الأولية للبيانات. بعد ان تم جمع البيانات توافرت مجموعة من المشاهدات تسمى البيانات المفردة او غير المبوبة، وعندما يكون عددها كبيراً جدا يكون من الضروري تنظيمها وتلخيصها لكي يمكن التعامل معها. وعندما يكون المطلوب تنظيم البيانات فهذا يعني الحصول على التوزيع التكراري للبيانات، ويعرف التوزيع التكراري للبيانات على انه جدول يتضمن عمودين يمثل العمود الأول فئات المتغير، والعمود الثاني التكرار الذي يمثل عدد المشاهدات في كل فئة من فئات المتغير، ويمكن الحصول عليه بواحدة من طريقتين اعتمادا على نوع البيانات.

3-2 الجدول التكراري للبيانات الوصفية.

Frequency table for qualitative data

يعرف الجدول التكراري في حالة البيانات الوصفية (qualitative Data) على انه جدول يتكون من عمودين أحدهما: يمثل صفات (فئات) الظاهرة والعمود الثاني:

يمثل التكرارات وهي عبارة عن عدد المشاهدات في كل صفة (فئة). ويعد تكوين التوزيع التكراري في حالة البيانات الوصفية (النوعية) أسهل منها في حالة البيانات الكمية ، لأنه يعتمد على صفات المتغير او الظاهرة وهذه الصفات واضحة ومحددة ومن ثم حساب عدد المشاهدات في كل صفة ومن ثم يتكون جدول التوزيع التكراري الاولي من ثلاثة اعمدة، الأول: يمثل الصفات والثاني: يمثل الاشارات (العلامات) والثالث: يمثل التكرارات او عدد المشاهدات في كل صفة، ولكن الجدول النهائي ويسمى الجدول التكراري البسيط يكون مكون من عمودين الأول: يمثل الصفات والثاني: يمثل التكرارات وكما موضح في المثال الاتي:

مثال 1

البيانات الاتية تمثل الحالة المرضية لمجموعة من 18 شخص وكالاتي:

بسيطة، متوسطة، شديدة، شديدة، شديدة، بسيطة، بسيطة، بسيطة، غير مريض، غير مريض، متوسطة، متوسطة، شديدة، متوسطة، بسيطة، متوسطة، غير مريض، بسيطة، متوسطة.

الحل:

نكون جدول يتألف من ثلاثة اعمدة يتضمن العمود الأول: يمثل الصفات والثاني: يتضمن الاشارات والثالث: يمثل التكرار كالاتي:

جدول (3–1) الجدول التكراري للبيانات الوصفية

الصفات	الاشارات	التكرار
غير مريض	|||	3
بسيطة	~~||||~~	5
متوسطة	~~||||~~ |	6
شديدة	||||	4
المجموع		18

والجدول النهائي يتضمن فقط العمودين الاول والثالث من الجدول ودائما مجموع التكرارات يساوي عدد المشاهدات.

3−3 الجدول التكراري للبيانات الكمية

Frequency table for quantitative data

يختلف تكوين الجدول التكراري في حالة البيانات الكمية عنه في حالة البيانات النوعية، لأنه في حالة البيانات النوعية فئات المتغير واضحة بينما في حالة البيانات الكمية، ولتكوين الجدول التكراري الذي يستعمل لتنظيم وتلخيص البيانات الكمية (Quantitative Data)، نحتاج الى ان نتبع الخطوات الاتية:

1− تحديد أكبر قيمة (A) وأصغر قيمة (B) في البيانات ومنها نحدد المدى (R) حسب الصيغة الاتية:

$$R = A - B \qquad (1-3)$$

2− تحديد عدد الفئات ويفضل أن يتناسب مع عدد البيانات والكثير من الكتاب يقترح ان يتراوح عدد الفئات من 5 الى 15 فئة، وهناك قاعدة مقترحة تسمى قاعدة سترجيوس تتمثل بالعلاقة رقم (3−2)، وتعد هذه العلاقة مجرد دليل أو مساعد على اعطاء فكرة تقريبية عن عدد الفئات:

$$k = 1 + 3.3 \log n \qquad (2\text{-}3)$$

حيث ان (n) تمثل عدد المشاهدات (البيانات)، (k) تمثل عدد الفئات.

3- نحدد طول الفئة (h) ويكون حسب العلاقة (3-3)

$$h = \frac{R}{k} \qquad (3\text{-}3)$$

4− تكوين الفئات إذ يبدأ باختيار أصغر قيمة او اي قيمة أصغر منها وتعتبر الحد الادنى للفئة الاولى. والحد الاعلى للفئة الاولى يساوي الحد الادنى للفئة + طول الفئة. والحد الاعلى للفئة الاولى يمثل الحد الادنى للفئة الثانية والحد الاعلى للفئة الثانية يساوي الحد الادنى للفئة الثانية زائدا طول الفئة وهكذا الى ان نصل الى أكبر قيمة او أكبر منها ليمثل الحد الاعلى للفئة الاخيرة.

5- بعد كتابة الفئات تفرغ البيانات على شكل علامات إذ تمثل كل قيمة بعلامة واحدة وبعد ذلك تعد هذه العلامات لتحويلها الى تكرارات. وكما هو موضح في المثال (2) الاتي:

مثال 2

قام طبيب بفحص مستوى الكولسترول لـ (20) مريضاً فكانت مقاسة بالملغرام/ مليلتر كالاتي:

210، 208، 210، 208، 203، 210، 207، 217، 208، 212، 209، 210، 214، 200، 218، 202، 218، 213، 221، 215، 199، 210.

والمطلوب: تلخيص هذه البيانات في جدول تكراري يتكون من 5 فئات.

1- أصغر قيمة هي 199 وأكبر قيمة هي 221. اذن مدى التغير يساوي

مدى التغير (R) = أكبر قيمة – أصغر قيمة=221- 199= 22

2- عدد الفئات تم تحديده (k=5) وعندما نستعمل القاعدة في العلاقة (3-2) يكون كالاتي:

$$k = 1 + 3.3log20$$
$$= 5.29 \approx 5$$

وهو الرقم نفسه المفروض مع المثال.

3- تحديد طول الفئة حسب العلاقة (3-3)

$$h = \frac{R}{k}$$
$$= \frac{22}{5}$$
$$= 4.4 \approx 5$$

يقرب طول الفئة دائما لأكبر عدد صحيح لسهولة العمل الحسابي.

4- تكوين الفئات كالاتي:

الفئة الأولى: تبدأ بأصغر قيمة وهي 199 او اقل منها، ويكون الحد الاعلى لها= الحد الادنى + طول الفئة = 199+5=204 والفئة الثانية: تبدأ بالحد الاعلى للفئة الاولى

204 ويكون الحد الاعلى لها هو 204+ 5=209 وهكذا حتى نصل الى أكبر قيمة او أكبر منها عند الحد الاعلى للفئة الاخيرة وكما موضح في الجدول الاتي:

جدول (3-2) الجدول التكراري للبيانات الكمية

الفئات	العلامات	التكرارات
204-199	||||	4
209-204	|||	3
214-209	𝍸 ||	7
219-214	𝍸	5
224-219	|	1
المجموع		20

وتكتب الفئات بعدة طرق، والطريقة في الجدول السابق تكون غير مقبولة، لأنه يصبح من المحير وضع الرقم 209 هل هو في الفئة الثانية ام في الفئة الثالثة، والجدول النهائي يتكون من عمودين هما الفئات والتكرارات. فاذا كانت البيانات متقطعة تكتب الفئات كما في الجدول الاتي:

جدول (3-3) الجدول التكراري للبيانات الكمية المتقطعة

الفئات	203-199	208-204	213-209	218-214	223-219
التكرار	4	3	7	5	1

ولكن في حال البيانات المستمرة والتي تأخذ كسور تكون غير صالحة لأنه هناك فاصلة بين الحد الاعلى للفئة السابقة والحد الادنى للفئة اللاحقة. لذلك تعد الطريقة الاتية أفضل من الطريقتين السابقتين والتي تستعمل غالبا بغض النظر عن نوع البيانات:

جدول (3-4) الجدول التكراري للبيانات الكمية

الفئات	-199	-204	-209	-214	224-219
التكرار	4	3	7	5	1

3-4 انواع الجداول التكرارية Types of frequency Tables

هناك عدة انواع من الجداول التكرارية يمكن تكوينها لتلخيص البيانات الكمية منها:

1-4-3 الجدول التكراري البسيط: (simple frequency table) ويتكون من عمودين الأول: يتضمن الفئات والثاني: يتضمن التكرارات ويكون على انواع على اساس حدود الفئات:

أ- **الجدول التكراري البسيط المغلق: (closed simple frequency table)** ويكون فيه الحد الادنى للفئة الاولى والحد الاعلى للفئة الاخيرة موجودان كما في الجدول (3-4) من المثال السابق.

ب- **الجدول التكراري المفتوح من الطرف الادنى:** وفيه يكون الحد الادنى للفئة الاولى مفقوداً والجدول التكراري المفتوح من الطرف الادنى يكون كما في الجدول الاتي:

جدول (3-5) الجدول التكراري البسيط المفتوح من الطرف الادنى

الفئات	اقل من 204	-204	-209	-214	224-219
التكرار	4	3	7	5	1

ت- **الجدول التكراري المفتوح من الطرف الاعلى:** وفيه يكون الحد الاعلى للفئة الاخيرة مفقوداً والجدول التكراري المفتوح من الطرف الاعلى يكون كما في الجدول الاتي:

جدول (3-6) الجدول التكراري البسيط المفتوح من الطرف الاعلى

الفئات	-199	-204	-209	-214	219 فأكثر
التكرار	4	3	7	5	1

ث- **الجدول التكراري المفتوح من الطرفين:** وفيه يكون الحد الادنى للفئة الاولى والحد الاعلى للفئة الاخيرة مفقودين والجدول التكراري المفتوح من الطرفين يكون كما في الجدول الاتي:

جدول (3−7) الجدول التكراري البسيط المفتوح من الطرفين

الفئات	اقل من 204	−204	−209	−214	219 فأكثر
التكرار	4	3	7	5	1

ويفضل في العمل الاحصائي، ولتسهيل العمل الحسابي استعمال الجداول التكرارية المغلقة والابتعاد عن الجداول التكرارية المفتوحة، وغالبا ما يستعمل التقدير لتحويل الجدول من جدول تكراري مفتوح الى جدول تكراري مغلق، بالاعتماد على طول الفئة. فضلا عن ذلك هناك تقسيم اخر للجداول التكرارية البسيطة وحسب طول الفئة تكون على نوعين: جداول تكرارية منتظمة، وتكون فيها الفئات متساوية الطول، والنوع الاخر جداول تكرارية غير منتظمة، وتكون فيها الفئات غير متساوية الطول.

3−4−2 الجدول التكراري البسيط النسبي Relative frequency table

وهو جدول تكراري يتكون من عمودين: الأول: يمثل الفئات، والثاني: يمثل النسبة العادية او النسبة المئوية للتكرارات ويكون على نوعين:

أ− **الجدول التكراري النسبي:** (Relative frequency table) ويتكون من عمودين: الأول: يتضمن الفئات، والثاني: يتضمن التكرار النسبي والذي يمثل نسبة تكرار كل فئة في الجدول الى مجموع التكرارات، فعلى سبيل المثال يكون الجدول التكرار النسبي للجدول (3−4) كما في الجدول الاتي:

جدول (3−8) جدول التكرار النسبي

الفئات	التكرارات	التكرار النسبي= تكرار الفئة/ مجموع التكرارات
−199	4	0.2=20/4
−204	3	0.15=20/3
−209	7	0.35=20/7
−214	5	0.25=20/5
224−219	1	0.05=20/1
المجموع	20	1

ب – الجدول التكراري النسبي المئوي:(Percentage frequency table)

والفرق عن الجدول السابق هو انه تتكون النسبة المئوية للتكرارات من خلال ضرب التكرار النسبي في 100 وجدول التكرار النسبي المئوي للجدول (3–4) يكون كما في الجدول الاتي:

جدول (3–9) الجدول التكراري النسبي المئوي

الفئات	التكرارات	التكرار النسبي المئوي = تكرار الفئة/ مجموع التكرارات × 100
199–	4	20=100×20/4
204–	3	15=100×20/3
209–	7	35=100×20/7
214–	5	25=100×20/5
219–224	1	5=100×20/1
المجموع	20	100

3–4–3 الجداول التكرارية المتجمعة Cumulative Frequency Tables

وهي جداول تكرارية تتكون من عمودين: الأول: يمثل الفئات، والثاني: يمثل التكرار المتجمع اما تصاعديا او تنازليا وتكون على نوعين:

أ_ الجدول التكراري المتجمع الصاعد:

(Upper Cumulative Frequency Table)

ويتكون من عمودين: الأول: يمثل الحدود العليا للفئات، والعمود الثاني: يمثل التكرار المتجمع تصاعديا والجدول التكراري المتجمع الصاعد للبيانات في الجدول (3–4) يكون كما في الجدول الاتي:

جدول (3−10) الجدول التكراري المتجمع الصاعد

الحدود العليات للفئات	التكرار المتجمع الصاعد
أقل من 204	4
أقل من 209	7 = 3+4
أقل من 214	14 = 7+7
أقل من 219	19= 5+14
أقل من 224	19 +1 = 20 (مجموع التكرارات)

ودائما يكون تكرار الحد المتجمع الاخير يساوي مجموع التكرارات.

ب−الجدول التكراري المتجمع النازل Lower Cumulative Frequency Table

ويتكون من عمودين: الأول: يمثل الحدود الدنيا للفئات، والعمود الثاني: يمثل التكرار المتجمع تنازليا وجدول التكرار المتجمع النازل للجدول (3−4) يكون كما في الجدول الاتي:

جدول (3−11) الجدول التكراري المتجمع الهابط

الحدود الدنيا للفئات	التكرار المتجمع النازل
أكثر من 199	20 (مجموع التكرارات)
أكثر من 204	20 −4 =16
أكثر من 209	16−3 =13
أكثر من 214	13−7 =6
أكثر من 219	6 −5 =1

ودائما يكون تكرار الحد المتجمع الاول يساوي مجموع التكرارات.

3-4-4 الجدول التكراري المزدوج (Joint Frequency (Contingency) Table):

ويمثل ظاهرتين مع بعضهما، ويتكون من اتجاهين الافقي يمثل فئات الظاهرة الأولى، والاتجاه العمودي يمثل فئات الظاهرة الثانية، وقد تكون هاتان الظاهرتان كمية او تكون احداهما كمية والاخرى وصفية او تكون الاثنتان وصفيتين وكما في الامثلة الاتية:

مثال 3

قام طبيب بفحص مستوى الكولسترول لـ (20) مريض فكانت بالملغرام/ مليلتر مع الوزن كغم لكل منهم كالاتي:

جدول (3−12) البيانات الأولية للمرضى

الكوليسترول	208	203	210	207	217	208	212	209	210	215
الوزن	75	62	56	58	65	88	72	67	55	70
الكوليسترول	199	221	213	210	210	214	200	218	202	218
الوزن	65	70	55	66	58	97	86	110	95	77

والمطلوب: تلخيص هذه البيانات في جدول تكراري يتكون من 5 فئات لكل من الظاهرتين.

الحل:

فئات الظاهرة الاولى التي تمثل مستوى الكولسترول تكون كما في المثال السابق

224−219 −214 −209 −204 −199

اما فئات الظاهرة الثانية والتي تمثل الوزن تكون كالاتي:

1− أكبر قيمة = 110 وأصغر قيمة = 55 اذن

المدى (R) = أكبر قيمة – أصغر قيمة = 110−55 = 55.

2− طول الفئة = المدى / عدد الفئات = 5/55 = 11.

3− اذن الفئات ستكون كما في الجدول (3−13) الاتي، ومن ثم تفرغ البيانات في الجدول إذ ان كل قيمتين تمثل مشاهدة واحدة ومن ثم تمثيلها بعلامة واحدة.

جدول (3−13) الجدول التكراري المزدوج لظاهرتين كميتين

فئات الوزن / فئات الكولسترول	55-	66-	77-	88-	99- 110	التكرار
199-	111		1			4
204-	1	1		1		3
209-	1111	11	1			7
214-		111		11		5
219- 224					1	1
التكرار	8	6	2	3	1	20

واحدة من مميزات الجدول التكراري المزدوج انه يمكن فصله الى جدولين تكرارين بسيطين كل جدول يمثل ظاهرة، ولكن العكس غير ممكن، إذ لا يمكن دمج جدولين تكرارين بسيطين لتكوين جدول تكراري مزدوج.

مثال4

البيانات الاتية تمثل حالة الاصابة بمرض معين مع مستوى الهيموجلوبين في الدم لمجموعة من الاشخاص. والمطلوب تلخيص هذه البيانات في جدول تكراري مزدوج.

بسيطة، 18، متوسطة، 14، شديدة، 12، شديدة، 11، شديدة، 13، بسيطة، 19، بسيطة، 17، غير مريض، 19، غير مريض، 18، متوسطة، 14، متوسطة، 15، متوسطة، 16، شديدة ،12، متوسطة،14، بسيطة،17، متوسطة،16، غير مريض،19، بسيطة،18، متوسطة، 13، غير مريض، 19.

الحل:

يلاحظ انه في ظاهرة الاصابة بالمرض تقسيمات الفئات واضحة، وهي تمثل صفات الظاهرة وهي أربع صفات ومن ثم تكون فئات الظاهرة بالشكل الاتي:

غير مريض، بسيطة، متوسطة، شديدة

اما في حالة الظاهرة الكمية والتي تمثل مستوى الهيموجلوبين فإننا نتبع الخطوات في المثال السابق لتكوين فئات الظاهرة وكالاتي:

1- تحديد أكبر قيمة = 19 وأصغر قيمة =11 ومن ثم تحديد مدى التغير كالاتي:
المدى (R) = أكبر قيمة – أصغر قيمة = 19-11=8.
2- نفرض عدد الفئات بانه يساوي 4 كما في الظاهرة المصاحبة. وهذا ليس شرطا وانما اختيارا.
3- نجد طول الفئة = مدى التغير/عدد الفئات= 8/4=2
4- تكوين الفئات كالاتي: 11- 13- 15- 17-19
وبعد ذلك نكون الجدول التكراري المزدوج وكما في الجدول (3-14) الاتي:

جدول (3-14) الجدول التكراري المزدوج لظاهرتين الاولى وصفية والثانية كمية

مستوى الهيموجلوبين / حالة الاصابة بالمرض	11-	13-	15-	17-19	التكرار
غير مريض				1111	4
بسيطة				11111	5
متوسطه		1111	111		7
شديدة	111	1			4
التكرار	3	5	3	9	20

وكما ذكرنا في بداية الفصل بان الجداول التكرارية هي الخطوة الأولى لوصف وتلخيص البيانات الخام بعد جمعها عبر تقليل اعداد المشاهدات وتحويلها الى جداول تسهل قراءتها واخذ المعلومات الأولية منها. اما الخطوة الثانية فهي تمثيل البيانات الأولية او الجداول التكرارية بأشكال بيانية كما يتضح في الفصل القادم.

3-5 تكوين الجدول التكراري البسيط باستعمال البرنامج (SPSS)
Create the Simple Frequency Tables using (SPSS)

تم استعراض كيفية تكوين أنواع الجداول التكرارية البسيطة للبيانات الوصفية والكمية في المباحث السابقة يدويا وفي هذا المبحث سوف نوضح كيفية استعمال البرنامج (SPSS) لإعداد الجداول التكرارية البسيطة لهذه البيانات بنوعيها الوصفي والكمي.

3-5-1 الجدول التكراري البسيط للبيانات الوصفية (النوعية)

اعداد الجداول التكرارية للبيانات الوصفية باستعمال برنامج (SPSS) يعد سهلا ومباشرا لان الفئات واضحة ولا تحتاج الى عملية تصنيف او تجزئة للبيانات كما هو واضح من خلال المثال الاتي:

مثال 5

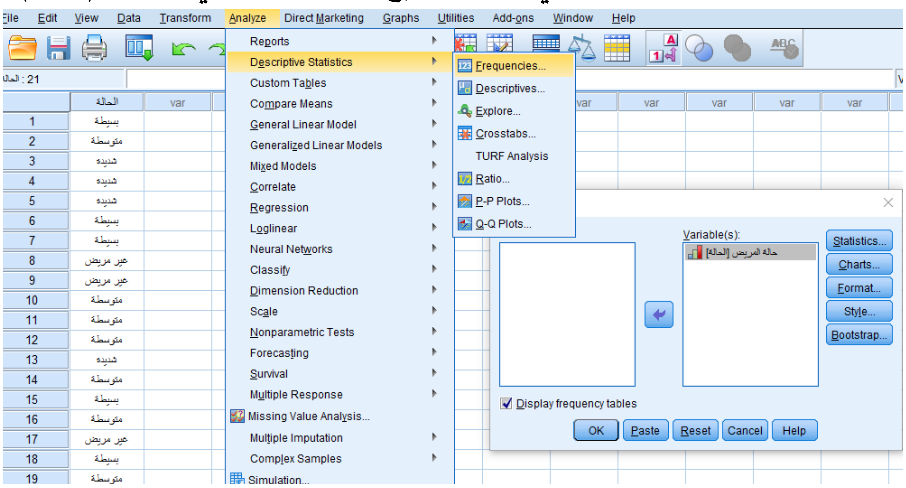
استعمل البيانات لحالة المرض في المثال (3-4) لتكوين جدول تكراري بسيط باستعمال البرنامج (SPSS)

الحل:

1- ادخال البيانات الى محرر البيانات في البرنامج بعد تسميته في صفحة المتغيرات كما في الشكل (3-1)

الشكل (3-1) ادخال البيانات الوصفية الى محرر البيانات

2- نستعمل الامر (Analyze) من قائمة الأوامر الرئيسة ومنها نستعمل الامر الفرعي (Descriptive Statistics) ومنه ننقر على الامر الفرعي (Frequencies) مع ملاحظة وضع علامة صح على المربع الذي يقابل اظهار الجدول التكراري في أسفل المربع الحواري كما في الشكل (3-2)



الشكل (3-2) المربع الحواري للتكرارات

3- بعد نقل اسم المتغير الى جهة اليمين في المربع ننقر على (OK) لنحصل على الجدول التكراري كما في الجدول (3-15)

الجدول (3-15) الجدول التكراري لحالة المريض

	Frequency	Percent	Cumulative Percent
غير مريض	4	23.8	23.8
بسيطة	5	23.8	47.6
متوسطة	7	33.3	81.0
شديدة	4	19.0	100.0
Total	21	100.0	

3-5-2 الجدول التكراري البسيط للبيانات الكمية

اما في حالة البيانات الكمية التي تمثل قيماً عددية للظاهرة، فان خطوات تكوين الجدول التكراري البسيط يحتاج الى خطوات أكثر، وتدخل من الباحث لتكوين الفئات من خلال اختيار عدد الفئات او طول الفئة والتدقيق لتضمين جميع البيانات ضمن الفئات وهذه الخطوات تكون كما موضحة في المثال الاتي.

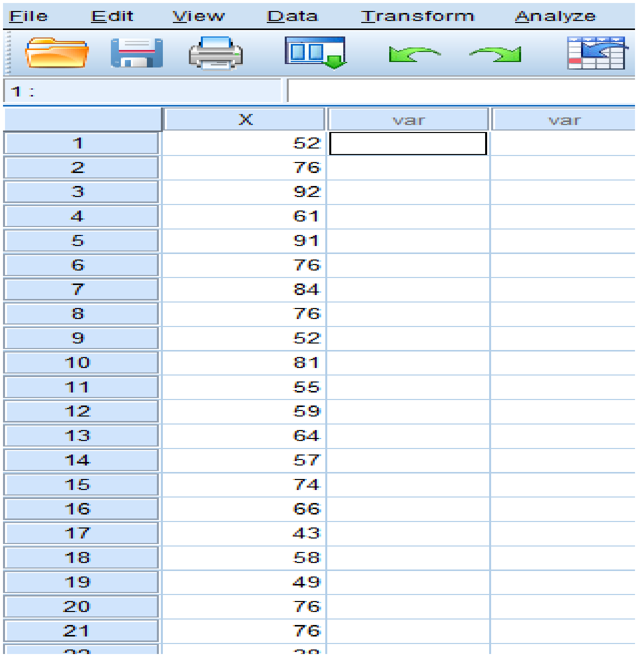
مثال 6

قام طبيب بقياس الوزن (كغم) لخمسين شخصاً اثناء مراجعتهم العيادة الخارجية في أحد المستشفيات خلال أحد أيام المراجعة فكانت:

43 66 74 57 64 59 55 81 52 76 84 76 91 61 92 76 52
76 90 64 62 52 56 46 42 54 64 38 53 38 76 76 49 58
.80 87 60 87 85 70 84 76 76 94 60 45 83 59 75 50

الحل:

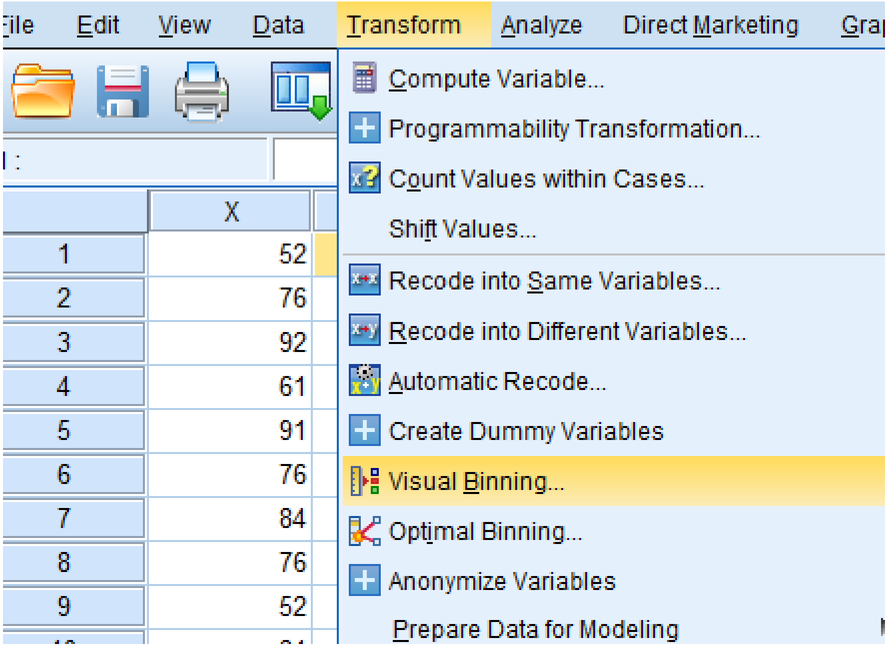
1- لتكوين جدول التوزيع التكراري البسيط للبيانات الكمية يتم ادخال البيانات الى محرر البيانات في برنامج (SPSS) كما في الشكل (3-3)



	X	var	var
1	52		
2	76		
3	92		
4	61		
5	91		
6	76		
7	84		
8	76		
9	52		
10	81		
11	55		
12	59		
13	64		
14	57		
15	74		
16	66		
17	43		
18	58		
19	49		
20	76		
21	76		

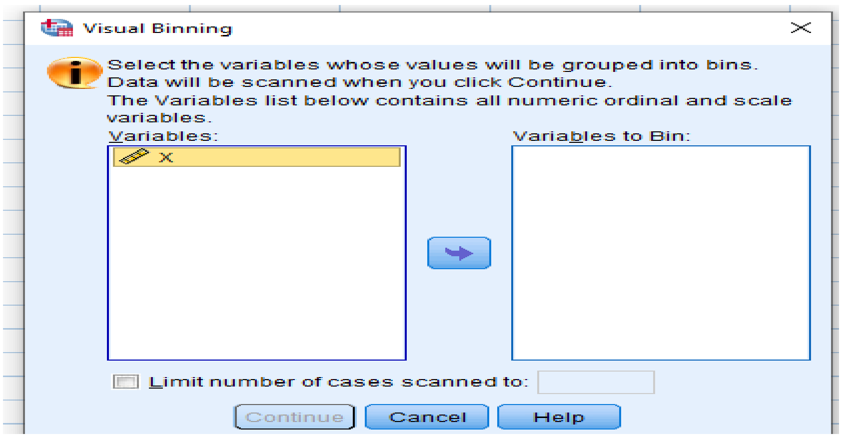
الشكل (3-3) ادخال البيانات الى محرر البيانات

2- الخطوة الأولى هي تكوين الفئات وتكون بالشكل الاتي من قائمة (Transformation) يتم اختيار الامر الفرعي (Visual Binning) كما الشكل (4-3)



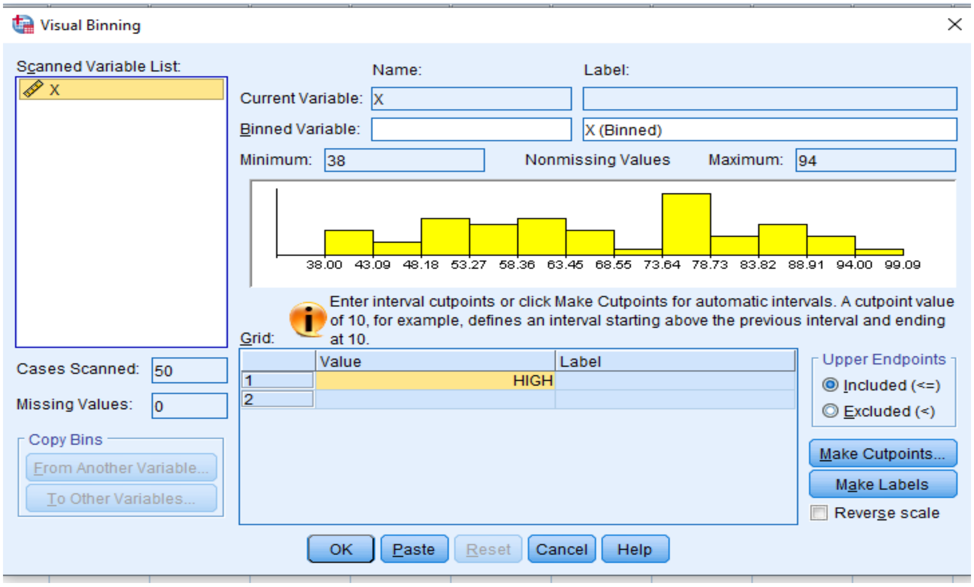
الشكل (3-4) تكوين الفئات

3- تتبع المسار في الشكل (3-4) تحصل منه على المربع الحواري في الشكل (3 -5)



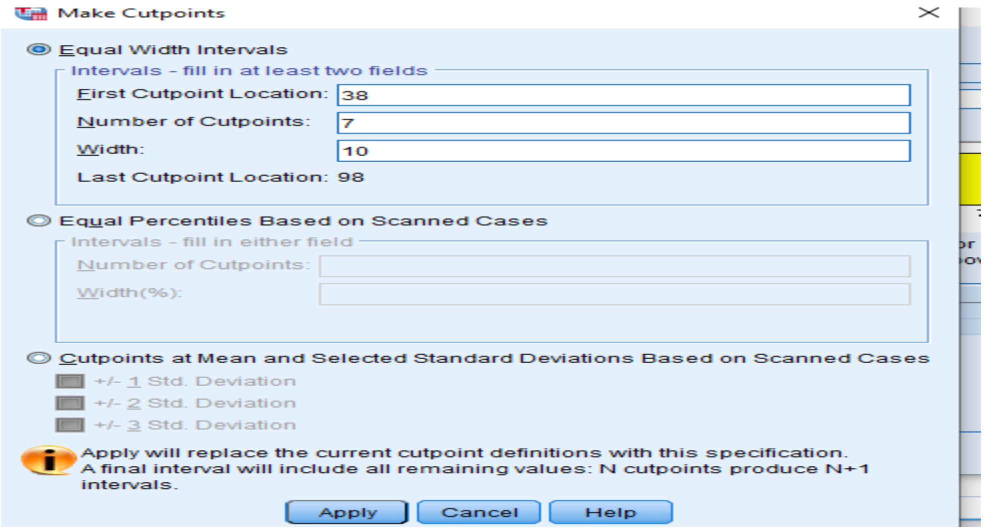
الشكل (3-5) المربع الحواري لإدخال اسم المتغير

4- يتم نقل المتغير من اليسار الى اليمين باستعمال السهم، ثم اختيار كلمة (Continue) لتحصل على المربع الحواري في الشكل (3-6).



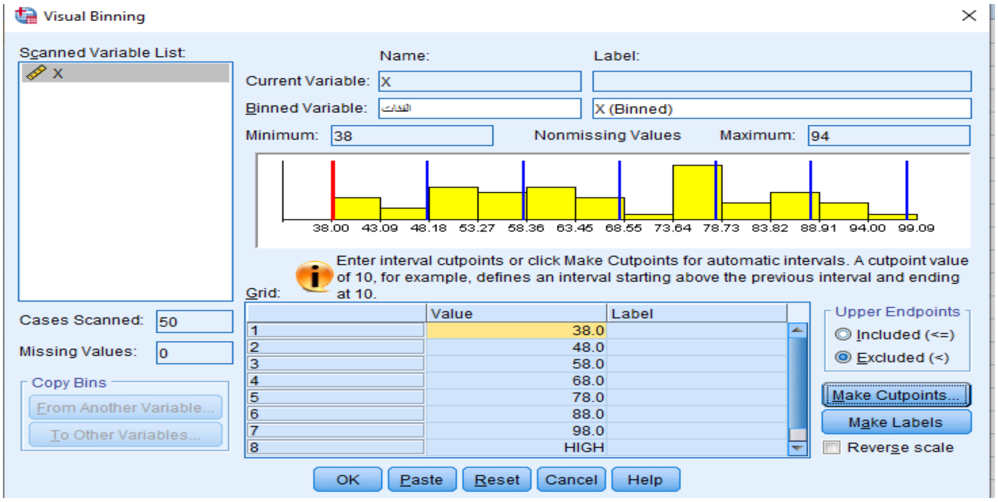
الشكل (3-6) المربع الحواري الخاص بتكوين الفئات

5- يلاحظ في المربع الحواري في الشكل (3-6) تحدد أصغر قيمة وأكبر قيمة في البيانات مع اسم المتغير المطلوب عمل فئات له، ويطلب منك إعطاء اسم جديد للفئات في الحقل (Binned variable)، وبعد كتابة الاسم نقوم بتحويل الخيار من (Included) الى (Excluded) ثم اختيار المربع (Make cutpoints) فتحصل على المربع الحواري في الشكل (3-7).



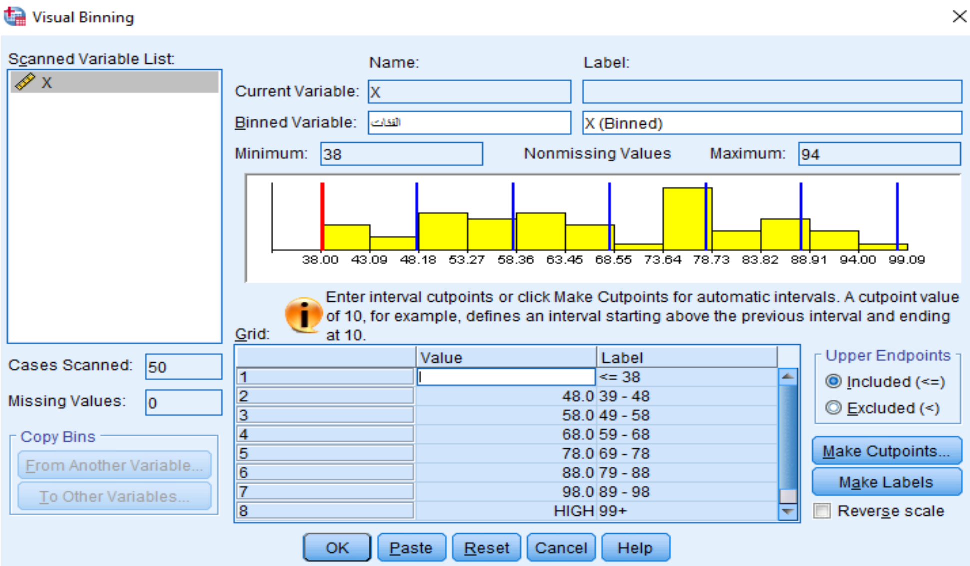
الشكل (3-7) المربع الحواري لاختيار اقل قيمة وطول الفئة او عدد الفئات

6- يطلب منك في المربع الحواري ادخال القيمة التي تمثل الحد الأدنى (First Cutpoint Location)، ومن ثم تحديد عدد الفئات (Number of Cutpoint)، او تحديد طول الفئة (Width)، فاذا حدد عدد الفئات سيحدد طول الفئة اتوماتيكيا، وكذلك إذا حدد طول الفئة سيحدد عدد الفئات، وبعد ذلك ينقر على كلمة (Apply) لتطبيق هذه المعلومات وتكون الفئات على المربع الحواري السابق كما في الشكل (3-8).



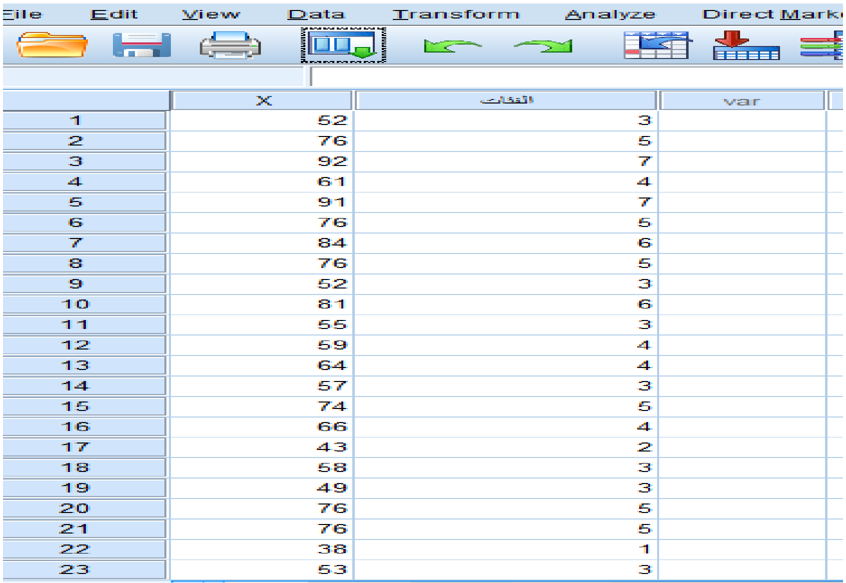
الشكل 3-8) تكوين حدود الفئات على المربع الحواري

7- ولغرض اظهار الفئات في مربع (Lebel) يختار المربع (Make Labels) فتتكون الفئات للتأكد من انها وصلت الى أكبر قيمة او أكبر منها كما في الشكل (3-9) وكالاتي.



الشكل (3-9) تكوين الفئات على الحقل (Label)

8- يلاحظ من الفئات تحت كلمة (Label) بدأت بأصغر قيمة وهي (38) وانتهت الفئة الأخيرة بالقيمة (98) وهي أكبر من أكبر قيمة بالبيانات والتي هي (92) اذن الفئات مقبولة وستتضمن القيم جميعها. وعند النقر على كلمة (OK) يتم ايجاد عمود جديد تحت اسم الفئات يوضح رقم الفئة التي يقع فيها الرقم كما في الشكل (3-10).

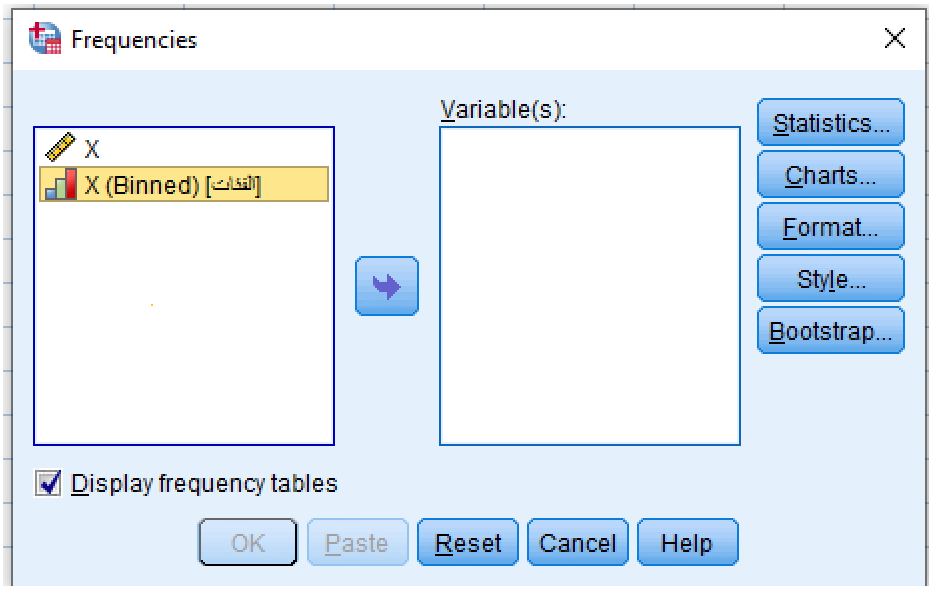


	X	الفئات	var
1	52	3	
2	76	5	
3	92	7	
4	61	4	
5	91	7	
6	76	5	
7	84	6	
8	76	5	
9	52	3	
10	81	6	
11	55	3	
12	59	4	
13	64	4	
14	57	3	
15	74	5	
16	66	4	
17	43	2	
18	58	3	
19	49	3	
20	76	5	
21	76	5	
22	38	1	
23	53	3	

الشكل (3-10) ظهور عمود يمثل الفئات على صفحة محرر البيان

9- يلاحظ في العمود الجديد ارقام تمثل رقم الفئة التي تقع ضمنها المشاهدة، ويمكن تغييرها الى الفئات بدلا من الأرقام عند النقر على المربع الخاص في شريط المهام، ولغرض اكمال الجدول التكراري يتم اختيار الامر (Analyze) من قائمة الأوامر ومنها يختار الامر الفرعي (Descriptive Statistics) ومنه يختار الامر الفرعي (Frequencies) نحصل على المربع الحواري

الاتي نحول المتغير الجديد الذي سميناه (الفئات) الى جهة اليمين كما في الشكل (3-11).



الشكل (3-11) المربع الحواري لتكوين التكرارات وإظهار الجدول التكراري

10- بعد تحويل المتغير الفئات الى جهة اليمين في المربع الحواري في الشكل (3-11) توضع علامة صح في المربع الذي يمثل اظهار الجدول التكراري ويتم اختيار كلمة (OK) للحصول على الجدول التكراري كما في الجدول (3-16)

الجدول (3–16) الجدول التكراري للبيانات التي تمثل الوزن

Intervals	Frequency	Percent	Cumulative Percent
38 – 47	6	12.0	12.0
48 – 57	10	20.0	32.0
58 – 67	11	22.0	54.0
68 – 77	11	22.0	76.0
78 – 87	8	16.0	92.0
88 – 97	4	8.0	100.0
Total	50	100.0	

يلاحظ بان الجدول (3–16) يتضمن التوزيع التكراري لظاهرة الوزن، إذ تضمن العمود الأول: الفئات والعمود الثاني: التكرار الذي يمثل عدد الأشخاص في كل فئة من فئات الوزن، ومن ثم التكرار النسبي في العمود الثالث، والعمود الأخير تضمن التكرار النسبي المتجمع. وما يهمنا في الجدول العمود الأول الذي يمثل الفئات للظاهرة والعمود الثاني الذي يمثل التكرار لكل فئة.

3–6 تكوين الجدول التكراري المزدوج باستعمال البرنامج (SPSS)

تم استعراض كيفية تكوين الجدول التكراري المزدوج للبيانات الوصفية والكمية في المباحث السابقة يدويا، وفي هذا المبحث سوف نوضح كيفية استعمال البرنامج (SPSS) لإعداد الجدول التكراري المزدوج لهذه البيانات بنوعيها الوصفي والكمي. وكما في الأمثلة الاتية:

3-6-1 الجدول التكراري المزدوج للبيانات الوصفية (النوعية)

يعد اعداد الجدول التكراري المزدوج للظواهر الوصفية سهلاً ومباشراً لان فئات (صفات) كل ظاهرة تكون واضحة ومعروفة، ولتوضيح ذلك نتبع الخطوات الاتية كما في المثال الاتي:

<u>مثال 7</u>

البيانات الاتية تمثل حالة الاصابة بمرض معين مع مستوى الهيموجلوبين في الدم لـ 20 شخص. والمطلوب تلخيص هذه البيانات في جدول تكراري مزدوج.

جدول (3-17) البيانات التي تمثل مستوى الإصابة مع مستوى الهيموجلوبين لـ 20 شخص

حالة الاصابة	بسيطة	متوسطة	شديدة	شديدة	شديدة	بسيطة	بسيطة	غير مريض	غير مريض	متوسطة
مستوى الهيموجلوبين	جيد جدا	مقبول	منخفض	منخفض	مقبول	مقبول	جيد جدا	جيد جدا	جيد جدا	جيد جدا
حالة الإصابة	متوسطة	متوسطة	شديدة	متوسطة	بسيطة	متوسطة	غير مريض	بسيطة	متوسطة	غير مريض
مستوى الهيموجلوبين	منخفض	جيد	جيد	منخفض	مقبول	جيد جدا	جيد	جيد جدا	جيد جدا	مقبول

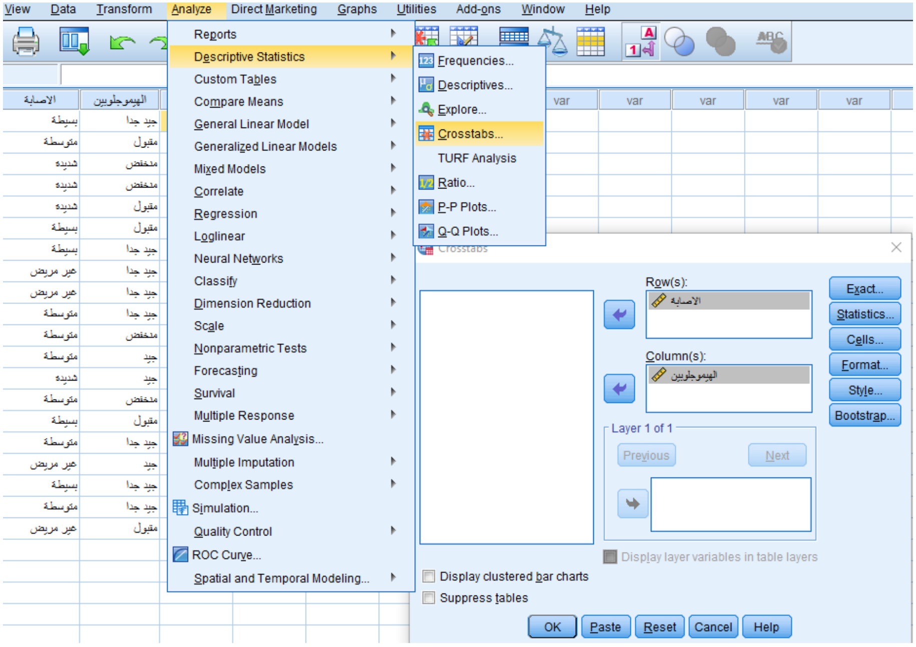
<u>الحل:</u>

أولا: إدخال البيانات الى محرر البيانات بعد تسميتها في صفحة محرر المتغيرات، كما في الشكل (3-12) الاتي:

	الاصابة	الهيموجلوبين	var	var
1	بسيطة	جيد جدا		
2	متوسطة	مقبول		
3	شديدة	منخفض		
4	شديدة	منخفض		
5	شديدة	مقبول		
6	بسيطة	مقبول		
7	بسيطة	جيد جدا		
8	غير مريض	جيد جدا		
9	غير مريض	جيد جدا		
10	متوسطة	جيد جدا		
11	متوسطة	منخفض		
12	متوسطة	جيد		
13	شديدة	جيد		
14	متوسطة	منخفض		
15	بسيطة	مقبول		
16	متوسطة	جيد جدا		
17	غير مريض	جيد		
18	بسيطة	جيد جدا		
19	متوسطة	جيد جدا		
20	غير مريض	مقبول		

الشكل (3-12) البيانات الوصفية لحالة الإصابة ومستوى الهيموجلوبين

ثانيا: بعد ادخال البيانات من قائمة الأوامر، نختار الامر (Analyze) ومنه نختار (Descriptive Statistics) ومن ثم اختيار الامر الفرعي (Crosstabs) للحصول على المربع الحواري كما في الشكل (3-13) الاتي:



الشكل (3–13) المربع الحواري الذي يمثل الجدول التكراري

ثالثا: تحويل المتغير الإصابة الى مربع (Rows) الذي يمثل الصفوف ومن ثم تحويل المتغير الهيموجلوبين الى مربع (Columns) الذي يمثل الاعمدة كما في الشكل (3–13) وبعد اختيار الامر (OK) نحصل على الجدول التكراري المزدوج الذي يمثل المتغيرين كما في الجدول (3–18) الاتي:

جدول (3-17) يمثل الجدول التكراري المزدوج لحالة الإصابة ومستوى الهيموجلوبين لمجموعة من الاشخاص

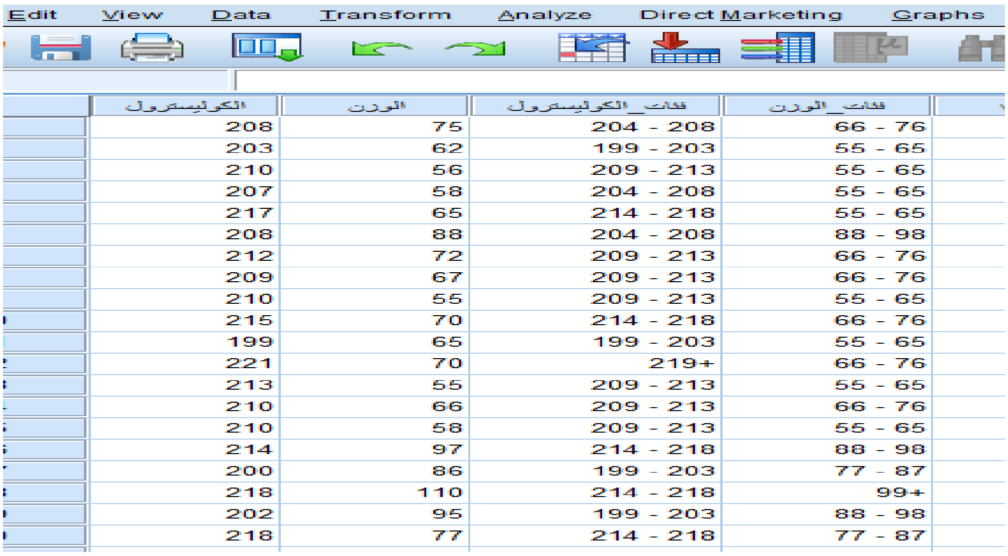
		الهيموجلوبين				Total
		منخفض	مقبول	جيد	جيد جدا	
الاصابة	غير مريض	0	1	1	2	**4**
	بسيطة	0	2	0	3	**5**
	متوسطة	2	1	1	3	**7**
	شديدة	2	1	1	0	**4**
Total		**4**	**5**	**3**	**8**	**20**

3-6-2 الجدول التكراري المزدوج للبيانات الكمية باستعمال البرنامج (SPSS)

يختلف تكوين الجدول التكراري المزدوج للبيانات الكمية عنه في حالة البيانات الوصفية، لان الفئات غير معروفة، كما في البيانات الوصفية لذلك يتطلب اتباع الخطوات الاتية عبر تطبيق المثال الاتي:

<u>**مثال 8:**</u> استعمل البيانات في المثال (3) لتكوين جدول تكراري مزدوج يمثل الظاهرتين الوزن ومستوى الكولسترول باستعمال البرنامج (SPSS).

أولا: الخطوة الأولى تكوين الفئات لكل متغير على حدة باستعمال الخطوات نفسها في حالة الجدول التكراري البسيط ومنها نحصل على متغيرين جديدين يمثلان الفئات لكل متغير كما في الشكل (3-14) الاتي:

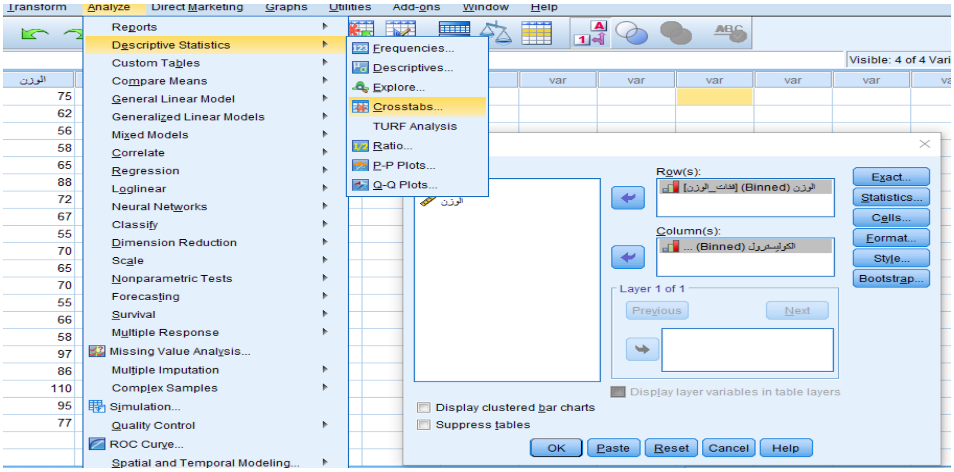


الكوليسترول	الوزن	فئات_الكوليسترول	فئات_الوزن
208	75	204 - 208	66 - 76
203	62	199 - 203	55 - 65
210	56	209 - 213	55 - 65
207	58	204 - 208	55 - 65
217	65	214 - 218	55 - 65
208	88	204 - 208	88 - 98
212	72	209 - 213	66 - 76
209	67	209 - 213	66 - 76
210	55	209 - 213	55 - 65
215	70	214 - 218	66 - 76
199	65	199 - 203	55 - 65
221	70	219+	66 - 76
213	55	209 - 213	55 - 65
210	66	209 - 213	66 - 76
210	58	209 - 213	55 - 65
214	97	214 - 218	88 - 98
200	86	199 - 203	77 - 87
218	110	214 - 218	99+
202	95	199 - 203	88 - 98
218	77	214 - 218	77 - 87

الشكل (3-14) المتغيرات التي تمثل فئات المتغيرين

ثانيا: نختار من الامر (Analyze) الامر الفرعي (Descriptive Statistics) ومنه نختار الامر (Crosstab) ثم نحصل على المربع الحواري الذي يمثل الصفوف

والاعمدة في الجدول التكراري المزدوج ثم يحول أحد المتغيرات الذي يمثل الفئات وليس المتغيرات الاصلية الى الصفوف والأخر الى الاعمدة كما في الشكل (3-15) الاتي:



الشكل (3-15) المربع الحواري الذي يمثل الجدول التكراري المزدوج

ثالثا: اختيار الامر (OK) للحصول على الجدول التكراري المزدوج الذي يمثل المتغيرين كما في الجدول (3-18) الاتي:

الجدول (3-18) الجدول التكراري المزدوج للمتغيرين الكوليسترول والوزن

		الكوليسترول					
		199-203	204-208	209- 213	214- 218	219-224	**Total**
الوزن	**55 - 65**	2	1	4	1	0	**8**
	66 - 76	0	1	3	1	1	**6**
	77 - 87	1	0	0	1	0	**2**
	88 - 98	1	1	0	1	0	**3**
	99-110	0	0	0	1	0	**1**
Total		**4**	**3**	**7**	**5**	**1**	**20**

وعند مقارنة هذا الجدول مع الجدول (3-13) تجد انها كانت النتائج نفسها.

تمارين الفصل الثالث

1- البيانات الاتية تمثل الوقت المصروف (دقيقة) من الطبيب مع المريض في اثناء مراجعته في العيادة الخارجية لعدد 28 مريض احدى المستشفيات الخاصة:

10، 12، 18، 22، 24، 10، 8، 11، 28، 9، 14، 27، 19، 14، 26، 15، 13، 23، 7، 9، 21، 20، 20، 18، 19، 12، 14، 10.

والمطلوب: تكوين جدول التوزيع التكراري المناسب لهذه البيانات باستعمال البرنامج (SPSS).

2- البيانات الاتية تمثل درجات مجموعة طلبة في أحد امتحانات مادة الإحصاء الحيوي. **والمطلوب:** تكوين جدول تكراري لهذه البيانات باستعمال البرنامج (SPSS).

79 72 73 66 61 63 56 72 74 87 69 85 63 53 66 64 59 47
64 70 69 55 45 75 71 60 57 71 79 71 73 62 68 65 76 55
52 70 62 52 78 44 80 68 65 68 69 63.

3- البيانات الاتية تمثل الوزن كصفة لمجموعة من المرضى. **والمطلوب:** هل يمكن وصف هذه البيانات بجدول تكراري باستعمال البرنامج (SPSS).

ثقيل، متوسط، خفيف، ثقيل، متوسط، خفيف جدا، خفيف، متوسط، ثقيل جدا، خفيف، متوسط، ثقيل، خفيف جدا، متوسط، ثقيل جدا، ثقيل، خفيف، متوسط، خفيف جدا، متوسط، ثقيل جدا، متوسط، خفيف.

4- البيانات الاتية تمثل عدد العمليات الجراحية التي اجريت في أحد المستشفيات حسب أيام الأسبوع، ماذا يسمى هذا التصنيف للبيانات؟

الأيام	السبت	الاحد	الاثنين	الثلاثاء	الاربعاء	الخميس	الجمعة
عدد العمليات	12	16	14	10	18	9	2

5- البيانات الاتية تمثل الوزن (كغم) والعمر (أيام)، **والمطلوب:** تنظيم هذه البيانات في جدول تكراري مزدوج واحد.

ت	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
الوزن	3.8	3.1	4.3	5.2	4.7	2.2	4.7	3.2	2.7	4.9
العمر	5	3	4	7	5	2	6	3	2	7
ت	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
الوزن	2.2	3.5	2.5	5.7	2.1	3.9	4.2	2.3	4.8	4.2
العمر	2	4	1	8	3	5	6	2	7	9
ت	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**
الوزن	2.9	5.3	5.6	3.9	4.3	4.1	2.8	3.5	5.1	4.0
العمر	4	6	7	5	7	8	2	3	9	8

6- قام مدير أحد المستشفيات بتسجيل حالة المريض بعد اجراء العملية الجراحية التي أجريت له بالمستشفى وعدد الأيام بعد اجراء العملية لمجموعة من المرضى كما في الجدول الاتي، **والمطلوب:** تنظيم هذه البيانات في جدول تكراري واحد.

ت	1	2	3	4	5	6	7	8	9	10
الحالة	متوسط	جيد	سيء	مقبول	جيد جدا	جيد	متوسط	مقبول	سيء	جيد
الايام	15	13	14	17	15	12	6	13	10	7
ت	11	12	13	14	15	16	17	18	19	20
الحالة	جيد	مقبول	متوسط	جيد جدا	ممتاز	جيد	سيء	جيد	مقبول	جيد
العمر	12	9	11	18	13	15	16	12	17	9

7- عرف التمثيل التكراري للبيانات؟

8- عدد أنواع الجداول التي يمكن ان تستعمل لوصف البيانات وتلخيصها.

الفصل الرابع

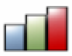
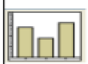
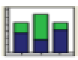
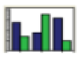
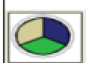
العرض البياني (الاشكال البيانية)

Graphical Presentation

4-1 تمهيد

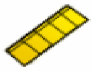
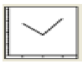
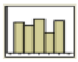
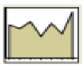
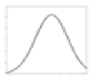
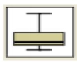
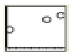
الخطوة الثانية في وصف البيانات وتلخيصها لإعطاء صورة أولية عن البيانات هي استعمال الاشكال البيانية او الصور لتوضيح طبيعة البيانات، والتي تكون على نوعين، اما بيانات وصفية (نوعية) (Qualitative data) او بيانات كمية (Quantitative data)، ومن ثم فان طرق عرض البيانات او الاشكال البيانية تعتمد على نوع البيانات والجدول (4-1) يلخص الاشكال المناسبة لنوع البيانات الوصفية (النوعية):

جدول (4-1) الاشكال المناسبة لنوع البينات الوصفية

الإسمى أو الترتيبي Nominal or Ordinal			
 أو 			
الأعمدة البيانية		Bars	تستعمل الأعمدة البيانية لعرض بيانات وصفية او عرض جدول تكراري بسيط لظاهرة واحدة
الأعمدة البيانية المجزأة		Sub-divided bars	تستعمل الأعمدة البيانية المجزأة لعرض بيانات كلية مقسمة الى اجزائها المتكاملة
الأعمدة البيانية المتجاورة		Multiple bars	تستعمل الأعمدة البيانية المجزأة لعرض بيانات متعارضة
الرسوم الدائرية		Pie	تستعمل الرسوم الدائرية لعرض بيانات كلية مقسمة الى اجزائها المتكاملة

اما الاشكال المناسبة للبيانات الكمية فهي كما في الجدول (4-2)

جدول (4-2) الاشكال البيانية المناسبة للبيانات الكمية

الكمي Scale			
			
الخط البيانى		Line	يستعمل الخط البيانى لعرض بيانات كمية لظاهرة مأخوذة على فترات زمنية
المدرج التكراري		Histo-grame	يستعمل المدرج التكرارى لعرض بيانات كمية لظاهرة بعد تحويلها الى فئات متساوية بشكل اعمدة متلاصقة
المضلع التكراري		Poly-gone	يستعمل المضلع التكرارى لعرض بيانات كمية لظاهرة بعد تحويلها الى فئات متساوية بشكل خطوط منكسرة
المنحنى التكراري		Curve	يستعمل المنحنى التكرارى لعرض بيانات كمية لظاهرة بعد تحويلها الى فئات متساوية بشكل خطوط منحنية
الرسم الصندوقى		Box-plot	يستعمل الرسم الصندوقى لعرض بيانات كمية لظاهرة بالاستعانة بالوسيط والربيعان.
شكل الانتشار		Scatter	يستعمل شكل الانتشار لعرض بيانات كمية لظاهرتان مرتبطتان عن طريق رسم نقاط.

4-2 عرض البيانات الوصفية (النوعية) (Qualitative Data presentation)

عادة ما تجمع البيانات الوصفية من المشاهدات اما حسب الزمن او حسب صفة كل مشاهدة او حسب المواقع الجغرافية. وهذه البيانات تكون وصفية، لذلك فان الطرق التي تستعمل لعرض البيانات لظاهرة واحدة هي الاعمدة البيانية المفردة او الدائرة البيانية، اما في حالة ان تكون البيانات لظاهرتين معا فان الاشكال المناسبة هي الاعمدة البيانية المجزأة او الاعمدة البيانية المزدوجة، وهذه البيانات الوصفية (النوعية) اما تكون على

شكل مشاهدات مفردة او على شكل جدول تكراري، وفي الحالتين تستعمل الأمثلة الآتية لتوضيح كيفية رسمها باستعمال برنامج (SPSS).

1-2-4 الاعمدة البيانية البسيطة Bars

لرسم الاعمدة البيانية باستعمال برنامج (SPSS) اما ادخال المشاهدات مباشرة بشكلها الاولي تحت اسم المتغير او الصفة او تحويل المشاهدات الأولية الى جدول تكراري بسيط يمثل الصفات والتكرار لكل متغير (صفه) ومن ثم رسم الاعمدة أو الأشرطة البيانية، والاعمدة البيانية هي مجموعة من المستطيلات متساوية القاعدة إذ تمثل القاعدة الصفات للظاهرة ومختلفة الارتفاع حيث ان ارتفاعها يمثل عدد المشاهدات (التكرار) لكل صفة، وكما موضحة في الأمثلة الاتية:

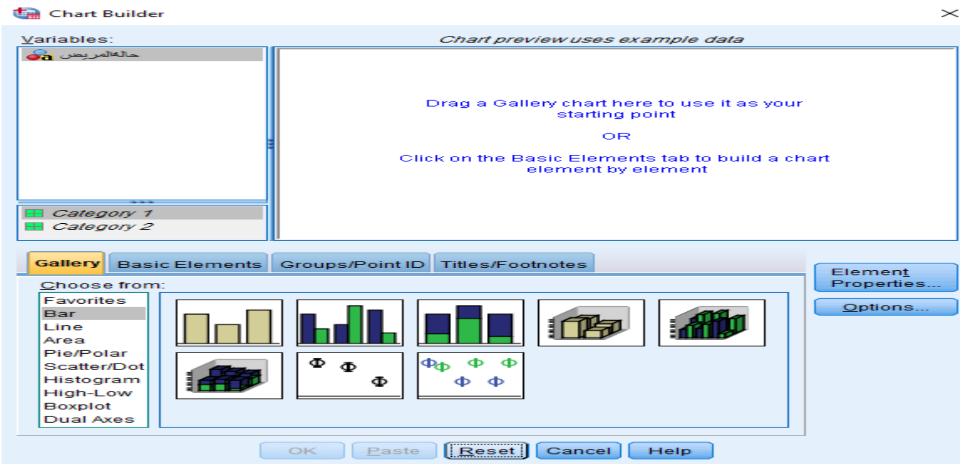
<u>مثال 1</u>

استعمل البيانات في المثال (1) في الفصل الثالث لرسم الاعمدة البيانية في الحالتين من البيانات مباشرة ومن الجدول التكراري البسيط.

أولا: يلاحظ أن البيانات هي صفة تسجل مباشرة ومن ثم ادخالها الى محرر البيانات في برنامج (SPSS) كما في الشكل الاتي:

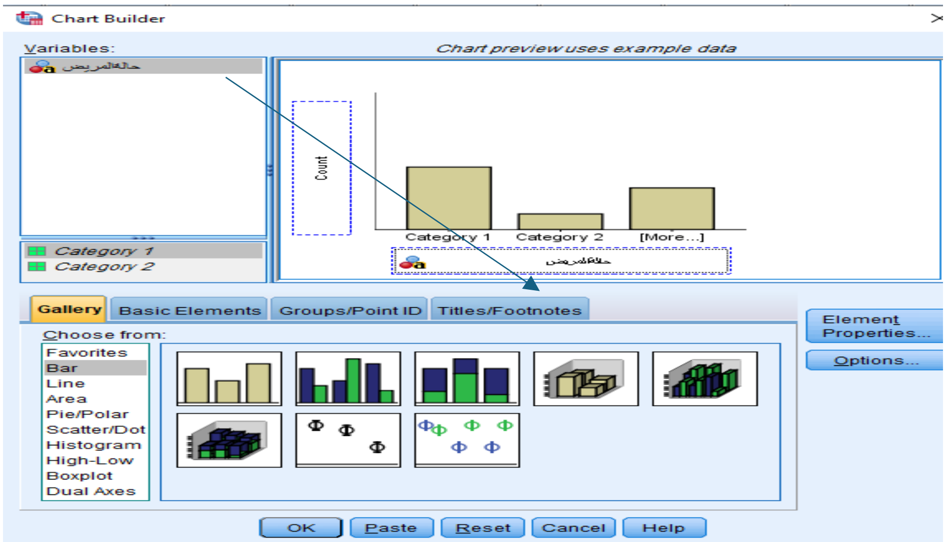
الشكل (4−1) ادخال البيانات في محرر البيانات في برنامج (SPSS)

ثانيا: وبعدها نختار من قائمة الأوامر الامر (Graph) ومن ثم منه نختار باني الاشكال (Chart Builder) فنحصل على مربع الحوار في الشكل (4−2).



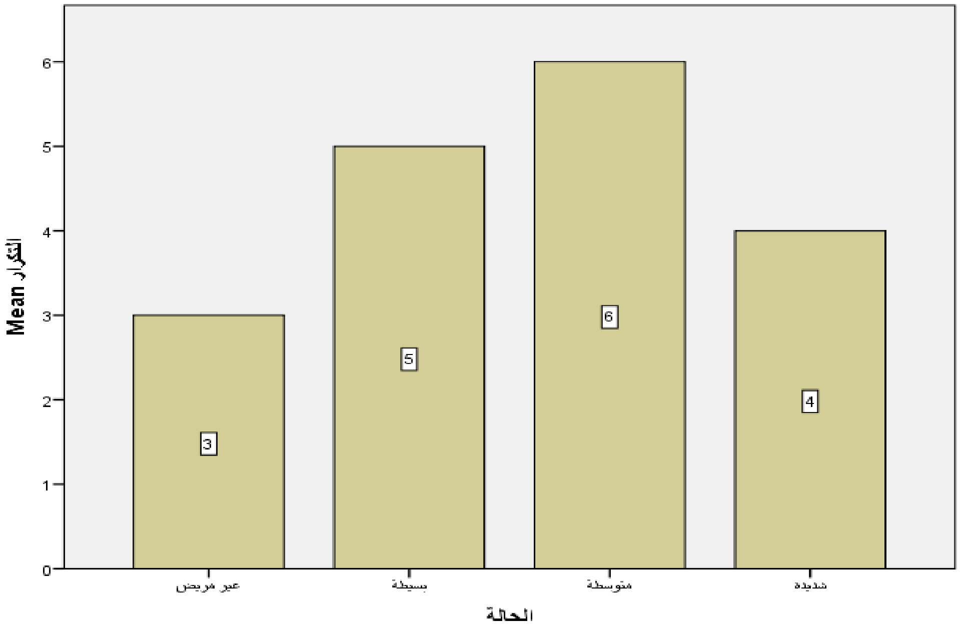
الشكل (4-2) مربع الحوار لباني الاشكال (Chart builder)

ثالثا: ومن القائمة في اقصى اليسار نختار (Bar) فتظهر لنا الاشكال في الوسط نختار منها الشكل الأول بالنقر عليه مرتين لينقل الى المربع الأعلى كما في الشكل (4-3).



الشكل (4-3) اختيار الشكل المطلوب ومن ثم تحويله الى الأعلى

رابعا: ثم ننقل اسم المتغير الذي يمثل حالة المريض الى مربع أسفل الشكل الذي يمثل المحور الافقي التي تشمل صفات الظاهرة ونلاحظ بانه تم تنشيط كلمة (OK) للنقر عليها لاحقا فنحصل على الشكل البياني الذي يمثل الاعمدة البيانية والتي تمثل صفات الظاهرة الأربعة (غير مريض، بسيطة، متوسطة، شديدة).



الشكل (4-4) الاعمدة البيانية التي تمثل الظاهرة حالة المريض

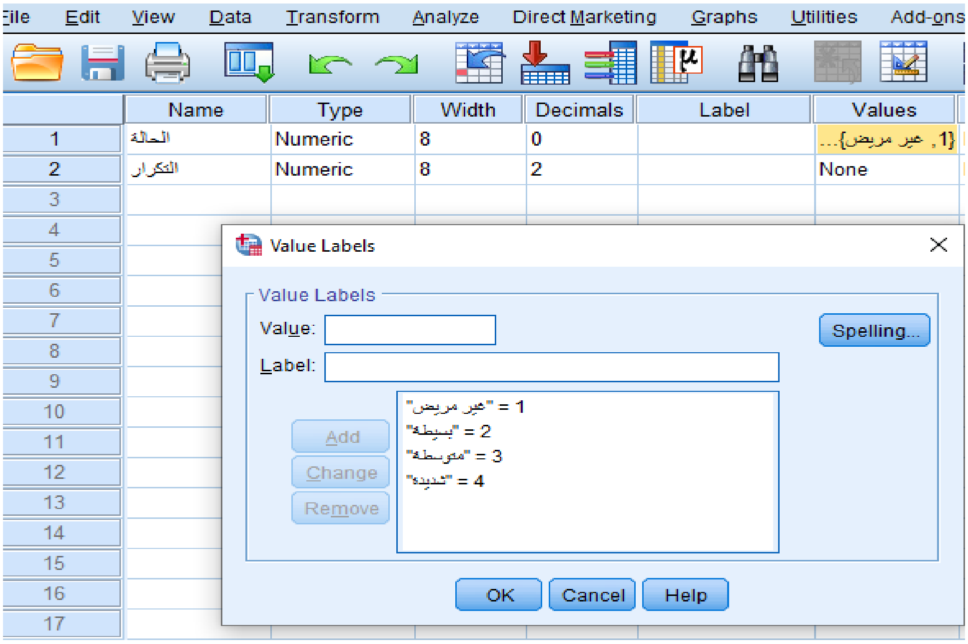
<u>مثال 2</u>

عادة عندما يكون حجم العينة او عدد المشاهدات كبيراً يفضل عرض البيانات في جدول تكراري بسيط يمثل صفات الظاهرة وتكرار كل صفة، ولتمثيل مثل هذه الظاهرة باستعمال الاعمدة البيانية او الدائرة، نتبع الخطوات الاتية من خلال تحويل بيانات المثال (1) اعلاه الى جدول تكراري، كما في الجدول الاتي:

جدول (4-3) حالة المريض حسب شدة المرض.

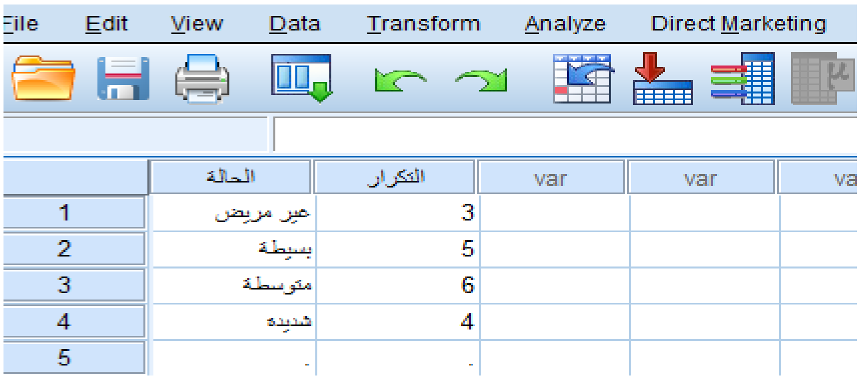
حالة المريض	غير مريض	بسيطة	متوسطة	شديدة
التكرار	3	5	6	4

أولا: ادخال البيانات الى محرر البيانات في البرنامج، وقبل ادخال البيانات تسمى المتغيرات، وهنا سيكون لدينا متغيران الأول يمثل حالة المريض وهو متغير نوعي، ومتغير التكرار، او عدد المرضى في كل صفة، وهو متغير كمي كما الشكل (4-7):



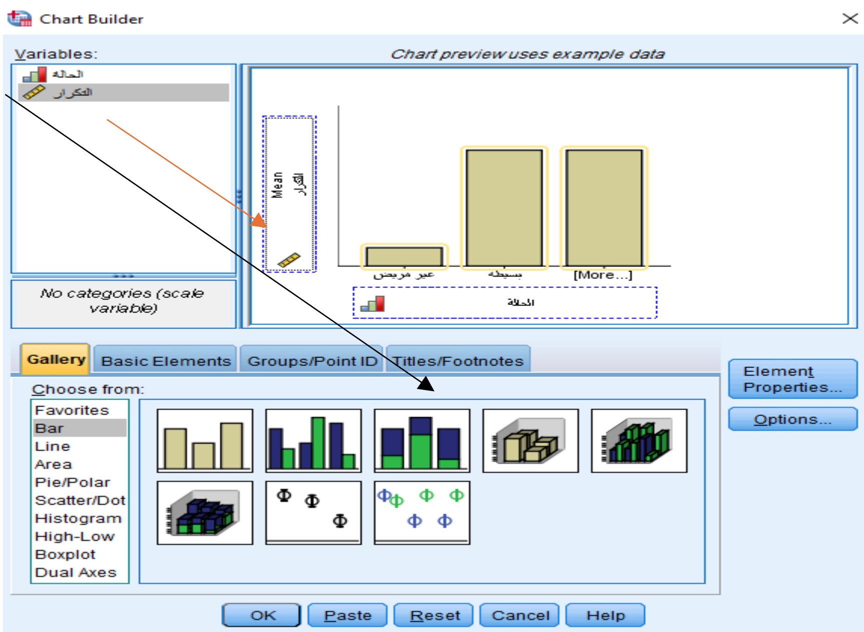
الشكل (4-7) تسمية المتغيرات قبل ادخال البيانات على صفحة المحرر

ثانيا: وبعد ذلك الانتقال الى محرر البيانات لإدخال البيانات على عمودين: الأول: يمثل الحالة والثاني: يمثل التكرار كما في الشكل (4-8)



الشكل (4–8) ادخال البيانات في محرر البيانات

ثالثا: نختار الامر (Graph) ومنه نختار الامر الفرعي (Chart Builder) ومن مربع الحوار نختار الشكل (Bar)، ومن ثم نختار الاعمدة البسيطة، فتتحول الى المربع الأعلى، ثم نسحب المتغير الحالة الى المحور الافقي، ومن ثم نسحب المتغير التكرار الى المحور العمودي كما في الشكل (4–9)

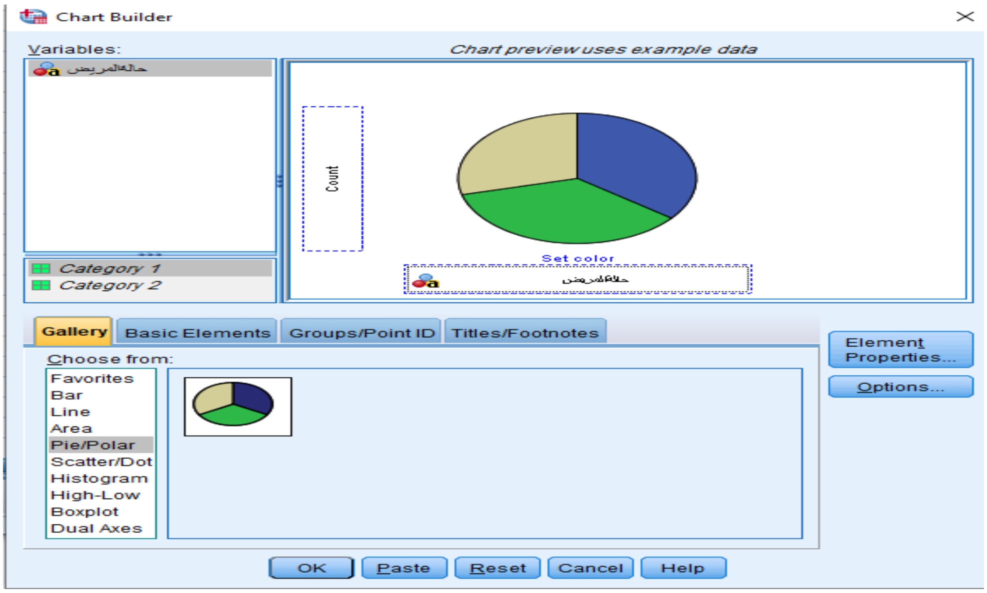


الشكل (4–9) مربع حوار رسم الاعمدة البيانية البسيطة

رابعا: وبعد النقر على كلمة (OK) نحصل على الشكل نفسه (4-4).

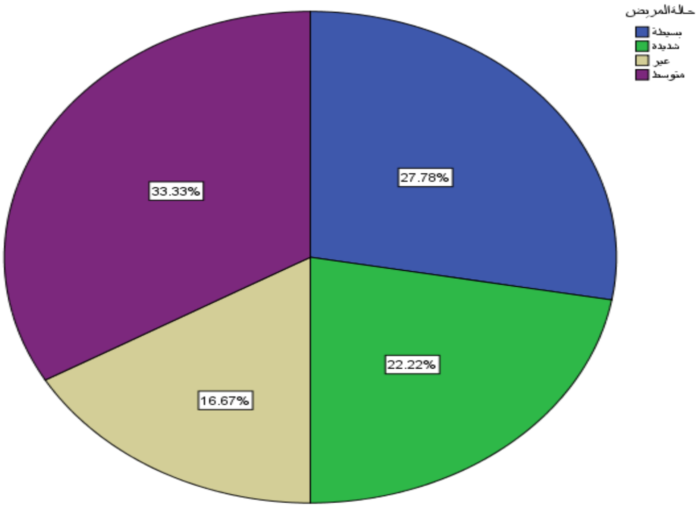
2-2-4 الدائرة البيانية Pie Chart

يمكن رسم الدائرة البيانية للظاهرة بنفس الخطوات التي استعملت لرسم الأشرطة البيانية لأنها تناسب الظاهرة المعروضة، والفرق هو في مربع الحوار في الشكل (2-4) نختار بدلا من الشكل (Bar) الشكل (Pie) وبالطريقة نفسها نختار الشكل الدائرة وننقر عليه مرتين، ومن ثم ننقل الظاهرة الى المربع أسفل الدائرة في المربع الأعلى كما في الشكل (5-4)



الشكل (5-4) مربع الحوار لرسم الدائرة البيانية

وبعد نقر كلمة (OK) يتكون رسم الدائرة البيانية ولإظهار قيم المتغير على الرسم ننقر على الرسم مرتين، ونستعمل مربع الحوار الذي نحصل وهو (Chart Editor) لإدخال التغييرات المطلوبة لنحصل على الشكل (6-4).



الشكل (4−6) الدائرة البيانية التي تمثل نسبة كل صفة في الظاهرة

4−2−3 الاعمدة المجزأة او المزدوجة Sub−divided or Multiple Bars

عند تمثيل ظاهرتين نوعيتين فإننا نستعمل اما الاعمدة المزدوجة، او الاعمدة المجزأة، وعادة ما تكون البيانات على شكل جدول تكراري للظاهرتين، ولتوضيح ذلك نأخذ المثال الاتي:

<u>مثال 3</u>

البيانات الاتية تمثل معدل الولادات والوفيات في عدد من المحافظات لاحدى السنوات لكل ألف من السكان، والمطلوب تمثيل هذه البيانات باستعمال برنامج (SPSS) والاشكال المناسبة.

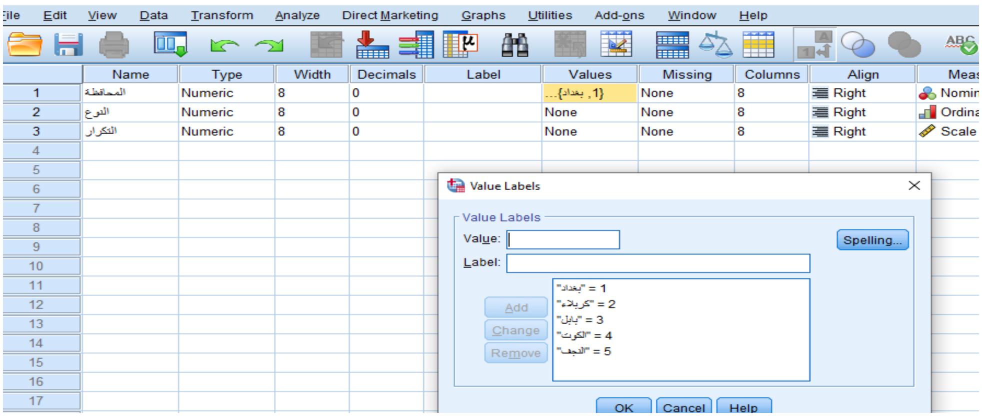
جدول رقم (4-4) معدلات الولادات والوفيات حسب المحافظات

المحافظة	بغداد	كربلاء	بابل	واسط	نجف
عدد الولادات	46	34	26	28	22
عدد الوفيات	20	16	14	16	10

الحل:

لتمثيل هذه البيانات باستعمال النوعين من الاعمدة البيانية باستعمال برنامج (SPSS) نتبع الخطوات الاتية:

1- تعريف المتغيرات واسمائها على محرر المتغيرات أولاً، إذ يكون لدينا ثلاثة متغيرات الأول: يمثل صفات الظاهرة وهو متغير نوعي يعرف باستعمال مستوى المتغير مع إعطاء رقم لكل صفة من صفات المتغير، كما لاحظنا عند رسم الاعمدة البسيطة في حالة الجدول التكراري، والثاني: متغير نوعي يمثل عدد المتغيرات ويأخذ الرقم 1 للمتغير الأول و2 للمتغير الثاني اما الثالث: فيكون متغير التكرار كما في الشكل (4-10).

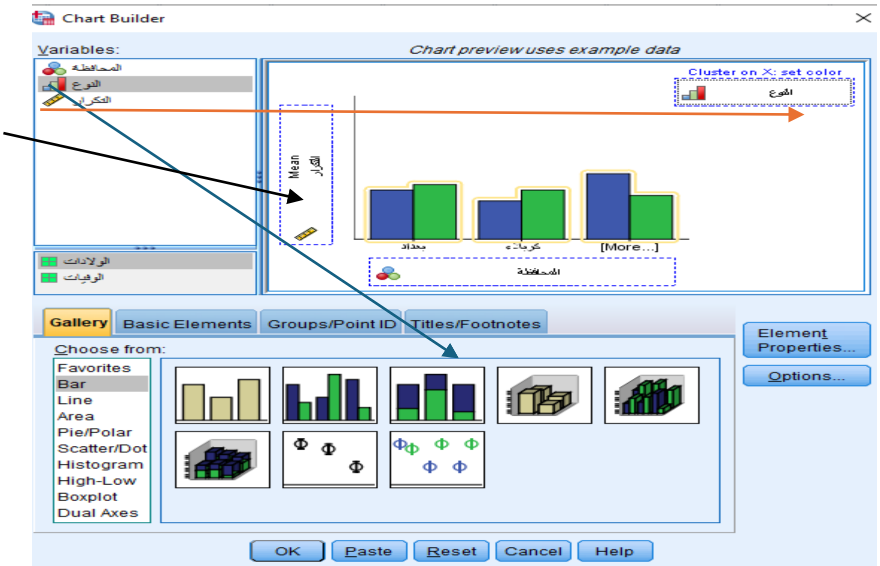


الشكل (4-10) تعريف المتغيرات على صفحة محرر المتغيرات

2- ندخل البيانات في صفحة محرر البيانات كما في الشكل (4-11)

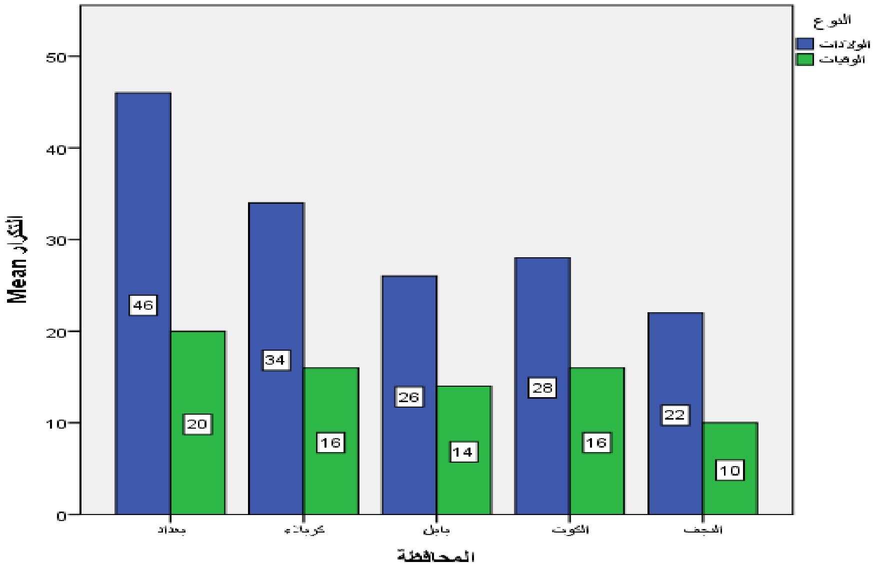
الشكل (4-11) ادخال البيانات على محرر البيانات

3- نختار من قائمة الأوامر الامر (Graph) ثم نختار منه الامر الفرعي (Chart Builder) ومن مربع الحوار نختار كلمة (Bar) ومن الاشكال التي تظهر على مربع الحوار نختار أحد الشكلين اما الاعمدة المتجاورة او الاعمدة المجزأة، فيظهر الشكل في المربع الأعلى بثلاثة مربعات على المحور الافقي يمثل المحافظة وعلى المحور العمودي يمثل التكرار وفي الأعلى لليمين يمثل النوع كما في الشكل (4-12).



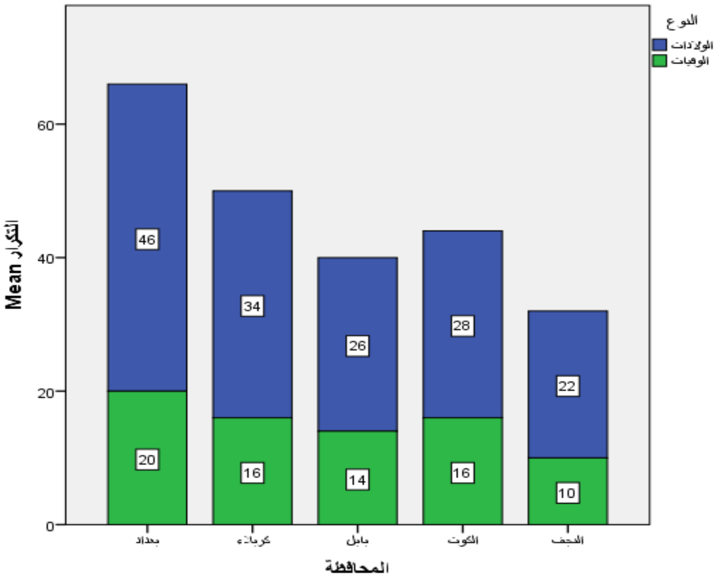
الشكل (4-12) مربع الحوار لاختيار الشكل المناسب

4- بعد النقر على كلمة (OK) يظهر لنا الشكل (4-13) المطلوب على صفحة النتائج ولإجراء اية تغييرات ننقر على الشكل مرتين فيظهر لنا مربع حوار (Chart Editor) يسمح لنا بإجراء التغييرات المطلوبة.



الشكل (4-13) الاعمدة المزدوجة التي تمثل معدل الولادات والوفيات لمجموعة من المحافظات

5- وباتباع الخطوات نفسها واختيار الاعمدة المجزأة من مربع الحوار (Chart Builder) نحصل على الشكل (4-14).



الشكل (4–14) الاعمدة المجزأة لتمثيل معد الولادات والوفيات

هناك تفاصيل أكثر عن كيفية اخراج هذه الاشكال عبر تغيير الألوان للمستطيلات او الخلفية مع بعض الإضافات الأخرى يمكن الاطلاع عليها من المصادر الخاصة بذلك، والمبحث التالي يوضح كيفية تمثيل البيانات الكمية.

4–3 عرض البيانات الكمية Quantitative Data Presentation

عرضنا في الجدول (4–2) الاشكال التي تناسب عرض البيانات الكمية ولتوضيح كيفية عرض كل شكل منها نأخذ الأمثلة الاتية:

4–3–1 المدرج التكراري Histogram

عادة بالعمل اليدوي لرسم المدرج التكراري نحتاج الى تكوين الجدول التكراري البسيط، ومن ثم الرسم ولكن باستعمال برنامج (SPSS) نحتاج الى ادخال البيانات الى محرر البيانات بشكلها الاولي، والبرنامج هو يكون الجدول التكراري ويرسم المدرج التكراري

بدون تدخل منا، وهناك طريقان لرسم المدرج التكراري بعد ادخال البيانات، اما من خلال الامر (Analyz)، ومن ثم اختيار الامر الفرعي (Descriptive Statistic)، ومن ثم اختيار الامر الفرعي (Frequency) يظهر مربع حوار منه نختار بعد نقل المتغير الى المكان المطلوب، ومن مربع الحوار ننقر على كلمة (chart) ومنه نختار رسم المدرج التكراري او عبر الامر (Graph) ومنه نختار الامر (Chart Builder)، ومن ثم نختار الرسم (Histogram) من مربع الحوار ونكمل الرسم، والاثنان تعطيان الشكل نفسه، ولتوضيح هذه الخطوات نأخذ المثال الاتي:

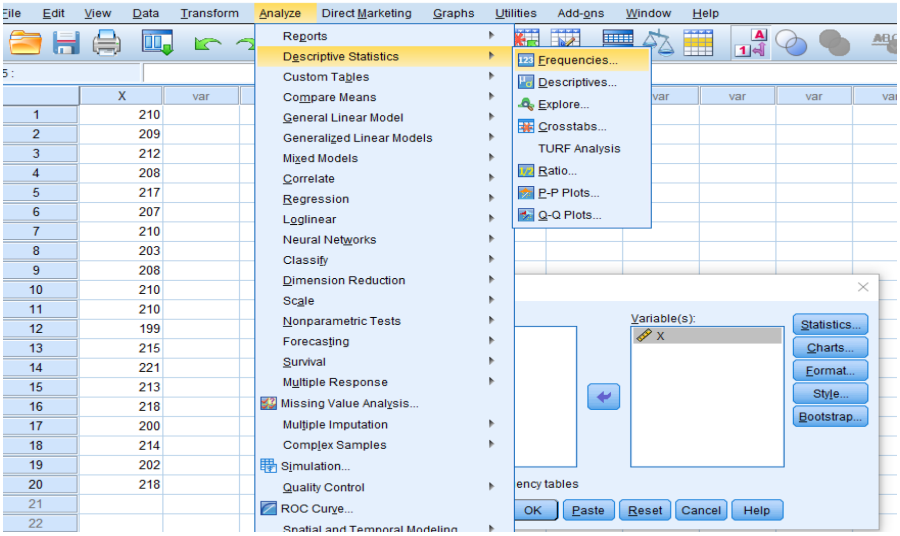
مثال 4

قام طبيب بفحص مستوى الكولسترول لـ (25) مريضا فكانت مقاسة بالملغرام/ مليلتر كالاتي:

210 209 212 208 217 207 210 203 208 210
210 199 215 221 213 218 202 218 200 214
205 204 203 220 198

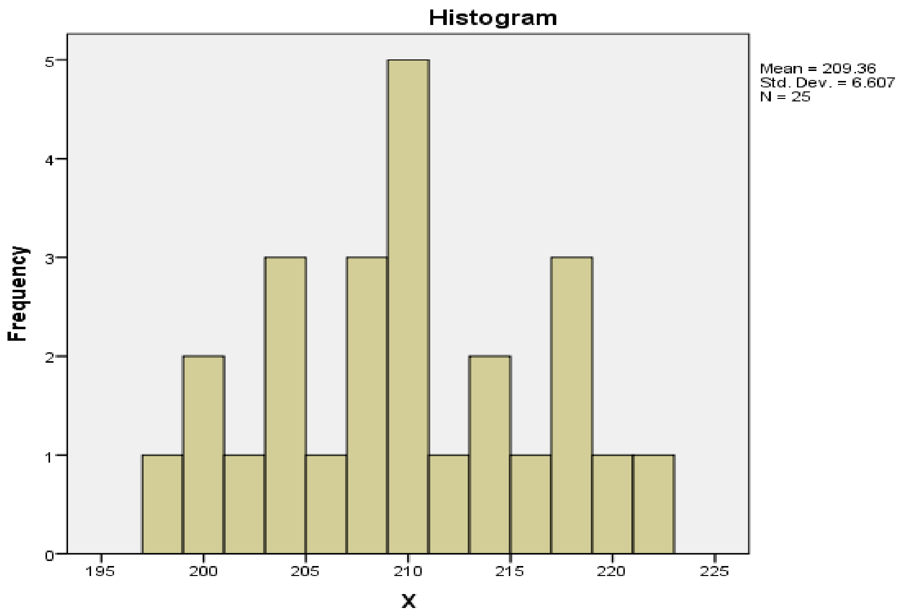
الحل:

1- ادخال البيانات واتباع المسار أعلاه لاختيار الرسم مباشرة كما في الشكل (4-15)



الشكل (4-15) ادخال البيانات واختيار المسار

2- نختار من مربع الحوار (Frequencies) وننقر على كلمة (Charts) ومن مربع الحوار الذي يظهر لنا نختار رسم المدرج التكراري (Histogram) فنحصل على الشكل (4-16).



الشكل 4−16 المدرج التكراري

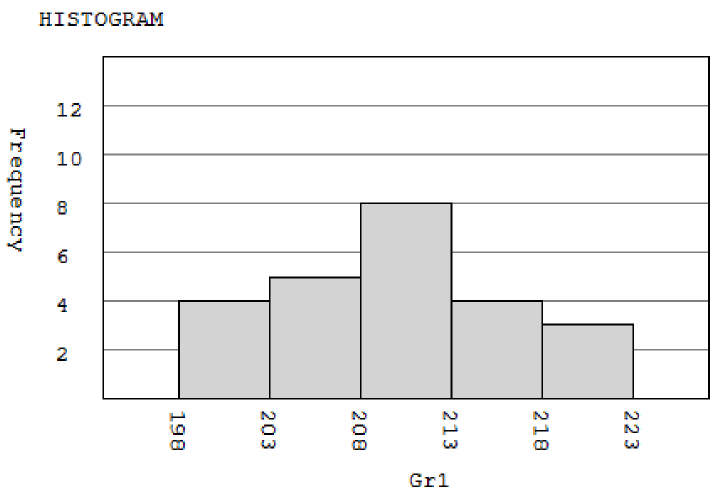
الملاحظ من الشكل (4–16) لا نجده يمثل المدرج التكراري بالشكل المتعارف عليه لان المدرج التكراري هو مستطيلات متجاورة طول قاعدة كل منها يساوي طول الفئة وارتفاعها يمثل تكرار الفئة، والشكل أعلاه لا يمثل هذه الصفات، وانما هو اشرطة بيانية للبيانات وكأنها بيانات وصفية وللمقارنة تم استعمال برنامج (Statext) لرسم المدرج التكراري لنفس البيانات، وسنلاحظ الفرق بين الاثنين عبر ملاحظة الجدول التكراري الذي اوجدناه باستعمال البرنامج (Statext) كما في الجدول (4–5).

جدول (4–5) الجدول التكراري للبيانات

Interval	Freq	R.Freq	Midpoint	C.Freq	C.R.Freq
198 <= X < 203	4	0.1667	200.5	4	0.1667
203 <= X < 208	5	0.2083	205.5	9	0.3750
208 <= X < 213	8	0.3333	210.5	17	0.7083
213 <= X < 218	4	0.1667	215.5	21	0.8750
218 <= X < 223	3	0.1250	220.5	24	1.0000
Sum	24	1.0000			

R.Freq = Relative Frequency
C.Freq = Cumulative Frequency
C.R.Freq = Cumulative Relative Frequency

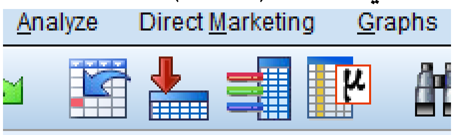
ولرسم المدرج التكراري باستعمال البرنامج نفسه نحصل على المدرج التكراري كما في الشكل (4–17) الاتي:



الشكل (4–17) المدرج التكراري للجدول التكراري (4–5)

2-3-4 المضلع التكراري Frequency Polygon

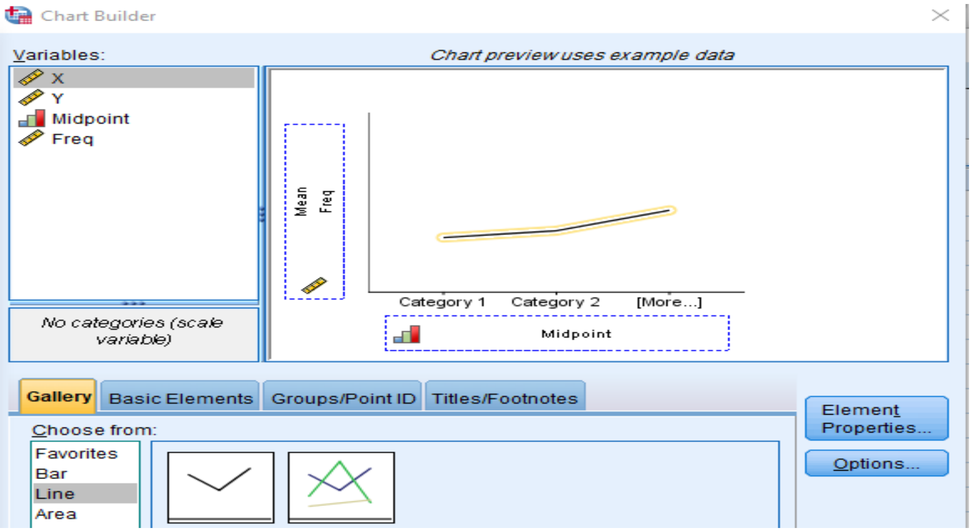
هو مجموعة من المستقيمات التي تصل بين النقاط المتكونة من التقاء مراكز الفئات على المحور الافقي والتكرار ممثلا على المحور العمودي، ويستعمل لعرض البيانات الكمية لظاهرة معينة بعد تحويلها الى فئات متساوية وعلى شكل خطوط منكسرة، ولرسم المضلع التكراري باستعمال برنامج (SPSS) نقوم بتكوين جدول تكراري للبيانات في الجدول (4-5) يتضمن مراكز الفئات (Midpoint) والتكرارات (Freq) كما في العمود الرابع والعمود الثاني، ومن ثم ننقل هذه البيانات الى محرر البيانات في برنامج (SPSS) في عمودين كما في الشكل (4-18).



Midpoint	Freq	var
200.5	4	
205.5	5	
210.5	8	
215.5	4	
220.5	3	

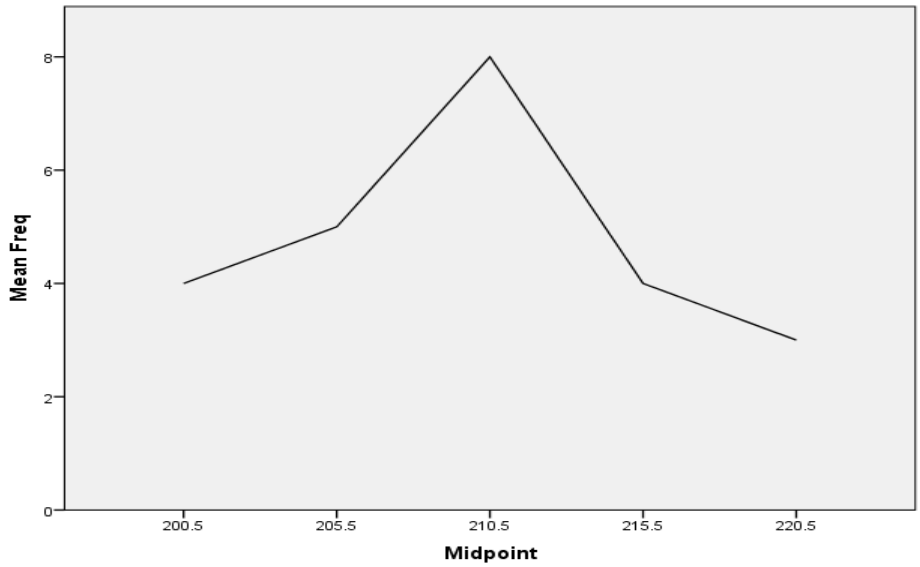
الشكل (4-18) مراكز الفئات مع التكرار

وبعد ذلك نختار الامر (Graph) من قائمة الأوامر الرئيسية، ومنه نختار الامر الفرعي (Chart Builder) ومنه نختار الامر (Line) وبعد ذلك نسحب متغير مراكز الفئـات الى العمـود الافقي والتكرارات الى المحـور العمودي كمـا في الشكل (4-19).



الشكل (4−19) مربع الحوار بناء الرسوم

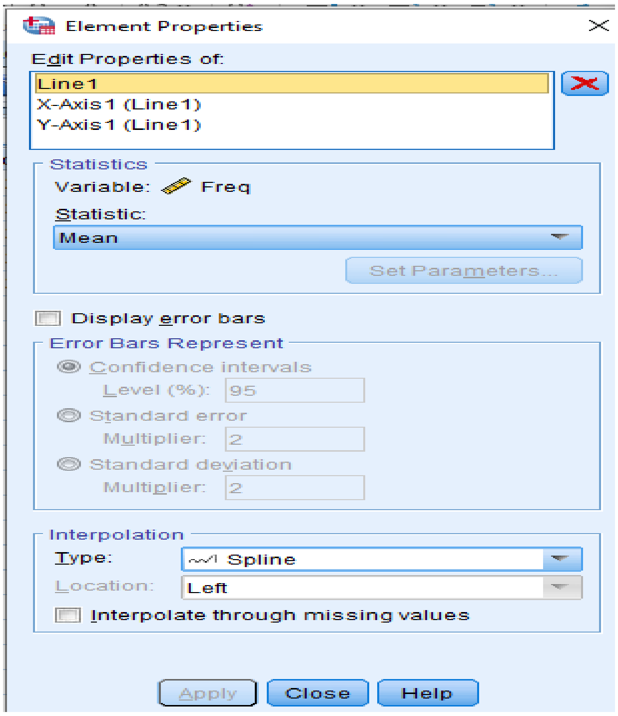
وبعد النقر على (Element Properties) يظهر لنا مربع حوار نختار منه نوع الخط (Stright) وبعد النقر على الامر (OK) نحصل على الشكل النهائي للمضلع التكراري كما في الشكل (4−20)



الشكل (4-20) المضلع التكراري

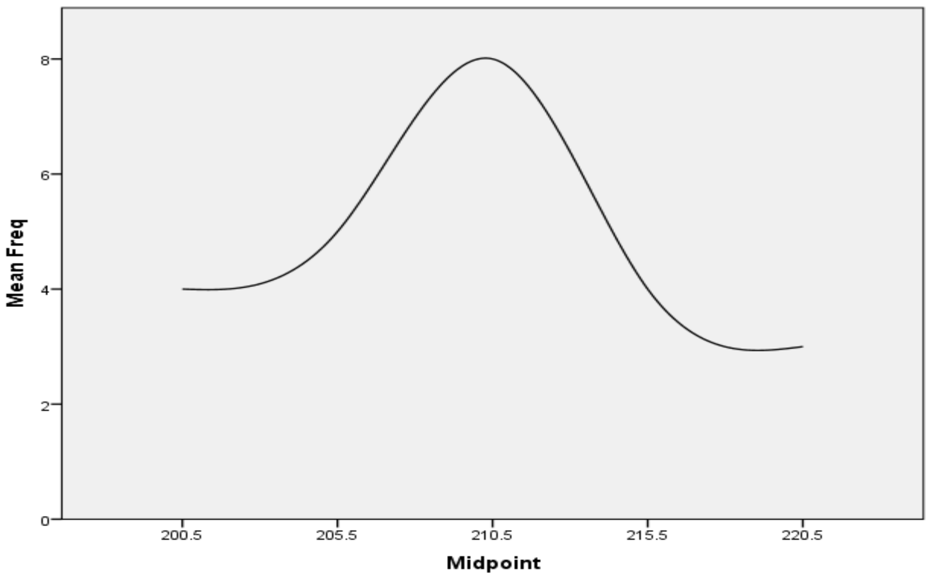
4-3-3 المنحنى التكراري Frequency Curve

هو خط ممهد باليد يصل بين النقاط المتكونة من التقاء مراكز الفئات على المحور الافقي والتكرار ممثلا على المحور العمودي. ويستعمل لعرض البيانات الكمية لظاهرة معينة بعد تحويلها الى فئات متساوية وعلى شكل خط ممهد، ولرسم المنحنى التكراري باستعمال برنامج (SPSS) نقوم بتكوين جدول تكراري للبيانات يتضمن مراكز الفئات (Midpoint) والتكرارات (Freq) كما في الجدول (4-5) العمود الرابع والعمود الثاني، ومن ثم ننقل هذه البيانات الى محرر البيانات في برنامج (SPSS) في عمودين، كما في الشكل (4-18)، ونتبع الخطوات نفسها في رسم المضلع التكراري مع فرق واحد هو اختيار نوع الخط الذي يصل بين النقاط، ففي المضلع التكراري تم اختيار الخط المستقيم (Stright) من المربع الحواري (Element Properties) وعند الرغبة في رسم المنحنى التكراري نختار نوع الخط (Spline) كما في الشكل (4-21).



الشكل (4-21) مربع الحوار خصائص الشكل

وبعد اختيار نوع الخط المطلوب تنشط كلمة (Apply)، وتعني تطبيق هذا الخيار، فننقر عليها كي يصبح هو الخيار، وبعد ذلك النقر على الامر (OK) على مربع الحوار الأصلي، فنحصل على المنحنى التكراري كما في الشكل (4−22).



الشكل (4−22) المنحنى التكراري

4−3−4 المنحنى التكراري المتجمع الصاعد او النازل

Ascending or Descending Cumulative Frequency

وهو خط ممهد يصل بين النقاط المتكونة من التقاء الحدود الدنيا والتكرار المتجمع الصاعد في حال المنحنى المتجمع الصاعد او الحدود العليا والتكرار المتجمع الهابط في حالة المنحنى التكراري المتجمع الهابط ونكتفي برسم المنحنى المتجمع الصاعد ونترك رسم المنحنى المتجمع الهابط للقارئ كتمرين. ولتوضيح كيفية الرسم نستعمل المثال الاتي:

مثال5

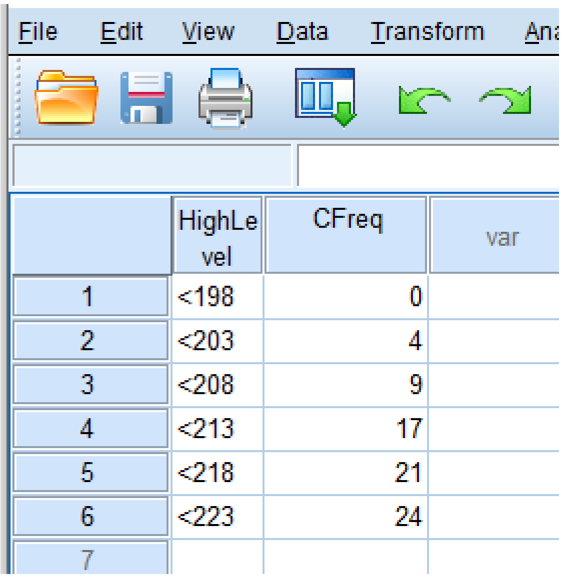
استعمل البيانات في الجدول التكراري (4−5) لرسم المنحنى المتجمع الصاعد.

الحل: تكوين الجدول التكراري المتجمع الصاعد كما في الجدول (4−6).

جدول (4−6) الجدول التكراري المتجمع الصاعد

الحدود العليا للفئات	<198	<203	<208	<213	<218	<223
التكرار المتجمع الصاعد	0	4	9	17	21	24

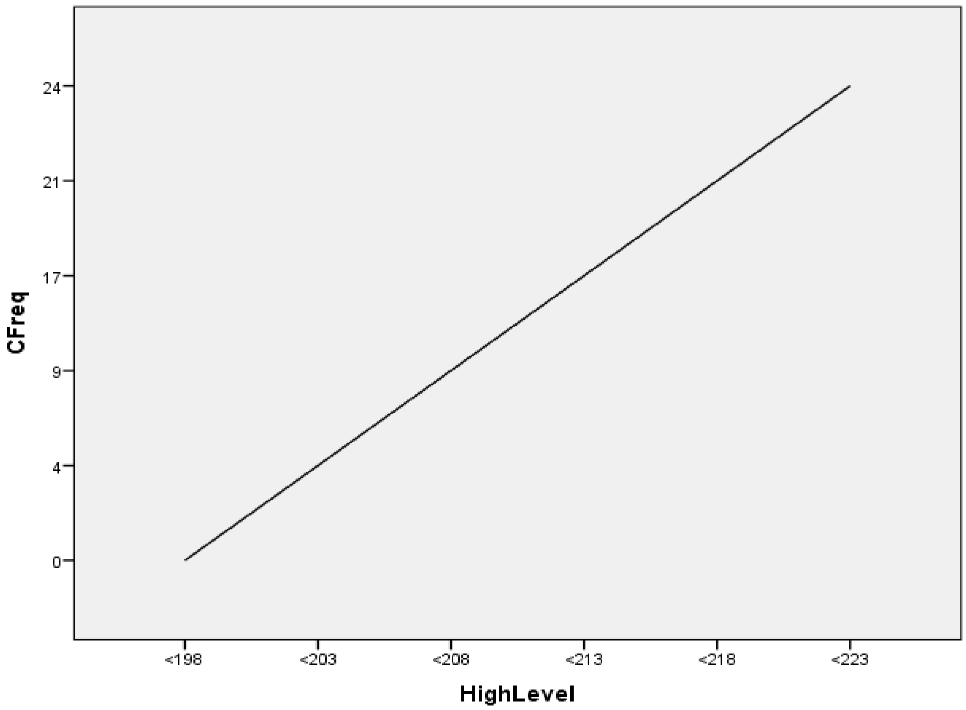
بعد ذلك ادخال البيانات على محرر البيانات في برنامج (SPSS) كما في الشكل (23−4)



	HighLevel	CFreq	var
1	<198	0	
2	<203	4	
3	<208	9	
4	<213	17	
5	<218	21	
6	<223	24	
7			

الشكل (4−23) ادخال البيانات الى محرر البيانات

وبعد ذلك اختيار الامر (Graph) ومنه نختار الامر الفرعي (Chart Builder) ونختار الشكل (Line) ونسحب الحدود العليا للفئات على المحور الافقي والتكرار المتجمع الصاعد على المحور العمودي، ثم ننقر الامر (OK) فنحصل على شكل المنحنى المتجمع الصاعد كما في الشكل (4−24)



الشكل (4-24) المنحنى التكراري المتجمع الصاعد

4-3-5 الخط البياني (المنحنى الزمني) Line Graph

وهو خط يصل النقاط التي تتكون من التقاء الزمن مع التكرار المقابل لكل فترة زمنية، ويستعمل الخط البياني لتمثيل البيانات التي تكون قد قيست على مدد زمنية سنوية، او شهرية، او أسبوعية، او يومية، او أصغر من ذلك من وحدات الزمن، وغالبا ما يمثل التطور الزمني للظاهرة ولتوضيح كيفية رسم الخط البياني باستعمال برنامج (SPSS) نورد المثال الاتي

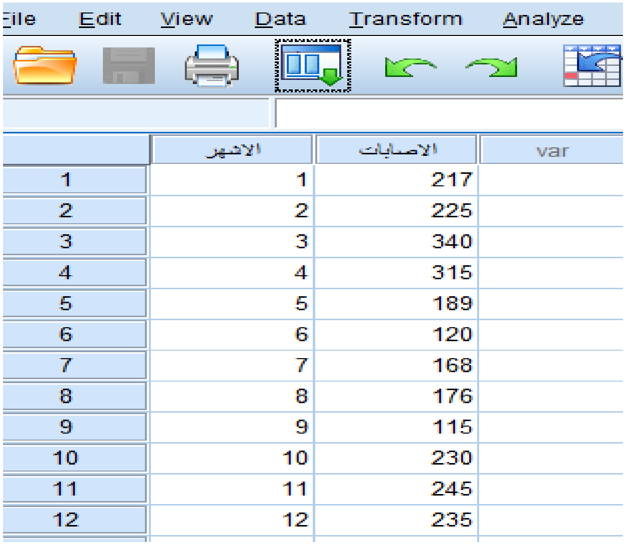
مثال 6 البيانات الاتية تمثل عدد الإصابات بفايروس كوفيد 19 خلال الأشهر من عام 2020 في احدى المحافظات.

جدول (4-7) عدد الإصابات الشهرية حسب الأشهر بفايروس كوفيد-19

الأشهر	ك2	شباط	اذار	نيسان	مايس	حزيران	تموز	اب	ايلول	ت1	ت2	ك1
عدد الاصابات	217	225	340	315	189	120	168	176	115	230	245	235

الحل:

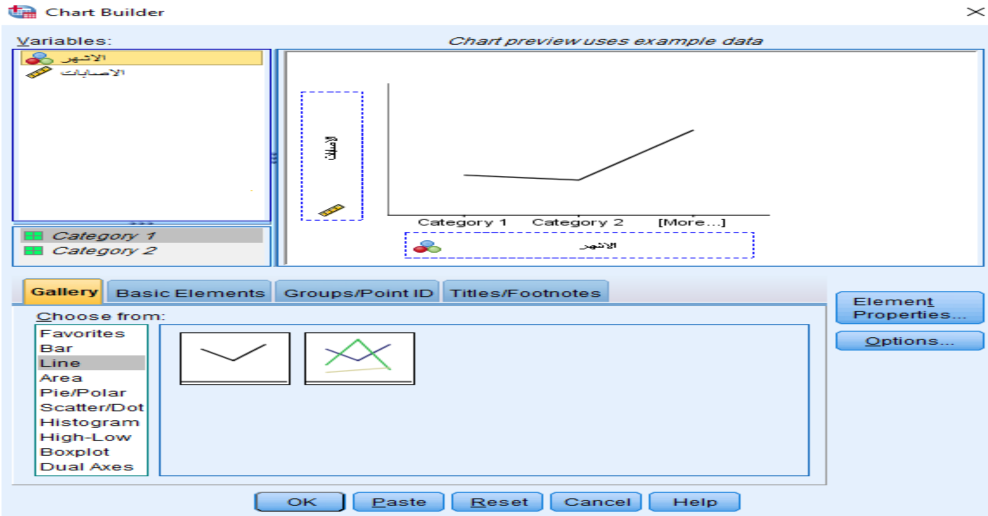
1- ادخال البيانات بعد تسمية المتغيرات على صفحة محرر المتغيرات كما في الشكل (4-25).



	الاشهر	الاصابات	var
1	1	217	
2	2	225	
3	3	340	
4	4	315	
5	5	189	
6	6	120	
7	7	168	
8	8	176	
9	9	115	
10	10	230	
11	11	245	
12	12	235	

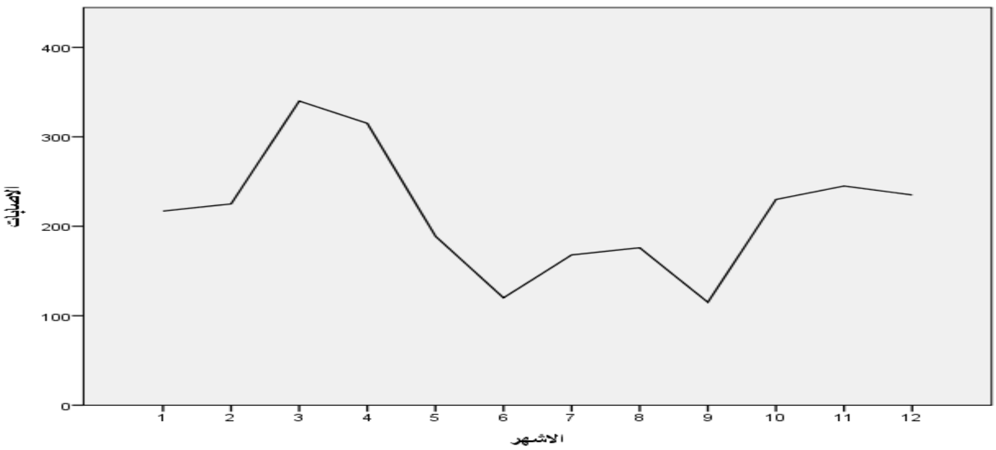
الشكل (4-25) عدد الإصابات بفايروس كوفيد-19 في محرر البيانات

2- نختار من الامر (Graph) الامر الفرعي (Chart Builder) ومنه نختار رسم الخط (Line) كما في الشكل (4-26).



الشكل (4−26) مربع الحوار لاختيار الخط البياني

3− بعد اختيار مربع الحوار ونقل المتغيرات في الأماكن المخصصة لها كالأشهر تكون في الخط الافقي والاصابات تكون في الخط العمودي واختيار الامر (OK) نحصل شكل الخط البياني كما في الشكل (4−27).



الشكل (4-27) الخط البياني الذي يمثل عدد الإصابات حسب عدد الأشهر

4-3-6 شكل الانتشار Scatter Plot

وهو مجموعة من النقاط المتكونة من التقاء قيمة الظاهرة الأولى على المحور الافقي مع القيمة المقابلة لها للظاهرة الثانية على المحور العمودي، ويستعمل شكل الانتشار لتمثيل العلاقة بين ظاهرتين كميتين او ظاهرة كمية وأخرى نوعية، ومنه نستطيع اكتشاف شكل التوزيع او العلاقة بين الظاهرتين هل هي موجبة او سالبة، وهل ان التوزيع متماثل او قريب من التماثل او يأخذ شكلا اخر. ولرسم شكل الانتشار باستعمال البرنامج (SPSS) نورد المثال الاتي:

مثال 7

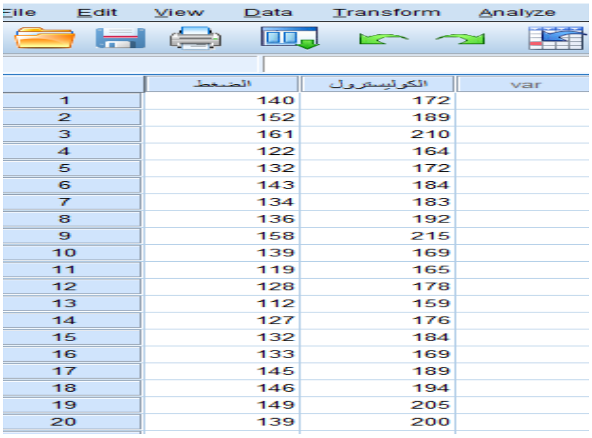
البيانات في الجدول (4-8) تمثل مستوى ضغط الدم العالي مع مستوى الكوليسترول لمجموعة من الأشخاص والمطلوب رسم شكل العلاقة بين الظاهرتين باستعمال شكل الانتشار.

جدول (4-8) بيانات ضغط الدم ومستوى الكوليسترول لمجموعة من الاشخاص

ت	1	2	3	4	5	6	7	8	9	10
ضغط الدم	148	152	161	122	132	143	134	136	158	139
الكوليسترول	172	189	210	164	172	184	183	192	215	169
ت	11	12	13	14	15	16	17	18	19	20
ضغط الدم	119	128	112	127	132	133	145	146	149	139
الكوليسترول	165	178	159	176	184	169	189	194	205	200

الحل:

1- ادخال البيانات الى محرر البيانات في البرنامج تحت اسم متغيرين وبوصفهما ظاهرتين كميتين كما في الشكل (4-28)

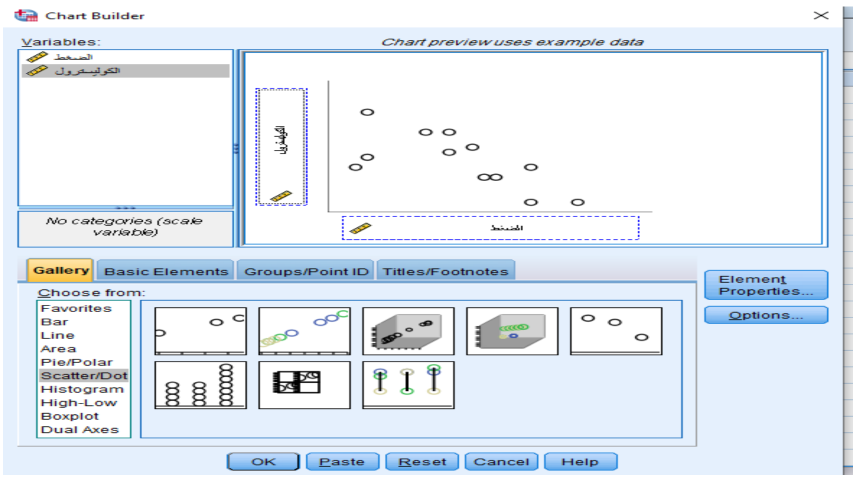


File Edit View Data Transform Analyze

	الضغط	الكوليسترول	var
1	140	172	
2	152	189	
3	161	210	
4	122	164	
5	132	172	
6	143	184	
7	134	183	
8	136	192	
9	158	215	
10	139	169	
11	119	165	
12	128	178	
13	112	159	
14	127	176	
15	132	184	
16	133	169	
17	145	189	
18	146	194	
19	149	205	
20	139	200	

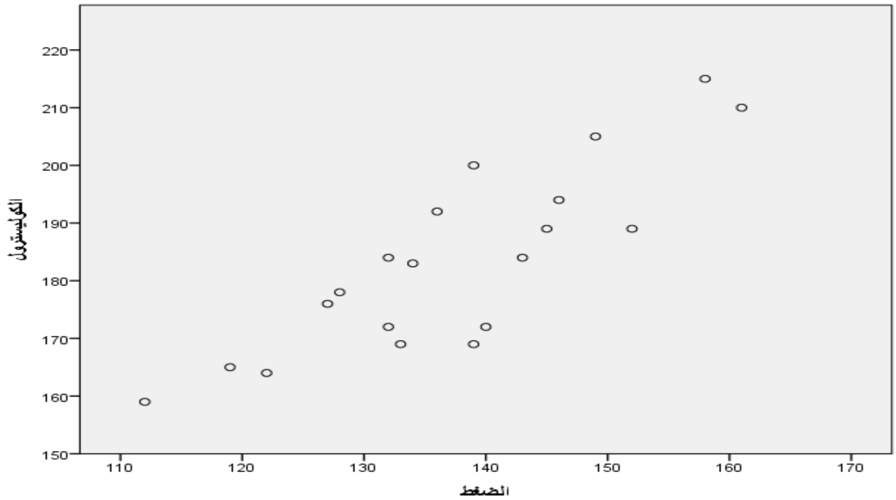
الشكل (4-28) ادخال البيانات الى محرر البيانات في برنامج (SPSS)

ننقر على الامر (Graph) ومنه نستعمل الامر الفرعي (Chart Builder) ومن مربع الحوار الذي يمثل الامر الفرعي نختار شكل الانتشار (Scatter/Dot)، ثم ننقر على الشكل المناسب مرتين كي ينقل الى المربع في الأعلى، وبعد ذلك نسحب اسم الظاهرة الأولى على المحـور الأفقـي والظاهرة الثانية على المحور العمودي كما في الشكل (4-29).



الشكل (4-29) اختيار شكل الانتشار

ثم بعد ذلك ننقر على الامر (OK) لنحصل على الشكل النهائي للانتشار كما في الشكل (4-30).



الشكل (4−30) شكل الانتشار بين ضغط الدم والكوليسترول لمجموعة من الاشخاص

نلاحظ من الشكل (4−30) بان العلاقة بين ظاهرة ضغط الدم ومستوى الكوليسترول تكون موجبة او تسمى طردية أي بزيادة الظاهرة الأولى تزداد الظاهرة الثانية والعكس صحيح، وهناك اشكال أخرى يمكن استعمال البرنامج نفسه الأوامر نفسها لرسم هذه الاشكال عند الحاجة لها.

تمارين الفصل الرابع

1- البيانات الآتية تمثل الوقت المصروف (دقيقة) من الطبيب مع المريض في اثناء مراجعته في العيادة الخارجية لأحد المستشفيات الخاصة:

10، 12، 18، 22، 24، 10، 8، 11، 28، 9، 14، 27، 19، 14، 26، 15، 13، 23، 7، 9، 21، 20، 20، 18، 19، 12، 14، 10.

والمطلوب: هل يمكن استعمال أحد الاشكال البيانية لتمثيل البيانات اعلاه.

2- البيانات الاتية تمثل جدول توزيع تكراري لقياسات ضغط الدم العالي لمجموعة من الاشخاص المدخنين.

120-	130-	140-	150-	160-	170-	180-190
3	4	9	10	13	8	5

والمطلوب: ارسم الاشكال البيانية المناسبة لهذا التوزيع التكراري.

3- كان اعداد المراجعين للعيادة الخارجية في أحد المستشفيات حسب ايام الاسبوع كما في الجدول الاتي:

السبت	الاحد	الاثنين	الثلاثاء	الاربعاء	الخميس	الجمعة
23	45	53	64	56	42	18

والمطلوب: رسم الاشكال البيانية المناسبة للجدول اعلاه.

4- البيانات الاتية تمثل اوزان الاطفال حديثي الولادة:

N	UW	UN	UW	N	UN	UW	UN	UN	UN	N	N
N	UN	UN	UW	N	UN	UW	N	N	UW	UN	UW
UN	UN	N	N	N	UW	N	UN	N	UW	N	UN

اذ ان (N) تمثل الوزن الطبيعي، (UN) تمثل الوزن فوق الطبيعي، (UW) تمثل الوزن اقل من الطبيعي.

والمطلوب: تمثيل البيانات في الجدول اعلاه بيانيا، واي الاشكال البيانية يكون مناسباً لتمثيلها.

5- البيانات الآتية تمثل معدل الولادات والوفيات لكل ألف من السكان لمجموعة من المحافظات لإحدى السنوات. والمطلوب اختيار الشكل المناسب لتمثيلها بيانيا

المحافظة	بابل	كربلاء	بغداد	واسط	نجف	الديوانية
معدل الولادات	18	17	42	15	14	12
معدل الوفيات	10	9	12	11	8	10

الفصل الخامس

مقاييس النزعة المركزية

Central Tendency Measurements

1.5 تمهيد

يعد الوصف العددي للبيانات الطريقة الثالثة لوصف البيانات الكمية او الوصفية فضلاً عن الجداول التكرارية والاشكال البيانية، وتختلف المقاييس العددية في حالة البيانات الوصفية عنها في حالة البيانات الكمية، ففي حالة البيانات الوصفية توجد لدينا تكرارات للصفات، لذلك يكون المقياس المناسب هو النسبة اما في حالة البيانات الكمية ننظر الى صفتين اساسيتين للبيانات، هما الموقع او المركز والصفة الثانية: هي التشتت او التباعد عن المركز، وعليه يكون لدينا نوعان من المقاييس العددية، وحسب نوع البيانات، فالنسب للبيانات الوصفية ومقاييس النزعة المركزية للبيانات الكمية وهما: موضوع هذا الفصل، اما التشتت فيكون موضوع الفصل القادم.

2.5 المقاييس العددية في حالة البيانات الوصفية (النوعية)

Numerical Measurements of Qualitative Data

لا توجد قياسات للبيانات الوصفية، وانما توجد اعداد (تكرار) لكل صفة من الصفات، ومن ثم فان المقاييس العددية المناسبة لهذا النوع من البيانات هي اما النسبة (Proportion) او النسبة المئوية (Percentage) لكل صفة في الظاهرة وتحسب هذه المقاييس على وفق الصيغ الآتية:

$$Proportion = \frac{f_i}{\sum f_i} \qquad (1-5)$$

$$Percentage = \frac{f_i}{\sum f_i} \times 100 \qquad (2-5)$$

حيث ان (f_i) تمثل عدد المشاهدات في الصفة (i) و$(\sum f_i)$ يمثل مجموع المشاهدات في الظاهرة، وكما موضح في المثال الآتي:

مثال 1

البيانات في الجدول الآتي توضح توزيع عينة من المرضى حسب حالة الاجابة بالمرض.

والمطلوب: ايجاد المقاييس العددية المناسبة للبيانات.

جدول رقم (5−1) توزيع المرضى حسب حالة المرض

حالة الاصابة بالمرض	غير مريض	بسيطة	متوسطة	شديدة	المجموع
التكرار (f_i)	3	5	6	4	18

الحل:

بما ان البيانات وصفية، إذن المقياس العددي المناسب هو النسبة كما في المعادلة (5−1) او النسبة المؤوية كما في المعادلة (5−2) وكما في الجدول الآتي:

جدول (5−2) حسابات النسبة والنسبة المؤوية للجدول (5−1)

حالة الاصابة بالمرض	التكرار (f_i)	النسبة (Proportion)	النسبة المؤوية (Percentage)
غير مريض	3	0.17=18/3	17 =100×18/3
بسيطة	5	0.28=18/5	28 =100×18/5
متوسطة	6	0.33=18/6	33 =100×18/6
شديدة	4	0.22=18/4	22 =100×18/4
المجموع	18	1	100

فضلاً عن هذين المقياسين هناك مقياس ثالث يمكن استعماله في حالة البيانات الوصفية هو المنوال، وهو أحد مقاييس النزعة المركزية كما سنلاحظ في المبحث الآتي.

3.5 المقاييس العددية في حالة البيانات الكمية

Numerical Measurements of Quantitative Data

البيانات الكمية هي قياسات كمية تعبر عن الظاهرة موضع الدراسة، لذلك فان المقاييس العددية في حالة البيانات الكمية هي تلك المقاييس التي تستعمل لقياس موضع تركز أو تجمع البيانات وتسمى مقاييس النزعة المركزية، لأن بيانات أي ظاهرة في الغالب تتركز او تتجمع حول قيمة او قيم معينة، تسمى هذه القيم بمقاييس النزعة المركزية (Central Tendency Measurements) وتستعمل لتلخيص البيانات عدديا، إذ أنها تعد قيماً إنموذجية أو مثالية للبيانات، كما أن هذه المقاييس تستعمل لوصف مجموعة البيانات، وكذلك لمقارنة المجموعات المختلفة للبيانات، وهناك عدد من المقاييس سنكتفي بأهم ثلاثة مقاييس هي: الوسط الحسابي أو المتوسط، الوسيط، والمنوال وهي كما في المباحث الفرعية الاتية:

5-3-1 الوسط الحسابي (المتوسط): Arithmetic Mean (Mean)

يعد الوسط الحسابي من أهم مقاييس النزعة المركزية وأفضلها، ومن أكثرها شيوعا واستعمالا في التحليل الإحصائي، وذلك لما يتمتع به من خصائص وصفات إحصائية جيدة، ولإيجاد الوسط الحسابي للبيانات فإننا لابد أن نفرق بين البيانات المفردة غير المبوبة في جدول تكراري والبيانات المبوبة الملخصة في جدول تكراري.

1- الوسط الحسابي للبيانات غير المبوبة (Mean of Ungrouped data) :

إذا كانت البيانات غير مبوبة في جدول تكراري وبالشكل الاولي للبيانات فان عدد البيانات لعينة بحجم هو n وكانت البيانات هي $x_1, x_2, \dots, x_n$ فإن مجموع هذه البيانات هو

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n; \qquad i = 1,2,\dots,n \qquad (3-5)$$

والوسط الحسابي لهذه البيانات يكون حاصل قسمة مجموع البيانات على عددها وكما في العلاقة الاتية:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad (4-5)$$

مثال 2

البيانات الاتية هي عبارة عن أوزان مجموعة مكونة من ثمانية أشخاص بالكيلوغرام أوجد المتوسط (الوسط الحسابي) للوزن (35, 37, 48, 45, 47, 55, 60, 65)

الحل:

$$x_1 = 35, x_2 = 37, x_3 = 48, x_4 = 45, x_5 = 47, x_6 = 55, x_7 = 60,$$

$$x_8 = 65, \qquad n = 8$$

$$\therefore \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{35+37+\cdots+65}{8} = \frac{392}{8} = 49$$ كيلوغرام

2- الوسط الحسابي للبيانات المبوبة (Mean of grouped data)

ينبغي في حالة البيانات الملخصة في جدول تكراري بسيط ملاحظة ان البيانات الأصلية غير معروفة بينما عدد البيانات في كل فئة (تكرار الفئة) معروف وان الجدول من النوع المغلق، وفضلاً عن ذلك فان مركز الفئة يستعمل بوصفه قيمة تقريبية لجميع البيانات في الفئة. فعلى سبيل المثال إذا كان لدينا بيانات عددها(n) وكانت هذه البيانات ملخصة في جدول تكراري بحيث أن (k) يمثل عدد الفئات كما في التوزيع التكراري البسيط ($j = 1,2,\dots,k$) كما في الجدول التكراري البسيط رقم (3−4). ولحساب الوسط الحسابي ($\bar{x}$) نحتاج الى اتباع الخطوات الاتية:

أ- إيجاد مركز كل فئة وحسب القاعدة الاتية:

$$x_j = \frac{A+B}{2} \qquad (5-5)$$

حيث ان (A) يمثل الحد الأدنى للفئة، (B) الحد الأعلى للفئة، (x_j) مركز الفئة (j).

ب - ايجاد حاصل ضرب مركز كل فئة (x_j) بالتكرار (f_j) المقابل له

ت - نوجد مجموع حاصل الضرب ($\sum_{j=1}^{k} x_j f_j$) ونقسم على مجموع التكرارات ($\sum_{j=1}^{k} f_j$) كما في الصيغة الاتية:

$$\bar{x} = \frac{\sum_{j=1}^{k} x_j f_j}{\sum_{j=1}^{k} f_j} \qquad (6-5)$$

ولتوضيح هذه الخطوات نورد المثال الاتي:

مثال 3

البيانات في الجدول رقم (5−3) تمثل مستوى الهيموجلوبين في الدم لعينة مكونة من خمسين شخصا، اوجد الوسط الحسابي لمستوى الهيموجلوبين.

جدول (5−3) التوزيع التكراري لمستوى الهيموجلوبين لعينة من الاشخاص

الفئات	−12.95	−13.95	−14.95	−15.95	−16.95	18.95−17.95
التكرار f_j	3	5	15	16	10	1

الحل:

لإيجاد الوسط الحسابي لمستوى الهيموجلوبين من الجدول التكراري (5−3) نتبع الخطوات أعلاه

أ− نوجد مراكز الفئات وفق القاعدة (5−5) وكما في العمود الثالث من الجدول (5−4).

ب− نجد حاصل ضرب مركز الفئة (x_j) في التكرار المقابل لها (f_j) وكما في العمود الرابع من الجدول (5−4) ونجد مجموع حاصل الضرب ($\sum x_j f_j$).

ت−نجد مجموع التكرارات ($\sum f_j$) ونطبق المعادلة رقم (5−6) لإيجاد الوسط الحسابي لمستوى الهيموجلوبين كما في الجدول (5−4) الاتي:

..

جدول (4-5) العمليات الحسابية لإيجاد الوسط الحسابي

فئات مستوى الهيموجلوبين	التكرار f_j	مراكز الفئات x_j	$x_j f_j$
12.95-	3	13.45	40.35
13.95-	5	14.45	72.25
14.95-	15	15.45	231.75
15.95-	16	16.45	263.20
16.95-	10	17.45	174.50
17.95-18.95	1	18.45	18.45
المجموع	$\sum f_j = 50$		$\sum x_j f_j = 800.5$

$$\bar{x} = \frac{\sum x_j f_j}{\sum f_j}$$

$$= \frac{800.5}{50} = 16.01$$

3– بعض الخصائص الجبرية للوسط الحسابي:

يتصف الوسط الحسابي ببعض الخصائص الجبرية لأنه يعتمد على العمليات الجبرية في حسابه منها:

$$\left.\begin{array}{lll} 1 - if & x_i = x_1,\ x_2, \cdots x_n & \therefore \sum_{i=1}^{n}(x_i - \bar{x}) = 0 \\ 2 - if & x \pm b & \therefore \overline{(x \pm b)} = \bar{x} \pm b, \\ 3 - if & ax & \therefore \overline{(ax)} = a\bar{x} \\ 4 - if & ax \pm b & \therefore \overline{(ax \pm b)} = a\bar{x} \pm b \end{array}\right]$$

(7–5)

حيث ان (a,b) كميات ثابتة. ولتطبيق هذه الخصائص نأخذ المثال الاتي:

مثال 4

توفرت لديك البيانات الاتية (5, 6, 4 , 3,7) والمطلوب تحقيق خصائص الوسط الحسابي، كما في المعادلة (5−7).

1− ايجاد الوسط الحسابي باستخدام المعادلة (5−4) كالاتي:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{5+6+4+7+3}{5} = 5$$

$$\therefore \sum (x_i - \bar{x})$$
$$= (5-5) + (6-5) + (4-5) + (7-5) + (3-5) = 0$$

2− نفرض ان قيمة (b=6) وبإضافة هذه القيمة للقيم ستكون القيم الجديدة كالاتي:

$$x_i = 9, \quad 13, \quad 10, \quad 12, \quad 11,$$

والوسط الحسابي الجديد يكون كالاتي:

$$\overline{x+b} = \frac{9+13+10+12+11}{5} = 11$$

$$\therefore \overline{(x+b)} = 11$$
$$= \bar{x} + 6$$
$$= 5 + 6 = 11$$

3−نفرض ان قيمة (a=4) وبضرب هذه القيمة في كل قيمة من القيم السابقة نحصل على قيم جديدة كالآتي:

$$x_1 = 5 \times 4 = 20, x_2 = 6 \times 4 = 24, x_3 = 4 \times 4 = 16, x_4 = 7 \times 4 = 28, \quad x_5 = 3 \times 4 = 12$$

والوسط الحسابي للقيم الجديدة باستخدام المعادلة (5−4) سيكون كالاتي:

$$\therefore \overline{ax} = \frac{20+24+16+28+12}{5} = 20$$
$$= 4 \times \bar{x}$$
$$= 4 \times 5 = 20$$

4−إذا تم ضرب القيم بقيمة (a) واضافة قيمة (b) الى القيمة الناتجة فان القيم الجديدة ستكون كالاتي:

$$x_1 = 5 \times 4 + 6 = 26, x_2 = 6 \times 4 + 6 = 30, x_3 = 4 \times 4 + 6$$
$$= 22, \qquad x_4 = 7 \times 4 + 6 = 34,$$
$$x_5 = 3 \times 4 + 6 = 18$$

والوسط الحسابي للقيم الجديدة باستخدام المعادلة (5−4) يكون كالاتي:

$$\therefore \overline{(ax+b)} = \frac{26 + 30 + 22 + 34 + 18}{5} = 26$$
$$= 4\bar{x} + b$$
$$= 4 \times 5 + 6 = 26$$

5− هناك خاصية اخرى للوسط الحسابي وهي إذا كان لدينا مجموعتان من البيانات بحيث أن عدد بيانات المجموعة الأولى هو n_1 ومتوسطها هو $\bar{x}_1$ وكان عدد بيانات المجموعة الثانية هو n_2 ومتوسطها هو $\bar{x}_2$ فإن متوسط المجموعة الكلية المكونة من دمج هاتين المجموعتين يمكن حسابه بالصيغة الأتية:

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2} \qquad (8-5)$$

وبشكل عام فإن الوسط الحسابي يعد من أفضل مقاييس النزعة المركزية ومن أكثرها شيوعا، وذلك لما يتمتع به من صفات جيدة، منها انه سهل الحساب، ويخضع للعمليات الجبرية في حسابه ودائما يكون قيمة وحيدة لمجموعة البيانات الواحدة، ويعتمد في حسابه على البيانات جميعها. وبالرغم من هذه الصفات فإن له بعض العيوب، منها انه يتأثر بالقيم الشاذة أو المتطرفة، ولا يمكن حسابه في حالة البيانات الوصفية (النوعية) إذ يمكن حسابه للبيانات الكمية فقط.

5−3−2 الوسيط Median

يعرف الوسيط لمجموعة من البيانات على أنه تلك القيمة التي تتوسط البيانات عند ترتيبها تصاعديا او تنازليا، أي أنه يمثل تلك القيمة التي تقسم البيانات بعد ترتيبها إلى جزئين متساويين من حيث العدد ومن ثم فان البيانات في الجزء الأول هي البيانات التي تكون أصغر من الوسيط والبيانات في الجزء الثاني تمثل البيانات التي تكون أكبر من الوسيط إذا كان الترتيب تصاعدي، أي أن 50% من البيانات أصغر من الوسيط

و50% من البيانات تزيد عن الوسيط، والعكس إذا كان الترتيب تنازلياً وكما موضح بالشكل رقم (5−1) ويرمز للوسيط بالرمز (Me).

شكل رقم (5−1) توزيع مجموعة البيانات حول الوسيط

ولإيجاد الوسيط للبيانات فإننا لابد أن نفرق بين البيانات المفردة غير المبوبة في جدول تكراري والبيانات المبوبة او الملخصة في جدول تكراري.

1− <u>الوسيط في حالة البيانات غير المبوبة</u> Median of ungrouped data

إذا توافرت عينة بحجم (n) من المشاهدات $(x_1, x_2, \dots, x_n)$ يمكن ايجاد الوسيط باتباع الخطوات الاتية:

1− نرتب هذه المشاهدات تصاعديا من أصغر قيمة الى أكبر قيمة في مجموعة المشاهدات.

2− نوجد ترتيب الوسيط ويعتمد على كون عدد المشاهدات فردي او زوجي، فاذا كان العدد فردي فان الترتيب (R_{Me}) يكون:

$$R_{Me} = \frac{n+1}{2} \qquad (9\text{−}5)$$

اما إذا كان عدد القيم زوجياً فهنا تكون للوسيط قيمتين وبالتالي نوجد ترتيبين كالاتي:

$$\left.\begin{aligned} R_{Me1} &= \frac{n}{2} \\ R_{Me2} &= R_{Me1} + 1 \end{aligned}\right] \qquad (10-5)$$

3- ايجاد قيمة الوسيط وهي القيمة المقابلة لترتيب الوسيط في حالة العدد الفردي للمشاهدات اما في حالة العدد الزوجي للمشاهدات فان الوسيط يساوي متوسط القيمتين الاولى والثانية واللتين تقابلان الترتيب الاول والثاني.

ولتوضيح ما تقدم نأخذ المثالين الآتيين:

مثال 5

البيانات الآتية تمثل الوزن (كغم) لمجموعة من الاطفال، (9.3، 5.7، 6.3، 3.4، 8.2، 4.5، 5.4). اوجد الوسيط للوزن.

الحل:

1- نرتب قيم المشاهدات تصاعديا من أصغر قيمة حتى أكبر قيمة كالاتي:
3.4، 4.5، 5.4، 5.7، 6.3، 8.2، 9.3.

2- بما ان عدد المشاهدات فردي ($n = 7$) لذلك فان ترتيب الوسيط (R_{Me}) يكون باستخدام المعادلة (5-9) كالاتي:

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

3- قيمة الوسيط تساوي القيمة والتي ترتيبها يساوي 4 من بين البيانات وهي (5.7). فإنها القيمة التي قسمت البيانات الى جزئين متساويين من حيث العدد فكانت ثلاث قيم هي الاصغر من الوسيط تقع قبل الوسيط وثلاث قيم هي الاكبر من الوسيط تقع بعد الوسيط.

مثال 6

البيانات الاتية تمثل الوزن (كغم) لمجموعة من الاطفال، (9.3، 5.7، 6.3، 3.4، 8.2، 4.5، 5.4، 7.3). اوجد الوسيط للوزن.

الحل:

1- نرتب قيم المشاهدات تصاعديا من أصغر قيمة وحتى أكبر قيمة كالاتي:
3.4، 4.5، 5.4، 5.7، 6.3، 7.3، 8.2، 9.3 .

2- بما ان عدد المشاهدات زوجي ($n = 8$) لذلك فانه تكون قيمتان للوسيط ترتيبهما يكون باستخدام المعادلة (5-10) كالاتي:

$$R_{Me1} = \frac{n}{2} = \frac{8}{2} = 4$$

$$R_{Me2} = R_{Me1} + 1 = 4 + 1 = 5$$

3- القيمة الاولى للوسيط هي القيمة التي ترتيبها يساوي 4 من بين البيانات وهي (5.7) والقيمة الثانية للوسيط هي القيمة التي تقابل الترتيب الثاني 5 من بين البيانات وهي (6.3) ومن ثم فان الوسيط هو متوسط هاتين القيمتين، اي ان:

$$Me = \frac{5.7+6.3}{2} = 6$$

فإنها القيمة التي قسمت البيانات الى جزئين متساويين، من حيث العدد، فكانت أربع قيم هي الاصغر من الوسيط تقع قبل الوسيط وأربع قيم هي الاكبر من الوسيط تقع بعد الوسيط.

2- <u>الوسيط في حالة البيانات المبوبة</u> (Median of Grouped data)

يمكن حساب الوسيط للبيانات الملخصة في جدول تكراري بطريقتين هما: طريقة حسابية، وطريقة بيانية، ويستعمل الجدول التكراري المتجمع الصاعد لإيجاد الوسيط حسابيا في حين يستعمل المنحنى التكراري المتجمع الصاعد لإيجاد الوسيط بيانيا.

1- حساب الوسيط حسابيا: لحساب الوسيط حسابيا نتبع الخطوات الاتية:

أ- ترتيب البيانات تصاعديا من خلال تكوين الجدول التكراري المتجمع الصاعد.

ب- ايجاد ترتيب الوسيط وفق القاعدة الاتية:

$$R_{Me} = \frac{\sum f_j}{2} \qquad (11-5)$$

ت-تحديد فئة الوسيط وهي الفئة التي تقابل ترتيب الوسيط، ومنها نحدد الحد الادنى لفئة الوسيط ونرمز له بالحرف (A)، ونحدد طول فئة الوسيط وهو الفرق بين الحد الاعلى والحد الادنى لفئة الوسيط ونرمز له بالحرف (L).

ث−نحدد التكرار المتجمع الصاعد السابق لترتيب الوسيط ونرمز له (F_1) وكذلك نحدد التكرار المتجمع الصاعد اللاحق لترتيب الوسيط ونرمز له (F_2).

ج− نحسب قيمة الوسيط (Me) حسب العلاقة الآتية:

$$Me = A + \left(\frac{R_{Me}-F_1}{F_2-F_1}\right) \times L \qquad (12-5)$$

ولتوضيح هذه الخطوات نأخذ المثال الآتي:

مثال 7

قام طبيب بقياس مستويات الكولسترول لمجموعة من المرضى الشباب فوجد ان القياسات كما في الجدول التكراري البسيط الاتي اوجد الوسيط حسابيا.

جدول رقم (5−5) جدول التوزيع التكراري لمستويات الكولسترول

التكرار f_i	مستويات الكولسترول (Mg/ml)
1	195−
3	200−
4	205−
7	210−
4	215−
1	220−225
20	المجموع

الحل:

−نكون جدول التوزيع التكراري المتجمع الصاعد من جدول التوزيع التكراري البسيط رقم (5−5) كالاتي:

جدول (5−6) الجدول التكراري المتجمع الصاعد

الحدود العليا للفئات	التكرار f_j	التكرار المتجمع الصاعد
اقل من 195	0	0
اقل من 200	1	1
اقل من 205	3	4
اقل من 210	4	8 F_1
اقل من 215	7	15 F_2
اقل من 220	4	19
اقل من 225	1	20
المجموع	20	

ح− نوجد ترتيب الوسيط وفق المعادلة (5−11) وكالاتي:

$$R_{Me} = \frac{\sum f_j}{2} = \frac{20}{2} = 10$$

خ− نحدد فئة الوسيط وهي الفئة (210−215) التي تقابل ترتيب الوسيط. ومنها يكون الحد الادنى لفئة الوسيط (A=210). وطول فئة الوسيط (L=215−210=5).

د− نحدد التكرار المتجمع السابق لترتيب الوسيط ($F_1 = 8$)، والتكرار المتجمع اللاحق لترتيب الوسيط ($F_2 = 15$).

ذ− نحسب قيمة الوسيط باستخدام المعادلة رقم (5−12) كالاتي:

$$Me = A + \left(\frac{R_{Me} - F_1}{F_2 - F_1}\right) \times L$$
$$= 210 + \left(\frac{10-8}{15-8}\right) \times 5$$
$$= 210 + 1.428$$
$$= 211.428$$

في هذا المثال نستطيع القول بأن 50% من الأشخاص (نصف الأشخاص) يقل مستوى الكوليسترول لهم عن Me =211.428 Mg/Ml .

2- **حساب الوسيط بيانيا:** يمكن إيجاد الوسيط بيانيا اما برسم المنحنى المتجمع الصاعد او الهابط اوكلاهما في شكل واحد وعند استعمال المنحنى الصاعد نتبع الخطوات الاتية:

أ- تكوين جدول التكرار المتجمع الصاعد من الجدول التكراري البسيط.

ب- رسم المنحنى التكراري المتجمع الصاعد من خلال رسم التكرار المتجمع على المحور العمودي والحدود العليا للفئات على المحور الافقي.

ت- تحديد ترتيب الوسيط باستخدام المعادلة رقم $(5-11)$ وتحديد موقعه على المحور العمودي الذي يمثل التكرار المتجمع، ونرسم خطاً افقياً موازياً للمحور الافقي ومن نقطة التقاء الخط الافقي مع المنحنى المتجمع الصاعد نرسم خطاً عمودياً موازياً للمحور العمودي ونقطة التقاء هذا الخط مع المحور الافقي تمثل قيمة الوسيط. وكما موضح بالمثال الاتي:

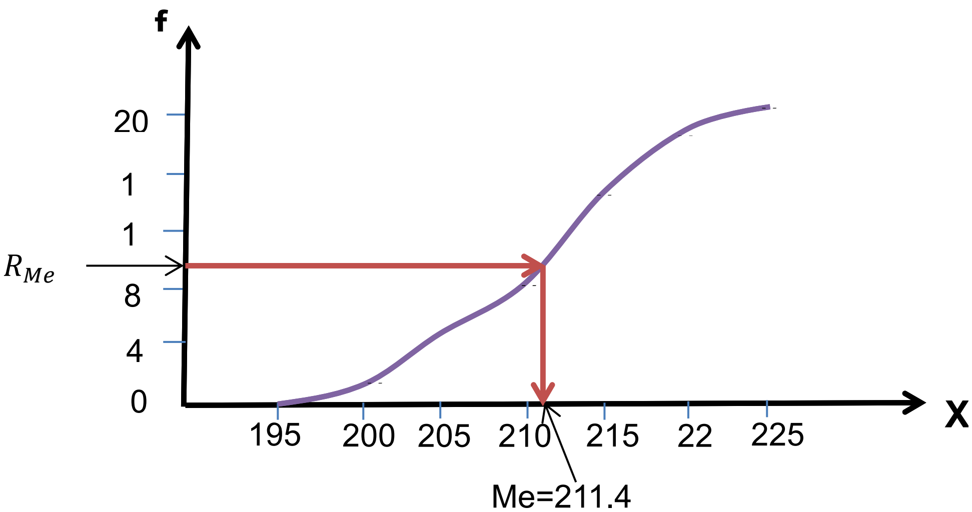
<u>مثال 8:</u> استعمل البيانات في المثال رقم (7) لإيجاد قيمة الوسيط بيانيا.

<u>الحل:</u>

أ- تكوين جدول التكرار المتجمع الصاعد كما في المثال السابق.

ب - ترتيب الوسيط تم حسابه في المثال السابق فكان $(R_{Me} = 10)$.

ت - نرسم المنحنى المتجمع الصاعد كما في الشكل رقم $(5-2)$ ونحدد الوسيط كما في الخطوات السابقة فتكون قيمة الوسيط مقاربة لقيمته عند حسابه بالطريقة الحسابية.



شكل رقم $(5-2)$ ايجاد الوسيط بيانيا باستخدام المنحنى التكراري االمتجمع الصاعد

3- بعض مميزات وعيوب الوسيط:

يعد الوسيط من مقاييس النزعة المركزية الشائعة، وذلك لما يتمتع به من بعض الصفات الجيدة .ومن مميزاته انه سهل التعريف والحساب، وانه وحيد لمجموعة البيانات الواحدة، وهو أقل تأثرا من المتوسط بالقيم الشاذة أو المتطرفة. وبالرغم من أن الوسيط يعتبر من مقاييس النزعة المركزية الجيدة فإن له بعض العيوب منها انه لا يأخذ في الاعتبار جميع البيانات في عملية احتسابه إذ أنه يعتمد فقط على القيم التي تقع بعد ترتيبها في المنتصف وبغض النظر عن قيمها، وبشكل عام لا يمكن حساب الوسيط للبيانات الوصفية (النوعية) غير القابلة للترتيب.

5-3-3 المنوال Mode

يعد المنوال أحد مقاييس النزعة المركزية شائعة الاستعمال، ولاسيما في حالة البيانات الوصفية ويعرف المنوال لمجموعة البيانات على أنه تلك القيمة التي تتكرر أكثر من غيرها، أي أنها القيمة ذات التكرار الأكبر إن وجدت، ويرمز للمنوال بالرمز (Mo). ولإيجاد المنوال للبيانات فإننا لابد أن نفرق بين البيانات المفردة غير المبوبة في جدول تكراري والبيانات المبوبة او الملخصة في جدول توزيع تكراري.

1- المنوال في حالة البيانات غير المبوبة Mode of ungrouped data

يتضح لنا من تعريف المنوال بانه في حالة البيانات المفردة تكون لدينا أنواع عدة من البيانات، النوع الأول: يمثل بيانات ليس لها منوال، وتسمى عديمة المنوال، والنوع الثاني: يمثل بيانات لها منوال واحد وتسمى وحيدة المنوال والنوع الثالث يمثل بيانات لها أكثر من منوال وتسمى متعددة المنوال. وكما هو موضح في المثال الاتي:

مثال 9

البيانات الاتية تمثل أربع صفات لمجموعة من الاشخاص، العمر مقاس بالسنوات والوزن مقاس بالكيلوغرام والطول مقاس بالسنتمتر وفصيلة الدم. اوجد المنوال لكل صفة من الصفات وبين نوع مجموعة البيانات حسب مقياس المنوال.

جدول رقم (5–7) البيانات التي تمثل الصفات الاربعة لمجموعة من الاشخاص

التسلسل	1	2	3	4	5	6	7	8	9	10
العمر	23	20	22	50	23	25	27	23	34	23
الوزن	70	68	66	70	57	86	68	70	68	55
الطول	167	176	180	193	178	182	170	160	188	175
فصيلة الدم	A	B	O	A	AB	B	A	O	A	AB

الحل:

القاعدة التي تستعمل لإيجاد المنوال في حالة البيانات غير المبوبة هي ان (المنوال = القيمة الاكثر تكرارا من بين البيانات) وعليه فان المنوال للصفات الاربعة الاتي:

أ- صفة العمر نجد انها وحيدة المنوال لان قيمة المنوال هي 23 لأنها تكررت أربع مرات وأكثر من باقي القيم.

ب- صفة الوزن هي متعددة المنوال (ثنائية المنوال) لان فيها قيمتين تكررتا أكثر من باقي القيم ولهما نفس التكرار ثلاث مرات وهما (68 و70).

ت- صفة الطول تعد عديمة المنوال لأنه لا توجد قيمة قد تكررت أكثر من باقي القيم وانما جميع القيم لها نفس التكرار.

ث- فصيلة الدم تعد وحيدة المنوال وقيمة المنوال هي الفصيلة (A) لأنها تكررت أكثر من باقي القيم وعلى الرغم من انه هناك صفات اخرى تكررت، ولكن هذه الصفة هي الاكثر تكرارا.

2- المنوال في حالة البيانات المبوبة Mode of grouped data

تعرف الفئة التي تحتوي على قيمة المنوال في حالة البيانات المبوبة في جداول تكرارية بأنها الفئة التي تقابل أكبر تكرار. وهنا قد يكون هناك فئة منواليه واحدة ومن ثم قيمة منواليه واحدة أو عدة فئات منواليه ومن ثم عدة قيم منواليه أو قد لا توجد فئة منواليه تقابل أكبر تكرار وانما جميع الفئات لها التكرار نفسه وعلية لا يوجد منوال.

ويمكن حساب المنوال للبيانات المبوبة في جدول تكراري بواحدة من طريقتين هما: الطريقة الحسابية، والطريقة البيانية. ويستعمل الجدول التكراري لإيجاد المنوال حسابيا في حين يستعمل المدرج التكراري لإيجاد المنوال بيانيا. وحسابيا هناك ثلاث طرق لحساب المنوال من الجدول التكراري البسيط هي:

أ- **الطريقة العامة (طريقة مركز الفئة):** وفيها يكون المنوال = مركز الفئة التي تقابل أكبر تكرار. ولحسابه نتبع الخطوات الاتية:

- ❖ نحدد فئة المنوال وهي الفئة التي تقابل أكبر تكرار ومنها نحصل على الحد الادنى للفئة ونرمز له بالرمز (A) وكذلك الحد الاعلى لفئة المنوال ونرمز له بالرمز (B).
- ❖ نحسب قيمة المنوال حسب العلاقة الاتية:

$$Mo = \frac{A+B}{2} \qquad (13\text{-}5)$$

ب- **طريقة الفروق لبيرسون:** وفي هذه الطريقة نتبع الخطوات الاتية:

- ❖ نحدد فئة المنوال وهي الفئة التي تقابل أكبر تكرار ومنها نحصل على الحد الادنى للفئة ونرمز له بالرمز (A) وكذلك طول الفئة = الحد الاعلى – الحد الادنى للفئة ونرمز له بالرمز (L).
- ❖ نحدد الفرق الاول= تكرار فئة المنوال – تكرار الفئة السابقة لفئة المنوال ونرمز له بالرمز (D_1).

❖ نحدد الفرق الثاني = تكرار فئة المنوال – تكرار الفئة اللاحقة لفئة المنوال ونرمز له بالرمز (D_2).

❖ نحسب المنوال وفق القاعدة الاتية:

$$Mo = A + \left(\frac{D_1}{D_1+D_2}\right) \times L \qquad (14\text{–}5)$$

ت – **طريقة الرافعة:** ولحساب المنوال في هذه الطريقة نتبع الخطوات الاتية:

❖ نحدد فئة المنوال وهي الفئة التي تقابل أكبر تكرار، ومنها نحصل على الحد الادنى للفئة ونرمز له بالرمز(A) وكذلك طول الفئة = الحد الاعلى – الحد الادنى للفئة ونرمز له بالرمز (L).

❖ نحدد تكرار الفئة السابقة لفئة المنوال ونرمز له بالرمز (f_1).

❖ نحدد تكرار الفئة اللاحقة لفئة المنوال ونرمز له بالرمز (f_2).

❖ نحسب المنوال على وفق القاعدة الاتية:

$$Mo = A + \left(\frac{f_1}{f_1+f_2}\right) \times L \qquad (15\text{–}5)$$

مثال 10

استعمل البيانات في جدول (5–5) لحساب المنوال بالطرائق الحسابية المختلفة وهل يوجد فرق كبير بين قيم المنوال حسب الطرائق الثلاثة.

أ- **الطريقة العامة (طريقة مركز الفئة)**

❖ نجد من الجدول (5–5) ان فئة المنوال التي تقابل أكبر تكرار (7) هي (210–215).

❖ اذن فقيمة المنوال حسب العلاقة (5–13) تكون كالاتي:

$$Mo = \frac{A+B}{2}$$
$$= \frac{210+215}{2}$$
$$= 212.5$$

ب- طريقة الفروق لبيرسون

- نجد من الجدول (5-5) ان فئة المنوال هي (210-215) اذن الحد الادنى للفئه هو (A=210) وطول فئة المنوال هو (L=215-210=5).
- نجد الفرق الاول وهو الفرق بين تكرار فئة المنوال الذي يساوي 7 وتكرار الفئة السابقة لها والذي يساوي 4 اذن ($D_1 = 7 - 4 = 3$).
- نجد الفرق الثاني وهو الفرق بين تكرار فئة المنوال الذي يساوي 7 وتكرار الفئة اللاحقة لها والذي يساوي 4 أيضا اذن ($D_2 = 7 - 4 = 3$).
- قيمة المنوال تحسب حسب العلاقة رقم (5-14) كالاتي:

$$Mo = A + \left(\frac{D_1}{D_1 + D_2}\right) \times L$$
$$= 210 + \left(\frac{3}{3+3}\right) \times 5 = 212.5 .$$

ج- طريقة الرافعة

- نجد من الجدول (5-5) ان فئة المنوال هي (210-215) اذن الحد الادنى للفئة هو (A=210) وطول فئة المنوال هو (L=215-210=5).
- نجد ان تكرار الفئة السابقة لها يساوي 4 وتكرار الفئة اللاحقة لها يساوي 4.
- اذن قيمة المنوال حسب العلاقة (5-15) تكون كالاتي:

$$Mo = A + \left(\frac{f_1}{f_1 + f_2}\right) \times L$$
$$= 210 + \left(\frac{4}{4+4}\right) \times 5 = 212.5$$

نلاحظ من النتائج بان الطرائق الثلاثة اعطت نفس القيمة للمنوال وذلك لكون التكرار السابق واللاحق لتكرار فترة المنوال متساويين، ولكن في حالة اختلافهما ستختلف قيمة المنوال قليلا في الطريقتين الاخيرتين عن الطريقة الاولى. وتعد الطريقة الأولى هي الاسهل في الاستعمال، ولكن تعد الطريقة الثانية هي الأفضل والأكثر دقة في تحديد قيمة المنوال لأنها تعطي أهمية للتكرار في الفئة السابقة واللاحقة.

ث – الطريقة البيانية او الهندسية

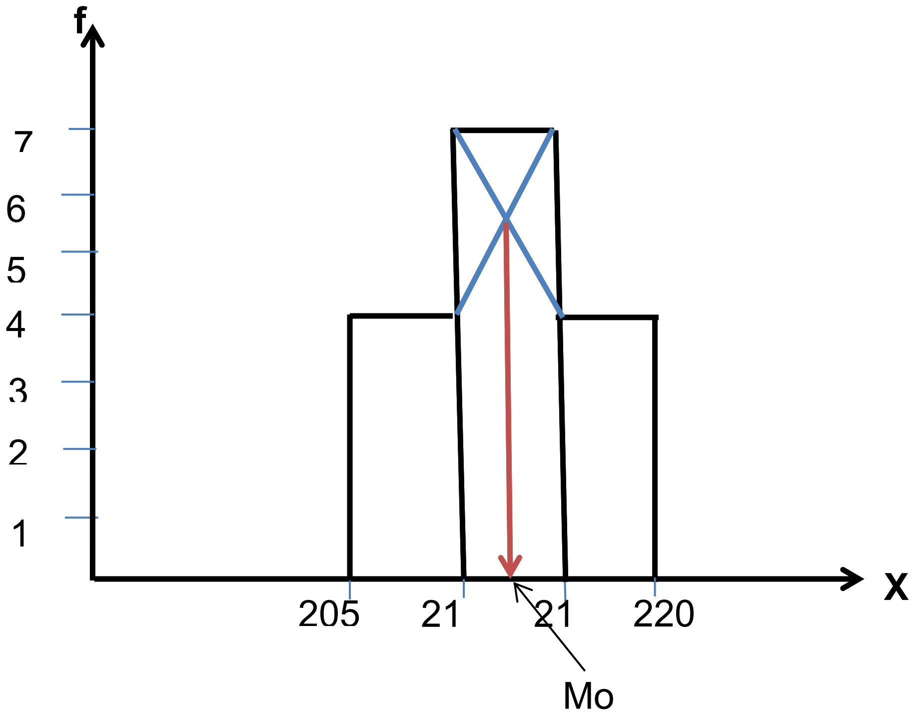
يستعمل المدرج التكراري لإيجاد المنوال بيانيا على وفق الخطوات الاتية:

- ❖ نحدد الفئة المنوالية في المدرج التكراري وهي الفئة ذات التكرار الأكبر (المستطيل الأطول).
- ❖ بعد تحديد الفئة المنوالية نحدد الفئتين السابقة واللاحقة للفئة المنوالية.
- ❖ نرسم خطاً مستقيماً يصل الزاوية اليمنى لقمة مستطيل الفئة المنوالية داخل مستطيل الفئة المنوالية بالزاوية اليمنى لقمة مستطيل الفئة السابقة لها.
- ❖ ونرسم خطاً مستقيماً يصل الزاوية اليسرى لقمة مستطيل الفئة المنوالية داخل مستطيل الفئة المنوالية بالزاوية اليسرى لقمة مستطيل الفئة اللاحق .
- ❖ نرسم عموداً على المحور الافقي عند نقطة تقاطع الخطين داخل مستطيل الفئة المنوالية الذي يمثل حدود الفئة المنوالية ونقطة التقاء الخطين تمثل قيمة المنوال كما موضح في المثال الاتي.

<u>مثال 11</u>

استعمل البيانات في الجدول (5–5) التي تمثل مستوى الكولسترول لمجموعة من الاشخاص لإيجاد قيمة المنوال بيانيا.

<u>الحل:</u> من الجدول رقم (5–5) نرسم المستطيلات الثلاثة التي تمثل الفئة المنوالية والفئة السابقة واللاحقة لها ونتبع الخطوات المذكورة انفاً لنحصل على قيمة المنوال كما هو موضح في الشكل (5–3).



شكل رقم (5-3) المنوال بيانيا من المدرج

❖ **بعض مميزات وعيوب المنوال:**

يعد المنوال من مقاييس النزعة المركزية الشائعة وينماز بانه سهل التعريف والحساب. ويكون المنوال أقل تأثرا من المتوسط بالقيم الشاذة أو المتطرفة، ويمكن حسابه للبيانات الكمية والوصفية (النوعية) وبالرغم من أن المنوال يعد من مقاييس النزعة المركزية الشائعة فإن اهم عيوبه انه لا يأخذ في الاعتبار جميع البيانات، إذا أنه يعتمد فقط على البيانات ذات التكرار الأكثر، وقد لا يوجد منوال لمجموعة من البيانات أو قد يكون هناك أكثر من منوال.

5-3-4 العلاقة بين المقاييس الثلاثة

أولا: لوحظ انه هناك علاقة خطية بين قيم المتوسطات الثلاثة في حالة التوزيعات غير المتماثلة ووحيدة المنوال وهي علاقة ليست دقيقة، ولكنها تقريبية تكون كالاتي:

الوسط الحسابي – المنوال = 3(الوسط الحسابي – الوسيط)

$$(\bar{x} - Mo) = 3(\bar{x} - Me) \tag{16-5}$$

تتمثل الفائدة من هذه العلاقة في امكانية تقدير قيمة اي من المتوسطات التي لا يمكن ايجادها من البيانات مباشرة فعلى سبيل المثال اذا كانت مجموعة من البيانات عديمة المنوال يمكن تقدير قيمته بعد حساب قيمة الوسط الحسابي والوسيط لهذه المجموعة من البيانات، وغيرها من الحالات الاخرى ولتوضيح ذلك نورد المثال الاتي:

مثال 12:

توافرت لديك المعلومات الاتية عن المقاييس لظواهر معينة في المجتمع:

1- إذا كان الوسط الحسابي للوزن = 65 كغم والوسيط = 64.5 اوجد قيمة المنوال.

2- إذا كان الوسط الحسابي لظاهرة الطول = 168 سم والمنوال =168.5 اوجد قيمة الوسيط.

3- إذا كان الوسيط لظاهرة ضغط الدم العالي = 125 وان المنوال = 124.5 اوجد الوسط الحسابي للظاهرة.

الحل:

1- باستعمال العلاقة (16.5) تكون قيمة المنوال لظاهرة الوزن كالاتي:

$$65 - Mo = 3(65 - 64.5)$$
$$\therefore Mo = 65 - 1.5 = 63.5$$

2- باستعمال العلاقة المذكورة انفاً تكون قيمة الوسيط لظاهرة الطول كالاتي:

$$168 - \ 168.5 = 3(168 - Me)$$
$$\therefore Me = \frac{504 + 0.5}{3} = 168.17$$

3- باستعمال العلاقة المذكورة انفاً تكون قيمة الوسط الحسابي لظاهرة ضغط الدم العالي كالاتي:

$$\bar{x} - 124.5 = 3(\bar{x} - 125)$$
$$\therefore 3\bar{x} - \bar{x} = 375 - 124.5$$
$$\therefore \ \bar{x} = \frac{250.5}{2} = 125.25$$

ثانيا: يمكن تحديد نوع التوزيع لمجموعة البيانات من العلاقة بين المتوسطات الثلاثة وبالشكل الاتي:

1- إذا كان الوسط الحسابي = الوسيط = المنوال فان توزيع البيانات متماثل.

$$\bar{x} = Me = Mo \qquad (17\text{-}5)$$

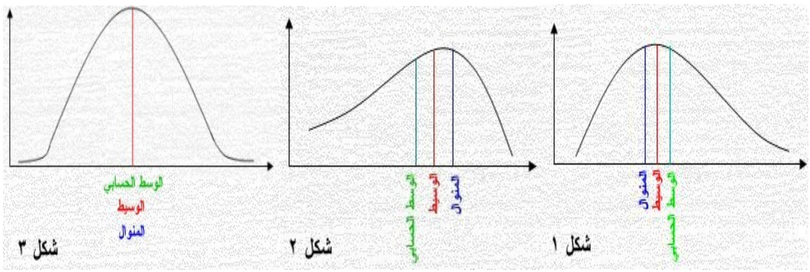
2- إذا كان الوسط الحسابي < الوسيط < المنوال فان توزيع البيانات ملتو نحو اليسار.

$$\bar{x} < Me < Mo \qquad (18\text{-}5)$$

3- إذا كان الوسط الحسابي > الوسيط > المنوال فان توزيع البيانات ملتو نحو اليمين.

$$\bar{x} > Me > Mo \qquad (19\text{-}5)$$

وكما موضحة بالشكل (5-4) وتعتبر هذه العلاقات واحد من الاختبارات السريعة للتعرف على نوع التوزيع للبيانات.



، المنوال

شكل 2: منحنى ملتوي نحو اليسار   شكل 1: منحنى متماثل   شكل 3: منحنى ملتو نحو اليمين

الوسط الحسابي < الوسيط < المنوال   الوسط الحسابي= الوسيط = المنوال   الوسط الحسابي > الوسيط > المنوال

شكل 5-4 نوع التوزيع حسب العلاقة بين مقاييس النزعة المركزية

5-4 حساب مقاييس النزعة المركزية باستعمال البرنامج (SPSS)

عادة عند استعمال الحاسبة لا نحتاج الى تبويب او تلخيص البيانات في جداول تكرارية لغرض حساب مقاييس النزعة المركزية او المقاييس في حالة البيانات الوصفية، بل تستعمل البيانات غير المبوبة لإدخالها الى محرر البيانات في البرنامج (SPSS) ومن ثم حساب المقاييس لها كما هو موضح في الأمثلة الاتية:

مثال 13

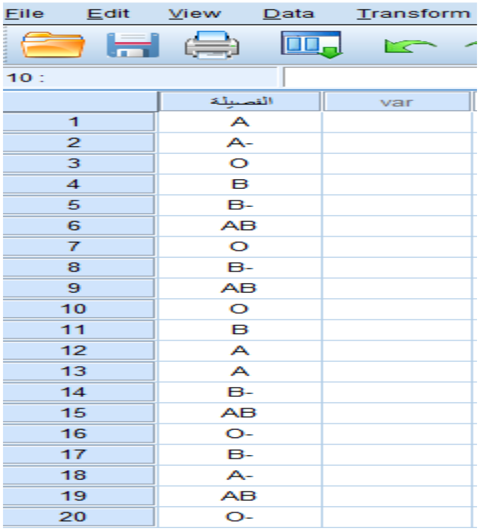
احسب المقاييس المناسبة للبيانات الاتية التي تمثل فصيلة الدم لمجموعة من الأشخاص.

جدول (5–8) فصيلة الدم لمجموعة من الأشخاص

ت	1	2	3	4	5	6	7	8	9	10
فصيلة الدم	A	A–	O	B	B–	AB	O	B–	AB	O
ت	11	12	13	14	15	16	17	18	19	20
فصيلة الدم	B	A	A	B–	AB	O–	B–	A–	AB	O–

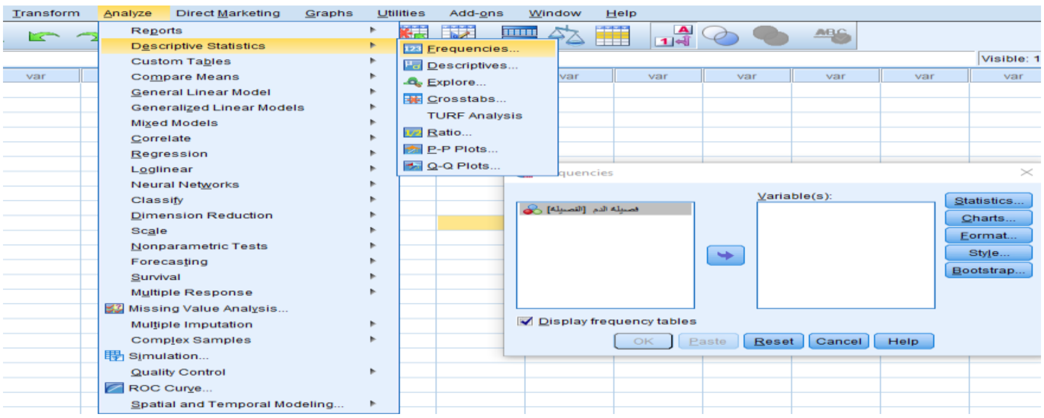
الحل

لحساب المقاييس المناسبة لهذه البيانات ندخلها الى محرر البيانات في برنامج (SPSS) كما في الشكل (5–5).



الشكل (5−5) البيانات في محرر البيانات

وبعد ذلك من خلال استعمال الامر (Analyze) من قائمة الأوامر الرئيسية ومنه نستعمل الامر (Descriptive Statistics) ومنه نستعمل الامر (Frequencies) فنحصل على المربع الحواري كما في الشكل (5−6).



الشكل (5−6) المربع الحواري بعد استعمال الأوامر

وننقل المتغير فصيلة الدم الى جهة اليمين ونضع علامة صح على المربع الخاص بإظهار الجدول التكراري أسفل المربع ومن ثم ننقر (OK) فنحصل على النتائج كما في الجدول (5–9).

جدول (5–9) المقاييس المناسبة للبيانات الوصفية

فصيلة الدم

		Frequency	Percent	Proportion	Cumulative Percent
Valid	A	3	15.0	0.15	15.0
	A−	2	10.0	0.10	25.0
	B	2	10.0	0.10	35.0
	B−	4	20.0	0.20	55.0
	AB	4	20.0	0.20	75.0
	O	3	15.0	0.15	90.0
	O−	2	10.0	0.10	100.0
	Total	20	100.0		

النتائج في الجدول (5–9) تظهر في العمود الأول فصائل الدم والعمود الثاني التكرار، أي عدد الأشخاص لكل فصيلة والعمود الثالث يمثل النسبة المؤوية للتكرار والعمود الرابع يمثل النسبة المطلقة للتكرار، في حين العمود الأخير يمثل النسبة المؤوية التجميعية للتكرار. اما ما يتعلق بالبيانات الكمية فيمكن توضيح كيفية حساب المقاييس المناسبة، كما في المثال الاتي.

مثال 14

البيانات الاتية تمثل مستوى الكوليسترول لدى مجموعة من الأشخاص والمطلوب حساب المقاييس العددية المناسبة باستعمال البرنامج (SPSS).

جدول (5-10) مستويات الكوليسترول لمجموعة من الاشخاص

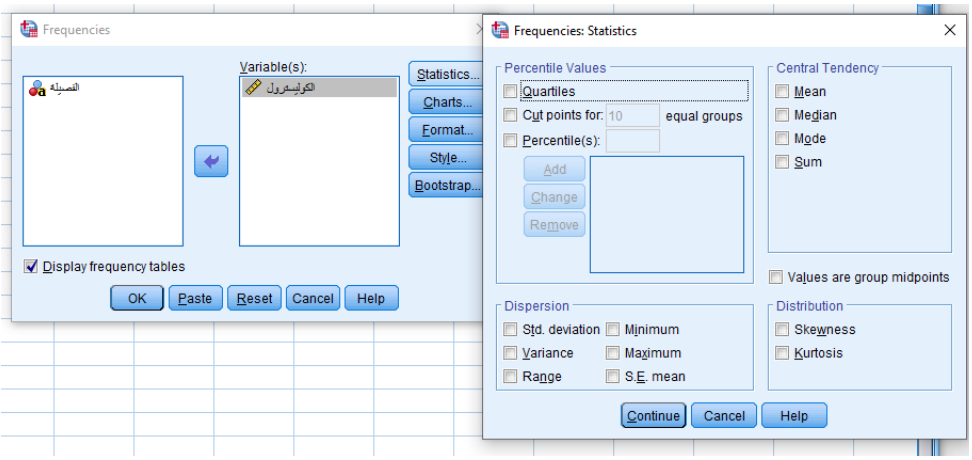
ت	1	2	3	4	5	6	7	8	9	10
الكوليسترول	178	169	170	210	189	192	220	198	230	215
ت	11	12	13	14	15	16	17	18	19	20
الكوليسترول	167	178	183	194	195	176	245	225	216	210

الحل:

يتم ادخال البيانات الى محرر البيانات في البرنامج كما في الشكل (5-7)

الشكل (5-7) ادخال البيانات الكمية الى محرر البيانات

وبعد ادخال البيانات نستخدم الامر (Analyze) من قائمة الأوامر الرئيسة، ومنه نستعمل الامر (Descriptive Statistics) ومن ثم الامر الفرعي (Frequencies) فنحصل على المربع الحواري، ومن المربع الحواري نفتح مربعاً حوارياً أخيراً بعد النقر على كلمة (Statistics) كما في الشكل (5–8).



الشكل (5–8) المربع الحواري لاختيار المقاييس المناسبة

ومن المربع الحواري الأخير نختار المقاييس من المربع (Central Tendency) وبعدها ننقر على كلمة (Continue) للعودة الى المربع الأصلي ونحذف علامة الصح من المربع الذي يمثل اظهار الجدول التكراري ثم ننقر على كلمة (OK) لنحصل على النتائج كما في الجدول (5–11).

جدول (5–11) مقاييس النزعة المركزية لمستوى الكوليسترول

N	Valid	20
Mean		198.00
Median		194.50
Mode		178

تمارين الفصل الخامس

1- عرف المصطلحات الاتية:

أ- النزعة المركزية، ب- الوسط الحسابي، ت - التوزيع التكراري.

2- البيانات الاتية تمثل الوقت المصروف (دقيقة) من الطبيب مع المريض في اثناء مراجعته في العيادة الخارجية لاحد المستشفيات الخاصة:

10، 12، 18، 22، 24، 10، 8، 11، 28، 9، 14، 27، 19، 14، 26، 15، 13، 23، 7، 9، 21، 20، 20، 18، 19، 12، 14، 10.

<u>والمطلوب:</u>

1- حساب مقاييس النزعة المركزية الممكنة لهذه البيانات.

2- كون التوزيع التكراري المناسب لهذه البيانات.

3- حساب مقاييس النزعة المركزية المناسبة للتوزيع التكراري.

3- احسب مقياس النزعة المركزية المناسب لكل مجموعة من البيانات الاتية، ولماذا اختير المقياس؟:

أ- 2، 2، 3، 4، 4، 4، 5، 6.

ب-2، 3، 4، 4، 4، 5، 20.

ت- -5، -4، -3، 0، 3، 4، 5.

ث-1.5، 1.7، 1.8، 1.9، 1.2، 1.3، 1.7، 1.6، 1.7.

4- البيانات الاتية تمثل قياسات ضغط الدم العالي لمجموعة من الاشخاص

-120	-130	-140	-150	-160	-170	-180
	4	9	10	13	8	5

<u>والمطلوب:</u>

حساب مقاييس النزعة المركزية لهذا الجدول التكراري.

5- كان اعداد المراجعين للعيادة الخارجية في أحد المستشفيات حسب ايام الاسبوع كما في الجدول الاتي:

السبت	الاحد	الاثنين	الثلاثاء	الاربعاء	الخميس	الجمعة
23	45	53	64	56	42	18

والمطلوب:

أ- حساب المقاييس المناسبة لهذا الجدول التكراري.

ب- حساب مقاييس النزعة المركزية الممكنة له.

.

6- احسب مقاييس النزعة المركزية المناسبة لكل من المعلومات الاتية:

a- $\sum x_i = 204, \quad \sum x_i^2 = 2099, \quad n = 20$

b- $\sum x_i = 9.78, \quad \sum x_i^2 = 0.97, \quad n = 100$

c- $\sum x_i = -246, \quad \sum x_i^2 = 1254, \quad n = 50$

d- $\sum x_i = 586, \quad \sum x_i^2 = 1844, \quad n = 30$

7- الجدول التكراري الاتي يلخص الوزن (كغم) لخمسين مريض في أحد المستشفيات:

الوزن (كغم)	-50	-55	-60	-65	75-70	المجموع
عدد الاشخاص (f_j)	5	10	20	8	7	50

والمطلوب:

حساب مقاييس النزعة المركزية الممكنة وبالطرق المختلفة.

8- البيانات الاتية تمثل اوزان الاطفال حديثي الولادة:

N	UW	UN	UW	N	UN	UW	UN	UN	UN	N	N
N	UN	UN	UW	N	UN	UW	N	N	UW	UN	UW
UN	UN	N	N	N	UW	N	UN	N	UW	N	UN

اذ ان (N) تمثل الوزن الطبيعي، (UN) تمثل الوزن فوق الطبيعي، (UW) تمثل الوزن اقل من الطبيعي.

والمطلوب: حساب المقاييس المناسبة لهذه المجموعة من البيانات. هل يمكن تمثيلها بيانيا واي الاشكال البيانية يكون مناسب لتمثيلها.

9- قياس الوزن لعينة من 5 اشخاص بالكيلوغرام فكانت (3.3، 4.8، 3.5، 3.7، 3.1) فالوسيط للعينة هو : أ- 3.5 ب- 3.3 ت- 3.7 ث- 3.6 ج- ولا واحد منها.

10- تسجيل فترة حضانة فايروس النقص المناعي لعينة من 7 اشخاص بالسنوات فكانت كالاتي (12.0، 9.5، 13.5، 7.2، 10.5، 6.3، 12.5)

والمطلوب:

1- احسب الوسط الحسابي للعينة. 2- احسب الوسيط للعينة. 3- اذا تم تغيير الرقم 6.3 الى 1.5 ماذا يحدث للوسط الحسابي، الوسيط، وضح هل تزداد قيمهم ام تنقص او لا يحدث تغيير.

4- اذا اضيفت سبعة اوزان لسبعة اشخاص اخرين لتصبح العينة 14 شخص كالاتي (12.0، 9.5، 13.5، 7.2، 10.5، 6.3، 12.5، 14.9، 7.9، 8.1، 5.3، 13.1، 10.7، 6.5) هل تتغير قيم الوسط الحسابي، الوسيط، عن القيم في حالة العينة السابقة وما هو نوع التغيير هل تزداد ام تنقص ام تبقى كما هي.

الفصل السادس

مقاييس التشتت

Dispersion Measurements

6-1 تمهيد

لقد سبق لنا دراسة طرائق عرض البيانات جدولياً وبيانياً والتعرف على أشكالها وتوزيعاتها الإحصائية المختلفة كذلك دراسة مقاييس النزعة المركزية (المتوسطات) وذلك لوصف البيانات عدديًا للبيانات الإحصائية المختلفة. ولكن طرق عرض البيانات وحساب المتوسطات للمجموعات المختلفة من البيانات غير كافٍ للمقارنة بين هذه المجموعات،

ولتوضيح ذلك نأتي بمثال لدراسة قياسات الكولسترول لثلاث مجموعات مختلفة من الطلاب Z وY و X التي كانت قياساتهم كالآتي:

X: 159, 161, 162, 158, 160;
Y : 150, 160, 166, 154, 170
Z : 119, 165, 146, 178, 172

بحساب المتوسط الحسابي للمجموعات الثلاث نجده مساوياً 160 درجة لكل منها، ولكن عند النظر إلى درجات المجموعة الأولى نجدها متقاربة، ودرجات المجموعة الثانية أقل تقاربًا من المجموعة الأولى، والمجموعة الثالثة أقل تقاربا من المجموعة الثانية، أي أن المجموعات الثلاث مختلفة التجانس على الرغم من أن المتوسط الحسابي لكل منها هو نفسه. وبذلك تكون مقاييس النزعة المركزية غير كافية للمقارنة بين طبيعة البيانات الإحصائية، ولهذا نشأت الحاجة إلى إيجاد مقاييس تقيس درجة تجانس (تقارب أو تشتت او تباعد) مفردات البيانات بعضها عن بعض، وهذه المقاييس هي ما تسمى بمقاييس التشتت. والتشتت لا يمكن حسابه في حالة البيانات الوصفية، اما في حالة البيانات الكمية هناك عدة مقاييس لقياس التشتت لمجموعة البيانات الكمية منها المدى (Range) والانحراف المتوسط (Mean Deviation) والتباين (Variance)

والانحراف المعياري (Standard Deviation) ونصف المدى الربيعي (Interquartile Range). ومعامل الاختلاف (التغيير) (Variation Coefficient) ومقاييس الالتواء ((Skewness Measurements ومقاييس التفلطح (Kurtosis Measurements) والمباحث الآتية لتوضيح كيفية حساب هذه المقاييس يدويا مع الأمثلة والمبحث الأخير يوضح كيفية حساب هذه المقاييس باستعمال البرامج الإحصائية.

6-2 المدى Range

يعرف المدى على انه الفرق بين أكبر قيمة وأصغر قيمة لمجموعة البيانات غير المبوبة في جدول توزيع تكراري، فاذا رمزنا لأكبر قيمة في المجموعة بالحرف (A) وأصغر قيمة للمجموعة بالحرف (B) يكون المدى وفق العلاقة رقم (3−1). اما في حالة البيانات المبوبة في جدول توزيع تكراري فان المدى يحسب بواحدة من طريقتين:

1− **طريقة حدود الفئات:** اعتمادا على التعريف المذكور انفاً للمدى فانه يساوي الفرق بين الحد الاعلى للفئة الأخيرة والحد الادنى للفئة الاولى على اساس ان الاول يمثل أكبر قيمة والثاني يمثل أصغر قيمة في مجموعة البيانات.

$$R = L_k - L_1 \qquad (1-6)$$

حيث ان (R) تمثل المدى، (L_1) تمثل الحد الأدنى للفئة الأولى، (L_k) يمثل الحد الأعلى للفئة الأخيرة، (k) يمثل عدد الفئات.

2− **طريقة مراكز الفئات:** احيانا يكون الحد الادنى للفئة الاولى او الحد الاعلى للفئة الاخيرة او كلاهما لا يعودان لقيم المجموعة من البيانات، ولكن ضرورة تكوين الفئات احيانا تتطلب البداية او النهاية في قيم خارج المجموعة، لذلك يفضل استخدام مركز الفئة الاولى كأصغر قيمة في مجموعة البيانات ومركز الفئة الاخيرة هو أكبر قيمة في مجموعة البيانات ومن ثم يكون المدى كالاتي:

$$R = X_k - X_1 \quad (2-6)$$

حيث ان (R) تمثل المدى، (X_1) تمثل الحد الأدنى للفئة الأولى، (X_k) يمثل الحد الأعلى للفئة الأخيرة، (k) يمثل عدد الفئات. وللتوضيح نورد الامثلة الاتية:

<u>مثال 1:</u>

البيانات الاتية تمثل الوزن (كغم) لمجموعة من الاطفال، (9.3، 5.7، 6.3، 3.4، 8.2، 4.5، 5.4، 7.3). اوجد المدى للوزن.

<u>الحل:</u>

باستعمال العلاقة (3−1) نجد أن أكبر قيمة لمجموعة البيانات هي (9.3) وأصغر قيمة في المجموعة هي (3.4) وبالتالي فان المدى يكون كالاتي:

$$R = B - A$$
$$= 9.3 - 3.4 = 5.9$$

<u>مثال 2:</u>

قام طبيب بقياس مستويات الكوليسترول لمجموعة من المرضى الشباب، فوجد ان القياسات كما في الجدول التكراري البسيط الاتي اوجد المدى بالطرق الممكنة.

جدول رقم (6−1) جدول التوزيع التكراري لمستويات الكولسترول

التكرار f_i	مستويات الكولسترول (Mg/ml)
1	195−
3	200−
4	205−
7	210−
4	215−
1	220−225
20	المجموع

الحل:

حسب الطريقة الاولى ومن جدول التوزيع التكراري (6−1) يكون لدينا الحد الادنى للفئة الاولى هو (195) والحد الاعلى للفئة الاخيرة هو (225) ومن ثم يكون المدى حسب الصيغة (6−1) كالاتي:

$$R = L_k - L_1$$
$$= 225 - 195 = 30$$

اما إذا استعملت الطريقة الثانية وحسب العلاقة (6−2) نحتاج الى حساب مركز الفئة الاولى ومركز الفئة الاخيرة وكالاتي:

$$X_1 = \frac{200+195}{2} = 197.5,$$
$$X_k = \frac{220+225}{2} = 222.5$$

ومن ثم فان المدى يكون كالاتي:

$$R = X_k - X_1$$
$$= 222.5 - 197.5 = 25$$

وينماز المدى بانه أسهل مقياس من حيث الحساب، ويعطي فكرة بسيطة عن تباعد البيانات، قد يكون مقياساً غير مناسب لمجموعة البيانات التي تتضمن قيماً متطرفة لأنه يتأثر بالقيم المتطرفة في طرفي البينات الكبيرة او الصغيرة، وأحد استعمالاته لتقييم العينات التي تتضمن مجموعة صغيرة من البيانات، والاستعمال الاخر هو في حسابات السيطرة النوعية لأنها تعتمد على عينات صغيرة متشابهة.

6−3 الانحراف المتوسط Mean Deviation (MD)

هو أحد مقاييس التشتت، ويعبر عنه بمتوسط الانحرافات المطلقة للقيم عن وسطها الحسابي أي هو عبارة عن متوسط انحرافات قيم المجموعة عن وسطها

الحسابي مع إهمال الإشارة، ويعتمد حسابه على نوع البيانات فيما إذا كانت بيانات أولية غير مبوبة في جدول تكراري او إذا كانت البيانات مبوبة في جدول تكراري وكالاتي:

6-3-1 الانحراف المتوسط في حالة البيانات غير المبوبة

إذا توافرت عينة حجمها (n) من المشاهدات $(x_1, x_2, \dots x_n)$ التي تم اخذها من ظاهرة معينة، فان الانحراف المتوسط (*MD*) لها يحسب حسب الخطوات الاتية:

أ- حساب الوسط الحسابي لها حسب المعادلة رقم (5-4) وكالاتي:

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

ب-حساب انحرافات القيم عن وسطها الحسابي واهمال الإشارة حسب القاعدة الاتية:

$$d_i = |x_i - \bar{x}| \qquad (3\text{-}6)$$

ت-حساب الانحراف المتوسط (MD) باستعمال القاعدة الاتية:

$$MD = \frac{\sum_{i=1}^{n} d_i}{n} \qquad (4\text{-}6)$$

ولتوضيح هذه الخطوات نورد المثال الاتي:

<u>مثال 3</u>

البيانات الاتية تمثل قراءات الكوليسترول لمجموعة من المرضى في أحد المختبرات **والمطلوب** حساب التشتت باستعمال الانحراف المتوسط لهذه البيانات

(164 172 198 189 177 168 177)

<u>الحل:</u>

أ- إيجاد الوسط الحسابي للبيانات وكالاتي:

$$\bar{x} = \frac{164 + 172 + 198 + 189 + 177 + 168 + 170}{7}$$

$$= \frac{1238}{7} = 176.9$$

ب- إيجاد الانحرافات المطلقة للبيانات عن الوسط الحسابي وكما في الجدول الاتي:

جدول (6-2) حسابات الانحراف المتوسط

x_i	$x_i - \bar{x}$	$\lvert x_i - \bar{x} \rvert$
164	−12.9	12.9
172	−4.9	4.9
198	21.1	21.1
189	12.1	12.1
177	0.1	0.1
168	−8.9	8.9
170	−6.9	6.9
$\sum \lvert x_i - \bar{x} \rvert$		66.9

ت-حساب الانحراف المتوسط باستعمال المعادلة رقم (6-4) وكالاتي:

$$MD = \frac{66.9}{7} = 9.557$$

وهذا يوضح بان البيانات فيها تشتت عن الوسط الحسابي، أي انها تعتبر غير متجانسة وكلما كبرت قيمة المقياس دل ذلك على تباعد البيانات عن وسطها الحسابي، وكلما كانت قيمة المقياس صغيرة كان ذلك أفضل، وخصوصا عند المقارنة بين المجموعات، اذ تكون المجموعة ذات المقياس الأصغر قيمة هي الأفضل والأكثر تجانسا.

6-3-2 الانحراف المتوسط في حالة البيانات المبوبة في جدول تكراري

لا تختلف خطوات حساب الانحراف المتوسط في حالة البيانات المبوبة في جدول تكراري الا باستعمال القواعد والمعادلات التي تتناسب مع البيانات المبوبة في الحساب وكما يأتي:

أ- حساب الوسط الحسابي $(\bar{x})$ للبيانات المبوبة وفق المعادلة رقم (5-6) بعد إيجاد مراكز الفئات (x_j) حسب المعادلة (5-5).

ب-إيجاد انحرافات مراكز الفئات (x_j) عن الوسط الحسابي $(\bar{x})$ وحسب المعادلة (6-3) ويرمز لها بالرمز (d_j):

ت- إيجاد حاصل ضرب كل انحراف مطلق (بعد اهمال الاشارة) $(|d_j|)$ في التكرار المقابل له (f_j) ومن ثم إيجاد مجموع حاصل ضرب الانحرافات المطلقة بالتكرارات $(\sum |d_j| f_j)$ وكذلك إيجاد مجموع التكرارات $(\sum f_j)$.

ث- حساب الانحراف المتوسط (MD) وفق الصيغة الاتية:

$$MD = \frac{\sum_{j=1}^{k} |d_j| f_j}{\sum_{j=1}^{k} f_j} \qquad (5-6)$$

ولتوضيح هذه الخطوات نورد المثال الاتي:

مثال 5

البيانات الاتية تمثل توزيع مجموعة من الاسر حسب الانفاق على الخدمات الصحية شهريا بمئة ألف دينار عراقي، اوجد الانحراف المتوسط لهذه البيانات.

جدول (6-3) توزيع الاسر حسب الانفاق على الجانب الصحي شهريا

الانفاق	2-	5-	8-	11-	17-14
عدد الاسر	1	8	13	10	8

الحل:

لحساب الانحراف المتوسط يتم تطبيق المعادلة (6−5) وكما في الجدول (6−4) الاتي لحساب مكونات المعادلة:

جدول (6−4) مكونات المعادلة (6−5)

Intervales	f_j	x_j	$x_j f_j$	$\|d_j\| = \|x_j - \bar{x}\|$	$\|d_j\| f_j$
2−	1	3.5	3.5	7.2	7.2
5−	8	6.5	52	4.2	33.6
8−	13	9.5	123.5	1.2	15.6
11−	10	12.5	125	1.8	18
14−17	8	15.5	124	4.8	38.4
$\sum$	40		428		112.8

$$\bar{x} = \frac{428}{40} = 10.7$$

$$MD = \frac{112.8}{40} = 2.82$$

6−3−3 مزايا وعيوب الانحراف المتوسط

من مزايا الانحراف المتوسط أنه يأخذ كل القيم في الاعتبار، ولكن يعاب عليه ما يلي:

1− يتأثر بالقيم الشاذة او المتطرفة لأنه يعتمد على القيم كلها في حسابه.

2- يصعب التعامل معه رياضيا بسبب اهمال الإشارة من جانب ومن جانب اخر إذا كان عدد المشاهدات كبيراً وقيمة الوسط الحسابي تحتوي على كسور عشرية، فان العمليات الحسابية يدويا تصبح صعبة.

6-4 التباين Variance

من الصعب التعامل رياضيا (تحليلياً) مع الانحراف المتوسط ولذلك دعت الحاجة إلى استعمال مقياس للتشتت بقوة الانحراف المتوسط أو أكثر ولكي يكون من السهل التعامل معه تحليلياً، وبما أن الفكرة هي التخلص من الإشارات للانحرافات فان تربيع الانحراف يخلصنا

من الإشارة لنحصل على مقياس يسمى التباين والذي يعرف على انه متوسط مربعات انحرافات جميع القيم عن وسطها الحسابي، ويرمز له في حالة المجتمع بالرمز (σ^2) اما في حالة العينة فيرمز له بالرمز (S^2). ولغرض إعادة المقياس الى الوحدات الاصلية للبيانات قبل التربيع نأخذ الجذر التربيعي للتباين لنحصل على مقياس يسمى الانحراف المعياري والذي بعرف على انه الجذر التربيعي للتباين ويحسب التباين والانحراف المعياري في حالتين هما:

6-4-1 التباين في حالة البيانات غير المبوبة:

لو توافرت عينة بحجم (n) من المشاهدات (x_i) اذ ان $(i = 1,2, \dots, n)$ فيمكن استعمال واحدة من الطرائق الاتية:

1- **طريقة الانحرافات عن الوسط الحسابي:** وتكون حسب الصيغة الاتية:

$$S^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1} \qquad (6-6)$$

او

2– **طريقة القيم الاصلية:** وتكون حسب واحدة من الصيغتين الاتيتين:

$$S^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} \qquad (7\text{–}6)$$

او

$$S^2 = \frac{\sum_{i=1}^{n} x_i^2}{n} - \left(\frac{\sum_{i=1}^{n} x_i}{n}\right)^2 \qquad (8\text{–}6)$$

جميع هذه الصيغ تعطي الناتج نفسه تقريبا والاختلاف هو بطريقة الحساب مع ملاحظة ان الصيغة (6–8) تستعمل في حالة العينات الكبيرة ويختلف مفهوم العينة الكبيرة من مجال الى اخر، ولكن بشكل عام، وهناك راي لأهل الاحصاء يعد العينة كبيرة إذا كان عدد مشاهداتها أكثر من 30 مشاهدة. ولتوضيح تطبيق هذه الصيغ نأخذ المثال الاتي:

مثال 6

البيانات الاتية هي عبارة عن أوزان مجموعة مكونة من ثمانية أشخاص بالكيلوغرام أوجد التباين للوزن (65 ,60 ,55 ,47 ,45 ,48 ,37 ,35)

الحل:

ولغرض تطبيق طريقة الانحرافات عن الوسط الحسابي وحسب الصيغة (6–6) تكون الخطوة الاولى هي حساب الوسط الحسابي للعينة باستعمال الصيغة (5–4) وكالاتي:

$$\therefore\ \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$= \frac{35+37+\cdots+65}{8}$$

$$= \frac{392}{8} = 49$$

بعد ذلك تكون العمليات الحسابية لتطبيق الصيغة (6–6) كما في الجدول الاتي:

جدول (6−5) العمليات الحسابية لتطبيق الصيغة (6−6)

x_i	65	60	55	47	45	48	37	35	المجموع
$x_i-\bar{x}$	16	11	6	−2	−4	−1	−12	−14	
$(x_i-\bar{x})^2$	256	121	36	4	16	1	144	196	774

$$\therefore S^2=\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}=\frac{774}{8-1}=110.57$$

اما عند استعمال طريقة القيم الاصلية وحسب الصيغة (6−7) فأننا لا نحتاج لحساب الوسط الحسابي، بل نحتاج الى العمليات الحسابية الاتية:

جدول (6−6) العمليات الحسابية لتطبيق الصيغة (6−7)

المشاهدات	1	2	3	4	5	6	7	8	المجموع
x_i	65	60	55	47	45	48	37	35	392
x_i^2	4225	3600	3025	2209	2025	2304	1369	1225	19982

$$\therefore S^2=\frac{\sum_{i=1}^{n}x_i^2-\frac{(\sum_{i=1}^{n}x_i)^2}{n}}{n-1}$$

$$=\frac{19982-\frac{(392)^2}{8}}{8-1}$$

$$=\frac{19982-19208}{7}=110.57$$

يلاحظ من النتائج بان الطريقتين أعطتنا النتيجة نفسها للتباين، اما إذا كان حجم العينة كبيرا ($n>30$) فنستعمل الصيغة الثالثة (6−8) لإيجاد التباين وباستعمال الخطوات نفسها في الطريقتين السابقتين.

6−4−2 التباين في حالة البيانات المبوبة:

توجد أربع طرق لحساب التباين في حالة البيانات المبوبة في جدول توزيع تكراري نختار منهما طريقتان وبالشكل الاتي:

1- **طريقة الانحرافات عن الوسط الحسابي:** وتكون وفق الخطوات الاتية:

أ- ايجاد مراكز الفئات وفق القاعدة (مركز الفئة (x_j)= (الحد الاعلى+ الحد الادنى للفئة)/2).

ب- ايجاد الوسط الحسابي $(\bar{x})$ وحسب الصيغة (6-5).

ت- ايجاد انحرافات مراكز الفئات عن الوسط الحسابي $(x_j - \bar{x})$ ومن ثم ايجاد مربعات الانحرافات $((x_j - \bar{x})^2)$.

ث- ايجاد حاصل ضرب مربعات الانحرافات $(x_j - \bar{x})^2$ في التكرارات (f_j) لكل فئة.

ج- ايجاد مجموع التكرارات لجميع الفئات $(\sum_{j=1}^{k} f_j)$. اذ ان (k) تمثل عدد الفئات، و$(j = 1,2, \dots , k)$.

ح- ايجاد التباين (S^2) على وفق الصيغة الاتية:

$$S^2 = \frac{\sum_{j=1}^{k}(x_j-\bar{x})^2 f_j}{\sum_{j=1}^{k} f_j} \qquad (9-6)$$

2- **طريقة القيم الاصلية:** ايجاد التباين على وفق الخطوات الاتية:

أ- ايجاد مراكز الفئات على وفق القاعدة (مركز الفئة (x_j)= (الحد الاعلى+ الحد الادنى للفئة)/2).

ب- ايجاد حاصل ضرب مراكز الفئات (x_j) في التكرارات (f_j) المقابلة لها ونوجد مجموع حاصل الضرب $(\sum_{j=1}^{k} x_j f_j)$.

ت- ايجاد مربعات مراكز الفئات (x_j^2) ومن ثم حاصل ضرب المربعات في التكرارات (f_j) وبعدها ايجاد مجموع حاصل الضرب $(\sum_{j=1}^{k} x_j^2 f_j)$.

ث- ايجاد مجموع التكرارات لجميع الفئات $(\sum_{j=1}^{k} f_j)$.

ج- ايجاد التباين (S^2) وفق الصيغة الاتية:

$$S^2 = \frac{\sum_{j=1}^{k} x_j^2 f_j}{\sum_{j=1}^{k} f_j} - \left(\frac{\sum_{j=1}^{k} x_j f_j}{\sum_{j=1}^{k} f_j}\right)^2 \qquad (10-6)$$

ولتوضيح هذه الصيغ وكيفية إيجاد التباين في حال البيانات المبوبة يمكن استعمل المثال (7). وبعد إيجاد التباين نحتاج الى إيجاد مقياس يقيس التشتت بوحدات البيانات نفسها لان التباين يقيس التشتت بوحدات مربعة للبيانات، ولهذا نحتاج الى مقياس يسمى الانحراف المعياري كما في المبحث الفرعي الاتي.

6-4-3 الانحراف المعياري Standard deviation:

ولغرض إعادة الوحدات الى الوحدات الاصلية للبيانات يسمى الجذر التربيعي للتباين بالانحراف المعياري (S) وباي طريقة يحسب التباين (S^2) اي:

$$S = \sqrt{S^2} \qquad (11-6)$$

بما ان التباين يقاس بمربع الوحدات للبيانات الاصلية، لذلك يفضل استعمال الانحراف المعياري لأنه يقاس بوحدات البيانات الاصلية نفسها لأنه الجذر التربيعي للتباين، ويجب التذكر بان التباين والانحراف المعياري قيم موجبة دائما. ولتوضيح كيفية حساب هذه المقاييس نأخذ المثال الاتي:

<u>مثال 7</u>

البيانات الاتية تمثل الوقت المستغرق (دقيقة) للفحص في إحدى العيادات الخارجية لعينة تتكون من 20 من المراجعين، **<u>والمطلوب:</u>** حساب التباين والانحراف المعياري للوقت المستغرق (دقيقة).

جدول (6-7) الوقت المستغرق دقيقة للفحص في العيادة الخارجية

الفئات	10-	14-	18-	22-	26-30
التكرار (f_j)	3	4	8	3	2

<u>الحل:</u>

1- طريقة الانحرافات عن الوسط الحسابي:

وتكون خطوات الحل كما في الجدول الاتي:

جدول (6-8) العمليات الحسابية لطريقة الانحرافات عن الوسط الحسابي

الفئات	f_j	x_j	$x_j f_j$	$x_j - \bar{x}$	$(x_j - \bar{x})^2$	$(x_j - \bar{x})^2 f_i$
10−	3	12	26	−6.9	47.61	142.83
14−	4	16	64	−2.9	8.41	33.64
18−	8	20	160	1.1	1.21	9.68
22−	3	24	72	5.1	26.01	78.03
26−30	2	28	56	9.1	82.81	165.62
Sum	20		378			429.8

$$\bar{x} = \frac{\sum_{j=1}^{k} x_j f_j}{\sum_{j=1}^{k} f_j} = \frac{378}{20} = 18.9$$

وباستعمال الصيغة (10.6) تكون قيمة التباين كالاتي:

$$S^2 = \frac{\sum_{j=1}^{k} (x_j - \bar{x})^2 f_j}{\sum_{j=1}^{k} f_j} = \frac{429.8}{20} = 21.49$$

والانحراف المعياري (S) سيكون وحسب الصيغة (6-11) كالاتي:

$$S = \sqrt{S^2} = \sqrt{21.49} = 4.636$$

2- طريقة القيم الاصلية

تكون خطوات حساب التباين كما في الجدول الاتي:

جدول (6–9) العمليات الحسابية حسب طريقة القيم الاصلية

الفئات	f_j	x_j	$x_j f_j$	x_j^2	$x_j^2 f_j$
10−	3	12	26	144	432
14−	4	16	64	256	1024
18−	8	20	160	400	3200
22−	3	24	72	576	1728
26−30	2	28	56	784	1568
Sum	20		378		7952

وبتطبيق الصيغة (6–10) تكون قيمة التباين كالاتي:

$$S^2 = \frac{\sum_{j=1}^{k} x_j^2 f_j}{\sum_{j=1}^{k} f_j} - \left(\frac{\sum_{j=1}^{k} x_j f_j}{\sum_{j=1}^{k} f_j}\right)^2$$

$$= \frac{7952}{20} - \left(\frac{378}{20}\right)^2 = 40.39$$

والانحراف المعياري بتطبيق الصيغة (6–11) يكون مساوياً الى:

$$S = \sqrt{S^2} = \sqrt{40.39} = 6.355$$

ويستعمل التباين والانحراف المعياري لقياس تشتت البيانات اي تباعدها عن المركز وعادة للمقارنة بين مجموعات البيانات لتحديد من هي المجموعة الاكثر تباينا، تكون أكثر تشتتا من المجموعات الاخرى. وهنالك استعمالات اخرى للمقياسين خصوصا في مجال الاستدلال الاحصائي.

6–4–4 بعض خصائص التباين والانحراف المعياري

1– عند أضافة أو طرح مقدار ثابت (a) الى او من جميع المشاهدات لاي مجموعة من البيانات فان التباين والانحراف المعياري للقيم الجديدة هو نفسه التباين والانحراف المعياري للقيم الأصلية.

2– عند ضرب جميع المشاهدات في مقدار ثابت (a) أو قسمتها على مقدار ثابت فان التباين والانحراف المعياري يتأثران بذلك، ففي حالة الضرب يكون

التباين للقيم الاصلية يساوي التباين للقيم الجديدة مضروبا في مربع الثابت، اما الانحراف المعياري للمشاهدات الأصلية يساوي الانحراف المعياري للمشاهدات الجديدة مضروبا في المقدار الثابت، والعكس في حالة القسمة حيث يقسم التباين على مربع الثابت والانحراف المعياري يقسم على قيمة الثابت.

3- مجموع الانحرافات عن الوسط الحسابي تساوي صفراً لاي مجموعة من البيانات اما مجموع مربعات الانحراف للمشاهدات عن وسطها الحسابي $(\bar{x})$ تكون أصغر من مجموع مربعات الانحراف للمشاهدات عن أي وسط فرضي آخر (d) بشرط ان $(d \neq \bar{x})$.

4- الانحراف المعياري لمجموعة من البيانات أكبر من الانحراف المتوسط لها.

5- إذا كانت هناك عينتان حجمهما هو (n_2, n_1) وتباينهما هو على الترتيب (S_2^2, S_1^2) ولهما المتوسط $(\bar{x})$ نفسه فإن التباين المشترك (Pooled variance) لهما هو:

$$S_p^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-1} \qquad (12-6)$$

يمكن اثبات هذه الخصائص من خلال البرهان الرياضي، ولكن أقرب للطالب نستعمل البيانات كمثال لإثبات هذه الخصائص وكما في المثال الاتي.

<u>مثال 8</u>

البيانات الاتية تمثل اعمار عينة من 8 من الأطفال في المدرسة الابتدائية (7، 8، 9، 7 ، 12 ، 8 ، 7 ، 6) اثبت الخصائص المذكورة انفاً.

<u>الحل:</u>

باستعمال المعادلة (6−7) لحساب التباين يحسب أولا الوسط الحسابي حسب الصيغة (6−5)

$$\bar{x} = \frac{64}{8} = 8$$

جدول (6−10) حسابات التباين والانحراف المعياري للقيم الاصلية

x_i	7	8	9	7	12	8	7	6	$\sum = 64$
$x_i - \bar{x}$	−1	0	1	−1	4	0	−1	−2	
$(x_i - \bar{x})^2$	1	0	1	1	16	0	1	4	$\sum = 24$

$$S^2 = \frac{\sum(x_i - \bar{x})^2}{n-1} = \frac{24}{8-1} = \frac{24}{7} = 3.429$$

$$S = \sqrt{3.429} = 1.852$$

1−نفرض بانه تم إضافة الثابت (a=5) للبيانات فتصبح القيم الجديدة (12، 13، 14، 12، 17، 13، 12، 11) فان حسابات التباين والانحراف المعياري للقيم الجديدة تكون على وفق الجدول الاتي:

2−

جدول (6−11) حسابات التباين والانحراف المعياري بعد إضافة الثابت

x_i	12	13	14	12	17	13	12	11	$\sum = 104$
$x_i - \bar{x}$	−1	0	1	−1	4	0	−1	−2	
$(x_i - \bar{x})^2$	1	0	1	1	16	0	1	4	$\sum = 24$

$$\bar{x} = \frac{104}{8} = 13$$

$$S^2 = \frac{\sum(x_i - \bar{x})^2}{n-1} = \frac{24}{8-1} = \frac{24}{7} = 3.429$$

$$S = \sqrt{3.429} = 1.852$$

النتائج في الجدول (6−11) تظهر بان التباين والانحراف المعياري لم يتأثر بإضافة أي قيمة ثابتة لمجموعة البيانات، ونترك للطالب اثبات حالة طرح مقدار ثابت كتمرين.

3–نفرض بانه هذه البيانات ضربت بالثابت (a=4) فتصبح القيم الجديدة (28، 32، 36، 28، 48، 32، 28، 24) وباستعمال المعادلة (6–6) لحساب التباين والانحراف المعياري للقيم الجديدة تكون النتائج كما يلي:

جدول (6–12) حسابات القيم الجديدة بعد ضرب بالثابت

x_i	28	32	36	28	48	32	28	24	$\sum = 256$
$x_i - \bar{x}$	−4	0	4	−4	16	0	−4	−8	
$(x_i - \bar{x})^2$	16	0	16	16	256	0	16	64	$\sum = 384$

$$\bar{x} = \frac{256}{8} = 32$$

$$S^2 = 384/7 = 54.857$$

$$= (4)^2 \times 3.429$$

$$S = \sqrt{54.857} = 7.407 = 4 \times 1.852$$

النتائج في الجدول (6–12) تظهر بان قيم التباين والانحراف المعياري تتأثر بحالة الضرب، فلهذا نجد بان قيمة التباين بعد الضرب تساوي مربع الثابت مضروب في قيمة التباين الأصلي وقيمة الثابت مضروبة في قيمة الانحراف المعياري الأصلي ونترك عملية القسمة للطالب كتمرين لإثبات انها تؤثر على القيم الاصلية للبيانات.

4– لإثبات الخاصية الثالثة كانت البيانات في الجدول (6–10) تشير بان مجموع البيانات كان يساوي 64 ومنه كان الوسط الحسابي يساوي 8 ومن ثم فان النتائج في الجدول (6–13) تثبت الخاصية الثالثة مع فرض بان الوسط الفرضي يساوي (c=7) وكما يلي:

جدول (6−13) حسابات اثبات الخاصية الثالثة

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$x_i - c$	$(x_i - c)^2$
7	−1	1	0	0
8	0	0	1	1
9	1	1	2	4
7	−1	1	0	0
12	4	16	5	25
8	0	0	1	1
7	−1	1	0	0
6	−2	4	−1	1
Total	0	24	8	32

يلاحظ من النتائج في الجدول (6−13) بان مجموع انحرافات القيم عن وسطها الحسابي يساوي صفراً في حين مجموعها عن الوسط الفرضي لا يساوي صفراً. أيضا يلاحظ بان مجموع مربعات الانحرافات عن الوسط الحسابي (24) هي أصغر من مجموع مربعات الانحرافات عن الوسط الفرضي (32).

5− لإثبات الخاصية الرابعة يلاحظ بان البيانات في الجدول (6−10) تشير بان مجموع البيانات كان يساوي 64 ومنه كان الوسط الحسابي يساوي 8، وكانت قيمة الانحراف المعياري يساوي (1.852) ومن ثم فان النتائج في الجدول (6−14) لحساب قيمة الانحراف المتوسط.

جدول (6−14) حسابات الانحراف المتوسط

x_i	7	8	9	7	12	8	7	6	Sum
$\lvert x_i - \bar{x} \rvert$	1	0	1	1	4	0	1	2	10

وباستعمال المعادلة (6−4) نجد بان قيمة الانحراف المتوسط تساوي

$$MD = \frac{\sum_{i=1}^{n} |d_i|}{n} = \frac{10}{8} = 1.25$$

ومن ثم فان النتائج في الجدول (6−14) تشير بان قيمة الانحراف المتوسط تساوي (1.25) وهي أصغر من قيمة الانحراف المعياري (1.852).

6− لإثبات الخاصية الخامسة نفرض بأن اختيار عينة من المستشفى الأول بحجم (10) مرضى وكان التباين لقياسات الكولسترول لهم يساو (168) مليمول/لتر تربيع واختيار عينة أخرى من مستشفى اخر بحجم 12 مريض وكان التباين لقياسات الكولسترول لهم تساوي (176) مليمول/ لتر تربيع. ما هو التباين المشترك للقياسات إذا علمت ان المتوسط للعينتين هو واحد. باستعمال المعادلة (6−12) يكون التباين المشترك يساوي

$$S_p^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-1} = \frac{(10-1)168+(12-1)176}{10+12-1} = 164.191 \text{ mm/L}$$

6−5 نصف المدى الربيعي (الانحراف الربيعي) Interquartile Range

توجد أحيانا قيم متطرفة اما صغيرة او كبيرة ضمن مجموعة البيانات، وكما عرفنا بان القيم المتطرفة تؤثر في قيم مقاييس التشتت السابقة لذلك ظهرت الحاجة لحساب مقياس لا يتأثر بهذه القيم وهذا المقياس يسمى المدى الربيعي، ويعرف بانه الفرق بين الربيع الثالث والربيع الأول، ومن ثم لمعرفة هذا المقياس يجب معرفة الربيعات وكيفية حسابها، ولهذا فان الربيع الاول يعرف على انه القيمة التي يسبقها ربع(4/1) البيانات(اي تكون هذه القيم اصغر منها) بعد ترتيبها تصاعديا، اما الربيع الثالث فانه القيمة التي يسبقها ثلاثة ارباع (4/3) القيم (اي تكون اصغر منها) ويليها او يقع بعدها ربع القيم (اي اكبر منها)، وبعبارة اخرى بعد ترتيب البيانات تصاعديا نقسمها الى اربعة اجزاء فالقيمة التي تقع بعد الجزء الاول تسمى بالربيع الاول ونرمز له (Q_1) والقيمة التي تقع بعد ثلاثة ارباع القيم تسمى الربيع الثالث ونرمز له (Q_3). ومن هنا

فان نصف المدى الربيعي او الانحراف الربيعي هو قسمة المدى الربيعي على 2 ولحساب الانحراف الربيعي نتبع الخطوات الاتية وحسب نوع البيانات:

6-5-1 حالة البيانات غير المبوبة (المفردة): نتبع الخطوات الاتية لحساب نصف المدى الربيعي:

أ- ترتيب البيانات تصاعديا من الاصغر الى الاكبر.

ب- ايجاد ترتيب الربيع الاول (RQ_1) ويكون وفق الصيغة الاتية:

$$RQ_1 = \frac{n}{4} \qquad (13-6)$$

ت- ايجاد قيمة الربيع الاول (Q_1) وهي القيمة التي تقابل ترتيب الربيع الاول إذا كان ناتج القسمة عدداً صحيحاً اما إذا كان ناتج القسمة يتضمن كسراً فهذا يعني بان قيمة الربيع الاول تساوي الوسط للقيمتين قبل ترتيب الربيع الاول وبعده.

ث- ايجاد ترتيب الربيع الثالث (RQ_3) ويكون على وفق الصيغة الاتية:

$$RQ_3 = \frac{3n}{4} \quad \text{او} \quad RQ_3 = 3RQ_1 \qquad (14-6)$$

ج- ايجاد قيمة الربيع الثالث (Q_3) وهو القيمة التي تقابل ترتيب الربيع الثالث إذا كان ناتج القسمة صحيحاً، اما إذا كان ناتج القسمة كسراً فان قيمة الربيع الثالث تساوي متوسط القيمة قبل الترتيب والقيمة بعد الترتيب.

ح- ايجاد قيمة نصف المدى الربيعي او الانحراف الربيعي (IQR) وتكون على وفق الصيغة الاتية:

$$IQR = \frac{Q_3 - Q_1}{2} \qquad (15-6)$$

ملاحظة: الربيع الثاني (Q_2) والذي يقع عند منتصف القيم يساوي قيمة الوسيط.

ولتوضيح هذه الخطوات نورد المثال الاتي:

مثال 9:

البيانات الاتية تمثل درجات الحرارة لاحد المرضى الراقدين في أحد المستشفيات بعد اجراء عملية جراحية ولمدى اسبوعين، اوجد الانحراف الربيعي لدرجات الحرارة:

(36.7 ،37.5 ،39 ،38 ،37 ،35 ،31 ،30 ،29 ،38 ،32 ،34 ،35،36)

الحل:

أ- ترتيب البيانات تصاعديا كالاتي:

(39 ،38 ،38 ،37.5 ،37 ،36.7 ،36 ،35 ،35 ،34 ،32 ،31 ،30 ،29)

ب- ايجاد ترتيب الربيع الاول حسب الصيغة (6-13) وكالاتي:

$$RQ_1 = \frac{n}{4} = \frac{14}{4} = 3.5$$

ت- ايجاد القيمة للربيع الاول: بما ان الترتيب يتضمن كسراً يعني ان قيمة الربيع الاول هو المتوسط للقيمتين الثالثة والرابعة وعليه فان:

$$Q_1 = \frac{31+32}{2} = 31.5$$

ث- ايجاد ترتيب الربيع الثالث حسب الصيغة (14.6) وكالاتي:

$$RQ_3 = \frac{3n}{4} = \frac{3(14)}{4} = 10.5$$

ج- ايجاد قيمة الربيع الثالث وبما ان قيمة الترتيب كانت تتضمن كسراً لذلك فان قيمة الربيع الثالث تكون عبارة عن المتوسط بين القيمتين العاشرة والحادية عشرة وكالاتي:

$$Q_3 = \frac{37+37.5}{2} = 37.25$$

ح- حساب الانحراف الربيعي (نصف المدى الربيعي): وحسب الصيغة (15.6) وكالاتي:

$$IQR = \frac{Q_3 - Q_1}{2}$$
$$= \frac{37.25-31.5}{2} = 34.38.$$

6-5-2 حالة البيانات المبوبة في جدول تكراري:

توجد طريقتان لإيجاد نصف المدى الربيعي في حالة البينات المبوبة في جدول تكراري هما الطريقة الرياضية والطريقة البيانية، والطريقتان تعملان بالفكرة نفسها في حالة البيانات غير المبوبة مع اختلاف الخطوات لاختلاف نوع البيانات.

اولا: الطريقة الرياضية: وتكون حسب الخطوات الاتية؟

أ- ترتيب البيانات تصاعديا، وفي حالة البيانات المبوبة، فان الترتيب يعني تكوين جدول تكراري متجمع صاعد.

ب- ايجاد ترتيب الربيع الاول (RQ_1) ويكون وفق الصيغة الاتية:

$$RQ_1 = \frac{\sum_{j=1}^{k} f_j}{4} \qquad (16-6)$$

ت- تحديد فترة الربيع الاول على الجدول التكرار المتجمع الصاعد وحسب الترتيب في الخطوة السابقة ومنها يتم تحديد الحد الادنى للفترة ويرمز له (A_1) ومن ثم تحديد التكرار المتجمع الصاعد السابق لترتيب الربيع الاول ويرمز له (f_{Q1-1}) وتحديد التكرار المتجمع الصاعد اللاحق لترتيب الربيع الاول ويرمز له (f_{Q1+1}) او يمكن الاستعاضة عنه بتكرار فئة الربيع الاول الاصلي ويرمز له (f_{Q1}) وبعدها تحديد طول فئة الربيع الاول ويرمز له (L_{Q1}) ومن هذه المقادير يمكن ايجاد قيمة الربيع الاول حسب الصيغة الاتية:

$$Q_1 = A_1 + \frac{RQ_1 - f_{Q1-1}}{f_{Q1+1} - f_{Q1-1}} \times L_{Q1} \qquad (17-6)$$

ث- ايجاد ترتيب الربيع الثالث (RQ_3) ويكون وفق الصيغة الاتية:

$$RQ_3 = \frac{3\sum_{j=1}^{k} f_j}{4} \quad \text{او} \quad RQ_3 = 3RQ_1 \qquad (18-6)$$

ج- تحديد فترة الربيع الثالث على الجدول التكرار المتجمع الصاعد وحسب الترتيب في الخطوة السابقة ومنها يحدد الحد الادنى للفترة ويرمز له (A_3) ومن ثم تحديد التكرار المتجمع الصاعد السابق لترتيب الربيع الثالث ويرمز له (f_{Q3-1}) وتحديد

التكرار المتجمع الصاعد اللاحق لترتيب الربيع الثالث ويرمز له (f_{Q3+1}) او يمكن الاستعاضة عنه بتكرار فئة الربيع الثالث الاصلي ويرمز له (f_{Q3}) وبعدها تحديد طول فئة الربيع الثالث ويرمز له (L_{Q3}) ومن هذه المقادير يمكن ايجاد قيمة الربيع الاول حسب الصيغة الاتية:

$$Q_3 = A_3 + \frac{RQ_3 - f_{Q3-1}}{f_{Q3+1} - f_{Q3-1}} \times L_{Q3} \qquad (19-6)$$

ح- حساب نصف المدى الربيعي او الانحراف الربيعي حسب العلاقة (6-15).

ثانيا: الطريقة البيانية: وتتضمن الخطوات الاتية:

أ- ترتيب البيانات تصاعديا من خلال تكوين جدول التكرار المتجمع الصاعد.

ب- رسم المنحنى المتجمع الصاعد كما مر معنا في الفصل السابق.

ت-ايجاد ترتيب الربيع الاول وترتيب الربيع الثالث كما مر في الطريقة الرياضية المذكورة آنفاً وحسب الصيغ (6-16) و(6-17).

ث-تحديد ترتيب الربيع الاول والثالث على المحور العمودي الذي يمثل التكرار المتجمع الصاعد ومنهما نرسم خطاً مستقيماً موازياً للمحور الافقي، ومن نقطة تقاطعهما نرسم خطاً مستقيماً على المحور الافقي لنحصل على قيمتي الربيع الاول والثالث.

ج-حساب قيمة الانحراف الربيعي وحسب العلاقة (6-15) من خلال قيمتي الربيع الاول والثالث التي حددت في الخطوة السابقة.

ولتوضيح هذه الخطوات في الطريقتين، نورد المثال الآتي:

<u>مثال 10</u>

الجدول التكراري الاتي يمثل اعمار مجموعة من المرضى المراجعين لاحد العيادات الخارجية خلال أحد الايام في الاسبوع، والمطلوب حساب الانحراف الربيعي بالطريقتين الرياضية والبيانية.

جدول (6−15) بيانات اعمار مجموعة من المراجعين لاحد العيادات الخارجية

فئات العمر	-5	-15	-25	-35	-45	65−55	المجموع
عدد الاشخاص(f_j)	5	10	20	22	13	3	73

الحل: أولا: حسب الطريقة الرياضية نتبع الخطوات الاتية:

أ- ترتيب البيانات تصاعديا من خلال ايجاد الجدول التكراري المتجمع الصاعد وكما في الجدول (6−16):

جدول (6−16) التكرار المتجمع الصاعد

الحدود العليا للفئات	التكرار المتجمع الصاعد
اقل من 15	5
اقل من 25	15
اقل من 35	35
اقل من 45	57
اقل من 55	70
اقل من 65	73

ب- تحديد ترتيب الربيع الاول وحسب العلاقة (6−16) وكالاتي:

$$RQ_1 = \frac{\sum_{j=1}^{k} f_j}{4} = \frac{73}{4} = 18.25$$

ت- من الترتيب اعلاه تحديد فترة الربيع الاول وتكون (25−35) ومنها نجد ان:

$$A_1 = 25,\ \ f_{Q1-1} = 15,\ \ f_{Q1+1} = 35,\ \ L_{Q1} = 10$$

ث- حساب قيمة الربيع الاول حسب الصيغة (6−17) وكالاتي:

$$Q_1 = A_1 + \frac{RQ_1 - f_{Q1-1}}{f_{Q1+1} - f_{Q1-1}} \times L_{Q1}$$

$$= 25 + \frac{18.25 - 15}{35 - 15} \times 10 = 26.625$$

ج- حساب ترتيب الربيع الثالث وحسب العلاقة (6-18) وكالاتي:

$$RQ_3 = 3RQ_1 = 3(18.25) = 54.75$$

ح- من الترتيب المذكور آنفاً نجد ان فئة الربيع الثالث هي (35-45) ومنها نجد ان:

$$A_3 = 35,\ \ f_{Q3-1} = 35,\ \ f_{Q3+1} = 57,\ \ L_{Q3} = 10$$

خ- حساب الربيع الثالث وحسب العلاقة (6-19) وكالاتي:

$$Q_3 = A_3 + \frac{RQ_3 - f_{Q3-1}}{f_{Q3+1} - f_{Q3-1}} \times L_{Q3}$$

$$= 35 + \frac{54.75 - 35}{57 - 35} \times 10 = 43.977$$

د- بعد تحديد قيمتي الربيع الاول والثالث نستعمل الصيغة (6-15) وكالاتي:

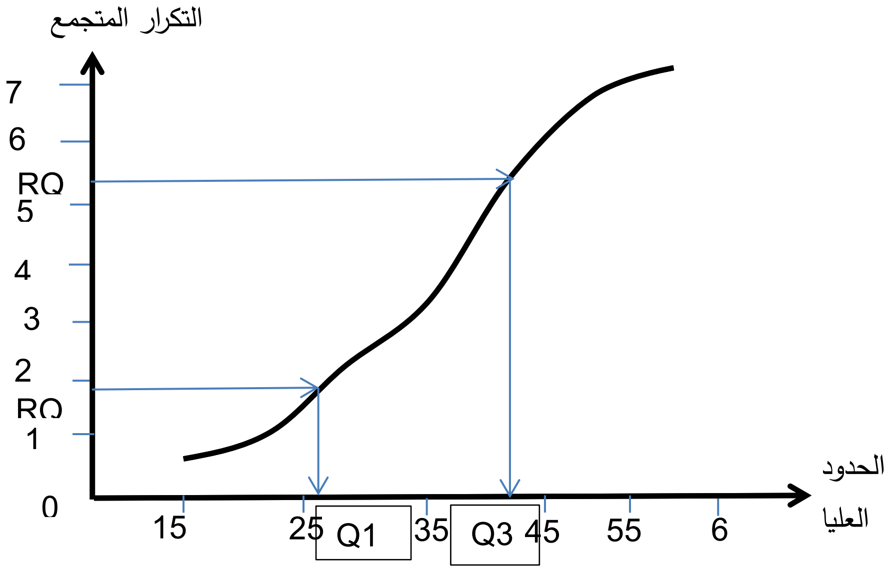
$$IQR = \frac{Q_3 - Q_1}{2} = \frac{43.977 - 26.625}{2} = 8.676$$

ثانيا: الطريقة البيانية: وتكون وفق الخطوات الاتية:

أ- ترتيب البيانات تصاعديا، اي تكوين جدول التكرار المتجمع الصاعد كما في الجدول المذكور آنفاً.

ب-ايجاد ترتيب الربيع الاول كما في الحل السابق ($RQ_1 = 18.25$) وكذلك ترتيب الربيع الثالث ($RQ_3 = 54.75$).

ت-رسم المنحنى المتجمع الصاعد وكما في الشكل الاتي:



الشكل $(6-1)$ إيجاد الانحراف الربيعي من المنحنى المتجمع الصاعد

ث- بعد تحديد القيم التقريبية للربيع الاول (Q_1) والربيع الثالث (Q_3) على الرسم تستخدم هذه القيم لحساب قيمة الانحراف الربيعي حسب الصيغة (16.6).

وفضلاً عن هذه المقاييس هناك مقاييس اخرى تسمى العشيرات، وتعتمد على تقسيم البيانات الى عشرة اجزاء بدلا من أربعة، وكذلك المئينات والتي تعتمد على تقسيم مجموعة البيانات الى مئة جزء.

6-6 معامل الاختلاف Coefficient of variation

عندما نرغب في المقارنة بين مجموعتين من المشاهدات من حيث تجانسهما، أي قرب البيانات من وسطها الحسابي نستعمل الانحراف المعياري لأنه يكون مقاسا بالوحدات الاصلية للمشاهدات فعلى سبيل المثال إذا توافرت عينتان الأولى تمثل قياسات درجة الحرارة فتكون مقاسة بالدرجة المئوية وعينة أخرى تمثل قياسات الكوليسترول يكون مقاسا بالملي مول/ لتر، فهنا لا يمكن استعمال الانحراف المعياري للمقارنة لأن وحدات القياس مختلفة، لذا دعت الحاجة الى إيجاد مقياس لا يعتمد على الوحدات وهذا المقياس يسمى بمعامل الاختلاف او معامل الاختلاف النسبي (التشتت النسبي) وهو نسبة او نسبة مئوية بين الانحراف المعياري والوسط الحسابي لأنه أيضا يكون مقاساً

بالوحدات الاصلية للمشاهدات او باستعمال الربيعات ومن ثم تحذف الوحدات بين البسط والمقام وتوجد صيغتان لحساب معامل الاختلاف كالاتي:

$$CV = \frac{S}{\bar{x}} \qquad or \qquad CV = \frac{S}{\bar{x}} \times 100$$
(20-6)

$$CV = \frac{Q_3 - Q_1}{Q_3 + Q_1} \qquad or \qquad CV = \frac{Q_3 - Q_1}{Q_3 + Q_1} \times 100$$
(21-6)

مثال 11

احسب معامل الاختلاف ومعامل الاختلاف النسبي للبيانات في المثال (9).

الحل:

1- استعمال الوسط الحسابي والانحراف المعياري للبيانات إذ ان الوسط الحسابي كان يساوي (34.871) والانحراف المعياري كان يساوي (3.221). وباستعمال الصيغة (6-20) نجد بان معامل الاختلاف ومعامل الاختلاف النسبي كالاتي:

$$CV = \frac{S}{\bar{x}} = \frac{3.221}{34.871} = 0.092$$
$$CV = \frac{S}{\bar{x}} \times 100 = \frac{3.221}{34.871} \times 100 = 9.2\%$$

2- اما في حالة استعمال الربيعات فإننا نجد بان الربيع الأول للبيانات كما في المثال كان يساوي ($Q_1 = 31.5$) والربيع الثالث كان يساوي ($Q_3 =$ 37.25) وعليه فان معامل الاختلاف ومعامل الاختلاف النسبي باستعمال الصيغة (6-21) يكون كالاتي:

$$CV = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{37.25 - 31.5}{37.25 + 31.5} = \frac{5.75}{68.75} = 0.084$$
$$CV = \frac{Q_3 - Q_1}{Q_3 + Q_1} \times 100 = \frac{37.25 - 31.5}{37.25 + 31.5} \times 100 = 8.4\%$$

ويلاحظ من النتائج بان معامل الاختلاف الذي يحسب عن طريق استعمال الربيعات يكون أصغر منه في حالة استعمال الوسط الحسابي والانحراف

المعياري لأنه يعتمد على قيمتين فقط بينما الثاني يعتمد على جميع القيم للمشاهدات.

7-6 معامل الالتواء Skewness coefficient

وهو مقياس يوضح درجة التواء توزيع البيانات حول مركزها، او مقدار تركز المشاهدات في المنحنى التكراري للبيانات. وعليه فان معامل الالتواء هو مقياس لبعد المنحنى التكراري للبيانات عن التماثل. والالتواء يمكن ان يكون نحو اليمين، وهذا يعني بان القيم الكبيرة تتركز في اليمين او هي الأكثر في مجموعة البيانات وهذا يؤثر في الوسط الحسابي ويسحبه نحو اليمين فيكون أكبر من الوسيط والمنوال، اما إذا كان المنحنى ملتوياً نحو اليسار فان الوسط الحسابي يكون أصغر من الوسيط والمنوال اما اذا كان المنحنى متماثلاً فان الوسط الحسابي يساوي الوسيط ويساوي المنوال كما هو موضح في الشكل (5-4). وتستخدم العلاقة بين المتوسطات والتي تم أشرنا إليها بالمعادلة (5-16) {الوسط الحسابي – المنوال= 3(الوسط الحسابي – الوسيط)} لحساب مقياسين لمعامل الالتواء(SK) وحسب العلاقات الاتية:

$$SK = \frac{3(\bar{X}-Med)}{S} \qquad (22-6)$$

$$SK = \frac{(\bar{X}-Mod)}{S} \qquad (23-6)$$

اذن لحساب معامل الالتواء (SK) يجب أولا حساب مقاييس النزعة المركزية للبيانات ومن ثم تطبيق واحدة من العلاقتين (6-22) او (6-23) وسوف يحسب من المثال التالي.

<u>مثال 12</u>

احسب معامل الالتواء للبيانات في المثال (9)

الحل:

توفرت لدينا المعلومات الاتية من البيانات:

$$\bar{X} = 34.87, \quad S = 3.22, \quad Mod = 35, \quad Med = 35.5$$

اذن معامل الالتواء حسب العلاقة (6−22) يكون:

$$SK = \frac{3(34.87 - 35.5)}{3.22} = -0.587$$

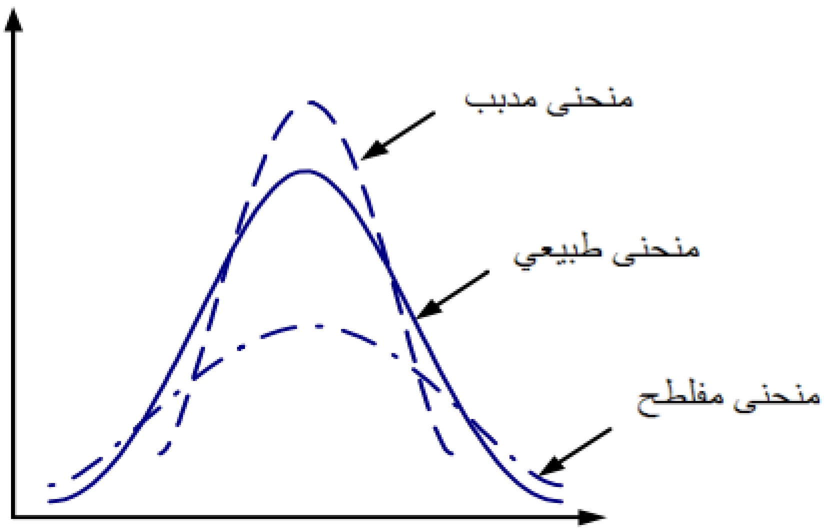
اما عند استعمال العلاقة (6−23) فان قيمة معامل الالتواء تكون:

$$SK = \frac{34.87 - 35}{3.22} = -0.040$$

أظهرت النتائج في كلتا الحالتين بان الالتواء سالب وبسيط لان القيم قريبة من الصفر فكلما كانت قيمة معامل الالتواء قريبة من الصفر يعني أن الالتواء بسيط والاشارة تمثل اتجاه الالتواء.

6−8 معامل التفرطح Kurtosis coefficient

وهو معامل يقيس مدى ارتفاع او انبساط قمة المنحنى التكراري، وبعبارة أخرى مدى تركز البيانات حول المركز او تباعدها عن المركز ويمكن ملاحظة ذلك من خلال تدبب المنحنى وتكون له قمة عالية او تسطح المنحنى وتكون له قمة واطئة مقارنة بالمنحنى المتماثل وكما في الشكل (6−2).



الشكل (6-2) اشكال المنحنيات من حيث التفرطح

ويقاس معامل التفرطح (Ku) باستعمال العلاقة الاتية:

$$Ku = \frac{M_4}{S^4} \qquad (6-24)$$

إذ ان (S) تمثل الانحراف المعياري و (M_4) تمثل العزم الرابع ويحسب بالشكل الاتي وحسب نوع البيانات فيما إذا كانت غير مبوبة او مبوبة في جدول تكراري.

$$M_4 = \frac{\sum(X-\bar{X})^4}{n} \quad \text{or} \quad M_4 = \frac{\sum f(X-\bar{X})^4}{n}$$

ولحساب معامل التفرطح من المثال السابق نحتاج الى حساب قيمة العزم الرابع وقيمة الانحراف المعياري وهما كالاتي:

$$M_4 = 186.286 \qquad S = 3.22$$

$$\therefore Ku = \frac{186.286}{(3.22)^4} = 1.733$$

6–9 حساب مقاييس التشتت باستعمال البرنامج (SPSS)

لتوضيح كيفية حساب مقاييس التشتت ومقاييس الالتواء والتفرطح نأخذ المثال الاتي:

<u>مثال 13:</u>

استعمل البيانات في المثال (14) في الفصل الخامس لحساب مقاييس التشتت والالتواء والتفرطح باستعمال البرنامج (SPSS).

<u>الحل:</u>

إدخال البيانات كما في الشكل (5–7) ومن ثم نتبع الأوامر للحصول على المربع الحواري كما في الشكل (5–8) ومن المربع الحواري نضع علامة صح في المربع المقابل لمقاييس الربيعات (Percentile Values) ومن ثم نضع علامة صح في المربعات التي تمثل مقاييس التشتت تحت مربع (Dispersion) ومن ثم نضع علامة صح في المربعات التي تمثل مقياس الالتواء والتفرطح تحت مربع (Distribution) فتكون النتائج كما في الجدول (6–17).

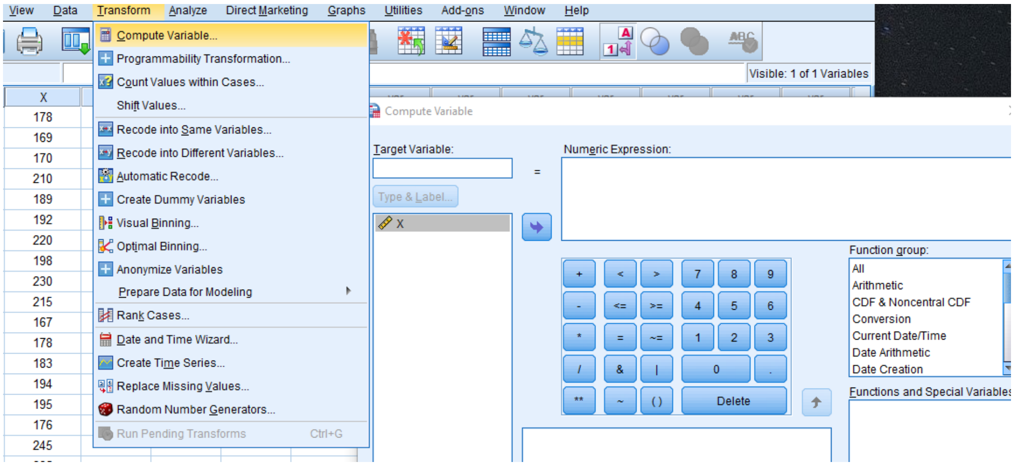
جدول (6–17) مقاييس التشتت والالتواء والتفرطح

N	Valid	20
Std. Deviation		22.459
Variance		504.421
Skewness		0.382
Kurtosis		-0.762
Range		78
Percentiles	Q1	178.00
	Q2	194.50
	Q3	215.75

النتائج في الجدول (6-17) لم تظهر المقاييس كلها لان قسماً منها يمكن حسابه من هذه النتائج مثل معامل الاختلاف، اما بالنسبة للانحراف المتوسط ليس محسوباً من ضمن هذه النتائج ويمكن حسابه من المثال حسب الخطوات الاتية:

أ- إدخال البيانات كما في الشكل (5-7) ومن ثم نتبع الأوامر للحصول على المربع الحواري كما في الشكل (5-8) ومن المربع الحواري ومنه يتم حساب قيمة الوسط الحسابي التي تساوي (198) كما في الجدول (5-11).

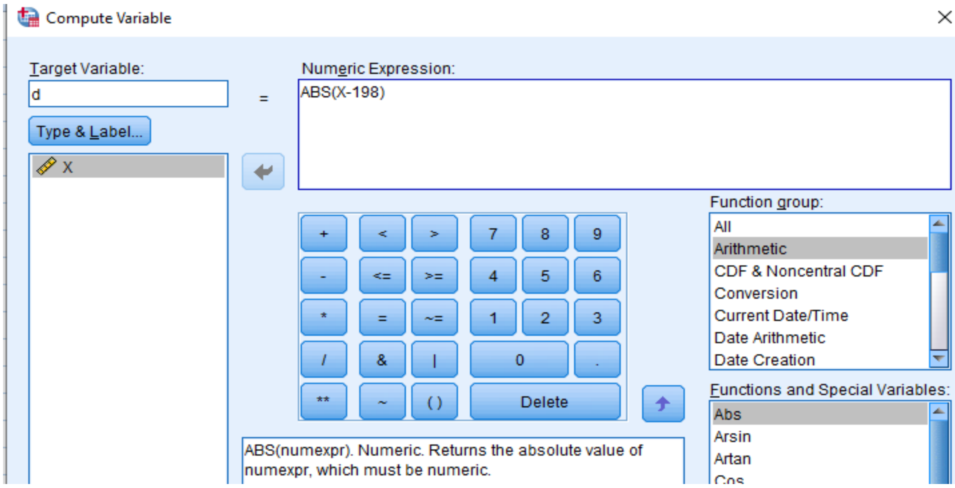
ب-إيجاد الانحرافات المطلقة عن الوسط الحسابي بعد اهمال الإشارة من خلال استعمال الامر الرئيسي (Transform) ومن الامر الفرعي (Compute variable) نحصل على مربع الحوار كما في الشكل (6-3).



الشكل (6-3) مربع الحوار لحساب الانحرافات المطلقة عن الوسط الحسابي

ت- نسمى المتغير الجديد الذي يمثل الانحرافات (d) في مربع (Target variable) ثم من مربع (Function group) نختار الدالة (Arithmetic) تظهر لنا في المربع الأسفل مجموعة دوال نختار الدالة (Abs) التي تمثل القيم المطلقة بعد اهمال الإشارة ثم ننقلها الى المربع الأعلى باستعمال السهم وداخل القوس امام الدالة ننقل اسم المتغير ونطرح منه قيمة الوسط الحسابي

لنحصل على عمود جديد يمثل الانحرافات المطلقة عن الوسط الحسابي كما في الشكل (4-6).



الشكل (4-6) مربع الحوار لحساب الانحرافات المطلقة

ث- حساب قيمة الانحراف المتوسط والتي تمثل متوسط الانحرافات المطلقة وذلك باستعمال الامر الرئيسي (Analyze) ومنه الامر الفرعي (Descriptive Statistics) ثم (Frequency) وننقل المتغير (d) الى المربع في الجانب الأيمن ومن مربع الحوار نختار الامر (Statistics) ثم نختار المتوسط (Mean) فنحصل على قيمة الانحراف المتوسط والتي تساوي (18.7).

تمارين الفصل السادس

1- عرف المصطلحات الاتية:

أ- التشتت، ب- التباين، ج- الانحراف المعياري.

2- البيانات الاتية تمثل الوقت المصروف (دقيقة) من الطبيب مع المريض في اثناء مراجعته في العيادة الخارجية لاحد المستشفيات الخاصة:

10، 12، 18، 22، 24، 10، 8، 11، 28، 9، 14، 27، 19، 14، 26، 15، 13، 23، 7، 9، 21، 20، 20، 18، 19، 12، 14، 10.

<u>والمطلوب: استعمال برنامج (SPSS) لحساب</u>

أ- مقاييس التشتت الممكنة لهذه البيانات.

ب- كون التوزيع التكراري المناسب لهذه البيانات.

ت-حساب مقاييس التشتت المناسبة للتوزيع التكراري.

3- احسب مقياس التشتت المناسب لكل مجموعة من البيانات الاتية، ولماذا تم اختيار المقياس؟:

ج- 2، 2، 3، 4، 4، 4، 5، 6.

ح- 2، 3، 4، 4، 4، 5، 20.

خ- -5، -4، -3، 0، 3، 4، 5.

د- 1.5، 1.7، 1.8، 1.9، 1.2، 1.3، 1.7، 1.6، 1.7.

4- البيانات الاتية تمثل قياسات ضغط الدم العالي لمجموعة من الاشخاص

-180	-170	-160	-150	-140	-130	-120
5	8	13	10	9	4	

<u>والمطلوب:</u>

حساب مقاييس التشتت الممكنة لهذا الجدول التكراري.

5- كان اعداد المراجعين للعيادة الخارجية في أحد المستشفيات حسب ايام الاسبوع كما في الجدول الاتي:

السبت	الاحد	الاثنين	الثلاثاء	الاربعاء	الخميس	الجمعة
23	45	53	64	56	42	18

<u>والمطلوب:</u>

حساب مقاييس التشتت المناسبة لهذا الجدول التكراري.

6- احسب مقاييس التشتت المناسبة وفقا للمعلومات الاتية:

e- $\sum x_i = 204, \quad \sum x_i^2 = 2099, \quad n = 20$

f- $\sum x_i = 9.78, \quad \sum x_i^2 = 0.97, \quad n = 100$

g- $\sum x_i = -246, \quad \sum x_i^2 = 1254, \quad n = 50$

h- $\sum x_i = 586, \quad \sum x_i^2 = 1844, \quad n = 30$

7- الجدول التكراري الاتي يلخص الوزن (كغم) لخمسين مريض في احد المستشفيات:

الوزن (كغم)	-50	-55	-60	-65	75-70	المجموع
عدد الاشخاص (f_j)	5	10	20	8	7	50

<u>والمطلوب:</u>

حساب مقاييس التشتت الممكنة وحساب معامل الالتواء والتفلطح.

الفصل السابع

مقاييس الارتباط

Correlation Measurements

7-1 تمهيد

بعد ان تم التعرف على كيفية وصف الظواهر الحياتية او غيرها من الظواهر من خلال مجموعة وسائل استعرضناها في الفصول السابقة التي تنوعت بين الوصف والتلخيص الجدولي والعرض البياني ومن ثم إيجاد مقاييس عددية غاية في الاختصار بحيث يتم وصف مجموعة بيانات من خلال رقم واحد يمثل المركز او التشتت والتباعد عن المركز. فمن الضروري التعرف بعدها على العلاقات التي تربط بين الظواهر المختلفة وهل توجد علاقة او لا توجد وما نوع العلاقة وهنا يكون دور مجموعة من المقاييس تسمى مقاييس الارتباط والتي ستكون موضوع هذا الفصل.

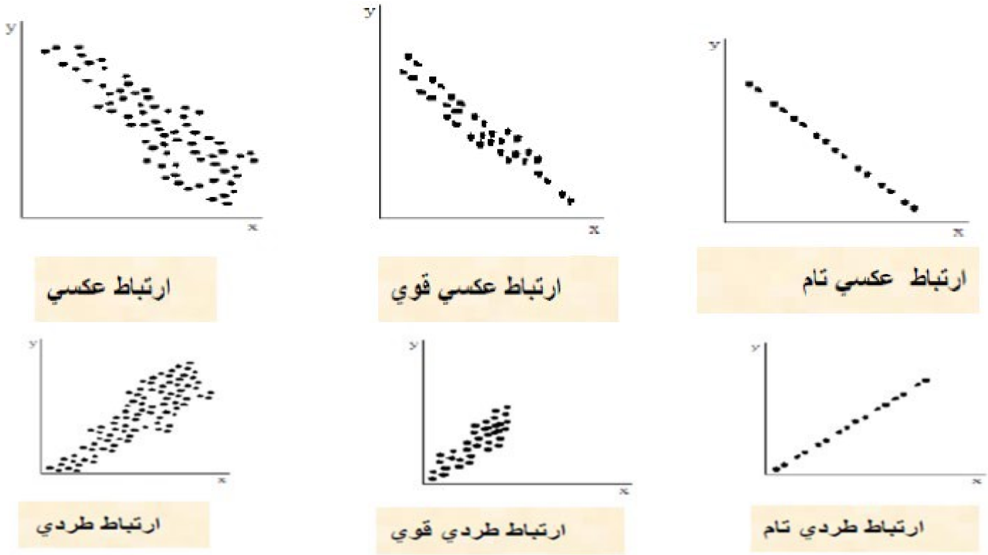
7-2 مفهوم الارتباط Correlation Definition

بالرغم من أن مقاييس العلاقة تختلف عما سبقها من مقاييس، فهي تتعلق بدراسة العلاقة بين ظاهرتين او أكثر، في حين أن المقاييس السابقة تهتم بدراسة صفات الظواهر. ويوفر تحليل الارتباط وسيلة للاستدلال على قوة العلاقة بين متغيرين أو أكثر، أي أن الارتباط هو مقياس للدرجة التي تتغير فيها قيم المتغير بأسلوب منتظم. وهو يعد مؤشراً كمياً لتحديد درجة الاعتماد على متغير أو أكثر في التنبؤ بقيم متغير آخر. من المهم معرفة ما يمكن أن يوفره تحليل الارتباط وبالأهمية بنفسها يتوجب معرفة مالا يمكن أن يوفره هذا النوع من التحليل. فتحليل الارتباط لا يقدم أية معلومات للتنبؤ بقيم متغير ما، كما انه لا يوفر أي مؤشر فيما لو كانت العلاقة بين المتغيرات سببية، انما يستطيع التحليل تحديد فيما لو كانت درجة التباين المشترك ذات دلالة

احصائية. ولذا تعرف العلاقة بين الظاهرتين او المتغيرين بالارتباط. وقد يكون الارتباط طرديا بمعنى ان تتغير الظاهرتين في الاتجاه نفسه بحيث إذا زادت احدى الظاهرتين تميل الثانية الى الزيادة وبالعكس. وقد يكون الارتباط عكسيا بمعنى ان الظاهرتين تتغير في اتجاهين متضادين بحيث إذا زادت احدى الظاهرتين تميل الثانية الى النقصان وبالعكس. ويلاحظ ان قيمة معامل الارتباط هي قيمة عددية تنحصر بين 1+ و 1− ولا تكون هذه القيمة 1+ و 1− الا اذا كان الارتباط تاما.

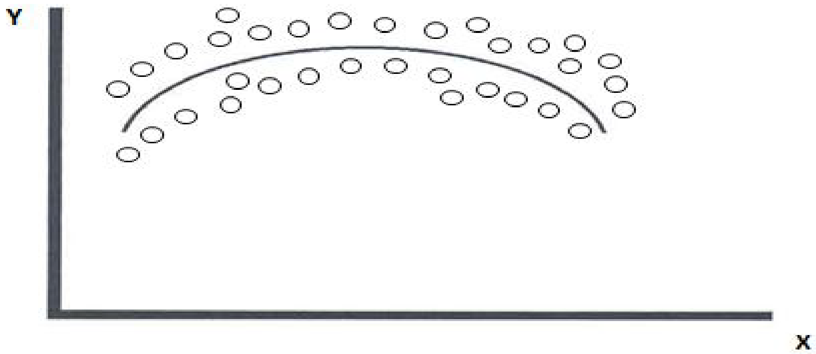
وهناك وسيلة مبدئية يعرف الباحث من خلالها نوع الارتباط بين المتغيرين، وما إذا كان الارتباط قويا او ضعيفا أو منعدما، وما إذا كانت العلاقة خطية أو غير خطية، موجبة أو سالبة. هذه الوسيلة هي " شكل الانتشار " (Scatter Diagram) التي تصلح إذا كان المتغيران كميين. وجدير بالذكر أن هذه الوسيلة مبدئية تساعد فقط على معرفة نوع الارتباط ولا تعد بديلا عن الطرق الإحصائية التي سوف نتناولها في هذا الفصل.

والمقصود بشكل الانتشار هو تمثيل قيم الظاهرتين بيانيا على المحورين، المتغير الأول (X) على المحور الافقي، والمتغير الثاني (Y) على المحور العمودي، إذ إن تمثيل كل زوج (pair) من القيم بنقطة، فنحصل على شكل يمثل كيفية انتشار القيم على المستوى، وهو الذي يسمى شكل الانتشار (Scatter Diagram). 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 النقاط 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:



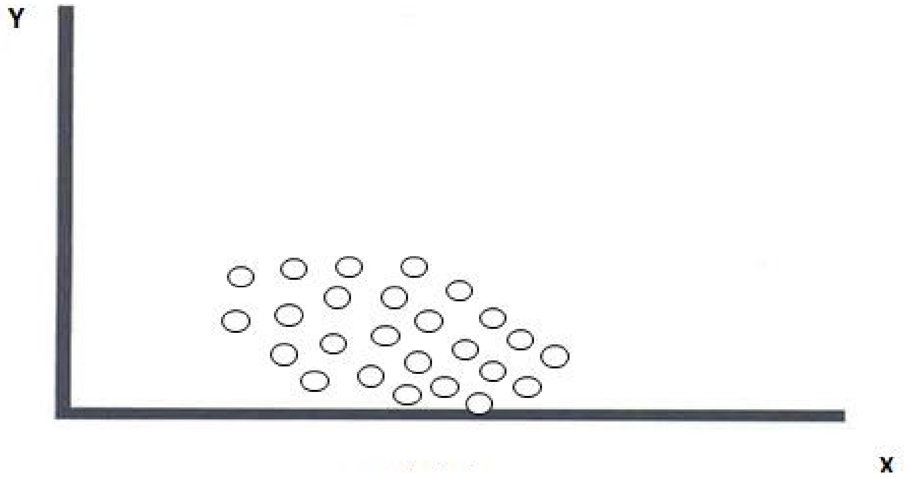
شكل (7-1) بعض اشكال الانتشار التي تمثل انواع الارتباط

واما إذا كانت العلاقة تأخذ شكلاً منحنياً فإن الارتباط لا يكون خطيا "ارتباط غير خطي (Nonlinear Correlation) كما في الشكل الاتي:



شكل (7–2) شكل الانتشار الذي يمثل الارتباط غير الخطي

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



شكل (7–3) شكل الانتشار عندما لا توجد علاقة بين المتغيرين

وبشكل عام كلما كانت العلاقة قوية بين المتغيرين كلما اقترب معامل الارتباط من + ١ أو – ١ فإذا وصلت قيمة المعامل إلى + ١ أو – ١ كان الارتباط تاما بين المتغيرين. وأنه كلما كانت العلاقة ضعيفة بين المتغيرين كلما اقترب معامل الارتباط من الصفر، فإذا وصلت قيمة المعامل إلى الصفر كان الارتباط منعدما بين المتغيرين، ومعنى ذلك أيضا أنه لا يوجد ارتباط بين المتغيرين. ويمكن تمثيل قوة العلاقة بالجدول الاتي:

جدول (7–1) قوة العلاقة من خلال قيم معامل الارتباط

ارتباط عكسي					ارتباط طردي					
قوي جدا	قوي	متوسط	ضعيف	ضعيف جدا	ضعيف جدا	ضعيف	متوسط	قوي	قوي جدا	
-1	-0.9	-0.7	-0.5	-0.3	0	0.3	0.5	0.7	0.9	1
تام					منعدمة					تام

7–3 أنواع الارتباط **Correlation Types**

يقسم الارتباط الى عدة أنواع اعتمادا على عدد المتغيرات التي تدخل في عملية الحساب، فإما ان يكون ارتباطاً بسيطاً ويمثل العلاقة بين ظاهرتين فقط بغض النظر عن الظواهر الأخرى، واما ان يكون متعدداً ويمثل العلاقة بين ظاهرة واحدة وعدد من الظواهر التي

تؤثر فيها لأنه في الحياة، وأحيانا يوجد اكثر من ظاهرة في الدراسة، ونرغب في معرفة العلاقة بين هذه الظواهر، او أحيانا أخرى نرغب في قياس العلاقة بين ظاهرتين، مع العلم بانه هناك ظواهر أخرى تؤثر في هاتين الظاهرتين ومن ثم نرغب في معرفة العلاقة بين تلك الظاهرتين مع ثبات تأثير الظواهر الأخرى، ويسمى الارتباط الجزئي ويمثل العلاقة بين متغيرين بشرط ثبات عدد اخر من المتغيرات وسنوضح كل نوع منها في المباحث الآتية. وخصص المبحث الأول للارتباط البسيط الذي يقسم الى نوعين اعتمادا على قياسات الظواهر المراد حساب معامل الارتباط لها، إذ ان المتغيرات التي تمثل الظواهر اما ان تكون مقاسة كميا اي استعمال وحدات القياس الكمي، او مقاسة باستعمال وحدات القياس النوعي اي الصفات اي ان قياسها لا يعتمد على كميات عددية ومن ثم فان معامل الارتباط يكون كمياً ويسمى معامل ارتباط بيرسون او يكون وصفياً ويسمى ارتباط سبيرمان كما المباحث الفرعية الاتية:

<u>7-3-1</u> الارتباط البسيط الكمي (معامل بيرسون)

Simple Quantitative Correlation (Pearson Coefficient)

ويشمل دراسة العلاقة بين الظواهر القابلة للقياس الكمي او العددي، وهذا يشمل الظواهر جميعها التي يمكن التعبير عنها بصورة رقمية، كالطول، والعمر، ودرجة الحرارة للمريض، وكمية الكوليسترول، وغير ذلك من الظواهر التي يمكن التعبير عنها رقميا. وهو معامل ارتباط يحدد مقدار او قوة العلاقة واتجاهها بين ظاهرتين اثنتين، وذلك في الحالات او الظواهر التي تقتصر فيها الدراسة على متغيرين. فعلى سبيل المثال قد يكون المطلوب التعرف على قوة العلاقة واتجاهها بين عدد السكائر التي يدخنها مجموعة من الأشخاص والاصابة بسرطان الرئة. او قد يكون الهدف مثلا التعرف على قوة واتجاه العلاقة بين كمية الكوليسترول في الجسم والوزن في مجتمع ما. إن الغرض من تحليل الارتباط الخطي البسيط هو تحديد نوع العلاقة وقوتها بين متغيرين، ويرمز له في حالة المجتمع بالرمز (ρ) (رو)، وفي حالة العينة بالرمز (r)، وحيث أننا في كثير من النواحي التطبيقية نتعامل مع بيانات عينة مسحوبة من المجتمع، سوف نهتم

بحساب معامل الارتباط في العينة كتقدير لمعامل الارتباط في المجتمع. ويحسب معامل ارتباط بيرسون من خلال العلاقة الآتية:

$$\boldsymbol{r} = \frac{\sum \boldsymbol{XY} - \frac{(\sum \boldsymbol{X})(\sum \boldsymbol{Y})}{\boldsymbol{n}}}{\sqrt{\left[\sum \boldsymbol{X}^2 - \frac{(\sum \boldsymbol{X})^2}{\boldsymbol{n}}\right]\left[\sum \boldsymbol{Y}^2 - \frac{(\sum \boldsymbol{Y})^2}{\boldsymbol{n}}\right]}} \qquad \textbf{(1-7)}$$

حيث ان:

$\sum XY$: تعني مجموع حاصل ضرب كل قيمة من المتغير (X) في كل قيمة مقابلة لها من المتغير (Y).

$\sum X$: تعني مجموع قيم المتغير (X).

$\sum Y$.تعني مجموع قيم المتغير (Y).

$\sum X^2$ تعني مجموع مربع قيم المتغير (X).

$(\sum X)^2$ تعني مربع مجموع قيم المتغير (X).

$\sum Y^2$ تعني مجموع مربع قيم المتغير (Y).

$(\sum Y)^2$ تعني مربع مجموع قيم المتغير (Y).

n تمثل عدد قيم الدراسة (عدد الأزواج المطلوب حساب الارتباط بينها).

<u>مثال1</u>

ترغب إحدى المستشفيات في معرفة العلاقة بين عدد ساعات العمل لموظفيها ومستوى الإنتاجية إذ أن مستوى الإنتاجية قيس بمؤشر من (1) الى (15)، فسجلت البيانات عن هذا الموضوع وكانت النتائج كما في الجدول الاتي:

جدول (7−2) بيانات ساعات العمل والإنتاجية لمجموع م الموظفين

الموظفين	1	2	3	4	5	6	7	8	9	10
ساعات العمل (X)	8	2	8	5	15	11	13	6	4	6
مستوى الإنتاجية (Y)	3	1	6	3	14	12	9	4	4	5

المطلوب: هل توجد علاقة بين الظاهرتين مع تحديد قوة العلاقة ونوعها.

الحل:

بما ان بيانات كلا المتغيرين كمية، ولرغبة المستشفى معرفة العلاقة بين المتغيرين موضوع الدراسة، فإن أنسب أسلوب إحصائي هنا هو معامل ارتباط بيرسون، ولغرض حساب معامل ارتباط بيرسون من خلال البيانات الخام وتطبيق المعادلة آنفة الذكر، فإننا نتبع الخطوات الآتية:

1- نربع كل قيمة من المتغير (X) لكي نحصل على (X^2).

2- نربع كل قيمة من المتغير (Y) لكي نحصل على (Y^2).

3- نضرب كل قيمة من المتغير (X) في القيمة المقابلة لها من المتغير (Y) لكي نحصل على (XY).

4- نقوم بعد ذلك بعرض البيانات المتحصل عليها في جدول (7−3) كالاتي وذلك حتى يسهل علينا استخلاص المجاميع التي نحتاجها لحساب معامل ارتباط بيرسون باستعمال المعادلة (7−1).

جدول (7−3) العمليات الحسابية لحساب معامل الارتباط البسيط

الموظفين ID	ساعات العمل X	مستوى الإنتاجية Y	X^2	Y^2	XY
1	8	3	64	9	24
2	2	1	4	1	2
3	8	6	64	36	48
4	5	3	25	9	15
5	15	14	225	196	210
6	11	12	121	144	132
7	13	9	163	81	117
8	6	4	36	16	24
9	4	4	16	16	16
10	6	5	36	25	30
المجموع	78	61	760	533	618

$$r = \frac{\sum XY - \frac{(\sum X)(\sum Y)}{n}}{\sqrt{\left[\sum X^2 - \frac{(\sum X)^2}{n}\right]\left[\sum Y^2 - \frac{(\sum Y)^2}{n}\right]}}$$

$$\therefore r = \frac{618 - \frac{(78)(61)}{10}}{\sqrt{\left[760 - \frac{(78)^2}{10}\right]\left[533 - \frac{(61)^2}{10}\right]}} = 0.91$$

وهذه النتيجة توضح أن درجة الارتباط = 0.91 وهذه النتيجة تعتبر مؤشراً على ان العلاقة إيجابية (طردية) قوية بين ساعات العمل ومستوى الإنتاجية. اما إذا كانت قياسات أحد الظاهرتين او كلاهما وصفية (نوعية) فانه لا يمكن تطبيق المعادلة $(1-7)$ لحساب معامل الارتباط لذلك نلجأ الى ما يسمى بارتباط الرتب او معامل سبيرمان كما في المبحث الفرعي الاتي.

7-3-2 ارتباط الرتب الوصفي (معامل سبيرمان)

Rank Correlation (Spearman Coefficient)

قد تكون الظواهر مقاسة بقياسات وصفية أي صفات وأنواع ويشترط بها ان تكون ترتيبية أي قابلة للترتيب (Ordinal variables) أو أن تكون أحد الظاهرتين مقاسة كميا، في حين الأخرى وصفية ترتيبية، أو أن تكون الظاهرتان كميتين، ولكن يرغب الباحث في حساب معامل الارتباط بين رتب المتغيرين وليس بين القيم ذاتها، ويكون اهتمام الباحث منصبا على الرتب أكثر من القيم. فعلى سبيل المثال في قياس كمية الكوليسترول يعتبر المريض الأول هو من سجل اعلى كمية بغض النظر عن مقدارها، والذي يحصل على كمية أقل منه مباشرة هو الثاني.. وهكذا. 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. ولحساب معامل ارتباط

الرتب هناك طرق مختلفة أهمها معامل ارتباط سبيرمان للرتب (Spearman Rank Correlation Coefficient). ولحساب معامل سبيرمان لارتباط الرتب نقوم بترتيب كل من المتغيرين ترتيبا تصاعديا أو تنازليا (اما تصاعديا لكلا المتغيرين أو تنازليا 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_i) ثم نقوم بتربيع هذه الفروق ونحصل على مجموعها أي نحصل على $(\sum d_i^2)$ ثم نعوض في معامل سبيرمان لارتباط الرتب والذي يأخذ الصيغة الاتية:

$$r_s = 1 - \frac{6(\sum d_i^2)}{n(n^2-1)} \qquad (2-7)$$

مما سبق نستطيع إجمال بعض الملاحظات كالاتي:

1- مجموع الفروق بين الرتب يساوي صفراً.

2- 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 + ١ (ارتباط طردي تام بين الرتب).

3- وإذا كانت الرتبة رقم ١ (أقل رتبة) للمتغير الأول تناظرها أعلى رتبة للمتغير الثاني وهكذا.. فإن معامل ارتباط الرتب يساوي – ١ (ارتباط عكسي تام بين الرتب).

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

<u>مثال 2</u>

البيانات التالية تمثل إجابات عينة من سبعة أشخاص حول برامج الضمان الاجتماعي، ومدى وملاءمتها لحاجات الناس، تم توجيه سؤالين للأشخاص وكانت اجاباتهم كالاتي:

جدول (7-4) إجابات عينة البحث حول برامج الضمان الاجتماعي

السؤال الأول	جيد	مقبول	ممتاز	جيد	جيد جدا	مقبول	جيد
السؤال الثاني	جيد جدا	مقبول	جيد جدا	جيد	جيد	جيد	ممتاز

<u>والمطلوب:</u> حساب معامل سبيرمان لارتباط الرتب بين هذه الاجابات

<u>الحل:</u>

ننظم البيانات في الجدول (7-5) مع ملاحظة ما يلي:

1- تم بالنسبة للسؤال الأول، ترتب البيانات تصاعديا فإن التقدير الأدنى سيحصل على الرتبة رقم 1 والأعلى منه مباشرة سيحصل على الرتبة رقم 2 وهكذا.. ونكرر العمل نفسه مع السؤال الثاني. عند حصول إجابتين أو أكثر على التقدير نفسه نعطي لكل إجابة مبدئيا رتبة كما لو كانوا مختلفين ثم نحسب متوسط هذه الرتب، وهذا المتوسط هو الذي يعطى لكل إجابة. فتكون الرتب كما في الجدول (7-5):

جدول (7-5) ترتيب البيانات بشكل تصاعدي مع اعطاء رتب لكل قيمة

السؤال الأول	مقبول	مقبول	جيد	جيد	جيد	جيد جدا	ممتاز
الرتب الأولية	1	2	3	4	5	6	7
الرتب النهائية	1.5	1.5	4	4	4	6	7
السؤال الثاني	مقبول	جيد	جيد	جيد	جيد جدا	جيد جدا	ممتاز
الرتب الأولية	1	2	3	4	5	6	7
الرتب النهائية	1	3	3	3	5.5	5.5	7

2- نستعمل البيانات في الجدول (7-4) كما ورد في المثال وللمحافظة على مواقع البيانات بدون تغيير ونضع الرتبة المقابلة لكل نوع من البيانات كما في الجدول

(5−7) ثم نحسب الفروق ونرمز لها بالحرف (d_i) ونجد مربعات الفروق لان مجموعها يساوي صفر ونرمز له بالرمز (d_i^2) ثم نجد مجموع مربعات الفروق ونرمز له $(\sum d_i^2)$ كما في الجدول (7−6):

جدول (7−6) حسابات الفروق ومربعاتها للرتب

السؤال الأول	جيد	مقبول	ممتاز	جيد	جيد جدا	مقبول	جيد	
الرتب	4	1.5	7	4	6	1.5	4	
السؤال الثاني	جيد جدا	مقبول	جيد جدا	جيد	جيد	جيد	ممتاز	
الرتب	5.5	1	5.5	3	3	3	7	
الفروق (d)	−1.5	0.5	1.5	1	3	−1.5	−3	$\sum d_i = 0$
d^2	2.25	0.25	2.25	1	9	2.25	9	$\sum d_i^2 = 26$

وبعد ذلك نستعمل الصيغة (7−2) لحساب قيمة معامل ارتباط سبيرمان كالاتي:

$$r_s = 1 - \frac{6\left(\sum d_i^2\right)}{n(n^2-1)} = 1 - \frac{6(26)}{7(7^2-1)} = 1 - 0.46 = 0.54$$

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

<u>مثال3</u>

البيانات التالية تمثل مؤشر كتلة الجسم (BMI) والوزن (كغم) لعشرة طلاب في أحد الأقسام.

جدول (7–7) بيانات كتلة الجسم والوزن

BMI(X)	10	26	12	14	11	16	19	16	31	28
Weight (Y)	60	48	83	76	58	74	98	69	97	89

<u>المطلوب:</u> أحسب معامل ارتباط الرتب (معامل سبيرمان)

<u>الحل:</u>

بما ان البيانات كمية والمطلوب حساب معامل ارتباط الرتب سبيرمان فلذلك يجب ان ترتب القيم تصاعديا او تنازليا وسوف يتم استعمال الترتيب التنازلي كما في الجدول (7–8):

جدول (7–8) الترتيب التنازلي للبيانات مع الرتب النهائية

BMI(X)	10	26	12	14	11	16	19	16	31	28	المجموع
ترتيب تنازلي (X)	31	28	26	19	16	16	14	12	11	10	
رتب (X)	**10**	**3**	**8**	**7**	**9**	**5.5**	**4**	**5.5**	**1**	**2**	
Weight (Y)	60	48	83	76	58	74	98	69	97	89	
ترتيب تنازلي (Y)	98	97	89	83	76	74	69	60	58	48	
رتب (Y)	**8**	**10**	**4**	**5**	**9**	**6**	**1**	**7**	**2**	**3**	
الفروق (d)	2	-7	4	2	0	-0.5	3	-1.5	-1	-1	
d^2	4	49	16	4	0	0.25	9	2.25	1	1	**86.5**

وبتطبيق المعادلة (7–2) نحصل على قيمة معامل ارتباط الرتب (سبيرمان) كالاتي:

$$r_s = 1 - \frac{6(\sum d^2)}{n(n^2-1)} = 1 - \frac{6(86.5)}{10(10^2-1)} = 1 - 0.524 = 0.476$$

مما يعني أننا أمام علاقة طردية ضعيفة بين مؤشر كتلة الجسم (BMI) والوزن.

<u>7–4 معامل الارتباط المتعدد Multiple Correlation Coefficient</u>

الارتباط المتعدد هو الارتباط الذي يبحث بالعلاقة بين متغير واحد من جهة وبين متغيرين إثنين مجتمعين معاً او أكثر من جهة أخرى دون استبعاد أي منهما على الإطلاق. فعلى سبيل المثال إذا كان لدينا المتغير (Y) يمثل المتغير المطلوب حساب

معامل الارتباط له (المتغير المعتمد) مع متغيرين مستقلين هما (X1) و (X2) فان معامل الارتباط المتعدد بالرموز يكون كما في الصيغة الاتية:

$$\boldsymbol{R_{Y.12}} = \sqrt{\frac{r_{Y1}^2+r_{Y2}^2-2r_{Y1}r_{Y2}r_{12}}{1-r_{12}^2}} \qquad (3-7)$$

حيث ان:

($R_{Y.12}$) يمثل معامل الارتباط المتعدد بين (Y) و (X_1) و(X_2).

(r_{Y1}) معامل الارتباط البسيط بين (Y) و (X_1)، (r_{Y2}) يمثل معامل الارتباط البسيط بين (Y) و (X_2)، (r_{12}) يمثل معامل الارتباط البسيط بين (X_1) و (X_2).

مثال 4

أراد الاطباء معرفة العلاقة بين (ضغط الدم) (Y) المتغير التابع وكل من (الوزن)(X1) المتغير المستقل الأول (والكوليسترول)(X2) المتغير المستقل الثاني وكانت معاملات الارتباط البسيط بين المتغيرات كما يأتي:

$$r_{Y1} = 0.82,\ \ r_{Y2} = 0.86,\ \ r_{12} = 0.80$$

الحل

عند تطبيق الصيغة (7−3) وتعويض المعاملات نحصل على قيمة الارتباط المتعدد كالاتي:

$$R_{Y.12} = \sqrt{\frac{r_{Y1}^2+r_{Y2}^2-2r_{Y1}r_{Y2}r_{12}}{1-r_{12}^2}} = \sqrt{\frac{(0.82)^2+(0.86)^2-2(0.82)(0.86)(0.80)}{1-(0.80)^2}} = 0.88$$

وهذا يعني بان العلاقة بين المتغير المعتمد (Y) والمتغيرات الاخرى قوية وطردية، ومن خصائص معامل الارتباط المتعدد

1- إن قيمة معامل الارتباط المتعدد تتراوح بين (1,0) بمعنى أن قيمته موجبة دائما أي إن العلاقة بين 3 متغيرات فأكثر هي دائما طردية.

2- إن قيمته تزداد كلما ازداد عدد المتغيرات الداخلة في الدراسة بمعنى إن قيمة معامل الارتباط المتعدد ل 4 متغيرات هي أكبر من قيمة معامل الارتباط المتعدد ل 3 متغيرات.

5−7 معامل الارتباط الجزئي **Partial Correlation Coefficient**

هو معامل ارتباط يقيس العلاقة بين متغيرين اثنين بافتراض ثبات تأثير المتغير الثالث على كل من المتغيرين. فعلى سبيل المثال لقياس قوة العلاقة بين متغير ضغط الدم (X) وقياس السكر في الدم (Y)، بافتراض ثبات مستوى الكوليسترول (Z) على العلاقة. فهذا يعني بان المطلوب قياس العلاقة بين متغيرين مع ثبات تأثير متغير ثالث ويرمز لمعامل الارتباط في مثل هذه الحالات بالرمز $r_{XY,Z}$ اما قانون حساب معامل الارتباط الجزئي بين المتغير Y والمتغير (X) مع ثبات تأثير المتغير (Z) فيكون حسب الصيغة الاتية:

$$r_{XY.Z} = \frac{r_{XY} - r_{XZ} r_{YZ}}{\sqrt{1-r_{XY}^2}\sqrt{1-r_{XZ}^2}} \qquad (4-7)$$

إذ ان (r) تشير الى معاملات الارتباط البسيط لبيرسون بين المتغيرات.

<u>مثال5</u>

أرادت إدارة الصحة في مدينة معينة معرفة العلاقة بين عدد المصابين بضغط الدم (Y) وعدد الأشخاص المصابين بالسكر (X) وعدد الأشخاص الذين لديهم سكر وضغط الدم (Z) في تلك المدينة وحصلت إدارة الصحة على النتائج الآتية:

$$r_{XZ} = 0.741, \qquad r_{YX} = 0.931, \qquad r_{YZ} = 0.909$$

بتطبيق المعادلة (7−4) لحساب معامل الارتباط الجزئي $(r_{XY,Z})$ وعلى النحو التالي:

$$r_{XY.Z} = \frac{r_{XY} - r_{XZ} r_{YZ}}{\sqrt{1-r_{XY}^2}\sqrt{1-r_{XZ}^2}} = \frac{0.931 - (0.741)(0.909)}{\sqrt{1-(0.931)^2}\sqrt{1-(0.741)^2}}$$
$$= 0.92$$

وهذا يعني بان العلاقة بين عدد المصابين بضغط الدم وعدد المصابين بالسكر كانت طردية وقوية، ومن خصائص معامل الارتباط الجزئي إن قيمته تتراوح بين (1−،1) وتفسر قيمته كما تفسر قيمة معامل الارتباط البسيط. وإن معامل الارتباط الجزئي لأي متغيرين تكون إشارته مماثلة لإشارة معامل الارتباط البسيط بينهما.

7-6 حساب معاملات الارتباط باستعمال برنامج (SPSS)

لتوضيح كيفية استعمال البرنامج (SPSS) لحساب معامل الارتباط نأخذ المثال الاتي:

مثال 6

البيانات الاتية تمثل مجموعة صفات لمجموعة من الأشخاص كالوزن (كغم) ودرجة الإصابة بالمرض (1=بسيطة، 2=متوسطة، 3= شديدة) ومستوى الكوليسترول (ملغم/دل) وعدد السكائر التي يدخنها الشخص ودرجة ضغط الدم الواطئ (بسيط، متوسط، شديد) ودرجة ضغط الدم العالي (ملم) كما في الجدول (7-9).

جدول (7-9) بيانات مجموعة من المرضى

درجة الضغط العالي	ضغط الدم الواطئ	عدد السكائر التي يدخنها	كمية الكوليسترول	درجة الإصابة بالمرض	الوزن (كغم)
132	بسيط	57	247	1	70
142	متوسط	61	163	2	68
124	متوسط	31	258	3	86
126	شديد	56	153	3	59
139	شديد	61	253	2	72
158	بسيط	61	163	2	86
160	بسيط	61	253	2	50
128	بسيط	36	239	2	69
146	متوسط	51	158	2	84
166	متوسط	51	250	2	48
156	متوسط	61	253	2	75
138	شديد	61	163	2	60
137	شديد	71	161	3	95
139	شديد	46	255	3	66
140	متوسط	56	231	1	58
130	شديد	56	250	1	69
129	بسيط	56	250	3	76
25	بسيط	56	158	2	95
126	بسيط	61	155	3	88
146	شديد	46	253	2	85
150	شديد	41	166	2	64
135	شديد	66	172	2	74

139	متوسط	56	255	2	41
148	متوسط	61	261	3	80
170	متوسط	46	239	2	62
136	متوسط	31	239	2	53
135	متوسط	66	161	3	54
120	شديد	46	158	2	78
128	بسيط	46	239	3	95
142	بسيط	41	155	1	69
143	شديد	51	247	2	67
152	متوسط	61	164	3	76
156	شديد	71	166	3	92
172	شديد	31	172	2	50
140	متوسط	61	261	3	99
146	متوسط	66	161	3	44
154	بسيط	66	166	2	85

المطلوب

1- حساب معامل الارتباط البسيط (معامل بيرسون) بين المتغيرات الكمية.

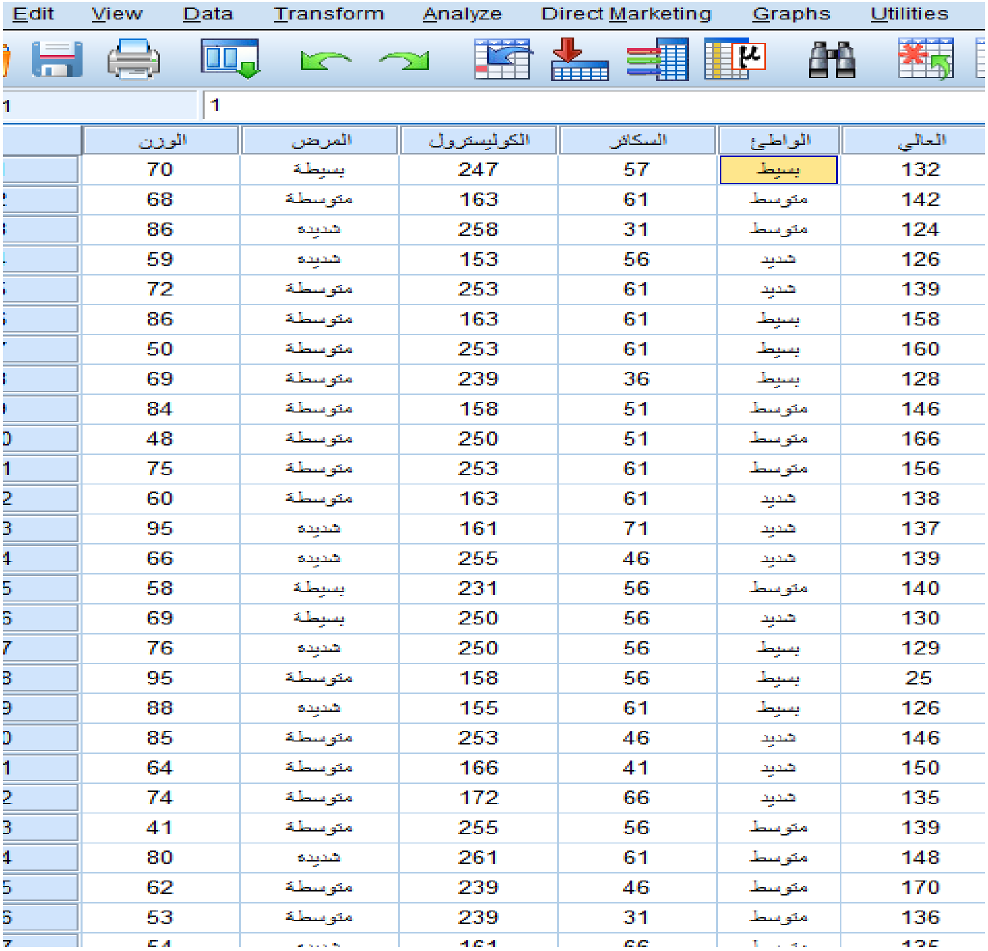
2- حساب معامل الارتباط للرتب بين المتغيرات النوعية (الوصفية).

3- حساب معامل الارتباط المتعدد على اعتبار ان الوزن هو المتغير المعتمد وباقي المتغيرات هي المستقلة.

4- حساب معامل الارتباط الجزئي بين المتغيرات الكمية على اعتبار ان متغير الوزن هو المتغير المسيطر.

الحل

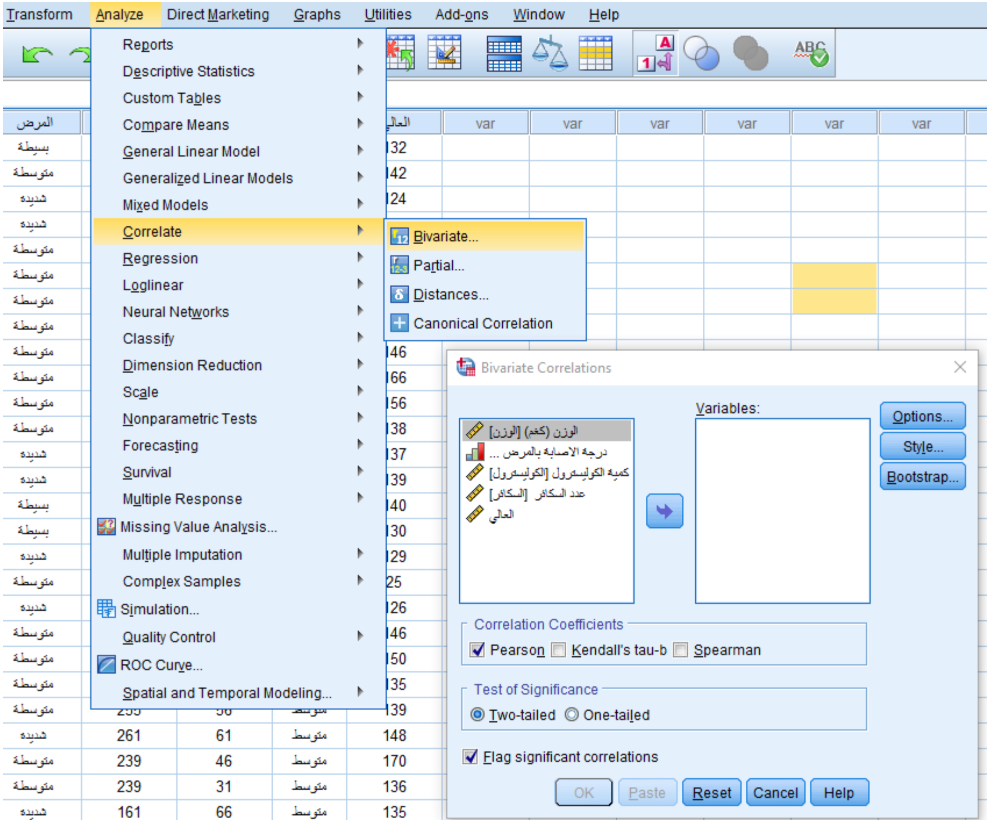
الخطوة الأولى هي نقل البيانات الى محرر البيانات في برنامج (SPSS) كما في الشكل (7-4).



الوزن	المرض	الكوليسترول	السكائر	الواطئ	العالي
70	بسيطة	247	57	بسيط	132
68	متوسطة	163	61	متوسط	142
86	شديدة	258	31	متوسط	124
59	شديدة	153	56	شديد	126
72	متوسطة	253	61	شديد	139
86	متوسطة	163	61	بسيط	158
50	متوسطة	253	61	بسيط	160
69	متوسطة	239	36	بسيط	128
84	متوسطة	158	51	متوسط	146
48	متوسطة	250	51	متوسط	166
75	متوسطة	253	61	متوسط	156
60	متوسطة	163	61	شديد	138
95	شديدة	161	71	شديد	137
66	شديدة	255	46	شديد	139
58	بسيطة	231	56	متوسط	140
69	بسيطة	250	56	شديد	130
76	شديدة	250	56	بسيط	129
95	متوسطة	158	56	بسيط	25
88	شديدة	155	61	بسيط	126
85	متوسطة	253	46	شديد	146
64	متوسطة	166	41	شديد	150
74	متوسطة	172	66	شديد	135
41	متوسطة	255	56	متوسط	139
80	شديدة	261	61	متوسط	148
62	متوسطة	239	46	متوسط	170
53	متوسطة	239	31	متوسط	136

الشكل (7-4) ادخال البيانات في محرر البيانات

1- حساب معامل الارتباط الكمي بين المتغيرات الكمية يكون عبر الامر (Analyze) من قائمة الأوامر الرئيسة ومن ثم الامر الفرعي (Correlate) ومنه نختار الامر الفرعي (Bivariate) فنحصل على مربع الحوار كما في الشكل (7-5).



الشكل (7-5) مربع الحوار لحساب الارتباط

بعد ذلك ننقل المتغيرات الكمية الى جانب (Variables) ثم نضع علامة صح على المربع الذي يمثل نوع الارتباط تحت عنوان (Correlation Coefficients) وإذا رغبنا في اختبار الفرضية الإحصائية حول معنوية معامل الارتباط احصائيا، نؤشر الاختبار باتجاهين تحت مربع (Test of Significance) وكذلك نضع علامة صح في المربع المقابل لتأشير معنوية المعامل من خلال وضع نجمة مقابل المعامل المعنوي احصائيا وحسب مستوى المعنوية (0.05) او (0.01) وبعد ذلك نختار الامر (OK) فتكون النتائج كما في الجدول (7-10).

جدول (7−10) قيم معامل الارتباط البسيط بين المتغيرات الكمية

Correlations

		الوزن (كغم)	كمية الكوليسترول	عدد السكائر	ضغط الدم العالي
الوزن (كغم)	Pearson Correlation	1	- 0.107	0.194	- 0.340*
	Sig. (2-tailed)		0.529	0.249	0.040
	N		37	37	37
كمية الكوليسترول	Pearson Correlation		1	- 0.261	0.144
	Sig. (2-tailed)			0.119	0.395
	N			37	37
عدد السكائر	Pearson Correlation			1	0.010
	Sig. (2-tailed)				0.955
	N				37

النتائج في الجدول (7−10) توضح قيم معامل الارتباط البسيط بين كل متغيرين، ومع اتجاه العلاقة هل هي سالبة ام موجبة وبشكل عام حسب هذه القيم تعتبر العلاقات ضعيفة بين المتغيرات لان قيم معامل الارتباط كانت اقل من 0.5

2− بعد نقل المتغيرات الوصفية تحت مربع (Variables) نحتاج الى وضع علامة صح مقابل معامل سبيرمان بدلا من بيرسون فتكون النتائج كما في الجدول (7−11).

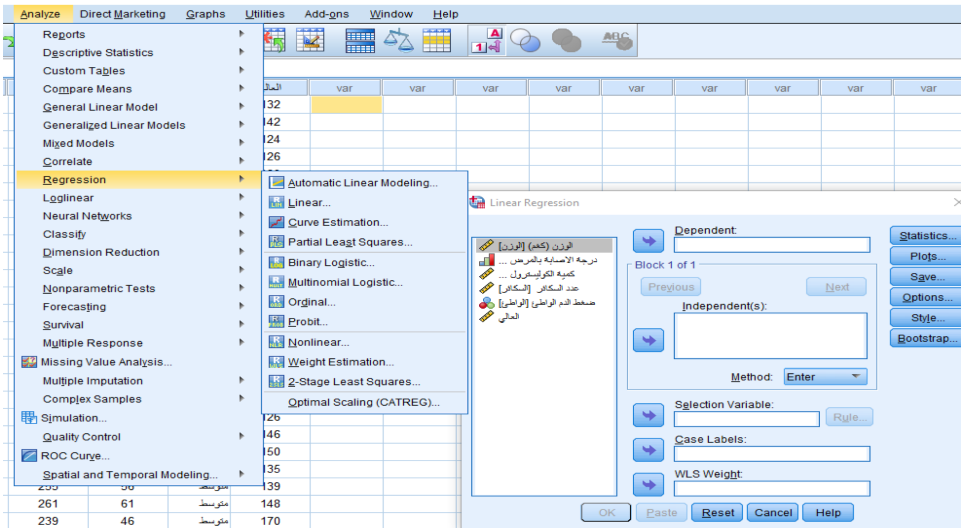
جدول (7-11) معامل ارتباط سبيرمان بين المتغيرات الوصفية

Correlations

			درجة الاصابة بالمرض	ضغط الدم الواطئ
Spearman's rho	درجة الاصابة بالمرض	Correlation Coefficient	1.000	.0490
		Sig. (2-tailed)	.	.7740
		N	37	37
	ضغط الدم الواطئ	Correlation Coefficient	.0490	1.000
		Sig. (2-tailed)	.7740	.
		N	37	37

نلاحظ من الجدول (7-11) بان العلاقة موجبة ضعيفة بين المتغيرين، لان قيمة معامل الارتباط اقل من 0.5.

3- حساب معامل الارتباط المتعدد يكون عن طريق استعمال الامر الرئيسي (Analyze) ومنه نأخذ الامر الفرعي (Regression) ومنه نأخذ الامر الفرعي (Linear) نحصل على مربع الحوار كما في الشكل (7-6).



الشكل (7-6) مربع الحوار لإيجاد معامل الارتباط المتعدد

بعد ذلك ننقل متغير الوزن في المربع (Dependent) والمتغيرات الأخرى في مربع (Independents) وننقر على الامر (OK) نحصل على الجدول (7-12).

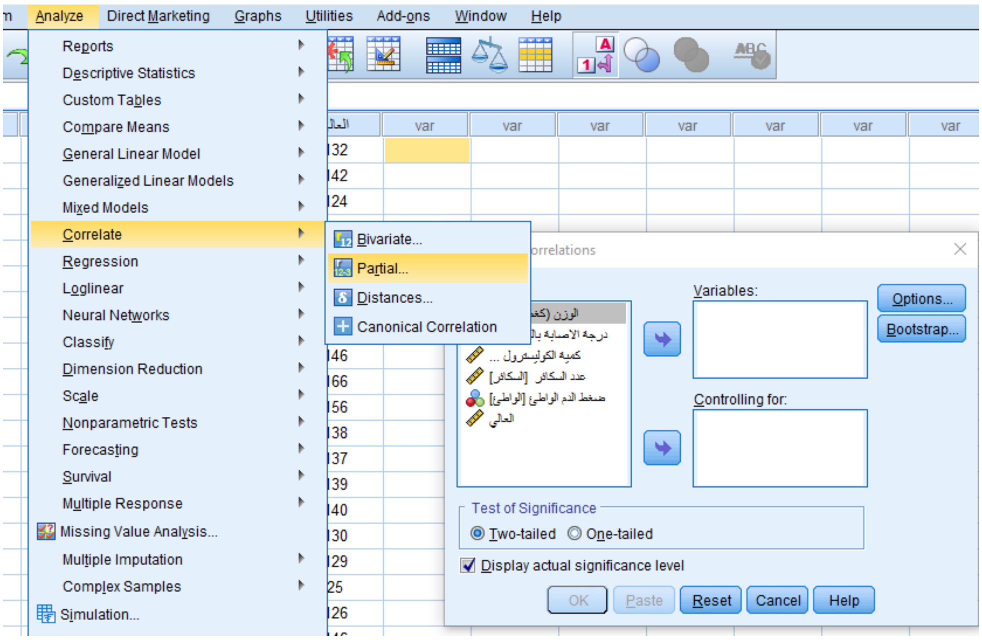
جدول (7-12) معامل الارتباط المتعدد

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	0.470[a]	0.221	0.095	14.817

نلاحظ بان العلاقة بين المتغير المعتمد الوزن وباقي المتغيرات كانت موجبة وضعيفة، لان معامل الارتباط المتعدد كان اقل من 0.5.

4- لحساب معامل الارتباط الجزئي بين المتغيرات الكمية يكون من الامر الرئيسي (Analyze) ومن ثم الامر الفرعي (Correlate) ومنه نختار الامر الفرعي (Partial) فنحصل على مربع الحوار كما في الشكل (7-7).



الشكل (7-7) مربع حوار حساب الارتباط الجزئي

ننقل متغير الوزن تحت مربع (Controlling for) وباقي المتغيرات الكمية تحت مربع (Variables) ثم ننقر على الامر (OK) فنحصل على الجدول (7–13)

جدول (7–13) معاملات الارتباط الجزئي بين المتغيرات الكمية

Partial Correlations

Control Variables			كمية الكوليسترول	عدد السكائر	ضغط الدم العالي
الوزن (كغم)	كمية الكوليسترول	Correlation	1.000	- 0.246	0.115
	عدد السكائر	Correlation	- 0.246	1.000	0.082
	العالي	Correlation	0.115	0.082	1.000

أظهرت النتائج في الجدول (7–13) بان معاملات الارتباط الجزئي بين متغير الوزن والمتغيرات الأخرى كانت ضعيفة وقسم موجبة والأخرى سالبه.

تمارين الفصل السابع

1- ما الفرق بين الارتباط حسب معامل بيرسون ومعامل ارتباط سبيرمان.

2- ما الفرق بين الارتباط الكمي البسيط والارتباط الجزئي.

3- ما الفرق بين الارتباط الجزئي والارتباط المتعدد.

4- البيانات التالية تمثل مجموعة من المتغيرات في مجموعة من الدول والمطلوب باستعمال البرنامج الاحصائي (SPSS):

أ- حساب معامل الارتباط البسيط بين المتغيرات المختلفة (معامل بيرسون).

ب- حساب معامل ارتباط الرتب سبيرمان بين المتغيرات المختلفة.

ت- حساب معامل الارتباط المتعدد بين المتغيرات (Y) و (X) و(Z).

ث- حساب معامل الارتباط الجزئي بين المتغيرات المختلفة.

نصيب الفرد من الدخل القومي (Y)	العمر المتوقع للفرد W	نسبة الذين تقل أعمارهم عن 15 سنة (R)	معدل الوفيات /1000 طفل (Z)	معدل الخصوبة /1000 نسمة (X)	البلد
2060	65	44	61	35	1
600	58	41	73	32	2
1420	65	35	44	27	3
370	48	45	114	46	4
240	41	44	147	48	5
120	46	46	139	47	6
940	54	46	85	44	7
20450	73	21	7.1	15	8
21700	72	22	9	16	9
910	75	36	15.3	27	10
590	62	46	69	40	11
22090	72	17	7.5	13	12
26070	71	19	5.8	13	13
960	65	41	57	31	14

5- تم اجراء اختبارات لقياس مستوى الاوستيرول (Estriol) (ملغم/24 ساعة) في الادرار للنساء في بداية الحمل، لأنه وجد في معظم الدراسات السابقة بان نسبة الاوستيرول لها تأثير على وزن الطفل عند الولادة. الاختبار يمكن ان يوضح السبب

غير المباشر للوزن غير الطبيعي للمولود. ولاختبار العلاقة بين مستوى الاوستيرول والوزن عند الولادة تم استعمال البيانات في الجدول الاتي لعينة من (31) من النساء بداية الحمل. والمطلوب استعمال برنامج (SPSS) لحساب معامل الارتباط المناسب لهذه البيانات.

ت	الاوستيرول (ملغم/24ساعة) (x_i)	الوزن عند الولادة (كغم) (y_i)	ت	الاوستيرول (ملغم/24ساعة) (x_i)	الوزن عند الولادة (كغم) (y_i)
1	7	2.5	16	16	3.2
2	9	2.5	17	17	3.2
3	9	2.5	18	25	3.2
4	12	2.5	19	27	3.4
5	14	2.7	20	15	3.4
6	16	2.7	21	15	3.4
7	16	2.7	22	15	3.5
8	14	2.4	23	16	3.5
9	16	3.0	24	19	3.4
10	16	3.0	25	18	3.5
11	17	3.1	26	17	3.6
12	19	3.0	27	18	3.7
13	21	3.0	28	20	3.8
14	24	2.8	29	22	4.0
15	15	3.2	30	25	3.9
			31	24	4.3

6- توافرت لديك البيانات في الجدول الآتي والتي تمثل ضغط الدم العالي ((SBP mm Hg (Y)) والوزن عند الولادة (كغم) (X_1) والعمر بالأيام (X_2) لمجموعة من الاطفال حديثي الولادة. **المطلوب:**

أ- إيجاد معامل الارتباط البسيط (معامل بيرسون) .

ب- (معامل سبيرمان) بين المتغيرات المتعددة وبين معنوية كل معامل.

ت-معامل الارتباط المتعدد.

ث- معامل الارتباط الجزئي.

الرقم	الوزن عند الولادة (X_1) (كغم)	العمر بالأيام (X_2)	ضغط الدم العالي (ملغم/دل) (Y)	الرقم	X_1	X_2	Y
1	3.8	3	89	17	2.8	3	79
2	3.4	4	90	18	2.5	4	82
3	2.8	3	83	19	2.8	3	91
4	3.0	2	77	20	3.2	2	83
5	3.7	4	92	21	3.1	5	78
6	3.5	5	98	22	2.9	6	78
7	3.5	2	82	23	2.9	2	76
8	3.0	3	85	24	3.4	4	84
9	3.4	5	96	25	3.4	5	92
10	2.6	4	95	26	4.6	8	93
11	3.4	2	80	27	4.2	7	76
12	2.7	3	79	28	3.8	2	85
13	3.4	3	86	29	3.6	3	92
14	4.3	4	97	30	3.4	5	73
15	4.5	3	92	31	3.5	4	82
16	3.5	3	88	32	4.1	6	91

7- درجة ابغار (Apgar) طورت في عام 1952 كمقياس للحالة الفيزيائية للمولود في الدقيقة الأولى والخامسة بعد الولادة. الدرجة يمكن الحصول عليها من خلال جمع خمسة عناصر الذي يمكن بإعطائها رتب 0 ،1، او 2 لتمثل عناصر مختلفة لحالة المولود عند الولادة، فاذا توافرت لديك البيانات الآتية

التي تمثل الدرجة عند الدقيقة الأولى والدقيقة الخامسة بعد الولادة لعدد من الولادات الحديثة والمطلوب: استعمال ارتباط سبيرمان لتقدير العلاقة بين هذه الدرجات مع تفسير النتيجة.

المولود	درجة ابغار 1 دقيقة	درجة ابغار 5 دقيقة	المولود	درجة ابغار 1دقيقة	درجة ابغار 5 دقيقة
1	10	10	**13**	6	9
2	3	6	**14**	8	10
3	8	9	**15**	9	10
4	9	10	**16**	9	10
5	8	9	**17**	9	10
6	9	10	**18**	9	9
7	8	9	**19**	8	10
8	8	9	**20**	9	9
9	8	9	**21**	3	3
10	8	9	**22**	9	9
11	7	9	**23**	7	10
12	8	9	**24**	10	10

الفصل الثامن

تحليل الانحدار

Regression Analysis

8–1 تمهيد

تطرقنا في الفصل السابق الى المقاييس التي تستعمل لقياس العلاقة بين الظواهر او المتغيرات التي تمثل هذه الظواهر، ومنها تم التعرف على كيفية تحديد قوة العلاقة واتجاهها، وفي الجانب الاخر تأتي الحاجة أحيانا الى صياغة علاقة رياضية تربط بين هذه المتغيرات لقياس الأثر لمتغير واحد او مجموعة من المتغيرات على متغير اخر او صفة أخرى تسمى متغير المعتمد (الاستجابة)، وهذا ما يسمى بتحليل الانحدار، ويعد الانحدار من المواضيع الواسعة الانتشار والمهمة في التحليل الاحصائي لإنه في النتيجة نحصل على نموذج او معادلة رياضية تمثل العلاقة بين متغير واحد يسمى المتغير المعتمد او الاستجابة (Dependent (Response) variable) ومتغير او عدة متغيرات تسمى المتغيرات التوضيحية او المستقلة (Explanatory (independent) variables)) وهذا هو موضوع هذا الفصل والمباحث التالية توضح بعض المفاهيم عن الانحدار لأنه موضوع واسع جدا.

8–2 تعريف الانحدار Regression Definition

تحليل الانحدار من أكثر الأدوات المستعملة في التحليل الاحصائي لذا يكون البدء بتحديد الخطوط العريضة لتحليل الانحدار. ونبدأ بالسؤال الأساسي: ما هو تحليل الانحدار؟ تحليل الانحدار يهتم بوصف العلاقة وتقويمها بين متغير (عادة يسمى المتغير المعتمد او الاستجابة) وواحد آو أكثر من المتغيرات الأخرى (تسمى عادة المتغيرات التوضيحية آو المتغيرات المستقلة). ويرمز للمتغير المعتمد بـ (Y) والمتغيرات التوضيحية بـ ($x_k \ldots x_3 \;\; x_2 \;\; x_1$).

يعد تحليل الانحدار أكثر طرق التحليل الإحصائي استخداما، إذ يمكن التنبؤ عبره بقيمة أحد المتغيرات (المتغير المعتمد) عند قيمة محددة لمتغير أو متغيرات أخرى (المتغيرات المستقلة). وتسمى العلاقة الرياضية التي تصف سلوك المتغيرات محل الدراسة التي من خلالها يكون التنبؤ بسلوك أحد المتغيرين عند معرفة الآخر بمعادلة خط الانحدار. ولذلك يعرف تحليل الانحدار على انه أداة إحصائية تقوم ببناء نموذج احصائي لتقدير العلاقة بين متغير كمي او وصفي واحد يسمى المتغير المعتمد ومتغير واحد او أكثر من المتغيرات الكمية او الوصفية والتي تسمى المتغيرات التوضيحية او المستقلة، إذ ينتج معادلة إحصائية توضح هذه العلاقة بين المتغيرات. ويمكن استعمال هذه المعادلة لمعرفة نوع العلاقة بين هذه المتغيرات وتقدير او التنبؤ بقيمة المتغير المعتمد اعتمادا على قيم المتغيرات التوضيحية.

ويجب ان تتوافر شروط أساسية لأجراء تحليل الانحدار حتى تكون النتائج دقيقة ويمكن الوثوق بها، إذ ينبغي ان يكون توزيع المتغيرين المستقل والمعتمد توزيعاً طبيعياً، كما ينبغي ان تكون العينة مختارة بشكل عشوائي.

8-3 أنواع نماذج الانحدار Types of Regression models

يمكن ان تقسم نماذج الانحدار الى أنواع عدة وحسب عدد من الاعتبارات وكالاتي:

أولا: الاعتبار الأول هو نوع البيانات التي تمثل المتغير المعتمد (Y). يقسم الانحدار على وفق هذا الاعتبار الى نوعين هما:

أ- **الانحدار الكمي** ويكون فيه المتغير المعتمد قياسات كمية مثل الوزن (كغم)، او الكوليسترول (مليمول /لتر) وغيرها من المتغيرات الكمية.

ب-**الانحدار الوصفي** ويكون فيه المتغير المعتمد صفات اما ثنائية او أكثر من صفتين مثل حالة المريض (مصاب او غير مصاب)، او مثل الإجابة على أسئلة الاستبانة بموافق جدا، موافق، محايد، غير موافق، غير موافق جدا. وغيرها من المتغيرات الوصفية.

ثانيا: الاعتبار الثاني يمثل عدد المتغيرات التوضيحية فان كلا النوعين السابقين سوف يقسمان الى نوعين حسب عدد المتغيرات التوضيحية في النموذج وكالاتي:

أ- انحدار كمي او وصفي بسيط وفيه يكون متغيراً توضيحياً واحداً.

ب-انحدار كمي او وصفي متعدد وفيه يكون أكثر من متغير توضيحي.

ثالثا: الاعتبار الثالث هو نوع العلاقة التي تربط المتغير المعتمد مع المتغير او المتغيرات التوضيحية ويقسم كلا النوعين الى:

أ- الانحدار الخطي وتكون فيه العلاقة تمثل خطاً مستقيماً.

ب-الانحدار غير الخطي وتكون فيه العلاقة بين المتغيرات غير خطية.

وسيقتصر هذا الفصل على انواع الانحدار الكمي الخطي ولمزيد من التفاصيل يمكن الرجوع الى مراجع تحليل الانحدار المختصة بهذا الموضوع للتعرف على الأنواع الاخرى.

8-4 نموج الانحدار الكمي البسيط Simple quantitative Regression model

الانحدار هو أسلوب يمكن بواسطته تقدير قيمة أحد المتغيرين بمعلومية قيمة المتغير الآخر عن طريق معادلة الانحدار، وتهدف دراسة الانحدار الى التنبؤ بقيمة متغير (Y)بمعرفة متغير آخر(X) ويعرف المتغير الأول بالمتغير المعتمد (dependent variable)وفي حين يعرف المتغير الآخر بالمتغير التوضيحي او المستقل (Explanatory or Independent variable) وجاءت التسمية للمتغير التوضيحي او المستقل لأنه يكون الأساس والمؤثر على المتغير المعتمد ، اما المتغير المعتمد فانه تتعين قيمته تبعاً لقيمة المتغير التوضيحي، وسمي الانحدار في هذه الحالة بالانحدار البسيط لوجود متغيرين فقط معتمد وتوضيحي، وعند ذكر كلمة الخطي نعني بها خط الانحدار. والغرض من استخدام أسلوب تحليل الانحدار الخطي البسيط، هو دراسة وتحليل أثر متغير كمي في متغير كمي آخر ومن الأمثلة النموذجية على تحليل الانحدار.

أ- اعتماد ضغط الدم Y على عمر الشخصX ،
ب - اعتماد الوزن لحيوانات التجربة Y على معدل التغذية اليومي X
ت - دراسة أثر كمية السماد على إنتاجية الدونم.
ث - دراسة أثر كمية البروتين التي تتناولها الأبقار على الزيادة في الوزن.
ويلاحظ من ذلك أن نموذج الانحدار يعتمد دائماً على العلاقة السببية بمعنى ان يكون التغير في المتغير المستقل مسبب رئيسي للتغير في المتغير المعتمد.

يعتبر تحديد المتغيرات التي تدخل في العلاقة او النموذج الخطوة الأولى في بناء نموذج الانحدار وإذا كانت هذه العلاقة بسيطة أي بين اثنين من المتغيرات ففي هذه الحالة عادة ما نفكر بأحد المتغيرين الأول وهو المتغير التوضيحي (X) والمتغير الآخر هو المتغير المعتمد أو متغير الاستجابة (Y) ويعتبر دالة إلى المتغير التوضيحي وكما في العلاقة الاتية:

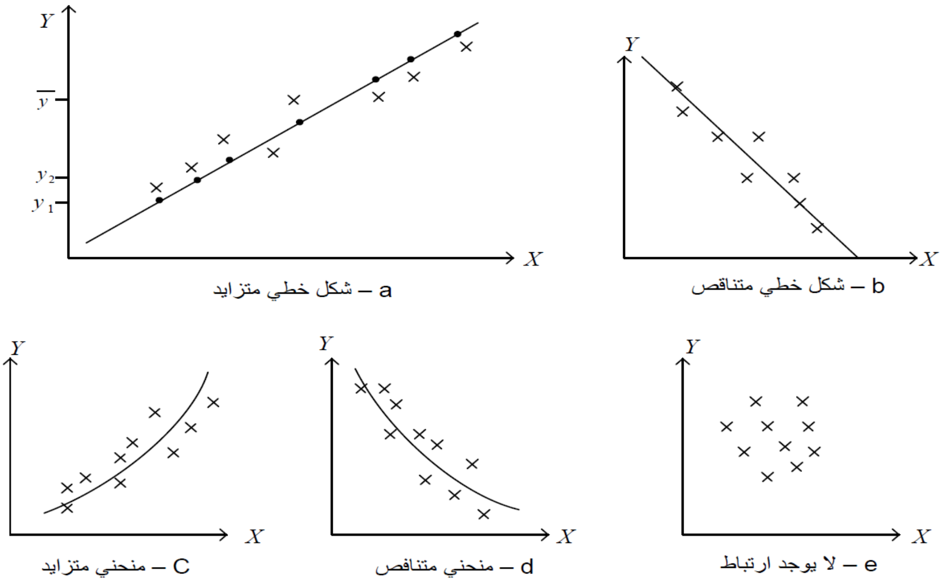
$$Y = f(X) \qquad (1-8)$$

والخطوة الثانية: هي تحديد شكل العلاقة بين (X ,Y) مثل هذه العلاقة من الصعوبة تحديدها بشكل دقيق، ولكن يمكن الاعتماد على النظرية التي بموجبها تم اختيار المتغيرات المهمة في العلاقة للاستفادة منها في اقتراح شكل العلاقة أو اقتراح بعض الشروط الجانبية حول معاملات الانحدار، مثل هذه الفرضيات يمكن أن تساعد في تحديد أشكال عدة للعلاقة، بعد ذلك يمكن الاستفادة من التحليل الإحصائي للمساعدة في الاختيار بين هذه الصيغ. ولدراسة مثل هذه العلاقات في المجتمع المفروض نسحب عينة عشوائية من عناصره بحجم n عنصراً، ونأخذ قيم كل منY وX . ونضعها في جدول منظم كما في الجدول (8-1):

جدول (8−1) عناصر العينة في نموذج الانحدار

i	1	2	3	…	i	…	n
X	x_1	x_2	x_3	…	x_i	…	x_n
Y	y_1	y_2	y_3	…	y_i	…	y_n

وعند رسم هذه النقاط المتقابلة (y_i, x_i) على المستوى (YOX) نحصل على ما يسمى بشكل الانتشار، والذي يأخذ أحد الأشكال التالية:



الشكل (8−1) اشكال الانتشار الممكنة لبيانات العينة

ومن الشكل (8−1) نلاحظ أنه يوجد لدينا عدة أشكال ممكنة للعلاقة بين Y و X هي:

أ- الانحدار الخطي المتزايد او المتناقص كما في الشكلين(a,b).

ب- الانحدار غير الخطي المتزايد والمتناقص كما في الشكلين (c,d).

ت-عدم وجود ارتباط بين المتغيرين Y و X كما في الرسم(e).

وسنركز اهتمامنا هنا على الانحـدار الخطي المبين في الرسمين a و b من الشكل $(1-8)$. وسنفترض أن الطبيعة العامة للعلاقة بين Y و X هي علاقة سببية، وان الصيغة الرياضية لها في المجتمع هي من الشكل الخطي الآتي:

$$Y = \beta_0 + \beta_1 X + \varepsilon \qquad (2\text{-}8)$$

إذ إن (ε) هو حد الخطأ العشوائي. وهو متغير عشوائي مؤلف من حدود مستقلة عن بعضها ومستقلة عن المتغير المستقلX ، وان توقعها يساوي الصفر $(E(\varepsilon) = 0)$، وان تباينها ثابت ويساوي (σ^2)، وهي تخضع للتوزيع الطبيعي $(N(0, \sigma^2))$.

أما المعلمات (β_1 و β_0) فهي أعداد حقيقية، تعمل على تحديد شكل المستقيم في المستوى (YoX) وبحيث يكون ميله مساوياً لـ (β_1) وقاطعه مع العمود (Y) مساوياً ل (β_0) ولكن بما ان هذه المعلمات مجهولة فإن وضع المستقيم في المستوى يبقى غير محدد.

ولتحديد الوضع المناسب لذلك المستقيم في المستوى نعتمد على بيانات العينة المسحوبة من ذلك المجتمع، ونرسم شكل الانتشار للنقاط (y_i, x_i)، فإذا كان شكلها يوحي لنا باتجاه خط مستقيم نفترض أن العلاقة بين Y و X في العينة هي علاقة خطية أيضاً وتأخذ الشكل الآتي:

$$Y = b_0 + b_1 X + e \qquad (4-8)$$

إذ إن (e) هو حد خطأ التقدير، وهو متغير عشوائي آخر يختلف عن (ε) ومؤلف من حدود مستقلة وخاضعة للتوزيع الطبيعي $(N(0, \sigma_e^2))$. لأن وضعية المستقيم تختلف من عينة لأخرى.

ولغرض تقدير هذا النموذج باستعمال البيانات المتوفرة من العينة عن المتغيرين المعتمد والتوضيحي يتم التأكد من توفر الافتراضات الآتية ولو بشكل تقريبي:

$1-$ أن تكون قياسات المتغير المستقل (X) دقيقة ومحددة (غير عشوائية).

2- أن يكون المتغير (X) مؤثرا على المتغير (Y) ويساهم في تفسير تغيراته (أي ان العلاقة بينهما سببية).

3-أن لا يقل عدد المشاهدات الزوجية للمتغيرين عن (30) مشاهدة.

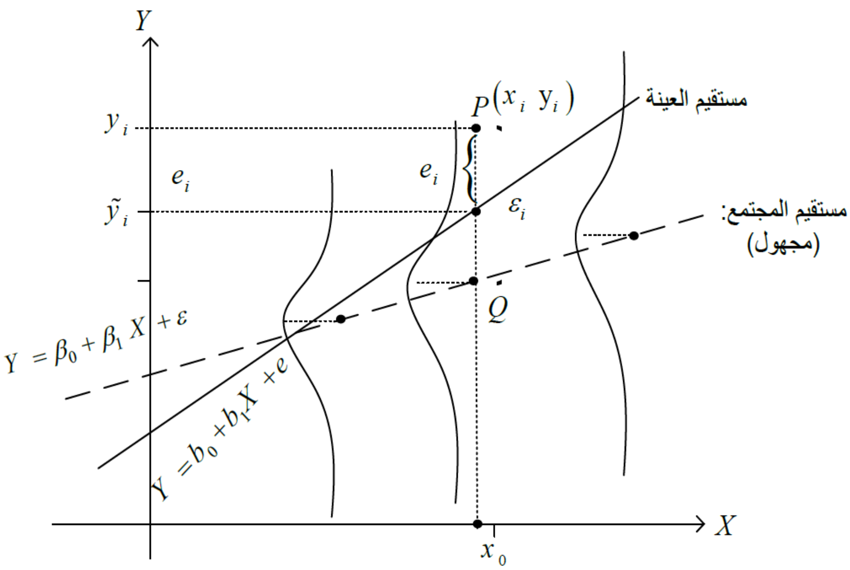
4-أن تكون القيمة المتوقعة لحدود الخطأ العشوائي (ε) مساوية للصفر ($E(\varepsilon)=0$).

5-أن يكون قيمة تباين حدود الخطأ العشوائي (ε) ثابتة عند كل قيم المتغير (X) وتساوي (σ_{ε}^{2}).

6-أن تكون حدود الخطأ العشوائي (ε) مستقلة عن بعضها البعض وتخضع للتوزيع الطبيعي ($N(0,\sigma^{2})$).

7-أن تكون حدود الخطأ العشوائي (ε) مستقلة عن المتغير المستقل (X).

ولتوضيح هذه الأمور نعرضها على الشكل البياني (8-2):



الشكل (8-2) مستقيما الانحدار فى المجتمع والعينة

وحتى يكون مستقيم الانحدار في العينة ممثلاً لشكل الانتشار يجب أن يأخذ أفضل وضعية ممكنة بين نقاط الانتشار، وهذه الوضعية هي التي تجعل مجموع مربعات أخطاء التقدير $(\sum_{i=1}^{n} e_i^2)$ أصغر ما يمكن، وهذا هو مبدأ طريقة المربعات الصغرى ونكتبه كما في المعادلة (8–5):

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}(y_i - b_0 - b_1 x_i)^2 \rightarrow Min. \quad (5\text{–}8)$$

ولتقدير معلمات النموذج حسب مبدأ المربعات الصغرى من بيانات العينة يمكن حل المعادلتين الاتيتين:

$$\hat{b}_1 = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2}, \quad (6\text{–}8)$$

$$\hat{b}_0 = \bar{y} - \hat{b}_1 \bar{x} \quad (7\text{–}8)$$

ولتقدير هذه المعلمات سوف نستعمل المثال الاتي:

مثال1:

يطلب طبيب الولادة اجراء اختبارات لقياس مستوى الاوستيرول (Estriol) (ملغم/24 ساعة) في الادرار للنساء في بداية الحمل لأنه وجد بان نسبة الاوستيرول لها تأثير على وزن الطفل عند الولادة. الاختبار يمكن ان يوضح السبب غير المباشر للوزن غير الطبيعي للوليد. ولاختبار العلاقة بين مستوى الاوستيرول والوزن عند الولادة تم استعمال البيانات في الجدول (8–2) لعينة من النساء بداية الحمل.

جدول $(2-8)$ مستوى الاوستيرول والوزن عند الولادة

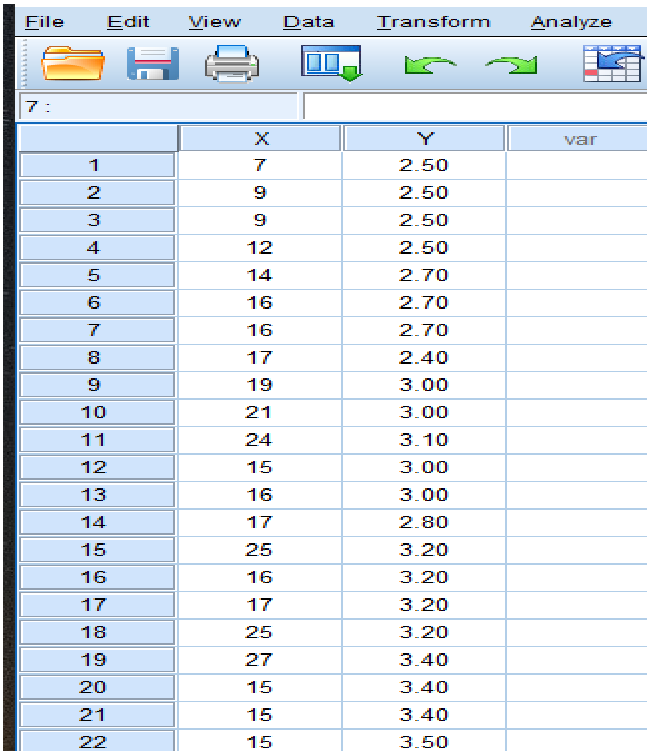
ت	الاوستيرول (ملغم/24ساعة) (x_i)	الوزن عند الولادة (كغم) (y_i)	ت	الاوستيرول (ملغم/24ساعة) (x_i)	الوزن عند الولادة (كغم) (y_i)
1	7	2.5	16	16	3.2
2	9	2.5	17	17	3.2
3	9	2.5	18	25	3.2
4	12	2.5	19	27	3.4
5	14	2.7	20	15	3.4
6	16	2.7	21	15	3.4
7	16	2.7	22	15	3.5
8	14	2.4	23	16	3.5
9	16	3.0	24	19	3.4
10	16	3.0	25	18	3.5
11	17	3.1	26	17	3.6
12	19	3.0	27	18	3.7
13	21	3.0	28	20	3.8
14	24	2.8	29	22	4.0
15	15	3.2	30	25	3.9
			31	24	4.3

الحل:

غالبا ما تكون الحسابات اليدوية غير سهلة وخصوصا عندما يكون حجم العينة كبيراً، واحتمال الخطأ يكون كبيراً في حالة الحسابات اليدوية، لذلك في الوقت الحاضر توفرت العديد من التطبيقات والبرامج الإحصائية التي تساعد على تسهيل الحل وعليه يمكن

حل المثال (1) باستعمال البرنامج الاحصائي (SPSS) الذي يعد واحد من البرامج الشائعة الاستعمال في مثل هذه المجالات وحسب الخطوات الاتية:

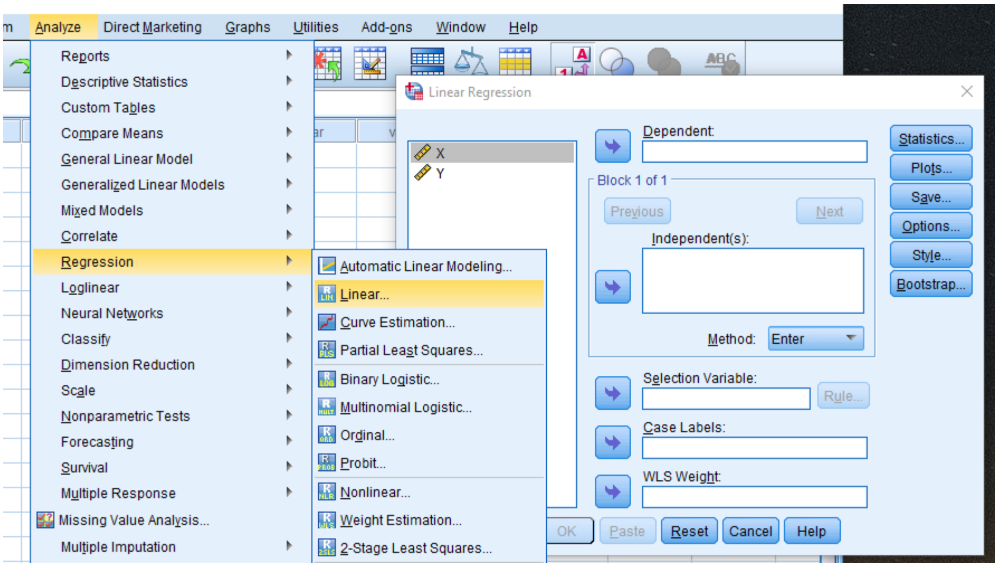
1- ادخال البيانات الى محرر البيانات بعد تسمية المتغيرات في محرر أسماء المتغيرات وكما في الشكل (8-3):



	X	Y	var
1	7	2.50	
2	9	2.50	
3	9	2.50	
4	12	2.50	
5	14	2.70	
6	16	2.70	
7	16	2.70	
8	17	2.40	
9	19	3.00	
10	21	3.00	
11	24	3.10	
12	15	3.00	
13	16	3.00	
14	17	2.80	
15	25	3.20	
16	16	3.20	
17	17	3.20	
18	25	3.20	
19	27	3.40	
20	15	3.40	
21	15	3.40	
22	15	3.50	

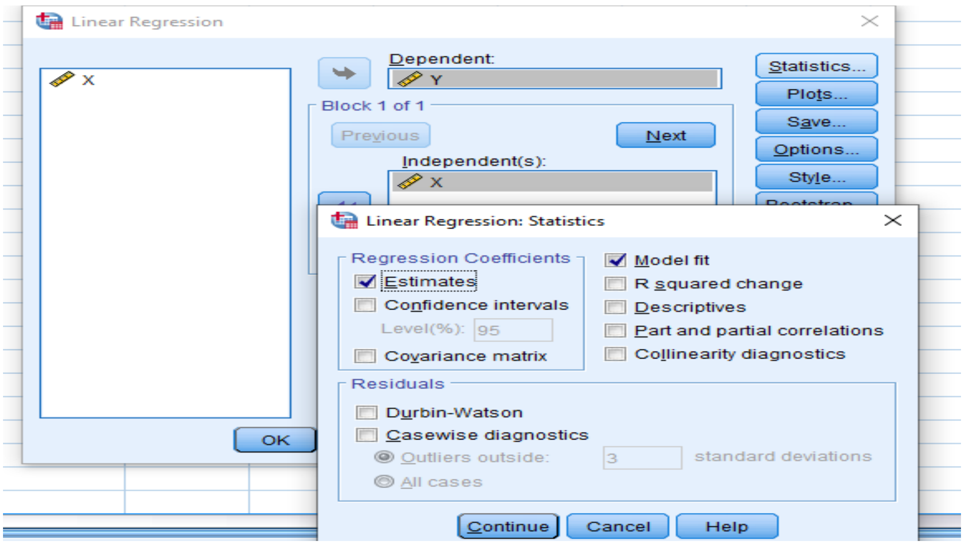
الشكل (8-3) طريقة ادخال البيانات في صفحة البيانات في الأعلى

2- من قائمة المهام في الأعلى نختار قائمة (Analyze) ثم نختار الامر الفرعي منها (Regression) ومنها نختار (Linear) وبعد الاختيار نحصل على صندوق الحوار كما في الشكل (8-4):



الشكل (8-4) اختيار قائمة التنفيذ من قائمة المهام

ثم ننقل المتغير المعتمد (Y) الى مربع المتغير (Dependent) وننقل المتغير التوضيحي (X) الى مربع المتغيرات التوضيحية (Independents) ثم نختار الامر (Statistics) لتحديد الاختبارات المطلوبة مع النموذج كما في الشكل (8-5)



الشكل (8-5) صندوق الحوار الخاص بكيفية رسم البيانات

فاذا كنا نرغب في الرسم نختار الاختيار المطلوب وإذا لم نختر نضغط على كلمة (OK) فتظهر لنا مجموعة من النتائج نختار منها ما يهمنا كالاتي:

أ- معادلة الانحدار البسيط اما نستعرضها بهذا الشكل كجدول او تكتب بالصيغ التي أسفل الجدول (8-3):

جدول (8-3) معادلة الانحدار البسيط

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.110	.2690		7.841	.0000
	X	.0610	.0150	.6080	4.128	.0000

$$\hat{y}_i = 2.11 + 0.061x_i$$

كذلك نستفيد من الجدول (8-3) لاختبار معنوية المعلمة المقدرة للنموذج إذ تستعمل الفرضية الإحصائية الاتية:

$$H_0: \beta = 0 \qquad VS \qquad H_1: \beta \neq 0$$

ومن الجدول (8-3) نجد بان قيمة دالة الاختبار كما في العمود الخامس ($t = 4.128$) وقيمة احتمال ان دالة الاختبار أكبر من قيمة محددة (P-value<0.001) كما في العمود السادس هي اقل من مستوى المعنوية ($\alpha = 0.05$) لذلك نرفض الفرضية الصفرية وبالتالي هذا يعني بان المعلمة المقدرة معنوية احصائيا أي ان المتغير التوضيحي له تأثير معنوي على المتغير المعتمد.

ب-اختبار جودة مطابقة النموذج حسب الفرضية الإحصائية التي تقول بان النموذج غير مناسب للاستعمال في التنبؤ او ان النموذج غير جيد ونستعمل جدول تحليل التباين مع اختبار (F) وكما في الجدول (8-4):

جدول (8-4) جدول تحليل التباين لحساب اختبار F

Model	Sum of Squares	df	Mean Square	F	Sig
Regression	2.668	1	2.668	17.038	<0.001
Residual	4.541	29	0.157		
Total	7.208	30			

ومن الجدول (8-4) نستعمل اختبار (F) لاختبار فرضية ان النموذج غير جيد من خلال استعمال قيمة ((P-value=P(>F)) وبما انها اقل من مستوى المعنوية 0.05 لذلك نرفض الفرضية الإحصائية التي تقول بان النموذج غير جيد وبالتالي فان النموذج جيد يمكن استعماله للتنبؤ في المستقبل.

ت – الاختبار الثالث هناك اختبار اخر كما في الجدول (8-5) وهو معامل التحديد $(R\ Sq\ =\ 37\%)$ او معامل التحديد المعدل $(Adj\ R\ Sq\ =\ 34.8\%)$ والذي تعني مقدار تفسير المتغير التوضيحي (X) للتغيرات التي تحصل في المتغير المعتمد (Y) ومن الجدول نجد بان قوة التفسير لهذا النموذج ضعيفة وهذا يعني بان المتغير التوضيحي يفسر من المتغير المعتمد ما نسبته حوالي (37%) والباقي يعود لعوامل اخرى لم يتم دخولها النموذج.

جدول (8-5) اختبار قوة تفسير النموذج

Model	R Square	Adjusted R Square
1	0.370	0.348

5-8 نموذج الانحدار الكمي المتعدد Multiple Quantitative Regression Model

يهتم تحليل الانحدار الخطي المتعدد بدراسة أثر عدة متغيرات مستقلة (توضيحية) كمية او مختلطة كمية مع وصفية على متغير واحد تابع كمي. فعلى سبيل المثال إذا كان (Y) يعبر عن المتغير المعتمد والمتغيرات $(X_1, X_2, \dots, X_k)$ تعبر عن (k) من المتغيرات التوضيحية، وان عدد المشاهدات هو(n) فان المشاهدة التابعة (y_i)، $(i = 1,2, \dots, n)$ يمكن التعبير عنها كدالة خطية في مجموعة المشاهدات التوضيحية $(x_{i1}, x_{i2}, \dots, x_{ik})$ كما يلي:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_{i2} x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i \qquad (8-8)$$

إذ ان $(\beta_0, \beta_1, \beta_2, \dots, \beta_k)$ تعبر عن معاملات الانحدار، (ε_i) يعبر عن الخطأ العشوائي للمشاهدة رقم (i)، $(i = 1,2, \dots, n)$ وحيث ان عدد المشاهدات (n)، اذن يكون لدينا (n) من المعادلات حسب الصيغة (8-8) وكالاتي:

$$y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \cdots + \beta_k x_{1k} + \varepsilon_1$$
$$y_2 = \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \cdots + \beta_k x_{2k} + \varepsilon_2$$
$$\dots\dots \qquad \dots\dots \qquad \dots\dots$$
$$y_n = \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \cdots + \beta_k x_{nk} + \varepsilon_n$$

يمكن صياغتها في صورة مصفوفات كما يلي:

$$\begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_i \\ \cdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} \dots & x_{1k} \\ 1 & x_{21} & x_{22} \dots & x_{2k} \\ \dots & \dots & \dots\dots & \dots \\ 1 & x_{i1} & x_{i2} \dots & x_{ik} \\ \dots & \dots & \dots\dots & \dots \\ 1 & x_{n1} & x_{n2} \dots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \dots \\ \beta_j \\ \dots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdots \\ \varepsilon_i \\ \cdots \\ \varepsilon_n \end{bmatrix} \qquad (9\text{-}8)$$

ومن ثم فان نموذج الانحدار المتعدد باستعمال رموز المصفوفات يكون كما في المعادلة (10-8).

$$\boldsymbol{Y} = \boldsymbol{XB} + \boldsymbol{\varepsilon} \qquad (10\text{-}8)$$

إذ ان: $(\boldsymbol{Y})$ يعبر عن متجه المشاهدات للمتغير التابع، وهو من درجة $(n \times 1)$ والعنصر رقم (i) في هذا المتجه هو (y_i). $(\boldsymbol{X})$ يمثل مصفوفة المشاهدات للمتغيرات التوضيحية، وهي من درجة $(n \times (k+1))$ والصف رقم (i) في هذه المصفوفة هو $(1\ x_{i1}\ x_{i2} \dots x_{ik})$، $(\boldsymbol{\beta})$ تعبر عن معامل الانحدار الجزئي لـ yعلى (x_i) Partial Regression Coefficientعندما تكون بقية المتغيرات المستقلة ثابتة. كما انها تمثل مقدار التغير في Y لزيادة وحدة واحدة في (X_j) عندما تكون بقية المتغيرات المستقلة ثابتة وهو من درجة $((k+1) \times 1)$. $(\boldsymbol{\varepsilon})$ يعبر عن متجه الاخطاء العشوائية وهو من درجة $(n \times 1)$ والعنصر رقم (i) هو (ε_i). ان المعادلة $(8-8)$ تدعى معادلة انحدار متعدد لأنها تحتوي على أكثر من متغير توضيحي وإنها خطية لان المعلمات هي من الدرجة الاولى أي ان اسها يساوي الواحد الصحيح. يستند نموذج الانحدار المبين في المعادلة $(8-8)$ على عدة افتراضات هي:

1- ان المتغير المعتمد $\boldsymbol{Y}$ هو متغير عشوائي وقيمه مستقلة احصائياً الواحدة عن الاخرى وتتوزع توزيعاً طبيعياً بوسط حسابي قدره $(\mu_{y/x_1x_{2,\dots}x_k})$ وتباين $(\sigma^2_{y/x_1x_{2,\dots,}x_k} = \sigma^2)$ أي ان متوسط Y هو دالة خطية:

$$\mu_{y/x_1x_{2,\dots}x_k} = E(y_i) = \beta_0 + \beta_1X_1 + \beta_2X_2 + \dots + \beta_kX_k$$

وان تباين Y هو: $(\sigma^2_{y/x_1\dots x_k} = \sigma^2_{\beta_0+\beta_1X_1+\dots+\beta_kX_k} = \sigma^2_\varepsilon)$ ان هذه الخاصية تسمى Homoscedasticity أي تجانس الخطأ.

2- مصفوفة المتغيرات التوضيحية $(\boldsymbol{X})$ محددة ومعطاة (Fixed) فهي مقاسة بدون اخطاء.

3- المتغيرات التوضيحية $(X_1, X_2, \dots, X_k)$ مستقلة احصائيا، وهذا يعني وجود استقلال خطي بين اعمدة المصفوفة ويعبر عن ذلك رياضيا بان $(Rank(X) = k+1 < n)$.

4- يوجد استقلال احصائي بين المتغيرات التوضيحية والخطأ العشوائي اي ان اعمدة المصفوفة $(\boldsymbol{X})$ مستقلة خطيا عن متجه الاخطاء العشوائية $(\boldsymbol{\varepsilon})$ ويعبر عن ذلك رياضيا كما يأتي:

$(Cov(X,\varepsilon) = E(X'\varepsilon) - [E(X)]'[E(\varepsilon)] = 0)$.

5- الاخطاء العشوائية (ε_i) لها توزيع طبيعي متوسطه صفر وتباينه ثابت (σ^2) من مشاهدة الى اخرى، اي ان $(\varepsilon_i \sim N(0,\sigma^2))$ وكذلك يفترض بان الاخطاء العشوائية مستقلة احصائيا ويعبر عن ذلك رياضيا كالاتي:

$$E(\varepsilon_i) = 0, \qquad Cov(\varepsilon_i, \varepsilon_j) = E(\varepsilon_i, \varepsilon_j) = \begin{cases} \sigma^2 & If \;\; i = j \\ 0 & if \;\; i \neq j \end{cases}$$

اي ان متجه الاخطاء العشوائية $(\boldsymbol{\varepsilon})$ له توزيع طبيعي متعدد متوسطه صفر وله مصفوفة تباين (Σ) حيث ان هذه المصفوفة متماثلة ومن الدرجة $(n \times n)$ ويعبر عنها كالاتي:

$$\Sigma = \begin{bmatrix} \sigma^2 & 0\dots & 0 \\ 0 & \sigma^2\dots & 0 \\ \dots & \dots & \dots \\ 0 & 0\cdots & \sigma^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & 0\dots & 0 \\ 0 & 1\dots & 0 \\ \dots & \dots & \dots \\ 0 & 0\cdots & 1 \end{bmatrix}$$

ويحتوي النموذج (8-8) على $(k+1)$ من المعاملات التي تمثل المتجه $(\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_k)')$ والذي يمكن تقديره باستعمال طريقة المربعات الصغرى (Ordinary Least Square(OLS)) ان أساس عمل طريقة المربعات الصغرى هو جعل مجموع مربعات البواقي او الخطأ اقل ما يمكن :

$$\varepsilon_i^2 = \left(y_i - (\beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik})\right)^2 \qquad (11-8)$$

وصولا الى المعادلات الطبيعية وبذلك يتكون لدينا متجه معلمات الانحدار المقدرة والذي يأخذ الصيغة الاتية:

$$\hat{B} = (X'X)^{-1}X'Y \qquad (12-8)$$

وعند تحقق الافتراضات السابقة الذكر يكون التقدير $(\hat{B})$ هو التقدير الخطي الافضل وغير المتحيز لمتجه معاملات الانحدار في النموذج (8-8). وبعد تقدير معاملات الانحدار نحصل على نموذج او معادلة يمكن استعمالها للتنبؤ بقيم المتغير المعتمد لاي قيمة من قيم المتغيرات التوضيحية ولغرض التأكد بان النموذج المقدر جيد يمكن الاعتماد علية في التقدير او التنبؤ يجب اجراء بعض الاختبارات وهناك مجموعة من المؤشرات او الاختبارات الاحصائية التي تؤيد او ترفض النموذج منها.

8-6 بعض مؤشرات جودة النموذج Some of Goodness of fit indicators

هناك مجموعة من المؤشرات الاحصائية او الاختبارات التي تستعمل للتأكد من جودة النموذج لغرض استعماله للتنبؤ منها:

1- **متوسط مربعات الخطأ (Mean Square Error (MSE)**: وهو يمثل تشتت الخطأ وهو يقيس مدى بعد القيم المقدرة عن القيم الحقيقية ويقاس حسب المعادلة الاتية:

$$MSE = \hat{\sigma}_e^2 = \frac{SSE}{n-k-1} = \frac{Y'Y - B'X'Y}{n-k-1} \qquad (13-8)$$

إذ ان: (SSE) يمثل مجموع مربعات الخطأ ، (n) يمثل عدد المشاهدات ، (k) يمثل عدد المتغيرات التوضيحية في النموذج.

2- **معامل التحديد (R^2) (Determinant Coefficient)**: يشير **معامل التحديد**، الذي يرمز له بـ R^2 أو"R squared" ، إلى نسبة التباين في المتغير التابع الذي يمكن التنبؤ به من خلال المتغير (أو المتغيرات) المستقلة. وهو يستعمل في النماذج الإحصائية التي يكون هدفها الرئيسي التنبؤ بالنتائج المستقبلية أو اختبار الفرضيات، على أساس المعلومات الأخرى ذات الصلة. يوفر معامل التحديد مقياسًا لنسبة التباين الكلي للنتائج التي أوضحها النموذج وكذلك يوضح بسبة مساهمة المتغيرات التوضيحية في تفسير الانحرافات الكلية للمتغير المعتمد. فانه عندما يكون النموذج مطابقا للبيانات الحقيقية فان قيمة معامل التحديد (R^2) تكون قريبة من الواحد اي ان القيم الحقيقية للمتغير التابع (Y) تكون قريبة او مطابقة للقيم المقدرة باستعمال النموذج ($\hat{Y}$). ويحسب حسب المعادلة الاتية:

$$R^2 = \frac{SSR}{SST} = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2} \qquad (14-8)$$

3- اختبار جودة النموذج

اختبار جودة نموذج الانحدار من خلال اختبار الفرضية الاحصائية التي تقول بان النموذج غير مناسب او غير جيد بمعنى ان جميع المتغيرات التوضيحية ليس لها تأثير على المتغير المعتمد وحسب الفرضية الاحصائية الاتية:

$$H_0: \beta_1 = \beta_2 = \beta_3 = \cdots = \beta_k = 0 \qquad \text{(النموذج غير جيد)}$$

مقابل الفرضية البديلة

$$H_1: At\ least\ one\ of\ them\ not\ equal\ zero \qquad \text{(النموذج جيد)}$$

ولاختبار هذه الفرضية يتم حساب إحصاءة الاختبار من خلال بناء جدول تحليل التباين كالاتي:

جدول (8–6) جدول تحليل التباين لحساب إحصاءة الاختبار (F)

مصدر التغير S.O.V	مجموع المربعات Sum Square (SS)	درجة الحرية df	متوسط المربعات MS	دالة الاختبار المحسوبة (F_c)
الانحدار	$SSR = \sum (\hat{y}_i - \bar{y})^2$	k	$MSR = SSR/k$	$F_c = MSR/MSE$
البواقي	$SSE = \sum (y_i - \hat{y}_i)^2$	n-k-1	$MSE = SSE/n-k-1$	
الكلي	$SST = \sum (y_i - \bar{y})^2$	n-1		

وبمقارنة (F_c) المحسوبة مع $(F_{(\alpha,k,n-k-1)})$ الجدولية التي يتم ايجادها من جدول (F) وحسب المعطيات اي عند مستوى معنوية (α) ودرجة حرية البسط (k) ودرجة حرية المقام $(n-k-1)$ نرفض الفرضية الصفرية إذا كانت المحسوبة أكبر من الجدولية والعكس صحيح، وإذا تم رفض الفرضية الصفرية فهذا يعني ان النموذج معنوي احصائيا اي يمكن الاعتماد عليه في التنبؤ المستقبلي والعكس عند قبول الفرضية الصفرية يعني ان النموذج غير معنوي احصائيا ولا يمكن استعماله للتنبؤ المستقبلي.

4– اختبار معنوية معاملات الانحدار

اختبار تأثير المتغيرات التوضيحية على المتغير المعتمد من خلال اختبار المعنوية الاحصائية لمعاملات الانحدار لكل متغير وحسب الفرضية الاحصائية الاتية:

فرض العدم: المتغير التوضيحي ليس له أثر معنوي على المتغير المعتمد $(H_0: \beta_j = 0)$.

الفرض البديل: المتغير التوضيحي ذو أثر معنوي على المتغير المعتمد $(H_1: \beta_j \neq 0)$.

وتستعمل إحصاءة الاختبار الاتية لاختبار رفض او عدم رفض الفرضية الصفرية:

$$t_c = \hat{b}_j / SE_{b_j} \qquad (15–8)$$

وبمقارنة قيمة (t_c) المحسوبة وقيمتها الجدولية التي يمكن ايجادها من جدول (t) وحسب المعطيات التي تمثل مستوى المعنوية (α) ودرجة الحرية $(n-k-1)$.

وعندما تكون المحسوبة أكبر قيمة من الجدولية نرفض الفرضية الصفرية وهذا يعني ان المتغير التوضيحية ذو أثر معنوي على المتغير التابع والعكس صحيح.

- **ارتباط المتغيرات المستقلة**

من الجوانب المهمة في إحصاءات تحليل الانحدار عملية تحديد مدى تداخل الارتباط بين المتغيرات المستقلة، فإذا كان الارتباط بين متغيرين مستقلين عالياً، فإن ذلك يعني أن هناك عوامل مشتركة كثيرة بينهما، بل ربما يكون المتغيرين هما تقريباً نفس المتغير مع اختلاف التسمية الظاهرية.

هذا الوضع يجعل نموذج الدراسة هشاً والنتائج التي يمكن أن يتوصل إليها الباحث غير دقيقة وغير موثوق بها. وبناء عليه فإن على الباحث اختبار العلاقة بين المتغيرات المستقلة في بحثه حيث يمكنه الاعتماد على ما يسمى بعامل تضخم التباين (VIF) أي Variance Inflation Factor والذي يستخرج من خلال تطبيق المعادلة التالية:

$$VIF = \frac{1}{1-R^2} \qquad (16\text{-}8)$$

هناك عدة اراء حول تحديد القيمة التي عندها يعد بان هناك تداخلاً بين المتغيرات فالرأي الأول يقول ان لا تزيد قيمة المؤشر عن الرقم 10 بينما بعضهم الاخر يقول ينبغي ألا تزيد قيمة VIF عن الرقم 5 والراي الثالث يقول ان لا تزيد عن 3 وسوف نعتمد الراي الثاني لأنه يمثل الوسط، فإن زادت قيمة المؤشر عن الرقم 5 فإن ذلك معناه أن هناك تداخلاً بين المتغيرات المستقلة في تأثيرها على المتغير التابع.

اصبح استعمال الحل اليدوي لحل مسائل الانحدار مضيعة للوقت والجهد مع احتمال الخطأ وخصوصا عندما تكون البيانات بأعداد كبيرة لذلك أصبحت مهمة الطالب او الباحث معرفة على الأقل استعمال احد التطبيقات الإحصائية العديدة والمتوافرة لحل مثل هذه المسائل، ومن ثم أصبحت مهمة الباحث هو معرفة كيفية اختيار الإجابات من خلال النتائج التي تعطيها مثل هذه التطبيقات وكيفية الوصول الى الإجابة الصحيحة والمناسبة ومن بين هذه التطبيقات تطبيق او برنامج (SPSS) والذي

استعمل لحل المسائل والامثلة في هذا الكتاب وسوف نستعمله لحل مسائل الانحدار الكمي البسيط والمتعدد وكما في الأمثلة الاتية.

<u>مثال 2:</u> توفرت لديك البيانات في جدول (8−7) والتي تمثل ضغط الدم العالي (SBP mm Hg (Y)) والوزن عند الولادة (كغم) (X_1) والعمر بالأيام (X_2) لمجموعة من الاطفال حديثي الولادة.

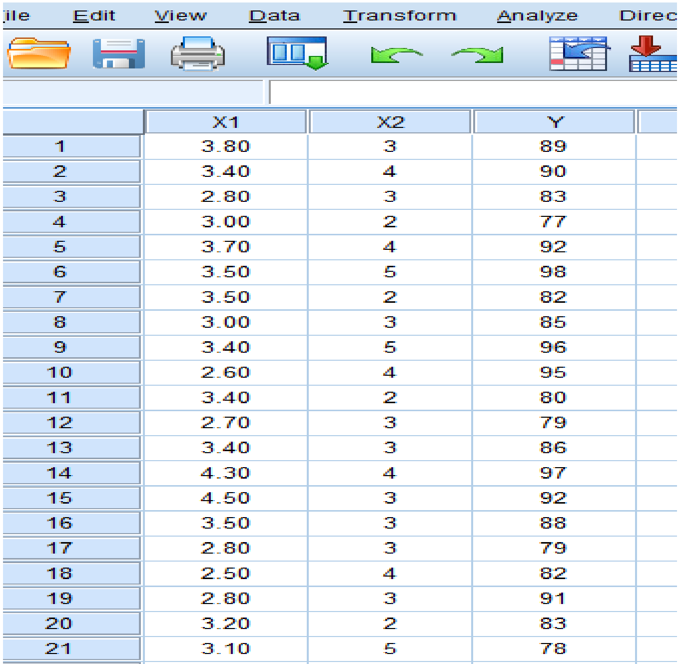
جدول (8−7) بيانات الأطفال حديثي الولادة

الرقم	الوزن عند الولادة (X1) (كغم)	العمر بالأيام (X2)	ضغط الدم العالي (ملغم/دل) (Y)	الرقم	X1	X2	Y
1	3.8	3	89	17	2.8	3	79
2	3.4	4	90	18	2.5	4	82
3	2.8	3	83	19	2.8	3	91
4	3.0	2	77	20	3.2	2	83
5	3.7	4	92	21	3.1	5	78
6	3.5	5	98	22	2.9	6	78
7	3.5	2	82	23	2.9	2	76
8	3.0	3	85	24	3.4	4	84
9	3.4	5	96	25	3.4	5	92
10	2.6	4	95	26	4.6	8	93
11	3.4	2	80	27	4.2	7	76
12	2.7	3	79	28	3.8	2	85
13	3.4	3	86	29	3.6	3	92
14	4.3	4	97	30	3.4	5	73
15	4.5	3	92	31	3.5	4	82
16	3.5	3	88	32	4.1	6	91

المطلوب: إيجاد معادلة الانحدار مع اجراء الاختبارات المناسبة لاختبار جودة النموذج.

الحل:

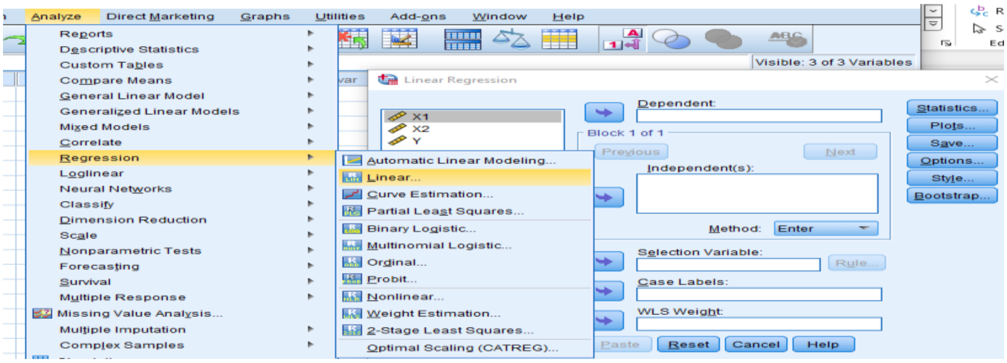
1- ادخل البيانات اعلاه في ثلاث متغيرات بالرموز (Y, X_1, X_2) كما في الشكل (8-6) في محرر البيانات لبرنامج (SPSS) بعد تحديد الأسماء للمتغيرات ضمن محرر الأسماء.



	X1	X2	Y
1	3.80	3	89
2	3.40	4	90
3	2.80	3	83
4	3.00	2	77
5	3.70	4	92
6	3.50	5	98
7	3.50	2	82
8	3.00	3	85
9	3.40	5	96
10	2.60	4	95
11	3.40	2	80
12	2.70	3	79
13	3.40	3	86
14	4.30	4	97
15	4.50	3	92
16	3.50	3	88
17	2.80	3	79
18	2.50	4	82
19	2.80	3	91
20	3.20	2	83
21	3.10	5	78

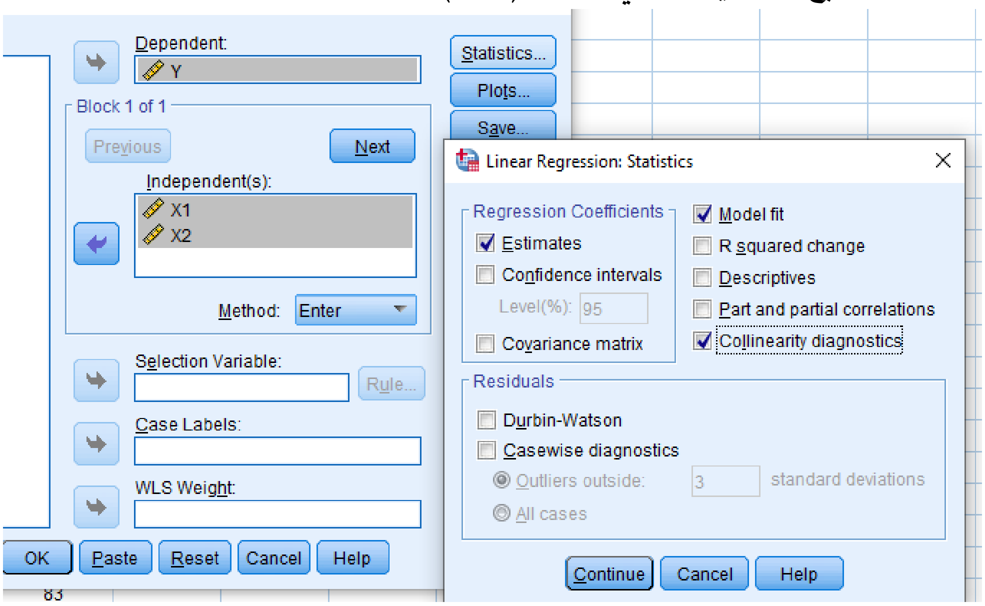
الشكل (8-6) ادخال البيانات في صفحة البيانات

2-من القائمة الرئيسية للأوامر نختار الامر (Analyze) ومنه نختار الامر الفرعي (Regression) ومن ثم نختار الامر الفرعي (Linear) فنحصل على المربع الحواري كما في الشكل (8-7).



الشكل (8-7) اختيار الامر المناسب للانحدار المتعدد

3- نحول المتغيرات في المربع الحواري فيكون المتغير التابع (Y) في المربع (Dependent) والمتغيرات التوضيحية في المربع (Independents) ثم نختار كلمة (Statistics) لتحديد الاختبارات المطلوبة مع النموذج من خلال المربع الحواري كما في الشكل (8-8).



الشكل (8-8) المربع الحواري لاختيار اختبارات جودة التوفيق للنموذج

4- ننقر على كلمة (Continue) لنرجع الى المربع الحواري الأصلي ومنه ننقر على كلمة (OK) لنحصل على النتائج كما في الجداول الاتية.

أ- معادلة الانحدار المتعدد المقدرة وكانت كما في الجدول (8-8):

جدول (8-8) معادلة الانحدار مع بعض الاختبارات

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	Constant	68.797	7.575		9.082	0.000		
	X1	4.781	2.384	0.371	2.006	0.054	0.854	1.171
	X2	0.186	0.852	0.041	0.219	0.828	0.854	1.171

ومن الجدول (8-8) يمكن كتابة معادلة الانحدار المقدرة بالشكل الاتي:

$$\hat{y}_i = 68.797 + 4.781x_{i1} + 0.186x_{i2}$$

ب-اختبارات جودة النموذج وكانت كالاتي

1- ومن الجدول نفسه ايضا يمكن اختبار معنوية المعلمات المقدرة حسب الفرضية في الفقرة 4 من المبحث ومن خلال استعمال المعادلة (8-15). ونجد في العمود الخامس قيمة دالة الاختبار (t) والعمود السادس يتضمن قيمة (P-value) وعند مقارنة قيمتها للمعلمتين نجد انهما اكبر من مستوى المعنوية 0.05 وبالتالي لا نرفض فرضية العدم والتي تعني بان المتغيرين ليس لهما تاثير معنوي احصائيا على المتغير المعتمد.

2- كذلك من الجدول (8-8) نحصل على الاختبار الثاني والذي يتعلق بمقدار الارتباط بين المتغيرات التوضيحية وحسب المعادلة (8-16) نجد بان المعاملات اقل من 5 وهذا يعني عدم وجود تعدد خطي بين المتغيرات التوضيحية.

3- اختبار جودة النموذج او المعنوية الإحصائية للنموذج وذلك باختبار الفرضية الاتية

$$H_0: \beta_1 = \beta_2 = \beta_3 = \cdots = \beta_k = 0 \qquad \text{(النموذج غير جيد)}$$

مقابل الفرضية البديلة

$$H_1: At\ least\ one\ of\ them\ not\ equal\ zero$$

(النموذج جيد)

ويتم ذلك باستعمال جدول تحليل التباين واختبار (F) كما في الجدول (8−9):

جدول (8−9) جدول تحليل التباين مع اختبار F

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	226.613	2	113.307	2.580	0.093
	Residual	1273.387	29	43.910		
	Total	1500.000	31			

أظهرت النتائج في الجدول (8−9) بان قيمة (P−value=0.093) للاختبار هي أكبر من مستوى المعنوية 0.05 وهذا يعني باننا لا نرفض الفرضية الصفرية وبالتالي هذا يعني بان النموذج غير معنوي احصائيا ولا يمكن الاعتماد علية في التنبؤ المستقبلي للمتغير المعتمد.

4− اختبار مقدار تفسير المتغيرات التوضيحية: أظهرت النتائج في الجدول (8−10) بان قوة التفسير للمتغيرات التوضيحية كانت ضعيفة وهذا ما يؤكد الاختبارات السابقة بان النموذج غير جيد لان قيمة معامل التحديد ($R^2 = 0.151$) تعتبر ضعيفة جدا وكذلك قيمته المعدلة كما نلاحظ في الجدول. كذلك كانت قيمة معامل الارتباط المتعدد بين المتغيرات التوضيحية والمتغير المعتمد ضعيفة (R=0.389).

جدول (8−10) معامل التحديد ومعامل التحديد المعدل

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	0.389	0.151	0.093	6.626

تمارين الفصل الثامن

1− يعد حجم الزفير القسري (FEV) مقياس معياري للأداء الرئوي. ولتحديد الأشخاص بالأداء الرئوي غير العادي يجب تحديد المعايير للأداء الرئوي الطبيعي للأشخاص. ومن الدراسات السابقة معروف بان الأداء الرئوي يرتبط بالطول والعمر. لذلك تم تجميع بيانات لمجموعة من الشباب بعمر يتراوح من 10 الى 30 سنة وكان قياس (FEV) متوسط باللتر لكل 12 سم من الطول. **والمطلوب:** أ− اوجد معادلة خط الانحدار ب− اختبر المعنوية الإحصائية للنموذج ت − ماهي نسبة التباين لـ (FEV) التي يمكن توضيحها من قبل المتغيرات:

ت	الطول سم	العمر سنة	متوسط الأداء لتر	ت	الطول سم	العمر سنة	متوسط الأداء لتر
1	134	10	1.7	11	174	21	3.8
2	138	12	1.9	12	180	22	3.9
3	142	13	2.0	13	182	23	4.0
4	146	14	2.1	14	185	24	4.2
5	150	15	2.2	15	188	25	4.3
6	156	16	2.5	16	190	26	4.4
7	158	17	2.7	17	192	27	4.5
8	162	18	3.0	18	194	28	4.6
9	166	19	3.1	19	196	29	4.7
10	170	20	3.4	20	198	30	4.9

2- ام باحث أكاديمي بدراسة عدد ساعات حضور طلبة مادة الإحصاء الحيوي عند عينة من الطلاب. بعد ذلك ظهرت نتيجة الطلاب في الامتحان وقام الباحث بتسجيل درجات كل طالب أمام ساعات الحضور كما أتي :

رقم الطالب	عدد ساعات الحضور	درجات الطالب
1	45	78
2	46	76
3	38	50
4	50	84
5	52	90
6	46	59
7	34	56
8	30	57
9	42	70
10	45	84
11	51	80
12	52	95
13	45	50
14	43	60

المطلوب :

- إيجاد معادلة الانحدار بين عدد ساعات الحضور ومستوى التحصيل العلمي للطالب مع التعليق على النتائج.

د - ماهي نسبة التفسير لتباين المتغير المعتمد وفقا در ة الطالب.

د - قدر درجة الطالب عندما يكون عدد ساعات حضوره 53 ساعة.

– البيانات الاتية تمثل نسبة الدهون في الجسم (Fat) وعدد من المتغيرات التي تؤثر فيها مثل الوزن (Weight) كغم، العمر (Age) سنوات، الطول (Height) سم ومحيط الرقبة (Neck) سم ومحيط الصدر (Chest) سم.

و لمطلوب:

– حساب معادلة الانحدار.

د - اجراء الاختبارات المناسبة لتحديد جودة النموذج.

د - ما نسبة التفسير للتباين المتغير المعتمد من قبل المتغيرات المستقلة.

د - قدر نسبة الدهون إذا علمت بان قيم المتغيرات المستقلة على التوالي تساوي: [illegible]8، [illegible]6، [illegible]89 ، [illegible]8.6، 15 [illegible].

ID	Fat	Age	Weight	Height	Neck	Chest	ID	Fat	Age	weight	Height	Neck	Chest
1	12.3	23	154.25	67.75	36.2	93.1	**26**	3.7	27	159.25	71.5	35.7	89.6
2	6.1	22	173.25	72.25	38.5	93.6	**27**	7.9	34	131.5	67.5	36.2	88.6
3	25.3	22	154	66.25	34	95.8	**28**	22.9	31	148	67.5	38.8	97.4
4	10.4	26	184.75	72.25	37.4	101.8	**29**	3.7	27	133.25	64.75	36.4	93.5
5	28.7	24	184.25	71.25	34.4	97.3	**30**	8.8	29	160.75	69	36.7	97.4
6	20.9	24	210.25	74.75	39	104.5	**31**	11.9	32	182	73.75	38.7	100.5
7	19.2	26	181	69.75	36.4	105.1	**32**	5.7	29	160.25	71.25	37.3	93.5
8	12.4	25	176	72.5	37.8	99.6	**33**	11.8	27	168	71.25	38.1	93
9	4.1	25	191	74	38.1	100.9	**34**	21.3	41	218.5	71	39.8	111.7
10	11.7	23	198.25	73.5	42.1	99.6	**35**	32.3	41	247.25	73.5	42.1	117
11	7.1	26	186.25	74.5	38.5	101.5	**36**	40.1	49	191.75	65	38.4	118.5
12	7.8	27	216	76	39.4	103.6	**37**	24.2	40	202.25	70	38.5	106.5
13	20.8	32	180.5	69.5	38.4	102	**38**	28.4	50	196.75	68.25	42.1	105.6
14	21.2	30	205.25	71.25	39.4	104.1	**39**	35.2	46	363.15	72.25	51.2	136.2
15	22.1	35	187.75	69.5	40.5	101.3	**40**	32.6	50	203	67	40.2	114.8
16	20.9	35	162.75	66	36.4	99.1	**41**	34.5	45	262.75	68.75	43.2	128.3
17	29	34	195.75	71	38.9	101.9	**42**	32.9	44	205	29.5	36.6	106
18	22.9	32	209.25	71	42.1	107.6	**43**	31.6	48	217	70	37.3	113.3
19	16	28	183.75	67.75	38	106.8	**44**	32	41	212	71.5	41.5	106.6
20	16.5	33	211.75	73.5	40	106.2	**45**	7.7	39	125.25	68	31.5	85.1
21	19.1	28	179	68	39.1	103.3	**46**	13.9	43	164.25	73.25	35.7	96.6
22	15.2	28	200.5	69.75	41.3	111.4	**47**	10.8	40	133.5	67.5	33.6	88.2
23	15.6	31	140.25	68.25	33.9	86	**48**	5.6	39	148.5	71.25	34.6	89.8
24	17.7	32	148.75	70	35.5	86.7	**49**	13.6	45	135.75	68.5	32.8	92.3
25	14	28	151.25	67.75	34.5	90.2	**50**	4	47	127.5	66.75	34	83.4

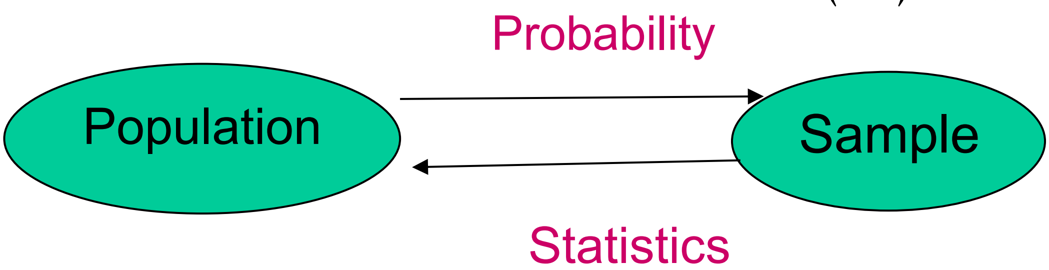
الفصل التاسع

مفاهيم اساسية في الاحتمالات

Principal Concepts of Probability

9-1 تمهيد

نحتاج في حياتنا اليومية الى اتخاذ قرار في كثير من الحالات التي لا نكون متأكدين من النتائج لهذه الحالات، تعد الاحتمالات الطريقة الكمية لتحويل عدم التأكد الى ارقام تعتمد للتعبير عن النتائج المحتملة لاي ظاهرة غير مؤكدة الحدوث. وكما تمت الدراسة في الفصول السابقة بان الاحصائيين يعتمدون على عنصر مهم لتفادي التحيز في جمع البيانات هذا العنصر يسمى العشوائية. والعشوائية تعني تخصيص عنصر معين لمعالجة معينة عشوائيا او اختيار شخص للمعاينة عشوائيا. وتستعمل العشوائية كذلك لمتغير الاستجابة عندما تكون النتائج الممكنة معروفة، ولكن الباحثين غير متأكدين من أي عنصر من النتائج سوف يتم اختياره. لا يوجد شيء في الظواهر الحياتية اكيد، بل كل شيء نعمله نخمن فيه فرصة نجاح العمل بدءا من إدارة الاعمال والطب والمناخ وغيرها من الظواهر الحياتية المختلفة. والاحتمال يمثل وصف كمي لفرصة النجاح لاي عمل، ويمثل حلقة وصل بين الإحصاء الوصفي والاحصاء الاستدلالي كما موضح بالشكل (9-1).



الشكل رقم (9-1) العلاقة بين الاحصاء

ان نظرية الاحتمالات تلعب دورا مهماً في نظريات وتطبيقات علم الإحصاء، ونظرية الاحتمالات تعنى بدراسة التجارب العشوائية. والاحتمال هو فرصة حدوث حدث معين مثل انتشار مرض الانفلونزا في مدينة ما وفرصة نجاح عملية جراحية لمريض محدد. وهناك مفاهيم أساسية في مجال الاحتمالات يجب التعرف عليها كي نفهم كيف نعبر عن الاحتمالات كما في المباحث الآتية.

9-2 المفاهيم الأساسية في الاحتمالات Basic Concepts of Probability

هناك مفاهيم أساسية يجب التعرف عليها كي نستطيع توضيح كيفية استعمال الاحتمالات لمختلف الظواهر الحياتية ومنها الظواهر الطبية واهم هذه المفاهيم ما تضمنته المباحث الفرعية الاتية.

9-2-1 التجربة العشوائية Random Experiment

تعرف التجربة العشوائية على انها كل عملية نعرف مسبقاً بجميع النتائج التي قد تترتب عليها وبعدد هذه النتائج، والتي ستكون على الأقل نتيجتان وكل نتيجة منهما محتملة الوقوع إلا انه لا يعلم أيا من هذه النتائج سوف يتحقق في محاولة معينة. مثال ذلك عند رمي قطعة نقود سليمة ومتزنة مرة واحدة تعد هذه تجربة عشوائية وذلك لأننا نعلم مسبقاً أن هناك نتيجتين لرمي القطعة وهي أما صورة أو كتابة لكننا لا نعلم أي النتيجتين سوف تظهر. وتجربة تسجيل اعمار طلبة الصف الذين تقل أعمارهم عن 22 سنة، تجربة رمي حجر النرد وتسجيل رقم الوجه الذي يظهر وغيرها من التجارب العشوائية. والنتائج التي تظهر من التجربة العشوائية تسمى **فضاء العينة أو المجموعة الشاملة** **(Sample Space)** وهو عبارة عن مجموعة النتائج التي يمكن أن تحدث عند إجراء التجربة العشوائية ونرمز له بالحرف (S). فعلى سبيل المثال عند رمي قطعة نقود متزنة مرة واحدة فإن المجموعة الشاملة هي الصورة أو الكتابة ويمكن كتابة الحالات الكلية بالشكل الآتي:

$$S=\{H, T\}$$

إذ نرمز للكتابة بالرمز (T) ونرمز للصورة بالرمز (H) ونرمز لفضاء العينة او المجموعة الشاملة بالرمز (S). اما عند رمي حجر نرد فإن الحالات الكلية او المجموعة الشاملة هي ظهور الرقم 1 أو 2 أو 3 أو 4 أو 5 أو 6 ومن الممكن كتابة المجموعة الشاملة بالشكل التالي:

$$S = \{1,2,3,4,5,6\}$$

9-2-2 الحدث Event

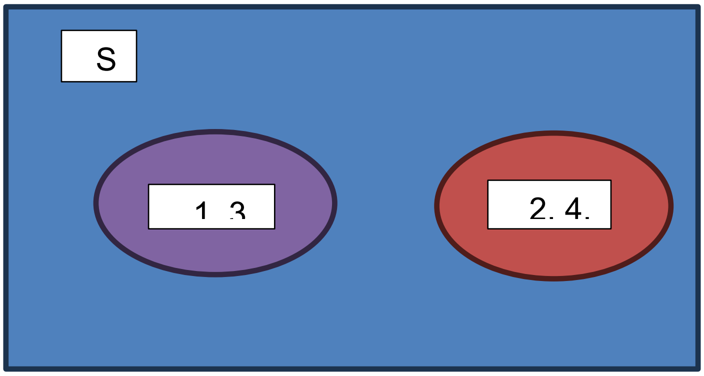
يعرف الحدث على انه مجموعة جزئية من المجموعة الشاملة. ويمكن أن يكون الحدث بسيط إذا كان يتكون من عنصر واحد أو يكون حدث مركب وذلك إذا كان يتكون من أكثر من عنصر. فعلى سبيل المثال إذا كان الحدث (A) يمثل ظهور كتابة عند رمي قطعة نقود مرة واحدة فإن الحدث يمكن كتابته بالشكل الاتي ويسمى حدث بسيط لأنه يتكون من عنصر واحد.

$$A= \{T\}$$

اما إذا كان الحدث (B) الذي يمثل ظهور رقم زوجي عند رمي حجر النرد مرة واحدة فإن الحدث يمكن كتابته بالشكل الاتي ويسمى حدث مركب لأنه يتكون من أكثر من عنصر.

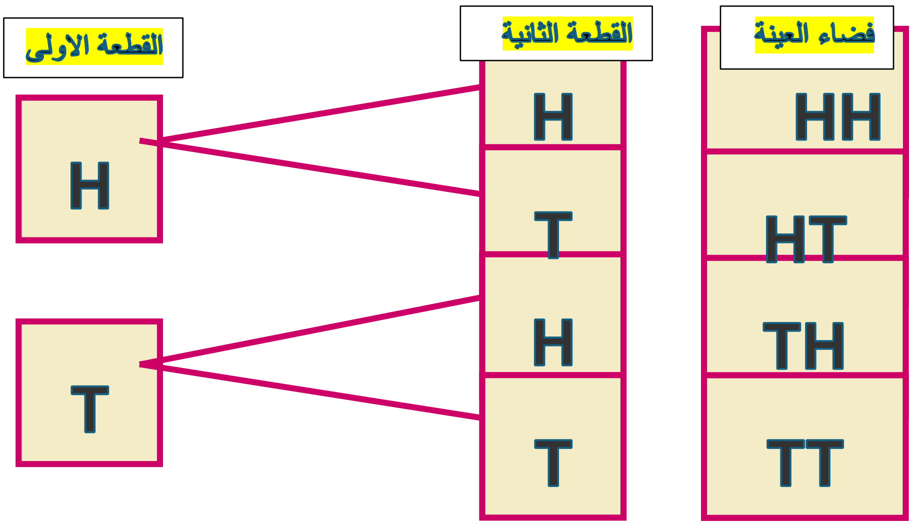
$$B = \{2, 4, 6\}$$

والاحداث عادة تكون على عدة أنواع فالنوع الأول هو **الاحداث المتنافية** (Mutually exclusive events) ويقال عن الحدثين، (A, B) انهما متنافيان (أي ان أحدهما يبعد الاخر) إذا استبعد حدوثهما معا، فعلى سبيل المثال عند رمي قطعة نقود مرة واحدة فمن المستحيل الحصول على صورة وكتابة في الوقت نفسه، وعند اختيار طالب من الصف من المستحيل ان يكون ذكر وانثى بنفس الوقت وعند رمي حجر النرد مرة واحدة يكون من المستحيل الحصول على رقم زوجي وفردي بنفس الوقت. ويمكن تمثيل الاحداث المتنافية كما في تجربة رمي حجر النرد مرة واحدة باستعمال اشكال فن (Ven Diagrams) كما في الشكل (9-2):



الشكل (9−2) الاحداث المتنافية فى تجربة رمى

والنوع الثاني هو **الاحداث المستقلة** **Independent Events** وهي الاحداث التي إذا وقع أحدها لا يمنع او يؤثر على وقوع الاحداث الأخرى، فعلى سبيل المثال عند رمي قطعتي نقود فالحصول على الصورة من القطعة الأولى لا يؤثر في نتيجة القطعة الثانية كما في الشكل (9−3)



الشكل (9−3) يمثل الاحداث المستقلة

والنوع الثالث هو **الاحداث غير المستقلة Non independent events** وهي الاحداث التي إذا وقع أحدها يؤثر في وقوع الأحداث الأخرى. فعلى سبيل المثال في حالة إذا كان لدينا صندوق يحتوي على كرات بيضاء وسوداء فعند سحب كرتين على التوالي بحيث لا تعاد الكرة الأولى فان نتيجة السحبة الثانية تتأثر بنتيجة السحبة الاولى لذا فالحدثان غير مستقلين .

وبشكل عام فان الحدث يمثل **بعدد مرات النجاح Number of success** في التجربة العشوائية وهي عدد الحالات التي يتحقق فيها الحدث أو عدد العناصر التي يتكون منها الحدث أو عدد مرات النجاح في الحصول على الحدث عند اجراء التجربة العشوائية. فعلى سبيل المثال إذا كان الحدث يمثل ظهور كتابة عند رمي قطعة نقود مرة واحدة فإن عدد حالات النجاح تساوي واحداً. اما إذا كان الحدث يمثل ظهور رقم زوجي عند رمي حجر النرد فإن عدد حالات النجاح تساوي ثلاث حالات. والمثال الاتي يوضح بعض هذه المفاهيم.

مثال 1

قام أحد المختبرات بإجراء تجربة عشوائية لمقارنة العلاج العشبي وعلاج بديل لمعالجة الانفلونزا. وتوفر لدى المختبر أربعة متطوعين اثنان من الرجال (احمد ومحمد) واثنان من النساء (فاطمة وزينب). وكانت متغيرات الاستجابة هي شدة المرض ومدته. اختير اثنان من المتطوعين لأخذ العلاج العشبي والأخران لأخذ العلاج البديل. والمطلوب:

1- تحديد فضاء العينة للتجربة
2- عدد نقاط الحدث (A) الذي يمثل رجل وامرأة في كل محاولة.
3- عدد النقاط في الحدث (B) الذي يمثل اختيار رجلين.
4- عدد النقاط في الحدث (C) الذي يمثل اختيار امرأتين.

الحل:

1- فضاء العينة = {(احمد، محمد)، (احمد، فاطمة)، (احمد، زينب)، (محمد، فاطمة)، (محمد، وزينب)، (فاطمة، زينب)}. وعليه فان عدد نقاط فضاء العينة هو 6.

2- عدد نقاط الحدث (A) = {(احمد، فاطمة)، (احمد، زينب)، (محمد، فاطمة)، (محمد، زينب)}.

3- عدد نقاط الحدث (B) = {(احمد، محمد)}.

4- عدد نقاط الحدث (C) = {(فاطمة، زينب)}.

وبعد ان تعرفنا على هذه المفاهيم يصبح من المهم تعريف الاحتمال والذي يكون كما في المبحث الآتي.

3-9 مفهوم الاحتمال Probability Definition

يعرف الاحتمال بطريقتين أولهما **التعريف التقليدي للاحتمال** ويعتمد هذا التعريف على تعريف المحاولة، وتعرف المحاولة بأنها التجربة التي تجرى ويتوقع أن ينتج عنها ناتج، ومثال على ذلك رمي حجر النرد، ورمي قطعة نقود، وسحبة اليانصيب، والنجاح او الفشل في الامتحان، وغيرها من المحاولات والمحاولات نوعان النوع الأول محاولات ينتج عنها نتيجتين فقط لا ثالث لهما وتسمى محاولات برنولي، مثل رمي قطعة النقود، ومن شروط محاولات برنولي ان تكون النتيجة واضحة ومستقلة وتنفي الحدث الاخر، وان يكون هناك احتمال بحصول الحدث. والنوع الثاني محاولات تكون نتائجها أكثر من اثنين. وبشكل عام فان المبدأ في حساب الاحتمال حسب هذا المفهوم هو مبدأ تساوي الفرص وهو أن احتمال أي حدث يساوي خارج قسمة عدد حالات النجاح لهذا الحدث على عدد الحالات الكلية وذلك بشرط تماثل جميع الحالات الكلية. لنفرض ان حدثا معينا(A) يمكن أن يتحقق او يحدث او ينجح في (n) من الحالات (وتسمى حالات النجاح (Success cases) من مجموعة (N) من الحالات الممكنة (Sample space) وعلى فرض ان جميع الحالات الممكنة هذه متساوية الحدوث)

(Equally likely cases فان احتمال ظهور الحدث (A) ويرمز له ((Pr(A)) :
احتمال الحدث = عدد حالات النجاح لوقوع الحدث (n)/ عدد الحالات الكلية للتجربة (N)

$$\boldsymbol{Pr(A) = n\,/N} \qquad (1-9)$$

بشرط تماثل الحالات الكلية ويقصد بتماثل الحالات الكلية هو أن يكون لكل عنصر من عناصر الحالات الكلية الفرصة المتساوية نفسها في النجاح. لذلك يقال إن قطعة النقود سليمة ومتوازنة بمعني أن فرصة ظهور الصورة تساوي فرصة ظهور الكتابة. وكذلك يقال إن حجر النرد سليم ومتوازن بمعنى أن فرصة ظهور كل وجه من الأوجه الستة للحجر متساوية.

مثال2

أحسب الاحتمالات الاتية حسب المعلومات في المثال السابق الذكر (1):
1- الحدث (A) الذي يمثل رجل وامرأة في كل محاولة.
2- عدد النقاط في الحدث (B) الذي يمثل اختيار رجلين.

الحل

عدد الحالات الكلية (فضاء العينة) لهذه التجربة (N=6) وهي اختيار شخصين.
1-عدد حالات النجاح للحدث (A) هو (n=4) لذا فان احتمال رجل وامرأة حسب المعادلة (9-1) هو:

$$Pr(A) = n/N = 4/6$$

2-عدد حالات النجاح للحدث (B) هو (n=1) لذا فان احتمال اختيار رجلين حسب المعادلة (9-1) هو:

$$\Pr(B) = \frac{n}{N} = \frac{1}{6}$$

مثال 3

إذا رمينا حجر نرد سليماً ومتوازناً مرة واحدة أحسب الاحتمالات التالية:

1-الحصول على الحدث (A) يمثل الرقم 4

2-الحصول على الحدث (B) يمثل رقم فردي

3-الحصول على الحدث (C) يمثل رقم أصغر من 3

4-الحصول على الحدث (D) يمثل رقم أكبر من صفر

5-الحصول على الحدث (E) يمثل الرقم 7

الحل

عدد الحالات الكلية فضاء العينة في هذه التجربة (N=6) وهي S= {1,2,3,4,5,6}

1- عدد حالات النجاح للحصول على الحدث (A={4}) والذي يمثل الرقم 4 تساوي (n=1) وعليه فإن الاحتمال هو:

$$Pr(A) = n/N = 1/6$$

2- عدد حالات النجاح للحصول على الحدث (B) والذي يمثل رقم فردي تساوي (n=3) وهي (B={1,3,5}) وعلية فأن الاحتمال هو:

$$Pr(B) = \frac{n}{N} = 3/6$$

3- عدد الحالات النجاح للحصول على الحدث (C) والذي يمثل رقم أصغر من 3 يساوي (n=2) وهي C = {1,2}: وعلية فإن الاحتمال هو:

$$Pr(C) = \frac{n}{N} = 2/6$$

4- عدد حالات النجاح للحصول على الحدث (D) والذي يمثل رقم أكبر من الصفر يساوي (n=6) وهي D= {1,2,3,4,5,6} : وعلية فإن الاحتمال هو:

$$Pr(D) = \frac{n}{N} = 6/6 = 1$$

ويلاحظ هنا أن الاحتمال يساوي الواحد الصحيح ويعد هذا حدثا مؤكداً لأنه من

المؤكد أن أي رقم نحصل علية عند رمي حجر نرد يكون أكبر من الصفر . ونستنتج من ذلك أن أكبر قيمة للاحتمال هي الواحد الصحيح وذلك عندما يكون الحدث مؤكداً.

5- وعدد حالات النجاح للحصول على الحدث (E) والذي يمثل الرقم (7) يساوي (n=0) لان الرقم 7 ليس من ضمن نتائج التجربة وعلية من كتابة الحدث =F {ϕ}وتسمي المجموعة الخالية وعلية يكون الاحتمال:

$$Pr(E)=\frac{n}{N}= 0/6 =0$$

يلاحظ أن التعريف السابق للاحتمال يشترط أن تكون الحالات الممكنة متماثلة (أي متساوية الاحتمال) الأمر الذي قد لا يتحقق دائماً. فعلى سبيل المثال قد تكون قطعة النقود متآكلة (أي غير متوازنة) بمعني أن فرصة ظهور الصورة لا تساوي فرصة ظهور الكتابة في هذه الحالة لا يمكن أن تساوي فرصة ظهور الصورة (0.5). ومثال اخر عند دراسة صفة التدخين في المجتمع ومن ثم يمكن تقسيم المجتمع إلى مدخنين وغير مدخنين ليس بالضرورة أن احتمال التدخين يساوي احتمال عدم التدخين ويساوي (0.5) وذلك لان الحالات الكلية غير متماثلة لأنه في الغالب عدد المدخنين لا يساوي عدد غير المدخنين. لذلك طور الإحصاء مفهوم الاحتمال إلى شكل (الاحتمال التجريبي) الذي يساوي التكرار النسبي لوقوع حدث ما عندما يكون عدد مرات أجراء التجربة كبيراً جدا. ولتوضيح هذا المفهوم نورد المثال الاتي.

مثال4

اختير 1000 شخص من مجتمع معين وفي كل مرة وسجل فيما إذا كان الشخص مصاب بالأنفلونزا خلال الشهر فكانت النتيجة ان (A=529) شخص قد أصيب خلال الشهر، فان احتمال الحدث (A) حسب مفهوم التكرار النسبي للإصابة بمرض الانفلونزا هو:

$$Pr(A) = 529/1000 = 0.529$$

فإذا إعيدت التجربة مرة اخرى وسجلنا ظهور الحدث (B) والذي يمثل عدد الأشخاص الذين لم يصابوا بمرض معين خلال الشهر فكان 493 مرة فان احتمال الحدث حسب مفهوم التكرار النسبي لعدد الأشخاص الذين لم يصابوا بمرض معين هو:

$$Pr(B) = 493/1000 = 0.493$$

وطبقا لتعريف الاحتمال التجريبي فإنه باستمرار هذه التجربة عدد كبير من المرات فان الاحتمال سيقترب أكثر فأكثر من رقم ندعوه احتمال ظهور الصورة في كل رمية لقطعة نقود سليمة وبتنفيذ ذلك سيكون هذا الاحتمال هو 0.5 ويمكن تعريف الاحتمال النسبي بأنه القيمة التي يستقر عندها التكرار النسبي لوقوع الحدث عندما تزيد عدد مرات إجراء التجربة بدرجة كافية لتحقق ذلك الاستقرار للتكرار النسبي. ولتوضيح هذا المفهوم أكثر نورد المثال الاتي.

مثال 5

الجدول التالي (9−1) يمثل توزيع عدد من الطلبة حسب النوع والعمر كالاتي:

جدول (9−1) توزيع عدد من الطلبة حسب النوع والعمر

النوع / فئات العمر	18−20	20−22	22−24	**المجموع (Total)**
الذكور (Male)	30	20	15	**65**
الاناث (Female)	10	15	10	**35**
المجموع (Total)	**40**	**35**	**25**	**100**

اختيرطالب بطريقة عشوائية أحسب الاحتمالات الآتية:

1− الحدث (A) أن يكون من الذكور

2− الحدث (B) أن يكون من فئة العمر 20−22

3− الحدث (C) أن يكون من فئة العمرية الأولى أو الفئة الثالثة.

الحل

1-أن يكون من الذكور فان احتمال الحدث (A) هو:

احتمال الحدث (A) = عدد الطلبة الذكور/ عدد الطلبة الكلي

$$Pr(A) = number\ of\ male\ students\ / total\ number\ of\ students\ = 65/100\ = 0.65$$

2- احتمال الحدث (B) أن يكون الطالب من الفئة العمرية الثانية هو:

احتمال الحدث (B) = عدد الطلبة في الفئة العمرية الثانية/ العدد الكلي للطلبة

$$Pr(B) = number\ of\ studentsin\ the\ scond\ intrval\ of\ Age\ / total\ number\ of\ students\ = 35/100\ = 0.35$$

3- احتمال الحدث (C) أن يكون الطالب من الفئة العمرية الأولى أو الثالثة هو:

احتمال الحدث (C)= (عدد الطلبة في الفئة الأولى + العدد في الفئة الثالثة) /عدد الطلبة الكلي

$$Pr(C) = (40 + 25)\ /\ 100\ = 65/100\ = 0.65$$

وهناك مفهوم ثالث للاحتمالات وهو التقدير الشخصي (Subjective Estimation) ويعتمد هذا المفهوم على الخبرة والمعلومات المتوافرة لدى الشخص عن الحدث او الظاهرة المطلوب حساب الاحتمال لها عندما لا يمكن اجراء التجربة او تحديد المحاولات لهذه التجربة وعندما يكون من غير الممكن معرفة النتائج مسبقا عن النتائج المتوقعة للظاهرة، ومن ثم نلجأ الى التقدير الشخصي للاحتمال والذي يعتمد على الاعتقاد الشخصي بالظاهرة فعلى سبيل المثال إذا اردت ان تشتري وثيقة تامين على سيارتك كم الاحتمال لأنك سوف تستعمل هذا التامين. ومثال اخر عن مدى اعتقاد علماء الفضاء بإمكانية هبوط مركبة الفضاء على القمر، وغيرها من الأمثلة التي يصعب اجراء تجربة لمعرفة النتائج على المدى الطويل وعدم توافر معلومات سابقة عنها لذلك نلجأ الى الاعتقاد الشخصي لتقدير الاحتمال والذي يعتمد على الخبرة السابقة والمعلومات المتوفرة لدى من يقوم بالتقدير.

4-9 قوانين الاحتمالات Probability Rules

ذكرنا في المباحث السابقة بان حساب الاحتمالات يعتمد على معرفة عدد حالات النتائج الممكنة للتجربة (فضاء العينة) وعدد حالات النجاح لكل حدث، لذلك لمعرفة عدد النتائج في التجربة وعدد حالات النجاح لكل حدث هناك العديد من القواعد والقوانين الرياضية التي تسهل وتساعد على حساب هذا العدد وخصوصا في التجارب الكبيرة التي يصعب رسم فضاء العينة او المجموعة الشاملة لها لأنها تتضمن عناصر كثيرة فعلى سبيل المثال عند رمي حجر نرد ثلاث مرات متتالية او رمي ثلاث حجرات نرد مرة واحدة فان عدد النتائج الممكنة في مثل هذه التجربة يساوي 216 عنصر او اذا رغبنا في ترتيب مجموعة من الأشياء وحساب عدد الترتيبات الممكنة وغيرها من الحالات ومن ثم نحتاج الى قاعدة لحساب هذا العدد تختلف حسب التجربة والطريقة التي يتم بها اجراء كل تجربة لذلك توجد قواعد او قوانين عدة لحساب الحالات الممكنة للتجربة او عدد عناصر فضاء العينة كما في المباحث الفرعية الآتية.

1-4-9 قانون العد Counting Rule

إذا كانت أي تجربة تجرى بمرحلتين وكان عدد الطرق التي تجرى بها التجربة في المرحلة الأولى هو (m) وكان عدد الطرق التي تجرى بها التجربة في المرحلة الثانية هو (n) فان عدد النتائج الممكنة او العناصر التي تمثل فضاء العينة للتجربة سيكون حاصل ضرب الرقمين لعدد الطرق للمرحلتين:

$$n(S) = m \times n \qquad (2\text{-}9)$$

وهكذا يمكن تعميم القانون الى (k) من التجارب والتي تجرى بالتتالي وبعدد طرق $(n_1, n_2, n_3, \dots, n_k)$ فان عدد العناصر في فضاء العينة (S) يكون كالاتي:

$$n(S) = n_1 \times n_2 \times n_3 \times \dots \times n_k \qquad (3\text{-}9)$$

<u>مثال 6</u>

قدم أحد الأساتذة ثلاثة أسئلة من الاختبارات المتعددة ذات الاجابتين (صح او خطأ) فما هو عدد النتائج الممكنة في فضاء العينة؟

<u>الحل:</u> بما ان كل سؤال له اجابتان فان عدد النتائج الممكنة حسب القاعدة (9−2) تكون كالاتي:

$$n(S) = 2 \times 2 \times 2 = 8$$

ويمكن رسم شجرة الاحتمالات لتمثيل النتائج الممكنة للتجربة والتي تمثل فضاء العينة كما في الشكل (9−3). وغالبا ما تستعمل شجرة الاحتمالات مع التجارب البسيطة وذات المراحل الصغيرة لأنه إذا زاد عدد المراحل يصبح من الصعب رسم هذه الشجرة فيستعاض عنها بقانون العد كما في المعادلة (9−2).

<u>مثال 7</u>

كرتان سحبا من صندوق يحتوي على كرتين حمراء وكرتين سوداء فما هو عدد النتائج الممكنة في فضاء العينة بغض النظر عن اللون؟

<u>الحل:</u>

1− إذا كان السحب بدون ارجاع فان سحب الكرة الأولى لدينا 4 طرق ولسحب الكرة الثانية لدينا ثلاث طرق وبالتالي فان عدد العناصر في فضاء العينة يكون حسب القاعدة (10−2) كالاتي:

$$n(S) = 4 \times 3 = 12$$

2− اما إذا كان السحب مع الارجاع فان عدد الطرق لسحب الكرة الأولى يساوي 4 وكذلك الكرة الثانية لأنه يتم ارجاع الكرة المسحوبة وبالتالي فان عدد النتائج الممكنة في فضاء العينة هي:

$$n(S) = 4 \times 4 = 16$$

2-4-9 قانون التوافيق Combinations Rule

عندما تكون التجربة اختيار مجموعة من الأشياء بغض النظر عن الترتيب. فعلى سبيل المثال اختيار مجموعة (r) من الأشياء من بين (n) من النتائج يسمى التوافيق ويرمز له:

$$\binom{n}{r} = C_r^n \qquad 0 \le r \le n$$

وبشكل عام فان عدد التوافيق لعدد من الاشياء المختلفة (n) يتم اختيار (r) منها في كل مرة بغض النظر عن الترتيب يحسب كالاتي:

$$\binom{n}{r} = C_r^n = \frac{n!}{r!(n-r)!}; \quad 0 \le r \le n \qquad (4-9)$$

حيث ان:

$$n! = \begin{cases} n(n-1)(n-2)\ldots \times 3 \times 2 \times 1 & where\ n \neq 0 \\ 1 & where\ n = 0 \end{cases}$$

وان (n) عدد موجب وكذلك فان:

$$\binom{\boldsymbol{n}}{\boldsymbol{r}} = \binom{\boldsymbol{n}}{\boldsymbol{n-r}} = \frac{\boldsymbol{n}!}{\boldsymbol{r}!(\boldsymbol{n-r})!}$$

مثال 8

بكم طريقة يمكن لطبيب اختيار 3 مرضى بغض النظر عن ترتيبهم من مجموعة تتكون من 4 مرضى مختلفين لغرض اجراء عملية جراحية؟

الحل: نرمز للمرضى بالحروف الاتية (A, B, C, D)

اذن الترتيبات او الطرق الممكنة هي (ABC, ABD, ACD, BCD)

وهذا يعني بان لدى الطبيب اربعة طرق يمكن ان يختار بها ثلاثة مرضى من بين اربعة بغض النظر عن الترتيب ويمكن كتابة ذلك بلغة التوافيق كالاتي:

$\binom{4}{3}=C_3^4 \quad =4$

مثال 9

كم لجنة طبية يمكن لمدير المستشفى تكوينها من 3 اطباء من بين 8 اطباء؟ او بكم طريقة يمكن اختيار هذه اللجنة؟

الحل: حسب قانون التوافيق لان الترتيب غير مهم وحسب الصيغة (9–4) يكون عدد الطرق:

$$\binom{8}{3}=\frac{8!}{3!\,(8-3)!}=\frac{8\times7\times6\times5!}{3\times2\times1\times5!}=56$$

مثال 10

بكم طريقة يمكن لطالب الاجابة عن 8 اسئلة في الامتحان الذي يتكون من 10 اسئلة وتحت الشروط الاتية:

1– بكم طريقة يمكن الاجابة بدون شرط؟

2– بكم طريقة يمكن الاجابة إذا كانت الاجابة اجبارية عن الاسئلة الثلاثة الاولى؟

3– بكم طريقة يمكن الاجابة إذا كان يجب عليه الاجابة على الاقل 4 اسئلة من الاسئلة الخمسة الاولى؟

الحل:

1) طالما لا يوجد شرط اذن (r=8) و (n=10) وعليه فان عدد الطرق الممكنة هو:

$$\binom{10}{8} = \frac{10!}{8!\,(10-8)!} = \frac{10\times 9\times 8!}{2\times 1\times 8!} = 45$$

2) الاجابة اجبارية عن 3 اسئلة الاولى اذن يتبقى لديه ان يختار 5 اسئلة من بين 7 اسئلة اي ان (r=5) و (n=7) وكالاتي:

$$\binom{7}{5} = \frac{7!}{5!(7-5)!} = \frac{7\times6\times5!}{2\times1\times5!} = 21$$

3) لديه خياران اما الاجابة عن الاسئلة الخمسة الاولى اجباري وعليه يتبقى لديه اختيار 3 اسئلة من الاسئلة 5 الباقية اي ان عدد الطرق هو:

$$\binom{5}{3} = \frac{5!}{3!\,(5-3)!} = 10$$

او الخيار الثاني وهو الاجابة عن 4 اسئلة من الخمسة الاولى وبالتالي يتبقى لديه اختيار 4 اسئلة من 5 الباقية وعليه فان عدد الطرق في الحالتين يكون كالاتي:

$$\binom{5}{4}\times\binom{5}{4}+\binom{5}{5}\times\binom{5}{3} = 25+10 = 35$$

3-4-9 قانون التباديل Permutations Rule

تكون التجربة اختيار او ترتيب مجموعة من الأشياء مع الاخذ بنظر الاعتبار الترتيب. او هي عدد الطرق التي يمكن بها اختيار (r) من الأشياء او العناصر من بين (n) من العناصر في كل مرة مع الاخذ بنظر الاعتبار الترتيب للعناصر. فاذا أردنا معرفة عدد الطرق الممكنة لترتيب مجموعة من الأشياء (n) مرة واحدة تكون كالاتي:

$$P_n^n = n! \qquad (5-9)$$

وكما موضحة في الأمثلة الاتية اما الحالة الثانية فهي اختيار جزء (r) من المجموعة (n) فيكون حسب القاعدة الاتية:

$$P_r^n = \frac{n!}{(n-r)!} \qquad (6\text{–}9)$$

مثال 11

بكم طريقة يمكن ترتيب ثلاثة اطباء (A,B,C) في وقت واحد في جدول الخفارة يتكون من ثلاث عيادات وبدون تكرار او مع التكرار؟

الحل: في حالة عدم التكرار فانه تكون لدينا ثلاثة خيارات في اليوم الاول واثنان في اليوم الثاني وخيار واحد في اليوم الثالث اما في حالة التكرار فمسموح به فان يكون لدينا في كل يوم ثلاثة خيارات كما في الجدول الاتي:

جدول (9–2) عدد الطرق الممكنة لترتيب ثلاث اطباء

الايام	1	2	3
بدون تكرار	3	2	1
مع التكرار	3	3	3

وعليه فان عدد الطرق في حالة عدم التكرار = $3 \times 2 \times 1 = 6$

اما في حالة التكرار فان عدد الطرق = $3^3 = 27 = 3 \times 3 \times 3$

وبشكل عام إذا كانت هناك (k) من التجارب التجربة الأولى تجرى بعدد (n_1) من الطرق والتجربة الثانية تجرى بعدد (n_2) من الطرق وهكذا الى اخر تجربة فان عدد الطرق الممكنة لإجراء هذه التجارب سوية يساوي:

$$P_{n1,n2,\dots,nk}^{n} = \frac{n!}{n_1! \times n_2! \times \dots \times n_k!} \qquad (7\text{–}9)$$

$$where\ n = n_1 + n_1 + \cdots + n_k$$

مثال 12

كم لوحة مرورية يمكن تكوينها باستخدام حرفين من حروف اللغة الانكليزية متبوعة بثلاثة ارقام مع التكرار؟

الحل:

يوجد في اللغة الانكليزية 26 حرفاً وعدد الارقام 10 وشكل اللوحة يكون كالاتي:

جدول (9–3) عدد النتائج الممكنة لكل مرتبة من اللوحة

26	26	9	10	10

عدد اللوحات = 26×26×9×10×10=60840 لوحة. والرقم الثالث كان 9 لان اللوحة لا يمكن ان تبدأ برقم صفر لذلك يبقى لدينا تسعة ارقام. اما باقي المراتب تقبل ان يكون الرقم الثاني او الثالث يساوي صفراً.

مثال 13

كم تشكيل يمكن تكوينه من حروف الكلمة (MISSISSIPI)؟

الحل:

الكلمة تتكون من ($n_1 = 1$) يمثل الحروف (M)، ($n_2 = 4$) يمثل الحرف (I) ، و($n_3 = 4$) يمثل الحرف (S) ، و ($n_4 = 1$) يمثل الحرف (P) وعدد الحروف الكلي يساوي 10 حروف ويمثل (n) . اذن عدد التشكيلات الممكنة حسب الصيغة (9–6) يساوي:

$$P^{10}_{1,4,4,1} = \frac{10!}{1! \times 4! \times 4! \times 1!} = 6300$$

مثال 14

يرغب طبيب في معرفة عدد الطرق التي يمكن بها ترتيب 5 اشخاص على صف واحد بجنب الاخر؟ وهل يختلف العدد إذا كان الجلوس على شكل دائرة؟

ملاحظة: إذا كان ترتيب (n) من الاشياء على شكل صف مستقيم فان عدد الطرق هو (n!) اما إذا كان الترتيب على شكل دائرة فان عدد الطرق الممكنة هو ((n−1)!)

الحل:

اولا: إذا كان الترتيب على شكل صف يكون عدد الطرق يساوي:

$$n! = 5! = 5 \times 4 \times 3 \times 2 \times 1 = 120 \times 2 = 240$$

ثانيا: إذا كان الترتيب على شكل دائري يكون عدد الطرق يساوي:

$$(n-1)! = (5-1)! = 4! = 4 \times 3 \times 2 \times 1 = 24 \times 2 = 48$$

ملاحظة: تم الضرب في 2 لأنه هناك احتمالين للجلوس (AB or BA) فتحسب مرتان.

بعد ان تعرفنا على القواعد او القوانين التي يمكن باستعمالها يمكن حساب النتائج للتجارب او الاحداث. بعد ذلك نحتاج ان نعرف العلاقات بين الاحداث كما في المبحث الآتي.

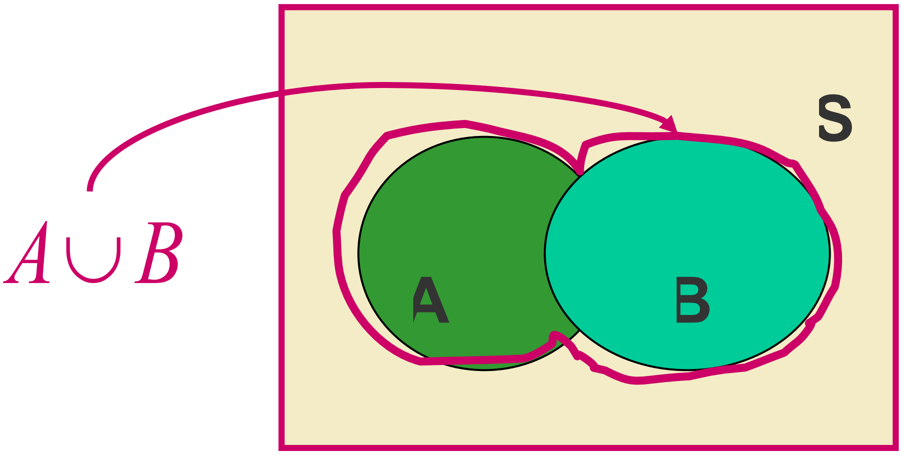
9-5 العلاقات بين الاحداث Events Relationships

هناك علاقات يمكن استعمالها لتركيب او دمج الاحداث البسيطة او قد تكون موجودة في الواقع بين الاحداث للحصول على احداث جديدة أكبر من الاحداث البسيطة باستعمال العمليات المنطقية و/ او. والمباحث الاتية توضح بعض هذه العلاقات

والقوانين التي تستعمل لحساب النتائج الممكنة لهذه الاحداث المركبة ومنها نتمكن من حساب الاحتمالات للأحداث الناتجة.

9-5-1 الاتحاد Union

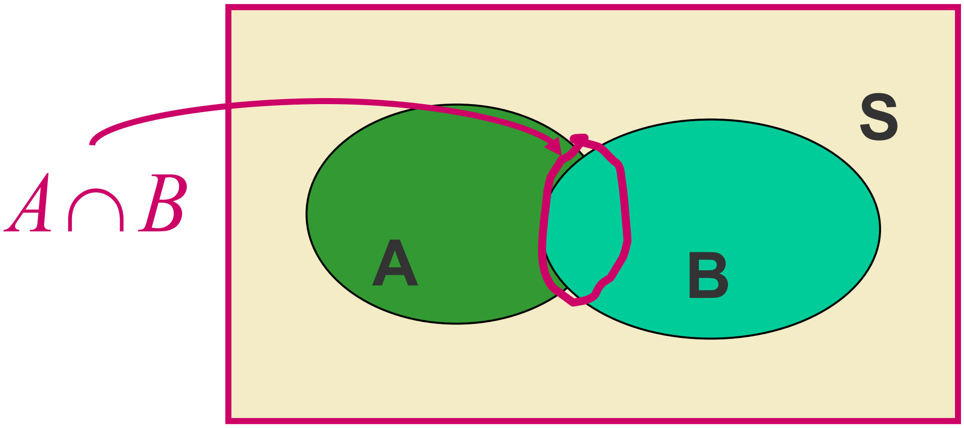
الاتحاد بين الحدثين (A and B) يرمز له بالرمز $(A \cup B)$ يمثل الحدث الذي يتكون من نقاط اما الحدث (A) او الحدث (B) او كلاهما عند اجراء التجربة ويمكن تمثيل ذلك حسب اشكال فن (Veen Diagram) كما في الشكل (9-4)



الشكل (9-4) الاتحاد بين الاحداث

9-5-2 التقاطع intersection

يعرف التقاطع بين الحدثين (A) و (B) على انه مجموعة العناصر الموجودة في الحدثين او المشتركة بين الحدثين عند اجراء التجربة ويرمز له بالرمز $(A \cap B)$ ويمكن تمثيله باستخدام اشكال فن كما في الشكل (9-5) الاتي:

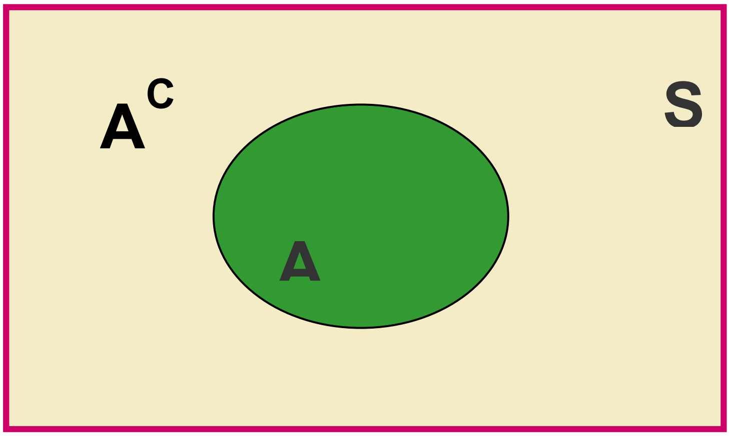


الشكل (9−5) التقاطع بين الحدثين (A) و (B)

وإذا كان الحدثان متنافيين فان التقاطع بينهما يكون عبارة عن المجموعة الخالية ويرمز له بالرمز $(A \cap B = \emptyset)$.

9−5−3 المكملة (المتممة) Complement

تعرف المكملة للحدث (A) على انها العناصر الموجودة جميعها في التجربة ولم تكن ضمن عناصر الحدث (A) ويرمز لها بالرمز (A^c) ويمكن تمثيلها كما في الشكل (9−6)



الشكل (9−6) المكملة للحدث (A)

مثال 15

كان في أحد المستشفيات 120 مريضاً وصنفوا حسب شدة المرض والنوع كما في الجدول الاتي:

جدول (9–4) بيانات المرضى حسب شدة المرض والنوع

النوع / شدة المرض	(متوسطة)	(شديدة)	المجموع
ذكور	20	40	60
اناث	30	30	60
المجموع	50	70	120

حدد عدد العناصر لكل من الاحداث الاتية:

1– الحدث (A) يمثل المرضى وحالتهم متوسطة.
2– الحدث (B) يمثل المرضى الذكور.
3– الحدث (C) يمثل المرضى الاناث.
4– الحدث (D) يمثل المرضى وحالتهم شديدة.
5– التقاطع بين الحدث (A) والحدث (B).
6– الاتحاد بين الحدث (A) والحدث (B).
7– المكملة للحدث (A).

الحل:

1– $n(A) = 50,$

2– $n(B) = 60$

3– $n(C) = 60$

4– $n(D) = 70$

5– $n(A \cap B) = 20$

$6- n(A \cup B) = n(A) + n(B) - n(A \cap B) = 50 + 60 - 20 = 90.$

$7- n(A^c) = 70.$

6-9 حساب الاحتمالات حسب العلاقات بين الاحداث

يمكن تصنيف الاحداث بشكل عام الى صنفين رئيسين هما الاحداث المتنافية والاحداث المستقلة، والاحداث المتنافية يمكن ان تكون متنافية او غير متنافية والمستقلة ممكن تكون مستقلة او غير مستقلة ففي الحالة الأولى وهي تنافي الاحداث يستعمل قانون الجمع لحساب الاحتمالات المتنافية وغير المتنافية ويستعمل قانون الضرب لحساب احتمالات الاحداث المستقلة او غير المستقلة وكما في المباحث الفرعية الاتية:

9-6-1 قانون الجمع في حالة الاحداث المتنافية Additive Rule

تعرف الأحداث المتنافية على انها الأحداث التي إذا حدث أحدها لا تحدث بقية الاحداث الأخرى أي هي الاحداث التي حدوث أحداها ينفي أو يمنع حدوث الاحداث الأخرى فعلى سبيل المثال عند رمي قطعة نقود فإن الحصول على الصورة أو الكتابة حدثان متنافيان لان الصورة والكتابة لا يمكن أن يحدثا معا. ففي حالة الاحداث المتنافية يكون التقاطع بين الاحداث يمثل المجموعة الخالية أي لا توجد عناصر مشتركة بين الاحداث وبالتالي لحساب احتمالات الاحداث المتنافية يستعمل قانون الجمع فاذا كان الحدثين B وA حدثين متنافيين فإن احتمال حدوث الأول أو الثاني (أي احتمال حدوث أحدهما) هو:

$$Pr(A \cup B) = \Pr(A) + \Pr(B) \qquad (8-9)$$

ويرمز ∪ إلى الاتحاد ويعني (أو) ويمكن تعميم القانون السابق لأي عدد من الأحداث المتنافية.

مثال16

في المثال السابق (15) أحسب احتمال أن يكون المريض من الذكور او من الاناث.

الحل

باستعمال قانون الأحداث المتنافية في المعادلة (9–8) لأنه من غير الممكن ان يكون المريض ذكر وانثى بنفس الوقت لذلك تم استعمال كلمة او للدلالة على ان الحدثان متنافيين وباستعمال التعريف التقليدي للاحتمال لحساب الاحتمالات للأحداث ولقد تم تعريف الحدثان في المثال السابق بالرمز(B,C) وكان عدد العناصر في كل حدث (n(B)=60) و(n(C)=60) وعدد نقاط فضاء العينة كان يساوي (n(S)=120) فإن احتمال حدوث الأول أو الثاني هو:

$$\Pr(B \cup C) = 60/120 + 60/120 = \frac{120}{120} = 1$$

وبما ان الاحتمال يساوي واحد فان الحدث يعتبر مؤكد الحدوث.

9–6–2 قانون الجمع في حالة الاحداث غير المتنافية

اما إذا كانت الاحداث غير متنافية وهي عبارة عن الاحداث التي يكون حدوث أحدها لا ينفي او لا يمنع حدوث الاحداث الأخرى، وتسمى الحالة العامة للاتحاد. فعلى سبيل المثال إذا كان الحدثان B وA حدثين غير متنافيين فإن احتمال حدوث الأول أو الثاني (أي احتمال حدوث أحدهما أو كليهما) فإن الاحتمال سيكون:

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B) \qquad (9\text{–}9)$$

ويلاحظ أن هذا القانون يختلف عن قانون الأحداث المتنافية بوجود احتمال التقاطع $P(A \cap B)$ وهو احتمال حدوث الحدثين معا وسبب طرح احتمال التقاطع هو أن الحدثان يمكن أن يحدثا معاً، لذلك فانه قد تم حساب احتمال التقاطع مرتين مرة مع الحدث الأول والمرة الثانية مع الحدث الثاني لذا يجب حذف أو طرح احتمال التقاطع مرة

واحدة. وهذا القانون يعتبر الحالة العامة للجمع لأنه لو كان الحدثان متنافيان فإنه لا يوجد تقاطع بينهما أي أن احتمال التقاطع يساوي الصفر.

<u>مثال17</u>

في المثال السابق (15) أوجد احتمال أن يكون المريض من الذكور أو من ذوي الحالة المتوسطة.

<u>الحل</u>

بما أن الحدثان غير متنافيان لأنه يمكن أن يكون المريض من الذكور ومن ذوي الحالة المتوسطة أي انه توجد عناصر مشتركة (تقاطع بين الحدثين) وكان عدد النقاط للحدث الأول المريض من ذوي الحالة المتوسطة (n(A)=50) وعدد النقاط للحدث الثاني المريض من الذكور (n(B)=60) وعدد النقاط للتقاطع بينهما أي ان المريض من الذكور ومن ذوي الحالة المتوسطة ($n(A \cap B) = 20$) وباستعمال التعريف التجريبي لحساب الاحتمال وحسب قانون الجمع العام في المعادلة (9−8) فان الاحتمال يكون كالاتي.

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$$

$$= \frac{50}{120} + \frac{60}{120} - \frac{20}{120} = \frac{90}{120} = 0.75$$

<u>مثال 18</u>

احسب الاحتمالات الاتية في تجربة اختيار عائلة لديها طفلان تم ترتيبهما حسب الاسبقية في الولادة:

1- الحدث (A) يمثل ولادة الطفل الأول انثى (F).
2- الحدث (B) يمثل ولادة الطفل الثاني انثى (F).
3- الحدث (C) يمثل ولادة الطفل الأول ذكر (M).
4- احتمال الحدث الطفل الأول انثى او الطفل الثاني انثى او الاثنين اناث.

<u>**الحل:**</u>

ان عدد النتائج الممكنة في فضاء العينة للتجربة

$n(S) = 2 \times 2 = 4$ {FF, FM, MF. MM}

اما عدد العناصر في كل حدث فيمكن ايجادها كالاتي:

$$n(A) = 2,\ \ n(B) = \ 2,\ \ n(C) = 2, \qquad n(A \cap B) = 1,$$
$$n(C \cap A) = 0.$$

$1-\ \Pr(A) = 2/4 = 0.5$

$2-\ \Pr(B) = 2/4 = 0.5$

$3-\ \Pr(C) = 2/4 = 0.5$

ومن هنا نجد بان الحدثين (A) و (B) غير متنافيين لوجود عناصر مشتركة بينهما، ولكن الحدثين (C) و(A) يعدان متنافيين لعدم وجود عناصر مشتركة بينهما وكان التقاطع يساوي صفراً وحسب قانون الاحتمال التجريبي فان احتمال حدوث الحدث (A) او الحدث (B) يكون كالاتي:

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - P(A \cap B)$$
$$= \frac{2}{4} + \frac{2}{4} - \frac{1}{4} = \frac{3}{4} = 0.75$$
$$\Pr(C \cup A) = \Pr(C) + \Pr(A) = \frac{2}{4} + \frac{2}{4} = \frac{4}{4} = 1$$

3-6-9 قانون الضرب في حالة الاحداث المستقلة Multiple Rule of Independent Events

تم إيجاد احتمال التقاطع بين الحدثين من الجدول الذي يمثل فضاء العينة، أحيانا يكون من غير الممكن استعمال مثل هذه الجداول لان التقاطع يعتمد على ما يعرف باستقلال الاحداث او عدم استقلالها. وتعرف الأحداث المستقلة على انها الأحداث التي إذا حدث أحدها لا يؤثر في احتمالات حدوث بقية الأحداث الأخرى. فعلى سبيل المثال عند

رمي قطعتي نقود مرة واحدة فإن نتيجة القطعة الأولى مستقل تماماً عن نتيجة القطعة الثانية. فاذا كانت القطعة الأولى صورة أو كتابة ليس لها تأثير على الإطلاق على القطعة الثانية . ومن ثم لحساب الاحتمالات لمثل هذه الاحداث نستعمل قانون الضرب فاذا كان الحدثان (A) و (B) مستقلين فإن احتمال حدوث الأول والثاني (أي حدوثهما معاً) ويرمز له بالتقاطع ($\Pr(A \cap B)$) كالاتي:

$$\Pr(A \cap B) = \Pr(A) \times \Pr(B) \qquad (10\text{–}9)$$

أي أن احتمال وقوع حدثين مستقلين معاً يساوي حاصل ضرب احتمال وقوع أي منهما بمفردة في احتمال وقوع الأخر بمفردة. ويمكن إعمام هذا القانون على عدد من الحوادث المستقلة.

<u>مثال 19</u>

قام أحد الباحثين بتصميم برنامج لمتابعة ارتفاع ضغط الدم الواطئ (Diastolic Blood Pressure (DBP)) للام والأب في أحد العوائل وعلى افتراض انه لا توجد علاقة جينية بين الاثنين. وكان يسجل القراءة عالياً (H) إذا كان (95< DBP) ويسجلها واطئاً (L) إذا كان (DBP<90) للاثنين. وتم تعريف الاحداث الاتية:

أ- الحدث (A) يمثل أن ضغط الدم يكون للام عالياً ويكون وللاب واطئاً.
ب- الحدث (B) يمثل أن يكون كليهما عالي.
<u>المطلوب:</u> احسب الاحتمالات للأحداث أعلاه

<u>الحل</u>

أ- ان يكون ضغط الدم للام عالياً وللاب واطئاً، فأن وجود الواو هنا يعني الضرب لأن الحدثين مستقلان (لأن نتيجة قياس ضغط الدم للام لا تؤثر في نتيجة قياس ضغط الدم للاب) فإن:

$$\Pr(H \cap L) = \Pr(H) \times Pr(L)$$

وذلك على اعتبار أن الحدث (H) يمثل قياس ضغط الدم للام عالياً والحدث (L) يمثل قياس ضغط الدم للاب واطئاً والاحتمال لكل منهما يساوي:

$$\Pr(H) = \frac{1}{2}.\ and\ \Pr(L) = \frac{1}{2}$$

$$\therefore \Pr(\mathrm{A}) = \Pr(\mathrm{H} \cap \mathrm{L}) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

ب-احتمال ان يكون ضغط الدم للاثنين عالياً وهذا يعني بان ضغط الدم للام وللاب يكون عالياً فاذا رمزنا للحدث الأول (H) والحدث الثاني (H) وبما ان الحدثين مستقلان فان الاحتمال هو:

$$\Pr(H) = \frac{1}{2}.\ and\ \Pr(H) = \frac{1}{2}$$

$$\therefore \Pr(\mathrm{B}) = \Pr(\mathrm{H} \cap \mathrm{H}) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

مثال 20

كان في مجتمع معين 20% من الناس يمكن تصنيفهم على انهم من المصابين بأمراض القلب، واختير ثلاثة اشخاص عشوائيا من هذا المجتمع، ما هو احتمال ان واحداً من هؤلاء الأشخاص مصاب بأمراض القلب.

الحل:

نفرض بان H: ترمز للمصابين بأمراض القلب، N: ترمز للأشخاص غير المصابين بامراض القلب.

احتمال ان شخصاً واحداً مصاب = احتمال الشخص الأول او احتمال الشخص الثاني او احتمال الشخص الثالث وبالرموز يمكن تمثيل الحالة كالاتي:

$$\Pr(H \cap N) = \Pr(HNN) + \Pr(NHN) + \Pr(NNH)$$

وبما ان الاحداث مستقله وان احتمال الشخص مصاب بمرض القلب هو ($Pr(H) =$ 0.2) واحتمال الشخص غير مصاب بالمرض هو ($Pr(N) = 0.8$) فان احتمال التقاطع يساوي:

$$\Pr(H \cap N) = (0.2 \times 0.8 \times 0.8) + (0.8 \times 0.2 \times 0.8) + (0.8 \times 0.8 \times 0.2) = 3(0.2)(0.8)^2 = 0.384$$

4-6-9 قانون الضرب للأحداث غير المستقلة (الحالة العامة): Multiple Rule of dependent Events

اما إذا كان الحدثان (A, B) غير مستقلين فإن احتمال حدوث الأول والثاني (أي حدوثهما معاً) وهذا يسمي قانون الضرب في الحالة العامة ويكتب:

$$\Pr(A \cap B) = \Pr(A) \times \Pr(B|A)$$

او

$$\Pr(A \cap B) = \Pr(B) \times \Pr(A|B) \qquad (11-9)$$

وهذا يعني احتمال التقاطع يساوي احتمال حدوث الحدث (A) مضروب في احتمال حدوث الحدث (B) بشرط حدوث الحدث (A) وهذا يسمى بالاحتمال الشرطي ويحسب بالشكل الاتي:

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} \qquad (12-9)$$

أو حدوث الحدث (B) مضروب في حدوث الحدث (A) بشرط حدوث الحدث (B) ويحسب هذا الاحتمال كالاتي:

$$\Pr\big((B|A)\big) = \frac{\Pr(B \cap A)}{\Pr(A)}; \quad A \neq 0 \qquad (13-9)$$

والفرق الوحيد بين الحالة العامة وحالة الحوادث المستقلة هو أن نراعي تأثير الحدث الأول على الثاني.

<u>مثال21</u>

البيانات في الجدول (9–5) توضح استطلاع حول القناعة بالعمل لمجموعة من العاملين في أحد المستشفيات:

جدول (9–5) بيانات الاستطلاع حول قناعة العاملين في المستشفى بالعمل

المستوى / القناعة بالعمل	مقتنع	غير مقتنع	**المجموع**
جامعي	74	43	**117**
ثانوي	224	171	**395**
ابتدائي	126	140	**266**
المجموع	**424**	**354**	**778**

<u>والمطلوب:</u>

1– حساب احتمال الحدث (C) والذي يمثل اختيار عامل من العاملين الجامعيين.

2– حساب احتمال الحدث (S) والذي يمثل اختيار عامل مقتنع بالعمل.

3– حساب احتمال الحدث الذي يمثل اختيار عامل جامعي ومقتنع بالعمل.

4– حساب احتمال الحدث الذي يمثل كون العامل جامعياً وشرط ان يكون مقتنعاً بالعمل.

5– حساب احتمال الحدث الذي يمثل كون العامل مقتنعاً وشرط ان يكون جامعياً.

6– هل ان الحدثين (C, S) مستقلان.

7– احسب احتمال الاتحاد بين الحدثين (C, S).

الحل:

أولا يمكن تحويل الجدول (9–5) من جدول تكرارات الى جدول احتمالات من خلال استعمال التعريف التجريبي للاحتمال وذلك بقسمة كل تكرار على مجموع التكرارات (778) ويكون كما في الجدول (9–6):

جدول (9–6) احتمالات البيانات في الجدول (9–5)

القناعة بالعمل / المستوى	مقتنع	غير مقتنع	**المجموع**
جامعي	0.095	0.055	**0.150**
ثانوي	0.288	0.220	**0.508**
ابتدائي	0.162	0.180	**0.342**
المجموع	**0.545**	**0.455**	**1.00**

1– من الجدول (9–6) نجد بان احتمال الحدث الأول (C)

$$\Pr(C) = \frac{117}{778} = 0.150$$

2– احتمال الحدث (S) كما في الجدول (9–6) يكون:

$$\Pr(S) = 0.545$$

3– احتمال الحدث الذي يمثل اختيار عامل جامعي ومقتنع بالعمل يمثل التقاطع بين الحدثين (C, S) ويساوي كما في الجدول (9–6)

$$\Pr(C \cap S) = 0.095$$

4– احتمال الحدث الذي يمثل كون العامل جامعياً وشرط ان يكون مقتنعاً بالعمل يمثل الاحتمال الشرطي بين الحدثين (C,S) ويمكن حسابه بالشكل الاتي:

$$\Pr(C|S) = \frac{\Pr(C \cap S)}{\Pr(S)} = \frac{0.095}{0.545} = 0.175$$

5– احتمال الحدث الذي يمثل كون العامل مقتنعاً وشرط ان يكون جامعياً يمثل الاحتمال الشرطي بين الحدثين (S,C) ويمكن حساب بالشكل الاتي:

$$Pr(S|C) = \frac{Pr(S \cap C)}{Pr(C)} = \frac{Pr(C \cap S)}{Pr(C)} = \frac{0.095}{0.150} = 0.632$$

6- هل ان الحدثين (C, S) مستقلان. يمكن اثبات ذلك من خلال القاعدة التي تقول أن الاحتمال الشرطي للحدثين يساوي احتمال أحدهما اذا كان الحدثان مستقلين ومن النتائج نجد ان:

$$Pr(C) = 0.150 \quad \text{and} \quad Pr(C|S) = 0.175$$

$$\therefore Pr(C) \neq Pr(C|S)$$

وبما ان الاحتمالين غير متساويين اذن الحدثان (C,S) غير مستقلين.

7- بما ان الحدثين (C, S) غير مستقلين اذن احتمال الاتحاد يكون كالاتي:

$$Pr(C \cup S) = Pr(C) + Pr(S) - Pr(C \cap S)$$
$$= 0.150 + 0.545 - 0.095 = 0.600$$

ويمكن حساب أي احتمال لاي يحدث يمكن تكوينه من الجدول المذكور آنفاً باستعمال قواعد الاحتمالات للعلاقات بين الاحداث.

تمارين الفصل التاسع

1- ثلاث سيدات ينتظرن الولادة، اكتب فراغ العينة لنوع المواليد إذا كانت كل منهن ستنجب مولوداً واحداً فقط.

2- كان فضاء العينة يمثل الأرقام الاتية وتم تعريف الاحداث على هذا الفضاء. والمطلوب حساب الاحتمالات لهذه الاحداث باستعمال التعريف التجريبي للاحتمال.

S = $\{0.5, 0, 3, 5, -2, -4, 6, -9\}$,   A= $\{0, 3, -2, -9\}$;

B= $\{0.5, 5, -2, -4\}$;   C= $\{0.5, 6, -4, 3\}$.

3- تحمل حقيبة اليد قفلاً رقمياً يتكون من ثلاث خانات متماثلة، كل خانة يمكن ان تحمل الأرقام: {0، 1، ... ، 9} بكم طريقة يمكن بها تكوين رقم سري يتكون من ثلاث اعداد اذا كان تكرار الرقم غير ممكن.

4- دخل عيادة الطبيب 8 مرضى لغرض العلاج والمطلوب:

أ- بكم طريقة يمكنهم الجلوس في العيادة في صف واحد به 8 مقاعد؟

ب-نفرض انهم لم يجدوا الا 6 مقاعد فبكم طريقة يمكنهم الجلوس؟

5- كان يجلس في عيادة دكتور 14 مريض، 8 منهم لديهم ضغط الدم عال. سحبنا شخصين في ان واحد من الجالسين. والمطلوب:

أ- ما هو احتمال ان يكونوا من ذوي الضغط العالي؟

ب-ما هو احتمال ان يكونوا من غير الذين لديهم ضغط الدم العالي؟

ت-ما هو احتمال ان يكون أحدهما لديه ضغط والأخر ليس لديه ضغط؟

6- إذا توافرت لديك المعلومات الاتية عن الاحتمالات للحدثين (A, B):

$$Pr(A) = \frac{3}{8}, \quad Pr(B) = \frac{1}{2}, \quad \Pr(A \cup B) = \frac{5}{8}$$

والمطلوب: احسب الاحتمالات لما يأتي:

$$1 - Pr(A^c); \quad 2 - \Pr(B^c); \quad 3 - \Pr(A \cap B);$$
$$4 - \Pr(A \cup B); \quad 5 - \Pr(A^c \cap B^c); \quad 6 - \Pr(A/B).$$

7- كانت احدى الدراسات تقول بان إصابة الشخص بمرض السكر 80% ونسبة اصابته بمرض ضغط الدم 70% ونسبة الإصابة بالمرضين معا تساوي 60%، اختير شخص بشكل عشوائي اوجد:

أ- احتمال ان يكون مصابا بأحد المرضين على الأقل.

ب-احتمال ان يكون مصابا بأحد المرضين على الأكثر.

ت-احتمال ان يكون مصابا بالسكر وغير مصاب بضغط الدم.

ث-احتمال ان يكون مصابا بالسكر علما انه مصاب بضغط الدم.

8- أعلن أحد المستشفيات عن حاجته الى عدد من الممرضين وبعد تصنيف المتقدمين لهذه الوظيفة على وفق المؤهل العلمي ولسنوات الخبرة وكانت نتائج المتقدمين كما في الجدول الاتي:

	لا يحمل شهادة جامعية (B)	يحمل شهادة جامعية (A)	المجموع
لديه خبرة (C)	40	20	60
ليس لديه خبرة (D)	30	10	40
المجموع	70	30	100

اخترنا ممرضا واحدا بصورة عشوائية والمطلوب:

أ- احتمال ان يكون ممن يحملون شهادة جامعية.

ب-احتمال ان يكون لديه خبرة ولا يحمل شهادة جامعية.

ت-احتمال ان لا تكون لديه شهادة وليس لديه خبرة.

9- في الجدول الاتي تصنيف 400 شخص حسب عادة التدخين ومستوى ضغط الدم

	يدخن (A)	لا يدخن (B)	المجموع
ضغط مرتفع (C)	40	10	50
ضغط متوسط (D)	70	130	200
ضغط منخفض (E)	55	95	150
المجموع	165	235	400

فاذا اختير أحد الأشخاص بشكل عشوائي، اوجد احتمال ان الشخص المختار:

أ- ضغط دمه مرتفع.

ب-مدخن.

ت-ضغط دمه مرتفع ويدخن.

ث-ضغط دمه مرتفع علما انه مدخن.

10- تحتوي ورقة الامتحان على 8 أسئلة وعلى الطالب ان يجيب عن 5 منها بشرط ان تتضمن سؤالين على الأقل من الأربعة الأولى، فكم طريقة يمكن بها الطالب اختيار الأسئلة.

الفصل العاشر

بعض التوزيعات الاحتمالية التي تناسب الظواهر الحيوية

Some probability distributions Appropriate for Biological Phenomena

10-1 تمهيد

ان دراسة المتغيرات العشوائية والتوزيعات الاحتمالية المصاحبة لها تساعد على الحصول على نتائج يمكن استخدامها في تقديرات معلمات المجتمع او في اختبار الفرضيات المتعلقة باتخاذ القرارات المهمة، وإذ يتخذ مثل هذه القرارات على اساس علمي صحيح. وبصفة عامة فان المتغير العشوائي هو المتغير الذي يأخذ قيما حقيقية مختلفة تعبر عن نتائج فضاء العينة لتجربة عشوائية، ومن ثم مجال هذا المتغير يشمل القيم الممكنة له كلها، ويكون لكل قيمة من القيم التي يأخذها المتغير احتمالاً معيناً، وتنقسم المتغيرات العشوائية الى نوعين، متغيرات عشوائية متقطعة وهي المتغيرات التي يكون مداها مجموعة محدودة (معدودة) من القيم، مثال ذلك عدد الاطفال في الاسرة، عدد المرضى في عيادة الطبيب، وغيرها من الأمثلة. والنوع الثاني هو المتغيرات العشوائية المتصلة (المستمرة) التي يكون مداها مجموعة من القيم غير محدودة (غير معدودة)، مثال ذلك قراءات ضغط الدم العالي او الواطئ او الوزن لمجموعة من الاشخاص. اما التوزيع الاحتمالي، فهو الذي يبين احتمالات حدوث القيم التي يمكن ان يأخذها المتغير العشوائي، والتي ترتبط باحتمالات النتائج الممكنة في فراغ العينة للتجربة العشوائية، وبمعنى آخر هو التوزيع التكراري النسبي للقيم التي يمكن أن يأخذها المتغير العشوائي. ومن ثم تكون لدينا نوعان من التوزيعات الاحتمالية كما ياتي:

2-10 التوزيعات الاحتمالية المتقطعة The Discrete Probability Distributions

المتغير العشوائي المتقطع هو الذي يأخذ قيم صحيحة ومنفصلة، ويرمز للمتغير العشوائي بشكل عام بحرف من الحروف الأبجدية الكبيرة مثل $X, Y, Z, \ldots$. ويرمز للقيم التي يأخذها المتغير بالحروف الأبجدية الصغيرة مثل $x, y, z, \ldots$. والتوزيع الاحتمالي، هو الذي يبين احتمالات حدوث القيم التي يمكن ان يأخذها المتغير، التي ترتبط باحتمالات النتائج الممكنة في فراغ العينة، وبمعنى آخر هو الجدول التكراري النسبي للقيم التي يمكن أن يأخذها المتغير. فعلى سبيل المثال إذا كان المتغير العشوائي المتقطع X يأخذ القيم، $X:\{x = x_1, x_2, \ldots, x_n\}$، وكان $Pr(X = x_i) = f(x_i)$ هو احتمال أن المتغير العشوائي يأخذ القيمة x_i، فإنه، يمكن تكوين جدول التوزيع الاحتمالي للمتغير العشوائي X، وهو جدول مكون من عمودين او صفين، الأول به القيم الممكنة للمتغير $X:\{x = x_1, x_2, \ldots, x_n\}$، والثاني به القيم الاحتمالية لهذا المتغير $Pr(X = x_i) = f(x_i)$، كمال في الجدول (1-10):

جدول (1.10) جدول التوزيع الاحتمالي للمتغير العشوائي المنفصل

x_i	x_1	x_2	$\cdots$	x_n	Sum.
$f(x_i)$	$f(x_1)$	$f(x_2)$	...	$f(x_n)$	1

وتسمى الدالة $f(x_i)$ بدالة التوزيع الاحتمالي للمتغير العشوائي X، ومن خصائص هذه الدالة ما ياتي: $0 < f(x_i < 1)$ و $\sum f(x_i) = 1$. **وفي** كثير من النواحي التطبيقية، تتبع بعض الظواهر توزيعات احتمالية خاصة، وهي التوزيعات التي يمكن حساب احتمالات قيم المتغير عن طريق معادلة رياضية، تسمى بدالة التوزيع الاحتمالي $f(x)$، وهذه المعادلة لها معلمات معينة، تسمى بمعلمات المجتمع الذي ينسب له هذا التوزيع، وهذه المعلمات ما هي إلا حقائق ثابتة مجهولة، وهي الأساس في حساب القيم الاحتمالية للتوزيع الاحتمالي للمجتمع محل الدراسة. **وفيما** يلي اهم التوزيعات الاحتمالية للمتغيرات العشوائية

المنفصلة (المتقطعة) والتي لها تطبيقات في الحياة العملية وفي مجال الاحصاء الحيوي مثل توزيع ذو الحدين وتوزيع بواسون.

10-2-1 توزيع ذو الحدين: . The Binomial Distribution

توجد العديد من الظواهر الحياتية تكون النتيجة الممكنة لها اما نجاحا او فشلا. فإذا كانت هناك تجربة عشوائية لها نتيجتان فقط هما ظهور حدث معين (نجاح) او عدم ظهور الحدث (فشل) مثل حالة اي شخص هل لديه مرض معين ام لا، وهذه التجربة إذا تكررت (n) من المرات فإننا نحصل في كل تكرار على نجاح (وجود المرض) او فشل (عدم وجود المرض) او العكس وبشرط نتيجة كل تكرار مستقلة عن نتيجة التكرار الاخر، بمعنى انه يتم اجراء التجربة (n) من التكرارات المستقلة. ويقال للتجربة التي تتكون من عدة تكرارات ثنائية بالتجربة ثنائية الحدين، مثل سؤال مجموعة من الطلبة ما إذا كانوا يدخنون أم لا. ويجب أن يتحقق في أي تجربة ثنائية الحدين الشروط الأربعة الآتية:

1- عدد التكرارات محدود.

2- لكل تكرار فقط نتيجتان ممكنتان وهما النجاح أو الفشل.

3- احتمال النجاح يكون ثابتاً في التكرارات كلها ولا يختلف من تكرار لآخر.

4- التكرارات مستقلة بعضها عن بعض.

وإذا افترضنا بأن احتمال النجاح في هذه التجربة هو(p) واحتمال الفشل فيها هو $(q = 1 - p)$ حيث ان $(p + q = 1)$. وفرضنا بأن (X) يمثل المتغير العشوائى المعرف على هذه التجربة ويكون عدد مرات النجاح في (n) من التكرارات، فان المتغير العشوائي (X) يسمى بمتغير ذي الحدين. ومعلمتا هذا التوزيع هما عدد المحاولات (n) واحتمال النجاح (p) ، ويمكن تمثيل دالة الكتلة الاحتمالية لتوزيع ذات الحدين في الجدول (10-2):

جدول رقم (2.10): قيم دالة الكتلة الاحتمالية لتوزيع ذات الحدين

x_i	0	1	⋯	n
$f(x_i) = Pr(X = x_i)$	$C_0^n p^0 q^{n-0} = q^n$	$C_1^n p^1 q^{n-1} = pq^{n-1}$	⋯	$C_n^n p^n q^{n-n} = p^n$

وبالتالي فان المعادلة التي تمثل دالة الكتلة الاحتمالية للتوزيع تأخذ الشكل الاتي:

$$f(x_i) = \Pr(X = x_i) = C_{x_i}^n p^{x_i} q^{n-x_i} \quad \text{حيث ان} \quad x_i = 0,1,2,\dots,n$$
$$= zero \qquad O/W$$

(1−10)

ومن خصائص هذه الدالة ما يلي:

$1-\ 0 < f(x_i) < 1 \qquad 2-\ \sum_{x=0}^{n} f(x_i) = 1$

ومعلمات هذا التوزيع والمؤشرات الإحصائية فيه تكون كما في الجدول (10−3):

جدول رقم (3.10): معلمات المجتمع والمؤشرات الاحصائية للعينة في التوزيع الثنائي

الاسم	معلمات التوزيع في المجتمع	المؤشرات الاحصائية في العينة
الوسط الحسابي	$\mu = N\pi$	$\bar{x} = np$
التباين	$\sigma^2 = N\pi Q$	$S^2 = npq$

ويستخدم هذا التوزيع في الحياة العملية في كثير من الظواهر، مثل انتشار مرض معين من عدمه، انتشار ظاهرة الاطفال حديثي الولادة وبأوزان اقل من الطبيعي من عدمه، الاصابة بمرض الانفلونزا من عدمه، وغيرها من الظواهر. ولتوضيح تطبيقات هذا التوزيع نورد الأمثلة الاتية:

مثال 1

اختيرت عينة تتكون من 20 شخص من محافظة معينة، فاذا علمت بان نسبة الاصابة بالأنفلونزا في الموسم الشتوي كانت (0.6)، فما هو احتمال ان يكون:

1- 5 اشخاص مصابون بالإنفلونزا.   2- جميعهم مرضى   3-جميعهم اصحاء

4-نصفهم مرضى   5- ما هو متوسط وتباين الاصابة بالمرض.

الحل:

المعطيات من المثال تكون كالاتي:

n=20   p= 0.6   q= 1-p = 1-0.6 = 0.4

$\Pr(X=x)=C_x^n p^x q^{n-x}$ حيث ان $C_x^n=\frac{n!}{x!(n-x)!}$

1- احتمال ان يكون 5 اشخاص مصابين بالمرض يحسب كالاتي:

$$\Pr(X=5)=C_5^{20}(0.6)^5(0.4)^{20-5}$$
$$=\frac{20\times19\times18\times17\times16\times15!}{5\times4\times3\times2\times1\times15!}\times(0.07776)\times(0.000001)$$
$$=15504\times0.07776\times0.000001\ =0.0013$$
$$\therefore Pr(X=5)=0.0013$$

2- احتمال ان يكون جميعهم مرضى

$$\Pr(X=20)=C_{20}^{20}(0.6)^{20}(0.4)^{20-20}$$
$$=1\times0.00004\times1=0.00004$$
$$\therefore Pr(X=20)=0.00004$$

3- احتمال ان يكون جميعهم اصحاء (هذا يعني ان عدد المرضى يساوي صفر)

$$Pr(X=0)=C_0^{20}(0.6)^0(0.4)^{20-0}$$
$$=1\times1\times0.0000=0.00$$
$$\therefore Pr(X=0)=1.0995E-8$$

4-احتمال ان يكون نصفهم مرضى

$$Pr(X=10)=C_{10}^{20}(0.6)^{10}(0.4)^{20-10}$$
$$=184756\times0.006046618\times0.0001$$
$$\therefore Pr(X=10)=0.111715$$

5-الوسط الحسابي والتباين للتوزيع يحسبان باستخدام العلاقات في الجدول (10-3) وكالاتي:

$$\bar{x}=np=20\times0.6=12$$
$$S^2=npq=20\times0.6\times0.4=4.8$$

10-2-2 حساب الاحتمالات باستعمال برنامج (SPSS)

نستعمل المثال الاتي لتوضيح كيفية استعمال البرنامج (SPSS) لحساب الاحتمالات في حالة التوزيعات المتقطعة.

<u>مثال 2</u> استعمل البيانات في المثال 1 لحساب الاحتمالات الاتية:

1- احتمال ان يكون 5 اشخاص مصابين بالمرض.

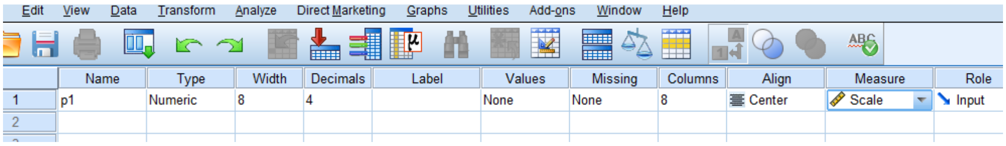
2- احتمال ان يكون على الأكثر 5 اشخاص مصابين بالمرض.

3- احتمال ان يكون على الأقل 5 اشخاص مصابين بالمرض.

<u>الحل:</u>

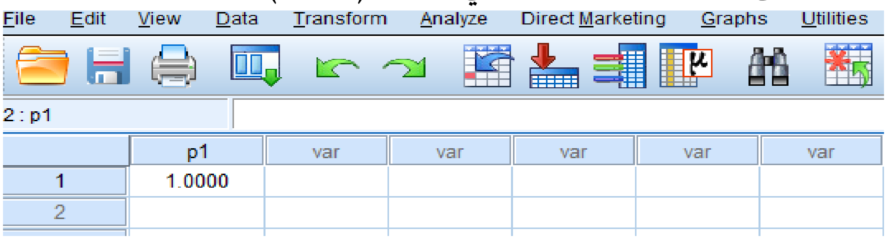
نتبع الخطوات الاتية لحساب الاحتمالات للتوزيع ذي الحدين باستعمال البرنامج الاحصائي (SPSS):

1- بعد فتح البرنامج نفتح صفحة أسماء المتغيرات (Variables view) ونثبت اسم متغير افتراضي على سبيل المثال (p1) ونحدد مواصفاته على تلك الصفحة كما في الشكل (10-1).



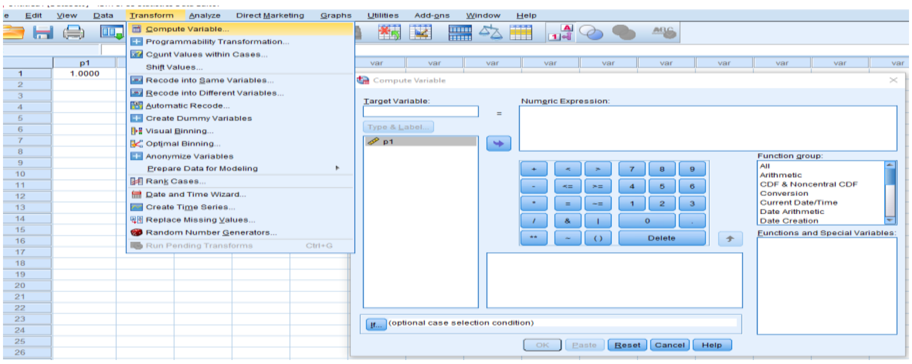
الشكل (10-1) تعريف اسم المتغير على صفحة أسماء المتغيرات

2- نرجع الى صفحة البيانات (Data view) نجد اسم المتغير نضع تحته اية قيمة افتراضية وذلك لان البرنامج لا يفعل قائمة الأوامر إذا لم تكن هناك بيانات على صفحة البيانات. كما في الشكل (10-2).



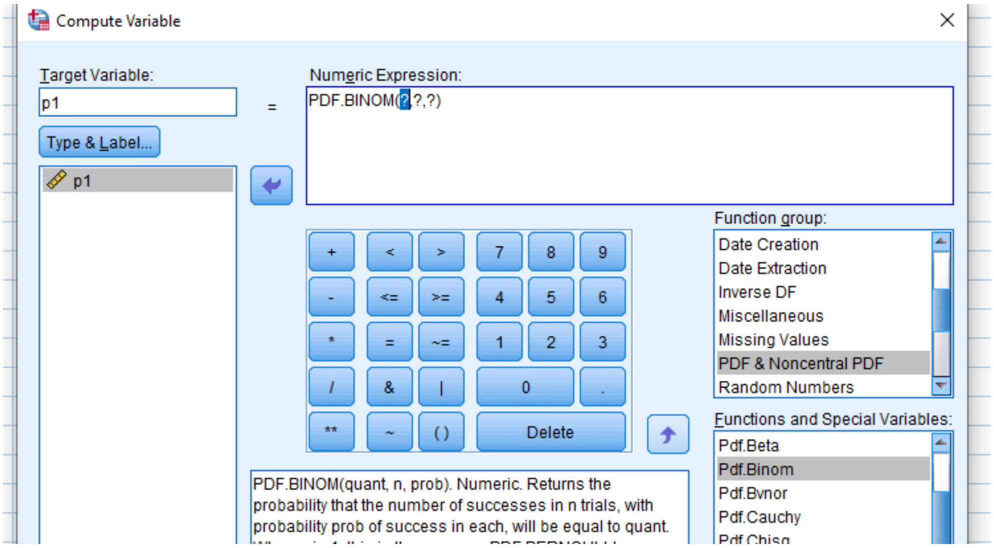
الشكل (10-2) تحديد قيمة افتراضية لقيمة المتغير

3- نختار من قائمة الأوامر الامر (Transform) ومن ثم الامر الفرعي (Compute variable فيفتح لنا المربع الحواري كما في الشكل (3-10).



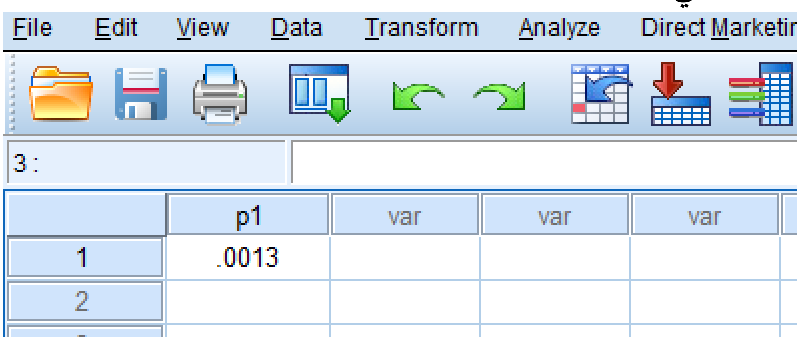
الشكل (3-10) مربع الحوار الخاص بحساب الاحتمالات من التوزيع ذو الحدين

4- يعتمد توزيع ذي الحدين على معلمتين هما (n, p) حيث ان (n) تمثل عدد المحاولات و(p) تمثل احتمال النجاح. وعليه فان المعلومات المتوافرة من المثال هي ان $(n = 20)$ وان احتمال النجاح هو $(p = 0.6)$ ولحساب الاحتمال الأول والذي يمكن تمثيله بالآتي: $(f(5) = Pr(X = 5))$ وهذا يعني حساب دالة الكتلة الاحتمالية عندما يكون المتغير $(X = 5)$. نكتب اسم المتغير $(p1)$ في المربع المسمى (Target variable) ثم نختار من المربع (Function group) دالة الكتلة الاحتمالية (pdf) سوف يفتح لنا في المربع الأسفل أسماء التوزيعات ومنها نختار (pdf Binom) والذي يمثل دالة الكتلة للتوزيع ذي الحدين وباستعمال السهم ننقل هذا الايعاز الى المربع الأعلى تحت اسم (Numeric Expression) كما في الشكل (4-10):



الشكل (10-4) اختيار دالة الكتلة الاحتمالية لحساب الاحتمال

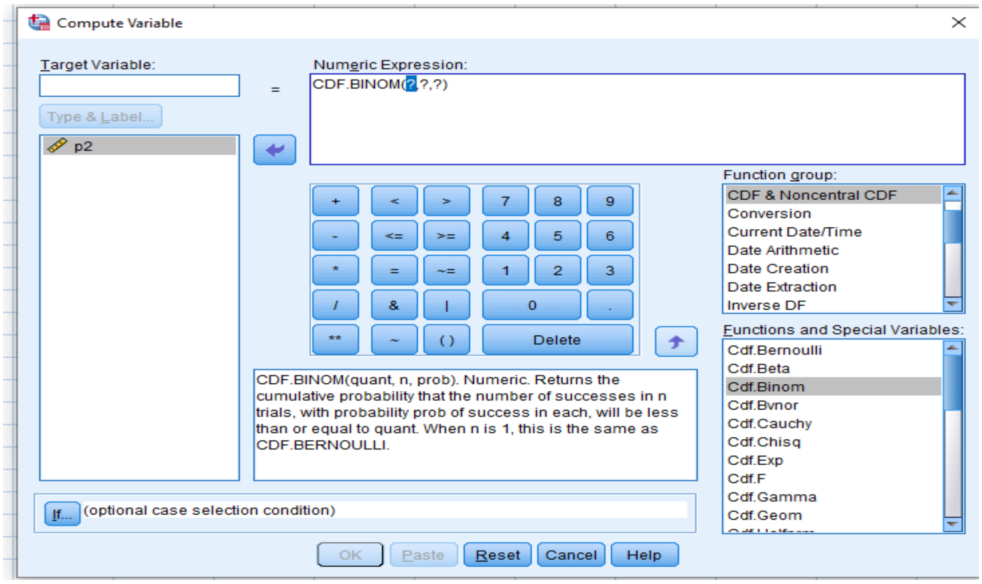
5- نجد ان العبارة (PDF.Binom(?,?,?) في المربع الأعلى تحتوي على ثلاث علامات استفهام كل منها تمثل قيمة يجب تعويضها فالأولى تمثل القيمة المطلوب حساب الاحتمال لها وهنا هي (5) والثانية تمثل عدد المحاولات (n=20) والثالثة تمثل احتمال النجاح(p=0.6) وهي كما موضحة في المربع أسفل السهم بالعبارة المؤشرة. وبعد تعويض القيم واختيار الامر (OK) يطلب منا قبول تغيير القيمة الافتراضية بالنتيجة المحسوبة فنقبل ذلك فنجد بان قيمة الاحتمال ظهرت تحت اسم المتغير (p1) كما في الشكل (10-5). وهي نفس النتيجة في المثال.



الشكل (10-5) نتيجة الاحتمال المحسوب

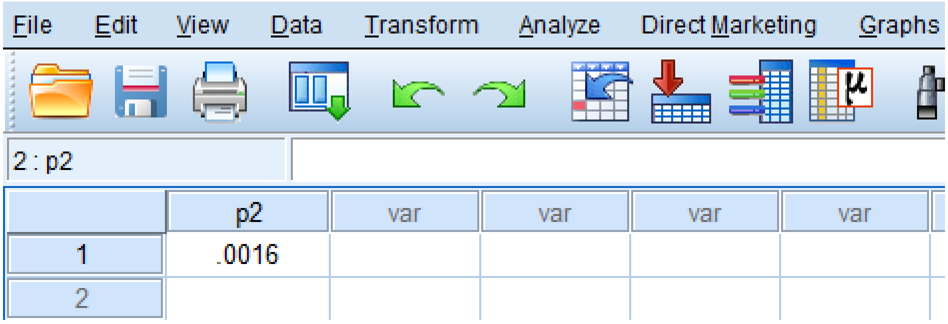
وهكذا تحسب الاحتمالات عندما يكون المتغير يساوي قيمة محددة $Pr(X = x)$ ولكن عندما يكون الاحتمال عبارة عن أكبر من او يساوي او أصغر من او يساوي قيمة معينة فيكون الحساب باستعمال البرنامج (SPSS) كما في الأمثلة الاتية:

2- احسب احتمال ان يكون على الأكثر 5 اشخاص مصابين بالمرض. يمكن الرمز لهذا الاحتمال بالشكل الاتي $(Pr(X \leq 5)$ ومن ثم فهو يمثل الاحتمال التراكمي او التجميعي للاحتمالات من 1الى 5، ومن ثم فإننا نتبع الخطوات الثلاث نفسها الأولى والخطوة الرابعة تكون اختيار دالة الكتلة التجميعية (CDF) من قائمة (Function group) ومن قائمة التوزيعات نختار (Cdf.Binom) ثم نحولها الى المربع (Numeric Expression) باستعمال السهم كما في الشكل $(6-10)$.



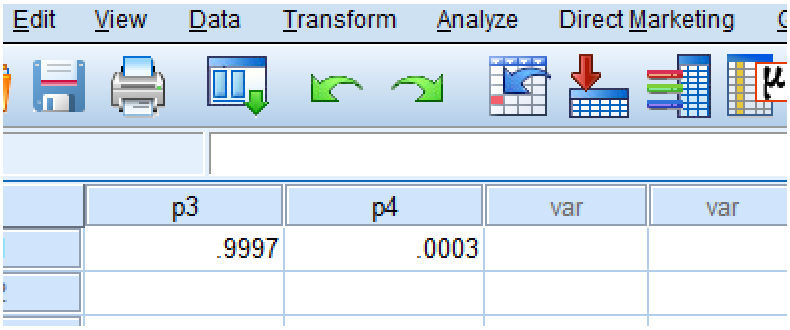
الشكل $(6-10)$ اختيار الدالة التجميعية

بعد تعويض القيم المطلوبة بدلا من علامات الاستفهام كما وضحنا في المثال السابق نجد ان قيمة الاحتمال تكون مساوية الى (0.0016). كما في الشكل $(7-10)$



الشكل (10−7) قيمة الاحتمال المطلوب

3− **احتمال ان يكون على الأقل 5 اشخاص مصابين بالمرض**. ويمكن ان نرمز لهذا الاحتمال على انه $(Pr(X \geq 5)$. وهذا يعني بانه الاحتمال التراكمي او التجميعي من القيمة 5 والى اخر قيمة بعدد المحاولات وفي المثال نجد بان اخر قيمة تساوي 20، وعليه نتبع الخطوات نفسها بعد ان نحول هذا الاحتمال الى الصيغة $(1 - Pr(X \leq 4)$ ونطرح قيم الاحتمال من 1 لان مجموع الاحتمالات يساوي 1. كما في الشكل (10−8).



الشكل (10−8) قيم الاحتمال المحسوب

تظهر النتائج في الشكل (10−8) الاحتمال الأول هو $(p_3 = 1 - \Pr(x \leq 4) =$ 0.9997) والاحتمال الثاني هو $(\Pr(X \leq 4) = 0.0003)$.

<u>مثال 3</u>

إذا كان من المعلوم أن نسبة الشفاء من مرض الربو باستخدام نوع معين من العقاقير الطبية هو (0.7)، إذا تناول هذا العقار 6 مصابين بهذا المرض. وعرفنا المتغير العشوائي X بأنه عدد المستجيبين (حالات الشفاء) لهذا العقار.

المطلوب:

1. ما هو نوع المتغير؟
2. اكتب دالة الكتلة الاحتمالية $f(x)$ لهذا المتغير.
3. احسب الاحتمالات التالية:
   أ- ما احتمال استجابة **3** مرضى لهذا العقار؟
   ب- ما هو احتمال استجابة مريض واحد على الأقل؟
   ت- ما هو احتمال استجابة **2** مرضى على الأكثر؟
4. احسب الوسط الحسابي، والانحراف المعياري لعدد حالات الاستجابة.
5. حدد شكل التوزيع.

الحـل:

1-عدد حالات الاستجابة X متغير كمي منفصل ثنائي، ومدى هذا المتغير في هذه الحالة هو: $X:\{x = 0,1,2,3,4,5,6\}$

2- دالة الكتلة الاحتمالية وحسب المعطيات الاتية: $n = 6$، $p = 0.7$،

$$q = 1 - p = 1 - 0.7 = 0.3$$
$$f(x_i) = Pr(X = x_i) = C_{x_i}^{n} p^{x_i} q^{n-x_i}$$
$$= C_{x_i}^{6}(0.7)^{x_i}(0.3)^{6-x_i} \qquad \text{حيث ان} \qquad x_i = 0,1,\dots,6$$

3-حساب الاحتمالات:

أ- حساب احتمال استجابة 3 مرضى لهذا الدواء:

$$f(3) = Pr(x_i = 3) = C_3^6(0.7)^3(0.3)^{6-3}$$
$$= \frac{6 \times 5 \times 4}{3 \times 2 \times 1} \times 0.343 \times 0.027$$
$$= 20 \times 0.00926 = 0.185$$

ب- حساب احتمال استجابة مريض واحد على الأقل:

$$Pr(x_i \geq 1) = f(1) + f(2) + f(3) + f(4) + f(5) + f(6)$$
$$= 1 - f(0)$$
$$= 1 - [C_0^6(0.7)^0(0.3)^6] = 1 - 1 \times 1 \times 0.0007 = 0.999$$

ت- حساب احتمال استجابة 2 مرضى على الأكثر:

$$Pr(x_i \leq 2) = f(0) + f(1) + f(2)$$
$$= C_0^6(0.7)^0(0.3)^6 + C_1^6(0.7)^1(0.3)^{6-1} + C_2^6(0.7)^2(0.3)^{6-2}$$
$$= 1 \times 1 \times 0.0007 + \frac{6}{1}(0.7)(0.0024) + \frac{6 \times 5}{2 \times 1}(0.49)(0.0081)$$
$$= 0.0007 + 0.0101 + 0.0595 = 0.0703$$

ملاحظة: يمكن استعمال نفس الخطوات السابقة لحساب الاحتمالات باستعمال البرنامج (SPSS) في هذا المثال، يترك تمريناً للقارئ للتطبيق.

4-حساب الوسط الحسابي، والانحراف المعياري لعدد حالات الاستجابة:

أ- الوسط الحسابي ($\bar{x}$) في حالة التوزيع ثنائي الحدين في حالة العينة يحسب بتطبيق العلاقة في الجدول (10-3)، إذن الوسط الحسابي هو:

$$\bar{x} = np = 6 \times 0.7 = 4.2$$

ب - الانحراف المعياري هو الجذر التربيعي الموجب للتباين، ولحساب التباين في التوزيع ثنائي الحدين يتم تطبيق العلاقة في الجدول (10-3)، ومنها يمكن التوصل إلى الصورة الآتية:

$$S^2 = npq$$

إذا تباين عدد حالات الاستجابة هو:

$$S^2 = 6 \times 0.7 \times 0.3 = 1.26$$

ومن ثم يأخذ الانحراف المعياري الصورة الآتية:

$$S = \sqrt{npq} = \sqrt{1.26} = 1.123$$

ويمكن حساب معامل الاختلاف النسبي، بتطبيق المعادلة الآتية:

$$V.C. = \frac{S}{\bar{x}} \times 100 = \frac{1.123}{4.2} \times 100 = 26.7\%$$

5-تحديد شكل التوزيع:

يتحدد شكل التوزيع ثنائي الحدين وفقا لقيمة احتمال النجاح p كما يأتي:

إذا كان $p = 0.5$ فإن التوزيع الاحتمالي ثنائي الحدين يكون متماثلاً.

إذا كان $p < 0.5$ فإن التوزيع الاحتمالي ثنائي الحدين يكون موجب الالتواء.

إذا كان $p > 0.5$ فإن التوزيع الاحتمالي ثنائي الحدين يكون سالب الالتواء.

وحيث أن $p = 0.7 > 0.5$ فإن توزيع عدد حالات الاستجابة سالب الالتواء.

يستخدم توزيع ثنائي الحدين لمعرفة عدد مرات وقوع الحدث موضع الاهتمام في عدد محدد من الاختبارات.

3-2-10 توزيع بواسون Poisson Distribution

يستخدم توزيع بواسون في التجارب التي تكون نتيجتها وقوع الحدث موضع الاهتمام عددا صحيحًا من المرات خلال وحده زمنية محددة (مثل الدقيقة أو الساعة أو اليوم ... الخ) أو على وحده أو منطقة محددة من الحيز (مثل الطول أو المساحة). وعليه فمن التطبيقات الشائعة لتوزيع بواسون التنبؤ بعدد الأحداث خلال زمن محدد او مساحة معينه، فضلا عن ذلك فأنه يصف الظواهر النادرة مثل عدد الزلازل السنوية أو عدد الحرائق الشهرية أو عدد الحوادث الأسبوعية، ومن أمثلة تطبيقات توزيع بواسون:

أ –المسائل التي تحتاج إلى الانتظار أو المرتبطة بالزمن مثل:

- عدد المرضى الذين يصلون إلى عيادة طبيب ما في ساعة واحدة.
- عدد القطارات التي تغادر المحطة في الساعة.
- عدد المراجعين في أحد مراكز الخدمة في اليوم.
- عدد المكالمات الهاتفية التي يتلقاها عامل بدالة في الدقيقة.

ب –المسائل التي ترتبط بمناطق الفراغ أو الحيز مثل المسائل المرتبطة بمراقبة الجودة ومنها:

- عدد العيوب الموجودة في قطعة قماش محددة المساحة.
- عدد العيوب الموجودة في سلك للهاتف طوله كيلومتران.
- عدد الاخطاء المطبعية في صفحة ما من كتاب.
- عدد الحوادث على إحدى الطرق وغير ذلك.

وهكذا فان المتغير البواسوني هو متغير متقطع يأخذ قيمة من القيم صفر أو 1 أو 2 أو ... في مدة زمنية متصلة أو في حيز أو منطقة متصلة ويجب أن يحقق الشروط الثلاثة الآتية:

1. عدد مرات النجاح مستقلة في فترتين متتاليتين من الزمن أو منطقتين متصلتين من الحيز.
2. احتمال النجاح خلال فترة قصيرة من الزمن أو منطقة صغيرة من الحيز يتناسب مع طول هذه الفترة أو مساحة هذا الحيز.
3. احتمال نجاحين أو أكثر خلال فترة زمنية قصيرة أو منطقة من الحيز صغيرة يكون صغيرًا جدًا بحيث يمكن إهماله.

ويعد توزيع بواسون من التوزيعات الاحتمالية المتقطعة التي تصف متغيرات عشوائية متقطعة لها الشروط السابقة. ويعد من اهم التوزيعات في المسائل المتعلقة ببعض الظواهر النادرة مثل الزلزال، والحرائق، الحوادث على الطرق وغير ذلك. إذا كان متوسط عدد مرات وقوع حادث على وفق معدل زمني معين هو (λ)، وكان المتغير العشوائي (X) يعبر عن عدد مرات وقوع الحادث على وفق هذا المعدل، فإن مدى المتغير العشوائي (X) هو: $x = 0,1,2, \dots$ ، وهذا المدى يمثل فئة مفتوحة من اليمين، فإن الاحتمال $f(x) = \Pr(X = x)$ الذي يعبر عن احتمال وقوع الحادث عدد (x) من المرات على وفق هذا المعدل، يسمى التوزيع الاحتمالي لمتغير بواسون العشوائي ويدعي بتوزيع بواسون ويعبر عن دالة الكتلة الاحتمالية بالصيغة (2−10):

$$\left.\begin{aligned} f(x) = \Pr(X = x) &= \frac{e^{-\lambda}\lambda^{x}}{x!} \quad , \quad x = 0,1,2,\dots \\ &= Zero \qquad\qquad O/W \end{aligned}\right] \qquad (2-10)$$

حيث ان:

e = 2.718 تمثل مقدار ثابت. λ = متوسط عدد مرات النجاح في الفترة الزمنية أو منطقة الحيز المحددة وتسمى معلمة التوزيع. وتأخذ X قيماً صحيحة موجبة اعتباراً من الصفر الى مالانهاية. ولإثبات انها دالة كتلة احتماليه يجب ان تحقق شروط دالة الكتلة الاحتمالية. وبما ان $\lambda > 0$ في دالة الكتلة الاحتمالية $f(x)$ لذلك فان الشرط

الاول هو $f(x) \geq 0$ اما الشرط الثاني فهو $\sum_{x=0}^{\infty} f(x) = 1$ وعليه فان الدالة $f(x)$ تحقق شروط دالة كثافة الاحتمال لمتغير عشوائي متقطع. والمعلمة (λ) فى توزيع بواسون هى الوسط الحسابى والتباين للتوزيع. وللتعرف على تطبيقات توزيع بواسون نورد الامثلة الاتية:

مثال 4

إذا كانت نسبة الاصابة بمرض الانفلونزا في مدينة معينة هي(0.002) وان (X) متغير عشوائي يمثل عدد الاشخاص المصابين بالمرض يتبع توزيع بواسون. نفرض أنه سحبنا عينة عشوائية من مائة شخص. **المطلوب**:

1. ايجاد دالة الكتلة الاحتمالية لهذا المتغير.
2. احتمال الحصول على شخص واحد مصاب بالمرض.
3. احتمال الحصول على شخص واحد مصاب بالمرض على الاكثر.

الحل: -

$$\lambda = np = 100 \times 0.002 = 0.2$$

1. بما أن X يتبع توزيع بواسون فان دالة الكتلة الاحتمالية هي:

$$f(x) = Pr(X = x) = \frac{e^{-0.2}(0.2)^x}{x!}, \quad x = 0,1,2,\ldots$$

2. احتمال الحصول على شخص واحد مصاب بالمرض هو:

$$f(1) = \Pr(X = 1) = \frac{e^{-0.2}(0.2)^1}{1!} = \frac{0.2}{e^{0.2}} = 0.1638$$

3. احتمال الحصول على شخص واحد مصاب على الاكثر:

$$\begin{aligned}\Pr(X \leq 1) &= f(0) + f(1)\\ &= \frac{e^{-0.2}(0.2)^0}{0!} + \frac{e^{-0.2}(0.2)^1}{1!}\\ &= e^{-0.2} + e^{-0.2}(0.2)\\ &= e^{-0.2}(1 + 0.2) = 0.9825\end{aligned}$$

مثال 5

افترض أن متوسط عدد الزبائن الذين يدخلون أحد العيادات الطبية هو ثلاثة زبائن في الساعة، باستخدام توزيع بواسون أوجد:

1. التوزيع الاحتمالي لهذا المتغير.
2. احتمال ان أربعة مرضى بالضبط سيدخلون العيادة خلال ساعة معينة.
3. احتمال على الاقل مريض واحد خلال الساعة يدخل الى العيادة.

الحل:

1- بما أن X يتبع توزيع بواسون فان دالة الكتلة الاحتمالية هي:

$$f(x) = Pr(X = x) = \frac{e^{-3}(3)^x}{x!}, \qquad x = 0,1,2, \dots$$

2- احتمال ان أربعة مرضى بالضبط سيدخلون العيادة خلال ساعة معينة

$$f(4) = Pr(X = 4) = \frac{e^{-3}(3)^4}{4!}$$
$$= \frac{0.05 \times 81}{24} = 0.1688$$

3- احتمال على الاقل مريض واحد خلال الساعة يدخل الى العيادة

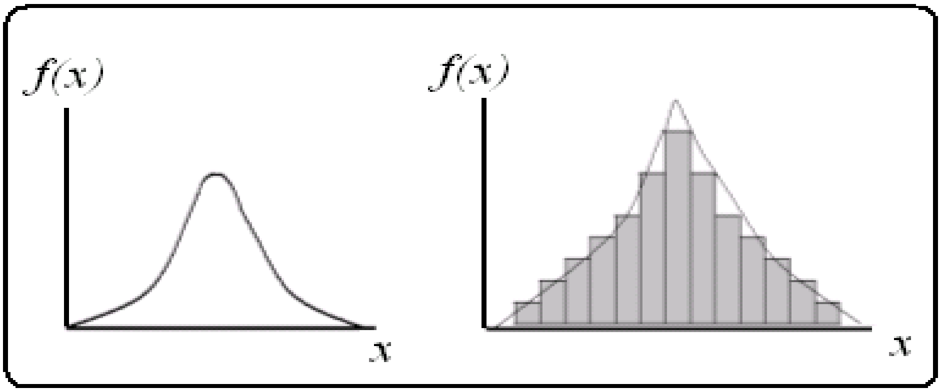
$$Pr(X \geq 1) = 1 - f(0) = 1 - \frac{e^{-3}(3)^0}{0!}$$
$$= 1 - e^{-3} = 1 - 0.05 = 0.95$$

ملاحظة: لغرض الاختصار وعدم التكرار، يمكن حساب الاحتمالات حسب توزيع بواسون باستعمال البرنامج (SPSS) وحسب الخطوات السابقة في حالة التوزيع ذو الحدين بعد تغيير التوزيع الى توزيع بواسون.

3-10 التوزيعات الاحتمالية المستمرة Continuous Probability distributions

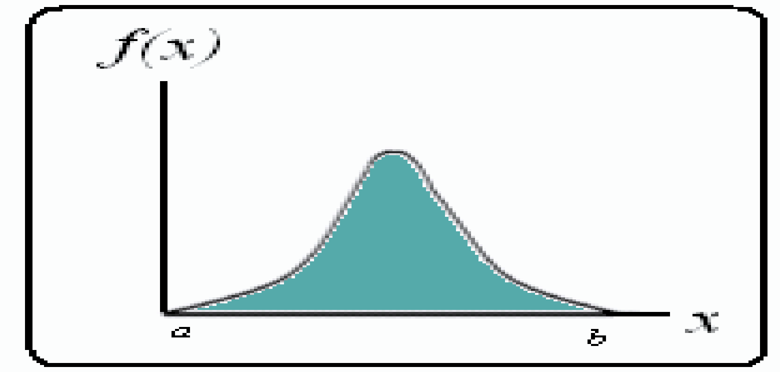
المتغير العشوائي المستمر، هو الذي يأخذ قيما متصلة، وبعدد غير نهائي من القيم الممكنة له داخل مجاله، فإذا كان X متغير عشوائي مستمر، ويقع في المدى (a, b)، أي أن:$\{X = x: a < x < b\}$ ، فإن للمتغير X عدداً غير

نهائي من القيم تقع بين الحدين الأدنى والأعلى (a, b)، ومن الأمثلة على المتغيرات الكمية المستمرة وزن الجسم بالكيلوغرام والطول بالسنتمتر وهكذا الأمثلة على المتغير الكمي المستمر كثيرة. وعند تمثيل بيانات المتغير الكمي المستمر في شكل المدرج التكراري النسبي، نجد أن شكل هذا المدرج هو أقرب وصف لمنحنى التوزيع الاحتمالي للمتغير المستمر، وكلما ضاقت الفترات بين مراكز الفئات، يمكن الحصول على رسم دقيق للمنحنى الخاص بدالة احتمال المتغير المستمر، كما هو مبين بالشكل $(10-9)$:



شكل (10-9): يمثل منحنى التوزيع الاحتمالي للمتغير العشوائي المستمر

والمساحة أسفل المنحنى تعبر عن مجموع الاحتمالات الكلية، ولذا فان هذه المساحة تساوي الواحد الصحيح، وتسمى الدالة $f(x)$ بدالة كثافة الاحتمال Probability Dencity Function *(pdf)*، ومن خلالها نستطيع إيجاد احتمالات الحوادث المعبر عنها بواسطة المتغير العشوائي . فالمساحة تحت منحنى هذه الدالة تعطي احتمال وقوع المتغير العشوائي X في الفترات المناظرة على المحور الأفقي. وبفرض المتغير العشوائي المستمر يقع في المدى: $\{X = x : a < x < b\}$، فأن منحنى هذه الدالة يأخذ الصورة كما في الشكل $(10-10)$:



شكل (10-10): يمثل المساحة تحت منحنى التوزيع للمدى $a < x < b$

ومن خصائص دالة كثافة الاحتمال $f(x)$ ما يلي:

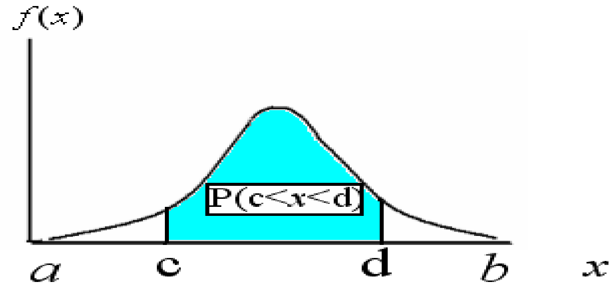
1- الدالة $f(x)$ موجبة داخل المدى (a, b) أي أن: $f(x) > 0$، $x \in (a, b)$

2- التكامل على حدود المتغير من الحد الأدنى a حتى الحد الأعلى b يعبر عن مجموع الاحتمالات الكلية، لذا يساوي الواحد الصحيح، أي أن:

$$\int_{x=a}^{x=b} f(x)dx = 1$$

إذ إن الشكل الرياضي المذكور آنفاً يسمى بالتكامل المحدود من $x = a$ حتى $x = b$، وهذا يعني إيجاد المساحة أسفل المنحني بين (a, b).

3- حساب احتمال أن المتغير العشوائي المستمر يقع في المدى (d, c) أي حساب الاحتمال $Pr(c < x < d)$، يجب حساب المساحة أسفل المنحنى من $x = c$ حتى $x = d$ كما هي مبينة في الشكل البياني (11-10):



شكل (11−10): يمثل المساحة تحت منحنى التوزيع للمدى $c < x < d$

ويتم ذلك بإيجاد التكامل المحدود في هذا المدى، كالاتي:

$$Pr(c < x < d) = \int_{x=c}^{x=d} f(x)dx = [g(x)]_c^d = g(d) - g(c)$$

4−يكون الاحتمال $Pr(x = c)$ في المتغير المستمر مساويا للصفر، أي أن:

$$Pr(x = c) = 0$$

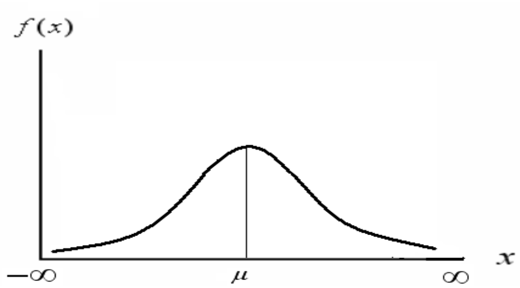
وهناك عدد من التوزيعات الاحتمالية المستمرة الخاصة، ولها دوال كثافة احتمال محددة، وفيما يأتي بعض هذه التوزيعات:

10−3−1 التوزيع الطبيعي The Normal Distribution

يعد هذا التوزيع من أكثر التوزيعات الاحتمالية استخداما في النواحي التطبيقية، ومنها الاستدلال الإحصائي شامل التقدير، واختبارات الفروض، كما أن معظم التوزيعات يمكن تقريبها إلى هذا التوزيع، وتتمثل دالة كثافة الاحتمال(pdf) في هذا التوزيع على فرض ان المتغير X متغير عشوائي له توزيع طبيعي، مداه هو $-\infty < x < \infty$ فإن دالة كثافة احتماله هي:

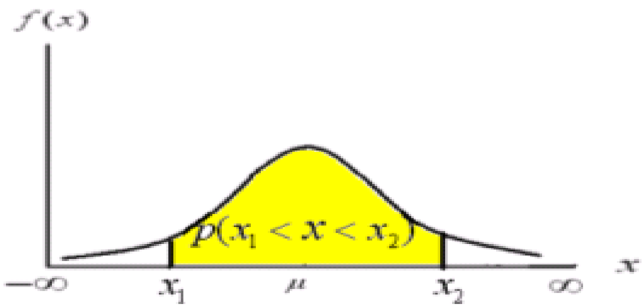
$$\left.\begin{array}{ll} f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} & ,-\infty < x < \infty, \quad \pi = 22/7 \\ \quad = Zero & \qquad O/W \end{array}\right] \quad (3-10)$$

وهذا التوزيع له منحنى متماثل على جانبي الوسط الحسابي μ يأخذ الشكل (12−10):



شكل (10-12) يمثل منحنى التوزيع الاحتمالي للمتغير العشوائي الطبيعي

وله معلمتان هما: الوسط الحسابي: $E(x) = \mu$ والتباين : $var(x) = \sigma^2$ ويعبر عن توزيع المتغير (X) بالرموز: $X \sim N(\mu, \sigma^2)$ ويعني ذلك أن المتغير العشوائي (X) يتبع التوزيع الطبيعي بمتوسط $\boldsymbol{\mu}$ ، وتباين $\boldsymbol{\sigma^2}$. وفي هذا التوزيع فإن :المتوسط = الوسيط = المنوال=μ . ويعتبر هذا التوزيع من أكثر التوزيعات الاحتمالية استعمالا، لأنه يشتق منه أكثر التوزيعات الاحتمالية الأخرى المستخدمة في الاستدلال الإحصائي. وتحسب الاحتمالات في هذا التوزيع باستعمال التكامل المحدود، فعلى سبيل المثال إذا كان المطلوب حساب الاحتمال الآتي $Pr(x_1 < X < x_2)$، فهذا الاحتمال يحدد بالمساحة كما في الشكل $(13-10)$:



شكل (10-13) يمثل المساحة تحت منحنى التوزيع الطبيعي للفترة $x_1 < x < x_2$

وحيث أن هذا التوزيع من التوزيعات المستمرة، فإن هذه المساحة (الاحتمال) تحسب بإيجاد التكامل في المعادلة (10−4):

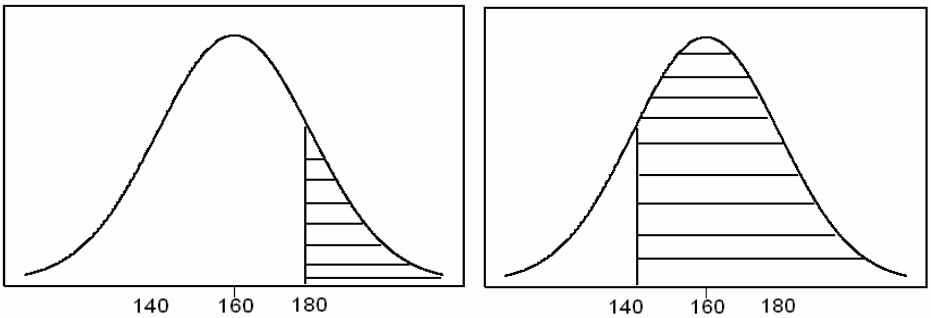
$$Pr(x_1 < X < x_2)$$
$$= \int_{x_1}^{x_2} f(x)dx = \int_{x_1}^{x_2} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \qquad (4-10)$$

<u>مثال 6</u>

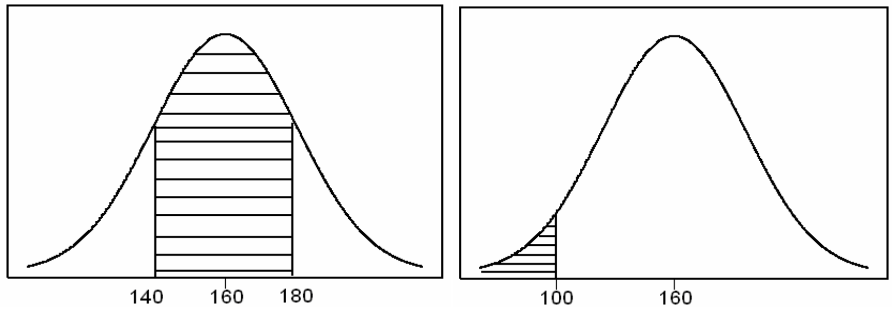
إذا كان طول الشخص X في مجتمع ما يتوزع تقريبا على وفق التوزيع الطبيعي بمتوسط 160 سم وانحراف معياري**5** سم. ارسم الاحتمالات الاتية بمساحات تحت منحنى التوزيع الطبيعي :

$$\Pr(X < 100), \Pr(140 < X < 180), \Pr(X > 180), \Pr(X > 140)$$

<u>الحل</u>: المساحات في الشكل (14.10) الاتية تمثل الاحتمالات المطلوبة:



$$Pr(x > 180) = \int_{180}^{\infty} f(x)dx \qquad , Pr(x > 140) = \int_{140}^{\infty} f(x)dx$$



$$Pr(140 < x < 180) = \int_{140}^{180} f(x)dx: \ \ Pr(x < 100) = \int_{-\infty}^{100} f(x)dx$$

شكل (10-14): يمثل المساحات تحت منحنى التوزيع الطبيعي وحسب الفترات المحددة

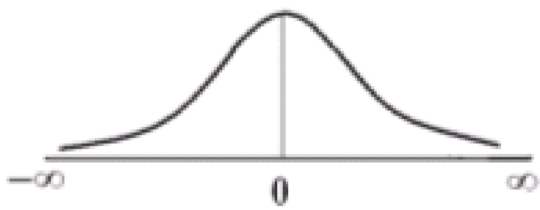
وهذه التكاملات يصعب حسابها، لذلك لجأ الإحصائيون إلى عمل تحويل رياضي Mathematical Transformation، يساعد على ايجاد توزيع احتمالي يمكن استخدامه لحساب مثل هذه الاحتمالات، يسمى بالتوزيع الطبيعي المعياري. ويمكن تحويل المتغير العشوائي الطبيعي الى متغير عشوائي طبيعي معياري باستخدام الصيغة الاتية:

$$Z = \frac{x-\mu}{\sigma} \qquad (5-10)$$

ويعرف المتغير الجديد (Z) بالمتغير الطبيعي المعياري Standard Normal Variable، وهذا المتغير له دالة كثافة احتمال تأخذ الشكل الاتي:

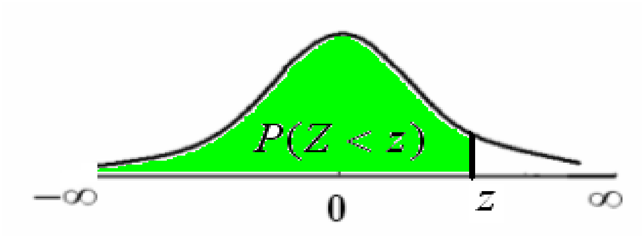
$$\left.\begin{aligned} f(z) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \ , \ -\infty < z < \infty \\ &= Zero \qquad\qquad o/w \end{aligned}\right] \qquad (6-10)$$

ومن خصائص هذا المتغير ان متوسطه يساوي $E(z) = 0$ وتباينه هو: $var(z) = 1$ ويعبر عن توزيع المتغير Z بالرموز: $Z \sim N(0,1)$ وهذا يعني بأن المتغير العشوائي X يتبع التوزيع الطبيعي المعياري بمتوسط (0)، وتباين (1) ويأخذ المنحنى شكل الناقوس المتماثل على جانبي الصفر كما في الشكل (15-10):



شكل (10-15): يمثل المساحة تحت منحنى التوزيع الطبيعي المعياري

ولقد تم ايجاد جداول إحصائية لحساب القيم الاحتمالية المقابلة لقيم المتغير العشوائي كما في الملحق رقم (1) ودالة التوزيع التجميعي تكون: $F(z) =$ $Pr(Z < z)$، كما هو مبين بالشكل (10−16):



شكل (10-16) يمثل المساحة تحت منحنى التوزيع الطبيعي المعياري التجميعي

ولهذا عند حساب الاحتمال $Pr(x_1 < X < x_2)$ نستعمل التحويل كما في المعادلة (10−5) لكي يتم تحويل القيم الطبيعية (x_1, x_2) إلى قيم طبيعية معيارية:

$$z_2 = (x_2 - \mu)/\sigma, \qquad z_1 = (x_1 - \mu)/\sigma,$$

ومن ثم يكون الاحتمال: $Pr(x_1 < X < x_2) = Pr(z_1 < Z < z_2)$ مما يساعد على استعمال جداول التوزيع الطبيعي المعياري، وهناك نوعان من الجداول التي تعطي المساحة تحت المنحنى الطبيعي المعياري، النوع الاول يعطي كامل المساحة التي تساوي الواحد الصحيح وبالتالي تعطي المساحة الخاصة بالاحتمال$F(z) = Pr(Z < z)$ التي تقع على يسار القيمة للمتغير العشوائي المعياري كما في الشكل المذكور آنفاً ولاستعمال هذا الجدول نضع في الاعتبار القواعد الاتية:

1. $Pr(Z < z) = F(z) =$ مباشرة من الجدول
2. $Pr(Z > z) = 1 - Pr(Z < z) = 1 - F(z)$
3. $\Pr(z_1 < Z < z_2)$
$= \Pr(Z < z_2) - \Pr(Z < z_1)$
$= F(z_2) - F(z_1)$
4. $Pr(Z < 0) = Pr(Z > 0) = F(0) = 0.5$
5. $Pr(Z = z) = 0$

والنوع الثاني يعطي نصف المساحة المحصورة بين الصفر وقيمة الاحتمال المطلوب. ولتوضيح الاستعمال المباشر لجدول الاحتمالات تحت منحنى التوزيع الطبيعي المعياري نأخذ الامثلة الاتية:

مثال7

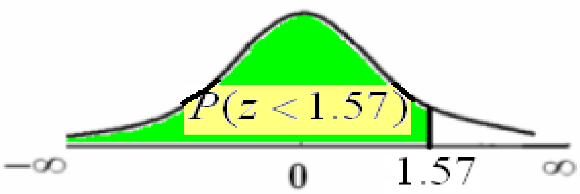
إذا كان المتغير العشوائي $Z \sim N(0,1)$ اوجد الاحتمالات الاتية باستعمال الجدول الخاص باحتمالات التوزيع الطبيعي المعياري:

أ- $Pr(z < 1.57)$ ب- $Pr(z < -2.33)$ ت- $Pr(z > 1.96)$

ث- $Pr(-2.01 < z < 1.28)$ ج - أوجد قيمة Z التي يسبقها مساحة مقدارها . 0.9505

الحل

أ- يحدد الاحتمال $Pr(z < 1.57) = \Phi(1.57)$ بالمساحة التي تقع على يسار القيمة (1.57) والتي تكون أسفل المنحنى مباشرة من الجدول في الملحق (1) كما في الاشكال الاتية:



شكل (10-17): يمثل المساحة تحت منحنى التوزيع الطبيعي المقابلة $Pr(z < 1.57)$

ويستعمل جدول الاحتمالات للتوزيع الطبيعي القياسي لإيجاد الاحتمال كما هو

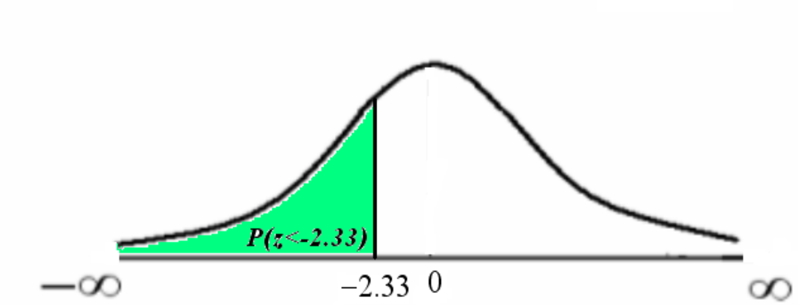
مبين في الجدول (4−10):

جدول (4−10) مقطع من جدول الاحتمالات للتوزيع الطبيعي المعياري

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
⋮										
1.00										
1.10										
1.20										
1.30										
1.40										
1.50								0.9418		
⋮										

ويكون الاحتمال المطلوب هو: $Pr(z < 1.57) = F(1.57) = 0.9418$

ب− تكون المساحة المعبرة عن الاحتمال $Pr(z < -2.33) = \Phi(-2.33)$ موضحة بالشكل الاتي ونوجدها من الجدول مباشرة:



شكل (18-10) يمثل المساحة تحت منحنى التوزيع الطبيعي المقابلة $Pr(z < -2.33)$

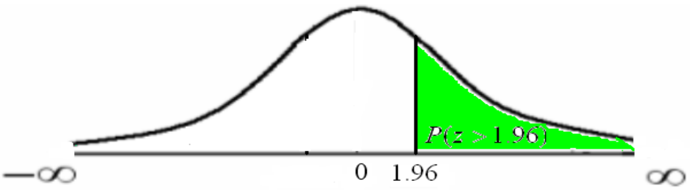
ويستعمل جدول التوزيع الطبيعي المعياري كما هو مبين في الجدول (5−10) لإيجاد الاحتمال:

جدول (10−5) مقطع من جدول الاحتمالات للتوزيع الطبيعي المعياري

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−2.50										
:										
−2.40										
:										
−2.30				**0.0099**						
:										

وعليه فان الاحتمال يكون: $Pr(z < -2.33) = \Phi(-2.33) = 0.0099$

ج− الشكل (10−19) يوضح المساحة التي تمثل الاحتمال $Pr(z > 1.96)$:



شكل (10-19) يمثل المساحة تحت منحنى التوزيع الطبيعي المقابلة $Pr(z > 1.96)$

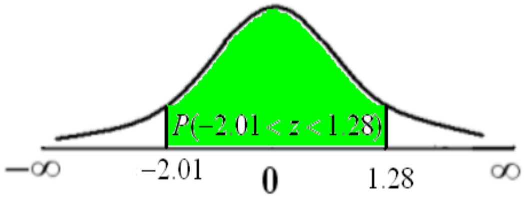
ولحساب هذا الاحتمال يجب استعمال خصائص دالة التوزيع التجميعي، حيث أن:

$$Pr(z > 1.96) = 1 - Pr(z < 1.96) = 1 - \Phi(1.96)$$

وبالبحث في جدول التوزيع الطبيعي المعياري بالطريقة السابقة نفسها عن الاحتمال المقابل للقيمة 1.96 نجد أن: $Pr(z < 1.96) = 0.9750$، وبالتالي فان الاحتمال المطلوب هو:

$$Pr(z > 1.96) = 1 - 0.9750 = 0.0250$$

د- الاحتمال $Pr(-2.01 < z < 1.28)$ يمكن تمثيله بالمساحة المحصورة بين القيمتين وكما موضح بالشكل (10−20):



شكل (10-20): يمثل المساحة تحت منحنى التوزيع المقابلة $Pr(-2.01 < z < 1.28)$

ويمكن حساب هذا الاحتمال باستعمال خصائص دالة التوزيع التجميعي، إذ إن:

$$Pr(-2.01 < z < 1.28) = Pr(z < 1.28) - Pr(z < -2.01)$$

وبالبحث في جدول التوزيع الطبيعي المعياري عن هاتين القيمتين، نجد أن:

$$Pr(z < 1.28) = 0.8997 \ ,$$
$$Pr(z < -2.01) = 0.0222$$

ومن ثم فان الاحتمال المطلوب يساوي:

$$Pr(-2.01 < z < 1.28)$$
$$= 0.8997 - 0.0222 = 0.8775$$

هـ- ايجاد قيمة Z $\quad \Pr(Z < z) = \Phi(z) = 0.9505$

$$\therefore z = 1.65$$

كما موضحه في الجدول (10-6):

جدول(10−6) مقطع من جدول الاحتمالات للتوزيع الطبيعي المعياري

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
:										
:										
1.4										
1.5										
1.6						0.9505				
1.7										
:										

مثال 8

لنفرض أن مستوى هيموجلوبين الدم في أحد المجتمعات البشرية يتوزع على وفق التوزيع الطبيعي بمتوسط 16 وانحراف معياري 0.9 .

1- إذا اخترنا أحد الاشخاص بشكل عشوائي فما هو احتمال أن يكون مستوى هيموجلوبين الدم لديه أكبر من 14.

2- ما نسبة الأشخاص الذين مستوى هيموجلوبين الدم لديهم أكبر من 14 .

3- ما نسبة الأشخاص الذين يتراوح مستوى هيموجلوبين الدم لديهم من 14 إلى 18.

الحل

ليكن مستوى هيموجلوبين الدم = X المعطيات:

$$\mu = 16 \quad \sigma = 0.9 \Leftrightarrow \sigma^2 = 0.81$$

$\therefore X \sim N(16\,, 0.81)$.

1- احتمال أن يكون مستوى هيموجلوبين الدم لديه أكبر من 14.

$$Pr(X > 14) = 1 - P(X < 14)$$

$$= 1 - Pr\left(Z < \frac{14-16}{0.9}\right) = 1 - Pr(Z < -2.22)$$

$$= 1 - F(-2.22) = 1 - 0.0132 = 0.9868$$

2- نسبة الأشخاص الذين مستوى هيموجلوبين الدم لديهم أكبر من 14 هي :

$$\Pr(X > 14) \times 100\% = 0.9868 \times 100\% = 98.68\%$$

3- نسبة الأشخاص الذين يتراوح مستوى هيموجلوبين الدم لديهم من 14 إلى 18.

$$Pr(14 < X < 18) = Pr(X < 18) - Pr(X < 14)$$

$$= Pr(Z < \frac{18-16}{0.9}) - Pr(Z < \frac{14-16}{0.9})$$

$$= \Pr(Z < 2.22) - \Pr(Z < -2.22)$$

$$= F(2.22) - F(-2.22) = 0.9868 - 0.0132 = 0.9736$$

وعليه فإن نسبة الأشخاص الذين يتراوح مستوى هيموجلوبين الدم لديهم من 14 إلى 18هي: 97.36%.

مثال 9

إذا كانت درجات الامتحان النهائي لمادة الاحصاء الحيوي تتبع التوزيع الطبيعي بمتوسط **80** درجة، وتباين **81** درجة. والمطلوب:

1- كتابة قيمة معالم التوزيع الاحتمالي للدرجات.

2- كتابة شكل دالة كثافة الاحتمال.

3- ما نسبة الطلبة الذين تقل درجاتهم عن 60 درجة؟

4- ما الدرجة التي أقل منها **0.975** من الدرجات؟

الحـل

1- كتابة قيمة معالم التوزيع الاحتمالي للدرجات.

نفرض أن X متغير عشوائي يعبر عن الدرجات، وهو يتبع التوزيع الطبيعي، ومعالمه هي: المتوسط $E(x) = \mu = 80$ والتباين هو: $var(x) = \sigma^2 = 81$ أي أن المتغير العشوائي يكون كالاتي: $X \sim N(80{,}81)$

2- شكل دالة كثافة الاحتمال

$$f(x) = \frac{1}{9\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-80}{9}\right)^2}, \quad -\infty < \boldsymbol{x} < \infty$$

3- نسبة الطلبة الذين تقل درجاتهم عن60 درجة هي:

$Pr(X < 60)$ ونتبع الخطوات المذكورة سابقا في حساب الاحتمال كما يلي:

$$Pr(x < 60) = Pr\left(z < \frac{x-\mu}{\sigma}\right)$$

$$= Pr\left(z < \frac{60-80}{9}\right) = Pr(z < -2.22) = F(-2.22)$$

وبالكشف مباشرة عن هذه القيمة في جدول التوزيع الطبيعي القياسي في الملحق 1، نجد أن

$Pr(x < 60) = Pr(z < -2.22) = 0.0132$

وبالتالي تكون نسبة الطلبة الذين تقل درجاتهم عن 60 درجة تساوي 1.32%.

4- الدرجة التي أقل منها (0.975) من الدرجات: في هذه الحالة يبحث عن قيمة المتغير (x) الذي أقل منه (0.975)، بفرض أن هذا المتغير هو x_1، فإن:

$$Pr(x < x_1) = Pr\left(z < \frac{x_1 - 80}{9}\right) = 0.975$$

بالكشف بطريقة عكسية، حيث نبحث عن المساحة 0.975 نجدها تقع عند تقاطع الصف 1.9، والعمود 0.06 أي أن قيمة $z = 1.96$، ويكون:

$1.96 = \frac{x_1 - 80}{9}, \quad then$

$x_1 = 9 \times 1.96 + 80 = 97.64$

وعليه فان الدرجة التي تحقق الاحتمال هي 97.64 درجة.

10-3-2 حساب احتمالات التوزيع الطبيعي باستعمال برنامج (SPSS)

لاحظنا في الأمثلة السابقة لحساب الاحتمال للمتغيرات التي تتبع التوزيع الطبيعي، أولا تحويل قيم المتغير الى القيم المعيارية حسب الصيغة (10-5) ومن ثم اما استعمال الجدول للتوزيع الطبيعي المعياري او استعمال البرنامج (**SPSS**) لحساب الاحتمالات كما في المثال الاتي.

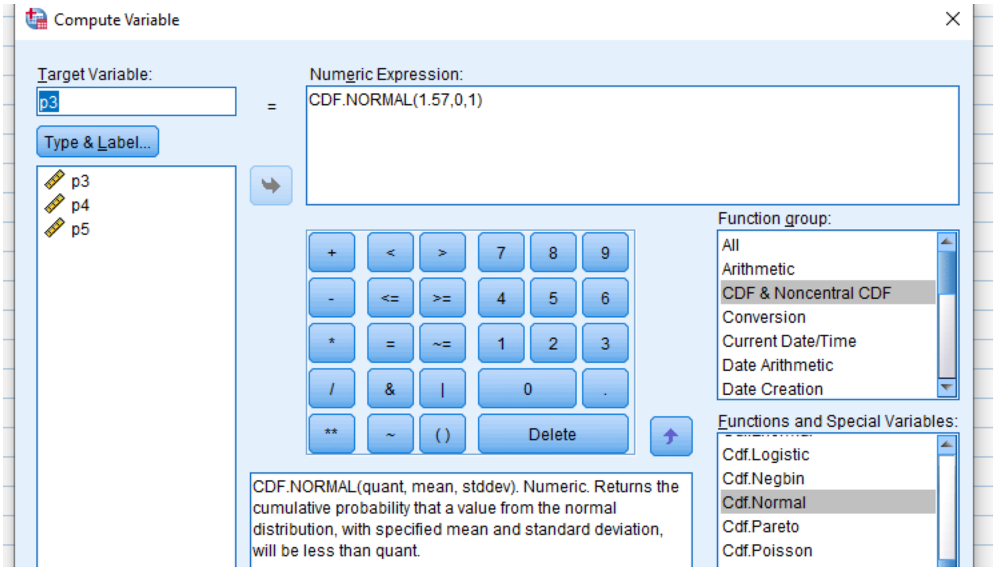
<u>مثال 10</u>

استعمل البيانات في المثال 6 لحساب الاحتمالات باستعمال البرنامج (**SPSS**).

<u>الحل:</u>

أ- نستعمل لحساب الاحتمال ($Pr(Z < 1.57)$) الخطوات نفسها من 1 الى 3 في حالة استعمال البرنامج لحساب الاحتمال للتوزيع ذي الحدين، وبعد ان

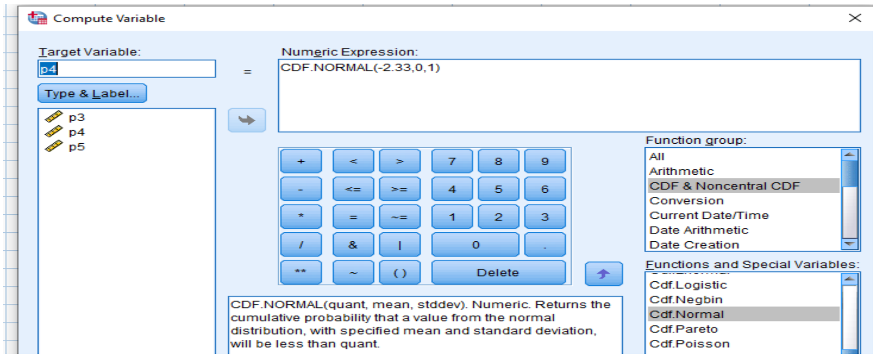
يكون لدينا الصندوق الحواري نختار (cdf) من المربع تحت عنوان (Function group) ومن المربع الذي يتضمن التوزيعات (CDF.Normal) كما في الشكل $(21-10)$



الشكل $(21-10)$ المربع الحواري لاختيار الدالة والتوزيع

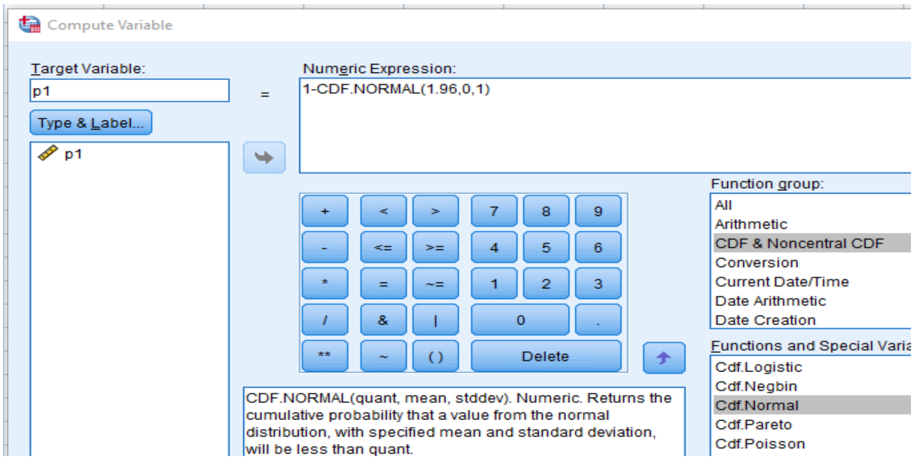
وبعد استبدال علامات الاستفهام بالقيم المناسبة حيث تمثل الأولى القيمة المعيارية المطلوب حساب الاحتمال لها وهي (1.57) والثانية تمثل المتوسط ويساوي صفراً للتوزيع الطبيعي المعياري والقيمة الثالثة تمثل الانحراف المعياري للتوزيع الطبيعي المعياري وتساوي 1 وبعد ذلك نختار الامر (OK) نحصل على قيمة الاحتمال المطلوب حسابه $(Pr(Z < 1.57 = 0.9418)$.

ب-تستعمل الخطوات المذكورة انفاً نفسها لحساب الاحتمال $(\Pr(Z < -2.33)$ كما في الشكل $(22-10)$ ونجد بانها تساوي (0.0099) وهي نفس القيم التي تم ايجادها في الحل اليدوي من الجداول للتوزيع الطبيعي المعيارية.



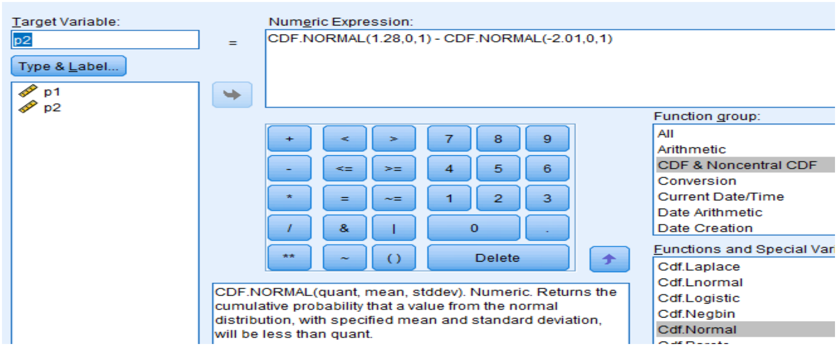
الشكل (10−22) إيجاد الاحتمال للقيمة (-2.33)

ت− نتبع نفس الخطوات لحساب الاحتمال ($\Pr(Z > 1.96)$)، ولكن نحول الاحتمال الى الصيغة الاتية ($1 - \Pr(Z < 1.96)$) ثم نستعمل مربع الحوار كما في الشكل (10−23) فتكون قيمة الاحتمال تساوي (0.025).



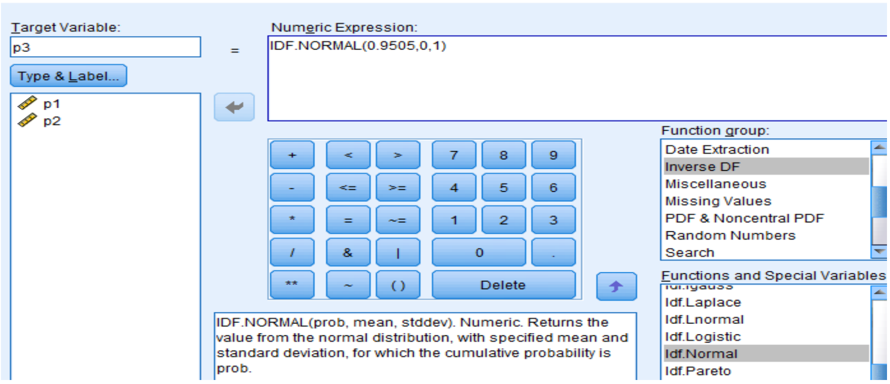
الشكل (10−23) حساب الاحتمال ($Pr(Z > 1.96)$)

ث−نتبع الخطوات نفسها لحساب الاحتمال ($\Pr(-2.01 < Z < 1.28)$) بعد تحويل الصيغة الى الشكل الاتي ($\Pr(Z < 1.28) - \Pr(Z < -2.01)$) ثم نستعمل مربع الحوار نفسه لنحصل على قيمة الاحتمال تساوي (0.8775) كما في الشكل (10−24).



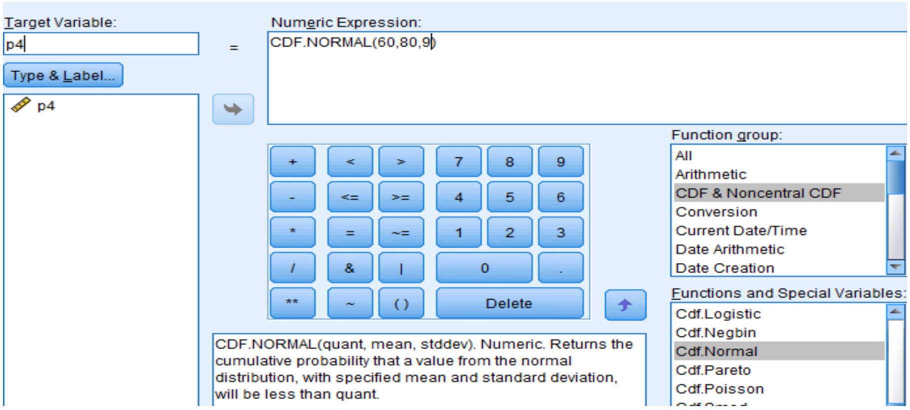
الشكل (24−10) حساب الاحتمال $Pr(-2.01 < Z < 1.28)$()

ج− نتبع الخطوات نفسها لإيجاد قيمة (Z) المقابلة لقيمة الاحتمال (0.9505) ولكن نستعمل معكوس الدالة التجميعية (Invers CDF) للتوزيع الطبيعي كما في الشكل (25−10) فتكون قيمة (Z=1.6497).



الشكل (25−10) حساب قيمة (Z)

ملاحظة: مراعاة للاختصار وعدم التكرار فان حسابات الاحتمالات والخطوات السابقة يمكن تطبيقها على القيم الاصلية للظاهرة دون الحاجة الى تحويلها الى القيم المعيارية. فعلى سبيل المثال إذا كانت الظاهرة تمثل الوزن والذي يخضع للتوزيع الطبيعي بمتوسط 80 كغم وتباين 81 كغم تربيع ما هو احتمال ان يكون الشخص وزنه اقل من 60 كغم. نستعمل الخطوات نفسها لحساب الاحتمال $(Pr(X < 60))$ كما في الشكل (26−10). فتكون قيمة الاحتمال تساوي (0.0131).



الشكل (10−26) حساب الاحتمال من القيم الاصلية للمتغير

وبشكل عام يعد التوزيع الطبيعي من أهم التوزيعات المستمرة وذلك لأن كثيراً من الظواهر الطبيعية تخضع لهذا التوزيع كما أن كثير من الظواهر الطبيعية التي لا تخضع لهذا التوزيع يمكن أن يقرب توزيعها بالتوزيع الطبيعي اعتمادا على نظرية النهاية المركزية. كذلك للتوزيع الطبيعي دور مهم آخر يتمثل في اشتقاق توزيعات احتمالية متعلقة بالمعاينة وذلك عند سحب عينة صغيرة حجمها (n<30) تسحب عادة من مجتمع طبيعي ومن هذه التوزيعات توزيع t والمبحث الاتي يتضمن نبذة عن هذا التوزيع.

3.3.10 توزيع t   t− Distribution

إذا كان X متغيراً عشوائياً له دالة الكثافة الاحتمالية الاتية:

$$f(t) = \left.\begin{array}{ll} \dfrac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)\sqrt{v\pi}}\left(1+\dfrac{t^2}{v}\right)^{\frac{-(v+1)}{2}} \sim t(v), & -\infty < t < \infty \\ = Zero & o/w \end{array}\right]$$

(7.10)

نقول إن المتغير يتبع توزيع t بدرجة حرية $v = n - 1$ والتي بدورها هي معلمة هذا التوزيع، ويلعب هذا التوزيع دورا مهماً عند سحب عينات صغيرة الحجم في مجتمع مجهول التباين. وهذا التوزيع يشبه لحد كبير التوزيع الطبيعي ويستعمل عوضا عن

التوزيع الطبيعي في حالة عدم معرفة الانحراف المعياري (σ) للمجتمع او حين يكون حجم العينة اقل من 30 اي ان ($n < 30$). والجدول الاتي يبين بعض المؤشرات الاحصائية للتوزيع.

جدول رقم (7.10) : خصائص توزيع t

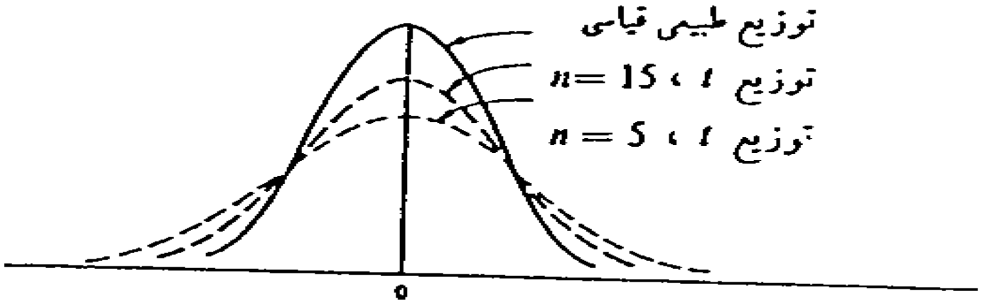
Mean (الوسط الحسابي)	$\mu = 0$
Variance (التباين)	$\sigma^2 = \frac{n}{n-2}, \quad n > 2$
Standard deviation (الانحراف المعياري)	$\sigma = \sqrt{\frac{n}{n-2}}$

ويشبه توزيع (t) التوزيع الطبيعي من حيث بعض الخصائص حيث له منحنى تكراري متماثل حول الوسط الحسابي، ولكنه أكثر امتدادا عند الأطراف من المنحنى الطبيعي. وهنالك معامل واحد يحدد شكل التوزيع يمثل درجات الحرية. وتعرف درجات الحرية على انها عدد المشاهدات المستقلة التي تستخدم في تعريف دالة الكثافة الاحتمالية للتوزيع او في تقدير معالم التوزيع. ولحساب الاحتمال نستعمل جدول توزيع (t) في الملحق (2) والمتغير المعياري الاتي:

$$t = \frac{(\bar{x}-\mu)}{s/\sqrt{n}} \sim t(n-1) \qquad (8.10)$$

حيث ان $\bar{x}$ يمثل الوسط الحسابي و s يمثل الانحراف المعياري للعينة. ولهذا التوزيع عدد $n-1$ من درجات الحرية. وتأتي أهمية هذا التوزيع من إمكانية استعماله في حالات العينات صغيرة الحجم التي تعتبر وضع واقعي كثير الحدوث في الدراسات الحيوية إذ ان المحددات العملية مثل قلة الامكانات وعدم توافر المستلزمات المطلوبة لاختيار العينات الكبيرة تجعل تصغير حجم العينة امرا ضرورياً. وهناك جداول لاحتمالات توزيع المتغير t توضح الاحتمالات لمستويات متعدده من قيمة t ولقيم مختلفة من درجات الحرية $n-1$ كما في الملحق (2). والشكل الاتي يوضح منحنى

توزيع t لدرجات حرية 5 و15 ومدى اقترابه من التوزيع الطبيعي كلما زادت درجات الحرية.



شكل (27.10) يمثل توزيع t حسب درجات حرية ومدى اقترابه من التوزيع الطبيعي المعياري

ونعبر عن قيمة المتغير العشوائي الذي يتبع توزيع t بالرمز $t_{(v,\alpha)}$ وهي القيمة التي يقع على يمينها مساحة α على منحى توزيع بدرجة حرية v وتحسب قيم هذا المتغير العشوائي والاحتمالات له باستعمال جدول (2) المرفق في نهاية الكتاب مع مراعاة القواعد الاتية وذلك بالاعتماد على تماثل منحنى التوزيع حول الصفر:

$$\left.\begin{array}{c} t(\alpha,v) = -t(1-\alpha,v) \\ Pr(t<-A) = Pr(t>A) \\ Pr(t<A) = 1-Pr(t>A) \end{array}\right]$$

(9.10)

ولتوضيح ذلك نأخذ الامثلة الاتية:

مثال 11

اوجد الاحتمالات الاتية: $t_{(0.995,9)}$، $t_{(0.99,11)}$، $t_{(0.05,3)}$

الحل: بالاعتماد على جدول الملحق رقم (2):

1- لإيجاد الاحتمال المقابل للقيمة $t_{(0.05,3)}$ نختار من العمود الايسر (عمود درجة الحرية df) القيمة 3 فنجد على يمينه صفاً من الاعداد ومن الصف العلوي (صف المساحات أي الاحتمالات α) نختار القيمة 0.05 فنجد أسفلها عمود من الاعداد فتكون القيمة الموجودة في تقاطع الصف والعمود هي القيمة المطلوبة وبالتالي فان الاحتمال يكون $t_{(0.05,3)} = 2.353$

2- لإيجاد الاحتمال ($t_{(0.99,11)}$) نلاحظ أن جدول التوزيع لا يحتوي على القيمة لذلك نستخدم القاعدة رقم 1 في المعادلة (10−9)، وعليه كما في المطلوب الاول من جدول توزيع t مباشرة يكون الاحتمال كما يأتي:

$$t_{(0.99,11)} = -t_{(0.01,11)} = -2.718$$

3- نستعمل الخطوات نفسها في المطلوب 2 لإيجاد الاحتمال المقابل للقيمة ($t_{(0.995,9)}$) فيكون كالاتي:

$$t_{(0.995,9)} = -t_{(0.005,9)} = -3.250$$

مثال 12

إذا كان $X \sim (8.5, \sigma^2)$، اوجد التوزيع الاحتمالي لمتوسط العينة الاتية (9,10,9,8,8,7,8,5,8) ثم احسب الاحتمال الاتي: $Pr(7 < \bar{x} < 9)$

الحل

بمان ان تباين مجتمع X مجهول ان حجم العينة n = 9 وهو اقل من 30 يكون توزيع الوسط الحسابي للعينة هو $\bar{x} \sim t(8)$، ولحساب الاحتمال المطلوب اذن يجب حساب متوسط وتباين العينة كالاتي:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{9+10+\cdots+8}{9} = 8,$$

$$S^2 = \frac{\sum (x_i-\bar{x})^2}{n-1} = \frac{(9-8)^2+\cdots+(8-8)^2}{9-1} = 2$$

وبالتالي يكون التوزيع الاحتمالي للمتغير $T=\frac{\bar{x}-8.5}{1.414/\sqrt{8}}\sim t(8)$ ويكون الاحتمال المطلوب كالاتي:

$$Pr(7<\bar{x}<9)=\frac{7-8.5}{1.414/\sqrt{8}}<\frac{\bar{x}-8.5}{1.414/\sqrt{8}}<\frac{9-8.5}{1.414/\sqrt{8}}$$

$$=Pr(-3<t<1)$$

$$=Pr(t<1)-Pr(t<-3)$$

وحسب القاعدة في المعادلة رقم (9.10) فان $Pr(t<-3)=Pr(t>3)$، وكذلك فان$Pr(t<1)=1-Pr(t>1)$ وبالتالي فان الاحتمال المطلوب يساوي:

$$Pr(7<\bar{x}<9)=1-Pr(T>1)-Pr(T>3)$$

$$=1-0.15-0.005=0.845$$

إن توزيع t له استعمالات كثيرة منها:

1-اختبار الفرضية حول متوسط مجتمع طبيعي مجهول التباين.

2-اختبار الفرضية حول الفرق بين متوسطي مجتمعين طبيعيين مجهولي التباين.

3-اختبار الفرضية حول معنوية معامل الارتباط البسيط.

4-اختبار الفرضية حول معنوية معاملات الانحدار في الانحدار الخطي المتعدد.

5-اختبار الفرضية حول معنوية معامل الارتباط الجزئي.

6-تكوين فترات الثقة لمتوسط مجتمع طبيعي مجهول التباين. وغيرها من الاستعمالات

<u>ملاحظة:</u> مراعاة للاختصار وعدم التكرار يمكن اتباع الخطوات نفسها في حساب الاحتمالات عندما يكون التوزيع طبيعياً لحساب الاحتمالات عندما يكون التوزيع (t). باستعمال البرنامج (SPSS) وذلك بتغيير التوزيع فقط من التوزيع الطبيعي الى توزيع (t).

تمارين الفصل العاشر

1- اذكر أي من الجداول التكرارية الاتية يمثل جدولاً توزيعياً احتمالياً وأيا منها لا يمثل ذلك، ولماذا؟

X	2	3	4	5
f(x)	1/8	3/8	5/8	−1/8

X	−1	0	1	2
f(x)	1/5	4/10	2/10	1/5

X	11	13	15	17	19
f(x)	1/7	1/7	3/7	2/7	2/7

2- اوجد قيمة الثابت (b) التي تجعل الجدول الاتي يمثل توزيعا احتماليات متقطعا.

X	2	3	4	5	6
f(x)	1/8	2/8	3/8	b	1/8

3- إذا علمت من التجارب السابقة بان 60% من الفئران التي تم تلقيحها بمصل معين هي محمية من المرض. وإذا اخترنا 5 فئران ملقحة باللقاح فما هو احتمال:

أ- 3 بالضبط ب - على الأقل 3 ت - على الأكثر 3 سوف تكون محمية من المرض.

4- إذا كانت دالة الاحتمال للمتغير العشوائي المتقطع (X) كما يأتي:

$$f(X) = \begin{cases} \frac{1}{k} & X = 1, 2,3,4,5 \\ 0 & OW \end{cases}$$

أ- اوجد قيمة (k)

ب- احسب الاحتمالات الاتية

$$\Pr(X = 6), \quad \Pr(0 < X \leq 2.5)\,; \quad \Pr(X \geq 4)\,; \quad \Pr(1 < X \leq 3)$$

5- إذا علمت بان عمر المرضى الراقدين في أحد المستشفيات الخاصة (X) يخضع للتوزيع الطبيعي بمتوسط 69 سنة وتباين 62 سنة، ماهي احتمالات ان (X) ستكون:

أ- اقل من 65 سنة؟

ب- أكبر من 60 سنة؟

ت- بين 70 و75 سنة؟

ث- بين 65 و80 سنة؟

ج- أكبر من 90 سنة؟

6- من المتعارف عليه بان ضغط الدم بين مجموعة العاملين (X) يخضع للتوزيع الطبيعي بمتوسط 122(mmHg) وانحراف معياري 6 (mmHg)، اختير شخص عشوائيا من هذا المجتمع، احسب احتمال ان ضغط الدم (X) له:

أ- اقل من 116 (mmHg)؟

ب- أكثر من 130 (mmHg)؟

ت- بين 113 و126 (mmHg)؟

7- اذا علمت بان كل طفل يولد لمجموعة من الاباء يكون بنسبة (0.25) يمتلك صنف (O) من الدم فاذا كان لهؤلاء الاباء 5 اطفال فما هو احتمال ان يكون:

أ- بالضبط طفلين منهم لهم فصيلة الدم O نفسها؟

ب- على الاقل طفلين منهم لهم فصيلة الدم O نفسها؟

ت- ما هو توقع وتباين العدد للأطفال اللذين لهم فصيلة الدم O نفسها؟

8- طلب من أحد منتجي المقاعد المدرسية أن يضع في كل غرفة دراسة بعض المقاعد الخاصة بالأعسرين) الايسرين. إذا كان يعلم أن % 10 من السكان أعسرين ما هو احتمال ان غرفة لعشرين طالبا:

أ- لا تحتوي على أي عسر؟

ب- اثنان منهم اعسران؟

ت- أكثر من أثنين منهم أعسرين؟

9- إذا علمت بان (X) تخضع للتوزيع الطبيعي واعطيت لك القيم المعيارية الاتية وتوزيعاتها احسب قيمة المتغير الاصلي لكل منها:

a- $Z = 2.324 \quad X \sim N(45;\ 3^2)$

b- $Z = -0.532 \quad X \sim N(0.75;\ 0.25^2)$

c- $Z = \pm 1.825 \quad X \sim N(10;\ 1.5^2)$

10- إذا كان العمر(X) لمجموعة من الاشخاص يتبع توزيع (t) وكان عددهم (11) شخص، اوجد احتمال ان العمر (X) يساوي:

أ- (t) اقل من 3.25؟

ب- (t) أكبر من 1.28؟

ت- (t) بين (-2.82) و (-1.28)؟

11- إذا كان المطلوب اختيار مريضين بطريقة عشوائية لإجراء العملية الجراحية من بين 4 مرضى ذكور و4 مريضات اناث، وكان المتغير العشوائي X يمثل عدد المرضى الذكور الذين سيتم اختيارهم. والمطلوب:

أ- أوجد التوزيع الاحتمالي لهذا المتغير العشوائي، ودالة توزيعه التراكمي.

ب - أوجد التوقع الرياضي والانحراف المعياري لهذا المتغير.

12- إذا كانت نسبة المصابين بعمى الألوان في مجتمع ما هي 15% اوجد:

أ- التوزيع الاحتمالي لعدد المصابين بعمى الألوان في اسرة مكونة من 4 افراد من هذا المجتمع.

ب-احتمال ان يكون عدد غير المصابين بعمى الألوان أكبر من 3 واقل من 4 في هذه الاسرة.

ت- ما هو متوسط التوزيع الاحتمالي وتباينه في الفقرة أ.

13- وجد في احدى الدراسات ان 20% من الناس يستعملون اليد اليسرى في الكتابة، سحبت عينة من عشرة اشخاص من هذا المجتمع، اوجد:

أ- احتمال شخصين يستعملان اليد اليسرى.

ب- احتمال على الأقل اثنان يستعملان اليد اليسرى.

ت- احتمال على الأكثر ثلاثة يستعملون اليد اليسرى.

ث- احتمال 8 اشخاص على الأكثر يستعملون اليد اليمنى.

14- بينت دراسة صحية ان نسبة العمال المصابين بمرض الربو من بين العاملين في مركز صحي هي 2%، سحبت عينة عشوائية حجمها 1500 شخص من العاملين. اوجد العدد المتوقع لغير المصابين بمرض الربو.

15- إذا كان (X) هو عدد الحالات الطارئة التي يستقبلها احد المستشفيات خلال ليلة واحدة متغير عشوائي له توزيع بواسون مع $(\mu = 3)$ فأوجد:

أ- التوزيع الاحتمالي للمتغير العشوائي.

ب-احتمال ان يستقبل المستشفى 4 حالات طارئة خلال ليلة واحدة.

ت-احتمال ان يستقبل المستشفى 4 حالات طارئة خلال 3 ليالي.

ث-احتمال ان يستقبل المستشفى اقل من حالتين طارئتين في ليلة واحدة.

ج- متوسط عدد الحالات التي يستقبلها المستشفى خلال ليلة واحدة.

ح- التباين والانحراف المعياري لعدد الحالات الطارئة التي يستقبلها المستشفى خلال ليلة واحدة.

Appendices المـلاحــــق

ملحق رقم 1 جدول الاحتمالات للتوزيع الطبيعي المعياري (أ)

Standard Normal Distribution Probabilities Table (A)

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
−3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
−3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
−3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
−3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
−2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
−2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
−1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
−1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
−1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
−1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
−1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
−1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
−1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
−1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
−1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
−0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
−0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
−0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
−0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
−0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
−0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
−0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
−0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
−0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
−0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

ملحق رقم 1 جدول الاحتمالات للتوزيع الطبيعي المعياري (ب)

Standard Normal Distribution Probabilities Table (B)

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

ملحق 2 القيم الاحتمالية لتوزيع (t)

Student t Distribution Probabilities Table

one-tail area	0.25	0.125	0.1	0.075	0.05	0.025	0.01	0.005	0.0005
two-tail area	0.5	0.25	0.2	0.15	0.1	0.05	0.02	0.01	0.001
confidence level	0.5	0.75	0.8	0.85	0.9	0.95	0.98	0.99	0.999
d.f. 1	1.000	2.414	3.078	4.165	6.314	12.706	31.821	63.657	636.619
2	0.816	1.604	1.886	2.282	2.920	4.303	6.965	9.925	31.599
3	0.765	1.423	1.638	1.924	2.353	3.182	4.541	5.841	12.924
4	0.741	1.344	1.533	1.778	2.132	2.776	3.747	4.604	8.610
5	0.727	1.301	1.476	1.699	2.015	2.571	3.365	4.032	6.869
6	0.718	1.273	1.440	1.650	1.943	2.447	3.143	3.707	5.959
7	0.711	1.254	1.415	1.617	1.895	2.365	2.998	3.499	5.408
8	0.706	1.240	1.397	1.592	1.860	2.306	2.896	3.355	5.041
9	0.703	1.230	1.383	1.574	1.833	2.262	2.821	3.250	4.781
10	0.700	1.221	1.372	1.559	1.812	2.228	2.764	3.169	4.587
11	0.697	1.214	1.363	1.548	1.796	2.201	2.718	3.106	4.437
12	0.695	1.209	1.356	1.538	1.782	2.179	2.681	3.055	4.318
13	0.694	1.204	1.350	1.530	1.771	2.160	2.650	3.012	4.221
14	0.692	1.200	1.345	1.523	1.761	2.145	2.624	2.977	4.140
15	0.691	1.197	1.341	1.517	1.753	2.131	2.602	2.947	4.073
16	0.690	1.194	1.337	1.512	1.746	2.120	2.583	2.921	4.015
17	0.689	1.191	1.333	1.508	1.740	2.110	2.567	2.898	3.965
18	0.688	1.189	1.330	1.504	1.734	2.101	2.552	2.878	3.922
19	0.688	1.187	1.328	1.500	1.729	2.093	2.539	2.861	3.883
20	0.687	1.185	1.325	1.497	1.725	2.086	2.528	2.845	3.850
21	0.686	1.183	1.323	1.494	1.721	2.080	2.518	2.831	3.819
22	0.686	1.182	1.321	1.492	1.717	2.074	2.508	2.819	3.792
23	0.685	1.180	1.319	1.489	1.714	2.069	2.500	2.807	3.768
24	0.685	1.179	1.318	1.487	1.711	2.064	2.492	2.797	3.745
25	0.684	1.198	1.316	1.485	1.708	2.060	2.485	2.787	3.725
26	0.684	1.177	1.315	1.483	1.706	2.056	2.479	2.779	3.707
27	0.684	1.176	1.314	1.482	1.703	2.052	2.473	2.771	3.690
28	0.683	1.175	1.313	1.480	1.701	2.048	2.467	2.763	3.674
29	0.683	1.174	1.311	1.479	1.699	2.045	2.462	2.756	3.659
30	0.683	1.173	1.310	1.477	1.697	2.042	2.457	2.750	3.646
35	0.682	1.170	1.306	1.472	1.690	2.030	2.438	2.724	3.591
40	0.681	1.167	1.303	1.468	1.684	2.021	2.423	2.704	3.551
45	0.680	1.165	1.301	1.465	1.679	2.014	2.412	2.690	3.520
50	0.679	1.164	1.299	1.462	1.676	2.009	2.403	2.678	3.496
60	0.679	1.162	1.296	1.458	1.671	2.000	2.390	2.660	3.460
70	0.678	1.160	1.294	1.456	1.667	1.994	2.381	2.648	3.435
80	0.678	1.159	1.292	1.453	1.664	1.990	2.374	2.639	3.416
100	0.677	1.157	1.290	1.451	1.660	1.984	2.364	2.626	3.390
500	0.675	1.152	1.283	1.442	1.648	1.965	2.334	2.586	3.310
1000	0.675	1.151	1.282	1.441	1.646	1.962	2.330	2.581	3.300
infinity	0.674	1.150	1.282	1.440	1.645	1.960	2.326	2.576	3.291

قائمة المراجـــــع

اولا: المراجـــع العربيـــة

ابو صالح، د. محمد صبحي، عدنان عوض، (1983) **مقدمة في الاحصاء**، دار جون وايلي وابناؤه، عمان.

الجبوري، شلال، (1990) **الاحصاء التطبيقي**، مطابع الجامعة المستنصرية، بغداد.

حسين، جاسم ناصر، (2019) **الإحصاء الحيوي التطبيقي**، ط1، مكتب الجزيرة للطباعة والنشر، بغداد.

خليل، شرف الدين، (بدون سنة نشر) **الاحصاء الوصفي**، شبكة الابحاث والدراسات الاقتصادية، WWW.RR4EE.NET

خواجة، خالد زهدي، (بدون سنة نشر) **اساسيات الاحتمالات**، المعهد العربي للتدريب والبحوث الاحصائية،

http://www.uobabylon.edu.iq/eprints/paper_12_15272_1049.pdf

الراوي / خاشع محمود، (1984) **المدخل الى الاحصاء**، دار الكتب، الموصل.

رشيد، ظافر حسين، المشهداني، كمال علوان خلف، (2016) **الاحصاء للتخصصات الادارية والمحاسبية**، مكتب الجزيرة للطباعة والنشر، بغداد.

رشيد، محمد حسين، (2008) **الاحصاء الوصفي والتطبيقي والحيوي**، دار صفاء للنشر والتوزيع، عمان.

رمضان، زياد، (2001) **مبادئ الاحصاء الوصفي والتطبيقي**، دار صفاء للنشر والتوزيع، عمان.

زغلول، يحيى سعد، (1988) **مقدمة في الاحصاء التطبيقي**، منشورات الدار الجامعية بيروت.

سالفاتور، دومينيك، ترجمة، منتصر، سعدية حافظ، (1982) **نظريات ومسائل في الاحصاء والاقتصاد القياسي**، ملخصات سلسلة شوم، الدار الدولية للنشر والتوزيع.

الشيحة، عبد الله، (بدون سنة نشر) مبادئ الإحصاء والاحتمالات (1) (المتغيرات العشوائية والتوزيعات الاحتمالية)، مجموعة محاضرات على الموقع الالكتروني: www.faculty.ksu.edu.sa

صالح، ابو عبد الله، (بدون سنة نشر) **مدخل الى الاحتمالات والاحصاء الرياضي** ، ج2 ، الإحصاء.

صالحين، صلاح العيادي، (بدون سنة نشر) مسائل وحلول في الإحصاء والاحتمالات، مجموعة تمارين على الموقع الالكتروني: www.utebooks.com/GS206

طبية، احمد عبد السميع، (2008) **مبادئ الاحصاء**، دار البداية ناشرون وموزعون، عمان.

العباسي، عبد الحميد محمد، (2013) **الاسلوب الاحصائي: التحليل والتفسير باستخدام الحاسب وبرنامج SPSS**، معهد الدراسات والبحوث الاحصائية، جامعة القاهرة، القاهرة.

العتوم، شفيق، العاروري، فتحي، (1995) الاساليب الاحصائية، دار المناهج، عمان.

اللافي، سعد مؤمن، (2003) **الاحصاء الاستنتاجي**، اكاديمية الدراسات العليا، طرابلس.

المشهداني، كال علوان خلف، (2010) **تصميم وتحليل التجارب**، مكتب الجزيرة للطباعة والنشر، بغداد.

نشوان، عماد، (2005) **الدليل العملي لمقررالاحصاء التطبيقي**، جامعة القدس المفتوحة، القدس.

النعيمي، محمد عبد العال، وطعمة، حسن ياسين، (2008) **الاحصاء التطبيقي**، دار وائل للنشر والتوزيع، عمان.

الهوبي، اياد محمد، (2014) **الاحصاء التطبيقي**، الكلية الجامعية للعلوم والتكنولوجيا، خان يونس.

ثانيا: المراجـــــع الاجنبيـــــة

Agresti, A., Franklin, C.A., Klingenberg, B., (2023) **Statistics, the Art and Science of Learning from Data**, 5th Ed., Pearson Education Limited.

Agresti, A., (2007) ***An Introduction to Categorical Data Analysis***,2nd Ed., John Wiley & Sons, Inc.

Agresti, A., (1996) ***An Introduction to Categorical Data Analysis***, New York, John Wiley & Sons.

Armitage, P., Berry, G., Matthews, J.N.S. (2002) *Statistical Methods in Medical Research*, 4th Ed., Blackwell Science.

Bury, K.V., (1975) ***Statistical Models in applied science***, Willy, New York.

Canavas, G.C., (1984) ***Applied probability and Statistical methods***, Little and Brown company, Toronto.

Casella, G., and Berger, R.L., (2002) ***Statistical Inference***, 2nd ed., Wadsworth, Inc., USA.

Chinna, K. and Krishnan, K., (2009) ***Biostatistics for the Health Sciences***, McGraw-Hill Sdn. Bhd., Malaysia.

Conover, W.J., (1980) ***Practical Nonparametric Statistics***, 2nd ed., John Wiley & Sons.

Conover, W.J., (1999) ***Practical Nonparametric Statistics***. 3rd ed., New York: John Wiley & Sons.

Conover, W.J. and Iman, R.L., (1976) On some alternative procedures using ranks for the analysis of experimental designs. *Commun. Statist.*, A5, 1349–1368.

COX, D.R., (2006) ***Principles of Statistical Inference***, Cambridge University Press, USA.

Day, S. and Talbot, D.J., Editorial, (2000) **Statistical guidelines for clinical trials**. *J. Roy. Statist. Soc. A.*, 163, 1–3.

Devore, J.L., (1995) ***Probability and Statistics for Engineers and Scientists***, Wadsworth, Inc., USA.

Dobson, A.J., (2002) ***An introduction to generalized linear models***, Chapman & Hall/CRC, USA.

Dudewicz and Mishra, (1998) ***Modern Mathematical Statistics***, Wiley.

Everitt, B.S. & Pickles, A. (2004) ***Statistical Aspects of the Design and Analysis of Clinical Trials****, (revised edition),* Imperial College Press.

Garth, A., (2008) **Analysing data using SPSS,** Sheffield Hallam University, https://students.shu.ac.uk/lits/it/documents/pdf/

Hayter, A.J., (1996) ***Probability and Statistics***, PWS Publishing Company, USA.

Hines, W.W., and Montgomery, D.C., (1990) ***Probability and Statistics in Engineering and Management Science***, John Wiley &Sons Inc. USA.

Hoel, P.G., (1971) ***Introduction to mathematical statistics***, 4th ed., Wiley, New York, USA.

Hogg,R.V. and Caraig, A.T., (1978) ***Introduction to Mathematical statistics***, 4th ed., MacMillon, New York, USA.

Kroshnamoorthy, K (2006) *Handbook of Statistical Distributions with Applications,* Chapman and Hall/CRC.

Landau, S. and Everitt, B. S., (2004) A Handbook of statistical Analyses Using SPSS, Chapman & Hall/CRC press LLC, New York.

Lindgren,B.W., (1976) ***Statistical theory***, 3rd Ed., MecMillon, New York. USA.

Lindsey, J.K., (1996) ***Applying Generalized Linear Models***, Springer - Verlag New York, 1996.

LE, C.T., (2003) ***Introductory Biostatistics***, John Wiley & Sons, Inc., USA.

Lohr, S. L., (without publishing year) ***Sampling: Design and Analysis***, Duxbury Press. USA.

Machin, D, Campbell, MJ and Walters, SJ (2007) *Medical Statistics: A textbook for the Health Sciences,* 4th Ed., Wiley.

Mann, P.S., (1998) ***Introduction Statistics***, John Willy &Sons, Inc. USA.

Marques de Sá, J.P., (2007) ***Applied Statistics using SPSS, STATISTICA, MATLAB and R***, Springer-Verlag Berlin Heidelberg.

Milton J., and Arnold, J., (1995) ***Introduction to Probability and Statistics***, McGraw-Hill.

Mood A. M., & others, (1974) ***Introduction to the theory of Statistics***, McGraw –Hill, Inc., Singapore.
Montgomery, D.C., Runger, G.C., (1994) ***Applied Statistics and Probability for Engineers***, John Wiley & Sons. In., USA.
Mukhopadhyay N., (2000) **Probability *and Statistical Inference***, Dekker.
Pallant. J., (2007) SPSS Survival Manual, A step by step Guide to data Analysis using SPSS for Windows, Open University Press, New York.
Pfeffermann, D. and Rao, C.R., (2009) ***Sample Surveys*: *Design, Methods and Applications***, Elsevier B.V., Hungary.
Rao, C.R., (1973) ***Linear Statistical Inference and its Applications,*** 2nd ed., John Wiley **&** Sons. Inc., USA.
Rosner, B., (2011) **Fundamentals of Biostatistics**, 7th ed., Brooks/cole, Cengage Learning.
Rowe, P., (2007) ***Essential Statistics for the Pharmaceutical Sciences***, John Wiley & Sons Ltd, UK.
Schervish, M., (1995) ***Theory of Statistics***, Springer-Verlag.
Snedeor and Cochran, (1980) ***Statistical Methods***, The Iowa State University Press, USA.
Soong, T.T., (2004) **Fundamental of Probability and Statistics for Engineers**, John Wiley & Sons Ltd, UK.
Sullivan, M., (2013) ***Statistics: Informed Decisions Using Data***, 4th ed., Pearson Education, Inc, USA.
Thompson, S.K., (2002) *Sampling.,* 2nd Ed., Wiley.

Wadsworth, H.M., (1998) ***Handbook of Statistical Methods for Engineers and Scientists***, McGraw- Hill Inc., USA.
Walck, C., (2007) ***Handbook on Statistical Distributions for Experimentalists***, Internal Report SUF–PFY/96–01, Stockholm.
Wassertheil-Smoller, S., (2004) ***Biostatistics and Epidemiology***, 3rd ed., Springer-Verlag New York, Inc., USA.
Weiss, L.A., (1994) *Elementary Statistics*, 2nd ed., Wesley published company.
Wegner, T., (2013) Applied Business Statistics, Methods and Excel-based Applications, Juta and company Ltd., South Africa.
Winner, L., (2004) *Introduction to Biostatistics*, University of Florida.