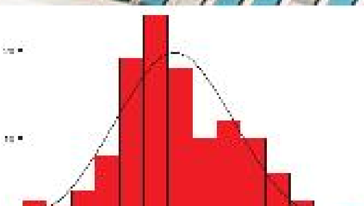
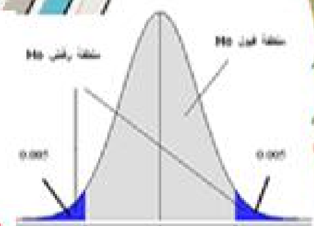
وزارة التعليم العالي والبحث العلمي
جامعة بغداد – كلية الزراعة
قسم المحاصيل الحقلية
المرحلة

المادة :. احصاء
مدرس المادة :.د.ناظم يونس عبد
رقم المحاضرة :.
العام الدراسي :. ٢٠١٧/٢٠١٦

<u>المحاضرات النظرية</u>

اساسيات علم الاحصاء





" وَكُل شَيئٍ أحْصَيْنَهُ كِتَباً "

صدق الله العظيم

سورة النبأ ، آية (٢٩)

الفهرس

الموضوع الصفحة

الفصل الاول

طبيعة علم الاحصاء

1.1 نبذة تاريخية عن علم الاحصاء

1.2 تعريف علم الاحصاء

1.3 علاقة علم الاحصاء بالعلوم الاخرى

1.4 تمارين

نبذة تاريخية عن علم الاحصاء

الإحصـاء، بمعنى العدّ والحصـر، فكرة قديمة يرجع منشـؤها إلى عهد بعيد في تاريخ المدنية الإنسـانية، فالحاجة إلى الحصـول على معلومات رقمية أو وصـفية عن المجتمعات وظروفها المادية وشـروط وجودها كانت حاجة ملحّة منذ أن وجدت المجتمعات البشـرية المنظمة، وهنالك إحصـاءات عند قدماء المصـريين والصـينيين والإغريق ، حيث أقتصـر اهتمام الحكومات منذ القدم بالمعلومات الاجتماعية وذلك لأغراض التنظيم والتخطيط ، واسـتخدم الإحصـاء في عصـره الأول في جمع البيانات عن السـكان وحصـرهم من قبل الدولة لأهداف معينة تتمثل في اسـتخدامهم في الجيوش أو توجيههم لتنفيذ بعض المباني أو لغرض فرض الضـرائب أو لتوزيع الأراضـي الزراعية على السـكان بطريقة عادلة ، ويعد قدماء المصريين أول من أستخدم هذا الأسلوب .

ولقد كان العرب المسلمون من أوائل من استعان بلغة الأرقام في إحصاء مواردهم وحصر غنائمهم وجندهم وأعطياتهم وأسلحتهم، ومعرفة الثروات لتحصيل الزكاة عنها. وكان لهم في الإحصاء اللغوي الباع الأطول فالكندي المتوفى سنة 260هـ، يصف في مؤلّفه «رسالة في استخراج المعمَّى» عملية إحصاء تواتر الحروف في لغة ما، وذلك بأخذ عينة كافية من الكلام المنثور في تلك اللغة، وقد أحصى نصاً مؤلفاً من ٣٦٦٧ حرفاً ثم استعمل تلك النتائج بعد ترتيبها في استنباط نص معمًّى وينبه فيها على أمر ذي بال، وهو أن النص المعمَّى ينبغي أن يكون ذا طول كافٍ يسمح بانطباق القواعد الإحصائية عليه، وهي فكرة رياضية على غاية من الأهمية، هي فكرة قانون الأعداد الكبيرة .

ولعل الكندي هو أول من أجرى ذلك الإحصاء في تاريخ الدراسات الكمية على اللغة، ولا شك في أنه أفاد من إحصائيات حروف القرآن الكريم التي سبقت عصره (وهي تعود إلى القرن الهجري الأول، وينسب بعضها إلى صدر الإسلام). كما كان للعرب في الإحصاء الاجتماعي أيضاً أثرٌ يجدر ذكره، وهو أن المفكر العربي ابن خلدون ربما كان أول من عالج قضايا السكان معالجة علمية، فبحث في عمران الدول واتساعها وتأخرها، وربط كل ذلك بنمو عدد السكان ونقصانهم .

وفي القرن السابع عشر والذي يمكن اعتباره العصر الإحصائي الثاني تم استخدام الطريقة الرقمية للدلالة على الظواهر موضـــوع البحث على اعتبار أن هذه الطريقة أدق وأقوى في التعبير عن هذه الظواهر وتركز الهدف من هذه الطريقة في معرفة عدد السكان وعدد المواليد وعدد الوفيات ومقدار الثروة والدخل ومقدار الضرائب المحصلة وكمية الناتج من المحاصيل الزراعية. وبذلك نشأت الحاجة الى تنظيم وتلخيص هذه البيانات ووضعها في صورة جداول او رسم بياني حتى يسهل الرجوع اليها والاستفادة منها باسرع وقت ممكن ، وقد اطلق على هذه الطرق " علم الدولة او علم الملوك ثم علم الاحصاء"

تاريخيا ارتبط ظهور نظرية الاحتمالات بألعاب الحظ والقمار التي كانت سائدة بكثرة في أوربا في القرن السابع عشر. لكن قلة انتشار طباعة الكتب والأجواء الدينية السائدة التي لا تبارك هذه الألعاب منعت انتشار الكتابات في هذا الشأن اما تطور علم الاحصاء في القرنين التاسع عشر والعشرين فقد تطور على ايدي :

- ادولف كيوتليت Adolph Quetelet (١٨٧٤ – ١٧٩٤) طبق علم الاحصاء بشكل فعال في علمي الاجتماع والتعليم.

- فرانسيس كالتون Francis Galton (١٨٢٢-١٩١١) الذي اشتهر بتطبيق علم الاحصاء في علوم الوراثة والتطور.
- كارل بيرسن karl Pearson (١٨٥٧-١٩٣٦) الفيزياوي الرياضي الذي عمل مع Galton في ايجاد نظرية الارتباط والانحدار والذي بذل اكثر من نصف قرن في الابحاث الاحصائية ، واسس مجلة Biometika التي لا زالت تصدر لحد الآن.
- فيشر R. A. Fisher (١٨٩٠-١٩٦٣) وهو من اشهر علماء القرن العشرين وقد اضاف الى الاحصاء وتطبيقاته هو وتلاميذه في علوم الزراعة والبايلوجي والوراثة والاقتصاد ووضع اسس تصميم وتحليل التجارب .

ان التطور السريع لعلم الاحصاء في هذا القرن يعود لسببين مهمين:

اولاً – ازدياد الطلب على استعمال الاحصاء في جميع العلوم.

ثانياً – تطور البرامجيات الاحصائية التي تستخدم في الحاسبات الالكترونية وسرعة انجاز العمليات الاحصائية وزيادة طاقتها الاستيعابية من البيانات.

ان كلمة الاحصاء (Statistics) مشتقة من عدة كلمات قديمة ، فهي مشتقة من الكلمة اللاتينية Status ، والانكليزية State ، والالمانية Statistik ، والايطالية Stato ، وكلها تعني شؤون الدولة او سياسة الدولة.

ان اصل علم الاحصاء وتطوره ينبع من مصدرين رئيسين هما :

١- السجلات الحكومية Governmental Records

ان كثير من الحضارات القديمة كالبابلية والاشورية والرومانية والفرعونية كانت لديهم سجلات تدون فيها بيانات لاغراض ادارة الدولة ، فاول تعداد سكاني مثلا اجره اليونانيون سنة ٤٣٥ قبل الميلاد. اما الاشوريون والفراعنة والرومان فقد وضعو احصائيات عن قوة الجيش ، الضرائب ، المواليد ، الوفيات ، وكميات الانتاج الزراعي وغيرها. وفي بداية القرن السادس عشر تم نشر عدة احصائيات عن الكوليرا في انكلترا.

٢- علم الرياضيات Mathematics

ان علم الاحصاء الاستنتاجي او الاستدلالي (Statistical Inference) يعتمد على نظرية الاحتمال التي تطورت في القرن السابع عشر على ايدي علماء الرياضيات نتيجة لانتشار لعب القمار بين نبلاء فرنسا وانكلترا. وقد جلب المقامرون انتباه اشهر علماء الرياضيات للحصول منهم على فرص الربح والخسارة ، وقد اكتشف كثير منهم التوزيعات الاحتمالية خلال تلك الفترة منها التوزيع الطبيعي ، وتوزيع برنولي ، وتوزيع ذي الحدين.

١,٢ تعريف علم الاحصاء

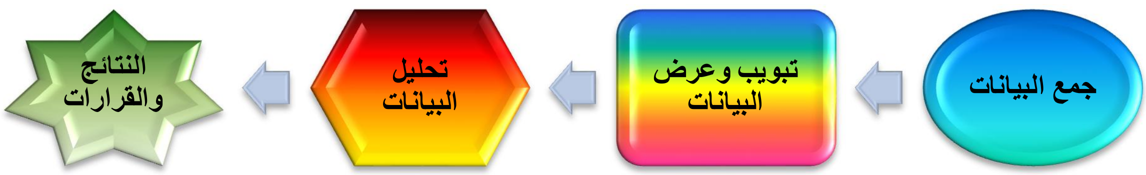
من المفاهيم الشائعة عن الاحصاء ، ما هي الا ارقام وبيانات رقمية فقط ، كأعداد السكان واعداد المواليد والوفيات ، واعداد المزارع والمزارعين ، واعداد الجيش واسلحته ...الخ. ومن ثم ارتبط مفهوم الاحصاء بانه العد والحصر للاشياء والتعبير عنها بارقام وهذا المفهوم المحدود لعلم الاحصاء . وعندما نتكلم عن علم الإحصاء لا نعنى بذلك البيانات الإحصائية وإنما نقصد حينئذ الطريقة الإحصائية . وهى الطريقة التى تمكننا من جميع الحقائق عن الظواهر المختلفة فى صورة قياسية رقمية وعرضها بيانيا ووضعها فى جداول تلخيصية بطريقة تسهل تحليلها بهدف معرفة اتجاهات هذه الظواهر وعلاقات بعضها ببعض .

ولقد كان الهدف الرئيسـى من علم الإحصـاء قديما هو عد أو حصـر الأشـياء المراد توفير بيانات إحصـائية عنها ، وكانت الجهة التى تقوم بإعداد الإحصـاءات على مسـتوى الدولة تعرف بمصـلحة التعداد ولذلك كان التعريف القديم لعلم الإحصـاء أنه علم العد ، أي العلم الذي يشـتمل على أسـاليب جمع البيانات الكمية عن المتغيرات والظواهر موضوع الدراسة .

ولكن مع تطور المجتمعات وتشـابه جوانب الحياة الاقتصـادية والاجتماعية الحديثة بها ، لم يعد مجرد توفير البيانات الكمية عن المتغيرات والظواهر موضـوع الدراسـة يفى بحاجات متخذى القرارات وصـانعى السـياسـة العامة إلى تكوين صـورة متكاملة الجوانب عن مجتمعهم والمجتمعات المحيطة به . فقام العلماء بتحديث نظريات علم الإحصـاء وأسـاليبه وأدواته لكى يعين الباحثين وغيرهم على اسـتخلاص اسـتنتاجات معينة من البيانات الكمية التى أمكن لهم جمعها عن طريق العد. لذا الاحصاء كعلم يمكن تعريفه كالآتي:

تعريف علم الاحصاء

هو ذلك العلم الذي يعمل على استخدام الاسلوب العلمي في طرق جمع البيانات وتبويبها



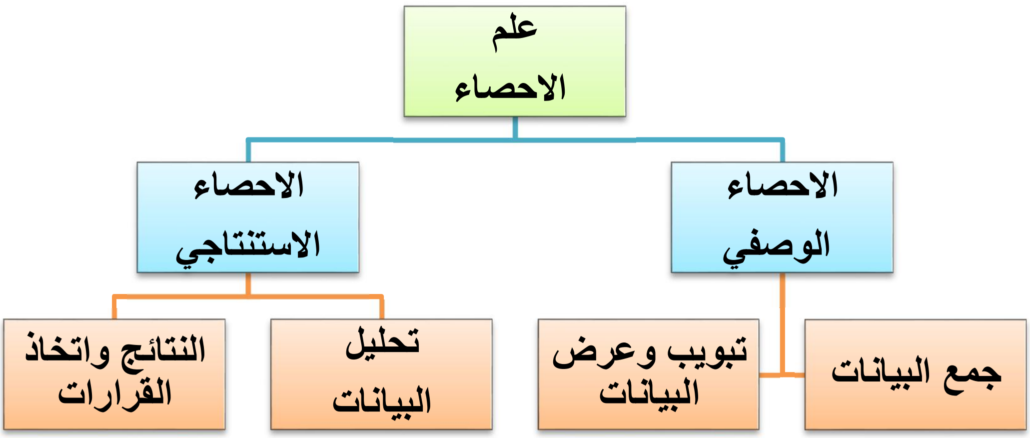
مما سبق يمكن تصنيف علم الاحصاء الى قسمين رئيسين هما:

١- الاحصاء الوصفي (Descriptive Statistics)

وهو ذلك الفرع من الاحصـــــاء الذي يتناول جمع وتنظيم وتلخيص وتبويب وعرض البيانات وحســـــاب بعض المقاييس الاحصائية المختلفة لها .

٢- الاحصاء الاستنتاجي او الاستدلالي (Statistical inference)

ويشمل الطرق الاحصائية التي تهدف الى عمل استنتاجات او استدلالات حول المصدر الذي جمعت منه البيانات والتوقعات عن المجتمع من خلال دراسة عينة من ذلك المجتمع.



علاقة الاحصاء بالعلوم الاخرى

يحتل علم الاحصــاء مكاناً مرموقاً بين العلوم مما له من اسـتعمالات واسـعة كأداة او وسـيلة للوصـول الى قرارات صـائبة لوصـف او تفسير الظواهر المختلفة في جميع العلوم . وقد شـملت تطبيقات علم الاحصـاء في علوم الزراعة ، الاحياء ، الكيمياء ، الفيزياء ، الفضاء والفلك ، الطب ، الصيدلة ، الاتصالات ، العلوم الهندسية ، البيئة ، التربية ، علم النفس ، العلوم السياسية وغيرها من العلوم الاخرى.

ومن التطبيقات العملية لعلم الاحصاء في العلوم الزراعية هي جمع البيانات اللازمة لمساحات الاراضي الزراعية وتصنيفها ، وجمع البيانات عن انواع المحاصيل المزروعة واعداد وانواع الثروة الحيوانية. وكذلك جمع البيانات عن كمية الانتاج وانتاجية وحدة المساحة للاراضي الصالحة للزراعة بهدف دراسة هذه البيانات وتحليلها للوصول الى استنتاجات عملية لتطوير واقع القطاع الزراعي من خلال دراسة الظواهر التي تسبب في انخفاض انتاجية المحاصيل كالجفاف والافات الزراعية وتملح التربة وغيرها بهدف وضع الخطط اللازمة للحد من هذه الظواهر والنهوض بواقع الانتاج الزراعي.

وجدت الامثلة والبداية في استخدام الرياضيات والإحصاء في القوانين الاحيائية على يد مؤسس علم الوراثة الراهب جريجور مندل عندما استنتج بعض القوانين الخاصة بعلم الوراثة بناءً على حسابات رياضية واحصائية. كان كل من التفكير والتمثيل الخاصين بالإحصاء الحيوي ذوي أهمية حاسمة في إيجاد نظريات الأحياء الحديثة. ففي أوائل القرن الماضي وبعد إعادة اكتشاف العمل الذي قام به مندل في مجال الوراثة فإن الفجوات الخيالية في الفهم والواقعة بين علم الوراثة والداروينية التطورية أدت إلى نقاش قوي بين الإحصائيين الحيويين مثل والتر ويلدن وكارل بيرسون وأتباع مندل مثل تشارلز دافينبورت وويليام باتسون وويلهلم جوهانسون .ساهم الإحصائيون والنماذج التي بنيت على الاستنتاج الإحصائي في حل هذه الاختلافات بحلول حقبة الثلاثينيات من القرن الماضي وذلك لإنتاج اصطناع الداروينية الجديدة التطوري الحديث.

لقد اعتمدت الشخصيات الرائدة جميعها في تأسيس هذا الاصطناع على الإحصائيات كما وطورت استخدامها في علم الأحياء ونورد فيما يلي بعضا من هؤلاء الشخصيات:

- طوّر سير رونالد أ. فيشر طرقا إحصائية أساسية عدة بهدف دعم كتابه النظرية الجينية للانتخاب الطبيعي.
- استخدم سيوال ج. رايت الإحصاءات في تطوير علم وراثة السكان الحديث.
- أعاد كتاب جي. بي. اس. هالدين :أسباب التطور تأسيس الانتخاب الطبيعي كآلية أصلية في التطور عن طريق شرحها من حيث عواقب الوراثة المندلية الرياضية.

ساعد العمل الذي قام به هؤلاء الأفراد إضافة إلى إحصائيين حيويين آخرين والمشتغلين بعلم الأحياء الرياضي وعلم الجينات الذي يميل إلى الإحصاء في جمع علم الأحياء التطوري وعلم الجينات في كلٍ متماسك ومتين والذي يمكن أن يبدأ تمثيله كميًا.

ويستعمل علم الاحصاء في السياسات المالية للدول من خلال جمع البيانات وعمل قواعد بيانات حقيقية لمقدار دخل الافراد واسعار السلع والخدمات ومقدار التضخم المالي ومصادر الدخل القومي وغيرها من البيانات ومن ثم تحليلها والوصول الى استنتاجات وقرارات مناسبة للنهوض بالواقع الاقتصادي والمالي للبلد. وكذلك يستخدم علم الاحصاء في التنبؤ بنتائج الانتخابات من خلال جمع الاستبيانات لشرائح مختلفة من الشعب وتحليلها والاستنتاج منها. وكذلك يستخدم علم الاحصاءفي مجال التربية والتعليم لرسم السياسة التربوية للبلدان من خلال جمع البيانات وتبويبها وتلخيصها حول الكثافات السكانية في المناطق والفئات العمرية والاجناس ومن ثم تحليل هذه البيانات والاستنتاج منها حول عدد المدارس المطلوبة في المدن والاقضية والنواحي والقرى واعداد المعلمين والمدرسين وغيرها من متطلبات السياسة التربوية.

الفصل الثاني

2.1 تعريف المتغير

2 .2 انواع المتغيرات

2.3 تعريف العينة والمجتمع

2.4 الرموز الاحصائية

تمارين

تعريف المتغير

عند دراسة صفة معينة مثل عدد الثمار في شجرة البرتقال لمجموعة معينة من اشجار البرتقال فسنجد اختلافات في عدد الثمار من شجرة لاخرى ، وفي هذه الحالة يطلق على صفة عدد الثمار بالشجرة بمصطلح المتغير. وكذلك عند دراسة صفة حاصل البذور لمحصول الرز في وحدة المساحة (كغم /دونم) لمجموعة من المزارع ، فسنجد اختلافات في كمية حاصل البذور في وحدة المساحة من مزرعة لاخرى ، وفي هذه الحالة نطلق على صفة الحاصل بمصطلح المتغير.

تعريف المتغير (Variable) **هو اي ظاهرة تظهر اختلافات بين مفرداتها**

أنواع المتغيرات

تنقسم المتغيرات إلى نوعين:

١ – متغيرات نوعية او وصفية (Qualitative Variable)

هي بيانات غير رقمية، أو بيانات رقمية مرتبة في شكل مستويات أو في شكل فئات رقمية، ومن ثم تقاس البيانات الوصفية بمعيارين هما:

أ -بيانات وصفية مقاسة بمعيار اسمي Nominal Scale

وهي بيانات غير رقمية تتكون من مجموعات : متنافية، كل مجموعة لها خصائص تميزها عن المجموعة الأخرى، كما أن هذه المجموعات لا يمكن المفاضلة بينها، ومن الأمثلة على ذلك:

- الجنس : متغير وصفي تقاس بياناته بمعيار اسمي " ذكر – أنثى. "

- الحالة الاجتماعية : متغير وصفي تقاس بياناته بمعيار اسمي " متزوج ، أعزب أرمل ، مطلق."

- أصناف التمور :متغير وصفي يقاس بياناته بمعيار اسمي " برحي ،خستاوي ، زهدي ، مكتوم".

- الجنسية : متغير وصفي يقاس بياناته بمعيار اسمي " عراقي غير عراقي"

وهذا النوع من البيانات يمكن اعطاء مجموعاته رموز او أرقام، فمثلا الجنسية يمكن إعطاء الجنسية " عراقي الرمز (1) ، والجنسية " غير العراقي " الرمز (2)

ب -بيانات وصفية مقاسة بمعيار ترتيبي Ordinal Scales

تتكون من مستويات، أو فئات يمكن ترتيبها تصاعديا أو تنازليا، ومن الأمثلة على ذلك:

- تقديرات النجاح للطالب :متغير وصفي تقاس بياناته بمعيار ترتيبي (مقبول ، متوسط ، جيد ، جيدجدا ، امتياز)

- المستوى التعليمي :متغير وصفي تقاس بياناته بمعيار ترتيبي" أمي ، يقرأ ويكتب، ابتدائية، متوسطة ، ثانوية ، جامعية ، أعلى من جامعية"

- المستوى الاقتصادي :متغير وصفي تقاس بياناته بمعيار ترتيبي: " غني ، متوسط ، فقير"

٢- متغيرات كمية (عددية) Quantitative Variables

وهي تلك الظواهر او الصفات التي يمكن قياسها مباشرة بارقام عددية وتنقسم إلى:

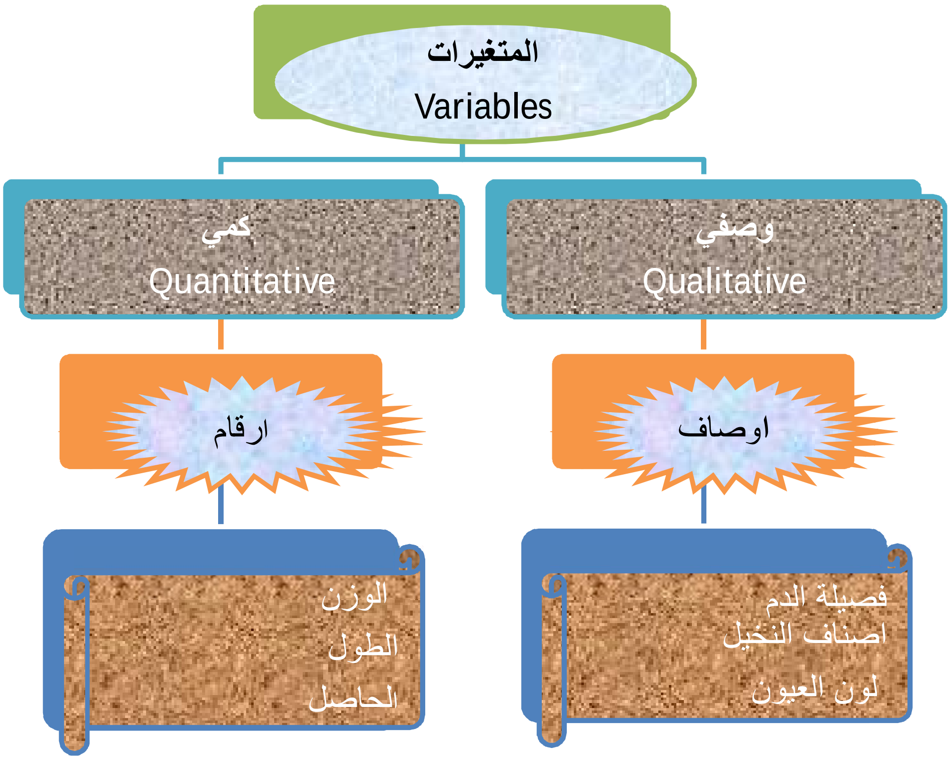
أ – متغير متقطع (غير مستمرة) Discrete Variable

وهو المتغير الذي يأخذ أعداد صحيحة، فمثلا إذا كان x متغير يمثل عدد أفراد الأسرة، فإنه يمكن أن يأخذ القيم ٢، ٣، ٤، ٥ ... ولا يمكن أن يأخذ x القيم 1.5 ، ٣,٢٥، ٥,١٧.وكأمثلة اخرى على المتغيرات المتقطعة هي عدد الثمار على النباتات ، عدد الطلبة في المدارس، عدد اشجار النخيل في محافظة ما.

بصورة عامة البيانات المستحصلة من طريقة العد (counting) فهي بيانات لمتغير متقطع او غير مستمر.

ب – متغير متصل (مستمر) Continuous Variable

وهو المتغير الذي يمكن أن يأخذ أي قيمة بين قيمتين معنيتين، وكأمثلة عن المتغيرات المتصلة:كمية الحاصل ، الطول، الوزن، الزمن، السرعة ...الخ، فإذا كان x هو متغير الطول فمثلا فإن x يمكن أن تأخذ القيم ١٥متر، ١١,٣ متر، ١٤,٧٥ متر، أي أن المتغير x يمكن أن نأخذ أي قيمة في فترة زمنية معينة. وبصورة عامة البيانات المستحصلة بطريقة القياس (Measurement) تعتبر لمتغير مستمر او متصل.



4– ٢ الرموز الاحصائية Statistical Notation

نظرا لاستخدام كافة المراجع العلمية العالمية الرموز والمعادلات الرياضية اللاتينية وذلك لكونها رموزا عالمية متفق عليها من جهة ولعدم وجود اتفاق تام بالوقت الحاضر على تعريبها من جهة اخرى ، لذا سوف نستخدم تلك الرموز وذلك لسهولة الاستفادة من المراجع الاجنبية والعلمية الاخرى.

يرمز للمتغير الرمز x او y او z ولكل قيمة له بالرمز x_i او y_i او z_i

ويمثل الرمز i رقم المفردة للمتغير.

فلو كانت اطوال اربعة اشجار نخيل هي ٦ ٤ ٣ ٥ متر فان

$$x_i = 6 , 4, 3, 5$$

اي ان $x_1 = 6$ القيمة الاولى للمتغير x

$x_2 = 4$ القيمة الثانية للمتغير x

$x_3 = 3$ القيمة الثالثة للمتغير x

$x_4 = 5$ القيمة الرابعة للمتغير x

ويرمز لمجموع قيم المتغير بالرمز

$$\sum_{i=1}^{n} x_i$$

الرمز $\sum$ هو حرف اغريقي يسمى Sigma ويعني مجموع والرمزان i و n هما حدا المجموع وعليه فالرمز

$$\sum_{i=1}^{n} x_i$$

يقرأ مجموع قيم x من المشاهدة الاولى وحتى الاخيرة

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + x_3 + \dots + x_n$$

اما الرمز

$$\sum_{i=3}^{5} x_i = x_3 + x_4 + x_5$$

فيقرأ مجموع قيم x الثالثة والرابعة والخامسة

اما الرمز

$$\sum_{i=1}^{n} x_i^2$$

فيعني مجموع مربعات قيم x من المشاهدة الاولى الى الاخير

$$\sum_{i=1}^{n} x_i^2 = x_1^2 + x_2^2 + x_3^2 + \dots + x_n^2$$

ملاحظة :

في حالة عدم ذكرحدي المجموع (i و n) فتعني اخذ جميع القيم من ١ الى n

$$\sum x_i^2 = x_1^2 + x_2^2 + x_3^2 + \dots + x_n^2$$

والرمز $(\sum_{i=1}^{n} x_i)^2$ يعني مربع مجموع المشاهدات

$$(\sum_{i=1}^{n} x_i)^2 = (x_1 + x_2 + x_3 + \dots + x_n)^2$$

ويرمز لحاصل ضرب مجموعين لقيم متغيرين بالرمز $(\sum x_i)(\sum y_i)$

$$(\sum x_i)(\sum y_i) = (x_1 + x_2 + x_3 + \dots + x_n)(y_1 + y_2 + y_3 + \dots + y_n)$$

مثال (١) : اذا كانت قيم المتغير x كالاتي xi = 6 2 3 5 4

وقيم المتغير y كالاتي yi = 1 3 7 8 6

اوجد (أ) $\sum x_i$ (ب) $\sum_{i=2}^{4} y_i$ (ج) $\sum X_i^2$ (د) $(\sum y_i)^2$

(هـ) $(\sum x_i)(\sum y_i)$ (و) $\sum x_i y_i$

الحل

(أ) $\sum x_i = x_1 + x_2 + x_3 + x_4 + x_5$

$= ٦ + 2 + 3 + 5 + 4$

$= 20$

(ب)

$$\sum_{i=2}^{4} y_i = y_2 + y_3 + y_4$$

$$\sum_{i=2}^{4} y_i = 3 + 7 + 8 = 18$$

(ج) $$\sum x_i^2 = x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2$$

$= 6^2 + 2^2 + 3^2 + 5^2 + 4^2$

$= 36 + 4 + 9 + 25 + 16$

$= 90$

(د) $$(\sum y_i)^2 = (y_1 + y_2 + y_3 + y_4 + y_5)^2$$

$= (1 + 3 + 7 + 8 + 6)^2$

$= (25)^2$

$= 625$

(هـ) $(\sum x_i)(\sum y_i)$ تعني حاصل ضرب مجموع x في مجموع y

$$(\sum x_i)(\sum y_i) = (x_1+x_2+x_3+x_4+x_5)\ (y_1+y_2+y_3+y_4+y_5)$$

$$= (6+2+3+5+4)(1+3+7+8+6)$$

$$= (20)\ (25)$$

$$= 500$$

(و) $\sum x_i y_i$ تعني مجموع حاصل ضرب قيم x في قيم y المناظرة

$$\sum x_i y_i = x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4 + x_5y_5$$

$$= (6x1)+(2x3)+(3x7)+(5x8)+(4x6)$$

$$= 6 + 6 + 21 + 40 + 24$$

$$= 97$$

مثال (٢) اذا علمت بان قيم المتغيرين x و y هما كالآتي:

$x = 5\ 3\ 6\ 2$

$y = 4\ 7\ 9\ 5$

اوجد (أ) $\sum x_i - 5$ (ب) $\sum (y_i - 2)$ (ج) $\sum (x_i + y_i)$

(د) $\sum (x_i - 1)(y_i - 2)$

الحل :

(أ) $$\sum x_i - 5 = (x_1 + x_2 + x_3 + x_4) - 5$$

$= (5 + 3 + 6 + 2) - 5$

$= (16) - 5$

$= 11$

$$\sum(y_i - 2) = (y_1 - 2) + (y_2 - 2) + (y_3 - 2) + (y_4 - 2)$$ (ب)

$= (4 - 2) + (7-2) + (9-2) + (5 -2)$

$= 2 + 5 + 7 + 3$

$= 17$

$\sum (xi + yi) = (x1+y1)+(x2+y2)+(x3+y3)+(x4+y4)$ (ج)

$= (5+4) + (3+7) + (6+9) + (2+5)$

$= 9 + 10 + 15 + 7$

$= 41$

$\sum$**(xi-1)(yi-2)** = (د)

$=(x1-1)(y1-2)+ (x2-1)(y2-2)+(x3-1)(y3-2)+(x4-1)(y4-2)$

$$= (5-1)(4-2)+(3-1)(7-2)+(6-1)(9-2)+ (2-1)(5-2)$$

$$= 8 + 10 + 35 + 3$$

56

تمارين الفصل الثاني

(١) عين نوع المتغير نوعي او كمي في كل من الحالات التالية:

أ - لون الازهار في نبات الداوودي.

ب- كمية حاصل الذرة الصفراء في الدونم الواحد.

ج- الحالة الزوجية.

د- لون العيون.

هـ- وزن العجول.

(٢) حدد نوع المتغير (مستمر ام متقطع) للحالات التالية:

أ- عدد افراد العائلة .

ب- مساحة المزرعة .

ج- عدد الكتب في مكتبة المدرسة.

د- عددالنخيل في متنزه الزوراء.

هـ- اطوال طلبة الاول زراعي في بابل.

(٣) اذكر بعض الامثلة للمجتمعات والعينات؟

(٤) ما الفرق بين المجتمع المحدود والمجتمع غير المحدود مع ذكر امثلة لكل منهما.

(٥) عرف ما يلي المجتمع ، العينة ، المعالم ، الاحصاءات.

(٦) تأمل سلسلتي الاعداد التاليتين عن المتغيرين x و y

xi = 2 , 4 , 5 , 8 , 3

yi = 1 , 2 , 3 , 2 , 3

جد

$(1)\sum_{1}^{5} x_i$

$(2)\sum_{1}^{5} y_i^2$

$(3)\sum_{i=1}^{n} 3x_i$

$(4)\sum_{i=1}^{5} (x_i - 4)$

$(5)\ (\sum_{i=1}^{5} x_i)^2$

$(6)\sum_{i=1}^{5} \frac{1}{x_i}$

$(7)\sum x_i y_i$

$(8)\sum_{i=1}^{5} (x_i - y_i)$

(٧) البيانات التالية تمثل قيم ظاهرة معينة

xi	2	6	5	10	7	9	8

المطلوب ايجاد

$$(1)\sum_{i=1}^{4} x_i \quad (2)\sum_{i=4}^{y} x_i^2 \quad (3)\sum_{i=2}^{5}(x_i-2)$$

$$(4)\sum_{i=1}^{4}(x_i-5)^2 \quad (5)\sum_{i=3}^{6}\frac{1}{x_i}$$

(٨) اذا كانت القيم ٦ ٤ ٢ ٧ ٣ تمثل اطوال خمسة نباتات ، ما هي قيمة:

أ- $\sum x_i$ ب- $\sum_{i=1}^{n} X_i^2$ ج- $(\sum_{i=1}^{n} X_i)^2$ د- $\sum (x_i - 3)$

(٩) اذا كانت القيم ١٠ ٥ ٨ ٧ تمثل عدد الفسائل في شجرة النخيل ، فما هي قيمة:

أ- $\sum (x_i + 2)$

ب- $\sum_{i=1}^{n} X_i^2 - \sum_{i=1}^{n} X_i$

(8) لو فرضنا قيم x و y هي

الطول (x)	الوزن (y)	الرمز الدليلي(i)
٤	٨	١
٧	١٢	٢
٥	١٠	٣
٢	٦	٤
٦	١٤	٥

اوجد أ- $\sum_{i=1}^{3} x_i y_i$

ب- $\sum_{i=1}^{3} X_i \sum_{i=1}^{3} y_i$

ج- $\sum_{i=2}^{4} y_i^2 - \sum_{i=2}^{4} x_i^2$

د- $\sum_{i=2}^{5} (x_i - 1)(y_i - 5)$

نظرية المعاينة

Sampling Theory

مقدمة

يعتبر اسلوب المعاينة من اهم الاساليب الاحصائية التي نستخدمها لدراسة مجموعة كبيرة من المفردات (تسمى مجتمع) بقصد التعرف على خواصها عن طريق دراسة مجموعة صغيرة من هذه المفردات (تسمى عينة). ان الصعوبات التي تصادف الباحثين عند دراسة جميع مفردات المجتمع ، وخاصة اذا كان المجتمع كبيراً ، تجعل الباحثين يلجأون عادة الى اختيار مجموعة صغيرة (عينة) يتم اختيارها من المجتمع بطريقة معينة بحيث تكون هذه المجموعة صورة مصغرة للمجتمع بقدر الامكان ثم يقومون بدراسة هذه العينة بدقة للتعرف على خواصها ومعرفة معالمها مثل الوسط الحسابي والوسيط والانحراف القياسي والتباين وغير ذلك من المقاييس الاحصائية ثم يقومون بتعميم هذه النتائج التي يحصلون عليها الى المجتمع الاصلي للتعرف على خواصه ومعالمه وهذا هو الهدف الاساسي من الدراسة. وبالطبع لا يكون هذا التعميم من العينة الى المجتمع له معنى او قيمة علمية الا اذا تم اختيار المفردات بطريقة نضمن بها ان تكون العينة ممثلة للمجتمع تمثيلاً حقيقياً ومثل هذه العينات يطلق عليها اسم العينات العشوائية.

تعريف المجتمع والعينة

المجتمع من الناحية الاحصائية يمثل جميع الافراد (او العناصر) التي تشترك في صفة متغيرة واحدة او اكثر تميزه تمييزا تاما عن بقية المجتمعات.

تعريف المجتمع (Population)

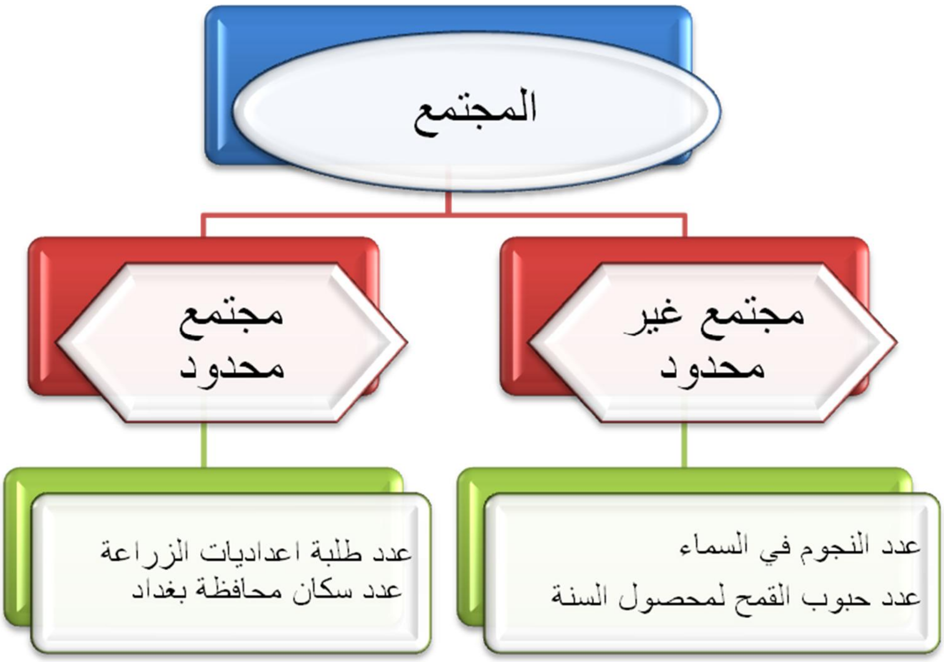
هو جميع القيم لمتغير ما

ويتعلق مفهوم المجتمع بالهدف المحدد للبحث الاحصائي .فمثلا اذا كان هدف البحث حساب عدد النخيل في العراق فعندها يكون المجتمع هو جميع مزارع النخيل في العراق بدون استثناء. وقد تشكل محافظة البصرة مجتمعا احصائيا اذا كان الهدف هو دراسة عدد النخيل في محافظة البصرة. وقد يشكل قضاء ابي الخصيب مجتمعا احصائيا اذا كان الهدف هو دراسة اعداد النخيل في قضاء ابي الخصيب.

وتختلف المجتمعات في احجامها (عدد مفرداتها) فبعضها صغير الحجم وبعضها كبيرة والبعض الاخر غير معروف الحجم .لذا فان المجتمعات تقسم الى:

أ- مجتمع محدود Finite population فاذا كان عدد افراد المجتمع محدود كما هي الحال في عدد اشجار النخيل في مزرعة ما ، او عدد الطلبة في كلية الزراعة. عدد حقول الدواجن في بغداد.

ب- مجتمع غير محدود Infinite Population اذا كان حجم المجتمع كبير جدا ولا يمكن حصره مثلا عددالاسماك في الخليج العربي ، عدد الحشرات على اشجار الحمضيات في محافظة ديالى ، عدد الطيور في الاهوار.



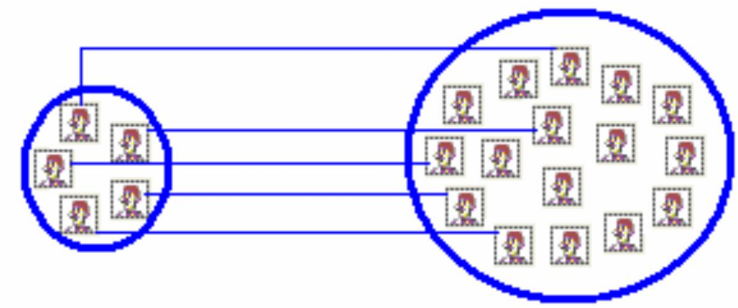
العينة (sample)

في حالة عدم امكانية الحصول على قيم او بيانات عن المجتع لاسباب مادية او فنية ، لذا نلجأ

الى اخذ عينة معينة من المجتمع بطريقة ما بحيث تمثله تمثيلا حقيقيا ، لذا تعرف العينة كالاتي:

تعريف العينة (Sample)

هي جزءمن المجتمع مأخوذة منه بطريقة عشوائية وتكون ممثلا له تمثيلا حقيقيا



عينة الدراسة   مجتمع الدراسة

شكل : الفرق بين المجتمع والعينة

يعتمد اسلوب العينات على معاينة جزء من المجتمع محل الدراسة ، يتم اختياره بطريقة علمية سليمة ، ودراسته ثم تعميم نتائج العينة على المجتمع ، ويتميز هذا الاسلوب بالآتي:

١ – تقليل الوقت والجهد.

٢ – تقليل التكلفة.

٣- الحصول على بيانات اكثر تفصيلا ، وخاصة اذا جمعت البيانات من خلال استمارة استبيان.

٤ – كما ان اسلوب المعاينة يفضل في بعض الحالات التي يصعب فيها اجراء حصر شامل مثل معاينة دم المريض ، اعداد الاسماك في البحر.

ولكن يعاب على هذا الاسلوب بان النتائج المستحصلة بهذا الاسلوب تكون اقل دقة من نتائج اسلوب الحصر الشامل وخاصة اذا كانت العينة المختارة لا تمثل المجتمع تمثيلا جيدا.

شروط اختيار العينة

١- يجب ان لا تتسم العينة التي تم اختيارها بالتحيز او المحاباة بمعنى ان تأخذها من بين مفردات المجتمع الاصلي عشوائياً.

٢- ان تكون الظاهرة المراد عمل معاينة لها سائدة ومنتشرة في المجتمع الاصلي وليست نادرة الحدوث.

٣- يجب ان تكون العينة ممثلة لجميع المجتمع الاصلي .

٤- ضرورة افتراض تجانس مفردات المجتمع الاصلب وفي حالة تعذر ذلك في المجتمعات غير المتجانسة يلجأ الباحث الى تقسيمها الى مجتمعات صغيرة متجانسة .

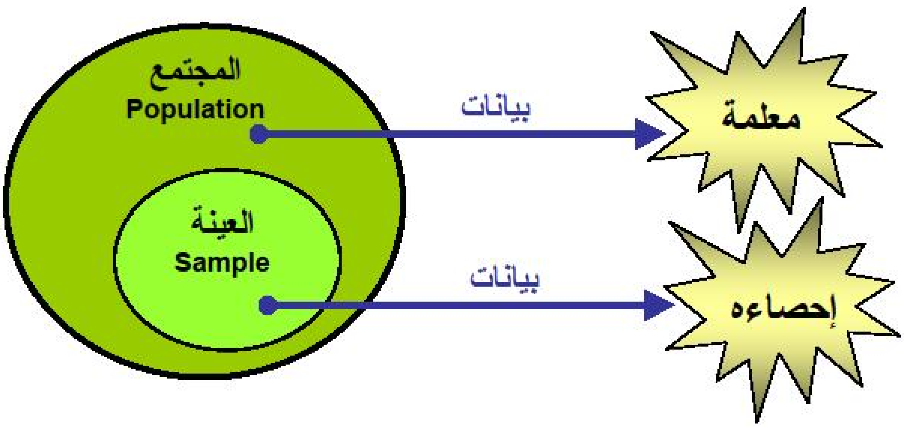
العوامل التي تحدد حجم العينة

١. حجم المجتمع الاحصائي الذي ستسحب منه العينة.
٢. درجة الاختلاف بين مفردات المجتمع الاحصائي .
٣. نسبة الخطأ المسموح به او القبول ودرجة الثقة التي يرغب الباحث في توافرها في النتائج التي يصل اليها من دراسته للعينة

الثوابت (اوالمعالم) والاحصاءات Parameters and Statistics

يطلق على المقاييس التي تحسب من المجتمع نفسه (اي من جميع القيم) مصطلح الثوابت او المعالم ، اما المقاييس المناظرة المحسوبة من العينة فتسمى الاحصاءات او التقديرات لانها لا تمثل سوى تقديرات لمعالم المجتمع الذي اخذت منه العينة.

وتجدر الاشارة الى ان معالم المجتمع محددة القيم (ثابتة) بينما الاحصاءات المناظرة تتغير بتغير العينة.



انواع البيانات الاحصائية

البيانات الاحصائية هي المواد الاولية للاحصاء ، والتي من خلالها يمكن تبويبها وتصنيفها واجراء التحاليل عليها واستنباط التوصيات منها. وتعتمد طبيعة البيانات الاحصائية على الاهداف المطلوبة منها ، لذا فان هذا الفصل سيتناول مصادر البيانات وطرق جمعها.

ويمكن تقسيم البيانات الإحصائية إلى نوعين أساسيين من البيانات هما:

١. البيانات النوعية Qualitative data:

وهي بيانات وصفية، تشمل الظواهر التي لا تخضع للقياسات الكمية، ويصعب التعبير عنها بصورة عددية. كما تعرف أحياناً باسم (البيانات التصنيفية) لأنها تصنف البيانات حسب الصفات، سواء أكانت ذلك من حيث النوع، مثل تصنيف التربة إلى طينية وغرينية ورملية، أم كان ذلك من حيث الدرجة، مثل تصنيف القوى العاملة بحسب الدرجة التعليمية، إلى أمي، وملم بالقراءة والكتابة، حاصــل على الشــهادة الابتدائية...الخ، حيث تشــتمل البيانات على عدد الأفراد الذين ينتمون إلى كل صفة من هذه الصفات.

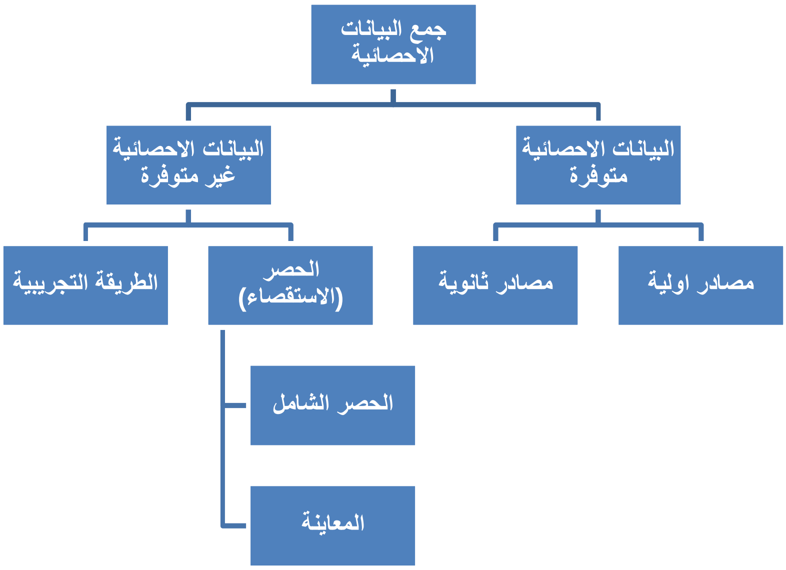
٢. البيانات الكمية Quantitative data

وهي بيانات رقمية، تشــمل الظواهر القابلة للقياســات الكمية، ويمكن التعبير عنها بصــورة عددية، مثل كميات الأمطار، وأعمار الســـكان وكميات الإنتاج.. الخ، وغير ذلك من البيانات التي تعكس القيم الفعلية للظواهر. ومن الجدير بالذكر أن معظم الأساليب الإحصائية تعنى بمعالجة البيانات الرقمية، وهذا على خلاف البيانات النوعية، التي لا تستخدم فيها سوى بعض الأساليب الإحصائية المعدودة.

طرق واساليب جمع البيانات الاحصائية

تعتبر طريقة جمع البيانات من أهم المراحل التي يعتمد عليها البحث الإحصائي، كما أن جمع البيانات بأسلوب علمي صحيح، يترتب عليه الوصول إلى نتائج دقيقة في التحليل، وقد تكون البيانات المطلوبة متوفرة سواء كانت منشورة اوغير منشورة نتيجة لجمعها من قبل جهات رسمية او غير رسمية وفي هذه الحالة فأن جمعها يتطلب الرجوع الى تلك المصادر . اما اذا كانت البيانات غير متوفرة فان الامر

يتطلب اجراء استقصاءات او تجارب ميدانية لتوفيرها. وسيتناول هذا الفصل مصادر البيانات المتوفرة والطرق الاحصائية لجمع غير المتوفر منها.



ولدراسة طرق جمع البيانات، يجب الإلمام بالنقاط التالية:

1- مصادر البيانات المتوفرة ٢- أسلوب جمع البيانات الغير متوفرة.

١- مصادر جمع البيانات المتوفرة

هناك مصدرين للحصول منها على البيانات هما:

أولا: المصادر الأولية Primary Source

وهي المصادر التي نحصل منها على البيانات بشكل مباشر، حيث يقوم الباحث نفسه بجمع البيانات من المفردة محل البحث مباشرة، فعندما يهتم الباحث بجمع بيانات عن الأسر الريفية، يقوم بإجراء مقابلة مع رب الأسرة، ويتم الحصول منه مباشرة على بيانات خاصة بأسرته، مثل بيانات المنطقة التابع لها،وعدد افراد الاسرة واعمارهم والمستوى التعليمي ، ومساحة الملكية الزراعية ، وانواع المحاصيل المزروعة، واعداد وانواع الثروةالحيوانية التي يمتلكها،وانواع المكننة الزراعية التي يستخدمها ونوع ملكيتها (ملك شخصي ام ايجار) والدخل الشهري، ... وهكذا. او قد تقوم جهات رسمية او شبه رسمية

بجمع انواع معينة من البيانات بشكل دوري او خلال فترة زمنية محددة كالجهاز المركزي للاحصاء او الهيئة العامة للانواء الجوية ، وقد تصدر هذه الجهات نشرات احصائية دورية.

ويتميز هذا النوع من المصادر بالدقة والثقة في البيانات، لأن الباحث هو الذي يقوم بنفسه بجمع البيانات من المفردة محل البحث مباشرة، ولكن أهم ما يعاب عليها أنها تحتاج إلى وقت ومجهود كبير، ومن ناحية أخرى فانها مكلفة من الناحية المادية.

ثانيا :المصادر الثانوية Secondary Source

وهي المصادر التي نحصل منها على البيانات بشكل غير مباشر، بمعنى آخر يتم الحصول عليها بواسطة أشخاص آخرين، أو أجهزة، وهيئات رسمية اخرى غير متخصصة بجمع البيانات الاحصائية، مثل نشرات منظمة الأغذية والزراعة الدولية....وهكذا.

ومن مزايا هذا النوع من المصادر، توفير الوقت والجهد والمال، إلا انها تحتوي على تفاصيل اقل من المصادر الاولية ، وأن درجة ثقة الباحث فيها ليست بنفس الدرجة في حالة المصادر الأولية.

2- أسلوب جمع البيانات الغير متوفرة

يتحدد الأسلوب المستخدم في جمع البيانات ، حسب الهدف من البحث، وحجم

المجتمع محل البحث، وهناك أسلوبين لجمع البيانات هما:

١- الاستقصاء (المسح) Survey Method

٢- الطريقة التجريبية Experimental Method

١- الاستقصاء (المسح) Survey

ان مصطلح الاستقصاء بالمفهوم الاحصائي يعني جمع البيانات حول صفات وخصائص الاشياء القائمة دون التحكم باي من العوامل التي تؤثر على المتغير قيد الدراسة. فمثلا لو اردنا على سبيل المثال جمع البيانات حول اعداد ومواصفات الاغنام العراقية ، فان عملية جمع البيانات الاحصائية تتركز على عدد، جنس ، عمر ، ونوع الاغنام دون التحكم بالعوامل التي تؤثر فيها (السلالات ، الاعلاف ، الخدمات البيطرية ، الحظائر ...الخ.

اما درجة تغطية الاستقصاء لعناصر المجتمع فقد تكون :

أولا :أسلوب الحصر الشامل (التعداد) Census

يستخدم هذا الأسلوب إذا كان الغرض من البحث هوحصر جميع مفردات المجتمع، وفي هذه الحالة يتم جمع بيانات عن كل مفردة من مفردات المجتمع بلا استثناء، كحصر جميع المزارع التي تنتج التمور، جميع الحقول المنتجة للدواجن، ويتميز أسلوب الحصر الشامل بالشمول وعدم التحيز، ودقة النتائج، ولكن يعاب عليه أنه

يحتاج إلى الوقت والجهد، والتكلفة العالية.

ويتم استخدام هذا الأسلوب في الحالات التالية:

١- إذا كان الغرض من البحث جمع بيانات عن مفردات المجتمع بصفة شخصية أو فردية فإن الأسلوب الذي يتبع في هذه الحالة هو أسلوب الحصر الشامل. فلو كان الغرض من البحث مثلاً، جمع بيانات عن مجتمع الآلات والمكائن الزراعية الموجودة في محافظة ما، وذلك للوقوف على الآلات التي تحتاج إلى إصـــلاح وصـــيانة، فالبيانات المطلوبة في هذه الحالة تخص كل آلة على حده، وبالتالي فإنها مطلوبة بصفة فردية وعليه فإنه لابد من استخدام أسلوب الحصر الشامل لحصر كل آلة على حده.

٢- يمكن استخدام الحصر الشامل عندما يريد الباحث الحصول على نتائج على مستوى عال من الدقة. فمثلاً تقوم الشركات المنتجة لأنابيب أفران الغاز بفحص هذه الأنابيب بأسلوب الحصر الشامل، وذلك للتأكد من ســـلامتها حتى لا يكون هناك أي احتمال لبيع أنابيب غير ســـليمة، وبالتالي تعريض حياة

المواطنين للخطر. وكذلك شركات إنتاج الأدوية التي تحمل طابع الخطورة يطبق عليها أسلوب الحصر الشامل حتى تتمكن الشركة من التأكد من سلامتها.

٣- ويمكن استعمال أسلوب الحصر الشامل في حالة ما إذا كانت مفردات المجتمع المراد دراسته غير متجانسة وإذا ما كان المجتمع صغير نسبيًا.

ثانيا :أسلوب المعاينة Sampling

يعتمد هذا الأسلوب على معاينة جزء من المجتمع محل الدراسة، يتم اختياره بطريقة علمية سليمة، ودراسته ثم تعميم نتائج العينة على المجتمع، ومن ثم يتميز هذا الأسلوب تقليل الوقت والجهد ، تقليل التكلفة ، والحصول على بيانات أكثر تفصيلا، وخاصة إذا جمعت البيانات من خلال استمارة استبيان.

كما أن أسلوب المعاينة يفضل في بعض الحالات التي يصعب فيها إجراء حصر شامل، مثل معاينة دم المريض، أو إجراء تعداد لعدد الأسماك في البحر او اعداد حشرات الحميرة والدوباس التي تصيب النخيل.

ولكن يعاب على أساوب المعاينة أن النتائج التي تعتمد على هذا الأسلوب أقل دقة من نتائج أسلوب الحصر الشامل، وخاصة إذا كانت العينة المختارة لا تمثل المجتمع تمثيلا جيدا.

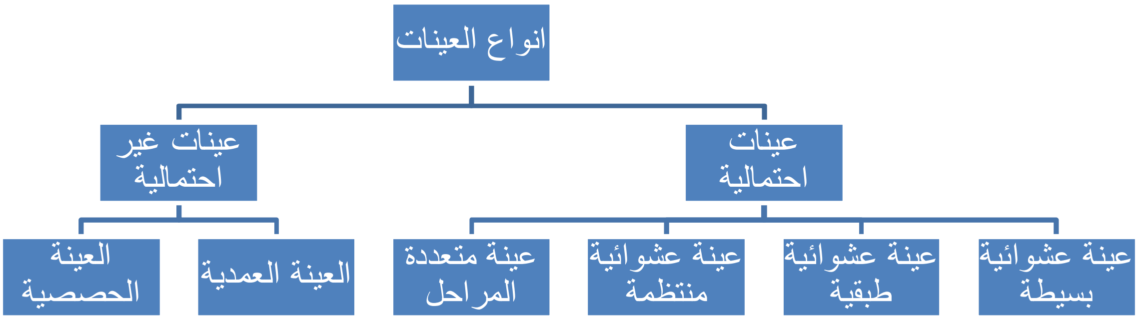
٢- الطريقة التجريبية Experimental Method

ان مصطلح التجربة يقصد به جمع البيانات عند ممارسة سيطرة فعلية على واحد او اكثر من العوامل التي تؤثر على المتغير قيد الدراسة ، وتستخدم هذه الطريقة لدراسة العلاقة السببية اي المسبب والتأثير ، عن طريق التحكم في العوامل الداخلة في التجربة لدراسة تأثيرها منفردة اومجتمعة. مثلا عند دراسة تأثير نوع معين ومستويات من السماد على انتاج محصول معين مع تثبيت العوامل الاخرى اللازمة لانتاج المحصول او تغيير بعضها كصنف البذور وطريقة الزراعة وعدد الريات ... الخ ، وتعتبر الطريقة التجريبية اقوى طرق البحث.

انواع العينات Samples

ويمكن تقسيم العينات وفقا لأسلوب اختيارها إلى نوعين هما:

أ -العينات الاحتمالية   ب -العينات غير الاحتمالية



أ- العينات الاحتمالية

هي العينات التي يتم اختيار مفرداتها وفقا لقواعد الاحتمالات، بمعنى آخر هي التي يتم اختيار مفرداتها من مجتمع الدراسة بطريقة عشوائية، بهدف تجنب التحيز الناتج عن اختيار المفردات،ومن أهم أنواع العينات الاحتمالية، ما يلي:

١ - العينة العشوائية البسيطة   Simple Random Sampling.

هي طريقة الاختيار التي تضمن احتمالا متساويا لكل عنصر من عناصر المجتمع بان يشمل في العينة واحتمالا متساويا لكل عينة من العينات التي يمكن تشكيلها من عناصر المجتمع.

مثال: اذا كان لدينا مجتمع مكون من٥ عناصر هي (أ ، ب ، ج ، د ، ه) واردنا اختيار عينة عشوائية بسيطة مكونة من ٣ عناصر، فان العينة التي يمكن اختيارها ستكون واحـــدة من العينات العشرة التي يمكن الحصول عليها من هذا المجتمع وهي:

أ ب ج   أ ب د   أ ب هـ   أ ج د   أ ج هـ

أ د هـ   ب ج د   ب ج هـ   ب د هـ   ج د هـ

ويجب الاشارة على ان العنصر لا يتكرر في اي عينةلان الترتيب غير مهم وذلك لان الترتيبات أ ب ج ، أ ج ب ، ب أ ج ، ب ج أ ، ج أ ب ، ج ب أ هي عينة واحدة تشمل على العنا صر أ . ب . ج

٢- العينة العشوائية الطبقية Stratified Random Sampling.

وهي طريقة اختيار عينة طبقية عن طريق تقسيم المجتمع الى اقسام متجانسة تعرف بالطبقات ، ثم اختيار عينة عشوائية فرعية بصورة عشوائية من كل طبقة ، وهذه العينات الفرعية مجتمعة تكون العينة الطبقية.

مثال: يراد اختيار عينة طبقية تمثل حاصل غلة محصول الحنطة في الدونم الواحد في العراق...

اولا – نقسم العراق الى ٣ مناطق متجانسة (طبقات) حسب الظروف المناخية وهي المنطقة الشمالية والوسطى والجنوبية.

ثانيا – ناخذ عينة عشوائية في كل من المنطقة الشمالية والوسطى والجنوبية.

ثالثا – العينات الثلاث اعلاه تمثل العينة الطبقية.

٣ – العينة العشوائية المنتظم Systematic Random Sampling.

وهي طريقة اختيار عينة منتظمة وذلك عن طريق اختيار وحدة المعاينة المرقمة K والتي تسمى نسبة المعاينة وهي حجم المجتمع الى حجم العينة ثم اختيار رقم عشوائي بين ١ و K ليكون رقم العينة الاولى ثم اضافة K ومضاعفاتها على رقم العينة الاولى الى ان يكمل حجم العينة.

مثال: يراد اختيار عينة منتظمة من شعبة أ للاول زراعي مكونة من خمسة طلاب علما بان عدد الطلبة في الشعبة ٤٠ طالبا.

اولا – نحسب نسبة المعاينة K

$$K = \frac{\text{حجم المجتمع}}{\text{حجم العينة}}$$

$$k = \frac{40}{5} = 8$$

ثانيا - نختار رقم عشوائي بين ١ و ٨ فليكن مثلا الرقم ٦

ثالثا _ نختار العنصر الاول من العينة الطالب الذي تسلسله ٦ ثم نضيف نسبة المعاينة (٨) ومضاعفاتها على الرقم ٦ الى ان نكمل حجم العينة ، اي يكون اختيار الطلبة الذين تسلسلهم في الصف ٦ ، ١٤، ٢٢ ، ٣٠ ، ٣٨.

٤– العينة المتعددة المراحل (العنقودية) Multi-Stage (Cluster) sampling

هي طريقة لاختيار عينة متعددة المراحل وذلك عن طريق اجراء الاختيار على مراحل متعددة . فاذا كان المجتمع مقسما الى اقسام فاننا في المرحلةالاولى نختارعشوائيا عينة من هذه الاقسام وفي المرحلة الثانية نختار عينة عشوائية من العينة التي اختيرت في المرحلة الاولى وهكذا حتى نحصل على حجم العينة المقرر.

مثال : لو اردنا معرفة غلة اصناف النخيل البرحي والخستاوي والمكتوم في محافظة ديالى . فلو اردنا ان نأخذ عينة مكونة من ١٠٠٠ نخلة من كل صنف من اصناف النخيل اعلاه وفق اسلوب المعاينة المتعددة المراحل ، نقوم بما يلي:

اولا- نختار ١٠ قرى من قرى محافظة ديالى التي تشتهر بزراعة النخيل.

ثانيا – نختار عشرة بساتين عشوائيا في كل قرية.

ثالثا – نختار ١٠ اشجار نخيل من كل صنف في كل بستان اختير اعلاه.

وبالتالي تكون حجم العينة لدينا مكونة من ١٠٠٠ نخلة لكل صنف

ب - العينات غير الاحتمالية

هي التي يتم اختيار مفرداتها بطريقة غير عشوائية، حيث يقوم الباحث باختيار مفردات العينة بالصورة التي تحقق الهدف من المعاينة وأهم أنواع العينات غير الاحتمالية:

١ - العينة العمدية Judgmental Sample

وهي الطريقة التي تستخدم لاخذ عينة صغيرة لمجتمع كبير ، ويلجأ الباحثون الى هذه الطريقة عندما يكون المجتمع على درجةعالية من التجانس. لذا يعمد بعض الباحثيين الى استخدام هذا الاسلوب معتقدين بتمثيل هذه العينة للمجتمع. وغالبا ما يلجأ الباحث على اختيار العينات القريبةمن متوسط المجتمع.

مثال: اختيار قرية في السماوة لحصر الامراض التي تصيب الجمال في العراق

وهنا يعتقد الباحث بان الامراض التي تصيب الجمال في السماوة هي نفسها التي تصيب الجمال في جميع مناطق العراق.

٢ -العينة الحصصية Quota Sample

وتستخدم عندما يكون المجتمع الاحصائي غير متجانس ومؤلف من طبقات. نقوم باختيار عينة من كل طبقة بصورة موضوعية (غير عشوائية) يتناسب حجمها مع حجم الطبقة في المجتمع ، ثم ندمج هذه العينات في عينة واحدة تسمى العينة الحصصية.

توزيعات المعاينة Sampling Distributions

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

توزيعات المعاينة للأوساط Sampling Distributions of Means

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

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \quad ، \quad \mu_{\bar{x}} = \mu$$

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

$$X \sim N(\mu, \sigma^2) \quad \Rightarrow \quad \bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$

ومن ثم يكون

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

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

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

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

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

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

حيث n_1, n_2 هما حجم العينة الأولى والثانية.

وإذا كان المجتمعين الأصـــليين طبيعيين فإن $(\bar{x}_1 \pm \bar{x}_2)$ يخضـــع لتوزيع طبيعي أيضـــاً بالمعلمات المعطاة وعليه فإنه في هذه الحالة

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

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

مثال : بفرض أنَّ لمنتج أجهزة A وسط عمر لها قدره ٨,٥ سنوات بانحراف معياري قدره ٠,٨ ، في حين لها وسط عمر قدره ٧ سنوات لمنتج آخر B بانحراف معياري قدره ٠,٩ ما احتمال أن يكون لعينة عشوائية مكونة من ٣٥ جهاز من المنتج A على الأقل وسط عمر أكبر بسنتين من وسط عمر عينة عشوائية مكونة من ٤٠ جهاز من المنتج B.

الحل : نلخص المعلومات في الجدول الآتي:

المنتج A	المنتج B
$\mu_1 = 8.5$	$\mu_2 = 7$
$\sigma_1 = 0.8$	$\sigma_2 = 0.9$
$n_1 = 35$	$n_2 = 40$

حسب قيمة z من القانون كالآتي:

$$z=\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}=\frac{2-(8.5-7)}{\sqrt{\frac{0.8^2}{35}+\frac{0.9^2}{40}}}$$

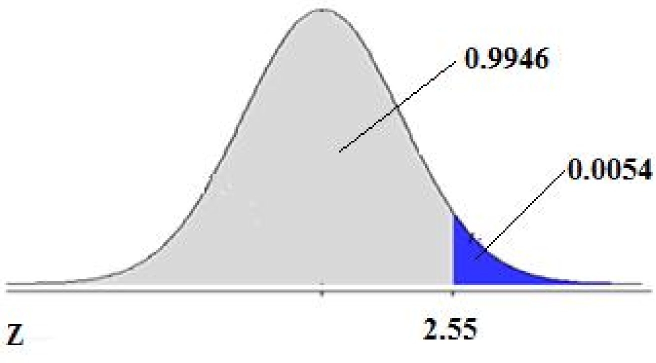
$$z=\frac{2-1.5}{\sqrt{\frac{0.64}{35}+\frac{0.81}{40}}}=\frac{0.5}{\sqrt{0.0183+0.0203}}$$

$$z=\frac{0.5}{\sqrt{0.386}}=\frac{0.5}{0.1965}=2.55$$

اي ان

$$P[(\bar{x}_1-\bar{x}_2)\geq 2]=P[z>2.55]$$
$$=1-P[z<2.55]=1-0.9946]=0.0054$$

$\therefore$ احتمال أن يكون لعينة عشوائية مكونة من ٣٥ جهاز من المنتج A على الأقل وسط عمر أكبر بسنتين من وسط عمر عينة عشوائية مكونة من ٤٠ جهاز من المنتج B هو ٠,٥٤ %



توزيع المعاينة للتباين: Sampling Distribution of The Variance

اذا كان $X_1, X_2, X_3, \ldots\ldots X_n$ مشـــاهدات من عينة لمتغيرعشـــوائي يتبع التوزيع الطبيعي $N(\mu,\sigma^2)$ ، فان:

$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ هو متوسط العينة لعدد n من المشاهدات،

$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_1 - \bar{X})^2$ هو تباين العينة لعدد n من المشاهدات

وان المتوسط للعينة $\bar{X}$ مستقل عن التباين S^2 للعينة

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

$$\frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2} \sim \chi^2(n-1)$$

وهنا يجدر الاشارة الى ان عندما تسحب عينة من مجتمع توزيع طبيعي فان:

$$\frac{\sum_{i=1}^{n}(X_i - \mu)^2}{\sigma^2} \sim \chi^2(n-1)$$

الفرق بين المعادلتين السابقتين هو ان الحالة الثانية هو جمع مربع الفرق بين المشاهدات ومتوسط المجتمع μ ، بينما في الحالة الاولى نحن جمعنا مربع الفرق بين المشاهدات ومتوسط العينة $\bar{X}$. والذي حدث عندما قدرنا متوسط المجتمع الغير معروف μ بمتوسط العينة $\bar{X}$ نحن سوف نخسر درجة حرية واحدة ، عموماً هذا صحيح ، اذ اننا نخسر درجة حرية واحدة لكل معلمة احصائية مقدرة في بعض المتغيرات العشوائية مثل مربع كاي.

مثال : سحبت عينة عشوائية بحجم 20 مشاهدة من مجتمع توزيع طبيعي تباينه 9، احسب احتمالية أن يزيد تباين العينة عن 15 .

الحل: n=20 ، $\sigma^2 = 9$ ، $S^2 = 9$

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

$$\chi^2 = \frac{(20-1)15}{9} = 31.67 \qquad p(\chi^2_{19} > 31.67) = 0.025$$

لاحظ أن القيمة 31.67 غير موجودة في الجدول الخاص بالتوزيع χ^2 فنأخذ أقرب قيمة لها وهي 32.825 وبذلك تكون الاحتمالية تساوي 0.025

جدول (3) جدول مربع كاي (Chi Square χ^2)

df	0.995	0.99	0.975	0.95	0.9	0.1	0.05	0.025	0.01	0.005
1	---	---	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.01	0.02	0.051	0.103	0.211	4.605	5.991	7.378	9.21	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.86
5	0.412	0.554	0.831	1.145	1.61	9.236	11.07	12.833	15.086	16.75
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.69	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.18	2.733	3.49	13.362	15.507	17.535	20.09	21.955
9	1.735	2.088	2.7	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.94	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.92	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.3
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.66	5.629	6.571	7.79	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.39	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.26	9.591	10.851	12.443	28.412	31.41	34.17	37.566	39.997
21	8.034	8.897	10.283	11.591	13.24	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.26	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.98	45.559

توزيع النسبة بين تباين عينتين

يعتبر هذا التوزيع من التوزيعات الهامة التي تبحث في تجانس المجتمعات ونلجأ لحساب النسب بين التباينات وليس الفرق بينهما لسهولة دراسة النسب و تفسيرها.

فإذا كان S_1^2 هو تباين عينه عشوائية حجمها n_1 مسحوبة من مجتمع طبيعي $N(\mu_1,\sigma_1^2)$ ، وكان S_2^2 هو تباين عينه عشوائية أخرى حجمها n_2 ومسحوبة من مجتمع طبيعي آخر $N(\mu_2,\sigma_2^2)$ وكانت العينتان مستقلتان فإن المتغير:

$$\frac{(n_1-1)S_1^2}{\sigma_1^2} \sim \chi^2_{(n_1-1)} \quad , \frac{(n_2-1)S_2^2}{\sigma_2^2} \sim \chi^2_{(n_2-1)}$$

$$\frac{\dfrac{(n_1-1)S_1^2}{\sigma_1^2}/(n_1-1)}{\dfrac{(n_2-1)S_2^2}{\sigma_2^2}/(n_2-1)} \sim F\,(n_1-1)(n_2-1)$$

وبعد الاختصار للطرف الايسر نحصل على

$$F=\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \quad \sim \quad F(n_1-1,n_2-1)$$

واذا تساوى تبايني المجتمعين فان:

$$F=\frac{s_1^2}{s_2^2} \quad \sim \quad F(n_1-1,n_2-1)$$

حيث أن $F(n_1-1,n_2-1)$ تسمى بتوزيع F بدرجتي الحريه n_1-1 و

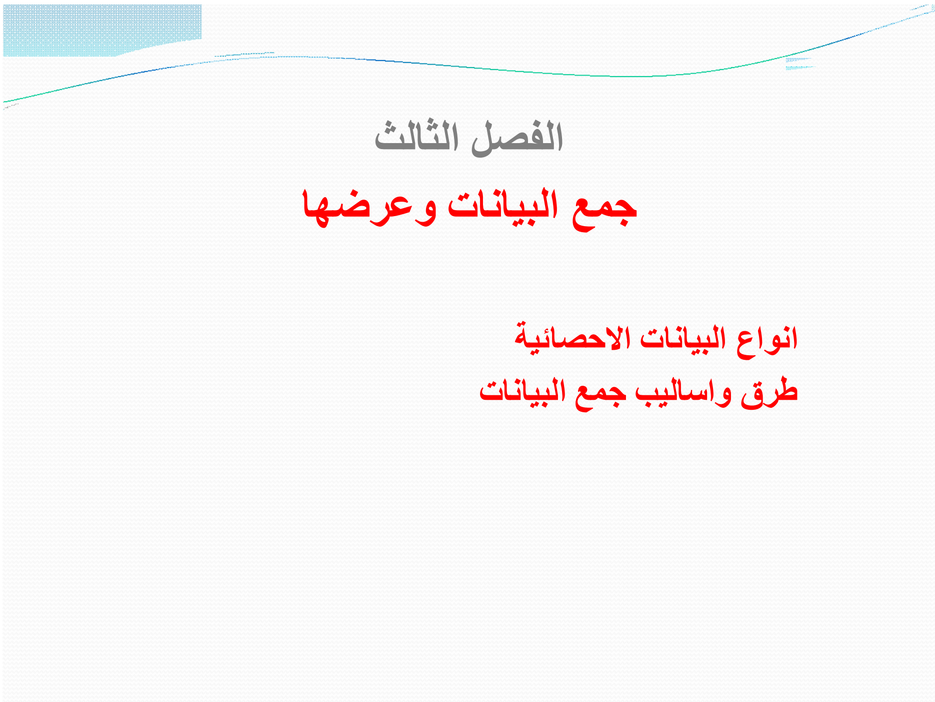
مثال : سحبت عينة حجمها 13 من مجتمع طبيعي تباينه ٩، وسحبت عينة أخرى حجمها21 من مجتمع طبيعي تباينه 25 ومستقل عن المجتمع الأول . اوجد احتمالية النسبة بين تبايني العينتين اقل من 0.8 .

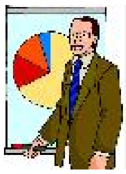
$$F=\frac{S_1^2/9}{S_2^2/25} \sim F_{(12,20)}$$

$$P\left(\frac{S_1^2}{S_2^2}<0.8\right)=P\left(\frac{S_1^2/9}{S_2^2/25}<0.8(\frac{25}{9})\right)$$

$$=P(F<2.22)=1-P(F>2.22)$$

$$=1-0.05=0.95$$





٣-١ انواع البيانات الاحصائية

البيانات الاحصائية هي المواد الاولية للاحصاء ، والتي من خلالها يمكن تبويبها وتصنيفها واجراء التحاليل عليها واستنباط التوصيات منها. وتعتمد طبيعة البيانات الاحصائية على الاهداف المطلوبة منها ، لذا فان هذا الفصل سيتناول مصادر البيانات وطرق جمعها.

ويمكن تقسيم البيانات الإحصائية إلى نوعين أساسيين من البيانات هما:

٣. **البيانات النوعية** Qualitative data:

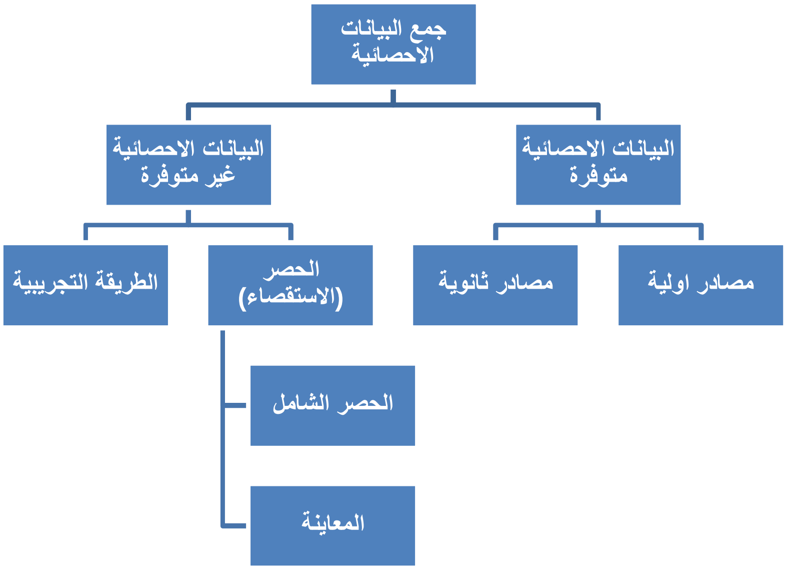
وهي بيانات وصفية، تشمل الظواهر التي لا تخضع للقياسات الكمية، ويصعب التعبير عنها بصورة عددية. كما تعرف أحياناً باسم (البيانات التصنيفية) لأنها تصنف البيانات حسب الصفات، سواء أكانت ذلك من حيث النوع، مثل تصنيف التربة إلى طينية وغرينية ورملية، أم كان ذلك من حيث الدرجة، مثل تصنيف القوى العاملة بحسب الدرجة التعليمية، إلى أمي، وملم بالقراءة والكتابة، حاصل على الشهادة الابتدائية...الخ، حيث تشتمل البيانات على عدد الأفراد الذين ينتمون إلى كل صفة من هذه الصفات.

٤. **البيانات الكمية** Quantitative data

وهي بيانات رقمية، تشمل الظواهر القابلة للقياسات الكمية، ويمكن التعبير عنها بصورة عددية، مثل كميات الأمطار، وأعمار السكان وكميات الإنتاج.. الخ، وغير ذلك من البيانات التي تعكس القيم الفعلية للظواهر. ومن الجدير بالذكر أن معظم الأساليب الإحصائية تعنى بمعالجة البيانات الرقمية، وهذا على خلاف البيانات النوعية، التي لا تستخدم فيها سوى بعض الأساليب الإحصائية المعدودة.

3-2 طرق واساليب جمع البيانات الاحصائية

تعتبر طريقة جمع البيانات من أهم المراحل التي يعتمد عليها البحث الإحصائي، كما أن جمع البيانات بأسلوب علمي صحيح، يترتب عليه الوصول إلى نتائج دقيقة في التحليل، وقد تكون البيانات المطلوبة متوفرة سواء كانت منشورة او غير منشورة نتيجة لجمعها من قبل جهات رسمية او غير رسمية وفي هذه الحالة فأن جمعها يتطلب الرجوع الى تلك المصادر . اما اذا كانت البيانات غير متوفرة فان الامر يتطلب اجراء استقصاءات او تجارب ميدانية لتوفيرها. وسيتناول هذا الفصل مصادر البيانات المتوفرة والطرق الاحصائية لجمع غير المتوفر منها.



ولدراسة طرق جمع البيانات، يجب الإلمام بالنقاط التالية:

1- مصادر البيانات المتوفرة

٢- أسلوب جمع البيانات الغير متوفرة.

١- مصادر جمع البيانات المتوفرة

هناك مصدرين للحصول منها على البيانات هما:

اولاً- **المصادر الأولية** Primary Source

ثانيًا- **المصادر الثانوية.** Secondary Source

أولا :المصادر الأولية

وهي المصادر التي نحصل منها على البيانات بشكل مباشر، حيث يقوم الباحث نفسه بجمع البيانات من المفردة محل البحث مباشرة، فعندما يهتم الباحث بجمع بيانات عن الأسر الريفية، يقوم بإجراء مقابلة مع رب الأسرة، ويتم الحصول منه مباشرة على بيانات خاصة بأسرته، مثل بيانات المنطقة التابع لها،وعدد افراد الاسرة واعمارهم والمستوى التعليمي ، ومساحة الملكية الزراعية ، وانواع المحاصيل المزروعة، واعداد وانواع الثروةالحيوانية التي يمتلكها،وانواع المكننة الزراعية التي يستخدمها ونوع ملكيتها (ملك شخصي ام ايجار) والدخل الشهري، ... وهكذا. او قد تقوم جهات رسمية او شبه رسمية بجمع انواع معينة من البيانات بشكل دوري او خلال فترة زمنية محددة كالجهاز المركزي للاحصاء او الهيئة العامة للانواء الجوية ، وقد تصدر هذه الجهات نشرات احصائية دورية.

ويتميز هذا النوع من المصادر بالدقة والثقة في البيانات، لأن الباحث هو الذي يقوم بنفسه بجمع البيانات من المفردة محل البحث مباشرة، ولكن أهم ما يعاب عليها أنها تحتاج إلى وقت ومجهود كبير، ومن ناحية أخرى فانها مكلفة من الناحية المادية.

ثانيا :المصادر الثانوية

وهي المصادر التي نحصل منها على البيانات بشكل غير مباشر، بمعنى آخر يتم الحصول عليها بواسطة أشخاص آخرين، أو أجهزة، وهيئات رسمية اخرى غير متخصصة بجمع البيانات الاحصائية، مثل نشرات منظمة الأغذية والزراعة الدولية....وهكذا.

ومن مزايا هذا النوع من المصادر، توفير الوقت والجهد والمال، إلا انها تحتوي على تفاصيل اقل من المصادر الاولية ، وأن درجة ثقة الباحث فيها ليست بنفس الدرجة في حالة المصادر الأولية.

2- أسلوب جمع البيانات الغير متوفرة

يتحدد الأسلوب المستخدم في جمع البيانات ، حسب الهدف من البحث، وحجم المجتمع محل البحث، وهناك أسلوبين لجمع البيانات هما:

١- الاستقصاء (المسح) Survey Method

٢- الطريقة التجريبية Experimental Method

١- الاستقصاء (المسح) Survey

ان مصطلح الاستقصاء بالمفهوم الاحصائي يعني جمع البيانات حول صفات وخصائص الاشياء القائمة دون التحكم باي من العوامل التي تؤثر على المتغير قيد الدراسة. فمثلا لو اردنا على سبيل المثال جمع البيانات حول اعداد ومواصفات الاغنام العراقية ، فان عملية جمع البيانات الاحصائية تتركز على عدد، جنس ، عمر ، ونوع الاغنام دون التحكم بالعوامل التي تؤثر فيها (السلالات ، الاعلاف ، الخدمات البيطرية ، الحظائر ...الخ.

اما درجة تغطية الاستقصاء لعناصر المجتمع فقد تكون :

أولاً- أسلوب الحصر الشامل (التعداد) Census

ثانياً - أسلوب المعاينة. Sampling

أولا :أسلوب الحصر الشامل (التعداد) Census

يستخدم هذا الأسلوب إذا كان الغرض من البحث هوحصر جميع مفردات المجتمع، وفي هذه الحالة يتم جمع بيانات عن كل مفردة من مفردات المجتمع بلا استثناء، كحصر جميع المزارع التي تنتج التمور، جميع الحقول المنتجة للدواجن، ويتميز أسلوب الحصر الشامل بالشمول وعدم التحيز، ودقة النتائج، ولكن يعاب عليه أنه

يحتاج إلى الوقت والجهد، والتكلفة العالية.

ويتم استخدام هذا الأسلوب في الحالات التالية:

٤- إذا كان الغرض من البحث جمع بيانات عن مفردات المجتمع بصفة شخصية أو فردية فإن الأسلوب الذي يتبع في هذه الحالة هو أسلوب الحصر الشامل. فلو كان الغرض من البحث مثلاً، جمع بيانات عن مجتمع الآلاتوالمكائن الزراعية الموجودة في محافظة ما، وذلك للوقوف على الآلات التي تحتاج إلى إصـــلاح وصـــيانة، فالبيانات المطلوبة في هذه الحالة تخص كل آلة على حده، وبالتالي فإنها مطلوبة بصفة فردية وعليه فإنه لابد من استخدام أسلوب الحصر الشامل لحصر كل آلة على حده.

٥- يمكن استخدام الحصر الشامل عندما يريد الباحث الحصول على نتائج على مستوى عال من الدقة. فمثلاً تقوم الشـركات المنتجة لأنابيب أفران الغاز بفحص هذه الأنابيب بأسلوب الحصـر الشـامل، وذلك للتأكد من ســـلامتها حتى لا يكون هناك أي احتمال لبيع أنابيب غير ســـليمة، وبالتالي تعريض حياة المواطنين للخطر. وكذلك شركات إنتاج الأدوية التي تحمل طابع الخطورة يطبق عليها أسلوب الحصر الشامل حتى تتمكن الشركة من التأكد من سلامتها.

٦- ويمكن استعمال أسلوب الحصر الشامل في حالة ما إذا كانت مفردات المجتمع المراد دراسته غير متجانسة وإذا ما كان المجتمع صغير نسبيًا.

ثانيا :أسلوب المعاينة Sampling

يعتمد هذا الأسلوب على معاينة جزء من المجتمع محل الدراسـة، يتم اختياره بطريقة علمية سليمة، ودراسته ثم تعميم نتائج العينة على المجتمع، ومن ثم يتميز هذا الأسلوب تقليل الوقت والجهد ، تقليل التكلفة ، والحصول على بيانات أكثر تفصيلا، وخاصة إذا جمعت البيانات من خلال استمارة استبيان.

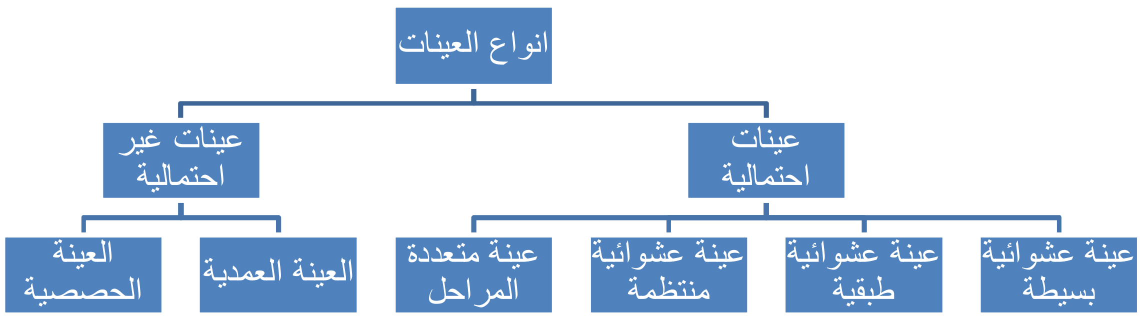
كما أن أسلوب المعاينة يفضل في بعض الحالات التي يصعب فيها إجراء حصر شامل، مثل معاينة دم المريض، أو إجراء تعداد لعدد الأسماك في البحر او اعداد حشرات الحميرة والدوباس التي تصيب النخيل.

ولكن يعاب على أساوب المعاينة أن النتائج التي تعتمد على هذا الأسلوب أقل دقة من نتائج أسلوب الحصر الشامل، وخاصة إذا كانت العينة المختارة لا تمثل المجتمع تمثيلا جيدا.

انواع العينات

ويمكن تقسيم العينات وفقا لأسلوب اختيارها إلى نوعين هما:

أ -العينات الاحتمالية ب -العينات غير الاحتمالية



أ- العينات الاحتمالية

هي العينات التي يتم اختيار مفرداتها وفقا لقواعد الاحتمالات، بمعنى آخر هي التي يتم اختيار مفرداتها من مجتمع الدراسة بطريقة عشوائية، بهدف تجنب التحيز الناتج عن اختيار المفردات،ومن أهم أنواع العينات الاحتمالية، ما يلي:

١- العينة العشوائية البسيطة Simple Random Sampling.

هي طريقة الاختيار التي تضمن احتمالا متساويا لكل عنصر من عناصر المجتمع بان يشمل في العينة واحتمالا متساويا لكل عينة من العينات التي يمكن تشكيلها من عناصر المجتمع.

مثال: اذا كان لدينا مجتمع مكون من٥ عناصر هي (أ ، ب ، ج ، د ، ه) واردنا اختيار عينة عشوائية بسيطة مكونة من ٣ عناصر، فان العينة التي يمـــكن اختيارها ستكون واحدة من العينات العشرة التي يمكن الحصول عليها من هذا المجتمع وهي:

أ ب ج أ ب د أ ب ه أ ج د أ ج ه

أ د ه ب ج د ب ج ه ب د ه ج د ه

ويجب الاشارة على ان العنصر لا يتكرر في اي عينةلان الترتيب غير مهم وذلك لان الترتيبات أ ب ج ، أ ج ب ، ب أ ج ، ب ج أ ، ج أ ب ، ج ب أ هي عينة واحدة تشمل على العنا صر أ . ب . ج .

٢- العينة العشوائية الطبقية Stratified Random Sampling.

وهي طريقة اختيار عينة طبقية عن طريق تقسيم المجتمع الى اقسام متجانسة تعرف بالطبقات ، ثم اختيار عينة عشوائية فرعية بصورة عشوائية من كل طبقة ، وهذه العينات الفرعية مجتمعة تكون العينة الطبقية.

مثال: يراد اختيار عينة طبقية تمثل حاصل غلة محصول الحنطة في الدونم الواحد في العراق...

اولا – نقسم العراق الى ٣ مناطق متجانسة (طبقات) حسب الظروف المناخية وهي المنطقة الشمالية والوسطى والجنوبية.

ثانيا – ناخذ عينة عشوائية في كل من المنطقة الشمالية والوسطى والجنوبية.

ثالثا – العينات الثلاث اعلاه تمثل العينة الطبقية.

٣ – العينة العشوائية المنتظمة Systematic Random Sampling.

وهي طريقة اختيار عينة منتظمة وذلك عن طريق اختيار وحدة المعاينة المرقمة K والتي تسمى نسبة المعاينة وهي حجم المجتمع الى حجم العينة ثم اختيار رقم عشوائي بين ١ و K ليكون رقم العينة الاولى ثم اضافة K ومضاعفاتها على رقم العينة الاولى الى ان يكمل حجم العينة.

مثال: يراد اختيار عينة منتظمة من شعبة أ للاول زراعي مكونة من خمسة طلاب علما بان عدد الطلبة في الشعبة ٤٠ طالبا.

اولا – نحسب نسبة المعاينة K

$$K = \frac{\text{حجم المجتمع}}{\text{حجم العينة}}$$

$$= \frac{\mathbf{40}}{\mathbf{5}} = \mathbf{8}$$

ثانيا - نختار رقم عشوائي بين ١ و ٨ فليكن مثلا الرقم ٦

ثالثا _ نختار العنصر الاول من العينة الطالب الذي تسلسله ٦ ثم نضيف نسبة المعاينة (٨) ومضاعفاتها على الرقم ٦ الى ان نكمل حجم العينة ، اي يكون اختيار الطلبة الذين تسلسلهم في الصف ٦ ، ١٤ ،٢٢ ، ٣٠ ، ٣٨.

٤– العينة المتعددة المراحل (العنقودية) Multi-Stage (Cluster) sampling

هي طريقة لاختيار عينة متعددة المراحل وذلك عن طريق اجراء الاختيار على مراحل متعددة . فاذا كان المجتمع مقسما الى اقسام فاننا في المرحلةالاولى نختارعشوائيا عينة من هذه الاقسام وفي المرحلة الثانية نختار عينة عشوائية من العينة التي اختيرت في المرحلة الاولى وهكذا حتى نحصل على حجم العينة المقرر.

مثال : لو اردنا معرفة غلة اصناف النخيل البرحي والخستاوي والمكتوم في محافظة ديالى . فلو اردنا ان نأخذ عينة مكونة من ١٠٠٠ نخلة من كل صنف من اصناف النخيل اعلاه وفق اسلوب المعاينة المتعددة المراحل ، نقوم بما يلي:

اولا- نختار ١٠ قرى من قرى محافظة ديالى التي تشتهر بزراعة النخيل.

ثانيا – نختار عشرة بساتين عشوائيا في كل قرية.

ثالثا – نختار ١٠ اشجار نخيل من كل صنف في كل بستان اختير اعلاه.

وبالتالي تكون حجم العينة لدينا مكونة من ١٠٠٠ نخلة لكل صنف

ب - العينات غير الاحتمالية

هي التي يتم اختيار مفرداتها بطريقة غير عشوائية، حيث يقوم الباحث باختيار مفردات العينة بالصورة التي تحقق الهدف من المعاينة وأهم أنواع العينات غير الاحتمالية:

١ - العينة العمدية   Judgmental Sample

وهي الطريقة التي تستخدم لاخذ عينة صغيرة لمجتمع كبير ، ويلجأ الباحثون الى هذه الطريقة عندما يكون المجتمع على درجةعالية من التجانس. لذا يعمد بعض الباحثيين الى استخدام هذا الاسلوب معتقدين بتمثيل هذه العينة للمجتمع. وغالبا ما يلجأ الباحث على اختيار العينات القريبةمن متوسط المجتمع.

مثال: اختيار قرية في السماوة لحصر الامراض التي تصيب الجمال في العراق

وهنا يعتقد الباحث بان الامراض التي تصيب الجمال في السماوة هي نفسها التي تصيب الجمال في جميع مناطق العراق.

٢ -العينة الحصصية   Quota Sample

وتستخدم عندما يكون المجتمع الاحصائي غير متجانس ومؤلف من طبقات. نقومباختيار عينة من كل طبقة بصورة موضوعية (غير عشوائية) يتناسب حجمها مع حجم الطبقة في المجتمع ، ثم ندمج هذه العينات في عينة واحدة تسمى العينة الحصصية.

٢- الطريقة التجريبية   Experimental Method

ان مصطلح التجربة يقصد به جمع البيانات عند ممارسة سيطرة فعلية على واحد او اكثر من العوامل التي تؤثر على المتغير قيد الدراسة ، وتستخدم هذه الطريقة لدراة العلاقة السببية اي المسبب والتأثير ، عن طريق التحكم في العوامل الداخلة فيالتجربة لدراسة تأثيرها منفردة اومجتمعة. مثلا عند دراسة تأثير نوع معين ومستويات من السماد على انتاج محصول معين مع تثبيت العوامل الاخرى اللازمة لانتاج المحصول او تغيير بعضها كصنف البذور وطريقة الزراعة وعدد الريات ... الخ ، وتعتبر الطريقة التجريبية اقوى طرق البحث.

تمارين الفصل الرابع

١- ما الفرق بين المصادر الاولية والثانوية للبيانات ، وأيهما يتصف بالدقة اكثر؟

٢- ما هي الاساليب المتبعة في جمع البيانات الغير متوفرة ، عددها واشرحها مفصلا ؟

٣- عدد اساليب اخذ العينات ، مع مثال لكل اسلوب؟

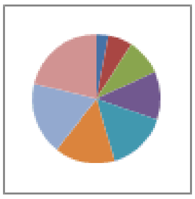
الفصل الرابع

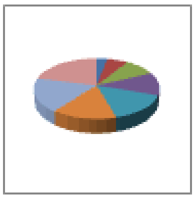
العرض الجدولي والتمثيل البياني

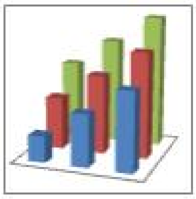
4.1 انواع الجداول

4.2 جداول التوزيع التكراري

4.3 التمثيل البياني







١ـ ٤ انواع الجداول

بعد جمع البيانات الاحصـائية الاولية (Raw data) لدراسـة ظاهرة معينة وفق الاساليب والطرق التي ذكرناها سابقا فانه غالبا لا يمكن الاستفادة منها وهي على هذه الصورة ، لذا فغالبا ما توضع في جداول مبسطة او يعبر عنها بصورة اشكال ورسوم بيانية لتسهيل عملية دراستها وتحليلها.

العرض الجدولي Tabular presentation

هناك نوعان رئيسيان من الجداول الاحصائية وهما:

١- الجدول البسيط ٢- الجدول المركب

الجدول البسيط : وهو الجدول الذي توزع فيه البيانات حسب صفة واحدة ويتألف من عمودين : الاول يمثل تقسيمات الصفة او الظاهرة الى فئات او مجموعات والثاني يبين عدد المفردات التابعة لكل فئة او مجموعة. الجدول ٩-١ و٩-٢ يمثل نموذج للجدول البسيط.

جدول توزيع الدرجات الفصلية لمادة الاحصاء للصف الاول زراعي.

فئات الدرجات	عدد الطلبة
٣١-٤٠	٢
٤١-٥٠	٤
٥١-٦٠	٧
٦١-٧٠	١٢
٧١-٨٠	١٨

٨١-٩٠	٦
٩١-١٠٠	١

جدول توزيع عدد الطلبة للمرحلة الاولى في كلية الزراعة حسب الاقسام العلمية

القسم العلمي	عدد الطلبة
علوم التربة	٧٠
علوم المحاصيل	٥٠
وقاية المزروعات	٦٠
البستنة	٥٠
الصناعات الغذائية	٣٠
الارشاد الزراعي	٢٥
الاقتصاد الزراعي	٢٠
المكننة الزراعية	٣٥
الثروة الحيوانية	٦٥

الجدول المركب: وهو الجدول الذي توزع فيه البيانات حسب صفتين او ظاهرتين او اكثر في نفس الوقت. فتكون الفئات او المجاميع للصفة او الظاهرة الاولى تمثل الصفوف والصفة او الظاهرة الثانية تمثل الاعمدة ، بينما تمثل الخلايا(المربعات) الداخلية اعداد المفردات لهذه الصفات او التكررات . والجدول ٩-٣ يمثل نموذج للجدول المركب المتكون من صفتين.

جدول توزيع عدد الطلبة في كلية الزراعة حسب المراحل الدراسية والاقسام العلمية

القسم العلمي	المرحلة الاولى	المرحلة الثانية	المرحلة الثالثة	المرحلة الرابعة
علوم التربة	٧٠	٤٥	٣٤	٣٦
علوم المحاصيل	٥٠	٤٠	٣٧	٤٦
وقاية المزروعات	٦٠	٥٥	٤٥	٤٨
البستنة	٥٠	٥٦	٤٤	٣٦
الصناعات الغذائية	٣٠	٢٥	٢٨	٣٢
الارشاد الزراعي	٢٥	١٨	٢٠	٢٤
الاقتصاد الزراعي	٢٠	٢٥	١٨	١٩
المكننة الزراعية	٣٥	١٨	١٧	١٢
الثروة الحيوانية	٦٥	٥٠	٤٠	٤٢

والآن سنشرح بالتفصيل كيفية انشاء او تكوين جدول من الجداول البسيطة الكثيرة والشائعة الاستعمال ويدعى بجدول التوزيع التكراري Frequency table.

4-2 جداول التوزيع التكراري Frequency table

تنظم وتلخص البيانات الاحصائية سواء كانت وصفية ام كمية فيما يسمى بالتوزيع التكراري (Frequency Distribution) وهو عبارة عن جدول يلخص البيانات الخام فيوزعها الى فئات ويحدد عدد الافراد الذين ينتمون الى كل فئة ويسمى هذا العدد بتكرار الفئة ويرمزله f ، ولاتمام ذلك ينبغي ان يصمم جدول آخر يسمى بجدول تفريغ البيانات الاحصائية. وهو يتكون من ثلاث اعمدة: العمود الاول يكتب فيه الصفة للبيانات الوصفية اوالفئة للبيانات الكمية ، وفي العمود الثاني توضع العلامات وهي عبارة عن حزم مكونة من خمسة خطوط اربعة منها رأسية والخامس مائل يربط الخطوط الاربعة الرأسية وبذلك تصبح الحزمة على الصورة (𝍸) وفي العمود الثاث يكتب مجموع العلامات امام كل صفة اوفئة ومجموع هذه العلامات في كل فئة يسمى التكرار لهذه الصفة او الفئة.

هناك عدة انواع من جداول التوزيع التكراري اهمها:

١- جدول التوزيع التكراري البسيط
٢- جدول التوزيع التكراري النسبي
٣- جدول التوزيع التكراري المتجمع
أ- جدول التوزيع التكراري المتجمع الصاعد
ب- جدول التوزيع التكراري المتجمع النازل

1- جدول التوزيع التكراري البسيط Simple Frequency table

تعريف جدول التوزيع التكراري

هوجدول بسيط يتكون من عمودين :

الاول : تقسم فيه قيم المتغير الى اقسام او مجموعات تسمى الفئات Classes

بعض التعاريف المهمة

البيانات غير المبوبة Ungrouped data

وهي البيانات الخام الاولية او الاصلية (Raw data) التي جمعت ولم تبوب.

البيانات المبوبة Grouped data

وهي البيانات التي تم تبويبها وتنظيمها في جدول التوزيع التكراري.

الفئات Classes

وهي الفترة التي نختارها لتقسيم بيانات المتغير الى مجموعات متساوية بحيث تكون لكل قسم او صنف صفة مميزة.

حدودالفئات Class limit

لكل فئة حدان حد ادنى وحد اعلى.

طول الفئة class length

وهو مقدار المدى بين حدي الفئة.

مركزالفئة Class midpoint

منتصف المدى بين حدي الفئة.

تكرار الفئة Class frequency

وهي عدد المفردات اوالقيم التي تقع تحت مدى تلك الفئة ويرمزلها fi

خطوات تكوين جدول تكراري:

أ- في حالة البيانات الوصفية

١- انشاء جدول تفريغ البيانات الذي يتكون من عمودين الاول يتضمن الصفة للبيانات الوصفية والثاني للعلامات والثالث للتكرار لكل صفة.

٢ – انشاء حدول التوزيع التكراري الذي يتضمن عمودين ، الاول للصفات الوصفية والثاني لتكرار الفئات .

مثال: اذا كانت لدينا بيانات انواع ٤٠ شجرة من الحمضيات في احد البساتين

برتقال	نارنج	برتقال	ليمون	لالنكي
نارنج	لالنكي	ليمون	نارنج	ليمون
برتقال	برتقال	ليمون	لالنكي	ليمون
برتقال	ليمون	برتقال	نارنج	برتقال
ليمون	برتقال	ليمون	نارنج	لالنكي
برتقال	برتقال	ليمون	لالنكي	ليمون

اولا: نعمل جدول تفريغ البيانات كالآتي

جدول 1-٢-9. جدول تفريغ البيانات لاشجار الحمضيات

انواع اشجار الحمضيات	العلامات	التكرار
نارنج	𝍸 |	٦
برتقال	𝍸 𝍸 𝍸	١٥
ليمون	𝍸 𝍸 |	١١
لالنكي	𝍸 |||	٨

ثانيا: جدول التوزيع التكراري لاشجار الحمضيات

جدول التوزيع التكراري لاشجار الحمضيات

انواع اشجار الحمضيات	التكرار
نارنج	٦
برتقال	١٥
ليمون	١١
لالنكي	٨

ب- في حالة البيانات الكمية

١- ترتيب البيانات تصاعديا اوتنازليا

٢- حساب قيمة المدى

قيمة المدى = اعلى قيمة - ادنى قيمة

٣- اختيار عدد مناسب للفئات :

هناك عدة طرق حسابية تقريبية لايجاد عدد الفئات سنذكرها للعلم فقط اهمها:

طريقة سترجس Sturges: عدد الفئات = ١ + (٣,٣ + لوغارتم عدد المفردات)

طريقة يول Yule: عدد الفئات = ٢,٥ x الجذر الرابع لعدد المفردات

ولكل من الطريقتين ميزات وعيوب ولن نستعمل اي منها بل سنختار الفئات اختيارا ويفضل ان يكون العدد بين خمسة الى خمسة عشر فئة.

٤- حساب طول الفئة

طول الفئة = المدى / عدد الفئات

ويفضل تقريبها الى اقرب عدد صحيح

٥- حساب مركز الفئة

مركز الفئة = (بداية الفئة + نهاية الفئة) / ٢

مثال

البيانات التالية تمثل درجات ٤٠ طالبا في مادة الرياضيات

٦٥	٧٢	٤٥	٥٥	٤٢	٨٧	٣٦	٨٤
٥٨	٦٦	٧٩	٢٨	٨٨	٧٦	٦٢	٩١
٢٩	٧٨	٦٧	٨٣	٩٣	٥٦	٦٤	٦١
٤٦	٧٤	٦٣	٣٣	٧١	٨٤	٧٣	٨٠

الحل : ترتيب البيانات تصاعديا او تنازليا

٢٥ ٢٨ ٢٩ ٣٣ ٣٦ ٤٢ ٤٥ ٤٦

٥٠ ٥١ ٥٥ ٥٦ ٥٨ ٥٨ ٦١ ٦٢

٦٣ ٦٤ ٦٤ ٦٥ ٦٧ ٦٦ ٧١ ٧٢

٧٣ ٧٤ ٧٤ ٧٦ ٧٨ ٧٩ ٨٠ ٨٣ ٨٤ ٨٤ ٨٥

٨٧ ٨٨ ٩١ ٩٣ ٩٤

المدى = ٩٤ - ٢٥ = ٦٩

عدد الفئات = ٧ (اختياري)

طول الفئة = ٦٩ / ٧ = ٩,٨٦ تقرب الى اقرب عدد صحيح (١٠)

جدول تفريغ البيانات لدرجات مادة الرياضيات

الفئات	مركز الفئات	العلامات	التكرار
٢٥ - ٣٤	29.5	||||	٤
٣٥ - ٤٤	39.5	||	٢

٤٥ - ٥٤	49.5		٤
٥٥ - ٦٤	59.5		٩
٦٥ - ٧٤	69.5		٨
٧٥ - ٨٤	79.5		٧
٨٥ - ٩٤	89.5		٦
المجموع			٤٠

جدول التوزيع التكراري البسيط لدرجات مادة الراضيات

الفئات	التكرار
٢٥ - ٣٤	٤
٣٥ - ٤٤	٢
٤٥ - ٥٤	٤
٥٥ - ٦٤	٩
٦٥ - ٧٤	٨
٧٥ - ٨٤	٧
٨٥ - ٩٤	٦
المجموع	٤٠

٢- جدول التوزيع التكراري النسبي Relative frequency distribution

وهو جدول يبين الاهميةالنسبية لكل فئة . ويحسب التكرار النسبي لكل فئة بالطريقة التالية:

$$\text{التكرار النسبي لأي فئة} = \frac{\text{تكرار تلك الفئة}}{\text{المجموع الكلي للتكرارات}} = \frac{f_i}{\sum f_i}$$

وعادة يوضع التكرار النسبي كنسبة مئوية وذلك بضرب كل تكرار نسبي x ١٠٠ وكما مبين في الجدول .

جدول التوزيع التكراري النسبي لدرجات مادة الرياضيات

الفئات	التكرار	التكرار النسبي	التكرار المئوي
٢٥ ــ ٣٤	٤	0.100	**١٠**
٣٥ ــ ٤٤	٢	0.050	**٥**
٤٥ ــ ٥٤	٤	0.100	**١٠**
٥٥ ــ ٦٤	٩	0.225	**22.5**
٦٥ ــ ٧٤	٨	0.200	**٢٠**
٧٥ ــ ٨٤	٧	0.175	**17.5**
٨٥ ــ ٩٤	٦	0.150	**١٥**
المجموع	٤٠	**1.000**	١٠٠

3- جدول التوزيع التكراري المتجمع **Cumulative frequenc Table**

في بعض الاحيان قد تكون هناك حاجة الى معرفة عدد القيم او المفردات التي تقل اوتزيد عن قيمة محددة ، والجداول التي تحوي على مثل هذه المعلومات تدعى بالجداول التكرارية المتجمعة ، وهناك نوعين من هذه الجداول

أ- جدول التوزيع التكراري المتجمع الصاعد

ب-جدول التوزيع التكراري المتجمع النازل

أ- جدول التوزيع التكراري المتجمع الصاعد

وهو الجدول الذي يعطينا عدد المفردات التي تقل قيمتها عن الحد الادنى لفئة معينة ، ويتكون من عمودين. العمود الاول نكتب فيه حدود الفئات كما موضح في الجدول ٩-٦ . والعمود الثاني نكتب فيه التكرار التجميعي التصاعدي بالشكل التالي:

تكرار ما قبل الفئة الاولى $= f_0 =$ صفر

تكرارالفئة الاولى $= f_1 = f_1$

تكرارالفئة الثانية $= f_2 = f_1 + f_2$

تكرارالفئة الثالثة $= f3 = f1 + f2 + f3$

وهكذا بحيث ان التكرار التجميعي التصاعدي للفئة الاخيرة $= f_n = \Sigma_f$

جدول التوزيع التكراري التجميعي التصاعدي لدرجات مادة الرياضيات.

حدود الفئات	التكرار التجميعي التصاعدي
اقل من ٢٥	صفر
اقل من ٣٥	٤
اقل من ٤٥	٦
اقل من ٥٥	١٠
اقل من ٦٥	١٩
اقل من ٧٥	٢٧
اقل من ٨٥	٣٤
اقل من ٩٥	٤٠

ب-جدول التوزيع التكراري المتجمع النازل

وهو الجدول الذي يعطينا عدد المفردات التي تزيد قيمتها عن الحد الادنى لفئة معينة ، ويتكون من عمودين. العمود الاول نكتب فيه حدود الفئات كما موضح في الجدول ١٠-٦ . والعمود الثاني نكتب فيه التكرار التجميعي التنازلي بالطريقة التالية:

تكرار الفئة الاولى = f_1 = $f_i\Sigma$

تكرار الفئة الثانية = f_2 = مجموع التكرارات - تكرار الفئة الاولى

= $f_i\Sigma$ - f_1

تكرار الفئة الثالثة = f_3 = $f_i\Sigma$ - f_1 - f_2

وهكذا بحيث ان التكرار التجميعي التصاعدي للفئة الاخيرة = صفر

جدول التوزيع التكراري التجميعي التنازلي لدرجات مادة الرياضيات.

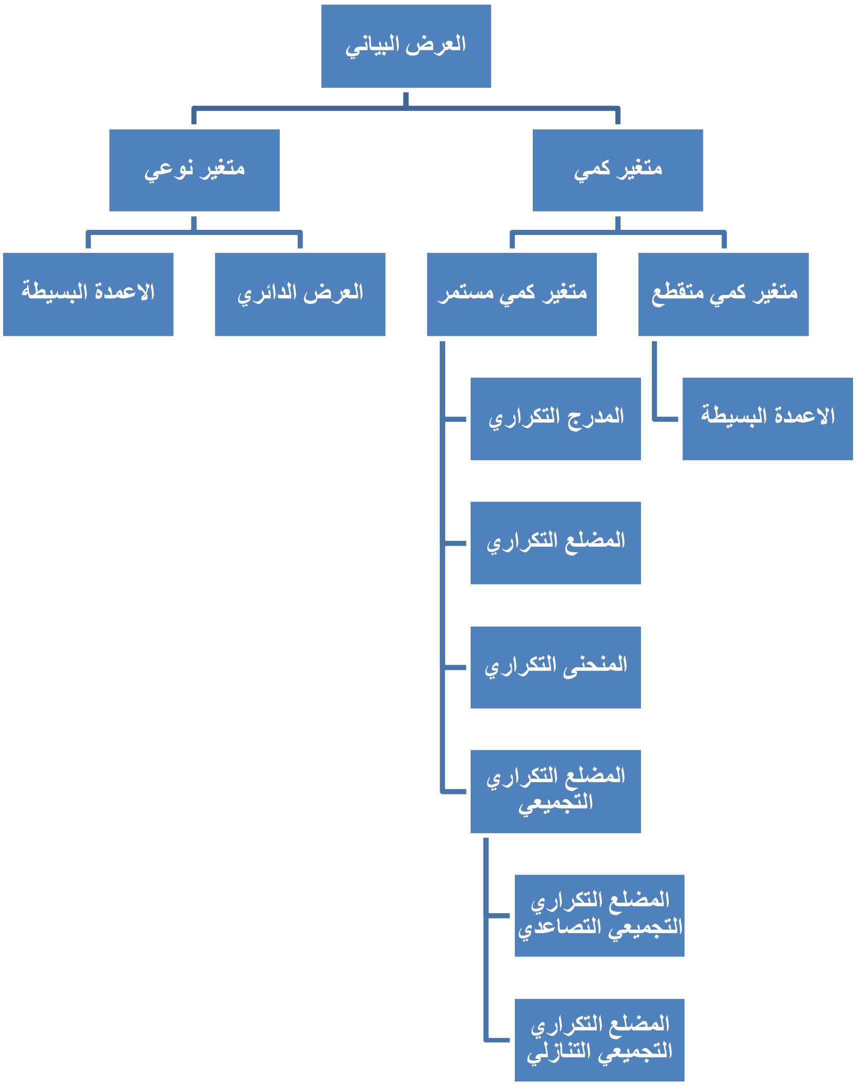
حدود الفئات	التكرار التجميعي التصاعدي
٢٥ فأكثر	٤٠
٣٥ فأكثر	٣٦
٤٥ فأكثر	٣٤
٥٥ فأكثر	٣٠
٦٥ فأكثر	٢١
٧٥ فأكثر	١٣

٨٥ فأكثر	٦
٩٥ فأكثر	صفر

4-3 التمثيل البيانى Graphic Presentations

بالرغم من أن التوزيع التكراري أساسي وفعال في إظهار طبيعة البيانات وعلاقاتها إلا أن الرسم البياني يبين طبيعة البيانات وأهميتها بصورة أسرع للقارئ بطريقة سهلة وجذابة وفعالة تساعده على فهم واستيعاب قيم الظاهرة او الصفة للمتغير تحت الدراسة ومقارنتها مع بعضها.

وتستخدم أنواع مختلفة للعرض البياني حسب نوعية المتغير المدروس. كما هوموضح في الشكل



مخطط لطريقةالتمثيل البياني حسب نوع المتغير

1-٣-4 العرض البياني في حالة متغير كمي

أ- العرض البياني في حالة متغير كمي متقطع

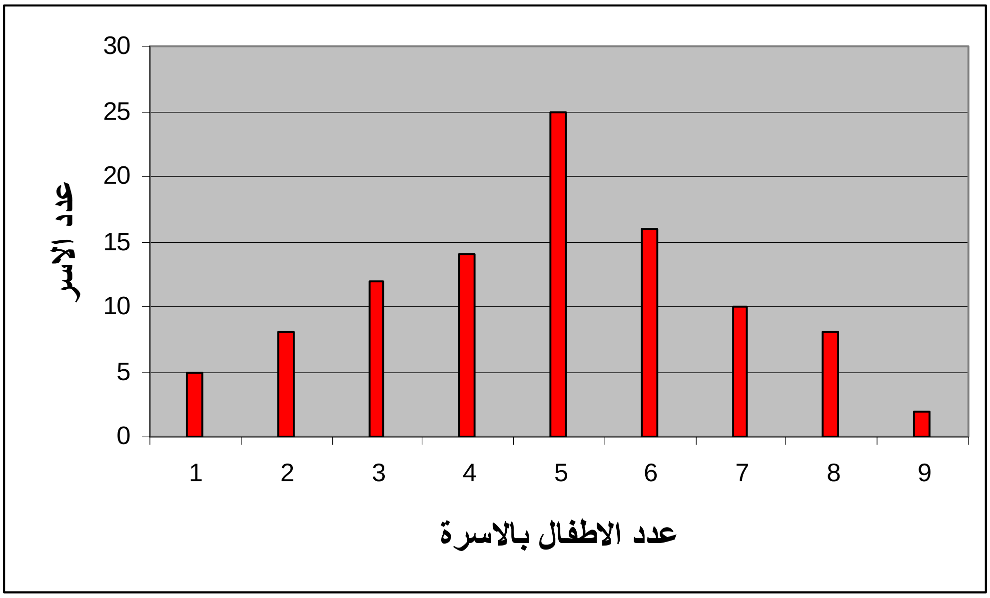
وهو المتغير الذي يأخذ أعداد صحيحة فقط مثل عدد افراد الاسرة ، عدد الطلبة ، عدد اشجار النخيل ، عدد الابقار في مزرعة ما ....الخ . وهنا يمكن استخدام الأعمدة البسيطة، وهي عبارة عن أعمدة بسيطة تتناسب أطوالها مع التكرار المقابل لقيمة العينة للمتغير المدروس.

مثال : يبين الجدول 1-٣-9 عدد الأطفال في العائلة لعينة تكون من ١٠٠ أسرة، المطلوب عرض هذه البيانات بطريقة العرض المناسب البسيط.

جدول1-٣-9. جدول بيانات عدد الاطفال في كل اسرة

عدد الأطفال في كل أسرة	عدد الأسر
١	٥
٢	٨
٣	١٢
٤	١٤
٥	٢٥
٦	١٦
٧	١٠
٨	٨
٩	٢
المجموع	١٠٠

الحل: افضل وابسط طريقة لعرض هذه البيانات هي باستخدام الاعمدة البسيطة



شكل 2-٣-9. الرسم البياني بطريقة الاعمدة البسيطة لعدد الاطفال بالاسرة

ب-العرض البياني في حالة متغير كمي مستمر

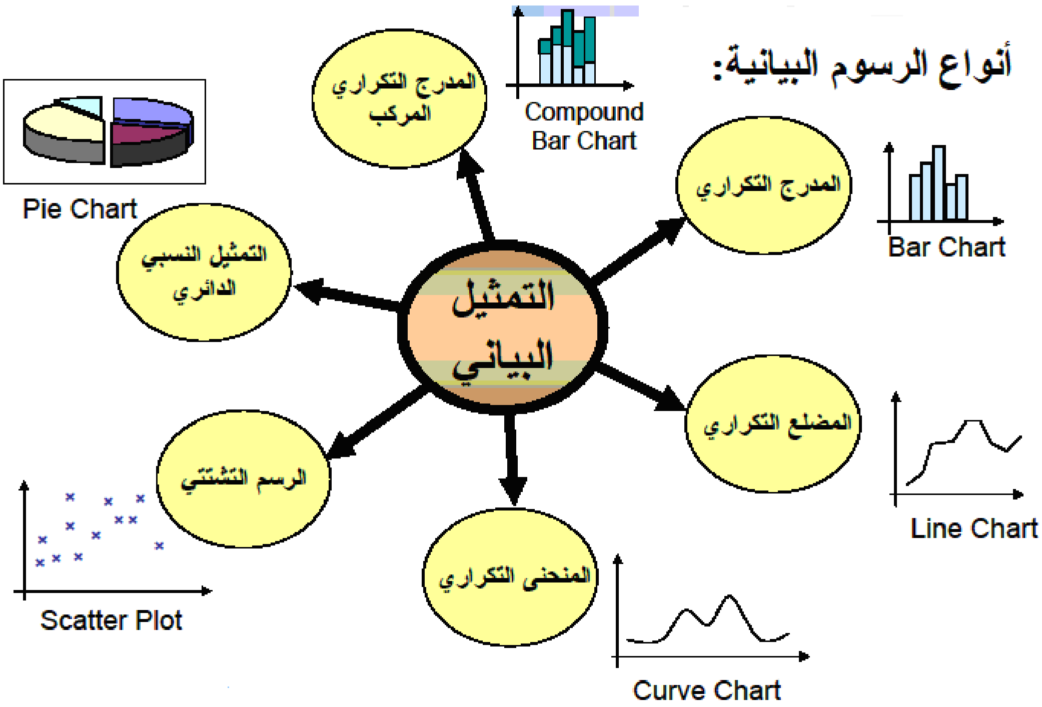
المتغير المســتمر وهو المتغير الذي يمكن أن يأخذ أي قيمة بين قيمتين معنيتين، وكأمثلة عن المتغيرات المتصلة: الطول، الوزن، الزمن، السرعة ...الخ . وهنا يمكن استخدام الأشكال التالية :

المدرج التكراري Frequency Histogram

المضلع التكراري Frequency Polygon

المنحنى التكراري Frequency Curve

المضلع التكراري المتجمع Cumulative Frequency Polygon



المدرج التكراري Frequency Histogram

وهو عبارة عن مستطيلات متلاصقة طول كل منها يتناسب مع التكرار المقابل، وقاعدة كل منها تساوي طول الفئة المقابلة. لأجل تمثيل البيانات بالمدرج التكراري ينبغي أولاً رسم محورين متعامدين الأفقي منها يمثل الفئات والرأسي يمثل التكرارات ، وعلينا أن نجزئ المحور الأفقي إلى وحدات متساوية

ونعين عليه الحدود الحقيقية للفئات ، ونجزئ المحور الرأســي على عدد التكرارات الواردة في الجدول .

والمدرج التكراري عبارة عن تمثيل كل فئة من الفئات بمســتطيل تمثل قاعدته الحدود الحقيقية لتلك الفئات وارتفاعه يســـاوي التكرار المقابل لها ، ومن الملاحظ أن الحد الأعلى الحقيقي للفئة الأولى هو نفس الحد الأدنى الحقيقي للفئة الثانية ، وبذا ترى جميع المستطيلات متلاصقة.

مثال

اخذت عينة متكونة من ١٠٠ دجاجة بعمر ٤٥ يوم اخذت من احد حقول الدواجن والجدول ٩-٣-٢ يبين التوزيع التكراري لأوزان الدجاج بالغرام.

جدول 2-٣-9. جدول التوزيع التكراري لاوزان الدجاج (غم)

فئات الوزن	التكرار
١٠٠٠-٨٠٠	٨
١٢٠٠-١٠٠٠	٢٤
١٤٠٠-١٢٠٠	٤٠
١٦٠٠-١٤٠٠	١٨
١٨٠٠-١٦٠٠	١٠
المجموع	١٠٠

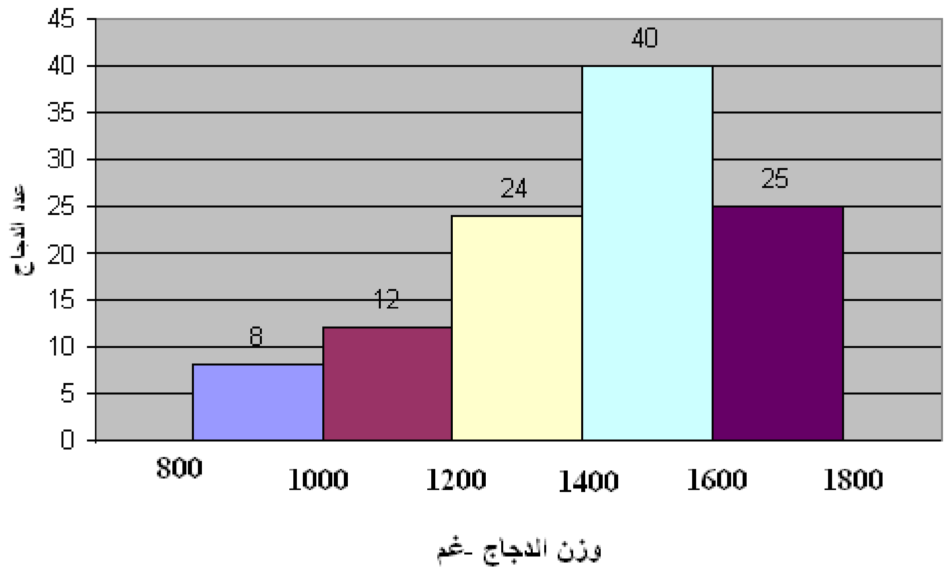
الحل : لرسم المدرج التكراري يتم إتباع الخطوات التالية:

رسم محوران متعامدان، الرأسي ويمثل التكرارات، الأفقي ويمثل الأوزان .

كل فئة تمثل بعمود ارتفاعه هو تكرار الفئة، وطول قاعدته هو طول الفئة .

كل عمود يبدأ من حيث انتهى به عمود الفئة السابقة

والشكل 3-3-9 يبين المدرج التكراري لأوزان الدجاج



المدرج التكراري لاوزان الدجاج

المضلع التكراري **Frequency polygone**

هو مجموع من قطع مستقيمة متصلة ومنكسرة تتحدد بنقاط أحداثياتها مركز الفئة والتكرارات المقابلة.

ملاحظة: الخط المنكسر يمثل المضلع التكراري، حيث أن المساحة التي تقع تحت المضلع التكراري تساوي المساحة التي تقع تحت المدرج التكراري، وحتى نحافظ على المساحة التي تقع أسفل هذا المضلع، وذلك بقفل المضلع بان نصل

بداية المضلع بالمحور الافقي بمركزفئة (خيالية) واقعة الى يسار اول فئة تكرارها صفرا□.ونصل نهاية المضلع بالمحور الافقي بمركز فئة (خيالية) واقعة الى يمين اخر فئةتكرارها ايضا□ صفر .

لرسم بيانات الجدول 2-3 -9 نقوم بما يلي:

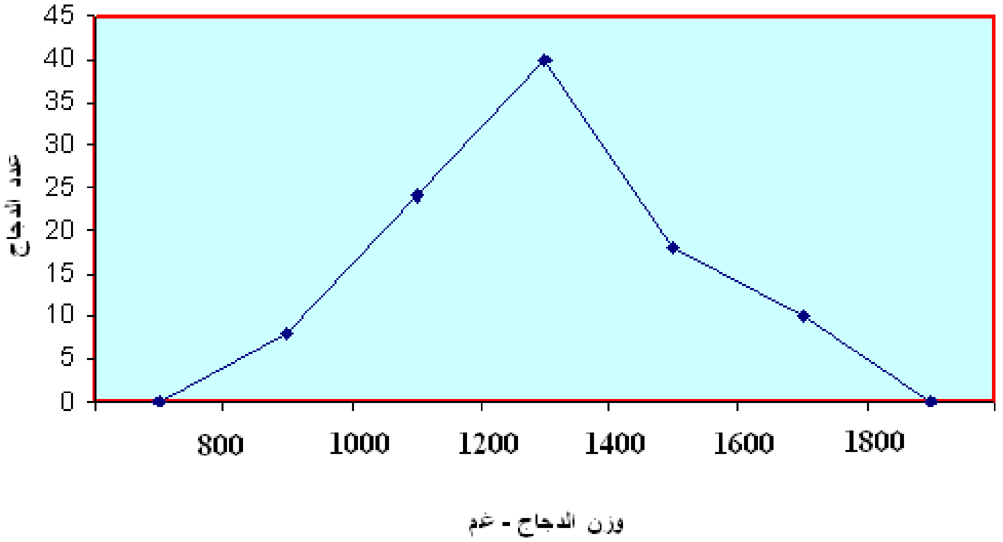
١- نقوم برسم المحور الافقي والعمودي

٢- تدريج المحور الافقي الى اقسام متساوية بحيث يشمل على جميع مراكز الفئات ، ويقسم المحور العمودي الى اقسام متساوية بحيث تشمل كل التكرارات.

٣- وضع نقطة امام مركزكل فئة وارتفاعها يعادل تكرار تلك الفئة.

٤- توصيل تلك النقاط بخطوط مستقيمة.

والشكل يمثل المضلع التكراري لبيانات الجدول 2-٣-9.

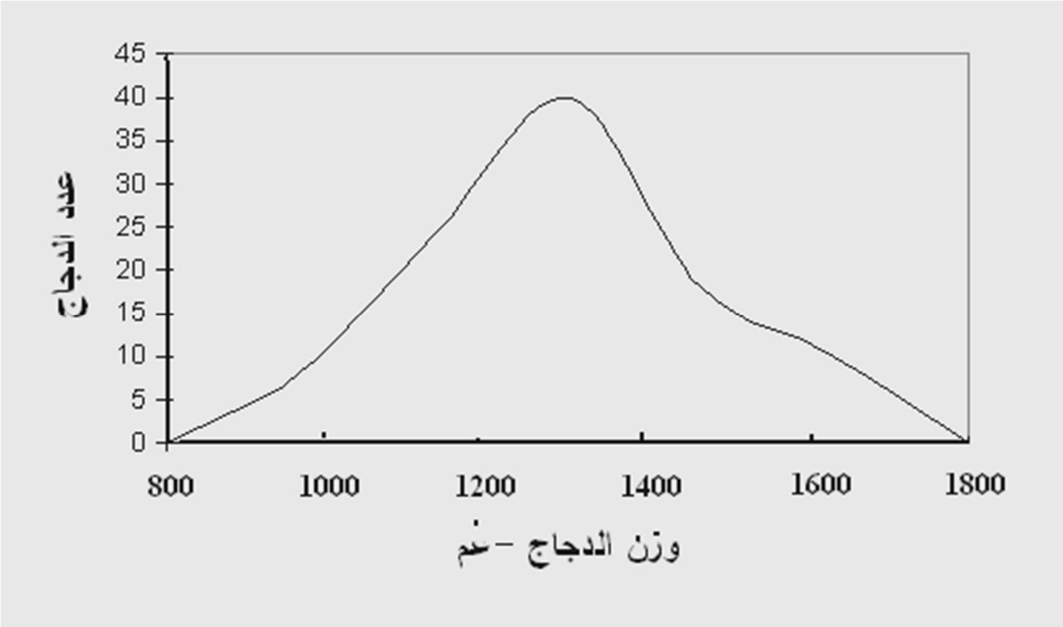


المضلع التكراري لبيانات الجدول 2-٣-9

المنحنى التكراري Frequency curve

بإتباع نفس الخطوات السابقة في رسم المضلع يمكن رسم المنحنى التكراري، ولكن يتم تمهيد الخطوط المنكسرة في شكل منحنى بحيث يمر بأكثر عدد من النقاط على مراكز الفئات والتي ارتفاعاتها تمثل تكرارات تلك الفئات.

وعادة يقفل المنحنى التكراري بان نصـــل بدايته بالحد الادنى للفئة الاولى ونهايته بالحد الاعلى للفئة الاخيرة. وتكون مساحة المنحنى مكافئة (وليست مساوية) للمضلع التكراري. كما في الشكل 5-٣-9.



شكل ٩-٣-٥ . المنحنى التكراري لبيانات الجدول ٩-٣-٢

المضلع التكراري المتجمع Cumulative Frequency Polygon

لتمثيل التكرار التجميعي بيانيا نستخدم المضلع التكراري التجميعي،وهوعبارة عن خطوط متكسرة تصل بين نقاط واقعةفوق الحدود الحقيقية للفئات وعلى ارتفاع تمثل التكرار التجميعي. وهناك نوعان من المضلع التكراري المتجمع:

أ- المضلع التكراري التجميعي الصاعد

ب-المضلع التكراري التجميعي النازل

أ- المضلع التكراري التجميعي التصاعدي

ولرسم المضلع التكراري التجميعي التصاعدي نتبع الخطوات التالية:

١- رسم المحور الافقي والعمودي.

٢- تدريج المحور الافقي الى اقسام متساوية تشمل على جميع حدود الفئات ويقسم المحور العمودي الى اقسام متساوية بحيث تشمل على اكبر التكرارات التجميعية وهي المجموع الكلي للتكرارات.

٣- وضع علامة امام كل حد فئة ارتفاعها يعادل التكرار التجميعي التصاعدي لذلك الحد.

٤- توصيل تلك النقاط بخطوط مستقيمة

مثال

ارسم المضلع التكراري التجميعي التصاعدي للجدول ٩-٣-٣

جدول ٩-٣-٣. جدول توزيع تكراري لاوزان ١٠٠ دجاجة.

فئات الوزن	التكرار
٨٠٠-١٠٠٠	٨

١٢٠٠-١٠٠٠	٢٤
١٤٠٠-١٢٠٠	٤٠
١٦٠٠-١٤٠٠	١٨
١٨٠٠-١٦٠٠	١٠
المجموع	١٠٠

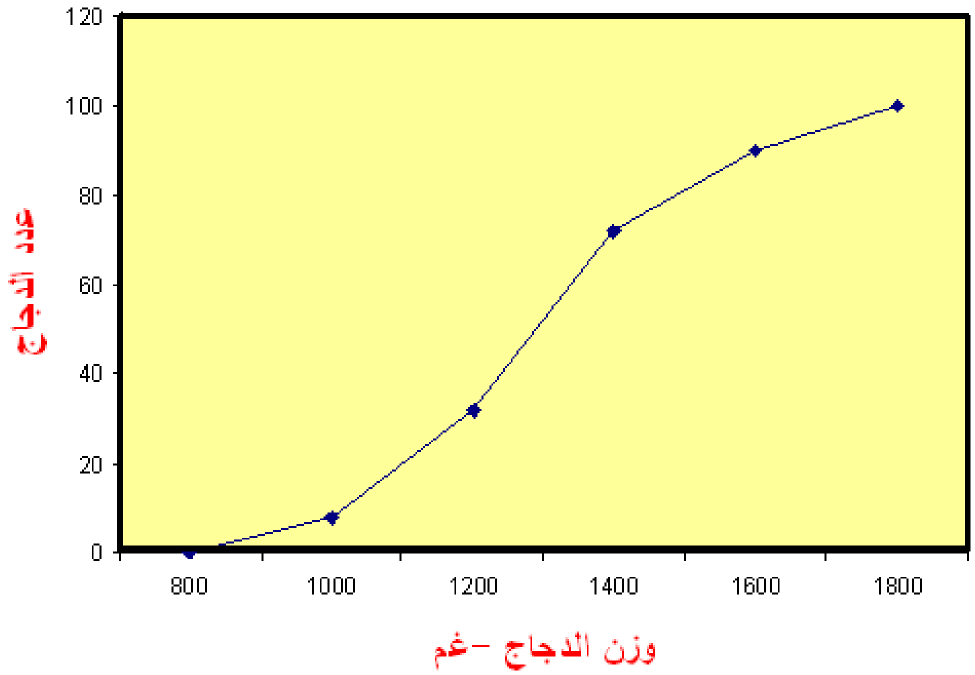
الحل: لتكوين الجدول التكراري المتجمع الصاعد، يتم حساب مجموع التكرارات (عدد القيم) التي تقل عن كل حد من حدود الفئات.وكما يلي

جدول ٩-٣-٤. جدول التكرار التجميعي التصاعدي

فئات الوزن	**التكرارالتجميعي التصاعدي**
اقل من ٨٠٠	٠

اقل من ١٠٠٠	٨
اقل من ١٢٠٠	٣٢
اقل من ١٤٠٠	٧٢
اقل من ١٦٠٠	٩٠
اقل من ١٨٠٠	١٠٠
المجموع	١٠٠

ثم رسم المضلع التكراري التجميعي التصاعدي وكما يلي:



شكل ٩-٣-٦ المضلع التكراري التجميعي التصاعدي لبيانات الجدول ٩-٣-٤.

ب- المضلع التكراري التجميعي التنازلي

ويرســم بنفس طريقة المضــلع التكراري التجميعي التصــاعدي ما عدا كون ارتفاع النقاط هنا هو التكرار التجميعي التنازلي ، ولذلك فيبدأ المضـلع التكراري التجميعي التنازلي من اعلى نقطة (مجموع التكرارات الكلي) وينتهي بالصفر اي عكس المضلع التكراري التجميعي التصاعدي تماما.

مثال

ارسم المضلع التكراري التجميعي التنازلي للجدول ٩-٣-٣

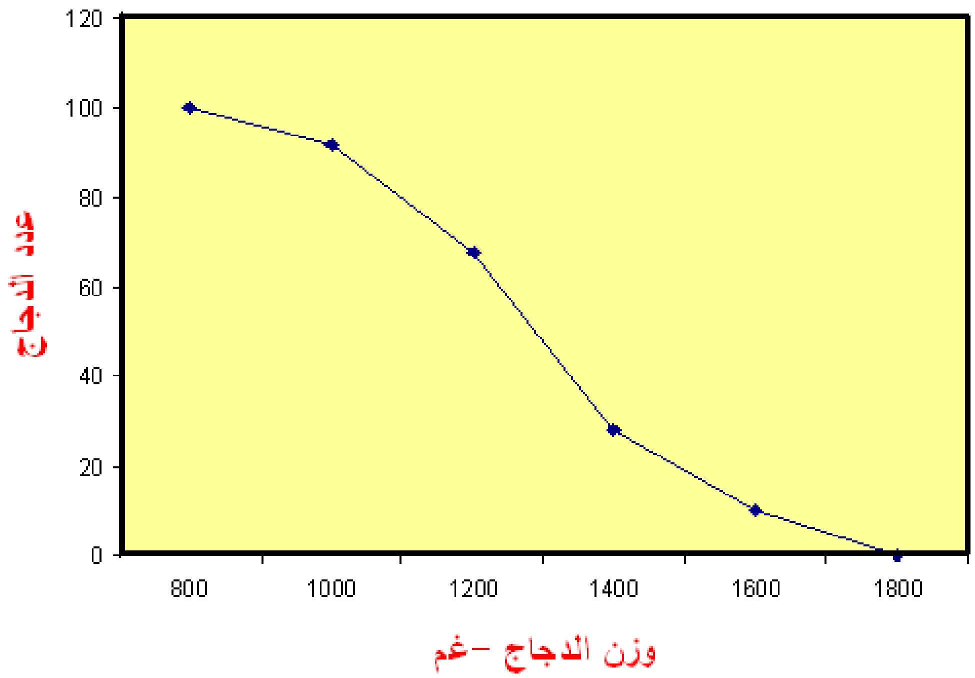
الحل: لتكوين الجدول التكراري المتجمع التنازلي، يتم حساب مجموع التكرارات (عدد القيم) التي تزيد عن كل حد من حدود الفئات.وكما يلي

جدول ٩-٣-٥. جدول التكرار التجميعي التنازلي

فئات الوزن	التكرارالتجميعي التنازلي
اكبر من ٨٠٠	١٠٠
اكبر من ١٠٠٠	٩٢
اكبر من ١٢٠٠	٦٨

اكبر من ١٤٠٠	٢٨
اكبر من ١٦٠٠	١٠
اكبر من ١٨٠٠	٠
المجموع	١٠٠

ثم رسم المضلع التكراري التجميعي التنازلي وكما يلي:



شكل ٩-٣- ٧ المضلع التكراري التجميعي التنازلي لبيانات الجدول ٩-٣-٥

٩-٣-٢ العرض البياني في حالة متغير نوعي

١) العرض الدائري **Pie Chart**

ويتمثل في دائرة مقسمة إلى عدة أجزاء كل جزء يقابل زاوية مركزية تتناسب مع التكرارت المقابلة لكل خاصية (صفة) من الخصائص المدروسة، ولتحقيق ذلك نضيف عمودا إلى جدول المعطيات يحتوي على الزوايا المركزية المقابلة لكل تكرار.

بما ان زوايا الدائرة (الزاوية القطرية) او الزاوية حول نقطة = ٣٦٠ °

لذا نحسب الزاوية المركزية لكل خاصية او صفة بالطريقة التالية:

$$\text{الزاوية المركزية} = \frac{\text{تكرار الخاصية}}{\text{التكرار الكلي}} \times ٣٦٠^\circ$$

نرسم دائرة ومن نقطة المركز نرسم نصف قطرها ، وباستعمال المنقلة نرسم الزوايا المركزية لكل خاصية بحيث تكون مجموع الزوايا ٣٦٠° . بعد ذلك نعطي كل زاوية لون يميزها عن البقية.

مثال: يبين الجدول التالي عدد النخيل لكل صنف في مزرعة.المطلوب عرض البيانات بطريقة العرض الدائري.

جدول ٩-٣-٦. جدول التوزيع التكراري لاصناف اشجار النخيل

اصناف النخيل	العدد
خستاوي	١٠٠٠
زهدي	٤٠٠

برحي	**٧٠٠**
مكتوم	**١٢٠٠**
خضراوي	**٣٠٠**
المجموع	**٣٦٠٠**

الحل:

١- نحسب الزوايا المركزية لكل صنف من النخيل وكالآتي

خستاوي = (١٠٠٠ / ٣٦٠٠) x ٣٦٠ = ١٠٠ °

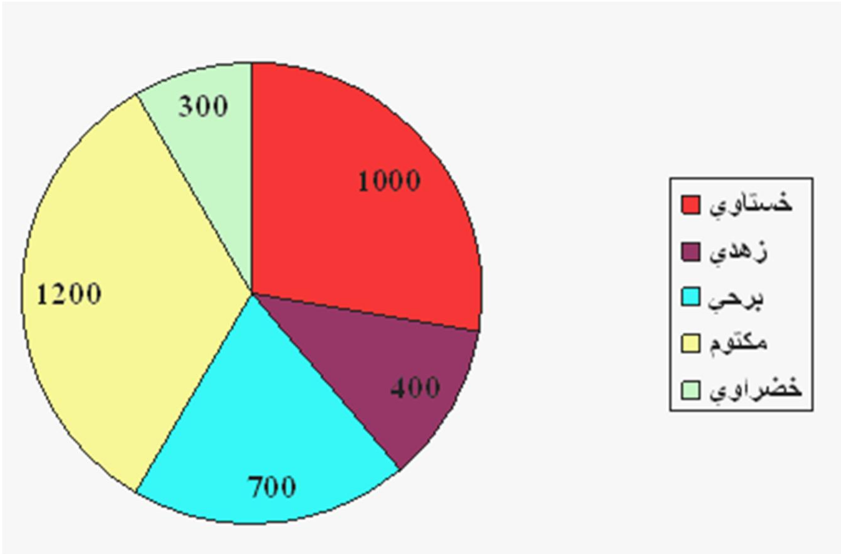
زهدي = (٤٠٠ / ٣٦٠٠) x ٣٦٠ = ٤٠ °

برحي = (٧٠٠ / ٣٦٠٠) x ٣٦٠ = ٧٠ °

مكتوم = (١٢٠٠ / ٣٦٠٠) x ٣٦٠ = ١٢٠ °

خضراوي = (٣٠٠ / ٣٦٠٠) x ٣٦٠ = ٣٠ °

٢- نرسم دائرة ثم باستعمال المنقلة نرسم الزوايا المركزية اعلاه فنحصل على الشكل التالي

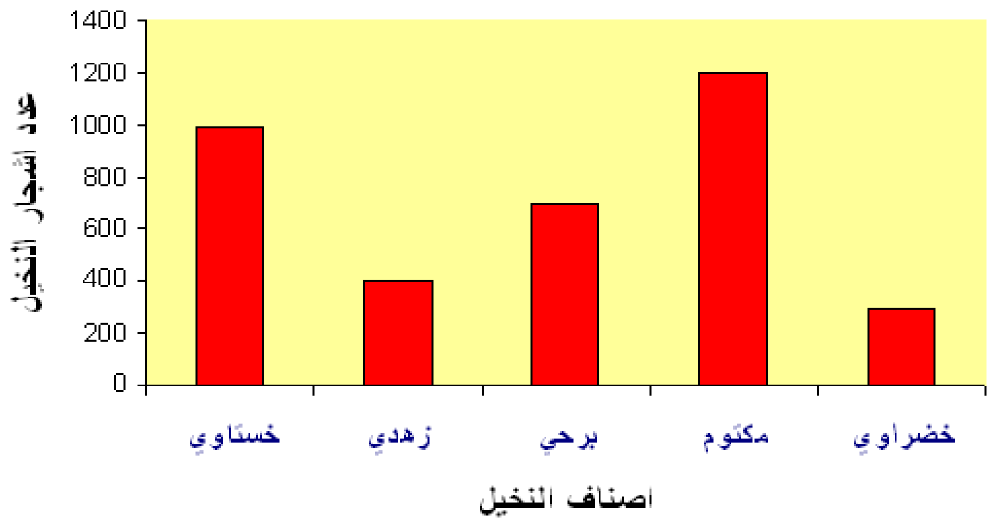


شكل ٩-٣-٨ . التمثيل الدائري لبيانات الجدول ٩-٣-٦

٢) الأعمدة Bar chart

وهو عبارة عن مجموعة من الأعمدة المتجاورة ذات القواعد المتساوية إلا أن ارتفاعها تتناسب مع تكرار كل خاصية، كما أن هذه الأعمدة تكون متباعدة بمسافات متساوية.

أعرض بيانات المثال السابق باستخدام الاعمدة البسيطة



شكل ٩-٣-٩. تمثيل بيانات اصناف اشجار النخيل بطريقة الاعمدة البسيطة

تمارين الفصل الرابع

١- البيانات التالية تبين عدد الغيابات التي سجلها طلبة اعدادية الزراعة في محافظة بابل الفصل الأول من السنة.

٣	٥	٢	٦	٢	٣	٧	٥	٣	٤	٦	١	٤	٥	٩
٠	٥	٢	١	٥	٠	٠	٤	٥	٥	٩	٤	٣	٣	٢
٤	١	٢	٣	٢	٠	٣	٥	١	١	١	٢	٢	٢	١

المطلوب::

- حدد المجتمع الاحصائي والمتغير الاحصائي ونوعه؟
- لخص هذه البيانات في جدول إحصائي؟.

٢ - البيانات التالية تمثل رواتب ٥٠ شخص في إحدى المؤسسات شهريا(بالاف الدنانير).

١٢٠	١٢٠	١٢٠	١٧٠	١٨٠	١٨٠	١٨٠	٢٣٠	٢٥٠	٢٠٠	٣٦٠	٣٧٠	٣٧٥

٣٧٥	٣٨٠	٣٨٠	٣٩٠	٩٠	١٠٠	١١٠	٤٦٠	٥٢٠	٥٢٠	٥٢٠	٢٨٠	٣٥٠
٣٥٠	٣٦٠	٣٦٠	٣٦٠	٦٤٠	٦٥٠	٣٩٠	٤٠٠	٤٠٠	٤٠٠	٤٢٠	٤٢٠	٤٤٠
٤٦٠	٥٤٠	٥٤٠	٦٠٠	٦٠٠	٦٤٠	٦٢٠	٦٢٠	٦٢٠	٦٢٠	٦٣٠		

المطلوب: لخص البيانات اعلاه في جدول توزيع تكراري من ٧ فئات متساوية الطول؟

٤- فيما يلي درجات ٦٠ طالب في الامتحان النهائي لمادة الرياضيات

٧١	٥٥	٨٨	٧٢	٩٢	٤٥	٦٧	٨٣	٥٦	٣٦	٨٥	٤٠
٥٥	٦١	٧٣	٦٤	٨٧	٩٣	٢٥	٤٧	٨٦	٩٤	٢٩	٣٥
٥٩	٤٨	٦٥	٦٧	٨٩	٥١	٢٩	٧٢	٨٧	٥٧	٦٠	٦٧
٩٠	٢٧	٨٨	٦٢	٧٤	٧٦	٤٣	٥٩	٧٩	٨٧	٩٢	٥٦
٣٤	٨١	٧٥	٧٠	٦٦	٨١	٥٣	٦٩	٧١	٤٠	٥٨	٧٦

المطلوب:

١. احسب المدى للبيانات اعلاه.
٢. تفريغ البيانات اعلاه في جدول توزيع تكراري لفئات متساوية الطول
٣. لخص البيانات في جدول توزيع تكراري
٤. ارسم المدرج التكراري
٥. ارسم المضلع التكراري
٦. ارسم المنحنى التكراري

٥- استخدم بيانات السؤال الرابع

١ - كون جدول توزيع تكراري متجمع تصاعدي لبيانات السؤال الرابع.

٢ - كون جدول توزيع تكراري متجمع تنازلي لبيانات السؤال الرابع.

٣- ارسم المضلع التكراري التصاعدي

٤- ارسم المضلع التكراري التنازلي

٦- البيانات التالية تمثل توزيع منتسبي احد المصانع حسب التخصص

التخصص	عدد العاملين
خبير	٥
رئيس مهندسين	١٠
مهندس	٣٠
عامل ماهر	٤٠
عامل غير ماهر	٢٠
اداري	١٥

المطلوب : ١- ما هو نوع المتغير

٢- مثل البيانات بطريقة التمثيل الدائري

٧- مثل البيانات في السؤال السادس بطريقة الاعمدة البسيطة

الفصل الخامس

مقاييس النزعة المركزية
(مقاييس التمركز او التوسط)
Measure of Central Tendency

الوسط الحسابي
الوسط الحسابي للبيانات غير المبوبة
الوسط الحسابي للبيانات المبوبة
الوسيط
المنوال
الوسط الهندسي

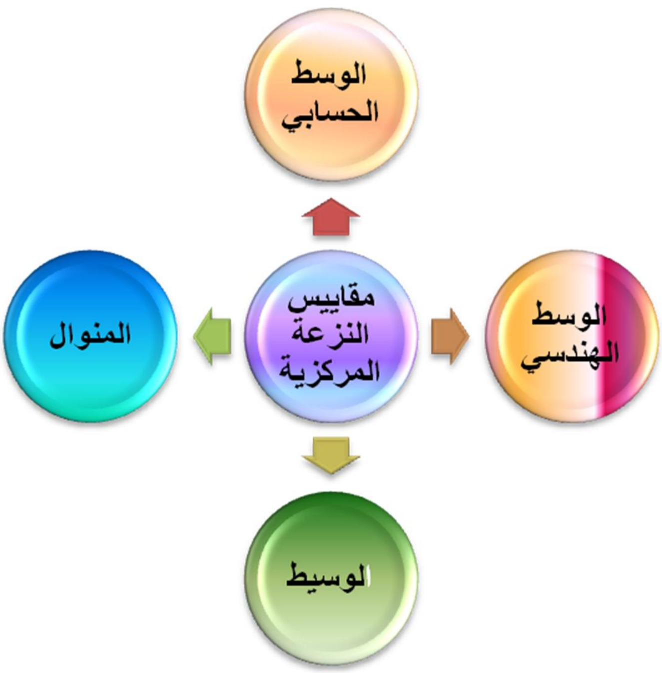
المقاييس الاحصائية

ان عرض البيانات بطريقتي الجداول اوالاشكا ل البيانية هو اجراء مفيد وفعال في تسهيل ادراك الصورة التي تعكسها البيانات. الا ان هناك طرق اخرى لعرض وتلخيص البيانات والتي تأخذ بنظر الاعتبار خصوصية تلك البيانات بما في ذلك تركيزها حول قيمة معينة ومدى تشتتها عن بعضها البعض. وفي كثير من النواحي التطبيقية يكون الباحث في حاجة إلى حساب بعض المؤشرات التي يمكن الاعتماد عليها في وصف الظاهرة من حيث القيمة التي تتوسط القيم أو تتركز حولها القيم ، ومن حيث التعرف على مدى تجانس القيم التي يأخذها المتغير، وأيضا ما إذا كان هناك قيم شاذة أم لا . والاعتماد على العرض البياني وحدة لا يكفى ، ولذا يتناول هذا الفصل، والذي يليه عرض بعض المقاييس الإحصائية التي يمكن من خلالها التعرف على خصائص الظاهرة محل البحث، وكذلك إمكانية مقارنة ظاهرتين أو أكثر ، ومن أهم هذه المقاييس:

- مقاييس النزعة المركزية (مقاييس التمركز او التوسط)
- مقاييس التشتت

مقاييس النزعة المركزية Measure of Central Tendency

عند التمعن في الظواهر التي حولنا والقيم التي تأخذها نلاحظ ان يحدث في اغلب التوزيعات التكرارية ان تتراكم اوتتمركز القيم عند نقطة متوســـطة ، وهو ما يعرف بظاهرة النزعة المركزية ، اي نزعة القيم المختلفة الى التمركز عند القيمة النموذجية او الممثلة لمجموعة القيم في التوزيع،ونظرا لأن مثل هذه القيمة تميل الى الوقوع في المركز داخل مجموعة البيانات لذا تسمى هذه القيمة بالقيمة المتوسطة اوالنزعة المركزية ، آخذين في الاعتبار انه يوجد عدة اسس لتحديد القيمة المتوسطة. وبالتالي فهناك عدة صور لهذه القيمة اهمها واكثرها شيوعا هي الوسط الحسابي ، والوسيط ، والمنوال ، والوسط الهندسي ولكل منهذه المقاييس مزاياه وعيوبه ، واختيار اي من هذه الطرق يعتمد على البيانات وعلى الهدف من دراستها.



٤-١ الوسط الحسابي Arithmetic Mean

وهو اكثر مقاييس النزعة المركزية شيوعا واستعمالا ، ويطلق عليه احيانا بالمعدل الحسابي او الوسط الحسابي ويرمز لهذا المقياس برمزين مختلفين هما:

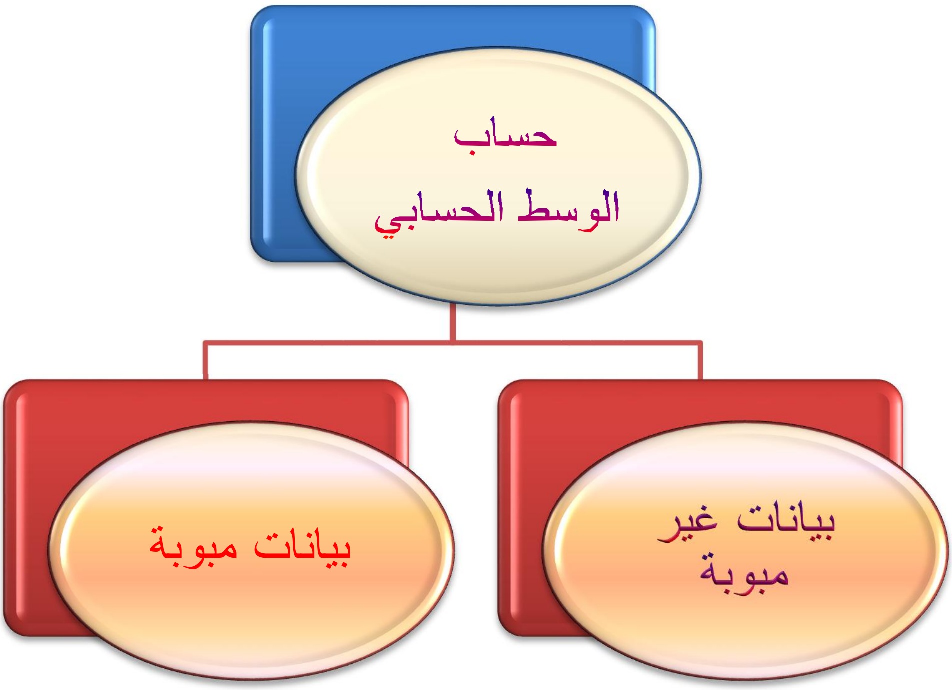
أ- الرمز (μ) ليمثل المتوسط الحسابي للمجتمع

ب- الرمز ($\bar{X}$) ليمثل المتوسط الحسابي للعينة

وتجدر الاشارة الى ان μ هي قيمة ثابتة لا تتغير ، ولهذا فأنه تعتبر من بين معالم او ثوابت (Parameter) المجتمع . اما $\bar{X}$ فانها تتغير من عينة لاخرى اعتمادا على العناصر التي تشملها كل عينة ولهذا فانها تعتبر من الاحصاءات (Statistics).

يعرف الوسط الحسابي لمجموعة من القيم **مجموع قيم المفردات مقسوما على عددها**

طرق حساب الوسط الحسابي



(أ)- الوسط الحسابى للبيانات غير المبوبة

اذا كان لدينا n من القيم او المشاهدات $x_1 , x_2 , ..., x_n$ فان الوسط الحسابي لها هو

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

$$\overline{X} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

مثال: اوجد الوسط الحسابي للبيانات التالية

٤ ، ١٢ ، ١ ، ٢ ، ٥ ، ٤ ، ١٠ ، ١٤ ، ١ ، ٢٣

عدد البيانات = n = ١٠

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{4+12+1+2+5+4+10+14+1+23}{10}$$

$$= \frac{76}{10} = 7.6$$

مثال ٤-١-٢: البيانات التالية تمثل كمية الامطار الساقطة سنويا (بالمليمترات) على محافظة نينوى خلال فترة سبع سنوات ، فما هو متوسط سقوط الامطار خلال تلك الفترة؟

البيانات: ٤٢٠ ٢٤٠ ٥١٠ ١٨٠ ٢٧٠ ٢٩٠ ٥٤٠

الحل

$$\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + x_3 ...... + x_n}{n}$$

$$= \frac{420 + 240 + 510 + 180 + 270 + 290 + 540}{7}$$

$$= \frac{2450}{7} = 350$$

مثال ٤-١-٣. البيانات ادناه تمثل الدرجات النهائية لاحد الطلبة في ستة مواد دراسية، احسب متوسط درجاته؟

٦٨ ٧٠ ٩٠ ٧٢ ٦٥ ٨٥

الحل:

$$\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + x_3 \ldots\ldots + x_n}{n}$$

$$= \frac{68 + 70 + 90 + 72 + 65 + 85}{6} = 75$$

(ب) الوسط الحسابى للبيانات المبوبة

من المعلوم ان القيم الاصـــليـة لا يمكن معرفتهـا من جـدول التوزيع التكراري ، حيـث ان هـذه القيم موضـوعة بشـكل فئات ، ولذا يتم التعبير عن كل قيمة من القيم التي تقع داخل حدود الفئة بمركز هذه الفئة، ومنثم يؤخذ في الاعتبار ان مركزالفئة هو القيمة التقديرية لكل مفردة تقع في هذه الفئة.

فاذا كانت k هي عدد الفئات ، وكانت $x_1 , x_2 , x_3, \ldots\ldots, x_k$ هي مراكزالفئات ، و $f_n \ldots . f_3 \quad f_2$, , f_1 هي التكرارات ، فان الوسط الحسابي يحسب بالطريقة التالية:

$$\overline{X} = \frac{\sum_{i=1}^{k} x_i . f_i}{\sum_{i=1}^{k} f_i}$$

مثال ٤-١-٤. احسب المتوسط الحسابي لدرجات ٥٠ طالب من جدول التوزيع التكراري ادناه

الدرجات	التكرار
49.5 – 59.5	٣
59.5 – 69.5	٥
69.5 – 79.5	١٨
79.5 – 89.5	١٦
89.5 – 99.5	٨
المجموع	٥٠

الحل: ١- يجب تعيين مركز الفئات كما يلي

مركزالفئة = (الحد الادنى +الحد الاعلى) / ٢

٢- ضرب تكرار كل فئة في مركز فئتها

الدرجات	مركز الفئة x_i	التكرار f_i	$x_i * f_i$
49.5 – 59.5	54.5	٣	163.5
59.5 – 69.5	64.5	٥	322.5
69.5 – 79.5	74.5	١٨	1341
79.5 – 89.5	84.5	١٦	١٣٥٢
89.5 – 99.5	94.5	٨	٥٦٧
المجموع		٥٠	٣٧٤٦

٣- استخراج مجموع (حاصل ضرب مركز الفئة x التكرار)

٤- تطبيق القانون التالي

$$\overline{X} = \frac{\sum_{i=1}^{k} x_i . f_i}{\sum_{i=1}^{k} f_i}$$

$$= \ (3746 / 50)$$

$$= 74.92$$

مثال ٤-١-٥. جدول التوزيع التكراري التالي يبين اوزان ٤٠ طالبا ، والمطلوب حساب المتوسط الحسابي.

فئات الوزن	٣٤-٣٢	34-46	36-38	38-40	40-42	42-44
عدد التلاميذ	٤	٧	١٣	١٠	٥	١

الحل: ١- تعيين مركزالفئات (x)

٢- ضرب تكرار كل فئة x مركزالفئة

٣- استخراج مجموع (حاصل ضرب مركز الفئة x التكرار)

٤- تطبيق القانون ٤-١-٢

فئات الوزن	٣٤-٣٢	34-36	36-38	38-40	40-42	42-44
عدد التلاميذ f_i	٤	٧	١٣	١٠	٥	١
مركزالفئة x	٣٣	٣٥	٣٧	٣٩	٤١	٤٣
x_i * f_i	١٣٢	٢٢٥	٤٨١	٣٩٠	٢٠٥	٤٣

$$\overline{x} = \frac{\sum_{i=1}^{6} x_i f_i}{\sum_{i=1}^{6} f_i} = \frac{1496}{40} = 37.4\ k.g$$

خواص الوسط الحسابى

١- مجموع انحرافات القيم عن الوسط الحسابي يساوي صفر

١_ مجموع انحرافا ت القيم عن متوسطها الحسابي يساوي صفر

مثال٤-١-٦

البيانات : ٥ ٦ ٢ ٣ ٤

الوسط الحسابي : $= \frac{20}{5}$ 4

الانحرافات عن الوسط الحسابي :

(4-4) + (3-4)+ (2-4) + (6-4) + (5-4)

0 + (-1) + (-2) + 2 + 1

= ٠

٢ - عند اضافة عدد ثابت (k) الى كل قيمة من المشاهدات فان الوسط الحسابي الجديد = الوسط الحسابي للقيم الاصلية + العدد الثابت (k)

مثال ٤-١-٧:

البيانات : ٥ ٦ ٢ ٣ ٤

الوسط الحسابي : $= \frac{20}{5}$ 4

اضافة العدد الثابت : ٣

البيانات الجديدة : ٨ ٩ ٥ ٦ ٧

الوسط الحسابي الجديد : $\frac{35}{5}$ = ٧

او الوسط الحسابي الجديد = الوسط الحسابي للقيم الاصلية + العدد الثابت (k)

= ٤ + ٣ = ٧

٣- اذا ضربت كل قيمة من المشاهدات في قيمة ثابتة فان

الوسط الحسابى الجديد = الوسط الحسابى للقيم الاصلية x العدد الثابت (k)

مثال ٤-١-٨:

البيانات : ٥ ٦ ٢ ٣ ٤

الوسط الحسابي : (5 / 20) = 4

نضرب x العدد الثابت : ٣

البيانات الجديدة : 15 18 6 9 12

الوسط الحسابي الجديد : (5 / 60) = 12

او الوسط الحسابي الجديد = الوسط الحسابي للقيم الاصلية x العدد الثابت (k)

= ٤ x ٣ = 12

٤- الوسط الحسابي لمجموع قيم متغيرين = مجموع الوسطين الحسابيين للمتغيرين

مثال ٤-١-٩.

اذا كانت قيم المتغير x تساوي ٧ ٣ ١٠ ١ ٩

اذا كانت قيم المتغير y تساوي 12 ٨ ١٥ ٦ ١٤

مجموع قيم المتغيرين ١٩ ١١ ٢٥ ٧ ٢٣

المتوسط الحسابي للمتغير x = $\frac{30}{5}$ = ٦

المتوسط الحسابي للمتغير y = $\frac{55}{5}$ = 11

المتوسط الحسابي للمتغيرين x و y = $\frac{85}{5}$ = ١٧

او الوسط الحسابي لمجموع قيم متغيرين = مجموع الوسطين الحسابيين للمتغيرين

الوسط الحسابي لمجموع قيم متغيرين = ٦ + ١١ = ١٧

مزايا وعيوب الوسط الحسابي

يتميز الوسط الحسابي بالمزايا التالية:

١- أنه سهل الحساب

٢- يأخذ في الاعتبار كل القيم

٣- أنه أكثر المقاييس استخداما وفهما

ومن عيوبه.

١- أنه يتأثر بالقيم الشاذة والمتطرفة

٢- يصعب حسابه في حالة البيانات الوصفية● .

٤-٢ الوسيط Median

تبين لنا عند دراستنا للمتوسط الحسابي أن هذا المتوسط يعطي نتيجة صحيحة ومنطقية عندما تكون البيانات التي حسب منها متجانسة ومتقاربة، أما إذا كانت تحتوي على قيم متطرفة في الصغر أو الكبر فإن النتيجة التي يعطيها تكون غير واقعية، في مثل هذه الحالات فقد وجد متوسط آخر سمي بالوسيط الذي هو أكثر واقعية ودلالة وصحة للحصول على فكرة عامة عن حالة البيانات التي بها قيم متطرفة.

الوسيط Median

هو القيمة التي تقع في الوسط ، وذلك بعد ترتيب القيم تصاعديا او تنازليا او بمعنى اخر هي القيمة التي تقسم مفردات العينة بعد ترتيبها تنازليا او تصاعدياالى قسمين متساويين ، بحيث تكون عدد المفردات الاصغر منها في القيمة مساويا لعدد المفردات الاكبر منها في القيمة.

طريقة حساب الوسيط

الحالةالاولى: اذا كان عدد البيانات فرديا تكون هناك قيمة واحدة فقط في الوسط وتكون هي قيمة الوسيط.

الحالة الثانية: اذا كان عدد البيانات زوجيا فتكون هناك قيمتان في الوسط ، وتكون قيمة الوسيط هي متوسط القيمتين

مثال ٤-٢-١ : اوجد قيمة الوسيط للبيانات التالية :

٩	٧	١	٥	١٠	٢	٨

لايجاد قيمة الوسيط : يجب ترتيب البيانات تصاعديا اوتنازليا

٣ مفردات ٣ مفردات

1	2	5	7	8	9	10

بما ان عدد المفردات اوالقيم عدد فردي (٧) فهناك قيمة واحدة في الوسط وهي ٧ وهي التي تمثل قيمة الوسيط.

الطريقة الثانية

هي استخراج مرتبة الوسيط في حالة اذا كان عدد القيم فرديا ، تساوي

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

n = عدد القيم او المفردات = ٧

لذا فان **مرتبة** الوسيط للبيانات اعلاه هي = $\frac{7+1}{2}$ = ٤

وفي **المرتبة** الرابعة بعد الترتيب التصاعدي هو الرقم ٧

اذن قيمة الوسيط تساوي ٧ ، لان هناك ثلاث قيم اقل من ٧ وثلاث قيم اكبر من ٧

مثال ٤-٢-٢ اوجد قيمة الوسيط للبيانات التالية:

١٢	٣	٢٠	١٤	٥	٢	١٨	٣	٩	١٨

لايجاد قيمة الوسيط : : يجب ترتيب البيانات تصاعديا اوتنازليا

٢	٣	٣	٥	٩	١٢	١٤	١٨	١٨	٢٠

ولاستخراج **قيمة** الوسيط في حالة اذا كان عدد القيم اوالمفردات زوجيا ، تساوي متوسط قيمتي المشاهدتين اللتين تقعان في المنتصف وهي ٩ و ١٢ ومتوسطهما ١٠,٥

والطريقة الثانية

هي بايجاد مرتبتي الوسيط في حالة كون عدد البيانات عدد زوجي

$$\frac{n}{2} \quad \text{و} \quad \frac{n}{2}+1$$

n = عدد القيم او المفردات = ١٠

لذا فان مرتبة الوسيط للبيانات اعلاه هي

١- المرتبة الاولى للوسيط = $\frac{n}{2} = \frac{10}{2} =$ ٥

٢- المرتبة الثانية للوسيط = $\frac{n}{2}+1 = \frac{10}{2}+1 =$ ٦

وفي المرتبة الخامسة الرقم ٩ بعد الترتيب التصاعدي ، وفي المرتبة السادسة الرقم ١٢

اذن قيمة الوسيط = $\frac{9+12}{2}$

$= 10.5$

مثال ٤-٢-٣. اوجد قيمة الوسيط للمجموعتين أ و ب ادناه:

(أ)	50 ، 60 ، 80 ، 70 ، 40 ، 100 ، 90
(ب)	40 ، 50 ، 60 ، 70 ، 100 ، 80

٣ مفردات ٣ مفردات

(أ) ترتب البيانات تصاعدياً: 40، 50، 60، 70، 80، 90، 100

القيمة الوسطى

M

الوسيط $M = 70$

القيمتان الوسيطان

مفردتان مفردتان

(ب) ترتب البيانات تصاعدياً: 40 ، 50 ، 70 ، 80 ، 90 ، 100

$$M = \frac{70+80}{2} = 75$$

الوسيط $M = 75$

خواص الوسيط:

يتصف الوسيط بعدة خصائص أهمها:

١ ــ لا يتأثر بالقيم المتطرفة وبالتالي فإنه يعتبر أصلح المقاييس عن وجود مثل هذه القيم.

٢ ــ يمكن إيجاد الوسيط من الرسم.

٣ ــ يمكن استخدامه في حالة البيانات النوعية.

٤-٣ المنوال Mode

يعرف المنوال بانه

القيمة الاكثرشيوعا اوتكرارا في البيانات

يمثل المنوال القيمة الآكثر تكرارا ، وفي القيم المنفصلة عن بعضها يسهل معرفته من التكرار ، (٢ ، ٣ ، ٤ ، ٥ ، ٥ ، ٦ ، ٧ ، ٨) ، وعندما تكون القيم مجدولة فان الفئة الاكثر تكرارا هي فئة المنوال . وعند اسقاط تكرار القيم على الورق البياني ، فان قمة المنحنى ، او العمود الأطول هو المنوال. والمنوال قد يكون مجموعة خالية (اي لايوجد منوال في البيانات لعدم وجود تكرار فيها) او وحيد القيمة (اي هناك منوال واحد للبيانات) كما قد يكون هناك أكثر من منوال في البيانات.ويمكن ايجاد قيمة المنوال للبيانات الوصفية والكمية.

مثال ٤-٣-١.

البيانات التالية تمثل التقديرات التي تحصل عليها ١٠ طلاب في مادة الرياضيات.

المطلوب: إيجاد المنوال

ممتاز ، جيد، جيد جدا، جيد، متوسط،، فوق المتوسط، جيد، ضعيف، جيد جدا، جيد.

الحل: المنوال هو جيد (لانه الاكثر تكرارا)

مثال ٤-٣-٢

اختيرت عينات عشوائية من طلاب بعض أقسام كلية الزراعة ، وتم رصد درجات هؤلاء الطلاب في مادة الإحصاء ، وكانت النتائج كالتالي:

قسم التربة	٨٠	٧٧	٧٥	٧٧	٧٧	٦٥	٧٠	٥٨	٦٧	٩٠
قسم المحاصيل	٥٣	٨٧	٦٧	٧٥	٥٥	٨١	٤٥	٨٨	٦٠	٧٠

قسم البستنة	٨٠	٨٦	٦٥	٧٦	٨٨	٦٥	٨٠	٦٩	٦٥	٨٠
قسم الثروة الحيوانية	٨٥	٧٢	٧٣	٦٩	٦٩	٧٣	٨٥	٦٩	٧٣	٨٥

المطلوب احسب المنوال للدرجات في كل قسم من الاقسام

الحل: المنوال هوالقيمة الاكثر تكرارا

القسم	القيمة الاكثر تكرارا	قيمة المنوال
قسم التربة	الدرجة ٧٧ تكررت ٣ مرات	٧٧
قسم المحاصيل	لا توجد درجة مكررة	لا يوجد منوال
قسم البستنة	الدرجة ٦٥ تكررت ٣ مرات الدرجة ٨٠ تكررت ٣ مرات	المنوال الاول ٦٥ المنوال الثاني ٨٠
قسم الثروة الحيوانية	الدرجة ٦٩ تكررت ٣ مرات الدرجة ٧٣ تكررت ٣ مرات الدرجة ٨٥ تكررت ٣ مرات	المنوال الاول ٦٩ المنوال الثاني ٧٣ المنوال الثالث ٨٥

خواص المنوال

١- لا يتأثر بالقيم او البيانات المتطرفة
٢- يمكن استخدامه في البيانات الوصفية

عيوب المنوال

غير دقيق وممكن وجود اكثر من منوال في البيانات

٤-٤ الوسط الهندسي Geometric Mean

في الحالات التي تكون فيها قيم الظاهرة المدروسة عبارة عن نسب أو معدلات، أي في الحالات التي نرغب فيها بدراسة معدل تغيرات ظاهرة ما فإن المتوسط الحسابي لن يصف هذه الظاهرة الوصف السليم، ولن يعطي أي فكرة صحيحة عن مثل هذه الظاهرة ولهذا دعت الضرورة إلى إيجاد متوسط آخر يصلح لوصف مثل هذه الظواهر سمي هذا المتوسط بالمتوسط الهندسي.

والمتوسط الهندسي واسع الاستخدام في الحياة الاقتصادية لأن التركيز يكون غالبا منصبا على إيجاد متوسط نسب التغير لبعض الظواهر مثل: تطور الدخول، زيادة الأجور، والنمو السكاني ... إلخ.

تعريف الوسط الهندسي Geometric Mean

إذا كانت لدينا القيم $x_1, x_2, x_3, x_4, \ldots\ldots x_n$ فإن متوسطها الهندسي يساوي الجذر النوني لحاصل ضرب هذه القيم في بعضها البعض.

$$G = \sqrt[n]{x\ x\ x \quad\ x}$$

وتجدر الاشارة الى ان المتوسط الهندسي لمجموعة من القيم يكون دائما اقل من المتوسط الحسابي ، كما ان المعادلة اعلاه لا تسمح بحساب المتوسط الهندسي اذا كانت احدى القيم تساوي صفر وكذلك بالنسبة لوجود عدد فردي من القيم السالبة.

في حالة كون البيانات كبيرة فإنه يفضل استخدام طريقة اللوغاريتم ويكون ذلك كالتالي:

$$G = \sqrt[n]{X1 . X2 . X3 .... Xn}$$

$$\log G = \frac{1}{N}[\log x_1 + \log x_2 + \log x_3 + ... + x_n]$$

$$\log G = \frac{1}{N}\sum_{i=1}^{n} \log x_i$$

وهذا يعني ان لوغاريتم المتوسط الهندسي ما هو الا المتوسط الحسابي للوغاريتمات القيم.

مثال ٤-٤-١

ما هو معدل سرعة احدى سيارات السباق اذا علمنا ان قياس السرعة قد سجل الارقام التالية على اساس كم / ساعة:

١٦٠ ، ١٧٠ ، ١٨٠ ، ١٩٠

الحل:

$$G = \sqrt[4]{160 \times 170 \times 180 \times 190}$$

وباستخدام اللوغاريتمات

$$\log G = (1/4) \times [\log 160 + \log 170 + \log 180 + \log 190]$$

$$\log G = (1/4) \times [2.2041 + 2.2305 + 2.2553 + 2.2788]$$

$$\log G = 8.9687 / 4$$

$$\log G = 2.2422$$

وباستعمال الجداول المقابلة للوغاريتمات

$$G = 147.7$$

اي المتوسط الهندسي للسرعة هو ١٤٧,٧ كم / ساعة

مثال ٤-٤-٢

اوجد الوسط الهندسي للقيم التالية

٣ ، ٥ ، ٨ ، ٣ ، ٧ ، ٢

الحل:

$$G = \sqrt[n]{X1 \cdot X2 \cdot X3 \ldots\ldots Xn}$$

$$G = \sqrt[6]{3 \times 5 \times 8 \times 3 \times 7 \times 2}$$

وباستخدام اللوغاريتمات

$$\log G = (1/6) [\log 3 + \log 5 + \log 8 + \log 3 + \log 7 + \log 2]$$

$$\log G = (1/6) [0.4771 + 0.6990 + \ldots\ldots\ldots + 0.3010]$$

$$\log G = 3.7024 / 6$$

$$\log G = 0.6171$$

وباستعمال الجداول المقابلة للوغاريتمات

$$G = 4.14$$

اي ان الوسط الهندسي للقيم اعلاه تساوي ٤,١٤

الربيعيات والعشيرات والمئينيات Quartile , Decile, and Percentile

اذا رتبت عينة من البيانات حسب قيمتها تصاعدياً او تنازلياً فان القراءة التي تكون في المنتصف والتي تقسم العينة الى مجموعتين متساويتين في العدد هي الوسيط كما سبق تعريفه. وبتعميم الفكرة وتقسيم البيانات بعد ترتيبها الى اربعة اجزاء متساوية ويرمز لها Q_1 ، Q_2 ، Q_3 فان Q_1 يسمى الربيع الاول و Q_2 يسمى الربيع الثاني (الوسيط) و Q_3 يسمى الربيع الثالث . وكذلك يمكن ايجاد القيم التي تقسم البيانات السابقة بعد ترتيبها الى عشرة اقسام ونرمز لها D_1 ، D_2 ، D_3 . . . . D_9 حيث D_1 يسمى العشير الاول وهو يمثل القيمة التي يسبقها عشر القراءات ، D_2 يسمى العشير الثاني وهو يمثل القيمة التي يسبقها 0.2 من القراءات وهكذا . كما يمكن ايجاد القيم التي تقسم

البيانات السابقة بعد ترتيبها الى مائة قسم ونرمز لها P_1 ، P_2 ، .... ، P_{99} حيث P_1 يسمى المئين الاول وهو يمثل القيمة التي يسبقها 0.01 من القراءات ، P_2 يسمى المئين الثاني وهو يمثل القيمة التي يسبقها 0.02 من القراءات وهكذا لباقي المئينات . ويعطي قانون حساب الربيعيات والعشريات والمئينات في حالة البيانات المبوبة مثل قانون الوسيط مع استبدال $\frac{n}{2}$ ب $\frac{n}{4}$ للربيع الأول ، $\frac{2n}{4}$ للربيع الثاني وهكذا ، كذلك استبدال ب $\frac{n}{2}$ $\frac{n}{100}$ للمئين الاول و $\frac{2n}{100}$ للمئين الثاني وهكذا.

- **الربيعيات (الاربعيات) Quartiles**

الأرباعيات او الربيعيات هي النقط التي تقسم التوزيع التكراري إلى أربعة أقسام

فالإرباعي الأول هو النقطة التي تسبقها ربع الدرجات وتليها ثلاثة أرباع الدرجات ، وبذلك تصبح رتبة الإرباعي الأول [n/4] ، حيث [n] عدد الحالات. والإرباعي الثاني هو النقطة التي تسبقها [2n/4 = n/2] ويساوي نصف الدرجات) قبلها ويليها نصف الدرجات . **أي أن الربيع الثاني هو الوسيط.**

الإرباعي الثالث هو النقطة التي تسبقها ثلاثة أرباع الدرجات وتليها ربع الدرجات ، وبذلك تصبح رتبة الإرباعي الثالث [3n/4] . هذا وقد أصطلح إحصائيا على قياس التشتت بتعين مدى الانحراف الإرباعي أي نصف (الإرباعي الثالث – الإرباعي الأول) . لا تختلف أهم الخواص الإحصائية للأرباعيات عن الخواص الإحصائية للوسيط .

تصلح الأرباعيات إلى حد ما لقياس التشتت وخاصة نصف مدى الانحراف الإرباعي . هذا وللأرباعيات أهمية قصوى في معرفة نقط التوزيع التكراري التي تحدد المستويات العليا والوسطى والدنيا للدرجات : فالإرباعي الأول مثلا يحدد النسبة المئوية المساوية لـــ [5%] والإرباعي الثاني يحدد النسبة المئوية المساوية [50%] والإرباعي الثالث يحدد النسبة المئوية المساوية لـــ [75%] أي أن الأرباعيات بهذا المعنى تحدد المستويات المختلفة للضعيف والمتوسط والممتاز ولهذا تعتبر الأرباعيات من أفضل الطرق لتقنين الاختبارات والمقاييس المختلفة .

- **المئينيات والإعشاريات Percentile and Decile**

المئينيات هي النقط التي تقسم التوزيع التكراري إلى أجزاء مئوية ، والإعشـاريات هي النقط التي تقســم التوزيع التكراري إلى أجزاء عشــرية ، كما قســمته الأرباعيات إلى أربعة أقسام : كل قسم يحدد ربع التوزيع التكراري .

لا تكاد تختلف الخواص الإحصـــائية للمئينيات والإعشـــاريات عن خواص الأرباعيات إلا في أمور يســيرة تقوم في جوهرها على كثرة عدد المئينيات والإعشـــــاريات في عدد المئينيات والإعشـــــاريات عن الأرباعيات ، ولهذه الكثرة أثرها في تغيير الصورة العامة النهائية للتقسيم المئيني أو الإعشاري .

D1 = 1rt Decile = 10th Percentile
D2 = 2nd Decile = 20th Percentile
Q1 = 1st Quartile = 95th Percentile
D3 = 3rd Decile = 30th Percentile
D4 = 4th Decile = 40th Percentile

Q2 = D5 = 2nd Quartile
= 5th Decile = Median = 50th Percentile

D6 = 6th Decile = 60th Percentile
D7 = 7th Decile = 70th Percentile

Q3 = 3rd Quartile = 75th Percentile

D8 = 8th Decile = 80th Percentile
D9 = 9th Decile = 90th Percentile

تمارين الفصل الرابع

١- اختر جواباً واحداً فقط من الاجوبة المقابلة لكل سؤال

اذا كانت لديك البيانات التالية ***٤ ٥ ٧ ٣ ٨ ٥***				
أ- الوسط الحسابي	**٦**	**٥,٣٣**	**٤**	**٦,٥**
ب- الوسيط	**٤,٥**	**٦**	**٥,٥**	**٥**
ج- المنوال	**٧**	**٨**	**٥**	**٣**

٢- استخرج الوسط الحسابي لاطوال النباتات من جدول التوزيع التكراري التالي:

الفئات	***التكرار***
٣١-٤٠	**١**

٤١-٥٠	٢
٥١-٦٠	٥
٦١-٧٠	١٥
٧١-٨٠	٢٥
٨١-٩٠	٢٠
٩١-١٠٠	١٢

٣- اختر جواباً واحداً فقط من الاجوبة التالية

مقياس النزعة المركزية الذي يتأثر بالقيم المتطرفة		
الوسيط	**الوسط الحسابي**	**المنوال**

٤- فيما يلي اعمار مجموعة من الطلبة في المرحلة المتوسطة

١٣،١٥ ،١٦،١٦،١٤،١٦،١٣،١٨،١٩،١٦،١٤،١٥،١٦،١٧

احسب أ- الوسط الحسابي

ب- الوسيط

ج- المنوال

٥- البيانات التالية تمثل اوزان ٩ عجول بالكيلوغرام

١٢٠ ١٦٠ ١٣٠ ١٢٠ ١٢٥ ١٨٠ ٢٢٠ ١٦٠ ٢٥٠

المطلوب احسب أ- الوسط الحسابي

ب- الوسيط

ج- المنوال

د- اذا اضفنا ٥ كيلوغرام لكل وزن فكم يكون المتوسط الحسابي الجديد؟

هـ- اذا ضربنا كل وزن في ٢ فماذا يكون الوسط الحسابي؟

٦- فيما يلي توزيع درجات ٦٠ طالبا في مادة اللغة الانكليزية

الفئات	٤٠-٤٩	٥٠-٥٩	٦٠-٦٩	٧٠-٧٩	٨٠-٨٩	٩٠-٩٩
عدد الطلبة	٦	١٠	١٧	١٧	٦	٤

احسب المتوسط الحسابي

٦-   احسب الوسط الهنسي للبيانات ادناه

الفصل السادس

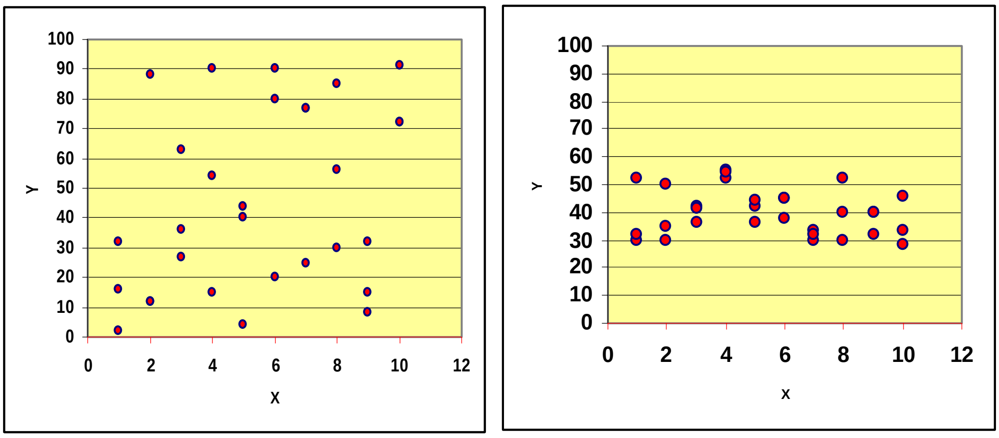
مقاييس التشتت او الاختلاف

Measure of Dispersion

التشتت المطلق
المدى
الانحراف المتوسط
التباين
الانحراف المعياري
الخطأ المعياري
لتشتت النسبي
معامل الاختلاف

المقدمة

تشـتت بيانات ظاهرة ما يقصـد به درجة أو مقدار التفاوت أو الاختلاف بين مفردات هذه الظاهرة، وتعتبر بيانات الظاهرة متجانســة عندما تكون قيمتها قريبة من بعضـــها البعض ونقول في هذه الحالة أنـها غير مشـتتة. أما إذا كانت بيانات الظاهرة متباعدة وغير متجانسـة فنقول أن مفردات الظاهرة مشـتتة وغير مركزة. ومقاييس التشـتت هي مقاييس لمدى تشـتت قيم المشـاهدات عن وسـطها. وعلى ذلك يمكننا ان نتخذ مقدار التشت كدليل على تجمع القيم وقربها من بعضها او على تفرقها وتباعدها عن بعضها ، وهكذا يكون لدينا مقياس لمقدار تجانس المجموعات الاحصـــائية او عدم تجانســـها.ويمكن ملاحظة الشـــكل ٥-١ والاستدلال عن الفرق بين المجموعات الاحصائية في مدى تجانسها.



(ب)   (أ)

الشكل ٥-١. درجة تشتت البيانات الاحصائية

تشتت (ب) > تشتت (أ)

وكما تعرفنا في الفصل السابق على مقاييس النزعة المركزية والتي اعطتنا فكرة اولية عن التوزيع التوزيع التكراري فمن الواضـح ان وصـف التوزيع التكراري باحد تلك المقاييس يعطينا فكرة ناقصـة عن حقيقة المجموعة التي يمثلها التوزيع ، كما ان المقارنة بين المجموعات بناءا على متوسطاتها فقط تكون ناقصة ، كذلك ان لم تكن مضـللة فعلا . فقد يحدث ان يتسـاوى متوسـطا مجموعتين ومع ذلك تكون مفرداتها مختلفة كل الاختلاف ، فربما تكون مفردات المجموعة الاولى قريبة في القيمة من متوسـطها اي مركزه حوله بينما تكون مفرداتة المجموعة الثانية بعيدة في القيمة وتختلف كثيرا عن متوسـطها فيكون بعضـها اكبر منه بكثير والاخر اقل منه بكثير. وكما يتضح من مقارنة المجموعتين التاليتين:

امجموعة الاولى	١٢	١٠	٨	١٣	٩	١١	٧
المجموعة الثانية	١٥	٤	٣١	٢	٧	٨	٣

فالوسط الحسابي لكلا المجموعتين يساوي ١٠ ولكن المجموعة الاولى تبدو اكثر تجانسا . من أجل ذلك لجأ الإحصائيون إلى استخدام مقاييس أخرى لقياس مدى تجانس البيانات، أو مدى انتشار البيانات حول مقياس النزعة المركزية، ويمكن استخدامها في المقارنة بين مجموعتين أو أكثر من البيانات، وسميت بمقاييس التشتت او الاختلاف. وهناك عدة مقاييس لتشتت اهمها :

اولا – التشتت المطلق

اي التي تكون وحداتها نفس وحدات القيم الاصلية واهمها:

١- المدى Range

٢ – متوسط الانحراف Mean Deviation

٣- الانحراف المعياري Standard Deviation

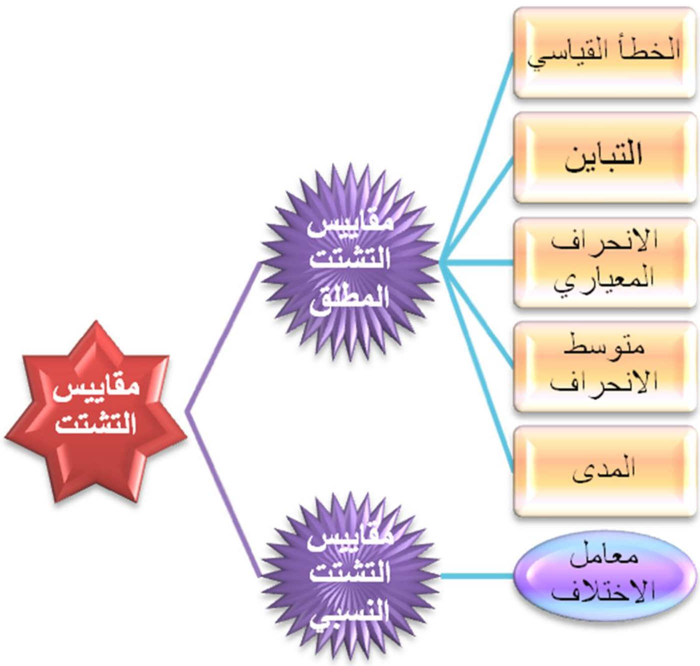
٤- التباين Variance

٥- الخطأ المعياري Standard Error

الثاني- التشتت النسبي

ان مقياس التشتت النسبي له اهميته عند مقارنة تشتت مجموعتين او اكثر تختلف في وحدات القياس لقيمها. لان مقياس التشتت النسبي تكون خالية من وحدات القياس ، واهم مقياس للتشتت النسبي هو:

- معامل الاختلاف Coefficient of Variation



٥-١ مقاييس التشتت المطلق

٥-١-١ المدى Range

تعريف المدى

المدى من القيم هو الفرق بين اعلى قيمة واقل قيمة في تلك المجموعة

ويرمز له R

يتميز هذا المقياس بسهولة حسابه واعطائه فكرة سريعة ومبسطة عن درجة تشتت قيم المجموعة. الا ان نقطة ضعفه انه يهمل جميع قيم المجموعة فيما عدا القيمتين العليا والدنيا وكثير التأثر بالقيم المتطرفة. ونتيجة لنقطة الضعف هذه فانه يعجز عن تمييز درجات تشتت المجموعات بشكل حازم بدليل ان قيمة المدى لمجموعة القيم

٢١	٢٣	٢٠	٢٣	٢٥	٨

تساوي ١٧ ومساوية لقيمة المدى لمجموعة القيم

١٨	٢٠	١٦	٣٠	٢٣	١٣

رغم الاختلاف الواضح في درجتي تشتت المجموعتين . لهذا السبب فان هذا المقياس محدود الاستخدام مقارنة بالمقاييس التي تأخذ بنظر الاعتبار جميع القيم وتقيس تشتتها عن قيمة معينة كاساس لقياس التشتت والتي عادة ما تكون المتوسط الحسابي.

مثال ٥-١-١-١

تم زراعة 9 وحدات تجريبية بمحصول القمح ، وتم تسميدها بنوع معين من الأسمدة الفسفورية، وفيما يلي بيانات كمية الإنتاج من القمح بالطن /هكتار

5.03	4.63	5.08	5.18	5.29	5.18	5.4	6.21	4.8

والمطلوب حساب المدى.

الحل : $R = X_{max} - X_{min}$

$$R = 6.21 - 4.63 = 1.58$$

اي ان المدى للحاصل يساوي 1.58 طن / هكتار

مثال ٥-١-١-٢

الجدول التالي يمثل مراقبة التقلبات السعرية لقيم اسهم شركتين (A و B) بالدينار، اوجد قيمة المدى لسعري السهمين في الشركتين

الشركة A	٣٨	٣٢	٢٤	٣٥	٣٠	٤٠
الشركة B	٥٠	٤٨	٤٩	٤٥	٤٧	٣٤

الحل:

المدى للشركة A	$R = 40 - 24 = 16$
المدى للشركة B	$R = 50 - 34 = 16$

وهذا يعني ان المدى للتغير في اسعار اسهم الشركتين متساوي

٥-١-٢ الانحراف المتوسط Mean Deviation

تعريف الانحراف المتوسط

اذا كانت لدينا n من المشاهدات $x_1, x_2, \ldots, x_n$ فان الانحراف المتوسط لها هو متوسط الانحرافات المطلقة لقيم المشاهدات (اي اهمال الاشارة) عن وسطها الحسابي ويرمزله

M.D

لأن مقاييس التشتت هي مقاييس لقوة تجمع البيانات حول بعضها، وحيث أن التجمع يكون حول القيم المتوسطة، فإنه إذا كان مقدار الاختلاف بين القيم ومتوسطها كبيرا دل ذلك على أن التشتت كبير والعكس صحيح. وحيث أن مجموع الاختلافات (الانحرافات) عن المتوسط يساوي صفرا ، فإنه لو حسبنا القيم

المطلقة لمقدار الاختلاف عن المتوسط يكون متوسط هذه الاختلافات مقياسا مناسبا لمقدار التشتت، يسمى هذا المقياس بالانحراف المتوسط.

ويمتاز هذا المقياس بانه يأخذ جميع القيم وسهل الحساب ولكن يعاب عليه بانه يتأثر بالقيم المتطرفة.

مثال ٥-١-٢-١

اوجد الانحراف المتوسط للبيانات التالية

٢	٨	٥	٦	٤

الحل: الخطوة الاولى استخراج الوسط الحسابي

الخطوة الثانية ايجاد انحراف كل قيمة عن الوسط الحسابي (مع اهمال الاشارة)

المتوسط الحسابي $= \frac{25}{5} =$ ٥

X_i	$\lvert X_i - \bar{X} \rvert$
٢	٣
٨	٣
٥	٠
٦	١
٤	١
المجموع	٨

$$M.D = \frac{\sum|x_i - \bar{x}|}{n}$$

الانحراف المتوسط $= \frac{8}{5} = 1.6$

مثال ٥-١-٢-٢

جد قيمة الانحراف المتوسط للبيانات التالية التي تمثل اوزان عشرة من اللهانة (كغم).

2.8	2.3	1.7	2.0	1.2	2.5	2.0	1.8	2.2	1.5

الحل:

المتوسط الحسابي $= \frac{20}{10} =$ ٢

القيم	1.5	2.2	1.8	2.0	2.5	1.2	2.0	1.7	2.3	2.8
المتوسط الحسابي	٢	٢	٢	٢	٢	٢	٢	٢	٢	٢
الانحراف المتوسط	0.5	0.2	0.2	٠	0.5	0.8	٠	0.3	0.3	0.8

$$M.D = \frac{\sum|x_i - \bar{x}|}{n}$$

0.5 +0.2 +0.2 +0 +0.5 +0.8 +0 +0.3 +0.3+0.8

الانحراف المتوسط = ____

١٠

الانحراف المتوسط = $\frac{3.6}{10} = 0.36$

٥-١-٣ التباين Variance

للتغلب على مشكلة الاشارات عند جمع انحرافات القيم عن وسطها الحسابي والتي تؤدي دائما لان يكون مجموع انحرافات اي عينة عن وسطها حسابي يساوي صفر. وبدلا من اخذ القيم المطلقة للانحرافات اي بدون اشارات فاننا نستطيع ان نتغلب على ذلك بطريقة اخرى وهي بتربيع قيم الانحرافات وبذلك تصبح جميعها موجبة. اي نحصل على مجموع مربعات الانحرافات والتي نرمز لها Sum of Square (SS) وعلى ذلك فان

$$SS = \sum_{i=1}^{n} (X_i - \bar{X})^2$$

ولكي نأخذ بنظر الاعتبار حجم العينة حتى يمكن مقارنة العينات المختلفة الاحجام فاننا نقسم مجموع مربعات الانحرافات على n في حالة حساب التباين للمجتمع وعلى n-1 في حالة حساب التباين للعينة ويسمى (n-1) بدرجات الحرية.

ويرمز لتباين المجتمع بالرمز (σ^2) ولتباين العينة بالرمز (S^2)

تعريف التباين هو متوسط مجموع مربعات انحرافات القيم عن وسطها الحسابي

ويحسب تباين المجتمع بالطريقة التالي

$$\sigma^2 = \frac{\sum (\mathbf{x_i} - \mu)^2}{\mathbf{N}}$$

حيث ان μ الوسط الحسابي للمجتمع

N عدد مفردات المجتمع

ويحسب تباين العينة بالطريقة التالية

$$S^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n - 1}$$

$$S^2 = \frac{\Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n}}{n-1}$$

ونظراً لاننا عند حساب التباين قمنا بتربيع الانحرافات فان قيمة التباين تكون مقاسة بمربع الوحدات المستخدمة في قياس المشاهدات. فاذا كانت المشاهدات مقاسة بالسنتمتر فان التباين يكون بالسنتمتر المربع . ولا توجد مشكلة في ذلك ، ولكن المشكلة عندما تكون وحدات المشاهدات كالاوزان بالكغم او عدد الاطفال في الاسر او عدد الموظفين في شركة ما، فالتباين عنده يقاس بالكغم المربع اوالطفل المربع او الموظف المربع وهذه كلها غير ذات معنى.

والحل لتلك المشكلة هي ارجاع الوحدات الى اصلها وذلك باخذ الجذر التربيعي للتباين لنحصل على ما يسمى بالانحراف المعياري (S) والذي سوف يكون مقاساً بالوحدات الاصلية.

مثال ٥-١-٣-١

اخذت عينة مؤلفة من ٦ نباتات بطاطا وحسبت عددالثمار بالنبات الواحد وكانت كما يلي:

٥	٤	٨	٧	١٠	٨

احسب تباين عدد الثمار بالنبات.

الخطوة الاولى : حساب الوسط الحسابي

$$\overline{X} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\overline{X} = \frac{42}{6} = 7$$

الخطوة الثانية :

عدد الثمار	٨	١٠	٧	٨	٤	٥
$(x_i-\overline{X})$	1	3	0	1	-3	-2

$(x_i-\bar{X})^2$	1	9	0	1	9	٤

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$$

$$\mathbf{S^2 = \frac{1+9+0+1+9+4}{6-1}}$$

$$\mathbf{S^2 = \frac{24}{5} = 4.8}$$

الطريقة الثانية

$$S^2 = \frac{\Sigma x_i{}^2 - \frac{(\Sigma x_i)^2}{n}}{n-1}$$

x_i	x_i^2
٨	٦٤
١٠	١٠٠
٧	٤٩
٨	٦٤
٤	١٦

٥	٢٥
$x_i\Sigma =$ ٤٢	$x_i^2 = 3\ 18\Sigma$

$$S^2 = \frac{318 - \frac{42^2}{6}}{5} = \frac{318 - 294}{5} = \frac{24}{5}$$

$$S^2 = 4.8$$

مثال ٥-١-٣-٢

اخذت عينة مؤلفة من ١٠ اشجار من العنب وحسب كمية الحاصل لكل شجرة بالكيلو غرام وكما يلي

٥	١١	٩	١٤	١٠	١٥	١٢	٦	٨	١٠

احسب التباين؟

الطريقة الاولى:

الوسط الحسابي

$$\overline{X} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\bar{X} = \frac{100}{10} = 10$$

X_i	٥	١١	٩	١٤	١٠	١٥	١٢	٦	٨	١٠
$(x_i - \bar{X})$	-5	١	-1	٤	٠	٥	٢	-4	-2	٠
$(x_i - \bar{X})^2$	25	1	1	16	0	25	4	16	4	٠

$$\sum_{i=1}^{n} (X_i - \bar{X})^2 = 92$$

$$S^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1} = \frac{92}{9} = 10.22$$

الطريقة الثانية

X_i	٥	١١	٩	١٤	١٠	١٥	١٢	٦	٨	١٠	$\Sigma x_i = 100$

X_i^2	٢٥	١٢١	٨١	١٩٦	١٠٠	٢٢٥	١٤٤	٣٦	٦٤	١٠٠	$\Sigma x_i^2 = 1092$

$$S^2 = \frac{1092 - \frac{100^2}{10}}{10-1} = \frac{1092 - 1000}{9} = \frac{92}{9} = 10.22$$

٥ - ١ - ٤ الانحراف المعياري Standard Deviation

يعتبر الانحراف المعياري من أهم المقاييس الإحصائية للتشتت، وهو الأكثر استخداما في النظريات والقوانين الإحصائية، لأنه يعطي فكرة سليمة ومنطقية عن ظاهرة التشتت. وتكون وحداته نفس وحدات الظاهرة او المتغير

تعريف الانحراف المعياري،

يعرف الانحراف المعياري بأنه الجذر التربيعي لمتوسط مجموع مربعات انحرافات القيم عن

ففي حالة حساب الانحراف المعياري للمجتمع فيرمز له σ (sigma)

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \mu)^2}{N}}$$

وفي حالة حساب الانحراف المعياري للعينة فيرمز له S

$$S = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}}$$

اي ان الانحراف المعياري يساوي الجذر التربيعي للتباين

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

ويمتاز الانحراف المعياري بسهولة حسابه وشموله على جميع قيم المشاهدات لذا فهو يعتبر من ادق معايير التشتت الاحصائية ، وله نفس وحدات القياس للظاهرة قيد الدراسة. ويعاب عليه تأثره بالقيم المتطرفة للبيانات ولا يمكن حسابه للقيم الوصفية.

مثال ٥-١-٤-١

احذت عينة مؤلفة من خمسة بساتين وحسبت اشجار الزيتون في كل بستان وكانت كما يلي :

٨	٦	١٠	٧	٩

احسب الانحراف المعياري لعدد اشجار الزيتون .

الحل:

الطريقة الاولى: نحسب الوسط الحسابي لقيم المتغير

الوسط الحسابي = مجموع القيم / عددها

الوسط الحسابي = $\frac{40}{5}$

الوسط الحسابي = ٨

X_i	٩	٧	١٠	٦	٨
$(x_i - \bar{x})$	١	-1	٢	-2	٠
$(x_i - \bar{x})^2$	١	١	٤	٤	٠

ثم نحسب التباين

$$S^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}$$

$$\mathbf{S^2 = \frac{10}{4} = 2.5}$$

ثم نأخذ الجذر التربيعي للتباين للحصول على الانحراف المعياري

$$\mathbf{S = \sqrt{2.5} = 1.581}$$

الطريقة الثانية

حساب الانحراف المعياري بالقانون التالي

$$S = \sqrt{\frac{\Sigma x_i^{\ 2} - \frac{(\Sigma x_i)^2}{n}}{n-1}}$$

x_i	٩	٧	١٠	٦	٨	x_i= 40 Σ
x^2	٨١	٤٩	١٠٠	٣٦	٦٤	x_i^2= 330 Σ

$$S = \sqrt{\frac{330 - \frac{40^2}{5}}{5-1}} = \sqrt{\frac{330 - \frac{1600}{5}}{4}} = \sqrt{\frac{330 - 320}{4}} = \sqrt{2.5}$$

$$S = 1.581$$

مثال :٥-١-٤-٢

اخذت عينة مكونة من ٨ أبقار وحسبت كمية انتاج الحليب (كغم) باليوم وكانت النتائج كما يلي:

X_i	١٠	١٢	٨	٩	٥	١٥	١٣	٨

احسب الانحراف المعياري لكمية انتاج الحليب.

الحل: الطريقة الاولى

١- حساب المتوسط الحسابي

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{10+12+8+9+5+15+13+8}{8} = \frac{80}{8} = 10$$

-٢

X_i	١٠	١٢	٨	٩	٥	١٥	١٣	٨
$(X_i - \bar{X})^2$	٠	٤	٤	١	٢٥	٢٥	٩	٤

$$\sum_{i=1}^{n} (X_i - \bar{X})^2 = 72$$

$$S = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}}$$

$$S = \sqrt{\frac{72}{7}} = \sqrt{10.29} = 3.21$$

الطريقة الثانية:

$$S = \sqrt{\frac{\Sigma x_i{}^2 - \frac{(\Sigma x_i)^2}{n}}{n-1}}$$

X_i	١٠	١٢	٨	٩	٥	١٥	١٣	٨
$X_i{}^2$	١٠٠	١٤٤	٦٤	٨١	٢٥	٢٢٥	١٦٩	٦٤

$$\sum_{i=1}^{8} X_i = 80$$

$$\sum_{i=1}^{8} X_i^2 = 872$$

$$S = \sqrt{\frac{872 - \frac{(80)^2}{8}}{7}} = \sqrt{\frac{872 - 800}{7}} = \sqrt{\frac{72}{7}} = \sqrt{10.29} = 3.21$$

٥-١-٥ الخطأ المعياري (الخطأ القياسي) Standard Error

كانت مقاييس التشتت السابقة عبارة عن إحصاءات لقياس تشتت المفردات داخل العينة وكان الانحراف القياسي هو أهم تلك المقاييس إذ أنه يقيس انحراف مفردات العينة حول متوسطها الحسابي.

والخطأ المعياري هو عبارة عن تقدير للانحراف المعياري للمتوسطات الحسابية المحسوبة من عدد العينات العشوائية الكبيرة الحجم (٣٠ فردا فأكثر) المأخوذة من مجتمع واحد متجانس.

استعمالات الخطأ المعياري:

١- تعتبر قيمة الخطأ المعياري كمقياس لدرجة الاعتماد علي متوسط العينة ، بمعني أن المتوسط الذي تكون قيمة الخطأ المعياري له صغيرة يمكن الاعتماد عليه عن المتوسط الذي تكون قيمة الخطأ المعياري له كبيرة.

٢- يفيد استعمال الخطأ المعياري في تحديد حجم العينة.

٣- يعطي الخطأ المعياري فكرة عن متوسط المجتمع ، فلقد وجد أن متوسطات العينات العشوائية الكبيرة المأخوذة من مجتمع واحد متجانس تتوزع توزيعا معتدلا تقريبا ، حتى ولو لم تكن مفردات المجتمع نفسها معتدلة التوزيع . وبذلك تكون النقطة على المنحني المعتدل التي تتركز حولها متوسطات العينات أحسن تقدير لمتوسط المجتمع.

٤- يمكن استعمال الخطأ المعياري لمقارنة متوسطين مختلفين لمعرفة حقيقة الفرق بينهما.

طريقة حساب الخطأ المعياري

يرمز للخطأ القياسي بالرمز $S_{\bar{x}}$ ويحسب بالطريقة التالية

$$S_{\bar{x}} = \frac{S}{\sqrt{n}}$$

او

$$S_{\bar{x}} = \sqrt{\frac{S^2}{n}}$$

مثال ٥-١-٥-١

اوجد الخطأ المعياري للبيانات التالية

٥	٤	٦	٩	٦

الحل

١- ايجاد قيمة الانحراف المعياري

$$S = \sqrt{\frac{\Sigma x_i^{\,2} - \frac{(\Sigma x_i)^2}{n}}{n-1}}$$

$$S = \sqrt{\frac{6^2 + 9^2 + 6^2 + 4^2 + 5^2 - \frac{(6+9+6+4+5)^2}{5}}{5-1}}$$

$$S = \sqrt{\frac{194 - \frac{30^2}{5}}{4}} = \sqrt{\frac{194 - \frac{900}{5}}{4}} = \sqrt{\frac{194-180}{4}}$$

$$S = \sqrt{\frac{14}{4}} = \sqrt{3.5} = 1.87$$

٢- ايجاد الخطأ المعياري

$$S_{\bar{X}} = \frac{S}{\sqrt{n}} = \frac{1.87}{\sqrt{5}} = \frac{1.87}{2.24} = 0.834$$

٥-٢ التشتت النسبي

كانت مقاييس التشـتت التي ذكرت سـابقا كلها مقاييس مطلقة ، تقدر بدلالة وحدات القياس المسـتعملة في قياس المتغير الموضـوع تحت البحث والدراسـة ، سـواء كانت هذه الوحدات مقاسـه بالسنتيمتر أو المتر أو الكيلو غرام وغيره ، وعلى ذلك فإذا أردنا مقارنة عينتين أو مجتمعين فقد يحول دون ذلك اختلاف وحدات القياس المســتعملة في كل منهما. لذا فان مقاييس التشــتت النســبي تكون خالية من وحدات القياس ، واهم مقاييس التشتت النسبي هي :

٥-٢-١ معامل الاختلاف Coefficent of Variation

ان درجة التشـتت بين قيم مفردات مجموعة معينة تختلف عادة عن درجة التشـتت لمجموعة اخرى ، وقد يكون هذا الاختلاف كبيرا ام صـغيرا. وبناءاً على ذلك لا بد من وجود وسـيلة لمقارنة درجات التشـتت بين المجموعات المختلفة. وخاصـــة اذا اختلفت هذه المجاميع بوحدات القياس ، وعليه فأن الامر يتطلب الاسـتعانة بمقياس تشـتت يحوّل قيمة الانحراف المعياري الى نسبة مئوية من المتوسـط الحسـابي وبذلك يأخذ بنظر الاعتبار التفاوت في القياسات الاصلية للبيانات ويتخلص من وحدات القياس ويوصلنا الى نسبة مئوية قابلة للمقارنة. ويطلق على هذا المقياس بمعامل الاختلاف ويرمز له (C.V)

تعريف معامل الاختلاف

اذا كان S و $\bar{X}$ هما الانحراف المعياري والوسط الحسابي لمجموعة من القيم على التوالي فان معامل الاختلاف لها

$$C.V = \frac{S}{\bar{X}} \times 100$$

مثال ٥-٢-١-١

كانت نتائج الامتحان النهائي لدرسي الرياضيات والكيمياء للصف الرابع كالاتي

	الرياضيات	الكيمياء
الوسط الحسابي	٨٠	٧٢

الانحراف المعياري	٨	٦

ففي اي الموضوعين كان التشتت اكبر؟

معامل الاختلاف لدرس الرياضيات $\mathbf{C.V = \frac{8}{80} \times 100 = 10\%}$

معامل الاختلاف لدرس الكيمياء $\mathbf{C.V = \frac{6}{72} \times 100 = 8.33\%}$

اي ان التشتت لدرجات الرياضيات كان اكثر

مثال ٥-٢-١-٢

اجريت مقارنة صفة ارتفاع النبات (سم) وكمية الحاصل (غم) لعينة مكونةمن ١٠٠ نبات من الذرة الصفراء وكانت النتائج كما يلي:

	الارتفاع	الحاصل

الوسط الحسابي	٢٠٠	٨٠٠
الانحراف المعياري	١٦	٣٦

قارن بين تشتت الصفتين.

معامل الاختلاف لارتفاع النبات $C.V = \frac{16}{200} \times 100 = 8\%$

معامل الاختلاف للحاصل $C.V = \frac{36}{800} \times 100 = 4.5\%$

اي ان تشتت صفة ارتفاع النبات اكبر من الحاصل

تمارين الفصل الخامس

١- عرف ما يلي: المدى ، الانحراف المتوسط ، التباين ، الانحراف المعياري

٢- احد المقاييس التالية هو مقياس للتشتت

المنوال	الوسيط	المدى	الوسط الحسابي

٣- احد المقاييس التالية ليس مقياسا للتشتت

الانحراف المتوسط	معامل الاختلاف	المنوال	التباين

٤- لكل من المجموعات التالية احسب:

(أ) المدى (ب) الانحراف المتوسط (ج) التباين (د) الانحراف المعياري

(هـ) الخطأ المعياري   (و) معامل الاختلاف

(a) $X_i = 3 , 6 , 8 , 10 , 4 , 5$

(b) $X_i = 1 , 8 , 9 , 5 , 12 , 8 , 10 , 7 , 3$

(c) $X_i = 5 , -3 , 2 , -4 , 10 , 12 , -8 , 11$

٥- اوجد قيمة المفقودة للبيانات التالية

	$\bar{X}$	S	C.V
١	؟	١٠	20 %
٢	٤٠	٢٥	؟
٣	٢٥	؟	5 %

الفصل السابع

مبادئ نظرية الاحتمالات

٧-١ مصطلحات وتعاريف

٧-٢ التوافيق

٧-٣ التباديل

مقدمة

كلمة" احتمال "هي كلمة ينطق بها الكثير من الناس، فبعض خبراء الأرصاد الجوية يقولون من المحتمل سقوط أمطار اليوم، احتمال ارتفاع في درجات الحرارة، وبعض خبراء البورصة يقولون احتمال ارتفاع قيمة الأسهم المتداولة في سوق المال لشركة معينة، خلال هذا اليوم، واحتمال نجاح طالب، واحتمال إصابة نوع معين من الفاكهة بنوع من البكتريا، وهكذا، يكثر نطق الأفراد بها وربما يجهلون معناها .فماذا تعني كلمة احتمال؟

يقصد بهذه الكلمة فرصة حدوث أو وقوع حادثة معينة، وتستخدم الاحتمالات في كثير من النواحي التطبيقية، مثل المجالات الاقتصادية، والتجارية، والزراعية، والطبية، والسلوكية، وغيرها، خاصة عند اتخاذ القرار في دراسات الجدوى، والتنبؤ بسلوك الظواهر المختلفة، ولكي يمكن فهم موضوع الاحتمال، وأهميته في النواحي التطبيقية، نقوم بعرض بعض المفاهيم الخاصة بالاحتمالات.

مصطلحات وتعاريف

معنى الاحتـمال

" هو مقياس لامكانية وقوع حدث (Event) معين " .

التجربة العشوائية

هي أي عملية تتم يمكن تحديد كل النتائج الممكنة لها، ولكن لا يمكن مسبقا تحديد النتيجة التي ستظهر أو تحدث،

ومثال على ذلك عند رمي حجر النرد (Dice) مرة واحدة وملاحظة الوجه الظاهري ، نعلم مسبقا ان الوجه الظاهري في الرمية سيكون احد الارقام 6 , 5 , 4 , 3 , 2 , 1 اي يمكن تحديد جميع النتائج الممكنة لكن من غير الممكن تحديد النتيجة بعينها ، لذا سميت هذه التجربة بالتجربة العشوائية. وكذلك عند إلقاء قطعة عملة معدنية مرة واحدة، فإن النتائج الممكنة لها نتيجتان هما" :ظهور الصورة "ويرمز لها بالرمز H . أو" ظهور الكتابة" ويرمز لها بالرمز T، أي أن النتائج الممكنة هي { H , T}، وقبل إلقاء القطعة ، لا يمكن تحديد أي من النتيجتين سوف تظهر.

فضاء العينة

Sample Space

فضاء العينة هو مجموعة من النقاط تمثل جميع النتائج الممكنة لتجربة ما حيث ان كل نتيجة تمثل بنقطة او عنصر في فضاء العينة ،ويرمز له S ويرمز الى عدد عناصر الفضاء بالرمز n(S)

ففي مثال حجرالنرد فان:

فضاء العينة S = { ٦ ، ٥ ، ٤ ، ٣ ، ٢ ، ١ }

عددعناصر الفضاء n(S) = ٦

الحدث

هو مجموعة جزئية من فضاء العينة S ، A حدث من فضاء العينة **S ⇔ S A**

الحدث هو نقطة او مجموعة نقاط في فضاء العينة

الاحداث

لتكن A و B و C احداث من فضاء العينة S يقال لهذه الاحداث شاملة اذا حققت الشروط التالية :

١- اتحاد الاحداث = S فضاء العينة

٢- تقاطعهما مثنى مثنى (كل اثنين منهما) = ø

٣- كل مجموعة منهما ليست خالية

مثال

ليكن { ٦ ، ٥ ، ٤ ، ٣ ، ٢ ، ١ } = S نأخذ بعض الاحداث من S

A_1 = { ١ ، ٤ } حدث مركب (Compound Effect) لان عدد عناصره اكبر من ١

A_2 = { ٣ } حدث بسيط (Simple Event) لان عدد عناصره = ١

A_3 = { ٦ } حدث بسيط

A_4 = { ٥ ، ٤ ، ٣ ، ٢ ، ١ } حدث مركب

A_5 عدد يقبل القسمة على ٥ ، ٢ في نفس الوقت $\leftarrow$ $A_5 = \emptyset$

A_5 حدث مستحيل (Impossible Event)

$A_6 = \{ ٥ ، ٢ \}$ حدث مركب

$A_7 = \{ ٦ ، ٥ ، ٣ ، ٢ \}$ حدث مركب

(لأن S=8SureEvent حدث مؤكد) $A_8 = \{$ ٦ ، ٥ ، ٤ ، ٣ ، ٢ ، ١ $\}$

نلاحظ A7 و A1 احداث شاملة من S

العمليات على الحوادث

١- $\mathbf{A \subseteq S}$ معناه حدث من S

٢- $\emptyset$ تسمى بالحدث المستحيل (الحدث الذي لا يمكن وقوعه)

٣- S فضاء العينة = الحدث المؤكد (يقع دائما)

٤- $A^c = S - A$ يسمى الحدث المكمل للحدث A (او عدم وقوع الحدث A)

A^c = (Complement Event)

٥- $\mathbf{A \cup B}$ يعني وقوع الحدث A أو B اي وقوع احد الحدثين على الاقل

٦- $\mathbf{A \cap B}$ يعني وقوع الحدث A أو B اي حدث وقوع احد الحدثين معًا

٧- $\mathbf{A \subseteq B}$ يعني حدث وقوع الحدث A يستلزم وقوع الحدث B

٨- $\mathbf{A \cap B = \emptyset \Leftrightarrow A, B}$ حدثين متنافيين Mutually Exclusive Events

٩- الحدث الذي يتكون من عنصر واحد يسمى حدث بسيط

١٠- الحدث الذي يتكون من عنصرين او اكثر يسمى حدث مركب

ملاحظة

اذا كانت تجربة مركبة من تجربتين متتاليتين وكان فضاء التجربة الاولى S_1 والثانية S_2 فان:

١- فضاء العينة للتجربة المركبة = $S_1 \times S_2$ (حاصل ضرب ديكارتي)

٢- $n(S) = n(S_1) \times n(S_2)$ ((مبدأ العد))

مثال

التجربة : القاء حجر نرد ثم قطعة نقود ثم حجر نرد مرة اخرى هنا مركبة من التجارب الثلاث الاتية :

الحل: $S_1 = \{ ١ ، ٢ ، ٣ ، ٤ ، ٥ ، ٦ \}$ حجر النرد الاول

$S_2 = \{ H , T \}$ حيث الصور H والكتابة T

$S_1 = \{ ١ ، ٢ ، ٣ ، ٤ ، ٥ ، ٦ \}$ حجر النرد الثاني

فأن $S = S_1 \times S_2 \times S_3$ (يمثل فضاء العينة للتجربة المركبة)

$\therefore$ عدد عناصر فضاء العينة للتجربة المركبة $n(S) = n(S_1) \times n(S_2) \times n(S_3)$

$n(S) = 6 \times 2 \times 6 = 72$

مثال

اقراص مرقمة من ١٠ الى21 سحب منها قرص واحد جد نسبة احتمال ان هذا القرص يحمل عددا زوجياً او عدداً يقبل القسمة على ٣ بدون اي باقي.

الحل:

$S = \{ 10 , 11 , 12 , 13, 14 , 15 , 16 , 17 , 18 , 19 , 20 , 21 \}$

$n(S) = 21 - 10 + 1 = 12$ (فضاء العينة يحوي ١٢ عنصر)

ليكن A حدث يحمل عدداً زوجياً $n(A) = 6$

$$P(A) = \frac{n(A)}{n(S)} = \frac{6}{12} = \frac{1}{2}$$

ليكن B حدث للعدد يقبل القسمة على ٣ بدون باقي

$B = \{ 12 , 15 , 18 , 21 \}$

$$P(B) = \frac{n(B)}{n(S)} = \frac{4}{12} = \frac{1}{3}$$

$$A \cap B = \{12,18\}$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A \cup B) = \frac{6}{12} + \frac{4}{12} - \frac{2}{12} = \frac{8}{12} = \frac{2}{3}$$

مثال

شركة افرادها ٦٠ رجلاً و ٢٠ امرأة ، من الرجال ٣٥ رجل متزوج ومن النساء ١٢ متزوجة من هذه الشركة اختير شخص واحد عشوائياً ، جد احتمال ان يكون:

١- هذا الشخص رجل

٢- هذا الشخص امرأة غير متزوجة

الحل:

١- ليكن A الحدث (لشخص رجل)

n(S) = 60 + 20 =80

$$\mathbf{P(A) = \frac{60}{80} = \frac{3}{4}}$$

٣- ليكن B الحدث (لشخص امرأة غير متزوجة)

$$\mathbf{P(B) = \frac{8}{80} = \frac{1}{10}}$$

مثال

القيا حجري نرد متمايزين مرة واحدة ، جد احتمال ان يكون مجموع العددين على الوجهين الظاهرين يساوي ١٠ او مجموع العددين على الوجهين الظاهرين يساوي ٩ .

الحل:

n(S) = 6 x 6 = 36

ليكن A = الحدث : مجموع العددين على الوجهين الظاهرين = ١٠

$$A = \{(5,5),(4,6),(6,4)\}$$

$$P(A) = \frac{n(A)}{n(S)} = \frac{3}{36}$$

ليكن B الحدث : مجموع العددين على الوجهين الظاهرين = ٩

$$B = \{(4,5),(5,4),(3,6),(6,3)\}$$

$$P(B) = \frac{4}{36}$$

$$P(A \cap B) = \varnothing$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A \cup B) = \frac{3}{36} + \frac{4}{36} = \frac{7}{36}$$

مثال

رمينا حجرين متمايزين من احجار النرد مرة واحدة ، ما احتمال ان يكون العدد على وجه احد الحجرين ضعف العدد على الوجه الآخر او العددين على الوجهين الظاهرين مجموعهما ٦.

الحل:

ليكن A = الحدث : العدد على الوجه الظاهري لاحد الحجرين ضعف العدد على الوجه الآخر .

$$A=\{(3,6),(6,3),(2,4),(4,2)(1,2),(2,1)\}$$

$$n(A)=6$$

$$P(A)=\frac{n(A)}{n(S)}=\frac{6}{36}$$

ليكن B = الحدث : مجموع العددين على الوجهين = ٦

$$B=\{(3,3),(2,4),(4,2),(5,1),(1,5)\}$$

$$n(B)=5$$

$$P(B)=\frac{n(B)}{n(S)}=\frac{5}{36}$$

$$A\cap B=\{(2,4),(4,2)\}$$

$$P(A\cap B)=P(A)+P(B)-P(A\cap B)$$

$$P(A\cap B)=\frac{6}{36}+\frac{5}{36}-\frac{2}{36}=\frac{9}{36}=\frac{1}{4}$$

الاحداث المتنافية

يقال عن الحادثين E1 وE2 انهما متنافيان اذا استحال حدوثهما معاً

فمثلاً اذا رمي زهر النرد مرة واحدة فمن المستحيل ظهور الوجه ٦ و ٥ في نفس الوقت

وكذلك اذا رميت قطعة نقود فمن المستحيل الحصول على صورة وكتابة في نفس الوقت

الاحداث المستقلة

يعتبر الحادثين A أو B حادثين مستقلين إذا كان وقوع إحداهما أو عدم وقوعه لا يؤثر في وقوع الآخر

. فمثلاً عند رمي قطعة نقود واحدة مرتين متتاليتين فإن نتيجة الرمية الثانية لا تتأثر بنتيجة الأولى.

الاحداث غيرالمستقلة

هي الحوادث التي اذا وقع احدها يؤثر في وقوع الاحداث الاخرى

فمثلاً في حالة صندوق به كرات صفراء وحمراء وبيضاء فعند سحب كراتان على التوالي بحيث لا تعاد الكرة الاولى فان نتيجة السحبة الثانية تتأثر بنتيجة السحبة الاولى لذا فالحادثان غير مستقلين.

الحالات المتماثلة

هي الحالات المتكافئة والمتساوية في امكانية حدوثها

فمثلاً عند رمي قطعة النقود فان الظروف المهيأة للحصول على اي وجه (صورة او كتابة) تكون متكافئة

الحالات المواتية

هي النتائج او الحالات التي تؤدي إلى تحقيق الحادث الذي هو موضع اهتمامنا

فمثلاً إذا كان الحادث هو الحصول على رقم فردي في حالة رمي زهرة النرد فإن الحالات التي تحقق هذا الحادث هي الحصول على ١ أو ٣ أو ٥ ، هذه الحالات الثلاثة تسمى الحالات المواتية.

الحالات الممكنة

هي الحالات أو النتائج المختلفة التي يمكن أن تظهر نتيجة لإجراء تجربة معينة

فمثلاً عند رمي قطعة نقود تكون نتيجتها صورة أو كتابة ، وعند رمي زهرة نرد تكون نتيجتها ١ أو ٢ أو ٣ أو ٤ أو ٥ أو ٦ فيقال أن عدد الحالات الممكنة ٢ في حالة رمي قطعة النقود و ٦ في حالة رمي زهرة النرد.

الاحداث المؤكدة

الحدث المؤكد هو الحدث الذي يضم كافة عناصر فضاء

كحدث ظهور عدد أقل من ٧ في تجربة إلقاء حجر النرد.

الاحداث المكملة

الاحداث التي عند اتحادها تساوي فضاء العينة

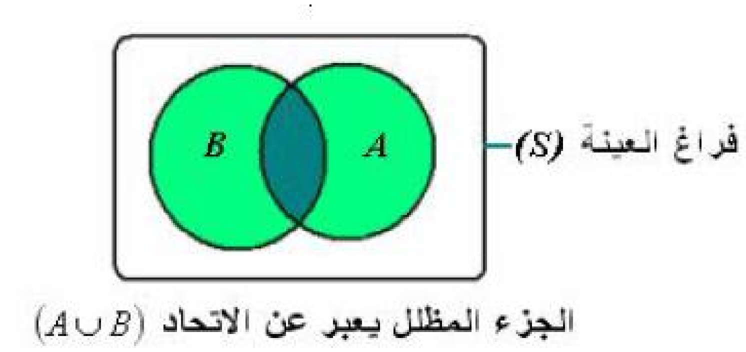
فمثلاً عند رمي زهر النرد فان احتمال الحصول على عدد فردي او زوجي فهي احداث مكملة وان اتحادهما يساوي فضاء العينة.

قوانين الاحتمالات

في البداية يجب ان نحدد بعض المفاهيم الجبرية التي سنستخدمها في قوانين الاحتمالات

- **الاتحاد $(\cup)$ Union**

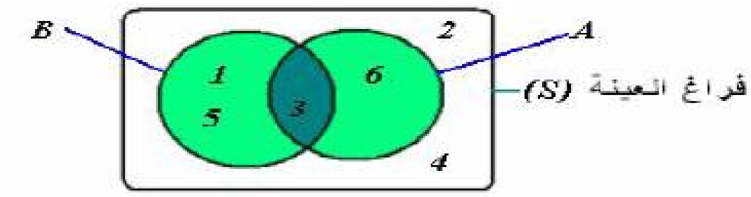
يعبر عن اتحاد الحدثين A و B عن وقوع احدهما على الاقل ، وبمنى اخر وقوع الاول او الثاني او كلاهما ويعبر عن ذلك رياضياً $(A \cup B)$ او $(A \text{ or } B)$ ، ويمكن الاستعانة بشكل " فن Ven Diagram " كما يلي:



ومثال على ذلك ، عند القاء زهرة النرد مرة واحدة ، وعرف الحادث A بانه ظهور وجه يقبل القسمة على ٣ ، والحادث B بانه ظهور عدد فردي ، يلاحظ ان

B:{1,3,5}　　A:{3,6}　　S:{1,2,3,4,5,6 }

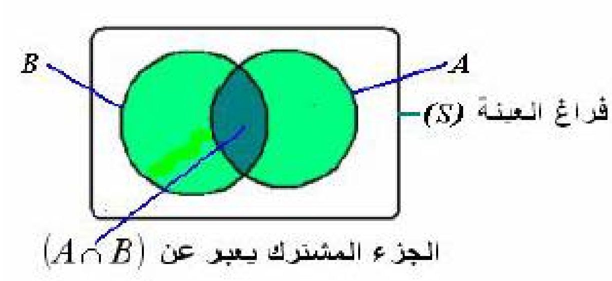
ويكون اتحاد الحادثان A و B هو {1,3,5,6} ويعبر عن ذلك في شكل Ven كما يلي :



$(A \cup B): \{1,3,5,6\}$

- **التقاطع** $(\cap)$ Intersection

يعبر تقاطع الحادثان A و B عن وقوع الاثنان في آن واحد ، ويشمل كل النتائج الممكنة المشتركة بين لحادثين ، ويعبر عن ذلك رياضياً $(A \cap B)$ او (A and B) ، ويظهر ذلك في شكل Ven كما يلي :



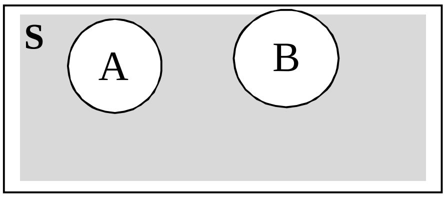
ففي المثال السابق نجد ان $(A \cap B) = \{3\}$

١ - قانون جمع الاحتمالات

أ - في حالة كون الحوادث متنافية Mutually Exclusive Events

احتمال وقوع أي حادث من الحوادث المتنافية يساوي مجموع احتمالات وقوع هذه الحوادث

فاذا كان الحدثين A و B حدثين متنافيين كما في الشكل ( )



شكل ( ) حوادث متنافية

فان $P\,(A \text{ or } B) = P\,(A) + P\,(B)$

ويرمز لها $P\,(AUB) = P\,(A) + P\,(B)$

مثال : في حالة رمي زهرة نرد ما هو احتمال الحصول على عدد فردي؟

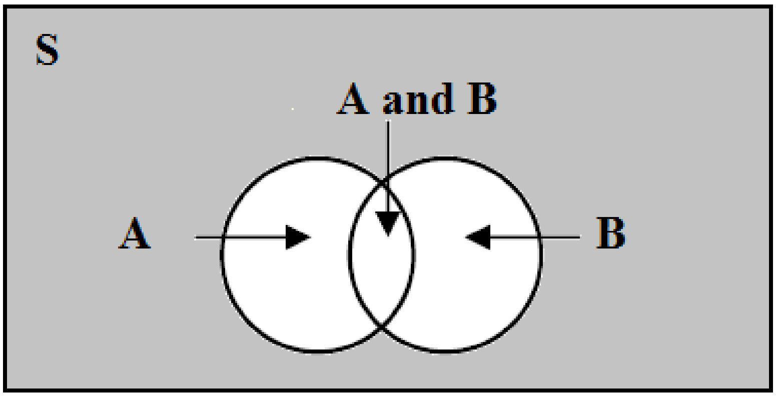
الحل : الحصول على عدد فردي معناه الحصول على١ أو٣ أو 5 وحيث

أن هذه الحوادث الثلاثة متنافية فإن

$$P(1 \text{ or } 3 \text{ or } 5) = P(1) + p(3) + P(5)$$
$$= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} \quad = \frac{1}{2}$$

ب - **في حالة كون الحوادث غير متنافية**

عند عدم اشتراط تنافي الحادثين A و B يكون المقصود بالحادث (B او A) وقوع A على انفراد او وقوع B على انفراد او وقوع الحادثين A و B معاً في وقت واحد معاً ، وكما يتضح من شكل ( )



شكل ( ) حوادث غير متنافية

الآن P (A) + P (B) تمثل مجموع الحالات المواتية للحادث A مضافاً اليها مجموع الحالات المواتية للحادث B ، ولكن يجب ملاحظة ان كل من الحالات المواتية للحادث A وتلك المواتية للحادث B تتضمن الحالات المواتية لوقوع A و B معاً ، وبهذا فانه في حالة جمع P (A) و P (B) فاننا نجمع

P (A و B) مرتين لهذا لا بد من طرح P (A و B) مرة واحدة للحصول على P (A أو B) وهذا هو :

$$P(\text{ A أو B }) = P(A) + P(B) - P(\text{ A و B})$$

او

$$P(AUB) = P(A) + P(B) - P(A \cap B)$$

مثال : ما هو احتمال ان يتم اختيار العضو A او D من بين اربعة اعضاء من مجلس ادارة احدى الشركات هم (A و B و C و D) لتمثيلهما في احد المؤتمرات ؟

الحل:

فضاء العينة (S) هو مجموع الحالات الممكننة

$$S = \{ AB , AC , AD , BC , BD , CD \}$$

- الحالات المواتية لاختيار A هي {AB , AC , AD }
- الحالات المواتية لاختيار D هي {AD , BD , CD }
- الحالات المواتية لاختيار A و D معاً هي {AD}

حسب قاعدة جمع الاحتمالات نحصل على النتيجة الآتية :

$$P(AUB) = P(A) + P(B) - P(A \cap B)$$

$$P(AUB) = \frac{3}{6} + \frac{3}{6} - \frac{1}{6} = \frac{5}{6}$$

مثال : إذا سحبت ورقة من مجموعة أوراق اللعب فما هو احتمال أن تكون الورقة عددًا أو من مجموعة معينة من المجموعات الأربعة المكونة (♠ ♣ ♦ ♥)لأوراق اللعب؟

الحل:

نفرض ان A يمثل الحصول على عدد ويمثل B الحصول على ورقة من مجموعة معينة ولتكن ♣

$$P(A) = \frac{40}{52}$$

$$P(B) = \frac{13}{52}$$

$$P(A \cap B) = \frac{10}{52}$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A \cup B) = \frac{40}{52} + \frac{13}{52} - \frac{10}{52} = \frac{43}{52}$$

مثال:

إذا كان احتمال نجاح محمد في أحد الاختبارات يساوي 0.25 واحتمال رسوب أحمد في هذا الاختبار يساوي 0.3 واحتمال نجاح محمد وأحمد معا في هذا الاختبار يساوي 0.1 فأوجد احتمال نجاح أحدهما على الأقل.

الحل: المعطيات نجاح محمد $P(A) = 0.25$

نجاح احمد $P(B) = 1- 0.3 = 0.70$

نجاح محمد واحمد $P(A \cap B) = 0.1$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.25 + 0.7 - 0.1 = 0.85$$

<u>مثال :</u>

إذا كان احتمال النجاح في مادة الرياضيات هو 0.45 واحتمال النجاح في مادة الإحصاء هو 0.65 واحتمال النجاح في المادتين معاً هو 0.37 أوجد احتمال النجاح في أحد المادتين على الأقل.

الحل: بتطبيق صيغة الاحتمالات للحوادث المتصلة بفرض أنَّ :

A : احتمال النجاح في مادة الرياضيات

B :   : احتمال النجاح في مادة الإحصاء

$A \cap B$ احتمال النجاح في المادتين معًا

فأنَّ:

$$\mathbf{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$$

$$\mathbf{= 0.45 + 0.65 - 0.37 = 0.73}$$

---

٢ - قانون الاحتمال الشرطي Conditional probability

يستند هذا الاحتمال على فرصة وقوع حادث، إذا توافرت معلومات عن وقوع حادث آخر له علاقة بالحادث الأول، كاحتمال نجاح الطالب في مادة الإحصاء إذا علم أنه من الناجحين في مادة الاقتصاد، وكاحتمال استخدام المزرعة لنوع معين من السماد، إذا علم أنه يقوم بزراعة محصول معين، وكاحتمال أن الخريجي يعمل بالقطاع الخاص، إذا علم أنه ممن تخرجوا من قسم معين من أقسام كلية الزراعة، والأمثلة على ذلك كثيرة.

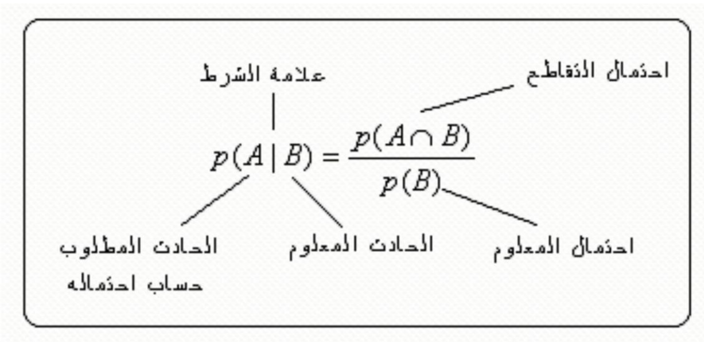
فإذا كان الحادث B حادث معلوم، والحادث A حادث آخر يراد حساب احتمال وقوعه، بشرط وقوع الحادث B فإن هذا الاحتمال يحسب بتطبيق المعادلة التالية:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

ويعرف الاحتمال $P(A|B)$ بقانون الاحتمال الشرطي، ويقرأ " احتمال وقوع الحادث A مشروطة بحدوث الحادث B" أو يقرأ" **احتمال وقوع الحادث** A بمعلومية الحادث B ". كما يمكن حساب احتمالية وقوع الحادث B بشرط وقوع الحادث A وذلك بتطبيق المعادلة الآتية:

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

ومن المعادلتين السابقتين يمكن ملاحظة بأن الاحتمال الشرطي هو نسبة بين حادث التقاطع الى الحادث المعلوم، حيث أن:



مثال :

فيما يلي توزيع تكراري لعينة عشوائية حجمها 100 من خريجي الكلية في العامين الماضيين،

حسب التخصص، ونوع المهنة:

التخصص	المهنة			
	عمل حر	قطاع خاص	قطاع حكومي	المجموع
علوم محاصيل	١٠	٥	١٥	٣٠
علوم اغذية	١٠	١٧	٨	٣٥
علوم تربة	١٣	١٠	١٢	٣٥
	٣٣	٣٢	٣٥	١٠٠

فإذا اختير أحد الخريجين بطريقة عشوائية، احسب الاحتمالات التالية:

1-ما احتمال أن يكون من خريجي قسم المحاصيل و يعمل بالقطاع الخاص.

2-ما احتمال أن يكون ممن يعملون بالحكومة أو من خريجي قسم علوم الأغذية.

3-ما احتمال أن يكون من خريجي قسم علوم الأغذية أو من قسم علوم التربة.

4-إذا علم أن الفرد من خريجي قسم عوم الأغذية، ما احتمال أن يكون ممن يعملون عملا حرا.

الحل:

أولا :نرمز لنوع المهنة بالرمز A ولنوع التخصص بالرمز B كما هو مبين بالجدول التالي

التخصص	المهنة			
	عمل حر A1	قطاع خاص A2	قطاع حكومي A3	المجموع
علوم محاصيل B1	١٠	٥	١٥	٣٠
علوم اغذية B2	١٠	١٧	٨	٣٥
علوم تربة B3	١٣	١٠	١٢	٣٥
المجموع	٣٣	٣٢	٣٥	١٠٠

ثانيا :التكرار في كل خلية يعبر عن عدد الخريجين الذين ينتمون لقسم معين و يعملون في مهنة

معينة، أي يعبر عن عدد تكرارات حوادث التقاطع الممكنة $A \cap B$. f

١- حساب احتمال أن يكون من خريجي قسم المحاصيل و يعمل بالقطاع الخاص.

$$P(B1 \cap A2) = \frac{f(B1 \cap A2)}{n} = \frac{5}{100} = 0.05$$

٢- حساب احتمال أن يكون ممن يعملون بالحكومة أو من خريجي قسم علوم الأغذية.

$$P(A_1 \cup B_2) = P(A_1) + P(B_2) - P(A_1 \cap B_2)$$

$$P(A_1 \cup B_2) = \frac{35}{100} + \frac{35}{100} - \frac{8}{100} = \frac{62}{100} = 0.62$$

٣- حساب احتمال أن يكون من خريجي قسم علوم الأغذية أو من قسم علوم التربة. هذان حادثان متنافيان، لأن تخرج الفرد من أحد الأقسام ينفي تخرجه من الأقسام الآخرى، وبمعنى آخر استحالة أن الفرد تخرج من قسمين في آن واحد، لذا يكون احتمال اتحادهما هو:

$$P(B_2 \cup B_3) = P(B_2) + P(B_3)$$

$$P(B_2 \cup B_3) = \frac{35}{100} + \frac{35}{100} = \frac{70}{100} = 0.70$$

٤- إذا علم أن الفرد من خريجي قسم علوم الأغذية، ما احتمال أن يكون ممن يعملون عملاً حراً، بشرط هذا احتمال شرطي، المطلوب هنا " حساب احتمال أن الفرد ممن يعملون عملا حراً A_3 بشرط أنه من خريجي قسم علوم أغذية B_2 أي أن الاحتمال المطلوب هو:

$$P(A_3 | B_2) = \frac{P(A_3 \cap B_2)}{P(B_2)}$$

$$P(A_3 | B_2) = \frac{\left[\frac{10}{100}\right]}{\left[\frac{35}{100}\right]} = \frac{10}{35}$$

٣- قانون ضرب الاحتمالات

أ - إذا كان لدينا الحادثين المستقلين A و B فان :

أن احتمال حدوث حادثين مستقلين أو أكثر معًا يساوي حاصل ضرب احتمال حدوث كل واحد من هذه الحوادث ببعضها بعضًا

$$\mathbf{P(A \cap B) = P(A)\, P(B)}$$

مثال

ما هو احتمال الحصول على الوجه (3 , 3) عند رمي زوج من أحجار النرد؟

الحل:

أن احتمال الحصول على الوجه 3 لدى رمي الحجر الأول من النرد هو $\frac{1}{6}$

وكذلك احتمال لحصول على الوجه ٣ لدى رمي الحجر الثاني من النرد هو $\frac{1}{6}$

$$\mathbf{P}\ (\mathbf{A} \cap \mathbf{B})\ =\ \mathbf{P}\ (\mathbf{A})\ \mathbf{P}\ (\mathbf{B})$$

$$\mathbf{P}\ (\mathbf{A} \cap \mathbf{B})\ = \frac{1}{6} \mathrm{X} \frac{1}{6}\ = \frac{1}{36}$$

ب - إذا كان لدينا الحادثين A و B غير مستقلين فان :

احتمال حدوثهما معاَ يساوي حاصل ضرب احتمال وقوع الحادث الاول في احتمال وقوع الحادث الثاني مشروطاً بحدوث الاول اي:

$$P(A \cap B) = P(A)\ P(B|A)$$

مثال:

مثال:

إذا كانت نسبة مزارع الخضروات التي تستخدم أسلوب معين للتسميد 60 % ، وإذا كان

نسبة المبيعات من إنتاج الخضروات المسمدة 70 % ، بينما نسبة المبيعات من الخضروات غير المسمدة (الخضروات العضوية) 80% ، إذا اختيرت أحد المزارع التي تنتج الخضروات عشوائيا ، فأوجد الآتي:

1-ما احتمال أن هذه المزرعة تستخدم أسلوب التسميد؟

2-إذا علم أن هذه المزرعة تستخدم أسلوب التسميد، ما احتمال أن تبيع إنتاجها؟

3-ما احتمال أن هذه المزرعة تستخدم أسلوب التسميد وتبيع إنتاجها؟

4-ما احتمال أن هذه المزرعة ممن لا يستخدمون أسلوب التسميد و تبيع إنتاجها؟

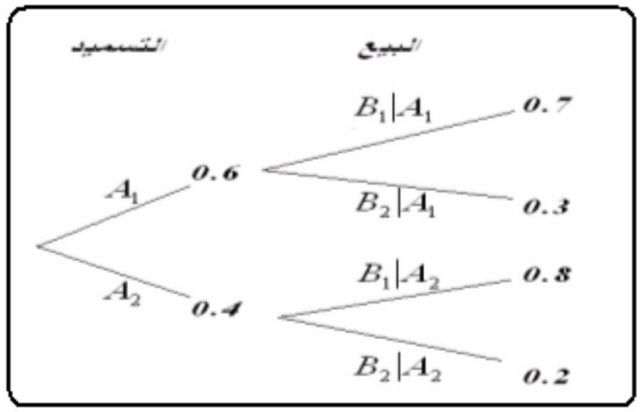
الحل

إذا فحصنا حال المزرعة المسحوبة، نجد أننا نتعامل مع نتيجتين متعاقبتين هما:

النتيجة الأولي ولها حالتان :{ المزرعة تستخدم طريقة التسميد (A_1) او المزرعة التي لا تستخدم التسميد(A_2)}

النتيجة الثانية ولها حالتان : { المزرعة التي تبيع الانتاج (B_1) ، او المزرعة التي لا تبيع الانتاج (B_2) }

لذا يمكن استنتاج شجرة الاحتمالات للحصول على النتائج الكلية كالتالي:



وفيما يلي حساب الاحتمالات:

١. احتمال أن المزرعة تستخدم أسلوب التسميد هو: P(A1)=0.60

٢- إذا علم أن هذه المزرعة تستخدم أسلوب التسميد، فإن احتمال أن تبيع إنتاجهاهو:

$$P(A_1 | B_1) = 0.70$$

٣- احتمال أن هذه المزرعة تستخدم أسلوب التسميد وتبيع إنتاجها عبارة عن احتمال وقوع حادثتان **معا** (A_1 and B_1) ، لذا يحسب هذا الاحتمال بتطبيق المعادلة **كما يلي:**

$$P(A_1 \cap B_1) = P(A_1)\ P(B_1 | A_1)$$
$$P(A_1 \cap B_1) = \quad (0.6)\ (0.70) \ = 0.42$$

٤- احتمال أن المزرعة لا تستخدم أسلوب التسميد وتبيع إنتاجها هو:

$$P(A_2 \cap B_1) = P(A_2)\ P(B_1 | A_2)$$
$$P(A_2 \cap B_1) = \quad (0.4)\ (0.80) \ = 0.32$$

<u>مثال</u>

إذا كان نسبة المزارع التي تنتج خضروات 60 % ، ونسبة المزارع التي تنتج فاكهه 75 % ونسبة المزارع التي تنتج الخضروات و الفاكهة 50 % ، أوجد الآتي:

1-ما احتمال أن مزرعة ما تنتج فاكهة أو خضروات؟

2-ما احتمال ألا تنتج المزرعة الفاكهة ؟

3-هل انتاج المزرعة للفاكهة مستقل عن إنتاجها للخضروات؟

الحل:

بفرض ا ن A حادث يعبر عن " المزرعة التي تنتج خضروات" ، و B هو حادث يعبر عن " المزرعة التي تنتج فاكهه" فان:

$$P(A) = 0.6 , \quad P(B) = 0.75 , \quad P(A \cap B) = 0.50$$

ويكون:

1-احتمال أن مزرعة ما تنتج فاكهة أو خضروات هو:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
$$P(A \cup B) = 0.60 + 0.75 - 0.50 = 0.85$$

٢- احتمال ألا تنتج المزرعة الفاكهة هو:

$P(B)$ احتمال تنتج فاكهه $\qquad$ $P(\overline{B})$ احتمال الا تنتج فاكهه (حدث مكمل)

$$P(\overline{B}) = 1 - P(B) = 1 - 0.75 = 0.25$$

٣- لمعرفة ما إذا كان إنتاج المزرعة للفاكهة مستقل عن إنتاجها للخضروات، يجب ان تكون المعادلة الاتية

$P(A \cap B) = P(A)\ P(B)$ صحيحة

وبما ان

$$P(A \cap B) = 0.50$$

$$P(A)\ \ P(B) = (0.60)(0.75) = 0.45$$

$$\therefore \quad P(A \cap B) \neq P(A)\ \ P(B)$$

فإن إنتاج المزرعة للفاكهة (A) غير مستقلة عن انتاجها للخضروات (B)

مبدأ العد

اذا امكن اجراء عملية باحدى الطرق المختلفة عددها (m) وكان لدينا في الوقت نفسه عملية اخرى يمكن اجراؤها بطرق عددها (n) فان عدد الطرق التي يمكن بها اجراء العمليتين معاً يساوي m x n .

مثال

يوجد لدى صاحب مخزن ثلاثة انواع من الدراجات الهوائية ومن كل نوع يوجد اربعة احجام ومن كل حجم يوجد ست دراجات فما عدد الدراجات ؟

الحل :

عدد الدراجات = ٣ x ٤ x ٦

= ٧٢ دراجة

مثال

كم عدد رمزه مكون من ثلاث مراتب يمكن تكوينه من مجموعة الارقام $\{1,2,5,7,8,9\}$

أ- التكرار مسموح

ب- التكرار غير مسموح

الحل: أ- التكرار مسموح

عدد اختيار الرقم الاول = ٦

عدد اختيار الرقم الثاني = ٦

عدد اختيار الرقم الثالث = ٦

عدد الاعداد = ٦ x ٦ x ٦ = ٢١٦

ب- التكرار غير مسموح

عدد اختيار الرقم الاول = ٦

عدد اختيار الرقم الثاني = ٥

عدد اختيار الرقم الثالث = ٤

عدد الاعداد = ٦ x ٥ x ٤ = ١٢٠

مثال

كم عدد رمزه مكون من رقمين واصغر من ٤٠ يمكن تكوينه باستخدام الارقام $\{1,2,3,4,5\}$

أ- تكرار الرقم مسموح

ب- تكرار الرقم غير مسموح

الحل: أ-

عدد اختيار رقم العشرات = ٣

عدد اختيار رقم الاحاد = ٥

عدد الاعداد = ٣ x ٥ = ١٥

ب-

عدد اختيار رقم العشرات = ٣

عدد اختيار رقم الاحاد = ٤

عدد الاعداد = ٣ x ٤ = ١٢

مثال

كم عدد مكون من ثلاثة مراتب واكبر من ٥٠٠ يمكن تكوينه من

الارقام $\{1,2,3,4,5,6,7\}$

أ- التكرار مسموح

ب - التكرار غير مسموح

الحل: أ-

عدد اختيار رقم المئات = ٣

عدد اختيار رقم العشرات = ٧

عدد اختيار رقم الآحاد = ٧

عدد الاعداد = ٣ x ٧ x ٧ = ١٤٧

ب-

عدد اختيار رقم المئات = ٣

دد اختيار رقم العشرات = ٦

عدد اختيار رقم الآحاد = ٥

عدد الاعداد = ٣ x ٦ x ٥ = ٩٠

يظهر في احيان كثيرة في الرياضيات ضرب الاعداد الصحيحة الموجبة من العدد (n) حتى (١) ويرمز له

$\underline{|n}$ = **n!** ويقرأ مضروب n

$$\boxed{n! = n(n-1)(n-2)..........1}$$

مثال

$$5! = 5\times4\times3\times2\times1 = 120$$

ملاحظة

$$1! = 1$$

اثبت ان $0! = 1$

$$n! = n(n-1)$$

نفرض ان n = 1

$$1!=1(1-1)!$$
$$1=1\times 0!$$
$$\therefore 0!=1$$

مثال

اذا كان $\dfrac{(n+1)!}{(n-1)!}=30$ جد قيمة (n)

الحل :

$$\frac{(n+1)!}{(n-1)!}=30 \rightarrow \frac{(n+1)n(n-1)!}{(n-1)!}=30$$
$$(n+1)n=30$$
$$n^2+n-30=0$$
$$(n-5)(n+6)=0$$
$$\therefore n=5 \qquad n=-\ 6$$

العدد السالب (٦-) يهمل لان n يجب ان تكون عدد صحيح موجب

مثال

اذا كان ٥٠٤٠ = $n!$ فما قيمة n

الحل :

$$n! = 5040$$
$$n! = 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1$$
$$n! = 7!$$
$$n = 7$$

٥٠٤٠	٢
٢٥٢٠	٣
٨٤٠	٤
٢١٠	٥
٤٢	٦
٧	٧
١	

التباديل والتوافيق

Permutations and Combinations

مقدمة

يطلق مسمى توافقيات على ذلك الفرع من الرياضيات الذي يتناول التباديل والتوافيق. وللتباديل استخدامات عديدة تشمل تحويل المكالمات الهاتفية عبر الأسلاك، وجدولة الإنتاج فى المصانع. ومع استخدام الحاسوب، غدت التباديل مجالاً خصبًا للأبحاث، وذلك لسرعة الحاسوب في القيام بالحسابات المتكررة. وكثيرا ما نحتاج في حياتنا الخاصة الى معرفة عدد الطرق المختلفة التي يمكن بها ترتيب مجموعة من الاشياء سواء بشروط او بدون شروط ودراسة التباديل والتوافيق قد تساعدنا في التعرف على طريقة الحساب الخاصة بتحديد عدد طرق الترتيب.

التباديل

Permutations

يقصد بالتباديل بانها عدد طرق الاختيار المرتب التي يمكن تكوينها من عدة اشياء بأخذها كلها او بعضها ويرمز له nPr او P_r^n اي تباديل r من n

وقانونه هو

$$n\,Pr = \frac{n!}{(n-r)!}$$

ملاحظة

قوانين التباديل

$$1 - P_n^n = n! = n(n-1)(n-2)........1$$

$$2 - P_0^n = 1$$

$$3 - P_r^n = \frac{n!}{(n-r)!}$$

$$4 - P_r^n = P(n,r) = n(n-1)(n-2).........(n-r+1)$$

مثال

احسب P_3^8 : الحل حسب القانون الثالث

$$P_3^8 = \frac{8!}{(8-3)!} = \frac{8\times7\times6\times5!}{5!} = 336$$

او حسب القانون الرابع $P_3^8 = 8\times7\times6 = 336$

مثال

احسب P_4^4

الحل : حسب القانون الاول

$$P_4^4 = 4! = 4\times3\times2\times1 = 24$$

مثال

احسب P_0^5

الحل: حسب القانون الثاني $P_0^5 = 1$

او حسب القانون الثالث

$$P_0^5 = \frac{5!}{(5-0)!} = \frac{5!}{5!} = 1$$

مثال

جد عدد التباديل للحروف أ ، ب ، ج المأخوذة منها اثنين في كل مرة.

الحل :

$$P_2^3 = \frac{3!}{(3-2)!} = \frac{3\times2\times1}{1} = 6$$

مثال

ما عدد طرق توزيع ٤ اشخاص على اربعة وظائف شاغرة بحيث كل شخص له فرصة عمل متساوية مع الاخرين.

الحل :

$$P_4^4 = \frac{4!}{(4-4)!}\frac{4\times3\times2\times1}{0!} = \frac{24}{1} = 24$$

مثال

بكم طريقة يمكن لمجموعة من سبعة اشخاص في حفل ان يرتبوا انفسهم بحيث يجلسون في صف مستقيم فيه سبعة مقاعد.

الحل:

$$P_7^7 = \frac{7!}{(7-7)!} = \frac{7\times6\times5\times4\times3\times2\times1}{1} = 5040$$

مثال

جد قيمة (n) اذا كان $P_2^n = 90$

الحل:

$$P_2^n = 90$$

$$P_2^n = \frac{n!}{(n-2)!} = \frac{n(n-1)(n-2)!}{(n-2)!} = \frac{n(n-1)}{1}$$

$$n(n-1) = 90$$

$$n^2 - n - 90 = 0$$

$$(n-10)(n+9) = 0$$

$$n = 10 \quad \text{or} \quad n = -9$$

العدد السالب (٩-) يهمل لان n يجب ان تكون عدد صحيح موجب

التوافيق

Combinations

يقصد بالتوافيق بانها عدد طرق الاختيار غير المرتب التي يمكن تكوينها من عدة اشياء باخذها كلها او بعضها ويرمز لها nC_r او $\binom{n}{r}$ او C_r^n وقانونه

(n) $n!$

مثال

احسب كل من C_2^5 و C_3^8

الحل:

$$C_2^5=\frac{5!}{2!(5-2)!}=\frac{5\times4\times3\times2\times1}{2\times1\times3\times2\times1}=10$$

$$C_3^8=\frac{8!}{3!(8-3)!}=\frac{8\times7\times6\times5!}{3\times2\times1\times5!}=56$$

مثال

كم لجنة ثلاثية يمكن تكوينها من ستة اشخاص .

الحل:

$$C_3^6=\frac{6!}{3!(6-3)!}=\frac{6\times5\times4\times3!}{3\times2\times1\times3!}=20$$

مثال

اذا كان عدد اسئلة امتحان الرياضيات هو (٨) والمطلوب حل (٥) اسئلة فقط. بكم طريقة يمكن الاجابة

الحل:

$$C_5^8=\frac{8!}{5!(8-5)!}=\frac{8\times7\times6\times5!}{5!\times3\times2\times1}=56$$

مثال

بكم طريقة يمكن اختيار لجنة من ثلاثة رجال وسيدتين من بين (٧) رجال و (٥) سيدات.

الحل:

$$C_3^7 \times C_2^5 = \frac{7\times6\times5\times4\times3!}{3!(7-3)!}\times\frac{5\times4\times3!}{2!(5-2)!}$$

$$C_3^7 \times C_2^5 = \frac{7\times6\times5\times4}{4\times3\times2\times1}\times\frac{5\times4}{2\times1} = 35\times10 = 350$$

مثال

كيس فيه (١٠) كرات حمراء و (٦) بيضاء سحبت منه (٤) كرات معا دون ارجاع . ما عدد الطرق التي تكون فيها الكرات المسحوبة من نفس اللون.

الحل:

$$C_4^{10} + C_4^6 = \frac{10!}{4!(10-4)!} = \frac{6!}{4!(6-4)!}$$

$$C_4^{10} + C_4^6 = \frac{10\times9\times8\times7\times6!}{4\times3\times2\times1\times6!}+\frac{6\times5\times4!}{4!\times2\times1}$$

$$C_4^{10} + C_4^6 = 210+15 = 225$$

تمارين الفصل

١- رمينا حجرين من احجار النرد جد

أ- عدد عناصر فضاء العينة $n(S)$

ب- اكتب فضاء العينة S

ج- اكتب الحدث الذي فيه قيمة مجموع العددين على وجهي الحجرين اكبر او يساوي ٩.

د- اكتب الحدث الذي قيمة العدد الذي على وجه احد الاحجار ضعف العدد على الوجه الاخر.

هـ- اكتب الحدث الذي قيمة مجموع العددين على وجهي الحجرين يقبل القسمة على ٦ بدون باقي.

٢- عند رمي حجر النرد مرة واحدة ، اكتب الاحداث التالية:

أ- ظهور عدد اولي

ب- ظهور عدد زوجي

ج- ظهور عدد فردي

٣- رميت حجرين متمايزين من احجار النرد

أ- ما هو احتمال العددين الظاهرين هما ٦

ب- ما هو احتمال الحصول على مجموع ٧ من ١١.

٤- كيس يحتوي على ٢٠ كرة متجانسة في جميع عناصرها مرقمة من ١ الى ٢٠ ، سحبت كرة واحدة ، جد:

أ- احتمال العدد الذي تحمله الكرة عددا اصغر من ٩

ب- احتمال العدد الذي تحمله الكرة عدداً اكبر من ٥

٥- صندوق يحتوي على ٢١ قرص مرقم من ١ الى ٢٠ سحبت قرصان جد نسبة احتمال:

أ- القرصان زوجيان

ب- الاول زوجي والثاني فردي

٦- لدينا ٥٠ بطاقة مرقمة من ١ الى ٥٠ ، جد احتمال العدد على البطاقة المسحوبة :

أ- يقبل القسمة على ٥

ب- يقبل القسمة على ٧

ج- يقبل القسمة على ٥ او ٧

٧- يراد اختيار لجنة طلابية تتكون من ثلاثة اشخاص بين ١٢ طالب و ٤ طالبات ، ما هو احتمال كل مما يأتي:

أ- ان تكون اللجنة جميعها طلاب

ب_ ان يكون في اللجنة طالب واحد فقط

٨-جد قيمة n اذا كان

$$\text{a) } P_2^n = 110$$
$$\text{b) } P_5^n = 8P_4^n$$
$$\text{c) } C_2^n = 55$$

٩- احسب قيمة $\frac{1}{210}\left[P_3^7 + P_4^7\right]$

١٠- كم كلمة مختلفة الحروف مكونة من حرفين يمكن تكوينها من كلمة (القدس).

١١- يراد تشكيل لجنة من ستة اعضاء من بين ٥ طلاب و ٨ مدرسين ، فبكم طريقة يمكن ان تكون اللجنة محتوية على مدرسين اثنين؟

١٢- اذا كان عدد اسئلة امتحان مادة الاحصاء هو ١٠ اسئلة وكان المطلوب حل ٧ اسئلة منها على ان تختار ٤ من الخمسة الاولى ، فبكم طريقة يمكن الاجابة.

١٣- صندوق يحتوي على ٤ كرات حمراء و ٨ كرات بيضاء ، سحبت ٣ كرات معاً ، جد عددطرق سحب اثنتان حمراء وواحدة بيضاء.

الفصل الثامن

التوزيعات الاحتمالية

- توزيع ذي الحدين
- توزيع بواسون
- التوزيع الطبيعي
- التوزيع الطبيعي القياسي

ما معنى التوزيع الاحتمالي Distribution Probability

يمكن فهم التوزيع (التوزيع الاحتمالي) كشكل مشابه للمدرج التكراري Histogram ولكن المدرج التكراري يصف توزيع البيانات الحقيقية بينما التوزيعات الرياضية (النظرية) مثل التوزيع الطبيعي وغيره هي توز يعات نظرية لها معادلات محددة وجداول تبين الاحتمالات المختلفة ولذلك تسمى توزيعات احتمالية. فعندما نرسم المدرج التكراري لمتغير ما فإننا نحاول أن نتعرف على التوزيع الاحتمالي الذي يُشبهه لكي نستخدم هذا التوزيع الاحتمالي في التحاليل الإحصائية.

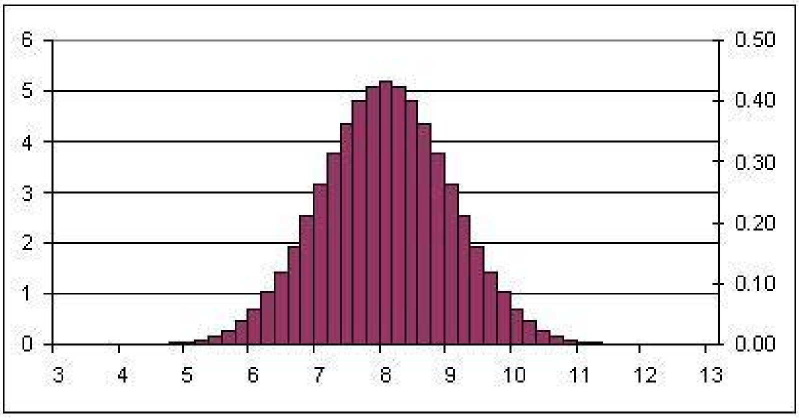
التوزيع يبين احتمالية أن يأخذ المتغير الذي ندرسه قيمة معينة أو أن يأخذ أقل أو أكثر من قيمة ما. فالتوزيع المنتظم Uniform يبين أن احتمالية أن يأخذ المتغير قيمة ما في مدى محدد متساوية بينما تجد الاحتماليات مختلفة في التوزيع الطبيعي. ففي التوزيع الطبيعي تكون الاحتمالية أعلى إذا كانت القيمة قريبة من المتوسط وتكون قليلة كلما ابتعدنا عن المتوسط. وهذه الاحتمالية يمكن تحديدها باستخدام الحاسوب أو الجداول.

افترض أنك تريد حساب محيط ومساحة منزلك. في البداية تقيس أبعاد الغرف ثم تقوم برسمها. بعد ذلك تبدأ في البحث عن أشكال هندسية تشابه أشكال الغرف مثل الشكل المستطيل أو المثلث أو شبه المنحرف أو المربع. وبعد تحديد الشكل الهندسي المشابه للغرفة تبدأ في حساب المحيط والمساحة باستخدام قوانين الهندسة الخاصة بكل شكل. هذا هو نفس الأمر بالنسبة لتغير متغير ما. إنك تقيس قيم هذا المتغير في فترة ما ثم تقوم برسمها كمدرج تكراري. بعد ذلك تبحث عن توزيع احتمالي يشبه هذا المدرج التكراري. وبعد تحديد التوزيع الاحتمالي المناسب تبدأ في استخدام جداوله أو استخدام الحاسوب للقيام ببعض التحاليل الخاصة بهذا المتغير.

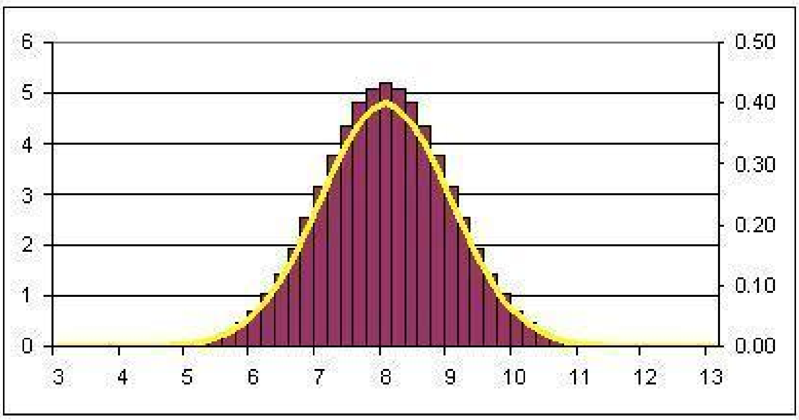
الكثير من التحاليل الإحصائية تعتمد على توزيع البيانات بنفس التوزيع الطبيعي ولذلك فإننا نرسم المدرج التكراري ونحاول مقارنته بمنحنى التوزيع الطبيعي. وهناك تطبيقات تفترض توزيع أسي Exponential

Distrintuion مثل نظرية خطوط الانتظار (الطوابير) أي أنها مبنية على افتراض أن زمن الخدمة يأخذ شكل التوزيع الأُسي.

والتوزيعات الاحتمالية لها أهمية في عمليات المحاكاة Simulation حيث نقوم بتحديد أقرب توزيع احتمالي للمدرج التكراري أي للتغيرات الحقيقية. وبناء عليه فإننا نستخدم هذا التوزيع في نموذج المحاكاة حيث يتم محاكاة التغير بنفس التوزيع ونفس القيم الحقيقية. افترض أننا قمنا برسم المدرج التكراري لمجموعة بيانات وحصلنا على الشكل التالي.



يمكننا البحث عن توزيع رياضي يشبه هذا المدرج التكراري والذي نرسمه بالخط الأصفر في الرسم التالي. في هذه الحالة فإن التوزيع المناسب هو التوزيع الطبيعي.



التوزيعات الاحتمالية Probabilities Distributions

- التوزيعات الاحتمالية المتقطعة
  - توزيع ذى الحدين **Binomial Distribution**
  - توزيع بواسون **Poisson Distribution**
- التوزيعات الاحتمالية المتصلة
  - التوزيع الطبيعي **Normal Distribution**

التوزيعات الاحتمالية المتقطعة

توزيع ذى الحدين Binomial Distribution

يستخدم هذا التوزيع في الحالات التي يكون للظاهرة محل الدراسة نتيجتان فقط متنافيتان، النتيجة محل الاهتمام وتسمى بحالة النجاح (p) ، والأخرى تسمى بحالة الفشل (q) ، ومن أمثلة ذلك

- : عند إعطاء مريض نوع معين من الأدوية، لها نتيجتان : استجابة للدواء، أو عدم استجابة
- · عند فحص عبوة بداخلها نوع معين من الفاكهة، لها نتيجتان إما أن تكون سليمة، أو تكون معيبة
- عند إلقاء قطعة عملة لها نتيجتان: ظهور الوجه الذي يحمل الصورة، أو الوجه الذي يحمل الكتابة
- نتيجة الطالب في الاختبار: نجاح، رسوب
- عند اختيار عينة عشوائية من إنتاج أحد المصانع وكنا نبحث عن المنتج المعيب
  - النجاح هنا هو الحصول على المنتج المعيب
  - الفشل هنا هو الحصول على المنتج السليم
- عند اختيار نتيجة طالبة من بين مجموعة من الطالبات فى مادة الإحصاء
  - النجاح هنا هو نجاح الطالبة فى مادة الإحصاء
  - الفشل هنا هو رسوب الطالبة فى مادة الإحصاء

⦿ تعريف محاولات برنولى Bernoulli Trials

هى سلسلة من المحاولات المكررة تحقق الآتى:

- كل محاولة لها نتيجتين فقط ونرمز لهما بالرمز p (نجاح) و q (فشل)
- كل المحاولات مستقلة أى أن نتيجة أى محاولة ليس لها أى تأثير على نتائج المحاولات الأخرى
- إحتمال النجاح واحتمال الفشل يظل ثابتا فى كل محاولة وهو p و q حيث p + q = 1

⦿ إذا عرفنا المتغير العشوائى X بأنه عدد مرات النجاح فى n من محاولات برنولى فإن قيم X الممكنة هى 0, 1, 2, 3, …, n

⦿ يكون للمتغير X توزيع يسمى توزيع ذى الحدين الاحتمالى وله الشكل العام

حيث p + q = 1

مفكوك ذى الحدين

$$p(X=x)=\binom{n}{x}p^{x}q^{n-x}\ ,\quad x=0,1,2,\dots,n$$

$$Where\quad \binom{n}{x}=\frac{n!}{x!(n-x)!}$$

<u>مثال</u>

إذا كان من المعلوم أن نسبة الشفاء من مرض معين باستخدام نوع معين من العقاقير الطبية هو 0.60

إذا تناول هذا العقار 5 مصابين بهذا المرض .إذا عرف المتغير العشوائي x بأنه عدد الذين يستجيبون

(حالات الشفاء) لهذا العقار.

المطلوب:

أ -ما هو نوع المتغير؟

ب -احسب الاحتمالات التالية:

ما هو احتمال استجابة 3 مرضى لهذا العقار؟

ما هو احتمال استجابة مريض واحد على الأقل؟

ما هو احتمال استجابة 2 مرضى على الأكثر؟

<u>الحل</u>

أ- عدد حالات الاستجابة x متغير كمي منفصل ، ومدى هذا المتغير في هذه الحالة هو

$$x = 0,1,2,3,4,5$$

ب- حساب احتمال استجابة 3 مرضى لهذا الدواء

$$p(X = x) = \binom{n}{x} p^x q^{n-x} \quad , \quad x = 0,1,2,\dots,n$$

$$Where \quad \binom{n}{x} = \frac{n!}{x!(n-x)!}$$

$$p(x=3) = \binom{5}{3}(0.6)^3(0.4)^{5-3}$$

$$p(x=3) = \frac{5x4x3x2x1}{3x2x1(2x1)}(0.216)(0.16)$$

$$p(x=3) = (10)(0.216)(0.16) = 0.3456$$

- **حساب احتمال استجابة مريض واحد على الأقل** $p(x \geq 1)$:

$$p(x \geq 1) = p(x=1) + p(x=2) + p(x=3) + p(x=4) + p(x=5)$$
$$p(x \geq 1) = 1 - p(x=0)$$
$$p(x \geq 1) = 1 - \binom{5}{0}(0.6)^0(0.4)^5$$
$$p(x \geq 1) = 1 - (1 \text{ x } 1 \text{ x } 0.01024) = 0.98976$$

- **حساب احتمال استجابة 2 مرضى على الأكثر** $p(x \leq 2)$:

$$p(x \leq 2) = p(x=2) + p(x=1) + p(x=0)$$
$$p(x \leq 2) = \binom{5}{2}(0.6)^2(0.4)^3 + \binom{5}{1}(0.6)^1(0.4)^4 + \binom{5}{0}(0.6)^0(0.4)^5$$
$$p(x \leq 2) = \frac{5\text{x}4}{2\text{x}1}(0.36)(0.064) + \frac{5}{1}(0.6)(0.0256) + 1(1)(0.01024)$$
$$p(x \leq 2) = 0.2304 + 0.0768 + 0.01024 = 0.31744$$

مثال ١

إذا كان احتمال إصابة الهدف لشخص ما هو 1/5 أتيحت له فرصة الرماية فى 10 محاولات

1- ما هو احتمال إصابة الهدف مرتين على الأكثر

2- احتمال إصابة الهدف مرة واحدة

الحل

X متغير عشوائى يمثل عدد مرات النجاح فى إصابة الهدف فى 10 محاولات

$n = 10$, $p = 1/5$, $q = 4/5$; $x = 0, 1, 2, \ldots, 10$

$$p(X=x)\binom{10}{x}(1/5)^x(4/5)^{10-x}, \quad x = 0,1,2,\ldots,10$$

1- احتمال إصابة الهدف مرتين على الأكثر

أى إحتمال $x = 0$ or $x = 1$ or $x = 2$

$$P(x \le 2) = p(x=0) + p(x=1) + p(x=2)$$

$$= \binom{10}{0}(1/5)^{0}(4/5)^{10} + \binom{10}{1}(1/5)^{1}(4/5)^{9} + \binom{10}{2}(1/5)^{2}(4/5)^{8}$$

٢- احتمال اصابة الهدف مرواحدة

<u>مثال</u>

اذا كان احتمال اصابة لاعب كرة القدم (A) الهدف في ضربة جزاء هو $\frac{3}{4}$ فما هو احتمال اصابة الهدف مرتين من اربعة ضربات جزاء.

$$p = \frac{3}{4} \quad , \quad q = \frac{1}{4} \quad , \; n = 4 \quad , \quad x = 2$$

$$p(x=2)=\binom{4}{2}(3/4)^2(1/4)^2$$

$$p(x=2)=\frac{4!}{2!(4-2)!}(\frac{9}{16})(\frac{1}{16})$$

$$p(x=2)=\frac{4x3x2x1}{2x1x2x1}(\frac{9}{16})(\frac{1}{16})=\frac{27}{128}=0.211$$

مثال

وجد في احد المصانع بان نسب العلب التالفة (المعابة) في معجون الطماطة التي ينتجها المصنع هي ٥% . فاذا اخذت عينة مؤلفة من ١٠ علب ، احسب احتمال :

(أ) ان تكون العينة كلها تالفة.

(ب) ان تكون العينة كلها جيدة.

(ت) ان تكون بالعينة ٣ علب تالفة فقط.

الحل

(أ)

$$p=0.05 \quad q=0.95 \quad n=10 \quad x=10$$

$$p(x=10)=\binom{10}{10}(0.05)^{10}(0.95)^{0}$$

(ب)

$$p=0.05 \quad q=0.95 \quad n=10 \quad x=0$$

$$p(x=0)=\binom{10}{0}(0.05)^{0}(0.95)^{10}$$

(ت)

$$p = 0.05 \quad q = 0.95 \quad n = 10 \quad x = 3$$

$$p(x = 3) = \binom{10}{3}(0.05)^3(0.95)^7$$

$$p(x = 3) = \frac{10!}{3!(10-3)!}(0.05)^3(0.95)^7$$

مثال

وجد فى إنتاج أحد المصانع أنه من بين 1000 وحدة إنتاج يوجد 150 وحدة معيبة. أخذت عينة بإرجاع مكونة من 5 وحدات، أوجد الاحتمالات التالية:

1- الوحدات المختارة كلها سليمة

2- على الأكثر توجد واحدة معيبة

3- على الأقل توجد وحدتان معيبتان

الحل

إحتمال النجاح (الحصول على وحدة معيبة) p = 150/1000 = 0.15

إحتمال الفشل (عدم الحصول على وحدة معيبة) q = 1-p= 1-0.15 =0.85

عدد المحاولات (عينة بإرجاع مكونة من 5 وحدات) n = 5

X متغير عشوائى يمثل عدد الوحدات المعيبة يأخذ القيم 0, 1, 2, 3, 4, 5

ويكون له توزيع ذى الحدين:

$$p(X=x)=\binom{5}{x}(0.15)^{x}(0.85)^{5-x}\ ,\quad x=0,1,2,3,4,5$$

(١) **الوحدات المختارة كلها سليمة** $x=0 \Leftarrow$

$$p(X=0)=\binom{5}{0}(0.15)^{0}(0.85)^{5-0}=\frac{5!}{0!(5-0)!}(1)(0.85)^{5}=0.4437$$

(٢) **على الأكثر توجد واحدة معيبة** $\Leftarrow$ $x \le 1$

$$p(X \le 1) = \binom{5}{0}(0.15)^0(0.85)^5 + \binom{5}{1}(0.15)^1(0.85)^4$$
$$= 0.4437 + \frac{5!}{1!5!}(0.15)(0.522)$$
$$= 0.4437 + 5 \times 0.0783 = 0.4437 + 0.3915 = 0.8352$$

(٣) **على الأقل توجد وحدتان معيبتان**

يعني أن $X \ge 2$

$$P(X \ge 2) = 1 - p(X < 2)$$

$$= 1 - [p(X = 0) + p(X = 1)]$$

$$= 1 - 0.8325 = 0.1648$$

تمارين

توزيع ذي الحدين

[١] اذا علمت بان نسبة العلب التالفة في مصنع كربلاءلمربى المشمش هو ١٠% ، فاذا اخذت عينة مؤلفة من ٦ علبوالمطلوب :

(أ) ايجاد احتمال ان تحتوي هذه العينة على اربعة علب تالفة.

(ب) ايجاد احتمال ان تحتوي هذه العينة على الاقل علبتين تالفتين.

[2]- في احدى البساتين الكبيرة كانت نسبة اصابة ثمار التفاح هي ٢٠% . فاذا اختيرت اربع تفاحات عشوائيا فما هو احتمال:

(أ) ان تكون واحدة مصابة.

(ب) ان تكون تفاحة واحدة على الاقل مصابة.

(ت) ان تكون هناك على الاكثر ثلاث تفاحات مصابة.

[3] - في عائلة مؤلفة من ٥ اطفال

(أ) ما هو احتمال ان يكون بينهم ذكر واحد.

(ب) ما هو احتمال انيكون بينهم على الاكثر ٣ بنات.

التوزيع الطبيعي Normal Distribution

التوزيع الطبيعي Normal Distribtion هو أشهر التوزيعات الاحتمالية وذلك لسببين.
السبب الأول هو أن الكثير من الظواهر تتبع منحنى التوزيع الطبيعي. السبب الآخر هو

أن هناك نظرية تقول أن متوسط قيم عينات متعددة يأخذ شكل التوزيع الطبيعي ولو لم يكن توزيع المتغير نفسه يتبع التوزيع الطبيعي. لذلك فإن التوزيع الطبيعي هو شيء محوري في علم الإحصاء.

يقال أن المتغير العشوائى المتصل X له توزيع طبيعى إذا كانت دالة كثافته لها الصورة

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

وهى دالة منحن له شكل الجرس متماثل حول المحور $X = \mu$ حيث:

π عدد ثابت يساوى تقريبا 3.1416

e عدد ثابت يساوى تقريبا 2.7183

μ هى القيمة المتوقعة لـ X يمكن أن يكون أى عدد حقيقى $E(X) = \mu$

$\sigma2$ هو التباين لـ X ويمكن أن يكون أى عدد حقيقى موجب $Var(X) = \sigma^2$

وهذه الدالة تحدد تماما متى علمت المعلمات μ, σ

ونقول أن X له توزيع طبيعى بتوقع μ وتباين σ^2

وتكتب $X \sim N(\mu, \sigma2)$

المنحنى متماثل حول خط المستقيم $X = \mu$

المحور X هو خط تقاربى للمنحنى الطبيعى

المنوال والوسيط والوسط الحسابى تساوى μ والتباين يساوى σ^2

نقط انقلاب المنحنى هى : $x = \mu - \sigma , x = \mu + \sigma$

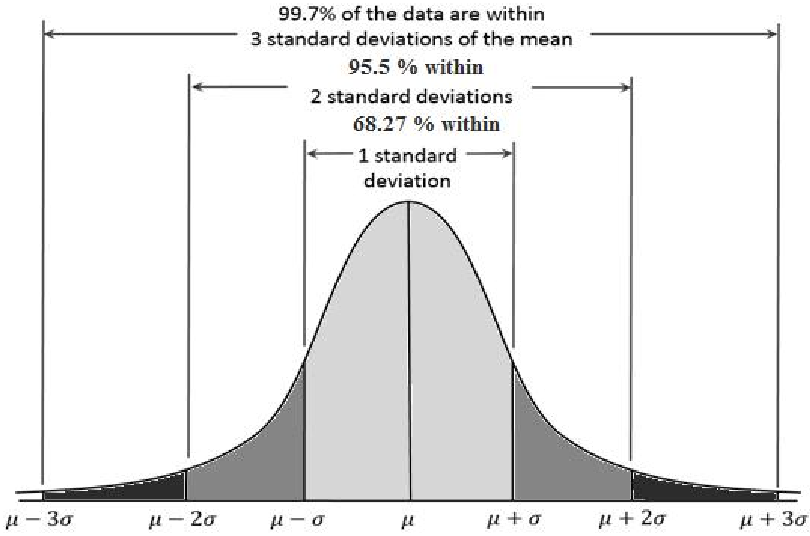
المساحة الكلية تحت المنحنى الطبيعى تساوى الواحد الصحيح

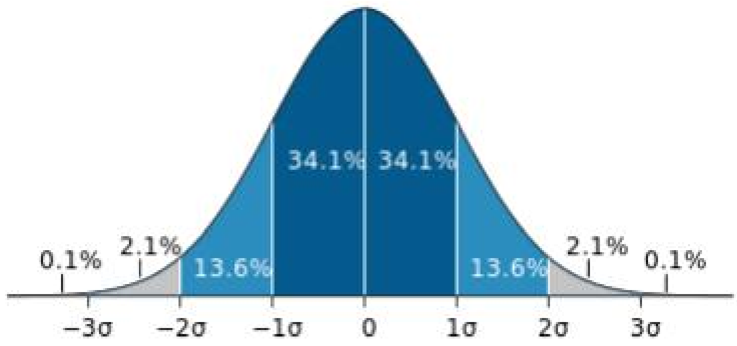
نسبة المساحة بين

$x = \mu - \sigma , x = \mu + \sigma$ هى 68.27%

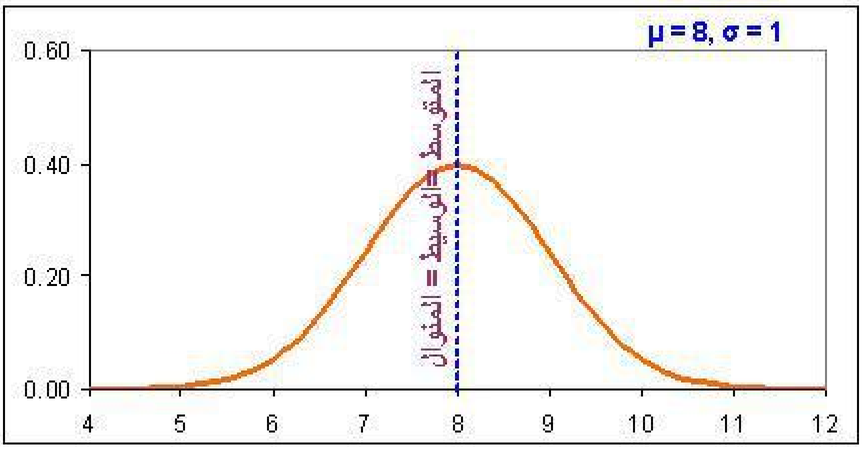
$x = \mu - 2\sigma , x = \mu + 2\sigma$ هى 95.5%

$x = \mu - 3\sigma , x = \mu + 3\sigma$ هى 99.7%



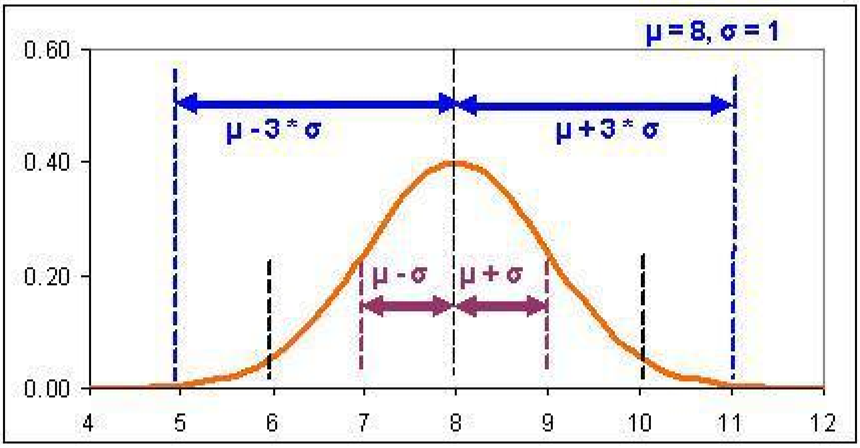


منحنى التوزيع الطبيعي يشبه الجرس (الناقوس) ويتميز بوجود تماثل بين جانبيه الأيمن والأيسر حول المتوسط. ومن سمات منحنى التوزيع الطبيعي أن المتوسط يساوي الوسيط ويساوي المنوال. يتم تعريف منحنى التوزيع الطبيعي بقيمتين: المتوسط والانحراف المعياري. ويرمز عادة للمتوسط بـ μ وللانحراف المعياري بـ σ الرسم التالي يبين شكل منحنى التوزيع الطبيعي وفي هذا المثال المتوسط 8 = μ لاحظ أن تماثل المنحنى يعني أن ٥٠% من القيم هي أقل من المتوسط و٥٠% من القيم هي أكبر من المتوسطو هذا يعني أن الوسيط يساوي المتوسط.



وكتذكرة سريعة فإن المتوسط هو مجموع القيم كلها مقسوما على عددها. والوسيط هو القيمة التي تكون ٥٠% منا لقيم أكبر منها. والمنوال هو القيمة الأكثر تكررا. والانحراف المعياري هو مقياس لبعد جميع القيم عن المتوسط أي مقياس لتشتت القيم.

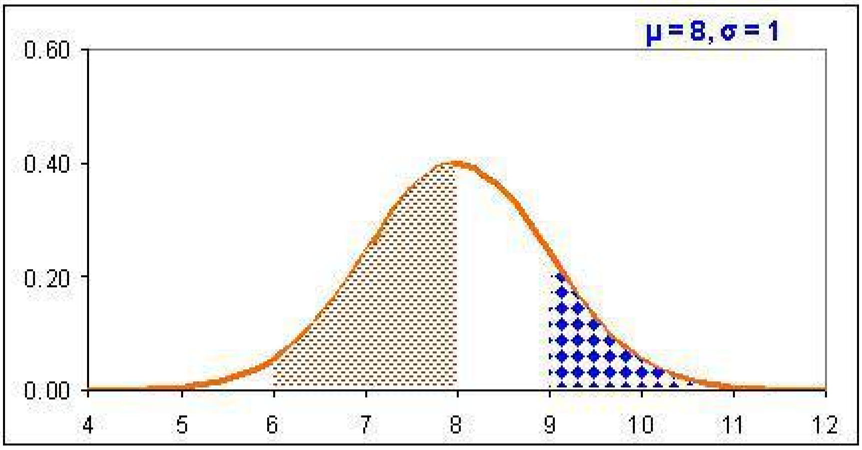
ولمنحنى التوزيع الطبيعي سمات رئيسية منها أن ٦٨% من الاحتمالات تقع في حدود المتوسط ±١ الانحراف المعياري. و٩٩,٧% من الاحتمالات تقع في حدود المتوسط ±٣ الانحراف المعياري. فلو عرفنا المتوسط والانحراف المعياري يمكننا حساب هذه الاحتمالات. لاحظ أن احتمال وقوع المتغير بين قيمتين تُمَثل بالمساحة تحت المنحنى بين هاتين القيمتين. ولذلك يمكننا بمجرد النظر أن نقول إن وقوع قيمة المتغير في الرسم أدناه بين ٨ و ٩ هي أعلى بكثير من وقوعه بين ١٠ و ١١ لأن المساحة تحت المنحنى بين ٨ و ٩ اكبر بكثير منها بين ١٠ و ١١.



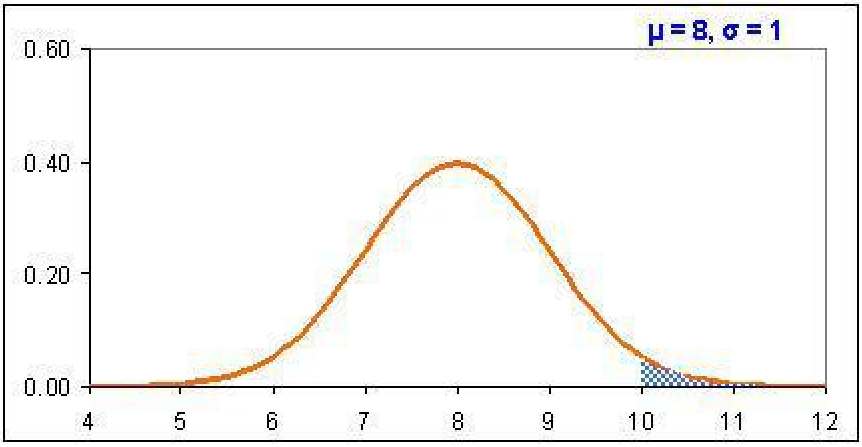
ففي الشكل أعلاه يمكننا أن نقول أن قيمة هذا المتغير في ٩٩,٧% من الحالات تقع بين ٥ و ١١. وأن قيمة هذا المتغير تتراوح بين ٧ و ٩ في ٦٨% من الحالات.

والتوزيع الطبيعي هو جزء أساسي من فكرة خرائط المراقبة. فالحدود القصوى والدنيا توضع عند $\mu \pm 3\sigma$. لماذا؟ لأنه في حالة التوزيع الطبيعي فإن احتمالية وقوع القيم في هذا المدى هي ٩٩,٧% كما ذكرنا منذ قليل. أي أن القيمة لو كانت خارج هذا المدى فهي لا تنتمي لنفس التوزيع أي أن شيئا غير طبيعي قد حدث.

ولذلك فإننا عندما نستخدم توزيع احتمالي مثل التوزيع الطبيعي أو المنتظم أو الأسي أوغيرهم فإننا نحدد الاحتماليات بالنظر للمساحة تحت المنحنى. فلو نظرنا للشكل أدناه لعلمنا أن وقوع هذا المتغير بين ٦ و ٨ (المساحة البنية اللون) هي أكبر بكثير من وقوعه بين ٩ و ١١ (المساحة الزرقاء اللون)

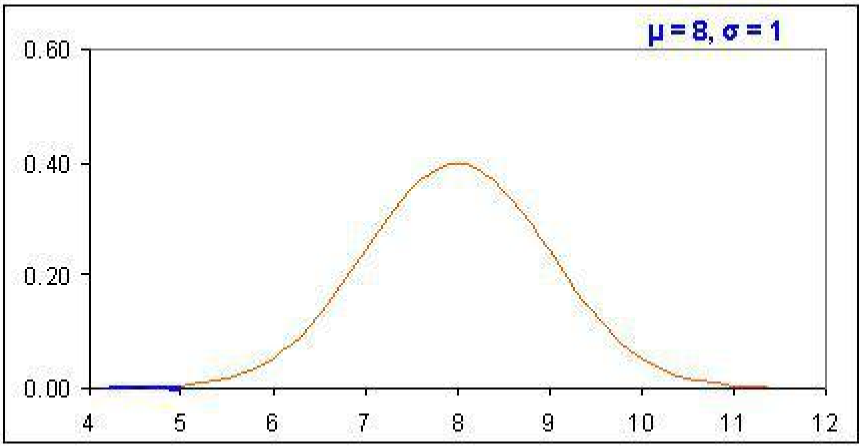


ويمكننا بنفس الطريقة تقدير احتمالية أن يتجاوز المتغير قيمة ما أو يقل عنها. فمثلا لو أحببنا أن نعرف احتمالية أن يزيد هذا المتغير عن ١٠ فإننا ننظر إلى المساحة المبينة في الشكل أدناه



.

ولو أحببنا أن نعرف احتمالية أن يقل هذا المتغير عن ٥ فإننا ننظر إلى المساحة تحت المنحنى من قيمة ٥ فما أقل وهي مساحة صغيرة جدا تقترب من الصفر (المساحة الزرقاء في الشكل أدناه.

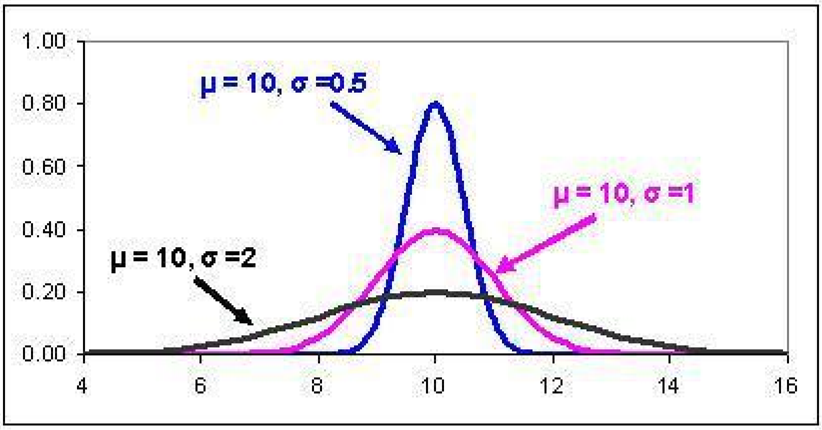


ومن هنا نعرف لماذا كانت معظم القيم (٩٩,٧%) في حدود $\mu \pm 3\ \sigma$ أي في هذا المثال من ٥ إلى ١١ لأن المساحة تحت المنحنى من ٥ إلى ١١ تكاد تكون هي المساحة كلها وتبقى مساحة ضئيلة جداً على الجانبين. وعملية حساب احتماليات وقوع المتغير بين

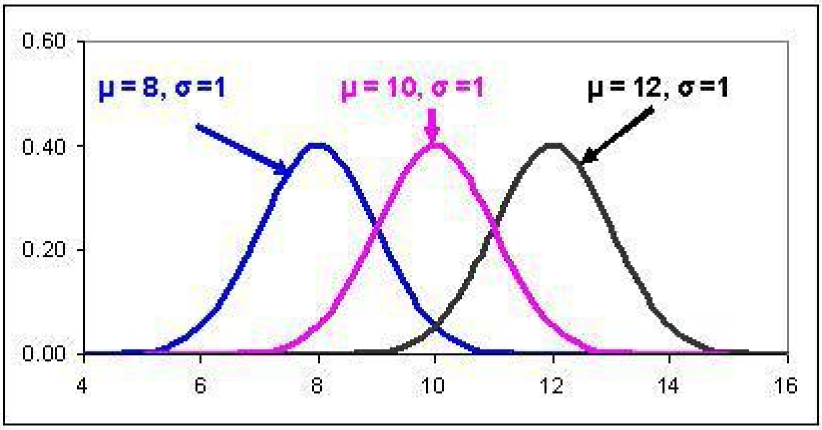
قيميتن أو أكبر من قيمة ما أو أقل من قيمة ما يتم تقديره على وجه الدقة باستخدام الجداول التي تعطي المساحة تحت المنحنى في كل جزء منه أو باستخدام الحاسوب.

تأثير تغير قيمة المتوسط أو الانحراف المعياري

الشكل التالي يبين تأثير تغير الانحراف المعياري مع ثبات المتوسط. إن ما يحدث هو أن المنحنى يقل انبعاجا كلما زادت قيمة الانحراف المعياري. وهذا مرتبط بأن الانحراف المعياري هو مقياس لتشتت المنحنى وبالتالي فكلما زاد الانحراف المعياري فإن هذا يعني أن المنحنى ينتشر على مدى أوسع. فعندما كان الانحراف المعياري يساوي ٠,٥ كان التوزيع قريب جدا من المتوسط بينما ازداد اتساعا عندما زادت قيمة الانحراف المعياري إلى ١ ثم ازداد اتساعا عندما وصلت قيمة الانحراف المعياري إلى ٢.



أما تغير المتوسط فيظهر في الرسم التالي. فالانحراف المعياري لكل منحنى من هذه المنحنيات متساوٍ بينما المتوسط مختلف. لاحظ أن المنحنيات الثلاثة متشابهة تماما ولكن كل منها يتوزع حول متوسط مختلف.



وترجع اهميته التوزيع الطبيعي الى اربع اعتبارات مهمة:

(١) ان كثيرا من المتغيرات تتوزع توزيعا طبيعيا فمعظم الصفات البيلوجية والاجتماعية وغيرها من الصفات المهمة يكون توزيعها مشابها للتوزيع الطبيعياومقاربا له.

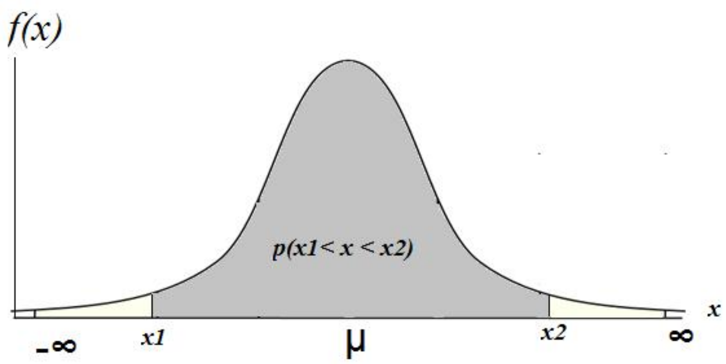
(٢) توزيعات المعاينة لمتوسطات العينات تكون مقاربة للتوزيع الطبيعي ويزداد هذا التقارب كلما زاد حجم العينة.

(٣) امكانية تحويل كثير من التوزيعات الى التوزيع الطبيعي.

(٤) ان معظم الاختبارات المستخدمة في الاستنتاج الطبيعي مبنية على كون المتغير يتوزع توزيعا طبيعيا.

كيفية حساب الاحتمالات

بفرض أن الاحتمال المطلوب حسابه هو $p(x_1 < x < x_2)$ **، وهذا الاحتمال يحدد بالمساحة التالية :**



وحيث أن هذا التوزيع من التوزيعات المستمرة، فإن هذه المساحة (الاحتمال)تحسب بإيجاد التكامل التالي:

$$p(x_1 < x < x_2) = \int_{x_1}^{x_2} f(x)dx = \int_{x_1}^{x_2} \frac{1}{\sigma\left(\sqrt{2\pi}\right)} e^{\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)} .dx$$

وهذا التكامل يصعب حسابه، ومن ثم لجأ الإحصائيين إلى عمل معادلة رياضية Transfom **يمكن استخدام توزيعها الاحتمالي في حساب مثل هذه الاحتمالات، وهذه المعادلة هي:**

$$z = \frac{x-\mu}{\sigma}$$

ويعرف المتغير الجديد z بالمتغير الطبيعي القياسي Standard Normal Variable **وهذا المتغير له دالة :**

$$f(z)=\frac{1}{\sqrt{2\pi}}\,e^{-\frac{1}{2}z^2},-\infty<z<\infty,\pi=\frac{22}{7}$$

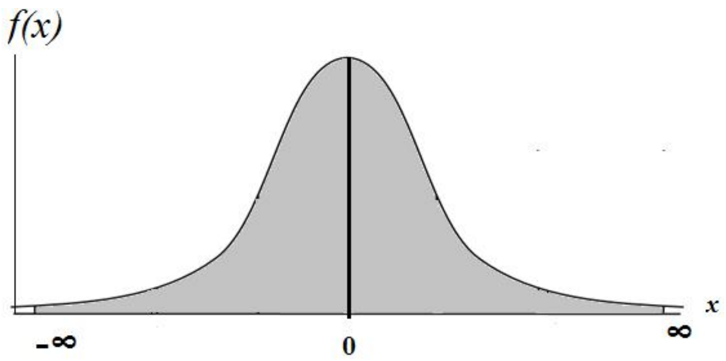
ومن خصائص هذا التوزيع ما يلي:

(١) متوسطه هو : $E(z)=0$

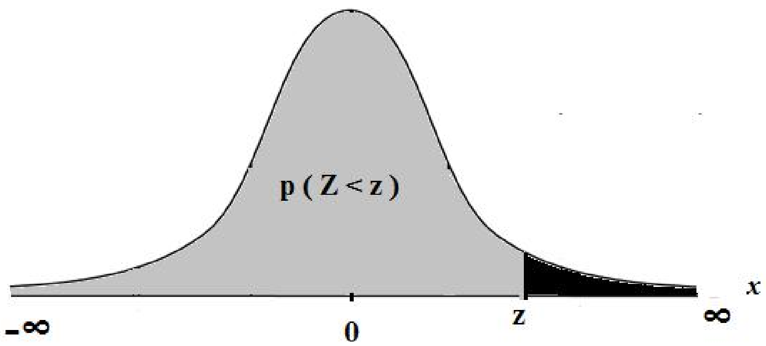
(٢) تباينه $\sigma^2(z)=1$

ومن ثم يعبر عن توزيع المتغير z بالرموز $z \sim N(0,1)$ ويعني ذلك ان المتغير العشوائي x يتبع التوزيع الطبيعي القياسي بمتوسط (٠) ، وتباين (١).

(٣) يأخذ المنحنى الشكل الناقوس المتماثل على جانبي الصفر.



وصمم الإحصائييون جداول إحصائية لحساب دالة التوزيع التجميعي $F(z)=p(Z<z)$ ، كما هو مبين في الرسم

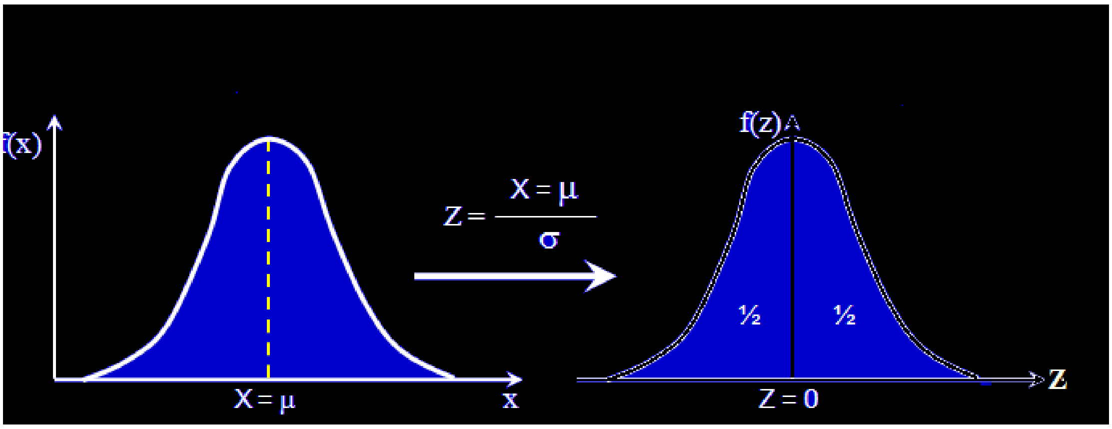


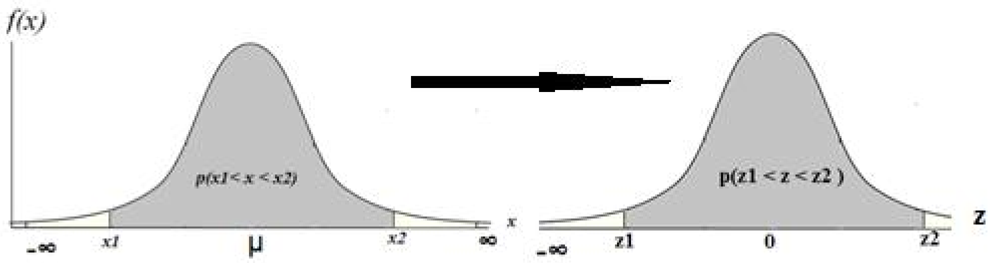
ونعود الآن إلى خطوات حساب الاحتمال $p(x_1<x<x_2)$ باستخدام المعادلة لتحويل المتغير العشوائي x الى متغير قياسي z وفق المعادلة الآتية $z=\frac{x-\mu}{\sigma}$

١. يتم تحويل القيم الطبيعية (x_1 , x_2) الى قيم طبيعية قياسية

$$z_1=\frac{x_1-\mu}{\sigma}\quad,\quad z_2=\frac{x_2-\mu}{\sigma}$$

٢-ومن ثم يكون الاحتمال $p(x_1 < x < x_2) = p(z_1 < z < z_2)$





٣- **تستخدم جداول التوزيع الطبيعي القياسي، والذي يعطي المساحة الخاصة بالاحتمال**

$$F(z) = P(Z < z)$$

ملاحظة: يمثل جدول التوزيع الطبيعي القياسي ما يلي

قيم المتغير القياسي z

المرتبة الثانية بعد الفارزة للمتغير Z

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
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

القيم داخل الجدول تمثل الاحتمالية

<u>**مثال**</u> : إذا كانت لدينا مجموعة من درجات الحرارة خلال شهر مارس هى 30, 33, 35, 40, 42 وإذا علم أن درجات الحرارة خلال هذا الشهر لها توزيع طبيعى بمتوسط مقداره 35 وانحراف معيارى 2. أوجد القيم المعيارية (القياسية) لدرجات الحرارة المعطاة.

الحل

إفرض أن X يمثل درجات الحرارة خلال شهر مارس ، $\sigma = 2$, $\mu = 35$

$$Z = \frac{x-\mu}{\sigma} = \frac{30-35}{2} = -2.5$$

$$Z = \frac{x-\mu}{\sigma} = \frac{33-35}{2} = -1$$

$$Z = \frac{x-\mu}{\sigma} = \frac{35-35}{2} = 0$$

$$Z = \frac{x-\mu}{\sigma} = \frac{40-35}{2} = 2.5$$

$$Z = \frac{x-\mu}{\sigma} = \frac{42-35}{2} = 3.5$$

مثال: إذا كان Z متغير عشوائى له التوزيع الطبيعى القياسي. إحسب الاحتمالات الآتية:

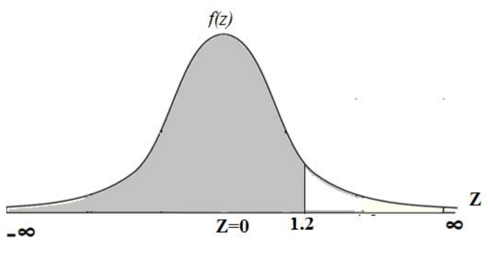
1. $P(Z < 1.2)$ 2. $P(Z < -0.11)$

3. $P(0.32 < Z < 1.24)$ 4. $P(Z < -0.15)$

الحل

١. $P(Z < 1.2)$

يمثل هذا الاحتمال المساحة المظللة تحت المنحنى



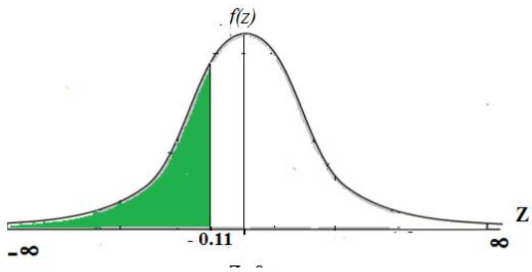
ويتم إيجاد هذه المساحة (الاحتمال) من جدول احصائى محسوب التكاملات فيه مسبقا لتوفير مجهود الحسابات كما هو موضح بالجدول:

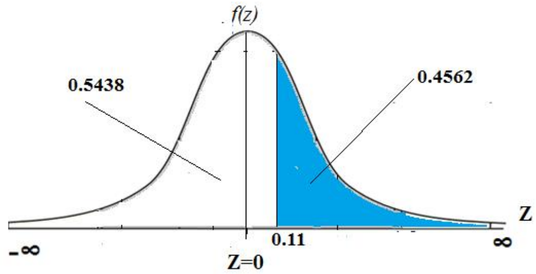
1. P(Z < 1.2) = 0.8849

Z	0.00	0.01	0.02	0.03	0.04	0.05			0.09
0.0		.5040							
0.1									
0.2									
1.2	.8849					.8944			

٢. P(Z < -0.11)

يمثل هذا الاحتمال المساحة المظللة تحت المنحنى





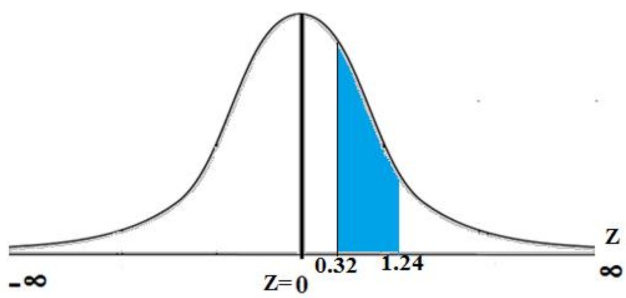
يمثل هذا الاحتمال المساحة المظللة الخضراء يسار المنحنى وهى تساوى المساحة الزرقاء يمين المنحنى وبالتالى فهى تساوى المساحة الكلية تحت المنحنى (تساوى الواحد) مطروحا منها المساحة الزرقاء (من الجدول)

ويتم إيجاد هذه المساحة الزرقاء (P(Z < 0.11)) من جدول احصائى محسوب التكاملات فيه مسبقا لتوفير مجهود الحسابات كما هو موضح بالجدول

2. $P(Z < -0.11) = 1 - P(Z < 0.11) = 1 - 0.5438 = 0.4562$

Z	0.00	0.01	0.02	0.03	0.04	0.05			0.09
0.0		.5040							
0.1		.5438							
0.2									
1.2	.8849					.8944			

٣. $P(0.32 < Z < 1.24)$



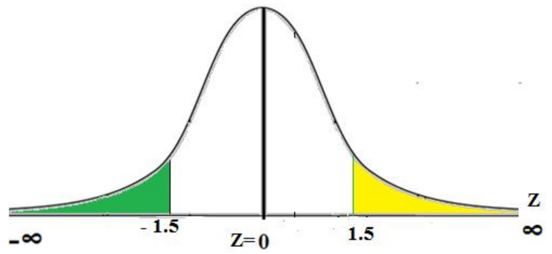
يمثل هذا الاحتمال المساحة المظللة الزرقاء ويتم إيجاد هذه المساحة ($P(Z < 1.24)$ والمساحة الزرقاء ($P(Z < 0.32)$ من جدول احصائى محسوب التكاملات فيه مسبقا لتوفير مجهود الحسابات كما هو موضح بالجدول

3. $P(0.32 < Z < 1.24) = P(Z < 1.24) - P(Z < 0.32) = 0.8925 - 0.6293 = 0.6270$

Z	0.00	0.01	0.02	0.03	0.04	0.05			0.09
0.0		.5040							
0.1		.5438							
0.2									
0.3			.6293						
1.2	.8849				.8925	.8944			

٤. $P(Z < -1.5)$

يمثل هذا الاحتمال المساحة المظللة الخضراء ونظرا لتماثل المنحنى فهى أيضا تساوى المساحة الصفراء أى أن :

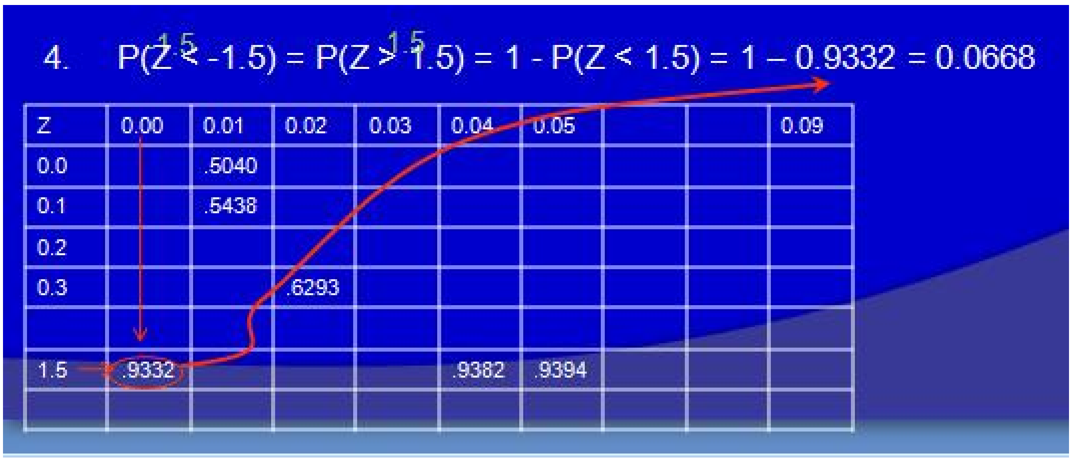


$$P(Z < -1.5) = P(Z > 1.5)$$

$$= 1 - P(Z < 1.5)$$

ويتم إيجاد هذه المساحة الخضراء $P(Z < 1.5)$ من جدول احصائى محسوب التكاملات فيه مسبقا لتوفير مجهود الحسابات كما هو موضح بالجدول ونطرحها من الواحد (المساحة الكلية تحت المنحنى) فتنتج المساحة الصفراء والتى تساوى بدورها المساحة الخضراء المطلوبة

$$P(Z < -1.5) = P(Z > 1.5) = 1 - P(Z < 1.5) = 1 – 0.9332 = 0.0668$$

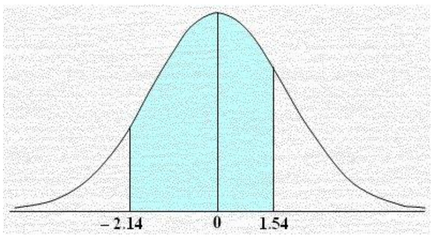


4. $P(Z < -1.5) = P(Z > 1.5) = 1 - P(Z < 1.5) = 1 – 0.9332 = 0.0668$

Z	0.00	0.01	0.02	0.03	0.04	0.05			0.09
0.0		.5040							
0.1		.5438							
0.2									
0.3			.6293						
1.5	.9332				.9382	.9394			

مثال(   )

احسب المساحة المحصورة بين1.54 , 2.14 – والواقعة تحت منحنى التوزيع الطبيعي والمبينة بالشكل المرفق.



الحل:

نعلم أن العدد 1.54 يقابله في جدول Z قيمة المساحة الواقعة يساره وكذلك العدد - ٢,١٤ تقابله مساحة في جدول Z والفرق بين المساحتين يعطينا المساحة المطلوبة.

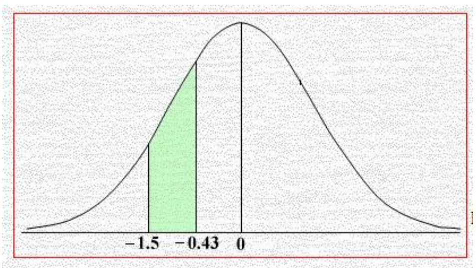
جدول Z

قيمة Z	الاحتمالية (المساحة)
1.54	0.9382
-2.14	0.0162

المساحة المطلوبة = **0.9382 – 0.0162= ٠,٩٢٢٠**

مثال ( )

احسب المساحة بين Z = – 0.43 , Z = – 1.5



الحل:

قيمة Z	الاحتمالية (المساحة)
-0.43	0.3336
-1.50	0.0668

المساحة المطلوبة = المساحة على يسار 0.43– مطروحاً منها المساحة على يسار1.5–

0.3336 – 0.0668 = 0.2668

مثال ( )

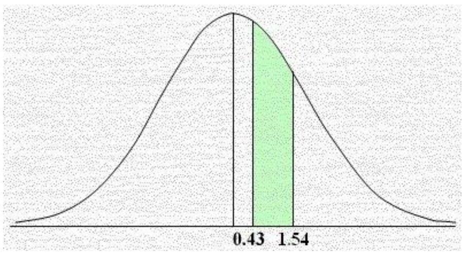
احسب المساحة بين Z = 0.43 , Z = 1.5

الحل:

المساحة المطلوبة = المساحة على يسار ١,٥ مطروحاً منها المساحة على يسار٠,٤٣

= 0.9332 – 0.6664

= 0.2668



أو

$P(0.43 < Z < 1.5) = P(Z < 1.5) - P(Z < 0.43)$

= 0.9332 – 0.6664

= 0.2668

مثال ( )

إذا كانت مجموعة مكونة من ٤٠٠ عضو في نادي تتوزع توزيعاً طبيعياً في العمر بمعدل ٤٠ سنة بانحراف معياري قدره ٥ فاحسب:

1) عدد الأعضاء الذين أعمارهم بين 35 إلى 45 سنة.
2) عدد الأعضاء الذين أعمارهم أقل من ٥٠
3) عدد الأعضاء الذين أعمارهم أقل من ٣٥ واكبر من ٤٥

الحل:

1- نحسب قيمة Z من القانون للعمر ٣٥ و ٤٥ :

$$z = \frac{x - \mu}{\sigma} = \frac{35 - 40}{5} = -1$$

$$z = \frac{x - \mu}{\sigma} = \frac{45 - 40}{5} = 1$$

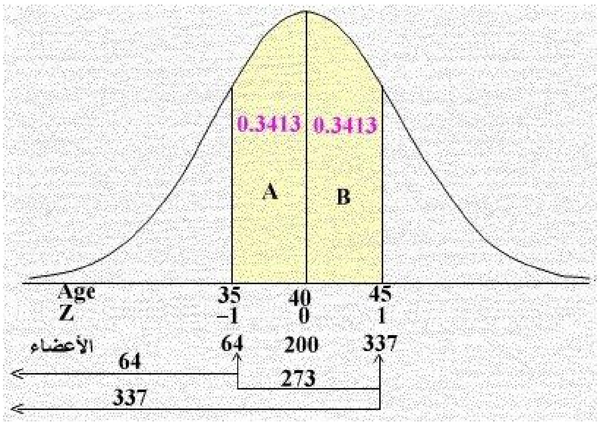
القيمة الجدولية المقابلة للعدد 1 = 0.8413 عدد الاعضاء = ٤٠٠ * ٠,٨٤١٣ = ٣٣٧

القيمة الجدولية المقابلة للعدد -1 = 0.1587 عدد الاعضاء = ٤٠٠ * ٠,١٥٨٧= ٦٤

المساحة الحصورة بين ١ و -١ = ٠,٨٤١٣ - ٠,١٥٨٧ = ٠,٦٨٢٦

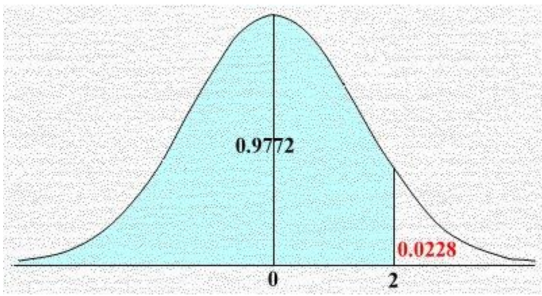
احتمالية عدد الاعضاء الذين اعمارهم بين ٣٥ الى ٤٥ سنة = ٠,٦٨٢٦

عدد الاعضاء الذين اعمارهم بين ٣٥ الى ٤٥ سنة = ٠,٦٨٢٦ * ٤٠٠ = ٢٧٣



2) نحسب قيمة Z من القانون للعمر ٥٠:

$$Z = (X - \mu) \div \sigma = (50 - 40) \div 5 = 2$$



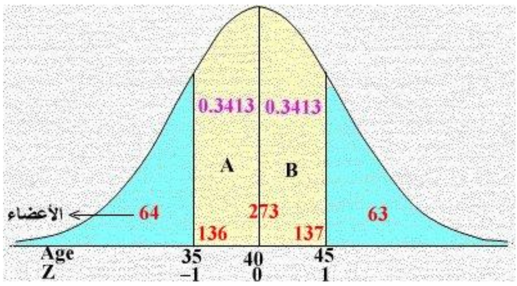
القيمة الجدولية المقابلة للعدد ٢ (المساحة) هي ٠,٩٧٧٢على يسار القيمة

فيكون عدد الذين تقل أعمارهم عن ٥٠ = ٠,٩٧٧٢ × ٤٠٠ ≈ ٣٨١ عضو

لاحظ:

الذين يزيد أعمارهم عن ٥٠ = (١ – ٠,٩٧٧٢) × ٤٠٠ = ٠,٠٢٢٨ × ٤٠٠ ≈ ٩

3) الأعضاء الذين أعمارهم أقل من ٣٥ واكبر من ٤٥ هم خارج الفترة العمرية للمطلوب ١



والمبينة بالشكل المقابل باللون الأزرق وهي تمثل ١ مطروحاً منه المساحة ٠,٦٨٢٦

أي : المساحة = ١– ٠,٦٨٢٦ = ٠,٣١٧٤

عدد الأعضاء = ٠,٣١٧٤ × ٤٠٠ ≈ 127

وتمثلهم المساحة المبينة باللون الأزرق . أنظر الشكل المقابل .

مثال : اختير طالب عشوائياً من مجتمع نسبة ذكاء أفراده تتبع توزيع طبيعي وبمتوسط حسابي ٨٠ وانحراف قياسي (معياري) ١٠ فأوجد:

(1)   احتمال أن تقل نسبة ذكاء الطالب المختار عن90

(2)   احتمال أن تزيد نسبة ذكاء الطالب المختار عن ١٠٥

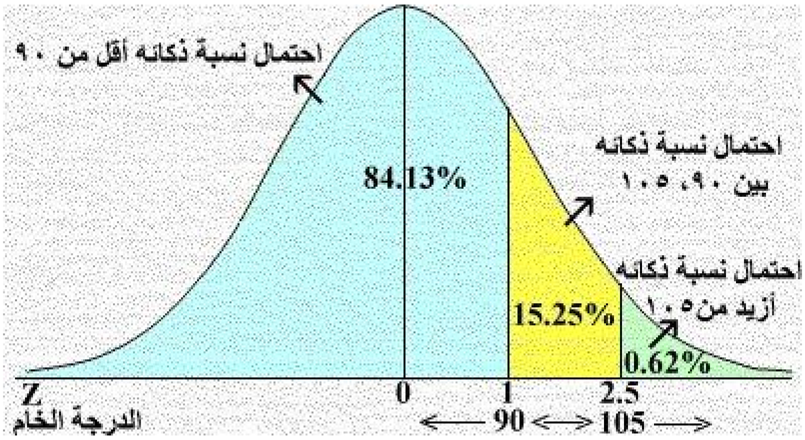
(3)   احتمال أن تتراوح نسبة ذكائه بين 90 ، 105

وضح ذلك بيانياً (المساحة تحت منحنى التوزيع الطبيعي).

الحل:

نحول قيم المتغير الطبيعي الى متغير طبيعي قياسي

نحسب القيمة القياسية (Z) التي تقابل القيمة ٩٠



$$z = \frac{x - \mu}{\sigma} = \frac{90 - 80}{10} = 1$$

من جدول Z نجد أن المساحة المقابلة (Z=1)   =   0.8413 وهو الاحتمال المطلوب

∴ احتمال أن تقل نسبة ذكاء الطالب المختار عن 90 = 0.8413 = ٨٤,١٣ %

2) نحسب القيمة القياسية (Z) التي تقابل القيمة ١٠٥

$$z = \frac{x - \mu}{\sigma} = \frac{105 - 80}{10} = 2.5$$

من جدول Z نجد أن المساحة المقابلة = 0.9938

وحيث المطلوب أن تزيد نسبة الذكاء فيكون الاحتمال المطلوب

=١ – ٠,٩٩٣٨ = ٠,٠٠٦٢ = ٠,٦٢ %

3) الاحتمال المطلوب = احتمال أقل من ١٠٥ مطروحاً منه احتمال أقل من ٩٠ أي:

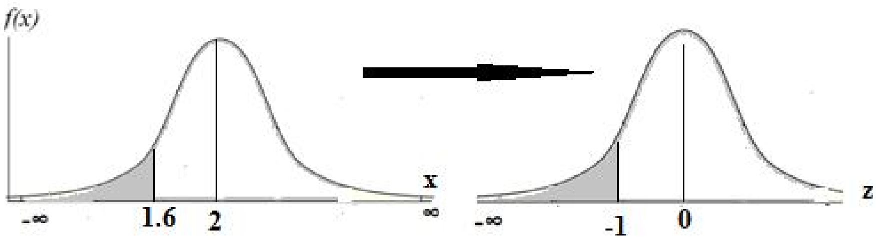
٠,٩٩٣٨ - 0.8413 = ٠,١٥٢٥ = ١٥,٢٥ %

مثال ( )

اذا علمت بان بطارية هاتف نقال مدة استهلاكها سنتان وانحرافها القياسي ٠,٤ سنة. فاذا علمت با مدة الاستهلاك تتبع التوزيع الطبيعي ، ما هو احتمال ان تستهلك بطارية معينة من نفس النوع باقل من ١,٦ سنة.

الحل: نحول المتغير الطبيعي الى قياسي وفق المعادلة:

$$z = \frac{x - \mu}{\sigma} = \frac{1.6 - 2}{0.4} = -1$$



ومن جدول Z نجد ان الاحتمالية تساوي ٠,١٥٨٧ ، اي ان احتمالية استهلاك البطارية باقل من ١,٦ سنة تساوي ١٥,٨٧ % .

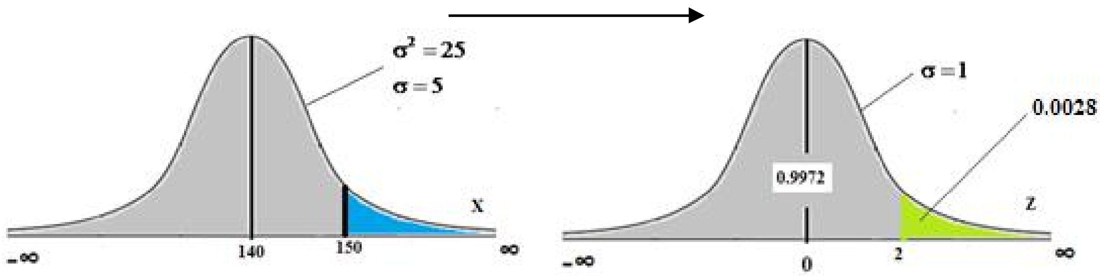
مثال:

لو فرضنا بان متوسط ارتفاع هجين الذرة الصفراء ١٤٠ سم وتباينه ٢٥ سم وانه متوزع توزيع طبيعي. فاذا اخترنا ١٠٠٠ نبات عشوائياً من هذا الهجين ، فما هو عدد النباتات المتوقع زيادة طولها عن ١٥٠ سم.

الحل : ان عدد النباتات المتوقع زيادة طولها عن ١٥٠ سم تساوي $p(x \ge 150)*(1000)$

نحول التباين σ^2 الى الانحراف القياسي σ $\quad \sigma=\sqrt{\sigma^2}=\sqrt{25}=5$

ونحول المتغير الطبيعي الى قياسي $\quad z=\frac{x-\mu}{\sigma}=\frac{150-140}{5}=2$



ومن جدول التوزيع الطبيعي نجد المساحة تحت المنحنى $p(z\ge 2)=1-p(z\le 2)$

$$p(z\ge 3)=1-0.9972$$
$$p(z\ge 3)=0.0028=0.28\%$$

∴ عدد النباتات التي يزيد ارتفاعها عن ١٥٥ سم = ١٠٠٠ * 0.0028 ≅ ٣

مثال: اذا كان معدل طول ١٠٠٠٠ نبات من القطن هو ٨٠ سم ، فاذا وجد بان ١٥٨٧ نبات يقل طولهم عن ٦٠ سم ، فما هي عدد النباتات التي يزيد طولها عن ٩٠ سم.

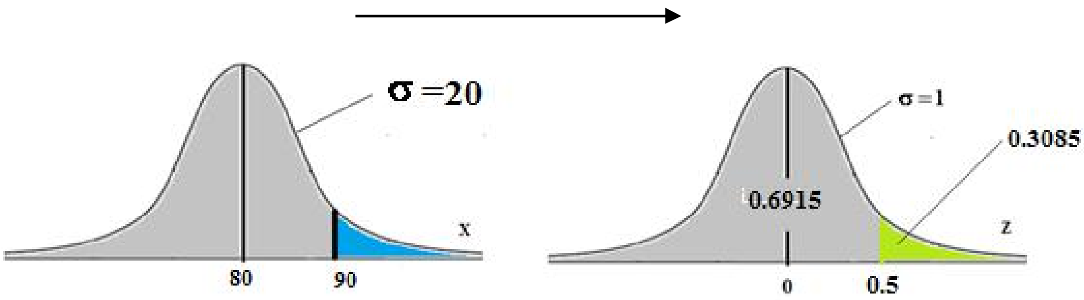
الحل: ∵ نسبة النباتات التي يقل طولها عن ٦٠ سم $=\frac{1587}{10000}=0.1587$ وهذه النسبة تعادل المساحة تحت المنحنى الطبيعي ، ومنها نجد قيمة Z التي تقابل هذه المساحة وتساوي- ١

$$z=\frac{x-\mu}{\sigma}$$

. وبعدها نستخرج قيمة σ

$$-1=\frac{60-80}{\sigma} \qquad \therefore \sigma=20$$

ولاستخراج عدد النباتات التي تزيد عن ٩٠ سم



$$\therefore p(x > 90) = 1 - p(x < 90)$$

$$\therefore p(x > 90) = 1 - p(z < \frac{x-\mu}{\sigma} = \frac{90-80}{20} = 0.5)$$

$$\therefore p(x > 90) = 1 - p(z < 0.5)$$

$$\therefore p(x > 90) = 1 - 0.6915$$

$$\therefore p(x > 90) = 0.3085$$

لذا فعدد النباتات التي يزيد طولها عن ٩٠ سم هو 10000 * 0.3085

اي ٣٠٨٥ نبات

الفصل التاسع

اختبار الفرضيات

Test of Hypothesis

اختبار الفرضيات الاحصائية

مقدمة

تعتبر اختبارات الفوضيات الإحصائية Testing Statistical Hypotheses واحدة من أهم التطبيقات التي قدمها علم الإحصاء كحل للمشاكل العلمية المختلفة بشتى فروع العلم. فباستخدام نظرية الاحتمالات وخصائص توزيعات العينة أمكن التعرف على ما يسمى باختبارات الفوضيات او الفروض الإحصائية ومن خلالها يمكن لأي شخص أن يتخذ قرار برفض أو قبول فرض معين أو مجموعة من الفروض المتعلقة بمشكلة معينة موجودة في الحياة العامة.

من أهداف علم الإحصاء هو الاستنتاج عن المجتمع وذلك استنادًا على بيانات العينة ، بما أن للمجتمع خصائص قياسية وصفية رقمية نسميها المعالم (Parameters) فإن الاستنتاج الإحصائي يختص بصنع الاستنتاج لمعالم المجتمع.

تعتبر اختبارات الفرضيات الإحصائية واحده من أهم التطبيقات التي قدمها علم الإحصاء كحل للمشاكل العلمية المختلفة بشتى فروع العلم ورغم أهميه موضوع تقدير المعالم الاحصائية إلا انه غالبا ما يكون الاهتمام مركز بشئ على مجرد تقدير المعالم ولكن على عمليه وضع قواعد تمكن من التوصل إلى قرار بقبول أو رفض خاصية أو بالمعنى الإحصائي فرض عن معالم مجتمع واحد أو أكثر ، وهذا ما يسمى باختبارات الفرضيات الإحصائية و تبدأ أي مشكله باهتمام الباحث بدراسة خصائص مجتمع ما ثم دراسة هذه الخصائص بعده طرق منها اختبار فرض (أو فروض) يتعلق بها كلما يتأكد الباحث من صحة أو عدم هذا الفرض فإنه لا يستطيع التعامل مع المجتمع ككل لان هذا إما مستحيلاً أو مكلفاً للغاية و يأخذ زمن كبير جدا ولذلك ليس أمام الباحث إلا اختيار عينه عشوائية من ذلك المجتمع .

يحاول الباحث اتخاذ قرار ما لمشكلة محددة بشأن خواص توزيع ما (المتوسط – التباين - النسبة) لعينة عشوائية تم سحبها من المجتمع "العينة المسحوبة لابد وأن تكون ممثلة تمثيلا جيدا للمجتمع محل الدراسة ". ولكي نصل إلى قرار إحصائي لابد من وضع فرضيات عن خواص المجتمع،

ومن هنا نختبر مدى صحة هذه الفرضيات من عدمها وذلك عن طريق العينة العشوائية التي تم سحبها من المجتمع. بفرض اننا مهتمين بتقدير الوسط الحسابى لمجتمع ما ونفرض ايضا اننا سحبنا جميع العينات الممكنه من ذلك المجتمع وحسبنا الوسط الحسابى لكل عينه وكنا نعلم مقدما قيمة الوسط الحسابى للمجتمع (فرضا) فاننا نلاحظ ان الوسط الحسابي لبعض العينات قد تتساوى او تقل او تزيد عن الوسط الحسابى للمجتمع المسحوب منه العينات. والفروق بين قيمة الوسط الحسابي المحسوب من العينات والمجتمع يسمى بخطأ المعاينه Sampling Error وهو متغير عشوائى يمكن التحكم فيه وحيث ان العينه عشوائيه فامكن اثبات ان خطأ المعاينه خطأ غير حقيقى.

وبذلك فان الوسط الحسابي لعينه واحده يصلح لان يكون تقديرا جيدا للوسط فى المجتمع وبالرغم من وجود بعض الفروق بين المعلمه والتقدير، فالاحصاءات التحليلية اعتبرت ان هذه الفروق فروقا غير حقيقية وترجع الى الصدفه وسميت بالفروق الغير معنويه Not Significant . فى نظرية التقديرات (Estimation Theory) قد تم اثبات انه اذا كان وسط المجتمع مجهول فان الوسط الحسابى لأى عينة هو تقدير جيد لوسط المجتمع بشرط ان تكون العينة المسحوبه عشوائية.

اختبارات الفوضيات ترتكز اساسا على هذه الفكره واشتقت اسمها منها حيث عن طريقها نستطيع ان نحدد وبسهوله هل الفروق بين المعلومات المحسوبة من العينة وبين المعلومات المفروضه لمجتمع معين فرقا يرجع الى الصدفه ام فرق حقيقى، وبأسلوب أخر هل هو فرق معنوى او فرق غير معنوى؟ وبذلك سميت هذه الاختبارات باسم اختبارات المعنويه Test of Significant.

الاختبارات الاحصائيه قد تدور حول معالم المجتمع المجهوله مثل الفوضيات المتعلقة بالوسط الحسابى، النسبه، التباين، معامل الارتباط،....الخ. وفى هذه الحاله يطلق على هذه الاختبارات اسم الاختبارات المعلميه Parametric Tests. وقد تكون فوضيات لا تتعلق بمعالم المجتمع ولكن تتعلق

بأشياء اخرى قد تكون وصفية مثل العلاقة بين التعليم والتدخين، خضوع نتائج معينه لنظريه معينه، العلاقة بين لون العينين ولون الشعر،.... وفى هذه الحالة يسمى الاختبار باسم الاختبار اللامعلمى Non Parametric Test .

مصادر الفرضيات

١. الدراسة الميدانية
٢. الدراسات السابقة
٣. الملاحظات الشخصية و الخبرة العملية
٤. التجارب المختبرية

الشروط التي يجب توافرها في الفرضيات

١. صياغة الفرضيات بعبارات سهلة و بسيطة و واضحة
٢. أن تكون جزءاً من خطة متكاملة للبحث العلمي
٣. إمكانية اختبارها و التأكد من صدقها و ثباتها
٤. أن لا تتعارض مع الحقائق العلمية
٥. أن يكون لها قدرة تفسيرية
٦. أن يكون لها نتيجة واحدة واضحة و محددة

الاخطاء التي ترتكب عند اختبار الفرضيات

الخطأ من النوع الأول Type I error : هو القرار برفض فرضية العدم عندما تكون صحيحة و يرمز له بالرمز **α** .

الخطأ من انوع الثاني Type II error : هو القرار بقبول فرضية العدم عندما تكون خاطئة فنرمز له بالرمز **β** .

هذان الخطآن موضحان في الجدول التالي :

الحالة الحقيقية الـــــقـرار	H0 صحيحة	H0 خاطئة

قبول H0	قرار صائب	الخطأ من النوع الثاني
رفض H0	الخطأ من النوع الأول	قرار صائب (قوة الاختبار)

مصطلحات ومفاهيم عامة

تعريف الفرضية الإحصائية Statisticel Hypothesis:

انواع الفرضيات

١. فرضية العدم (او الفرضية الصفرية) Null Hypothesis (H_0)

هي الفرضية التي يضعها الباحث على امل ان يرفضها.

هي الفرضية حول معلمة المجتمع التي نجري اختبار عليها باستخدام بيانات من عينة والتي تشير أن الفرق بين معلمة المجتمع والإحصائي من العينة ناتج عن الصدفة ولا فرق حقيقي بينهما. وهي الفرضية التي ننطلق منها ونرفضها عندما تتوفر دلائل على عدم صحتها، وخلاف ذلك نقبلها وتعني كلمة Nul انه لا يوجد فرق بين معلمة المجتمع والقيمة المدعاة (إحصائية العينة).

٢. الفرضية البديلة (H_A) Alternative Hypothesis

هي الفرضية التي يريد الباحث أن يتوصل اليه بما جمع من البيانات .

هي الفرضية التي يضعها الباحث كبديل عن فرضية العدم و نقبلها عندما نرفض فرضية العدم باعتبارها ليست صحيحة بناء على المعلومات المستقاة من العينة.

صياغة الفرضية الاحصائية

دائماً نصيغ الفرضية الاحصائية بصورة معاكسة للحالة التي نريد اختبارها ...فمثلاً في حالة مقارنة انتاجية صنفين من القطن وكان هناك ادعاء ان متوسط انتاج الصنف الثاني اكبر من متوسط انتاج لصنف الاول ويراد اجراء اختبار احصائي لهذا الادعاء فاننا نفترض حسن النية ونبدأ بوضع الفرضية الاحصائية الآتية :

مثلاً " نفترض عدم وجود اختلاف بين متوسطي الانتاج للصنفين من القطن " وتسمى هذه الفرضية بفرضية العدم او الفرضية الصفرية ويرمز لها H_0 ثم نجري الاختبار اما قبول H_0 او رفضها ، فاذا كان القرار قبول H_0 كان معنى ذلك انه لا يوجد اختلاف بين متوسطي الانتاج للصنفين من القطن وان الاختلاف الموجود لدينا هو اختلاف ظاهري نتيجة للصدفة وحدها ودائماً في الواقع يقابل فرضية العدم فرضية معاكسة لها تسمى الفرضية البديلة وترمز لها H_A ، فاذا كانت فرضية العدم هي عدم وجود اختلاف تكون الفرضية البديلة ان متوسط انتاجية الصنف الثاني اكبر من متوسط انتاجية الصنف الثاني وتصاغ الفرضيتين كالآتي : $H_0 : \mu1 = \mu2$ $H_A: \mu2 > \mu1$. وعند قبول فرضية العدم يكون معناه انه لا يوجد لدينا من المبررات ما يكفي لرفض فرضية العدم وفي هذه الحالة لا يكون في وسعنا الا قبولها. وعند رفض فرضية العدم يكون معناه انه لا يوجد لدينا من المبررات ما يكفي لرفض الفرضية البديلة وفي هذه الحالة لا يكون في وسعنا الا قبولها.

امثلة اخرى على صياغة الفرضيات:

2) أن متوسط مقياس الممارسات العنيفة أقل لدى تلاميذ الصف السادس الإبتدائي عن اللذين لا يمارسون ألعاب الكمبيوتر العنيفة μ_1 بالمقارنة مع تلاميذ الصف السادس الإبتدائي الذين يمارسون هذه الألعاب μ_2. $H_0 : \mu1 = \mu2$ $H_A: \mu1 < \mu2$

3) يوجد فرق بين معلمي ومعلمات مدارس التعليم العام في الدولة في متوسط مقياس الضغوط النفسية . $H_A: \mu1 \neq \mu2$ $H_0 : \mu1 = \mu2$

اختبار النوع الأول من الفرضيات في المثالين (١) ،(٢) اختبار من طرف واحد One - sided test واختبار النوع الثاني في المثال اختبار ذو طرفين Two - sided test

مستوى المعنوية (α) Significance Level :

احتمال رفض فرضية العدم H_0 بينما هي في الواقع صحيحة .(احتمال الوقوع في الخطأ من النوع الاول).

وهو يمثل احتمال الوقوع في الخطأ من النوع الأول ، و هو يمثل مستوى عدم الثقة في التقدير الذي نحصل عليه ، أو يمثل احتمالية أن نكون مخطئين عند رفضنا لفرضية العدم و قبولنا للفرضية البديلة ، و تستخدم أغلب الدراسات البايولوجية والزراعية مستوى الدلالة ٠,٠٥ او ٠,٠١.

ومن الملاحظات المهمة هنا هو أن " مستوى المعنوية " والذي يسمى أحياناً " مستوى الدلالة " هو المكمل لدرجة الثقة " بمعنى أن مجموعها يساوي 100% أو واحد صحيح. فإذا كانت درجة الثقة 95% فإن مستوى المعنوية يساوي 5%. والعكس صحيح فإذا كان مستوى المعنوية 5% فإن هذا يعني أن درجة الثقة 95 %. ولعل من أهم الملاحظات هنا هو استخدام تعبير "مستوى المعنوية" في حالات اختبارات الفروض، بينما يستخدم مصطلح "درجة أو مستوى الثقة" في حالات التقدير.

والفكرة الأساسية في اختبار الفرض هي تقسيم المساحة تحت المنحنى إلى منطقتين: أحداهما تسمى " منطقة القبول " أي منطقة قبول فرضية العدم . والأخرى تسمى " منطقة الرفض"، أي منطقة رفض فرضية العدم والتي تسمى أحيانا " بالمنطقة الحرجة Critical region ". والنقطة الجديرة بالملاحظة هنا هي أن منطقة القبول تمثل درجة الثقة، بينما تمثل منطقة الرفض مستوى المعنوية.

- يمكن توضيح مستوى المعنوية من خلال المثال التالي:

- شركة تنتج نوع معين من السيارات ادعت أن سياراتها توفر كثيرا في كمية الوقود وان متوسط المسافة التي تقطعها السيارة حوالي ٦٠ ميل لكل جالون بنزين بانحراف معياري ١٠ميل لكل جالون. وقام الباحثون باختبار جميع السيارات المنتجة حيث وجدوا أن متوسط المسافة بلغ ٦١ ميل/جالون هل هذا الفرق معنوي (حقيقي) مقارنة بالمتوسط الذي حددته الشركة ؟

- الإجابة نعم بالطبع.

- لماذا: لان الفرق مبني على الإحصاء الشامل لجميع السيارات ولا يوجد أخطاء معاينة يمكن أن يعزى لها هذا الفرق، وهذا يشير بشكل واضح أن متوسط المجتمع تغير من ٦٠ ميل إلى ٦١ ميل.

المختبر الاحصائى Test Statistic :

هو متغير عشوائي تحسب قيمته من العينة و تستخدم في اتخاذ القرار برفض أو عدم رفض فرضية العدم .

وهو نوع الاختبار الاحصائي التي يتم حسابه من بيانات العينة بافتراض أن فرضية العدم صحيحة.

ويتوقف نوع المختبر الاحصائي على العوامل التالية:

أ- توزيع المجتمع، وهل هو طبيعي أم لا، وهل تباينه معروف أم لا.

ب- وحجم العينة، وهل هو كبير أم صغير.

جـ- وفرضية العدم المراد اختبارها وهل هي عن الوسط أو النسبة أو التباين أو الارتباط... الخ.

والفكرة الأساسية (غالباً) في المختبر الاحصائي هي: حساب الفرق بين قيمة المعلمة التي نفترضها للمجتمع (في فرضية العدم) والقيمة المقابلة لها في العينة أي التابع الإحصائي، ثم نقسم (أو ننسب) هذا الفرق إلى الخطأ القياسي للتابع الإحصائي. فمثلاً: إذا كان الاختبار عن الوسط الحسابي فإنه يتم حساب الفرق بين قيمة الوسط الحسابي للمجتمع التي نفترضها وقيمة الوسط الحسابي للعينة، ثم نقسم هذا الفرق على الخطأ القياسي للوسط. وهكذا مع باقي الإحصائيات. فلو أراد الباحث اختبار

فرضية أن متوسط عمر الناخب في دولة ما هو مثلاً 30 سنة ولاختبار مدى صحة هذه الفرضية فإنه عادة ما تسحب عينة عشوائية من المجتمع، ولنفرض أن متوسط عمر الناخب في هذه العينة كان 31 سنة، فالفرق هنا هو سنة واحدة وهو فرق صغير بين الافتراض والعينة الحقيقية فالباحث عادة ما يميل إلى قبول فرضيه العدم.

أما إذا كان متوسط عمر الناخب في العينة مثلاً هو 45 سنة، فالفرق هنا كبير بين الفرض والعينة، ولذا فإن احتمال رفض فرضيه العدم هي احتمال كبير نظراً لكبر الفرق بين قيمة الفرض وقيمة العينة. من هنا نستطيع القول بأن المختبر الاحصائي يعتمد على حساب الفرق بين قيمة الوسط المفترض وقيمة متوسط العينة.

هنا قد يثار تساؤل عن المعيار الذي يستطيع من خلاله الباحث الحكم على هذا الفرق ومدى كبره أو صغره. والإجابة الإحصائية عليه تتم من خلال قسمة هذا الفرق على الخطأ القياسي للوسط، ثم مقارنة خارج القسمة بالقيمة الجدولية أو ما يسمى بحدود منطقتي القبول والرفض كما سوف نرى لاحقاً.

وفيما يأتي نموذج لبعض نماذج المختبرات الاحصائية واستخداماتها

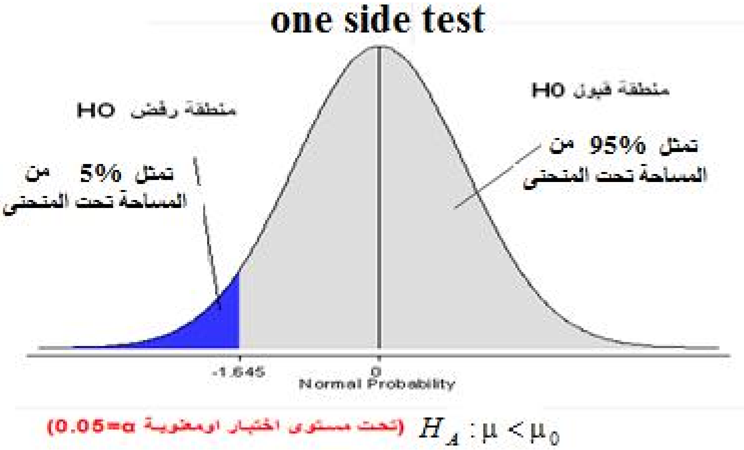
ت	الاختبار	العينات	المختبر الاحصائي	توزيع المختبر
١	متوسط مجتمع غير محدود ومعلوم التباين	كبيره	$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$	N(0,1)
٢	متوسط مجتمع محدود ومعلوم التباين وحجمهn	كبيرة	$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}}$	N(0,1)
٣	متوسط مجتمع طبيعي مجهول التباين	صغيرة n<30	$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$	$t_{df} = n-1$
٤	تباين مجتمع طبيعي وانحرافه المعياري معلوم	صغيرة	$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$	$\chi^2_{df} = n-1$
٥	النسبة بين تباين مجتمعين طبيعيين	صغيرة	$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$	$F_{(n_1-1\ ,\ n_2-1)}$
٥	الفرق بين متوسطي مجتمعين معلومين التباين	صغيره أو كبيره	$Z = \frac{(\overline{x_1} - \overline{x_2}) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$	N(0,1)

منطقة الرفض أو المنطقة الحرجة Critical region :

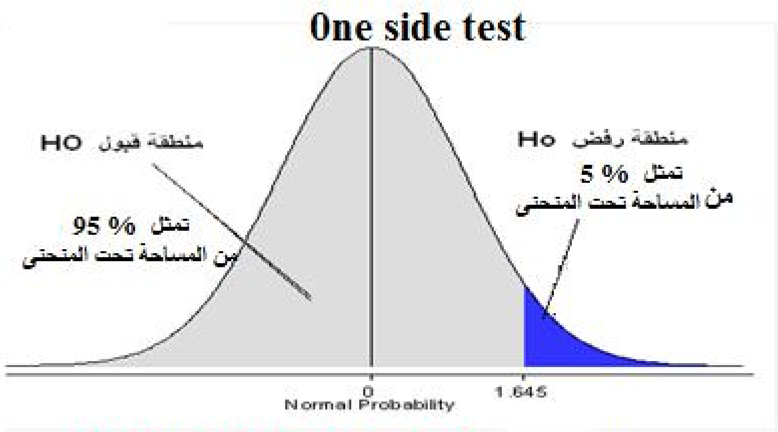
> هي المنطقة التي اذا وقع فيها قيم المختبر الاحصائي تؤدي إلى رفض فرضية العدم.

اذا وقعت قيمة المختبر الاحصائي والمحسوبة من بيانات العينة في منطقة قبول H_0 فاننا نقبل فرضية العدم بدرجة الثقة المحددة وعلى ذلك نرفض الفرضية البديلة ، اما اذا وقعت قيمة المختبر الاحصائي في منطقة الرفض كان معنى ذلك اننا رفض فرضية العدم وبمنى آخر لدينا المبرر الكافي لرفض فرضية العدم وفي هذه الحالة نقبل الفرضية البديلة H_A. . وهناك ثلاث حالات مختلفة لمنطقتي القبول والرفض هي :

١. إذا كانت الفرضية البديلة تأخذ شكل " أقل من " $H_A: \mu < \mu_0$ فإن منطقة الرفض تكون مركزة بالكامل في الطرف الأيسر للمنحنى. ويسمى الاختبار في هذه الحالة اختبار الطرف الأيسر. والشكل التالي يوضح ذلك :

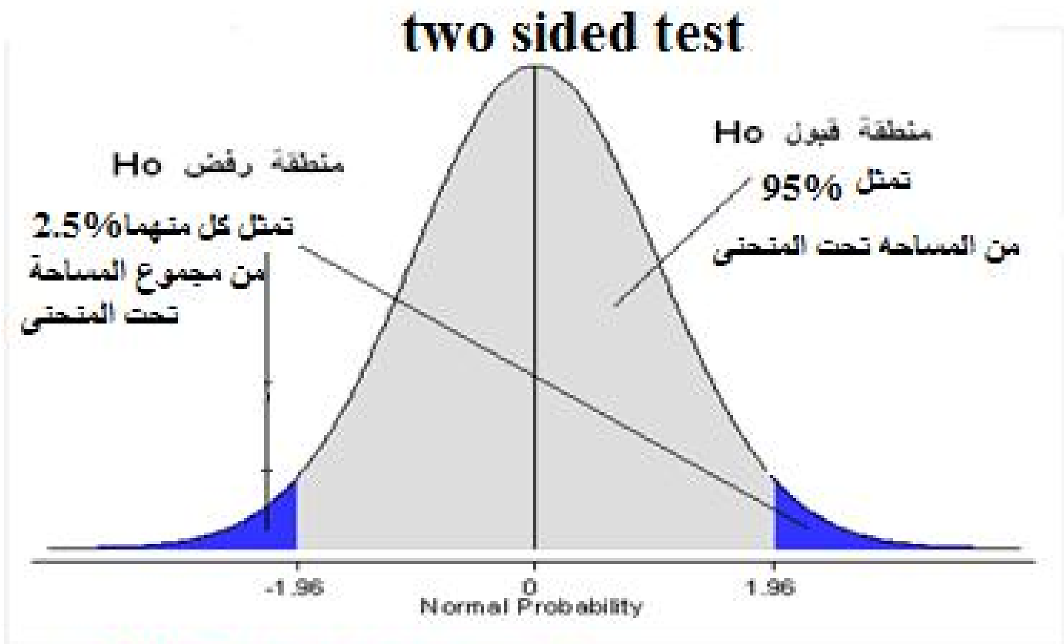


٢. إذا كانت الفرضية البديلة تأخذ شكل "أكبر من" $H_A: \mu > \mu_0$ فإن منطقة الرفض تكون مركزة بالكامل في الطرف الأيمن للمنحنى. ويسمى الاختبار في هذه الحالة اختبار الطرف الأيمن. والذي يأخذ الشكل التالي أدناه :



$H_A : \mu > \mu_0$ (تحت مستوى اختبار او معنوية α=0.05)

٣. إذا كانت الفرضية البديلة تأخذ شكل " لا يساوي " $H_A: \mu \neq \mu_0$ كأن تكون الفرضية في هذه الحالة هي أن متوسط دخل الفرد لا يساوي 200 دولاراً فإن منطقة الرفض تكون موزعة على طرفي المنحنى بالتساوي، ويسمى الاختبار في هذه الحالة " اختبار الطرفين "، والذي يأخذ الشكل التالي (بافتراض أن $\alpha = 5\%$) :



$H_A : \mu \neq \mu_0$ (تحت مستوى اختبار او معنوية α=0.05)

الخطوات الأساسية فى اختبارات الفرضيات

تعتبر اختبارات الفرضيات الإحصائية أسلوباً لتحديد ما إذا كانت البيانات التي تتضمنها العينة كافية لرفض فرضية العدم H_0 و يتضمن هذا الأسلوب أساساً الخطوات التالية :

١.صياغة الفرضيات فرضية العدم و الفرضية البديلة H_0 و H_A و تحديد الفرضيات المتعلقة بتوزيع المجتمع

٢.تحديد مستوى المعنوية للاختبار α .

٣.صياغة قاعدة القرار بتحديد منطقة الرفض بناءً على مستوى المعنوية و اتجاه الفرضية البديلة .

٤.تحديد إلمختبر الاحصائي المستخدم و تحديد توزيع المعاينة لهذا الإحصاء في ضوء الافتراضات المشار إليها في الخطوة (١) . حساب قيمة إلمختبر الاحصائي باستخدام بيانات العينة .

٥. اتخاذ القرار برفض أو قبول فرضية العدم H_0 بناءً على قيمة إلمختبر الاحصائي و منطقة الرفض و تفسير نتيجة الاختبار في ضوء الهدف من البحث .

بعض الملاحظات

١. الخطوات الثلاثه الاولى والخطوه الخامسة تحدد بمعرفة الباحث نفسه ولا تحتاج لمعلومات احصائيه

٢. الاختبارات التى سنتعامل معها فى هذا الفصل هى اختبارات معلميه تتعلق بعينة واحده وايضا عينتين.

٣. توجد طريقتين لاتخاذ قرار فى الاختبارات الاحصائية

(i) حساب المختبر الاحصائي ومقارنته بقيمة جدوليه وتحدد القيمة الجدوليه بناء على نوع الاختبار ذو طرف واحد One Tail Test أو ذو طرفين Two Tail Test

(ii) حساب ما يسمى بالقيمة الاحتماليه P-value ويرمز لها فى بالرمز .Sig فاذا كان الاختبار ذو طرف واحد تقارن .Sig بالقيمة α لكن اذا كان الاختبار ذو طرفين تقارن بالقيمة $\alpha/2$