جامعة بغداد

كلية التربية للعلوم الصرفة / ابن الهيثم

منهج البحث العلمي

قسم الرياضيات

مدرس المادة

م.م . اريج صلاح محمد

الفصل الاول: العلم والبحث العلمي

المقدمة:

مر الإنسان بمراحل عدة، وعلى مدى طويل من الزمان، حتى وصل إلى ما هو عليه اليوم من التطور في أساليب التفكير، والحصول على المعرفة. فالإنسان منذ نشأته أحاطت به المشكلات بشتى أنواعها، وتطلب منه مواجهتها وإيجاد الحلول المناسبة لها بإمكاناته المحدودة. اذ بدأ يتأمل بما حوله، ويتساءل عن أسباب ما يحدث له، وما يحدث من حوله في بيئته التي يعيش فيها، وكثيراً ما كان يواجه الصعوبات للإجابة عن تساؤلاته، وإيجاد الحلول لها. لكنه استمر في ذلك عن طريق المحاولة والخطأ، فضلاً عن (ملاحظته للحيوانات، فيصنع كما تصنع في بعض المواقف الحياتية) ولكن كانت أكثر إجاباته، وحلوله قاصرة لقلة خبراته ومعارفه، وضعف إمكاناته، ومع الوقت صار يكتسب المعرفة، والخبرة الشخصية، وتحولت لتصبح معارفه وخبراته أعرافاً وتقاليد، وتطورت لمراحل أكثر تقدماً من التفكير والتأمل إلى التفكير الاستنباطي، والاستقرائي، ثم كان اكتشافه واستعماله للمنهج العلمي في التفكير والبحث، باستعماله أساليب الملاحظة العلمية الدقيقة للوقائع، وفرض الفرضيات، وإجراء التجارب للوصول إلى الحقائق.

فالبحث هو السبيل الأمثل للتوصل للحقيقة، اذ ليس هناك علم أو تقدم علمي إلا عن طريق البحث العلمي، كما أن تقدم البحث العلمي يعتمد على المنهج العلمي. **فما المقصود بكل منها؟**

العلم واهدافه

العلم : نشاط يهدف إلى زيادة قدرة الانسان على السيطرة على الطبيعة من خلال فهم الظواهر وايجاد العلاقات التي تحكم هذه الظواهر .

ولما كانت ظواهر الكون عديدة فان العلاقات بينها عديدة ومتشابكة، لذلك لجأ العلماء الى تقسيمها في مجموعات لتسهيل دراستها فنشأت العلوم المختلفة ، فالظواهر الخاصة بالفلك كانت موضوعاً خاصاً لعلم الفلك ، والظواهر الخاصة بالسلوك الانساني كانت موضوعاً لعلم النفس وهكذا ....

اما أهداف العلم فتتمثل بالآتي:

١) **الفهم** : وهو الغرض الأساسي للعلم، فالعلم كنشاط انساني يهدف إلى فهم الظواهر المختلفة وتفسيرها، اي فهم الأسباب والعوامل التي أدت إلى حدوث الظاهرة وليس الاكتفاء بتعداد صفاتها وخصائصها بل التعرف على علاقة الظاهرة بالظواهر الأخرى التي أدت إلى وقوعها.

٢) التنبؤ: وهي قدرة الباحث على ان يستنتج من فهمه للظاهرة وقوانينها نتائج أخرى مرتبطة بهذا الفهم. فالتنبؤ هو تصور النتائج التي يمكن أن تحدث إذا طبقنا القوانين التي اكتشفناها على مواقف جديدة . وتزداد القدرة على التنبؤ بزيادة درجة التشابه بين الظاهرة التي درسناها وبين الظواهر التي سنطبق عليها فهمنا للظاهرة الأولى ، وأي استنتاج لا يعد صحيحاً إلاّ إذا تم اثباته تجريبياً .

٣) الضبط والتحكم : يساعد العلم والبحث العلمي في عملية الضبط والتحكم في الظواهر والأحداث والوقائع والامور والسيطرة عليها وتوجيهها التوجيه المطلوب واستغلال النتائج لخدمة الانسانية، وبذلك تمكن الانسان بفضل العلم من التحكم والضبط مثلاً في مسار الأنهار الكبرى ومياه البحار والمحيطات والتحكم بالجاذبية الأرضية واستغلال ذلك لخدمة البشرية، كما أصبح اليوم بفضل العلم امكانية التحكم بالأمراض والسلوكيات البشرية وضبطها وتوجيهها نحو الخير .

مثال: اذ فهم التربويون ظاهرة الذكاء عند الانسان فإنهم يستطيعون التنبؤ بعلاقة الذكاء بالتحصيل الدراسي، كما يستطيعون التحكم بهذه الظاهرة عن طريق تنظيم دراسات خاصة للطلبة الذين ترتفع نسبة ذكائهم ودراسات اخرى للطلبة الذين تقل نسبة ذكائهم كثيراً.

والمخطط الآتي يبين ذلك بوضوح.

مفهوم البحث العلمي

يتكون مصطلح البحث العلمي من مقطعين الأول "البحث" وهو كلمة مشتقة من مصدر الفعل الماضي بحث ومعناه: حاول، تتبع، سعى، تحرى ...الخ، والمقطع الثاني "العلمي" وهو كلمة مشتقة من كلمة العلم ومعناه: الحقيقة، المعرفة، التجريب ... الخ.

فهو وسيلة للاستعلام والاستقصاء المنظم والدقيق الذي يقوم به الباحث بهدف الوصول الى حلول لمشكلات معينة، واكتشاف معلومات أو علاقات جديدة، فضلاً عن تطوير أو تصحيح المعلومات الموجودة فعلاً، على ان يتبع في هذا الفحص والاستعلام خطوات المنهج العلمي .

ويمكن **تعريف البحث العلمي بأنه:** عرض مفصل أو دراسة متعمقة تمثل كشفاً لحقيقة جديدة، أو التأكيد على حقيقة قديمة سبق بحثها، وإضافة شيء جديد لها، أو حل لمشكلة كان قد تعهد بها شخص باحث بتقصيها وكشفها وحلها.

وكذلك يوجد تعريف آخر مفاده بأن: البحث العلمي هو نشاط علمي منظم، وطريقة في التفكير واستقصاء دقيق يهدف إلى اكتشاف الحقائق معتمداً على مناهج موضوعية من أجل معرفة الترابط بين هذه الحقائق واستخلاص المبادئ العامة والقوانين التفسيرية.

مفهوم المنهج العلمي

يتمثل **المنهج العلمي** بالدراسة الفكرية الواعية للمناهج المختلفة التي تطبق في مختلف العلوم وفقاً لاختلاف موضوعات هذه العلوم.

فهو خطوات منظمة يتبعها الباحث او الدارس في معالجة الموضوعات التي يقوم بدراستها الى ان يصل الى نتيجة معينة، اي هو الطريق المؤدي الى الكشف عن الحقيقة في العلوم المختلفة عن طريق اتباع جملة من القواعد العامة التي تسيطر على سير العقل.

يلعب منهج البحث دوراً أساسيا في تكوين معلومات البحث، فهو يلزم الباحث على عدم إبداء رأيه الشخصي دون تعزيزه بآراء لها قيمتها، والتقيد بإخضاع أي رأي للنقاش مهما كانت درجة الثقة به، إذ لا توجد حقيقة راهنة بذاتها، وضرورة تقيد الباحث بالدقة في الاعتماد على الروايات والاقتباسات أو التواريخ غير الواضحة أو غير الدقيقة، وكذلك ضرورة الدقة في شرح المدلولات التي يسوقها الباحث.

أهمية البحث العلمي

ان الحاجة الى الدراسات والابحاث العلمية اضحت اليوم مهمة اكثر من اي وقت مضى، اذ نجد الأمم في سباق دائم للوصول الى اكبر قدر ممكن من المعرفة الدقيقة المستمدة من العلوم التي تكفل الرفاهية للإنسان وتضمن له التفوق على غيره ، واذا كانت الدول المتقدمة تولي اهتماماً كبيراً للبحث العلمي فذلك يرجع الى انها ادركت ان عظمة الامم تكمن في قدرات ابنائها العلمية والفكرية والسلوكية .

والبحث العلمي ميدان خصب ودعامة اساسية لاقتصاد الدول وتطورها وبالتالي يحقق الرفاهية لشعوبها والمحافظة على مكانتها بين الدول. واصبحت منهجية البحث العلمي وأساليب القيام بها من المسلمات في المؤسسات الاكاديمية ومراكز البحوث فضلاً عن انتشار استعمالها في معالجة المشكلات التي تواجه المجتمع بنحو عام.

ويمكن تلخيص أهمية البحث العلمي بالنقاط الاتية:

١) يضمن البحث العلمي استمرارية التقدم والتطور.

٢) يفيد في التغلب على الصعوبات الحياتية والتنموية، والانتفاع بفوائده التطبيقية.

٣) يسهم في احياء الموضوعات القديمة عن طريق التحقيق فيها وتطويرها وصولاً لاكتشافات جديدة.

اما أهمية البحث العلمي بالنسبة للطلبة تتضح في النقاط الاتية :

١) أثراء معلومات الطالب في مواضيع معينة.

٢) تعود الطلبة على الاعتماد على النفس في دراسة المشكلات واصدار الاحكام بشأنها.

٣) أتباع الاساليب والقواعد العلمية المعتمدة في كتابة البحوث.

٤) التعود على استعمال الوثائق والكتب ومصادر المعلومات الاخرى والربط بينها للوصول الى نتائج جديدة.

٥) تدريب الطلبة على معالجة الموضوعات بنزاهة وموضوعية ونظام في العمل.

خصائص البحث العلمي

يتميز البحث العلمي بعدة خصائص أساسية أهمها:

(١) الموضوعية: وتعني الموضوعية هنا، أن الباحث يلتزم في بحثه المقاييس العلمية الدقيقة، ويقوم بإدراج الحقائق والوقائع التي تدعم وجهة نظره، وكذلك الحقائق التي تتضارب مع منطلقاته وتصوراته، فالنتيجة يجب أن تكون منطقية ومنسجمة مع الواقع ولا تناقضه.

(٢) استخدام الطريقة الصحيحة والهادفة: ويقصد بذلك، أن الباحث عندما يقوم بدراسة مشكلة أو موضوع معين، ويبحث عن حل لها، يجب أن يستعمل طريقة علمية صحيحة وهادفة للتوصل إلى النتائج المطلوبة.

(٣) الاعتماد على القواعد العلمية: يتعين على الباحث الالتزام بتبني الأسلوب العلمي في البحث عن طريق احترام جميع القواعد العلمية المطلوبة لدراسة كل موضوع، حيث إن تجاهل أو إغفال أي عنصر من عناصر البحث العلمي، يقود إلى نتائج خاطئة أو مخالفة للواقع.

(٤) الانفتاح الفكري: ويقصد بذلك، انه يتعين على الباحث الحرص على التمسك بالروح العلمية والتطلع دائما إلى معرفة الحقيقة فقط، والابتعاد قدر الإمكان عن التزمت والتشبث بالرؤية الأحادية المتعلقة

بالنتائج التي توصل إليها من خلال دراسته للمشكلة، ويجب أن يكون ذهن الباحث منفتحاً على كل تغير في النتائج المحصول عليها والاعتراف بالحقيقة وإن كانت لا تخلو من مرارة .

(٥) **الابتعاد عن إصدار الأحكام النهائية:** لا شك أن من أهم خصائص البحث العلمي التي ينبغي على الباحث التقيد بها، هي ضرورة التأني وعدم إصدار الأحكام النهائية، إذ يجب أن تصدر الأحكام استنادا إلى البراهين والحجج والحقائق.

صفات البحث الجيد

١) **العنوان الواضح والشامل للبحث:** إن الاختيار المناسب لعنوان البحث أو الرسالة أمر ضروري للتعريف بالبحث منذ الوهلة الأولى لقراءته من قبل الآخرين.

٢) **تركيز البحث على مجال معين:** اذ لا يجوز البحث في اكثر من مجال من مجالات الحياة في آن واحد مثال: لا يصح البحث في النثر و الشعر في آن واحد.

٣) **تخطيط حدود البحث:** ضرورة صياغة موضوع البحث ضمن حدود موضوعية وزمنية ومكانية واضحة المعالم، وتجنب التخبط والمتاهة في أمور لا تخص موضوع البحث، لأن الخوض في العموميات غير محددة المعالم والأهداف تبعد الباحث عن البحث بعمق بموضوع بحثه المنصوص عليه في العنوان.

٤) **الإسناد:** ضرورة اعتماد الباحث في كتابة بحثه على الدراسات السابقة والآراء الأصلية المسندة، وأن يكون دقيقا في سرد النصوص وإرجاعها لكاتبها الأصلي، والاطلاع على الآراء والأفكار المختلفة المتوفرة في مجال البحث لاسيما المذكورة في المصادر الحديثة. (فالأمانة العلمية بالاقتباس ونقلها أمر في غاية الأهمية في كتابة البحوث).

٥) **وضوح الأسلوب:** يجب أن يكون البحث الجيد مكتوب بأسلوب واضح، ومقروء، ومشوق، مع مراعاة السلامة اللغوية، وان تكون المصطلحات المستعملة موحدة في متن البحث.

٦) **الترابط بين أجزاء البحث:** ضرورة ترابط أقسام البحث وأجزاءه المختلفة وانسجامها، كما يجب أن يكون هناك ترابط تسلسل منطقي، وتاريخي أو موضوعي، يربط الفصول ما بينها، ويكون هناك أيضا ترابط وتسلسل في المعلومات ما بين الفصول.

٧) **الإسهام والإضافة إلى المعرفة في مجال تخصص الباحث:** الباحث الجيد هو الذي يبدأ من حيث أنتهي الآخرون بغرض مواصلة المسيرة البحثية وإضافة معلومات جديدة في المجال نفسه.

٨) **أن لا يكون البحث محكومًا عليه مسبقًا:** بمعنى ألا يقدر الباحث النتيجة و أن لا يبحث في شيء معروف مثال: كأن يبحث في حكم الصلاة .

صفات الباحث الجيد

قبل الولوج في صفات الباحث الجيد ينبغي **تعريف الباحث** :هو من يستقصي المعرفة من اجل اكتشاف حقائق يفيد بها نفسه ويفيد بها غيره. **وينبغي أن تتوافر في الباحث العديد من الصفات أهمها:**

١) **الرغبة في العمل:** وذلك بأن تتوفر لدى الباحث الرغبة الأكيدة في القيام بعمل البحث والمواصلة في الموضوع الذي اختاره ، وهي من أهم عوامل نجاح البحث، اذ أن الرغبة تعطي الباحث حافزًا على العمل وتعطيه الطاقة من أجل مواصلة العمل.

٢) **الصبر و الجلد:** وهو يعني (المتابعة و المثابرة، كثرة القراءة وعدم التذمر ، مواجهة الصعاب). لأن الاستقصاء والتفتيش ليس بالأمر السهل لذلك يحتاج الى صبر، والصبر يوصل الباحث إلى أكبر قدر ممكن من المعلومات.

٣) **التتبع وحب الاطلاع وبذل الجهد و القراءة بفهم وعمق:** وهذا يعني أن يلم الباحث بكل ما كتب عن موضوعه من بحوث ودراسات و أراء ، سواء أكانت مدونة في كتب أم مخطوطات أو مجلات، ومهما تكن تلك الآراء بسيطة أو مخالفة فإن سعة الاطلاع دليل على استقصاء المادة والسيطرة عليها ، ومن العيب أن يواجه باحث يبحث في موضوعه أو كتابًا منشورًا أو رأي لم يطلع عليه.

٤) **الدقة في فهم النصوص وآراء الغير:** وذلك بنقل العبارات بدقة وفهم، وعدم التسرع في النقل دون فهم، فإن ذلك يؤدي إلى نتائج غير صحيحة.

٥) **الثقة بالنفس:** وذلك بعدم الاستهانة بالكفاءة الشخصية والمهارات الذاتية، فإنها تنمو بالعمل وتتدرب بالمران على أن تبنى هذه الثقة على العمل والحفظ والذكاء والتذكر.

٦) **العقلية العلمية المنطقية المنظمة** التي تعنى بترتيب الأفكار وتحليلها وترتيبها وتنظيمها، وتدريب هذه العقلية على النقد والشك والتثبت وعدم الاستسلام للبديهيات والأفكار العامة بمعنى الشك العلمي، فتقرأ الخبر مرة واثنتين ، نشك فيه ونتأمل ونناقش ونتثبت من أجل الفهم والاطمئنان للوصول إلى الحقيقة واليقين.

٧) **الأمانة العلمية:** في النقل وعرض الأفكار وعزوها إلى أصحابها والإشارة إلى المصادر والمراجع التي انتفعت بها أو تعلمت منها فإن ذكر الفضل والإشادة بأهله دليل المروءة والأمانة وسعة الاطلاع أيضًا.

الفصل الثاني: التفكير العلمي

التفكير العلمي: هو منهج أو طريقة منظمة يمكن استعمالها في حياتنا اليومية أو في اعمالنا ودراستنا، فهو تفكير غير متخصص بموضوع معين بل يمكن ان يوجه في معالجة الموضوعات والقضايا والاحداث التي تواجهنا.

بينما تفكير العلماء يقوم على أساس دراسة مشكلة محددة متخصصة مستخدمين في ذلك لغة ورموزاً علمية خاصة.

خطوات التفكير العلمي

١) الشعور بالمشكلة.

٢) تحديد المشكلة.

٣) جمع المعلومات والبيانات المتعلقة بها.

٤) فرض الفروض

٥) اختبار صحة هذه الفروض والوصول إلى نتائج وحلول للمشكلة.

وهذا مثال على تطبيق خطوات الطريقة العلمية في التفكير من حياتنا اليومية :

اكتشف رجل بعد عودته من إجازته أن حديقته قد تعرضت للتلف الشعور بالمشكلة، أخذ يفكر فوجد باب الحديقة مكسورًا والزهور مقطوعة تحديد المشكلة، اعتقد أن أطفال الحي دخلوا وخربوا الحديقة وضع الفروض، لكنه لم يشاهد الأطفال، ومع ذلك فالاحتمال وارد، ثم قال: ألا يمكن أن تكون الحديقة قد تعرضت لعاصفة ؟ وضع فرض آخر؛ أخذ يسأل عن الأطفال فاكتشفت أنهم يعملون في نادٍ صيفي خارج المدينة منذ أسبوع ، فألغى هذا الفرض، ثم نظر فوجد أن حدائق جيرانه قد خربت أيضًا ، وبينما هو مستغرق في قراءة الجريدة ، عرف أن عاصفة قد هبت منذ يومين ؛ ثم سأل جيرانه فحدثوه عن عاصفة قوية ؛ هنا تأكد الرجل من صحة الفرض الثاني، فوصل إلى النتيجة ، وهي: أن العاصفة دمرت الحديقة.

خصائص التفكير العلمي

١) التراكمية: ينطلق التفكير العلمي من الواقع ، فالمعرفة بناء يسهم فيه كل الباحثين والعلماء ، وكل باحث يضيف شيء جديدًا إلى المعرفة، وتتراكم المعرفة وينطلق الباحث مما توصل إليه من سبقه من الباحثين، فيصحح أخطاءهم ويكمل خطواتهم أو قد يلغي معرفة سابقة ويبطل نظرية عاشت فترة من الزمن.

٢) التنظيم : إن وسيلة العلم هي إتباع منهج علمي، فالعلم معرفة منهجية تبدأ بالملاحظة ووضع الفروض واختبارها عن طريقة التجريب ثم الوصول إلى النتائج ، فالتفكير العلمي ليس منهجًا في تنظيم أفكارنا وعدم تركها حرة طليقة دون إلزامها بقواعد وقوانين فحسب، بل هو منهج في تنظيم العالم الخارجي أيضًا. فالباحث العلمي لا يناقش ظواهر متباعدة أو مفككة ، بل يدرس الظاهرة في علاقاتها بالظواهر الأخرى، فيكشف العلاقة بين الأسباب والنتائج، ويكشف الصلات والارتباط بين ظاهرة وأخرى ، ويميز ما بين التجاور الزماني والمكاني لظاهرات معينة تحدث معًا بالصدفة وما بين ظواهر مترابطة تظهر معًا نتيجة علاقات علمية ؛ فالحقيقة العلمية حين تكتشف تأخذ مكانها بين مجموعة الحقائق المكتشفة فتدمج معها أو تتفاعل معها ، وقد تعدل فيها أو تلغي بعضها والحقيقة العلمية بهذا المعنى ليست مستقلة عن الحقائق الأخرى.

٣) البحث عن الأسباب: يهدف العلم إلى فهم الظواهر التي يدرسها ، ولا يتم هذا الفهم من خلال الوصول إلى المعلومات والحقائق ، بل لا بد من تفسير هذه الظواهر وتحليلها عن طريق معرفة أسبابها وعوامل نشوئها وتطورها. إن معرفة أسباب ظاهرة ما هو الذي يمكن الإنسان من السيطرة عليها وضبطها والتأثير فيها وزيادتها أو إنقاصها، وبالتالي التحكم فيها وإخضاعها للتجربة والتعديل والتطوير. والعلم يبحث عن الأسباب كغيره من النشاطات الإنسانية؛ ولكن ما يميز التفكير العلمي هو بحثه عن الأسباب المباشرة لا الأسباب البعيدة، ويؤدي البحث عن الأسباب غرضين أساسيين هما: إرضاء حب الإنسان للاستطلاع والمعرفة والفهم ، وزيادة قدرة الإنسان على السيطرة على الظواهر عن طريق معرفة أسبابها والتحكم فيها.

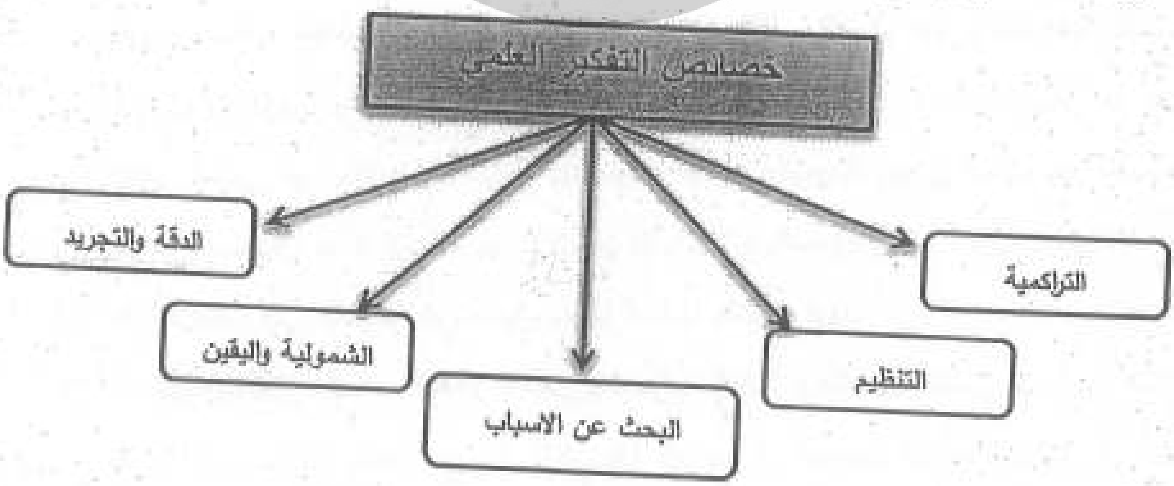
٤) الشمولية واليقين: يتصف التفكير العلمي بالشمولية واليقين ، فالباحث العلمي لا يدرس مشكلة محددة كهدف ، بل ينطلق من دراسة المشكلة المحددة أو الموقف الفردي للوصول إلى نتائج وتعميمات تشمل الظواهر المشتركة أو المواقف المشتركة مع موضوع دراسته ؛ لذلك فإن هدف العلم هو الوصول إلى تعميمات ونتائج تتسم بالشمولية وتنطبق على أكثر من فرد وأكثر من ظاهرة وأكثر من موقف. ويرتبط بالشمولية : شمولية الموضوع ؛ وشمولية من يتقبلون هذا

الموضوع ، صفة أخرى من صفات الحقيقة العلمية وهي" اليقينية " ،أي استناد الحقيقة العلمية على مجموعة كافية من الأدلة الموضوعة المقنعة بحيث لا يبقى هناك شك في صدقها ؛ واليقين العلمي يختلف عن اليقين الذاتي حين يقتنع شخص ما بفكرة معينة ، لأنها تبدو له واضحة صادقة ، أو لأنه يحس بصدقها ويشعر بصحتها دون وجود أدلة عليها ؛ إن هذا اليقين ليس علميًا لعدم استناده إلى أدلة محسوسة

٥) **الدقة والتجريد** : يتسم التفكير العلمي بالدقة والتجريد، وهذا ما يميزه أيضًا عن أنماط التفكير الأخرى، فالباحث العلمي يسعى إلى تحديد مشكلته بدقة، وتحديد إجراءاته بدقة ولا يستعمل سوى كلام دقيق محدد، ولكي ينجح الباحث العلمي في أن يكون دقيقا ويحدد مشكلاته وإجراءاته وفروضه بدقة، فإنه يستعمل اللغة الرياضية التي تقوم على أساس القياس المنظم الدقيق والتحدث بلغة الأرقام والرموز والعلاقات الرياضية المحددة. واستعمال هذه اللغة يؤدي إلى فهم دقيق للظواهر، فالأحكام الكيفية لا تساعد على فهم الظواهر بل قد تعطي فهما خاطئا لها. فإن استخدام كلمات مثل "ذكي" أو "ذكي جدا" أو "غبي" لا تعني شيئا محددا كأن نقول "تبلغ نسبة ذكاء الشخص (٩٠) أو (١٠٠) أو (١٢٠) بالمائة. فالأرقام تسمح بالمقارنة .والتفكير العلمي حين يستعمل الأرقام والقياس الكمي أو حين يستعمل لغة رياضية فإنه يجرد الأشياء من مادتها، فحين نقول ٣ + ٤ = ٧ فإننا لا نعني ثلاثة أو أربعة أشياء معينة، بل كل ثلاثة وكل أربعة مهما كان موضوع هذا العدد.

فالتجريد هو وسيلة الباحث العلمي للسيطرة على الواقع وفهم قوانينه وحركاته وتغيراته بنحو أفضل.

والمخطط الآتي يوضح خصائص التفكير العلمي:



عوائق التفكير العلمي

إن من أبرز العوائق التي تواجه المجتمعات في سعيها نحو البحث العلمي واستعمال الأساليب العلمية هي :

١) **انتشار الفكر الاسطوري والفكر الخرافي**: يلاحظ في مجتمعاتنا العربية والمجتمعات النامية أن الفكر الخرافي ما زال قوياً ويقف موقفاً معادياً للعلم والتفكير العلمي فالخرافات والاعتقاد بالقوى الخارقة لدى بعض الأشخاص وتحضير الأرواح ما زالت منتشرة وسيمر وقت طويل قبل أن يتخلص الانسان في مجتمعنا من هذا التفكير .

٢) **الالتزام بالأفكار الذائعة**: الانسان حمل الكثير من الأفكار والتقاليد القديمة والتي ما زالت حية حتى الآن، فالأفكار التي ابتكرها أجدادنا وآباؤنا والحكمة التي ورثناها من الأجيال القديمة ما زال ينظر إليها نظرة احترام وتقديس وما زال كثيرون يرفضون مجرد مناقشتها بل يؤمنون بها بنحو تام لا يقبل النقاش، ويزداد التمسك بهذه الأفكار القديمة كلما واجهت الانسان ظروف ومصاعب وكلما عاشت في ظروف تمنعها من التعبير الحر والتفكير العلمي .

٣) **إنكار قدرة العقل** : ينظر إلى العقل كأداة محددة في كشف الظواهر، أو كأداة عاجزة عن الوصول إلى الحقيقة، ولذلك كان الناس يبحثون عن أداة أخرى غير العقل وعن وسيلة أخرى غير المعرفة العلمية . ولكن هذه الاتهامات سرعان ما تتبدد حين نرى أن العقل الانساني يتطور باستمرار وأن المعرفة العلمية تتفجر بنحو هائل.

وفي الختام واذا كانت هناك سبلاً لتنمية القدرة على التفكير العلمي ، فان من بين هذه السبل اعداد المناهج الدراسية التي توفر المواقف التعليمية التي تسعى الى هذا الهدف والاعتماد على طرائق تدريس حديثة تتيح الفرص للأشخاص في ممارسة التفكير العلمي واستعمال التقنيات الحديثة لتقريب الحقائق الى عقول الطلبة واثارة اهتمامهم تحفيزهم على المشاركة في الانشطة المختلفة.

١-١ المقدمة :

لاشك بأن البحث العلمي هو الطريق الأمثل لإكتساب المعرفة وتفهم الظواهر الطبيعية المتداخلة، ويعتمد البحث العلمي على الدراسة الموضوعية للظاهرة، وعلى إعتماد الرقابة الإحصائية في تفسير العلاقات بين المتغيرات، فيرفض الباحث الأعتماد على الفطرة السليمة في الحكم على الظواهر، ويلجأ الى الطرق العلمية في سعيه الدائم نحو المعرفة.

١-٢ طريقة البحث العلمي :

يمكن تلخيص طريقة البحث العلمي بالخطوات الآتية :

أولا- تحديد مشكلة البحث :

لاشك بأن الخطوة الأولية في البحث العلمي هي ان يحدد الباحث المشكلة قيد الدراسة فبدون هذه الخطوة لن يتمكن الباحث من القيام ببحثه ، فعليه وقتئذ في بداية الامر أن يحدد مشكلة البحث ويضعها في شكل مقبول حتى ولو كان شكلا تجريبيا غير نهائي وعلى الرغم من ان المعضلة الكبرى في البحث العلمي تكون قد حلت فيما لو عزف الباحث ماذا يريد أن يبحث فعلا، فانه في أغلب الأحيان يصعب على الباحث في بداية الأمر أن يحدد مشكلة البحث في شكل مبسط، واضح وتام، إذن على الباحث في بداية الأمر أن يبدأ بقراءة البحوث والدراسات المتصلة بموضوعه الى ان يصبح متمكنا من الموضوع فعندئذ يجد في كل بحث يقرأه مشكلة أو أكثر جديرة بالبحث، كما ان مثل هذه القراءة ستعرف الباحث على ماتم انجازه في بحثه، فبذلك يتمكن من تضييق مشكلة البحث ويتعرف على تجارب الآخرين في تجميع معطياتهم .

توجد مواصفات معينه يمكن صياغة مشكلة البحث وفقا لها، وهذه المواصفات هي :

١- ان تصاغ مشكلة البحث في شكل جملة استفهامية، فيسأل الباحث ماهي العلاقة الموجودة بين متغيرين أو أكثر ؟ ويكون الهدف الاساسي من البحث هو الاجابة على هذا السؤال.

٢- اذا كانت مشكلة البحث هي مشكلة علمية، فحينئذ لابد أن تدرس العلاقة بين متغيرين أو أكثر ، ومثال ذلك السؤال الذي يبحث عن وجود علاقة بين المتغير X والمتغير Y ، أو يبحث عن العلاقة التي تربط المتغيرين X و Y بالمتغير Z ، أو يبحث عن وجود علاقة بين المتغير X والمتغير Y تحت الظروف A-B-C وتستثنى من ذلك البحوث التصنيفية Taxonomic Research والبحوث التاريخية Historical Research .

٣- يجب أن تتضمن مشكلة البحث امكانية الملاحظة والاختبار وهذا يعني ان المتغيرات قيد الدراسة قابلة للقياس ولاتتعلق بأسئلة مبنية على الاجتهاد الشخصي والتي لايستطيع العلم ان يجد جوابا لها.

ثانيا – صياغة فروض البحث Formulating testable hypotheses :

يبدأ البحث العلمي عادة بمشكلة قابلة للبحث، وبعد ان يعرف الباحث ماذا يريد ان يبحث فعلا اذن عليه ان يصيغ فروض بحثه. فالباحث يلاحظ ظاهرة ما ،ومن ثم يتوقع مسببات ونتائج لهذه الظاهرة، ولذا فان فروض البحث هي في الحقيقة توقعات ورهان.

وتجدر الاشارة هنا الى ان الباحث في بداية الامر قد يفترض فرضا عاما وواسعا، واذا كان هذا الفرض جيدا ، حينئذ باستطاعة الباحث ان يستنتج أو يستدل الى فروض ثانوية مبنية على الفرض العام، اما من حيث صياغة فروض البحث فيجب ان تتوافر فيها المواصفات الآتية:

١- يجب ان تصاغ فروض البحث في شكل جمل تصريحية Declarative sentence .

٢- يجب ان تربط فروض البحث بين متغيرين او أكثر.

٣- يجب ان تحتوي فروض البحث على مفهوم ضمني Implications مؤداة امكانية قياس المتغيرات المستقلة والمتغيرات التابعة وبالتالي امكانية اجراء اختبارات احصائية على العلاقات تحت الدرس.

٤- يجب ان لاتتعلق فروض البحث بالقيم والاجتهاد الشخصي ومثال ذلك أنه لايمكن صياغة فرض بحث المجرمون ليسوا أسوأ من رجال الاعمال لان مثل هذه الفروض تقوم على الاجتهاد الشخصي.

٥- يجب على فروض البحث ان تكون محددة، فعندما تكون التوقعات محددة يصغر احتمال الحصول على نتائج أو فروقات جوهرية في البحث عن طريق الصدفة، فعلى سبيل المثال لو أردنا ان نتوقع أثر هجرة أبناء الريف الى المدينة على انخفاض الانتاج الزراعي، فعلينا ان نحدد ونتوقع خصوصيات هذا الموضوع مثلا المكان والزمان ونسبة المهاجرين ... الخ . ويجب التنويه هنا الى أهمية التمييز بين فروض البحث Research or Scientific hypotheses وبين الفروض الاحصائية Statistical hypotheses ففروض البحث تكون عامة ومبنيه على نظرية ما Theory أو مبنيه على ملاحظات سابقة Prior observation أو مبنيه على اسس منطقية Logical basis ، ومثل هذه الفروض تتضمن عادة توقعات لنتائج البحث، علما انه بالاستدلال الاستنتاجي يمكن تحويل مثل هذه الفروض الى فروض احصائية قابلة للاختبار الاحصائي، أما الفروض الاحصائية فهي فرض العدم (H_0) والفرض البديل (H_1) والتي تصاغ عادة لتقييم فروض البحث وعليه فيمكن تقسيم الفروض كما يلي :

❖ فروض البحث : وهي فروض علمية مبنيه على نظرية، أو أسس منطقية أو ملاحظات سابقة تستخدم كدليل في البحث وتتضمن توقعات.

❖ فروض احصائية : هي تعبير عن واحد أو أكثر من مقاييس المجتمع التي سحبت منها العينة وفرض العدم والفرض البديل هما شكلان من الفروض الاحصائية .

١- فرض العدم Null hypotheses (H_0) : هو تعبير يتضمن واحد أو أكثر من المقاييس الخاضعة لاختبار أحصائي.

٢- الفرض البديل Alternative hypotheses (H_1) :وهو أي فرض يختلف عن فرض العدم H_0 ،ونلاحظ ان فرض العدم غالبا مايكون في الاتجاه المعاكس لفرض البحث. فعلى سبيل المثال قد نتوقع فروقات بين تأثير دوائيين على المرضى، لكن لاسباب إحصائية نفترض فرض العدم بأنه لاتوجد فروقات ثم بعد تحليل المعطيات فلو وجدنا فروقات جوهرية فحينئذ تعتبر بأن المعطيات تدعم فروض البحث، علما بان فرض العدم هو الفرض الذي يمكن رفضه لكن لايمكن برهنته.

ثالثا - أنواع المتغيرات:

بعد ان يحدد الباحث الفروض الاحصائية، عليه ان يحدد المتغيرات المستقلة والمتغيرات التابعة وكيفية قياسها، ويرى (Fer Guson 1976) أنه يمكن تصنيف المتغيرات الى أربع فئات:

أ- المتغير الاسمي A nominal Variable :

هو خاصية لاعضاء مجموعة ما محددة بطريقة تسمح لعمل تصريحات عن المساواة أو الفروق بين أعضاء المجموعة، فعلى سبيل المثال يمكن تصنيف الاشخاص بالنسبة الى لون عيونهم، في هذه الحالة تعتبر اللون متغير أسمي، فنقول بأن شخص ما له نفس لون عين أو لون مختلف عن عيني شخص آخر.

ب- المتغير الترتيبي An ordinal Variable :

هو خاصية لاعضاء مجموعة ما محددة بطريقة تسمح لعمل تصريحات عن نظام تراتيب الاعضاء في المجموعة، وعليه يمكن عمل تصريحات عن المساواة أو الفروق كما في حالة المتغير الا سمي بالاضافة الى تصريحات من نوع (أكثر أو أقل من) فعلى سبيل المثال يمكن

ترتيب العمال بالنسبة الى مهارة العمل، فنقول ان العامل X هو أكثر أو أقل مهارة من العامل Z ، أو ان العاملين متساويان من حيث المهارة في العمل.

ت- المتغير الفنوي An interval Variable :

هو خاصية لاعضاء مجموعة ما محددة بطريقة تسمح لعمل تصريحات عن تساوي الفئات بالاضافة الى الخاصيات المتعلقة بالمساواة أو الفروق والمتعلقة الاكثر او الأقل من. هذا النوع من المتغير لايحتوي على نقطة صفر حقيقي ((A true zero point)) وقد تحدد هذه النقطة عشوائيا للسهولة، مثال ذلك درجات الحرارة، فلو كان لدينا ثلاثة مركبات أ ، ب ، جـ بدرجات حرارة ١٢٠°، ٢٤٠° ، ٣٦٠° فحينئذ يمكن القول بأن الفرق بين درجات الحرارة لـ أ او ب يساوي الفرق بين ب و جـ كذلك يمكن القول بان الفرق في درجات الحرارة بين أ وجـ يساوي ضعف الفرق بين أ و ب أو ب و جـ ، علما بانه لايمكن القول بأن درجة الحرارة ب هي ضعف درجة حرارة أ ،لأن نقطة الصفر هي عشوائية وليست حقيقية .

ث- المتغير النسبي A ratio Variable :

بالاضافة الى التصريحات الممكنه في الفقرات أ وب وت فهذا المتغير يسمح بتصريحات عن المساواة بين النسب لان هذا المتغير يحتوي على الصفر الحقيقي مثال ذلك الوزن ، الطول ، ... فيمكن القول بأن طول شئ ما هو ضعف طول أو ثلاثة أضعاف طول شئ آخر، أذن الفرق بين المتغير النسبي والمتغير الفنوي هو ان القياسات في حالة المتغير النسبي تقاس من نقطة الصفر الحقيقية.

كما يمكننا تصنيف المتغيرات الى متغيرات متصلة أو متغيرات متقطعة، بدلا من تصنيف المتغيرات بالطريقة السابقة، حيث ان *المتغير المتصل* ***Continuous Variable*** *هو المتغير الذي يمكنه اخذ اية قيمة بين قيمتين معينتين، ومثال ذلك العمر* ، فعلى سبيل المثال من الممكن ان يكون عمر شخص ٤٠ سنة أو ٤٠,٢ (أربعين سنة وشهران) ، *أما المتغير المتقطع* ***Discrete Variable*** *فلا يمكن ان ياخذ اية قيمة بين قيمتين ، ومثال ذلك عدد أطفال أسرة* ، فلا يمكن القول بانه لدى أسرة ما ٢,٥ طفل .

رابعاً : تجميع المعطيات (البيانات) Data Collection:

في طريقنا لمعرفة أي شئ عن الناس ، الظواهر الطبيعية ، الحوادث .... الخ علينا ان نبدأ بتحديد المجتمع Population or Universe موضوع الدراسة ثم نختار عينه ممثله لهذا المجتمع Representative Sample وذلك بالاعتماد على احدى الطرق المعروفة في أختيار عينه عشوائيه . وفي حالة التجارب الاحصائية فيجب توزيع الوحدات تحت الدراسة بشكل عشوائي على المجموعات. وهنا تجدر الاشارة الى ان الخطوة الثانية الواجب اتباعها هي تحديد طرق تجميع المعطيات حيث يوجد تفاعل مشترك بين مشكلة البحث وطرق تجميع المعطيات ، فأحيانا توجد مشكلات جيدة للبحث الاانه

لاتوافر طرق جيدة لتجميع البيانات، لكن في أغلب الأحيان توجد طرق عديدة متاحة لتجميع البيانات وهي أمتداد من نظرية وطرق القياسات **Measurement Theory and Methods** وتعتمد بشكل أساسي على الملاحظة وأعطاء الرموز أو الاعداد للوحدات أو لمجموعات من الوحدات تحت الدراسة. ولاشك بأن الخبرة تلعب دورا مهما في تجميع البيانات.

بعد تجميع البيانات ، فان الباحث يعمل على تحليل البيانات الى أجزاء متكاملة حتى يتمكن من الأجابة على أسئلة البحث ، ويتمكن من أختيار فروض البحث ، فعلى الباحث تفسير النتائج والاجابه على أسئلة وفروض البحث ومن هنا يظهر لنا الترابط الوثيق بين تحليل البيانات وتفسير النتائج **Analysis and interpretation** ، وهنا نشير الى ان الباحث قبل تجميع المعطيات وعند تحديد مشكلة البحث وفروض أو اسئلة البحث عليه ان يحدد طرق تصنيف وترتيب وتلخيص البيانات، وتحديد النماذج الاحصائية لتحليل البيانات، ثم يحلل البيانات ويعمل على تفسيرها وربطها بالنظرية التي أشتقت منها مشكلة البحث، أو ربطها بنتائج البحوث الاخرى، علما بان لكل باحث حرية الاختيار من حيث طرق الملاحظة والقياس والتجربة والتحليل لكن عليه ان يختار الطرق الملائمة والمناسبة لمشكلة البحث وإلا كان تفسير الباحث لنتائج التحليل خاطئا نظرا لخطأ الطرق المستخدمة في تجميع وتحليل البيانات.

وتجدر الاشارة هنا على ان نتائج البحث قد تكون ايجابية ومتوافقة مع فروض البحث، وقد تكون النتائج سلبية ومخالفة لما جاء في النظرية أو لما جاء في بحوث سابقة، بيد ان كلا النتائج الايجابية والسلبية تعد مفيدة ، علما بان تفسير النتائج السلبية هو أصعب على الباحث من تفسير النتائج الايجابية، حيث يفترض بالباحث في هذه الحالة أن يتأكد بأن طرق تجميع بياناته وتحليلها كانت مناسبة وملائمة لمشكلة البحث ولايوجد خطأ في القياسات وبعد أن يتأكد من صحة الطرق المستخدمة في بحثه ويتأكد من انها لم تكن السبب في الحصول على النتائج السلبية ، حينئذ تعد نتائج بحثه اسهاما في المعرفة العلمية، شأنها في ذلك شأن النتائج الايجابية، أما اذا كانت نتائج بحثه ايجابية، فعلى الباحث في هذه الحالة ألا يعتقد بأن النتائج تبرهن فروض البحث لأن البحث العلمي لايبرهن أي شئ وأنما يفسره.

١-٣ أساليب جمع البيانات :-

ان أي بحث علمي يستند في تحليله الى الطريقة الاحصائية، يحتاج الى بيانات ومعلومات حول موضوع البحث قيد الدراسة. ويمكن للباحث الحصول على هذه البيانات والمعلومات من أحد المصدرين الآتيين :

أ- المصادر التاريخية **Historical Sources** :
وهي البيانات والمعلومات المحفوظة والمتجمعة لدى أجهزة ومؤسسات ودوائر الدولة المختلفة نتيجة لأستقصاءات أو مسوحات قامت بها هذه الجهات أو هيئات معينه لاغراض خاصة بها أو تجمعت لديها بحكم وظائفها الادارية والفنية مثال على ذلك

البيانات المتجمعة عن تعدادات السكان في العراق، احصاءات الانتاج الزراعي والصناعي،احصاءات التجارة الداخلية والخارجية،احصاءات الطلبة المتخرجين من الجامعات العراقية وغيرها .

ب- مصادر الميدان Field Sources :
وهذه تمثل بيانات ومعلومات يمكن الحصول عليها من مصادرها الاصلية بطريقة المراسلة أو المواجهة أو أية طريقة اتصال أخرى. مثال على ذلك نتائج التعداد العام لسكان العراق عام ١٩٨٧ ، تسجيل حوادث الطرق خلال شهر معين، تسجيل وقوعات الزواج والطلاق خلال عام ٢٠٠٩ ، وغيرها.

ان اختيار هذا المصدر دون ذاك في جمع البيانات والمعلومات يعتمد بالاساس على طبيعة البحث والنتائج المتوخاة منه، وهنالك اسلوبان يمكن من خلالهما جمع البيانات والمعلومات أياً كان مصدرهما، هذان الاسلوبان هما أسلوب التسجيل الشامل وأسلوب العينات، علماً ان اختيار أي من هذين الاسلوبين يعتمد وبشكل أساس على طبيعة البحث أو الدراسة والدقة المطلوبة لنتائجها.

أولاً - أسلوب التسجيل الشامل Census :

يقصد بأسلوب التسجيل الشامل، جمع البيانات والمعلومات عن كافة المفردات التي تؤلف المجتمع الاحصائي للظاهرة أو الظواهر قيد البحث، وفي هذه الحالة يجب ان يكون هذا المجتمع محدداً، أي مامعنيه انه يمكن مواجهة وملاحظة كل مفردة من مفرداته. مثال على ذلك عملية التعداد العام للسكان في العراق لعام ١٩٨٧ حيث تم تسجيل البيانات والمعلومات عن كل فرد (المفردة الاحصائية) دون استثناء ويعتبر أسلوب التسجيل الشامل أفضل أسلوب في جمع البيانات كونه يجهز الباحث ببيانات كاملة عن كافة مفردات مجتمع الدراسة إلا إنه يحتاج الى وقت وجهد وموارد مادية وبشرية كبيرة في انجاز مهمة جمع البيانات بالاضافة الى احتمال الوقوع في اخطاء نتيجة التعامل مع مفردات كثيرة.

ثانياً - أسلوب العينات Samples :

يقصد بأسلوب العينات عملية جمع البيانات والمعلومات عن مجموعة معينه من مفردات مجتمع الدراسة، هذه المجموعة من المفردات تسمى ((عينة Sample))

مثال على ذلك حصر نشاط بعض الوحدات الصناعية في العراق، أستفتاء رأي بعض الاشخاص بالبرامج التلفزيونية، دراسة فاعلية دواء معين على بعض الاشخاص المصابين بمرض السرطان، وغيرها من الامثلة.
ويمتاز أسلوب العينات بأنه يحتاج الى وقت وجهد وموارد مادية وبشرية أقل مما يحتاجه أسلوب التسجيل الشامل.

٤-١ المجتمع والعينة Population & Sample :

المجتمع Population: هو عبارة عن جميع القيم أو المفردات التي يمكن ان يأ خذها المتغير ،أو هو مجموعة من الافراد أو المشاهدات التي تشترك في صفة متغيرة واحدة أو أكثر تميزه تماما عن بقية المجتمعات مثل قسم الرياضيات في كلية التربية ابن الهيثم، ويقسم المجتمع الى قسمين:

i. مجتمع محدد **Finite population:**
أي من الممكن جمع عدد مفرداته أو مشاهداته مثل عدد التدريسيين في كلية ما.

ii. مجتمع غير محدد **Infinite population :** المحصوره لا حقل ما
أي من غير الممكن جمع عدد مفرداته أو مشاهداته مثل عدد ذرات الارز، عدد الاسماك في نهر دجلة .

العينة Sample: هي جزء من المجتمع يجري اختيارها وفق قواعد خاصة لكي تمثل المجتمع تمثيلا صحيحا.

٥-١ التفكير العلمي وخصائصه:

التفكير العلمي منهج أو طريقة منظمة يمكن إستخدامها في معالجة جميع الموضوعات والقضايا التي تواجهنا في حياتنا اليومية أو في أعمالنا ودراستنا دون إعتبار للتخصص ، ويقوم التفكير العلمي على المبدأين التاليين:

١- لايمكن إثبات الشئ ونقيضه في نفس الوقت ، فالشئ إما ان يكون موجودا أو غير موجود ،فالتفكير العلمي لايجمع بين النقائض في سمة واحدة.

٢- إن لكل حادثة سبب ، وان هذا السبب يؤدي الى ظهور النتيجة مالم يكن هناك عائق.

٦-١ السمات المميزة للتفكير العلمي :

أولاً - التراكمية: لقد تراكمت المعارف العلمية عبر القرون ، واستفاد فيها اللاحق من جهد السابق وأستكمل الطالب عمل المعلم حتى غدونا نعيش في عصر العلم، والمتتبع لتاريخ العلم يجد بذور المعارف العلمية تمتد الى أيام الحضارات الأولى.

وينطلق الباحث مما توصل اليه من سبقه من الباحثين، فيصحح أخطاءهم ويكمل خطواتهم ،أو قد يلغي معرفة سابقة.

ثانياً – التنظيم : ان الحقائق العلمية تتكامل على صورة منظومات أو أبنية متناسقة ، فموضوعات العلم الواحد تكون مترابطة بعضها مع بعض بعلاقات بحيث يبدو وكأن كل قانون إنما يدخل في إطار قانون أعم، وهذا القانون العام يدخل في إطار قانون أكثر عمومية وهكذا.

٢_١ علم الأحصاء :-

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

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

١) الاحصاء الوصفي **Descriptive Statistic** :
ويتضمن هذا الفرع الطرق والاساليب المستخدمه في جمع البيانات والمعلومات عن ظاهرة معينه أو مجموعة ظواهر وكيفية تنظيم وتصنيف وتبويب هذه البيانات مع امكانية عرضها في جداول ورسومات بيانية وحساب بعض المؤشرات الاحصائية منها.

٢) الاحصاء الاستدلالي **Inferential Statistic** :
وهو الشطر الآخر من علم الاحصاء الذي يتم عادة بموضوعي التخمين **Estimation** واختبار الفرضيات **Testing hypotheses** .

ويعتبر علم الاحصاء أحد الوسائل المهمة في البحث العلمي من خلال استخدام قواعده وقوانينه في جمع البيانات والمعلومات اللازمة للبحث العلمي وتحليل هذه البيانات والمعلومات بغية الوصول الى النتائج التي يهدف لها البحث. كما وان للاحصاء دورا بارزا في وضع الخطط المستقبلية عن طريق التنبؤ بالنتائج ولكافة القطاعات سواء كانت انتاجية أم خدمية، وحيث ان الاحصاء بحد ذاته يعتبر وسيلة وليس غاية فذلك يعني ان

مجالات تطبيق علم الاحصاء ممكنه سواء كان ذلك في مجال العلوم الصرفة أو العلوم الانسانية وغيرها.

٢ ـ ٢ الطريقة الاحصائية في البحث العلمي :

ان أستخدام الأسلوب الاحصائي في البحث العلمي يعني توفر بيانات ومعلومات عن الظاهره أو الظواهر المطلوب دراستها في ذلك البحث، وهذا يعني ان امكانية تطبيق الطريقة الاحصائية مرهون بامكانية التعبير عن هذه الظاهرة أو تلك تعبيرا كميا.

وفيما يلي المراحل الرئيسية للطريقة الاحصائية في البحث العلمي :-

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

٢ـ٣ أسلوب تصميم البحوث :

هناك اعتبارات كثيرة يتوقف عليها تصميم البحث، ففي كل تصميم يتوجب على الباحث الأخذ بنظر الاعتبار مسألة الحصول على البيانات والمعلومات بأقصر وقت وأقل جهد وكلفة وهذا يعني انه يجب مراعاة مايلي عند تصميم البحث :

أ. تحديد الغرض من البحث :

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

ب. تحديد امكانية التنفيذ الفعلي للبحث :

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

ت- تحديد اطار البحث :

ان أحد الامور الهامة قبل البدء بتنفيذ البحث هو تحديد نوع وطبيعة مجال ذلك البحث، أي تحديد المجتمع الاحصائي Statistical Population على نحو واضح ودقيق، والمجتمع الاحصائي عبارة عن مجموعة من الوحدات أو المفردات Units التي تشترك بصفة أو صفات معينه والتي غالبا مايتم الحصول منها على البيانات والمعلومات المطلوبة فمثلا لو كان الهدف من البحث هو التعرف على نسبة عدد الافراد المصابين بأحد الامراض المتوطنه في قرية معينه فان المجتمع الاحصائي هنا هو كافة الافراد المقيمين في هذه القرية والمفردة الاحصائية هنا تمثل الفرد المقيم في هذه القرية الممكن اخضاعه لعملية الفحص الطبي.

ث- تحديد أسلوب جمع البيانات والمعلومات :

بغية الحصول على البيانات والمعلومات التي يتطلبها البحث أو الدراسة هنالك أسلوبان يمكن من خلالهما جمع هذه البيانات والمعلومات كل منهما له ميزاته وعيوبه ،هذان الاسلوبان هما أسلوب التسجيل الشامل عن كافة مفردات المجتمع الاحصائي وأسلوب التسجيل عن طريق العينات.

٢_٤ طبيعة البيانات الإحصائية :

عند جمع بيانات حول ظاهرة ما فاننا نرمز للظاهرة بالرمز (X) وكل مفردة أو مشاهدة منها نرمز لها بالرمز (x_i) ، فمثلا عند دراسة درجات الطلبة في إحدى الجامعات فاننا نرمز لصفة الدرجة بالرمز (X) ودرجة أي طالب بالرمز (x_i) وتسمى المشاهدة أو المفردة (Observation) وان قيمة (x_i) قد تختلف من طالب الى آخر ولهذا نقول بأن (X) هو متغير (Variable) ، وتقسم المتغيرات الى قسمين:

أ- متغيرات وصفية (نوعية) Qualitative Variable :

وهي المتغيرات أو الصفات التي لايمكن قياسها بوسائل قياس مألوفة كالعد أو القياس إنما تكون صفات لذلك المتغير مثل لون العين كمتغير يمكن ان يكون (أسود، عسلي، أزرق) ، الجنس (ذكر،أنثى) ، الحالة الاجتماعية (أعزب ، متزوج، مطلق، أرمل) وغيرها من الحالات.

ب- متغيرات كمية Quantitative Variable:

وهي المتغيرات التي يمكن قياسها مباشرة بأرقام عددية أو وسائل قياس مألوفة مثل عدد الطلبة في مرحلة ما، عدد أشجار البرتقال في بستان ، طول الشخص بالسنتمتر ، وهذا القسم من المتغيرات يكون على نوعين:

١- متغيرات كمية متصلة **Continuous Variable** :

وهي تلك المتغيرات التي تأخذ فيها المفردة أو المشاهدة أي قيمة رقمية في مدى معين (فترة) مثل درجات الحرارة خلال اليوم، أطوال الطلبة في قسم ما.

٢- متغيرات كمية متقطعة **Discrete Variable** :

وهي تلك المتغيرات التي تأخذ فيها المفردة أو المشاهدة أعداداً صحيحة فقط مثل عدد الطلاب في قسم ما، عدد أفراد الاسرة، كما تم شرح ذلك في الفقرات السابقة . ① ②

---

① كتاب (الانحدار الخطي المتعدد) تأليف د. عبد الرزاق محمد صلاح شربجي / وزارة التعليم العالي والبحث العلمي/ جامعة الموصل / ١٩٨١ .

② كتاب (أساليب البحث العلمي) د. جودت عزت عطوي / دار الثقافة للنشر والتوزيع / ٢٠٠٩.

Chapter 3 – Descriptive Statistics

Frequency Distributions

A frequency distribution is a table used to describe a data set. A frequency table lists intervals or ranges of data values called **data classes** together with the number of data values from the set that are in each class. This number is called the **frequency** of the class.

Example: Statistics exam grades. Suppose that 20 statistics students' scores on an exam are as follows:

97, 92, 88, 75, 83, 67, 89, 55, 72, 78, 81, 91, 57, 63, 67, 74, 87, 84, 98, 46

We can construct a frequency table with classes 90-99, 80-89, 70-79 etc. by counting the number of grades in each grade range.

Class	Frequency (f)
90-99	4
80-89	6
70-79	4
60-69	3
50-59	2
40-49	1

Note that the sum of the frequency column is equal to 20, the number of test scores.

2- General Rule for construction frequency table:

1) Total Range (T.R): Is the difference between the largest value and smallest value in the group of data.

$$T.R = x_L - x_s + 1$$

or

$$T.R = X_L - X_s$$

2) Number of classes : it is the number of groups in the frequency table.

a) Yule formula :

$$m = 2.5\sqrt[4]{n}$$

b) Sturges formula:

$$m = 1 + 3.322\log_{10}(n)$$

Where:

m:The number of classes

n: The number of data.

3) Length of a class: It is also called the class with, which is the difference between two consecutive lower (or upper) class limits.

$$L = \frac{T.R}{m}$$

Where:

L: Length of class

T.R: Total Range

m: Number of classes

4- Lower class Limits: Are the smallest numbers that belong to each class.

5- Upper class Limits: Are the largest numbers that belong to each class.

6- Class midpoints: Are the middle numbers of each classes.

$$X = \frac{L.C + UC}{2}$$

Where:

X: class midpoint.

L.C: lower class limit.

U.C: upper class limit.

7- Class Frequency: The frequency is the number of times that each of the different values in the given class occurs in the data, and the sum of the frequencies in a frequency table should equal the number of values in the data set as: $\sum_{i=1}^{m} f_i = n$

$$\sum_{i=1}^{m} f_i = f_1 + f_2 + f_3 + \cdots + f_m = n$$

f_i: The frequency of each class.

n: The number of data.

The general shape of the frequency table of the equal class is set as:

Seq.	*lower limit*	*upper limit*	*Class len.* (L)	(X) *midpoint*	**Freq.** (f)
1	X_s	$X_s + L$	L	$X_s + (1/2)L$	f_1
2	$X_s + L$	$X_s + 2\,L$	L	$X_s + (3/2)L$	f_2
3	$X_s + 2\,L$	$X_s + 3\,L$	L	$X_s + (5/2)\,L$	f_3
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
m	$X_s + (m-1)\,L$	$X_s + m\,L$	L	$X_s + (m-(1/2))\,L$	f_n
sum	....	......	......	....	n

3- Class frequency tables for Discrete Quantitative date.

In this case we write the limit as the follow frequency distribution, so as to ensure that each value of the data values within one class of the distribution classes without any repeat could happen at any of the distribution classes.

Seq.	Lower limit (from)	Upper limit (To)	Classes limit	Freq. (f)
1	X_s	X_s+L-1	$(X_s)-(X_s+L-1)$	f_1
2	X_s+L	X_s+2L-1	$(X_s+L)-(X_s+2L-1)$	f_2
3	X_s+2L	X_s+3L-1	$(X_s+2L)-(X_s+3L-1)$	f_3
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
m	$X_s+(m-1)L$	$X_{s+}mL-1$	$X_s+(m-1)L)-(X_{s+}mL-(1)$	f_n
Sum.	....	....	....	n

Example ① :

The following sample data set list the grades of 30 students in statistic , construct a frequency distribution table.

56	65	68	52	63	55	79	73	42	61
72	47	82	87	70	63	48	43	73	44
45	53	75	60	68	82	76	53	46	70

Sol.

1- T.R.= $X_L - X_S + 1$

T.R. $= 87 - 42 + 1 = 46$

2- $m = 2.5\sqrt[4]{n} = 2.5\ \sqrt[4]{30} = 5.8508 \approx 6$

3- $L = \frac{T.R.}{m} = \frac{46}{6} = 7.67 \approx 8$

So the frequency distribution table will have 6 classes as follows:

Seq.	Lower limit (from)	Upper limit (to)	Classes limit	frequency f
1	42	42+08-1	42-49	7
2	42 + 08	42+16-1	50 – 57	5
3	42+16	42+24-1	58-65	5
4	42+24	42+32-1	66-73	7
5	42+32	42+40-1	74-81	3
6	42+40	42+48-1	82-89	3
Sum.				30

We can write the frequency distribution table as follow:

Seq.	classes	(f) frequency	(x) midpoint
1	42-49	7	45.5
2	50-57	5	53.5
3	58-65	5	61.5
4	66-73	7	69.5
5	74-81	3	77.5
6	82-89	3	85.5
Sum.		30	

Class midpoints are the middle numbers of each class, so for the first class the midpoint set as:

$$X_1 = \frac{L.C + UC}{2} = \frac{(42 + 49)}{2} = 45.5$$

The class midpoint are: 45.5 , 53.5 , 61.5 , 69.5 , 77.5 , 85.5

4 - Class frequency tables for Continuous Quantitative data.

In this case we write the limit as the follow frequency distribution, so as to ensure that each value of the data values within one class of the distribution classes without any repeat could happen at any of the distribution classes.

Seq.	*Lower limit*	*Upper limit*	*classes*	Freq. (f)
1	X_s	$X_s + L$	$(X_s) - < (X_s + L)$	f_1
2	$X_s + L$	$X_s + 2L$	$(X_s + L) - < (X_s + 2L)$	f_2
3	$X_s + 2L$	$X_s + 3L$	$(X_s + 2L) - < (X_s + 3L)$	f_3
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
m	$X_s + (m-1)L$	$X_{s} + mL$	$(X_s + (m-1)L) - < (X_s + m)L$	f_n
Sum.	....	......	....	n

Example ②:

The following sample data set list the weighted of (16) students , construct a frequency distribution table.

90	73.5	88.5	66.5	55.5	46.5	46	80.5
87.5	69.5	77.3	60.8	48.2	51.3	49.2	49.5

As shows the random variable is a continuous variable , so the frequency distribution table set as :

1- $T.R. = X_L - X_S + 1 = 90-46+1=45$

2- $m = 2.5 \sqrt[4]{n} = 2.5 \sqrt[4]{16} = 5$

3- $L = \frac{T.R.}{m} = \frac{45}{5} = 9$

Seq.	Class limit	(f) frequency	(X) Midpoint
1	46 - < 55	6	50.5
2	55 - < 64	2	59.5
3	64 - < 73	2	68.5
4	73 - < 82	3	77.5
5	82 - < 91	3	86.5
Sum.	....	16	......

5- Cumulative Frequency distribution :

The cumulative frequency of a data is the number of data elements in that class and all previous classes. You get the cumulative frequency from standard frequency.

1)) Ascending cumulative frequency: It is the summation for the frequency from the first class to the last class, and it's symbol is (FV).The last value will always be equal to the total for all observations, since all frequencies will already have been added to the previous total.

Upper limit classes	Cumulative sentence	Freq.	(Fv)
X_s+L-1 X_s+2L-1 X_s+3L-1 . . . X_s+mL-1	Less than or equal upper limit	f_1 f_2 f_3 . . . f_m	$Fv_1=f_1$ $Fv_2=f_1+f_2=Fv_1+f_2$ $Fv_3=f_1+f_2+f_3=Fv_2+f_3$ . . . $Fv_m=f_1+f_2+\ldots+f_m=Fv_{m-1}+f_m=n$
sum		n	

Ascending cumulative frequency table for discrete random variable.

Upper limit classes	Cumulative sentence	Freq.	(Fv)
X_s+L X_s+2L X_s+3L . . . X_s+mL	Less than upper limit	f_1 f_2 f_3 . . . f_m	$Fv_1=f_1$ $Fv_2=f_1+f_2=Fv_1+f_2$ $Fv_3=f_1+f_2+f_3=Fv_2+f_3$ . . . $Fv_m=f_1+f_2+\ldots+f_m=Fv_{m-1}+f_m=n$
sum		n	

Ascending cumulative frequency table for continuous random variable.

Example ③: For the following frequency distribution table , find the ascending cumulative frequency.

Number of trees	60-74	75-89	90-104	105-119	120-134	135-149	150-164
Number of families	4	5	10	12	16	7	6

Note : the random variable in this example is discrete variable.

classes	Upper limit classes	f_i	Fv
60-74	Less than or equal 74	4	4
75-89	Less than or equal 89	5	9
90-104	Less than or equal 104	10	19
105-119	Less than or equal 119	12	31
120-134	Less than or equal 134	16	47
135-149	Less than or equal 149	7	54
150-164	Less than or equal 164	6	60
sum		60	

Example ④ : The following distribution frequency for a sample of 100 students weighted , find the ascending cumulative frequency.

classes	46-	53-	60-	67-	74-	81-	88-	95-<102
Freq.	7	15	27	21	14	8	5	3

Note : the random variable in this example is continuous variable.

classes	The upper limit	f_i	Fv
46-	Less than 53	07	07
53-	Less than 60	15	22
60-	Less than 67	27	49
67-	Less than 74	21	70
74-	Less than 81	14	84
81-	Less than 88	08	92
88-	Less than 95	05	97
95-<102	Less than 102	03	100
sum		100	

2)) Descending cumulative frequency: It is the decreasing for the frequencies from the first class to the last class, and it's symbol is (FF).

lower limit classes	Cumulative sentence	Freq.	(FF)
X_s	Largest or equal lower limit of the class	f_1	$FF_1=f_1+f_2+\ldots+f_m=n$
X_s+L		f_2	$FF_2=n-f_1=FF_1-f_1$
X_s+2L		f_3	$FF_3=n-f_1-f_2=FF_2-f_2$
.		.	.
.		.	.
.		.	.
$X_s+(m-1)L$		f_m	$FF_m=n-f_1-f_2-\ldots-f_{m-1}=FF_{m-1}-f_{m-1}=f_m$
sum		n	

Descending cumulative frequency table for discrete random variable.

lower limit classes	Cumulative sentence	Freq.	(FF)
X_s X_s+L X_s+2L . . . $X_s+(m-1)L$	Largest or equal lower limit of the class	f_1 f_2 f_3 . . . f_m	$FF_1=f_1+f_2+....+f_m=n$ $FF_2=n-f_1=FF_1-f_1$ $FF_3=n-f_1-f_2=FF_2-f_2$ . . . $FF_m=n-f_1-f_2-....-f_{m-1}=FF_{m-1}-f_{m-1}=f_m$
sum		n	

Descending cumulative frequency table for continuous random variable.

Example ⑤ : Back to example (3),find the descending cumulative frequency distribution.

Sol.:

Classes	Lower limit classes	f_i	FF
60-74	≥ 60	4	60
75-89	≥ 75	5	56
90-104	≥ 90	10	51
105-119	≥ 105	12	41
120-134	≥ 120	16	29
135-149	≥ 135	7	13
150-164	≥ 150	6	6
sum	....	60	

cumulative frequency distribution.

Sol.:

classes	Lower limit classes	f_i	FF
46-	≥ 46	07	100
53-	≥ 53	15	93
60-	≥ 60	27	78
67-	≥ 67	21	51
74-	≥ 74	14	30
81-	≥ 81	08	16
88-	≥ 88	05	08
95-<102	≥ 95	03	03
sum	....	100	

7 – A Relative Frequency Distribution has the same class limits as the frequency distribution but instead of listing frequency list relative frequencies.

Relative frequency = class frequency / sum of all frequency

$$rf = f / n$$ (can be written as a percent)

Example: The following distribution are exam scores for 90 students, find the relative frequency distribution.

classes	f_i	*rf*
10-	2	(2/90) *100 =2.222%
20-	*3*	(3/90)*100 =3.333%
30-	*5*	(5/90)*100 =5.556%
40-	10	(10/90)*100 =11.111%
50-	18	(18/90)*100 =20%
60-	25	(25/90)*100 =27.778%
70-	15	(15/90)*100 =17.778%
80-90	12	(12/90)*100 =14.444%

6 – Graphing a simple Frequency Distribution:

To visualize data we graph the pictures of distributions. Among the different types of graph in this chapter (a histogram , a polygon), we start with histogram.

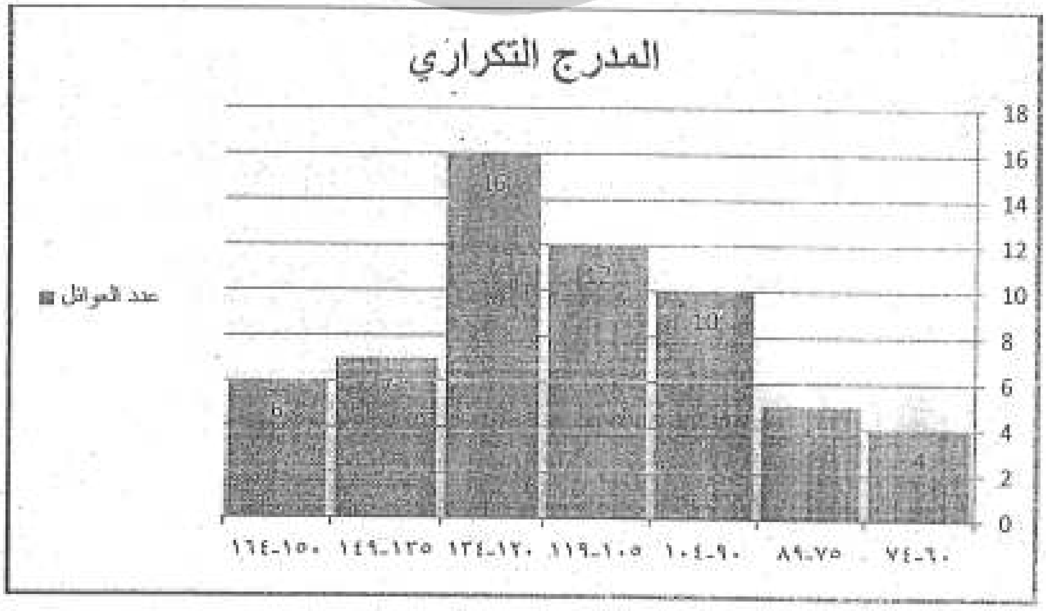
1- A histogram:

Def.: A histogram is a graph that uses bars to portray the frequency or the relative frequency of the possible outcomes for the numerical data. In which the horizontal of the bars correspond to the frequency values, and the bars are drawn adjacent to each other

Example ⑦: Draw a frequency histogram for the frequency table in example (3).

عدد الاشجار	60-74	75-89	90-104	105-119	120-134	135-149	150-164
عدد العوائل	4	5	10	12	16	7	6

Sol.:



Example ⑧: Draw a frequency histogram for the frequency table in example (4).

الفئات	46-	53-	60-	67-	74-	81-	88-	95 -< 102
التكرارات	7	15	27	21	14	8	5	3

Sol.:

2 – Polygon :A frequency polygons are very similar to histogram, except instead of using bars to represent the frequency of each class, a single plotted point over the class is used. These points are then jointed by line segments.

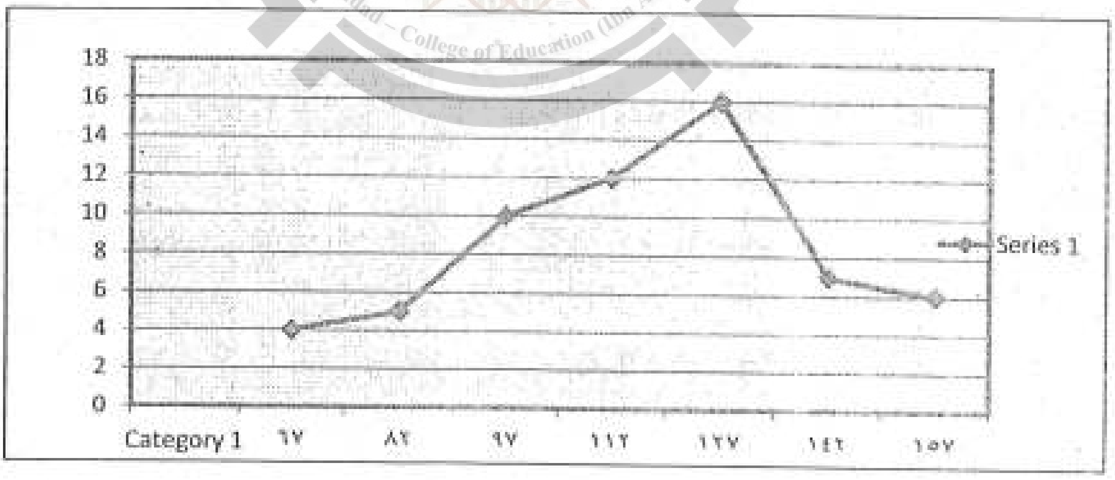
Example ⑨: Draw a frequency Polygon for the frequency table in example (3).

classes	60-74	75-89	90-104	105-119	120-134	135-149	150-164
Freq.	4	5	10	12	16	7	6

Sol.: First we have to find the midpoint for each class.

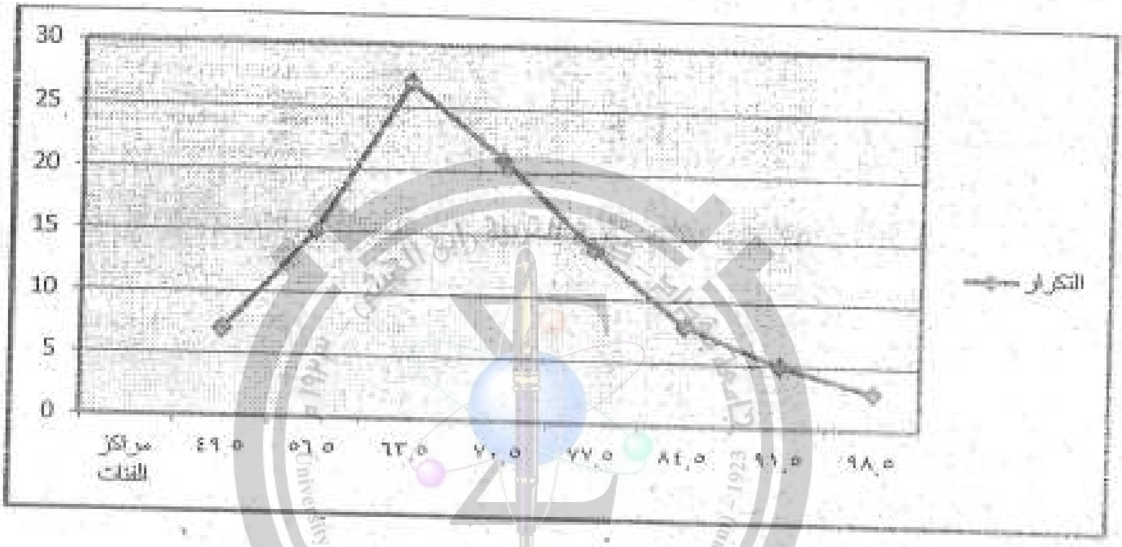
Midpoint (first class) = (60 +74) / 2 = 67 ....etc. for last class .

Freq.	4	5	10	12	16	7	6
Midpoint class(x)	*67*	*82*	*97*	*112*	*127*	*142*	*157*



Example ⑩ : Draw a frequency Polygon for the frequency table in example (4).

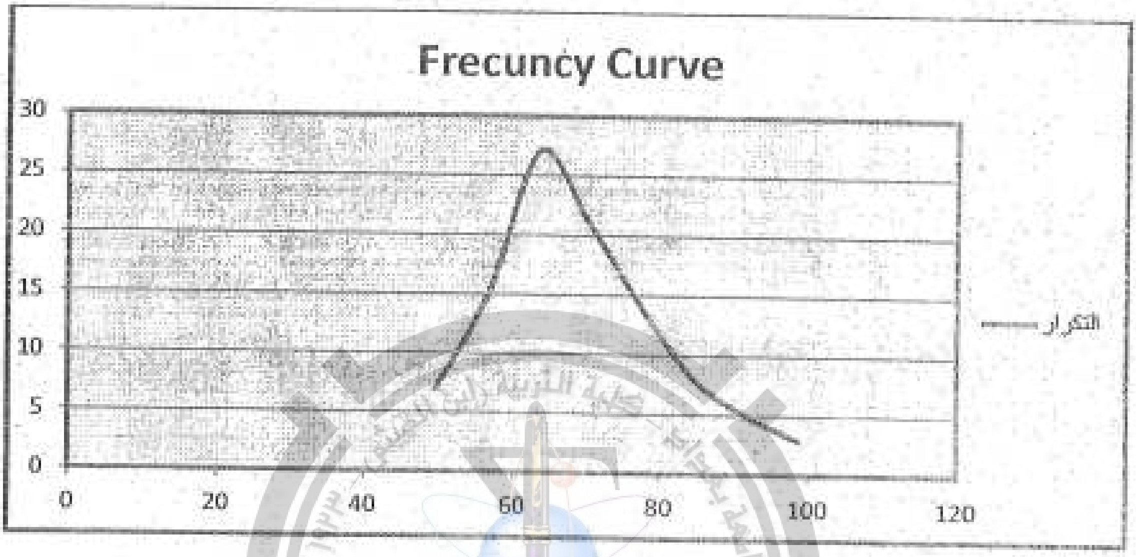
Freq.	7	15	27	21	14	8	5	3
Midpoint class	49.5	56.5	63.5	70.5	77.5	84.5	91.5	89.5



3- Frequency Curve : A frequency Curve are very similar to frequency Polygon, except instead of using line to join the points we use curves. This type of graph is only for continuous variables data.

Example ⑪: Draw a frequency Curve for the frequency table in example (4).

Freq.	7	15	27	21	14	8	5	3
Midpoint class	49.5	56.5	63.5	70.5	77.5	84.5	91.5	89.5



« Exercises»

1 – The following data are the members of 60 families, find:

*** The Frequency distribution table.**

*** The Ascending Cumulative Frequency distribution.**

***The Descending Cumulative Frequency distribution.**

***The Relative frequency distribution.**

***Graph the frequency polygon.**

3	13	3	4	2	2	7	5	6	8	20	6	11	14	15	18	9	10	12	15
16	11	7	8	5	6	9	7	15	3	18	4	5	17	6	12	11	9	8	10
7	2	4	6	9	5	8	6	10	8	9	5	4	7	8	14	20	7	10	12

2 - The following data are the average scores for 50 students, find:

*** The Frequency distribution table.**

*** The relative frequency distribution.**

*** The Ascending Cumulative Frequency distribution.**

***The Descending Cumulative Frequency distribution.**

***Graph the histogram frequency .**

***Graph the Curve for the frequency distribution .**

62.1	64.1	55	68.1	79	95	82.3	53	68	63.2	86	89	70.7	71
75.1	81.6	52	82.2	69	74	62.5	66	58.9	76.3	54	80	66.2	71.8
69.3	92.1	60	56	54	77	63.8	73	66.5	60.3	51	62	93	86.6
90	66.2	71	87	61	70	64.6	52	91.4	59.6	64	74	88.4	72.4

جامعة بغداد

كلية التربية للعلوم الصرفة / ابن الهيثم

منهج البحث العلمي

قسم الرياضيات

مدرس المادة

م.م . اريج صلاح محمد

Chapter Four // Measures of Central Tendency

1- Some rules in summation: The symbol (Σ) means to sum (add) the scores, and it is called sigma or summation of .

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

2- $$\sum_{i=1}^{n} xi^2 = x_1^2 + x_2^2 + \ldots\ldots + x_n^2$$

3- If (a) is a constant then:

$$\sum_{i=1}^{n} a = na$$

Proof:

$$\sum_{i=1}^{n} a = a + a + \ \ldots\ + a = na$$

4 – If (a) is a constant then :

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

Proof:

$$\sum_{i=1}^{n} a\, x_i = ax_1 + ax_2 + \cdots + ax_n$$

$$= a\,(x_1 + x_2 + \ldots\ldots + x_n)$$

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

5 – If x and y two variables :

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

Proof:

$$\sum_{i=1}^{n} (x_i \mp y_i) = (x_1 \mp y_1) + (x_2 \mp y_2) + \cdots + (x_n \mp y_n)$$

$$= (x_1 + x_2 + \cdots + x_n) \mp (y_1 + y_2 + \cdots + y_n)$$

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

6 -

$$\left(\sum_{i=1}^{n} x_i\right)^2 = (x_1 + x_2 + \cdots + x_n)^2$$

7-

$$\sum_{i=1}^{n} x_i\, y_i = x_1 y_1 + x_2 y_2 + \cdots + x_n\, y_n$$

8 –

$$\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right) = (x_1 + x_2 + \cdots + x_n)(y_1 + y_2 + \cdots + y_n)$$

9 –

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

10 –

$$\frac{1}{\sum x_i} = \frac{1}{x_1 + x_2 + \cdots + x_n}$$

11 –

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

12 –

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

13 –

$$\sum_{i=1}^{n} (x_i \mp a) = \sum_{i=1}^{n} x_i \mp n\,a$$

Proof :

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

$$= (x_1+x_2+\ldots\ldots+x_n) \mp (a+a+\ldots\ldots+a)$$

$$= \sum_{i=1}^{n} x_i \mp n\,a$$

Note:

1- $\sum_{i=1}^{n} x_i \, y_i \neq \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i$

2- $\sum_{i=1}^{n} {x_i}^2 \neq (\sum_{i=1}^{n} x_i)^2$

3- $\sum_{i=1}^{n} \frac{xi}{yi} \neq \frac{\sum_{i=1}^{n} xi}{\sum_{i=1}^{n} yi}$

Homework:

If x and y two variables where:

Xi = 2 , 6 , 3 , 1

Yi = 3 , 9 , 6 , 2

Compute :

a- $\sum_{i=1}^{n}(y_i - x_i)^2$

b- $\sum_{i=1}^{n}(x_i - 3)(y_i - 5)$

c- $\sum_{i=1}^{n} x_i \, y_i \neq \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i$

d- $\sum_{i=1}^{n} x_i \, y_i - \frac{(\sum xi)\,(\sum yi)}{n}$

e- $\sum_{i=1}^{n} x_i^2 \neq (\sum x_i)^2$

2- Some rules in Multiply (π) :

1- If $x_1, x_2, \ldots\ldots, x_n$ is a random variables then the multiples of these variables is:

$$\prod_{i=1}^{n} x_i = x_1 * x_2 * \ldots \ldots * x_n$$

2- If (a) is a constant then:

$$\prod_{i=1}^{n} a = a^n$$

Proof :

$$\prod_{i=1}^{n} a = a * a * \ldots . * a = a^n$$

3 –

$$\prod_{i=1}^{n} a\, x_i = a^n \prod_{i=1}^{n} x_i$$

Proof:

$$\prod_{i=1}^{n} a\, x_i = a\, x_1 * a x_2 * \ldots . * a\, x_n = a^n \prod_{i=1}^{n} x_i$$

4 -

$$\prod_{i=1}^{n} a\, x_i\, b\;\; y_i = \; a^n b^n \prod_{i=1}^{n} x_i \prod_{i=1}^{n} y_i$$

Proof :

$$\prod_{i=1}^{n} ax_i\, by_i$$

$$= (ax_1 * by_1) * (ax_2 * by_2) * \ldots .* (ax_n * by_n)$$

$$= (ab*ab*\ldots.. *ab) * (x_1y_1*x_2y_2*\ldots\ldots*x_ny_n)$$

$$= a^n b^n \prod_{i=1}^{n} x_i \prod_{i=1}^{n} y_i$$

5 –

$$\prod_{i=1}^{n} \frac{1}{x_i} = \frac{1}{\prod_{i=1}^{n} x_i}$$

6 –

$$\prod_{i=1}^{n} \frac{a}{x_i} = \frac{a^n}{\prod_{i=1}^{n} x_i}$$

7 -

$$\prod_{i=1}^{n} \sqrt{x_i} = \sqrt{x_1} * \sqrt{x_2} * \dots\dots * \sqrt{x_n}$$

$$= \sqrt{x_1 * x_2 * \dots . * x_n} = \sqrt{\prod_{i=1}^{n} x_i}$$

8 –

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

4 - Measures of Central Tendency :

A score that indicates where the center of the distribution tends to be located to describes most typical values, and tell us about the shape and nature of the distribution, it is depends on level of measurement. Or it is a value to represent the typical or " average " value in a data set.

4-1 The Arithmetic Mean:

The score located at the mathematical center of a distribution, it is used to summarize interval or ratio data in situations when the distribution is symmetrical and unimodal.

4-1-2 Compute the Arithmetic mean for ungrouped data:

The mean is the sum of all data values $x_1, x_2, \ldots\ldots\ldots, x_n$ divided by the number of values in the data set . The mean of a sample data set in denoted by $\overline{(X)}$ and the mean of population data set by the Greek letter μ.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad i = 1, 2, \ldots, n$$

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

$$\mu = \frac{\sum_{i=1}^{N} xi}{N}$$

Example ①: Find the Mean for the following data set.

400 , 380 , 450 , 350 , 520

Sol.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{400+380+450+350+520}{5} = 420$$

Example ②: The following data is a members of families, compute the average of the member of family.

3 , 4 , 7 , 8 , 10 , 9 , 2 , 5 , 6 , 9 , 7 , 5

Sol.

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{3+4+7+8+10+9+2+5+6+9+7+5}{12} = 6.25 \approx 6$$

The average is nearly to (6) persons.

4-1-3 Compute the Arithmetic mean for grouped data:

Estimating a Mean from a frequency table, given the frequency distribution of data set, we can make the best estimate of the mean for the data set.

1- Calculate the class midpoint for each data class. These will be our data values for calculating the mean.

2- Use the Frequency of the data class for each data class midpoint.

3-Calculate the Mean by the following formula:

$$\overline{x} = \frac{\sum_{i=1}^{n} f_i \, x_i}{\sum_{i=1}^{n} f_i} \quad \dots \dots$$

Where:

X_i : the midpoint for each class.

f_i : The frequency for each class.

Example ③: The following table gives the frequency distribution of the temperatures in a city for 99 days, find the average of the temperature.

Temperatures	1-	2-	3-	4-	5-	6-	7-	8-9
Number of days	4	8	12	16	20	25	6	8

Sol.

To find the mean we must compute the midpoint (x) for each class.

Midpoint (x) = (L.L + U.L) /2

Midpoint (x_1) = (1+2)/2 = 1.5

Midpoint (x_2) = (2+3) /2= 2.5 ……. etc.

classes	f_i	x_i	$f_i\ x_i$
1 –	4	1.5	6
2 –	8	2.5	20
3 –	12	3.5	42
4 –	16	4.5	72
5 –	20	5.5	110
6 –	25	6.5	162.5
7 –	6	7.5	45
8 – 9	8	8.5	68
sum	99		525.5

$$\overline{x} = \frac{\sum_{i=1}^{n} fi\ xi}{\sum_{i=1}^{n} fi} = \frac{525.5}{99} = 5.308$$

Example ④:

Find the arithmetic mean for the following frequency distribution.

classes	31-40	41-50	51-60	61-70	71-80	81-90	91-100
Freq.	1	2	5	15	25	20	12

Sol.

classes	f_i	Midpoint (x_i)	$f_i\ x_i$
31 - 40	1	35.5	35.5
41 – 50	2	45.5	91
51 – 60	5	55.5	277.5
61 – 70	15	65.5	982.5
71 – 80	25	75.5	1887.5
81 – 90	20	85.5	1710
91 – 100	12	95.5	1146
sum	80		6130

Where the Arithmetic mean is:

$$\overline{X} = \frac{\sum_{i=1}^{n} xi\ fi}{\sum_{i=1}^{n} fi} = \frac{6130}{80} = 76.62$$

4-1-4 Some properties of the mean:

1-

a) For ungrouped data:

$$\sum_{i=1}^{n}(x_i - \bar{x}) = 0$$

Proof:

$\sum_{i=1}^{n}(x_i - \bar{x}) = \sum_{i=1}^{n} x_i - n\bar{x}$ Since $\bar{x} = \frac{\sum_{i=1}^{n} xi}{n}$

$\Rightarrow \sum_{i=1}^{n} xi = n\bar{x}$

Now $\sum_{i=1}^{n}(x_i - \bar{x}) = \sum_{i=1}^{n} x_i - n\bar{x}$

$= n\bar{x} - n\bar{x} = 0$

b) For grouped data:

$$\sum_{i=1}^{n} fi(xi - \bar{x}) = 0$$

Proof:

$\sum_{i=1}^{n} f_i(x_i - \bar{x}) = \sum_{i=1}^{n} x_i f_i - \bar{x}\sum_{i=1}^{n} f_i$

Since $\bar{x} = \frac{\sum_{i=1}^{n} xifi}{\sum_{i=1}^{n} fi} \Rightarrow \sum_{i=1}^{n} x_i f_i = \bar{x}\sum_{i=1}^{n} f_i$

Now

$\sum_{i=1}^{n} f_i (x_i - \bar{x}) = \sum_{i=1}^{n} x_i f_i - \bar{x}\sum_{i=1}^{n} f_i$

$=$

$\bar{x}\sum_{i=1}^{n} f_i - \bar{x}\sum_{i=1}^{n} f_i = 0$

2 – If (K) is a constant then:

$$Y_i = k\,x_i \;\Rightarrow\; \bar{y} = k\,\bar{x}$$

Proof :

$Y_i = k\,x_i$ **(take summation both sides)**

$\sum_{i=1}^{n} y_i = k \sum_{i=1}^{n} x_i$ **(Divided by (n) to get the mean)**

$\frac{\sum_{i=1}^{n} yi}{n} = k\,\frac{\sum_{i=1}^{n} xi}{n} \;\Rightarrow\; \bar{y} = k\,\bar{x}$

3 - If (K) is a constant then:

$$Y_i = x_i + k \;\Rightarrow\; \bar{y} = \bar{x} + k$$

Proof:

$Y_i = x_i + k$ **(take summation both sides)**

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} (x_i + k)$$

$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} x_i + n\,k$ **(divided by (n) to get the mean)**

$$\frac{\sum_{i=1}^{n} yi}{n} = \frac{\sum_{i=1}^{n} xi}{n} + k \;\Rightarrow \bar{y} = \bar{x} + k$$

4-2 Weighted mean :

a) For ungrouped data:

If $x_1, x_2, \ldots\ldots, x_n$ is a sample of n observations, $w_1, w_2, \ldots\ldots, w_n$ is a weighted for these observations, then the weighted mean ($\overline{x}_w$) is:

$$\overline{x}_w = \frac{\sum_{i=1}^{n} xi\, wi}{\sum_{i=1}^{n} wi} = \frac{x1w1+x2w2+\cdots+xnwn}{w1+w2+\cdots+wn}$$

b) For grouped data:

We have to compute.....

1- The midpoint (x_i) for each class.

2- Multiply each midpoint with its frequency (f_i) and its weight (w_i).

3- Apply the following formula to find the weighted mean:

$$\overline{x}_w = \frac{\sum_{i=1}^{n} wi\, fi\, xi}{\sum_{i=1}^{n} wi\, fi} = \frac{w1f1x1+w2f2x2+\cdots+wnfnxn}{w1f1+w2f2+\cdots+wnfn}$$

Example ⑤: For the following data (students' scores x_i and the hours w_i for each lesson) , find the weighted mean for the student score.

(x_i) scores	62	80	75	88	84	86	90
(w_i)Hours	2	2	2	3	3	3	3

Sol.

$$\overline{x}_w = \frac{\sum_{i=1}^{n} xiwi}{\sum_{i=1}^{n} wi} = \frac{(62*2)+(80*2)+(75*2)+(88*3)+(84*3)+(86*3)+(90*3)}{2+2+2+3+3+3+3}$$

$$= \frac{1478}{18} = 82.11$$

Example ⑥:

For the following frequency distribution table, find the weighted mean.

classes	2 -<4	4 -<6	6-<8	8-< 10	10-< 12
(fi) Freq.	4	5	6	3	2
(wi) weight	6	5	6	4	4

Sol.

First we have to compute the midpoint (x_i) for each class then we had to compute the measurements as shown in the following table:

classes	Fi	W_i	X_i	$W_i f_i$	$X_i w_i f_i$
2-< 4	4	6	3	24	72
4-<6	5	5	5	25	125
6-<8	6	6	7	36	252
8-<10	3	4	9	12	108
10-<12	2	4	11	8	88
Sum	20			105	745

$$\overline{x}_w = \frac{\sum_{i=1}^{n} xi\, fi\, wi}{\sum_{i=1}^{n} fi\, wi} = \frac{745}{105} = 7.095$$

4 – 3 The Harmonic Mean:

a) For ungrouped data:

If $x_1, x_2, \ldots\ldots, x_n$ is a sample of n observations, then to compute the Harmonic Mean (H) we had to apply the following formula:

$$H = \frac{1}{\left(\sum_{i=1}^{n} \frac{1}{xi}\right)/n} = \frac{n}{\sum_{i=1}^{n} \frac{1}{xi}} \qquad ; x_i \neq 0$$

b) For grouped data:

Estimating a Harmonic Mean from a frequency table, given the frequency distribution of data set, where $x_1, x_2, \ldots\ldots, x_n$ are the midpoints and $f_1, f_2, \ldots\ldots, f_n$ are the frequencies, then the formula to compute the Harmonic Mean is:

$$H = \frac{\sum_{i=1}^{n} fi}{\sum_{i=1}^{n} \frac{fi}{xi}}$$

Where:

f_i / x_i : divide the frequency for each class by the midpoint for the class.

Example ⑦: For the following data, find the Harmonic Mean.

6 , 7 , 6 , 5 , 3 , 10 , 12

Sol.

$$H = \frac{n}{\sum_{i=1}^{n} \frac{1}{xi}} = \frac{7}{\frac{1}{12} + \frac{1}{10} + \frac{1}{3} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{6}} = 5.87$$

Example ⑧: For the following frequency table, find the Harmonic mean.

classes	50-	60-	70-	80-	90-	100-	110-< 120
(fi) freq.	8	10	16	14	10	5	5

Sol.

To find the Harmonic mean we must compute the measurement as shown in the following table:

classes	Fi	Xi	fi / xi
50 -< 60	8	55	0.145
60 -< 70	10	65	0.154
70 -< 80	16	75	0.213
80 -< 90	14	85	0.165
90 -< 100	10	95	0.105
100 -< 110	5	105	0.048
110 -< 120	5	115	0.043
sum	68		0.873

$$H = \frac{\sum_{i=1}^{n} fi}{\sum_{i=1}^{n} \frac{fi}{xi}} = \frac{68}{0.873} = 77.892$$

4 – 4 The Quadratic Mean :

This kind of Mean use in physics science, and it denoted by (Q).

a) For ungrouped data:

If $x_1, x_2, \ldots\ldots, x_n$ is a sample of n observations, then to compute the Quadratic Mean (Q) we had to apply the following formula:

$$Q = \sqrt[2]{\frac{\sum_{i=1}^{n} x_i^2}{n}}$$

b) For grouped data:

Estimating a Quadratic Mean from a frequency table, given the frequency distribution of data set, where $x_1, x_2, \ldots\ldots, x_n$ are the midpoints and $f_1, f_2, \ldots\ldots, f_n$ are the frequencies, then the formula to compute the **Quadratic** Mean is:

$$Q = \sqrt{\frac{\sum_{i=1}^{n} f_i x_i^2}{\sum_{i=1}^{n} f_i}}$$

Example ⑨: Find the Quadratic mean for the following data.

$$2 , 3 , 4 , 5 , 6$$

Sol.

$$\sum_{i=1}^{n} xi^2 = 2^2 + 3^2 + \cdots + 6^2 = 90$$

$$Q = \sqrt{\frac{\sum_{i=1}^{n} xi^2}{n}} = \sqrt{\frac{90}{5}} = 4.243$$

Example ⑩:The following frequency distribution represent members of patients ,compute the Quadratic mean for the age of patient.

Age classes	10-	20-	30-	40-	50-	60 -< 70
Number of patients (f_i)	2	3	4	10	36	14

Sol.

classes	f_i	X_i	X_i^2	$f_i x_i^2$
10-	2	15	225	450
20-	3	25	625	1875
30-	4	35	1225	4900
40-	10	45	2025	20250
50-	36	55	3025	108900
60 -< 70	14	65	4225	59150
sum	69			195525

$$Q = \sqrt{\frac{\sum_{i=1}^{n} fi\, xi^2}{\sum_{i=1}^{n} fi}} = \sqrt{\frac{195525}{69}} = 53.232$$

4 – 5 The Geometric Mean :

This central tendency is important in census survey, it denoted by (G).

a) **For ungrouped data:**

If $x_1, x_2, \ldots\ldots, x_n$ is a sample of n observations where $x_i > 0$, then to compute the Geometric Mean (G) we had to apply the following formula:

$$G = \sqrt[n]{\prod_{i=1}^{n} x_i} = \left(\prod_{i=1}^{n} x_i\right)^{1/n}$$

Also we can compute the Geometric mean by using the following formula:

$$\log G = [\log(x_1) + \log(x_2) + \ldots\ldots + \log(x_n)] / n$$

Example ⑪: For the following data, find the Geometric mean.

10 , 20 , 30 , 40 , 50 , 60

Sol.

$$G = \sqrt[n]{\prod_{i=1}^{n} x_i} = \left(\prod_{i=1}^{n} x_i\right)^{1/n}$$
$$= (10*20*30*40*50*60)^{1/6}$$
$$= (720000000)^{1/6} = 29.938$$

b) For grouped data:

Estimating a Geometric Mean from a frequency table, given the frequency distribution of data set, where $x_1, x_2, \ldots\ldots, x_n$ are the midpoints and $f_1, f_2, \ldots, f_n$ are the frequencies, then the formula to compute the Geometric Mean is:

$$G = \sqrt[\sum fi]{\prod_{i=1}^{n} x_i^{f_i}} = \left(\prod_{i=1}^{n} x_i^{f_i}\right)^{1/\sum f_i}$$

Or we can use the following formula .....

$$\log G = \frac{\sum f_i (\log x_i)}{\sum f_i} = \frac{f_1 \log x_1 + f_2 \log x_2 + \ldots\ldots + f_n \log x_n}{f_1 + f_2 + \ldots\ldots + f_n}$$

Example ⑫: Find the Geometric mean for the following frequency distribution table.

classes	10-	20-	30-	40-	50-	60-	70-< 80
Freq.	3	5	7	8	6	5	2

Sol.

First we had to compute the midpoint for each class.

classes	10-	20-	30-	40-	50-	60-	70-< 80
Freq.	3	5	7	8	6	5	2
(Xi) midpoint	15	25	35	45	55	65	75

$$G = \sqrt[\sum fi]{\prod_{i=1}^{n} {x_i}^{f_i}}$$

Where $\sum fi = 36$

$$G = \sqrt[36]{15^3 * 25^5 * 35^7 * 45^8 * 55^6 * 65^5 * 75^2}$$

$$G = 40.35$$

4 – 6 The Mode :

The Mode is the value that has the highest frequency in a data set.

a) Find the Mode for ungrouped data:

If $x_1, x_2, \ldots\ldots, x_n$ is a sample of n observations, then the Mode (Mo) is the value that has the highest frequency in a data set, if the data have exactly two Modes, the data are bimodal, if the data have more than two Modes the data are multimodal, and sometimes the data has no Mode .

Example ⑬: Compute the Mode of the following sample.

2, 3, 2, 4, 2, 5, 4, 4, 5, 4, 6, 8, 9, 4, 7, 3, 7, 6

Sol.

The value that has the highest frequency in the sample is the number (4).

$$\therefore Mo = 4$$

Example ⑭: Compute the Mode of the following sample.

2, 4, 3, 6, 8, 7, 10, 12, 5, 9, 13, 11

Sol.

The data have no Mode.

b) Find the Mode for grouped data:

i)) IF the Random Variable is discrete:

IF $x_1, x_2, \ldots\ldots, x_n$ are the midpoints for a classes of a frequency distribution and $f_1, f_2, \ldots\ldots, f_n$ are the frequencies,

then the Mode is the midpoint class for the largest frequency for the distribution.

Example ⑮: Compute the Mode for the following distribution frequency.

classes	60-74	75-89	90-104	105-119	120-134
fi	2	6	14	10	8

Sol.

The largest frequency is (14) for the third class.

$\therefore$ Mode = (90 + 104) / 2 = 97

ii)) IF the Random Variable is Continuous:

IF $x_1, x_2, \ldots\ldots, x_n$ are the midpoints for a classes of a frequency distribution, $f_1, f_2, \ldots\ldots, f_n$ are its frequencies, to find the Mode use the following formula:

$$Mo = L_1 + \left(\frac{d1}{d1+d2} \right) w$$

Where:

class Mode : is the class the highest frequency.

L_1 : is the lower limit (lower boundary) of class mode.

d_1 : is the difference between the frequency of class mode and the frequency of the class after the class mode.

d_2 : is the difference between the frequency of class mode and the frequency of the class before the class mode.

W : is the class width.

Example ⑯: Find the Mode for the following frequency distribution.

classes	60-	63-	66-	69-	72-75	sum
fi	5	18	42	27	8	100

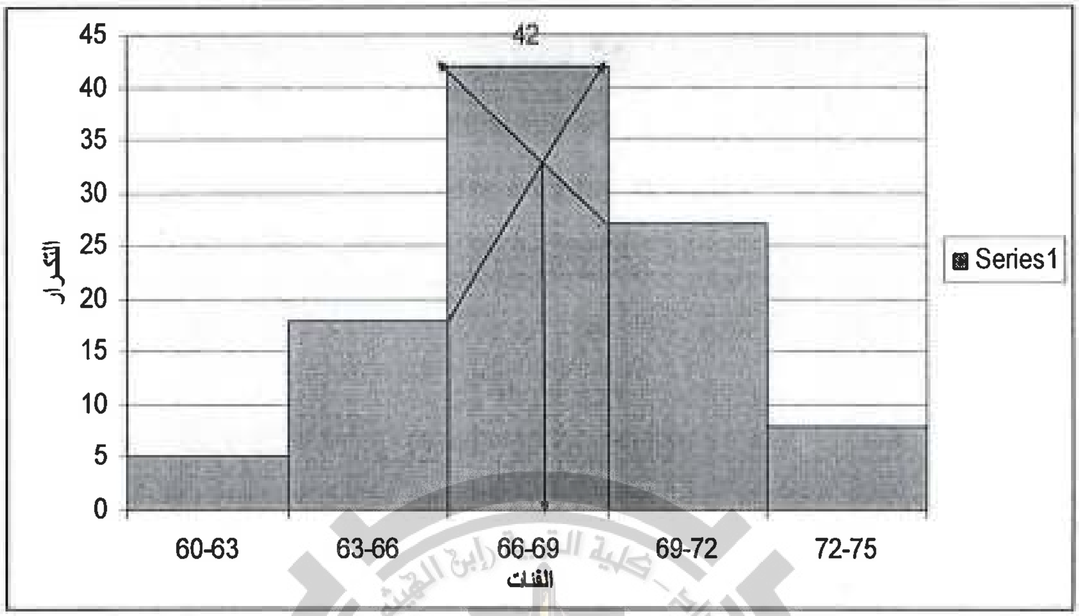
Sol.

The third class (66 – 69) is the class mode with highest frequency (42).

$L_1 = 66$ $\quad d_1 = 42\text{-}18 = 24$ $\quad d_2 = 42 - 27 = 15$ $\quad w = 3$

$$Mo = L_1 + \left(\frac{d1}{d1+d2}\right) w = 66 + \left(\frac{24}{24+15}\right) 3 = 67.8$$

Mode paragraph



4 – 7 The Median :

4 – 7 – 1 Compute the Median for ungrouped data:

The median of a data set is the value in the middle when the data items are arranged in ascending order.

i)) For odd number of observations:

The median is the middle value $\left[\frac{n+1}{2} \right]$

ii)) For even number of observations:

The median is the average of the middle two value

Example ⑰: Find the median for the following data.

55 , 62 , 53 , 70 , 68 , 65 , 63 , 79 , 80

Sol.

First we arranged the data in an ascending order

53 , 55 , 62 , 63 , 65 , 68 , 70 , 79 , 80

$$\textbf{median order} = \frac{n+1}{2} = \frac{10}{2} = 5$$

$\therefore$ the median is the fifth value = 65

Example ⑱: The following data is the age of 12 peoples, find the median for the people age.

20 , 22 , 19.5 , 26 , 24.5 , 27 , 28 , 29 , 18 , 20 , 23 , 25

Sol.

First we arranged the data in an ascending order

18 , 19.5 , 20 , 20 , 22 , 23 ,24.5 , 25 , 26 ,27 , 28 ,29

Median order = $\frac{n}{2} = \frac{12}{2} = 6$

Median order = $\frac{n}{2} + 1 = \frac{12}{2} + 1 = 7$

The median is the average of the two values whose sequence are (6) and (7):

$$\therefore \text{ Me} = \frac{24.5+23}{2} = 23.75$$

4 – 7 – 2 Compute the Median for grouped data:

i)) For the discrete random variable:

In this situation we had to compute:

1- The median order = $\frac{\sum_{i=1}^{n} f_i}{2}$

2- Compute the ascending cumulative frequency (FV).

3- Compare the median order with the ascending cumulative frequency:

$$F_k \leq \frac{\sum fi}{2} \leq F_{k+1}$$

Then the median is the midpoint for the median class that have (K+1) sequence.

Example ⑲: The following frequency distribution table is a sample of (80) families, compute the median of families members.

Number of persons	2-4	5-7	8-10	11-13	14-16	17-19	20-22
Numbers of families (fi)	6	9	12	20	14	11	8

Sol.

First we had to compute the ascending cumulative frequency.

Classes	2-4	5-7	8-10	11-13	14-16	17-19	20-22
(fi)	6	9	12	20	14	11	8
Upper limit classes	4	7	10	13	16	19	22
Cumulative freq. (FV)	6	15	27	47	61	72	80

Median order = $\frac{\sum_{i=1}^{m} fi}{2}$

$$= \frac{80}{2} = 40$$

Compare the median order with the cumulative frequency (FV), $27 \leq 40 \leq 47$, so the median class is the fourth class (11 – 13), then the median is the midpoint for the fourth class:

$$Me = \frac{11+13}{2} = 12$$

ii)) For the continuous random variable:

Assume that $f_1, f_2, \ldots\ldots, f_m$ is a frequencies for a distribution table, $F_1, F_2, \ldots\ldots, F_m$ is the cumulative frequency for the classes upper limits, if:

$$F_{k-1} < \frac{\sum fi}{2} < F_k$$

Where :

median order : $\frac{\sum fi}{2}$

Then to find the median we had to apply the following formula:

$$Me = L_k + \frac{h_k}{f_k}\left(\frac{\sum fi}{2} - F_{k-1}\right)$$

Where:

L_k : lower limit for median class.

f_k : median class frequency.

h_k : median class length.

F_{k-1} : previous ascending cumulative frequency for median class.

Example 20: Find the median for the following distribution table.

classes	100-	120-	140-	160-	180-	200-	220-< 240
Freq.	3	7	14	20	18	12	6

Median order $= \frac{\sum fi}{2} = \frac{80}{2} = 40$

classes	fi	Upper limit Classes	F.V
100-	3	120	3
120-	7	140	10
140-	14	160	24
160-	20	180	44
180-	18	200	62
200-	12	220	74
220-<240	6	240	80

Compare the median order with the ascending cumulative frequency $24 < 40 < 44$ then the median class is the fourth class (160 – 180) …..

$L_4 = 160$ $h_4 = 20$ $f_4 = 20$ $F_{4-1} = 24$

$$Me = L_k + \frac{h_k}{f_k}\left(\frac{\sum fi}{2} - F_{k-1}\right)$$

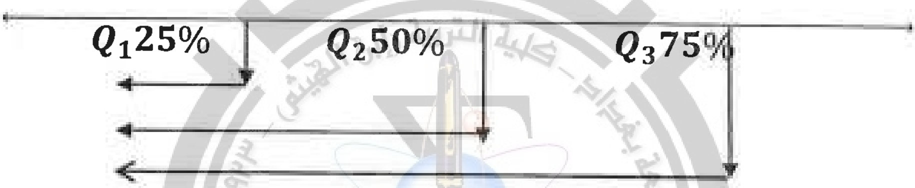
$$\therefore\ Me = 160 + \frac{20}{20}(40 - 24) = 160 + 16 = 176$$

4 - 8 Partition measures :

4 – 8 -1 The Quartiles :

Quartiles are specific percentiles .....

First Quartile = 25^{th} percentile
Second Quartile = 50^{th} percentile = Median
Third Quartile = 75^{th} percentile



i)) Compute the Quartile for ungrouped data:

If $x_1, x_2, \ldots\ldots\ldots, x_n$ is a sample of n observations arranged in ascending order, then to compute the Quartile:

Q_1 is the value that sequence = $\frac{n}{4}$

Q_2 is the value that sequence = $\frac{n}{2}$

Q_3 is the value that sequence = $\frac{3\,n}{4}$

Example 21 : Find the Quartiles for the following data.

2,7,3,5,8,6,10,12,9,11,5,1,6,13,16,14

Sol.

1- First we must arranged the data in an ascending order.

1, 2,3,5,5,6,6,7,8,9,10,11,12,13,14,16

2- The sequence of $Q_1 = \frac{n}{4} = \frac{16}{4} = 4$

$$\therefore Q_1 = 5$$

The sequence of $Q_2 = \frac{n}{2} = 16/2 = 8$

$$\therefore Q_2 = 7$$

The sequence of $Q_3 = \frac{3n}{4} = \frac{3*16}{4} = 12$

$$\therefore Q_3 = 11$$

ii)) Compute the Quartile for grouped data:

1- For discrete random variable :

Assume that $F_1 , F_2 , \ldots\ldots , F_m$ is an ascending cumulative frequency for frequency distribution with m classes, then if

$1 \leq i \leq k \leq j \leq m$ is a real numbers:

$F_{i-1} \leq \frac{n}{4} \leq F_i$ then Q_1 is the midpoint of class (i).

$F_{k-1} \leq \frac{n}{2} \leq F_k$ then Q_2 is the midpoint of class (k).

$F_{j-1} \leq \frac{3n}{4} \leq F_j$ then Q_3 is the midpoint of class (j).

Example (22) : The following frequency distribution table is a sample of 124 families, find the quartile for the distribution.

classes	2-4	5-7	8-10	11-13	14-16	17-19	20-22
fi	4	15	22	36	28	14	5

Sol.

First we had to find the ascending cumulative frequency....

classes	2-4	5-7	8-10	11-13	14-16	17-19	20-22
fi	4	15	22	36	28	14	5
F.V	4	19	41	77	105	119	124

The order of the first Quartile = $\frac{124}{4} = 31$

$\because$ $19 < 31 < 41$ then the value of Q_1 is the midpoint of first class

$\therefore$ $Q_1 = (8 + 10) / 2 = 9$

The order of the second Quartile = $\frac{124}{2} = 62$

$\because$ $41 < 62 < 77$ then the value of Q_2 is the midpoint of second class

$\therefore$ $Q_2 = (11 + 13) / 2 = 9$

The order of the third Quartile = $\frac{3*124}{4} = 93$

$\because$ $77 < 93 < 105$ then the value of Q_3 is the midpoint of third class

$\therefore$ $Q_3 = (14 + 16) /2 = 15$

2- For continuous random variable:

Assume that F_1 , F_2 , …….. , F_m is an ascending cumulative frequency for frequency distribution with m classes for a continuous random variable, then to compute the Quartile we must apply the following formula:

$$Q_j = L_j + \frac{h_j}{f_j}\left(\frac{n*j}{4} - F_{j-1}\right) \qquad , \quad j = 1,2,3$$

Where:

L_j : Lower limit for Quartile class.

h_j : Quartile class length.

f_j : Quartile class frequency.

Quartile class order = $\frac{n*j}{4}$, where $n = \sum_{i=1}^{m} f_i$

Example (23): The following frequency distribution table set the age for (772) students, find the Quartile.

classes	5-	6	7-	8-	9-	10-	11-	12-<13
Freq.	**16**	**120**	**131**	**145**	**122**	**115**	**101**	**22**

Sol.

First we had to compute the ascending cumulative frequency ....

classes	5 -	6 -	7 -	8 -	9 -	10 -	11 -	12-< 13
fi	16	120	131	145	122	115	101	22
F.V	16	136	267	412	534	649	750	772

First Quartile order = 772 /4 = 193

$\because$ 136 < 193 < 267

$\therefore$ First Quartile class is (7 – 8)

$$Q_j = L_j + h_j / f_j \,(n_j / 4 - F_{j-1})$$

$$\therefore \quad Q_1 = 7 + 1 / 131 \,(193 - 136) = 7.457$$

Second Quartile order = 772 /2 = 386

$\because$ 267 < 386 < 412

$\therefore$ Second Quartile class is (8 – 9)

$$\therefore \quad Q_2 = 8 + 1 / 145 \,(386 - 267) = 8.821$$

Third Quartile order = (3 * 772) / 4 = 579

$\because$ 534 < 579 < 649

$\therefore$ Third Quartile class is (10 - 11)

$$\therefore\ Q_3 = 10 + 1 / 115 (579 - 534) = 10.391$$

Example (24): For the following frequency distribution table find the Quartile.

classes	2 - < 4	4 - < 6	6 - < 8	8 - < 10	10 - < 12	12 -< 14
fi	3	5	8	11	4	9

Sol.

First we had to compute the ascending cumulative frequency ....

classes	2 - < 4	4 - < 6	6 - < 8	8 - < 10	10 - < 12	12 -< 14
fi	3	5	8	11	4	9
F.V	3	8	16	27	31	40

First Quartile order = 40 /4 = 10

$\because$ 8 < 10 < 16

$\therefore$ First Quartile class is (6 – < 8)

$$Q_j = L_j + h_j / f_j (n_j / 4 - F_{j-1})$$

$Q_1 = 6 + 2 / 8 (10 - 8) = 6.5$

Second Quartile order = 40 /2 = 20

$\because$ 16 < 20 < 27

$\therefore$ Second Quartile class is (8 – < 10)

$Q_2 = 8 + 2/11 (20 - 16) = 8.72$

Third Quartile order = (3 * 40) / 4 = 30

$\because$ 27 < 30 < 31

$\therefore$ Third Quartile class is (10 - < 12)

$Q_3 = 10 + 2 / 4 (30 - 27) = 11.5$

*Exercises *

Q1:Find the arithmetic mean and the Harmonic mean for the following data:

50 , 52 , 60 , 59 , 62 , 63 , 65 , 54 , 57

Q2:The following frequency distribution set for the age of a student's sample, find: 1 the student age average 2 The Harmonic mean 3 Graph the frequency polygon .

classes	6-	12-	15-	18-	24-	26-30
Freq.	30	55	80	90	66	14

Q3: For the following data, find the Quadratic mean and the Geometric mean.

10 , 15 , 18 , 22 , 19 , 14 , 11 , 21 , 20 , 19 , 13

Q4 : The following frequency distribution is a sample of students' scores , find: 1 Quadratic mean for the student score 2 Geometric mean for the student score.

classes	10-	20-	30-	40-	50-	60-	70-	80-	90-100
Freq.	2	3	6	8	24	20	26	14	10

Q5 : For the following data, find : 1 The Mode 2 The Median.

2 , 5 , 3 , 5 , 25 , 3.3 , 2.52 , 3.4 , 3.29 , 3.6

Q 6: For the following frequency distribution table, find:

① The Mode ② The Median.

classes	400-	420-	440-	460-	480-	500-520
Freq.	14	20	36	45	26	15

Q 7 : The following frequency distribution is for a some production factory by (ton), in such a day distributed by working machines and the numbers of working hours, find the average of the machine product.

Product classes	2 -	4 -	6 -	8 -	10- - 12
Number of working machine	4	5	6	3	2
Machine working hours	6	5	6	4	4

Chapter Five

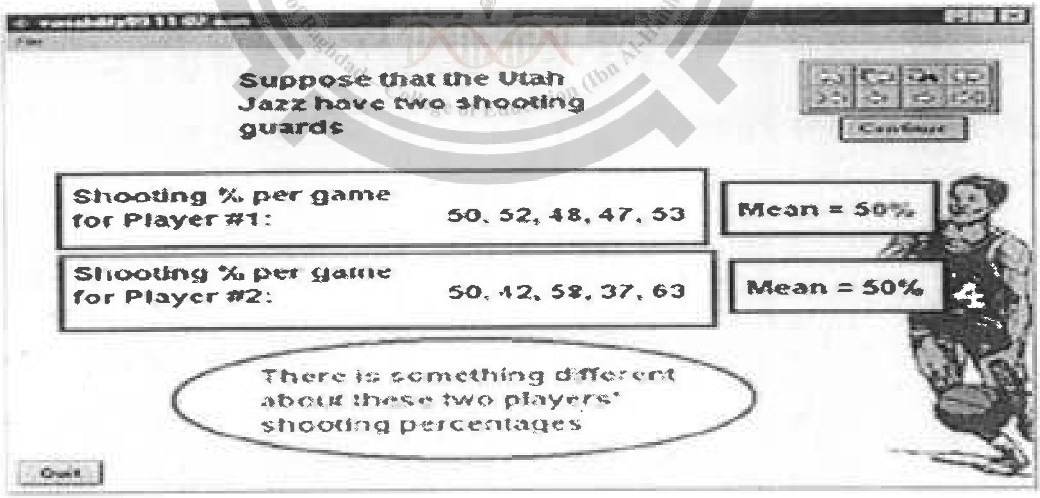
Measures of Variability

Chapter Five // Measures of variability :

It is often desirable to consider measures of variability (dispersion), as well as measures of location.

Variability addresses the issue of how spread out the scores are rather than what the center of numbers is. Numbers in a set can be tight and compact around their center or they can be dispersed and spread out a long way from their center. We're going to look at several descriptive statistics which will describe the degree of spread in a set of numbers around their center.

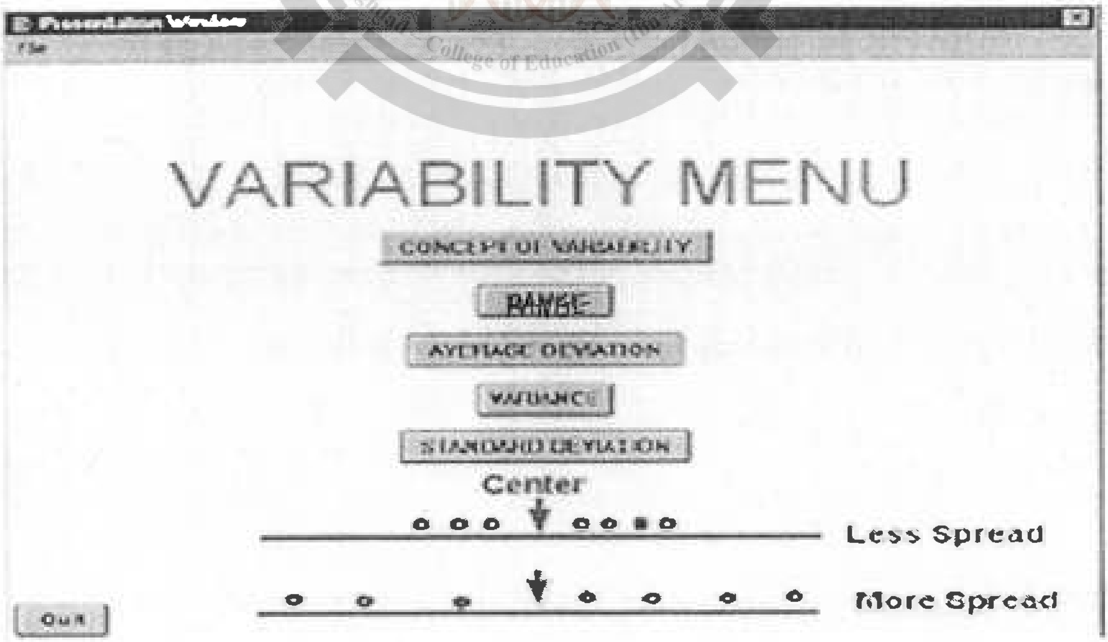
Example: Here is an example of two basketball players, with their shooting percentages.



We have player #1 and player #2. Let's say they each play in five games. In these games, player #1 shoots many shoots in each game. His averaging about 50 percent. Player #2 his

averaging about 50 percent also. Each one of these players is shooting 50% in five games. So for these five games they appear to be about the same on the average. Player #1 his shooting percentages have low variability, that is the scores don't vary much around their central tendency. They're not very dispersed, as opposed to player #2 who shows high variability. This second person has shooting percentages that are quite varied, or dispersed around their center.

This is a basic example of the idea of variability. This idea is different than the notion of central tendency. Central tendency statistics provide a set of measures that describe the center of a set of numbers. Statistics also has a set of other measures that we'll get into now, that describe how spread out things are from the center, or how diverse they are. Variability is measure of diversity, which is an interesting thing to know about.



If all values are the same the dispersion is zero.

If the values are homogenous and close to each other the dispersion is small.

If the value are so different the dispersion is large.

Measures of dispersion:

We'll concentrate on [illegible] measures of variability, the Range, average deviation, standard deviation, variance

5 – 1 The Range:

i) Find the range for ungrouped data:

The simplest measure of variability is the range. It is simply describes the difference between the largest and the smallest value.

$$R = X_L - X_S$$

Where:

R : The range.

X_L : the largest value.

X_S : the smallest value.

Example ①: Compute the Range for the following data.

$$2 , 5 , 3 , 8 , 7 , 10 , 9 , 12 , 15$$

Sol.

$$X_L = 15 \quad , \quad X_S = 2$$

$$R = 15 - 2 = 13$$

ii) Find the Range for grouped data:

The range is the difference between the upper real limit of the largest (maximum) X value and the lower real limit of the smallest (minimum) X value.

$$R = U.L_{\,L.C} - L.L_{\,F.C}$$

Where:

$U.L_{L.C}$: Upper limit last class.

$L.L_{F.C}$: Lower limit first class.

OR

Range = midpoint for the first class – midpoint for the last class

Example ②: Find The Range for the following frequency distribution table is the score of 40 student's.

classes	40-49	50-59	60-69	70-79	80-89	90-99
Freq.	2	9	15	11	2	1

Sol.

$$R = U.L_{L.C} - L.L_{F.C}$$

$$R = 99 - 40 = 59 \quad \text{degree}$$

OR

$$R = [\,(90+99)/2\,] - [\,(40+49)/2\,]$$

$$\therefore \quad R = 94.5 - 44.5 = 50 \quad \text{degree}$$

Properties of the range:

^ Simple to calculate.

^ Easy to understand.

^ It neglect all values in the center and depend on the extreme value, extreme value are dependent on sample size.

^ It is not based on all observations.

^ It is not amenable for further mathematic treatment.

Example ③: Find the Range for the hemoglobin levels of 50 peoples on the following frequency distribution table.

Hemoglobin level	fi	(Xi) midpoint
12.95 – 13.95	3	13.45
13.95 – 14.95	5	14.45
14.95 – 15.95	15	15.45
15.95 – 16.95	16	16.45
16.95 – 17.95	10	17.45
17.95 – 18.95	1	18.45

Sol.

$$\text{Range} = X_{max} - X_{min} = 18.45 - 13.45 = 5$$

5 – 2 Quartile deviation (Semi- Interquartile range):

The Quartile deviation (Q.D) is the mean between the third Quartile and the first Quartile for a sample of data, it's best from the Range because it is used 50% from the data:

$$Q.D = \frac{Q_3 - Q_1}{2}$$

Example ④: Back to example (24) P.60 , find the Quartile deviation.

Sol.

$$Q.D = \frac{Q_3 - Q_1}{2}$$

$$Q.D = \frac{10.391 - 7.433}{2} = 1.47$$

5 – 3 Mean deviation:

i)) Find the Mean deviation for ungrouped data:

The Mean deviation (M.D) is a measure of dispersion which calculate the distance between each data point and the mean, and then finds the average of these distances.

$$M.D = \frac{\sum_{i=1}^{n} |x_i - \overline{x}|}{n}$$

Example ⑤: For the following data, find the Mean Deviation.

$$X_i = 9 , 8 , 6 , 5 , 7$$

Sol.

$$\bar{x} = \frac{\sum_{i=1}^{n} xi}{n} = \frac{35}{5} = 7$$

X_i	$X_i - \bar{X}$	$\lvert x_i - \bar{x} \rvert$
9	2	2
8	1	1
6	-1	1
5	-2	2
7	0	0
		6

$$\text{M.D} = \frac{\sum_{i=1}^{n} \lvert xi - \bar{x} \rvert}{n}$$

$$= \frac{6}{5} = 1.2$$

ii)) Find the Mean deviation for grouped data:

If $x_1, x_2, \ldots\ldots, x_m$ is a midpoint for a frequency distribution of m classes and $f_1, f_2, \ldots\ldots, f_m$ is the frequencies', then to find the mean deviation we had to apply the following formula:

$$M.D = \frac{\sum_{i=1}^{n} f_i \, |x_i - \bar{x}|}{\sum_{i=1}^{n} f_i}$$

Example ⑥: The following frequency distribution table is the score of a group of students, compute the Mean deviation for the student's score.

classes	0-	10-	20-	30-	40-	50-	60-	70-	80-	90-100
Freq.	2	4	8	16	25	60	42	35	18	10

Sol.

classes	f_i	X_i	f_iX_i	$\lvert xi - \bar{x}\rvert$	$f_i\lvert xi - \bar{x}\rvert$
0-	2	5	10	54.5	109
10-	4	15	60	44.5	178
20-	8	25	200	34.5	276
30-	16	35	560	24.5	392
40-	25	45	1125	14.5	362.5
50-	60	55	3300	4.5	270
60-	42	65	2730	5.5	231
70-	35	75	2625	15.5	542.5
80-	18	85	1530	25.5	459
90-100	10	95	950	35.5	355
sum	220		13090		3175

$$\bar{x} = \frac{\sum_{i=1}^{n} fi\, xi}{\sum_{i=1}^{n} fi} = \frac{13090}{220} = 59.5$$

$$\text{M.D} = \frac{\sum_{i=1}^{n} fi\ |xi - \bar{x}|}{\sum_{i=1}^{n} fi} = \frac{3175}{220} = 14.4318$$

5 – 4 Standard deviation:

The standard deviation is the most popular and most important measure of variability. It takes into account all of the individuals in the distribution.

i)) Find the standard deviation for ungrouped data:

If X_1 , X_2 , …… ,X_n is a set of a sample (n) and ($\overline{x}$) is the mean, so to compute the standard deviation we had to apply the following formula:

$$S = \sqrt{\frac{\sum_{i=1}^{n}(xi - \overline{x})^2}{n-1}}$$

$$S = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 - 2\overline{x}\sum_{i=1}^{n} x_i + n\,\overline{x}^2}{n-1}}$$

$\because$ $\sum_{i=1}^{n} xi = n\,\overline{x}$, then ….

$$S = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 - n\,\overline{x}^2}{n-1}}$$

Or we could compute the standard deviation by the following formula:

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

Example ⑦: Compute the standard deviation for the following data:

7.1 , 2.5 , 2.5 , 5.4 , 8.3

Sol.

x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	X_i^2
7.1	1.94	3.7636	50.41
2.5	- 2.66	7.0756	6.25
2.5	- 2.66	7.0756	6.25
5.4	0.24	0.0576	29.16
8.3	3.14	9.8596	68.89
25.8	0.00	27.832	160.96

Where:

$n = 5$ $\sum_{i=1}^{n}(xi - \bar{x})^2 = 27.832$

$\sum x_i = 25.8$

$$\overline{x} = \frac{\sum_{i=1}^{n} xi}{n} = \frac{25.8}{5} = 5.16$$

$$S = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}} = \sqrt{\frac{27.832}{5-1}} = 2.6378$$

Example ⑧: The following data are the weighted for 10 students, Compute the standard deviation.

$$X_i = 56 , 62 , 69 , 71 , 68 , 65 , 63 , 72 , 68 , 56$$

Sol.

$$\overline{x} = \frac{\sum_{i=1}^{n} xi}{n} = 65 \quad KG$$

i)) Find the standard deviation for ungrouped data:

Assume that $X_1, X_2, \ldots\ldots, X_m$ is a midpoint for a frequency distribution for m classes, then to compute the standard deviation we had to apply the following formula:

$$s = \sqrt{\frac{\sum_{i=1}^{n} f_i\,(x_i - \overline{x})^2}{n-1}}$$

Where:

$$n = \sum_{i=1}^{n} f_i \quad , \quad \bar{x} = \frac{\sum_{i=1}^{n} fixi}{\sum_{i=1}^{n} fi}$$

$$S = \sqrt{\frac{1}{n-1}\left(\sum_{i=1}^{n} f_i\, x_i^2 - \frac{(\sum_{i=1}^{n} fi\, xi)^2}{n}\right)}$$

Example ⑨:Find the standard deviation for the following frequency distribution table.

classes	40-49	50-59	60-69	70-79	80-89	90-99
f_i	2	9	15	11	2	1

Sol.

First we had to compute the midpoint for each class.

Classes	40-49	50-59	60-69	70-79	80-89	90-99	sum
f_i	2	9	15	11	2	1	40
(midpoint)X_i	44.5	54.5	64.5	74.5	84.5	94.5	
$f_i x_i$	89	490.5	967.5	819.5	169	94.5	2630
$X_i - \overline{x}$	-21.25	-11.25	-1.25	8.75	18.75	28.75	
$(x_i - \overline{x})^2$	451.5625	126.5625	1.5625	76.56	351.56	826.56	
$fi\,(x_i - \overline{x})^2$	903.13	1139.06	23.44	842.16	703.12	826.56	4437.5

$$\overline{x} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} = \frac{2630}{40} = 65.75$$

$$s = \sqrt{\frac{\sum_{i=1}^{n} f_i\,(x_i - \overline{x})^2}{n-1}} = \sqrt{\frac{4437.5}{40-1}} = 10.667$$

Example ⑩: The following frequency distribution tables a score for a sample of students, find the standard deviation for the scores.

classes	f_i	(midpoint)X_i	f_ix_i	$f_i\, x_i^2$
0 –	2	5	10	50
10 –	4	15	60	900
20 –	8	25	200	5000
30 –	16	35	560	19600
40 –	25	45	1125	50625
50 –	60	55	3300	181500
60 –	42	65	2730	177450
70 –	35	75	2625	196875
80 –	18	85	1530	130050
90 -100	10	95	950	90250
المجموع	220		13090	852300

$$S = \sqrt{\frac{1}{n-1}\left(\sum_{i=1}^{n} f_i\, x_i^2 - \frac{(\sum_{i=1}^{n} fi\, xi\,)^2}{n}\right)}$$

$$S = \sqrt{\frac{1}{220-1}\left(852300 - \frac{(13090\,)^2}{220}\right)}$$

=18.313

Properties of standard deviation:

1- $S \geq 0$, that's mean the value of standard deviation is always positive and it's zero in special case.

2- If (a) is a constant and $y_i = a\, x_i$ then

$$S_y = |a|\, S_x$$

3- Adding a constant (a) to each score in the distribution will not change the standard deviation:

$$y_i = a \mp x_i \qquad \text{then} \qquad S_y = S_x$$

5 – 5 The Variance:

The variance is the average of the squared differences between each data value and the mean, and it's computed as follows:

i)) For ungrouped data:

If $X_1 , X_2 , \ldots\ldots ,X_n$ is a set of a sample (n) and ($\overline{x}$) is the mean, then to compute the variance:

$$S^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1} \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad ❶$$

$$S^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^{\,2} - \frac{(\sum_{i=1}^{n} x_i)^2}{n}\right) \quad \text{………… ❷}$$

Example ⑪: Find the variance for the following age students sample.

8 , 9 , 7 , 6 , 5

Sol.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{8+9+7+6+5}{5} = \frac{35}{5} = 7$$

x_i	$X_i - \bar{x}$	$(X_i - \bar{x})^2$
8	1	1
9	2	4
7	0	0
6	-1	1
5	-2	4
35	0	10

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{5-1}(10) = 2.5$$

ii)) Find the Variance For grouped data:

Assume that $X_1, X_2, \dots\dots, X_m$ is a midpoint for a frequency distribution for m classes, then to compute the variance we had to apply the following formula:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} f_i\,(x_i - \bar{x})^2 \quad \dots\dots\dots \text{③}$$

$$S^2 = \frac{1}{n-1} \left(\sum_{i=1}^{n} f_i\, x_i^{\,2} - \frac{(\sum_{i=1}^{n} f_i x_i)^2}{n} \right) \quad \dots\dots\dots \text{④}$$

Example ⑫: Compute the variance for the following students Score.

classes	40-49	50-59	60-69	70-79	80-89	90-99
fi	2	9	15	11	2	1

Sol.

classes	40-49	50-59	60-69	70-79	80-89	90-99
fr.	2	9	15	11	2	1
X_i	44.5	54.5	64.5	74.5	84.5	94.5
$f_i x_i$	89	490.5	967.5	819.5	169	94.5
$X - \overline{x}$	-21.25	-11.25	-1.25	8.75	18.75	28.75
$(x - \overline{x})^2$	451.5625	126.5625	1.5625	76.56	351.56	826.56
$fi\,(x - \overline{x})^2$	903.13	1139.06	23.44	842.16	703.12	826.56

Where the arithmetic mean is :

$$\overline{x} = \frac{\sum_{i=1}^{n} f_i\, x_i}{n} = \frac{2630}{40} = 65.75$$

And the Variance is

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} f_i\,(x_i - \overline{x})^2 = \frac{1}{40-1}(4437.5) = 113.78$$

5 – 6 Coefficient of Variation (C.V) :

* Measure of Relative Variation

*Always in percentage (%)

*Shows Variation Relative to Mean

Used to Compare two or more sets of data Measured in Different units

$$C.V = \frac{s}{\bar{x}} \cdot 100\%$$

Example: Suppose that, the average of a score for a sample of students in mathematic exam is 69 with standard deviation 19.3, and the average in statistic exam is 75 with standard deviation 25.5, in which exam the students has the GREATER VARIBILITY?

Sol.

The coefficient of variation quickly provides the answer:

$$C.V(\text{stat.}) = \frac{25.5}{75} \cdot 100 = 34\%$$

$$C.V(\text{math}) = \frac{19.3}{69} \cdot 100 = 28\ \%$$

From these calculations, it is immediately obvious that spread of scores in mathematics is greater than statistic.

*Exercises *

Q1:Find the: ① Range ② Mean deviation ③ Standard deviation ④ The Variance, for the following data:

6 , 3 , 5 , 5 , 9 , 4 , 6 , 7 , 1 , 2 , 4 , 8

Q2 : The following frequency distribution is the weighted of 50 students', find : ① the Range ② Mean deviation ③ Standard deviation ④ The Variance for the students weight.

Weighted classes	58 - 60	61 - 63	64 - 66	67 - 69	70 – 72	73 -75
Number of students	2	7	14	15	8	4

Q3 : The following table is a research on 5 persons , to find the weight or the measures of the lengths has the GREATER VARIBILITY?

Sequence of people	1	2	3	4	5
The weighted in (Kg)	65	67	65	59	69
The length In (cm)	158	165	155	162	164

Q4: The following frequency distribution is the age of a sample of people, find : ① the Range ② the Mean deviation ③the Standard deviation ④ The Variance.

classes	20-	30-	40-	50-	60-	70-	80-90
fr.	13	61	120	133	120	51	2

الفصل السادس

"إعداد مسودة البحث

إعداد أصل البحث"

الفصل السادس // إعـــداد مســـودة البحث – إعــداد أصل البحث ـ

6 - 1: إعــداد مســودة البحث:

تدوين البحث :

إن أسس كتابة البحث العلمي المتفق عليها تقضي بان يدون البحث مقسماً الى الأقسام الآتيه حسب تسلسلها:

عنوان البحث - الموجز (الملخص) – المقدمة – المواد والطرق – النتائج – المناقشة – الخاتمة (المختصر) - الشكر – قائمة المصادر – الملاحق (الجدول ، الخطوط البيانية ، الصور التوضيحية ، ....الخ).

وقبل ان يشرع الباحث في تدوين مسودة بحثه (وهي نسخته الأولى التي يكتبها بيده) عليه ان يراعي النقاط الآتيه ويتقيد بها جهد المستطاع، علماً ان هذه النقاط تتعلق بتدوين البحث كوحدة واحدة متكاملة:

١- تحديد المصادر.
٢- إختيار المجلة الناشرة.
٣- لغة البحث.
٤- عناوين البحث .
٥- الاقتباس .
٦- الهوامش .
٧- الأسماء العلمية .
٨- الأرقام .
٩- الرموز والمصطلحات .

١- تحديد المصادر: على الباحث ان يحدد المصادر التي سيستعرضها في مقدمة البحث ويقارن بها نتائج بحثه في قسم ((المناقشة))، وان اغلب تلك المصادر هي بحوث منشورة تتناول موضوع البحث نفسه أو مواضيع قريبه منه، وقد تكون تلك المصادر نشرات مستقلة أو كتب أو مطبوعات أخرى.

٢- إختيار المجلة الناشرة: قبل ان يدون الباحث مسودة بحثة عليه ان يحدد المجلة التي سيختارها لنشر ذلك البحث ، ليستطيع تدوين بحثه وفقا لتعليمات المجلة المختارة، وبذلك يوفر كثيرا من الجهد والوقت في تعدبل طريقة كتابة البحث فيما بعد لتطابق تعليمات المجلة، فأسس كتابة البحث واحدة لكن التفصيلات تختلف من مجلة لأخرى.

٣- لغة البحث : من المعتاد ان تدون البحوث وتنشر بإحدى اللغات الأجنبية الحيه كالانكليزية مثلاً، وهي اللغة الاجنبية الأكثر إستخداما في العراق وأقطار عديدة أخرى في مجال النشر العلمي.

٤- عناوين البحث : يشتمل البحث على عدة عناوين هي عنوان البحث أو العنوان الرئيسي، وعناوين فرعيه تمثل أقسام البحث (الموجز ، المقدمة ، .... الخ) وهذه العناوين الفرعية إما ان تكون وسطية أو جانبية ، وفي الحالة الأخيرة اما ان توضع بعد العنوان الجانبي علامة شارحه (:) أو فاصلة (-) أو نقطة (.) أو لايوضع شئ إطلاقاً.

٥- الأقتباس : يضطر الباحث احياناً لاقتباس سطور أو فقرات منشورة لباحث آخر، تأييدا لوجهة نظره في موضوع معين من بحثه أو توضيحاً لبعض الجوانب التي يجدها واضحة في ذلك البحث.

٦- الهوامش : لاتستعمل الهوامش في البحوث العلمية إلا عند الضرورة، وتوضع الارقام (١، ٢ ، ٣ ،....... الخ) فوق أواخر الكلمات أو الجمل التي يريد الباحث كتابة الهوامش لها، ثم تدوين الارقام نفسها بالتسلسل في ورقة مستقلة تلحق بآخر البحث.

٧- الأسماء العلمية : كل نوع من الكائنات الحيه له أسم علمي باللغة اللاتينيه عادة، متفق عليه علمياً وهو الأسم الذي يجب ان يستخدمه الباحث في كتابة بحثه.
يتألف الأسم العلمي من كلمتين هما أسم الجنس وأسم النوع يليها أسم المؤلف كاملاً أو مختصراً، يبدأ أسم الجنس دائما بحرف استهلالي كبير Capital ، بينما يبدأ أسم النوع دائما بأسم صغير، فالاسم العلمي للفيل الافريقي Loxodonta africana (نلاحظ ان أسم النوع هنا مشتق من أسم علم هو أفريقيا لكنه لايكتب Africana بل africana) أما أسم المؤلف الذي يلي أسم النوع فيطبع بحروف رومانية اعتيادية.

٨- الأرقام :

١- في البحوث المدونه باحدى اللغات الأوربية ، لاتبدأ الجملة برقم أبدا ، بل يكتب الرقم بالحروف، مثال ذلك لانكتب :

((12 chrysomelid species have been recorded ......))

بل يكتب :

((Twelve chrysomelid species ......))

واذا بدات الجملة باحدى الاعداد المحصورة بين 21 و 99 فان فاصلة (-) توضع بين الكلمتين الدالتين على العدد ، مثال ذلك:

eighty-six

Twenty – one , eighty-six

٢- الاعداد من (1-9) تكتب عادة بالحروف أيا كان موقعها ومازاد عن ذلك فيكتب كأرقام ، مثال ذلك:

((واتضح ان للحشرة أربعة أجيال في السنة))

((بلغ معدل عدد البيض للانثى الواحدة 38 بيضة))

أما اذا كانت تمثل وحدات قياس ، أو دونت في جدول أو خط بياني ، فانها تكتب كأرقام مهما كان مقدارها مثال ذلك : 20 غم ، 7 ملم ، 100 سم^3 ، ...... الخ.

٣- اذا كان الرقم يمثل وحدات تمثل بدورها عدة وحدات قياس معين، فانها تكتب بالحروف وتبين وحدة القياس بالارقام بعد ان يفصل بينها وبين الرقم بفاصلة (-).

٤- يوحد عدد المراتب العشرية في الكسور اينما وردت في سياق البحث وفي الجدول.

٥- العلامة (/) لاتعتبر فارزة كسر عشرية (كما اعتاد بعض الباحثين استخدامها في بحوثهم المدونه بالعربية) بل تعني نسبة أو علامة قسمة.

٦- من الافضل استخدام الوحدات العشرية (ملم ، غم ، كغم ، .... الخ) في متن البحث وجداوله بدلا من الوحدات الانكليزية (أنج ، قدم ، فهرنهايت ،..... الخ) .

٧- اذا كانت الارقام تؤلف جزءا من الجملة فلا توضع بين قوسين ، مثال ذلك لاتكتب :

((وكانت درجة الحرارة السائدة (22- 24 °م)))

بل تكتب ((وكانت درجة الحرارة السائدة 22- 24 °م))

٩- الرموز والمصطلحات : هناك بعض الرموز المستخدمة في الكتابة العلمية باللغات الاجنبية، وهي في حقيقتها مختصرات لكلمات لاتينيه أو يونانية أو انكليزية أو غيرها، تدل على معاني محددة ومن المستحسن للباحث ان يكون ملما ببعضها ليستطيع قراءة بحوث الآخرين قراءة صحيحة وليدون بحوثه الخاصة على أفضل وجه.

6 – 2 أقســـــام البحث :

بعد استعراض النقاط التسع المتعلقة بتدوين البحث وتوضيحها، نبين الآن أقسام البحث :

١- عنوان البحث The Title :

يتضمن عنوان البحث نص العنوان وأسم الباحث (أو أسماء الباحثين) وعنوان المؤسسة التي اجري فيها البحث.

نص العنوان : يجب ان يتضمن العنوان الاسم العلمي للكائن الحي موضوع البحث ومرتبته التصنيفيه (العائلة ، الرتبة ، .... الخ) فالاسم العلمي يكتب كاملا في نص العنوان ، يليه أسم مؤلفه كاملا أو مختصرا.

أسم الباحث : يدون أسم الباحث أو أسماء الباحثين تحت نص العنوان مباشرة ، وتكون الاسماء خالية من ذكر الشهادات او الالقاب العلمية او الاجتماعية مثل (الدكتور ، الاستاذ ، السيد ،......) ومع ذلك فهناك بعض المجلات الاجنبية التي تضع كلمة (السيدة .Mrs) أو (الآنسة Miss) بين قوسين بعد أسم الباحثة مباشرة للدلالة على جنسها وعلى وضعها الاجتماعي.

عنوان المؤسسة : يكتب عنوان المؤسسة التي اجري فيها البحث تحت أسم الباحث (أو أسماء الباحثين) مباشرة ، ومع ذلك هنالك بعض المجلات تضع عنوان المؤسسة فوق نص عنوان البحث، وبعضها الاخر يضعها كهوامش أسفل الصفحة.

خص) Abstract :

ـوجز عادة نصف صفحة من صفحات المجلة الناشرة لذا فانه يجب ان يكون مركزا في تقديم محتواه.

٣- المقدمـــة Introduction :

مقدمة البحث مدخل لصلب البحث والتعريف به، لذا فان الباحث يشير في المقدمة الى الكائن الحي الذي هو موضوع البحث، ثم يشير الى بعض المصادر التي تذكر أهميته وانتشاره في العالم.

٤- المواد والطرق Materials and methods :

في هذا القسم يذكر الباحث بايجاز شديد المواد التي استخدمها في اجراء بحثه، وطرق استخدامها والمكان والزمان اللذين استخدم فيها تلك المواد.

٥- النتائج Results :

يدون الباحث في هذا القسم نتائج بحثه التي حصل عليها من غير ان يعلق عليها بشئ أو يقارنها بنتائج بحوث أخرى مماثله، لأن الشرح والتعليق والمقارنة لها مكان آخر في البحث هو قسم ((المناقشة)).

٦- المناقشة Discussion :

في هذا القسم من البحث يقارن الباحث نتائج بحثه بنتائج باحثين آخرين بحثوا الموضوع نفسه أو موضوعا قريبا منه ، داخل العراق أو خارجه، ويقوّم نتائجه على ضوء تلك المقارنه، فأما ان تتفق معها أو تخالفها.

٧- الخاتمة Conclusion :

يدون الباحث في هذا القسم النتائج النهائية والمعلومات الجديدة التي استخلصها من بحثه ،واذا كان البحث تطبيقيا له ان يدونها كتوصيات مكتوبة بشكل جمل قصيرة متتابعة، أو فقرات مرقمة بالتسلسل.

٨- الشكر Acknowledgement :

في هذا القسم يشكر الباحث الاشخاص أو المؤسسات الذين ساعدوه بصورة أو بأخرى في البحث.

٩- قائمة المصادر List of references :

يدرج الباحث في آخر بحثه قائمة بالمصادر التي رجع اليها فعلا واشار اليها في سياق بحثه، ولايجوز له إدراج مصدر لم يرجع اليه فعلا وان كان يحوي معلومات تتعلق بموضوع بحثه.

6 - 3 : إعـداد أصـل البحث

بعد انتهاء الباحث من تدوين مسودة بحثه بشكلها النهائي، عليه ان يعد أصل البحث manuscript وهو نسخته المطبوعة بالطابعة والتي سترسل للنشر، إذ لايجوز إرسال أصل البحث للنشر مكتوبا باليد. وقبل الشروع في طباعة أصل البحث على الباحث ان يراعي النقاط الآتيه :

١- أختيار الورق : يكون الورق المستخدم عادة في طباعة أصل البحث ورق A4 أبيض مالم تنص التعليمات على غير ذلك.

٢- طباعة أقسام البحث : من الأفضل ان يطبع كل قسم من أقسام البحث – التي ذكرت سابقا – على ورقة منفصلة .وترقم الاوراق في الزاوية العليا اليسرى من كل ورقة (اذا كان البحث بالعربية) أو الزاوية العليا اليمنى منها (اذا كان البحث بلغة أوربية)، تبعا لتسلسل هذه الاقسام.

٣- طريقة الطباعة : يطبع نص البحث باقسامه المختلفة على وجه واحد من الورقة.

٤- انواع الحروف وعلاماتها : في البحوث والمقالات والكتب المنشورة باللغات الاوربية (كالانكليزية مثلا) تستخدم الحروف ذات الاستقامة العمودية ويكون الحرف باللغة الانكليزية Times New Romans وبالعربية يستخدم Arabic Transparent أو غير ذلك حسب التعليمات المنصوص عليها.

٥- المعادلات الرياضية والكيميائية : يخصص للمعادلات الرياضية والكيميائي ونحوها، فراغ كاف حولها يميزها عما فوقها وتحتها من نص البحث. ومن الافضل ان تكتب المعادلة بمستوى سطر واحد، تسهيلا لعملية الطباعة. مثال ذلك :

بدلا من : $X = \frac{\pi}{2}$ نكتب ⇦ $X = {}^{\pi}/_{2}$

٦- ملاحق البحث :

أ- الجداول : تطبع الجداول على ورقة مستقلة ويدون رقم الجدول فوقه مباشرة.
ب- الخطوط البيانية والمدرجات التكرارية.
ت- الصور التوضيحية .

6 -4 إرســال أصل البحث للنشــر

بعد ان ينتهي الباحث من طباعة بحثه وملحقاته ، يكون أصل هذا البحث جاهزا للارسال لنشره في إحدى المجلات العلمية المتخصصة.

الفصل السابع

Samples العينات

الفصل الســـابع // العينـــــات Samples

مقدمـــــة :

تقسـم العينات الى نوعين ، الأول هو العينات الاحتمالية Probabilistic Samples وهو النوع الذي يعتمد على نظرية الاحتمالات، أي ان عملية الاختيار تجري بحيث يكون لكل مفردة في المجتمع احتمال معلوم، وان احتمال ظهور تلك المفردة في العينة غير مساوٍ للصفر، وهذا الاحتمال يمكن ان يكون متساوياً لكافة المفردات كما يحصل ذلك في العينة العشوائية البسيطة، التي هي أحد أنواع العينات الاحتمالية، أو يكون غير متســاوٍ كما في الأنواع الأخرى من العينات الأحتمالية، كالطبقية ، المنتظمة، العنقودية ... الخ.

أمـــا النوع الثاني فهي العينات غير الاحتمالية، وفي هذا النوع يتم اختيار مفردات العينة عمدياً ومن أمثلتها العينة العمدية والعينة الحصصية وعينات المجتمعات المتحركة. ان الشائع في الاستخدام العملي هي العينات الاحتمالية نظرا لما تتمتع به من دقة واعتمادية وعليه فإننا سنركز على هذا النوع.

7 – 1 العينة معناها وأسباب اختيارها Sample and Reasons for Sampling

إننا نعيش في عصر يصعب الحكم فيه على صحة الأشياء بدون الأدلة الموضوعية والبحث الواسع، عصر ملئ بالمشاكل العلمية المختلفة والظواهر المتعددة. ولدراسة أية مشكلة علمية نحتاج الى جمع كل مايتعلق بتلك المشكلة من معلومات. وتسمى مجموعة العناصر المتعلقة بتلك المشكلة المجتمع الإحصائي. من هنا تبدأ دراسة إحصائية بالبيانات الخام (raw data) المتوفرة عن الدراسة والتي يتم جمعها بإحدى الطرق الآتية:

1- طريقة المسح الشامل (Census) : وفيها تجمع البيانات من جميع أفراد المجتمع الإحصائي، فمثلا إذا أردنا القيام بتعداد للسكان في بلد قليل السكان كالأردن مثلا – نقوم بعد أفراد جميع العائلات في المجتمع الاردني ، واذا أردنا التعرف على مستوى الطلاب في جامعة بغداد في قسم الرياضيات نقوم برصد علامات جميع هؤلاء الطلاب في مادة الرياضيات المتوفرة لدى الكليات.

2- طريقة العينات (Sampling Method) : هناك عدة حالات يتعذر فيها المسح الشامل وعندها نلجأ الى دراسة جزء من المجتمع الاحصائي يسمى العينة Sample وحجم العينة Sample size هوعدد عناصرها .

7 – 2 أنواع العينات الاحتمالية

7 – 3 العينة العشوائية البسيطة Simple Random Sample :

ان مســـوح العينة تتعامل مع عينة تتعامل مع عينة مختاره من مجتمع فيه عدد محدد من الوحدات المختلفة ولنقل N ، وإذا كان بالامكان التمييز بين واحدة وأخرى من هذه الوحدات، فان عدد العينات المختلفة من حجم (n) التي يمكن سحبها من المجتمع الذي عدد وحداته (N) يمكن ايجادها باستخدام قاعدة التوافيق الآتية:

$$C_n^N = \frac{N!}{n!(N-n)!}$$

فلو كانت لدينا مجتمع يتكون من الوحدات (A, B, C, D) أي ان N = 4 ، فان هناك ســت عينات مختلفة من حجم n = 2 ممكن سحبها وكالآتي:

$$C_2^4 = \frac{4!}{2!(4-2)!} = \frac{4*3*2*1}{2*1\,(2*1)} = 6$$

وإذا أردنا ترجمة ذلك العدد الى صيغة الحروف فان العينات الممكنة والبالغ عددها (6) هي كالآتي:

AB , AC , AD , BC , BD , CD.

ويمكن ملاحظة ان (نفس المفردة) لاتظهر مرتين في عينة واحدة من العينات الممكنة .

إن العينة العشوائية البسيطة Simple Random Sample هي طريقة لاختيار (n) من الوحدات من مجتمع حجمه (N) من الوحدات بحيث يكون لكل مفردة من المفردات المختارة نفس الاحتمال في الظهور في العينة.

تعتبر هذه الطريقة من أسهل طرق المعاينة الاحتمالية، ولكن ليست أكثرها استخداما، بل انها أقل أنواع العينات استخداما في الحياة العملية، بالرغم من سهولتها ودقتها عندما يكون المجتمع متجانساً وفي حالة توفر التجانس، فان الأمر يغنينا عن اللجوء الى إستخدام أنواع اخرى من العينات.

تصنف أساليب المعاينه من حيث طريقة التعامل مع المفردات التي تسحب من العينه الى نوعين:

1. المعاينه بدون إعاده : أي ان المفردة التي تختار ضمن العينه في أي مرحلة من مراحلها، لاتعاد الى المجتمع كي لاتعطي فرصة جديدة في الظهور ثانيه في العينه.

2. المعاينه مع الإعادة : أي ان المفردة التي تختار ضمن العينه تعطي فرصة أخرى للظهور ثانية في العينه ضمن مراحلها المتعاقبة، وهذا النوع لايستخدم إلا لظروف وحالات خاصة وهو نادر الاستخدام.

7 – 4 العينة العشوائية الطبقية Stratified Random Sample :

أشرنا سابقاً الى ان استخدام العينة العشوائية البسيطة يعتمد على حقيقة ان المجتمع متجانس، أي ان مفرداته متجانسة من حيث الخاصية (أو الخواص) التي يراد معاينتها، وان عدم توفر التجانس سيفضي الى مقدرات لاتمثل خصائص المجتمع بشكل واقعي، لذا يتم استخدام الاسلوب الطبقي (أي تقسيم المجتمع الى طبقات (أجزاء)) بحيث تكون مفردات كل جزء أو طبقة متجانسة فيما بينها ومختلفة عن الطبقات الأخرى، ويتم التعامل مع كل طبقة وكأنها مجتمع مستقل تسحب منه عينه عشوائية ذات حجم معين.

في ضوء ماأشرنا إليه في إعلاه فان حجم المجتمع البالغ عدد مفرداته (N) سيمثل مجموع عدد المفردات في الطبقات كافه وان عدد المفردات في كل طبقة سنرمز له (N_h) .

حيث ان $h = 1 , 2 , ... , L$ ، أي ان عدد الطبقات في المجتمع = L وان

$$N = \sum_{h=1}^{L} N_h = N_1 + N_2 + \dots\dots + N_L$$

وان $N_1 , N_2 , ... , N_L$ تسمى طبقات (Strata) وان الجزء الواحد يسمى طبقة (Stratum) ، أما حجم العينه التي نسحبها هو (n) أي ان :

n = عدد المفردات في العينه

حجم العينه الكلي $n = n_1+n_2+ ...... +n_L$

7 – 5 العينه المنتظمة Systematic Sample :

تشير تسمية هذا النوع من العينات الى انه يتبع أسلوبا منتظما لاختيار وحدات المجتمع. فاذا كان لدينا مجتمع يتكون من (N) من الوحدات ورغبنا ان نأخذ عينه حجمها (n) من هذا المجتمع فان الاسلوب المنتظم يكون بالشكل الاتي:

1- نستخرج قيمة (K) والتي تساوي ($\frac{N}{n}$) وبذلك نقسم المجتمع الذي حجمه (N) الى مجاميع عددها (K) وكل مجموعة تضم (n) من المفردات.

2- نختار وحده واحدة بشكل عشوائي من أول مجموعة.

3- بعد ان يتم تحديد الوحدة الأولى يمكن ان تؤخذ الوحدات المتبقية، حيث يكون تسلسل الوحدة الثانية بعد (K) من الوحدات (ابتداءا من الوحده المختاره من المجموعة الأولى) فاذا كان لدينا مجتمع حجمه (N = 200) وأردنا أختيار عينه حجمها n = 10 فان قيمة K هي :

$$K = \frac{N}{n} \quad \Longrightarrow \quad k = \frac{200}{10} = 20$$

نقوم بأختيار وحدة واحدة عشوائيا من أول (20) وحدة، ولنقل إنها كانت الوحدة رقم (6) بهذا فان الوحدة الثانية ستكون (26) والثالثة (46) ثم (66) ، (86) ، (106) ، (126) ،(146) ، (166) ، (186) وبهذا تكون العينه المطلوبه التي حجمها n = 10 قد تحققت.

تستخدم العينه المنتظمة في المجتمعات الكبيرة وفي المشاهدات الميدانيه، علماً ان العينه المنتظمة أكثر تمثيلاً من العينه العشوائية البسيطة حيث يجري تقسيم المجتمع الى طبقات عددها (K) وكل طبقة حجمها (n) يتم اختيار واحدة منها، وكذلك ان العينه المنتظمة تتوزع على المجتمع توزيعاً شاملاً أكثر مما يحدث في العينة العشوائية البسيطة.

<u>مثال :</u> نفرض لدينا مجتمع حجمه (N = 20) ورغبنا بسحب عينه حجمها (n = 5) عليه فان :

$$K = \frac{N}{n} = \frac{20}{5} = 4$$

أي ان عدد العينات الممكنه هي (4) وكما يأتي :

العينه الأولى	العينه الثانية	العينه الثالثة	العينه الرابعة
1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16
17	18	19	20

تمارين ٣-٥

س١: مجتمع مكون من (50) طالب ، ما هي عدد العينات العشوائية البسيطة الممكن سحبها من هذا المجتمع بحجم ١٠ طلاب .

س٢: ... من مناطق مدينة بغداد تحتوي على (40) مكان ، ما هي عدد العينات العشوائية المنتظمة الممكن سحبها من هذا المجتمع، علماً ان حجم العينة المراد سحبها (8) مكان .

جامعة بغداد

كلية التربية للعلوم الصرفة / ابن الهيثم

منهج بحث

قسم الرياضيات

المرحلة الثانية

نظرية الاحتمال Probability Theory

1-8 المقدمة :

تعد نظرية الاحتمال، أو قوانين المصادفة Law of chances ، من النظريات الرياضية المهمة التي تتخصص في دراسة الحوادث والمتغيرات والظواهر التي تتميز بعدم تأكد حدوثها.

ولنظرية الاحتمال تاريخ حافل يبدأ من منتصف القرن السابع عشر عندما أقترح النبيل الفرنسي Chevalier De Mere مسألة معينة في لعب الورق على العالم الفرنسي باسكال Blaise Pascal ومن خلال مراسلة باسكال الرياضي الفرنسي فرمات Pierre De Fermat وضعت الاسس الاولى لنظرية الاحتمال.

في حياتنا اليومية نستعمل بكثرة كلمة إحتمال، فيقول طالب إحتمال ان انجح في الامتحان 60 % ، ويقول شخص احتمال سقوط الامطار غدا 20% ، ويقول الطبيب احتمال ان تنجح العملية الجراحية 97% ، ان الحوادث التي تمر علينا يوميا يمكننا ان نصنفها الى ثلاثة اصناف:

1- حوادث مؤكدة الحدوث Certain events : وهي حوادث نجزم بحدوثها جزما مؤكدا، والسبب ان هذه الحوادث قد حدثت بالتأكيد في جميع الفترات الزمنية ، مثل حادثة " موت الانسان " هي حادثة مؤكدة.

2- حوادث مستحيلة الحدوث Impossible events : وهي حوادث نجزم بعدم حدوثها ولانتوقع ان تحدث على الاطلاق، مثل حادثة " شروق الشمس ليلا في مدينة بغداد " هي حادثة مستحيلة، وكذلك حادثة " ان تعيش سمكة خارج الماء" هي حادثة مستحيلة.

3- حوادث محتملة الحدوث (غير مؤكدة) Probable (uncertain) events : وهي حوادث لايمكننا ان نجزم جزما مؤكدا بحدوثها أو عدم حدوثها، أي انها حوادث غير مؤكدة الحدوث ولامستحيلة الحدوث، مثال ذلك " سقوط الامطار يوم غد " " ارتفاع اسعار النفط بعد خمس سنوات من الآن ".

ولما كنا نعرف سلفاً نتائج الحوادث المؤكدة والمستحيلة فاننا لانهتم كثيرا بدراسة مثل تلك الحوادث، اما الحوادث المحتملة فاننا لايمكننا معرفة نتائجها معرفة قاطعة قبل حدوثها، لهذا فاننا في نظرية الاحتمال نركز اهتمامنا على دراسة الحوادث المحتملة لكي نتمكن (بشكل احتمالي) من التكهن بنتائجها ولمعرفة القوانين الاحتماليه المتحكمة بحدوثها. وبعبارة اخرى فان نظرية الاحتمال هي نظرية رياضية تتركز على دراسة الحوادث المحتملة وقياس مقدار احتمال حدوثها.

8 – 2 التجربة العشوائيــة Random Experiment :

يمكننا تعريف التجربة بانها حالة خاصة يمكننا من خلالها ملاحظة نتيجة ما، هنالك نوعان أساسيان من التجارب : محددة Deterministic وغير محددة يطلق عليها غالبا عشوائية Stochastic . وتكون التجربة محددة اذا كانت نتائجها ثابته ومستقرة ولاتخضع للمصادفة، وإذا اعيدت التجربة أكثر من مرة واحدة وبنفس الظروف في كل مرة فانها تعطي دائما نفس النتيجة، (مثال ذلك اذا كان لديك ميزان حساس وطلب منك وزن كتله معينه، ولو أعيدت التجربة (وزن الكتله) وتحت نفس الظروف ولأكثر من مرة واحدة فان النتيجة ستكون نفسها).

أما في التجارب العشوائية فان النتيجة تخضع دائماً لتأثيرات عشوائية(ولهذا سميت تجارب عشوائية)، فاذا أعيدت التجربة العشوائية أكثر من مرة واحدة فانها في كل مرة سوف تعطي نتائج مختلفة ولو بقدر بسيط عن المرات السابقة (مثال ذلك إذا رميت قطعة نقود أكثر من مرة فانك سوف لاتحصل دائما على نفس الوجه).

ان نظرية الاحتمال تتخصص بدراسة التجارب العشوائية، وباختصار فان التجربة العشوائية هي عملية تحقيق شروط معينه بحيث يمكنها اعطاء ناتج واحد من بين أكثر من نتيجة ممكنة.

<u>مثال (1) :</u> ان عملية رمي قطعة نقود هي تجربة عشوائية وان نتائج هذه التجربة هي أما ظهور الصورة H أو الكتابة T (Head or Teal) .

Tail

<u>مثال (2) :</u> ان عملية رمي قطعة زهر النرد هي تجربة عشوائية وان النتائج الممكنة لهذه التجربة ان يكون عدد النقاط على الوجه العلوي لقطعة النرد (6 , 5 , 4 , 3 , 2 , 1) .

<u>مثال (3) :</u>ان فحص فصيلة الدم هي تجربة عشوائية نتائجها أحد الأصناف (A , B , AB , O)

8 – 3 فضاء العينه Sample Space : ان فضاء العينه S المتعلق بالتجربة العشوائية هو المجموعة الجامعة المؤلفة من جميع النتائج الممكننة للتجربة.

- في المثال (1) يتكون فضاء العينه من العنصرين H و T أي ان : $S=\{ H ,T \}$
- في المثال (2) ان فضاء العينه المتعلق بالتجربة العشوائية هو: $S=\{ 1,2,3,4,5,6 \}$
- في المثال (3) ان فضاء العينه المتعلق بالتجربة هو : $S= \{ A , B , AB , O \}$

8 – 4 الحادث (أو الحدث) The event :

الحادث هو نقطة أو عدة نقاط في فضاء العينه ويرمز له بـ (Ei) . والحادث قد يكون بسيطا Simple event اذا تكون من نقطة واحدة في فضاء العينه (أي حالة واحدة من الحالات التي تظهر نتيجة التجربة) أو يكون حادثا مركبا Compound event اذاشمل حالتين أو أكثر من الحالات التي تظهر نتيجة التجربة .

مثال : الحصول على (عدد زوجي) في رمية الزار (النرد) يسمى حـادثا وهو يتـكون من النــقاط 2,4 , 6) من مجموعة نقاط فضاء العينه { 1 , 2, 3 , 4 , 5 , 6 } .

مثال : الحصول على الصورة (H) من رمي قطعة نقود معدنية مرة واحدة يسمى حادثا وهو يتكون من نقطة واحدة { H } من مجموعة نقاط فضاء العينه { H , T } .

الحوادث المتنافيه (المتناقضه) Mutually events :

يقال عن الحادثين E_1 , E_2 انهما متنافيان اذا استحال حدوثهما معا.

مثال : عند رمي قطعة نقود فمن المستحيل الحصول على (صورة وكتابة) في نفس الوقت .

الحوادث المستقلة Independent events :

الحوادث المستقلة هي الحوادث التي اذا وقع أحدها لايمنع أو يؤثر على وقوع الحوادث الاخرى.

مثال : عند رمي قطعتي نقود فالحصول على صورة في القطعة الاولى مثلا لايؤثر في نتيجة القطعة الثانية.

مثال : صندوق فيه عدد معين من الكرات البيضاء والسوداء (المتماثلة وزنا وحجما)، فعند سحب كرتين بحيث تعاد الاولى قبل سحب الثانية فان نتيجة السحبة الاولى لاتؤثر في نتيجة السحبة الثانية لذا فالحادثان مستقلان .

الحوادث غير المستقلة Non Independent events :

الحوادث غير المستقلة هي الحوادث التي اذا وقع أحدها يؤثر في وقوع الحوادث الاخرى.

مثال: حالة الصندوق وفيه الكرات البيضاء والسوداء فعند سحب كرتان على التوالي بحيث لاتعاد الكرة الاولى فان نتيجة السحبة الثانية تتأثر بنتيجة السحبة الاولى لذلا فالحادثان غير مستقلين.

الحالات الممكنة Possible Cases :

الحالات الممكنة هي جميع الحالات المختلفة التي يمكن ان تظهر في تجربة معينه.

مثال: عند رمي قطعة النقود فعدد الحالات الممكنة هنا حالتين (صورة Head أو كتابة Tail).

مثال : عند رمي زهر النرد عدد الحالات الممكنة هي 6 حالات ، أما عند رمي زهري نرد فعدد الحالات الممكنة هي $36 = 6 * 6$ حالة ، من ذلك نرى بأن الحالات الممكنة هي نفسها فضاء العينة.

الحالات المواتية Favorable Cases :

هي الحالات التي تحقق ظهور الحادث المراد دراسته وتسمى ايضا بحالات النجاح.

مثال: عند رمي زهر النرد فاذا كان الحادث هو الحصول على عدد زوجي فالحالات التي تحقق ظهور هذا الحادث هي الحصول على (2) أو (4) أو (6) فهذه الحالات الثلاثة تسمى الحالات المواتية.

الحالات المتماثلة Equally Likely Cases :

هي الحالات المتكافئة والمتساوية في امكانية حدوثها.

مثال : عند رمي قطعة النقود فان الظروف المهيأه للحصول على أي وجه (صورة أو كتابة) تكون متكافئة، فيقال بان الحالتين التي تنتج عن تجربة رمي قطعة النقود حالتان متماثلتان.

8 – 5 المجمـــوعات Sets :

المجموعة هي تجمع أي عدد من العناصر وسنمثل المجموعات بحروف لاتينية كبيرة مثل A,B,C,... وعناصر المجموعة بحروف صغيرة ....,a,b,c ، المجموعة $A = \{ a , b ,c \}$ فيقال ان $a \in A$

(العنص a ينتمي الى المجموعة A) وان $k \notin A$ (العنصر k لاينتمي الى المجموعة A).

المجموعة الخالية Empty set : هي المجموعة التي لايوجد فيها عناصر مطلقا ويعبر عنها بالرمز ϕ أو $\{ \}$.

المجموعة الجزئية Subset : نقول ان Aمجموعة جزئية من B ونرمز لها $A \subset B$ اذا كان كل عنصر في A منتميا للمجموعة B .

التساوي Equality : المجموعة A تساوي المجموعة B أي $A = B$ اذا كانت A , B تحتويان على نفس العناصر، ويكون ذلك اذا تحقق الشرطان $A \subset B$ و $B \subset A$.

المجموعة الكلية Universal set : هي المجموعة التي تدرس جميع المجموعات تحت البحث باعتبارها مجموعات جزئية منها ويرمز لها (S) .

المجموعة الكلية Universal set : هي المجموعة التي تدرس جميع المجموعات تحت البحث باعتبارها مجموعات جزئية منها ويرمز لها (S) .

الاتحاد Union : اتحاد المجموعتين A و B أي $A \cup B$ هو مجموعة العناصر الموجودة في A أو في B أو في كليهما وعبارة أخرى:

$$A \cup B = \{ x / x \in A \text{ or } x \in B \}$$

التقاطع Intersection : تقاطع المجموعتين B وA أي $A \cap B$ هي مجموعة العناصر المشتركة بين A و B أي هو مجموعة العناصر المنتمية الى A والى B في نفس الوقت، بعبارة أخرى:

$$A \cap B = \{ x / x \in B \text{ and } x \in A \}$$

أحيانا يستخدم الرمز AB للدلالة على $A \cap B$

المتممة Complement : المتممة A (بالنسبة الى المجموعة الكلية S) هي مجموعة العناصر المنتمية الى المجموعة الكلية S ولاتنتمي الى A ويرمز لها A^c ، بعبارة أخرى:

$$A^c = \{ x/x \in S \ , \ x \notin A \}$$

التباديل Permutation : يقصد بالتباديل بأنها عدد طرق الأختيار المرتب التي يمكن تكوينها من عدة أشياء يأخذها كلها او بعضها ويرمز له ((nPr)) أي تباديل r من n وقانونه هو :

$$nPr = n! / (n-r)!$$

حيث ان :

$$n! = n (n-1) (n-2) \ .... \ 3*2*1$$

$$0! = 1 \quad , \quad 1! = 1$$

مثال : إذا كان لدينا أربعة حروف A , B , C , D وأختير منها حرفان، فما هي عدد طرق الاختيار المرتب التي يمكن بها أختيار هذين الحرفين ؟

$$nPr = n! / (n-r)!$$

$$4 P2 = 4 ! / (4-2)! = (4*3*2*1) / 2*1 = 12$$

أي ان عدد الطرق = 12 وهذه الطرق هي :

AB	AC	AD	BC	BD	CD
BA	CA	DA	CB	DB	DC

حيث ان كلاً منهما يمثل ترتيباً مختلفاً للحرفين

التوافيق Combinations :

يقصد بالتوافيق بأنها عدد طرق الاختيار غير المرتب التي يمكن تكوينها من عدة أشياء يأخذها كلها أو بعضها ويرمز له $\binom{n}{r}$ أو C_r^n وقانونه هو :

$$C_r^n = \frac{n!}{r!(n-r)!}$$

حيث ان :

$$n! = n(n-1)(n-2) \;.....\; 3\;2\;1$$

$$0! = 1 \quad , \quad 1! = 1$$

مثال : ماعدد طرق الاختيار التي يمكن الحصول عليها لاختيار لجنة مؤلفة من 5 أشخاص من مجموع 9 أشخاص ؟

$$C_r^n = \frac{n!}{r!(n-r)!}$$

$$= 9! \,/\, 5!\,(9-5)! = 9! \,/\, 5! * 4! = 126$$

8 – 6 قاعدتان أساسيتان يعتمد عليهما كل من التباديل والتوافيق :

1- إذا كان عدد الطرق الممكنة لوقوع الحادث E_1 هو n وان عدد الطرق الممكنة لوقوع الحادث E_2 هو m ، وان E_1 و E_2 حادثان متنافيان فإن عدد الطرق الممكنة لوقوع الحادث E_1 أو E_2 هو (n + m) من الطرق. [illegible] (n+m)

ما هو عدد الطرق الممكنة لسحب ورقة من أوراق اللعب بحيث تكون الورقة [illegible] أو القلب ؟

مثال : عدد أوراق اللعب هو 52 ورقة وان ورقة (السنك ♣) يمكن أن تحدث 13 طريقة وان ورقة (القلب ♥) يمكن أن تحدث 13 طريقة ايضاً فعند سحب ورقة عشوائياً من أوراق اللعب فإن عدد الطرق الممكنة لإختيار السنك أو القلب = 13 + 13 = 26 طريقة

2- إذا كان عدد الطرق الممكنة لوقوع الحادث E_1 هو n وإن عدد الطرق الممكنة لوقوع E_2 هو m وكان E_1 و E_2 حادثان مستقلان فان عدد الطرق المكنة لوقوع الحادثان E_1 و E_2 هو $(n\ m)$ من الطرق.

<u>مثال :</u> اذا سحبت ورقتان من مجموعة أوراق اللعب بحيث ان احداهما تكون (سنك) والاخرى (قلب) فان هناك $13 * 13 = 169$ طريقة لعمل ذلك .

<u>مثال :</u> صندوق فيه 6 كرات حمراء و 4 سوداء و 2 بيضاء فبكم طريقة يمكن أختيار 5 كرات بحيث تكون 3 منها حمراء و 2 سوداء ؟

الحل : عدد طرق اختيار 3 كرات حمراء هو $C_3^6 = 20$

عدد طرق أختيار 2 كرة سوداء هو $C_2^4 = 6$

$\therefore$ عدد الطرق لاختيار 3 كرات حمراء و 2 سوداء هو $C_3^6\ C_2^4 = 20 * 6 = 120$

8 – 7 قياس الاحتمال Measuring of Probability :

أفرض إن هنالك محاولة يمكن ان تقع بعدد من الحالات الكلية المتنافية (أي أنه لايمكن وقوع حادثتين أو أكثر في آن واحد لنفس المحاولة، مثال ذلك عند رمي قطعة نقود فان نتيجة الرمية هي أما صورة أو كتابة ولايمكن ظهور الاثنين معا) التي تملك نفس الفرصة في الظهور وأفرض إن هذا العدد هو n ، وأفرض إن m من هذه الحالات $m \le n$ ممكنة لوقوع حادثة معينة لتكن E ، عندئذ فإن إحتمال وقوع E يعرَف على النحو الآتي:

$$\text{Probability}\ (E) = \Pr(E) = \frac{\text{الحالات الممكنة لوقوع الحادثة } E}{\text{الحالات الكلية لـ } E}$$

$$P = m / n$$

بعبارة أخرى لتكن E حدث يمكن أن يحدث بـ m طريقة وكانت n عدد جميع الحالات الكلية والتي لها نفس الفرصة في الظهور ، وبهذا فإن إحتمال حدوث الحدث (E) $\Leftarrow$ Pr (E)

$$Pr (E) = m / n = \bar{P}$$

إن الاحتمال (P) غالباً مايطلق عليه بأحتمال نجاح الحادثة E . إن عـدد الحالات الممكنة لعدم وقوع الحادثة E هو n – m وعندئذ فإن احتمال عدم وقوع E هو :

$$q = (n - m) / n$$

وغالباً مايطلق عليه باحتمال فشل الحادثة E وعليه فان :

$$q = (n - m) / n = 1 - m /n = 1 - p \Rightarrow p + q =1$$

8 – 8 بعض الملاحظات التي تساهم في توضيح فكرة ومفهوم الاحتمال :

1- ان كل من p و q كميات غير سالبة non – negative quantities ولايمكن ان تزيد قيمة أي منها عن الواحد وهذا يعني ان :

$$0 \leq P \leq 1 \quad , \quad 0 \leq q \leq 1$$

2- اذا كان $Pr (E) = 1$ فذلك يعني ان الحادثة E ((مؤكدة)) الوقوع، (مثال ذلك ماهو إحتمال وفاة شخص يوما ما؟ إن هذا الاحتمال مساوٍ للواحد لانه مؤكد وفاة أي شخص يوماً ما عاجلاً أو آجلاً).

3- إذا كان $Pr (E) = 0$ فذلك يعني ان الحادثة E ((مستحيلة)) الوقوع (مثال ذلك ماهو أحتمال ان يعيش شخص توقف قلبه عن العمل ؟ واضح إن هذا الاحتمال مساوٍ للصفر).

4- بشكل عام فإن إحتمال وقوع الحادثة E محصور في الفترة (1 , 0) أي ان $0 \leq Pr(E) \leq 1$

وعليه يمكن القول إنه إذا كانت $Pr (E) > ½$ فذلك يعني ان عدد الحالات الممكنة لوقوع E أكبر من عدد حالات عدم وقوعها، في حين إذا كانت $Pr (E) < ½$ فذلك يعني ان عدد الحالات الممكنة لوقوع E أقل من عدد حالات عدم وقوعها.

5- إن صغر أو كبر إحتمال وقوع الحادثة E يعتمد بطبيعة الحال على عدد الحالات الممكنة لوقوع E وعلى عدد الحالات الكلية المتاحة.

مثال 1 : سحبت قطرة دم من شخص ، ماهو إحتمال أن يكون صنف دم هذا الشخص هو AB ؟

الحل : ان عدد الحالات الكلية لعملية فحص الدم هي أربعة حالات وإن عدد الحالات الممكنة للصنف AB هي حالة واحدة، وعليه فإن إحتمال أن يكون صنف دم هذا الشخص AB هو:

$$Pr(AB) = 1/4 = 0.25$$

مثال ②: رمي زهرا نرد مرة واحدة،ماهو احتمال أن يكون مجموع النقاط على الوجهين الظاهرين :

A – مساوِ الى 9 B- مساوِالى 7 D- – أقل من 8 E- – أكبر من أو يساوي 6 .

الحل : بهدف التوضيح سنكون الجدول الآتي الذي يبين الحلات الكلية عند رمي زهرا النرد:

1,1	1,2	1,3	1,4	1,5	1,6
2,1	2,2	2,3	2,4	2,5	2,6
3,1	3,2	3,3	3,4	3,5	3,6
4,1	4,2	4,3	4,4	4,5	4,6
5,1	5,2	5,3	5,4	5,5	5,6
6,1	6,2	6,3	6,4	6,5	6,6

من الواضح ان عدد الحالات الكلية الممكنة هي 36 حالة:

A – إن الحالات الممكنة التي فيها مجموع النقاط مسـاوٍ الى 9 هي أربع حالات وعليه فإن :

$$Pr(E = 9) = 4 / 36 = 0.111$$

B – الحالات الممكنة التي فيها مجموع النقاط مسـاوٍ الى 7 هي ســت حالات وعليه فان :

$$Pr(E = 7) = 6 / 36 = 0.167$$

D – إن عدد الحالات التي يكون فيها المجموع أقل من ثمانية هو 21 وعليه فان :

$$Pr(E < 8) = 21 / 36 = 0.583$$

E – عدد الحالات التي يكون فيها المجموع أكبر من أو يساوي 6 هو 26 عليه فان :

$Pr (E \geq 6) = 26 / 36 = 0.722$

مثال :

صندوق يحتوي على 12 كرة، ثلاث منها حمراء وأربعة بيضاء وخمس سوداء، أختيرت ثلاث كرات من هذا الصندوق بشكل عشوائي، ماهو أحتمال أن تكون :

-1- جميعها حمراء -2- جميعها بيضاء -3- جميعها سوداء -4- أثنين منها حمراء والاخرى بيضاء -5- واحدة بيضاء والبقية سوداء.

الحل : ان عدد الحالات الكلية لاختيار ثلاث كرات من هذا الصندوق هو :

$$c_3^{12} = \frac{12!}{3!(12-3)!} = 220$$

-1- عـدد الحالات الممكنة لان تكون الكرات الثلاث المختارة حمراء هو :

$$c_3^3 \; c_0^4 \; c_0^5 = 1$$

∴ إحتمال أن تكون الكرات الثلاث حمراء هو :

$$Pr (3 \text{ balls are red}) = 1 / 220 = 0.0045$$

-2- عـدد الحالات الممكنة لأن تكون الكرات الثلاث المختارة بيضاء هو :

$$c_0^3 \; c_3^4 \; c_0^5 = 4$$

∴ احتمال ان تكون الكرات الثلاث بيضاء هو : $Pr (3\text{balls are white})= 4/220 = 0.0182$

-3- عدد الحالات الممكنة لأن تكون الكرات الثلاث المختارة سوداء هو :

$$c_0^3 \; c_0^4 \; c_3^5 = 10$$

∴ احتمال ان تكون الكرات الثلاث سوداء هو: $Pr(3\text{balls are black}) = 10/220 = 0.045$

-4- عدد الحالات الممكنة لان تكون كرتين منها حمراء والثالثة بيضاء:

$$c_2^3 \; c_1^4 \; c_0^5 = 12$$

∴ احتمال ان تكون هنالك كرتين حمراء والثالثة بيضاء هو:

Pr(2 red balls and the other white) = 12 / 220 = 0.055

-5- عدد الحالات الممكنة لان تكون كرة بيضاء والبقية سوداء هو :

$$c_0^3 \; c_1^4 \; c_2^5 = 40$$

∴ احتمال ان تكون هنالك كرة بيضاء والبقية سوداء هو:

Pr(one white ball and the others black) = 40/220 = 0.182

8 – 9 قواعد عامة في نظرية الاحتمالات General rules in probability theory

أولاً : قاعدة الجمع The addition rule :

أ-عندما تكون الحوادث متنافية: لتكن E_1 ، E_2 حادثتين متنافيتين في محاولة معينة، عندئذ فان احتمال حدوث E_1 أو E_2 يمثل حاصل جمع احتمال E_1 مع احتمال E_2 ، أي أن :

$$Pr (E_1 \text{ or } E_2) = Pr (E_1) + Pr (E_2)$$

بشكل عام اذا كانت E_1 , E_2 ,........, E_k تمثل K من الحوادث المتنافية في محاولة معينة، عندئذ فان احتمال حدوث E_1 أو E_2 أو ...... E_k يمثل حاصل جمع احتمالات هذه الحوادث ، أي أن:

$$Pr (E_1 \text{ or } E_2 \text{ or } .... \text{ or } E_k) = \sum_{i=1}^{k} Pr (E_i)$$

مثال : تأمل تجربة رمي زهر النرد ، واضح إن إحتمال ظهور الوجه الذي يجعل عدد النقاط 3 أو الوجه الذي يجعل عدد النقاط 5 أو الذي عدد النقاط 6 هو :

Pr (3 or 5 or 6) = Pr (3) + Pr (5) + Pr (6) = 1/6 + 1/6 + 1/6

طالما إن هذه الحوادث متنافية = 3/6 = 0.5

مثال : سحبت قطرة دم من شخص يرغب في فحص دمه ، ماهو إحتمال ان يكون دم هذا الشخص من

صنف غير الصنف O .

$$Pr\ (A\ or\ B\ or\ AB) = Pr\ (A) + Pr\ (B) + Pr\ (AB)$$

$$= 1/4 + 1/4 + 1/4 = 3/4 = 0.75$$

Or $Pr\ (A\ or\ B\ or\ AB) = 1 - Pr\ (O) = 1 - 1/4 = 3/4 = 0.75$

ب – *عنـدما تكون الحوادث غير متنافية* :

لتكن E_1 و E_2 حادثتين ممكنتي الوقوع معاً في محاولة معينة، عندئذ فان احتمال وقوع E_1 أو E_2 هو حاصل جمع احتمال وقوع E_1 مع احتمال وقوع E_2 مطروحا من ذلك احتمال وقوعهما معاً، أي :

$$Pr\ (E_1\ or\ E_2) = Pr\ (E_1) + Pr\ (E_2) - Pr\ (E_1E_2)$$

<u>مثال :</u> إذا علمت إن احتمال سـقوط المطر في احد أيام فصل الربيع هو 0.52 وان إحتمال كون الطقس مشمس هو 0.36 وإن إحتمال كون الطقس مشمس وممطر في آن واحد هو 0.18 ، ماهو إحتمال ان يكون الطقس مشمس أو ممطر.

الحل : لنفرض ان E_1 تمثل حادثة سقوط المطر في ذلك اليوم ، وان E_2 تمثل حادثة كون الطقس مشمس في ذلك اليوم.

$$\therefore\ Pr\ (E_1\ or\ E_2) = Pr\ (E_1) + Pr\ (E_2) - Pr\ (E_1E_2)$$

$$= 0.52 + 0.36 - 0.18 = 0.70$$

<u>ثانياً : قاعـدة الضرب The multiplication rule :</u>

أ- *عندما تكون الحوادث مستقلة :* لتكن E_1 و E_2 حادثتين مستقلتين، عندئذ فان إحتمال حدوث

E_1 و E_2 معاً هو :

$$Pr (E_1 \text{ and } E_2) = Pr(E_1) . Pr(E_2)$$

بشكل عام اذا كانت E_1 , E_2 ,……. E_k تمثل K من الحوادث لمحاولة معينة وإن هذه الحوادث مستقلة عن بعضها ، عندئذ فان احتمال حدوثها معا هو :

$$Pr (E_1 \text{ and } E_2 \text{ and } \ldots\ldots \text{ and } E_k) = \prod_{i=1}^{k} Pr (E_i)$$

مثال : تأمل تجربة رمي قطعتي نقود معدنية، وأفرض إن E_1 تمثل ظهور الوجه الذي يحمل صورة للقطعة الأولى، وإن E_2 تمثل ظهور الوجه الذي يحمل كتابة للقطعة الثانية.

الحل : واضح ان احتمال E_1 هو 0.5 وإن احتمال E_2 هو 0.5 ، عليه فان احتمال ظهور صورة في القطعة الاولى وكتابة في القطعة الثانية هو :

$$Pr (E_1 \text{ and } E_2) = Pr (E_1) . Pr (E_2) = 0.5 * 0.5 = 0.25$$

طالما إن الحادثان مستقلان

ب - *عندما تكون الحوادث مشـــروطة* ***Conditional events:***

لتكن E_1 حادثة في محاولة معينة وإن E_2 حادثة أخرى وقوعها مشروط بوقوع الحادثة E_1 ، عندئذ فان احتمال وقوع الحادثتين في آن واحد مســاوٍ لحاصل ضرب إحتمال وقوع E_1 في احتمال وقوع E_2 المشروطة بوقوع E_1 ، ويرمز لاحتمال وقوع E_2 المشروطه بوقوع E_1 بالشكل [$Pr (E_2/E_1)$] وعندئذ فان :

$$Pr (E_1 \text{ and } E_2) = Pr (E_1 . E_2) = Pr(E_1) . Pr (E_2 / E_1)$$ تصحح

وهذا يعني إن :

(احتمال وقوع الحادثة E_2 المشروط بوقوع الحادثة E_1) $Pr(E_2/E_1) = Pr(E_1.E_2)/Pr(E_1)$

كذلك فان :

(احتمال وقوع الحادثة E_1 المشروط بوقوع الحادثة E_2) $Pr(E_1/E_2) = Pr(E_1.E_2)/Pr(E_2)$

مثال : تأمل تجربة رمي زهري النرد ولتكن E_1 حادثة تتمثل في أن يكون مجموع النقاط على وجهي الزهرين مســاوٍ الى 6 ، ولتكن الحادثة E_2 تتمثل في أن يكون وجه أحد الزهرين يحمل نقاط عددها 2 ، جــد إحتمال حدوث E_2 المشروط بوقوع الحدث E_1 .

الحل: واضح ان عدد الحالات الممكنة لحدوث E_1 هي خمس حالات وهي :

$(1,5),(5,1),(2,4),(4,2),(3,3)$

عدد الحالات الكلية عند رمي زهري نرد = 36

$\therefore$ احتمال حدوث E_1 = عدد الحالات الممكنة / عددالحالات الكلية

$\therefore \; Pr(E_1) = 5/36$

$\because$ إن عدد الحالات الممكنة لحدوث E_2 بحيث يكون المجموع مســاوٍ الى 6 هو حالتين فقط هما $(2,4)$ و $(4,2)$ ، عندئذ فان احتمال حدوث E_1 و E_2 معاً هو:

$$Pr(E_1 . E_2) = 2/36$$

عليه فإن إحتمال حدوث E_2 مشــروط بحدوث E_1 ، لذا فان :

$$Pr(E_2/E_1) = Pr(E_1.E_2)/Pr(E_1) = (2/36)/(5/36) = 2/5 = 0.4$$

Some Probability distribution بعض التوزيعات الاحتمالية 8- 10

أستعرضنا في الفقرات السابقة مفهوم الاحتمال وكيفية قياسه وبعض القواعد الهامة المتعلقه به، في هذه الفقرة سوف نتطرق الى دراسة موجزة لبعض التوزيعات ا لاحتمالية الشائعة الاستخدام في التطبيقات الاحصائية من خلال عرض مبسط لكل توزيع منها دون الدخول في التفاصيل.

The Binomial distribution أولاً : توزيع ذي الحدين

أفرض ان هنالك تجربة يمكن تكرارها بعدد معين من المحاولات المستقلة ليكن n ، بحيث ان نتائج كل محاولة تتمثل بنتيجتين ممكنتين فقط هما نجاح المحاولة أو فشلها، ولنفرض ان P يمثل احتمال نجاح المحاولة وان $q = 1-p$ يمثل احتمال فشلها، عندئذ فان احتمال الحصول على X من المحاولات الناجحة ($X \leq n$) من أجمالي هذه المحاولات هو:

$$Pr(X) = C_x^n \, p^x q^{n-x} \qquad x = 0,1,2, \dots\dots ,n$$

يلاحظ من هذا الا ستنتاج ان هنالك متغير عشوائي من النوع المتقطع x (يشير الى المحاولات الناجحة)

وهنالك دالة بدلالة هذا المتغير هي p (x) هذه الدالة تعرف بدالة الكتلة الاحتمالية للمتغير x Probability mass function وتعرف هذه الدالة بأسم توزيع ذي الحدين وغالباً مايتم الاصطلاح لهذا التوزيع بالشكل $x \sim b(n,p)$ ، ويمكن صياغة التعريف أعلاه بالشكل، حيث يقال ان المتغير العشوائي x يتوزع وفق توزيع ذي الحدين إذا كانت دالة الكتلة الاحتمالية تأخذ الشكل الآتي :

$$P(x) = \begin{cases} C_x^n \, p^x q^{n-x} & x = 0,1,2,\dots\dots,n \\ 0 & other\ wise \end{cases}$$

علما ان $0 < p < 1$ وان $q = 1 - p$

الجدول الآتي يبين بعض المقاييس الخاصة بهذا التوزيع :

المقياس	الصيغة
1- الوسط الحسابي لقيم x في هذا التوزيع هو	$\mu_x = np$
2- التباين لقيم x في هذا التوزيع هـــو	$\sigma_x^2 = npq$
3- الانحراف المعياري لقيم x في هذا التوزيع هو	$\sigma_x = \sqrt{npq}$

خصائص الدالة :

1- انها دالة وحيدة القيمة Single valued function.

2- انها دالة موجبة تتراوح بين $0 < pr(x) < 1$.

3- مجموع القيم الاحتمالية (الكتل الاحتمالية) المقترنة بكافة قيم المتغير x يجب ان يكون مســاوٍ للواحد، أي ان :

$$\sum_{x=0}^{n} C_x^n P^x q^{n-x} = 1$$

مثال : أفرض ان $x \sim b(5, 0.8)$ ، جــد مايأتي :

1- دالة الكتلة الاحتمالية p.m.f الى x .

2- رســم دالة الكتلة الاحتمالية لـ x .

3- حساب الوسط الحسابي والتباين لـ x .

4- $Pr(x=5)$, $Pr(x=2)$, $Pr(x \le 4)$, $Pr(x>2)$

الحل:

1- بما ان $x \sim b\ (5,0.8)$ لذا فان دالة الكتلة الاحتمالية الى x هي :

$$P(x) = C_x^5\,(0.8)^x(0.2)^{5-x}$$

2- لغرض رسم هذه الدالة سنوجد قيم x علما ان x=0,1,2,3,4,5

$P(x=0) = C_0^5\ 0.8^0\ 0.2^{5-0} = 0.00032$

$P(x=1) = C_1^5\ 0.8^1 0.2^{5-1} = 0.0064$

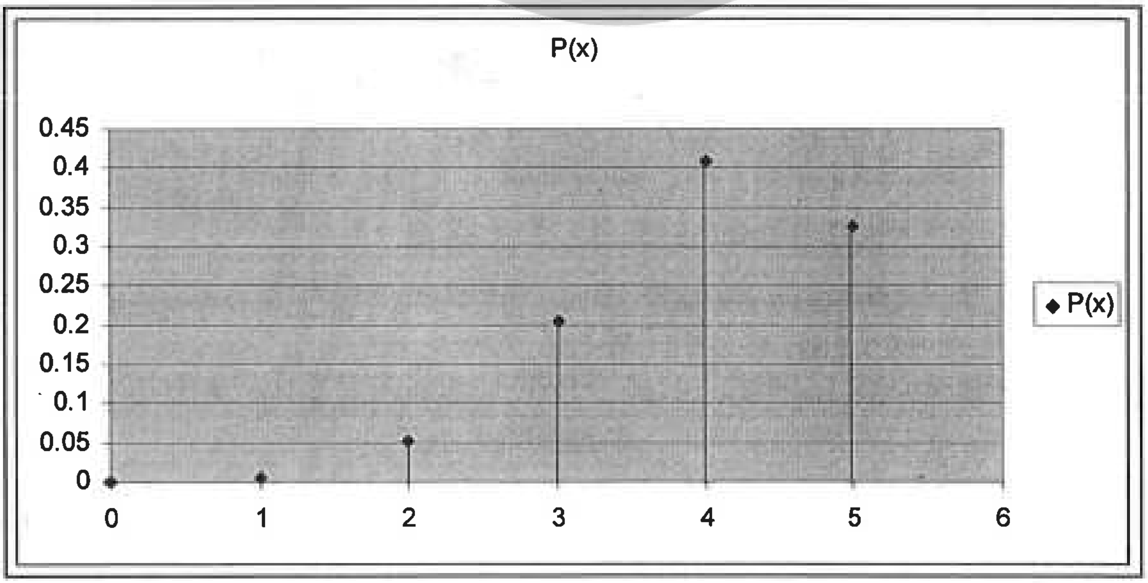
$P(x=2) = C_2^5\ 0.8^2\ 0.2^{5-2} = 0.0512$

$P(x=3) = C_3^5\ 0.8^3\ 0.2^{5-3} = 0.2048$

$P(x=4) = C_4^5\ 0.8^4\ 0.2^{5-4} = 0.409$

$P(x=5) = C_5^5\ 0.8^5\ 0.2^{5-5} = 0.327$

ثم تعيين احداثيات النقاط (x , P(x)) وكما موضح في الشكل :



$$\mu_x = n\,p = 5 * 0.8 = 4 \qquad -3$$

$$\sigma_x^{\,2} = n\,p\,q = 5 * 0.8 * 0.2 = 0.8$$

4- من الجدول الســـابق نلاحظ ان

$$Pr\,(X = 5) = P\,(5) = 0.3276$$

$$Pr = (X = 2) = P\,(2) = 0.0512$$

$$Pr\,(X \le 4) = P(0) + P(1) + P(2) + P(3) + P(4)$$

$$= 1 - Pr\,(x = 5) = 1 - 0.32768 = 0.67232$$

$$Pr\,(x > 2) = P(3) + P(4) + P(5) = 0.94208$$

ثانياً : توزيع بواسون Poisson distribution

يعتبر توزيع بواسون واحد من التوزيعات المتقطعة ذات أهمية كبيرة في التطبيقات الاحصائية. ويمثل هذا التوزيع حالة خاصة من توزيع ذي الحدين عندما يكون احتمال نجاح المحاولة P صغيراً جــداً (نظرياً يقترب من الصفر) وان عدد المحاولات n كبيرجداً (نظرياً يقترب من ∞) بحث ان nP وفق هذين الشرطين لتستقر نحو عدد ثابت مثل λ ، عليه ووفق هذه الشروط فان توزيع بواسون هو توزيع تقاربي من توزيع ذي الحــدين ، أي ان :

$$\lim_{\substack{p \to 0 \\ n \to \infty}} C_x^n\, p^x q^{n-x} = \frac{\lambda^x e^{-\lambda}}{x!} \quad , \quad x = 0, 1, 2, \ldots$$

يلاحظ من هذا الاستنتاج ان هنالك متغير عشوائي متقطع X وهنالك دالة بدلالة هذا المتغير هي:

$P\,(X) = \dfrac{\lambda^x e^{-\lambda}}{x!}$ ، هذه الدالة تعرف بأســـم دالة توزيع بواسون ، والتي تعبر عن احتمال الحصول على x من المحاولات الناجحة.

ان المتغير العشوائي x يتوزع وفق دالة توزيع بواسون اذا كانت دالة الكتلة الاحتمالية لهذا المتغير تأخذ الشكل التالي :

$$P(x) = \begin{cases} \dfrac{\lambda^x \, e^{-\lambda}}{x!} & , x = 0, 1, 2, \dots\dots\dots \\ 0 & \text{other wise} \end{cases}$$

حيث ان :-

$e = 2.71828$ ، $\lambda > 0$

أي أن $x \sim Po(\lambda)$ ، ويتم تشخيص التوزيع من خلال تحديد قيمة λ . ان هذه الدالة تتميز بنفس خصائص دالة توزيع ذي الحدين من حيث كونها دالة وحيدة القيمة وانها دالة موجبة وان :

$$\sum \frac{\lambda^x \; e^{-\lambda}}{x!} = 1$$

غالباً مايطلق على هذا التوزيع تسمية توزيع الحوادث النادرة كونه يتعامل مع حالات تطبيقية تتصف بكون ان احتمال نجاح المحاولة صغيرة جداً، على سبيل المثال حوادث سقوط الطائرات. الجدول الآتي يبين بعض المقاييس الخاصة بهذا التوزيع:

المقياس	الصيغة
1- الوسط الحسابي لقيم x في هذا التوزيع هو	λ
2- التباين لقيم x في هذا التوزيع هــو	λ
3- الانحراف المعياري لقيم x في هذا التوزيع هو	$\sqrt{\lambda}$

مثال : لوحظ في منطقة معينة ان من بين كل 1000 شخص هنالك شخص واحد مصاب بمرض معين، ماهو احتمال ان من بين 2000 شـــخص هنالك :

1- ثلاثة أشخاص مصابين بهذا المرض.

2- أكثر من شخصين مصابين بهذا المرض.

الحل : (طالما ان احتمال نجاح المحاولة صغير جدا وان عدد المحاولات كبير نستخدم توزيع بواسون)

ان احتمال الاصابة بهذا المرض هو 0.001 (احتمال نجاح المحاولة) ، عليه فان احتمال اصابة X شخص بهذا المرض هو :

$$P(x) = \frac{\lambda^{x} \ e^{-\lambda}}{x!}$$

وان $\lambda = np = 2000 * 0.001 = 2$

1- احتمال ان هنالك ثلاث أشخاص مصابين بهذا المرض هو :

$$P(x = 3) = \frac{2^{3} \ e^{-2}}{3!} = 0.18045$$

2- احتمال وجود أكثر من شخصين مصابين بهذا المرض هو:

$$Pr(x>2) = 1 - Pr(x \leq 2) = 1 - [P(0) + P(1) + P(2)]$$

$$= 1 - 0.67668 = 0.32332$$

مثال : اذا علمت ان $x \sim Po(3)$ ، جـد الوسط الحسابي والتباين، ثم أرســم دالة هذا التوزيع مع حساب الاحتمالات الآتية: $Pr(x=2)$, $Pr(1 \leq x \leq 3)$, $Pr(x \leq 2)$

الحل : حيث ان $\lambda = 3$

$$P(x) = \frac{3^x \, e^{-3}}{x!}$$

- الوسط الحسابي لقيم x هو ⇦ $\mu_x = \lambda = 3$
- تبايـــن قيـــم x هو ⇦ $\sigma_x^2 = \lambda = 3$

لغرض رســـم دالة هذا التوزيع نعمل الجدول الآتي :

X :	0	1	2	3	4	5	6	7	8	...
P(x) :	0.05	0.15	0.22	0.22	0.17	0.1	0.05	0.03	0.01	....

Where $P(x=0) = \frac{\lambda^x e^{-\lambda}}{x!} = \frac{3^0 e^{-3}}{0!} = \frac{1}{(2.71828)^3} = \frac{1}{20.0855} = 0.05$

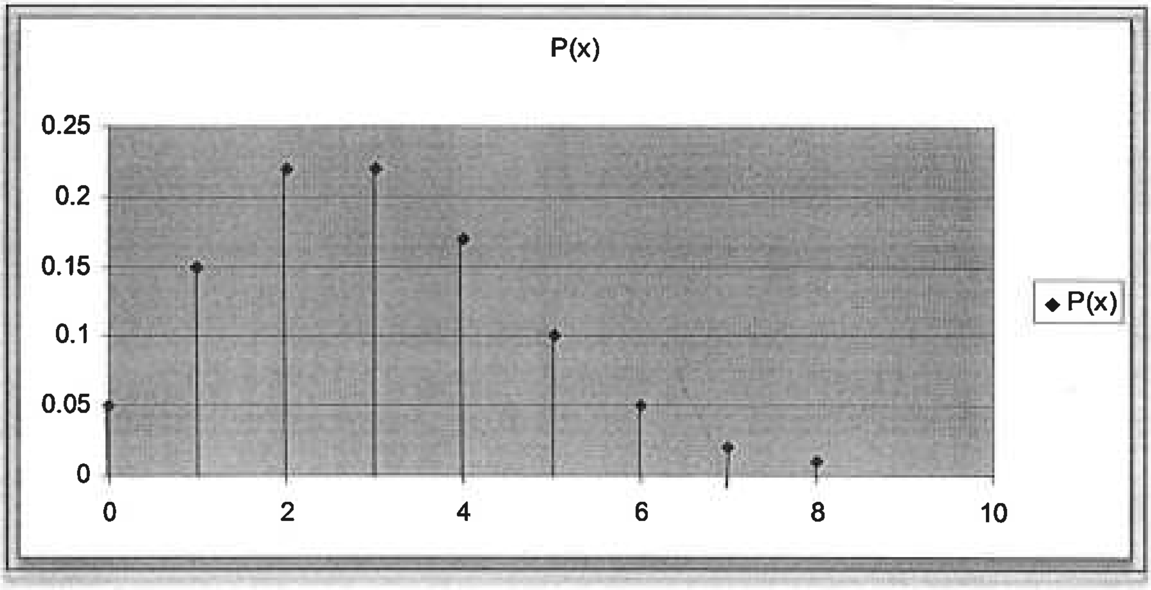
$P(x=1) = \frac{3^1 \; e^{-3}}{1!} = \frac{3}{(2.71828)^3} = 0.15$

$P(x=2) = \frac{3^2 \; e^{-3}}{2!} = \frac{9}{2 * (2.71828)^3} = 0.22$

$P(x=3) = \frac{3^3 \; e^{-3}}{3!} = \frac{27}{(3*2*1)*(2.71828)} = 0.22$

وهكذا يتم التعويض بنقاط المتغير(x) الى 8 = x أو أكثر ليتسنى لنا رسم دالة التوزيع .....

ومن ثم نعين احداثيات النقاط (x , P(x)) كما موضح في الشكل الآتي :



$$Pr\ (1 \le x \le 3) = p(1) + p(2) + p(3)$$

$$= \frac{3^1\ e^{-3}}{1!} + \frac{3^2\ e^{-3}}{2!} + \frac{3^3\ e^{-3}}{3!} = 0.59$$

$$Pr\ (x \le 2) = p(0) + p(1) + p(2)$$

$$= \frac{3^0\ e^{-3}}{0!} + \frac{3^1\ e^{-3}}{1!} + \frac{3^2\ e^{-3}}{2!} = 0.42$$

ثالثاً : التوزيع الطبيعي The Normal distribution :

يعتبر التوزيع الطبيعي من أهم التوزيعات الاحتمالية الشائعة الاستخدام في النظرية الأحصائية، حيث ان غالبية الظواهر الطبيعية تسلك وفق هذا التوزيع، وهو من التوزيعات الاحتمالية المستمرة.

يقال ان المتغير العشوائي المستمر (x) هو ذا توزيع طبيعي اذا كانت الدالة الاحتمالية لهذا التوزيع تأخذ الشكل الآتي:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\ e^{\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)} \qquad -\infty < x < \infty$$

$$0 < \sigma < \infty \quad , \quad -\infty < \mu < \infty$$

حيث ان $\pi = 3.14159$, $e = 2.71828$

إن μ تمثل الوسط الحسابي لقيم (x) في التوزيع، σ الانحراف المعياري لقيم x ، وتسميان بمعلمتي التوزيع، وغالباً مايتم اعتماد الشكل الآتي للدلالة على هذا التوزيع $x \sim N(\mu, \sigma^2)$ الذي يعني ان المتغير x يسلك (~) وفق دالة توزيع طبيعي (N) بوسط قدره (μ) وتباين مقداره (σ^2) ، وحيث ان دالة هذا التوزيع هي دالة احتمالية فانها يجب أن تتصف بما يلي:

1- ان f(x) دالة وحيدة القيمة، أي مانعنيه ان لأية من قيم x المعوضة في f(x) تعطي قيمة واحدة فقط الى f(x) .

2- ان f(x) دالة موجبة دائما مهما كانت قيمة x سالبة أم موجبة.

3- ان المساحة تحت منحني الداله f(x) ضمن الفترة ($-\infty,\infty$) تساوي واحـد، أي ان :