



وزارة التعليم العالي والبحث العلمي

جامعة تكريت

كلية الإدارة والاقتصاد

قسم ادارة الاعمال

المرحلة الاولى

محاضرات في مادة الاحصاء

مدرس المادة

م.د ثائر جاسم مُحمَّد

2023 م

1445 هـ

الفصل الأول

المبحث الأول

((معنى الإحصاء وأهميته في البحوث والدراسات العلمية))

1- تعريف الإحصاء.

علم الإحصاء : هو الطريقة العلمية التي تختص بجمع البيانات والحقائق عن ظاهره او فرضيه معينه وتنظيم وتبويب هذه البيانات والحقائق بالشكل الذي يسهل عملية تحليلها وتفسيرها ومن ثم استخلاص النتائج باتخاذ القرار على ضوء ذلك ، وبشكل عام فإن علم الإحصاء يقسم الى القسمين الرئيسيين الآتيين :

1- **الإحصاء الوصفي** : وهو الفرع الذي يتضمن الطرق والأساليب المستخدمة في جمع البيانات عن ظاهرة معينة او مجموعه ظواهر وكيفية تنظيمها وتصنيفها وتبويبها وعرضها وحساب المؤشرات الإحصائية منها.

2- **الإحصاء الاستدلالي** : وهو الفرع الآخر لعلم الإحصاء الذي يهتم بموضوع التقدير واختبار الفرضيات.

2- اهميه علم الاحصاء ومجالات تطبيقاته

يعتبر علم الاحصاء احد الوسائل المهمة في البحث العلمي من خلال استخدام قواعده وقوانينه وطرقه في جمع البيانات اللازمة للبحث العلمي وتحليل هذه البيانات بغية الوصول الى النتائج التي يهدف لها البحث . ويمكن تمثيل مجالات تطبيق علم الاحصاء حسب ما هو موضح ادناه:

- 1- البحوث الزراعية التطبيقية.
- 2- البحوث الصناعية التطبيقية.
- 3- البحوث النفسية والتربوية.
- 4- البحوث الطبية.
- 5- بحوث التربية الرياضية.
- 6- البحوث الاقتصادية والإدارية.
- 7- البحوث الهندسية.
- 8- بحوث علم الأرض ، علم الأحياء (المايولوجيا) ، الصيدلة ، علم الفيزياء ، معالجة الصور الرقمية ... الخ.

3- الطرق الإحصائية في البحث العلمي

ان استخدام الاسلوب الإحصائي في البحث العلمي يعني توفر بيانات ومعلومات عن الظاهرة أو الظواهر المطلوب دراستها في البحث العلمي ، لذلك فإن إمكانية تطبيق الطرق الإحصائية مرهون بإمكانية التعبير عن هذه الظاهرة تعبيراً كمياً.

4- المراحل الرئيسية للطرق الإحصائية في البحث العلمي :

- أ- تحديد مشكلة البحث أو فرضية البحث أو الدراسة.
- ب- جمع البيانات والمعلومات عن الظاهرة أو الظواهر ذات العلاقة بالبحث أو الدراسة.
- ج- تصنيف البيانات وتبويبها وعرضها.
- د- تحليل معطيات الدراسة للوصول إلى النتائج في ضوء فرضية البحث أو الدراسة.
- هـ- حساب المؤشرات الإحصائية كتقديرات لمعالم مجتمع البحث أو الدراسة.
- و- تفسير النتائج وعملية اتخاذ القرار بشأن فرضيات البحث.

5- أسلوب تصميم البحث :

توجد اعتبارات كثيرة يتوقف عليها تصميم البحث العلمي يتوجب على الباحث اخذها بنظر الاعتبار وأهمها:

- أ- تحديد الغرض من البحث : يجب تحديد الغرض من البحث بشكل واضح ودقيق لغرض الاستفادة من النتائج التي يتوصل إليها الباحث ، فمثلاً إذا كان البحث يدرس نمط استهلاك الفرد من سلعة ما مثلاً (لحم الغنم) فإنه يجب دراسة الطلب على لحم الغنم من قبل المستهلكين.
- ب- تحديد اطار البحث: احد الامور المهمة قبل البدء بتنفيذ البحث هو تحديد نوع وطبيعة مجال ذلك البحث ، اي ما نعينه المجتمع الاحصائي على نحو واضح ودقيق، والمجتمع الاحصائي هو عبارته عن مجموعه من الوحدات او المفردات التي تشترك بصفه او صفات معينه والتي غالبا ما يتم الحصول منها على البيانات والمعلومات المطلوبه . فمثلا اذا كان البحث يهدف الى التعرف على نسبة المرضى المصابين بمرض معين في قرية ما ، فان المجتمع الاحصائي هو افراد القرية ، اما ما يسمى بالمفردة الاحصائية فهي تمثل الفرد الذي يمكن فحصه من افراد القرية . والمجتمع الاحصائي يمكن ان يكون محدد او غير محدد. فالمجتمع الاحصائي المحدد هو الذي يمكن ملاحظة او مواجهة كل مفردة من مفرداته اي يمكن حصر كافة المفردات التي تنتمي له . مثلا طلبه جامعه تكريت . اما المجتمع الاحصائي الغير محدد هو الذي لا يمكن ملاحظة او مواجهة كل مفرداته اي لا يمكن حصر كافة مفرداته. مثلا كريات الدم الحمراء في جسم الانسان.

ت- 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 لتحقيق بعض فقرات البحث كذلك التأكد من مدى توفر البيانات والمعلومات الدقيقة عن مشكلة البحث.

ث- تحديد أسلوب جمع البيانات والمعلومات : لغرض الحصول على البيانات والمعلومات التي يتطلبها البحث أو الدراسة يجب تحديد نوع أو أسلوب جمع هذه البيانات والمعلومات .

6- أساليب جمع البيانات : هنالك نوعان من أساليب جمع البيانات هي :

أ- أسلوب التسجيل الشامل لكافة مفردات المجتمع الإحصائي/

وهو جمع البيانات والمعلومات عن كافة المفردات التي تؤلف المجتمع الإحصائي الظاهرة (أو الظواهر) قيد البحث ، ويجب أن يكون المجتمع محدد مثل عملية التعداد العام للسكان في العراق ، وهو أفضل أسلوب لكونه يجهز الباحث ببيانات كاملة عن مفردات المجتمع والدراسة ، إلا أنه يحتاج إلى موارد مادية وبشرية كبيرة ، وكذلك احتمال الوقوع في أخطاء أكبر لكبر العينة وضخامة البيانات.

ب- أسلوب التسجيل عن طريق العينات /

يقصد بأسلوب العينات عملية جمع البيانات والمعلومات عن مجموعة معينة من مفردات مجتمع الدراسة ، وهذه المجموعة من المفردات تسمى (عينة) مثل استفتاء رأي ، دراسة فاعلية علاج معين للمصابين بمرض القلب ، إن هذا الأسلوب يحتاج إلى جهد وموارد مالية وبشرية أقل من التسجيل الشامل ، وكذلك فإنه مفيد عند دراسة المجتمعات غير المحددة ، إلا أن دقته تعتمد على النتائج المستخلصة من بيانات العينة.

7- وسائل جمع البيانات :

بعد تحديد حجم العينة وأسلوب المعاينة الملائم في اختيار مفردات هذه العينة من مجتمع ما ، يتطلب اختيار الوسيلة الملائمة في جمع البيانات وهنالك عدة وسائل لجمع البيانات أهمها :

أ- أسلوب الجمع المباشر :

وفق هذا الأسلوب يتم جمع البيانات والمعلومات المتوفرة لدى جهات معينة كأجهزة الدولة أو الهيئات ذات العلاقة مثل الجهاز المركزي للإحصاء ، وكذلك البيانات ذات الطابع المختبري أو الحقلية التي يتم من قبل الباحث.

ب- الاستبيان :

هو عبارة عن استمارة تحتوي على مجموعة من الأسئلة يتم من خلالها جمع البيانات أو المعلومات من مفردات (أو بعض مفردات) مجتمع الدراسة من خلال مواجهة الباحث شخصياً للمفردة الإحصائية أو عن طريق المراسلة ، كما في التعداد العام للسكان.

لغرض تصميم استمارة الاستبيان يتوجب على الباحث مراعاة ما يلي :

أ- إعداد مقدمة إيضاحية يتم من خلالها الإطلاع على هدف الاستمارة والبحث لغرض تسهيل الإجابة على أسئلة الاستمارة وبصورة صحيحة.

ب- أن تكون فقرات الاستمارة متسلسلة ومتكاملة ويكون كل جزء منها يحقق غرض معين . ويجب عند وضع الأسئلة مراعاة ما يلي:

1- أن تكون الأسئلة متوسطة العدد.

2- أن تكون الأسئلة واضحة المعنى ليس فيها غموض وقصور في جانب معين.

3- أن تكون الإجابة على الأسئلة بـ نعم أو لا أو إشارة وأن تكون الأجوبة أمام السؤال لكي يجيب عليها الفرد بسهولة.

4- أن يؤخذ بنظر الاعتبار عند وضع السؤال ظروف تفريغ وتصنيف وتبويب وترميز الأجوبة لتسهيل وسرعة إنجاز البحث.

8- الأخطاء الشائعة في جمع البيانات :

يحدث في بعض الأحيان أخطاء يقع بها الباحث عند جمعه للبيانات والمعلومات التي يتطلبها البحث وهذه الأخطاء تحدث نتيجة سوء استخدام الطريقة الإحصائية أهمها :

أ- **خطأ التحيز:** وهو الخطأ الذي يحدث عندما يأخذ الباحث المعلومات أو البيانات من مصادر ثانوية وليست المصادر الأصلية.

ب- **أخطاء الصدفة :** يحدث هذا الخطأ بسبب الباحث نفسه كأن يقوم الباحث باستيفاء بعض المعلومات والبيانات بالاعتماد على معلوماته الشخصية أو التعمد في جمع البيانات من بعض المفردات دون الأخرى المحددة ، أو جمع بيانات ناقصة لسبب أو آخر ، هذه الأخطاء وغيرها لها أثر في الحصول على نتائج غير دقيقة لتلك الدراسة.

9- النقاط التي يجب تحديدها قبل الشروع بعملية جمع البيانات هي :

1- تحديد الغرض من البحث.

2- تحديد مجال البحث.

3- تحديد المصدر.

4- تعيين الطريقة التي تتبع في الحصول على البيانات الإحصائية.

5- تحديد الوحدات القياسية المناسبة المستخدمة في عملية القياس.

6- تحديد الميزانية اللازمة للبحث.

7- تحديد الوقت

10- المصادر الإحصائية في جمع البيانات :

1- **المصادر التاريخية :** وهي البيانات والمعلومات المحفوظة والمتجمعة لدى أجهزة ومؤسسات دوائر الدولة المختلفة نتيجة الاستقصاءات أو المسوحات التي قامت بها هذه الجهات أو الهيئات مثال ذلك الجهاز المركزي للإحصاء مثال ذلك، التعداد العام للسكان في العراق.

2- **المصادر الميدانية :** وهي المصادر التي يقوم الباحث بجمعها والحصول عليها بنفسه من مصادرها الأصلية أما عن طريق المقابلة الشخصية أو المراسلة بالبريد أو محاكاة الحاسوب والانترنت.

أسلوب العينات

أ- تعريف المجتمع الاحصائي : هو عبارة عن جميع القيم أو المفردات التي يأخذها المتغير ، فمثلاً إذا كانت الصفات متعلقة بعمر الموظفين العاملين بإحدى الوزارات فإن المجتمع في هذه الحالة هو جميع الموظفين في تلك الوزارة والمجتمع أما أن يكون :

1- مجتمعاً محدداً : أي من الممكن حصر عدد مفرداته كما هو الحال في حصر عمر عدد من الموظفين في وزارة الصناعة مثلاً أو عدد الوحدات الإنتاجية من نوع معين في يوم معين.

2- المجتمعات غير المحددة : وهي المجتمعات التي يتعذر حصرها أو عدّها مثل عدد البكتيريا في مائه معينه.

ملاحظة : المجتمع ليس بالضرورة أن يكون مجتمعاً بشرياً.

ب- العينة : تعرّف العينة بأنها مجموعه معينة من مفردات مجتمع الدراسة ندرس صفاتها وخواصها وتعتبر ممثلة للمجتمع الأصلي وتعمم النتائج على المجتمع عند الحصول عليها ، حيث تكون ممثلة ومعبرة له ، وذلك أن دراسة المجتمع ككل قد يكون ضخماً يحتاج إلى وقت وجهد ومال ، لذا فقد تم الاستعانة عن دراسة المجتمع بدراسة العينة .

ان اسلوب العينات : هو عملية جمع البيانات والمعلومات عن مجموعة معينة من مفردات مجتمع معين للدراسة (العينة) مثل استفتاء رأي ، دراسة فاعلية علاج جديد معين لمرضى القلب، هذا الاسلوب يحتاج الى جهد وموارد مالية وبشرية اقل من التسجيل الشامل ، وهو ايضا مفيد عند دراسة المجتمعات غير المحددة الا ان دقته تعتمد على النتائج المستخلصة من البيانات.

تقسم العينات إلى قسمين

اولاً : العينات الإحصائية (العشوائية) : ويقصد بها تلك المجموعة من المفردات المختارة من مجتمع الدراسة ، بحيث أنه ليس للباحث أي دخل في اختيار هذه المفردة دون الأخرى أي أن هناك مبدأً تساوي الفرصة لظهور أية مفردة من مفردات المجتمع ضمن هذه العينة ، وتقسم العينات العشوائية (الاحتمالية) إلى ما يلي وحسب طريقة الاختيار :

1- العينة العشوائية البسيطة : وهي عملية اختيار عينة عشوائية من مجتمع الدراسة ، بحيث تملك كل مفردة من مفردات المجتمع نفس الفرصة والاحتمال في الظهور ضمن مفردات العينة ويتوجب عند اتباع هذا الأسلوب ملاحظة مسألة تجانس مفردات المجتمع من حيث الصفة والصفات ذات العلاقة بموضوع البحث ، ويتم اختيار العينة العشوائية البسيطة كما يلي :

أفرض أن مجتمع الدراسة متجانس ومحدود وأن عدد مفرداته N ، وأن هناك دراسة تتطلب اختيار عينة من هذا المجتمع قوامها n ، بمعنى أن عدد العينات ذات الحجم n تمثل عدد مفردات العينة المختارة من مجتمع حجمه N ، وأن احتمال الظهور أي مفردة ضمن العينة هو $\frac{1}{N}$ وأن عدد العينات الممكنة الاختيار من هذا المجتمع هي r وتكون وفق القانون الاحتمالي التالي :

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

$$n! = n(n-1)(n-2) \dots 1$$

حيث أن $n!$ تسمى المفكوك النوني أو مضروب n
على سبيل المثال

$$\begin{aligned} 4! &= 4(4-1)(4-2)(4-3)+ \\ &= 4(3)(2)(1) = 24 \end{aligned}$$

$$\text{وإن } 1! = 1 \text{ و } 0! = 1$$

مثال : أفرض أنه تم اختيار ثلاثة منشآت من أربعة منشآت في قطاع وزارة الصناعة ، المطلوب : ما هو عدد العينات الممكنة الاختيار من هذا القطاع ؟ بين توزيعها بدون تكرار ؟

$$\text{الحل : } N = 4 , n = 3$$

$$r = C_3^4 = \frac{4!}{3!(4-3)!} = \frac{4*3*2*1}{3*2*1*1} = 4$$

التوزيع : نفرض أن رموز هذه المنشآت هي (A , B , C , D) . اذا يكون توزيع هذه المنشأة حسب ما هو موضح:

عدد العينات (4)

التوزيع الأول $\rightarrow (A, B, C)$

التوزيع الثاني $\rightarrow (A, B, D)$

التوزيع الثالث $\rightarrow (A, C, D)$

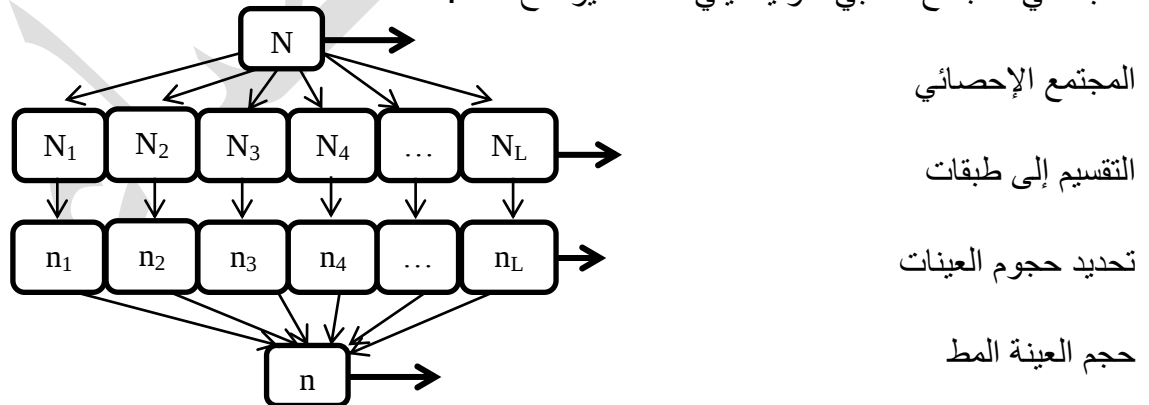
التوزيع الرابع $\rightarrow (B, C, D)$

2- العينة الطبقية العشوائية : تعتبر العينات المختارة وفق هذا الأسلوب أفضل أنواع العينات وأكثرها دقة في تمثيل المجتمع ، حيث أنه وفي أحوال كثيرة يلاحظ أن مفردات المجتمع الإحصائي غير متجانسة ، أي متكونه من طبقات تختلف فيما بينها مثال لو كنا بصدد دراسة المستوى العلمي لطلبة كلية الإدارة والاقتصاد (مجتمع احصائي فيه الطلبة هم المفردة الإحصائية) حيث ينقسم المجتمع الاحصائي الى طبقة قسم الإدارة وطبقه قسم المحاسبة وطبقه قسم الاقتصاد.

إن اختيار هذه العينة يتم على مرحلتين :

المرحلة الأولى : يتم في هذه المرحلة تقسيم المجتمع إلى طبقات متجانسة (مجتمعات جزئية).

المرحلة الثانية : يتم في هذه المرحلة أخذ عينة من كل طبقة بطريقة عشوائية ، أي أن كل طبقة تكون مجتمعاً بحد ذاته والشرط هو يجب ملاحظة التركيب النسبي ، أي يجب أن يتناسب حجم العينة مع حجم الطبقة في المجتمع النسبي ، وفيما يلي مخطط يوضح ذلك :



مخطط المعاينة العشوائية البسيطة

مثال 2 : يراد اختيار عينة عشوائية بحجم 20 مهني من مجتمع حجمه 100 مهني موزعين على النحو التالي 25 مهندس ، 35 موظف ، 40 عامل.

المطلوب : بيان عدد المهندسين والموظفين والعمال المكونين لهذه العينة.

الحل :

$$n = 20 \quad N = 100$$

نستخرج كسر المعاينة f وهي نسبة العينة من التجمع

$$f = \frac{n}{N} = \frac{20}{100} = \frac{2}{10} = \frac{1}{5}$$

$$\text{المهندسين} \rightarrow n_1 = \left(\frac{1}{5} \right) (25) = \frac{25}{5} = 5$$

$$\text{الموظفين} \rightarrow n_2 = \left(\frac{1}{5} \right) (35) = \frac{35}{5} = 7$$

$$\text{العمال} \rightarrow n_3 = \left(\frac{1}{5} \right) (40) = \frac{40}{5} = 8$$

$$n = n_1 + n_2 + n_3$$

$$= 5 + 7 + 8 = 20$$

3- **العينة العشوائية المنتظمة** : وهي أقل استخداماً من النوعين الأولين وعند اختيارها يتبع أسلوب معين في الاختيار ، وفي هذه الحالة يكون على الأغلب المجتمع متجانس ، أفرض أن مفردات مجتمع الدراسة البالغ حجمها N مفردة مرتبة وفق ترتيب معين ، أما تصاعدي أو تنازلي أو وفق أي معيار آخر للترتيب (مثلاً درجات طلبية من أدنى درجة لأعلى درجة) ، وأفرض أن دراسة معينة تتطلب اختيار عينة من المفردات حجمها n مفردة ، عندئذ فأن ذلك يتم على النحو التالي :

1- يتم تقسيم مفردات المجتمع المرتبة إما تصاعدياً أو تنازلياً إلى عدد من المجاميع كل مجموعة منها تضم عدد من المفردات وفق القانون $k = \frac{N}{n}$.

2- بافتراض أن مفردات المجتمع متسلسلة في الترتيب فأن هذه المجاميع ستكون وفق السلسلة

$$1, 2, 3, \dots, k \quad k+1, k+2, k+3 \quad (n-1)k, \dots, nk$$

3- من تقسيم السلسلة في الخطوة (2) ولغرض تحديد مفردات العينة المطلوبة للدراسة فإنه يتم وبشكل عشوائي اختيار مفردة واحدة من المجموعة الأولى.

4- بعد معرفة تسلسل هذه المفردة يتم تلقائياً تحديد بقية المفردات الأخرى من خلال إضافة العدد k إلى تسلسل المفردة الأولى لنحصل على تسلسل المفردة الثانية ويضاف العدد k لتسلسل المفردة الثانية لتحصل على تسلسل المفردة الثالثة وهكذا.

وبذلك نضمن أن تكون كل مجموعة قد ساهمت بمفردة واحدة من إجمالي حجم العينة.

مثال 3 : في امتحان الطلبة صف معين (مجتمع إحصائي) عددهم 24 طالب رتبت أسماؤهم حسب تسلسل درجاتهم تنازلياً وبهدف التعرف على أسباب انخفاض مستواهم في الامتحان تطلب استقراء رأي ستة طلاب منهم أي حجم العينة = 6 ، المطلوب تحديد تسلسل هؤلاء الطلاب وبشكل عشوائي.

الحل : واضح في هذا المثال أن الطالب الأول في القائمة يحمل أعلى درجة والطالب الأخير فيها يحمل أوطأ درجة وعليه فأن هناك ستة مجاميع كل منها تضم أربعة طلاب.

1 2 3 4	5 6 7 8	9 10 11 12	13 14 15 16	17 18 19 20	21 22 23 24
المجموعة الأولى	المجموعة الثانية	المجموعة الثالثة	المجموعة الرابعة	المجموعة الخامسة	المجموعة السادسة

$$K = \frac{N}{n} = \frac{24}{6} = 4$$

الآن وبشكل عشوائي تم اختيار مفردة واحدة من المجموعة الأولى ولنفرض أنه الطالب الذي يحمل التسلسل 2 على ضوء ذلك تتحدد بقية تسلسلات مفردات العينة من خلال إضافة العدد $k = 4$ وفقاً ما يلي :

اختيار عشوائي المفردة الأولى = 2 ، $2 + 4 = 6$ ، المفردة الثانية من المجموعة الثانية

المفردة الثالثة $6 + 4 = 10$

$14 = 10 + 4$ المفردة الرابعة ، $18 = 14 + 4$ المفردة الخامسة ، $22 = 18 + 4$ المفردة السادسة

وبذلك فأن العينة المختارة في هذا الصف تمثل الطلبة الذي تسلسلاتهم 2 , 6 , 10 , 14 , 18 , 22

حجم العينة 6

ومثال ذلك عدد الموظفين والعمال في إحدى الوزارات.

4- **العينة أو المعاينة متعددة المراحل:** يتم اختيار هذا النوع من العينة عن طريق تقسيم المجتمع إلى وحدات تدعى بالوحدات الأولية يتم اختيار عينة عشوائية من هذه الوحدات الأولية كمرحلة أولى ثم تقسم كل وحدة أولية إلى وحدات أصغر تدعى وحدات ثانوية وهكذا يتم اختيار المفردات المطلوبة في هذا النوع.

ثانياً : العينة غير العشوائية : وهي تلك المجموعة من المفردات المختارة من المجتمع التي يكون للباحث دور أو دخل في اختيارها أي (يتدخل الباحث باختيارها) ، وذلك لاعتبارات تتعلق بطبيعة تلك الدراسة (وليس بطريقة عشوائية).

وتقسم هذه العينات إلى :

1- **العينة الحصصية :** يتم حسب هذا النوع تقسيم مجتمع الدراسة إلى عدة طبقات استناداً إلى معايير معينة تتعلق بطبيعة الدراسة ، فمثلاً استطلاع رأي على برنامج تلفزيوني معين يتم تقسيم المجتمع إلى ذكور وإناث ثم اختيار عينة من الذكور وعينة من الإناث.

2- **العينة العمدية :** وهو أسلوب اختيار عينة من المجتمع بشكل متعمد أي تعتمد مسبقاً على أن المفردات المختارة من هذه العينة هي خير من يمثل مجتمع الدراسة ، فمثلاً عند دراسة واقع التعليم الجامعي في العراق يجب أن تكون العينة المختارة هي أساتذة الجامعة وطلبة الجامعة.

عرض البيانات الإحصائية : عند جمع البيانات الأولية الخاصة بظاهرة ما فإنه عادة لا يمكن الاستفادة منها وهي بهذه الصورة لذلك غالباً ما توضع في جداول مبسطة أو يعبر عنها في صور أشكال ورسوم بيانية لكي تسهل دراستها وتحليلها ولذلك هناك وسائل تعرف البيانات هي :

أ- وضع البيانات بشكل تقديري.

ب- تقديم البيانات في جداول.

ج- رسمها في صور وأشكال هندسية لتساعد على فهمها.

د- تمثيل البيانات وتلخيصها بمؤشرات ومقاييس إحصائية معروفة.

العرض الجدولي : سوف تخصص هذه الفقرة لدراسة أساليب عرض البيانات المصنفة في جداول خاصة تدعى بـ جداول التوزيعات التكرارية التي تتخذ أشكالاً متعددة حسب نوع المتغير العشوائي الذي صنفت على أساسه البيانات الأولية (الخام).

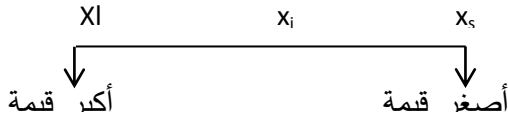
بعض التعاريف :

1- البيانات غير المبوبة : وهي البيانات الأولية أو الأصلية التي جمعت ولم تبوب.

2- البيانات المبوبة : وهي البيانات التي نظمت وبوبت ضمن جداول توزيع تكراري.

3- التوزيع التكراري: هو عبارة عن تلخيص وترتيب لبيانات المتغير العشوائي التي سبق ان جمعت وصنفت ، مقسمة إلى عدد من المجاميع كل منها تسمى بـ (الفئة class) ، هذه الفئات قد تكون مرتبة تصاعدياً أو تنازلياً حسب طبيعة البيانات ، ويسمى توزيع عدد قيم (X) حسب الفئات بـ (التوزيع التكراري) ، وقد تكون فئات التوزيع التكراري متساوية في الطول أو غير متساوية وذلك يعتمد على طبيعة الدراسة ومتطلباتها.

- الفئات : وهي المجاميع التي قسمت إليها قيم المتغير وكل فئة تأخذ مدى معين من قيم المتغير.
- حدود الفئات : لكل فئة حدان حد أعلى وحد أدنى مثلاً الفئة من خمسين إلى ستين ، الخمسين تمثل الحد الأدنى والستين تمثل الحد الأعلى.
- الحدود الحقيقية للفئات : لكل فئة عدنان حقيقيان حد أعلى وحد أدنى.
- طول الفئة : وهو مقدار المدى بين حدي الفئة ، هذا ويستحسن أن تكون أطوال الفئات متساوية لتسهيل العمليات الحسابية وسنرمز لطول الفئة بالرمز L .
- مركز الفئة : لكل فئة مركز وسنرمز له (x_i) وهو عبارة عن منتصف المدى بين حدي الفئة ، فمثلاً الفئة من 50 إلى 60 مركزها = الحد الأعلى + الحد الأدنى تقسيم (2)



$$X_i = \frac{50+60}{2} = 55$$

- تكرار الفئة : يمثل تكرار الفئة جزء من مفردات العينة التي تتصف بكونها تقع من حيث القيمة العددية ما بين حدي الفئة ، بحيث أن مجموع هذه الأجزاء يشكل عدد مفردات العينة n فإذا رمزنا لتكرارات الفئات بالرمز $f_1, f_2, \dots, f_m$ فإن $f_1 + f_2 + \dots + f_m = n$

$$f_1 + f_2 + \dots + f_m = n$$

الفصل الثاني

التوزيعات التكرارية وأساليب عرض البيانات

تصنيف وتبويب البيانات :

بعد جمع البيانات فإن هذه البيانات التي تم الحصول عليها بخصوص ظاهرة معينة بشكلها الأولي تسمى بيانات خام raw data أو البيانات الأولية elementary data أو البيانات غير المصنفة ، وتكون هذه البيانات في الغالب غير منظمة وبالتالي يتعذر الاعتماد عليها لأغراض التحليل الإحصائي للوصول إلى النتائج المطلوبة ، لذلك فإن أولى الخطوات بعد جمع البيانات هي :

- أ- مراجعة البيانات : وهي العملية التي تأتي بعد عملية جمع البيانات حيث يتم مراجعة وتدقيق البيانات.
 - ب- تصنيف البيانات : تبدأ هذه العملية بتصنيف البيانات على أساس الظواهر التي جمعت عنها فقد يكون للتصنيف على أساس العمر ، الجنس ، الوزن ... الخ.
 - ج- تبويب البيانات : ويقصد بالتبويب عملية تفريغ البيانات المصنفة في جداول خاصة بحيث أن كل جزء من البيانات المصنفة عن الظاهرة المعنية يعود إلى مستوى معين لتلك الظاهرة.
- طبيعة البيانات الإحصائية والرموز الإحصائية :

عند توفر بيانات حول ظاهرة ما فأننا نرمز للظاهرة بالرمز (Y) وكل مفردة أو مشاهدة منها نرمز لها بـ (y) فمثلاً عند دراسة عدد الشركات في القطاع الصناعي فأننا نرمز لصفة القطاع الصناعي بالرمز (Y) وأي شركة تابعة له بالرمز (y_i) وتسمى y_i المشاهدة أو المفردة ، هذا وأن قيمة (y_i) قد تختلف من منشأة إلى أخرى ولهذا نقول أن (Y) متغير.

المتغيرات العشوائية Random variable :

يعرف المتغير العشوائي بأنه دالة ذات قيمة حقيقية real – valued function معرفة على مجال العينة sample space وغالباً ما يرمز للمتغير العشوائي بأحد الأحرف الكبيرة مثل X , Y , Z ... الخ ، ولقيم المتغير عند تنفيذ التجربة بأحد الأحرف الصغيرة مثل x , y , z ... الخ.

وتقسم المتغيرات إلى :

1- المتغيرات النوعية (الوصفية) : وهي المتغيرات التي لا يمكن قياسها بوسائل قياس مألوفة كالعدد أو التقييس إنما تشكل صفات لذلك المتغير مثل الحالة الإنتاجية لإحدى الشركات (رديئة ، متوسطة ، جيدة ، جيدة جداً) وغيرها من الأمثلة.

2- المتغيرات الكمية : وهي المتغيرات التي يمكن قياسها بوسائل قياس مثل الموظفين في وزارة معينة وهذا النوع من المتغيرات يكون على نوعين أهمها :

أ- المتغيرات المستمرة:

إذا كانت مجموع القيم الممكنة للمتغير (X) مجموعة غير قابلة للعد سواء كانت مجموعة محدودة أو غير محدودة عندئذ يقال أن (X) متغير عشوائي مستمر ، بمعنى أن المتغير المستمر هو المتغير الذي تأخذ لمشاهدة أو المفردة فيه أي قيمة رقمية في مدى وجيز ، فلو فرضنا أطوال طلبة كلية معينة تتراوح بين 130,5–180,5 فعندئذ يقول بأن $130,5 \leq x \leq 180,5$ أي أن المتغير (X) يمكن ان يأخذ أي قيمة بين 130,5 و 180,5 وهنالك أمثلة أخرى مثل الوزن وكمية المحصول والزمن لأنها ممكن قياسها بأجزاء صغيرة جداً.

ب- المتغيرات المتقطعة أو المنفصلة:

إذا كانت مجموعة القيم الممكنة للمتغير (X) مجموعة قابلة للعد سواء كانت مجموعة محددة أو غير محددة عندئذ يقال أن (X) متغير عشوائي متقطع.

الخطوات العامة في إنشاء جدول التوزيع التكراري :

لتكوين جدول التوزيع التكراري يجب إتباع ما يلي :

1- استخراج المدى :

المدى = أكبر قيمة – أقل قيمة + 1

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

2- اختيار وتحديد عدد الفئات : تمثل عدد الفئات عدد المجاميع التي يتألف منها التوزيع التكراري ، وهناك صيغ تقريبية يمكن من خلالها تحديد عدد فئات التوزيع أهمها :

أ- صيغة يول Yule وهي :

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

حيث أن n تمثل عدد المفردات.

ب- صيغة سترجس وهي :

عدد الفئات = m

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

وعند التطبيق يتم تقريب الناتج لأقرب عدد صحيح.

لكل من الطريقتين مزايا وعيوب ولكن سوف تستخدم في العمل الحسابي عدداً من الفئات لا يقل عن

5 ولا يزيد عن 15.

3- إيجاد طول الفئة :

$$\text{طول الفئة} = \frac{\text{المدى}}{\text{عدد الفئات}}$$

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

ويستحسن أن تكون أطوال الفئات متساوية.

مثال : البيانات التالية تمثل أعمار 40 موظف في إحدى الشركات التابعة لوزارة التجارة ، المطلوب :
تكوين جدول التوزيع التكراري لهذه البيانات :

19 , 31 , 22 , 22 , 27 , 28 , 28 , 28 , 26 , 25 , 26 , 25 , 26 , 26 , 24 , 22 , 25 ,
24 , 24 , 22 , 24 , 24 , 24 , 23 , 21 , 22 , 22 , 29 , 29 , 24 , 24 , 19 , 20 , 24 ,
24 , 24 , 24 , 24 , 24 , 24

الحل :/ الخطوة الأولى للحل نرتب القيم تصاعدياً لسهولة الحل :

19 , 19 , 20 , 21 , 22 , 22 , 22 , 22 , 22 , 22 , 23 , 24 , 24 , 24 , 24 , 24 , 24 , 24 ,
24 , 24 , 24 , 24 , 24 , 24 , 24 , 24 , 25 , 25 , 25 , 26 , 26 , 26 , 26 , 27 , 28 , 28 ,
28 , 29 , 29 , 31

2- المدى : $T.R = x_L - x_S + 1$

$$= 31 - 19 + 1 = 13$$

3- عدد الفئات : $m = 1 + 3.322 \log(n)$

$$= 1 + 3.322 \log(40) = 1 + 3.322 (1.6) = 1 + 5.322 = 6.322 = 6 \text{ بالتقريب}$$

4- طول الفئة :

$$L = \frac{T.R}{m} = \frac{13}{6} = 2.16 = 2 \text{ بالتقريب}$$

1- الحد الأدنى للفئة الأولى يمثل أصغر قيمة حسب المثال هنا = 19

2- يضاف إلى الحد الأدنى طول الفئة $19 + 2 = 21$