علـم الإحصـاء (Statistics Science): هـو الطريقـة العلميـة التـي تخـتص بجمـع البيانـات والحقائق عن ظـاهرة أو مجموعـة ظـواهر معينـة وتنظيم وتصـنيف وتبويـب هذهِ البيانـات والحقـائق بطريقة يسهل عملية تحليلها وتفسيرها ومن ثم استخلاص النتائج واتخاذ القرار على ضوء ذلك.

1- علم الاحصاء يتضمن فرعين رئيسيين هما:

أ- **الإحصاء الوصفي (Descriptive statistics)**

يتضمن هذا الفرع الطرائق والأساليب المستخدمة في جمع البيانات والمعلومات عن ظـاهرة معينـة أو مجموعة ظـواهر وكيفيـة تنظيم وتصـنيف وتبويـب هذهِ البيانـات وعرضـها فـي جداول و رسوم بيانية وحساب بعض المؤشرات الإحصائية.

ب- **الإحصاء الاستدلالي (Inferential statistics)**

وهذا الفرع يهتم بموضوعين التقدير (**Estimation**) وهو تقدير مؤشرات المجتمع الاحصـائي واختبار الفرضيات (**Testing hypotheses**) .

2- **مجالات تطبيق علم الاحصاء**:

1- البحوث الطبية. 2- البحوث الهندسية. 3- البحوث الزراعية. 4- البحوث الصناعية.

5- البحوث الاقتصادية. 6- البحوث الادارية. 7- البحوث النفسية و الاجتماعية.

8- بحوث الرياضة والشباب.

3- **مراحل الطريقة الإحصائية في البحث العلمي** :

1- تحديد مشكلة أو فرضية البحث أو الدراسة.

2- جمع البيانات والمعلومات عن الظاهرة أو الظواهر ذات العلاقة بالبحث أو الدراسة.

3- تصنيف البيانات وتبويبها وعرضها.

4- حساب المؤشرات الإحصائية كتقديرات لمعالم مجتمع البحث أو الدراسة.

5- تحليل معطيات الدراسة والتوصل للنتائج على ضوء فرضية البحث أو الدراسة.

6- تفسير النتائج وعملية اتخاذ القرار بشأن فرضيات البحث.

المجتمع الإحصائي (Statistical Population): هو جميع مفردات الظاهرة موضوع البحث او الدراسة وتلك المفردات يفترض ان تشترك بصفة او صفات معينة وقد تكون تلك المفردات كائن حي او اي شيء اخر ، والمجتمع الاحصائي قد يكون محدد او غير محدد.

المفردة الاحصائية (Statistical Unit): اصغر وحدة في المجتمع الاحصائي.

جمع وتصنيف وتبويب البيانات

4- أساليب جمع البيانات (Data Collection Methods):

أن أي بحث علمي يستند في تحليلهِ الى الطريقة الاحصائية يحتاج الى بيانات ومعلومات حول موضوع البحث قيد الدراسة ، ويمكن الحصول على هذهِ البيانات والمعلومات من مصدرين هما:

أ- المصادر التأريخية (Historical Sources):

وهي البيانات والمعلومات المحفوظة والمتجمعة لدى أجهزة ومؤسسات و دوائر الدولة المختلفة نتيجة لمسوحات قامت بها هذهِ الجهات أو هيئات معينة 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 من الأمثلة الاخرى.

ب- المصادر الميدانية (Field Sources):

في حالة تعذر الحصول على البيانات من المصادر التأريخية نلجأ الى الميدان للحصول على تلك البيانات اي الحصول على البيانات من مصادرها الاصلية ، ويتم ذلك باتباع احد الاسلوبين الآتيين:

1) اسلوب التسجيل الشامل (Census):

بموجب هذا الاسلوب يتم جمع البيانات والمعلومات عن كافة مفردات المجتمع الاحصائي وفي هذهِ الحالة يجب أن يكون المجتمع الاحصائي محدداً ، أي يمكن ملاحظة كل مفردة فيهِ. مثال على ذلك، التعداد العام للسكان ، او حصر نشاطات الوحدات الصناعية في العراق. ويعتبر اسلوب التسجيل الشامل افضل اسلوب في جمع البيانات كونهُ يجهز الباحث ببيانات كاملة عن كافة مفردات

مجتمع الدراسة الا انهُ يحتاج الى وقت وجهد موارد مادية ، بشرية كبيرة في انجاز مهمة جمع البيانات ، فضلاً عن احتمال الوقوع في اخطاء نتيجة التعامل مع مفردات المجتمع الاحصائي بشكل كامل.

2) اسلوب العينات (Samples):

ونقصد بهِ عملية جمع البيانات والمعلومات عن مجموعة معينة من مفردات المجتمع الاحصائي تدعى هذهِ المجموعة من المفردات بالعينة " **Sample** " ، بحيث يتم اختيارها بطريقة تضمن تمثيلها للمجتمع الاحصائي بشكل دقيق ، ويمتاز أسلوب العينات بما يلي:

1- لا تحتاج الى وقت و جهد موارد مادية ، بشرية كبيرة.
2- يمكن استخدامهُ في حالة المجتمعات الغير محدودة.
3- إمكانية الحصول على بيانات حول صفات اكثر مما لو استخدم اسلوب الحصر الشامل.
4- إمكانية اختبار دقة النتائج.

انواع العينات

بهدف الحصول على أفضل دقة ممكنة للنتائج فان ذلك يستوجب اختيار أسلوب المعاينة (**Sampling technique**) الملائم الذي يكفل أن تكون العينة المختارة تمثل المجتمع قيد الدراسة أفضل تمثيل بالاضافة إلى أن حجم العينة يجب أن يكون ضمن حدود معقولة بحيث لا يكون كبير بحيث يؤدي ذلك إلى صرف وقت وجهد وموارد مادية كبيرة ولا يكون صغير بحيث يكون ذلك على حساب الدقة في النتائج . ويقصد بـ (المعاينة) أسلوب لاختيار مفردات من مجتمع الدراسة تؤلف العينة. وتنقسم العينات بشكل عام إلى قسمين رئيسيين **عينات عشوائية وعينات غير عشوائية**. وفيما يلي استعراض لكل قسم.

أولاً : العينات العشوائية (**Random samples**)

نقصد بالعينة العشوائية بأنها تلك المجموعة من المفردات المختارة من المجتمع الاحصائي بحيث ليس للباحث أي تدخل في عملية الاختيار بمعنى أن لكل مفردة نفس الفرصة في الظهور في العينة. ومن انواع العينات العشوائية ما يلي:

1- العينة العشوائية البسيطة (Simple Random Sampling)

وهي تلك العينة المختارة بشكل عشوائي يضمن لان تمتلك اية مفردة من مفردات المجتمع الاحصائي نفس الفرصة في الظهور ضمن مفردات العينة ويراعى عند استخدام هذا النوع من العينات ان يكون المجتمع الاحصائي متجانس من حيث الصفة او الصفات ذات العلاقة بالبحث او الدراسة ، وفيما يلي وصف لأسلوب المعاينة العشوائية البسيطة:

بافتراض أن المجتمع الاحصائي متجانس ومحدود وأن عدد مفرداتهِ يساوي $(\mathbf{N})$. نجد ان كل مفردة في هذا المجتمع الاحصائي لها نفس الاحتمال في الظهور في العينة وهذا الاحتمال هو $\left(\frac{1}{N}\right)$ ، فعند اختيار عينة عشوائية بسيطة من هذا المجتمع قوامها $(\mathbf{n})$. ومن الملاحظ ان كل مفردة من مفردات المجتمع الاحصائي ستظهر بعدد متساوي من المرات في كل العينات العشوائية البسيطة وان عدد هذهِ المرات يحسب باستخدام التوافيق $\left[C_{n-1}^{N-1}\right]$ ، اما عدد العينات العشوائية البسيطة التي يمكن اختيارها من هذا المجتمع هو (r) اذ يمكن حسابها باستخدام قانون التوافيق $\left[C_{n}^{N}\right]$ ايضاً وأن احتمال اختيار اي عينة هو $\left(\frac{1}{r}\right)$

عدد العينات العشوائية البسيطة $= r = C_{n}^{N} = \frac{N!}{n!(N-n)!}$

عدد مرات ظهور مفردة معينة في كل العينات $= C_{n-1}^{N-1}$

$N! = N(N-1)(N-2) \dots 3.2.1$

$6! = 6.5.4.3.2.1 = 720$

$1! = 1$

$0! = 1$

مثال (1): مجتمع احصائي متجانس عدد مفرداتهِ يساوي (**4**) هي $[\mathbf{A, B, C, D}]$ ، يراد اختيار عينة عشوائية بسيطة قوامها (**3**) مفردات.

المطلوب: 1. ما هو عدد العينات العشوائية البسيطة التي يمكن اختيارها من هذا المجتمع. ، 2.ما هو احتمال اختيار اي مفردة. ، 3. ما هو احتمال سحب اي عينة. ، ما هو عدد مرات ظهور المفردة **B** في كل العينات.

الاجابة: $(\mathbf{n = 3})$ ، $(\mathbf{N = 4})$

عدد العينات العشوائية البسيطة $= \mathbf{r = C_n^N = \frac{N!}{n!\,(N-n)!}}$

عدد العينات العشوائية البسيطة $= \mathbf{r = C_3^4 = \frac{4!}{3!\,(4-3)!} = \frac{4.3.2.1}{(3.2.1)*1!} = 4}$

احتمال سحب أي مفردة $= \mathbf{\frac{1}{N} = \frac{1}{4}}$

احتمال سحب أي عينة $= \mathbf{\frac{1}{r} = \frac{1}{4}}$

عدد مرات ظهور المفردة **B** في كل العينات $= \mathbf{\left[C_{n-1}^{N-1}\right] = \left[C_2^3\right] = \frac{3!}{2!*1!} = 3}$

العينات العشوائية البسيطة
$[\mathbf{A, B, C}]$
$[\mathbf{A, B, D}]$
$[\mathbf{A, C, D}]$
$[\mathbf{B, C, D}]$

مثال (**H.W**): مجتمع احصائي متجانس عدد مفرداتهِ يساوي (**5**) هي $[\mathbf{A, B, C, D, E}]$ ، يراد اختيار عينة عشوائية بسيطة قوامها: **فرع (أ)** (2) من المفردات، **فرع (ب)** (3) مفردات.

1- ما هو عدد العينات العشوائية البسيطة التي يمكن اختيارها من هذا المجتمع.

2- ما هو عدد مرات ظهور المفردة **B** في جميع العينات.

3- ما هو احتمال اختيار اي مفردة .

4- ما هو احتمال سحب اي عينة.

2- العينة العشوائية الطبقية (Stratified Random Sampling)

تعتبر العينة العشوائية الطبقية أفضل أنواع العينات وأكثرها دقة في تمثيل المجتمع الاحصائي الغير متجانس ، حيث أنهُ في كثير من الاحوال تكون مفردات المجتمع الاحصائي غير متجانسة من حيث الصفة أو الصفات المدروسة. ففي حالة كون المجتمع الاحصائي ذو مفردات غير متجانسة لا يجوز سحب عينة عشوائية بسيطة تمثل هذا المجتمع. فعلى سبيل المثال اذا اريد دراسة دخل الاسرة فأننا نجد ان هناك اسر ذات دخول عالية واخرى ذات دخول متوسطة واخرى ذات دخول منخفضة ، إذاً المجتمع الاحصائي هنا غير متجانس من حيث الصفة المدروسة ولا يجوز سحب عينة عشوائية بسيطة لأننا سنحصل على تقدير لمتوسط الدخل يكون منحازاً لاحدى الفئات الثلاثة ، وعليهِ يجب تقسيم المجتمع الاحصائي الى ثلاث طبقات (فئات ، مجموعات) ، الطبقة الاولى تضم الاسر ذات الدخول المرتفعة والطبقة الثانية تضم الاسر ذات الدخول المتوسطة والطبقة الثالثة تضم الاسر ذات الدخول المنخفضة ، وبعد ذلك يتم سحب عينة عشوائية بسيطة من كل طبقة يتناسب حجمها وحجم الطبقة في المجتمع ، وان مجموع حجوم العينات العشوائية الثلاثة تؤلف (تكون) العينة العشوائية الطبقية.

وفيما يلي وصف لأسلوب المعاينة العشوائية الطبقية ، بافتراض أن المجتمع الاحصائي مؤلف من $(\mathbf{N})$ من المفردات والذي يمكن تجزئتهِ الى $[\mathbf{L}]$ من الطبقات حجمها $[\mathbf{N_1}, \mathbf{N_2}, \dots, \mathbf{N_L}]$ ، حيث ان:

$$\sum_{i=1}^{L} N_i = N_1 + N_2 + \dots + N_L = N$$

وان تقسيم المجتمع الاحصائي الى $[\mathbf{L}]$ من الطبقات يتم على اساس اشتراك مفردات كل طبقة بصفة او عدة صفات مشتركة، فإذا أُريد اختيار عينة طبقية عشوائية ذات حجم $(\mathbf{n})$ فان هناك عدة طرائق لاختيار هذهِ العينة منها ما تسمى بطريقة التوزيع المتناسب (**Proportional allocation**) وبموجب هذهِ الطريقة يتم سحب عينة عشوائية بسيطة من كل طبقة يتناسب حجمها وحجم الطبقة في المجتمع ، فإذا كانت نسبة كل طبقة في المجتمع (وزن الطبقة) يساوي:

$$W_h = \frac{N_h}{N} \quad , h = 1, 2, \dots, L$$

حيث أن:

W_h **: وزن الطبقة.**

h **: يمثل دليل (مؤشر) لتسلسل الطبقة .**

$$\sum_{h=1}^{L} W_h = W_1 + W_2 + \dots + W_L = 1$$

وأن حجم العينة العشوائية البسيطة المسحوبة من الطبقة الاولى تساوي $[\mathbf{n_1 = W_1 * n}]$ ، وأن حجم العينة العشوائية البسيطة المسحوبة من الطبقة الثانية تساوي $[\mathbf{n_2 = W_2 * n}]$ ،... ، وهكذا لأخر طبقة فأن حجم العينة العشوائية البسيطة المسحوبة من الطبقة $[\mathbf{L}]$ تساوي $[\mathbf{n_L = W_L * n}]$ وبذلك فأن حجم العينة العشوائية الطبقية يساوي:

$$\sum_{i=1}^{L} n_i = n_1 + n_2 + \ldots + n_L = n$$

مما سبق نلاحظ ان $\left[\frac{n_h}{n} = \frac{N_h}{N}\right]$ ، اي ان نسبة مساهمة الطبقة **h** في العينة تساوي نسبة مساهمة الطبقة **h** في المجتمع.

مثال (2): مجتمع احصائي مؤلف من (2200) اسرة ، يراد دراسة متوسط الدخل للأسرة، علماً ان هذا المجتمع الاحصائي يضم اسر ذات دخول مرتفعة وذات دخول متوسطة وذات دخول منخفضة ، فإذا علمت ان عدد الأسر ذات الدخول المرتفعة يساوي (**700**) وعدد الاسر ذات الدخول المتوسطة يساوي (900) وعدد الاسر ذات الدخول المنخفضة يساوي (600). يراد سحب عينة عشوائية طبقية ذات حجم (**110**) ، المطلوب: ما هو حجوم عينات الطبقات باعتماد طريقة التوزيع المتناسب.

الاجابة: **N=2200** ، **n=110** ، $\mathbf{N_1 = 700}$ ، $\mathbf{N_2 = 900}$ ، $\mathbf{N_3 = 600}$

$$W_h = \frac{N_h}{N} \quad , h = 1, 2, \ldots, L$$

$$W_1 = \frac{700}{2200} = \frac{7}{22} \quad \Rightarrow n_1 = W_1 * n \Rightarrow n_1 = \frac{7}{22} * 110 = \frac{770}{22} = 35$$

$$W_2 = \frac{900}{2200} = \frac{9}{22} \quad \Rightarrow n_2 = W_2 * n \Rightarrow n_2 = \frac{9}{22} * 110 = \frac{990}{22} = 45$$

$$W_3 = \frac{600}{2200} = \frac{6}{22} \quad \Rightarrow n_3 = W_3 * n \Rightarrow n_3 = \frac{6}{22} * 110 = \frac{660}{22} = 30$$

نلاحظ ان:

$$\sum_{i=1}^{L} N_i = N_1 + N_2 + \ldots + N_L = N \quad \Rightarrow 700 + 900 + 600 = 2200$$

$$\sum_{i=1}^{L} n_i = n_1 + n_2 + \dots + n_L = n \quad \Rightarrow 35 + 45 + 30 = 110$$

$$\sum_{h=1}^{L} W_h = W_1 + W_2 + \cdots + W_L = 1 \quad \Rightarrow \frac{7}{22} + \frac{9}{22} + \frac{6}{22} = \frac{22}{22} = 1$$

$$\left[\frac{n_h}{n} = \frac{N_h}{N}\right] \quad \Rightarrow \frac{35}{100} = \frac{700}{2200} \quad ,\Rightarrow \frac{45}{100} = \frac{900}{2200} \quad ,\Rightarrow \frac{30}{100} = \frac{600}{2200}$$

مثال (**H.W**): يراد سحب عينة عشوائية طبقية ذات حجم (**75**) ، المطلوب: ما هو حجوم عينات الطبقات باعتماد طريقة التوزيع المتناسب. إذا علمت الاتي:

الطبقة	N_h
1	500
2	800
3	1200

3- العينة العشوائية المنتظمة (Systematic Random Sampling)

بموجب هذا النوع من المعاينة يتم تقسيم مفردات المجتمع الاحصائي البالغ عددها (**N**) الى عدد من المجاميع كل مجموعة تحتوي على (**K**) من المفردات حيث ان $k = \frac{N}{n}$ ، علماً بأن مفردات المجتمع مرتبة وفق نظام معين كأن يكون ترتيباً تصاعدياً او تنازلياً أو مثلاً ترتيب الدور السكنية حسب تسلسلها في شارع معين. وعند اختيار مفردات العينة المنتظمة يتم اولاً اختيار مفردة واحدة بصورة عشوائية من المجموعة الاولى وعلى ضوء تسلسل المفردة المختارة من المجموعة الاولى يتم اختيار مفردة من المجموعة الثانية بعد اضافة العدد (**K**) على تسلسل المفردة الاولى وهكذا يضاف العدد (**K**) على تسلسل المفردة الثانية لنحصل على المفردة الثالثة وهكذا وصولاً لأخر مفردة. بمعنى اخر يتم اختيار مفردة من المجموعة الاولى بطريقة عشوائية اما بقية المفردات يتم اختيارها على ابعاد متساوية من المفردة الاولى.

مثال (3): تم ترتيب (**24**) طالب حسب تسلسل درجاتهم تنازلياً ، يراد اختيار عينة عشوائية منتظمة بحجم (**6**) طلاب للتعرف على اسباب انخفاض مستواهم في الامتحان. المطلوب تحديد تسلسل هؤلاء الطلبة.

الاجابة: **N=24** ، **n=6** ، **k=4**

أ- يتم تقسيم الطلبة الى (**6**) مجاميع كل منها تحتوي على اربعة طلاب وكمايلي:

1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
13	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**

ملاحظة: 1. الرقم يشير الى تسلسل الطالب وليس درجتهِ حيث ان الطالب رقم (**1**) لهُ اعلى درجة في الامتحان.

2. تم تحديد عدد الطلبة في كل مجموعة كالآتي: $\mathbf{k = \frac{N}{n} = \frac{24}{6} = 4}$.

ب- يتم اختيار طالب بصورة عشوائية من المجموعة الاولى وليكن مثلاً الطالب ذو تسلسل رقم (3).

ج- اما بقية الطلبة فنحصل عليهم بإضافة (**k=4**) الى التسلسل (3) لنحصل على الطالب الثاني في العينة وهو التسلسل (7) وهكذا ... وصولاً لأخر طالب في العينة وهو التسلسل (23).

إذاً العينة العشوائية المنتظمة تكون على اساس ان الطالب الذي تم اختياره عشوائياً من المجموعة الاولى والذي يحمل التسلسل (3):

$$[3 \quad 7 \quad 11 \quad 15 \quad 19 \quad 23]$$

<u>ملاحظة:</u> عدد العينات العشوائية المنتظمة التي يمكن اختيارها من المجتمع الاحصائي يساوي عدد مفردات كل مجموعة (**K**).

4- العينة متعددة المراحل (Cluster Random Sampling)

بموجب اسلوب المعاينة المتعددة المراحل يتم تقسيم المجتمع الاحصائي الى وحدات اولية ثم يتم اختيار عينة عشوائية بسيطة من هذهِ الوحدات كمرحلة اولى ، ثم يتم تقسيم الوحدات الاولية المختارة الى وحدات اصغر تدعى بالوحدات الثانوية ويتم اختيار عينة عشوائية بسيطة من الوحدات الثانوية لكل وحدة اولية كمرحلة ثانية ، ثم تقسم الوحدات الثانوية المختارة الى وحدات اصغر يتم اختيار عينة عشوائية بسيطة منها كمرحلة ثالثة وتستمر عملية التقسيم والاختيار لحين الوصول الى المفردات التي يتم جمع البيانات منها. **فمثلاً عند إجراء دراسة لتقدير متوسط استهلاك العائلة العراقية لمادة السكر** فإن الوحدة الاحصائية التي يمكن الحصول على البيانات منها هي العائلة العراقية فعند اختيار عينة عنقودية يتم تقسيم **العراق** الى محافظات كوحدات اولية نختار منها عشوائياً عينة من **المحافظات** كمرحلة اولى ثم نقسم المحافظات المختارة في المرحلة الاولى الى **اقضية** كوحدات ثانوية يتم اختيار عينة عشوائية بسيطة منها كمرحلة ثانية ثم تقسم الوحدات الثانوية المختارة في المرحلة الثانية الى **نواحي** يتم اختيار عينة عشوائية بسيطة منها كمرحلة ثالثة ثم تقسم النواحي المختارة في المرحلة الثالثة الى **محلات سكنية** يتم اختيار عينة عشوائية بسيطة منها كمرحلة رابعة ثم تقسم المحلات السكنية المختارة الى **ازقة** ومنها يتم اختيار عينة عشوائية بسيطة والتي تقسم بدورها الى **الدور السكنية** التي يتم اختيار عينة عشوائية بسيطة منها وبهذا يتم الوصول الى العائلة التي منها يتم جمع البيانات.

ثانياً : العينات غير العشوائية (Non – Random samples)

وهي تلك المجموعة من المفردات المختارة من المجتمع الاحصائي بطريقة يكون للباحث دخل في اختيار تلك المفردات وذلك لاعتبارات تتعلق بطبيعة البحث أو الدراسة ، وهذهِ العينات نوعين هما:

1- العينة العمدية: وهي العينة المختارة بشكل متعمد نعتقد مسبقاً ان مفردات هذهِ العينة هي خير من يمثل مجتمع الدراسة. فمثلاً عند دراسة السبل الكفيلة للارتقاء برياضة كرة القدم فمن الافضل اختيار عينة من المختصين برياضة كرة القدم وبشكل عمدي (مقصود) كون أن هذهِ العينة هي ذات خبرة بشؤون هذهِ الرياضة.

2- العينة الحصصية: (**The quota sample**)

بموجب هذا النوع من المعاينة يتم تقسيم المجتمع الاحصائي الى عدة طبقات بالاستناد الى معايير تتعلق بطبيعة الدراسة ثم يتم اختيار عينة عمدية من كل طبقة (بشكل غير عشوائي) يتناسب حجمها وحجم الطبقة في المجتمع ومجموع حجوم هذهِ العينات العمدية يمثل حجم العينة الحصصية.

5- وسائل جمع البيانات

يعد تحديد حجم العينة وأسلوب المعاينة الملائم يتطلب الامر اختيار الوسيلة الملائمة في جمع البيانات والمعلومات عن الظاهرة او الظواهر المتعلقة بتلك الدراسة ، واهم وسائل جمع المعلومات ما يلي:

1. اسلوب جمع البيانات المباشر (**Direct data collection method**) :

بموجب هذا الاسلوب يتم جمع البيانات والمعلومات المتوفرة لدى اجهزة معينة ذات علاقة بالدراسة ، فمثلاً عند دراسة تطور الدخل القومي فيمكن الرجوع الى الجهاز المركزي للإحصاء مثلاً لاختيار سلسلة زمنية من البيانات عن الدخل القومي في العراق. كذلك وفقاً لهذا الاسلوب يمكن توفير البيانات ذات الطابع المختبري والحقلي التي يتم عملها من قبل الباحث فمثلاً عند إجراء دراسة حول اختبار فاعلية سماد معين على زيادة كمية محصول الحنطة فإن الباحث يقوم بنفسهِ بعمل تصميم لهذهِ التجربة وتسجيل نتائجها التي تمثل البيانات اللازمة لهذا البحث.

2. الاستبيان (**Questionnaire**):

وهو عبارة عن استمارة يتم من خلالها جمع البيانات والمعلومات من مفردات مجتمع الدراسة او من عينة البحث من خلال مواجهة الباحث الشخصية للمفردة الاحصائية او عن طريق المراسلة ويجب مراعاة النقاط التالية عند تصميم الاستمارة:

1) اعداد مقدمة ايضاحية تكتب في بداية استمارة الاستبيان.
2) ان تكون فقرات استمارة الاستبيان متسلسلة وغير مبعثرة بحيث ان كل صنف منها يحقق غرض معين.

3) ان تكون الاسئلة متوسطة العدد.

4) ان تكون الاسئلة واضحة المعنى.

5) ان تهيء الاسئلة بالشكل الذي يجب ان تكون اجابة الفرد عليها محددة وقصيرة.

6) ان يؤخذ بنظر الاعتبار ظروف تفريغ وتصنيف وتبويب وترميز الاجابات.

6- الاخطاء الشائعة في جمع البيانات

يحدث في بعض الاحيان ان يقع الباحث في بعض الاخطاء لدى جمعهِ للبيانات وهذهِ الاخطاء تحدث نتيجة سوء استخدام الطريقة الاحصائية وهذهِ الاخطاء هي:

أ- خطأ التحيز (**bias error**)

عند جمع البيانات يفترض ان تجمع من مصادرها الاصلية الا أنهُ في بعض الاحيان تجمع البيانات من مصادر اخرى غير مصادرها الاصلية ، فمثلاً عند دراسة رغبات الاطفال على اساس عينة من اطفال مجموعة من الاسر فإن المصدر الصحيح لجمع البيانات حول هذهِ الظاهرة هو الام كونها تمتلك معلومات كافية عن طفلها بسبب معايشتها لهُ ، ولكن في حالة اخذ معلومات من غير هذا المصدر فمن المحتمل جداً ان نقع في خطأ في تسجيل رغبات الطفل وبالتالي تأثر نتائج الدراسة بمجمل هذهِ الاخطاء.

ب- خطأ الصدفة (**the error of chance**)

يحصل هذا النوع من الاخطاء عندما يقوم الباحث باستيفاء بعض البيانات والمعلومات بالاعتماد على معلوماتهِ الشخصية او التعمد في جمع هذهِ البيانات من بعض المفردات دون الاخرى او يقوم بجمع بيانات ناقصة لسبب او لأخر.

تصنيف وتبويب البيانات

ان البيانات المستحصل عليها بخصوص الظاهرة المعنية تسمى بالبيانات الخام او الاولية وتلك البيانات تكون غير منظمة ويصعب على الباحث تكوين فكرة عن الظاهرة المدروسة كذلك يصعب الاعتماد عليها لاغراض التحليل الاحصائي ، لذا فإن الخطوة الهامة التي تلي عملية جمع البيانات هي عملية مراجعة وتصنيف البيانات وتبويبها.

1. **مراجعة البيانات:** بعد اتمام عملية جمع البيانات يجب مراجعة تلك البيانات وتدقيقها لغرض التأكد من مطابقتها وتكاملها لمتطلبات الدراسة.
2. **تصنيف البيانات:** بعد التأكد من تكامل و وضوح ودقة البيانات التي تم الحصول عليها ، نبدأ بعملية تصنيف على اساس الظواهر التي جمعت عنها البيانات حيث يتم فرز البيانات على هيئة مجموعة فقد يكون التصنيف على اساس ظاهرة العمر او الجنس او المهنة او الوزن او الطول او محل الاقامة وغيرها.
3. **تبويب البيانات:** بعد اتمام عملية تصنيف البيانات نبدأ بعملية تبويب البيانات ويقصد بتبويب البيانات تفريغ البيانات المصنفة في جداول خاصة بحيث ان كل جزء من البيانات المصنفة عن الظاهرة المعنية يعود الى مستوى معين لتلك الظاهرة. والهدف من عملية التبويب هو ابراز البيانات وتوضيحها في اضيق حيز ممكن كي نتمكن من تكوين فكرة عنها ، وهناك اربعة اشكال للتبويب هي:

أ- **التبويب الزمني:** عبارة عن تجميع البيانات المصنفة وترتيبها في جداول على اساس ان كل جمع منها يعود لوحدة زمنية معينة ، كاليوم او الاسبوع او الشهر او السنة.

ب- **التبويب الجغرافي:** عبارة عن تجميع البيانات المصنفة وترتيبها في جداول على اساس ان كل جمع منها خاص بوحدة جغرافية معينة ، او تقسيم اداري معين.

ج- **التبويب الكمي:** عبارة عن تجميع البيانات المصنفة وترتيبها في جداول على اساس ان كل جمع منها خاص بوحدة كمية معينة ، كوحدات الوزن ، الطول ، العمر ، المسافة ، ... الخ.

د- **التبويب على اساس صفة معينة:** عبارة عن عملية تجميع للبيانات المصنفة وترتيبها في جداول خاصة على اساس ان كل جمع منها يشترك بصفة معينة ، كالجنس، الحالة الاجتماعية، عنوان الوظيفة، القومية.

التوزيعات التكرارية واساليب عرض البيانات

المتغيرات العشوائي (Random Variables)

يعرف المتغير العشوائي بأنهُ دالة ذات قيمة حقيقية معرفة على فضاء يدعى فضاء العينة (**Sample Space**) ، يرمز للمتغير بالحروف الكبيرة ولقيم المتغير عند تنفيذ التجربة بالحروف الصغيرة.

مثال: عند رمي قطعة من زهرة النرد وملاحظة العدد الذي سوف يظهر على وجه الزهر بعد رمية ، هنا المتغير **X** هو العدد الذي سوف يظهر على وجه الزهر بعد رمية. ان الحالات الممكنة الوقوع اي الظهور على وجه الزهر هي مجموعة الاعداد الحقيقية $\{1, 2, 3, 4, 5, 6\}$ ، ولا يمكن اطلاقاً ظهور غيرها هذهِ المجموعة يطلق عليها فضاء العينة للمتغير **X** وعناصرها تمثل القيم الممكنة للمتغير **X** ، هذا يعني ان المتغير **X** دالة معرفة على هذا الفضاء ، وحيث ان تجربة رمي الزهر هي تجربة عشوائية اي ان عملية رمي الزهر تتم عشوائياً دون تحيز للمجرب لهذا الوجه او ذاك ، عليهِ فإن **X** متغير عشوائي ، وغالباً ما يتم الرمز الى مجموعة القيم الممكنة للمتغير (فضاء العينة) بالرمز **Ω** ، ويلفظ اوميكا (**Omega**) ، اي ان:

$$\Omega = \{X: x = 1, 2, 3, 4, 5, 6\}$$

تقسم المتغيرات العشوائية الى قسمين رئيسيين هما:

أ- المتغيرات النوعية (الوصفية): وهي المتغيرات التي لا يمكن قياسها بوسائل القياس المألوفة ، وانما تشكل صفات لذلك المتغير مثل لون العين كمتغير، يمكن ان يكون (اسود، ازرق، عسلي).

ب- المتغيرات الكمية: وهي المتغيرات التي يمكن قياسها بوسائل القياس المألوفة ، مثل عدد الطلبة في صف معين او عدد اشجار البرتقال في بستان او وزن حمولة من الاسمنت بالطن.

❖ المتغيرات الكمية نوعين هما:

1) المتغيرات المتقطعة (**Discrete Variables**): اذا كانت مجموعة القيم الممكنة للمتغير **X** مجموعة قابلة للعد سواء كانت مجموعة محدودة ام مجموعة غير محدودة عندئذ يقال ان **X** متغير عشوائي متقطع.

2) المتغيرات المستمرة (**Continuoes Variables**): اذا كانت مجموعة القيم الممكنة للمتغير **X** مجموعة غير قابلة للعد سواء كانت مجموعة محدودة ام مجموعة غير محدودة عندئذ يقال ان **X** متغير عشوائي مستمر.

❖ **اساليب عرض البيانات : وهي نوعين وكالآتي:**

<u>اولاً: العرض الجدولي للبيانات:</u>

أ- التوزيع التكراري (**Frequency distribution**): وهو عبارة عن تلخيص وترتيب البيانات الخاصة بالمتغير العشوائي التي سبق ان جمعت وصنفت في جداول مقسمة الى مجاميع كل منها تسمى الفئات (**classes**) ، هذه الفئات قد تكون مرتبة تصاعدياً او تنازلياً حسب طبيعة البيانات.

ولشرح مكونات التوزيع التكراري ، نفرض ان المتغير العشوائي **X** يأخذ القيم $(x_1, x_2, \dots, x_n)$ من عينة عشوائية ذات حجم n ، ولنفرض (X_L, X_S) ، تعني اصغر قيمة واكبر قيمة في تلك البيانات على الترتيب ، ويراد عمل توزيع تكراري عدد فئاتهِ (m) ، نتبع الخطوات الاتية في عمل التوزيع التكراري.

1) المدى الكلي (**Total Range**) : وهو الفرق بين اكبر قيمة واصغر قيمة في البيانات مضافاً إليهِ واحد ، ويتم حساب المدى الكلي للتوزيع وفق الصيغة الاتية:

$$\mathbf{T.R = X_L - X_S + 1}$$

2) عدد الفئات (**Number of classes**) : لتحديد عدد فئات التوزيع التكراري يمكن اعتماد احدى الصيغتين التقريبيتين ، لحساب عدد فئات التوزيع وكما يلي:

1. صيغة يول (**Yule**) وهي:

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

2. صيغة سترجس (**Sturges**)

$$\mathbf{m = 1 + 3.322 \log (n)}$$

ملاحظة: عند التطبيق يتم تقريب الناتج لأقرب عدد صحيح.

3) طـول الفئـة (**Length of a classes**) : ويسـمى احيانـاً المـدى الفئـوي ، يمثـل طـول الفئـة مقدار المسافة بين الحد الادنى والحد الاعلى ويحسب كالاتي:

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

4) الحـد الادنـى والحـد الاعلـى للفئـة (**Lower and upper of a classes**) : لكل فئـة حد ادنى وحد اعلى ، الحد الادنـى يمثل بدايـة الفئـة والحـد الاعلـى يمثل نهايـة الفئـة ، ويمكـن تحديد الحد الادنى والحد الاعلى للفئات (**اذا كانت اطوالها متساوية**) كالآتي:

تسلسل الفئة	الحد الادنى	الحد الاعلى
$\mathbf{1}$	$\mathbf{X_S}$	$\mathbf{X_S + L}$
$\mathbf{2}$	$\mathbf{X_S + L}$	$\mathbf{X_S + 2L}$
$\mathbf{3}$	$\mathbf{X_S + 2L}$	$\mathbf{X_S + 3L}$
$\vdots$	$\vdots$	$\vdots$
$\mathbf{m}$	$\mathbf{X_S + (m-1)L}$	$\mathbf{X_S + mL}$

وهنـاك طرائـق مختلفـة لكتابـة حـدود الفئـات للتوزيـع التكـراري استناداً الـى نـوع المتغيـر العشـوائي وكالاتي:

1- **المتغيـرات العشـوائية المتقطعـة:** فـي حالـة المتغيـر العشـوائي مـن النـوع المتقطـع، تكتـب الفئـات كالاتي:

تسلسل الفئة	الحد الادنى	الحد الاعلى
$\mathbf{1}$	$\mathbf{X_S}$	$\mathbf{X_S + L - 1}$
$\mathbf{2}$	$\mathbf{X_S + L}$	$\mathbf{X_S + 2L - 1}$
$\mathbf{3}$	$\mathbf{X_S + 2L}$	$\mathbf{X_S + 3L - 1}$
$\vdots$	$\vdots$	$\vdots$
$\mathbf{m}$	$\mathbf{X_S + (m-1)L}$	$\mathbf{X_S + mL - 1}$

ملاحظة رقم (1): يجب ضمان كل قيمة من قيم البيانات ان تظهر في فئة واحدة دون ان تتكرر في الفئات الاخرى.

ملاحظة رقم (2): ان طول الفئة في حالة المتغيرات العشوائية المتقطعة يحسب كالاتي:

$$\mathbf{L = U.L - L.L + 1}$$

حيث ان، $\mathbf{U.L}$: الحد الاعلى ، وان، $\mathbf{L.L}$: الحد الادنى.

2- المتغيرات العشوائية المستمرة: في حالة المتغير العشوائي من النوع المستمر، تكتب الفئات كالاتي:

الحد الاعلى	الى اقل من	الحد الادنى	تسلسل الفئة
$X_S + L$	الى اقل من	X_S	1
$X_S + 2L$	الى اقل من	$X_S + L$	2
$X_S + 3L$	الى اقل من	$X_S + 2L$	3
$\vdots$		$\vdots$	$\vdots$
$X_S + mL$	الى اقل من	$X_S + (m-1)L$	m

ملاحظة رقم (1): بافتراض ان الحد الاعلى لكل فئة هو الحد الادنى للفئة التالية لهُ ، لذا من الممكن ان نكتب الفئات بشكل مختصر بذكر الحد الادنى لكل فئة مع امكانية كتاباتها بشكل غير مختصر وكالاتي:

الفئات بدون اختصار
X_S —— $X_S + L$
$X_S + L$ —— $X_S + 2L$
$X_S + 2L$ —— $X_S + 3L$
$\vdots$
$X_S + (m-1)L$ —— $X_S + mL$

الفئات اختصارا
X_S ——
$X_S + L$ ——
$X_S + 2L$ ——
$\vdots$
$X_S + (m-1)L$ —— $X_S + mL$

ملاحظة رقم (2): ان طول الفئة في حالة المتغيرات العشوائية المستمرة يحسب كالاتي:

$$\mathbf{L = U.L - L.L}$$

حيث ان، U.L : الحد الاعلى ، وان، L.L : الحد الادنى.

5) تكرار الفئة (**Class frequency**): يمثل تكرار الفئة ذلك الجزء من مفردات العينة التي تقع ضمن تلك الفئة ، بحيث ان مجموع هذه الاجزاء (تكرارات الفئات) يشكل عدد مفردات العينة ، ويرمز لتكرار الفئة بـ (f).

حيث ان:

$$\sum_{i=1}^{m} f_i = f_1 + f_2 + \cdots + f_m = n \quad , \qquad n: \text{حجم العينة}$$

6) **مركز الفئة** (**Center of a classes**): يمثل مركز الفئة قيمة من قيم المتغير العشوائي **X** التي تتوسط المسافة بين الحد الادنى والحد الاعلى للفئة ، ويرمز لمركز الفئة بالرمز **X** ، و يحسب كالاتي:

$$\mathbf{X = \frac{L.L + U.L}{2}}$$

حيث ان، U.L : الحد الاعلى ، وان، L.L : الحد الادنى. ، وان L : طول الفئة.

وتجدر الملاحظة هنا ان قيم المفردات التي تقع ضمن تلك الفئة الواحدة ينظر لها وكأنها متساوية من حيث القيمة العددية لمركز الفئة.

انواع التوزيعات التكرارية

ان التوزيع التكراري قد يكون توزيع **مغلق** او توزيع **مفتوح** ، وذلك يعتمد على طبيعة الدراسة ويمكن تعريف التوزيع التكراري المغلق والتوزيع التكراري المفتوح كما يلي:

التوزيع التكراري المغلق : يعرف التوزيع التكراري المغلق بأنهُ ذلك التوزيع الذي يمتلك حداً ادنى للفئة الاولى وحداً اعلى للفئة الاخيرة.

التوزيع التكراري المفتوح: يعرف التوزيع التكراري المفتوحة بأنهُ ذلك التوزيع الذي لا يمتلك حداً ادنى للفئة الاولى **او** حداً اعلى للفئة الاخيرة **او** كليهما معاً. ويمكن تقسيم التوزيع التكراري المفتوح الى نوعين هما:

أ- التوزيع التكراري المفتوح من احد طرفيهِ: يعرف التوزيع التكراري المفتوح من احد طرفيهِ بأنهُ ذلك التوزيع الذي لا يمتلك حداً ادنى للفئة الاولى او حداً اعلى للفئة الاخيرة.

ب- التوزيع التكراري المفتوح من الطرفين: يعرف التوزيع التكراري المفتوح من الطرفين بأنهُ ذلك التوزيع الذي لا يمتلك حداً ادنى للفئة الاولى وكذلك لا يمتلك حداً اعلى للفئة الاخيرة.

ملاحظة : يمكن تقسيم التوزيعات التكرارية من حيث طول الفئة **(L)** ، الى نوعين هما:

1) توزيع تكراري منتظم: وفيهِ تكون اطوال الفئات متساوية.

2) توزيع تكراري غير منتظم: وفيهِ تكون اطوال الفئات غير متساوية.

مثال (4): البيانات التالية تمثل عدد اشجار البرتقال المملوكة من قبل 60 عائلة فلاحية ، يطلب تبويب هذهِ البيانات في جدول توزيع تكراري ، وحساب مراز الفئات.

60	**76**	**80**	**120**	**132**	**82**	**90**	**65**	**68**	**72**
150	**142**	**157**	**164**	**88**	**90**	**98**	**101**	**103**	**110**
119	**116**	**120**	**126**	**109**	**114**	**120**	**122**	**111**	**116**
90	**78**	**93**	**95**	**98**	**104**	**120**	**113**	**121**	**119**
125	**126**	**130**	**131**	**136**	**118**	**120**	**142**	**150**	**154**
122	**123**	**139**	**125**	**156**	**154**	**136**	**137**	**110**	**136**

الحل: من خلال معطيات السؤال نجد بأن المتغير العشوائي الذي يمثل عدد اشجار البرتقال هو متغير عشوائي متقطع لذا نستخدم توزيع تكراري خاص بالمتغير العشوائي المتقطع.

$$\mathbf{n = 60 \quad , \quad X_S = 60 \quad , \quad X_L = 164}$$

ولغرض تكوين جدول توزيع تكراري لغرض تبويب البيانات ، يجب حساب المؤشرات الاتي:

1. المـدى الكلـي. 2. عـدد الفئـات. 3. طـول الفئـة. 4.الحـد الادنـى والحـد الاعلـى لجميـع الفئات. 5. تكرار كل فئة. 6. مراكز الفئات.

1) (المدى الكلي) $\mathrm{T.R} = \mathrm{X_L} - \mathrm{X_S} + 1 \rightarrow \mathrm{T.R} = 164 - 60 + 1 = 105$

2. عدد الفئات: ويتم حساب عدد الفئات وفقاً لاحدى الصيغتين التقريبيتين التاليتين:

(صيغة يول) $\mathbf{m = 2.5 * \sqrt[4]{n}}$

$\mathbf{m = 2.5 * \sqrt[4]{60} = 2.5(2.783) = 6.958 \cong 7}$ (فئات)

او

(صيغة سترجس) $\mathbf{m = 1 + 3.322 * \log(n)}$

$\mathbf{m = 1 + 3.322 * \log(60) = 1 + 3.322 * (1.778) = 6.907 \cong 7}$ (فئات)

3. طول الفئة: ويتم حساب طول الفئة وفقاً للصيغة التالية:

$$\mathbf{L} = \frac{\mathbf{T.R}}{\mathbf{m}} = \frac{105}{7} = 15$$

4. الحد الادنى والحد الاعلى للفئة: ويتم حساب الحد الادنى والحد الاعلى للفئة ، كما مبين بالجدول الآتي:

الفئات	الحد الاعلى	الحد الادنى	تسلسل الفئة	الحد الاعلى	الحد الادنى	تسلسل الفئة
60 – 74	60+15-1=74	60	1	$X_S + L - 1$	X_S	1
75 – 89	75+15-1=89	75	2	$X_S + 2L - 1$	$X_S + L$	2
90 – 104	90+15-1=104	90	3	$X_S + 3L - 1$	$X_S + 2L$	3
105 – 119	105+15-1=119	105	4	⋮	⋮	⋮
120 – 134	120+15-1=134	120	5			
135 – 149	135+15-1	135	6			
150 – 164	150+15-1	150	7	$X_S + mL - 1$	$X_S + (m-1)L$	m

الفئات	التكرار بالإشارات	f	مراكز الفئات
60 – 74	xxxx	4	60+74/2=67
75 – 89	xx xx x	5	82
90 – 104	xxxxx xxxxx	10	97
105 – 119	x x xxx x xxxx x	11	112
120 – 134	x xx xx xxxxx xxxxx x	16	127
135 – 149	xxxxx xx	7	142
150 – 164	x xxx x x x	7	157

والجدول التكراري اعلاه ، يشير الى ان اربعة عوائل لديها عدد اشجار البرتقال تتراوح ما بين 60 الى 74 ، وخمسة عوائل لديها عدد اشجار البرتقال تتراوح ما بين 75 الى 89 ، وهكذا.

وفي حالة طلب حساب مراكز الفئات فإنها تحسب كالاتي:

$$\mathbf{X = \frac{L.L + U.L}{2}}$$

$$\mathbf{X = \frac{60 + 74}{2} = 67}$$

$$\mathbf{X = \frac{75 + 89}{2} = 82}$$

$$\mathbf{X = \frac{90 + 104}{2} = 97}$$

وهكذا الى اخر فئة.

تسلسل الفئة	الفئات	f	X
1	60 – 74	4	67
2	75 – 89	5	82
3	90 – 104	10	97
4	105 – 119	11	112
5	120 – 134	16	127
6	135 – 149	7	142
7	150 – 164	7	157
المجموع		60	

مثال (5): البيانات التالية تمثل اوزان عينة من طلبة احدى الكليات قوامها (20) طالب ، يطلب تبويب هذهِ البيانات في جدول توزيع تكراري ، ومن ثم حساب مراكز الفئات.

65.3	**70.5**	**83**	**94**	**55**	**46**	**47.8**	**62.3**	**77.2**	**61.3**
66.5	**95**	**68.3**	**80.2**	**78.3**	**80.1**	**76.3**	**81.5**	**51.8**	**74.1**

الحل: من خلال معطيات السؤال نجد بأن المتغير العشوائي الذي يمثل اوزان الطلبة هو متغير عشوائي مستمر لذا نستخدم توزيع تكراري خاص بالمتغير العشوائي المستمر.

$$\mathbf{n = 20 \quad , \quad X_S = 46 \quad , \quad X_L = 95}$$

ولغرض تكوين جدول توزيع تكراري لغرض تبويب البيانات ، يجب حساب المؤشرات الاتي:

2. المدى الكلي. 2. عدد الفئات. 3. طول الفئة. 4.الحد الادنى والحد الاعلى لجميع الفئات. 5. تكرار كل فئة. 6. مراكز الفئات.

1. (المدى الكلي) $\mathrm{T.R} = \mathrm{X_L} - \mathrm{X_S} + 1 \;\rightarrow \mathrm{T.R} = 95 - 46 + 1 = 50$

2. عدد الفئات: ويتم حساب عدد الفئات وفقاً لاحدى الصيغتين التقريبيتين التاليتين:

(صيغة يول) $\mathbf{m = 2.5 * \sqrt[4]{n}}$

$\mathbf{m = 2.5 * \sqrt[4]{20} = 2.5(2.115) = 5.288 \cong 5}$ (فئات)

او

(صيغة سترجس) $\mathbf{m = 1 + 3.322 * \log(n)}$

$\mathbf{m = 1 + 3.322 * \log(20) = 1 + 3.322 * (1.301) = 5.322 \cong 5}$ (فئات)

3. طول الفئة: ويتم حساب طول الفئة وفقاً للصيغة التالية:

$$\mathbf{L = \frac{T.R}{m} = \frac{50}{5} = 10}$$

الحد الاعلى	الى اقل من	الحد الادنى	تسلسل الفئة	الحد الاعلى	الى اقل من	الحد الادنى	تسلسل الفئة
56	الى اقل من	46	1	$X_S + L$	الى اقل من	X_S	1
66	الى اقل من	56	2	$X_S + 2L$	الى اقل من	$X_S + L$	2
76	الى اقل من	66	3	$X_S + 3L$	الى اقل من	$X_S + 2L$	3
86	الى اقل من	76	4	$\vdots$		$\vdots$	$\vdots$
96	الى اقل من	86	5	$X_S + mL$	الى اقل من	$X_S + (m-1)L$	m

الفئات بدون اختصار	التكرار بالإشارات	f	x
46 – 56	xxxx	4	51
56 – 66	xxx	3	61
66 – 76	xxxx	4	71
76 – 86	xxxxx xx	7	81
86 – 96	xx	2	91
المجموع		20	

الفئات اختصارا
46 –
56 –
66 –
76 –
86 – 96

والجدول التكراري اعلاه ، يشير الى ان اربعة طلاب اوزانهم تتراوح ما بين 46 الى 56 ، وثلاثة طلاب اوزانهم تتراوح ما بين 56 الى 66 ، وهكذا.

وفي حالة طلب حساب مراكز الفئات فإنها تحسب كالاتي:

$$X = \frac{L.L + U.L}{2}$$

$$X = \frac{46 + 56}{2} = 51$$

$$X = \frac{56 + 66}{2} = 61$$

وهكذا الى اخر فئة.

ب- التوزيع التكراري النسبي(**Proportionate frequency distribution**): وهو توزيع تكراري تكون تكرارات فئاتهِ معبراً عنها بنسب مئوية فإذا رمزنا للتكرار النسبي بـ f^* فإن التكرار النسبي يحسب كالاتي:

$$f^* = \frac{f_i}{\sum_{i=1}^{m} f_i} * 100 \quad , i = 1, 2, \dots, m$$

حيث ان: f_i : تكرار الفئة i.

$\sum_{i=1}^{m} f_i = n$: مجموع تكرارات الفئات.

مثال (6): كون توزيع تكراري نسبي للبيانات المبوبة في التوزيع التكراري الوارد في المثال رقم (4).

تسلسل الفئة	الفئات	f	X	f^*
1	60 – 74	4	67	$\frac{4}{60} * 100 = 6.67\%$
2	75 – 89	5	82	$\frac{5}{60} * 100 = 8.33\%$
3	90 – 104	10	97	$\frac{10}{60} * 100 = 16.67\%$
4	105 – 119	11	112	$\frac{11}{60} * 100 = 18.33\%$
5	120 – 134	16	127	$\frac{16}{60} * 100 = 26.67\%$
6	135 – 149	7	142	$\frac{7}{60} * 100 = 11.67\%$
7	150 – 164	7	157	$\frac{7}{60} * 100 = 11.67\%$
	المجموع	60		100%

وهذا يعني ان (**6.67%**) من العوائل تتراوح عدد اشجار البرتقال لديهم ما بين 60 الى 74 شجرة وان (**8.33%**) من العوائل تتراوح عدد اشجار البرتقال لديهم ما بين 75 الى 89 شجرة ، وهكذا.

مثال (**H.W**): كون توزيع تكراري نسبي للبيانات المبوبة في التوزيع التكراري الوارد في المثال رقم (**5**).

مثال(H.W): الاتي توزيع تكراري لدرجات (90) طالب وطالبة في امتحان معين. يطلب تكوين التوزيع التكراري النسبي.

الفئات	10 –	20 –	30 –	40 –	50 –	60 –	70 –	80 – 90
التكرارات	2	3	5	10	18	25	15	12

ج- التوزيع التكراري المتجمع (**Cumulative frequency distribution**): وهو التوزيع الذي يبين كمية التكرار المتجمع عند قيمة معينة من قيم المتغير العشوائي وكذلك تتجمع التكرارات فيهِ تصاعدياً او تنازلياً وهذا التوزيع على نوعين هما:

1) **التوزيع التكراري المتجمع الصاعد**: وهو التوزيع الذي يبين تراكم التكرارات ابتداءً من الفئة الاولى وانتهاءً بالفئة الاخيرة ، ويرمز لهُ بـ (F).

ويتم حساب التوزيع التكراري المتجمع الصاعد باتباع مايلي:

1. عمل توزيع تكراري كما ذكر سابقاً
2. يتم إعادة كتابة الفئات بذكر الحدود العليا لها مع ذكر عبارة اقل من او يساوي بجانب الحدود العليا في حالة المتغيرات المتقطعة وعبارة اقل من في حالة المتغيرات المستمرة.
3. يتم تجميع التكرارات تصاعدياً حيث يكون تكرار الفئة الاولى مساوي الى تكرارها، وتكرار الفئة الثانية مساوي الى مجموع تكرارات الفئة الاولى والفئة الثانية ، وهكذا لغاية الوصول الى تكرار الفئة الاخيرة التي يكون تكرارها المتجمع يساوي مجموع تكرارات كل الفئات (**أي حجم العينة** = (**n**)).

مثال (7): كون توزيع تكراري متجمع صاعد ، يعرض توزيع اوزان عينة من (20) طالب وبالاعتماد على البيانات المبوبة الواردة في المثال رقم (5).

الفئات	التكرار بالإشارات	f	x	الحدود العليا	F
46 – 56	xxxx	4	51	اقل من 56	4
56 – 66	xxx	3	61	اقل من 66	7
66 – 76	xxxx	4	71	اقل من 76	11
76 – 86	xxxxx xx	7	81	اقل من 86	18
86 – 96	xx	2	91	اقل من 96	20
المجموع		20			

مثال (**H.W**): كون توزيع تكراري متجمع صاعد للبيانات المبوبة في التوزيع التكراري الوارد في المثال رقم (4)**.**

2) **التوزيع التكراري المتجمع النازل**: وهو التوزيع الذي يبين تناقص التكرارات ابتداءً من الفئة الاولى وانتهاءً بالفئة الاخيرة ، ويرمز لهُ بـ ($\mathbf{F'}$).

ويتم حساب التوزيع التكراري المتجمع النازل باتباع ما يلي:

1. عمل توزيع تكراري كما ذكر سابقاً.
2. يتم إعادة كتابة الفئات بذكر الحدود الدنيا لها مع ذكر عبارة اكبر من او يساوي بجانب الحدود العليا في حالة المتغيرات المتقطعة وفي حالة المتغيرات المستمرة على حد سواء.
3. يكون التكرار المتجمع النازل للفئة الاولى مساوي الى المجموع الكلي للتكرارات اما التكرار المتجمع النازل للفئة الثانية فيكون مساوياً للتكرار المجتمع النازل للفئة الاولى مطروحاً منهُ تكرار الفئة الاولى الاصلية ، اما تكرار الفئة الثالثة المجتمع النازل فيساوي التكرار المتجمع النازل للفئة الثانية مطروحاً منهُ تكرار الفئة الثانية وهكذا يبدأ التكرار المتجمع النازل بالتنازل لغاية الفئة الاخير حيث يكون تكرارها المتجمع النازل مساوي لتكرارها الاعتيادي.

مثال (8): كون توزيع تكراري متجمع نازل للبيانات المبوبة في التوزيع التكراري الوارد في المثال رقم (4).

تسلسل الفئة	الفئات	f	الحدود الدنيا	F′
1	60 – 74	4	اكبر من او يساوي 60	60
2	75 – 89	5	اكبر من او يساوي 75	56
3	90 – 104	10	اكبر من او يساوي 90	51
4	105 – 119	11	اكبر من او يساوي 105	41
5	120 – 134	16	اكبر من او يساوي 120	30
6	135 – 149	7	اكبر من او يساوي 135	14
7	150 – 164	7	اكبر من او يساوي 150	7
	المجموع	60		

ثانياً: العرض الهندسي للبيانات: (Engineering view of data)

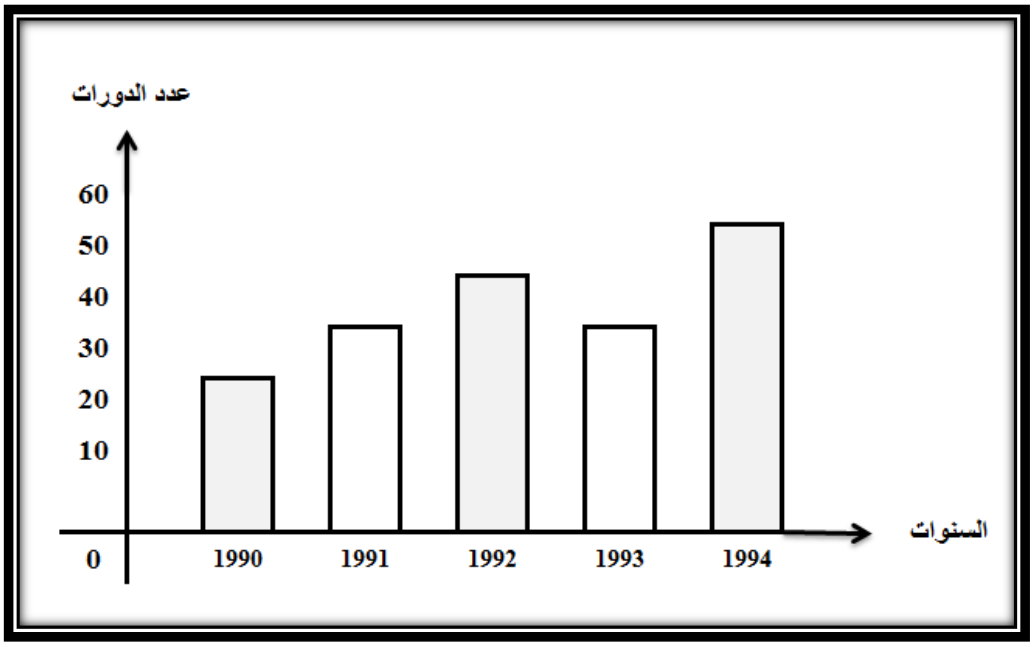
بغية اعطاء فكرة واضحة وسريعة عن البيانات المبوبة ، فأنهُ يمكن عرض هذهِ البيانات بهيئة رسوم بيانية واشكال هندسية متعددة الاشكال والتصاميم ، ومن اهم وسائل العرض الهندسي للبيانات هي:

1) الاشرطة البيانية (**Bar – charts**) : وهي عبارة عن مستطيلات رأسية او افقية قواعدها متساوية وتمثل الصفة التي تم على اساسها التبويب (سنة ، شهر ، محافظة ، قطر ، صنف دم ، ... الخ) وارتفاعاتها تمثل البيانات المقابلة لتلك الصفة (كميات النفط المستخرجة خلال سلسلة زمنية سابقة ، درجات الحرارة حسب اشهر السنة ، كميات المحصول من الحنطة حسب المحافظات ، عدد المرضى حسب صنف الدم، ... الخ).

مثال (9): بلغ عدد الدورات التدريبية التي نفذت من قبل كليات احدى الجامعات العراقية للفترة من (**1990 – 1994**).

السنوات	1990	1991	1992	1993	1994
عدد الدورات	25	35	45	35	55

المطلوب: تمثيل هذهِ البيانات بأشرطة بيانية.



الشكل رقم (1): عدد الدورات التدريبية المنفذة من قبل كليات احدى الجامعات العراقية للفترة من (1990 – 1994).

2) المستطيل البياني (**Rectangular – chart**) : عبارة عن شكل هندسي يستخدم في تمثيل بيانات ظاهرة معينة يمكن تجزئتها الى عدد من الاصناف القابلة للتجميع مثل تكاليف انتاج سلعة معينة، عدد الطلبة موزعين حسب المراحل الدراسية ، وفكرة هذا الشكل بسيطة جداً وتتلخص باختيار مستطيل ذو قاعدة مناسبة، هذا المستطيل يمثل مجموع البيانات الكلية عن تلك الظاهرة (**مجموع التكاليف، عدد الطلبة ، ... الخ**) ، وبعد ذلك يتم تمثيل كل صنف من البيانات بمستطيل جزئي داخل المستطيل الكبير بحيث ان مجموع مساحة المستطيل الكبير ، وتتم عملية اختيار قواعد المستطيلات الجزئية وفقاً لما يلي:

$$\text{طول قاعدة المستطيل الجزئي} = \frac{\text{عدد بيانات الصنف}}{\text{مجموع البيانات الكلية}} * \text{طول قاعدة المستطيل الكبير}$$

مثال (10): بلغ عدد الطلبة في احدى الكليات (**2000**) طالب وطالبة ، منهم (**800**) في المرحلة الاولى ، و (**500**) في المرحلة الثانية ، و (**400**) في المرحلة الثالثة ، و (**300**) في المرحلة الرابعة. المطلوب تمثيل هذهِ البيانات بمستطيل بياني.

الحل: نختار مستطيل ذو قاعدة مساوية الى 10 سم ، وعندئذ فإن طول قاعدة كل مستطيل جزئي الخاص بتلك المرحلة هو:

المرحلة الاولى $= \frac{800}{2000} * 10 = 4$ سم

المرحلة الثانية $= \frac{500}{2000} * 10 = 2.5$ سم

المرحلة الثالثة $= \frac{400}{2000} * 10 = 2$ سم

المرحلة الرابعة $= \frac{300}{2000} * 10 = 1.5$ سم

وبعد ذلك نبدأ بعملية رسم الشكل

الشكل رقم (2): يمثل عدد الطلبة بشكل مستطيل بياني

3) الدائرة البيانية (**Pie – chart**) : عبارة عن شكل هندسي مثل المستطيل البياني حيث انهُ وبدلاً من تمثيل الاصناف بمستطيلات جزئية فانهُ يتم تمثيلها بقطاعات داخل دائرة بحيث ان مجموع مساحات القطاعات تمثل مساحة الدائرة وبهدف تحديد كل قطاع فأنهُ يتوجب تحديد زاوية كل منهما ، ان ذلك يتم وفقاً لما يلي:

زاوية القطاع $= \frac{\text{عدد بيانات الصنف}}{\text{مجموع البيانات الكلية}} * 360^{\circ}$

مثال (11): بلغ عدد الطلبة في احدى الكليات (**2000**) طالب وطالبة ، منهم (**800**) في المرحلة الاولى ، و (**500**) في المرحلة الثانية ، و (**400**) في المرحلة الثالثة ، و (**300**) في المرحلة الرابعة. المطلوب تمثيل هذهِ البيانات بدائرة بيانية.

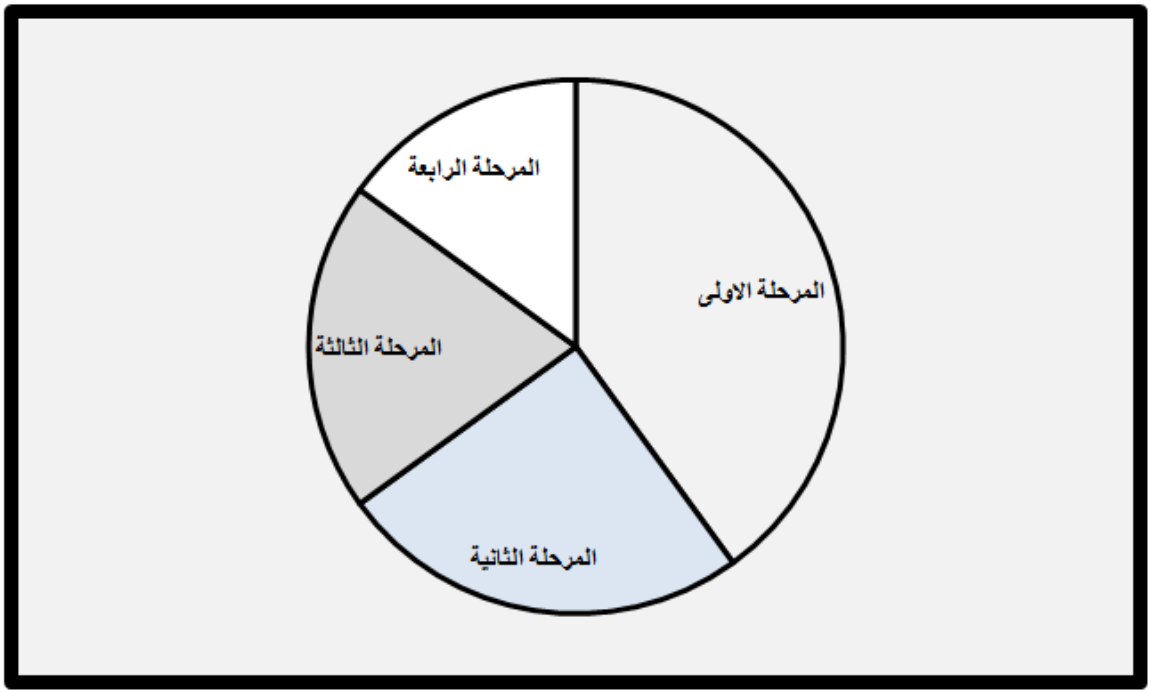
الحل:

$$\textbf{زاوية القطاع التي تمثل طلبة المرحلة الاولى} = \frac{800}{2000} * 360^{\circ} = 144^{\circ}$$

$$\textbf{زاوية القطاع التي تمثل طلبة المرحلة الثانية} = \frac{500}{2000} * 360^{\circ} = 90^{\circ}$$

$$\textbf{زاوية القطاع التي تمثل طلبة المرحلة الاولى} = \frac{400}{2000} * 360^{\circ} = 72^{\circ}$$

$$\textbf{زاوية القطاع التي تمثل طلبة المرحلة الاولى} = \frac{300}{2000} * 360^{\circ} = 54^{\circ}$$



الشكل رقم (3): يمثل عدد الطلبة بشكل دائرة بيانية

4) الخط البياني (**Line – chart**) : عبارة عن شكل بياني يوضح التغيرات الحاصلة في ظاهرة معينة عبر الزمن ، وهو شكل نافع في حالة اجراء مقارنة بين ظاهرتين او اكثر مقاسة بنفس وحدات القياس، مثل مقارنة التغيرات الحاصلة بين تكاليف انتاج سلعة معينة والارباح المتحققة من مبيعات تلك السلعة خلال فترة زمنية معينة.