الفصل الرابع: معالجة وتحليل وتفسير البيانات.

البيانات الإحصائية

هي مجموعة المشاهدات أو الملاحظات المأخوذة أثناء دراسة معينة وقد تكون بيانات رقمية (كمية) مثل أطوال وأوزان مجموعة من الطلاب ودخول مجموعة من الأسر أو بيانات غير رقمية (وصفية) مثل لون البشرة والجنس... إلخ

أ-المتغير (Variable)

هو مقدار له خصائص رقمية (كمية) وغير رقمية (وصفية) تتغير قيمته من عنصر إلى آخر من عناصر المجتمع أو العينة. يرمز له بحرف (Y أو x).

ب-أنواع البيانات

1-بيانات وصفية أو نوعية: وهي تلك الظواهر أو الصفات التي لا يمكن قياسها مباشرة بالأرقام العددية مثل صفة لون الازهار (ازرق، أصفر، بني) والحالة القيمة الغذائية لنوع نباتي (غني، متوسط، فقير) والجنس (ذكر، أنثى) .....الخ.

2-بيانات كمية: وهي تلك الظواهر أو الصفات التي يمكن قياسها مباشرة بالأرقام العددية مثل صفة الطول والوزن والعمر وكمية المحصول.....الخ.

وتنقسم المتغيرات الكمية إلى:

أ-متغيرات مستمرة أو متصلة:

وهو المتغير الذي تأخذ المشاهدة أو المفردة فيه أي قيمة رقمية في مدى معين وأمثلتها الوزن والطول ودرجة الحرارة.... الخ لأنه يمكن قياسها بأجزاء صغيرة جدا وتأخذ أية قيمة تقع في حدود معينة (أي انها تأخذ أرقام مستمرة مثل 1, 1.2, 2.3). وبصورة عامة كل البيانات التي تقاس تعتبر بيانات لمتغير مستمر.

ب-متغيرات غير مستمرة (منفصلة):

وهو المتغير الذي تأخذ المشاهدة أو المفردة فيه قيما متباعدة أو متقطعة غير مستمرة مثل أعداد أفراد الأسرة أو عدد المتزوجين أو العزاب في مجتمع ما (أي في الغالب تكون أعداد صحيحة).

ج-مصادر البيانات:

1-المصدر المباشر: النزول للميدان وجمع المعلومات مباشرة

2-المصدر غير المباشر يندرج تحت هذا المصدر ما يلي

- السجلات أو الوثائق التاريخية
- الاستبيان
- المقابلات الشخصية
- الاختبارات الخاصة

يجب الآن معالجة جميع المعلومات التي تم جمعها في الميدان. أولاً، سيتم التحقق من صحة الملاحظات، ثم سيتم إدخال البيانات لاستخدامها في الحسابات المختلفة والتمثيلات الرسومية اللازمة للتحليلات.

2. أدخال البيانات

وتصبح المعالجة الحاسوبية ضرورية بمجرد زيادة حجم البيانات، أو بمجرد سعينا لاستخدام تقنيات التحليل الإحصائي، التي تتطلب الكثير من العمليات الحسابية مثل التحليلات متعددة الأبعاد (Scherrer, 1984). يجب أن يتم إدخال البيانات بطريقة منطقية للسماح بالمعالجة الإحصائية إذا لزم الأمر. قبل كل شيء، يجب أن يكون لدى الباحث فهم واضح جدًا لكيفية تحليل وتلخيص كل نوع من البيانات. ومن هنا تأتي أهمية التفكير قبل إنشاء الملف أو برنامج إدخال البيانات.

3. تنظيم البيانات

تتطلب أي معالجة حاسوبية أو يدوية للبيانات أولاً معرفة نوع البيانات المستخدمة. أول شيء يجب فعله هو مراجعة البيانات للحكم على أهمية العمل الذي يتعين القيام به، واكتشاف أنواع المتغيرات التي تلعب دورًا، وما هي المعالجة التي يجب تنفيذها. إن وزن الفرد ولون بتلات الزهرة ليسا بيانات من نفس الطبيعة. ومن المهم التعرف على أنواع المتغيرات لأن "الكل متغير حساباته" سواء كانت كمية أو نوعية.

4. استغلال البيانات باستخدام المؤشرات البيئية

ستتم معالجة النتائج من خلال المؤشرات البيئية للتكوين والبنية. هذه الجوانب تجعل من الممكن معرفة نوعية العينات ووفرة وتشتت الأنواع المدروسة. كما تمت مناقشة المقارنة بين تكوين السكان في المحطات التي تمت دراستها.

1.4مؤشرات البيئية التركيبية

تهتم هذه المؤشرات في دراسة الوفرة النسبية والنوعية لكل من الأنواع ومقارنتها، تقارن بين المجتمعات في نفس المكان والزمان، كذلك تصف وتقدم البنية العامة للمجتمع الحيواني او النباتي في محطات الدراسة.

1.الغنى النوعي (S)

وهو يمثل العدد الكلي الإجمالي للكائنات الحية المتواجدة في مجتمع معين أو في بيئة معينة

أو في منطقة أو محطة معينة. ولهذا يعتبر الغنى النوعي من أبسط معايير قياس التنوع البيولوجي داخل بيئة أو منطقة معينة

و حسب دراستنا يمثل الغنى النوعي(S) العدد الكلي (إن كانت عائلات أو أجناس أو أنواع) بغض النظر عن نسبتها أو طريقة انتشارها داخل هذه المحطات

2.الوفرة المطلقة AR%

هو نسبة عدد الأفراد من الأنواع الموجودة على العدد الكلي لجميع الأنواع في بيئة معينة.

$$AR\% = Pi \times 100$$

حيث

$$P_i = \frac{ni}{N}$$

Pi: الوفرة النسبية.

ni : عدد أفراد النوع.

N : العدد الكلي للأنواع.

2.4 المؤشرات البيئية البنيوية

1. مؤشر: *Schannon-Weaver*

يعتبر معامل Weaver-Schannon من أهم المؤشرات التنوع استخداما لأنها تأخذ بعين الاعتبار الثراء النوعي (S) والوفرة النسبية للأنواع في نفس الوقت هذا المؤشر يحدد من خلال إحصاء عدد أفراد كل نوع ضمن العينة.

وحسب معادلة مؤشر التنوع **(H')** التي وضعت من قبل **Weaver-Schannon** سنة 1949 م.

$$H' = -\sum_{i=1}^{S} p_i (Log_2 P_i)$$

حيث:

$$Log_2 P_i = \frac{Ln\, Pi}{Ln2}$$

يقاس مؤشر التنوع بوحدة بت/ الفرد

2.مؤشر Simpson 1-D

هذا المؤشر يسمى بمؤشر **Simpson** للتنوع الإيكولوجي في هذه الحالة يمكن لإثنين من الأحياء المنتخبين عشوائيا ينتميان إلى فئتين مختلفتين، ويمكن حسابه على شكل الصيغة التالية:

$$1 - D = 1 - \sum_{i=1}^{S} p_i^2$$

بحيث تتراوح قيمته ما بين (0 و1) حيث كلما كانت قيمة مؤشر **Simpson** للتنوع الإيكولوجي أكبر كلما كان التنوع أكثر، فالقيمة (1) تعني لا نهاية للتنوع.

3.مؤشر الثبات (Ci%)

هو النسبة المئوية لعدد الخرجات التي تحتوي على الأنواع المدروسة **(ri)** على العدد الكلي للخرجات R التي أجريت. يتم حسابه لعينة أو لجميع العينات المأخوذة من المجتمع الحيوي .(Dajoz ,1985) حسب المعادلة التالية

$$Ci\% = \frac{ri}{R} \times 100$$

ri : عدد الخرجات التي تحتوي النوع المدروس.

R : عدد العينات الكلي.

حسب Bigot and boot (1972) يمكن التمييز بين مجموعات الأنواع وفقا لتكرار حدوثها.

Espèce omniprésent	Ci=100%
Espèce constante	%75 < Ci< 100%
Espèce régulière	%50 <Ci <75%
Espèce accessoire	%25 <Ci <50%

Espèce accidentelle %5<Ci <25%

Espèce Rare Ci < 5 %

4. مؤشر التوازن E:

مؤشر التوازن أو الوفرة النسبية للأنواع المختلفة التي تعكس الثراء أو الغنى النوعي ضمن مجتمع ما، يتم قياس مؤشر التوازن بالاعتماد على الغنى ال نوعيS ومؤشر (H') Schannon-Weaver، حيث تتراوح قيمة مؤشر التوازن E بين (0 و1).

ويمكن حساب أعلى نسبة تنوع لأكثر ا الانواع انتشارا H' max وهذا بافتراض أن الأنواع متساوية في الكثافة حيث تعطى الصيغة التالية:

$$E = \frac{H'}{Hmax} = \frac{H'}{Log_2 S}$$

حيث:

$$.Hmax = Log_2 S = \frac{Ln\, S}{Ln\, 2}$$

H' : مؤشر. **Schannon-Weaver**

S: النوعي الثراء.

5. مؤشر Jaccard للتشابه

يسمح مؤشر **Jaccard** بتحديد درجة التشابه أو التداخل بين الأنواع الموجودة في مجتمعين أو منطقتين أو محطتين حيث:

$$Jc = \frac{c}{a + b - c}$$

Jc : مؤشر**Jaccard**

c : عدد العائلات أو الأجناس أو الأنواع المتواجدة أو المشتركة في كلا المحطتين.

a : عدد العائلات أو الأجناس أو الأنواع المتواجدة في المحطة الأولى.

b :عدد العائلات أو الأجناس أو الأنواع المتواجدة في المحطة الثانية.

5.الطرق الإحصائية :

5-1-التصنيف التدريجي التصاعدي: CHA

اقترحت هذه الطريقة من طرف **(**Benzekri , 1973**)**، بحيث نحاول أن نرتب مجموعات من الأنواع بتتابع، للحصول في النهاية على شجرة بعقدتين التي تشكلان مجموعة انطلاق . وقد استعملت هذه الطريقة من طرف عدة باحثين.
تسمح هذه الطريقة بتفسير الجدول مكون من عمودين: الأنواع والمحطات، الذي يعتمد على مؤشر المسافة ما يعرف بمؤشر التشابه يتم بجمع الأنواع في المحطات المدروسة المتقاربة ثم يضم الأكثر تقاربا بتتابع المتكرر.

2.5 التحليل بالمكونات الاساسية (ACP)

تم استخدام تحليل المكونات الاساسية A.C.P بهدف الكشف عن العوامل المشتركة المتعددة المؤثرة في عدد من الظواهر المختلفة , اي تكثيف اعداد كبيرة من المتغيرات تبعا لعدد علاقاتها الارتباطية في عدد من المحاور او العوامل.

3.5 ما هو الرسم البياني في الإحصاء؟

الرسم البياني في الإحصاء هو تمثيل مرئي للبيانات يهدف إلى توضيح الاتجاهات، العلاقات، والفروقات بين القيم المختلفة. تساعد هذه الرسوم في تبسيط الأرقام والمعلومات المعقدة لتسهيل تحليلها واتخاذ القرارات بناءً عليها

أنواع الرسم البياني في الإحصاء

تُعد الرسوم البيانية في الإحصاء أدوات حيوية لتحليل البيانات وتقديمها بطرق تسهل فهم المعلومات المعقدة. تتنوع هذه الرسوم لتناسب مختلف أنواع البيانات والأهداف التحليلية. فيما يلي أبرز أنواع الرسوم البيانية المستخدمة في الإحصاء:

1. الرسم البياني الشريطي (Bar Graph): يُستخدم لتمثيل البيانات الفئوية، حيث تمثل كل شريط فئة معينة، ويعكس طول الشريط قيمة تلك الفئة. يُعد هذا النوع مثاليًا لمقارنة كميات بين فئات مختلفة.

2. الرسم البياني الخطي (Line Graph): يُستخدم لعرض البيانات المتغيرة على مدار الزمن، حيث تُرسم النقاط المتصلة بخطوط لبيان الاتجاهات والتغيرات في البيانات.

3. المدرج التكراري (Histogram):يستخدم لتمثيل توزيع البيانات الكمية، حيث تُقسم البيانات إلى فئات متساوية، ويُظهر ارتفاع كل عمود عدد القيم في كل فئة. يساعد هذا النوع في فهم توزيع البيانات والكشف عن الأنماط.

4. الرسم البياني الدائري (Pie Chart): يُستخدم لعرض النسب المئوية للفئات المختلفة من مجموعة بيانات معينة، حيث يمثل كل قطاع نسبة فئة معينة من الكل. يُفضل استخدامه عندما يكون عدد الفئات محدودًا.

5. الرسم البياني الصندوقي (Box Plot): يستخدم لتلخيص مجموعة من البيانات من خلال خمسة إحصاءات: الحد الأدنى، الربع الأول، الوسيط، الربع الثالث، والحد الأقصى. يساعد هذا النوع في الكشف عن القيم المتطرفة وتحديد مدى تشتت البيانات.

6. الرسم البياني المبعثر (Scatter Plot): يُستخدم لدراسة العلاقة بين متغيرين كميين، حيث تُرسم النقاط على المحاور لتحديد ما إذا كان هناك ارتباط أو نمط معين بين المتغيرين.

7. الرسم البياني ذو المضلع التكراري (Frequency Polygon): يُشبه المدرج التكراري، لكنه يستخدم النقاط المتصلة بخطوط لتمثيل تكرار الفئات. يُفيد في مقارنة توزيعات متعددة على نفس الرسم.

8. الرسم البياني ذو السلاسل الزمنية (Time Series Graph): يُستخدم لعرض البيانات التي تم جمعها على فترات زمنية متتالية، مما يساعد في تحليل الاتجاهات والتنبؤ بالتغيرات المستقبلية.

9. الرسم البياني الحراري (Heat Map): يُستخدم لتمثيل البيانات من خلال الألوان، حيث تشير الألوان المختلفة إلى قيم مختلفة. يُفيد في عرض الكثافات والأنماط في مجموعات البيانات الكبيرة.

10. الرسم البياني الشجري (Tree Map): يُستخدم لعرض البيانات الهرمية من خلال مستطيلات متداخلة، حيث يمثل كل مستطيل فئة معينة، ويعكس حجمه حجم تلك الفئة.

اختيار النوع المناسب من الرسم البياني يعتمد على طبيعة البيانات والهدف من التحليل. يجب على المحلل اختيار الرسم البياني الذي يقدم المعلومات بأكثر الطرق وضوحًا ودقة.

أهمية الرسم البياني في الإحصاء

الرسم البياني في الإحصاء ليس مجرد وسيلة عرض جذابة، بل هو أداة تحليلية قوية تُسهل فهم البيانات واتخاذ القرارات بناءً على أنماط واتجاهات واضحة. فهو يساعد على:

1. تبسيط البيانات المعقدة: يعرض كميات ضخمة من البيانات بصيغة بصرية سهلة الفهم.
2. تحديد الأنماط والاتجاهات: يوضح العلاقات بين المتغيرات ويساعد في التنبؤ بالمستقبل.
3. تحسين دقة التحليل: يتيح للمحللين فحص البيانات بشكل سريع، مما يقلل من أخطاء التفسير.

كيفية اختيار نوع الرسم البياني المناسب؟

يعتمد اختيار الرسم البياني المناسب على طبيعة البيانات والهدف من التحليل. إليك بعض الإرشادات:

- للمقارنة بين الفئات ← استخدم الرسم البياني الشريطي أو الرسم البياني الدائري.
- لمتابعة التغيرات عبر الزمن ← استخدم الرسم البياني الخطي أو الرسم البياني ذو السلاسل الزمنية.
- لفهم توزيع البيانات ← استخدم المدرج التكراري أو الرسم البياني الصندوقي.
- لدراسة العلاقة بين متغيرين ← استخدم الرسم البياني المبعثر.

أخطاء شائعة عند استخدام الرسوم البيانية في الإحصاء

على الرغم من فوائدها، قد تؤدي بعض الأخطاء إلى تفسيرات خاطئة، مثل:

- استخدام مقياس غير مناسب: قد يؤدي تغيير مقياس المحور إلى تضليل القارئ.
- الإفراط في استخدام البيانات: قد تجعل البيانات الكثيفة الرسم البياني معقدًا وغير مفهوم.
- اختيار نوع رسم غير مناسب: مثل استخدام رسم بياني دائري لمقارنة بيانات غير نسبية.
- إغفال توضيح المصدر والوحدات: قد يؤدي إلى التباس في فهم البيانات المعروضة

البرامج والأدوات المستخدمة لإنشاء الرسوم البيانية

يمكن إنشاء الرسوم البيانية باستخدام العديد من الأدوات، منها:

1. Excel: مناسب للمبتدئين ويوفر مجموعة متنوعة من الرسوم البيانية.
2. Tableau: أداة قوية لتحليل البيانات وإنشاء رسوم تفاعلية.
3. Google Charts: أداة مجانية لإنشاء رسوم بيانية تفاعلية عبر الإنترنت.
4. Python (Matplotlib, Seaborn): تُستخدم في تحليل البيانات الكبيرة وإنشاء رسوم بيانية متقدمة.
5. Power BI: مثالي لتحليل البيانات الضخمة وإنشاء تقارير تفاعلية.

6. تفسير النتائج

ويجب ألا ننسى أن أي دراسة يتم إجراؤها بهدف تقديم معلومات مفيدة لفهم الديناميكيات الطبيعية أو تلك المرتبطة بإدارة الموقع. إذا كان من المهم معرفة كيفية الحساب، فمن المهم بنفس القدر معرفة كيفية تفسير النتائج التي تم الحصول عليها (أو على الأقل اقتراح تفسير). يجب أن يكون لديك المهارات اللازمة لاستغلال النتائج.

تفسير النتائج ليس دائما فوريا. ويجب أن يتم ذلك في حدود الطريقة، بهدف تحقيق الهدف الذي يسعى إليه الباحث، ويجب أن يقدم إجابات على الفرضيات التي تمت صياغتها في البداية. إذا تم التخطيط للمنهجية بعناية، فمن الممكن تفسير واحد فقط وتكون مصادر الخطأ صفرًا أو صغيرة (Scherrer، 1984).