

المحور السابع: تحليل البيانات إحصائياً

أولاً: تعريف التحليل الإحصائي للبيانات:

التحليل الإحصائي للبيانات هو عملية استخدام الأساليب الإحصائية لفهم البيانات واستخلاص استنتاجات ذات مغزى منها. يتضمن التحليل الإحصائي عدة خطوات تشمل التخطيط، التصميم، جمع البيانات، التحليل، وتفسير النتائج بشكل دقيق.

ثانياً: الأهداف الأساسية للتحليل الإحصائي للبيانات:

- التعرف المبدئي على البيانات: أي بناء فهم شامل ومبدئي لطبيعة البيانات من خلال استكشافها وتحليل خصائصها الأساسية لتوجيه مسار التحليل الإحصائي.
- اختبار جودة البيانات: التركيز على تحديد المشكلات مثل القيم المفقودة، القيم الشاذة، أو التكرار غير المبرر في البيانات.
- اختبار صحة فروض البحث: تقييم مدى تماشي البيانات مع الافتراضات الأساسية للاختبارات الإحصائية، مثل التوزيع الطبيعي.
- الكشف عن الأنماط والعلاقات: يهدف التحليل إلى الكشف عن الأنماط والاختلافات والعلاقات التي قد لا تكون واضحة من خلال الفحص المباشر للبيانات.
- تحديد العلاقات السببية: يساعد التحليل في قياس انتشار القيم وتحديد العلاقات السببية وتقييم العوامل المؤثرة على الأداء، مما يساهم في تطوير استراتيجيات فعالة.
- التنبؤ واتخاذ القرارات: استخدام الأدوات الإحصائية لوصف البيانات واستنتاج القيم والتنبؤ بها، مما يساهم في اتخاذ قرارات دقيقة ومبنية على البيانات.

ثالثاً: خطوات التحليل الإحصائي للبيانات

1. مرحلة تنظيم البيانات وعرضها:

وتهدف إلى إعطاء صورة سريعة عن الظاهرة المدروسة ومدلولاتها بشكل مبسط ومختصر، بحيث يسهل على الباحث بعد ذلك دراستها وتحليلها دون مزيد من العناء والتفكير. وتستخدم الجداول الترتيبية والتكرارية والرسومات البيانية بما فيها المردجات التكرارية والمنحنيات البيانية في هذه المرحلة.

2. مرحلة وصف البيانات:

تستخدم في هذه المرحلة مقاييس التمرکز والتشتت وبعض مقاييس الارتباطات لتوضيح العلاقة بين المتغيرات.

3. مرحلة تحليل البيانات:

تهدف هذه المرحلة إلى اختبار الفروض اختباراً علمياً مبرهنًا بأدلة إحصائية، وبناءً على ذلك يتم قبول أو رفض الفرض.

رابعاً: الأساليب الإحصائية الشائعة الاستخدام في تحليل البيانات:

1. الإحصاء الوصفي (Descriptive Statistics):

أ. مقاييس النزعة المركزية (Measures of Central Tendency):

- المتوسط الحسابي (Mean): هو مجموع القيم مقسوماً على عددها، ويعطي فكرة عامة عن القيمة الوسطية للبيانات.

مثال: إذا كانت لدينا البيانات التالية: 5، 10، 15، 20، و25. فإننا نحسب مجموع القيم كالتالي $5+10+15+20+25=75$ ، ثم نقسم المجموع على عدد القيم وهو 5. إذن المتوسط الحسابي هو $75/5=15$ وبالتالي، فإن المتوسط الحسابي للبيانات هو 15، وهو يعطي فكرة عن القيمة الوسطية للبيانات إذا تم توزيعها بالتساوي.

- الوسيط (Median): القيمة التي تقسم مجموعة البيانات إلى نصفين بعد ترتيبها تصاعدياً أو تنازلياً.

مثال: إذا كانت البيانات 5، 10، 15، 20، و25 مرتبة تصاعدياً، نجد أن عدد القيم فردي، لذا فإن الوسيط هو القيمة التي تقع في المنتصف مباشرة، وهي 15. أما إذا كانت البيانات 5، 10، 15، و20،

فعدد القيم زوجي، وبالتالي نحسب متوسط القيمتين الوسطيتين. إذن، الوسيط في الحالة الأولى هو 15 وفي الحالة الثانية هو 12.5، مما يعكس المركز الترتيبي للبيانات.

- **المنوال (Mode):** القيمة الأكثر تكراراً في البيانات.

مثال: إذا كانت البيانات 5، 10، 10، 15، 20، و25، فإن القيمة الأكثر تكراراً هي 10، وبالتالي فإن المنوال هو 10. وإذا كانت البيانات 5، 5، 10، 10، 15، و20، فإن القيمتين 5 و10 تتكرران بنفس العدد، لذا فإن المنوال هنا هو 5 و10 (منوالان). المنوال يعكس القيم الأكثر شيوعاً في البيانات.

ب. **مقاييس التشتت: (Measures of Dispersion):**

- **المدى (Range):** يعبر عن الفرق بين أكبر وأصغر قيمة.

مثال: إذا كانت البيانات هي 5، 10، 15، 20، و25، فإن أكبر قيمة هي 25 وأصغر قيمة هي 5. بحساب الفرق بينهما: $25 - 5 = 20$ المدى هو 20، وهو يعبر عن مدى انتشار البيانات بين أكبر وأصغر قيمة فيها.

- **الانحراف المعياري (Standard Deviation):** يقيس مقدار تشتت البيانات حول المتوسط.

$$\sqrt{(\sum (x_i - \bar{x})^2 / n)} = \text{الانحراف المعياري}$$

√: الجذر التربيعي، $\sum$: رمز المجموع (Sum)، x_i : القيم الفردية، $\bar{x}$: المتوسط الحسابي، n : عدد القيم.

مثال: إذا كانت البيانات هي 5، 10، 15، فإننا نحسب المتوسط الحسابي أولاً

$$\bar{x} = 35 + 10 + 15 = 10$$

ثم نحسب انحراف كل قيمة عن المتوسط:

$$(5-10)^2=25, (10-10)^2=0, (15-10)^2=25$$

مجموع الانحرافات المربعة:

$$25+0+25=50$$

بقسمة هذا المجموع على عدد القيم:

$$16.67 = \frac{50}{3}$$

وأخذ الجذر التربيعي:

$$\sqrt{16.67} \approx 4.08$$

- التباين (**Variance**): مربع الانحراف المعياري ويستخدم لفهم درجة التغير في البيانات.

مثال: باستخدام نفس البيانات السابقة

التباين هو 16.67، وهو مربع الانحراف المعياري 4.08^2

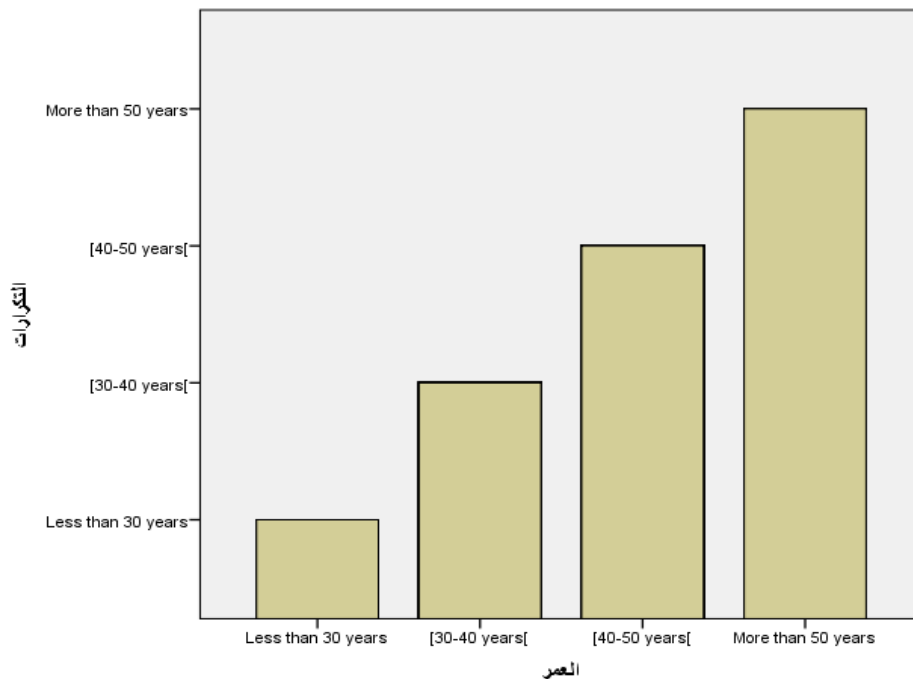
ت. التمثيلات البيانية (**Graphical Representations**):

- الرسوم البيانية الشريطية (**Bar Charts**): تستخدم الرسوم البيانية الشريطية لتمثيل القيم

الفئوية (Categorical Data) عن طريق مستطيلات متساوية العرض تختلف في الارتفاع بناء

على القيم التي تمثلها.

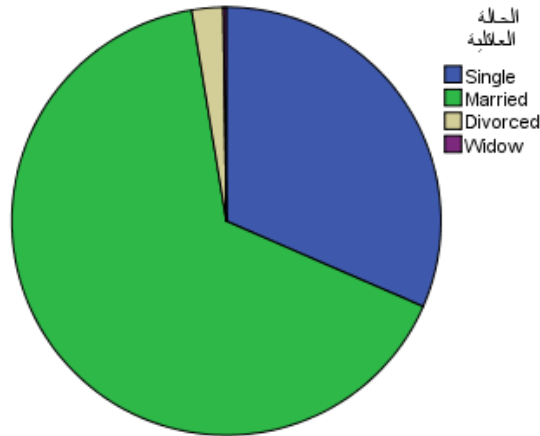
الشكل (1): نموذج للرسم البياني الشريطي



المصدر: من إعداد الأستاذ

- الرسوم البيانية الدائرية (Pie Charts): لتوضيح النسب المئوية لفئات مختلفة ضمن مجموعة كلية.

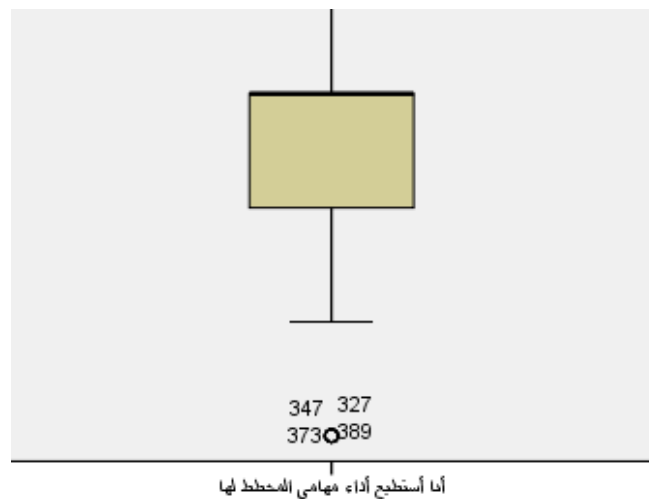
الشكل (2): نموذج للرسم البياني الدائري



المصدر: من إعداد الأستاذ

- المخططات الصندوقية (Box Plots): تُستخدم المخططات الصندوقية لفحص توزيع البيانات والكشف عن القيم المتطرفة (Outliers).

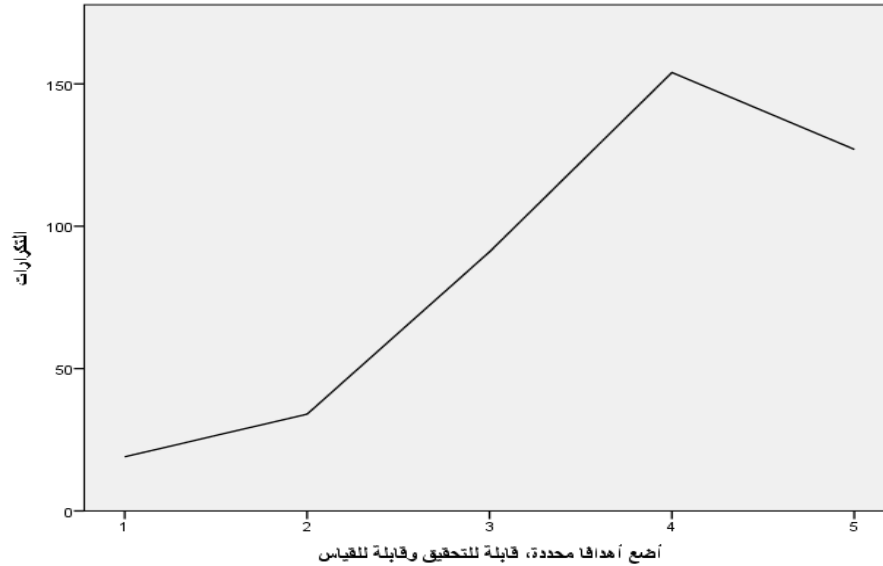
الشكل (3): نموذج للرسم البياني لمخطط صندوقي



المصدر: من إعداد الأستاذ

- المخططات الخطية (Line Charts): تستخدم المخططات الخطية لتوضيح التغير في القيم بمرور الوقت أو عبر نطاق مستمر.

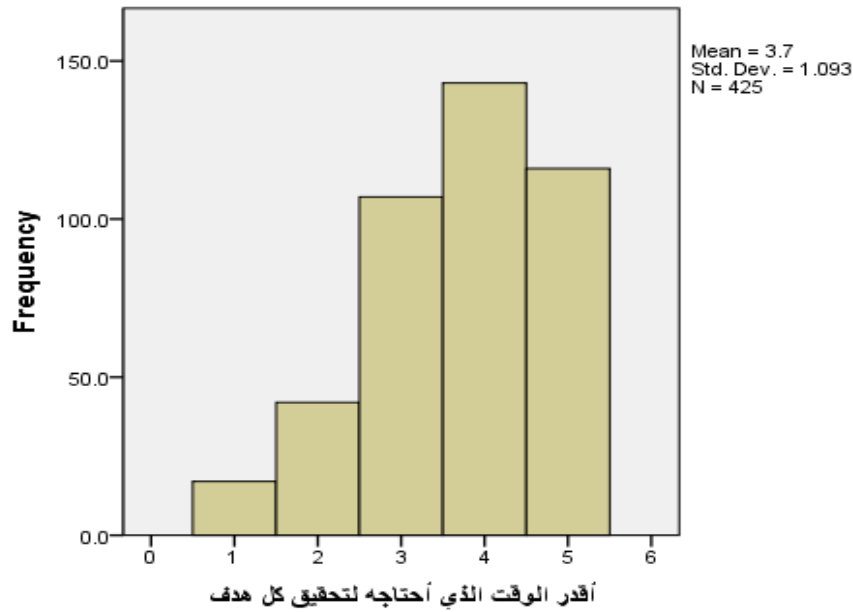
الشكل (4): نموذج للرسم البياني للمخطط الخطي



المصدر: من إعداد الأستاذ

- المدرج التكراري/الهستوغرام (Histograms): تستعمل في تمثيل التوزيع التكراري للبيانات العددية (Continuous Data) باستخدام أعمدة متجاورة. ويستفاد منها في تحليل التوزيع الطبيعي للبيانات، وكشف التكدسات أو الفجوات في البيانات .

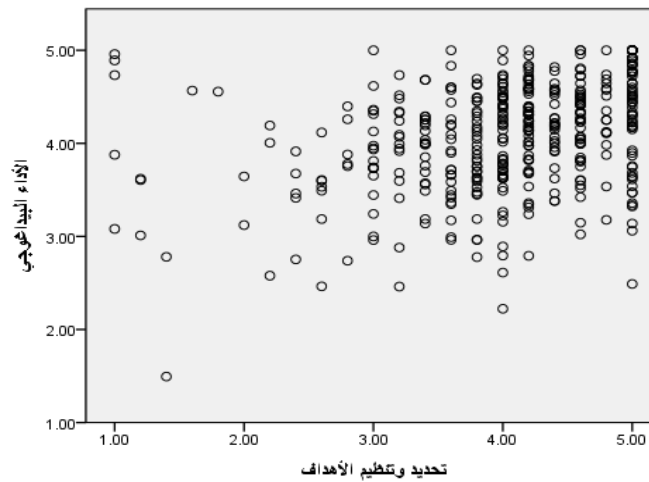
الشكل (5): نموذج للرسم البياني للمدرج التكراري



المصدر: من إعداد الأستاذ

- المخططات النقطية (Scatter Plots): تستعمل بشكل أساسي في فحص العلاقة بين متغيرين.

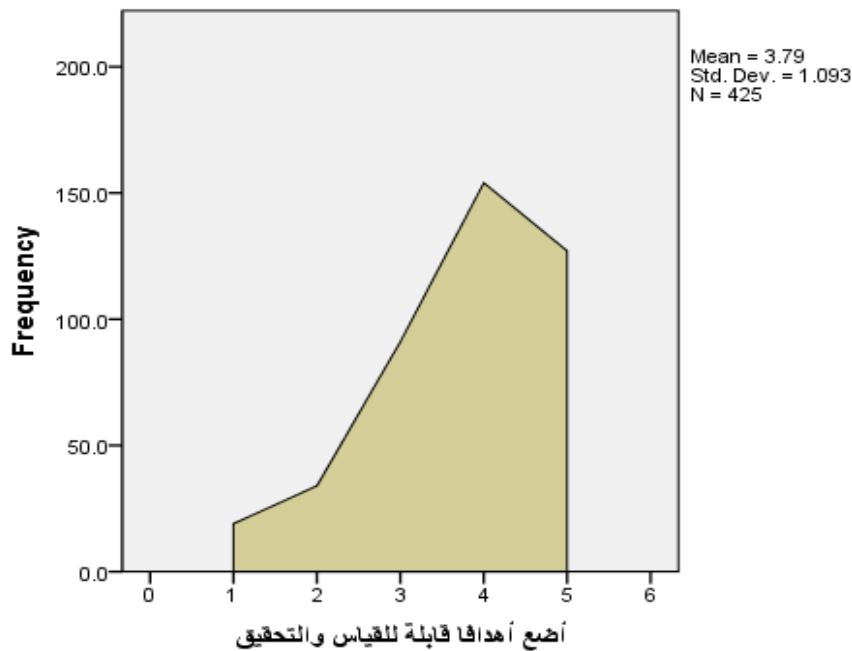
الشكل (6): نموذج للرسم البياني للمخطط النقطي



المصدر: من إعداد الأستاذ

- الرسم البياني المساحي (Area Chart): تعرض اتجاه القيم المتغيرة بمرور الوقت أو مقارنة مكونات القيم الإجمالية. وتمثل القيم المتراكمة مثل النسبة المئوية أو الكميات عبر الزمن.

الشكل (7): نموذج للرسم البياني المساحي



المصدر: من إعداد الأستاذ

2. الإحصاء الاستدلالي (Inferential Statistics):

- أ. اختبارات توزيع البيانات (Distribution Tests): اختبار مدى تطابق البيانات مع توزيع معين (التوزيع الطبيعي أو توزيع غاما مثلا)، وهي مهمة لتحديد الاختبار الإحصائي المناسب. من أهمها:
 - اختبار شابيرو-ويلك (Shapiro-Wilk Test): يستخدم للكشف عن مدى انحراف البيانات عن التوزيع الطبيعي. مثال: دراسة هل علامات الطلبة تتبع التوزيع الطبيعي.
 - اختبار أندرسون-دارلينج (Anderson-Darling Test): اختبار تطابق البيانات مع توزيع معين.
 - اختبار كولموغوروف-سميرنوف (Kolmogorov-Smirnov Test): يقارن توزيع البيانات بالتوزيع الطبيعي أو توزيعات أخرى.
 - اختبار جارك-بيرا (Jarque-Bera Test): لاختبار التوزيع الطبيعي بناء على التفرطح والالتواء.

ب. اختبارات العلاقات (Correlation Tests): تستخدم لقياس قوة واتجاه العلاقة بين متغيرين أو أكثر، ويمكن أن تكون هذه العلاقات خطية أو غير خطية. يعتمد اختيار نوع الاختبار المناسب على طبيعة البيانات ومستوى قياسها.
من أهم هذه الاختبارات:

- معامل الارتباط بيرسون (Pearson Correlation): يقيس العلاقة الخطية بين متغيرين مستمرين، هذا مع افتراض أن البيانات موزعة توزيعاً طبيعياً، والعلاقة بين المتغيرات خطية. مثل: العلاقة بين ساعات الدراسة والدرجات الأكاديمية.

- معامل ارتباط سبيرمان (Spearman Correlation): يعتمد عليه عندما تكون البيانات ترتيبية أو عندما تكون العلاقة غير خطية. مثل: العلاقة بين ترتيب الطلبة حسب الأداء وترتيبهم حسب النشاط المدرسي.

- اختبار كاي-تربيع (Chi-Square Test) للعلاقات: يعتمد عليه لاختبار وجود علاقة بين متغيرين فئويين. مثل: دراسة العلاقة بين الجنس والتخصص الجامعي.

ت. اختبارات المقارنات (Comparison Tests): لمقارنة المتوسطات أو التوزيعات بين مجموعات مختلفة بهدف تحديد الفروق الإحصائية. تتنوع هذه الاختبارات بين المعلمية وغير المعلمية بناءً على طبيعة البيانات وتوزيعها.

ب. 1. الاختبارات المعلمية (Parametric Tests): من أهمها ما يلي:

- اختبار T للعينات المستقلة (Independent Samples T-Test): لمقارنة متوسط مجموعتين مستقلتين. مثل: مقارنة متوسط دخل الرجال والنساء في قطاع معين.

- اختبار T للعينات المرتبطة (Paired Samples T-Test): لمقارنة متوسط نفس المجموعة قبل وبعد التدخل. مثل: مقارنة أداء الطلاب قبل وبعد دورة تدريبية.

- تحليل التباين (ANOVA): وينقسم إلى:

• One-Way ANOVA مقارنة متوسطات أكثر من مجموعتين بناءً على عامل واحد.

• **Two-Way ANOVA** مقارنة تأثير عاملين مستقلين على متغير تابع.

مثال: مقارنة متوسط درجات الطلبة بين عدة كليات أو جامعات.

ب.2. الاختبارات غير المعلمية (Non-Parametric Tests): من أهمها ما يلي:

- اختبار مان-ويتني (Mann-Whitney U Test): يعد بديلاً لاختبار T للعينات المستقلة.

مثل: مقارنة مستوى الرضا الوظيفي بين شركتين.

- اختبار ويلكوكسون (Wilcoxon Signed-Rank Test): يعتبر بديلاً لاختبار T للعينات

المرتبطة. مثل: مقارنة رضا العملاء قبل وبعد تحسين الخدمة.

- اختبار كروسكال واليس (Kruskal-Wallis Test): بديل لـ ANOVA في حالة البيانات تتبع

التوزيع غير الطبيعي. مثل: مقارنة رضا العملاء بين ثلاثة فروع لشركة.

- اختبار كاي-تربيع (Chi-Square Test) للمقارنات: مقارنة توزيع الترددات بين مجموعتين أو

أكثر. مثل: دراسة توزيع التخصصات الأكاديمية بين الطلاب الذكور والإناث.

ث. اختبارات التأثير (Effect Tests): يعتمد عليها في تحديد حجم ومدى تأثير متغير أو أكثر

على متغير تابع. وغالباً ما تستخدم هذه الاختبارات في التحليل السببي. من أهمها:

- الانحدار الخطي (Linear Regression): يستخدم البسيط في قياس تأثير متغير مستقل واحد

على المتغير التابع، أما المتعدد فيقيس تأثير عدة متغيرات مستقلة على متغير تابع. مثل: تأثير

ساعات الدراسة والمسافة على التحصيل العلمي.

- الانحدار اللوجستي (Logistic Regression): يعتمد عليه في قياس تأثير المتغيرات على

حدث ثنائي (مثل النجاح/الفشل). مثل: تأثير التدريب على احتمال اجتياز الاختبار.

- تحليل التباين المشترك (ANCOVA): يستعمل لقياس تأثير متغير مستقل مع التحكم في متغيرات

أخرى. مثل تأثير نوع التعليم على الأداء الأكاديمي مع التحكم في العمر.

- تحليل المسار (Path Analysis): يستخدم في فحص العلاقات السببية بين المتغيرات. مثل: تأثير

الدخل على الأداء الأكاديمي عبر مستوى التعليم.

- اختبار التأثيرات المتداخلة (Interaction Effects): يعتمد عليها في فحص كيفية تفاعل متغيرين مستقلين للتأثير على متغير تابع. ويعتمد على النماذج أو المعادلات الخطية العامة في دراستها (GLM or GEE). مثل: تأثير التفاعل بين التدريب ونوع التعليم على الأداء الوظيفي.
- ج. اختبارات المتغيرات الزمنية (Time Series Analysis): يعتمد عليها في تحليل البيانات التي تعتمد على الزمن لفهم الأنماط أو التنبؤ بالمستقبل. منها:
 - اختبار ديكي-فولر الموسع: (Augmented Dickey-Fuller Test) : لاختبار استقرار السلاسل الزمنية.
 - اختبار كوسرانو: (Cochrane-Orcutt Test) : لتصحيح الارتباط الذاتي في البيانات الزمنية.
 - تحليل التباين الزمني: (Time Variance Analysis) : لتحديد التغيرات الزمنية ومدى استقرار الظاهرة عبر الزمن.

خامسا: معايير اختيار الاختبار الإحصائي المناسب:

إن اختيار الاختبار الإحصائي المناسب يعتبر خطوة محورية لضمان دقة وصحة نتائج التحليل. ويعتمد هذا الاختيار على فهم طبيعة البيانات وأهداف الدراسة بالأساس. ومن المعايير الأساسية التي يجب أن يأخذها الباحث بعين الاعتبار عند اختيار الاختبارات الإحصائية، ما يلي: (17) (18)

- تحديد نوع البيانات
- معرفة التوزيع : البيانات موزعة توزيعا طبيعيا أم لا.
- تحديد الغرض من التحليل :دراسة العلاقات، الفروق، أم التأثيرات.
- تصميم البحث: يجب أن يتماشى تصميم البحث مع الاختبار الإحصائي المختار لضمان صحة النتائج وعموميتها.
- الفرضيات الإحصائية: يجب أن تتماشى الفرضيات الصفرية والبديلة مع الاختبار الإحصائي المختار لضمان دقة التحليل.
- استخدام برامج التحليل المناسبة: مثل SPSS ، R ، Python، أو EViews...الخ.

6. البرمجيات الإحصائية الشائعة في تحليل البيانات:

توفر البرمجيات الإحصائية إمكانيات واسعة لتحليل البيانات وعرضها بطرق تسهل فهمها واتخاذ القرارات المستندة إليها. ومن الضروري اختيار البرنامج بناء على احتياجات الدراسة ومستوى تعقيد التحليل المطلوب، وفيما يلي أهم البرامج التي يمكن الاعتماد عليها في عملية التحليل الإحصائي: (19) (20)

- **MsExcel**: يستخدم بشكل شائع في التحليلات الأساسية وإدارة البيانات. على الرغم من أنه ليس مخصصا للتحليل الإحصائي المتقدم، إلا أنه يوفر أدوات تحليلية مفيدة مثل الجداول والرسوم البيانية.
- **SPSS**: يعتبر من أكثر البرمجيات استخدامًا في التحليل الإحصائي، خاصة في الأبحاث الاجتماعية والسلوكية. يتميز بواجهة مستخدم سهلة الاستخدام ويدعم العديد من التحليلات الإحصائية.
- **EViews**: هو أحد البرمجيات الإحصائية التي تُستخدم بشكل خاص في التحليل الاقتصادي والمالي. يُعتبر أداة قوية لتحليل السلاسل الزمنية والنماذج الاقتصادية.
- **R**: برمجية مفتوحة المصدر تستخدم على نطاق واسع في الأوساط الأكاديمية والبحثية. تتميز بقدرتها على إجراء تحليلات إحصائية متقدمة ومعقدة، وتدعم البرمجة الإحصائية.
- **Stata**: يستعمل بشكل رئيسي في الأبحاث الاقتصادية والاجتماعية، وتوفر أدوات قوية لتحليل البيانات الإحصائية والاقتصادية.
- **Minitab**: يعتمد عليه بشكل شائع في الأبحاث الصناعية والهندسية، وتوفر أدوات لتحليل البيانات الإحصائية وتحسين العمليات.
- **Python**: برمجية تستخدم بشكل متزايد في التحليل الإحصائي بفضل مكتباتها القوية مثل Pandas و NumPy و SciPy، وتعد خيارا جيدا لتحليل البيانات الكبيرة والمعقدة.