تحليل بالبيانات إحصائيا

المحور السابع

2026-2025

د/ إكرام قمري

تُعدّ عملية تحليل البيانات خطوة أساسية في أي دراسة علمية أو بحث تطبيقي، إذ تتيح تحويل البيانات الخام إلى معلومات ذات معنى تساعد في اتخاذ القرارات وبناء الاستنتاجات. وتشمل هذه العملية مراحل متعددة تبدأ بالمعالجة الأولية للبيانات لضمان جودتها ودقتها، ثم الانتقال إلى مراحل التحليل المختلفة التي تهدف إلى اكتشاف العلاقات بين المتغيرات، وفحص الفروقات، وتقدير التأثيرات، وصولاً إلى التنبؤ بالاتجاهات المستقبلية. وتختلف أساليب التحليل المستخدمة حسب طبيعة البيانات، وأهداف الدراسة، والمنهج المتبع.

أولا: المعالجة الأولية للبيانات:

تُعدّ المعالجة الأولية للبيانات مرحلة أساسية تسبق أي عملية تحليل، وتهدف إلى إعداد البيانات وتحسين جودتها حتى تصبح صالحة للاستخدام. وتشمل هذه المرحلة ثلاث خطوات رئيسية: تفريغ البيانات، ترميز البيانات، تنظيف البيانات.

وتتم المعالجة الأولية للبيانات عبر إتباع المراحل التالية:

1. تفريغ البيانات:

يقصد بتفريغ البيانات عملية نقل البيانات المجمّعة من أدوات جمع البيانات المختلفة مثل الاستبيانات، المقابلات، الملاحظات الميدانية، السجلات الإدارية، أو قواعد البيانات وإدخالها في برنامج التحليل الإحصائي أو الجدولي مثل Excel، SPSS، Stata، R أو غيرها.

وتمثل هذه المرحلة الخطوة الأولى في المعالجة الأولية للبيانات، وهي أساسية لأنها تؤثر مباشرة في دقة النتائج النهائية.

وتهدف هذه العملية إلى:

- تحويل البيانات من الشكل الورقي أو الخام إلى ملف قابل للمعالجة؛
- تنظيم المعلومات في متغيرات وصفوف وأعمدة؛
- تقليل أخطاء الإدخال.

وتتطلب عملية تفريغ البيانات مايلي:

- مراجعة كل قيمة مدخلة لتجنب أخطاء الكتابة أو الأرقام؛
- وضع أسماء واضحة للمتغيرات (Variable Names)؛
- وضع تسميات تفسيرية للمتغيرات (Variable Labels) ولمستويات الإجابة (Value Labels).

2. ترميز البيانات:

يُعدّ ترميز البيانات مرحلة أساسية في المعالجة الأولية للبيانات، ويُقصد به تحويل الإجابات أو المتغيرات النصية إلى رموز رقمية بهدف تسهيل إدخالها، تحليلها، ومعالجتها باستخدام البرامج الإحصائية مثل SPSS، Stata، R، Excel.

وتهدف عملية ترميز البيانات إلى تحقيق ما يلي:

- جعل البيانات قابلة للتحليل الإحصائي، لأن معظم البرامج تتعامل مع الأرقام؛
- تنظيم الإجابات وتوحيد شكلها؛
- تقليل الأخطاء الناتجة عن الاختلاف في كتابة النصوص؛
- تبسيط عمليات المقارنة، التحليل، والنمذجة.

وتتصف عملية الترميز الناجحة بالخصائص التالية:

- الوحدة والاتساق: نفس الفئة يجب أن تحمل نفس الرقم في كل الجداول؛
- الوضوح: توثيق جدول الترميز في ملف منفصل أو داخل البرنامج؛
- عدم التحيز: الأرقام تُستخدم فقط لتمثيل الفئات، وليس لافتراض تفوق فئة على أخرى (إلا إذا كان المتغير ترتيبيًا)؛
- قابلية الفهم: أن يفهم الباحث والمتلقي الترميز بسهولة

3. تنظيف البيانات:

وهي المرحلة الثالثة والأخيرة قبل الانتقال إلى تحليل البيانات، وهي عملية تهدف إلى التأكد من أن البيانات صحيحة وكاملة ومتسقة وخالية من الأخطاء.

وتتضمن عدة إجراءات أهمها:

1.3 معالجة القيم المفقودة:

تُعد القيم المفقودة إحدى المشكلات الشائعة في قواعد البيانات، وتظهر عندما تكون بعض الخانات فارغة أو لم يتم تسجيلها أثناء جمع البيانات. وقد يؤدي تجاهلها إلى انحراف النتائج، لذلك يجب التعامل معها بطريقة علمية قبل البدء في التحليل الإحصائي.

وهناك عدة عامل تتسبب في ظهور القيم المفقودة من بينها:

- عدم إجابة المبحوث على بعض الأسئلة؛
- خطأ في إدخال البيانات أو تفريغها؛
- فقدان جزء من الاستمارة أو عدم وضوح بعض الإجابات؛
- رفض المبحوث تقديم معلومات خاصة أو حساسة؛
- أعطال تقنية في أجهزة جمع البيانات.

ويتم معالجة القيم المفقودة من خلال إحدى الطرق التالية:

- **الحذف:** يستخدم عندما تكون نسبة القيم المفقودة قليلة جدًا؛
- **التعويض البسيط:** استبدال كل قيمة مفقودة بمتوسط المتغير، يتناسب بشكل كبير مع المتغيرات العددية؛

- **التعويض المتقدم:** يفضّل عندما تكون القيم المفقودة كثيرة أو غير عشوائية، حيث يتم تقدير القيم المفقودة بإستخدام نموذج إنحدار يعتمد على متغيرات أخرى.

2.3 اكتشاف القيم الشاذة:

القيم الشاذة هي القيم التي تختلف بشكل كبير عن باقي القيم في المتغير، وقد تكون نتيجة خطأ في الإدخال، أو اختلاف طبيعي في العينة، أو ظاهرة اقتصادية/سلوكية غير مألوفة.

وجود قيم شاذة قد يؤدي إلى انحراف المتوسطات، إضعاف دقة النماذج الإحصائية، وتغيير اتجاه العلاقات بين المتغيرات، لذلك يجب اكتشافها قبل تحليل البيانات.

من أسباب ظهور مشاكل القيم الشاذة ما يلي:

- أخطاء في إدخال البيانات (مثل: 220 بدلًا من 22)؛
- خطأ في قياس المتغير أو أداة جمع البيانات؛
- سلوك غير طبيعي أو حالة استثنائية في البيانات؛
- اختلاف جوهري بين بعض المبحوثين وبقية العينة؛
- تغيرات اقتصادية مفاجئة (تضخم، أزمة مالية...).

ويتم التعامل مع القيم الشاذة بعد إكتشافها كما يلي:

- حذف القيمة إذا كانت خطأ إدخال واضح؛
- تعديل القيمة إذا كان الخطأ واضحًا؛
- الاحتفاظ بالقيم إذا كانت تمثل واقعًا اقتصاديًا أو ظاهرة مهمة؛
- استخدام طرق تحليل مقاومة للقيم الشاذة مثل الوسيط بدل المتوسط، نماذج الانحدار.

3.3 إكتشاف التكرارات:

التكرارات هي السجلات أو الصفوف التي تتكرر في قاعدة البيانات، سواء بشكل كامل أو جزئي، فوجود تكرارات يؤدي إلى انحراف النتائج، تضخيم الأوزان النسبية لبعض الأفراد، والتأثير على المتوسطات والانحدارات والتحليلات الاقتصادية.

ومن أسباب حدوث التكرارات ما يلي:

- إدخال نفس الاستمارة مرتين؛
- جمع البيانات من أكثر من مصدر بدون دمج صحيح؛
- أخطاء في تفريغ البيانات في Excel أو SPSS؛
- ضغط زر الإدخال مرتين دون قصد.

4.3 اكتشاف القيم غير المنطقية:

هي القيم التي تختلف بشكل كبير عن بقية القيم في مجموعة البيانات، بحيث تبدو غير متوقعة أو غير متسقة مع النمط العام. قد تكون:

- نتيجة خطأ في إدخال البيانات؛
- قيمة حقيقية لكنها نادرة جدًا؛
- إشارة إلى ظاهرة خاصة يجب تحليلها.

نقرر أن القيمة غير منطقية إذا كانت:

- بعيدة جدًا عن بقية القيم عددياً؛
- غير متسقة مع السياق (مثلاً: عمر شخص = 250 سنة)؛
- تؤثر سلبًا على المتوسط الحسابي، النتائج الإحصائية، النماذج التنبؤية.

ويتم معالجة القيم غير المنطقية حسب طبيعة الدراسة كما يلي:

- تصحيحها إذا كان الخطأ واضحًا؛
- تعويضها بمتوسط المجموعة أو الوسيط؛
- استبعادها إذا كانت نتيجة خطأ في التسجيل؛
- الإبقاء عليها إذا كانت تمثل ظاهرة مهمة (مثل قيم عالية في التكاليف أو المبيعات).

ثانيا: التحليل الإحصائي للبيانات:

تُعدّ التحليلات الإحصائية من أهم الأدوات التي يعتمد عليها الباحثون في مختلف العلوم الاجتماعية والاقتصادية، لأنها تسمح بفهم البيانات، واختبار الفرضيات، واكتشاف العلاقات بين المتغيرات، والقدرة على التنبؤ بالسلوك المستقبلي.

وتنقسم هذه التحليلات إلى أربعة محاور أساسية: تحليل الفروقات بين المجموعات، تحليل العلاقات بين المتغيرات، تحليل التأثيرات، ثم التحليلات الخاصة بالتنبؤ.

1. التحليلات الإحصائية الخاصة بالفروقات بين المجموعات:

تركّز على مقارنة متوسطات المجموعات، بهدف معرفة هل توجد فروقات معنوية بين عينات مختلفة.

يتم إستخدام عدة نماذج إحصائية لتحديد الفروقات بين المجموعات من بينها:

1.1 اختبار T للعينات المستقلة:

يقيس الفرق بين متوسطَيْ مجموعتين مستقلتين لمعرفة هل الفروق بينهما معنوية. ومن أهم شروطه:

- المجموعتان مستقلتان تمامًا؛

- التوزيع طبيعي في كل مجموعة؛
- تجانس تباين المجموعتين؛
- المتغير التابع كمي.

2.1 اختبار T للعينات المترابطة:

يقيس الفرق بين متوسطَيْ قياسين لنفس المجموعة: قبل/بعد، أو قياسين لنفس الأفراد. من شروطه:

- القياسات مرتبطة (نفس الأشخاص)؛
- القياسات مرتبطة (نفس الأشخاص)؛
- المتغير التابع كمي.

3.1 تحليل التباين الأحادي ANOVA:

يقيس وجود فروقات معنوية بين ثلاث مجموعات فأكثر من حيث متوسط المتغير التابع. تتمثل شروط استخدامه في:

- المتغير التابع كمي؛
- المتغير المستقل فئوي (3 مجموعات أو أكثر)؛
- تجانس التباين بين المجموعات؛
- التوزيع الطبيعي.

2. التحليلات الإحصائية الخاصة بالعلاقات بين المتغيرات:

تهدف إلى قياس قوة واتجاه العلاقة بين متغيرين أو أكثر، وهل العلاقة خطية أو رتبية. ويعتمد في قياس هذه العلاقة على مختلف معاملات الإرتباط والتي من بينها:

1.2 معامل الارتباط بيرسون:

يقيس قوة واتجاه العلاقة الخطية بين متغيرين كميين. وتتمثل شروطه في:

- المتغيران كميّان؛
- العلاقة خطيّة؛
- التوزيع الطبيعي؛
- عدم وجود قيم شاذة جدًا.

2.2 معامل الارتباط سبيرمان:

يقيس قوة واتجاه العلاقة الرتبية أو غير الخطية بين متغيرين. من شروطه:

- المتغيرات قد تكون رتبية أو كميّة؛
- مناسب إذا لم يتحقق شرط الطبيعية؛
- مناسب إذا كانت العلاقة غير خطية.

3. التحليلات الإحصائية الخاصة بالتأثيرات بين المتغيرات:

تركّز على دراسة تأثير متغير أو عدة متغيرات مستقلة على متغير تابع، وتفسير حجم هذا التأثير.

1.3 الانحدار الخطي البسيط:

يقيس تأثير متغير مستقل واحد على متغير تابع، ويعطي القدرة على التنبؤ بالنتيجة. شروط إستخدامه تتمثل:

- المتغير التابع كمي؛
- العلاقة بين المتغيرين خطية؛
- لا توجد قيم شاذة مؤثرة؛
- استقلالية الأخطاء؛
- تجانس التباين.

2.3 الانحدار الخطي المتعدد:

يقيس تأثير عدة متغيرات مستقلة على متغير تابع واحد، ويحدد المتغيرات الأكثر تفسيرًا. من شروطه:

- المتغير التابع كمي؛
- عدم وجود Multicollinearity بين المتغيرات المستقلة؛
- استقلالية الأخطاء؛
- العلاقة خطية؛
- فحص القيم المتطرفة.

3.3 تحليل التباين المشترك ANCOVA:

يقيس الفروق بين المجموعات مع التحكم في أثر متغيرات أخرى تسمى Covariates، ويعطي تأثيرًا "صافيًا" للمتغير المستقل على المتغير التابع. من شروطه:

- وجود متغيرات مستقلة فئوية + متغيرات مرافقة كميّة؛
- العلاقة بين المتغيرات المرافقة والمتغير التابع خطية؛
- تجانس انحدار المتغيرات المرافقة عبر المجموعات.

4. التحليلات الإحصائية الخاصة بالتنبؤ:

تهدف إلى بناء نماذج تقدّر القيم المستقبلية للمتغيرات اعتمادًا على البيانات الحالية أو الماضية.

1.4 نماذج الانحدار التنبؤية:

تقدّر القيم المستقبلية للمتغيرات باستخدام بيانات حالية. تتمثل شروط إستخدامه في:

- جودة النموذج (اختبارات: R^2 ,MAE ,RMSE)؛
- ثبات العلاقة بين المتغيرات؛
- كفاية حجم العينة.

2.4 التنبؤ بالانحدار التدريجي:

يختار أفضل المتغيرات التي تؤثر على المتغير التابع بطريقة آلية. شروطه تتمثل في:

- البيانات كميّة؛
- عدم وجود Multicollinearity حادة؛
- حجم عينة كافٍ.

3.4 تحليل السلاسل الزمنية:

يحلل البيانات عبر الزمن لاكتشاف الاتجاه (Trend)، الموسمية (Seasonality)، الدورية (Cyclic)، التذبذب غير المنتظم، ويُستخدم للتنبؤ بالقيم المستقبلية.

شروطه تتمثل في:

- البيانات يجب أن تكون مرتبة زمنيًا؛
- ثبات السلسلة أو تحويلها إلى Stationary (مثل ARIMA)؛
- عدم وجود فجوات زمنية كبيرة.