

المحاضرة النظرية الثانية في الإحصاء

التحليل الإحصائي DATA ANALYSIS

إعداد: د. حيدر الحسن

18/10/2020

تعريف ومصطلحات:

التحليل الإحصائي: يعني تحليل البيانات وهو ترتيب وتبويب البيانات ومن ثمّ حساب المؤشّرات الوصفية لها بالإضافة إلى تطبيق تحاليل واختبارات أخرى من أجل تفسير النتائج وعرضها بطريقة مفهومة بالاستعانة بوسائل العرض البياني وأخيرا تعميمها ونشرها.

المراحل الأساسية للبحث (أو التجربة) الإحصائي:

أولاً- مرحلة التخطيط: وضع الخطة لإجراء البحث.

ماذا تتضمن خطة البحث؟:

1- عرض المشكلات (مشكلات البحث):

وفيها يتم تقديم عرض مختصر ومركز حول مشكلة البحث بحيث يتم تسليط الضوء على المشكلة العلمية التي يتمحور حولها البحث.

2- تحديد أهداف البحث:

هنا يتم صياغة الأهداف بدقة علمية إلى أقصى حدّ ممكن ويفضّل أن تتم صياغة الأهداف على شكل أسئلة واضحة ومختصرة ودقيقة ولا يوجد عدد محدد للأهداف من الممكن أن يكون عدد الأهداف (1 - 2 - 3 - 4 -) وذلك حسب رغبة الباحث.

3- تحديد المادة التجريبية:

هي مادة التجربة التي تطبق عليها أهداف الاختبارات والمعاملات المطلوبة.

4- الفرضيات الإحصائية:

يتم تحديد الفرضيات الإحصائية طبقاً لأهداف البحث وهي في الواقع عبارة عن تعبير رياضي أو تعبير بطريقة رياضية عن الأهداف ويتمّ هنا تحديد الفرضيات الأولية والبديلة لها في حال تمّ رفضها.

5- تحديد المتغيرات أو العوامل المطلوب دراستها من البحث:

تحدد العوامل في البحث من خلال تحديد المؤثرات على المادة التجريبية **وتحديد نوعها (هل هي كمية أم نوعية) وهل هي عشوائية أم لا عشوائية.**

6- المعاملات:

هي عبارة عن أشكال أو كميات أو تدرجات أو أنواع أو رموز للعوامل والمتغيرات في البحث.

7- المكررات:

هي عبارة عن عدد مرات تكرار المعاملات في البحث أي كم مرة نحتاج إلى إعادة المعاملات في التجربة.

8- الصفة المختارة (الصفة المدروسة):

هي الصفة التي يختارها الباحث، يراقب تغيرات قيم هذه الصفة خلال فترة البحث. وهي إما أن تكون صفة كمية أو نوعية وعادةً يتم تحويل النوع الثاني (النوعي) إلى نوع كمي حتى يصبح التحليل والتقويم لها ممكناً.

9- المجال التجريبي:

هو المجال الذي يتم تجربة المعاملات ضمنه، وهذا المجال يمكن أن يكون كمّي أو نوعي بحسب نوع المتغيرات أو العوامل المدروسة للبحث.

10- حجم العينة أو التجربة:

يرمز له بالرمز (n) وهو يساوي جداء عدد المعاملات (t) في عدد المكررات (r).

$$n = r * t$$

11- الدقة:

هي الدقة المطلوبة التي يجب أن تتوفر خلال البحث أو التجربة وهذه الدقة تختلف بحسب نوع البحث أو التجربة وهي عادةً إمّا 1% أو 5% ولكن نادراً ما تكون 10 أو 25% .

12- التحاليل والاختبارات الإحصائية اللازمة:

يتم اختيار التحاليل والاختبارات الإحصائية اللازمة حسب أهداف التجربة وهذه تحتاج عادة الى خبرة أو بالاستعانة بالاختصاصيين في التحليل الإحصائي.

13- التصميم الرياضي:

لكل تجربة تصميم، وسيتم دراسة التصاميم التجريبية في النصف الأخير من هذا المقرر (تصميم التجارب).

14- الملاحظات Notice:

يتم هنا كتابة أسم الباحث وفريق العمل بالإضافة إلى مواعيد تتعلق بالبحث وملاحظات أخرى.

ثانياً- تحليل البيانات:

يوجد نوعان للتحليل الإحصائي:

1- التحليل الإحصائي الأولي Primary:

نحتاج دائماً إلى إجراء تحليل إحصائي أولي لكل البيانات التجريبية أو بيانات بحث معيّن.

ماذا يشمل التحليل الإحصائي الأولي:

1- وصف البيانات بشكل نظري.

2- حساب المؤشرات الإحصائية الأساسية **Descriptive Statistics**

3- إنشاء جدول التوزيع التكراري.

4- العرض البياني للبيانات.

تتضمن المؤشرات الإحصائية الأساسية ما يلي:

- أ- المتوسط أو المعدل الحسابي Average أو Main.
- ب- الوسيط الحسابي Median.
- ج- المنوال أو القمة Mode.
- د- الخطأ القياسي Standard Error.
- هـ- التباين أو التشتت Variation.
- و- معامل الاختلاف Coefficient of Variation.

يشمل التحليل الإحصائي الأولي:

1- وصف البيانات بشكل نظري: وهو يعني وصف نوع وطريقة جمع البيانات بالإضافة إلى الأساليب والطرق المستخدمة في ترتيب وجمع البيانات كما يشمل آليات الوصول إلى مصادر هذه البيانات إن كانت تقليدية أم أنها متطورة.

2- التوزيع التكراري: هو عملية ترتيب وتنظيم البيانات في جداول نطلق عليها جدول التوزيع التكراري بحيث يمكن إدارة ومعالجة البيانات بطريقة سهلة ودقيقة بدون فقدان أي تفاصيل مهمة في البيانات.

3- حساب المؤشرات الإحصائية للعينة أو العينات المتوفرة قيد التحليل ومن ثم حساب مؤشرات وصفية تقديرية للمجتمع الذي أخذت منه العينات.

4- إنشاء العروض البيانية: وهنا يفضل استخدام الرسوم البيانية التي توضح الغاية من البحث أو التجربة بشكل واضح، وذلك حسب المجال الذي ستستخدم البيانات فيه سواء المجال العلمي أم الإرشادي أم الدعاية والإعلان.

2- التحليل الإحصائي المتقدم Advanced ويشمل:

أ- البحث عن التوزيع الرياضي المناسب للبيانات واختبار مدى صلاحية البيانات باستخدام الاختبارات المناسبة ومعالجة القيم الشاذة والمفقودة.

ب- تطبيق التحاليل الإحصائية اللازمة مثل: (تحليل التباين المشترك – التحليل العنقودي – تحليل الصفوف).

ج- استخراج أو استنتاج النموذج الرياضي المناسب للبيانات.

د- تطبيق الاختبارات الإحصائية اللازمة مثل: (الاختبارات المعنوية – اختبارات مقارنة المتوسطات – اختبارات إضافية متعددة).

هـ- تحليل أو دراسة العلاقات باستخدام التحاليل المعروفة مثل: (تحليل الانحدار والارتباط).

ملاحظة:

دائماً يلزمنا إجراء التحليل الإحصائي الأولي أما التحليل الإحصائي المتقدم فيتم استخدامه عند الضرورة إذا كانت البيانات أو أهداف البحث تحتاج لذلك.

1- مقاييس النزعة المركزية: المؤشرات الإحصائية

النزعة المركزية طبيعة عامّة لكل المجتمعات أو ميزة تميّز كل المجتمعات الطبيعية وتعني النزعة المركزية رغبة أو نزوع الأفراد في المجتمع حول المركز أي في كل مجتمع من المجتمعات الطبيعية يحاول أفراد هذا المجتمع التجمع حول المركز.

مثال: لو أخذنا أطوال المجتمع البشري في سوريا: لو أخذنا أطوال (500) شخص نلاحظ أنّ الأطوال تتراوح حول مجال معيّن (150 – 180) سم بشكل عام. المتوسط لهذه الأطوال أو المركز هو (165) سم.

إذا كان لدينا (500) شخص نلاحظ أنّ (400) شخص طوله قريب من (165) سم ونلاحظ عدد قليل طوله (180) سم.

مثال آخر: لو أخذنا ضغط الدم مثلاً عند (500) شخص نلاحظ أنّ حوالي (400) شخص يكون ضغط الدم عندهم (12/8) ونلاحظ حوالي (30) شخص ضغط الدم عندهم (10/7) و(70) شخص عندهم ضغط الدم (14/9).

مقاييس النزعة المركزية:

يوجد مقاييس خاصة لقياس النزعة المركزية وهي:

- 1- المتوسط أو المعدل الحسابي Avarage أو Main.
- 2- الوسيط Median.
- 3- المنوال أو القمة Mode.

1- المتوسط الحسابي أو المعدل الحسابي:

هو من أهم مقاييس النزعة المركزية وهو الأكثر استخداماً ويتمتع بدلالة قوية أو هو مؤشر للنزعة المركزية.

مميزات المتوسط الحسابي:

- أ- سهل جداً للحساب حيث يمكن حسابه بسهولة كبيرة فهو يساوي المجموع/العدد.
- ب- يعطي دلالة قوية جداً عن العينة التي تمّ الحساب منها.
- ج- يمكن حسابه من أي نوع من العينات سواء كانت عينات بسيطة أو غير ذلك ولكن لا يمكن حسابه للمجتمع الطبيعي إلا إذا عرفنا العدد الكلي.

عيوب المتوسط الحسابي:

- يتأثر المتوسط الحسابي بالقيم المتطرفة أي أنه يتأثر بمدى التباين أو التفاوت بين القيم.

مثال:

لدينا عينة مؤلفة من (3 - 4 - 5).

ولدينا عينة أخرى (1 - 2 - 9).

ولدينا عينة ثالثة مؤلفة من (1 - 1 - 10).

لو قمنا الآن بحساب المتوسط الحسابي لهذه العينات نلاحظ أنّ المتوسط هو نفسه بالرغم من اختلاف الأرقام المكونة لكل عينة من العينات والمتوسط يكون للعينات الثلاث السابقة (4).

فمشكلة المتوسط الحسابي يتأثر بالقيم المتطرفة ولا يتأثر بمدى اختلاف أفراد العينة.

مقارنة بين العينة والمجتمع:

المجتمع	العينة
عدد الأفراد غير معروف وغير محدد ويرمز لعدد الأفراد بالرمز (N). لا يمكن حساب متوسط المجتمع (μ) لأن العدد الكلي غير معروف.	عدد الأفراد معروف يرمز له ب (n) وبالتالي متوسط العينة معروف x' ويمكن حسابه.
المتوسط في المجتمع تقديري أي يتم تخمينه ضمن مجال يسمى مجال الثقة.	المتوسط في العينة هو حسابي أي يتم حسابه.

كيف يمكن حساب المتوسط أو المعدل العام للمجتمع:

بالنسبة للمجتمع يوجد لدينا متوسطين:

(μ) متوسط المجتمع وهو المتوسط الحقيقي.

(m) متوسط المجتمع التقديري أو التقريبي والذي يمكن حسابه تقديراً أو تقريباً ولا يمكن معرفة المتوسط الحقيقي للمجتمع.

- نقوم بحساب المتوسط التقريبي ثم نحسب مجال المتوسط الحقيقي.

المتوسط التقريبي نقدره بالطريقة التالية:

- 1- نأخذ أكبر عدد ممكن من العينات الممثلة للمجتمع.
- 2- نحسب متوسط كل عينة.
- 3- نحسب متوسط متوسطات للعينة.
- 4- نحصل بذلك على المتوسط الحسابي التقريبي.

درجات الثقة:

- مجال الثقة عند احتمال خطأ (1 % أو 5 % أو 10 % أو غير ذلك).

- هناك مقولة لعالم إحصاء ألماني يقول فيها: "إن دراسة أي ظاهرة من ظواهر الطبيعة عرضة للأخطاء واحتمال كل حدث مرهون بالخطأ".

- كذلك أنشتاين يقول كل شيء في الكون نسبي ولا يوجد شيء مطلق.

وإلى اللقاء في المحاضرة القادمة