## المحاضرتين التاسعة والعاشرة

# تجهيز وتحليل البيانات

- تجهيز البيانات

- مراجعة البيانات
- ترميز البيانات
- إدخال البيانات
- جدولة البيانات

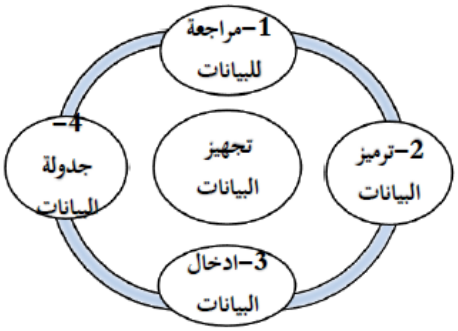
- تحليل البيانات

- أساليب تحليل البيانات
- محددات اختيار أسلوب تحليل البيانات
- تحليل البيانات بمتغير واحد
- تحليل البيانات بمتغيرين
- التحليل متعدد المتغيرات

- تجهيز البيانات

بعد تجميع البيانات باســـتخدام المعاينة تأتي مرحلة تجهيز البيانات وتتكون هذه المرحلة من عدة خطوات نبرزها في التالي:

- **مراجعة البيانات**
- **ترميز البيانات**
- **ادخال البيانات الى الحاسوب**
- **جدولة البيانات**




**1.مراجعة البيانات:Data Review**

تعد عملية المراجعة الخطوة الأولى في إعداد البيانات، وتتمثل هذه العملية في **التدقيق وفحص قوائم الأســئلة المجمعة** والإجابات التي عليها وتنقيحها والتصــرف في البيانات غير الكاملة أو غير الصــحيحة حتى تصــبح هذه القوائم كاملة موثوقا فيها بقدر الإمكان والاعتماد عليها في المراحل التالية من ترميز وتحليل، وتحتاج عملية المراجعة إلى خبرة ودراية خاصة بها.

- **هي عملية يتم من خلالها التأكد من مدى دقة وصحة وكفاية البيانات التي تم جمعها.**
- **ومن ثم مدى صلاحيتها كأساس يعتمد عليه في المراحل التالية للبحث.**
- **أي أن عملية المراجعة تهدف إلى تحديد الإجابات السليمة وحذف البيانات الغير صحيحة.**

### 2.ترميز البيانات:Data Encoding

تتضـــمن عملية ترميز البيانات **قيام الباحث بإعطاء المتغيرات محل الدراســة بأبعادها المختلفة قيما رقمية** أو حروفا هجائية بشـكل متسـلسـل، للتعبير عن الإجابات الواردة بطريقة مختصـرة ومقبولة **يسـهل إدخالها إلى الحاسوب ومعالجتها بواسطته، بحيث تصبح هذه القيمة الرقمية المفتاح للوصول إلى تلك الفئة أو أية معلومات عنها** و للتمييز عن باقي الإجابات المحتملة من نفس السـؤال، بحيث تأخذ كل إجابة رمزا مخصـصـا لها دائما في كل الاستمارات، ويمكن الاستغناء عن عملية الترميز عندما يكون عدد الأسئلة محدودا وعدد الإجابات على هذه الأسئلة محدودا أيضا.

**كيفية ترميز البيانات.**

- **الترميز المسبقPre-encoding** : وهو الترميز الذي يستخدم في حالة الأسئلة ذات النهاية المغلقة، حيث يتيح للباحث فرصـــة التحديد المســـبق للرموز أي إمكانية تجهيز أو بناء ورقة الترميز قبل إجراء الدراســة الميدانية، وذلك للعلم المســبق بعدد الأســئلة وعدد احتمالات الإجابة، حيث يتم ترميز كل قسم من الأسئلة تسلسليا.

- **الترميز اللاحق:Subsequent coding** : يخص هذا النوع من الترميز الأسئلة المفتوحة، ففي هذه الحالة لا يمكن إعداد نظام الترميز مســبقا لعدم القدرة على معرفة احتمالات الإجابة وبالتالي فان ترميز الإجابات يتم بعد الانتهاء من البحث والحصـول على كل الإجابات، حيث يكون على الباحث تصـــميم النظام المناســـب بترميز البيانات بصـــورة تجعل عملية تحليلها ســـهلة وممكنة يدويا وآليا.

### 3.إدخال البيانات الى الحاسوب:Enter data to the computer

القيام بإدخال البيانات التي تم ترميزها إلى الحاســـب الآلي باســـتخدام البرامج الجاهزة المناســـبة وفي هذا الصدد يفضل استخدام برنامج إدخال البيانات المتوفر على حزمة الأساليب الإحصائية والمعروف باسم spss.

4. **جدولة البياناتData Tabulation**: حصر المستقصى منهم بالنسبة لكل خاصية أو متغير ويمكن أن تكون هذه الجدولة يدويا أو آليا.

أ. **الجدولة البسيطة**: هي **إعداد جدول لمتغير واحد** وتعرف أيضا بالجدولة ذات اتجاه واحد وهي التي تقوم على إعداد الجداول لمتغير واحد وكذلك بعض الإحصائيات الوصفية مثل التكرارات وعادة تستخدم لأغراض أخرى هامة في مرحلة التحليل لإجراء تلخيص إحصائي وصفي.

ب. **الجدولة المتقاطعة أو الجدولة المتعامدة: هي إعداد جدول لمتغيرين أو أكثر، وبالرغم** من أن الجدولة في اتجاه واحد تعتبر مفيدة عند تحليل بيانات المتغير الواحد وتوفير ملخص إحصائي مبسط، إلا أن الجدولة المتعامدة تعد الطريقة الأنسب لفحص وتحليل البيانات الخاصة بين متغيرين أو أكثر والعلاقة بينهما ولكن بالرغم من شيوع استخدام الجدولة المتقاطعة في البحوث التسويقية، إلا أنها تعتبر محدودة الفائدة وقاصرة على أساليب التحليل الإحصائية البسيطة فضلا عن صعوبة تصميمها في حالة العلاقة بين أكثر من متغيرين اثنين.

## ـ تحليل البيانات

يقصد بتحليل البيانات استخدام الاساليب الاحصائية لدراسة العلاقات والتأثيرات الموجودة بين متغيرات الدراسة.

1. **أساليب تحليل البيانات:** هناك نوعان من أساليب التحليل:

أ. **التحليل النوعي:Qualitative Analysis**

- ينطلق من مقاربات شخصية ترى الفرد وحدة مركبة يجب دراستها وتفسير سلوكها.
- يمكن التحليل النوعي من معرفة اللاشعور عند المستهلك وحوافزه.

ب. **التحليل الكمي:Quantitative Analysis**

- يعتبر التحليل الكمي أن الفرد شخص منطقي يتأثر سلوكه بعدد من المتغيرات التي يجب دراستها وتوضيحها.
- يستخدم التحليل الكمي تقنيات التحليل الإحصائي.

2. **محددات اختيار أسلوب تحليل البيانات:** تتمثل هذه المحددات في:

أ. **نوعية البيانات:**

- **البيانات الوصفية:** وهي البيانات التي تكون في صورة غير كمية مثل لون السيارة.
- **البيانات الكمية:** وهي البيانات التي تكون في صورة رقمية مثل سعر السلعة.

ب. **تصميم البحث** يتوقف على:

- عينة واحدة أو عدد من العينات المستقلة.
- عدد المتغيرات الداخلة في التحليل.

**3. تحليل البيانات بمتغير واحد:Analyze data with one variable في هذا النوع من الأساليب نحاول دراسة بحث يحتوي على متغير وحيد.**

إن البيانات الخام التي يتم جمعها عن ظاهرة معينة لا يمكن وصفها وتفسيرها وهي في هيئتها الاولى كبيانات اولية، الامر الذي يتطلب تلخيص ووضع هذه البيانات في جداول خاصة او رسوم بيانية هندسية بهدف تسهيل دراستها وإجراء التحليل الاحصائي عليها.

- **عرض البيانات جدوليا:** وهو وضع البيانات في جدول إحصائي، ويختلف شكل الجدول الاحصائي باختلاف نوع البيانات.
- **عرض البيانات بيانيا:** وهو التعبير عن البيانات بالشكل البياني حسب طبيعة المتغير.

◄ **المقاييس الإحصائية للبيانات:**

**يمكننا حصر أهم المقاييس الإحصائية للبيانات في مجموعتين كما يلي:**

ا. **مقاييس النزعة المركزية (Measures of Central Tendency)** هي قيم مركزية (متوسطة) تتمركز أو تتوزع حولها البيانات.

❑ **الوسط الحسابي mean مجموع القيم على عددها.**

$$\bar{x} = \frac{\sum x}{n}$$

- **بعض مميزات الوسط الحسابي:**
  - مقياس سهل حسابه ويخضع للعمليات الجبرية بسهولة.
  - يأخذ في الاعتبار جميع القيم محل الدراسة.
  - أكثر المقاييس استخداما في الإحصاء.
- **بعض عيوب الوسط الحسابي :**
- يتأثر بالقيم الشاذة (المتطرفة) وهي القيم الكبيرة جدا أو الصغيرة جدا مقارنة بباقي القيم .
- يصعب حسابه في حالة الجداول التكرارية المفتوحة، حيث يتطلب ذلك معرفة مركز كل فئة .
- لا يمكن حسابه في حالة البيانات الوصفية.

❑ **الوسيط Median** القيمة التي يقل عنها 50% من مفردات العينة.

$$\frac{n+1}{2}$$

- إذا كان الناتج عدد صحيح فان الوسيط هو القيمة التي تقع في هذا الموقع مباشرة.
- إذا كان الناتج كسر فان الوسيط هو متوسط القيمتين التي وقع الوسيط بينهما.

❑ **المنوال Mode هو المفردة ذات القيمة الأكثر شيوعا أو تكراراً.**

ب. **مقاييس التشـــتت Dispersion** : هي درجة تقارب أو تباعد البيانات عن بعضـــها البعض، وهي مقاييس عددية تستخدم لقياس درجة انحراف البيانات عن القيمة الوسطية.

- **الانحراف المعياري Standard Deviation** مقدار تشتت القيم عن وسطها الحسابي مقاسا بوحدات المتغير نفسها.

$$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu_X)^2}{N}$$

- **التباين Variance** مربع الانحراف المعياري.

$$\sigma = \sqrt{\sigma^2}$$

- **المدى Range**: هو الفرق بين أكبر قيمة وأقل قيمة من البيانات، ويرمز له بالرمز (R).

- **الخطأ المعياري S.E.mean** مقدار الخطأ الموجود في الوسط الحسابي وهو دلالة على دقة الوسط الحسابي كتقدير لوسط المجتمع.

4.تحليل البيانات بمتغيرين:Data analysis with two variables في هذا النوع من الأســـاليب نحاول دراسة بحث يحتوي على متغيرين فقط.

❑ تحليـل معـامـل التوافق: Analysis of the compatibility coefficientيقيس درجـة الارتباط بين ظاهرتين من المتغيرات النوعية او الوصفية.

$$C = \sqrt{\frac{x^2}{x^2 + N}}$$

❑ أسلوب تحليل الانحدار الخطي البسيط: Linear regression analysisيستخدم لقياس أثر أحد المتغيرين على الأخر، دراســـة وتحليل أثر متغير كمي على متغير كمي آخر، معادلة خط الانحدار تأخذ المعادلة التالية:

$$Y = ax + b$$

حساب قيمة a، b، من خلال المعادلات الرياضية التالية:

$$a = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2}$$

$$b = \bar{y} - a.\bar{x}$$

☐ **تحليل الارتباط الخطي بين متغيرين:** Correlation linear يستخدم تحليل الارتباط في تقدير درجة الارتباط الخطي (مدى وجود علاقة خطية) بين متغيرين كميين، وتقاس تلك العلاقات بمقياس يسمى معامل الارتباط ويرمز له بالرمز r ويأخذ القيم من –1 إلى 1.

- يكون الارتباط طردي تام إذا كانت قيمة معامل الارتباط تساوي 1.
- يكون الارتباط عكسي تام إذا كانت قيمة معامل الارتباط تساوي–1.
- لا يوجد ارتباط إذا كانت قيمة معامل الارتباط تساوي صفر.
- كلما كانت القيمة المطلقة لمعامل الارتباط قريبة من الواحد كان الارتباط قويا.
- كلما كانت القيمة المطلقة لمعامل الارتباط قريبة من الصفر كان الارتباط ضعيفا.

**من مقاييس الارتباط:**

- **معامل بيرسون Pearson:** لقياس درجة الارتباط بين **متغيرين كميين.**
- **معامل سبيرمان: Spearman** لقياس درجة الارتباط بين **متغيرين رتبيين، او بين المتغير الرتبي والمتغير الكمي.**
- **الارتباط الجزئي: Partial Linear Correlation**

يستخدم الارتباط الجزئي لقياس الارتباط **بين متغيرين بمعزل عن تأثير المتغيرات الأخرى**، فقد يبدو معامل الارتباط البسيط بين متغيرين (على غير الواقع) كبيراً وذا دلالة إحصائية، لأن متغيراً ثالثاً أو مجموعة من المتغيرات تؤثر في المتغيرات معاً.

**يقيس معامل الارتباط الجزئي الارتباط الفعلي بين المتغيرين، بعد أن يعزل تأثير المتغيرات الأخرى عنهما.**

△ الفرق بين ضـــرورة اســـتخدام الانحدار واســـتخدام الارتباط: إذا كان الغرض من التحليل هو تحديد نوع وقوة العلاقة بين متغيرين، يســـتخدم تحليل الارتباط، وأما إذا كان الغرض هو دراســـة وتحليل أثر أحد المتغيرين على الآخر، يستخدم تحليل الانحدار.

5. **التحليل متعدد المتغيرات: Multivariate Analysis في** هذا النوع من الأســـاليب نحاول دراســـة بحث **يحتوي على أكثر من متغيرين**، ونريد دراســـة العلاقة والارتباط او الأثر او مدى تأثير متغير على المتغيرات الأخرى.

**أ. التحليل العاملي: Factor Analysis** هو أسـلوب إحصـائي يسـتخدم في دراسـة الظواهر **بهدف إرجاعها إلى العوامل المؤثرة فيها**، وهو عملية رياضـية تسـتهدف **تفسـير معاملات الارتباط الموجبة– التي لها دلالة إحصـــائية–بين مختلف المتغيرات ، يلخص المتغيرات في عدد أقل من العوامل** الرئيسية التي يمكن أن تفسر الظاهرة، ويهدف **الى ابراز مجموعة العناصر الكامنة** التي يصعب الكشف عنها والتي **يمكن أن يكون لها دور في تفسير العلاقات** بين عدد كبير من المتغيرات.

**ب. تحليل الانحدار المتعدد:** يهتم بدراسة وتحليل أثر عدة متغيرات مستقلة كمية على متغير تابع كمي.

**ت. تحليل التباين:** دراســة وتحليل أثر متغير او أكثر من المتغيرات الاسمية او الوصــفية على متغير كمي، من اهداف هذا التحليل المقارنة بين متوسطات مجموعات كل متغير من المتغيرات الوصفية محل الدراسة.

.....