



جمع البيانات، نظرية المعاينة

د. مصطفى بابكر



جمع البيانات، نظرية المعاينة والإحصاء الوصفي

جمع البيانات ونظرية المعاينة (Data Collection and Samples Theory)

يستعرض هذا الجزء الطرق المختلفة لجمع البيانات باستخدام العينات الإحصائية والغرض من جمع البيانات.



1. المجتمعات والعينات (Populations and Samples)

• يعرف المجتمع الإحصائي على أنه مجموعة أو تجمع كل المفردات الممكنة والتي تشترك في أو تربطها خصائص وسمات محددة جغرافية كانت أو سياسية أو اجتماعية أو اقتصادية أو غيرها. وتعرف العينة على أنها جزء من المجتمع يتم اختياره بطرق معينة لتمثيل المجتمع المعني والاستدلال على خواصه.

• تعرف مجموعة الطرق والتدابير التي يتم اتباعها في اختيار العينة وتحليلها بتصميم المعاينة (Sampling Design) وتشتمل على تحديد المشكلة، تحديد المجتمع المراد معانيته، تحديد البيانات المطلوبة وطريقة جمعها وقياسها، تحديد وحدة المعاينة ونوع وحجم العينة، ترتيب العمل الميداني، وتلخيص وتحليل البيانات. ويعتبر الهدف الأساسي وراء تصميم المعاينة هو الحصول على عينة تعطي نتائج ذات دقة معينة بأقل تكاليف ممكنة.



● المعاينة ليس مجرد استخدام جزء من المجتمع بدلاً منه كله فحسب بل علم وفن التحكم وقياس دقة المعلومات الإحصائية عن طريق استخدام بعض النظريات الرياضية. وقد أصبحت العينات أساساً في كثير من الدراسات النظرية والعملية كالتخطيط النوعي للإنتاج، البطالة وحجم القوى العاملة، التخطيط الاجتماعي والصحي، التعليم، أبحاث السوق وتحسين الكفاءة وعمليات مراقبة الإنتاج وضبط الجودة على سبيل المثال.

● يجب التنويه أنه ليس في أخذ العينة ما يوحي بأن هذه العملية أقل كفاية أو دقة من عملية التعداد الشامل كما يتبادر للذهن بل على العكس فقد تأتي العينة بنتائج أكثر دقة من التعداد الشامل في عدة ظروف تتعرض لبعضها في سردنا لميزات البحث عن طريق العينة.



2. ميزات البحث عن طريق العينة على المسح الشامل

للبحث عن طريق العينة عدة ميزات منها:

- أ. أن جملة تكاليف البحث من حيث الجهد، الوقت، والمال عموماً أقل في حالة استخدام العينة عنه في حالة استخدام المسح الشامل للمجتمع.
- ب. في حالة الإحتياج للمعلومة لاتخاذ قرارات مهمة وعاجلة فإن العينة هي الطريق الأفضل وذلك لسرعة جمع وتلخيص وتحليل البيانات. أيضاً لسهولة المتابعة في حالة العينة، تزيد معدلات الاستجابة عنها في حالة المسح الشامل.



ج. نظراً لسهولة توفير الإعداد الجيد والتدريب الكافي للعاملين في الميدان والمكتب وسهولة الإشراف على مختلف عمليات المسح يمكن الحصول على بيانات ونتائج أكثر دقة باستخدام العينة.

د. هنالك حالات لا يمكن فيها إجراء البحث عن طريق الحصر الشامل مما يحتم استخدام أسلوب العينة كما في حالة الأسماك والطيور والحيوانات المفترسة.

هـ. لقياس وتقييم دقة نتائج المسح تستخدم نظريات الإحصاء الاستدلالي والتي تركز أساساً على نظرية المعاينة. لذلك فإن الطريقة المتبعة في تقييم دقة نتائج المسح الشامل هي إجراء المسح على عينة مختارة ومقارنة نتائجها بنتائج المسح الشامل.



3. جمع البيانات

- بعد تحديد البيانات المراد جمعها يقوم الباحث بصياغة الأسئلة وتصميم الاستبيان أو الاستمارة طبقاً للطريقة التي يود استخدامها في جمع البيانات. وهناك قواعد خاصة لترتيب وصياغة الأسئلة لضمان الحصول على بيانات موحدة وبدرجة عالية من الثقة من كل أفراد العينة.
- هنالك عدة طرق لجمع البيانات في الدراسات الإحصائية منها طريقة الاتصال المباشر أو المقابلات الشخصية وطرق الاتصال غير المباشر كالبريد والتلفون.



• تستخدم المقابلات الشخصية استمارة التسجيل وفيها يقوم معدّ الاستمارة بطرح الأسئلة على المستجيب وتسجيل الإجابات على الاستمارة، وتستخدم طرق الاتصال غير المباشر أسلوب الاستبيان في حالة البريد حيث يرسل للمستجيب الاستبيان مع مجموعة من التعليمات ويقوم المستجيب بتعبئة الاستبيان وإيداعه في البريد مستخدماً الظرف المرفق مع الاستبيان. أما في حالة التلفون فيستخدم أسلوب استمارة التسجيل وفي كثير من الأحيان يستخدم التلفون أيضاً في شرح وتوضيح التعليمات ومتابعة المستجيب في حالة الاستبيان.



• تعتبر القدرة على تقليل نسبة عدم الاستجابة وتقليل الإجابات الخاطئة من أهم ميزات طريقة الاتصال المباشر وتمثل أهم عيوب الطريقة في ارتفاع تكلفة تحصيل المعلومة وعرضتها لأخطاء العداد كالتسجيل الخاطئ للإجابات والإيجاء والتضليل في طرح الأسئلة. في المقابل تتميز طريقة الاتصال غير المباشر في حالة البريد بخلوها من أخطاء العداد ولكن يعيبها قلة نسبة الاستجابة والفهم الصحيح لأسئلة الاستبيان.



4. التحيز والتغيرات العرضية (Biases and Chance Variation)

- هنالك نوعان من الخطأ يواجههما الباحث عند استخدام المسوحات الإحصائية هما التغيرات العرضية والتحيز. تنتج التغيرات العرضية من أخطاء وظروف خارج إرادة الباحث، العداد أو المستجيب كاحوال الطقس والحالة النفسية للعداد وسوء الطالع وعامل الصدفة وهي بطبيعتها تغييرات عشوائية النمط ولا يمكن التحكم فيها. أما التحيز فينتج من أخطاء إجرائية في تنفيذ المسح كاختيار العينة مثلاً ويمكن تقليله والتحكم فيه بالتصميم الجيد للمسح الإحصائي.



5. أنواع العينة

- هنالك نوعان من العينات هما العينة الاحتمالية والعينة غير الاحتمالية. يتم اختيار المفردات في العينة الاحتمالية حسب قانون الاحتمالات ويكفل لكل فرد نفس الفرصة في الاختيار وتعرف أيضا بالعينة العشوائية (Random Sample) لاستخدامها طرق الاختيار العشوائي المبني على الاحتمالات. يتم اختيار المفردات في العينات غير الاحتمالية باستخدام الطرق الذاتية كالاختيار حسب المتاح من أفراد المجتمع أو الاختيار حسب قناعة الباحث ومعرفة بأفراد المجتمع.



• تتميز العينة الاحتمالية بموضوعيتها وخلوها من تحيز المعاينة من جانب وبإمكانية استخدام نظرية الاحتمالات وطرق الاستدلال الإحصائي في تقييم وتعميم نتائج البحوث المبنية عليها في الجانب الآخر. وفي المقابل تتميز العينة غير الاحتمالية بالسهولة وقلة التكلفة ولكن يعيبها ذاتيتها وعدم إمكانية تطبيق نظريات الاحتمالات والإحصاء الاستدلالي عليها.



- تمثل أهم أنواع العينات الاحتمالية في: العينة العشوائية البسيطة (Simple Random Sample)، العينة العشوائية النظامية (Systematic Random Sample)، العينة العشوائية الطبقية (Stratified Random Sample)، العينة العنقودية (Cluster Sample)، المعاينة عشوائية الاستجابة (Randomized Response Sampling)، والمعاينة المتتابعة (Sequential Sampling).



6. مراحل دراسة العينة

هناك ثلاثة مراحل أساسية في دراسة العينة هي خطة المعاينة، جمع البيانات، وتحليل وتلخيص البيانات.

- مرحلة الخطة تشمل على تحديد المجتمع الإحصائي، تصميم الاستمارة أو الاستبيان، اختيار أسلوب المعاينة، تحديد المعالم المراد تقديرها أو الفرضيات المراد اختبارها واختيار الطرق الإحصائية الملائمة، ومن ثم تحديد حجم العينة.
- تشمل مرحلة جمع البيانات على اختيار أفراد العينة وإجراء المسح.



- مرحلة التحليل والتلخيص تشتمل على إفراغ البيانات، وصف البيانات في صورة جداول وأشكال، حساب القيم الممثلة للعينة كالوسط الحسابي والانحراف المعياري، تقدير معالم المجتمع، اختبار الفرضيات وتلخيص النتائج.



الإحصاء الوصفي

د. مصطفى بابكر



الإحصاء الوصفي

يعني هذا الجزء بتعريف المشارك بالطرق الإحصائية المستخدمة في وصف البيانات. وتتمثل هذه الطرق في الرسومات البيانية، التوزيعات التكرارية ومقاييس النزعة المركزية ومقاييس التشتت.



1. الرسومات البيانية

- تعتبر الرسومات البيانية أكثر الطرق الإحصائية استخداماً في وصف وتلخيص البيانات، وذلك لتميزها بالبساطة والسهولة والوضوح.
- هنالك أنواع عديدة من الأشكال والرسومات يمكن استخدامها لوصف البيانات لكل منها مميزات وعيوبه من حيث درجة البساطة والوضوح ومن حيث إظهار معالم البيانات. وتبعاً لذلك يختلف استخدامها حسب نوع البيانات وأحجامها، عدد المتغيرات والغرض من وصف البيانات.



• أشهر أنواع الأشكال والرسومات البيانية مرتبة حسب البساطة هي الأشكال المصورة (Pictograms)، الدوائر المجزأة (Pie Chart)، الأعمدة (Bar Charts) وتشمل الأعمدة البسيطة (Simple Bar Charts)، الأعمدة المتلاصقة (Multiple Bar Charts)، والأعمدة المركبة (Component Bar Charts) وأخيرا الخطوط البيانية (Graphs).

• تعطي الأشكال المصورة انطباعا بصريا جيدا عن مجموعة البيانات المبحوثة ولكن يعيبها عدم الدقة وإخفاء التفاصيل.



● تستخدم الدوائر الجزأة عندما يكون الهدف مقارنة الأجزاء المختلفة بالنسبة للمجموع الكلي وعدد الأجزاء المقارنة قليل نسبيا .

● تستخدم الأعمدة عندما تكون أجزاء الظاهرة المقارنة كثيرة العدد نسبيا وعندما نرغب في توضيح قيم الأجزاء المقارنة .

● تستخدم الخطوط البيانية عندما يكون عدد المفردات كبير نسبياً أو عندما يكون الغرض توضيح العلاقة بين المتغيرات لفترات زمنية متعاقبة كما في حالة السلاسل الزمنية .

● يصعب استخدام الرسومات لتوضيح المقارنات البيانية عندما يزيد عدد المتغيرات عن حد معين وفي تلك الحالة تستخدم الجداول الإحصائية لعرض البيانات ومقاييس النزعة المركزية والتشتت لتلخيصها .



2. مقاييس النزعة المركزية (Measures of Central Tendency)

- تستخدم مقاييس النزعة المركزية الطرق الرياضية للتعبير عن قيم التمرکز في الظواهر وذلك لتحديد الاتجاهات العامة وتلخيص البيانات من خلال هذه القيم.
- المتوسط هو القيمة النموذجية الممثلة لمجموعة من البيانات وأكثر المتوسطات شيوعاً هي الوسط الحسابي (Arithmetic Mean)، الوسط الهندسي (Geometric Mean)، الوسط التوافقي (Harmonic Mean)، الوسيط (Median) والمنوال (Mode).



- يعرف الوسط الحسابي رياضياً بأنه يساوي مجموع قيم البيانات مقسوماً على عدد مفردات البيانات، أي:

$$\bar{X} = \frac{\sum_{i=1}^N x_i}{N}$$

حيث \bar{X} هو الوسط الحسابي، x_i قيم المفردات و N عدد المفردات.



• يرمز للوسط الهندسي بـ G ويساوي مجموع لوغاريتمات القيم مقسوما على عددها، أي:

$$G = \frac{\sum_{i=1}^N \ln x_i}{N}$$



- يرمز للوسط التوافقي بـ H ويساوي مقلوب الوسط الحسابي لمقلوبات القيم، أي:

$$H = \frac{1}{\sum_{i=1}^N \frac{1}{x_i}}$$

- يستخدم الوسط الهندسي والوسط التوافقي عندما تكون قيم البيانات في شكل نسب.



- الوسيط هو القيمة التي تتوسط المفردات البيانية ويحسب بترتيب البيانات حسب القيم ثم أخذ المفردة ذات الترتيب $\frac{N}{2}$ حيث N هي عدد المفردات.
- المنوال هو القيمة الأكثر شيوعاً أو القيمة ذات التكرار الأكبر في البيانات. وقد لا يكون للقيم منوال وقد يكون هنالك أكثر من منوال واحد للقيم.
- يتميز الوسيط والمنوال عن الوسط الحسابي بعدم تأثرهما بوجود القيم الشاذة (Outliers) في البيانات.
- للتمييز بين العينة والمجتمع يرمز للوسط الحسابي للعينة بـ \bar{X} وللمجتمع بـ μ (ميو).



3. مقاييس التشتت (Measures of Dispersion)

- تستخدم مقاييس التشتت للتعبير عن مدى التقاف أو بعثرة البيانات حول القيم المتوسطة وأكثر مقاييس التشتت شيوعا هي الانحراف المتوسط (Mean Deviation) والانحراف المعياري (Standard Deviation).
- الانحراف المتوسط هو متوسط انحرافات المفردات عن الوسط الحسابي، أي:

$$MD = \frac{\sum_{i=1}^N |x_i - \bar{x}|}{N}$$



- الانحراف المعياري هو الجذر التربيعي لمتوسط مربعات انحرافات المفردات عن الوسط الحسابي للمفردات، أي:

$$S = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N}}$$

حيث S ترمز للانحراف المعياري.



● للتمييز بين العينة والمجتمع يرمز للانحراف المعياري للعينة بـ S وللمجتمع بـ σ (سيجما) ويسمى مربعيهما (أي S^2 و σ^2) بالتباين (Variance).

● يستخدم الانحراف المعياري بكثرة في حساب وتقييم معاملات الخطأ للعينة وفي الإحصاء الاستدلالي لتقييم كفاءة المقدرات واختبار الفرضيات.



تطبيقات باستخدام برنامج SPSS

د. مصطفى بابكر



تطبيقات باستخدام برنامج SPSS

يعتبر البرنامج الإحصائي للعلوم الاجتماعية أو SPSS من أكثر البرامج الإحصائية شيوعاً ويستخدم في تحليل المسوحات الإحصائية ابتداءً من مرحلة تفرغ البيانات وحتى مرحلة اختبار الفرضيات وتلخيص النتائج.



تطبيقات على الإحصاء الوصفي

في إطار سعيها لتحسين الأوضاع المعيشية لشريحة العمالة الصناعية، قامت مصلحة التخطيط في بلدية كازاكا بإجراء المسح العيني التالي. تم اختيار عينة عشوائية بسيطة مكونة من 50 شخص من جملة العمالة الصناعية في البلدية والبالغ قدرها 1000 شخص.



قامت مصلحة الإحصاء في البلدية بتدريب فريق المسح وتصميم الاستبيان. بعد التشاور مع مصلحة التخطيط قرر أن تشمل البيانات المجموعة الآتي: الجنس، التعليم، العمر، عدد أفراد الأسرة التي يعولها الشخص بالإضافة إلى شخصه، المرتب الشهري، وعدد سنوات الخبرة. تم تنفيذ المسح بكفاءة وبلغت نسبة الاستجابة 100%.



تم إعداد نظام المصطلحات وأفرغت بيانات الاستبيان في الجدول أدناه.
ترغب بلدية كازاكا التعاون مع فريق "برنامج التحليل الإحصائي لأخذ
القرار" في تحليل المسح وإعداد تقرير عن مستوى الأجور وأعباء المعيشة
لقطاع العمالة الصناعية في البلدة توطئة لأخذ القرارات اللازمة بشأن
سياسات الأجور في القطاع.



No.	Sex	Education	Age	Household	Salary	Experience
الرقم	الجنس	التعليم	العمر	الأسرة	المرتب	سنوات الخبرة
1	0	1	35	3	150	8
2	1	2	31	2	198	5
3	1	2	51	5	240	14
4	0	1	57	6	192	15
5	0	2	28	1	190	3
6	1	3	64	6	305	20
7	1	2	30	1	192	3
8	1	2	35	2	197	5
9	0	2	26	1	190	3
10	1	2	39	4	225	9
11	1	2	22	1	184	1
12	1	2	24	1	188	2
13	0	2	41	3	184	6
14	1	2	58	4	260	17
15	1	3	45	4	500	10
16	0	1	52	5	160	8
17	0	1	33	1	130	3
18	0	2	46	5	185	12
19	1	2	49	4	218	10
20	1	2	21	1	182	1
21	1	2	43	4	258	12
22	1	3	35	2	300	7
23	1	2	33	1	208	7
24	1	2	64	6	278	21
25	1	2	55	5	268	18
26	0	1	33	2	148	6
27	1	2	55	4	260	16
28	0	1	42	1	140	4
29	0	1	30	1	127	2
30	0	30	44	3	240	10
31	1	2	23	1	185	1
32	1	2	35	3	302	7
33	0	1	40	1	189	2
34	1	2	49	3	350	13
35	1	2	44	4	240	12
36	0	1	54	5	191	12
37	1	2	32	1	204	6
38	0	1	41	1	137	4
39	1	2	26	1	192	3
40	1	3	57	6	330	14
41	1	2	43	4	236	11
42	1	2	26	1	188	2
43	1	2	47	3	250	15
44	1	2	32	3	194	4
45	0	3	34	1	200	3
46	1	2	51	4	248	14
47	0	1	28	1	140	4
48	0	1	28	2	142	5
49	1	2	64	5	500	17
50	1	3	31	3	240	2

المصطلحات:

الجنس: 0 أنثى 1 ذكر
 التعليم: 1 ابتدائي 2 ثانوي 3 جامعي



هذا وقد تم إدخال البيانات في الحاسب الآلي وتخزينها في الملف
"KAZA.XSL". المطلوب إنجاز المهمة الأولى في إعداد التحليل
والتقرير وهي الوصف الإحصائي للبيانات.