المعاينة واختيار العينة

فصل المعاينة واختيار العينة

- تعرف **العينة** بأنها "جزء من المجتمع يتم اختياره لتمثيل المجتمع" أما **المعاينة** فتعرف بأنها "عملية اختيار جزء من المجتمع الإحصائي للاستدلال على خواص المجتمع بأكمله عن طريق تعميم نتائج العينة". وتتنوع طرق اختيار العينة حسب خصائص المجتمع المدروس ومزايا وعيوب كل طريقة.

- ويوجد نوعان من العينات: العينات الاحتمالية حيث تسحب وحداتها بطرق عشوائية وتخضع بالتالي لقوانين الاحتمالات، ثم العينات أللاحتمالية حيث مبدأ اختيار أفرادها لا يخضع لقوانين موضوعية.

العينات الاحتمالية: Probability Sampling

- تعتمد على المساواة بين احتمالات اختيار أي فرد من أفراد المجتمع الأصلي.
- العشوائية هي طريقة الباحث في تحقيق التكافؤ بين الأفراد.
- تعطي الباحث عينة ممثلة لمجتمعها الأصلي بتكلفة أقل مع تجنب تحيز الباحث في الاختيار، وما ينتج عنها من مشكلات تشكك في صحة النتائج.
- تشترك العينات الاحتمالية في تحديد مجتمع الدراسة، وإعداد قائمة بعناصره، ثم اختيار عينة بحجم يكفي لتمثيل خصائص المجتمع.

وتتعدد طرق اختيار العينات الاحتمالية ومن أهمها:

أولا: السحب العشوائي البسيط

- يعتبر السحب العشوائي البسيط من التصاميم الأولى لكل الاحتمالات في اختيار العينة المدروسة. وتعتمد هذه الطريقة على خطوة وحيدة في الانتقاء تعطي لكل عينة ممكنة ذات حجم معين n نفس الاحتمال في الانتقاء.

- نتيجة لذلك فان كل وحدة من العينة لها نفس احتمال الإدراج $= \frac{n}{N}$ (Inclusion Probability) حيث N هو عدد الوحدات في المجتمع.
- وقد يتم السحب بإحلال (بإرجاع) (with replacement) أو بدون إحلال (بدون إرجاع) (without replacement)، حيث تتاح للوحدة فرصة الاختيار أكثر من مرة في حالة السحب مع إحلال. وتعتبر الطريقتان تقريبا متطابقتان عمليا إذا كان حجم العينة صغيرا أمام حجم المجتمع n<<<N (الاحتمال $\pi = \frac{\eta}{N} \leftarrow \frac{1}{N}$).
- عادة ما ينتج السحب بدون إحلال نتائج عملية أكثر دقة ومقنعة.

في كل ما يلي، نعتبر السحب بدون إحلال ما لم يشار إلى غير ذلك.

- مثال 1: المجتمع N=5، العينة n=3 الطريقة هي السحب العشوائي البسيط بدون إحلال.
- بالنسبة للمجموعة N= {1,2,3,4,5} هناك عشر سحوبات ممكنة للعينة من الحجم n=3 : {1,3,5} {1,3,4} {1,2,5} {1,2,4} {1,2,3} {3,4,5} {2,4,5} {2,3,5} {2,3,4} {1,4,5}.
- كل هذه العينات لها نفس احتمال السحب وكل وحدة (فرد) لها احتمال 6 من 10 ليتم سحبها أي أن احتمال إدراجها: $\pi = \frac{n}{N} = \frac{6}{10} = \frac{3}{5}$

- مثال 2: نعتبر مدرسة من N =1530 طالب. مطلوب اختيار عدد من أربعة أرقام (عدد أرقام 1530 4digits) بعد ذلك يتم اختيار الأعداد التسعين الأولى في السحب والتي لا تتعدى 1530.

- عمليا يتم السحب العشوائي في Excel عن طريق

Tools → Data Analysis→ Random Number→Generation

السحب العشوائي في برنامج SPSS: راجع التطبيق العملي.

تتسم طريقة السحب العشوائي البسيط بالمزايا التالية:

- فهي أسهل وأبسط تقنية للسحب.
- لا تتطلب معلومات إضافية لسحب العينة. كل ما في الأمر هو توفر قائمة كاملة عن المجتمع ومعلومات الاتصال.
- لا تحتاج تقنية متطورة حيث تعتمد على صيغ رياضية وقواعد سهلة الاستعمال (تحديد حجم العينة، تقدير السكان، تقدير التغاير).

ومن عيوب هذه الطريقة:

- لا تستعمل أي معلومات إضافية عن إطار المسح ولو كانت متوفرة، مما قد يؤثر على النتائج.
- بما أن كل العينات لها نفس احتمال الإدراج، فان سحب عينة غير جيدة في تمثيل المجتمع ستكون له عواقب على مستوى النتائج.

- مثال للسحب العشوائي البسيط ,N=36, n=12

o	o	o	**o**	o	**o**
o	o	**o**	o	o	o
o	**o**	o	**o**	**o**	o
o	**o**	o	o	o	o
o	o	**o**	o	**o**	o
o	o	o	**o**	o	o

ثانيا: السحب المنتظم

- في العينة المنتظمة، يتم انتقاء وحدات من المجتمع على فترات منتظمة. وتستخدم هذه الطريقة عادة إذا تعذر وجود قائمة الأفراد كاملة أو إذا كانت القائمة نفسها عشوائية من حيث الترتيب. آنذاك يصبح السحب المنتظم أبسط إجراء من السحب العشوائي البسيط.

نتبع الخطوات التالية:

- باستعمال قائمة لمجتمع ذو حجم N مضاعف صحيح لحجم العينة n،
- يتم اختيار الحالة الأولى من العينة بطريقة عشوائية ثم يمضى الباحث في اختيار بقية الحالات على أبعاد رقمية منتظمة أو متساوية بين الحالات، بحيث تكون المسافة بين أي وحدتين متتاليتين ثابتة في جميع الحالات.

1. تحديد المجتمع الأصلي (N)
2. تحديد حجم العينة المرغوب فيه (n)
3. تحديد المسافة بين أفراد العينة من خلال k= N/n
4. اختر عشوائيا عددا ينحصر بين (1 والقيمة k)
5. أضف إلى العدد المختار قيمة k بشكل منتظم، حتى تحصل على العينة التي تريدها .

- العينات المسحوبة هي من المجالات : k ;2k ;3k ;..; (n-1)k

حيث رقم الوحدات المسحوبة بالتوالي:

r+k ; r+2k; r +3k;…; r+(n-1)k

- مثال : إذا كان المجتمع الأصلي يتكون من (100) فرد ، ونريد الحصول على عينة منه عددها (20) فردا، كيف يمكنك تحقيق ذلك باستخدام العينة العشوائية المنتظمة؟

$N=100$ و $n=20$ إذن $k=100/20=5$

- اختر الآن عددا ينحصر بين (1و5) وليكن (4) وذلك بشكل عشوائي، ونجعل العدد (4) نقطة الانطلاق مثلا، نضيف له (5) بشكل ثابت منتظم إلى أن يصبح حجم العينة المختارة (20) فردا .
- علما بأن: الحالة الأولى تم اختيارها عشوائيا وهى (4)، فان:
- الحالة الثانية = الحالة الأولى + طول الفئة K = 4+5 = 9
- الحالة الثالثة = الحالة الثانية + طول الفئة = 9+5 =14
- الحالة الرابعة = الحالة الثالثة + طول الفئة = 14+5= 19

- وهكذا يتضح أنه لتحديد أي حالة يجب أن نعرف الحالة التي تسبقها ثم نضيف إليها طول الفئة.
- على غرار السحب البسيط، كل الوحدات لها نفس احتمال الإدراج $\pi = n/N$ غير أن العينات n ليس لها نفس الاحتمال ليتم اختيارها:
- في السحب المنتظم يتم اختيار العينات حيث الأفراد/الوحدات مفصولة بسعة المجال k.
- مثال توضيحي: $N = 36; n = 12 : k = N/n = 3$ مثلا r = 1

o	o	o	o	o	o
o	o	o	o	o	o
o	o	o	o	o	o
o	o	o	o	o	o
o	o	o	o	o	o
o	o	o	o	o	o

مميزات العينات العشوائية المنتظمة:

- تعد من أسهل العينات العشوائية في التطبيق.
- لا تحتاج إلى عملية إعداد مسبق لمفردات الدراسة خاصة إذا كانت مجموعات داخل مجتمع الدراسة.
- لا تحتاج إلى الرجوع في كل مرة يتم فيها سحب المفردات إلى مرجع أو دليل فيكتفي بالمفردة الأولى أما باقي المفردات فتحدد تلقائيا عن طريق صيغة رياضية سهلة ومبسطة.

عيوب العينات العشوائية المنتظمة:

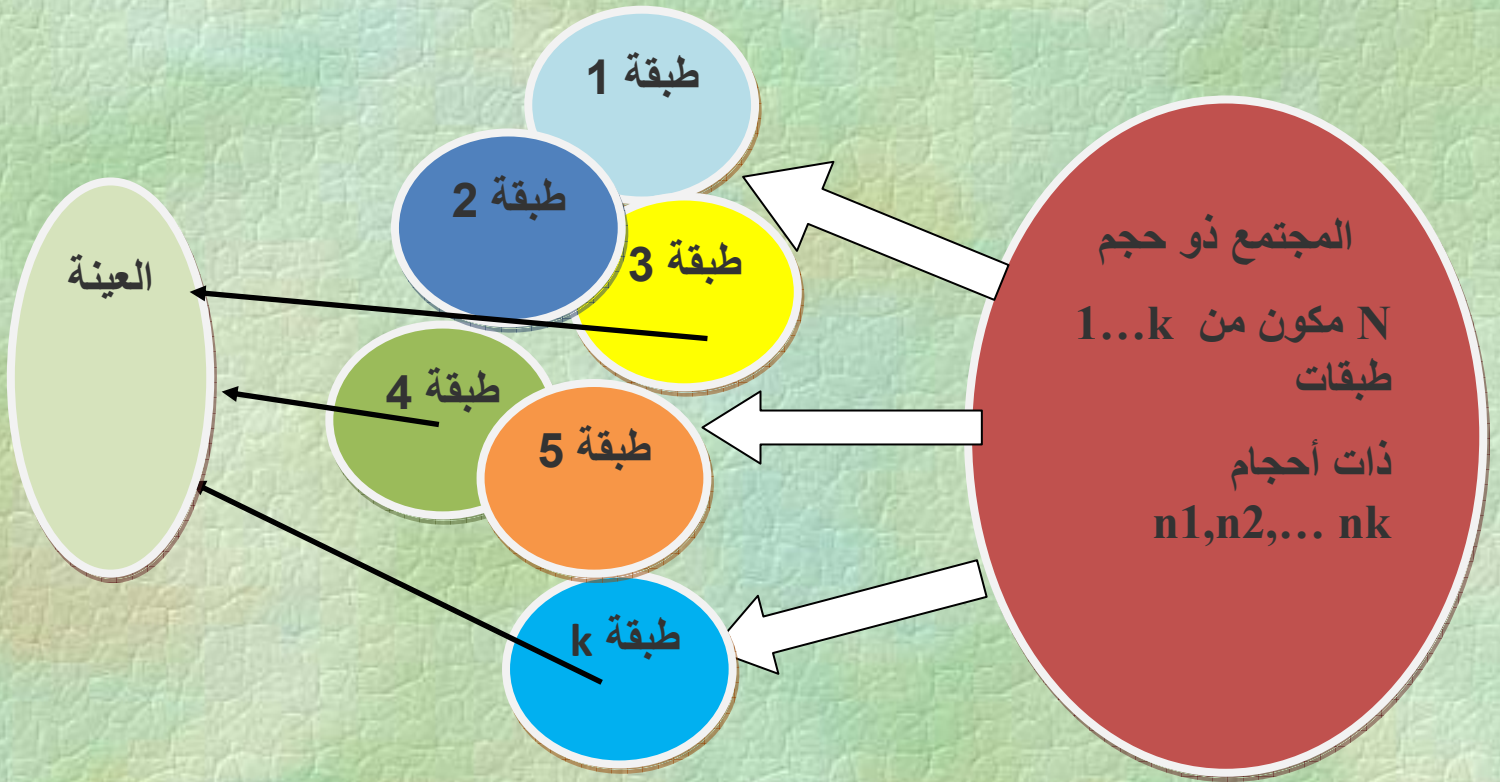
- تستلزم توفر قائمة حديثة تشمل كافة أسماء مفردات المجتمع الأصلي.
- قد تكون العينة المختارة غير متجانسة ، وذلك حينما تختار مفردات على أبعاد منتظمة يصادف أن يكونوا من طبقة معينة أو من ذوى خصائص وصفات مميزة وغير متشابهة مع بقية المفردات.
- يشترط في المجتمع الأصلي أن يكون الأفراد في تسلسل منسق وتدرج من حيث التنوع.

- لا تحدث احتمالية فرصة التمثيل لمفردات مجتمع الدراسة إلا مرة واحدة وهى عند اختيار المفردة الأولى.
- في حالة كون طول الفئة كبيرا وهناك مجموعات داخل مجتمع الدراسة عددها أقل من طول الفئة فإن احتمال تمثيل هذه المجموعة في العينة يكون محدودا .

ثالثا: السحب العنقودي

- هو طريقة للسحب العشوائي لمجموعة كاملة (عنقود) من مجموع وحدات المجتمع. عادة، هذه الطريقة أقل كفاءة من السحب العشوائي البسيط، غير أن استعمالها مُبرر أحيانا نظرا لقلة تكلفتها أو عندما يتعذر عمليا مسح العينات واحدة بواحدة كما في السحب البسيط.
- مثال: سحب عينات أسر بدل سحب عينات أفراد أسهل وأقل تكلفة، إذ يقلل من كلفة التنقل مثلا لشمل عينات الأفراد.

- السحب العنقودي يتم عبر مرحلتين: يتم تجميع وحدات المجتمع في شكل مجموعات (قد تكون مجمعة طبيعيا كما في مثال الأسرة أو المدرسة. .) في المرحلة الأولى. في الثانية، يتم سحب عينة من هذه المجموعات ثم استقراء كل وحدات هذه العينات.
- سحب العناقيد تتم بالطرق السابقة مثل السحب العشوائي البسيط أو السحب المنتظم أو السحب بالتناسب مع حجم العينة وغالبا ما يفضل هذا الأخير نظرا لاختلاف الأحجام بالنسبة للمجموعات الطبيعية (الأسرة، الجامعة).


طبقة 1
طبقة 2
طبقة 3
طبقة 4
طبقة 5
طبقة k
العينة
المجتمع ذو حجم
N مكون من 1...k
طبقات
ذات أحجام
n1,n2,... nk

- يعتبر هذا النوع من السحب أقل كفاءة من السحب البسيط خصوصا إذا كانت المجموعات (العناقيد) غير متجانسة.
- فسحب مجموعة من قاعدة مجموعات غير متجانسة سوف لن يمثل المجتمع. بينما سحب وحدات مختلفة من كل المجموعات عن طريق السحب البسيط سيكون له كفاءة أكبر في تمثيل المجتمع.
- هناك عناصر أخرى تلعب دورا مهما في تحديد الكفاءة الإحصائية وهي عدد الوحدات في كل عنقود وعدد العناقيد في المجتمع.

- قد تعترض السحب العنقودي عراقيل لوجستيكية خصوصا إذا كانت قاعدة السحب معدة بطريقة عشوائية. آنذاك سيكون من الصعوبة تحديد أي الوحدات تنتمي إلى عنقود محدد ما لم توضع طريقة للتعرف على هذه الوحدات سالفا .

من مزايا السحب العنقودي:

- تقليل تكلفة المسح الميداني عن طريق تقليل تكلفة التنقل بين الوحدات وربح الوقت أيضا .
- سهولة سحب العينة مقارنة مع الطرق السابقة الذكر خصوصا إذا كانت قاعدة المسح مجمعة على شكل مجموعات طبيعية (الأسر، المدارس، الخ) .
- يتيح تقدير الإحصائيات للعناقيد نفسها . مثال: حساب عدد الطلبة في كل فصل في المتوسط إذا اعتبرنا الفصول هي العناقيد .

من سلبيات هذا السحب:

- ارتباط الكفاءة الإحصائية بمدى تجانس العناقيد أو الوحدات.
- تقدير التغاير قد يكون معقدا على غرار السحب البسيط خصوصا إذا سحبت العناقيد بدون إحلال.

رابعا: سحب العينات باحتمال متناسب مع حجمها

Probability Proportional-to- Size Sampling (PPS)

- سحب العينات الاحتمالية–النسبي إلى الحجم (PPS) هو أسلوب يستخدم البيانات المساعدة والنتائج هي أن احتمالات الإدراج ليست متساوية. إذا كانت الوحدات تختلف في الأحجام التي هي معروفة، مثل هذه المعلومات يمكن استخدامها أثناء سحب العينات لزيادة الكفاءة الإحصائية.

- ومن الأمثلة الجيدة على الأرض: المسوح الزراعية وغالبا ما تستخدم هذه الطريقة، حيث قياس الحجم هو حجم المزرعة بالهكتار. حجم مزرعة يمكن أن بنمو أو ينكمش إذا كان المزارع يشتري أو يبيع الأرض، ولكن بالنسبة للجزء الأكبر، حجم المزرعة هو ثابت من سنة إلى أخرى. وبالإضافة إلى ذلك، غالبا ما ترتبط الأسئلة النموذجية للمسوح الزراعية، بأسئلة مثل الدخل، وإنتاج المحاصيل، وحيازات الثروة الحيوانية والنفقات مع حيازات الأراضي. تدابير أخرى لاستقصاءات الأعمال التجارية وتشمل: عدد العاملين، والمبيعات السنوية وعدد من المواقع، على الرغم من أن هذه المتغيرات هي أكثر عرضة للتغيير من سنة إلى أخرى.

- في حين أن حجم الوحدة يحدد احتمال إدراجها، وباستخدام المزارع كمثال فان: مزرعة تبلغ مساحتها 200 هكتار لديها ضعف احتمال سحب مزرعة بـ 100 هكتار.

- ولتوضيح ذلك، نفترض أن هناك ستة مزارع وأن الباحث يهتم في تقدير مجموع نفقات هذه الفئة عن طريق أخذ عينة مزرعة واحدة. (يتم استخدام عينة من حجم واحد لغرض التوضيح؛ في الممارسة، نادرا ما تختار وحدة واحدة فقط) . لنفترض أن هناك مقياس ثابت لحجم كل مزرعة (حجم المزرعة بالهكتار) . و لتوضيح مكاسب الكفاءة افترض أن كل مصاريف المزرعة معروفة.

نعتبر القائمة التالية من المزارع:

الوحدة	معلومات مساعدة: حجم المزرعة بالهكتارات	المتغير مسح الاهتمام: المصروفات
1	50	26000
2	1000	470000
3	125	63800
4	300	145000
5	500	230000
6	25	12500
إجمالي	2000	947300

- مقارنة الكفاءة الإحصائية بين طريقة السحب البسيط وطريقة تناسب الاحتمال والحجم من خلال الجدول السابق:
- يمكن اختيار عينة عشوائية بسيطة، حيث تحتوي كل عينة على وحدة واحدة وكل وحدة لديها احتمال إدراج $\frac{1}{6}$. العينة من حجم واحد، وتقدر مجموع النفقات للسكان عن طريق ضرب نفقات الوحدة في وزنها . هذا الوزن هو متوسط عدد الوحدات في عدد السكان = 6 وهو عكس احتمال الإدراج.
- N=6, n=1, method SRS

العينة المسحوبة	الحجم	احتمال الإدراج	الوزن	المصروفات	المصروفات المقدرة
1	50	1/6	6	26,000	156,000
2	1,000	1/6	6	470,000	2,820,000
3	125	1/6	6	63,800	382,800
4	300	1/6	6	145,000	870,000
5	500	1/6	6	230,000	1,380,000
6	25	1/6	6	12,500	75,000

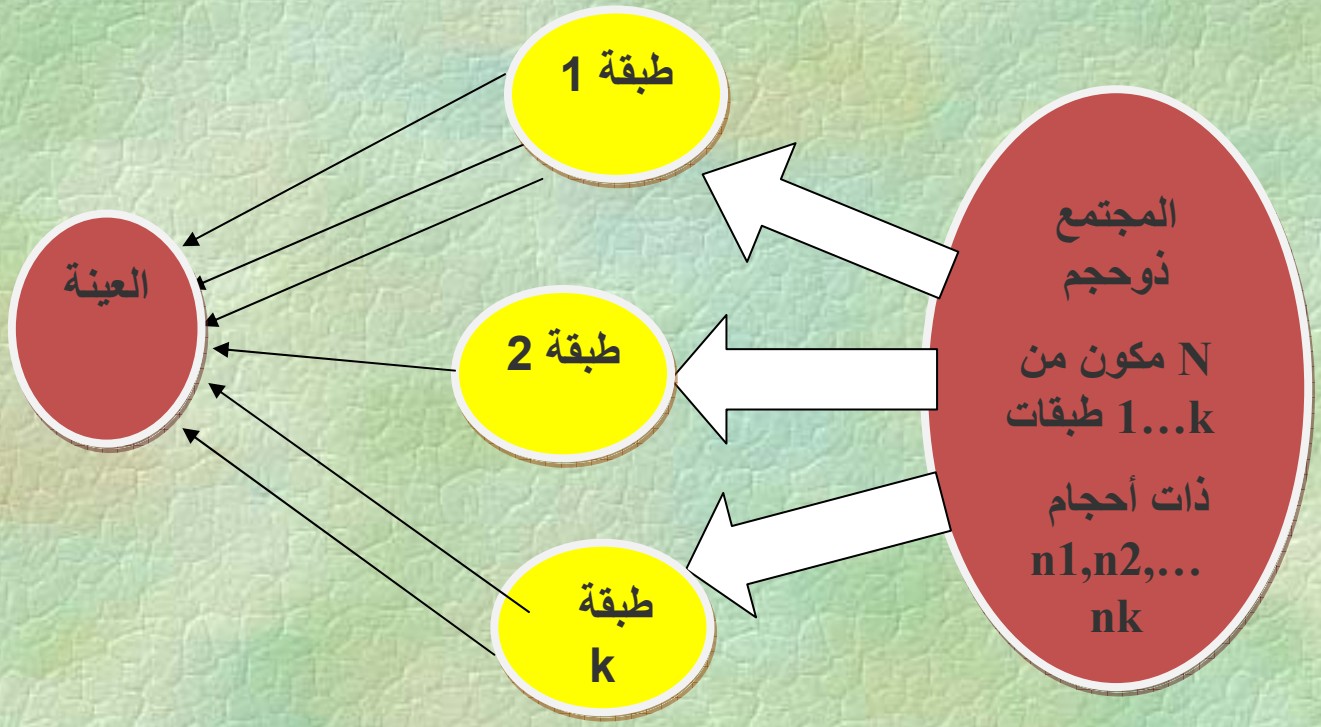
- لاحظ أن تقلب سحب العينات كبير في تقديرات SRS، تتراوح بين $75000 و 2.8 مليون دولار.
- بالنسبة للطريقة الثانية (PPS)، تقلب سحب العينات هو أقل من ذلك بكثير. تقديرات التغاير منخفضة من $920000 إلى 1.04 مليون دولار – أفضل بكثير من طريقة السحب البسيط. انظر الجدول الموالي: N=6, n=1, method PPS .

العينة المسحوبة	الحجم	احتمال الادراج	الـوزن	المصروفات	المصروفات المقدرة
1	50	50/2,000	2,000/50	26,000	1,040,000
2	1,000	1,000/2,000	2,000/1000	470,000	940,000
3	125	125/2,000	2,000/125	63,800	1,020,800
4	300	300/2,000	2,000/300	145,000	966,667
5	500	500/2,000	2,000/500	230,000	920,000
6	25	25/2,000	2,000/25	12,500	1,000,000

- يتم احتساب احتمال الإدراج ويبلغ حجم المزرعة مقسوما على مجموع حجم جميع المزارع. أما الوزن فهو عكس الاحتمال.
- إذن الميزة الرئيسية لأخذ العينات بهذه الطريقة هو تحسين الكفاءة الإحصائية لسحب العينات باستخدام معلومات مساعدة. هذا يمكن أن يؤدي إلى انخفاض كبير في التباين مقارنة مع السحب العشوائي البسيط أو العينات الطبقية.

خامسا: سحب العينات العشوائية الطبقية (Stratified sample):

- يستعمل هذا النوع من العينات في الحالات التي يكون فيها مجتمع الدراسة مكونا من طبقات مختلفة، وهنا يضع الباحث شروطاً معينة لاختيار أفراد العينة بحيث تمثل العينة جميع أفراد المجتمع المدروس، وبنفس نسبه وجودها في المجتمع. أي أن الباحث يختار لكل طبقة وبطريقة عشوائية عددا من المفردات، يتناسب مع حجمها الحقيقي في المجتمع الأصلي.


طبقة 1
العينة
طبقة 2
طبقة
k
المجتمع
ذوحجم
N مكون من
1…k طبقات
ذات أحجام
n1,n2,…
nk

- يمكن تقسيم السكان إلى مستويات مختلفة حسب أي من المتغيرات التي تتوفر لجميع الوحدات في الإطار المحدد للمسح. على سبيل المثال، قد تكون هذه المعلومات مجرد عنوان الوحدة، مما يتيح التقسيم الطبقي حسب المحافظة، أو قد تكون هناك بيانات الدخل مثلا لكل وحدات الإطار، مما يسمح التقسيم الطبقي حسب فئات الدخل. ومن المتغيرات الشائعة استخداما لتقسيم المجتمع إلى طبقات ما يلي: العمر، والجنس، والجغرافيا (على سبيل المثال، المحافظات)، والدخل أو الإيرادات، حجم الأسرة، وحجم الأعمال التجارية، ونوع الأعمال، وعدد الموظفين، الخ.

- هناك ثلاثة أسباب رئيسية للعينات الطبقية: الأول هو لجعل إستراتيجية سحب العينات أكثر كفاءة من الطرق السابقة: مثلا السحب البسيط أو المنتظم. والثاني هو لضمان أحجام عينة كافية للمجالات ذات فائدة في التحليل. والثالث هو للحماية من سحب عينة ضعيفة التمثيل.

- باعتبار حجم عينة معين ومقدر، قد يؤدي التقسيم الطبقي إلى انخفاض الخطأ في اختيار العينات. بالنسبة للعينات الطبقية، يتم سحب عينات داخل كل طبقة، بينما لأخذ العينات العنقودية، يتم سحب عينات من مجموعات والجميع في هذه العينات يشملها الاستطلاع. التقسيم الطبقي بشكل عام يزيد من دقة التقدير مقارنة مع السحب البسيط، في حين أن العينات العنقودية تقلل من دقة التقدير والكفاءة الإحصائية مقارنة مع السحب البسيط.

- من أجل تحسين الكفاءة الإحصائية، يجب أن يكون هناك تجانس قوي داخل الطبقة الواحدة، وضعيف فيما بين الطبقات نفسها بالنظر إلى المتغير المشترك والمعتمد في تقسيم الطبقات: (كوشران Cochrane,1977) .

مزايا العينة الطبقية هي كالتالي :

- يمكن زيادة دقة التقديرات مما يؤدي إلى نتائج أكثر كفاءة.
- يمكن أن نضمن أن مجموعات فرعية هامة، ممثلة تمثيلا جيدا في العينة، مما يرفع الكفاءة إحصائيا .
- عينة أصغر يمكن أن توفر مبلغا كبيرا على المسح، لاسيما خلال جمع البيانات.
- يمكن أن تكون عمليا أو إداريا مريحة عندما تكون الطبقات متوفرة طبيعيا .

- يمكن أن تحمي ضد اختيار عينة 'سيئة'.
- لأنها تتيح أطر أخذ العينات وإجراءات مختلفة ليتم تطبيقها على طبقات مختلفة (على سبيل المثال، السحب البسيط في طبقة، والمنتظم في أخرى).

مساوئ أخذ العينات الطبقية هي:

- يتطلب أن يحتوي الإطار لسحب العينات على معلومات مساعدة ذات جودة عالية لجميع وحدات الإطار، وليس فقط لتلك الموجودة في العينة، التي يمكن استخدامها للطبقية.
- خلق الإطار هو أكثر كلفة وتعقيدا من الطريقتين البسيطة والمنتظمة، نظرا لأن هذا الإطار يتطلب الحصول على معلومات مساعدة جيدة.
- يمكن أن يؤدي إلى كفاءة إحصائيا أقل من السحب البسيط للمتغيرات التي لا ترتبط بالمتغيرات الطبقية.
- تقدير المعالم قد يكون نوعا ما أكثر تعقيدا من الطريقتين البسيطة والمنتظمة.

العينات الاحتمالية (Non Probability samples)

- وفيها يتم اختيار العينة بشكل غير عشوائي، حيث يتدخل فيها حكم الباحث وذلك باستثناء بعض عناصر الدراسة من الظهور في العينة لأسباب معينة. عدم توافر المعلومات المطلوبة، أو استحالة الوصول إلى هذه العناصر، أو كبر حجم وحدات مجتمع الدراسة، وهي عدة أنواع:

عينة الصدفة (Accidental Sample):

- أو تسمى العينة الملائمة (Convenience sample) وفي هذا النوع من العينات يعطى لعنصر مجتمع الدراسة الأصلي حرية الاختيار في المشاركة في الدراسة، بحيث لا يكون هناك تحديد مسبق لمن تشملهم العينة، بل يتم اختيار أفراد العينة من بين أول مجموعة يقابلهم الباحث، بحيث يوافق هؤلاء على المشاركة. يتميز هذا النوع من العينات بالسهولة في اختيار عينة الدراسة وانخفاض التكلفة والوقت والجهد المبذول وبسرعة الوصول إلى أفراد الدراسة والحصول على نتائج. ويؤخذ على هذا النوع من العينات أنه لا يمكن أن يمثل المجتمع الأصلي بدقة ومن هنا يصعب تعميم نتائج البحث على المجتمع كله.

العينة الفرضية أو العقدية أو العمدية (Purposive sample):

- وتسمى أيضا العينة الهدفية لأن الباحث يختار هذا النوع من العينات لتحقيق غرضه، بحيث يقدر حاجته من المعلومات، ويقوم باختيار عينة الدراسة اختيارا حرا ، أي أن هذا النوع من العينات لا يكون ممثلاً لأحد .

العينة الحصصية (Quota Sample):

- تشبه العينة الحصصية العينة الطبقية العشوائية من حيث تقسيم مجتمع الدراسة الأصلي إلى فئات أو شرائح ضمن معيار معين، لكنها تختلف عنها في أن الباحث في العينة العشوائية لا يختار الأفراد كما يريد، بينما في هذا النوع يقوم الباحث بهذا الاختيار بنفسه، دون أية شروط.