تصميم الأوزان، أخطاء المسح وتقدير المعالم

تصميم الأوزان، أخطاء المسح وتقدير المعالم

أولا: المقدمة

- الدافع وراء المسح عادة هو دراسة خصائص مجتمع ما . و تقدير المعالم هو الوسيلة التي يمكننا بها الحصول على قيم للمجتمع المدروس حيث يتم استخلاص استنتاجات حول السكان بالاعتماد فقط على المعلومات التي تم جمعها من العينة الممثلة للمجتمع .

- المبدأ الكامن وراء التقدير في المسح الاحتمالي هو أن كل وحدة من العينة لا تمثل نفسها فقط، ولكن أيضا عدة وحدات من مجتمع المسح. ومن الشائع اعتبار متوسط عدد الوحدات -التي تمثلها هذه الوحدة- في عدد سكان المجتمع يمثل وزن هذه الوحدة. تحديد هذا الوزن هو جزء مهم من عملية التقدير لأن تصميم الأوزان يمكن من تقدير المعالم (التغاير، المتوسط...) بدقة أعلى. معظم الدراسات الاستقصائية تنتج مجموعة من الطرق لتصميم الأوزان، والسبب يتجلى في إجراء تعديلات تأخذ في الحسبان عدم الإجابة والاستفادة من البيانات المساعدة.

حال الانتهاء من تقدير الأوزان النهائية، يتم تطبيقها على البيانات المجمعة من المسح لحساب التقديرات. وعادة ما يتم تقدير مقاييس موجزة من السكان مثل المجاميع والمتوسطات والنسب لمجموعة واسعة من الخصائص التي تم جمعها من وحدات العينة. هذه الخصائص، غالبا ما يشار إليها بالمتغيرات في النظرية الإحصائية، قد تكون نوعية – على سبيل المثال، الجنس أو الحالة الاجتماعية – أو كمية – على سبيل المثال، العمر أو الدخل. اعتمادا على نوع البيانات، هناك صيغ مختلفة مناسبة لتقدير معالم المجتمع.

- تقدير حجم الخطأ في اختيار العينات يعد جزءا هاما من التقدير. وهذا يوفر مقياسا لنوعية تقديرات المسح لعينة محددة. لا يمكن إلا أن يقدر هامش الخطأ إلا في حالات العينات الاحتمالية.
- يهدف هذا الفصل إلى توضيح كيفية حساب الأوزان (الترجيح)، وحساب تقديرات الخطأ في اختيار العينات الاحتمالية وتحديد حجم العينة و تقدير بعض من معالم المجتمع.

ثانياً: الترجيح (صياغة الأوزان)

- الخطوة الأولى في تقدير معالم المجتمع هي تحديد الوزن لكل العينات أو لكل وحدات العينات المستجوبة. يمكن اعتبار الوزن متوسط عدد الوحدات – التي تمثلها هذه الوحدة – في عدد سكان المجتمع يمثل وزن هذه الوحدة. وزن وحدة واحدة من العينة هو عكس احتمال إدراجها، π.

لاحظ أن في التصميم متعدد المراحل مثلا، الاحتمال النهائي لانتقاء وحدة في العينة هو احتمال سحبها في المرحلة الأولى مضروب في احتمال سحبها في المرحلة الثانية مضروب في احتمال سحبها في المرحلة الثالثة. . .الخ. وهكذا، فان وزن تصميم وحدة العينة هو:

$$W_d = \frac{1}{\pi_1 . \pi_2 \pi_3 .... \pi_n}$$

حيث n يمثل عدد مراحل السحب.

يتم حفظ بيانات عينة المسح عموما في ملف مع سجل واحد (one record unit) لكل وحدة من العينة. تذكر أنه مع أخذ العينات الاحتمالية، كل وحدة لديها احتمال معروف، π. وإذا كان هذا الاحتمال، على سبيل المثال، واحد في الخمسين، فان كل وحدة مختارة تمثل في المتوسط 50 وحدة من سكان المسح. و بالتالي فالوزن هو 50. إذا كان الوزن هو عدد صحيح، طريقة واحدة لإنتاج تقديرات للسكان سيكون لتكرار كل سجل بحيث يكون هناك 50 نسخة من كل واحد وبعد ذلك حساب أي إحصائيات عن المجتمع ذات فائدة (مثل المتوسطات، المجاميع والنسب، الخ) من هذا الملف أخذا بالاعتبار الأوزان.

لكن الأمر يصبح أكثر صعوبة عندما الأوزان ليست قيمة عددية صحيحة. على سبيل المثال، إذا تم اختيار وحدتين من أصل خمسة باستخدام عينة عشوائية بسيطة، فالوزن المرجح هو 2.5 = 5/2 .

في ما يلي نبدأ مناقشة الترجيح مع العينات متساوية احتمال الإدراج الذي هو أبسط حالات الترجيح.

ترجيح العينات متساوية الاحتمال:

- تسمى العينات متساوية الاحتمال ذاتية الترجيح عند تصميم الأوزان التي هي نفسها لجميع الوحدات في العينة. يحدث هذا عندما تكون كل وحدة لها نفس احتمال إدراج . بالنسبة للترجيح الذاتي ، إذا لم يتم إجراء أي تعديلات لاحقة على تصميم الأوزان، على سبيل المثال ، لعدم الإجابة أو عند توفر البيانات الإضافية، يمكن تجاهل الأوزان في إنتاج بعض الإحصائيات، مثل النسب والمعدلات.

الآن أي العينات العشوائية هي ذاتية الترجيح ؟

- من الواضح أن العينات العشوائية البسيطة (SRS) والعينات المنتظمة هي ذاتية الترجيح لأن كل الوحدات لديها فرصة متساوية ليتم تضمينها في العينة. في حالة سحب الطبقات، يتم الحصول على عينة ذاتية الترجيح، على سبيل المثال، عندما يتم السحب العشوائي البسيط SRS في كل طبقة وعندما يكون حجم العينة في كل طبقة يتناسب مع حجم السكان. وهكذا، فإن سحب العينات هو نفسه في كل طبقة وجميع الوحدات في المجتمع لديها نفس احتمال الإدراج.

مثال:

- لنفترض أن عدد أفراد المجتمع N=1000ينقسم الأشخاص فيه إلى طبقتين في إطار المسح. وتتألف الطبقة الأولى من N1 = 400 رجل والثانية من N2= 600 امرأة. نختار عينة من حجم n= 250. بحيث يتم سحب عينة من الطبقة الأولى بحجم n1=100 ومن الطبقة الثانية n2 = 150 .

الطبقة	حجم المجتمع	حجم العينة
رجال	N1=400	n1=100
نساء	N2=600	n2=150
الاجمالي	N=1000	n=250

$$\pi_1 = n_1 / N_1 = \pi_2 = n_2 / N_2 = \pi = n / N = 1/4$$

ولذلك، فالجميع لديه نفس احتمال اختياره ونفس الوزن = 4 .

لتصميم متعدد المراحل، يتحقق الترجيح الذاتي من خلال تحديد مجموعات مع احتمال يتناسب والحجم في جميع المراحل باستثناء المرحلة النهائية. في المرحلة النهائية، يتم اختيار عدد محدد من الوحدات ضمن كل طبقة. وكثيرا ما يستخدم السحب المتناسب مع الحجم (PPS) في تصاميم متعددة المراحل لأنه يمكن أن يؤدي إلى عينة ذاتية الترجيح.

- **الترجيح لعدم المساواة في الاحتمال**
- على الرغم من بساطة التصاميم ذاتية الترجيح، فإنه ليس من الممكن دائما أو مرغوبا فيه اختيار عينة ذاتية الترجيح. على سبيل المثال، في حالة وجود تصميم طبقي لدراسة استقصائية وطنية لمناطق صغيرة، فالتخصيص النسبي قد يؤدي إلى عدم كفاية حجم العينة. كما أنه بوجود مناطق أوسع، فإنه قد يؤدي إلى عينات كبيرة دون داع.
- يوضح المثال التالي تصميم الأوزان لتصميم عينة طبقية حيث حجم العينة في كل طبقة ليس متناسبا مع حجم السكان في الطبقة.

- مثال : العينات الطبقية مع توزيع غير متناسب
- لإجراء مسح على وسائل النقل العام، ينقسم سكان المجتمع N= 1100 إلى طبقتين جغرافيتين: الريف والحضر. بما أنه من المتوقع أن الأشخاص الذين يعيشون في المناطق الحضرية قد يختلفون جدا فيما يتعلق بالمعلومات التي تم جمعها من المسح عن قاطني المناطق الريفية، تم استخدام تصميم عينة طبقية. الطبقة الحضرية هي من الحجم N1=1000 في حين أن الطبقة الريفية من حجم N2=100. يتم اختيار عينة من n = 250 شخصا: n1=200 في الطبقة الحضرية و n2=50 في الطبقة الريفية. ما هو تصميم الأوزان لعينات الأشخاص في كلا المنطقتين؟

- نفترض أولا السحب العشوائي البسيط مع عدم تناسب الأحجام:
- احتمالات إدراج الأشخاص في كل طبقة هي:
- الطبقة 1، الحضر: $\pi_U = \frac{n_U}{N_U} = 200 / 1000 = 1/5 = 20\%$
- الطبقة 2، الريف: $\pi_R = \frac{n_R}{N_R} = 50 / 100 = 1/2 = 50\%$
- كل شخص في الطبقة الحضرية لديه وزن $W_U = \frac{1}{\pi_U} = \frac{1}{0.20} = 5$
- و كل شخص في الطبقة الريفية لديه وزن $W_U = \frac{1}{\pi_R} = \frac{1}{0.50} = 2$

- نفترض ثانيا السحب العشوائي متعدد المراحل (هنا مرحلتين) مع عدم تناسب الأحجام:

- لأخذ العينات متعددة المراحل، يتم حساب الوزن الكلي بضرب احتمالات الاختيار في كل مرحلة ببعضها وعكس الحاصل يمثل الوزن.

على سبيل المثال، لنفترض عينة عنقودية twostages على مجتمع بحجم N=6000 بحيث يقسم المجتمع في المرحلة الأولى إلى 100 طبقة ويختار بالسحب البسيط منها عشر طبقات، n1=10 (من N1=100 طبقة /عنقود) وفي المرحلة الثانية نسحب n2=30 وحدة داخل كل طبقة من الطبقات العشر المنتقات في المرحلة الأولى، حيث عدد الوحدات داخل كل طبقة N2=60=(6000/100).

- احتمال الاختيار في المرحلة الأولى هو: $\pi_1 = \frac{n_1}{N_1} = 10/100 = 1/10$
- واحتمال الاختيار في المرحلة الثانية هو: $\pi_2 = \frac{n_2}{N_2} = 30/60 = 1/2$
- وبالتالي فإن الوزن هو: $W_d = \frac{1}{\pi_1}.\frac{1}{\pi_2} = 10 \times 2 = 20$

تعديل الوزن لعدم الإجابة

- تعاني معظم المسوحات من عدم الإجابة، والذي يحدث عندما تكون جميع أو بعض من المعلومات المطلوبة من وحدات العينات غير متوفرة لسبب ما. هناك نوعان رئيسيان من عدم الإجابة: جزئي أو كلي. يحدث عدم الإجابة الجزئي عندما تتوفر المعلومات فقط لبعض البنود في الاستمارة، أي عندما تتم الإجابة فقط على جزء من الاستبيان. في هذه الحالة، فإن النهج الأكثر شيوعا هو أن تحتسب القيم المقابلة لها على أنها مفقودة (Missing Values).

- في حالة عدم الإجابة الكلي: أي عندما تكون جميع أو كل البيانات تقريبا لوحدة أو لوحدات العينات مفقودة. يمكن أن يحدث هذا عندما يرفض المستجوب المشاركة، أو لا يمكن أن يكون موجودا أو المعلومات التي تم الحصول عليها هي غير صالحة للاستعمال. أسهل طريقة للتعامل مع مثل عدم الإجابة الكلي هو تجاهل ذلك. في بعض الظروف الاستثنائية، تقدير نسب أو المعدلات دون التعديل لعدم الإجابة الكلية هي تقريبا نفس تلك التي تنتج بتعويض عدم الإجابة. ومع ذلك، عدم تعويض الوحدات الناقصة يؤدي إلى التقليل من المجاميع (على سبيل المثال، حجم السكان، مجموع الدخل أو مساحة الأراضي المزروعة بالمحاصيل..).

الطريقة الأكثر شيوعا في التعامل مع عدم الإجابة الكلي هو تعديل تصميم الأوزان على أساس الافتراض أن الوحدات المكتملة الإجابة تمثل أيضا الوحدات التي لم تتم الإجابة عنها. في ظل هذا الافتراض، يتم إعادة توزيع وتصميم الأوزان بين المشاركين . و يتم ذلك بحساب معامل عدم الإجابة يتم ضربه في الوزن لإنتاج الوزن المعدل لعدم الإجابة، كما هو موضح في المثال أدناه.

عادة ما يتم تعريف عامل عدم الإجابة كنسبة من مجموع الأوزان في العينة الأصلية إلى مجموع الأوزان لوحدات الاستجابة. في تصميم الترجيح الذاتي، هذا ما يعادل نسبة عدد الوحدات في العينة الأصلية إلى عدد الوحدات المستجيبة (أي الكاملة الإجابة) .

- مثال لعامل عدم الإجابة:
- تم اختيار 25 شخصا من سكان 100 شخص عن طريق السحب البسيط. لنفترض أن فقط 20 شخصا قدمت جميع المعلومات المطلوبة. ما هو تعديل الأوزان لعدم الإجابة لوحدات العينة الخمسة؟
- الخطوة الأولى هي حساب احتمالات الإدراج: $\pi = \frac{n}{N} = \frac{25}{100} = \frac{1}{4}$
- الوزن هو إذن: $W_d = 1/\pi = 4$
- والخطوة الثانية هي حساب عامل عدم الإجابة:

- بما أن فقط 20 شخصا من 25 أدلوا بالبيانات الكاملة، فإن الحجم النهائي للعينة هو 20. وبافتراض وحدات الاستجابة يمكن استخدامها لتمثيل كل من الاستجابة وعدم الاستجابة، فإن عامل التعديل لعدم الإجابة هو:

$$\frac{n}{n_r} = \frac{25}{20} = 1.25$$

- الخطوة الأخيرة هي لحساب الوزن المعدل لعدم الإجابة:

$$W_{anr} = W_d \cdot \frac{n}{n} = 4 \times 1.25 = 5$$

- يمكن تطبيق هذه الطريقة في حالة ما إذا كان هناك انسجام بين خصائص المستجيبين والممتنعين عن الإجابة. و غير ذلك يتم اللجوء إلى طرق تراعي الاختلاف كما في المثال الآتي:
- أثناء جمع البيانات، فقط $n_{r1} = 150$ شخصا في الطبقة الحضرية و $n_{r2} = 40$ شخصا في الطبقة الريفية توفرت لديهم المعلومات المطلوبة. ما هي الأوزان المعدلة لعدم الإجابة في هذه العينة؟
- يتم تطبيق عاملين مختلفين لعدم الإجابة لكل من الطبقتين في المناطق الحضرية والريفية. السبب في ذلك هو أن الخصائص المدروسة تختلف بين الطبقات.

الطبقة	حجم المجتمع	حجم العينة	عدد الإجابات الكاملة
الحضر	N1=1000	n1=200	n_{r1}=150
الريف	N2=100	n2=50	n_{r2}=40

- وزن التصميم في كل طبقة : يساوي 5 للطبقة الحضرية و 2 للطبقة الريفية.
- يحسب عامل التعديل لعدم الإجابة بالنسبة لكل طبقة كما يلي:

الطبقة الحضرية:

$$\frac{n_1}{n_{r1}} = \frac{200}{150} = 1.33$$

الطبقة الريفية:

$$\frac{n_2}{n_{r2}} = \frac{50}{40} = 1.25$$

الوزن المعدل لعدم الإجابة عن كل طبقة هي نتاج تصميم الوزن الأولي مضروب في عامل التعديل:

الطبقة الحضرية: $$W_{anr1} = W_U \times \frac{n_1}{n_{r1}} = 5 \times \frac{200}{150} = 1.33 \times 5 = 6.67$$

الطبقة الريفية: $$W_{anr\,2} = W_R \times \frac{n_2}{n_{r\,2}} = 2 \times \frac{50}{40} = 2.5$$

في النهاية، يتم إعطاء كل بيانات الطبقة الحضرية وزنها النهائي 6.67 ويتم إعطاء كل من في الطبقة الريفية الوزن النهائي 2.5.

ثالثا: أخطاء المعاينة:

- إن الأخطاء التي قد نقع فيها عند استخدام أسلوب المعاينة كأسلوب لجمع البيانات تسمى أخطاء المعاينة الكلية ويمكن تقسيمها إلى نوعين من الأخطاء: خطأ المعاينة العشوائي وخطأ التحيز.

خطأ المعاينة العشوائي Randam sampling Errors

- عند اختيار عينة عشوائية حجمها n وحدة من مجتمع حجمه N وحدة معاينة نجد أن هناك خطأ ينتج عن الاختلاف بين قيم الوحدات التي تتكون منها العينة وتلك التي لم تشأ الصدف أن ندخلها في العينة وهذا الخطأ يسمى خطأ المعاينة العشوائي .
- ويمكننا باستخدام الطريقة المناسبة لاختيار الوحدة تحديد متوسط أخطاء المعاينة العشوائية من نتائج العينة وتوزيعها . إن الحجم المتوسط لهذه الأخطاء يعتمد على حجم العينة ومدى تشتت مفرداتها والإجراءات التي استخدمت لاختيار الوحدات.

- وإذا عالجنا موضوع الأخطاء بعيدا عن أخطاء التحيز، فان الطريقة الأسهل لزيادة دقة نتائج العينة هي زيادة حجمها وذلك للتقليل من خطأ المعاينة العشوائي. ويمكن القول أن خطأ المعاينة العشوائي، يتناسب عكسيا مع الجذر التربيعي لحجم العينة.
- ويمكننا تقدير خطأ المعاينة العشوائي، إذا كنا نقدر أحد معالم المجتمع بحساب الانحراف المعياري لمتوسطات العينات الممكنة الذي يسمى الخطأ المعياري Standard Error ونستخدمه للحكم على دقة الوسط الحسابي في المعاينات العشوائية وتقدير حجم العينة.

أخطاء التحيز وأنواعها Bias Errors

- عندما نستخدم أسلوب المعاينة لتقدير معلمات المجتمع، فان متوسط جميع التقديرات المحسوبة باستخدام مقدر معين للعينات الممكنة، يجب أن يساوي قيمة المعلمة التي نقوم بتقديرها، وفي حالة وجود فروق بينها فان هذا الفرق يسمى خطأ التحيز . ويعرف خطأ التحيز بأنه انحراف متوسط جميع تقديرات معلمه المجتمع للعينات الممكنة عن القيمة الحقيقية لهذه المعلمة. ويتصف التحيز بأنه ثابت القيمة وتوجد صعوبة في التقليل أو التخلص منه.
- إن خطأ التحيز لا يقل إذا ازداد عدد وحدات العينة بينما نجد أن خطأ المعاينة العشوائي يقل إذا ازداد حجم العينة كما ذكرنا .

I. خطأ التحيز في الاختيار

يوجد عدة طرق لاختيار وحدات العينة، تؤدي إلى ارتفاع خطأ التحيز:

- الاختيار غير العشوائي لوحدات العينة الذي يعتمد على مزاج الباحث دون إتباعه للتعليمات المعطاة له وعدم إتباع طرق الاختيار العشوائي.
- تعتمد بعض طرق الاختيار على خاصية معينه قد تكون مرتبطة بالخاصية المدروسة كالاعتماد على دليل الهاتف لاختيار عينه من السكان لدراسة الدخل والإنفاق، حيث نجد أن من لديهم الهاتف هم من أصحاب الدخول الجيدة. لذا يؤدي استخدام هذه الطريقة من الاختيار إلى وقوع خطأ التحيز .

- التحيز المقصود أو غير المقصود في اختيار وحدات العينة، إذ قد يقوم الباحث باختيار بعض الوحدات متعمداً إدخالها أو غير متعمد إدخالها نتيجة لأسباب متعددة. إنّ آثار هذا النوع من الأخطاء خطيرة ولا تظهر مباشرة.
- استبدال وحدة بوحدة أخرى غير مدرجة ضمن قائمة أسماء الوحدات المختارة، قد يجد الباحث صعوبة في جمع بيانات من وحدة فيأخذ وحدة أخرى (اختيار موظف عوضاً عن الموظف المحدد بالعينة لعدم وجوده).

- عدم التمكن من استكمال وصول جميع الاستمارات، على الرغم من المتابعة المستمرة والزيارات المتكررة للوحدات، خاصة إذا استخدمت طريقة المراسلة كطريقة لجمع البيانات. وللتقليل من هذه الأخطاء المتعلقة بالتحيز في الاختيار، أو التخلص منها يمكننا إتباع ما يلي :

- اختيار جميع وحدات العينه عشوائيا باستخدام إحدى طرق الاختيار العشوائي.
- عدم استبدال أية وحدة تم اختيارها بالعينة بوحدة أخرى.

- استكمال الإجابات لجميع الأسئلة ، واستلام جميع الاستمارات والقيام بالمتابعة المستمرة بالهاتف أو بالزيارات للعمل على استكمال استلام جميع الاستمارات .
- إجراء البحث التجريبي (العينة الاستطلاعية) لكشف التحيز المقصود وغير المقصود والتخلص منه أو التقليل من حجمه.
- تدريب الباحثين بشكل جيد على جمع البيانات والتقيد بالتعليمات المحددة المتعلقة بالوحدات المختارة.

2. خطأ التحيز في التقدير:

- إضافة للأخطاء التي قد تقع فيها عند اختيار وحدات العينة، هناك خطأ قد نقع فيه يتعلق بطريقة التقدير، أو طرق التحليل المناسبة يسمى خطأ التحيز في التقدير .

- كمثال هذا النوع من الأخطاء نفترض أن لدينا ثلاث حيازات زراعية، تزرع الخضراوات، وكان متوسط محاصيلها على التوالي 15، 20، 25 طنا للهكتار الواحد. إن تقدير متوسط محصول الهكتار بجمع المتوسطات الثلاثة وقسمتها على ثلاثة أي: $\frac{15 + 20 + 25}{3} = 20$ يعطي متوسط المحصول للهكتار 20 طنا. إن هذه الطريقة المستخدمة تؤدي إلى خطأ التحيز في التقدير إذ يجب أن ترجح المتوسطات السابقة بالمساحات المزروعة في المزارع الثلاث. فإذا كانت المساحات المزروعة في هذه المزارع على التوالي هي 12، 8، 14 هكتاراً فإن متوسط محصول الهكتار يساوي: $\frac{15 \times 12 + 20 \times 8 + 25 \times 14}{12 + 8 + 14} = 20.29$

3. خطأ التحيز الناتج عن التعريف الخاطئ لوحدة المعاينة:

- عندما نقوم بتحديد وحدة المعاينة ، يجب تعريفها تعريفا واضحا بشكل يقلل من أخطاء التحيز التي تنتج إذا كانت هذه الوحدة غير محددة وغير معرفه تعريفا واضحا . مثلا عندما نحدد الموظف كوحدة إحصائية لجمع البيانات عن سنوات خبرته ومدى رضاه الوظيفي، يجب أن نعرف الموظف تعريفا واضحا . ويجب توضيح ما إذا كان الموظف المتعاقد الأجنبي سيعد من وحدات المعاينة، وتبرز هذه المشكلة بشكل واضح عند اختيار وحدات لها مساحات أو قياسات معينه تختلف من تلك التي يغطيها البحث وذلك بسبب عدم تعريفها تعريفا واضحا .

(4) الأخطاء الأخرى الشائعة في العينات

- توجد أخطاء أخرى تقع عند استخدام المعاينة كأسلوب لجمع البيانات (وتقع أيضا عند استخدام أسلوب الحصر الشامل) نلخصها فيما يلي :

- أخطاء عدم الاستجابة وقد تعود إلى عدم تحديث الإطار وشموله لجميع الوحدات أو عدم إمكانية الوصول إلى الوحدات المختارة، أو عدم تواجد المدلين بالبيانات، ويؤدي ذلك إلى زيادة أخطاء المعاينة العشوائية نتيجة انخفاض حجم العينه ويؤدي ذلك أيضا إلى زيادة الأخطاء الأخرى.
- أخطاء التبويب ومعالجة البيانات، وذلك بدءا من تدقيق البيانات إلى عرضها بشكل جداول ويمكن التقليل من هذه الأخطاء عن طريق التدقيق وتصحيح الأخطاء.
- أخطاء الطباعة التي يجب تصحيحها.
- أخطاء تفسير النتائج على الرغم من صحة طرق التقدير وأساليب التحليل.

ثالثا: تقدير معالم المجتمع باستخدام المعاينة الطبقية العشوائية

I. تقدير حجم العينة:

يعتمد حجم العينة على عدد من العوامل يمكن تلخيصها في:

- الغرض من البحث.
- حجم مجتمع الدراسة.
- مدى تباين وتجانس الظاهرة المراد دراستها في المجتمع.
- درجة الدقة المطلوبة في نتائج البحث.
- البيانات الثانوية المتاحة والتي يمكن أن يستخدمها الباحث في البحث.
- الإمكانيات المادية والبشرية المتوفرة.

- هنالك اتجاهان رئيسيان في تحديد حجم العينة. الاتجاه الأول مرتبط بخبرة الباحث والخبرة هنا من شقين خبرة الباحث البحثية وخبرة الباحث بالظاهرة المراد دراستها وبمنطقة الدراسة. الاتجاه الثاني مرتبط بنظريات الاحتمالات المختلفة والذي يعتمد تنفيذه على تحديد المتغيرات والعوامل التي تؤثر في الدراسة والتي تتوفر غالبا من خلال معالم مجتمع الدراسة. يساعد وجود معلومات سابقة عن مجتمع الدراسة في تحديد حجم العينة وفي حالة عدم توفر مثل هذه المعالم يلجأ الباحث للعينات الاسترشادية أو الاستطلاعية لتقدير بعض معالم المجتمع الضرورية لتحديد حجم العينة.

- حجم العينة الأمثل هو الحجم الذي يعطينا أقصى دقة بأقل ما يمكن من التكاليف، ولكن عملياً نجد أن حجم العينة الأمثل هو الذي يعطي أعلى دقة ممكنة بتكاليف محددة بصورة مسبقة. ولتوزيع حجم العينة الإجمالي على مختلف الطبقات بحيث يعطي أقل ما يمكن من أخطاء المعاينة يوجد عدة طرق تسمى طرق توزيع العينة Sample allocation وتتلخص فيما يلي:
  - يتم توزيع حجم العينة الإجمالي على مختلف الطبقات بشكل متساو:

- يوزع حجم العينة الطبقية على الطبقات على أساس درجة تجانس هذه الطبقة وإدخال عامل التكاليف. فإذا كانت مفردات الطبقة متجانسة فإننا نختار عددا أقل من الوحدات، وكلما قل التجانس في مفردات الطبقة ازداد عدد الوحدات التي نختارها من الطبقة، وذلك للتقليل من أخطاء المعاينة. ويمكننا القول أنه عند استخدام طريقة التوزيع الأمثل يكون حجم العينة من الطبقة كبيراً عندما يكون حجم الطبقة من المجتمع كبيراً أو تباين هذه الطبقة كبيراً أو يكونان كلاهما معا كبيرين. وعند إدخال عامل التكاليف في تحديد حجم العينة في الطبقة نجد أن هذا الحجم يقل إذا كانت تكاليف الوحدة كبيرة والعكس بالعكس وذلك إضافة لحجم وتباين الطبقة في المجتمع.

2. تقدير متوسط المجتمع والقيمة الكلية للمجتمع:

- إن الغاية الأساسية من استخدام أسلوب المعاينة، تعميم نتائج العينة على المجتمع الذي اختيرت منه ولنوضح الآن كيفية تقدير كل من متوسط المجتمع والقيمة الكلية لمفردات المجتمع من بيانات العينة الطبقية. إذا سحبنا عينة طبقية من مجتمع مكون من L طبقة يكون لدينا L متوسطاً للطبقات وهي: $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_L$ حيث:

متوسط الطبقة ذات الرتبة h:	$\bar{x}_h = \frac{x_h}{n_h}$
(حيث $x_1, x_2, \ldots, x_h$ هي مجموع قيم العينة للطبقات 1,2,...,h على التوالي)إن متوسط العينة في الطبقة h هو مقدر غير متحيز ومتسق لمتوسط الطبقة h في المجتمع، أي أن $\bar{x}_h$ هو مقدر غير متحيز ومتسق لـ $\bar{X}_h$	
تقدير القيمة الكلية للمجتمع : بترجيح متوسطات الطبقات في العينة بأحجامها في المجتمع وذلك كما يلي:	$\hat{X}_h = N_h \bar{x}_h$
مقدار القيمة الكلية للطبقة ذات الرتبة h يساوي:	$\hat{X}_{st} = \sum_{h=1}^{L} \hat{X}_h$
مقدر القيمة الكلية للمجتمع من عينة طبقية ولنرمز له بالرمز $\hat{X}_{st}$ يساوي مجموع تقديرات القيمة الكلية للطبقات أي يساوي:	أي أن: $\hat{X}_{st} = \sum_{h=1}^{L} N_h \bar{x}_h$

3. تقدير متوسط المجتمع على أساس عينة طبقية

إذا رمزنا لمقدر متوسط المجتمع على أساس عينة طبقية بالرمز $\bar{x}_{st}$ فإنه يساوي متوسطات الطبقات من العينة مرجحة بنسبة حجم الطبقة في المجتمع إلى إجمالي حجم المجتمع، أي:	$\bar{x}_{st} = \sum_{h=1}^{L} \frac{N_h \bar{x}_h}{N}$
أي يساوي:	$\bar{x}_{st} = \sum_{h=1}^{L} w_h \bar{x}_h$
حيث:	$w_h = \frac{N_h}{N}$

- ويعد متوسط العينة الطبقية $\bar{x}_{st}$ مقدراً غير متحيز ومتسقاً لمتوسط المجتمع، حيث نعلم أن توقع المقدر يجب أن يساوي متوسط المجتمع لكي يعد غير متحيز.

مثال:

- يتكون مجتمع من 6 موظفين يعملون في الادارتين (أ) و (ب) وكانت سنوات الخبرة لديهم:

$$X_{11}=2,\quad X_{12}=4,\quad X_{13}=6$$
$$X_{21}=8,\quad X_{22}=12,\quad X_{23}=16$$

المطلوب:

1) الوسط الحسابي لسنوات الخبرة للموظف في كل ادارة
2) الوسط الحسابي لسنوات الخبرة للموظفين
3) اجمالي عدد سنوات الخبرة لدى الموظفين

الحل:

عدد سنوات الخبرة في كلتا الادارتين:

$$X_h = \sum_{i=1}^{N_h} X_{hi}$$

نستخدم الصيغة التالية:

$$X_1 = X_{11} + X_{12} + X_{13}$$
$$= 2 + 4 + 6 = 12$$
$$X_2 = X_{21} + X_{22} + X_{23}$$
$$= 8 + 12 + 16 = 36$$

ويكون عدد سنوات الخبرة في الادارة (أ)

وعدد سنوات الخبرة في الادارة (ب)

$$X = \sum_{h=1}^{L} X_h = 12 + 36 = 48$$

اجمالي عدد سنوات الخبرة في الادارتين:

$$\overline{X}_h = \frac{X_h}{N_h} = \sum_{i=1}^{N_h} \frac{X_{hi}}{N_h}$$

الوسط الحسابي للطبقة h في المجتمع:

$$\overline{X}_1 = \frac{12}{3} = 4$$

ومتوسط سنوات الخبرة في الادارة (ب):

$$\overline{X}_2 = \frac{36}{3} = 12$$

متوسط المجتمع أي متوسط سنوات الخبرة للموظف سواء كان في الادارة (أ) أو الادارة (ب):

$$\overline{X} = \frac{X}{N} = \sum_{h=1}^{L} \frac{X_h}{N} = \frac{12 + 36}{6} = 8$$

ونحصل على النتيجة نفسها باستخدام الصيغة التالية:

$$\overline{X} = \sum_{h=1}^{L} \frac{N_h \overline{X}_h}{N} = \frac{N_1 \overline{X}_1 + N_2 \overline{X}_2}{N} = \frac{(3 \times 4) + (3 \times 12)}{6} = 8$$