

دليل تقنيات المعاينة

مقدمة

1. أنواع العينات الإحصائية
1.1- العينة العشوائية البسيطة (Simple Random Sampling)

1.2- العينة العنقودية (Cluster Sampling)

1.3- العينة الطبقيّة (Stratified Sampling)

1.4- العينة المتعددة المراحل (Multistage Sampling)

1.5- العينة الأساسية (Master Sample)

1.6- العينة على مرحلتين (Two phases sampling)

1.7- العينات المتكررة (Rotation Sampling)

2. حجم العينة

2.1- حجم العينة العشوائية البسيطة (مع الإعادة) لتقدير المتوسط أو النسبة

2.2- حجم العينة العشوائية البسيطة (بدون إعادة) لتقدير المتوسط أو النسبة

2.3- حجم العينة العشوائية البسيطة لتقدير عدد من المؤشرات

2.4- حجم العينة المركبة

2.5- حجم العينة للمسوح الأسرية

2.6- توزيع العينة على الطبقات

3- أوزان العينة (Sampling Weights)

3.1- سحب وحدات العينة

3.2- الوزن الأساسي (المطلق)

3.3- التعويض عن عدم الاستجابة

3.4- التعويض عن عدم التغطية

3.5- التعويض عن الاستجابات غير المكتملة (Imputation)

3.6- تخفيض الأوزان (Trimming of Weights)

3.7-الأوزان النسبية

4-أخطاء العينة (Sampling Errors)

4.1-أخطاء العينة العشوائية البسيطة

4.2-أخطاء العينة الطبقية

4.3-أخطاء العينة العنقودية

4.4-أخطاء العينة المتعددة المراحل (المركبة)

المراجع

مقدمة

تعتبر المعاينة الإحصائية إحدى طرق جمع البيانات الإحصائية، والتي تقتصر على جزء من المجتمع المستهدف، تمييزاً لها عن طريقة الحصر الشامل لجميع أجزاء المجتمع المستهدف والمعروفة بالتعداد. وقد تطورت طريقة المعاينة إلى ما هو معروف اليوم خلال فترة طويلة من الزمن. وارتبط تطورها بنظرية الاحتمالات. وكان لتطور قانون الأعداد الكبيرة، ونظرية النزعة المركزية، وقانون التوزيع الطبيعي، وتوزيع (t-student) الخاص بالعينات الصغيرة، والتطبيقات العملية التي رافقت ذلك، خاصة في النصف الثاني من القرن العشرين، أهمية كبيرة في تطور أساليب وطرق المعاينة الإحصائية وخاصة ما يتعلق بطرق حساب الأخطاء العينية وتعميم نتائج العينة على المجتمع. ونظراً للترابط الوثيق بين نظرية الاحتمالات ونظرية العينات، يميز بين نوعين مختلفين من العينات: **النوع الأول**، ويسمى **العينات الاحتمالية**، وهي العينات التي يتم اختيارها بالطريقة العشوائية، ويكون لكل وحدة من وحداتها احتمالاً محددًا مختلفًا عن الصفر. ويتم التركيز في التطبيقات العملية على هذا النوع، حيث يمكن الاعتماد على نتائج هذه العينات لتعميمها على المجتمعات التي سحبت منها، وتحديد درجة الثقة بالنتائج وحساب الأخطاء المرتكبة. **النوع الثاني**، ويسمى **بالعينات غير الاحتمالية**، وهي العينات التي لا تعتمد على مبادئ الاحتمالات، حيث يتم اختيار الوحدات المكونة للعينة بطريقة تحكومية أو شخصية. وهي على عكس ما هو في العينات الاحتمالية فلا يمكن الدفاع بثقة عن دقة نتائجها وتمثيلها للمجتمع. ولهذا السبب وندرة استخدام طريقة العينات غير الاحتمالية، سيتم اقتصار العرض لاحقاً على تقنيات العينات الاحتمالية. وقبل عرض التقنيات الخاصة بهذه الطريقة سيتم عرض موجز لبعض المفاهيم الخاصة بالمعاينة.

- المجتمع المستهدف، ويقصد بذلك الوحدات التي ينبغي تغطيتها بالمسح، ففي استقصاء متعلق بسكان منطقة أو بلد معين، يكون المجتمع المستهدف هو، جميع سكان هذه المنطقة أو البلد، والمجتمع المستهدف في المسح الاقتصادي هو جميع المنشآت الاقتصادية والاجتماعية.

- مجتمع الاستقصاء، في الواقع العملي، قد توجد بعض القيود لعدم تغطية بعض وحدات المجتمع المستهدف، بسبب التكلفة الكبيرة لتغطيتها أو بسبب صعوبات الوصول إلى هذه الوحدات، مثل السكان الذين يقيمون في الخارج، أو الوحدات التي تقع في مناطق الكوارث، وبناء على ذلك ينبغي حذف هذا النوع من الوحدات من المجتمع المستهدف بالاستقصاء، وتسمى الوحدات المتبقية بعد هذا الاجراء، "مجتمع الاستقصاء" أو (مجتمع المسح).
- إطار العينة، ويقصد به قائمة الوحدات التي يتم سحب العينة منها في كل مرحلة من المراحل، ويمثل الإطار حجر الزاوية في تصميم العينة وتنفيذ جميع مراحلها اللاحقة، وان نجاح برنامج المعاينة يتوقف بشكل كبير على توفر أطر حديثة ومناسبة لتصميم العينة. إن الإطار النموذجي للمعاينة هو أحدث إطار يضم جميع الوحدات. ومن الأمثلة على الأطر الأكثر استخداما في التطبيقات العملية، إطار الأسر، وإطار المنشآت. إن من أهم الميزات التي يجب أن تتوفر في الإطار هو أن يكون كاملاً بحيث يشمل جميع الوحدات موضوع الدراسة. ويجب أن ترد كل وحدة من وحدات المسح المرغوب مرة واحدة وواحدة فقط في الإطار أو القوائم التي تشمل هذه الوحدات، مع عدم وجود أية وحدات غريبة. وفي هذا السياق، إن توفر الخرائط المرافقة لهذه الاطر يحتل أهمية خاصة في تحديد مواقع وحدات العينة، وتسهيل الاعمال الميدانية، وتقليل التكلفة، وبالتالي رفع كفاءة المسح. وتم عرض الموضوعات الخاصة بهذا المنشور بالاعتماد على العديد من منشورات تقنيات المعاينة، كما في أعمال العديد من المختصين أمثال (Kish)، (Cochran)، (Hansen) وغيرهم، وفي منشورات العديد من المؤسسات الدولية وخاصة منشورات الأمم المتحدة الصادرة حديثاً عن قسم الإحصاء في الأمم المتحدة مثل: استقصاء الأسر المعيشية في البلدان النامية والبلدان التي تمر اقتصاداتها بمرحلة انتقالية (Household sample surveys in Developing and transition Countries) 2005 وتصميم عينات مسوح الأسر المعيشية، ومبادئ توجيهية عامة 2010 (Designing Household Survey Samples: Practical Guidelines) و"دليل مسوح الأسر" الصادرة عام 1987 (Handbook of Household Surveys)، بالإضافة إلى أهم الممارسات العالمية التي نفذت في العديد من بلدان العالم منذ الثمانينات

من القرن الماضي، ومن أهمها مسح الخصوبة العالمي (WFS) (World Fertility Survey) الذي نفذ في العديد من البلدان منذ عام 1974، والمسح الديموغرافي الصحي (DHS) (Demographic and Health Survey) الجاري تنفيذه في العديد من بلدان العالم منذ 1984.

1. أنواع العينات الاحتمالية

1.1- العينة العشوائية البسيطة (Simple Random Sampling)

تعرف العينة العشوائية البسيطة بأنها عدد من الوحدات، التي يتم سحبها بالطريقة العشوائية من إطار العينة الذي يشمل جميع الوحدات (أفراد، أسر، منشآت، وغيرها)، ومن خلال المعلومات التي يتم جمعها عن وحدات العينة، تتم جميع التقديرات المقابلة والخاصة بمجتمع المسح، وتقديرات الأخطاء المرتكبة فيها، ومستويات ومجالات الثقة بهذه التقديرات. إن العامل الحاسم في العينات الاحتمالية هو سحب وحدات العينة بالطريقة العشوائية، والشرط الاساسي لتحقيق العشوائية هو أن يكون لكل وحدة من وحدات إطار العينة فرصة للسحب في العينة، ويعني ذلك عدم وجود أية قيود لعدم وقوع أية وحدة في العينة، ويعبر عن هذا المبدأ في نظرية الاحتمالات، بأن لكل وحدة احتمالٍ مختلفٍ عن الصفر لسحبها في العينة. وتعتبر العينة العشوائية البسيطة الأساس الذي بنيت عليه طرق العينات الأخرى، لذا سيتم عرض موجز للأساس النظري للعينة العشوائية، والمتمثل بقانون الأعداد الكبيرة (LLN) (The Law of Large Numbers) ونظرية النزعة المركزية (CLT) (The Central Limit Theorem).

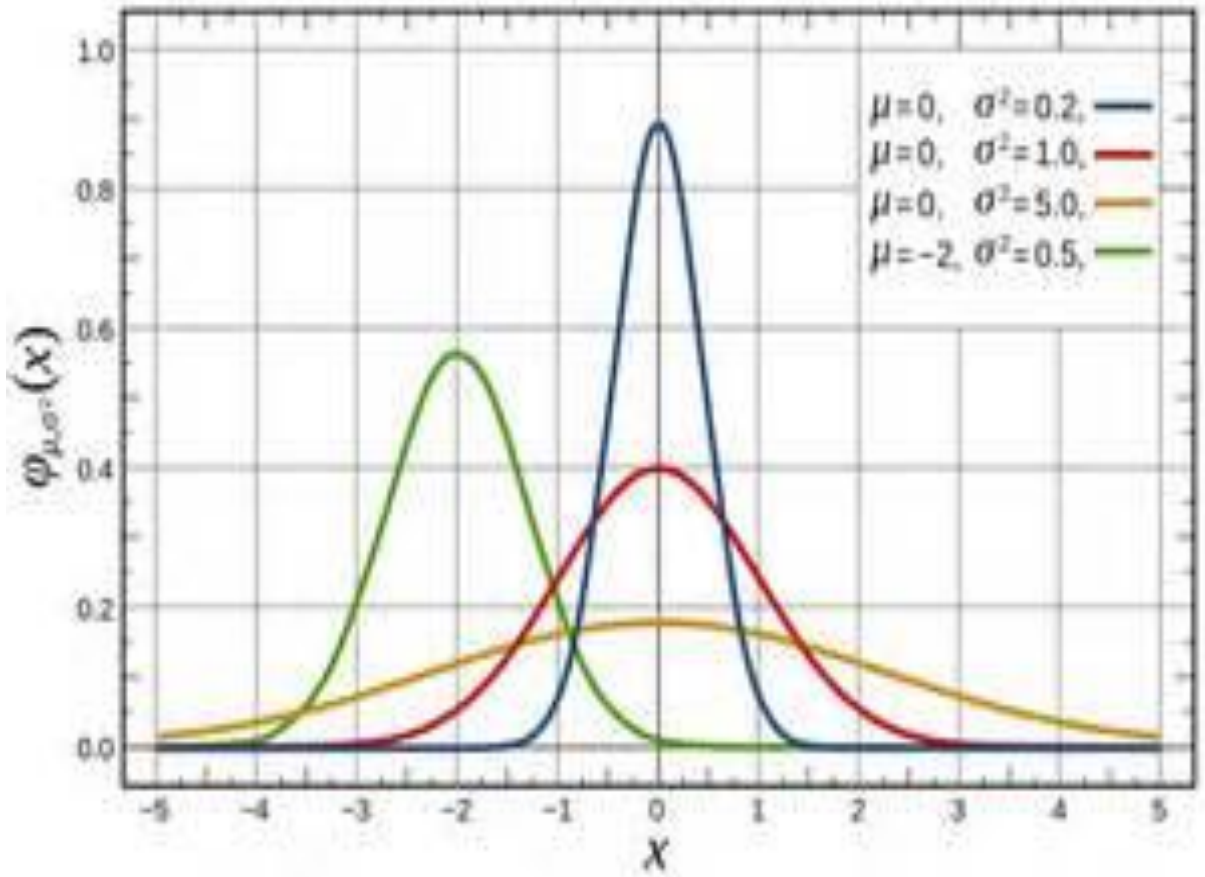
لوحظ منذ القدم، أنه من خلال عدد محدود من المشاهدات الخاصة بظاهرة معينة يمكن التعرف على بعض خصائص هذه الظاهرة، مثل المتوسط الحسابي أو النسبة، وعندما يتم تكرار هذه العملية على عدد محدود آخر من المشاهدات (أكبر أو يساوي 30)، فإن هذه الصفات ستختلف اختلافات صغيرة عن الصفات المحسوبة من المشاهدات السابقة، وستختلف اختلافات طفيفة عن الصفات الحقيقية في المجتمع. وكان (Bernoulli) أول من لاحظ ذلك من خلال سحب عدد محدود من الكرات من صندوق يحوي نوعين من الكرات بنفس الظروف ودون أي تدخل مقصود (وهذا ما أطلق عليه لاحقاً اسم السحب العشوائي)، فوجد أن نسبة الكرات من نوع معين بين الكرات المسحوبة قريبة من نسبة هذا النوع بين إجمالي الكرات في الصندوق، وتتحقق هذه النتيجة عندما يتم سحب الوحدات مع الإعادة (With Replacement)، حيث تعاد

الوحدة المسحوبة إلى الصندوق قبل عملية السحب التالية، أو بدون إعادة (Without Replacement)، حيث لا تعاد الوحدة المسحوبة إلى الصندوق بعد سحبها. وكان ذلك مقدمة لأول صياغة رياضية لقانون الأعداد الكبيرة. وعلى هذا المنوال، يمكن التعرف على نسبة المتعطلين بين السكان في منطقة معينة، من خلال معاينة عدد كبير من السكان، وكذلك يمكن التعرف على متوسط دخل الأسرة في مدينة محددة من خلال معاينة عدد كبير من الأسر. إن ما يجمع الأمثلة السابقة، وغيرها من الأمثلة العديدة المشابهة للظواهر الطبيعية أو الاجتماعية، هو العدد الكبير من المشاهدات. وقد سمي هذا المبدأ "بقانون الأعداد الكبيرة"، وكان مقدمة لنظرية النزعة المركزية.

إن من أهم النظريات في تقنية المعاينة الاحصائية، هي نظرية النزعة المركزية، ذات التطبيق الواسع. وكان العالمان (Laplace) و (De Moivre) أول من صاغوا هذه النظرية في النصف الأول من القرن الثامن عشر. وكانت الصيغة النهائية لنظرية النزعة المركزية في أعمال العالمين (Chebyshev) و (Lyapunov) في نهاية القرن التاسع عشر وبداية القرن العشرين، فقد تبين في ضوء نتائج التجارب العملية لقانون الأعداد الكبيرة، أنه إذا تم تشكيل جميع العينات العشوائية التي حجم كل منها n من مجتمع حجمه N ، وكان حجم العينة أكبر أو يساوي 30 (وهذا ما يعرف بالعينات الكبيرة)، فإن متوسطات هذه العينات، تتوزع وفقاً لقانون **التوزيع الطبيعي (Normal Distribution)** بغض النظر عن قانون توزيع المجتمع. وكما هو معروف، فإن دالة التوزيع الطبيعي هي:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

وجميع هذه التوزيعات تشترك بمنحنيات متماثلة ولكنها تختلف من حيث المتوسط \bar{X} (μ) والتباين σ^2 . كما في الشكل التالي:



المصدر: <http://www.bing.com/>

ولاحقاً في عام 1908 توصل العالم البريطاني (Student) الى إن هذا المبدأ صحيح، عندما يكون حجم العينة أصغر من 30 ويكون توزع المجتمع طبيعياً، فإن متوسطات جميع العينات يتوزع وفقاً لقانون التوزيع الخاص بالعينات الصغيرة، المعروف بتوزيع (t-Student)، وهو يشبه التوزيع الطبيعي، ويتطابق معه عندما يكون حجم العينة أكبر من (30).

وبهدف تسهيل استخدامات التوزيعات الطبيعية المختلفة، تم إيجاد توزيعاً جديداً سمي **التوزيع الطبيعي المعياري (Standard Normal Distribution)**، بناء على متوسط وتباين التوزيع الطبيعي، وذلك

باستبدال المتحول العشوائي x في دالة التوزيع الطبيعي بالمتحول العشوائي Z التالي:

$$z = \frac{x - \bar{X}}{\sigma}$$

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

وتصبح دالة التوزيع الجديد:

ويسمى التوزيع الناتج أحياناً، توزيع Z ، ومن أهم خصائص التوزيع الطبيعي المعياري إن متوسط هذا التوزيع يساوي (0)، وتباينه يساوي (1) مهما كان متوسط وتباين التوزيع الطبيعي الذي تم تحويله، ولهذا السبب سمي التوزيع الجديد بالمعياري لأنه سيكون واحداً مهما اختلفت التوزيعات الطبيعية التي استخدمت لتشكيله، وشكله مشابه لشكل التوزيع الطبيعي، كما مبين باللون الأحمر في الشكل السابق. وقد أعد جدول خاص بقيم هذا التوزيع الموافقة لقيم Z والذي يمكن إيجاده في العديد من المراجع الإحصائية والبرامج الجاهزة مثل برنامج (Excel).

ولهذه النتيجة أهمية كبيرة في نظرية العينات، حيث يتم تقدير القيم الحقيقية في المجتمع (مثل المتوسط أو النسبة) عن طريق التقديرات المقابلة لها في العينة. فإذا تم تشكيل جميع العينات العشوائية التي حجم كل منها n من مجتمع حجمه N وكانت متوسطات هذه العينات التالية

$$\bar{x}_1, \bar{x}_2, \dots, \bar{x}_L$$

حيث: L عدد العينات العشوائية التي يمكن تشكيلها، وبناء على نظرية النزعة المركزية، فإن متوسطات العينات العشوائية تتوزع وفقاً للتوزيع الطبيعي، ويتحدد هذا التوزيع بالمتوسط الذي يساوي متوسط

المجتمع \bar{X} وتباينه $\sigma^2(\bar{x})$ وتنطبق هذه النتيجة على جميع أنواع المجتمعات. ويمكن البرهان على أن:

$$\frac{\sum_{i=1}^{i=L} \bar{x}_i}{L} = \bar{X}$$

• متوسط هذا التوزيع يساوي متوسط المجتمع

$$\sigma^2(\bar{x}) = \frac{\sigma^2}{n}$$

تباين متوسط هذا التوزيع مع الإعادة (مع الإرجاع) يساوي

• تباين متوسط هذا التوزيع بدون إعادة (بدون إرجاع) يساوي

$$\sigma^2(\bar{x}) = \frac{\sigma^2}{n} * \frac{N - n}{N - 1}$$

ويلاحظ أن تباين المتوسط في حالة بدون إعادة يختلف عنه في حالة مع الإعادة بالمعامل، $\frac{N - n}{N - 1}$

وعندما يكون حجم المجتمع كبيراً، فإن الفرق بين القيمة $N - 1$ والقيمة N صغيراً جداً، وباستبدال ذلك في عبارة التباين بدون إعادة تصبح

$$\sigma^2(\bar{x}) = \frac{\sigma^2}{n} * \frac{N - n}{N} = \frac{\sigma^2}{n} * \left(1 - \frac{n}{N}\right) = \frac{\sigma^2}{n} * (1 - f)$$

حيث $f = \frac{n}{N}$ ويسمى كسر المعاينة (Sampling Fraction). ويسمى $1 - f$ معامل تصحيح المجتمع

(PCF) (Population Correction Factor).

وبناء على ذلك يمكن إيجاد التوزيع الطبيعي المعياري لهذا المجتمع باعتبار أن $Z = \frac{\bar{x} - \bar{X}}{\sigma(\bar{x})}$ ويمثل \bar{X}

متوسط إحدى العينات التي تم تشكيلها من هذا المجتمع، ويمثل $\sigma(\bar{x})$ الانحراف المعياري لتباين هذا المتوسط، ويمكن البرهان على أن متوسط المتحول الجديد يساوي الصفر، وتباينه يساوي (1) ويتوزع وفق التوزيع الطبيعي المعياري، وبناء على خواص هذا التوزيع، يمكن حساب احتمال أن تأخذ Z قيمة محصورة

بين أي قيمتين متناظرتين، فإذا كانت $f(Z)$ دالة التوزيع الطبيعي المعياري، فإن هذا الاحتمال يحسب وفقاً للعلاقة التالية

$$\Pr\left(-Z \leq \frac{\bar{x} - \bar{X}}{\sigma_{(\bar{x})}} \leq +Z\right) = \int_{-Z}^{+Z} f(Z) * dZ = 2 \int_0^{+Z} f(Z) * dZ$$

وترتبط العلاقة الأخيرة بين متوسط المجتمع غير المعروف ومتوسط إحدى العينات التي حجم كل منها n وفي الواقع العملي يتم التعامل مع عينة واحدة من بين جميع العينات المتشكلة من المجتمع، ومن ذلك يمكن حساب الفرق بين متوسط العينة ومتوسط المجتمع، وقد يكون هذا الفرق موجباً أو سالباً، وتعتبر القيمة المطلقة لهذا الفرق عن الخطأ المرتكب في تقدير متوسط المجتمع، ويسمى: هامش الخطأ (Margin of Error)، ويرمز له بالرمز e

$$e = |\bar{X} - \bar{x}|$$

وبناء على ذلك يمكن أن يحسب احتمال تحقق العلاقة التالية

$$-Z \leq \frac{\bar{X} - \bar{x}}{\sigma_{(\bar{x})}} \leq +Z \Leftrightarrow -Z * \sigma_{(\bar{x})} \leq e \leq +Z * \sigma_{(\bar{x})}$$

فإذا كانت $Z = 1$ فإن هامش الخطأ يكون محصوراً بين قيمتين متناظرتين، كما في العلاقة التالية

$$-\sigma_{(\bar{x})} \leq e \leq +\sigma_{(\bar{x})}$$

ويسمى هامش الخطأ عندئذ بالخطأ العشوائي (Standard Error) (SE)، ويعادل مثلاً واحداً من الانحراف المعياري لتباين متوسط العينة، ويسمى أحياناً الخطأ المعياري أو خطأ الصدفة. وبالاعتماد على جدول التوزيع الطبيعي المعياري المعروف يمكن حساب قيمة الاحتمالات المقابلة لقيم Z المختلفة. ويسمى هذا الاحتمال مستوى الثقة (Confidence Level)، وأحياناً يسمى مجال الثقة (Confidence Interval)، وفي أغلب الدراسات والمسوح الإحصائية يتم الاعتماد على مستوى ثقة أكبر من 90%، وغالباً 95% والتي يقابلها في جدول التوزيع الطبيعي قيمة $Z = 1.96$ ، وبشكل عام يمكن كتابة العلاقة التالية التي تربط بين متوسط المجتمع ومتوسط إحدى العينات، وهامش الخطأ

$$-Z * \sigma_{(\bar{x})} \leq \bar{X} - \bar{x} \leq +Z * \sigma_{(\bar{x})}$$

وبناء على ذلك يمكن تحديد المجال الذي يقع فيه متوسط المجتمع، والثقة بهذه النتيجة التي يعبر عنها بالاحتمال أو مستوى الثقة. ويمكن التعبير عن هذا المجال بالعلاقة

$$\bar{x} - e \leq \bar{X} \leq \bar{x} + e$$

ويسمى الطرف الايمن لهذه المترجحة بالحد الاعلى لمتوسط المجتمع، والطرف الايسر بالحد الادنى، ويسمى التقدير لمتوسط المجتمع وفقاً لذلك **(التقدير بمجال)**، تمييزاً عن التقدير الذي يعتمد على متوسط العينة فقط، دون اعتبار لهامش الخطأ، والذي يسمى **(التقدير بنقطة)**.

وعلى سبيل المثال، سحبت عينة حجمها 100 وحدة من مجتمع حجمه 1000 وحدة لتقدير متوسط إحدى الظواهر، فإذا كان تباين هذه الظاهرة في المجتمع 2000، والمتوسط في العينة يساوي 80، فماذا يمكن القول عن هامش الخطأ في حالتي الإعادة (الارجاع)، ودون إعادة (بدون إرجاع)؟ وما هو الحد الاعلى والادنى لمتوسط المجتمع بمستوى ثقة 95%.

هامش الخطأ مع الاعادة:

$$e_0^2 = Z^2 * \frac{\sigma^2}{n_0} = (1.96)^2 * \frac{2000}{100} = 76.83 \Rightarrow e_0 = 8.77$$

$$\bar{x} - e \leq \bar{X} \leq \bar{x} + e$$

ويكون تقدير متوسط المجتمع

$$80 - 8.77 \leq \bar{X} \leq 80 + 8.77$$

$$71.23 \leq \bar{X} \leq 88.77$$

وتعني هذه النتيجة أن متوسط المجتمع سيقع في هذا المجال بمستوى ثقة 95%. ويمكن أن يقع متوسط المجتمع خارج هذا المجال بثقة 5% وفي هذه الحالة يكون متوسط المجتمع أكبر من 88.77 أو أصغر من 71.23. ويتعلق ذلك بالعينة التي تم اختيارها من بين جميع العينات التي يمكن تشكيلها من المجتمع. وينبغي مراعاة هذه النتيجة الهامة لدى تصميم العينة بحيث تكون عينة ممثلة للمجتمع، عن طريق اختبار توافق عدد من المؤشرات المعروفة للمجتمع (متوسط حجم الأسرة، التركيب العمري والنوعي وغيرها) مع مثيلاتها في العينة التي تم اختيارها، حيث يتم تغيير العينة إذا لم تتوافق خصائصها العامة مع الخصائص المعروفة في المجتمع.

هامش الخطأ بدون إعادة:

$$e^2 = Z^2 * \frac{\sigma^2}{n} * \frac{N - n}{N - 1} = 69.2 \Rightarrow e_0 = 8.32$$

وهو أصغر منه في حالة الاعادة، مع ملاحظة أن هامش الخطأ في حالة الاعادة لا يتأثر بحجم المجتمع. وفي حالة المجتمعات الكبيرة فإن هامش الخطأ واحد تقريباً في حالتها الاعادة وبدون إعادة. فإذا كان حجم المجتمع 1,000,000، فإن هامش الخطأ بدون إعادة يساوي 8.77 أيضاً، وهو هامش الخطأ مع الاعادة. وبطريقة مشابهة لتقدير متوسط المجتمع، يمكن أن يعمم ذلك على نسبة ظاهرة في مجتمع، ويعود ذلك إلى أن نسبة الظاهرة في المجتمع يمكن اعتبارها متوسط، وذلك بإسناد القيمة (1) لكل قيمة توافق الظاهرة في

المجتمع، والقيمة (0) للقيم التي لا توافق الظاهرة. وعندئذ يكون متوسط القيم التي توافق الظاهرة يحسب، كما هو معروف، وفق العلاقة التالية:

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_N}{N}$$

وهو ما يسمى نسبة الظاهرة في المجتمع Proportion ويرمز لها بالرمز P للتمييز بينها وبين المتوسط، وتكون النسبة التي لا توافق الظاهرة المدروسة (المعكوسة)، ويرمز لها بالرمز $Q = 1 - p$ ونظراً لأن أي قيمة في المجتمع تأخذ القيمة (0) أو (1)، فإن تباينها وفقاً لقانون توزيع (Bernoulli) $\sigma^2 = PQ$ وبطريقة مشابهة لما تم بخصوص المتوسط، يمكن كتابة مربع الخطأ الخاص بتقدير نسبة الظاهرة في مجتمع، وفق ما يلي:

مع الاعادة

$$e_0^2 = Z^2 * \frac{PQ}{n_0}$$

بدون إعادة

$$e^2 = Z^2 * \frac{PQ}{n} * \frac{N - n}{N - 1}$$

ومن ذلك يمكن استنتاج النتائج الهامة التالية:

- يتناسب هامش الخطأ، (في حالتي الاعادة، وبدون إعادة)، طردياً مع تباين الظاهرة المدروسة، ومستوى الثقة المرغوب بالبيانات.
- يتناسب هامش الخطأ عكساً مع حجم العينة (في حالتي الاعادة، وبدون إعادة).
- إن هامش الخطأ لا يتعلق بحجم المجتمع في حالة (مع الاعادة)، ووجود هذه العلاقة في حالة (بدون إعادة).

- إن هامش الخطأ بدون إعادة أصغر منه في حالة الاعادة، من أجل حجم عينة واحد في

الحالتين، لأن النسبة $\frac{N - n}{N - 1}$ تكون أصغر من (1) دوماً، مع الأخذ بالاعتبار أن حجم

العينة أكبر من (1) دوماً.

1.2-العينة العنقودية (Cluster Sample)

إن سحب وحدات العينة بطريقة العينة العشوائية البسيطة يؤدي الى تناثر وحدات العينة، مما يزيد من تكلفة الاعمال الميدانية في العديد من المسوح الإحصائية، وبهدف تقليل تكلفة العمل الميداني فقد تم تطوير أسلوباً آخر هو أسلوب العينة العنقودية. ويقوم هذا الأسلوب على دراسة وحدات المعاينة في مناطق محدودة بدلاً من دراسة الوحدات المتناثرة في المجتمع. ويسمى كل من هذه المناطق عنقوداً. إذ بدلاً من دراسة وحدات عينة حجمها /1000/ أسرة، على سبيل المثال في إحدى المدن، فيمكن تجميع وحدات العينة في عناقيد حجم كل منها يساوي 50 وحدة، وبدلاً من جمع البيانات عن الأسر المتناثرة في جميع أنحاء المدينة، يتم الاقتصار على 20 عنقود بحيث تغطي جميع وحدات العنقود أو جزء منها في المسح. ومن الواضح التكلفة التي يمكن تخفيضها وفقاً لذلك. وعلى الرغم من أن أسلوب المعاينة العنقودية يقلل التكلفة إلى درجة كبيرة بالمقارنة مع أسلوب العينة العشوائية البسيطة، إلا أن هذا الأسلوب، يزيد من أخطاء المعاينة بالمقارنة مع أخطاء المعاينة العشوائية البسيطة، وهذا أحد العيوب الرئيسة لهذا الأسلوب. وقد تم تطوير هذا النوع من طرق المعاينة، واستخدمت أساليب متنوعة في التطبيقات العملية لجهة حجم العناقيد وتقليل كل من الخطأ والتكلفة. فيمكن استخدام عناقيد ذات أحجام متساوية ودراسة جميع مفردات العنقود، وقد تتم الدراسة على عدد من مفردات العنقود، ويمكن أن يكون عدد هذه المفردات واحداً في كل عنقود، أو يمكن أن يكون مختلفاً.

1.3- العينة الطبقية (Stratified Sample)

من المرغوب غالباً، في العديد من الدراسات الإحصائية، الحصول على بيانات مستقلة لبعض المناطق الجغرافية لأغراض إدارية أو تنموية مثل: الحضر والريف أو الاقليم أو المحافظة أو المنطقة أو المواطنين وغير المواطنين. وفي الواقع العملي، هناك مبررات قوية للتمييز بين هذه المناطق، فسكان الريف، على سبيل المثال، يتميزون عن سكان الحضر في العديد من الخصائص الاجتماعية والاقتصادية والديموغرافية. ومن جهة أخرى أثبتت التجربة والواقع العملي، بأنه يمكن تقليل الخطأ العشوائي بتقسيم المجتمع إلى عدد محدد من الاقسام، بحيث تكون وحدات كل منها متماثلة أو متجانسة، بحيث لا يوجد فروق كبيرة بين وحدات كل قسم، بينما توجد مثل هذه الاختلافات بين الاقسام، وقد أطلق اسم الطبقة على كل قسم من هذه الاقسام، لذا سمي هذا النوع الجديد من المعاينة: **بالمعاينة الطبقية**. وتتلخص هذه الطريقة بتقسيم المجتمع إلى عدد من الاقسام يسمى كل منها طبقة. ومن أهم خصائص العينة الطبقية أنها تعطي دقة أكثر ضمن شروط معينة، كما أنها تتيح إمكانية جمع المعلومات عن كل طبقة باعتبارها مجتمعاً قائماً بذاته.

1.4- العينة المتعددة المراحل (Multistage Sample)

إن طرق المعاينة التي تم استعراضها سابقاً، يتطلب استخدامها توفر أطر حديثة للمجتمع المدروس بحيث يتضمن هذا الإطار قوائم بجميع وحدات المجتمع. وفي الواقع العملي قد لا تتوفر مثل هذه الأطر في بعض الأحيان. وعادة ما يكون توفير إطار حديث مكلفاً للغاية. فدراسة ظاهرة البطالة تتطلب توفر إطار حديث للأسر، خاصة في الفترات البعيدة عن تاريخ إجراء آخر تعداد، وان توفير إطار حديث لجميع الأسر سيكون أيضاً مكلفاً للغاية. وفي ضوء هذه الاعتبارات، تم تطوير طريقة جديدة من طرق المعاينة للتغلب على الصعوبات والمشاكل المتعلقة بالإطار، وذلك بتقسيم المجتمع إلى مجموعات جزئية، ثم تقسيم هذه المجموعات إلى مجموعات جزئية أخرى.... وهكذا حتى يتم الوصول إلى الوحدات المرغوب تغطيتها، وتحديث الأخيرة وذلك بوقت قصير وتكلفة مقبولة، على أن يتم سحب هذه الاجزاء في كل مرحلة بالطريقة العشوائية.

وقد سميت هذه الطريقة بطريقة المعاينة المتعددة المراحل. حيث يتم سحب وحدات المعاينة على عدة مراحل. ففي المرحلة الأولى يتم تقسيم المجتمع إلى وحدات كبيرة نسبياً (مناطق إدارية، أو جغرافية)، ويتم اختيار بعض هذه الأجزاء بالطريقة العشوائية وتسمى الوحدات التي يتم اختيارها في المرحلة الأولى بوحدات المعاينة الأولية (PSU_S) Primary Sampling Units. وفي المرحلة الثانية يتم تقسيم الوحدات الأولية التي اختيرت في المرحلة الأولى إلى وحدات مناسبة، كوحدات العد التي استخدمت في التعداد السابق على سبيل المثال، أو أي تقسيم آخر مناسب. ثم يتم اختيار عدد محدد من هذه الأجزاء بالطريقة العشوائية أيضاً. وتسمى الوحدات التي اختيرت في هذه المرحلة بوحدات المعاينة في المرحلة الثانية، أو وحدات المعاينة الثانوية ($SSUs$) Secondary Sampling Units) وهكذا، ويتم في المرحلة الثالثة تقسيم الوحدات التي اختيرت في المرحلة الثانية (SSU_S) إلى وحدات مناسبة، واختيار عدداً محدداً منها في المرحلة الثالثة وهكذا في المرحلة الرابعة حتى المرحلة الأخيرة، حيث يتم سحب وحدات المعاينة النهائية. ويمكن استخدام طريقة المعاينة الطبقية أو العنقودية أو الطريقتين معاً في كل مرحلة من مراحل العينة المتعددة المراحل. ففي المسوح الخاصة بالأسر، على سبيل المثال، يمكن تقسيم المجتمع إلى طبقات مثل الحضر، والريف، أو الأقاليم، وتقسيم كل طبقة إلى وحدات مناسبة، مثل الأحياء في المدن في المناطق الحضرية والقرى في المناطق الريفية، في المرحلة الأولى، وفي المرحلة الثانية يمكن تقسيم الأحياء والقرى إلى وحدات أصغر، كوحدات العد التي استخدمت في التعداد الأخير. ويمكن الاعتماد على طريقة المعاينة العشوائية البسيطة أو المنتظمة لاختيار الوحدات الأولية في كل طبقة من الطبقات التي قسم إليها المجتمع في المرحلة الأولى، أو الوحدات الثانوية في المرحلة الثانية، وفي المرحلة الثالثة يمكن تغطية جميع الأسر في الوحدات التي اختيرت في المرحلة الثانية، أو اختيار عدد محدد من كل وحدة بالطريقة العشوائية، فالعينة في هذا المثال هي عينة طبقية عنقودية ذات ثلاث مراحل.

1.5 العينة الأساسية (Master Sample)

يعتبر تعداد المساكن والسكان مناسبة لتوفير أطر حديثة وشاملة للمباني والوحدات السكنية والأسر بتاريخ التعداد، ويمثل الإطار الذي يوفره التعداد أساساً لتصميم عينات المسوح الاجتماعية والديموغرافية المختلفة في الفترة التالية للتعداد، إلا أنه وبسبب تطور الحركة العمرانية وحركة السكان المستمرة والهجرة، لا تلبث أن تصبح هذه الأطر قديمة بعد انقضاء عامين أو ثلاثة أعوام على الأكثر على تنفيذ التعداد، وتصبح هذه المسألة أكثر تعقيداً مع مرور الوقت، بحيث يصبح من غير الممكن تصميم عينات ممثلة للمجتمع قبل تحديث إطار التعداد مع ما تطلبه هذه العملية من جهد ووقت ومال. ولهذا السبب اعتمدت الأمم المتحدة نهج العينة الأساسية في الثمانينات من القرن الماضي، في الوثيقتين الصادرتين عن المكتب الإحصائي في الأمم المتحدة، الأولى بعنوان "البرنامج المتكامل للمسح الوطني الأسري" الصادرة عام 1986 (National Household Survey Capability Program) والثانية بعنوان "دليل مسوح الأسر" الصادرة عام 1987 (Handbook of Household Surveys). ويقوم المبدأ الأساسي لهذا النهج على تصميم عينة ذات حجم مناسب اعتماداً على أحدث إطار للأسر، وتسمى **العينة الأساسية (Master Sample)**. ويمكن الاعتماد على هذا النهج في الفترات البعيدة عن تاريخ إجراء التعداد لإنشاء أطر مناسبة لتصميم العينة الأساسية، وسيكون الإطار الخاص بالعينة الأساسية بديلاً عن الإطار الشامل الذي وفره التعداد، وتستخدم هذه العينة لتصميم عينات جميع المسوح الأسرية خلال فترة زمنية محددة، على أن يتم تحديث وإصلاح إطار هذه العينة كل سنتين أو ثلاث سنوات على الأكثر لإدخال التعديلات التي قد تكون حدثت بعد انقضاء هذه الفترة. وقد تم اعتماد هذا النهج وتنفيذه بنجاح في العديد من دول العالم. وتجمع هذه المنشورات والممارسات العملية على استخدام أسلوب العينة الطبقيّة العنقودية المتعددة المراحل (مرحلتين أو أكثر) لتصميم العينة الأساسية. فيتم في البداية تقسيم المجتمع المستهدف إلى فئتين أو أكثر تسمى كل منها طبقة، كطبقة سكان الريف وطبقة سكان الحضر أو طبقة السكان المواطنين وطبقة غير المواطنين، وغيرها من الأمثلة، وفي المرحلة الثانية يتم سحب عينة عشوائية من الوحدات (ذات حجم محدد من السكان) التي

تُشكل في كل طبقة والتي تسمى كل منها وحدة عد أو عنقود، وتعتبر وحدات العد (Enumeration Areas) (EAs) التي استخدمت في أحدث تعداد للسكان أكثر استخداماً في التطبيقات العملية، والتي يتراوح حجمها عادة بين 100-200 أسرة تقريباً، ويمكن استخدام أي طريقة أخرى مناسبة شريطة أن يكون حجم وحدة العد محدوداً، بحيث يكون انتشار وحدات العينة كبيراً في الطبقة لزيادة تمثيل العينة، وتقليل تكلفة التنقل ضمن وحدة العد. وفي المرحلة الأخيرة يتم تغطية عينة من الأسر في كل وحدة عد من وحدات العينة، وأحياناً يتم تغطية جميع الأسر في وحدة العد.

ويتم تصميم العينة الأساسية بحجم كبير من وحدات العد، بحيث تكون أساساً لتشكيل عينات جزئية غير متداخلة تلبي حاجة المسوح المتوقع تنفيذها خلال فترة عامين على الأكثر، ويتحدد عدد وحدات العد وفقاً لعدد المسوح التي ستنفذ خلال فترة سنة أو سنتين، فإذا كان من المتوقع تنفيذ خمسة مسوح، وكان حجم العينة في كل منها 1000 أسرة، على سبيل المثال، فيمكن تشكيل العينة الأساسية وفق أحد النهجين التاليين:

- **النهج الأول:** تشكيل عينة أساسية حجمها 50 وحدة عد حجم كل منها 100 أسرة، وهي كافية لسحب خمس عينات جزئية من كل وحدة عد حجم كل منها 20 أسر وغير متداخلة، لتخفيف العبء عن الأسر بعدم تكرارها في المسوح أكثر من مرة.
- **النهج الثاني:** تشكيل عينة أساسية حجمها 150 وحدة عد، وتشكل أساساً لسحب ثلاث عينات جزئية غير متداخلة حجم كل منها 50 وحدة عد، ويسمى كل منها (مكررة). ويمكن عندئذ استخدام المكررة الواحدة لسحب عينات جزئية غير متداخلة أيضاً لأكثر من مسح، كما في النهج الأول. وفي هذه الحالة يمكن تنفيذ خمسة مسوح غير متداخلة لكل مكررة.

1.6- العينة على مرحلتين (Two phases sample)، ينصب الاهتمام في بعض المسوح على ظواهر نادرة في

المجتمع المستهدف مثل المعاقين، او ذوي الدخل العالية وغيرها، وعلى الرغم من أن حجم الفئة المستهدفة قد يكون صغيراً إلا أن الوصول إلى هذا العدد يتطلب تكلفة كبيرة لتغطية عدد كبير من الأسر، وبهدف تقليل التكلفة يتم تنفيذ مسح استطلاعي على عينة كبيرة من الأسر يتضمن عدد محدود من الأسئلة الخاصة بالفئة المستهدفة، وقد يتم ذلك عن طريق إلحاق هذه البنود بأي من المسوح الأخرى الجاري تنفيذها، وتصنيف الوحدات التي توجد بها الفئات المستهدفة، ولاحقاً في المرحلة الثانية يتم اختيار العينة من بين هذه الوحدات، ولهذا السبب سمي هذا النهج بالعينة على مرحلتين، وتسمى أحياناً العينة المزدوجة. فإذا كان حجم العينة المقرر 1000 معاق، لمسح خاص بالمعاقين على سبيل المثال، وكان من المقرر تنفيذ مسحاً للقوى العاملة حجم عينته 15000 أسرة عندئذ يمكن أن يدرج سؤال واحد فقط حول وجود معاق في الأسرة من عدمه وهذا لا يشكل عبئاً على استبيان المسح، وبفرض تم تصنيف 1500 معاق بنتيجة هذا المسح، عندئذ يتم اختيار عينة ذات حجم 1000 معاق، لجمع البيانات وفق استبيان تفصيلي خاص بالمعاقين.

1.7 العينات المتكررة (Rotation Samples)، قد يكون الهدف من الاستقصاء تقدير التغير أو دراسة الاتجاه لبعض الظواهر، ويتم ذلك عن طريق المسوح المتكررة أو المستمرة لتحقيق ذلك، حيث يتم تكرار المسح خلال فترات متعاقبة سنوية أو ربع سنوية، ووفقاً للمنشورات التي سبق ذكرها وفي ضوء الممارسات العملية، ينصب الاهتمام في هذا النوع من المسوح على هامش الخطأ المتوقع لتقدير الفرق بين تقديرات دورة وأخرى، ويتعلق ذلك بخيارات تكرار العينة التي تنحصر في ثلاثة طرق:

- تكرار نفس عينة المسح السابق، وهو النهج الأمثل.
- تكرار جزئي للعينة السابقة 50% على سبيل المثال، واستكمال العينة بعينة جزئية جديدة.
- تكرار المسح على عينة جديدة.

2- حجم العينة

يعتبر تحديد حجم العينة الخطوة الأولى في التحضير والاعداد لتنفيذ المسح، إذ يتوقف على القرار الخاص بحجم العينة تقرير العديد من المسائل المتعلقة بالنتائج المتوقعة من المسح من حيث الدقة، وأخطاء العينة، ومستويات الثقة المتوقعة بالنتائج، ويتحدد على هذا الأساس أيضاً المستلزمات المادية والبشرية، والزمن اللازم لتنفيذ المسح.

2.1 حجم العينة العشوائية البسيطة (مع الإرجاع) لتقدير المتوسط والنسبة

في التطبيقات العملية يتم التركيز على حساب حجم العينة لتقدير متوسط أو نسبة ظاهرة في المجتمع المستهدف، مثل حساب حجم العينة لتقدير متوسط الانفاق أو الدخل السنوي للأسرة أو الفرد، أو لتقدير الإنتاج في القطاعات الاقتصادية، أو لتقدير نسبة البطالة بين السكان، أو نسبة الرضا في استطلاعات الرأي وغيرها. وتعتبر العينة العشوائية البسيطة، كما هو معروف، الأساس الذي بنيت عليه طرق العينات الأخرى، وهي تطبيق مباشر لنظرية النزعة المركزية (CLT) التي تم عرضها سابقاً.

• حجم العينة (مع الإرجاع) لتقدير المتوسط

يهدف التمييز بين حجم العينة مع الإرجاع وبدون إرجاع سيتم الرمز بـ n_0 لحجم العينة مع الإرجاع، e_0 لهامش الخطأ مع الإرجاع. إن العلاقة التي تربط بين حجم العينة ومربع هامش الخطأ في العينة العشوائية البسيطة (مع الإرجاع) وفقاً لنظرية النزعة المركزية، كما سبق عرضه في الفقرة 1.1 السابقة، هي العلاقة التالية:

$$e^2_0 = Z^2 * \frac{\sigma^2}{n_0}$$

وبذلك يحسب حجم العينة مع الإرجاع لتقدير المتوسط بالعلاقة التالية

$$n_0 = Z^2 * \frac{\sigma^2}{e_0^2} \quad (1)$$

حيث:

- n_0 حجم العينة العشوائية البسيطة مع الإرجاع

- Z القيمة المقابلة لمستوى الثقة (Confidence Level) المعتمد في جدول التوزيع الطبيعي، وتساوي

1.96 عند مستوى ثقة 95% المعتمدة في أغلب المسوح الإحصائية.

- σ^2 تباين (Variance) الظاهرة المدروسة في المجتمع (أفراد، أسر، منشآت، وغيرها) ويرمز له أحياناً

بالرمز V أو الرمز S^2 ويحسب بالعلاقة المعروفة التالية

$$V = S^2 = \sigma^2 = \frac{\sum_{i=1}^{i=N} (X_i - \bar{X})^2}{N}$$

حيث N حجم المجتمع المستهدف، \bar{X} متوسط قيم الظاهرة المدروسة X_1, X_2, \dots, X_N في

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

المجتمع المستهدف، ويحسب بالعلاقة التالية

- e_0 هامش الخطأ (Margin of Error) المتوقع في تقدير متوسط المجتمع، ويعبر عن إجمالي الفرق

المتوقع بين القيمة الحقيقية في المجتمع المستهدف والقيمة المماثلة المقدرة عن طريق العينة، وقد يكون

هذا الفرق موجباً أو سالباً، ويحدد هامش الخطأ بدلالة Z وبما يسمى الخطأ المعياري (Standard Error)

(SE) بالعلاقة التالية

$$e_0 = \pm Z * SE$$

ويعرف الخطأ المعياري بأنه الانحراف المعياري لتباين التقدير (المتوسط أو النسبة) وفقاً لنظرية النزعة المركزية. ويحدد هامش الخطأ في التطبيقات العملية، بحيث تكون النسبة المئوية بين هامش الخطأ والتقدير لا تزيد عن 15%، وهي النسبة المقبولة لتحليل النتائج، وتلبية حاجة مستخدمي البيانات، مع الإشارة إلى اعتماد هامش خطأ أعظمي 20% أحياناً في بعض المسوح. وتسمى النسبة بين هامش الخطأ والتقدير (المتوسط أو النسبة) هامش الخطأ النسبي (Margin of Relative Error) ويرمز له بالرمز $MoRel$

ويحسب بالعلاقة $MoRel_{(\bar{x})} = \frac{e}{\bar{x}}$ لهامش الخطأ النسبي للمتوسط، والعلاقة المشابهة

$MoRel_{(p)} = \frac{e}{p}$ لهامش الخطأ النسبي للنسبة. وبشكل مشابه يمكن حساب النسبة بين الخطأ المعياري

(SE) (Standard Error) والمتوسط أو النسبة. وتسمى الخطأ المعياري النسبي Relative Standard

(Error)، ويرمز له بالرمز Rel ويحسب بالعلاقة $Rel_{(\bar{x})} = \frac{SE}{\bar{x}}$ للمتوسط والعلاقة $Rel_{(p)} = \frac{SE}{p}$

للنسبة. فإذا كان مستوى الثقة 95% فينبغي ألا يزيد الخطأ المعياري النسبي عن 7.5% (0.15/1.96) وفي حالات نادرة 10% (0.20/1.96).

ويمكن ملاحظة أن حجم العينة مع الإرجاع لا يتعلق بحجم المجتمع، ويتناسب طردياً مع تباين الظاهرة ومستوى الثقة، ويتناسب عكساً مع مربع هامش الخطأ.

مثال (1): إذا كان تباين ظاهرة $V = \sigma^2 = 4000$ في مجتمع حجمه 1000 ومتوسط هذه الظاهرة من

التعداد الأخير أو من دراسات سابقة $\bar{X} = 150$ وكان هامش الخطأ المتوقع 5% من قيمة المتوسط

بمستوى ثقة 95%، أي أن هامش الخطأ $e_0 = .05 * \bar{X} = 7.5$ فيكون حجم العينة بتطبيق العلاقة (1)

يساوي 274

$$n_0 = Z^2 * \frac{\sigma^2}{e_0^2} = 1.96^2 * \frac{4000}{7.5^2} = 273.18$$

وتجدر الإشارة الى أن حجم العينة المحسوب، يكون من نفس نوع الظاهرة المدروسة (أسر، أفراد فئة من السكان، منشآت، وغيرها من الظواهر).

ويمكن حساب حجم العينة بدلالة الخطأ المعياري

$$n_0 = Z^2 * \frac{\sigma^2}{Z^2 * SE^2} = \frac{\sigma^2}{SE^2} = \frac{4000}{\left(\frac{7.5}{1.96}\right)^2} = \frac{4000}{3.8265^2} = 273.18$$

$$Rel_{(\bar{x})} = \frac{e}{\bar{x}} = \frac{7.5}{150} = 0.05 \text{ أو بدلالة هامش الخطأ النسبي الذي يساوي } 0.05$$

$$n_0 = Z^2 * \frac{\sigma^2}{Rel^2(\bar{x}) * \bar{X}^2} = 1.96^2 * \frac{4000}{0.05^2 * 150^2} = 273.18$$

أو يحسب بدلالة معامل الاختلاف (Coefficient of Variation) CV حيث \bar{X} متوسط الظاهرة ومعامل

$$CV = \frac{\sigma}{\bar{X}} = \frac{\sqrt{4000}}{150} = 0.42164 \text{ أو الذي يساوي } CV = \frac{\sigma}{\bar{X}}$$

$$n_0 = Z^2 * \frac{CV^2(\bar{x})}{Rel^2(\bar{x})} = 1.96^2 * \frac{0.42164^2}{0.05^2} = 273.18$$

والنتيجة واحدة في جميع الحالات.

• حجم العينة (مع الإرجاع) لتقدير النسبة

وينطبق العرض السابق على حساب حجم العينة لتقدير النسبة (Proportion)، لأن نسبة ظاهرة هي عبارة عن حالة خاصة من المتوسط حيث أن أية قيمة في المجتمع إما أن تكون 1 عندما توافق الظاهرة ذات الاهتمام، أو تساوي الصفر عندما تكون غير ذلك، وتحسب النسبة P بعلاقة المتوسط السابقة بحيث يكون البسط يساوي عدد الحالات التي توافق الظاهرة ويكون المقام يساوي عدد الحالات في المجتمع (الموافقة للظاهرة وغير الموافقة)، وتكون النسبة المتممة Q تساوي $Q = 1 - P$ ويكون تباين النسبة وفقاً

$$V = \sigma^2 = P * Q \text{ (Bernoulli) لدستور}$$

وبالتبديل في علاقة مربع هامش الخطأ السابقة، يمكن إيجاد العلاقات المشابهة التالية لحساب حجم العينة لتقدير النسبة بدلالة هامش الخطأ

$$n_0 = Z^2 * \frac{P * Q}{e_0^2} \quad (2)$$

مثال (2): إذا كان المرغوب حساب حجم العينة لتقدير نسبة ظاهرة، وكانت نسبة هذه الظاهرة من دراسات سابقة $P = 0.30$ (نسبة المدخنين بين السكان في فئة العمر 18-40 على سبيل المثال)، وكان هامش الخطأ المتوقع 10% من النسبة (هامش الخطأ النسبي) بمستوى ثقة 95%، أي أن هامش الخطأ $e_0 = 0.10 * 0.30 = 0.03$ فيكون حجم العينة مع الإرجاع بدلالة هامش الخطأ بتطبيق العلاقة (2) يساوي 897 فرداً في الفئة العمرية المستهدفة.

$$n_0 = Z^2 * \frac{P * Q}{e_0^2} = 1.96^2 * \frac{0.30 * (1 - 0.30)}{0.03^2} = 896.37$$

2.2- حجم العينة العشوائية البسيطة (بدون إرجاع) لتقدير المتوسط والنسبة

• حجم العينة (بدون إرجاع) لتقدير المتوسط

إن العلاقة التي تربط بين حجم العينة ومربع هامش الخطأ في العينة العشوائية البسيطة (بدون إرجاع)، وفقاً لنظرية النزعة المركزية، كما سبق عرضه في الفقرة 1.1 السابقة، هي العلاقة التالية:

$$e^2 = Z^2 * \frac{\sigma^2}{n} * \frac{N - n}{N - 1}$$

وفي التطبيقات العملية وبهدف المقارنة، ولتسهيل الحسابات، يتم عادة الربط بين حجم العينة مع الإرجاع وحجم العينة بدون إرجاع، وفقاً لهامش خطأ ومستوى ثقة واحد، وبتبديل مربع هامش الخطأ في العلاقة الأخيرة بهامش الخطأ مع الإرجاع تصبح

$$Z^2 * \frac{\sigma^2}{n_0} = Z^2 * \frac{\sigma^2}{n} * \frac{N - n}{N - 1} \Leftrightarrow \frac{1}{n_0} = \frac{1}{n} * \frac{N - n}{N - 1} = \frac{1}{n} * \frac{N}{N - 1} - \frac{1}{N - 1}$$

وبضرب طرفي المعادلة الأخيرة بـ $\frac{N - 1}{N}$ تصبح

$$\frac{1}{n_0} * \frac{N - 1}{N} = \frac{1}{n} - \frac{1}{N} \Leftrightarrow \frac{1}{n_0} * \frac{N - 1}{N} + \frac{1}{N} = \frac{1}{n}$$

ومن العلاقة الأخيرة، يحسب حجم العينة بدون إرجاع بالعلاقة التالية:

$$n = \frac{1}{\frac{1}{n_0} * \frac{N - 1}{N} + \frac{1}{N}}$$

وبضرب البسط والمقام في الطرف الثاني بـ n_0 تصبح العلاقة

$$n = \frac{n_0}{\frac{N-1}{N} + \frac{n_0}{N}} \quad (3)$$

وهي العلاقة العامة لحجم العينة بدون إرجاع وعلاقته بحجم العينة مع الإرجاع. ويلاحظ أن حجم العينة

يتعلق بحجم المجتمع المستهدف N على خلاف حجم العينة مع الإرجاع.

فحجم العينة بدون إرجاع لتقدير المتوسط، في المثال (1) بتطبيق العلاقة (3)، يساوي 216، وهو أصغر من

حجم العينة مع الإرجاع الذي بلغ 274، وهي الحالة العامة من أجل نفس الشروط.

$$n = \frac{n_0}{\frac{N-1}{N} + \frac{n_0}{N}} = \frac{274}{\frac{1000-1}{1000} + \frac{274}{1000}} = 215.24$$

وعندما يتم التعامل مع مجتمعات كبيرة الحجم، وهي الحالة الشائعة في التطبيقات العملية، فإن النسب

تكون قريبة من الواحد الصحيح، وفي هذه الحالة يمكن كتابة العلاقة التالية لحجم العينة بدون

إرجاع، ويمكن تسميتها (العلاقة المختصرة للعينات الكبيرة).

$$n = \frac{n_0}{1 + \frac{n_0}{N}} \quad (4)$$

وبتطبيق العلاقة المختصرة للمجتمعات الكبيرة الحجم يكون حجم العينة يساوي 216 ايضاً، لأن

$$\frac{N-1}{N} = 0.999 \text{ يساوي } 1 \text{ تقريباً}$$

$$n = \frac{n_0}{1 + \frac{n_0}{N}} = \frac{274}{1 + \frac{274}{1000}} = 215.07$$

وعندما تكون النسبة $\frac{n_0}{N}$ أصغر من 5% ويتحقق ذلك في المجتمعات الكبيرة جداً، فإن الفرق بين حجم

العينة مع الإرجاع وبدون إرجاع يكون صغيراً، فإذا كان حجم المجتمع في هذا المثال يساوي 100000 فإن

$$\text{حجم العينة بدون إرجاع يساوي } 274 \text{ أيضاً، (} n = \frac{n_0}{1 + \frac{n_0}{N}} = \frac{274}{1 + \frac{274}{100000}} = 273.25 \text{) وهو حجم العينة مع}$$

الإرجاع المحسوب سابقاً، لأن النسبة $\frac{n_0}{N}$ صغيرة جداً وأقل من 5% بكثير.

• حجم العينة (بدون إرجاع) لتقدير النسبة

وبنفس الخطوات السابقة لحساب حجم العينة للمتوسط، يمكن إيجاد علاقة مطابقة للعلاقتين الأخيرتين

(العامة والمختصرة) لحساب حجم العينة للنسبة، آخذين بالاعتبار، أن n_0 يمثل حجم العينة (مع الإرجاع)

لتقدير النسبة، والذي يعطى بالعلاقة (2) السابقة

$$n_0 = Z^2 * \frac{pQ}{e_0^2}$$

ويكون حجم العينة (بدون إرجاع) للنسبة يعطى بالعلاقة العامة التالية

$$n = \frac{n_0}{\frac{N-1}{N} + \frac{n_0}{N}} \quad (5)$$

أو العلاقة المختصرة للمجتمعات الكبيرة الحجم

$$n = \frac{n_0}{1 + \frac{n_0}{N}} \quad (6)$$

فحجم العينة بدون إرجاع بتطبيق العلاقة المختصرة للعينات الكبيرة في المثال (2) السابق يساوي 473 وهو أصغر من حجم العينة مع الإرجاع الذي بلغ 897، وهي الحالة العامة، وبالتالي فإن تكلفة المسح أقل.

$$n = \frac{n_0}{1 + \frac{n_0}{N}} = \frac{897}{1 + \frac{897}{1000}} = 472.85$$

ومن جهة أخرى في الواقع العملي لا يتم تغطية حجم العينة كاملاً بسبب عدم الاستجابة، ولتجنب ذلك ينبغي زيادة حجم العينة في جميع العلاقات السابقة بضربها بمعامل يسمى معامل عدم الاستجابة k الذي يساوي مقلوب نسبة الاستجابة المتوقعة $\frac{1}{RR}$ حيث RR نسبة الاستجابة (Response Rate)، ويستفاد عادة من المسوح السابقة لتقدير نسبة الاستجابة المتوقعة في المسح قيد التنفيذ. وبناء على ذلك ينبغي ضرب حجم العينة المحسوب بالعلاقات السابقة بمعامل عدم الاستجابة للحصول على حجم العينة النهائي.

فإذا كانت نسبة عدم الاستجابة المتوقعة في المثال (1) تساوي 10% فإن نسبة الاستجابة تكون 90%

ويكون معامل عدم الاستجابة يساوي $k = \frac{1}{RR} = \frac{1}{0.90}$ وبذلك يكون حجم العينة النهائي، مع الأخذ

بالاعتبار عدم الاستجابة، يساوي حجم العينة المحسوب سابقاً مضروباً بمعامل عدم الاستجابة أي إن حجم العينة ينبغي أن يكون 304 للحصول على حجم العينة المطلوب 274 (304*0.90).

$$n = n_0 * k = Z^2 * \frac{\sigma^2 * k}{e_0^2} = 1.96^2 * \frac{4000}{7.5^2} * \frac{1}{0.90} = 303.5$$

$$\frac{e}{\bar{X}} = e_0 \Leftrightarrow e = e_0 * \bar{X}$$

$$n = Z^2 * \frac{\sigma^2}{e_0^2 * \bar{X}^2}$$

وبالتبديل تصبح علاقة حجم العينة لتقدير المتوسط

$$CV^2 = \frac{\sigma^2}{\bar{X}^2}$$

وكما هو معروف، إن مربع معامل الاختلاف للمتوسط

$$n = Z^2 * \frac{CV^2}{e_0^2}$$

وبالتبديل تصبح علاقة حجم العينة لتقدير المتوسط

$$\text{وباعتبار } a^2 = \frac{Z^2}{e_0^2} \text{ تصبح العلاقة الأخيرة}$$

$$n = a^2 CV^2 \quad (8)$$

وهي دالة لحجم العينة لتقدير المتوسط باعتبار معامل الاختلاف هو المتحول المستقل $CV \geq 0$ ، وتمثل معادلة قطع مكافئ أيضاً، له نهاية صغرى عندما يكون معامل الاختلاف يساوي الصفر. وبهدف المقارنة تم التمثيل البياني لجزء منه في المجال $0 \leq CV \leq 1$ وعندما تكون $Z = 1.96$ و $e_0 = 0.05$ أيضاً كما في

حالة النسبة، كما مبين في الشكل أدناه. ويمكن الحصول على العديد من الأشكال المشابهة لهذا الشكل

بتغيير هامش الخطأ ومستوى الثقة، ويمكن ملاحظة الخصائص والنتائج التالية:

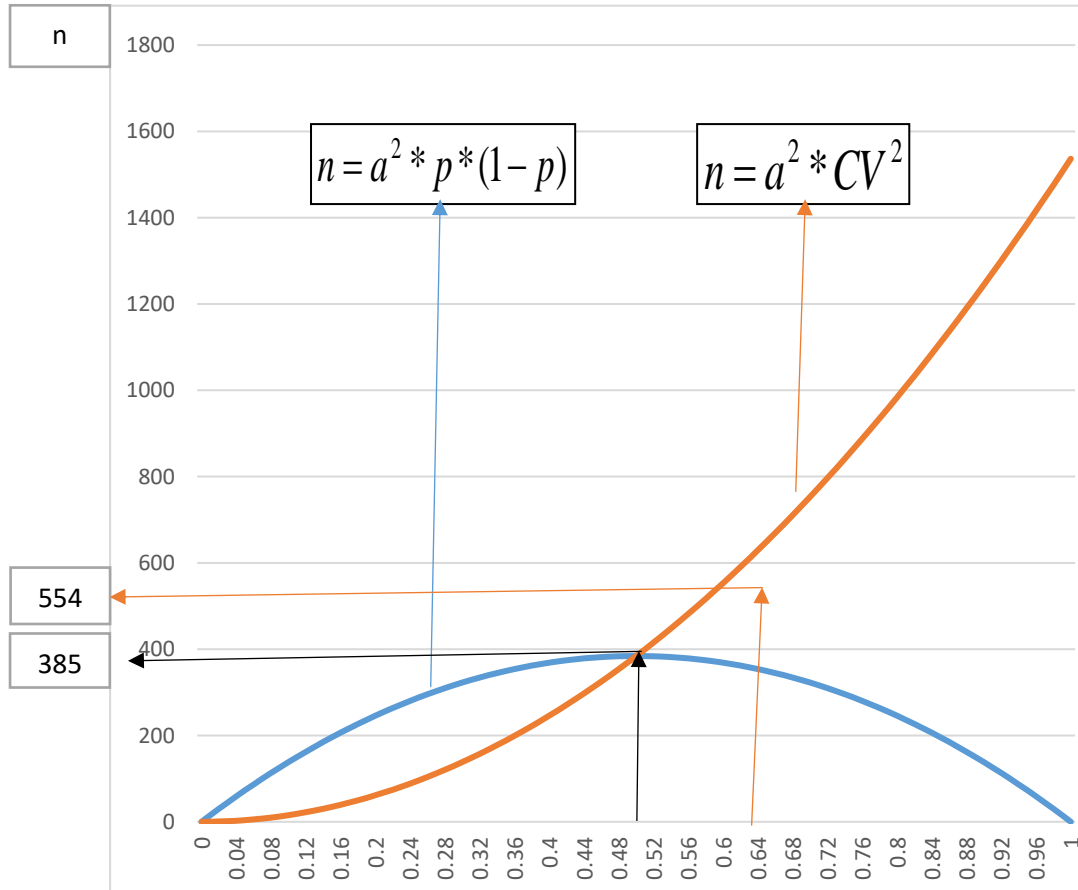
1- إن حجم العينة لتقدير النسبة المحسوب بالعلاقة (7) يبلغ نهايته العظمى ويساوي 385 عندما تكون النسبة $P = 0.50$ وسيكون تباين النسبة أعظمي $V = P * (1 - P) = 0.25$ فإذا كانت النسبة تساوي 0.20 فإن حجم العينة الموافق يساوي 264 وإذا كانت النسبة 0.60 فإن حجم العينة الموافق يساوي 369

وكلاهما أصغر من 385 وينطبق ذلك على جميع النسب الأخرى. فيكون حجم العينة الأعظمي مناسب لتقدير جميع النسب التي هي أصغر من أو أكبر من 0.50 من أجل نفس هامش الخطأ ومستوى الثقة. ويمكن ملاحظة أن حجم العينة المقابل للنسبة 0.40 يساوي حجم العينة المقابل للنسبة 0.60 وكل منهما يساوي 369 لأن تباين النسبة واحد في الحالتين $V = P * (1 - P) = 0.40 * 0.60 = 0.24$.

وتستخدم هذه النتيجة على نطاق واسع في التطبيقات العملية وخاصة في حساب حجم العينة لاستطلاعات الرأي، حيث يكون الهدف من الاستطلاع غالباً تقدير العديد من المؤشرات النسبية، وتعتمد العديد من مؤسسات استطلاعات الرأي الدولية هذه النتيجة لحساب حجم عينة الاستطلاع.

2- إذا كان معامل الاختلاف للمتوسط يساوي نفس القيمة أيضاً $CV = 0.50$ يكون حجم العينة المحسوب بالعلاقة (8) يساوي 385 ويكون حجم العينة واحد للنسبة والمتوسط، وفي جميع الحالات التي يكون فيها معامل الاختلاف أقل من 0.50 يكون حجم العينة الأعظمي للنسبة أكبر من حجم العينة للمتوسط، فحجم العينة للمتوسط من أجل $CV = 0.40$ ومستوى ثقة 95% وهامش خطأ 0.05 يساوي 246، وهو أصغر من حجم العينة الأعظمي للنسبة. وهي الحالة العامة، شريطة أن يكون معامل الاختلاف للمتوسط أصغر من 0.50. وسيكون حجم العينة للنسبة ملائماً لتقدير جميع النسب، كما تم توضيحه سابقاً، وهو أيضاً مناسباً لتقدير جميع المتوسطات التي معامل اختلافها أصغر من 0.50. وفي هذه الحالة يكفي حساب حجم العينة لتقدير النسبة بالعلاقة (7). ولهذه النتيجة أهمية كبيرة في التطبيقات العملية. فإذا كان

معامل الاختلاف أصغر أو يساوي 50% فإن حجم العينة الأعظمي المحسوب لتقدير النسبة سيكون مناسباً لتقدير المتوسط أيضاً. وعندئذ يمكن الاستغناء عن تقدير تباين الظاهرة (المتوسط أو النسبة) في المجتمع المستهدف والذي لا يتوفر في أغلب الحالات، لأن حجم العينة الموافق للنسبة التي تساوي 50% هو حجم أعظمي ومناسب لتقدير النسبة والمتوسط معاً.



الشكل (1)

3- ومن جهة أخرى، عندما يكون معامل الاختلاف للمتوسط أكبر من 0.50، سيكون حجم العينة لتقدير المتوسط أكبر من حجم العينة الاعظمي لتقدير النسبة، وسيكون عندئذ حجم العينة المحسوب لتقدير المتوسط بالعلاقة (8) مناسباً لتقدير جميع المؤشرات النسبية. فإذا كان معامل الاختلاف للمتوسط 60% فإن حجم العينة لتقدير المتوسط يساوي 554، وهو أكبر من حجم العينة لتقدير النسبة الذي يساوي 396.

4- إن الخاصة (3) مناسبة لتقدير متوسط واحد لإحدى الظواهر، وقد لا يتناسب حجم العينة لتقدير متوسطات أخرى ذات الاهتمام، ومن الواضح عدم وجود حجم عينة أعظمي لتقدير جميع المتوسطات كما في حجم العينة الأعظمي لتقدير جميع النسب، لأن دالة توزيع العينة للمتوسط في العلاقة (8) متزايد دوماً، ونظراً لتعذر حساب حجم العينة لجميع المتوسطات المرغوب تقديرها بسبب صعوبة الحصول على تباين الظواهر الخاصة بكل منها، فيمكن تقسيم المجتمع المستهدف إلى طبقات متجانسة، بحيث يكون معامل الاختلاف للمتوسطات ذات الاهتمام أصغر أو يساوي 0.50 في جميع الطبقات، وعندئذ يمكن تطبيق الخاصة (1) في كل طبقة، باعتماد حجم العينة الأعظمي في كل طبقة لتقدير النسبة.

2.4- حجم العينة للمسوح الأسرية

في العديد من المسوح الاجتماعية والديموغرافية يكون المجتمع المستهدف فئة محددة من السكان (أفراد من القوى العاملة في مسح البطالة، الأطفال دون سن الخامسة، النساء في سن الانجاب 15-49 وغيرها من الأمثلة) ، ويكون حجم العينة العشوائية البسيطة (مع الإرجاع) المحسوب لتقدير النسبة أو المتوسط بالعلاقين (1) و(2) يعبر عن حجم العينة الخاص بالظاهرة المدروسة وهو من نفس أفراد الفئة المستهدفة، فإذا كان المرغوب على سبيل المثال تقدير نسبة البطالة فإن حجم العينة يكون أفراداً من القوى العاملة، وإذا كان المرغوب تقدير نسبة إحدى العادات الصحية لدى الشباب فإن حجم العينة يكون أفراداً من الشباب وغيرها من الأمثلة. وغالباً لا تتوفر أطر مناسبة للفئة المستهدفة لسحب وحدات العينة المطلوبة، بينما تتوفر أطر للأسر من التعداد السابق أو من مصادر أخرى كالسجلات الإدارية، ويمكن الاعتماد على الأسرة كوحدة إحصائية قبل النهائية لسحب عينة الأسر المقابلة لحجم عينة الأفراد للفئة المستهدفة من السكان، وعندئذ ينبغي تحويل حجم عينة الأفراد إلى أسر، لذا ينبغي معرفة نسبة الظاهرة بين السكان \bar{N} ومتوسط حجم

الاسرة \bar{n} فيكون حجم عينة الأسر يساوي حجم عينة الأفراد مقسوماً على $r * \bar{n}$ (والذي يدل على حجم الظاهرة المدروسة في الأسرة الواحدة).

وبالتعويض عن ذلك في العلاقتين (1) و (2) السابقتين مع الأخذ بالاعتبار معامل عدم الاستجابة k تصبحان على الشكل التالي:

حجم عينة الأسر مع الإرجاع المقابل لعينة الأفراد لتقدير المتوسط يحسب بالعلاقة التالية:

$$n = Z^2 * \frac{\sigma^2 * k}{e^2_0 * r * \bar{n}} \quad (9)$$

والعلاقة المشابهة لحجم عينة الأسر مع الإرجاع لتقدير النسبة

$$n = Z^2 * \frac{p * Q * k}{e^2_0 * r * \bar{n}} \quad (10)$$

ولحساب حجم العينة (بدون إرجاع) يتم تطبيق العلاقات من (3) إلى (6) العامة أو المختصرة للمجتمعات الكبيرة.

مثال (3): إذا كان حجم العينة المحسوب بالمثال (1) يمثل 274 فرداً من فئة محددة نسبتها 10% من السكان، أي أن $r = 0.10$ وإذا كان متوسط حجم الأسرة $\bar{n} = 5$ وكانت نسبة عدم الاستجابة 15%، أي ان نسبة الاستجابة المتوقعة 85% فإن حجم عينة الأسر بتطبيق العلاقة (9) يساوي 643 أسرة.

$$n = Z^2 * \frac{\sigma^2 * k}{e^2_0 * r * \bar{n}} = 1.96^2 * \frac{4000}{7.5^2 * 0.10 * 5 * 0.85} = 642.8$$

مثال (4) إذا كانت نسبة الفئة العمرية المستهدفة في المثال (2) بين السكان تساوي 40% وكان متوسط حجم الأسرة 5 أفراد ونسبة عدم الاستجابة 5% فإن حجم عينة الأسر بتطبيق العلاقة (10) يساوي 1820 أسرة.

$$n' = Z^2 * \frac{P * Q}{e^2_0 * r * \bar{n} * RR} = 1.96^2 * \frac{0.10 * 0.90}{0.01^2 * 0.40 * 5 * 0.95} = 1819.7$$

ويمكن حساب حجم العينة (بدون إرجاع) بتطبيق العلاقات من (3) إلى (6) العامة أو المختصرة للمجتمعات الكبيرة، بعد معرفة حجم المجتمع N فإذا كان حجم المجتمع المستهدف في هذا المثال 10000 أسرة فإن حجم العينة العشوائية البسيطة (بدون إرجاع) في المثال (3) بتطبيق العلاقة (3) يساوي 605 أسر. وهو أصغر قليلاً من حجم العينة (مع الإرجاع) بسبب أن حجم المجتمع المستهدف كبير، وهي الحالة العامة، لذا يتم الاعتماد على حجم العينة مع الإعادة في أغلب التطبيقات العملية في المجتمعات الكبيرة الحجم.

$$n = \frac{n_0}{\frac{N-1}{N} + \frac{n_0}{N}} = \frac{643}{\frac{10000-1}{10000} + \frac{643}{10000}} = 604.2$$

2.5- حجم العينة المركبة

إن حجم العينة المحسوب بالعلاقات السابقة يمثل حجم العينة العشوائية البسيطة، وفي الواقع العملي يتم الاعتماد على عدد آخر من أنواع العينات مثل الطباقية أو العنقودية أو المتعددة المراحل، أو الاعتماد على بعض منها أو جميعها بآن واحد وهذا ما يسمى نهج العينات المركبة. وتوفر نظرية العينات العلاقات الرياضية الخاصة بحساب حجم العينة لكل نوع من هذه العينات كما في العلاقات الرياضية لحساب حجم العينة العشوائية البسيطة التي تم عرضها في الفقرات السابقة، إلا أن التعقيدات المتعلقة بتطبيق هذه العلاقات تحول دون استخدامها في التطبيقات العملية على نطاق واسع. ويستعاض عن ذلك في التطبيقات

العملية بحساب ما يسمى معامل أثر التصميم (Design Effect) ويرمز له بالرمز $deff$ والذي يمثل النسبة بين تباين التقدير في إحدى العينات من الأنواع الأخرى والتباين المماثل في العينة العشوائية البسيطة لنفس حجم العينة (KISH,1965)، مع ملاحظة أن كل منهما يمثل مربع الخطأ العشوائي للتقدير. فإذا كان تباين التقدير في العينة العشوائية البسيطة V_{srs} والتباين المماثل V_s في عينة من الأنواع الأخرى فإن أثر التصميم يحسب بالعلاقة التالية:

$$deff = \frac{V_s}{V_{srs}}$$

وعلى سبيل المثال، إذا استخدمت العينة العنقودية لتقدير نسبة إحدى الظواهر وكان تباين النسبة

$V_s = 0.0025$ وكان تباين نفس النسبة لنفس الحجم في العينة العشوائية البسيطة

$V_{srs} = 0.0016$ فإن أثر التصميم $deff = \frac{0.0025}{0.0016} = 1.56$ ومن الواضح أن التباين في

العينة العنقودية أكبر من التباين في العينة العشوائية البسيطة لنفس الحجم بقيمة تساوي قيمة أثر

التصميم، ولما كان تباين التقدير يتناسب عكساً مع حجم العينة، فإن حجم العينة من الأنواع الأخرى يرتبط

بحجم العينة العشوائية البسيطة بالعلاقة التالية

$$n' = n * deff \quad (11)$$

ولهذه النتيجة أهمية كبيرة لحساب حجم العينة لأنواع أخرى بدلالة حجم العينة العشوائية البسيطة (الذي

يحسب بالعلاقات السابقة)، وبدلالة أثر التصميم الذي يحسب من مسوح سابقة، وتوصي العديد من

المنشورات باعتبار قيمة أثر التصميم 1.5 أو 2 إذا لم تتوفر قيمته سابقاً. فإذا كان حجم العينة العشوائية

البسيطة في المثال السابق 350 فإن حجم العينة العنقودية ينبغي أن يكون 546 ($350 * 1.56$) للحصول

على دقة واحدة.

2.6-توزيع العينة على الطبقات

تم في العرض السابق حساب حجم العينة لجميع أنواع العينات، لتقدير المتوسط أو النسبة على مستوى المجتمع المستهدف وفق مستوى ثقة وهامش خطأ محددين. وفي العينة الطباقية، كما هو معروف، يقسم المجتمع المستهدف إلى طبقات، وعندئذ ينبغي تحديد حجم العينة على مستوى كل طبقة. ويتم ذلك في التطبيقات العملية بإحدى الطريقتين التاليتين:

الطريقة الأولى، حساب حجم العينة على مستوى كل طبقة، باعتبار كل طبقة مجتمعاً قائماً بذاته، وعندئذ يتم حساب حجم العينة بإحدى الطرق المعروضة سابقاً وفق مستوى ثقة وهامش خطأ محددين على مستوى كل طبقة. وسيتضاعف حجم العينة بالمقارنة مع حجم العينة على مستوى المجتمع المستهدف، وقد لا يتناسب ذلك مع الموارد البشرية والمادية المتاحة.

الطريقة الثانية، حساب حجم العينة على مستوى المجتمع المستهدف، وفقاً لمستوى ثقة وهامش خطأ محددين وبما يتناسب مع الموارد المتاحة، وتوزيع حجم العينة بعدئذ على الطبقات التي عددها h . فإذا

كان N حجم المجتمع المستهدف و N_i حجم الطبقة i فإن $N = \sum N_i$ وإذا كان σ_i^2 تباين

الظاهرة ذات الاهتمام في الطبقة i فإن توزيع حجم عينة محدد n على الطبقات التي قسم إليها

المجتمع المستهدف فإن توزيع العينة يتم بإحدى طرق التوزيع التالية المقترحة في منشورات العينات.

1. **التوزيع الأمثل أو توزيع نيومان (Optimal or Neyman Allocation)**، يعتبر هذا التوزيع الأكثر

استخداماً في التطبيقات العملية، ويتم توزيع العينة على الطبقات، وفقاً لهذه الطريقة، بحيث يكون

2. تباين التقديرات المختلفة σ_{st}^2 على مستوى المجتمع المستهدف أصغر ما يمكن. ويتحقق

ذلك عندما يكون حجم العينة n_i في الطبقة i متناسباً مع الناتج $N_i * \sigma_i$ ويحسب حجم العينة في الطبقة بالعلاقة التالية:

$$n_i = n * \frac{N_i * \sigma_i}{\sum_{i=1}^h N_i * \sigma_i}$$

3. التوزيع المتناسب مع الحجم (Proportional Allocation)، ويتم توزيع العينة وفقاً لهذه

الطريقة باعتبار أن كسر المعاينة $f_i = \frac{n_i}{N_i}$ واحد في جميع الطبقات ويساوي $\frac{n}{N}$ ويصبح

حجم العينة في كل طبقة متناسباً مع حجمها بالنسبة لحجم المجتمع المستهدف، ويحسب بالعلاقة التالية، وغالباً ما يستخدم هذا التوزيع في التطبيقات العملية للسهولة. كما في المثال أدناه.

$$n_i = n * \frac{N_i}{N}$$

4. توزيع (Bankier) أو الرفع إلى قوة (Power allocation)، إن توزيع العينة على الطبقات بالتوزيع

الأمثل أو بالتناسب مع الحجم، يكون مناسباً للتقديرات المختلفة على مستوى المجتمع المستهدف وفق هامش خطأ ومستوى ثقة محددين، وعندما يكون الهدف من المسح الحصول على تقديرات ذات دقة مقبولة على مستوى كل طبقة، كما في أغلب المسوح، فإن حجم العينة في الطبقات الصغيرة الحجم وفق التوزيعين السابقين سيكون صغيراً وقد لا يتناسب مع الدقة المرغوبة للتقديرات على مستوى كل طبقة، وعندئذ يتم توزيع العينة على الطبقات بالتناسب مع

الناتج $N_i^\alpha * CV_i$ حيث $0 \leq \alpha \leq 1$ و CV_i معامل الاختلاف في الطبقة i

لإحدى الظواهر ذات الاهتمام. ويحسب حجم العينة n_i في الطبقة i بالعلاقة التالية:

$$n_i = n * \frac{N_i^\alpha * CV_i}{\sum_{i=1}^{i=h} N_i^\alpha * CV_i}$$

ويتم وفقاً لذلك تخصيص حجم عينة مناسب للطبقات صغيرة الحجم، مع بقاء حجم العينة المخصص للطبقات المتوسطة أو الكبيرة الحجم مقبولاً للتقديرات ذات الاهتمام. ويبين المثال التالي توزيع عينة حجمها 300 وحدة على ثلاث طبقات من مجتمع حجمه 850 وحدة وفقاً للتوزيعات المختلفة باعتبار $\alpha = 0.5$ وكما يبين الجدول أدناه فإن التوزيع بطريقة (Bankier) قد تم تخصيص الطبقة الأولى (صغيرة الحجم) بعينة حجمها 52 وحدة مقابل تخصيصها بعينة حجمها 21 وحدة بطريقة نيومان و 35 وحدة بطريقة التوزيع المتناسب مع الحجم.

| الطبقة | حجم الطبقة N_i | الانحراف المعياري σ_i | المتوسط \bar{x}_i | معامل الاختلاف CV_i | توزيع Neyman | التوزيع المتناسب | توزيع Bankier |
|---------|------------------|------------------------------|---------------------|-----------------------|--------------|------------------|---------------|
| 1 | 100 | 12 | 3 | 4 | 21 | 35 | 52 |
| 2 | 250 | 25 | 5 | 5 | 107 | 88 | 103 |
| 3 | 500 | 20 | 4 | 5 | 172 | 177 | 145 |
| المجموع | 850 | | | | 300 | 300 | 300 |

3- أوزان العينة

3.1- طرق سحب وحدات العينة عشوائياً واحتمالات السحب

إن الشرط الأساسي لتحقيق نظرية النزعة المركزية هو العشوائية في اختيار وحدات العينة، بحيث لا يوجد أية قيود تمنع وقوع أية وحدة في العينة، ويتم التعبير عن هذا المبدأ في نظرية الاحتمالات بسحب وحدات العينة بحيث يكون لكل وحدة مسحوبة في العينة احتمالاً مختلفاً عن الصفر، ويعتبر هذا المبدأ الأساس الذي بنيت عليه الطرق المختلفة للعينات الاحتمالية، لذا ينبغي توفر قائمة حديثة بوحدات المجتمع المستهدف بالمسح أو الاستطلاع وهو ما يسمى إطار العينة. وكما هو معروف فإن الاحتمال هو مقياس يعبر عن درجة النجاح في وقوع حدث محدد، أو يقيس النجاح في الحصول على نتيجة معينة في تجربة محددة، وتحسب قيمة الاحتمال بالنسبة بين الحالات المتوافقة مع الحدث وعدد الحالات الممكنة لحدوثه، فاحتمال سحب كرة واحدة من بين 10 كرات يساوي $\frac{1}{10}$ ، واحتمال الحصول على أحد الوجوه لدى القاء حجر النرد يساوي $\frac{1}{6}$ (لأن عدد الحالات الممكنة الحدوث 6) واحتمال الحصول على ورقة حمراء من بين أوراق اللعب يساوي $\frac{26}{52}$ (لأن عدد الحالات الموافقة 26 وعدد الحالات الممكنة 52)، وبشكل مشابه فإن احتمال سحب وحدة واحدة من مجتمع مؤلف من N وحدة يساوي $P = \frac{1}{N}$ واحتمال سحب وحدة واحدة في عينة حجمها n سحبت من هذا المجتمع يساوي $P = \frac{n}{N}$ (لأن عدد الحالات الموافقة n وعدد الحالات الممكنة N)، وغيرها من الأمثلة، وبشكل عام إذا كانت عدد الحالات الموافقة n لوقوع الحدث وكانت عدد الحالات الممكنة

$$P = \frac{n}{N}$$

فإن احتمال حدوث هذا الحدث ويرمز له بالرمز P يحسب بالعلاقة التالية:

وكما يلاحظ فإن عدد الحالات المتوافقة مع الحدث أكبر أو يساوي الصفر وهو أصغر أو يساوي عدد الحالات الممكنة، لذا فإن الاحتمال هو أكبر أو يساوي الصفر وأصغر أو يساوي 1 دوماً، ويعبر عن ذلك بالعلاقة التالية:

$$0 \leq P \leq 1$$

وفي الواقع العملي يوجد طيف واسع من الوحدات يغطي مصطلح الوحدة. فالوحدة في المسوح الإحصائية قد تكون فرداً من السكان، أو أسرة، أو منشأة، أو وحدة عد، أو حي أو قرية، وغيرها من الأمثلة، وينطبق التعريف السابق لاحتمال سحب أية وحدة من هذه الوحدات على جميع هذه الأمثلة. وتسمى طريقة سحب هذه الوحدات وفقاً لهذا المبدأ بطريقة السحب العشوائي البسيط (Simple random selection)، وتعتبر هذه الطريقة هي التطبيق المباشر لقانون الأعداد الكبيرة (LLN) ونظرية النزعة المركزية (CLT)، وتسمى العينة المسحوبة بهذه الطريقة عينة عشوائية بسيطة (SRS (Simple Random Sampling) والتي تعتبر الأساس الذي بنيت عليه جميع أنواع العينات الأخرى.

وفي التطبيقات العملية، يتم سحب وحدات العينة عشوائياً بعدة طرق، منها ما هو تقليدي كطريقة القرعة المعروفة أو طريقة جداول الأرقام العشوائية، وقد تم استخدام هاتين الطريقتين على نطاق واسع قبل استخدام الطرق الحديثة التي تعتمد على الحاسبات الشخصية والعديد من البرامج الجاهزة Minitab، SASE، EXCEL وغيرها. وسيتم عرض أهم طرق السحب العشوائي في التطبيقات العملية، واحتمالات سحب وحدات العينة.

1-طريقة السحب المنتظم (النمطي) (Systematic Sampling)

تعتبر طريقة السحب المنتظم من أكثر الطرق استخداماً في التطبيقات العملية لسحب وحدات العينة عشوائياً، ويتم ذلك بتقسيم المجتمع المستهدف الذي حجمه N إلى عدد من المجموعات يساوي حجم العينة n ، وتتحدد وحدات العينة باختيار وحدة واحدة من كل مجموعة، ويتم ذلك وفق الخطوات التالية:

الخطوة الأولى: ترقيم وحدات المجتمع من (1) إلى (N) .

الخطوة الثانية: يتم تقسيم المجتمع المستهدف إلى n من المجموعات المتساوية في الحجم، ولتحقيق ذلك يتم تقسيم حجم المجتمع على حجم العينة، ويسمى الناتج فترة السحب (Selection Interval) ويرمز له بالرمز I أي أن

$$I = \frac{N}{n}$$

وقد تكون فترة السحب عدداً صحيحاً مثل 10 أو 7 وغيرها، وقد تكون عدداً مؤلفاً من جزء صحيح وكسر عشري مثل 15.12، ويتم سحب وحدات العينة وفقاً لذلك كما يلي:

1.1-فترة السحب عدداً صحيحاً

- يتم تحديد رقماً عشوائياً محصوراً بين الواحد الصحيح والعدد الدال على فترة السحب، ويرمز له بالرمز R ويسمى البداية العشوائية، وتعتبر الوحدة التي ترتيبها في المجتمع يساوي R هي الوحدة الأولى في العينة. ويتم الحصول على ترتيب الوحدة الثانية في العينة بإضافة فترة السحب إلى البداية العشوائية، فيكون ترتيب الوحدة الثانية في العينة هو الترتيب المقابل في المجتمع. وهكذا بالنسبة للوحدة الثالثة، حتى يتم سحب عدداً من الوحدات يساوي حجم العينة. ويمكن ملاحظة أن

ترتيب كل وحدة يحسب بإضافة فترة السحب الى ترتيب الوحدة التي تسبقها، وبناء على ذلك فإن ترتيب وحدات العينة تحسب كما يلي:

$$R, R + I, R + 2 * I, \dots, R + (n - 1) * I$$

فسحب 3 طلبة من بين 15 طالباً على سبيل المثال بالطريقة المنتظمة، يتم ذلك كما يلي:

- ترقيم الطلبة من 1 على 15

- يتم حساب فترة السحب والتي تساوي $I = \frac{15}{3} = 5$

- يتم اختيار رقماً عشوائياً محصوراً بين الواحد والرقم 5، بطريقة القرعة او باستخدام جداول الأرقام

العشوائية أو باستخدام الحاسبات بالعلاقة (1,5) =RANDBETWEEN وليكن 2، فيكون هو

البداية العشوائية، ويكون الطالب الذي ترتيبه 2 في المجتمع هو أول طالب في العينة.

- للحصول على ترتيب الطالب الثاني، تضاف فترة السحب إلى البداية العشوائية، فيتم الحصول على

الرقم 7، وهو الطالب الثاني في العينة، وهكذا يتم الحصول على ترتيب الطالب الثالث بإضافة فترة

السحب إلى ترتيب الطالب السابق ويساوي 12. وبذلك تكون العينة هي الطلبة ذوي الأرقام

2,7,12.

1.2-فترة السحب تحوي رقماً عشرياً

إذا كان عدد الطلبة في المثال السابق 16 طالباً كما في الجدول التالي:

فإن فترة السحب تكون مؤلفة من عدد صحيح ورقم عشري، ويتم سحب عينة حجمها 3 طلاب كما يلي:

- يتم حساب فترة السحب التي تساوي $I = \frac{16}{3} = 5.3$

- يتم اختيار رقم عشوائي محصور بين 0,1 باستخدام العلاقة () = RAND() وليكن الرقم 0.12

- يتم ضرب الرقم العشوائي بفترة السحب، فإذا كان الناتج عدداً صحيحاً فيكون هو ترتيب الوحدة الأولى في العينة، وإذا كان الناتج عدداً يحوي كسراً عشرياً فيكون ترتيب الوحدة الأولى هو العدد الصحيح الذي يلي العدد الناتج مباشرة في المجتمع، فالناتج في هذا المثال يساوي 0.636 (0.12*5.3)، فيكون ترتيب الطالب الأول في العينة هو العدد الصحيح في المجتمع الذي يلي العدد الناتج، وهو الرقم 1، ويقابل الطالب الذي رقمه 1 في هذا الصف.
 - وللحصول على ترتيب الطالب الثاني في العينة، يضاف الرقم 1 إلى الرقم العشوائي ويضرب الناتج بفترة السحب فيكون الناتج هو ترتيب الطالب الثاني في العينة، لذا فإن ترتيب الطالب الثاني هو 5.936 (5.3*1.12) ويقابل الطالب الذي رقمه 6 في هذا الصف.
 - وهكذا يضاف الرقم 2 إلى الرقم العشوائي ويضرب بفترة السحب للحصول على ترتيب الطالب الثالث الذي يساوي 11.236 ويكون ترتيب الطالب الثالث هو الطالب رقم 12 في هذا الصف.
- وإذا تم الرمز للرقم العشوائي بالرمز r فإن الترتيب العشوائي لوحدات العينة يحسب بالعلاقة التالية:

$$I * (r + i)$$

$$i = 0,1,2,\dots,n - 1$$

حيث

ويمكن ملاحظة أن الفرق بين الترتيب العشوائي لوحدتين متتاليتين في العينة يساوي فترة السحب أيضاً كما إذا كانت فترة السحب عدداً صحيحاً، أي أن ترتيب وحدات العينة يتم بإضافة فترة السحب إلى ترتيب الوحدة السابقة. ويمكن الحصول على عدد من العينات بتغيير فترة السحب، وهي عينات مختلفة وغير متداخلة، فلا تتكرر الوحدات في العينات المختلفة. ولهذه النتيجة أهمية كبيرة في تخفيف العبء عن المستهدفين في المسوح المختلفة.

ويمكن تطبيق هذه الطريقة لسحب عدد من العناقيد من المجتمع المستهدف في العينة العنقودية، أو سحب عدد من الوحدات الأولية (Primary Sampling Units) (PSUs) في العينة المتعددة المرحل، وقد تكون هذه الوحدات متساوية في الحجم أو غير متساوية في الحجم وهي الحالة العامة (العناقيد في العينة

العنقودية، أو وحدات العد في التعداد، أو القرى في الريف، أو الأحياء في المدن، وغيرها). فإذا كان عدد العناقيد أو الوحدات الأولية A في المجتمع المستهدف (متساوية في الحجم أو غير متساوية في الحجم) وكان عدد الوحدات المطلوب سحبها في هذه المرحلة a فإن سحب هذه الوحدات يمكن أن يتم بالطريقة المنتظمة

$$I = \frac{A}{a}$$

بإحدى الطريقتين السابقتين أيضاً، حيث فترة السحب تساوي

وبناء على التعريف السابق لاحتمال سحب وحدة من وحدات العينة، وبناء على طريقة السحب المنتظم فقد

قسم المجتمع المستهدف إلى n مجموعة حجم كل منها يساوي فترة السحب I فاحتمال سحب

وحدة واحدة من كل مجموعة ويرمز له بالرمز P وفقاً لما سبق، يساوي عدد الوحدات المسحوبة (1)

مقسوماً على حجم المجموعة (فترة السحب) وفقاً للعلاقة التالية:

$$P = \frac{1}{I} = \frac{1}{\frac{N}{n}} = \frac{n}{N} = f$$

ويكون احتمال سحب وحدات العينة ثابت f ويسمى كسر المعاينة (Sampling Fraction) وتسمى

العينة في هذه الحالة عينة متوازنة ذاتياً (Self-weighted) نظراً لتساوي احتمالات سحب وحدات العينة.

فاحتمال سحب الطالب في المثال الأول السابق يساوي 5 واحتمال سحب الطالب في المثال الثاني يساوي

5.3

فعندما تكون الوحدات الأولية متساوية في الحجم أو غير متساوية يكون سحبها مشابهاً لسحب وحدات العينة

المنتظمة فبدلاً من سحب عنصر واحد يتم سحب وحدة أولية مؤلفة من عدد من العناصر، ويكون احتمال سحب

الوحدة الأولية يساوي حاصل قسمة العدد 1 على فترة السحب، وتكون احتمالات سحب الوحدات متساوية والعينة

متوازنة ذاتياً أيضاً، ويحسب هذا الاحتمال بالعلاقة التالية

$$P = \frac{1}{I} = \frac{1}{\frac{A}{a}} = \frac{a}{A}$$

وبالمقابل يمكن أن تكون احتمالات سحب الوحدات غير متساوية، وعندئذ تكون العينة غير متوازنة ذاتياً، كما سيتم عرضة في الفقرة التالية.

| رقم الطالب | فترة السحب | الرقم العشوائي | الترتيب العشوائي | ترتيب وحدات العينة |
|------------|------------|----------------|------------------|--------------------|
| | I | r | $I*(r+i)$ | |
| 1 | 5.3 | 0.12 | 0.636 | 1 |
| 2 | | | 5.936 | 6 |
| 3 | | | 11.236 | 12 |
| 4 | | | | |
| 5 | | | | |
| 6 | | | | |
| 7 | | | | |
| 8 | | | | |
| 9 | | | | |
| 10 | | | | |
| 11 | | | | |
| 12 | | | | |
| 13 | | | | |
| 14 | | | | |
| 15 | | | | |
| 16 | | | | |

1.3- السحب المنتظم باحتمالات متناسبة مع الحجم (PPS) Probability Proportional to Size

في الواقع العملي نادراً أن تكون الوحدات الأولية متساوية في الحجم، وعندئذ يتم سحب الوحدات الأولية بطريقة السحب باحتمالات متناسبة مع الحجم (PPS) بالطريقة العشوائية المنظمة، وهو الأسلوب الأكثر استخداماً في التطبيقات العملية،

فإذا كان M_i حجم الوحدة الأولية i فيتم سحب a من الوحدات الأولية بهذه الطريقة وفق الخطوات التالية:

- إعداد التجميع المساعد لحجم الوحدات الأولية.

$$I = \frac{\sum_{i=1}^{i=A} M_i}{a}$$

- تحديد فترة السحب والتي تساوي

- يتم تطبيق الخطوات الواردة في الفقرة 1.1 السابقة، إذا كانت فترة السحب عدداً صحيحاً، والفقرة 1.2 إذا كانت فترة السحب تحوي رقماً عشرياً، لتحديد ترتيب الوحدات العشوائية.
- يتم تحديد قيمة في التجميع الصاعد تساوي أو أكبر مباشرة من القيمة التي يتم الحصول عليها بإحدى الطريقتين السابقتين، فتكون الوحدة المقابلة هي الوحدة العشوائية المطلوبة.
- وعلى سبيل المثال، يبين الجدول التالي سحب 4 وحدات عد من بين 19 وحدة عد غير متساوية في الحجم. وتم ذلك كما يلي:

- إعداد التجميع الصاعد لحجم الوحدات الأولية، كما في العمود الثالث من الجدول.

$$I = \frac{\sum_{i=1}^{i=A} M_i}{a} = \frac{2386}{4} = 596.5$$

تحديد فترة السحب، التي تساوي 596.5

- يتم اختيار رقم عشوائي محصور بين 0,1 باستخدام العلاقة $RAND()$ وليكن الرقم 0.013
- حساب الترتيب العشوائي للوحدة الأولى في العينة باستخدام العلاقة السابقة في 1.2 والذي يساوي

$$I * (r + i) = 596.5 * (0.013 + 0) = 7.75$$

- إن القيمة الأولى في الترتيب الصاعد التي هي أكبر مباشرة من هذه القيمة هي القيمة 104 وهي تقابل وحدة العد الأولى، فتكون وحدة العد التي رقمها 1 هي الوحدة الأولى في العينة.

- ويكون الترتيب العشوائي للوحدة الثانية يساوي

$$I * (r + i) = 596.5 * (0.013 + 1) = 604.25$$

فتكون الوحدة الثانية في العينة هي وحدة

العد التي رقمها 6.

- وهكذا يحدد ترتيب الوحدة الثالثة والرابعة، وتكون وحدات العينة هي وحدات العد ذات الأرقام: 1,6,10,15 ويمكن الحصول على عدد من العينات المختلفة، وذلك بتغيير البداية العشوائية في كل مرة. من الواضح أن احتمال سحب الوحدة الأولية i التي حجمها M_i يساوي حجمها مقسوماً على فترة السحب

$$P = \frac{M_i}{I} = \frac{M_i}{\sum_{i=1}^a M_i} = \frac{a}{\sum_{i=1}^a M_i} * M_i = \beta * M_i$$

حيث $\beta = \frac{a}{\sum M_i}$ وهو عدد ثابت. ويتضح من هذه العلاقة أن احتمالات سحب الوحدات مختلفة والعينة غير

متوازنة ذاتياً، واحتمال سحب الوحدة الأولية متناسب مع حجمها بالمعامل β لذا سميت هذه الطريقة "السحب باحتمالات متناسبة مع الحجم".

| ترتيب وحدات العينة | الترتيب العشوائي | الرقم العشوائي | فترة السحب | التجميع الصاعد | عدد الأسر | وحدات العد |
|--------------------|------------------|----------------|---------------|----------------|-----------|------------|
| | $I*(r+i)$ | r | $I=\sum Mi/4$ | Comul | Mi | |
| 1 | 7.75 | 0.013 | 596.5 | 104 | 104 | 1 |
| 2 | | | | 234 | 130 | |
| 3 | | | | 369 | 135 | |
| 4 | | | | 503 | 134 | |
| 5 | | | | 604 | 101 | |
| 6 | 604.25 | | | 704 | 100 | 6 |
| 7 | | | | 818 | 114 | |
| 8 | | | | 930 | 112 | |
| 9 | | | | 1080 | 150 | |
| 10 | 1200.75 | | | 1226 | 146 | 10 |
| 11 | | | | 1335 | 109 | |
| 12 | | | | 1481 | 146 | |
| 13 | | | | 1612 | 131 | |
| 14 | | | | 1759 | 147 | |
| 15 | 1797.25 | | | 1862 | 103 | 15 |
| 16 | | | | 1997 | 135 | |
| 17 | | | | 2129 | 132 | |
| 18 | | | | 2268 | 139 | |
| 19 | | | | 2386 | 118 | |
| Total | | | | | 2386 | |

2-طريقة السحب العشوائي المتسلسل Sequential Simple Random Sampling

ويتم سحب وحدات العينة بهذه الطريقة وفق الخطوات التالية:

- ترقيم وحدات الإطار التي عددها N من 1 إلى N
- يسند لكل وحدة رقماً عشوائياً محصوراً بين 0,1 باستخدام برنامج EXCEL بالعلاقة $() = RAND$
- ترتب هذه الأرقام تصاعدياً، ويتم بعد ذلك اختيار رقماً عشوائياً محدداً محصوراً بين 1 و N باستخدام برنامج EXCEL بالعلاقة $() = RANDBETWEEN(1, N)$ فيكون الرقم المقابل له في الترتيب العشوائي هو بداية السحب العشوائي، أو الرقم المقابل في الترتيب المتسلسل للوحدات.
- يتم تحديد عدد من الأرقام المتسلسلة بعد الرقم العشوائي يساوي حجم العينة المطلوب، وإذا تم بلوغ نهاية الترتيب قبل سحب حجم العينة المطلوب فينبغي استكمال حجم العينة بالتسلسل من بداية الترتيب. وأحياناً، للسهولة، يتم اعتماد بداية الترتيب التصاعدي هي بداية السحب العشوائي، ويتم تحديد عدد من الأرقام المتسلسلة بعد هذا الرقم يساوي حجم العينة المطلوب.

فعلى سبيل المثال، لسحب 3 طلاب عشوائياً من بين 15 طالباً في أحد الصفوف يتم اتباع الخطوات التالية:

- يتم ترقيم الطلبة من 1 إلى 15، كما في العمود الأول من الجدول التالي.
- يتم تحديد رقم عشوائي محصوراً بين 0,1 مقابل لكل طالب بالعلاقة السابقة، كما في العمود الثاني.
- يتم ترتيب الأرقام العشوائية تصاعدياً، ويتبع ذلك ترتيب الطلبة كما في العمودين الثالث والرابع.
- فإذا كان الرقم 5 هو الرقم العشوائي الذي تم اختياره بين 1 و 15 بالعلاقة السابقة، فيتم تحديده في الترتيب العشوائي ويكون هو رقم الطالب الأول في العينة، ويكون رقم الطالب الثاني في العينة هو الرقم التالي له مباشرة في الترتيب العشوائي، وهكذا للطالب الثالث، وتكون وحدات العينة هم الطلبة ذوي الأرقام

5,6,12

- وإذا كانت بداية السحب العشوائي 15 فإن وحدات العينة هم الطلبة 14،7،15.
- ويمكن إجراء عملية الاختيار باعتبار أن الرقم 5 الذي تم اختيار عشوائياً هو رقم الطالب في تسلسل ترتيب الطلبة وعندئذ يكون الطالب المقابل له في الترتيب العشوائي هو الطالب الأول في العينة، وهو الطالب ذو الرقم 3 ويكون رقم الطالب الثاني هو الرقم 9 والثالث 5 وتكون العينة عندئذ 5،9،3 ومن الواضح أنه يمكن سحب العديد من العينات العشوائية بهذه الطريقة. لذا، وللسهولة أحياناً، يتم اعتبار البداية العشوائية هي بداية الترتيب العشوائي، وتكون وحدات العينة في هذا المثال هم الطلبة 11،8،14. ومن الواضح أن احتمال سحب وحدة واحدة بهذه الطريقة يساوي $\frac{1}{N}$

| رقم الطالب | الرقم العشوائي | الترتيب العشوائي | |
|------------|----------------|------------------|-------|
| 1 | 0.669 | 14 | 0.191 |
| 2 | 0.829 | 8 | 0.200 |
| 3 | 0.367 | 11 | 0.309 |
| 4 | 0.752 | 10 | 0.352 |
| 5 | 0.556 | 3 | 0.367 |
| 6 | 0.577 | 9 | 0.404 |
| 7 | 0.891 | 5 | 0.556 |
| 8 | 0.2 | 6 | 0.577 |
| 9 | 0.404 | 12 | 0.658 |
| 10 | 0.352 | 1 | 0.669 |
| 11 | 0.309 | 13 | 0.713 |
| 12 | 0.658 | 4 | 0.752 |
| 13 | 0.713 | 2 | 0.829 |
| 14 | 0.191 | 15 | 0.854 |
| 15 | 0.854 | 7 | 0.891 |

3-طريقة (KISH) لاختيار فرد عشوائي من بين أفراد الاسرة

يطلب عادة في العديد من المسوح الاسرية، ومسوح استطلاعات الرأي، والمسوح التسويقية وغيرها، اختيار فرد عشوائي من بين أفراد الاسرة المؤهلين للمسح. ويتم الاعتماد على الجدول أدناه الذي أعده (Kish) عام 1965، وسمي جدول كيش (Kish Table) والذي يتألف من ثمانية جداول. ويتم اختيار فرد عشوائي من بين مجموعة من أفراد الاسرة المؤهلين، (18) سنة فأكثر من العمر على سبيل المثال، كما يلي:

- يتم تسجيل أفراد الأسرة الذكور المؤهلين (18) سنة فأكثر من العمر من الأكبر إلى الأصغر ويتبع ذلك تسجيل الإناث المؤهلات من الأكبر إلى الأصغر سناً أيضاً، ويتم ترقيم الأفراد بالتسلسل.
 - يتم اختيار رقم عشوائي محصوراً بين الرقم 1 والرقم 8 فيكون الرقم المقابل في الجدول في العمود الأول هو رقم الجدول العشوائي الذي سيتم على أساسه تحديد رقم الفرد المؤهل وفقاً لعدد أفراد الأسرة.
 - فإذا تم اختيار الجدول رقم (4) وكان عدد أفراد الأسرة المؤهلين (3) أفراد، فإن الفرد رقم (2) في مسلسل الأفراد هو الفرد المطلوب لإجراء المقابلة معه.
 - بعد أن يتم تحديد رقم الجدول العشوائي للأسرة الأولى، فإن الجدول العشوائي للأسرة الثانية هو الجدول الذي يليه، وهو الجدول ذي الرقم (5) في هذا المثال، وينبغي تسجيل الأفراد المؤهلين في الأسرة الثانية، وتحديد الفرد المؤهل كما في الخطوة السابقة.
 - وهكذا يتم تحديد الجدول العشوائي لكل أسرة في العينة بأخذ الجدول الذي يلي الجدول المقابل للأسرة التي تسبقها بالتسلسل، الى أن يتم تغطية جميع أسر العينة. ويلاحظ أن الحد الأقصى لعدد أفراد الأسرة المؤهلين هو 6 أفراد، وفي الواقع العملي يمكن أن يكون عدد أفراد الأسرة المؤهلين أكثر من ستة أفراد، وعندئذ ينبغي استخدام طريقة أخرى لسحب الفرد عشوائياً.
- ويكون احتمال سحب الفرد يساوي $\frac{1}{K}$ ، حيث K عدد أفراد الأسرة المؤهلين .

KISH TABLE

| | | الأفراد المؤهلين في الأسرة | | | | | |
|------------|---|----------------------------|---|---|---|---|----|
| رقم الجدول | | 1 | 2 | 3 | 4 | 5 | +6 |
| | | رقم الفرد المؤهل | | | | | |
| A | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| B1 | 2 | 1 | 1 | 1 | 1 | 2 | 2 |
| B2 | 3 | 1 | 1 | 1 | 2 | 2 | 2 |
| C | 4 | 1 | 1 | 2 | 2 | 3 | 3 |
| D | 5 | 1 | 2 | 2 | 3 | 4 | 4 |
| E1 | 6 | 1 | 2 | 3 | 3 | 3 | 5 |
| E2 | 7 | 1 | 2 | 3 | 4 | 5 | 5 |
| F | 8 | 1 | 2 | 3 | 4 | 5 | 6 |

4- طريقة اختيار رقم عشوائي محصور بين رقمين

ترقم وحدات المجتمع من 1 إلى N ويتم الحصول على عدد من الأرقام العشوائية يساوي حجم العينة باستخدام

$$= \text{randbetween}(1, N) \quad \text{العلاقة}$$

فإذا كان حجم المجتمع 50 أسرة، وكان حجم العينة المطلوب 5 أسر، فيمكن أن نسحب العديد من العينات العشوائية، التي حجم كل منها 5 أسر باستخدام العلاقة السابقة، فعينة الأسر ذات الأرقام 49,3,10,13,24 هي إحدى هذه العينات. وينبغي التمييز بين السحب مع الإعادة (With Replacement) حيث يمكن أن يتكرر الرقم المسحوب، وبين السحب بدون إعادة (without Replacement) حيث لا يمكن أن يتكرر الرقم المسحوب، وإذا تكرر الرقم في السحب بدون إعادة فيؤخذ رقماً عشوائياً إضافياً دون تكرار نفس الرقم. إن احتمال سحب الأسرة في هذا المثال يساوي $1/10$ (5/50).

5- طريقة الأرقام العشوائية للاتصال الهاتفي (Random Digit Dialing)

تشكل كشوف مشتركي الهاتف الثابت أو الهاتف المتحرك إطاراً لاختيار أفراد العينة بالطريقة العشوائية. وفي الوقت الراهن تستخدم طريقة ما يعرف بطريقة الاتصال العشوائي (Random Digit Dialing) على نطاق واسع وخاصة في استطلاعات الرأي التي تتم عن طريق الهاتف الثابت أو المتحرك. وملخص هذا الأسلوب اختيار أرقام عشوائية بالطريقة العشوائية وإجراء اتصال برقم الهاتف الأرضي أو المتحرك المطابق للرقم العشوائي الذي تم اختياره. وبحسب احتمال سحب وحدة أو عدة وحدات كما تم في الطرق السابقة أيضاً. في التطبيقات العملية، قد تنفذ عملية سحب الوحدات خلال عدة مراحل، فيمكن اتباع إحدى طرق السحب العشوائي السابقة في أية مرحلة من المراحل، وسيكون احتمال السحب في كل مرحلة مشروطاً (Conditional) باحتمال السحب في المرحلة السابقة. فعلى سبيل المثال، إذا سحبت وحدة عد في المرحلة الأولى في عينة أسرية ذات مرحلتين باحتمال يساوي P_1 فإن احتمال سحب أسرة من هذه الوحدة في

المرحلة الثانية يكون مشروطاً باحتمال سحب وحدة العد في المرحلة الأولى، وسيكون الاحتمال النهائي

لسحب الأسرة P يساوي $P = p_1 * p_2$ حيث P_2 هو احتمال سحب الأسرة في المرحلة الثانية.

3.2- الوزن الأساسي (المطلق)

إن الهدف الأساسي من استخدام طريقة المعاينة، هو الحصول على تقديرات للمتغيرات على مستوى المجتمع بناء على نتائج العينة. إلا إن التقديرات المحسوبة من نتائج العينة مباشرة تتباين عن مثيلاتها في المجتمع لعدة أسباب:

- سحب وحدات العينة باحتمالات غير متساوية، وهي الحالة العامة تقريباً، إذ يندر أن تكون احتمالات سحب وحدات العينة متساوية.
- عدم استجابة Non-response بعض وحدات العينة، وهذا الذي يحدث في أغلب الحالات، ويحدث ذلك عند الفشل في الحصول على البيانات المطلوبة من بعض وحدات العينة، بسبب الرفض، أو بسبب عدم تواجد الوحدة بتاريخ المسح، أو بسبب وجود وحدات غريبة عن المجتمع المستهدف بالدراسة وغيرها.
- عدم التغطية Non-coverage لبعض أجزاء المجتمع أو الوحدات، بسبب قدم الإطار، أو بسبب عدم تغطية بعض أجزاء المجتمع لأسباب قاهرة، بسبب الكوارث الطبيعية، أو بسبب استبعاد بعض أجزاء المجتمع لأسباب تتعلق بالتكلفة أو صعوبة الوصول إلى هذه الأجزاء (المناطق الصحراوية، الجبال الوعرة وغيرها). وينبغي تعديل النتائج النهائية بحيث تتوافق التوزيعات الأساسية (إجمالي السكان، العمر، الجنس) مع مثيلاتها من التوزيعات في المجتمع.

وتستخدم الأوزان لتعديل نتائج العينة، لإزالة أو تخفيض أثر هذه الأسباب على الفروقات بين تقديرات العينة والقيم الحقيقية لهذه التقديرات في المجتمع المستهدف. مع الإشارة إلى أن تقديرات العينة ستكون مختلفة عن القيم الحقيقية بسبب أنواع أخرى من الأخطاء، والتي تصنف إلى نوعين من الأخطاء: **النوع الأول الأخطاء غير العينية (non sampling error)**، وتندرج بعنوان أخطاء التحيز، مثل أخطاء: الأطر، والقياس، والاستجابة، ومعالجة البيانات وغيرها، وتفيد التجربة العملية إمكانية تخفيض هذا النوع من الأخطاء وليس إزالتها، بسبب تعدد مصادرها وعدم القدرة على التحكم بها بشكل كامل. **النوع الثاني الأخطاء العينية (sampling error)**، والمتمثلة بالخطأ الناتج عن الاقتصار على عينة من المجتمع المستهدف وليس كامل المجتمع، ويندرج هذا الخطأ تحت عنوان الأخطاء العشوائية، وهي ملازمة لأيّة عملية سحب عشوائي. ويمكن حسابها لكل نوع من أنواع العينات العشوائية والتحكم بمستوى هذا النوع من الأخطاء. تمثل العينة المجتمع المستهدف الذي سحبت وحدات العينة من بين وحداته، وبذلك تمثل كل واحدة من وحدات العينة عدداً من وحدات المجتمع المستهدف، والذي يسمّى وزن الوحدة، فإذا كان حجم المجتمع المستهدف 20 وحدة (أفراد، أو أسر، أو منشآت، أو مناطق، وغيرها)، وسحبت منه عينة حجمها 4 وحدات، فإن كل وحدة في العينة تمثل 5 وحدات ($20/4=5$) من وحدات المجتمع المستهدف، ويمثل العدد 5 وزن كل وحدة من وحدات العينة، ويعبر عن الوزن بلغة الاحتمالات بأنه مقلوب احتمال سحب الوحدة، فاحتمال سحب أية وحدة من عينة حجمها 4 وحدات سحبت من مجتمع حجمه 20 وحدة، كما هو معروف، يساوي $1/5$ ($4/20=1/5$)، ومقلوب هذا الاحتمال يساوي 5 والذي يمثل وزن كل وحدة من وحدات العينة. وللحصول على النتائج على مستوى المجتمع المستهدف يتم ضرب نتائج العينة بوزن كل وحدة من وحداتها، وتسمى هذه العملية توزيع نتائج العينة أو تكبير نتائج العينة على مستوى المجتمع المستهدف. وينبغي أن يكون مجموع الأوزان بعد عملية تكبير النتائج يساوي حجم المجتمع المستهدف، ويعتبر ذلك إحدى القواعد التي يجب أن تكون محققة دوماً.

وبشكل عام، وباستخدام العلاقات الرياضية، إذا كان P هو احتمال سحب الوحدة فإن مقلوب هذا الاحتمال

يسمى **الوزن الأساسي (Basic Weight)** ويرمز له بالرمز W_b ويساوي

$$W_b = \frac{1}{P}$$

وتعتبر الأوزان معاملات تضخيم لقيم الوحدات في العينة، ولهذا السبب تسمى أحياناً **معاملات التكبير**

(Raising Factors). وقد تكون احتمالات سحب وحدات العينة متساوية وستكون عندئذ أوزان الوحدات

متساوية وتسمى العينة عندئذ عينة ذاتية التوزيع أو عينة متوازنة ذاتياً (Self-Weighted). ومن خصائص

هذا النوع من العينات إن التقديرات (النسب والمتوسطات) على مستوى العينة تكون متطابقة قبل توزيع وبعد توزيع النتائج، لأن الوزن سيكون مضروباً في كل من المقام والبسط لكل من النسبة والمتوسط وبالتالي

سيحذف، وليس هناك حاجة لتوزيع نتائج العينة، ولهذا السبب يفضل استخدام هذا النوع من العينات

عندما يكون الهدف هو الحصول على تقديرات خاصة بالمجتمع مثل النسبة أو المتوسط، وعندما يكون من

المرغوب الحصول على المجاميع النهائية فينبغي استخدام الأوزان.

وبناء على ما تم عرضه في الفقرة السابقة، إذا سحبت عينة حجمها n من مجتمع حجمه N فإن

احتمال سحب أية وحدة يساوي كسر المعاينة $f = \frac{n}{N}$ ويكون وزن الوحدة

الأساسي في العينة يساوي مقلوب احتمال سحب الوحدة أو مقلوب كسر المعاينة، وقد يتم سحب وحدات

العينة بطريقة السحب المنتظم (Systematic Selection)، ويكون وزن الوحدة يساوي فترة السحب

وبشكل عام يحسب وزن الوحدة الأساسي بأحد أشكال العلاقة التالية $I = \frac{N}{n}$

$$W_b = \frac{1}{P} = \frac{1}{\frac{n}{N}} = \frac{N}{n} = \frac{1}{f} = I$$

وعلى سبيل المثال: بهدف تقدير متوسط دخل الأسرة الشهري بين مجموعة من الأسر عددها 20 أسر، سحبت عينة حجمها 5 أسر بإحدى طرق السحب العشوائي، فكان دخل الأسرة الشهري في العينة كما مبين في الجدول التالي.

| مسلسل الأسرة | دخل الأسرة وحدة نقدية | فترة السحب | احتمال سحب الأسرة | وزن الأسرة | تكبير (توزين) العينة | |
|--------------|-----------------------|-------------------|-------------------|-----------------------|----------------------|-----------|
| | | $I = \frac{N}{n}$ | $P = \frac{n}{N}$ | $w = \frac{1}{p} = I$ | الدخل | عدد الأسر |
| 1 | 85000 | 4 | 0.25 | 4 | 340000 | 4 |
| 2 | 60000 | | | | 240000 | 4 |
| 3 | 10000 | | | | 40000 | 4 |
| 4 | 120000 | | | | 480000 | 4 |
| 5 | 50000 | | | | 200000 | 4 |
| المجموع | 325000 | | | | 1300000 | 20 |
| المتوسط | 65000 | | | | 65000 | |

من الواضح أن فترة السحب تساوي $I = \frac{N}{n} = \frac{20}{5} = 4$ واحتمال سحب الأسرة يساوي

وهو كسر المعاينة ايضاً. فيكون الوزن الأساسي لكل أسرة في العينة يساوي فترة السحب $p = \frac{n}{N} = \frac{5}{20} = \frac{1}{4}$

أو مقلوب احتمال سحب الأسرة أو مقلوب كسر المعاينة، وفي جميع الحالات يكون الوزن الأساسي لكل أسرة ثابت ويساوي 4، والعينة متوازنة ذاتياً، ويتم تكبير نتائج العينة بضرب وزن كل أسرة بدخلها للحصول

- على إجمالي دخل الأسر في المجتمع المستهدف (20 أسرة)، وضرب وزن كل أسرة ب 1 للحصول على عدد الأسر في المجتمع كما مبين في العمودين الأخيرين من الجدول. ويتضح من هذا الجدول ما يلي:
- إن متوسط دخل الأسرة في المجتمع المستهدف بعد عملية توزيع نتائج العينة يساوي 65000 وحدة نقدية، وهو نفس متوسط دخل الأسرة في العينة قبل عملية التوزيع، ويعود السبب في ذلك إلى أن العينة متوازنة ذاتياً، كما اشير إلى ذلك سابقاً.
 - ولكن مجموع الدخول في المجتمع المستهدف يقدر ب 1300000 وحدة نقدية، وهو مختلف عن مجموع دخول أسر العينة والبالغ 325000 وحدة نقدية، ولهذا السبب ينبغي تكبير نتائج العينة للحصول على المجاميع النهائية. ويمكن الحصول على مجموع الدخول في المجتمع المستهدف في هذا المثال بضرب حجم المجتمع بمتوسط العينة (20*65000)، ويعود السبب في ذلك إلى أن أوزان أسر العينة متساوية والعينة متوازنة ذاتياً، ولا يصح ذلك عندما تكون أوزان أسر العينة مختلفة. فإذا كانت أوزان أسر العينة (افتراضياً مختلفة) كما في الجدول التالي، فإن متوسط الدخل الشهري للأسرة في المجتمع يساوي 67000 وحدة نقدية بعد عملية توزيع نتائج العينة، بينما يساوي 65000 في العينة.

| مسلسل الأسرة | دخل الأسرة | وزن الأسرة | تكبير (توزين) العينة | |
|--------------|------------|------------|----------------------|-----------|
| | | | الدخل | عدد الأسر |
| 1 | 85000 | 2 | 170000 | 2 |
| 2 | 60000 | 4 | 240000 | 4 |
| 3 | 10000 | 3 | 30000 | 3 |
| 4 | 120000 | 5 | 600000 | 5 |
| 5 | 50000 | 6 | 300000 | 6 |
| المجموع | 325000 | | 1340000 | 20 |
| المتوسط | 65000 | | 67000 | |

و قد يتم سحب الوحدة النهائية (المستهدفة) على عدة مراحل كما في العينة المتعددة المراحل، وعندئذ سيكون لكل مرحلة من المراحل الوزن الخاص بها، ويكون وزن الوحدة النهائية هو حاصل ضرب الأوزان في جميع المراحل، فإذا كان W_1 وزن سحب الوحدة في المرحلة الأولى والتي قد تكون وحدة عد، على سبيل المثال، وكان W_2 وزن سحب الأسرة في المرحلة الثانية من أسر وحدة العد التي سحبت في المرحلة الأولى، وكان W_3 وزن سحب أحد أفراد الأسرة (المؤهلين للمسح) التي سحبت في المرحلة الثانية فإن الوزن

الأساسي للوحدة النهائية (الفرد في هذا المثال) W_b يعطى بالعلاقة التالية

$$W_b = W_1 * W_2 * W_3$$

وقد تكون العينة متوازنة ذاتياً في إحدى مراحل سحب العينة، وقد لا ينطبق التوازن الذاتي على المرحلة التالية، فقد تكون العينة متوازنة ذاتياً لدى سحب عينة من الأسر، وإذا تم سحب فرد عشوائي من بين أفراد الأسرة، فتصبح العينة غير متوازنة ذاتياً بالنسبة للفرد العشوائي، حيث سيختلف وزن الفرد من أسرة إلى أخرى بسبب اختلاف عدد الأفراد بين الأسر. وفي الواقع العملي، نادراً أن تكون العينة متوازنة ذاتياً، بسبب عدم استجابة بعض وحدات العينة.

3.3- التعويض عن عدم الاستجابة

في التطبيقات العملية نادراً ما تبقى الأوزان الأساسية ثابتة بسبب عدم استجابة بعض وحدات العينة لأسباب متعددة (الرفض، عدم التواجد، الوحدات غير المؤهلة، الوحدات المغلقة وغيرها)، وينبغي تعديل الأوزان الأساسية للتعويض عن عدم الاستجابة، بحيث يعدل وزن الوحدة بما يتوافق مع الوحدات المستجيبة، ويتم ذلك بحساب نسبة الاستجابة، ولحساب نسبة الاستجابة ينبغي تصنيف وحدات العينة إلى الأنواع التالية:

1. **وحدات مستجيبة**، وهي الوحدات المؤهلة التي أعطت إجابات مكتملة (كلياً أو جزئياً)، ويرمز لها

بالرمز H_r ويقصد بالمؤهلة أنها تنتمي إلى المجتمع وليست غريبة.

2. **وحدات غير مستجيبة**، وهي الوحدات المؤهلة التي لم يتم الحصول منها على البيانات المطلوبة،

لأسباب عدة، مثل الرفض، أو عدم التواجد، أو المرض وغيرها، ويرمز لها بالرمز (H_{nr}).

3. **وحدات غير مؤهلة**، يتبين من خلال العمل الميداني أن بعض الوحدات n' غريبة لا تنتمي إلى

المجتمع المستهدف، لذا هي غير مؤهلة، ويندرج ضمن هذه الفئة المساكن الخالية، والمساكن المهدامة.

4. **وحدات غير معروفة فيما إذا كانت مؤهلة أو غير مؤهلة** n_{un} كما في حالة الوحدات المغلقة أو

عدم التواجد بتاريخ المسح. وبعد الانتهاء من العمل الميداني، يتم التأكد من وجود بعض هذه الحالات أو جميعها، ويتعلق ذلك بحداثة الأطر المستخدمة في سحب وحدات العينة. وفي الواقع العملي يمكن مواجهة الحالات التالية:

• **عدم وجود وحدات غير مؤهلة أو غير معروفة الأهلية.** وعندئذ تصنف وحدات العينة إلى وحدات

مستجيبة ووحدات غير مستجيبة، وتحسب نسبة الاستجابة بالعلاقة التالية:

$$R = \frac{H_r}{H_r + H_{nr}}$$

فإذا كان حجم العينة 200 أسرة وكان عدد الأسر غير المستجيبة 20 أسرة، وعدد الأسر المستجيبة 180

$$R = \frac{H_r}{H_r + H_{nr}} = \frac{180}{180 + 20} = 0.90 \quad \text{أسرة فإن نسبة الاستجابة تساوي}$$

ولما كان الوزن الأساسي للوحدة $\frac{N}{n}$ قد حسب على أساس حجم العينة الأساسي، ولم يتم استجابة

جميع الوحدات بل 90% فقط منها، فينبغي تعديل حجم العينة الأساسي ليتوافق مع حجم العينة

المستجيبة، وبذلك يكون حجم العينة الجديد يساوي حجم العينة الأساسي مضروباً بنسبة الاستجابة، أي أن

$$n_r = 0.90 * n \quad \text{حجم العينة المستجيبة يساوي}$$

$$W = \frac{N}{n_r} = \frac{N}{0.90 * n} = \frac{N}{n} * \frac{1}{0.90}$$

ويصبح الوزن الفعلي لكل وحدة

ومن الواضح أن الوزن الجديد لكل وحدة يساوي الوزن الأساسي مضروباً بمقلوب نسبة الاستجابة، وهي

القاعدة العامة في حالة وجود عدم استجابة. فإذا كان مقلوب نسبة الاستجابة W_{rr} فإن

$$W_{rr} = \frac{1}{R}$$

$$W = W_b * W_{rr}$$

ويصبح الوزن الجديد

فإذا كان الوزن الأساسي للوحدة 10 في هذا المثال، فإن الوزن الجديد يساوي

$$W = 10 * \frac{1}{0.90} = 1.11$$

• وجود حالات غير مؤهلة، فيكون حجم العينة الفعلي يساوي $n_r = n - n'$ ونسبة الاستجابة

$$R = \frac{H_r}{n - n'}$$

فإذا كان حجم العينة 250، وصنفت بعد العمل الميداني كما في الجدول أدناه، فإن نسبة الاستجابة

$$R = \frac{H_r}{n - n'} = \frac{200}{250 - 20} = 0.8696$$

فإذا كان الوزن الأساسي للوحدة 10 فإن الوزن الجديد يساوي $W = 10 * \frac{1}{0.8696} = 1.15$

| | |
|-----|-----------------------|
| 200 | أسر مستجيبة |
| 30 | أسر مؤهلة غير مستجيبة |
| 20 | أسر غير مؤهلة |
| 250 | المجموع |

- وجود حالات غير معروفة الأهلية، وفي هذه الحالة ينبغي تقدير عدد الحالات المؤهلة من بينها.

ويتم ذلك عن طريق تقدير نسبة الوحدات المؤهلة ε والتي تساوي:

$$\varepsilon = \frac{H_r + H_{nr}}{H_r + H_{nr} + n'}$$

وعدد الحالات المتوقع أن تكون مؤهلة في هذا النوع من الوحدات يساوي

$$H_{el} = \varepsilon * n_{un}$$

بعد أن يتم تصنيف هذه الحالات، يصبح حجم العينة الفعلي

$$n_r = H_r + H_{nr} + H_{el}$$

ويتم حساب نسبة الاستجابة بالعلاقة

$$R = \frac{H_r}{H_r + H_{nr} + H_{el}}$$

وللتعويض عن عدم الاستجابة يتم ضرب الوزن الأساسي بمقلوب نسبة الاستجابة

$$W_{rr} = \frac{1}{R} = \frac{1}{\frac{H_r}{H_r + H_{nr} + H_{el}}} = \frac{H_r + H_{nr} + H_{el}}{H_r}$$

ويصبح الوزن الجديد

$$W = W_b * W_{rr}$$

وكمثال على ذلك، سحبت عينة حجمها 300 أسرة، وبعد انتهاء العمل الميداني تبين أن توزيع الأسر التي تم زيارتها كما يلي:

| | |
|-----|----------------------------------|
| 200 | أسر مستجيبة |
| 30 | أسر مؤهلة غير مستجيبة |
| 25 | أسر غريبة، أو مسكن خالي أو مهدوم |
| 45 | أسر غير معروف أهليتها |
| 300 | المجموع |

فتكون نسبة الأسر المؤهلة بين الأسر غير المعروف أهليتها

$$\varepsilon = \frac{H_r + H_{nr}}{H_r + H_{nr} + n'} = \frac{200 + 30}{200 + 30 + 25} = 0.902$$

ويكون عدد الاسر المتوقع أن تكون مؤهلة بين الأسر غير المعروف أهليتها

$$H_{el} = \varepsilon * n_{un} = 0.902 * 45 = 41$$

وتكون نسبة الاستجابة تساوي

$$R = \frac{H_r}{H_r + H_{nr} + H_{el}} = \frac{200}{200 + 30 + 41} = 0.738$$

وينبغي تعديل الوزن الأساسي بضربه بمقلوب نسبة الاستجابة، فإذا كان الوزن الأساسي 10 فإن الوزن

$$W = W_b * W_{rr} = 10 * \frac{1}{0.738} = 13.55$$

المعدّل يساوي

3.4-التعويض عن عدم التغطية

عندما يتم الاعتماد على أطر حديثة تمثل حجم المجتمع المستهدف أو حجم الطبقات المكوّنة له، سيكون حجم المجتمع بعد عملية التوزيعين مطابقاً لحجم المجتمع الفعلي، وفي الواقع العملي يتم الاعتماد على أطر قديمة نسبياً لا تعكس الواقع الفعلي للتغيرات التي طرأت على حجم المجتمع المستهدف، وينتج عن ذلك

عدم تطابق المجاميع النهائية بعد عملية توزيع النتائج مع المجاميع الفعلية بتاريخ المسح، لذا يتم الاعتماد على أحدث التقديرات الخاصة بتوزيعات المجتمع الإجمالية، بحيث تتوافق نتائج العينة بعد عملية التوزيع مع هذه التوزيعات. فإذا كان حجم المجتمع الأساسي N وحجم المجتمع المقدر بتاريخ المسح (N')، ينبغي ضرب الوزن الأساسي بالنسبة ($\frac{N'}{N}$). ويرمز له بالرمز W_c أي أن

$$W_c = \frac{N'}{N}$$

إن الوزن النهائي W هو حاصل ضرب الأوزان السابقة

$$W = W_b * W_{rr} * W_c$$

وينبغي تعديل النتائج بضرب نتائج العينة بالوزن النهائي، وبذلك يمكن الحصول على التقديرات المطلوبة غير المتحيزة (إن الفروق بينها وبين القيم الحقيقية تعزى إلى الأخطاء العشوائية فقط). وعلى سبيل المثال، سحبت عينة حجمها 100 بالطريقة العشوائية من مجتمع حجمه 1500، فمن الواضح

$$P = \frac{n}{N} = \frac{100}{1500} = \frac{1}{15} \quad \text{أن احتمال سحب وحدات العينة}$$

$$W_b = \frac{1}{P} = \frac{1}{\frac{1}{15}} = 15 \quad \text{ويكون الوزن الأساسي لكل وحدة}$$

ويعني ذلك أن كل وحدة من العينة تمثل 15 وحدة من المجتمع.

وبفرض أن عدم الاستجابة بلغ 10 وحدات من العينة، وبافتراض عدم وجود وحدات غريبة $n' = 0$ وعدم وجود وحدات غير معروفة $n_{un} = 0$ عندئذ سيكون معدل الاستجابة يساوي

$$R = \frac{H_r}{H_r + H_{nr}} = \frac{90}{90 + 10} = \frac{90}{100} = \frac{9}{10}$$

ويكون وزن التعويض عن عدم الاستجابة W_{rr} والذي يساوي مقلوب نسبة الاستجابة، أي

$$W_{rr} = \frac{1}{R} = \frac{1}{\frac{9}{10}} = \frac{10}{9}$$

ويحسب الوزن الجديد (W')، بضرب الوزنين السابقين

$$W' = W_b * W_{rr} = 15 * \frac{10}{9} = 16.67$$

فأصبحت كل وحدة في العينة تمثل 16.67 في المجتمع بدلاً من 15. وينبغي ملاحظة أن حجم المجتمع يبقى ثابتاً بعد عملية التوزيع. فإذا تم الاقتصار على الوزن الأساسي (وفي هذه الحالة لا يوجد عدم استجابة)، فإن ناتج ضرب العينة بالوزن الأساسي يساوي $15 * 100 = 1500$ وهو حجم المجتمع.

وإذا كان إطار المجتمع المستهدف حديثاً (ويتحقق ذلك نادراً)، ووجد عدم استجابة يتم تكبير نتائج العينة بالوزن W' ويكون حجم المجتمع الناتج عن ضرب الوزن الجديد بعدد الوحدات المستجيبة

$$15 * \frac{10}{9} * 90 = 1500 \quad (90) \text{ مطابقاً لحجم المجتمع أيضاً}$$

وإذا كان إطار العينة قديماً نسبياً، وبفرض أن حجم المجتمع وفق أحدث التقديرات هو 1700 فإن وزن

التعويض عن عدم التغطية W_c يساوي

$$W_c = \frac{N'}{N} = \frac{1700}{1500}$$

ويكون الوزن النهائي W هو جداء الأوزان السابقة

$$W = W_b * W_{rr} * W_c = 15 * \frac{10}{9} * \frac{1700}{1500} = 18.89$$

ويكون ناتج ضرب الوزن النهائي بحجم العينة المستجيبة يساوي 1700 وهو حجم المجتمع وفق أحدث التقديرات.

عندما يتم تقسيم المجتمع الى طبقات، كما في حالة الحضر والريف، على سبيل المثال، فينبغي حساب الأوزان المختلفة السابقة في كل طبقة وضرب نتائج العينة في كل طبقة بالأوزان الخاصة بتلك الطبقة، بحيث يكون حجم كل طبقة بعد عملية التوزين مساوياً لحجمها الأساسي قبل عملية التوزين وفق أحدث التقديرات لحجم كل طبقة كما تم توضيحه سابقاً، أو ينبغي أن يكون التركيب النسبي لأحجام الطبقات بعد عملية التوزين مطابقاً للتركيب النسبي الحقيقي لهذه الطبقات في المجتمع المستهدف، وما لم يتحقق ذلك فإن التقديرات على مستوى إجمالي المجتمع المستهدف ستكون متحيزة لإحدى الطبقات وغير مقبولة. فعلى سبيل المثال، يبين الجدول التالي مجتمع مؤلف من طبقتين، وقد سحبت عينة من كل طبقة لتقدير نسبة الذكور والاناث في كل طبقة وفي المجتمع المستهدف، وتم توزيع نتائج العينة في كل طبقة وفقاً لما سبق، وكانت النتائج النهائية لكل من الذكور والاناث في كل طبقة بعد توزيع نتائج العينتين كما مبين في الجدول، وبناء على أحدث التقديرات السكانية لحجم كل طبقة فإن حجم الطبقة الأولى 10400 فرداً، وحجم الطبقة الثانية 2600 فرداً. من الواضح اختلاف حجم كل طبقة بعد توزيع النتائج عن حجمها في المجتمع المستهدف، واختلاف مجموعهما عن حجم المجتمع المستهدف، ويختلف أيضاً التركيب النسبي للطبقتين بعد عملية التوزين عن تركيبهما النسبي في المجتمع المستهدف، لأن التركيب النسبي لهما 70%، 30% على التوالي بعد عملية التوزين، وهو مختلف عن تركيبهما النسبي في المجتمع المستهدف 80%، 20% على التوالي. ويعود ذلك الى السببين التاليين:

- عدم التغطية
- اختلاف أوزان وحدات العينة في الطبقة أو بين الطبقات، أي أن العينة غير متوازنة ذاتياً.

| المجتمع | | | العينة موزونة | | | | | |
|---------|-------|--------|------------------|--------|--------|--------|--------|--------|
| الطبقة | N_i | النسبة | N'_i | النسبة | الذكور | النسبة | الإناث | النسبة |
| 1 | 10400 | 0.8 | 7700 | 0.7 | 4235 | 0.55 | 3465 | 0.45 |
| 2 | 2600 | 0.2 | 3300 | 0.3 | 1683 | 0.51 | 1617 | 0.49 |
| Total | 13000 | 1 | 11000 | 1 | 5918 | 0.538 | 5082 | 0.462 |

وهذا يتناقض مع المبدأ الأساسي لنتائج عملية توزيع نتائج العينة، فيجب أن تكون أحجام الطبقات بعد عملية التوزيع مساوية لأحجامها في المجتمع، وبناء على هذا المبدأ ينبغي تعديل الأوزان بحيث يتحقق ما يلي:

- حجم كل طبقة مساوياً لحجمها في المجتمع المستهدف.
- أن يكون التركيب النسبي للطبقتين متوافقاً مع التركيب النسبي الحقيقي للمجتمع المستهدف.

ويتم ذلك بطريقتين:

1. الطريقة الأولى:

تعديل النتائج بحيث تكون مجاميع الطبقتين بعد التوزيع مساوية لمجاميعهما في المجتمع المستهدف، وذلك بالتعويض عن عدم التغطية وفقاً لما تم عرضه سابقاً، وسيكون بالتالي تركيبهما النسبي مماثلاً لتركيبهما النسبي في المجتمع المستهدف. ويتم ذلك بضرب الأوزان في كل طبقة بالنسبة بين حجم الطبقة الحقيقي وحجمها بعد التوزيع، كما مبين في العمود الثالث من الجدول التالي. ويتضح من الجدول أن حجم الطبقتين بعد تعديل الأوزان أصبح مجموعهما يساوي مجموع المجتمع المستهدف، وحجمهما يساوي حجمهما في المجتمع المستهدف، وتركيبهما النسبي أيضاً كما في العمود الخامس من الجدول، وقد أصبحت نسبة الذكور على مستوى المجتمع المستهدف 0.542 بدلاً من 0.538 قبل تعديل الأوزان، وأصبحت نسبة الإناث 0.458 بدلاً من 0.462 قبل تعديل الأوزان، وهي اختلافات طفيفة في هذا المثال، ويمكن عرض أمثلة

أخرى تكون فيها هذه الاختلافات كبيرة. ويلاحظ أن تعديل الأوزان لم يغير نسبة الذكور ونسبة الإناث على مستوى كل طبقة، ويعود السبب في ذلك إلى أن البسط والمقام في معادلة النسبة سيضربان بنفس معامل التعديل، ويؤدي ذلك إلى حذفه، ويمكن التعبير عن هذه النتيجة بأن خصائص كل طبقة (النسب والمتوسطات) لا تتغير بإجراء هذه التعديلات، وهو مبدأ صحيح دوماً، وينبغي التحقق من ذلك لدى إجراء هذا النوع من التعديلات.

| المجتمع | | تعديل الاوزان | | | | | | | |
|---------|-------|---------------|--------------|---------------------|--------|--------|--------|--------|--------|
| الطبقة | N_i | N'_i | N_i / N'_i | $N'_i * N_i / N'_i$ | النسبة | الذكور | النسبة | الإناث | النسبة |
| 1 | 10400 | 7700 | 1.351 | 10400 | 0.8 | 5720 | 0.55 | 4680 | 0.45 |
| 2 | 2600 | 3300 | 0.788 | 2600 | 0.2 | 1326 | 0.51 | 1274 | 0.49 |
| Total | 13000 | 11000 | 1.182 | 13000 | 1 | 7046 | 0.542 | 5954 | 0.458 |

2. الطريقة الثانية:

في التطبيقات العملية، قد يكون الهدف من المسح تقديرات الخصائص (النسب والمتوسطات) فقط دون المجاميع الإجمالية، كما في العديد من المسوح الديموغرافية، واستطلاعات الرأي وغيرها، وفي الواقع العملي قد لا تتوفر تقديرات حديثة لحجم كل طبقة، وفي الحالتين يمكن الاعتماد على طريقة أخرى لتعديل الأوزان بحيث يكون التركيب النسبي للطبقات بعد التوزين مطابقاً للتركيب النسبي للطبقات في المجتمع المستهدف وفق التقديرات النسبية المعروفة سابقاً للطبقتين في المجتمع المستهدف (من التعداد الأخير على سبيل المثال)، ويتم ذلك بضرب الأوزان بمعامل يساوي النسبة بين النسبة الحقيقية للطبقة ونسبة الطبقة بعد التوزين، كما في الجدول أدناه. وبناء على ذلك أصبح التركيب النسبي للطبقتين مطابقاً للتركيب

النسبي لهما في المجتمع المستهدف، وتكون نسبة الذكور والاناث في المجتمع المستهدف مساوية للنسبة باستخدام الطريقة الأولى.

| النسبة بين نسبي الطبقة P | N' | $N' * P$ | النسبة | الذكور | النسبة | الإناث | النسبة |
|----------------------------|-------|----------|--------|--------|--------|--------|--------|
| 1.142857 | 7700 | 8800 | 0.8 | 4840 | 0.55 | 3960 | 0.45 |
| 0.666667 | 3300 | 2200 | 0.2 | 1122 | 0.51 | 1078 | 0.49 |
| Total | 11000 | 11000 | 1 | 5962 | 0.542 | 5038 | 0.458 |

ويمكن الوصول الى نفس النتائج على المستوى الإجمالي، وذلك اعتماداً على المبدأ المعروف لحساب المتوسط أو النسبة على مستوى مجموع عدة طبقات، فإذا كانت P_i النسبة في الطبقة i وكانت N_i حجم الطبقة i فإن النسبة P على المستوى الإجمالي تحسب بالعلاقة التالية، بالاعتماد على حجم كل طبقة.

$$P = \frac{\sum N_i * P_i}{\sum N_i}$$

ويمكن حساب النسبة على المستوى الإجمالي بالاعتماد على نسبة حجم كل طبقة من الحجم الإجمالي

$$W_i = \frac{N_i}{\sum N_i}$$

للطبقات W_i ويمكن كتابة العلاقة السابقة على الشكل التالي

$$P = \sum W_i * P_i$$

ويمكن كتابة علاقيتين مشابهتين لحساب المتوسط للمجموع. وتطبيق ذلك على هذا المثال فإن نسبة الذكور على المستوى الإجمالي تساوي 0.542 وهي نفس النسبة المحسوبة بالطريقتين السابقتين، $P = 0.8 * 0.55 + 0.2 * 0.51 = 0.542$ وسيتم الحصول على نفس نسبة الاناث بتطبيق هذه العلاقة. مما سبق يمكن أن نصل إلى النتيجة الهامتين التاليتين:

- ينبغي تعديل الأوزان بحيث تكون أحجام الطبقات أو التركيب النسبي في العينة التطبيقية مماثلاً للطبقات في المجتمع المستهدف.
- يمكن الاستفادة من نتائج العينة لتعديل بعض خصائص المجتمع التي اعتمدت على تقديرات تم إعدادها بناء على نتائج تعداد قديم، ومن المتوقع أن تكون قد حدثت عليها تغيرات كبيرة. بحيث يتم تعديل التركيب النسبي للفئات المستهدفة في المجاميع المقدره بما يتوافق مع التركيب النسبي لهذه الفئات في نتائج العينة الموزونة، بطريقة عكسية للطريقة الثانية الموضحة أعلاه.

3.5-التعويض عن الاستجابات غير المكتملة (Imputation)، ويقصد بذلك التعويض عن بعض البنود غير المكتملة بسبب رفض الإجابة من قبل بعض وحدات العينة، وعندئذ يمكن التعويض عن هذه البنود بالبيانات المشابهة لوحدات مماثلة لهذه الوحدات، من حيث الموقع الجغرافي أو الحالة الاجتماعية أو النشاط الاقتصادي وغيرها من المعايير المناسبة. ويمكن استخدام طريقة الانحدار للتعويض عن القيم المفقودة.

3.6-تخفيض الأوزان (Trimming of Weights)

بعد استكمال تعديل الأوزان الأساسية، ينبغي التدقيق في بعض أوزان الوحدات ذات الأوزان المفرطة في الكبر، حيث سيؤدي إلى زيادة كبيرة في تباين التقديرات وبالتالي تضخيم خطأ التقدير، ويمكن أن يتحكم بالتقديرات الإجمالية، ويتم عادة تخفيض هذا النوع من الأوزان لعدد محدود بحيث لا تتجاوز حد معين، وبحيث لا يتغير إجمالي الأوزان، فإذا كان N إجمالي الأوزان قبل التخفيض وكان N' إجمالي الأوزان بعد التخفيض، فينبغي ضرب أوزان جميع وحدات العينة بما فيها الأوزان الجديدة المخفّضة بالنسبة $\frac{N}{N'}$ ، وتجدر الإشارة إلى أن هذا الاجراء يساعد على تخفيض التباين وبالتالي تخفيض أخطاء التقديرات، إلا أنه يزيد من التحيز في بعض التقديرات، لذا ينبغي الحذر في تخفيض الأوزان وعدم اللجوء إلى هذا الإجراء إلا في حالات خاصة ومحدودة.

3.7- الأوزان النسبية (Relative Weights)

في بعض الأحيان، وبقصد السهولة، في حساب بعض المؤشرات مثل المتوسط أو النسبة، تستخدم الأوزان النسبية بدلاً من الأوزان المطلقة، ويتم ذلك بحيث يكون مجموع أوزان العينة يساوي (1)، وعندئذ يكون الوزن النسبي لأحد العناصر مساوياً لنتاج ضرب الوزن المطلق لهذا العنصر بمقلوب متوسط جميع

الأوزان. فإذا كان W_i الوزن المطلق للعنصر i وكان W_{ir} الوزن النسبي له، فإن الوزن النسبي

يحسب بالعلاقة التالية

$$W_{ir} = W_i * \frac{1}{\sum W_i} = W_i * \frac{n}{\sum W_i}$$

وينبغي أن يكون متوسط الأوزان النسبية يساوي 1 كما مبين في المثال التالي، حيث يمثل العمود الأول ارقام أسر العينة، ويمثل العمود الثاني الأوزان الأساسية لعينة حجمها 5 أسر. ويتم حساب الأوزان النسبية باتباع الخطوات التالية.

| رقم الأسرة | الوزن- W_i | دخل الأسرة- I | $w_i * 1 / \sum w_i / n$ | $W_i * I$ | $(w_i * 1 / \sum w_i / n) * I$ |
|------------|-----------------|-----------------|--------------------------|-----------|--------------------------------|
| 1 | 120 | 10 | 1.5 | 1200 | 15 |
| 2 | 75 | 20 | 0.9375 | 1500 | 18.75 |
| 3 | 15 | 35 | 0.1875 | 525 | 6.5625 |
| 4 | 90 | 50 | 1.125 | 4500 | 56.25 |
| 5 | 100 | 70 | 1.25 | 7000 | 87.5 |
| المجموع | 400 | | 5 | 14725 | 184.0625 |
| المتوسط | 80 | | | 36.8125 | 36.8125 |

- **الخطوة الأولى:** حساب متوسط الأوزان الأساسية (المطلقة) والذي يساوي 80
- **الخطوة الثانية:** يحسب الوزن النسبي لكل وحدة بضرب الوزن الأساسي بمقلوب متوسط الأوزان الأساسية. كما مبين في العمود الرابع.

ويتم توزيع نتائج العينة بالأوزان النسبية الجديدة، لحساب المتوسطات والنسب المئوية فقط، التي لن تختلف عما هي عليه إذا تم توزيع النتائج بالأوزان المطلقة. ويعود السبب في ذلك إلى أن قيمة المؤشر في العينة يبقى ثابتاً قبل عملية التوزيع وبعدها لأن كل من البسط والمقام سيضرب بنفس الوزن. فإذا تم تكبير النتائج بالأوزان الأساسية كما في العمود الخامس، وتم تكبير النتائج بالأوزان النسبية كما في العمود السادس، فإن متوسط دخل الأسرة يساوي 36.8125 في الحالتين. وعندما يكون الهدف الحصول على المجاميع الإجمالية للمجتمع، فينبغي استخدام الأوزان الأساسية، فيكون إجمالي دخل المجتمع في هذا المثال 14725 باستخدام الأوزان الأساسية، أو بضرب متوسط دخل الأسرة بمجموع الأوزان (400*36.8125)، وهذا يختلف عن إجمالي الدخل باستخدام الأوزان النسبية الذي يساوي 184.0625.

4- أخطاء العينة

يُميز بين نوعين من الأخطاء في العينات الاحتمالية، هي الأخطاء غير العينية (Non-sampling errors)، والأخطاء العينية (Sampling errors). وكما هو معروف فإن الأخطاء غير العينية هي أخطاء شائعة في الحصر الشامل وكذلك لدى استخدام أسلوب المعاينة، وهي أخطاء عديدة ومتنوعة، كأخطاء الشمول، وأخطاء عدم الاستجابة وأخطاء الإجابة وغيرها من الأخطاء المعروفة في المسوح والتعدادات على حد سواء. أما الأخطاء العينية فهي الأخطاء الخاصة بالعينة، ولا توجد مثل هذه الأخطاء في التعداد أو الحصر الشامل، وهي ناتجة عن دراسة جزء من المجتمع وليس كل المجتمع، وتسمى هذه الأخطاء أحياناً، الأخطاء العشوائية أو أخطاء الصدفة. وتأتي أهمية حساب الأخطاء المرتكبة في التقديرات المختلفة، كونها تساعد في تحديد المجالات التي يمكن أن تقع فيها هذه التقديرات ومستويات الثقة المرغوبة، وإجراء الاختبارات العديدة على نتائج العينة. ويمكن حساب هذا النوع من الأخطاء لجميع التقديرات التي تغطي عدداً كبيراً من المؤشرات، وفي التطبيقات العملية يتم الاقتصار عادة على الأخطاء المتعلقة بالتقديرات الرئيسية مثل المجاميع الاجمالية والمتوسطات والنسب المئوية. ويحسب هامش الخطأ المرتكب في التقدير، كما تم توضيحه سابقاً وفقاً لنظرية النزعة المركزية، بالاعتماد على تباين التقدير ومستويات الثقة المرغوبة، وتختلف أخطاء التقديرات المختلفة باختلاف تصميم العينة المستخدم (العينة العشوائية البسيطة، والعنقودية، والطبقية، والعينات المركبة). وتتضمن العديد من المنشورات الخاصة بتقنيات المعاينة، طرق حساب الأخطاء المرتكبة في العينات بناءً على تباين التقدير (حيث يعتبر الخطأ المرتكب في التقدير يساوي الجذر التربيعي لهذا التباين)، كما في أعمال (Cochran) و(Kish)، و(Hansen)، وغيرهم، وفي منشورات الأمم المتحدة في هذا المجال. وسيتم في الفقرات التالية عرض العلاقات الرياضية والأمثلة المناسبة لحساب هامش الخطأ للتقديرات الرئيسية في كل نوع من أنواع العينات.

4.1 - أخطاء العينة العشوائية البسيطة (Simple Random Sampling errors)

تعتبر العينة العشوائية البسيطة تطبيقاً مباشراً لنظرية النزعة المركزية التي عرضت سابقاً في القسم (1.1)، وهي الأساس الذي بنيت عليه طرق العينات الأخرى، فإذا تم سحب عينة عشوائية بسيطة من مجتمع حجمه N وكانت X_1, X_2, \dots, X_N قيم الظاهرة المدروسة في هذا المجتمع، فإن متوسط هذه الظاهرة يحسب بالعلاقة التالية

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

$$V = \sigma^2 = \frac{\sum_{i=1}^{i=N} (X_i - \bar{X})^2}{N} \quad \text{ويحسب تباينها (Variance) بالعلاقة المعروفة التالية}$$

ووفقاً لنظرية النزعة المركزية فإن التقدير المحسوب عن طريق العينة يتوزع وفقاً للتوزيع الطبيعي (Normal Distribution)، كما أن الفرق بين القيمة الحقيقية للتقدير في المجتمع المستهدف والتقدير المحسوب عن طريق العينة ويساوي الانحراف المعياري لتباين التقدير، وقد يكون هذا الفرق موجباً أو سالباً، ويرمز لتباين التقدير بالرمز $V(\bar{x})$ أو الرمز $\sigma_{(\bar{x})}^2$ لتباين المتوسط، وتباين النسبة بالرمز $V(p)$ أو $\sigma_{(p)}^2$ والرمز $V(T)$ أو $\sigma_{(T)}^2$ لتباين المجموع، وبناء على ذلك يعبر عن الخطأ المعياري لتقدير المتوسط بالعلاقة التالية، والعلاقات المشابهة الخاصة بالخطأ المعياري للنسبة وللمجموع.

$$SE = \pm \sigma_{(\bar{x})} = \pm \sqrt{V(\bar{x})}$$

ويمكن حساب هذا النوع من الأخطاء لجميع التقديرات التي تغطي عدداً كبيراً من المؤشرات، وفي التطبيقات العملية يتم الاقتصار عادة على الأخطاء المتعلقة بالتقديرات الرئيسية مثل المتوسطات والنسب المئوية والمجاميع النهائية.

4.1.1- مربع الخطأ العشوائي لمتوسط العينة

- تباين المتوسط (مع الإعادة)

$$V_{(\bar{x})} = \sigma^2_{(\bar{x})} = \frac{\sigma^2}{n}$$

- تباين المتوسط (بدون إعادة)

$$V_{(\bar{x})} = \sigma^2_{(\bar{x})} = \frac{\sigma^2}{n} * \left(\frac{N-n}{N-1} \right)$$

وعندما يكون حجم المجتمع كبيراً، فإن الفرق بين القيمة $N-1$ والقيمة N صغيراً جداً، وباستبدال ذلك في عبارة التباين بدون إعادة، تصبح

$$\sigma^2_{(\bar{x})} = \frac{\sigma^2}{n} * \frac{N-n}{N} = \frac{\sigma^2}{n} * \left(1 - \frac{n}{N} \right) = \frac{\sigma^2}{n} * (1-f)$$

حيث $f = \frac{n}{N}$ ويسمى كسر المعاينة (Sampling Fraction). ويسمى $1-f$ معامل تصحيح

المجتمع (PCF) (Population Correction Factor).

ويمكن ملاحظة أن التباين بدون إعادة أصغر من التباين مع الإعادة، ولهذا السبب يفضل استخدام العينة العشوائية بدون إعادة في التطبيقات العملية لأن أخطاءها أقل من أخطاء العينة العشوائية مع الإعادة.

وعندما يكون كسر المعاينة أصغر من 5% ويتحقق ذلك في المجتمعات الكبيرة، فيمكن اعتبار $1 - f = 1$ وعندئذ يكون تباين التقدير واحد في الحالتين.

4.1.2- مربع الخطأ العشوائي للنسبة في العينة

وينطبق ما سبق على نسبة ظاهرة (Proportion)، لأن النسبة هي حالة خاصة من المتوسط، حيث أن أية قيمة في المجتمع إما أن تكون 1 عندما توافق الظاهرة ذات الاهتمام، أو تساوي الصفر عندما تكون غير ذلك، وتحسب النسبة P بعلاقة المتوسط السابقة بحيث يكون البسط يساوي عدد الوحدات التي توافق الظاهرة ويكون المقام يساوي إجمالي عدد وحدات المجتمع (الموافقة للظاهرة وغير الموافقة)، وتكون النسبة المتممة Q تساوي $Q = 1 - P$ ويكون تباين النسبة وفقاً لدستور (Bernoulli)

$$V = \sigma^2 = P * Q$$

عندئذ يحسب تباين النسبة المقدر بشكل مشابه لتباين المتوسط المقدر، بإحدى العلاقتين التاليتين

- تباين النسبة (مع الإعادة)

$$V_{(p)} = \sigma^2_{(p)} = \frac{V}{n} = \frac{P * Q}{n}$$

- تباين النسبة (بدون إعادة)

$$V_{(p)} = \sigma^2_{(p)} = \frac{P * Q}{n} * \left(\frac{N - n}{N - 1} \right)$$

وعندما يكون حجم المجتمع كبيراً، فإن الفرق بين القيمة $N-1$ والقيمة N صغيراً جداً، وباستبدال ذلك في عبارة التباين بدون إعادة، تصبح

$$V_{(p)} = \sigma^2_{(p)} = \frac{P * Q}{n} * (1 - f)$$

4.1.3- مربع الخطأ العشوائي للمجموع

يقدر مجموع المجتمع T بناء على متوسط العينة \bar{x} بالعلاقة التالية

$$T = N * \bar{x}$$

وبالاعتماد على خواص التباين (إذا تم ضرب إحدى القيم بعدد ثابت فإن تباين القيمة الجديدة يساوي تباين القيمة الأساسية مضروباً بمربع العدد الثابت) يمكن كتابة العلاقة التالية لتباين المجموع المقدر بدلالة تباين المتوسط المحسوب بالعلاقات السابقة، الذي يعبر عن مربع الخطأ العشوائي لمجموع المجتمع

$$V(T) = V(N * \bar{x}) = N^2 * V(\bar{x})$$

ويمكن كتابة العلاقة المشابهة التالية لتباين مجموع T' لإحدى الظواهر ذات الاهتمام التي نسبتها p المقدر في العينة

$$V(T') = V(N * p) = N^2 * V(p)$$

مثال: بهدف تقدير متوسط الأجر اليومي للعامل في إحدى الشركات التي يعمل بها 10000 عامل، سحبت

عينة عشوائية بسيطة حجمها 50 عامل وكانت أجورهم اليومية كما يلي

| | | | | |
|-----|-----|-----|-----|-----|
| 104 | 129 | 146 | 121 | 123 |
| 122 | 131 | 147 | 150 | 129 |
| 121 | 145 | 127 | 149 | 147 |
| 102 | 122 | 138 | 138 | 142 |
| 150 | 145 | 131 | 111 | 132 |
| 132 | 123 | 103 | 140 | 126 |
| 135 | 114 | 130 | 129 | 139 |
| 122 | 106 | 146 | 102 | 102 |
| 145 | 115 | 108 | 124 | 119 |
| 134 | 115 | 113 | 107 | 101 |

وبتطبيق العلاقات السابقة سيتم تقدير متوسط أجر العامل في هذه الشركة والخطأ المرتكب في تقدير

المتوسط عن طريق العينة، علماً أن تباين أجر العامل في هذه الشركة من دراسات سابقة يساوي 450

$$\bar{x} = \frac{\sum_{i=1}^{i=50} x_i}{n} = 126.64$$

إن متوسط أجر العامل في العينة يساوي

ولحساب الخطأ المرتكب في تقدير المتوسط، يتم تطبيق العلاقات السابقة لحساب تباين المتوسط مع

الإرجاع وبدون إرجاع كما يلي

$$V_{(\bar{x})} = \sigma^2_{(\bar{x})} = \frac{\sigma^2}{n} = \frac{450}{50} = 9$$

تباين المتوسط مع الإرجاع

ويكون الخطأ المعياري المرتكب بتقدير المتوسط (مع الإرجاع) يساوي

$$SE = \pm \sigma_{(\bar{x})} = \sqrt{9} = 3$$

وتباين المتوسط بدون إرجاع

$$V_{(\bar{x})} = \sigma^2_{(\bar{x})} = \frac{\sigma^2}{n} * \left(\frac{N-n}{N-1} \right) = \frac{450}{50} * \left(\frac{10000-50}{10000-1} \right) = 8.96$$

ويكون الخطأ المعياري المرتكب بتقدير المتوسط (بدون إرجاع) يساوي 2.99 وهو أصغر من الخطأ

المعياري مع الإرجاع كما أشير إلى ذلك أعلاه، ولكن الفرق بينهما صغير جداً ($f = \frac{n}{N} = \frac{50}{10000} = 0.005$)،

بسبب حجم المجتمع الكبير وصغر كسر المعاينة كما أشير إلى ذلك سابقاً.

$$SE = \pm \sigma_{(\bar{x})} = \pm \sqrt{8.96} = 2.99$$

4.2- أخطاء العينة الطباقية (Stratified Sampling Errors)

يتم حساب أخطاء التقديرات المختلفة في كل طبقة بنفس الطريقة التي حسبت فيها أخطاء العينة العشوائية البسيطة التي تم عرضها في الفقرة السابقة باعتبار كل طبقة مجتمعاً قائماً بذاته. ويتم حساب أخطاء التقديرات المختلفة على مستوى المجتمع المستهدف بالاستفادة من خواص التباين لكل من المتوسط والنسبة على المستوى الإجمالي بدلالة التباين على مستوى الطبقات كم مبين في الفقرة التالية:

4.2.1- الخطأ المعياري في تقدير المتوسط

من المعروف أن المتوسط المقدر للمجتمع ويرمز له بالرمز $\bar{x}_{(st)}$ هو متوسط مرجح لمتوسطات

العينات المسحوبة من الطبقات، ويحسب بالعلاقة التالية

$$\bar{x}_{(st)} = \frac{\sum_1^l N_i * \bar{x}_i}{\sum_1^l N_i} = \frac{\sum_1^l N_i * \bar{x}_i}{N} = \sum_1^l W_i * \bar{x}_i$$

$$W_i = \frac{N_i}{N} \quad \text{و} \quad i = 1, 2, \dots, l \quad \text{حيث}$$

\bar{x}_i المتوسط المقدر في الطبقة i

ونظراً لأن الطبقات مستقلة فإن تباين متوسط المجتمع المقدر يحسب بالعلاقة التالية:

$$\sigma^2(\bar{x}_{st}) = \sigma^2\left(\sum W_i * \bar{x}_i\right)$$

وبالاعتماد على خواص التباين (إذا تم ضرب عدد من القيم بعدد ثابت فإن تباين القيم الجديدة يساوي

تباين القيم الأساسية مضروباً بمربع العدد الثابت) يمكن كتابة العلاقة التالية

$$\sigma^2(\bar{x}_{st}) = \sum W_i^2 * \sigma^2(\bar{x}_i)$$

وبناء على الفقرة السابقة، فإن $V(\bar{x}_i)$ هو تباين المتوسط المقدر في الطبقة i ويحسب في حالة بدون

إرجاع بالعلاقة السابقة

$$\sigma^2(\bar{x}_i) = \frac{\sigma_i^2}{n_i} * (1 - f_i)$$

وبناء على ذلك يتم حساب تباين المتوسط بدلالة تباين المتوسط في الطبقات بالعلاقة التالية

$$\sigma^2(\bar{x}_{st}) = \sum W_i^2 * \sigma^2(\bar{x}_i) = \sum W_i^2 * \frac{\sigma_i^2}{n_i} * (1 - f_i)$$

وتمثل هذه العلاقة مربع الخطأ العشوائي للمتوسط المقدر للمجتمع. وهو يتناسب طردياً مع التباين في الطبقات. فكلما كان التباين في الطبقات كبيراً، كلما كان مربع الخطأ العشوائي كبيراً، وبالعكس كلما كان التباين في الطبقات صغيراً، كلما كان مربع الخطأ العشوائي صغيراً. وهذا يقود إلى النتيجة الهامة التالية التي تعتبر الأساس الذي يبرر استخدام أسلوب المعاينة الطبقيّة، والتي يمكن التعبير عنها بما يلي: بغية تقليل الخطأ العشوائي باستخدام أسلوب المعاينة الطبقيّة ينبغي تقسيم المجتمع إلى طبقات متجانسة، وذلك لتقليل التباين في الطبقات. وكمثال على حساب الخطأ العشوائي في العينة الطبقيّة، يبين الجدول أدناه حجم العينة في مجتمع مقسم إلى طبقتين والتباين والمتوسط والنسبة في العينتين.

إن مربع الخطأ العشوائي يساوي 10.35 وفق العلاقة السابقة

$$\sigma_{(\bar{x}_{st})}^2 = \left(\frac{70}{100}\right)^2 * \frac{200}{10} \left(1 - \frac{10}{70}\right) + \left(\frac{30}{100}\right)^2 * \frac{100}{4} \left(1 - \frac{4}{30}\right) = 10.35$$

$$e = \sqrt{\sigma_{(\bar{x}_{st})}^2} = \sqrt{10.35} = 3.22 \quad \text{ويكون الخطأ العشوائي (المعياري) يساوي}$$

وبحساب المتوسط المقدر بالعلاقة السابقة الذي يساوي

$$\bar{x}_{(st)} = \frac{\sum_1^l N_i * \bar{x}_i}{\sum_1^l N_i} = \frac{70 * 15 + 30 * 22}{100} = 17.1$$

ويكون متوسط المجتمع وفقاً لمستوى ثقة 95% يقع في المجال

$$17.1 - 1.96 * 3.22 \leq \bar{X} \leq 17.1 + 1.96 * 3.22$$

$$10.8 \leq \bar{X} \leq 23.4$$

| P_i | \bar{x}_i | σ_i^2 | n_i | N_i | الطبقة |
|-------|-------------|--------------|-------|-------|--------|
| 0.20 | 15 | 200 | 10 | 70 | 1 |
| 0.35 | 22 | 100 | 4 | 30 | 2 |

ويمكن اتباع نفس الخطوات التي اتبعت في تقدير المتوسط لتقدير نسبة الظاهرة في المجتمع. فتقدير نسبة الظاهرة في المجتمع يعطي بالعلاقة التالية:

$$P_{(st)} = \frac{\sum_{i=1}^L N_i P_i}{\sum_{i=1}^L N_i} = \sum W_i * P_i$$

حيث P_i هي النسبة المقدرة لنسبة الظاهرة في الطبقة (i)، وبطريقة مشابهة يمكن البرهان على أن تباين نسبة الظاهرة المقدرة يعطى بالعلاقة التالية

$$\sigma_{(P_{st})}^2 = \sum W_i^2 * (1 - f_i) * \frac{P_i * Q_i}{n_i}$$

$$Q_i = 1 - P_i \quad \text{و} \quad f_i = \frac{n_i}{N_i} \quad \text{حيث}$$

وبتطبيق العلاقة الأخيرة يمكن حساب مربع الخطأ العشوائي للنسبة المقدرة للمجتمع

$$\sigma^2_{(P_{st})} = \left(\frac{70}{100}\right)^2 * \left(1 - \frac{10}{70}\right) * \frac{0.20 * 0.80}{10} + \left(\frac{30}{100}\right)^2 * \left(1 - \frac{4}{30}\right) * \frac{0.35 * 0.65}{4} = 0.0067 + 0.0044 = 0.0111$$

$$e = \sqrt{\sigma^2_{(P_{st})}} = \sqrt{0.0111} = 0.105 \quad \text{ويكون الخطأ العشوائي (المعياري) يساوي}$$

وتحسب النسبة المقدرة للمجتمع بالعلاقة السابقة

$$P = \frac{70}{100} * 0.20 + \frac{30}{100} * 0.35 = 0.14 + 0.105 = 0.245$$

ويمكن تحديد الحد الأعلى والأدنى لتقدير النسبة في المجتمع وفقاً لمستوى ثقة 95% بالعلاقة التالية

$$0.245 - 1.96 * e \leq P \leq 0.245 + 1.96 * e$$

4.3- أخطاء العينة العنقودية (Cluster Sampling Errors)

في التطبيقات العملية يمكن استخدام عناقيد ذات أحجام متساوية، أو استخدام عناقيد ذات أحجام غير متساوية، وقد تتم الدراسة على جميع وحدات العنقود أو على عدد محدد من وحدات العنقود. وسيتم في هذه الفقرة حساب الأخطاء العشوائية للتقديرات المختلفة (المتوسط، النسبة، والمجموع) للعناقيد المتساوية الحجم، وسيتم حساب الأخطاء العشوائية للعناقيد المختلفة الحجم في الفقرة اللاحقة.

4.3.1- العناقيد متساوية في الحجم

بفرض أن مجتمعا مؤلفا من A عنقود وعدد العناصر في كل عنقود يساوي B وقد تم اختيار a عنقودا

من هذا المجتمع، بالطريقة العشوائية بدون إرجاع. فيكون $n = a * B$ و $N = A * B$ ويمكن

اعتبار هذه العينة هي عينة عشوائية بسيطة بدون إعادة وتكون وحداتها هي العناقيد وعددها a . ونظراً لأن

العناقيد متساوية بالحجم فإن متوسط العينة يساوي متوسط القيم في جميع عناقيد، أو يساوي متوسط

المتوسطات في جميع العناقيد. كما في العلاقة التالية

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{\sum_{a=1}^a \bar{x}_a}{a}$$

وهي حالة مشابهة للعينة العشوائية البسيطة، بحيث يمكن اعتبار متوسطات العناقيد والتي عددها a هي وحدات العينة العشوائية البسيطة، ويكون تباين المتوسط (مربع الخطأ العشوائي) يساوي

$$\sigma^2(\bar{x}) = (1 - f) * \frac{s^2(a)}{a}$$

$$s^2(a) = \frac{\sum_{a=1}^a (\bar{x}_a - \bar{x})^2}{a - 1} \quad f = \frac{a}{A} \quad \text{حيث}$$

وبالتبديل يحسب مربع الخطأ العشوائي بالعلاقة التالية

$$\sigma^2(\bar{x}) = (1 - f) * \frac{\sum_{a=1}^a (\bar{x}_a - \bar{x})^2}{a * (a - 1)}$$

مثال: يبين الجدول التالي أجور 40 عاملاً موزعة في أربعة عناقيد حجم كل منها 10 عمال

| | C1 | C2 | C3 | C4 | |
|-------------------------|------------|------------|------------|------------|-------------|
| | 17 | 17 | 19 | 12 | |
| | 29 | 34 | 15 | 10 | |
| | 40 | 34 | 41 | 41 | |
| | 26 | 29 | 39 | 18 | |
| | 40 | 18 | 35 | 27 | |
| | 13 | 35 | 22 | 28 | |
| | 39 | 31 | 34 | 33 | |
| | 12 | 48 | 17 | 17 | |
| | 20 | 16 | 21 | 20 | |
| | 43 | 41 | 11 | 17 | |
| Total | 279 | 303 | 254 | 223 | 1059 |
| \bar{x} | 27.9 | 30.3 | 25.4 | 22.3 | 26.475 |
| $(\bar{X} - \bar{x})^2$ | 2.030625 | 14.63063 | 1.155625 | 17.43063 | 35.2475 |

بفرض أن $f = \frac{a}{A}$ صغيرة بحيث يمكن إهمالها وتطبيق العلاقة السابقة يكون تباين متوسط الأجر

$$\sigma^2(\bar{x}) = (1 - f) * \frac{\sum_{a=1}^a (\bar{x}_a - \bar{x})^2}{a * (a - 1)} = \frac{35.2475}{4 * (4 - 1)} = 2.937$$

ويكون الخطأ العشوائي في تقدير المتوسط يساوي 1.714 وهو الجذر التربيعي لتباينه.

ويحسب الخطأ العشوائي لتقدير النسبة بالعلاقة التالية المشابهة

$$\sigma^2(\bar{x}) = (1 - f) * \frac{\sum_{a=1}^a (p_a - p)^2}{a * (a - 1)}$$

فإذا كانت القيم الملونة (بالأحمر) في المثال السابق تمثل الاناث فإن نسبة الاناث تساوي 45%(18/40)،

| | C1 | C2 | C3 | C4 | |
|----------|--------|--------|--------|--------|-------|
| | 17 | 17 | 19 | 12 | |
| | 29 | 34 | 15 | 10 | |
| | 40 | 34 | 41 | 41 | |
| | 26 | 29 | 47 | 18 | |
| | 40 | 18 | 35 | 27 | |
| | 13 | 35 | 22 | 28 | |
| | 39 | 31 | 34 | 33 | |
| | 12 | 48 | 17 | 17 | |
| | 20 | 16 | 21 | 20 | |
| | 43 | 41 | 11 | 17 | |
| pi | 0.3 | 0.5 | 0.4 | 0.6 | 0.45 |
| (pi-p)^2 | 0.0225 | 0.0025 | 0.0025 | 0.0225 | 0.050 |

إن مربع الخطأ العشوائي لنسبة الاناث العاملات في هذا المصنع يساوي 0.00417 والخطأ العشوائي يساوي 0.065

$$\sigma^2(\bar{x}) = (1 - f) * \frac{\sum_{a=1}^a (p_a - p)^2}{a * (a - 1)} = \frac{0.05}{4 * (4 - 1)} = 0.00417$$

4.4- أخطاء العينة المركبة

في الواقع العملي، من النادر أن يتم استخدام طريقة من طرق المعاينة المعيارية، البسيطة أو الطبقيّة أو العنقودية، بشكل مستقل كما تم عرضه سابقاً. بل يتم استخدام أكثر من طريقة واحدة، وهو ما يطلق عليه اسم العينات المركبة، كما في العينة المتعددة المراحل، حيث يتم استخدام بعض أو كل هذه الطرق في آن واحد. إن حساب أخطاء العينة في هذه الحالة يصبح أكثر تعقيداً، ونظراً للتعقيدات الخاصة بحساب التباين، في

مثل هذه الحالات، فقد تم تطوير بعض المنهجيات والبرامج الخاصة بحساب هذا النوع من الأخطاء من قبل بعض المنظمات الدولية، ومراكز البحث العلمي وبعضها متاح بشكل مجاني. ومن هذه المنهجيات: برنامج (CENVAR) وبرنامج (CSPRO) وبرنامج (SUDAAN) وغيرها. ومن أكثر الطرق استخداماً لتقدير التباين في العينات المركبة، ما يعرف بطريقة العنقود النهائي (Ultimate Cluster Method)، حيث يتم التركيز على حساب التباين في المجموعات النهائية التي أنشئت من وحدات المعاينة الأولية، في العينة المتعددة المراحل، مهما كان عدد مراحل السحب العشوائي التي تمت قبل ذلك. وسيتم الاعتماد على المنشور الصادر عن الأمم المتحدة بعنوان (Designing Household Survey Samples: Practical Guidelines, 2008. F No.98) لحساب الخطأ العشوائي للتقديرات المختلفة (المجموع والمتوسط والنسبة والتناسب) بهذه الطريقة.

4.4.1- الخطأ العشوائي للمجموع

يتم حساب تباين تقدير المجموع الإجمالي على مستوى كل طبقة من الطبقات التي قسم إليها المجتمع المستهدف. بفرض أن C عدد العناقيد التي تم اختيارها في إحدى الطبقات، وأن X_c المجموع الموزون في العنقود C فيمكن كتابة المجاميع المختلفة للعناقيد في هذه الطبقة على الشكل التالي:

$$X_1, X_2, \dots, X_c$$

ومجموعها X على مستوى الطبقة يساوي:

$$X = \sum_{c=1}^C X_c$$

ويكون متوسط العناقيد \bar{X} يساوي مجموعها مقسوماً على عدد العناقيد أي:

$$\bar{X} = \frac{\sum_{c=1}^C X_c}{C}$$

ويحسب التباين في المجاميع السابقة، كما هو معروف، بالعلاقة التالية:

$$V(X_c) = \frac{\sum_{c=1}^C (X_c - \bar{X})^2}{C - 1}$$

ويكون تباين المجموع المقدر في هذه الطبقة من جميع العناقيد C يساوي:

$$V(X) = C * \frac{\sum_{c=1}^C (X_c - \bar{X})^2}{C - 1} * \left(1 - \frac{C}{T}\right)$$

حيث T عدد العناقيد في الطبقة، وأحياناً يكون $\left(\frac{C}{T}\right)$ صغيراً جداً بحيث يمكن إهماله من علاقة تباين المجموع.

4.4.2- حساب الخطأ العشوائي للتناسب

يمكن حساب المتوسطات والنسب عن طريق التناسب بين قيمتين X, Y ويعبر عن ذلك رمزياً بالرمز:

وفي حالة العينة المتعددة المراحل تكون المتوسطات والنسب حالات خاصة لهذا التناسب، ففي حالة $R = \frac{Y}{X}$

المتوسط يكون البسط مساوياً لقيم الخاصة الجاري تقديرها، ويكون المقام يساوي 1 لكل عنصر وبذلك

يكون المقام يساوي مجموع الترجيحات (الأوزان) لجميع العناصر. وفي حالة النسبة فإن المقام أيضاً يساوي

1 لجميع العناصر، ويكون كل عنصر في البسط إما يساوي 1 عندما يكون العنصر يوافق الخاصة الجاري

تقديرها ويساوي الصفر إذا كان غير ذلك. ويحسب تباين التناسب R في إحدى الطبقات بالعلاقة التالية:

$$V(R) = \frac{1}{X^2} \{V(Y) + R^2V(X) - 2 * R * cov(Y, X)\}$$

ويحسب تباين المجموع $V(X), V(Y)$ كما في الفقرة السابقة، ويحسب معامل التغير cov بالعلاقة التالية

$$cov(Y, X) = \frac{c}{c-1} \left\{ \sum \left(Y_i - \frac{Y}{c} \right) \left(X_i - \frac{X}{c} \right) \right\} * \left(1 - \frac{c}{T} \right)$$

أو العلاقة

$$cov(Y, X) = \frac{c}{c-1} \left\{ \sum Y_i X_i - \frac{(\sum Y_i)(\sum X_i)}{c} \right\} * \left(1 - \frac{c}{T} \right)$$

المراجع

- 1 - Cochran, W.G, Sampling Techniques, New York 1977.
- 2- Dalenus Elements of Survey Sampling, Swedish Agency for Research Cooperation with Developing Countries, 1985.
- 3 - Hansen, M.H, Hurwitz, WN. and Madow. Sample Survey Methods and Theory, Vol. I. Methods and Applications, Vol, II. Theory, Willey, New York, 1953.
- 5 - Kish, L., Survey Sampling, Willey New York, 1965.
- 6 - International Statistical Institute , World Fertility Statistical Institute , World Fertility Survey , guidelines for Country Report No. 1 , Basic Documentation , The Hague , 1977 .
- 7 - Yates, F., Sampling Methods for Census and Surveys, 4th, Griffin, London, 1981.
- 8 - Sampling frames and sample designs for integrated household survey programmes 1986
Household Survey Capability Program, New York United Nations. Department of Technical Cooperation for Development and Statistical Office.
- 9 - League of Arab States (1990), Sampling Manual, Arab Maternal and Child Health Survey
- 10 - Macro International Inc. (1996). Sampling Manual. DHS-III Basic
- 11 - Household Sample Surveys in Developing and transition Countries UN 2005
- 12-Designing Household survey Samples: Practical Guidelines. United Nation, New York, 2008.