حكومة دبي
GOVERNMENT OF DUBAI

مركز دبي للإحصاء
Dubai Statistics Center

دليل حجم العينة

رؤيتنا | إحصاء داعم لصناعة السعادة والتنمية المستدامة.
Our Vision | Statistics that support creating happiness and sustainable development.

dsc.gov.ae

المحتويات

مقدمة

في إطار سعي مركز دبي للإحصاء لتحقيق اعلى درجات الجودة للبيانات الإحصائية، أعد المركز العديد من الأدلة والمنهجيات الإحصائية التي تغطي جميع مجالات العمل الاحصائي، ويأتي إعداد دليل حجم العينة في هذا السياق، بهدف توثيق المنهجيات المتعلقة بحساب حجم العينة للمسوح الاقتصادية والاجتماعية والديموغرافية واستطلاعات الرأي التي ينفذها المركز، ولتكون دليلاً للعاملين في وحدة العينات في المركز وتعريف مستخدمي البيانات الإحصائية بالأسس العلمية لحساب حجم العينة للمسوح المختلفة.

إن تحديد حجم العينة يشكل الخطوة الأولى في التحضير والاعداد لتنفيذ المسوح، 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 الاساس أيضاً المستلزمات المادية والبشرية، والزمن اللازم لتنفيذ المسح. ويتوفر العديد من المنشورات والمطبوعات الخاصة بالعينات الإحصائية والتطبيقات العملية لحساب حجم العينة والتي تم الاعتماد عليها في هذا الدليل، كما في أعمال العديد من المختصين أمثال (Kish)، (Cochran)،(Nyman)، (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 2005 (Household sample surveys in Developing and transition Countries) 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 (World Fertility Survey) الذي نفذ في العديد من بلدان العالم منذ عام 1974، والمسح الديموغرافي الصحي (DHS) (Demographic and Health Survey) الجاري تنفيذه في العديد من بلدان العالم منذ 1984.

1. حجم العينة العشوائية البسيطة

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 المعاينة الأخرى، ويمكن الرجوع إلى دليل المعاينة الإحصائية في موقع مركز دبي للإحصاء للتعرف على الأساس النظري للعينة العشوائية البسيطة، كمقدمة للموضوعات الخاصة بحساب حجم العينة، التي سيتم عرضها في هذا الدليل. ويتم التمييز عادة بين طريقة سحب وحدات العينة، مع الإعادة (مع الارجاع) حيث تعاد الوحدة المسحوبة إلى المجتمع قبل السحب التالي (With Replacement)، وبدون إعادة (بدون إرجاع) حيث لا تعاد الوحدة المسحوبة إلى المجتمع في السحب التالي (Without Replacement). بناء على نظرية النزعة المركزية (Central Limit Theory)، إذا تم تشكيل جميع العينات العشوائية التي حجم كل منها n من مجتمع حجمه N وكانت متوسطات هذه العينات التالية

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

حيث: L عدد العينات العشوائية التي يمكن تشكيلها، وبناء على نظرية النزعة المركزية، فإن متوسطات العينات العشوائية تتوزع وفقاً للتوزيع الطبيعي (Normal Distribution)، 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:

متوسط هذا التوزيع يساوي متوسط المجتمع

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

تباين متوسط هذا التوزيع مع الإعادة (مع الارجاع) يساوي $\sigma^2(\bar{x}) = \frac{\sigma^2}{n}$ حيث σ^2 تباين

المجتمع تباين متوسط هذا التوزيع بدون إعادة (بدون إرجاع) $\sigma^2(\bar{x}) = \frac{\sigma^2}{n} * \frac{N-n}{N-1}$

ويلاحظ أن تباين المتوسط في حالة بدون إعادة يختلف عنه في حالة مع الاعادة 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 (Population Correction Factor) (PCF).

1.1-حجم العينة (مع الارجاع) لتقدير المتوسط والنسبة

- **حجم العينة لتقدير المتوسط**

بهدف التمييز بين حجم العينة مع الإرجاع وبدون إرجاع سيتم الرمز ب 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 الارجاع) وفقاً لنظرية النزعة المركزية هي العلاقة التالية:

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

وبذلك يحسب حجم العينة مع الإرجاع لتقدير المتوسط بالعلاقة التالية

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

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 $X_1, X_2, \ldots\ldots\ldots, X_N$ في المجتمع المستهدف، ويحسب بالعلاقة التالية $\bar{X} = \frac{\sum_{i=1}^{i=N} X_i}{N}$

- e_0 هامش الخطأ (Margin of Error) المتوقع في تقدير متوسط المجتمع، ويعبر عن إجمالي الفرق المتوقع بين القيمة الحقيقية في المجتمع المستهدف ونفس القيمة المقدرة عن طريق العينة وقد يكون هذا الفرق موجباً أو سالباً، ويحدد هامش الخطأ بدلالة Z وبما يسمى الخطأ المعياري (Standard Error) (SE) بالعلاقة العلاقة التالية

$$e_0 = \pm Z * SE$$

ويعرف الخطأ المعياري بانه 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 $Mo\,\mathrm{Re}\,l$

ويحسب بالعلاقة $Mo\,\mathrm{Re}\,l_{(\bar{x})} = \frac{e}{\bar{x}}$ لهامش الخطأ النسبي للمتوسط، والعلاقة المشابهة

$Mo\,\mathrm{Re}\,l_{(P)} = \frac{e}{p}$ لهامش الخطأ النسبي للنسبة. وبشكل مشابه يمكن حساب النسبة بين الخطأ المعياري

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

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

للنسبة. فإذا كان مستوى الثقة 95% فينبغي الاّ 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 طرداً مع تباين الظاهرة ومستوى الثقة، ويتناسب عكساً مع مربع هامش الخطأ.

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

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

وتجدر الاشارة 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

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

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

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

أو يحسب بدلالة معامل الاختلاف CV (Coefficient of Variation) حيث $\bar{X}$ متوسط الظاهرة ومعامل الاختلاف $CV = \frac{\sigma}{\bar{X}}$ الذي يساوي $CV = \frac{\sigma}{\bar{X}} = \frac{\sqrt{4000}}{150} = 0.42164$

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

والنتيجة واحدة في جميع الحالات

- **حجم العينة 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 لدستور (Bernoulli) $V = \sigma^2 = P * Q$

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

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

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 18-40على سبيل المثال)، وكان هامش الخطأ المتوقع 10% من النسبة 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.

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

1.2-حجم العينة (بدون إرجاع) لتقدير المتوسط أو النسبة:

- **حجم العينة لتقدير المتوسط**

إن العلاقة التي تربط بين حجم العينة ومربع هامش الخطأ في العينة العشوائية البسيطة (بدون إرجاع)، وفقاً لنظرية النزعة المركزية 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 الاخيرة ب $\frac{N-1}{N}$ 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 الاخيرة، 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 ب n_0 تصبح العلاقة

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

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 الارجاع.

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 النسبة $\frac{N-1}{N}$ تكون قريبة من الواحد الصحيح، وفي هذه الحالة يمكن كتابة العلاقة التالية لحجم العينة بدون إرجاع، ويمكن تسميتها (المختصرة للعينات الكبيرة)

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

وبتطبيق العلاقة المختصرة للمجتمعات الكبيرة الحجم يكون حجم العينة يساوي 216 ايضاً، لأن $\frac{N-1}{N} = 0.999$ يساوي 1 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 الاعادة وبدون إعادة يكون صغيراً، فإذا كان حجم المجتمع في هذا المثال يساوي 100000فإن

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

وهو حجم العينة مع الارجاع المحسوب سابقاً، لأن النسبة $\frac{n_0}{N}$ صغيرة جدً وأقل من 5% بكثير.

- **حجم العينة لتقدير النسبة**، وبنفس الخطوات السابقة لحساب حجم العينة للمتوسط، يمكن إيجاد علاقة مطابقة للعلاقتين الاخيرتين (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

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

أو العلاقة المختصرة للمجتمعات الكبيرة الحجم

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

فحجم العينة بدون إرجاع بتطبيق العلاقة المختصرة للعينات الكبيرة في المثال 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 عد م الاستجابة، يساوي حجم العينة المحسوب سابقاً مضروباً بمعامل عدم الاستجابة أي إن حجم العينة ينبغي أن يكون 304 للحصول على حجم العينة المطلوب 274.

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

1.3-حجم العينة لتقدير عدد من المؤشرات

في الفقرة السابقة، تم حساب حجم العينة لتقدير المتوسط أو النسبة لظاهرة محددة، ومن النادر، في التطبيقات العملية، اقتصار الهدف من المسح الاحصائي على تقدير متوسط أو نسبة ظاهرة وحيدة، بل يغطي الهدف من المسح تقدير العديد من المؤشرات التي يمكن أن تكون عادة على شكل نسب مئوية أو

متوسطات. وسيكون حجم العينة المناسب لتقدير أحد هذه المؤشرات مختلفاً عن حجم العينة لتقدير المؤشرات الاخرى، والسؤال الذي يطرح نفسه، ما هو حجم العينة المناسب لبعض هذه المؤشرات أو

لجميعها؟ وقد تم معالجة بعض جوانب هذه المسألة في المنشورات الخاصة بتقنيات المعاينة، كما في أعمال (Kish)، (Cochran) وغيرهما، وذلك عن طريق حساب حجم العينة لعدد محدود من المؤشرات الرئيسية، واعتماد حجم العينة الأكبر وفق مستوى ثقة محدد.

وفي هذا السياق يمكن الاستفادة من خصائص تباين النسبة أو المتوسط لتقدير حجم العينة المناسب لجميع المؤشرات المتوقعة من المسح، وليس لعدد محدود فقط، وسيتم معالجة ذلك في حالة الاعادة، نظراً للعلاقة التي تربط بين حجم العينة مع الاعادة وحجم العينة بدون إعادة، من أجل نفس مستوى الثقة وهامش الخطأ، لأن حجم المجتمع المستهدف (العامل الثاني في هذه العلاقة ثابت).

وبالعودة إلى حجم العينة (مع الإرجاع) لتقدير النسبة

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

وبفرض أن $a^2 = \frac{Z^2}{{e^2}_0}$ وبما أن $Q = 1 - P$ وبتبديل ذلك يصبح حجم العينة

$$n_0 = a^2 * p * (1 - p) \qquad (7)$$

والعلاقة الأخيرة هي دالة لحجم العينة باعتبار النسبة كمتحول مستقل، وتمثل معادلة قطع مكافئ يبلغ نهايته العظمى عندما تكون $p = 0.50$ ، مع الأخذ بالاعتبار أن النسبة اقل أو تساوي الواحد وأكبر او تساوي الصفر $0 \leq p \leq 1$ ويبين الشكل (1) أدناه الخط البياني لهذه الدالة الموافق للقيمتين $Z = 1.96$ و $e_0 = 0.05$

وبالعودة إلى حجم العينة لتقدير المتوسط $n = Z^2 * \frac{\sigma^2}{e^2}$ وبهدف المقارنة مع حجم العينة للنسبة، وبفرض أن هامش الخطأ النسبي لتقدير المتوسط $\frac{e}{\overline{X}}$ يساوي هامش الخطأ لتقدير النسبة أي e_0

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

وبالتبديل تصبح علاقة حجم العينة لتقدير المتوسط

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

وكما هو معروف، إن مربع معامل الاختلاف للمتوسط

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

وبالتبديل تصبح علاقة حجم العينة لتقدير المتوسط

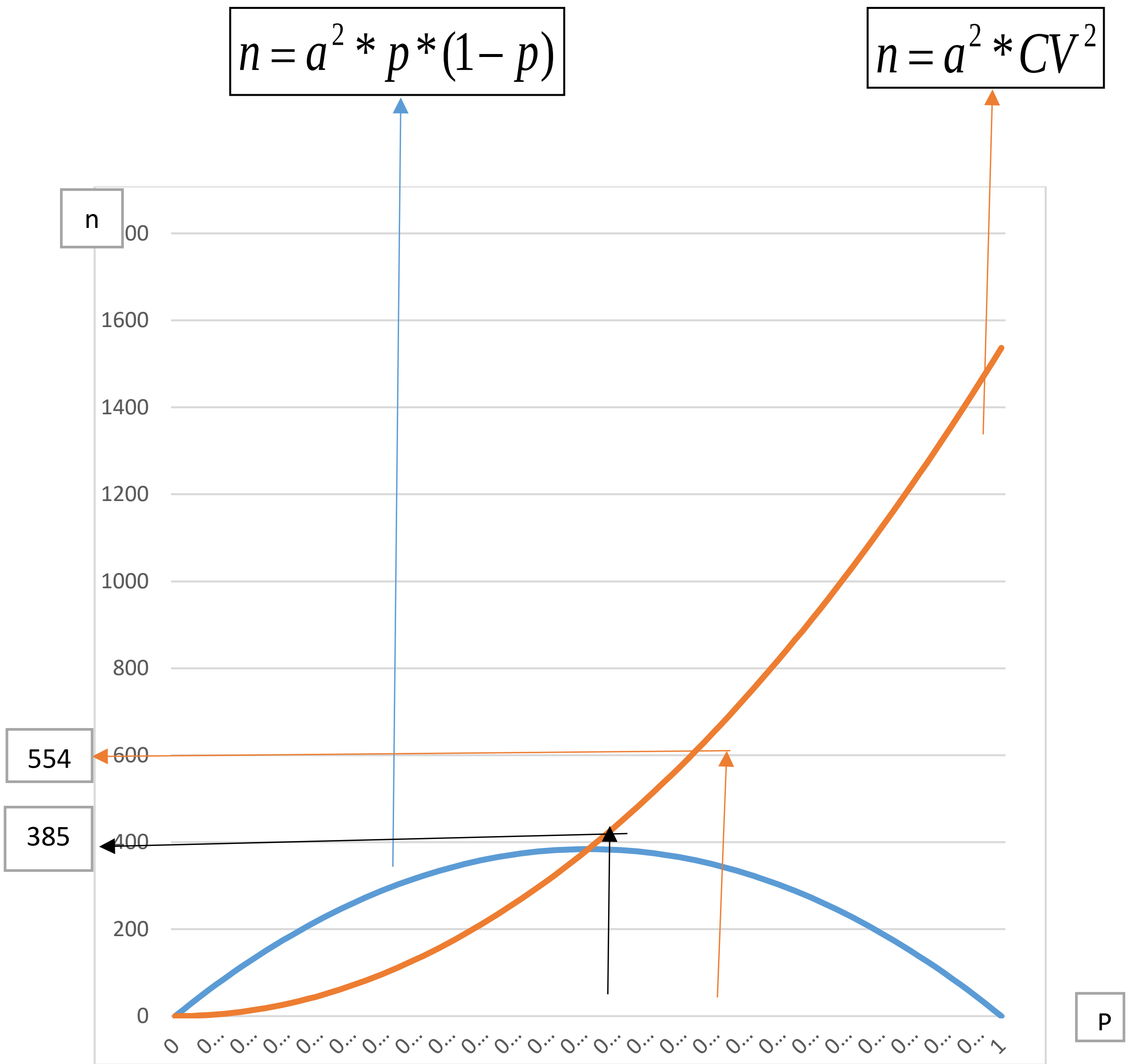
$$n = Z^2 * \frac{CV^2}{e_0^{\ 2}}$$

وباعتبار $a^2 = \frac{Z^2}{e^2_0}$ كما سبق تصبح العلاقة الأخيرة

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

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 (1). ويمكن الحصول على العديد من الأشكال المشابهة للشكل (1) بتغيير هامش الخطأ ومستوى الثقة، ومن الشكل (1) يمكن 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 الكلي 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 $V = P*(1-P) = 0.40*0.60 = 0.24$



الشكل (1)

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.

كما تقوم العديد من المواقع الالكترونية بنشر العلاقة الرياضية السابقة لحساب حجم العينة مباشرة، ويمكن الاطلاع على بعض هذه المواقع على العنوان التالي (Sample Size Calculation).

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 العظمى 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 الاعظمي 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 العظمى 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 أعظم ومناسب لتقدير النسبة والمتوسط.

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 أعظم لتقدير جميع المتوسطات كما

في حج م 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 الأعظم في كل طبقة لتقدير النسبة.

2-حجم العينة للمسوح الأسرية

في العديد من المسوح الاجتماعية والديموغرافية يكون المجتمع المستهدف فئة محددة من السكان(أفراد من القوى العاملة في مسح البطالة، الأشخاص في عمر معين، 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 بالعلاقتين (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 r ومتوسط حجم الاسرة $\overline{n}$ فيكون حجم عينة الأسر يساوي حجم عينة الأفراد مقسوماً على $r*\overline{n}$ (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:

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

والعلاقة المشابهة لحجم عينة الأسر مع الإرجاع لتقدير النسبة

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

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 $\overline{n} = 5$ وكانت نسبة عدم الاستجابة 15%، أي ان نسبة الاستجابة المتوقعة 85% فإن حجم عينة الأسر بتطبيق العلاقة (9) يساوي 643 أسرة.

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

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

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

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 الارجاع) 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.

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

3-حجم العينة لأنواع العينات الأخرى

إن حجم العينة المحسوب بالعلاقات السابقة يمثل حجم العينة العشوائية البسيطة، وفي الواقع العملي يتم الاعتماد على عدد آخر من أنواع العينات مثل الطبقية 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، الا 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

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

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.

4-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 الطبقية، 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. وقد يتضاعف 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 خطا 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 $N=\sum_{i=1}^{h}$ وإذا كان σ_i^2 تباين الظاهرة ذات الاهتمام في الطبقة i فإن توزيع حجم عينة محدد n على الطبقات التي قسم اليها المجتمع المستهدف فإن توزيع العينة على الطبقات يتم بإحدى طرق التوزيع التالية المقترحة في منشورات العينات.

1. **التوزيع الأمثل أو توزيع نيمان (Optimal or Neyman Allocation)**، يعتبر هذا التوزيع الأكثر استخداماً في التطبيقات العملية، ويتم توزيع العينة على الطبقات، وفقاً لهذه الطريقة، بحيث يكون تباين التقديرات المختلفة $\sigma^2{}_{st}$ على مستوى المجتمع المستهدف أصغر ما يمكن.

2. 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 or 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 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 $0 \le \alpha \le 1$ 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 (بانكير) قد تم تخصيص الطبقة الأولى (صغيرة الحجم) بعينة حجمها 52 وحدة مقابل تخصيصها بعينة حجمها 21 وحدة بطريقة نيمان 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 نيمان	التوزيع المتناسب	توزيع بانكير
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

المـراجع

1-Cochran, W.G, Sampling Techniques, New York 1977.

2-Dalenius, Elements of Survey Sampling, Swedish Agency for Research Cooperation with Developing Countries, 1985

3-Kish. L Survey Sampling, Willey New York, 1965.

4-International Statistical Institute، World Fertility Statistical Institute، World Fertility Survey guidelines for Country Report No. 1 , Basic Documentation , The Hague , 1977 .

5-Yates، Sampling Methods for Census and Surveys، London, 1981.

6-Sampling frames and sample designs for integrated household survey programmes1986

7-Household Survey Capability Program، New York United Nations. Department of Technical Cooperation for Development and Statistical Office.

8-Macro International Inc. (1996). Sampling Manual. DHS-III Basic

9-Household Sample Surveys in Developing and transition Countries UN 2005

10-Designing Household survey Samples: Practical Guidelines. United Nation, New York, 2008.