كيفية حساب حجم العينة n من مجتمع طبيعي حجمه N عنصراً

إن حجم العينة n يرتبط بعدة أمور وهي:

أ- بنوع المعاينة: المعاينة البسيطة- المعاينة الطبقية- المعاينة العنقودية ...الخ

هنا سنقوم بحساب حجم العينة n في المعاينة العشوائية البسيطة فقط .

ب- بطريقة السحب: السحب مع الإعادة- السحب بدون إعادة .

ج- بنوع المؤشر المدروس: ونقصد به المؤشر الذي نريد تقديره من معلومات العينة وهو غالباً ما يكون أحد المؤشرين التاليين:

- متوسط المجتمع μ والذي سنقدره بمتوسط العينة $\bar{x}$.
- النسبة في المجتمع R والتي سنقدرها بالنسبة في العينة r .

د- بدرجة تجانس المجتمع: ونعبر عنه من خلال تباين المجتمع σ^2 فإذا كان التباين كبيراً (التجانس ضعيفاً) فإن حجم العينة n يجب أن يكون كبيراً أي أن حجم العينة n يتناسب طرداً مع التباين σ^2

هـ-بمقدار الدقة المطلوبة في التقدير: ويرمز له بـ d وهو الحد الأعلى للخطأ المسموح به عند تقدير المؤشرات: ويتم تحديده مسبقاً من قبل المسؤولين حسب نوع المؤشرات المدروس . فإذا كان المؤشر المطلوب تقديره هو متوسط الدخل في المجتمع μ فإننا نضع مثلاً مقدار الدقة (d=1000) أو أي رقم آخر .

وهذا يعني أنه عند تقدير متوسط الدخل يجب أن يكون الخطأ المرتكب فيه (الخطأ المعياري) أقل أو يساوي المقدار المحدد للدقة d .

أما عند تقدير النسبة R في المجتمع فإن الدقة تحدد على شكل نسبة مئوية مثل d=6% أو d=3% أو d=2% . وهذا يعني أنه عند تقدير النسبة R يجب أن يكون الخطأ المرتكب فيه (الخطأ المعياري) أقل أو يساوي المقدر المحدد للدقة d .

و- بالخطأ المعياري للمؤشر المدروس: وهو يحسب لكل مؤشره على حدة ويبرهن في نظرية العينات على ما يلي:

- إن الخطأ المعياري المرتكب في تقدير متوسط المجتمع μ من خلال متوسط العينة $\bar{x}$ والذي سنرمز له بـ $\sigma_{\bar{x}}$ يساوي حسب حالتي السحب ما يلي:

(1) (في حالة السحب مع الإعادة) $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \approx \frac{s}{\sqrt{n}}$

حيث s^2 هو تباين العينة ويحسب من العلاقة : $s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1}$

$$\sigma_{\bar{x}} = \sqrt{\frac{N-n}{N}} * \frac{\sigma}{\sqrt{N}} \approx \sqrt{\frac{N-n}{N}} * \frac{s}{\sqrt{N}} \qquad \text{(2)} \quad \text{(في حالة السحب بدون إعادة)}$$

- إن الخطأ المعياري المرتكب في تقدير النسبة R في المجتمع من خلال النسبة في العينة r والذي سنرمز له بـ σ_r يساوي حسب حالتي السحب ما يلي:

$$\sigma_r = \sqrt{\frac{R*(1-R)}{n}} \approx \sqrt{\frac{r*q}{n}} \qquad \text{(3)} \quad \text{(في حالة السحب مع الإعادة)}$$

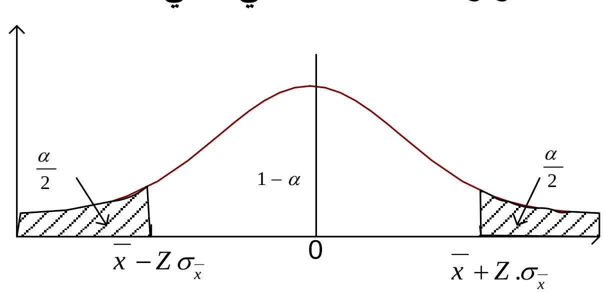
حيث r هي النسبة في العينة أو في أية عينة تجريبية و q=1-r

$$\sigma_r = \sqrt{\frac{N-n}{N}} * \sqrt{\frac{r.q}{n}} \qquad \text{(4)} \quad \text{(في حالة السحب بدون إعادة)}$$

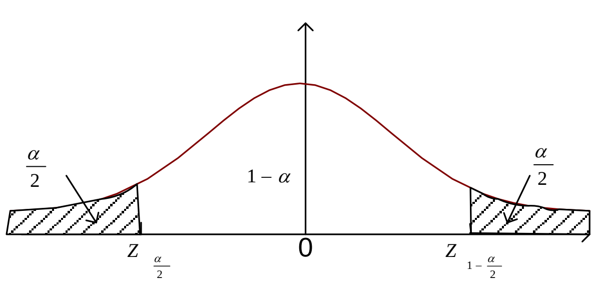
ز– بمستوى الدلالة المطلوب ∝ : ويتم تحديده مسبقاً من قبل المسؤولين وهو يعبر عن احتمال أن يكون التقدير المأخوذ من العينة مرفوضاً، أو أن يكون مقبولاً باحتمال $(1-\propto)$، ويسمى الاحتمال $(1-\propto)$ باحتمال الثقة في التقدير، لأنه يضمن لنا أن يكون المؤشر المطلوب تقديره واقعاً ضمن مجال محدد يسمى مجال الثقة باحتمال $(1-\propto)$. وعندما نحدد مستوى الدلالة ∝ نوزعه على طرفي المجال ونضع لكل طرف $\frac{\propto}{2}$. وعندها تتحدد فعنا القيمتين العدديتين لمتحول التوزيع الطبيعي المعياري $Z_{\frac{\propto}{2}}$ و $Z_{1-\frac{\propto}{2}}$ الفاصلة بين منطقتي القبول والرفض ويكون لدينا $Z_{1-\frac{\propto}{2}} = -Z_{\frac{\propto}{2}}$ لذلك نأخذ القيمة $Z_{1-\frac{\propto}{2}}$ بمجال الثقة المقابل لذلك . وتكون منطقة الرفض خارج مجال الثقة ولإنشاء مجال الثقة لمتوسط المجتمع μ نجعل مركزه متوسط العينة $\bar{x}$ ونصف طوله يساوي $(Z * \sigma_{\bar{x}})$ فنحصل على مجال الثقة الذي يحقق احتمال الثقة المطلوب وهو المجال التالي:

$$\mathrm{P} = \left[\bar{x} - Z_{1-\frac{\propto}{2}} * \sigma_{\bar{x}} \le \mu \le \bar{x} + Z_{1-\frac{\propto}{2}} * \sigma_{\bar{x}}\right] = 1-\propto \qquad (5)$$

وهو يأخذ الشكل البياني التالي :



الشكل (1) الشكل العام لمجال الثقة للمتوسط



الشكل (2) الشكل المعياري لمجال الثقة للمتوسط

ولإنشاء مجال الثقة للنسبة R في المجتمع نجعل مركزه النسبة في العينة r ونصف طوله $(Z - \sigma_r)$ فحصل على مجال الثقة الذي يحقق احتمال الثقة $(1-\propto)$ التالي:

$$P = \left[r - Z_{1-\frac{\propto}{2}} * \sigma_r \le R \le r + Z_{1-\frac{\propto}{2}} * \sigma_r\right] = 1-\propto \quad (6)$$

ح- ولحساب حجم العينة n في المعاينة العشوائية البسيطة علينا أن نراعي جميع الشروط المذكورة وخاصة شرط الدقة d (الحد الأعلى للخطأ المسموح به عند التقدير) وهنا سنميز بين حالتي السحب (إعادة وبدون إعادة) وسنحسب حجم العينة n لكل مؤشر على حدة (المتوسط المجتمع μ والنسبة فيه R) .

أولاً: حالة السحب مع الإعادة :

1- إذا كان المطلوب تقدير متوسط المجتمع الطبيعي μ وكان السحب مع الإعادة وكان فقط الدقة المطلوبة d . فإننا نجعل نصف طول مجال الثقة للمتوسط μ أصغر أو يساوي مقدار الدقة المحدد d فحصل على ما يلي:

$$Z_{1-\frac{\propto}{2}} * \sigma_r \le d \quad (7)$$

حيث أن $\left(\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}\right)$ نجد أن :

$$Z^2 * \sigma_{\bar{x}}^2 \le d^2$$

$$Z^2 * \sigma_{\bar{n}}^2 \le d^2$$

ومنها نجد أن :

$$n \ge \frac{Z^2 * \sigma^2}{d^2} \approx \frac{Z^2 s^2}{d^2} \quad (8)$$

(وهو الحد الأدنى لحجم العينة)

حيث أن s^2 هو تباين العينة أو أي عينة تجريبية عشوائية .

2- كان المطلوب تقدير النسبة R في المجتمع وكان السحب مع الإعادة فإننا نجعل نصف طول مجال الثقة أصغر أو يساوي الدقة المحددة d لنسبة مئوية صغيرة مثل (2% أو 3% ....) .

$$Z_{1-\frac{\propto}{2}} * \sigma_r \le d \quad (9)$$

وبما أن $\left(\sigma_{\bar{x}}^2 = \frac{R(1-R)}{n}\right)$ نجد ن:

$$Z^2 * \sigma_r^2 \le d^2$$

$$d^2 \ge Z^2 \frac{R(1-R)}{n} \approx Z^2 \frac{r*q}{n} \quad (10)$$

$$n \ge \frac{Z^2 * r * q}{d^2}$$

حيث أن r هي النسبة في العينة و q=1-r وهو الحد الأدنى لحجم العينة .

ثانياً: حالة السحب بدون إعادة :

1- إذا كـان المطلـوب تقدير متوسـط المجتمـع الطبيعـي μ وكـان السـحب بـدون إعـادة وكـان مقدار الدقـة المطلوبة d . فإننا نجعل نصف طول مجال الثقة أصغر أو يساوي مقدار الدقة d فحصل على ما يلي:

$$Z_{1-\frac{\alpha}{2}} * \sigma_{\bar{x}} \le d \qquad (11)$$

$$Z^2 * \sigma_{\bar{x}}^2 \le d^2$$

وبما أن $\sigma_{\bar{x}}^2 = \frac{N-n}{N} * \frac{\sigma n^2}{n}$ نجد أن

$$Z^2 * \frac{N-n}{N}\frac{\sigma^2}{n} \le d^2$$

ومنها نجد أن :

$$n * N * d^2 \ge Z^2 * N\sigma^2 - Z^2 * n\sigma^2$$

$$nNd^2 + Z^2 * n\sigma^2 \ge Z^2 * N\sigma^2$$

$$n(Nd^2 + Z^2) \ge Z^2 * N * \sigma^2$$

$$n \ge \frac{N * Z^2 * \sigma^2}{Nd^2 + Z^2 * \sigma^2} \qquad (12)$$

وباستبدال σ^2 بتباين العينة s^2 نحصل على مايلي:

$$n \ge \frac{NZ^2 * s^2}{Nd^2 + Z^2 s^2} \qquad (13)$$

وهو الحد الأدنى لحجم العينة (بدون إعادة)

ملاحظـة: مـن مقارنـة العلاقتين (8) و (13) نجد أن حجـم العينـة n المحسوبة فـي حالـة السحب بدون إعادة أصغر من حجمها n_0 في حالة السحب مع الإعادة وهما يرتبطان بالعلاقة $n = \frac{N * n_0}{N + n_0}$.

وهذه العلاقـة هـي العلاقـة الأساسـية لحسـاب حجـم العينـة فـي حالـة السـحب بـدون إعـادة ولكـن بعـض المؤلفين والباحثين قام بتحويرها إلـى أشكال أخرى . فقام أحدهم بتقسيم البسط والمقام علـى σ^2 (أو علـى s^2) فحصل على أن:

$$n \ge \frac{NZ^2}{N\left(\frac{d^2}{\sigma^2}\right) + Z^2} \approx \frac{NZ^2}{Ne^2 + Z^2} \qquad \text{حيث أن } \left(e^2 = \frac{d^2}{s^2}\right) \quad (14)$$

وقام بعضهم بتقسيم البسط والمقام على ($Z^2\sigma^2$) فحصل على أن :

$$n \ge \frac{N}{N\left(\frac{d^2}{Z^2\sigma^2}\right) + 1} \approx \frac{N}{NE^2 + 1} \qquad \text{حيث أن } \left(E^2 = \frac{d^2}{Z^2 * s^2}\right) \quad (15)$$

وقام بعضهم بتقسيم البسط والمقام على $\bar{x}^2$ وادخل معامل الاختلاف في المعادلة (15) فحصل على أن:

$$N \ge \frac{NZ^2\frac{s^2}{\bar{x}^2}}{N\left(\frac{d^2}{\bar{x}^2}\right) + Z^2 * \frac{s^2}{\bar{x}^2}} = \frac{NZ^2 * CV^2}{N\sigma^2 + Z^2 * CV^2} \qquad \text{حيث أن (} CV = \frac{s}{\bar{x}} \text{ وأن } \sigma = \frac{d}{x} \text{ الدقة النسبية)} \quad (16)$$

وقام بعضهم بتقسيم البسط والمقام على N فحصل على أن:

$$n \ge \frac{Z^2\sigma^2}{d^2 + \frac{Z^2\sigma^2}{N}} \approx \frac{Z^2 s^2}{d^2 + \frac{Z^2 s^2}{N}} \qquad (17)$$

ومنها استنتج أنه عندما يكون حجم المجتمع N كبيراً فإن المقدار $\frac{Z^2s^2}{N}$ يصبح مهملاً والعلاقة السابقة تأخذ الشكل التالي :

$$n \geq \frac{Z^2s^2}{N} \tag{18}$$

وهي نفس العلاقة لحالة السحب مع الإعادة . ومنها يمكننا أن نستنتج مايلي:

نتيجة هامة. عندما يكون حجم المجتمع N كبيراً فإن حالة السحب بدون إعادة تصبح مقاربة ومعادلة لحالة السحب مع الإعادة وبناءً على ذلك يمكننا حساب حجم العينة في كلتا حالتي السحب من العلاقة :

$$n \geq \frac{Z^2s^2}{N} \tag{19}$$

حيث s^2 هو تباين العينة أو أي عينة تجريبية .

علماً بأنه يمكننا الحصول على هذه النتيجة من قيمة الخطأ المعياري للمتوسط حيث نجد أنه عندما يكون N كبيراً وn صغيراً فإن كسر معامل التصحيح يصبح قريباً من الواحد لأن

$$\frac{N-n}{N} \approx 1$$

وعندها نجد أن الخطأ المعياري لحالة السحب بدون إعادة يصبح مساوياً للخطأ مع الإعادة

$$\sigma_{\bar{x}} = \sqrt{\frac{N-n}{N}} * \frac{\sigma}{\sqrt{n}} \approx \frac{\sigma}{\sqrt{n}} \approx \frac{s}{\sqrt{n}} \tag{20}$$

2- إذا كان المطلوب تقدير النسبة R في المجتمع وكان السحب بدون إعادة فإننا نجعل نصف طول مجال الثقة أصغر أو يساوي نسبة الدقة d فنجد أن

حيث أن $\sigma_r = \sqrt{\frac{N-n}{n}}\sqrt{\frac{R(1-R)}{n}}$

$$Z_{1-\frac{\alpha}{2}} * \sigma_r \leq d$$

$$Z^2 * \frac{N-n}{N} * \frac{R(1-R)}{n} \leq d^2$$
$$n * Nd^2 \geq N * Z^2R(1-R) - nZ^2 * R(1-R)$$
$$n\left(Nd^2 + Z^2R(1-R)\right) \geq N * z^2 * R(1-R)$$

ومنها نجد أن:

$$n \geq \frac{N*Z^2*R(1-R)}{Nd^2+Z^2*R(1-R)} \tag{21}$$

وباستبدال النسبة R بالنسبة r في العينة أو في أي عينة تجريبية نحصل على أن

$$n \geq \frac{N*Z^2*r*q}{Nd^2+Z^2*r*q} \tag{22}$$

حيث أن q=1-r

وهي العلاقة الأساسية لحساب حجم العينة n عند تقدير النسبة R ويمكن تحويرها إلى أشكال أخرى كما رأينا في حالة المتوسط .

ملاحظة: عند حساب حجم العينة يراعى أن يتم حساب النسبة r من أي عينة تجريبية وهي قد تكون 20% أو 30% أو 40% أو غير ذلك وبعد تعويضها في العلاقة السابقة نحصل على حجم العينة الصحيح

ويمكن عندما تكون النسبة متوازنة في المجتمع المدروس (كنسبة الذكور التي تأخذ حوالي r=0.50) فإن حجم العينة n المحسوبة من تلك العلاقة يبلغ أكبر قيمة له . وذلك لأن الجداء(r * q) يأخذ أكبر قيمة له عندما تكون (r = 0.50) و (q= 0.50) ويساوي r *q= (0.50)(0.50)=(0.25)

وبناءً على هذه الخاصة يقوم بعض الباحثين بحساب حجم n من العلاقة الأخيرة بافتراض أن النسبة متوازنة ويضعون فيها (r = 0.50) و (q= 0.50) ويعوضون باقي المؤشرات ,d, N, Z بقيمها فيحصلون على أكبر حجم للعينة n ويستخدمونه في البحث المقابلة لإحجام مختلفة لمجتمعات متوازنة (وفي حالة r= 0.50 ولقد قام بعضهم بتطبيق العلاقة (22) وحساب أحجام العينات والمقابلة لنسب معينة من الدقة d وعند مستوى دلالة محدد $\propto= 0.05$ والمقابل لـ Z= 1.96 فحصلوا على الجدول التالي :

الجدول (1) أحجام العينات لأحجام مختلفة من مجتمعات متوازنة الإحصائية وحسب قيم مختلفة للدقة بالمائة من الثقة أي أن $\propto= \mathbf{0.05}$ وأن قيمة $\mathbf{Z_{1-\frac{\propto}{2}} = 1.96}$ وفي حالة السحب بدون إعادة

حجم المجتمع N	نسبة الدقة =d			
	%5	%3	%2	%1
50	44	48	49	50
100	79	91	96	99
150	108	132	141	148
200	132	168	185	196
250	151	203	226	244
300	168	234	267	291
400	196	291	343	384
500	217	340	414	475
750	254	440	571	696
1000	278	516	706	906
2000	322	696	1091	1655
5000	357	879	1622	3288
10000	370	964	1936	4899
100000	383	1056	2345	8762
1000000	384	1066	2395	9513
10000000	384	1067	2400	9595

المصدر: sounders lewis & thornhl, 2009

وكمثال على كيفية حساب أحجام هذه العينات نفترض أننا نريد أن نسحب عينة بحجم n من مجتمع طبيعي حجمه معلوم N= 1000 لتقدير نسبة الذكور فيه وبدقة محددة d=0.03 وباحتمال ثقة قدره 95% .

(أي بمستوى دلالة $\propto = 0.05$ وهو يقابل قيمة حرجة تساوي $Z_{1-\frac{\propto}{2}} = 1.96$) وللحصول على أكبر حجم للعينة المطلوبة نفترض أن هذا المجتمع متوازن جنسياً ونضع نسبة الذكور r=0.50 فتكون q=0.50 ونعوض في العلاقة الأخيرة (   ) فنحصل على أن:

$$n \geq \frac{1000(1.96)^2(0.50)(0.50)}{1000(0.03)^2 + (1.96)^2(0.50)(0.50)} = \frac{960.4}{1.8604} = 516.23 \approx 517$$

ولكن إذا كنا نريد تقدير نسبة المدخنين في المجتمع المذكور وبنفس الشروط السابقة فإننا نكون في مجتمع غير متوازن من حيث التدخين، لذلك يجب أن نبحث عن تقدير أولي لنسبة المدخنين فيه، ولهذا السحب عينة تجريبية بحجم معقول (وليكن n=100) ونستخلص منها تقدير نسبة المدخنين الأولية في المجتمع ولنفترض إنها كانت تساوي r= 0.35 . وعندها نقوم بحساب حجم العينة النهائي من العلاقة السابقة (   ) فنجد أن:

$$n \geq \frac{1000(1.96)^2(0.35)(0.65)}{1000(0.03)^2 + (1.96)^2(0.35)(0.65)} = \frac{873.964}{1.77396} = 492.48 \approx 493$$

وهو حجم العينة الكافي لتحقيق شروط المسألة وهي أن يكون احتمال الثقة أكبر من 95% وأن تكون نسبة الخطأ المعياري (المرتكب) في تقدير النسبة لا تزيد عن مقدار الدقة المحددة بـ 3%. وكل ذلك بفرض أن جميع عناصر هذه العينة ستستجيب للمعاينة وتعطينا البيانات الصحيحة عن أحوالها .

لأما إذا كانت الاستجابة غير كاملة نقسم حجم العينة على نسبة الاستجابة فنحصل على حجم العينة اللازم للمعاينة . وفي جميع الأحوال يجب على الباحث أن يقوم بإعادة حساب حجم العينة اللازم للمعاينة اعتماداً على بيانات العينة نفسها للتأكد من حجمها ومن صلاحيتها ومن أنها تحقق الشروط المفروضة على البحث.

ملاحظة: لحساب حجم العينة الكلية في المعاينة الطبيعية يمكننا استخدام نفس العلاقات وتطبيقها على المجتمع فنحصل على حجم العينة الكلية n

ثم نقوم بتوزيعها على الطبقات حسب أحجامها النسبية من المجتمع كما يمكن حساب أحجام العينات الطبقية (في كل طبقة كمجتمع مستقل) باستخدام نفس العلاقات على كل طبقة ثم دمج هذه العينات الطبقية فنحصل على العينة الكلية . كما يمكن أخذ التكلفة بعين الاعتبار ...الخ وهناك أمور كثيرة أخرى لا مجال للبحث فيها الآن .