الباب الثاني

العينة و خصائص مجتمعها

الفصل الاول : الأسس العامة

1 - المقدمة :

يعتمد الجغرافيون العينات Samples في الكثير من دراساتهم لاستحالة القيام بمسح شامل لمجتمع الدراسة ، واستكمالا للبيانات المطلوبة لانجاز البحث ، كأن يكون مجتمع الدراسة مياه نهر أو رمال شاطئ أو صخور أو مجتمع بشري . فإذا أراد جغرافي دراسة نسبة الأحماض في مياه المطر ، أو درجة ملوحة مياه نهر معين فانه لا يقوم بتحليل جميع مياه المطر أو النهر ، بل يأخذ عينة لتمثل المجموع . كذلك الحال عند دراسة اتجاهات التبضع (التسوق) لدى سكان مدينة معينة على سبيل المثال . فاخذ العينات ممارسة ميدانية لجمع المعلومات يجب أن يتدرب الطلبة عليها ، وان يتعرفوا على قواعدها ويستوعبوها وان يلتزم بها الباحثين توخيا للدقة والأمانة العلمية .

تنقسم العينات حسب نوعها إلى فئتين رئيسيتين : الأولى احتمالية ، وهي الأكثر شيوعا ، والثانية غير احتمالية . ولكثرة استخدام الفئة الأولى فقد تنوعت طرق جمعها تحقيقا لأعلى تمثيل لمجتمعها وتوافقا مع أهداف البحوث والدراسات .

2 - الاستدلال على خصائص مجتمع الدراسة :

اعتاد البعض على معاملة القيم التي يحصل عليها من العينة وكأنها قيم مجتمعها ، وهذا خطأ فادح . فلكي يستدل على خصائص مجتمع الدراسة تعتمد معادلات عديدة ، ومتنوعة حسب نوع العينة. ما سيرد هنا الشائع منها والضروري لإعطاء النتائج صفة العلمية . و تعتمد جميعها على القياسات الآتية :-

(أ) المعدل والانحراف المعياري :

إن الهدف من اخذ العينة هو معرفة خصائص مجتمعها ، ونادرا ما يكون تمثيل العينة لمجتمعها دقيقا . فعند اخذ مائة قيمة من جداول الأرقام العشوائية وترتيبها بعشرة أعمدة ومثلها اسطر وحساب معدل قيم كل سطر ومعدل قيم كل عمود ، وحساب قيمة الانحراف المعياري للقيم عن معدلاتها وجد أن قيم معدلات الأسطر قد تراوحت بين (6ر41) و (6ر59) وبين (7ر16) و (7ر33) لقيم الانحراف المعياري لها . أما معدلات قيم الأعمدة فقد تراوحت بين (6ر43) و (1ر66) وقيم انحراف معياري بين (7ر19) و(4ر32) ، في وقت كان معدل جميع القيم (المعدل الحقيقي) يساوي (61ر50) وبانحراف معياري قدره (055ر26) .

وبقيام خمس طلبة بدراسة هذا المجتمع مع اختلاف في حجم العينة ، حيث اختاروا (10 ، 20 ، 30 ، 40 ، و 50) على التوالي ، تبين ان المعدلات كانت (1ر66 ، 1ر49 ، 5ر55 ، 3ر47 و 1ر48) على التوالي ، وكانت قيم الانحراف المعياري لها (8ر24 ، 8ر26 ، 6ر26 ، 1ر26 و 6ر25) على التوالي (العمر 1989) . يعني هذا ان أي منهم لم يمثل معدل مجتمع الدراسة ، بل بقيت التقديرات تدور حوله .

يشير المثال أعلاه إلى أن العينة لا تعرض الواقع بل تقترب من معدله ، لهذا من الضروري أن يعرف الباحث درجة القرب منه . بعبارة أدق ، إن كل عينة لم تؤخذ

بموضوعية ولم تحسب احتمالات تمثيلها لمجتمعها بدقة لا يمكن الركون إليها لأنها تمثل نفسها دون مجتمعها . فاخذ العينات ليس القصد منه العينة لذاتها بل المجتمع الذي أخذت منه، فهي وسيلة وليست الهدف . ويأتي هنا دور نظرية العينات لتساعد في تقدير قيم خصائص مجتمع الدراسة وضمن مجال محدد للخطأ ، وحسب تنظير الحد المركزي Central Limit Theorem ، فانه **عند اخذ عينات بحجم كبير من أي مجتمع فان معدلات العينات ستتوزع بصورة طبيعية Normal ، وان معدل معدلات العينات سيقترب من معدل مجتمع الدراسة** (Cohen & Holliday 1983)

تقدم العينات تقديرات لخصائص مجتمعها ، وهذه التقديرات تدور حول المعدل الحقيقي لمجتمع الدراسة . أي **ان معدل العينة هو ليس معدل مجتمعها ، وليس تقديرا له ، بل قيمة تمثل العينة ذاتها ، وتعتمد في تقدير القيمة المحتملة لمعدل المجتمع وفق حدود معينة للثقة** . ولهذا من الضروري حساب معدل قيم العينة و درجة التباين Variance فيها وفق المعادلة :-

التباين = مجموع (تربيع الفرق بين قيم العينة عن معدلها) \ حجم العينة

$$Var = \Sigma (x - u)^2 / n$$

وباشتقاق الجذر ألتربيعي لقيم التباين نحصل على قيمة الانحراف المعياري :-

$$SD = \sqrt{\sum(x-u)^2/(n-1)}$$

إن ارتفاع قيمة الانحراف المعياري يدلل على التباين الكبير بين قيم العينة ، ولهذا أهمية خاصة عند تحديد حجم العينة وعند تحديد درجة الثقة بالتقديرات التي توفرها عن مجتمعها . **وكلما كان التباين كبيرا في خصائص المجتمع كانت معدلات العينة بعيدة عن معدل مجتمعها** ، ولهذا فان قيمة الخطأ المعياري SE . Standard Error of Sample في قيمة الانحراف المعياري للعينة ستكون كبيرة . لهذا فان اخذ Standard Deviation العينات بعدد قليل قد لا يعكس الصورة الصحيحة للتباين في خصائص مجتمع الدراسة . وقد تتشابه قيم العينات عن طريق الصدفة ، ولهذا يفضل اعتماد عدد (حجم) كبير للعينات .

(ب) الخطأ المعياري :

بمعرفة قيمة الانحراف المعياري لقيم العينة يمكن تقدير قيمة الخطأ المعياري في الانحراف المعياري للعينة باعتماد المعادلة الآتية :-

الخطأ المعياري = (الانحراف المعياري \ جذر ع) X جذر (1 - (ع \ ن))

حيث ان : (ن) (N) = حجم مجتمع العينة و (ع) (n) = حجم العينة

$$SE = (SD/\sqrt{n}) \times (\sqrt{1-(n/N)}$$

وفق هذه المعادلة تؤخذ نسبة العينة إلى مجتمعها ، وكلما كبرت هذه النسبة تحسن تمثيلها لمجتمع الدراسة . أما عندما يكون حجم مجتمع العينة مجهولا ، وهي حالة أكثر شيوعا في الدراسات الجغرافية ، حينها تعتمد المعادلة الآتية :- (Dixon & Leach 1978)

$$SE = Sd\sqrt{n}$$ الخطأ المعياري = الانحراف المعياري لقيم العينة \ جذر حجم العينة

أما عندما يكون حجم العينة أكثر من (100) فتعتمد المعادلة أدناه :-
الخطأ المعياري = الانحراف المعياري \ جذر(2 حجم العينة)

$$SE = SD / \sqrt{2n}$$ (Shaw & Wheeler 1985)

ويقدم كريكوري تعديل بيسل عندما يكون حجم العينة اقل من (30) :-

$$SE = SD \times \sqrt{(n / (n-1))}$$ (Gregory 1978)

يتمثل هذا التعديل بأخذ الجذر التربيعي لحاصل قسمة حجم العينة على (حجم العينة - 1) الهدف منه الحصول على أفضل تقدير للخطأ المعياري . بعبارة أخرى ، تستخدم القيمة المشتقة عن معادلة بيسل كبديل لقيمة الخطأ المعياري في المعادلات الأخرى عندما يكون حجم العينة صغيرا!

قدم كريكوري مثالا عمليا يعتمد قيمتي المعدل والانحراف المعياري وحجم العينة فقد

اختار (200) مزرعة فوجد أن معدل حجمها يساوي (90) هكتارا وبانحراف معياري قدره (7) هكتار عن المعدل . ولتقدير معدل حجم المزرعة في منطقة الدراسة اعتمد الصيغة:-

المعدل المتوقع = معدل العينة ± (الانحراف المعياري \ جذر العينة)

$89ر5 = (\sqrt{200} \ \backslash \ 7) - 90 =$ $\quad$ $90ر5 = (\sqrt{200} \ \backslash \ 7) + 90 =$

وباحتمالية قدرها (68%) ((ضمن درجة معيارية واحدة)) فان معدل مساحة المزرعة في منطقة الدراسة يقع بين (5ر89 و 5ر90) هكتار .

المعدل المتوقع = $91 = (\sqrt{200} \ \backslash \ 7) \times 2 + 90$

وباحتمالية قدرها (90%) فان معدل مساحة المزرعة يقع بين (89) و (91) هكتار.
وباحتمالية عالية قدرها (7ر99%) (ثلاث درجات معيارية) يكون المعدل المتوقع لمساحة المزرعة بين (5ر88) و (5ر91) هكتار . (Gregory 1978)

ان الانحراف المعياري للتوزيع النظري لمعدلات العينات يقيس خطأ المعاينة ويسمى بالخطأ المعياري للمعدل . ومن الضروري التذكر دوما **ان معدل المجتمع قيمة محددة تقع ضمن مجال محدد Certain Interval ، والباحث غير متأكد من قيمتها ، ولكنه بحسب احتمالية وجودها ضمن المجال المحدد وبمستوى ثقة إحصائية معلوم**. (Cohen & Holliday 1983)

(ج) مستوى الثقة وحدودها :

إذا أخذت جميع العينات المحتملة من مجتمعها فيتوقع أن تكون معدلات العينات موزعة بالتساوي حول معدل مجتمع الدراسة . بعبارة أخرى ، إن معدل معدلات العينات يساوي معدل مجتمعها . وتتوزع معدلات العينات دائما بصورة متماثلة Normal Distribution ، الذي يمتاز رياضيا ، وبغض النظر عن قيمتي المعدل والانحراف المعياري بالابتعاد بنسب ثابتة عن المعدل مع كل درجة معيارية . أي مهما كان المجتمع متباينا في خصائصه ، وبالتالي تباينت معدلات العينات المأخوذة منه فانه يتوقع أن يقع معدله وباحتمالية قدرها (90%) ضمن مدى لا يزيد عن (96ر1) من قيمة الخطأ المعياري.

وباحتمالية قدرها (99%) يقع معدل مجتمع الدراسة بين (+ و - 58ر2) من قيمة الخطأ المعياري . وتوفر الجداول الإحصائية نسبا منوية أخرى لقياس بعد معدل العينات عن معدل مجتمعها، إلا أن هاتين النسبتين (95%) و (99%) هما الأكثر استخداما. وتسمى مستوى الثقة Confidence Level و يعبر عنها بإشارة النسبة المئوية (%) بان تكون التقديرات صحيحة ، أو باحتمالية (01ر0) أو (05ر0) أن تكون خاطئة .

بمستوى ثقة إحصائية قدره (26ر68%) أو باحتمالية قدرها (6827ر0) يقع معدل مجتمع الدراسة بين قيمة معدل معدلات العينات و (+ و -) درجة واحدة من الخطأ المعياري. وبثقة إحصائية قدرها (95%) ، أو باحتمالية (95ر0) يقع معدل مجتمع الدراسة بين معدل معدلات العينات و (+ و - 69ر1) من قيمة الخطأ المعياري . وبثقة إحصائية قدرها (99%) أو باحتمالية قدرها (99ر0) يقع معدل مجتمع الدراسة بين قيمة معدل معدلات العينات و (+ و - 58ر2) من قيمة الخطأ المعياري . وبثقة قدرها (73ر99%) أو باحتمالية قدرها (9973ر0) فان معدل مجتمع الدراسة يقع بين معدل معدلات العينات و(+ و -) ثلاث درجات من الخطأ المعياري .

وتحسب حدود الثقة لاعتمادها في تقدير معدل مجتمع الدراسة وكما يأتي :-

حدود الثقة = ± (درجة معيارية * الخطأ المعياري) $= C \pm SE * Z$

وبافتراض قيام جغرافي بدراسة مائة مزرعة ، وجد ان معدل مساحة المزرعة الواحدة (53) هكتارا، وبانحراف معياري عن المعدل بقيمة (26) هكتار ، عندها تحسب حدود الثقة في تقدير معدل مساحة المزرعة في منطقة الدراسة :-

حدود الثقة = ± 96ر1 $\times$ (26 \ $\sqrt{100}$) = 1ر5

عندها يمكن القول بأنه بثقة إحصائية قدرها (95%) أن معدل مساحة المزرعة في منطقة الدراسة يقع بين (53 + 1ر5 = 1ر58) هكتار و (53 - 1ر5 = 9ر47) هكتار . أما إذا أراد الباحث أن يكون مستوى الثقة الإحصائية (99%) حينها يستبدل (96ر1) ب (58ر2) في المعادلة أعلاه ليكون المعدل المتوقع بين (3ر46) و (7ر59) هكتار .

تعتمد قيمتي الخطأ المعياري و حدود الثقة على درجة التباين في خصائص المجتمع قيد الدرس . فقد يوفر عدد قليل من العينات تقديرات جيدة لخصائص المجتمع عندما يكون هذا متجانسا نسبيا ، والعكس صحيح أيضا . أي أن تكون التقديرات غير واقعية عندما تكون التباينات كبيرة في خصائص المجتمع وحجم العينة صغيرا .

3 - حجم العينة :

لحجم العينة أهمية كبيرة في تحديد الثقة بالنتائج ، لذا من الضروري أن يسلط الضوء عليه بشيء من التفصيل وحسب التوزيعات المعروفة للقيم .

(أ) التوزيع الطبيعي للقيم :

كلما كبر حجم العينة ازدادت دقة تمثيلها لمجتمعها واقترب توزيع القيم فيها من التوزيع الطبيعي (المتماثل الجانبين) وأصبحت عملية الاستدلال أكثر دقة . وللتوضيح نورد مثالا ، إذا أريد معرفة نسبة طلبة قسم الجغرافيا إلى مجموع طلبة الكلية فان عينة من عشرة طلبة قد لا تفي بالغرض، ولكن عينة من مائة طالب تفي بالغرض حتما . بعبارة أخرى، إن حجم العينة أساسي لإعطاء صورة عن مجتمع الدراسة وليس النسبة المئوية للعينة قياسا

بحجم مجتمعها . **فكلما ازداد حجم العينة ازدادت الثقة بتقديرات خصائص المجتمع وصغرت معه حدود الثقة** .

وقد لا يعرف الباحث الكثير عن مجتمع الدراسة ، فهو بحاجة إلى بعض المؤشرات عن تباين خصائص المجتمع قبل أن يحدد حجم العينة . يتطلب منه هذا القيام بمسح تجريبي لاستخراج قيمة المعدل والانحراف المعياري للعينة ، وبتطبيق المعادلة المذكورة في أدناه يستطيع تحديد حجم العينة المناسبة لدراسته :- $n = (Z * SE / C)^2$

حجم العينة = ((الدرجة المعيارية × الخطأ المعياري) \ حدود الثقة)^2

أما **إذا كان اهتمام الباحث منصبا على نسبة الخصائص وليس قيمتها ، لذا لا يمكن اشتقاق قيمة الانحراف المعياري ، بل يستعاض عنه بقيمة التباين في النسب** ، وتحسب :-

التباين = جذر (ن × (100 - ن)) $Var = \sqrt{(P(100-P))}$

حيث تمثل (ن) (p) النسبة المئوية للخاصية قيد الدرس ، ولهذا تكون معادلة حجم العينة :

حجم العينة = ((درجة معيارية × التباين) \ حدود الثقة)^2 $n = (Z * Var / C)^2$

فإذا أراد باحث دراسة ملكية السيارات في مدينة ما ، واختار (2%) حدا للثقة وثقة إحصائية قدرها (95%) ، ويتوقع أن يمتلك نصف السكان وسائط نقل خاصة ، حينها:-

التباين = جذر (50 × (100 - 50)) = 50

العينة = ((1ر96 × 50) \ 2)^2 = 1ر24

أي انه بحاجة إلى عينة بحجم 24 لضمان الدقة المرجوة في تمثيل خصائص المجتمع. وقد يبحث الجغرافي في مواضيع ذات مجتمعات صغيرة الحجم وبهذا تشكل النسبة المئوية حجما كبيرا فتكون النتائج مضللة ، في بعض الأحيان .

(ب) توزيع (ت) للقيم :

من الضروري اخذ الحذر عندما يكون حجم العينة صغيرا ، اقل من (30) وذلك لأنها تتطلب إجراءات خاصة عند التحليل . فعندما يكون حجم العينة أكثر من (30) يتجه توزيع قيمها نحو التوزيع الطبيعي وبغض النظر عن التوزيع الحقيقي لقيم مجتمع الدراسة. وبالنسبة للعينات الصغيرة الحجم فان توزيع قيمها يتأثر بطبيعة توزيع قيم المجتمع المأخوذة منه. **وعندما يكون توزيع قيم المجتمع معروفا أو متوقعا أن لا يكون طبيعي حينها يجب اعتماد حجم كبير للعينة** . أما إذا كان توزيع قيم المجتمع طبيعيا عندها يمكن اخذ عينات بحجم صغير ويعتمد توزيع (ت) (T) في التحليل و المقارنة .

يتشابه توزيع قيم (ت) مع شكل الجرس بزيادة عددها حتى يتطابق معه عندما يتعدى العدد (30) . بعبارة أخرى ، لا يختلف شكل توزيع (ت) للقيم كثيرا عن التوزيع الطبيعي إلا في الأعداد القليلة (خاصة عند ذيل الجرس) ، وكلاهما متماثل النصفين لذا يعتمد كبديل له في القيم القليلة العدد .

وكما للتوزيع الطبيعي جدول للقيم المعيارية (الحرجة) كذلك لتوزيع (ت) منظمة على شكل اسطر اعتمادا على درجة الحرية التي تقاس ب(حجم العينة - 1) . أما الأعمدة فتمثل درجة الاحتمالية Probability ، وتتناقص القيم الحرجة بتزايد درجة الحرية (حجم العينة). **ودرجة الحرية تفضل على حجم العينة في الأحجام الصغيرة للعينة لأنها تقلل من**

الانحياز في تقدير خصائص مجتمع الدراسة . لذلك يفضل القسمة على (ع - 1) (1 - n) في المعادلات التي يكون حجم العينة مقاما لها.

ليس هناك اختلاف في حساب قيم المعدل والانحراف المعياري بين العينات الكبيرة الحجم أو الصغيرة ، يبدأ الاختلاف مع تحديد حدود الثقة حيث تستبدل قيمة (Z) لتحل مكانها قيمة (T) . وللتوضيح نعتمد مثالا أورده شو و ويلر عن دراسة خصائص تربة منطقة هونك كونك حيث أخذا (20) عينة فوجدا أن معدل حموضتها (4ر75) وبانحراف معياري قدره (0ر12) ، وبحساب قيمة الخطأ المعياري اتبعت الخطوات الآتية :-

الخطأ المعياري = 0ر12 \ جذر (20) = 0ر027

وبالعودة إلى جداول (ت) وبدرجة حرية (20 - 1 = 19) وباحتمالية إحصائية قدرها (95%) وجد أن القيمة الجدولية تساوي (2ر09) ، وتقابلها في التوزيع الطبيعي القيمة (1ر96):-

حدود الثقة = قيمة (ت) × الخطأ المعياري = 2ر09 × 0ر027 = 0ر05643

أي إن معدل حموضة تربة هونك كونك يقع بين :- (Shaw & Wheeler 1985)

معدل العينة + حدود الثقة = 4ر75 + 0ر05643 = 4ر80643

معدل العينة - حدود الثقة = 4ر75 - 0ر05643 = 4ر69357

أي إن درجة الحموضة تقع بين (4ر8) و (4ر6) PH باحتمالية إحصائية قدرها (95%) تكون النتيجة صحيحة . إن ارتفاع القيم الجدولية في توزيعات (ت) قياسا بنظائرها في توزيع (Z) تعكس الدرجة العالية لعدم ضمان تمثيل العينات الصغيرة لخصائص مجتمعها.

(ج) التوزيع الثنائي :

معظم الأمثلة الإحصائية عن بيانات تقاس بالفاصلة Interval أو بمقياس النسبة Ratio وهي ذات توزيع طبيعي للقيم ، ولكن هناك حالات عديدة تكون البيانات التي يعتمدها الجغرافي بقياسات اسمية Nominal ، وهي تتبع التوزيع الثنائي Binomial Distribution للقيم، وتكون بياناتها على شكل تكرارات أو نسب Proportions. يتشابه التوزيع الثنائي مع التوزيع الطبيعي للقيم عندما يكون عددها كبيرا .

بافتراض وجود مجتمع يضم أفرادا يمكن تصنيفهم إلى مجموعتين على أساس خاصية معينة (استخدام تقنية معينة ، مهنة ما ، صفة، .... الخ) وتعد حالة وجود هذه الصفة (في الفئة) نجاحا بلغة الإحصاء ، إذا يرمز لها ب(ن) (P) ، وحالة عدم الوجود بالفشل (1 - ن) (1 - P) أو الرمز (ف) (Q) ويحسب المعدل كنسبة مئوية لتكرار حالة النجاح إلى مجموع العينة ((ن × 100) \ ع) (P*100/n) . ويحسب التباين بالصيغة (ن × ف) \ع Var = PQ/n وبأخذ الجذر التربيعي للتباين نحصل على قيمة الانحراف المعياري. وطالما هناك تشابه بين التوزيعين الطبيعي والثنائي عندما يكون حجم العينة كبيرا فان تكرار حالة النجاح في مجتمع الدراسة يقع وباحتمالية قدرها (95%) بين درجتين معياريتين على جانبي المعدل :

نسبة النجاح = ن ± 2 جذر ((ن × ف) \ ع) $pp = p \pm 2 * \sqrt{(pq/n)}$

وللتوضيح قدم كونوي المثال الآتي : في دراسة شملت (144) مزرعة وجد أن (33) منها تربي الأغنام و (41) تربي الأبقار ، فما هي تقديرات النسب في منطقة الدراسة ؟

نسبة المزارع التي تربي الأغنام = ((33 × 100) \ 144) = 9ر22%

قيمة الانحراف المعياري = جذر ((9ر22 × 1ر77) \ 144) = 5ر3%

إذا باحتمالية قدرها (95%) تقع نسبة المزارع التي تربي الأغنام في منطقة الدراسة بين :

9ر22 + (2 × 5ر3) = 9ر29%   9ر22 - (2 × 5ر3) = 9ر15%

(Conway 1967)

4 - العوامل المحددة لحجم العينة :

لا يتحدد حجم العينة اعتباطيا ، بل يجب اعتماد الأسس الآتية :-

(أ) درجة التباين في خصائص مجتمع الدراسة :

يلعب التباين في خصائص مجتمع الدراسة دورا مهما في تحديد درجة دقة نتائج العينة ، فكلما كان التباين كبيرا تطلب الأمر زيادة حجم العينة ليكون تمثيلها للتباين في المجتمع صحيحا . لتوضيح هذه النقطة يورد معظم الإحصائيين المثال الآتي : ورق لعبة الميسر (القمار) يضم لونين وأربعة أصناف و (13) قيمة ، ولتمثيل الألوان يؤخذ عدد قليل من الأوراق ، ويزاد العدد لتمثيل الأصناف الأربعة ، ولا تمثل القيم الثلاث عشر إلا بعدد كبير من أوراق اللعب (العينات) .

في الواقع هناك حاجة لأخذ عينات بحجم كبير إذا كان التباين في خصائص المجتمع كبيرا ، أما عندما يكون مشروع البحث غير واسع حينها لابد من البحث عن وسائل تقلل من حجم العينة والتركيز على الأهداف الرئيسية والنقاط الجوهرية المطلوب تحليلها . وعندما يراد الوصول إلى عمومية عن سلوكية التبضع (التسوق) من الضروري زيادة حجم العينة لتغطية التباين الكبير في هذا المجال . أما إذا أريد دراسة سلوكية التبضع لدى العوائل ذات الأطفال (أوائل دورة حياة العائلة) حينها يعتمد حجم صغير للعينة لصغر حجم مجتمع الدراسة وتجانسه النسبي . وبمضاعفة حجم العينة عشر أضعاف تنقص قيمة الخطأ المعياري إلى الثلث (Gregory 1978) .

(ب) طريقة التحليل المعتمدة :

عند إقرار حجم العينة ، من الضروري تحديد الحجم الأصغر المقبول للعينة في المجاميع الثانوية ضمن مجتمع الدراسة ، إذ أن بعض الاختبارات الإحصائية تتطلب عددا معينا كحد أدنى لكل فئة أو صنف لتكون النتائج ذات معنى (مربع كاي على سبيل المثال) .

(ج) حجم المعلومات المطلوبة :

كلما كانت المعلومات المطلوبة من العينة (الواحدة) كثيرة وتفصيلية (سواء أكان هذا عن طريق الاستبانة أو التحليلات المختبرية) كان حجم العينة صغيرا ، ما لم يكن المشروع البحثي كبيرا وتتوفر له المصادر البشرية والمادية اللازمة . **إن الدقة في المعلومات المطلوبة من العينة أهم بكثير من حجم العينة** ، فحجم العينة لا يتحدد بحجم مجتمع الدراسة فقط ، بل وبالدقة المتوخاة والتفاصيل المطلوبة (Theakston & Harrison 1978) .

(د) المصادر المالية والبشرية المتوفرة :

تتطلب الدراسة الميدانية توفر مصادر مالية وبشرية لتغطية تكاليفها التي تكون في الغالب باهظة ، حتى وان كان مشروع البحث صغيرا . لذا تكون المصادر المالية والبشرية المتاحة للباحث من ابرز النقاط وأكثرها فاعلية عند مناقشة خطة البحث لتأثيراتها على تحديد حجم منطقة الدراسة ، مجتمع الدراسة وبالتالي حجم العينة . إن مضاعفة حجم العينة

يتطلب زيادة في كمية المصادر المالية والجهد البشري بنسبة (80%) (Dixon & Leach 1978) .

(هـ) حدود الثقة في تقديرات خصائص مجتمع الدراسة :

لزيادة الدقة في النتائج يعمد البعض إلى تقليص حدود الثقة (المدى الذي يفترض أن يقع ضمنه المعدل المتوقع للمجتمع) . إن إنقاص حدود الثقة من (6%) إلى (4%) يتطلب زيادة حجم العينة بنسبة (225%) ، وكلما كان المدى كبيرا كان حجم العينة صغيرا، والعكس صحيح . من الناحية العملية فان حجم العينة التي يعتمدها الطلبة في مشاريع بحوثهم تعتمد الوقت المتاح والإمكانات المادية المتوفرة فقط . مع هذا ، من الضروري تحديد حدود الثقة لتأشير حجم العينة . **فإذا كان حجم العينة لا يساعد في الخروج بعمومية واقعية عن خصائص مجتمع الدراسة حينها يفضل عدم القيام بجمع العينات أساسا** . فحتى المشاريع البحثية البسيطة التي هدفها الخروج بأفكار عامة يشترط فيها تحديد حدود الثقة لتقديراتها، ويفضل تحديد حجم العينة ومدى الثقة قبل القيام بالدراسة الميدانية وليس بعدها .

(و) حالات الإخفاق و عدم الاستجابة :

العامل الآخر الذي يحدد حجم العينة هو حالات الإخفاق في الحصول على المعلومات وعدم الاستجابة أو المعلومات غير الوافية (Kalton 1983) . ويمكن تحديد نسبة هذه الحالات مسبقا من خلال المعرفة التفصيلية بمنطقة الدراسة و مجتمعها ليتسنى زيادة حجم العينة تلافيا للنقص المتوقع .