**جامعة وهران**

**كلية العلوم الإقتصادية، علوم التسيير، والعلوم التجارية**

**مذكرة تخرج لنيل شهادة الماجستير في العلوم الإقتصادية**

**تخصص: التقنيات الكمية المطبقة**

الموضـــــــوع

# إشكاليـــة تحديـد حجـــم العينـــة فـــي الدراســـات الإقتصاديـــة والإجتماعيـــة


**من إعداد الطالب:**

موسى نبيل سمير

**تحــت إشـــــراف:**

د.عامر عامر أحمد


أعضاء لجنة المناقشة:

| | | |
|---|---|---|
| **ـ فقيه عبد الحميد** | **أستاذ محاضر-أ- جامعة وهران** | **رئيسا** |
| **ـ عامر عامر أحمد** | **أستاذ محاضر-أ- جامعة مستغانم** | **مقررا** |
| **ـ بوكعبار بوجلال** | **أستاذ محاضر-أ- جامعة وهران** | **مناقشا** |
| **ـ كفيف محمد بن عودة** | **أستاذ محاضر-أ- جامعة وهران** | **مناقشا** |

**السنة الجامعية: 2010- 2011**

## المقدمة:

أصبح اللجوء إلى المسوحات الإحصائية بالمعاينة العشوائية (Sondages) عملية جد ضرورية في معظم الدراسات التي تجري في مختلف العلوم و الميادين، كعلم الإجتماع، والديمغرافيا، والعلوم الإقتصادية، ودراسات السوق، ومراقبة الانتاج و الجودة ...الخ. تستعمل المسوحات عند استحالة القيام بالحصر الشامل نظرا لعدة أسباب سوف نأتي لذكرها بالتفصيل لاحقا.

يعتمد المسح الإحصائي على دراسة جزء فقط من المجتمع و الذي يسمى عينة، وهو أسلوب قليل التكاليف كما يتم إنجازه في وقت أقصر ما يسمح لنا بالحصول على معلومات في مدة وجيزة قصد اتخاذ قرارات حاسمة لا تحتمل التأجيل.

لذا سوف نحاول في هذا البحث معالجة موضوع تحديد حجم العينة بصفة عامة و حجم العينة الأمثل بصفة خاصة و عرض الصيغ الرياضية التي تمكننا من تحديد حجم العين الواجب سحبه حتى نتمكن من انجاز دراسة إحصائية في وقت قصير، و بتكاليف أقل مع مراعاة دقة النتائج و تقليص الأخطاء.

يعتبر موضوع تحديد حجم العينة مهما جدا، لأنه إذا كانت العينة صغيرة أكثر من اللازم، فإننا نفشل في الوصول إلى أهداف الدراسة. أما إذا كانت العينة أكبر مما ينبغي فإننا نبدد الوقت و الموارد، لأن التكلفة تكون أكبر كلما كانت العينة كبيرة. لذا فعند تصميم الدراسة بالمعاينة، يجد الباحث نفسه في مرحلة تستوجب عليه الإجابة على السؤال الجوهري التالي :" ما هو حجم العينة الذي ستشمله الدراسة حتى نتمكن من الحصول على نتائج دقيقة؟" أو" ما هو تأثير حجم العينة على دقة النتائج، و دوره في تقليص الأخطاء ؟".

يدفعنا الإشكال المطروح إلى محاولة الإجابة على ما يلي :

- ما هي مزايا وعيوب كل من الحصر الشامل و المعاينة؟
- ما هي مختلف أساليب السحب العشوائي ؟
- ما هي المعلومات التي يمكن للعينة أن توفرها لنا حول المجتمع؟
- كيف يمكن تعميم النتائج المحصل عليها في العينة على المجتمع؟

من أجل الإجابة على هذه الأسئلة، سوف نعرض في الفصل الأول مزايا وعيوب كل من الحصر الشامل و المعاينة بالإضافة إلى عرض أهم خطوات إنجاز بحث ميداني. سوف نعرض في الفصل الثاني كيفية تحديد حجم العينة في أسوب المعاينة العشوائية البسيطة. يتحدث الفصل الثالث عن عملية تحديد حجم العينة في أسلوب المعاينة العشوائية الطبقية. سوف نعرض في الفصل الرابع أسلوب المعاينة العشوائية العنقودية. سوف نعرض في الفصل الأخير نتائج البحث الميداني الذي خص المؤسسات الصغيرة و المتوسطة في ولاية وهران ونبين فيه كيفية تحديد حجم العينة الأمثل الخاص بمثل هذا النوع من البحوث و الدراسات.

# الفصل الأول:

# التعداد الشامل و المسح بالمعاينة

قبل الخوض في الحديث عن مختلف المعايير الواجب مراعاتها من أجل تحديد حجم العينة الضروري لإجراء دراسة، لا بأس أن نعرض أولا هذا الفصل التمهيدي و الذي يحتوي على شرح لبعض المفاهيم و التعاريف الأساسية المستعملة في الإحصاء الاستقرائي مثل المجتمع الإحصائي، والعينة، و غيرها من المصطلحات. كما نعرض في هذا الفصل تعريفا لكل من الحصر الشامل و المسح بالمعاينة، مع ذكر مزايا و عيوب استعمال كل واحد منهما و كذا مجالات تطبيقهما. يعرض هذا الفصل كذلك بعض أساليب المعاينة العشوائية و غير العشوائية، مثل أسلوب المعاينة العشوائية البسيطة، و الطبقية، و العنقودية، و بالحصص و غيرها من الأساليب. كما نتحدث في الفقرة الأخيرة من هذا الفصل عن أهم الخطوات الواجب اتباعها عند إنجاز بحث ميداني.

## I-1 مصطلحات وتعاريف:

### I-1-1- المجتمع الإحصائي و الوحدة الإحصائية:

نقصد بكلمة مجتمع إحصائي مجموعة من الوحدات تكون موضع الدراسة والتي نريد معرفة بعض خصائصها وتحديد حقائق عنها. والخاصية (caractéristique)هي ظاهرة عشوائية قيد الدراسة، يمكن أن تكون هذه الخاصية نوعية مثل الذكر، والأنثى عند دراسة ظاهرة الجنس، أو أسود، وأصفر في حالة دراسة لون الشعر. كما يمكن أن تكون الخاصية كمية، فمثلا عند دراسة التحصيل العلمي في مادة الرياضيات يمكن أن تكون الخاصية عن العلامة المحصل عليها من طرف التلاميذ، وعند دراسة إنتاج مصنع لصناعة المصابيح، يمكن أن تكون الخاصية هي المصابيح التالفة أو الضائعة.

تسمى الوحدات قيد الدراسة كذلك بالعناصر، والعنصر هو جزء صغير في المجتمع لا يمكن تجزئته، وهو لوحده عبارة عن وحدة دراسة قائمة بذاتها، كما نجد في المراجع الإحصائية مرادفات أخرى للوحدة الإحصائية، فمصطلح المفردة(Individu) مثلا، وكذلك وحدة الاختيار. إن الوحدات المسحوبة من المجتمع الإحصائي و التي نقوم بدراستها تسمى وحدات الدراسة.

يمكن أن يكون المجتمع مجموع الطلبة في كلية العلوم الاقتصادية، وعلوم التسيير، والعلوم التجارية أو مجموع المؤسسات الصغيرة والمتوسطة في ولاية وهران، أو مجموع سكان دائرة السانية، أو مجموع أساتذة التعليم العالي بالجزائر.

إن المجتمعات المذكورة يمكن عدها وإحصاء مجموعها في هذه الحالة نقول عن المجتمع أنه معلوم أو منته، أما إذا لم يكن بالإمكان عد وحصر وحدات مجتمع معين كالبحر، أوعدد النجوم، أوعدد الأشجار في غابة ما، فإن المجتمع هنا يسمى غير معلوم أو غير منته. ولكل مجتمع حدوده الزمنية والمكانية أو الجغرافية، كأن نقول مثلا بأننا نود دراسة الطلبة الجامعيين وتكون الحدود المكانية هي ولاية وهران والحدود الزمنية تحدد الطلبة المسجلين في الموسم الجامعي **2008/2007**.

## I-1-2- العينة العشوائية:

يشكل مجموع وحدات الدراسة المسحوبة من مجتمع ما يسمى بالعينة. ونقول عن هذه الأخيرة أنها عشوائية إذا قمنا بسحب وحداتها، بحيث يكون لكل وحدة في المجتمع الإحصائي نفس الاحتمال في السحب ($P=1/N$). يجب على العينة العشوائية أن تمثل المجتمع أحسن تمثيل: أي يجب عليها أن تكون صورة مصغرة للمجتمع قيد الدراسة، حتى نتمكن من الحصول على نتائج وتقديرات تقترب من النتائج الحقيقية.

نقول عن عينة عشوائية أنها تمثل المجتمع إذا تم سحب وحداتها باستعمال الأساليب الإحصائية الملائمة وفقا لخصوصيات المجتمع وحسب المتغيرات المدروسة، وإذا تم إعداد الدراسة والتحضير لها بطريقة جيدة. لا تعني التمثيلية بالضرورة أن تتطابق التقديرات المحصل عليها من العينة مع المعالم الحقيقي في المجتمع، بل نقصد بالتمثيلية أن يحقق الباحث مستوى معين من الدقة يكون قد قبل بها مسبقا، كأن يقبل مثلا باحتمال تحقيق خطأ معين، ونسمي مستوى الخطأ المقبول هذا بخطأ الاختيار.

سوف نرى في كل الفصول المقبلة كيف أن لحجم العينة والمجتمع تأثير كبيرا على درجة تمثيلية العينة للمجتمع وكذا تأثيرها على دقة تقديرات المعالم المحصل عليها. يقصد بالمعالم (Paramètres) الإحصائية، كل من المتوسط أو التباين أو النسبة...إلخ.

## I-1-3- حجم المجتمع وحجم العينة:

نسمي عدد الوحدات الإحصائية (وحدات الاختيار) المشكلة للمجتمع الإحصائي بحجم المجتمع، والذي نرمز له بـــ N. أما عدد وحدات الدراسة التي يتم سحبها في العينة فتسمى بحجم العينة والذي يرمز له بــ n. وحاصل قسمة عدد وحدات العينة n على كل وحدات المجتمع N يسمى بمعدل الاختيار، والذي يرمز له بـــ $f$.

## I-1-4- إطار المعاينة:

عند سحب عينة من مجتمع إحصائي معين لا بد أن يوجد إطار ما لهذا المجتمع بكامله حتى يمكن سحب العينة على أساسه. والإطار إما أن يكون قائمة بأسماء وحـدات المجتمـع موضوع الدراسة (أسر أو مؤسسات صناعية أو أراضي زراعية...). مرتبة حسب الأحرف الأبجدية أو حسب الأرقام المرفقة بها، بحيث لا تتكرر أسماء الوحدات مع عدم إهمال بعض الأسماء. ويعني ذلك أن القائمة يجب أن تكون شاملة وليس فيها أي ازدواج حتى لا يحـدث التحيزعند سحب العينة. إن إهمال بعض الأسماء لا يعطيها فرصة الـدخول فـي العينـة، وتكرار بعض الأسماء يعطيها فرصة أكثر من الأسماء الأخرى التي لم تكـرر، وبـذلك لا تعطي وحدات المجتمع الإحصائي نفس الفرص للدخول في العينة، وهـو أسـاس السـحب العشوائي السليم للعينات.

وقد يكون الإطار خريطة أو صورة شمسية مكبرة للأراضي الزراعية. وتنقسم أطـر المعاينة إلى قسمين؛ القسم الأول يتعلق بأطر معاينة المجتمعات الصغيرة مثل أطـر تتعلـق بمؤسسات، وقوائم تضم أسماء طلبة الجامعات والمعاهد وأسماء المواطنين المسـموح لهـم بالانتخاب، وأطر المنظمات والجمعيات التي تنشط في منطقة ما. أما القسم الثـاني للأطـر، فهو متعلق بالمجتمعات الكبيرة، كسجلات الأحوال المدنية التي تضم جميع أفـراد المجتمـع والسكنات والتي يتحصل عليها انطلاقا من التعداد الشامل للسكان.

## I-2 ماهية التعداد الشامل وأهميته:

يهدف التحليل الإحصائي إلى إعطاء معلومات عن ظواهر غير معروفة بشكل واضـح ودقيق، وذلك بهدف وضع مجموعة من الاستنتاجات، أو اتخاذ قرارات أكثر دقة وعقلانيـة حول الظاهرة المدروسة. ويقصد بكلمة إحصاء في صيغته المفردة، ذلك العلم الذي يهـدف إلى دراسة كيفية جمع المعلومات من المجتمعات الإحصائية المختلفة وكيفيـة تحويـل هـذه المعلومات إلى بيانات في جداول إحصائية، واستخدام مختلف الأساليب لتحليل هذه البيانات قصد استنتاج المقاييس المختلفة مثل المتوسط، والانحراف المعياري، والنسبة... أو حسـاب مختلف المعلومات كمعامل الارتباط ومعامل الانحدار في الاقتصاد القياسي. كما يهدف علـم الإحصاء إلى تقديم النتائج المحصل عليها ضمن تقرير عن موضوع الدراسة والبحث.

يتم التحليل الإحصائي بدراسة مجموعة من الوحدات الإحصائية، وفي كثير من الأحيان نحصل على هذه المجموعة من الوحدات من خلال دراسة جزء فقط من المجتمع قيد الدراسة، والذي سبق تسميته بالعينة، وفي حالات أخرى قليلة نقوم بدراسة كل وحدات المجتمع الإحصائي، تسمى الحالة الأولى بالمسح بالمعاينة، أما الثانية فتسمى بالتعداد أو الحصر الشامل.

توجد أمثلة ومجالات كثيرة تستعمل فيها هاتان الطريقتان في جمع البيانات، وكثيرا ما نندهش للدقة الكبيرة التي تحصل عليها بعض المؤسسات المتخصصة في المسح بالمعاينة، خاصة في الدول المتقدمة، إذ تقترب كثيرا النتائج المحصل عليها من العينة مقارنة بنتائج الحصر الشامل. ومثال ذلك تلك المسوحات بالمعاينة التي تجرى قبيل الانتخابات والتي كثيرا ما تقترب تقديراتها ونتائجها من نتائج فرز الأصوات بعد الانتهاء من عملية الانتخاب.

يتم اختيار أحد الطريقتين الشائعتين في جمع البيانات حسب أهمية وخطورة القرارات الواجب اتخاذها انطلاقا من نتائج الدراسة، وحسب عدة اعتبارات أخرى كالإمكانيات المادية والبشرية المتوفرة وكذلك عامل الزمن.

ولكل من الحصر والشامل والمعاينة إيجابياته وسلبياته وهذا ما سوف نعرضه بالتفصيل فيما يلي.

### I-2-1- تعريف الحصر( التعداد) الشامل:

هو عملية جمع البيانات الإحصائية من خلال دراسة كل وحدات المجتمع الإحصائي قيد الدراسة. إذ تتطلب بعض القرارات الاقتصادية أو السياسية معرفة رأي كل المواطنين في المجتمع، ولا يمكن اتخاذ هذه القرارات إلا بعد استجواب كل فرد على حدا. فمثلا في حالة إجراء انتخابات رئاسية في أي دولة ما يتم الآخذ بعين الاعتبار برأي كل منتخب مدلي بصوته حتى يتعين تحديد المرشح الفائز بغالبية الأصوات.

إن الحاجة إلى جمع معلومات شاملة وكاملة حول ظاهرة ما، هي حاجة طبيعية فرضت نفسها منذ العهد القديم. حيث كان الحكام يسعون إلى الحصول على معلومات كافية حول عدد مواطنيهم وقدراتهم العسكرية، والإنتاج الزراعي في البلد وثرواته المختلفة وكذا محاولة منهم لجمع الضرائب على السكان. يعتقد أن أول تعداد للسكان تم إجراؤه في الحضارة السامرية،

وكان منذ 2000 إلى 5000 سنة قبل الميلاد، حيث وجدت فيه أسماء للرجال، كما وجدت فيه سلع مسجلة على ألواح من الطين. كما وجدت تعدادات أخرى للسكان والسلع خلال فترة MISOPATAMIE وكان ذلك 3000 سنة ق.م. إلى جانب التطور الكبير الذي عرفه هذا المجال خلال الحضارة الفرعونية في مصر، حيث كان حكامها أول من قاموا بإحصاءات منتظمة ومتكررة وكان ذلك على الأقل منذ 2900 سنة ق.م. في الصين قام الإمبراطور "YAO" بإصدار قائمة كاملة للأفراد والإنتاج الزراعي سنة 2238 سنة ق.م.

يجرى في الجزائر تعداد وحصر شامل للسكان مرة كل عشر سنوات تقريبا، وذلك بغرض معرفة عدد المواطنين الإجمالي ومعرفة بعض خصائص المجتمع، مثل نسبة الفقر ومستوى الأمية وغيرها من أجل اتخاذ إجراءات كفيلة بتحسين الأوضاع والتخفيف من حدة المشاكل التي يعاني منها المواطنون، وحتى تتمكن الدولة من وضع سياسات اقتصادية واجتماعية فعالة وفقا للمتطلبات والأولويات الملحة.

أما إجراء أول تعداد شامل للسكان في الجزائر المستقلة فكان سنة 1966م[1]، ثم تلاه الإحصاء الثاني للسكان سنة 1977 ثم بعد ذلك في سنة 1987، ومؤخرا آخر تعداد للسكان أجري سنة 2001. ويقوم بإنجاز هذا النوع من المهام جهاز إحصائي كبير تابع للدولة يسمى الديوان الوطني للإحصائيات والذي أنشئ غداة الاستقلال في سنة 1964 تحت اسم المحافظة الوطنية لإحصاء السكان. بالإضافة إلى مهمته في إنجاز التعداد الشامل للسكان فإن هذا الديوان يسهر على إنجاز مسوحات إحصائية سنوية متعددة تخص جميع الميادين والتي تمكن الأعوان والمتعاملين الاقتصاديين والباحثين من استعمال مختلف البيانات المحصل عليها قصد استغلالها في الدراسات واتخاذ القرارات المناسب.

قام هذا الجهاز بإجراء مسح ديموغرافي سنة 1972–1979، ومسح خرائطي في نفس الفترة ومسح استهلاك الأسر في سنة 1975–1979 و1980 بالإضافة إلى المسوحات السنوية التي ينجزها حول أسعار المواد الاستهلاكية والمؤشرات الاقتصادية حول مختلف النشاطات الإنتاجية للمؤسسات.

تظهر جليا الأهمية الكبيرة للحصر الشامل (التعداد الشامل) في جميع الميادين الاقتصادية والاجتماعية والثقافية، فبعدما كانت وظيفة الحصر الشامل قديما تقتصر فقط على خدمة أهداف الدولة، والمتمثلة في معرفة عدد الرجال والجيوش في محاولة لمقارنتهم مع جيوش العدو،

[1] موقع الانترنيت للديوان الوطني للإحصاءات www.ons.dz .

أصبح الحصر الشامل (التعداد) يلعب دورا جوهريا في إعطائنا المزيد من البيانات والمعلومات التفصيلية عن جميع المجالات بهدف التخطيط والتنمية الاقتصادية للبلاد. لم يعد الحصر الشامل كذلك يقتصر على الميادين التقليدية كتعداد السكان، والزراعة بل أصبح يمكننا من اكتشاف ميادين جديد بتفصيل أكثر من خلال الإحصاءات المتعلقة بالقوة العاملة، وإحصاءات التجارة الخارجية، والداخلية، وإحصاءات البنوك والمواصلات، والدخل وغيرها من الظواهر.

## I-2-2- مزايا الحصر الشامل:

في حالة انعدام أخطاء في القياس و الحساب، يمدنا الحصر الشامل بمعلومات شامل وكاملة عن كل وحدات المجتمع قيد الدراسة، وذلك وفق الأسئلة المطروحة، وهو يشكل وثيقة تاريخية ذات أهمية كبير جدا والتي يمكن استغلالها في اعداد بحوث ودراسات تخص نفس المجتمع. كما يمكّن الحصر الشامل من توفير معلومات دقيق حول وحدات إحصائية نادرة لا توجد أو لا تتوفر عنها معلومات مسبقا، كما يمكن الإحصاء الشامل من وضع إطار شامل للمعاينة. إذ يمكن مثلا من وضع سجلات للحالة المدنية وقوائم بأسماء جميع المواطنين، والمؤسسات بمختلف أنواعها، وتعتبر مسألة وجود أطر معاينة كاملة وشاملة وحديثة عن مجتمع ما، مسألة جد هامة، وضرورية تساهم كثيرا في إنجاح دراسات تعتمد على المسوحات بالمعاينة والتي عادة ما تؤدي إلى الحصول على نتائج فعالة.

## I-2-3- عيوب الحصر الشامل وحدوده:

يتطلب إعداد الحصر الشامل إمكانيات مادية وبشرية كبيرة جدا، بحيث يستدعي الأمر معالج الملايين من الوثائق والمعلومات وهذا ما يفسر لجوء الدولة إلى هذا الأسلوب في جمع البيانات فقط عند الحاجة الملحة جدا والاستراتيجية، وهذا ما يفسر كذلك إجراء الحصر الشامل مر كل 5 سنوات أو 10 سنوات.

كما يتطلب الحصر الشامل توفر معدات حاسبة ضخمة، ومتخصصين كثيرين، وهياكل قاعدية معينة، مثل وسائل النقل والمكاتب، ولا يمكن للعديد من الدول خاصة الفقير منها والنامية من جمع هذه الإمكانيات.

ويتطلب الحصر الشامل إجراء تربصات مكلفة للعدادين المساعدين في جمع المعلومات، والذين ينحدرون من مستويات اجتماعية مختلفة، ومنهم الموظفون الحكوميون، والطلبة،

والعمال المتقاعدين، والتجار، والفلاحين. وهذا يعني أنهم لا يملكون المعلومات الكافية والمؤهلات التي تمكنهم من الإلمام التام والجيد بمحتويات الأسئلة المطروحة.

كما يستحيل القيام بالحصر الشامل إذا كانت المجتمعات غير منتهية، كدراسة نوعية مياه نهر، أو دراسة حول الأشجار في مدينة ما، أو دراسة تشمل طيور من صنف معين.

نظرا للكم الهائل من الوحدات الواجب استجوابها في الحصر الشامل فهذا يؤدي حتما ورغم انتباه العدادين إلى حدوث أخطاء كثيرة نذكر من أهمها أخطاء التسجيل وأخطاء السهو من طرف العداد بسبب التعب، وأخطاء الحساب وأخطاء إدخال المعلومات في الحواسيب...الخ.

## I-3.المسح بالمعاينة:

### I-3-1-أسباب استعمال أسلوب المسح بالمعاينة:

1. إذا كان المجتمع غير منته فإنه ليس بإمكاننا دراسة كل وحداته، لدا نكتفي بدراسة جزء فقط منه ، وتوجد عدة أمثلة عن ذلك نذكر منها مثلا تلك المتعلقة بدراسة الحشرات في غابة ما، أو نوعية المياه في نهر، ونجد كذلك الأبحاث المتعلقة بالكائنات الحية في البحر كالأسماك والنباتات...الخ.

2. المسح بالمعاينة يتطلب جهدا أقل وبالتالي تكلفة أقل مقارنة باستعمال الحصر الشامل، بعض القضايا والمسائل لا تتطلب إنفاق تكاليف باهضة بهدف الحصول على نتيجة لا تتناسب مع هذه التكاليف. لنفرض مثلا أننا بحاجة في مدينة وهران إلى معرفة رأي المواطنين حول إنجاز مركب رياضي أو مصنع كيماوي، فالأمر سيكون متعبا جدا ومكلفا إذا ما حاولنا استجواب كل مواطن من أجل معرفة رد فعله.

3. نلجأ إلى المعاينة بصفة مؤكدة إذا كانت عملية مشاهدة أو دراسة الوحدة الإحصائية تؤدي إلى إتلافها نهائيا، فإذا كان مثلا بصدد دراسة مدة حياة مصابيح في مصنع ما، في هذه الحالة فإن مشاهدة أو دراسة كل مصابيح المصنع سوف تؤدي إلى ضياعها كلها.

4. نحن بحاجة في بعض الأحيان إلى الحصول على معلومات حول مسألة ما في أقرب وقت ممكن بهدف اتخاذ قرار لا يمكن تأجيله، في هذه الحالة يمكننا المسح بالمعاينة من تحقيق هدفنا بسرعة أكبر مما يسمح لها الحصر الشامل.

5. يتطلب الحصر الشامل تجنيد عدد كبير من الخبراء الإحصائيين و العدادين الذين يكون أغلبهم غير متخصصين في مسألة جمع البيانات الإحصائية، وبالتالي يتعين على الإحصائي إعطاؤهم بعض التعليمات والإرشادات التي تساعدهم في الحصول على المعلومات بشكل جيد، إلا أن حجم الاستثمارات المستعملة في الحصر الشامل وكثرة أسئلتها تجعلها عرضة لتأويلات مختلفة من طرف العداد، وهذا ما يزيد من نسبة الوقوع في خطأ المعاينة وأخطأ التسجيل أو عدم الإجابة (الأجوبة الناقصة)، أو سوء فهم السؤال وبالتالي سوء الإجابة...الخ. أحيانا تكون نتائج المسح بالمعاينة أكثر نجاعة ومصداقية من نتائج الحصر الشامل.

6. في الأبحاث الطبية حيث يتعين تجريب أدوية جديدة على المرضى فإنه ونظرا لأسباب أخلاقية، لا يمكننا تجريب هذه الأدوية على كل المرضى، حيث من الممكن أن يكون لهذه الأخيرة مضاعفات خطيرة كما أنه للمريض رأيه الخاص في قبول العلاج أو رفضه، لذا نستعين بعينة فقط من المرضى من أجل الحصول على نتائج البحث.

### I-3-2- لمحة تاريخية عن المعاينة ( سبر الرأي):

لقد ظهرت عملية سبر الرأي انطلاقا من الرغبة في الحصول على معلومات كمية حول الحالة النفسية للمجتمع، من أجل تحقيق هذا الهدف نجد أن كلا من « Daniel Defor » في إنجلترا والكونت « Lafayette » بفرنسا وضعا شبكات الاتصال والمراسلة موزعة جغرافيا وكان ذلك في أوائل القرن التاسع عشر.

لكن المكاتب الحقيقية لسبر الآراء وجدت في الولايات المتحدة الأمريكية، وذلك بمناسبة التغطية الإعلامية للانتخابات الرئاسية. منذ سنة 1824، قامت كل من " Harrisbourg Pennsylvania" و "Raleigh Star" بمسوحات عينة بسبر الآراء وذلك بهدف إعطاء تقديرات حول نتائج الانتخابات، وذلك انطلاقا من استجواب عينة من الناخبين باستعمال مبدأ الانتخاب بالقشة (Vote de paille).

انتشرت بعد ذلك هذه الوسيلة على مستوى الجرائد الأخرى نظرا للشعبية التي نالتها في أوساط المواطنين ونذكر من أهمها " American Chicago"،"New York Herald"، " Columbus Dispatch"، و"Chicago Journal " ،وخاصة "Literary Digest" ابتداء من سنة 1916.

لم تكن العينات المسحوبة ترتكز على أي حد أدنى من معايير التمثيلية لأن هذا المصطلح ظهر فقط سبعين سنة من بعد ذلك التاريخ (1916)، لكن هذه الجرائد كانت تعتمد على الحجم الكبير للعينة، فنجد مثلا أن جريد New York Harald استجوبت أكثر من 30.000 مواطن في سنة 1905، واستجوب Literary Digest أكثر من مليوني (02) ناخب وذلك في سنة 1936.

وتعتبر هده السنة الأخيرة حدثا تاريخيا ومنعرجا في مسألة سبر الآراء وهو تاريخ الإعلان عن نتائح الانتخابات الرئاسية في الولايات المتحدة الأمريكية. في حين تنبأت Literary Digest بفوز المشرح London باستعمال تقنية الاقتراع بالقشة Vote de paille، كان الفوز حليفا المشرح F.D Roosevelt، والذي تنبأت بفوزه كل من:George Gallup ، و Elmo Roper ، وArchibald Crossley . هذه النتائج الأخيرة كانت نتيجة استعمال مبدأ التمثيلية عن طريق السحب العمدي للأفراد بطريقة الحصص.

رغم أن العينة المستعمل في الدراسة التي قامت بها جريد Literary Diges كانت كبيرة إلا أنه تبين بعد ذلك أنها كانت متحيزة.

تم التخلي عن طريقة الاقتراع باستعمال مبدأ القشة لتأخذ مكانها تقنيات العينة الممثلة للمجتمع والمستعملة من قبل مؤسسات متخصصة في عملية سبر الآراء.

كما ظهر في سن 1938 أول معهد لدراسة الآراء في بريطانيا، ثم تبعه في فرنسا أول معهد فرنسي للرأي العام IFOP، بعدها سارت العديد من الدول على هذا المنهج.

## I-3-3- أساليب المعاينة العشوائية:

نظرا لاختلاف طبيعة وخصائص مجتمع البحث من حالة إلى أخرى من ناحية، واختلاف الهدف من الدراسة من ناحية أخرى، ونظرا لأن الهدف من استخدام أسلوب المعاينة هو الرغبة في الحصول على بيانات عن مجتمع الدراسة بأقل تكلف ممكنة و في وقت قصير ومحدود، وجعل الأخطاء الناتجة عن المعاينة عند حدها الأدنى حتى تكون النتائج دقيقة. لكل ما تقدم هناك عدة أساليب للمعاينة العشوائية، من أهمها نجد أسلوب المعاينة العشوائية البسيطة، و أسلوب المعاينة العشوائية الطبقية و أسلوب المعاينة الانتظامية (المنتظمة) و أسلوب المعاينة العنقودية و أسلوب المعاينة المتعددة المراحل.

## أ-العينة العشوائية البسيطة:

يمثل أسلوب المعاينة المسمى بالمعاينة العشوائية البسيطة أهم الأسس التي يبنى عليها الإحصاء الاستقرائي كما تسحب العينة في هذا النوع من الأساليب بطريقة عشوائية محضة، بحيث يكون لكل وحدة اختيار في المجتمع نفس احتمال الانتماء إلى العينة المسحوبة والذي يساوي $\frac{1}{N}$. ولكل العينات الممكن تشكيلها من هذا المجتمع نفس احتمال للسحب.

يستعمل أسلوب المعاينة العشوائية البسيطة في الحالات التي تكون فيها وحدات المجتمع قيد الدراسة متجانسة فيما بينها، ونقصد بكلمة تجانس اقتراب صفات وخصائص وحدات الاختيار من حيث التشابه والتماثل.

وعلى الرغم من المزايا التي يتصف بها هذا النوع من الأساليب إلا أن هناك عدة عيوب ترد عليه لعل من أهمها ما يلي:

1- إذا كان المجتمع الإحصائي يتألف من وحدات اختيار غير متجانسة من حيث الظاهر أو الظواهر التي نعتبرها رئيسة، فإنه من الممكن جدا أن نحصل على عينة لا تتوزع فيها الوحدات بنفس الكيفية التي تتوزع فيها هذه الوحدات في المجتمع. وفي هذه الحالة فإن الصورة التي تعطيها العينة عن المجتمع الإحصائي قد تكون متحيزة. فمثلا، إذا كان المجتمع الإحصائي عبارة عن طلبة أحد الكليات، وكان يتألف من 60% ذكور و40% إناث، فإذا كنا بصدد اختيار 100 طالب فمن الممكن جدا أن يكون كل هؤلاء الطلبة من الذكور، أو يكون كلهن إناثا ومن الممكن كذلك أن تتألف العينة من 30% ذكور و70% إناث، في كل هذه الحالات وغيرها إذا كانت الدراسة ترتكز على معالم لها علاقة وطيدة بالجنس، فإن العينة المأخوذة تعطي صورة متحيزة عن هذا المجتمع.

2- إذا كان المجتمع الإحصائي كبيرا جدا، وكان حجم العينة المراد سحبها كبيرا كذلك، فإن استخراج مفردات العينة العشوائية سوف يحتاج مجهود كبير، كما يمكن أن تكون مفردات المجتمع منتشرة بصورة واسعة في مناطق شاسعة، بحيث أنه يمكن أن يقع الاختيار على عدد قليل من الوحدات الموجودة في مناطق نائية فإن تكاليف التنقل إلى هذه الوحدات تكون كبيرة ناهيك عن امتداد مدة إتمام العمل الميداني نظرا لطول المسافة الفاصلة بين مختلف وحدات الدراسة التي وقع عليها الاختيار، بالإضافة إلى كل هذا، يصعب كذلك للمراقب الإحصائي متابعة العمل الميداني للعدادين في مثل هذا النوع من الحالات.

## ب- العينة العشوائية الطبقية:

إذا كان المجتمع الإحصائي يتألف من وحدات اختيار غير متجانسة فيما بينها من حيث الظاهرة المدروسة، فإنه يتعين علينا تقسيم هذا المجتمع إلى مجتمعات جزئية تكون الوحدات المنتمية إليها متجانسة نوعا ما فيما بينها، نسمي هذه المجتمعات الجزئية بالطبقات. بعد ذلك نسحب من كل طبقة عينة عشوائية بسيطة، بحيث يشكل مجموع العينات الجزئية المسحوبة من كل طبقة ما يسمى بحجم العينة **n** المراد سحبه.

توجد أسباب أخرى تفرض نفسها تدفعنا إلى استخدام هذا النوع من الأساليب بدلا من المعاينة العشوائية البسيطة، لعل من أهمها رغبة الباحث في دراسة كل طبقة من طبقات المجتمع على حدا في محاولة لمقارنتها فيما بينها. فمثلا إذا كنا نريد دراسة المستوى المعيشي لسكان مدينة وهران، فإنه من المفيد جدا أن ندرس هذه الظاهرة في كل من الريف، والمدينة، وهما طبقتان أساسيتان يتكون منهما المجتمع الوهراني.

فيما يخص عدد الوحدات الواجب سحبها من كل طبقة، يمكننا مثلا أن نأخذه متناسبا مع حجم الطبقة، كما يمكن أن نسحب من كل طبقة نفس عدد الوحدات، كما يمكن أن نسحب عدد الوحدات في الطبقة بدلالة تكاليف البحث أو بدلالة شدة التباين والتشتت بين الوحدات في كل طبقة معبر عنها بالانحراف المعياري، أو أي عامل آخر.

## ج- أسلوب المعاينة الانتظامي (systématique):

يوجد أسلوب آخر من أساليب المعاينة العشوائية والذي يمكن أن نعتبره أكثر نجاعة وفاعلية من أسلوب المعاينة العشوائية البسيطة، ويسمى هذا الأسلوب بالمعاينة المنتظمة، يمتاز هذا النوع من العينات بسهولة تطبيقه خاصة في المجتمعات الكبيرة الحجم، كما يمتاز بسهولة استخراج وحدات العينة من القوائم والسجلات المسجلة فيها.

يتم اختيار الوحدات الإحصائية بحيث يفصل ما بين كل وحدة وأخرى مجال معين وثابت، يمكن لهذا المجال أن يكون زمنيا أو فضاء معبر عنه بمسافة أو يكون عبارة عن ترتيب أو تصنيف رقمي، فمثلا إذا كنا بحاجة إلى دراسة جودة الإنتاج في مصنع لإنتاج المصابيح، وكان هذا المصنع ينتج 60 مصباحا في الساعة، يمكن أن نسحب كل نصف ساعة مصباح حتى نحصل مثلا على 16 مصباح ليشكل العينة المدروسة في نهاية اليوم، وإذا كنا مثلا بحاجة إلى اختيار 10 طلبة من قائمة تتكون من 100 طالب، فإنه يمكننا سحب الطالب

العاشر، والعشرين، والثلاثين ورقم 100. يمكن تطبيق الأسلوب الانتظامي من السحب فقط إذا كانت الوحدات عبارة عن منازل في مدينة ما، أو أسماء مواطنين في قوائم انتخابية، أو ترتيب جنود في صفوف...إلخ.

إذا كان حجم العينة المطلوب هو n، وحجم المجتمع الإحصائي N فإن الكسر $\frac{N}{n}$ يمثل المجال الفاصل بين كل وحدة وأخرى على أن يتم اختيار الوحدة الابتدائية بصفة عشوائية، إذا تحققت العشوائية في سحب الوحدة الأولى في السحب الانتظامي، فإن لكل وحدة من المجتمع نفس الاحتمال في الانتماء إلى العينة. إذا كان المجتمع غير متجانس فإنه بإمكان السحب الانتظامي إعطائنا عينة أحسن تمثيلية من العينة المحصل عليها بالمعاينة العشوائية البسيطة، فمثلا إذا كان المجتمع الإحصائي مكون من مجموعات تحتوي على وحدات لها نفس الخصائص، وكانت هذه المجموعات مرتبة مع بعضها في الإطار، فإنه من المنتظر أن تمثل كل مجموعة منها بنفس وزنها في المجتمع الإحصائي المأخوذة منه.

وعلى الرغم من المزايا الكثيرة التي يتصف بها على النوع من الأساليب في سحب العينات والمرتكزة خاصة في سهولة اختيار الوحدات في المجتمعات الكبيرة وقلة تكاليف النقل بين الوحدات خاصة إذا كانت هذه الأخيرة مساكن أو مصانع موجودة بجوار بعضها البعض في مساحات جغرافية محدودة. إلا أن أهم العيوب التي ترد عليه وتجعله متحيزا نوعا ما هو ما إذا كان ترتيب الوحدات داخل الإطار بطريقة مرتبطة مع البيانات المرغوب جمعها وكذا ظاهرة الدورية (Phénomène de périodicité) الممكن وقوعها في المجتمع؛ فمثلا إذا كانت آلة معينة تنتج بطريقة منتظمة مصباح فاسد بعد كل أربعة مصابيح جيدة و طلب منا سحب مصباح بعد كل 9 مصابيح منتجة، أي أننا تختار المصباح ذو الرتبة 10، ثم 20... وعليه فلن تكون العينة ممثلة للمجتمع أحسن تمثيل.

**<u>د- أسلوب المعاينة العنقودية:</u>**

في كثير من الأحيان يصعب على الباحث الحصول على إطار معاينة للمجتمع المراد دراسته، بل تسحيل هذه العملية في أحيان أخرى نظرا لأسباب متعددة، فمثلا إذا لم تتوفر لدينا قائمة حديثة تحمل أسماء كل الأسر الساكنة في مدينة وهران فإنه من المكلف جدا ومن الصعب إعادة تشكيلها من جديد، بهدف الشروع في دراسة بسيطة محدودة الميزانية. من أجل تفادي هذا المشكل يوجد أسلوب آخر من أساليب المعاينة العشوائية والمسمى أسلوب

المعاينة العنقودية. في هذا النوع من الأساليب يكون المجتمع الإحصائي مقسما طبيعيا إلى مجموعات تحتوي كل واحدة منها على مجموعة من الوحدات الإحصائية، تسمى المجموعات بالعناقيد، ويمكن أن تكون هذه العناقيد عمارات تحتوي على مساكن، هذا إذا كانت الوحدة الإحصائية هي المسكن، أو بلديات تحتوي على أسر، أو يكون العنقود عبارة عن مدرسة بها مجموعة من التلاميذ، في كل هذه الحالات وعند سحب وحدات العينة نقوم باختيار مجموعة من العناقيد بشكل عشوائي، ثم نسحب من كل عنقود مختار عدد معين من الوحدات حسب حجم العينة المطلوب.

حتى وإن يبدو للواحد منا أنه في كل من أسلوب المعاينة العشوائية الطبقية والمعاينة العنقودية يتم تقسيم المجتمع الإحصائي إلى مجتمعات جزئية، فإن الأسلوبان يطبقان في حالات مختلفة تماما عن بعضها البعض، إذ نقوم في المعاينة الطبقية بتشكيل طبقات تحتوي على وحدات متجانسة فيما بينها، أما الطبقات فيكون الاختلاف بينها كبيرا.في حين تحتوي العناقيد على وحدات غير متجانسة فيما بينها، بينما تكون العناقيد متجانسة نوعا ما ومتشابهة، كما أن العناقيد تكون مقسمة و موجودة بشكل طبيعي.

من أجل حجم عينة معطى، يمدنا أسلوب المعاينة العنقودية بنتائج أقل دقة من أسلوب المعاينة العشوائية البسيطة، والمعاينة الطبقية. لكننا نعوض النقص في الدقة بالتكاليف القليلة التي تساعد الباحثين والبحوث ذات الميزانية المحدودة جدا.

**ه- المعاينة العشوائية المتعددة المراحل:**

توجد أساليب أخرى أكثر تعقيدا يمكن للباحث تطبيقها، فنجد مثلا أسلوب المعاينة العشوائية على مراحل. في هذا النوع من العينات يتم اختيار وحدات الدراسة على مرحلتين أو أكثر، فإذا كنا مثلا بصدد إنجاز دراسة حول الأسر في مدينة وهران، نقوم في مرحلة أولى باختيار عينة من الدوائر ثم في مرحلة ثانية نختار من كل دائرة عينة أخرى من الأسر، وهنا نكون قد مررنا بمرحلتين من أجل اختيار الوحدات

1- المرحلة الأولى: اختيار عينة من الدوائر.

2- المرحلة الثانية: اختيار عينة من الأسر.

أما إذا كانت وحدة الاختيار هي الأفراد فإنه يمكننا كمرحلة ثالثة سحب عينة من الأفراد في كل أسرة.

يهدف هذا الأسلوب إلى تخفيض التكاليف وربح الوقت خاصة إذا كان المجتمع كبيـــرا جدا وكانت الوحدات متباعدة ومشتتة في مساحات شاسعة.

## I-3-4- أهم أنواع العينات غير العشوائية أو غير الاحتمالية:

كل الأساليب المعروضة في الفقرات السابقة كانت تسمى عشوائية، وذلك نظرا لإمكانية معرفة احتمال انتماء أي مفردة إلى العينة المسحوبة، وكذا استعمال مبدأ العشوائية في اختيار الوحدات.

تحتاج هذه الأساليب إلى توفر أطر للمعاينة تكون نوعا ما كاملة وشاملة وتحتوي على كل وحدات المجتمع الإحصائي. إلا أن الباحث يمكن أو يجد نفسه في كثيرا مـــن الأحيـــان مواجها لمشكلة صعوبة الحصول على هذا النوع من الأطر، سواء كان هذا المشكل مادي أو مشكل يتعلق بالبيروقراطية التي تشتهر بها إدارتنا المركزية، أو نظرا لأسباب أمنية تـــدخل ضمن إطار الحفاظ على حقوق الأفراد في إخفاء معطياتهم وبياناتهم الشخصية، مـــن أجـــل تفادي هذه العقبات أوجد الباحثون والإحصائيون أساليب أخرى للمعاينـــة تســـمى بأســـاليب المعاينة غير الاحتمالية أي غير العشوائية، ونظرا لكوننا نجهل احتمال انتماء أي مفردة إلى العينة فإنه من غير الممكن تطبيق نظرية الاحتمالات ولا حتى تطبيـــق أصـــول الاســـتدلال الإحصائي. ويدعى هذا النوع من العينات بالعينات الشخصية.

ترتكز مختلف الأساليب غير الاحتمالية على المبدأ التالي: لنفرض أننـــا علـــى درايـــة بخصائص متغيرة ما x من المجتمع، توجد طرق تمكننا من الحصول على عينة مثلى حـــول قيم هذه المتغيرة وذلك إذا درسنا متغيرة y يكون سلوكها مرتبطا كثيرا بسلوك المتغيـــرة x، ونضع فرضية كون العينة مثلى كذلك بالنسبة للمتغيرة y.

إن من أهم عيوب هذا الأسلوب هو عدم إمكانية حساب أخطاء المعاينة، لكن يمكنها أن تمدنا بنتائج قريبة من النتائج المحصل عليها باستعمال أساليب عشوائية، وهذا في حالة ما إذا قمنا بتطبيقها بصفة جيدة.

ولعل من أهم أنواع أساليب السحب غير العشوائي نجد المعاينـــة بالحصـــص (Quota) والمعاينة المنتقاة (sélective) والمعاينة كبيرة الحجم (Extensive samples).

## أ- العينات بالحصص:

ينقسم المجتمع الإحصائي إلى مجموعات متجانسة من الوحدات حسب الخصائص والمميزات من وجهة نظر الباحث نفسه، ويتحكم هذا الأخير في عدد كل مجموعة (حصة) على حسب أهميتها الخاصة ودرجة تجانس الوحدات الإحصائية المنتمية إليها. تستعمل هذه الأساليب عادة في استطلاع الرأي، وفي البحوث التسويقية لمعرفة الاتجاه العام، والسلوك الشائع لدى الوحدات الإحصائية، بحيث تكون الدقة الكبيرة جدا هي للمعيار الأساسي في هذا النوع من الأساليب. إذ يكتفي العداد بالحصول على العدد المطلوب من الوحدات الإحصائية بغض النظر عن كيفية الحصول عليها ولا عن استعمال مبدأ العشوائية. كأن يطلب من كل عداد مثلا استجواب 150 فردا من سكان مدينة وهران أو زيارة 10 أسر في عمارة معينة من عمارات وسط المدينة.

عادة ما يسعى الباحث إلى أن تتناسب عدد المفردات المشكلة للعينة مع عدد المفردات الموجودة فعلا في المجتمع الإحصائي، كأن يحافظ مثلا في العينة على نسبة الذكور والإناث في جامعة ما، أو في مدينة ما. وهو بذلك (الباحث) يسعى إلى بلوغ الأمثلية حتى وإن لم يستعن بمبدأ العشوائية.

لقد أثبتت بعض التجارب فاعلية هذا الأسلوب لكونها قد أعطت نتائج جيدة، ومع ذلك فإن هناك احتمالا كبيرا لأن تكون العينة مشوبة بشوائب التحيز لأنه عادة ما يسعى العداد إلى الاتصال بالوحدات الإحصائية التي تكون سهلة الاستجواب، فمثلا من المحتمل أن تقتصر الإجابات المأخوذة من سكان المدن الكبرى على أولئك الذين يسكنون في الطوابق السفلية والتي يسهل الوصول إليها، أو على سكان وسط المدينة بدلا من أطرافها، ونظرا لكل ذلك فإن هذا النوع من الأساليب لا يمكننا من الحصول على عينات تمثل المجتمع أحسن تمثيل، وبالتالي لا تعطي هذه العينات صورة مصغرة عن المجتمع الإحصائي المأخوذة منه.

## ب- العينات المنتقاة:

في هذا النوع من أساليب المعاينة غير الاحتمالية يقوم الباحث بصفة عمدية باختيار الوحدات الواجب انتماؤها إلى العينة، وذلك وفقا لخبرته ومعرفته حول المجتمع المدروس، وترتبط جودة العينة المحصل عليها على مهارة الباحث ومدى إلمامه بالظاهرة قيد الدراسة. في العينات المنتقاة وكغيرها من العينات غير الاحتمالية لا يمكننا قياس أخطاء المعاينة ولا

استعمال نظرية الاستدلال الإحصائي نتيجة لعدم قدرتنا على تحديد احتمالات انتماء كل وحدة إحصائية إلى العينة.

لقد أثبتت العينات المنتقاة فعاليتها في بعض المجالات مثل حساب مؤشر الأسعار والذي يرتكز فيه الباحث على اختيار مجموعة من السلع الاستهلاكية والتي يراها هو الأحسن تمثيل لمجمل السلع الموجودة في السوق، وتطبق هذه الطريقة في اختيار السلع من أجل حساب مؤشر الأسعار في أغلبية الدول إن لم نقل كلها.

كما يستعمل هذا النوع من الأساليب خاصة في البحوث الاجتماعية التي تشتد فيها درجة الاختلاف بين الوحدات الإحصائية، بحيث يضطر الباحث إلى تحديد مجموعة الوحدات التي تصلح لإجراء الدراسة وهذا من وجهة نظره.

**<u>ج- أسلوب المعاينة للعينات كبيرة الحجم:</u>**

يسمى هذا النوع من العينات كذلك بالعينات الواسعة (Extensive samples) وذلك نظرا لعمله بمبدأ سحب عينة ذات الحجم الكبير جدا بطريقة غير منتظمة وكيفما كان، ويعتقد العاملون بمثل هذا النوع من الأساليب أنه كلما كانت العينة كبيرة فإن النتائج المحصل عليها سوف تقترب بالضرورة من النتائج الحقيقية، وذلك بغض النظر عن استعمال الأساليب الاحتمالية (العشوائية) في اختيار الوحدات.

إن هذا الاعتقاد محقق فقط في الحالة التي يكون فيها المجتمع منتهيا، بحيث أننا نسحب حجم كبيرة جدا من الوحدات بحيث يقترب حجم العينة من حجم المجتمع نظرا لأن أخطاء المعاينة لا بد وأن تتناقص في حالات كهذه، أما إذا كان المجتمع غير منته فإنه لا يمكن أبدا التحقق من هذا الافتراض ما لم يقترب حجم العينة من حجم المجتمع.

وأخيرا فإن طبيعة الدراسة ونوعية المجتمع موضع البحث وخبرة الباحث ووجهة نظره هي الأمور التي يعتمد عليها في تحديد نوع أسلوب المعاينة المختار وكيفية استعماله من أجل تقليص الأخطاء.

نظرا لكون الدراسة في أسلوب المعاينة تقتصر فقط على جزء فقط من وحدات المجتمع، فإنها لا تمكننا من إعطاء بيانات ومعلومات شاملة حول الظاهرة المدروسة.

في حالة وجود أخطاء تتعلق بتصميم البحث أو تقديم معالم المجتمع، فإنه يمكننا الحصول على نتائج غير دقيقة في بعض الحالات خلافا للحصر الشامل الذي لا يحتاج إلى استعمال تقنيات التقدير إذ يكتفي فقط بحساب معالم المجتمع ووصفها كما هي.

رغم العيوب الموجودة في أسلوب المعاينة، إلا أنه أصبح الأكثر انتشارا واستعمالا من طرف الباحثين في جميع الميادين مقارنة بالأساليب الأخرى في جمع البيانات.

## I-4- خطوات تصميم البحث الميداني بالمعاينة:

تمر عملية الإعداد للبحث الميداني وإنجازه بعدة مراحل مهمة يجب على كل باحث الإلمام بها حتى يتمكن من بلوغ أهدافه المسطرة بشكل علمي ومنهجي وفعال، ولعل من أهم هذه الخطوات نذكر ما يلي:

### I-4-1- تحديد المشكلة:

تبدأ العملية الإحصائية بمشاهدة الظواهر التي نرغب في دراستها، ومن هنا يتولد الإحساس بالمشكلة ووضع فرض مبدئي لتفسير الظاهرة موضوع البحث، فإذا كانت المشكلة تهم الباحث فيتطلب الأمر منه تقسيمها وتحديد أبعادها وتصور الحلول الممكنة لها. ويتأتى ما سبق بوضع فرض مبدئي لتفسير الظاهرة، ولكي يصل الباحث إلى تفسيرها عليه أن يدرس الظاهرة ويحللها ويستخدم نتائجها ثم يقرر الباحث إما قبول جزئي أو كلي للفرض المبدئي.

ويمكن للباحث صياغة المشكلة بعبارة لفظية تقريرية، أو يطرح سؤالا واحدا أو عدة أسئلة يحاول الإجابة عليها.

### I-4-2- أهداف البحث:

تعتبر عملية تحديد أهداف البحث مهمة جدا، وذلك قصد تحديد البيانات المطلوب جمعها واستخدامها من قبل الباحث.

ويقوم الكثير من الباحثين بصياغة الأهداف بشكل مفصل بغرض توضيح النقاط الأساسية التي يرغب الباحث في الوصول إليها.

### I-4-3- عنوان البحث:

لا يمكن تصور أي بحث بدون وضع عنوان له، شريطة أن يكون هذا العنوان واضحا ومختصرا وبلغة سهلة تعبر عن المشكلة التي نقوم بدراستها.

## I-4-4- حدود البحث:

على الباحث أن يضع حدودا زمانية ومكانية للبحث، مع عدم تجاوزها، بهدف تحقيق الأهداف المرجوة من البحث، ونقصد بالحدود الزمنية الفترة التي سوف يستغرقها إنجاز البحث، كأن نحدد مثلا تاريخ الشروع في الدراسة وتاريخ الانتهاء منها، فمثلا يمكن للواحد منا أن يدرس مشكلة التحصيل العلمي في الجامعات و يحدد الموسم الجامعي المقصود كأن يكون مثلا 2007/2006. أما المقصود بالحدود المكانية، فهي المناطق الجغرافية التي يشملها البحث، وفي مثالنا السابق يجب أن نحدد أي الجامعات سوف تدخل في نطاق الدراسة، فمثلا يمكن أن نحدد الكليات الموجودة في جامعة وهران، وبذلك تكون المساحة الجغرافية لولاية وهران هي الإطار المكاني للبحث. أما الجامعات الأخرى الموجودة في بقية ولايات الوطن فلا تدخل في البحث.

## I-4-5- تعريف وحدة المعاينة والمجتمع الإحصائي:

يجب على الباحث أن يعرف وحدة المعاينة (الاختيار) والمجتمع تعريفا دقيقا لا التباس فيه حتى يتمكن من جمع البيانات من الوحدات ذات العلاقة بدقة كبيرة، ويقصي جميع الوحدات التي لا تدخل ضمن إطار المعاينة الخاص بالمجتمع قيد الدراسة.

فمثلا عند دراسة الإنتاج والمبيعات في قطاع الحليب والأجبان، نختار عينة من المؤسسات في منطقة ما، وتشكل المؤسسة وحدة المعاينة في هذا المثال، وهي الشركة أو المؤسسة التي تقوم بإنتاج الحليب والأجبان بمختلف أنواعه.

أما المجتمع الإحصائي فهو يشمل جميع المؤسسات التي تنتج الحليب والأجبان في مدينة وهران مثلا.

## I-4-6- إعداد إطار المعاينة:

تتطلب عملية اختيار عينة من مجتمع ما توفر قائمة بأسماء الوحدات الإحصائية وعناوينها وأهم المعلومات المتعلقة بها. تسمى هذه القوائم أو السجلات التي تحتوي على كل وحدات المجتمع قيد الدراسة بإطار المعاينة، ويمكن أن يكون الإطار على شكل قائمة أو بطاقات أو سجل أو خريطة أو غير ذلك.

حتى تكون الدراسة الإحصائية بالمعاينة فعالة ودقيقة، من الضرورة جدا أن يتوفر للباحث إطار جيد تتوفر فيه الشروط التالي:

1. أن يحتوي الإطار على جميع وحدات المجتمع بدون استثناء.
2. أن لا تتكرر الوحدة الإحصائية في الإطار
3. استحداث إطار المعاينة قبل الشروع في المسح بإدخال تعديلات عليه بحذف أو إضافة وحدات جديدة.
4. عدم تداخل أو وجود إطارات للمعاينة في إطار واحد شامل، كأن نجد مثلا إطار للمؤسسات التجارية مع إطار للمؤسسات الصناعية.
5. أن يحتوي الإطار على عناوين وأماكن وحدات المعاينة وترتيبها بشكل جيد يساعد على الوصول إلى الوحدات بسهولة.

## I-4-7- تحديد البيانات المطلوب جمعها:

يتم تحديد البيانات والمعلومات المطلوب جمعها وفقا لأهداف البحث وذلك بعد استشارة الباحث بالإضافة إلى النصائح التي يمكن أن يقدمها الإحصائي المتمرن، والتي من شأنها أن تؤدي إلى جمع أكبر عدد ممكن من المعلومات حول الظاهرة المدروسة بهدف تفسيرها تفسيرا دقيقا.

## I-4-8- تحديد أسلوب المعاينة المتبع وحجم العينة:

وفقا لتجانس وحدات المجتمع من حيث الظاهرة المدروسة، وكذا الإمكانيات المادية والبشرية المتوفرة، يقوم الباحث بتحديد نوع أسلوب المعاينة الواجب تطبيقه في هذا المسح، سواء كانت معاينة عشوائية بسيطة أو طبقية أو عنقودية أو معاينة غير احتمالية، ويجب اختيار أسلوب المعاينة بدقة لاختلاف النتائج من أسلوب إلى آخر وكما رأينا في الفقرات السابقة، فإن لكل أسلوب معاينة مجالات استخدامه وخصائصه، لذا على الباحث أن يكون حذرا وملما بتقنيات المعاينة حتى يتسنى له اختيار وتطبيق الأسلوب الملائم.

كما يقوم الباحث بتحديد حجم العينة الضروري لإنجاز بحثه، وذلك حسب نوع أسلوب المعاينة المتبع، باستخدام الصيغة الرياضية التي تختلف من أسلوب إلى آخر، وذلك بمستوى ثقة معين بعد تحديد خطأ التقدير وعوامل أخرى سوف نتعرض لها بالتفصيل في جميع الفصول المقبلة والتي تمثل جوهر بحثنا هذا.

**I-4-9- طرق جمع البيانات:**

توجد عدة طرق تستعمل لجمع البيانات، ويمكن للباحث اختيار الأداة المستخدمة لـذلك، فمثلا للحصول على معلومات حول مؤسسة ما يمكن استجواب صاحبها مباشرة أو من خلال بيانات المؤسسات الإحصائية في الديوان الوطني للإحصاء أو مديرية الضرائب، وهنا نتحدث عن المصادر المباشرة وغير المباشرة في جمع البيانات.

ويمكن استخدام الهاتف أو الفاكس أو التلكس أو الانترنيت أو حتى المقابلة الشخصية أو البريد كطرق لجمع البيانات. ولكل طريقة مزاياها وعيوبها. ولعل من أهم وأبرز الوسائل في جمع البيانات والتي تتصدر طليعتها نجد الاستمارة الإحصائية.

**I-4-10- إعداد الاستمارة الإحصائية:**

**1-تعريف:**

الاستمارة الإحصائية أداة تستخدم لجمع البيانات، وهي عبارة عن وعاء كتابي يحتوي على الأسئلة التي تمكن الباحث من الوصول إلى تفسير الظاهرة المدروسة، وهي تحتوي على فراغات لتدوين الإجابات الممكنة والرموز، وتتطلب عناية فائقة أثناء تصميمها حتى تمكن من الوصول إلى الإجابات بدقة كبيرة.

فإذا أحسن الباحث وضع الأسئلة وأتقن صياغتها وراعى كل الشروط والاعتبارات في ذلك، تصبح الاستمارة في هذه الحالة أداة جيدة لنقل المعلومات بطريقة صحيحة. أما إذ لم يحسن الباحث ذلك فإنه يحصل على بيانات بطريقة غير سليمة لا تصلح لأن تكون أساسا للتحليل الإحصائي واستخلاص النتائج واتخاذ القرارات.

الاستمارة الإحصائية نوعان رئيسان، هما الاستمارة وصحيفة الاستبيان.

**أ-الاستمارة:**

في الاستمارة يقوم المستجوب بتدوين الإجابات مباشرة فور تلقيها من الوحدة الإحصائية من خلال المقابلة الشخصية أو الملاحظة، ولهذا النوع من الاستمارات عدة مزايا وعيوب.

*** مزايا الاستمارة:**

- الحصول على إجابات كاملة لجميع الأسئلة.
- توضيح الأسئلة غير المفهومة من قبل المستوجب.

- إمكانية استخدمها في حالات جمع البيانات عن طريق الملاحظة.
- توفر الوقت الكافي بالنسبة للمدلي بالبيانات.

*** عيوبها:**

- تتطلب تكاليف مادية وبشرية كبيرة نظرا لكونها تستدعي إجراء مقابلة شخصية لجميع وحدات العينة.
- تسبب إحراجا للمدلي بالبيانات خاصة إذا كانت الأسئلة شخصية أو محرجة.

رغم عيوبها تبقى هذه الاستمارة مستعملة في العديد من المسوحات بالمعاينة خاصة في الدول النامية نظرا للمستوى الثقافي والوعي الإحصائي المتدني في هذه المجتمعات.

**ب- صحيفة الاستبيان:**

تستعمل صحيفة الاستبيان في الحالات التي يتم فيها جمع البيانات بطريقة غير مباشرة، بحيث يتم فيها استعمال البريد أو الفاكس أو الانترنيت كوسائل للاتصال بالوحدات الإحصائية، ونظرا لكون الباحث لا يقوم بالمقابلة المباشرة مع المستجوبين فإن الاستبيان يستدعي احتوائه على بعض المعلومات الإضافية، كالخطاب الذي يوضح فيه الباحث أهمية الموضوع وكيفية ملء الاستبيان، وكذا تاريخ إرجاع. بالإضافة إلى إعطاء بعض التعاريف والشروحات حول الأسئلة التي تستدعي ذلك.

*** مزايا صحيفة الاستبيان:**

- تتطلب تكاليف مادية وبشرية منخفضة، كما توفر الكثير من الوقت في الحصول على البيانات.
- حتى في الحالات التي تكون فيها عدد وحدات العينة كبيرة، فإنه يمكن بلوغها مهما كانت درجة انتشارها وتباعدها عن بعضها البعض.
- يمكن للمدلي ببياناته من الإجابة عن الأسئلة بدون أي إحراج.

*** عيوب صحيفة الاستبيان :**

- عدم وصول عدد كبير من صحف الاستبيان إلى المرسل نظرا لأهمالها من طرف بعض المستجوبين (المدلي بالبيانات).
- تأخر وصول الإجابات، وفقدان بعضها خلال انتقالها عبر البريد.

- عدم اكتمال الإجابة على كل الأسئلة الموجودة فيها في بعض الحالات.

**<u>2- الشروط الواجب مراعاتها عند تصميم الاستمارة الإحصائية:</u>**

عند تصميم الاستمارة يجب على الباحث مراعاة عدة شروط حتى يتمكن من الحصول على استمارة تفي بالغرض المطلوب، و هي كالتالي:

**<u>1. من حيث شكل الاستمارة:</u>**

يلعب شكل الاستمارة دورا هاما في مساعدة وحث المدلي ببياناته على الإجابة والإمعان في الأسئلة المطروحة. حيث أن الراحة البصرية متصلة مباشرة مع الاستعداد النفسي، ولعل من أهم المواصفات التي يجب أن تتوفر في شكل الاستمارة ما يلي:

- أن يكون الورق المستعمل في الاستمارة من نوع جيد وشديد المقاومة للظروف الكثيرة التي يمكن أن تتعرض لها أثناء عملية فتحه وطيه واستخراجه عدة مرات متتالية، أي أن يكون الورق المستعمل لا يتلف بسهولة حتى لا يجد المستجوب نفسه مضطر لاتخاذ الحذر والخوف من إتلاف الاستمارة، كما يجب أن يتحمل الورق استعمال أي نوع من الأقلام أو السيالة.
- يجب أن يساعد لون الاستمارة القارئ على حسن ملاحظة الكتابة والحروف المستعملة، ويفضل أن يكون هذا اللون أبيضا لأنه يساعد نسبة كبيرة من الأفراد، خاصة ذوي البصر الضعيف أو حتى المصابين بمعمى الألوان.
- يجب أن يكون حجم الاستمارة معقولا وغير مبالغ فيه، فإذا كان عدد صفحات الاستمارة كثيرة جدا فهذا يدفع المستجوب إلى عدم التركيز في الأسئلة المطروحة، كما يؤدي به إلى الملل ومحاولة الانتهاء من الإجابة على الأسئلة بسرعة.
- يجب ترك أماكن ومساحات كافية للإجابة والترميز.
- كما يجب ترقيم الأسئلة حتى تساعد الباحث على تحليل البيانات وربط الأسئلة ببعضها البعض في محاولة منه لإبراز تأثير المتغيرات على بعضها البعض.
- يجب أن يحتوي غلاف الاستمارة على الجهة أو الهيئة المنفذة للبحث بالإضافة إلى ضرورة وجود عنوان للبحث وكذا خطاب يطمئن المدلي ببياناته حول سرية البيانات وعدم استخدامها في أغراض أخرى غير البحث العلمي المحض.

## 2. من حيث مضمون الاستمارة:

- يجب أن تصاغ الأسئلة بعبارة سهلة وواضحة وبمصطلحات شائعة الاستعمال.
- الاكتفاء بالأسئلة الهامة والمنطقية وتجنب الأسئلة الإضافية التي لا علاقة لها بأهداف الدراسة.
- يجب أن تتطلب الأسئلة إجابات واضحة ومحددة وبسيطة، وأن لا تحتاج هذه الإجابات إلى حسابات مطولة أو ذاكرة قوية من طرف المدلي بالبيانات.
- يجب تجنب طرح الأسئلة المحرجة قدر المستطاع مثل السؤال عن الدين، أو الجنس أو الانتماء السياسي...الخ.
- تجنب طرح أسئلة مفتوحة تحتمل عدة إجابات مطولة وينصح أن يقدم الباحث جميع الإمكانيات المتوقعة للإجابة وذلك في شكل وضع عدة أجوبة ممكنة يطلب فيها من المدلي بالبيانات باختيار إحداها وذلك بوضع علامة أمام الجواب المناسب.
- يستحسن وضع أسئلة للمراقبة تمكن الباحث من التحقق من مدى صدق المستجوب في إدلائه بالأجوبة، كأن تحتوي الأسئلة مثلا على سؤال حول عمر المستجوب، وسؤال آخر حول مستواه التعليمي، إذ لا يعقل أن يكون المستجوب متحصلا على دكتوراه دول في تخصص ما ويكون عمره لا يتعدى عشرين عاما.
- أن لا يكون السؤال موحيا بالإجابة المتوقعة، كأن يسأل مثلا عن رأي المستجوب في تفضيله للفقر أو الغنى، أو عن رأيه في إلغاء الضريبة عن الدخل.
- يجب ترتيب الأسئلة بشكل منطقي يساعد الباحث على استخلاص النتائج وكذا الربط بين متغيرات الدراسة.
- كما يجب ترميز الأسئلة وإجاباتها الممكنة حتى يتمكن الباحث من إدخال البيانات في الحاسب بسهولة.

بعد التعرف على أهم الخطوات و المراحل الواجب إتباعها عند إنجاز بحث ميداني و التي تضمنت في إحدى خطواتها مسألة اختيار أسلوب المعاينة وتحديد حجم العينة الضروريين لجمع المعلومات حول الظاهرة المدروسة، يجب على الباحث الآن الإلمام بكل أسلوب معاينة على حدا و ما يتضمنه من تقدير لبعض معالم المجتمع الإحصائي، ثم بعد ذلك تحديد المعايير الواجب مراعاتها من أجل تحديدي حجم العينة الضروري سحبه في مختلف أساليب المعاينة الإحصائية.

# الفصل الثانـي:
# تحديد حجم العينـة في أسلوب المعاينـة العشوائيـة البسيطة

سوف نتحدث في هذا الفصل عن كيفية تحديد حجم العينة عند استعمال أسلوب المعاينة العشوائية البسيطة لجمع البيانات. في الفقرة الأولى لا بأس أن نذكر بمفهوم المعاينة العشوائية البسيطة و أسباب استعمالها. كما سوف نعرض من الفقرة الثانية (2) إلى الفقرة الخامسة(5) بعض التعاريف والرموز، وكذا تقدير بعض معالم المجتمع كالمتوسط والمجموع و تبايناتها. تعرض الفقرة السادسة مقدر النسبة أما الفقرة السابعة فهي مخصصة للمعاينة العشوائية البسيطة في حالة النسب المئوية. الفقرة الأخيرة مخصصة لعرض تفصيلي عن كيفية تحديد حجم العينة من اجل تحقيق أهداف يحددها الباحث.

## 2-1المعاينة العشوائية البسيطة و أسباب دراستها :

المعاينة العشوائية البسيطة أسلوب يهدف إلى سحب n وحدة احصائية من بين N وحدة مكونة للمجتمع قيد الدراسة، بحيث يكون لكل من الـــ NCn عينة ممكنة نفس الفرصة في أن يتم سحبها، و احتمال سحب أي عينة من العينات متساوي ويقدر بـــ 1/NCn.

هناك عدة أسباب تدفعنا إلى مناقشة أسلوب المعاينة العشوائية البسيطة في هذا البحث:

1- يتم تفسير العديد من المبادئ الأساسية للمعاينة انطلاقا من العينة العشوائية البسيطة ثم بعد ذلك تتوسع التحاليل لتمس الأساليب الأخرى للمعاينة الأكثر تعقيدا.

2- تحت بعض الشروط الخاصة، يمكن أن يمنح أسلوب المعاينة العشوائية البسيطة بعض التسهيلات فيما يخص الدقة المرجوة من بعض الأساليب الأخرى للمعاينة والتي تكون فيها الصيغ الرياضية أكثر تعقيدا.

3- حاليا، يستعمل أسلوب المعاينة العشوائية البسيطة بدون أي تغيير في العديد من المسائل والدراسات الاقتصادية، والاجتماعية، وغيرها.

توجد الكثير من الأمثلة عن المعاينة العشوائية البسيطة، كأن تبحث مؤسسة ما عن الحصول على معلومات تخص عمالها في اطار دراسة حول الظروف المهنية أو المعيشية و مدا تأثيرها على المردودية في العمل، أو أن يود مجمع تجاري الحصول على معلومات تتعلق بزبائنه عند قيامه بدراسة تسويقية لمعرفة سلوك المستهلك. في هذه الحالات عادة ما تتوفر لدينا قوائم كاملة، بحيث يمكن بسهولة إجراء معاينة عشوائية بسيطة. مثال آخر بارز عن الطرق التي تقترب من أسلوب المعاينة العشوائية البسيطة وهو استعمال المعاينة

بالاستناد إلى الحصر الشامل. فمثلا لقد وجد أنه في الحصر الشامل للسكان في كل من الولايات المتحدة الأمريكية في سنة 1940 و1950 وكذا في كندا في سنة 1951، من الأحسن أخذ عينة من القوائم الشاملة للسكان، ثم القيام ببعض القياسات والتقسيمات على جزء من المجتمع وهو العينة بدلا من جميع الأفراد، وذلك بهدف الإسراع في الحصول على النتائج[2] وكذلك لأغراض اقتصادية.

فيما يخص كيفية سحب عينة عشوائية بسيطة، فإن الأمر يتمثل هنا في محاولة لتحقيق العشوائية في حد ذاتها، بحيث يكون لكل وحدة في المجتمع نفس احتمال الظهور أو السحب. لا يعتبر هذا الإجراء سهلا كما يمكن للبعض أن يتوقعه، لذا تم وضع مجموعة من الجداول للأعداد العشوائية تحت تصرف الباحثين. يمكن الحصول على هذه الجداول عن طريق إجراء الآلاف من عمليات السحب لأعداد محصورة بين 0 و99999 بحيث يكون لكل عدد من هذه الأعداد نفس الاحتمال في الظهور والذي يساوي 1/100000 .من بين الجداول الأكثر انتشارا نجد تلك التي نشرتها Rand corporation لسنة 1950, والتي تحتوي على مليون رقم, وتلك التي نشرتها Kendall et Smith (1938) والتي تحتوي على 100000 رقم. كما تتوفر العديد من الجداول في الكتب الإحصائية قديمة كانت أم حديثة. كما يمكن القيام بالسحب عن طريق اللجوء إلى برامج الحاسب بمختلف أنواعها (SPSS,EXCEL).أو باستعمال مختلف تقنيات السحب العشوائي مثل مبدأ اللوطو الرياضي أو ترقيم كريات و وضعها في حاوية...الخ.

## 2-2 تعاريف ورموز:

إن الهدف من إجراء المسوحات الإحصائية بالمعاينة هو التمكن من تقدير قيمة خاصية أو مجموعة من خصائص المجتمع والتي تسمى بالمعالم، ولكن قبل الخوض في عملية التقدير يتعين علينا وضع مجموعة من الرموز وإعطاء مجموعة من المفاهيم الأساسية التي سوف تسهل علينا المهمة. الرموز التي سوف نستعملها في هذا الفصل لا تخص فقط أسلوب

[2] Hansen, Hurwitz and Madow , "Sample survey Method and theory". Vol 1 et 2 , John Wiley and Sons, 1993, p110.

المعاينة العشوائية البسيط بل وكذلك جميع الأساليب التي سنعرضها في الفصول القادمة. سنشير إلى الصفات المميزة للمجتمع بأحرف كبيرة، ولتلك المتعلقة بالعينة بأحرف صغيرة ، نرمز بـ N إلى وحدات المجتمع، وب n إلى وحدات العينة ، كما نرمز بالحرف $f$ إلى كسر المعاينة والذي يعبر عن نسبة حجم العينة إلى حجم المجتمع، ويسمى كذلك معدل الاختيار وهو يساوي $f=n/N$

### 2-2-1 بعض القيم في المجتمع:

القيم الموافقة لوحدات المجتمع: $y_i \quad (i = 1, 2, 3, \ldots\ldots N)$

مجموع قيم وحدات المجتمع: $\mathrm{Y} = \sum_{i=1}^{N} y_i = y_1 + y_2 + \ldots + y_N$

متوسط قيم وحدات المجتمع: $\overline{Y} = \frac{Y}{N} = \frac{\sum^{N} y_i}{N}$

نسبة مجموعين أو متوسطين في المجتمع: $\mathrm{R} = \frac{Y}{X} = \frac{\overline{Y}}{\overline{X}}$

تباين وحدات المجتمع يأخذ هذا التباين صيغتين مختلفتين:

$$S_y^2 = \frac{\sum^{N}(y_i - \overline{Y})^2}{N-1}$$

أو

$$\sigma_y^2 = \frac{\sum^{N}\left(y_i - \overline{Y}\right)^2}{N}$$

تستعمل الصيغة الأولى $S_y^2$ لتسهيل وتبسيط العديد من النتائج، أما الصيغة الثانية $\sigma_y^2$ فتعتبر صيغة التباين لمتغير ما بدون إجراء أي تعديل عليها.

و يرمزللانحراف المعياري لوحدات المجتمع بـ: $\sigma_y$ أو $S_y$.

### 2-2-2 بعض القيم في العينة:

القيم الموافقة لوحدات العينة: $y_i \quad (i = 1, 2, 3, \ldots, n)$

مجموع قيم وحدات العينة هو $y$: $y = \sum_{i=1}^{n} y_i$

متوسط العينة: $\overline{y} = \frac{\sum_{i=1}^{n} y_i}{n}$

التباين في العينة:

$$s_y^2 = \frac{\sum^{n} (y_i - \bar{y})^2}{n-1}$$

وسوف نرى لاحقا أنه تقدير غير متحيز للتباين في مجتمع $S^2$.

نتطرق في هذا الفصل إلى مسألة التقدير وكيفية حساب هذه المقدرات والتي تخص ثلاث معالم أساسية يرتكز الاهتمام حولها في معظم الأحيان وهي:

1- المتوسط $\bar{Y}$: كأن نحاول مثلا تقدير معدل أعمار سكان مدينة ما في فترة معينة.

2- المجموع Y: فمثلا يمكن أن نحاول التعرف على مجموع دخل أفراد يعيشون في منطقة معينة انطلاقا من نتائج العينة المدروسة في إطار دراسة السوق.

3- نسبة مجموعين أو متوسطين $R = \frac{Y}{X} = \frac{\bar{Y}}{\bar{X}}$، مثل تقدير نسبة الذكور إلى الإناث في مدينة ما. أو نسبة الأفراد المدخنين إلى مجموع الأفراد في مكان ما.

في هذا الفصل نعالج فقط التقديرات البسيطة بانتهاج طريقة السحب بدون إعادة لأنها الأكثر استعمالا في البحوث والأكثر واقعية. نرمز بـ: $\bar{y}$ للمتوسط في العينة والذي يعتبر مقدرا لمتوسط المجتمع $\bar{Y}$، أي $\bar{y} = \hat{\bar{Y}}$، و نرمز بـ $N\bar{y}$ لمقدر المجموع الكلي Y في المجتمع،أي $N\bar{y} = \hat{Y}$ وأخيرا نرمز بـ $R$ للنسبة في العينة والتي نستعين بها لتقدير النسبة R في المجتمع ، حيث :

$$R = \frac{\bar{y}}{\bar{x}} = \sum^{n} y_i \Big/ \sum_{i=1}^{n} x_i$$

## 3-2 خواص المقدرات في المعاينة العشوائية البسيطة:

بعد تقديره لمختلف الخصائص والمعالم في المجتمع، يطرح الإحصائي أو الباحث على نفسه سؤالا جوهريا وهو: هل المقدرات المتحصل عليها من العينة تمثل تمثيلا صحيحا قريناتها في المجتمع؟

الإجابة على هذا السؤال تدفعنا إلى ذكر مجموعة من الخصائص التي أعدها Fisher RA (1925)، والتي بموجبها نقرر فيما إذا كان التقدير جيد أو لا. لقد طورت نظرية التقدير مجموعة من الخصائص، إن هي توفرت في تقدير ما فنقول عنه أنه جيد وجدير لأن يمثل معلمة معينة في المجتمع. في بحثنا هذا سوف نكتفي بذكر خاصيتان فقط وهما:

**1- الاتساق:**

يكون المقدر متسقا إذا كان احتمال أن يتجاوز خطؤه أي قيمة معطاة، يؤول إلى الصفر عندما تؤول العينة إلى ما لا نهاية (عينة كبيرة جدا).

يمكن البرهنة بسهولة على أن كلا المقدران $\bar{y}$ و $N\bar{y}$ متسقان لمتوسط المجتمع ومجموعه على الترتيب.

**2- التحيز:**

نقول عن مقياس إحصائي أنه تقدير غير متحيز أو تقدير أمين إن كان معدل قيم التقدير محسوبا فوق جميع العينات الممكنة ذات الحجم n مساويا تماما للقيمة الحقيقة في المجتمع.

متوسط العينة $\bar{y}$ تقدير غير متحيز لمتوسط المجتمع $\bar{Y}$ . للبرهان، يكفي أن نحسب قيمة $\bar{y}$ في جميع العينات الممكنة NCn، ثم نحسب معدل هذه القيم، يدل الرمز E على التوقع الرياضي أو عملية حساب المعدل.

$$E(\bar{y}) = \frac{\sum \bar{y}}{{}_N C_n} = \frac{\sum (y_1 + y_2 + y_3 + \dots y_n)}{n.[N!/n!(N-n)!]} \quad \dots\dots\dots(1\text{-}2)$$

من أجل حساب المجموع في البسط $\left(\sum_{i=1}^{{}_N C_n} \bar{y}\right)$، يجب علينا معرفة عدد المرات التي يتكرر فيها قيمة متغير ما $y_i$ في جميع العينات الممكنة.

وبما أن عدد وحدات المجتمع هو N وأن عدد العينات الممكنة هو NCn وبما أن كل عينة من هذه الأخيرة تحتوي على n وحدة فإن عدد الوحدات في جميع العينات الممكنة هو NCn.n. يكفي أن نقسم هذه القيمة على N لنتحصل على عدد المرات التي تتكرر فيها أي قيمة معينة وعليه:

$$\frac{{}_N C_n . n}{N} = \frac{N!}{n!(N-n)!} . \frac{n}{N}$$
$$= \frac{(N-1)!}{(n-1)!(N-n)!}$$

ومنه فإن

$$\sum_{i=1}^{{}_N C_n} (y_1 + y_2 + y_3 + \dots y_n) = \frac{(N-1)!}{(n-1)!(N-n)!} (y_1 + y_2 + y_3 + \dots y_N) \dots\dots\dots(2\text{-}2)$$

واستنادا إلى العلاقة 2−1 فإن هذا يعطي:

$$E(\overline{y}) = \frac{(N-1)!}{(n-1)!(N-n)!} \cdot \frac{n!(N-n)!}{n.N!}(y_1 + y_2 + y_3 + .....y_N)$$

$$= \frac{1}{N}\sum_{i-1}^{N} y_i = \overline{Y} \qquad (3\text{-}2)$$

و هو المطلوب

و كنتيجة لما سبق، يمكن أن نبين بسهولة أن $N\overline{y}$ هو مقدر غير منحاز لمجموع المجتمع Y أي $\hat{Y} = N\overline{y}$.

## 2−4 تباينات التقديرات في المعاينة العشوائية البسيطة:

لقد رأينا سابقا أن للتباين صيغتان وهما:

$$S^2 = \frac{\sum\left(y_i - \overline{Y}\right)^2}{N-1} \qquad (4\text{-}2)$$

$$\sigma^2 = \frac{\sum\left(y_i - \overline{Y}\right)^2}{N}. \qquad (5\text{-}2)$$

الأكيد أن كلتا الصيغتان تعطيان نتائج متساوية تقريبا، ولهذا سوف نحتفظ نحن في هذا العمل على الصيغة (2−5) وذلك من أجل جميع الفصول المقبلة.

يمكن أن نكتب كقاعدة أن تباين المتوسط $\overline{y}$ في العينة العشوائية البسيطة هو:

$$Var(\overline{y}) = E(\overline{y} - \overline{Y})^2 = \frac{S^2}{n}\left(\frac{N-n}{N}\right)$$

$$= \frac{S^2}{n}(1-f) \qquad (6\text{-}2)$$

من خلال العلاقة (2-6) يمكن استخلاص النتائج التالية:

أولا: الخطأ المعياري لـــ $\overline{y}$ هو:

$$\sigma_{\overline{y}} = \frac{S}{\sqrt{n}}\sqrt{\frac{N-n}{N}} = \frac{S}{\sqrt{n}}\sqrt{1-f} \qquad (7\text{-}2)$$

ثانيا: تباين $\hat{Y} = N\overline{y}$ كمقدر لمجموع المجتمع Y هو:

$$Var(\hat{Y}) = E(\hat{Y} - Y)^2 = \frac{N^2 S^2}{n}\left(\frac{N-n}{N}\right)$$
$$= \frac{N^2 . S^2}{n}(1-f) \qquad (8-2)$$

وخطأه المعياري هو:

$$\sigma_{\hat{Y}} = \frac{S}{\sqrt{n}}\sqrt{\frac{N-n}{N}} = \frac{S}{\sqrt{n}}\sqrt{1-f} \qquad (9-2)$$

إن الهدف من حساب العلاقتان الخاصتان بكل من الخطأ المعياري لمقدر متوسط المجتمع وكذا الخطأ المعياري لتقدير المجموع في المجتمع هو:

1- مقارنة الدقة الناتجة عن معاينة عشوائية مع تلك الناتجة عن طرق أخرى للمعاينة.

2- تقدير الدقة الفعلية التي بلغناها في مسح تم إنجازه.

3- تقدير حجم العينة الضروري لمسح نقوم بالتخطيط لإنجازه، وهذا هو موضوع بحثنا.

تحتوي العلاقات السابقة لتباينات التقديرات على تباين المجتمع $S^2$، لكننا في الواقع نجهل قيمته لذا يتم تعويضه بالتباين في العينة $s^2$ بحيث نتحصل على التقديرات التالية لتباين $\bar{y}$ و $\hat{Y}$ :

$$var(\bar{y}) = s_{\bar{y}}^2 = \frac{s^2}{n}\left(\frac{N-n}{N}\right) = \frac{s^2}{n}(1-f) \qquad (10-2)$$

$$var(\hat{Y}) = s_{\hat{Y}}^2 = \frac{N^2 s^2}{n}\left(\frac{N-n}{N}\right) = \frac{N^2 s^2}{n}(1-f) \qquad (11-2)$$

حيث يكون التباين في العينة:

$$s^2 = \frac{\sum^{n}(y_i - \bar{y})^2}{n-1} \qquad (12-2)$$

تقديرا غير منحازا للتباين في المجتمع:

$$S^2 = \frac{\sum^{N}(y_i - \bar{Y})^2}{N-1} \qquad (13-2)$$

ونأخذ في حالة الأخطاء المعيارية :

$$s_{\bar{y}} = \frac{s}{\sqrt{n}}\sqrt{1-f} \qquad (14-2)$$

$$s_{\hat{Y}} = \frac{Ns}{\sqrt{n}}\sqrt{1-f} \qquad (15\text{-}2)$$

## 2-5 عامل التصحيح لمجتمع منته (ت.م.م):

$$1-f = \frac{N-n}{N}$$

تحتوي صيغ التباين (2-6) والخطأ المعياري (2-7) على العامل $1-f$ و $\sqrt{1-f}$ على التوالي و الذي يسمى بمعامل تصحيح المجتمع المنتهي (ت.م.م). في المعاينة بدون إرجاع يظهر هذا المعامل كمصحح لأهم صيغ التباين وهي $S^2/n$ فإذا كان حجم العينة (n) ثابتا وطبقناه على مجتمعات ذات عدد وحدات N متزايد، فإن الكسر $f = n/N$ يؤول إلى الصفر، بالتالي يقترب المعامل $1-f$ من الواحد.

هذا يعني أنه لا يوجد لمعامل التصحيح لمجتمع منته أي تأثير على النتائج إذا كان المجتمع أكبر كثيرا عن العينة، في التطبيقات يمكن إهمال الـ ت.م.م إذا كان كسر المعاينة لا يتجاوز 5% ولأهداف عديدة حتى إذا كان عاليا[3] إلى10%، من أجل مجتمع غير منته، أين يكونN مجهول أو كبير جدا يختفي معامل التصحيح $1-f$ من صيغة التباين، ولما يكون السحب بإرجاع يكون المعامل $1-f$ مساويا للواحد[4].

## 2-6 تقدير النسبة:

يعتبر مقدر النسبة بين متغيرين اثنين من أهم خصائص المجتمع الإحصائي التي تهدف مسوح العينة إلى تقديرها، خاصة إذا اشتمل هذا المجتمع على وحده معاينة تحتوي بدورها على مجموعة من العناصر، كأن تكون مثلا وحدة المعاينة هي الأسرة، ويكون كل من الجنس والسن والمستوى التعليمي عناصر هذه الوحدة، فكثير من المسوح الإحصائية المنزلية تلجأ إلى حساب نسبة المصاريف المخصصة لشراء الملابس إلى مجموع مصاريف الأسر، أو تقدير متوسط عدد الأحذية التي يقتنيها كل فرد بالغ خلال مدة سنة مثلا.، كما توجد عمليا عدة تطبيقات لتقدير النسب، فمثلا نسبة النساء المصوتات لصالح مترشح معين في

[3] وليام كوكران. « تقنية المعاينة الإحصائية » ، ترجمة الدكتور 'أنيس كنجو'، عمادة شؤون المكتبات، جامعة الملك سعود، المملكة العربية السعودية، 1995.

[4] Leslie Kish. « Survey Sampling ». John Wiley and Son INC.USA, 1965.p44

الانتخابات إلى مجموع المواطنين المنتخبين، أو نسبة الأفراد الذين يفضلون لونا معينا إلى مجموع أفراد تلك المنطقة بهدف إصدار منتوج يتلاءم وميولاتهم، أو حتى نسبة المساحة المزروعة قمحا إلى مجموع المساحة المزروعة في بلد ما.

يعتمد المقدر النسبة على متغيرين، فإذا أخذنا المثال الأخير الخاص بمساحة الأرض المزروعة قمحا، يمكن أن نضع $\sum_{i=1}^{N} y_i$ مجموع المساحة المزروعة قمحا في المجتمع (البلد)، و $\sum_{i=1}^{N} x_i$ مجموع المساحة المزروعة في ذلك البلد، وتكون النسبة التي نريد تقديرها في المجتمع هي:

$$R = \frac{\sum_{i=1}^{N} y_i}{\sum_{i=1}^{N} x_i} \qquad (16\text{-}2)$$

وتقدير العينة الموافق هو:

$$\hat{R} = \frac{\sum^{n} y_i}{\sum^{n} x_i} = \frac{\bar{y}}{\bar{x}} \qquad (17\text{-}2)$$

لما تكون العينة المدروسة صغير يكون توزيع $\hat{R}$ متناظر كما يشكل هذا المقدر انحيازا طفيفا بالنسبة ل R. أما في العينات الكبيرة يقترب توزيع $\hat{R}$ من التوزيع الطبيعي إلا أن الانحياز يصبح مهمل[5].

**تباين مقدر النسبة ($\hat{R}$):**

إذا كنا بصدد دراسة عينة عشوائية بسيطة حجمها n، وإذا توفرت لدينا معلومات حول كل من المتغيران $y_i$ ، و $x_i$ يكون تباين $\hat{R}$ معطى بالعلاقة التالية:

$$Var(\hat{R}) = \frac{1-f}{n.\bar{x}^2} \cdot \frac{\sum_{i=1}^{N}(y_i - Rx_i)}{N-1} \qquad (18\text{-}2)$$

بحيث أن :

[5] كوكران، مرجع سبق ذكره.

$$R = \frac{\sum^{N} y_i}{\sum^{N} x_i} = \frac{\overline{Y}}{\overline{X}}$$

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

ويكون المقدر في العينة هو الآتي:

$$\operatorname{var}(\hat{R}) = \frac{\sum^{n} (y_i - \hat{R}x_i)}{n-1} \qquad (19\text{-}2)$$

$$\frac{\sum^{N} (y_i - Rx_i)}{N-1}$$

مقدار منحازا نوعا ما بمرتبة 1/n للصيغة التالية في المجتمع:

وعند تقدير الخطأ المعياري لـــ $\hat{R}$ يعطى هذا:

$$S(\hat{R}) = \frac{\sqrt{1-f}}{\sqrt{n.\overline{\overline{X}}}} \sqrt{\frac{\sum^{n} (y_i - \hat{R}x_i)}{n-1}} \qquad (20\text{-}2)$$

إذا كان $\overline{X}$ غير معلوما في المجتمع، فإننا نعوضه بـــ $\overline{x}$ كتقدير له في العينـــة، كمـــا يمكن أن نحصل على صيغة أخرى لـــ $S(\hat{R})$ بعد فتح الأقواس الموجودة في الطرف الأيمن من العلاقة، فنتحصل على ما يلي:

$$S(\hat{R}) = \frac{\sqrt{1-f}}{\sqrt{n.\overline{\overline{X}}}} \sqrt{\frac{\sum^{n} y_i^2 - 2\hat{R}.2y_i x_i + \hat{R}^2 \sum^{n} x_i^2}{n-1}}$$

## <u>2-7 معاينة النسب والنسب المئوية:</u>

في كثير من الأحيان، تقدم نتائج مسوحات احصائية بالمعاينة على شكل نســـب مئويـــة لوحدات مجتمع إحصائي ذات صفة أو خاصية مميزة، أو انتماء إلى صف معـــين. تعـــرف النسبة على أنها متوسط حسابي لمتغير منفصل (أي متقطع)، بحيث يأخذ هذا الأخير القيمـــة $y_i = 1$ إذا كانت الوحدة تنتمي إلى الصف و $y_i = 0$ إذا لم تكن كذلك. عادة ما يســـمى هـــذا المتغير بالثنائي. عمليا، تأتي العديد من التحاليل والعروض الإحصائية في شكل نسب، فمثلا نسب الشباب المدخنين للسجائر في منطقة ما، أو النسبة المئوية للأفراد ذوي الانتماء العرقي أو العقائدي المعين، أو النسبة المئوية للبطالة في مجتمع ما.

في مثل هذا النوع من المعاينة، يقسم المجتمع الإحصائي إلى صفين $C$ و $C^{/}$، بحيث تقع كل وحدة من المجتمع في أحدهما.

A=NP=Y هو عدد وحدات المجتمع التي تنتمي إلى الصف C، و $P\overline{Y}$ هي نسبهم في المجتمع. أما Q=1-P فهي نسبة الوحدات التي لا تنتمي إلى الصف C، و $\overline{A} = NQ$ هو عدد هذه الوحدات.

$$A = NP = Y \qquad (22\text{-}2)$$

$$p = \overline{y} = \frac{\sum^{n} y}{n} \qquad (23\text{-}2)$$

p هو تقدير العينة لـــ P الدي يمثل نسبة وحدات الصف C التي تنتمي إلى العينة.

$$a = np = \sum_{i=1}^{n} y_i \qquad (24\text{-}2)$$

هو عدد وحدات العينة التي تنتمي إلى الصف C.

$$\hat{A} = Np = N\frac{a}{n} \qquad (25\text{-}2)$$

هو تقدير العينة لـــ:A .

في معاينة عشوائية بسيطة للنسب، يمكننا بسهولة استنتاج تباينات تقديرات العينة، انطلاقا من النتائج المحصل على في الفقرة (5). إذ يكفينا فقط كتابة كل من $S^2$ و $s^2$ بدلالة P وp على التوالي، ثم نعوض النتائج المحصل عليها في تباينات تقديرات العينة.

لدينا:

$$Y = \sum_{i=1}^{N} y_i = A = NP$$

$$\sum_{i=1}^{N} y_i^2 = \sum_{i=1}^{N} y_i = NP \qquad (26\text{-}2)$$

$$\sum_{i=1}^{N} (y_i - \overline{Y})^2 = \sum^{N} y_i^2 - N\overline{Y}^2 = NP - NP^2 = NP(1-P) \qquad (27\text{-}2)$$

وبصورة مماثلة في العينة:

$$y = \sum_{i=1}^{n} y_i = a = np \qquad (28\text{-}2)$$

$$\sum_{i=1}^{N} (y_i - \overline{y})^2 = \sum^{N} y_i^2 - n\overline{y}^2 = np(1-p) \qquad (29\text{-}2)$$

وعليه نكتب كل من تباين المجتمع $S^2$ وتباين العينة $s^2$ كما يلي:

$$S^2 = \frac{\sum^{N}(y_i - \overline{Y})^2}{N-1} = \frac{N}{N-1}PQ \qquad (30\text{-}2)$$

$$s^2 = \frac{\sum^{N}(y_i - \overline{y})^2}{n-1} = \frac{n}{n-1}pq \qquad (31\text{-}2)$$

بتعويض هاتان الصيغتان (2–30) و(2–31) في تباينات التقديرات للمتوسط والمجموع في الفقرة (5) سوف نتحصل على تباينات تقديرات كل من P وA في معاينة النسب ونلاحظ أن:

تباين p هو:

$$Var(P) = \frac{S^2}{n}\left(\frac{N-n}{N}\right) = \frac{PQ}{n}\left(\frac{N-n}{N-1}\right) \qquad (32\text{-}2)$$

وهو نفسه تباين p في العينة، كما أنه يمكن البرهنة بسهولة على أن التقدير غير المتحيز لتباين p المحسوب في العينة هو:

$$\text{var}(p) = s^2 p = \frac{N-n}{N(N-1)}pq \qquad (33\text{-}2)$$

تباين المجموع المقدر لوحدات الصف C ، أي تباين $\hat{A} = Np$ يساوي:

$$Var(\hat{A}) = \frac{N^2 PQ}{n}\left(\frac{N-n}{N-1}\right) \qquad (34\text{-}2)$$

وتقديره غير المتحيز هو:

$$\text{var}(\hat{A}) = S^2 Np = N\left(\frac{N-n}{n-1}\right)pq \qquad (35\text{-}2)$$

التوزيع الاحتمالي لـــ p في معاينة عشوائية بسيطة هو توزيع ثنائي عندما يكون السحب بإرجاع .أما إذا كان السحب بدون إرجاع يكون التوزيع ما فوق الهندسي[6].

في حالة المعاينة مع الإعادة تتألف عملية سحب العينة من سلسلة من n تكرارات، واحتمال سحب وحدة من C يساوي في كل منها p. وفي هذه الحالة يمكننا بسهولة استنتاج دالة التكرارات الثنائية لعدد وحدات العينة التي تنتمي إلى الصف C. و حيث أن المتغير $y_i$ يأخذ فقط القيمة 1 أو 0 فإن توزيع p هو ثنائي وعليه فإن احتمال أن تحتوي العينة على $a$ من وحدات الصف C هو:

[6] 1965,Kish. مرجع سبق ذكره.

$$\Pr(a)=\frac{n!}{a!(n-a)!}p^{a}Q^{n-a}$$

أما في حالة معاينة بدون إعادة كما هو الحال في جميع الفصول المقبلة، يمكن إيجاد توزيع p ، وذلك بحساب عدد العينات الممكن تشكيلها والتي تتضمن $a$ وحدة بالضبط من C، و $a^{/}$ من $C^{/}$.

حيث:

$$a+a^{/}=n \qquad\qquad A+A^{/}=N$$

توجد ${}_{A}C_{a}$ طريقة مختلفة لاختيار $a$ وحدة من ألـــ A الموجودة في الصف C، و ${}_{A^{/}}C_{a^{/}}$ طريقة لاختيار $a^{/}$ وحدة من ألـــ $A^{/}$ الموجودة في الصف $C^{/}$من أجل تشكيل عينة ما، فإنه يمكن ضم أي اختيار من النوع الأول لأي اختيار من النوع الثاني، ويكون العدد الكلي من العينات من الشكل المطلوب هو:

$${}_{A^{/}}C_{a^{/}}\ {}_{A}C_{a}$$

ومنه فإن احتمال أن تكون العينة المسحوبة من النوع المطلوب هو عدد العينات من الشكل المطلوب على عدد العينات الممكنة:

$$\Pr(a,a'/A,A')=\frac{{}_{A}C_{a}\cdot{}_{A'}C_{a}}{{}_{N}C_{n}} \qquad (36\text{-}2)$$

هذا هو التوزيع التكراري لـــ$a$ ومنه يمكن استنتاج توزيع p الذي يدعى على بالتوزيع فوق الهندسي.

من خلال الصيغة التقديرية لتباين P ، يمكن أن نعرض أحد أشكال التقريب الطبيعي(Cochran1977) لحدود الثقة لـــ P وهو:

$$P\pm\left[t\sqrt{1-f}\sqrt{pq/n-1}+1/2n\right] \qquad (37\text{-}2)$$

حيث

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

$t$ : قيمة المتغير الطبيعي المعياري الموافق لاحتمال الثقة المرغوب.

½n : عامل التصحيح من أجل الاستمرار الدي يؤدي إلى تحسين دقة التقريب نوعا ما.

## 8-2 تحديد حجم العينة في أسلوب المعاينة العشوائية البسيطة:

### تمهيد:

لكي تكون أي دراسة إحصائية الأحسن، يجب أن يتم التخطيط لها بحذر، وللتخطيط الجيد عدة أوجه، إذ يجب أن يحدد المشكل بوضوح، كما يجب أن نجعله أكثر عمليا. من المهم جدا أن تسحب وحدات التجربة أو الوحدات المشاهدة من المجتمع المناسب، كما يجب أن تتبع الإجراءات بحذر وأن تستعمل الوسائل الضرورية من أجل الحصول على القياسات المطلوبة. أخيرا، من المهم جدا أن يكون للدراسة حجم عينة يتناسب وأهداف هذه الدراسة، إذ يجب أن يكون حجم العينة هذا كبير لدرجة أن تكون للنتائج دلالة علمية وإحصائية في نفس الوقت.

إن مسألة تحديد حجم العينة مسألة مهمة جدا نظرا لأسباب اقتصادية وإحصائية في نفس الوقت، فإذا أجريت دراسة بحجم عينة صغير أكثر من اللازم، فإن هذا سوف يعتبر مضيعة وهدرا للموارد، لأنها لا تملك القدرة على إنتاج معطيات ونتائج مفيدة. في حين أن دراسة بحجم عينة أكبر من اللازم تستعمل موارد أكثر من المفروض استعماله، ولا تضيف إلى معرفتنا شيئا آخر نجهله. في التجارب الكلينيكية مثلا، والتي تتضمن أفراد أو حيوانات كوحدات معاينة، يعتبر حجم العينة موضوعا محوريا نظرا لأسباب أخلاقية، لأن التجارب التي تستعين بحجم قليل من الوحدات سوف تعرض هذه الأخيرة إلى أخطار محتملة للعلاج بدون إحداث أي تقدم في المعرفة. أما في التجارب التي تعتمد على حجم أكبر للعينة فهي تعرض عدد معين من الوحدات الإضافية التي نحن في غنا عنها إلى ذلك الخطر المحتمل.

إن لحجم العينة عدة مداخل يمكن التقرب بها، فمثلا يمكن للواحد منا أن يحدد الدقة المرغوب فيها ويحسب حجم العينة الذي يحقق هذا الهدف. كما توجد النظرة البايبزية (Bayesian) عندما نريد تعظيم بعض دوال الهدف، وهي تدمج كلا من دقة التقدير والتكاليف مع بعض. كما نجد من بين أهم الطرق المستعملة في تحديد حجم العينة تلك المتعلقة باختبار الفرضيات والمتضمنة للقوة. سوف نركز في هده الفقرة على الطريقة الأولى المرتبطة بدقة التقديرات.

من أجل التبسيط نفترض أولا أن دقة النتائج تتمثل في الحصول على تباين متوسط $Var(\bar{y})$ معين. ثانيا؛ أن تكاليف الدراسة تتجسد في حجم العينة n.

يمكننا إذا محاولة الإجابة على السؤال التالي: ما هو حجم العينة الضروري للحصول على تباين متوسط $Var(\bar{y})$ معين؟

يأتي الجواب على هذا السؤال من الصيغة رقم (2–6).

$$Var(\bar{y}) = \frac{S^2}{n}\left(1 - \frac{n}{N}\right)$$

نضع :

$$n' = \frac{n}{1 - n/N} \qquad (38\text{-}2)$$

و بتعويض $n'$ في الصيغة (2–6) نحصل على

$$Var(\bar{y}) = \frac{S^2}{n'} \qquad (39\text{-}2)$$

و منه

$$n' = \frac{S^2}{Var(\bar{y})} \qquad (40\text{-}2)$$

من (38.2) نجد :

$$n = \frac{n'}{1 - n'/N}$$

هذا إذا اعتبرنا أن الدقة متمثلة في تباين المتوسط المرغوب الحصول عليه، أما إذا كانت الدقة المرجوة هي خطأ التقدير المسموح به وليكن مثلا الفرق بين القيمة المقدرة لمتوسط المجتمع $\bar{y}$ والقيمة الحقيقية المتوسطة في المجتمع $\bar{Y}$ فإن الأمر يختلف.

لنستعن بالمثال التوضيحي التالي: نفترض أنه تريد مؤسسة ما لصناعة ألبسة الأطفال عرض منتجاتها في سوق جديد بالنسبة لها وأنها تود معرفة نسبة الأطفال التي تتراوح أعمارهم بين سن الثامنة والثانية عشر، من أجل هذا الغرض، و من أجل تحقيق هدفها تلجأ هذه المؤسسة إلى خدمات مكتب دراسات وتطلب منه إجراء مسح إحصائي لصالحها. فما هو حجم العينة الضروري لإجراء هذه الدراسة؟

من أجل الإجابة على هذا السؤال، سوف يطرح العاملون في مكتب الدراسات على هذه المؤسسة مجموعة من الأسئلة التي سوف تساعدهم على أخذ قرار تحديد حجم عينة معين، و من أهم هذه الأسئلة: ما هي الدقة التي ترغب بها المؤسسة في معرفة نسبة هؤلاء الأطفال؟ تجيب المؤسسة بأنه لا مانع إذا كانت النسبة المئوية صحيحة في حدود ± 5 بالمائة،

وهذا يعني أنه إذا كانت النسبة المئوية للأطفال المقدرة هي 30% فإن قيمتها الحقيقية في المجتمع سوف تكون محصورة بين 25% و35%. في الواقع نحن لن نستطيع أن نضمن لهذه المؤسسة تلك النتائج لأننا نجهل قيمة المعلمة في المجتمع، ولو كنا على دراية بها لما كنا بحاجة لإجراء مثل هذه الدراسة. إن السبيل الوحيد للتأكد من تلك النتائج هو إجراء مسح شامل(recensement) يمس جميع وحدات المجتمع الإحصائي، الشيء الذي يعتبرا أمرا مستحيلا إذا كنا نتحدث عن مجتمع إحصائي كبير جدا وكانت ميزانية الدراسة محدودة. من المهم جدا أن نشير إلى أن اختيار المؤسسة لدقة مسموح بها قدرها ±5% كان اختيار شخصيا لأنه كان من الممكن أن تختار قيما أخرى للدقة وليكن مثلا 3% أو 4%. تأتي هذه الخيارات حسب أهداف الدراسة وأهمية القرارات التي سوف تتخذ انطلاقا من النتائج المحصل عليها من العينة. فكلما كان القرار هاما جدا وحساس كلما كانت الدقة المرجوة أكبر وبالتالي مجال الخطأ المسموح به أصغر. كما ترتكز عملية اختيار الدقة المرجوة على نوعية الدراسة، فاستطلاع الرأي (سبر الآراء) لأهداف سياسية كالانتخابات ليس كاستطلاع الرأي لأهداف أخرى ثقافية أو تجارية. في الحالة الأولى، تهدف السلطة إلى الحصول على نتائج دقيقة جدا لأنها تتعلق بمصير أمة بكاملها، أما في الحالة الثانية فإنه لا يهم كثيرا إذا كانت النتائج أقل دقة نوعا ما. بصفة عامة، تثبت التجربة أنه يتم تحديد مجال الخطأ من طرف صاحب الدراسة الذي يعتبر أن هذا المجال معقول بالنسبة له.

## 2-8-1 خطوات تحديد حجم العينة:

الخطوات الرئيسية[7] التي يتضمنها اختيار العينة هي كما يلي:

1- يجب أن يكون هناك تصور ما حول ما نتوقعه من العينية، ويمكن أن يكون هذا التصور بدلالة حدود الخطأ المرغوبة، كما في المثال السابق أو بدلالة قرار ما سنتخذه أو عمل سنقوم به عندما تصبح نتائج العينة معروفة. وتبقى مسؤولية تأطير التصور بصورة رئيسية على عاتق الأشخاص الذين يرغبون في استخدام نتائج المسح الإحصائي، علما بأنهم يحتاجون في الغالب للإرشاد كما توضع رغباتهم في شكل عددي.

[7] "كوكران". مرجع سبق ذكره

2- يجب إيجاد معادلة ما تربط n بالدقة المرغوبة، وستتغير المعادلة وفقا لمحتوى الدقة المرغوبة ووفقا لنوع المعاينة الذي سيجري تطبيقه .أحد فوائد المعاينة الاحتمالية أنها تمكننا من وضع مثل هذه المعادلة.

3- ستحوي هذه المعادلة بعض الخواص المجهولة للمجتمع على شكل معالم.

4- يحدث غالبا أن تنشر معلومات إحصائية تتعلق بأجزاء رئيسة معينة من المجتمع، وتوضح حدود الخطأ المرغوبة لكل جزء ونقوم بحسابات منفصلة لقيم n في كل جزء ثم نحسب الحجم الكلي n بالجمع.

5- تقاس عادة أكثر من مفردة أو خاصية في مسح احصائي يالمعاينة، ويكون عدد المفردات أحيانا كبيرا، وإذا وضعنا الدرجة المرغوبة من الدقة لكل مفردة، فقد تقود الحسابات إلى سلسلة من القيم المتعارضة لــ n، واحدة لكل مفردة، ويجب إيجاد طريقة ما للتوفير بين هذه القيم.

6- وأخيرا يجب تثمين القيمة المختارة لــ n لرؤية ما إذا كانت تتلاءم مع المصادر المتوافرة لأخذ العينة، وهذا يتطلب تقديرا للتكلفة والعمل والوقت والمواد المطلوبة للحصول على عينة الحجم المقترح، ويصبح باديا للعيان أنه لا بد من تخفيض كبير في قيمة n. ولا بد عندئذ من مواجهة قرار صعب، فإما أن نمضي بعينة ذات حجم أصغر بكثير، وبالتالي تخفض الدقة، أو أن نهجر المشروع حتى تتوافر لنا موارد أكثر. سنناقش في الفقرات القادمة بعض هذه المسائل بتفصيل أكبر.

## 2-8-2 حجم العينة في حالة بيانات من طبيعة مستمرة:

قبل الخوض في عملية تحديد حجم العينة تجدر الإشارة إلى التذكير بأننا نفترض أن البيانات المدروسة ($\overline{Y}$ في هذه الحالة) تتبع التوزيع الطبيعي. وأننا في معاينة عشوائية بسيطة. سوف نتطرق في الفصول القادمة إلى توزيعات احتمالية أكثر تعقيدا، وذلك في مختلف أساليب المعاينة العشوائية الأخرى.

من الصيغة (2-6) يكون تباين متوسط العينة:

$$Var(\overline{y}) = \frac{S^2}{n} \cdot \frac{N-n}{n}$$

وخطأه المعياري هو:

$$\sigma_{(\bar{y})} = \sqrt{\frac{N-n}{N}} \cdot \frac{S}{\sqrt{n}}$$

نرمز بـــ d لـــمجال الخطأ الممكن تقبله والسماح به عند تقدير متوسط العينة $\bar{y}$، ونحن نرضى بأن يكون الخطأ الحقيقي $(\bar{y}-\bar{Y})$ أكبر من الخطأ المقبول d، ولكن باحتمال صغير قدره α. يمكن صياغة ما ذكرناه على النحو التالي:

$$\Pr\left(\bar{y}-\bar{Y} \geq d\right) = \alpha \qquad (41\text{-}2)$$

يمكن حساب الدقة المرجوة كما يلي:

$$d = t.\sigma_{(\bar{y})} \qquad (42\text{-}2)$$

$$d = t\sqrt{\frac{N-n}{N}} \cdot \frac{S}{\sqrt{n}} \qquad (43\text{-}2)$$

حيث يمثل t قيمة المتغير الطبيعي المعياري الموافق لمستوى معنوية معين.

بعد تربيع طرفي المعادلة (2−43) نتحصل على قيمة n والتي تساوي:

$$n = \left(\frac{tS}{d}\right)^2 \Big/ 1 + \frac{1}{N}\left(\frac{ts}{d}\right)^2 \qquad (44\text{-}2)$$

في حالة ما إذا كان N كبيرا نسبيا أو غير معلوم يكون التقريب الأولى لحجم العينة هو:

$$n = n_0 = \left(\frac{ts}{d}\right)^2 \qquad (45\text{-}2)$$

لأن مقام العلاقة (2−44) سوف يؤول إلى (1). يكون $n_0$ حجما كافيا إذا كان $n_0/N$ صغيرا أو غير معلوم .أما كان $n_0/N$ كبيرا فأننا نحسب n كما في العلاقة (2−44) وهي:

$$n = \frac{n_0}{1 + \left(n_0/N\right)} \qquad (47\text{-}2)$$

يمكن كتابة الصيغة رقم (2−45) للتقريب الأولى لحجم العينة كما يلي:

$$n_0 = \left(\frac{tS}{d}\right)^2 = \frac{S^2}{V}$$

حيث يمثل $V = \left(\frac{d}{t}\right)^2 = \frac{S^2}{n_0}$ التباين المرغوب لمتوسط المجتمع.

غالبا ما نرغب في التحكم في الخطأ النسبي r في تقديرات مجموع أو متوسط المجتمع بدلا من الخطأ المطلق كما رأيناه سابقا. في هذه الحالة يمكن صياغة المسألة المطروحة كما يلي:

$$\Pr\left(\left|\frac{\bar{y}-\bar{Y}}{\bar{Y}}\right| \geq r\right) = \Pr\left(\left|\frac{N\bar{y}-N\bar{Y}}{N\bar{Y}}\right| \geq r\right) = \alpha$$

$$= \Pr\left(\left|\bar{y}-\bar{Y}\right| \geq r\bar{Y}\right) = \alpha \qquad (47\text{-}2)$$

في هذه الحالة يصبح مجال الخطأ المقبول هو:

$$d = r\bar{Y} = t\sigma_{\bar{y}}$$

وبالتالي:

$$r\bar{Y} = t\sqrt{\frac{N-n}{N}}.\frac{S}{\sqrt{n}} \qquad (48\text{-}2)$$

وبحل هذه العلاقة نجد أن:

$$n = \left(\frac{tS}{r\bar{Y}}\right)^2 \Big/ \left[1+\frac{1}{N}\left(\frac{tS}{r\bar{Y}}\right)^2\right] \qquad (49\text{-}2)$$

وكتقريب أولي نأخذ:

$$n'_0 = \left(\frac{tS}{r\bar{Y}}\right)^2 \qquad (50\text{-}2)$$

فإذا كانت النسبة$n_0'/\text{N}$ 1 مهملة أو صغيرة، نعتبر $n_0'$ تقريبا مرضيا لــ n ، أما إذا كانت النسبة$n_0'/\text{N}$ 1 كبيرة فإننا نحسب n كما في العلاقى (2–49) وهي:

$$n = \frac{n'_0}{1+\left(n'_0/N\right)} \qquad (51\text{-}2)$$

## <u>2–8–3 تحديد حجم العينة الضروري عند معاينة النسب:</u>

نفس طريقة التحليل المستعملة مع متوسط العينة، أي في حالة بيانات مستمرة يمكن تطبيقها لإيجاد عدد المشاهدات أو الوحدات الضروري عند تقدير النسبة P لوحدات صف معين في المجتمع من خلال العينة المسحوبة. إن الدور الذي يلعبه كل من الخطأ المقبول d

و مستوى المعنوية $\alpha$ في التأثير على n متشابه، يمثل الخطأ المقبول d أقصى انحراف يمكن قبوله بين P ومقدره p. أما α ، فهو يمثل احتمال أن يختلف p عن P بقيمة تفوق d.

يمكن صياغة ما سبق على النحو التالي:

$$\Pr\left(p-P\geq d\right)=\alpha \qquad (52\text{-}2)$$

ونفترض لأن p يتوزع طبيعيا، كما أن المعاينة عشوائية بسيطة . من العلاقة (2−32) في الفقرة (7) الخاصة بمعاينة النسب، كان تباين P هو :

$$V(P)=\frac{S^2}{n}\left(\frac{N-n}{N}\right)=\frac{PQ}{n}\left(\frac{N-n}{N-1}\right)$$

وخطأه المعياري:

$$\sigma_P=\sqrt{\frac{PQ}{n}}.\sqrt{\frac{N-n}{N-1}}$$

باتباع نفس الطريقة المستعملة عند تحديد حجم العينة في حالة البيانات المستمرة، يمكن أن نحصل على n الضروري بدلالة الدقة المرجوة والذي يساوي :

$$n=\frac{\dfrac{t^2PQ}{d^2}}{1-\dfrac{1}{N}\left(\dfrac{t^2PQ}{d^2}-1\right)} \qquad (52\text{-}2)$$

فإذا كان N كبيرا فإن التقريب الأولي لعدد وحدات العينة الضروري هو:

$$n_0=\frac{t^2PQ}{d^2}=\frac{PQ}{V} \qquad (53\text{-}2)$$

حيث التباين المرغوب لنسبة العينة هو :

$$V=\frac{PQ}{n_0}$$

عند التطبيق، إذا كان $n_0/N$ مهملا أو صغيرا، فإننا نعتبر $n_0$ تقريبا مرضيا لــــ n، أما إذا لم يتوافر ذلك، فمن الظاهر أن نستعمل العلاقة (2−52).

وعند صياغة هذه الأخيرة بدلالة $n_0$ فإننا نتحصل على:

$$n=\frac{n_0}{1+{(n_0-1)}/{N}}\approx\frac{n_0}{1+{n_0}/{N}} \qquad (54\text{-}2)$$

هذا هو تقدير النسبة P بدلالة p، أما إذا كنا بصدد تقدير العدد الكلي لوحدات الصف C في المجتمع A=NP، فبما أننا نرغب أحيانا في التحكم في الخطأ النسبي r بدلا من الخطأ المطلق وعليه:

$$\Pr\left(\frac{|Np-NP|}{NP}\geq r\right)=\Pr\left(|p-P|\geq rP\right)=\alpha \qquad (55\text{-}2)$$

من أجل هذا التجديد يكفي أن نعوض rP بدلا d في العلاقتين (2−25) و(2−53) حتى نتحصل على:

$$n=\frac{\dfrac{t^2PQ}{(rp)^2}}{1+\dfrac{1}{N}\left(\dfrac{t^2PQ}{(rP)^2}-1\right)} \qquad (56\text{-}2)$$

وكذلك :

$$n_0=\frac{t^2PQ}{r^2P^2}=\frac{t^2}{r^2}.\frac{Q}{P} \qquad (57\text{-}2)$$

إن صيغ حجم العينة هذه خاصة فقط بالمسوحات التي تدرس خاصية واحدة من المجتمع فقط، أما إذا كان المسح يدرس عدة خصائص فإن الأمر يختلف قليلا.

## 2−8−4 حجم العينة الضروري لمسح إحصائية بالمعاينة ذات الأهداف المتعددة:

تدرس أغلبية المسوحات الإحصائية بالمعاينة أكثر من خاصية واحدة في المجتمع، وفي هذه الحالة يحدد حجم العينة النهائي بعد تحديد مجال الخطأ لكل خاصية مدروسة على حدى، فنتحصل على عدة أحجام للعينة حسب عدد الخصائص المدروسة.

بعد الحصول على n لكل مفردة نختار حجم العينة النهائي، فقد يحدث أن تكون هذه أحجام $n_i$ $(i=1,2,3,...,k)$ و $k$ هو عدد الخصائص) متقاربة فيما بينها. ففي هذه الحالة نختار حجم العينة الأكبر شريطة أن تغطي ميزانية البحث تكاليفه. وقد يحدث أن يكون الاختلاف كبيرا بين أحجام العينة $n_i$ الشيء الذي يعتبر الأكثر شيوعا في العموم. فإنه ونظرا لاعتبارات اقتصادية وأخرى إحصائية (إذا كان n كبيرا تكون تكاليف الدراسة اكبر وتكون الدقة المتحصل عليها أكبر بكثير من الدقة المرجوة) سوف نتساهل في معايير الاختيار ونكتفي بحجم أصغر لــ n.

## 2-8-5 البحث عن أحسن القيم لمعالم المجتمع:

تحتوي أغلبية الصيغ الخاصة بتحديد حجم العينة على معلمة أو عدة معالم من المجتمع مثل النسبة P والتباين $S^2$، ويعتبر هذا الأخير من أهمها.

نحن نجهل عمليا قيمة $S^2$، لذا يجب علينا تقدير قيمته أو التخمين في قيمة نظن نحن أنها تقترب إلى حد ما من قيمته الحقيقية في المجتمع. لكن، على أي أساس أو معيار نرتكز لإعطاء قيمة للتباين $S^2$ بدلا من قيمة أخرى؟ أو ما هي مصادر تخميناتنا؟

1- يمكننا أن نلجأ إلى معيطات وبيانات لدراسة سابقة مشابهة للدراسة التي نحن بصدد إجراءها، أو يمكننا اللجوء إلى رأي خبير إحصائي في المعاينة، نظرا لمعرفة وقدرته على اكتشاف مظاهر وخصائص معاينات سابقة. يستطيع الإحصائي أن يطرح على الباحث المتخصص مجموعة من الأسئلة، وعلى ضوء إجاباته نستطيع نحن إنشاء نموذج لتوزيع المجتمع، وشكله وكذا حدوده المحتملة، ثم نستنتج من خلال كل هذه المعلومات قيمة التباين $S^2$.

2- إذا كنا نعرف تباين المتوسط $\text{var}(\overline{y})$ لمعاينة عشوائية بسيطة سابقة بحجم عينة $n'$، فإنه يمكننا استعمال الصيغة الموالية لتقدير قيمة $S^2$ وهي :

$$S'^2 = \frac{Var'(\overline{y}).n'}{(1-f)}$$

فإذا كانت المعاينة بأسلوب آخر غير أسلوب المعاينة العشوائية البسيطة، فإننا نلجأ إلى صيغة أثر التصميم Deff الذي سوف نتطرق إليه لاحقا.

3- بدلا من التخمين في قيمة $S^2$، نستطيع التخمين في التغاير $V = S/\overline{Y}$، لأن هذا الأخير لا يتغير كثيرا من فترة إلى أخرى مثلما يتغير التباين $S^2$، وعليه نستطيع بسهولة استغلال نتائج مسوحات سابقة بمتغيرات مشابهة لكي نحصل على تقدير لـ V وكذا لـ $\overline{Y}$ ثم نقدر قيمة $S = V.\overline{Y}$.

4- تعتبر إشكالية تقدير النسبة في المجتمع (P) سهلة نوعا ما، لأن قيمة التباين P(1-P) لا تتغير كثيرا إذا كانت النسبة P محصورة بين P=0,2 وP=0,8. في العموم يمكننا إجراء تخمين عقلاني لـ P، بالطبع إن الاختيار الأسهل هو لما يكون $\sigma = 0,25$ أي عندما تكون P=0,5.

5- من أجل رسم عينة كبيرة الحجم في الميدان دراسة مجهول, يمكننا إجراء معاينة الاستطلاعية (Pilot survey)، قبل الشروع في المعاينة الكبيرة قيد الدراسة. تهدف هذه الدراسة الاستطلاعية إلى الحصول على بعض المعلومات التي سوف تساعدنا على تشكيل دراستنا جيدا، لكن أغلب البحوث صغيرة الحجم، ولا تحتمل إجراء معاينة استطلاعية، فإذا كانت هذه الأخيرة صغيرة جدا فإن نتائجها لن تجدي نفعا ولا يمكننا الاعتماد عليها. وفي هذه الحالة نكتفي برأي الباحث الإحصائي وتخميناته بدلا من هدر المال.

إن أي تقدير أو تخمين لقيمة معمله ما من المجتمع سوف يخضع لا محالة إلى مسألة جد هامة، وهي عدم اليقين (Uncertainty)، لذا نسعى دائما إلى التقرب من القيمة الحقيقية في المجتمع، ولا يمكننا أبدا الجزم بأننا سوف نتوصل إليها بصفة مطلقة. توجد بعض القراءات التي تتحدث عن كيفية التقدير والتخمين في قيمة التباين $S^2$ في المسوحات الاستطلاعية، ونجد من أهمها تلك التي عرضها Taylor and Muller (1992) وكذلك Muller and Benignus (1995)[8] وThomas (1997)[9] الذين عرضوا بعض الطرق السهلة لحل هذا المشكل.

أخيرا، إذا جمعنا معطيات وبيانات الدراسة فإنه من المستحسن والمفيد مقارنة التباينات الحالية المشاهدة بتلك التي استعملناها في حساب حجم العينة، هذا لن يساعد على رسم (Design) الدراسة الحالية، وإنما يساعد على إنجاح دراسة ومعاينات مستقبلية.

## 2-8-6 حجم العينة والتكنولوجيا الحديثة للمعلوماتية:

لقد غزت التكنولوجيا الحديثة للمعلوماتية جميع الميادين، وحتى الميدان الإحصائي، بحيث يوجد تصاعد دائم في وفرة البرامج الإحصائية بصفة عامة والمتعلقة بتحديد حجم العينة بصفة خاصة. من بين هذه البرامج نجد Nquery Advisor (Elashoff, 200)، وكذا برنامج PASS2000 لمؤسسة NCSS [10] الموجود على مواقع الانترنيت.

[8] Muller, K.E and Benignus, V.A. " scientific power with statistical power ". Neuro-toxicology and Teratology , n°14, 211-219. 1995.
[9] Thomas, L ." retrospective Power Analysis". Conservation Biology, n°11, pp276-280. 1997.
[10] http://www.ncss.com

كما نجد كذلك برنامج Power and précision 2000 لمؤسسة Biostat المجود عالى الموقع (http://www.Powerand precision.com) الذي يعرض خدماته للباحثين في الميدان الكلينيكي- البيولوجي. تتضمن صفحات الواب (Web) على الانترنيت كذلك قوائم واضحة لحساب حجم العينة، إذ يكفي أن يدخل الواحد منا كلمة Sample size calculator على لوحة البحث حتى يتحصل على المئات من المواقع التي تعرض خدمات مجانية للحساب المباشر لحجم العينة الضروري للحصول على دقة معينة. في العموم، تتشابه هذه الصفحات من حيث مضمونها، لذا سوف نكتفي في هذه الفقرة بعرض حاسب واحد توفره مؤسسة Creative research system على موقعها في الانترنيت المعروض في الملحق رقم 01.

يعرض هذا الحاسب بعض المصطلحات والمفاهيم الإحصائية الأساسية التي تساعد غير المتمرسين على الفهم، من بين هذه المصطلحات نجد: مستوى المعنوية ومجال الثقة المسموح به، وكذلك حجم العينة وحجم المجتمع وأخيرا النسبة. بعد إدخال المعلومات الضرورية على الصفحة، وبعد نقر زر التشغيل أو الحساب يعطيها هذا الجانب حجم العينة الضروري لتحقيق مجال ثقة محدد أو مجال الثقة الضروري والمناسب مع تثبيت حجم العينة، في الحالة الأولى نختار مستوى المعنوية الذي يناسبنا (99% أو 95%) ثم ندخل قيمة مجال الثقة المقبول وكذا حجم المجتمع إذا كان معلوما، ويمكن إهمال هذا الأخير إذا كان كبيرا جدا أو مجهولا. ثم ننقر زر الحساب (Calculate) لكي نتحصل على حجم العينة الضروري.

في الحالة الثانية، نتبع نفس خطوات المرحلة الأولى إلا أننا نثبت كل من قيمة حجم العينة وقيمة النسبة حتى نتحصل على مجال الثقة الذي يحقق هذه الشروط.

في بعض الحالات تكون وحدات الاختيار في المجتمع الإحصائي غير متجانسة فيما بينها وفقا للخصائص المدروسة، وعليه فإنه يتعذر علينا استعمال طرق تحديد حجم العينة التي عرضناها في أسلوب المعاينة العشوائية البسيطة، حيث تفرض علينا الضرورة استعمال أسلوب آخر من أساليب المعاينة قصد التقليل من أخطاء التقدير و التحيز. و من أهم الأساليب المستعملة في حالة عدم تجانس وحدات المجتمع هو أسلوب المعاينة العشوائية الطبقية.

# الفصل الثالث:
# تحديد حجم العينة في أسلوب المعاينة العشوائية الطبقية

يعرض هذا الفصل أسلوب المعاينة العشوائية الطبقية و كيفية تحديد حجم العينة الضروري عند استعمال هذا الأسلوب. نعرض في الفقرة الأولى أهم أسباب الانتقال من أسلوب المعاينة العشوائية البسيطة إلى أسلوب المعاينة العشوائية الطبقية. كما نعرض بعض التعاريف و الرموز التي تساعدنا على فهم مقدرات بعض معالم المجتمع و تبايناتها في كل من الفقرات الثانية و الثالثة و الرابعة. الفقرة الخامسة مخصصة للحديث عن أسلوب المعاينة العشوائية الطبقية في حالة النسب المئوية. أما الفقرة السادسة فتعرض كيفية تحديد حجم العينة وأهم المعايير الواجب مراعاتها في ذلك بهدف بلوغ دقة معينة يحددها الباحث أو تدنية تكاليف الدراسة أو بهدف تحقيق أغراض أخرى. كما نتحدث في الفقرة الأخيرة عن مسألة تحديد حجم العينة عند المقارنة بين متوسطات ميادين دراسة.

**3-1 المعاينة العشوائية الطبقية:**

المعاينة العشوائية الطبقية أسلوب يتمثل في تقسيم المجتمع الإحصائي إلى عدة مجتمعات جزئية متجانسة تسمى طبقات. تحتوي كل طبقة على مجموعة من الوحدات الإحصائية المرتبطة فيما بينها بخاصية أو عدة خصائص من المجتمع، والتي بدورها (أي الخصائص) تربط بالمتغير قيد الدراسة. بعد إجراء عملية التقسيم، تسحب من كل طبقة عينة عشوائية بسيطة، ثم نجمع كل العينات الجزئية حتى نتحصل على حجم العينة الكلي n الضروري لإنجاز الدراسة.

توجد عدة أسباب تدفنا إلى استعمال أسلوب المعاينة العشوائية الطبقية بدلا من المعاينة العشوائية البسيطة. أولا، إذا كانت وحدات المجتمع غير متجانسة فيما بينها فمن الضروري تصنيفها وتجيعها في طبقات أكثر تجانسا بهدف الحصول على صورة أوضح وأحسن عن المجتمع، فمثلا إذا قمنا بدراسة سلوك المستهلكين في منطقة ما، فإنه من الضروري تقسيمهم إلى طبقات حسب الجنس (ذكر أو أنثى)، أو حسب السن (ليكن مثلا من 20 إلى 25 سنة، ومن 25 إلى 30 سنة، ومن 30 إلى 35 سنة...إلخ)، أو حسب الدخل (مثلا: أقل من 10000دج، ومن 10000دج إلى 15000دج، وأكثر من 15000دج). ثانيا، نستعين كذلك بالمعاينة العشوائية الطبقية إذا كنا نسعى إلى الحصول عل معلومات دقيقة حول فئات معينة في المجتمع، فمثلا إذا قمنا بدراسة تهدف إلى تقدير معدل النمو السكاني للبلد، فإنه بالإضافة إلى كوننا نريد التوصل على المعدل الإجمالي فإنه يهمنا كذلك الحصول على معدلات النمو الخاصة بكل

منطقة جغرافية معينة. ثالثا، يمكن أن تفرض المعاينة العشوائية الطبقية نفسها نظرا لأسباب إدارية، فمثلا يمكن أن تملك أحد المصالح الإدارية مكتبا في كل منطقة من مناطق البلد كمصلحة الضرائب مثلا. فمن السهل اعتبار كل منطقة كطبقة.

## 3-2 رموز:

سوف نستعمل في هذا الفصل نفس الرموز السابقة، ولكن سوف نضيف الرمز h الذي يدل على الطبقة، والمؤشر i الذي يدل على الوحدة ضمن الطبقة، و $k$ الذي يرمز لعدد الطبقات.

### 3-2-1 بعض القيم في الطبقة:

عدد وحدات الطبقة h : $N_h \qquad (h = 1,2,3,...,k)$

بحيث:

$$N_1 + N_2 + ... + N_h + ... + N_k = N$$

قيمة الوحدة i ضمن الطبقة h : $y_{hi}$

مجموع قيم وحدات الطبقة h : $Y_h = \sum_{i=1}^{Nh} y_{hi}$

متوسط قيم وحدات الطبقة h : $\overline{Y}_h = \dfrac{Y_h}{N_h}$

نسبة مجموعين أو متوسطين : $R_h = \dfrac{Y_h}{X_h} = \dfrac{\overline{Y}_h}{\overline{X}_h}$

التباين:

$$S_h^2 = \frac{\sum_{i=1}^{N_h} (y_{hi} - \overline{y_h})^2}{N_h - 1}$$

### 3-2-2 بعض القيم في العينة الجزئية:

عدد وحدات العينة من الطبقة h :.................... $n_h \quad (h = 1,2,3,...,k)$

حيث أن : $n = n_1 + n_2 + ... + n_h + ... + n_k$

قيمة الوحدة i ضمن العينة $n_h$ : $y_{hi} \quad (i = 1,2,3,...,n_h)$

متوسط العينة:

$$\bar{y}_h = \frac{\sum_{i=1}^{n_h} y_{hi}}{n_h}$$

التباين في العينة، وهو تقدير غير منحاز للتباين في الطبقة h كما سوف نراه لاحقا:

$$S_h^2 = \frac{\sum^{n_h}(y_{hi} - \bar{y}_h)^2}{n_h - 1}$$

## 3-3 إجراءات سحب عينة عشوائية طبقية:

بعد تحديد الطبقات في المجتمع، وكذا حجم العينة الذي سوف نسحبه من كل طبقة، تتم عملية السحب بنفس الطريقة المستعملة في أسلوب المعاينة العشوائية البسيطة، حيث أننا نعتبر كل مجتمع جزئي (الطبقة) كمجتمع قائم بذاته، وبالتالي نستعمل نفس الطرق التي رأيناها في الفصل السابق لسحب عينة عشوائية بسيطة، كاستعمال مبدأ اللوطو أو استعمال جدول الأرقام العشوائية أو حتى إدخال الوحدات الإحصائية للطبقة في الحاسب...الخ.
تكون عملية السحب في أي طبقة مستقلة تماما عن عملية السحب في الطبقات الأخرى. إذا قمنا بسحب كل العينات الجزئية الـــ k التي سوف يشكل مجموعها العينة الكلية n ، نقول أننا أجرينا معاينة عشوائية طبقية بسيطة. تصمم العينة أحيانا من الطبقات بحيث تكون نسبة الاختيار متساوية في جميع الطبقات (أي أنه $f_h = n_h / N_h = f^{/}$، و $f^{/}$قيمة ثابتة مهما كان $(h = 1,2,3,..........,k)$. في هذه الحالة، نسمي المعاينة العشوائية الطبقية بالنسبية. كما توجد حالات أخرى للمعاينة أين تتغير فيها نسبة الاختيار تغيرا طفيفا أو كبيرا من طبقة إلى أخرى.

## 4-3 خصائص المقدرات وتبايناتها:

سوف نستعين في هذه الفقرة بخصائص المقدرات التي رأيناها في أسلوب المعاينة العشوائية البسيطة (فصل 2) وذلك لعرض خصائص مقدرات كل من المتوسط $\bar{Y}$ والمجموع Y في حالة المعاينة العشوائية الطبقية، كما سوف نحسب تباين هاذان المقدران.

## 3-4-1 تقدير متوسط المجتمع ($\overline{Y}$):

في معاينة عشوائية طبقة، تكتب صيغة متوسط وحدات المجتمع ككل على الشكل التالي:

$$\overline{Y} = \frac{Y}{N} = \frac{\sum_{i=1}^{N} y_i}{N} = \sum_{h=1}^{k} \frac{N_h . \overline{Y_h}}{\sum_{h=1}^{k} N_h} \qquad (1\text{-}4)$$

حيث أن $\overline{Y}_h$ يمثل متوسط وحدات الطبقة h. يمكن أن نلاحظ أن المتوسط $\overline{Y}$ هو متوسط مرجح لمتوسطات الطبقات، حيث أن الترجيح هو $N_h$، عدد الوحدات في الطبقة $h$.

اعتبرنا أن عملية السحب في كل طبقة كانت عشوائية بسيطة وعليه، فنحن نعلم من خلال الفصل 2 أن:

$$\overline{y}_h = \frac{\sum_{i=1}^{n_h} y_{hi}}{n_h}$$

مقدر غير متحيز للمتوسط الحقيقي في الطبقة h ($\overline{Y}_h$) (الفصل2، العلاقة 3.2). إن هذه الخاصية تدفعنا إلى استخلاص النظرية التالية:

إذا كان المتوسط $\overline{y}_h$ في العينة الجزئية تقدير غير متحيز للمتوسط $\overline{Y}_h$ في الطبقة h، فهذا يعني أن المقدر:

$$\overline{y}_{st} = \frac{\sum^{k} N_h \overline{y}_h}{\sum^{k} N_h} = \sum^{k} W_h \overline{y}_h \qquad (2\text{-}4)$$

هو كذلك مقدر غير متحيز للمتوسط الحقيقي في المجتمع $\overline{Y}$.

تأتي البرهنة على هذه النظرية من واقع أن المقدار $N_h \overline{y}_h$ تقدير غير متحيز لـــ $N_h \overline{Y}_h$ وأن التوقع الرياضي للمجموع $\sum N_h \overline{y}_h$ هو نفسه مجموع التوقعات الرياضية للحدود $N_h \overline{y}_h$.

**نظرية**:

إذا كان $\overline{y}_{st}$ تقدير غير متحيز لـــ $\overline{Y}$ فإن تباينه هو:

$$Var(\bar{y}_{st}) = \sum_{h=1}^{k} W_h^2 Var(\bar{y}_h)... \qquad (3\text{-}4)$$

نعلم من خلال الفصل السابق(علاقة 6.2) أن:

$$Var(\bar{y}_h) = \frac{S_h^2}{n_h}(1 - f_h)$$

وعليه تصبح العلاقة (4−3) كما يلي:

$$Var(\bar{y}_{st}) = \sum_{h=1}^{k} W_h^2 \frac{S_h^2}{n_h}(1 - f_h). \qquad (5\text{-}4)$$

أو:

$$Var(\bar{y}_{st}) = \frac{1}{N^2}\sum_{h=1}^{k} N_h (N_h - n_h)\frac{S_h^2}{n_h}. \qquad (6\text{-}4)$$

والتقدير غير المتحيز للتباين $Var(\bar{y}_{st})$ هو:

$$\text{var}(\bar{y}_{st}) = \frac{1}{N^2}\sum N_h (N_h - n_h)\frac{s_h^2}{n_h}. \qquad (7\text{-}4)$$

إذا كانت المعاينة العشوائية تناسبية، فإن صيغة تباين مقدر المتوسط ($\bar{y}_{st}$) تصبح كما يلي:

$$Var(\bar{y}_{st}) = \sum W_h^2 \frac{S_h^2}{n_h}(1 - f_h) = .\sum W_h S_h^2 \frac{(1 - f)}{n}. \qquad (8\text{-}4)$$

### 3−4−2 تقدير مجموع المجتمع Y:

في معاينة عشوائية طبقية يكون مجموع المجتمع Y مساويا لــ :

$$Y = \sum_{h=1}^{k} N_h \bar{Y}_h \qquad (9\text{-}4)$$

وكنتيجة لما سبق، فإن المقدار $\hat{Y} = \sum N_h \bar{y}_h = N\bar{y}_{st}$ هو تقدير غير متحيز لهذا المجموع.

**نظرية :**

تباين تقدير مجموع المجتمع هو:

$$Var(\hat{Y}_{st}) = \sum N_h (N_h - n_h)\frac{S_h^2}{n_h} \qquad (10\text{-}4)$$

## 5-3 المعاينة العشوائية الطبقية في حالة النسب:

في المعاينة العشوائية الطبقية للنسب، سوف نحاول تقدير النسبة P للوحدات التي تنتمي إلى الصف C في المجتمع بعدها سوف نتطرق إلى صيغة التباين لهذا التقدير.

### 1-5-3 تقدير نسبة وحدات الصف C:

نرمز بـــ: $P = \frac{A}{N}$ إلى نسبة الوحدات التي تنتمي إلى الصف C في كل المجتمع، وبـــ:

$P_h = \frac{A_h}{N_h}$ إلى نفس النسبة ولكن ضمن الطبقة $h$. $p_h = \frac{a_h}{n_h}$ ، هو نسبة وحدات الصف C ضمن العينة المأخوذة من هذه الطبقة $h$.

$P_{st}$ هو تقدير النسبة في كامل المجتمع الموافق لمعاينة عشوائية طبقية وهو:

$$P_{st} = \frac{\sum_{h=1}^{k} N_h . p_h}{N} \qquad (1\text{-}5)$$

نظرية:

إذا كان $P_{st}$ تقدير غير متحيز للنسبة P في المجتمع فإن تباينه يعطى بالعلاقة :

$$Var(P_{st}) = \frac{1}{N^2} \sum \frac{N_h^2 (N_n - n_n)}{N_h - 1} . \frac{P_h Q_h}{n_n} \qquad (2\text{-}5)$$

ويمكن البرهنة على ذلك بسهولة كوننا بينا في الفقرة (4) بالعلاقة (4−6) أن:

$$Var(\bar{y}_{st}) = \frac{1}{N^2} \sum N_h (N_h - n_h) \frac{S_h^2}{n_h}$$

ومن العلاقة (2−30) في الفصل 2 فان:

$$S_h^2 = \frac{N_h}{N_h - 1} P_h . Q_n$$

إذا كانت المعاينة العشوائية الطبقية بحصص متناسبة (تناسبية)، فإن صيغة تباين مقدر النسبة تصبح:

$$VarP_{st} = \frac{1}{N^2} \sum \frac{N_h^2}{N_h - 1} P_h Q_h \left( \frac{N_h - n_h}{n_h} \right) \qquad (3\text{-}5)$$

$$= \frac{1}{N^2} \sum \frac{N_h^2}{N_h - 1} P_h Q_h \left( \frac{N - n}{n} \right)$$

$$= \frac{N-n}{N^2}.\frac{1}{n}\sum\frac{N_h^2}{N_h-1}P_h.Q_h \qquad (4\text{-}5)$$

من أجل التبسيط ، نفترض أن $\frac{1}{N_h-1} \approx \frac{1}{N_h}$ ،و علما أن $W_h = \frac{N_h}{N}$ ، فإن العلاقة (4-4) تصبح:

$$VarP_{st} = \frac{1-f}{n}\sum W_h P_h Q_h \qquad (5\text{-}5)$$

### 3-6 تحديد حجم العينة في المعاينة العشوائية الطبقية.

### 3-6-1 حجم العينة الضروري للحصول على دقة معينة:

لقد رأينا في الفصل السابق (الفقرة 1) مراحل تحديد حجم العينة الضروري لبلوغ دقة معينة والتي سميناها مجال الخطأ المقبول d (فصل 2 العلاقة 2-43). سوف نستعمل هنا نفس المنهجية السابقة في تحديد حجم العينة في المعاينة العشوائية الطبقية.

من العلاقة (4.3) يكون تباين متوسط العينة هو:

$$Var(\bar{y}_{st}) = \sum_{h=1}^{k} W_h^2 \frac{S_n^2}{n_h}(1-f_h)$$

يمكن الحصول على صيغة أخرى لهذا التباين، وذلك باعتبار $N_h = W_h N$ و $n_h = w_h n$ وهي:

$$Var(\bar{y}_{st}) = \sum^{k} \frac{W_h S_h^2}{n_h} - \sum_{h=1}^{k}\frac{W_h^2 S_h^2}{N_h}$$

$$Var(\bar{y}_{st}) = \frac{1}{n}\sum_{h=1}^{k}\frac{W_h^2 S_h^2}{w_h} - \frac{1}{N}\sum W_h S_h^2 \qquad (1\text{-}6)$$

وتكتب الدقة المرجوة $d$ باحتمال صغير $\alpha$ كما يلي في معاينة عشوائية طبقية:

$$d = t\sigma_{\bar{y}_{st}} \qquad (2\text{-}6)$$

حيث تمثل $t$ قيمة المتغير الطبيعي المعياري الموافق لمستوى معنوية معين α،

و $\sigma_{\bar{y}_{st}}$ هو الانحراف المعياري للمتوسط في المعاينة العشوائية الطبقية.

من العلاقة (6-1) و (6-2) نستخرج الصيغة العامة لــ n و هي:

$$n = \frac{\sum \frac{W_h^2 . S_h^2}{w_h}}{(d/t)^2 + \frac{1}{N}\sum W_h S_h^2} \qquad (3\text{-}6)$$

إذا كان N كبيرا نسبيا، أو غير معلوم فإن التقريب الأولى لحجم العينة هو من الشكل

$$n_0 = \frac{1}{\left(d/t\right)^2} \sum \frac{W_h^2 S_h^2}{w_h} \qquad (4\text{-}6)$$

أما إذا كان N معلوم، وغير كبير جدا مقارنة بحجم العينة الأولى، أي أن n0/N غير مهمل، فيمكن حساب n كما في العلاقة (6−3) و الصيغة المشابهة لها هي

$$n = \frac{n_0}{1 + \frac{1}{N\left(d/t\right)^2} \sum_{h=1}^{k} W_h S_h^2} \qquad (5\text{-}6)$$

إن صيغ حجم العينة التي توصلنا إليها حتى الآن، تخص بصفة عامة المعاينة العشوائية الطبقية مهما كان نوعها، بحصص نسبية أو بحصص غير متناسبة.

توجد صيغ أخرى لــ n في عدة حالات أخرى ونذكر منها تلك المتعلقة بمعاينة بحصص متناسبة أين يكون $\frac{N_h}{N} = \frac{n_h}{n}$ ،أي أن $W_h = w_h$.

في هذه الحالة، نعوض $W_h$ بدلا من $w_h$ في العلاقة (6−3)، فنتحصل على الصيغة التالية لــ n:

$$n = \frac{\sum W_h S_h^2}{\left(d/t\right)^2 + \frac{1}{N} \sum W_h S_h^2} \qquad (6\text{-}6)$$

فإذا وضعنا:

$$n_0 = \frac{\sum W_h S_h^2}{\left(d/t\right)^2} \qquad (7\text{-}6)$$

تصبح العلاقة (6−6) كما يلي:

$$n = \frac{n_0}{1 + \frac{n_0}{N}} \qquad (8\text{-}6)$$

وهذا هو حجم العينة الضروري في معاينة عشوائية طبقية ذات حصص متناسبة.

بنفس الكيفية السابقة، يكون حجم العينة الضروري عند تقدير مجموع المجتمع $\hat{Y}_{st}$ كما يلي:

$$n = \frac{\sum \frac{N_h^2 S_h^2}{w_h}}{Var\left(\hat{Y}_{st}\right) + \sum N_h S_h^2} \qquad (9\text{-}6)$$

حيث $\left(d/t\right)^2 = Var\left(\hat{Y}_{st}\right)$ هي صيغة تباين مقدر مجموع المجتمع كما في العلاقة(4−10).

في حالة ما إذا كانت المعاينة العشوائية الطبقية بحصص متناسبة، نتحصل على n كما يلي:

$$n = \frac{n_0}{1 + \frac{n_0}{N}} \qquad (10\text{-}6)$$

حيث أن:

$$n_0 = \frac{N}{\left(d/t\right)^2} \sum N_h S_h^2 \qquad (11\text{-}6)$$

عمليا، ونظرا لعدم معرفتنا لقيمة التباين في الطبقة $S_h^2$، فإنه يتم تعويضه بـــ $s_h^2$ في العينة أو تقديره من خلال إجراء مسح استطلاعي، أو عن طريق بيانات ودراسات سابقة بنفس المتغيرات المدروسة حاليا أو حتى نلجأ إلى رأي أخصائي في المعاينة.

3−6−2 حجم العينة الأمثل في المعاينة العشوائية بحصص غير متناسبة أو التقسيم الأمثل لحجم العينة ضمن الطبقات:

3−6−2−1 التقسيم الأمثل للعينة ضمن الطبقات (المحاصة المثلى):

لقد رأينا سابقا كيفية تحديد حجم العينة الضروري لبلوغ دقة معينة، وذلك دون أخذ تكاليف الدراسة بعين الاعتبار، الشيء الذي قد يعتبر غير واقعي، إذ يمكن أن نتحصل على حجم عينة معين بحيث لا تستطيع الميزانية المخصصة لبحثنا تغطية تكاليفه، هنا نجد نفسنا أمام أحد الخيارات؛ الأول، أن نتخلى عن القيام بهذه الدراسة. الثاني، التريث حتى نتحصل على أموال إضافية تكفي لإنجاز البحث. في كلتا الحالتين، يمكن أن يكون عامل الزمن جد هام وضروري في العديد من الحالات.

تفاديا لهذه المشاكل سوف نعرض في هذه الفقرة كيفية تحديد حجم العينة الأمثل، والضروري للحصول على أدنى قيمة لتباين المقدر عندما تكون التكاليف محدودة ،أو أدنى تكلفة عندما تكون الدقة محددة عند قيمة معينة.

من أجل إدخال تكاليف الدراسة بعين الاعتبار، يجب علينا تحديد الاختلافات الموجودة في تكاليف معاينة عدة طبقات. سوف نفترض في هذه الفقرة (من أجل التبسيط فقط) أن تكلفة معاينة أي وحدة من وحدات الطبقة h هي متساوية، ولكنها تختلف من طبقة إلى أخرى.

بالإضافة إلى ذلك، توجد تكلفة ثابتة لا تتغير بتغير عدد وحدات العينة المسحوبة في كل طبقة. يمكن صياغة هذه الافتراضات في شكل دالة خطية بسيطة على النحو التالي:

$$C = C_0 + \sum_{h=1}^{k} n_h c_h \qquad (12\text{-}6)$$

حيث تمثل C التكلفة الإجمالية، $c_h$ تكلفة وحدة واحدة مسحوبة من العينة ضمن الطبقة h، و C0 التكلفة الثابتة.

سوف نرى في الفصول المقبلة أشكالا أخرى لدالة التكاليف أكثر تعقيدا ونكتفي في هذه الفقرة بعرض العلاقة الخطية البسيطة لدالة التكاليف والتي سوف تساعدنا على استخلاص النتائج إذا كنا في حالات أخرى أكثر تعقيدا.

يمكن أن نبرهن بسهولة، أن التقسيم الأمثل للعينة على الطبقات أو الحجم الأمثل للعينة الجزئية في الطبقة h الذي يحقق :(1) أدنى تكلفة لما يكون التباين ثابت ،أو (2) أدنى تباين لما تكون التكلفة محدودة هو من الشكل:

$$n_h = n.\frac{N_h.S_h / \sqrt{C_h}}{\sum\left(N_h S_h / \sqrt{C_h}\right)} \qquad (13\text{-}6)$$

في الحالة الأولى ، يكون حجم العينة الإجمالي الأمثل كما يلي:

$$n = \frac{C - C_0}{\sum^{k} N_h S_h \sqrt{C_h}} . \sum^{k} \frac{N_h S_h}{\sqrt{C_h}} \qquad (14\text{-}6)$$

في الحالة الثانية:

$$n = \frac{\sum^{k}\left(N_h S_h \sqrt{C_h}\right) \sum\left(N_h S_h / \sqrt{C_h}\right)}{N^2 Var\left(\bar{y}_{st}\right) + \sum N_h S_h^2} \qquad (15\text{-}6)$$

يكون التباين الأمثل للمتوسط في كلا الحالتان مساويا لـــ :

$$Var\left(\bar{y}_{st}\right)_{\min} = \frac{1}{nN^2} \sum^{k} N_h S_h \sqrt{C_h} \sum^{k} \frac{N_h S_h}{\sqrt{C_h}} - \frac{\sum N_h S_h^2}{N^2} \qquad (16\text{-}6)$$

برهان:

تباين المتوسط $\bar{y}_{st}$ في معاينة عشوائية طبقية هو:

$$Var\left(\bar{y}_{st}\right) = \sum_{h=1}^{k} W_h^2 \frac{S_h^2}{n_h} (1 - f_h).$$

أو

$$\mathrm{var}\left(\bar{y}_{st}\right) = \frac{1}{N^2} \sum N_h^2 S_h^2 \left(\frac{1}{n_h} - \frac{1}{N_h}\right). \qquad (17\text{-}6)$$

تكلفة المعاينة هي:

$$C = C_0 + \sum_{h=1}^{k} n_h c_h$$

يمكن تحديد قيم $n_h$ التي تدنى التباين تحت قيد التكاليف المثبتة أو تدنية التكاليف علما أن التباين محدد عن طريق استعمال متراجحة "كوشي –شوارتز" (Stuart, 1945) أو استعمال مضاعف لا غرانج (M.H,Hansen .W.N Hurwitz and W.G,Madow (1993)).كلتا الطريقتين تؤديان إلى نفس النتائج. سوف نكتفي هنا بعرض طريقة مضاعف لاغرانج.

1- من أجل تحديد قيم $n_h$ التي تدني التباين لما تكون التكلفة الإجمالية مثبتة، نشكل دالة لاغرانج F:

$$F = \frac{1}{N^2}\sum^{k} N_h^2 S_h^2 \left(\frac{1}{n_h} - \frac{1}{N_h}\right) + \lambda\left(\sum C_h n_h - C'\right)$$

حيث:

C’ = C-C0

$$\frac{\partial f}{\partial n_h} = -\frac{1}{N^2}.\frac{N_h^2 S_h^2}{n_h^2} + \lambda C_h = 0 \qquad \text{(I)}$$

نحل (I) من أجل $n_h$، فنجد:

$$n_h = \frac{N_h S_h}{N\sqrt{C_h}}.\frac{1}{\sqrt{\lambda}} \qquad \text{(II)}$$

لدينا:

$$n = \sum_{h=1}^{k} n_h \qquad \text{(III)}$$

بتعويض العلاقة II في III نتحصل على :

$$n = \frac{\sum^{k}\left(N_h S_h / \sqrt{C_h}\right)}{N\sqrt{\lambda}} \qquad \text{(IV)}$$

نستخرج $\sqrt{\lambda}$ من (IV) ونعوضه في (II)، فنجد:

$$n_h = \frac{N_h S_n / \sqrt{C_h}}{\sum^{k}\left(N_n S_h / \sqrt{C_h}\right)}.n \qquad \text{(18-6)}$$

لما تكون التكلفة الإجمالية محددة، نعوض $n_h$ في $C=C_0+\Sigma C_h n_h$ فنحصل على:

$$C=C_0+n\frac{\sum N_h S_h\sqrt{C_h}}{\sum\left(N_h S_h/\sqrt{C_h}\right)} \qquad (19\text{-}6)$$

ويعطى حجم العينة n الأمثل كما في العلاقة (6–14).

2– من أجل تحديد قيم $n_h$ التي تدني التكلفة الإجمالية علما أن التباين محدد عند القيمة $Var^*$، نشكل دالة لاغرانج F كالآتي:

$$F=\Sigma C_h n_h+\lambda\left[\frac{1}{N^2}\sum^{k}N_h S_h^2\left(\frac{1}{n_h}-\frac{1}{N_h}\right)-Var^*\right]$$

باستعمال نفس الكيفية كما في (1)، نجد أن القيم المثلى لـــ nh معطاة كما في العلاقة (6–13) والملاحظ أن دالة لاغرانج في كلا الحالتان تعطي لنا نفس النتيجة.

لما يكون التباين مثبت عند القيمة:

$$Var\left(\bar{y}_{st}\right)=\frac{1}{N^2}\sum^{k}N_h^2 S_h^2\left(\frac{1}{n_h}-\frac{1}{N_n}\right)=Var^*$$

بتعويض القيمة المثلى لـــ $n_h$ في هذا التباين، نحصل على التباين الأمثل كما في العلاقة (6–16). كذلك نستخرج حجم العينة الكلي n الأمثل والمعطى بالعلاقة (6–15).

نلاحظ من خلال الصيغة (6–13) أنه كلما كان حجم الطبقة h أكبر وكانت التغيرات فيها أكبر، وتكلفة سحب وحدة واحدة إضافية منها أصغر، كان حجم العينة المسحوب من هذه الطبقة أكبر، والعكس صحيح. كما نلاحظ كذلك أن المحاصة المثلى تتأثر بالجذر التربيعي للتكاليف الوحدوية، فإذا كانت التغيرات في هذه التكاليف طفيفة بين الطبقات فإن المحاصة لا تتأثر كثيرا، كما يمكننا الاستغناء عن استعمال دالة التكلفة في حساب المحاصة المثلى، ما لم تتعدى هذه التغيرات في التكاليف المعامل 3 فما فوق[11].

[11] Hansen, Hurwitz and Madow , "Sample survey Method and theory". Vol 1 , John Wiley and Sons, 1993, p222.

في الكثير من الأحيان تختلف التكلفة الوحدوية من طبقة إلى أخرى، إلا أن Neyman وفي سنة 1934 قام بعرض حالة خاصة، أين تكون فيها هذه التكلفة نفسها في جميع الطبقات، فتصبح التكلفة الإجمالية.

$$C = C_0 - cn \qquad (20\text{-}6)$$

حيث:

$$C_h = c \qquad \forall (h = 1,2,\dots\dots, k)$$

وتصبح المحاصة المثلى في العلاقة (6-13) كما يلي:

$$n_h = n \frac{N_h S_h}{\sum_{h=1}^{k} N_h S_h} \qquad (21\text{-}6)$$

لأنه يتم اختزال $\sqrt{C_h}$ في البسط والمقام لكونهما متساويان.

تجدر الإشارة إلى أن المحاصة المثلى في حالة تكلفة مثبتة تصبح نفسها المحاصة المثلى من أجل حجم عينة ثابت، وتسمى هذه المحاصة بمحاصة نيمان على أسم مبتكرها.

في هذه الحالة، تكون صيغة التباين الأمثل (الأصغر) من أجل تكلفة مثبتة، أو n مثبت هي:

$$Var_{\min}(\bar{y}_{st}) = \frac{\left(\sum W_h S_h\right)^2}{n} - \frac{\sum W_h S_h^2}{N} \qquad (21\text{-}6)$$

أو

$$Var_{\min}(\bar{y}_{st}) = \frac{1}{N^2}\left[\frac{\left(\sum N_h S_h\right)^2}{n} - \sum N_h S_h^2\right] \qquad (22\text{-}6)$$

### 3-6-2-2 المحاصة التي تحتاج إلى معاينة تزيد عن 100%:

أحيانا يجد الإحصائي نفسه مجبرا على إجراء التعديلات على القيم المثلى لـــ $n_h$، وذلك في بعض الظروف الخاصة، إذ يمكن أن تعطي المحاصة المثلى قيمة لـــ $n_h$ تكون أكبر من حجم الطبقة $N_h$، وهذا يعني أننا مطالبون بسحب أكثر من 100% من الطبقة h الشيء الذي يعتبر أمرا مستحيلا. يحدث هذا عندما يكون كسر المعاينة الإجمالي كبير نوعا ما، ويكون تغير إحدى الطبقات أكثر بكثير من الطبقات الأخرى (Cochran, 1977)، يكمن الحل في وضع $N_n = \tilde{\tilde{n}}_h$ في جميع الطبقات التي يكون فيها هذا المشكل مطروحا، ثم

يوزع باقي حجم العينة على الطبقات المتبقية عن طريق صيغة المحاصة المثلى. لنوضح ذلك بمثال بسيط:

لنفرض أنه عند تطبيقنا للمحاصة المثلى وجدنا أن $N_1 \langle n_{1opt}$ ، حيث أن عدد الطبقات 2≤h. في هذه الحالة المحاصة المثلى المحسنة تكون كما يلي:

$$\tilde{\tilde{n}}_1 = N_1$$

$$\tilde{\tilde{n}}_2 = (n - N_1)\frac{W_h S_h}{\sum W_h S_h}$$

$$\tilde{\tilde{n}}_h = (n - N_1)\frac{W_h S_h}{\sum W_h S_h} \qquad (23\text{-}6)$$

تكون العلاقة (6-23) صحيحة إذا كان كل $N_h \rangle n_h$مهما كان 2<h، أي في كل الطبقات المتبقية. أما إذا كان $N_2 < \tilde{\tilde{n}}_2$ فإن المحاصة المحسنة تصبح:

$$\tilde{\tilde{n}}_1 = N_1$$

$$\tilde{\tilde{n}}_2 = N_2$$

$$\tilde{\tilde{n}}_h = (n - N_1 - N_2)\frac{W_h S_h}{\sum W_h S_h} \qquad (24\text{-}6)$$

حيث أن عدد الطبقات أكبر من أو يساوي 3 (3≤h)، ونواصل العملية حتى نتحصل على الشكل العام الموالي:

$$\tilde{\tilde{n}}_{h'} = N_{h'}, \qquad \forall (h' = 1,2,.....,m)$$

حيث m هو عدد الطبقات التي يكون فيها المشكل مطروحا k≥m ،عدد الطبقات.

$$\tilde{\tilde{n}}_h = \left(n - \sum_{h'=1}^{m} N_{h'}\right)\frac{W_h S_h}{\sum W_h S_h} \qquad (25\text{-}6)$$

في مثل هذه الحالات أين تكون فيها المحاصة المثلى مصححة، يجب أخذ الحذر في حساب التباين الأمثل لأن صيغة التباين (6-21) تصبح غير صحيحة. يقترح كل من " Hanssen " و "Hurwitz " و "Madow" (1993) أن لا نأخذ بعين الاعتبار إلا الطبقات التي تستوفي الشرط $N_h \rangle n_h$. عند تطبيقنا للصيغة (6-21). في حين يرى (Cochran، 1977) أن نأخذ بعين الاعتبار الأحجام المصححة لــ $n_h$ المثلى التي رمزنا لها بــ $\tilde{\tilde{n}}_h$ في حسابنا للتباين الأمثل وتعطى صيغة التباين الجديدة كما يلي:

$$Var_{\min}(\overline{y}_{st}) = \frac{\sum{}^{/}(W_h S_h)^2}{n^{/}} - \frac{\sum{}^{/} W_h S_h^2}{N} \quad (26-6)$$

حيث تمثل $n^{/}$ مجموع الأحجام المثلى المصححة ضمن الطبقات أي حجم العينة الإجمالي المصحح، ويرمز المجموع $\sum^{/}$ إلى ذلك المتعلق بالطبقات التي يكون فيها $N_h \geq \tilde{\tilde{n}}_h$.

## 3-6-2-3 المحاصة المثلى في حالة معاينة أكثر من خاصية واحدة:

لقد رأينا حتى الآن كيفية حساب حجم العينة الأمثل في كل طبقة باستعمال المحاصة المثلى في حالة معاينة خاصية واحدة من المجتمع (المتوسط ، أو المجموع ، أو النسبة).

لكن في الكثير من الأحيان، تهدف الدراسات إلى معاينة أكثر من خاصية واحدة إن لم نقل عدد كبير من الخصائص. في هذه الحالة، تصبح نتائج المحاصة لخاصية واحدة تختلف عن نتائج المحاصة للخصائص الأخرى. ويمكن أن يكون هذا الاختلاف طفيفا كما يمكن أن يكون كبيرا لدرجة أنه تستعصي علينا عملية اختيار قيمة محاصة ما دون أخرى.

في الواقع، لا يوجد لدينا حل بسيط و واضح، وموحد لحل هذا النوع من المحاصة المعقدة، بل تتوفر لدينا مجموعة من الاقتراحات من هنا وهناك ترمي كلها إلى تقريب المحاصة إلى الحل الأنسب والأوسط. من أهمها نذكر تلك التي جاء بها Cochran (1977)، والتي تقتضي حساب المحاصة المثلى لكل خاصية على حدى. ولكن قبل ذلك تختزل الخصائص المدروسة، بحيث لا نحتفظ في حساب المحاصة إلا بالمتغيرات الأكثر أهمية.

بعد حصولنا على النتائج المختلفة لـ $n_{hj}$، حيث يرمز j إلى المتغير أو الخاصية قيد الدراسة، نحسب المتوسط الحسابي لهذه المحاصات في كل طبقة والذي يساوي:

$$\overline{n}_h = \frac{\sum_{j=1}^{m} n_{hj}}{m} \quad (27-6)$$

حيث m هو عدد الخصائص التي تدخل في الحسبان لحساب المحاصة في كل طبقة h. وقد لاحظ Cochran في المثال الذي عرضه أن نتائج هذه المحاصة الوسطية لا تختلف كثيرا

عن المحاصة التناسبية سواء تعلق الأمر بالأحجامأأااتا المثلى للعينات في الطبقات، أو فيما يخص الدقة المترتبة عن هذه النتائج.

ويرى كل من Hansen, Hurwitz and Madow (1993) أنه في حالة ما إذا كانت نتائج المحاصة للخصائص المختلفة لا تتغير كثيرا فيما بينها، يكفي عندئذ الاحتفاظ بنتيجة المحاصة لخاصية واحدة، إذا كانت هذه الأخيرة محققة وجيدة بالنسبة للخصائص الأخرى. أما إذا كانت قيم المحاصات متباعدة نوعا ما، فإننا نعاين أولا من أجل خاصية واحدة ونحصل على $n_{hj}$ المثلى، ثم نستعمل هذه الأخيرة في تقدير خاصية أخرى. فإذا كانت النتائج جيدة فلا بأس وإلا فإنه يستوجب علينا إضافة وحدات أخرى من الطبقة التي لم تكن تمتلك تمثيلية أكبر في المحاصة السابقة. كما يرى هؤلاء الباحثون أنه ما لم تتم معاينة خاصية واحدة فقط من المجتمع، فإنه من الأحسن الاكتفاء باستعمال المعاينة الطبقية بحصص متناسبة.

يرى Chatterjee (1967) أنه من الأحسن حساب محاصة وسطية بديلة عن كل المحاصات الجزئية المحسوبة لكل خاصية، والتي تهدف إلى تدنية معدل الزيادة النسبية الناجم عن الانحراف عن المحاصة المثلى، عند حساب التباين عاديا. وتعطى صيغة هذا المعدل كالآتي:

$$\frac{Var\left(\overline{y}_{st}\right)-Var_{\min}\left(\overline{y}_{st}\right)}{Var_{\min}\left(\overline{y}_{st}\right)}=\frac{1}{n}\sum_{h=1}^{k}\frac{\left(\hat{n}_h-n'_h\right)^2}{\hat{n}_h} \quad (28\text{-}6)$$

حيث، nh هو الحجم الأمثل، و $\hat{n}_h$ هو الحجم الفعلي للعينة في الطبقة h، و $Var_{\min}(\overline{y}_{st})$ هو التباين الأمثل عند استعمال المحاصة المثلى علاقة (6–21) و $Var(\overline{y}_{st})$ هو التباين الناتج عن عدم استعمال المحاصة إطلاقا.

مشكل تدنية معدل الزيادة النسبية في التباين وفق (6–28) على أن نأخذ هذا المعدل فوق المتغيرات (الخصائص) المختلفة يؤدي إلى الإختيار التالي:

$$n_h = n\sqrt{\sum_{j}^{m} n'^2_{hj}} \Bigg/ \sum_h \sqrt{\sum_{j}^{m} n'^2_{hj}} \quad (29\text{-}6)$$

حيث أن $n'_{hj}$ هو حجم العينة الأمثل في الطبقة h وفق المتغير j.

اقترح Yates (1960) طريقتين مهمتين لحساب الحجم الأمثل في حالة وجود اختلاف كبير بين المحاصات المثلى. تهدف الطريقة الأولى إلى وضع الخسارة الكلية المتوقعة كدالة خطية بالنسبة لتباينات المتوسطات التقديرية للمجتمع، والتي أعطاها الصيغة التالية:

$$L=\sum_{j}^{m} a_j Var\left(\bar{y}_{jst}\right)=\sum_{j}^{m} a_j \sum_{h}^{k} W_h^2 S_{jh}^2\left(\frac{1}{n_h}-\frac{1}{N_h}\right) \qquad (30\text{-}6)$$

حيث أن $S_{jh}^2$ هو تباين المتغير في الطبقة h ، و $Var(\bar{y}_{jst})$ هي التباينات المرغوبة لكل متغير j على حدى، وأن القيم $a_j$ تتناسب عكسيا مع هده التباينات . يعطي YATES مثالا بسيطا أين نكون فيه في حالة متغيرين ويكون $a_1=V_2/(V_1+V_2)$، $a_2=V_1/(V_1+V_2)$، حيث تصبح

$$L=\frac{2V_1V_2}{V_1+V_2}.$$

يمكن للصيغة ( 30-6 ) أن تأخذ شكلا آخر وذلك بعد إجراء عملية النشر و هو :

$$L=\sum_{j}\frac{W_h^2}{n_h}\left(\sum_{j} a_j S_{jh}^2\right)-\frac{1}{N}\sum W_h\left(\sum a_j S_{jh}^2\right) \qquad ({}^{/}30-6)$$

وعلما أن دالة التكلفة هي خطية ومن الثكل:

$$C=C_0+\sum C_h n_h$$

ومن أجل L محددة ومع افتراض تجاهل معامل التصحيح لمجتمع منته، تكون التكلفة أصغر ما يمكن إذا كان الحجم الأمثل nh كما يلي:

$$n_h=\frac{n\left(W_h A_h/\sqrt{C_h}\right)}{\sum\left(W_h A_h/\sqrt{C_h}\right)} \qquad (31\text{-}6)$$

حيث

$$A_h=\sqrt{\sum a_i S_{jh}^2}$$

بتعويض ( 6-31) في $({}^{/}30-6)$ نحصل على حجم العينة الكلي الأمثل:

$$n=\frac{1}{L}\left(\sum_{h}\frac{W_h A_h}{\sqrt{c_h}}\right)\left(\sum_{h} W_h A_h\sqrt{c_h}\right) \qquad (32-6)$$

وتتمثل الطريقة الثانية في تدنية التكلفة الإجمالية C تحت القيود $Var\,\bar{y}_{jst}\le V_j$، حيث $j=1,2.....m$ هو عدد المتغيرات. والمسألة هنا هي إحدى مسائل البرمجة غير الخطية.

## 3-6-3 حجم العينة في حالة معاينة النسب:

يمكن بسهولة استنتاج القوانين المتعلقة بتحديد حجم العينة في معاينة النسب انطلاقا من تلك التي تطرقنا إليها سابقا في حالة متغير من طبيعة مستمرة ($\bar{y}$)، وكما رأينا في الفقرة (6-1) فإن حجم العينة الضروري لبلوغ دقة معينة معطى بالعلاقة (6-3) كما يلي:

$$n = \frac{\sum \frac{W_h^2 S_h^2}{w_h}}{\left(d/t\right)^2 + \frac{1}{N}\sum W_h S_h^2}$$

إذا يكفي أن نعوض $S_h^2$ بالتباين في حالة النسب والذي يساوي $S_h^2 = \frac{N_h}{N_h - 1} P_h Q_h$ من العلاقة (2-30 فصل 2) حتى نحصل على حجم العينة الضروري وهو:

$$n = \frac{\sum \frac{W_h^2}{w_h} \cdot \frac{N_h}{N_h - 1} P_h Q_h}{\left(d/t\right)^2 + \frac{1}{N}\sum W_h \frac{N_h}{N_h - 1} P_h Q_h} \qquad (1\text{-}3\text{-}6)$$

فإذا اعتبرنا أن $\frac{N_h}{N_h - 1}$ يساوي الواحد، تصبح العلاقة (6-3-1) كما يلي:

$$n = \frac{\sum \frac{W_h^2}{w_h} . P_h Q_h}{\left(d/t\right)^2 + \frac{1}{N}\sum W_h P_h Q_h} \qquad (2\text{-}3\text{-}6)$$

في حالة إهمال عامل الـــ ت.م.م نحصل كتقريب أولي لحجم العينة على:

$$n_0 = \frac{1}{\left(d/t\right)^2} \sum \frac{W_h}{w_h} P_h Q_h \qquad (3\text{-}3\text{-}6)$$

فإذا كانت المعاينة بحصص متناسبة، فإن الصيغ (6-3-2) و (6-3-3) تصبح كما يلي:

$$n = \frac{n_0}{1 + \frac{n_0}{N}} \qquad (4\text{-}3\text{-}6)$$

حيث:

$$n_0 = \frac{\sum W_h P_h Q_h}{\left(d/t\right)^2} \qquad (5\text{-}3\text{-}6)$$

أما فيما يخص القوانين المتعلقة بالمحاصة المثلى في حالة النسب، فيكفي أن نعوض $S_h^2$ بـــ $P_h Q_h$ في جميع الصيغ التي رأيناها في الفقرة (6-2-1) وذلك في مختلف المحاصات، سواء كانت بتكلفة ثابتة في جميع الطبقات (Nayman) أو بتكاليف متغيرة.

## 3-6-4 تحديد حجم العينة عند مقارنة متوسطات الميادين:

في كثير من الأحيان، يمكن للمجتمع الإحصائي أن يشتمل على عدة خصائص تكون ممثلة في جميع الطبقات. فقد نرغب في إجراء تقديرات حول هذه الخصائص، ويمكن أن تكون هذه الخصائص متمثلة في الجنس (ذكر أو أنثى) أو فئة الأعمار (أطفال، شباب، كهول...الخ) أو لون البشرة (بيضاء، سمراء ،أو سوداء) أو حتى نوع المصاريف (مصاريف تتعلق بالغذاء، وأخرى بالثياب،...الخ). في الإحصاء الاستقرائي(المعاينة) تسمى هذه الخصائص بالميادين.

تهدف العديد من الدراسات إلى إجراء مقارنة بين ميدانين أو عدة ميادين في المجتمع، فيمكن مثلا أن تسعى دراسة ما حول سلوك المستهلكين إلى مقارنة مجموع مصاريف اللباس عند الذكور مع مجموع هذه المصاريف عند الإناث، أو تهدف دراسة أخرى إلى مقارنة متوسط وزن الذكور مع متوسط وزن الإناث الذين تتراوح أعمارهم بين 20 و25 سنة في مدينة ما. سوف نعالج في هذه الفقرة مسألة تحديد حجم العينة عند مقارنة متوسطا ميداني دراسة أو أكثر.

ليكن لدينا العدد k من الطبقات و D ميدان دراسة ويتم سحب عينات عشوائية بسيطة من كل طبقة. فإذا كان $n_h$ حجم العينة المسحوبة من الطبقة h، فإن $n_{hj}$ يمثل عدد الوحدات التي تنتمي إلى الميدان j ضمن هذه العينة، حيث $\sum_{j=1}^{D} n_{hj} = n_h$ وتعتبر ألــ $n_{hj}$ متغيرات عشوائية لأنه لا يمكن تحديد ومعرفة ميادين الدراسة إلا بعد إجراء المعاينة.

في هذه الفقرة، وخلافا للمنهجية المتبعة السابقة في حساب المحاصة المثلى، سوف نسعى (أ) إلى تثبيت حجم العينة n بهدف تعظيم الاحتمال P(n) لتحقيق دقة مرجوة للمقارنة، علما أن $\sum n_h = n$، و(ب) تدنية حجم العينة n لما يكون P(n)=P* أين P* هو الاحتمال الثابت المرجو.

لنرمز بـــ $N_{hj}$ إلى حجم المجتمع في الطبقة h والميدان j، ومنه يكون $N_j = \sum_{h=1}^{k} N_{hj}$ هو حجم المجتمع الخاص بالميدان j فوق جميع الطبقات. يمثل كل صف $N_{h.}$ و $n_{h.}$ على التوالي حجم المجتمع والعينة في الطبقة h. $\bar{Y}_{hj}$ و $\bar{y}_{hj}$ هما كذلك على التوالي متوسط

المجتمع و متوسط العينة في الطبقة h، الميدان j . يمثل $P_{hj}=\frac{N_{hj}}{N_h}$ نسب الوحدات التي تنتمي إلى الميدان j في الطبقة h و $\sigma_{hj}^2$ هو التباين في الطبقة h الميدان j.

نريد دراسة الفرق بين متوسطا ميداني دراسة $\overline{Y}_{.i}-\overline{Y}_{.j}$.حيث أن $\overline{Y}_{.i}=\sum N_{hi}\overline{Y}_{hi}/N_{.i}$ . و كتقدير للمتوسط $\overline{Y}_{.i}$ نجد $\overline{y}'_{.i}=\sum_n \frac{\hat{N}_{hi}\overline{y}_{hi}}{\hat{N}_{.i}}$ ، حيث يكون $\hat{N}_{.i}=\sum_h^k \hat{N}_{hi}$ تقدير غير متحيز لــ $N_{.i}$ ، حيث أن: $\hat{N}_{hi}=N_{h.}.n_{hi}/n_{h.}$ و $\overline{Y}_{hi}=\overline{y}_{hi}$ .

سوف نفترض (من أجل التبسيط) أن أحجام العينة ضمن الطبقات كبيرة لدرجة أنه يمكننا تجاهل خطأ تقدير $N_{.i}$ . في بعض الحالات يمكن أن تكون قيم $N_{.i}$ معلومة من معطيات الحصر الشامل، وعليه يعطى مقدر $\overline{Y}_{.i}$ كما يلي:

$$\overline{y}_{.i}=\frac{\sum_{h=1}^{k}N_h n_{hi}\overline{y}_{hi}}{n_{h.}N_{.i}} \qquad (1\text{-}4\text{-}6)$$

أحيانا، يمكن أن يريد الواحد منا الحصول على تباين محدد V والذي يحقق:

$$Var\{\ (\overline{y}_{.i}-\overline{y}_{.j})|n_1,n_2,...,n_k\}\leq V$$

حيث أن:

$$n_h=\sum_{i=1}^{D}n_{hi}$$

إذا كنا بصدد دراسة مجتمع بميداني دراسة، يكون الاحتمال P(n) المذكور سابقا معطى بالصيغة التالي:

$$P(n)=\Pr\left[\{\ Var(\overline{y}_{.1}-\overline{y}_{.2})|n_1,...,n_k\ \}\leq V\right] \qquad (2\text{-}4\text{-}6)$$

من الجدير بالذكر أنه قد تم سابقا التعرض إلى هذا النوع من القضايا، وذلك بهدف تدنية القيمة المتوقعة للتباين $E\{V(|.)\}$ المذكور أعلاه، علما أن حجم العينة الإجمالي n مثبت، أو تدنيه n لما يكون التوقع $E\{V(|.)\}$ ثابت[12] .

من أجل تقييم الاحتمال P(n)، يجب علينا إرفاق جميع القيم بالمعالم التالية: $\sigma_{hi}^2$ و $P_{hi}$. يمكن تقدير هذه المعالم جيدا، انطلاقا من دراسات سابقة أو من خلال إجراء مسوحات استطلاعية كما ذكرنا سابقا، ويمكن للواحد منا أن يعتقد بأن التقدير المسبق للنسب Phi سوف يكون ضعيفا نوعا ما، ولقد أوضح كل من H.liao and Sedrank[13] في تجربة أعداها أنه إذا كان الخطأ ضئيلا في تقدير Phi فإن هذا سوف ينتج محاصة مثلى لـــ n تقترب من القيمة الحقيقية والمثلى لـــ n.

### 3-6-4-1 تحديد حجم العينة لتعظيم P(n):

#### 1-1- ميدانين للدراسة:

لنعتبر المقارنة بين متوسطا ميداني دراسة، حيث أن الميدانين مستقلين عن بعضهما البعض وينتميان إلى مجتمع منته.

من أجل حل المسألة (أ) المطروحة في التمهيد، يجب علينا حساب أحجام $n_h$ التي تعظم الاحتمال P(n):

$$P(n) = \Pr\left\{Var(\overline{y}_1 - \overline{y}_2) \leq V\right\} \qquad (3\text{-}4\text{-}6)$$

ولقد بين كوكران[14] أن تباين الفرق بين متوسطا ميداني دراسة هو كما يلي:

$$Var(\overline{y}_{.i} - \overline{y}_{.j}) = Var(\overline{y}_{.i}) + Var(\overline{y}_{j}) \qquad (4\text{-}4\text{-}6)$$

$$= \frac{S_i^2}{n_i} + \frac{S_j^2}{nj}$$

[12] Sedrank, KJ. " A double sampling scheme for Analytical surveys".JASA, n°60, 985-1004 .1965, p520.

[13] H Liao , J. Sedrank ." Selection of strata sample sizes for the comparison of domain means". JASA, vol 78 , n°384, 12/1983. p873.

[14]مرجع سبق ذكره ، ص56.

وذلك عند تجاهل معامل الـــ ت.م.م

تبقى خصائص المقدرات (الفصل 2) محققة من أجل متوسط العينة $\bar{y}_{hi}$ ، وعليه يكون تباين الفرق بين المتوسطين كما يلي:

$$Var(\bar{y}_{.1}-\bar{y}_{.2})=\sum_{i=1}^{2}Var(\bar{y}_{.i}) \qquad (5\text{-}4\text{-}6)$$

وبتعويض قيمة $\bar{y}_{hi}$ كما في العلاقة (6-4-1) في هذان التباينان نحصل على :

$$Var(\bar{y}_{.1}-\bar{y}_{.2})=\sum_{i=1}^{2}Var\sum_{h=1}^{k}\left(N_h n_{hi}.\bar{y}_{hi}/n_h.N_{.i}\right)$$

$$=\sum_{i=1}^{2}\sum_{h=1}^{k}\left(k'_{hi}\Big/n_{h.}\right)^2 n_{hi} \qquad (6\text{-}4\text{-}6)$$

حيث :

$$k_{hi}=\frac{N_h.\sigma_{hi}}{N_{.i}}$$

تصبح صيغة الاحتمالي P(n) بتعويض (6-4-6) في (6-4-3) كما يلي:

$$p(n)=\Pr\left\{\sum_{i=1}^{2}\sum_{h=1}^{k}\left(k_{hi}\Big/n_{h.}\right)^2 n_{hi}\leq V\right\} \qquad (7\text{-}4\text{-}6)$$

سوف نفترض من أجل التسهيل فقط أن السحب يتم بإرجاع، لذا تتبع الـــ nhi التوزيع الثنائي ذو المعالم $b(n_{h.},p_{h1.})$ ، وذلك من أجل h=1,2,...k . تكون الـــ $\{n_{hi}\}$ مستقلة فيما بينها.

بعد إجراء بعض التعديلات على (6-4-7) بتعويض $n_{h2}$ بـــ $n_h-n_{h1}$ و وضع $V'=V-\sum_h K_{h2}^2\Big/n_h$ سوف نحصل على:

$$p(n)=\Pr\left\{\sum_{h}\left(k_{h1}^2-k_{h2}^2\right).n_{h1}\Big/n_{h.}^2\leq V'\right\} \qquad (8.4.6.)$$

حيث أن: $k_{h1}\neq k_{h2}$ $\forall$ h .

يعرض[15] (LIAO, 1975) الحل في حالة ما إذا كان $k_{h1}=k_{h2}$. أما نحن فنضع $G_h=K_{h1}^2-K_{h2}^2\Big/n_h^2$ ، فيكتب الاحتمال P(n) كما يلي:

[15] H.Liaho, some sampling designs for comparison of domain means. Unpublished PHD dissertation, university of Wisconsin, Madison, dep of statistics.

$$Pn = \sum_{n_{h1} \in S} \left\{ \prod_{h=1}^{k} \binom{n_h}{n_{h1}} P_{h1}^{n_{h1}} q_{h1}^{n_h - n_{h1}} \right\}$$

حيث:

$$q_{h1} = 1 - p_{h1}$$

$$S = \left\{ \{n_{h1}\} \Big/ \sum_{h=1}^{k} G_h n_{h1} \le V', 0 \le n_{h1} \le n_h \quad h = 1,2,...,k \right\} \qquad (9\text{-}4\text{-}6\ )$$

$$0 \le n_{h1} \le n_{h.} \,........ (h = 1,...,k)$$

يمكن التوصل إلى حل المسألة (أ) بتقييم P(n) كما في العلاقة ( 9-4-6) من أجل كل القيم لـــ n مع $\sum n_h = n$. تعتبر هذه الطريقة هدرا للوقت والمال وذلك حتى باستعمال الحاسوب، لدى سوف نستعين ببعض التقريبات.

باستعمال التقريبات الطبيعية للتوزيعات الثنائية برهن (1983) LIAO على التقريب التالي للاحتمال P(n):

$$P_{NA}(n) = \Phi \left[ \frac{V - \sum b_h n_{h.}}{\left\{ \sum (d_h / n_{h.}^3) \right\}^{1/2}} \right] \qquad (10 - 4 - 6)$$

حيث:

$$\Phi(z) = \int_{-\infty}^{Z} (2P)^{-\frac{1}{2}} \exp(-t^2/2)dt,$$

$$k_{hi} = N_h \sigma_{hi} / N_{.i}$$

$$b_h = P_{h1} k_{h1}^2 + q_{h1} k_{h2}^2, \qquad (11.4.6\ \ )$$

$$d_h = P_{h1} q_{h1} (k_{h1}^2 - k_{h2}^2)^2, \qquad (12.4.6\ \ )$$

كون $\Phi(T)$ دالة متزايدة بالنسبة لـــ

$$T = \left[ \frac{V - \sum b_h n_{h.}}{\left\{ \sum (d_h / n_{h.}^3) \right\}^{1/2}} \right] \qquad (13\text{-}4\text{-}6)$$

فإن تعظيم (10-4-6) مع $\sum n_h = n$ مثبت هو نفسه تعظيم T بالنسبة لـــ $n = \sum n_h$،

كون T دالة معقدة بالنسبة لـــ nh فإنه لا يوجد حل واضح لمسألة تحديد القيم المثلى لـــ nh.

يمكن عرض طريقة عملية تعتمد على أسلوب التكرار أو الإعادة (Itération) كما يلي:

(أ) لتكن $Q = T - \lambda\left(\sum n_h - n\right)$، حيث يمثل λ معامل "لاغرانج". بعد أخذ المشتقات الجزئية ووضعها مساوية للصفر، وبعد إجراء بعض التعديلات والتحويلات الرياضية، نتحصل على الدالة التالية:

$$n_h^{-2} = \frac{D_n}{3d_h B_n}\left[-b_h + \left\{b_h^2 + \frac{3d_h B_n\left(2V + B_n\right)}{nD_n}\right\}^{1/2}\right] \qquad (14\text{-}4\text{-}6)$$

حيث أن:

$$D_n = \sum d_h / n_h^3$$

$$B_n = V - \sum b_n / n_h$$

(ب) ليكن $f\left(\{n_{h.}\}\right)$ رمزا للطرف الأيمن من العلاقة (6-4-14).

نضع:

$$m_{h.}^{(i)} = \left\{f\left(\left\{n_h^{(i-1)}\right\}\right)\right\}^{-\frac{1}{2}}$$

و:

$$n_{h.}^{(i)} = n.m_{h.}^{(i)} \Big/ \sum_h m_{h.}^{(i)}$$

حيث يرمز i إلى عمليات التكرار والتي تأخذ القيم ......,i=1,2,3

نبدأ عملية التكرار (itération) بقيم ابتدائية لـــ $n_h^{(0)}$ والتي تعطي لنا $\left\{m_h^{(1)}\right\}$ و $\left\{n_h^{(1)}\right\}$. نواصل عملية حساب هذه القيم إلى أن لا تتعدى قيمتين متتاليتين $n_h^{(i)}$ و $n_h^{(i+1)}$ قيمة ثابتة محددة مسبقا.

يوجد حل آخر لـــ $n_h$ نحصل عليه من خلال تعظيم بسط العلاقة (6-4-13) وهو:

$$n_h^* = n\sqrt{b_h} \Big/ \sum \sqrt{b_h} \qquad (15\text{-}4\text{-}6)$$

يعتبر هذا الحل من التقريبات الجيدة للحل الأمثل، في العديد من الحالات.

**1-2- مقارنة أكثر من ميدانين دراسي:**

عندما يكون لدينا D مجال، حيث 2<D فإن القيد الخاص بتباينات الفرق بين المتوسطات يصبح:

$$\sum \alpha_{ij} Var\left\{\left(\bar{y}_i - \bar{y}_j\right) / n_1,....,n_k\right\} \leq V$$

حيث تمثل $\alpha_{ij}$ الترجيحات التي تبين الأهمية النسبية لمقارنة متوسطات الميادينi وj. من أجل كل قيم i و j يكون $\alpha_{ij} = \alpha_{ji}$ وكذا $\sum_{i<j} \alpha_{ij} = 1$ ، و V هو تباين محدد وثابت.

الهدف هو إيجاد المحاصة المثلى لــ n ضمن k من الطبقات من أجل تعظيم الاحتمال p(n) حتى يكون القيد حول التباين محققا. يعطى (LIAO 1983) حلا جيدا لهذه المسألة باستعمال التقريبات الطبيعة المتعددة المتغيرات للتوزيعات المتعددة الحدود، ويمكن تقريب P(n) بـــ :

$$\Phi\left[\frac{V^* - \sum\left(b_n^* / n_h\right)}{\left\{\sum\left(d_h^* / n_h^3\right)\right\}^{\frac{1}{2}}}\right] \qquad (16\text{-}4\text{-}6)$$

$$\alpha_k = \sum_{j=1}^{D} \alpha_{kj} - \alpha_{kk}$$

$$V^* = V - \sum_{h}^{L} \alpha_D k_{hD}^2 / n_h$$

$$b_h^* = \sum_{j=1}^{D-1} \left(\alpha_j k_{hj}^2 - \alpha_D k_{hD}^2\right) P_{hj}$$

$$d_h^* = \sum_{j=1}^{D-1} \left(\alpha_j k_{hj}^2 - \alpha_D k_{hD}^2\right) P_{hj} - b_h^*$$

و $k_{hi}$ هي معطاة كما في العلاقة (6-4-10)

بما أن العلاقة (6-4-16) تشبه العلاقة (6-4-10) فإنه يمكننا استعمال نفس المنهجية المتبعة سابقا من أجل إيجاد القيم المثلى لــ $n = \left(n_1, n_2, ..., n_L\right)$.

### 3-6-4-2 تحديد حجم العينة الكلي n الأدنى:

لقد عرضنا سابقا مسألة إيجاد التقسيم الأمثل (المحاصة المثلى) لحجم العينة المعطى n ضمن الطبقات من أجل تعظيم الاحتمال p(n)، في هذه الفقرة سوف نعالج المسألة الثنائية للمحاصة (Dual allocation)، والمتمثلة في :

$$\begin{cases} Min\ n = \sum n_h & \text{إيجاد القيمة الدنيا لــ} \\ S/C\ \ P(n) = P^* & \text{تحد القيد} \end{cases}$$

حيث أن P* هو احتمال مرجو مثبت، 1>P*>0.

لقد بينا سابقا أنه في ميداني دراسة يمكن تقريب P(n) جيدا باستعمال التقريب الطبيعي $\Phi(T)$ كما في الصيغة(6-4-10). باستعمال هذا التقريب يمكن أن نصيغ المسألة السابقة بالطريقة التالية:

$$\begin{cases} Min\ \sum n_h \\ S/C\ \ T = C \end{cases} \qquad (17.4.6\ )$$

حيث $C = \Phi^{-1}(P^*)$ ، وT هي معطاة بالعلاقة (6-4-13). في الحالة الخاصة التي تكون فيها C=0 (أي لما P*=0,5) ، يمكن بسهولة توضيح أن المحاصة المثلى لــ $\{n_h^*\}$ وكذا القيمة الدنيا لــ n* هي المعطاة بــ :

$$n_h^* = \sqrt{b_h}\left(\sum \sqrt{b_h}\right)/V$$

$$n^* = \left(\sum \sqrt{b_h}\right)^2 /V \qquad (18\text{-}4\text{-}6\ )$$

عندما تكون $C \neq 0$، فإن العلاقة ( 6-4-18) لا تصبح الأمثل، ولكنها تعطينا فكرة عن الحجم الذي يجب أن تكون عليه العينة، وكذلك تساعدنا على حساب القيمة الابتدائية للحلول بالتكرار (itérative) الخاصة بالعلاقة ( 6-4-17).

إجراءات حل هذه المسألة لما يكون $C \neq 0$ هي مشابهة لتلك التي رأيناها في الفقرة السابقة (1-1).

نضع:

$$Q = \sum n_h + \lambda\left( \sum b_h / n_h + C\left(\sum d_h / n_h^3\right)^{1/2} - V\right)$$

بتطبيق الطريقة المعتادة لمضاعف لاغرانج، يمكن للواحد منا أن يحصل على الدالة التالية.

$$n_h^{-2} = \left(\sqrt{D_h}/3Cd_n\right)\left[-b_h + \left\{b_h^2 + \left(3Cd_h / \sum n_h\right).\left(C + 2V/\sqrt{D_n}\right)\right\}^{1/2}\right] \quad (19\text{-}4\text{-}6)$$

باستعمال نفس الطريقة التكرار المتبعة في الفقرة (1−1)، نضع $f(\{h_h\})$ هو الحد الأيمن من العلاقة (6-4-19) ونرمز بــ $m_h^{(k)}$ و $n_h^{(k)}$ إلى كل من التكرار رقم الــ $k$ لــ $m_h$ والقيمة لــ $n_h$.

نأخذ كقيمة ابتدائية $n_{h.}^{(0)} = n_h^*$ المعطاة بالعلاقة (6-4-18) ونعوضها بدلا من nh في الجهة اليمنى من الصيغة (6-4-19). عند كل مرحلة نحصل على القيم المتتالية لــ $\{m_h^{(k)}\}$ من الصيغة:

$$m_h^{(k)} = \left[f\left(\left\{n_h^{k-1}\right\}\right)\right]^{-\frac{1}{2}}$$

ونعدلها مع $n_h^{(k)}$ حتى نحقق:

$$\frac{V - \sum\left(b_h / n_h^{(k)}\right)}{\left\{\sum d_h / \left(n_{h.}^{(k)}\right)^3\right\}^{1/2}} = C \quad (20\text{-}4\text{-}6)$$

ليكن $T\left(\left\{n_h^{(k)}\right\}\right)$ يمثل الجهة اليسرى للعلاقة (6-4-20) .لنفرض أنه في المرحلة $k$، يكون $T\left(\left\{m_h^{(k)}\right\} = C\delta_k\right)$ حيث $\delta_k$ هو عدد حقيقي، وعليه يمكن تحديد القيم المعدلة لــ $n_h^k$ كما يلي:

$m_h^{(k)} = n_h^{(k)}$ إذا كان $T\left(\left\{m_h^{(k)}\right\}\right) = C$

$n_h^{(k)} = q_k m_h^{(k)}$ ما عدا ذلك.

حيث نختار qk التي تسمح بإبقاء العلاقة (6-4-20) محققة يمكن البرهنة على أن $q_k$ هي الحلول الجذرية لكثير حدود من الدرجة الثالثة (أنظر الملحق رقم:10) .ثم نكرر إجراءات العملية إلى أن نحصل على تغير صغير بين $n^{(k)}$ و $n^{(k+1)}$.

لكي نحصل على عينة عشوائية باستعمال أسلوب المعاينة العشوائية البسيطة أو الطبقية، يجب أن تتوفر لدينا قوائم أو سجلات تحتوي على جميع وحدات المجتمع الإحصائي. لكن في بعض الحالات، تكون عملية الحصول على هذه القوائم صعبة أو مستحيلة، نظرا لعدة أسباب مختلفة. إذ أن الكثير من الدول وخاصة منها المتخلفة، لا تتوفر على معطيات وبيانات حديثة حول السكان، أو المنازل، أو المزارع، أو المؤسسات الصناعية مثلا، كما أن إعداد مثل هذه القوائم لإجراء بحثنا سوف يكون أمرا غاية في الصعوبة والتعقيد، كما أنه سوف يكلفنا مبالغ ضخمة، بالإضافة إلى المدة الطويلة التي سوف يستغرقها مثل هذا العمل. إن استحالة القيام بهذا الأمر (إعداد القوائم) يعتبر أحد عيوب المعاينة العشوائية البسيطة و المعاينة العشوائية الطبقية، ومن أهم عيوبها كذلك، نجد تلك المتعلقة بإمكانية تباعد وتشتت الوحدات المصحوبة فيما بينها بشكل كبير جدا. فمثلا، إذا طلب منا استجواب 500 شخص في الجزائر حول أمر معين، فإن تكاليف التنقل بين هؤلاء الأشخاص من ولاية إلى أخرى سوف تكون كبيرة جدا، وخاصة إذا كانت هذه الولايات متباعدة، كما أن مثل هذا الأمر سوف يستغرقنا وقتا طويلا.

من أجل تخطي هذه المشاكل والعقبات المذكورة أعلاه، نستعين بأسلوب آخر في المعاينة العشوائية والذي يسمى بأسلوب المعاينة العنقودية.

# الفصل الرابع:

# تحديد حجم العينة في أسلوب المعاينة العشوائية العنقودية

يعرض هذا الفصل أسلوب المعاينة العشوائية العنقودية و كيفية تحديد حجم العينة الضروري عند استعمال هذا النوع من الأساليب. نتحدث في الفقرة الأولى عن مفهوم المعاينة العشوائية العنقودية. كما نعرض بعض التعاريف و الرموز و كذا مقدرات بعض معالم المجتمع و تبايناتها في الفقرتين الثانية و الثالثة على التوالي. الفقرة الرابعة مخصصة للحديث عن أثر المعاينة العشوائية العنقودية. أما الفقرة الخامسة فتعرض مختلف صيغ دوال التكلفة. نعرض في الفقرتين السادسة و السابعة كيفية تحديد حجم العينة و أهم المعايير الواجب مراعاتها في ذلك من أجل بلوغ أهداف معينة يحددها الباحث أو تدنية تكاليف دراسة أو بهدف تحقيق أغراض أخرى.

## 4-1المعاينة العنقودية :

المعاينة العشوائية العنقودية أسلوب يهدف إلى سحب عينة من مجتمع إحصائي يكون مقسما بشكل طبيعي إلى مجموعات أو أفواج جزئية تسمى عناقيد. و يحتوي كل عنقود بدوره على وحدات إختيار بحيث تنتمي كل وحدة من المجتمع إلى عنقود واحد فقط.

يتم تحديد العناقيد حسب أهداف الدراسة. فمثلا، يمكن للعنقود أن يمثل قسما يحتوي على مجموعة من التلاميذ، أو عمارة تحتوي على شقق، أو مصلحة بها مجموعة من الموظفين...إلخ.

بعد تحديد كل العناقيد في المجتمع، نسحب من بينها عينة من العناقيد باستعمال اسلوب المعاين العشوائية البسيطة، بحيث يصبح العنقود وحدة معاينة يطلق عليها اسم وحدة الدراسة الأولية.

في بعض الحالات، نحتفظ بجميع الوحدات التي تنتمي إلى العناقيد المسحوبة ونشكل بها العينة قيد الدراسة. نسمي هذا النوع من المعاينة بالمعاينة العنقودية على مرحلة واحدة. أما في حالات أخرى، فإننا نسحب من كل عنقود تم اختياره عينة عشوائية بسيطة جزئية من الوحدات والتي تسمى بوحدات الدراسة الثانوية، ونسمي هذا النوع من المعاينة بالمعاينة العنقودية على مرحلتين.

لا يمكن تطبيق كل من المعاينة العشوائية الطبقية والمعاينة العنقودية في نفس الظروف، إذ نستعمل أسلوب المعاينة الأول إذا لم تكن هناك تغيرات كبيرة بين وحدات

الطبقة (تجانس وحداتها) بينما يكون التغير من طبقة إلى أخرى كبيرا. في حين نستعمل أسلوب المعاينة الثاني إذا كان التغير بين وحدات العنقود كبيرا (عدم تجانس الوحدات) بينما يكون التغير من عنقود إلى آخر ضئيلا نوعا ما. بالإضافة إلى ذلك فإنه في أسلوب المعاينة الطبقية يتم تقسييم المجتمع إلى طبقات، بينما حسب أسلوب المعاينة العشوائية العنقودية يكون المجتمع مقسما سلفا وبشكل طبيعي إلى عناقيد.

ونظرا لعدم تجانس الوحدات في المعاينة العنقودية فإن نتائجها تكون أقل دقة من المعاينة العشوائية البسيطة والمعاينة الطبقية. كما أن تكاليف ومسائل التحليل الإحصائي للمعطيات يكون أكبر وأكثر تعقيدا. لكن تتم تغطية هذان العيبان بكون تكاليف دراسة كل وحدة إحصائية أكثر انخفاضا في المعاينة العنقودية مقارنة بالأسلوبان الآخران.

## 4-2 بعض المفاهيم والرموز :

سوف نستعمل في هذا الفصل رموزا إضافية لم يسبق أن عرضناها في أسالوب المعاينة العشوائية البسيطة و المعاينة العشوائية الطبقية، نرمز بـــ N إلى عدد الوحدات الإحصائية في المجتمع (تسمى كذلك بالوحدات الثانوية)، وبـــ $N_i$ إلى عدد الوحدات التي تنتمي إلى العنقود $i$. توزع الوحدات ألـــ N على مجموعة من العناقيد، التي يبلغ عددها M، والتي تسما بالوحدات الأولية، نظرا لكونها تسحب في المرحلة الأولى من المعاينة. ومن أهم الرموز في هذا الفصل نجد :

M : عدد العناقيد في المجتمع أوعدد الوحدات الأولية.

m : عدد العناقيد في العينة.

$N_i$ : عدد وحدات العنقود $i$ في المجتمع.

$n_i$ : عدد وحدات العينة المسحوبة من العنقود $i$.

$Y_{ij}$ : قيمة الخاصية $Y$ ذات المرتبة j في العنقود i.

حيث i = 1، 2، .... M.

j = 1، 2، .... $N_i$.

$y_{ij}$ = قيمة الخاصية Y ذات المرتبة j في العينة الجزئية المسحوبة من العنقود $i$،

بحيث:

$i = 1، 2، ...، m،$

و $j = 1، 2، ...، n_i$.

استنادا إلى هذه الرموز، نعرض بعض المجاميع التالية :

المجموع في العنقود $i$ ، $Y_i = \sum_{j=1}^{N_i} Y_{ij}$

المجموع في المجتمع، $Y = \sum_{i=1}^{M} Y_i$

عدد وحدات المجتمع، $N = \sum_{i=1}^{M} N_i$

ونجد في العينة ما يقابل هذه المجاميع على الترتيب، فيكون $y_i$ يمثل مجموع العينة في العنقود i ، و $y$ مجموع العينة ، وأخيرا n الذي يمثل حجم العينة.

## 3.4. المقدرات في المعاينة العنقودية وتبايناتها :

### *1.3.4. مقدر المجموع والمتوسط في المجتمع :*

في معاينة عشوائية عنقودية على مرحلة أو على مرحلتين يكون المقدر $\hat{Y}$ مقدر غير متحيز[16] لمجموع قيم المجتمع Y، وهو يعطى بالشكل الآتي :

$$\hat{Y} = \frac{1}{f} y = N \frac{y}{n} \quad \dots \ (1.1.3)$$

حيث يمثل $f$ معدل الاختيار الإجمالي، والذي يساوي بدوره حاصل ضرب معدل الاختيار في المرحلة الأولى $f_1$ و معدل الاختيار في المرحلة الثانية $f_2$، بحيث :

$$f = f_1 . f_2$$

و : $f_1 = \frac{m}{M}$

$$f_2 = \frac{\bar{n}}{\bar{N}} = \frac{n_i}{N_i}$$

حيث تمثل كل من $\bar{n}$، و $\bar{N}$ متوسط عدد الوحدات في كل عينة جزئية من كل عنقود ومتوسط عدد الوحدات في كل عنقود من المجتمع على التوالي.

[16] Hensen (1993 ) مرجع سبق ذكر ، فصل رقم 06 ص 253ه

هذا عن مجموع المجتمع أما عن متوسطه فإننا نرمز لمتوسط العينة بالرمز $\bar{y}$،

بحيث $\bar{y}=\frac{\sum y_i}{n}=\frac{y}{n}$

في هذا الفصل بـ

$$r=\frac{y}{n}=\frac{1}{n}\sum_{i=1}^{m}y_i=\frac{\sum_{i=1}^{m}y_i}{\sum_{i=1}^{m}n_i} \quad \dots (2.1.3)$$

حيث يمثل $n$ عدد وحدات العينة.

يمثل هنا $n_i$ حجم العينة المسحوبة من العنقود i.

بنفس الكيفية، يمثل $y=\sum_{i=1}^{m}y_i$ مجموع قيم وحدات العينة بالنسبة للمتغير أو الخاصية $Y_i$، ويمثل $y_i=\sum_{j=1}^{n_i}y_{ij}$ مجموع كل قيم العينة في العنقود $i$، كما يمثل $y_{ij}$ قيمة المتغير ذو الرتبة (j) في هذه العينة من العنقود (i).

يعتبر كل من $y$ و $n$ متغيران عشوائيان في سحب باحتمالات متساوية، لذلك، بما أن $n$ غير ثابت، كونه متغير، فإنه r مقدر متحيز نوعا ما لـ $\bar{Y}=\frac{Y}{N}$ [17]، بحيث N، وهو عدد الوحدات في المجتمع. وبالتالي فإن $\bar{Y}$ هو متوسط قيم كل الوحدات. يسمى المتوسط r كذلك بمتوسط النسبة (RATION MEAN)، والذي يمكن أن نعبر عنه بمتوسط الترجيحات لـ m متوسط يتم حسابه في كل عنقود:

$$r=\frac{y}{n}=\frac{1}{m}\sum_{i}^{m}\frac{n_i}{n/m}\bar{y}_i \quad \dots (3.1.3)$$

حيث $\bar{y}_i$ تساوي $\bar{y}_i=\frac{y_i}{n_i}$ و هو متوسط قيم العينة من العنقود i :

تعتبر المتوسطات مرجحة بالنسبة لأحجامها النسبية، أي $n_i/(\frac{n}{m})$ حيث يمثل $\frac{n}{m}$ متوسط عدد الوحدات في كل عنقود والذي نرمز له ب $\bar{n}$. فإذا كان هذا الحجم متساويا في جميع العناقيد، فإن المتوسط المذكور أعلاه يصبح يمثل المتوسط في حالة أحجام عناقيد متساوية.

[17] Kish, Chap 6, P186

$r$ هو مقدر للمتوسط في المجتمع، والذي بدوره يمكن أن نعبر عنه بالمتوسط المرجح :

$$\overline{Y} = \frac{Y}{N} = \frac{1}{M}\sum_{i}^{M} \frac{N_i}{(N/M)} \overline{Y}_i \qquad (4.1.3)$$

بحيث :

$$\overline{Y}_i = \frac{Y_i}{N_i} \qquad (4.1.4)$$

### *2.3.4. تباينات المجموع و المتوسط :*

في معاينة عشوائية عنقودية على مرحلتين، أين يتم في المرحلة الأولى سحب m عنقود بطريقة عشوائية بسيطة من مجموع Mعنقود ، ثم $n_i$ عينة جزئية من كل عنقود $i$ كذلك بمعاينة عشوائية بسيطة يكون تباين مقدر المجموع $\hat{Y}$ هو :

(1.2.3)

$$Var\left(\hat{Y}\right) = \frac{M^2}{m}\frac{M-m}{M}S_{1Y}^2 + \frac{M}{m}\sum_{i}^{M}\frac{N_i^2}{n_i}\frac{N_i-n_i}{N_i}S_{2iY}^2$$

حيث

$$S_{1Y}^2 = \sum_{i}^{M}\frac{\left(Y_i - \overline{Y}\right)^2}{M-1} \qquad (2.2.3)$$

$$S_{2iY}^2 = \sum_{j}^{Ni}\frac{\left(Y_{ij} - \overline{\overline{Y}}_i\right)^2}{N_i - 1} \qquad (3.2.3)$$

و

$$Y_i = \sum_{j}^{N_i} Y_{ij} \quad , \quad \overline{Y} = \frac{Y}{M} \quad , \quad \overline{\overline{Y}}_i = \frac{Y_i}{N_i} \qquad (4.2.3)$$

يمثل $S_{1y}^2$ تباين المجاميع بين العناقيد أو الوحدات الأولية. ويمثل $S_{2iy}^2$ تباين الوحدات الثانوية في العنقود i . كما يمثل كل من $\overline{Y}$ ، و $\overline{\overline{Y}}$ متوسط وحدات المجتمع و متوسط وحدات العينة على التوالي.

إذا كانت نسبة المعاينة (معدل الاختيار) في كل عنقود متساوية، أي $f_2 = \frac{n_i}{N_i}$ ،

تصبح صيغة تبايني مقدر المجموع هي :

$$Var\left(\hat{Y}\right)=M^{2}\frac{M-m}{Mm}S_{1y}^{2}+N^{2}\frac{\overline{N}-\overline{n}}{\overline{N}mn}S_{2Y}^{2} \qquad (5.2.3)$$

وحيث تمثل $S_{2y}^{2}$ التباين في المجتمع بين الوحدات داخل العناقيد و الذي يساوي:

$$S_{2y}^{2}=\frac{1}{N}\sum_{i}^{M}\frac{N_i}{N_i-1}\sum_{j}^{N_i}\left(Y_{ij}-\overline{\overline{Y}}_i\right)^{2}=\frac{1}{N}\sum_{i}^{M}N_i\,S_{2iy}^{2} \qquad (6.2.3)$$

توجد معلمة (خاصية) أخرى ألا وهي مربع التغاير (relavariance)، والتي تمثل مربع تباين المعلمة على المعلمة . في حالة المجموع يكون

$$\left(V(\hat{Y})\right)^{2}=M^{2}\;\frac{M-m}{M}\;\frac{B_{Y}^{2}}{m}+\frac{\overline{N}-\overline{n}}{\overline{\overline{N}}}\;\frac{W_{Y}^{2}}{m\overline{\overline{n}}} \qquad (7.2.3)$$

هو مربع التغاير للمجموع المقدر حيث :

$$\left.\begin{aligned} B_{Y}^{2}&=\frac{S_{1Y}^{2}}{\overline{Y}^{2}}\\ W_{Y}^{2}&=\frac{S_{2Y}^{2}}{\overline{\overline{Y}}^{2}}\end{aligned}\right\} \qquad (8.2.3)$$

بحيث يمثل كل من $B_{Y}^{2}$ ، و $W_{Y}^{2}$ مربع تغاير المجاميع بين العناقيد، ومربع تغاير الوحدات داخل العناقيد على التوالي.

بنفس الكيفية يمكن البرهنة بسهولة، أن تباين المقدر r في حالة تساوي معدل الاختيار $\frac{n_i}{N_i}=f_2$ هو :

$$Var\left(r\right)=\frac{M-m}{M}\,\frac{S_1^{2}}{m}+\frac{\overline{N}-\overline{n}}{\overline{\overline{N}}}\,\frac{S_2^{2}}{m\overline{\overline{n}}} \qquad (9.2.3)$$

ومربع تغاير المتوسط هو :

$$\left[V\left(r\right)\right]^{2}=\frac{M-m}{M}\;\frac{S_1^{2}}{m\overline{Y}^{2}}\;+\;\frac{\overline{N}-\overline{n}}{\overline{\overline{N}}}\;\frac{S_2^{2}}{mn\overline{\overline{Y}}^{2}} \qquad (10.2.3)$$

بحيث :

$$S_1^2 = S_{1Y}^2 + R^2\, s_{1X}^2 - 2\,R\,S_{1XY} \qquad (11.2.3)$$

مع

$$\left.\begin{aligned} R &= \frac{Y}{X} \\ S_{1XY}^2 &= \sum_{i}^{M} \frac{\left(Y_i - \overline{Y}\right)\left(X_i - \overline{X}\right)}{M-1} \end{aligned}\right\} \qquad (12.8.3)$$

$$S_{2i}^2 = S_{2iY}^2 + R^2\, S_{2iX}^2 - 2\,RS_{2iXY}$$

نضع $B^2 = \frac{S_1^2}{\overline{\overline{Y}}^2}$ ، و $W^2 = \frac{S_2^2}{\overline{\overline{Y}}^2}$، وبتعويض هاتان القيمتان في 11.8.3 نحصل على صيغة أخرى لمربع التغاير وهي :

$$\left[V\,(r)\right]^2 = \frac{M-m}{M}\;\frac{B^2}{m} + \frac{\overline{N}-\overline{n}}{\overline{\overline{N}}}\;\frac{W^2}{m\overline{\overline{n}}} \qquad (12.2.3)$$

### 3.3.4. مقدرات التباينات من العينة :

نظرا لأننا نجهل بعض معلمات المجتمع فإننا نسعى إلى تقديرها فقط من العينة. وتقدير تباين مجموع المجتمع $Var\left(\hat{Y}\right)$ (العلاقة 1.2.3) المتسق(*) هو :

$$S'^2\left(\hat{Y}\right) = \frac{\sum_{i}^{m'} \left(\hat{y}_i - \hat{y}\right)^2}{m\,(m'-1)} \qquad (1.3.3)$$

حيث :

$$\left.\begin{aligned} \hat{y}_i &= M\,\frac{N_i}{n_i}\sum_{j}^{n_i} y_{ij} = M\,\hat{Y}_i \\ \hat{y} &= \frac{\sum \hat{y}_i}{m'} \end{aligned}\right\} \qquad (2.3.3)$$

حيث يمثل $m'$ عدد العناقيد المسحوبة بطريقة المعاينة العشوائية البسيطة من أجل تقدير التباين.

ويمثل $\hat{Y}_i$ مقدر مجموع العنقود.

مقدر مربع تغاير مجموع المجتمع $\left[V\left(\hat{Y}\right)\right]^2$ [ علاقة 7.2.3 ] هو:

* من أجل البرهان ارجع إلى المراجع المتخصصة: .Madow, Hanrsen Vol II. Sec 3. P. 153,

$$\left[v\left(\hat{Y}\right)\right]^2 = \frac{S'^2_{\left(\hat{Y}\right)}}{\hat{Y}^2}$$

وهو كذلك مقدر متسقا.

أما بالنسبة لمقدر المتوسط r فإن مقدر مربع تغاير المتوسط من العينة هو :

$$v^2(r) = \frac{S'^2_{\hat{Y}}}{\hat{Y}^2} + \frac{S'^2_{\hat{X}}}{\hat{X}} - 2\frac{S'_{\hat{X}\hat{Y}}}{\hat{X}\hat{Y}} = \frac{1}{\hat{Y}}\ \frac{M^2}{m}\ S^2 \qquad (8.2.3)$$

حيث :

$$S^2 = S_Y^2 + r^2 S_X^2 - 2rs_{XY} \qquad (9.2.3)$$

إذا كانت نسبة الإختيار واحدة في جميع الطبقات و كانت $m' = m$ يصبح المقـــدر $v_r^2$كما يلي :

$$v_r^2 = \frac{S_{CY}^2}{m\bar{y}^2} + \frac{S_{CX}^2}{m\bar{x}^2} - 2\,\frac{S_{CXY}}{m\bar{y}\bar{x}} \qquad (10.2.3)$$

حيث :

$$S_{CY}^2 = \frac{\sum\left(y_i - \bar{y}\right)^2}{m-1} \qquad (11.2.3)$$

يمثل $y_i$ مجموع وحدات العينة من العنقود $i$، ويمثل $\frac{y}{m} = \bar{y}$ متوســـط مجمـــوع وحدات العينة في كل عنقود.

ويمكن أن نتحصل على مقدر لــ $S^2$ تباين الوحدات فـــي المجتمـــع، مـــن العينـــة العنقودية، كما يلي :

$$s^2 = \frac{\sum^{m}\sum^{n_i}\left(y_{ij} - \bar{\bar{y}}\right)}{n-1} \qquad (12.2.3)$$

ولكن يعتبر هذا المقدر متحيزا نوعا ما، وينقص هذا التحيز كلما كان عدد العناقيـــد كبيرا في العينة.

## 4.4. أثر المعاينة العنقودية :

حتى نتمكن من معرفة مدى دقة التقديرات المتحصل عليهـا باسـتعمال أسـلوب المعاينة العنقودية البسيطة، يجب علينا مقارنتها بدقة تقديرات أسـلوب أو أسـاليب أخـرى للمعاينة. المعاينة العشوائـية البسيطة أحسن وأبسط أسلوب يمكننا من فعل ذلك.

لذا يمكن أم نكتب صيغة التباين لأحد المقدرات بدلالة تبـاين المعاينـة العشـوائية البسيطة لنفس المقدر.

ولقد عرض كـل مـن Cochran (1977) و Hansen Heurwitz end Madow (1993) الصيغة التالية لتباين مقدر المتوسط.

$$Var\left(\overline{\overline{y}}\right)=Var\left(r\right)=\frac{1-f}{m\overline{n}}S^{2}\left[1+\rho\left(\overline{n}-1\right)\right] \qquad (1.4.4)$$

يستعمل Cochran الرمز $\overline{\overline{y}}$ ، في حين يسـتعمل Hansen ,Hurwitz end Madow الرمز r للدلالة على مقدر المتوسط. يمثل $S^2$ تباين الوحدات في المجتمع، كما يرمز كل من $f$ إلى معدل الاختيار، و m إلى عدد العناقيد المسحوبة، و $\overline{n}$ إلى متوسط عدد الوحدات في كل عنقود. وكذلك :

$$S^{2}=\frac{\sum_{i=1}^{M}\sum_{j=1}^{\overline{N}}\left(Y_{ij}-\overline{\overline{Y}}\right)^{2}}{\left(N-1\right)} \qquad (2.4.4)$$

الذي يمثل تباين الوحدات في المجتمع، حسب الخاصية المدروسة.

$\overline{\overline{y}}$ و $Y_{ij}$ و N معلمات ثم شرحها سابقا. أما ρ فهو معامـل الارتبـاط بـين الوحدات داخل العناقيد، وهو يعتبر مقياسا للتجانس بين الوحـدات. فـي معاينـة عشـوائية عنقودية على مرحلة أو مرحلتين تعطى صيغة هذا المعامل تعريفا كما يلي :

$$\rho=\frac{\frac{M-1}{M}B^{2}-\frac{S^{2}}{\overline{N}}}{\left(\overline{N}-1\right)S^{2}/\overline{N}} \qquad (3.4.4)$$

تصبح هذه الصيغة في حالة عناقيد متساوية الحجم كما يلي :

$$\rho=\frac{\sigma_{b}^{2}-\sigma^{2}/\overline{N}}{\left(\overline{N}-1\right)\sigma^{2}/\overline{N}} \qquad (4.4.4)$$

حيث يمثل $\sigma_b^2$ التباين بين متوسطات العناقيد، و $\sigma^2$ تباين وحدات المجتمع بمقـام يساوي N بدلا من (N-1).

يمكن أن يأخذ معامل الارتباط ρ عدة قيم، أقصاها هي 1، أي ارتباط وتجانس تام بين الوحدات في كل عنقود، وهذا يعني أن كل الوحدات في العنقود متساوية ومتشابهة حسب الخاصية المدروسة. يأخذ ρ أدنى قيمة له وهي $\frac{-1}{\bar{N}-1}$ إذا كانت المتوسطات في جميع العناقيد متساوية (أي $0=\sigma_b^2$)، وهذا يعني أن الوحدات غير متجانسة تماما ضمن العناقيد. ولقد عرض Hansen end Hurwitz ، Madow (1993) تحليلا دقيقا حول مختلف القيم التي يمكن أن يأخذها هذا المعامل ρ.

بالعودة إلى الصيغة (1.4.4) يمكن أن نلاحظ بأن تباين المتوسط يحتوي على حدين الأول $\frac{(1-f)S^2}{m\bar{n}}$، والثاني $1+\rho(\bar{n}-1)$. يمثل الحد الأول تباين مقدر المتوسط باستعمال المعاينة العشوائية البسيطة. وبالتالي يمثل الحد الثاني المعامل الذي يجب أن نضربه في تباين المعاينة العشوائية البسيطة حتى نتحصل على تباين المعاينة العنقودية. يسمى KISH[18] هذا المعامل بأثر التصميم Deff. إذا كان $\bar{n}=1$، أي أننا نسحب من كل عنقود وحدة واحدة فإن تباين المتوسط في المعاينة العنقودية هو نفسه تباين المتوسط في المعاينة العشوائية البسيطة. ويمكن أن نستنتج كذلك أنه كلما ارتفع عدد الوحدات $\bar{n}$ في كل عنقود كلما كان هذا التباين Var ( r ) أكثر ما دام معامل الارتباط موجب. كما أنه إذا كانت الوحدات في العناقيد متجانسة ( ρ موجب ويقترب من الواحد) كان التباين كبيرا.

[18] Leslie Kish, Survey Sampling, London . John Wiley and son, 1965, p 162

## 4-5 دالة التكاليف في أسلوب المعاينة العنقودية :

### *1.5.4. الشكل العام لدالة التكلفة :*

تحتل مسألة دراسة وتحديد التكاليف الخاصة بالمسوحات الإحصائية بالمعاينة مكانة جد هامة وجوهرية خاصة إذا كانت هذه المسوحات واسعة المدى بحيث تشمل مناطق جغرافية كبيرة. إن عملية حساب القيم المثلى للوحدات المسحوبة أو التباينات تقتضي التوصل إلى إيجاد دوال تكلفة جيدة ومفصلة تحدد بحذر مختلف عناصر التكلفة وكيفية دخولها في التكاليف الإجمالية وكذا تغيرها مع تغير أسلوب المعاينة المستعمل. في كثير من الأحيان، تكون التباينات الناتجة عن استعمال أسلوب المعاينة العنقودية أكبر من مثيلاتها في المعاينة العشوائية البسيطة، ولكن من ناحية أخرى يحقق أسلوب المعاينة العنقودية تكاليف وحدوية أقل بكثير من التكاليف الوحدوية في المعاينة العشوائية البسيطة. لذا نلجأ إلى استعمال العنقدة(*) إذا كان تأثير التكاليف المتدنية أكبر من تأثير زيادة تباينات المقدرات.

سوف نعرض في هذه الفقرة تدريجيا شكلا عاما لدالة التكاليف، ثم نتطرق إلى صيغة مبسطة لدالة التكلفة، ثم بعد ذلك صيغة أكثر تعقيدا تأخذ بعين الاعتبار كل العناصر التي يمكن أن تدخل في التكلفة الإجمالية للمسح الإحصائي بالمعاينة.

سوف نحاول تبسيط المقارنة بين مختلف الأساليب وذلك بعرض صيغة عامة لدالة التكلفة والتي يمكن تطبيقها على مختلف الأساليب المستعملة. تقسيمنا الموالي لعناصر التكلفة يمكن أن يكون ناقصا إلا أنه يساعد على فهم تقسيمات أخرى أكثر تعقيدا. وعليه، يمكن أن نقسم التكلفة الإجمالية C إلى أربعة أقسام أساسية، كل واحد منها يحتوي على عدة عناصر مختلفة :

$$. \ \mathbf{C = C_F + C_V} \qquad \mathbf{(1.5.4)}$$

1- يمثل $C_F$ قسم التكاليف الثابتة (couts fixes) التي لا تتغير بتغير عدد الوحدات الإحصائية المسحوبة. تنقسم هذه التكاليف بدورها إلى قسمين. القسم الأول منها هو التكاليف الثابتة التي لا ترتبط بأسلوب المعاينة المستعمل وهي تشمل كلا من : مناقشة وتحديد أهداف الدراسة، و إعداد الاستمارة، و تحليل النتائج والمعطيات، و تكاليف تحضير وطبع التقرير

* عندقة : المعاينة العنقودية.

النهائي للدراسة. أما القسم الثاني منها، فهو يخص التكاليف الثابتة التي ترتبط مباشرة بأسلوب المعاينة المستعمل وهي تشمل كل من : حساب مقدرات العينة وتبايناتها، و بعض أعمال مصلحة المعاينة مثل شراء معدات المعاينة كالخرائط، وكذا تحضير التعليمات،و تكوين وتدريب المستجوبين أو كما يسمون عادة بالعدادين لإنجاز الدراسة على أحسن وجه.

2- يمثل $C_V$ ، التكاليف المتغيرة (couts variables) التي ترتبط مباشـرة بعـدد الوحدات المسحوبة في العينة وهي بدورها تنقسم إلى قسمين أساسيين. يشمل القسم الأول، التكاليف الخاصة بعدد وحدات العينة، والذي لا يتغير بتغير أسلوب المعاينة المستعمل. تتضمن هذه التكاليف كل من : تسجيل المشاهدات والاستجوابات، و ترميز وإدخال هذه المشاهدات في الحاسب أو آلة أخرى، و بعض الأعمال الإضافية في مصلحة المعاينة. أما القسم الثاني فهو يخص التكاليف التي تتغير بتغير كل من عدد الوحدات في العينة وأسلوب المعاينة المستعمل، وهي تشمل كل من : تكاليف التنقل بين الوحدات وتحديدها، و تكاليف تحضير معدات المعاينة وكيفية السحب، و تكاليف التأطير الجيد وهي تختلف كذلك من أسلوب إلى آخر. تجدر الملاحظة إلا أنه يمكن أن نكون قد أهملنا بعض العناصر ولكن سوف نتطرق إليها بالتفصيل في الفقرات المقبلة.

### ***2.5.4. دالة تكلفة مبسطة خاصة بالمعاينة العنقودية :***

لقد كانت الصيغة (1.5.4) تعبر فقط عن الشكل العام لدالة التكاليف. فهي لا تتطرق بالتفصيل لكيفية دخول كل عنصر تكلفة في مجال خاص به. لذا سوف نعرض في هذه الفقرة دالة تكلفة بسيطة ولكنها أثبتت فعاليتها في كثير من البحوث والدراسات. في هذه الدالة المبسطة تعطى صيغة التكلفة الإجمالية C كما يلي :

$$C=C_F+C_1 m+C_2 m\bar{n} \quad \mathbf{(2.5.4)}$$

يمثل m عدد عناقيد العينة، و $C_1$ تكلفة معاينة كل عنقود، وهي تضم تكاليف الانتقال، وتحديد العناقيد، وتكاليف تحضير قوائم الوحدات الإحصائية في كل عنقود، بالإضافة إلى تكاليف الاتصال بالجماعات المحلية بهدف الحصول على القوائم وعلى المعلومات الخاصة بالسكان في منطقة ما في حالة الدراسات التي تخص الأفراد كوحدات دراسة، وبالتالي تشمل $m\bar{n}$عدد وحدات العينة. أما $C_2$ فهي تمثل التكلفة الخاصة بكل وحدة

إحصائية مسحوبة. تضم هذه التكلفة العناصر التي تطرقنا إليها سابقا لما تحدثنا عن التكلفة ا لمتغيرة $C_V$ في الصيغة (1.5.4).

### 3.5.4. دالة تكاليف أكثر تعقيدا وشمولية :

في العديد من الحالات، يمكن أن نلاحظ بأن العلاقة المبسطة لدالة التكـاليف بعيـدة جدا عن الحقيقة. فمثلا إذا كان لدينا عدد كبير من العناقيد منتشرة في مجـال معـين، فـإن تكاليف التنقل والعمل الميداني بين العناقيد سوف تكون كبيرة كلما كانت هذه العناقيد متشـتتة ومتباعدة فيما بينها، وتقل هذه التكاليف كلما كانت العناقيد متمركزة ومتقاربة. فإذا طلب منـا مثلا معاينة 500 أسرة بهدف دراسة سلوكياتهم الاستهلاكية، وكانت هذه الأسر موزعة على 10 أحياء في مدينة ما بحيث يحتوي كل حي على 50 أسرة. فإنه مـن البـديهي أن تكـون تكاليف التنقل بين هذه الأحياء وبين المنازل في كل حي أقل بكثير من تكاليف التنقل في حالة ما إذا كانت تلك المنازل موزعة على 100 حي، يحتوي كل واحد منها على 5 منازل فقط. نظرا لأهميتها في مثل هذه الحالات، سوف نعرض في هذه الفقرة تكاليف التنقل بشكل أكثـر تفصيلا. تعطى التكلفة الإجمالية C بعد إدماج تكاليف التنقل كما يلي :

$$C = C_F + C_0\sqrt{m} + C_1 m + C_2 m\bar{n} \qquad \textbf{(3.5.4)}$$

"تمثل $C_0\sqrt{m}$ تكلفة التنقل من عنقود إلى آخر. وتبين الاختبارات علـى خريطـة أن هذه التكلفة تتغير، في حالة مجتمع مثبت، وفقا للجذر التربيعي لعدد العناقيد تقريبا"[19].

يمثل $C_0$ معامل $\sqrt{m}$ الذي يمكن أن يتغير حسب المسـاحة التـي تشـتمل عليهـا الدراسة، وكذا كيفية توزيع العناقيد فيها، وعدد العناقيد في العينـة، وكـذا وسـائل النقـل المتوفرة، بالإضافة إلى عدة عوامل أخرى. يجب على المعاين أن يحدد متوسط المسافة بين العناقيد، والمساحة الكلية التي سوف يقطعها لإنجاز عمله. ثم يحدد بدقـة تكلفـة الكيلـومتر الواحد ويضربها في المسافة الكلية للتنقل.

إذا كانت العناقيد تتوزع بتساوي في المساحة المدروسة. أي أنها متباعدة فيما بينهـا بنفس المسافة، فإن هذه الأخيرة سوف تكون مساوية لـــ $d=\sqrt{\frac{A}{m}}$ ([20])، حيـث تمثـل A المساحة المدروسة بالكم المربع. وتكون المسافة الإجمالية المقطوعة هي $\sqrt{A}\sqrt{m}=md$.

[19] "كوكران" 1995 .ص 352. مرجع سبق ذكره.

[20] Hansen, Hurwitz and Madow, Sample survey Methodes and theory,Vol I ch06 p275 .Wiley

تضم تكلفة الكلم الواحد كلا من : تكلفة استعمال السيارة الخاصة، أو سيارة أجـــرة أو أي وسيلة نقل أخرى، وكذا تكلفة عمل العداد (المعاين) في الساعة الواحـــدة، وتكـــاليف أخرى. فإذا رمزنا لكل هذه العناصر بـــ K ، فإن $C_0$ تصبح تساوي $C_0 = K.\sqrt{A}$. يمكن أن تضم $C_0$ عدة تكاليف تختلف حسب نوعية وظروف العمل والتي يمكن أن يضيفها المعـــاين بشكل أكثر دقة. الحقيقة أن كل ما ذكرناه سابقا هو عبارة عن مجرد تقريب فقط للواقـــع. إذ يمكن في كثير من الأحيان أن لا تتوزع العناقيد بتساوي وتجانس في المســـاحة المدروســـة، كما يمكن أن لا ينتقل المعاين من عنقود إلى آخر، وإنما ينتقل من مكتب أو مصلحة المعاينة إلى العنقود مباشرة ثم يعود في اليوم الموالي وينتقل من المصلحة إلى العنقود الثاني ويرجع بعد انتهاء عمله في هذا العنقود وهكذا دواليك حتى ينتهي من جميع العناقيد. في هذه الحالة، يجب تحديد المسافة التي تفصل بين كل عنقود والمصلحة. ورغم كل هذه التعقيدات، ســـوف نستعمل نحن في فقراتنا الموالية كلا الصيغتان (2.5.4)، و(3.5.4) من أجـــل حســـاب القـــيم المثلى لأحجام العينات وكذا التباينات المثلى الخاصة بالمعاينة العنقودية.

**6.4. تحديد حجم العينة الأمثل في المعاينة العنقودية البسيطة على مرحلة أو على مرحلتين :**

***<u>4-6-1- تحديد الحجم الأمثل لعدد العناقيد m وعدد الوحدات في كل عنقود n باستعمال دالـــة تكلفة بسيطة :</u>***

كما سبق وأن رأينا في الفصول السابقة، فإننا نسعى دائما من خلال بحثنا إلى إيجاد القيم المثلى لحجم العينة المسحوبة علما أن ميزانية الدراسة (تكاليف الدراسة) مثبتة. أو إلـــى تدنية تكاليف البحث عند تثبيت حجم العينة. في المعاينة العنقودية البسيطة، يقتضـــي تحديـــد حجم العينة تحديد كل من عدد العناقيد m ومتوسط عدد الوحدات في كل عنقود $\bar{n}$ لأن حجم العينة هو $n = m.\bar{n}$ .

بانتهاج نفس الطريقة المستعملة في الفصل السابق، يمكننا بسهولة البرهنة على أنـــه إذا كانت تكاليف البحث الإجمالية معطاة كما في الصيغة (2.5.4) وهي:

$$C = C_F + C_1 m + C_2 m\bar{n} \quad . \qquad (1.6.4)$$

وكان التباين معطى كما في العلاقة (10.2.3)، فإن الحجم الأمثل لمتوسط عدد الوحدات في كل عنقود والحجم الأمثل لعدد العناقيد m في العينة إذا كانت التكلفة الإجمالية مثبتة هو :

$$\bar{n}_{opt}=\bar{N}\sqrt{\frac{C_1}{C_2}\frac{S_2^2}{S_1^2-\bar{N}S_2^2}}=\sqrt{\frac{C_1}{C_2}\frac{W^2}{B^2-W^2/\bar{N}}}=\sqrt{\frac{C_1}{C_2}\frac{1-\rho}{\rho}} \qquad \textbf{(2.6.4)}$$

و

$$m_{opt}=\frac{C-C_F}{C_1+C_2\bar{n}} \qquad \textbf{(3.6.4)}$$

حيث يمثل ρ معامل الارتباط كما هو معطى بالعلاقة (3.4.4) فقرة 4.4 أما المعالم المتبقية فقد تم شرحها من قبل.

**<u>برهان :</u>**

خلافا للفصول السابقة، سوف نستعمل في برهاننا هذا صيغة مربع تغاير المتوسط $V^2(r)$ بدلا ن صيغة التباين Var ( r ) والتي هي معطاة كما في العلاقة (10.2.3) أو (12.2.3) وهي :

$$V^2(r)=\left(\frac{M-m}{M}\right)\frac{B^2}{m}+\left(\frac{\bar{N}-\bar{n}}{\bar{N}}\right)\frac{W^2}{m\bar{n}} \qquad \text{(I)}$$

من أجل تحديد القيم المثلى لـ n و m والتي تجعل $V^2$ ( r ) أصغر ما يمكن (أحسن دقة) من أجل تكاليف إجمالية ثابتة نستعين كما فعلنا ذلك سابقا بدالة لاغرانج F.

تحت هذه الشروط السابقة تعطى دالة F بالعلاقة التالية :

$$F=V^2(r)+\lambda\left[C_1m+C_2m\bar{n}-(C-C_F)\right] \qquad \text{... (II)}$$

من أجل إيجاد القيم المثلى، فإنه يتعين علينا جعل المشتقات الجزئية لكل من m و $\bar{n}$ مساوية للصفر، وهي كالآتي :

$$\frac{\partial F}{\partial m}=-\frac{B^2}{m^2}-\frac{W^2}{m^2\bar{n}}+\frac{W^2}{m^2\bar{N}}+\lambda\left(C_1+C_2\bar{n}\right)=0 \qquad \text{(III)}$$

و

$$\frac{\partial F}{\partial \bar{n}}=-\frac{W^2}{m\bar{n}^2}+\lambda C_2m=0 \quad \text{...} \qquad \text{(IV)}$$

بعد حل جملة المعادلات هذه، نحصل على القيمة المثلى لـــ $\bar{n}$ كمـــا هـــي معطـــاة بالعلاقة (2.4.6)، ثم نحصل على القيمة المثلى لعدد العناقيد m كما هـــي معطـــاة بالعلاقـــة (3.4.6) بتعويض $\overline{n_{opt}}$ في دالة التكلفة.

تساعدنا الصيغ الرياضية للقيم المثلى في استنتاج وملاحظة كيفية تغير هـــذه القـــيم بتغير أحد مكوناتها ملاحظة مباشرة، فمن خلال الصيغة (2.4.6). يمكن أن نستنتج أنه كلمـــا كانت التكلفة $C_1$ كبيرة بالنسبة للتكلفة $C_2$ فإن عدد الوحدات المسحوبة من كل عنقـــود $(\bar{n}_{opt})$ يزداد، في حين ينخفض عدد العناقيد المسحوبة ضمن العينة، وكلما كانت التكلفة $C_1$ صغيرة بالنسبة لـــ $C_2$فإن $\overline{n_{opt}}$ ينخفض وبالتالي يزداد عند العناقيد $m_{opt}$ المسحوبة.

تجدر الإشارة إلى أن القيمة المثلى لـــ $\bar{n}$ تتغير بتغير الجـــذر التربيعـــي لـــــ $\frac{C_1}{C_2}$ وبالتالي قإن $\bar{n}_{opt}$ لا تتأثر كثيرا بالتغيرات الطفيفة في $\sqrt{\frac{C_1}{C_2}}$ وإنما تتأثر فقـــط إذا كـــان هـــذا التغيركبيرا نوعا ما. نلاحظ كذلك في الصيغة (2.4.6) أن $\bar{n}_{opt}$ معبر عنـــدما بدلالـــة $\sqrt{\frac{1-\rho}{\rho}}$ وهذا يعني أنه كلما كان معامل الارتباط ρ كبيرا، أي أن الوحدات ضمن العناقيد متجانســـة فيما بينها، كان عدد الوحدات المسحوبة في كل عنقود $\bar{n}_{opt}$ صغيرا ، وبالتالي يـــزداد عـــدد العناقيد $m_{opt}$ المسحوبة في العينة والعكس صحيح. يمكن أن نلاحظ كذلك بأن $\bar{n}_{opt}$ لا تتغيـــر كثيرا إذا كان التغير في $\sqrt{\frac{1-\rho}{\rho}}$ طفيفا.

و فيما يخص الصيغة (3.6.4) يمكن أن نلاحظ أنه كلما كانت ميزانية البحث كبيرة، كان عدد العناقيد $m_{opt}$ كبيرا وكلما قلت قل معها عدد هذه العناقيد، كما يمكن أن نلاحظ كذلك أنه كلما كانت التكلفة $C_1$كبيرة قل $m_{opt}$; وبالتالي يرتفع $\bar{n}_{opt}$ .

تقود هذه الحقائق كلها إلى استنتاج أن الحجم الأمثل للعناقيد المســـحوبة $\bar{n}_{opt}$ يكـــون كبيرا عندما:

- تنخفض تكلفة المقابلة $C_2$ (تكلفة سحب وحدة إضافية).
- ترتفع التكلفة $C_1$لمعاينة عنقود إضافي.
- تنخفض قيمة المعامل ρ للارتباط، أي عندما تكون الوحدات ضـــمن العنقـــود غير متجانسة.

- متوسط عدد الوحدات $\overline{N}$ في كل عنقود كبيرا.

ويكون العدد الأمثل $m_{opt}$ لعدد العناقيد كبيرا إذا كانت:

- التكلفة $C_1$ لمعاينة عنقود إضافي قليلة.
- التكلفة الإجمالية للدراسة كبيرة.
- معامل الارتباط $\rho$ كبيرا.

مهما كانت النتائج التي سوف نتوصل إليها باستعمال تقديراتنا للقيم المثلى لكل من $\overline{n}_{opt}$ و $m_{opt}$ فإنها لا يمكن أبدا أن تكون نتائج حقيقية مثلى وإنما نحن نسعى إلى التقرب من مجال الأمثلية، ذلك لأنها لا يمكننا معرفة تكاليف الدراسة بدقة أو حتى التباينات والقياسات الأخرى للتجانس قبل إجراء المعاينة.

إن الاستنتاجات التي توصلنا إليها حتى الآن فيما يخص تحديد القيم المثلى لـــ $\overline{n}_{opt}$ و $m_{opt}$ تنطبق فقط في حالة ما إذا كانت دالة التكلفة بسيطة، في الواقع توجد عدة دوال أخرى أشمل و أكثر واقعية من دالة التكلفة التي عرضناها في تحديد القيم المثلى. سوف نحاول استعمال الصيغة (3.5.4) بهدف الحصول على نتائج أحسن وتقديرات أدق لكل من عدد الوحدات المثلى في كل عنقود وكذا العدد الأمثل للعناقيد المسحوبة.

## 4-6-2 تحديد الحجم الأمثل لـــ $\overline{n}$ و m في حالة دالة تكلفة أكثر شمولية:

في معاينة عشوائية عنقودية بسيطة، إذا كانت التكلفة معطاة كما في الصيغة (3.5.4) وهي:

$$C=C_F+C_0\sqrt{m}+C_1m+C_2m\overline{n} \qquad (4.6.4)$$

وكان التباين معطى كما في العلاقة (10.2.3) فإن الحجم الأمثل للعناقيد المسحوبة $\overline{n}_{opt}$ هو:

$$\overline{n}_{opt}=\sqrt{\frac{\frac{C_0}{a}+C_1}{C_2}\frac{W_2}{B_2-W^2/\overline{N}}}=\sqrt{\frac{\frac{C_0}{a}+C_1}{C_2}\frac{1-\rho}{\rho}} \qquad (5.6.4)$$

وذلك بهدف تدنية التباين علما أن التكلفة الإجمالية محددة. في هذه الحالة يكون

$$m_{opt}=\frac{1}{4}a^2 \qquad (6.6.4)$$

$$a=\frac{\sqrt{1+4\frac{C}{C_0}\frac{C_1+C_2\bar{n}}{C_0}}}{\frac{C_1+C_2\bar{n}}{C_0}} \qquad (7.6.4)$$

أما في حالة تدنية التكلفة الإجمالية لما يكون التباين مثبت عند قيمة معينة $\varepsilon$، فإن الصيغة $a$ تصبح كما يلي:

$$a=\sqrt{4\frac{B^2-\frac{W^2}{N}+\frac{W^2}{\bar{n}}}{\varepsilon+\frac{B^2}{M}}}=\sqrt{4\frac{\hat{V}^2\rho}{\varepsilon}\left(1+\frac{1-\rho}{\rho}\frac{1}{\bar{n}}\right)} \qquad (8.6.4)$$

حيث يمثل $\hat{V}^2$ مربع تغاير وحدات المجتمع والذي يساوي $\frac{S^2}{\bar{Y}^2}$ .

نحصل على الحل الأمثل لـ $\bar{n}$ و m عندما تكون التكلفة مثبتة بتعويض قيمة افتراضية لـ $a$ في العلاقة (5.6.4) ثم تعوض النتيجة المحصل عليها لـ $\bar{n}$ في العلاقة (7.6.4)، ولتكن مثلا $\bar{n}_1$ حتى نحصل على $a_1$، ثم نعوض $a_1$ في العلاقة (5.6.4)، فنحصل على $\bar{n}_2$ التي بدورها نعوضها في العلاقة (7.6.4) فنحصل على $a_2$ ثم نعوض هذه القيمة الجديدة في العلاقة (5.6.4) وهكذا دواليك حتى نحصل على حلول متتابعة لـ $\bar{n}$ و $a$ متساوية تقريبا، ولا يتعدى الفرق بينهما أكثر من 2%.[21]

بعدها نعوض القيمة الأخيرة لـ a في العلاقة (6.6.4) من أجل حساب القيمة المثلى لـ m.

بنفس الكيفية هذه يمكن أن نجد القيم المثلى إذا كانت الدقة مثبتة، باستعمال الصيغة (8.6.4) بدلا من الصيغة (7.6.4) واتباع نفس المراحل المذكورة أعلاه.

**<u>برهان:</u>**

من أجل تحديد القيم المثلى لـ $\bar{n}$ و m والتي تدني الدقة كما في الصيغة (10.2.3) علما أن التكلفة الإجمالية المعطاة بالصيغة (3.5.4) مثبتة، نشكل دالة لاغرانج F.

[21] Hansen, Hurwitz and Madow, Sample survey Methodes and theory,Vol I ch06 p291 .Wiley

$$F = V^2(r) + \lambda\left[C_0\sqrt{m} + C_1 m + C_2 m\bar{n} - (C - C_F)\right]\text{..............}(\mathbf{I})$$

أما إذا كنا نهدف إلى تدنية التكلفة الإجمالية، علما أن الدقة مثبتة عند القيمـــة $V^2(r)$ فإن دالة لاغرانج تصبح كما يلي:

$$F' = \alpha\left(V_{(r)}^2 - V_{(r)}^{2*}\right) + C' \text{..............}(\mathbf{II})$$

بحيث يمثل $C'$ الفرق بين التكلفة الإجمالية C والتكلفة الثابتة $C_F$ أي:

$$C' = C - C_F$$

المشتقات الجزئية لكل من $\bar{n}$ و m متساوية في كلا الحالتين، وعليه:

$$\frac{\partial F}{\partial \bar{n}} = \frac{\partial F'}{\partial \bar{n}} = -\frac{W^2}{m\bar{n}^2} + \lambda . C_2 m = 0$$

...........III

مع وضع

$$\lambda = \frac{1}{\alpha} \qquad (1\text{-III})$$

$$\frac{\partial F}{\partial m} = \frac{\partial F'}{\partial m} = -\frac{B^2}{m^2} - \frac{W^2}{m^2\bar{n}} + \frac{W^2}{m^2\bar{N}} + \lambda\left(\frac{C_0}{2\sqrt{m}} + C_1 + C_2\bar{n}\right) \qquad (\text{IV})$$

من العلاقة III نحصل على:

$$\lambda m^2 = \frac{W^2}{C_2\bar{n}^2} \qquad (2\text{-III})$$

يمكن أن نحصل على نفس الطرف الأيسر للعلاقة (2-III) وذلك بضرب العلاقة III بـــ $\bar{n}$ وطرح النتيجة من حاصل ضرب العلاقة (1-III) بـــ m فنحصل على :

$$\lambda m^2 = \frac{B^2 - W^2/\bar{N}}{(C_0/2\sqrt{m}) + C_1} \qquad (3\text{-III})$$

بعد جعل مساواة بين العلاقتين (2-III) و(3-III) نحل هذه المساواة من أجـــل $\bar{n}$ ثـــم نعوض $m=\frac{a^2}{4}$ فنحصل على القيمة المثلى لـــ $\bar{n}$ كما هي معطاة في العلاقة (5.6.4) ونحصل على قيمة $a$ كما هي معطاة في العلاقة (7.6.4) في الحالة التي تكون فيها التكلفة الإجماليـــة

مثبتة، كما نحصل على $a$ كما هي معطاة في العلاقة الثانية (8.6.4) في الحالة التي يكون فيها الدقة مثبتة ومساوية لـ $V_{(\bar{r})}^{2*}$ ونستخرج صيغة m ثم نحصل على $a$ مساوية لـ $a=\sqrt{4m}$ .

نعوض هذه الأخيرة في صيغة القيمة المثلى لـ $\bar{n}$ (5.6.4) فنحصل على :

$$\bar{n}_{opt} = \sqrt{\frac{\frac{C_0}{2\sqrt{m_{opt}}} + C_1}{C_2} \frac{1-\rho}{\rho}} \qquad (9.6.4)$$

نلاحظ من خلال هذه الصيغة أن كل من $C_1$ و$C_2$ و $\rho$ يؤثرون على $\bar{n}_{otp}$ بنفس الكيفية التي رأيناها في حالة دالة تكلفة بسيطة لا تحتوي على تكاليف التنقل أي على $C_0\sqrt{m}$، كما يمكن أن نلاحظ بأن للمعاملان $C_1$، $\frac{C_0}{2\sqrt{m_{opt}}}$ نفس اتجاه التأثير على $\bar{n}_{opt}$، أي كلما كان $\frac{C_0}{2\sqrt{m_{opt}}}$ كبيرا كلما كان $\bar{n}_{opt}$ أكبر، والعكس صحيح، هذا من جهة، ومن جهة أخرى نلاحظ أن تأثير واحد وحدة نقدية من $C_1$ أكبر من تأثير واحد وحدة نقدية إضافية من $C_0$(تكلفة التنقل بين العناقيد) لأن هذه الأخيرة مقسومة على $2\sqrt{m}$، الأمر الذي ينقص من مفعولها في التأثير على $\bar{n}_{opt}$.

في العموم، يمكن أن نقول بأنه إذا كانت تكاليف التنقل بين العناقيد كبيرة، فإن عدد الوحدات الأمثل في كل عنقود سوف يكون أكبر وبالتالي القيمة المثلى لعدد العناقيد $m_{opt}$ تنقص، ولكن تأثير التغيير في تكاليف التنقل ليس تأثيرا كبيرا على $\bar{n}_{opt}$. إلا في الحالات التي تكون فيها تكاليف التنقل تمثل نسبة كبيرة من مجمل التكاليف، تجدر الإشارة إلى أنه في حالة التكاليف البسيطة، لم يكن للميزانية الإجمالية للبحث أي تأثير على $\bar{n}_{opt}$ ولكن بعد إدماج تكاليف التنقل في الصيغة العامة لدالة التكاليف من خلال المعامل $C_0\sqrt{m}$ أصبح للميزانية الإجمالية تأثير معتبر على القيم المثلى.

# الفصل الثالث:

## تحديد حجم العينة الضروري لإجراء دراسة حول المؤسسات الصغيرة و المتوسطة في ولاية وهران

## دراسة حالة حول المؤسسات التابعة لحي سيدي البشير

## مقدمة

سوف نقوم في هذا الفصل بتحديد حجم العينة الضروري للقيام بدراسة حول المؤسسات الصغيرة و المتوسطة، والتي تهدف إلى تقدير معالم بعض المتغيرات كرقم الأعمال، الربح، والقيمة المضافة، كما نحاول كذلك توضيح مدى تأثر حجم العينة بالدقة المرغوب بلوغها، ومدى تأثير حجم العينة على دقة النتائج المحصل عليها.

لا شك أن المسح بالمعاينة هو الوسيلة المفضلة بل و الأساسية المعتمدة من قبل أغلبية الباحثين بهدف دراسة المؤسسات الاقتصادية من مختلف الجوانب كما و كيفا. كما أن أول خطوة يتعين على الباحث تجاوزها هي مسألة تحديد حجم العينة الضروري لإجراء دراسته. إذ تعتبر هذه المسألة جد هامة و ضرورية لأنه إذا كانت العينة أصغر من اللازم فإن نتائج الدراسة تكون بعيدة عن الحقيقة في الواقع، هذا من ناحية، و من ناحية أخرى، إذا كانت العينة أكبر مما ينبغي فإننا نبدد الوقت و الموارد لأن التكلفة تكون أكبر كلما كان حجم العينة كبيرا.

نود التأكيد في هذا البحث على ضرورة تشجيع الباحثين في الجزائر كل في مجاله من أجل الإكثار من إجراء البحوث الميدانية حول المؤسسات الاقتصادية في العموم و بالأخص الصغيرة و المتوسطة منها بهدف معرفة مدى ملاءمتها مع الظروف الاقتصادية الجديدة، خاصة في ظل العولمة و ظروف انخراط بلادنا في الشراكة الأورو- متوسطية التي تتحكم فيها قوانين المنافسة الحرة و الشرسة. حيث لا يمكننا متابعة هذه المؤسسات إلا من خلال جمع معلومات واقعية و دقيقة نضعها بين يدي الرجل السياسي و الاقتصادي حتى يتسنى له اتخاذ قرارات فعالة. و أساس هذه المعلومات بيانات تنتجها الدوائر و المكاتب المتخصصة في جمع البيانات حيث تشكل بنكا للمعلومات حول أنشطة و محاسبة المؤسسات.

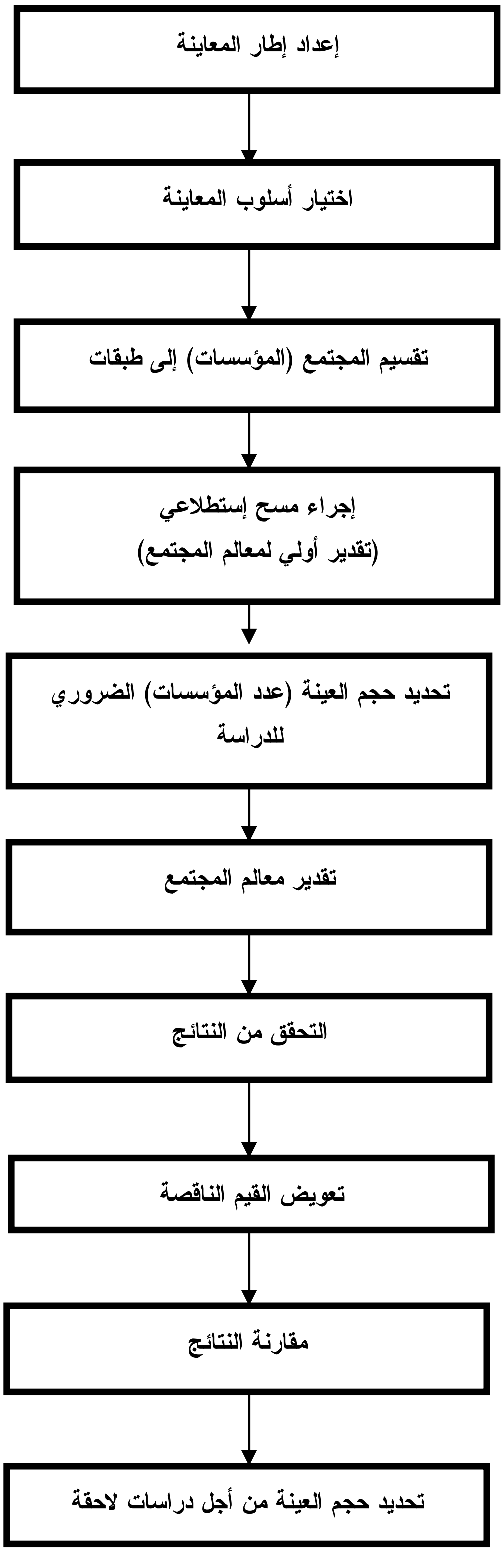
من أجل تنفيذ بحثنا الميداني قمنا باتباع مجموعة من الخطوات الأساسية التي تعكس في مجملها مراحل المنهاج الإحصائي، بغية التوصل إلى نتائج تعكس حقيقة الظاهرة قيد الدراسة، أي تقديد معالم المجتمع الإحصائي بصورة جيدة و فعالة. في الفقرة الأولى نقوم بتعريف الوحدة الإحصائية و المجتمع الإحصائي موضوع الدراسة في الظرفين الزماني و المكاني. كما نعرض في الفقرة الثانية الإمكانات المادية و البشرية المتاحة لدينا . إذ يلعب هذان العاملان دورا أساسيا في تحديد حجم الدراسة. لأنه كلما كانت الموارد متوفرة كلما

أمكننا توسيع مجال البحث ليشمل عددا كبيرا من الوحدات الإحصائية. الفقرة الثالثة خصصت لعرض المراحل الأساسية التي تمر بها عملية تحديد حجم العينة، حيث نتطرق في هذه الفقرة إلى كيفية إنجاز مسح استطلاعي يساعدنا على تقدير تباينات المتغيرات و التي ندرجها في الصيغ الرياضية لحجم العينة. نقوم في الفقرة الرابعة بتقدير معالم المجتمع المدروس. أما الفقرة الخامسة فهي مخصصة لإنجاز حصر شامل للوحدات الإحصائية نتمكن من خلاله من التحقق من النتائج التي تحصلنا عليها. سوف نبين في الفقرة السادسة مدى تأثير حجم العينة على دقة النتائج أما الفقرة الأخيرة فهي مخصصة لتحديد حجم العينة الضروري لإنجاز بحوث لاحقة.

شمل المسح الميداني المؤسسات الصغيرة و المتوسطة المسجلة على مستوى مصلحة الضرائب بمقاطعة سيدي البشير التابعة لبلدية وهران. بلغ عدد هذه المؤسسات المشكلة للمجتمع الإحصائي 454 مؤسسة تجارية و خدمية بمختلف أنواعها. و نظرا لأسباب منهجية مررنا بعدة مراحل، بدءا من مرحلة التحضير و الإعداد مرورا بمرحلة اختيار أسلوب المعاينة الضروري لهذه الدراسة ثم انتقلنا إلى مرحلة تقسيم المجتمع الإحصائي (المؤسسات) إلى طبقات، فإعداد الاستمارة ثم تعيين العدادين و إعدادهم. في مرحلة التنفيذ شرعنا في إجراء مسح استطلاعي حتى نتمكن من أخذ فكرة عن المؤسسات قيد الدراسة و كذا الحصول على بيانات تساعد في تحديد حجم العينة الضروري لإجراء المسح الميداني النهائي. في المرحلة الأخيرة، قمنا بتقدير بعض المعالم و حساب مختلف المقاييس و تقييمها ثم مقارنتها مع النتائج المحصل عليها من خلال إجراء حصر شامل للمؤسسات في المجتمع الإحصائي نهدف من خلالها إلى التحقق من فاعلية أسلوبنا المتبع.

على ضوء ما سبق يعرض المخطط الموالي المراحل المتبعة في إجراء مسحنا الميداني باستعمال أسلوب المعاينة العشوائية الطبقية التناسبية، و الذي يتماشى مع المجتمع قيد الدراسة.

**شكل رقم 01: مراحل المسح الميداني بالمعاينة**




لا يخلو أي بحث ميداني من مشاكل وعراقيل عدة تؤدي إلى إعاقة عمل الباحث و خلط برنامجه الزمني المحدد. في هذا البحث واجهتنا عدة أنواع من هذه المشاكل نذكر من أهمها نقص الإمكانيات المادية و البشرية الذي يؤدي بطريقة مباشرة إلى تقليص حجـــــم

الدراسة. أما المشكلة الثانية فتمثلت في صعوبة تجاوب أصحاب المؤسسات و موظفيهم مع العدادين الميدانيين نظرا لعدة أسباب من بينها:

1- تخوف المستجوبين من إمكانية عمل العداد في أحد مصالح المراقبة التابعة للدولة.

2- عدم اكتراث المستجوب بالبحث إطلاقا.

3-خوف بعض العمال من تحمل المسؤولية في غياب صاحب المؤسسة.

4-جهل المستجوب لبعض أو أغلبية مصطلحات المتغيرات الواجب التصريح بها للعداد.

5-تضييع الاستمارة و الإطالة في إرجاعها للعداد.

كل هذه المشاكل أدت إلى تمديد فترة الدراسة المخطط لها. قصد تجاوز هذه العقبات بهدف إتمام الدراسة الميدانية على الباحث و العدادين التحلي بالصبر و المثابرة من اجل انتزاع البيانات و لو بشق الأنفس مع ما يقتضيه ذلك من تكرار الزيارات لنفس المؤسسات المتعنتة و كذا الإلحاح على أهمية هذا النوع من البحوث و كذا المساهمة الفعالة في نشر ثقافة الاستبيانات (les sondages) لدا أفراد المجتمع و المواطنين.

## 5-1المجتمع الإحصائي و إطاره المكاني والزماني:

### 5-1-1 الوحدة الإحصائية:

تمثل الوحدة الإحصائية في هذه الدراسة المؤسسات التي تمارس نشاط مهني معين سواء كانت فردية أو جماعية، أو مؤسسات ذات الشخص المعنوي أو المادي وهي مقسمة إلي مالي:

1- المؤسسات الفردية

2- شركات ذات مسؤولية محدودة SARL :

3- مؤسسات أحادية الجانب ذات المسؤولية المحدودة (EURL)

4- مؤسسات تضامن SNC

تشمل الدراسة كل مؤسسة أو شركة مسجلة في مصلحة من مصالح الضرائب والحاملة لسجل تجاري مهما كان نشاطها.

### 5-1-2 حدود المجتمع الإحصائي في المكان:

شمل مجال بحثنا كل أشكال المؤسسات الخاصة المذكورة سابقا والمسجلة بمصالح الضرائب لمفتشية "سيدي البشير" ببلدية وهران والتي تتحصر في مساحة جغرافية يحدها

شرقا كل من نهج "عدة بن عودة"، ونهج "زيغوت يوسف"، وجزء من نهج "الأمير عبد القادر. ويحدها غربا نهج "زبانة"، وجزء من نهج "الدكتور بن زرجب" جنوبا، ومن الشمال ساحة 1 نوفمبر 1954، وجزء من نهج "معطى محمد الحبيب"، و من الجنوب شارع "الإخوة بوشاقور". لم تذكر هنا إلا أهم الشوارع الكبرى ناهيك عن عدد كبيرة من الشوارع الصغيرة التي تحدد هذه المساحة جغرافيا. إذ تعتبر هذه الأخيرة الواقعة تحت مسؤولية مفتشة سيدي البشير كبيرة نوعا ما مقارنة بمفتشات أخرى في الولاية.

## 5-1-3 حدود المجتمع المدروس زمنيا:

يلعب الزمن دورا هاما في البحوث الإحصائية وخاصة إذا تعلق الأمر بدارسة معالم تتغير بصفة دائمة وفي فترات قصيرة الأمد، لأنه لا يمكن أن يكون للمعلومة مصداقية إلا إذا تم استغلالها في الوقت المناسب، فما الفائدة من إجراء سبر الرأي قبيل انتخابات معينة إذا لم تتحصل على النتائج قبل الإجراء الفعلي لهذه الانتخابات؟ و ما الفائدة من إجراء استطلاع للرأي لمواطني منطقة ما حول إنجاز مشروع بناء مصنع مواد كيماوية بجوارهم إذا جاءت نتائج هذا الاستطلاع بعد عملية الشروع في الإنجاز؟. إن وضع الحدود الزمنية لبحث ما يتمثل في تحديد الفترة التي سوف تمتد فيها مدة الدراسة والتي يجب احترامها مهما كانت الظروف والأسباب الخارجية بهدف الحصول على بحوث ذات مصداقية كبيرة.

كان من المفترض أن تدوم فترة الدراسة حوالي شهرين بحيث تمتد من **01** أبريل **2010** إلى نهاية ماي من نفس السنة، لكن نظرا للعراقيل التي تم التطرق إليها سابقا و إلى مشاكل تتعلق بتوفر الوقت لدى العدادين المتطوعين فإن مدة الدراسة امتدت إلى غاية شهر ديسمبر **2010**.

## 5-1-4إطار المعاينة:

قمنا بإعداد إطار المعاينة '( Base de Sondage ) انطلاقا من سجلات مفتشية الضرائب "سيدي البشير " والتي ضمت المؤسسات المسجلة بالمصلحة إلى غاية شهر ديسمبر 2009. نحن نعتبر أن هذا الإطار حديث ويفي بالغرض المطلوب نظرا لكوننا قد أقصينا المؤسسات التي أوقفت نشاطها قبل تاريخ انطلاق المسح بالمعانية ( Survey Sampling).

## 5-1-5تقسيم المجتمع إلى طبقات:

قسم المجتمع الإحصائي حسب متغيرة واحدة أساسية (Variable de Contrôle)، والمتمثلة في النشاط التجاري للمؤسسة نظرا لإيماننا بأن نشاط المؤسسة يلعب دورا كبيرا في تحديد حجم المتغيـرات المراد تقديرها في هذا البحث، وهي رقم الأعمال، والقيمة المضافة، والأرباح. يعرض الجـدول التـالي تقسيم وحدات المجتمع (المؤسسات) إلى طبقات و كذا حجم كل واحدة من هذه الأخيرة:

**جدول رقم 01**

**تقسيم وحدات مجتمع الدراسة إلى طبقات حسب النشاط التجاري**

| رقم الطبقة $h$ | النشاط التجاري | *حجم الطبقة* $N_h$ |
|---|---|---|
| 01 | بيع بالجملة | 32 |
| 02 | حراسة | 03 |
| 03 | الطب | 48 |
| 04 | أشغال عمومية | 38 |
| 05 | مكتبة ووراقة | 21 |
| 06 | مكتب دراسات | 11 |
| 07 | شركة نقل بحري | 4 |
| 08 | وكالة سفر | 5 |
| 09 | بيع منتجات كهرومنزلية | 8 |
| 10 | خردوات عامة | 7 |
| 11 | أدوات منزلية | 9 |
| 12 | استيراد و تصدير | 12 |
| 13 | نقل المسافرين | 5 |
| 14 | صناعة الحفاظات | 2 |
| 15 | بيع المجوهرات | 17 |
| 16 | مطعم | 05 |
| 17 | صيدلية | 15 |
| 18 | بيع الملابس | 36 |
| 19 | بيع الحليب و مشتقاته | 06 |
| 20 | جمعيات | 01 |
| 21 | مستحضرات التجميل | 14 |
| 22 | موثق | 02 |
| 23 | محاماة | 23 |
| 24 | محاسب | 05 |
| 25 | فندقة | 05 |
| 26 | بيع مشروبات | 05 |
| 27 | قطع الغيار | 02 |
| 28 | صيانة صناعية | 08 |
| 29 | مواد غذائية عامة | 11 |
| 30 | وكيل عبور | 01 |
| 31 | مؤسسة نظافة | 01 |
| 32 | مدرسة تكوين | 02 |
| 33 | مصور | 01 |
| 34 | انتاج مواد صناعية | 02 |
| 35 | تلحيم و ترصيص | 04 |
| 36 | تبغ و جرائد | 37 |
| 37 | مخبزة | 16 |
| 38 | قصابة | 07 |
| 39 | مقهى انترنيت | 02 |
| 40 | حلاقة | 02 |
| 41 | كشك متعدد الخدمات | 02 |
| 42 | انتاج المعدات | 02 |
| 43 | كراء العقارات | 01 |
| 44 | بيع ألعاب بلاستيك | 03 |
| 45 | كراء السيارات | 01 |
| 46 | أدوات خياطة | 08 |
| 47 | وكالة بريد و طرود | 01 |
| 48 | تأمين | 01 |
| | مجموع المؤسسات (N) | 454 |

## 2-5 الإمكانيات المادية والبشرية المتوفرة:

إن عملية إنجاح أي دراسة بالمعانية لا تقتصر فقط على التخطـيط الجيـد والعلمـي للمسح، والذي يأخذ في عين الاعتبار كل الخصائص الأساسية للمجتمع قيد الدراسـة، بـل يحتاج كذلك إلى عمل لوجستي ومنظم في الميدان الأمر الذي يتطلب توفير امكانات ماديـة وبشرية كافية من أجل إنجاح البحث بتكاليف معينة وفي أقرب الآجال.

### 1.2.5الإمكانيات والمعدات المادية:

**وسائل التنقل:** نظرا لمحدودية البحث من حيث الموارد المادية، والزمـان والمكـان فإنـه، وعكس غالبية البحوث المتعلقة بالمؤسسات، لا داعي هنا للحديث عن وسائل نقـل خاصـة كتلك التي يتم توفيرها للفرق الإحصائية التابعة للهيئات العمومية كالديوان الوطني للإحصاء أو مكاتب الدراسات الخاصة، حيث اكتفينا نحن فقط بوسيلة نقـل واحـدة خاصـة و كـذا الاستعانة في كثير من الأحيان بالمواصلات العمومية كالحافلات وسيارات الأجرة.

**تجهيزات المكتب و المعدات الحاسبة:** تم توفير كل المستلزمات المكتبية الضرورية للأفراد المساهمين في هذا المسح ( أقلام، ممحاة، مساطر أوراق، ملفات الخ...)، كما وضع تحـت تصرفنا مجموعة من الحواسيب وآلات الطبع الموفرة من طرف كلية العلـوم الاقتصـادية علوم التسيير والعلوم التجارية، إذ لا يفوتنا أن ننوه بالمجهودات الكبيرة والمعتبرة التي قـام و لازال يقوم بهام إطارات الكلية من أجل تهيئة وتوفير أحسن مـا أمكـن مـن المعـدات والتجهيزات التي تساعد الطلبة على القيام بأعمالهم في أحسن الظروف.

### 2.2.5الإمكانيات البشرية:

يلعب الجانب البشر دورا جوهريا في عملية جمع المعلومات الضرورية مـن لـدى المؤسسات (وحدات الدراسة) ، لأنه من الضروري أن يتمتع العدادين بنـوع مـن اللباقـة و القدرة على الإقناع اللتان تمكنهم من كسب ثقة و ارتيـاح الأفـراد المسـتجوبين قصـد الحصول على معلومات صحيحة و دقيقة و ذات مصداقية.

نظرا لعدم توفر الامكانات المادية اللازمة لتوظيف عدد كافي من العدادين كما تجرى العادة في البحوث الكبرى فإن الأمر اقتصر في بحتنا هذا على بضعة مساعدين متطـوعين تراوح عددهم مابين 2 إلي 4 أفراد طيلة فتر البحث.

### 3.5 مراحل تحديد حجم العينة:

لقد رأينا سابقا أن عملية تحديد حجم العينة من أهم و أصـــعب القـــرارات الواجـــب إتحادها من طرف الباحث, و التي تسبق إنجاز أي مسح بالمعاينة. لا مجال للحـــديث عـــن تدنية دالة التكاليف نظرا لكونها مسألة غير هامة جدا و ضرورية في مثل هدا النـــوع مـــن البحوث الصغيرة و المحدودة جدا من حيث الإمكانيات المادية و البشرية المسخرة لإنجازها. لذا سوف يرتكز عملنا على محاولة لتحديد حجم العينة الضروري لتحقيق مستوى معين مـــن الدقة يرمز له ب $d$ باحتمال معين قدره $\alpha$, ودلك في معاينة عشوائية طبقيـــة بحصـــص متناسبة ثم باستعمال التخصيص الأمثل لنايمن (Nayman Allocation).

تحتاج صيغة حجم العينة في كلتا الحالتين إلى تحديد تباينات المتغيرات $s_h^2$ في كـــل الطبقات إما عن طريق الرجوع إلى دارسات سابقة، الأمر الذي لا يتوفر لدينا للأسف بسبب انعدام مثل هدا النوع من البحوث المتعلقة بتقدير رقم أعمال المؤسسات أو القيمة المضـــافة، أو الأرباح في ولاية وهران، أو عن طريق محاولة التخمين (Guess) في القـــيم التقريبيـــة لهده التباينات بالرجوع إلى خبرة الإحصائيين الميدانيين، و للأسف لم تتسنى لنا كذلك هـــده الإمكانية بسبب تعدد الطبقات و تزايد احتمال استعمال قيم بعيدة جدا عن الحقيقة. و أخيرا لم يبقى لنا إلا أن نلجأ إلى المسح الاستطلاعي (Pilot Survey) كوسيلة جيدة لتحقيق هذا الغرض.

#### 1.3.5 إنجاز المسح الاستطلاعي و تقدير التباينات :

إن المتغيرات التي سوف نقوم بتقدير بعض معالمها في بحثنا هي رقم الأعمال الـــذي نرمز له ب(CA)، و القيمة المضافة (VA)، و الربح (RE). نحن نحتاج إلى تقدير تباينـــات هده المتغيرات في كل الطبقات، لذا نقوم بتحديد حجم عينة ضروري لبدأ المسح الاستطلاعي والدي يبلغ $n_{pilot} = 120$ مؤسسة، أي ما يعادل نسبة اختيار قدرها حوالي 26.5% من مجمـــل وحدات المجتمع الإحصائي الذي يبلغ 454 مؤسسة.

وزعت وحدات العينة على الطبقات بحصص متناسبة، بحيث يكون $f = \frac{N_h}{N} = \frac{n_h}{n}$، كما هو مبين في الملحق رقم8.

قمنا باختيار المؤسسات الواجب سحبها في هذا المسح باستعمال أسلوب المعاينة العشوائية البسيطة في كل طبقة بعد إدخال إطار المعاينة في برنامج الحاسب SPSS، ثم يقوم هذا الأخير بتوليد الأرقام العشوائية بشكل أوتوماتيكي و يختار الوحدات.

بعد ذلك قمنا بزيارة ميدانية لهده المؤسسات من أجل جمع البيانات، ونظرا لتعرض العدادين لمشكلة عدم الإجابة من قبل بعض المؤسسات، وبعد الإلحاح بتكرار الزيارة للمرة الثالث على الأقل، نقوم باستبدال الوحدات حتى نحصل على حجم العينة المحدد للمسح.

بعد جمع البيانات المتمثلة في تصريحات كل من رقم الأعمال (CA)، و القيمة المضاف (VA)، و الأرباح (RE)، المحصل عليها عن طريق الإطلاع العيني من طرف العداد على الميزانية الرسمية المصرح بها لسنة 2009 لدى مصلحة الضرائب، قمنا بتقدير تباينات المتغيرات قيد الدراسة.

يعرض الجدول التالي تباينات رقم الأعمال المقدرة في كل الطبقات. أما باقي النتائج فهي موجودة كاملة في الملحق رقم 2.

**جدول رقم:2**

**تباين رقم الأعمال في الطبقات**

| الطبقة $h$ | $S_h^2$ | الطبقة $h$ | $S_h^2$ |
|---|---|---|---|
| 1 | ([22]) 4.3941E+13 | 25 | 7.3487E+11 |
| 2 | 2.82251E+14 | 26 | 4.1922E+12 |
| 3 | 7.33814E+11 | 27 | 0 |
| 4 | 2.54998E+14 | 28 | 15257304.7 |
| 5 | 1.08921E+13 | 29 | 1.6997E+11 |
| 6 | 1.01618E+12 | 30 | 0 |
| 7 | 7.97649E+16 | 31 | 0 |
| 8 | 1.23941E+14 | 32 | 0 |
| 9 | 9.55317E+13 | 33 | 0 |
| 10 | 2.32441E+12 | 34 | 0 |
| 11 | 1.32914E+11 | 35 | 0 |
| 12 | 3.62909E+17 | 36 | 1.8765E+11 |
| 13 | 1.60252E+11 | 37 | 2.0627E+12 |
| 14 | 0 | 38 | 4.4746E+11 |
| 15 | 80891664693 | 39 | 0 |
| 16 | 1.38195E+12 | 40 | 0 |
| 17 | 1.85951E+14 | 41 | 0 |
| 18 | 0 | 42 | 0 |
| 19 | 7.75633E+13 | 43 | 0 |
| 20 | 0 | 44 | 1.5351E+13 |
| 21 | 5.50417E+11 | 45 | 0 |
| 22 | 0 | 46 | 1.3359E+12 |
| 23 | 12564347445 | 47 | 0 |
| 24 | 2.65516E+11 | 48 | 0 |

**المصدر:أنظر إلى الملحق رقم 2**

نلاحظ من خلال هذا الجدول أن قيمة التباينات المقدرة كبيرة نوعا، فعلى سبيل المثال يبلغ تباين رقم الأعمال في الطبقة 1 أكثر من ***4.3941E+13*** دينار جزائري[2]، وهو ما يعادل انحراف معياري قدره حوالي 6.6 مليون دج، و في الطبقة 12 يصل حتى أكثر من $3.63 \times 10^{17}$ دج، أي ما يعادل انحراف معياري قدره حوالي 602 مليون دج. يمكن أن يكون سبب هذه القيم الكبيرة راجع إما إلى سوء تقديراتنا الناتج عن محدودي المسح الاستطلاعي كونه لم يشمل عدد كبير من المؤسسات في المجتمع، أو إلى حقيقة أنه يوجد بالفعل اختلاف و تباين كبير بين تصريحات رقم الأعمال لهذه المؤسسات في كل طبقة ( حسب النشاط التجاري).

[22] تعني E+13 وجود 13 رقم بعد النقطة.

لا يمكن التحقق مما نقول إلا عن طريق إجراء حصر شامل لجميع المؤسسات، وبالنظر إلى القيم الكبيرة لهده التباينات، نحن نتوقع أن تفرز نتائج استعمال الصيغ الرياضية لتحديد حجم العينة عن قيم كبيرة كذلك لهده الأخيرة لأنه كلما كان المجتمع الإحصائي غير متجانس كان حجم العينة الضروري لتمثيله كبيرا بهدف الحصول على نتائج تقترب من الواقع أكثر فأكثر.

## 2.3.5 تحديد حجم العينة:

في معاينة عشوائية طبقية بحصص متناسبة، أي بنفس معدلات الاختيار(الفصل03) و من أجل تقدير المتوسط رقم الأعمال مثلا لكل المؤسسات، تعطى صيغ حجم العينة الضروري من أجل قبول مجال خطأ معين قدره $d$ باحتمال قدره $\alpha$ المسمى كذلك مستوى المعنوية وفقا للصيغة رقم 6.6 .

يعرض الجدول رقم 3 ، و الشكل رقم 1 التاليين التغيرات الحاصلة في حجم العينة بدلالة تغير كل من مجال الخطأ المقبول $d$ ، و مستوى المعنوية $\alpha$:

**جدول رقم:2**

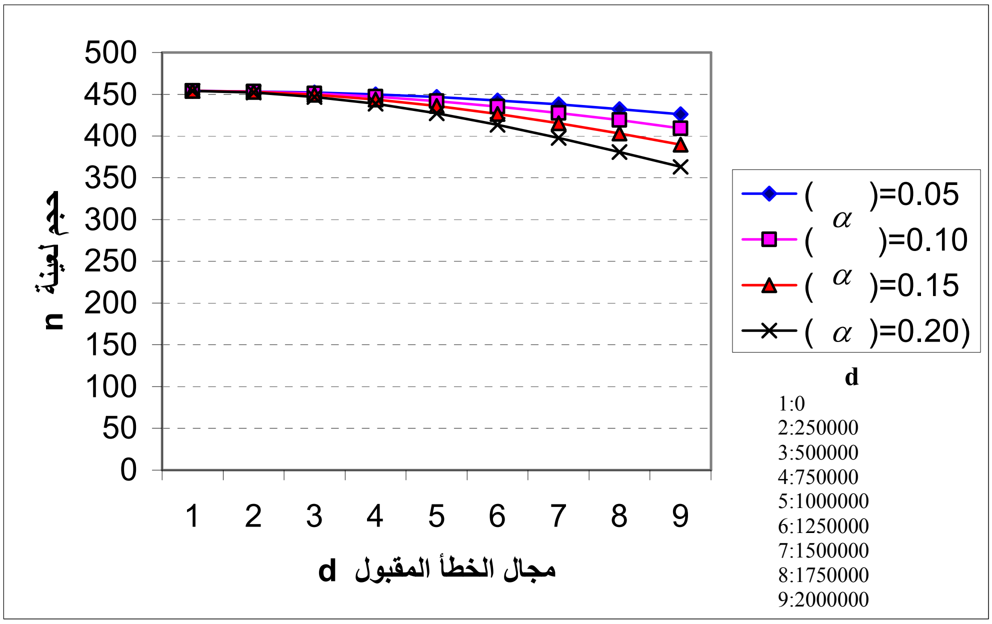
**تغير حجم العينة بدلالة تغير مجال الخطأ المقبول $d$ ، و مستوى المعنوية $\alpha$**

**الوحدة: مؤسسة**

| 0.05 | 0.1 | 0.15 | 0.2 | مستوى المعنوية $\alpha$ / مجال الخطأ $d$ |
|---|---|---|---|---|
| 454 | 454 | 454 | 454 | 0 |
| 453.5 | 453.2 | 452.8 | 452.2 | 250000 |
| 452.2 | 450.9 | 449.4 | 447 | 500000 |
| 449.9 | 447.2 | 443.7 | 438.6 | 750000 |
| 446.7 | 442.0 | 436 | 427.3 | 1000000 |
| 442.7 | 435.5 | 426.5 | 413.6 | 1250000 |
| 437.9 | 427.8 | 415.4 | 398 | 1500000 |
| 432.3 | 419.1 | 403.1 | 381 | 1750000 |
| 426.1 | 409.5 | 389.7 | 363.2 | 2000000 |

**شكل رقم:1**

**تغير حجم العينة بدلالة تغير مجال الخطأ المقبول $d$ ، و مستوى المعنوية $\alpha$**




يمكن قراءة الجدول بطريقين مختلفتين: الأولى عمودية، بحيث نلاحظ مثلا أنه من أجل مستوى معنوية ثابت قدره $\alpha = 0.05$، يتغير حجم العينة بتغير مجال الخطأ المقبول $d$ من $n = N = 454$ مؤسسة في حالة $d = 0$ دينار جزائري، أي في حالة التطابق التام بين نتائج المسح بالمعاينة مع النتائج الحقيقية في المجتمع وهدا لا يتحقق بالطبع إلا عن طريق الحصر الشامل لكل مؤسسات المجتمع، إلى $n = 446$ مؤسسة في حالة $d = 1000000$ دينار جزائري، ثم يستمر حجم العينة في التناقص حتى يبلغ $n = 426$ مؤسسة لما يكون $d = 2000000$ دينار جزائري.

أما الثانية فهي أفقية، في هده الحالة نثبت مجال الخطأ $d$ ونغير مستويات المعنوية $\alpha$. فعلى سبيل المثال، نلاحظ أنه من أجل $d = 1500000$ دينار جزائري ينخفض حجم العينة كلما رفعنا من مستوى المعنوية بحيث يبلغ كل من 437، و 427، و 415، و 398 من أجل $\alpha$ يساوي 0.05، و 0.10، و 0.15، و 0.20 على التوالي .

نستنتج مما سبق، أنه كلما كان مجال الخطأ المقبول $d$ و مستوى المعنوية $\alpha$ كبيرين كان حجم العينة صغيرا، و العكس صحيح. و هذا يعني أنه كلما تسامحنا في الدقة المرغوب

بلوغها اكتفينا بأحجام عينة صغيرة. أما إذا كنا بحاجة ماسة إلى بلوغ دقة كبيرة في النتـــائج فإن الأمر يتطلب سحب حجم عينة أكبر.

إن المثير للانتباه و فيما سبق، و من خلال ملاحظة دقيقة للمنحنيات في الشكل رقم 1 هو أحجام العينة الكبيرة المسحوبة. حيث أنه حتى عند أسوء الحالات التي نقبل فيها مجـــال خطأ كبير نوعا ما قدره **2.000.000** دج، و بمستوى معنوية كبير كذلك و البالغ $\alpha = 0.20$ فــإن حجم العينة لم ينخفض تحت $n = 363$ ، أي ما يعادل معدل اختيار يساوي $f = n / N = 0.79$ . يعتبر هذا المعدل كبيرا جدا مقارنة بما تجري به العادة في المسوحات الإحصائية بالمعاينـــة، حيث لا يتعدى فيها معدل الاختيار 30 % من مجمل وحدات المجتمع.

يوجد تفسيران ممكنان للنتائج الكبيرة لأحجام العينة التي حصلنا عليهـــا. الأول، هـــو احتمال أن تكون تقديرات تباينات رقم الأعمال $s_h^2$ في الطبقـــات- الناتجـــة عـــن المســـح الاستطلاعي([23])- كبيرة جدا و مختلفة عن التباينات الحقيقية $S_h^2$ في المجتمع نظرا لصـــغر حجم العينة المستخدم مقارنة بالعدد الكبير للطبقات. أما الثـــاني، فهـــو احتمـــال أن تكـــون المؤسسات في كل المجتمع غير متجانسة جدا لدرجة أنه يتعين علينا استجواب ما لا يقل عن 70% من المجتمع حتى نحصل على نتائج ذات معنوية إحصائية.

من أجل إثبات صحة أحد الاحتمالين المذكورين سابقا، تنازلنا أكثر فيما يخص الدقـــة المرغوب تحقيقها من أجل الشروع في المسح قصد تقدير بعض معالم المجتمـــع، و قبلنـــا مبدئيا بمجال خطأ قدره 2000000 دج عند مستوى معنوية يصل حتى $\alpha = 0.30$.فحصلنا علـــى حجم عينة قدره $n = 265$ مؤسسة، و هو ما يعادل نسبة اختيار تساوي 58.3 %.

## 4.5 إجراءات سحب وحدات العينة وتقدير المعالم:

### 1.4.5 سحب وحدات العينة:

قمنا بإتباع نفس الخطوات المستعملة في المسح الاستطلاعي ودلك بتوزيـــع وحـــدات العينة $n$ في كل الطبقات بحصص متناسبة كما هو موضح في الملحق رقم9 ، كما ســـحبت

[23] أنظر الملحق رقم2

المؤسسات الواجب استجوابها في كل طبقة $h$ بإتباع أسلوب المعاينة العشوائية البسيطة عـــن طريق مولد الأرقام العشوائية في برنامج الحاسب SPSS.

تمت عملية جمع البيانات عن طريق المقابلة الميدانية للمؤسســـات المختـــارة. حيـــث واجهتنا في هذا البحث و كغيره من البحوث الميدانية مشكلة عـــدم الإجابـــة و التـــي قمنـــا بتجاوزها عن طريق استبدال المؤسسات الرافضة للتصريح ببياناتها بمؤسسات مـــن نفـــس الطبقة بعد تكرار الزيارة للمرة الثالثة على الأقل كما سبق ذكره.

**2.4.5 تقدير معالم المجتمع:**

قمنا في هذه الفقرة بتقدير بعض المعالم الخاص برقم الأعمال الذي نرمز له ب CA، و القيمة المضافة VA، و الربح RE بحيث تشمل هذه المعالم كلا من: مقدر المتوسط $\overline{y}_{st}$، و المجموع $\hat{Y}$، و تبايناتهما، وكذا مقدر تباينات المتغيرات $S_h^2$.

باستعمال صيغ المقدرات و تبايناتها كما هي معروضة في الفصل الثالث، الفقرة 4

بحيث أن:

$N=454$
$n=265$
$h=1, 2,3,.....,48$

بتعويض النتائج المعروضة في الملحق رقم 3 و المتعلقة بمتوسطات المتغيـــرات $\overline{y}_h$ في كل طبقة $h$، والتباينات المقدرة $s_h^2$ ، نحصل على النتائج المعروضة فـــي الجـــدول الموالي:

**جدول رقم3**

**تقدير متوسط و مجموع رقم الأعمال، والربح، و القيمة المضافة**

الوحدة: دج

| المتغيرات / المقدرات | رقم الأعمال CA | الربح RE | القيم المضافة VA |
|---|---|---|---|
| المتوسط $\bar{y}_{st}$ | 4993660,423 | 613334,5671 | 516834,39 |
| المجموع $\hat{Y}$ | 2267121932 | 278453893,5 | 234642813 |

نلاحظ من خلال هدا الجدول، أن متوسط رقم الأعمال يبلغ تقريبا 5 مليون دج في حين يبلغ مجموعه أكثر من 2,26 مليار دج . أما فيما يخص متوسط كل من الربح و القيمة المضافة فهو على التوالي: 613334,5671 دج، و 516834,39 دج. ومجموعهما هو على التوالي: 278,4 ، و 234,6 مليون دج.

إن الهدف من هده الدراسة وكما سبق ذكره هو توضيح أثر حجم العينة على دقة النتائج، و حتى نتمكن من فعل دلك قمنا في خطوة ثانية من هده الدراسة بمحاولة لجمع كل بيانات المجتمع عن طريق الحصر الشامل لكل الوحدات ( المؤسسات).

**5.5 الحصر الشامل للمؤسسات و حساب المعالم الحقيقية في المجتمع:**

في خطوة ثانية من هده الدراسة، قمنا بزيارة كل المؤسسات الموجودة في إطار المعاينة حتى نتمكن من الحصول على البيانات الحقيقية للمجتمع. فبلغت نسبة الإجابة حوالي 65 %، و دلك بعد جهد كبير في محاولة منا لرفع هده النسبة عن طريق الإلحاح و تكرار الزيارة لمرات عديدة.

يعرض الجدول الآتي رقم (4) النتائج الحقيقي لبعض المعالم في المجتمع و كذلك تقديراتنا للمعالم عن طريق محاولة لتعديل القيم الناقصة (Missing Data)، بتعويضها

بمتوسطات المتغيرات في كل طبقة، كأن نعوض مثلا رقم أعمال مؤسسة ما لم تجب على تساؤلاتنا بمتوسط رقم أعمال المؤسسات التي تنتمي إلى نفس الطبقة $h$ والتي صرحت لنا ببياناتها.

**جدول رقم:4**

**نتائج الحصر الشامل**

الوحدة: دج

| المتغيرات / المعالم | CA | CA_1 [24] | RE | RE_1 | VA | VA_1 |
|---|---|---|---|---|---|---|
| المتوسط $(\overline{Y})$ | 5127974,7 | 5214340 | 233241 | 533397 | 785640,15 | 1075162,2 |
| المجموع $(Y)$ | 1507624569 | 2367310347 | 68572854.19 | 242162236,5 | 230978203.9 | 488123660,1 |
| التباين $(S^2)$ | 3.507E+14 [25] | 2.269E+14 | 1.597E+13 | 1.050E+13 | 9.294E+12 | 6.166E+12 |

**المصدر: حسبت النتائج باستعمال بيانات الملحق رقم 1و5.**

نلاحظ من خلال هذا الجدول أن كل قيم معالم رقم الأعمال، و الربح، و القيمة المضافة تتغير بعد تعويض القيم الناقصة. فعلى سبيل المثال إرتفع متوسط رقم أعمال المؤسسات من **5127974,7** دج إلى **5.214.340** دج، و مجموعه من 1,5 مليار دج إلى 2,36مليار دج. كما نلاحظ كذلك أن تباين رقم الأعمال (CA) انخفض من 3.507E+14 إلى 2.269E+14 كما حدث نفس الشيء مع كل من الربح و القيمة المضافة.

[24] يدل المؤشر CA_1 على القيم الجديدة للمتغير بعد تعديل القيم الناقصة (Missing Data)

[25] يدل الرمز E+14 على عدد الأرقام على يمين النقطة، فمثلا تساوي القيمة 3.507E14 أكثر من .3.50700000000000

**6.5 أثر تغيير حجم العينة على دقة النتائج:**

أثناء إجرائنا لهدا البحث، مرت عملية تحديد حجم العينة بعدة مراحل: في المرحلة الأولى قمنا بسحب 120 مؤسسة من أجل إجراء المسح الاستطلاعي، وفي المرحلة الثانية وقع اختيارنا على 265 مؤسسة من أجل القيام بمختلف التقديرات، و في آخر مرحلة قمنا بحصر شامل لكل المؤسسات و التي بلغ عددها 454 مؤسسة. إن كل مرحلة من المراحل السابقة تحتوي على نتائج تقديرات خاصة بها، والتي نعرضها في الجدول الموالي:

**جدول رقم:5**

**أثر تغيير حجم العينة على دقة النتائج**

الوحدة: دج

| حجم العينة / المعالم | n=120 | n=265 | n=N=454 ([26]) |
|---|---|---|---|
| $(\overline{y}_{st})$ | 3606266,686 | 4993660,423 | 5214340 |
| $(\hat{Y})$ | 1637240535 | 2267121832 | 2367310347 |
| $Var\,\overline{y}_{st}$ (الوحدة: دج$^2$) | 6.06E+13 | 9.11E+10 | 0 |
| $Var\,\hat{Y}_{st}$ (الوحدة: دج$^2$) | 1.25E+19 | 1.88E+16 | 0 |
| $\lvert\overline{y}_{st}-\overline{Y}\rvert$ | 1608073,31 | 220679,57 | 0 |
| $\lvert\hat{Y}-Y\rvert$ | 730069812 | 100188515 | 0 |

يعرض الجدول أعلاه تغير حجم العينة ومدى تأثيره على أخطاء التقدير وتباينات المقدرات. نلاحظ أنه من أجل حجم عينة قدره 120 مؤسسة بلغ خطأ تقدير متوسط رقم الأعمال ومجموع رقم الأعمال على التوالي 1608073,31 و 730069812 دج. في حين ينخفض هدا الخطأ بشكل كبير و معتبر عند ارتفاع حجم العينة إلى 265 مؤسسة ليبلغ فقط 220679,57 دج، و 100188515 دج لكل من المتوسط والمجموع على التوالي. ثم ينعدم تماما خطأ التقدير عند معاينة كل مؤسسات المجتمع قيد الدراسة. كما نلاحظ كذلك انخفاض في

[26] نتائج الحصر الشامل بعد تعويض القيم الناقصة أنظر إلى الجدول رقم (04)

تباينات المقدرات كلما ارتفع حجم العينة، فعلى سبيل المثال ينخفض تباين مقدر المتوسط من 6.06E+13 إلى 9.11E+10 عند ارتفاع حجم العينة من 120 إلى 265 مؤسسة على التوالي.

### 7.5 تحديد حجم العينة الضروري لإجراء دراسة لاحقة:

لقد ذكرنا في الفقرة 2.3.4 من هدا الفصل أن حجم العينة الكبير المحصل عليه والبالغ 265 مؤسسة والناتج عن قبول مجال خطأ كبير يبلغ 2 مليون دج بمستوى معنوية $\alpha = 0.30$ كان نتيجة لأحد الاحتمالين:

1-إحتمال أن تكون تقديرات تباينات رقم الأعمال $s_h^2$ المستعملة في الصيغة الرياضية لحجم العينة كبيرة جدا عن التباينات الحقيقية في المجتمع.

2-إحتمال وجود تباين وتشتت كبير حقيقي بين رقم الأعمال المصرح به في كل طبقة ناهيك عن إمكانية وجود تباين كبير بين كل المؤسسات في المجتمع.

بعد حصولنا على كل البيانات الناتجة عن الحصر الشامل تبين[27] تحقق الاحتمال الأول بشكل كبير و الاحتمال الثاني نوعا ما[28] .

سوف نحاول في هده الفقرة استغلال النتائج الحقيقية لتباينات رقم الأعمال $S_h^2$ في محاولة لتحديد مختلف أحجام العينة الضرورية لإجراء دراسات مستقبلية حول نفس المجتمع باستعمال أسلوب المعاينة العشوائية الطبقية،أولا (1) بحصص متناسبة بهدف بلوغ دقة معينة، و ثانيا(2) عن طريق التخصيص الأمثل لنايمن **( Nayman Allocation ).**

#### 1.7.5 تحديد حجم العينة الضروري لقبول مجال خطأ قدره $d$ بمستوى معنوية $\alpha$:

باستعمال صيغة حجم العينة رقم (6.6)صفحة (60) و بتعويض تباينات رقم الأعمال الحقيقية $S_h^2$ في كل طبقة، نحصل على الجدول و البيان التاليين واللذان يوضحان تغير حجم العينة بدلالة مجال خطأ مقبول $d$ عند مستوى معنوية $\alpha$.

[27] بعد إجراء مقارنة بين تباين رقم الأعمال الناتج عن المسح الاستطلاعي (ملحق رقم2 ) و التباين الحقيقي (ملحق رقم4).

[28] أنظر إلى الملحقين رقم 04 و 05.

## جدول رقم:07

## تغير حجم العينة بدلالة تغير مجال الخطأ المقبول $d$ ، و مستوى المعنوية $\alpha$

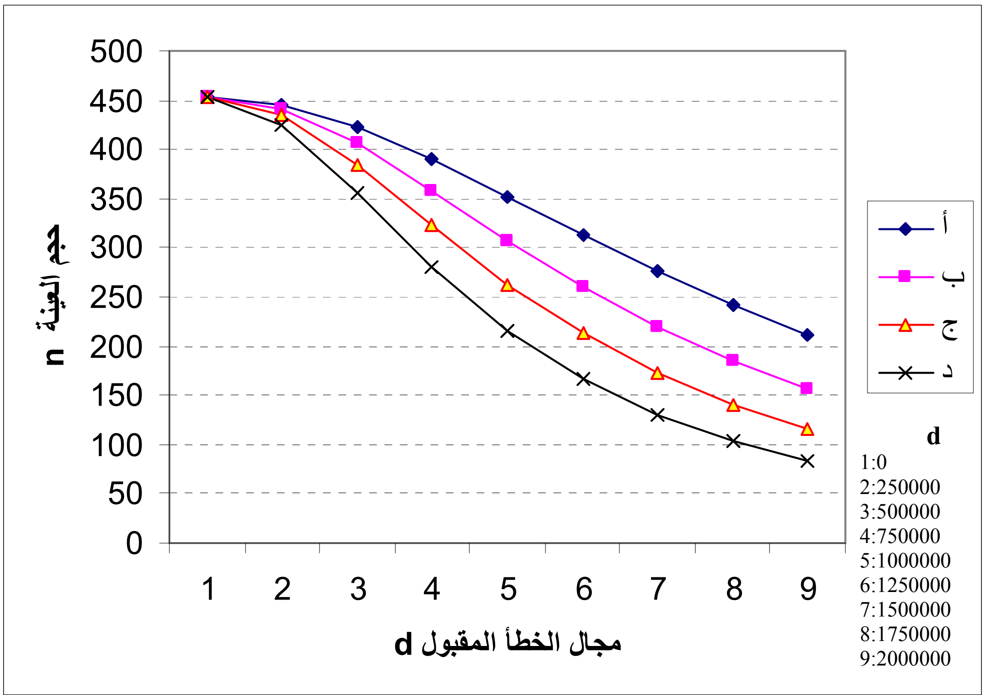
الوحدة: مؤسسة

| 0.5 | 0.1 | 0.15 | 0.2 | $\alpha$ / $d$ (دج) |
|---|---|---|---|---|
| 454 | 454 | 454 | 454 | 0 |
| 446.0 | 440.8 | 434.3 | 424.8 | 250000 |
| 423.6 | 405.6 | 384.3 | 356.2 | 500000 |
| 390.8 | 357.8 | 322.5 | 280.6 | 750000 |
| 352.6 | 307.2 | 263.2 | 216.3 | 1000000 |
| 313.3 | 260.0 | 212.9 | 167.1 | 1250000 |
| 275.7 | 218.8 | 172.5 | 130.8 | 1500000 |
| 241.4 | 184.3 | 141 | 104.0 | 1750000 |
| 211.2 | 156.0 | 116.4 | 84.2 | 2000000 |

## شكل رقم:2

## منحنيات تبين تغير حجم العينة بدلالة مجال الخطأ المقبول *d*

## عند مستوى معنوية $\alpha$ ثابت

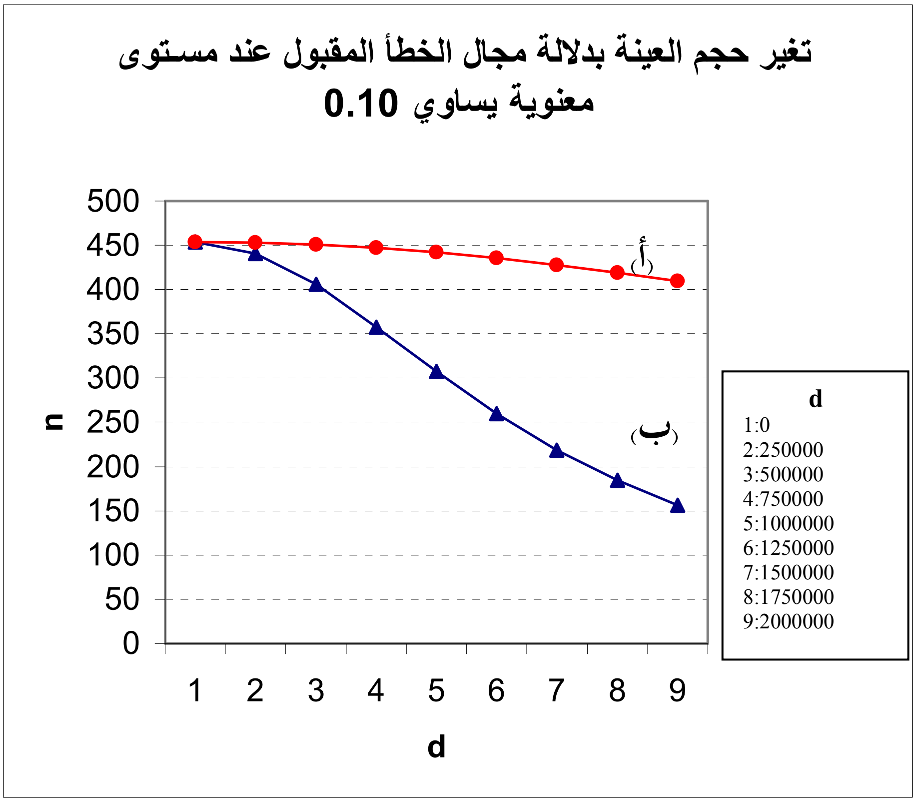


يعرض المنحنى ( أ ) تغير حجم العينة عند مستوى معنوية ثابت قدره $\alpha = 0.05$، وقيم مختلفة لمجال الخطأ المقبول $d$ كما تعرض المنحنيات الأخرى نفس التغير ولكن من أجل مستويات المعنوية 0.10، و0.15، و0.20 على التوالي.

نلاحظ من خلال كل المنحنيات أنه لا يمكن بلوغ دقة كاملة، أي $d = 0$ مهما تغير مستوى المعنوية $\alpha$ إلا بالحصر الشامل لكل المؤسسات (n=N=454).كما نلاحظ كذلك أنه كلما ارتفع مجال الخطأ المقبول ينخفض حجم العينة و كلما ارتفع $\alpha$ مع بقاء $d$ ثابت ينخفض حجم العينة كذلك.

عند مقارنة هده النتائج مع تلك المحصل عليها في الفقرة 2.3.5 صفحة(113) نلاحظ أن أحجام العينة المحسوبة و الموافقة لنفس الظروف $(d,\alpha)$ مختلفة كثيرا، حيث أنها في هده الحالة أكثر دقة وواقعية، ويظهر دلك جليا من خلال المنحنيين التاليين:

**شكل رقم:3**



يمثل المنحنى ( أ ) تغير قيم أحجام العينة المحصل عليها باستعمال تقديرات تباينـــات رقـــم الأعمال $S_h^2$. أما المنحنى (ب) فهو يمثل نفس التغير ولكن باستعمال التباينات الحقيقيـــة $S_h^2$. نلاحظ أن كل قيم حجم العينة في المنحنى (أ) أكبر من قيم حجم العينة فـــي المنحنـــى (ب) وهدا يعني أنه في الحقيقة، إذا كانت تقديراتنا للتباينات تقترب من الواقع فإننا سنوفر الكثيـــر في مسألة اختيار حجم العينة الضروري لبلوغ دقة معينة. فعلى سبيل المثال، كان يكفي مثلا أن نسحب 260 مؤسسة ( حسب المنحنى أ) بدلا من سحب 435 مؤسسة (حسب نتائج المنحنى ب) من أجل تحقيق مجال خطأ مقبول قدره $d = 1250000$ دج عند مستوى معنوية $\alpha = 0.10$.

في الفقرة 2.3.5 صفحة (115) كنا قد حصلنا على حجم عينة قدره 265 مؤسسة، وكـــان الهدف هو بلوغ الدقة التالية:

$$d = 2000000$$
$$\alpha = 0.30$$

من خلال ملاحظة للجدول رقم 6 صفحة (119) نلاحظ أنه كان بالإمكان بلوغ نفـــس هده الدقة فقط بسحب أقل من 84 مؤسسة. أما 265 مؤسسة فهي تحقق الدقة التالية:

$$\begin{cases} d = 1250000 \\ \alpha = 0.10 \end{cases}$$
$$\begin{cases} d = 1000000 \\ \alpha = 0.15 \end{cases}$$

وهذا يرجعنا إلى التأكيد على أهمية توفر دراسات سابقة من أجل الاستفادة من نتائجها في إجراء بحوث لاحقة. الامر الذي يدفعنا إلى حث الباحثين على إجراء دراسات من هـــذا القبيل و في جميع الميادين حتى و إن كانت هذه الدراسات بسيطة لكنها تمكننا لاحقـــا مـــن التمهيد لإجراء بحوث تكون نتائجها اكثر دقة و مصداقية.

### 2.7.5 تحديد حجم العينة حسب التخصيص الأمثل "لنايمن"(Nayman Allocation):

في هذه الفقرة سوف نعرض كيفية توزيع وحدات العينة $n$ على الطبقات بهدف تدنية تباين مقدر المتوسط $Var\bar{y}_{st}$ لما تكون التكلفة مثبتة حسب التخصيص الأمثل "لنـــايمن". وهـــذا يكافئ نفس التخصيص ولكن من أجل حجم عينـــة ثابـــت إذا كانـــت تكلفـــة معاينـــة كـــل مؤسسة $(c_h = c \forall h = 1,2,.....,k)$ ثابتة و متساوية في كل الطبقات[29] .

[29] أنظر الفصل (03 ) الفقرة ( 2.6.3).

تعطى صيغة حجم العينة حسب التخصيص الأمثل لنايمن حسب العلاقة (21.6)من الفصل (3) ، و بعد تعويضنا لحجم عينة مثبت قدره 265 مؤسسة، و تباينات رقم الأعمال $S_h^2$ كما هي موضحة في الملحق رقم (4)، و الترجيحات $W_h = \frac{N_h}{N}$ كما هي معروضة في الملحق رقم 7 في هذه الصيغة، حصلنا على أحجام عينة مثلى في كل طبقة كما هو مبين في الجدول رقم: 07 الموالي، لكن معظمها كان أكبر من حجم الطبقة في حد ذاتها، وهذا ما سبق تسميته في الفقرة 2.2.6 من الفصل الثالث "بالمحاصة التي تحتاج معاينة تزيد عن 100%، يكمن الحل في مثل هذه الحالات بوضع حجم العينة الأمثل مساويا لحجم الطبقة $h$ أي $\tilde{\tilde{n}}_h = N_h$ ، ثم نوزع باقي حجم العينة على الطبقات المتبقية عن طريق صيغة المحاصة المثلى رقم (25.6) و التي جاءت نتائجها كما يلي:

جدول رقم:7

| $\tilde{\tilde{n}}_h$ | $n_{h.opt}$ | $N_h$ | طبقة $h$ | $\tilde{\tilde{n}}_h$ | $n_{h.opt}$ | $N_h$ | طبقة $h$ | $\tilde{\tilde{n}}_h$ | $n_{h.opt}$ | $N_h$ | طبقة $h$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 00 | 0 | 1 | 33 | 15 | 22 | 15 | 17 | 32 | 45 | 32 | 1 |
| 00 | 0 | 2 | 34 | 13 | 22 | 36 | 18 | 03 | 6 | 3 | 2 |
| 00 | 0 | 4 | 35 | 03 | 5 | 6 | 19 | 02 | 3 | 48 | 3 |
| 01 | 2 | 37 | 36 | 00 | 0 | 1 | 20 | 38 | 54 | 38 | 4 |
| 01 | 2 | 16 | 37 | 14 | 46 | 14 | 21 | 04 | 7 | 21 | 5 |
| 00 | 0 | 7 | 38 | 00 | 0 | 2 | 22 | 00 | 0 | 11 | 6 |
| 00 | 0 | 2 | 39 | 00 | 0 | 23 | 23 | 00 | 0 | 4 | 7 |
| 00 | 0 | 2 | 40 | 00 | 0 | 5 | 24 | 05 | 5 | 5 | 8 |
| 00 | 0 | 2 | 41 | 00 | 0 | 5 | 25 | 04 | 7 | 8 | 9 |
| 02 | 11 | 2 | 42 | 01 | 2 | 5 | 26 | 01 | 1 | 7 | 10 |
| 00 | 0 | 1 | 43 | 02 | 3 | 2 | 27 | 02 | 3 | 9 | 11 |
| 01 | 1 | 3 | 44 | 00 | 0 | 8 | 28 | 00 | 0 | 12 | 12 |
| 00 | 0 | 1 | 45 | 01 | 1 | 11 | 29 | 00 | 0 | 5 | 13 |
| 02 | 3 | 8 | 46 | 00 | 0 | 1 | 30 | 00 | 0 | 2 | 14 |
| 00 | 0 | 1 | 47 | 00 | 0 | 1 | 31 | 00 | 0 | 17 | 15 |
| 00 | 0 | 1 | 48 | 00 | 0 | 2 | 32 | 00 | 0 | 5 | 16 |

نلاحظ من خلال هذه النتائج أن حجم العينة الأمثل $n_{h.opt}$ في الطبقة رقم 01 مثلا يساوي 45 مؤسسة في حين أن عدد المؤسسات في نفس الطبقة لا يتعدى 32 مؤسسة، لذا نقوم بتحديد حجم العينة في هذه الطبقة بحيث يكون مساويا لعدد وحدات الطبقة كاملة $(\tilde{\tilde{n}}_h = N_h)$ ، بنفس الطريقة نتعامل مع

الطبقات رقم : 42,27,21,17,09,04 ثم يوزع باقي حجم العينة على الطبقات اللاحقة باستعمال صيغة المحاصة المثلى.

## نتائج و استنتاجات

لقد أثبتت نتائج البحث الميداني حول المؤسسات أن التحديد الجيد لحجم العينة، أي عدد المؤسسات الواجب سحبها يتطلب تقديرا جيدا لتباينات المتغيرة قيد الدراسة. إذ تبين أنه كان بالإمكان الاقتصاد في الجهد و الوقت بمعاينة حوالي 84 مؤسسة بدلا من 265 مؤسسة[30] بهدف تحقيق الدقة المطلوبة، وذلك بسبب جهلنا التام لمعطيات سابقة حول المجتمع المدروس، وكذلك نتيجة للقدرة المحدودة للمسح الاستطلاعي على منح تقديرات جيدة للتباينات نظرا لقلة المؤسسات المسحوبة فيه. كما تبين أنه كلما كان حجم العينة كبيرا ( عدد المؤسسات) كانت النتائج المحصل عليها أكثر دقة و فاعلية.

لاحظنا كذلك مدى تأثر حجم العينة بمجال الخطأ المقبول و مستوى المعنوية، حيث تبين أنه كلما تساهلنا في مسألة الدقة المراد تحقيقها كان حجم العينة صغيرا، و كلما كانت الدقة المراد تحقيقها كبيرة وجب علينا سحب عدد أكبر من الوحدات(المؤسسات).

تجدر الإشارة هنا إلى مسألة جد هامة، وهي ضرورة تصميم دوال تكلفة جيدة تمكننا في بحوث أكبر حجما من الحصول على معاينة مثلى تهدف إلى تدنية التكاليف أو تباينات التقديرات. رأينا كذلك أهمية استغلال النتائج الحالية للبحث حتى و إن كانت ناقصة، في التمهيد لإجراء دراسات لاحقة .

[30] أنظر إلى الصفحة 123 من هذا الفصل.

## الخلاصة:

تعتبر المسوحات الإحصائية بالمعاينة (échantillonnage) بمختلف أنواعها و أساليبها جد ضرورية كونها تساعد الباحثين في العديد من المجالات سواء كانت اقتصادية، اجتماعية، سياسية أو حتى كلينيكية (طبية)...إلخ، في إعداد دراساتهم الميدانية، و بشكل عام، تمكن صانعي القرارات كل في مجاله من اعتماد سياسات وتدابير موضوعي و عقلانية و ذات مصداقية كبيرة.

كان الهدف من إعداد هذه المذكرة الإسهام و لو بالشيء القليل في إثراء مكتبتنا الجامعية بهذا النوع من البحوث و المناهج الإحصائية التي تمكن زملائنا الباحثين من الإلمام بتقنيات المعاينة الإحصائية بطريقة مبسطة، و كذا تسليط الأضواء على مسألة جد هامة غالبا ما يتم تجاهلها أو تناسيها من قبل الباحثين ألا و هي كيفية تحديد عدد وحدات العينة الضروري قبل الشروع في إجراء أي دراسة على أن يتم ذلك بطريقة علمية و عقلانية تجنب الباحث العشوائية في اتخاذ مثل هذه القرارات. حيث أوضحنا في هذا العمل بأن حجم العينة يتغير بدلالة كل من مجال الخطأ المقبول، و مستوى المعنوية، و حتى حجم الميزانية المتوفرة لإجراء دراسة ما.

من خلال هذا العمل البسيط تجدر الإشارة إلى النقاط الأساسية التالية:

- الدور الكبير الذي يلعبه المسح بالمعاينة في مساعدة الباحثين من أجل إجراء بحوث ميدانية قليلة التكلفة لكنها تؤدي إلى الحصول على نتائج دقيقة و فعالة.
- أهمية وجود بيانات إحصائية (بنك للمعلومات) في جميع الميادين من أجل إنجاح دراسات ميدانية.
- الحاجة الماسة في بلادنا إلى توفر بحوث ميدانية يمكن من خلالها إنشاء بنك للمعلومات يساعد الباحثين في مختلف المجالات و كذا صناع القرار من أجل التخطيط الجيد و الموضوعي قصد إنجاح البرامج التنموية. إذ أن الدول المتقدمة تولي أهمية كبيرة لهذا النوع من التقنيات الكمية من أجل تحديد الخصائص و إبراز الاتجاه العام لمختلف الظواهر و التمكن من تحليل العلاقات المتشابكة و المتبادلة بين المعالم على أسس موضوعية.

تم التطرق في هذا البحث إلى أساليب المعاينة العشوائية كما يمكن أن يتوسع مجال العمل في مجال تحديد حجم العينة مستقبلا ليشمل حالات أكثر تعقيد مثل نظرية اختبار الفرضيات، أو الطريقة البايزية في اتخاذ القرارات و حتى القياس الاقتصادي.

# قائمة المراجع


## المراجع باللغة العربية:

(1 ) عبد الرزاق أمين أبو شعر. « العينات و تطبيقاتها في البحوث الاجتماعية » . الإدارة العامة للبحوث . الكلية العربية السورية. 1997 .

(2 ) وليام كوكران. « تقنية المعاينة الإحصائية » ، ترجمة الدكتور 'أنيس كنجو'، عمادة شؤون المكتبات، جامعة الملك سعود، المملكة العربية السعودية، 1995.

## المراجع باللغة الفرنسية:

(3 ) Anne-Marie D , J.P.Indjehagopian. « Méthodes Statistiques Appliquées à la Gestion ». Ed d'ORGANISATION, 1981.

(4 ) Gourieroux, Ch et Roy, G. « Enquete deux vagues, renouvellement de l'échantillon ». Annales de l'INSEE, VOL29, pp115-134. 1978.

(5 ) Gourieroux, Ch et Roy, G. « Renouvellement tomporel d'un échantillon » .Anales de l'INSEE.1978.

(6 ) Jean-Jacque Drosberge. Bernard Fichet et Philiphe Tassie , « Les Sondages ». ED ECONOMICA. Paris.1987.

(7 ) Jean-Jacque Drosberge. « Eléments de Statistiques ». OPU .Algérie, 1988.

(8 ) Jean .M. Martel , et Raymond Nadeau . « Statistiques en gestion et en économie ». GAETAN MARIN édition. 1993.

## المراجع باللغة الإنجليزية:

(9 ) Goldstein. H (1995), multi level statistical models, Edward Arnold.

(10 ) Harry Frank , Steven C Althoem. « Statistics Concepts and applications ». Cambridge University Press, 1994.

(11 ) Gurney and Daly. " A multivariate approach to estimation in periodic sample survey". ASA, Social statistics section, pp242-257 .1965.

(12 ) H Liao , J. Sedrank ." Selection of strata sample sizes for the comparison of domain means". JASA, vol 78 , n°384, 12/1983.

(13 ) Javed Ahmed , Charles. D. Bonham. "Optimum Allocation in two stage sampling ".Journal of range management.35(8), November 1982.

(14 ) Lawrance .L. Lapin. « Statistics for Modern Business Decisions ». HARCOUT BRACE JOVANOVICH , INC, 1973.

(15 ) Leslie Kish. « Survey Sampling ». John Wiley and Son INC.USA, 1965.

(16) Magdalena Mok, . Note from "multi-level models" project of "Magdalena.M". university. of. California 1995 unpublished paper .

(17 ) Muller, K.E and Benignus, V.A. " scientific power with statistical power ". Neuro-toxicology and Teratology , n°14, 211-219. 1995.

(18 ) Paterson . "Sampling on successive occasions with partial replacement of units". J ROY Stat Doc, VOL12, pp241-255. 1950.

(19) Russell V Lent . « Somme Practical guidelines for sample size determination ». University of IOWA. 03/2001. unpublished document .

(20 ) Sedrank, KJ. " A double sampling scheme for Analytical surveys".JASA, n°60, 985-1004 .1965.

(21 ) Taylor, D and Muller, K.E ." computing confidence bouns for power and sample size for the general linear univariate model".The American statitician, n°49, pp 43-47.1995.

(22 ) Thomas, L ." retrospective Power Analysis". Conservation Biology, n°11, pp276-280. 1997.

(23) Thomas, L. "Statistical power analysis software". http.www.foresty.ubc.ca/conservation/power/ .1998

(24 ) Walter, K.M. "Estimation in finite populations". JASA, Vol 74, pp604-613. 1971.

(25) Hansen, Hurwitz and Madow , "Sample survey Method and theory". Vol 1 et 2 , John Wiley and Sons, 1993

## الفهـــــرس

**موســى نبيــل سميــــر**

## ملــخــص

تعتبر المسوحات الإحصائية بالمعاينة (échantillonnage) بمختلف أنواعها و أساليبها جد ضرورية كونها تساعد الباحثين في العديد من المجالات سواء كانت اقتصادية، اجتماعية، سياسية أو حتى كلينيكية (طبية)...إلخ، في إعداد دراساتهم الميدانية، و بشكل عام، تمكن صانعي القرارات كل في مجاله من اعتماد سياسات وتدابير موضوعية و عقلانية و ذات مصداقية كبيرة. و لكي تكون أي دراسة إحصائية الأحسن، يجب أن يتم التخطيط لها بحذر، وللتخطيط الجيد عدة أوجه، إذ يجب أن يحدد المشكل بوضوح، كما يجب أن نجعله أكثر عمليا. من المهم جدا أن تسحب وحدات التجربة أو الوحدات المشاهدة من المجتمع المناسب، كما يجب أن تتبع الإجراءات بحذر وأن تستعمل الوسائل الضرورية من أجل الحصول على القياسات المطلوبة. أخيرا، من المهم جدا أن يكون للدراسة حجم عينة يتناسب وأهداف هذه الدراسة، إذ يجب أن يكون حجم العينة هذا كبير لدرجة أن تكون للنتائج دلالة علمية وإحصائية في نفس الوقت.

**الكلمات المفتاحية:**

حجم عينة؛ مسح إحصائي؛ دقة التقدير؛ المحاصة المثلى؛ الانحراف المعياري؛ دالة التكلفة؛ الطبقية؛ المعاينة العنقودية؛ المسح الاستطلاعي؛ نايمن.