مـــركـــز الإحـــصـــاء
STATISTICS CENTRE

دليل المعاينة الإحصائية

أدلـــة المنهجيـــة والجـــودة - دليل رقم (1)

www.scad.ae

قائمة المحتويات

المقدّمة 3

1 | مفاهيم أساسية في العينات 4

2 | أساليب المعاينة الإحصائية 9

3 | تقدير حجم العينة 20

4 | أساليب المعاينة في مركز الإحصاء – أبوظبي 28

المراجع 36

المقـدّمــة

يأتـي اعـداد دليـل المعاينـة الإحصائيـة فـي مركز الإحصاء – أبوظبي، ضمـن إطـار عمـل إدارة المنهجيـات والجـودة والتحليـل علـى توثيـق الأدلـة الخاصـة بالعمليـات الإحصائيـة. ويهـدف هـذا الدليـل إلى اطـلاع الإحصائييـن الفنييـن داخـل المركـز، ومسـتخدمي البيانـات خـارج المركـز علـى تفاصيـل إجـراءات تصميم وسـحب عينـات المسـوح الإحصائيـة التـي ينفذهـا المركـز سـواء كانـت هـذه المسـوح اقتصاديـة اوأسرية أو غيرها.

يتضمـن هـذا الدليـل أربعـة فصـول رئيسـية، الأول يسـتعرض المفاهيـم والمصطلحـات الإحصائيـة الخاصـة بأسـاليب وطـرق المعاينـة الإحصائيـة، تأتـي هـذه المفاهيـم مـن واقـع الاسـتعراض المرجعـي لنظريـة العينـات ومـن واقـع أدلة ومعاجـم المصطلحـات الإحصائيـة المعتمـدة لـدى المنظمـات الدوليـة والإقليميـة.

أما الفصـل الثانـي، فيتعـرض إلى أسـاليب وطـرق المعاينـة الإحصائيـة المختلفـة، هـذا إضافـة إلى محـددات ومزايـا اسـتخدام كل منهـا ومعاييـر اختيـار الأسـلوب الأمثـل. فـي حيـن، يتنـاول الفصـل الثالـث تقديـرات حجـم العينـة بحسـب أسـاليب تصميـم المعاينـة المسـتخدمة كمـا ويعـرض المتغيـرات الرئيسـية التـي يجـب توفرهـا لتحديـد حجـم العينـة.

أخيـرا يتنـاول الفصـل الرابـع وصفـا إحصائيـا لأطـر المعاينـة الإحصائيـة المعتمـدة فـي المركـز سـواء كانـت هـذه الأطـر اقتصاديـة أو أسرية أو غيرهـا. إضافـة إلى تصاميـم المعاينـة الإحصائيـة المسـتخدمة فـي المركـز وآلية تحديـد أحجام العينـات، إضافـة إلى الطـرق والأسـاليب العشـوائية لسـحب العينـات.

1 | مفاهيم أساسية في العينات

يتناول هذا الفصل المفاهيم الأساسية والتعاريف التي تتعلق بالجوانب النظرية والتطبيقية الخاصة بتصميم وسحب العينات، وهي تنسجم مع المفاهيم والتعاريف الدولية في هذا المجال:

المجتمع الإحصائي (Population):

جميع الوحدات الإحصائية التي يراد إجراء البحث الإحصائي عليها، ومن الضروري تعريف هذه الوحدات بشكل واضح بحيث تجمعها صفة واحدة أو صفات مشتركة. ومعظم المجتمعات الإحصائية مؤلفة من وحدات إحصائية تتغير حسب الزمن (مجتمعات متجددة)، وبعضها الأخر مجتمعات ثابتة لا تتغير حسب الزمن.

المجتمع المستهدف (Target Population):

هو جميع الوحدات الإحصائية التي يراد إجراء البحث الإحصائي عليها، ومن الضروري تعريف هذه الوحدات بشكل واضح بحيث تجمعها صفة واحدة أو صفات مشتركة. ومعظم المجتمعات الإحصائية مؤلفة من وحدات إحصائية تتغير حسب الزمن (مجتمعات متجددة)، وبعضها الأخر مجتمعات ثابتة لا تتغير حسب الزمن.

المسح الإحصائي (Statistical Survey):

هو عمل إحصائي منظم مبني على أسس علمية ويقوم على مبدأ شمول جزء من المجتمع الإحصائي وتختار المفردات في الغالب باعتماد أحد أساليب المعاينة الاحتمالية، أو شمول جميع وحدات المجتمع وإخضاعها للمشاهدة من خال المسح الشامل.

العد الشامل (complete Census):

هو العمل الإحصائي المنظم الذي يقوم على مبدأ الشمول لكل مفردات المجتمع الإحصائي بعملية جمع البيانات وإخضاعها للمشاهدة الإحصائية، وعادة يجري العد الشامل في التعددات كالتعداد السكاني والتعداد الزراعي والتعداد الصناعي، وأحيانا يتم العد الشامل إذا كان المجتمع المنوي دراسته مجتمع صغير حيث يكون أسلوب المعاينة غير فعال. أيضا إذا كان الباحث يجهل طبيعة المجتمع من الممكن أن يجري عد شامل له بدلا من المعاينة.

أسلوب المعاينة (Sampling Techniques):

هو أسلوب يستخدم لاختيار مفردات من المجتمع وإخضاعها للعمل الإحصائي، بحيث تكون النتائج التي يتم التوصل إليها بناء على معطيات العينة تمثل مؤشرات المجتمع المراد تقديرها.

الاختيار العشوائي (Random Selection):

هي عملية اختيار مفردات من المجتمع الإحصائي بطريقة تبعد أي تحكم شخصي للتدخل في اختيار أو استبعاد أي مفردة من مفردات المجتمع، مع ضمان إعطاء فرصة متساوية للمفردات كافة لأن تظهر في العينة المنتقاة.

الإطار (Frame):

قائمة أو سجل يشمل جميع وحدات المجتمع الإحصائي، ويتضمن عادة أسماء وعناوين الوحدات الإحصائية وبعض المعلومات المتعلقة بها، والإطار هو الدليل أو مجموعة الوثائق التي تساعدنا في الوصول إلى الوحدات الإحصائية لجمع البيانات عنها.

تصميم العينة (Sample Design):

هي عملية اختيار التركيب المناسب من عدة أنواع من العينات للوصول إلى العينة التي تحقق النتائج المرجوة منها.

العينة (Sample):

جزء من المجتمع الإحصائي يتم اختياره وفق أساليب المعاينة الإحصائية ويشترط أن تكون ممثلة للمجتمع الذي نقوم بدراسته، ولكي تكون العينة ممثلة للمجتمع يجب أن تتضمن خصائص المجتمع بشكل يمكننا تعميم نتائجها لتقدير أهم معالم المجتمع الإحصائي.

أنواع العينات (Types of Samples):

تقسيم العينات الإحصائية إلى قسمين رئيسيين:

أولاً: عينات احتمالية (Probability Samples):

يتم سحبها على أساس قانون الاحتمالات حيث يتم سحب مفرداتها بشكل متتالي وباحتمال معروف، ومن أنواعها العينة العشوائية البسيطة Random Sample والعينة الطبقية Stratified Sample والعينة المنتظمة Systematic Sample والعينة العنقودية Cluster Sample.

إن أهم ما يميز هذا النوع من العينات هو أنه يمكن تعميم نتائج العينة على كافة وحدات المجتمع من خلال حساب معاملات رفع أو أوزان، بحيث يعتمد مقدار الوزن لوحدة العينة على احتمال سحب تلك الوحدة من المجتمع. أيضا تمكننا العينات الاحتمالية من تحليل نتائج العينة وحساب الأخطاء المعيارية، ومعامات التغير إضافة إلى أثر التصميم، وبالتالي فأن العينة الاحتمالية تعطي نسبة الخطأ في التقديرات الناتجة عنها وتعطي مقدار الثقة في التقديرات.

وعليه فإن مؤشرات الإحصاءات الرسمية تعتمد على تصاميم عينات احتمالية ترفع نتائجها على مستوى المجتمع ككل وتقدر قيم ونسب الأخطاء فيها.

ثانياً: عينات غير احتمالية (Non-Probability Samples):

يتم اختيار وحداتها وفقا لمعايير يضعها الباحث دون التقيد بقوانين الاحتمالات ومنها العينة العمدية والعينة الحصصية quota Sample. هذا النوع من العينات يطبق غالبا في مسوح استطلاعات الرأي، وفي دراسات بعض الظواهر المحدودة ضمن المجتمعات، وهي تعطي نتائج تأشيريه تعتمد على بيانات تمثل وحدات العينة وليس المجتمع ككل.

العينة المتناسبة مع الحجم (Sampling Proportional to the Size):

هي العينة التي يكون احتمال ظهور كل وحدة معاينة فيها يتناسب مع حجم تلك الوحدة للصفة المدروسة، فمثا المنشأة الاقتصادية يقاس حجمها بعدد العاملين، وعند اختيار عينة المنشآت الاقتصادية بأسلوب المعاينة المتناسبة مع الحجم تعطى احتمالية أو فرصة أكبر لظهور المنشآت ذات الحجم الكبير أي ذات عدد العمال الأكبر.

العينة متعددة الأغراض (Multi-Purpose Sample):

هي العينة التي يتم من خلالها جمع بيانات لأكثر من موضوع في مسح إحصائي واحد مثل الدخل والإنفاق والصحة والتغذية.

العينة المتوالية (Successive Sample):

هو أحد أساليب المعاينة الذي يغطي المجتمع على عدة سنوات حيث يقسم المجتمع إلى عدة أجزاء، بحيث يغطى كل جزء في سنة، وعادة يستخدم هذا الأسلوب في المجتمعات الصغيرة كبديل عن التعدادات.

العينة الاستطلاعية (البحث التجريبي) Pilot Survey:

اختيار عدد من الوحدات الإحصائية وجمع البيانات عنها، وتدوينها في استمارات متخصصة لهدف اختيار دقة تصميم الاستبانة، والوقوف على الصعوبات التي قد يواجهها الباحثون عند تنفيذ البحث.

العينة المتطابقة (Matched Sample):

هي العينات التي لها نفس المساهمة أو المتغير والذي يقاس مرتين لكل موضوع وتحت ظروف مختلفة وبصورة عامة تدعى بالقياسات المكررة. أي بمعنى آخر تكون العينتان متطابقتان عند وضعهما كأزواج متشابهة أو متطابقة.

تصميم العينة الموزونة ذاتيا (Self-Weighting Sample):

هو تصميم المعاينة الذي يكون فيه وزن وحدة المعاينة (معامل التكبير) متساوي لجميع الوحدات في العينة.

العينة الملائمة (Convenient Sample):

هي أسلوب معاينة غير احتمالي، يتم من خلاله سحب عينة ممن يحملون صفة معينة في المجتمع، ويؤدي هذا الأسلوب إلى وجود حذف أو تكرار في إطار المعاينة، يمكن التقليل من مخاطره باستخدام أسلوب المعاينة الحصصية.

وحدة المعاينة (Sampling unit):

هي المفردة أو الوحدة التي تشكل عنصرا في المجتمع الإحصائي الذي يخضع لعملية العد أو عملية المعاينة، أي هي الوحدة التي تجمع عنها البيانات أو المعلومات الإحصائية المطلوبة.

وحدة المعاينة الأولية (Primary Sampling Unit):

هي وحدات المعاينة التي تسحب في المرحلة الأولى من تصميم عينة متعددة المراحل، وغالبا ما تمثل وحدة المعاينة الأولية عنقودا، وتكون وحدة المعاينة الأولية مجموعة من وحدات المعاينة الثانوية.

وحدة المعاينة الثانوية (Secondary Sampling Unit):

هي وحدات المعاينة التي يتم سحبها في المرحلة الثانية من تصميم عينة متعددة المراحل وتعتبر كل وحدة معاينة ثانوية جزءا من وحدات المعاينة الأولية.

وحدة التحليل (Analyzing Unit):

هي الوحدة التي تستخدم في تحليل البيانات الإحصائية التي يتم جمعها لتحقيق اهداف المسح الإحصائي. ومن الممكن أن تكون وحدة التحليل هي ذاتها وحدة العد المستخدمة في المعاينة أو قد تكون غير ذلك.

أخطاء عدم التغطية (Non – coverage Errors):

أخطاء تحصل نتيجة عوامل عديدة أهمها ما يحصل اثناء اعداد الإطار من نقص في الشمول أو زيادة فيه، أو ادخال مفردات غريبة عن المجتمع أو عدم توصيف المفردات في الإطار بشكل صحيح. تنقسم أخطاء عدم التغطية إلى نوعين: الأول النقص في التغطية أي عدم شمول مفردات ينبغي شمولها. والثاني أخطاء الزيادة في التغطية أي شمول مفردات ينبغي عدم شمولها.

أخطاء عدم الاستجابة (Non-Response Errors):

هي عدم استجابة بعض المستجوبين في العينة للإجابة على الاستمارة ككل، وتكون عدم الاستجابة نتيجة للرفض أو اية أسباب أخرى، ويندرج هذا النوع من عدم الاستجابة تحت اسم عدم استجابة كلية لوحدة المعاينة، أما عدم الاستجابة الجزئية فهي تكون عندما يرفض المستجوب الإجابة على أسئلة معينة دون غيرها.

الخطأ العشوائي (Random Error):

ويمكن التعرف عليه من مشاهدة انتشار نتائج البحث إذا تكرر إجراءه بنفس الأسلوب وتحت نفس الظروف. وهذا الخطأ لا يختفي عند استخدام أسلوب الحصر الشامل وذلك لأنه ينتج عن اختلاف العدادين أو اختلاف الدافع الشخصي للإجابة على أسئلة البحث وفي معظم الأحيان يكون هذا الخطأ ضئيا ويمكن قياسه ومعرفة حدوده. ويتوقف مقدار هذا الخطأ على عاملين أساسيين هما مدى الاختلاف أو التباين بين وحدات المجتمع وحجم العينة بالنسبة للمجتمع الذي سحبت منه فكلما ازداد التباين بين وحدات المجتمع ازداد احتمال الوقوع في الخطأ العشوائي، أما بالنسبة لحجم العينة فكلما كبر حجم العينة انخفض احتمال الوقوع في هذا الخطأ.

خطأ التحيز (Bias Error):

(أ) خطأ التحيز في التقدير:

وهو انحراف متوسط جميع التقديرات الممكنة لمعلمة المجتمع عن قيمتها الحقيقية، ومن الصعب اكتشاف هذا الخطأ والتخلص منه إلا بأجراء تعديات جذرية على تصميم الدراسة أو طريقة جمع البيانات أو تعديل النتائج.

(ب) خطأ التحيز في المعاينة:

وهو التحيز الذي يكون مقصودا، وينشأ بسبب الإدلاء بمعلومات لا تطابق الواقع من قبل المزود بالبيانات، أو بسبب خلل من قبل مصممي المسح وفقا لميول أو أغراض مقصودة، وأما أن يكون التحيز غير مقصود، وهو الذي يتسرب إلى المسح لعدم فهم المبحوث للبيانات المطلوب تقديمها أو لعدم إتاحة الفرصة لتحضير إجابات صحيحة.

الخطأ المعياري (Standard Error):

هو عبارة عن الجذر التربيعي لتباين العينة المقدر. وتباين العينة هو عبارة عن متوسط مربعات الفروق ما بين قيم وحدات العينة وقيمة المتوسط الحسابي لتلك الوحدات.

خطأ العينة (Sample error):

هو الفرق بين تقديرات العينة ومعلمات المجتمع الحقيقية، وكلما كان اختيار العينة جيدا كلما قل خطا العينة وأصبحت أكثر تمثيل للمجتمع.

الخطأ المعياري النسبي (Relative Standard Error):

هو الخطأ المعياري مقسوما على المعلمة المحسوب لها الخطأ المعياري وهو مساوي لمعامل التباين.

دقة المسح بالعينة (Accuracy of the sample Survey):

هو الفرق بين تقدير الصفة المدروسة عن طريق العينة والقيمة الحقيقية للصفة المدروسة في المجتمع، وكلما كان الفرق قليا كلما كانت الدقة أعلى.

نقطة البداية العشوائية (Random Start):

هو رقم عشوائي يتم اختياره من فترة المعاينة)من 1 إلى K (وذلك باستخدام جداول الأرقام العشوائية ليمثل نقطة البداية العشوائية لاختيار عينة عشوائية منتظمة.

التوزيع الأمثل (Optimum allocation):

هو أحد أساليب توزيع وحدات المعاينة الطبقية على مختلف الطبقات بحيث تكون حصة كل طبقة تتناسب طرديا مع حجم الطبقة والتباين داخل الطبقة وعكسيا مع كلفة جمع بيانات وحدة المعاينة في تلك الطبقة.

توزيع نيمان (Nyman Allocation):

هو أحد أساليب توزيع المعاينة الطبقية على مختلف الطبقات بحيث تكون حصة كل طبقة تتناسب طرديا مع كل من حجم الطبقة ومقدار التباين ضمن الطبقة الواحدة.

التوزيع المتناسب (Proportional Allocation)

هو أحد أساليب توزيع وحدات المعاينة الطبقية على مختلف الطبقات بحيث تكون حصة كل طبقة تتناسب طرديا مع حجم تلك الطبقة من حيث عدد وحدات المعاينة.

الثقة في تقدير العينة (confidence in Sample Estimate):

تقيس مدى الاعتماد على النتائج المقدرة عن طريق العينة، وتزداد الثقة في النتائج كلما ازداد حجم العينة، وتقترب من المتوسط الحقيقي أو معلمات المجتمع الأخرى.

حد الخطأ (Bound of error):

هو قيمة الخطأ المعياري مضروبة بقيمة Z أو قيمة t الجدولية عند حدود ثقة معينة.

تأثير التصميم (Design Effect):

هو نسبة تباين أحد التقديرات في تصميم عينة معينة إلى تباين عينة عشوائية بسيطة من نفس الحجم.

الترجيح أو التوزين (Weighting):

هو إجراء العمليات الحسابية للوصول إلى قيم المعلمات المقدرة للمجتمع باستخدام نتائج المسح بالعينة.

2 | أساليب المعاينة الإحصائية

تتضمن نظرية العينات عدة طرق وأساليب للمعاينة الإحصائية، وجميع هذه الأساليب تركز على الحصول على عينة إحصائية تفرز مؤشرات وتقديرات بأقل خطأ معاينة ممكن، مع الأخذ بالاعتبار أيضا حجم العينة المقدر. يتناول هذا الفصل طرق وأساليب المعاينة الإحصائية.

1.2 أسلوب المعاينة العشوائية البسيطة

يعتبر أسلوب المعاينة العشوائية البسيطة من أبسط الطرق وأكثرها انتشارا في أساليب المعاينة، يمتاز هذا الأسلوب بأنه يعطي كل وحده من وحدات المعاينة الموجودة في المجتمع فرص أو احتمالات متساوية للاختيار أو الظهور بالعينة.

هذا ويندرج ضمن المعاينة العشوائية البسيطة أسلوبين لسحب العينة، الأول في حالة سحب العينة وإرجاعها إلى المجتمع بعد عملية السحب لإعطائها فرصة الظهور مرة أخرى ويسمى أسلوب سحب العينة مع الإرجاع، أما الأسلوب الثاني فهو في حال سحب العينة واستثناء كل عينة أخرى مسحوبة من المجتمع يسمى سحب عينة دون إرجاع حيث أن هذا الأسلوب لا يعطي فرصة لتكرار ظهور العينة الواحدة أكثر من مره.

1.1.2 شروط استخدام العينة العشوائية البسيطة:

نظرا لسهولة وبساطة تطبيق هذا الأسلوب من العينات فهو منتشر بشكل كبير، ولكن قبل استخدام هذا الأسلوب يجب ملاحظة أن هناك شروط يجب أن تتوفر في وحدات المعاينة للمجتمع المستهدف في الدراسة، إذ يجب أن تكون وحدات المعاينة في المجتمع متجانسة بالنسبة للصفة المدروسة، أي أن التباين بين وحدات المعاينة في المجتمع للصفة المدروسة قليل نسبيا، هذا بالإضافة إلى المحددات الأخرى من قبل مستخدمي البيانات مثل مستويات النشر، فمثا إذا كان مطلوب استخراج النتائج على مستوى المنطقة فا يجوز سحب عينة عشوائية بسيطة على مستوى الإمارة ككل، كذلك وجود محدد التكاليف والإمكانيات الأخرى مثل توفر الكوادر الميدانية المدربة وتكاليف التنقل داخل المجتمع، مما يحدد من استخدام هذا الأسلوب. إضافة إلى ما سبق أن تأثر هذا الأسلوب بوحدة تكلفة المسح قد يكون أكثر من معظم أنواع العينات الأخرى، لذا يجب أخذ جانب الحيطة عند استخدام أسلوب المعاينة العشوائية البسيطة.

2.1.2 طريقة سحب العينة العشوائية البسيطة.

عند اختيار أسلوب العينة العشوائية البسيطة، لابد من اختيار طريقة تضمن العشوائية في اختيار وحدات المعاينة، وتعطي كل وحدة من وحدات المجتمع فرصة متساوية للظهور. في الجانب التطبيقي هناك عدة طرق مستخدمة لسحب عينات عشوائية بسيطة، أحد هذه الطرق إعطاء رقم متسلسل لكافة وحدات المجتمع الخاضعة للمعاينة، ومن ثم اختيار رقم عشوائي من جدول الأرقام العشوائية أو اليا. ومن ثم يتم سحب وحدة المعاينة التي تتطابق مع الرقم العشوائي، أما إذا كانت قيمة الرقم العشوائي خارج إطار تسلسل وحدات المعاينة فيتم اعادة سحب رقم آخر... وهكذا.

مثال (1):

إذا كان لدينا مجتمع مكون من 60 منشأة واردنا سحب عينة عشوائية بسيطة مقدارها 4 منشات. في هذه الحالة نعطي المنشات في المجتمع ارقام متسلسلة بدءا من 01، 02، 03،..... 60. ومن ثم يتم اختيار رقم عشوائي ذو منزلتين عشريتين من جدول الأرقام العشوائية أو باستخدام الحاسوب، وليكن الرقم 45، في هذه الحالة تختار المنشاة ذات التسلسل 45 ضمن العينة، وفي السحبة الثانية إذا كان الرقم العشوائي 73 في هذه الحالة يتم تجاهل هذا الرقم ويعاد سحب رقم آخر... وهكذا، حتى نحصل على 4 منشات كعينة عشوائية مطلوبة.

2.2 أسلوب المعاينة العشوائية المنتظمة

أسلوب العينة العشوائية المنتظمة هو أحد أساليب المعاينة العشوائية التي تمتاز بالسهولة والبساطة في التطبيق، إضافة إلى أنه يضمن انتشار العينة على أكبر مساحة من المجتمع بسبب أن أسلوب السحب يتم وفق انتظام متسلسل. تعتبر المعاينة المنتظمة الخطية هي الأسلوب الأكثر شيوعا في العينات المنتظمة، ويتلخص أسلوب تطبيقها بما يلي:

افرض أن المجتمع يتكون من N من وحدات المجتمع وأن حجم العينة المطلوب سحبها هو n فإذا ما قسمنا حجم المجتمع N على حجم العينة المطلوب n نحصل على المقدار k حيث $nk = N$ ويعرف إحصائيا مقدار k بفترة الانتظام، بعد ذلك يجري اختيار رقم عشوائي يقع بين 1 و k يسمى هذا الرقم برقم البداية العشوائية ويرمز له بالرمز I، يكون الرقم المتسلسل للعينة الأولى هو (I) والعينة الثانية هي $k + I$ والثالثة $2k + I$..الخ.

مثال (2):

لتقدير عدد عمال المنشات ضمن إقليم معين يحتوي على 400 منشأة سحبت عينة منتظمة من تلك المنشات حجمها 16.

يكون أسلوب سحب العينة كما يلي:

$$N = 400 \qquad n = 16$$

$$k = N / n = 400 / 16 = 25$$

يتم سحب رقم عشوائي يقع بين 1 و 25 وليكن الر قم 14، بذلك تكون الأرقام المتسلسلة لوحدات المعاينة المسحوبة في العينة هي:

$$14,\ 14 + 25,\ 14 + 2 \times (25),\ 14 + 3 \times (25) \ldots$$

وتساوي: 14، 39، 64، 89....

1.2.2 مزايا ومحددات استخدام العينة العشوائية المنتظمة.

- من أهم مزايا العينة المنتظمة هو سهولة سحب العينات، وتوزيع العينة على المجتمع بشكل جيد، وتعتبر العينة المنتظمة فعالة مقارنة بالعينة العشوائية البسيطة لكثير من المجتمعات وخاصة إذا كان ميل الصفة المدروسة في المجتمع خطي.
- أما من محددات هذه العينة هي صعوبة الحصول على تقدير متحيز للتباين، كذلك في حالة وجود صفة دورية في المجتمع قد يؤدي إلى وجود التحيز في العينات المسحوبة وفي التقدير فمثا ضمن مجموعة من الأسر تم سحب عينة فرد واحد من كل أسرة وكان رقم البداية هو 1 وفترة العينة هي 2 وكان الأفراد الذي تسلسلهم 3 هم المختارين في العينة، سنلاحظ أن جميع افراد العينة هم ذوو الترتيب الثالث في الأسرة، وإذا كان ترتيب الأفراد في الأسرة الأب، الأم، الابن، الابنة، سنلاحظ أن العينة جميعها أبناء مما يؤدي إلى تحيز في التقديرات المطلوبة.

3.2 أسلوب المعاينة المتناسبة مع الحجم

كما ذكر سابقا فإن أهم ما يميز العينة العشوائية البسيطة هو أن احتمالية ظهور كل وحدة من وحدات المعاينة للصفة المدروسة تكون متساوية لجميع وحدات المجتمع، فعلى سبيل المثال عند سحب عينة من المنشآت الاقتصادية بأسلوب المعاينة العشوائية البسيطة، تكون جميع المنشآت وبأحجامها المختلفة سواء كانت متناهية الصغر أو كبيرة لها نفس الفرصة في الظهور بالعينة. عند دراسة كثير من الصفات بمختلف الظواهر قد تتطلب طبيعة الصفة المدروسة إعطاء احتمال أكبر لظهور وحدات معينة من المجتمع في العينة المسحوبة. فمثا قد تكون طبيعة الصفة المدروسة في المنشآت الاقتصادية تتطلب إعطاء فرصة أكبر لاختيار المنشآت ذات الحجم الأكبر)عدد العمال الأكثر(، في هذه الحالة يتم اتباع أسلوب المعاينة المتناسبة مع الحجم.

كذلك يعرف أسلوب سحب العينة المتناسبة مع الحجم بأسلوب التجميع التراكمي، ويتلخص هذا الأسلوب بإعطاء كل وحدة معاينة رقم يوازي الصفة التي تحملها.

مثال (3):

ضمن إطار المنشات الاقتصادية، المنشأة التي تتضمن 1000 عامل يعتبر وزنها 1000 أي أنها تحتوي على 1000 وحدة معاينة فرضية والمنشأة التي تتضمن 100 عامل يكون وزنها 100 وهكذا.

وللتوضيح من خلال الجدول التالي:

المنشأة	عدد العمال	عدد العمال التجميعي	الأرقام التي ترافقها
1	1000	1000	1-1000
2	700	1700	1700 - 1001
3	1200	2900	2900-1701
4	500	3400	3400 - 2901
5	300	3700	3700 - 3401
6	800	4500	4500 - 3701

لاختيار ثلاث منشات يتم سحب ثلاثة أرقام عشوائية تقع بين 1 و4500 باستخدام الجدول العشوائي أو الحاسوب، ومن ثم يتم حصر الرقم العشوائي على عمود عدد العمال التجميعي في الجدول أعاه ويتم اختيار المنشأة التي يكون فيها عدد العمال التجميعي أكبر من أو يساوي الرقم العشوائي. ففي هذا المثال إذا اختيرت الأرقام العشوائية 4000، 2000، 75 فإن منشات العينة تكون المنشآت ذات الأرقام 6، 3، 1. على التوالي.

عندما يكون المجتمع كبير نسبيا فإن عملية سحب العينة بالطريقة أعلاه تستغرق وقتا طويلا، لذا هناك أسلوب أو طريقة أخرى لسحب العينة المتناسبة مع الحجم، وتعرف بطريقة لاهير (Lahiri) حيث يتضمن هذا الأسلوب سحب أزواج من الأرقام العشوائية، يمثل الرقم الأول من كل زوج رقم وحده المعاينة بحيث يسحب الرقم العشوائي بين 1 وN حيث N تمثل عدد وحدات المعاينة الإجمالية في المجتمع ويمثل الرقم الثاني في الزوج حجم وحدة المعاينة بحيث يسحب الرقم العشوائي الثاني بين 1 وM حيث M تمثل حجم أكبر وحدة معاينة موجودة في المجتمع للصفة التي يتم الوزن على أساسها.

4.2 أسلوب المعاينة العشوائية الطبقية

كما بينا سابقا، يعتمد تطبيق أسلوب المعاينة العشوائية البسيطة على شرط التجانس بين جميع وحدات المجتمع وغير ذلك فإن العينة ستؤدي إلى نتائج متحيزة وغير دقيقة، أضف إلى ذلك فإن العينة ستنتشر بطريقة عشوائية على منطقة واسعة من المجتمع مما يؤدي إلى زيادة حجم العبئ والكلفة في عملية جمع البيانات. ونظرا لصعوبة تحقق التجانس في كثير من المجتمعات، يلجأ إلى تطبيق أسلوب المعاينة العشوائية الطبقية، حيث يقسم المجتمع إلى عدد من المجموعات غير المتداخلة، كل مجموعة تكون متجانسة للصفة المدروسة وتسمى طبقة، وذلك بهدف الحصول على نتائج أكثر دقة. فمثلا عند دراسة متوسط دخل الأسرة أو المستوى التعليمي يمكن تقسيم المجتمع إلى ريف وحضر، وللحصول على نتائج جيده واستخدام أسلوب المعاينة الطبقية بفاعلية عالية يجب أن يراعى الدقة وخاصة عند إجراء الأمور التالية:

- تكوين الطبقات.
- عدد الطبقات المراد عملها.
- حجم العينة في كل طبقة.
- تحليل البيانات لتصميم العينة الطبقية.

مثال (4): إذا كان لدينا ثماني منشآت وكان عدد العمال في هذه المنشآت كما يلي:

رقم المنشأة	عدد العمال	رقم المنشأة	عدد العمال
1	3000	5	6000
2	1500	6	1200
3	7000	7	4500
4	2500	8	5000

للحصول على تقدير معدل عدد العمال في المنشأة الواحدة. يمكن من خال أسلوب المعاينة البسيطة سحب ثاث عينات، فإذا كانت العينة المسحوبة هي المنشآت التي أرقامها 6، 2، 1 فإن معدل عدد العمال هو 1900 عامل بينما المعدل الحقيقي للمجتمع هو (3838) أي ما يزيد على ضعف المعدل المقدر بالعينة، من الواضح وجود خلل نتيجة استخدام العينة العشوائية البسيطة لمجتمع غير متجانس.

إذا ما ذهبنا باتجاه تطبيق أسلوب المعاينة الطبقية، من خلال تقسيم هذا المجتمع إلى طبقات حسب فئات عدد العمال كما يلي:

الطبقة الأولى	اقل من 3000
رقم المنشأة	عدد العمال
2	1500
4	2500
6	1200
الطبقة الثانية	5000 - 3000
رقم المنشأة	عدد العمال
1	3000
7	4500
الطبقة الثالثة	فأكثر 5000
رقم المنشأة	عدد العمال
3	7000
5	6000
8	5000

فإذا ما تم سحب عينة واحدة من كل طبقة من الطبقات أعلاه وكما هو موضح ادناه:

رقم الطبقة	رقم المنشأة	عدد العمال
1	2	1500
2	7	4500
3	5	6000

فإذا ما تم سحب عينة واحدة من كل طبقة من الطبقات أعاه وكما هو موضح ادناه:

رقم الطبقة	رقم المنشأة	عدد العمال
1	2	1500
2	7	4500
3	5	6000

إن المعدل المقدر لعدد العمال = معدل العينة = $(3/8 \times 1500) + (2/8 \times 4500) + (3/8 \times 6000) = 3938$

يلاحظ أن هذه النتيجة قريبة جدا من المعدل الحقيقي للمجتمع لذا فإن استخدام العينة الطبقية في مثل هذه الحالات يعتبر ضرورة البد منها.

1.4.2 مبادئ تقسيم المجتمع إلى طبقات:

عند تقسيم المجتمع إلى طبقات يجب مراعاة النقاط التالية:

- أن يكون مجموع وحدات المعاينة لجميع الطبقات مساوي لمجموع وحدات المعاينة للمجتمع ويجب ألايكون تداخل بين الطبقات.
- أن تكون وحدات المعاينة في داخل كل طبقة متجانسة بالنسبة للصفة المدروسة.
- الحصول على نتائج حسب تقسيمات جغرافية أو إداريه (إقليم، منطقة) يترتب عليه تقسيم المجتمع إلى طبقات جغرافية بحسب مستويات التمثيل المطلوبة.
- مراعاة الطرق العلمية عند تقسيم المجتمع إلى فئات حيث تعتبر كل فئة طبقة.

أسلوب تقسيم المجتمع إلى طبقات:

تعتبر عملية تقسيم المجتمع إلى فئات كنوع من التقسيم الطبقي للمجتمع، حيث يمكن اعتبار كل فئة من الفئات طبقة، ويهدف هذا الأسلوب إلى تقليل التباين بين مفردات المجتمع للصفة المدروسة، فمثا إذا كانت الدراسة عن معدل دخل الأسرة فيمكن تقسيم الأسر إلى فئات حسب مستويات الدخل، وإذا كانت الدراسة عن عدد العاملين في المصانع يمكن تقسيم المصانع إلى فئات حسب عدد العاملين وهكذا.

أما تحديد عدد الفئات المقترح فهناك عدة طرق نذكر منها استخدام القانون التالي:

عدد الفئات = $3 + 1 \ (Log\ 1850)$

مثال (5):

إذا كان عدد المنشآت في قطاع معين في إمارة أبوظبي هو 1850 منشأة فإن عدد الفئات التي يمكن أن يقسم المجتمع إليها هي:

$3 + 1 \ (Log\ 1850)$

$= 10.8 => 11$ فئة

أما الأسلوب الآخر لتقسيم المجتمع إلى طبقات فهو استخدام أسلوب Cochran والذي يأخذ عدد وحدات المعاينة ووزن كل وحدة بالاعتبار، عندما يكون عدد الطبقات المطلوب محدد مسبقا. ويمكن توضيح ذلك بالمثال التالي:

مثال (6):

إذا كان عدد المنشآت في المجتمع هو 1850 مقسمة إلى فئات مبدئية حسب عدد العمال إلى 12 فئة والمطلوب إعادة تقسيم هذه المنشآت إلى أربع فئات فإذا كان لدينا البيانات التالية:

الفئة	عدد المنشات	عدد العمال (c)	التراكمي C	√C
1 - 5	200	720	720	26.8
6 - 10	250	2000	2720	52.2
11 - 15	300	3900	6620	81.4
16 - 20	300	5250	11870	108.9
21 - 25	250	5750	17620	132.7
26 - 30	200	5400	23020	151.7
31 - 35	100	3650	26670	163.3
36 - 40	80	3000	29670	172.2
41 - 45	80	3360	33030	181.7
46 - 50	50	2400	35430	188.2
51 - 55	30	1590	37020	192.4
56 - فأكثر	10	950	37970	194.9

نقسم الجذر التربيعي لعدد العمال التجميعي (√C) على عدد الفئات المقترحة

$$194.4/4 = 48.7$$

الفئة الأولى هي التي تقابل الجذر التربيعي لعدد العمال التجميعي (48.7) وبالتقريب يمكن اعتبار الفئة الأولى هي 1 - 10 عمال، وباستخدام نفس الأسلوب لبقية الفئات نحصل على الفئات التالية:

الفئة الأولى $48.7 \times 1 <= 1 - 10$

الفئة الثانية $48.7 \times 2 <= 11 - 20$

الفئة الثالثة $48.7 \times 3 <= 21 - 30$

الفئة الرابعة $48.7 \times 4 <= 31 - 40$

2.4.2 ميزات وفوائد العينة الطبقية:

1. في العينة الطبقية يكون المجتمع متجانسا في كل طبقة ويكون المجتمع ممثل بشكل جيد حيث تؤخذ عينات من مختلف الطبقات، والتي قد يكون لبعضها أهمية خاصة، وقد تستثنى هذه الطبقات عند استخدام أساليب معاينة أخرى.
2. استخدام العينة الطبقية فعال أكثر من العينات الأخرى خاصة في حالة وجود مجتمع غير متجانس وفي حالة وجود قيم متطرفة لبعض وحدات المعاينة.
3. تؤدي إلى تخفيض الكلفة لأنها تخفض عدد وحدات المعاينة المطلوب تغطيتها عند ثبات الدقة.
4. يمكن استخدام العينة الطبقية للحصول على نتائج على مستويات إدارية معينة (إقليم، قطاع، منطقة....).
5. تكون السيطرة على العمل والإشراف عليه وتنظيمه وتحديد منطقة عمل كل مجموعة بشكل أفضل وذلك عند تقسيم المجتمع إلى طبقات حسب المناطق الإدارية.

3.4.2 توزيع العينة على مختلف الطبقات:

هناك عدة طرق لتوزيع العينة الإجمالية (n) على مختلف الطبقات بحيث يكون حجم العينة في الطبقة هو N_h ومن أهمها:

أسلوب التوزيع المتساوي من العينات لكل طبقة:

تستخدم هذه الطريقة عادة عند الحاجة للحصول على نتائج على مستوى المناطق الإدارية (في حالة كون الطبقة تمثل منطقة إدارية) أو في حالة توزيع العمل بشكل متساوي بين جميع الطبقات (حسب توفر الإمكانيات للعمل الميداني) كذلك عندما يكون عدد وحدات المعاينة متقارب في جميع الطبقات. في هذه الحالة يكون عدد العينات في الطبقة الواحدة حسب المعادلة التالية:

$$n_h = {}^{n}/_{L} \quad .....(1)$$

حيث هو L عدد الطبقات.

مثال (7):

إذا كان حجم العينة 2000 وحدة معاينة وعدد الطبقات يساوي (8) فإن عدد العينات في الطبقة h يساوي:

$${}^{2000}/_{8} = 250$$

أسلوب التوزيع المتناسب مع حجم الطبقة:

إن هذا الأسلوب من أكثر الأساليب شيوعا نظرا لسهولته، إذ عندما لا توجد معلومات سوى عدد وحدات المعاينة في كل طبقة فيمكن استخدام المعادلة التالية لتقدير عدد العينات في الطبقة (h)

$$n_h = n({}^{N_h}/_{N}) \quad .....(2)$$

وهذا يعني أن كسر المعاينة)نسبة العينة في الطبقة الواحدة(متساوي لجميع الطبقات مما يؤدي إلى عينة موزونة ذاتيا لا حاجة عندها إلى حساب أوزان المعاينة عند بناء المؤشرات والتقديرات. ويعطي قدرة على إجراء التقديرات بسرعة ودقة عالية.

مثال (8):

لتقدير عدد العمال في إمارة أبوظبي سحبت عينة منشآت من كل إقليم بأسلوب العينة الطبقية وباستخدام التوزيع المتناسب وبحجم عينة إجمالي قدره 35 منشأة.

الطبقة (الإقليم)	عدد المنشات الكلي
1	200
2	100
3	50

عدد عينات الطبقة الأولى يساوي $n({}^{N_h}/_{N}) = 35({}^{200}/_{350}) = 20$

عدد عينات الطبقة الثانية يساوي $n({}^{N_h}/_{N}) = 35({}^{100}/_{350}) = 10$

عدد عينات الطبقة الثالثة يساوي $n({}^{N_h}/_{N}) = 35({}^{50}/_{350}) = 5$

أسلوب توزيع نيمان (Nyman):

أهم ما يميز هذا الأسلوب أنه يستخدم للتقليل من حجم التباين وزيادة دقة وكفاءة البيانات، حيث يؤخذ في الاعتبار بالإضافة إلى حجم الطبقة عند توزيع العينة الكلية على الطبقات تباين كل طبقة. حيث يكون حجم العينة في الطبقة يتناسب طرديا مع الانحراف المعياري لتلك الطبقة وذلك من أجل جعل تصميم المعاينة أكثر فعالية من أسلوب التوزيع المتناسب، وعادة يستخدم هذا الأسلوب عندما يكون الانحراف المعياري مختلف من طبقة لأخرى، وعندما يكون حجم العينة ثابت وكلفة العينة ثابتة لمختلف الطبقات، ويقدر حجم العينة للطبقة (h) حسب المعادلة التالية:

$$n_h = \frac{N_h S_h}{\sum_{h=1}^{L} N_h S_h} \quad .....(3)$$

أما الانحراف المعياري على مستوى كل طبقة فيمكن الحصول عليه من خال تعداد سابق أو يمكن تقديره من المسوح السابقة أو مسوح مشابهة أخرى.

مثال (9):

في مسح دخل وإنفاق الأسرة لتقدير معدل دخل الأسرة في أحد الأقاليم، قسم المجتمع إلى ثاث طبقات حسب فئات الدخل وسحبت عينة من كل فئة بطريقة توزيع نيمان وبحجم عينة إجمالي قدره 15 أسرة فإذا كانت لدينا البيانات التالية:

الطبقة (فئة الدخل)	عدد الأسر في N_h	الانحراف المعياري لكل طبقة	$N_h S_h$
أقل من 1000 دينار	30	20	600
دينار 1000 - 3000	50	30	1500
أكثر من 3000 دينار	20	20	1000
			3100

إن عدد العينات للطبقة الأولى $= n_h = \frac{N_h S_h}{\sum_{h=1}^{L} N_h S_h} = 15 \times \frac{600}{3100} = 3$

عدد العينات للطبقة الثانية $= 15 \times \frac{1500}{3100} = 7$

عدد العينات للطبقة الثالثة $= 15 \times \frac{15 \times 1000}{3100} = 5$

أسلوب التوزيع الأمثل:

يهدف هذا التوزيع إلى تخفيض التباين لأقل قدر ممكن بكلفة محددة، أو لتقليل الكلفة اقل ما يمكن بمستوى دقة معين؛ حيث يدخل عامل الكلفة في توزيع العينات على الطبقات، وعادة يستخدم هذا الأسلوب عندما يكون تفاوت في كلفة جمع البيانات بين الطبقات فمثا كلفة جمع البيانات من المناطق البعيدة أعلى بكثير من كلفة جمع البيانات من المناطق القريبة وهناك عدة معادلات تستخدم لهذا الغرض نعرض منها المعادلة التالية:

$$n_h = \frac{N_h S_h / \sqrt{C_h}}{\sum_{h=1}^{L} N_h S_h / \sqrt{C_h}} n \qquad .....(4)$$

حيث ($\sqrt{C_h}$) تشكل كلفة إحصاء وحدة المعاينة الواحدة للطبقة h.

مثال (10):

في دراسة لتقدير معدل إنتاج المنشأة في مناطق مختلفة أخذت عينة طبقية حيث شكلت كل منطقة طبقة وكان عدد العينات الإجمالي هو 200 عينة وكانت البيانات التالية عن المناطق:

الطبقة (المنطقة)	عدد الحقول	الانحراف المعياري للطبقة S_h	N	كلفة الوحدة	C
1	200	10	2000	4	1000
2	100	20	2000	6	816
3	150	10	1500	6	612
4	100	15	1500	9	500
5	50	20	1000	12	287
6	50	20	1000	20	224
					3439

$$n_h = \frac{N_h S_h / \sqrt{C_h}}{\sum_{h=1}^{L} N_h S_h / \sqrt{C_h}} n$$

عدد عينات الطبقة الأولى: $200 \times 1000/3439 = 58$

عدد عينات الطبقة الثانية: $(816/3439)\ 200 = 47$

عدد عينات الطبقة الثالثة: $(612/3439)\ 200 = 36$

عدد عينات الطبقة الرابعة: $(500/3439)\ 200 = 29$

عدد عينات الطبقة الخامسة: $(287/3439)\ 200 = 17$

عدد عينات الطبقة السادسة: $(224/3439)\ 200 = 13$

أسلوب التوزيع المتناسب مع صفة معينة:

في بعض الحالات تكون الأهمية لصفة معينة والتوزيع على أساسها أهم من عدد مفردات المجتمع ولتوضيح ذلك يمكن أخذ المثال التالي: إذا كان الهدف تقدير الإيرادات، في مختلف المناطق (الطبقات) وكان لدينا البيانات التالية:

الطبقة	عدد المنشآت	كمية الإنتاج
1	1000	20000
2	500	30000
3	1500	10000

في حالة استخدام أسلوب التوزيع المتناسب فإن وزن الطبقة الثالثة يساوي 0.5 على الرغم من أنها تحتوي على 0.1667 فقط من كمية الإنتاج، لذلك في حالة استخدام عدد المنشآت لوزن الطبقة فإنه يعطي نتائج مضللة ويكون نصيب الطبقة الثالثة 50% من حجم العينة المسحوبة، بينما يكون نصيب الطبقة الثانية 17% من العينة المسحوبة على الرغم من أنها تحتوي على 50% من إجمالي كمية الإنتاج مما يقلل من فعالية العينة ويزيد في التكاليف.

لذلك فإن الأسلوب الأفضل هو توزيع العينة على الطبقات حسب كمية الإنتاج كونها الصفة المراد دراستها ويتم توزيع العينة بين الطبقات بناء عليها بدلا من عدد المنشآت. وباستخدام المعادلة التالية لتوزيع العينة بالأسلوب المتناسب مع صفة معينة.

$$n_h = \frac{X_h}{\sum_{n=1}^{L} X_h} n \qquad .....(5)$$

Xh: الصفة التي تحملها الطبقة (h) وهي كمية الإنتاج في المثال الوارد لدينا.

الطبقة	حجم العينة
1	33
2	50
3	17

فإذا كان مثال حجم العينة المطلوب هو 100 منشأة بالأسلوب المتناسب وحسب صفة كمية الإنتاج فيكون توزيع العينة على الطبقات كما يلي:

5.2 أسلوب المعاينة العنقودية

يقوم أسلوب المعاينة العنقودية على مبدأ تقسيم المجتمع إلى مجموعات بشكل مناسب بحيث تكون هذه المجموعات متقاربة بالحجم ومتجانسة بالنسبة للصفة المدروسة، حيث كل مجموعة من هذه المجموعات تسمى عنقود، وتشكل العناقيد المجتمع كاملا دون حذف أو تكرار.

من ميزات العينة العنقودية إنها فعالة بالنسبة لوحدة التكاليف حيث تعطي دقة أكثر لوحدة الكلفة، كذلك يلجأ إلى هذا الأسلوب في كثير من الأحيان خاصة في المجتمعات التي لا يتوفر لديها اطر معاينة أو يصعب توفير إطار حديث بكل مفردة من مفردات المجتمع، ولكن من الممكن توفير إطار بالعناقيد مما يوفر بالجهد والوقت. كذلك يوجد هناك ميزة أخرى لتطبيق هذا الأسلوب وهو التوفير في تكاليف التنقل اثناء العمل الميداني بين وحدات المعاينة. لكن يجب أن لا ننسى أن من عيوب العينة العنقودية أنها اقل فاعلية من العينة العشوائية البسيطة، كونها اقل انتشارا.

وعند استخدام العينة العنقودية يجب مراعاة ما يلي:

1. أن يكون حجم العنقود صغير وعدد العناقيد كبير.
2. عند تكوين العناقيد تؤخذ مفردات المجتمع المتجاورة أو ضمن منطقة معينة حيث تكون غالبا متشابهة للصفة المدروسة.
3. أن تكون أحجام العناقيد متقارب قدر الإمكان.
4. يجب أن يكون كل عنقود موضّح ومعّرف لجامع البيانات.

6.2 أسلوب المعاينة متعددة المراحل

التحدي الرئيسي في كثير من المسوح هو عدم توفر إطار حديث لوحدات المعاينة الرئيسية مثل المنشآت أو المساكن وغيرها، ويكون من الصعب إعداد إطار حديث لها، وفي نفس الوقت يتوفر قائمة أو إطار بمتغير على مستوى تجميعي وليس تفصيلي مثل تجمعات سكانية أو مناطق رئيسية، حيث هذه التجمعات بطيئة التغير. في هذه الحالة يمكن استخدام أسلوب المعاينة متعددة المراحل.

1.6.2 ميزات العينة متعددة المراحل

- توفر في الوقت والمال حيث يكفي بأعداد الإطار بوحدات المعاينة الثانوية لوحدات المعاينة الرئيسية المسحوبة بالعينة.
- هذا التصميم مرن حيث أنه من الممكن استخدام أسلوب سحب العينات في كل مرحلة مختلف عن المراحل الأخرى.

هذا ويفضل ضمن أسلوب المعاينة المتعددة المراحل تقسيم المجتمع إلى وحدات معاينة رئيسية متساوية وذلك:

- عندما يكون حجم وحدات المعاينة الرئيسية كبير نسبيا يحتاج وقت كبير في إعداد إطار بوحدات المعاينة الثانوية.
- عندما يكون حجم وحدات المعاينة الرئيسية صغير يحتاج إلى وقت في عملية التنقل بين العينات.

2.6.2 أسلوب اختيار وحدات المعاينة الرئيسية:

1. إذا كانت وحدات المعاينة الرئيسية متجانسة يمكن استخدام العينة العشوائية البسيطة.
2. إذا وجد تفاوت يمكن تقسيم وحدات المعاينة الرئيسية إلى طبقات وسحب عينة من كل طبقة.
3. إذا كان التفاوت كبير يمكن استخدام العينة المتناسبة مع الحجم.
4. يمكن استخدام العينة المنتظمة ولكن قد يكون من الصعوبة الحصول على تقدير غير متحيز لخطأ المعاينة.

أما وحدات المعاينة الثانوية فيمكن اختيارها أيضا بأي من أساليب المعاينة: العشوائية البسيطة، المنتظمة، الطبقية، العنقودية، أو المتناسبة مع الحجم. وعند استخدام العينة العنقودية يجب مراعاة ما يلي:

3 | تقدير حجم العينة

يتناول هذا الفصل موضوع تقدير حجم العينة الازم عند تنفيذ مسح إحصائي معين، وما هي المتطلبات الأساسية التي يجب توفرها للوصول إلى تقدير أمثل لحجم العينة. إضافة إلى استعراض آلية تقدير حجم العينة وفق طرق وأساليب المعاينة الإحصائية المختلفة.

1.2 تحديد المؤشرات اللازمة لحساب حجم العينة

إن عملية تقدير حجم العينة تقوم على أساس صيغ ومعادلات رياضية تعتمد على عدد من المتغيرات التي يجب توافرها عند إجراء عملية حساب حجم العينة المناسب. إن هذه المتغيرات هي كما يلي:

1. معرفة مستوى الثقة بالتقديرات التي ستبنى بالاعتماد على هذا الحجم من العينة، كأن يكون (99%، 95%، 90%...) وهي تمثل إحصائيا مساحات التوزيع الطبيعي تحت المنحنى الطبيعي القياسي عندما تكون قيم (1.64، 1.96، 2.58، ...) Z على التوالي. وترتبط درجة الثقة بقيمة التقدير ارتباطا موجبا مع حجم العينة، أي كلما زاد حجم العينة يزاد مستوى الثقة بالتقدير.
2. مستوى دقة التقدير، وهو عبارة عن قيمة الخطأ المسموح به أي الفرق بين القيمة الحقيقية والقيمة التقديرية للمعلمة المطلوب إيجاد تقدير لها باستخدام بيانات العينة، هذا ويتناسب حجم العينة طرديا مع مستوى دقة التقدير أي يزداد التقدير دقة (أي يقل الخطأ) كلما زاد حجم العينة.
3. مقدار التباين في وحدات المجتمع الخاص بالمعلمة المنوي تقديرها، فإذا كانت قيمة تباين المجتمع غير معروفة للمؤشر المطلوب تقديره فيجب ايجاد تقدير مناسب لها. أحيانا قد يكون الهدف من المسح تقدير أكثر من مؤشر واحد. فإذا كان الهدف من المسح تقدير مؤشرات عديدة. لابد من اختيار مؤشر مناسب (key indicator) لتقدير حجم عينة يعتمد عليها في تقدير كافة المؤشرات المطلوبة بمستوى كاف من الدقة.
4. لابد للباحث هنا من دراسة لاختيار كل من)مستوى الثقة ومستوى دقة التقدير(في حساب حجم العينة، فكلما كانت قيمة مستوى دقة التقدير أي حد الخطأ المطلوب عند التقدير (d) صغيرا، وكلما كان مستوى الثقة بتجاوز الخطأ المسموح به القيمة (d) عالي، كلما احتاج حجم عينة أكبر. إن دراسة هذا الموضوع تكون من خال دراسة عدة قيم لكل من مستوى الثقة (z) وحد الخطأ (d) وعلى أساس ذلك يتم وضع عدة سيناريوهات مختلفة لحجم العينة، لتقوم إدارة المسح بالموازنة بينها وفقا للتكاليف ومستلزماتها المادية والبشرية المتاحة.

أما موضوع اختيار المؤشر الإحصائي المناسب من بين عدة مؤشرات للمسح بحيث يتوفر له مقدار تباين معلوم مسبقا، فيكون وفق أحد الطرق التالية:

2.2 التقدير المسبق لتباين المجتمع

إن بناء وتركيب معادلات تقدير أحجام العينة تقوم على أساس توفر قيمة لتباين المجتمع (σ^2) للمؤشرات الرقمية، (من هنا لابد من إجراء عملية تقدير لهذه التباينات. إحصائيا هنالك عدة أساليب يمكن من خلالها تقدير تباين المجتمع لأغراض حساب حجم العينة:

الأسلوب الأول: تجزئة العينة على مرحلتين.

ضمن هذه الطريقة يتم تقسيم العينة إلى جزئين يتم تنفيذهما على مرحلتين، في المرحلة الأولى يتم سحب عينة عشوائية بسيطة بحجم (n_1) تقدر على أساسها قيم التباين بالمقدارين $S1^2$ أو $p1\hat{}$. هذا ويستخدم هذا التقدير لحساب حجم العينة النهائي (n).

وفي المرحلة الثانية بعد تقدير قيمة التباين يحسب حجم العينة اعتمادا على هذا التقدير. يتم سحب وحدات المعاينة المتبقية من حجم العينة الكلي.

مثال (1):

إذا كان حجم العينة المتوقع لدراسة متوسط إنفاق الأسرة في منطقة أبوظبي هو 2000 أسرة. وكان هذا الحجم ليس مبنيا على أساس وجود قيمة معروفة لتباين الأسر في الإنفاق على السلع. يتم هنا الاكتفاء باختيار (500) أسرة اختيارا عشوائيا من مجتمع الدراسة ويتم على أساسها تقدير التباين. ثم يستخدم هذا التقدير لحساب حجم العينة، فقد يظهر أن حجم العينة الكافي)بمستوى الثقة والدقة المحددتين مسبقا(هو (2300) أسرة. عندها اكمال المسح بوحدات العينة المتبقية وهي الفارق ما بين 2300 و500 أي 1700 منشأة.

إن أهم ما يميز هذا الأسلوب هو توفير تقديرات موثوق بها لمعلمة (S^2) أو p، أما محددات استخدام هذا الأسلوب فهي أنه يحتاج إلى جهد كبير ووقت طويل. إضافة إلى أن أسلوب سحب العينة انقسم إلى جزئين كل جزء تم من خلاله سحب وحدات عينة بكسر معاينة أي باحتمالية مختلفة عن الآخر، أي لم توظف نظرية الاحتمالات في اختيار العينة بشكلها الأمثل.

الأسلوب الثاني: إجراء مسح تجريبي

يقوم هذا الأسلوب على الاستفادة من بيانات المسوح التجريبية التي يتم تنفيذها قبيل المسوح الرئيسية لتخدم اهداف أخرى كاختبار استمارات المسح وقواعد التدقيق والضبط إضافة إلى تقديرات اعدا د الكوادر اللازمة لتنفيذ المسح. إن واحدا من الأهداف الأخرى التي يمكن وضعها للمسوح التجريبية هو الاستفادة من بياناتها في تقدير تباين المجتمع وحساب الأخطاء المعيارية اللازمة لتقدير حجم عينة المسح الرئيسي.

الأسلوب الثالث: الاعتماد على نتائج مسوح سابقة

يعتبر هذا الأسلوب الأكثر شيوعا في الجانب التطبيقي، إذ عند تقدير حجم العينة المناسب يتم الرجوع إلى مسوح سابقة تم تنفيذها على نفس المجتمع أو حتى على مجتمع آخر مشابه وذلك لتقدير قيمة الخطأ المعياري (Standard error) وعلى الرغم من أن التباين يعتبر أكثر استقرار من التغيرات التي تحصل في مؤشرات النزعة المركزية للظاهرة قيد الدراسة إلا أن مقدار الانحراف المعياري المعتمد على هذا التباين قد يتطلب الوضع أحيانا إجراء تعديل مناسب عليه لاحتمال حصول تغيرات في سلوك الظاهرة مع الزمن.

الأسلوب الرابع: تقدير النسبة للخصائص الوصفية

إن مؤشرات النسبة (عدد الحالات التي تتصف بها الظاهرة المدروسة إلى عدد الحالات الكلي) هي من المؤشرات الأساسية في الدراسات والمسوح. فمن الممكن إعطاء تقدير للنسبة p ومن المعلوم أن كلما كانت النسبة قريبة من الواقع كلما كان تقدير التباين دقيقا. فإذا كانت مثلا النسبة الحقيقية في المجتمع ($p = 0.3$) فإن التباين هو: $(S^2 = p(1-p) = 0.3 \times 0.7 = 0.21)$

يلاحظ أن مقدار التباين يكون في حده الأعلى عندما تكون قيمة النسبة تساوي 0.5، عندها تكون: $(S^2 = p(1-p) = 0.5 \times 0.5 = 0.25)$

وهذا الافتراض يتطلب اختيار أكبر حجم عينة في ظل مستوى الثقة والدقة المعتمدين مسبقا.

من هنا إذا كان مقدار التباين غير معلوم لظاهرة معينة، وأردنا تقدير حجم عينة مناسب لها يتم تحديد مقدار التباين بالمقدار 0.25 فهو يعطي أكبر حجم عينة ممكن كإجراء احترازي.

3.2 اختيار المتغير المناسب لتقدير حجم العينة

من المعلوم في الجانب التطبيقي أن الهدف من تنفيذ مسح إحصائي لا يقتصر على عملية جمع بيانات خاصة بمؤشر واحد فقط، إذ لابد وأن يكون هناك عدة مؤشرات تجمع عنها بيانات المسح وذلك لتحليل الظاهرة المدروسة من مختلف الجوانب. من جانب آخر أصبح هناك تركيزا من قبل معظم المؤسسات الإحصائية الوطنية والدولية نحو تنفيذ مسوح إحصائية تكون ذات مؤشرات متعددة، ففي ظل تطور أساليب جمع البيانات ومعالجتها آليا أصبحت المسوح متعددة المؤشرات (Multi-indicators Surveys) شائعة الاستخدام. حيث توفر نتائج مثل هذه المسوح مؤشرات عديدة ومختلفة. من أمثلة ذلك المسح العنقودي متعدد المؤشرات (Multi-indicators cluster Surveys) الذي يهدف إلى جمع مؤشرات تفصيلية عن وضع الطفل والأم وما يحيط بهما من ظروف وعوامل.

إن التحدي الرئيسي الذي يمكن أن نواجهه عند تحديد حجم عينة هذا النوع من المسوح هو آلية وأسلوب اختيار المؤشر المناسب الذي يمكن أن يفرز حجم عينة كاف لإعطاء تقدير لقيمة هذا المؤشر، وتقديرات للمؤشرات الأخرى في المسح بحيث تتسم جميعها بالدقة والكفاءة.

إن هذا التحدي يقودنا إلى نتيجة وهي أن الأسلوب الأمثل في اختيار المؤشر لتحديد حجم العينة وفقا لما يلي:

- تحديد جميع المؤشرات المهمة التي يتضمنها المسح، واختيار مؤشر مهم من بينهما يتطلب أكبر حجم عينة
- الحرص على ألا تقل نسبة الخطأ (عن حد معين) للمؤشرات المختلفة التي يجري العمل على جمع بيانات عنها.
- تحديد الفئات المستهدفة، واختيار الفئة الأقل نسبة في المجتمع مع مراعاة أهمية تلك الفئة في مجمل أهداف المسح.

مثال (2):

إذا كان مؤشر نسبة الأسر التي لديها إنفاق على السلع المعمرة في مسح دخل وإنفاق الأسرة هو (82%) وكان مؤشر نسبة الأسر التي لديها طاب في المدارس (82%) على التوالي. فإن حجم العينة المسحوبة من المؤشر الأول ينبغي أن يكون هو المعتمد على الرغم من أن المؤشر الثاني يتطلب حجم عينة أكبر، كما يتضح من العمليات الحسابية الآتية بافتراض أن $Z=1.96$ و $d=0.05$

$$n_1 = \frac{(1.96)^2\,(0.82)(0.18)}{(0.05)^2} = 202$$

$$n_2 = \frac{(1.96)^2\,(0.6)(0.4)}{(0.05)^2} = 369$$

يعتبر المؤشر الأول في هذا المسح من المؤشرات الرئيسية وأهميته تتفوق كثيرا على المؤشر الثاني، وعليه فإن زيادة حجم العينة من **202** إلى **369** ليست مبررة للوصول إلى مؤشر ال يعد أساسيا في مسح دخل وإنفاق الأسرة.

4.2 تقدير حجم العينة عند اعتماد مؤشر النسبة:

إن تقدير حجم العينة المناسب يرتبط بطبيعة ونوع المؤشر الرئيسي الذي سوف تستخدم بيانات العينة في تقديره، فقد تكون المؤشرات رقمية كالمتوسط الرقمي، كمتوسط الطول (بالسنتمتر)، أو متوسط وزن العبوة (بالغرام)، أو متوسط درجات الطلبة (بالدرجة)، أو قد تكون المؤشرات تهتم بنسبة حدوث الظاهرة (أي نسبة تكراراها)، مثل نسبة المدخنين البالغين.

تقوم القواعد الإحصائية في تقدير حجم العينات سواء كانت هذه العينة تهتم في تقدير مؤشرات رقمية أو مؤشرات نسبة، تقوم على أساس خصائص التوزيع الطبيعي.

حجم العينة لتقدير نسبة المجتمع بدلالة مقدار نقاط الخطأ النسبي (h) المحدد مسبقا

إذا كان مقدار النسبة المقدر مسبقا بالاعتماد على مسح سابق أو بناء على تجربة قبلية أو بافتراضه قيمة معينة هو (p) وكان توزيع المعاينة التوزيع الطبيعي التقريبي بوسط وتباين

$$E(\hat{p}) = P_1 \quad \text{وتباين} \quad var(\hat{p}) = \frac{p(1-p)}{n}$$

تمثل القيمة **(1)** المسافة (الفرق) عن نسبة المجتمع، ويتم التعبير عنها بالصيغة الآتية:

$$D = \frac{Z_{1-a}}{2} \sqrt{\frac{p(1-p)}{n}} \quad .....(1)$$

وتعبر في العلاقة **(1)** عن عدد الأخطاء المعيارية التي تبتعد فيها النسبة المقدرة من العينة (p)، (أي الفرق بينهما) أي أن قيمة (d) تعطي دلالة لمستوى الدقة، للحصول على قيمة صغيرة ل (d) يتطلب زيادة حجم العينة، فإذا اختيرت القيمة $Z = 1.96$ فإن تفسير ذلك هو **95%** من نسب العينات الممكنة ستقع ضمن ما مقداره **1.96** من الأخطاء المعيارية standard errors.

إن ما سبق يقود إلى القول إن الحرص على إن يكون الفرق بين النسبة الحقيقية والنسبة التي نقدرها من العينة صغيرا جدا يعني اختيار حجم عينة كبير. وللحصول على حجم العينة يمكن حل المعادلة **(1)** أعلاه بدلالة قيمة (n) لنحصل على

$$D = \frac{\frac{Z_{1-a^2}}{2}}{d^2} p(1-p) \quad .....(2)$$

يتضح من المعادلة أعلاه أن حجم العينة يزداد بزيادة قيمة البسط في العلاقة **(2)**، عند قيمة معينة ل Z فإن n تتغير تبعا لتغير قيمة P

إن أعلى قيمة ل $P(1-P)$ هي **0.25** التي تقابل النسبة $P = 0.5$ وعليه إن أكبر حجم للعينة التي نختارها لتقدير النسبة هو عندما تكون النسبة في المجتمع مساوية ل **(0.5)** ويلاحظ أن الانخفاض في قيمة **(1-P)** هو انخفاض بطيء مع ابتعاد قيمة P عن **0.5**.

– حجم العينة لتقدير نسبة المجتمع بدلالة الفرق النسبي ε

أظهرت العلاقة **(2)** أن مقدار حد الخطأ المسموح به تم تحديده بمقدار فرق نسبي من النقاط. فعندما نقول إن **(d=%5)** فهذا يعني أن تقدير نسبة المجتمع مقبولا إذا وقع بين قيمة **P** المقدرة مطروحا منها ومضافا إليها المقدار **d= %5** في كثير من الأحيان يحبذ أن يكون تقدير النسبة **P** واقعا ضمن نسبة من قيمته بدلا من وقوعه ضمن عدد من النقاط النسبية كما عكستها الحالة السابقة. فعلى سبيل المثال إذا كانت قيمة **P** المقدرة هي، **0.2** وأن قيمة التقدير المطلوبة هي ضمن **10%** من قيمة **P** فإن قيمة الخطأ المطلق بالنقاط هي

$$d = 0.2 \times 0.10 = 0.02$$

إن مقدار حجم العينة في هذه الحالة يمكن أن يعبر عنه بالعلاقة التالية:

$$n = \frac{(1-p)\,\frac{Z_{1-\alpha^2}}{2}}{(p)(\varepsilon)^2} \qquad .....(3)$$

5.2 تقدير حجم العينة عند اعتماد المتوسط

إذا كانت المؤشرات المطلوب تقديرها هي ليست نسب وإنما متوسطات أو مجاميع، في هذه الحالة إن التطبيق الإحصائي لحساب حجم العينة اللازم اختياره لا يختلف عن حساب حجم العينة في حال تقدير المؤشرات التي تكون على شكل نسب.

حجم العينة لتقدير متوسط المجتمع بدلالة الخطأ بالنقاط (**d**)

إن حجم العينة الازم للوصول إلى تقدير مناسب للمتوسط (**μ**) لمتغير معين، يحسب حجم العينة باعتماد الصيغة التالية:

$$n = \frac{Z_{1-\alpha^2/2}\,\sigma^2}{d^2} \qquad .....(4)$$

حيث أن:

σ^2: هو تباين المجتمع، ويمكن تقديره باستخدام مقدار S^2 لعينة من مسح سابق مشابه أو بالاعتماد على نتائج مسح قبلي.

d: تمثل حد الخطأ المسموح به في التقدير وهي: القيمة المطلقة للفرق المرغوب عدم تجاوزه بين متوسط المجتمع الحقيقي (**μ**) والمتوسط المقدر من معطيات العينة.

مثال (3):

تشير المعلومات إلى أن التباين في إنتاج عدد من المزارع هو **(250)** كغم... عند تقدير متوسط الإنتاج في الحيازة الزراعية الواحدة بحيث لا يزيد الفرق بين المتوسط المقدر من معطيات العينة وبين المتوسط الحقيقي عن (كغم واحد)، بمستوى ثقة **95%**. فإن حجم العينة الازم (عدد الحيازات) لتقدير متوسط عدد إنتاج الحيازة كالتالي:

المعطيات هي:

أن الخطأ المطلق المسموح به هو **d=1**

قيمة **Z** التي تحقق ثقة قدرها **95%** هي **Z= 1.96**

وتباين المجتمع هو $\sigma^2 = 140$

$$n = \frac{(140)(1.96)^2}{1^2} = 538$$

إن حجم العينة البالغ **(538)** هو حجم كبيرا نسبيا لأننا تشددنا في تضييق فجوة الفرق بين القيمة الحقيقية والقيمة التقديرية لمتوسط إنتاج المزرعة، فإذا قبلنا بفرق لا يتجاوز **2** كغم (أي **d=2**) فإن حجم العينة سينقص بشكل كبير، حيث:

$$n = \frac{(250)(1.96)^2}{2^2} = 240$$

حجم العينة لتقدير متوسط المجتمع بدلالة الخطأ مقاسا بنسبة من المعلمة (ε)

إذا ما تم اعتماد فرق نسبي للخطأ المسموح به بدلا من الفرق بالنقاط، عند ذلك تستبدل الصيغة **(4)** بالصيغة الآتية:

$$n = \frac{Z^2 \sigma^2}{\varepsilon^2 (\mu)^2} \quad .....(5)$$

حيث أن:

ε نسبة الخطأ المسموح به:

μ معلمة المجتمع (الوسط الحسابي):

لكن القيم الحقيقية μ .σ^2 قد لا تكون متوفرة، عند ذلك يستعاض عنها بقيم تقديرية مناسبة سواء من مسوح أو دراسات سابقة، أو من مؤشرات لدول أخرى، أو الاستعانة ببعض المسوح التجريبية المحدودة.

مثال (4):

في المثال السابق إذا كان الخطأ المسموح به هو أن يقع متوسط العينة ضمن **5%** من القيم الحقيقية للمتوسط، التي تشير دراسة سابقة إلى أنها تبلغ **(80)** كغم (أي $\mu = 80$) فإن حجم العينة الازمة سيكون:

$$n = \frac{(250)(1.96)^2}{(0.05)^2(80)^2} = 60$$

إذا ما نظرنا إلى المعادلة **(6)** نلاحظ أن نسبة σ^2 إلى μ^2 من بسط الصيغة ومقامه على التوالي ما هي إلا قيمة مربع معامل الإختلاف (Variation of Coefficient) حيث أن:

$$C.V = \frac{\sigma}{\mu}$$

إن هذا المعامل في الغالب غير متوفر يتم تقديره من العينة.

فإذا كان المرغوب فيه اعتماد قيمة معامل الاختلاف لتقدير حجم العينة، فيمكن التعبير عن ذلك بالصيغة التالية:

$$n = \frac{Z_{1-a^2/2^2} (C.V)^2}{\varepsilon^2} \quad .....(6)$$

6.2 تقديرات حجم العينة في المعاينة الطبقية

إن تقديرات حجم العينة السابقة تنطبق بشكل أساسي عند أسلوب تطبيق المعاينة العشوائية البسيطة، ولكن وكما بينا سابقا أن أسلوب المعاينة الطبقية يختلف عنه في أسلوب العشوائية البسيطة، فهو يقوم على أساس تقسيم المجتمع إلى طبقات مستقلة يتم تقدير حجم العينة ومعاينة كل طبقة بشكل مستقل تبعا لعدد من المتغيرات وهي مقدار التباين في الطبقة الواحدة وكلفة معاينة الوحدة الواحدة وحجم الطبقة. إن تقدير حجم العينة يكون ضمن الحالات التالية ادناه:

لكلفة محددة مسبقا، وكلفة وحدة المعاينة للطبقات المتساوية مع ثبات التباين:

إذا كانت الميزانية المتاحة للمسح (بالعينة) هي المحدد الوحيد لحجم العينة، وكانت كلفة العمل الميداني المرتبطة بحجم العينة محددة الكمية ***(C)*** وهي مبينة على أساس كلفة الوحدة الواحدة لجمع البيانات C_h أي أن كلفة الوحدة الواحدة ثابتة لكل الطبقات(، ويحسب حجم العينة بموجب الصيغة ***(3)***.

$$n = \frac{C}{C_h} \quad .....(3)$$

مثال (5):

إذا كانت التكاليف الكلية المتاحة للعمل الميداني = ***160000*** درهم وكانت كلفة جمع البيانات من وحدة واحدة = ***100*** درهم فإن حجم العينة سيكون

$$n \quad \frac{C}{C_h} = \frac{160000}{100} = 1600 \text{ وحدة}$$

الكلفة الكلية محددة مسبقا وتباينات الطبقات ثابتة مع اختالف الكلف بين الطبقات

بافتراض أن الكلفة C_h تختلف من طبقة لأخرى، فإن حجم العينة الكلي يحسب كالتالي.

$$n = C\frac{\sum \frac{N_h}{\sqrt{C_h}}}{\sum N_h \sqrt{C_h}} \; n \quad .....(7)$$

مثال (5):

إذا حددت الكلفة الكلية بـ *160000* درهم، وكانت تفاصيل الحجم والكلفة، كما في الجدول:

الطبقات				$N_h\sqrt{C_h}$	$N_h/\sqrt{C_h}$
Stratum No.	N_h	C_h	$\sqrt{C_h}$		
1	4000	36	6	24000	667
2	2700	81	9	24300	300
3	1600	100	10	16000	160
المجموع	8300			64300	1127

$$n = C\frac{\sum \frac{N_h}{\sqrt{C_h}}}{\sum N_h \sqrt{C_h}} = \frac{(160000)(1127)}{64300} = 2805$$

الكلفة الكلية محددة مسبقا مع اختالف كل من حجوم الطبقات وتبايناتها:

في حال أن الكلفة الكلية للمسح معلومة مسبقا، بينما كان كل من حجوم الطبقات، ومقدار التباين لكل منها قيم مختلفة فأن تقدير حجم العينة يكون وفق المعادلة التالية:

$$n = C\frac{\sum L_{h=1} N_h \hat{S_h}/\sqrt{C_h}}{\sum L_{h=1} N_h \hat{S_h}\sqrt{C_h}} \quad .....(8)$$

الكلف والتباين محددة مسبقا على مستوى الطبقة، عند مستوى من الدقة معلوم مسبقا:

كان الاهتمام في الحالات السابقة منصبا على ربط حجم العينة بالكلف والموارد المتاحة للمسح، ولكن إذا ما تم ربط كل من حجم العينة وكلفة المسح المعلومة مسبقا بمستوى الدقة المطلوب عند تقدير حجم العينة، فإن المعادلة التالية يتم من خلالها

$$n = C\frac{\sum (N_h \hat{S_h}\sqrt{C_h})}{N^2 B^2 + N_h \hat{S_h}} \sum \frac{N_h \hat{S_h}}{\sqrt{C_h}} \quad .....(9)$$

6.2 تقدير حجم العينة في أسلوب المعاينة العنقودية باستخدام أثر

التصميم Design effect

يعتبر أسلوب المعاينة العشوائية البسيطة حجر الأساس في تصاميم المعاينة الأخرى، ونتيجة لمحددات مختلفة مثل:

عدم توفر التجانس في المجتمع، الكلفة العالية، عدم توفر إطار معين دقيق أحيانا، يلجأ إلى اساليب معاينة أكثر تعقيد ولكنها بنيت على أساس العينة العشوائية البسيطة. ففي المعاينة العنقودية هناك قيم احتمالية مختلفة ترتبط بخصوصية كل عنقود وهذه تؤدي إلى خصائص قد لا تنطبق مع افتراض الاستقلالية والتطابق في التوزيع الاحتمالي للمجتمع وإنما إلى توزيعات أكثر تعقيداً الأمر الذي يؤدي إلى ظهور مشاكل وتعقيدات في التحليلات الإحصائية.

إذا ما رغبنا بتطبيق المعاينة العنقودية في مسح معين. لتحديد حجم العينة المطلوب ضمن مستوى ثقة وحد خطأ معلومين يلجأ في هذه الحالة إلى افتراض الإستقلالية والتطابق في البيانات وأنها تتوزع توزيع طبيعي وبالاعتماد على المعادلات المشتقة وفق هذا الفرض نحصل على حجم العينة المقدر.

إلا أن حجم العينة المسحوب هو مقرر لعينة عشوائية بسيطة وليست عنقودية، ولكن من المعلوم أن تباين المعاينة العنقودية أعلى من تباين المعاينة العشوائية البسيطة مما يعني أننا بحاجة إلى التغلب على ذلك من خلال زيادة حجم العينة في المعاينة العنقودية إلى حد معين يكون فيه التباين اقرب ما يكون إلى تباين العشوائية البسيطة.

بناء على ما سبق جاءت فكرة استخدام معامل نسبي بين التأثير المتوقع لاستخدام تصميم معاينة معين بدلا من استخدام العينة العشوائية البسيط سمي هذا المعامل بتأثير التصميم.

إذا تأثير التصميم: هو نسبة تباين التصميم الفعلي إلى تباين التصميم العشوائي البسيط، وهو مقياس للكفاءة النسبية ويعبر عنه رياضيا:

$$deff\,(\hat{\Theta}) = \frac{Var_{true}\,(\hat{\Theta})}{Var_{SRS}\,(\hat{\Theta})} \quad .....(9)$$

مما سبق فإن التضخم في التباين في أسلوب المعاينة العنقودية يعود إلى التصميم العنقودي، وعليه يمكن تفسير أثر التصميم على أنه مقدار التضخم في التباين المقدر الذي سببه استخدام التصميم العنقودي بدلا من العينة العشوائية البسيطة.

من جانب آخر فإن عدد وحدات المعاينة الذي يجب اختياره من منطقة العد الواحدة يرتبط ارتباطا مباشرة بقيمة أثر التصميم وفق المعادلة التالية:

$$DEFF = 1 + \delta_x \times (\bar{n} - 1)$$

حيث أن:

$DEFF$ = هو عبارة عن مقدار أثر التصميم

δ_x = هو مقدار الارتباط الداخلي بين الأسر في منطقة العد الواحدة

$\bar{n}$ = هو متوسط عدد الأسر المسحوب من منطقة العد الواحدة

يتضح من المعادلة أعلاه أن مقدار أثر التصميم يتناسب طرديا مع كل من عدد الأسر المسحوب من منطقة العد الواحدة ومع مقدار الارتباط الداخلي بين وحدات المعاينة في منطقة العد الواحدة.

مثال (6):

لدراسة متوسط إنتاج المنشأة الصناعية من مادة معينة، اختيرت عينتين عشوائيا الأولى عنقوديى والثانية عشوائية بسيطة، وكان التباين الناتج في العينتين هو $S^2_{SRS} = 370$ في حين كان تباين العينة العنقودية هو $S^2_C = 520$ أن حجم العينة اللازم للوصول إلى مقدار متوسط الإنتاج وفق $Z = 1.96$ وحد الخطأ $d = 1.5$ هو وفق العينة العشوائية البسيطة:

$$n = \frac{Z^2 S_{SRS}}{d^2} = \frac{1.96^2 370}{1.5^2} = 632$$

إذا حجم العينة العشوائية البسيطة المطلوب ضمن مستوى ثقة 95% هو 632 منشأة، إذا استخدمنا العينة العنقودية بالتأكيد فإن مقدار التباين سيكون أعلى منه في العينة البسيطة وعليه لابد من تعديل حجم العينة للاحاطة بمقدار التباين والمحافظة على التقدير ضمن حد الخطأ المحدد. هنا لابد من تعديل حجم العينة بتأثير التصميم أي أن: $n_C = n_{SRS}.deff$

4 | أساليب المعاينة في مركز الإحصاء – أبوظبي

ضمن إطار المهام الموكل بتنفيذها مركز الإحصاء – أبوظبي، والتي تتناول تنفيذ المسوح الإحصائية بكافة أنواعها الاقتصادية والأسرية والزراعية والبيئية التي تهدف إلى توفير إحصاءات رسمية خاصة بإمارة أبوظبي. هذا إضافة إلى دوره في تصميم وسحب العينات التي قد تطلب من مختلف مؤسسات ودوائر حكومة أبوظبي.

يقوم المركز ببناء اطر المعاينة بأنواعها المختلفة ويتابع بشكل دوري ومستمر عمليات تحديث هذه الأطر، بحيث تؤدي إلى عينات إحصائية قادرة على تمثيل المجتمع الإحصائي بأعلى درجة من الكفاءة والدقة. وبناء على هذه الأطر تتم إجراءات تقدير أحجام العينات المطلوبة، واعداد تصميم المعاينة اللازمة، وسحب العينات الإحصائية بمختلف أنواعها.

1.4 أطر المعاينة في مركز الإحصاء – أبوظبي

وكما بينا سابقا، فإن إطار المعاينة هو عبارة عن قائمة بجميع وحدات المجتمع قيد البحث والدراسة، أو قد تكون على شكل خرائط جغرافية توضح كافة وحدات المجتمع. هذا وتتضمن قائمة الإطار متغيرات تعريفية وجغرافية يستدل من خلالها على وحدات المجتمع، هذا إضافة إلى متغيرات فنية أخرى تساعد في دراسة طبيعة وخصائص وحدات المجتمع ضمن الإطار، كعدد العمال في المنشأة الاقتصادية مثلا في إطار المنشات الاقتصادية، أو عدد الأسر في منطقة العد الواحدة ضمن إطار الوحدات السكنية والأسر.

يعتبر كل من إطار المنشات الاقتصادية وأطار الوحدات السكنية والأسر، إطاران يعتمد عليهما بشكل أساسي في تصميم وسحب عينات المسوح الاقتصادية والأسرية في إمارة أبوظبي.

1.1.4 إطار المنشات الاقتصادية:

المحتوى: يتضمن إطار المنشات الاقتصادية، قائمة بجميع المنشات الاقتصادية العاملة التي تمارس نشاط اقتصادي واحد أو أكثر على أرض إمارة أبوظبي، وذلك بحسب كيانها القانوني، إذ قد تكون مركزا رئيسيا أو منشأة مفردة أو فرع يمسك حسابات لمنشأة مركزها الرئيسي إمارة أبوظبي، أو فرع المنشأة لا تمسك حسابات ومركزها الرئيسي خارج إمارة أبوظبي.

هذا وتصنف ضمن هذا الإطار المنشات الاقتصادية بحسب النشاط الاقتصادي ISIC -4 على أربعة حدود، وعلى مستوى حجم المنشاة وفقا لعدد العمال فيها، وفق التصنيف الخاص بالمنشات الاقتصادية لحكومة أبوظبي:

1. منشات متناهية الصغر: عدد العمال فيها أقل من 5 عمال.
2. منشات صغيرة: يتراوح عدد العمال بها من 5 إلى 19 عامل.
3. منشات متوسطة: يتراوح عدد العمال فيها من 20 إلى 49 عامل.
4. منشات كبيرة: يتراوح عدد العمال بها من 50 عاملا فأكثر.

أما أهم المتغيرات التي يتضمنها الإطار فيمكن تصنيفها على النحو التالي:

أنواع المتغيرات	
المتغيرات الجغرافية	تتناول التعريف بالجانب الجغرافي لوحدات المعاينة (المنشأة) الخاصة بالإطار كالإقليم، المنطقة، رقم الحوض، موقع المنشأة بالمبنى، عنوان المنشأة، وما إلى ذلك.
المتغيرات التعريفية للمنشأة	هذه المتغيرات تتناول التعريف بالمنشأة، مثل اسم المنشأة، ورقم الرخصة، اسم صاحب المنشأة والمدير المسؤول ورقم هاتف المنشأة.
متغيرات خصائص المنشأة	تتناول هذه المتغيرات خصائص حالة المنشأة، والكيان القانوني، والنشاط الاقتصادي، وصفة المنشأة، ورأس المال المدفوع، والإيرادات.

مصـدر البيانـات: يعـود إطـار المنشـات الاقتصاديـة إلى عمليـة حصـر شـامل للمنشـآت الاقتصاديـة فـي إمارة أبوظبي، وذلـك من خال تنفيـذ تعداد شـامل للمنشـآت الاقتصاديـة التـي تمـارس أنشـطة علـى أرض إمارة أبوظبي.

تحديـث إطار المنشات الاقتصاديـة: بهـدف مواكبـة الإطار لعمليـة التغطيـة والشـمول لكافـة المنشات الاقتصاديـة التـي تسـتحدث، أو تغلـق، أو تغيـر نشـاطها، تتـم عمليـات تحديـث مسـتمرة تعتمـد علـى نتائج المسـوح الاقتصاديـة السـنوية التـي ينفذهـا المركـز، إذ يتـم سـنويا إجـراء التحديـث علـى حالـة المنشات فيما إذا توقـف نشـاطها أو غيـرت مـن نـوع النشـاط، أو إذا تغيـر عـدد العمـال فـي المنشـاة، هـذا إضافـة إلـى تحديـث المتغيـرات التعريفيـة والجغرافيـة الدالـة علـى المنشـأة.

إضافـة لمـا سـبق مـن عمليـات تحديـث، يقـوم المركـز بتنفيـذ مشـروع تحديـث إطـار المنشـات الاقتصاديـة بالاعتمـاد علـى بيانـات السـجلات الإداريـة المتوفـرة مـن مؤسسـات حكومـة أبوظبـي لحصـر قوائـم المنشـات التـي أصبحـت تمـارس أنشـطة اقتصاديـة خـال فتـرة 2010 إلـى 2013 أي مـا بعـد فتـرة بنـاء الإطـار الحالـي. ومـن ثـم زيارتهـا ميدانيـا لجمـع البيانـات اللازمـة واضافتهـا إلـى إطـار المنشـات وذلـك بهـدف الوصـول إلـى أكبـر شـمول ممكـن فـي الإطـار الحالـي.

2.1.4 إطار الوحدات السكنية والأسر:

المحتـوى: يتضمـن قائمـة بجميـع الوحـدات السـكنية القائمـة علـى أرض إمـارة أبوظبـي، وهـذه الوحـدات قـد تكـون مأهولـة بأسـر عنـد حصرهـا، أو قـد تكـون خاليـة ولكنهـا معـدة أصـا للسـكن أو تكـون تحـت الإنشـاء.

مصـدر البيانـات: يتـم بنـاء الإطـار بالاعتمـاد علـى عمليـات حصـر المبانـي والمسـاكن والأسـر مـن واقـع تعـداد السـكان والمسـاكن، بحيـث تـدرج جميـع الوحـدات السـكنية التـي كانـت مأهولـة بأسـر وقـت تنفيـذ التعـداد أو كانـت خاليـة ولكنهـا معـدا أصـا للسـكن.

التنظيـم الجغرافـي للإطـار: يأتـي تنظيـم الإطـار جغرافيـا منسـجما مـع التقسـيمات الإداريـة المعتمـدة لـدى حكومـة أبوظبـي، إضافـة إلـى تقسـيمات إحصائيـة تفصيليـة تسـتخدم لأغـراض المعاينـة:

أمـا مسـتويات التقسـيم الإداريـة المعتمـدة فهـي: الإقليـم Region، القطـاع district، الحـوض Sector. فـي حيـن أن مسـتويات التقسـيم الإحصائـي هـي مناطـق العـد الإحصائيـة Enumeration Area.

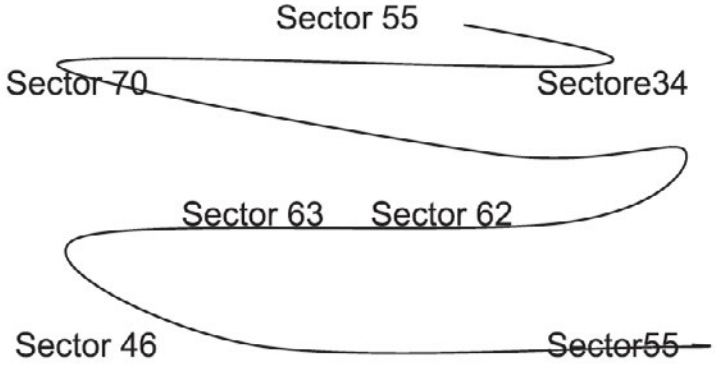
تعريـف: منطقـة العـد هـي عبـارة عـن منطقـة جغرافيـة ذات حـدود طبيعيـة أو صناعيـة تتضمـن مبانـي ووحـدات سـكنية وأسـر، بحيـث يكـون متوسـط عـدد الأسـر فيهـا وقـت تعـداد السـكان والمسـاكن يتـراوح مـا بيـن 100 إلـى 200 أسـرة بالمتوسـط، مـع بعـض الاسـتثناءات الخاصـة بالمناطـق الممتـدة علـى مسـاحات كبيـرة وذات كثافـة سـكانية قليلـة.

بنـاء علـى مـا سـبق، يأتـي بنـاء وتنظيـم إطـار الوحـدات السـكنية والأسـر علـى مرحلتيـن، بحيـث ينسـجم مـع تصاميـم المعاينـة الإحصائيـة التـي يمكـن تطبيقهـا مسـتقبلا:

المرحلـة الأولـى: بنـاء)مناطـق العـد Enumeration areas (بالاعتمـاد علـى الخرائـط الجغرافيـة والبيانـات الخاصـة بالمبانـي والمسـاكن والأسـر مـن التعـداد العـام للسـكان والمسـاكن، تـم تقسـيم إمـارة أبوظبـي إلـى مناطـق عـد مسـتقلة بحيـث تحتـوي المنطقـة الواحـدة 100 – 200 أسـرة.

بنـاءا علـى ذلـك، مـن الممكـن أن تشـكل منطقـة العـد الواحـدة حـوض أو جـزء مـن حـوض، أو أحيانـا قـد يتـم دمـج مجموعـة مـن الأحـواض لتشـكيل منطقـة عـد واحـدة، وذلـك بحسـب حجـم وتوزيـع الأسـر والوحـدات السـكنية فـي المنطقـة. ومـن الجديـر بالذكـر أن هـذه المناطـق بعـد عمليـة بنـاء الإطـار يتـم اعتمادهـا كوحـدات معاينـة أوليـة Primary Sampling units (PSUs).

بعـد إجـراء عمليـات تشـكيل وبنـاء مناطـق العـد تـم ترتيبهـا بحسـب التقسـيمات الإداريـة التـي وردت أعـاه بـدءا مـن الإقليـم ومـن ثـم القطـاع. وضمـن القطـاع الواحـد تـم ترتيـب الأحـواض ترتيبـا حلزونيـا بـدءا مـن اتجـاه الشـمال وانتهـاء بالجنـوب وذلـك لضمـان انتشـار العينـة مسـتقبلا.



أما ترتيب مناطق العد ضمن الحوض الواحد فكان أيضا ترتيبا حلزونيا بحيث يضمن أكبر انتشار داخل الحوض خاصة في الأحواض التي تتضمن عدد كبير من مناطق العد.

بالإضافة إلى القائمة التي تتضمن جميع الوحدات السكنية والأسر ضمن منطقة العد الواحدة هناك خرائط جغرافية توضح المواقع التفصيلية للمباني والوحدات السكنية ضمن منطقة العد الواحدة.

المرحلة الثانية: هي عبارة عن الوحدات السكنية التي كانت مأهولة بأسر وقت التعداد، يتضمن إطار الوحدات والأسر ضمن منطقة العد الواحدة قائمة بأسماء وعناوين جغرافية تفصيلية يمكن من خلالها الوصول إلى الوحدة السكنية والأسر المقيمة فيها. هذا إضافة إلى أسماء ارباب الأسر المقيمين في الوحدات السكنية وقت التعداد (إنشاء الإطار) ونوع الأسرة المقيمة في الوحدة بحسب جنسية رب الأسرة (أسرة مواطنة، أسرة غير مواطنة، أسرة جماعية). من هنا تصنف الأسر ضمن الإطار إلى نوعين رئيسيين:

الأسرة الخاصة: هي عبارة عن فرد أو كثر يعيشون ضمن وحدة سكنية واحدة ويتشاركون في وجبات الطعام، وهناك شخص متعارف عليه من بينهم على أنه رب للأسرة وهو مسؤول عن الترتيبات المعيشية للأسرة، وغالبا ما ينفق افراد الأسرة من دخل رب الأسرة. وإذا كانت جنسية رب الأسرة مواطن تصنف الأسرة على أنها أسرة خاصة مواطنة، أما إذا كانت جنسية رب الأسرة غير مواطنة فتعتبر الأسرة خاصة غير مواطنة.

الأسرة الجماعية: هي مجموعة من الأفراد يقيمون معا في وحدة سكنية واحدة، لا تربطهم علاقة قرابة ولا يوجد رب للأسرة، وغالبا لا يشتركون في تناول وجبات الطعام والإنفاق عليها معا.

ويجب التمييز هنا بين الأسرة الجماعية وفق هذا المفهوم ومعسكرات العمال، فمعسكرات العمال تكون وحدات سكنية كبيرة وتكون اداراتها من قبل مؤسسة أو شركة معينة تشغل الأفراد المقيمين في المعسكر، بينما الأسرة الجماعية يكون الأفراد المقيمين في الوحدة السكنية هم المسؤولين عن الترتيبات المعيشية وإدارة المسكن الذي يقيمون فيه.

تحديث إطار الوحدات السكنية والمساكن:

إن عملية المحافظة على تحديث إطار المعاينة تعتبر واحدة من الأوليات الرئيسية التي تبنى عليها دقة الأطار وبالتالي كفاءة وجودة المسوح التي تبنى على هذا الإطار.

إن عملية تحديث الإطار تكون ضمن منهجين مستقلين يمكن الاعتماد على أي منهما بحسب الإمكانيات والموارد المتوفرة والوقت المتاح.

المنهج الأول: يعتمد على استخدام بيانات السجلات الإدارية التي توفر معلومات تفصيلية عن تطورات الوحدات السكنية والأسر المقيمة فيها، وغالبا ما توفر سجلات الخدمات كالماء والكهرباء مثلا بيانات يمكن الاستفادة منها في تحديث مناطق العد الجديدة التي ظهرت بعد فترة أو وقت التعداد، فقد تكون هناك مدن جديدة وإسكانات أنشئت بعد التعداد أو كانت تحت الإنشاء والآن أصبحت مأهولة، فبالاعتماد على بيانات السجلات الإدارية يمكن إضافة هذه الوحدات إلى الإطار بهدف زيادة مستوى الشمول والتغطية.

المنهج الثاني: يكون من خال إجراء عمليات حصر شاملة لمناطق العد يتم خلالها تحديث المباني الوحدات السكنية والأسر المقيمة ضمنها، وفي الغالب لا تتم عملية التحديث وفق هذا الأسلوب لكافة مناطق العد وذلك نظرا للوقت والتكلفة الباهظة التي سوف تترتب على ذلك. من هنا تتجه بعض الدول إلى إجراء عمليات تحديث جزئية للمناطق التي يتوقع أن يكون قد حصل فيها تغيرات معنوية بإعداد المباني والوحدات والأسر.

من جانب آخر هناك منهج آخر يتناول تحديث جزئي فقط لمناطق العد الواقعة ضمن العينة الرئيسية Master Sample، والعينة الرئيسية هي عبارة عن عينة كبيرة نسبية يتم سحبها بهدف خدمة عدة مسوح تكون مقسمة إلى مجموعات تسمى مكررات Replicates بحيث تسحب المكررة الواحدة بشكل مستقل وتكون ممثلة للمجتمع بأكمله. عند إجراء أي مسح يتم اختيار مكررة واحدة أو أكثر بحسب حجم عينة المسح المقررة، ويتم تنفيذ المسح بناءا عليها.

يقوم مركز الإحصاء – أبوظبي بتحديث إطار المساكن والمباني الحالي اعتمادا على المنهاج السابقة وذلك من خال الاستفادة من بيانات الماء والكهرباء في إضافة ودمج المناطق والتجمعات السكنية الجديدة التي تظهر بين الحين والأخر. كذلك يتم تحديث مناطق عد العينة الرئيسية والتي تشكل نسبة كبيرة من عينة مسح القوى العاملة السنوي.

2.4 تصميم عينات المسوح الإحصائية

1.2.4 تصميم عينات المسوح الاقتصادية:

يقوم مركز الإحصاء – أبوظبي بتنفيذ مسوح اقتصادية ربعية وسنوية وذلك لتحقيق اهداف موضوعة مسبقا، إن كلا النوعين من المسوح يتطلب تصاميم وأحجام عينات مختلفة بحسب اختلاف الأهداف.

المسوح الاقتصادية الربعية:

تهدف المسوح الاقتصادية الربعية إلى الحصول على مؤشرات اقتصادية على مستوى الحد الأول من النشاط الاقتصادي، وهي تأشيريه لأغراض متابعة تطورات الحسابات القومية.

إن أسلوب المعاينة المعتمد لهذا المسح هو أسلوب المعاينة غير الاحتمالية (الغرضية) وهو يقع ضمن أسلوب عينة القطع Cut off sample، تم الاعتماد على متغير مجموع الإنتاج من بيانات المسوح الاقتصادية السنوية لعامي 2010 و2011 كإطار للمعاينة، وتم حصر جميع المنشات التي يشكل مجموع إنتاجها على مستوى النشاط الاقتصادي الواحد 80% فأكثر من مجمل إنتاج النشاط.

المسوح الاقتصادية السنوية:

ضمن اهداف المسوح الاقتصادية السنوية المتمثلة في توفير مؤشرات اقتصادية على مستوى الأنشطة الاقتصادية على حدين، إضافة إلى التمثيل ضمن النشاط الواحدة لحجم المنشاة (كبير، متوسط، صغير، صغير جدا)، فإن أسلوب المعاينة الأمثل المعتمد في تصميم هذه العينة هو أسلوب المعاينة الطبقية العشوائية، إذ يتم تقسيم المنشات إلى طبقات مستقلة:

- بحسب النشاط الاقتصادي على حدين بحيث تتضمن كل طبقة عدد من المنشات الاقتصادية
- بحسب حجم المنشأة الاقتصادية الواقع ضمن 4 مجموعات، وهي كبيرة، متوسطة، صغيرة، متناهية الصغر.

مما سبق يتم التعامل مع كل طبقة على أنها مجتمع إحصائي مستقل يتم تحديد حجم العينة لها وفقا للصيغ والمعادلات السابق ذكرها في الفصل الثالث، وذلك بناء على مستوى ثقة وحد خطأ محدد مسبقا بحيث لا يتجاوز 10% لمتغير الطبقية (حجم العمالة) على مستوى النشاط الاقتصادي على حدين.

2.2.4 تصميم عينات المسوح الأسرية:

لقد تم اعداد وتنظيم إطار المسوح الأسرية بما يتناسب مع اهداف تلك المسوح بحيث يضمن تمثيل كفؤ للنتائج. وكأحد الإجراءات الأساسية التي اعتمدت عند بناء الإطار تم تقسيم مجتمع الأسر في إمارة أبوظبي إلى أربعة طبقات مستقلة وذلك لضمان الحد الأدنى من التباين في وحدات المجتمع بما يكفل التقليل ما أمكن من حجم العينة مع المحافظة على مستوى عال من الدقة، أما الطبقات فقد كانت بالشكل التالي:

الطبقة الأولى: تتضمن جميع مناطق العد التي أسر مواطنة بنسبة تقل عن 25% من مجموع الأسر.

الطبقة الثانية: تحتوي على جميع مناطق العد التي فيها أسر مواطنة بنسبة 25% إلى 50%.

الطبقة الثالثة: تحتوي على جميع مناطق العد التي فيها أسر مواطنة بنسبة 50% إلى 75%.

الطبقة الرابعة: تحتوي على جميع مناطق العد التي فيها أسر مواطنة بنسبة 75% إلى 100%.

بناء على ذلك فقد تم ترتيب مناطق العد ضمن الإطار وفق التسلسل التالي: الإقليم، القطاع، الحوض، الطبقة. بحيث تصنف ضمن القطاع الواحد جميع مناطق العد إلى 4 طبقات بحسب التصنيف أعاه. بناءا على ذلك وعلى مستوى أقاليم إمارة أبوظبي نكون قد حصلنا على 12 طبقة ضمنية مشكلة على النحو التالي:

الرمز	الطبقة	الرمز	الطبقة	الرمز	الطبقة
11	الطبقة الأولى أقليم أبوظبي	21	الطبقة الأولى أقليم العين	31	الطبقة الأولى أقليم الغربية
12	الطبقة الثانية أقليم أبوظبي	22	الطبقة الثانية أقليم العين	32	الطبقة الثانية أقليم الغربية
13	الطبقة الثالثة أقليم أبوظبي	23	الطبقة الثالثة أقليم العين	33	الطبقة الثالثة أقليم الغربية
14	الطبقة الرابعة إقليم أبوظبي	24	العين الطبقة الرابعة إقليم	34	الطبقة الرابعة إقليم الغربية

أما أسلوب المعاينة المستخدم هو المعاينة الطبقية العنقودية المسحوبة على مرحلتين Stratified Two Stage cluster Sample Design. فضمن المرحلة الأول يتم سحب عينة مناطق العد من كل طبقة من الطبقات المذكورة أعاه، وفي هذه المرحلة تسمى مناطق العد بوحدات المعاينة الأولية PSUs، وفي المرحلة الثانية يتم سحب عينة الوحدات السكنية المأهولة بالأسر، من كل وحدة معاينة أولية تم اختيارها في المرحلة الأولى.

مثال (1):

إذا كان حجم العينة المقرر لتنفيذ مسح القوى العاملة هو 3360 أسرة، وافضت الدراسات الأولية للعينة بأن اختيار 14 أسرة من كل منطقة عد يعطي نتائج تتسم بالدقة المطلوبة، فإن عدد مناطق العد التي سوف يتم سحبها في المرحلة الأولى يكون 3360 مقسوما على 14، أي 240 منطقة عد.

أما تحديد حجم عينة المسوح الأسرية فيعتمد على مستوى الثقة وحد الخطأ المطلوب مسبقا، فعلى مستوى الإقليم يكون حد الخطأ لا يتجاوز 5% من قيمة التقدير وبمستوى ثقة مقداره 95%، ويزداد حد الخطأ نسبيا كلما نزلنا في تمثيل العينة على مستويات أكثر تفصيل، وعلى أساس هذه المعطيات يتم تحديد حجم العينة بناء على المعادلات والصيغ السابقة في الفصل الثالث. بعد ذلك يتم تعديل حجم العينة الناتج من المعادلات بمقدار تأثير التصميم Design effect الناتج من نفس المسح أو مسوح شبيهة، تم تنفيذها خال السنة أو السنوات السابقة.

3.4 أساليب سحب العينات

1.3.4 أسلوب سحب عينة المسوح الاقتصادية:

كما بينا سابقا يتم التعامل مع كل طبقة من طبقات مجتمع المسوح الاقتصادية في إمارة أبوظبي بشكل مستقل، فيتم سحب العينة المقررة من كل طبقة بأسلوب المعاينة العشوائية المتناسبة مع الحجم. إذ يتم ترتيب المنشات الاقتصادية في الطبقة تصاعديا بحسب عدد العمال، ومن ثم يتم إنشاء عمود يتضمن المجموع التراكمي لأعداد العمال. وبقسمة مجموع عدد العمال في الطبقة على عدد المنشات المقرر سحبه من الطبقة يتم تحديد فترة الانتظام ومن ثم تختار نقطة البداية العشوائية، وعلى ضوء هذه المعطيات يتم سحب عينة المنشات من كل طبقة.

أما بالنسبة لمنشات الطبقة التي تتضمن 50 عاما فأكثر، فيتم التعامل معها بطريقة مختلفة، إذ يتم تجزئتها إلى طبقتين ضمنيتين، بحيث تتضمن الطبقة الأولى جميع المنشات ذات الأكبر عدد من العمال ويتم في هذه الحالة مسح كافة المنشات في هذه الطبقة حيث يكون في الغالب عددها محدودا، أما الطبقة الضمنية الثانية التي تتضمن منشات فيها عدد عمال اقل مقارنة بالطبقة الأولى فيتم سحب العينة المقررة منها بأسلوب العينة المتناسبة مع الحجم.

2.3.4 أسلوب سحب عينات المسوح الأسرية:

بعد تحديد حجم العينة وتوزيعها بحسب الإقليم والطبقة، تتم إجراءات السحب ضمن مرحلتي المعاينة بالطرق التالية:

- يتم سحب مناطق العد من قائمة المناطق في الإطار بأسلوب المعاينة العشوائية المتناسبة مع الحجم (Proportional to the size)، إن هذا الأسلوب يعطي فرصة أكبر لمناطق العد ذات عدد الأسر الكبير نسبيا بالظهور في العينة مما يرفع من الكفاءة والجودة.
- يتم سحب عينة الأسر من كل منطقة عد، بأسلوب المعاينة العشوائية المنتظمة – Systematic Sampling، إن هذا الأسلوب يضمن انتشار توزيع عينة الأسر المختارة أن تكون منتشرة جغرافيا في منطقة العد الواحدة. مما يعطي تباين عالي نسبيا بين الأسر داخل منطقة العد الواحدة. وبنفس الوقت يعطي تباين قليل نسبيا ما بين الطبقات. مما يؤدي إلى تفعيل دور هذا النوع من تصاميم المعاينة في إعطاء نتائج ذات دقة عالية.

4.4 تطبيقات (تصميم عينة مسح القوى العاملة 2012)

ضمن مهام عمل مركز الإحصاء – أبوظبي يقوم سنويا بتنفيذ مسح القوى العاملة في إمارة أبوظبي. الهدف: يهدف هذا المسح بشكل كلي إلى توفير بيانات تتعلق بالقوى العاملة وبخصائص المتعطلين والمشتغلين والنشيطين اقتصاديا، إضافة إلى قياس مستويات العمالة والبطالة والمعروض من العمالة المحتملة، والتغيرات التي تطرأ عليها ومن ثم التعرف على ديناميكيتها. أما أهم المؤشرات التي يعرضها هذا المسح فهي:

- اعداد وتوزيع السكان النشيطين اقتصاديا بحسب حالة النشاط (مشتغل، متعطل).
- ربط حالة النشاط الاقتصادي بمختلف الخصائص الاجتماعية والسكانية كالعمر والحالة الاجتماعية والجنس وغيرها.
- اعداد وتوزيع السكان غير النشيطين اقتصاديا بحسب العلاقة بالنشاط الاقتصادي (كالطلاب، ربات البيوت....الخ).
- التعرف على معدلات البطالة حسب الأقاليم وحسب الجنسية (مواطن/غير مواطن) وحسب مختلف المتغيرات الاجتماعية.

مستوى التغطية: ضمن الأهداف المحددة لهذا المسح، يغطي جغرافيا السكان المقيمين في أقاليم إمارة أبوظبي الثلاثة (أبوظبي، العين، المنطقة الغربية)، كما ويشمل في التغطية ضمن هذه الأقاليم جنسية الأفراد (مواطن/ غير مواطن) إضافة إلى النوع الاجتماعي للفرد (ذكور/ إناث).

المتغيرات الرئيسية لتحديد حجم العينة:

كما ذكرنا سابقا، هناك عدة متغيرات رئيسية يغطيها هذا المسح منها، عدد المتعطلين، عدد المشتغلين، معطل البطالة.... الخ. بهدف تحديد حجم العينة لابد من اعتماد متغير واحد تتم على أساسه عملية تقدير الحجم. وبما أن متغير معدل البطالة هو مركب من اعداد المشتغلين واعداد المتعطلين فيعتمد هنا لتقدير حجم العينة.

تحديد مقدار التباين في القوى العاملة. من المعلوم أن مركز الإحصاء – أبوظبي قام في عام 2008 بتنفيذ مسح القوى العاملة. أما أسلوب تصميم المعاينة الذي اعتمد لهذا المسح آنذاك فقد كان أسلوب المعاينة العشوائية الطبقية العنقودية المسحوبة على مرحلتين، بحجم عينة مقداره 3631 أسرة موزعة على الأقاليم الثلاثة، حيث في المرحلة الأولى تم سحب مناطق عد، وفي المرحلة الثانية سحبت عينة من الأسر بواقع 16 أسرة من كل منطقة من مناطق العد.

بالاعتماد على قاعدة البيانات لهذا المسح يمكن التعرف على الأخطاء المعيارية واخطاء المعاينة ومعامل التغير النسبي لمعدل البطالة وذلك بحسب الإقليم وكما يبينه الجدول التالي:

جدول (1) خطأ المعاينة والأخطاء المعيارية لمعدل البطالة بحسب الإقليم، 2008

			Confidence Interval 95%			No. of Observations
Domain	SE	CV	Lower	Upper	DEFF	
Abu Dhabi Emirate Region	0.004	0.084	0.035	0.050	2.91	9,139
Abu Dhabi	0.004	0.086	0.034	0.048	1.64	4,819
Al Ain	0.008	0.168	0.030	0.060	4.63	3,479
Al Gharbiah	0.012	0.320	0.014	0.063	2.90	841

حيث أن:

SE: هو مقدار الخطأ المعياري.

CV: هو مقدار معامل التغير النسبي أو يعرف بمقدار خطأ المعاينة النسبي.

Interval Confidence 95%: هي عبارة عن فترة الثقة التي تحتوي على معدل البطالة الحقيقي باحتمال 95%.

DEFF: هو عبارة عن مقدار أثر التصميم، وكما تم تعريفه سابقا.

تقدير حجم العينة:

ضمن اهداف مسح القوى العاملة، المطلوب هو تقديرات لحجم العينة على مستوى أقاليم إمارة أبوظبي الثلاثة للحصول على مؤشرات القوى العاملة في كل أقليم، بحيث يكون مقدار الخطأ المعياري على مستوى الإقليم اقل ما يمكن ضمن الموازنة والإمكانيات المتاحة، عند تقدير حجم العينة، تم التعامل مع كل إقليم على أنه طبقة أو مجتمع مستقل لابد من تقدير حجم العينة له بحد خطأ مقبول. قبل ذلك لا بد من الاطلاع على توزيع وحدات المعاينة (مناطق العد والأسر) في الإطار وذلك بحسب الإقليم. والتي يوضحها الجدول التالي:

جدول (2) التوزيع النسبي للأسر ومناطق العد في اطار السكان والمساكن 2011

الإقليم	العد مناطق عدد	نسبة الأسر الكلية	نسبة الأسر المواطنة	المواطنة الخاصة غير نسبة الأسر	الجماعية نسبة الأسر
أبوظبي	1,161	64.4	58.0	70.2	52.8
العين	680	29.2	37.0	23.6	38.4
الغربية	126	6.4	5.0	6.2	8.8
المجموع	1,967	100	100	100	100

تم الرجوع إلى نتائج مسح القوى العاملة 2008 لغايات تقدير قيمة التباين لمعدل البطالة، وذلك لاستخدامها في تقدير حجم العينة في مسح القوى العاملة 2013، وتم استخدام المعادلات السابقة التي من خلالها يتم حساب حجم العينة عند معرفة مقدار حد أو هامش الخطأ وعند معرفة مقدار التباين من مسح سابق. والجدول التالي يبين التقديرات التفصيلية لحجم العينة:

جدول (3) حسابات عينة مسح القوى العاملة 2008 نحسب الإقليم

المستوى	2008 المعياري الخطأ	2008 العينة حجم	تقدير التباين	الخطأ المتوقع حد أو هامش	تقدير حجم العينة 2013	2013 لحجم العينة التقدير النهائي
أقليم أبوظبي	0.0040	2135	0.03416	0.008	2142	2254
أقليم العين	0.008	1089	0.069696	0.014	1428	1512
المنطقة الغربية	0.012	407	0.058608	0.018	672	714
أمارة أبوظبي	0.004	3631	0.058096	0.007	4242	4480

من جانب آخر لوحظ لارتفاع مقدار أثر التصميم على مستوى الأقاليم، إن رفع كفاءة تمثيل وانتشار العينة يكون من خال التقليل من مقدار أثر التصميم وذلك عن طريق تقليل عدد الأسر المختارة في العنقود الواحد. ففي عام 2008 كان عدد أسر العينة المختارة من العنقود الواحد 16 أسرة، باستخدام المعادلة التالية تم تقدير أثر التصميم في حال تقليل حجم العينة في العنقود الواحد إلى 14 أسرة بدلا من 16 أسرة كما كان في عام 2008. حيث لوحظ أن التقدير الناتج لأثر التصميم مقبول إحصائيا.

على ضوء المعطيات السابقة، يكون العدد الكلي لوحدات المعاينة الأولية المسحوبة في العينة حيث لوحظ أن التقدير الناتج لأثر التصميم مقبول إحصائيا.

$$DEFF_{LFS2012} = 1+ (\bar{n}_{LFS2012} -1) \times \frac{DEFF_{LFS2008} -1}{\bar{n}_{LFS2008} -1} \quad .....(1)$$

حيث أن:

$DEFF_{LFS2012}$: هو أثر التصميم المقدر لعام 2013.

$\bar{n}_{LFS2012}$: هو عدد عينة الأسر في العنقود الواحد في عام 2013 ويساوي (14) أسرة.

$DEFF_{LFS2008}$: هو مقدار أثر التصميم في مسح القوى العاملة 2008.

$\bar{n}_{LFS2008}$: هو عدد عينة الأسر في العنقود الواحد في عام 2008.

على ضوء المعطيات السابقة، يكون العدد الكلي لوحدات المعاينة الأولية المسحوبة في العينة (العناقيد) هو العد 4480 أسرة مقسوماً على العدد (14) حجم العنقود، أي 320 وحدة معاينة أولية أو عنقود، موزعة على إقليم أبوظبي بواقع 160، العين 108، والغربية 52 عنقود.

توزيع العينة على الطبقات:

أما توزيع عينة العناقيد ضمن الإقليم الواحد على الطبقات الأربعة التي تم تعريفها سابقا ضمن الإطار، فقد كان تناسبيا بحسب الإطار، وكما يبينه الجدول التالي:

جدول (4): توزيع عينة وحدات المعاينة الأولية (العناقيد) على الطبقات، حسب الإقليم

الإقليم	الكلي في العينة عدد العناقيد	الطبقة الأولى	الطبقة الثانية	الطبقة الثالثة	الطبقة الرابعة
أبوظبي	160	77	29	34	20
العين	108	43	16	26	23
الغربية	52	28	16	5	2
المجموع	320	148	61	66	45

أسلوب سحب العينة:

لقد استخدم أسلوب المعاينة العشوائية المتناسبة مع الحجم لسحب العناقيد من كل طبقة ضمن الإقليم الواحدة بطريقة مستقلة.

أما اختيار الأسر وعددها (14) من منطقة العد الواحدة فقد تم اختيارها من كل عنقود باستخدام أسلوب المعاينة العشوائية المنتظمة.

حساب أوزان المعاينة:

يمتاز تصميم المعاينة في مسح القوى العاملة أنه مسح غير موزون ذاتيا، أي بمعنى آخر أن مقدار احتمال سحب وحدات المعاينة يختلف بين طبقة وأخرى وأن كسر المعاينة $\frac{n}{m}$ يتباين بين الطبقات، من هنا لابد من حساب أوزان المعاينة لوحدات المعاينة في كل طبقة من الطبقات.

يأتي حساب الأوزان على مرحلتين، الأولى تكون بعد تصميم العينة مباشرة وتدعى الأوزان في هذه المرحلة بالأوزان الأولية، أما الأوزان النهائية فيتم حسابها اعتمادا على الأوزان الأولية بعد جمع البيانات، وتكون من خال تعديل الأوزان الأولية بنسبة الاستجابة على مستوى العنقود.

المعادلة التالية تبين احتمال سحب الأسرة i من منطقة العد h:

$$p_{hi} = \frac{n_h \times M_{hi}}{M_h} \times \frac{m_{hi}}{M_{hi}} = \frac{n_h \times 12}{M_h} \quad \ldots\ldots\ldots\ldots (2)$$

حيث أن:

p_{hi}: احتمال سحب الأسرة i من الطبقة h.

n_h: عدد وحدات المعاينة الأولية في الطبقة h.

M_h: عدد وحدات المعاينة الأولية في الطبقة h.

M_{hi}: عدد الأسر في العنقود j في الطبقة h.

M_h: عدد الأسر المسحوب من العنقود j في الطبقة h، ويساوي 14 أسرة في هذا المسح.

إن الوزن الأولي للأسرة هو عبارة عن معكوس احتمال سحب الأسرة أي من المعادلة أعاه يكون:

$$W_{hi} = \frac{M_h}{n_h \times m_{hi}} = \frac{M_h}{n_h \times 12} \quad \ldots\ldots\ldots\ldots (3)$$

حيث W_{hi}: هو وزن للأسرة i في الطبقة h.

أخيرا، إن الوزن النهائي للأسرة W_{hi} يعبر عنه بالمعادلة التالية:

$$W_{hi} = W_{hi} \times \frac{14}{m_{hi}} \quad \ldots\ldots\ldots\ldots (4)$$

حيث m_{hi}: هو عدد الأسر المستجيبة للمسح في العنقود i بالطبقة h وهو بالطبع اقل من أو يساوي 14.

المراجع:

- معجم المصطلحات الإحصائية، المعهد العربي للتدريب والبحوث الإحصائية 2005.
- الأساليب الإحصائية في ميدان التطبيق، د.مهدي العلاق، أ ذ.عدنان شهاب حمد، 2001.
- Theory and Analysis of Sample Survey Designs, Caroga Singh, 1986.
- Sampling Techniques, William. Cochran, 1953.