الجمهورية الجزائرية الديمقراطية الشعبية
وزارة التعليم العالي والبحث العلمي
جـامـعة أمحمد بوقرة بومـرداس
كلية العلوم الاقتصادية، التجارية وعلوم التسيير
قسم العلوم الاقتصادية

مطبوعة بيداغوجية تحت عنوان:

محاضرات في تقنيات الاستقصاء

من إعداد: د. جمعاسي ابراهيم

موجهة لطلبة: السنة اولى ماستر

تخصص: اقتصادي كمي

السنة الجامعية :2017/2016

مقدمـة عامـة 1

المحور الأول : مدخل لتقنيات الاستقصاء 3

1-0-مقدمة 3
1-1-مراحل الدراسة الاستقصائية 3
1-2-دورة حياة الدراسة الاستقصائية 4
1-2-1-التخطيط 4
2-2-1-التصميم و التطوير 5
3-2-1- التنفيذ 5
4-2-1- التقييم 5

خلاصة 6

المحور الثاني : صياغة الاهداف 7

2-0-مقدمة 7
2-1- خطوات إعداد الأهداف..... 7
2-2- القيود التي تؤثر على اعلان الأهداف 8

خلاصة 9

المحور الثالث : صياغة المسح.....10

3-0- مقدمة 10
3-1- إحصاء شامل / استطلاع الرأي 10
3-2- المجتمع المستهدف / الدراسة 11
3-3- اطار المسح 12
1-3-3-أنواع أطر الدراسة 13
1-1-3-3- القائمة 13
2-1-3-3- المنطقة.....13
3-1-3-3- الاطار المتعدد 14
2-3-3- عيوب الاطار 14
3-3-3- النوعية الجيدة للاطار 14
3-4- أخطاء الاستقصاء 14
1-3-4- أخطاء المتعلقة بالعينة 15

2-3-4-أخطاء الغير متعلقة بالعينة 15

خلاصة 16

المحور الرابع : طرق جمع البيانات..... 17

4-0-مقدمة 17

4-1- طرق جمع البيانات 17

4-2- جمع المعلومات بمساعدة الحاسوب 18

4-3- طرق أخرى لجمع البيانات :..... 19

خلاصة 19

المحور الخامس : صياغة الاستبيان 21

5-0 مقدمة : 21

5-1- خطوات تصميم الاستبيان 21

5-1-1- التشاور/اتحدث مع المستخدمين /المجيبين 21

5-1-2- دراسة الاستبيانات السابقة 22

5-1-3- كتابة الاستبيان.....22

5-1-4- مراجعة الاستبيان 22

5-1-5- اختبار و اعادة الصياغة 23

5-1-6- انهاء الاستبيان (اللمسة الاخيرة) 23

5-2- أنواع الاسئلة..... 23

5-2-1- الاسئلة الثنائية الاختيارية 25

5-2-2- الأسئلة ذات الاختيارات المتعددة / أو القائمة 25

5-2-3-الاسئلة المصنفة : 25

5-2-4- الاسئلة التفضيلية 26

5-3- توجيهات لصياغة الاسئلة..... 26

5-3-1- البساطة في الصياغة 26

5-3-2-التعريف بالرموز المستعملة (المختصرات) 27

5-3-3- قابلية تطبيق الأسئلة 27

5-3-4- التخصص و الدقة 27

5-3-5- تجنب الأسئلة المزدوجة 27

5-3-6- تجنب الأسئلة الايحائية 27

5-3-7- تجنب الأسئلة المزدوجة النافية 28
5-3-8- التخفيف من الاسئلة الحساسة 28
5-3-9- التأكد من القراءة الحسنة للأسئلة 28
5-4- الأخطاء المتعلقة بالإجابات 28
5-4-1- مصادر الاخطاء المتعلقة بالإجابة 28
5-4-2- تقنيات الحد/تقليص الاخطاء المتعلقة بالإجابة 29
خلاصة 30
المحور السادس : شكل العينة.....32
6-0- مقدمة 32
6-1- العينة الغير احتمالية.....32
6-1-1- العينة الفجائية (hasard sampling).....34
6-1-2- العينة التطوعية 34
6-1-3- العينة التحكيمية 34
6-1-4- العينة حسب الحصص 34
6-1-5- العينة الاحتمالية المعدلة 35
6-2- العينة الاحتمالية.....35
6-2-1- الفاعلية الاحصائية (الجودة) 36
6-2-2- العينة العشوائية البسيطة Simple Random Sampling (SRS)..37
6-2-3-العينة المنتظمة Systematic Sampling (SYS) 39
6-2-4- العينة التناسبية مع الحجم Probability-Proportional-to-Size (PPS) Sampling 41
6-2-4-1- طرق العينة PPs 44
6-2-5- العينة العنقودية Cluster sampling.....45
6-2-6- العينة الطبقية Stratified Sampling (STR).....46
6-2-7- العينة المتعددة الخطوات Multi-Stage Sampling 48
6-2-8- عينات متعددة المراحل Multi-Phase Sampling.....49
6-2-9- العينة المعادة Replicated Sampling 50
خلاصة 50
المحور السابع : التقديرات.....52

7-0- مقدمة 52
7-1- الترجيح 52
7-1-1- الترجيح في شكل العينة الاحتمالية المتساوية 53
7-1-2- الترجيح في شكل العينة الاحتمالية الغير متساوية 54
7-1-3 تعديل القوة بالنسبة لعدم الإجابة 56
7-1-4- استعمال البيانات الثانوية لتعديل القوة 59
7-1-4-1- بعد انجاز الطبقية 59
7-1-4-2- النسب التقديرية 62
7-2- التقديرات البسيطة (الاجمالية، المتوسطات، النسب) 63
7-2-1- التقديرات لمختلف أنواع البيانات 63
7-2-2-أمثلة عن التقديرات للقيم الكلية، المتوسطات، النسب 64
7-3- تقدير أخطاء العينات للتقديرات الخاصة بالدراسة 66
7-3-1- تباين العينة 66
7-3-1-1- حساب التباين الحقيقي للعينة..... 68
7-3-2- مقادير أخرى لأخطاء العينات 70
7-3-2-1 الانحراف المعياري، و معامل التغير..... 70
7-3-2-2- هامش الخطأ و مجال الثقة 71
7-3-2-3- تقدير تباين العينات في عينة طبقية 72
7-3-3- أثار نمط العينة 73
7-3-4- تقدير تباين العينات بالاستعانة بالعينة المعادة 74
خلاصة 75
المحور الثامن : تحديد حجم العينة 76
8-0- مقدمة..... 76
8-1- تحديد حجم العينة..... 76
8-1-1- بعض الشروط المتعلقة بالدقة 77
8-1-2- العوامل المؤثرة في الدقة 78
8-1-2-1- التغير على مستوى المجتمع..... 78
8-1-2-2- حجم المجتمع 80
8-1-2-3- مثال التصميم و التقدير..... 81
8-1-2-4- نسبة الاستجابة 81

-1-3-8- أمثلة عن صيغ الحجم 81
-1-3-1-8- مثال في تحديد حجم العينة 83
8-2- توزيع العينة في حالة العينة الطبقية 87
1-2-8- معايير التوزيع 87
1-1-2-8- حجم ثابت للعينة 87
2-1-2-8- معامل التباين الثابت.....88
2-2-8- طرق توزيع العينات 89
1-2-2-8 التوزيع النسبيN (المتكافئ) 89
2-2-2-8- التوزيع الغير نسبي (غير متكافئ).....91
3-2-8- بعض الاعتبارات الحاصة عند التوزيع 96
خلاصة 96
المحور التاسع : تمارين و تطبيقات.....98
9.0. السلسلة الاولى حول انماط اختيار العينات 98
9-1- السلسلة الثانية حول التقديرات و الترجيح.....99
9-2- السلسلة الثالثة حول تحديد حجم العينة.....100

مقدمـة عامـة

يندرج هذه المطبوعة الجامعية في إطار البرنامج المسطر لطلبة الماستر تخصص اقتصاد كمي، فهو يتناول أحد المقاييس المهمة في هذا التكوين ألا وهو تقنيات الاستقصاء. فهي تعتبر كدليل تطبيقي من اجل التخطيط والتحضير لإنجاز دراسة استقصائية ميدانية. فهي تشمل على مجموعة من الأسئلة المتعلقة بعملية سبر للأراء وطرق أخرى التي يمكن أن تضاف لعملية إنجاز الدراسة الاستقصائية.

إن الهدف الرئيسي من هذه المطبوعة هو مساعدة الطالب على إنجاز دراسة استقصائية على أسس علمية. فهي تجيب عن أهم الأسئلة التي يطرحها الطالب عند إنجاز مذكرة التخرج خاصة ما تعلق بعملية جمع البيانات ومعالجتها. كما انها تهدف إلى نقل المعرفة للطلبة من خلال مجموعة من المحاضرات المدعمة بأمثلة وتمارين تطبيقية والتي جاءت في عدة فصول منها ما هو تقني ومنها ما هو نظري.

ولقد تم تنظيم هذه المطبوعة على الشكل التالي: خمسة فصول الأولى تعالج الامور العامة لتحضير عملية سبر الآراء خاصة:

مدخل لبعض المصطلحات الخاصة بالدراسة الاستقصائية والتخطيط لها (الفصل1).

كيفية تحضير وإعلان الاهداف (الفصل 2).

عموميات لتحضير وانجاز الدراسة (الفصل 3) مثل:

- الاختيار بين عملية سبر للآراء ومسح شامل.
- كيفية تحديد المجتمع المدروس.
- مختلف أطر سبر للآراء.
- مصادر الاخطاء.

طرق جمع البيانات (الفصل 4) مثل:

- الجمع التلقائي/الذاتي.
- استجواب شخصي/هاتف.
- استجواب بمساعدة الحاسوب.

كيفية تهيئة وصياغة الاستبيان (الفصل 5).

أما الفصول 6، 7، 8، فهي مخصصة للجانب التقني لإنجاز الدراسة عن طريق سبر للآراء:

كيفية اختيار العينة (الفصل 6).

كيفية تقدير خصائص المجتمع (الفصل 7)

كيفية تحديد حجم العينة وتوزيعها (الفصل 8).

والفصل 9 فهو عبارة عن تمارين وتطبيقات وهذا حتى يتسنى للطالب تجريب ما تم رؤيته والتعود على التقنيات الخاصة بالعينات والاستبيان. والتي يمكن أن يعتمد عليها من أجل إنجاز مذكرة نهاية الدراسة.

كما تجدر الاشارة الى ان هذه المطبوعة هي عبارة عن ترجمة وبتصرف للدليل العلمي المسمى " Survey Method and practices" المنجز من طرف .canada statistique ولقد تم تدعيمه بتمارين وأمثلة وكذا بعض المصطلحات من مراجع اخرى.

المحور الأول : مدخل لتقنيات الاستقصاء

1-0-مقدمة

ماذا نقصد بالاستقصاء؟ الاستقصاء هو كل عمل يحقق جمع المعلومات بطريقة منظمة وممنهجة حسب خصائص بعض أو كل المجتمع بالاستعانة بمفاهيم محددة، مناهج وطرق. في الاخير تقدم كل المعلومات على شكل ملخص ذو فائدة.

يبدأ الاستقصاء عادة من الحاجة للمعلومة أو عدم وجود بيانات أو أن تلك البيانات تكون ناقصة.

في بعض الاحيان يمكن الحصول على البيانات لدى المنظمات الاحصائية الرسمية والتي نتحصل هي كذلك لدى منظمات حكومية أو وزارات أو حتى منظمات خاصة.

تتكون الدراسة الاستقصائية من عدة خطوات مترابطة فيما بينها والتي تشمل:

تعريف بالأهداف، تحديد قاعدة المسح، تحديد العينة، تحضير الاستبيان، جمع ومعالجة البيانات، تحليل ونشر البيانات وفي النهائية حفظ ملف الدراسة الاستقصائية.

كما أنه يمكن تقسيم مدة صلاحية الدراسة إلى عدة مراحل: مرحلة التخطيط، والتي تتبع بمرحلة التصميم و التطوير ثم القيام الفعلي بالدراسة و أخيرا تقييم مسار الدراسة الاستقصائية.

إن الهدف من هذا الفصل هو إعطاء لمحة شاملة عن مجمل الاعمال (النشاطات) المتعلقة بإنجاز دراسة احصائية و هذا مع التفاصيل التي سنتطرق اليها في الفصول اللاحقة.

1-1-مراحل الدراسة الاستقصائية

إن القيام بعملية الاستقصاء يمكن أن تظهر على أنها مسار بسيط عبارة عن طرح بعض الاسئلة ثم معالجة الاجوبة من أجل الحصول على احصائيات. إلا أن الاستقصاء يكون خطوة بخطوة انطلاقا من طرق محددة حتى يكون هناك نتائج محددة و معنوية و يمكن تلخيص هذه الخطوات في:

- كتابة الاهداف
- تحديد قاعدة المسح
- تحديد و معرفة العينة

- صياغة الاستبيان
- جمع البيانات
- إدخال البيانات و عملية الترميز
- المراقبة
- التقدير
- تحليل البيانات
- نشر البيانات و النتائج
- كتابة التقرير

سنعود إلى كل هذه الخطوات بشيء من التفصيل.

2-1-دورة حياة الدراسة الاستقصائية

إن خطوات الدراسة الاستقصائية المذكورة سالفا ليست بالضرورة متتابعة: فيمكن أن تكون هناك خطوات منجزة في نفس الوقت، و البعض الاخر مثال الكتابة فيمكن ان تتكرر في مختلف الخطوات.

كل خطوة يجب أن تكون مخططة أولا ثم تنشئ و تتطور و تنجز على أرض الواقع و في الاخير تقيم. أن مراحل حياة الدراسة هي:

- التخطيط
- التصميم و التطوير
- التنفيذ
- التقييم و التقدير

1-2-1-التخطيط

إن أول خطوة في مسار الدراسة الاستقصائية هي التخطيط، لكن قبل أي تخطيط لابد من وجود هيكل للتسيير و التخطيط و عادة ما يكون عبارة عن مجموعة من الخبراء التي تقترح ضرورة إنجاز دراسة استقصائية أولا هل نكتفي بالبيانات المتوفرة؟ ما هي طبيعة الاسئلة التي يمكن إضافتها؟ هل يمكن العودة الى دراسات سابقة...

من هنا يجب على الفريق تحديد أهداف الدراسة، نمط جمع البيانات، المجال الجغرافي، النوعية، التكلفة، التوقيت،.... كل خطوة من الخطوات تحتاج الى تخطيط مسبق، صياغة وتطوير ثم الانجاز و يمكن اجراء تغييرات على التخطيط.

1-2-2-التصميم و التطوير

بعد ان نضع الإطار المنهجي يمكن القيام بأعمال تفصيلية على مختلف الخطوات و هو ما يعرف بمرحلة التصميم و التطوير. أن الهدف الرئيسي لهذه الخطوة هو إيجاد مختلف الطرق والمناهج التي تسمح بالقيام بتوازن بين الاهداف و العراقيل (المصادر).

في هذه الخطوة يمكن إجراء ما يعرف بالتجربة (اختبار تمهيدي) حتى يمكن معرفة مدى مطابقة الاستبيان مع الدراسة. كما يتم تحضير كل ما له علاقة بالدراسة مثل (كتيبات، التدريب و التكوين على الاستبيان، إعداد وثائق التحكم في العينة...) برامج إعلام ألي، الترميز....

1-2-3- التنفيذ

بعد التأكد من أن كل الامور متوفرة يمكن القيام بتنفيذ الدراسة الاستقصائية، طبع الاستبيان إذا كان لازما، تحضير المقابلات، تحديد العينة.

تشمل عمليات المعالجة الحصول على البيانات، الترميز، التحرير و الطباعة و التكون النتيجة هنا عبارة عن مجموعة من البيانات المهيكلة و التي يمكن استخراج جداول و تحليلها. و يجب مراقبة نوعية البيانات في كل مرحلة استنادا إلى طرق و مناهج متطورة.

1-2-4- التقييم

إن التقييم هو عبارة عن مسار مستمر على طول الدراسة. كل خطوة يجب أن تقيم على حسب فعاليتها، التكلفة....

هذا و يتطلب تقييم الطرق المستعملة و هذا ما يمكن أن يعتبر كتقييم لأداة مستعملة وهذا ما يساعد في المستقبل في استعمال طريقة على حساب أخرى. كما أن تقييم الاستقصاء التجريبي أو التمهيدي جد مهم عند القيام بأعمال احصائية كما نعطي نظرة على نوعية البيانات و النتائج.

خلاصة

ما هي الدراسة الاستقصائية؟ الدراسة الاستقصائية هي نشاط الذي يجمع المعلومات بشكل منتظم و ممنهج. و هو عموما محفز من الحاجة لدراسة الخصائص لمجتمع بناء قاعدة بيانات من اجل التحليل أو اختبار فرضية.

ما هي خطوات الدراسة الاستقصائية؟ اهي عبارة عن مسار جد معقد و ليس عملية طرح للأسئلة و الحصول على أجوبة من اجل انتاج احصائيات. عدة خطوات يجب ان تنجز على أساس طرق و مسار محدد إذا كانت النتائج يجب أن تعطى معلومات محددة. هذه الخطوات تشمل صياغة الأهداف (الدراسة) تحديد نمط العينة، صياغة الاستبيان، جمع البيانات، معالجة البيانات و نشر النتائج.

كيف يمكن انجاز كل الخطوات؟ تنفيذ دراسة استقصائية يمكن أن بعبر كدورة حياة من أربع خطوات: المنهجية، الميزانية و جدول الزمني لعمليات الدولة. الخطة التالية هي تهيئة وتطور الخطوات الدراسة، الخطوة الثالثة هي تنفيذ الخطوات الدراسة. عند تنفيذ الدراسة تقدير و تراقب حتى تطمن من اي مسار يسيير بشكل جيد و في الاخير يتم معالجة و تقييم خطوات الدراسة.

المحور الثاني : صياغة الاهداف

2-0-مقدمة

ان اول مهمة عند التخطيط لدراسة استقصائية هي تحديد أهداف الدراسة (واضحة و على أكمل وجه). إن الاعلان الواضح للأهداف يقود و يوجه كل مراحل الدراسة فيما بعد.

لنفترض أننا نريد القيام بدراسة حول ظاهرة الفقر في بلد ما فلا يجب أن تقتصر في الاعلان عن الاهداف أننا نريد معرفة شروط سكن الفقراء. هذا الاعلان واسع يمكن أن يكون كموضوع للدراسة لكن يجب أن يطور مثل ماذا نعني بشروط السكن؟ الانجاز، السن أو عدد الافراد في المتر المربع..؟ ماذا نعني بالفقر؟ هل نقيس الفقر بالدخل؟ بالمصاريف؟ بالديون؟ أو كل هذه الامور؟.

إذن يجب ان نعرف من يستفيد من هذه البيانات؟ و ما هو الهدف منها؟ إذا كان الذي يريد الدراسة هو هيئة رسمية، بلدية مثلا من أجل بناء مساكن جديدة مدعمة، و يريد معرفة كم العدد و ما هي التكلفة؟

يجب كذلك تحديد ما يمكن مشاهدته و ما يمكن تقديره في حالة الفقر. تعريف العائلة الفقيرة فيجب تحديد المجتمع المستهدف. كل فئة أو عائلة لها دحل أقل من دينار؟ ما هي المنطقة الجغرافية، و ما هي الفترة المرجعية (الاسبوع الماضي، السنة الماضية...الخ). هل يريد احصائيات على فئات الدخل، نوع المسكن، السن، عدد الافراد.

إن الهدف الرئيسي من هذا الفصل هو معرفة كيفية صياغة الاهداف.

2-1- خطوات إعداد الأهداف

يعتمد تحضير الاعلان عن الاهداف على عدة خطوات نذكر منها :

الحاجة للمعلومة : حيث أنه يجب أولا معرفة المشكل و لماذا ننصح بالقيام بالاستقصاء، و ما هو المحتوى العام لإنجاز هذه الدراسة. و بعد أن نحدد الاحتياجات من المعلومات الخاصة بالدراسة الاستقصائية من المهم العودة (الرجوع) إلى هذه الاهداف في كل الخطوات حتى نجيب عتها و نحققها في الميدان.

- المستخدمين (استخدام البيانات) : الاسئلة التي يجب طرحها الان هي من هم أهم مستخدمي البيانات التي سنجمعها؟ و هل تعتبر المعلومة مفيدة؟

- تحديد المفاهيم : من اجل تحديد البيانات اللازمة التي تجيب على الاهداف لابد من تحديد بعض المفاهيم بوضوح. مثل من هم الاشخاص المعنيين بالدراسة؟ (مثال من هو الفقير؟) و هذا ما يساعد على فهم الدراسة فيما بعد. كما يجب تحديد بعض المفاهيم العامة مثل الاسرة، المسكن، العائلة...الخ. و من أجل تحديد المفاهيم العملية هناك ثلاث أسئلة : من/ماذا؟ ، أين و متى؟ . أول مصطلح يجب تحديده هو المجتمع المستهدف و هو المجتمع الذي تكون فيه المعلومة مفضلة و منه يجب معرفة الوحدة المستعملة في الدراسة (الفرد، الاسرة، المدرسة، النساء، المؤسسات،...).
- محتوى الدراسة الاستقصائية: بعد تحديد الاهداف، معرفة المستخدمين يجب تحديد محتوى الدراسة و التي تتطلب معرفة خصائص أخرى مثل السن، الجنس، الدخل...الخ. إن المواضيع الخاصة تحتاج إلى تحديد المتغيرات التي يجب جمعها، صياغة الاستبيان و تحديد العينة و هذا يمكن ان يتأثر بنمط الاستقصاء (الاستعانة بأشخاص مثلا) و هذا يتطلب تكلفة و كل هذا يكون من خلال أسئلة الاستبيان.
- خطة التحليل (الجداول المقترحة) : بعد أن يتم تحديد كل العناصر فإن الخطوة التالية تتطلب تحديد مستوى التفصيل (التدقيق) لكل عنصر و نمط النتائج. ما هي التقديرات، التعداد، المؤشرات....التي تكون ضرورية؟ .

إن خطة التحليل تساعد كثيرا في صياغة الاستبيان مثال هل من الضروري معرفة أو الحصول على النتائج حسب فئات العمر؟ حسب الرجال و النساء؟ بين مختلف أنواع وسائل النقل (دراجة، سيارة، حافلة، قطار؟) هل البيانات مستمرة أو فئوية؟ مثال حالة الدخل. هل نبحث عن الدخل الكلي للأسرة أم الشخص المجيب؟.

2-2- القيود التي تؤثر على اعلان الأهداف

هناك العديد من القيود التي تؤثر على حالة الأهداف لدراسة ما.

كل ما تعلق بنوعية التقديرات أي ما هي الدقة المطلوبة في النتائج؟ هذا ما يرتبط بهامش الخطأ في العينة التي يمكن قبولها بالنسبة للمتغيرات المهمة و هذا يمكن أن يكون له علاقة بالحصول على عينات كبيرة. و من العوامل التي تؤثر على الدقة و منه على حجم العينة نجد:

- متغيرات المتعلقة بخصائص المجتمع
- حجم المجتمع
- تكوين العينة و طريقة التقدير
- نسبة الاستجابة

إضافة إلى قيود عملية و التي تؤثر على الدقة مثل :

- ما هو حجم العينة التي يقبلها الزبون؟
- ما هي الفترة المتاحة لإنجاز الدراسة؟

- ما هو عدد أفراد المحققين (المستجوبين)؟

خلاصة

بدون فكرة واضحة حول الحاجة للبيانات، فإن الهيئة الاحصائية يمكن أن تعالج المشكل الخاطئ و انتاج نتائج غير مكتملة و غير فعالة و فقدان أو تبذير للوقت و الامكانيات. عمليات سبر للأراء يمكن ان تنتج عدة مستجوبين بدون انتاج معلومات مفيدة من أجل هذه الاسباب فإن أهداف الدراسة يجب أن تكون واضحة أثناء مرحلة التخطيط القائمة التالية تلخص الاسئلة المهمة و النقاط التي يجب أخذها بعين الاعتبار أثناء الانجاز:

- الاهداف لسبر الآراء و الحاجة للمعلومات
- ما هي الحاجة للمعلومة للدراسة؟
- من يستعمل البيانات و كيف نستعملها؟
- ما هي المواضيع الخاصة التي تغطيها الدراسة؟
- ما هي الدقة المطلوبة في التقديرات؟.

المحور الثالث : صياغة المسح

3-0- مقدمة

ما ان يتم تحديد الاهداف من الدراسة الاستقصائية تبدأ صياغة و تصميم نمط الدراسة الاستقصائية و من أهم الامور التي يجب طرحها هي ما هل نريد إنجاز دراسة شاملة أو سبر للآراء (أي عينة من المجتمع)، و ما هي أهم مصادر الخطأ التي يمكن أن تكون في الدراسة؟ و ما هو تأثيرها على النتائج؟

إن القرار لاختيار نمط و طريقة إنجاز الدراسة بين الشاملة و مسح لعينة يتحكم فيها عدة عوامل منها : الميزانية، المصادر المتاحة، حجم المجتمع، قاعدة المسح . تجدر الاشارة إلى أن هناك نوعين من الخطأ : الخطأ الناتج عن العينة غي حالة الدراسة المسحية و الخطأ الخارج عن العينة.

3-1- إحصاء شامل / استطلاع الرأي

تجدر الاشارة الى وجود نوعين من الدراسات الاستقصائية : استطلاع للرأي sondage, sample أو احصاء شامل recensement, census . الفرق بينهما أن الأول يشمل معلومات جزء من وحدات المجتمع أما الثاني فيشمل معلومات كل وحدات المجتمع . و في كلتا الحالتين فإن البيانات تستعمل في حساب احصائيات عن المجتمع و العينة.

يوفر لنا الاستقصاء على عينة معلومات سريعة و التي يحتاجها صاحب الطلب (الدراسة) و في بعض الاحيان فإن المسح الشامل يكون ضروري مثل احصاء السكان كل عشر سنوات. و هناك عدة عوامل تؤثر في اختيار نمط الاستقصاء منها :

- أخطاء الاستقصاء و هنا نجد الأخطاء المتعلقة بالعينة و الاخطاء الغير متعلقة بالعينة أو خارج العينة. إن النوع الاول من الاخطاء مرتبط بكل عملية استقصائية على العينات (سبر للاراء) و هذا ناتج عن التقديرات الناتجة من عينة و ليس كل المجتمع. إن الخطأ يتناقص مع زيادة حجم العينة و بطريقة التقدير كذلك. أما الخطأ الثاني فليس له صلة بالعينة لأنه يستعمل كل وحدات المجتمع في تقديراته. إلا أن كل الدراسات الاستقصائية تعاني من الاخطاء الغير المتعلقة بالعينة مثل أخطاء في التقدير، أخطاء في المعالجة،...الخ.
- التكلفة فالمسح الشامل يتطلب تكلفة كبيرة مقارنة بالمسح على عينة محددة خاصة فيما تعلق بعملية جمع البيانات.

- التوقيت، ففي كثير من الاحيان فإن جمع، معالجة، تقدير، نشر البيانات و النتائج تحتاج إلى وقت كبير وفي كثير من الاحيان يكون الوقت ضيق لذا فأن مدة المسح الشامل تكون أكبر من العمل على عينة محددة.
- حجم المجتمع، إذا كان حجم المجتمع صغير فمن الافضل إجراء مسح شامل و هذا من أجل الحصول على نتائج بهامش خطأ صغير. و في حالة حجم مجتمع كبير نفضل العمل بالعينات حتى نقلص من التكلفة و الوقت.
- التقديرات لمناطق صغيرة، و هذا له علاقة بالنقطة السابقة عندما نحتاج إلى إجراء تقديرات على أساس مناطق جغرافية ذات عدد سكان محدد فإننا نفضل المسح الشامل مثل إجراء دراسة حسب البلديات على المستوى الوطني.
- انتشار الخصائص (الميزات)، إذا كان الاستقصاء من أجل تقدير خصائص معينة و كانت هذه الخصائص شائعة في المجتمع فإن استعمال عينة يكون كافي، أما إذا كانت الخصائص نادرة فمن الضروري إجراء مسح شامل. مثال ذلك تحديد نسبة كبار السن بين السكان و أن هذه النسبة تقدر ب 15 بالمائة فإن دراسة عينة يمكن ان يقدر هذه النسبة بهامش خطأ صغير. لذا ينصح بالقيام بدراسات قبلية أو تمهيدية لمعرفة امكانية إنجاز الدراسة أم لا.
- البيانات الخاصة ففي بعض الاحيان نحتاج إلى بيانات خاصة لا يمكن أن نطلبها مباشرة من الشخص مثل زمرة الدم، بعض الامراض، الشروط الصحية و التي عادة ما نحتاجها في دراسات متخصصة.
- عوامل أخرى، في كثير من الاحيان تقوم الدول بإنجاز احصاء شامل للسكان كل خمس أو عشر سنوات و منه يمكن استعمال النتائج المتحصل عليها في دراسات أخرى و هذا ما يمكن اعتباره بيانات مرجعية. مثال عدد الرجال و النساء في المجتمع من أجل تقدير نسبة تمثيلية على عينة.

3-2- المجتمع المستهدف / الدراسة

راينا فيما سبق أن صياغة الاهداف تستلزم معرفة المجتمع المستهدف و هذه بعض العوامل التي تحدد المجتمع:

- الوحدة التي تكون المجتمع و خصائصها (من؟)
- الموقع الجغرافي للوحدة (أين؟)
- التوقيت (متى؟)

و من اجل تحديد المجتمع المستهدف نبدأ بالمجتمع الذي لا توجد فيه أي قائمة مثال : كل الفلاحين ثم نحدد مفهوم الفلاح. هل هو الشخص الذي يمتلك مزرعة؟ يعتبر فلاح كل شخص ....

مثال : دراسة حول دحل و مصاريف الاسر. المجتمع المستهدف هو السكان المقيمين في الجزائر في 3 ديسمبر 2016. اما المجتمع المدروس فهو سكان الجزائر في 3 ديسمبر 2016 ما عدا الذين ليس لهم عنوان ثابت.

3-3- اطار المسح

بعد تحديد المجتمع المستهدف و تعريفه تكون بعض أدوات الوصول الى الوحدات جد ضرورية مثل قاعدة الاستبيان (قاعدة المسح) و التي تسمح بالوصول الى المجتمع و هذا الاطار هو الذي يحدد مجتمع الدراسة.

مثال احصاء عام للمؤسسات الصناعية فالمجتمع المستهدف هي كل المؤسسات الناشطة الى غاية مارس 2016 في الجزائر. أما المجتمع المدروس فهو كل المؤسسات الناشطة مع موظفيها حت مارس 2016 في الجزائر.

هذا معناه أن المؤسسات بدون عمال لا تدخل في الدراسة (أي هناك مؤسسات تعمل بملاكها فقط).

تجدر الاشارة الى وجود ثلاث أنواع للوحدات :

- وحدة العينة (المستعملة في العينة)
- الوحدة المرجعية (الوحدة التي تكون المعلومة فيها متوفرة و تستعمل لحساب بعض النسب)
- الوحدة المستغلة (الوحدة التي توفر لنا المعلومات لأي هناك بعض الوحدات لا توفر لنا كل المعلومات اللازمة فيتم تجاهلها في الدراسة).

في بعض الاحيان لا نمكن التفريق بين هذه الوحدات. كما ان في إطار دراسة ما فإن بعض أو كل العناصر يجب أن تتوفر.

- بيانات تعريفية مثل الاسم، العنوان، رقم تعريفي من اجل تعريف بشكل وحيد للوحدة.
- بيانات الاتصال و هي ضرورية من أجل تحديد (تموقع) وحدات العينات عند الجمع مثال الرمز الولائي، رقم الهاتف،...
- بيانات التصنيف و هي مهمة من اجل تصنيف العينة و كذا من أجل التقديرات مثل نمط السكن، عدد العمال في مؤسسة، نجد كذلك تصنيف جغرافي، مهني.
- بيانات الصيانة و هي بيانات ضرورية نحتاجها إذا أردنا العودة الى الدراسة مرة أخرى مثل التاريخ...
- بيانات الربط و التي تسمح بربط وحدات الدراسة بمصدر البيانات و هذا حتى نتمكن من القيام بعملية التحيين لقاعدة المسح مثلا.

في دراسة ما يتطلب على الاقل بيانات تعريفية، بيانات الاتصال، التصنيف و الصيانة (مستحسنة).

3-3-1-أنواع أطر الدراسة

هناك نوعين من أطر الدراسة الاستقصائية: القائمة أو المنطقة الجغرافية.

3-3-1-1- القائمة

و هي القائمة الرسمية أو المادية لكل وحدات المجتمع و هي القائمة الفعلية أثناء إنجاز الدراسة. مثال قائمة كل السيارات التي تدخل الى حضيرة ما من 9 صباحا الى غاية 8 مساءا في اليوم. أو قائمة الطلبة المسجلين في طور الماستر. و يمكن الحصول على القوائم من عدة مصادر مثل الادارات، المؤسسات،...الخ.

أمثلة : - احصائيات الحالة المدنية (المواليد، الوفيات،...)

- فهرس أرقام الهواتف
- قائمة زبائن مؤسسة ما
- قائمة أعضاء منظمة ما.

عند استعمال هذه القوائم الصادرة من الادارات يجب الاخذ بعين الاعتبار:

- التكلفة
- التغطية (تغطية المجتمع المستهدف)
- التحيين (بيانات جديدة و حديثة)
- التعاريف (نفس التعريف المستعمل)
- النوعية
- استقرارية البيانات من المصدر (مستقرة في الزمن أي كل المتغيرات و كذا المفاهيم)
- العلاقة القانونية و الرسمية (هل هناك عقد عمل؟ هل يستلزم تصاريح لإنجاز الدراسة،...)
- سهولة الاستخدام/ الوصول (متاحة الكترونيا،...)

3-3-1-2- المنطقة

و هذا في حالة أن تكون وحدة الدراسة مناطق جغرافية (مثال قائمة للحيوانات المتواجدة في منطقة جغرافية معينة بالكيلومتر المربع، قائمة العائلات المستفيدة من إعانات حسب البلديات). و يمكن تقسيمها الى عدة فروع جزئية.

3-1-3-3- الاطار المتعدد

و هو عبارة عن تركيبة من النوعين السابقين مثال نحتاج الى قائمة مع المنطقة كحالة دراسة مسألة الصحة في الولايات في مناطق معينة. و من اجل الاخذ بعين الاعتبار مشكلة التداخل : يجب حذف التداخل و التكرارات و تصحيح التقديرات.

3-3-2- عيوب الاطار

هناك عدة عيوب منها :

- التعداد الناقص
- التعداد الزائد
- التكرار
- الخطأ في التصنيف.

3-3-3- النوعية الجيدة للاطار

يجب تقييم نوعية إطار الدراسة عن طريق أربعة معايير :

- الملاءمة Pertinence
- الدقة Exactitude
- الفرصة السانحة Opportunité
- التكلفة Coût

كما يستحسن توفر هذه الخصائص :

- نفس القيم، المفاهيم، الاجراءات.
- امكانية تحيين اطار الدراسة بالرجوع الى المصادر الادارية
- سهولة الاستعمال.

3-4- أخطاء الاستقصاء

في عالم مثالي فإنه من الممكن وجود عينة مثالية و تصميم استبيان مثالي و تكون النتائج كذلك مثالية و جمع البيانات كاملا و مثاليا و لن يكون هناك أي خطأ، إلا أن عالم الدراسات الاستقصائية ليس مثالي فهناك عدة مشاكل تظهر و عدة اخطاء ترتكب لذلك يجب الاخذ بعن الاعتبار بعض الامور في الحسبان عند إنجاز التقديرات خاصة ما تعلق بمسألة نمط تصميم الاستبيان و التحليل و الادوات (و يمكن تصنيف الاخطاء الى نوعين كما ذكرنا سالفا).

4-3-1- أخطاء المتعلقة بالعينة

لقد تم تعريف هذا فيما سبق و هو الخطأ الناتج عن تقدير خصائص المجتمع و هذا بقياس جزء من المجتمع. و بما أن كل الدراسات الاستقصائية (سبر للاراء) معرضة لهذا النوع من الخطأ فلابد من اعطاء نسبة الخطأ للمستعملين. ففي الدراسات الاستقصائية الاحتمالية هناك طرق لحساب الخطأ المتعلق بالعينة.

المقدر الاكثر استعمالا من اجل تقدير خطا العينة هو تباين العينة. فالتباين الذي يقيس تقدير خصية انطلاقا من عينات مختلفة ممكنة من نفس الحجم. فكلما كان التباين مهم فأن التقدير يكون ضعيف و غير دقيق. هناك عدة عوامل تؤثر على حجم التباين في العينة:

- حجم المجتمع
- تغير في المتغير المهم في الدراسة
- نسبة الاستجابة: فالتباين يزيد مع تناقص حجم العينة. و عذم الاجابة يتناقص مع حجم العينة. و عدم الاجابة يزيد التباين و يمكن ان تؤدي الى التحيز.
- تصميم العينة و طريقة التقدير: تصميم واحد يمكن أن تؤدي الى التباين هذا إذا كانت العينات مختلفة كما أن نموذج لاستبيان قد يؤدي الى التباين...

4-3-2-أخطاء الغير متعلقة بالعينة

الاخطاء خارج العينة هو أخطاء المتعلقة بنشاطات الدراسة ما عدا العينة و هي موجودة في نوعي الدراسات (الشاملة و العينة) و يمكن تصنيفها في شكلين:

- أخطاء عشوائية و هي أخطاء التي تؤثر في حالة استخدام العينات الكبيرة و هي تزيد في التباين.
- أخطاء تلقائية و هي الاخطاء التي تسير في اتجاه واحد و تمس جل العينة و هذا ما يؤدي الى التغير في النتائج و هي السبب الرسمي في نوعية البيانات.

تنشأ الاخطاء الغير متعلقة بالعينة من عدة امور منها:

- التغطية : و التي تنتج من السهو، التكرار، تصنيف خاطئ للوحدات...
- القياس: و الناتج عن الفرق بين الجواب المسجل و الجواب الحقيقي مثال حالة السؤال التي تكون فيه الاجابة بنعم أو لا . أما فيما يخص طريقة جمع البيانات فتؤثر على خطأ التقدير مثال : الاستعانة بالسائل أو اجابة تلقائية ذاتية. و يكون مصدر هذه الاخطاء إما السائل، المجيب، الاستبيان في حد ذاته، طريقة جمع البيانات و الاداة المستعملة في التقدير. و تعتبر أهم نقطة هي تلك المتعلقة بعدم فهم الاستبيان من طرف الشخص المجيب و هذا نظرا ل :

- استعمال مفاهيم تقنية
- عدم وضوح المفاهيم
- السؤال الغير واضح أو صياغة خاطئة
- معلومة خاطئة
- مشكلة اللغة
- ترجمة خاطئة عند استعمال عدة لغات
- عدم الاستجابة . و في حالة عدم الاجابة (الخطأ الناتج عنها) هناك نوعين : جزئي و كلي. فإذا كانت نسبة عدم الاجابة كبيرة تزيد من احتمال وجود نتائج خاطئة
- مسار الدراسة و الذي ينتج أخطاء في المعالجة الناتجة عن سوء الترميز، إدخال البيانات...

خلاصة

لقد تم عرض بعض النقاط المهمة في هذا الفصل و التي يجب أن تأخذ بعين الاعتبار عند التخطيط للدراسة الاستقصائية. فاقد تم التطرق الى الفرق بين دراسة استقصائية على العينة و الاحصاء الشامل و كذا ايجابيات و سلبيات كل طريقة.

كما تم التطرق الى الفرق بين المجتمع المستهدف و مجتمع الدراسة. و شمل هذا الفصل كذلك مفاهيم قاعدة المسح و مختلف أنواعها و التي يمكن أن تستعمل في الدراسة الميدانية إضافة الى كل هذا فأن مسألة الاخطاء المتعلقة بالدراسة تم عرضها بشيء من التعديل فأخطاء العينة (أخطاء التغطية، أخطاء المعالجة، أخطاء عدم الاجابة).

فمسائل كيفية تخطيط و تسيير دراسة استقصائية، مراحل التخطيط للدراسة هي أهم مواضيع الفصل الثالث.

المحور الرابع : طرق جمع البيانات

4-0-مقدمة

ان عماية جمع البيانات هي عبارة عن مسار لجمع المعلومات اللزمة عن كل وحدة مختارة في الدراسة. و عند جمع المعلومات فان الافراد محددين مكانا و يتم الاتصال بهم من أجل المشاركة في الدراسة. ثم بعدها تتم عملية توزيع الاستبيان و حفظ الاجابات. ان هذا المسار يستلزم تكلفة، وقت و مصادر مهمة و هذا ما يكون له اثر على نوعية البيانات.
كما انه اثناء مرحلة التخطيط يجب تحديد اداة و طريقة اجراء الدراسة.
هل يتم اجراء الاستبيان ذاتيا .....؟
مقابلة شخصية.
هل استبيان ورقي أو الكتروني؟
هل ندمج عدة دراسات؟

ان طريقة جمع البيانات يجب ان تختار حتى تسمح بالحصول على اكبر نسبة استجابة.

ان الهدف من هذا الفصل هو تقديم مختلف طرق جمع البيانات.

4-1- طرق جمع البيانات

الطرق الاساسية: و هي:

(ا) الاستجواب الذاتي يقوم الشخص المستقصي بمليء الاستبيان بدون مساعدة من أحد. هناك عدة طرق لإرجاع الاستبيان اما عن طريق البريد، فاكس، الكترونيا (مهيكلة، سهلة، تعليمات محددة و دقيقة ، شكل جيد) قليل التكلفة، نسبة الرجوع – 70 بالمئة.

(ب) المقابلة بالمساعدة (شخصية/هاتف)

المقابلات الشخصية: المستجوب يساعد المستقصي على معنى الاستبيان و عادة ما يكون في البيت / العمل/ مكان عمومي/ مطار، مركز تجاري ...

المقابلات عبر الهاتف : المستجوب يساعد الشخص على معنى الاستبيان عبر الهاتف (ورقي أو الكتروني).

هذه الطريقة : - تزيد من نسبة الاستجابة و نوعية الدراسة

طمأنة المستقصي من أهداف الدراسة

تسريع في العمل حسب الوقت
التحكم و معرفة من المجيب على الاسئلة عكس الاول
من السلبيات التأثر بالمستجوب و ضعف تكوين المستجوب.

أما الاختيار بين هذه الطرق فهناك عدة أسئلة يجب طرحها و أخذها بعين الاعتبار :

المعلومات المتاحة في قاعدة المسح
خصائص المجتمع المستهدف
طبيعة الاسئلة التي تسأل
الموارد المتاحة (المقابلات)
سهولة الاستبيان من أجل استكماله
اعتبارات خاصة
متطلبات جودة البيانات

	الذاتي	المساعدة	
		وجه لوجه	عبر الهاتف
التكلفة	منخفضة	عالية	متوسطة
الوقت	طويل	متوسط	قصير
نسبة الاستجابة	منخفضة	عالية	متوسطة/عالية

2-4- جمع المعلومات بمساعدة الحاسوب

إن الطريقة التقليدية (الورقية) لها عدة سليبيات :
يتطلب الجمع اليدوي وقتا كبيرا
يتطلب طباعة الاستبيان مما يزيد في التكلفة
يتطلب التوزيع مما يزيد في التكلفة
الاستبيان المملوء يخزن و يحفظ بكل امان و سرية.
اما طريقة الاستعانة بالحاسوب فان :
جمع البيانات يكون اليا، سريعا و يمكن ان تكون اما ذاتيا، بالهاتف او وجها لوجه.

4-3- طرق أخرى لجمع البيانات :

هناك عدة طرق أخرى :

1- المشاهدة المباشرة : هذه الطريقة تتطلب المشاهدة او التقدير المباشر للخصائص في الواقع(مثل بيانات صحية، الصور ، المسافة،...).
2- التصريح الالكتروني للبيانات : و يستعمل عندما لا يمكن استعمال الطرق الاخرى.
3- البيانات الادارية : هناك دراسات يمكن أن تحصل على البيانات مباشرة من الادارات (مصلحة الضرائب) .
4- الطرق المتعددة (المزج) : في كثير من الاحيان نحتاج الى استعمال عدة طرق لجمع البيانات و هذا ما يسمح بزيادة نسبة الاستجابة (نسبة الرجوع) لكن يمكن أن تنشأ مشكلة البيانات الغير متناسقة.
5- الطرق التكميلية و الجامعة ، استطلاعات : في بعض الاحيان يمكن ان تشترك عدة دراسات فيما بينها.

خلاصة

هناك ثلاث طرق رئيسية لجمع البيانات: استجواب ذاتي، استجواب في عين المكان و استجواب عبر الهاتف. فالاستجواب الذاتي خاصة عن طريق البريد هو عادة الطريقة جمع البيانات الاقل تكلفة لكن يعطي نسبة استجابة (الرد) ضعيفة كما انها تتطلب وقتا أكبر خاصة إذا ما كان الرد عن طريق البريد.

أما الاستجواب في عين المكان فيعطي نسبة استجابة أكبر لكن يمكن أن يتطلب تكلفة كبيرة. و هو عادة مطبقا في حالات الاسئلة المعقدة التي تتطلب تفسيرا أو زيارة العينة و تحديدها خاصة إذا كانت نسبة الامية كبيرة في المجتمع.

و في حالة الاستجواب عبر الهاتف فهو يعطي نسبة استجابة متوسطة و تكون التكلفة أقل من تكلفة الطريقتين السابقتين، كما انها الطريقة الاسرع لجمع البيانات و هي مفيدة خاصة إذا كان المجتمع/العينة موزع و مشتت جغرافيا. أن المشكلة التي تعاني منها هذه الطريقة هي العينة التي تعتمد على قائمة أرقام الهواتف التي تكون في كثير من الاحيان غير متحينة (أرقام خارج الخدمة).

كل هذه الطرق لجمع البيانات يمكن أن تكون على الورق أو الكترونيا. إن الشيء الإيجابي للاستبيان الذي يعتمد على مساعدة الحاسوب هي في جمع و إدخال البيانات اما الشيء السلبي فهي تتطلب التحكم في برمجيات خاصة.

كما ان هناك بعض الطرق الاخرى لجمع البيانات منها : المشاهدة المباشرة، التصريح الالكتروني للبيانات و كذا البيانات الادارية.

المحور الخامس : صياغة الاستبيان

5-0 مقدمة :

ان الاستبيان هو عبارة عن مجموعة من العناصر المكونة من أسئلة مصممة من أجل الحصول على معلومات خاصة بموضوع ما من المستقصي.

و يحتوي الاستبيان المستعمل في الدراسة الشاملة أو على عينة على الشكل الكتابي الاداري. و يلعب دورا أساسيا في خطوات جمع البيانات. كما له أثر مهم على نوعية البيانات المتحصل عليها.

كل سؤال في الاستبيان يجب أن يكون له مكان في الدراسة. فيجب ان تكون صياغة السؤال واضحة و مرتبة ترتيبا منطقيا، سهلة الفهم. فالاستبيان الجيد يجب ان يكون :

لجمع لبيانات بشكل فعال مع نسبة ضئيلة من الاخطاء
تقليص التكلفة
المقابلة ودية.

ان الهدف الرئيسي من هذا الفصل هو اعطاء نظرة شاملة على خطوات تصميم الاستبيان (أنواع الاسئلة، الاخطاء في الاجابات و كيفية المعالجة...)

5-1- خطوات تصميم الاستبيان

الخطوة الاولى في تصميم الاستبيان هي التعريف بأهداف الدراسة و كذا معلومات عن الهيئة المنجزة للاستبيان:

1- الاستطلاع على البيانات / المستعميلين و المجبين
2- مراجعة و قراءة استبيانات سابقة
3- كتابة الاستبيان
4- اعادة القراءة و تصحيح الاستبيان
5- اختبار و مراجعة الاستبيان
6- الاستبيان النهائي.

5-1-1- التشاور/اتحدث مع المستخدمين /المجبيين

ان الفحص أو الاتصال مع المستخدمين تبدأ منذ دراسة الاهداف في خطوة التخطيط و تتبع على طول خطوات الدراسة. كما يجب الاتصال مع المستجبين ،

الخبراء و الذين قاموا بدراسات مماثلة في الماضي. ان ملاقاة المستجبين يمكن ان تساعد على تحديد بعض الاسئلة، تحديد اللغة التي يجب أن في صياغة الاستبيان/ كما تكون المقابلات على شكل أفواج محادثات.

5-1-2- دراسة الاستبيانات السابقة

تعتبر الدراسات السابقة المماثلة مصدر مهم للمعلومة حتى يمكن تطوير الدراسة. و منه يجب دراسة الاسئلة التي يمكن ان تكون نقطة بداية في صياغة أسئلة الاستبيان (تسمية، الاسئلة المطروحة...). و في بعض الحالات هناك بعض الاسئلة يجب ان تستعمل في الدراسة و هذا حتى يتسنى لنا مقارنة الدراسات فيما بينها.

5-1-3- كتابة الاستبيان

الخطوة الموالية هي تطوير صياغة الاستبيان و التي تأخذ بعين الاعتبار العوامل التالية:

- طرق جمع البيانات
- خصائص الاشخاص المستجوبين
- صعوبات في جمع البيانات
- سرية وحساسية المعلومات
- الترجمة
- مقارنة النتائج مع الدراسات السابقة
- عوامل اخرى نذكر منها:
  - قابلية المستجوبين على الرد على الاسئلة
  - قابلية عدم الاجابة
  - قواعد الادارية و القانونية
  - انواع الاسئلة
  - كلمات و مفردات كل سؤال
  - مصدر التقدير و اخطاء في الاجابة
  - مراحل الاستبيان

5-1-4- مراجعة الاستبيان

من الاحسن مراجعة الاستبيان داخليا قبل أن يتم اختباره و هذه المراجعة يجب أن تحدد المشاكل الموجودة في الاستبيان (اخطاء املائية، نحوية، الصياغة، كما يستحسن أن تكون المراجعة من طرف أشخاص غير معنيين بالاستبيان و كذا من طرف خبراء في الميدان.

5-1-5- اختبار و اعادة الصياغة

يجب اختبار كل الصيغ الاستبيان (اللغات) على أفراد قبل البداية في عملية ادارة و جمع البيانات (سن محدد ، الجنس، المستوى...)

يجب على المجيب فهم السؤال أولا ثم البحث عن المعلومات للإجابة يسمح الاختبار ن:

معرفة صحة ترتيب الاسئلة
وضوح التعليمات
تقليص الوقت و التكلفة

و نعتمد على عينة صغيرة غير احتمالية.

الطرق هي :

اختبار قبلي رسمي (prétest) .
الاساليب المعرفية (cognitive)
مجموعات التركيز (من 6 الى 12 فرد في المجموعة)
استخلاص المعلومات /المقابلات
الترميز سلوك التفاعلات اثناء المقابلات
اختبارات تقسيم العينة
الاختبارات التجريبية (pilot)

5-1-6- انهاء الاستبيان (اللمسة الاخيرة)

ان تصميم الاستبيان هي عملية تكرارية طول اعداد الاستبيان و الاختبارات و نقوم بالتغيرات و بعد المراجعة و منه صياغة الاستبيان النهائي ثم طباعته و برمجته.

5-2- أنواع الاسئلة

هناك نوعان من الاسئلة المفتوحة و المغلقة. **الاسئلة المفتوحة** هي تلك التي لا تتوفر على اقتراحات للإجابة. و يقدم هناك المستقصى رقم/عدد أو يجيب على سؤال خاص بكلمات خاصة و يجب ترك مساحة كافية لكتابة الاجابة.

مثال: ما هي أهم مشكلة يوجهها الطالب في مساره الدراسي؟

أما **الاسئلة المغلقة** فهي تلك التي تقترح فيها اجابات و يتم الاجابة في مربع أو دائرة للجواب المناسب (فئات).

مثال : مما يلي ما هي أهم المشاكل التي يوجهها الطالب في مساره الدراسي؟

الاقامة ☐ التدريس ☐ التوجيه ☐
الاطعام ☐ النقل ☐ المنحة ☐ اخر ☐

الاسئلة المفتوحة لا تؤثر على المجيب عكس السؤال المغلق و هذا ما يساعد على الحصول على اجابات مختلفة. السؤال المغلق يقيد المجيب.

كما يسمح بالحصول على اجابات دقيقة خاصة الرقمية أو العددية مثل ما هو سنك؟

كما تستعمل لمتابعة سؤال مغلق :

مثال : هل لديك اقتراح لتحسين خدماتنا لعملائنا؟

نعم ☐ ***لا*** ☐

اذا كان بنعم، ما هي مقترحتك؟**.....**

.....

كما تستعمل في نهاية الاستبيان "تعليقات اضافية؟

هناك بعض السلبيات :

- ثقيلة لان الشخص يجب ان يعبر عن الاجابة بدون مساعدة (الاقتراحات)؟
- يستهلك الوقت للكتابة.
- صعوبة ترميز الاجابات.
- زيادة في التكلفة، الوقت، الاخطاء.

اما من ناحية الاسئلة المغلقة فهناك عدة إيجابيات :

- سهلة للمجيب.
- تقليص التكلفة.
- سريعة و سهلة لان المجيب يختار من بين عدة اختيارات و سهلة التحليل.
- سهلة للمقارنة مع الدراسات الاحرى.

كما يوجد عدة سلبيات للسؤال المغلق :

- يتطلب جهد لإعداد الاقتراحات (الفئات)
- الترميز صعب بالنسبة للفئات.
- اجبار على اعطاء رأي من بين الفئات لذا يستحسن إضافة "بدون رأي" أو "غير معني" أو "لا أدري".

هناك عدة انواع للسؤال المغلق :

- السؤال المغلق الوحيد (الثنائي) نعم او لا
- المتعدد
- القائمة
- التصنيف (الترتيب)

5-2-1- الاسئلة الثنائية الاختيارية

و هي الصيغة البسيطة للسؤال المغلق و تكون عادة على شكل نعم، لا أو ذكر/أنثى.

5-2-2- الأسئلة ذات الاختيارات المتعددة / أو القائمة

تتطلب اختيار من بين قائمة جواب واحد أو عدة اختيارات. و يمكن اضافة أخرى حدد.

مثال1 : ما هو نوع المسكن ؟ اختيار واحد.

☐ ***عمارة*** ☐ ***منزل فردي*** ☐ ***منزل جماعي*** ☐ ***منزل تقليدي*** ☐ ***أخر حدد.....***

مثال2 : أثناء العطلة، ما هو نمط الاقامة الذي تختاره؟ اختيار متعدد.

☐ ***فندق*** ☐ ***اقامة عند العائلة*** ☐ ***مخيم*** ☐ ***أخر حدد .....***

مثال 3 : ما هو سنك ؟

☐ ***30-20*** ☐ ***40-31*** ☐ ***50 -41*** ☐ ***أكثر من 51***

5-2-3-الاسئلة المصنفة :

في هذه الحالة يجب ترتيب الاجابات.

مثال : هذه بعض سبل الحصول على العمل. من فضلك رتب حسب الفاعلية 1 امام الطريقة التي تضن انها الافضل، 2 امام الطريقة الثانية و هكذا.

ـــــ الانترنت

ـــــ الورق

ـــــ ادارة

ـــــ اصدقاء

ـــــ الخواص

ـــــ مباشرة

ـــــ اخرى حدد .....

5-2-4- الاسئلة التفضيلية

يطلب من الشخص تقييم الاجابات أو اعطاء درجة التفضيل .

مثال : هل انت راضي عن الخدمة المقدمة لك في المستشفى؟

☐ ***جد راضي***

☐ ***راضي***

☐ ***غير راضي***

☐ ***غير راضي تماما***

أو اعطاء علامات من 1 الى 10 (مهم، غير مهم.....)

5-3- توجيهات لصياغة الاسئلة

يجب أن تكون الاسئلة مصوغة بشكل واضح، مفهوم بالنسية للأشخاص المستجبين. و حتى تكون بيانات الدراسة ذات نوعية جيدة يجب ان يستطيع المجيب أن يعطي اجابته بكل سهولة و يفهم معنى الكلمات المستعملة و ما هو مطلوب منه. كما يمكن ان ينتج من صياغة خاطئة نتائج مغلوطة أو غير دقيقة نتيجة ل :

إذا لم يفهم المستجيب معنى الكلمات.
إذا فسر الكلمات تفسيرا اخر.
إذا لم يكن متعودا على استعمال مصطلح ما.

و من اجل تجنب هذه المشاكل هذه بعض التوجيهات.

5-3-1- البساطة في الصياغة

أفضل طريقة للتواصل مع المستقصى ان تستعمل كلمات بسيطة: كلمات كل يوم ، مفهومة من طرف الجميع.

مثال : هل أنت واعي بالاندماج الشامل للقطاعات البيئية في منطقة ما؟

فكثير من الافراد لا يفهمون كلمات الاندماج، القطاعات....

5-3-2-التعريف بالرموز المستعملة (المختصرات)

عادة ما نجد الرموز المختصرة في الكتابات التقنية/القانونية/العلمية. فينصح بعدم استعمالها في الاستبيان.

مثال : TEOM و التي تعني الرسم على رفع النفايات المنزلية

5-3-3- قابلية تطبيق الأسئلة

يجب طرح أسئلة ذات فائدة و قابلة للتطبيق حتى نقلص الوقت و التكلفة.

مثال : ما هي وظيفة الحالية؟.....

ما رأيك في محرقة ذات 1600 درجة لمدة 30 دقيقة ....؟

5-3-4- التخصص و الدقة

يجب ان تكون الاسئلة دقيقة و متخصصة حتى يفهما المستقصى.

مثال : ما هو دخلك؟ في الوهلة الاولى يتبين ان السؤال واضح لكن غير دقيق، هل نقصد بالدخل الشهري، السنوي دخل الفرد ام دخا الاسرة، الدخل الصافي.....

5-3-5- تجنب الأسئلة المزدوجة

و نقصد هنا طرح سؤالين في سؤال واحد.

مثال هل كنت تنوي ترك سيارة في المنزل و أخذ حافلة للذهاب للعمل خلال الشهر المقبل؟

إذن يجب تقسيم السؤال الى سؤالين أفضل.

5-3-6- تجنب الأسئلة الايحائية

و هو السؤال الذي يحتوي على ايحاءات للإجابة و هذا ما قد يؤثر على النتائج.

مثال : من فضلك هل أنت متفق/تختلف أو لديك رأي حول " السياحة مفيدة بالنسبة للمقاطعة (س) و بالتالي ينبغي تعزيزها.

السؤال: هل يجب تشجيع السياحة في المقاطعة (س)؟

يجب أن تحتوي الأسئلة على جميع البدائل الممكنة.

مثال : هل تظن أن تسريح العمال يخفض نت أثار الازمة الاقتصادية ؟

☐م ☐لا بدون رأي☐

5-3-7- تجنب الأسئلة المزدوجة النافية

يجب تجنب الأسئلة المزدوجة النافية لأن الشخص لا يعرف هل هو موافق أو غير موافق.

مثال " هل أنت مع أو ضد عدم زيادة الضريبة .....؟ النفي المزدوج "ضد" و "عدم"

5-3-8- التخفيف من الاسئلة الحساسة

يمكن أت تؤدي الأسئلة الحساسة الشخصية الى تحيز الشخص المجيب.

مثال هل هناك أوقات لم تستطع فيها توفير حاجيات لأسرتك؟

يمكن أن يعطي جواب غير صحيح نظرا لحساسية المواضيع الاجتماعية (سؤال عن الدخل كذلك..)

5-3-9- التأكد من القراءة الحسنة للأسئلة

يجب أن تكون الأسئلة مختصرة و تقرأ قدر الإمكان بتمعن.

5-4- الأخطاء المتعلقة بالإجابات

في الفصل الثالث الخاص بمدخل لتحضير الدراسة هي من بين مصادر الأخطاء و هو ما يسمى بالأخطاء المتعلقة بالإجابات.

5-4-1- مصادر الاخطاء المتعلقة بالإجابة

ان الاخطاء المتعلقة بالإجابة يمكن أن تنتج غي جميع المراحل (سؤال-جواب) و يمكن أن يكون مصدر الخطأ الاستبيان، المجيب، السائل، كريقة جمع البيانات أو الوسيلة المستعملة للتقدير (في حالة التقدير المباشر)..

ان مصادر الاخطاء الناجمة عن الاستبيان تم التطرق اليها سابقا مثال : الاسئلة المغلقة تؤدي الى الجواب بنعم أو لا.

على العموم يمكن ذكر :

نمط السؤال (مغلق/مفتوح)
صيغة السؤال
طول الاستبيان (يؤدي الى تعب المجيب..)
كيفية تهيئة الورقة
معالجة الاستبيان.

كما يمكن ان يكون للمجيب عدة صعوبات لنذكر سلوكيات سابقة أو حوادث ماضية و هذا المصدر للخطأ يسمى **خطأ الذاكرة** و هذا ما يؤدي الى سوء التقدير ((sous estimations (لا يتذكر الشخص سلوك في الفترة المرجعية).

أما الحالة المعاكسة فيمكن للفرد ذكر حوادث أو نشاطات كانت خرج الفترة المرجعية و يسمى هذا المصدر للخطأ **بخطأ التداخل** (télescopage) و قد يكون التداخل الى الامام أو الى الخلف.

كما يمكن ان يكون الاخطاء ناتجة عن **مشكلة الالتماس** (الابلاغ بعدد كبير من التغيرات في فترات الالتماس بين الفترات المرجعية (حدود الفترات).

مثال في ظل 12 شهر الماضية، كم مرة زرت الطبيب؟

في الشهر الماضي، ما هي المجلات التي قرأتها؟

في الاسبوع الماضي ما هي البرامج التلفزيونية التي شاهدتها؟

كما يمكن ان يكون مصدر الخطأ السائل . فكل سائل يجب ان يطرح السؤال بنفس الطريقة، فإذا كان هناك العديد من السائلين فتعير الصيغة قد تؤدي الى خطأ.

كما يمكن ان يقوم السائل بتسجيل الاجابة عن سؤال أو سوء فهم الاجابة.

2-4-5- تقنيات الحد/تقليص الاخطاء المتعلقة بالإجابة

بعد التعرف على مصادر الاخطاء يمكن الحد منها و هذا باستعمال بعض التقنيات : (كتابة أسئلة صغيرة، واضحة حتى تساعد المجيب على التذكر و الاجابة.

مثال سؤال قصير : ما هو المشكل الصحي الذي عانيت منه في السنة الماضية؟

سؤال طويل : السؤال الموالي يتعلق بالمشاكل الصحية في السنة الماضية. و هذا السؤال يطرح على الجميع في هذه الدراسة. ما هي المشاكل الصحية الذي عانيت منها في السنة الماضية؟

اما الحد من الاخطاء الناجمة عن السائلين فتكون بصياغة السؤال جيدا و كذا اليات للمراقبة مثل اعادة السؤال مرة أخرى.

أما الاستبيان فيجب تجنبها حسب ما تم التطرق اليه. في حالة ما إذا تم اكتشاف حالة خطأ الذاكرة فيمكن استعمال التقنيات التالية :

تقليص الفترة المرجعية (صعوبة في التذكر)
.جدول زمني أو نقاط معينة مثل العطلة يمكن ان تساعد على تقليص أخطاء الذاكرة.
التذكير بالحدود أو المجالات الزمنية للحد من أخطاء التداخل.
.يمكن الاطلاع على ملفات خاصة من أجل تحديد اجابة معينة لم يستطع المجيب الاجابة عنها مثل الدخل.

خلاصة

لقد تم التطرق في هذا الفصل الى كيفية صياغة و تحضير الاستبيان. فالمرحلة الاولى تهتم بصياغة الاهداف المرجوة من الدراسة و كذا معرفة و إعلام الاشخاص المستجوبين و المستعملين للبيانات إضافة الى تقييم الاستبيان.

تأتي فيما بعد مرحلة اختبار الاستبيان من طرف مختصين أو مجموعة من الاشخاص و هذا ما يسمح لفهم و ترميز البيانات.

عموما هناك نوعين من الاسئلة: المغلقة و المفتوحة. الاسئلة المغلقة يمكن أن تكون أسئلة ثنائية أو متعددة بالترتيب أو سلم تقييمي. أما الاسئلة المفتوحة فتسمح بالتعبيير الشخصي لكنها صعبة التحليل و تحتاج الى وقت أطول فيجب احترام بعض النصائح عند صياغة الاستبيان:

البساطة
تحديد معنى الرموز و الاختصارات
الفائدة من السؤال المطروح
الدقة
تجنب الاسئلة المزدوجة
السهولة في القراءة و الفهم.

فصياغة الاستبيان يجب أن تسمح بتقليص الاخطاء المتعلقة بالاجابة. كما أن شكل و تهيئة الاستبيان لها دور أساسي م مهم في نجاح العملية. فشكل المقدمة و ترتيب الاسئلة

يمكن ان يحفز أولا الشخص على الاجابة، فيجب استعمال بعض عبارات الانتقال، نصائح للمستحيب، نصائح للسائل،...

و في النهاية يجب الاخذ بعين الاعتبار معالجة الاستبيان فيجب أن يكون سهل لجمع البيانات و إدخالها في البرامج الاحصائية الخاصة.

المحور السادس : شكل العينة

6-0- مقدمة

لقد رأينا في الفصل الثالث انه يجب تحديد نوعية الدراسة، هل هي احصاء شامل أو دراسة على عينة. إذا كانت الدراسة من النوع الثاني فلابد من تحديد نوع العينة. فالعينة هي عبارة عن وسيلة لتحديد جزء من المجتمع من أجل الحصول على معلومات و استنتاج نتائج للمجتمع الكلي.

و هناك نوعان من العينات : الغير احتمالية و الغير احتمالية. يعتمد الاختيار اساس على ما إذا كانت الاستدلالات موثوقة أو لا. فالعينات الغير احتمالية تستخدم اسلوب شخصي لاختيار وحدات المجتمع. و من أجل أن تكون النتائج جيدة يجب أن تكون العينة ممثلة للمجتمع. و هذا صعب في العينات الغير الاحتمالية.

أما العينة الاحتمالية فتتطلب تحديد وحدات من المجتمع على أساس مبدأ العشوائية أو الحظ. فهو جد معقد و طويل و مكلف عموما. و بما انه يمكن حساب احتمال ادراج كل وحدة فإنه يمكن الحصول على تقديرات موثوقة و يمكن استنتاج اسقاطات على المجتمع.

هناك عدة طرق تمكننا من اختيار عينة عشوائية : هناك عدة عوامل :

قاعدة المسح المتاحة
تغير وحدات المجتمع
التكلفة

الهدف من هذا الفصل هو تقديم أنواع العينات و العوامل التي يجب أن نأخذ بعين الاعتبار.

6-1- العينة الغير احتمالية

العينة الغير احتمالية هي طريقة لاختيار وحدات من المجتمع و هذا باستخدام أسلوب شخصي (غير عشوائي). و بما ان العينة الغير احتمالية لا تتطلب قاعدة للمسح كاملة، فهي سريعة، سهلة و غير مكلفة.

المشكل في العينة الغير احتمالية هي صعوبة معرفة هل يمكن تعميم النتائج على المجتمع أم لا.

مثال : الشخص صاحب الدراسة هو الذي يقرر من يكون ضمن العينة و هذا حسب السهولة، العلاقة الشخصية و منه فإن عدد كبير من المجتمع من المستحيل لأن يكون ضمن العينة المختارة و هذه الاخيرة يمكن أن تختلف عن الافراد المختارين.

و نظرا لمشكلة التحيز في التحديد (عموما) نفترض أن العينة تمثيلية حتى يمكن أن نستنتج نتائج عامة. و يمكن استعمال هذا النوع من العينات في الدراسات:

- كأداة للحصول على فكرة عامة/شاملة
- مرحلة المتابعة من أجل المساعدة على فهم نتائج لدراسة على عينة احتمالية.
- مثال : لفهم عادات و سلوكيات...

في بعض الاحيان تكون العينة الغير احتمالية الوسيلة الوحيدة مثال في حالة العينة التطوعية في حالات التجارب الصحية.

ايجابيات العينة الغير احتمالية هي :

سريعة و تطبيقية، واقعية (طرح أسئلة على أول 100 شخص في الشارع).
قليلة التكلفة (بعض ساعات في المكان الواحد).
لا تستحق قاعدة المسح.
مفيدة للدراسات الاستكشافية / تطوير الدراسات.

أما السلبيات فيمكن تلخيصها في :

من أجل الحصول على نتائج للمجتمع يتطلب فرضيات قوية للتمثيل العينة و هذا قد يؤثر على النتائج.
مستحيل تحديد احتمال أن تكون وحدة ما ضمن العينة و منه تقدير خطأ العينات.

و الان سنرى حمس أنواع للعينات الغير احتمالية :

1 – العينة العشوائية

2 – العينة التطوعية

3 - العينة التحكيمية

4 – العينة حسب الحصص

5- العينة الاحتمالية المعدلة.

6-1-1- العينة الفجائية (hasard sampling)

يتم تحديد الوحدات بطريقة عشوائية تحكيمية بدون تحظير أو تخطيط جزئي من قبل. و العينة الفجائية يفترض ان تكون متناسقة (homogène) (أي ان أي وحدة يمكن أن تكون ضمن العينة). مثال شخص في الشارع، فالسائل يختار الشخص الذي يصل مشيا لكن العينة تعاني من مشكلة التحييز للسائل.

6-1-2- العينة التطوعية

في هذه الطريقة الافراد المستجوبين يكونون متطوعين. و بصفة عامة المتطوعون يجب ان يكون هناك نوع من المعاينة حتى نحصل على مجموعة من الخصائص ملاءمة لمتطلبات الدراسة مثال (أشخاص لهم مرض معين).

تعاني هذه طريقة من مشكلة التحيز في التحديد لكن في بعض الاحيان تكون ضرورية.

مثال حصة راديو تطالب من المستمعين الاجابة عن سؤال بالاتصال بالحصة فالأشخاص الذين لهم رغبة في الاجابة هم من يتطوع للرد.

6-1-3- العينة التحكيمية

تكون العينة في هذه الطريقة على أساس أفكار سابقة لسلوكيات الافراد. فالخبير الذي له معرفة بالمجتمع هو من يقرر أي وحدات تكون في العينة.

فالخبير هو من يحكم من يكون ضمن العينة. يمكن أن نكون هذه الطريقة مفيدة للدراسات الاستكشافية (exploratrice) (تحديد مجموعة المحادثة من أجل اختيار جوانب الاستبيان).

6-1-4- العينة حسب الحصص

تعتبر هذه الطريقة من الطرق الاكثر استعمالا في العينات الغير احتمالية. فهي تضم عددا من الافراد معنيين (حصة) من اجل تقسيم المجتمع. و يهدف الى تحقيق هده الحجم للعينة للمجتمع التحتي (الجزئي).

يمكن ان تكون الحصص على أساس نسب (مقارنة بالمجتمع).

مثال : إذا كان هناك 100 رجل و 100 امرأة في المجتمع و العينة المختارة تضم 20 فرد. يجب أخذ 10 رجال و 10 امرأة.

و هو مفضل مقارنة بالأنواع الاخرى السابقة لأنها تفرض وجود أفراد لكل المجموعات الجزئية (الفرعية). العينة حسب الحصص تشبه نوعا ما العينة الطبقية (stratifié) (نفس الوحدات في المجموعات).

كما ان العينة حسب الحصص تختلف في طريقة تحديد الوحدات، ففي العينة الاحتمالية، الوحدات تختار عشوائيا اما في الحصص فالطريقة غير عشوائية. فالمجيب هو من يختار الافراد و يمكن ان يعوض الوحدات او تجاهل وحدات أخرى.

ان الدراسات السوق كثيرا ما تستعمل هذه الطريقة (خاصة الدراسات بالهاتف). و هو قيلي التكلفة مقارنة بالعينة الطبقية، و سهلة الادارة و الانجاز لكن تحتوي على مشكلة التحيز في التحديد. و من اجل اسقاط النتائج يجب وضع فرضيات ان الافراد متشابهون و هذا قليل الحصول.

6-1-5- العينة الاحتمالية المعدلة

و هي عبارة عن طريقة مزدوجة بين الاحتمالية و الغير احتمالية. الخطوة الاولى تكون عموما على اساس عينة احتمالية (انظر النقطة الموالية). و الخطوة الاخيرة تكون عينة غير احتمالية عادة عينة حسب الحصص.

مثال : المناطق الجغرافية يمكن ان تكون محددة باستعمال مفهوم احتمالاتي ثم يتم اختيار في كل منطقة عينة حسب الحصص.

6-2- العينة الاحتمالية

ان العينة الاحتمالية هي طريقة لاختيار العينات التي تسمح بتعميم الموضوع على المجتمع حسب المشاهدات انطلاقا من العينة. و حتى يمكن القيام بالتعميم (الاسقاط) لا بد من ان تكون العينة غير متحيزة (أي خالية من مشكلة التحديد). و تجتنب العينة الاحتمالية هذا المشكل باختيار عشوائي لوحدات المجتمع (بالاستعانة بالكومبيوتر أو بجدول الاعداد العشوائية). و تجدر الاشارة الى ان عشوائي لا يعني تحكيمي، فهي لا تترك للشخص السائل القرار في اختار من يكون ضمن العينة ام لا.

هناك معيارين اساسيين بالنسبة للعينة الاحتمالية :

الاولى هي ان الوحدات تختار على شكل فجائي.

الثاني كل وحدات مجتمع الدراسة لها احتمال الانتماء inclusion صفر في العينة. و ان هذه الاحتمالات يمكن حسابها. ليس من الضروري ان تكون نفس الاحتمال الانتماء. هناك العديد من الانواع للعينة الاحتمالية:

1- العينة العشوائية البسيطة
2- العينة الطبقية
3- العينة النسبية.

كل نوع من هذه الانواع لحالة معينة. و من اهم المزايا نذكر:

كل وحدة مختارة فجائيا و الانتماء للعينة ممكن حسابه، ويمكن حساب الخطأ المتعلق بالعينة.
يمكن القيام بالإسقاط على المجتمع.

اما العيوب فهو صعب و يتطلب وقت كبير +تكلفة مرتفعة.

1-2-6- الفاعلية الاحصائية (الجودة)

تستعمل العينة العشوائية البسيطة كمرجع من أجل تقدير الفعالية (الجودة) مقارنة مع استراتيجيات أخرى. وحتى نفهم معنى الفاعلية يجب اعطاء بعض التعاريف :

- المعلمة : هي البيانات التي نبحث عن تقديرها بالنسبة لخصية الفرد/المجتمع مثال : متوسط المجتمع، النسب.....
- المقدر و هي الصيغة التي من خلالها نقوم بتقدير المعلمة، و تحسب انطلاقا من العينة و تقدير القيمة بالاستعانة ببيانات العينة المنجزة.
- مثال : معلمة المتغير التابع (موضوع الدراسة) يمكن ان تكون متوسط المجتمع Y حسب : $\bar{Y} = \sum_{i \in U} \frac{y_i}{N}$

حيث ان :

y_i يمثل قيمة المتغير y للوحدة i

U هو مجموع الوحدات المجتمع.

N هناك N وحدة للمجتمع.

بالنسبة ل العينة العشوائية البسيطة و في حالة الاجابة بنسبة 100 بالمئة فمقدر المعلمة هو :

$$\bar{Y} = \sum_{i \in Sr} \frac{y_i}{n}$$

حيث ان Sr هو مجموع المجبيين من العينة.

القيمة $\sum_{i\in}\frac{y_i}{n}$ تسمى التقدير.

و منه فإن التقديرات تختلف حسب العينات. و المقدرات لها بغض الخصائص المفضلة.

يجب ان يكون المقدر غير جزئي (كامل) impartiale
المقدر متحيز اذا كانت قيمة المتوسط المينة الممكنة = القيمة الحقيقية للمعلمات.
يكون توزيع العينات ممركز قريب من المتوسط (خطأ العينة صغير).
خطأ العينة لمقدر يتم تقديره عن طريق تباين العينة.

استراتيجية العينات نكون فعالة مقارنة باستراتيجيات أخرى اذا كان تباين العينة اصغر في حالة لأخرى.

6-2-2- العينة العشوائية البسيطة *Simple Random Sampling (SRS)*

ان نقطة انطلاقة كل النماذج الاحتمالية هي العينة العشوائية البسيطة و هي طريقة تعتمد على خطوة واحدة و ان كل العينات الممكنة لحجم (n) لها نفس الحظ لتكون مختارة و منه كل وحدة من العينة لها نفس احتمال الانتماء. هذا الاحتمال π يساوي n/N حيث ان N هو عدد الوحدات المجتمع.

يمكن اختيار العينات ب او بدون تعويض. العينة بالتعويض هي ان نسمح لوحدة باختيارها اكثر من مرة. اما بدون تعويض فنعني ان الوحدة تختار مرة واحدة.

العينة العشوائية البسيطة بالتعويض/ بدون تعويض هما في الواقع مماثلتان اذا كان حجم العينة هو جزء صغير من المجتمع.

العينة العشوائية بدون تعويض هي أدق و أكثر عملية.

مثال :

نفترض مجتمع من 5 أفراد و أن العينة تتكون من 3 أفراد مختارة. نقوم بإعطاء الارقام 1، 2، 3، 4، 5

إذن هناك 10 عينات ممكنة

{1.2.4} , {1.2.5} , {1.3.4} , {1.3.5}, {1.4.5}, {2.3.4}, {2.3.5}, {2.4.5}, {3.4.5}
كل عينة من العينات لها نفس الحظ ان تكون مختارة. و كذا اي وحدة مختارة في 6 عينات ممكنة في كل عينة فان الفرد له احتمال $\pi=n/N = 3/5$

و من أجل تحديد عينة عشوائية بسيطة نقوم:

أولا بإعطاء أرقام للأفراد من 1 الى .N
N وحدات القائمة نختار عشوائيا باستعمال جدول الاعداد العشوائية أو بالاستعانة بالحاسوب.
مثال : نريد اجراء مسح على مستوى التلاميذ في مدرسة
لنفترض ان القائمة الخاصة بالتلاميذ متوفرة و هي قاعدة المسح. و ان المجتمع N=1530 تلميذ حيث ان حجم العينة n=90 . الان كيف نختار 90 تلميذ.
أولا : اختيار عدد من 4 ارقام (4 لان لدينا 1530)
العينة تبدأ باختيار عدد موجود في الجدول، ثم التوجه في اي اتجاه. 90 الاوائل المرقمين بأربع اعداد و لا تفوق 1530 يتم اختيارهم.
نفترض السطر 1 و العمود 85-89 تم اختيارها كنقطة بداية. في هذا العمود نختار 0189، 0256، 0984، 0749، 1441 هكذا الى غاية 90 انظر الجدول .

الجدول

Table 1: Excerpt of a Table of Random Numbers

	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85-89	90-94	95-99
00	59311	58030	52098	87024	14194	82848	04190	96574	90464	29065
01	98567	76364	77204	27062	53402	96621	43918	01896	83991	51141
02	10363	97518	51400	98342	24830	61891	27101	37855	06235	33516
03	86852	19558	64432	99612	53537	59798	32803	67708	15297	28612
04	11258	24591	36863	31721	81305	94335	34936	02566	80972	08188
05	95068	84628	35911	33020	70659	80428	39936	31855	34334	64865
06	54463	47437	73804	36239	18739	72824	83671	39892	60518	37092
07	16874	62677	57412	31389	56869	62233	80827	73917	82402	84420
08	92484	63157	76593	03205	84869	72389	96363	52887	01087	66591
09	15669	56689	35682	53256	62300	81872	35213	09840	34471	74441
10	99116	75486	84989	23476	52967	67104	39495	39100	17217	74073
11	15696	10703	65178	90637	63110	17622	53988	71087	84148	11670
12	97720	15369	51269	69620	03388	13699	33423	67453	43269	56720
13	11666	13841	71681	98000	35979	39719	81899	07449	47985	46967
14	71628	73130	78783	75691	41632	09847	61547	18707	85489	69944
15	40501	51089	99943	91843	41995	88931	73631	69361	05375	15417
16	22518	55576	98215	82068	10798	82611	36584	67466	69377	40054
17	75112	30485	62173	02132	14878	92879	22281	16783	86352	00077
18	08327	02671	98191	84342	90813	49268	95441	15496	20168	09271
19	60251	45548	02146	05597	48228	81366	34598	72856	66762	17002
20	57430	82270	10421	00540	43648	75888	66049	21511	47676	33444
21	73528	39559	34434	88596	54086	71693	43132	14414	79949	85193
22	25991	65959	70769	64721	86413	33475	42740	06175	82758	66248
23	78388	16638	09134	59980	63806	48472	39318	35434	24057	74739
24	12477	09965	96657	57994	59439	76330	24596	77515	09577	91871
...	...	...	...	...	...	...	...	...	...	...
45	12900	71775	29845	60774	94924	21810	38636	33717	67598	82521
46	75086	23537	49639	33595	31484	97588	28617	17979	78749	35234
47	99445	51434	29181	09993	38190	42553	68922	52125	91077	40197
48	26075	31671	45386	36583	93459	48599	52022	41330	60650	91321
49	13636	93596	23377	51133	95126	61496	42474	45141	46660	42338

ايجابيات عينة عشوائية بسيطة :

هي طريقة العينات الاكثر سهولة.
لا تحتاج الى اى معلومة اضافية (نحتاج فقط الى القائمة)
لا تحتاج الى تطوير تقني

السلبيات :

لا تحتاج الى معلومات اضافية قد يؤدي الى فعالية اقل.
تكلفة كبيرة اذا كانت المقابلة فردية أو جغرافية
يمكن ان تكون أخطاء في العينة اي ان نختار عينة لا تمثل المجتمع.

3-2-6-العينة المنتظمة Systematic Sampling (SYS)

في العينات المنتظمة تكون الوحدات مختارة من المجتمع على مجالات منتظمة. تستعمل العينات المنتظمة في بعض الاحيان عندما نريد استعمال العينة العشوائية البسيطة لكن لا تتوفر لدينا قائمة أو ان القائمة مختارة بطريقة فجائية. في هذه الحالة العينة المنتظمة أسهل من العشوائية. ان مجال العينات هو بداية عشوائية ضروري عندما نستعمل N حجم العينة هو ضرب العينة n .

كل وحدة k^{th} تختار عند المجال k يساوي N/n

الانطلاقة العشوائية (r) هي عدد بسيط عشوائي بين 1 و k، الوحدات المختارة هي إذن :

r, r+k , r+2k,....r+(n-1)k.

مثال: الحالة العينة العشوائية البسيطة. فإن كل الوحدات لها احتمال π يساوي n/N لكن عكس ل SRS كل التركيبات الثنائية n وحدات لها فرصة متساوية لتكون مختارة. أما SYS فلا يختار العينات التي تكون الوحدات منفصلة ب k. في هذه الطريقة فقط k عينة ممكنة الاختيار من المجتمع.

مثال : نفترض مجتمع يحتوي على N=54 وحدة. العينة n=9 مجال العينات k=N/n=54/9=6

ثم عدد عشوائي بين 1 و 6 ، لنقول 2 (يتم اختياره).

الوحدات المختارة من المجتمع لتكوين العينة هي الارقام: 2، 8، 14، 20، 26، 32، 38، 44، و 50 مع مجال للعينات يساوي 6 إذن هناك 6 عينات ألية ممكنة لمن في SRS لحجم عينة =6 هناك 25 مليون عينة ممكنة.

ايجابيات SYS : يمكن ان تستعمل عند غياب قائمة للوحدات المكونة للمجتمع.

المشكل هو ان حجم العينة n غير معروف حتى نتعرف على العينة أو نختارها.

اضافة الى مشكلة عندما يكون المجال يصادف حالات موسمية مثال حركة النقل حيث k=7 فالتقديرات تكون مختلفة اذا كان اليوم هو الاحد او الجمعة لكن الموسمية لا يمكن معرفتها من قبل.

إذا كان N لا يمكن ان يكون حد يقسم على n فالعدد k ليس عدد صحيح.

مثال N=55 n=9 k=N/n=55/9=6.1

إذا افترضنا k=6 و إذا كان r=2 فالعينة تحتوي على {2,8,14,20,26,32,38,44,50}

اما إذا كان r=1 فإن {1,7,13,19,25,31,37,43,49,55}

في هذه الحالة العينة ذات حجم 10و ليس 9 .

طريقة أخرى :

تحديد القيم r, r+k, r+2k, r+(n-1)k الى القيم القريبة. العينة تكون ثابتة هنا.

مثال : N=55 n=9 r=1 k=6.1

{1, 7, 13, 19, 25, 31, 38, 44, 50}

أما إذا كان Nلا يقبل كذلك القسمة على n ، و حتى نجتنب حجم العينة متغير، العينات المنتظمة الدائرية circulaire يمكن ان تكون منجزة بهذه الطريقة ، تعتبر الوحدات المجتمع موجودة على الدائرة و يستعمل تعداد الوحدات

قيمة k تساؤي عدد حقيقي entier الاقرب الى N/n لكن نقطة البداية عشوائية r تكون بين 1 و N و ليس بين 1 و k بمعنى الوحدة الاولى يمكن ان تكون داخل القائمة. الوحدات المختارة هي r, r+k, r+2k, r+(n-1)k

اذا كانت الوحدة j هي r+(j-1).k>N إذن الوحدة المختارة هي r+(j-1).k-N بمعنى عند نهاية القائمة تم الوصول اليها. العينة نبدأ من بداية القائمة.

ايجابيات الطرقة الدائرية هي لن وحدة لها فرصة متساوية لتكون في العينة.

مثال : نفترض N=55 n=9 k=6

البداية r تكون بين 1 و 55 لنقول r=42 وحدات المجتمع المختارة هي:

{42, 48, 54, 5, 11, 17, 23, 29, 35}

كما يمكن ذكر بعض الايجابيات:

تعتبر كحل لطريقة العينة العشوائية البسيطة عندما لا تكون هناك إطار عمل.
لا تستلزم معلومات إضافية
يمكن ان تعطي عينة أفضل تشتت
هناك نظرية معتمدة و تقديرات سهلة الحساب
سهل لأن عدد واحد عشوائي ضروري.

السلبيات :

يمكن أن تعطي عينة سيئة إذا كان المجال له علاقة بالموسمية.
عدم اعتمادها على المعلومات الاضافية يؤدي الى عينة غير فعالة.
حجم العينة النهائية لا نعرفها مسبقا.
لا يعتبر مقدر غير متحيز للتباين العينة.

6-2-4- العينة التناسبية مع الحجم Probability-Proportional-to-Size (PPS) Sampling

تعتبر هذه التقنية من التقنيات التي تستعمل بيانات اضافية (ثانوية) و تعطي احتمالات غير متساوية للانتماء. إذا كانت الوحدات المجتمع تتغير في الحجم و هذه الاحجام معروفة، هذه المعلومات يمكن استعمالها عند العينات من أجل زيادة الفعالية (الجودة) الاحصائية. فهذه الطريقة يمكن لن تزيد في الدقة إذا كانت قيم الحجم دقيقة و متغيرات الدراسة مرتبطة مع حجم الوحدات.

في حالة قيم أقل دقة للحجم، يستحسن انشاء مجموعات للأحجام و انجاز العينات طبقية (أنظر 6-2-6).

أحسن مثال ل PPS هي دراسة منطق فلاحية تستعمل عادة هذه الطريقة عندما تكون قيم الحجم هي حجم المزرعة بالهكتار. حتما حجم المزرعة يمكن ان يتطور (يزيد او ينقص) إذا كان الفلاح قد اشترى أو باع أراضي لكن اجمالا حجم الاستغلال بقي ثابتا من سنة لأخرى.

إضافة الى أسئلة خاصة بالدراسة الفلاحية مثل (الدحل، الانتاج الفلاحي، الاستغلال، التكلفة، دائما مرتبطة مع العقار).

بالنسبة للمؤسسات : عدد العمال، رقم الاعمال السنوي،... لكنها تتغير من سنة لأخرى.

في هذه الطرقة حجم الوحدات يحدد احتمال الانتماء. استعمال المزرعة كمثال يعني أن مزرعة ذات مساحة 200 هكتار لها 2 احتمال لتكون مختارة مثل المزرعة ب 100 هكتار.

مثال : نفترض ان هناك مجتمع يتكون من 6 مزارع ، نريد تقدير المصاريف الاجمالية لهذا المجتمع الفلاحي باستعمال عينة واحدة (مزرعة) (عينة ذات حجم 1 تستعمل من أجل التمثيل) نفترض ان هناك قيمة حجم مستقرة لكا مزرعة (الحجم = الهكتار) في الواقع الربح في الجودة على العينة العشوائية البسيطة نفترض ان المصاريف لكل مزرعة معروفة. الان لدينا الجدول التالي:

المزرعة	المعلومات الثانوية (الحجم الهكتار)	المصاريف (دج)
1	50	26000
2	1000	470000
3	125	63800
4	300	145000
5	500	230000
6	25	12500
المجموع	**2000**	**947300**

لهذا المجتمع المتكون من 6 مزارع، المصاريف الحقيقية هي 947300 دج.

عينة عشوائية بسيطة يمكن ان تختار حيث كل عينة بسيطة تحتوي على وحدة و كل وحدة لها احتمال 1/6 للانتماء. 6 عينات في SRS تختلف في الحجم n=1 ممكنة.

نفترض النتائج ل SRS (جدول اسفل) عن كل عينة.

لعينة ذات حجم 1، المصاريف الاجمالية للمجتمع تقدر بقرب المصاريف الوحدة العينة في قوة الوحدة. هذه القوى هي عكس احتمال الانتماء.

للعينة PPS تغير العينات هو جد صغير، تقديرات 6 عينات ممكنة هي أقل(min) 920000 الى max 1.04 مليون أفضل من SRS (الجدول 2).

احتمال الانتماء PPS يحسب حجم المزرعة قسمة حجم الكلي بكل المزارع. في هذا المثال نفترض ان هناك علاقة بين المصاريف و حجم المزرعة.

الشيء الايجابي لهذه الطريقة هي تحسين الجودة الاحصائية لهذه الاستراتيجية للعينات باستعمال المعلومات الاضافية و هذا يؤدي الى تقليص التباين.

المزرعة (العينة)	احتمال الانتماء π	قوة الترجيح 1/π	المصاريف دج	Pop estiamte of total expe
1	1/6	6	26 000	156 000
2	1/6	6	470 000	2 820 000
3	1/6	6	63 800	382 800
4	1/6	6	145 000	870 000
5	1/6	6	230 000	1380 000
6	1/6	6	12 500	75 000
			المتوسط	947 300

نلاحظ التغيرات الكبيرة في العينة عند التقدير ب SRS من 75000 الى 48 مليون دج.
اما PPT تعطي تقديرات ضعيفة لتغير العينة

المزرعة (العينة)	**الحجم**	**احتمال الانتماء π**	**قوة الترجيح 1/π**	**المصاريف دج**	**تقدير المصاريف الكلية للمجتمع**
1	50	50/2000	2000/50	26000	1 040 000
2	1000	1000/2000	2000/1000	470000	940 000
3	125	125/2000	2000/125	63800	1 020 800
4	300	300/2000	2000/300	145000	966 667
5	500	500/2000	2000/500	230000	920 000
6	25	25/2000	2000/25	12500	100 000
				المتوسط	947 300

السلبيات :

تستلزم قاعدة للمسح تحتوي على معلومات نوعية، معلومات ثانوية متحينة.
غير مطابقة اذا كانت وحدات الحجم غير صحيحة.
غير مطبقة دائما لأن كل العينات لها قيم حجم ثابتة مرتبطة مع المتغير التابع .
أقل فعالية (-)
تقدير التباين جد معقد.
تكلفة مرتفعة.

1-4-2-6- طرق العينة PPs

كيف يتم اختيار عينة PPS هناك عدة انظمة و تقنيات مستعملة : الطريقة العشوائية، الالية و الطريقة الالية العشوائية.

(أ) الطريقة العشوائية لPPS

لكل وحدة من المجتمع، تجمع قيم الحجم للوحدات حتى تضم هي .

تحديد الصف المتعلق بكل وحدة للمجتمع انطلاقا من مجموع التراكمي للوحدات السابقة الى مجموع التراكمي للوحدة الحالية.

اختيار عدد عشوائي بين 1/0 و حجم الفعلي التراكمي و اختيار الوحدة في ....التي تحتوي على العدد العشوائي.

اعادة المرحلة السابقة الى ان n وحدة تكون مختارة.

مثال

المجال	الحجم التراكمي	الحجم	المزرعة (العينة)
1-50	50	50	1
51-1050	1050	1000	2
1051-1175	1175	125	3
1176-1475	1475	300	4
1476-1975	1975	500	5
1976-2000	2000	25	6

لعينة تضم 3 وحدات، 3 أعداد عشوائية بين 1 و 2000 تم اختيارها. نفترض أن هذه الاعداد هي كالتالي: 1967، 629، 1109. ثم أن المزارع المختارة هي 5 ، 2 و 3. في حالة

الطريقة العشوائية بدون تعويض /ارجاع، إذا كان أكثر من وحدة مختارة هناك عدة تعقيدات عند محاولة الحفاظ على احتمالات التناسبية مع حجم العينة و تقديرات التباين. و هذا يصبح أكثر تعقيدا عند اختيار أكثر من 2 أو 3 في حالة PPS بدون تعويض/بدون ارجاع.

(ب) الطريقة المنتظمة ل PPS

لكل وحدة من المجتمع نجمع قيم الحجم لكل الوحدات الى غاية الوحدة المعنية.
نحدد الصف المتعلق بكل وحدة من المجتمع (غير معنية) انطلاقا من المجموع التراكمي للوحدات السابقة بالنسبة لمجموع التراكمي للوحدة الحالية.
تحديد مجال العينة n/ (الحجم الكلي التراكمي) k
تحديد نقطة البداية العشوائية r بين 0 (حجم غير حقيقي) أو 1 (إذا كان الحجم حقيقي) و k.
تحديد الوحدات في الصف الذي يحتوي الاعداد العشوائية r, r+k, r+2k, r+(n-1)k

(ج) الطريقة المنتظمة العشوائية لPPS

في هذه الحالة، القائمة عشوائية فبل تطبيق العينة المنتظمة، مثل العينة المنتظمة. غذا كانت القائمة المستعملة في الاطار الاصلي: بعض العينة الممكنة يمكن حذفها. و من أجل جعل القائمة عشوائية، العدد العينات الممكنة التي يمكن اختيارها تزيد.

5-2-6- العينة العنقودية Cluster sampling

هي عبارة عن مسار لتحديد المجموعات كاملة (عنقود) بطريقة فجائية لوحدات من المجتمع من قاعدة المسح. و هي عبارة عن طريقة للعينة أقل جودة من العينة البسيطة العشوائية. و تستعمل لعدة أسباب :

المجموعات يمكن ان تقلل من تكلفة جمع البيانات خاصة إذا كان المجتمع مشتت و يجب القيام بمقابلات شخصية.
في الواقع لا يوجد دائما عينات من وحدات فردية من المجتمع. في بعض الاحيان المجموعات تكون سهلة (مثل عائلة كلها).
تسمح بانتاج تقديرات للمجموعات (مثل دخل المتوسط للأسرة)

العينة العنقودية هي مسار بمرحلتين:

الاولى : العينات تكون في مجموعات (مدارس، أسر،...)

الثانية : تحديد عينة من المجموعات و استجواب كل الوحدات داخل العنقود المختار.

يجب ان تكون الوحدات داخل المجموعة مختلفة. و إلا الناتج خاطئة. لكن للأسف الوحدات داخل العنقود تكون دائما متشابهة.

مثال : نفترض مدينة من 100000 فرد ثم تم اختيار عينتين. بالنسبة للأولى : العينة العنقودية استعملت و مجموعة مدن تحتوي على 400 فرد اختيرت عشوائيا.

بالنسبة للثانية : استعملنا طريقة العينة العشوائية البسيطة لتحديد 400 فرد من قائمة 100000 فرد من العينة. يمكن ان تكون أكثر اختلافا في الدخل، السن، المهنة، المستوى الدراسي. أما 400 فرد في العينة العنقودية تأتي من نفس المجموعة.

الجودة الاحصائية للعينة العنقودية تكون انطلاقا من الشكل المتجانس للوحدات في المجموعات. كم عدد من الوحدات تكون في كل مجموعة. وكم هو عدد المجموعات.

عندما تكون الوحدات المختارة متشابهة احصائيا تكون أكثر جودة إذا تم اختيار عدد كبير لوحدات صغيرة من اختيار بعض الوحدات الكبيرة.

الايجابيات :

تقلل من التكلفة.
سهلة الاستعمال من العينة العشوائية البسيطة أو المنتظمة للمجتمعات طبيعيا توكن في مجموعات.
يسمح بإنجاز تقدير بالنسبة للمجموعات
جودة احصائية.

6-2-6- العينة الطبقية Stratified Sampling (STR)

في العينة الطبقية، المجتمع مقسم الى مجموعات متجانسة تمس طبقة. ثم عينات مستقلة تختار في كل طبقة. كل الطرق السابقة يمكن أن تطبق في هذه الحالة.

يمكن للمجتمع أن يكون على شكل طبقات بالنسبة لكل المتغيرات المتوفرة لكل الوحدات. مثال هذه المعلومات يمكن أن تكون عنوان الوحدة و هذا يمكن جعل الطبقات حسب المنطقة، أو الدخل و جعل طبقات بين الدحل (المتغيرات التي تستعمل عادة هي السن، الجنس، المنطقة الجغرافية، الدخل، الحجم، عدد العمال، نوع النشاط،.....

هناك ثلاث أسباب من أجل الطبقية:

1 – جعل استراتيجية للعينات أكثر فعالية من العينة العشوائية البسيطة أو المنتظمة.

2 – السهر على حجم العينة المناسب للدراسة حتى تكون التحليل ذو جودة.

3 – الحفاظ من العينة السيئة.

كما ان حجم العينة معطاة، الطبقية قد تؤدي الى خطأ في العينة صغير أو العكس.

تجدر الاشارة الى أن الوحدات في العينة العنقودية/الطبقية أن في الطبقية تختار العينات في كل طبقة عكس العنقودية أين يتم اختار العينات في العنقود و يتم استجواب الكل.

هناك تجانس داخل الطبقات.

العينات الطبقية مهمة في حالات المجتمعات المتناظرة مثال : دراسة على مستوى المؤسسات التجارية و الفلاحية هي مجتمعات متناظرة فالمؤسسات التجارية و الفلاحية لهم قيم كبيرة بالنسبة للمتغير المدروس مثل الدخل، التكلفة، عدد العمال.

في مثل هذه الحالات بعض وحدات المجتمع يمكن أن يؤثر أكبر على التقديرات إذا ما تمت اختيارها في العينة (زيادة في التقدير) و إذا لم تختار فقد تقلص من التقديرات.

إذن هذه الوحدات يجب أن تكون في طبقة لها نفس الخصائص.

و من اجل جعل المؤسسات في طبقات نأخذ مثلا متغير الحجم حسب عدد العمال.

إذا كان متغير الحجم له 3 قيم : صغيرة، متوسطة، و كبيرة. الجودة الاحصائية تكون أحسن إذا كانت المؤسسات الكبيرة لها نفس المبيعات و كذا بالنسبة للمؤسسات الصغيرة و المتوسطة.

الطبقية تسهر على حجم العينة المناسب لمجال الدراسة. عادة عند انجاز دراسة الهدف هو تقدير الكل. كم شخص كان في البطالة الشهر الماضي؟ كم كانت المبيعات الشهر الماضي؟

إضافة الى هذا، نحتاج الى تقديرات الى الافواج الجزئية للمجتمع تسمى مجالات. مثال الزبون يريد معرفة كم عدد الرجال بدون عمل مقارنة بالنساء. أو معرفة المبيعات الشهر الماضي في مجالات الالبسة.

حماية من العينات السيئة: في حالة العينة العشوائية البسيطة اختيار العينة يكون فجائي أما العينة الطبقية نحاول تحديد العينات الممكنة عند الذين أقل حدود حتى يكون الكل ممثلا.

بعد أن يتم جعل المجتمع في طبقات لابد من تحديد عدد وحدات العينة في كل طبقة. هذه الخطوة تسمى توزيع العينات répartition (سنرى اكثر في الفصل الخاص بتحديد حجم العينة).
احتمال الانتماء يختلف من طبقة لأخرى و تعتمد على الطريقة التي انجز فيها العينة في كل طبقة. و من اجل حساب الاحتمال الانتماء في الكثير من خطط العينات، الحجم العينة و حجم المجتمع في كل طبقة يجب أن يكون معتبرا، مثل:
نفترض مجتمع N=1000 و هناك طبقتين N1=250 و N2=750
نفترض استعمال طريقة العشوائية البسيطة لتحديد الوحدات n1=50 من الطبقة الاولى و n2=50 من الطبقة الثانية.
الاحتمال π1 هو احتمال ان تكون وحدة مختارة في العينة 1 هي π1=50/250=1/5
الاحتمال π2 هو احتمال ان تكون وحدة مختارة في العينة 2 هي π2=50/750=1/15
الوحدات لهم احتمالات مختلفة للانتماء، وحدة في الطبقة الاولى لها احتمال أكبر لتكون مختارة من الطبقة الثانية.

الايجابيات :

يمكن زيادة دقة التقديرات بالنسبة للمجتمع الكلي أي هناك عينة فعالة ذات جودة.
يضمن تمثيل المجموعات الثانوية المهمة عندما تكون محدد في طبقات.
أكثر عملية و إدارية.
يقلص من الوقوع في العينات السيئة
استعمال مختلف طرق العينات.
السلبيات:
تتطلب وجود قاعدة للمسح ذات جودة كبيرة من المعلومات.
إنشاء اطار الدراسة يتطلب تكلفة كبيرة و معقد من العينة العشوائية البسيطة و المنتظمة.
التقديرات معقدة مقارنة بالعينة العشوائية البسيطة و المنتظمة.

7-2-6- العينة المتعددة الخطوات Multi-Stage Sampling

الى حد الان رأينا نماذج العينات في مشهد واحد، أما الان فسنرى العينات في عدة طوابق (étage) و هو عبارة عن مسار لتحديد العينة في خطوتين متتاليتين أو أكثر.

الوحدات المختارة في الخطوة الاولى تسمى الوحدات الاساسية للعينة (PSU).

الوحدات المختارة في الخطوة الثانية تسمى هي وحدات الطابق الثاني (SSU).

وحدات كل طبقة تختلف في الهيكلة و تكون هرمية (hiérarchique) مثل الاشخاص يعشون في مساكن، المساكن تكوّن حي من المدينة، الاحياء في المدينة تكوّن مدينة،....الخ.

في العينة بطبقتين، PSU هي الوحدات الفردية للمجتمع. شكل العينة متعدد في الخطوة الاولى لتحديد المناطق، ثم عينة منتظمة للمساكن (SSU) في المنطقة في الخطوة الثانية.

مع عينة بطوابق بدرجة واحدة، كل وحدة في طبقة عينية تدمج في العينة.

في العينة بطابقين فقط بعض الوحدات في كل طابق مختارة في عينات جزئية.

تستعمل هذه الطريقة في أخذ العينات في اطار المناطق من أجل التغلب على عدم فعالية (جودة) العينات العنقودية من مجموعة واحدة لذا فهي قليلة الاستعمال.

يمكن استعمال الطرق السابقة في كل خطوة من خطوات الطريقة.

الايجابيات:

فعالية احصائية مقارنة بالعينة العنقودية
تخفيض في الوقت و التكلفة
لا تحتاج الى قائمة لكل المجتمع

السلبيات:

عموما ليس فعال أكثر من العينة العشوائية البسيطة
حجم العينة النهائية غير معروف مسبقا
تنظيم الدراسة معقد
صيغ حساب التقديرات معقدة

8-2-6- عينات متعددة المراحل Multi-Phase Sampling

رغم ان هناك تشابه في التسمية، الا ان العينة متعددة المراحل مختلفة تماما عن السابقة. يعتمد على أخذ عينتين أو أكثر، كل العينات اختيرت من نفس الاطار و الوحدات لها نفس الهيكل في كل مرحلة. بجمع معلومات قاعدية انطلاقا من وحدات العينة، ثم من عينة جزئية هذه الوحدات تجمع معلومات أكثر دقة.

مثال: نفترض أن المعلومة ضرورية حول مربي الأبقار.

دراسة بسيطة يمكن ان تنجز حول السؤال التالي " هل كل أو جزء من المزرعة مخصص لتربية الأبقار" بسؤال واحد هذه الدراسة يكون لها تكلفة ضعيفة/مجيب.

بعد ان يتم تحديد العينة الاولى، عينة صغيرة يمكن أن تختار من العينة الاولى و هذا بطرح أسئلة جزئية لمربي الابقار فقط مثال: دراسة حول المبيعات بالتجزئة في الثلاثي الاول.

المرحلة الاولى هي دراسة شهرية حول البيع بالتجزئة و الجملة. كل شهر تطلب المبيعات بالجملة/التجزئة لكل المواد (الاولية/المواد...).

الايجابيات :

رفع من دقة التقديرات (مقارنة بالعينة العشوائية البسيطة)
يستعمل من اجل الحصول على معلومات جزئية/ثانوية
يستعمل عندما تكون التكلفة لبعض المتغيرات أو متغيرات حساسة.

السلبيات:

يحتاج الى الكثير من الوقت
تحتاج الى استجواب الشخص مرتين
معقدة الإدارة و التقدير.

9-2-6- العينة المعادة Replicated Sampling

العينة المعادة تستلزم تحديد عدد معين للعينات المستقلة انطلاقا من المجتمع بدل عينة واحدة. عوضا عينة احتمالية، عدد صغير من العينات ذات حجم متساوي تسمى معادة. تحتار بطريقة مستقلة: ل واحدة تعتمد على نفس خطة العينات.

يستعمل عندما تكون النتائج أولية ضرورية و سريعة. هذه النتائج الاولية يمكن أن تعتمد على تحليل عينة واحدة معادة. السبب الرئيسي للعينة المعادة هو تسهيل حساب التباين العينات.

مثال : عوض اختيار عينة ذات حجم 1000، نقوم باختيار 10 عينات مستقلة ذات حجم 1000 يمكن أن تختار.

خلاصة

لقد رأينا في هذا الفصل المفاهيم الاساسية للعينات. فالنوعين الاساسيين هما العينة الاحتمالية و العينة الغير احتمالية. فالفائدة من العينة الغير احتمالية هي محددة للدراسات الهيئات الاحصائية لأن التحديد المتحيز للوحدات لا يسمح لصياغة انية لفروقات مجتمع الدراسة و هو سهل و سريع و ذو فائدة لدراسات البحث أو في مرحلة من مراحل الدراسة الاستقصائية (اختبار العينة مثلا).

أما العينة الاحتمالية فيمكن أن تكون مهمة عندما يجب كتابة الفروقات حول المجتمع و هذا انطلاقا من نتائج الدراسة. ففي العينة الاحتمالية كل وحدة من قاعدة المسح لها

احتمال يختلف عن الصفر ليكون ضمن القائمة و تحديد العينة هو عشوائي. ان التحديد هنا ليس متحيز فيمكن حساب احتمال الانتماء و التباين للعينة. أما سلبيات العينة الاحتمالية فهو يحتاج الى وقت و تكلفة كبرتين إضافة الى نوعية قاعدة المسح. من بين أنواع العينات الاحتمالية نجد العينة العشوائية البسيطة، النظامية، الطبقية،...الخ.

عند اختيار نمط العينة الممكنة يجب تحديد العينة الممكنة التطبيق اعتمادا على قاعدة المسح، الوحدات، مجال الدراسة، نسبة الاجابة، نمط جمع البيانات و التكلفة.

بعض النقاط المهمة :

البحث عن وجود بيانات اضافية تساعد في إنجاز الطبقات.
هل المجتمع موزع طبيعيا بين المجموعات؟ هل هي موزعة جغرافيا؟

و من أجل معرفة كيفية تحديد حجم العينة سنرى هذا في الفصل الثمن بعد عرض بعض التقديرات المهمة التي تطبق على العينات في الفصل الموالي (السابع).

المحور السابع : التقديرات

7-0- مقدمة

في الفصل الاول رأينا أن الزبون يكون محفزا عموما بالرغبة في دراسة خصائص مجتمع ما. فتفسير التقديرات الوسيلة من خلالها يمكن للهيئة الحصول على قيم لهذا المجتمع من اجل الوصول الى نتائج.

فالأساس لتقديرات في دراسة احتمالية هي ان كل وحدة من العينة لا تمثل فقط نفسها ة لكن وحدات أخرى من مجتمع الدراسة. فمن الطبيعي استعمال عدد المتوسط للوحدات، القوة، الترجيح....فتحديد هذه القوة هي مرحلة مهمة في الدراسة. وهذا من أجل القيام بتعديلات. وهذه التعديلات يجب أن تأخذ بعين الاعتبار نقطتنهما: - عدم الاجابة – استعمال بيانات ثانوية.

بعد حساب قوة التقدير النهائي يمكن ان نستعمل في البيانات العينة من أجل حساب التقديرات.

إن خصائص العينة تسمى بالمتغيرات و تكون إما متغيرات كيفية مثال (الجنس، الحالة العائلية،....) أو كمية مثل (السن، الدخل، ...). وحسب طبيعة المتغير هناك صيغ ونماذج للتغيير. كما أن تقدير خطأ العينة هو بمثابة الجزء المهم في التقدير و هذا ما يسمح بالحصول على نوعية جيدة في التقديرات.

الهدف من هذا الفصل هو معرفة كيفية حساب قوة الترجيح، القيام ببعض التقديرات و تقدير الخطأ المتعلق بالعينة من أجل دراسة مسحية احتمالية.

7-1- الترجيح

المرحلة الاولى في التقدير هي إعطاء قوة لكل وحدة من العينة أو لكل وحدات العينات. و تعرف القوة على انها عدد متوسط للمجتمع الدراسة. القوة هي wd حيث ان d هو رمز للوحدة من العينة، وهي معاكس احتمال الانتماء π.

نذكر أن من اجل انجاز عينة متعددة الطوابق/ الخطوات، فان الاحتمال لتحديد وحدة هي احتمال المزدوج لتحديد الوحدة في كل طابق/أو خطوة. و في حالة عينة من مرحلتين أين

يكون احتمال التحديد لوحدة هو π1 في المرحلة الاولى، π2 في المرحلة الثانية، فإن القوة لتلك الوحدة من العينة هي :

$$wd = \frac{1}{\pi 1} * \frac{1}{\pi 2}$$

فالبيانات لعينة الدراسة تكون عموما متوفرة في ملف التسجيلات لكل وحدة من العينة. تذكير في العينة الاحتمالية كل وحدة لها احتمال معروف π ليكون مختار (محدد).
إذا كانت قيمة احتمال الانتماء (مثلا 50) إذن كل وحدة مختارة تمثل في المتوسط 50 وحدة من مجتمع الدراسة و قوتها هي $wd = 50$
إذا كانت القوة عدد حقيقي فان وسيلة انتاج التقديرات للمجتمع هي اعادة انتاج كل تسجيل حتى تكون 50 نسبة لكل واحد، ثم حساب قيم الدراسة من متوسط، المجموع، النسب،....انطلاقا من ذلك الملف.
أما عندما تكون القيم غير حقيقية (مثل إذا كانت وحدتين من 5 مختارة باستعمال الطريقة العشوائية البسيطة فان القوة $wd = 2.5$)
عموما من الصعب ادخال متغير القوة من أجل تسجيل كل الوحدات.

1-1-7- الترجيح في شكل العينة الاحتمالية المتساوية

الترجيح التلقائي يكون عندما تكون القوة نفسها لكل وحدات العينة. و هذا ينتج عندما تكون كل وحدة لها نفس احتمال الانتماء.

بالنسبة للنماذج ذات الترجيح الذاتي (التلقائي) إذا كان التعديل السابق قد طبق على قوة (مثال بالنسبة لعدم الاجابة أو البيانات الثانوية) القوة يمكن أن تهمل في ايجاد بعض الاحصائيات مثل النسب، المتوسطات،...الخ.

ما هو نمط ذو مرحلة واحدة يكون ذو ترجيح ذاتي؟

العينة العشوائية البسيطة و المنتظمة تكون ذات ترجيح ذاتي لأن كل وحدة لها فرصة متساوية لتكون مدمجة في العينة. أما في حالة النمط الطبقي فالترجيح الذاتي هو المختار. مثال عندما نختار العينة العشوائية البسيطة في كل طبقة و أن حجم العينة في كل طبقة يكون نسبي بالنسبة لحجم المجتمع في الطبقة.

كما أن قسمة العينة نفسها في كل طبقة و ان كل الوحدات للمجتمع لها نفس احتمال الانتماء (توزيع العينة بين الطبقات بهذه الطريقة تسمى N-repartition proprtionnelle).

مثال 7-1- الطبقية في العينة العشوائية البسيطة ذات N توزيع نسبي.

نفترض مجتمع متكون من N=1000 فرد مقسم الى طبقتين حسب اطار الدراسة. الطبقة الاولى مكونة من N1=400 رجل و الثاني N2=600 امرأة. العينة الاجمالية ذات حجم n=250 بالنسبة للطبقتين قد اختيرت. و ان توزيع العينة لكل طبقة يكون النسبي. تقسيم العينات في طبقة يكون إذن متساوي.n/N=250/1000=1/4

جدول عينة طبقية بالطريقة العشوائية البسيطة التناسبية N

الطبقة	حجم المجتمع	حجم العينة
رجال	400	100
نساء	600	150
المجموع	1000	250

قيمة احتمال الانتماء في كل طبقة هو :

طبقة 1 رجال $\pi 1 = \frac{n1}{N1} = \frac{100}{400} = \frac{1}{4}$

طبقة 2 نساء $\pi 2 = \frac{n2}{N2} = \frac{150}{600} = \frac{1}{4}$

و منه فإن الكل لها نفس احتمال ليكون مختار و نفس القوة wd=1/π=4.

أما بالنسبة لنمط متعدد الخطوات فان الترجيح الذاتي يتحصل عليه باختيار الطبقات ذات احتمال نسبي بلالة الحجم في كل الخطوات ما عدا الاخيرة. في المرحلة الاخيرة عدد ثابت من الوحدات في الطبقة يختار (مثال دائما تختار n=5 في المرحلة الاخيرة). العينة PPS يستعمل دائما في انماط المتعددة الخطوات لأنه يمكن أن تؤدي الى عينة ذات ترجيح ذاتي و هذا بالتحكم في حجم العينة.

مثال لنمط لخطوتين ذات ترجيح ذاتي يكون SRS أو عينة منتظمة أو عينة طبقية ذات توزيع نسبي N مختار في كل مرحلة.

2-1-7- الترجيح في شكل العينة الاحتمالية الغير متساوية

رغم أن النماذج ذات الترجيح الذاتي و مفضلة مثال في حالة أنماط طبقية في الدراسات الوطنية للمناطق الصغيرة، التوزيع N النسبي يمكن أن يؤدي الى حجم العينة غير ملاءم. أما بالنسبة للمناطق الكبيرة يؤدي الى عدم فعالية العينات الكبيرة.

المثال التالي يبين قوة العينات الطبقية أين حجم العينة في كل طبقة ليس نسبي مع حجم المجتمع في الطبقة.

مثال 7-2- عينة SRS طبقية ذات توزيع غير نسبي

مثال دراسة عن النقل العمومي. لدينا مجتمع من N=1100 فرد مقسم الى طبقتين (2) جغرافيتين. و بما أن السكان يعيشون في المناطق الحضرية و الريفية جد مختلفة فيما يخص المعلومات المجمعة انطلاقا من الدراسة تصميم العينة الطبقية.

الطبقة الحضرية ذات حجم N1=1000 إذن الطبقة الريفية N2=100 عينة من n=250 مختارة فان n1=200 في الطبقة الاولى، n2=50 في الطبقة الثانية. ما هي قوة المعطاة للوحدات المختارة.

طبقية SRS ذات توزيع غير نسبي

الطبقة	حجم المجتمع	حجم العينة
الحضرية	N1=1000	n1=200
الريفية	N2=100	n2=50
المجموع	N=1100	n=250

احتمال الانتماء

طبقة 1 حضري $\pi 1 = \frac{n1}{N1} = \frac{200}{1000} = \frac{1}{5}$

طبقة 2 ريفي $\pi 2 = \frac{n2}{N2} = \frac{50}{100} = \frac{1}{2}$

في ملف العينة كل مجيب في الطبقة الحضرية له قوة wd1=5 ، كل مجيب في الطبقة الريفية له قوة wd2=2

اما بالنسبة لعينات ذات خطوات متعددة القوة الاجمالية تحسب بضرب احتمال الاختيار لكل طبقة (خطوة) ثم نأخذ العكس : مثال نفترض عينة عنقودية ذات مستويين مختارة عن طريق العينة العشوائية البسيطة n1=10 من N=100 في المرحلة الاولى و n2=30 في كل فوج في المرحلة الثانية حيث أن عدد الوحدات في كل عنقود (طبقة) هو N2=60 .

احتمال الاختيار في المرحلة 1 هي $\pi 1 = \frac{n1}{N1} = \frac{10}{100} = \frac{1}{10}$

احتمال الاختيار في المرحلة 2 هي $\pi 2 = \frac{n2}{N2} = \frac{3}{60} = \frac{1}{2}$

حيث أن قوة التصميم

$$wd = \frac{1}{\pi 1} * \frac{1}{\pi 2} = 10 * 2 = 20$$

7-1-3 تعديل القوة بالنسبة لعدم الإجابة

ما يمكن ملاحظته أن كل الدراسات تعاني من مشكلة عدم الاجابة و هذا ما يؤدي الى نقص في المعلومات المطلوبة انطلاقا من العينة. و تجدر الاشارة الى أن هناك نوعين من عدم الاجابة: عدم الاجابة لعنصر واحد او لكل العناصر.

في هذه الحالة المقاربة المعمول بها تتطلب فرض القيم الناقصة.

سنركز هنا على عدم الاجابة الكلية (حالة كل او جزء كبير من البيانات مثل وحدة من العينة ناقصة). و هذا ينتج عندما نرفض وحدة من الاجابة (المشاركة) لا يمكن معرفة تمركز الوحدة. أحسن طريقة لمعالجة عدم الاجابة هنا هو تجاهلها. في بعض الاحيان تكون (النسب/المتوسطات) الاحصائية نفسها سواءا تم التعديل أو لا لعدم الاجابة. لكن عدم تعويض الوحدات عدم الاجابة بالاستعانة بالتعديل ajuster يؤدي الى استنقاص من التقدير الكلي (مثل حجم المجتمع، الدخل الاجمالي، او المساحة الكلية المزروعة،...الخ).

الطريقة الاكثر استعمالا لمعالجة عدم الاجابة الكلية هي تعديل قوة التصميم على أساس فرضية أن وحدات المجيبين تمثل الوحدتين المجيبين و الغير المجيبين.

و هذا ما يعتبر منطقي تحت الفرضية أن بالنسبة للخصائص المقدرة في الدراسة، عدم الاجابة هي مثل المجيبين، قوة التصميم لعدم الاجابة يعاد توزيعها على المجيبين و هذا يستعمل عادة عن طريق عامل التعديل لعدم الاجابة و الذي يضرب في القوة من أجل انتاج قوة معدلة او مصححة لعدم الاجابة. بعض الدراسات يكون قوة التصميم الاولية wd=1 .

فعامل التعديل لعدم الاجابة يكون معرفا كعلاقة مجموع قوى العينة الاولية في مجموع قوى وحدات المجيبين. في حالة الترجيح الذاتي هو العلاقة بين عدد الوحدات في العينة الاصلية بعدد وحدات المجيبين.

مثال 3-7- في حالة العينة العشوائية البسيطة n=25 وحدة في العينة اختيرت من مجتمع N=1000 لنقول أن عدد وحدات المجيبين (nr) فقط nr=20 شخص اعطوا إجابات صحيحة و كاملة. ما هي قوة المعدلة لعدم الاجابة لوحدات العينة؟

1- المرحلة الاولى تعتمد على حساب احتمالات انتماء العينة العشوائية البسيطة :

$$\pi = \frac{n}{N} = \frac{25}{100} = \frac{1}{4}$$

إذن قوة التصميم لكل وحدة في العينة wd=4 .

2- المرحلة الثانية تعتمد على حساب عامل التعديل لعدم الاجابة. لدينا فقط nr=20 من n=25 مختارة أعطت معلومات كاملة، حجم العينة النهائية هو 20. لنفترض أن الوحدات المجيبة يمكن أن تستعمل لتمثيل وحدتين، فعامل التعديل لعدم الاجابة هو :

$$\frac{n}{nr} = \frac{25}{20} = 1.25$$

3- المرحلة الاخيرة تعتمد على حساب قوة التعديل لعدم الاجابة. قوة التصميم لعدم الاجابة المعدل Wnr هو ناتج قوة التصميم و عامل التعديل لعدم الاجابة

$$Wnr = wd\frac{n}{nr} = 4 * 1.25 = 5$$

إذن كل مجيب يمثل 5 أفراد في مجتمع الدراسة. القوة النهائية 5 تعطى لكل وحدة في البيانات.

إذا كانت الفرضية أن كل عدم المجيبين لهم مثل المجيب فيما يخص الخصائص المقدرة في الدراسة صحيحة، نفس عامل التصحيح (التعديل) لعدم الاجابة يمكن أن يستعمل لكل وحدات المجيبين. (المثال السابق).

إلا ان هناك في بعض الأحيان مجموعات ثانوية لها نسب اجابة مختلفة و خصائص مختلفة، في هذه الحالة التعديل للجميع يمكن أن يغير النتائج (خاطئة).

مثال : الاسر التي لها شخص واحد لها نسبة الاجابة ضعيفة و خصائص مختلفة من أسرة لها عدة أفراد . في هذه الحالة التعديل لعدم الاجابة مختلف يجب ان ينجز.

في المثال الاتي عامل التعديل لعدم الاجابة يطبق في كلتا الطبقتين(حضري، ريفي)

مثال -2-7- تكميلي

أثناء جمع البيانات هناك فقط nr1=150 فرد في الطبقة الحضرية و nr2=40 فرد في الطبقة الريفية و تتوفر على المعطيات كاملة.

ما هي القوى المعدلة لعدم الاجابة لهؤلاء الافراد؟

الطبقة	حجم المجتمع	حجم العينة	عدد المجيبين
الحضرية	N1=1000	n1=200	nr1=150
الريفية	N2=100	n2=50	nr2=40
المجموع	N=1100	n=250	nr=190

1- قوة التصميم في كل طبقة هي wd1=5 و wd2=2

2- عامل التعديل لعدم الاجابة

الطبقة الاولى $\frac{n1}{nr1} = \frac{200}{150} = 1.33$

الطبقة الثانية $\frac{n2}{nr2} = \frac{50}{40} = 1.25$

3- القوة المعدلة لعدم الاجابة لكل طبقة هو ناتج قوة التصميم مع عامل التعديل.

الطبقة الاولى $wnr1 = wd1\frac{n1}{nr1} = 5 * 1.33 = 6.67$

الطبقة الثانية $wnr2 = wd2\frac{n2}{nr2} = 2 * 125 = 2.5$

في ملف العينة كل مجيب في الطبقة الحضرية له قوة نهائية تساوي 6.67 و كل مجيب في الطبقة الريفية له قوة نهائية تساوي 2.5.

كما انه عند حساب معمل التعديل لعدم الاجابة، يجب الاخذ بعين الاعتبار انه هناك بعض وحدات العينة تكون بعيدة (مثل وحدة لا تنتمي الى مجتمع المستهدف) . ففي مثال الدراسة على كل مستويات البيع بالتجزئة هناك بعض المعلومات يمكن ان الصورة غير صحيحة و أن وحدة هي العينة يمكن أن تكون خطأ كبير. تعديل عدم الاجابة لا يجب أن يحسب مقارنة بوحدات داخل مجال الدراسة. هذا لأن الوحدات خارج مجال الدراسة يمثل عموما وحدات أخرى خارج الاطار. لأن وحدات خارج المجال لم يكنوا ليجيبوا على الدراسة. نسبة الاجابة يمكن أن تكون 100 بالمائة.

المثال السابق يفترض أن كل عدم المجيبين هم داخل المجال، فعوامل التعديل المتحصل عليها يمكن ان تكون مختلف حسبما نظن أنه العدد الصحيح لقيم المجيبين على الدراسة.

7-1-4- استعمال البيانات الثانوية لتعديل القوة

إن قوى التصميم المضروب في عامل التعديل لعدم الاجابة يمكن أن يستعمل من أجل انتاج قوى نهائية و تقدير الخصائص المراد دراستها. لكن في بعض الاحيان البيانات حول مجتمع الدراسة تكون متوفرة انطلاقا من دراسات أخرى سابقة أو مصادر أخرى، هذه المعلومات يمكن كذلك إدماجها في مسار الترجيح.

هناك سببين من اجل استعمال البيانات الثانوية :

أولا: لمعرفة التقديرات الاجمالية (المعروفة و المقدرة) مثال عدة دراسات اجتماعية تعدل تقديراتها من اجل أن تكون ملاءمة مع تقديرات (السن، الجنس،...) مع الاحصاء السكان الاخير.

المعلومات الثانوية يمكن ان نتحصل عليها كذلك من الادارات أو دراسات أخرى نظرا لحجم العينة الكبير أو أن التقديرات المنشورة يجب ان تحترم.

ثانيا : استعمال البيانات الثانوية تزيد من تحسين دقة و جودة التقديرات.

مثال : إذا كان سن الشخص غير متوفر في الاطار العام للدراسة لا يمكن ان يستعمل في انجاز الطبقات حسب السن للمجتمع. إذن إذا كانت المعلومة قد تم جمعها في الدراسة و ان تقديرات الدراسة يمكن ان تعدل من اجل أن تتلاءم مع توزيع السن في الاحصاء. إذا كان السن مرتبط مع متغيرات أخرى في الدراسة (الصحة) و استعملت بيانات ثانوية يمكن أن تحسن جودة التقديرات.

كما ان هناك 3 شروط لاستعمال البيانات الثانوية في مرحلة التقدير:

- ارتباط جيد للبيانات الثانوية مع متغير الدراسة
- مصادر المعلومات صحيحة و دقيقة
- معلومات كاملة لكل وحدات العينة المجيبة.

7-1-4-1- بعد انجاز الطبقية

تستعمل مرحلة بعد الطبقية من أجل تعديل قوى الدراسة و هذا باستعمال متغيرات ملاءمة للطبقية لكن لا يمكن أن تستعمل في مرحلة التصميم لأن البيانات غير ملاءمة أو غير متحينة. فالمعلومات الجيدة للمجتمع أصبحت متوفرة بعد تحديد العينة.

و تستعمل مرحلة بعد الطبقية عندما تكون البيانات الثانوية على شكل تعداد مثل (عدد الرجال، النساء في المجتمع). و تجدر الاشارة الى ان استعمال الطبقية في مرحلة التصميم أفضل من مرحلة بعد الطبقية.

مثال التالي يبين كيفية استعمال بعد الطبقية من أجل تحسين التقديرات لعدد من المدخنين في المؤسسة.

مثال-4-7- تعديل القوى بعد الطبقية باستعمال SRS

نفترض أن دراسة أنجزت من أجل الحصول على بيانات حول عادات المدخنين في مؤسسة صغيرة. و قد تم استعمال SRS لاختيار n=25 فرد مختارة من قائمة N=78 عامل. في مرحلة التصميم لا توجد أي معلومة ثانوية متوفرة يمكن أن تستعمل من أجل انجاز الطبقية. إضافة الى بيانات حول عادات التدخين، السن و الجنس لكل مجيب متوفرة في الاصل n=25 و n_r=15 مجيب حسب التوزيع التالي:

	قبل الطبقية 1 رجال	قبل الطبقية 2 النساء	عدد المجبيين
عدد العمال	3	12	15
عدد المدخنين	1	7	8

1- احتمال الانتماء لكل وحدة في العينة هو :

$$\pi = \frac{n}{N} = \frac{25}{78} = 0.32$$

قوة التصميم

$$wd = \frac{1}{\pi} = \frac{1}{0.32} = 3.12$$

2- معامل التعديل لعدم الاجابة، نفترض أن الجميع في الدراسة لهم نفس احتمال الاجابة في الدراسة (بمعنى المجموع يساوي 1 لعدم الاجابة) هي:

$$\frac{n}{nr} = \frac{25}{15} = 1.67$$

3- قوة التعديل لعدم الاجابة هي :

$$wnr = wd * \frac{n}{nr} = 3.12 * 1.67 = 5.2$$

إذن كل مجيب له نفس القوة المعدلة لعدم الاجابة wnr=5.2 استعمال هذه القوة يعطي لنا التقديرات التالية:

	رجال	النساء	المجموع
عدد العمال	15.6	62.4	78
عدد المدخنين	5.2	36.4	41.6
نسبة المدخنين	0.33	0.59	0.53

إن القوة المعدلة لعدم الاجابة تؤدي الى تقدير حوالي 16 رجل و 62 امرأة تعمل في المؤسسة مع حوالي 33بالمائة رجال و 59 بالمائة نساء في الشركة مدخنين.

نفترض أن بعد انجاز الدراسة توفرت معلومات اضافية : هناك 42 رجال و 36 امرأة تعمل في المؤسسة، التقديرات انتجت عن طريق الدراسة مختلفة عن البيانات الحقيقية، و منه فالهيئة المعنية بالدراسة تريد ان تكون النتائج حقيقية. كما أن الهيئة تشك في عادات التدخين المرتبطة بعنصر الجنس. ان الطبقية البعدية يمكن ان تحسن من دقة التقديرات لو ان هذه البيانات الثانوية كانت متوفرة قبل بداية الدراسة من أجل القيام بدراسة على أساس الجنس. ماذا يمكن القيام به الان؟

يمكن انجاز الطبقية للعينة بعد حدوث الدراسة و تحديد القوة الترجيحية بعد الطبقية (البعدية) و التي يمكن ان نستعمل أثناء التقديرات و نرمز له ب wpst و هو ناتج قوة الترجيح المعدلة لعدم الاجابة wnr و معامل التعديل المعدل للطبقية البعدية.

فمعامل التعديل الطبقية البعدية يحسب لكل طبقة و هذا المعامل هو العلاقة بين وحدات الرجال في المجتمع في الطبقية البعدية N و عدد الوحدات المقدرة للمجتمع في الطبقية البعدية $\widetilde{N}$ عدد الوحدات المقدرة للمجتمع باستعمال قوة التعديل لعدم الاجابة.

4- في هذا المثال فمعامل التعديل للطبقية البعدية

الطبقة الاولى $\frac{Nmen}{\widetilde{N}men} = \frac{42}{15.6} = 2.69$

الطبقة الثانية $\frac{Nwomen}{\widetilde{N}women} = \frac{36}{62.4} = 0.58$

ملاحظة استعمال SRS في هذا المثال لنفس الصيغة $\frac{N}{\widetilde{N}}$

و عند استعمال القوة المعدلة لعدم الاجابة تنتج القوة النهائية للطبقية البعدية.

الطبقة الاولى $\text{wpst/men} = wnr * \frac{Nmen}{\widetilde{N}men} = 5.2 * 2.69 = 14$

الطبقة الثانية wpst/women= $wnr * \frac{Nwo}{\tilde{N}women} = 5.2 * 0.58 =$

3- باستعمال القوة للطبقية البعدية اصبحت التقديرات على الشكل التالي:

	رجال	النساء	المجموع
عدد العمال	42	36	78
عدد المدخنين	14	21	35
نسبة المدخنين	0.33	0.59	0.45

أصبحت الان تقديرات الخاصة بالرجال و النساء ملاءمة مع المجاميع المعلومة للرجال و النساء في المؤسسة. و في حالة ما إذا كان الجنس مرتبط فإن نسبة و عدد المدخنين فإنه يمكن الحصول على تقديرات دقيقة.

نلاحظ ان نسبة المدخنين في كل طبقة بعدية لم تتغير لكن نسبة المدخنين في المجتمع الكلي.

2-4-1-7- النسب التقديرية

إحدى الطرق المستعملة في الدراسات من أجل إضافة البيانات الثانوية من أجل تحسين التقديرات هي النسب التقديرية. عندما تكون البيانات على شكل أعداد فإن النسبة التقديرية هو الطبقية البعدية. بصفة عامة ، مع النسبة التقديرية الترجيح للبيانات لمجموعة التصنيفات تكون معدلة بمعامل الضرب هذا المعامل هو النسبة التقديرية انطلاقا من البيانات الاضافية من أجل تقدير الدراسة لنفس المتغير لمجموعة التصنيف.

مثال : إذا كان هدف الدراسة هو تقدير مساحة القمح في مناطق محددة ، و أن المساحة الكلية في كل منطقة معروفة و يمكن أن تكون بيانات اضافية مفيدة. إذا كانت مساحة القمح مرتبطة مع المساحة الكلية للمنطقة، هذه المعلومة الاضافية يمكن أن تحسن التقديرات لمساحة القمح. ان استعمال النسب التقديرية ، فعامل التعديل لكل منطقة يمكن ان يكون المساحة الكلية لأراضي المنطقة قسمة المساحة الكلية المقدرة للمنطقة $\left(Y/\hat{Y}\right)$.

في المثال 4-7- : النسبة هي عدد الرجال قسمة عدد الرجال المقدر و هو نفسه بالنسبة للنساء.

يمكن تطبيق النسبة التقديرية لكل طبقة إذا كانت المعلومة الاضافية متوفرة على هذا المستوى و حجم العينة ملاءم في كل طبقة.

7-2- التقديرات البسيطة (الاجمالية، المتوسطات، النسب)

لقد رأينا في الامثلة السابقة كيفية حساب القوى القاعدية و تعديلها في حالة عدم الاجابة و كذا في حالة البيانات الاضافية. سنرى الان كيفية الخصول غلى تقديرات باستعمال القوى النهائية (قوة التقدير).

7-2-1- التقديرات لمختلف أنواع البيانات

في جل الدراسات فإنه يتم تقدير الاحصائيات الوصفية (الاجمالية، المتوسطات، النسب،...)كما انه يتم كذلك حساب و تقدير احصائيات جد معقدة لمختلف المتغيرات سواءا كانت كيفية او كمية.

ان بعض الاحصايات تكون مناسبة للمتغيرات الكيفية مثل الجنس، الخالة المدنية، ...أما فيما يخص أسئلة الرأي فإن الاجوبة تمون عادة على شكل سلم القبول مثال موافق تماما، غير موافق، غير موافق... فالاجوبة للأسئلة التالية يمكن ان تساعد على تحديد كيفية حساب التقديرات:

ما هو نوع الاحصائيات المطلوبة؟ كلية ، متوسطة، نسب؟
ما هي القوى النهائية؟
ما هي مجالات الفائدة ؟

إن التقديرات التالية يمكن ان تستعمل في كل أنماط العينات الاحتمالية (البسيطة/ المنتظمة) أو المعقدة.

تقدير كلي للمجتمع : ان تقدير العدد الاجمالي لوحدات المجتمع المدروس بالنسبة للبيانات الكيفية و الكمية يحسب بجمع القوى النهائية (المعدلة) للوحدات المجيبة.

$$\widetilde{N} = \sum_{i \in S_r} w_i y_i$$

حيث أن (i) هو الوحدة المجيبة في العينة
w_i هي القوة النهائية
S_r مجمل الوحدات المجيبة.

تقدير العينة الاحمالية للبيانات الكمية (المصاريف الاجمالية) هي ناتج القوى النهائية w_i و قيمة y_i لكل وحدة مجيبة أين قمنا بمجموع الوحدات المجيبة .

$$\hat{y} = \sum_{i \in S_r} w_i y_i$$

تقدير متوسط المجتمع : تقدير قيمة متوسطة للمجتمع الخاص بالبيانات الكمية نتحصل عليه بجمع ناتج القيم المشاهدة و القوى النهائية لكل وحدة مجيبة و تقسيمها على القوى، بعبارة أخرى المتوسط في المجتمع هو تقسيم القيمة الكلية القيم الكمية على عدد الوحدات في المجتمع.

$$\hat{\bar{y}} = \frac{\sum_{i \in S_r} w_i y_i}{\sum_{i \in S_r} w_i} = \frac{\hat{Y}}{\hat{N}}$$

تقدير النسب في المجتمع : غن تقدير النسب في المجتمع المدروس و له خصي معينة بالنسبة للبيانات الكيفية هو جمع قوى الوحدات ذات نفس الخصية و تقسيمها على مجموع الوحدات لكل المجبيين و بعبارة أخرى هي تقدير عدد الكلي للوحدات ذات نفس الخصية و تقسيمها على العدد الكلي للوحدات المجتمع.

$$\hat{p} = \frac{\sum_{i \in S_r \cap C} w_i}{\sum_{i \in S_r} w_i} = \frac{\widehat{N_c}}{\widehat{N}}$$

حيث C هي الوحدات التي لها نفس الخصية

2-2-7-أمثلة عن التقديرات للقيم الكلية، المتوسطات، النسب

سنرى المثال -6-7- التقديرات في حالة وجود قوى نهائية متفاوتة

تمت دراسة من أجل الحصول على المعلومات بالنسبة لمجتمع فلاحي (مزارع) العينة n=10 مزرعة محددة باستعمال عينة طبقية. العشر مزارع تجيب و لا يوجد تعديل للقوة. القوة النهائية إذن مساوية للقوة المرجعية (القاعدية).

لدينا نوع المزرعة 1= فلاحة ، 2= تربية المواشي

حجم العينة الفلاحية $n_1=6$ و حجم الطبقة الفلاحية $N_1=34$

حجم العينة تربية المواشي $n_2=4$ حجم الطبقة تربية المواشي $N_2=66$

الطبقة	القوة النهائية	نوع المزرعة	الدخل
1	5.67	1	75000
1	5.67	2	15000
1	5.67	1	125000
1	5.67	1	67000

80000	2	5.67	1
40000	1	5.67	1
30000	1	16.5	2
14000	1	16.5	2
48000	2	16.5	2
22000	1	16.5	2

فعدد المزارع الفلاحية في العينة هو:

$$\widetilde{N} = \sum_{i \in S_r} w_i = 5.67 + 5.67 + \cdots + 16.5 + 16.5 + 16.5 + 16.5 = 1000$$

ملاحظة : إذا ما تم حذف (إهمال) القوة فالتقدير يكون 10.

أما عدد المزارع المتخصصة في تربية المواشي فهو:

$$\widetilde{N_a} = \sum_{i \in S_r \cap a} w_i = 5.67 + 5.67 + 16.5 = 28$$

نستنتج نسبة مزارع تربية المواشي :

$$\hat{p} = \frac{\sum_{i \in S_r \cap C} w_i}{\sum_{i \in S_r} w_i} = \frac{28}{100} = 0.28$$

الدخل الكلي للمجتمع كليا هو :

$$\hat{y} = \sum_{i \in S_r} w_i y_i$$

$$= 5.67 * 75000 + 5.67 * 15000 + \cdots + 16.5 * 40000 + \cdots + 16.5 * 22000 = 4160340$$

الدخل المتوسط لكل المجتمع

$$\hat{\bar{y}} = \frac{\sum_{i \in S_r} w_i y_i}{\sum_{i \in S_r} w_i} = \frac{4160340}{100.02} = 41595$$

تقدير الدخل الكلي للمزارع المتخصصة في تربية المواشي :

$$\widehat{Y_a} = \sum_{i \in S_r \cap a} w_i y_i = 5.67 * 84000 + 5.67 * 48000 + 16.5 * 23000 = 1330650$$

تقدير الدخل المتوسط للمزارع المتخصصة في تربية المواشي :

$$\widehat{\overline{Y_a}} = \frac{\sum_{i \in S_r \cap a} w_i y_i}{\sum_{i \in S_r \cap a} w_i} = \frac{5.67*84000+5.67*48000+16.5*23000}{5.67+5.67+16.5} = \frac{1330650}{27.8} = 47796$$

ملاحظة : إذا ما تم إهمال القوى في العينات فإن التقدير يكون غير دقيق. الجدول التالي يبين المقارنة:

المعامل المقدر	التقدير الدقيق مع القوى	التقدير الغير دقيق بدون القوى
N	100	10
N_a	28	3
p	0.28	0.30
Y	4160340 $	516000 $
$\bar{Y}$	41595 $	51600 $
$\hat{Y}_a$	330650 $	155000 $
$\bar{Y}_a$	47796 $	51667 $

7-3- تقدير أخطاء العينات للتقديرات الخاصة بالدراسة

يمكن للأخطاء أن تكون أثناء القيام بالتقديرات في دراسة ما و لقد ذكرنا سابقا ان هناك نوعين من الاخطاء: الاخطاء المتعلقة بالعينة و الاخطاء الغير متعلقة بالعينة. فيما يخص الاخطاء الغير متعلقة بالعينة فهي عادة بمشكلة التحيز و هي صعبة التقدير. أما الاخطاء المتعلقة بالعينة فتعطي تغير و تباين. فهي تقدير الى أي نقطة يكون تقدير لمختلف العينات الممكنة لنفس الحجم و لنفس نمط العينات و بالاستعانة بنفس المقدر تعطي نتائج مختلفة من عينة لأخرى.

إن الهدف من هذه النقطة هي شرح كيفية تقدير تباين العينات و اهمية أخذها بعين الاعتبار نمط العينة المستعمل. نقدم هنا فقط المقدرات التباين للمتوسط و للمجموع المقدر لعينة بسيطة أو منتظمة و لنفترض أنه لا يوجد تعديل للقوة.

7-3-1- تباين العينة

من الناحية الرياضية فإن تباين العينة لتقدير ما هو تغير المربع لمتوسط مقارنة بقيم متوسط للمقدر لكل العينات الممكنة.

هناك عدة عوامل تؤثر على أهمية التباين منها :

- التباين في الخصائص المجتمع
- حجم المجتمع
- نسبة الاستجابة
- نمط العينة و طريقة التقدير

نوضح الان تأثير كل عامل على تباين العينة بالاستعانة بالمقدر التباين العينات لمتوسط مقدر للمجتمع انطلاقا من عينة عشوائية بسيطة مع نسبة استجابة 100 بالمائة.

قيمة كل متغير yi تتغير عادة من وحدة لأخرى في المجتمع. فتباين المجتمع σ^2 لكل الوحدات y_i في المجتمع محددة كما يلي :

$$\sigma^2 = \frac{(N-1)}{N} S^2$$

حيث أن

$$S^2 = \sum_{i \in U} \frac{(y_i - \bar{Y})^2}{N-1}$$

حيث أن U مجموع الوحدات المجتمع و هناك N وحدة في المجتمع. فالمتوسط الغير متحيز مع نننسبة استجابة 100% هي

$$\hat{\bar{Y}} = \sum_{i \in S_r} \frac{y_i}{n}$$

حيث S_r هي مجموع المجبين في العينة و هناك n وحدات في العينة. ان تقدير $\hat{\bar{Y}}$ يتغير من عينة لأخرى، فتباين العينة ل $\hat{\bar{Y}}$ للعينة البسيطة من حجم n يمكن كتابته على الشكل التالي

$$var\left(\hat{\bar{Y}}\right) = (1 - \frac{n}{N}) \frac{S^2}{n}$$

فعادة S^2غير معروف و يجب تقديره.

كما أن حجم المجتمع له تأثيرات على تباين العينة : فالمعادلة $f = \frac{n}{N}$ تسمى *fraction de sondage* و المعادلة $(1 - f) = \frac{1-n}{N}$ هو معامل التصحيح للعينة.

تباين العينة يتناقص في حالة كون العينة n يتزايد و في حالة الاحصاء الشامل حيث ($n=N$) فلا يوجد تباين في العينة.

اما إذا كانت نسبة سبر الاراء صغيرة (بمعنى حجم العينة صغير مقارنة بالمجتمع) يمكن ان نهمل معامل التصحيح. هذا المعامل يمكن إهماله إذا لم يتعدى 5 بالمائة و في كثير من الحالات يهمل حتى مع نسبة 10 بالمائة.

كما أن عدم الاجابة يزيد من تباين المجتمع و هذا يتناقص حجم العينة.

1-1-3-7- حساب التباين الحقيقي للعينة

المثال التالي يوضح كيفية حساب التباين الحقيقي للعينة الخاص بالمصاريف المتوسطة المقدرة لسلعة ما، حجم $n=2$

مثال 7-7 مصاريف سلعة ما لمجتمع مكون من أربع أسر مسجلة في جدول مع استعمال العينة العشوائية بدون إرجاع حجم العينة =2. ما هو التباين الحقيقي للمصاريف المتوسطة المقدرة للعينة.

الاسرة	مصاريف السلعة
1	10
2	20
3	30
4	40

متوسط مصاريف المجتمع

$$\bar{Y} = \sum_{i \in U} \frac{y_i}{N} = \frac{10 + 20 + 30 + 40}{4} = 25$$

فهذا المقدر المعتاد للمتوسط المقدر للعينة البسيطة

$$\hat{\bar{Y}} = \sum_{i \in S_r} \frac{y_i}{n} = \sum_{i \in S_r} \frac{y_i}{2}$$

يمكن حساب التباين الحقيقي للعينة للمتوسط المقدر $var\left(\hat{\bar{Y}}\right)$ لعينة بسيطة عشوائية من حجم n=2

العينة	وحدة العينة	تقدير العينة $	$(\hat{\bar{Y}} - \bar{Y})$	$\left(\hat{\bar{Y}} - \bar{Y}\right)^2$
1	(1,2)	15	-10	100
2	(1,3)	20	-5	25

3	(2,3)	25	0	0
4	(1 ;4)	25	0	0
5	(2,4)	30	5	25
6	(3,4)	35	10	100
	المتوسط	25	0	41,7

1- حساب المتوسط لكل المتوسطات الممكنة للعينة :

$$\bar{\hat{Y}} = \frac{\hat{Y}_1+\hat{Y}_2+\hat{Y}_3+\hat{Y}_4+\hat{Y}_5+\hat{Y}_6}{6} = \frac{10+20+25+25+30+35}{6} = 25 = \bar{Y}$$

نلاحظ أن قيمة المتوسط لكل العينات الممكنة يساوي متوسط المجتمع $\bar{Y}$ و هذا لأن المقدر ل $\hat{\bar{Y}}$ غير متحيز.

2- ثم حساب الفرق بين كل تقدير للعينة و التقدير المتوسط لكل العينات (أي $\hat{\bar{Y}}_j - \bar{Y}$ لكل وحدة من العينة) مسجلة في العمود الرابع في الجدول أعلاه.

3- حساب مربع هذه المتوسطات أي $\left(\hat{\bar{Y}}_j - \bar{Y}\right)^2$ المسجل في العمود الخامس من الجدول.

4- كل المجموع J الكل العينات في المجتمع، حساب متوسط مربع الفروقات :

$$var\left(\hat{\bar{Y}}\right) = \sum_{j\in J} \frac{\left(\hat{\bar{Y}} - \bar{Y}\right)^2}{6} = \frac{1000 + 25 + 0 + 0 + 25 + 100}{6} = 41.7$$

تباين العينة الحقيقي للمصاريف المتوسطة المقدرة لسلعة ما في عينة عشوائية بدون ارجاع من حجم n=2 لهذا المجتمع هي إذن $var\left(\hat{\bar{Y}}\right) = 41.7$

كما يمكن استعمال المعادلة السابقة مباشرة.

$$var\left(\hat{\bar{Y}}\right) = (1 - \frac{n}{N})\frac{S^2}{n}$$

$$S^2 = \sum_{i\in U} \frac{(y_i - \bar{Y})^2}{N-1}$$

$$= \frac{(10 - 25^2) + (20 - 25^2) + (30 - 25^2) + (40 - 25^2)}{(4-1)}$$

$$= 166.7$$

$$var\left(\hat{\bar{Y}}\right) = \left(1 - \frac{n}{N}\right)\frac{S^2}{n} = \left(1 - \frac{2}{4}\right)\frac{166.7}{2} = 41.7$$

في هذه المعادلة يجب تقدير التغير في المجتمع S^2 من خلال عينة واحدة.

إن المقدر الغير متحيز للتباين العينة للمتوسط المقدر $\hat{\bar{Y}}$ للعينة العشوائية بدون إرجاع هي :

$$\hat{var}\left(\hat{\bar{Y}}\right) = (1 - \frac{n}{N})\frac{\hat{S}^2}{n}$$

$$\hat{S}^2 = \sum_{i \in S} \frac{(y_i - \bar{y})^2}{n-1}$$

$$\bar{y} = \sum_{i \in S} \frac{y_i}{n}$$

هذه المعادلة تستعمل في 3-2-3-7.

المقدر الغير متحيز لتباين العينة الاجمالية المقدرة $\hat{Y}$ باستعمال العينة البسيطة العشوائية بدون إرجاع هي :

$$\hat{var}\left(\hat{Y}\right) = \hat{var}\left(N * \widehat{\bar{Y}}\right) = N^2\, \hat{var}\left(\hat{\bar{Y}}\right) = N^2(1 - \frac{n}{N})\frac{\hat{S}^2}{n}$$

7-3-2- مقادير أخرى لأخطاء العينات

سنرى هنا الانحراف المعياري، هامش الخطأ، مجال الثقة.

7-3-2-1 الانحراف المعياري، و معامل التغير

الانحراف المعياري هو الجذر التربيعي للتباين العينة اما معامل التغير فهو يمثل العلاقة بين الانحراف المعياري و القيم المتوسطة لكل العينات الممكنة و هي تعطى عادة بالنسبة المئوية (10 بالمائة عوض 0.1).

مثال : لدينا *n=10* أشخاص (نسبة الاستجابة 100 بالمائة). مختارة من *N=500* شخص سن كل وحدة من العينة مسجل في الجدول أسفله (مختارة حسب السن)

ما هو تباين العينة المقدر للسن المتوسط؟
ما هو الانحراف المعياري؟ و معامل التغير المقدر؟

ما هو هامش الخطأ و مجال الثقة عند مستوى 95 بالمائة؟

2-2-3-7- هامش الخطأ و مجال الثقة

في كثير من الدراسات نقرأ أن نتائج الدراسة دقيقة بأكثر أو أقل من 5 بالمائة من هامش الخطأ. ما ذا نعني بهذا ؟

الشخص	سن الوحدة	$(y_i - \bar{y})$	$(y_i - \bar{y})^2$
1	21	-13.4	179.56
2	26	-8.4	70.56
3	27	-7.4	54.76
4	32	-2.4	5.76
5	34	-0.4	0.16
6	37	2.6	6.76
7	38	3.6	12.96
8	40	5.6	31.36
9	42	7.6	57.76
10	47	12.6	158.76

1- حساب السن المتوسط للمجتمع

$$\hat{\bar{Y}} = \sum_{i \in S_r} \frac{y_i}{n} = \frac{21 + 26 + 27 + 32 + 34 + 37 + 38 + 40 + 42 + 47}{10}$$

$$= 34.4$$

السن المتوسط هو 34.4 سنة

2- حساب التباين للعينة $\hat{\bar{Y}}$

$$\hat{var}\left(\hat{\bar{Y}}\right) = \left(1 - \frac{n}{N}\right)\frac{\hat{S}^2}{n}$$

$$\hat{S}^2 = \sum_{i \in U} \frac{(y_i - \bar{y})^2}{n - 1}$$

$$= \frac{179.56 + 70.56 + 54.67 + \cdots + 31.36 + 57.76 + 158.76}{10 - 1}$$

$$= 64.3$$

إذن

$$\hat{var}\left(\hat{\bar{Y}}\right) = \left(1 - \frac{n}{N}\right)\frac{\hat{S}^2}{n} = \left(1 - \frac{10}{500}\right)\frac{64.3}{10} = 6.3$$

تباين المقدر العينة هو 6.3

3- حساب الانحراف المعياري

$$\widehat{\mathrm{SE}}\left(\hat{\bar{Y}}\right) = \sqrt{\hat{var}\left(\hat{\bar{Y}}\right)} = \sqrt{6.3} = 2.5$$

$$\widehat{\mathrm{CV}}\left(\hat{\bar{Y}}\right) = \frac{\widehat{\mathrm{SE}}\left(\hat{\bar{Y}}\right)}{\bar{y}} = \frac{2.5}{34.4} = 0.073 = 7.3\%$$

4- حساب هامش الخطأ و مجال الثقة (95 بالمائة)

$$\text{هامش الخطأ} = 1.96 * \widehat{\mathrm{SE}}\left(\hat{\bar{Y}}\right) = 1.96 * 2.5 = 4.9$$

$$IC = \hat{\bar{Y}} \pm 1.96 * \widehat{\mathrm{SE}}\left(\hat{\bar{Y}}\right) = 34.4 \pm 4.9 = (29.5, 39.3) \text{ مجال الثقة}$$

إذن السن المتوسط الحقيقي للمجتمع حسب مجال الثقة موجود بين سن 29.5 و 39.3.

3-2-3-7- تقدير تباين العينات في عينة طبقية

سنرى الان الاثار إذا لم نأخذ بعين الاعتبار نمط العينة المستعمل و ان تباين تم حسابه فقط بالاستعانة بالمعادلة الخاصة بنمط العينة البسيطة .

مثال6-7 (تكملة) تقدير تباين العينة

إذا كانت العينة هي نتيجة عينة عشوائية بسيطة من حجم n=10 (نسبة الاستجابة 100بالمائة) مأخوذة من حجم مجتمع N=100 إذن يمكن الحصول على تقديرات للدخل المتوسط و التبيان المقدر لدخل المتوسط للعينة.

$$\hat{\bar{Y}} = \sum_{i \in S_r} \frac{y_i}{n} = 51600$$

مقارنة ب 41595 حيث تم أخذ بعين الاعتبار نمط العينة.

التباين المقدر للعينة (بالالاف)

$$\hat{var}_{EAS}\left(\hat{\bar{Y}}\right) = \left(1 - \frac{n}{N}\right)\frac{\hat{S}^2}{n} = \left(1 - \frac{10}{100}\right)\frac{1247}{10} = 112.2$$

أما الانحراف المعياري (بالالاف)

$$\sqrt{\hat{var}\left(\hat{\bar{Y}}\right)} = \widehat{\mathrm{SE}}_a\left(\hat{\bar{Y}}\right) = 10.6$$

و من أجل تقدير صحيح لتباين العينة للمتوسط انطلاقا من عينة طبقية يجب تحديد تباين المقدر للعينة لكل طبقة (h) و جعل مجموع النتائج لكل طبقة من أجل الحصول على تقديرات كاملة (بالالاف)

$$\hat{V}_{STR}\left(\hat{\bar{Y}}\right) = \frac{1}{N^2}\sum_h N_n^2\left(1 - \frac{n_h}{N_h}\right)\frac{\hat{S}_h^2}{n_h}$$

$$= \frac{1}{100^2}\left\{34^2\left(1 - \frac{6}{34}\right)\frac{1406}{6} + 66^2\left(1 - \frac{4}{66}\right)\frac{211.6}{4}\right\} = 44$$

و الانحراف المعياري

$$\sqrt{\widehat{var}_{STR}\left(\hat{\bar{Y}}\right)} = \widehat{\mathrm{SE}}_{STR}\left(\hat{\bar{Y}}\right) = 6.6$$

فإذا تم تقدير تباين للعينة بدون الاخذ بعين الاعتبار نمط العينة و ان المقدر للعينة البسيطة استعمل فإن تباين العينة يكون مرتفع، و بالمقابل إذا كان في حالة العينة الطبقية فإن استعمال معادلة العينة البسيطة فإن التباين يكون منخفض .

3-3-7- أثار نمط العينة

يقارن اثار النمط تباين المقدرات بين نمط العينات و العينة العشوائية البسيطة، فهو عبارة عن العلاقة بين تباين العينات لمقدر ما حسب نمط معين و تباين العينات للعينة البسيطة من نفس الحجم.

في المثال السابق

$$deff = \frac{\widehat{var}_{STR}\left(\hat{\bar{Y}}\right)}{\widehat{var}_{EAS}\left(\hat{\bar{Y}}\right)} = \frac{44000}{112200} = 0.39$$

و هذا معناه أن الطبقية تحسن من دقة المتوسط المقدر في الدراسة. كما يستعمل في تحديد حجم العينة الممكنة للدراسة.

7-3-4- تقدير تباين العينات بالاستعانة بالعينة المعادة

هناك طريقة التبديل لتقدير التباين في حالات العينات المعقدة منها العينة المعادة. ففي هذه الحالة فبدل تحديد عينة من حجم n، k عينات مستقلة من حجم n/k مختارة.

يتم تقدير خصية الدراسة لكل العينات k.

فالتغير في التقديرات للعينات k يستعمل فيما بعد لتقدير التباين العينات. إن تقدير t لخصية الدراسة نتحصل عليها بالاستعانة بمتوسط التقديرات المنجزة لكل عينة :

$$t = \sum_{j \in k} \frac{t_j}{k}$$

حيث k هو مجموع العينات k هو عدد العينات و t_j هي تقدير ل j^{eme} عينة.

التباين المقدر للعينة t ، $\widehat{\text{var}}(t)$ هو نتيجة المعادلة

$$\widehat{\text{var}}(t) = \sum_{j \in K} \frac{(t_j - t)^2}{k(k-1)}$$

مثال 9-7- تقدير تباين العينة ل $\hat{\bar{Y}}$ بالاستعانة بالعينة المعادة *EAS*

في المثال -8-7 عوض تجديد عينة من حجم *n=10* و تطبيق المعادلة السابقة من أجل تقدير $var(t) = var\left(\hat{\bar{Y}}\right)$ نستعمل عينتان من حجم *n=5* مختارة النتنائج في الجدول.

العينة 1	السن للوحدة	العينة 2	السن للوحدة
1	21	1	26
2	27	2	32
3	34	3	37
4	38	4	40

5	42	5	47
متوسط السن	132.4	متوسط السن	36.4

1- حساب متوسط المجتمع

$$\hat{\bar{Y}} = \sum_{j \in K} \frac{\hat{\bar{Y}}_j}{k} = \frac{32.4 + 36.4}{2} = 34.4$$

2- تباين للعينة المقدر للسن المتوسط بالاستعانة بطريقة العينة المعادة

$$\widehat{\text{var}}\left(\hat{\bar{Y}}\right) = \sum_{j \in K} \frac{\left(\hat{\bar{Y}}_j - \hat{\bar{Y}}\right)^2}{k(k-1)} = \frac{(32.4 - 34.4)^2 + (36.4 - 34.4)^2}{2} = 4$$

3- الانحراف المعياري المقدر هو :

$$\widehat{\text{SE}}\left(\hat{\bar{Y}}\right) = 2$$

خلاصة

رأينا في هذا الفصل مجموعة من النقاط المتعلقة بالتقديرات و نلخصها كمايلي.

يجب أ ن يراعي التقدير نمط العينات المستعمل في الدراسة إذ يجب إدراج القوى في مسار التقدير.
يجب تعديا القوى في حالة وجود عدم الاجابة في الدراسة.
يجب استعمال البيانات الاضافية من أجل تحسين دقة التقديرات.
يجب استعمال أنماط العينات و توزيعها من أجل الاجابة لمتطلبات مجال البحث، و في حالة استحالتها في مرحلة التحضير في نمط العينة فيجب استعمال طرق خاصة للتقدير في مرحلة التقديرات.
يجب معالجة القيم الشاذة التي يمكن ان تعطي تباين في العينات أثناء التقدير.
أثناء التقدير، يجب إضافة نسبة الخطأ المتعلقة بالعينة، تباين العينة، معامل التغير، هامش الخطأ، مجال الثقة،....

المحور الثامن : تحديد حجم العينة

8-0- مقدمة

ان من بين الاسئلة التي يطرحها الاحصائي هي ما هو حجم العينة المناسب للدراسة؟ فالمسير يهمه الحصول على الجواب في مرحلة التخطيط للدراسة لأنها تؤثر في الامور العملية مثل عدد المستجوبين اذ لا توجد حل سحري لهذا السؤال نظرا لشروط الدقة في التقدير و كذا الامور المالية المتوفرة. للتذكير فأن العوامل التي تؤثر على الدقة في التقديرات نجد: التغير في المجتمع، حجم المجتمع، نمط العينة، نسبة الاستجابة. فالصيغة لحساب حجم العينة الضرورية يجب أن تأخذ بعين الاعتبار هذه العوامل. إضافة الى هذه العوامل هناك بعض القيود العملية التي تؤثر في حجم العينة.

و في حالة استعمال عينة طبقية فيجب معفرة حجم العينة الكلية و كيف توزع على مختلف الطبقات و هذا ما يسمى ب توزيع العينة. هناك استراتيجيتين ممكنتين : الاولى تتعلق بتحديد حجم العينة الكلي ثم يعاد توزيعها على الطبقات. الثانية تتطلب تثبيت الدقة المطلوبة ثم تحديد حجم العينة الضرورية لكل طبقة.

هناك عدة طرق ممكنة للتوزيع. الهدف من هذا الفصل هو معرفة كيفية حساب حجم العينة مع مراعاة بعض شروط الدقة و كذا كيفية توزيعها على الطبقات.

8-1- تحديد حجم العينة

مثلما أشارنا في الفصل السابع أن دقة التقديرات في دراسة استقصائية و حجم العينة مرتبطين فيما بينهما. بما أن تباين العينة يتناقص مع زيادة حجم العينة، فكلما كانت التقدير تتطلب الدقة كلما كانت حجم العينة مهم و منه فإن حجم العينة مرتبط بالدقة المطلوبة في التقديرات. هذه الدقة لتقديرات (t) يمكن أن يعبر عنها الانحراف المعياري ((SE(t)، هامش الخطأ zxSe(t) ،و معامل التغير SE(t)/t . إن تحديد حجم العينة لدراسة ما تأخذ عادة احدى هذه القيم أو أكثر من ذلك.

للتذكير فإن تحديد حجم العينة نحاول مراقبة أخطاء المتعلقة بالعينة و عدم الاجابة و التي تنتج فجأة فهي لا تحاول مراقبة الاخطاء خارج العينة. و من أجل ضمان نتائج الدراسة يجب تحديد ما هي المشاكل المسببة لهذه الاخطاء غير المتعلقة بالعينة و التي يجب ان تنقص الى أكبر قدر ممكن.

يجب التذكير ببعض الامور المتعلقة بالدقة/الجودة قبل البدأ في عملية حساب حجم العينة.

8-1-1- بعض الشروط المتعلقة بالدقة

ان الهيئة الاحصائية يجب أن تاخذ بعين الاعتبار الاسئلة المهمة قبل اتخاذ القرار حول المستوى الملائم للدقة التقديرية. يجب معرفة ما هو مطلوب من الدراسة. تقديرات متعلقة بالبيانات، تحليل و ما هي القرارات التي يمكن أن تتخذ.

(أ) كيف نستعمل تقديرات الدراسة؟ ما هو التباين المقبول في الدراسة،؟ ما هي درجة الشك (الخطأ) المقبولة مثل ±6% مع 95 مجال الثقة.

(ب) ما هي التقديرات المطلوبة لمجموعات الثانوية لمجتمع الدراسة؟ ان نتائج الدراسة يمكن أن تحتوي على تقديرات عديدة لمجموعات الثانوية/مجالات مختلفة مثال اضافة الى نتائج الدراسة يمكن أن تكون مطلوبة أو تقديرات خاصة بالولايات أو البلديات يمكن ان تكون مطلوبة أو تقديرات حسب السن، الجنس، المستوى الدراسي لمجتمعات ثانوية.

(ت) كما يجب تحديد درجة الدقة/أي هامش الخطأ المقبول. بالنسبة لدراسة شاملة وطنية فيمكن أن تكون ±3% و ±5% بالنسبة للدراسات الولائية (المناطق) و ±10% للبلديات.

(ث) ما هو حجم تباين العينة بالنسبة لتقديرات الدراسة؟ بعد دراسة حجم العينة لابد من تحديد شروط الدقة. مقال من أجل سياسة جديدة الوزارة مطالبة بتوفير خدمة لفئة قلة حسب اللغة . إذا كان P=0.05 أو (0.5بالمائة) من الطلبات تكون ضمن هذه الفئة.

لنفترض أن الوزارة قررت إجراء دراسة على مستوى الزبائن حتى تقوم بتقدير الطلب على الخدمة بلغة معينة للأقلية. فمن الوهلة الاولى يتضح أن هامش الخطأ ±0.05 يتضح انه كبير مقارنة بالتقديرات الدراسة التي تكون في مجال 0.05. في هذه الحالة هامش الخطأ أكبر، لا يكون اكبر من ±0.01 الى ±0.02 بحيث ان يحدد (مثال مجال الثقة يجب أن يكون ±0.01 ±0.05 و ±0.02 ±0.05.

إذن الزبون يجب أن يأحذ بعين الاعتبار حجم التقديرات المطلوبة الضعيفة عند تحديد شروط او متطلبات الدقة.

لنفترض أن الدراسة هي تقدير لنسب، بعض هذه النسب يمكن ان تكون P=0.50 أو أكبر و لكن البعض الاخر يمكن لأن تكون أقل P=0.05 أو P=0.10.

إذا كانت النسبة الضعيفة للتقدير تكون P=0.05 و هذه النسبة مهمة من أجل أهداف الدراسة، فالهيئة الاحصائية تريد هامش الخطأ 0.05.

(ح) ما هي النتائج الميدانية من جراء تلك المتطلبات الخاصة بالدقة؟ كم نربح من الدقة إذا زاد حجم العينة؟ أكيد ان الدقة تتحسن إذا زاد حجم العينة، إلا أن الربح في الدقة لا يزيد بنفس النسبة مع تزايد حجم العينة.

لنفترض متغير كيفي يتكون من فئتين ممكنتين A و B (مثال الرجال و النساء) و يهتم الزبون بتقدير نسبة المجتمع في الفئة الاولى.

يبين الجدول الاتي هامش الخطأ في النسبة المقدرة الناتجة عن عينة عشوائية بسيطة لمختلف أحجام العينة و في حدود مجال الثقة 95 بالمائة أو نسبة حقيقية لتكون المجتمع في الفئة الاولى هو P=0.5 و حجم المجتمع N=100000 .

حجم العينة	هامش الخطأ
50	0.139
100	± 0.098
500	± 0.044
1000	± 0.031

نلاحظ أنه عندما تكون يتضاعف حجم العينة من 50 الى 100 فإن هامش الخطأ لتقدير النسب ينتقل من 0.139 الى 0.098 ±، فهامش الخطأ لم يتنافص بالنصف ليكون 0.070 ± كما كن ننتظر. إذن لا توجد علاقة خطية بين حجم العينة و هامش الخطأ.

هذا يجعل الهيئة تقرر هل من الفائدة استجواب 1000 شخص بدل 500 من اجل زيادة الدقة بهامش خطأ يقدر ب 0.032 ± عوض 0.045 ± .

إذن في بعض الاحيان اختيار حجم العينة كبير لا يعطي هامش خطأ صغير.

8-1-2- العوامل المؤثرة في الدقة

سنبين في هذه النقطة أثار بعض العوامل عند تحديد حجم العينة على الدقة و الجودة

8-1-2-1- التغير على مستوى المجتمع.

ان الخصية المدروسة أو المتغير الفائدة يختلف من شخص لأخر، من أسرة لأخرى، من مؤسسة ..... في مجتمع الدراسة الميدانية. و بما أن هذا المتغير لا يمكن أن يتحكم فيها و

مراقبتها فإنها يمكن أن تؤثر على حجم العينة الضرورية للحصول على درجة من الدقة لهذا المتغير.

لنأخذ المثال التالي و نفترض ان دراسة تريد للمرة الاولى تقدير نسبة الزبائن الذين يرضون بخدمات بعض المؤسسات و ان هناك قيمتين فقط متوفرتين من أجل رضا الزبائن (راضي/غير راضي).

	راضي	غير راضي
1	% 100	%0
2	% 90	% 10
3	% 80	% 20
4	% 70	% 30
5	% 60	% 40
6	% 50	% 50
7	% 40	% 60
8	% 30	% 70
9	% 20	% 80
10	% 10	% 90
11	% 0	% 100

القيم التي تمثل الحالات الممكنة للمتغير حول رضا الزبائن في المجتمع من 1 الى 11. فالأعداد 2 و 10 من القائمة تمثل تباين صغير (تغير) مع 90 بالمائة من الزبائن لهم رأي ايجابي و 10 بالمائة رأي سلبي (أي غير راضي) و العكس صحيح. كما ان الاعداد 3 و 9 ، 4 و 8 ، 5 و 7 لهم نفس التباين. أما الاعداد من 1 الى 6 أو من 11 الى 6 فتباين خصية الرضا تزيد. أما العدد 6 يمثل التوزيع 50-50 حيث 50 بالمائة من الزبائن راضين و 50 بالمائة غير راضين.

كما لو أن كل الزبائن كانوا راضين عن الخدمة المقدمة فإنه لا يكون هناك تباين إذن شخص واحد يمكن أن يعطي لنا تقديرات صحيحة حول الرضا.

و بما ان التباين الحقيقي في المتغير المدروس يزيد في المجتمع الدراسة، كذلك بالنسبة لحجم العينة حتى تعطي تقديرات ذات جودة جيدة.

2-2-1-8- حجم المجتمع

يلعب حجم المجتمع دورا كبيرا في تحديد حجم العينة مثال استعمال دراسة حول رضا الزبائن، النسبة الحقيقية للزبائن الراضين هي P=0.5. نفترض أن الهيئة الاحصائية تستعمل العينة العشوائية البسيطة و أن تقدير P بل هامش خطأ 0.05 ± مع مجال ثقة 95 بالمائة.

الجدول و يبين حجم العينة المناسب لمختلف أحجام المجتمع.

حجم العينة	حجم العينة المناسب
50	44
100	80
500	222
1000	286
5000	370
10000	385
100000	398
1000000	400
10000000	400

نلاحظ أن من أجل تحقيق درجة الجودة المطلوبة، حجم العينة يزيد الى نسبة التي تناقص عندما حجم المجتمع يرتفع. إلا أن الهيئة تريد 44 استبيان مملوء لعينة من المجتمع N=50 ، فإنه ليس من الحاجة مضاعفة حجم العينة الى 88 (44*2) مع مضاعفة حجم مجتمع الدراسة (50*2).

كما أن حجم العينة يقترب بسرعة عند عينة n=400 مع مجتمع الدراسة N=5000 و أكثر. و منه استعمال عينة عشوائية بسيطة لملئ 400 استبيان لا تعطي شروط الجودة المطلوبة لمجتمعات أكبر من 5000 عندما تكون النسبة الحقيقية للمجتمع هي .P=0

في المجتمعات الصغيرة عادة نسبة كبيرة مطلوبة حتى نحصل على جودة جيدة.

8-1-2-3- مثال التصميم و التقدير

إن استراتيجية العينات أي نمط تصميم العينة و المُقدَر المستعمل لهم تأثيرات حول الدقة. إن تقنيات حساب حجم العينة لدرجة معينة من الدقة تطبق عادة صيغة تباين العينات لعينة بسيطة عشوائية.

عموما إذا كانت صيغة حساب حجم العينة نفترض أن العينة البسيطة العشوائية لكن إذا كان نمط التصميم لعينة جد معقد و قد استُعمل فإن حجم العينة يتطلب من أجل الحصول على درجة من الدقة استعمال جداءها في معمل يسمى أثار نمط التصميم و يرمز له ب deff و هو عبارة عن العلاقة بين تباين العينة لمقدر حسب نمط التصميم للعينة المستعمل و تباين العينة لمقدر في حالة عينة عشوائية بسيطة لها نفس الحجم. في تصميم العينات البسيطة فإن deff=111 ، في العينات الطبقية deff<=1، و في العينات مراحل يكون أكبر من 1.

8-1-2-4- نسبة الاستجابة-

من أجل الوصول الى الدقة المطلوبة للتقديرات، يجب على الهيئة الاحصائية تعديل حجم العينة مع نسبة الاستجابة السابقة. و هذا باختيار عينة أكبر انطلاقا من نسبة الاستجابة المرجوة اعتمادا على دراسات مماثلة.

مثال : إذا كان الحجم الابتدائي للعينة محسوب ب 400 و ننتظر نسبة استجابة 75 بالمائة، الهيئة الاحصائية يجب ان تحدد العينة

$$n = \frac{400}{0.75} = 533$$

نسبة استجابة ضغيفة تؤدي الى حجم عينة صغير و هذا قد يؤدي الى نقص في الدقة المطلوبة و العكس صحيح. كما ان نسبة استجابة كبيرة تؤدي الى عينة كبيرة و قد يزيد من عدم الاجابة.

8-1-3- أمثلة عن صيغ الحجم

الصيغ التالية يمكن ان تستعمل لحساب حجم العينة المطلوبة للحصول على الدقة.

(أ) الدقة لمتوسط مقدرY لعينة عشوائية (100 بالمائة اجابة) و هامش خطأ،....تستعمل لتحديد العينة.

و من أجل تحديد n بعض العوامل ضرورية :

هامش الخطأ المقبول e
مجال الثقة المطلوب z
حجم المجتمع N

تقدير التباين المجتمع S^2 (هذا هو الاصعب و يمكن الاعتماد على الدراسات السابقة).

(ب) الدقة لنسبة المقدر P لعينة عشوائية بسيطة (مع 100 بالمائة استجابة). في هذه الدقة المطلوبة تكون محددة حسب هامش الخطأ و خصية الدراسة تكون نسبة مع المجتمع P و تكون في إحدى النوعين.

(ت) و من المعروف أنه في العينات الكيرة فان نسبة P المقدرة هي عادة توزيع طبيعي و التباين خصية المجتمع التالي yi يقدر حسب :

$$\hat{S}^2 = \hat{P}(1-\hat{P})$$

و منه

$$n = \frac{Z^2\hat{P}(1-\hat{P})}{e^2 + \frac{Z^2\hat{P}(1-\hat{P})}{N}}$$

إذا كانت قبل الدراسة قيمة P متوفرة، يجب أن يستعمل في المعادلة أعلاه، و إلا فإننا تستعمل P=0.5 .

(ث) تحديد حجم العينة خطوة بخطوة:

1. حجم العينة الاولية)الابتدائية) :

تجدر الاشارة إلى أن المعادلة السابقة تستعمل معامل التصحيح للمجتمع (1-n/N) من اجل تصحيح حجم المجتمع. و في حالة إهمال هذا المعامل فإن التقدير حجم العينة الاولية n1 يمكن أن نتحصل علية فقط ب :

$$n_1 = \frac{Z^2\hat{P}(1-\hat{P})}{e^2}$$

2. التعديل لحجم العينة باستعمال الصيغة التالية:

$$n_2 = n_1 \frac{N}{N + n_1}$$

3. إذا كان تحضير العينة ليس عينة عشوائية بسيطة، فإنه يمكن استعمال الصيغة التالية من أجل تعديل العينة.

$$n_3 = deff * n_2$$

حيث أن deff هو أثار نمط العينة و عادة ما تكون :

deff=1 في حالة العينة العشوائية البسيطة.

deff>1 في حالة العينة العنقودية او متعددة المراحل.

deff<1 في حالة العينة الطبقية.

4. في النهاية تعديل الاجوبة من أجل تحديد حجم العينة n

$$n = \frac{n_3}{r}$$

حيث أن r هو نسبة الاستجابة المرجوة.

-1-3-1-8- مثال في تحديد حجم العينة

نستعمل طريقة خطوة بخطوة من اجل تحديد حجم العينة.

مثال 1-8- استعمال نمط العينة العشوائية البسيطة.

تريد إحدى المجلات دراسة تقدير رضا القراء مع اعلم أنه يمكن للمجلة أن تستجوب كل المشتركين البالغ عددهم 25000 مشترك عن طريق البريد، إلا أن ضيق الوقت جعل المجلة تقرر استجواب عينة عشوائية عن طريق الهاتف.

ما هم عدد القراء الذين يجب استجوابهم في هذه الحالة؟

الفرضيات في هذه الحالة :

تكون المجلة راضية إذا كانت النسبة الحقيقية للمجتمع هي 10± من النسبة المقدرة للمجتمع اعتمادا على نتائج العينة، مع العلم أن هامش الخطأ e هو 10 بالمائة.

تريد المجلة مجالا للثقة يقدر ب ,% 95 في التقدير (هذا معناه أن هناك 1 من 20 للحصول على عينة تنتج تقديرات خارج نطاق P±10 و منه Z=1.96.

نستعمل نمط العينة العشوائية البسيطة في الدراسة.

ننتظر نسبة استجابة تقدر ب 65بالمائة و منه r=0.65

لا توجد تقديرات من قبل P و منه P=0.5.

حجم العينة يكون الاتي:

1. حساب حجم العينة الابتدائية

$$n_1 = \frac{Z^2 \hat{P}(1-\hat{P})}{e^2} = \frac{1.96^2(0.5)(1-0.5)}{0.10^2} = 96$$

2. تعديل حجم المجتمع مع الاخذ بعين الاعتبار حجم المجتمع

$$n_2 = n_1 \frac{N}{N+n_1} = 96 \frac{2500}{(2500+96)} = 92$$

3. تعديل حجم العينة حسب نمط اختيار العينة deff=1 لأننا استعملنا العينة العشوائية البسيطة.

$$n_3 = deff * n_2 = n_2 = 92$$

4. تعديل حجم العينة حسب نسبة الاستجابة المرجوة

$$n = \frac{n_3}{r} = \frac{92}{0.65} = 142$$

إذن بعد هذه الخطوات المطبقة يجب أن نختار 142 مشترك من اصل 2500 باستعمال نمط العينة العشوائية البسيطة لتقدير رضا المشتركين في المجلة مع هامش خطأ 0.10 بالمائة و مجال ثقة 95 بالمائة و نسبة استجابة 65 بالمائة.

مثال 2-8- استعمال نمط العينة الطبقية

عملية استطلاع للرأي من اجل تحديد نسبة المجتمع الموافق لإنشاء محمية جديدة. المجتمع يتكون من كل البالغين القاطنين في مدينتين و منطقة ريفية.

عينة عشوائية بسيطة للأشخاص البالغين في كل مدينة و في المنطقة الريفية.

حجم العينة المطلوب في كل طبقة يجب أن يحدد ب .*P*

حجم العينة هو 657500 مع التوزيع التالي :

الرمز	الطبقة	المجتمع *Ni*
1	المدينة 1	*400 000*
2	المدينة 2	*250 000*
3	المنطقة الريفية	*7 500*
المجموع		*657 500*

حجم العينة المطلوب يتحكم فيه شروط الدراسة : هناك حالتين :

الحالة الاولى : هامش الخطأ بالنسبة للمجتمع الكلي.

نفترض هامش خطأ هو 0.05± مع مجال الثقة *95%* لكل منطقة. تقدير أولي للنسبة ضروري، إذن *P=0.5* يفترض نسبة استجابة *50%* منتظرة.

1 . حساب حجم العينة الأولية *n1*

$$n_1 = \frac{Z^2 \hat{P}(1-\hat{P})}{e^2} = \frac{1.96^2(0.5)(1-0.5)}{0.05^2} = 384$$

2. حساب العينة المعدلة *n2*

$$n_2 = n_1 \frac{N}{N+n_1} = 384 \frac{657500}{(657500+384)} = 384$$

إذا تم إهمال (n1/N) فأن n1=n2

3. تعديل حجم العينة حسب نمط العينة *n3*

$$n_3 = deff * n_2 = n_2 = 384$$

بالنسبة للعينة العشوائية الطبقية عادة تكون *deff<1* هنا لا توجد أي تقدير ل *deff* إذن نفترض أن قيمتها تساوي 1 حتى نأخذ أكبر حجم للعينة.

5. التعديل حسب نسبة الاستجابة n

$$n = \frac{n_3}{r} = \frac{384}{0.50} = 768$$

إذن حجم العينة هو 768 ، سنرى فيما بعد كيفية توزيعها على الطبقات.

الحالة الثانية : هامش خطأ لكل طبقة

نفترض أن الزبون يريد نتائج بهامش 0.05± و مجال ثقة 95 بالمائة لكل طبقة. الان حساب وحدات كل عينة لكل طبقة تكون ضرورية عكس الحالة الاولى. (كل طبقة تعالج على حدى).

مع العلم أن المدينة 1 و 2 لهم عدد سكان أكبر إذن نسبة حجم المجتمع ليس له أي أثر على حجم العينة. و منه إذا كانت نفس الفرضيات محققة، فإن حجم العينة في كل طبقة هو 768. أما المنطقة الريفية صغيرة يمكن أن تؤثر على حجم العينة.

المنطقة الريفية :

$$n_1 = \frac{Z^2 \hat{P}(1-\hat{P})}{e^2} = \frac{1.96^2(0.5)(1-0.5)}{0.05^2} = 384$$

$$n_2 = n_1 \frac{N}{N+n_1} = 384 \frac{7500}{(7500+384)} = 366$$

$$n_3 = deff * n_2 = n_2 = 366$$

$$n = \frac{n_3}{r} = \frac{366}{0.50} = 732$$

إذن حجم العينة الكلية هو بالنسبة للمدينة الاولى 768+ المدينة الثانية 768 +الريف 732 = 2268 .

مقارنة الحالة الاولى و الحالة الثانية، فإن حجم العينة الاجمالي هو 2268 في الحالة الثانية و هو ثلاث أضعاف الحالة الاولى 768.

8-2- توزيع العينة في حالة العينة الطبقية

الشيئ المهم لتحديد فعالية العينة الطبقة هو شكل الذي تكون فيه العينة الكلية *n* موزعة على الطبقات. للتذكير فإن الحجم الكلي *N* موزع على الطبقات من *N1* إلى *NL* إذن *N=N1+N2+....NL*.

العينة تكون إذن *n=n1+n2+....nl*

توزيع العينة *n* على الطبقات *L* يمكن أن يكون باستعمال أحد المعيارين:

حجم الكلي للعينة يمكن ان يحدد باستعمال الانماط السابقة ثم توزيع على الطبقات (حجم ثابت للعينة).
أو حجم العينة يطلب في كل طبقة من أجل الرد لمتطلبات الدقة يمكن أن يحدد و تجمع من اجل تحديد الحجم الكلي للعينة (يسمى معامل التباين الثابت إذا كانت متطلبات الدقة تعبر بمعامل التباين).

8-2-1- معايير التوزيع

سنرى هنا الفرق بين حجم ثابت للعينة و معامل التباين لمعلمات التباين.

8-2-1-1- حجم ثابت للعينة

في هذه الحالة حجم العينة ثابت n يعطى للطبقات بشكل خاص ، نسبة العينة المعطاة لكل طبقة h يكتب

$$ah = nh/n$$

حيث كل *ah* محصور بين 0 و 1 و أن مجموع *ah=1*

و منه كل طبقة h حجم عينة ثابت nh يساوي ناتج حجم الكلي للعينة n و نسبة ah للعينة الثانية عن هذه الطبقة.

مثال إذا كانت طبقة لها نسبة ah=1/2 إذن نصف العينة الكلية معطاة لهذه الطبقة.
حسب هذا المعيار(n معروف مسبقا) حجم العينة nh لكل طبقة يمكن حسابه بمجرد معرفة قيمة ah هناك عدة طرق لتحديد ah ايجاد قيم ah التي تقلص تباين خصائص العينة (تحديد ah في 8-2-2-)

2-1-2-8- معامل التباين الثابت

البديل هو تثبيت حجم العينة n هو تحديد حجم العينة المطلوبة في كل طبقة nh مع مستوى للدقة من أجل التقدير الكلي. و هذا يتطلب ايجاد حجم العينة nh (h=1, 2,...l) لكل طبقة بحيث أن معامل التباين للتقدير الكلي لا تتعدى قيم CV consigne.

مثال : نفترض التقدير الكلي $\widehat{Y}$ من خلال عينة عشوائية بسيطة طبقية. دالة معامل التباين الكلي يقدر من خلال عينة طبقية باتباع الكتابة التالية لحجم الكلي n.

$$n = \frac{\sum_{h=1}^{L} N_h^2 S_h^2 / ah}{CV^2 Y^2 + \sum_{h=1}^{L} N_h S_h^2}$$

حيث N_h هي حجم الطبقة .

S_h^2 تباين وحدات المجتمع في كل طبقة

ah نسبة العينة المخصصة لكل طبقة

CV معامل التغير الواجب ل Y و Y يمثل المجموع.

للتذكير أن الصياغة أعلاه نفترض أن nh=n*ah <Nh أي أن حجم العينة في كل طبقة أصغر من حجم مجتمع الطبقة (سنرى أكثر في 3-2-8) تباين المجتمع S_h^2 يمكن أن يقدر ب $\hat{S}_h^2$ حيث :

$$\hat{S}_h^2 = \frac{\sum_{i=1}^{nh} (y_{hi} - \bar{y}_h)^2}{n_h - 1}$$

حيث أن $\bar{y}_h$ يمثل متوسط عينة الطبقة مع العلم ان $\bar{y}_h = \frac{\sum_{i=1}^{h} y_{hi}}{n_h}$

$$n_h = a_h \frac{\sum_{h=1}^{L} N_h^2 S_h^2 / ah}{CV^2 \hat{Y}^2 + \sum_{h=1}^{L} N_h \hat{S}_h^2}$$

مع العلم أنه إذا كان y_{hi} متغير ثنائي، إذن متوسط الطبقة هو نسبة ، بمعنى $\bar{y}_h = \hat{P}_h$

$$\hat{S}_h^2 = \hat{P}_h (1 - \hat{P}_h)$$

بتعويض $n_h = n * a_h$ ، $\hat{S}_h^2$ و $\hat{y}$ في المعادلة السابقة ل n نتحصل على المعادلة التالية ل nh

$$n_h = a_h \frac{\sum_{h=1}^{L} N_h^2 S_h^2 / ah}{CV^2 Y^2 + \sum_{h=1}^{L} N_h S_h^2}$$

و منه بعد تحديد قيمة n_h لكل طبقة نحسب حجم العينة n_h كما ذكرنا سابقا لتحديد هذا الحجم يستلزم توفر شروط الدقة (هنا معامل التباين)، التباين المقدر للمجتمع $\hat{S}_h^2$ و حجم المجتمع N_h.

تعديل بالنسبة لعدم الاجابة يجب ان يدخل في حساب النهائي ل n_h.

للتذكير فإن مقاربة معامل التباين الثابت من أجل توزيع العينات جد معقد من المقاربة العينة الثابت و ان هذه الاخيرة هي التي تستعمل في توزيع العينات.

8-2-2- طرق توزيع العينات

المعادلة 4 و 5 هما الاداة الاساسية لتوزيع العينات الطبقية و أن المعادلة تستعمل عندما نحدد قيم *ah* لكل طبقة. اختيار *ah* لكل طبقة يمكن ان يكون بإحدى الطريقتين:

التوزيع النسبي
التوزيع الغير نسبي

تعتمد هذه الطرق على بعض اقيم : حجم المجتمع في كل طبقة، قيم أخرى لحجم الطبقة، التباين في المجتمع او التكلفة في كل طبقة.

8-2-2-1 التوزيع النسبي N (المتكافئ)

مع التوزيع النسبي أو التوزيع النسبي *N* ، حجم العينة n_h في كل طبقة هو تناسبي مع حجم مجتمع الطبقة N_h. و منه فإن الطبقات الكبيرة تستقبل عينة أكبر و الطبقات الصغيرة تستقبل عينات أقل. ينتج جزء من العينة $f_h = {}^{n_h}/_{N_h}$ لكل طبقة. و هو يساوي $f = {}^{n}/_{N}$ إذن الكتابة التالية نتحصل على :

$$n_h = \frac{N_h}{N} n$$

نتيجة التوزيع النسبي *N* هي إذت. $a_h = \frac{n_h}{n} = \frac{N_h}{N}$

بمعنى اخر معامل التوزيع *ah* لكل طبقة يساوي العلاقة حجم المجتمع في الطبقة لحجم المجتمع الكلي. نستعمل هذه الطريقة عندما لا نملك معلومات على التباين في الطبقات المجتمع. و منه فهي لا تستعمل في حساب أحجام العينات لمعامل التباين الثابت لان هذه المقاربة تستلزم معرفة التباين للطبقات.

يستعمل التوزيع النسبي *N* من اجل استعمال الترجيح التلقائي للعينة (كل وحدات لها نفس احتمال الانتماء) و منه نفس القوة.

العينات الطبقية التي تستعمل التوزيع النسبي *N* تعتبر أكثر فاعلية من العينة العشوائية البسيطة لكل مجتمع إذا كانت متوسط الطبقة $\overline{Y}_h$ يختلف من طبقة لأخرى.

أما في حالة ما إذا كان متوسط الطبقات متقارب، فإن الطبقة باستعمال التوزيع النسبي *N* فإنها لا تقلص التباين لإلا بدرجة صغيرة.

مثال 2-8 تكميلة التوزيع النسبي *N* مع استعمال حجم عينة ثابت.

في المثال السابق الحالة 1 فإن حجم العينة *n=768* كيف نوزع 768 فرد على ثلاث طبقات باستعمال التوزيع النسبي *N* لجدم هينة ثابت.

1- باستعمال التوزيع النسبي N، أحسب قيم معامل التوزيع a_h لطل طبقة.

الطبقة	معامل a	n_h
المدينة 1	$a_1 = \frac{N_1}{N} = \frac{400000}{657000} = 0.6084$	$n_1 = 467$
المدينة 2	$a_2 = \frac{N_2}{N} = \frac{250000}{657000} = 0.3802$	$n_2 = 292$
الريف	$a_3 = \frac{N_3}{N} = \frac{7500}{657000} = 0.0114$	$n_3 = 9$

2- حساب حجم العينة n_h لكل طبقة

$$n_1 = n * a_1 = 768 * 0.6084 = 467$$

$$n_2 = n * a_2 = 768 * 0.3802 = 292$$

$$n_3 = n * a_3 = 768 * 0.0114 = 9$$

نلاحظ ان أغلبية العينة توزع للطبقتين 1 و 2 حيث أن 467 و 292 فرد لكل طبقة، أما الطبقة 3 فلها 9 أفراد.

الطبقة	المجتمع N	a_h	n_h	$f_h = {n_h}/{N_h}$
المدينة 1	*400 000*	*0.6084*	*467*	*0.0012*
المدينة 2	*250 000*	*0.3802*	*292*	*0.0012*
الريف	*7 500*	*0.114*	*9*	*0.0012*
المجموع	*657 500*	*1*	*768*	*0.0012*

نلاحظ ان التوزيع النسبي N يعطي ترجيح تلقائي لأن $f_h = 0.0012$ في الطبقات الثلاث.

2-2-2-8- التوزيع الغير نسبي (غير متكافئ)

مع التوزيع الغير تناسبي (غير متكافئ) تقسييم العينات يختلف حسب الطبقات. هناك عدة طرق للتوزيع الغير تناسبي يمكن ان تطبق : التوزيع التناسبي Y ، التوزيع التناسبي $\sqrt{N}$ ، التوزيع التناسبي $\sqrt{Y}$، التوزيع الأمثل، توزيع نيمان (*Neyman*) و التوزيع الامثل في حالة تساوي الفروقات.

تجدر الإشارة أنه رغم ان بعض الطرق تسمى تناسبية فإنه بمجرد أن التوزيع يختلف بين الطبقات فإنها تعتبر غير تناسبية.

1-2-2-2-8- التوزيع التناسبي Y

بما أن متغير الدراسة Y_{hi} و هو عبارة عن مقدر لحجم معين لوحدة *(i)* في الطبقة *(h)* فإن أحجام العينات n_h يمكن أن تحسب كنسب ل Y_h و هي مقدر كلي لحجم العينة n هذا النوع من التوزيع يسمى التوزيع التناسبي Y .

$$a_h = {Y_h}/{Y} \qquad \text{هنا}$$

و هذا معناه ان معامل التوزيع a_h لكل طبقة يساوي العلاقة بين مقدر حجم الطبقة و مقدر حجم المجتمع الكلي. هذا التوزيع جد معروف في الدراسات التجارية.

2-2-2-2-8- التوزيع التناسبي $\sqrt{N}$

كل الطرق السابقة للتوزيع تهتم أكثر لموضوع الدقة في التقديرات الكلية، لكن عادة يهتم الزبون بمعرفة جودة و دقة النسب المقدرة ل y_h .

مثال:

إذا كانت الطبقات عبارة عن ولايات فإن التقديرات الولائية أو الوطنية تكون مهمة، التوزيع على الطبقات الذي يستعمل التوزيع التناسبي $\sqrt{N}$ يمكن أن يحسن من دقة التقديرات.

في هذه الحالة معامل التوزيع a_h يحسب على الشكل التالي:

$$a_h = \frac{\sqrt{N_h}}{\sum_{h=1}^{L}\sqrt{N_h}}$$

بمعنى آخر معامل التوزيع $\sqrt{N_h}$ يساوي العلاقة بين الجذر التربيعي لحجم مجتمع العينة و مجموع الجذور التربيعية لحجم المجتمع لكل الطبقات.

مثال 2-8- تكملة

في المثال السابق لدينا عينة ثابتة 768 موزعة على ثلاث طبقات باستعمال التوزيع النسبي *N*. الان سنستعمل التوزيع التناسبي $\sqrt{N}$.

1- حساب قيم معامل التوزيع a_h

$$a_1 = \frac{\sqrt{N_1}}{\sum_{h=1}^{3}\sqrt{N_h}} = \frac{632.46}{121906} = 0.5188$$

$$a_2 = \frac{\sqrt{N_2}}{\sum_{h=1}^{3}\sqrt{N_h}} = \frac{500}{121906} = 0.4102$$

$$a_3 = \frac{\sqrt{N_3}}{\sum_{h=1}^{3}\sqrt{N_h}} = \frac{86.60}{121906} = 0.0710$$

2- حساب حجم العينة

$$n_1 = n * a_1 = 768 * 0.5188 = 398$$

$$n_2 = n * a_2 = 768 * 0.4102 = 315$$

$$n_3 = n * a_3 = 768 * 0.0710 = 55$$

الجدول التالي يلخص المقارنة بين التوزيعين للعينة على الطبقات.

	التوزيع التناسبي N				التوزيع التناسبي $\sqrt{N}$.			
الطبقة	N	a_h	n_h	f_h	$\sqrt{N}$	a_h	n_h	f_h
المدينة 1	400 000	0.6084	**467**	0.0012	632.46	0.5188	398	0.0016
المدينة 2	250 000	0.3802	**292**	0.0012	500	0.4102	315	0.0013
الريف	7 500	0.114	**9**	0.0012	86.60	0.710	55	0.073
المجموع	**657 500**	**1**	**768**	**0.0012**			**768**	**0.0012**

التوزيع التناسبي $\sqrt{N}$ يعطي لنا حجم عينة أصغر في المدينة 1 مقارنة بالتوزيع التناسبي N . و عكس ذلك في المدينة 2 و المنطقة الريفية حيث نسجل عينتين أكبر. و منه فمن المستحيل أن نتحصل على تقديرات جيدة ب 9 عينات في المنطقة الريفية. إذ أن نقصان حجم العينة في المنطقة 1 من 467 الى 398 له تأثير ضئيل على دقة التقديرات. و زيادة حجم العينة في المنطقة 2 من 292 إلى 315 له كذلك تأثير ضئيل على التقديرات. أما زيادة حجم العينة في المنطقة الريفية من 9 إلى 55 فيؤثر في جودة التقديرات بشكل كبير و هذا يؤدي إلى تعويض الدقة المفقودة في المنطقة 1.

3-2-2-2-8- التوزيع التناسبي $\sqrt{Y}$

طريقة أخرى لضمان جودة التقديرات للطبقات و كذا التقديرات الكلية هي استعمال التوزيع التناسبي $\sqrt{Y}$ أين y_{hi} هو مقدر لحجم . و هو عبارة عن طريقة بديلة التوزيع التناسبي $\sqrt{N}$ و هو أدق من التوزيع التناسبي $\sqrt{N}$ *في تقديرات الدراسة التي تكون مرتبطة مع* حجم Y_h أكثر من حجم الطبقة $\mathbf{N_h}$.

يحسب معامل التوزيع على الشكل التالي :

$$a_h = \frac{\sqrt{Y_h}}{\sum_{h=1}^{L} \sqrt{Y_h}}$$

هذا معناه ان معامل التوزيع يساوي العلاقة بين الجذر التربيعي لمقدر حجم الطبقة مع مجموع الجذور التربيعية المقدرة لحجم كل الطبقات.

التوزيعين التناسبيين $\sqrt{N}$ و $\sqrt{Y}$ تسمى في بعض الاحيان توزيع القوى و تكتب على الشكل التالي :

$$a_h = \frac{Y_h^P}{\sum_{h=1}^{L} Y_h^P}$$

حيث P يمثل عادة جزء مثال $\frac{1}{2}$.

4-2-2-2-8- التوزيع الأمثل

في حالة اختلاف تكلفة المقابلة من طبقة لأخرى و ان فروقات المجتمعات S_h^2 تختلف كثيرا. هناك طريقة للتوزيع الغير تناسبي تسمى التوزيع الامثل و هي الطريقة الوحيدة للتوزيع التي تأخذ بعين الاعتبار التكلفة في توزيع حجم العينة.

و من أجل استعمال هذه الطريقة نحتاج إلى دالة لنمذجة التكلفة و التي تكتب على الشكل التالي:

$$co\hat{u}t = C = c_0 + \sum_{h=1}^{L} c_h n_h$$

حيث أن c_h تمثل تكلفة المقابلة لوحدة في طبقة حيث أن (h=1,2,...,L).

c_0 تمثل التكلفة الثابتة.

هذه الدالة مفضلة عندما يكون عنصر تكلفة المقابلة هو العنصر المهم من أجل الحصول على قيم لكل وحدة.

معامل التوزيع a_h المستعمل في التوزيع الأمثل يكتب :

$$a_h = \frac{N_h S_h / \sqrt{c_h}}{\sum_{h=1}^{L} N_h S_h / \sqrt{c_h}}$$

التوزيع الامثل يقلص تباين التقدير لتكلفة معطاة، أو يقلص تكلفة العينة الكلية من اجل تباين كلي معطاه.

و من اجل الوصول إلى الهدف، العينة موزعة حول الطبقات التي تمون التكلفة كبيرة.

بصفة عامة، التوزيع الأمثل ...العينة المختارة في كل طبقة مهمة إذا كان :

- الطبقة كثيرة.

- تباين داخلي كبير الطبقة.
- تكلفة المقابلة ضئيل في الطبقة.

و من اجل تطبيق هذه الطريقة لابد من توفر معلومات حول الفروقات بين الطبقات و التكلفة الاحادية.

و في الواقع التكلفة و الفروقات قد تكون غير متوفرة و هنا يمكن تقديرهما لعينة الاولية أو عن طريق سبر للاراء سابق. و في حالة ما إذا كانت الفروقات و التكلفة متساوية لجميع الطبقات فإن التوزيع الامثل يصبح توزيع التناسبي N. و في حالة التساوي في التكلفة في جميع الطبقات نقوم باستعمال توزيع نيمان Neyman .

8-2-2-2-5- توزيع نيمان (Neyman)

هذه الحالة الخاصة للتوزيع الامثل الذي يسمى توزيع نيمان و الذي ينتج عندما تكون تكلفة المقابلة متساوية لكل الطبقات و الذي يوزع الحجم الكلي للعينة على الطبقات و الذي يقلص التباين الكلي للتقدير. فتوزيع نيمان يعطي وحدات أكثر للعينة في الطبقات الكبيرة أو الطبقات ذات تباين كبير مثله مقل التوزيع السابق فيمكن أن لا نعرف الفروقات إذن يمكن استعمال تقديرات .

معامل التوزيع يكتب على الشكل التالي:

$$a_h = \frac{N_h S_h}{\sum_{h=1}^{L} N_h S_h}$$

هذا معناه أن معامل التوزيع يساوي العلاقة ناتج حجم الطبقة و الجذر التربيعي للتباين مع ناتج حجم الطبقة و الجذر التربيعي لتباين كل الطبقات.

و عندما لا يمكن الحصول على قيم دقيقة للتباين يمكن اعتبار أن العلاقة تباين الطبقة لمتوسط الطبقية $S_h/\bar{Y}_h$ ثابت بين الطبقات. هذه الفرضية تقلص توزيع نيمان إلى توزيع تناسبي *Y*.

8-2-2-2-6- التوزيع الامثل في حالة تساوي الفروقات

حالة خاصة أخرى للتوزيع الأمثل عندما يمون قيم التباين متساوية في كل الطبقات- و هذا غير معتاد- و هي ضئيلة الاستعمال. كما أنها تستعمل عندما لا نعرف الفروقات المجتمعات و تفترض أنها متساوية. و هذا يعطي أكثر وحدات للطبقات الكبيرة :

$$a_h = \frac{N_h/N_h/\sqrt{c_h}}{\sum_{h=1}^{L} N_h/\sqrt{c_h}}$$

3-2-8- بعض الاعتبارات الحاصة عند التوزيع

يجب أخذ بعين الاعتبار مجموعة من النقاط عند توزيع الوحدات :

1. استعمال البيانات الثانوية لتوزيع التناسبي Y و التوزيع التناسبي $\sqrt{Y}$: عند استعمال هذين التوزيعين في قيم Y غير معروفة لكل وحدات المجتمع، إذن قيم المتغيرات الثانوية المرتبطة مع Y يمكن ان تستعمل انطلاقا من دراسات سابقة أو بيانات إدارية.
2. التوزيع القوي : في حالة العينة الطبقة باستعمال التوزيع الامثل، التوزيع نيمان، التوزيع التناسبي Y و التوزيع التناسبي $\sqrt{Y}$ ، يمكن أن n_h يكون أكبر من حجم المجتمع الطبقي N_h و هذا ما يسمى بالتوزيع القوي أو الزائد. في هذه الحالة يجب احصاء الطبقات التي تكون فيها توزيع الوحدات زائد، إذن حجم العينة الكلي الناتج عن هذه التوزيع الكلي يكون أقل من حجم العينة الأولي. الحل هو زيادة العينات في الطبقات الباقية أبن n_h أقل من N_h باستعمال الفائض في الاحجام العينات في الطبقة الزائدة.
3. الحجم الادنى للعينة الطبقية : في كل طبقة عدد يطلب على الاقل وحدتين إلا أن كل الطرق السابقة للتوزيع تمكن أن تعطي قيم أقل من 2 و حتى أقل من 1. لذا في هذه الحالة يمكن إعطاء على الاقل 2. إحدى الحلول هي إعطاء وحدتين لكل طبقة ثم توزيع الباقي باستعمال إحدى الطرق السابقة.

خلاصة

ان تحديد حجم العينة هو عبارة عن مسار معقد بين الاختبارات الميدانية و بين احتياجات الدقة و التي تكون عادة مفقيدة بأمور عملية مثال الميزانية، التكلفة، الوقت المتاح، عدد المستجوبين. إن قرارات التي يجب اتخاذها على حجم العينة يمكن ان يطلب تحليل جديد او تعديلات على الاهداف، المتطلبات و الحاجيات المطلوبة، درجة الدقة، عناصر الدراسة، النشاطات الميدانية...الخ، المحددة مسبقا.

إن الهيئة الإحصائية و الزبون يحددان عادة المردودية التي من خلالها يمكن للزبون الحصول على حجم العينة اللازم، فهم عادة ما ينظمون استجواب سريع، تطبيق طريقة اخرى لحجم العينة،...

فإذا تم استعمال العينة الطبقية، فالعينة يجب أن توزع بين الطبقات. هناك طريقتين: تحديد حجم العينة الاجمالي و توزيعها بين الطبقات من اجل تقليص التباين أو نظرا لمتطلبات الدقة المطلوبة ، تحديد حجم العينة اللازم في كل طبقة فيجب تحديد معامل التوزيع في كل طبقة لهذين الطريقتين.

هناك عدة طرق للتوزيع: التوزيع التناسبي مع حجم المجتمع و هي الطريقة التي تعطي تقسيمات للعينة متساوية في كل طبقة. أما طريقة التوزيع الغير تناسبية فهي توزع العينة بين الطبقات مع الاخذ بعين الاعتبار حجم المجتمع في كل طبقة أو التكلفة وكذا التباين في كل طبقة.

المحور التاسع : تمارين و تطبيقات

9.0. السلسلة الاولى حول انماط اختيار العينات

التمرين الأول

نريد اختيار عينة تساوي 5 بالمائة من طلاب الجامعة البالغ عددهم 20000 طالب و طالبة. بالاستعانة بجداول الارقام العشوائية أو برنامج Excel .

ما هو حجم المجتمع المستهدف؟
ما هو حجم العينة المراد دراستها؟
أذكر الخطوات التي يجب اتباعها لتحديد العينة؟

التمرين الثاني

ليكن لدينا مجتمع مكون من 300 ذكور و 200 إناث و يريد باحث إجراء دراسة تضم 10 بالمائة من كل جنس.

ما هو حجم العينة لكل جنس باستعمال العينة العشوائية البسيطة؟
بعدها أراد الباحث إجراء الدراسة على أساس المستوى الدراسي و هذا باستعمال العينة الطبقية مع العلم أن :
المستوى الاول يضم 400 فرد
المستوى الثاني يضم 300 فرد
المستوى الثالث يضم 200 فرد
المستوى الرابع يضم 100 فرد.

العينة المراد دراستها تضم 200 فرد ممثلة لمجتمع البحث و هذا بالاستعانة بعينة عشوائية غير تناسبية (غير متكافئة).

ما هو حجم العينة في كل طبقة؟
أعد حساب حجم العينة لكل طبقة و هذه المرة باستعمال طريقة العينة التناسبية مع حجم كل طبقة؟

التمرين الثالث

نريد إجراء دراسة استقصائية على مجتمع مستهدف موجود في منطقة جغرافية محددة (مثال قرية او حي) و الوحدة المستعملة هي الاسرة. مع العلم أننا لا نملك قائمة شاملة اعدد الاسر. المعلومة الوحيدة التي نملكها هي أن عدد السكان الاجمالي هو 15000 نسمة و ان

متوسط أفراد الاسرة هو 5 افراد في الاسرة. بالاستعانة بالعينة المنتظمة نريد إجراء دراسة على عينة مكونة من 500 أسرة.

أحسب حجم المجتمع المستهدف؟
ما هي الخطوات التي يجب اتباعها لتحديد العينة (أي اختيار 500 أسرة)؟

1-9- السلسلة الثانية حول التقديرات و الترجيح

التمرين الأول

نريد انشاء عينة ل 150 عائلة من مجتمع يتكون من 408 أسرة استفادت من اعانات الدولة على مستوى بلدية تحتوي على 10 أحياء حيث :

الحي											المجموع
عدد الاسر	3	8	5	5	0	5	06	3	0	7	1	408

المطلوب:

1. أحسب احتمال الانتماء الوحدات في كل حي؟
2. أحسب قوة كل وحدة لكل حي؟
3. ما هي الخطوات التي يجب اتباعها لاختيار 150 عائلة باستعمال نمط العينة العشوائية البسيطة؟
4. كم تمثل كل عائلة يتم اختيارها من الحي (F) ؟
5. أحسب الوحدات التراكمية للمجتمع؟
6. أجد مجالات للوحدات الخاصة بكل حي؟
7. باستعمال برنامج Excel تتتم اختيار الاعداد التالية : 77، 36، 106، 205، 388، 38، 45. إلى أي حي تنتمي هته العائلات؟

التمرين الثاني

لتكن لدينا مجتمع متكون من 1000 طالب ماستر مقسم الى أربع تخصصات اقتصاد كمي، مالية و بنوك، تأمينات ، محاسبة. حيث أن كل تخصص يضم عدد من طلبة 120، 260، 220، 400 على التوالي.

نريد عينة احتمالية ذات حجم n=250.

1. باستعمال توزيع التناسبي مع كل طبقة، ما هو حجم العينة في طل طبقة؟
2. ما هي قوة الترجيح وكم يمثل كل طالب في تخصص؟
3. نعيد حساب قوة الترجيح مع العلم أن n1=60 , n2=70, n3=45, n4=75

التمرين الثالث

يريد قسم العلوم الاقتصادية إجراء دراسة حول عوامل إعادة الطلبة للسنة أثناء مسارهم الدراسي. و باستعمال عينة عشوائية بسيطة تضم n=250طالب تم اختيارهم من قائمة N=850. و من أصل 250طالب أجاب على الاستبيان n_r=200 . أثناء القيام بالدراسة لا تتوفر أي معلومة إضافية يمكن استعمالها من أجل إنشاء عينة طبيقية.

نتائج الدراسة كانت على النحو التالي:

	قبل إجراء الطبيقية		
	ذكور	إناث	عدد المجيبين
كل العينة	30	170	200
الطلبة المعيدين	11	21	32

الأسئلة

1- كم يمثل كل طالب في العينة؟
2- أرسم جدول يمثل الإحصائيات السابقة باستعمال قوة الترجيح؟

بعد إنجاز الدراسة، توفر لدينا معلومة حول عدد الذكور و الإناث المسجلين في القسم. 620 إناث، 230 ذكور.

تريد إدارة القسم أن تكون النتائج دقيقة تأخذ بعين الاعتبار الطبقية حسب الجنس.

3- كم يمثل كل طالب في طبقته ؟ (W_{pst})
4- أعد رسم الجدول الأول باستعمال قوة الترجيح الجديدة ؟

2-9- السلسلة الثالثة حول تحديد حجم العينة

التمرين الاول

تريد إدارة قسم العلوم الاقتصادية دراسة مدى رضا طلبة الماستر عن مسار التكوين. هناك أربع طبقات في المجتمع المستهدف بالدراسة حسب البيانات التالية.

الطبقة	حجم المجتمع Ni
ماستر اقتصاد كمي	$N_1=50$
ماستر مالية و بنوك	$N_2=85$
ماستر محاسبة	$N_3=60$
ماستر تأمينات	$N_4=45$
المجموع	N=240

الفرضيات

- هامش الخطأ هو 5 بالمائة للمجتمع الكلي (الحالة الاولى فقط).
- مجال الثقة هو 95 بالمائة Z=1.96
- نفترض أن p=0.5
- نسبة الاستجابة المنتظرة و المقبولة 85 بالمائة.

الأسئلة

1- ما هو حجم العينة المطلوب في كل طبقة؟
2- باستعمال طريقة التوزيع التناسبي N و التوزيع التناسبي $\sqrt{N}$. قم بتوزيع العينة على الطبقات الأربع؟ (كل النتائج المتحصل عليها تضع في جدول واحد).
3- كم يمثل كل طالب في تخصص اقتصاد كمي إذا علمت أن عدم الإجابة في هذا التخصص تقدر ب 5 طلاب لا يجيبون على الاستبيان.

المراجع :

- Statistics Canada, **(2003) Survey Methods and Practices** Catalogue no. 12-587-X October 2003.
- Lehtonen Risto and Erkki Pahkinen (2004) **Practical Methods for Design and Analysis of Complex Surveys**, second edition, John Wiley & Sons, Ltd England.
- De Leeuw & Joop J. Hox & Don A. Dillman (2008), **International Handbook of Survey Methodology.**
- Hansen M.H., W., Hurwitz, W.G. Madow (1953) **Sample survey methods and theory**, Volume I, Wiley publications in statistics, Walter A. Shewhart, Editor.

Entente de licence ouverte de Statistique Canada

Octroi de licence

Sous réserve des conditions de la présente entente, Statistique Canada vous accorde une licence mondiale, libre de redevances et non exclusive vous permettant :

- d'utiliser, **de reproduire, de publier**, de diffuser gratuitement ou de vendre l'information;

- d'utiliser, de reproduire, de publier, de diffuser gratuitement ou de vendre les produits à valeur ajoutée;

- d'accorder des sous licences conférant une partie ou la totalité de ces droits, conformément aux conditions de cette entente.

Mention de la source

(a) Pour tout exercice de vos droits d'utilisation de l'information, vous devez inclure et maintenir la mention suivante :

Source : Statistique Canada, nom du produit, date de référence. Reproduit et diffusé « tel quel » avec la permission de Statistique Canada.

(b) Pour toute information contenue dans un produit à valeur ajoutée, vous devez inclure dans ce produit la mention suivante :

Adapté de Statistique Canada, nom du produit, date de référence. Cela ne constitue pas une approbation de ce produit par Statistique Canada.

La présente entente entre en vigueur à la date et à l'heure où vous accédez à l'information et est résiliée automatiquement si vous enfreignez l'une de ses conditions.