



المعهد العربي للتدريب والبحوث الإحصائية

Arab Institute for Training and Research in Statistics

دورة متوسطة حول تصميم عينات المسوح الإحصائية

10 مايو / أيار - 5 تموز / يوليو 2022

المحاضرة (3): اطر المعاينة الإحصائية

www.aitrs.org
info@aitrs.org

إعداد: نايف عابد
17 أيار - 2022



المحتويات

- 1- طرق إعداد إطار المعاينة
- 2- أنواع اطر المعاينة
- 3- أخطاء إطار المعاينة
- 4- الإطار الثنائي والإطار المتعدد
- 5- حساب خطأ الشمول في التعدادات Coverage Error

طرق إعداد إطار المعاينة

في العادة يتم بناء قواعد بيانات تسمى اطر المعاينة من المصادر التالية:

1- التعدادات والمسوح الشاملة

2- السجلات المركزية (السجل السكاني، سجل الاعمال، سجل الحيازات

الزراعية...الخ)

3- قائمة الوحدات الجغرافية (مناطق العد، المربعات (البلوكات)، التجمعات

المأهولة) بحيث يتوفر خرائط جغرافية لهذه الوحدات.

أنواع اطر المعاينة

أمثلة على أهم أنواع الأطر المستخدمة في المؤسسات الإحصائية الرسمية

1- إطار الأسر المعيشية: قائمة تحتوي على أسماء أرباب الأسر وعناوينهم وطريقة الوصول لهم

بالإضافة إلى معلومات أساسية وديمغرافية عن الأسرة.

2- إطار المنشآت الاقتصادية (الوحدات الاقتصادية) وفيها كل البيانات التعريفية وعنوانها وبيانات

أخرى أهمها النشاط الاقتصادي الرئيسي، وعدد العاملين، والكيان القانوني لها، تصنيفها ومدى

استخدام التسجيل فيها....الخ.

3- إطار الحيازات الزراعية: قائمة تحتوي على كل ما يمكن تصنيفه كحيازة زراعية، واسماء

وعناوين الحائزين، ونوع الحيازة اذا كانت نباتية او حيوانية وحجمها وغير ذلك من المعلومات

الأساسية.

مزايا اطر المعاينة الجيدة

- ان يتصف بشموله العالي لمجتمع الهدف
- ان يحتوي على معلومات مساعدة كثيرة
- ان تتصف البيانات المساعدة بالحدائة والدقة

إطار العينة الشاملة Master Sample

هو عبارة عن عينة كبيرة الحجم ممثلة للمجتمع بدرجة عالية يتم تحديثها من اجل سحب عينات منه.

مزايا إطار العينة الشاملة:

1- حداثة البيانات لوحدات المعاينة مقارنة مع الإطار الأصلي

2- إمكانية استعماله للعينات الدورانية Rotational Samples

3- تكلفة إنشاءه قليلة والجهد المبذول لذلك محدود مقارنة مع إجراء تعداد جديد

عيوب إطار العينة الشاملة:

1- زيادة العبء على وحدات المعاينة نظرا لكثرة استخدامها في عملية التحديث وفي تصميم عينات المسوح التي يتم تنفيذها.

2- هناك احتمالية بوجود تحيز اكبر في العينات المصممة منه اذا كانت نسبة عدم الاستجابة متوسطة او عالية

3- بحاجة مستمرة للتحديث للإبقاء على كفاءته

4- قد لا يغطي المناطق او وحدات المعاينة التي طرأ عليها تغيرات طارئة بعد عملية إنشائه، وبالتالي يصبح غير ممثل لمجتمع الدراسة.

الإطار الجغرافي Area Frame

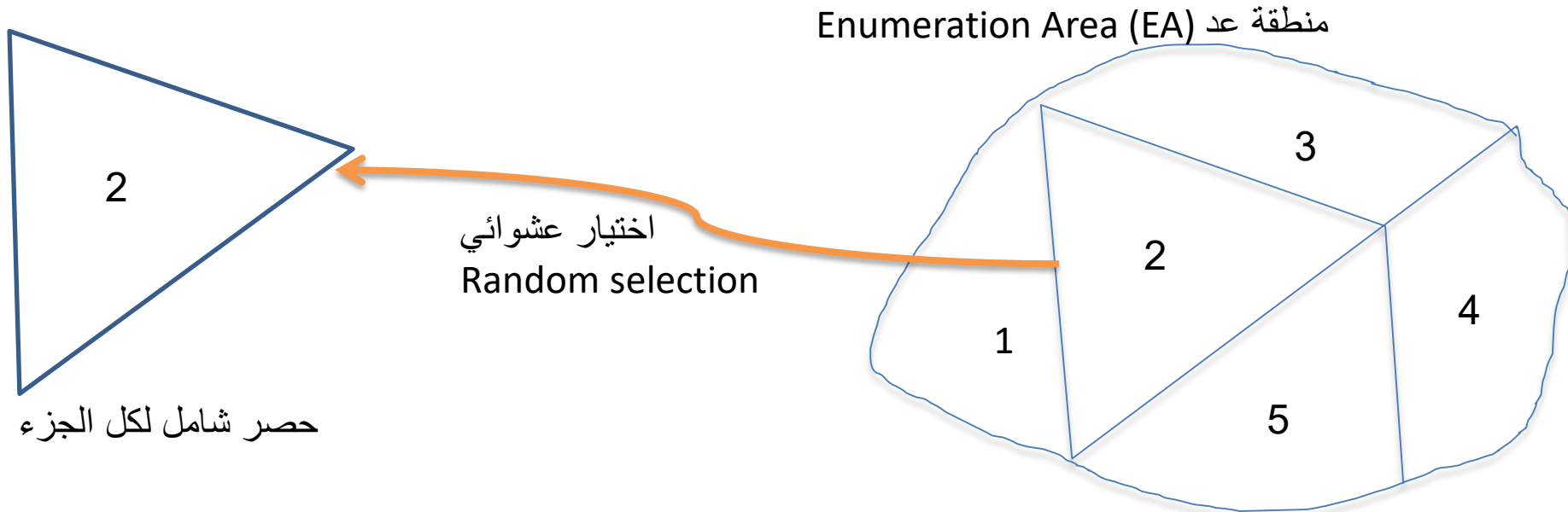
1- يتم بناءه بالاعتماد على الوحدات الجغرافية المقسمة في التعدادات العامة

2- في العادة تكون الوحدات الجغرافية هي مناطق العد او البلوكات Enumeration Areas or Blocks

3- يتم الاعتماد عليه في تصميم العينات المساحية Area Samples

4- في العادة يتم الاعتماد على الخرائط في سحب العينات.

ويتم حصر كافة المباني في جزء من منطقة العد مسحوب بطريقة عشوائية او في كل منطقة العد.



إطار الأرقام التليفونية Telephone numbers Frame

1- يتم بناءه من ارقام التليفونات الثابتة للاسر او المنشات

2- تقسيم المناطق فيه يكون بحسب رقم مقدمة التالفون ويختلف في العادة عن تقسيم

المناطق حسب التعداد

3- قد لا يتوفر اي معلومات عن الاسرة/المنشاة التي تمتلك التالفون باستثناء معلومات

الشخص المسجل الرقم باسمه.

4- اخطاء الاطار فيه كثيرة مثل نقص وزيادة الشمول والتكرار والارقام غير المستخدمة

للاتصال وغير ذلك.

5- سهل الاستخدام وقليل التكلفة لسحب العينات قد يتم استخدامه لسحب عينات عشوائية

حسب المنطقة في حال عدم توفر اطر محدثة اخرى تفي بالغرض مثل المسح الاسري

باستخدام الاتصال على الأرقام العشوائية

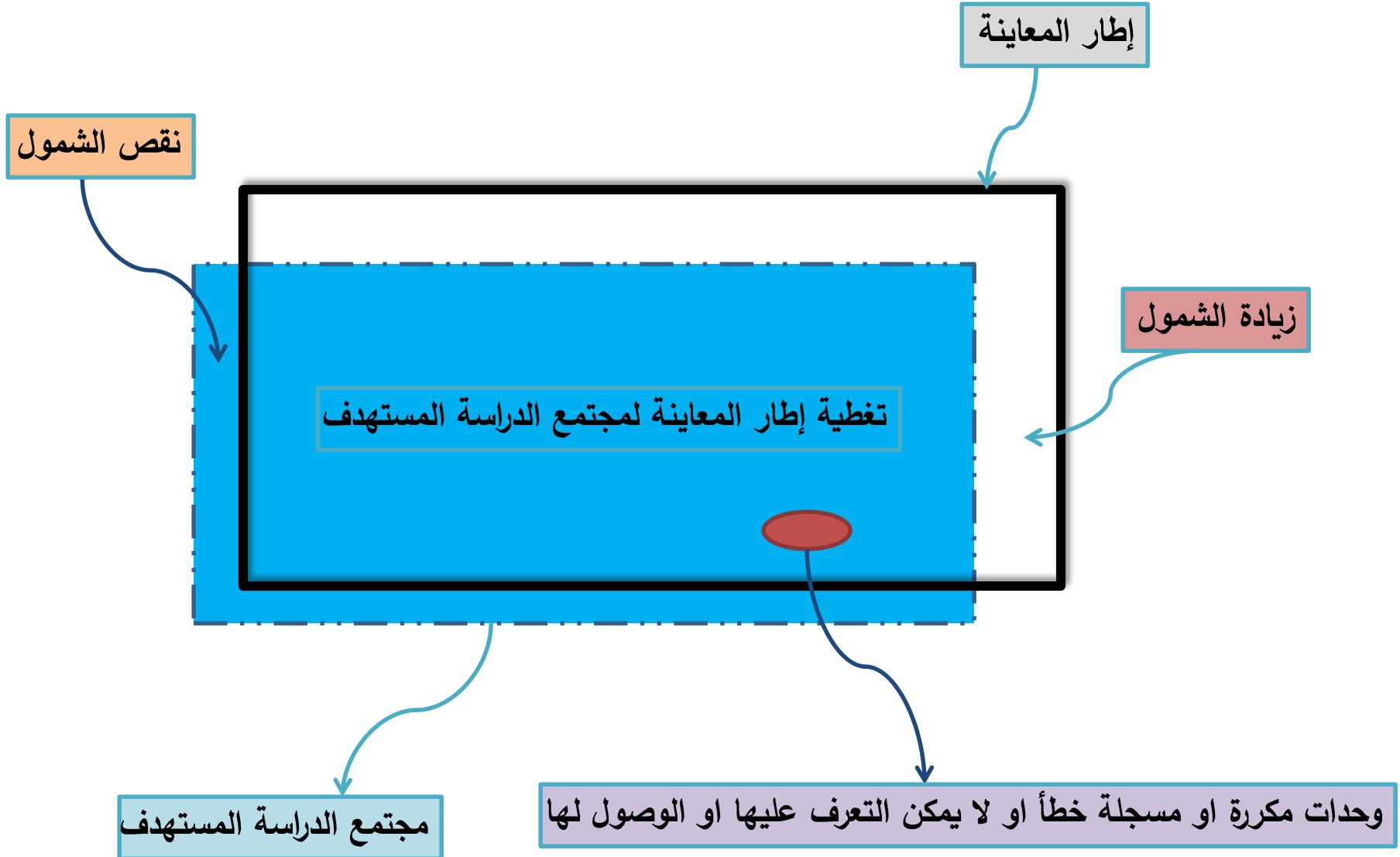
household survey random digit dialing(RDD)

بعض الاطر الاخرى

- اطار عناوين البريد العادي
- اطار العناوين الالكترونية (البريد الالكتروني)

ملاحظة: اذا كان المجتمع المستهدف غير محدد Infinite population فمن الصعب ان يكون هناك اطار معاينة.

قياس أخطاء إطار المعاينة:



تقدير نسبة زيادة الشمول Over coverage

(1) من بيانات العينة: لنفرض أن حجم العينة = n ، وكان عدد الوحدات التي تم

تصنيفها أنها زيادة شمول = n_{ov}

ونسبة زيادة الشمول = p_{ov}

$$p_{ov} = \frac{n_{ov}}{n}$$

(2) من بيانات تحديث الإطار: لنفرض أن حجم الإطار = N ، وكان عدد الوحدات

التي تم تصنيفها في عملية التحديث أنها زيادة شمول = N_{ov}

ونسبة زيادة الشمول = P_{ov}

$$P_{ov} = \frac{N_{ov}}{N}$$

*ملاحظة : يتم تقدير نسبة التكرار والوحدات التي لم يستدل عليها بنفس الطريقة من بيانات العينة او التحديث.

قياس نسبة نقص الشمول Under Coverage

يمكن قياس نسبة نقص الشمول كالتالي:

- من خلال بيانات حديثة بعد إجراء تعداد أو حصر شامل جديد
- من مصادر أخرى مرتبطة بالإطار مثل السجلات الإدارية والمركزية
- يمكن تقدير نسبة نقص الشمول من بيانات العينة إذا كانت العينة مساحية

معالجة وتقليل أخطاء الإطار

فيما عدا إجراء تعداد جديد، هناك ممارسات عملية للتقليل من أخطاء الإطار كما يلي:

- تحديث بيانات مستمر للإطار من خلال بيانات المسوح المنفذة
- استخدام العينات المساحية وتحديث كل مناطق العد في العينة
- إضافة وحدات معاينة حديثة من مصادر أخرى مرتبطة بالإطار مثل السجلات الإدارية والمركزية.

• استخدام اطر متعددة للعينات multiple frame sampling

• استخدام مصادر ذات دقة لمعالجة الأوزان وتعديلها ومعايرتها

المعاينة من الأطر المتعددة Multiple Frame Sampling

- حينما لا يتوفر إطار معاينة شامل يحتوي على كل وحدات مجتمع الهدف.
- عند وجود عدة اطر جزئية تشكل في مجموعها تغطية وحدات مجتمع الهدف
- قد يكون هناك تقاطع بين اطر المعاينة جزئية.

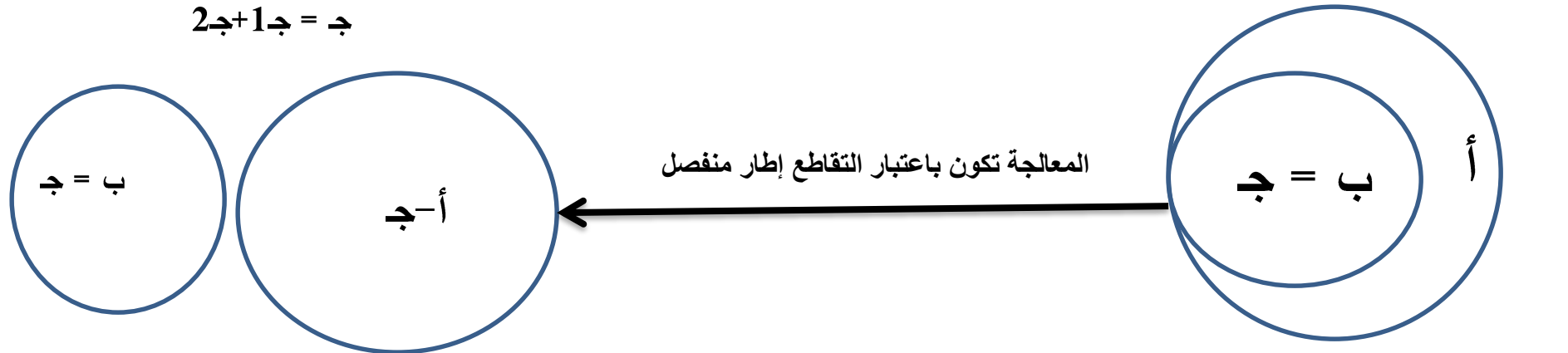
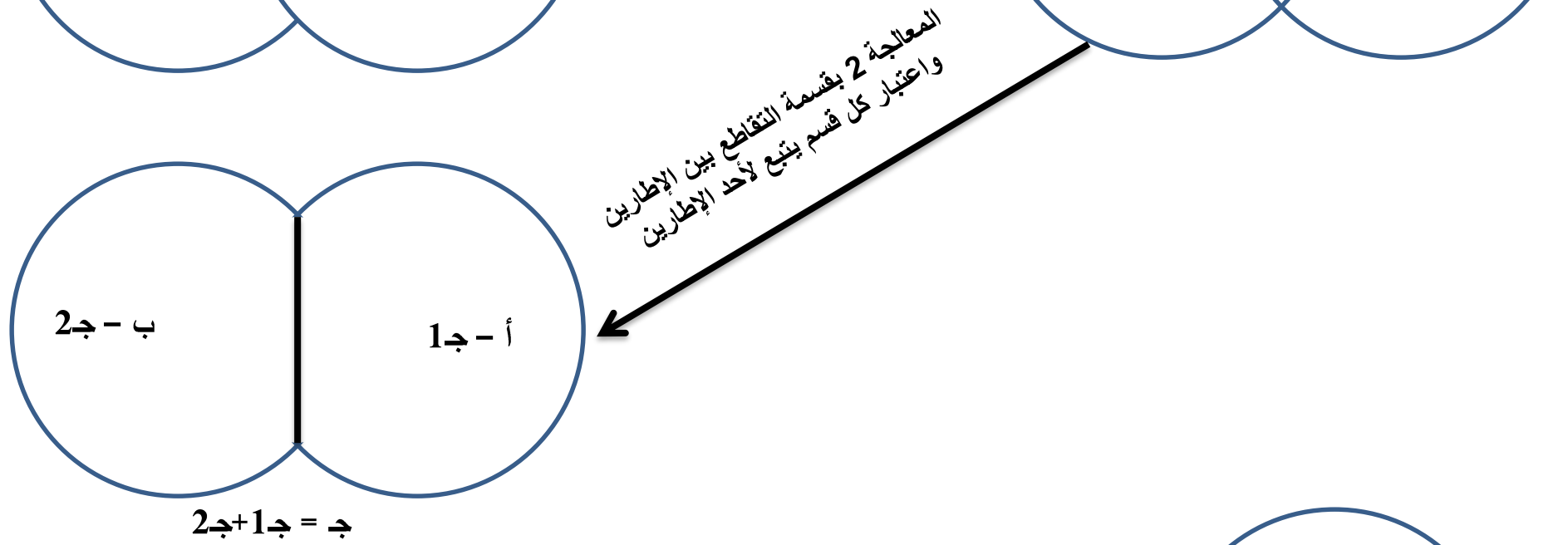
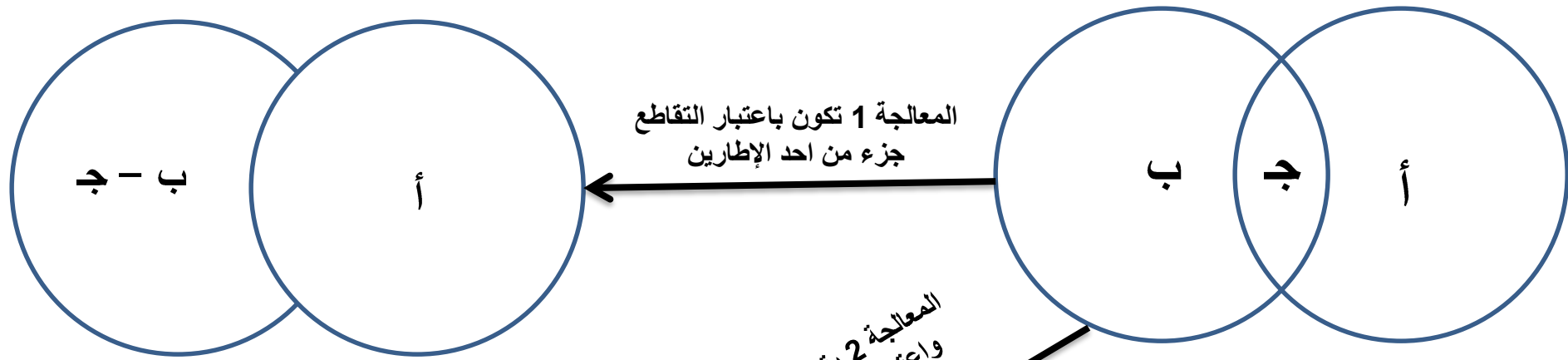
• إذا كان هناك إطارين جزئيين تسمى الطريقة Dual Frame Sampling

لنفرض أن هناك إطارين: أطار أ وإطار ب (overlap) بينهما هو ج المؤسسات

الصحية ويشكلان في مجموعها تغطية لمجتمع الهدف

مثال: الإطار الأول هو قائمة المؤسسات الصحية، والإطار الثاني هو قائمة الأطباء، ومجتمع الهدف

هو الممرضين.



التقديرات في المعاينة من الأطر المتعددة

- يجب الانتباه في حساب اوزان التصميم لكل عينة حسب الإطار المسحوبة منه
- يتم ايجاد التقاطع والتكرار بين الأطر بتعريف كل وحدة إلى اي اطار تنتمي
- يتم تعريف مجالات التحليل Domains حسب حجمهم في كل إطار وتوحيدهم
- استخدام البيانات المساعدة المتوفرة Auxiliary data ضروري جدا بعض أنواع التقديرات

حساب خطأ الشمول في التعدادات Coverage Error

• بعد إجراء التعدادات يكون هناك نقص في شمول كل الوحدات التي يجب عدّها والوصول لها.
• من أسباب نقص العد: هو عدم القدرة على الوصول للوحدات المطلوبة أثناء فترة العد. أو عدم الوصول لها بالخطأ.

• من اجل تقدير نقص العد يجب تنفيذ الدراسة البعدية للتعداد Post-Enumeration Survey
PES

• تعتمد الدراسة البعدية للتعدادات على إعادة العد لعينة ممثلة من المجتمع

• يتم تنفيذ الدراسة البعدية خلال فترة لا تتجاوز شهر من فترة تنفيذ التعداد نفسه

• الهدف هو قياس نسبة خطأ شمول التعداد Coverage error وقياس مدى جودة بيانات التعداد

او خطأ المحتوى Content error

المطابقة بين التعداد والدراسة البعدية (Matching)

- بشكل عام ، تتضمن عملية المطابقة الأساسية مقارنة جميع الافراد الواردين في التعداد والدراسة البعدية وتحديد حالة المطابقة لديهم.
- تكون سجلات الدراسة البعدية هي الاساس في عمليات المطابقة، حيث يتم تحديد كل فرد ورد في الدراسة البعدية وتحديد حالته اذا كان متطابق ما بين التعداد وما بين الدراسة البعدية ام انه غير متطابق.
- اجراء المطابقة يعتمد على مقارنة العناوين والأسماء والخصائص الديموغرافية نتائج الدراسة البعدية PES وبين نتائج التعداد
- يتم وضع مجموعة من قواعد المطابقة تشمل الحالات التي يمكن ان تواجه المدققين القائمين على عملية المطابقة

انواع المطابقة

- **المطابقة اليدوية:** وهي التي تعتمد على العمل المكتبي وقيام المدقق بمقارنة البيانات التعريفية لكل فرد
- **المطابقة الآلية:** تعتمد على المقارنة الآلية لكل فرد من خلال برامج ولو غارثميات تأخذ بعين الاعتبار مدى تطابق البيانات التعريفية لكل فرد (تعتمد على احتمالية التطابق)

التعريفات والمفاهيم والتصنيفات للحالات الناتجة من عمليات المطابقة
Matching Operations

الرمز	تصنيف الفرد في الدراسة البعدية	المعنى
1	Non-mover غير متنقل	هو الفرد المقيم في اسرة خاصة لم تتغير ما بين فترة الدراسة البعدية والتعداد
2	Out-mover المنفصل	هو الفرد الذي كان مقيما في فترة التعداد مع اسرة معينة ولم يعد يقيم مع نفس الاسرة في فترة الدراسة البعدية
3	In-mover الوافد	هو الفرد الذي كان مقيما في فترة الدراسة البعدية مع اسرة معينة ولم يكون يقيم مع نفس الاسرة في فترة التعداد، حيث اقام هذا الفرد في الاسرة بعد اجراء التعداد
4	Out-of-scope خارج النطاق	هو الفرد الذي لا ينتمي لمجتمع الهدف في فترة التعداد، وعلى سبيل المثال: الطفل المولود بعد التعداد، او الفرد المقيم خارج البلد فترة التعداد وعاد بعد التعداد ليقوم مع اسرة معينة في فترة الدراسة البعدية.

أخطاء شمول التعداد:

• خطأ الشمول يتألف من ثلاثة انواع:

• **Omissions error** خطأ بسبب الحذف (الاهمال)، ويحصل هذا بسبب عدم الوصول

الى وحدات سكنية او اسر او افراد ضمن الاسرة، او بسبب عدم تغطية جزء من مناطق جغرافية اثناء مراحل الحصر الاولي في التعدادات، او فقدهم لاي سبب اخر. (وهذا جزء من نقص الشمول)

• **Duplications** خطأ بسبب تكرار العد للفرد او الاسرة او الوحدة السكنية لمرتين او

اكثر سواء من خلال الوصول لهم اكثر من مرة، وقد يكون هذا بسبب اخطاء في ترسيم وحدود مناطق العد، ويحصل التكرار احيانا بسبب عد بعض الاشخاص في وحدات سكنية مختلفة (وهذا جزء من زيادة الشمول)

• **Erroneous inclusions** التضمين او الاحتواء الخاطيء لبعض الحالات التي لا

ينطبق عليها تعريف التعداد مثل الوفيات قبل فترة الاسناد الزمني، والولادات بعد فترة الاسناد الزمني، وقد تشمل غير المقيمين الذين لا ينطبق عليهم العد الفعلي. (وهذا جزء من زيادة الشمول)

طريقة تقدير النظام الثنائي (DSE) Dual system estimation method

		Census		المجموع
		التعداد		
		تم عدّهم	لم يتم عدّهم	
الدراسة PES البعديّة	تم عدّهم	N_{11}	N_{12}	N_{1+}
	لم يتم عدّهم	N_{21}	N_{22}	N_{2+}
المجموع		N_{+1}	N_{+2}	N

$$\hat{N} = \frac{N_{+1} * N_{1+}}{N_{11}}$$

N_{1+} = مجموع التعداد
 N_{+1} = مجموع الدراسة البعدية

تقدير المجتمع الحقيقي

$$R_{uc} = \frac{\hat{N} - N_{1+}}{\hat{N}} * 100\%$$

تقدير تقريبي لنسبة نقص العد

تعديل نتائج التعداد حسب الدراسة البعدية

- في التعدادات السكانية يتم إضافة نسبة نقص العد الى التعداد الفعلي، ولكن لا يستفاد من هذه النسبة في تشكيل الإطار.
- لا يتم أي تعديل على الخصائص والمتغيرات الأخرى في التعدادات
- اذا كانت نسبة الشمول منخفضة فهذا يعني ان التعداد لم يكن شاملا بما يكفي

تقييم خطأ المحتوى Content Error

يتم قياس جودة البيانات الناتجة من التعداد من خلال بعض المقاييس:

- معدل الفرق الصافي Net difference rate
- مؤشر عدم الاتساق Index of inconsistency
- معدل التوافق Rate of agreement

المراجع

- <https://www2.census.gov/programs-surveys/decennial/2020/program-management/planning-docs/PES-detailed-operational-plan.pdf>
- [https://unstats.un.org/unsd/demographic/standardmeth/handbooks/Manual PESen.pdf](https://unstats.un.org/unsd/demographic/standardmeth/handbooks/Manual_PESen.pdf)

شكراً لحسن الاستماع