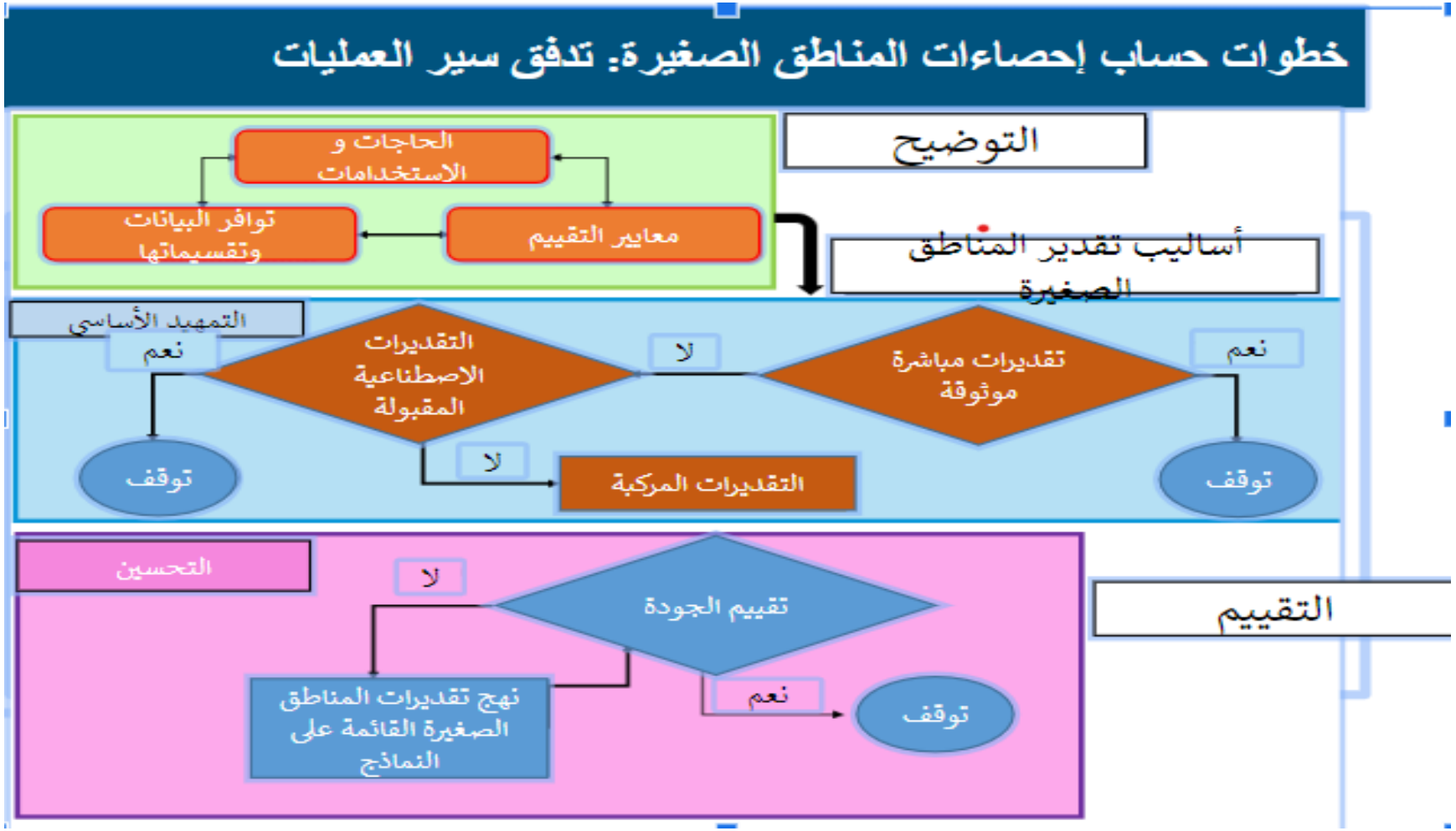
EU Twinning Project on
Statistics in Jordan

دائرة الإحصاءات العامة
Department of Statistics

النظرية وأفضل الممارسات المتعلقة بتقديرات المناطق الصغيرة
المعيار 2: منهجية إنتاج إحصاءات المناطق الصغيرة

26 شباط 2024

خطوات احتساب إحصاءات المناطق الصغيرة



الاستدلال القائم على التصميم

تصميم العينة	العامل
	الوحدة السكانية
	العينة
	توزيع الاحتمالية
	بارامترات
	المُقدر

المُقدر المباشر

- U المجتمع المحدود (finite population) للحجم N
- $(y_1 \ldots, y_N)$ قياسات المتغير المستهدف على الوحدات السكانية
- الكمية المستهدفة: مثال على المتوسط الحسابي للوحدة السكانية

$$h\ (y_1 \ldots, y_N) = \sum_{i=1}^{N} y_i\ /N$$

- s عينة عشوائية للحجم n التي تم سحبها من الوحدة السكانية U
- r =U-s الوحدة غير العينية للحجم (N-n)

المقدر المباشر الأساسي

- π_j احتمالية الدمج للوحدة j في العينة

- $d_j = 1/\pi_j$ وزن العينة للوحدة j
- مقدر Horvitz-Thompson (HT) للمتوسط الحسابي

$$\hat{\bar{Y}}_{DIR} = \frac{1}{N}\sum_{i=1}^{n} d_i y_i$$

- مقدر التباين غير المتحيز في التصميم (تحت بند التقريب):

$$\widehat{V}(\hat{\bar{Y}}_{DIR}) \cong \frac{1}{N}\sum_{i=1}^{n} d_i(1-d_i)y_i^2$$

تقدير المجال

- U مقسمة إلى D مجالات U_1، ..... ،U_d للأحجام N_1،.....،N_D
- S_d العينة الفرعية للحجم n_d التي تم سحبها من U_d $\left(n = \sum_{d=1}^{D} n_d\right)$
- $r_d = U_d - s_d$ عينة مكملة، للحجم $N_d - n_d$.
- البارامتر المستهدف: المتوسط الحسابي للمجال

$$\bar{Y}_d = \sum_{i=1}^{N_d} y_i \, / N_d$$

المقدر المباشر الأساسي (المجال)

- المقدر للمتوسط الحسابي للمجال Horvitz-Thompson (HT)

$$\hat{\bar{Y}}_{DIR}^{d} = \frac{1}{N_d} \sum_{i \in s_d} d_i y_i$$

- مقدر التباين التصميمي غير المتحيز (تحت بند التقريب)

$$\widehat{var}(\hat{\bar{Y}}_{DIR}^{d}) \cong \frac{1}{N_d} \sum_{i \in s_d} d_i (1 - d_i) y_i^2$$

- يستخدم HTفقط المتغير المستهدف و بيانات العينة الخاصة بمنطقة معينة

لننتقل إلى برنامج R

التقديرات الأساسية باستخدام المتغيرات المساعدة (الإضافية)

- p المتغيرات المساعدة (الإضافية) $k = 1, \ldots, p$ و X_{ik}, $i = 1,\ldots,n$

- المجاميع السكانية المعروفة لمتغيرات p في المجالات d

$$\boldsymbol{X_d} = (X_{1d}, \ldots, X_{pd}), \quad \mathrm{d} = 1, \ldots, \mathrm{D}$$

التقديرات الأساسية باستخدام المتغيرات المساعدة (الإضافية)

- محاولات لتحسين دقة مقدر HT التقليدي باستخدام معامل الارتباط بين المتغير المستهدف والتغاير من خلال تعديل أوزان أخذ العينات الأولية.

- لا يزال هذا المقدر غير متحيز تقريبيًا من حيث التصميم، وينبغي أن يسمح بتقليل التباين في التصميم.

- لا يزال مقدرًا مباشرًا، لأنه يستخدم معلومات المجال فقط

مثال 1. مقدر النسب

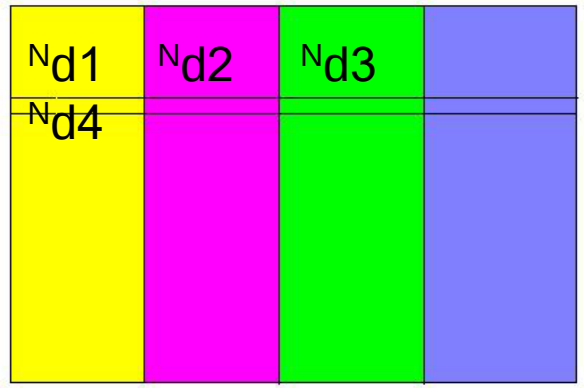
- مقدر HT ل $\bar{X}_d$ $\quad \hat{\bar{X}}_{DIR}^{d} = \frac{1}{N_d}\sum_{i\in s_d} d_i x_i$
- عامل التعديل: $g_d = \frac{\bar{X}_d}{\hat{\bar{X}}_{DIR}^{d}}$
- مقدر النسبة مع المتغير المساعد (الإضافي) X

$$\hat{\bar{Y}}_R^d = \frac{\bar{X}_d}{\hat{\bar{X}}_{DIR}^{d}} \hat{\bar{Y}}_{DIR}^{d} = \frac{1}{N_d}\sum_{i\in s_d} d_i g_d y_i$$

مثال 2. المقدر لمرحلة ما بعد التقسيم الطبقي

- J مرحلة ما بعد تقسيم الطبقي $(j = 1, \ldots, J)$ مقسمة عبر المجالات.
- N_{dj} القيمة المعروفة في تقاطع المجال d و مرحلة ما بعد التقسيم الطبقي j.
- المتوسط الحسابي للمجال d: $\bar{Y}_d = \frac{1}{N_d}\sum_{j=1}^{J} N_{dj}\bar{Y}_{dj}$

$$\hat{\bar{Y}}_{PST}^{d} = \sum_{j=1}^{J} \frac{N_{dj}}{N_d}\hat{\bar{Y}}_{DIR}^{dj}$$



المنطقةd

$^{N}d=^{N}d4d^{N+}3d^{N+}2d^{N+}1$

مثال 3. مقدر الانحدار العام

- نموذج الانحدار الخطي

- $y_j = x_j^T \beta + e_j \;\; E(e_j) = 0, \;\; E(e_j^2) = \sigma^2, j = 1, \ldots, N$

- مقدر الانحدار العام (GREG)

$$\hat{\bar{Y}}^d_{GREG} = \hat{\bar{Y}}^d_{DIR} - (\boldsymbol{X_d} - \hat{\bar{\boldsymbol{X}}}^d_{DIR})\hat{\boldsymbol{B}}_d = \frac{1}{N_d}\sum_{i \in s_d} w_i y_i$$

لننتقل إلى برنامج R

تلخيص المقدرات المباشرة

متطلبات البيانات

- أوزان التصميم المخصصة لوحدات العينات عبر المنطقة المحددة
- مقدر Horvitz-Thompson (HT) : إجمالي عدد الوحدات السكانية المجال Nd))
- مقدر الانحدار العام GREG)): المجاميع السكانية للمتغيرات المساعدة ضمن المجالات المحددة.
- المقدر لمرحلة ما بعد التقسيم: المجاميع السكانية للمتغيرات المساعدة ضمن المجالات المحددة ومرحلة ما بعد التقسيم الطبقي.

تلخيص المقدرات المباشرة

الإيجابيات:

- النهج غير الباراميتري: التحرر من الاعتماد على افتراضات نموذجية محددة
- دمج أوزان أخذ العينات: يسمح بالتصميم التقريبي وعدم التحيز واتساق التصميم مع زيادة حجم العينة (n)
- الإضافة (خاصية القياس المنهجي/المرجعي): توضح الفعالية في المقارنات القياسية.

السلبيات:

- زيادة تباين المقدر ((Y) V) مع انخفاض حجم العينة (n)، مما يجعله غير فعال للغاية بالنسبة للمجالات الصغيرة.
- لا يمكن حسابها للمناطق غير المأخوذة من العينات (Ond = 0).

المقدر غير المباشر

المقدرات الاصطناعية (التركيبية):

- يستخدم مقدر مباشر موثوق به لمنطقة واسعة، يغطي عدة مناطق صغيرة، لاشتقاق مقدر غير مباشر لمنطقة صغيرة.
- ينتج على افتراض أن المناطق الصغيرة لها نفس خصائص المنطقة الواسعة.

المقدرات المركبة:

- مزيج خطي بين المقدر المباشر والمقدر الاصطناعي باستخدام نهج قائم على التصميم أو بافتراض وجود نموذج واضح على مستوى المنطقة أو الوحدة
- يمثل حلا وسطا جيدا من حيث الكفاءة بين خصائص العنصرين

المقدرات التركيبية

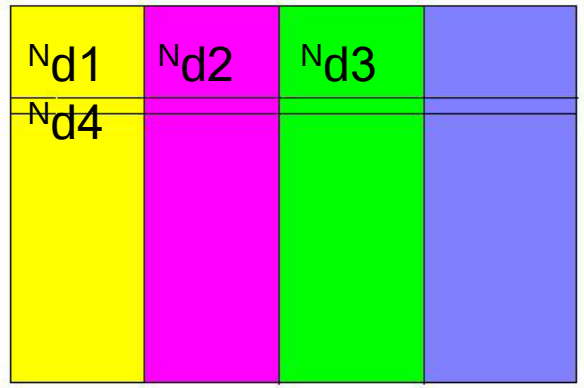
مثال بسيط:

- الهدف: $\bar{Y}_d = \sum_{i=1}^{N_d} y_i / N_d$
- الافتراض: $\bar{Y}_d = \bar{Y}$
- المقدر التركيبي ل $\bar{Y}_d$:

$$\hat{\bar{Y}}_{SYN}^{d} = \frac{1}{N} \sum_{i \in s} d_i y_i$$

المقدرات التركيبية لمرحلة ما بعد التقسيم الطبقي

- J مرحلة ما بعد التقسيم الطبقي $(j = 1, \ldots, J)$ مقسمة عبر المجالات
- N_{dj} القيمة المعروفة في تقاطع المجال d و مرحلة ما بعد التقسيم الطبقي j.
- المتوسط الحسابي للمجال d: $\bar{Y}_d = \frac{1}{N_d} \sum_{j=1}^{J} N_{dj} \bar{Y}_{dj}$
- النموذج الضمني: $\bar{Y}_{dj} = \bar{Y}_j$ لكل من d و j



area d

$^{N}d=^{N}d4d^{N+}3d^{N+}2d^{N+}1$

المقدرات التركيبية لمرحلة ما بعد التقسيم الطبقي

- المقدر التركيبي لمرحلة ما بعد التقسيم الطبقي

$$\hat{\bar{Y}}^{d}_{SYN} = \sum_{j=1}^{J} \frac{N_{dj}}{N_d} \hat{\bar{Y}}^{j}_{DIR}$$

الحاجة إلى:

- التقديرات المباشرة الموثوق بها ل $\hat{\bar{Y}}^{j}_{DIR}$
- التجانس داخل كل طبقة لاحقة.

الخطأ التربيعي المتوسط (MSE) للمقدر التركيبي

- ويعتمد تباين التقديرات التركيبية ، على تباين $\hat{\bar{Y}}_{DIR}^{j}$ أصغر نسبيا مقارنة بتباين المقدر المباشر في المجال.
- تعتمد التقديرات التركيبية على افتراضات قوية وقد تظهر تحيزًا عند انتهاك هذه الافتراضات.
- لذلك، يحتاج إلى تقدير الخطأ التربيعي المتوسط (MSE)، مع مراعاة كل من التحيز والتباين.

لننتقل إلى برنامج R

تلخيص المقدرات التركيبية

الإيجابيات:

- تيسر إعداد التقديرات حتى في المناطق غير المشمولة بالعينات.
- يمكن أن تقلل من التباين في التقديرات المباشرة.
- إنها سهلة التنفيذ.

السلبيات:

- إنها لا تأخذ في الحسبان عدم التجانس بين المناطق، مما يؤدي إلى تحيز كبير.
- يستلزم الافتراض التحقق من صحته
- التصميم المستقر والمخصص للمنطقة يعني أن تقديرات الخطأ التربيعي المتوسط غير متوفرة.
- ولا غنى عن إدخال تعديلات على المعايير المرجعية.

المقدرات المركبة

يُعرف بأنه مزيج خطي من المقدر المباشر والمقدر التركيبي . يهدف هذا النهج إلى موازنة تحيز المقدر التركيبي مع تباين المقدر المباشر داخل مجال معين.

$$\hat{\bar{Y}}_{CE}^{d} = \phi_d \hat{\bar{Y}}_{DIR}^{d} + (1 - \phi_d)\hat{\bar{Y}}_{SYN}^{d}$$

حيث أن:

- $\hat{\bar{Y}}_{DIR}^{d}$ هو المقدر المباشر للمنطقة الصغيرة d
- $\hat{\bar{Y}}_{SYN}^{d}$ مقدر التركيبي للمنطقة الصغيرة d
- ϕ_iهو الوزن المختار بشكل مناسب، مع $0 \leq \phi_d \leq 1$

لننتقل إلى برنامج R

تلخيص المقدرات المركبة

الإيجابيات:

- لا يمكن أن تظهر تباينًا في التصميم أعلى من المقدر المباشر أو تحيزًا أكبر من المقدر الا التركيبي

السلبيات:

- لا يمكن حسابها للمجالات غير المأخوذة من العينات.
- التصميم المستقر والمخصص للمجال متوسط تقديرات الخطأ التربيعي المتوسط غير متوفرة.
- ويلزم إجراء تعديل للمقارنة المرجعية.

الاستدلال القائم على التصميم مقابل الاستدلال القائم على النموذج

قيد وضع النموذج	قيد التصميم	العامل
		الوحدة السكانية
		العينة
		توزيع الاحتمالية
		بارامترات
		المُقدر

تقديرات المناطق الصغيرة – الأساليب القائمة على النماذج

نماذج على مستوى المنطقة

- النماذج محددة على مستوى المنطقة.
- الاعتماد على البيانات على مستوى المنطقة التي يتم الحصول عليها من المسوح، سواء التقديرات المباشرة أو الدقة النسبية، وكذلك التغاير.
- الوصول إلى البيانات أقل تعقيدًا مقارنة بالحصول على بيانات على مستوى الوحدة.

النماذج على مستوى الوحدة

- النماذج محددة على مستوى الوحدة.
- استخدام البيانات على مستوى الوحدة، مثل بيانات المسح، لأغراض توفيق النماذج.
- دمج متغيرات مستوى المنطقة كمتغيرات تنبؤية.
- قد يكون الوصول إلى البيانات على مستوى الوحدة أمرًا صعبًا بسبب مخاوف السرية المحتملة.

النماذج على مستوى المنطقة: نموذج FAY-HERRIOT

1. نموذج العينة

$$\hat{\theta}_d = \theta_d + e_d \quad d = 1, \dots, D$$

$\hat{\theta}_d$ هو مقدر مباشر غير متحيز للتصميم (مثل: HT)
e_d هو الخطأ المعروف في أخذ العينات من المقدر المباشر

2. نموذج الربط

$$\theta_d = \boldsymbol{X}^T\beta + u_d \quad d = 1, \dots, D$$

$u_d \sim N(0, \sigma_u^2)$ مع σ_u^2 غير معروف

3. النموذج المدمج: نموذج مختلط خطي

$$\hat{\theta}_d = \boldsymbol{X}^T\beta + u_d + e_d$$

النماذج على مستوى المنطقة: نموذج FAY-HERRIOT

- يتم الحصول على EBLUPبموجب نموذج Fay-Herriot (FH) عن طريق المعادلة التالية:

$$\hat{\theta}_d^{FH} = \boldsymbol{X}^T\hat{\beta} + \hat{u}_d = \gamma\hat{\theta}_d^{DIR} + (1-\gamma)\,\boldsymbol{X}^T\hat{\beta}$$

- مقدر الخطأ التربيعي المتوسط الرئيسي لمقدر المساحة الصغيرة للمتوسط

$$MSE\left(\hat{\theta}_d^{FH}\right) = g_1 + g_2 + g_3$$

g_1 و g_2 عدم اليقين بشأن BLUP، ومعالجة مكونات التباين كما هو معروف
g_3 عدم اليقين بسبب تقدير عناصر التباين

لننتقل إلى برنامج R

تلخيص نموذج FAY-HERRIOT

الإيجابيات:

- يعتمد فقط على البيانات المساعدة على مستوى المنطقة
- يخصص تلقائيًا وزنًا أكبر لمقدر الانحدار في المناطق ذات أحجام العينات المحدودة ويستخدم المقدر المباشر مع زيادة حجم عينة المجال
- غالبًا ما تظهر كفاءة فائقة مقارنة بالمقدر المباشر
- يعالج عدم التجانس بين المناطق غير المبرر

تلخيص نموذج FAY-HERRIOT

السلبيات:

- هناك فقدان للمعلومات مع تجميع المتغيرات المساعدة
- النموذج مزود بملاحظات D فقط
- فحص النموذج ضروري، وإدخال قضايا الخطية المحتملة للبارامترات غير الخطية.
- من الضروري إجراء تقدير أولي للتباين في أخذ العينات
- لا يمكن تصنيفها حسب المجالات الفرعية
- تحتاج التقديرات إلى تعديل قياسي / منهجي

نماذج على مستوى الوحدة: نموذج BATTESE-HARTER-FULLER

نموذج الآثار العشوائية
(i =individual, d =domain)

$$y_{id} = x_{id}^T \beta + u_d + e_{id} \;\; i = 1, \dots, n \,, \;\; d = 1, \dots, D$$

الآثار العشوائية
$u_d \sim N(0, \sigma_u^2)$
حد الخطأ
$e_{id} \sim N(0, \sigma_e^2)$

المفهوم الأساسي: يُشار إلى هذا النموذج المختلط الخطي على أنه نموذج تقاطع عشوائي، يُسمح للتقاطعات بالاختلاف بين المجالات الصغيرة، في حين أن تأثيرات المتغيرات متساوية لجميع المجالات.

نماذج على مستوى الوحدة: نموذج BATTESE-HARTER-FULLER

- يتم الحصول على EBLUPبموجب نموذج Fay-Herriot (FH) من خلال المعادلة التالية:

$$\hat{\theta}_d^{BHF} = \frac{1}{N_d}\left(\sum_{i\in s} y_{id} + \sum_{i\in r} \hat{y}_{id}\right) = \frac{1}{N_d}\left(\sum_{i\in s} y_{id} + \sum_{i\in r} (x_{id}^T\hat{\beta} + \hat{u}_d)\right)$$

- متوسط تقدير الخطأ التربيعي للمتوسط (MSE)لمقدر المساحة الصغيرة للمتوسط

$$MSE(\hat{\theta}_d^{BHF}) = g_1 + g_2 + g_3$$

g_1 و g_2 عدم اليقين بشأن BLUP، ومعالجة مكونات التباين كما هو معروف
g_3 عدم اليقين بسبب تقدير عناصر التباين

لننتقل إلى برنامج R

Delegation of the European Union to Jordan
TWINNING
دائرة الإحصاءات العامة Department of Statistics
STATISTICS DENMARK
Statistisches Bundesamt
Istat
STATISTICS LITHUANIA STATE DATA AGENCY
Statistics Finland

تلخيص نموذج BATTESE-HARTER-FULLER

الإيجابيات:

- معلومات مساعدة على مستوى الوحدة، تستغل الارتباط بالمتغير المستهدف بفعالية أكبر من البيانات على مستوى المنطقة.
- عادةً ما يكون حجم العينة الإجمالي كبيرًا
- يعين بصورة دينامية وزنًا أعلى لمقدر الانحدار في المناطق ذات أحجام العينات الأصغر، وينتقل إلى المقدر المباشر مع نمو حجم عينة المجال
- يمكن تصنيف التقديرات للمناطق الفرعية، مع تقديم أفكار مفصلة
- يمكن تطبيق المكون الاصطناعي على المناطق غير المأخوذة من العينات، مما يعزز التغطية والشمولية

تلخيص نموذج BATTESE-HARTER-FULLER

السلبيات:

- غالبًا ما يكون الحصول على المعلومات المساعدة على مستوى الوحدة أمرًا صعبًا
- يقتصر على البارامترات الخطية
- لا يتضمن أوزان أخذ العينات
- عرضة للقيم المتطرفة و/أو الانحرافات عن الحالة الطبيعية
- فحص دقيق للنموذج
- وتتطلب التقديرات إجراء تعديل مرجعي لضمان قابلية المقارنة والموثوقية

EU Twinning Project on
Statistics in Jordan

دائرة الإحصاءات العامة
Department of Statistics

شكرًا لحسن استماعكم