



THE WORLD BANK



المسار الفني الجلسة الأولى الاستدلال السببي

داميين دي ووك
عمان 2009

Human Development
Network

South Asia Region

World Bank Institute
Evaluation Group

- أسئلة البحث التي تشكل الدافع وراء معظم الدراسات في العلوم الصحية هي أسئلة سببية بطبيعتها.
- على سبيل المثال:
 - ماهي فعالية دواء ما بالنسبة لمجموعة سكانية؟
 - كم من حالات الوفاة بسبب أحد الأمراض كان يمكن تجنبها بإتباع علاج ما أو سياسة ما؟

□ تتضمن أصعب الأسئلة التجريبية في علم الاقتصاد علاقات الأثر السببي:

- هل تحسن لامركزية المدارس من جودة التعليم؟
- هل يتسبب عام إضافي من التعليم في زيادة الدخل؟
- هل تؤدي التحويلات النقدية المشروطة إلى نتائج صحية أفضل بالنسبة للأطفال؟

□ ما يحفز الإهتمام بهذه المواضيع هو:

■ إهتمامات السياسة العامة

□ هل تقلل البرامج الحكومية من الفقر؟

■ إعتبرات نظرية

■ مشكلات تواجه صناع القرار المنفردين

الإحساس بالمشكلة: مثال افتراضي

- برنامج تحويل نقدي مشروط للنساء الحوامل
 - سيخضعن لفحوص طبية دورية
 - يركز البرنامج على توفير تغذية أفضل
 - عقد اجتماعات دورية بشأن مخاطر التدخين وتناول الكحول
- سؤال رئيسي: هل للبرنامج أي أثر على وزن المواليد عند الولادة؟

الحوالات النقدية (الحوالات النقدية المشروطة) والوزن عند الولادة



□ لنفترض أن بياناتك هي: نساء في بلدة بحوالات نقدية مشروطة ونساء في بلدة قريبة بدون حوالات نقدية مشروطة

متوسط الوزن عند الميلاد بالجرامات				
	النساء في بلدة قريبة مع حوالات نقدية مشروطة	النساء في بلدة قريبة بدون حوالات نقدية مشروطة	الفرق	فرق المضاعف
قبل البرنامج	?	?		
بعد البرنامج	3,250	3,100	150	?

الحوالات النقدية (الحوالات النقدية المشروطة) والوزن عند الولادة

□ بيانات افتراضية: النساء المشاركات بالبرنامج والنساء في أقرب بلدة (بدون البرنامج)

(متوسط) الوزن عند الميلاد بالجرامات

	النساء في البلدة مع حوالات نقدية مشروطة	نساء في بلدة قريبة بدون حوالات نقدية مشروطة	الفرق	الفرق المضاعف
قبل البرنامج	3,025	2,840	185	
بعد البرنامج	3,250	3,100	150	-35

التحليل الإحصائي المعياري

- الأدوات : تقنيات الإحتمالية، وتقنيات تقدير أخرى
- الهدف : استنتاج ابعاد توزيع من العينات المأخوذة من هذا التوزيع
- استخدامات : بمناسبة الأبعاد، يمكن للمرء :
 1. استنتاج العلاقة بين المتغيرات،
 2. تقدير احتمالات الأحداث الماضية والمستقبلية،
 3. تحديث احتمال الأحداث في ضوء أدلة جديدة أو مقاييس جديدة.
- شرط أن يعمل هذا جيدا هو أن ظروف التجربة ينبغي ان تبقى كما هي.

التحليل السببي يذهب خطوة أبعد التحليل الإحصائي المعياري: □

- هدفه هو استخلاص جوانب عملية توليد البيانات.
- وبمساعدة هذه الوجوه المختلفة استنتاج ما يلي :
- احتمالات الحدوث في ظل الظروف الثابتة (كما هو في التحليل الإحصائي المعياري)
- وكذلك ديناميكية الأحداث تحت الظروف المتغيرة

وتشمل "ديناميكية الأحداث في ظل الظروف المتغيرة" : □

1. التنبؤ بتأثير التدخلات
2. التنبؤ بتأثير التغيرات التلقائية
3. التعرف على أسباب الأحداث المنقولة

السببية مقابل الارتباط

- نظرية التحليل الإحصائي المعياري/الإحتمالات
 - كلمة "السبب" لا وجود لها بين مفرداتها
 - تسمح لنا بأن نقول أن حدثين مترابطان بصورة متبادلة أم يعتمد أحدهما على الآخر
 - * إذا وجدنا أحدهما، فلنا أن نتوقع إيجاد الآخر
- هذا لا يكفي لوضعي السياسات
 - يبحثون عن الأسباب الموجبة لقرارات السياسة العامة، إذا قمنا ب
 - س س س، فهل سنحصل على ص ص ص؟
- علينا بالتالي أن نتمم لغة الإحتمالات بمفردات للسببية

نموذج روبن السببي

□ يُرمز للسكان بحرف U . ويُرمز لكل وحدة من وحدات U بحرف u .

□ لكل وحدة $u \in U$ توجد قيمة مرتبطة $Y(u)$ للمتغير المعني Y ، وهو ما نسميه: متغير الاستجابة.

□ ولذا، فإن روبن يتبنى موقفاً يعتبر أن الأسباب ما هي إلا تلك الأشياء التي يمكن أن تكون معالجات في تجارب افتراضية.

□ للتبسيط، نفترض أن هناك سببين فقط أو مستويين للمعالجة.

□ لنفترض أن D متغير يدل على السبب الذي تتعرض له كل وحدة من وحدات U

$$D = \begin{cases} 1 & \text{if unit } u \text{ is exposed to treatment} \\ 0 & \text{if unit } u \text{ is exposed to control} \end{cases}$$

المعادلة:

$D = 1$ إذا تعرضت الوحدة للمعالجة
 $D = 0$ إذا تعرضت الوحدة للمعالجة

في حالة الدراسة الموجهة، يحدد القائم بالتجربة قيمة D . أما في حالة الدراسة غير الموجهة فإن قيمتها تحددها عوامل خارجة عن سيطرته.

□ يحتمل ان تتأثر استجابة Y بما إذا كانت u تتلقى العلاج ام لا.

□ ولذلك، نحتاج إلى متغيرين من متغيرات الاستجابة:

$$Y_1(u), Y_0(u)$$

□ حيث Y_1 هي النتيجة إذا تعرضت الوحدة u للمعالجة

□ و Y_0 هي النتيجة إذا تعرضت u للرقابة

$$D = \begin{cases} 1 & \text{if unit } u \text{ is exposed to treatment} \\ 0 & \text{if unit } u \text{ is exposed to control} \end{cases}$$

المعادلة:

$D = 1$ إذا تعرضت الوحدة للمعالجة
 $D = 0$ إذا تعرضت الوحدة للمعالجة

□ حيث $Y_1(u)$ هي النتيجة إذا تعرضت الوحدة u للمعالجة

□ و $Y_0(u)$ هي النتيجة إذا تعرضت u للرقابة

ويمكن بالتالي كتابة نتيجة كل وحدة كالتالي:

□ $Y(u) = DY_1(u) + (1-D)Y_0(u)$

ملاحظة: يفترض هذا التعريف أن وضع المعالجة لوحدة لا يؤثر على النتائج المحتملة للوحدات الأخرى

□ **تعريف:** بالنسبة لكل وحدة u فإن المعالجة تسبب الأثر

$$\delta_u = Y_1(u) - Y_0(u)$$

□ **مشكلة جوهرية للاستدلال السببي:** يستحيل رصد قيمة $Y_1(u)$ أو $Y_0(u)$ على نفس الوحدة، و ← لذا يستحيل رصد تأثير المعالجة على u ذاتها.

□ **الموضوع:** ليس لدينا الإثبات **المغاير للحقيقة** ل u اي: ماذا كان ليحدث ل u في غياب المعالجة.

□ بالنظر إلى تعذر رصد الأثر السببي بالنسبة لوحدية U واحدة، فإننا نسعى لتحديد متوسط أثر المعالجة للمجموعة السكانية U أو لمجموعات سكان فرعية.

□ متوسط أثر المعالجة على U (أو أي مجموعة سكانية فرعية لـ U):

$$TE_u = \delta_u = Y_1(u) - Y_0(u)$$



$$\begin{aligned} ATE_U &= E_U [Y_1(u) - Y_0(u)] \\ &= E_U [Y_1(u)] - E_U [Y_0(u)] \\ &= \bar{Y}_1 - \bar{Y}_0 \\ &= \bar{\delta} \end{aligned} \quad (1)$$

□ الحل الإحصائي يستبدل بأثر المعالجة المستحيل رصده للسبب t على وحدة معينة U مع وجود متوسط الأثر السببي الممكن تقديره للسبب t على مجموعة من الوحدات U .

□ على الرغم من تعذر حساب كل من $E(Y_0)$ و $E(Y_1)$ ، فإن من الممكن تقديرهما.

□ معظم وسائل القياس الاقتصادي تسعى لأن تستخلص من بيانات الرصد تقديرات متسقة لقيمة كل من:

$$E_U(Y_1) = \bar{Y}_1 \quad \text{and} \quad E_U(Y_0) = \bar{Y}_0$$

□ فحنن إذن نحاول ان نقدر :

$$\begin{aligned}
 ATE_U &= E_U [Y_1(u)] - E_U [Y_0(u)] \\
 &= \bar{Y}_1 - \bar{Y}_0
 \end{aligned} \tag{1}$$

□ لتأمل المعادلة البسيطة التالية لتقدير متوسط الأثر السببي:

$$\hat{\delta} = [\hat{Y}_1 | D = 1] - [\hat{Y}_0 | D = 0] \tag{2}$$

- يلاحظ أن المعادلة (1) تتعلق بالمجموعة السكانية ككل،
- تمثل المعادلة (2) تقديراً ينبغي تقييمه من خلال عينة من السكان.

ليما : إذا افترضنا أن :

Lemma: If we assume that

$$[\bar{Y}_1 | D = 1] = [\bar{Y}_1 | D = 0]$$

and $[\bar{Y}_0 | D = 1] = [\bar{Y}_0 | D = 0]$

then

$$\hat{\delta} = [\hat{Y}_1 | D = 1] - [\hat{Y}_0 | D = 0]$$

هو مقدر ثابت لـ is a consistent estimator of

$$\bar{\delta} = \bar{Y}_1 - \bar{Y}_0$$

المعالجة

π تساوي نسبة السكان الذين انضموا لمجموعة

Proof (to be tried at home)

Let π be the proportion of units that are treated. Then

$$\bar{\delta} = \bar{Y}_1 - \bar{Y}_0$$

$$= \pi \left[(\bar{Y}_1 - \bar{Y}_0) \mid D = 1 \right] + (1 - \pi) \left[(\bar{Y}_1 - \bar{Y}_0) \mid D = 0 \right]$$

$$= \left[\pi [\bar{Y}_1 \mid D = 1] + (1 - \pi) [\bar{Y}_1 \mid D = 0] \right] -$$

$$\left[\pi [\bar{Y}_0 \mid D = 1] + (1 - \pi) [\bar{Y}_0 \mid D = 0] \right]$$

\downarrow if $[\bar{Y}_1 \mid D = 1] = [\bar{Y}_1 \mid D = 0]$ and $[\bar{Y}_0 \mid D = 1] = [\bar{Y}_0 \mid D = 0]$

$$= \left[\pi [\bar{Y}_1 \mid D = 1] + (1 - \pi) [\bar{Y}_1 \mid D = 1] \right] -$$

$$\left[\pi [\bar{Y}_0 \mid D = 0] + (1 - \pi) [\bar{Y}_0 \mid D = 0] \right]$$

$$= [\bar{Y}_1 \mid D = 1] - [\bar{Y}_0 \mid D = 0]$$

وبالتالي الظرف الكافي ليقدر المقدر البسيط متوسط اثر المعالجة هو

$$[\bar{Y}_1 | D = 1] = [\bar{Y}_1 | D = 0]$$

and

$$[\bar{Y}_0 | D = 1] = [\bar{Y}_0 | D = 0]$$

• متوسط النتائج تحت المعالجة هو نفسه للمعالجة (D=1) ومجموعات الرقابة (D=0)

• متوسط النتائج تحت الرقابة هو نفسه للمعالجة (D=1) ومجموعات الرقابة (D=0)

مى تتم تلبية هذه الشروط؟

- يكفي أن تكون مهمة المعالجة D غير مترابطة مع توزيعات النتيجة المحتملة لـ $Y0$ و $Y1$.
- حدسيا: لا يمكن أن يكون هناك ترابط بين
 - ما إذا كان أحدهم يحصل على المعالجة
 - كم يستفيد ذلك الشخص من المعالجة
- أسهل طريقة لتحقيق عدم الترابط هي من خلال الإحالة العشوائية للمعالجة

وطريقة أخرى للنظر الى ذلك هي :

□ بعد إجراء القليل من عمليات الجبر، يمكن إظهار أن:

$$\hat{\delta} = \bar{\delta} + \underbrace{\left([\bar{Y}_0 | D = 1] - [\bar{Y}_0 | D = 0] \right)}_{\text{Baseline Difference}} + (1 - \pi) \underbrace{\left(\bar{\delta}_{\{D=1\}} - \bar{\delta}_{\{D=0\}} \right)}_{\text{Treatment Heterogeneity}}$$

Baseline difference:

فرق خط الأساس (الخط القاعدي)

Treatment Heterogeneity:

تفاوت المعالجة

طريقة أخرى للنظر في ذلك (بالكلمات)

□ هناك مصدري تحيزات يتعين إزالتها من تقديرات الآثار السببية المستمدة من دراسات الملاحظة.

1. اختلاف خط الأساس (التحيز في الاختيار)

2. عدم تجانس المعالجة.

□ معظم الأساليب المتاحة لا تتعامل إلا بالتحيز في الاختيار.

معالجة من تلقوا علاجاً

□ متوسط أثر المعالجة ليس دائماً محور الاهتمام.

□ كثيراً ما يكون متوسط أثر المعالجة للذين عولجوا هو ذو اهتمام جوهري:

$$\begin{aligned} \text{TOT} &= E [Y_1(u) - Y_0(u) | D = 1] \\ &= E [Y_1(u) | D = 1] - E [Y_0(u) | D = 1] \end{aligned}$$

معالجة من تلقوا علاجاً

إذا كنا بحاجة لتقدير TOT

$$TOT = E [Y_1(u) | D = 1] - E [Y_0(u) | D = 1]$$

فيكون المقدر البسيط (2)

$$\hat{\delta} = [\hat{Y}_1 | D = 1] - [\hat{Y}_0 | D = 0]$$

يتم تقدير TOT بثبات إذا :

$$[\bar{Y}_0 | D = 1] = [\bar{Y}_0 | D = 0]$$

لا يوجد اختلاف في خط الأساس بين المجموعات المعالجة ومجموعات الرقابة”

- جوديا بيرل (2000): السببية: النماذج، والمنطق، والاستدلال، 5, 1, Chapters CUP. and 7.
- تريغفي هافيلمو (1944): "المدخل الاحتمالي في القياس الاقتصادي" *Econometrica* 12, pp. iii-vi+1-115.
- آرثر غولدبرغ (1972): "أساليب المعادلات الهيكلية في العلوم الاجتماعية"، *Econometrica* 40, pp. 979-1002.
- دونالد ب. روبن (1974): "تقدير الآثار السببية للمعالجة في التجارب العشوائية وغير العشوائية"، *Journal of Educational Psychology* 66, pp. 688-701.
- بول و. هولاند (1986): "الإحصاءات والاستدلال السببي" *Journal of the American Statistical Association* 81, pp. 945-70, with discussion.