Evaluation de projets de développement

Lire les travaux d’évaluation d’impact

Björn Nilsson

bjorn.nilsson@universite-paris-saclay.fr

Faculté Jean Monnet
Université Paris-Saclay

M2 GPDS, 1er semestre
Année universitaire 2024-2025

Présentation à faire pour janvier

- Vous aurez à faire une présentation en binôme pour le mois de janvier.
  - Il s'agira de présenter de manière critique un article de recherche.
  - L'article utilisera une méthode d'évaluation d'impact pour répondre à une question liée à la lutte contre la pauvreté.
- 20 minutes (+ 10 minutes de discussion)
- Ce qui est attendu:
  - Présentation de la motivation, de la question de recherche qui est posée, et de la stratégie mise en oeuvre pour y répondre.
  - Veillez particulièrement à identifier les hypothèses derrière le résultat obtenu et de discuter leur crédibilité.
  - Discutez les résultats (portée et limites)

Plan du cours

universite PARIS-SACLAY

- Séance 1 : Introduction
- Séance 2 : Mesurer la pauvreté
- **Séance 3 : Lire des travaux d'évaluation d'impact**
- Séance 4 : La différence de différence
- Séance 5 : La régression par discontinuité
- Séance 6 : Les variables instrumentales
- Séance 7 : Les essais aléatoires randomisés
- Séance 8 : Présentations d'articles
- Séance 9 : Présentations d'articles
- Séance 10 : Examen

Plan du chapitre

Concepts clés

Pourquoi évaluer ?

- L'évaluation de politiques publiques répond à des enjeux de redevabilité et d'allocation des ressources.
- Ils permettraient aussi de justifier des politiques et les inscrire au-delà des périodes électorales.
    - Dorsch & Mareek (2021): augmentation des RCT en Inde dans les élections serrées.
    - Pour 262 lois votées en France entre 2008 et 2020, 18 travaux évaluatifs sont cités en moyenne avant le vote (Baïz, 2022).
- Qu'est-ce qui constitue une bonne évaluation de politiques publiques ?
    - Attribution : est-ce que la politique mise en place est la cause de ce que j'observe ?

Les résultats potentiels

- Comment penser l'effet que l'on cherche à mesure ? Sur qui/quoi ? Par rapport à qui/quoi ?
  - Besoin de formalisme.
  - $\Rightarrow$ "Potential outcomes framework" de **Neyman-Rubin**.
- Nous avons déjà vu que l'idéal pour l'évaluation est de pouvoir observer deux univers : l'univers réel où une politique a été mise en place ; et un univers parallèle où elle ne l'a pas été.
- Le cadre conceptuel des "résultats potentiels" est basée sur cette idée.
  - Permet de définir différents effets que l'on rencontre dans la littérature : effet moyen du traitement, effet moyen sur les traités, sur les non-traité, effet moyen local du traitement, intention de traiter, etc...

Les résultats potentiels

Les différents effets

- Supposons qu'une population est composée en personnes recevant un traitement (T=1), et personnes ne le recevant pas (T=0).
- Le résultat d'un individu (outcome) s'écrit Y(1) s'il est traité, et Y(0) s'il ne l'est pas.
  - Tel que dans le monde réel, on n'observera donc jamais Y(1) et Y(0) pour une même personne.
  - Dans ce monde réel, soit Y(1), soit Y(0) est observé, et l'autre est le **contrefactuel**.
- Pour un être omniscient, l'évaluation est facile ! L'effet d'un programme est simplement **Y(1) - Y(0)**.

Les résultats potentiels

Les différents effets

Individu	T	Y(0)	Y(1)	Effet
1	1	0	7	7
2	0	3	0	-3
3	0	2	2	0
4	1	1	8	7
5	0	4	1	-3
6	1	3	10	7

- Supposons qu'on soit omniscient. Le tableau nous donne les effets de traitement individuels.
- Ceci nous permet de définir plusieurs paramètres à estimer.

Les résultats potentiels

Les différents effets

Individu	T	Y(0)	Y(1)	Effet
1	1	0	7	7
2	0	3	0	-3
3	0	2	2	0
4	1	1	8	7
5	0	4	1	-3
6	1	3	10	7

- Average Treatment Effect (**ATE**) : l'effet de traitement moyen dans la population.
- ATE : $\frac{7-3+0+7-3+7}{6} = 2,5$

Les résultats potentiels

Les différents effets

Individu	T	Y(0)	Y(1)	Effet
1	1	0	7	7
2	0	3	0	-3
3	0	2	2	0
4	1	1	8	7
5	0	4	1	-3
6	1	3	10	7

- Average Treatment Effect on the Treated (**ATT**) : l'effet de traitement moyen sur les traités.
- ATT : $\frac{7+7+7}{3} = 7$

Les résultats potentiels

Les différents effets

Individu	T	Y(0)	Y(1)	Effet
1	1	0	7	7
2	0	3	0	-3
3	0	2	2	0
4	1	1	8	7
5	0	4	1	-3
6	1	3	10	7

- Average Treatment Effect on Control (**ATC**) : l'effet de traitement moyen sur les non-traités.
- ATC : $\frac{-3+0-3}{3} = -2$

Les résultats potentiels

Les différents effets

- Hélas, nous ne sommes pas omniscients, et observons plutôt :

Individu	T	Y(0)	Y(1)	Effet
1	1		7	
2	0	3		
3	0	2		
4	1		8	
5	0	4		
6	1		10	

- Comment récupérer ATE, ATT et ATC dans ces conditions ?
- On voit bien que $E[Y(1)] - E[Y(0)] = \frac{7+8+10}{3} - \frac{3+2+4}{3}$ =**5,33** donne ni ATE, ni ATT, ni ATC.
- Si cette expression est proposée comme estimateur d'ATE, ATT ou ATC, l'estimateur est **biaisé**.

Les résultats potentiels

Les différents effets

- C'est ici que les techniques expérimentales ou quasi-expérimentales rentrent en jeu.
  - En exploitant une variation *exogène* dans l'accès au traitement, elles arrivent à contourner le **problème de sélection**.
- La méthode DiD estime en général un effet **ATT**.
- Les RDD et les variables instrumentales fournissent un effet moyen local (**LATE**).
- Les RCT fournissent un effet moyen **ATE**, identique à l'**ATT** sous certaines conditions.

La régression pour débutants

L'ensemble de ces méthodes s'appuient sur de la **régression**.

- Le terme vient de Francis Galton (regression towards the mean), pour décrire l'évolution des tailles dans la population.
- Il a ensuite pris un sens statistique avec les travaux de Karl Pearson et Udny Yule.
- En pratique: on choisit un modèle dont on veut estimer des paramètres. Ces paramètres sont souvent notés par des lettres de l'*alphabet grec*.
- On choisit ensuite une **méthode** de régression.

La régression pour débutants

universite PARIS-SACLAY

Exemple du rendement de l'éducation. L'équation de Mincer postule la relation suivante entre log-salaire et scolarisation :

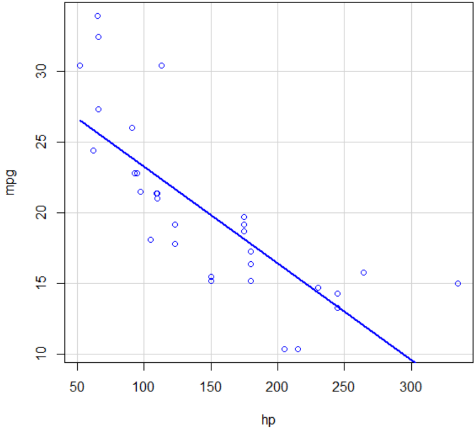
$$\ln W(S) = \ln W(0) + rS \tag{1}$$

- Le modèle statistique équivalent est: $W_i = \alpha + \beta S_i + \epsilon_i$
- α correspond à $\ln W_0$, β correspond à r, le rendement de l'éducation.
- ϵ_i est là car on ne peut pas supposer que cette relation est parfaite.
- La régression vise surtout à déterminer la valeur de β, l'effet moyen d'une année d'éducation supplémentaire sur le salaire.

La régression pour débutants

Les moindres carrés ordinaires

On peut (dans le cadre d'une régression univariée) voir la régression comme un problème graphique :



La régression pour débutants

Les moindres carrés ordinaires

- Comment trouver la ligne qui décrit le mieux la relation entre X et Y ?
- Les **moindres carrés ordinaires** (MCO) : on minimise la somme des carrés des erreurs.
    - Parmi toutes les valeurs possibles de β, on prend celle qui minimise la somme des carrés des distances entre la ligne et les points.
- Cette valeur est égale à la pente de la **droite de régression**.
    - Sur l'exemple précédent, on a donc une valeur de β qui est négative, et qui (à vue d'oeil) a l'air d'être proche de -0.07.
    - Autrement dit, quand la variable X augmente de 1, la variable Y diminue de 0.07.
    - Le graphique montre aussi l'utilité du terme α qui permet à la droite de régression de ne pas passer par 0.

La régression pour débutants

Les moindres carrés ordinaires

Il existe d'autres méthodes de régression dont vous pourrez rencontrer les noms :

- Le maximum de vraisemblance (Maximum likelihood ou ML).
- La méthode des moments (Method of Moments).
- Ces méthodes sont plutôt interchangeables, produisant des résultats identiques ou du moins très très proches.
- MCO/OLS : convient pour des **modèles linéaires**.
    - D'autres modèles conviennent quand la réponse est distribuée de manière non-linéaire (Poisson, Logistic, Probit, etc.)

La régression pour débutants

OK, mais encore ?

Les différentes méthodes aboutissent donc à une valeur estimée pour β. Cette estimation est appelée **coefficient**.

- Pour l'exemple précédent, on dira que le coefficient (β) du nombre d'années de scolarisation sur le salaire est **-0.07**.
- Dans les travaux d'évaluation d'impact, le coefficient estimé correspondra aux différents effets que nous avons vus : à l'**ATE**, l'**ATT**, le **LATE** etc.
- On verra plus tard ce que chaque méthode estime comme effet.
- Pour le moment, regardons un **tableau de régression**.

La régression pour débutants

OK, mais encore ?

- Pour l'exemple précédent, on dira que le coefficient (β) du nombre d'années de scolarisation sur le salaire est **-0.07**.
- Dans les travaux d'évaluation d'impact, le coefficient estimé correspondra aux différents effets que nous avons vus : à l'**ATE**, l'**ATT**, le **LATE** etc.
- On verra plus tard ce que chaque méthode estime comme effet.
- Pour le moment, regardons un **tableau de régression**.

La régression pour débutants

Exemple de Jayachandran & Pentha (2017)

université PARIS-SACLAY

TABLE 2—INDIA'S DIFFERENTIAL BIRTH ORDER GRADIENT IN CHILD HEIGHT AND RELATED OUTCOMES

	HFA z-score					Stunted	WFA z-score	Hb level	Deceased
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
India	−0.082	0.092							
	[0.011]	[0.018]							
India × 2nd child		−0.144	−0.161	−0.110	−0.243	0.051	−0.146	−0.094	0.003
		[0.025]	[0.027]	[0.063]	[0.048]	[0.007]	[0.020]	[0.030]	[0.004]
India × 3rd+ child		−0.377	−0.227	−0.193	−0.436	0.064	−0.198	−0.159	0.002
		[0.024]	[0.032]	[0.092]	[0.085]	[0.009]	[0.024]	[0.036]	[0.004]
2nd child		0.023	−0.011	−0.097	−0.167	0.009	0.009	−0.011	−0.014
		[0.015]	[0.017]	[0.053]	[0.027]	[0.004]	[0.012]	[0.022]	[0.002]
3rd+ child		−0.066	−0.118	−0.169	−0.334	0.036	−0.063	−0.037	−0.011
		[0.013]	[0.019]	[0.074]	[0.044]	[0.005]	[0.014]	[0.025]	[0.003]
Africa mean of outcome	−1.351	−1.351	−1.351	−1.351	−1.351	0.375	−0.877	10.150	0.071
Child's age dummies × India	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mother's literacy × India	No	No	Yes	Yes	No	Yes	Yes	Yes	Yes
Mother's age at birth × India	No	No	Yes	Yes	No	Yes	Yes	Yes	Yes
PSU fixed effects	No	No	Yes	Yes	No	Yes	Yes	Yes	Yes
Mother fixed effects	No	No	No	No	Yes	No	No	No	No
Completed fertility sample	No	No	No	Yes	No	No	No	No	No
Observations	168,108	168,108	167,737	66,566	83,228	167,737	167,737	88,838	199,514

La régression pour débutants

Exemple de Jayachandran & Pentha (2017)

- Ce tableau est issu d'un article qui étudie les retards de croissance des enfants indiens versus les enfants africains.
- La première colonne donne le nom des différentes variables.
- Les colonnes d'après correspondent aux coefficients estimés via différentes **spécifications économétriques**.
  - Une spécification: un modèle statistique, e.g. $Y_i = \alpha + \beta_0 X_i + \beta_1 Z_i + \epsilon$.
  - Souvent, pour être convaincant, on veut des résultats robustes à une variété de spécifications.

La régression pour débutants

Exemple de Jayachandran & Pentha (2017)

- Que disent ces résultats ?
- Commençons par la première colonne. La spécification y est simple : $Z = \alpha + \beta \times India + \epsilon$
- Le z-score (Z) est une mesure de la taille comparée à la taille 'normale' d'un enfant de son âge.
- Le coefficient β de la variable **India** vaut -**0.082**.
  - Autrement dit, un enfant d'Inde a en moyenne une valeur inférieure de Z qu'un enfant africain, de 0.082.
  - Est-ce beaucoup? Est-ce significatif ?

La régression pour débutants

Significativité d'un coefficient

- Que veut-dire **significatif** dans ce contexte ?
- La plupart des études sont faites sur des données d'enquête statistique.
- Une enquête statistique interroge un sous-ensemble de la population, **aléatoirement choisi**.
  - Mais alors, comment être sûr que l'échantillon sur lequel on a fait l'enquête est une bonne approximation de la population ?
  - Si j'estime une relation en utilisant mon échantillon, comment savoir si cette même relation tient en population générale ?
  - Je ne peux jamais l'affirmer à 100%, mais la statistique nous permet de faire une affirmation probabiliste.
  - De type, "avec 95% de certitude, le coefficient de la population est supérieur à 0".

La régression pour débutants

Erreur-type

- Il y a plusieurs façons d'indiquer la significativité dans un tableau de régression.
- Cela relève de préférences — l'information communiquée est exactement la même, et connaissant une des valeurs citées, on peut retrouver les autres.
- Jayachandra & Pandhe utilisent ici **l'erreur-type** (standard error).
  - Elle est renseignée entre crochets en dessous de chaque coefficients.
  - Qu'est-ce que ça veut dire ? C'est l'équivalent de l'écart-type du vrai paramètre dans la population.
  - Si en s'écartant de 2 erreurs-types du coefficients, vers le haut ou vers le bas, on ne rencontre pas 0, alors le coefficient est dit significatif au seuil de **95%**.

La régression pour débutants

Intervalle de confiance

université PARIS-SACLAY

- On voit aussi souvent un **intervalle de confiance**.
  - Le vrai paramètre dans la population a 95% de chance de se situer dans cet intervalle de confiance.
- Lié à l'erreur-type:
- Intervalle de confiance: $[\beta$-1.96 x σ,β + 1.96 x $\sigma]$
  - Si l'intervalle de confiance ne contient pas **0**, alors le coefficient est significatif au seuil de 95%.
  - La valeur 1.96 (≈ 2) est
- Calculons l'intervalle de confiance pour le premier coefficient de Jayachandra & Panche :
  - 1.96 x 0.011 = 0.022, qu'il faut ajouter/supprimer à 0.082.
  - $[-0.104, -0.060]$

La régression pour débutants

Le T de Student

universite PARIS-SACLAY

- D'où vient ce 1.96 ?
- C'est en fait une valeur critique de la distribution de Student, correspondant à une probabilité de 95%.
- L'erreur-type (qui, pour rappel on ne peut pas observer dans la population) est distribuée selon cette loi.
- 1.96 nous dit alors que 95% de la distribution du coefficient est située à moins de 1.96 erreurs-types du coefficient estimé.
- Parfois, le **t de Student** est indiqué dans un tableau de régression. S'il est supérieur à 1.96 pour une variable, cela veut dire que la variable est significative à 95%.

La régression pour débutants

Le T de Student

- A chaque valeur du t de Student correspond donc une probabilité. Ainsi, pour 1,645 la probabilité est de 90%.
  - Autrement dit, 90% de la distribution du coefficient dans la population se trouve à moins de 1,645 erreurs-types du coefficient estimé.
- Souvent la **P-valeur** est présentée.
- C'est **1 - la probabilité associée au t de Student**.
- Ainsi, pour une valeur du **t** de 1.96, la P-valeur est de 0.05.
- Une P-valeur inférieure à 0.05 veut dire que le coefficient est significatif au seuil de 95%.

La régression pour débutants

Les astérisques

- Mais l'indicateur le plus courant et le plus facile à analyser est l'**astérisque** !
  - Il s'agit d'indiquer les coefficients significatifs à l'aide de symboles, souvent des astérisques.
- On définit les astérisques comme on veut, mais souvent la convention suivante est adoptée :
- *** significatif à $<1\%$; ** significatif à $<5\%$; * significatif à $<10\%$.

La régression pour débutants

Exemple d'output de Stata

Source	SS	df	MS
Model	.031310582	1	.031310582
Residual	7.89843695	909	.00868915
Total	7.92974753	910	.008714008

Number of obs	=	911
F(1, 909)	=	3.60
Prob > F	=	0.0580
R-squared	=	0.0039
Adj R-squared	=	0.0029
Root MSE	=	.09322

beat	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
age	-.0006501	.0003425	-1.90	0.058	-.0013222	.000022
_cons	.031385	.0123014	2.55	0.011	.0072425	.0555275

Ici j'ai essayé d'expliquer le fait de penser qu'il est légitime qu'un mari batte sa femme si elle sort sans lui dire par l'âge de l'enquêtée, auprès d'un échantillon de 911 femmes d'Abidan.

La régression pour débutants

Wooclap

Wooclap
https://app.wooclap.com/events/UQOCNA/

La régression pour débutants

Quelques mises en garde

- Un coefficient peut (presque) toujours être estimé.
- On peut sans problème estimer le rendement de l'éducation sur un échantillon avec trois individus.
  - $\Rightarrow$ la présence d'un coefficient ne garantit pas qu'il soit précis, ni qu'il soit sans biais.
- Vous savez maintenant comment juger de la significativité d'un coefficient.
- Cela suffit-il à évaluer la qualité d'une évaluation ?
  - Non! La significativité ne dit rien en soi de la qualité de l'évaluation. En revanche, il faut que le discours de l'auteur soit cohérent avec les tableaux.

La régression pour débutants

Quelques mises en garde

- Pour chaque méthode d'évaluation d'impact, le contrefactuel est construit sur la base d'hypothèses sur le comportement inobservé du groupe de contrôle.
- Il n'y a la plupart du temps **pas de test formel** permettant de valider ces hypothèses.
  - Pour convaincre le lecteur, l'évaluateur a recours a des indices semblant aller dans le sens de son interprétation des choses.
  - Parfois, il prend les devants, identifie une menace pour l'évaluation, et montre par le raisonnement ou en mobilisant d'autres sources de données que cette menace n'est probablement pas importante.

La régression pour débutants

Quelques mises en garde : exemple

- Dans une étude sur l'impact de la taille des familles sur le bien-être, un chercheur mobilise le sexe des deux premiers enfants comme instrument de la taille de la fratrie.
  - Il conclut : faire des enfants rend heureux !
- Vous lui faites remarquer que :
  - L'**ordre sexuel de naissance** n'est pas le même dans les deux groupes; dans les traités les deux premiers enfants sont toujours du même sexe, cela peut peut-être avoir une incidence directe sur le bonheur ?
  - Il répond en mobilisant des statistiques sur les ménages avec deux enfants. Il y montre qu'il n'y a pas de différence de bonheur entre les ménages avec deux enfants de même sexe et les ménages avec deux enfants de sexe opposés.

La régression pour débutants

Quelques mises en garde : exemple

- Vous lui dites ensuite : mais quand même, c'est pas tout le monde qui a les moyens d'avoir un troisième enfant. Les avantages fiscaux du troisième enfant s'appliquent aux familles relativement aisées, donc les plus modestes devraient moins réagir à deux enfants de même sexe! Ca, quand même, ça va fausser les résultats !
  - Le chercheur rétorque que certes, son estimation est une estimation LATE: local average treatment effect. Elle évalue l'effet sur un groupe particulier, ceux qui ont un troisième enfant **parce que** les deux premiers étaient de même sexe. Ce groupe n'est pas forcément similaire au ménage moyen dans la société.
  - Il mobilise d'autres statistiques, montrant que l'effet est présent pour tous les ménages, mais qu'il est plus fort chez les ménages aisés.
- $\Rightarrow$ Qu'est-ce qu'on apprend sur l'effet général?
- Il faudrait creuser davantage, peut-être distinguer les ménages selon leur niveau de revenu, si les données le permettent.