

Analyse de l'impact de la connaissance du statut VIH sur le comportement sexuel

MOOC – Introduction aux méthodes d'évaluation des politiques publiques.

Dans cet exercice, il vous est demandé d'analyser les liens entre connaissance du statut VIH et comportement sexuel à partir des données provenant d'une évaluation d'impact à assignation aléatoire qui a déjà été présentée dans le cas pratique RCT. Ces données proviennent de l'article de Rebecca Thornton "The Demand and Impact of Learning HIV Status" (American Economic Review, 2008). Dans cet exercice nous mettrons en œuvre l'approche par les variables instrumentales pour répondre à la question de l'impact de la connaissance du statut HIV sur l'achat de préservatifs.

Informations générales :

Dans la seconde partie de son article de 2008, Rebecca Thornton évalue si la connaissance du statut VIH influence le comportement sexuel. Répondre à cette question à partir de données d'observation pose problème dans la mesure où les individus qui ont connaissance de leur statut HIV sont ceux qui ont fait la démarche d'aller passer un test de dépistage du VIH. Dès lors, on peut s'attendre à ce qu'ils aient des caractéristiques différentes de ceux qui n'ont pas entrepris cette démarche et, en particulier, un comportement sexuel différent. On peut imaginer par exemple que ceux qui ont fait le test de dépistage ont en moyenne un comportement sexuel plus risqué. Cet effet de sélection est susceptible de biaiser les résultats.

Variables clés :

- Variable de résultat :
 - *bought* = 1 si la personne a acheté des préservatifs
- Variable de traitement :
 - *got* = 1 si la personne est allée récupérer ses résultats de test VIH¹
- Variable instrumentale :
 - *any* = 1 si la personne a reçu une incitation monétaire pour aller chercher ses résultats de test VIH
- Autre variable clé :
 - *hiv2004* = 1 si la personne est séropositive.

Par la suite, on appellera « groupe de traitement » les individus qui ont connaissance de leur statut sérologique (*got*=1). Les autres seront appelés « groupe de contrôle » (*got*=0).

¹ C'était la variable de résultat dans le cas pratique RCT.

Première partie : Statistiques descriptives

Cette partie vous permet de décrire votre échantillon. Quelle est l'importance du problème considéré dans votre échantillon (statut sérologique, achat de préservatifs) ? Quelles sont les caractéristiques des individus de l'échantillon (en retenant notamment des variables considérées comme importantes par rapport au phénomène étudié : âge, genre, éducation, activité sexuelle, richesse, etc.) ?

1. Présentez les statistiques descriptives de la variable de résultat (*bought*).
2. Présentez les statistiques descriptives des variables âge, genre, éducation, activité sexuelle, etc. Commentez vos résultats.
3. Présentez séparément ces statistiques pour le groupe de traitement (*got=1*) et le groupe de contrôle (*got=0*). Pour chaque variable, faites un test statistique pour évaluer s'il existe des différences significatives (au seuil de 5%) entre les groupes de traitement et de contrôle. Commentez vos résultats. En particulier, si vous trouvez des différences significatives, indiquez en quoi est-ce que cela peut affecter les résultats de l'évaluation d'impact.

Deuxième partie : Analyse graphique

Des graphiques simples peuvent permettre une première visualisation des résultats. C'est également la façon la plus adéquate de communiquer vos résultats auprès de leurs utilisateurs éventuels (gouvernements, bailleurs de fonds, ONG etc.).

1. Faites un histogramme, où figurent en abscisse les catégories « traitement » et « contrôle » et, en ordonnée, le pourcentage des individus de chacun de ces groupes qui a acheté des préservatifs. Commentez vos résultats.
2. Refaites le même graphique, mais en séparant l'échantillon selon le statut sérologique. Commentez vos résultats.

Troisième partie : Analyse MCO

De simples régressions MCO sont le moyen le plus couramment utilisé pour estimer l'impact d'une intervention. Si la variable dépendante est binaire, certains choisissent d'utiliser les MCO (modèle de probabilité linéaire), quand d'autres préfèrent utiliser des modèles de type Logit ou Probit².

1. Faites une première estimation de l'effet de connaître son statut sérologique (variable *got*) sur la probabilité d'avoir acheté des préservatifs (variable *bought*). Commentez vos résultats du point de vue de la significativité statistique mais également de l'interprétation du niveau du coefficient.
2. Refaites la même estimation, mais cette fois en ajoutant des variables de contrôle comme l'âge, l'éducation, ou le genre. Comment cela affecte-t-il vos résultats ?

² Les résultats sont similaires entre les deux méthodes. Les MCO sont en général préférés car les coefficients obtenus sont plus directement interprétables en termes d'effet marginal.

Quatrième partie : Analyse d'impact VI

Afin de redresser le biais de sélection, nous allons exploiter le caractère aléatoire de l'attribution d'une incitation monétaire à aller chercher ses résultats et donc à connaître son statut sérologique. Le caractère aléatoire de cette attribution fait de la variable *any* une variable instrumentale dont le caractère exogène est indiscutable. Nous allons tout d'abord procéder en deux étapes puis en une étape avec la commande *ivreg2* de Stata.

1. Approche en 2 étapes :
 - a. Première étape : Faites une estimation de l'effet d'avoir obtenu une incitation monétaire (variable *any*) sur la probabilité d'être allé chercher ses résultats (variable *got*). Commentez vos résultats du point de vue de la significativité statistique mais également de l'interprétation du niveau du coefficient. Calculez la variable prédite (*got_hat*) par cette équation de première étape avec la commande *predict* de Stata.
 - b. Deuxième étape : Faites une estimation de l'effet de la variable prédite d'obtenir une incitation monétaire (variable *got_hat*) sur la probabilité d'avoir acheté des préservatifs (variable *bought*). Commentez vos résultats du point de vue de la significativité statistique mais également de l'interprétation du niveau du coefficient.
 - c. Comparez vos résultats avec ceux obtenus par les MCO.
2. Refaites la même estimation mais en une seule étape avec la commande *ivreg2* de Stata.
 - a. Comment cela affecte-t-il vos résultats par rapport à l'approche en 2 étapes?
 - b. Comparez vos résultats avec ceux obtenus par les MCO.

Cinquième partie : Hétérogénéité de l'impact

Si l'impact moyen est un résultat important, il se peut qu'il y ait de fortes disparités d'impact entre différentes catégories de la population. Cette hétérogénéité peut être déterminante au moment de définir le ciblage de programmes de développement.

1. Y a-t-il un effet différent entre les hommes et les femmes ? Estimez la relation entre le fait de connaître son statut sérologique et le fait d'avoir acheté des préservatifs pour les 2 groupes. Commentez vos résultats.
2. Y a-t-il un effet différent entre les individus selon leur statut sérologique en 2004? Estimez la relation entre le fait de connaître son statut sérologique et le fait d'avoir acheté des préservatifs pour les 2 groupes. Commentez vos résultats.