

Application 1 (Python) : évaluation d'impact à assignation aléatoire (randomisée)

Cette application consiste à analyser les données provenant d'une évaluation d'impact à assignation aléatoire (randomisée).

Les données associées à cette application se trouvent sur e-campus. Elles proviennent de l'article de Rebecca Thornton "*The Demand and Impact of Learning HIV Status*" (*American Economic Review*, 2008).

Les questions suivantes sont des étapes standards pour l'analyse de données d'évaluations randomisées.

Informations générales :

Le but est d'évaluer si une baisse des coûts liés à l'obtention des résultats de tests VIH amène à une hausse du nombre de personnes qui cherchent à obtenir leurs résultats. Il y a deux interventions :

- 1) un paiement en cash aux personnes qui viennent chercher leurs résultats,
- 2) une variation dans le montant de l'incitation monétaire promise aux personnes qui viennent chercher leurs résultats.

Chacune des interventions a été allouée de manière aléatoire au niveau individuel.

Variable clés :

any =1 si la personne a été sélectionnée (aléatoirement) pour recevoir une incitation monétaire afin qu'elle aille chercher ses résultats de test.

Ti : Niveau (aléatoirement déterminé) de l'incitation monétaire reçu par la personne si elle va chercher ses résultats. A l'époque, le taux de change entre le Kwacha (Malawi) et le dollar Américain était de 107 Kwacha pour 1 dollar.

got =1 si la personne est allée chercher ses résultats de tests après l'intervention. C'est la principale variable de résultat

Le tableau suivant donne le descriptif de l'ensemble des variables de la base de données.

Variables	Description
got	est allé cherché son résultat - endline
Ti	valeur de l'incitation monétaire reçue
any	a reçu une incitation monétaire
age	age en années révolues - baseline
male	=1 si homme, 0 si femme - baseline

educ2004	niveau d'étude atteint, en années - baseline
hadsex12	A eu des rapports sexuels au cours des 12 derniers mois - baseline
eversex	A déjà eu des rapports sexuels - baseline
tb	a déjà été cherché un résultat de test VIH - baseline
land2004	possède une terre agricole - baseline
hiv2004	résultat de test HIV - baseline

Par la suite, on appellera « groupe de traitement » les individus qui ont reçu une incitation ($any=1$). Les autres seront appelés « groupe de contrôle ».

Note : dans les variables présentes dans la base de données, le terme « *baseline* » ou « 2004 » indique que la mesure a été faite avant l'intervention, et « *endline* » ou « *final* » indique qu'elle a été faite après.

I. Première partie - Statistiques descriptives

Cette partie vous permet de décrire votre échantillon. Quelle est l'importance du problème considéré dans votre échantillon (prévalence HIV) ? Quels types d'individus sont dans cet échantillon (en considérant notamment des variables considérées comme importantes par rapport au phénomène étudié (âge, genre, éducation, activité sexuelle, richesse, etc.) ? Est-ce que l'allocation aléatoire du traitement a bien produit deux groupes identiques (en moyenne) par rapport à ces variables ?

1. Présentez les statistiques descriptives des variables âge, genre, éducation, activité sexuelle, dépistage avant l'intervention, statut VIH avant l'intervention etc. Commentez vos résultats.
2. Présentez séparément ces résultats pour le groupe de traitement et le groupe de contrôle. 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.

II. 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 figure en abscisses les catégories « traitement » et « contrôle », et en ordonnées le pourcentage des individus de chacun de ces groupes qui est allé chercher ses résultats après l'intervention. Commentez vos résultats.
2. Refaites le même graphique, mais cette fois les catégories en abscisses sont les différents montants d'incitation monétaires. Commentez vos résultats.

III. Troisième partie - Analyse sur la base de régressions linéaires

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 néanmoins d'utiliser les MCO (modèle de probabilité linéaire), quand d'autres préfèrent utiliser des modèles de type Logit ou Probit. Vous avez le choix d'utiliser l'estimateur que vous jugez le plus approprié. Cependant, si les résultats sont similaires de ceux obtenus avec les MCO, ces derniers sont en général préférés car les coefficients sont plus directement interprétables.

1. Faites une première estimation de l'effet d'avoir reçu une incitation monétaire (quel que soit son montant) sur la probabilité d'être allé chercher son résultat de test (variable « got »). 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 ou le genre. Comment cela a-t-il affecté vos résultats ?
3. Refaites la même estimation, mais cette fois en utilisant le niveau de l'incitation à la place de la variable binaire. Pour plus de précision, refaites une fois encore cette estimation, en utilisant le caractère catégoriel de cette variable. Commentez vos résultats.
4. Trouvez-vous que l'incitation monétaire a un effet important ? Y a-t-il une grande différence entre une incitation monétaire de 1\$ et de 2\$?

IV. Quatrième 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.

Y a-t-il un effet différent entre les hommes et les femmes ?

Note : cette application est tirée d'un MOOC de Luc Behaghel, Anne-Sophie Robilliard et Philippe de Vreyer