Rappel TD_7

Marion Brouard, Pauline Leveneur

November 20, 2025

- **Objectif du rappel:**
  1. Comment tester si on se trouve dans le cadre d'une expérience randomisée ?
  2. Faire un point sur les "confounders"
  3. Revoir quelques propriétés calculatoires

1 Point cours

1.1 Tester si l'expérience est aléatoire

Lorsqu'une expérience est randomisée, le biais de sélection (B) est nul (i.e. le groupe de contrôle et le groupe de traitement sont similaires.)

$$B = E(Y(0)|D = 1) - E(Y(0)|D = 0) \tag{1}$$

(Rappel: Le biais de sélection est nul lorsque $Cov(Y(0), D) = 0$ *(cf section 2 pour le calcul)*)

Cependant, en pratique, il est impossible de tester directement si le biais de sélection est nul puisqu'on n'observe jamais Y(0)|D=1. Il est toutefois possible de se faire une idée en testant si les individus dans le groupe de contrôle et le groupe de traitement sont similaires en testant la similitude de leurs caractéristiques observables. L'idée est de tester si D a, ou non, un effet sur un set de variables X, elles-mêmes corrélées avec Y(0) (Contrairement à Cov(Y(0),D), on peut tester Cov(X,D)!).

Il existe deux types de caractéristiques observables:

- Les caractéristiques qui varient avec les temps. (ex: le salaire)
- Les caractéristiques fixes. (ex: le sexe)

Il s'agit de comparer les caractéristiques entre groupe de contrôle et groupe de traitement avant la date du traitement.

Voici les trois principaux tests:

1. **Test de Kolmogorov-Smirnov:** On cherche à tester si la distribution des variables continues de X sont similaires dans les deux groupes.
   $\Rightarrow H_0^{KS}$: La variable considérée est identiquement distribuée dans les deux groupes.

   Exemple STATA:
   On réalise un test Kolmogorov-Smirnov sur la variable "age".

Figure 1: Kolmogorov Smirnov, STATA

Smaller group	D	P-value
0:	0.0459	0.479
1:	-0.0366	0.626
Combined K-S:	0.0459	0.856

Exemple R:

Figure 2: Kolmogorov Smirnov, R

```
        Asymptotic two-sample Kolmogorov-Smirnov test

data:  age_treated and age_control
D = 0.045855, p-value = 0.8558
alternative hypothesis: two-sided
```

Cette sortie Stata/R nous permet de conclure que nous ne pouvons pas rejeter l'hypothèse nulle (p-value = 0.856). On ne peut pas rejeter que l'age est identiquement distribué entre les deux groupes. (Les deux groupes semblent similaires, au niveau des distributions de l'age).

2. **Test du Chi-2**: De la même manière que le test de Kolmogorov-Smirnov, on souhaite tester la similitude des distributions de X, mais pour les variables catégorielles (ex: éducation)
   $\Rightarrow H_0^C$: La variable considérée est identiquement distribuée dans les deux groupes.

3. **Test de Student:** On veut vérifier si les moyennes de X sont les mêmes dans les deux groupes (avec hypothèse de variance inégale entre les deux groupes)
   $\Rightarrow H_0^S$: La différence de moyenne (de la variable considérée) des deux groupes est nulle.

Attention: Ce n'est pas parce qu'un test rejette l'hypothèse nulle que cela signifie forcément que nous ne sommes pas dans le cas d'une expérience randomisée.

1.2 Confounders

Nous avons vu plusieurs exemples de traitement D qui ne vérifient pas $Y(0), Y(1) \perp\!\!\!\perp D$ et qui entraînent un biais de sélection. Pour vérifier si nous sommes dans ce cas, nous pouvons regarder si les groupes de traitement et de contrôle sont similaires (cf section 1.1). Les expériences randomisées permettent normalement d'éviter ce biais de sélection. Mais même dans certains cas d'expériences randomisées, il est possible que la probabilité de recevoir le traitement soit différente entre plusieurs groupes d'individus. Par exemple, parmi le groupe de traitement choisi initialement, on re-sélectionne aléatoirement les personnes qui recevront vraiment le traitement.

Une solution est d'inclure dans le modèle des variables de contrôles G qui influencent à la fois la variable de traitement D et l'outcome Y (et donc Y(0)). Il est important que ce soit des variables de **pré-traitement** afin qu'elles n'aient pas été influencées par le traitement D. Leur inclusion permet d'estimer l'effet du traitement de manière non-biaisée. On parle de "variables de contrôle" car on ne va pas regarder leur effet directement sur l'outcome Y mais elles seront incluses dans le "toutes choses égales par ailleurs" (par exemple on compare l'effet du traitement entre les traités et les non-traités à niveau d'éducation, de revenu et d'âge donnés). Ces variables de contrôle doivent permettre de prendre en compte la totalité des facteurs déterminant Y(0) et D, ie on doit trouver les variables de contrôle G telles que $Y(0), Y(1)|G \perp\!\!\!\perp D$.

Autrement dit, on doit avoir les hypothèses suivantes (cf cours chapitre 3, slide 28):
$Y(d) = Y(0) + \Delta d$ avec $E\big(\Delta|D,G\big) = \delta_0$ (effet moyen du traitement est le même pour tous les "groupes d'individus" définis par G)
$Y(0) = \zeta_0 + G'\gamma_0 + \eta$ avec $E\big(\eta|D,G\big) = 0$ (plus de sélection dans le traitement lorsque l'on contrôle par G dans Y(0), $cov(D, Y(0)|G = 0)$)

2 Propriétés calculatoires

- **Simplification espérances conditionnelles**
  - $E\big(Y|D=0\big) = E\big(Y(1)D + (1-D)Y(0)|D=0\big) = E\big(Y(0)|D=0\big)$
    $\Rightarrow$ Le conditionnement sur D disparaît uniquement lorsque $Y(0) \perp\!\!\!\perp D$, auquel cas $E\big(Y(0)|D=0\big) = E\big(Y(0)\big)$
  - $E\big(Y|D=1\big) = E\big(Y(1)D + (1-D)Y(0)|D=1\big) = E\big(Y(1)|D=1\big)$
    $\Rightarrow$ Le conditionnement sur D disparaît uniquement lorsque $Y(1) \perp\!\!\!\perp D$, auquel cas $E\big(Y(1)|D=1\big) = E\big(Y(1)\big)$

- **Pourquoi dit-on que B=0 si Cov(Y(0),D)=0 ?**

$$
\begin{aligned}
\frac{Cov(Y(0),D)}{V(D)} =& \frac{E(DY(0)) - E(D)E(Y(0))}{E(D^2) - E(D)^2} \\
=& \frac{\overbrace{E(Y(0)|D=1)\times P(D=1)}^{E(DY(0))} - \overbrace{P(D=1)}^{E(D)}\Big[\overbrace{E(Y(0)|D=1)\times P(D=1) + E(Y(0)|D=0)\times P(D=0)}^{E(Y(0))}\Big]}{E(D) - E(D)E(D)} \\
=& \frac{P(D=1)(1-P(D=1))E(Y(0)|D=1) - P(D=1)(1-P(D=1))E(Y(0)|D=0)}{E(D)(1-E(D))} \\
=& \frac{P(D=1)(1-P(D=1))\Big[E(Y(0)|D=1) - E(Y(0)|D=0)\Big]}{P(D=1)(1-P(D=1))} \\
=& E(Y(0)|D=1) - E(Y(0)|D=0)
\end{aligned}
$$