Estimation and Hypothesis Testing 2 | *Les estimateurs et les tests d'hypothèses 2*

Nahomi Ichino and/et Yannick Kouogueng

22 May/mai 2025

Covariate Adjustment | *Ajustement des covariables*

Cluster Randomization | *Randomisation par grappe*

Experiments with Multiple Arms | *Les éxperiences avec plusieurs bras*

Factorial Design | *La conception factorielle*

A Quick Reminder | *Un pétit rappel*

- Remember: Analyze as you randomize
- We prefer estimators that are unbiased and have greater precision
- Hypothesis testing can be simple with linear regression

- N'oubliez pas : Analysez comme vous randomisez
- Nous préférons les estimateurs non biaisés et plus précis
- Test d'hypothèse peut être simple avec la régression linéaire

Covariate Adjustment | *Ajustement des covariables*

Estimator: Linear regression with covariates | Estimateur : La régression linéaire avec des covariables

$$Y_i = \beta_0 + \beta_1 Z_i + \gamma X_i + e_i$$

- We can include a **pre-treatment covariate** X that is *predictive of* the outcome variable in our regression model.
- Think of X as fixed before the randomization. For example: pre-treatment measure of the outcome.
- Careful: This can bias our estimates, but improve their precision.

- Nous pouvons inclure une **covariable pré-traitement** X qui est *prédictive* de la variable de résultat dans notre modèle de régression.
- Pensez que X est fixé avant la randomisation. Par exemple : un mesure du résultat avant le traitement.
- Attention : Cela peut biaiser nos estimations, mais améliorer leur précision.

Estimator: Linear regression with covariates | *Estimateur : La régression linéaire avec des covariables*

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 Z_i + \hat{\gamma} X_i + e_i$$

- The estimated coefficient on the treatment variable ($\hat{\beta}_1$) is again our $\widehat{ATE}$.
- The estimated coefficient on the covariate ($\hat{\gamma}$) is *not* the causal effect of that variable.

- Le coefficient estimé sur la variable de traitement ($\hat{\beta}_1$) est encore notre $\widehat{ATE}$.
- Le coefficient estimé de la covariable ($\hat{\gamma}$) n'est *pas* l'effet estimé causal de cette variable.

Estimator: Linear regression with covariates | *Estimateur : La régression linéaire avec des covariables*

	Reject FM	
	(1)	(2)
EFM Treat	0.093***	0.095***
Standard Error	0.027	0.020
RI *p*-value	0.001	<0.001
Hypothesis	+	+
Control Mean	0.82	0.82
Control SD	0.16	0.16
DV Range	[0-1]	[0-1]
Blocked FE	Yes	Yes
Controls	No	16
Adj-R^2	0.09	0.23
Observations	998	998

Note: * p < .1, ** p < 0.05, and *** p < 0.01

egap EVIDENCE IN GOVERNANCE AND POLITICS

Estimator: Linear regression with covariates | *Estimateur : La régression linéaire avec des covariables*

```r
library(estimatr)
# lm_robust(Y ~ treatment + Language + Gender)
```

Cluster Randomization | *Randomisation par grappe*

Estimator: Regression with cluster-robust standard errors | *Estimateur : La régression avec des erreurs types robustes au niveau du cluster*

$$Y_{ic} = \hat{\beta}_0 + \hat{\beta}_1 Z_c + e_{ic}$$
$$Y_{ic} = \hat{\beta}_0 + \hat{\beta}_1 Z_c + \hat{\gamma} X_{ic} + e_{ic}$$

- Our analysis has to take into account the fact that treatment is assigned at the cluster level with *cluster-robust standard errors*.
- $\hat{\beta}_1$ is the $\widehat{ATE}$ of the treatment on individual units.
- We can also do covariate adjustment at the same time.

- Notre analyse doit prendre en compte le fait que le traitement est attribué au niveau du cluster avec des *erreurs types robustes au niveau du cluster*.
- $\hat{\beta}_1$ est $\widehat{ATE}$ du traitement sur les unités individuelles.
- Nous pouvons également effectuer un ajustement covariable en même temps.

Cluster Randomization | *Randomisation par grappe*

```r
library(estimatr)

# lm_robust(Y ~ treatment, clusters=cluster_variable)

# lm_robust(Y ~ treatment + covariate, clusters=cluster_variable)
```

Experiments with Multiple Arms | *Les éxperiences avec plusieurs bras*

EVIDENCE IN GOVERNANCE AND POLITICS

Estimator 1: Difference-in-Means | *Estimateur 1 : La différence en moyennes*

Z_A only	Z_B only	Neither (control)

- We can always take the difference-in-means between any two groups.

- Nous pouvons toujours tenir compte de la différence de moyens entre deux groupes.

Estimator 2: Linear regression | *Estimateur 2 : La régression linéaire*

$$Y_i = \hat{\alpha} + \hat{\beta}_A Z_{Ai} + \hat{\beta}_B Z_{Bi} + e_i$$
$$Y_i = \hat{\alpha} + \hat{\beta}_A Z_{Ai} + \hat{\beta}_B Z_{Bi} + \hat{\gamma} X_i + e_i$$

- Regression with an indicator variable for each of the two treatment arms.
  - $Z_{Ai} = 1$ if unit i has treatment Z_A, 0 otherwise
  - $Z_{Bi} = 1$ if unit i has treatment Z_B, 0 otherwise
- We can also do covariate adjustment at the same time.

- Régression avec une variable indicatrice pour chacun des deux bras de traitement.
  - $Z_{Ai} = 1$ si unité i a traitement Z_A, sinon 0
  - $Z_{Bi} = 1$ si unité i a traitement Z_B, sinon 0
- Nous pouvons également effectuer un ajustement covariable en même temps.

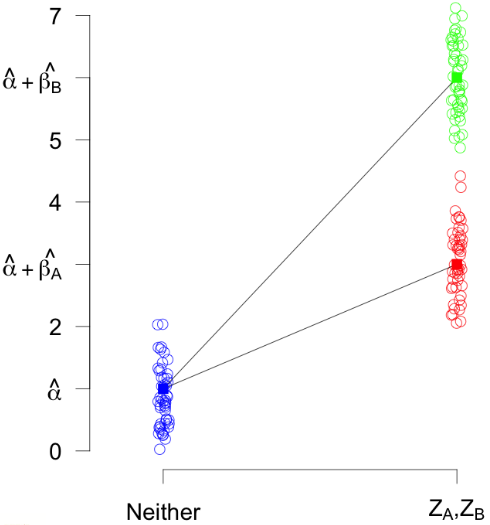
Estimator 2: Linear regression | *Estimateur 2 : La régression linéaire*

$$Y_i = \hat{\alpha} + \hat{\beta_A} Z_{Ai} + \hat{\beta_B} Z_{Bi} + e_i$$

- $\hat{\beta_A}$ is the $\widehat{ATE}$ of Z_A (compared with control).
- $\hat{\beta_B}$ is the $\widehat{ATE}$ of Z_B (compared with control).
- How do we estimate the effect of Z_B compared to Z_A?

- $\hat{\beta_A}$ est $\widehat{ATE}$ de Z_A (par rapport au contrôle).
- $\hat{\beta_B}$ est $\widehat{ATE}$ de Z_B (par rapport au contrôle).
- Comment estimer l'effet de Z_B par rapport à Z_A ?

Estimator 2: Linear regression | *Estimateur 2 : La régression linéaire*



Z_A only
Z_B only
Neither (control)

$Y_i = \hat{\alpha} + \hat{\beta}_A Z_{Ai} + \hat{\beta}_B Z_{Bi} + e_i$

Estimators for Multi-arm Designs | *Les estimateurs pour les éxperiences avec plusiers bras*

```
# library(estimatr)
# difference_in_means(Y ~ treatment,
#         condition1="T1",
#         condition2="T2")
#
#
# library(car)
# M <- lm_robust(Y ~ as.factor(treatment))
# linearHypothesis(M, "T1=T2")
```

Factorial Design | *La conception factorielle*

Estimator 1: Difference-in-Means | *Estimateur 1 : La différence en moyennes*

		Z_2 = 1	Z_2 = 0
Z_1	1	Both Z_1 and Z_2	Z_1 only
	0	Z_2 only	Neither

- We use factorial design, when we are interested in interaction effects.
- If we have a 2*2 factorial design, we have four groups.
- We can always take the difference-in-means between any two groups.

- Nous utilisons un plan factoriel quand nous nous intéressons aux effets d'interaction.
- Si nous avons une conception factorielle 2*2, nous avons 4 groupes.
- Nous pouvons toujours tenir compte de la différence de moyens entre

Estimator 2: Linear Regression with an Interaction Term | *Estimateur 2 : La régression linéaire avec un terme d'interaction*

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 Z_{1i} + \hat{\beta}_2 Z_{2i} + \hat{\beta}_3 Z_{1i} * Z_{2i} + e_i$$

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 Z_{1i} + \hat{\beta}_2 Z_{2i} + \hat{\beta}_3 Z_{1i} * Z_{2i} + \hat{\gamma} X_i + e_i$$

- Indicator variables for Z_1 and Z_2.
- We can also do covariate adjustment at the same time.

- Variables indicatrices pour Z_1 et Z_2.
- Nous pouvons également effectuer un ajustement covariable en même temps.

Estimator 2: Linear Regression with an Interaction Term | *Estimateur 2 : La régression linéaire avec un terme d'interaction*

		Z_2 = 1	Z_2 = 0
Z_1	1	Both Z_1 and Z_2	Z_1 only
	0	Z_2 only	Neither

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 Z_{1i} + \hat{\beta}_2 Z_{2i} + \hat{\beta}_3 Z_{1i} * Z_{2i} + e_i$$

- Estimand | Paramètre: $E[Y(Z_1 = 1)|Z_2 = 0] - E[Y(Z_1 = 0)|Z_2 = 0]$
  - $\hat{\beta}_1$ is the $\widehat{ATE}$ of Z_1 conditional on $Z_2 = 0$ | $\widehat{ATE}$ de Z_1 conditionnel à $Z_2 = 0$.

Estimator 2: Linear Regression with an Interaction Term | *Estimateur 2 : La régression linéaire avec un terme d'interaction*

		Z_2: 1	Z_2: 0
Z_1	1	Both Z_1 and Z_2	Z_1 only
	0	Z_2 only	Neither

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 Z_{1i} + \hat{\beta}_2 Z_{2i} + \hat{\beta}_3 Z_{1i} * Z_{2i} + e_i$$

- Estimand | Paramètre: $E[Y(Z_1 = 1) - Y(Z_1 = 0) | Z_2 = 1]$
  - $\hat{\beta}_1 + \hat{\beta}_3 = \widehat{ATE}$ of Z_1 conditional on $Z_2 = 1$ | $\widehat{ATE}$ de Z_1 conditionnel à $Z_2 = 1$
- β_3 is called the interaction effect. β_3 est appelé l'effet d'interaction.

Estimator 2: Linear Regression with an Interaction Term | *Estimateur 2 : La régression linéaire avec un terme d'interaction*

Antisocial behaviors, z-score	
Intercept	0.151
Therapy	-0.249**
	(0.088)
Cash	-0.079
	(0.091)
Cash*Therapy	-0.308**
	(0.089)

NOTE: * p<0.05; ** p<0.01; *** p< 0.001 . The table reports intent-to-treat estimates of the effect of each treatment arm, controlling for covariates and block fixed effects. Taken from Table 2 of Blattman, Jamison, and Sheridan (2017)

$$\hat{Y} = \hat{\alpha} - 0.249\,Therapy - 0.079\,Cash - 0.308\,Therapy * Cash$$

Estimator 2: Linear Regression with an Interaction Term | *Estimateur 2 : La régression linéaire avec un terme d'interaction*

$$\hat{Y} = \hat{\alpha} - 0.249\,Therapy - 0.079\,Cash - 0.308\,Therapy * Cash$$

Neither: $\hat{Y} = \hat{\alpha}$

Cash only: $\hat{Y} = \hat{\alpha} - 0.079$

Therapy only: $\hat{Y} = \hat{\alpha} - 0.249$

Therapy and Cash: $\hat{Y} = \hat{\alpha} - 0.249 - 0.079 - 0.308$

$\widehat{ATE}$ of Cash conditional on No Therapy $= \hat{\alpha} - 0.079 - \hat{\alpha} = -0.079$

$\widehat{ATE}$ of Cash conditional on Therapy $=$
$\hat{\alpha} - 0.249 - 0.079 - 0.308 - (\hat{\alpha} - 0.249) = -0.079 - 0.308 = -0.387$

Interaction effect $= -0.079 - 0.308 - (-0.079) = -0.308$

Estimator 2: Linear Regression with an Interaction Term | *Estimateur 2 : La régression linéaire avec un terme d'interaction*

```r
library(estimatr)
# lm_robust(Y ~ Z1 + Z2 + Z1*Z2)

# lm_robust(Y ~ Z1*Z2)

# lm_robust(Y ~ Z1*Z2 + covariate)
```