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

Name

Date

Covariate Adjustment | *Ajustement des covariables*

Cluster Randomization | *Randomisation par grappe*

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

Factorial Design | *La concéption factorielle*

A Quick Reminder | *Un pétit rappel*

- Remember: Analyze as you randomize
- We prefer estimators that are unbiased and have greater precision

- N'oubliez pas : Analysez comme vous randomisez
- Nous préférons les estimateurs non biaisés et plus précis.

Covariate Adjustment | *Ajustement des covariables*

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

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

- Including a **pre-treatment covariate** X that is *predictive of* the outcome variable in our regression model is called *covariate adjustment*.
- For example: pre-treatment measure of the outcome.
- This can bias our estimates, but improve their precision.

- L'inclusion d'une **covariable pré-traitement** X qui est *prédictive* de la variable de résultat dans notre modèle de régression est appelée *ajustement des covariables*.
- Par exemple: un mesure du résultat avant le traitement.
- Cela peut biaiser nos estimations, mais améliorer leur précision.

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

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

- The coefficient on the treatment variable (β_1) is again our estimate of the ATE.
- The coefficient on the covariate (γ) is *not* an estimate of the causal effect of that variable.

- Le coefficient sur la variable de traitement (β_1) est encore notre estimation de l'ATE.
- Le coefficient de la covariable (γ) n'est *pas* une estimation de l'effet causal de cette variable.

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

Table 4. Adoption for Parents Sampled for SAFI & Subsidy Programs

	Used Fertilizer Season 1	
Panel A. 2004 Season 1 Treatments	(1)	(2)
SAFI Season 1	0.114	0.143
	(0.035)***	(0.038)***
Starter Kit Farmer	0.059	0.080
	(0.042)	(0.046)*
Starter Kit Farmer * Demonstration Plot School	-0.026	-0.061
	(0.060)	(0.066)
Demonstration Plot School	0.006	0.441
	(0.314)	(0.435)
Household had Used Fertilizer Prior to Season 1	0.369	0.315
	(0.031)***	(0.035)***
Male		0.012
		(0.033)
Home has mud walls		-0.193
		(0.081)**
Education primary respondent		0.004
		(0.004)
Income in past month (in 1,000 Kenyan shillings)		0.004
		(0.003)
Mean Usage Among Season 1 Comparison	0.244	0.240
Mean Usage Among Pure Comparison Group	0.296	0.227
Observations	876	716

Cluster Randomization | *Randomisation par grappe*

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

$$Y_{ic} = \beta_0 + \beta_1 Z_c + e_{ic}$$

- Our analysis has to take into account the fact that treatment is assigned at the cluster level with *cluster-robust standard errors*.
- β_1 is the ATE of the treatment at the individual level.
- 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*.
- β_1 est l'ATE du traitement au niveau individuel.
- Nous pouvons également effectuer un ajustement covariable en même temps.

Cluster Randomization | *Randomisation par grappe*

- Remember: we have assumed non-interference (part of SUTVA).
- If we are concerned about interference within clusters (but can assume non-interference across clusters), we can shift the analysis to the cluster level by using the average or other summary measure at the cluster level as our outcome variable.

- N'oubliez pas : nous avons supposé la non-contamination (qui fait partie de la SUTVA).
- Si nous sommes préoccupés par la contamination au sein des clusters (mais pouvons supposer qu'il n'y a pas de contamination entre les clusters), nous pouvons déplacer l'analyse au niveau du cluster en utilisant la moyenne ou une autre mesure récapitulative au niveau du cluster comme variable de résultat.

Cluster Randomization | *Randomisation par grappe*

Table 2 | Effect of treatment on movement outcomes

			Mean (95% CI)		OLS model			Number of days × counties
Campaign	**Outcome**	**Period**	**High-intensity county**	**Low-intensity county**	**High-intensity county (95% CI)**	***P* value**	**RI *P* value**	
Both campaigns	Distance traveled	From day − 3 to day − 1	−4.384 (−4.973,−3.796)	−3.603 (−4.254, −2.952)	−0.993 (−1.616, −0.371)	0.002	0.002	4,059
Both campaigns	Share ever left home	Thanksgiving (26 November) or Christmas (24–25 December)	72.326 (72.012, 72.639)	72.381 (72.092, 72.670)	0.030 (−0.361, 0.420)	0.881	0.911	2,017
Thanksgiving	Distance traveled	From day − 3 to day − 1	−6.082 (−6.822, −5.341)	−5.320 (−6.113, −4.527)	−0.924 (−1.785, −0.063)	0.035	0.030	2,072
Thanksgiving	Share ever left home	Thanksgiving (26 November)	71.308 (70.885, 71.731)	71.468 (71.071, 71.866)	0.012 (−0.438, 0.461)	0.959	0.966	689
Christmas	Distance traveled	From day − 3 to day − 1	−2.603 (−3.279, −1.927)	−1.823 (−2.588, −1.057)	−1.041 (−1.847, −0.235)	0.011	0.008	1,987
Christmas	Share ever left home	Christmas (24–25 December)	72.859 (72.507, 73.210)	72.852 (72.520, 73.185)	0.095 (−0.289, 0.479)	0.629	0.580	1,328

The control and treatment means at the county level and different periods, in addition to the estimate of the treatment coefficient in Eq. (1). Standard errors are clustered at the county level. 95% CIs are reported in parentheses. *P* values are based on a two-sided test. RI *P* values are computed using randomization inference, accounting for the two-stage design.

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

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.
- But hypothesis testing is simpler with regression.

- Nous pouvons toujours tenir compte de la différence de moyens entre deux groupes.
- Mais le test d'hypothèse est plus simple avec la régression.

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

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

- Regression with an indicator variable for each of the two treatment arms.
- We can also do covariate adjustment at the same time.

- Régression avec une variable indicatrice pour chacun des deux bras de traitement.
- Nous pouvons également effectuer un ajustement covariable en même temps.

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

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

- β_A is the ATE of Z_A (compared with control).
- β_B is the ATE of Z_B (compared with control).

- β_A est l'ATE de Z_A (par rapport au contrôle).
- β_B est l'ATE de Z_B (par rapport au contrôle).

Factorial Design | *La concéption factorielle*

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

Neither	Z_2 only
Z_1 only	Both Z_1 and Z_2

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

- Si nous avons une concéption factorielle 2*2, nous avons 4 groupes.
- Nous pouvons toujours tenir compte de la différence de moyens entre deux groupes.

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

$$Y_i = \gamma_0 + \gamma_1 Z_{1only,i} + \gamma_2 Z_{2only,i} + \gamma_3 Z_{both1\&2,i} + e_i$$

- Regression makes it easier to do hypothesis testing.
- An indicator variable for each of the three treatment conditions.
- We can also do covariate adjustment at the same time.

- La régression facilite les tests d'hypothèses.
- Régression avec une variable indicatrice pour chacun des trois bras de traitement.
- Nous pouvons également effectuer un ajustement covariable en même temps.

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

$$Y_i = \gamma_0 + \gamma_1 Z_{1only,i} + \gamma_2 Z_{2only,i} + \gamma_3 Z_{both1\&2,i} + e_i$$

- γ_1: ATE of Z_1 conditional on $Z_2 = 0$ | *l'ATE de* Z_1 conditionnel à $Z_2 = 0$, $(E[Y(Z_1 = 1) - Y(Z_1 = 0)|Z_2 = 0])$
- γ_2: ATE of Z_2 conditional on $Z_1 = 0$ | *l'ATE de* Z_2 conditionnel à $Z_1 = 0$, $(E[Y(Z_2 = 1) - Y(Z_2 = 0)|Z_1 = 0])$
- γ_3: ATE of having both Z_1 and Z_2 compared with having neither | *l'ATE d'avoir à la fois* Z_1 et Z_2 par rapport à n'avoir ni l'un ni l'autre

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

Table 2. Short- and Medium Run Impacts: Roll Call Data

	(1)	(2)	(3)	(4)	(5)	(6)
	Impacts after 3 years					
Panel A. Girls	Dropped Out of Primary School	Attendance rate (while enrolled)	Ever Married	Ever Pregnant[a]	Ever Pregnant but Never Married	Ever Married but Never Pregnant
Stand-Alone Education Subsidy (S)	-0.031	-0.002	-0.026	-0.027	-0.004	-0.002
	(0.012)**	(0.006)	(0.010)**	(0.011)**	(0.006)	(0.003)
Stand-Alone HIV Education (H)	0.003	-0.008	0.011	-0.007	-0.014	0.005
	(0.011)	(0.006)	(0.009)	(0.011)	(0.006)**	(0.003)*
Joint Program (SH)	-0.016	0.000	0.000	-0.011	-0.013	-0.001
	(0.012)	(0.006)	(0.009)	(0.010)	(0.006)**	(0.003)
Observations	9116	8232	9107	9072	9072	9072
Mean of Dep. Var. (Control)	0.188	0.939	0.128	0.160	0.046	0.011
p-val (Test: S = SH)	*0.245*	*0.712*	*0.012***	*0.149*	*0.088**	*0.812*
p-val (Test: H = SH)	*0.097**	*0.201*	*0.227*	*0.728*	*0.872*	*0.057**
p-val (Test: S = H)	*0.005****	*0.327*	*0****	*0.083**	*0.092**	*0.021***
p-val (Test: SH = S + H)	*0.484*	*0.235*	*0.285*	*0.137*	*0.544*	*0.309*

Estimator 3: Linear Regression with an Interaction Term | *Estimateur 3 : La régression linéaire avec un terme d'interaction*

$$Y_i = \beta_0 + \beta_1 Z_{1i} + \beta_2 Z_{2i} + \beta_3 Z_{1i} * Z_{2i} + e_i$$

- 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 3: Linear Regression with an Interaction Term | *Estimateur 3 : La régression linéaire avec un terme d'interaction*

$$Y_i = \beta_0 + \beta_1 Z_{1i} + \beta_2 Z_{2i} + \beta_3 Z_{1i} Z_{2i} + e_i$$

- β_1 is the ATE of Z_1 conditional on $Z_2 = 0$ | *l'ATE de* Z_1 conditionnel à $Z_2 = 0$, $(E[Y(Z_1 = 1) - Y(Z_1 = 0)|Z_2 = 0])$
- β_2 is the ATE of Z_2 conditional on $Z_1 = 0$ | *l'ATE de* Z_2 conditionnel à $Z_1 = 0$, $(E[Y(Z_2 = 1) - Y(Z_2 = 0)|Z_1 = 0])$

Estimator 3: Linear Regression with an Interaction Term | *Estimateur 3 : La régression linéaire avec un terme d'interaction*

$$Y_i = \beta_0 + \beta_1 Z_{1i} + \beta_2 Z_{2i} + \beta_3 Z_{1i} Z_{2i} + e_i$$

- $\beta_1 + \beta_3 =$ ATE of Z_1 conditional on $Z_2 = 1$ | *l'ATE de* Z_1 conditionnel à $Z_2 = 1$, $(E[Y(Z_1 = 1) - Y(Z_1 = 0)|Z_2 = 1])$
- $\beta_2 + \beta_3 =$ ATE of Z_2 conditional on $Z_1 = 1$ | *l'ATE de* Z_2 conditionnel à $Z_1 = 1$, $(E[Y(Z_2 = 1) - Y(Z_2 = 0)|Z_1 = 0])$
- β_3 is called the interaction effect. β_3 est appelé l'effet d'interaction.

Estimator 3: Linear Regression with an Interaction Term | *Estimateur 3 : La régression linéaire avec un terme d'interaction*

Table 4. Adoption for Parents Sampled for SAFI & Subsidy Programs

	Used Fertilizer Season 1	
Panel A. 2004 Season 1 Treatments	(1)	(2)
SAFI Season 1	0.114	0.143
	(0.035)***	(0.038)***
Starter Kit Farmer	0.059	0.080
	(0.042)	(0.046)*
Starter Kit Farmer * Demonstration Plot School	-0.026	-0.061
	(0.060)	(0.066)
Demonstration Plot School	0.006	0.441
	(0.314)	(0.435)
Household had Used Fertilizer Prior to Season 1	0.369	0.315
	(0.031)***	(0.035)***
Male		0.012
		(0.033)
Home has mud walls		-0.193
		(0.081)**
Education primary respondent		0.004
		(0.004)
Income in past month (in 1,000 Kenyan shillings)		0.004
		(0.003)
Mean Usage Among Season 1 Comparison	0.244	0.240
Mean Usage Among Pure Comparison Group	0.296	0.227
Observations	876	716