28) Matching Methods in Practice

Vitor Kamada

April 2018

Reference

Imbens (2015). Matching Methods in Practice. Journal of Human Resources, Vol 50:2, pp. 373-419

Assessing Overlap: Normalized Differences in Average Covariates

$$\Delta_{X,k} = \frac{\bar{X}_{t,k} - \bar{X}_{c,k}}{\sqrt{\frac{(s^2_{X,t,k} + s^2_{X,c,k})}{2}}}$$

$$t_{X,k} = \frac{\bar{X}_{t,k} - \bar{X}_{c,k}}{\sqrt{\frac{s^2_{X,t,k}}{N_t} + \frac{s^2_{X,c,k}}{N_c}}}$$

Summary Statistics for Experimental Lalonde Data

	Experimental Controls ($N_c = 260$)		Trainees ($N_t = 185$)			
Covariate	Mean	(Standard Deviation)	Mean	(Standard Deviation)	t-stat	nor-dif
Black	0.83	(0.38)	0.84	(0.36)	0.5	0.04
Hispanic	0.11	(0.31)	0.06	(0.24)	–1.9	–0.17
Age	25.05	(7.06)	25.82	(7.16)	1.1	0.11
Married	0.15	(0.36)	0.19	(0.39)	1.0	0.09
No degree	0.83	(0.37)	0.71	(0.46)	–3.1	–0.30
Education	10.09	(1.61)	10.35	(1.97)	1.4	0.14
E'74	2.11	(5.69)	2.10	(4.89)	–0.0	–0.00
U'74	0.75	(0.43)	0.71	(0.46)	–1.0	–0.09
E'75	1.27	(3.10)	1.53	(3.22)	0.9	0.08
U'75	0.68	(0.47)	0.60	(0.49)	–1.8	–0.18

Estimated Parameters of Propensity Score for the Lalonde Experimental Data

Variable	Estimated	(Standard Error)
Intercept	–3.48	(0.10)
Preselected linear terms		
Earn '74	0.03	(0.05)
Unemployed '74	–0.24	(0.39)
Earn '75	0.06	(0.05)
Unemployed '75	–3.48	(1.65)
Additional linear terms		
No degree	7.33	(4.25)
Hispanic	–0.65	(0.39)
Education	0.29	(0.37)
Second-order terms		
No degree × education	–0.67	(0.35)
Earn '74 × no degree	–0.13	(0.06)
Unemployed '75 × education	0.30	(0.16)

Sample Sizes for Subsamples with the Propensity Score between α and $1-\alpha$ ($\alpha = 0.1299$)

	Low $e(x) < \alpha$	Middle $\alpha \leq e(X) \leq 1 - \alpha$	High $1 - \alpha < e(X)$	All
Controls	4	256	0	260
Treated	1	182	2	185
All	5	438	2	445

Propensity Score: Blocking with Regression

$[0, 1]$ **into** J **intervals** $[b_{j-1}, b_j)$

$$\min_{\alpha,\tau,\beta} \sum_{i=1}^{N} B_i(j)[Y_i - \alpha - \tau W_i - \beta' X_i]^2$$

$$\tau_{block,treat}(Y, W, X) = \sum_{j=1}^{J} \frac{N_{tj}}{N_t} \hat{\tau}_j$$

Simple Matching using Mahalanobis Metric

$$||x, x'|| = (x - x')' \hat{\Omega}_X^{-1} (x - x')$$

$$\hat{Y}_i(0) = \begin{cases} Y_i^{obs} & if \quad W_i = 0 \\ Y_{m(i)} & if \quad W_i = 1 \end{cases} \qquad \hat{Y}_i(1) = \begin{cases} Y_{m(i)} & if \quad W_i = 0 \\ Y_i^{obs} & if \quad W_i = 1 \end{cases}$$

$$\hat{X}_i(0) = \begin{cases} X_i & if \quad W_i = 0 \\ X_{m(i)} & if \quad W_i = 1 \end{cases} \qquad \hat{X}_i(1) = \begin{cases} X_{m(i)} & if \quad W_i = 0 \\ X_i & if \quad W_i = 1 \end{cases}$$

$$\hat{\tau}_{sm} = \frac{1}{N} \sum_{i=1}^{N} [\hat{Y}_i(1) - \hat{Y}_i(0)]$$

Bias-Adjusted Matching Estimator [Abadie and Imbens (2006, 2010)]

$$\hat{Y}_i(0) = \alpha_c + \beta_c \hat{X}_i(0) + \epsilon_{ci}$$

$$\hat{Y}_i(1) = \alpha_t + \beta_t \hat{X}_i(1) + \epsilon_{ti}$$

$$\hat{Y}_i^{adj}(0) = \begin{cases} Y_i^{obs} & if \quad W_i = 0 \\ \hat{Y}_i(0) + \hat{\beta}_c(X_i X_{l(i)}) & if \quad W_i = 1 \end{cases}$$

$$\hat{Y}_i^{adj}(1) = \begin{cases} \hat{Y}_i(1) + \hat{\beta}_t(X_i X_{l(i)}) & if \quad W_i = 0 \\ Y_i^{obs} & if \quad W_i = 1 \end{cases}$$

$$\hat{\tau}_{adj} = \frac{1}{N} \sum_{i=1}^{N} [\hat{Y}_i^{adj}(1) - \hat{Y}_i^{adj}(0)]$$

Assessing Unconfoundedness: Estimates of Average Treatment Effects for Pseudo Outcomes

	Blocking		Matching	
Pseudo Outcome	Estimated	(Standard Error)	Estimated	(Standard Error)
Earn '75	0.22	(0.22)	0.03	(0.27)
(Earn '74 + Earn'75)/2	0.03	(0.36)	–0.08	(0.41)

Subclasses for the Experimental Lalonde data

Subclasss	p-score		Number of Controls	Number of Treated	Average p-score		Average Difference in p-score	t-stat
	Minimum	Maximum			Controls	Treated		
1	0.07	0.38	152	67	0.32	0.33	0.01	0.8
2	0.38	0.49	52	42	0.42	0.42	0.01	1.0
3	0.49	0.85	52	73	0.56	0.58	0.02	1.4

$$E[e(x)|W = 1] = 0.45$$

$$E[e(x)|W = 0] = 0.39$$

Experimental Lalonde Data: Estimates of Average Treatment Effects

	Full Sample		Trimmed Sample		
Covariate	1 Block	Match	1 Block	2 Blocks	3 Blocks
Number	1.79	2.21	1.69	1.49	1.48
	(0.67)	(0.82)	(0.66)	(0.68)	(0.68)
Few	1.74	2.15	1.60	1.54	1.52
	(0.67)	(0.82)	(0.66)	(0.66)	(0.68)
All	1.67	2.11	1.56	1.56	1.46
	(0.64)	(0.82)	(0.65)	(0.64)	(0.65)

Summary Statistics for Nonexperimental Lalonde Data

	CPS Controls (N_c = 15,992)		Trainees (N_t = 185)			
Covariate	Mean	(Standard Deviation)	Mean	(Standard Deviation)	t-stat	nor-dif
Black	0.07	(0.26)	0.84	(0.36)	28.6	2.43
Hispanic	0.07	(0.26)	0.06	(0.24)	–0.7	–0.05
Age	33.23	(11.05)	25.82	(7.16)	–13.9	–0.80
Married	0.71	(0.45)	0.19	(0.39)	–18.0	–1.23
No degree	0.30	(0.46)	0.71	(0.46)	12.2	0.90
Education	12.03	(2.87)	10.35	(2.01)	–11.2	–0.68
E'74	14.02	(9.57)	2.10	(4.89)	–32.5	–1.57
U'74	0.12	(0.32)	0.71	(0.46)	17.5	1.49
E'75	13.65	(9.27)	1.53	(3.22)	–48.9	–1.75
U'75	0.11	(0.31)	0.60	(0.49)	13.6	1.19

Estimated Parameters of Propensity Score for the Lalonde Nonexperimental (CPS) Data

Variable	Estimated	(Standard Error)
Intercept	–16.20	(0.69)
Preselected linear terms		
Earn '74	0.41	(0.11)
Unemployed '74	0.42	(0.41)
Earn '75	–0.33	(0.06)
Unemployed '75	–2.44	(0.77)
Additional linear terms		
Black	4.00	(0.26)
Married	–1.84	(0.30)
No degree	1.60	(0.22)
Hispanic	1.61	(0.41)
Age	0.73	(0.09)
Second-order terms		
Age × age	–0.007	(0.002)
Unemployed '74 × unemployed '75	3.41	(0.85)
Earn '74 × age	–0.013	(0.004)
Earn '75 × married	0.15	(0.06)
Unemployed '74 × earn '75	0.22	(0.09)

Normalized Differences Before and After Matching for Nonexperimental Lalonde Data

	Full Sample nor-dif	Matched Sample nor-dif	Ratio of nor-dif
Black	2.43	0.00	0.00
Hispanic	–0.05	0.00	–0.00
Age	–0.80	–0.15	0.19
Married	–1.23	–0.28	0.22
No degree	0.90	0.25	0.28
Education	–0.68	–0.18	0.26
E'74	–1.57	–0.03	0.02
U'74	1.49	0.02	0.02
E'75	–1.75	–0.07	0.04
U'75	1.19	0.02	0.02

Estimated Parameters of Propensity Score for the Matched Lalonde Nonexperimental (CPS) Data

Variable	Estimated	(Standard Error)
Intercept	–0.15	(0.11)
Preselected linear terms		
Earn '74	0.03	(0.04)
Unemployed '74	–0.00	(0.42)
Earn '75	–0.06	(0.05)
Unemployed '75	0.26	(0.36)
Additional linear terms		
Married	–0.52	(0.55)
No degree	0.26	(0.26)
Second-order terms		
Unemployed '75 × married	–1.24	(0.55)
Married × no degree	1.10	(0.55)

Assessing Unconfoundedness for the Nonexperimental Lalonde Data: Estimates of Average Treatment Effects for Pseudo Outcomes

	Blocking		Matching	
Pseudo Outcome	Estimated	(Standard Error)	Estimated	(Standard Error)
Earn '75	–1.22	(0.25)	–1.24	(0.30)
(Earn '74 + earn '75)/2	–6.13	(0.49)	–6.37	(0.67)

Subclasses for the nonexperimental Lalonde data

Subclass	p-score		Number of Controls	Number of Treated	Average p-score		Difference in Average p-score	t-stat
	Minimum	Maximum			Controls	Treated		
1	0.00	0.37	31	7	0.20	0.25	0.05	1.75
2	0.37	0.43	5	7	0.39	0.40	0.00	0.39
3	0.43	0.46	26	22	0.44	0.44	0.00	0.18
4	0.46	0.53	36	36	0.50	0.50	0.00	0.51
5	0.53	1.00	87	113	0.57	0.58	0.01	1.14

$$E[e(x)|W = 1] = 0.53$$

$$E[e(x)|W = 0] = 0.47$$

Nonexperimental Lalonde Data: Estimates of Average Treatment Effects

	Full Sample		Trimmed Sample			
Covariate	1 Block	Match	1 Block	2 Blocks	4 Blocks	Match
Number	–8.50	1.72	1.72	1.81	1.79	1.98
	(0.58)	(0.90)	(0.74)	(0.75)	(0.76)	(0.85)
Few	0.69	1.73	1.81	1.80	2.10	1.98
	(0.59)	(0.90)	(0.73)	(0.73)	(0.75)	(0.85)
All	1.07	1.81	1.97	1.90	1.93	2.06
	(0.55)	(0.90)	(0.66)	(0.67)	(0.70)	(0.85)