

Métodos de puntuación de propensión y técnicas de aprendizaje automático para el equilibrio de covariables en un estudio observacional retrospectivo

VIII Jornadas de Usuarios de R

Elvira Ferre Jaén¹ Antonio J. Fernández L.² Antonio Maurandi L.³

{elvira¹, antoniojose.fernandez², amaurandi³}@um.es

Sección de Apoyo Estadístico, Servicio de Apoyo a la Investigación, Universidad de Murcia

17 de noviembre de 2016, Albacete



Propensity Score Analysis



Motivación

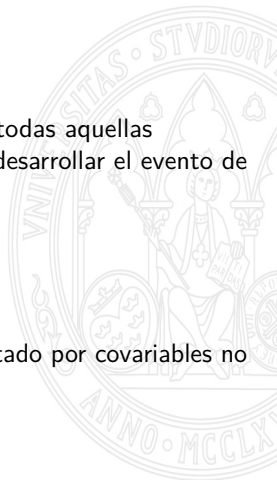
- Los estudios observacionales carecen de aleatorización
- Grupos de estudio no homogéneos en la distribución de sus características basales
- Posible existencia de variables de confusión o modificadoras de efecto
- Las estimaciones obtenidas pueden verse afectadas por
 - sesgos de confusión y selección

¿Qué es un índice de propensión?

- **Probabilidad** que tiene cada participante de ser asignado a cada rama del estudio según sus características basales
- **Estudios experimentales**
 - probabilidad fija y conocida de antemano
- **Estudios observacionales**
 - probabilidad desconocida
 - se calcula a partir de los datos disponibles



- Técnica estadística que trata de
 - formar grupos homogéneos en la distribución de todas aquellas características que puedan influir en el riesgo de desarrollar el evento de estudio
- Permittiéndonos
 - controlar y reducir los sesgos
 - evitar que el efecto de un tratamiento se vea afectado por covariables no controladas



Pasos del análisis

- ① Estimación del Índice de Propensión
- ② Control de sesgos
 - de confusión
 - de selección
- ③ Efecto del tratamiento



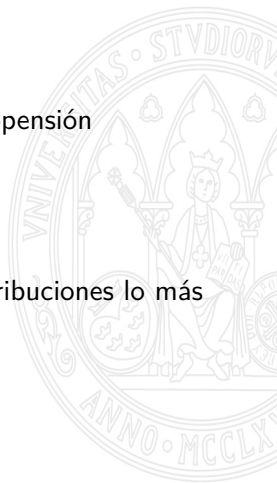
(1) Estimación del índice de propensión

- Técnicas principales:
 - Support Vector Machine
 - Regresión logística
 - Bagging y Random forest
 - Árboles de clasificación y regresión (CARTS)
- Aplicación:
 - comprobar si ambas distribuciones son muy diferentes
 - existencia de un solapamiento sustancial entre ambos grupos



(2) Control de sesgos

- Agrupar los participantes según sus índices de propensión
- Métodos:
 - Estratificación
 - Emparejamiento
- **Objetivo:** conseguir grupos homogéneos con distribuciones lo más similares posibles en las características basales.



(3) Efecto del tratamiento

- Tras lograr un **equilibrio** aceptable en las covariables
- Se pueden estimar los **efectos causales del tratamiento** entre los grupos emparejados



Aplicación del análisis



Estudio observacional retrospectivo

- Comparar dos técnicas quirúrgicas:
 - 239 pacientes técnica BGL
 - 61 pacientes técnica GVL
- Comparativa de las características basales:
 - T-student, χ^2 , diferencias estandarizadas
 - desequilibrio entre los grupos de tratamiento
 - sesgos de covarianza en las variables



Cálculo del índice de propensión: técnicas

Tabla 1: Comparativa técnicas PSM. Emparejamiento 1:2

	Librería de R	Tamaño muestral	Balanceo	ASAM	OVL	AUC
SVM	kernlab	154	27	0.0805	476	835
Reg. logística	glm	182	21	0.1679	428	839
Random Forest	randomForest	183	15	0.233	0	0
Baggin	ipred	183	16	0.2143	0	0
CART	twang	113	17	0.2488	219	961

Gráfico comparativo técnicas para calcular los IP

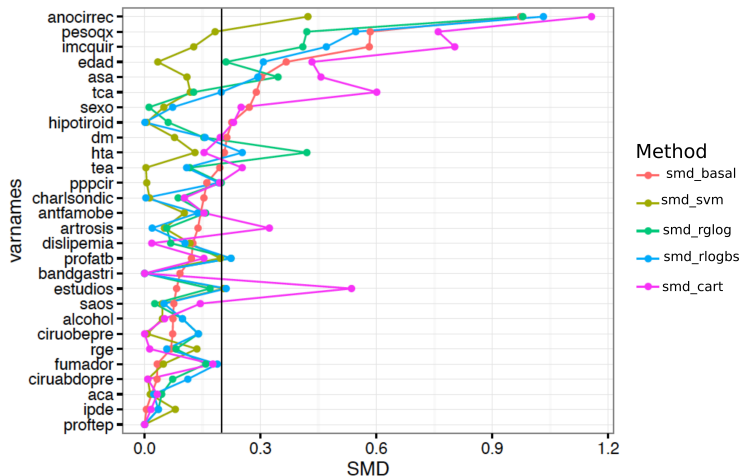


Figura 1: Comparativa técnicas PSM

Cálculo del IP mediante SVM

```
library( kernlab )  
model.ksvm <- ksvm( tecnica ~ ., data = dfpsm,  
                    kernel = "rbfdot", C = 25,  
                    kpar = list( sigma = 0.001),  
                    prob.model = TRUE )  
  
# Probabilidades predichas  
predProb <- predict( model.ksvm, newdata = dfpsm,  
                     type = "probabilities" )  
  
# Crear IP y guardarlo  
psSVM <- predProb[1:300, 2] # Prob. de GVL  
dfpsm <- cbind( dfpsm, psSVM )
```

Solapamiento

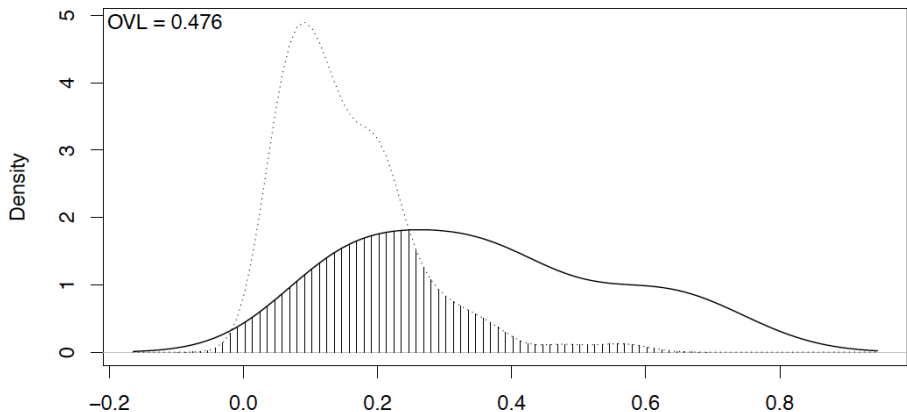


Figura 2: Densidades Índices de Propensión

Emparejamiento

- **Objetivo:** crear conjuntos de sujetos de cada grupo con IP similares.
- Realizamos un emparejamiento de 2 BGL por cada 1 GVL
- Técnica del “vecino más próximo”
 - caliper de 0.2

```
library( nonrandom )  
fCirc_mat <- ps.match( dfpsm, ratio = 2,  
                        matched.by = "psSVM",  
                        caliper = 0.2,  
                        treat = "tecnica",  
                        setseed = 270156 )
```


Resultado del emparejamiento

Tabla 2: Emparejamiento mediante SVM

	BGL	GVL	Total
Total	239	61	300
Emparejados	101	53	154
No emparejados	138	8	146

Equilibrio

- Diferencia de medias estandarizada **antes y después** del emparejamiento
- Punto de equilibrio: 20

```
fCirc_bal <- ps.balance( fCirc_mat,  
                        sel = c(1:28),  
                        treat = "tecnica",  
                        method = "stand.diff",  
                        cat.levels = 10,  
                        alpha = 20,  
                        equal = TRUE )  
  
plot.std(fCirc_bal )
```

Gráfico equilibrio

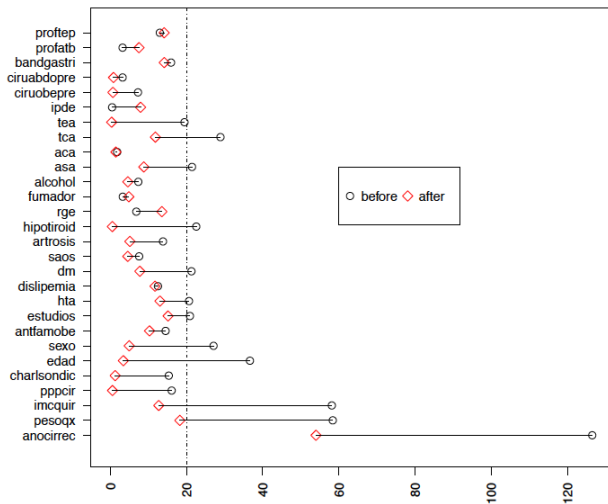
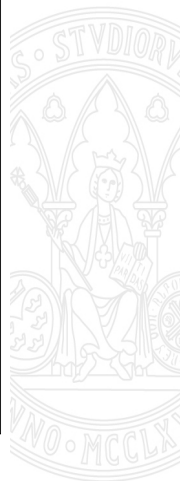


Figura 3: Equilibrio de las covariables



Gracias por su atención

Métodos de puntuación de propensión y técnicas de aprendizaje automático para el equilibrio de covariables en un estudio observacional retrospectivo

VIII Jornadas de Usuarios de R

Elvira Ferre Jaén¹ Antonio J. Fernández L.² Antonio Maurandi L.³

`{elvira1, antoniojose.fernandez2, amaurandi3}@um.es`

Sección de Apoyo Estadístico, Servicio de Apoyo a la Investigación, Universidad de Murcia

17 de noviembre de 2016, Albacete

Referencias

Karatzoglou, Alexandros, Alex Smola, Kurt Hornik, and Achim Zeileis. 2004. "Kernlab – An S4 Package for Kernel Methods in R." *Journal of Statistical Software* 11 (9): 1–20.

<http://www.jstatsoft.org/v11/i09/>.

Stampf, Susanne. 2014. *Nonrandom: Stratification and Matching by the Propensity Score*.

<https://CRAN.R-project.org/package=nonrandom>.