MEDICIÓN DE LA DISPARIDAD SALARIAL EN LAS REGIONES DEL DEPARTAMENTO DE LA PAZ

M. Sc. Abalos Choque Melisa
melisa.abalos@gmail.com

RESUMEN

Las disparidades salariales han sido objeto de estudio en países europeos y americanos con el fin de aplicar políticas que permitan reducir dichas diferencias. El presente artículo desarrolla teóricamente los métodos de Propensity Score Matching (PSM) y Coarsened Exact Maching (CEM) a fin de mostrar al investigador una alternativa para medir el impacto en investigaciones de tipo social. En tal sentido, se utiliza la teoría de inferencia causal para medir la diferencia salarial entre la población que reside en la región Metropolitana versus las demás regiones del departamento de La Paz mediante el método CEM con datos de la encuesta Sociodemográfica 2010-2012. Estos resultados serán útiles para futuras investigaciones que busquen determinar las causas teóricas de los resultados aquí encontrados.

PALABRAS CLAVE

Disparidad, Inferencia Causal, CEM

1. INTRODUCCIÓN

La literatura ha distinguido la existencia de diferencias espaciales de salario en países como; Chile, España y México, entre otros. Las disparidades salariales podrían deberse a la composición de habilidades de los trabajadores, dotaciones no humanas (como las características geográficas) e interacción entre trabajadores y empresas (Combes, Duranton & Gobillon, 2007).

Para el caso del departamento de La Paz, aún no existe evidencia de estudios realizados para medir las diferencias de salario en espacios geográficos definidos. Medir la existencia o no de diferencias espaciales es un reto que podría ser afrontado con técnicas estadísticas.

La inferencia causal es una rama de la Estadística que estudia las técnicas de estimación de una variable de interés que puede variar gracias a la aplicación de un tratamiento específico y no a otros factores. Éstas técnicas podrían aplicarse mediante datos experimentales y no experimentales.

Los diseños experimentales son métodos estadísticos que nos permiten comparar dos grupos a fin de medir la causalidad de una variable específica. Estos métodos aseguran la equivalencia inicial la cual implica que los dos grupos son similares entre sí al momento de iniciarse el experimento. Sin embargo, en muchos casos, los diseños experimentales demandan costos elevados que no podrían ser asumidos por el investigador. Winship & Morgan 1999, sugieren utilizar datos observables de censos, encuestas, o registros administrativos para medir la causalidad cuando es imposible realizar experimentos aleatorios.

Los diseños transeccionales, que son parte de los diseños no experimentales, nos permiten obtener información en un momento dado. La debilidad de utilizar éste tipo de información sin un criterio previo es que las diferencias salariales podrían deberse a aspectos propios del individuo y no así al tratamiento.

En tal sentido, se debe garantizar la homogeneidad de los individuos estudiados antes de realizar las mediciones en la variable de interés. Rubin, Rosenbaum y Heckman, entre otros conceptualizan la utilización de un modelo contrafactual para resolver el problema de la heterogeneidad existente entre el grupo de tratamiento y el de control. Así los métodos contrafactuales se conciben como una solución factible para el fenómeno de no comparabilidad.

Un modelo contrafactual consiste en medir la diferencia de la variable de interés con un individuo que esté tanto en el grupo de tratamiento como en el de control al mismo tiempo, lo que es físicamente imposible. Sin embargo, podemos hacer esta comparación entre dos individuos, del grupo de tratamiento y del grupo de control que tengan características similares. En este contexto se utilizan los métodos mátching los cuales consideran éstos aspectos.

Los métodos llamados matching o emparejamiento consisten en parear el resultado de la variable de interés del individuo tratado con el resultado de otro individuo perteneciente al grupo de control mismo que comparte características similares o cercanas al individuo tratado. El presente artículo desarrolla teóricamente los métodos de Propensity Score Matching (PSM) y Coarsened Exact Maching (CEM). Por otro lado, mide la diferencia salarial existente entre la población que reside en la región Metropolitana versus las demás regiones del departamento de La Paz mediante el CEM el cual garantiza un matching exacto. Estos resultados serán útiles para futuras investigaciones que busquen determinar las causas teóricas de los resultados aquí encontrados.

La sección 2 del artículo muestra la teoría de la inferencia causal incluyendo los métodos matching a desarrollar PSM y CEM. La sección 3 desarrolla las virtudes de la información utilizada. La sección 4 muestra los resultados obtenidos y finalmente la sección 5 presenta las conclusiones.

2. INFERENCIA CAUSAL

Los métodos de inferencia causal estudian como estimar el efecto en una variable de interés que se debe exclusivamente al tratamiento y no a otros factores. Entonces, se puede definir a la inferencia causal como una rama de la Estadística que estudia la obtención de efectos causales los cuales podrían realizarse mediante datos experimentales y no experimentales.

Los datos experimentales se obtienen mediante la realización de experimentos aleatorios. Un experimento aleatorio es controlado por el investigador, el cual asigna, de manera aleatoria, a los individuos a ser estudiados, al grupo de tratamiento o a un grupo de control[1]. La asignación aleatoria garantiza la equivalencia inicial que deben tener los dos grupos, o sea, que los grupos son similares al inicio del experimento. Sin embargo, la realización de experimentos aleatorios en muchos casos implica altos costos.

La estimación de efectos causales mediante datos observables, tales como, encuestas, censos y registros administrativos, es utilizada como una alternativa cuantitativa a la imposibilidad de diseñar experimentos aleatorios (Winship & Morgan, 1999). En tal sentido, consideremos que y_i es el valor del salario para el individuo i.

1 El grupo de control esta formado por los individuos que no han recibido el tratamiento.

Se generan dos resultados potenciales, el primero en ausencia del tratamiento al individuo *i* denotado por y_{0i} y el segundo en presencia del tratamiento para el individuo *i* denotado por y_{1i}. Los resultados potenciales se presentan en situaciones contrafactuales ya que y_{1i} representa el salario hipotético que hubiera obtenido el individuo *i* (que no ha sido sometido al tratamiento) si hubiera sido sometido al tratamiento.

El efecto causal del tratamiento para el individuo *i* se define como la diferencia entre ambos resultados potenciales dado en (1).

$$\Delta = y_{1i} - y_{0i} \qquad (1)$$

En la realidad, no es posible observar ambos resultados potenciales, sino que únicamente se observamos el resultado y_i que se puede expresar en (2), donde D_i es una variable dummy que indica si el individuo *i* ha recibido el tratamiento o no.

$$y_i = D_i y_{1i} + (1 - D_i) y_{0i} \qquad (2)$$

En caso de que existiera homogeneidad[2] entre todos los individuos incluidos en el estudio, el problema podría estar resuelto. Sin embargo, generalmente existe alto grado de heterogeneidad en las características de los individuos y de sus respuestas. Por lo que no es garantizado que la diferencia refleje el efecto del tratamiento. Estudios realizados en esta área muestran que se puede calcular efectos individuales de tratamiento mediante el llamado efecto medio del tratamiento (ATE) y el efecto medio del tratamiento seleccionado (SATE).

La homogeneidad entre los grupos se garantiza mediante un vector *X* de variables observables, tal que el tratamiento es independiente de los resultados potenciales condicionados en dichas variables. En tal sentido, por independencia condicional entre el tratamiento y los resultados potenciales, la esperanza matemática (E[ ]) de la diferencia del salario dado el vector X está dado por (3).

$$\begin{aligned} E\left[y_1 - \frac{y_0}{X}\right] &= E\left[y_1 - \frac{y_0}{X}, D = 1\right] \\ &= E\left[\frac{Y}{X}, D = 1\right] \\ &= 1 - E\left[\frac{Y}{X}, D = 0\right] \end{aligned} \qquad (3)$$

Los métodos llamados matching o emparejamiento son métodos no paramétricos que han sido estudiados recientemente y consisten en emparejar el resultado de cada individuo tratado con el resultado de otro individuo que no pertenece al grupo de tratamiento, pero que comparte características similares o cercanas al individuo tratado, dichas características se pueden incluir en el vector *X*. En el presente artículo se exponen dos estrategias econométricas relacionadas con los métodos matching los cuales son: Coarsened Exact Maching (CEM) y Propensity Score Matching (PSM) los cuales garantizan la comparabilidad entre grupos (Enriquez & Paredes, 2014).

2.1 PROPENSITY SCORE MATCHING (PSM)

El Propensity Score Matching, *e(x)*, es la probabilidad condicional de recibir tratamiento (*D=1*) dado un vector de variables observables *X*, es decir, *e(x)=P(D=1/X)*. El propensity Score generalmente es desconocido, sin embargo, se puede estimar mediante una regresión logística binaria logit o probit (Winship & Morgan, 1999). El modelo logit asociado esta dado en (4), donde D_i es la condición de tratamiento que toma un valor 1 si el individuo está en el grupo de tratamiento y 0 si el individuo está en el grupo de control para cada individuo *i=1,2,...,N,*

2 Esto es lo que se pretende en un experimento aleatorio.

el vector de variables observables es X_i y el vector de regresión de parámetros es β_i.

$$P\left(\frac{D_i}{X_i} = x_i\right) = \frac{e^{x_i\beta_i}}{1 + e^{x_i\beta_i}} = \frac{1}{1 + e^{-x_i\beta_i}} \quad (4)$$

Aunque D_i no es una función lineal de X_i ésta es una variable transformada a través de una función Logit. La condición es que el tratamiento es independiente a los resultados potenciales, el cual se logra condicionando un vector X de variables observables.

El proceso se puede resumir en tres pasos:

Primero, se estima el Propensity Score (PS) mediante el modelo de elección binaria. Se espera que la distribución conjunta entre todas las variables observables sea igual para cada grupo (tratamiento y control). Con ese propósito, los valores estimados del PS se dividen en estratos, contrastando las diferencias existentes entre los grupos. Si las diferencias no son significativas se acepta la distribución conjunta, caso contrario, se añaden potencias de orden superior de las variables observables e interacción entre las mismas hasta que se acepte la igualdad de la distribución conjunta de variables observables entre ambos grupos.

Segundo, las observaciones se ordenan en forma ascendente de acuerdo al valor estimado del PS. Los valores estimados extremos del PS del grupo de comparación son descartados. Con las observaciones restantes, se procede al emparejamiento de cada individuo del grupo de tratamiento con el grupo de control que tenga Propensity Score más próximo.

Finalmente, se calcula el efecto medio del tratamiento ATE mostrado en (5), donde Y_i es el salario de los individuos pertenecientes al grupo de tratamiento y de control.

$$E[\Delta] = E\left[\frac{Y_1}{\hat{e}(x)}(x), D = 1\right] = E\left[\frac{Y_0}{\hat{e}(x)}(x), D = 0\right] \quad (5)$$

2.2 COARSENED EXACT MATCHING (CEM)

CEM es parte de los métodos matching conocidos como "Monotonic Imbalance Bounding" (MIB) que muestra mejores rendimientos que los otros métodos matching (Iacus, King, Porro, 2011). El balance entre el grupo de control y tratamiento implica que la distribución conjunta de las covariantes X en más similar entre los dos grupos. CEM trabaja en muestras y no requiere supuestos sobre el proceso generador de datos, es un método que mejora el balance entre el grupo de control y tratamiento, por tanto, el desequilibrio no será mayor a la que el investigador puede definir ex ante.

Para medir el imbalance se construye un histograma multidimensional de las celdas generadas por un producto cartesiano $H(X_1)x\ldots xH(X_k)=H(X)$. El imbalance multivariado se muestra en (6), donde, f y g son las distribuciones de frecuencia relativa para el grupo de tratamiento y control respectivamente.

$$L_1(f,g) = \frac{1}{2}\sum_{l_1\ldots l_k \epsilon H(X)} |f_{l_1\ldots l_k} - g_{l_1\ldots l_k}| \quad (6)$$

Si, f^m y g^m son las distribuciones de frecuencia relativa para emparejar el grupo de tratamiento y control respectivamente. Entonces, un buen método matching cumple que $\mathcal{L}_1(f^m, g^m) \leq \mathcal{L}_1(f,g)$. Si $\mathcal{L}_1 = 1$ se dice que las dos distribuciones de datos son completamente diferentes y si $\mathcal{L}_1 = 0$, las distribuciones son exactamente iguales. Si por ejemplo $\mathcal{L}_1 = 0{,}7$, solamente el 30% de la densidad de los dos histogramas se solapan.

El algoritmo CEM puede resumirse en tres pasos. *Primero*, se recodifica cada variable de tal manera que los valores indistinguibles se agrupan y se asigna el mismo valor numérico (coarsar). *Segundo*, el algoritmo CEM crea un set de estratos, sϵS, cada uno con el mismo valor coarsado de *X* y con por lo menos una unidad de tratamiento y una unidad de control. Se aplica el algoritmo matching exacto a los datos coarsados para determinar los emparejamientos. *Tercero*, los estratos solamente con unidades de control o tratamiento se eliminan y se toman en cuenta solamente los estratos que tienen al menos una unidad de tratamiento y una unidad de control. El algoritmo CEM asigna los pesos mostrados en (7), donde, T^s son las unidades tratadas en el estrato *s*, m_T^s es la cantidad de unidades tratadas en el estrato *s*, C^s son las unidades de control en el estrato s, m_C^s es la cantidad de unidades de control en el estrato s, m_T, m_C son el número de unidades emparejadas en grupo de tratamiento y control respectivamente. Los valores no emparejados reciben un peso igual a cero, o sea, $w_i=0$, por tanto son eliminados. En tal sentido, CEM impone un matching exacto entre individuos evitando problemas con imposiciones arbitrarias de medidas de distancia.

$$w_i = \begin{cases} 1, & i \in T^s \\ \dfrac{m_C}{m_T}\dfrac{m_T^s}{m_C^s}, & i \in C^s \end{cases} \qquad (7)$$

El efecto medio de tratamiento de la muestra (SATT) y de la población (PATT) se presentan en (8), donde, $TE_i=Y_i(1)-Y_i(0)$, es el efecto de tratamiento de la unidad *i*, n_T es la cantidad de elementos en el set de índices de unidades tratadas *T* en la muestra, *T** es el set de índices de unidades tratadas de la población y N_T es la cantidad de elementos en *T**. *SATT* es un estimador insesgado de *PATT* de tal forma que *E(SATT)=PATT*.

$$SATT = \frac{1}{n_T}\sum_{i\in T} TE_i \ ,$$

Tabla1
Departamento de La Paz: Distribución de municipios por región

REGIÓN	MUNICIPIO			
Amazonia	Ixiamas	San Buenaventura	Caranavi	Alto Beni
	Guanay	Tipuani	Mapiri	Teoponte
	Apolo			
Metropolitana	Viacha	Laja	La Paz	Palca
	Mecapaca	Achocalla	El Alto	
Yungas	Coroico	Coripata	Chulumani	Irupana
	Yanacachi	Palos Blancos	La Asunta	
Altiplano Sur	Sica	Umala	Ayo Ayo*	Calamarca
	Patacamaya	Colquencha	Collana	Papel Pampa*
	S.P. de Curahuara*	Santiago de Machaca*	San Andrés de Machaca*	Jesús de Machaca
	Chacarilla*	Catacora*	Coro Coro	Caquiaviri
	Calacoto	Comanche	Charaña*	Waldo Ballivián*
	Nazacara de Pacajes	Santiago de Callapa		
Altiplano Norte	Puerto Acosta	Puerto Carabuco	Humanata*	Escoma
	Guaqui	Tiahuanacu	Desaguadero	Taraco
	Pucarani	Batallas	Puerto Pérez	Copacabana
	San Pedro de Tiquina	Tito Yupanqui	Achacachi	Ancoraimes
	Huatajata	Huarina	Santiago de Huata	Chua Cocani
Valles Norte	Charazani	Curva	Mocomoco	Pelechuco
	Sorata	Tacacoma	Quiabaya	Combaya
	Chuma	Ayata	Aucapata	
Valles Sur	Inquisivi	Quime	Cajuata	Colquiri
	Ichoca	Villa Libertad Licoma	Luribay	Sapahaqui
	Yaco	Malla	Cairoma	

Fuente: Plan de Desarrollo del Departamento Autónomo de La Paz al 2020
*Municipios que no fueron encuestados
Elaboración Propia

$$PATT = \frac{1}{N_T} \sum_{i \in T^*} TE_i \qquad (8)$$

3. DATOS

El presente artículo utiliza los datos de la encuesta sociodemográfica realizada por la Universidad Mayor de San Andrés y el Gobierno Autónomo Departamental de La Paz entre las gestiones 2010 y 2012.

El objetivo de la encuesta fue conformar una línea base de información estadística que permita realizar un diagnóstico de la situación socio – demográfica y productiva para el planteamiento de políticas y la planificación de los municipios de las 7 regiones del Departamento de La Paz.

Según el Plan de Desarrollo del Departamento Autónomo de La Paz al 2020, el departamento de La Paz tiene 7 regiones de planificación las cuales son; Altiplano Norte, Altiplano Sur, Yungas, Amazonía, Valles Norte, Valles Sur y Metropolitana. La Tabla 1 muestra la conformación de los municipios incluidos en cada una de las regiones.

La información contenida en la Base de datos de la encuesta Sociodemográfica es representativa para hallar resultados de los municipios, provincias y regiones del Departamento de La Paz, en tal sentido proporciona estimadores robustos a los niveles mencionados. La muestra es de 33.011 personas del departamento de La Paz que forman parte de la población económicamente activa[3] y que declaran tener algún salario. De la muestra total del departamento de La Paz, 4.833 corresponden a la región Altiplano Sur, 5.341 a Altiplano Norte, 4.295 a Yungas, 4.154 a Amazonía, 3.657 a Valles Norte, 3.932 a Valles Sur y 6.799 a la región Metropolitana. El total de población incluida en la base de datos de la encuesta es de 103.054 personas.

4. RESULTADOS

Las disparidades salariales se remontan a diferencias en la composición de habilidades de la fuerza laboral (Combes, Duranton & Gobillon, 2008). En tal sentido, el vector de variables observables X está conformada por variables cuantitativas y cualitativas que suponemos que inciden en las variables salario las cuales son; edad, años promedio de escolaridad, años de experiencia[4] , genero, estado civil, área, relación de parentesco y categoría ocupacional. El grupo de tratamiento está conformado por los individuos que viven en la región Metropolitana y los grupos de control son los individuos que viven en otras regiones del departamento de La Paz.

La Tabla 2 muestra la diferencia de la media y la desviación estándar entre la región Metropolitana y el resto de las regiones, de las variables utilizadas para el emparejamiento. A primera vista el promedio del salario[5] es mayor en la región metropolitana al resto de las regiones. Las variables cuantitativas presentan diferencias aparentemente significativas. Las variables Dummy relacionadas con el género, estado civil, relación de parentesco con el jefe de hogar y la categoría ocupacional presentan, de la misma manera, muestran diferencias

3 Se considera a la Población Económicamente Activa a las personas mayores a 10 años que declararon trabajar al menos una hora la semana pasada a la encuesta y aquellos que no trabajaron por licencia o que atendieron o ayudaron en cultivos agrícolas o en algún negocio familiar o buscaron trabajo.

4 La experiencia (Exp) es igual a la edad (Edad) menos los años promedio de escolaridad (Esc) menos seis (Exp=Edad-Esc-6)..

5 Se eliminaron valores atípicos encontrados por debajo del percentil 1 y por encima del percentil 99.

Tabla 2
Departamento de La Paz: Media y Desviación estándar de las variables para emparejamiento por región

Variable	Metropolitana				Resto Regiones			
	Media/ proporción	Std. Dev.	Min	Max	Media/ proporción	Std. Dev.	Min	Max
Salario	2.399,44	1.843,55	170	10.500	1.477	1.719	17,32	21.217
Edad	39,79	13,54	12	93	42	16	10	98
Escolaridad	10,97	4,64	0	25	7	5	0	25
Experiencia	23,25	15,69	0	93	29	19	0	93
Hombre	56,1%	0,496	0	1	62,3%	0,485	0	1
Mujer	43,9%	0,496	0	1	37,7%	0,485	0	1
Soltero	24,7%	0,432	0	1	17,4%	0,379	0	1
Casado	56,1%	0,496	0	1	59,4%	0,491	0	1
Jefe de Hogar	50,5%	0,500	0	1	55,4%	0,497	0	1
Esposa(o)	22,2%	0,416	0	1	22,8%	0,420	0	1
Hijo	22,1%	0,415	0	1	18,3%	0,387	0	1
Obrero/Empleado	46,6%	0,499	0	1	20,7%	0,405	0	1
Cuentapropista	50,3%	0,500	0	1	75,6%	0,429	0	1
Patrón Socio o Empleador	1,5%	0,123	0	1	2,2%	0,148	0	1

Fuente: Elaboración propia con datos de la encuesta sociodemográfica (2010 - 2012)

entre ambos contextos geográficos. Las diferencias apreciadas en las variables para el emparejamiento son evidencia preliminar de que la diferencia del salario podría deberse a otros factores asociados a la productividad (Enriquez & Paredes, 2014) y no así a la condición de vivir en una región específica.

Las diferencias en media pueden ser encontradas con un análisis de regresión simple. La Tabla 3 muestra los resultados de la regresión lineal simple utilizando como variable dependiente al salario y como variable independiente a la variable dummy (Reg 7) que toma el valor 1 si la persona pertenece a

Tabla 3
Diferencia de salario de la población entre la región metropolitana vs otras regiones sin emparejamiento

Source	SS	df	MS
Model	1.1673e+09	1	1.1673e+09
Residual	6.4425e+11	33011	19516225.8
Total	6.4542e+11	33012	19550995.9

Number of obs	= 33013
F(1, 33011)	= 59.81
Prob > F	= 0.0000
R-squared	= 0.0018
Adj R-squared	= 0.0018
Root MSE	= 4417.7

salario_mes	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Reg7	465.0006	60.12447	7.73	0.000	347.1545	582.8467
_cons	1894.962	27.28546	69.45	0.000	1841.481	1948.442

Fuente: Elaboración propia con datos de la encuesta sociodemográfica (2010 - 2012)

la región metropolitana y 0 si pertenece a otra región. Los resultados indican que el vivir en la región metropolitana tiene una prima positiva de 465 Bs a diferencia de vivir en otras regiones. Sin embargo, estos resultados no toman en cuenta las variables propias de cada individuo que podrían estar afectando la variable salario. El presente artículo utiliza un método de emparejamiento que considera a individuos con características similares con el propósito de eliminar las diferencias entre covariantes.

Tabla 4
Cálculo del imbalance antes y después del emparejamiento

Covariantes	Antes de emparejar		Después de emparejar	
	L1	Diferencia en Media/ proporción	L1	Diferencia en Media/ proporción
Edad	0,08205	-1,31760	0,03018	-0,01409
Escolaridad	0,26994	2,49870	0,03925	0,0024
Experiencia	0,10223	-3,81630	0,01595	0,01143
Hombre	0,03423	-0,03423	1,4E-14	-2E-14
Mujer	0,03423	0,03423	1,4E-14	-7,4E-15
Soltero	0,04208	0,04208	2,2E-14	-1,7E-14
Casado	0,01235	0,01235	2E-14	-1,5E-14
Jefe de Hogar	0,01054	-0,01054	1,9E-14	-1,2E-14
Esposa(o)	0,00507	-0,00507	3E-14	1,7E-15
Hijo	0,01298	0,01298	3,4E-14	-5,3E-14
Obrero/Empleado	0,17113	0,17113	2,9E-14	-2,4E-15
Cuentapropista	0,15518	-0,15518	2,8E-14	-1E-14
Patrón Socio o Empleador	0,01586	-0,01586	3E-14	4,9E-17
Distancia Multivariada		**0,5573**		**0,2576**
Estratos (Total-pareados)		9.222		2.719

	Metropolitana	Otra Región	Total	%
Total	18.280	84.774	103.054	100%
Pareados	17.080	70.070	87.150	85%
No pareados	1.200	14.704	15.904	15%

Fuente: Elaboración propia con datos de la encuesta sociodemográfica (2010 - 2012)

El estadístico $\mathcal{L}_1$ mide el imbalance entre la distribución conjunta de las covariantes entre el grupo de control y tratamiento. La Tabla 4 muestra el cálculo de los imbalances antes y después del emparejamiento, el valor de la distancia multivariada reduce 0,56 a 0,26 lo que muestra que un buen emparejamiento produce una reducción en el sesgo de la diferencia de medias. Por otro lado, los imbalances univariados reducen drásticamente después del emparejamiento. También se aprecia que el total de individuos pareados es de 87.150 que corresponde al 85% de la muestra de la población total. Una vez realizado el emparejamiento, utilizamos los pesos generados por el CEM para medir el SATT con los valores pareados. La tabla 5 muestra la diferencia de salario entre el

Tabla 5
Diferencia de salario de la población entre la región metropolitana vs otras regiones con emparejamiento

Source	SS	df	MS
Model	389821150	1	389821150
Residual	3.1226e+11	29171	10704334
Total	3.1265e+11	29172	10717329.9

Number of obs	= 29173
F(1, 29171)	= 36.42
Prob > F	= 0.0000
R-squared	= 0.0012
Adj R-squared	= 0.0012
Root MSE	= 3271.7

salario_mes	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Reg7	282.7978	46.8618	6.03	0.000	190.9465	374.649
_cons	2032.655	21.57889	94.20	0.000	1990.36	2074.951

Fuente: Elaboración propia con datos de la encuesta sociodemográfica (2010 - 2012)

grupo de tratamiento y control después del emparejamiento. El SATT es de Bs. 283, mismo que es significativo, lo cual expresa que el individuo que vive en la región metropolitana gana más que un individuo que vive en otras regiones del departamento de La Paz.

5. CONCLUSIONES

La inferencia causal estudia la estimación de efectos causales de una variable de interés. La medición de efectos causales podría realizarse mediante datos experimentales y no experimentales. La estimación de efectos causales mediante datos observables, tales como, encuestas, censos y registros administrativos, es utilizada como una alternativa cuantitativa a la imposibilidad de diseñar experimentos aleatorios (Winship & Morgan, 1999). Los métodos matching son una alternativa para medir efectos causales mediante datos observables.

El presente artículo aplica el método de Coarsened Exact Matching (CEM) que son parte de los métodos matching para medir la diferencia del salario entre los habitantes de la Región Metropolitana (Tratamiento) versus los habitantes de otras regiones (Control) del departamento de La Paz. Los resultados descritos en la sección 4 muestran que el CEM es un método de emparejamiento que en primer lugar disminuye el desbalance entre covariantes propias del individuo, lo que asegura una homogeneidad entre grupos, mostrando una mejora en el estadístico L_1. Lo cual implica que utilizar este método puede reducir los sesgos de la diferencia de medias.

En segundo lugar, nos permite utilizar los datos de la encuesta Socio demográfica para medir la diferencia de medias que no sean afectadas por variables particulares de cada individuo y que pueden influir en la variable salario por tanto se usa datos transeccionales y no se depende de un experimento aleatorio. En tercer lugar, se verifica que existe una diferencia en el salario de la población que vive en la región Metropolitana versus otras regiones. El vivir en la región metropolitana, tiene un premiun de Bs. 283 más que vivir en otras regiones.

La Región Metropolitana contiene a las ciudades de La Paz y El Alto, que según el Censo Nacional de Población y Vivienda 2012 juntos tienen una población de 1.605.725 habitantes y representa 59% de la población total del departamento. Esta aglomeración de los habitantes podría estar generando economías a escala, que podrían explicar el premio de vivir en la región metropolitana, hecho que puede ser estudiado posteriormente mediante las teorías de economías de aglomeración.

BIBLIOGRAFIA

Plan de Desarrollo del Departamento Autónomo de La Paz al 2020.

Blackwell, Iacus, King & Porro (2010), "*Coarsened Exact Matching in Stata"*.

Castro (2005), "*Salarios y desigualdad territorial en las áreas urbanas de México, 1992-2002"*, Tesis de doctorado en economía, Universidad Autónomo de Barcelona.

Combes, Duranton & Gobillon, (2008), "*Spatial wage disparities: Sorting matters*", Journal of Urban Economics 63, pp. 723-742.

Enriguez & Paredes (2014), "*Migración Interna y Diferenciales de Ingreso: Evidencia para Bogotá (Colombia) a Partir de Métodos de Emparejamiento*", Revista de economía & Administración, Vol. 11, N°1, pp. 65-83.

García, I (2009), "*Metodología y diseño de estudios para la evaluación de políticas públicas"*, Barcelona – España.

Garza & Quintana (2014), "*Determinantes de la desigualdad salarial en las regiones de México 2005-2010, una visión alternativa a la teoría del capital humano*", paradigma económico, año 6, núm. 1, pp. 33-48.

Guo & Fraser (), "A*dvanced Quantitative Techniques in the Social Sciences Series"*, Propensity Score Marching, pp. 127-208.

Iacus, King & Porro, 2011, "*Causal Inference without Balance Checking: Coarsened Exact Matching"*, Oxford University Press on behalf of the society for political methodology.

Iacus, King & Porro, 2011, "*Multivariate Matching Methods That Are Monotonic Imbalance Bounding"*, Journal of the American Statistical Association, Vol. 106, No. 493, Theory and Methods, pp. 345-361

Khandker, Koolwal & Samad, 2010, "*Handbook on Impact Evaluation-Quantitative Methods and Practices*", The International Bank for Reconstruction and Development / The World Bank, Washington DC.

King, Nielsen, Coberley, Pope & Wells, (2011), "*Comparative effectiveness of Marching Methods for causal inference*".

Larraz & Pavia (2014), "*Concentración salarial en España, un análisis espacial*", International Conference On Regional Sciences.

Rosembaum & Rubin (1985), "*Constructing a Control Group Using Multivariate Marched Sampling Methods That Incorporate de Propensity Score*", American Stadistical Association, Vol. 39, N°1, pp. 33-38.

Winship & Morgan (1999), "*The estimation of causal effects from observational data"*, Rev. Social 25, pp. 659-706.

Winship & Morgan (2007), "*Conterfactuals and Causal Inference – Methods and Principles for Social Research"*, Cambridge University Press, United States of America.