Revista Electrónica de Metodología Aplicada
2004, Vol. 9 nº 2, pp. 1-14

Estimaciones bootstrap para el coeficiente de determinación: un estudio de simulación

López Jáuregui, A.[1] y Elosua Oliden, P.
Facultad de Psicología. Universidad del País Vasco.

RESUMEN

La aplicación del enfoque bootstrap permite obtener estimaciones de medidas de precisión así como la realización de contrastes de hipótesis en aquellas situaciones en las que no se dispone de información acerca de las distribución muestral de un estadístico o en casos en los que la distribución muestral es dependiente de parámetros desconocidos. Este es el caso del coeficiente de determinación, el índice de evaluación más generalizado en el contexto empírico, utilizado para evaluar el ajuste del modelo lineal de regresión. El objetivo de este trabajo es la exploración de la bondad de este procedimiento en la estimación del error de medida de R2 así como en la determinación de los intervalos de confianza, mediante el método percentil y el método Bca. Este análisis se lleva a cabo mediante una simulación Monte Carlo.

Palabras clave: Bootstrap, Estimación, Intervalos de confianza, coeficiente de determinación, simulación montecarlo.

ABSTRACT

The application of bootstrap approach allows the estimations of the measurements of precision and carry out contrast of hypothesis in situations in which we haven't got information about the sample distribution of one statistic or in situations in which the sample distribution is dependent on unknown parameters. This is the case for the determination coefficient, the most usual index in the empirical context used to evaluate the adjustment for the linear regression model. The aim of this study is to explore how well this procedure works in the estimation of the measurement error of R^2, and also, in the determination of confidence intervals using the percentile method and the Bca method This is achieved by applying Monte Carlo simulation.

Keywords: Bootstrap, estimation, detemination coefficient, confidence interval, Monte Carlo simulation..

[1] Alicia López Jáuregui; Facultad de Psicología, Avda. Tolosa, 70. 20018. San Sebastian

e-mail: psplojaa@sc.ehu.es

1.- Introducción

La metodología bootstrap debe su nombre y su formulación original a Bradley Efron (1979). Constituye la línea más desarrollada, tanto desde el punto de vista teórico como aplicado, de una variedad de técnicas para la inferencia estadística denominadas genéricamente "métodos de remuestreo" (Simon, 1969) entre las que se encuentran la permutación estocástica , el jackniffe (Quenouille, 1956; Tukey,1958) y la validación cruzada (Mosier, 1951). Son básicamente técnicas de simulación que reutilizan los datos observados para constituir un universo del cual extraer repetidas muestras. El requerimiento de gran potencia computacional común a todas ellas ha llevado a denominarlas técnicas de "computación intensiva" (Noreen, 1989).

La idea subyacente al bootstrap es simple: Los datos muestrales son tratados como si constituyesen los datos de toda la población, es decir se utilizan como el universo del que se extraerán muestras con reemplazamiento. Para cada remuestreo se calculará el valor del estimador bootstrap que se utilizará para estimar la variabilidad muestral. Tal y como los estudios teóricos han demostrado, este enfoque proporciona una buena aproximación de la distribución de los estimadores (Diaconis y Efron, 1983; Efron, 1981; Lunneborg, 1987) lo cual permitirá describir algunas de sus propiedades muestrales, así como el cálculo de intervalos de confianza y la realización de contrastes de hipótesis .

Al construirse empíricamente la distribución del estimador sobre la base de todas las características de la distribución original de los datos, incluyendo aquellos factores considerados como contaminantes (colas pesadas, outliers, etc.), el enfoque bootstrap está especialmente indicado en los casos en que los datos no siguen una distribución normal; hecho, que es común a la mayor parte de las medidas utilizadas habitualmente en las ciencias del comportamiento (Micceri, 1989).

El procedimiento bootstrap es útil para la descripción de la distribución muestral de aquellos estimadores con propiedades muestrales desconocidas o difícilmente obtenibles por medios analíticos. Por ejemplo, el bootstrap ha resultado efectivo en la estimación de la variabilidad de los coeficientes de estructura en el contexto del análisis discriminante (Dalgleish, 1994; Ramírez, 1996), o de la variabilidad de los valores propios, la estabilidad de los pesos factoriales y del coeficiente de congruencia dentro del análisis factorial (Chatterjee, 1984; Scott, Thompson y Sexton, 1989; Thompson, 1988; Chan y otros, 1999); también ha sido utilizado para medir la estabilidad del índice de sesgo (Harris y Kolen, 1989), o la distribución del coeficiente alpha (Yuan, Guarnaccia y Hayslip, 2003) en teoría de tests, o para evaluar diferentes índices de ajuste en el marco de los modelos de estructuras de covarianza (Bollen y Stine, 1993; Yung y Bentler, 1994). Además de sus aplicaciones inferenciales, el bootstrap se ha contemplado como una herramienta útil para describir la estabilidad y replicabilidad de los resultados estadísticos (Thompson, 1995).

El objetivo de este trabajo es presentar y evaluar el método bootstrap no paramétrico para obtener medidas de precisión, en concreto errores estándares, e intervalos de confianza para el coeficiente de determinación (R^2), índice comúnmente utilizado en el contexto de los modelos de regresión para evaluar la bondad de ajuste del modelo. Dada la relevancia de este índice en los estudios aplicados, es innegable el interés de contar con procedimientos que permitan generalizar los resultados a la población.

2.- Método

Consideremos el modelo de regresión lineal, $\mathbf{y} = \mathbf{XB} + \boldsymbol{\varepsilon}$ donde $\mathbf{y}$ es un vector de observaciones en una variable dependiente o criterio, $\mathbf{X}$ es una matriz de observaciones de variables independientes (VI), $\mathbf{B}$ es el vector de coeficientes de regresión y $\boldsymbol{\varepsilon}$ es un vector de términos de error, que se distribuyen normalmente, con esperanza de cero, varianza común (σ^2) y covarianza nula; $\varepsilon \approx N\left(0, \sigma^2 I_n\right)$. La estimación mínimo cuadrática del vector de coeficientes se obtiene mediante $\mathbf{b} = \left(\mathbf{X'X}\right)^{-1}\mathbf{X'y}$ y el vector de residuales se corresponderá con $\mathbf{e} = \mathbf{y} - \mathbf{Xb}$.

En este contexto se define el coeficiente de determinación R^2, como $R^2 = \dfrac{\mathbf{b'X'Xb}}{\mathbf{b'X'Xb} + \mathbf{e'e}}$, función de la matriz **X** de **VI**, y de los vectores **b** y **e**. Se interpreta como la proporción de varianza explicada por el modelo respecto a la varianza total. Como índice de bondad de ajuste exige un valor alto como criterio o argumento para mantener el modelo evaluado como plausible.

La aplicación del procedimiento bootstrap particularizado para el caso de la regresión (Wu, 1986; Stine, 1989) se esquematiza en el gráfico siguiente:

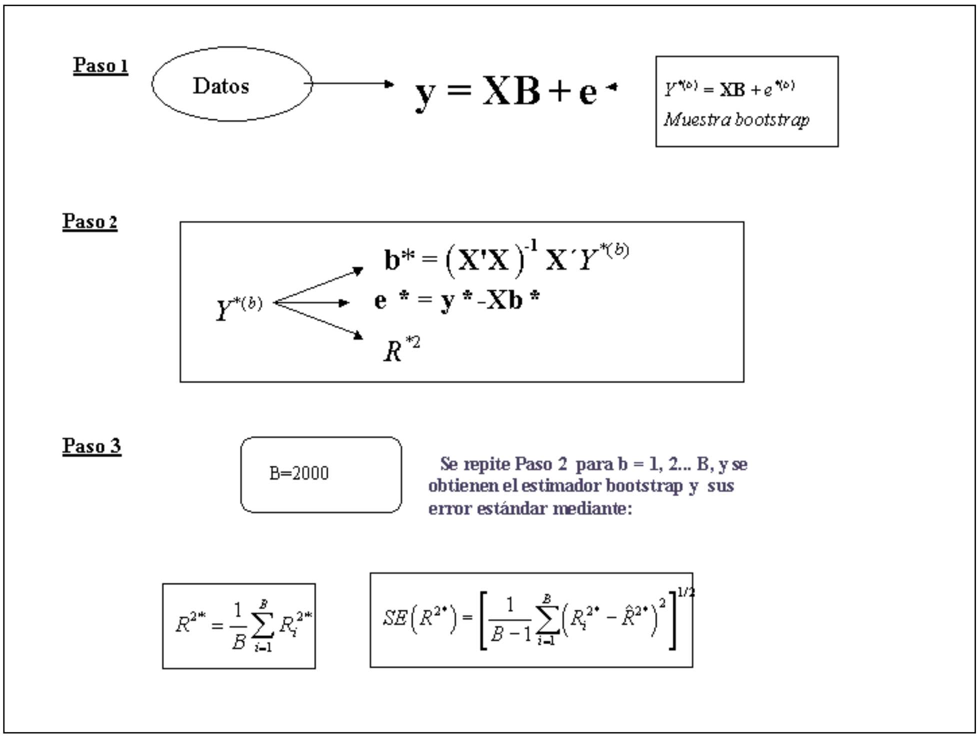

GRAFICO 1. Esquema del procedimiento para la obtención del estimador bootstrap y su error estándar

En el primer paso, se estima el ajuste mínimo cuadrático para la muestra original, y se crea la muestra bootstrap añadiendo al ajuste anterior los residuales remuestreados. En el segundo paso, se obtienen las estimaciones mínimo cuadráticas y sus correspondientes residuales a partir de la muestra bootstrap. Con estos valores se calculan los valores bootstrap de R^2. Se repiten los pasos 1 y 2 , en nuestro caso un número de veces igual a 2000.

Se utilizan dos procedimientos para la estimación de los intervalos de confianza; el método percentil (Efron, 1979) y el método corregido para el sesgo y acelerado (Bca) (Efron 1987). El primero asigna como extremos inferior y superior del intervalo de confianza (1-α), los percentiles $\alpha/2$ y 1-$\alpha/2$ de la distribución bootstrap del estimador. El segundo compensa las limitaciones del método percentil ante la ausencia de simetría en la distribución del estimador, y en aquellas situaciones en que la forma de la distribución cambia dependiendo de los valores del parámetro. Los límites del intervalo de confianza según el método BCa se obtienen al igual que en el método percentil a partir de los cuantiles de la distribución bootstrap, pero dependerán además de una constante de aceleración ***a*** y de la corrección para el sesgo[1].

Las propiedades asintóticas de los intervalos bootstrap han sido bien establecidas (Hall, 1988, 1992; DiCiccio y Efron, 1996); En este aspecto el método BCa es superior al método percentil (precisión de segundo orden) (Hall, 1988) cuando el tamaño de la muestra es elevado; Sin embargo en muestras pequeñas la evaluación de la superioridad relativa de unos métodos sobre otros es más compleja, dependiendo del estadístico evaluado y de las características de la población muestreada, de ahí el interés en comparar el comportamiento de los dos procedimientos en las diferentes condiciones que planteamos en nuestra simulación.

2.1 Simulación Montecarlo

Para estudiar el comportamiento de las estimaciones bootstrap, sus errores estándares y los dos diferentes intervalos de confianza, diseñamos un estudio Montecarlo en el que las muestras aleatorias se generan bajo condiciones poblacionales conocidas y controladas y a partir de distribuciones normales multivariadas.

Se manipularon tres factores con tres niveles cada uno. El tamaño muestral, ***n*** (50, 100 y 200), el valor del coeficiente de determinación poblacional, $\boldsymbol{\rho^2}$ (0.90, 0.60, 0.30) y el número de variables independientes (3,5,8). Tenemos así un total de 27 condiciones y un número de réplicas por condición igual a 1000. Los promedios de los resultados en esas mil iteraciones

1 Para una descripción detallada del intervalo Bca Se puede consultar Davison y Hinckley (1997)

(las estimaciones bootstrap, errores estándar e intervalos de confianza bootstrap) constituirán las estimaciones empíricas .[2]

Para realizar las simulaciones se creó un programa en lenguaje R (Ihaka y Gentleman, 1996), versión 1.6.2 utilizando la biblioteca boot que recoge los métodos presentados en el texto de Davison y Hinckley (1997) así como la función de generación de números aleatorios normales implementada en R.

3.- Resultados

Las tres primeras columnas de la tabla siguiente (Tabla 1) resumen las condiciones de la simulación; p (número de variables) n (tamaño de la muestra) y ρ^2 (coeficiente de determinación poblacional).

Dado que el coeficiente de determinación es un estadístico que sobreestima el verdadero valor del parámetro (Cramer, 1987) es una práctica habitual la obtención del "coeficiente de determinación corregido" R_c^2 (Wherry, 1931) ,

$$R_c^2 = 1 - \frac{n-1}{n-k}\left(1 - R^2\right)$$

En nuestro trabajo estudiamos ambos coeficientes. Se ha estudiado el comportamiento del coeficiente corregido, aplicando el procedimiento de remuestreo y la simulación en modo análogo al ya explicado para R^2. La tabla 1 muestra los resultados referidos a los dos coeficientes, R^2 y R_c^2

[2] Advirtamos en este punto del alto coste computacional de las simulaciones que involucran a su vez remuestreo. En nuestro caso cada una de las 27 condiciones implica un total de 1000 simulaciones y cada una de ellas implica 2000 remuestras, lo que supone un total de 54 millones de regresiones. A título orientativo señalemos que en un Pentium III a 800 Mhz, la media por condición fue de aproximadamente 6 horas.

Cond	p	n	ρ^2	R^{2*}	$E.S.\left(R^{2*}\right)$	$R_c^{\,2*}$	$E.S.\left(R_c^{\,2*}\right)$
1	**3**	**50**	.9	.9094	.0182	.9056	.019
2			.6	.625	.0702	.6091	.0732
3			.3	.3608	.0951	.3336	.0991
4		**100**	.9	.9045	.0135	.9025	.0138
5			.6	.6166	.0503	.6087	.0514
6			.3	.3353	.0689	.3216	.0703
7		**200**	.9	.9023	.0098	.9013	.0099
8			.6	.6063	.0362	.6023	.0366
9			.3	.3151	.0493	.3081	.0498
10	**5**	**50**	.9	.9138	.0179	.9061	.0195
11			.6	.6536	.0671	.6228	.073
12			.3	.4053	.0946	.3525	.103
13		**100**	.9	.9054	.0135	.9014	.0141
14			.6	.6271	.0495	.6114	.0516
15			.3	.36	.0686	.333	.0715
16		**200**	.9	.9033	.0097	.9013	.0099
17			.6	.6106	.036	.6027	.0367
18			.3	.3295	.0493	.3158	.0503
19	**8**	**50**	.9	.9203	.0171	.907	.02
20			.6	.6859	.0636	.6335	.0742
21			.3	.4708	.092	.3826	.1073
22		**100**	.9	.9094	.0132	.9025	.0142
23			.6	.6435	.0483	.6164	.0519
24			.3	.3851	.0682	.3384	.0734
25		**200**	.9	.9051	.0096	.9016	.0099
26			.6	.6202	.0355	.6064	.0368
27			.3	.3443	.0492	.3204	.051

Tabla 1. Medias y errores estándares bootstrap para R^2 y R^2_c

Como se aprecia en la tabla el estimador bootstrap R^{2*} reproduce el sesgo de R^2, de modo que los resultados mejores se obtienen para las condiciones 7, 8 y 9 en las que los valores de las estimaciones bootstrap están muy próximas al verdadero valor ρ^2 (0,9; 0,6 y 0,3) mientras que en las condiciones 19, 20 y 21 en las que el número de variables es el más numeroso (p=8) y el tamaño de la muestra es el menor (n=50) su diferencia con el parámetro es notable.

Utilizando el estimador bootstrap para el coeficiente de determinación corregido ($R_c^{\,2*}$) se atenúa el sesgo. Las estimaciones resultan muy aceptables especialmente para valores medios y altos del parámetro, y para ratios elevados de n/p.

En cuanto a los errores estándares, sus valores fluctúan en función de las condiciones: se elevan para ρ^2 bajo (.3) y n bajo (50), aunque incluso bajo estas condiciones el error estándar no llega a .1. En el resto de casos se mantienen por debajo de .07. Los valores de los errores estándares bootstrap para la versión corregida y la original del coeficiente son muy similares; aunque ligeramente inferior el correspondiente al coeficiente sin corregir.

Los resultados referidos a los intervalos de confianza se muestran en la tabla 2. Junto a las condiciones de las simulaciones se recogen los límites inferior y superior de los intervalos obtenidos por los dos procedimientos evaluados, método percentil y el intervalo corregido para el sesgo y acelerado (BCa), obtenidos a partir de las distribuciones bootstrap de las dos expresiones del coeficiente de determinación, la original y la corregida.

p	n	ρ^2	R cuad				R cuad corregido			
			PERCENTIL		BCA		PERCENTIL		BCA	
			li	ls	li	ls	li	ls	li	ls
3	50	,9	,8709	,9422	,8408	,9270	,8655	,9397	,8340	,9239
		,6	,4792	,7539	,3887	,7003	,4571	,7434	,3627	,6876
		,3	,1795	,5490	,1000	,4639	,1446	,5298	,0617	,4410
	100	,9	,8763	,9294	,8619	,9208	,8738	,9279	,8590	,9192
		,6	,5135	,7108	,4697	,6817	,5035	,7048	,4587	,6752
		,3	,2024	,4718	,1587	,4280	,1859	,4609	,1413	,4163
	200	,9	,8822	,9205	,8755	,9159	,8810	,9197	,8743	,9151
		,6	,5330	,6751	,5127	,6598	,5283	,6718	,5077	,6563
		,3	,2195	,4129	,1980	,3911	,2116	,4069	,1898	,3849
5	50	,9	,8758	,9458	,8419	,9246	,8648	,9410	,8279	,9179
		,6	,5137	,7759	,3974	,6982	,4705	,7560	,3439	,6714
		,3	,2211	,5897	,1096	,4562	,1518	,5533	,0305	,4079
	100	,9	,8771	,9302	,8560	,9183	,8720	,9273	,8500	,9149
		,6	,5255	,7196	,4612	,6775	,5055	,7078	,4385	,6639
		,3	,2267	,4951	,1544	,4259	,1941	,4739	,1188	,4017
	200	,9	,8833	,9215	,8741	,9152	,8809	,9199	,8716	,9134
		,6	,5377	,6790	,5078	,6566	,5282	,6724	,4977	,6495
		,3	,2337	,4270	,1975	,3915	,2180	,4153	,1811	,3790
8	50	,9	,8838	,9506	,8510	,9200	,8644	,9424	,8262	,9066
		,6	,5523	,8010	,4390	,6868	,4776	,7679	,3455	,6345
		,3	,2877	,6464	,1665	,4479	,1690	,5875	,0275	,3559
	100	,9	,8819	,9335	,8582	,9168	,8729	,9285	,8474	,9104
		,6	,5444	,7334	,4631	,6715	,5097	,7132	,4222	,6465
		,3	,2519	,5187	,1612	,4108	,1950	,4821	,0974	,3659
	200	,9	,8854	,9230	,8718	,9142	,8812	,9202	,8672	,9110
		,6	,5482	,6876	,5022	,6546	,5317	,6762	,4840	,6420
		,3	,2485	,4412	,1899	,3860	,2211	,4208	,1604	,3636

Tabla 2. Límites inferior y superior de los intervalos de confianza bootstrap percentil y BCa para R^2 y R^2_c

La evaluación del comportamiento de estos intervalos, se basa en los criterios de recubrimiento y equilibrio (Efrony Tsibirani,1993). El recubrimiento (*c)* se define como la proporción de veces que el intervalo incluye al parámetro dividido entre el número de réplicas o iteraciones (*I*).

$$c = \#\left(l_i < \Phi < l_s\right)/I$$

El equilibrio, (*e*), se refiere al grado en que los errores en la captura del parámetro se producen en proporciones similares en las dos colas de la distribución del estimador. Beran (1990) expresa el desequilibrio mediante la diferencia absoluta entre dos proporciones de recubrimiento. El problema de esta formulación es que no permite la comparación entre valores de diferentes tests cuando éstos tienen un valor de recubrimiento diferente; Para obviarlo, modificaremos ligeramente la métrica con el fin de hacer comparables los diferentes intervalos, de modo que la diferencia se divide entre la suma o proporción total de fallos o falta de recubrimiento;

$$e = \frac{P_i - P_s}{P_i + P_s} \times 100 \quad \forall \{P_i + P_s > 0\}$$

Siendo $P_i = \#(\Phi < l_i)/I$ y $P_s = \#(\Phi > l_s)/I$

El cero como resultado de la fórmula indica el equilibrio, o igual número de fallos en cada cola. El resultado vendrá afectado por el signo negativo en caso de que el mayor número de casos se den en la cola derecha, mientras que el signo positivo indicará un mayor número de rechazos por la izquierda.

Idealmente el recubrimiento debería ser exactamente el recubrimiento nominal del intervalo deseado (.95) y el desequilibrio igual a α.

Tal y como se aprecia en la tabla 3 el intervalo BCa utilizando el estimador bootstrap de R^2 arroja un nivel de recubrimiento muy bueno con valores cercanos al recubrimiento nominal de .95. Este el método que mejores resultados proporciona para la mayoría de las condiciones, (valores resaltados en negrita).

Cond	p	n	ρ^2	R cuad PERCENT	R cuad BCA	R cuad corregido PERCENT	R cuad corregido BCA
1	3	50	.9	.864	**.941**	.9	.92
2			.6	.908	**.929**	.926	.908
3			.3	.888	**.94**	.932	.917
4		100	.9	.901	**.937**	.923	.924
5			.6	.913	**.953**	.939	.94
6			.3	.928	**.95**	.947	.937
7		200	.9	.916	**.934**	.92	.929
8			.6	.938	**.95**	.942	.948
9			.3	.946	**.95**	.948	.955
10	5	50	.9	.825	.906	.911	.842
11			.6	.84	**.919**	.908	.857
12			.3	.82	.916	.933	.858
13		100	.9	.91	.924	.931	.881
14			.6	.888	**.942**	.921	.909
15			.3	.868	**.949**	.928	.908
16		200	.9	.92	.933	.934	.916
17			.6	.922	.938	.944	.916
18			.3	.903	**.945**	.93	.936
19	8	50	.9	.702	.791	.885	.626
20			.6	.701	.868	.901	.701
21			.3	.577	.858	.88	.686
22		100	.9	.845	.902	.931	.785
23			.6	.81	.918	.93	.822
24			.3	.764	.917	.923	.799
25		200	.9	.886	**.933**	.929	.877
26			.6	.905	**.948**	.95	.899
27			.3	.856	.939	.93	.892

Tabla 3. Valores de recubrimiento de los intervalos de confianza bootstrap percentil y BCa para R^2 y R^2_c

Estos aspectos se muestran con mayor nitidez en el gráfico 2. Dado que el método BCa ha sido ideado específicamente para los casos en que el estimador se vea afectado por el sesgo, no es de extrañar que su rendimiento sea muy superior frente al método percentil cuando ambos se obtienen para R^2 . Por otro lado se observa que si el intervalo bootstrap se construye para R^2 corregido, en este caso el método percentil da excelentes resultados; mejor incluso que el BCa. De hecho para aquellas condiciones más desfavorables (19,20,21) su nivel de recubrimiento es mejor que éste para R^2. Ambos métodos se mantienen en niveles de recubrimiento entre .87 y .95.

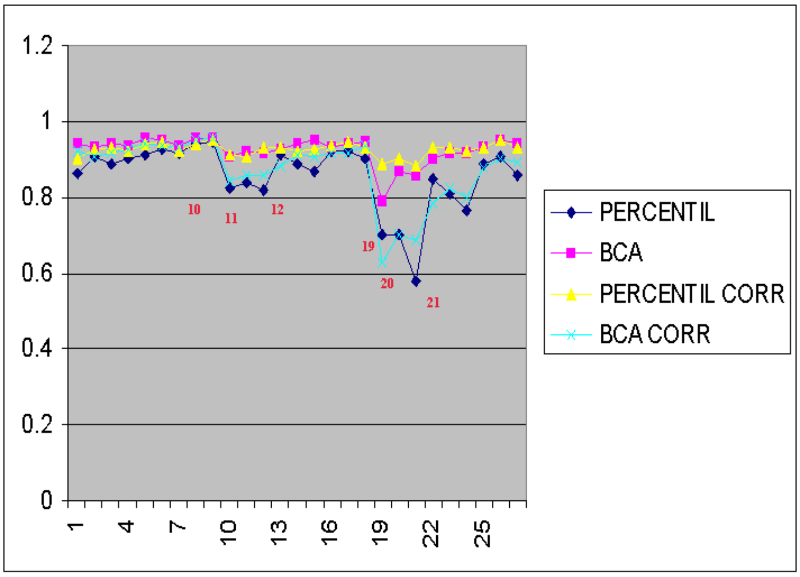

Gráfico 2. Recubrimiento de los intervalos de confianza bootstrap percentil y BCa para R^2 y R^2_c

Otro aspecto a destacar es el empeoramiento de los niveles de recubrimiento para las condiciones 19, 20, 21 (p = 8, n.= 50), produciéndose el pico más bajo para ρ^2 = .30 (condición 21), y también, aunque en menor cuantía, para las condiciones 10, 11 y 12 (p = 5, n = 50) y para la 24, lo que corrobora el hecho de que ratios sujetos/variables bajos, y valores bajos de ρ^2 afectan negativamente a la bondad de la estimación.

En cuanto al equilibrio, tal y como se aprecia en el gráfico 3 el intervalo percentil tanto para R^2 como para R^2 corregido manifiesta un desequilibrio con signo positivo, es decir, la proporción de falsos rechazos por la derecha es mayor; el límite superior del intervalo está desplazado hacia la izquierda.

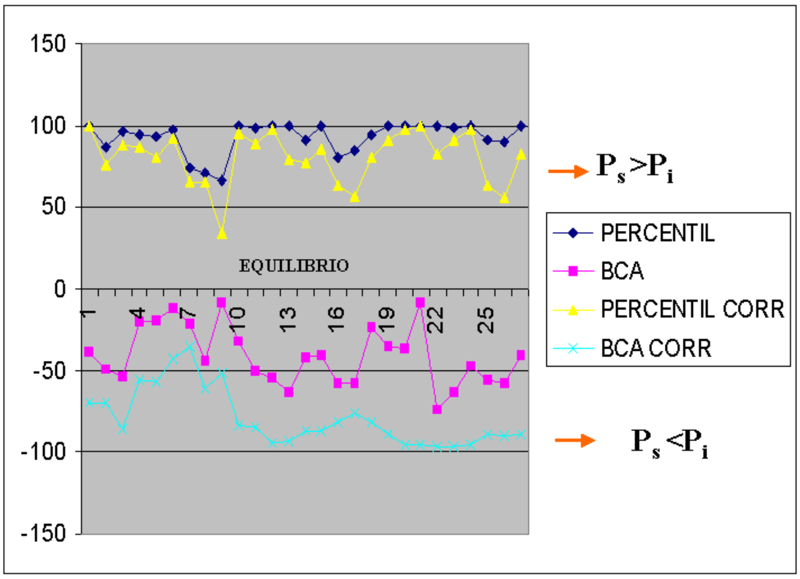

Gráfico 3: Equilibrio de los intervalos de confianza bootstrap percentil y BCa para R^2 y R^2_c

Los intervalos BCa por el contrario muestran un desequilibrio negativo, la proporción de falsos rechazos es mayor en la cola izquierda. El intervalo BCa para R^2 muestra un nivel de equilibrio mayor en todas las condiciones comparándolo con el intervalo percentil, oscilando en la mayoría de ellas entre los valores de 0 y –5.

4.- Discusión

A partir de los resultados de la simulación podemos concluir que los métodos boostrap constituyen un adecuado enfoque para obtener estimaciones del error estándar e intervalos de confianza para el coeficiente de determinación. Las estimaciones son razonablemente precisas especialmente en aquellos casos en los que el ratio entre el número de sujetos y el de variables es adecuado, y los valores del coeficiente son medios o elevados. Por otro lado, la eficiencia de los intervalos de confianza es también sensible a las condiciones mencionadas.

Se han comparado asimismo los resultados referidos al coeficiente de determinación R^2, y al coeficiente de determinación corregido R_c^2, y dos procedimientos bootstrap de construcción de intervalos de confianza. Las magnitudes de los errores estándares son similares para R^2 y para R_c^2. En cuanto a los intervalos de confianza, a tenor de los resultados se puede concluir la conveniencia de utilizar el método percentil en el caso de que el intervalo de confianza se construya a partir de la distribución bootstrap de R_c^2 o alternativamente el método BCa si el intervalo de confianza se obtenga a partir del coeficiente R^2 sin corregir. Los valores de recubrimiento en ambos casos son similares, aunque el intervalo obtenido por el segundo procedimiento se muestra más equilibrado.

Estos resultados apoyan el uso de los procedimientos bootstrap en el ámbito de la regresión, ahora bien, nuestro análisis se ha limitado a la comprobación de la eficacia de la técnica en el supuesto de normalidad multivariable; Sería de interés ampliar el estudio introduciendo condiciones más realistas de no normalidad y analizando su impacto.

Para finalizar subrayemos que este trabajo se inscribe en la línea de los recientes esfuerzos por proporcionar métodos que permitan ir más allá del mero contraste de la hipótesis nula, enfatizando la obtención y el uso de los intervalos de confianza (Wilkinson y APA Task Force, 1999). Estos últimos son más informativos y posibilitan en mayor medida la acumulación del conocimiento (Thompson, 2002).

5.- Referencias

Beran, R. (1990). Refining bootstrap simultaneous confidence sets. *Journal of the American Statistical Association, 85*(410), 417-426.

Bollen, K. A., y Stine, R. (1993). Bootstrapping goodness-of-fit measures in structural equation models. In K. A. Bollen y J. S. Long (Eds.), *Testing structural equation models*. Newbury Park, CA: Sage.

Chan, W., Ho, R., Leung, K., Chan, D., y Yung, Y. (1999). An alternative method for evaluating congruence coefficients with procrustes rotation: A bootstrap procedure. *Psychological Methods, 4*, 378-402.

Chatterjee, S. (1984). Variance estimation in factor analysis: An application of the bootstrap. *British Journal of Mathematical and Statistical Psychology, 37*, 252-262.

Cramer, J. S. (1987). Mean and variance of R2 in small and moderate samples. *Journal of Econometrics, 35*, 253-266.

Dalgleish, L. I. (1994). Discriminant analysis: Statistical inference using the jacknife and bootstrap procedures. *Psycological Bulletin, 116*, 498-508.

Davison, A.C. and Hinckley, D.V. (1997). *Bootstrap methods and their application*. New York: Cambridge University Press

Diaconis, P., y Efron, B. (1983). Computer intensive methods in statistics. *Scientific American, 248*(5), 116-13.

DiCiccio, T. J., y Efron, B. (1996). Bootstrap confidence intervals (with discussion). *Statistical Science, 11*, 189-228.

Efron, B. (1979). Bootstrap methods. *Annals of statistics, 7*, 1-26.

Efron, B. (1981). Nonparametric estimates of standard error: the jacknife, the bootstrap, and other resampling methods. *Biometrika, 68*, 589-599.

Efron, B. (1987). Better bootstrap confidence intervals (with discussion). *Journal of the American Statistical Association, 82*, 171-20.

Efron, B., y Tsibirani, R. J. (1993). *An introduction to the Bootstrap*. N.Y.: Chapman & Hall.

Hall, P. (1988). Theoretical comparison of bootstrap confidence intervals. *Ann. Statist. 16*, 927-985.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. New York: Springer-Verlag.

Hall, P., y Martin, M. (1988). On bootstrap re-sampling and iteration. *Biometrika, 75*(4), 661-671.

Harris, D. J., y Kolen, M. J. (1989). Examining the stability of Angoff's Delta item bias statistic using the bootstrap. *Educational and Psychological Measurement, 49*, 81-87.

Ihaka, R., y Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics, 5*, 299-314.

Lunneborg, C. E. (1987). Bootstrap applications for the behavioral sciences. *Educational and Psychological Measurement, 47*, 627-629.

Micceri, T. (1989). The unicorn, the normal curve and other improbable creatures. *Psycological Bulletin, 105*(1), 156-166.

Mosier, C. (1951). Problems and Designs of Cross-Validation. *Educational and Psychological Measurement, 11*.

Noreen, E. (1989). *Computer intensive methods for testing hypotheses*. New York: John Wiley & Sons, Ltd.

Quenouille, M. H. (1956). Notes on Bias in Estimation. *Biometrika, 43*, 353-36.

Ramírez Santana, G. M. (1996). *Estudio de la robustez del Análisis Discriminante Multivariado: propuesta de un ajuste bootstrap para la significación estadística de los coeficientes de la técnica.* Universidad de la Laguna.

Scott, R. L., Thompson, B., y Sexton, D. (1989). Structure of a short form of the questionnaire on resources and stress: A bootstrap factor analysis. *Educational and Psychological Measurement, 49*, 409-419.

Simon, J. L. (Ed.). (1969). *Basic Research Methods in Social Science*. (Vol. acceso en Marzo 2003). New York: Random House.

Stine, R. (1989). An Introduction to Bootstrap Methods. *Sociological Methods and Research, 18*(2 y 3), 243-291.

Thompson, B. (1988). Program factrap: A program that computes bootstrap estimates of factor structure. *Educational and Psychological Measurement, 48*, 681-686.

Thompson, B. (1995). Exploring the replicability of a study's results: Bootstrap statistics for the multivariate case. *Educational and Psychological Measurement, 55*, 84-94.

Thomson, B. (2002). What Future Quantitative Social Science Research Could Look Like: Confidence Intervals for Effect Sizes. *Educational Researcher, 31*(3), 25-33.

Tukey. (1958). Bias and Confidence in Not-Quite Large Samples (Abstract). *Annals of Mathematical Statistics, 29*, 614.

Wherry, R. J., Sr. (1931). A new formula for predicting the shrinkage of the coefficient of multiple correlation. *Annals of Mathematical Statistics, 2*, 440-457.

Wilkinson, L., y Force., A. T. (1999). Statistical Methods in Psychology Journals: Guidelines and Explanations. *American psychologist, 54*(8), 594-604.

Wu, C. F. J. (1986). Jackknife, bootstrap and the resampling methods in regression analysis. *Annals of Statistics, 14*, 1261-135.

Yuan, K., Guarnaccia, C., y Hayslip, B. (2003). A study of the distribution of sample coefficient alpha with the Hopkins symptom checklist: Bootstrap versus asymptotics. *Educational and Psychological Measurement, 63*(1), 5-23.

Yung, Y., y Bentler, P. M. (1994). Bootstrap-corrected ADF statistics in covariance structure-analysis. *British Journal of Mathematical and Statistical Psychology, 47*, 63-84.