QÜESTIIÓ, Vol 20, 3 pp. 549-562, 1996

SENSIBILIDAD DEL ANÁLISIS DE LA «ESTRUCTURA PROPIA» EN LA DETECCIÓN DE COLINEALIDAD

ESTARELLES, R.*

PRIETO, J.*

Universidad de Valencia

Muchos son los criterios, tanto de carácter formal como informal, que han ido proponiéndose para la detección de la colinealidad. Sin embargo, muchos de ellos no ofrecen un diagnóstico eficaz que permita evaluar el nivel de alteración que se produce en las estimaciones obtenidas, variables realmente implicadas... En el presente trabajo se presentan las ventajas de utilizar índices basados en la «estructura propia» de la matriz de predictores frente a otras estrategias formales de uso más frecuente. Este objetivo se pone de manifiesto mediante el análisis de datos reales.

Sensibility analysis of the proper structure in detecting collinearity

Keywords: Diagnóstico de Colinealidad, Factor de Inflación de la Varianza, Descomposición en Valores Singulares, Número de Condición, Descomposición de la Varianza de los Coeficientes en Proporciones.

*Estarelles, R. y Prieto, J. Metodología de las Ciencias del Comportamiento. Facultad de Psicología. Universidad de Valencia.

–Article rebut el gener de 1996.

–Acceptat el desembre de 1996.

1. INTRODUCCIÓN

La presencia de un elevado nivel de colinealidad es un hecho frecuente en gran número de investigaciones, y sus repercusiones afectan tanto a nivel estadístico como de cálculo en los resultados. Entre las repercusiones estadísticas más importantes se encuentran la imposibilidad para obtener estimaciones únicas para los coeficientes de regresión, incrementos en la variabilidad de los coeficientes estimados, errores de tipo II más elevados, sensibilidad ante pequeñas variaciones en la variable respuesta, dificultades en la interpretación de los resultados, entre otras. Dadas las implicaciones que suelen presentarse, el investigador deberá tener un claro conocimiento del nivel de incidencia y alteraciones que este fenómeno pudiera estar produciendo, número de relaciones colineales y su intensidad, su existencia de forma simultánea o independiente, su implicación en relaciones equivalentes o dominantes, así como si las alteraciones son lo suficientemente nocivas para los objetivos propuestos como para plantearse la utilización de estrategias alternativas. Por todo ello, el estudio de esta deficiencia y sus implicaciones ha sido el aspecto central de gran número de manuales y artículos sobre regresión desde hace años (Gunst y Mason, 1977; Belsley, Kuh y Welsch, 1980; Chatterjee y Price, 1977; Estarelles, 1989; Fox, 1991; Belsley, 1991, etc), habiéndose propuesto un amplio número de estrategias para su detección y tratamiento.

Desde un enfoque aplicado han ido ofreciéndose acercamientos basados fundamentalmente en la matriz de correlaciones, tal como el propio análisis de los elementos dicha matriz, su determinante, evaluación de correlaciones múltiples y parciales, o bien la consideración del *Factor de Inflación de la Varianza «VIF»* (Marquardt, 1970), así como otros índices derivados de este factor, como son el «Índice de Tolerancia» o el «Valor Promedio para VIF» (Berk, 1977). Sin embargo, desde la perspectiva del Análisis Numérico se han ido presentando estrategias, que aunque no pueden considerarse como independientes de las anteriores, han demostrado ser bastante más precisas (Berk, 1977; Mandel, 1982; Stewart, 1987, entre otros). Estos acercamientos, surgieron como un medio para disminuir los errores de cálculo derivados de la presencia de matrices próximas a la singularidad, y se han basado en el análisis de la «estructura propia» de la matriz de datos (Golub y Reinsch, 1970; Stewart, 1973; Lawson y Hanson, 1974; Mandel, 1982; etc). Esta línea de trabajo ha permitido profundizar en el conocimiento de las propiedades de la «estructura propia» de la matriz de predictores, ofreciendo una magnífica herramienta para la identificación y estudio de la dependencia entre variables, como es la descomposición de la matriz X en valores singulares, tal que $X = UDV'$, siendo U los «vectores propios» o «eigenvectores» de la matriz XX', V los «vectores propios» de la matriz $X'X$ y D los valores singulares, o lo que es lo mismo, las raíces cuadradas de los «eigenvalues» o «raíces características» para la matriz $X'X$.

1.2. «Estructura propia» de la matriz de predictores

El análisis de la «estructura propia» de la matriz X, a pesar de no ser un enfoque reciente para el diagnóstico de la colinealidad (Kendall, 1957; Silvey, 1969; Chatterjee y Price, 1977, etc), dista mucho de ser uno de los procedimientos más utilizados, y algunas de las estrategias de él derivadas, como es la «Descomposición de la Varianza de los Coeficientes en Proporciones», a pesar de haber demostrado su gran utilidad, han sido prácticamente ignoradas desde una perspectiva aplicada (Estarelles, 1989), siendo lamentable la falta de interés mostrada hacia enfoques que resultaron innovadores al inicio de la década de los 80 (Belsley, Kuh y Welsch, 1980), como es el análisis conjunto de dicha descomposición respecto al desglose de la matriz de predictores en valores singulares.

Posiblemente el procedimiento más utilizado desde esta perspectiva ha venido siendo la consideración de las «raíces características», y de los «vectores propios» para la matriz de correlaciones. Este criterio, a pesar de tener la ventaja de permitir analizar el número de relaciones colineales en un conjunto de datos (un valor de cero para los valores singulares estaría determinando la nulidad de X), no cuenta, sin embargo, con un punto de corte formal, a partir del cual pueda inferirse que se producirán graves alteraciones en los resultados. Asimismo, valores elevados para el vector propio suelen estar asociados con problemas de colinealidad, en cambio, si estos valores fuesen reducidos no podría desprenderse la conclusión opuesta (Belsley, 1984). Igualmente, una matriz bien condicionada podría incorporar raíces características con valores reducidos, por tanto, más que el análisis individual de raíces y vectores característicos, se precisará de otros acercamientos que permitan comparar la dispersión entre las «raíces características», intensidad con que se presenta el condicionamiento de los datos, nivel de perturbación en los resultados, así como análisis de las variables implicadas en cada relación colineal y tipo de dependencia establecido, independientemente de que, de forma complementaria, puedan incorporarse otras estrategias, tal como considerar el inverso de las «raíces características» (Johnston, 1984; Fox, 1991), la obtención de «Componentes Principales» (Kloeck y Mennes, 1960; Jollife, 1982), o el análisis de la estructura, incluyendo la variable respuesta, tal como si de un modelo de «Raíces Latentes» se tratara (Estarelles, Prieto, Castro y Aragón, 1995), entre otras.

Teniendo en cuenta que el hecho de que un valor singular sea reducido puede no ser indicativo de la presencia de colinealidad, el condicionamiento deficiente de una matriz podrá manifestarse a partir del producto de su máximo valor singular, es decir, de la norma espectral para aquella matriz, por el máximo valor singular de su inversa (Belsley, 1991). Este resultado conocido como «Índice de Condicionamiento» o «Número de Condición», fue introducido por Turing en 1948 para delimitar perturbaciones en la relación de sistemas lineales, y aplicado posteriormente para soluciones mínimo cuadráticas por Golub y Wilkinson (1966), con el fin de detectar la sensibilidad de la ecuación ante modificaciones en los elementos del sistema. Teniendo en cuenta que puede demostrarse que $\|X\| = \mu_{\text{máx}}$, así como que si X es una matriz cuadrada $\|X^{-1}\| = 1/\mu_{\text{mín}}$ (Wilkinson, 1965), la relación entre ambos términos mostrará el condicionamiento deficiente de una matriz (para una exposición detallada ver Belsley, 1984):

(1) $$IC = \frac{\mu_{\text{máx}}}{\mu_{\text{mín}}} = \sqrt{\frac{\lambda_{\text{máx}}}{\lambda_{\text{mín}}}} \geq 1$$

siendo $\lambda_{\text{máx}}$ y $\lambda_{\text{mín}}$ las raíces características para la matriz de predictores de tamaño superior e inferior, respectivamente.

Para conocer el número de dependencias, la lógica presentada podrá extenderse para el resto de valores singulares, pudiendo obtenerse tantos índices como predictores incorporados. Estos cocientes se conocen como «Índices de Colinealidad». No obstante, aún se desconocerá el nivel de distorsión que pueden presentar los resultados. Una estrategia que permite acometer con cierta garantía de éxito este objetivo es la «Descomposición de la Varianza de los Coeficientes de Regresión en Proporciones», introducida por Silvey en 1969.

Teniendo en cuenta que la matriz de Varianza-Covarianza para los coeficientes de regresión mínimo-cuadráticos se obtiene de: $V_{(b)} = \sigma^2(X'X)^{-1}$, esta matriz podrá expresarse equivalentemente a partir de la descomposición de X en valores singulares, tal que $V_{(b)} = \sigma^2 V D^{-2} V'$, siendo la varianza para cada coeficiente:

(2) $$\text{var}_{(b_i)} = \sigma^2 \sum_j \frac{v_{ij}^2}{u_j^2}$$

siendo $V \equiv v_{ij}$ y u_j los valores singulares obtenidos. Por tanto, cuanto más reducidos sean estos valores, previsiblemente mayor será la variabilidad del coeficiente, aunque de la expresión anterior igualmente se deduce que una varianza del error reducida disminuirá los efectos de la colinealidad en la variabilidad de los coeficientes. No obstante, otras situaciones especiales también influirán en el mismo sentido, tal sería el caso de la presencia de colinealidad en dos conjuntos de variables que resultasen ser ortogonales entre sí (Belsley, 1991). La «Descomposición de la Varianza de los Coeficientes en Proporciones» se obtendrá a partir de:

(3) $$\pi_{ji} = \frac{\omega_{ij}}{\omega_i}, \quad \text{siendo } \omega_i = \Sigma\, \omega_{ij} \quad \text{y} \quad \omega_{ij} = \frac{v_{ij}^2}{u_j^2}$$

La representación mediante una tabla conjunta de «Índices de Colinealidad» y «Descomposición de la Varianza» ofrece un medio para evaluar la degradación que sufren los coeficientes estimados, informando del número y tipo de dependencias mostrado por las variables. Cuando esta identificación se torna compleja, como en el caso de que se presentasen relaciones equivalentes, el papel desempeñado por cada uno de los predictores podrá clarificarse mediante la utilización de «Regresiones Auxiliares» (Belsley, 1991; Estarelles, Castro y Prieto, 1995), aunque si bien estas regresiones solo deberán plantearse con fines descriptivos, no inferenciales, teniendo sentido su aplicación cuando tras el proceso de diagnóstico pudiera cuestionarse el papel desempeñado por las variables implicadas.

En el presente trabajo se considera el comportamiento de algunas de las estrategias más usuales en la detección de la colinealidad, destacándose las ventajas de aplicar estrategias basadas en la «estructura propia» respecto a otros acercamientos de uso más frecuente en investigación aplicada. Asimismo, se ha considerado de interés resaltar el comportamiento de los índices estudiados a partir de una base de datos reales, ya que éste es el punto de referencia más habitual en investigación aplicada, en donde difícilmente se encuentran aquellas situaciones teóricas ideales que pueden plantearse a partir de una simulación. Las estrategias seleccionadas en el presente estudio no representan todo el abanico de posibilidades que incorpora la literatura, ya que la inclusión de muchas de ellas ofrecería información redundante, no encontrándose entre los objetivos de este estudio el ofrecer una revisión exhaustiva de los numerosos índices que han ido desarrollándose. Para una exposición detallada ver Estarelles y Prieto (1995).

2. APLICACIÓN A PARTIR DE DATOS REALES

A partir de una muestra de 83 jóvenes, de edades comprendidas entre 20 y 23 años, compuesta por aproximadamente igual número de sujetos de ambos sexos, se pretendió explicar el nivel de *Auto-Satisfacción con el Comportamiento Personal*, a través de variables como: *Percepción de Afecto Familiar*, *Auto-Concepto Ético-Moral*, *Auto-Concepto Físico*, *Pensamientos Depresivos*, *Ausencia de Sentimientos de Culpabilidad* y *Problemas Sociales*.

Tras la realización de los análisis preliminares pertinentes, no se detectó la presencia de puntuaciones excesivamente extremas, en ninguna de las variables (ni aparentemente de forma conjunta), ni ninguna otra violación a los supuestos esenciales del modelo, pudiendo asumirse una relación lineal entre la variable criterio y los predictores. Sin embargo, podrían considerarse objeto de un análisis más detallado las correlaciones relativamente elevadas que se observan entre los predictores (Tabla 1), haciendo sospechar la posible presencia de relaciones colineales que pudieran estar afectando los resultados. No obstante, este aspecto, posiblemente por no tener connotaciones alarmantes, podría haber pasado erróneamente desapercibido en una investigación aplicada que no tuviera unos intereses esencialmente metodológicos.

Tabla 1

Correlaciones entre las variables

VARIABLE	A.Sat.(Y)	A.Fam.(X_1)	A.Et.(X_2)	A.Fis.(X_3)	P.Dep.(X_4)	Pr.Soc.(X_5)	A.S.C(X_6)
A.Sat.(Y)	1.000	–	–	–	–	–	–
A.Fam.(X_1)	0.620	1.000	–	–	–	–	–
A.Et.(X_2)	0.650	0.580	1.000	–	–	–	–
A.Fis.(X_3)	0.703	0.699	0.633	1.000	–	–	–
P.Dep.(X_4)	−0.24	−0.12	−0.34	−0.02	1.000	–	–
P.Soc.(X_5)	0.804	0.752	0.726	−0.24	0.571	1.000	–
A.S.C.(X_6)	0.777	0.681	0.645	−0.26	0.500	0.707	1.000

Tras el estudio de cada uno de los predictores independientemente, así como tras el análisis de todos los subconjuntos posibles, se procedió a comparar el comportamiento de los coeficientes estimados y su variabilidad, observándose como los valores para los coeficientes iban modificándose a medida que se alteraba el conjunto de predictores, comprobándose igualmente la falta de significación de variables relevantes, así como cambios de signos no justificados en el valor de algunos coeficientes. Como puede observarse, los signos para los coeficientes de las variables X_1 y X_2 resultan modificados, y los valores para el resto de coeficientes se ven distorsionados. Sin embargo, tal como puede comprobarse en la tabla 2, no se observa un incremento notable en la variabilidad de los coeficientes. Tal como Myers (1990) demuestra, van a ser todos los coeficientes los que de alguna manera van a verse afectados por la colinealidad de los datos, lo cual no implica que esta deficiencia no esté alterando la estimación e interpretación del modelo, especialmente en casos en los que la colinealidad se da de forma simultánea (Estarelles, 1989), tal como veremos ocurre en el caso estudiado.

En este sentido, y como resumen de los sucesivos análisis realizados, a continuación se exponen los resultados para cada predictor respecto al criterio, y tras el análisis global del conjunto de predictores seleccionados.

Para la detección de la presencia de posibles relaciones colineales, es usual el considerar el «Factor de Inflación de la Varianza» *(VIF)* para cada uno de los predictores:

$$VIF_{(x_j)} = \frac{1}{1 - R_j^2} \tag{4}$$

siendo R_j^2 el *Coeficiente de Determinación* resultante de regresar la variable X_j respecto al resto de predictores incluidos en el modelo. También es común la utilización de otros índices derivados como es el nivel de «Tolerancia» $(1 - R_j^2)$. Estos valores suelen incluirse de forma automática por los principales paquetes de Software. En el presente estudio se ha incluido igualmente el valor promedio para el *Factor de Inflación*. La realización de los análisis sucesivos se ha efectuado a partir de una matriz de predictores homogeneizada, tal como se estima pertinente para este tipo de situaciones (Stewart, 1987; Myers, 1990; Belsley, 1991, etc).

Tabla 2

Análisis de regresión para cada uno de los predictores y resultados globales

ANÁLISIS INDIVIDUAL				
Variables	Coef.	Err.Est.	t	p
Af.Fam.(X_1)	0.773	0.109	7.111	0.0000
A.Et.M.(X_2)	0.737	0.096	7.698	0.0000
A.Fís.(X_3)	1.260	0.142	8.898	0.0000
P.Depr.(X_4)	−0.622	0.273	−2.275	0.0255
Pr.Soc.(X_5)	1.347	0.111	12.167	0.0000
A.St.C.(X_6)	1.318	0.119	11.095	0.0000
ANÁLISIS CONJUNTO ($R^2 = 0.79$)				
Variables	Coef.	Err.Est.	t	p
Af.Fam.(X_1)	−0.375	0.111	−3.376	0.0012
A.Et.M.(X_2)	−0.153	0.096	−1.600	0.1137
A.Fís.(X_3)	0.737	0.144	5.136	0.0000
P.Depr.(X_4)	−0.197	0.143	−1.374	0.1736
Pr.Soc.(X_5)	0.856	0.158	5.418	0.0000
A.St.C.(X_6)	0.789	0.133	5.911	0.0000

Tabla 3

Inflación de la varianza «VIF»

Af.Fam.(X_1)	A.Et.(X_2)	A.Fís.(X_3)	P.Depr.(X_4)	Pr.Soc.(X_5)	A.St.C.(X_6)
3.10397	2.78129	2.51663	1.24895	3.48798	2.4248

Si se asume el punto de corte habitualmente establecido en la literatura de 10, aparentemente ninguna de las relaciones en el conjunto de datos parece mantener una relación excesivamente conflictiva para la estimación del modelo. Este resultado guarda relación con la falta de aparentes alteraciones en los errores estándar que se observaron (tabla 2), y que muestra la falta de fiabilidad de «*VIF*» para determinados conjuntos de datos, especialmente cuando las dependencias se dan de forma simultánea. Por otro lado, el valor promedio para este índice es igual a 2.59. La superioridad de este resultado respecto a la unidad no es excesivamente elevada, sin embargo, las alteraciones anteriormente observadas aconsejan un diagnóstico más riguroso, tal como el que se desprende del análisis de la «estructura propia». En la tabla siguiente se incluyen las raíces características para la matriz de predictores.

Tabla 4

Raíces características de la matriz de predictores

λ_0	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6
0.0047	0.0044	0.0063	0.0075	0.0021	0.0444	6.9348

Tal como puede observarse, existen cinco raíces con valores muy próximos a cero, lo cual permite identificar el condicionamiento deficiente de los datos, especialmente si tenemos en cuenta que, aún más importante que la proximidad de este valor a cero, es la gran disparidad existente entre el valor máximo respecto al resto de valores para las raíces características inferiores. Los «Índices de Colinealidad» para cada uno de los predictores se incluyen en la tabla siguiente. Se ha estimado conveniente la inclusión de la constante, con el fin de comprobar si existe algún tipo de relación colineal con este término.

Tabla 5

Índices de colinealidad

IC_0	IC_1	IC_2	IC_3	IC_4	IC_5	IC_6
38.56	39.78	33.33	30.44	56.88	12.50	1.00

En la literatura se han sugerido puntos de corte aproximados para este índice (Belsley, Kuh y Welsch, 1980; Fox, 1991, etc), habiéndose asumido que valores superiores a 30 pondrán de manifiesto la presencia de distorsiones importantes. El número de «Índices de Colinealidad» que resulten superiores al valor de corte asumido indicará el número de dependencias existentes. Como puede observarse, la presencia de cinco «Índices de Colinealidad» por encima de 30 ratifican el problema de dependencia ya apuntado.

Como puede comprobarse, el análisis de la «estructura propia» permite analizar con mayor precisión que el «Índice de Inflación de la Varianza» la presencia de relaciones colineales, así como la intensidad de esta deficiencia. No obstante, un aspecto de interés, aún no claramente determinado, será la identificación y vinculación entre aquellas variables principalmente implicadas. Para ello se procederá a la «Descomposición de la Varianza» de cada uno de los coeficientes en proporciones, analizándose la incidencia de estos valores respecto a los «Índices de Colinealidad» previamente obtenidos.

Tabla 6

Diagnóstico de la colinealidad mediante la descomposición de la varianza

IC_j	Descomposición de la Varianza en Proporciones						
	b_0	b_1	b_2	b_3	b_4	b_5	b_6
1	0.001	0.003	0.009	0.012	0.003	0.000	0.007
12.50	0.072	0.049	0.045	0.026	0.028	0.051	0.092
30.44	0.206	0.278	0.312	0.123	0.017	0.093	0.108
33.33	0.280	0.214	0.170	0.274	0.039	0.201	0.093
38.56	0.176	0.149	0.163	0.192	0.056	0.082	0.068
39.78	0.168	0.210	0.159	0.235	0.073	0.102	0.113
56.88	0.097	0.097	0.142	0.138	0.784	0.471	0.519

Teniendo en cuenta que son cinco los «Índices de Colinealidad» superiores a 30, el análisis de la tabla se centrará especialmente en estos valores. El estudio de los resultados obtenidos permite identificar que cuatro de las relaciones mostradas son equivalentes, lo cual dificulta la lectura e interpretación del papel desempeñado por cada predictor en las mismas. En estos casos la distribución de la variabilidad, a lo largo de los índices con valores similares, suele presentarse de forma arbitraria, por lo que el papel de cada una de las variables en dichas dependencias queda obscurecido. En este caso, la variabilidad de los coeficientes podría quedar distribuida a lo largo de los «Índices de Colinealidad» con relaciones equivalentes, sugiriéndose en tal caso, como punto de corte aproximado para pensar en la degradación de los coeficientes, que el sumatorio de dichas proporciones de varianza ascienda a .50 (Belsley, 1991), siendo evidente la alteración que sufre la variabilidad del término constante.

Por otro lado, partiendo de que el «Número de Condición» asciende a 56.88, podría considerarse que en cierta medida este índice está mostrando una depedencia dominante respecto a los anteriores, quedando claramente implicados el cuarto, quinto y sexto predictor del modelo. Este hecho puede dificultar la ubicación de aquellos predictores respecto a su posible implicación en otras dependencias. En este caso, por reducido que fuese el porcentaje de varianza que aportasen respecto a otras relaciones, no podría descartarse sin más su papel en aquellas dependencias. Evidentemente ninguno de los predictores está exento de formar parte de las dependencias que se presentan, pudiendo inferirse claramente la simultaneidad en la colinealidad del modelo. El estudio descriptivo que se llevó a cabo posteriormente se resume en la tabla siguiente.

Tabla 7

Resumen regresiones auxiliares

PREDICTOR	VARIABLES CON PESO REGRESION AUXILIARES	R^2
X_1	2, 3, 5 y 6	0.57
X_2	1, 3, 5 y 6	0.52
X_3	1 y 2	0.39
X_4	5 y 6	0.31
X_5	1, 2, 4 y 6	0.63
X_6	1, 2, 4 y 5	0.59

Como puede observarse, el análisis de datos reales se aleja considerablemente de aquellas situaciones ideales que se ofrecen como prototipo en los manuales, y en donde, sin excesivo problema, es factible seleccionar las variables más representativas en cada dependencia para su posterior estudio mediante análisis adicionales. Sin embargo, el enfoque utilizado permite analizar el número y tipo de dependencias así como la degradación sufrida por los coeficientes, poniéndose en evidencia el ciclo de estrechas interacciones que de forma tan usual se presentan en gran número de investigaciones, con la consiguiente alteración en los resultados, pero no siempre con incrementos notables en la variabilidad de los coeficientes.

De los resultados expuestos se desprende la importancia de la relación entre la variable «Pensamientos Depresivos» y las variables «Problemas Sociales» y «Culpabilidad», y cómo estas dos variables, a su vez, están relacionadas con «Percepción de Afecto Parental» y «Auto-Concepto Ético-Moral», y por tanto, la relación parcial de estas variables con «Pensamientos Depresivos». De igual manera, la interacción entre la variable «Auto-Concepto Físico» con las variables «Percepción de Afecto Parental» y «Auto-Concepto Ético- Moral» es destacada, de ahí la relación de esta variable con «Problemas Sociales» y «Culpabilidad». De todo ello se deduce, por tanto, la simultaneidad de las interacciones observadas, así como las alteraciones observadas como consecuencia del condicionamiento de los datos, y que podría sugerir su estudio a través de la consideración de efectos directos e indirectos.

3. CONCLUSIONES Y DISCUSIÓN

El tipo de datos que se han considerado en el presente estudio, y las relaciones que entre las variables se presentan, son un aspecto habitual en gran número de investigaciones sociales. A pesar de ello, no es usual que el investigador se detenga a analizar los posibles efectos de las relaciones entre los predictores, y menos aún si el nivel de alteración ocasionado pudiera cuestionar los resultados y conclusiones buscados.

Ahora bien, cuando en los resultados se hace referencia al empleo de un índice de diagnóstico para detectar la presencia de colinealidad, el más habitual es sin duda el *Factor de Inflación de la Varianza*, a pesar de contar con gran número de limitaciones, como es la ausencia de un punto de corte formal, su falta de precisión cuando la matriz de predictores es casi singular, su imposibilidad para facilitar la detección del número de conjuntos de variables implicados en dicho fenómeno, así como el nivel de alteración que estuviera afectando la estimación del modelo. La consideración de las raíces características de una matriz es un criterio más adecuado en la medida en que permite analizar con mayor precisión el número de relaciones colineales. Sin embargo, tampoco esta estrategia cuenta con un punto de corte formal para la identificación de dependencias, siendo de mayor utilidad la obtención de los valores singulares para la matriz de predictores, y a partir de aquí calcular otros instrumentos de diagnóstico más precisos como son los «Índices de Colinealidad» y la «Descomposición de la Varianza de los Coeficientes de Regresión en Proporciones». Los «Índices de Colinealidad» permiten analizar el número de relaciones colineales y su importancia, sin embargo, tienen el inconveniente de ser muy susceptibles ante variaciones en el tipo de escala, así como ante ligeras modificaciones en las observaciones que componen el conjunto muestral. Por tanto, de forma complementaria será conveniente proceder al análisis conjunto de estos índices respecto a la «Descomposición de la Varianza» de los coeficientes, con el fin de identificar el tipo y número de dependencias, implicación de los predictores en cada relación colineal, así como nivel de alteración sufrido por las estimaciones. Con este fin podrán incluirse análisis de regresión adicionales, así como la evalución de efectos directos e indirectos, como en la situación analizada en el presente estudio.

Dependiendo del tipo de interés del analista este proceso de diagnóstico podría mejorarse mediante la utilización de estrategias que no sólo detecten la perturbación que presentan los coeficientes estimados, sino las nefastas consecuencias que se derivan. Ello requiere de estudios más sofisticados, como es la consideración de «elipsoides de isodensidad γ» (Belsley, 1991; Estarelles, Castro y Prieto, 1995), pudiendo complementarse esta información con un análisis pormenorizado de la influencia ejercida en el condicionamiento de los datos por las observaciones que componen el conjunto muestral (Estarelles, 1995). Estos criterios permitirán decidir con mayor precisión la posibilidad de utilizar estrategias de análisis alternativas. Las ventajas y aplicación de estos criterios de forma accesible para el usuario serán objeto de atención en futuros trabajos.

REFERENCIAS

[1] **Belsley, D.A.** (1984). «Eigenvector Weaknesess and Other Topics for Assessing Conditioning Diagnostics. Letters». *Technometrics*, **26**, 297–299.

[2] **Belsley, D.A.** (1991). *Conditioning Diagnosis*. New York. John Wiley & Sons.

[3] **Belsley, D.A., Kuh, E.** y **Welsch, R.E.** (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York. John Wiley & Sons.

[4] **Berk, K.N.** (1977). «Tolerance and Condition in Regression Computations». *J. of the American Statistical Association*, **72**, 863–866.

[5] **Chatterjee, S.** y **Price, B.** (1977). *Regression Analysis by Example*. New York. Wiley.

[6] **Estarelles, R.** (1989). *Colinealidad en el Análisis de Datos*. I Symposium de Metodología de las Ciencias del Comportamiento. Salamanca.

[7] **Estarelles, R.** (1995). *Identificación de Observaciones con Influencia en Datos Debilmente Condicionados*. 5ª Conferencia de Biometría. Valencia.

[8] **Estarelles, R., Castro, J.** y **Prieto, J.** (1995). *Criterios formales y gráficos en la detección de colinealidad*. 5ª Conferencia de Biometría. Valencia.

[9] **Estarelles, R.** y **Prieto, J.** (1995). *Colinealidad: Efectos, Detección y Tratamiento*. Ed. Orenga. Valencia.

[10] **Estarelles, R., Prieto, J., Castro, J.** y **Aragón, J.L.** (1995). *Componentes Principales* vs *Raíces Latentes como Técnicas Alternativas a OLS*. 5ª Conferencia de Biometría. Valencia.

[11] **Fox, J.** (1991). *Regression Diagnostics*. Newbury Park. Sage Publications.

[12] **Golub, G.H.** y **Reinsch, C.** (1970). «Singular Value Decomposition and Least-Squares Solutions». *Numerische Mathematik*, **14**, 403–420.

[13] **Golub, G.H.** y **Wilkinson, J.H.** (1966). «Note on the Iterative Refinement of Least Squares Solution». *Numer. Math.*, **9**, 139–148.

[14] **Gunst, R.F.** y **Mason, R.L.** (1977). «Advantages of Examining Collinearities in Regression Analysis». *Biometrics*, **33**, 249–260.

[15] **Johnston, J.** (1984). *Econometric Methods.* 3ª Ed. New York. MacGraw-Hill.

[16] **Jolliffe, I.T.** (1982). «A note on the Use of Principal Components in Regression». *Appl. Statist.*, **31**, 300–303.

[17] **Kendall, M.G.** (1957). *A Course in Multivariate Analysis.* Griffin. London.

[18] **Kloeck, T.** y **Mennes, L.B.M.** (1960). «Simultaneous Equations Estimation Based on Principal Components of Predeter mined Variables». *Econometrika*, **28**, 45–61.

[19] **Lawson, C.R.** y **Hanson, R.J.** (1974). *Solving Least-Squares Problems.* Prentice Hall. Englewood Cliffs, N.J.

[20] **Mandel, J.** (1982). «Use of Singular Decomposition in Regression Analysis». *Amer. Stat.*, **36**, 15–24.

[21] **Marquardt, D.W.** (1970). «Generalized Inverses, Ridge Regression, Biased Linear Estimation and Non-linear Estimation». *Technometrics*, **12**, 591–612.

[22] **Myers, R.H.** (1990). *Classical and Modern Regression with Applications.* Belmont. Duxbury Press.

[23] **Silvey, S.D.** (1969). «Multicollinearity and Imprecise Estimation». *J. of the Royal Stat. Soc.*, **Series B 31**, 539–552.

[24] **Stewart, G.W.** (1973). *Introduction to Matrix Computations.* Academic Press. New York.

[25] **Stewart, G.W.** (1987). «Collinearity and Least Squares Regression». *Statistical Science*, **1**, 68–100.

[26] **Turing, A.M.** (1948). «Rounding-off Errors in Matrix Processes». *Quart. J. Mech. Appl. Math.*, **1**, 287–308.

[27] **Wilkinson, J.H.** (1965). *The Algebraic Eigenvalue Problem.* Oxford University Press.

ENGLISH SUMMARY

SENSIBILITY ANALYSIS OF THE PROPER STRUCTURE IN DETECTING COLLINEARITY

ESTARELLES, R.*

PRIETO, J.*

Universidad de Valencia

The collinearity diagnosis has already been widely treated and due to this many strategies have been suggested. However, many proposals have been critized because of their drawbacks, given that they can neither determine statistical harm, nor identify properly the variables that cause the conditioning, among other things. The aim of this paper is to point out the advantages of using indexes based on the eigen-system of X'X, instead of some of the strategies more usually applied. This objective has been verified by means of an empirical research.

Keywords: Collinearity Diagnostic, The Variance Inflation Factor, Singular Decomposition, The Condition Number, Regression-Coefficient Variance Decomposition.

*Estarelles, R. y Prieto, J. Metodología de las Ciencias del Comportamiento. Facultad de Psicología. Universidad de Valencia.

–Received january 1996.

–Accepted december 1996.