Descomposiciones del R^2 y clasificación por cosenos cuadrados

CÓDIGO: 7097

Junio 2014

Resumen

Se presenta un estudio de los métodos propuestos por Fields y Cupé para descomponer el coeficiente de determinación R^2 del modelo de regresión lineal y se introduce un nuevo método basado en la estandarización multivariante, mediante el cual se obtiene una descomposición del R^2 que es suma de sólo valores positivos. También se estudian los cosenos cuadrados entre observaciones y su base ortonormal en el análisis de componentes principaless y se muestra que pueden servir para clasificar a las observaciones por relaciones lineales. Se ejemplifica el método clasificando países por sus características socioeconómicas.

Palabras clave: Coeficiente de determinación R^2, Análisis de componentes principales, Estandarización multivariante, Clasificación.

Clasificación JEL: C00, C10, C38.

I. Introducción

Cuando lo que se quiere en el modelo de regresión lineal múltiple es sopesar la importancia que una variable explicativa tiene para explicar la variabilidad de la variable dependiente entonce se puede pensar en separar el coeficiente de determinación R^2 por aportes de las variables explicativas del modelo, éste objetivo se se han trazado Fields (2000) y Cupé (2008) obteniendo diferentes descomposiciones del R^2.

La descomposición de Fields proporciona aportes de cada variable explicativa pero con algunos valores que pueden ser negativos, si lo que se pretende es cuantificar algo así como que cuánta de varianza aporta una variable explicativa entonces un aporte negativo no parece ser apropiado, pero si se lo puede utilizar para sopesar que tanta importancia puede tener una variable independiente en el modelo para explicar a la variable dependiente.

En la misma línea Cupé desarrolla métodos que él denomina 'descomposición ortogonal del R^2' y 'descomposición dual del R^2 por variables' además de otra técnica para observaciones que la llama 'descomposición dual del R^2 por observaciones' (en éste trabajo se denominarán primera, segunda y tercera descomposición de Cupé respectivamente), los resultados que obtiene brindan aportes individuales posititvos de las variables explicativas y aportes conjuntos de pares de variables explicativas que en algunos casos pueden ser negativos; en todo caso los aportes individuales que obtiene no pueden interpretarse como proporciones porque no lo son y en la segunda descomposición algunos pueden ser mayores a uno. Sin embargo al igual que la descomposición de Fields, pueden servir para sopesar que variables explicativas y/u observaciones influyen en la varianza de la variable dependiente proyectada.

En éste trabajo se presentan un método de descomponer el R^2 de tal manera que cada término se corresponda con cada variable explicativa, sin aportes conjuntos y con todos los aportes positivos; la descomposición propuesta se basa en la estandarización multivariante que se obtiene a través de una rotación en el análisis de componentes principales.

El coeficiente de determinación es la correlación al cuadrado de la variable dependiente con una combinación lineal de las variables explicativas, geométricamente es el coseno cuadrado del ángulo entre el vector que representa a la variable dependiente y el espacio que generan las variables explicativas, ésto ha dado pie a estudiar también los cosenos cuadrados de los ángulos entre las variables explicativas y sus componentes principales, y los cosenos cuadrados de los ángulos entre las observaciones y los vectores propios; de ésta manera se ha encontrado una manera alternativa al análisis de conglomerados para clasificar las observaciones considerando las relaciones lineales entre variables.

En la siguiente sección se analizan las descomposiciones del R^2 propuestas por Fields y Cupé, en la tercera sección se presenta la descomposición del R^2 por componentes principales y la descomposición por estandarización multivariante, en la cuarta se explica cómo clasificar las observaciones utilizando los cosenos cuadrados de los ángulos entre observaciones y su base ortonormal, finalmente en la quinta sección se presentan aplicaciones.

II. Descomposiciones del R^2

II.1. El modelo de regresión lineal

El modelo lineal ajustado se expresa como

$$\widehat{\mathbf{y}} = \mathbf{X}\mathbf{b} \tag{1}$$

o también como

$$\widehat{\mathbf{y}} = \mathbf{Q}\mathbf{y} \tag{2}$$

donde $\mathbf{X} = (x_{ij})$ es la matriz $n \times p$ de variables explicativas centradas respecto a sus medias correspondientes, $\mathbf{y} = (y_i)$ es el vector $n \times 1$ que contiene a la variable dependiente también centrada respecto a su media, $\widehat{\mathbf{y}} = (\widehat{y}_i)$ denota el ajuste de la variable dependiente, el vector de coeficientes es

$$\mathbf{b} = \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}'\mathbf{y} \tag{3}$$

y la matriz de proyección en el espacio que generan las variables explicativas es

$$\mathbf{Q} = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}' \tag{4}$$

que se observa es idempotente; multiplicando a (3) por $(n-1)/(n-1)$ o por n/n se obtiene la expresión

$$\mathbf{b} = \mathbf{S}_{xx}^{-1}\mathbf{s}_{xy} \tag{5}$$

donde $\mathbf{S}_{xx} = (s_{jk})$ es la matriz de varianzas-covarianzas de $\mathbf{X}$, con $s_{jj} = s_j^2$ como la varianza de x_j y s_{jk} $(j \neq k)$ como la covarianza entre x_j y x_k, $\mathbf{s}_{xy} = (s_{jy})$ es el vector de covarianzas cruzadas entre $\mathbf{X}$ y $\mathbf{y}$ respectivamente, con s_{jy} como la covarianza entre x_j y y. El R^2 es por definición

$$R^2 = \frac{\widehat{\mathbf{y}}'\widehat{\mathbf{y}}}{\mathbf{y}'\mathbf{y}} = \frac{s_{\widehat{y}}^2}{s_y^2} \tag{6}$$

donde s_y^2 es la varianza de y y $s_{\widehat{y}}^2$ la varianza de $\widehat{y}$. Estandarizando las variables se obtiene el vector de coeficientes estandarizados

$$\mathbf{c} = \mathbf{R}_{xx}^{-1}\mathbf{r}_{xy} \tag{7}$$

donde $\mathbf{R}_{xx} = (r_{jk})$ es la matriz de correlaciones de $\mathbf{X}$, con $r_{jj} = 1$ y r_{jk} $(j \neq k)$ como la correlación entre x_j y x_k, $\mathbf{r}_{xy} = (r_{jy})$ es el vector de correlaciones cruzadas entre $\mathbf{X}$ y $\mathbf{y}$ respectivamente, con r_{jy} como la correlación entre x_j y y, la varianza residual de ésta regresión con variables estandarizadas es $1 - R^2$ y la relación entre los coeficientes usuales y los estandarizados es

$$\mathbf{c} = \frac{1}{s_y}\mathbf{D}_x\mathbf{b} \tag{8}$$

donde $\mathbf{D}_x$ es una matriz diagonal que contiene las desviaciones estándar de las variables explicativas en la diagonal principal, el j-ésimo coeficiente estandarizado es

$$c_j = \widetilde{\mathbf{r}}_j'\mathbf{r}_{xy} = \sum_{k=1}^{p} \widetilde{r}_{jk} r_{ky} \tag{9}$$

donde $\widetilde{\mathbf{r}}_j'$ con términos $\widetilde{r}_{jk}$ es la j-ésima fila de $\mathbf{R}_{xx}^{-1}$, Peña (2002) muestra que el término

$$\widetilde{r}_{jj} = \frac{1}{1 - R_{j.r}^2}$$

donde $R_{j.r}^2$ es el coeficiente de determinación de la regresión de x_j sobre el resto de las variables explicativas (r representa el resto de variables independientes, es decir sin x_j), y los términos

$$\widetilde{r}_{jk} = \frac{-c_{jk}}{1 - R_{j.r}^2}$$

para $k \neq j$, donde c_{jk} son los coeficientes de x_k en la regresión múltiple de x_j sobre el resto de las explicativas, utilizando éstos resultados se reescribe (9) como

$$c_j = \frac{r_{jy} - \mathbf{c}_j'\mathbf{r}_{ry}}{1 - R_{j.r}^2} \tag{10}$$

donde $\mathbf{r}_{ry}$ es el vector de correlaciones cruzadas del resto de las independientes con y, $\mathbf{c}_j = \mathbf{R}_{rr}^{-1}\mathbf{r}_{rj}$ con $\mathbf{R}_{rr}^{-1}$ como la inversa de la matriz de correlaciones del resto de las explicativas y $\mathbf{r}_{rj}$ como el vector de correlaciones cruzadas del resto de las independientes con x_j, utilizando éste hecho y reescribiendo adecuadamente

$$c_j = \frac{r_{jy} - \mathbf{r}_{jr}\mathbf{R}_{rr}^{-1}\mathbf{r}_{ry}}{\sqrt{1-R_{j.r}^2}\sqrt{1-R_{y.r}^2}}\frac{\sqrt{1-R_{y.r}^2}}{\sqrt{1-R_{j.r}^2}} = r_{jy.r}\frac{\sqrt{1-R_{y.r}^2}}{\sqrt{1-R_{j.r}^2}} \tag{11}$$

donde $r_{jy.r}$ es el coeficiente de correlación parcial entre x_j y y controlando por el resto de las explicativas, $R_{y.r}^2$ es el coeficiente de determinación de la regresión de y sobre el resto de las explicativas y $\mathbf{r}_{jr} = \mathbf{r}'_{rj}$. Por tanto se tiene que el j-ésimo coeficiente estandarizado es proporcional al coeficiente de correlación parcial correspondiente y conserva el signo de la correlación parcial dado que la razón de desviaciones estándar de los residuales de ambas regresiones es positiva, finalmente el coeficiente de regresión parcial es

$$b_j = \frac{s_y}{s_j}c_j = r_{jy.r}\frac{s_y}{s_j}\frac{\sqrt{1-R_{y.r}^2}}{\sqrt{1-R_{j.r}^2}}$$

también proporcional al coeficiente de correlación parcial.

II.2. Descomposición de Fields

La descomposición de Fields se obtiene reemplazando apropiadamente (1) y (2) en (6)

$$R^2 = \frac{\mathbf{b}'\mathbf{X}'\mathbf{Q}\mathbf{y}}{\mathbf{y}'\mathbf{y}} = \frac{\mathbf{b}'\mathbf{X}'\mathbf{y}}{\mathbf{y}'\mathbf{y}} = \frac{\mathbf{b}'\mathbf{s}_{xy}}{s_y^2} = \mathbf{c}'\mathbf{r}_{xy} \tag{12}$$

luego escribiendo (12) en notación escalar se obtiene la descomposición

$$R^2 = \sum_{j=1}^{p} c_j r_{jy} \tag{13}$$

los términos de ésta suma son utilizados para medir la importancia que cada variable independientes tiene en la explicación de la variación de la variable dependiente; reemplazando (11) en los sumandos de (13) se tiene que

$$c_j r_{jy} = r_{jy.r} r_{jy}\frac{\sqrt{1-R_{y.r}^2}}{\sqrt{1-R_{j.r}^2}}$$

lo que da una expresión para cada aporte que puede ser estudiada mejor, se puede observar así que alguna participación puede ser negativa cuando la correlación entre una variable explicativa con la variable dependiente sea de un signo dado y la correlación parcial entre las mismas variables controlando por el resto de las explicativas sea del signo contrario, en otras palabras se puede decir también que un aporte será negativo cuando la relación lineal sin controlar por las otras variables explicativas sugiera lo contrario a lo que señala la relación controlando por las demás variables explicativas, ésto generalmente se da cuando dos o más variables explicativas son colineales.

II.3. Descomposiciones de Cupé

La primera descomposición de Cupé se obtiene reemplazando (1) en (7)

$$R^2 = \frac{\mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b}}{\mathbf{y}'\mathbf{y}} = \frac{\mathbf{b}'\mathbf{S}_{xx}\mathbf{b}}{s_y^2} = \mathbf{c}'\mathbf{R}_{xx}\mathbf{c} \tag{14}$$

ésta no es de hecho la expresión que el autor utiliza pero proporciona los mismos resultados[1], escribiendo en notación escalar se obtiene la expansión

$$R^2 = \sum_{j=1}^{p} c_j^2 + 2\sum_{j=1}^{p-1}\sum_{k=j+1}^{p} c_j c_k r_{jk} \tag{15}$$

los p primeros términos de la suma se utilizan para medir el grado de participación que cada variable independiente tiene individualmente en la explicación de la variabilidad de la variable dependiente y los $p\left(p-1\right)/2$ últimos términos proporcionan las participaciones combinadas de cada par de variables explicativas, en ésta descomposición los aportes individuales son positivos pero no

[1] La expresión que utiliza Cupé en el numerador es $\mathbf{b}'\mathbf{X}'\mathbf{U}\mathbf{U}'\mathbf{X}\mathbf{b}$, donde $\mathbf{U}$ es $n \times p$ que proviene de la descomposición de valor singular $\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}'$, pero $\mathbf{Q} = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}' = \mathbf{U}\mathbf{U}'$, luego $\mathbf{U}\mathbf{U}'\mathbf{X} = \mathbf{Q}\mathbf{X} = \mathbf{X}$ y por tanto $\mathbf{b}'\mathbf{X}'\mathbf{U}\mathbf{U}'\mathbf{X}\mathbf{b} = \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b}$.

se garantiza lo mismo para los aportes conjuntos, reemplazando (11) en los sumandos de (15) se obtienen las ecuaciones

$$c_j^2 = r_{jy.r}^2 \frac{1 - R_{y.r}^2}{1 - R_{j.r}^2}$$

$$c_j c_k r_{jk} = r_{jy.r} r_{ky.r} r_{jk} \frac{1 - R_{y.r}^2}{\sqrt{1 - R_{j.r}^2}\sqrt{1 - R_{k.r}^2}}$$

se observa en éste caso que las participaciones conjuntas serán negativas cuando una o las tres correlaciones tengan el signo negativo.

La segunda descomposición de Cupé se obtiene reemplazando primero (3) y luego (4) en (7)

$$R^2 = \frac{\mathbf{y}'\mathbf{Q}\mathbf{y}}{\mathbf{y}'\mathbf{y}} = \frac{\mathbf{y}'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}}{\mathbf{y}'\mathbf{y}} = \frac{\mathbf{s}_{yx}\mathbf{S}_{xx}^{-1}\mathbf{s}_{xy}}{s_0^2} = \mathbf{r}_{yx}\mathbf{R}_{xx}^{-1}\mathbf{r}_{xy} \tag{16}$$

ésta tampoco es la expresión que el autor utiliza, pero igual que en el anterior caso proporciona los mismos resultados[2], escribiendo (12) en notación escalar se tiene

$$R^2 = \sum_{j=1}^{p} \widetilde{r}_{jj} r_{jy}^2 + 2\sum_{j=1}^{p-1}\sum_{k=j+1}^{p} \widetilde{r}_{jk} r_{jy} r_{ky} \tag{17}$$

siendo los p primeros términos de la suma las participaciones individuales y los $p\left(p-1\right)/2$ siguientes términos las participaciones combinadas, reemplazando (11) en los sumandos de (17) se obtienen

$$\widetilde{r}_{jj} r_{jy}^2 = \frac{r_{jy}^2}{1 - R_{j.r}^2}$$

$$\widetilde{r}_{jk} r_{jy} r_{ky} = \frac{-c_{jk} r_{jy} r_{ky}}{1 - R_{j.r}^2} = \frac{-r_{jk.r\cap s} r_{jy} r_{ky}}{1 - R_{j.r}^2} \frac{\sqrt{1 - R_{j.r\cap s}^2}}{\sqrt{1 - R_{k.r\cap s}^2}}$$

[2]Cupé utiliza en el numerador $\mathbf{y}'\mathbf{X}\mathbf{V}\mathbf{D}^{-2}\mathbf{V}'\mathbf{X}'\mathbf{y}$ pero por la descomposición espectral $\mathbf{V}\mathbf{D}^{-2}\mathbf{V}' = \left(\mathbf{X}'\mathbf{X}\right)^{-1}$.

donde $r_{jk.s}$ es el coeficiente de correlación parcial de x_j y x_k controlando por el resto de explicativas (en éste caso s representa el conjunto de variables independientes excluyendo a x_k); en ésta descomposición los aportes individuales pueden ser mayores a uno cuando la correlación al cuadrado entre y y x_j sea lo suficientemente mayor a la varianza residual de la regresión estandarizada de x_j sobre el resto de las explicativas, ésto indicaría un alto grado de colinealidad entre variables independientes; en cuanto a los aportes conjuntos, éstos serán negativos cuando dos o tres correlaciones sean positivas.

Por último en esta sub-sección se presenta la tercera descomposición de Cupé, la misma se obtiene reemplazando (3) en (7)

$$R^2 = \frac{\mathbf{y}'\mathbf{Q}\mathbf{y}}{\mathbf{y}'\mathbf{y}} = \sum_{i=1}^{n} \frac{q_{ii} y_i^2}{\sum_{k=1}^{n} y_k^2} + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{q_{ij} y_i y_j}{\sum_{k=1}^{n} y_k^2} \tag{18}$$

donde q_{ij} es el elemento ij-ésimo de $\mathbf{Q}$, no es la expresión que utiliza Cupé pero proporciona los mismos resultados[3], ésta descomposición es parecida a las anteriores, los aportes conjuntos puede proporcionar valores negativos, sin embargo en éste trabajo no se la analizará más.

III. Descomposición del R^2 por estandarización multivariante

Si la regresión de y se realiza sobre los componentes principales de las variables explicativas entonces la descomposición del R^2 se puede expresar como la suma de las correlaciones al cuadrado entre la variable dependiente y los componentes, para ver esto se expresa la descomposición espectral de la matriz

[3] El numerador en la expresión de Cupé (presentado aquí en notación matricial) es $\widehat{\mathbf{y}}'\widehat{\mathbf{y}} = \mathbf{y}'\mathbf{U}\mathbf{U}'\mathbf{y}$ pero como se vio $\mathbf{Q} = \mathbf{U}\mathbf{U}'$.

de correlaciones de las variables explicativas

$$\mathbf{R}_{xx} = \mathbf{A}\mathbf{D}_\lambda\mathbf{A}' \tag{19}$$

donde $\mathbf{D}_\lambda$ es una matriz diagonal $p \times p$ que contiene los valores propios de $\mathbf{R}_{xx}$ en la diagonal principal ordenados de mayora a menor, $\mathbf{A}$ es la matriz que en columnas contiene los vectores propios de $\mathbf{R}_{xx}$ asociados a los valores propios respectivos, se trata de una matriz ortogonal puesto que $\mathbf{A}'\mathbf{A} = \mathbf{I}_p$, los componentes principales son

$$\mathbf{Z} = \mathbf{X}\mathbf{D}_x^{-1}\mathbf{A} \tag{20}$$

puesto que al obtener la descomposición de la matriz de correlaciones entonces se utilizan los datos estandarizados para obtener los componentes, el vector de los coeficientes de regresión estandarizados es

$$\mathbf{g} = \mathbf{R}_{zz}^{-1}\mathbf{r}_{zy} = \mathbf{r}_{zy} \tag{21}$$

puesto que la matriz de correlaciones de los componentes $\mathbf{R}_{zz}^{-1}$ es la identidad y $\mathbf{r}_{zy}$ es el vector de correlaciones entre los componentes y la variable dependiente, entonces cada coeficiente de regresión es

$$g_k = \underline{r}_{ky} \tag{22}$$

donde $\underline{r}_{ky}$ es la correlación de z_k con y. Reemplazando (19) en (14) donde corresponde se obtiene una descomposición del R^2 que es suma de sólo valores positivos que se corresponden con cada componente

$$R^2 = \mathbf{c}'\mathbf{A}\mathbf{D}_\lambda\mathbf{A}'\mathbf{c} = \mathbf{g}'\mathbf{R}_{zz}^{-1}\mathbf{g} = \mathbf{g}'\mathbf{g} \tag{23}$$

por tanto $\mathbf{g} = \mathbf{D}_\lambda^{1/2}\mathbf{A}'\mathbf{c}$, expandiendo ésta suma y utilizando (22) se tiene

$$R^2 = \sum_{k=1}^{p} g_k^2 = \sum_{k=1}^{p} \underline{r}_{ky}^2 \tag{24}$$

con lo que se obtiene que el R^2 es la suma de los cuadrados de las correlaciones de cada componente con la variable dependiente, es decir, es la suma de los coeficientes de determinación simples de y con cada z_k.

En ésta descomposición los términos son positivos y se encuentran entre cero y uno, no existen aportes conjuntos porque los componentes están incorrelacionados, luego se pueden interpretar como la proporción de la variación de la variable dependiente que es explicada por cada componente; si cada componente captura aspectos comunes de las variables que pueden considerarse como factores latentes, entonces se tiene una descomposición por éstos factores que además son ortogonales, es decir, cada uno explica aspectos muy distintos de la varianza de la variable dependiente.

En vez de $\mathbf{R}_{xx}$ se puede utilizar también $\mathbf{S}_{xx}$ en la descomposición espectral y reemplazar en (14) donde corresponde, sin embargo los resultados de las participaciones de los componentes serán diferentes dado que los valores y vectores propios cambiarán. Si las variables están en diferentes unidades es más conveniente utilizar la matriz de correlaciones. Un problema adicional a la descomposición por componentes principales (que también se da en la regresión sobre los componentes) es que no siempre será claro que es lo que están representando tales componentes, luego de hecho no se obtienen las participaciones de las variables originales.

Para obtener una descomposición donde las participaciones estén más relacionadas con las variables originales podría considerarse la estandarización multivariante, la cual aplica una rotación a la matriz $\mathbf{DA}'$ utilizando la misma $\mathbf{A}$ y permite así obtener una ortogonalización de las variables originales de tal manera que las nuevas variables tengan matriz de covarianzas identidad y la correlación de cada una con la original que le corresponde sea alta al mismo tiempo que la correlación con las otras que no les corresponden sean bajas.

Se muestra a continuación la estandarización multivariante, se parte de la siguiente factorización de la matriz de correlaciones de las variables explicativas

$$\mathbf{R}_{xx} = \mathbf{A}\mathbf{D}_{\lambda}^{1/2}\mathbf{A}'\mathbf{A}\mathbf{D}_{\lambda}^{1/2}\mathbf{A}' = \mathbf{P}\mathbf{P} \tag{25}$$

donde

$$\mathbf{P} = \mathbf{A}\mathbf{D}_{\lambda}^{1/2}\mathbf{A}' \tag{26}$$

es una matriz simétrica, las variables estandarizadas de manera multivariante son

$$\mathbf{W} = \mathbf{X}\mathbf{D}_{x}^{-1}\mathbf{P}^{-1} \tag{27}$$

cuya matriz de covarianzas es identidad

$$\frac{1}{n}\mathbf{W}'\mathbf{W} = \frac{1}{n}\mathbf{P}^{-1}\mathbf{D}_{x}^{-1}\mathbf{X}'\mathbf{X}\mathbf{D}_{x}^{-1}\mathbf{P}^{-1} = \mathbf{P}^{-1}\mathbf{R}_{xx}\mathbf{P}^{-1} = \mathbf{P}^{-1}\mathbf{P}\mathbf{P}\mathbf{P}^{-1} = \mathbf{I}_{p}$$

asimismo la matriz de correlaciones cruzadas entre las variables originales y las estandarizadas de manera multivariante es

$$\frac{1}{n}\mathbf{D}_{x}^{-1}\mathbf{X}'\mathbf{W} = \frac{1}{n}\mathbf{D}_{x}^{-1}\mathbf{X}'\mathbf{X}\mathbf{D}_{x}^{-1}\mathbf{P}^{-1} = \mathbf{R}_{xx}\mathbf{P}^{-1} = \mathbf{P}\mathbf{P}\mathbf{P}^{-1} = \mathbf{P}$$

donde p_{jk} es la correlación entre x_j con w_k, que es la misma que p_{kj} dado que $\mathbf{P}$ es simétrica. La correlación entre x_j y w_j es

$$p_{jj} = \underline{\mathbf{a}}_j'\mathbf{D}_{\lambda}^{1/2}\underline{\mathbf{a}}_j = \sum_{k=1}^{p} a_{jk}^2\sqrt{\lambda_k}$$

donde $\underline{\mathbf{a}}_j$ es la j-ésima fila de $\mathbf{A}$ expresada como vector columna, en la práctica ésta correlación tiende a uno especialmente si existe baja multicolinealidad, luego

$$p_{jk} = \underline{\mathbf{a}}_j'\mathbf{D}_{\lambda}^{1/2}\underline{\mathbf{a}}_k = \sum_{l=1}^{p} a_{jl}a_{kl}\sqrt{\lambda_l}$$

es la correlación entre x_j y w_k con $j \neq k$, en la práctica ésta correlación tiende a cero más que todo cuando la multicolinealidad no es severa.

El vector de los coeficientes estandarizados de la regresión de y sobre las w_k es

$$\mathbf{h} = \mathbf{R}_{ww}^{-1}\mathbf{r}_{wy} = \mathbf{r}_{wy} \tag{28}$$

donde al igual que en el anterior caso la matriz de correlaciones $\mathbf{R}_{ww}^{-1} = \mathbf{I}_p$ y $\mathbf{r}_{wy}$ es el vector de correlaciones entre las variables estandarizadas de forma multivariante y la variable dependiente, entonces

$$h_k = \widetilde{r}_{ky} \tag{29}$$

donde $\widetilde{r}_{ky}$ es la correlación de w_k con y. Reemplazando (25) en (14) donde corresponde se obtiene la descomposición del R^2 por estandarización multivariante

$$R^2 = \mathbf{c}'\mathbf{P}\mathbf{P}\mathbf{c} = \mathbf{h}'\mathbf{R}_{ww}^{-1}\mathbf{h} = \mathbf{h}'\mathbf{h} \tag{30}$$

de donde se deduce que $\mathbf{h} = \mathbf{P}\mathbf{c}$, expandiendo ésta suma

$$R^2 = \sum_{k=1}^{p} h_k^2 = \sum_{k=1}^{p} \widetilde{r}_{ky}^2 \tag{31}$$

por tanto el R^2 es al igual que con la anterior descomposición, una suma de cuadrados de correlaciones, en éste caso es la suma de los coeficientes de determinación simples de y con cada w_k, ésta última que es la x_k estandarizada de manera multivariante.

De ésta manera se obtiene una descomposición del R^2 donde también pueden interpretarse los aportes de cada variable como la proporción en que explican a la variable dependiente, pero a diferencia de la regresión sobre componentes, en éste caso cada variable ortogonal w_k está altamente correlacionada con x_k, es ortogonal a las otras variables estandarizadas y tiene baja correlación con las otras explicativas, por tanto su aporte puede interpretarse como la proporción en que la variable explicativa correspondiente explica a y, al menos aproximadamente o mediante su versión ortogonalizada.

Se pueden probar otras factorizaciones de $\mathbf{R}_{xx}$ o de $\mathbf{S}_{xx}$ en la descomposición del R^2, entre ellos la factorización de Cholesky u otra rotación a la matriz $\mathbf{DA}$, sin embargo los resultados serán distintos, en el caso de Cholesky habrá que seleccionar a una de las variables la cual será solo reescalada, las otras serán combinaciones lineales cada vez con más variables; mediante la rotación como por ejemplo Varimax, las variables ortogonalizadas se agruparán en grupos de variabilidad común; en todos esos casos no siempre se corresponderán las variables explicativas con cada variable ortogonal, o lo harán por grupos cuando haya alta multicolinealidad.

La descomposición del R^2 se puede extender al caso en que haya dos grupos de variables y el objetivo sea encontrar la correlación máxima que existe entre ellas, lo cual es el objetivo en el análisis de correlaciones canónicas, de hecho el coeficiente de determinación de la regresión múltiple es un caso particular del análisis de correlaciones canónicas, así el R^2 es una correlación canónica al cuadrado de dos grupos de variables donde uno de ellos solo contiene a una variable.

Para aplicar en el análisis de correlaciones canónicas, la descomposición del R^2 estudiada, se realiza la estandarización multivariante a ambos grupos de variables y así se obtienen las proporciones en que participan las variables de ambos grupos en las variables canónicas.

IV. Clasificación por cosenos cuadrados

El estudio de los coeficientes de determinación conduce a considerar los cuadrados de las correlaciones entre variables originales y componentes, los mismos son los cosenos cuadrados del ángulo entre las variable y los componentes en el espacio $\mathbb{R}^n$, más concretamente en el subespacio generado por las variables o los componentes; así los componentes principales son una base ortogonal de las variables y cuando se norman a uno son una base ortonormal que genera el subespacio en el que se encuentran las variables; ésto conduce a la descomposición de valor singular

$$\mathbf{X} = \mathbf{FDA}'$$

donde como ya se conoce $\mathbf{X}$ es $n \times p$ con $n > p$ y de rango completo, $\mathbf{F}$ son los componentes normados y $\mathbf{D}$ la matriz diagonal que contiene los valores singulares en la diagonal principal, si $\mathbf{F}$ es una base ortonormal para las columnas de $\mathbf{X}$

se tiene que $\mathbf{A}$ es una base ortonormal para las filas de $\mathbf{X}$, es decir, para las observaciones, las mismas que se encuentran en el espacio $\mathbb{R}^p$, én éste caso los vectores propios generan todo el espacio $\mathbb{R}^p$.

Los cosenos cuadrados de los ángulos de variables y observaciones respecto a sus bases proporcionarán una medida del grado de cercanía que tienen respecto a los vectores propios respectivos, a partir de esto se pueden clasificar las observaciones por su cercanía a las direcciones ortogonales, particularmente interesan las direcciones de máxima variabilidad; para precisar esto sean $\mathbf{x}_i$ las observaciones (filas de $\mathbf{X}$) expresadas como vectores columna

$$\mathbf{x}_i = \mathbf{A}\mathbf{z}_i \tag{32}$$

donde $\mathbf{z}_i$ es la i-ésima observación en los componentes, luego los cosenos del ángulo entre las observaciones y los vectores propios son

$$\cos\left(\theta_{ik}\right) = \frac{\mathbf{x}_i'\mathbf{a}_k}{\sqrt{\mathbf{x}_i'\mathbf{x}_i}\sqrt{\mathbf{a}_k'\mathbf{a}_k}} = \frac{z_{ik}}{\sqrt{\mathbf{z}_i'\mathbf{A}'\mathbf{A}\mathbf{z}_i}} = \frac{z_{ik}}{\sqrt{\sum_{k=1}^{p} z_{ik}^2}} \tag{33}$$

y los cosenos cuadrados son

$$\cos^2\left(\theta_{ik}\right) = \frac{z_{ik}^2}{\sum_{k=1}^{p} z_{ik}^2} \tag{34}$$

considerando el valor que toma $\cos^2(\theta_{ik})$ pueden clasificarse las observaciones de acuerdo a su proximidad a las direcciones de máxima variabilidad, es decir, se pueden establecer m grupos asociados a los m mayores valores propios de $\mathbf{R}_{xx}$[4], luego una observación se clasifica a un grupo si su coseno cuadrado del ángulo con el vector propio asociado al grupo es el mayor con respecto a los otros, las observaciones que no logran ser clasificadas a ninguno de los grupos preestablecidos pueden pertenecer a un "grupo residual" asociado a todos los vectores propios de los componentes de mínima varianza. Dentro de cada grupo las observaciones pueden clasificarse en grupos más pequeños mediante el signo de su correlación, así en total se estimarán $2m+1$ grupos.

Luego se pueden interpretar los grupos considerando la matriz de correlaciones cruzadas entre variables y componentes y las correlaciones al cuadrado de las mismas, la matriz de correlaciones cruzadas es

$$\frac{1}{n}\mathbf{D}_x^{-1}\mathbf{X}'\mathbf{Z}\mathbf{D}_\lambda^{-1/2} = \frac{1}{n}\mathbf{D}_x^{-1}\mathbf{X}'\mathbf{X}\mathbf{D}_x^{-1}\mathbf{A}\mathbf{D}_\lambda^{-1/2} = \mathbf{R}_{xx}\mathbf{A}\mathbf{D}_\lambda^{-1/2} = \mathbf{A}\mathbf{D}_\lambda^{1/2}$$

donde se ha utilizado (20), cada correlación es

$$\overline{r}_{jk} = a_{jk}\sqrt{\lambda_k}$$

y su cuadrado es

$$\overline{r}_{jk}^2 = a_{jk}^2\lambda_k$$

Para interpretar la clasificación se tiene que los dos primeros grupos se caracterizan por ser aquéllos cuyas observaciones tienen alta puntuación absoluta en las variables con mayor correlación al cuadrado con el primer componente, luego éstos dos grupos se diferencian en que unos tiene alta puntuación positiva y los otros alta puntuación negativa en el primer componente; los primeros

[4]La regla para seleccionar m puede ser alguna de las utilizadas tradicionalmente, es decir, la regla del promedio, la del codo o la del porcentaje.

tienen altos valores positivos en las variables con alta correlación positiva con el primer componente y altos valores negativos en las variables con alta correlación negativa con el primer componente; los otros son los que tienen altos valores positivos en las variables con alta correlación negativa con el primer componente y altos valores negativos en las variables con alta correlación positiva con el primer componente. Para el tercer y cuarto grupo es igual, sólo que con las variables que les corresponden según las correlaciones al cuadrado, y así sucesivamente.

V. Aplicaciones

Para ilustrar los métodos estudiados se han utilizado las siguientes variables: $Bien_i$ que es una medida del bienestar subjetivo promedio del i-ésimo país, IDH_i que es el índice de desarrollo humano del país i, $Democ_i$ que es el índice de democracia del i-ésimo país y Paz_i que es un indicador de paz global del país i, con $i = 1, 2, \ldots, 144$ países para el año 2012.

Los datos de bienestar se han obtenido de la encuesta de Gallup, el IDH del informe del PNUD, el índice de democracia de The Economist y el índice de paz de Paz Global, en anexos se presentan los valores de las variables para cada país.

V.1. Aplicación de la descomposición del R^2

Se han aplicado las distintas descomposiciones del R^2 a la regresión

$$\widehat{Bien}_i = 2{,}128 + 4{,}028 IDH_i + 0{,}104 Democ_i - 0{,}019 Paz_i$$

El objetivo de la regresión es verificar si el bienestar subjetivo de la población de un país puede ser explicado por su nivel de desarrollo humano, su calidad de democracia y/o el ambiente de paz en el que viven.

Las descomposiciones del $R^2 = 0{,}5651$ obtenidas se encuentran en el cuadro 1.

Cuadro 1: Descomposiciones del R^2

	Fields	Cupé 1	Cupé 2	Est Mult
IDH	$0{,}4551$	$0{,}3808$	$1{,}0008$	$0{,}3876$
$Democ$	$0{,}1134$	$0{,}0374$	$0{,}7186$	$0{,}1322$
Paz	$-0{,}0034$	$0{,}0001$	$0{,}3981$	$0{,}0453$
$IDH * Democ$		$0{,}1536$	$-0{,}7494$	
$IDH * Paz$		$-0{,}0051$	$-0{,}3419$	
$Democ * Paz$		$-0{,}0018$	$-0{,}4611$	

Fuente: Elaboración propia

Se obtiene que el desarrollo humano es la variable que más influye a la variación del bienestar subjetivo, pero sólo mediante la descomposición por la estandarización multivariante se puede interpretar en términos de porcentaje, así se tiene que la variación del bienestar subjetivo se explica en un 38,76 % por el nivel de desarrollo humano ortogonalizado, en un 13,22 % por la calidad de la democracia ortogonalizada y en un 4,53 % por el ambiente de paz ortogonalizado.

V.2. Aplicación de la clasificación por cosenos cuadrados

Se ha aplicado la clasificación por cosenos cuadrados a los mismos datos considerados en el anterior acápite, introduciendo en un análisis de componentes principales a las cuatro variables: Bienestar subjetivo, Nivel de desarrollo humano, Calidad de la democracia y Ambiente de paz, la clasificación obtenida es

Grupo 1 (observaciones con altos cosenos cuadrados en el primer componente y coseno positivo):

Argentina, Australia, Austria, Bélgica, Brasil, Canadá, Chile, Costa Rica, Croacia, Chipre, República Checa, Dinamarca, Ecuador, Estonia, Finlandia, Francia, Alemania, Grecia, Islandia, Irlanda, Italia, Japón, Corea del Sur, Lituania, Malasia, Mauricio, Moldavia, Holanda, Nueva Zelanda, Noruega, Panamá, Paraguay, Perú, Polonia, Portugal, Singapur, Eslovaquia, Eslovenia, España, Suecia, Suiza, Trinidad y Tobago, Reino Unido, Estados Unidos y Uruguay.

Grupo 2 (observaciones con altos cosenos cuadrados en el segundo componente y coseno positivo):

Botswana, Bulgaria, Ghana, Hungría, Indonesia, Mongolia, Namibia, Rumanía, Senegal y Tanzania.

Grupo 3 (observaciones con altos cosenos cuadrados en el primer componente y coseno negativo):

Afganistán, Argelia, Angola, Armenia, Azerbaiján, Bangladesh, Benin, Burkina Faso, Burundi, Camboya, Camerún, República Centroafricana, Chad, China, República del Congo, República Democrática del Congo, Costa de Marfil, Yibuti, Egipto, Etiopía, Guinea, Haití, Irán, Irak, Kenia, Kirguistán, Liberia, Madagascar, Malí, Mauritania, Marruecos, Mozambique, Myanmar, Nepal, Níger, Nigeria, Pakistán, Ruanda, Sierra Leona, Sudán, Siria, Tayikistán, Uganda, Uzbekistán, Yemen y Zimbabwe.

Grupo 4 (observaciones con altos cosenos cuadrados en el segundo componente y coseno negativo):

Belarus, Colombia, El Salvador, Honduras, Israel, Líbano, Libia, México, Rusia, Arabia Saudí, Tailandia, Turquía, Turkmenistán y Venezuela.

Grupo 5 (observaciones con altos cosenos cuadrados en los otros componentes):

Albania, Baréin, Bolivia, Bosnia Herzegovina, Cuba, República Dominicana, Georgia, Guatemala, Guyana, India, Jamaica, Jordania, Kazajstán, Kuwait, Laos, Latvia, Malawi, Nicaragua, Filipinas, Catar, Serbia, Sudáfrica, Sri Lanka, Macedonia, Túnez, Ucrania, Emiratos Árabes Unidos, Vietnam y Zambia.

Son dos los componentes que explican al menos el 80 % de la varianza, los vectores propios están en el cuadro 2.

Cuadro 2: Vectores propios

	Z_1	Z_2	Z_3	Z_4
Bienestar	$0{,}4957$	$-0{,}5887$	$0{,}1219$	$0{,}6268$
IDH	$0{,}5257$	$-0{,}3228$	$0{,}2073$	$-0{,}7593$
Democracia	$0{,}5081$	$0{,}2724$	$-0{,}8170$	$0{,}0129$
Paz	$0{,}4688$	$0{,}6892$	$0{,}5241$	$0{,}1747$

Fuente: Elaboración propia

Y las correlaciones al cuadrado entre variables y componentes están en el cuadro 3.

Cuadro 3: Cuadrados de las correlaciones entre variables y componentes

	Z_1	Z_2	Z_3	Z_4
Bienestar	$0{,}6941$	$0{,}2029$	$0{,}0051$	$0{,}0979$
IDH	$0{,}7807$	$0{,}0610$	$0{,}0146$	$0{,}1436$
Democracia	$0{,}7294$	$0{,}0434$	$0{,}2271$	$0{,}0000$
Paz	$0{,}6208$	$0{,}2782$	$0{,}0935$	$0{,}0076$

Fuente: Elaboración propia

Mediante los valores de las correlaciones al cuadrado y los signos de los valores de los vectores propios, se observa que el primer grupo de países se caracteriza por elevadas puntuaciones en las cuatro variables, el tercer grupo de países tiene bajas puntuaciones en las cuatro variables, el segundo grupo tiene valores altos en Paz pero bajos en Bienestar, el cuarto es al revés, tiene altos valores en Bienestar pero bajos en Paz, finalmente el grupo quinto son países que no tienen valores muy altos ni muy bajos en general en las variables de estudio.

V.3. Código en E-views

En esta parte se muestra el código de los programas para obtener la descomposición del R^2 y la clasificación por cosenos cuadrados.

Código de la Descomposición del R^2 (donde se debe ingresar una ecuación estimada por mínimos cuadrados ordinarios)

```
if @uiedit(%ecua,"Ingrese el nombre de la ecuaci\U{f3}n") =-1 then
stop
endif
{%ecua}.makeregs gav
!p=gav.@count-1
stom(gav,yx)
sym correls=@cor(yx)
sym rxx=@subextract(correls,2,2)
vector rxy=@subextract(correls,2,1,!p+1,1)
vector cest=@inverse(rxx)*rxy
table(1+!p+!p*(!p-1)/2,6) desc_r2
desc_r2(1,2)="Fields"
desc_r2(1,3)="Cup\U{e9} 1"
desc_r2(1,4)="Cup\U{e9} 2"
```

```
desc_r2(1,5)="Est Mult"
vector aportes1=@emult(cest,rxy)
for !i=1 to !p
desc_r2(!i+1,1)=gav.@seriesname(!i+1)
desc_r2(!i+1,2)=aportes1(!i)
next
sym aportes2=@emult(cest*@transpose(cest),rxx)
for !i=1 to !p
desc_r2(!i+1,3)=aportes2(!i,!i)
next
!c=!p
for !i=1 to !p-1
for !j=!i+1 to !p
!c=!c+1
desc_r2(!c+1,1)=gav.@seriesname(!i+1)+"*"+gav.@seriesname(!j+1)
desc_r2(!c+1,3)=2*aportes2(!i,!j)
next
next
aportes2=@emult(rxy*@transpose(rxy),@inverse(rxx))
for !i=1 to !p
desc_r2(!i+1,4)=aportes2(!i,!i)
next
!c=!p
for !i=1 to !p-1
for !j=!i+1 to !p
!c=!c+1
desc_r2(!c+1,4)=2*aportes2(!i,!j)
```

```
next
next
matrix autoval=@makediagonal(@eigenvalues(rxx))
matrix autovec=@eigenvectors(rxx)
cest=autovec*@sqrt(autoval)*@transpose(autovec)*cest
aportes1=@emult(cest,cest)
for !i=1 to !p
desc_r2(!i+1,5)=aportes1(!i)
next
```

Código para la clasificación por cosenos cuadrados (donde se debe ingresar un grupo con las variables de interés)

```
if @uiedit(%grupo,"Ingrese el nombre del grupo") =-1 then
stop
endif
!p={%grupo}.@count
%comps="comp_1"
for !i=2 to !p
%comps=%comps+" comp_"+@str(!i)
next
{%grupo}.makepcomp(eigval=aval,eigvec=avec) {%comps}
matrix corr2=@emult(avec,avec)*@makediagonal(aval)
!suma=0
for !i=1 to !p
!suma=!suma+aval(!i)/@sum(aval)
if !suma>0.8 then
!m=!i
```

```
exitloop
endif
next
genr sumacuads=comp_1^2
for !i=2 to !p
%var=@word(%comps,!i)
sumacuads=sumacuads+{%var}^2
next
for !i=1 to !p
%var=@word(%comps,!i)
genr coscuad_!i={%var}^2/sumacuads
next
%coscuads="coscuad_1"
for !i=2 to !p
%coscuads=%coscuads+" coscuad_"+@str(!i)
next
group grupocomps {%comps}
group grupocos2 {%coscuads}
stom(grupocomps,comp)
stom(grupocos2,cos2)
vector(@rows(comp)) clasif=2*!m+1
for !i=1 to @rows(comp)
for !j=1 to !m
if cos2(!i,!j)=@max(@rowextract(cos2,!i)) then
if comp(!i,!j)>0 then
clasif(!i)=!j
else
```

```
clasif(!i)=!j+!m
endif
exitloop
endif
next
next
smpl @all if not @isna(comp_1)
mtos(clasif,clasific)
smpl @all
```

Referencias

[1] Cuadras, C. (1981). *Métodos de Análisis Multivariante,* 3ra Ed., Eunibar, Barcelona, España.

[2] Cupé, E. (2007). "Descomposición en Regresión Lineal: Un Nuevo Método para el Análisis de Determinantes y la Toma de Decisiones", *Investigación & Desarrollo No 7*, Universidad Privada Boliviana, Cochabamba, Bolivia.

[3] Cupé, E. (2008). "Descomposición Dual de R2 en Modelos de Regresión Lineal", *Investigación & Desarrollo No 8*, Universidad Privada Boliviana, Cochabamba, Bolivia.

[4] Fields, G. (2004) Regression-Based Decompositions: A New Tool for Managerial Decision-Making", Departamente of Labor Economics, Cornell University.

[5] Peña, D. (2002). *Análisis de Datos Multivariantes*, Mc-Graw Hill Interamericana, Madrid, España.

Anexos

Datos utilizados

PAIS	BIENESTAR	IDH	DEMOC	PAZ
Afganistán	4.75838086	0.374	2.48	0.748
Albania	5.2689366	0.749	5.67	2.073
Argelia	5.23696071	0.713	3.83	1.745
Andorra		0.846		
Angola	4.20609164	0.508	3.35	1.895
Antigua y Barbuda		0.76		
Argentina	6.4410672	0.811	6.84	2.237
Armenia	4.36781129	0.729	4.09	1.762
Australia	7.40561615	0.938	9.22	2.506
Austria	7.34603596	0.895	8.62	2.672
Azerbaiján	4.21861082	0.734	3.15	1.64
Bahamas		0.794		
Baréin	4.54946585	0.796	2.53	1.753
Bangladesh	4.98564918	0.515	5.86	1.929
Barbados		0.825		
Belarus	5.52592342	0.793	3.04	1.792
Bélgica	6.85351408	0.897	8.05	2.624
Belice	6.45064432	0.702		
Benin	3.66713955	0.436	6	1.769
Bhutan		0.538	4.65	2.519
Bolivia	5.78062029	0.675	5.84	1.979
Bosnia Herzegovina	4.66851741	0.735	5.11	2.077
Botswana	3.55302015	0.634	7.85	2.379
Brasil	6.83733119	0.73	7.12	1.983
Brunei Darussalam		0.855		
Bulgaria	4.22036745	0.782	6.72	2.301
Burkina Faso	4.03556044	0.343	3.52	2.119
Burundi	3.79168079	0.355	3.6	1.476
Camboya	4.16122537	0.543	4.96	1.793
Camerún	4.4338852	0.495	3.44	1.887
Canadá	7.65034626	0.911	9.08	2.683
Cabo Verde		0.586	7.92	

PAIS	BIENESTAR	IDH	DEMOC	PAZ
República Centroafricana	3.56789246	0.352	1.99	1.128
Chad	3.74287115	0.34	1.62	1.329
Chile	6.6356557	0.819	7.54	2.384
China	4.65273664	0.699	3	1.939
Colombia	6.40811371	0.719	6.63	1.375
Comores	3.92351296	0.429	3.52	
República del Congo	3.81979221	0.534	2.89	1.852
República Democrática del Congo	3.98384857	0.304	1.92	0.927
Costa Rica	7.27105381	0.773	8.1	2.341
Costa de Marfil	4.19718152	0.432	3.25	1.581
Croacia	5.59557522	0.805	6.93	2.352
Cuba	5.41786848	0.78	3.52	2.049
Chipre	6.38654611	0.848	7.29	2.043
República Checa	6.15220361	0.873	8.19	2.604
Dinamarca	7.77051528	0.901	9.52	2.761
Yibuti	5.00581087	0.445	2.74	2.119
Dominica		0.745		
República Dominicana	4.73502123	0.702	6.49	1.932
Ecuador	5.83805151	0.724	5.78	1.972
Egipto	3.87692247	0.662	4.56	1.78
El Salvador	6.73991115	0.68	6.47	1.78
Guinea Ecuatorial		0.554	1.83	1.961
Eritrea		0.351	2.4	1.736
Estonia	5.13773884	0.846	7.61	2.285
Etiopía	4.37004393	0.396	3.72	1.496
Fiji		0.702	3.67	
Finlandia	7.39326421	0.892	9.06	2.652
Francia	6.79790111	0.893	7.88	2.29
Gabón		0.683	3.56	2.028
Gambia		0.439	3.31	2.039
Georgia	4.10183709	0.745	5.53	1.459
Alemania	6.72453096	0.92	8.34	2.576

PAIS	BIENESTAR	IDH	DEMOC	PAZ
Ghana	4.60625182	0.558	6.02	2.193
Grecia	5.83955864	0.86	7.65	2.024
Granada		0.77		
Guatemala	6.28974871	0.581	5.88	1.713
Guinea	4.04456972	0.355	2.79	1.927
Guinea-Bissau		0.364	1.43	
Guyana	5.99282647	0.636	6.05	2.063
Haití	3.76599873	0.456	3.96	1.821
Honduras	5.86613151	0.632	5.84	1.661
Hong Kong	5.6428346	0.906	6.42	
Hungría	4.7251327	0.831	6.96	2.524
Islandia	6.88828417	0.906	9.65	2.887
India	4.9892774	0.554	7.52	1.451
Indonesia	5.45729941	0.629	6.76	2.087
Irán	4.76750686	0.742	1.98	1.676
Irak	5.01692081	0.59	4.1	0.808
Irlanda	7.25738964	0.916	8.56	2.672
Israel	7.3589161	0.9	7.53	1.158
Italia	6.35423798	0.881	7.74	2.31
Jamaica	6.20788195	0.73	7.39	1.778
Japón	6.05675259	0.912	8.08	2.674
Jordania	5.69618056	0.7	3.76	2.096
Kazajstán	5.51428658	0.754	2.95	1.849
Kenia	4.25585915	0.519	4.71	1.748
Kiribati		0.629		
Corea del Sur	6.11602441	0.909	8.13	2.266
Corea del Norte			1.08	1.068
Kuwait	6.58116458	0.79	3.78	2.208
Kirguistán	4.99641081	0.622	4.69	1.641
Laos	5.04409897	0.543	2.32	2.338
Latvia	4.66891042	0.814	7.05	2.226
Líbano	5.18167627	0.745	5.05	1.541

PAIS	BIENESTAR	IDH	DEMOC	PAZ
Lesotho		0.461	6.66	2.136
Liberia	4.19606317	0.388	4.95	1.869
Libia	4.9206923	0.769	5.15	1.17
Liechtenstein		0.883		
Lituania	5.0658249	0.818	7.24	2.259
Luxemburgo	7.09725176	0.875	8.88	
Madagascar	4.64007892	0.483	3.93	1.876
Malawi	5.1482395	0.418	6.08	2.106
Malasia	5.5802816	0.769	6.41	2.515
Maldivas		0.688		
Malí	3.76230498	0.344	5.12	1.868
Malta	5.77387482	0.847	8.28	
Mauritania	4.97416914	0.467	4.17	1.699
Mauricio	5.47707298	0.737	8.17	2.513
México	6.80238863	0.775	6.9	1.555
Micronesia		0.645		
Moldavia	5.58973661	0.66	6.32	2.073
Mongolia	4.5855235	0.675	6.35	2.116
Montenegro		0.791	6.05	1.994
Marruecos	4.38324734	0.591	4.07	2.133
Mozambique	4.65358324	0.327	4.88	2.204
Myanmar	5.32197694	0.498	2.35	1.475
Namibia	4.88558712	0.608	6.24	2.196
Nepal	3.80944466	0.463	4.16	1.999
Holanda	7.50187584	0.921	8.99	2.394
Nueva Zelanda	7.2237564	0.919	9.26	2.761
Nicaragua	5.68669935	0.599	5.56	1.994
Níger	4.10101619	0.304	4.16	1.759
Nigeria	4.76027579	0.471	3.77	1.199
Noruega	7.63228755	0.955	9.93	2.52
Palestina	4.84521638	0.67	4.8	
Omán		0.731	3.26	2.113

PAIS	BIENESTAR	IDH	DEMOC	PAZ
Pakistán	5.26718614	0.515	4.57	1.167
Palau		0.791		
Panamá	7.32146749	0.78	7.08	2.101
Papúa Nueva Guinea		0.466	6.32	1.924
Paraguay	5.84117412	0.669	6.26	2.027
Perú	5.6127852	0.741	6.47	2.005
Filipinas	4.94151414	0.654	6.3	1.585
Polonia	5.78033017	0.821	7.12	2.476
Portugal	4.87272077	0.816	7.92	2.53
Catar	6.591604	0.834	3.18	2.605
Rumanía	4.90916595	0.786	6.54	2.373
Rusia	5.46477911	0.788	3.74	1.062
Ruanda	4.02976189	0.434	3.36	1.75
Saint Kitts y Nevis		0.745		
Santa Lucía		0.725		
San Vincente y las Granadinas		0.733		
Samoa		0.702		
Santo Tomé y Príncipe		0.525		
Arabia Saudí	6.72691831	0.782	1.71	1.822
Senegal	3.83420149	0.47	6.09	2.006
Serbia	4.46130441	0.769	6.33	2.08
Seychelles		0.806		
Sierra Leona	4.13395606	0.359	4.71	2.145
Singapur	6.5314018	0.895	5.88	2.479
Eslovaquia	6.05222314	0.84	7.35	2.41
Eslovenia	6.08255519	0.892	7.88	2.67
Islas Salomón		0.53		
Somalia				0.608
Sudáfrica	4.65242858	0.629	7.79	1.679
España	6.18826265	0.885	8.02	2.452
Sri Lanka	4.18056926	0.715	5.75	1.855
Sudán	4.37509714	0.414	2.38	0.807

PAIS	BIENESTAR	IDH	DEMOC	PAZ
Suriname		0.684	6.65	
Swazilandia		0.536	3.2	1.972
Suecia	7.49601906	0.916	9.73	2.581
Suiza	7.52452064	0.913	9.09	2.651
Siria	4.06582428	0.648	1.63	1.17
Taiwán			7.57	2.398
Tayikistán	4.38063633	0.622	2.51	1.876
Tanzania	3.22912903	0.476	5.88	2.127
Tailandia	6.21670314	0.69	6.55	1.697
Macedonia	4.18020186	0.74	6.16	2.065
Timor-Leste		0.576	7.16	
Togo	2.80785514	0.459	3.45	
Tonga		0.71		
Trinidad y Tobago	6.69644393	0.76	6.99	1.918
Túnez	4.68598074	0.712	5.67	2.045
Turquía	5.49034716	0.722	5.76	1.656
Turkmenistán	6.56771327	0.698	1.72	1.758
Uganda	4.19288222	0.456	5.16	1.879
Ucrania	5.05756132	0.74	5.91	2.047
Emiratos Árabes Unidos	7.19680309	0.818	2.58	2.215
Reino Unido	7.02936431	0.875	8.21	2.391
Estados Unidos	7.16361619	0.937	8.11	1.942
Uruguay	6.06201095	0.792	8.17	2.372
Uzbekistán	5.09534226	0.654	1.72	1.781
Vanuatu		0.626		
Venezuela	7.47845437	0.748	5.15	1.722
Vietnam	5.76734461	0.617	2.89	2.359
Yemen	3.92414192	0.458	3.12	1.399
Zambia	5.26036083	0.448	6.26	2.17
Zimbabwe	4.84564185	0.397	2.67	1.462