ciento en la esperanza de vida femenina, esto se debe a que las tres variables independientes tienen un alto grado de colinealidad.

Este alto grado de colinealidad se puede apreciar directamente de la matríz de correlación que se presenta en el anexo, y de los errores estandar de ambas ecuaciones que, como ya dijimos, son los números que figuran entre paréntesis en las dos ecuaciones. La variable porcentaje de población rural en ambas ecuaciones tiene un error estandar mucho mayor que el valor del coeficiente, y las otras dos variables los errores estandar son mayores que la mitad de los coeficientes de regresión respectivos. La regla práctica para aceptar si una variable está realmente aportando algo en la explicación de la variable dependiente es que el error estandar de su coeficiente sea menor que la mitad de dicho coeficiente. Como acabamos de ver, esto no se da en ninguno de los casos de las ecuaciones que hemos ajustado, por lo que tendríamos que llegar a la conclusión que ninguna de las variables incluídas nos sirve. Aquí entramos en una aparente contradicción porque ya se había mencionado que el modelo se aceptaba con nivel inferior al uno al millar y ahora llegamos a la conclusión que ninguna variable nos sirve, esto obedece al alto grado de colinealidad que presentan las variables explicativas que hemos ya mencionado. En el siguiente punto volveremos sobre este tema.

APLICACION DEL METODO DE COMPONENTES PRINCIPALES

Acabamos de mencionar que las tres variables que hemos escogido como las explicativas del modelo de regresión tienen un alto grado de colinealidad, y por lo tanto, se viola el supuesto de "no multicolinealidad" que anteriormente dejamos asentado con el consiguiente problema de imprecisión en las estimaciones de los coeficientes parciales de regresión, cuyos errores estandar ya vimos son muy grandes.

Acerca de la violación de los otros supuestos que pueden acarrear problemas en los estimadores mínimo cuadráticos, para el supuesto de homosquedasticidad no se tiene suficiente información que permita verificarlo, y en cuanto al supuesto de no autocorrelación de los errores, dado que utilizamos una mezcla de datos transversales con datos longitudinales, existen ordenamientos adecuados que pasan la prueba de que los errores se comportan aleatoriamente.

Por lo antes dicho, nos centramos en un procedimiento para corregir el alto grado de colinealidad que presentan los datos, para lo cual vamos a utilizar el método de Componentes Principales.

Dicho método consiste en encontrar combinaciones lineales de las variables explicativas que se estén considerando, y que capten la mayor variación de la matriz de covarianzas de dichas variables, y al mismo tiempo sean linealmente independientes. El número de componentes principales es igual al número de variables. En nuestro caso son tres las variables, por lo que se tendrán tres combinaciones lineales de las mismas que son independientes.

Para la obtención de las componentes principales se procede mediante el cálculo de las raíces características de la matriz de covarianzas de las variables ya citadas. Se ordenan dichas raíces características en forma descendente de sus valores, y la mayor de ellas constituye la primer raíz característica y asi sucesivamente. Se puede demostrar que la traza de la matriz de covarianzas corresponde a la suma de las raíces características y el determinante de la matriz de covarianzas, que recibe el nombre de varianza generalizada, es igual al producto de las raíces características.

Tomando como variabilidad la suma de las varianzas de las combinaciones lineales que se buscan, y que corresponde a la traza de la matriz de covarianzas, entonces las raíces características mas grandes captan mayor proporción de esta variación, por lo tanto, si la primera raíz característica representa un alto porcentaje de dicha suma o variación, bastará tomar solamente esta primer raíz característica y descartar las restantes. En caso de que el porcentaje que capte de la variación no sea muy grande, se incluye también la segunda u otras más.

En nuestro caso, hemos tomado las tres variables explicativas del modelo, o sea el porcentaje de población rural, el porcentaje de fuerza de trabajo no agrícola y el porcentaje de población analfabeta, las cuales ya sabemos tienen un alto grado de colinealidad.

Al obtener las tres raíces características ha resultado que la primer raíz característica capta el 92.91% de la variación -en el anexo se presenta la información- por lo que hemos utilizado solamente esta raíz para nuestros objetivos.

A cada una de las raíces características está asociado un vector característico, cuyos elementos representan los coeficientes de la combinación lineal de las variables que están tratando. Dichos vectores característicos son ortonormales, y las combinaciones lineales que se obtienen de los mismos constituyen las componentes principales. Nosotros, como ya hemos dicho, trabajaremos solamente con la primera componente principal, cuya expresión es la siguiente:

$$Z_i = 0.6529FTNoA_i - 0.5229PAN_i - 0.5480PR_i$$

La variable Z_i que hemos obtenido, podemos interpretarla como un índice sintético de las condiciones socio-económicas, y sus valores representan la situación prevaleciente en el período que va de 1940 a 1970. La mayor ponderación corresponde a la Fuerza de Trabajo no Agrícola con signo positivo, el porcentaje de población analfabeta y el porcentaje de población rural tienen ponderaciones muy semejantes y ambas con signo negativo. Recordemos las dificultades que presenta la interpretación de las componentes principales la cual no es siempre clara, y que desde el punto de vista estadístico, sabemos que los vectores característicos representan los ejes de un elipsoide de concentración.

Conocidos los valores de las tres variables, se pueden obtener los correspondientes valores de Z_i, los cuales se presentan en un cuadro del anexo.

Con la variable Z_i, podemos establecer la regresión entre la esperanza de vida al nacimiento y dicha variable, de esta manera se elimina el problema de colinealidad, sin embargo, este procedimiento no nos permite contar con estimaciones de los coeficientes del modelo ori ginal.

Las ecuaciones resultantes son:

Para la esperanza de vida masculina:

$$e^{o}_{M} = 57.67 + 0.33Z_i$$
$$(0.039) \qquad r^2 = 0.837$$

Para la esperanza de vida femenina:

$$e^{o}_{F} = 60.46 + 0.36Z_i$$
$$(0.051) \qquad r^2 = 0.783$$

ambas ecuaciones se aceptan a niveles inferiores al uno al millar y también se aprecia que el error estandar de los coeficientes de regresión -número entre paréntesis en las ecuaciones- son muy pequeños, además, dichos coeficientes, considerando pruebas individuales, se acepta que su diferencia -entre 0.33 y 0.36- no es estadísticamente significativa por lo que aceptamos que estas rectas provienen de poblaciones con igual pendiente. En ambos casos, el valor del coeficiente de determinación es prácticamente igual al que se obtuvo en las ecuaciones de regresión múltiple del modelo original.

APLICACION

Tratando de encontrar una aplicación inmediata del modelo, se pensó en utilizar sus resultados para obtener estimaciones de las esperanzas de vida de algunos estados y compararlas con las que se tienen de tablas de mortalidad construidas para los mismos.