ANÁLISIS DEL PARO Y LA EMIGRACIÓN EN LA PROVINCIA DE LEÒN.

INTRODUCCIÒN:

La idea de este trabajo surge de las preocupaciones surgidas en los últimos años en la provincia de León, donde se observa un éxodo de la población hacia otras zonas de España, debida probablemente al desempleo que sufre la región. Esta situación es de sobra conocida por toda España, ya que es noticia habitualmente en los medios de comunicación los problemas que estamos teniendo con la reestructuración del sector de la minería. Pero no es un problema solamente contemporáneo pues ya en estudios referentes al siglo XIX pueden encontrarse referencias a una distribución de la población centrífuga en todo el país. Concretamente una superpoblación de las zonas periféricas o costeras (con mayores ventajas de comunicación y transporte), y una despoblación creciente del centro, con excepción naturalmente de Madrid, que era centro político. El análisis tratado en esta aplicación, intenta determinar si, ciertamente, la merma (por otro lado patente) de nuestra población, se debe al alto nivel de desempleo que sufrimos. O si por otra parte el desempleo no juega un papel tan importante como la gente opina, en la despoblación de la provincia de León, y la verdadera influencia proviene de otros factores.

Según Joaquín Leguina en sus Fundamentos de Demografía, esta disciplina tiene por objeto: El estudio de la estructura y la reproducción de la fuerza del trabajo. Afirmación que resulta interesante para un economista, ya que está hablando de uno de los dos factores fundamentales de la producción (capital y trabajo). Al tratar a la población desde el punto de vista de su utilidad para la producción, tratándola directamente según el concepto de capital humano. Por esta razón fundamentalmente el estudio de la población resulta tan interesante en economía.

El tema ha sido ampliamente tratado en diversos libros, como el de Luis Sastre, Distribución de la renta, mercados regionales de empleo y migraciones en España. En el cual examina la relación entre el desempleo y las migraciones interprovinciales, en nuestro país.

Este análisis se realiza partiendo de una serie de supuestos, ya contrastados en países de elevada movilidad del factor trabajo y que concreta en tres:

− Primero la relación con la actividad de un trabajador afecta a su movilidad; un trabajador empleado tiene menos probabilidad de moverse que un trabajador desempleado.

− Segundo las diferencias interprovinciales en el desempleo favorecen la movilidad; la probabilidad de que un trabajador emigre es más alta si el trabajador vive en provincias con elevado desempleo que si el trabajador vive en provincias con un desempleo inferior.

−Tercero las diferencias salariales deben favorecer la movilidad interprovincial del factor trabajo.

Es decir, que su análisis parece indicar en principio que la situación de desempleo de un individuo sería una causa de movilidad interregional, precisamente para la búsqueda de ese empleo. Aunque el cambio podría también producirse por causa de una diferencia salarial, aunque esto sería menos probable, ya que en los tiempos que corren pocos son los privilegiados que se arriesgan a un cambio de provincia por el mero echo de un salario mayor, y más aún teniendo en cuenta que no solo un nivel mayor en el sueldo garantizaría un nivel mayor de vida. Dependería directamente del coste de la vida en el país de origen. Estos análisis contrastados a los que hace referencia Luis Sastre parecen respaldar la teoría objeto de este estudio, es probable que el desempleo sea un sector desencadenante de la emigración.

La probabilidad de la emigración individual, es función de características personales y de variables de

mercado. Las características personales influyen en la decisión de emigrar, principalmente a través del coste subjetivo que genera la emigración. Las variables de mercado influyen a través de los beneficios netos que origina la emigración. Sin embargo, las características personales pueden, también, afectar en las decisiones de emigración a través de sus efectos en las ganancias potenciales.

En los supuestos teóricos, subyacentes en los modelos de capital humano, se supone que un emigrante tomará la decisión de trasladarse de una provincia a otra, si los beneficios de la emigración exceden del coste personal que conlleva la decisión de moverse. El coste de emigrar depende de un número de características observables (circunstancias familiares, edad, relación con la actividad, etc.) y algunos no observables.

Asumiendo que las características inobservables se distribuyen normalmente en el conjunto de la población, podemos describir la probabilidad de emigrar como una función logística de un vector de características personales observadas y de variables de mercado. El modelo ha sido desarrollado, en diversos estudios similares realizados en Inglaterra y Estados Unidos.

Se utiliza un modelo de Logit, que trata de medir las relaciones entre el empleo y las migraciones interprovinciales en España. Concluye que la probabilidad de no haber cambiado de provincia en el período intercensal 1.981−1.991 depende de muchos factores: en general la probabilidad calcula que es alta, en concreto de 0´98 %. Lo cual indicaría que hay escasa movilidad interprovincial en España en dicho período.

Por edades, emigran más los jóvenes, pues en una edad de 16 a 24 años, la probabilidad de no cambiar de lugar de residencia es 0´98 %. Circunstancia lógica teniendo en cuenta que el sector que busca el primer empleo tiene muchas dificultades para conseguirlo y parece normal que tengan que cambiar incluso de provincia para encontrarlo. Los más estables son los de 50 a 64 años ya que tienen una probabilidad de 0´99 %, de no cambiar de provincia este caso también es evidente pues a estas edades la vida está consolidada y cuesta más desprenderse de las raices y por lo tanto emigrar.

Con relación al nivel de estudios los niveles primarios son los más estáticos, con un 0´99 %, a continuación los niveles de estudios medios, con 0´97 % y por último los estudios superiores, con 0´95 %. En este sentido se justifica que los trabajadores más cualificados sean los más escasamente requeridos y por tanto con mayores necesidades de emigrar para conseguir un trabajo.

El análisis de ocupación revela que los habitantes de la provincia de León tienen una probabilidad de tener trabajo, de estar activos, entre los 16 y los 25 años de 0´28 %, de 0´36 % entre 25 y 36 años y 0´67 % entre 50 y 64 años. Lo cual concuerda con lo anterior de que serían los jóvenes los más propensos a la emigración, precisamente por esa causa de la mayor falta de empleo relativa y también probablemente por las circunstancias personales. Conforme va avanzado la edad las personas tienen más consolidada su situación laboral y personal, en general, y por lo tanto será menos probable que emigren que personas más jóvenes y con menos ataduras a priori.

Con nivel de estudios primarios hay un 0´62 % de probabilidad de tener empleo, y lo mismo con estudios medios, pero en el caso de estudios superiores es de 0´59 %. Lo cual también concordaría con el echo comentado antes de que sean los titulados en estudios superiores los que más posibilidades tengan de emigrar. Pues los trabajos altamente especializados son por la contra de los menos demandados.

Los varones tienen un 0´76 % de posibilidades de encontrar empleo. Y más probabilidades de emigrar que las mujeres, entre otros motivos quizás por circunstancias más bien sociales y culturales, que por las derivadas del mero echo de buscar un trabajo, cuestión que no se ve en principio influida por la diferencia de sexos.

Las conclusiones que establece Luis Sastre son:

−El desempleo aumenta la probabilidad de emigrar, lo cual se cumple para, prácticamente, todas las

provincias españolas, así como para el Total nacional. Aunque el valor absoluto es pequeño, es decir, el desempleo influye de manera muy ligera en la decisión de emigrar. Este aspecto hace pensar que quizás el desempleo no sea el único condicionante de la emigración.

−Los habitantes de provincias con elevado desempleo tienen una mayor probabilidad de emigrar que los habitantes de provincias con bajo nivel de desempleo. Esta relación, aunque significativa estadísticamente, no es muy consistente.

−Los salarios reales tienen, a nivel nacional, un impacto prácticamente nulo sobre la probabilidad de emigrar. Esto es debido seguramente a la escasa diferencia por provincias en cuanto al nivel de sueldos, y en caso de haberlos, es evidente que también habrá diferencias en cuanto al nivel de vida, en el sentido de que los precios y el coste de la vida en las regiones con mejores sueldos serán habitualmente más caras, por lo cual un aspecto contrarrestaría al otro. Por otro lado, los subsidios por desempleo y el Plan de Empleo Rural, tiene un efecto desincentivador, respecto de la movilidad interprovincial de la fuerza de trabajo.

En esta aplicación pretendo encontrar la justificación del descenso acusado del nivel de población en la provincia de León, a través del impacto causado en concreto por el aumento del desempleo, sin tener en cuenta otros posibles factores.

EXPOSICIÓN

Hay tres pasos considerados fundamentales para la medición de cualquier teoría económica:

1− Verificar con datos si el signo de la pendiente es negativo,

2− establecer un valor de la misma,

3− y fundamentar su constancia en el tiempo.

La ecuación de regresión que he considerado más adecuada para explicar la aplicación, es la siguiente:

$Y = a + b\,X$

Es una ecuación de regresión simple, y las variables son:

- Y, la variable explicada, que representa en este caso la población total de la provincia de León.
- X, la variable explicativa, que en este caso se identifica con el número de desempleados de la misma provincia.

Es claro por tanto que, si este modelo lo que pretende es explicar un descenso en la población total de la provincia basándose en el nivel del desempleo, entonces el signo de la pendiente (signo de b) debería ser negativo. La teoría a demostrar sería en este caso que, al aumentar el número de desempleados (X), entonces el nivel de población (Y), debería disminuir, por causa de un aumento de la emigración. Ya que lo que pretendo explicar es que el aumento del desempleo se reflejará en la población, reduciéndola, debido a que la población emigrará hacia otras provincias con más posibilidades de encontrar empleo.

Con el resultado de un signo negativo para b, se conseguiría verificar el punto número 1 exigido para la medición de cualquier teoría económica. También hallaremos un valor concreto para b y verificaremos su constancia en el tiempo, para completar los otros dos puntos requeridos.

Hay otras posibles causas de la disminución de la población aparte de la emigración por la alta tasa del desempleo, como pueden ser una baja tasa de natalidad, unos salarios bajos o los costes adyacentes de la

emigración, aspecto que ya indique en la introducción, pero he decidido no incluirlos en el modelo por considerarse en estos momentos que el factor más importante de la emigración en nuestra provincia es el paro.

LOS DATOS

Respecto a los datos usados para los cálculos de esta aplicación cabe hacer algunas aclaraciones. Los datos de la población se refieren a la población total, ya que considero que la población activa no sería un buen indicador de la emigración, pues el emigrante normalmente no emigra solo, sino que suele arrastrar a su entorno familiar directo con él. Y aquí lo que se trata de justificar es una disminución de la población total, no de la población activa. Y el uso de estos datos de población total podría compensar desde mi punto de vista la no utilización de más variables en el modelo, es decir, que el resultado final no se encontrará tan lejos de la realidad. Los datos que figuran en la tabla de datos simples tienen como fuente el INE, no han sufrido ninguna transformación, son simplemente la serie original. Los datos aparecen en la segunda columna de la tabla de datos simples bajo el título de: población.

En el caso del desempleo, he usado la media anual de los parados, ya que aunque los datos de que disponía eran trimestrales, me parecía más adecuado usar una media del año para evitar que los datos pudiesen estar influenciados por una época del año determinada. De todos es conocido que existen profesiones estacionales debidas por ejemplo a los períodos estivales o de cosechas, que no deben condicionar en ningun caso el número de parados que existe realmente en dicho año. En este caso la fuente de los datos también es el INE, que a través de la encuesta de población activa, facilita los siguientes resultados:

Las medias anuales se han calculado de manera simple, como medias aritméticas sencillas, de acuerdo con la fórmula: X = ð (xi ððn.

Los datos aparecen en la última columna de la tabla de datos simples bajo el epígrafe: desempleados.

Los datos de la siguiente tabla, denominada tabla de datos simples, son por tanto los referentes a la población y al número medio anual de desempleados de la provincia de León, en el período de 1.977 a 1.999, procedentes del Instituto Nacional de Estadística (INE).

Si realizamos un análisis a priori de la evolución de la población total, teniendo en cuenta los datos de la tabla anterior, nos encontramos con el siguiente gráfico:

Como puede apreciarse en el gráfico, la evolución del nivel de la población en la provincia de León es desalentadora.

Se refleja en él una profunda caída del nivel de la población en los años 1.977 al 1.981 y a continuación una recuperación de 5 años, aunque sin llegar a recuperar niveles como los de años anteriores al 77, en el cual la población era mucho mayor. Lo más negativo se encuentra en los últimos l3 años, donde se está produciendo un descenso progresivo pero continuado de la población. Estos son los datos que seguramente han alarmado tanto a la población y han hecho del tema poblacional la conversación diaria de la provincia.

Predicciones del Instituto Nacional de Estadística apuntan a que esta tendencia va a continuar igual que estos últimos años, por lo cual se acrecienta la preocupación sobre este tema. En concreto, la predicción para el año 2.005 es de 499.545 habitantes, lo que supone una pérdida de 11.065 habitantes en 5 años, dato que resulta ilustrativo del pesimismo reinante en este tema.

CALCULOS

Se pretende en este apartado encontrar solución a los tres pasos fundamentales antes citados de verificar el signo de la pendiente, que en este caso hemos considerado que debería ser negativa, encontrar un valor para la

misma y por último verificar su constancia en el tiempo.

El modelo usado en este análisis es el de MCO (Método de los Mínimos Cuadrados Ordinarios). Que se basa en encontrar una recta de regresión que minimice las discrepancias respecto a una nube de puntos que representan los datos del modelo. Las discrepancias se miden por las distancias verticales de los valores observados con respecto a la línea teórica. Diferentes líneas (determinadas por diferentes valores dados para los parámetros a y b) darán lugar a diferentes sumas de discrepancias. El criterio MCO elige la línea (y por lo tanto los valores de los parámetros), que minimizan la suma de dichas discrepancias. Elegir la línea del mejor ajuste equivale a elegir los valores de los parámetros que la determinan, o sea, dar unos valores concretos a los parámetros a y b.

Estos parámetros y esta línea de la que hablamos es la que formulamos anteriormente: Y = a + b X. Aquí Y era la población total y X el nivel medio de desempleo anual, los parámetros son a y b, que son los valores que dan forma a la recta de regresión, son desconocidos en este momento y por lo tanto el objeto de la búsqueda de esta aplicación.

El valor de b es el más importante ya que representa la pendiente de la recta de regresión mientras que el valor de a sirve para mejorar el ajuste de la regresión y representa el valor que tendría Y en el caso de que X fuese 0.

En la siguiente tabla se muestra un análisis de la dispersión a priori de los pares de valores mostrados en la tabla de datos simples.

En esta representación de los valores de la dispersión, tomando como datos los de la tabla de datos simples, con pares de valores (x , y), se vaticina a priori una pendiente negativa para la recta de regresión. Esta representación es atemporal, ya que no tiene en cuenta el paso del tiempo de la serie histórica que se toma como base para el cálculo de la regresión, sino que solo se toman los pares de valores de las variables.

En cuanto al modelo especificado en la recta Y = a + b X, hay que establecer a priori un serie de hipótesis para poder usar el método MCO. Una vez hallados los resultados del modelo, procederemos a contrastar dichas hipótesis y a verificar que son verdaderas. Por el momento nos limitaremos a enumerarlas, pasando más adelante (una vez hechos los cálculos del modelo), a verificar si se cumplen.

En la fundamentación metodológica, se justifica la naturaleza aleatoria de la variable endógena, Y, objeto de explicación de la teoría, estableciendo su dependencia respecto a la perturbación aleatoria V.

El modelo de regresión simple: Yt = ð + ð Xt + Vt, se denomina modelo estadístico. El objetivo del modelo estadístico es describir el proceso de muestreo por el cual han sido generados los valores observados de Y, dados los valores de X.

Vt representa el efecto de las demás causas inobservables distintas de la representada por X, es decir, representa las demás causas que no han sido incluidas en el modelo pero que son englobadas en el ceteris paribus. Vt es la variable aleatoria o estocástica, representa las discrepancias o valores residuales entre los valores observados de Yt y los estimados Yt por la recta de regresión a + b Xt . La introducción de V implica explicitar dos tipos de causas, una económica, supuestamente conocida, X y otra aleatoria V.

Las hipótesis fundamentales que permiten especificar la distribución de las V son:

a) <u>hipótesis I ó de esperanza nula</u>: E(Vt) = 0.

Esta hipótesis implicaría que las V no guardan relación con las X, siendo por lo tanto variables totalmente independientes.

b) hipótesis II ó de homocedasticidad: E (Vt 2) = 2

En este caso indica que los valores de Vt tienden a distribuirse en torno a la recta de regresión con dispersión constante.

c) hipótesis III ó de no autocorrelación: Cov(Vt, Vt´) = 0

Implica esta hipótesis que los valores sucesivos de V tienen correlación nula. Es la introducción formal de la aleatoriedad.

d) hipótesis IV ó de normalidad: Vt v (0, 2).

Por último ya, esta hipótesis implica que las perturbaciones V siguen una distribución normal.

Si estas hipótesis no se cumpliesen, deberían usarse otros métodos de cálculo, como el de los mínimos cuadrados generalizados (MCG) o el de los mínimos cuadrados ponderados (MCP). Estos se realizan haciendo transformaciones en las variables originales para solucionar los problemas de no cumplimiento de las hipótesis. Y luego se aplicaría MCO, pero en este caso en principio no es necesario y comenzaremos usando el método MCO.

Una vez enumeradas las hipótesis puede pasarse a la estimación de los parámetros del modelo. En este caso, como usamos el método de los mínimos cuadrados ordinarios, y se trata de una regresión simple los parámetros resultan de la aplicación de las siguientes fórmulas:

$a = Y - b\,X$

$b = \text{cov}(x,y) / \text{var}(x)$

Usando estas fórmulas estadísticas y otras fundamentales, hallamos la tabla de los parámetros estimados para la regresión.

Para comenzar el análisis, y una vez hallado el valor de los coeficientes, resaltar que la ecuación de la regresión resultaría:

$Y = 539.711{,}086 - 0{,}52180\,X$

El valor calculado para el parámetro a es un valor lógico para este parámetro, pues esa cantidad podría ser perfectamente la de la población en circunstancias en las que no influyese el desempleo (porque su valor fuese 0), sino otros factores.

El valor de b, como ya había predicho la teoría, es negativo, puesto que al aumentar el desempleo se supone que habrá de disminuir la población. Queda por tanto verificado el signo de la teoría y hallado el valor cuantitativo de los parámetros.

El valor del coeficiente de determinación puede considerarse un poco bajo, ya que implica que el desempleo solo explica un 43 % de la disminución de la población.

INFERENCIAS ESTADÍSTICAS

Una vez especificado el valor de los parámetros y el signo de b, puede pasarse a confirmar la significatividad de dichos parámetros. Esto consiste en la realización de inferencias estadísticas con los parámetros estructurales hallados.

− Contraste individual:

Para ello nos servimos del análisis del estadístico t de Student, que permite realizar un contraste individual de la significatividad de cada parámetro, primero lo realizaremos para el parámetro b y a continuación para el a.

1) En el caso de b

La hipótesis nula viene dada por:

H(0); ð = 0.

y la hipótesis alternativa por:

H(1); ð ð 0.

La cuestión es dilucidar si se acepta o no la hipótesis nula. Si se aceptase resultaría que ð=0, con lo cual la variable explicada sería igual al término independiente y X no tendría ninguna relación de influencia sobre Y. El caso alternativo en el cual ðð0, si se daría esa relación de dependencia, es decir, se confirma que la variable que representa el parámetro b es significativa.

El valor de la t de Student, ya calculado en la tabla, viene dado por la ecuación: (b−ð)/D(b). Que puede simplificarse a efectos de cálculo, puesto que en la hipótesis nula = 0, tenemos entonces que la t de Student se calcula mediante la fórmula b / D(b). El numerador de esta expresión es el valor calculado para el parámetro, y el denominador, la desviación típica de dicho parámetro. Que a su vez se calcularía mediante la fórmula D2(b) = /(nS2x1), donde el numerador es el estimador de la varianza de las discrepancias, y el denominador n que multiplica a la varianza de la variable a la cual se refiere el parámetro. Una vez calculado ese valor de la varianza, habría que realizarle la raiz cuadrada positiva para hallar el valor de la desviación típica. En este caso t = − 3,97, para el parámetro b, y como solo se tiene en cuenta el valor absoluto sería t = 3,97.

La condición a considerar es que ese valor calculado para la t, a de ser superior al valor correspondiente en la tabla, para los grados de libertad que halla que considerar y para la probabilidad de 0,05, ya que se requiere una significatividad del 95%. Para lo cual 1−ð = 0,95 y por lo tanto ð=0,05.

De esta forma la fórmula a tener en cuenta sería:

Pr(−A < b−ð / D(b)< A) = 95%

Al comprobar en una tabla de t de Student, para n−2 = 23−2 = 21 grados de libertad y en la probabilidad de 0,05, se encuentra que el valor de la t = 1,721 < 3,97, por lo cual puede rechazarse la hipótesis nula y considerar que el parámetro b es significativo. Ya que el valor calculado para el parámetro cae en la región critica, en la cual es rechazada la hipótesis nula.

2) En el caso de a.

Para el parámetro a se seguiría el mismo proceso de verificación. Proceso ya llevado a cabo en la tabla de parámetros estimados, donde puede verse que la probabilidad > t (en valor absoluto) es inferior a 0,05, por lo cual también se aceptaría la significatividad de dicho parámetro.

−Contraste global:

La causa más importante para no haber rechazado el modelo es el buen valor de la F de Snedekor, que se usa

para la contrastación de hipótesis conjuntas de b y a. La fórmula, en una ecuación de regresión simple, para su cálculo es: (bððx²t) / D(n−2), que representa el cociente entre la variación explicada y la no explicada, corregidas por sus respectivos grados de libertad. Este estadístico permite determinar si el modelo en conjunto es adecuado.

La hipótesis nula viene dada por:

H(0); ð = ð = 0

y la hipótesis alternativa por:

H(1); ð " ð " 0

El valor calculado para nuestra F es 15,77, lo bastante alejado de 1 para rechazar la hipótesis y apoyar la conclusión ya adelantada por el análisis de la t de student, que aseguraba la significatividad del modelo. Comprobándolo en una tabla de distribución de la F de Snedecor, para una significatividad del 5%, y teniendo en cuenta los grados de libertad del numerador (1) y del denominador (n − 2 = 23 − 2 = 21), hallamos un valor F = 4,32 < 15,77. Por lo cual nuevamente comprobamos que el valor calculado es superior al tabulado y por lo tanto cae dentro de la región crítica y se rechaza la hipótesis nula.

Si el valor del estadístico fuese próximo a 1, significaría que la variación explicada es la misma que la puramente aleatoria, de manera que X no explicaría nada. Como muestra la tabla, la probabilidad dada 0,0007, nos permite aceptar el modelo para un nivel de confianza superior al 95%.

R^2, representa la bondad del ajuste a nivel muestral, la fórmula para su cálculo es

R^2= cov (x; y)/ (x)(y).

Oscilando su valor en todo caso entre 0 y 1. Supone un mayor ajuste cuanto más cercano sea su valor a 1. Considerándose un ajuste adecuado para $R^2 > 0,7$. Por lo tanto, en nuestro modelo el valor de R2, resulta un poco bajo R2 = 0,4 < 0,7. Lo cual indica un ajuste a nivel muestral un poco bajo, como ya indicamos en el apartado de cálculos. Esto quiere decir que la nube de puntos que representa los valores de X e Y, no se ajusta bien a la recta. Hay bastante dispersión en torno a la que sería la recta de regresión como puede verse en el gráfico de la dispersión ya mostrado en páginas anteriores de esta aplicación.

CONTRASTE DE HIPÓTESIS

Como ya se anunció en apartados anteriores, a continuación comprobaremos si se cumplen las hipótesis que se introdujeron a priori para poder usar el método MCO.

1− Hipótesis de no−autocorrelación

La autocorrelación implica que alguna de las causas incluidas en el ceteris paribus no a sido especificada, con lo cual aparecerá un componente sistemático en las V. Puede deberse a muchas circunstancias diferentes, como un tratamiento erróneo de los datos, la omisión de alguna variable explicativa, una forma funcional inadecuada, la presencia de variables retardadas, etc.

Para decidir si un modelo concreto presenta perturbaciones autocorrelacionadas, se examinan las discrepancias de la regresión, mediante el método gráfico, en un análisis atemporal se examina el ajuste de las discrepancias a la recta de regresión. Cuando no es visible gráficamente, se usan contrastes como el de Von Neuman y el de Durbin−Watson. La fórmula de cálculo para este último estadístico es :

D−W = d = "(dt − dt−1)2 / " dt2.

No hay autocorrelación de primer orden si el valor empírico del estadístico calculado se mantiene en torno a 2. Observando la tabla de estimación de los parámetros de la regresión, puede comprobarse que en este caso esa predicción no se cumple y el valor D−W = 0,2020, que está lo suficientemente alejado de 2 como para poder admitir la posibilidad de que el modelo presente problemas de autocorrelación.

La autocorrelación, como hemos visto, implica que alguna causa no ha sido especificada explícitamente. Por lo cual, al darse una especificación inadecuada del modelo, quizás resultaría conveniente especificar una nueva hipótesis. Por ello podría resultar necesario incluir nuevas variables explicativas en el modelo para paliar estos problemas. Variables de las que ya se ha hablado con anterioridad, como una tasa de natalidad, etc. Quizás estableciendo un modelo de regresión múltiple se paliarían estos problemas.

2− Hipótesis de normalidad:

El contraste de normalidad de la variable población, como puede verse en la siguiente tabla, no permite aceptar a priori que se comporte como una variable normal.

El coeficiente de asimetría, elaborado por Fisher tiene por fórmula de cálculo:

= / , donde el numerador representa el momento respecto a la media de orden 3 (= E[(−)3], y el denominador es la desviación típica elevada al cubo. El valor de dicho coeficiente de asimetría debería estar más cercano a 0 para que la distribución pudiese ser considerada como normal. En esta situación puede decirse que es positiva hacia la derecha al ser el valor del coeficiente superior a 0.

En el caso del coeficiente de curtosis, también debido a Fisher, y que mide el grado de achatamiento o apuntamiento de una distribución de probabilidad, su fórmula de cálculo es: = (/)−3, donde el numerador representa el momento respecto a la media de orden 4 y el denominador la desviación típica elevada a la potencia cuarta. Este es mayor que 0, por lo cual la distribución se considera como leptocurtica, o lo que es lo mismo, más apuntada de lo normal.

Y el análisis del estadístico Chi−cuadrado nos dice que la distribución tiene a su derecha un área de 0,0216.

Sin embargo, a pesar de los datos de los coeficientes de asimetría y de curtosis, con un nivel de significación del 1% si que podríamos aceptar la hipótesis de normalidad.

3−) Hipótesis de homocedasticidad.

La heterocedasticidad, o incumplimiento de la hipótesis de homocedasticidad, significa que, la variación de Y atribuible a V es mayor de lo que supone la especificación del modelo.

Uno de los contrastes más sencillos y de aplicación más general, es el contraste de la razón de verosimilitud, para el cual se calcula el estadístico:

= − 2 ln = −2 ("Ti ln i − T ln) ; = ((i) Ti / T

En muestras grandes el estadístico , se distribuye como una con n−1 grados de libertad, bajo la hipótesis de homocedasticidad. Cuanta más heterocedasticidad exista, mayor será el valor de este estadístico, y por ello el contraste apropiado es una contraste de una sola cola:

si < (p−1), aceptamos la hipótesis de homocedasticidad.

si > (p−1), rechazamos dicha hipótesis y aceptamos la hipótesis de heterocedasticidad.

Siendo el nivel crítico de la distribución , es decir el valor tal que:

p(>) = . Para calcular el estadístico , tenemos que dividir la muestra en p grupos de acuerdo con algún criterio que nos haga sospechar la presencia de heterocedasticidad.

En nuestra tabla de datos de la regresión obtenemos un valor grande del estadístico de la razón de verosimilitud, por lo que cabe sospechar la presencia de heterocedasticidad.

CONCLUSIÓN

Como conclusión, resaltar que aunque se considera el desempleo como la principal variable causa del decremento de la población en la provincia de León, el análisis demuestra que dicha causa no explica completamente, ni mucho menos, el descenso de la población. Tan solo lo hace en un 43%. Habiendo por tanto otras múltiples causas que podrían considerarse en el modelo como causantes de dicho decremento.

Esta era una cuestión que ya se sospechaba a priori, pues la despoblación, es un fenómeno que evidentemente no afecta solamente a la provincia de León. Es cuestión de sobra conocida que en los países desarrollados se tiende hacia un nivel decreciente del nivel de la población. El envejecimiento de la población es un fenómeno preocupante en todo el mundo, sobre todo en lo relativo a la financiación del sistema de pensiones. Por estás razones era evidente que un análisis de la influencia del desempleo en la disminución de la población no iba a mostrar una influencia unitaria de esta variable.

Además el modelo no da esperanzas de que los resultados puedan ser muy fiables debido a la posible presencia de heterocedasticidad , autocorrelación, y el no cumplimiento además de la hipótesis de normalidad. Esto puede ser debido a un tratamiento equivocado de los datos de las series originales o a un mal ajuste por la falta de más variables explicativas. Sea cual sea el motivo, es evidente que el resultado de la regresión podría haberse mejorado.

PREDICCIÓN

En la predicción para el año 1976, se conocen los valores de las variables endógenas y exógenas para el período de predicción. La comparación con los valores obtenidos es lo que permite evaluar la capacidad predictiva del modelo.

El valor de X1976 es por lo tanto conocido, debido a que, evidentemente es un dato del pasado. Realizando los cálculos necesarios se llegaría a:

Y 1976 = a + b X 1976

X 1976 = 4.560

Y 1976 = 539.711,086 + 4520 (−0,521806)

Y 1976 = 537.352,52

Predicción que resulta por defecto pues el verdadero valor de Y para 1976 fue de 542.130. Luego el error es de 4.777 habitantes, un 0,8% de la población real.

Considerando que una de las principales fuentes de error de las predicciones es una posible especificación inadecuada del modelo y teniendo en cuenta que en el caso del modelo expuesto hemos supuesto que podrían

faltar variables que serían interesantes a la hora de explicar Y, puede justificarse este pequeño error en la predicción.

16