

Índices de pobreza en Galicia

María José Ginzo Villamayor¹

¹Departamento de Estadística e Investigación Operativa (Universidad de Santiago de Compostela)

RESUMEN

Se analiza el perfil de la pobreza de la población gallega. Se estiman índices de la familia de Foster, Greer y Thorbecke que permiten estudiar la incidencia, intensidad y desigualdad de la pobreza y se determinan los intervalos de confianza para los distintos indicadores utilizando técnicas de inferencia estadística clásica basadas en distribuciones asintóticas y metodología Bootstrap.

Palabras y frases clave: Bootstrap, escala de equivalencia, umbral de riesgo de pobreza, Foster, Greer y Thorbecke.

1. INTRODUCCIÓN

En este estudio se analiza el perfil de la pobreza de la población gallega. Se han utilizado los microdatos de la “Encuesta de Condiciones de Vida” (ECV) del Instituto Gallego de Estadística (IGE), año 2006. La ECV es una actividad estadística anual que realiza el IGE desde 1999. Es una encuesta dirigida a los hogares gallegos cuyo objetivo es obtener información sobre sus características socioeconómicas. Se estiman índices de la familia **Foster, Greer y Thorbecke**, (véase en Foster, Greer and Thorbecke (1984)). Para analizar la pobreza, se utilizó la inferencia estadística para estimar los niveles de pobreza de cada provincia y Galicia. Concretamente, se determinaron los intervalos de confianza para distintos indicadores usando inferencia clásica basada en distribuciones asintóticas y metodología Bootstrap. Se obtiene un mapa de la distribución provincial de la pobreza en Galicia y compara la precisión de las técnicas.

Los resultados se muestran tanto para el nivel de pobreza del 60% como del 50% de la mediana de ingresos equivalentes y se observará que Ourense es la provincia gallega con mayor nivel de pobreza, en cambio, con menor nivel la provincia de Pontevedra (al 60%) y la provincia de A Coruña (al 50%).

2. PRELIMINARES

Para medir la pobreza hay que tomar una serie de decisiones previas que son: la elección del indicador monetario de bienestar, la determinación de la línea de pobreza y la elección del objeto de estudio (hogar o individuo).

En cuanto al indicador monetario de bienestar, se pueden usar los ingresos o los gastos. La decisión de cuál utilizar dependerá de las características particulares de la fuente de datos utilizada en cada estudio concreto. En este trabajo, dado que la fuente utilizada es la ECV del IGE, se considerará como indicador el ingreso total, ya que es la variable, que se recoge con un mayor grado de detalle.

En el caso de las líneas de pobreza relativas, para determinar el umbral de pobreza se suele tomar una medida de posición central de los ingresos o gastos equivalentes. Es habitual usar el 60% de la mediana.

Se define el **umbral de riesgo de pobreza** que se corresponde con el 60%-50%, (en este trabajo, se han realizado ambos análisis), de la mediana de los ingresos equivalentes de todas las personas, (Eurostat usa el 60%). La **tasa de riesgo de pobreza** se define como el porcentaje de personas con un ingreso equivalente inferior al umbral del riesgo de pobreza.

Una vez determinado que se usarán los ingresos y no los gastos para el estudio, hay que decidir, también, si se medirá la pobreza a nivel de hogar o individuo. En los microdatos de la encuesta utilizada en este trabajo, se dispone de la información del hogar y también a nivel de individuo. Se ha optado por usar la de los individuos, es decir, sus ingresos equivalentes. Para comparar ingresos hay que usar las economías de escalas que se producen dentro del hogar, es decir, ver cómo se distribuyen los recursos entre los distintos miembros. Aparecen las escalas de equivalencia que tratan de llevar a cabo una normalización que permita el análisis comparativo de hogares con distintas composiciones. El **ingreso equivalente** corrige el efecto del número de personas en el volumen de ingresos del hogar, haciendo comparables los ingresos de hogares con distinto número de miembros. Se calculan las unidades de consumo (U. C.) del hogar empleando la escala de equivalencia de la OCDE modificada, $[1 + 0.5(a - 1) + 0.3(b)]$. En la expresión anterior a es el número de adultos (≥ 14 años) y b el número de niños (< 14 años) del hogar. Notar que debido a la forma de los microdatos, se ha tomado $a > 15$ y $b \leq 14$ años. Se define **ingreso equivalente del hogar** como el ingreso medio mensual del hogar dividido por las unidades de consumo. Como **ingreso equivalente de cada persona** se toma el equivalente del hogar al cual pertenecen. Se entenderá por **hogar pobre** aquél en el que viven personas pobres. Una persona se considera pobre atendiendo a un umbral.

3. METODOLOGÍA

Una vez decidido el escenario previo, hay que seleccionar el índice para evaluar la pobreza. Para ello, en este trabajo, se ha empleado la familia de índices de Foster, Greer y Thorbecke, que sirven para evaluar las distintas dimensiones de la pobreza: **incidencia, intensidad y desigualdad**. Se han seleccionado por la necesidad de abarcar todas las dimensiones de la pobreza y por sus propiedades. Una de las principales propiedades es que son aditivamente descomponibles, lo que permite determinar la aportación de cada provincia al índice general, así como facilitar la determinación

asintótica de su distribución en la aplicación de inferencia estadística. Se definen:

$$FGT_{\alpha}(y, z) = \frac{1}{n} \sum_{i=1}^q \left[\frac{z - y_i}{z} \right]^{\alpha-1}, \alpha > 0,$$

donde según el valor de α (parámetro de aversión a la desigualdad) se obtienen distintos índices. En la expresión de FGT_{α} , q es el número de pobres, n el total de elementos, y_i es la renta del individuo i -ésimo y z es el umbral de pobreza. A mayores valores de α más importancia relativa se le da a los desniveles de pobreza relativa mayores. Para $\alpha = 1$, se tiene la proporción de pobres y para $\alpha = 2$ el índice es igual al producto de la proporción de pobres por el desnivel de pobreza. Es decir:

$$FGT_{\alpha}(y, z) = \frac{1}{n} \sum_{i=1}^q \left[\frac{z - y_i}{z} \right]^1 = H \cdot \frac{1}{n} \sum_{i=1}^q \left[\frac{z - y_i}{z} \right] = H \cdot I$$

$$\text{donde } \frac{q}{n} = H, I = \frac{1}{n} \sum_{i=1}^q \left[\frac{z - y_i}{z} \right], I \text{ es el nivel de pobreza.}$$

Por otra parte, Atkinson (1987) ofrece las condiciones bajo las cuales todos los miembros de la familia FGT_{α} dan la misma ordenación de las distribuciones de renta.

Estos índices de pobreza serán calculados a partir de los datos de la ECV. A continuación se describen las dos técnicas de inferencia utilizadas.

1. Inferencia Clásica. Se han usado las distribuciones asymptóticas de los estimadores de las medidas de pobreza para el análisis inferencial, (Bishop, J.A., Chow, K.V. y Zheng, B. (1995) y Rongve (1997)). Se determinarán las distribuciones asymptóticas para cualquier índice de la familia FGT_{α} . Los índices FGT_{α} son aditivamente descomponibles, de la forma:

$$G = \int_{\alpha}^z g(z, y) f(y) dy$$

donde z es el umbral de pobreza, $f(y)$ es la función de densidad de los ingresos, $g(z, y) = \left(\frac{z-y}{z}\right)^{\alpha-1}$ función de la variable y y del umbral. Por tanto, si se define G como una medida de pobreza, \hat{G} será el estimador de esta medida para una muestra $Y_n = \{y_1, y_2, \dots, y_n\}$ de ingresos equivalentes. Por el Teorema Central del Límite, se verifica:

$$\sqrt{n}(\hat{G} - G) \rightarrow N(0, \sigma^2(G))$$

Si además $\sigma^2(\hat{G})$ es un estimador consistente de $\sigma^2(G)$, se obtiene un estadístico t que permite realizar análisis inferenciales para la medida de pobreza G , i.e.,

$$t = \frac{\hat{G} - G}{SE(\hat{G})} \rightarrow N(0, 1), \text{ donde } SE(\hat{G}) = \frac{\sigma^2(\hat{G})}{\sqrt{n}}$$

Mediante distribuciones asintóticas para cualquier índice, la estimación muestral será:

$$\begin{aligned}\hat{G} &= \hat{G}(y, z) = \frac{1}{n} \sum_{i=1}^n g(y_i, z) = \frac{1}{n} \sum_{i=1}^q g(y_i, z) + \frac{1}{n} \sum_{i=q+1}^n g(y_i, z) \\ &= \frac{1}{n} \sum_{i=1}^q g(y_i, z) = \frac{1}{n} \sum_{i=1}^q \left[\frac{z - y_i}{z} \right]^{\alpha-1}, \alpha > 0\end{aligned}$$

$g(y_i, z) = 0$ si $i > q$, ya que al estar ordenados los ingresos, cualquier ingreso y_i es mayor que la línea de pobreza fijada. Se tiene entonces para los estadísticos más notables lo siguiente:

$$\begin{aligned}E(\hat{G}) &= E\left(\frac{1}{n} \sum_{i=1}^n g(y_i, z)\right) = \frac{1}{n} n E(g(y_i, z)) = E(g(y_i, z)) = G, \text{ i.e. es insesgado.} \\ \text{Var}(g(y_i, z)) &= E[g(y_i, z) - G]^2 = \left[\int_0^z g^2(y, z) f(y) dy \right] - G^2 \\ \hat{\text{Var}}(g(y_i, z)) &= \left[\frac{1}{n} \sum_{i=1}^q g^2(y_i, z) \right] - \hat{G}^2\end{aligned}$$

$\sqrt{n}(\hat{G} - G) \rightarrow N(0, \text{Var}(g(y_i, z)))$ por el TCL, por la forma de G y a partir de esta distribución asintótica se determina la distribución del estadístico t que permitirá realizar análisis inferenciales para los índices G :

$$t = \frac{\hat{G} - G}{SE(\hat{G})} = \frac{\sqrt{n}(\hat{G} - G)}{\sqrt{\hat{\text{Var}}(g(y_i, z))}} \rightarrow N(0, 1)$$

Sustituyendo se obtiene:

$$\hat{\text{Var}}(g) = F\hat{G}T_{2(\alpha-1)+1}^2 - F\hat{G}T_{\alpha}^2$$

Así, un intervalo de confianza de nivel $(1 - \alpha)$, para $F\hat{G}T_{\alpha}$ será:

$$I = \left[F\hat{G}T_{\alpha} \mp \frac{1.96 \sqrt{F\hat{G}T_{2(\alpha-1)+1}^2 - F\hat{G}T_{\alpha}^2}}{\sqrt{n}} \right]$$

2. Metodología Bootstrap. El método Bootstrap (Efron y Tibshirani 1994) es un método de remuestreo desde los propios datos, a partir del cual se pueden obtener medidas de precisión sobre las estimaciones estadísticas realizadas, o intervalos de confianza para los estadísticos objeto de estudio. En este estudio de simulación, para cada remuestra Bootstrap se obtienen los índices de pobreza para $\alpha = 1, 2, 3$, para Galicia y provincias. El estadístico de interés será la media de cada índice. La estimación del estadístico Bootstrap se realiza de modo no paramétrico, ya que no se establece ningún supuesto sobre la distribución de los datos. Para la comparación de

dichas estimaciones con la inferencia clásica, se analizan los intervalos de confianza mediante los métodos percentil, percentil-t y simetrizado. Dada una muestra $Y_n = \{y_1, y_2, \dots, y_n\}$ de ingresos equivalentes de F desconocida, el objetivo es construir un intervalo de confianza de nivel $(1 - \alpha)$ para $G = G(F)$ a partir de un estimador $\hat{G} = G(Y)$. Se usan las estimaciones obtenidas para la varianza, es decir,

$$\begin{aligned} \hat{Var}(g) &= F\hat{GT}_{2(\alpha-1)+1}^2 - F\hat{GT}_\alpha^2 \\ \text{Método percentil. } I &= \left[\hat{G} - n^{-\frac{1}{2}}y_{1-\frac{\alpha}{2}}, \hat{G} - n^{-\frac{1}{2}}y_{\frac{\alpha}{2}} \right]. \\ \text{Método percentil-t. } I &= \left[\hat{G} - \hat{Var}(g)n^{-\frac{1}{2}}y_{1-\frac{\alpha}{2}}, \hat{G} - \hat{Var}(g)n^{-\frac{1}{2}}y_{\frac{\alpha}{2}} \right]. \\ \text{Método simetrizado. } I &= \left[\hat{G} - \hat{Var}(g)n^{-\frac{1}{2}}y_{1-\alpha}, \hat{G} + \hat{Var}(g)n^{-\frac{1}{2}}y_{1-\alpha} \right]. \end{aligned}$$

4. RESULTADOS DE SIMULACIÓN Y CONCLUSIONES

Para realizar el análisis empírico se utiliza el fichero de microdatos de la “Encuesta de Condiciones de Vida” para el año 2006, con una muestra de 6.337 hogares (17.619 personas). En este estudio se ha optado por tomar como indicador de bienestar el ingreso equivalente de las personas. El umbral de pobreza considerado es el 60% y 50% de la mediana de la distribución del ingreso equivalente de las personas. Se calcularon los índices para Galicia y provincias, seguidamente se comparan los resultados al aplicar la metodología clásica y la Bootstrap y los intervalos de confianza para cada índice. Para determinar el perfil de la pobreza en Galicia se estudia el valor de los índices en las distintas provincias. Para no hacer el documento muy extenso, no se presentan tablas de resultados, sólo los resultados y conclusiones más generales.

Para el índice FGT_1 , que indica la proporción de pobres, se obtiene que el 13.85% (para el 60%) o el 7.24% (para el 50%) de las personas están por debajo del umbral de pobreza. Su distribución a lo largo del territorio gallego no es homogénea. Estos resultados se pueden observar en la Tabla 1, (del mismo modo se han calculado para el 50% del umbral de pobreza). El índice FGT_2 , que además de la incidencia de la pobreza, tiene en cuenta la intensidad de la misma, proporciona resultados similares. Al considerar FGT_3 , que tiene en cuenta, además de la intensidad y la incidencia, la desigualdad entre los pobres nos encontramos con una ordenación similar a los índices anteriores. La inferencia clásica para estos índices, cuando se dispone de muestras pequeñas, puede presentar problemas a la hora de asumir la convergencia del estadístico a la distribución normal.

Tabla 1: Porcentaje de personas por debajo del umbral de pobreza al 60%. (Ingreso mediano equivalente: 929.64, umbral: 557.78)

Galicia	A Coruña	Lugo	Ourense	Pontevedra
13.85%	13.20%	13.70%	18.10%	12.95%

Se realiza un estudio de simulación Bootstrap (realizándose 1000 réplicas de las muestras) para comparar los resultados obtenidos con la inferencia clásica. La utilización de Bootstrap supone una ventaja al no establecer supuestos a priori sobre la función de distribución. Los resultados tras aplicar Bootstrap son muy similares entre Galicia y las provincias para la incidencia de pobreza. En el caso de la brecha y severidad de pobreza los resultados son similares en cuanto al orden pero su valor no es tan próximo con en el caso del FGT₁.

Se calcularon los intervalos de confianza con ambas metodologías, con el propósito de determinar cuál de los métodos los proporciona de menor amplitud para un mismo nivel de confianza. Respecto a la amplitud de los intervalos de confianza, a nivel de provincia, con el umbral al 60%, se observa que no hay una reducción con la metodología Bootstrap, lo que implicaría una mejora en la precisión de las estimaciones. Para algún indicador se observa una reducción en la amplitud del intervalo de confianza con Bootstrap en el caso de Galicia. Con el método no paramétrico nunca se obtienen resultados fuera del rango de posibles valores del indicador. Las distribuciones empíricas generadas por el remuestreo representan de forma más adecuada el comportamiento de los datos originales. Para Galicia y provincias el intervalo de confianza clásico es el de menor longitud para la *incidencia* y el de mayor se obtiene con el simetrizado. Cabe destacar que, en el caso de Lugo y Ourense para el percentil-t se obtienen intervalos de confianza amplios, en el caso de Lugo la longitud es 21.66 y para Ourense 11.79. En el caso del percentil-t se obtienen también intervalos de confianza amplios comparados con el clásico o el percentil. Para las provincias el intervalo de confianza de menor longitud para la *brecha* es el clásico mientras que en Galicia es el percentil. Para la *severidad* en el caso de Galicia el intervalo de confianza de menor longitud es el percentil, seguido del clásico, que es además para las provincias el que ofrece el intervalo de confianza más pequeño. En todos los casos el de mayor longitud es el simetrizado. Resultados análogos se obtienen con el umbral al 50%. Por tanto, aunque la metodología Bootstrap no presenta intervalos de confianza menos amplios de manera general, sí que ofrece resultados más adecuados que la inferencia clásica basada en distribuciones asintóticas.

Se han representado los estimadores tipo núcleo de la función de densidad para los índices generados por las muestras Bootstrap, para determinar en qué medida las distribuciones de los índices se aproximan a la campana de Gauss. Se observa que las distribuciones no presentan normalidad debido a la asimetría y/o apuntamiento, lo que justifica el uso del Bootstrap para la construcción de intervalos de confianza, en vez de los derivados de la metodología clásica basados en la convergencia a distribuciones normales, ya que cuando se dispone de muestras reducidas pueden alejarse de la realidad.

REFERENCIAS

- Atkinson, A. B.(1987) On the Measurement of Poverty, *Econometrica*, 55(4), 749-764.
Bishop, J.A., Chow, K.V. y Zheng, B. (1995) Statistical inference and decomposable poverty measures. *Bulletin of Economic Research*, 47(4), 329-340.
Efron, B. y Tibshirani, R. (1994) *An introduction to the Bootstrap*. Chapman & Hall.

Encuesta de Condiciones de Vida del Instituto Galego de Estadística.
Foster, J., Greer, J., y Thorbecke, E. (1984) A class of decomposable poverty measures, *Econometrica*, 52(3), 761-766.
Rongve, I. (1997) Statistical inference for poverty indices with fixed poverty lines. *Applied Economics*, 29, 387-392.