

## Proyecciones de tasas de actividad

### METODOLOGÍA

---

#### 1. Introducción

La *Encuesta de población activa (EPA)* define la tasa de actividad como el porcentaje de activos respecto a la población de 16 o más años, siendo la población activa el conjunto de personas de 16 o más años que suministran mano de obra para la producción de bienes y servicios o que están sin trabajo, disponibles para trabajar y buscando activamente un empleo. Así pues, la población activa representa los recursos de fuerza de trabajo de los que dispone una economía y la tasa de actividad mide la propensión de las personas en edad de trabajar a participar en el mercado laboral.

La proyección del comportamiento futuro de la población en relación con el mercado laboral resulta fundamental en la planificación de las políticas públicas, ya que es determinante en el crecimiento potencial de la economía, sustentabilidad de los sistemas públicos...

Si para el año  $t$ , sexo  $s$  ( $s=hombres,mujeres$ ) y grupo quinquenal de edad  $i$  ( $i=16-19,20-24...65-69,70+$ ) denotamos:

$P_{s,i}^t$ : población en el año  $t$  de sexo  $s$  y grupo quinquenal de edad  $i$

$TA_{s,i}^t$ : tasa de actividad en el año  $t$  de la población de sexo  $s$  y grupo quinquenal de edad  $i$

tenemos que para cualquier grupo poblacional  $G$  que sea agregación de grupos elementales, la tasa de actividad y la población activa en el año  $t$  vienen dadas por:

$$TA_G^t = \sum_{(s,i) \in G} TA_{s,i}^t \cdot \frac{P_{s,i}^t}{P_G^t}$$
$$PA_G^t = TA_G^t \cdot P_G^t = \sum_{(s,i) \in G} TA_{s,i}^t \cdot P_{s,i}^t$$

donde  $P_G^t$  denota la población en el año  $t$  del grupo poblacional  $G$ :

$$P_G^t = \sum_{(s,i) \in G} P_{s,i}^t$$

Por tanto, la proyección de la tasa de actividad y la población activa de cualquier grupo poblacional que se obtenga como agregación de grupos elementales se reduce a disponer de las proyecciones de población y a proyectar las tasas de actividad correspondientes a los grupos elementales.

Las *Proyecciones de población a corto plazo* nos permiten conocer la evolución futura del volumen y estructura de la población por sexo y grupo quinquenal de edad. Ahora bien, la población objeto de estudio en esas proyecciones de población es la población residente en Galicia a 1 de enero, tanto en

viviendas familiares como en establecimientos colectivos, pero la *EPA* va dirigida a la población residente en viviendas familiares y obtiene los resultados anuales como la media de los cuatro trimestres. Por tanto, tendremos que adaptar estas proyecciones de población a nuestro contexto para que hagan referencia a la media anual de población residente en viviendas familiares.

La proyección de las tasas de actividad por sexo y grupo quinquenal de edad la haremos a partir de las series históricas de tasas de actividad proporcionadas por la *EPA*. Entre los organismos nacionales e internacionales que elaboran proyecciones de tasas de actividad, existen cuatro enfoques diferentes para proyectar las tasas de actividad:

- Métodos de juicio. Un grupo de expertos emite un juicio sobre el futuro de las tasas de actividad en base a las series históricas, datos observados en países próximos cultural y estructuralmente pero más avanzados económicamente...
- Modelos de regresión. Se ajusta un modelo de regresión para las tasas de actividad en función de variables explicativas de carácter demográfico, socioeconómico e institucional. Una vez fijado el escenario futuro para las variables explicativas, se utiliza el modelo de regresión ajustado para proyectar las tasas de actividad.
- Modelos de extrapolación en el tiempo. Se ajusta en función del tiempo una curva sigmoide para las tasas de actividad, normalmente la curva logística, lo que permite extrapolar sus valores futuros. Las curvas sigmoides son muy útiles para modelar procesos en los que se espera alcanzar un estado estable a largo plazo (tasas de actividad, inflación, productividad...).
- Modelos basados en una visión generacional. La proyección de las tasas de actividad se realiza a partir de las probabilidades de entrada o salida del mercado laboral correspondientes a los últimos años, que se estiman considerando la información transversal como longitudinal (generación ficticia).

Las oficinas de estadística se decantan sobre todo por modelos de extrapolación en el tiempo o modelos basados en una visión generacional, ya que tanto los métodos de juicio como los modelos de regresión requieren establecer una opinión sobre el comportamiento futuro de las tasas de actividad o de las variables explicativas de los modelos de regresión, lo cual resulta complejo y subjetivo.

En el mercado laboral gallego se está produciendo un proceso de convergencia de las tasas de actividad femeninas con las masculinas, por lo que resulta especialmente adecuado el uso de modelos de extrapolación en el tiempo para proyectar las tasas de actividad, pues permiten extrapolar directamente las tasas de actividad masculinas y extrapolar indirectamente las tasas de actividad femeninas extrapolando los ratios entre las tasas de actividad femeninas y masculinas.

Al igual que acontece con las proyecciones de población, es importante tener en cuenta que las proyecciones de tasas de actividad constituyen una simulación estadística de su evolución futura en el caso de mantenerse las tendencias observadas en el pasado. Así pues, no pretenden adivinar el futuro, sino proporcionar una herramienta de apoyo para la toma de decisiones en base a una

simulación estadística del rumbo que tomaría la fuerza de trabajo bajo la hipótesis de que se mantuviesen las tendencias pasadas.

## 2. Objetivo

En esta actividad estadística se trata de obtener información sobre la evolución futura de las tasas de actividad y del volumen y estructura de la población activa por sexo y edad.

## 3. Ámbitos de investigación

**Ámbito poblacional:** La población objeto de estudio es la población residente en viviendas familiares.

**Ámbito territorial:** El ámbito territorial es el conjunto de la Comunidad Autónoma de Galicia.

**Ámbito temporal:** Con periodicidad bienal, se ofrecen resultados para cada año del período de proyección.

## 4. Fuentes de información

Las tasas de actividad en los grupos elementales para el período de observación las obtenemos de la *EPA* y para calcular la población residente en viviendas familiares de los grupos elementales en el período de proyección utilizamos las *Proyecciones de población a corto plazo*.

### 4.1. Encuesta de población activa

La *EPA* es una investigación por muestreo de periodicidad trimestral dirigida a la población residente en viviendas familiares con la finalidad de conocer las características de esa población en relación con el mercado laboral.

Desde sus inicios en 1964, sufrió numerosos cambios en su metodología y elaboración, entre los que cabe destacar:

- En 1987 se introdujo un nuevo cuestionario para adaptarse a las recomendaciones de la *Conferencia internacional de estadísticos del trabajo* de la Organización Internacional del Trabajo de 1982, incorporar las variables de estudio de la *Encuesta de fuerza de trabajo* de la Unión Europea e incorporar información sobre las novedades del mercado laboral español (incremento del trabajo a tiempo parcial, aumento de la contratación temporal...).
- En 2002 se pasó a utilizar la definición de parado establecida en el *Reglamento 1987/2000 de la Comisión de las Comunidades Europeas, del 7 de septiembre de 2000*.

El modelaje temporal de las tasas de actividad anuales por sexo y grupo quinquenal de edad requiere disponer de series homogéneas de tasas de actividad anuales, por lo que restringiéndonos a los resultados trimestrales que proporciona la *EPA* desde 1987 tenemos que resolver el problema que

supone el cambio en la definición de parado que se empezó a aplicar en 2002 y convertir la información trimestral en información anual.

Según la definición de parado que se venía utilizando hasta el cuarto trimestre de 2001 se clasificaban como paradas a personas que según la nueva definición pasan a ser inactivos, por lo que para disponer de series homogéneas desde el primer trimestre de 1987 hay que corregir el número de parados y en consecuencia de activos e inactivos anteriores al primer trimestre de 2002.

Para llevar a cabo este enlace en las series trimestrales de parados utilizamos el documento *Enlace de las series de paro 1976-2000 según la definición EPA-2002 (J. Trejo y L. Ortega. INE)*, en el que estimaron las series trimestrales de parados bajo la nueva definición mediante un modelo de regresión probit con variable respuesta binaria que indica si una persona es parada según ambas definiciones o si con la nueva definición deja de ser parada y variables explicativas el sexo, la edad y la interacción entre el sexo y la edad. Para ajustar este modelo de regresión probit utilizaron los microdatos de la *EPA* correspondientes al período 2001-2004 en el que se recogió la información necesaria para saber si una persona es parada tanto con la antigua como con la nueva definición. Esta es la razón por la que realmente solo necesitaron recalcular las series trimestrales de parados según la nueva definición hasta el cuarto trimestre de 2000.

Una vez llevado a cabo este enlace ya disponemos de series trimestrales homogéneas de ocupados, parados, activos, inactivos y población por sexo y grupo quinquenal de edad desde el primer trimestre de 1987. Así pues, podemos obtener las series homogéneas de tasas de actividad anuales desde 1987 dividiendo el número medio anual de activos entre la población media anual, entendiendo que las medias anuales de activos y población se calculan como la media aritmética de los cuatro trimestres.

## **4.2. Proyecciones de población a corto plazo**

Las proyecciones de población por sexo y grupo quinquenal de edad que ofrecen las *Proyecciones de población a corto plazo* se refieren a la población residente en Galicia a 1 de enero, tanto en viviendas familiares como en establecimientos colectivos.

Ahora bien, nosotros necesitamos proyecciones de población por sexo y grupo quinquenal de edad que aludan a la media anual de población residente en viviendas familiares y que sean coherentes con las estimaciones anuales de población que proporciona la *EPA* para el último año.

Para obtener las proyecciones de población que vamos utilizar partimos de las estimaciones anuales de hombres y mujeres entre los 16 y 64 años y 16 o más años que proporciona la *EPA* para el último año y proyectamos la media anual de población residente en viviendas familiares para esos grupos poblacionales utilizando las tasas de variación que presentan en las *Proyecciones de población a corto plazo* después de transformar la información a 1 de enero en información media anual. Una vez tenemos proyectadas las medias anuales de población residente en viviendas familiares para los grupos poblacionales considerados, las desagregamos por grupo quinquenal de edad usando la

estructura por grupo quinquenal de edad que presentan en las *Proyecciones de población a corto plazo* referidas a la media anual.

## 5. Proyecciones de tasas de actividad en los grupos elementales

A partir de las series homogéneas de tasas de actividad anuales desde 1987, proyectamos las tasas de actividad en los grupos elementales utilizando modelos de extrapolación logística. Para los grupos elementales masculinos proyectamos directamente las tasas de actividad mientras que para los grupos elementales femeninos proyectamos las tasas de actividad indirectamente, ya que no proyectamos las tasas de actividad femeninas sino los ratios entre las tasas de actividad femeninas y masculinas.

Como ya adelantamos, las proyecciones de las tasas de actividad en los grupos elementales nos permiten calcular como agregación ponderada la proyección de la tasa de actividad de cualquier grupo poblacional que se obtenga como unión de grupos elementales, siendo el coeficiente de ponderación de cada grupo elemental la proporción que representa en las proyecciones de población su población respecto a la población total del grupo poblacional de interés.

### 5.1. Proyecciones de tasas de actividad masculinas por grupo quinquenal de edad

Si queremos extrapolar en el tiempo las tasas de actividad masculinas para el grupo quinquenal de edad  $i$  ( $i \in \{16-19, 20-24, \dots, 65-69, 70+\}$ ) entre los valores asintóticos inferior  $AI$  y superior  $AS$  resulta interesante considerar la transformación logística:

$$TAT_{hombres,i}^t = \ln \left( \frac{TA_{hombres,i}^t - AI}{AS - TA_{hombres,i}^t} \right)$$

ya que si modelamos la transformación logística de las tasas de actividad como la adición de una tendencia temporal  $f(t)$  y un residuo  $\varepsilon$ :

$$TAT_{hombres,i}^t = f(t) + \varepsilon$$

tenemos que:

$$TA_{hombres,i}^t = \frac{AI + AS \cdot e^{TAT_{hombres,i}^t}}{1 + e^{TAT_{hombres,i}^t}} = \frac{AI + AS \cdot e^{f(t)+\varepsilon}}{1 + e^{f(t)+\varepsilon}}$$

Así pues, si a partir de las tasas de actividad masculinas para el grupo quinquenal de edad  $i$  en el período de observación obtenemos las correspondientes tasas de actividad transformadas logísticamente y ajustamos el modelo de tendencia temporal para la transformación logística de las tasas de actividad:

$$\widehat{TAT}_{hombres,i}^t = \hat{f}(t)$$

podemos extrapolar en el tiempo las tasas de actividad para cada año del período de proyección mediante:

$$\widehat{TA}_{hombres,i}^t = \frac{AI + AS \cdot e^{\widehat{TA}_{hombres,i}^t}}{1 + e^{\widehat{TA}_{hombres,i}^t}} = \frac{AI + AS \cdot e^{\hat{f}(t)}}{1 + e^{\hat{f}(t)}}$$

Debemos tener en cuenta que existen tantos modelos de extrapolación logística para las tasas de actividad masculinas para el grupo quinquenal de edad  $i$  como posibles valores asintóticos para las tasas de actividad y tendencias temporales para la transformación logística de las tasas de actividad.

El procedimiento considerado para obtener las proyecciones de tasas de actividad masculinas para el grupo quinquenal de edad  $i$  consiste en:

1. Para el período de observación calculamos las tasas de actividad transformadas logísticamente mediante tres transformaciones logísticas:

Transformación logística 1. Simplemente garantizamos que no se llegue a resultados absurdos, en el sentido de que las proyecciones de tasas de actividad siempre queden acotadas entre 0 y 1 por muy bruscas que sean las tendencias ascendentes o descendentes que se observaran en el pasado:

$$AI = 0$$

$$AS = 1$$

Transformación logística 2. Permitimos que las proyecciones de tasas de actividad queden por debajo del mínimo histórico o por encima del máximo histórico hasta una cantidad  $\mu$  que representa la volatilidad absoluta media de la serie histórica de tasas de actividad en períodos temporales de 15 años, es decir:

$$AI = \min\{TA_{hombres,i}^t, t \text{ en el período de observación}\} - \mu$$

$$AS = \max\{TA_{hombres,i}^t, t \text{ en el período de observación}\} + \mu$$

donde:

$$\mu = \frac{\sum_{t=t_{ini}}^{t_{fin}-15} |TA_{hombres,i}^{t+15} - TA_{hombres,i}^t|}{t_{fin} - t_{ini} - 14}$$

siendo  $t_{ini}$  y  $t_{fin}$  el primer y último año del período de observación.

Transformación logística 3. Nos aseguramos de que las proyecciones de tasas de actividad siempre queden acotadas entre el mínimo histórico y el máximo histórico:

$$AI = \min\{TA_{hombres,i}^t, t \text{ en el período de observación}\}$$

$$AS = \max\{TA_{hombres,i}^t, t \text{ en el período de observación}\}$$

2. Ajustamos por mínimos cuadrados ordinarios tres modelos de tendencia temporal para cada transformación logística de las tasas de actividad:

Tendencia temporal 1. Las tasas de actividad transformadas logísticamente son independientes del tiempo:

$$f(t) = a$$

Tendencia temporal 2. Las tasas de actividad transformadas logísticamente presentan una tendencia lineal en el tiempo:

$$f(t) = a + b \cdot t$$

Tendencia temporal 3. Las tasas de actividad transformadas logísticamente muestran una tendencia cuadrática en el tiempo:

$$f(t) = a + b \cdot t + c \cdot t^2$$

3. Utilizando las transformaciones inversas asociadas a las transformaciones logísticas obtenemos los valores ajustados de las tasas de actividad para el período de observación correspondientes a los nueve modelos de extrapolación logística ajustados y nos quedamos con aquel que presente un menor error absoluto medio entre la serie ajustada y la serie histórica en los últimos años del período de observación.
4. Mediante el modelo de extrapolación logística ajustado que seleccionamos, extrapolamos la tasa de actividad para cada año del período de proyección.

## 5.2. Proyecciones de tasas de actividad femeninas por grupo quinquenal de edad

La herramienta estadística utilizada para proyectar las tasas de actividad femeninas también son los modelos de extrapolación logística, pero en vez de emplearlos para proyectar directamente las tasas de actividad femeninas, los usamos para proyectar los ratios de actividad entre mujeres y hombres. A partir de las proyecciones de ratios de actividad y tasas de actividad masculinas podemos obtener las proyecciones de tasas de actividad femeninas.

El procedimiento utilizado para obtener las proyecciones de tasas de actividad femeninas para el grupo quinquenal de edad  $i$  ( $i \in \{16-19, 20-24, \dots, 65-69, 70+\}$ ) es:

1. Para el período de observación calculamos los ratios de actividad entre mujeres y hombres, es decir, los ratios entre las tasas de actividad femeninas y masculinas:

$$RA_i^t = \frac{TA_{mujeres,i}^t}{TA_{hombres,i}^t}$$

y las transformamos logísticamente mediante la transformación logística asociada a los valores asintóticos 0 y 1:

$$RAT_i^t = \ln\left(\frac{RA_i^t}{1 - RA_i^t}\right)$$

Solo consideramos la transformación logística asociada a los valores asintóticos 0 y 1, pues en las últimas décadas se está produciendo un proceso de convergencia de las tasas de actividad femeninas con las masculinas y si bien las ratios de actividad pueden ser superiores a 1 no parece razonable esperar que las tasas de actividad femeninas tiendan a estabilizarse en un nivel superior al correspondiente a las tasas de actividad masculinas.

2. Ajustamos por mínimos cuadrados ordinarios tres modelos de tendencia temporal para la transformación logística de las ratios de actividad (constante:  $f(t) = a$ , lineal:  $f(t) = a + b \cdot t$  y cuadrática:  $f(t) = a + b \cdot t + c \cdot t^2$ ).
3. Utilizando la transformación inversa asociada a la transformación logística obtenemos los valores ajustados de las ratios de actividad para el período de observación correspondientes a los tres modelos de extrapolación logística ajustados y nos quedamos con aquel que presente un menor error absoluto medio entre las series ajustada e histórica de las ratios de actividad en los últimos años del período de observación.
4. Mediante el modelo de extrapolación logística ajustado que seleccionamos, extrapolamos la ratio de actividad para cada año del período de proyección:

$$\widehat{RA}_i^t = \frac{e^{\hat{f}(t)}}{1 + e^{\hat{f}(t)}}$$

5. Para cada año del período de proyección obtenemos la proyección de la tasa de actividad como:

$$\widehat{TA}_{mujeres,i}^t = \widehat{RA}_i^t \cdot \widehat{TA}_{hombres,i}^t$$

## 6. Indicadores derivados de las proyecciones de población activa

A partir de las proyecciones de población activa podemos proyectar el comportamiento futuro de una serie de indicadores que permiten analizar el mercado laboral desde diferentes perspectivas (sustentabilidad, feminidad y juventud). Recordemos que la proyección de la población activa de cualquier grupo poblacional se obtiene multiplicando su tasa de actividad proyectada por su proyección de población.

### 6.1. Índices de sustentabilidad

El **índice de sustentabilidad de los mayores** definido como el cociente entre la población activa y la población inactiva de 65 o más años:

$$\frac{\text{Activos}}{\text{Inactivos de 65 o más años}}$$

representa el número de potenciales trabajadores por cada jubilado, entendiendo el término jubilado en un sentido amplio como toda persona inactiva de 65 o más años.

Ahora bien, la población inactiva no solo está formada por los jubilados sino también por los niños, estudiantes, incapacitados permanentes..., por lo que para evaluar la sustentabilidad global de los

sistemas públicos también resulta interesante el **índice de sustentabilidad global** definido como el cociente entre la población activa y la población inactiva:

$$\frac{\text{Activos}}{\text{Inactivos}}$$

## 6.2. Índice de feminidad

Una de las características del mercado laboral en las últimas décadas es el progresivo acercamiento de las tasas de actividad femeninas a las masculinas, por lo que resulta interesante disponer de un indicador de la estructura por sexo de la población activa.

El **índice de feminidad** definido como el cociente entre la población activa femenina y la población activa masculina:

$$\frac{\text{Mujeres activas}}{\text{Hombres activos}}$$

señala el número de mujeres activas por cada hombre activo.

## 6.3. Índice de juventud

La población está experimentando un proceso de envejecimiento que tiene su reflejo en la población activa. Así pues, resulta interesante disponer de un indicador de la estructura por edad de la población activa.

El **índice de juventud** definido como el cociente entre la población activa entre los 16 y 39 años y la población activa entre los 40 y 64 años:

$$\frac{\text{Activos entre los 16 y 39 años}}{\text{Activos entre los 40 y 64 años}}$$

representa el número de activos entre los 16 y 39 años por cada activo entre los 40 y 64 años.

## 7. Presentación de resultados

Con periodicidad bienal, el IGE publica en su página web ([www.ige.eu](http://www.ige.eu)) dentro del apartado temático de **Trabajo** la siguiente información para el período de proyección:

- Tasas de actividad por sexo y grupos decenales de edad
- Población activa por sexo y grupos decenales de edad
- Tasas de actividad de la población de 16 o más años por sexo
- Población activa de 16 o más años por sexo
- Tasas de actividad de la población entre los 16 y 64 años por sexo
- Población activa entre los 16 y 64 años por sexo
- Índices de sustentabilidad, feminidad y juventud