

Article

Evolución de la actividad, ocupación para grupos de edad mayores de 16 años

Jeremy Joel Aguilar Marin, Jose Aguilar Milla, Javier Herrero Pérez

* Correspondence:

Abstract: Este estudio realiza un análisis exploratorio y estadístico de la Encuesta de Población Activa (EPA) para evaluar la evolución de la brecha de género en el mercado laboral español entre el 2006 y 2024. Los resultados clave validados mediante pruebas de t de student, demuestran que la tasa de paro y de actividad dependen del sexo.

Keywords: Brecha de Género; Tasa de Actividad; Tasa de Paro; Análisis de Series de Tiempo; t de Student.

1. Introducción

Este artículo presenta un enfoque en R para el Análisis Exploratorio de Datos de la Encuesta de Población Activa (EPA), centrándose en la evolución histórica de la brecha de género en el mercado laboral español. Utilizando el ecosistema *tidyverse*, el estudio comienza con una limpieza y transformación de los datos empleando *dplyr* y el cálculo de Tasa de paro y actividad, claves para el análisis. Seguido de la visualización con *ggplot2* para ilustrar la evolución temporal. Por último, se validan los resultados con pruebas de *t Student* para confirmar la significancia estadística de los resultados.

La cuestión que queremos resolver en el trabajo es ver si hay relación entre la tasa de paro y de actividad con el sexo basándonos en la hipótesis de que existe una brecha de género en el mercado. Para este análisis nos hemos basado principalmente en las tasas de paro y de actividad por considerarlas los indicadores más representativos del mercado laboral. Son los mejores indicadores porque permite entender tanto quién participa en el mercado laboral mediante la tasa de actividad como quién logra empleo dentro de él con la tasa de paro.

Los datos se han obtenido del INE a fecha de 21/10/2025.

El código fuente completo y reproducible de este análisis se encuentra en este repositorio de [Github](#).

2. Exploración inicial

2.1. Importación y verificación inicial

Los datos utilizados pertenecen a la [Encuesta de Población Activa \(EPA\)](#). Se ha comprobado que la codificación de los datos brutos es UFT-8 con una confianza del 100%, el archivo está delimitado por tabulaciones. Para una correcta lectura de los datos numéricos, se indica que el separador decimal es la coma.

Las variables originales del dataframe son: **Sexo, Edad, Relación con la actividad económica, Periodo y Total**. Es importante resaltar que los valores ausentes (NA) están codificados como "...", lo cual ha sido tratado directamente con la importación.

2.2. Valores faltantes

Los valores NA se concentran en la variable **Total**, mostrando una ausencia específica en las categorías de situación laboral relacionadas con el desempleo, como se muestra en la siguiente figura:

Citation: . Evolución de la actividad, ocupación para grupos de edad mayores de 16 años. *Journal Not Specified* **2024**, *1*, 0. <https://doi.org/>

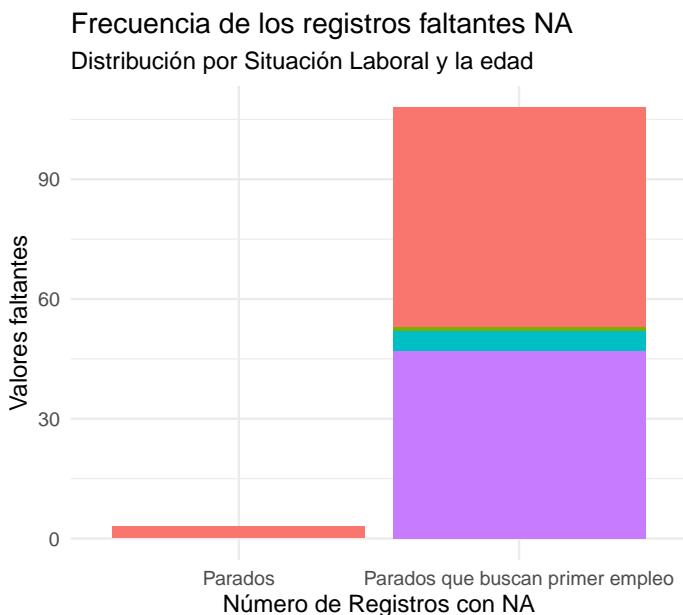
Received:

Revised:

Accepted:

Published:

Copyright: © 2025 by the authors. Submitted to *Journal Not Specified* for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).



Se observa que los datos faltantes son del tipo MAR (Missing At Random). La ausencia de datos viene asociada a grupos de edad altos con una relación con la actividad económica de "Parados que buscan primer empleo" o "Parados". Esta dependencia permite concluir que este valor NA se debe a que es imposible, o extremadamente improbable, que una persona en ese rango de edad busque su primer empleo.

3. Definición de variables

El valor de la columna Total viene definido con la siguiente expresión:

$$\text{Total} = \text{Activos} + \text{Inactivos} \quad (1)$$

Y los **Activos** se dividen en Ocupados y Parados:

$$\text{Activos} = \text{Ocupados} + \text{Parados} \quad (2)$$

3.1. Fórmulas para el análisis (Tasas)

Las tasas permiten normalizar los datos y son la base para el análisis de brechas y series temporales:

La **tasa de paro** mide la proporción de la población activa que está desempleada:

$$\text{Tasa de paro} = \frac{\text{Parados}}{\text{Activos}} \cdot 100 \quad (3)$$

La **tasa de actividad** mide la participación de una población en el mercado laboral:

$$\text{Tasa de actividad} = \frac{\text{Activos}}{\text{Total}} \cdot 100 \quad (4)$$

4. Limpieza de los datos

La limpieza de los datos ha consistido en estos puntos:

- **Renombrar columnas:** Utilizando `rename()` para acortar "Relación con la actividad económica" por "Situacion_Laboral" para simplificar el manejo de columnas.
- **Filtrado de doble conteo:** Se eliminan las instancias donde las variables representan la suma de otras instancias (`Sexo="Ambos sexos"` y `Edad="Total"`) para evitar doble conteo y así verificar que cada registro sea una observación única.

- **Se aplica pivot_wider()** utilizando los valores de la columna Situacion_Laboral para crear las nuevas columnas (names_from) y los valores numéricos de la columna Total. 59
60
61
- **Renombrar variables:** Se renombra la columna “parados que buscan primer empleo” a un formato más corto “ParadosSinExperiencia”. 62
63
- **Imputación de NA a 0:** Los valores faltantes se concentran en categorías como personas mayores en paro o que buscan su primer empleo. El INE suprime el dato por ser un valor insignificante, por lo que se opta por remplazar los valores NA a 0 para no perder el resto de información válida de la instancia. 64
65
66
67
- **Creación de tasas:** Se añaden las columnas Tasa_paro y Tasa_actividad utilizando mutate() y las expresiones de la EPA. 68
69
- **Ordenación en factores:** Se ordena la columna Edad como un factor ordinal y la columna Sexo como factor nominal. 70
71

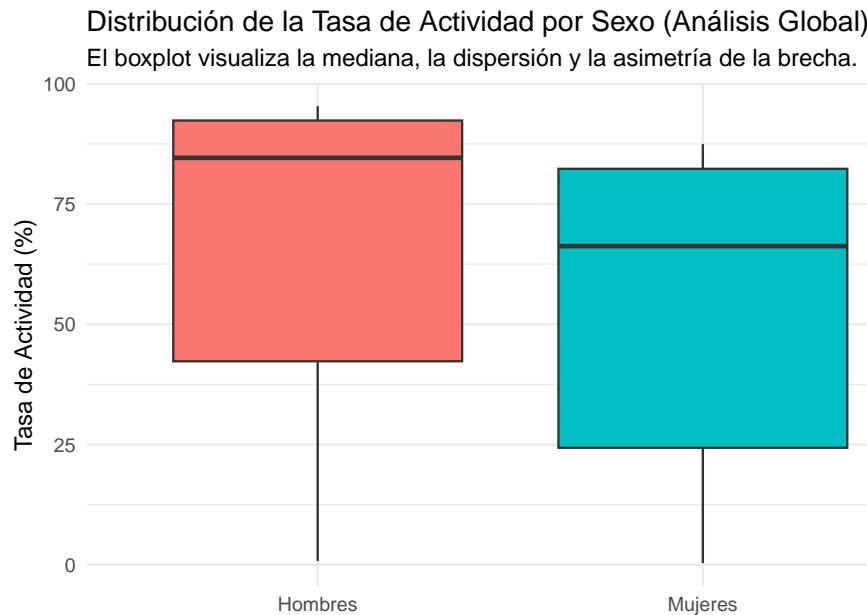
4.1. Outliers

Para identificar outliers se va a utilizar el método del rango intercuartílico (IQR) ya que permite la detección de outliers sin asumir la normalidad de la distribución. Se obtiene que para la tasa de actividad no hay valores outliers, sin embargo, para la tasa de paro sí. Se tendrá en cuenta este resultado para evitar trabajar con estos valores atípicos cuando se estudie la tasa de paro.

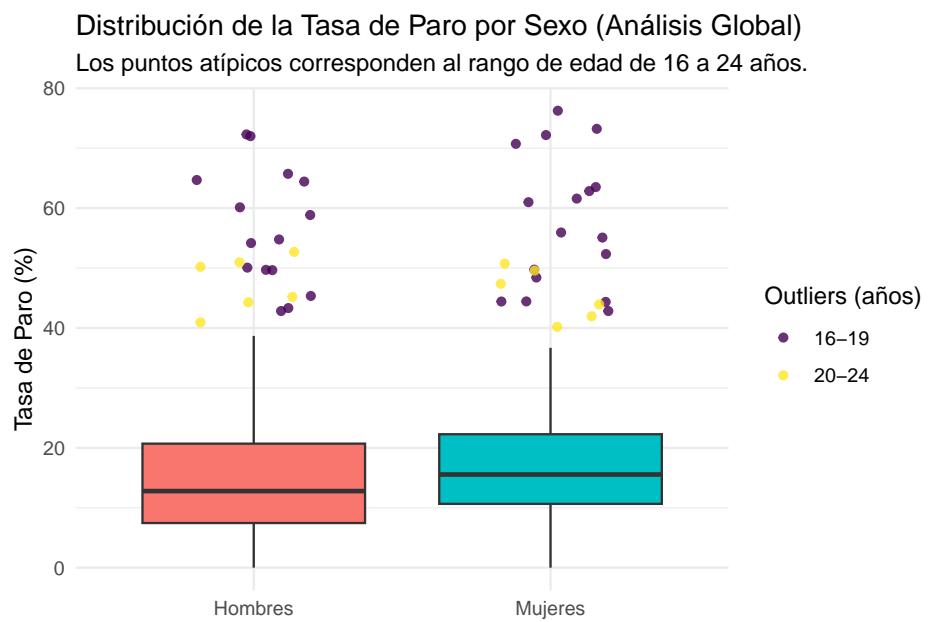
5. Análisis de las tasas

5.1. Distribución de la tasa de actividad y paro por sexo

En las siguientes gráficas se va a estudiar la distribución general de las tasas a lo largo de todo el periodo y separado por sexo.



En este primer gráfico se muestra la tasa de actividad para hombres y mujeres. Se observa una clara y pronunciada separación entre estas distribuciones. La mediana de la tasa de actividad para hombres es superior a la de las mujeres. Además, la caja del grupo masculino se sitúa por encima del grupo femenino.



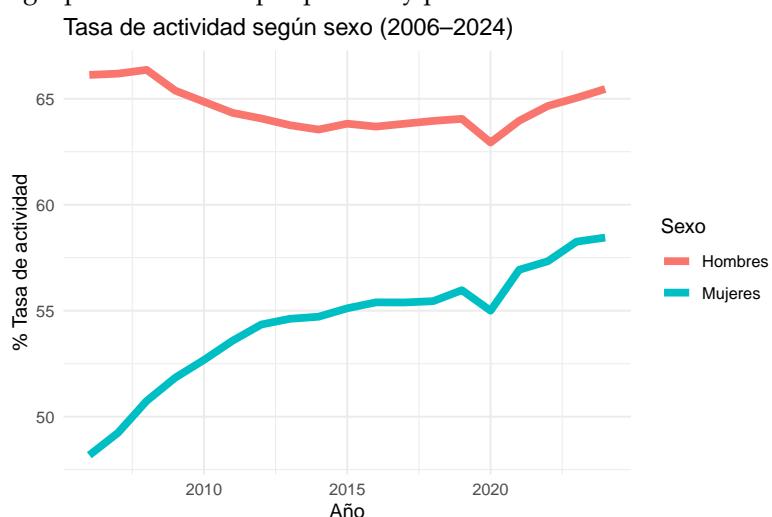
Mientras que la tasa de actividad no se han detectado outliers, la tasa de paro muestra una alta concentración de valores atípicos en ambos sexos. Los puntos atípicos se han filtrado y coloreado según el rango de edad.

Los valores atípicos en la tasa de paro están ligados a la variable de edad. Estos valores no son errores de registro sino la muestra del paro juvenil. Estos valores outliers se omitirán cuando se haga análisis relacionados con la tasa de paro.

5.2. Evolucion de la tasa de paro y de la brecha de paro por género

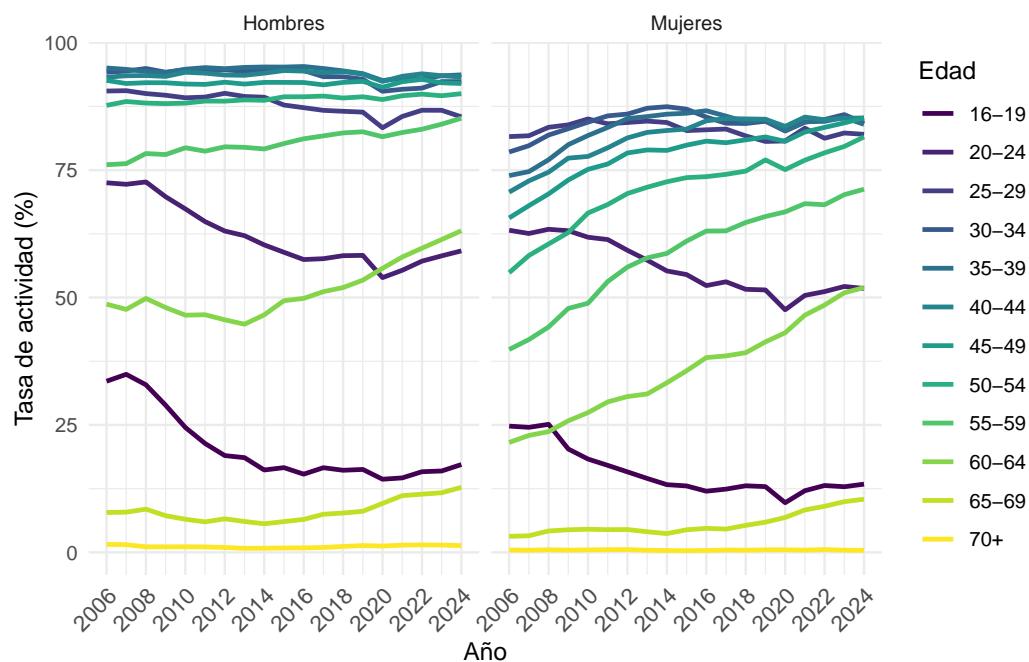
En este bloque de código se construyen varias representaciones gráficas con el fin de analizar la evolución de la tasa de paro y la brecha de género a lo largo del tiempo. Para ello, se parte de una tabla en la que se calcula la variable *brecha_paro*, definida como la diferencia entre la tasa de paro de mujeres y la de hombres. Posteriormente, los datos se agrupan por periodo para obtener las tasas medias anuales de paro por sexo y la brecha media correspondiente.

Agrupación de datos por periodo y por sexo

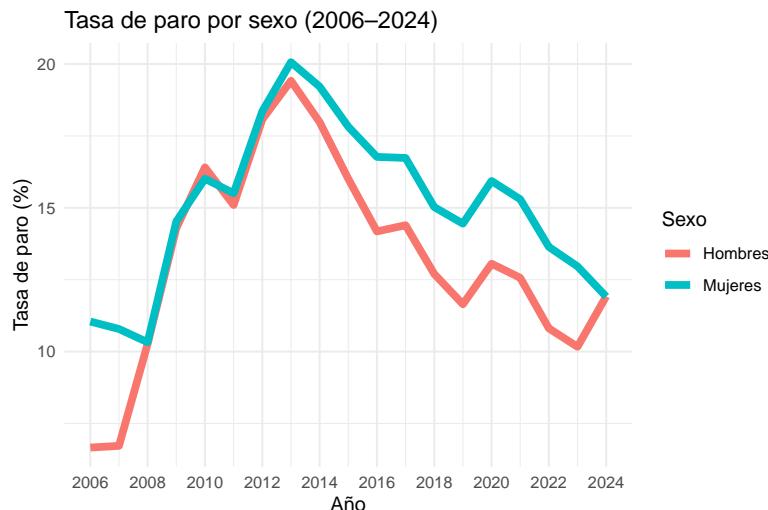


La gráfica muestra la evolución de la tasa de actividad en España entre los años 2006 y 2024, diferenciada por sexo. Por un lado, la proporción de hombres es mucho mayor a la de mujeres, cosa que va ligada a que los hombres siempre han tenido mayor participación en muchos más sectores, como lo puede ser la construcción, y a que las mujeres se ligan mas al cuidado familiar. Pese a esto, la actividad masculina en 2008 se desacelera economicamente, esto provocado por la crisis financiera que se vivió en la época. Es en 2017/18 que comienza su recuperación, pero nunca a los niveles previos a la crisis. En 2020, debido al Covid-19, se observa otra caída en la tendencia (porque el Covid-19 afectó a muchos sectores, como la industria por ejemplo). Aunque se observa también su recuperación, tampoco volvió a los niveles previos de la crisis de 2008 por el otro lado, la tasa de actividad femenina se muestra, en el periodo, ascendente ya que han ido incorporándose cada vez más en el mercado laboral español.

Tasa de actividad por edad y sexo (2006–2024)



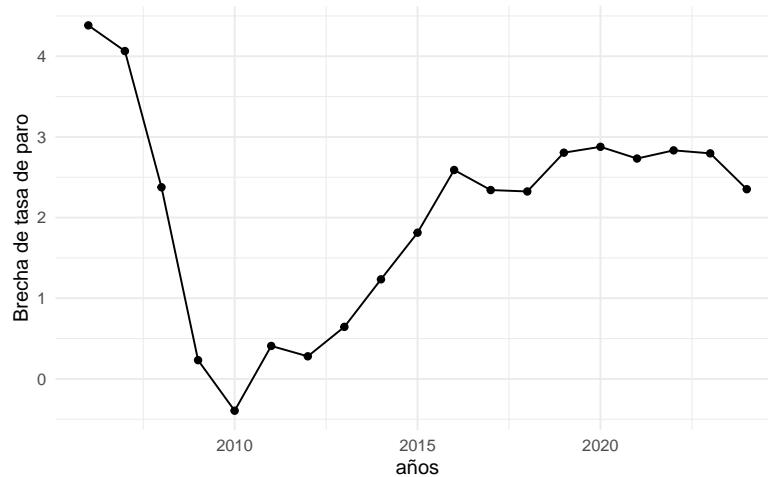
Como se observa en la gráfica, los hombres presentan tasas de actividad mucho más elevadas que las mujeres, sobretodo en el rango de 30 y 54 años. En ambos sexos, la tasa decrece a partir de los 60 años, reflejando consigo la desigualdad en el mercado cuando se trata de envejecimiento.



Desde la crisis de 123
2008, la tasa de paro en ambos sexos ha ido aumentado progresivamente hasta 2013, año 124
en el que empezó a decaer. Lo que se observa, además, es como desde 2014 el paro en 125
mujeres es mucho mas alto que el de hombres, reflejando la brecha persistente. Por tanto, 126
aunque ambos sexos sufren paro, las mujeres tienen mas dificultades para acceder (o 127
incluso mantenerse) a trabajos.

El gráfico a continuación representa la evolución de la brecha de paro total. 129

Evolución de la brecha de paro total

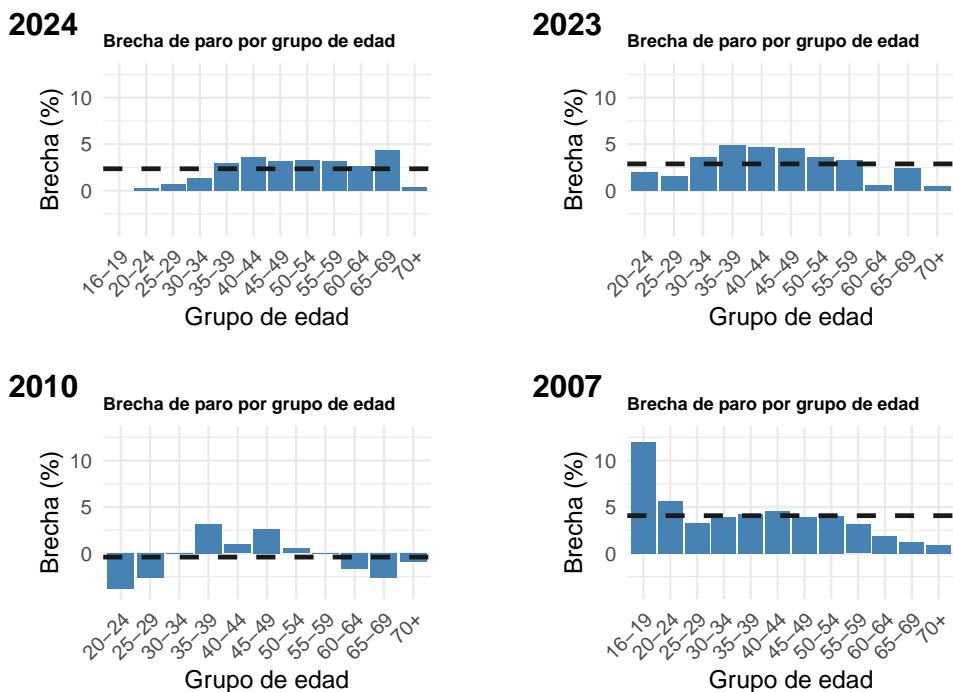


El gráfico previo representa la evolución de la brecha de paro total. En él se observa 130
cómo la diferencia entre hombres y mujeres en la tasa de paro varía a lo largo del tiempo, 131
con un aumento notable en los años posteriores a la crisis de 2008 y una ligera reducción 132
en la última etapa analizada. 133

6. Evolucion de la tasa de paro y su brecha por grupo de edad

6.1. Gráfica de Barras comparación grupos de edades con mayor brecha en 2006 y en 2024

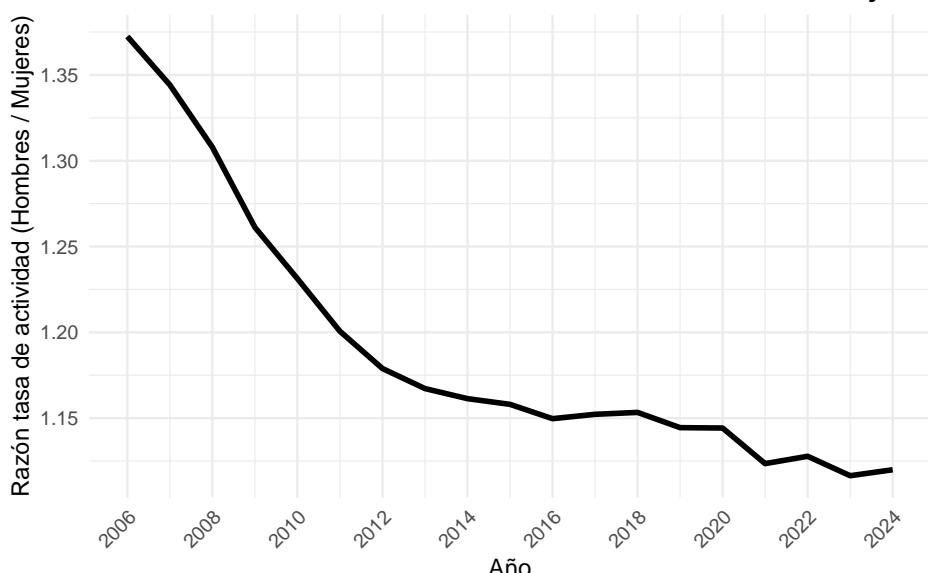
En este apartado se elaboran una serie de gráficos de barras destinados a analizar la 137
evolución de la brecha de paro por grupos de edad a lo largo del tiempo. En cada uno 138
de ellos, las barras representan la magnitud de la brecha en los distintos grupos de edad, 139
mientras que la línea discontinua negra indica la media general de la brecha para ese año. 140



Los gráficos resultantes muestran, de forma clara, cómo la intensidad de la brecha de género en el desempleo varía según la edad y el momento temporal. En los primeros años analizados, la brecha es especialmente elevada entre los jóvenes, lo que evidencia una mayor vulnerabilidad del empleo femenino en los jóvenes. En cambio, en los años más recientes para esos tramos de edad, las diferencias tienden a moderarse y presentan un patrón más equilibrado entre grupos, aunque persisten desigualdades estructurales en los segmentos de mayor edad y los más jóvenes.

Este bloque busca ofrecer una visión desagregada y evolutiva de la brecha de paro por edad. Gracias a la representación mediante diagramas de barras, se facilita la comparación entre grupos generacionales y se identifican con claridad los grupos que más contribuyen a mantener las disparidades laborales entre hombres y mujeres a lo largo del tiempo.

Evolución de la razón de la tasa de actividad Hombres/Mujeres



Esta gráfica muestra cómo ha cambiado la relación entre la tasa de actividad de los hombres y la de las mujeres a lo largo del tiempo. Cada punto representa el cociente entre ambas

141
142
143
144
145
146
147
148
149
150
151
152

153
154
155

tasas en un año determinado. Una razón superior a 1 indica que los hombres tienen mayor participación en el mercado laboral que las mujeres. Desde 2006, la razón ha descendido de forma sostenida, lo que refleja una reducción progresiva de la brecha de género. Aunque no se ha alcanzado la igualdad, las mujeres han incrementado su participación relativa.

156
157
158
159

7. Distribución de la población en el mercado laboral

160

7.1. Composición del mercado laboral

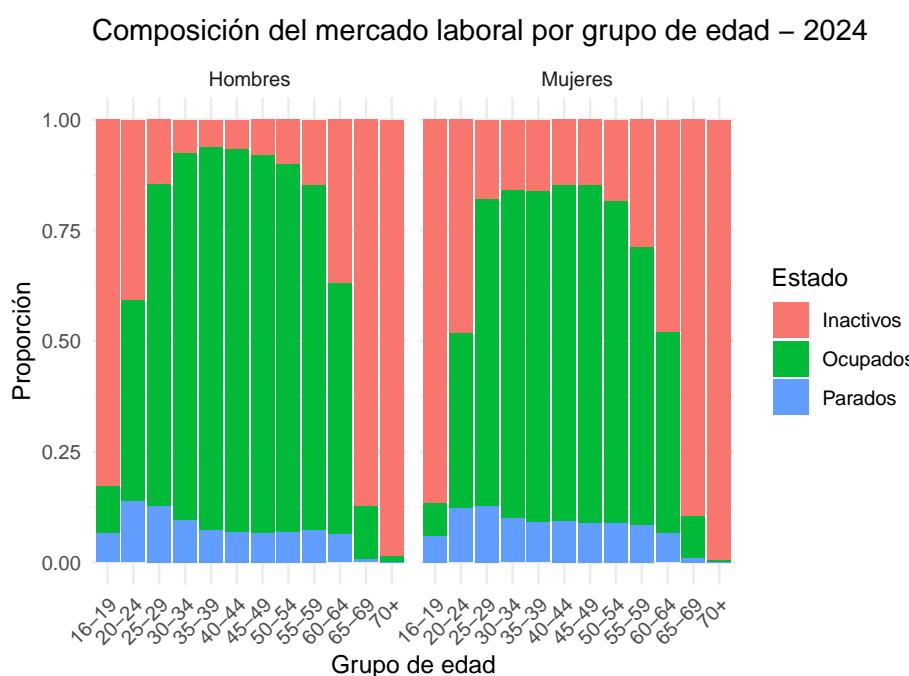
161

En el siguiente gráfico se muestra la composición del mercado laboral, separando entre: Parados, Ocupados e Inactivos para cada grupo de edad, separado por sexo.

162
163

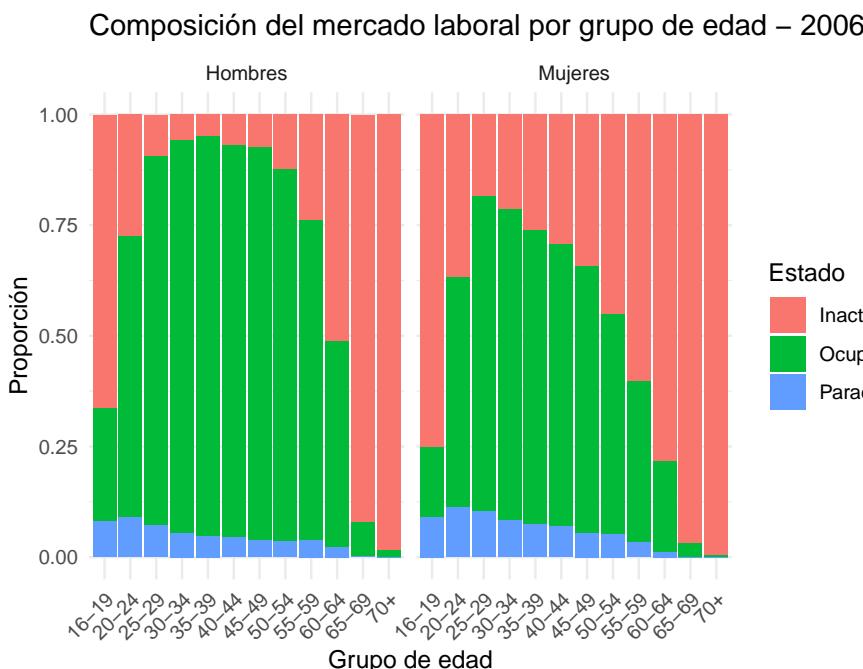
```
## [[1]]
```

164



```
##  
## [[2]]
```

165
166
167



Del análisis de los gráficos se observa una estructura bastante estable del mercado laboral por edad y sexo a lo largo del tiempo, con ligeras variaciones asociadas a los ciclos económicos. En ambos sexos, el grupo central de edad (30-54 años) concentra la mayor proporción de personas ocupadas, mientras que los extremos jóvenes y mayores presentan una mayor presencia de inactivos. Las mujeres muestran sistemáticamente una menor proporción de ocupación y una mayor de inactividad en comparación con los hombres, aunque esta diferencia tiende a reducirse en los años más recientes. Asimismo, la proporción de personas paradas es más significativa en los grupos jóvenes, reflejando una mayor vulnerabilidad del empleo juvenil. En conjunto, los gráficos ponen de manifiesto que, pese a los avances en participación femenina y la estabilidad general del empleo en edades medias, persisten desigualdades de género e importantes diferencias intergeneracionales dentro del mercado laboral.

Estas gráficas, también ponen de manifiesto un proceso de envejecimiento del desempleo y una convergencia progresiva entre hombres y mujeres en la estructura interna del paro, aunque todavía persisten diferencias derivadas de los roles laborales y sociales tradicionales. Mientras los hombres mayores ganan peso dentro del paro total por el retraso en la jubilación, las mujeres siguen mostrando una mayor vulnerabilidad en las edades centrales, reflejando las desigualdades de género que aún caracterizan el mercado laboral español.

8. Análisis estadístico

Para validar estadísticamente las diferencias observadas entre las series temporales por sexo se va a realizar una prueba de t de Student para las dos principales métricas del mercado laboral, la tasa de actividad y la de paro. Los resultados se pueden observar en la siguiente tabla:

Table 1. Resultados de las pruebas t para comparar las tasas entre sexos

| Tasa | t | df | p.value |
|-----------|--------|--------|---------|
| Actividad | 3.298 | 449.48 | 0.0010 |
| Paro | -2.288 | 408.85 | 0.0227 |

Como se observa en esta tabla, la tasa de actividad presenta un valor *p* extremadamente bajo (0.001), lo que permite rechazar la hipótesis nula y por lo tanto, hay indicios estadísticos de que la tasa de actividad depende del sexo.

Por otro lado, la tasa de paro muestra un valor de *p* también bajo (0.023) rechazando de la misma forma la hipótesis nula.

Por lo tanto, el análisis muestra argumentos estadísticos suficientes como para afirmar con un 95% de nivel de confianza que existe una dependencia significativa entre el sexo y la tasa de paro y actividad.

9. Conclusión

Mediante el análisis gráfico se ha observado que la tasa de actividad masculina es superior, aunque se va reduciendo progresivamente. También se aprecia que las mayores diferencias se dan en los grupos de mediana edad que coinciden con las etapas de mayor carga familiar y laboral. En términos de la composición del mercado laboral, se aprecia claramente las mujeres siguen mostrando una mayor inactividad y ocupación en el empleo aunque las diferencias se reducen gradualmente. Tras el análisis de outliers se ha comprobado gráficamente que la tasa de paro entre grupos de edad de 16 a 24 años es muy superior al resto. Por esta razón se decidió excluir estos valores para el análisis de la tasa de paro.

Para complementar las gráficas de una forma rigurosa se ha hecho una prueba de *t* de student en la que se ha obtenido como resultado que la dependencia entre la tasa de actividad y paro con el sexo es estadísticamente significativa, esto nos permite concluir que la brecha de género existe en el mercado laboral.

Por tanto, con este análisis sobre la brecha de genero se ha podido detectar como esta brecha ha sido real en el intervalo de tiempo estudiado.

La sociedad actual debería poner mas atención a esta brecha y, por ende, promover diferentes políticas que beneficien la conciliación laboral y fomenten la inclusión laboral femenina.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.