



FACULTAD DE ESTUDIOS ESTADÍSTICOS

GRADO EN ESTADÍSTICA APLICADA

Curso 2025/2026

Trabajo de Fin de Grado

TÍTULO: Determinantes de la participación deportiva federada en España: un análisis territorial y temporal

Alumno: Claudia Suano Gil

Tutor: Alicia Pérez Alonso

Enero de 2026



UNIVERSIDAD COMPLUTENSE
MADRID



**Declaración Responsable sobre Autoría y Uso Ético de
Herramientas de Inteligencia Artificial (IA)**

Yo, **SUANO GIL CLAUDIA**

declaro de manera responsable que el/la presente:

- Trabajo de Fin de Grado (TFG)
- Trabajo de Fin de Máster (TFM)
- Tesis Doctoral

Titulado/a

Determinantes de la participación deportiva federada en España: un análisis territorial y temporal

es el resultado de mi trabajo intelectual personal y creativo, y ha sido elaborado de acuerdo con los principios éticos y las normas de integridad vigentes en la comunidad académica y, más específicamente, en la Universidad Complutense de Madrid.

Soy, pues, autor del material aquí incluido y, cuando no ha sido así y he tomado el material de otra fuente, lo he citado o bien he declarado su procedencia de forma clara -incluidas, en su caso, herramientas de inteligencia artificial-. Las ideas y aportaciones principales incluidas en este trabajo, y que acreditan la adquisición de competencias, son mías y no proceden de otras fuentes o han sido reescritas usando material de otras fuentes.

Asimismo, aseguro que los datos y recursos utilizados son legítimos, verificables y han sido obtenidos de fuentes confiables y autorizadas. Además, he tomado medidas para garantizar la confidencialidad y privacidad de los datos utilizados, evitando cualquier tipo de sesgo o discriminación injusta en el tratamiento de la información.

En Madrid a **23 de enero de 2026**

AGRADECIMIENTOS

A mi familia y amigos, por confiar siempre en mí y apoyarme incondicionalmente.

RESUMEN

Este trabajo trata de analizar la participación deportiva federada en España desde un punto de vista territorial, económico y de género, centrándose en las diferencias que existen entre comunidades autónomas y en la brecha de género en la práctica deportiva. A partir de la información de las diversas fuentes oficiales, se ha construido una base de datos a nivel autonómico que permite observar la evolución de la federación deportiva desde el 2015 al 2024, incorporando además otras variables que contienen información sobre población, empleo deportivo y nivel de renta per cápita a nivel regional.

El objetivo del estudio consiste en analizar si factores como el nivel económico regional y el peso del empleo deportivo, entre otros, influyen, y en qué medida lo hacen en la participación federada. Además, se explora cómo se agrupan las diferentes comunidades autónomas según su desarrollo deportivo y económico, para facilitar identificar distintos modelos regionales de práctica deportiva federada.

Hemos obtenido evidencia de que el nivel económico regional no influye en la participación federada, factores como el peso del empleo deportivo, influyen de otra manera, aunque no siempre está asociado a modelos deportivos necesariamente vinculados a la federación. Asimismo, se identifican distintos patrones territoriales de práctica deportiva y se podrá ver las variaciones temporales comunes, especialmente a partir de 2020, que han afectado de manera generalizada a la participación federada.

Resumiendo, los resultados muestran que la participación deportiva federada responde a muchos factores desde estructurales, institucionales, hasta sociales que van más allá del crecimiento económico, lo que señala la utilidad de diseñar políticas deportivas adaptadas a las necesidades de cada comunidad autónoma y orientadas a reducir las desigualdades de género en el deporte federado.

ABSTRACT

This study analyzes federated sports participation in Spain from a territorial, economic, and gender perspective, focusing on the differences between autonomous communities and the gender gap in sports practice. Using information from various official sources, a regional-level database was constructed to observe the evolution of sports federation from 2015 to 2024, incorporating additional variables such as population, sports-related employment, and regional per capita income levels.

The objective of the study is to analyze whether factors such as regional economic level and the weight of sports employment, among others, influence federated participation and to what extent. Furthermore, the study explores how different autonomous communities cluster according to their sports and economic development to help identify distinct regional models of federated sports practice.

The evidence obtained suggests that regional economic levels do not influence federated participation. Factors such as the weight of sports employment influence it differently, though they are not always linked to sports models necessarily associated with federation. Likewise, distinct territorial patterns of sports practice are identified, alongside common temporal variations—especially from 2020 onwards—that have broadly affected federated participation.

In summary, the results show that federated sports participation responds to a multitude of factors—ranging from structural and institutional to social—that go beyond economic growth. This highlights the utility of designing sports policies tailored to the specific needs of each autonomous community and aimed at reducing gender inequalities in federated sports.

INDICE

AGRADECIMIENTOS	iv
RESUMEN	v
ABSTRACT	vi
INTRODUCCIÓN	1
1. DATOS Y CONSTRUCCIÓN DE LA BASE DE DATOS.	3
1.1 FUENTES DE INFORMACIÓN	3
1.2 PROCESO DE INTEGRACIÓN Y DEPURACIÓN DE DATOS	4
1.3 DEFINICIÓN Y DESCRIPCIÓN DE VARIABLES	5
2. METODOLOGÍA	8
3. ANÁLISIS DESCRIPTIVO UNIVARIANTE	15
3.1 ANÁLISIS DE LA TASA DE FEDERACIÓN	15
3.1.1 DISTRIBUCIÓN GLOBAL.....	15
3.1.2 DIMENSIÓN DE LA TIPOLOGÍA DEL DEPORTE	16
3.1.3 PERSEPECTIVA DE GÉNERO	20
3.1.4 DIMENSIÓN TERRITORIAL	23
3.1.5 DINÁMICA TEMPORAL.....	26
3.2 ANÁLISIS DEL EMPLEO DEPORTIVO	29
3.3 ANÁLISIS DEL PIB PER CÁPITA	31
4. ANÁLISIS DESCRIPTIVO MULTIVARIANTE	33
4.1 ANÁLISIS DE COMPONENTE PRINCIPALES	33
4.1.1 CONTRIBUCIÓN LAS DE VARIABLES Y DEFINICIÓN DE DIMENSIONES ..	34
4.1.2 REPRESENTACIÓN GRÁFICA	35
4.2 ANÁLISIS CLUSTER	37
4.2.1 IDENTIFICACIÓN DE PERFILES REGIONALES	38
4.3 INCORPORACIÓN DE GÉNERO	40
4.3.1 BRECHA DE GÉNERO EN LA TASA DE FEDERACIÓN DEPORTIVA	40
4.3.2 RELACIÓN ENTRE LA ACTIVIDAD ECONÓMICA Y PARTICIPACIÓN DEPORTIVA POR SEXO	43
5. ESTIMACIÓN Y DISCUSIÓN DE RESULTADOS	45
5.1 MODELO DE REGRESIÓN LINEAL	45
5.2 ANÁLISIS DE DATOS PANEL	47
5.2.1 ESTACIONARIEDAD DE LAS VARIABLES	48
5.2.2 MODELO BASE	49
5.2.3 MODELO EFECTOS FIJOS VERSUS EFECTOS ALEATORIOS.....	50
5.2.4 ELECCIÓN DEL MODELO	52
5.3.5 MODELO DE EFECTOS ALEATORIOS CORRELACIONADOS.....	53
5.3.6 VALIDACIÓN DEL MODELO	55

6. CONCLUSIONES57
7. REFERENCIAS BIBLIOGRÁFICAS59

ÍNDICE DE FIGURAS

Figura 1: Distribucion de la tasa de federación en España	16
Figura 2: Top 15 deportes con mayor número de federados en España (2015–2024)	17
Figura 3: Diagrama de cajas: Top 20 deportes por tasa de federación en España (Ordenado por Mediana)	18
Figura 4: Gráficos caja con dispersión de la tasa de federación por tipo de deporte (equipo vs. individual).....	19
Figura 5: Gráfico de cajas: Tasa de federación por categoría funcional (Nivel 2)	20
Figura 6: Evolución de la tasa de federados por sexo.	21
Figura 7: Top 15 deportes por número total de federados según género .	22
Figura 8: Tasa de federación por tipo de deporte y sexo	23
Figura 9: Tasa de federados promedios por CC.AA.	24
Figura 10: Tasa de federación por CC.AA. en España.....	24
Figura 11: Gráfico de cajas: Tasa de federación por CCAA y Sexo	25
Figura 12: Evolución del número total de federados (2015-2024)	27
Figura 13: Evolución de la tasa de federación según Nivel1	28
Figura 14: Evolución temporal de la tasa de federación según Nivel2.....	29
Figura 15: Diagrama de cajas: peso de empleo deportivo por sexo.	30
Figura 16: Evolución del empleo deportivo (2015-2024)	30
Figura 17: Evolución del empleo deportivo por CC.AA.....	31
Figura 18: Evolución de totalPerCapita por CC.AA.	32
Figura 19: Evolución de totalPerCapita transformada.....	33
Figura 20 Scree plot: Dimensiones de componentes principales	35
Figura 21: Mapa conjunto de variables y PCA.....	36
Figura 22: Círculo de las correlaciones de PCA.....	37
Figura 23: Perfil medio de cada Clúster	40
Figura 24: Distribución geográfica de la brecha de género deportiva.	41
Figura 25: Evolución temporal de la Brecha Media de Género	42
Figura 26: Gráfico de dispersión: Relación entre ihs_PIB y tasaFed por sexo	44

INTRODUCCIÓN

La práctica deportiva es un elemento fundamental del bienestar social, con repercusiones fuertes en la salud pública, unión social y desarrollo económico.

En España, el deporte federado ha sido tradicionalmente una de las principales formas de participación deportiva organizada. Sin embargo, en las últimas décadas se han producido cambios significativos a la hora de hacer deporte, caracterizados por el auge del deporte recreativo y de modalidades no necesariamente vinculadas a estructuras federativas, como puede ser el caso de deportes en colegios o clubes no oficiales. Tras la crisis sanitaria, el deporte, para mucha gente, se volvió una vía de escape. Paso a verse desde una perspectiva distinta. Empezó a ser más accesible. (El aumento del deporte después de la pandemia, 2024)

A pesar de esto, estas transformaciones no se han producido de la misma manera en todo el territorio español, sino que en cada comunidad autónoma se distribuye diferente. Al mismo tiempo, el crecimiento económico y empleo vinculado al sector deportivo no siempre se traducen en un aumento de la participación federada, lo que plantea dudas sobre la relación entre desarrollo económico, empleo deportivo y los diferentes deportes.

En este contexto, el presente trabajo analiza la participación deportiva federada en España mediante un enfoque empírico que combina distintas herramientas estadísticas y econométricas, con el objetivo de ofrecer una visión integral del fenómeno y de sus determinantes territoriales y temporales.

El resto del trabajo se organiza como sigue. En la sección 1 se explica cómo se han obtenido los datos, así como la construcción de la base de datos correspondiente. En la sección 2, se describe la metodología empleada para llevar a cabo el análisis, detallando las herramientas estadísticas aplicadas a lo largo del trabajo. La sección 3 se dedica al análisis univariante, en el que se estudia de forma descriptiva la distribución y el comportamiento individual de las principales variables de interés.

A continuación, en la sección 4, se desarrolla el análisis multivariante, que incluye técnicas como el Análisis de Componentes Principales (PCA) y el análisis clúster, con el objetivo de identificar patrones y agrupaciones territoriales en la participación deportiva federada.

En la sección 5 se presentan los resultados de los modelos de regresión, tanto lineales como de datos de panel, que permiten analizar de forma más rigurosa la relación entre la tasa de federación, el nivel económico y el empleo deportivo. Por último, la sección 6 recoge las principales conclusiones del estudio. Además, luego encontraremos las referencias con las que hemos contrastado y completado la información.

1. DATOS Y CONSTRUCCIÓN DE LA BASE DE DATOS.

Este apartado describe el proceso de construcción de la base de datos que se ha utilizado en el análisis empírico del trabajo. Para ello, se detallan las fuentes de información empleadas, el procedimiento seguido para su formación y depuración, así como la definición de las principales variables consideradas.

1.1 FUENTES DE INFORMACIÓN

Para la elaboración de la base de datos se ha recurrido a las fuentes de información oficiales y acceso público, que proporcionan distintos conjuntos de datos reales, seleccionadas por su fiabilidad y adecuación a los objetos del presente estudio.

En concreto, se usaron como fuentes, el “Ministerio de Educación, Formación Profesional y Deportes”, exactamente el apartado de la web, acerca del deporte, cuyo nombre es DEPORTEDATA (DEPORTEDATA, s.f.) así como el Instituto Nacional de Estadística. (INE, s.f.)

En el caso de la plataforma **DEPORTEDATA**, se recopilaron los conjuntos de datos relativos a la participación deportiva federada, desagregados por comunidad autónoma (CC.AA.), año y sexo. Por su parte, del **Instituto Nacional de Estadística (INE)** se extrajeron las variables de carácter económico y sociodemográfico necesarias para el desarrollo del análisis.

Como se ha señalado previamente, la información utilizada combina datos de naturaleza deportiva y económica a nivel autonómico, excluyéndose las ciudades autónomas de Ceuta y Melilla, dado que cuentan con un régimen competencial y estatutario diferente al de las comunidades autónomas. El periodo temporal analizado abarca desde 2015 hasta 2024 y toda la información se encuentra desagregada por sexo, distinguiendo entre hombres y mujeres.

No obstante, los datos proceden de distintos archivos independientes y presentan estructuras heterogéneas, lo que hace necesario realizar

previamente un proceso de integración, homogeneización y depuración de la información antes de abordar el análisis estadístico y econométrico.

1.2 PROCESO DE INTEGRACIÓN Y DEPURACIÓN DE DATOS

Una vez recopiladas las distintas bases de datos, se procedió a su fusión y limpieza con el objetivo de construir un conjunto de datos único, coherente y adecuado para el análisis. Este proceso incluyó la unificación de los distintos archivos anteriormente mencionados, la comprobación de la veracidad de la información y el tratamiento de posibles valores perdidos o incoherentes.

En particular, de DEPORDATA (2025), la variable seleccionada fue Número de personas federadas (*federados*), y de INE (2025) las variables seleccionadas fueron *empleados*, *personas*, *totalPerCapita*. A partir de *personas* y *federados* se construye la variable *tasaFed*. Después, a partir de *empleados* se obtiene *totalEmpleoDep*, por la cual, junto con *personas* se construye *pesoEmp*.

El proceso de construcción de la base de datos se inició a partir de la información sobre *federados*, que se toma como factor principal del análisis. En primer lugar, se seleccionaron únicamente las observaciones correspondientes a hombres y mujeres y el periodo comprendido entre 2015 y 2024, con el fin de garantizar la coherencia temporal del conjunto de datos.

A continuación, se incorporó la información sobre la población total (*personas*), procedente de dos conjuntos de datos distintos, uno correspondiente al periodo 2015–2020 y otro al periodo 2021–2024, que fueron integrados en una única estructura, manteniendo el nivel de desagregación por CC.AA.. En este proceso fue necesario tareas de depuración, como la corrección de espacios, signos de puntuación y denominaciones territoriales, para asegurar la correcta vinculación entre fuentes.

Posteriormente, se añadieron los datos acerca de los *empleados*, seleccionando las observaciones correspondientes a hombres y mujeres a nivel autonómico.

Finalmente, se incorporaron las variables económicas relativas al total per cápita, previa una depuración de datos que eliminó símbolos no numéricos, reconstruyó nombres de las variables y unificó denominaciones territoriales duplicadas.

En síntesis, todas las fuentes de información integradas de forma secuencial en una única base de datos. En primer lugar, se combinaron las variables económicas y demográficas por CC.AA. y periodo temporal; posteriormente, se incorporaron las variables de *empleo* y *totalEmpDep* desagregadas por *Sexo*, *CCAA* y *Periodo*; y, por último, se añadieron las variables relativas a las licencias federativas. Como resultado de este proceso, la base de datos final utilizada en el análisis está compuesta por 13 variables y un total de 22.440 observaciones.

Para completar el estudio, además de las variables que se construyeron adicionalmente para facilitar el futuro análisis como: *tasaFed* y *pesoEmp*, se añade una clasificación deportiva, para simplificar su interpretación definiendo dos variables categóricas, según criterios explicados posteriormente establecidos, que son: *Nivel1* y *Nivel2*.

Todo el proceso de integración y depuración de los datos se llevó a cabo mediante el software estadístico R, lo que garantiza la reproducibilidad y coherencia del análisis.

1.3 DEFINICIÓN Y DESCRIPCIÓN DE VARIABLES

Este apartado consiste en una explicación clara y concisa de cada variable incluida en la base de datos final, centrándose en el definición y finalidad de cada una de ellas.

Las variables son las siguientes:

1) Variables indicadoras:

- **Sexo:** variable cualitativa nominal que nos permite estructurar e identificar la información ya que distingue entre hombres y mujeres.
- **CCAA:** variable cualitativa nominal que hace referencia a las Comunidades Autónomas, recoge las 17 españolas.

- Periodo: variable cuantitativa discreta que corresponde al año de referencia, puede abarcar desde el 2015 al 2024.
- 2) Variables desagregadas a partir de los indicadores anteriores:
 - Personas: variable cuantitativa discreta que recoge la población total
 - empleados: variable cuantitativa discreta que cuantifica el total de personas ocupadas.
 - totalEmpleoDep: variable cuantitativa discreta que marca la cantidad de personas empleadas en el sector deportivo.
 - totalPerCapita: variable económica que mide el Producto Interior Bruto (PIB) per cápita (calculado sobre la población total).
 - Federados: variable cuantitativa discreta que representa la cantidad de licenciados que hay por deporte.
 - tasaFed: variable cuantitativa continua que describe la tasa de federados por cada 1.000 habitantes.
 - pesoEmp: variable cuantitativa continua que describe la tasa de empleo en el sector deportivo por cada 1.000 empleados totales.
 - Deportes: variable cualitativa nominal que recoger todos los tipos de deporte, 65 en total.
 - ihs_PIB: variable *totalPerCapita* transformada mediante el seno hiperbólico inverso (IHS por sus siglas en inglés, Inverse Hyperbolic Sine)

Para posteriormente reducir la dimensionalidad de la variable Deportes (65 categorías), añadimos dos columnas adicionales:

- Nivel1: variable cualitativa nominal que clasifica las modalidades deportivas en deportes de equipo y deportes individuales, siguiendo los criterios establecidos por el Consejo Superior de Deportes (CSD, 2022)
- Nivel2: variable cualitativa nominal que agrupa las modalidades deportivas según su lógica funcional, distinguiendo entre deportes de invasión o territorio, deportes de red o muro, deportes de combate, deportes de precisión, deportes de carácter estético-expresivo y deportes de resistencia cíclica u *outdoor*, de acuerdo con la literatura

de didáctica y teoría del deporte. (Hernández Moreno, 2000), (Parlebas, 2001)

Contamos con 16 modalidades deportivas según su lógica funcional:

- Adaptado: Deporte Discapacidad Física, Deporte Discapacidad Intelectual, Deporte para Ciegos, Deporte para Sordos, Deporte Parálisis Cerebral y DCA.
- Animales: Colombicultura, Colombófila, Galgos, Hípica, Polo.
- Aventura: Espeleología, Montaña y Escalada, Orientación.
- Bate y Campo: Béisbol y Sófbol.
- Combate: Boxeo, Esgrima, Judo, Karate, Kickboxing y Muaythai, Luchas Olímpicas, Taekwondo.
- Combinado: Pentatlón Moderno.
- Estrategia: Ajedrez.
- Estética: Baile Deportivo, Gimnasia, Patinaje.
- Fuerza: Halterofilia.
- Invasión: Baloncesto, Balonmano, Fútbol, Fútbol Americano, Hockey, Rugby.
- Motor: Automovilismo, Motociclismo, Motonáutica.
- Precisión: Aeronáutica, Billar, Bolos, Caza, Golf, Pesca y Casting, Petanca, Tiro a Vuelo, Tiro con Arco, Tiro Olímpico.
- Red/Muro: Bádminton, Pádel, Pelota, Squash, Tenis, Tenis de Mesa, Voleibol.
- Rescate: Salvamento y Socorrismo.
- Resistencia: Actividades Subacuáticas, Atletismo, Ciclismo, Natación, Piragüismo, Remo, Triatlón.
- Técnico: Deportes de Hielo, Deportes de Invierno, Esquí Náutico y Wakeboard, Surfing, Vela.

De esta estructura de datos nos sale una base de datos constituida 22.440 filas y 13 columnas.

Con el fin de facilitar la comprensión de la estructura de la base de datos, a continuación, se presentan, a modo ilustrativo, las tres primeras y las tres últimas observaciones del conjunto de datos. Esta información nos permite

visualizar de forma clara la organización de las variables y sirve como apoyo para la correcta interpretación del análisis descriptivo posterior.

CCAA	Periodo	totalPerCapita	Sexo	personas	empleados	totalEmpleoDep	Deportes	federados	tasaFed	pesoEmp	Nivel1	Nivel2
Andalucía	2024	2.85	Hombres	4249355	2152.2	78.6	Actividades Subacuáticas	3.379	0.0007951795	36.52077	Individual	Resistencia
Andalucía	2024	2.85	Hombres	4249355	2152.2	78.6	Aeronáutica	719.000	0.1692021495	36.52077	Individual	Precisión
Andalucía	2024	2.85	Hombres	4249355	2152.2	78.6	Ajedrez	3.135	0.0007377590	36.52077	Individual	Estrategia
Rioja, La	2015	4.04	Mujeres	160320	68.0	6.2	Triatlón	21.000	0.130988024	91.17647	Individual	Resistencia
Rioja, La	2015	4.04	Mujeres	160320	68.0	6.2	Vela	0.000	0.0000000000	91.17647	Individual	Técnico
Rioja, La	2015	4.04	Mujeres	160320	68.0	6.2	Voleibol	583.000	3.636477046	91.17647	Equipo	Red/Muro

2. METODOLOGÍA

En este apartado se definen las técnicas utilizadas para los diferentes tipos de análisis con lo que se ha desarrollado el estudio. Se van a explicar en profundidad todas las herramientas necesarias para abordarlo.

- IHS: inverso del Seno Hiperbólico. Calcula el seno hiperbólico de un ángulo. Las funciones trigonométricas se basan en la hipérbola con la ecuación $x^2 - y^2 = 1$. Estas funciones difieren de las utilizadas en la trigonometría estándar (también llamada trigonometría circular), cuyas funciones se basan en el círculo unitario con la ecuación $x^2 + y^2 = 1$. Sin embargo, comparten muchas identidades similares, tales como $\sinh^2 x + \cosh^2 x = 1$, donde h representa hiperbólico. (Función Seno hiperbólico.)
- Histograma: es un gráfico estadístico que representa la distribución de un conjunto de datos continuo a través de barras trazadas, cada una de las cuales representa una categoría o intervalo de clase particular. La altura de la barra refleja la frecuencia o el recuento de puntos de datos dentro de cada grupo. En esencia, los histogramas revelan patrones, tendencias y conocimientos ocultos dentro de datos sin procesar y desordenados, exhibiendo una imagen organizada que hace que la interpretación y el análisis sean mucho más manejables y efectivos. (Histograma.)
- Diagrama de cajas: es un gráfico utilizado para representar una variable cuantitativa (variable numérica). El gráfico es una

herramienta que permite visualizar, a través de los cuartiles, cómo es la distribución, su grado de asimetría, los valores extremos, la posición de la mediana, etc. Se compone de:

Un rectángulo (*caja*) delimitado por el primer y tercer cuartil (Q_1 y Q_3). Dentro de la caja una línea indica dónde se encuentra la mediana (segundo cuartil Q_2).

Dos *brazos*, uno que empieza en el primer cuartil y acaba en el mínimo, y otro que empieza en el tercer cuartil y acaba en el máximo.

Los datos atípicos (o valores extremos) que son los valores distintos que no cumplen ciertos requisitos de heterogeneidad de los datos. (Universoformulas.com)

- Test de Normalidad (Shapiro-Wilk): el test de *Shapiro-Wilks* plantea la hipótesis nula que una muestra proviene de una distribución normal. Elegimos un nivel de significancia, por ejemplo 0,05, y tenemos una hipótesis alternativa que sostiene que la distribución no es normal. (Dietrichson)

Tenemos:

h_0 : La distribución es normal

h_1 : La distribución no es normal

- Test de Kruskal-Wallis: la prueba de Kruskal-Wallis es una prueba no paramétrica que se utiliza para comparar si tres o más grupos independientes provienen de la misma distribución. Es la alternativa no paramétrica al ANOVA cuando no se puede asumir la normalidad en los datos. (Kruskal-Wallis.)

Tenemos:

h_0 : no hay diferencias significativas entre las distribuciones de los grupos

h_1 : al menos uno de los grupos tiene una distribución diferente a los otros

- Test de Wilcoxon (Mann-Whitney): la prueba U de Mann-Whitney es una prueba no paramétrica alternativa a la prueba t de muestras independientes (una prueba de hipótesis estadística utilizada para determinar si una media poblacional desconocida es diferente de un valor específico). La prueba U de Mann-Whitney se utiliza para comparar dos medias muestrales que provienen de la misma población, así como para probar si dos medias muestrales son iguales o no. (Narvaez, 2022)

Tenemos:

h_0 : no existen diferencias significativas entre los grupos o que la mediana de las diferencias es igual a cero.

h_1 : si existen diferencias significativas.

- PCA: el PCA, por sus siglas en inglés Principal Component Analysis, es una técnica estadística que se utiliza para reducir la dimensionalidad de los datos. El PCA te ayuda a reorganizar todo ese caos para enfocarte solo en las combinaciones más relevantes.

En términos más técnicos, el PCA transforma tus variables originales en un conjunto más pequeño de variables llamadas componentes principales, que capturan la mayor parte de la variabilidad de tus datos. Es como condensar la información de un libro en un resumen, pero manteniendo lo esencial. (Navarro, 2022)

- k-means: método de agrupación no jerárquico, que divide un conjunto de n observaciones en k grupos (clústeres) predefinidos, donde cada observación pertenece al grupo cuya media es el más cercano.
- Analisis clúster: el análisis Clúster, conocido como análisis de Conglomerados, es una técnica estadística multivariante que busca agrupar elementos (o variables) tratando de lograr la máxima homogeneidad en cada grupo y la mayor diferencia entre los grupos. (cluster)

Coeficiente de correlación de Pearson: la correlación de Pearson evalúa la relación lineal entre dos variables continuas. Una relación es lineal

cuando un cambio en una variable se asocia con un cambio proporcional en la otra variable. (Minitab.com)

- Coeficiente de correlación de Spearman: la correlación de Spearman evalúa la relación monótona entre dos variables continuas u ordinales. En una relación monótona, las variables tienden a cambiar al mismo tiempo, pero no necesariamente a un ritmo constante. El coeficiente de correlación de Spearman se basa en los valores jerarquizados de cada variable y no en los datos sin procesar. La correlación de Spearman suele utilizarse para evaluar relaciones en las que intervienen variables ordinales.
(Minitab.com)

- Coeficiente de determinación: el coeficiente de determinación es una medida estadística que examina cómo las diferencias en una variable pueden ser explicadas por la diferencia en una segunda variable, al predecir el resultado de un evento determinado. En otras palabras, este coeficiente, que se conoce más comúnmente como R-cuadrado (o R^2), evalúa la fuerza de la relación lineal entre dos variables, y es muy utilizado por los investigadores cuando realizan análisis de tendencias. Mide la proporción de la variabilidad en la variable dependiente que se explica por las variables independientes. Cuanto más cercano a 1. mejor explica la variabilidad de la variable dependiente. (Arenas, 2018)

- F-test: el test F de Fisher permite contrastar si dos muestras presentan igualdad de varianzas (homocedasticidad). Al igual que el test de Shapiro-Wilk que sirve para contrastar la normalidad de los datos, la contrastación de la homocedasticidad es de vital importancia porque otros muchos análisis estadísticos requieren la igualdad de varianzas de los datos para poder llevarlos a cabo. (Test F de Fisher en R Commander: homocedasticidad de dos muestras., 2015)

Tenemos:

H_0 : las dos muestras presentan varianzas iguales (homocedasticidad)

h1: las dos muestras presentan varianzas distintas (heterocedasticidad)

- Test de raíz unitaria: para ver si se trabaja con las variables en diferencias en lugar de trabajar en niveles.
- Test de Augmented Dickey-Fuller (ADF): el test ADF es la herramienta estadística que se utiliza para ejecutar un test de raíz unitaria. Su resultado te dice si debes diferenciar la variable (si sale $p_valor > 0.05$) o dejarla en niveles (si sale $p_valor < 0.05$).

Tenemos:

h0: existe raíz unitaria (no es estacionaria)

h1: no existe raíz unitaria (es estacionaria)

- Test de Im–Pesaran–Shin (IPS): una prueba de raíz unitaria diseñada específicamente para datos de panel. Su función principal es determinar si las variables son estacionarias (sus propiedades estadísticas no cambian en el tiempo) o no.

Tenemos:

h0: Todas las series del panel tienen raíz unitaria (son no estacionarias).

h1: Al menos una parte de las series son estacionarias.

- Métodos de estimación:

- Modelo *pooled OLS*: estima parámetros del modelo, con el objetivo de minimizar la suma de los cuadrados de las diferencias entre los valores observados y los valores predichos por el modelo. Sencillo y sensible a outliers. Da estimaciones insesgadas. Es consistente.

Se caracteriza por: linealidad, independencia, homocedasticidad, normalidad y no multicolinealidad. No se puede aplicar cuando: existe heterocedasticidad, endogeneidad (errores correlacionados con los regresores o variable explicativas), cuando no consta normalidad y multicolinealidad (variables explicativas correlacionadas entre sí)

- Modelo de regresión: el modelo de regresión se suele utilizar con el fin de determinar si existe, o no, relación causal entre una variable dependiente (Y) y un conjunto de otras variables explicativas (X). Asimismo, el modelo busca determinar cuál será el impacto sobre la variable Y ante un cambio en las variables explicativas (X). (Roldán, 2016)
 - Modelo de efectos fijos: se refiere a variables o factores cuyos niveles son específicamente elegidos por el investigador y son de gran interés para el estudio. Sus efectos se consideran constantes y no aleatorios, por lo tanto, sus conclusiones solo son aplicables a los niveles específicos de la variable.
 - Modelo de efectos aleatorios: se refiere a variables o factores cuyos niveles son aleatoriamente seleccionados de una población de efectos que son el objeto del estudio y al investigar le interesa generalizar los efectos más allá de los niveles incluidos del estudio. Sus efectos se consideran variables aleatorias, por lo tanto, sus conclusiones se pueden generalizar al resto de niveles de las variables.
- Test F de efectos fijos: sirve para contrastar si los efectos temporales son necesarios en el modelo de panel, comparando una especificación con efectos individuales y temporales frente a otra con solo efectos individuales.
Tenemos:
h0: el modelo pooled es suficiente (no hace falta el FE)
h1: el modelo pooled no es suficiente (hace falta el FE)
 - Test de Breusch–Pagan: evalúa si la varianza del efecto individual es significativamente distinta de cero.
Tenemos:
h0: homocedasticidad (varianza constante)
h1: heterocedasticidad (varianza no constante)

- Test de Hausman: contrasta si los efectos individuales no observados están correlacionados con las variables explicativas.

Tenemos:

h_0 : modelo de efectos aleatorios es consistente y eficiente

h_1 : modelo de efectos fijos es el modelo adecuado

- Test de Wald: suele usarse especialmente para contrastar si es cero o no un determinado coeficiente que multiplica a una variable independiente en una regresión. Si el p-valor, como siempre, es menor que 0.05, se rechaza esa hipótesis nula que afirma que ese coeficiente es cero, y se entiende entonces que ese coeficiente no es cero y que, por lo tanto, el modelo es útil para representar una determinada relación. Si, por el contrario, el p-valor es mayor que 0.05 eso significa que el valor del coeficiente podría ser perfectamente cero y estar viendo lo que vemos, por lo tanto, esa variable no influye a la hora de determinar la variable dependiente (o también denominada, a veces, variable respuesta) del modelo de regresión. (Estadisticallopis, 2013)

Tenemos:

h_0 : los coeficientes son iguales a 0 ($\beta=0$)

h_1 : los coeficientes son distintos a 0 ($\beta\neq 0$)

- Test de Pesaran: es una prueba estadística fundamental en econometría para detectar la dependencia transversal (o correlación entre unidades) en modelos de datos panel. Esta prueba verifica si los residuos de un modelo están correlacionados entre sí a través de las diferentes unidades, ayudando a determinar si es necesario usar modelos de efectos fijos o aleatorios para obtener estimaciones consistentes, ya que la dependencia transversal puede sesgar los resultados. (Pesaran test.)

Tenemos:

h_0 : independencia transversal

h_1 : dependencia transversal

3. ANÁLISIS DESCRIPTIVO UNIVARIANTE

El análisis de los datos, se inicia con un análisis descriptivo univariante, que consiste en la descripción del comportamiento individual de las variables de estudio, haciendo referencia a los estadísticos descriptivos básicos y a las representaciones gráficas de las características de las variables.

3.1 ANÁLISIS DE LA TASA DE FEDERACIÓN

La tasa de federación (*tasaFed*) es nuestra variable principal del estudio. Analizaremos diferentes aspectos sobre ella:

- Distribución global
- Dimensión de la tipología del deporte
- Perspectiva de género
- Dimensión territorial
- Dinámica temporal

3.1.1 DISTRIBUCIÓN GLOBAL

La [Figura 1](#) muestra la distribución de *tasaFed* para el global de los datos, esto es, considerando todas las observaciones disponibles en las dimensiones de *Periodo*, *CCAA*, *Sexo* y *Deporte*.

Observamos que la distribución global de los datos presenta una asimetría positiva, muy marcada a la derecha, con una media de 0'217 y una mediana de 0'032 (bastante pequeña), lo que indica que pocos deportes o regiones tienen tasas altas de federación, como se puede ver a simple vista.

La mayor parte de las observaciones se concentran en valores muy bajos, cercanos a cero, representado por la barra más alta, lo que representa, al ser prácticamente la barra que destaca, es que pocos deportes presentan una tasa de federación alta, como se podría dar en el caso, de los deportes más famosos como fútbol o baloncesto.

Además, el análisis de cuartiles, puede reforzar la interpretación anterior ya que, se puede decir que el 75% de los registros (0'206) se encuentra por debajo de la media (0'206 < 0'217), lo que indica que la media está condicionada por la existencia de valores elevados y poco frecuentes sobre un número reducido de observaciones.

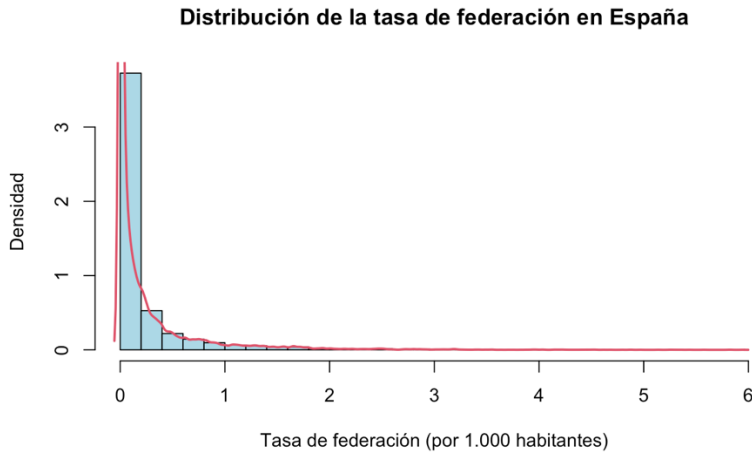


Figura 1: Distribución de la tasa de federación en España

3.1.2 DIMENSIÓN DE LA TIPOLOGÍA DEL DEPORTE

A continuación, se procede a descomponer *tasaFed* por deporte, ya que es la variable de desagregación, para ver cuáles son los que ejercen mayor peso sobre la tasa de federación, contando para cada categoría con todas las observaciones disponibles de *Periodo*, *Sexo* y *CCAA*.

El "Top 15 de deportes con más federados en España", [Figura 2](#), muestra cuales son aquellos deportes que cuentan con mayor número de licenciados. En este caso, los deportes que sobresalen frente al resto son: el Taekwondo (106.579'868 federados), el Triatlón (104.314'34 federados) y la Natación (98.079'46 federados).

Esta concentración de los licenciados en algunos deportes concretos es condicionante y la explicación de la asimetría positiva explicada en la [Figura 1](#), ya que la mayoría de las 65 modalidades restantes da lugar a registros mucho menores, con un bajo volumen de licenciados, como por ejemplo Motonáutica (4.257 federados), Pentatlón moderno (4.603 federados) y Deportes para sordos (7.983 federados), entre otros, lo que puede generar mayor variabilidad, ruido, permitiendo afirmar que la estructura federativa española está fuertemente concentrada, no existe un reparto equilibrado de la práctica federada entre modalidades, sino que unos pocas disciplinas acumulan la mayoría de federados. En particular, estas 15 disciplinas acumulan el 40% de federados.

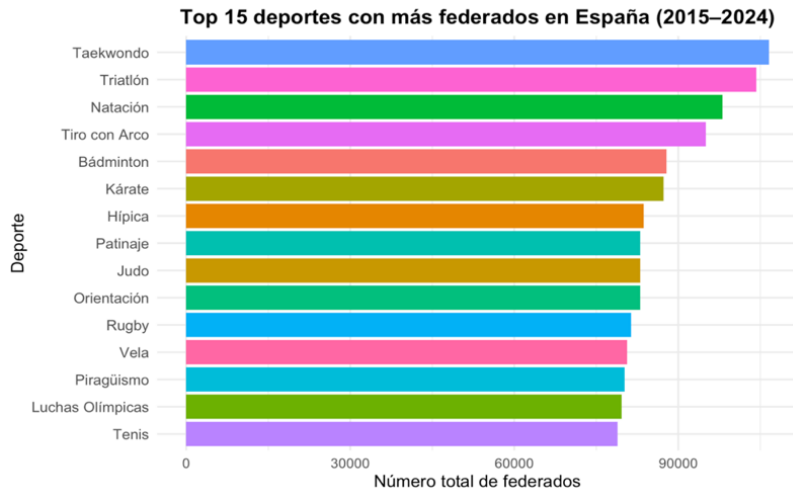


Figura 2: Top 15 deportes con mayor número de federados en España (2015–2024)

La [Figura 3](#), muestra el diagrama de cajas correspondiente al “Top 20 de deportes ordenados según la mediana de la tasa de federación” donde la longitud horizontal de la caja representa el rango intercuartílico (Q1–Q3), es decir, la amplitud que contiene el 50% central de los datos, mientras que la mediana se muestra mediante una línea vertical en el interior de la caja. Los puntos negros individuales situados fuera de los bigotes corresponden a valores atípicos, asociados a observaciones excepcionalmente altas de la tasa de federación.

En la parte superior del gráfico se sitúan disciplinas como Taekwondo y Triatlón, que presentan las medianas más elevadas, lo que indica una intensidad relativa de práctica federada sistemáticamente superior en comparación con el resto de deportes analizados, presentando además una dispersión moderada, lo que sugiere cierta estabilidad a lo largo del tiempo. Por el contrario, deportes como Tenis, Patinaje o Luchas Olímpicas destacan por la presencia de numerosos valores atípicos y un rango intercuartílico amplio, lo que indica que, existen años o comunidades autónomas concretas en las que la participación federada se incrementa de forma notable, generando picos puntuales en la distribución.

En el extremo inferior del gráfico se encuentran disciplinas como Petanca o Tenis de Mesa, caracterizadas por cajas estrechas y una mediana reducida, lo que sugiere una participación federada estable, con escasa variabilidad territorial o temporal.

Además, en muchos deportes se aprecia una asimetría positiva caracterizada por medianas reducidas y la presencia de numerosos valores extremos elevados, indicando que la posición de algunos deportes en el “top” por tasa puede estar influida por la cola derecha de la distribución, es decir, por observaciones puntuales con tasas excepcionalmente altas.

Este comportamiento es habitual cuando la práctica federada se encuentra concentrada territorialmente o cuando el tamaño es reducido, ya sea por un número limitado de federados o por poblaciones pequeñas.



Figura 3: Diagrama de cajas: Top 20 deportes por tasa de federación en España (Ordenado por Mediana)

En este apartado también se analiza cuáles son los deportes que presentan mayores valores según la variable *Nivel1* que como se explicó distingue entre las modalidades de práctica, deportes individuales frente a los que se realizan en equipo.

La [Figura 4](#) representa un gráfico de cajas el cual muestra la “Tasa de federación según el tipo de deporte (*Nivel1*)”. Se observa que existen diferencias tanto en la mediana, aunque mínimas (mediana deportes individuales = 0’04, mediana deportes equipo = 0’008), como en la variabilidad de la tasa de federación entre ambas categorías.

En los deportes individuales se observa que la caja es más alta, y la nube de puntos revela que el registro es mayor, situado en torno a 2 licencias por cada 1000 habitantes, según el grosor de dicha nube. Mientras que los deportes de equipo, aunque las modalidades son más populares (fútbol,

baloncesto), la mediana es inferior al igual que su nube de puntos (es muy cerca al cero).

Esto puede ser porque en los deportes individuales, a la hora de competir, es necesaria, mayoritariamente, la formalización federativa, al contrario que los deportes de equipo que la práctica es más común realizarla sin ningún tipo de federación de manera más recreativa e informal.

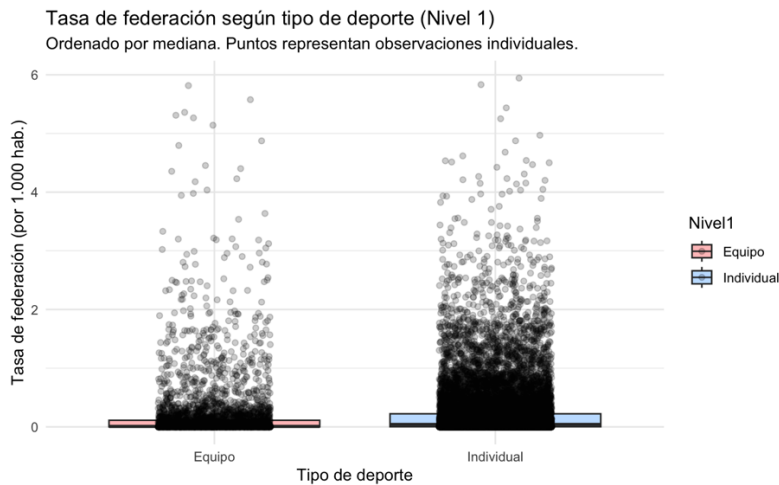


Figura 4: Gráficos caja con dispersión de la tasa de federación por tipo de deporte (equipo vs. individual)

A continuación, se hace lo mismo, pero teniendo en cuenta la variable *Nivel2*, que hace referencia a el carácter de cada deporte, es decir, la lógica funcional de cada uno de ellos.

La [Figura 5](#) representa los gráficos de cajas sobre la "Tasa de federación por categoría funcional (*Nivel2*)" ordenador según la mediana, donde la categoría que encabeza es Estrategia (0'1222), seguida de Rescate (0'1208) y Resistencia (0'1207), Combate (0'1091) y Aventura (0'0714). Además, Resistencia y Combate presentan una larga racha de valores atípicos (puntos negros) confirmando que los 3 deportes que pertenecen a estas categorías son los más destacados entre los "Top 15 de deportes con más federados en España", [Figura 2](#), que son Triatlón, Taekwondo y Natación.

Sin embargo, categorías como Fuerza, Adaptado y Combinado, se visualiza que las cajas tienen un tamaño mínimo por lo que su efecto es minoritario.

La categoría de Invasión (donde se sitúan muchos deportes de equipo tradicionales, entre ellos baloncesto, balonmano, fútbol) muestra una

mediana baja pero una cantidad elevada de valores atípicos, lo que refuerza la idea de que estos deportes dependen mucho del contexto regional o del volumen, por lo que su práctica no está generalizada en todo el territorio.

Para verificar todo lo observado, se realiza un análisis de contraste de hipótesis. Se comienza con una prueba para ver la normalidad, el test de Shapiro-Wilk, comprobando que *tasaFed* para cada uno de las categorías de *Nivel2*, presenta para todos los niveles un *p_valor* menor a 0'05 lo que lleva a rechazar la hipótesis de normalidad, llevándonos a utilizar una prueba no paramétrica, el test de Kruskal-Wallis, para comprobar si las categorías del *Nivel2* afectan por igual a *tasaFed*.

Esta prueba indica que existen diferencias significativas entre los niveles impuestos en el *Nivel2* y la variable *tasaFed* (*p_valor*<0'05, rechaza la hipótesis nula de igualdad entre grupos), es decir, al menos uno de los 16 grupos (*df*=15) tiene una distribución diferente de federados.

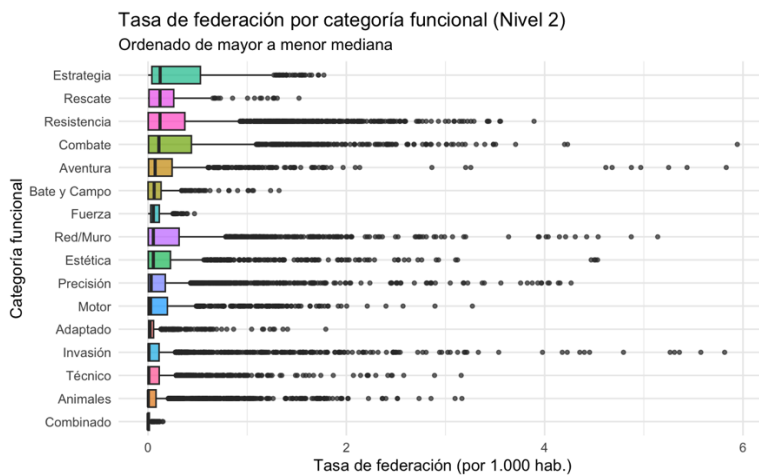


Figura 5: Gráfico de cajas: Tasa de federación por categoría funcional (Nivel 2)

3.1.3 PERSEPECTIVA DE GÉNERO

En este apartado se analiza la tasa de federación por género. Dado que el criterio de desagregación aplicado agrupa los datos por *Periodo* y *Sexo*, el valor medio de cada categoría se calcula promediando las observaciones correspondientes a las *CCAA* y *Deportes*.

En primer lugar, se realizaron diferentes test para confirmar las siguientes hipótesis. Los resultados de test de normalidad, Shapiro-Wilk, aplicados a ambos grupos (hombres y mujeres), dan como resultado un $p_valor < 0'05$ por lo que no se asume una distribución normal, por lo que habrá que optar por pruebas no paramétricas, como en este caso, el test de Wilcoxon (Mann-Whitney) para comparar los grupos de géneros.

Para este último test, se repite el resultado ($p_valor < 0'05$, rechazamos h_0) por lo tanto sí que hay diferencias significativas entre los hombres y mujeres en cuanto a la tasa de federación.

De todas formas, en la [Figura 6](#), se observa que el número de federados hombres es superior al de mujeres. Los hombres presentan niveles de federación más altos a lo largo de todo el periodo de tiempo seleccionado, sobre todo hasta 2022, donde alcanza su pico máximo cercano a 0'27 federados por cada 1000 habitantes. Además, es destacable que el impacto de la pandemia en el 2020 no afectó a este grupo, a diferencia que las mujeres, que muestran un ligero descenso entre el 2020 y 2021, aunque seguido de una recuperación inmediata.

Las mujeres mantienen durante todo el periodo seleccionado niveles inferiores, pero más constantes, entre 0'17-0'19, sin variaciones como en el caso de los hombres.

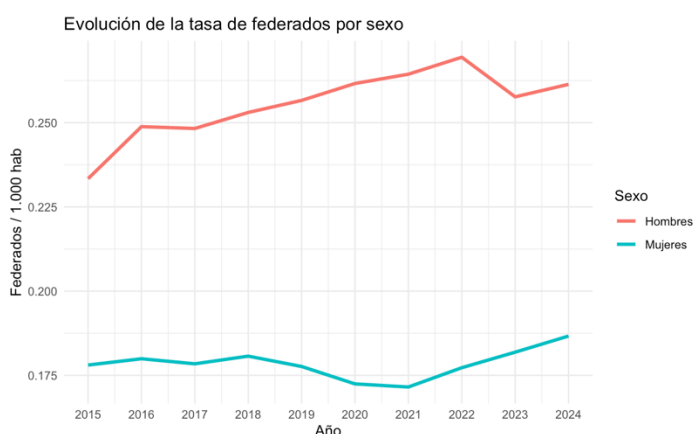


Figura 6: Evolución de la tasa de federados por sexo.

Si analizamos los deportes con mayor número de federados por cada grupo en cuanto al género, se obtienen un resultado interesante:

- Para los hombres, el deporte con mayor número de federados, con un total de 54.943'37, es Luchas Olímpicas.
- Y para las mujeres, el deporte con el mayor volumen de licenciadas, con un total de 77.456'63, es Judo.

La [Figura 7](#), muestra el "top 15 de deportes según el número total de federados", diferenciando por sexo.

En el caso de los hombres, destacan principalmente deportes como: Luchas Olímpicas, Motociclismo, Hípica o Taekwondo. En las mujeres, el ranking está encabezado por disciplinas como Judo, Taekwondo o Kárate, junto con deportes individuales como Triatlón, Tenis o Natación.

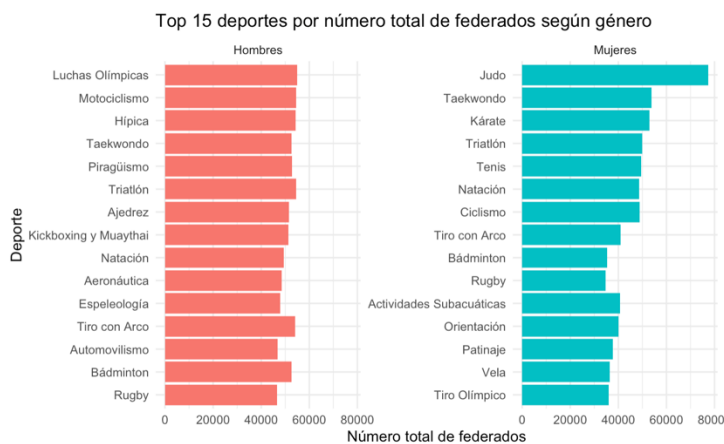


Figura 7: Top 15 deportes por número total de federados según género

Finalmente, la [Figura 8](#) muestra un gráfico de cajas, donde se observa la tasa de federación por tipo de deporte, Nivel1, y Sexo

Se puede observar que las tasas de federación son bajas en ambos tipos de deporte y sexos, con medianas aproximadamente cero. Se observa una mayor dispersión en los deportes individuales, especialmente entre los hombres. En cambio, en los deportes de equipo, hay una menor dispersión, pero destacan sus valores extremos, sobre todo en los hombres, alcanzando por ello las cifras máximas más elevadas.

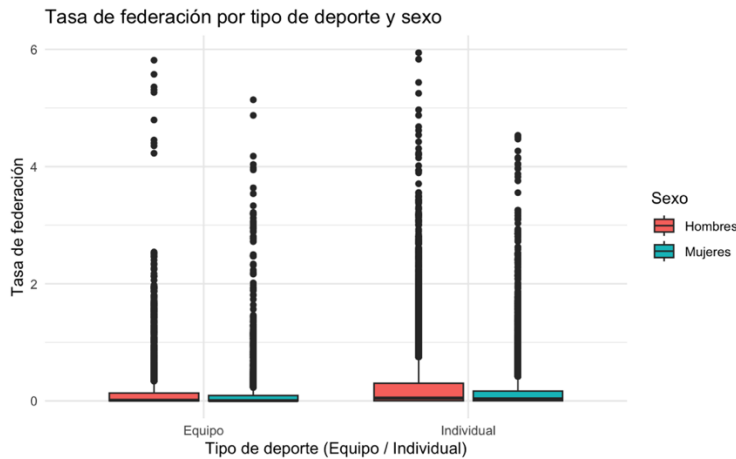


Figura 8: Tasa de federación por tipo de deporte y sexo

3.1.4 DIMENSIÓN TERRITORIAL

3.1.4.1 FEDERATIVA POR COMUNIDAD AUTÓNOMA

La [Figura 9](#) muestra el promedio de la tasa de federados por CC.AA. En concreto, para cada CC. AA. se calcula la media no ponderada de la tasa de federación a lo largo de todas las combinaciones de Periodo, Sexo y Deporte. Se puede confirmar que la tasa promedio en este ranking no depende de volumen de población, ya que podemos observar que las comunidades que cuentan con una elevada población, como por ejemplo la Comunidad de Madrid, Cataluña o Andalucía, no ocupan las primeras posiciones, al contrario, son las tres últimas.

Las CC.AA. que van en cabeza son aquellas que tienen menor censo, pero una especialización deportiva mayor, como es el caso de La Rioja (tasa de federación > 0'6), Cantabria (cerca 0'5) y Comunidad Foral de Navarra (0'4 aproximadamente), que presentan las tasas de federación promedio más altas, e incluso Principado de Asturias, Islas Baleares y Extremadura, con tasas más moderadas (entre 0'25 y 0'35). Sin embargo, en el extremo opuesto, como se indica anteriormente, se sitúa la Comunidad de Madrid, así como Cataluña y Andalucía (por debajo de 0'1).

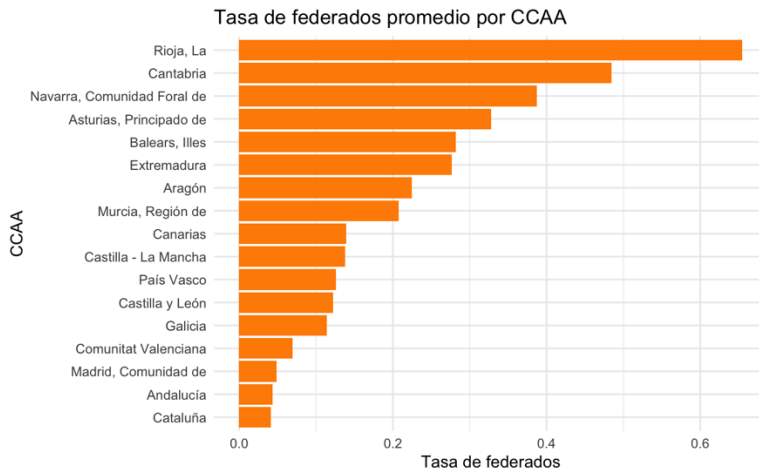
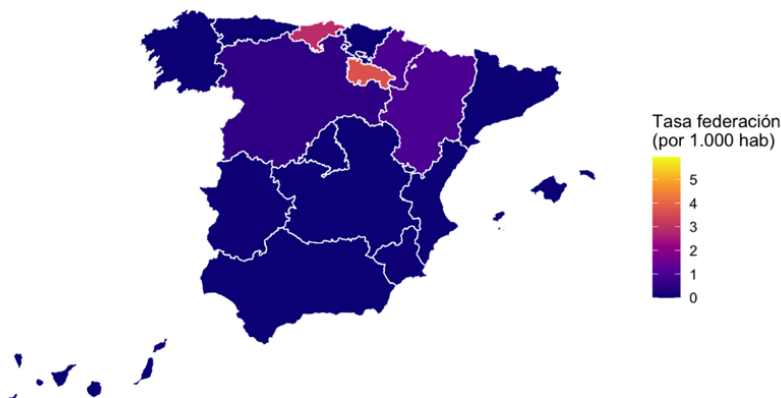


Figura 9: Tasa de federados promedios por CC.AA..

La [Figura 10](#), muestra la representación cartográfica por CC.AA. de la tasa de federación donde la intensidad del color viene determinada por la tasa de federados por cada 1.000 habitantes, confirmándose los mismos resultados que en el ranking anterior.

Las comunidades con niveles altos se concentran en el norte de país (Cantabria, la Rioja, Comunidad Foral de Navarra), representados por colores cálidos o de tono más claros, alcanzando tasas más elevadas por cada 1000 habitantes. Por el contrario, la zona sur, centro y arco mediterráneo presenta tasas bajas, representadas por colores fríos y oscuros.

Tasa de federación por CCAA en España



Fuente: elaboración propia con datos deportivos y población

Figura 10: Tasa de federación por CC.AA. en España.

3.1.4.2 INTERACCIÓN TERRITORIAL Y FEDERATIVA

La [Figura 11](#) muestra los diagramas de caja, asociados a la tasa de federación por CC.AA. y sexo, a partir de las observaciones agregadas por CCAA, *Periodo* y *Deporte*, ordenados según la mediana masculina de la tasa de federación.

Para ambos sexos, las CC.AA. que presentan una mayor tasa de federación son: La Rioja, Cantabria y Navarra, ya que son las que tienen las cajas más alargadas (presentan mayor dispersión) y medianas más elevadas. Además, en todas las CC.AA., excepto en Cataluña, la tasa de federación masculina supera a la femenina.

Cuanto más alta es la tasa de federación, más diferencia hay entre la participación entre hombres y mujeres, pero al haber tantos valores atípicos, se acentúa esa variabilidad.

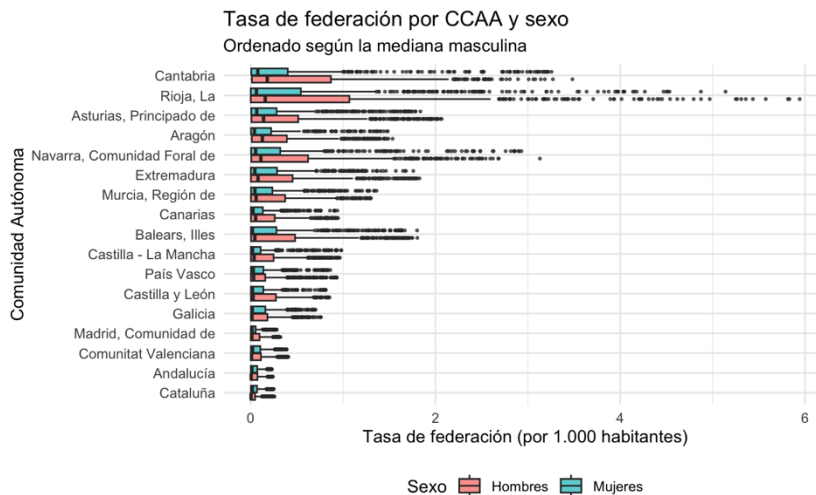


Figura 11: Gráfico de cajas: Tasa de federación por CCAA y Sexo

Por otra parte, la participación federada por CC.AA. varía según la tipología del deporte, cómo se practica, *Nivel1*.

Para cada CCAA y *Nivel1* (Individual o Equipo), se calcula la tasa media de federación, promediando las observaciones correspondientes a las dimensiones de *Periodo*, *Sexo* y *Deportes*. Finalmente, se obtiene la diferencia mediante la resta aritmética entre la tasa media Individual y la tasa media de Equipo, lo que permite determinar qué modalidad de práctica predomina en cada región.

Este comportamiento se ve en casos como Comunidad Foral de Navarra (0'204), Extremadura (0'188) y Castilla-La Mancha (0'116), seguidas por La Rioja (0'088) y Castilla y León (0'074).

Sin embargo, otras regiones presentan tasas medias relativamente mayores en deportes por equipo (diferencia negativa) como en el caso de Cantabria (-0.196), el más extremo en España, y de forma más moderada en Murcia (-0.025), la Comunidad Valenciana (-0.008) y Cataluña (-0.010). Este resultado podría estar asociado a una mayor implantación del deporte de equipo, por ejemplo, a través de una red de clubes densa.

Finalmente, CC.AA. como el Principado de Asturias, Galicia, la Comunidad de Madrid o Andalucía presentan diferencias cercanas a cero, lo que refleja un sistema deportivo equilibrado donde ambas tipologías de deporte existen con pesos similares.

3.1.5 DINÁMICA TEMPORAL

3.1.5.1 EVOLUCIÓN HISTÓRICA E HITOS

La [Figura 12](#), muestra la evolución del número de personas federadas entre el 2015 y el 2024, donde el total de federados se calcula sumando, para cada *Periodo*, el número de federados por *CCAA*, *Sexo* y *Deporte*.

Se observa una tendencia creciente a lo largo de todo el periodo.

Desde el 2015 se produce un ascenso hasta el 2018 que se estabiliza.

En el año 2020, marcado por la crisis sanitaria del COVID-19, no se observa una caída inmediata, sino que la tendencia se mantiene e incluso crece hasta alcanzar un pico en 2021. Este fenómeno, contrario a lo esperado dadas las restricciones y la suspensión de competiciones, sugiere que la tramitación de licencias se mantuvo por inercia administrativa o se renovaron antes del confinamiento.

Sin embargo, el impacto negativo parece venir más tarde, visualizándose un descenso entre 2021 y 2023. Tras este periodo, la situación se recupera con fuerza en el último año, disparándose la cifra en 2024, donde se alcanza el máximo de la serie. En resumen, la práctica deportiva federada en España ha resistido la crisis y retoma ahora un camino de fuerte crecimiento.

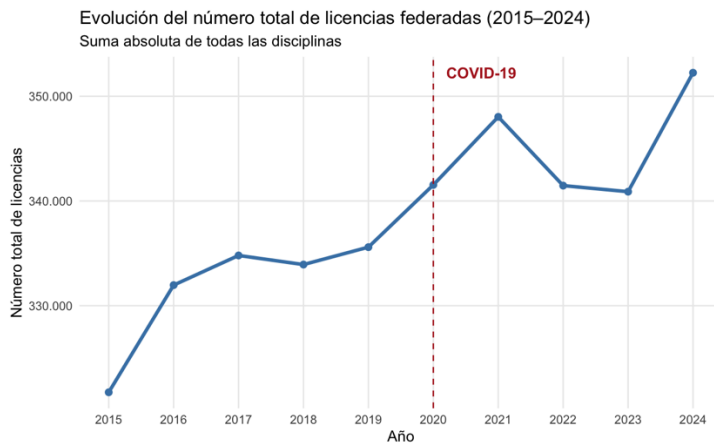


Figura 12: Evolución del número total de federados (2015-2024)

3.1.5.2 TENDENCIAS POR NIVELES DE DEPORTE

La [Figura 13](#) muestra la evolución temporal de la media de la tasa de federación según *Nivel1*. Para cada *Periodo* y *Nivel1*, dicha media se calcula como el promedio de las tasas de todas las observaciones correspondientes a las dimensiones *CCAA*, *Sexo* y *Deportes*. De este modo, se integra en un único valor anual el comportamiento de todas las disciplinas que integran la categoría correspondiente (ya sea Individual o de Equipo), neutralizando las diferencias territoriales.

Se observa que hay un crecimiento constante a lo largo del tiempo en el deporte individual, sin que se viera afectado en ningún momento por la crisis sanitaria del COVID-19, ya que era más fácil practicarlo sin mantener contacto con otras personas.

Mientras que los deportes de equipo hay una clara caída desde 2016, aunque se observan recuperaciones puntuales en 2019, 2022. Esto se sucede justo pre y postpandemia cuando había restricciones en el contacto físico, y la mayoría de las instalaciones permanecían cerradas y por eso los deportes individuales destacaron frente al resto. En 2024 se observa un nuevo repunte.

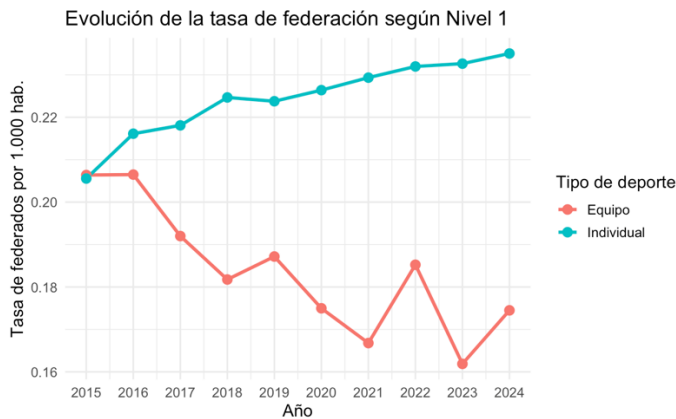


Figura 13: Evolución de la tasa de federación según Nivel 1

La [Figura 14](#) representa la evolución temporal de la media de la tasa de federación según *Nivel 2*, esto es, la variable que clasifica a los deportes en función de su categoría funcional.

Para cada *Periodo Y Nivel 2* se calcula el promedio de la tasa de federación correspondiente al conjunto de deportes que componen cada categoría integrando en dicho cálculo las observaciones de las dimensiones CCAA y Sexo para obtener un valor representativo global.

Observamos que la mayoría de las categorías tienen tendencias crecientes. Mientras que categorías como Combate o Estrategia lideran con tendencias crecientes y sostenidas en el tiempo, alcanzando las tasas más elevadas en 2024, el caso de Aventura presenta un patrón opuesto.

Esta categoría experimentó un crecimiento notable hasta 2020, pero sufrió un descenso brusco y continuado inmediatamente después de la pandemia (2021-2022), probablemente vinculado a las restricciones de movilidad y acceso a espacios naturales. La categoría no ha logrado recuperar sus niveles previos.

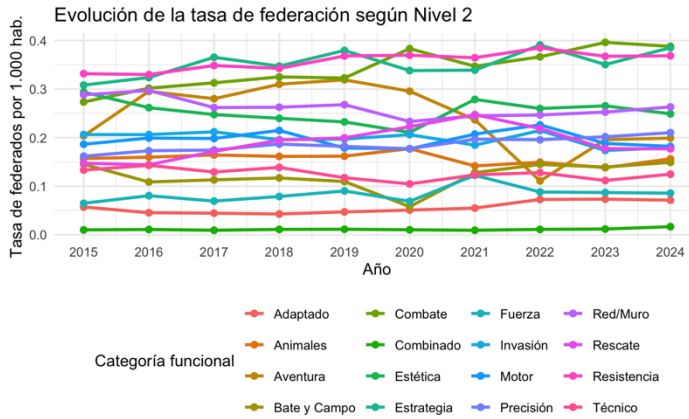


Figura 14: Evolución temporal de la tasa de federación según Nivel2

3.2 ANÁLISIS DEL EMPLEO DEPORTIVO

En este apartado se analizan los estadísticos (media y mediana) de la variable *pesoEmp* en función del género. Dado que el criterio de desagregación aplicado agrupa los datos por Sexo, los valores correspondientes se calculan procesando todo el conjunto de observaciones de *Periodo*, *CCAA* y *Deportes*.

Si analizamos el empleo deportivo según el género, en particular, la *pesoEmp* por cada 1000 empleados, se obtiene que la media de empleo masculino se sitúa en 33'94 (con una mediana de 31'54) y el empleo femenino alcanza una media de 100.71 (mediana de 100.32). Esta diferencia implica que la *pesoEmp* de las mujeres en el sector deportivo triplica a la de los hombres.

La [Figura 15](#) muestra el diagrama de cajas de la *pesoEmp* por sexo. La caja de las mujeres está situada por encima de la de los hombres la mediana de las mujeres es superior a la de los hombres, y estas presentan una mayor cantidad de valores extremos (bigotes son más largos, pero dentro del rango) lo que quiere decir que su distribución es más dispersa. La distribución de los hombres es más compacta y se concentra en valores bajos, aunque presenta algunos valores atípicos que muestran que en algunas zonas el empleo en el sector masculino por parte de los hombres es superior.

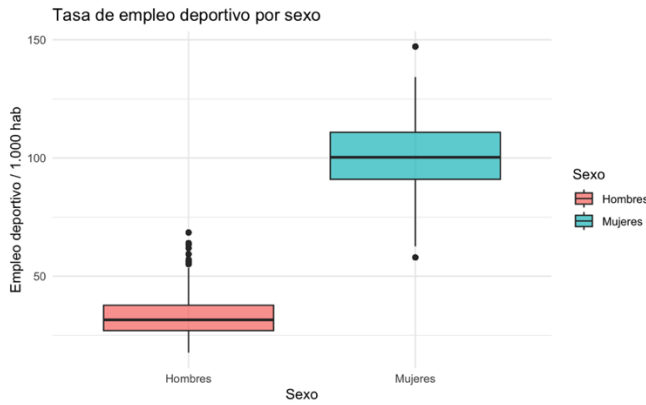


Figura 15: Diagrama de cajas: peso de empleo deportivo por sexo.

La [Figura 16](#) muestra la evolución temporal del empleo deportivo desde el 2015 al 2024. Para cada *Periodo*, la media de *pesoEmp* se calcula promediando los valores correspondientes al total de observaciones registradas en dicho año, integrando las dimensiones *CCAA*, *Sexo* y *Deportes*. De este modo, se obtiene un indicador anual agregado que sintetiza el comportamiento del empleo en todo el territorio nacional, independientemente de la región o deporte.

A diferencia de la *tasaFed*, el *pesoEmp* nacional ha sufrido una tendencia descendente entre el 2015 y 2022. La serie comenzó en 2015 con valores cercanos a los 73 empleos en el sector deportivo por cada 1.000 personas empleadas hasta situarse por debajo de los 63 empleados en el deporte por cada 1.000 empleados en 2022, a partir del 2022 podemos ver que experimenta una ligera recuperación, llegando a 65 empleados en el sector deportivo en 2023, coincidiendo con el periodo postpandemia. Tras este periodo, el empleo vuelve a caer en picado.

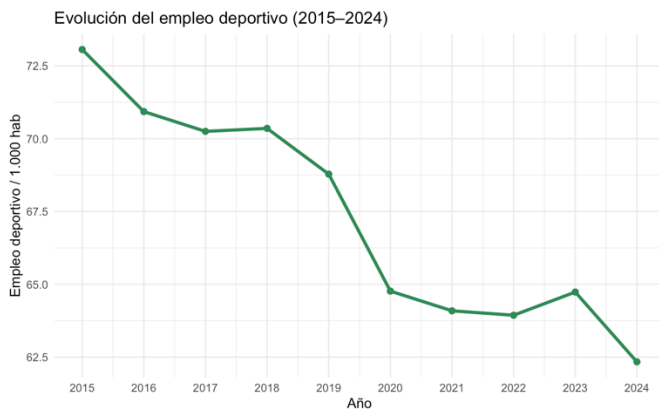


Figura 16: Evolución del empleo deportivo (2015-2024)

La [Figura 17](#) muestra la evolución del empleo deportivo por CC.AA.

Para cada Periodo y CCAA, la media de *pesoEmp* se calcula promediando los valores de *pesoEmp* entre el total de observaciones registradas, considerando las dimensiones *Sexo* y *Deportes*. De este modo, se obtiene la evolución específica de cada territorio, independientemente de los deportes o el género.

Observamos que, aunque la evolución no es uniforme en todas las CC.AA., sí se aprecia una tendencia general descendente alineada con el comportamiento nacional, manifestándose con distinta intensidad según la región.

Destaca negativamente el caso de La Rioja, que registra las cifras de *pesoEmp* más bajas, llegando a descender por debajo de los 40 empleos en el sector deportivo por cada 1.000 empleados en el año 2022.

Por el contrario, las Islas Baleares y la Comunidad de Madrid se sitúan en el extremo opuesto, liderando la serie con los valores de *pesoEmp* más elevados, esto es porque son CC.AA. que tienen una alta oferta y demanda de empleo; sobre todo el archipiélago canario, que, durante los periodos vacacionales, en temporada alta, tiene una alta demanda de empleo en actividades deportivas.

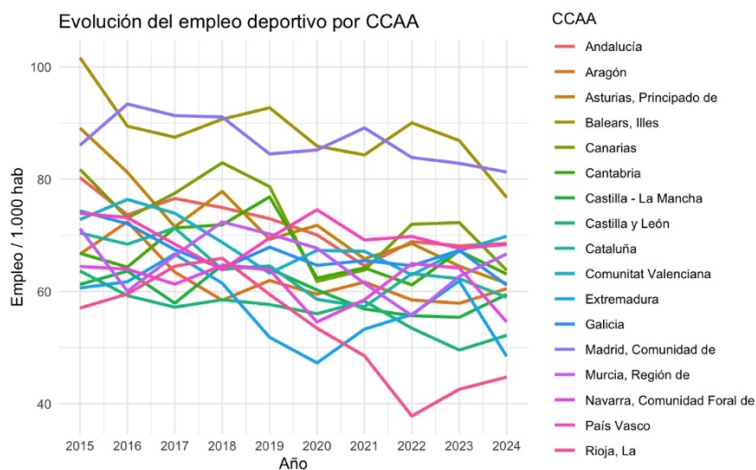


Figura 17: Evolución del empleo deportivo por CC.AA.

3.3 ANÁLISIS DEL PIB PER CÁPITA

La [Figura 18](#), representa la evolución temporal de *totalPerCapita* desglosada por CC.AA.. Para cada *Periodo* y *CCAA*, la media de *totalPerCapita* se calcula promediando los valores correspondientes al total de observaciones registradas, considerando las dimensiones *Sexo* y *Deportes*. De este modo, se obtiene la trayectoria económica específica de cada territorio a lo largo del tiempo.

Se observa que, tras un periodo de estabilidad, del 2015 al 2019, el año 2020 marca un desplome generalizado hacia valores negativos, seguido de una recuperación que dispara las cifras en 2021 y 2022, para finalmente moderarse hacia la estabilización en el último año.

A nivel territorial, destacan las Islas Baleares y Canarias, que registraron tanto la caída más profunda en 2020 como la recuperación más intensa en el 2022.

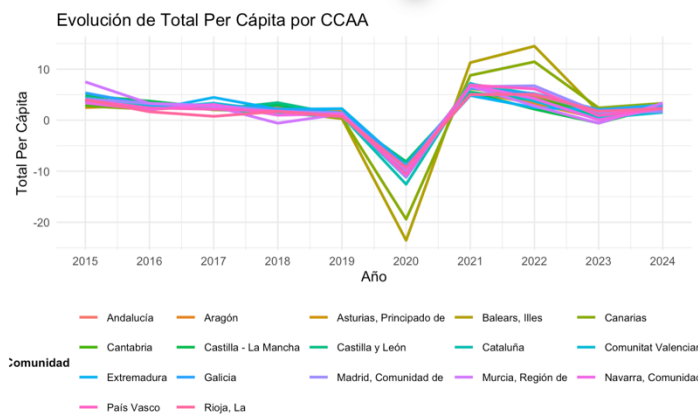


Figura 18: Evolución de *totalPerCapita* por CC.AA.

Además, en el pico más bajo de toda la serie, en 2020, se puede ver que llega a valores negativos, por lo que, para evitar errores, se optó por emplear la transformación del seno hiperbólico Inverso (*IHS sus siglas en inglés*), que da lugar a la variable transformada de *totalPerCapita* denominada *ih_s_PIB*, definida como: $asinh(x) = e^{x + \sqrt{x^2 + 1}}$, que permite trabajar con dichos valores y presenta un comportamiento similar al logaritmo para valores elevados, como muestra la [Figura 19](#), ya que la transformación logarítmica no es adecuada si hay valores negativos porque genera datos ausentes.

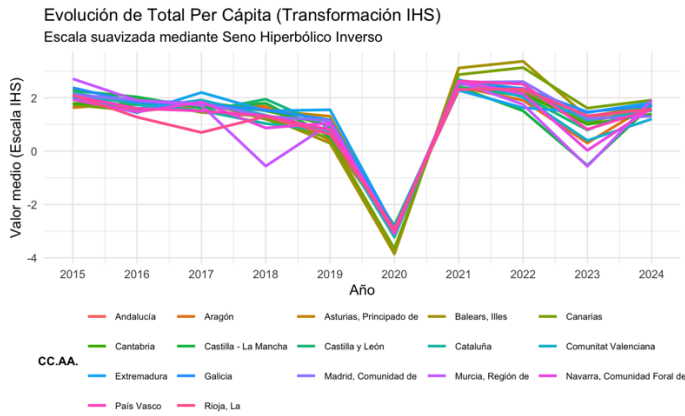


Figura 19: Evolución de totalPerCapita transformada

4. ANÁLISIS DESCRIPTIVO MULTIVARIANTE

Tras el análisis univariante —en el que se ha estudiado la variable principal, la *tasaFed*, junto con *pesoEmp* y *ihs_PIB*—, se aborda un análisis multivariante para examinar conjuntamente la relación entre estas tres variables. Este enfoque permitirá identificar asociaciones y patrones mediante un análisis de componentes principales y técnicas de clúster, y, posteriormente, cuantificar dichas relaciones a través de una estimación con datos de panel.

4.1 ANÁLISIS DE COMPONENTE PRINCIPALES

El objetivo del análisis de componentes principales (PCA por sus siglas en inglés) es identificar patrones globales y la relación entre las variables *tasaFed*, *pesoEmp* y *ihs_PIB*.

Para ello, se ha transformado la base de datos original agregando la información por CC.AA.. Específicamente, se han calculado los valores medios de cada variable para el periodo analizado, obteniendo así un perfil representativo de cada región. Finalmente, las variables fueron estandarizadas (media 0 y varianza 1) para evitar que las diferencias de escala distorsionen los resultados.

CCAA	tasaFed	pesoEmp	ihs_PIB
1 Andalucía	0.04382338	71.86057	1.2627914
2 Aragón	0.22435003	62.09525	1.0686206
3 Asturias, Principado de	0.32779490	72.10355	1.2706443
4 Balears, Illes	0.28199315	88.59699	1.2673785
5 Canarias	0.13910624	72.75978	1.2631040
6 Cantabria	0.48477802	66.92443	1.2044583
7 Castilla – La Mancha	0.13776483	59.91592	1.0591351
8 Castilla y León	0.12250444	56.57674	1.2423463
9 Cataluña	0.04145604	63.88907	1.1679149
10 Comunitat Valenciana	0.06969498	68.96602	1.0972793
11 Extremadura	0.27695050	56.90311	1.3017789
12 Galicia	0.11379666	66.90090	1.3855729
13 Madrid, Comunidad de	0.04865814	86.87524	1.2894626
14 Murcia, Región de	0.20743949	65.44407	0.9527183
15 Navarra, Comunidad Foral de	0.38728210	61.49683	1.0507694
16 País Vasco	0.12591006	69.84511	1.1776858
17 Rioja, La	0.65474074	53.35028	1.0631323

El objetivo es explorar cómo se agrupan las diferentes comunidades autónomas según su desarrollo deportivo y económico a través del análisis de las variables *tasaFed*, *pesoEmp* y *ihs_PIB*.

4.1.1 CONTRIBUCIÓN LAS DE VARIABLES Y DEFINICIÓN DE DIMENSIONES

Para ver los pesos de los distintos bloques, es decir, los diferentes grupos de CC.AA. según su desarrollo deportivo y económico, se realiza una tabla de contribuciones sobre las componentes principales. Estos resultados mostrarán posteriormente los ejes donde se clasificará cada CC.AA. dependiendo de sus características.

Las tres dimensiones, componentes principales, son las siguientes:

- PC1 (dimensión 1) – Eje de desarrollo deportivo-económico. Recoge de manera bastante equilibrada la estructura de los diferentes datos, ya que las contribuciones tienen valores parecidos (*tasaFed* 28%, *pesoEmp* 38%, *ihs_PIB* 34%), por eso el eje se puede interpretar como un eje general del desarrollo deportivo-económico.
- PC2 (dimensión 2) – Eje de participación federada. Esto es así porque está claramente dominado por la tasa de federación, ya que su

contribución es alta (*tasaFed* 67%), mientras que el resto de variables analizadas tienen una contribución reducida.

- PC3 (dimensión 3) – Eje económico-laboral. Viene explicado prácticamente por el peso del empleo deportivo (*pesoEmp* 58%) y en menor medida por el nivel económico regional.

4.1.2 REPRESENTACIÓN GRÁFICA

Además, se cuenta con una serie de gráficos que sirven de apoyo para reforzar el análisis.

En primer lugar, el gráfico de la varianza explicada, en la [Figura 20](#), donde el eje Y explica la varianza explicada y el eje X indica cada componente. Este gráfico no muestra ni relaciones entre variables ni diferencias entre las CC.AA.. La primera componente predomina frente al resto, ya que explica un 55'9%, mientras que la segunda y tercera componentes explican un 24'9% y un 19'2%, respectivamente. Tanto a partir de la línea como de las barras, observamos una caída pronunciada entre la primera y la segunda componente. Estas dos primeras componentes recogen un 80% de la varianza explicada por lo que se podría optar por retener estas dos y no tener en cuenta la tercera, pero la tercera recoge información específica relacionada con el eje económico-laboral por lo que se consideraron las tres componentes principales en el análisis.

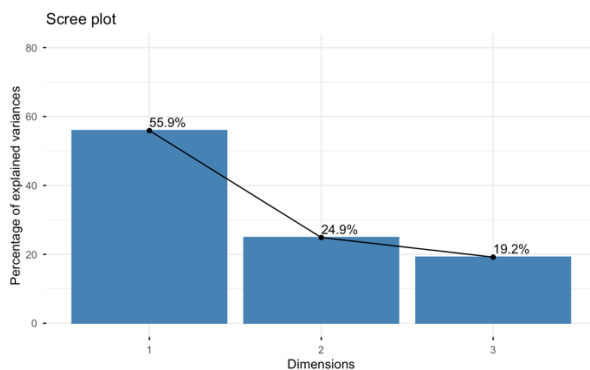


Figura 20 Scree plot: Dimensiones de componentes principales

En segundo lugar, el mapa conjunto de variables y CC.AA., [Figura 21](#), que muestra las variables y las CC.AA. en un mismo plano. Cuenta con dos ejes, que representan las dos primeras dimensiones, que explican un 80% de la varianza total, y diferentes flechas que se atribuyen a las tres variables

consideradas. La dirección de la flecha indica la relación con las componentes principales, la longitud indica la contribución y su ángulo la correlación entre variables. Además, se aprecian puntos que son las CC.AA. que, dependiendo de a que distancia estén unas de otras, tendrán comportamientos parecidos (cerca) o diferentes (lejos).

Como se observa, *pesoEmp* y *ihs_PIB* están positivamente asociadas con la Dimensión 1 (eje deportivo-económico), ya que apuntan hacia la derecha, mientras que *tasaFed* apunta hacia la izquierda.

Las CC.AA. que están situadas a la derecha del eje vertical (dimensión 1) son aquellas que presentan mayor *ihs_PIB* y mayor peso del empleo deportivo, como es el caso de las Islas Baleares (punto 4), ya que se sitúa en lo más alto de la trayectoria de la flecha de *pesoEmp* y *ihs_PIB*. Esto puede ser porque tiene una industria turística y de ocio inmensa, lo que podría disparar el empleo deportivo por encima de la tasa de federación. La Comunidad de Madrid (punto 13) también sería otra CC.AA. donde podría considerarse que hay una mayor *ihs_PIB*.

Por el contrario, las CC.AA. situadas a la izquierda del eje vertical son aquellas que presentan mayor intensidad en la tasa de federación, como es el caso de La Rioja (punto 17), seguida de la Comunidad Foral de Navarra (punto 15) y Cantabria (punto 6).

Por otra parte, las CC.AA. más alejadas del origen muestran perfiles más específicos, frente a aquellas próximas al centro, que presentan valores intermedios en las variables consideradas.

Las agrupaciones observadas en el gráfico sugieren la existencia de grupos similares de CC.AA., lo que justifica la aplicación del análisis clúster posterior.

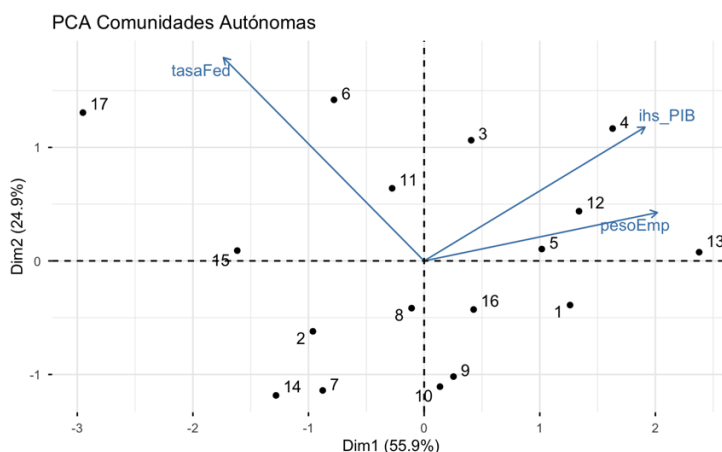


Figura 21: Mapa conjunto de variables y PCA

Y, por último, la [Figura 22](#) corresponde al círculo de las correlaciones del PCA, que permite obtener las relaciones entre las variables y componentes. Al igual que la [Figura 21](#), cuenta con los dos ejes que se refieren a las dos primeras dimensiones, y la circunferencia representa la máxima correlación posible entre las variables originales y las componentes principales.

En este caso, se puede observar que las flechas son muy próximas a la circunferencia por lo que están bien representadas. Las correlaciones de las variables son iguales que las de la [Figura 21](#), es decir, *pesoEmp* y *ihs_PIB* tiene una relación positiva con la primera componente debido a la dirección creciente de la flecha, mientras que *tasaFed*, al contrario.

Por otra parte, la segunda componente está asociada principalmente a la tasa de federación, mientras que el *ihs_PIB* y el *pesoEmp* presentan una contribución muy reducida en esta dimensión.

De este modo, la dimensión 1 puede entenderse como un eje que separa el desarrollo económico y la relevancia del empleo deportivo de la intensidad de la práctica federada.

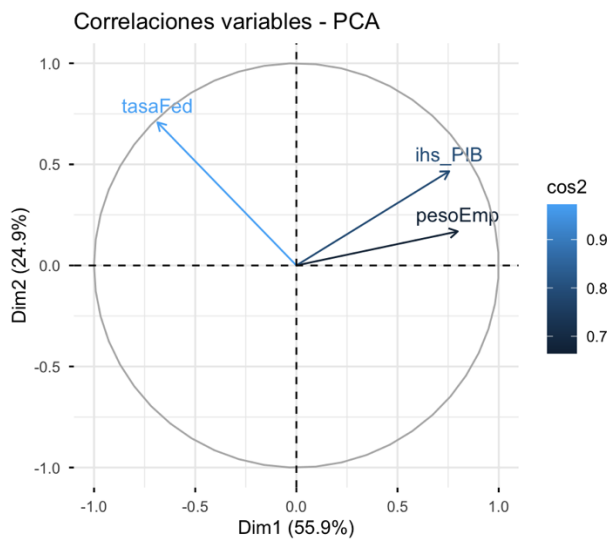


Figura 22: Círculo de las correlaciones de PCA

4.2 ANÁLISIS CLUSTER

El análisis de clúster se aplica con el objetivo de identificar grupos de CC.AA. en función de la similitud en cuanto a su perfil deportivo, económico y laboral,

facilitando la identificación de distintos modelos regionales de práctica deportiva.

El análisis se llevó a cabo utilizando las tres variables más representativas de las dimensiones identificadas previamente en el PCA: *tasaFed*, *pesoEmp* y *ihs_PIB*. Al igual que en el análisis de PCA, en la base de datos agruparon los registros por CCAA y se calculó la media aritmética de cada variable para el *Periodo* analizado, eliminando los valores ausentes. Además, fueron estandarizadas para evitar diferencias en la escala.

Se empleó un método de agrupación no jerárquico (k-means), fijando el número de clústeres en cuatro grupos. Para garantizar la estabilidad de los resultados se fijó una semilla aleatoria.

CCAA	tasaFed	pesoEmp	ihs_PIB	cluster
1 Andalucía	0.04382338	71.86057	1.2627914	4
2 Aragón	0.22435003	62.09525	1.0686206	3
3 Asturias, Principado de	0.32779490	72.10355	1.2706443	4
4 Balears, Illes	0.28199315	88.59699	1.2673785	2
5 Canarias	0.13910624	72.75978	1.2631040	4
6 Cantabria	0.48477802	66.92443	1.2044583	1
7 Castilla - La Mancha	0.13776483	59.91592	1.0591351	3
8 Castilla y León	0.12250444	56.57674	1.2423463	4
9 Cataluña	0.04145604	63.88907	1.1679149	4
10 Comunitat Valenciana	0.06969498	68.96602	1.0972793	3
11 Extremadura	0.27695050	56.90311	1.3017789	4
12 Galicia	0.11379666	66.90090	1.3855729	4
13 Madrid, Comunidad de	0.04865814	86.87524	1.2894626	2
14 Murcia, Región de	0.20743949	65.44407	0.9527183	3
15 Navarra, Comunidad Foral de	0.38728210	61.49683	1.0507694	1
16 País Vasco	0.12591006	69.84511	1.1776858	4
17 Rioja, La	0.65474074	53.35028	1.0631323	1

4.2.1 IDENTIFICACIÓN DE PERFILES REGIONALES

A partir del número de clústeres que se fijaron anteriormente se cuenta con:

- Clúster 1: agrupa CC.AA. con una participación federada alta, un nivel económico intermedio y un peso del empleo reducido, comparado con otros clústeres. Tienden a ser CC.AA. con una fuerte cultura de clubes donde la práctica federada es elevada sin la necesidad de un sector deportivo laboral fuerte. (Cantabria, Comunidad Foral de Navarra y La Rioja)

- Clúster 2: reúne las CC.AA. con el más alto nivel económico, con mucho empleo vinculado al deporte y una tasa de federación baja comparando con el primer clúster. Esto sugiere un modelo deportivo en donde la actividad económica vinculada al deporte tiene un papel esencial. (Comunidad de Madrid e Islas Baleares)
- Clúster 3: contiene las CC.AA. que cuentan con niveles *ihs_PIB* más bajos y con un empleo deportivo bastante desarrollado, sin embargo, la tasa de federación es escasa, lo que sugiere un modelo deportivo orientado a un consumo recreativo y no tanto a clubes. (Castilla – La Mancha, Aragón, Comunidad Valenciana y Murcia)
- Clúster 4: recoge a CC.AA. cuya tasa de federación es baja, pero con un peso del empleo deportivo y un nivel económico elevado. Esto se suele ver en CC.AA. en donde este fuerte componente económico está vinculado al turismo, los servicios o actividades recreativas sin necesidad de licencias federativas. (Andalucía, Canarias, Principado de Asturias, Castilla y León, Cataluña, Extremadura, Galicia y País Vasco)

Esta información, además se muestra en la [Figura 23](#), un gráfico de barras agrupado por cada una de las variables.

En conjunto, el análisis de clúster confirma la existencia de distintos modelos regionales de práctica deportiva en España. Los resultados demuestran que el *ihs_PIB* y el *pesoEmp* no explican por sí solos la *tasaFed*, y que coexisten modelos regionales de práctica deportiva con lógicas diferentes: federativa, económica y recreativa. Estos patrones refuerzan los resultados obtenidos en el PCA y aportan una visión de las diferentes asociaciones deportivas en España.

cluster	tasaFed	pesoEmp	ihs_PIB
1	0.5089336	60.59051	1.106120
2	0.1653256	87.73611	1.278421
3	0.1598123	64.10531	1.044438
4	0.1489178	66.35485	1.258980

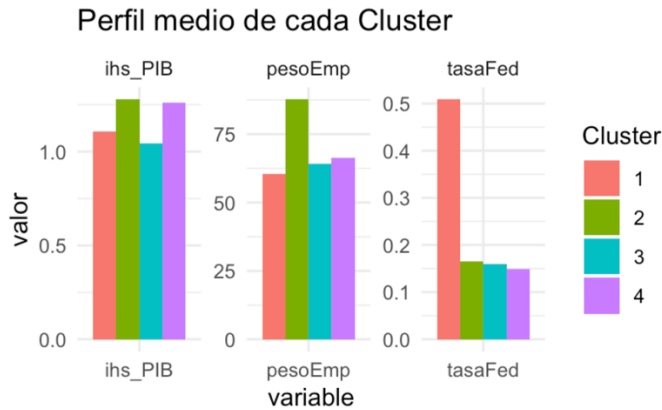


Figura 23: Perfil medio de cada Clúster

4.3 INCORPORACIÓN DE GÉNERO

4.3.1 BRECHA DE GÉNERO EN LA TASA DE FEDERACIÓN DEPORTIVA

4.3.1.1 ANÁLISIS DE LA DESIGUALDAD POR COMUNIDADES AUTÓNOMAS

Para ver cómo de equilibrado es el sistema deportivo entre CC.AA. en función del *Sexo*, el análisis se desarrolló en dos etapas secuenciales.

En primer lugar, se generó la brecha de género, una base de datos desagregada agrupando los registros por *CCAA*, *Periodo* y *Deportes*. En esta fase, se calculó la tasa media de federación diferenciada para hombres (*tasaH*) y mujeres (*tasaM*) y se definió la brecha de género como la diferencia aritmética entre ambas ($tasaH - tasaM$), con el fin de captar las diferencias específicas entre disciplinas y años.

Sin embargo, dado que este nivel de desagregación puede generar una alta variabilidad, especialmente en deportes en los cuales la participación no es muy alta, se procedió a sintetizar la información.

Para ello, el análisis final se centra en medidas agregadas, calculadas agrupando únicamente por CCAA y obteniendo el promedio de la brecha para todo el periodo. Este procedimiento permite ofrece una visión más estable de la desigualdad de género en la práctica federada.

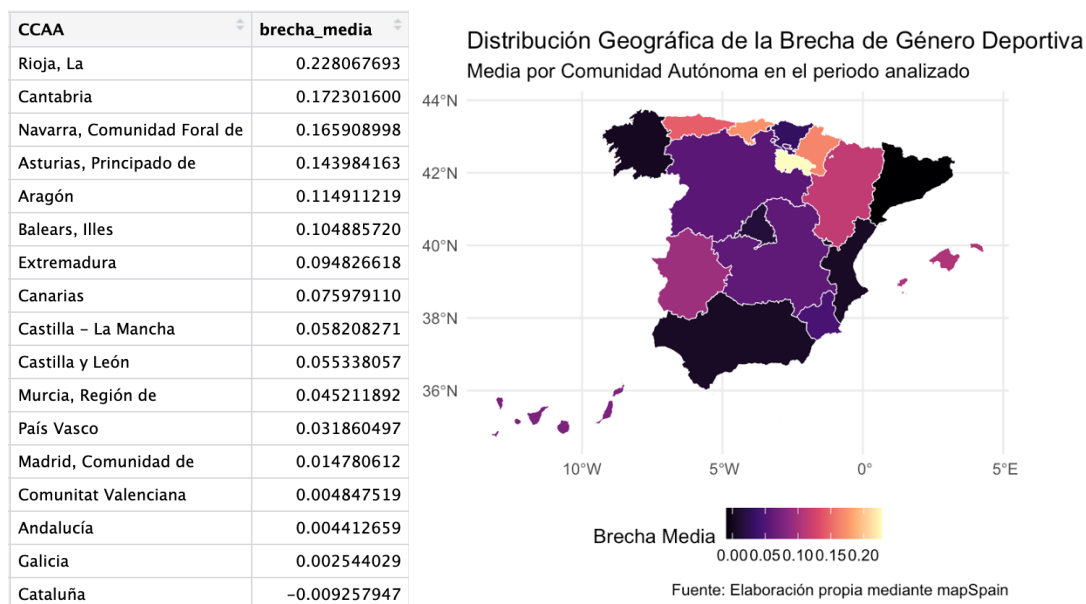


Figura 24: Distribución geográfica de la brecha de género deportiva.

El análisis por CC.AA. revela que las regiones del norte de España presentan las mayores desigualdades: La Rioja encabeza la brecha, la tasa media de federación masculina supera en 0'23 puntos a la femenina, seguida de Cantabria (0'17) y el bloque de la Comunidad Foral de Navarra, el Principado de Asturias y Aragón (0'11–0'16). En cambio, las Islas Canarias y Castilla y León (0'05–0'07) registran las diferencias más bajas.

Esta distribución queda reflejada visualmente en el mapa, [Figura 24](#), mediante una escala de colores que identifica los distintos niveles de desigualdad: las zonas claras (amarillo y naranja), representan los valores más altos de la brecha, en cuanto a las zonas oscuras (morado y negro), identifican a las regiones con menores diferencias, donde la participación federativa es más equilibrada y los tonos rosados el resto, representa los valores medios de la brecha.

El patrón sugiere una mayor desigualdad de género en el norte, coherente con que, en análisis previos, [Figura 11](#), estas CC.AA. presentan tasa de

federación más elevadas. Mientras que en algunos territorios del sur y las Islas Canarias presentan una menor brecha, posiblemente por mayor diversidad de deportes y peso del turismo deportivo en el territorio insular, si bien esta interpretación debe tomarse como hipótesis y requeriría análisis adicional para confirmarse.

4.3.1.2 ANÁLISIS DE LA DESIGUALDA TEMPORAL

A continuación, con el fin de analizar la evolución temporal de la brecha de género en la práctica federada, se calculó también la brecha media anual.

Para ello, se agruparon los registros únicamente por *Periodo* y se obtuvo el promedio de la brecha para cada año. Este procedimiento implica agregar la información de todas las *CC.AA.* y *Deportes* en un único valor anual, lo que permite observar la tendencia general de la desigualdad a lo largo del tiempo, como se puede observar en el [Figura 25](#).

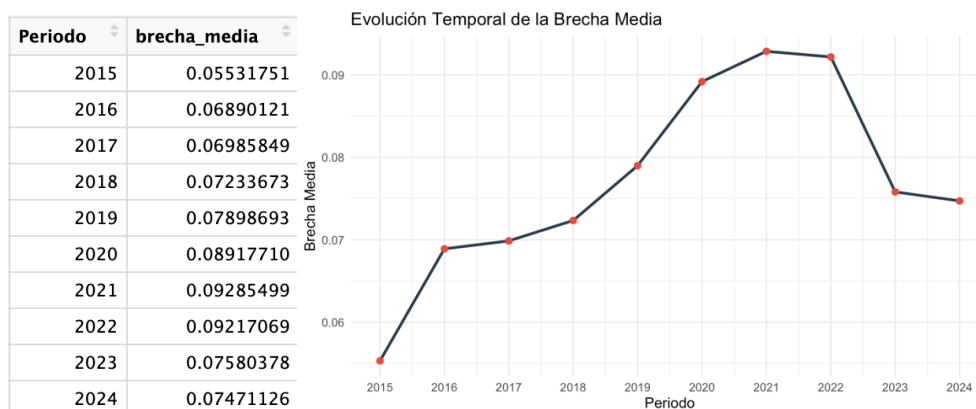


Figura 25: Evolución temporal de la Brecha Media de Género

Al inicio de periodo de estudio, en 2015, la brecha de género no contaba con valores muy altos (0'0553). Con el paso de los años, el valor sube hasta llegar a su punto máximo en 2021 (0'0929), donde comienza a caer suavemente hasta el final del periodo en 2024 (0'0747).

4.3.1.3 ANÁLISIS DE DESIGUALDAS POR CLÚSTER

En cuanto a los modelos regionales, que tienen como unidad de observación los clústeres identificados en la sección [4.2.1](#), se calculó el promedio de la brecha media de las CC.AA. que integran cada grupo. Aquel que tienen una mayor desigualdad de género en la práctica federada es Clúster 1 (0'19), mientras que los tres restantes tiene una participación más equilibrada al contar con una brecha media más suave (0'05 aproximadamente).

cluster	brecha_media_cluster
1	0.18875943
2	0.05983317
3	0.05579473
4	0.04996090

4.3.2 RELACIÓN ENTRE LA ACTIVIDAD ECONÓMICA Y PARTICIPACIÓN DEPORTIVA POR SEXO

El PCA relaciona *ihs_PIB*, *pesoEmp* y *tasaFed* sin distinguir hombres de mujeres. El análisis clúster agrupa CC.AA. por sus perfiles en base a esas tres variables, pero no por sexo.

Este apartado explora si cambia la relación actividad económica - participación deportiva según el sexo. Y Siguiendo el enfoque usado en PCA, se decidió trabajar con *ihs_PIB* para evitar trabajar con valores negativos en el *totalPerCapita*. En este proceso, se desagregaron los datos originales agrupando por *CCAA*, *Periodo* y *Sexo*, calculando nuevamente los promedios de la *tasaFed* y el *ihs_PIB* para cada grupo.

La [Figura 26](#), muestra un diagrama de dispersión que relaciona *ihs_PIB* (eje X) con la tasa de federación (eje Y, por 1.000 habitantes), desagregando por sexo (puntos rojos: hombres; puntos turquesa: mujeres). Las líneas representan la tendencia lineal para cada sexo. Se observando que la relación observada sigue siendo débil y las pendientes de las rectas de regresión para hombres y mujeres son prácticamente nulas lo que indica que no existe una correlación línea clara, el nivel de *ihs_PIB* no determina la *tasaFed*, además de ser paralelas lo que confirma que la brecha de género es un problema que no se corrige con el crecimiento económico, incluso en niveles altos

de *ihs_PIB*, la tasa media de federación femenina se sigue manteniendo por debajo de la masculina.

Asimismo, el análisis de la nube de puntos revela una elevada dispersión de los datos, evidenciando una gran heterogeneidad entre las CC.AA. Especialmente en los tramos de mayor *ihs_PIB*, encontramos puntos situados en extremos opuestos del gráfico, lo que demuestra que regiones con la misma riqueza pueden tener niveles de práctica deportiva muy dispares. Por último, se aprecia un 'techo' visible en la participación femenina: mientras que los puntos rojos (hombres) alcanzan valores máximos superiores a 0,8, la nube de puntos turquesa (mujeres) rara vez supera el umbral de 0,5, independientemente del desarrollo económico de la región.

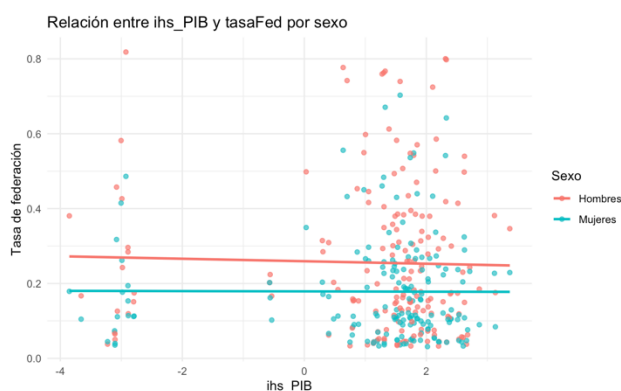


Figura 26: Gráfico de dispersión: Relación entre *ihs_PIB* y *tasaFed* por sexo

Esta débil relación entre la tasa de federación y el *ihs_PIB* se corrobora mediante los coeficientes de correlación.

Tanto la correlación de Pearson (hombres: -0.026 ; mujeres: -0.004) como la de Spearman (hombres: -0.037 ; mujeres: -0.037) arrojan valores muy próximos a cero. En consecuencia, no se observa evidencia de una asociación relevante entre *ihs_PIB* y *tasaFed*, para ninguno de los dos sexos, en el conjunto de observaciones analizadas. Se reportan Pearson y Spearman para contrastar, respectivamente, asociaciones lineales y monótonas y evaluar la robustez frente a posibles valores extremos.

Que los valores sean tan cercanos a cero implica una independencia estadística, lo que significa que la variabilidad del *ihs_PIB* no explica la

variabilidad de la *tasaFed*. En la práctica, esto nos dice que el hecho de que una CC.AA. sea más rica no implica que sus ciudadanos compitan más oficialmente, es decir, que se federen; es probable que en las zonas ricas el deporte se canalice hacia el ocio privado (gimnasios) en lugar de hacia la licencia federativa.

5. ESTIMACIÓN Y DISCUSIÓN DE RESULTADOS

Hasta ahora, el análisis se ha centrado en el estudio de la participación deportiva desde una perspectiva descriptiva y multivariante, mediante el uso de diferentes técnicas como el Análisis de Componentes Principales y el análisis de clúster.

En este apartado se analizan los determinantes de la tasa de federación mediante modelos de regresión en dos enfoques complementarios. En primer lugar, se estima una regresión de sección cruzada, utilizando observaciones agregadas por CCAA, con el fin de evaluar las asociaciones entre la *tasaFed* y las variables explicativas en un corte temporal representativo. En segundo lugar, se emplean modelos con datos de panel, que explotan simultáneamente la variación entre CC.AA. y a lo largo del tiempo, permitiendo controlar la heterogeneidad no observada específica de cada territorio y obtener estimaciones más informativas sobre la relación entre variables.

5.1 MODELO DE REGRESIÓN LINEAL

Con el fin de analizar conjuntamente el efecto del *ihs_PIB* y el *Sexo* sobre la *tasaFed*, y para dar soporte al resultado que muestra la [Figura 26](#), se estimó un modelo de regresión lineal utilizando la base de datos de la sección [4.3.2](#), con el objetivo de ver si el efecto de *ihs_PIB* es distinto para hombres y mujeres.

Para ver que estas variables no son estacionarias se hace un el test ADF, el cuál obtiene p_valores mayores a 0'05, por lo que se acepta la hipótesis nula

de no estacionariedad, raíz unitaria, por lo tanto, hay que trabajar con las variables en diferencias.

$$\Delta tasaFed_i = \beta_0 + \beta_1 * \Delta ihs_{PIB_i} + \beta_2 * Sexo_i + \beta_3 * (\Delta ihs_{PIB_i} * Sexo_i) + \varepsilon_i \quad (1)$$

donde i denota la CCAA. $\Delta tasaFed_i$ es la *tasaFed* diferenciada, es decir, la variación interanual media de la tasa de federación en el periodo t para el $Sexo_i$. Δihs_{PIB_i} es *ihs_PIB* diferenciada y $Sexo_i$ es una variable indicadora que toma el valor 1 para mujeres y 0 para hombres. El término ε_i el error idiosincrático.

La tasa de federación se explica en función del *ihs_PIB*, el *Sexo* y la interacción entre ambas variables.

En primer lugar, el coeficiente asociado a la variación del *ihs_PIB* es positivo (0,00012), pero no estadísticamente significativo (p-valor = 0,888 > 0,05). Este resultado indica que los cambios interanuales en el *ihs_PIB* no tienen un efecto significativo sobre las variaciones en la tasa de federación, una vez eliminadas las tendencias de las series mediante el uso de primeras diferencias.

En cuanto a las diferencias por *Sexo*, el coeficiente correspondiente a la categoría de mujeres es negativo (-0,00215), aunque no resulta estadísticamente significativo (p-valor = 0,463 > 0,05). Esto implica que no se observan diferencias sistemáticas en la evolución interanual de la tasa de federación entre hombres y mujeres, una vez controlados los cambios en el *ihs_PIB*. En este contexto dinámico, no puede afirmarse la existencia de una brecha de género significativa en términos de variaciones anuales de la participación federada.

Por su parte, el término de interacción entre la variación del *ihs_PIB* y el *Sexo* tampoco resulta estadísticamente significativo (p-valor = 0,937 > 0,05). Este hecho sugiere que el efecto de los cambios en el *ihs_PIB* sobre la tasa de federación no difiere entre hombres y mujeres. En otras palabras, la

relación entre las variaciones del *ihs_PIB* y los cambios en la tasa de federación es similar para ambos sexos y, en términos generales, no significativa. Esta conclusión es coherente comparando con el anterior gráfico, la [Figura 26](#), en la que las rectas de regresión para hombres y mujeres presentan pendientes prácticamente paralelas.

En relación con el ajuste del modelo, el coeficiente de determinación ($R^2 = 0,045$) indica que el modelo explica una proporción muy reducida de la variabilidad observada en las variaciones de la tasa de federación. Asimismo, el contraste global del modelo (F-test) no resulta estadísticamente significativo ($p\text{-valor} = 0,881 > 0,05$), lo que sugiere que el conjunto de variables incluidas no aporta una capacidad explicativa relevante sobre los cambios interanuales de la tasa de federación. Este resultado es habitual en el análisis de fenómenos sociales y deportivos cuando se analizan dinámicas de corto plazo, en las que intervienen múltiples factores no observados.

En conjunto, los resultados del modelo de regresión en primeras diferencias confirman que el *ihs_PIB* no ejerce un efecto estadísticamente significativo sobre los cambios en *tasaFed*, ni en hombres ni en mujeres.

Sin embargo, se identifica una brecha de género clara y persistente, con tasas de federación inferiores en mujeres en todos los niveles *ihs_PIB*. Asimismo, la relación entre *Sexo* y *ihs_PIB* no es significativa indicando que la relación entre *ihs_PIB* y *tasaFed* es similar para ambos sexos y, en general, débil.

5.2 ANÁLISIS DE DATOS PANEL

El análisis panel nos permite captar variaciones tanto entre las CC.AA. como a lo largo del tiempo, además de controlar la heterogeneidad no observada en los diferentes territorios.

Los datos construyen un panel corto, ya que, tiene muchas unidades ($N=17$) y pocos periodos ($T=10$), además de ser un panel equilibrado porque todas las CC.AA. en todos los años cuentan con información.

Previamente a cualquier análisis, se estudia el comportamiento de las tres variables principales: *tasaFed*, *pesoEmp* y *ihs_PIB*. Los gráficos de cada uno de ellas sugieren no estacionariedad, ya que crecen y decrecen en el tiempo, por lo que se va a realizar un test de raíz unitaria para ver si se trabaja con las variables en diferencias en lugar de trabajar en niveles.

Esto se puede observar para *pesoEmp* en la [Figura 16](#), para *ihs_PIB* en la [Figura 18](#) y *tasaFed* en la [Figura 12](#).

Dado que el criterio de desagregación aplicado agrupa los datos por CCAA y *Periodo*, los valores se calculan promediando las diferencias observadas en el conjunto de datos de las tres variables principales. Adjuntamos las 3 primeras y últimas observaciones en este caso.

CCAA	Periodo	tasaFed	pesoEmp	ihs_PIB
Andalucía	2015	0.03829424	80.28505	2.0192607
Andalucía	2016	0.03865747	73.63750	1.5702785
Andalucía	2017	0.04226659	76.57361	1.6692870

Rioja, La	2021	0.67094989	48.52993	2.3143976
Rioja, La	2022	0.71985003	37.79625	2.3280081
Rioja, La	2023	0.71894068	42.56428	1.3258978
Rioja, La	2024	0.72124656	44.73541	1.5702785

5.2.1 ESTACIONARIEDAD DE LAS VARIABLES

Antes de la estimación de los distintos modelos, se realizaron pruebas de raíz unitaria para determinar si las variables debían tratarse en niveles o en diferencias.

Se empleó el test de Im–Pesaran–Shin (IPS) como contraste principal y el de Fisher-ADF (Maddala–Wu) como prueba de robustez.

<p>Im-Pesaran-Shin Unit-Root Test (ex. var.: Individual Intercepts)</p> <p>data: ihs_PIB ~ 1 Wtbar = -5.1907, p-value = 1.048e-07 alternative hypothesis: stationarity</p>	<p>Im-Pesaran-Shin Unit-Root Test (ex. var.: Individual Intercepts)</p> <p>data: tasaFed ~ 1 Wtbar = -4.1339, p-value = 1.783e-05 alternative hypothesis: stationarity</p>
<p>Maddala-Wu Unit-Root Test (ex. var.: Individual Intercepts)</p> <p>data: ihs_PIB ~ 1 chisq = 105.66, df = 34, p-value = 2.845e-09 alternative hypothesis: stationarity</p>	<p>Maddala-Wu Unit-Root Test (ex. var.: Individual Intercepts)</p> <p>data: tasaFed ~ 1 chisq = 123.87, df = 34, p-value = 3.789e-12 alternative hypothesis: stationarity</p>

Im-Pesaran-Shin Unit-Root Test (ex. var.: Individual Intercepts)

data: pesoEmp ~ 1
 Wtbar = -3.5591, p-value = 0.0001861
 alternative hypothesis: stationarity

Maddala-Wu Unit-Root Test (ex. var.: Individual Intercepts)

data: pesoEmp ~ 1
 chisq = 106.21, df = 34, p-value = 2.34e-09
 alternative hypothesis: stationarity

Ambos test permitieron rechazar la hipótesis nula de raíz unitaria, confirmando que las series son estacionarias en niveles. Esto valida el uso de las variables originales en los modelos de panel, garantizando que los resultados obtenidos no requieren el uso de diferencias.

5.2.2 MODELO BASE

El modelo base de datos de panel es:

$$tasaFed_{it} = \alpha + \beta_1 * ihs_{PIB_{it}} + \beta_2 * pesoEmp_{it} + \varepsilon_{it} \quad (2)$$

donde i denota la CC.AA. y t el periodo. α es el término constante, $tasaFed_{it}$ es la variable dependiente $tasaFed$, $ihs_{PIB_{it}}$ es la variable independiente que recoge ihs_PIB y $pesoEmp_{it}$ representa el $pesoEmp$. El término ε_{it} el error idiosincrático.

Como primer paso se estima un modelo *pooled OLS*, el cual trata la información de todas las CC.AA. y los distintos años de manera conjunta, sin introducir rasgos específicos individuales ni temporales, lo que permite hacerse una idea inicial acerca de la relación entre la $tasaFed$, ihs_PIB y el $pesoEmp$.

Los resultados de este modelo muestran que el coeficiente asociado al $pesoEmp$ presenta un signo negativo y resulta estadísticamente significativo, al 0'1%, pero tiene relación con la tasa de federación (p_valor aproximadamente 0). Este resultado sugiere que un mayor peso del empleo deportivo no se traduce necesariamente en una mayor participación federada, sino que podría estar vinculado a modelos deportivos más orientados al ocio o incluso a actividades recreativas no federadas.

El intercepto es positivo y significativo ($p_valor = 1.855e-10$), no obstante, su interpretación económica es limitada, ya que representa el nivel de la tasa de federación sin tener en cuenta los efectos de las variables explicativas. Por otro lado, la variable *ih*s_PIB no resultó ser estadísticamente significativa ($p_valor = 0.9258$), lo que indica que, bajo este modelo, el *totalPerCapita* no explica variaciones en la tasa de federados.

En cuanto al ajuste del modelo, el coeficiente de determinación indica que el modelo explica un porcentaje moderado de la variabilidad total de *tasaFed*, en torno al 9% ($R^2=0.0908$).

Sin embargo, este el modelo asume que todas las CC.AA. comparten un comportamiento homogéneo por lo que necesario contrastar los resultados con otros modelos como el de efectos fijos y de efectos aleatorios, que se analizan a continuación.

5.2.3 MODELO EFECTOS FIJOS VERSUS EFECTOS ALEATORIOS

Previamente, con el fin de evaluar la necesidad de incorporar efectos temporales al modelo de panel (de cada *Periodo*), se realizó un contraste F comparando un modelo de efectos fijos bidimensionales, es decir, efectos individuales (de cada CC.AA.) y temporales, con un modelo que incluye únicamente efectos fijos individuales.

Como resultado se obtiene un p_valor mayor a 0.05 ($p_valor=0.95$), aceptando la hipótesis nula, por tanto, los efectos temporales no significativos.

Primero estimamos un modelo de efectos fijos. En este modelo cada CC.AA. se va a comparar con ella misma a lo largo del tiempo, por lo que va a eliminar las diferencias que existen entre ellas.

Los resultados muestran que el coeficiente asociado al *pesoEmp* es negativo (coeficiente= -0.00098), pero resulta estadísticamente significativo ($p_valor=0.0044 < 0.05$), indicando que, dentro de una misma CC.AA., un aumento del

pesoEmp a lo largo del tiempo presenta una relación inversa relevante con la *tasaFed*. No obstante, el signo negativo sigue sugiriendo que el crecimiento del empleo en el sector podría estar más vinculado a actividades recreativas, de ocio o turísticas que no requieren necesariamente una afiliación federada.

Por su parte, el coeficiente de la variable *ihs_PIB* presenta un signo positivo (coeficiente = 0,00056), aunque no alcanza la significatividad estadística (p-valor = 0,6297 > 0,05). Este resultado sugiere que, si bien existe una tendencia a que el crecimiento económico regional se relacione positivamente con la federación deportiva, dicho impacto no es lo suficientemente robusto como para ser considerado un determinante significativo.

En cuanto al ajuste del modelo, el coeficiente de determinación indica que el modelo explica aproximadamente un 5'2% de la variabilidad de la *tasaFed*. El contraste global del modelo resulta significativo (p-valor= 0'0171 < 0'05), lo que sugiere que lo que sugiere que las variables introducidas poseen, en conjunto, capacidad explicativa sobre la variable dependiente.

A continuación, se estima el modelo de efectos aleatorios. Este modelo mantiene la heterogeneidad propia de cada CC.AA., pero aquellas características que no se observan en los datos no se tiene en cuenta que están relacionadas con las variables explicativas.

Los resultados muestran, al igual que en el modelo de efectos fijos, que el coeficiente asociado al *pesoEmp* es negativo (coeficiente = -0'0010) y estadísticamente significativo al 5 % (p_valor= 0'0028 < 0'05), indicando que a mayor peso del empleo deportivo se asocia una menor tasa de federación, una vez controladas las diferencias estructurales entre CC.AA. y las variaciones temporales, reforzando el resultado anterior de que el crecimiento del *pesoEmp* no implica necesariamente una mayor *tasaFed*, sino que puede estar vinculado a otros factores.

Por su parte, el *ihs_PIB* no resulta estadísticamente significativo en esta especificación ($p\text{-valor} = 0'6242$), lo que confirma que el *ihs_PIB* no ejerce un efecto directo y significativo sobre la tasa de federación en este contexto.

Este resultado es coherente con los obtenidos en el análisis univariante, el PCA y los modelos de regresión de la sección [6.1](#).

Respecto al ajuste del modelo, el coeficiente de determinación indica que el modelo explica aproximadamente un 5'1 % de la variabilidad total de la tasa de federación. En este caso, el contraste global del modelo resulta estadísticamente significativo ($p\text{-valor} = 0,0115 < 0,05$), lo que indica que, bajo la especificación de efectos aleatorios, el conjunto de variables explicativas logra aportar información relevante para explicar la tasaFed, a diferencia de lo observado en especificaciones previas menos robustas.

5.2.4 ELECCIÓN DEL MODELO

Una vez estimadas los diferentes modelos de panel, se procede a seleccionar el modelo más adecuado mediante los contrastes estadísticos adecuados para cada uno de ellos. En particular, se van a emplear el test F de efectos fijos, el test de Breusch-Pagan y el test de Hausman.

5.2.4.1 CONTRASTE ENTRE EL MODELO POOLED OLS Y EL MODELO DE EFECTOS FIJOS

En primer lugar, se contrasta el modelo pooled OLS frente a un modelo de efectos fijos mediante un test F.

Este contraste evalúa si los efectos individuales y temporales son conjuntamente significativos, donde los resultados muestran que hay que rechazar la hipótesis nula de ausencia de efectos fijos, el modelo pooled no sería suficiente ($p\text{-valor} = 0 \text{ aprox.} < 0'05$), lo que indica la presencia de heterogeneidad no observada significativa entre CC.AA. y a lo largo del tiempo, es mejor el modelo de efecto fijos, controla esta heterogeneidad.

5.2.4.2 CONTRASTE ENTRE EL MODELO POOLED OLS Y EL MODELO DE EFECTOS ALEATORIOS

Para confirmar el resultado anterior, se aplica complementariamente el test de Breusch-Pagan para contrastar el modelo pooled OLS y el modelo de efectos aleatorios, con el objetivo de evaluar si la varianza del efecto individual es significativamente distinta de cero.

Como resultado sale un p_valor que es aproximadamente cero, por lo tanto, permite rechazar la hipótesis nula de no haber efectos individuales, confirmando la existencia de heterogeneidad individual y permitiendo concluir que en este caso el modelo pooled tampoco es adecuado.

5.2.4.3 CONTRASTE ENTRE EL MODELO DE EFECTOS FIJOS Y EL MODELO DE EFECTOS ALEATORIOS

Una vez se ha descartado el modelo pooled OLS, los modelos que se van a considerar para la elección final van a ser el modelo de efectos fijos y efectos aleatorios.

Para esta elección, se emplea el test de Hausman, que contrasta si los efectos individuales no observados están correlacionados con las variables explicativas.

El resultado de la prueba no permite rechazar la hipótesis nula ($p_valor = 0,66 > 0,05$), lo que indica que no existe evidencia de correlación entre los efectos individuales y las variables explicativas, es decir, el modelo de efectos aleatorios es eficiente y consistente, es preferible al de efectos fijos.

Tras la ejecución de las diversas pruebas, se selecciona el modelo de efectos aleatorios (RE por sus siglas en inglés) para el análisis de la tasa de federación.

5.3.5 MODELO DE EFECTOS ALEATORIOS CORRELACIONADOS

Con el fin de reforzar la elección del modelo de efectos aleatorios se estimó un modelo de efectos aleatorios correlacionados (CRE por sus siglas en inglés) ya que permite comprobar la posible correlación entre los efectos individuales no observados y las variables explicativas incorporando sus medias temporales.

Los resultados indican que el coeficiente asociado al *pesoEmp* tiene signo negativo (coeficiente=-0'00098) y además se puede rechazar la hipótesis nula de ausencia de efecto de dicha variable ($p_valor=0'00389 < 0'05$).

Esto indica que, una vez controladas las diferencias entre CC.AA. en los diferentes años, un mayor *pesoEmp* se asocia con una menor *tasaFed*, lo que es coherente, volviendo a los resultados obtenidos con anterioridad tanto en el modelo de efectos fijos como en el modelo de efectos aleatorios.

Por su parte, la media temporal del *pesoEmp* no resulta estadísticamente significativa ($p\text{-valor} = 0,436 > 0'05$) lo que supone que no se puede demostrar que las diferencias entre CC.AA. en el *pesoEmp* estén correlacionadas con los efectos individuales no observados.

En cuanto al *ihs_PIB*, no muestra significación estadística ($p\text{-valor} = 0'629$), lo que confirma que el nivel económico regional no influye significativamente sobre la tasa de federación deportiva en esta especificación.

Por otra parte, el modelo explica aproximadamente un 5'96 % de la variabilidad total de la tasa de federación, coincidiendo con el obtenido en el modelo de efectos aleatorios, además, el contraste global del modelo resulta estadísticamente significativo ($p\text{-valor} = 0,033 < 0'05$), por lo que las variables explicativas, consideradas conjuntamente, aportan información relevante al modelo.

También se realiza el test de Wald, que compara el modelo CRE y las medias temporales de las variables explicativas, cuyo resultado muestra que no se rechaza la hipótesis nula ($p_valor=0'4608 > 0'05$), por tanto, las medias temporales no son significativas, no hay evidencias de correlación entre los efectos individuales no observados y las variables explicativas.

Esto a su vez se puede traducir como que el modelo de efectos aleatorios es el adecuado.

Finalmente, el enfoque CRE confirma que el modelo de efectos aleatorios es el correcto, al no detectarse correlación entre los efectos individuales no

observados y las variables explicativas, permitiendo así analizar de forma adecuada la relación entre la *tasaFed*, *pesoEmp* y *ihs_PIB*.

5.3.6 VALIDACIÓN DEL MODELO

Por último, se realizaron diferentes test en el modelo de efectos aleatorios.

5.3.6.1 TEST DE BREUSCH-PAGAN

El test de Breusch-Pagan se rechaza la hipótesis nula ($p_valor=0'0086 > 0'05$), por lo que la varianza de los errores no es constante, existe heterocedasticidad, el grado del error del modelo no es similar en todas las observaciones.

5.3.6.2 TEST DE PESARAN

Sin embargo, en el test de dependencia cruzada de Pesaran, no se rechaza la hipótesis nula de independencia transversal ($p_valor=0'3618 > 0'05$), por tanto, no existen evidencias estadísticas de dependencia cruzada entre las comunidades autónomas. Es decir, los errores del modelo de efectos aleatorios no parecen estar correlacionados entre regiones, lo que ocurre en una CC.AA.: es completamente independiente a lo que ocurre en las otras.

Dado el resultado anterior, se procedió a corregir los errores estándar del modelo utilizando una matriz de varianzas-covarianzas robusta, garantizando así la fiabilidad de la inferencia estadística ante posibles problemas de heterocedasticidad o autocorrelación, sin afectar a los estimadores del modelo.

Una vez corregidos, los resultados muestran una reducción generalizada de la significación estadística de algunos coeficientes respecto a la estimación original. El coeficiente asociado al *pesoEmp* mantiene un signo negativo (coeficiente=-0'0010), aunque deja de ser estadísticamente significativo ($p_valor=0'2803 > 0'05$).

Por su parte, el *ihs_PIB* tampoco muestra significación estadística (p-valor = 0'9044 > 0'05), reforzando la idea de que el nivel económico regional no influye sobre la tasa de federación deportiva en este modelo.

En resumen, el uso de errores robustos garantiza una inferencia estadística fiable sin modificar los coeficientes estimados del modelo, aunque en este caso conlleva que *ihs_PIB* y *pesoEmp* pierdan su capacidad explicativa individual bajo este nivel de exigencia estadística.

6. CONCLUSIONES

Este estudio analiza la evolución y los determinantes de la *tasaFed* en España, contrastándola con *pesoEmp* y *ihs_PIB*, mayoritariamente, además del resto variables como *Nivel1*, *Nivel2* y *Sexo*, a través de un enfoque descriptivo, multivariante y de modelos de datos de panel.

Concluye que la estructura federativa española es altamente concentrada y asimétrica. Un número reducido de deportes (como Taekwondo, Triatlón y Natación) concentra la mayor parte de las licencias, mientras que la mayoría de disciplinas presenta tasas muy bajas.

Además, predominan los deportes individuales sobre los de equipo; estos últimos, aunque populares, tienden a una práctica más recreativa e informal fuera del ámbito federativo.

Temporalmente, el sistema ha mostrado resiliencia frente al COVID-19, alcanzando su máximo histórico en 2024.

Territorialmente, el Norte (La Rioja, Cantabria, Comunidad Foral de Navarra), presenta las tasas de federación más altas. Se caracteriza por una "cultura de clubes" tradicional, donde la práctica deportiva se formaliza a través de licencias, a pesar de tener un peso de empleo deportivo menor. Sin embargo, grandes ciudades y zonas turísticas (Comunidad de Madrid, Islas Baleares, Cataluña), muestran tasas de federación bajas, pero lideran en el *pesoEmp* y *ihs_PIB*. Esto sugiere un modelo orientado al consumo de servicios deportivos (gimnasios, turismo activo, ocio) que no requiere federarse.

Por otro lado, la brecha de género es estructural. La tasa media de federación masculina supera sistemáticamente a la femenina en casi todas las CC.AA y años. Los análisis de correlación y regresión confirman que el crecimiento económico no corrige esta desigualdad: incluso en las regiones con mayor *ihs_PIB*, la participación federada deportiva femenina sigue estancada por debajo de la masculina.

El análisis econométrico (mediante modelos de efectos aleatorios) revela dos conclusiones fundamentales:

1. El *ihs_PIB* no es determinante: No existe evidencia estadística de que un mayor desarrollo económico incremente la tasa de federación.
2. El *pesoEmp* tiene un efecto negativo: Existe una relación inversa significativa entre el empleo deportivo y la federación. Un mayor peso del sector laboral deportivo se asocia con una menor tasa de licencias.

En resumen, el estudio demuestra que la riqueza de una región no implica mayor deporte federado. En España coexisten lógicas diferentes: un modelo asociativo y de competición (predominante en el norte) frente a un modelo comercial y de ocio (predominante en zonas ricas y turísticas).

Por tanto, el crecimiento del "negocio" deportivo (*pesoEmp* y *ihs_PIB*) tiende a canalizar la práctica hacia el consumo privado a costa de la licencia federativa tradicional.

7. REFERENCIAS BIBLIOGRÁFICAS

- CSD. (2022). Obtenido de <https://www.csd.gob.es/sites/default/files/media/files/2022-12/Encuesta%20de%20Hábitos%20Deportivos%20en%20España%202022%20Resultados%20detallados.pdf>
- DEPORTEDATA. (s.f.). <https://www.educacionfpydeportes.gob.es/mc/deportedata/portada.html>.
- El aumento del deporte después de la pandemia.* (17 de Diciembre de 2024). Obtenido de <https://sportrade.es/pandemia-deporte/>
- Hernández Moreno, J. (2000). *La iniciación a los deportes desde su estructura y dinámica.* Obtenido de https://books.google.es/books?id=3ahOm_pB0xYC&pg=PA52&hl=es&source=gbs_selected_pages&cad=1#v=onepage&q&f=false
<https://www.educacionfpydeportes.gob.es/mc/deportedata/portada.html>. (s.f.). Obtenido de DEPORTEDATA.
- INE. (s.f.). Obtenido de <https://www.ine.es>
- Minitab.com . (s.f.). Obtenido de <https://support.minitab.com/es-mx/minitab/help-and-how-to/statistics/basic-statistics/supporting-topics/correlation-and-covariance/a-comparison-of-the-pearson-and-spearman-correlation-methods/>
- Navarro, S. (11 de Noviembre de 2022). *PCA en estadística: ¿Qué es y cómo calcularlo?* . Obtenido de <https://keepcoding.io/blog/que-es-pca-en-estadistica/>
- Parlebas, P. (2001). *Juegos, deporte y sociedad. Léxico de praxiología motriz.* Obtenido de <https://books.google.com.cu/books?id=vxDwXPRBnuoC&printsec=frontcover#v=onepage&q&f=false>
- Universoformulas.com. (s.f.). Obtenido de <https://www.universoformulas.com/estadistica/descriptiva/diagrama-caja/>