



Universidad de Oviedo

FACULTAD DE CIENCIAS

INFERENCIA ESTADÍSTICA

# ESTUDIO INFERENCIAL SOBRE EL GASTO EN VIAJES REALIZADOS EN JULIO DE 2024: PERSPECTIVAS DESDE LOS DATOS DEL I.N.E.

EQUIPO  $\varepsilon$

CORTE FERNÁNDEZ, CÉSAR

FERNÁNDEZ VEGA, NURIA

GARCÍA PESCADOR, ISABEL

LLORIÁN GONZÁLEZ, PABLO

RIVERA MENÉNDEZ, CARMEN ZIYI

RODRÍGUEZ CRISTÓBAL, MANUEL

RUBIO ROBLES, PABLO

SANTOS LÓPEZ, CANDELA DE LOS

30 de abril de 2025

# Índice

<b>1</b>	<b>Introducción</b>	<b>3</b>
1.1	Material y filtrado de datos . . . . .	3
1.2	Objetivos . . . . .	3
1.3	Variables . . . . .	4
1.4	Material y métodos . . . . .	4
<b>2</b>	<b>Análisis sobre la influencia del acompañante en el gasto total por día</b>	<b>5</b>
2.1	Estudio descriptivo . . . . .	5
2.2	Contrastes . . . . .	6
2.3	Conclusiones . . . . .	7
<b>3</b>	<b>Análisis sobre la influencia de los ingresos del hogar en el gasto total del viaje por día y en el destino elegido</b>	<b>7</b>
3.1	Relación con el gasto total por día . . . . .	8
3.2	Relación con el destino del viaje . . . . .	8
<b>4</b>	<b>Análisis de la relación entre el destino de viaje y el gasto total por día</b>	<b>10</b>
4.1	Estudio descriptivo . . . . .	11
4.2	Contrastes . . . . .	12
4.3	Conclusiones . . . . .	13
<b>5</b>	<b>Análisis de la relación entre el género del viajero y el gasto total</b>	<b>14</b>
5.1	Estudio descriptivo . . . . .	14
5.2	Contrastes . . . . .	15
5.3	Conclusión . . . . .	16
<b>6</b>	<b>Análisis de la influencia del sitio dónde se reservó el alojamiento en el gasto por día y por persona del mismo</b>	<b>16</b>
6.1	Estudio descriptivo . . . . .	17
6.2	Contrastes . . . . .	18
6.3	Conclusiones . . . . .	19
<b>7</b>	<b>Conclusiones globales</b>	<b>19</b>
<b>8</b>	<b>Reparto de tareas</b>	<b>20</b>

# 1. Introducción

España es un país con un fuerte componente turístico, tanto a nivel internacional como nacional. En el año 2024, el turismo representó en torno al 13,1 % del Producto Interior Bruto (PIB), lo que subraya su importancia estratégica para la economía nacional. Sin embargo, además de ser un destino turístico de relevancia global, España es también un país con una gran movilidad ya que sus propios residentes realizan muchos viajes, especialmente durante los meses de verano.

El presente estudio se basa en los datos recogidos por el Instituto Nacional de Estadística (INE) mediante su encuesta sobre viajes a los residentes en España en el mes de julio de 2024. A partir del conjunto de datos que proporciona, se desarrollará un análisis estadístico con el objetivo de identificar tendencias relevantes y realizar inferencias que nos permitan entender mejor las características de estos viajes.

## 1.1. Material y filtrado de datos

La muestra utilizada en este estudio está compuesta por un total de 6180 individuos residentes en España que realizaron algún tipo de viaje en el mes de julio de 2024. Para poder realizar el análisis, se ha reestructurado el conjunto de datos de la siguiente forma:

- Juntamos todas aquellas referentes a los acompañantes con los que se realizaba el viaje en una única variable.
- Para adaptarnos a la notación de R, cambiamos en las variables numéricas de gastos (totales y de alojamiento) los puntos por comas. Además creamos nuevas variables a partir de estas dividiendo por el número de días (y personas en el caso del gasto en alojamiento).
- En el conjunto de datos algunas variables cualitativas se habían recogido con un código numérico, asociamos cada valor numérico a la clase cualitativa correspondiente.

Este tratamiento de los datos facilita el análisis posterior.

## 1.2. Objetivos

El objetivo general de este estudio es analizar los patrones de viaje de los residentes en España durante el mes de julio de 2024, utilizando para ello métodos estadísticos inferenciales aplicados al conjunto de datos. Este análisis permitirá obtener conclusiones sobre variables clave como los destinos elegidos, el perfil sociodemográfico de los viajeros y las modalidades de reserva utilizadas en relación al gasto realizado en el viaje o el gasto realizado por día.

Más específicamente, analizaremos:

- Si existe una relación significativa entre el género del viajero y el gasto total en euros.
- Abordar la relación entre el destino del viaje y el gasto total diario.
- Determinar la influencia del destino en el gasto diario, proporcionando así información valiosa

para la toma de decisiones en el ámbito del turismo y la planificación de viajes.

- Describir el perfil sociodemográfico de los individuos que viajaron en julio de 2024, incluyendo variables como el género, ingresos del hogar y tipo de acompañamiento.
- Estudiar el gasto medio diario y total de los viajes en función de distintas categorías de ingreso, destino y tipo de acompañamiento.
- Evaluar si existen diferencias significativas en el gasto total según el género o el método de reserva empleado.
- Comprobar la normalidad de algunas variables clave y aplicar los contrastes de hipótesis adecuados según los resultados obtenidos.

### 1.3. Variables

En el estudio se han recogido tanto variables cualitativas como cuantitativas, y tanto discretas como continuas. A continuación, se indican las más relevantes con las que vamos a trabajar y los valores que toman:

#### Variables discretas:

- **Destino del viaje:** España, Resto de Europa, África, América, Resto del mundo.
- **Género:** Hombre, Mujer.
- **Tipo de acompañamiento:** Con amigos, Con compañeros de trabajo, Con hijos, Con otros familiares, Desplazamiento al Centro de Estudios/trabajo, En pareja, Solo.
- **Ingresos del hogar:** Hasta 999 euros, De 1000 a 1499 euros, De 1500 a 2499 euros, De 2500 a 3499 euros, De 3500 a 4999 euros, 5000 euros o más, No contesta.
- **Modo de reserva:** Con el hotel, Con una agencia de viajes, Directamente con el propietario, En página web especializada, En una agencia inmobiliaria, No procede, No sabe.

#### Variables continuas:

- Gasto en alojamiento por día/persona
- Gasto total por día
- Gasto total del viaje

Todas las variables continuas toman valores reales positivos.

### 1.4. Material y métodos

En primera instancia, se llevará a cabo un análisis descriptivo de los datos con el fin de familiarizarse con las variables y su distribución. Para esto, usaremos diagramas de cajas, de barras apiladas, de tipo “violín” y tablas descriptivas.

Posteriormente, se aplicarán diversas técnicas estadísticas inferenciales, entre ellas tests de contrastes como *Shapiro-Wilk*, *Kruskal-Wallis*, *Wilcoxon*, *Levenne*, *Holm*, *Dunn*, *Kolmogorov-Smirnov*, *Spearman* y la corrección *Lilliefors*. También usaremos técnicas como el *Q-Q plot*.

La elección adecuada y justificada de estos métodos estadísticos, basados en las características específicas de nuestros datos, garantiza la validez y robustez de los test que usaremos en los análisis posteriores.

Todas las pruebas estadísticas se realizarán utilizando un nivel de significación del 5 % ( $\alpha = 0,05$ ), con el fin de asegurar un alto grado de confianza en los resultados obtenidos.

## 2. Análisis sobre la influencia del acompañante en el gasto total por día

El objetivo de este análisis es estudiar si existe alguna relación entre el tipo de acompañantes con los que se realiza un viaje y el gasto total diario. Para ello, trabajamos con las variables *Acompañantes* y *GastoTotalDía*.

### 2.1. Estudio descriptivo

Antes de comenzar con la inferencia estadística, nos paramos a analizar los datos. Para ello, realizamos una representación gráfica mediante un diagrama de cajas, lo que nos permite tener una visión general de cómo varía el gasto diario dependiendo de con quién se realiza el viaje (Figura 1).

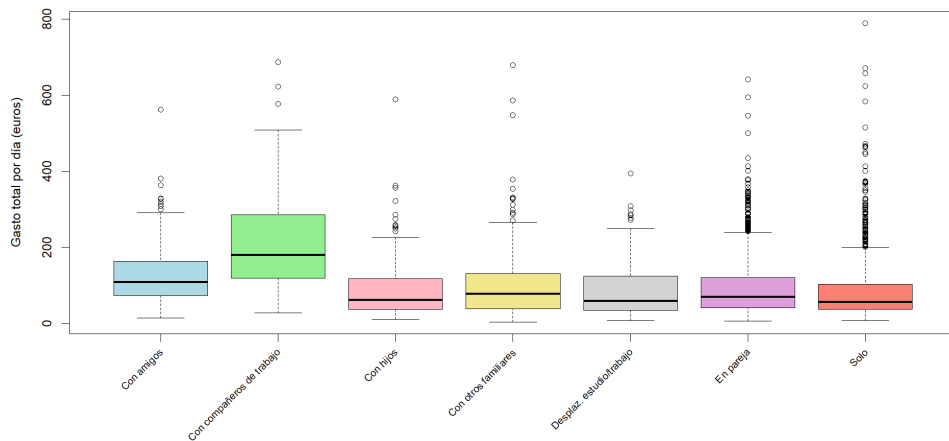


Figura 1: Diagrama de cajas del gasto diario por tipo de acompañantes.

Observamos que en esta representación destaca especialmente el grupo de viajes realizados con compañeros de trabajo, cuyo gasto total por día es visiblemente más elevado que el del resto. Además, muestra una gran dispersión en los datos y valores atípicos bastante elevados. Los viajes con amigos también presentan un gasto medio relativamente más alto que los otros grupos, aunque de forma no tan notable. Por el contrario, los viajes en solitario reflejan gastos diarios bastante más bajos.

De forma general, notamos que todos los casos presentan una mayor concentración de los datos en torno a valores bajos, por debajo de los 300€, y una gran asimetría, con colas más largas hacia gastos más elevados, aunque esto se debe fundamentalmente a valores atípicos.

## 2.2. Contrastes

Tras observar diferencias aparentes mediante el estudio descriptivo, el siguiente paso es comprobar de forma rigurosa si son significativas.

Lo primero sería realizar contrastes para comprobar si se cumplen las hipótesis necesarias para aplicar los test más usuales (ANOVA). Es decir, comprobaremos la normalidad y la homocedasticidad. Podemos asumir, por cómo se han obtenido los datos, que las variables aleatorias *GastoTotalDía* en función de los acompañantes son independientes.

Para analizar si la variable *GastoTotalDía* sigue una distribución normal dentro de cada grupo de acompañantes, aplicamos el test de Shapiro-Wilk en cada uno de ellos. Este test busca detectar fundamentalmente problemas de asimetría en el ajuste a una normal, por tanto, es el más indicado.

Los resultados muestran que los p-valores son extremadamente pequeños en todos los grupos ( $p < 3,9 \cdot 10^{-5}$ ). Esto implica que, al nivel de significación fijado ( $\alpha = 0,05$ ), debemos rechazar la hipótesis nula de normalidad en todas las categorías de acompañantes.

Para respaldar esta prueba numérica, se elaboraron gráficas *Q-Q* para cada grupo, que permiten detectar de forma visual las desviaciones en las colas de las distribuciones. En cada una de ellas se observa lo que ya anticipábamos, es decir, las variables no se ajustan a una distribución normal. Aquí lo mostraremos en el caso de viajar solo, que aparentemente era el caso en el que menos dispersión de los datos había. En la Figura 2, como era de esperar, los puntos muestrales tienden a separarse de la línea teórica sobre todo en las colas superiores.

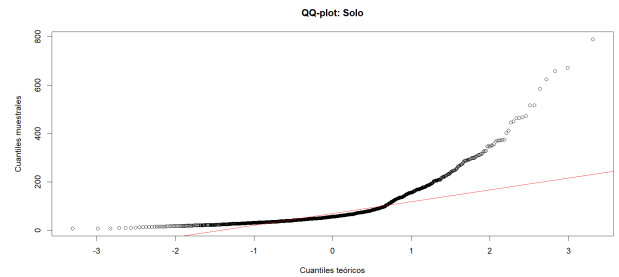


Figura 2: Gráfica Q-Q.

Aunque hayamos descartado la normalidad, aún debemos estudiar la homocedasticidad. Aplicando el test de Levene a nuestros datos, obtenemos un p-valor muy pequeño ( $p < 2,2 \cdot 10^{-16}$ ), lo que nos indica que debemos rechazar la hipótesis nula, igualdad de varianzas. Por tanto, la variabilidad del gasto diario sí depende del acompañante.

Este resultado refuerza lo que habíamos intuido en el estudio descriptivo, que había grupos con mayor dispersión que otros.

La falta de normalidad y de homocedasticidad nos impide el uso del ANOVA. Por tanto, debemos recurrir a otros test, como el de Kruskal-Wallis, un contraste no paramétrico que nos permite comparar las distribuciones de varios grupos. Es decir,  $H_0 =$  “*las distribuciones del gasto total diario son iguales entre todos los grupos de acompañantes*”.

Al aplicar el test de Kruskal-Wallis, el p-valor es muy bajo ( $p < 2,2 \cdot 10^{-16}$ ). Por tanto, rechazamos la hipótesis nula, lo que indica que hay diferencias significativas en la distribución entre los grupos.

Ahora, nos podemos preguntar cuáles son las que difieren entre sí; para ello, haremos un contraste

a posteriori, en concreto el test de Dunn con la corrección de Holm. Ver Tabla 1 donde “Sí” hace referencia a que hay diferencias significativas y “No” a que no hay diferencias significativas en la distribución.

	Compañeros	Hijos	Otros familiares	Pareja	Desplz. Estudio/Trabajo	Solo
Con amigos	Sí	Sí	Sí	Sí	Sí	Sí
Compañeros de trabajo		Sí	Sí	Sí	Sí	Sí
Con hijos			No	No	No	No
Con otros familiares				No	No	No
Pareja					No	Sí
Desplz. Estudio/Trabajo						No

Tabla 1: Comparaciones de gasto diario entre tipos de acompañantes

## 2.3. Conclusiones

El resultado del contraste respalda de manera clara lo que ya intuíamos con el análisis descriptivo: el tipo de acompañantes influye de forma significativa en el gasto total diario.

Estas diferencias tienen un sentido desde el punto de vista del comportamiento social y económico, ya que viajar con compañeros de trabajo suele implicar gastos elevados asociados a estancias profesionales, reembolsos de empresa o un menor control del gasto personal; viajar con amigos suele estar relacionado con actividades de ocio que tienden a aumentar el gasto y, además, puede influir el contagio del gasto, es decir, gastas más si el resto gasta más. Mientras que en los viajes en solitario o en los desplazamientos por estudios o trabajo se suele controlar más el gasto diario, ya sea porque es más fácil adaptarse al presupuesto que te propongas (si vas solo), no dispongas de tanto dinero (estudiante) o no quieras gastar de más (laboral).

Además, el hecho de que haya una gran dispersión en algunos grupos implica que hay viajeros con un comportamiento de gastos distinto. Lo que invita a considerar que el gasto no depende solo del acompañante, sino que también se ve afectado por otros factores que estudiaremos a continuación.

## 3. Análisis sobre la influencia de los ingresos del hogar en el gasto total del viaje por día y en el destino elegido

En el presente análisis se busca estudiar si los ingresos del hogar de los residentes en España influyen en dos aspectos clave relacionados con los viajes, como son el gasto total diario realizado durante el mismo y la elección del destino. Para ello, hemos excluido en este estudio los datos de aquellos viajeros que no han proporcionado información sobre el nivel de ingresos de su hogar, ya que no aportan valor al análisis.

Intuitivamente parece razonable suponer que mayores ingresos del hogar podrían estar asociados a

un mayor gasto durante el viaje, así como a la preferencia por destinos potencialmente más costosos o lejanos. Vamos a respaldar esta intuición con un análisis estadístico y descriptivo de los datos.

### 3.1. Relación con el gasto total por día

En primer lugar, veamos si existen diferencias significativas en el gasto total diario según el nivel de ingresos del hogar del viajero. En la Tabla 2 podemos ver cómo a medida que aumentan los ingresos, tanto la media como la mediana del gasto tienden a incrementarse. Asimismo, las desviaciones típicas también aumentan con el ingreso, lo que indica una mayor dispersión del gasto entre los grupos de mayores ingresos. Esta tendencia sugiere que los hogares con mayor capacidad económica realizan un gasto más elevado y variable durante sus viajes.

Ingresos del Hogar	Media	Mediana	Desv. Típica	Mínimo	Máximo
Hasta 999 euros	73.5	49.4	64.0	8.48	679.0
De 1000 a 1499 euros	85.1	64.9	63.1	4.37	562.0
De 1500 a 2499 euros	91.8	71.0	68.7	8.20	590.0
De 2500 a 3499 euros	93.8	74.1	71.8	9.14	790.0
De 3500 a 4999 euros	105.0	83.7	79.1	7.92	643.0
5000 euros o más	115.0	83.3	95.4	7.80	688.0

Tabla 2: Estadísticas descriptivas del gasto total por día en euros según el nivel de ingresos del hogar

Procedemos entonces con el estudio inferencial, planteando como hipótesis nula  $H_0 = \text{"No hay diferencias en los gastos totales según los niveles de ingresos del hogar"}$ . Al verificar las condiciones para aplicar un ANOVA, los test de Shapiro-Wilk y Levene arrojaron p-valores muy cercanos a cero, por lo que optamos por el test no paramétrico de Kruskal-Wallis para contrastar la homogeneidad. El contraste da un p-valor inferior a  $2,2 \cdot 10^{-16}$ , lo que nos lleva a rechazar la hipótesis nula, es decir, los gastos totales no se distribuyen de manera homogénea entre los diferentes niveles de ingresos del hogar.

A continuación, con el objetivo de comprobar si además existe una relación monótona creciente entre los ingresos del hogar y el gasto total diario, aplicamos el test de Spearman, transformando los ingresos en una escala ordinal. En el análisis se obtiene un coeficiente  $\rho = 0,124$  con un p-valor inferior a  $2,2 \cdot 10^{-16}$ , indicando una relación positiva y estadísticamente significativa entre ambas variables. Así, aunque esta relación directa es débil, el resultado respalda la existencia de una tendencia creciente en el gasto conforme aumentan los ingresos del hogar.

### 3.2. Relación con el destino del viaje

Antes de analizar la relación entre las variables *Destino* y el *GastoTotalDía*, puede ser pertinente estudiar la posible dependencia entre el destino y el nivel de ingresos del hogar, pues, tal como se



ha observado en las primeras conclusiones de esta sección, los ingresos del hogar influyen de manera clara en los patrones de gasto durante los viajes. Por tanto, comprender si existe una asociación significativa entre el nivel de ingresos y el destino elegido por los viajeros, puede aportar un contexto clave para interpretar adecuadamente los resultados del análisis posterior sobre el gasto total.

Comenzamos haciendo un estudio previo a través de una tabla de porcentajes (Tabla 3) y un gráfico de barras apiladas proporcionales (Figura 3).

Ingresos del Hogar	España	Resto de Europa	África	América	Resto del Mundo
Hasta 999 euros	93	6	1	1	0
De 1000 a 1499 euros	91	7	2	0	1
De 1500 a 2499 euros	91	8	1	0	0
De 2500 a 3499 euros	89	9	1	0	1
De 3500 a 4999 euros	85	13	1	0	1
5000 euros o más	87	10	0	1	1

Tabla 3: Distribución de ingresos del hogar según el destino

En primer lugar, en la tabla se aprecia que España y el resto de Europa concentran un mayor porcentaje de viajeros de todos los niveles de ingresos, de hecho, visualmente, estos destinos muestran una distribución bastante equilibrada entre los distintos tramos. Intuitivamente, esta tendencia podría explicarse por razones de accesibilidad geográfica, fruto de las ventajas que proporciona la Unión Europea, o una mayor familiaridad cultural, especialmente en el caso de España; aunque también por la posibilidad de unos menores costes en el viaje a nivel general. De hecho, este porcentaje es especialmente alto en España entre los niveles de ingresos más bajos (hasta los 999 euros) lo que parece respaldar esta idea.

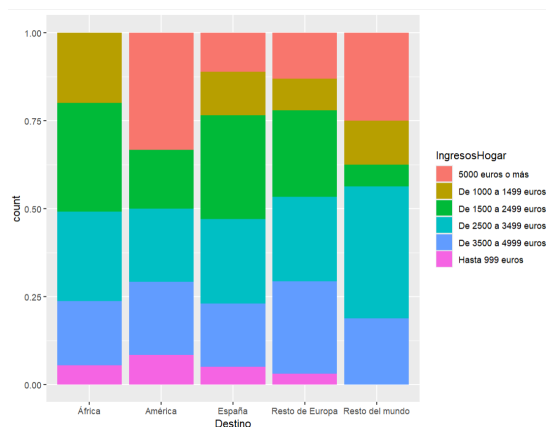


Figura 3: Relación proporcional de los Ingresos del Hogar respecto del Destino

Por otro lado, gráficamente se puede ver que las personas con ingresos medios en sus hogares suponen la mayor proporción de viajeros que escogen como destino África, mientras que en América, los ingresos de la mayor parte de viajeros son más altos. De hecho, resulta especialmente llamativo en la gráfica de barras la ausencia de viajeros con ingresos superiores a los 5000 euros entre los que eligen África. Esta falta de representación contrasta con lo que ocurre en destinos como América, donde la proporción de viajeros pertenecientes a esa franja de ingresos es considerablemente mayor al resto.

Este fenómeno puede estar relacionado con varios factores externos. En primer lugar, África, aunque geográficamente cercana a España, no suele ocupar una posición

destacada en el imaginario turístico de los hogares con rentas más altas, posiblemente debido a una percepción generalizada de limitaciones en infraestructura turística, seguridad o comodidad. En contraste, este grupo de viajeros suele buscar destinos que ofrezcan una experiencia más estructurada, con mayores garantías y servicios de alto nivel, características que se identifican con regiones como América del Norte o ciertos países de Sudamérica. La consolidada promoción turística y la oferta de experiencias exclusivas pueden hacer que estos destinos, a pesar de implicar viajes más largos y costosos, resulten más atractivos entre las personas con mayor capacidad de gasto. En cambio, muchos viajes a África desde España suelen estar ligados al turismo solidario, de aventura o a vínculos familiares, opciones más habituales entre viajeros de ingresos medios o medios-altos, pero no necesariamente entre los de renta más alta.

En cuanto al Resto del mundo, se destaca una presencia también relevante de los tramos de ingresos altos y medios-altos (más de 2500 euros), mientras que los hogares con los ingresos más bajos no aparecen siquiera en la representación. Esto refuerza la idea de que los viajes de larga distancia son elegidos preferentemente por hogares con mayor capacidad adquisitiva.

En términos generales, estas observaciones ya sugieren una relación clara entre el nivel de ingresos y la elección del destino. Para validarlo formalmente, se plantea un contraste de homogeneidad cuya hipótesis nula es  $H_0 = \text{"La variable Destino no depende de la variable IngresosHogar"}$ . El test  $\chi^2$  rechaza esta hipótesis, con un p-valor de  $4,9 \cdot 10^{-4}$ , confirmando la existencia de una relación significativa entre ambas variables. Aunque inicialmente parecía que la lejanía podía ser un factor clave, el coeficiente de Spearman obtenido ( $\rho = 0,058$ ), considerando el orden de lejanía de los destinos, indica una relación directa pero débil. El análisis de los residuos tipificados revela patrones más específicos: destaca el residuo de  $-2,50$  para África en los hogares de mayores ingresos (más de 5000 euros), así como una sobrerrepresentación de estos mismos hogares en América (residuo de  $3,18$ ). También resalta el residuo de  $4,02$  para Europa en el tramo de ingresos entre 3500 y 4999 euros, indicando que este grupo viaja más a esta región de lo esperado. En conjunto, los resultados muestran que las diferencias no responden solo a la distancia, sino también a otros factores económicos.

## 4. Análisis de la relación entre el destino de viaje y el gasto total por día

En el presente análisis, hemos abordado la relación entre el destino del viaje y el gasto total diario. Intuitivamente parece razonable suponer que destinos potencialmente más costosos o lejanos podrían estar asociados a un mayor gasto durante el viaje.

El estudio se estructura en dos fases principales. Inicialmente, hemos realizado un análisis descriptivo para caracterizar las variables y obtener una visión preliminar de su comportamiento. Posteriormente, hemos aplicado algunas de las pruebas estadísticas, con el objetivo de determinar la influencia del destino en el gasto diario.

## 4.1. Estudio descriptivo

En esta sección, realizamos un análisis descriptivo de la variable *GastoTotalDía* en función del *Destino* del viaje. El objetivo es obtener una primera aproximación a la distribución y características de los gastos diarios según el destino del viaje.

Destino	N	Media	Mediana	Desv. Tip.
España	5511	85.26	64.97	62.71
Resto de Europa	554	181.21	164.79	104.67
África	55	101.27	62.46	73.92
América	26	232.85	237.56	151.65
Resto del mundo	34	205.05	201.14	93.53

Tabla 4: Estadísticos descriptivos del gasto total por día según destino

La Tabla 4 resume los principales estadísticos descriptivos para cada destino, los cuales aparecen ordenados por orden de cercanía, pues en el estudio previo observamos que los datos sugerían una relación entre el nivel de ingresos y el destino: a mayor nivel de ingresos, mayor probabilidad de elegir destinos más lejanos. Se observa que el gasto medio diario es notablemente inferior en España (85,26€) en comparación con el resto de destinos, mientras que América presenta el gasto medio más elevado (232,85€). La mediana del gasto sigue un patrón similar, siendo también más baja en España (64,97€) y más alta en América (237,56€). La desviación típica es considerablemente mayor en América (151,65€) y Resto de Europa (104,67€), lo que sugiere una mayor variabilidad en los gastos diarios en estos destinos. Es importante destacar que el tamaño de la muestra para España es más elevado en comparación con los otros destinos, lo cual podría influir en los resultados. Un tamaño de muestra tan grande puede llevar a una mayor precisión en la estimación de los parámetros poblacionales para España, pero también podría hacer que pequeñas diferencias sean significativas.

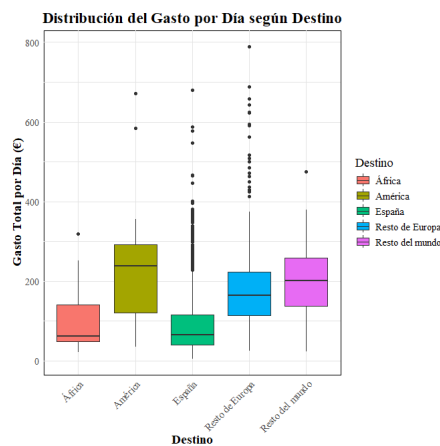


Figura 4: Diagrama de cajas del gasto diario por destino

En la Figura 4 podemos ver que España presenta una distribución más concentrada del gasto, mientras que Resto de Europa manifiesta una mayor dispersión. En resumen, los resultados sugieren que el destino del viaje influye en el gasto total diario, siendo España el destino más económico y

América el más costoso. A través de este estudio preliminar, hemos identificado patrones que nos puedan servir para el análisis inferencial posterior.

## 4.2. Contrastes

Antes de seleccionar un contraste estadístico para evaluar la influencia del destino en el gasto total diario, hemos intentado verificar el cumplimiento de las hipótesis necesarias para la aplicación del test ANOVA: la normalidad y la homocedasticidad.

En primer lugar, para contrastar si la variable *GastoTotalDía* sigue una distribución normal en cada destino, hemos aplicado el test de Kolmogorov-Smirnov con la corrección de Lilliefors, siendo la hipótesis nula  $H_0$ : la variable *GastoTotalDía* sigue una distribución normal.

Destino	p-valor
España	$8.87 \times 10^{-255}$
Resto de Europa	$6.74 \times 10^{-15}$
África	$7.29 \times 10^{-12}$
América	$1.90 \times 10^{-1}$
Resto del mundo	$8.36 \times 10^{-1}$

Tabla 5: p-valores del Test de Kolmogorov-Smirnov con corrección de Lilliefors

Los resultados del test, presentados en la Tabla 5, muestran los valores p-valores obtenidos para cada destino. Según los resultados obtenidos, al nivel de significación fijado ( $\alpha = 0,05$ ), se rechaza la hipótesis de normalidad para los destinos España, Resto de Europa y África. En cambio, no se rechaza la hipótesis nula para América y Resto del Mundo, lo que sugiere que la distribución del gasto total por día podría ser normal en estos destinos.

Los resultados del test de Kolmogorov-Smirnov con corrección de Lilliefors se confirman visualmente con las gráficas del tipo *Q-Q plot* (Figura 5).

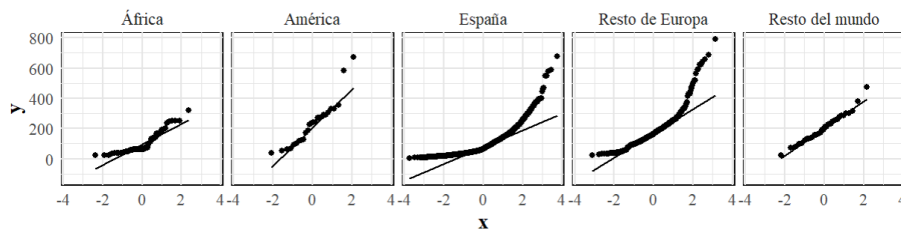


Figura 5: Gráficas *Q-Q plot*

A continuación, hemos contrastado el supuesto de igualdad de varianzas mediante el test de Levene, cuya hipótesis nula es  $H_0$ : Las varianzas de la variable *GastoTotalDía* son iguales en todos los destinos. Los resultados de este test nos dan un p-valor extremadamente pequeño ( $p < 2, 2 \cdot 10^{-16}$ ), lo que nos lleva al rechazo de  $H_0$  a nivel de significación  $\alpha = 0,05$ .

Dado que ninguna de las hipótesis (normalidad y homocedasticidad) necesarias para aplicar el test ANOVA se verifica, hemos optado por usar el test no paramétrico de Kruskal-Wallis para evaluar si existen diferencias significativas en el gasto diario total en los diferentes destinos. La hipótesis de dicho test es:  $H_0$ : Las distribuciones de la variable *GastoTotalDía* son iguales en todos los destinos.

El p-valor obtenido es extremadamente pequeño ( $p < 2,2 \cdot 10^{-16}$ ), lo que nos indica que existen diferencias significativas en el gasto total diario entre al menos dos destinos.

Los resultados obtenidos son coherentes con lo esperado, ya que el gasto diario varía según el destino. Por ejemplo, destinos asociados a viajes de negocios o turismo de lujo suelen implicar un mayor gasto cotidiano debido a precios más altos en alojamiento, comidas y actividades. En cambio, algunos destinos más turísticos pueden presentar gastos diarios menores. Estas diferencias reflejan las características y el tipo de turismo predominante en cada destino.

Después de haber rechazado el contraste de igualdad en las distribuciones, nos surgió la pregunta de a qué pares de destinos se debe el rechazo.

Podemos reformular la hipótesis  $H_0$  del siguiente modo  $\bigcap_{i,j} H_{0ij}$ : Las medias de *GastoTotalDía* son iguales en los destinos  $i$  y  $j$ . Podemos resolver tal contraste mediante el test de Dunn con la región crítica de Bonferroni. Tras aplicar el test, hemos obtenido los siguientes p-valores:

	Resto de Europa	África	América	Resto del mundo
España	$9.36 \times 10^{-138}$	$9.41 \times 10^{-1}$	$6.68 \times 10^{-9}$	$4.06 \times 10^{-13}$
Resto de Europa		$2.97 \times 10^{-9}$	1.00	1.00
África			$3.39 \times 10^{-4}$	$8.63 \times 10^{-6}$
América				1.00

Tabla 6: Resultados del test de Dunn con el ajuste de Bonferroni

Como se observa en la Tabla 6, el test (a nivel  $\alpha = 0.05$ ) muestra diferencias significativas en el gasto diario entre África y América, África y Resto de Europa, América y España, España y Resto de Europa, África y Resto del mundo; y España y Resto del mundo (p-valores muy pequeños). No se encontraron diferencias significativas entre América y Resto de Europa, América y Resto del mundo, ni Resto de Europa y Resto del mundo.

### 4.3. Conclusiones

En conclusión, el análisis revela que España presenta un patrón de gasto diario significativamente distinto al resto de destinos. Las diferencias entre América, Resto de Europa y Resto del mundo no son significativas tras la realización del contraste a posteriori, lo que sugiere que el gasto en estos destinos podría estar influenciado por factores como la lejanía. Por otro lado, el gasto turístico en España puede estar condicionado por otras causas, como por ejemplo, la tipología turística predominante

(sol y playa), la duración media de las estancias o la procedencia de los turistas (mayoritariamente europeos y con presupuestos normales).

## 5. Análisis de la relación entre el género del viajero y el gasto total

Ahora veremos si se puede observar una relación entre el género de los individuos de la muestra con el gasto total que realizan en el viaje.

### 5.1. Estudio descriptivo

Para comenzar este estudio, realizamos una pequeña tabla que recoge la cantidad de encuestados que son hombres y que son mujeres y, en función de eso, los valores de algunos estadísticos relevantes sobre el gasto total que han tenido en el viaje, la media, la mediana y la desviación típica.

Sexo	N	Media	Mediana	Desv. Típica
HOMBRE	2849	413.5890	239.91	509.2585
MUJER	3331	401.5793	244.98	473.1077

Tabla 7: Estadísticos descriptivos de *GastoTotal* por género

En la Tabla 7 vemos que toman valores similares. Esto nos da indicios de que el género podría no afecta en el gasto del viaje, pero hará falta un estudio inferencial para asegurarlo.

La muestra de individuos encuestados es aleatoria, por tanto, para conocer un poco más los datos con los que trabajaremos, vamos a estudiar si las proporción de hombres y de mujeres es igual o si por el contrario vemos diferencias significativas. Para esto recurriremos al EMV de la proporción de hombres, es decir, a la proporción muestral de hombres (número de hombres encuestados entre número total de individuos). Realizamos un contraste de proporciones con:  $H_0 : p = 0,5$ . Bajo la hipótesis nula, la proporción muestral se debería comportar como una binomial de parámetros  $n$  (el tamaño de la muestra) y  $p = \frac{1}{2}$ . A partir de esta información, obtenemos un p-valor prácticamente nulo, por tanto se rechaza la hipótesis nula, concluyendo que la proporción de hombres y de mujeres es significativamente diferente.

Para seguir con el estudio de la relación entre las variables, realizaremos un diagrama de tipo “violín”.

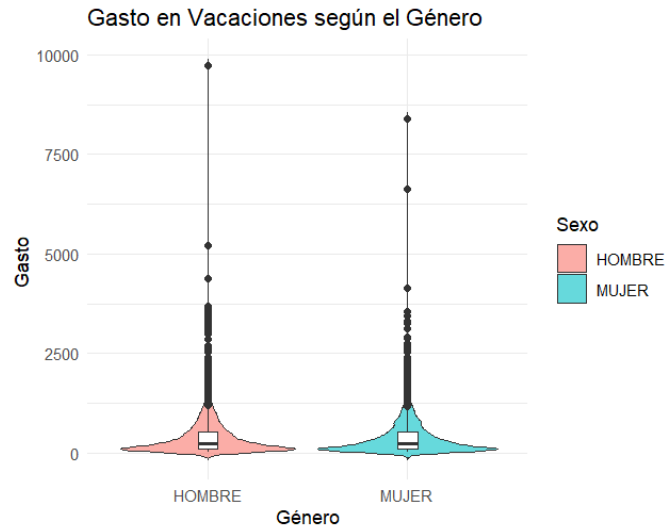


Figura 6: Diagrama de tipo "violín" del gasto de los viajeros según su género.

En la Figura 6 podemos observar una asimetría hacia la derecha tanto en el gasto total correspondiente a los hombres como en el de las mujeres. Además, vemos que tienen una representación muy similar. Tras el estudio descriptivo nos puede parecer que el género no influye en el gasto total del viaje, lo comprobaremos mediante contrastes.

## 5.2. Contrastes

Antes de elegir un contraste de homogeneidad, comprobamos si se cumplen las hipótesis necesarias (normalidad y homocedasticidad) para aplicar los test ANOVA. Para comprobar la normalidad, estudiaremos si la variable *GastoTotal* sigue una distribución normal tanto en hombres como en mujeres. Esto lo realizaremos aplicando el test de Shapiro-Wilk para cada género.

Una vez aplicado, observamos que los p-valores obtenidos tanto para hombres como para mujeres son muy pequeños ( $p < 0.05$ ), de hecho son prácticamente nulos, lo que indica una desviación muy significativa respecto a la normalidad. Luego se rechaza la hipótesis nula en ambos grupos. Como el test Shapiro-Wilk suele rechazar la hipótesis nula en muestras de tamaño "grande", vamos a respaldar estos resultados con *Q-Q* plot. Una vez representadas (Figura 7), podemos observar que confirman lo mismo que el test de Shapiro-Wilk, ya que para ambos sexos, su distribución de gastos se desvía de la distribución normal teórica, lo que se nota fundamentalmente en las colas superiores.

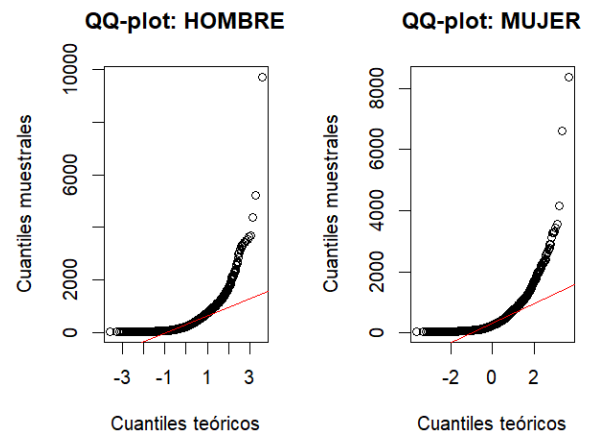


Figura 7: *Q-Q* plot

A continuación, analizaremos si se verifica la condición de **homocedasticidad**. Para ello, aplicaremos el test de **Levene**. Una vez realizado el test, observamos que el p-valor es significativamente grande (con respecto al  $\alpha$  fijado,  $p > 0,05$ ). Por tanto, no hay evidencias suficientes para decir que las varianzas del gasto total difieran entre hombres y mujeres. De esta forma se puede asumir la homocedasticidad de varianzas respecto al factor *Sexo*.

Por tanto, tenemos que se cumplen las condiciones de homocedasticidad pero no las de la normalidad, luego no se cumplen las hipótesis necesarias para aplicar ANOVA. En estas circunstancias, podemos realizar un test no paramétrico como el de Kruskal-Wallis, para comparar si hay homogeneidad entre las distribuciones del gasto total entre los distintos grupos. Como en este caso solo tenemos dos grupos (hombres y mujeres), aplicamos el test de Wilcoxon.

Tras aplicar el test de Wilcoxon, obtenemos un p-valor de 0.65 , es decir, bastante grande ( $p > 0,05$ ), que implica que no hay evidencias para rechazar la hipótesis nula. Luego no podemos asegurar que tengan distribuciones distintas.

### 5.3. Conclusión

Tras el análisis realizado podemos concluir que no hay evidencias para pensar que el género influye en el gasto total del viaje. Las similitudes entre los gastos totales de hombres y mujeres no solo se reflejan en los valores descriptivos (medias, medianas, desviaciones), sino también mediante contrastes inferenciales.

## 6. Análisis de la influencia del sitio dónde se reservó el alojamiento en el gasto por día y por persona del mismo

El objetivo de esta sección es determinar si el modo de reserva del alojamiento para un viaje influye de alguna forma en el coste por día y por persona del mismo, es decir, en el gasto diario en alojamiento por persona.

Para el estudio que se va a realizar, se filtrarán los datos, ya que no interesa tener en cuenta los valores "No sabe", "No procede" de la variable *ModoReserva*, debido a que estos se corresponden con aquellos individuos que no saben cómo reservaron el alojamiento o que no tenían asignada una respuesta a esta pregunta. Bajando así a una muestra de 2191 individuos.

Antes de comenzar con el estudio se puede destacar el hecho de que el modo de reserva de alojamientos no se distribuye de forma uniforme entre todas las posibilidades, lo cual puede verse reflejado en la Tabla 8. Como consecuencia, los test de hipótesis que se realicen con aquellos grupos cuyo tamaño es menor serán menos potentes que para los demás.



Hotel	Agencia	Propietario	Web	Inmobiliaria
896	486	256	503	50

Tabla 8: Tabla de frecuencias absolutas de la v.a. ModoReserva

## 6.1. Estudio descriptivo

Con el fin de tener una idea inicial sobre el estudio, se realizará un breve análisis descriptivo de los gastos diarios en alojamiento por persona en función del modo de reserva. Los resultados se ajustan en líneas generales al diagrama de cajas de la Figura 8.

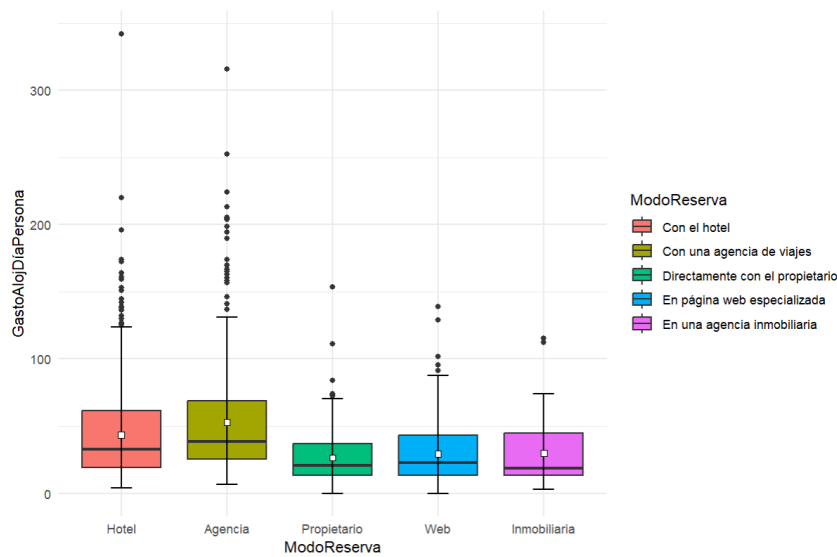


Figura 8: Diagrama de cajas de la v.a. *GastoAlojDíaPersona* en función de la v.a. *ModoReserva*

Por una parte, se puede observar que casi todos los grupos muestran una asimetría hacia la derecha, con colas más largas en la parte de los valores altos de gasto. Sin embargo, esta asimetría está causada principalmente por valores que son considerados atípicos, es decir, representan valores extremos.

Por otra parte, se pueden percibir algunas diferencias entre los diagramas. Por ejemplo, los gastos en alojamiento por día y por persona de aquellos individuos que reservaron con el hotel o con una agencia de viajes parecen tener medianas mayores que las de los demás. Más adelante se comprobará si dichas diferencias son realmente significativas o no.

En la Tabla 9 se puede observar la media muestral de los gastos en alojamiento por día y por persona en función del modo de reserva del mismo, junto con un intervalo de confianza a nivel  $1 - \alpha = 0.95$ . Esto también permite extraer una idea inicial de la homogeneidad que hay entre las muestras de cada uno de los modos de reserva; basta notar que los dos primeros grupos parecen ser algo más dispares al resto al tener medias de gasto en alojamiento por día y por persona más altas.

Modo de Reserva	Media muestral	IC para $\mu$
Con el hotel	43.2	(43.158, 43.305)
Con una agencia de viajes	52.9	(52.720, 53.052)
Directamente con el propietario	26.4	(26.292, 26.581)
En página web especializada	29.4	(29.347, 29.508)
En una agencia inmobiliaria	30.0	(28.999, 30.969)

Tabla 9: Intervalos de confianza para la media de la v.a. *GastoAlojDíaPersona* en función de la v.a. *ModoReserva*

## 6.2. Contrastes

Tal y como se han recopilado los datos, los valores que toman las variables aleatorias *GastoAlojDíaPersona* en función del modo de reserva son independientes entre sí. Basta notar que cada dato proviene de un individuo diferente y que cada individuo solo puede haber recurrido a un único modo de reserva.

Antes de entrar en el estudio principal de esta sección, se puede comprobar la normalidad y la homocedasticidad de las variables aleatorias *GastoAlojDíaPersona* en función del *ModoReserva*, para determinar qué test podemos aplicar.

Tras aplicar el test de Shapiro-Wilk para comprobar la normalidad, todos los p-valores son menores que  $2 \cdot 10^{-4}$ , luego hay evidencias suficientes para rechazar la hipótesis nula en todos los casos, es decir, queda descartado que las variables sigan una distribución normal.

Por otra parte, tras aplicar el Test de Levene para comprobar la homocedasticidad, el p-valor que se obtiene es menor que  $2 \cdot 10^{-16}$  y por tanto, hay evidencias suficientes para rechazar la hipótesis nula, es decir, queda descartado que las variables sean homocedásticas.

En la Figura 9 se pueden ver las funciones de densidad empíricas de las variables aleatorias *GastoAlojDíaPersona* en función del *ModoReserva*.



Figura 9: Densidad de la v.a. *GastoAlojDíaPersona* en función de la v.a. *ModoReserva*

Nótese que las funciones de densidad empíricas representadas en la Figura 9 parecen tener una distribución similar, ya que la mayor parte de la densidad de los gastos se concentra en el intervalo  $(0, 100)$ .

Para comprobar si efectivamente hay homogeneidad o no, se aplica el test de Kruskal-Wallis. El

p-valor que se obtiene es menor que  $2,2 \cdot 10^{-16}$  y, por tanto, hay evidencias suficientes para rechazar la hipótesis nula, es decir, las variables aleatorias consideradas no siguen una misma distribución.

Puede ser interesante hacer un contraste a posteriori por pares, para determinar cuáles son las que difieren entre sí y dan problemas en el test de homogeneidad. En este caso, se ha aplicado el test de Wilcoxon, obteniendo p-valores muy bajos entre todos los pares salvo en la pareja que está asociada a las reservas “Directamente con el propietario” y “En una agencia inmobiliaria” cuyo p-valor es de 1,00, la que está asociada a las reservas “En página web especializada” y “En una agencia inmobiliaria” cuyo p-valor es de 1,00 y la que está asociada a las reservas “En página web especializada” y “Directamente con el propietario” cuyo p-valor es de 0,34. Esto podría deberse a que los test de hipótesis que se realicen con el grupo “En una agencia inmobiliaria” cuyo tamaño es menor son menos potentes que para los demás, obteniendo así p-valores mayores.

### 6.3. Conclusiones

Tras haber realizado los contrastes que consideramos pertinentes se puede concluir que las variables no están distribuidas de manera homogénea, tomando valores en general mayores los gastos en alojamiento por día y por persona asociados a las reservas “Con el hotel” y “Con una agencia de viajes”. Esto podría deberse a que al reservar el alojamiento desde una agencia de viajes no solo se paga por el alojamiento en sí, sino que también se paga por los servicios de gestión de la agencia. En el caso de los gastos en alojamiento reservados directamente con el hotel, puede deberse a que se desea estar en un hotel en concreto sin darle tanta importancia al precio, ya que es más tedioso ir mirando los precios de los hoteles uno a uno.

Por otra parte, el hecho de que para los demás modos de reserva salga más barato en general el alojamiento podría deberse a que al reservar directamente con el propietario se ahorran los gastos de gestión y otras tasas; al reservar en páginas web especializadas, a veces hay descuentos, se pueden comparar y filtrar los alojamientos por precio; y al reservar en una agencia inmobiliaria, se suelen buscar alojamientos para estancias más largas, lo que puede abaratar el coste del alojamiento por día y por persona.

## 7. Conclusiones globales

Este análisis realizado a partir de datos del INE nos permite concluir que en el gasto en viajes de los residentes en España durante julio de 2024 influyen de forma significativa variables como el acompañante, el nivel de ingresos del hogar, el destino del viaje y el modo de reserva del alojamiento. En particular, se ha observado un mayor gasto diario en viajes realizados con compañeros de trabajo y una relación positiva entre ingresos y gasto. Además, los hogares con mayor capacidad adquisitiva tienden a elegir destinos más lejanos el gasto asociado al alojamiento varía según cómo se hiciera la reserva, siendo más elevado en reservas gestionadas por hoteles o agencias de viajes.

Por otro lado, no se encontraron diferencias significativas en el gasto total del viaje en función del género. En conjunto, los resultados obtenidos muestran patrones coherentes con las expectativas sociales y económicas.

## 8. Reparto de tareas

La división del trabajo se presenta a continuación:

- *Planteamiento* Realizado por César Corte Fernández, Manuel Rodríguez Cristóbal y Candela De Los Santos López
- *Análisis unidimensional variables y metodología* Realizado por César Corte Fernández y Candela De Los Santos López.
- *Análisis influencia de con quién se realiza el viaje en el gasto total por día y conclusiones generales* Realizado por Carmen Ziyi Rivera Menéndez.
- *Análisis de la influencia de los ingresos del hogar en el gasto total del viaje por día y la elección del destino* Realizado por Nuria Fernández Vega.
- *Análisis de la influencia del destino en el gasto total del viaje por día* Realizado por Pablo Llorián González
- *Análisis de la influencia de género y el gasto total del viaje* Realizado por Pablo Rubio Robles.
- *Filtrado de datos y análisis influencia del modo de reservay gasto por día y persona* Realizado por Isabel García Pescador
- *Coherencia, cohesión y maquetado* Realizado por César Corte Fernández, Nuria Fernández Vega, Isabel García Pescador y Carmen Ziyi Rivera Menéndez.

Tanto el informe como la presentación se han realizado de manera conjunta por todos los integrantes del equipo.

## Referencias

- [1] Instituto Nacional de Estadística (INE). (2024). *Encuesta de turismo de residentes. Resultados*. Recuperado el 27 de febrero de 2025, de [https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica\\_C&cid=1254736176990&menu=resultados&secc=1254736195369&idp=1254735576863](https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176990&menu=resultados&secc=1254736195369&idp=1254735576863)
- [2] RTVE.es. (2025, 16 de enero). *El turismo bate récord histórico en 2024 y aporta casi 208.000 millones de euros a la economía española*. Recuperado el 17 de abril de 2025, de <https://www.rtve.es/noticias/20250116/turismo-bate-record-historico-2024-aporta-casi-208000-millones-euros-a-economia-espanola-16409618.shtml>