

Casos de estudio de contrastes no paramétricos

Antonio Fernández Morales,

Departamento de Economía Aplicada (Estadística y Econometría)
Universidad de Málaga, 2024



Esta obra se encuentra bajo una Licencia Creative Commons Reconocimiento-NoComercial-SinObraDerivada. Puede copiar, distribuir y comunicar públicamente la obra bajo las condiciones siguientes:

- Reconocimiento: Debe reconocer los créditos de la obra citando al autor.
- No comercial: No puede utilizar esta obra para fines comerciales.
- Sin obras derivadas: No se puede alterar, transformar o generar una obra derivada a partir de esta obra.

1. Introducción

Las técnicas estadísticas no paramétricas han experimentado un importante desarrollo en los últimos decenios, debido en parte a algunas ventajas que proporcionan frente a otras técnicas clásicas [1]: requieren supuestos menos exigentes acerca de las poblaciones subyacentes de las que se obtienen los datos, muestran una mayor robustez frente a observaciones extremas y pueden ser más fáciles de interpretar.

En este trabajo se presentan varios casos de estudio de contrastes no paramétricos de utilidad para cursos de grado y posgrado en los ámbitos de la economía, administración de empresas, finanzas y turismo. Para ello, se ha optado por un enfoque aplicado, sin entrar en los conceptos teóricos, que pueden ser consultados en textos como [2], [3], [4], [5], [6]. Estos casos se añaden a una colección más amplia disponible en disciplinas como las ciencias actuariales [7], [8], [9], [10], o las técnicas cuantitativas aplicadas al turismo [11], [12] con los que se pueden combinar para un resultado más completo.

La resolución de los casos propuestos puede realizarse utilizando diversos recursos y *software*, siendo recomendable acudir a programas genéricos como SPSS [13], [14], Excel, [15], R [16] o Python [17] y complementarlos con otros recursos específicos disponibles online, como [18] o [19].

2 Se ha tomado una muestra aleatoria de 150 trabajadores en Andalucía para estudiar la distribución de probabilidad de la variable X: “Salario anual declarado de los trabajadores andaluces (10^3€)”. Verifique, con un nivel de significación $\alpha = 0,05$, si dicha variable sigue una distribución normal. Los datos resumen de la muestra son:

X: Salario anual declarado (10^3€)	n_i
0 a 5	25
5 a 10	55
10 a 15	46
15 a 20	12
20 a 25	12
	150

3 En el estudio “Desigualdad y pobreza en los ámbitos rural y urbano en España: un enfoque paramétrico” [21], se ha estimado para la variable X: “ingreso familiar anual familiar ajustado (en el ámbito urbano en 1996) en miles de euros” el siguiente modelo de probabilidad, definido por la función de distribución de la variable:

$$F(x) = (1 + 474x^{-3,27})^{-1,25}, \quad x > 0.$$

Verifique con un nivel de significación $\alpha = 0,05$ si la muestra aleatoria siguiente del ingreso familiar ajustado en miles de euros constantes de 1996 de cinco hogares procede del modelo citado:

9,6 2,4 5,1 38,0 11,2

4 En un estudio sobre el deporte en las empresas andaluzas [22] se obtuvieron los datos siguientes sobre la cuantía del patrocinio de actividades deportivas (10^3€) que realizan las empresas andaluzas en tres muestras aleatorias independientes.

	Cuantía anual de los patrocinios (10^3€)
Muestra 1 (empresas con 150 a 300 trabajadores)	0,9 3 5 8 10 13 14 17 20 30
Muestra 2 (empresas con 300 a 500 trabajadores)	9 19 23 27 29 35
Muestra 3 (empresas con más de 500 trabajadores)	2 10 12 16 18 24

A. ¿Se puede considerar que los datos de la muestra 1 proceden de una población normal, con un nivel de significación 0,10?

B. ¿Se puede considerar que los datos de la muestra 2 proceden de una población $N(\mu = 23, \sigma = 9)$, con un nivel de significación $\alpha = 0,05$?

C. ¿Se puede considerar que los datos de la muestra 3 proceden de una población normal, con un nivel de significación $\alpha = 0,05$?

5 Una empresa de telefonía móvil ha creado una sección de atención telefónica gratuita al cliente. Un estudio detallado de su primer año de funcionamiento revela que la duración de las llamadas recibidas sigue un modelo exponencial de media 10 minutos y que la cuantía de la factura mensual de consumo telefónico de los clientes que han usado el servicio sigue una distribución normal de media 25€y desviación típica 5€. Verifique con un nivel de significación $\alpha = 0,05$ con la muestra siguiente de 100 clientes atendidos si ambas variables son independientes.

Duración de la llamada (min.)	Cuantía de la factura mensual (€)		
	Menos de 20	20 a 25	25 y más
Menos de 4	10	12	8
4 a 11	12	18	6
11 y más	10	12	12

6 Verifique si las variables X: “edad” e Y: “duración del contrato” son independientes con un nivel de significación $\alpha = 0,05$, basándose en una muestra aleatoria de 850 trabajadores con contrato temporal cuyos resultados figuran en la tabla siguiente:

X: Edad	Y: Duración del contrato			
	Menos de 1 año	1 a 3 años	Más de 3 años	
16 a 29 años	10	12	8	370
30 a 49 años	12	18	6	417
50 y más años	10	12	12	63
	470	282	98	850

7 Verifique con un nivel de significación $\alpha = 0,02$ si la distribución del nivel educativo del cabeza de familia de los hogares españoles e italianos es homogénea empleando los datos de dos muestras que figuran en el estudio [23].

	Años de estudios completados					
	Menos de 5	5 a 8	8 a 13	13 a 16	Más de 16	
España	5491	8372	4287	2087	913	21150
Italia	646	2887	2186	1839	630	8188
	6137	11259	6473	3926	1543	29338

8 Sospechamos que los datos de las muestras siguientes de la variable gasto semanal en actividades de ocio de los adolescentes de una ciudad (€), recogidas por tres encuestadores contratados, no son aleatorias. Verifique con un nivel de significación $\alpha = 0,05$ la aleatoriedad de dichas muestras.

Muestra 1 30 55 25 22 55 60 20 50 60 20

Muestra 2 60 20 50 60 20 60 20 55 60 20 30 55 25

Muestra 3 60 55 50 60 50 60 60 20 20 20 30 25 25 25

9 Verifique con un nivel de significación $\alpha = 0,05$ la hipótesis consistente en que la duración del periodo de desempleo observado de los parados de una comunidad autónoma presenta una mediana de 14 meses, con los datos muestrales:

Duración del periodo de desempleo (meses): 1 10 14 11 2 15 4 2 16

A. Utilice el test de los signos

B. Utilice el test de Wilcoxon de rangos-signos

10 Un gran centro comercial ha introducido un cambio en el sistema de crédito para las compras de sus clientes con tarjeta. Para verificar si el nivel de compras semanal de sus clientes con tarjeta se ha visto modificado con la implantación de este cambio, ha tomado una muestra aleatoria de 8 clientes, registrando su nivel de compra semanal antes y después del cambio. Realice la verificación con $\alpha = 0,10$, empleando un contraste no paramétrico.

		Gasto semanal (€)							
		Cliente							
		1	2	3	4	5	6	7	8
Antes		210	12	21	40	226	480	60	65
Después		216	15	17	29	235	488	65	72

11 Una empresa de trabajo temporal asegura en su publicidad que trata a sus nuevos clientes con la misma diligencia que a sus clientes con más antigüedad, lo que se traduce en contratos de trabajo con una duración similar. Para verificar esta hipótesis con un nivel de significación $\alpha = 0,01$, se ha tomado una muestra de 25 clientes con más de un contrato firmado, registrando la duración de su primer y segundo contrato de trabajo. Realice el contraste.

Muestra 1	Muestra 2
9	1
1	6
5	1
6	7
1	14
1	2
1	24
8	1
6	3
4	3
35	3
2	2
1	12
1	30
14	2
3	4
3	1
1	15
10	1
4	3
5	15
3	2
7	1
20	5
2	3

12 Para analizar el comportamiento de los precios de bienes de consumo se ha tomado una muestra aleatoria de los precios de varios productos en 30 puntos de venta de una provincia. Realice una verificación de la aleatoriedad de la muestra correspondiente a los precios de una conocida bebida refrescante en formato 2 litros.

Precio observado en euros:

1,4; 1; 0,8; 1; 1,4; 1,3; 0,9; 0,9; 1,2; 1,1; 0,8; 1,2; 1,1; 1,3; 1,3;

1; 0,8; 0,9; 1; 1,3; 1,2; 1,2; 1,3; 1,1; 1,3; 1,3; 1,4; 1,2; 1,1; 0,8

13 En las investigaciones sobre turismo social en España [24], [25], [26], se ha analizado el papel que el programa del IMSERSO tiene en la resiliencia de los destinos turísticos costeros españoles frente a *shocks* económicos externos y en su efecto contra la estacionalidad [27]. Con los datos siguientes del número de hoteles que ofertan este programa vacacional en 23 destinos turísticos en 2008 y 2019, verificar mediante un contraste no paramétrico si ha cambiado entre ambas fechas el nivel del número de hoteles con este programa.

Punto turístico	Número de hoteles	
	2008	2019
1	39	32
2	2	1
3	0	1
4	14	11
5	13	8
6	18	12
7	6	6
8	14	12
9	4	2
10	4	4
11	1	0
12	1	0
13	10	9
14	2	5
15	0	2
16	6	4
17	0	3
18	0	1
19	1	0
20	4	0
21	24	17
22	8	4
23	10	9

14 Existe una preocupación creciente por las repercusiones de la especialización productiva en el sector turístico, que ha promovido la creación de observatorios de turismo sostenible en los destinos turísticos más relevantes, como es el caso de Málaga, dedicados a la monitorización de los principales indicadores de sostenibilidad económica, social y ambiental [28]. Con los datos de una muestra de 1000 contratos de trabajo firmados en Málaga en 2022 en los sectores que componen el núcleo de la actividad turística, realice un contraste de independencia con nivel de significación $\alpha = 0,05$ entre las variables tipo de contrato (indefinido o temporal) y género.

	Tipo de contrato	
	Temporal	Indefinido
Mujer	308	230
Hombre	235	227
	543	457

Referencias

- [1] Hollander, M., Wolfe, D. A., Chicken, E. (2014). *Nonparametric Statistical Methods. Third Edition*. New Jersey: Wiley & Sons.
- [2] Gibbons, J. D., Chakraborti, S. (2021). *Nonparametric Statistical Inference, 6th Edition*. Abingdon, Oxon: CRC Press.
- [3] Dickhaus, T. (2018). *Theory of Nonparametric Tests*. Cham, Switzerland: Springer.
- [4] Neuhauser, M. (2012). *Nonparametric Statistical Tests A Computational Approach*. Boca Raton: CRC Press Chapman and Hall.
- [5] Sheskin, D. (2011). *Handbook of Parametric and Nonparametric Statistical Procedures, Fifth Edition*. Abingdon, Oxon: CRC Press.
- [6] Desu, M. M., Raghavarao, D. (2004). *Nonparametric Statistical Methods For Complete and Censored Data*. New York: Chapman and Hall.
- [7] Fernández-Morales, A. (2023). *Aplicaciones actuariales de modelos multi-estados y tablas de decrementos múltiples*. RIUMA, Universidad de Málaga. Disponible en: <https://hdl.handle.net/10630/28165>
- [8] Fernández-Morales, A. (2023). *Aplicaciones actuariales de graduación paramétrica de tablas de supervivencia y mortalidad*. RIUMA, Universidad de Málaga. Disponible en: <https://hdl.handle.net/10630/28019>
- [9] Fernández-Morales, A. (2022). *Casos de estudio de estimación de funciones de supervivencia para Ciencias Actuariales*. RIUMA, Universidad de Málaga. Disponible en: <https://hdl.handle.net/10630/25644>
- [10] Fernández-Morales, A. (2022). *Casos de estudio de modelos de supervivencia para Ciencias Actuariales*. RIUMA, Universidad de Málaga. Disponible en: <https://hdl.handle.net/10630/25377>
- [11] Fernández-Morales, A. (2022). *Investigaciones en turismo con análisis cuantitativo*. RIUMA, Universidad de Málaga. Disponible en: <https://hdl.handle.net/10630/25455>
- [12] Fernández-Morales, A. (2021). *Técnicas Estadísticas para Dirección y Planificación del Turismo: Cuestiones y aplicaciones*. RIUMA, Universidad de Málaga. Disponible en: <https://hdl.handle.net/10630/25455>
- [13] Hinton, P., McMurray, I., Brownlow, C., Terry, P. C. (2024). *SPSS Explained. 3rd Edition*. New York: Routledge.

- [14] Ho, R. (2014). *Handbook of Univariate and Multivariate Data Analysis with IBM SPSS. Second Edition*. Boca Raton, Florida: CRC Press.
- [15] Panneerselvam, R. (2024). *Business Statistics Using Excel A Complete Course in Data Analytics*. New York: Routledge.
- [16] John Kloke, J., McKean, J. (2024). *Nonparametric Statistical Methods Using R. 2nd Edition*. Abingdon: CRC Press, Chapman & Hall Texts in Statistical Science.
- [17] Agresti, A., Kateri, M. (2022). *Foundations of Statistics for Data Scientists With R and Python 1st Edition*. Abingdon: CRC Press, Chapman & Hall Texts in Statistical Science.
- [18] Fernández-Morales, A. (2009). *CALCUPROB An on-line interactive calculator of probabilities*. RIUMA, Universidad de Málaga. Disponible en: <http://hdl.handle.net/10630/5071>
- [19] Fernández-Morales, A. (2002). *Tamaños muestrales: Instrumentos interactivos on-line para la formación estadística en el sector turístico*. RIUMA, Universidad de Málaga. Disponible en: <http://hdl.handle.net/10630/5075>
- [20] Fernández Morales, A., Lacomba Arias, B. (2003). *Técnicas estadísticas para el turismo*, Ed. Ágora Universidad, Málaga.
- [21] Fernández Morales, A. (2003). Desigualdad y pobreza en los ámbitos rural y urbano en España: un enfoque paramétrico. *Estadística*, Vol. 55, nº 164-165, pp. 351-365, Instituto Interamericano de Estadística, Madrid.
- [22] Otero Moreno, J. M., Fernández Morales, A., Isla Castillo, F. (2002). *Workplace sport in Andalusia*, 10th European Sport Management Congress, Jyväskylä, Finlandia.
- [23] Costa, M., Fernández Morales, A. (1997). Livello di istruzione e disuguaglianza nella distribuzione dei redditi: La scomposizione degli indici di Gini e di entropia generalizzata, *Statistica*, 57, 309-32.
- [24] Fernández-Morales, A., McCabe, S., Cisneros-Martínez, J. D. (2024). Is Social Tourism a Vector for Destination Resilience to External Shocks? Evidence From Spain. *Journal of Travel Research*, 63(7), 1606-1625, <https://doi.org/10.1177/00472875231200493>.
- [25] Cisneros-Martínez, J. D., McCabe, S., Fernández-Morales, A. (2023). Assessing the contribution of different markets in combatting destination seasonality: The case of Benidorm, Spain. *Journal of Destination Marketing and Management*, 29, <https://doi.org/10.1016/j.jdmm.2023.100802>.

- [26] Cisneros-Martínez, J. D., Fernández-Morales, A. (2020). The social tourism programmes in Spain. En Diekmann, A., McCabe, S. (eds) *Handbook of Social Tourism*, 72-82. Cheltenham: Edward Elgar Publishing.
- [27] Fernández-Morales, A. (2021). Tourism seasonality across markets. En Ferrante, M., Fritz, O., Öner, Ö. (eds.) *Regional Science Perspectives on Tourism and Hospitality*. Advances in Spatial Science. Springer, Cham, pp 125-141. https://doi.org/10.1007/978-3-030-61274-0_7
- [28] Fernández-Morales, A. (2023, November 23-24). Knowledge management in the sustainability of the destination. From data to the sustainable transformation of the destination: The University at the service of knowledge of the Smart Destination [Conference presentation]. *Present and future of sustainable tourism destinations*, STO Malaga and UNWTO International Network of Sustainable Tourism Observatories, Málaga, Spain.