

Casos de estudio de muestreo en poblaciones finitas

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

En el ámbito de la economía y la administración de empresas, rara vez tenemos información estadística completa y detallada del problema que queremos analizar. Lo más frecuente es tomar decisiones con información parcial, que generalmente proviene de algún tipo de muestra de la población en la que estamos interesados. Por ello es muy importante conocer cómo evaluar la calidad de las estimaciones procedentes de una muestra y cómo diseñar y analizar los tipos, muy variados, de muestras [1].

Los casos de estudio de muestreo en poblaciones finitas que se incluyen en esta obra se han elaborado para su uso en programas de estudios de grado y posgrado en titulaciones de las áreas económica y empresarial, entre los que se contemplan también títulos de ciencias actuariales, finanzas o turismo. El enfoque seleccionado es eminentemente práctico, dejando los aspectos teóricos para su consulta en manuales como [1], [2], [3], [4], [5] o [6]. Se puede ampliar este conjunto de casos con una colección disponible en abierto, para ciencias actuariales ([7], [8], [9], [10], entre otros) o relacionado con técnicas cuantitativas aplicadas al turismo (por ejemplo [11] o [12]).

Es recomendable resolver los casos prácticos empleando recursos y software actualizados, ya sean de tipo genérico como Excel, [13], SPSS [14], [15], R [16] y Python [17], o recursos para objetivos más específicos, como [18] o [19] disponibles en abierto.

2. Casos

Caso 1

Para conocer diversas características laborales de los titulados en Ciencias Actuariales y Financieras por la Universidad de Málaga, [20], [21], se ha tomado una muestra aleatoria simple sin reemplazamiento de 112 titulados. Se ha registrado para cada uno de ellos el valor de la variable X: “tiempo transcurrido desde la licenciatura hasta encontrar el primer empleo”, con los siguientes resultados:

$$\sum_1^{112} X_i = 248, \sum_1^{112} X_i^2 = 1990 \quad (1)$$

A Calcule la media muestral

$$\bar{x} =$$

B ¿Es la media muestral un estimador insesgado de μ ?

Sí No

C Calcule los siguientes estimadores de σ^2 :

$$s^2 = \frac{\sum_1^n (X_i - \bar{x})^2}{n} =$$
$$\hat{s}^2 = \frac{\sum_1^n (X_i - \bar{x})^2}{n - 1} =$$

D ¿Es s^2 un estimador insesgado de σ^2 ?

Sí No

E ¿Es \hat{s}^2 un estimador insesgado de σ^2 ?

Sí No

F Para un número total de licenciados en Ciencias Actuariales y Financieras por la Universidad de Málaga de 257, estime la varianza del estimador puntual del tiempo medio hasta encontrar el primer empleo de esta población.

G ¿Cuál o cuáles de estas ecuaciones son correctas?

$E\left(\frac{\hat{s}^2}{n} \frac{N-n}{N}\right) = \sigma_{\bar{x}}^2$

$E(\hat{s}^2) = \sigma^2 \frac{N}{N-1}$

$E\left(\frac{\hat{s}^2}{n} \frac{N-n}{N}\right) = \sigma^2$

$E(\hat{s}^2) = \sigma_{\bar{x}}^2 \frac{N}{N-1}$

H Estime un intervalo de confianza con $1-\alpha=0,95$ para el tiempo medio en encontrar empleo de los licenciados en Ciencias Actuariales y Financieras por la Universidad de Málaga. Cuantifique el error de estimación en meses.

Caso 2

En un estudio relativo a diversas características y preferencias de los alumnos de la Universidad de Málaga, [22], se ha tomado una muestra aleatoria simple sin reemplazamiento de tamaño 400 entre los 10372 alumnos del Campus de El Ejido. El número de alumnos de la muestra que usan habitualmente el coche propio para acceder a la Universidad es igual a 168.

A Obtenga el estimador puntual de la proporción de alumnos del Campus de El Ejido que usan el coche propio para desplazarse a la Universidad.

$$\hat{p} =$$

B ¿Es la proporción muestral un estimador insesgado de p ?
 Sí No

C ¿Cuál de las siguientes es la varianza del estimador \hat{p} ?

$Var(\hat{p}) = \frac{\hat{p}(1-\hat{p})}{n} \frac{N-n}{N-1}$

$Var(\hat{p}) = \frac{p(1-p)}{n} \frac{N-n}{N-1}$

$Var(\hat{p}) = \frac{\hat{p}(1-\hat{p})}{n-1} \frac{N-n}{N-1}$

$Var(\hat{p}) = \frac{p(1-p)}{n-1} \frac{N-n}{N-1}$

D Calcule el error estándar de la estimación puntual que ha realizado en el epígrafe A.

E Estime un intervalo de confianza con $1-\alpha=0,90$ para el porcentaje de alumnos del Campus de El Ejido que usan el coche propio para acceder a la Universidad.

Caso 3

En el estudio realizado para el Observatorio del Deporte Andaluz (cuyos principales resultados pueden consultarse en [23]) se realizó una encuesta sobre deporte escolar. Dicha encuesta se basó en una muestra representativa de los centros escolares andaluces no universitarios, tanto públicos, como privados y concertados. La muestra, de tamaño 349, se seleccionó aleatoriamente entre los 3664 centros de todo el territorio andaluz, estratificando según tipo de centro. Deseamos realizar una estimación de la proporción de centros que están abiertos para competiciones extraescolares fuera del horario escolar para equipos que no son del centro.

A Teniendo en cuenta que el tamaño poblacional de cada estrato se muestra en la tabla 1, ¿cuál es el tamaño muestral seleccionado en cada estrato?

Tabla 1: Número de centros escolares en Andalucía no universitarios

Estrato	Centros n_i
C. Público	3011
C. Privado/Concertado	653
Total	3664

B Los resultados de la muestra se reflejan en la tabla 2. Con dicha información, realice una estimación puntual de la proporción de centros que están abiertos para competiciones extraescolares fuera del horario escolar para equipos que no son del centro.

Tabla 2: Centros abiertos en horario extraescolar en la muestra

Estrato	Centros abiertos
C. Público	193
C. Privado/Concertado	32
Total	235

$$\hat{p} =$$

C Teniendo en cuenta que \hat{p} es el estimador de p obtenido con este diseño muestral estratificado, ¿cuál de las siguientes es la varianza de \hat{p} ?

$$\textcircled{O} \text{Var}(\hat{p}) = \sum_{i=1}^2 \frac{N_i^2}{N^2} \frac{N_i - n_i}{N_i - 1} \frac{\hat{p}_i(1 - \hat{p}_i)}{n_i}$$

$$\textcircled{O} \text{Var}(\hat{p}) = \sum_{i=1}^2 \frac{N_i^2}{N^2} \frac{N_i - n_i}{N_i - 1} \frac{p_i(1 - p_i)}{n_i}$$

$$\textcircled{O} \text{Var}(\hat{p}) = \sum_{i=1}^2 \frac{N_i^2}{N^2} \frac{N_i - n_i}{N_i} \frac{p_i(1 - p_i)}{n_i}$$

$$\textcircled{O} \text{Var}(\hat{p}) = \sum_{i=1}^2 \frac{N_i^2}{N^2} \frac{N_i - n_i}{N_i} \frac{\hat{p}_i(1 - \hat{p}_i)}{n_i}$$

D Complete la tabla 3 realizando una estimación de la varianza de los estimadores de la proporción en cada estrato y de la proporción de la población.

Tabla 3: Centros abiertos en horario extraescolar en la muestra

Estrato	Varianza estimada
C. Público	$\hat{\sigma}_{\hat{p}_1}^2 =$
C. Privado/Concertado	$\hat{\sigma}_{\hat{p}_2}^2 =$
Total	$\hat{\sigma}_{\hat{p}}^2 =$

E Estime un intervalo de confianza con un nivel $1 - \alpha = 0,90$ para la proporción de centros que están abiertos para competiciones extraescolares fuera del horario escolar para equipos que no son del centro.

Caso 4

La Consejería de Turismo y Deporte de la Junta de Andalucía promovió, dentro de la fase de diagnóstico del Plan General del Deporte Andaluz 2004-2008, la realización de un amplio estudio sobre las prácticas relacionadas con el deporte en Andalucía, [24]. La encuesta realizada siguió un esquema estratificado por provincias, siendo la población objeto de estudio los andaluces de 16 y más años en 2002 (tabla 4). Para estimar el número medio de horas semanales de consumo televisivo relacionado con el deporte en Andalucía, se ha estimado en cada muestra provincial la media de esta variable (tabla 4).

Tabla 4: Consumo televisivo semanal (horas) de espacios deportivos

Estrato	T. Población	T. Muestra	\bar{x}_i	$\frac{N_i}{N} \bar{x}_i$	$\hat{\sigma}_{\bar{x}_i}^2$	$\frac{N_i^2}{N^2} \hat{\sigma}_{\bar{x}_i}^2$
Almería	447724	72	2,54	0,18	0,1177	0,00061
Cádiz	951413	153	3,66	0,56	0,1093	0,00255
Córdoba	646712	104	2,66	0,28	0,1253	0,00135
Granada	690241	111	2,44	0,27	0,1105	0,00136
Huelva	385540	62	4,02	0,25	0,2445	0,00094
Jaén	541000	87	2,08	0,18	0,1082	0,00082
Málaga	1100654	177	2,22	0,39	0,0334	0,00104
Sevilla	1467539	236	2,62			
Total	6230823	1002				

A Estime el número medio de horas de consumo televisivo semanal de espacios deportivos en Andalucía con un estimador insesgado.

$$\bar{x} =$$

B Estime el error estándar de la estimación del epígrafe A.

C Obtenga un intervalo de confianza con un nivel $1 - \alpha = 0,95$ para la media del número de horas semanales de consumo televisivo de espacios deportivos en Andalucía.

Caso 5

El decanato de la Facultad de Ciencias Económicas y Empresariales quiere realizar un estudio similar al analizado en el caso 2. En este caso, se desea una estimación de la proporción de alumnos del Campus de la Facultad que usan el coche propio para acceder a la Universidad, con un error máximo de $\pm 0,04$, con un nivel de confianza de $1 - \alpha = 0,95$. ¿Cuál es el tamaño muestral mínimo (para una muestra aleatoria simple sin reemplazamiento), conociendo que la población es 3908 alumnos? Dado que ha pasado un cierto tiempo respecto del estudio del caso 2, supondremos completamente desconocida la proporción poblacional.

Caso 6

El turismo senior es un segmento que puede contribuir positivamente en el desarrollo de algunos destinos [25], [26], [27], particularmente en la lucha contra la estacionalidad, [28]. Por ello, estamos interesados en conocer la proporción de establecimientos hoteleros de la provincia de Málaga que realizan ofertas específicas para este segmento turístico mediante una muestra aleatoria simple sin reemplazamiento, con un error máximo de $\pm 0,05$, y un nivel de confianza de $1 - \alpha = 0,95$. Teniendo en cuenta que el número de establecimientos hoteleros en la provincia de Málaga asciende a 671, calcule el tamaño muestral mínimo (supondremos desconocida la proporción poblacional).

Caso 7

Una de las principales preocupaciones de los observatorios de sostenibilidad turística de destinos urbanos, como el la ciudad de Málaga, [29], se encuentra en los efectos sobre el tejido económico y social del crecimiento del mercado de viviendas turísticas, [30], [31], [32]. Para profundizar en esta problemática, queremos estimar el precio medio por noche (en €) en este tipo de alojamiento. Calcule el tamaño muestral mínimo para estimar la media de X : “precio del alojamiento por noche en vivienda turística en Málaga en €” con un error máximo de ± 10 €, con un nivel de confianza de $1 - \alpha = 0,95$ (considere que $Var(X) = 1600€^2$).

Caso 8

Basándonos en el estudio citado en el caso 1, se va a realizar una encuesta similar (con diseño muestral basado en una muestra aleatoria simple sin reemplazamiento) en otra Universidad española que imparte una titulación similar, de la que se han graduado 850 alumnos. Si se asume que la varianza de la variable X : “tiempo transcurrido desde la licenciatura hasta encontrar el primer empleo” es igual a 13 meses², obtenga el tamaño muestral mínimo para estimar la media de X con un error máximo de $\pm 0,5$ meses (con un nivel de confianza de $1 - \alpha = 0,95$).

Referencias

- [1] Lohr, S. L. (2022). *Sampling Design and Analysis. 3rd Edition*. Abingdon, Oxon: CRC Press.
- [2] Tille, Y. (2020). *Sampling and Estimation from Finite Populations*. New York: Wiley and Sons.
- [3] Arnab, R. (2017). *Survey Sampling Theory and Applications*. Oxford: Academic Press.
- [4] Fuller, W. A. (2009). *Sampling Statistics*. New York: Wiley and Sons.
- [5] Barnett, V. (2009). *Sample Survey Principles and Methods. 3rd Edition*. New York: Wiley and Sons.
- [6] Valliant, R., Dorfman, A. H., Royall, R. M. (2000). *Finite Population Sampling and Inference: A Prediction Approach*. New York: Wiley and Sons.
- [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] Panneerselvam, R. (2024). *Business Statistics Using Excel A Complete Course in Data Analytics*. New York: Routledge.
- [14] Hinton, P., McMurray, I., Brownlow, C., Terry, P. C. (2024). *SPSS Explained. Third Edition*. New York: Routledge.

- [15] Ho, R. (2014). *Handbook of Univariate and Multivariate Data Analysis with IBM SPSS. Second Edition*. Boca Raton, Florida: CRC Press.
- [16] John Kloke, J., McKean, J. (2024). *Nonparametric Statistical Methods Using R. Second 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. First Edition*. Abingdon: CRC Press, Chapman & Hall Texts in Statistical Science.
- [18] Fernández-Morales, A. (2009). CALCUPROB An on-line interactive calculator of probabilities. 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. Universidad de Málaga. Disponible en: <http://hdl.handle.net/10630/5075>
- [20] Fernández-Morales, A., Moreno-Ruiz, R. (2008). Estudio sobre la inserción laboral de los licenciados en Ciencias Actuariales y Financieras por la Universidad de Málaga (1997-2008). *Cuadernos de Ciencias Económicas y Empresariales. Papeles de Trabajo*, 36, pp. 1-22.
- [21] Fernández Morales, A., Moreno Ruiz, R. (2011). Situación laboral y profesional de los licenciados en Ciencias Actuariales y Financieras por la Universidad de Málaga. *III Congreso Ibérico de Actuarios*. Madrid: Instituto de Actuarios Españoles.
- [22] Fernández Morales, A. (1997). *Características del alumnado de la Universidad de Málaga*, Universidad de Málaga.
- [23] Otero Moreno, J. M., Fernández Morales, A., Isla Castillo, F. (2002). *El deporte andaluz en cifras 2002*. Sevilla: Consejería de Turismo y Deporte, Junta de Andalucía.
- [24] Otero Moreno, J. M., Fernández Morales, A., Isla Castillo, F. (2004). *Hábitos y actitudes de los andaluces ante el deporte 2002*. Málaga: Observatorio del Deporte Andaluz.
- [25] 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>.

- [26] 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>.
- [27] 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.
- [28] 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
- [29] 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.
- [30] Fernández-Morales, A., Mayorga-Toledano, M. C. (2018). New accommodation models in city tourism: The case of Airbnb in Málaga. III Spring Symposium on Challenges in Tourism Development. Instituto Universitario de Turismo y Desarrollo Económico Sostenible, Universidad de Las Palmas de Gran Canaria.
- [31] Fernández-Morales, A., Mayorga-Toledano, M. C. (2020). Characterization and concentration of Airbnb supply in Málaga. *Revista de Estudios Regionales*, 119, pp. 141-170.
- [32] Mayorga-Toledano, M. C. (2019). La intermediación en línea de las plataformas: el caso de Airbnb. In *Las viviendas vacacionales: Entre la economía colaborativa y la actividad mercantil*, 51-70. Madrid: Dykinson. DOI: 10.2307/j.ctvk8vzdt.7