

Casos de Estudio de Verificación de Hipótesis para Turismo

Antonio Fernández Morales

Departamento de Economía Aplicada (Estadística y Econometría)
Universidad de Málaga, 2016



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 este trabajo se incluyen diversos casos de estudio que utilizan las técnicas básicas de verificación de hipótesis estadísticas paramétricas más comunes. La elaboración de estos casos se ha realizado con un enfoque centrado en los estadísticos de mayor interés en la práctica profesional y la investigación en turismo.

Las hipótesis que se trabajan en los casos de estudio hacen referencia a los parámetros de la población más comunes: la media, la varianza y la proporción. Se proponen casos con poblaciones normales o con muestras grandes que permitan la aproximación normal. Además, se hace referencia a varios escenarios muestrales:

- Una muestra
- Dos muestras independientes
- Dos muestras relacionadas

La aproximación metodológica que se ha empleado es esencialmente aplicada, con la principal pretensión de facilitar una comprensión contextualizada en la experiencia práctica, tanto para asignaturas de grado como de posgrado. De forma complementaria, se puede completar el contenido teórico en manuales como [1]. Respecto a la ampliación de técnicas multivariantes, se puede acudir a [2] para un contenido teórico y práctico adicional.

Para la resolución de los cálculos es recomendable acudir a algún software estadístico disponible, aunque también se ofrecen varias aplicaciones interactivas de libre acceso que sirven de ayuda en diversos aspectos concretos de los casos propuestos, [3], [4], [5].

Para concluir, la estructura de los casos se ha elaborado con un diseño dirigido a la actividad individualizada. No obstante, es posible emplear un enfoque grupal para fomentar las competencias de tipo colaborativo entre el alumnado, [6].

Caso 1

Un equipo de técnicos utiliza para su modelo de impactos económicos del turismo en la ciudad de Málaga una distribución de probabilidad normal para la variable X ="Gasto diario en la ciudad de los cruceristas, medido en euros", con una media de 88 euros [7] y una desviación típica de 40 euros:

$$X \sim N(\mu = 88, \sigma = 40)$$

No obstante, dados los recientes cambios en este mercado, desea revisar algunos elementos del modelo, entre los que se encuentra la media de X . Para ello desea realizar una verificación de hipótesis para contrastar si la media de gasto diario de los cruceristas se mantiene en 88 euros o por el contrario ha cambiado, utilizando una muestra aleatoria de tamaño $n = 25$.

A Establezca las hipótesis del test y diga de qué tipo son

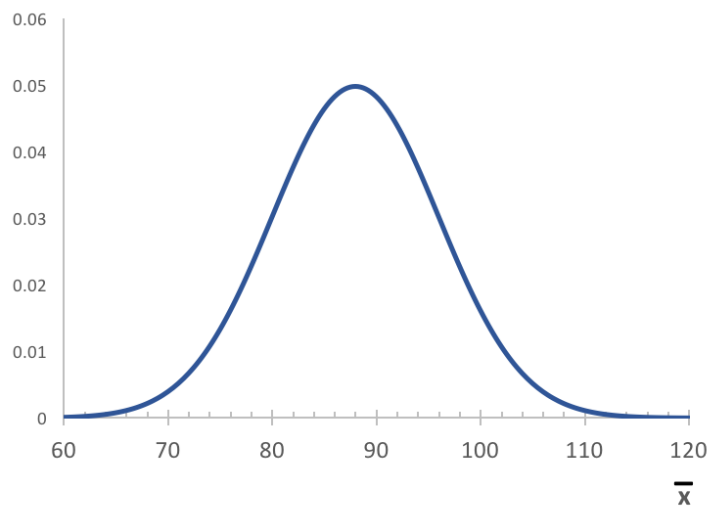
	Hipótesis	
H_0 :	<input type="checkbox"/> Hipótesis paramétrica	<input type="checkbox"/> Hipótesis no paramétrica
H_1 :	<input type="checkbox"/> Hipótesis paramétrica	<input type="checkbox"/> Hipótesis unilateral izquierda
	<input type="checkbox"/> Hipótesis no paramétrica	<input type="checkbox"/> Hipótesis unilateral derecha
		<input type="checkbox"/> Hipótesis bilateral

B ¿Que distribución de probabilidad tiene la media de la muestra si H_0 es cierta?

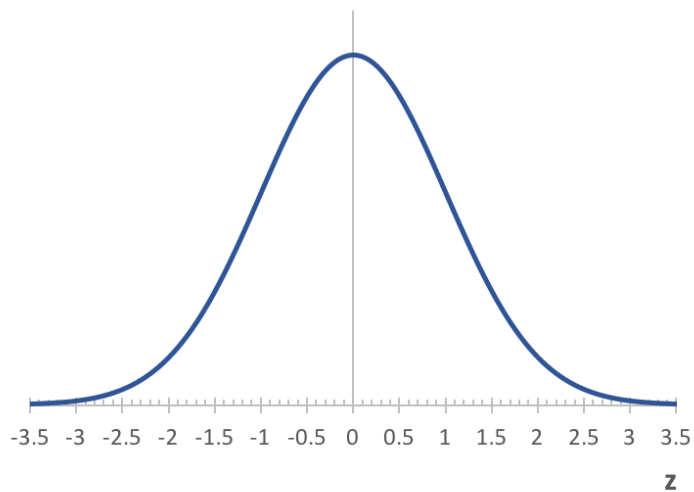
- $N(\mu_{\bar{x}} = 88, \sigma_{\bar{x}} = 40)$
- $N(\mu_{\bar{x}} = 88, \sigma_{\bar{x}} = \frac{40}{25})$
- $N(\mu_{\bar{x}} = 88, \sigma_{\bar{x}} = \frac{40}{5})$
- Ninguna de las anteriores

C ¿Cuáles son el 5% de los valores de la media muestral menos probables en caso de ser cierta la hipótesis nula? Dibújelos en el gráfico.

Sabemos que si H_0 es cierta, $P(\bar{x} < 73,32) = 0,025$ $P(\bar{x} < 103,68) = 0,975$



D Establezca la región crítica del test sobre la distribución $N(0,1)$ con un nivel de significación $\alpha = 0,05$ y señale sus valores en el gráfico.



E Con la siguiente información sobre la región crítica del test que hemos definido, C , y señale la(s) respuesta(s) correcta(s).

$$P(\bar{x} \in C | \mu = 88) = 0,05$$

$$P(\bar{x} \in C | \mu = 70) = 0,61$$

$$P(\bar{x} \in C | \mu = 60) = 0,94$$

- La probabilidad de cometer error de tipo I es 0,95
- La probabilidad de cometer error de tipo II si $\mu = 70$ es 0,95
- La potencia del test para detectar que $\mu = 70$ es 0,61
- La potencia del test para detectar que $\mu = 60$ es 0,94
- La probabilidad de cometer error de tipo II si $\mu = 70$ es 0,39
- El error de tipo I es 0,05
- La probabilidad de cometer error de tipo II si $\mu = 70$ es 0,61
- La probabilidad de cometer error de tipo II si $\mu = 60$ es 0,94

F Se ha obtenido la muestra de 25 cruceristas con el resultado de una media muestral igual a 80. Calcule el valor del estadístico de prueba tipificado y señale la(s) respuesta(s) correcta(s)

$z =$ _____

- Podemos rechazar la hipótesis nula
- Queda demostrado que el gasto medio diario es de 88 euros
- No hay suficiente evidencia empírica en contra de la hipótesis nula para rechazarla
- Queda demostrado que el gasto diario medio es diferente a 88 euros
- No podemos aceptar que el gasto medio diario es de 88 euros
- No podemos rechazar que el gasto medio diario es de 88 euros

Caso 2

La repercusión que el retraso en la recogida del equipaje del transporte aéreo tiene sobre la responsabilidad de las compañías ha sido estudiada en [8]. Los datos disponibles en [9] nos permiten suponer que la variable X: "Tiempo de espera del pasajero para la recogida del equipaje, medido en minutos" sigue una distribución normal con media 18 minutos ($\mu = 18$). Sin embargo, las recientes inversiones en nuestro aeropuerto nos hacen dudar de dicha cifra, que nos parece muy elevada, por lo que tomaremos una muestra aleatoria de tamaño $n = 20$ para contrastar dicha hipótesis.

Los estimadores muestrales obtenidos son

$$\bar{x} = 15,500 \qquad s = 12,684$$

A Establezca las hipótesis a contrastar:

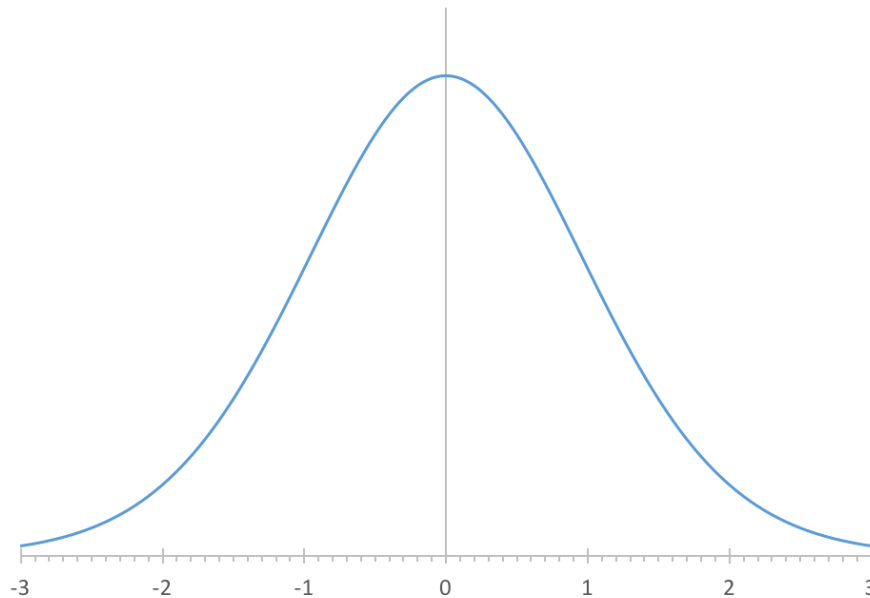
- H_0 : _____
- H_1 : _____

B ¿Cuál es el estadístico adecuado para este contraste y qué distribución de probabilidad sigue si la hipótesis nula es cierta?

C Dibuje en el gráfico la región crítica para este contraste con un nivel de significación $\alpha = 0,05$.

Conocemos las siguientes probabilidades

$$P(t_{19} < -2,093) = 0,025 \quad P(t_{19} < -1,729) = 0,05 \quad P(t_{19} < -1,328) = 0,10$$



D Calcule el valor observado del estadístico del contraste e indique su posición en la figura anterior

E La conclusión de este contraste es:
Señale la(s) respuesta(s) correcta(s)

- Podemos rechazar la hipótesis nula
- Queda demostrado que el retraso medio es 18 minutos
- No hay suficiente evidencia empírica en contra de la hipótesis nula para rechazarla
- Queda demostrado que el retraso medio es inferior a 18 minutos
- Aceptamos que el retraso medio es 18 minutos

F Realice el contraste para $\alpha = 0,10$ calculando el *p-value*.

p-value = _____

Caso 3

En el artículo [10] se estudia el turismo receptor en las regiones de Inglaterra, y se menciona que el 13% de los visitantes de la región de Yorkshire lo hacen por motivo de visita a familiares y amigos (VFR).

Queremos verificar con datos más actualizados si esta cifra se mantiene o ha cambiado. Para ello se toma una muestra aleatoria simple de tamaño $n = 200$ de visitantes de la región de Yorkshire y denominamos \hat{p} a la proporción de individuos de la muestra cuya visita es por motivo de visita a familiares y amigos (VFR). El resultado obtenido es $\hat{p} = 0,19$.

A Establezca las hipótesis a contrastar:

- H_0 : _____
- H_1 : _____

B ¿Cuál es el estadístico de prueba de este contraste y cómo se distribuye si la hipótesis nula es cierta?

C Sabiendo que el estadístico muestral tipificado es igual a 0,024 y que su *p-value* bilateral es igual a 0,012, ¿cuál es la conclusión de este contraste?

Caso 4

La duración de la visita a las ciudades de Palermo y Dubrovnik por los cruceristas que hacen escala en sus respectivos puertos han sido estudiados en [11], [12]. Deseamos verificar si hay diferencias significativas entre la duración media de la visita en cada ciudad (asumiremos normalidad en ambas poblaciones). Para ello se ha tomado una muestra independiente en ambas ciudades, cuyos resultados muestrales figuran en la tabla.

Datos muestrales			
Ciudad	n	\bar{x}	s
Palermo	278	3,87	1,80
Dubrovnik	51	3,77	1,15

A Establezca las hipótesis a contrastar:

- H_0 : _____
- H_1 : _____

B Verifique con $\alpha = 0,05$ si existe evidencia empírica suficiente en contra de la hipótesis de igualdad de varianzas en ambas poblaciones.

C ¿Cuál es el estadístico adecuado para este contraste y qué distribución de probabilidad sigue si la hipótesis nula es cierta?

D Calcule el valor muestral del estadístico del contraste.

E Obtenga el *p-value* del estadístico muestral.

F ¿Se puede rechazar la hipótesis nula con un nivel de significación $\alpha = 0,05$? Interprete el resultado.

Caso 5

Se ha realizado un estudio acerca del comportamiento de gasto de los visitantes que acuden a un evento deportivo en Miami (*ING Miami Marathon and Half Marathon*) dependiendo del grado de familiaridad con el mismo [13]. Queremos verificar si el gasto total realizado en compras es mayor entre los visitantes que repiten su experiencia. Para ello tomamos una muestra aleatoria de 30 visitantes y se registra para cada uno de ellos dos variables:

X: "Gasto realizado en compras en la visita actual"

Y: "Gasto realizado en compras en la primera visita"

Asumimos que X e Y siguen modelos de probabilidad normales.

Los resultados muestrales son:

Datos muestrales			
Ciudad	n	\bar{x}	s
X	30	310	156
Y	30	229	150

A Establezca las hipótesis a contrastar:

- H_0 : _____
- H_1 : _____

B ¿Cuál es el estadístico adecuado para este contraste y qué distribución de probabilidad sigue si la hipótesis nula es cierta?

C Calcule el valor muestral del estadístico del contraste.

D Obtenga el *p-value* del estadístico muestral.

E ¿Se puede rechazar la hipótesis nula con un nivel de significación $\alpha = 0,05$?
Interprete el resultado.

Caso 6

Un reciente estudio sobre la actitud hacia el medio ambiente de los turistas de golf [14] realizado con información en Andalucía menciona un mayor porcentaje de individuos con estudios superiores en este segmento que en la población general. Deseamos verificar si hay diferencias significativas en la proporción de turistas de golf con estudios superiores en Andalucía y en otra región española. Para ello se ha tomado una muestra aleatoria de 600 turistas de golf en cada una de las regiones con los resultados siguientes:

Datos muestrales		
Región	n	Porcentaje con estudios superiores
Andalucía	600	68
Otra	600	62

A Establezca las hipótesis a contrastar:

- H_0 : _____
- H_1 : _____

B ¿Cuál es el estadístico adecuado para este contraste y qué distribución de probabilidad sigue si la hipótesis nula es cierta?

C Calcule el valor muestral del estadístico del contraste.

D Obtenga el *p-value* del estadístico muestral.

E ¿Se puede rechazar la hipótesis nula con un nivel de significación $\alpha = 0,05$?
Interprete el resultado.

Referencias

- [1] Newbold, P., Carlson, W.L., Thorne, B.M. (2008). *Estadística para administración y economía*. Madrid: Pearson.
- [2] Fernández Morales, A. (2009). *Técnicas de análisis multivariante aplicadas al turismo*. Málaga: Canales 7.
- [3] Fernández Morales, A. (2009). CALCUPROB An on-line interactive calculator of probabilities. University of Málaga. Disponible en: <http://hdl.handle.net/10630/5071>
- [4] 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>
- [5] Fernández Morales, A. (2016). Introduction to measures of inequality and concentration in tourism. University of Málaga. Disponible en: <http://hdl.handle.net/10630/11035>
- [6] Fernández Morales, A., Mayorga Toledano, M.C. (2013). Developing Creativity and Innovation through Collaborative Projects. *Interdisciplinary Studies Journal 2* (3), 70-82.
- [7] Fernández Morales, A., Martín Carrasco, Y. (2014). Concentración e impacto estacional del turismo de cruceros en Málaga. *Revista de Estudios Regionales* 101, 43-70.
- [8] Mayorga Toledano, M.C. (2015). El transporte de equipaje: ámbito y extensión de la responsabilidad del porteador aéreo por pérdida, destrucción, daños o retraso. En Guerrero Lebrón, M.J. (dir.) *La responsabilidad del transportista aéreo y la protección de los pasajeros*. Madrid: Marcial Pons, 163-192.
- [9] SITA(2016). 2016 Air transport industry insights. The baggage report. Disponible en <https://es.sita.aero/resources/type/surveys-reports/baggage-report-2015>
- [10] Fernández Morales, A., Cisneros Martínez, J.D., McCabe, S. (2016). Seasonal concentration of tourism demand: Decomposition analysis and marketing implications. *Tourism Management* 56, 172-190.
- [11] Ferrante, M., De Cantis, S., Kahani, A., Shoval, N. (2016). Cruise passengers' behavior at the destination: Investigation using GPS technology. *Tourism Management* 52, 133-150.

- [12] Ferrante, M., De Cantis, S., Shoval, N. (2016). A general framework for collecting and analysing the tracking data of cruise passengers at the destination. *Current issues in Tourism*, 1-26.
- [13] Lee, S.K., Jee, W.S., Funk, D.C., Jordan, J.S. (2015). Analysis of attendees' expenditure patterns to recurring annual events: Examining the joint effects of repeat attendance and travel distance. *Tourism Management* 46, 177-186.
- [14] López-Bonilla, L.M., López-Bonilla, J.M. (2015). From the new environmental paradigm to the brief ecological paradigm: a revised scale in golf tourism. *Anatolia*, 27 (2), 227–236.