

Solución de problemas de prueba de hipótesis para promedio y proporción usando R *

Wilson Rodríguez Calderón
Myriam R. Pallares Muñoz

Escuela Superior de Administración Pública - sede Neiva
Universidad Surcolombiana

Resumen: Este trabajo presenta el planteamiento, planeación, montaje y desarrollo de tres casos de estudio del tema de prueba de hipótesis para promedio y proporción en el aula para estudiantes universitarios. Se trata del desarrollo de tres problemas modelo de prueba de hipótesis estadística para promedio (2) y proporción (1) usando como herramienta computacional el paquete estadístico libre R y como herramienta de verificación o comparación la hoja de cálculo Excel. En la actualidad a nivel mundial existen paquetes estadísticos especializados comerciales de gran capacidad, no obstante, se busca incentivar el uso de herramientas libres de disposición inmediata y amigables con los estudiantes que permitan descargar las fases de cálculo de los problemas, para privilegiar el esfuerzo y la atención en el análisis más que en el cálculo mismo. Para los casos se usó un procedimiento de revisión del estado del conocimiento del tema, planteamiento de los problemas, desarrollo de los códigos R y las hojas de cálculo en Excel por parte de los investigadores, prueba en el aula y trabajo en casa de los estudiantes y por último las conclusiones y recomendaciones. Los resultados muestran ventajas importantes de usar herramientas libres como R en el apoyo didáctico y de trabajo autónomo del alumno para apalancar experiencias de investigación formativa en el aula, experiencias de educación STEM y aprendizaje basado en problemas.

Palabras Clave: Pruebas de hipótesis, R, Estadística Inferencial.

Resultados

Problema 1. Aplicación en prueba de hipótesis para descarga promedio en máquina expendedora de gaseosas

Se trata de la aplicación de R al estudio de una prueba de hipótesis para descarga promedio en onzas por taza. La pregunta fundamental es: ¿Los datos presentan suficiente evidencia para indicar que la descarga media difiere de 7 onzas? (Wackerly et al., 2010). Si bien R posee funciones para pruebas de hipótesis estadística, para la investigación se desarrolló una herramienta que pudiera atender de manera general las variantes de los problemas de prueba de hipótesis para promedio y proporción. A continuación se muestra la solución en R y Excel encontrada y verificada.

En la Figura 1 se muestra la solución en R con su respectiva gráfica de las zonas de rechazo de la hipótesis nula marcadas en verde, por otra parte, la Figura 2 muestra la verificación satisfactoria en Excel que muestra como la herramienta desarrollada se adapta a la solución del problema. El R muestra gran versatilidad y la capacidad de rápidamente graficar las zonas de rechazo de la hipótesis nula, aspecto muy útil para la didáctica y comprensión del problema.

*Autor de contacto: wilson.rodriguezc@esap.edu.co

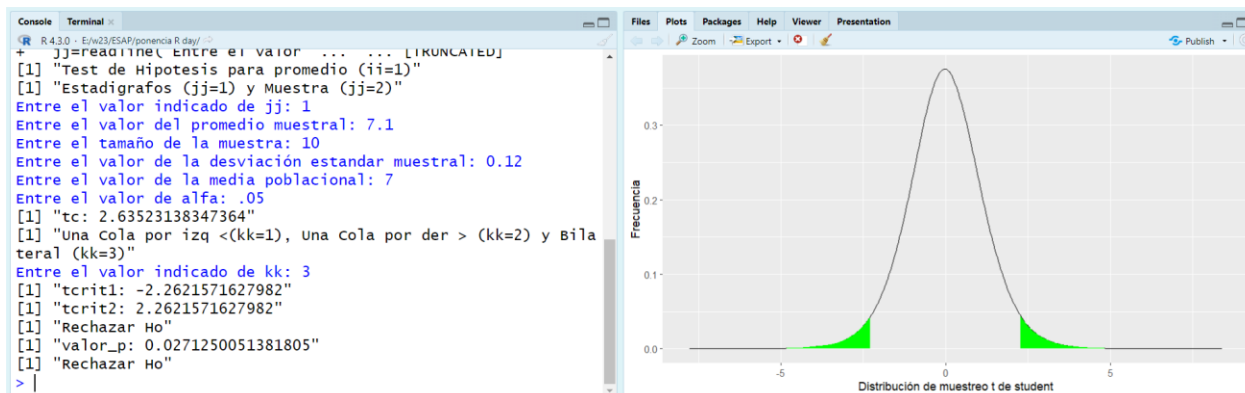


Figura 1: Problema de prueba de hipótesis para descarga promedio en máquina expendedora de gaseosas resuelto en R. Fuente: Autores

MÁQUINA EXPENDEDORA DE GASEOSAS									
X: Descarga en onzas de líquido por taza									
Ho: $\mu = 7$ (La descarga promedio de la máquina expendedora de gaseosas es de 7 onzas/taza)									
Ha: $\mu \neq 7$ (La descarga promedio de la máquina expendedora de gaseosas es diferente de 7 onzas/taza)									
\bar{x}	7.100	Media de la muestra				\bar{x}	7.100	Media de la muestra	
n	10	Tamaño de la muestra				n	10	Tamaño de la muestra	
s	0.12	Desviación estándar de la muestra				s	0.12	Desviación estándar de la muestra	
μ_0	7	Media de la población				μ_0	7	Media de la población	
α	0.05	Nivel de significancia				α	0.05	Nivel de significancia	
$\alpha/2$	0.025					$\alpha/2$	0.025		
t_c	2.635231383					t_c	2.635231383		
$t_{\alpha/2;n-1}$	2.262157163								
$t_{1-\alpha/2;n-1}$	-2.262157163								
prueba	VERDADERO					Valor-p	0.027125005	$p=2*P(t>t_c)$	
	Rechazar Ho					prueba	VERDADERO	$p<\alpha$	
							Rechazar Ho		
<div> <p>La conclusión en el contexto del problema implica que, al nivel de significancia de 5%, hay evidencia suficiente para creer que descarga promedio de la máquina de gasesosas es diferente a 7 onzas/taza.</p> </div>									

Figura 2: Problema de prueba de hipótesis para descarga promedio en máquina expendedora de gaseosas resuelto en Excel. Fuente: Autores

Problema 2. Aplicación en prueba de hipótesis para capacidad promedio de difusión de monóxido de carbono en fumadores

Se trata de la aplicación de R al estudio de una prueba de hipótesis para capacidad promedio de difusión de monóxido de carbono en fumadores. La pregunta fundamental es: ¿Estos datos indican que la lectura media de DL para fumadores actuales es menor que 100, el promedio de lectura de DL para los no fumadores? (Wackerly et al., 2010).

En la Figura 3 se muestra la solución en R con su respectiva gráfica de la zona de rechazo de la hipótesis nula marcada en verde, por otra parte, la Figura 4 muestra la verificación satisfactoria en Excel. En este caso se carga una muestra de 20 datos, en lugar de suministrar los estadígrafos de la muestra, que para el caso son calculados y usados para la solución.

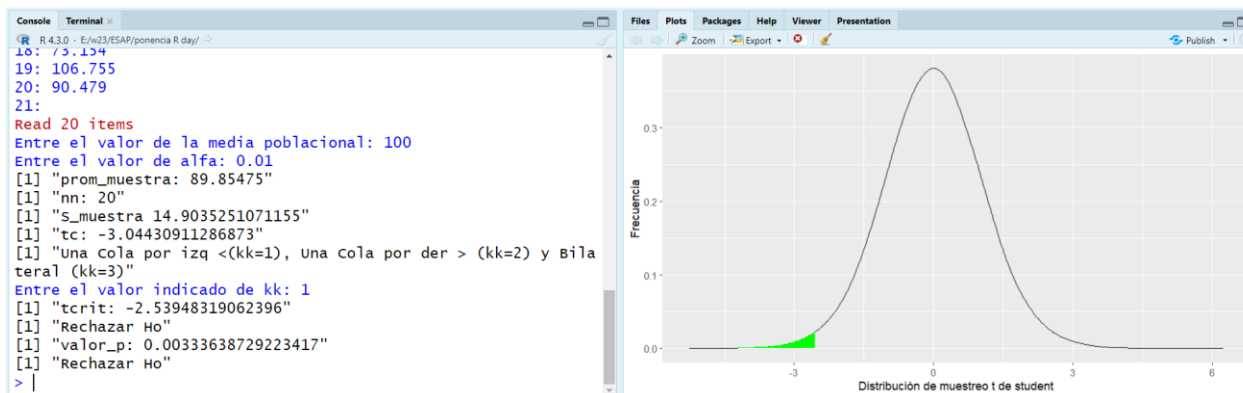


Figura 3: Problema de prueba de hipótesis para capacidad promedio de difusión de monóxido de carbono en fumadores resuelto en R. Fuente: Autores

CAPACIDAD DE DIFUSIÓN DE MONOXIDO DE CARBONO (DL) EN FUMADORES									
X: Capacidad de difusión de monoxido de carbono (DL) en fumadores									DL
$H_0: \mu = 100$ (El promedio de capacidad de difusión de monóxido de carbono (DL) en fumadores es de 100)									1 103.768
$H_a: \mu < 100$ (El promedio de capacidad de difusión de monóxido de carbono (DL) en fumadores es menor que 100)									2 88.602
\bar{x} 89.855	Media de la muestra	\bar{x} 89.855	Media de la muestra						3 73.003
n 20	Tamaño de la muestra	n 20	Tamaño de la muestra						4 123.086
s 14.90352511	Desviación estándar de la muestra	s 14.90352511	Desviación estándar de la muestra						5 91.052
μ_0 100	Media de la población	μ_0 100	Media de la población						6 92.295
α 0.01	Nivel de significancia	α 0.01	Nivel de significancia						7 61.675
t_c -3.04430911		t_c -3.044309113							8 90.677
$t_{1-\alpha/2, n-1}$ -2.53948319	Punto crítico en la t de student $t_{0.95; 99}$	Valor-p 0.003336387	$p = P(T_{n-1} < t_c)$						9 84.023
prueba VERDADERO	$t_c < t_{1-\alpha/2, n-1}$	prueba VERDADERO	$p < \alpha$						10 76.014
Rechazar H_0		Rechazar H_0							11 100.615
La conclusión en el contexto del problema implica que, al nivel de significancia de 1%, hay evidencia suficiente para creer que la capacidad de difusión de monoxido de carbono (DL) promedio en fumadores es menor que 100 (el promedio de lectura del DL para no fumadores).									12 88.017
									13 71.210
									14 82.115
									15 89.222
									16 102.754
									17 108.579
									18 73.154
									19 106.755
									20 90.479
									prom 89.855
									s 14.90353

Figura 4: Problema de prueba de hipótesis para capacidad promedio de difusión de monóxido de carbono en fumadores resuelto en Excel. Fuente: Autores

Problema 3. Aplicación en prueba de hipótesis para proporción de percepción sobre la calidad del medio ambiente del país

Se trata de la aplicación de R al estudio de una prueba de hipótesis para la proporción de personas con percepción regular o mala sobre la calidad del medio ambiente del país. La pregunta fundamental es: ¿Hay evidencia convincente de que una mayoría de adultos de la nación piensa que la calidad del medio ambiente es regular o mala? (Wackerly et al., 2010).

En la Figura 5 se muestra la solución en R con su respectiva gráfica de la zona de rechazo de la hipótesis nula marcada en verde, por otra parte, la Figura 6 muestra la verificación satisfactoria en Excel. En este caso se entrega el estadígrafo de proporción muestral directamente, junto con el tamaño de muestra.

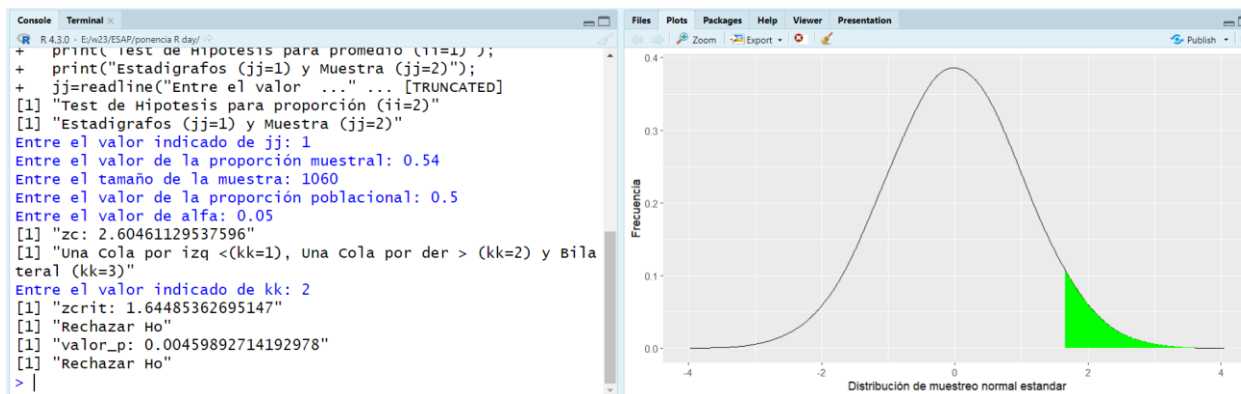


Figura 5: Problema de prueba de hipótesis para proporción de percepción sobre la calidad del medio ambiente del país resuelto en R. Fuente: Autores

PERCEPCIÓN SOBRE LA CALIDAD DEL MEDIO AMBIENTE DEL PAÍS					
X: Número de personas que consideran que la calidad del medio ambiente es regular o mala					
Ho: $\pi = 0.5$ (La proporción de personas que consideran que la calidad del medio ambiente del país es regular o mala es del 50%)					
Ha: $\pi > 0.5$ (La proporción de personas que consideran que la calidad del medio ambiente del país es regular o mala es mayor del 50%)					
x	572	# individuos que cumplen la característica x	x	572	# individuos que cumplen la característica x
n	1060	Tamaño de la muestra	n	1060	Tamaño de la muestra
p	0.54	Proporción de la muestra	p	0.54	Proporción de la muestra
π_0	0.5	Proporción de la población	π_0	0.5	Proporción de la población
α	0.05	Nivel de significancia	α	0.05	Nivel de significancia
Z_c	2.604611295		Z_c	2.604611295	
Z_α	1.644853627				
			Valor-p	0.004598927	$p=P(Z>Z_c)$
prueba	VERDADERO	$Z_c > Z_\alpha$	prueba	VERDADERO	$p < \alpha$
	Rechazar Ho			Rechazar Ho	
<div> $Z_c = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1 - \pi_0)}{n}}}$ </div>					
<div> <p>La conclusión lleva a creer que, al nivel de significancia de 5%, que los datos proporcionan evidencia suficiente para pensar que la mayoría de personas piensan que la calidad del medio ambiente es regular o mala.</p> </div>					

Figura 6: Problema de prueba de hipótesis para proporción de percepción sobre la calidad del medio ambiente del país resuelto en Excel. Fuente: Autores

Referencias

R Core Team. 2019. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

URL: <https://www.R-project.org/>

Wackerly, D., L. Scheaffer, R. y Mendenhall, W. (2010). *Estadística matemática con aplicaciones*. Cuajimalpa de Morelos, Mexico: Cengage Learning Editores.