Punto seis de la tarea uno de Estadística No Paramétrica

Simón Cuartas Rendón

Septiembre 28 de 2022

Punto seis

- **Enunciado.** A veintidós clientes en una tienda de comestibles se les preguntó por el gusto entre dos tipos de queso y estos declararon sus preferencias así: siete clientes prefirieron un tipo, doce clientes prefirieron el otro tipo y tres no tuvieron preferencia. ¿Esto indica una diferencia significativa en la preferencia de los dos tipos de queso? Usar un nivel de significancia de $\alpha = 0.05$ e interpretar.

Para este ejercicio se va a etiquetar al queso que fue preferido por siete clientes como el queso A, mientras que aquel que gustó más a doce clientes se le catalogará como el queso B. Con esto presente, es posible emplear la **relación entre los tests de McNemar y del signo** para determinar si existe una diferencia significativa en la preferencia entre los quesos tipo A y tipo B.

Para ello, se considerará que "$+$" representa el evento en que el queso tipo B es más apeticido que el queso tipo B, "$-$" la situación contraria y "0" cuando se presenta un empate, que puede darse cuando no se prefiere ninguno de los quesos especialmente.

- **Hipótesis.**

Teniendo en cuenta que se quiere probar si hay diferencia significativa entre la preferencia hacia un tipo particular de queso, y no determinar si uno es más o menos preferido que el otro, entonces basta con emplear un test de dos colas, de forma que el juego de hipótesis es el que se muestra a continuación:

$$\begin{cases} H_0: \ Pr(+) = Pr(-) \ \forall \ (X_i, \ Y_i) \\ H_1: \ \exists \ i \mid Pr(+) \neq Pr(-), \ i = 1, \ldots, 100 \end{cases}$$

- **Estadística de prueba**

De acuerdo con la información del enunciado, el evento denotado con "$+$" sucede doce veces, el evento denotado con "$-$" ocurre en siete ocasiones y los empates se dan en tres veces. Además, como el estadístico de prueba T denota la cantidad de veces en las que el queso tipo B es preferido sobre el queso tipo A, esto es, el evento "$+$", entonces $T = 12$. Por otro lado, se denota a n como el número de veces en que no ocurre un empate o se elige un queso sobre otro, de tal suerte que $n = 12 + 7 = 19$.

- **Regón crítica**

Como $n = 19 < 20$, entonces se debe emplear la región crítica definida por la estadística de prueba como regla de decisión, la cual se define como sigue:

$$R.C. = \{T \mid T \leq n \quad \vee \quad T \geq n - t\}$$

donde t es el cuantil de una $Y \sim Binomial(n, \; ^1\!/_2)$ tal que $Pr(Y < t) \leq {}^\alpha\!/_2$

- **Cálculos**

Se va iniciar calculando el valor t:

```
n <- 19
prob <- 1/2
alfa <- 0.05

t <- qbinom((alfa/2), size = n, prob = 1/2, lower.tail = TRUE)
t
```

```
## [1] 5
```

Es decir, $t = 5$. Si bien $T = 12 < 14 = n - t$, se cumple que $T = 12 < 19$, por lo que se está en la región crítica, lo que implica que se debe rechazar la hipótesis nula. Así, se tiene que no hay evidencia muestral suficiente para sugerir que la preferencia por los quesos tipo A y tipo B es la misma, con una significancia del cinco por ciento, lo que en la práctica significa que uno de los dos tipos de queso es más preferido que el otro.

Implementación en R

Es posible verificar los cálculos anteriores al emplear la función `binom.test()` del paquete **`stats`* de R, ya que el test del signo se puede ver como un caso particular del test binomial, llegando al siguiente resultado:

```
# Parámetros importantes ==============================================
exitos <- 12        # Número de veces que ocurrió el evento 'siete'.
ensayos <- 15       # Número de lanzamientos o ensayos realizados.
prob_teorica <- 0.5 # Probabilidad teórica de que el evento de interés
                    # suceda en 180 ensayos

# Función ==============================================================
binom.test(x = exitos, n = ensayos, p = prob_teorica,
           alternative = 'two.sided')
```

```
##
##  Exact binomial test
##
## data:  exitos and ensayos
## number of successes = 12, number of trials = 15, p-value = 0.03516
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.5191089 0.9566880
## sample estimates:
## probability of success
##                    0.8
```

Como se observa, $V_p = 0.0351 < 0.05 = \alpha$, por lo que se rechaza la hipótesis nula y se deriva la misma conclusión que con los cálculos manuales.