

Ejercicios de estadística no paramétrica

Simón Cuartas Rendón

2022-09-06

Punto uno

- **Enunciado.** En un juego se lanzó un par de dados 180 veces, de las cuales se produjo el evento ‘siete’ 38 veces. Si los dados no están cargados, la probabilidad de sacar ‘siete’ es $1/6$, mientras que si se cargan, la probabilidad es mayor. ¿Hay evidencia muestral para sugerir que los dados están cargados?***

En este caso es razonable emplear el **test de la binomial**, ya que se desea realizar inferencia sobre una proporción o una probabilidad, y es que esta situación puede verse como un experimento binomial, en donde se realizan X_1, X_2, \dots, X_{180} lanzamientos de dos dados (ensayos Bernoulli), cada uno de estos independientes entre sí, siendo el evento de interés que la suma de los puntos resultantes en las caras superiores de los dados sea siete, de tal suerte que en el caso de dos dados no cargados esta probabilidad es de $P^* = 1/6$, de forma que esta última probabilidad se puede considerar como la probabilidad de éxito. En ese sentido, se cumple que $X_i \sim \text{Bernoulli}(p = 1/6)$, $i = 1, 2, \dots, 180$, y por tanto, se puede definir la variable aleatoria Y binomial asociada a estas variables aleatorias Bernoulli como $Y \sim \text{Binomial}(n = 180, p = 1/6)$. Así pues, sea P la probabilidad de obtener un ‘siete’ al lanzar los dos dados usados en el juego.

- **Hipótesis.**

$$\begin{cases} H_0 : P = \frac{1}{6} \\ H_1 : P > \frac{1}{6} \end{cases}$$

Nótese que se está considerando un test de cola izquierda ya que el enunciado indica que si los dados están cargando, la probabilidad de obtener un siete es mayor a $1/6$, y esto es justamente lo que se quiere probar.

- **Estadística de prueba.** Sea T la estadística de prueba para este test y O_1 el número de éxitos observados en el ensayo (o el juego). Así, se tiene:

$$T = O_1$$

- **Región crítica.**

$$R_c = \{T : T > t\}, \text{ donde } t \text{ es tal que } P(Y > t) \leq \alpha, \text{ donde } Y \sim \text{Binomial}(180, 1/6)$$

Y como $n = 180 > 20$, entonces el valor p V_p asociado está dado por:

$$V_p \approx Pr \left(Z \geq \frac{T_{obs} - nP^* + 0.5}{\sqrt{nP^*(1 - P^*)}} \right)$$

* **Cálculos.** Así pues, se va a considerar un nivel de significancia de $\alpha = 0.05$. Así, se va a comenzar hallando el t de la región crítica, teniendo entonces que:

```
t <- qbinom(0.05, 180, 1/6, lower.tail = FALSE); t
```

```
## [1] 38
```

$$R_c = \{T : T \leq 38\}$$

Pero se tiene que $T_{obs} = O_1 = 38$, y como es falso que $T_{obs} = 38 > 38 = t$, entonces no se está en la región crítica, por lo que no se rechaza la hipótesis nula. De igual forma, usando el valor p :

```
# Parámetros
t_obs <- 38
n <- 180
p_star <- 1/6

# Probabilidad
z_calc <- (t_obs - n * p_star + 0.5) / sqrt(n * p_star * (1 - p_star))
Vp <- pnorm(z_calc, mean = 0, sd = 1, lower.tail = FALSE)
Vp
```

```
## [1] 0.04456546
```

Como se puede ver, se tiene que $V_p = 0.04 < \alpha = 0.05$. Se tiene entonces una situación contradictoria entre la estadística de prueba y el valor p . Nótese que esto se debe a que se está en una situación límite. En este caso es razonable considerar que los datos no están cargados, esto es no, no rechazar la hipótesis nula, ya que típicamente se considera que cometer el error tipo I (rechazar la hipótesis nula cuando debería hacerse) es peor que el error tipo II (no rechazar la hipótesis nula cuando no debería hacerse).

- **PREGUNTAR.**

Punto dos

- **Enunciado.** Un grupo de personas reportó al consejo del pueblo que por lo menos el 60 % de los residentes del pueblo estaban a favor de la emisión de un bono. El consejo del pueblo tomó una muestra aleatoria de cien personas y les preguntó si estaban a favor o no de la emisión del bono, a lo que 48 personas dijeron que sí. ¿Es el reporte del grupo cívico razonable? Usar $\alpha = 0.05$ e interpretar.

En este caso es nuevamente adecuado emplear el **test de la binomial**, ya que se desea realizar inferencia sobre una proporción o una probabilidad, y es que esta situación puede verse como un experimento binomial, en donde se realizan X_1, X_2, \dots, X_{100} cuestionarios a ciudadanos del pueblo (experimentos Bernoulli), cada uno de estos independientes entre sí, siendo el evento de interés que los ciudadanos respondan de forma afirmativa a la pregunta asociada a la emisión del bono, siendo la afirmación del grupo de ciudadanos que esto sucederá un $P^* = 6/10 = 60$ de las veces, de forma que esta última probabilidad se puede considerar como la probabilidad de éxito. En ese sentido, se cumple que $X_i \sim \text{Bernoulli}(p = 6/10)$, $i = 1, 2, \dots, 100$, y por tanto, se puede definir la variable aleatoria Y binomial asociada a estas variables aleatorias Bernoulli como $Y \sim \text{Binomial}(n = 100, p = 6/10)$. Así pues, sea P la probabilidad de que un ciudadano diga que sí está de acuerdo con la emisión del bono en el pueblo.

- **Hipótesis.**

$$\begin{cases} H_0 : P \leq 0.6 \\ H_1 : P > 0.6 \end{cases}$$

Nótese que se está considerando un test de cola izquierda ya que el enunciado indica que el grupo cívico dice que *al menos* el 60 % de la ciudadanía del pueblo está a favor de la emisión del bono.

- **Estadística de prueba.** Sea T la estadística de prueba para este test y O_1 el número de éxitos observados en el ensayo (o el juego). Así, se tiene:

$$T = O_1$$

- **Región crítica.**

$$R_c = \{T : T > t\}, \text{ donde } t \text{ es tal que } P(Y > t) \leq \alpha, \text{ donde } Y \sim \text{Binomial}(100, 60)$$

Y como $n = 100 > 20$, entonces el valor p V_p asociado está dado por:

$$V_p \approx Pr \left(Z \geq \frac{T_{obs} - nP^* + 0.5}{\sqrt{nP^*(1 - P^*)}} \right)$$

- **Cálculos.** Así pues, se va a considerar un nivel de significancia de $\alpha = 0.05$. Así, se va a comenzar hallando el t de la región crítica, teniendo entonces que:

```
t <- qbinom(0.05, 100, 0.6, lower.tail = FALSE); t
```

```
## [1] 68
```

$$R_c = \{T : T \leq 68\}$$

Pero se tiene que $T_{obs} = O_1 = 48$, y dado que $T_{obs} = 48 < 68 = t$, entonces no se está en la región crítica, lo que implica que no se rechaza la hipótesis nula. De igual forma, usando el valor p :

```
# Parámetros
t_obs <- 48
n <- 100
p_star <- 0.6

# Probabilidad
z_calc <- (t_obs - n * p_star + 0.5) / sqrt(n * p_star * (1 - p_star))
Vp <- pnorm(z_calc, mean = 0, sd = 1, lower.tail = FALSE)
Vp
```

```
## [1] 0.9905482
```

Como se puede ver, se tiene que $V_p = 0.99 > \alpha = 0.05$, coincidiendo con la conclusión de la estadística de prueba en cuanto a no rechazar la hipótesis nula, pudiendo determinar pues que *no* existe evidencia muestral suficiente para sugerir que la proporción de ciudadanos que está a favor de la emisión del bono sea mayor a (o por lo menos del) 60 %, de forma que la afirmación del grupo de ciudadanos **carece de sustento estadístico**.