

Metodos No Paramétricos - Final

Gonzalo Barrera Borla

8/11/2019

Ejercicios

Practica 1, ej. 4

Un empresario de la industria alimenticia asegura que menos del 10% de sus frascos de café instantáneo contiene menos café del que garantiza la etiqueta. Para probar esta afirmación se eligen al azar 15 frascos de café y se pesa su contenido. Su afirmación es aceptada si a lo sumo dos frascos contienen menos café del garantizado.

- ¿Qué hipótesis se deben testear?
- ¿Cuál es el nivel de la regla de decisión planteada? ¿Le parece razonable?
- Encuentre la probabilidad de que la afirmación del empresario sea aceptada cuando el porcentaje real de frascos que contienen menos café del garantizado en la etiqueta es 5%, 10% y 20%.
- Grafique la función de potencia del test planteado inicialmente. Muestre que es insesgado.
- Con el tamaño de muestra dado, ¿es posible obtener un test de nivel 0.05? Hallar el tamaño de muestra mínimo para obtener un test de nivel 0.05, manteniendo la misma región de rechazo que el test anterior.

Sea $X_i = 1$ si el i -ésimo frasco de café tiene menos café del que garantiza la etiqueta, y 0 en caso contrario, de manera que bajo H_0 , $X_i \stackrel{iid}{\sim} \text{Ber}(p = 0.1)$ y $\sum_i X_i = T \sim \text{Bi}(15, 0.1)$. Las hipótesis a testear serán:

$$H_0 : p \geq 0.1 \quad \text{vs.} \quad H_1 : p < 0.1$$

con región de rechazo $RR = \{0, 1, 2\}$. Los puntos (b) y (c) piden el nivel de significación y la probabilidad de error de tipo II, para lo cual ya hemos generado funciones. Para el punto (c), calculamos la función de potencia en una grilla de valores de p , y la graficamos con `ggplot`:

```
n <- 15
p_null <- 0.1 # Usamos el p_null de la igualdad para calcular la significacion
RR <- 0:2
prob_rechazo <- function(n, p, RR) {
  return(sum(dbinom(RR, n, p)))
}
potencia <- prob_rechazo
significacion <- prob_rechazo
prob_etII <- function(n, p_alt, RR) { 1 - potencia(n, p_alt, RR) }
alfa <- significacion(n, p_null, RR)
plej4 <- list(
  b = alfa,
  c = list(
    etII.05 = prob_etII(n, 0.05, RR),
    etII.1 = prob_etII(n, 0.1, RR),
    etII.2 = prob_etII(n, 0.2, RR))
)
```

Notese que para esta RR , como $Pr(\text{Bi}(15, 0.1) \in \{0, 1, 2\}) = 0.816$, cuando $p = 0.1$ rechazaremos la hipótesis nula alrededor del 81.59% de las veces, lo cual no suena muy razonable. Aun así, el test Φ es insesgado cuando

$$\begin{aligned} Pr(\text{rech } H_0 | H_0 \text{ Verdadera}) &< Pr(\text{rech } H_0 | H_0 \text{ Falsa}) \\ Pr(\Phi = 1 | H_0 \text{ V}) &< Pr(\Phi = 1 | H_0 \text{ F}) \\ Pr_{\theta_0}(\Phi = 1) &< Pr_{\theta_1}(\Phi = 1) \quad \forall \quad \theta_0 \in \Theta_0, \theta_1 \in \Theta_1 \end{aligned}$$

y como Φ es una unción indicadora, la probabilidad de que valga 1 es igual a su esperanza. Considerando que la desigualdad se tiene que cumplir para todo par de elementos en la hipótesis nula y la alternativa, concluimos que un test es insesgado cuando:

$$\begin{aligned} \forall \theta_0 \in \Theta_0, \quad Pr_{\theta_0}(\Phi = 1) &\leq \sup_{\theta \in \Theta_0} E_{\theta}(\Phi) \\ \forall \theta_1 \in \Theta_1, \quad Pr_{\theta_1}(\Phi = 1) &\geq \inf_{\theta \in \Theta_1} E_{\theta}(\Phi) \\ \Phi \text{ es insesgado} &\Leftrightarrow \sup_{\theta \in \Theta_0} E_{\theta}(\Phi) < \inf_{\theta \in \Theta_1} E_{\theta}(\Phi) \end{aligned}$$

Y recordemos que la función `prob_rechazo` es, justamente, la esperanza del test, bajo H_0 y H_1 . El siguiente gráfico muestra claramente la insesgadez del test, a pesar de su pesima significación.

```
densidad <- 0.001
plej4$d <- tibble(
  p = seq(0, 1, densidad),
  potencia = map_dbl(p, ~prob_rechazo(n, ., RR))) %>%
  ggplot(aes(p, potencia)) +
  geom_line() +
  geom_vline(xintercept = p_null, alpha = 0.3) + # Referencia: p_null
  geom_hline(yintercept = alfa, alpha = 0.3) # Referencia: signif. de la RR
```

Manteniendo la RR propuesta, el punto (e) nos pide $\min n : Pr(Bi(n, 0.1) \in \{0, 1, 2\}) \leq 0.05$. Escribamos la función que lo busca:

```
n_requerido <- function(RR, alfa, p_null, max_n = 10000) {
  n_req <- 0
  while (n_req < max_n) {
    if (significacion(n_req, p_null, RR) <= alfa) {
      return(n_req)
    } else {
      n_req <- n_req + 1
    }
  }
  return(NA) # Devuelve NA si no encuentra n antes de max_n
}
plej4$e <- n_requerido(RR, 0.05, p_null)
```

Practica 2, ej. 2

Construya el gráfico de $S(\theta)$ para n impar y deduzca el estimador y el intervalo de confianza para θ .

Practica 3, ej. 8

Suponga que $\forall i = 1, \dots, n, X_i \stackrel{iid}{\sim} F \in \Omega_s$ (distribuciones simétricas con única mediana en 0).

- Usando que $g(X_1, \dots, X_n)$ y $g(-X_1, \dots, -X_n)$ tienen la misma distribución, muestre que si $g(X_1, \dots, X_n) + g(-X_1, \dots, -X_n) = \mu_0$, entonces $g(X_1, \dots, X_n)$ está simétricamente distribuída alrededor de $\mu_0/2$. Hint: Muestre que $P(g(X_1, \dots, X_n) \leq \mu_0/2 - t) = P(g(X_1, \dots, X_n) \leq \mu_0/2 + t)$.
- Aplique a) al estadístico del test de rangos signados de Wilcoxon para demostrar que T^+ tiene distribución simétrica bajo la hipótesis nula. ¿Cuál es el punto de simetría?

Demostraciones

Consistencia del test de Wilcoxon

Teorema de Proyeccion

MWW: Distribución exacta de los estadístico del test de Mann-Whitney-Wilcoxon (MWW) bajo H_0