Análisis Estadístico en Inferencia para Investigación Médica

Wilson Eduardo Jerez-Hernández

2024-03-02

Abstract

En este ejercicio, exploramos la inferencia estadística aplicada a una investigación médica centrada en determinar la probabilidad de éxito de un nuevo tratamiento médico. El objetivo es asegurar al menos un 55% de efectividad, y se utiliza la fracción de pacientes que responden positivamente como estimador clave. Analizamos las características del estimador, su distribución, y aplicamos la Aproximación de la Binomial por la Normal. Este enfoque proporciona una perspectiva estadística en el contexto médico.

keywords: Inferencia Estadística, Investigación Médica, Probabilidad de Éxito ,Fracción de Pacientes, Aproximación de la Binomial.

1. Introducción

En este ejercicio, nos enfocaremos en la inferencia estadística aplicada a una investigación médica. El objetivo es determinar la probabilidad de éxito de un nuevo tratamiento médico, asegurando al menos un 55% de efectividad. Exploraremos el proceso de inferencia y sus implicaciones en este escenario específico.

2. Descripción de la Población

Supongamos que se considera exitoso un tratamiento médico si al menos el 55% de los pacientes tratados muestran mejoría. La pregunta clave es: ¿Cuál es la probabilidad de alcanzar este umbral de efectividad cuando se selecciona aleatoriamente un grupo de $n = 100$ pacientes?

Contexto Adicional: Se estima que aproximadamente el 50% de los pacientes responderán positivamente al tratamiento.

```r
set.seed(42)
n <- 100 # Muestra
p <- 0.5 # Estimación
Datos <- rbinom(n, 1, p) # Distribución Bernoulli
Datos
```

```
##   [1] 1 1 0 1 1 1 1 0 1 1 0 1 1 0 0 1 1 0 0 1 1 0 1 1 0 1 0 1 0 1 1 1 0 1 0 1 0
##  [38] 0 1 1 0 0 0 1 0 1 1 1 1 1 0 0 0 1 0 1 1 0 0 1 1 1 1 1 1 0 0 1 1 0 0 0 0 0
##  [75] 0 1 0 0 1 0 1 0 0 1 1 1 0 0 0 0 1 0 0 1 1 1 0 1 1 1
```

```r
table(Datos)
```

```
## Datos
##  0  1
## 45 55
```

```r
Frecuencia <- mean(Datos)
Frecuencia
```

```
## [1] 0.55
```

2.1 Solución Detallada

- **Objetivo:** Determinar la probabilidad de éxito del tratamiento médico, garantizando al menos un 55% de efectividad en un grupo aleatorio de $n = 100$ pacientes.
- **Parámetros de la Población:**
  - Población de Interés: Pacientes sometidos al tratamiento médico.
  - Umbral de Éxito: 55% de pacientes que deben mostrar mejoría.
  - Distribución Poblacional: Binomial con $n = 100$, $p = 0.5$ (probabilidad de éxito estimada) y $q = 1 - p = 0.5$.

3. Identificación del Estimador Apropiado

- **Función de Verosimilitud:**

La función de verosimilitud para la distribución binomial se define como: $L(p; x_1, x_2, ..., x_n) = \prod_{i=1}^{n} \binom{n}{x_i} p^{x_i} (1-p)^{n-x_i}$

Donde: p es la probabilidad de éxito, x_i es el número de éxitos en la i-ésima observación. - **Maximización de la Verosimilitud:**

Para encontrar el valor de p que maximiza la verosimilitud, se deriva la función de verosimilitud con respecto a p e iguala a cero.

$\frac{dL(p;x_1,x_2,...,x_n)}{dp} = 0$ Al resolver la ecuación anterior, se obtiene que el valor que maximiza la verosimilitud es: $p = \frac{\sum_{i=1}^{n} x_i}{n}$

- **Estimador Clave:** La fracción de pacientes que responden positivamente, representada por Y/n (Y el número de pacientes que responden positivamente al tratamiento).
- **Método:** Calcular la proporción muestral de pacientes que responden positivamente al tratamiento y que es mayor o igual al 55%, representada por Y/n.

```
Y <- sum(Datos) # Y es el exito de la muestra (Cantidad de 1 en la distribución de bernoulli)
estimador_muestra <- Y / n
cat("Estimador de la proporción de éxito:", estimador_muestra)
```

```
## Estimador de la proporción de éxito: 0.55
```

- **Definición del Estimador:** $\frac{Y}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i$, donde $X_i = 1$ si el paciente i responde positivamente, y $X_i = 0$ en caso contrario.
- **Independencia de X_i:** Suponiendo independencia de las respuestas de los pacientes.

4. Demostración de Características del Estimador

4.1 Esperanza del Estimador

La esperanza del estimador se calcula como la media ponderada de las posibles respuestas, considerando la probabilidad de éxito p:

$$E\left(\frac{Y}{n}\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n} p = p$$

4.2 Varianza del Estimador

La varianza del estimador se calcula como la media ponderada de las varianzas individuales, considerando la independencia de las respuestas:

$$Var\left(\frac{Y}{n}\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2}\sum_{i=1}^{n} pq = \frac{pq}{n}$$

Estas propiedades respaldan la validez y utilidad del estimador en el contexto de la inferencia estadística.

```
m <- 10000 # número simulaciones
esperanzas <- numeric(m)
varianzas <- numeric(m)
for (i in 1:m) {datos_simulacion <- rbinom(n, 1, p) # Distribución Bernoulli
  estimador_simulacion <- mean(datos_simulacion)
  esperanzas[i] <- mean(datos_simulacion)
  varianzas[i] <- var(datos_simulacion) / n}
cat("Varianza:", mean(varianzas),"Esperanzas:",mean(esperanzas))
```

```
## Varianza: 0.002499921 Esperanzas: 0.50036
```

5. Distribución del Estimador

En el libro *Estadística Matemática con Aplicaciones* (Wackerly, Mendenhall, and Scheaffer 2014) se explica que el teorema del límite central implica que $X = \frac{Y}{n}$ se distribuye aproximadamente de manera normal con media $p = 0.5$ y varianza $\frac{pq}{n} = \frac{(0.5)(0.5)}{100} = 0.0025$ para muestras grandes, como se muestra en `"(Figure 1)"`.

```
m <- 10000 # número de simulaciones
datos <- matrix(nrow = m, ncol = n)
for (i in 1:m) {
  datos[i,] <- rbinom(n, 1, p)
}
datos <- data.frame(datos) # DataFrame con 1000 obs, y 100 muestras.
SumaXi <- apply(datos,1,sum)
SumaXi <- data.frame(SumaXi)
media <- apply(datos, 1, mean)
SumaXi <- cbind(SumaXi, media)
head(SumaXi)
```

```
##   SumaXi media
## 1     44  0.44
## 2     52  0.52
## 3     49  0.49
## 4     46  0.46
## 5     53  0.53
## 6     51  0.51
```

6. Aproximación de la Binomial por la Normal

- **Justificación:** Basada en el Teorema del Límite Central.
- **Condiciones de Validez:** Validez para muestras grandes.

7. Cálculos y Resultados

Probabilidad de Éxito: $P\left(\frac{Y}{n} \geq 0.55\right) = P\left(\frac{Y/n - 0.5}{\sqrt{0.0025}} \geq \frac{0.55 - 0.50}{0.05}\right)$

Aproximación por la Normal: $\approx P(Z \geq 1) = 0.1587$

```
# Calcular P(Z >= 1) utilizando pnorm
prob <- 1 - pnorm(1)

# Imprimir el resultado
cat("P(Z >= 1) =", round(prob, 4))
```

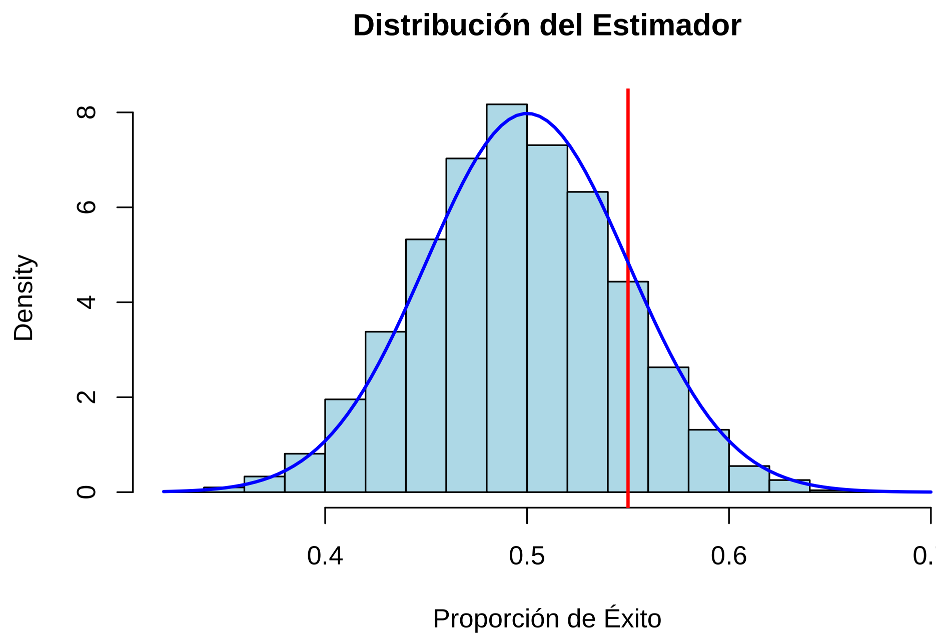

Figure 1: Distribución del Estimador

```
## P(Z >= 1) = 0.1587
```

8. Conclusión

- Se identificó la probabilidad de éxito del tratamiento mediante el análisis de la fracción de pacientes que responden positivamente.
- La aplicación del Teorema del Límite Central respalda la aproximación de la distribución binomial por la normal.
- La distribución del estimador fue demostrada mediante simulaciones y visualización gráfica.
- La probabilidad de éxito del tratamiento muestra una tendencia a ser muy baja.

Estos hallazgos destacan la importancia de una evaluación continua de la efectividad del tratamiento en busca de alternativas más efectivas.

Referencias

Wackerly, Dennis D., William Mendenhall, and Richard L. Scheaffer. 2014. *Estadística Matemática Con Aplicaciones*. 7th edition. México: Cengage Learning Editores.