

# Actividad 2 Metodos y Simulación Estadística-Problema 2

Javier\_Caicedo

2023-03-06

## PROBLEMA 2 VALIDACION DE RESULTADOS

La comparación de tratamientos es una práctica fundamental en las ciencias agropecuarias y para ello a nivel estadístico se cuenta con herramientas para apoyar el proceso de toma de decisiones y así poder lograr concluir con un alto grado de confianza sobre los resultados observados en una muestra. A través de la simulación es posible evaluar estimadores y sus propiedades, que nos permitan usarlos con toda tranquilidad.

Suponga un escenario en el cual se aplicó tratamientos diferentes a dos lotes de una misma plantas y se desea analizar si alguno de los dos tratamientos presenta un mejor desempeño en el control de una plaga presente en ambos al momento inicial. Para ello utilizará como criterio de desempeño el tratamiento, el menor porcentaje de plantas enfermas presente después de un tiempo de aplicación (es decir, si se presentan o no diferencias en las proporciones de enfermos  $p_1$  y  $p_2$  - proporciones poblacionales-).

- a) Realice una simulación en la cual genere dos poblaciones de  $N_1=1000$  (Lote 1) y  $N_2=1500$  (Lote 2), para los cuales se asume que el porcentaje de individuos (plantas) enfermas en ambos lotes es del 10% (es decir, sin diferencias entre los tratamientos).

```
## [1] 0.08245237 0.12069092
## attr(,"conf.level")
## [1] 0.95
```

Para una población de 1000 plantas y asumiendo que el porcentaje de plantas enfermas es del 10%, se puede afirmar con una confianza del 95%, que entre el 8.24% y el 12.06% de las plantas tienen alguna enfermedad

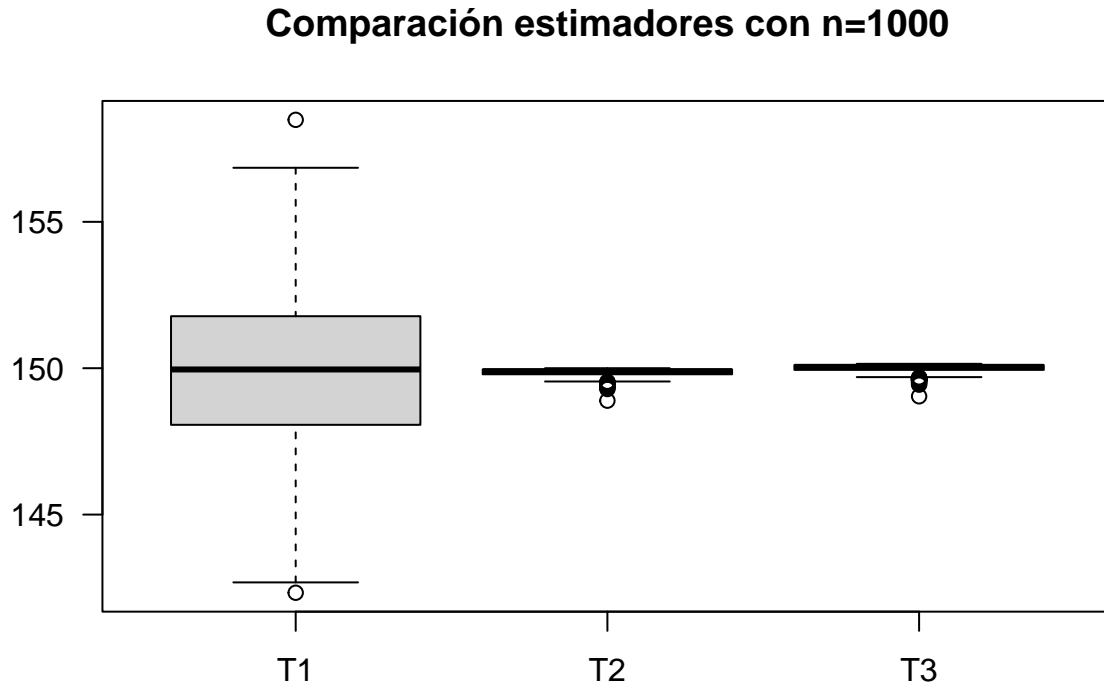
```
## [1] 0.08551487 0.11657338
## attr(,"conf.level")
## [1] 0.95
```

Para una población de 1500 plantas y asumiendo que el porcentaje de plantas enfermas es del 10%, se puede afirmar con una confianza del 95%, que entre el 8.55% y el 11.65% de las plantas tienen alguna enfermedad

- b) Genere una función que permita obtener una muestra aleatoria de los lotes y calcule el estimador de la proporción muestral para cada lote ( $\hat{p}_1$  y  $\hat{p}_2$ ) para un tamaño de muestra dado  $n_1=n_2$ . Calcule la diferencia entre los estimadores ( $\hat{p}_1 - \hat{p}_2$ ).

```
## [1] -0.02840566 0.05840566
## attr(,"conf.level")
## [1] 0.95
```

- c) Repita el escenario anterior (b) 500 veces y analice los resultados en cuanto al comportamiento de los 500 estimadores (diferencias  $\hat{p}_1 - \hat{p}_2$ ). ¿Qué tan simétricos son los resultados?, ¿Son siempre cero las diferencias?



- d) Realice los puntos b y c para tamaños de muestra  $n_1=n_2=5, 10, 15, 20, 30, 50, 60, 100, 200, 500$ . Compare los resultados de los estimadores ( $\hat{p}_1 - \hat{p}_2$ ) en cuanto a la normalidad. También analice el comportamiento de las diferencias y evalúe. ¿Considera que es más probable concluir que existen diferencias entre los tratamientos con muestras grandes que pequeñas, es decir, cuál considera usted que es el efecto del tamaño de muestra en el caso de la comparación de proporciones?

```
## Warning in prop.test(c(d1[2], d2[2]), c(sum(d1), sum(d2))): Chi-squared
## approximation may be incorrect
```

```
##
## 2-sample test for equality of proportions with continuity correction
##
## data: c(d1[2], d2[2]) out of c(sum(d1), sum(d2))
## X-squared = 0, df = 1, p-value = 1
## alternative hypothesis: two.sided
## 95 percent confidence interval:
## -0.5334346 1.0000000
## sample estimates:
## prop 1 prop 2
## 0.3333333 0.0000000
```

```

## Warning in prop.test(c(d1[2], d2[2]), c(sum(d1), sum(d2))): Chi-squared
## approximation may be incorrect

##
## 2-sample test for equality of proportions with continuity correction
##
## data:  c(d1[2], d2[2]) out of c(sum(d1), sum(d2))
## X-squared = 7.994e-32, df = 1, p-value = 1
## alternative hypothesis: two.sided
## 95 percent confidence interval:
## -0.4275413  0.2053191
## sample estimates:
##      prop 1      prop 2
## 0.0000000 0.1111111

## Warning in prop.test(c(d1[2], d2[2]), c(sum(d1), sum(d2))): Chi-squared
## approximation may be incorrect

##
## 2-sample test for equality of proportions with continuity correction
##
## data:  c(d1[2], d2[2]) out of c(sum(d1), sum(d2))
## X-squared = 2.6002e-32, df = 1, p-value = 1
## alternative hypothesis: two.sided
## 95 percent confidence interval:
## -0.1698866  0.3517048
## sample estimates:
##      prop 1      prop 2
## 0.09090909 0.00000000

## Warning in prop.test(c(d1[2], d2[2]), c(sum(d1), sum(d2))): Chi-squared
## approximation may be incorrect

##
## 2-sample test for equality of proportions with continuity correction
##
## data:  c(d1[2], d2[2]) out of c(sum(d1), sum(d2))
## X-squared = 3.6312e-32, df = 1, p-value = 1
## alternative hypothesis: two.sided
## 95 percent confidence interval:
## -0.1582985  0.1916318
## sample estimates:
##      prop 1      prop 2
## 0.06666667 0.05000000

## Warning in prop.test(c(d1[2], d2[2]), c(sum(d1), sum(d2))): Chi-squared
## approximation may be incorrect

##
## 2-sample test for equality of proportions with continuity correction
##
## data:  c(d1[2], d2[2]) out of c(sum(d1), sum(d2))

```

```

## X-squared = 1.2457e-31, df = 1, p-value = 1
## alternative hypothesis: two.sided
## 95 percent confidence interval:
## -0.10163616 0.09899066
## sample estimates:
##      prop 1      prop 2
## 0.03571429 0.03703704

## Warning in prop.test(c(d1[2], d2[2]), c(sum(d1), sum(d2))): Chi-squared
## approximation may be incorrect

##
## 2-sample test for equality of proportions with continuity correction
##
## data:  c(d1[2], d2[2]) out of c(sum(d1), sum(d2))
## X-squared = 2.5965e-31, df = 1, p-value = 1
## alternative hypothesis: two.sided
## 95 percent confidence interval:
## -0.1106866 0.1047607
## sample estimates:
##      prop 1      prop 2
## 0.03703704 0.04000000

## Warning in prop.test(c(d1[2], d2[2]), c(sum(d1), sum(d2))): Chi-squared
## approximation may be incorrect

##
## 2-sample test for equality of proportions with continuity correction
##
## data:  c(d1[2], d2[2]) out of c(sum(d1), sum(d2))
## X-squared = 2.541e-32, df = 1, p-value = 1
## alternative hypothesis: two.sided
## 95 percent confidence interval:
## -0.06655316 0.05850307
## sample estimates:
##      prop 1      prop 2
## 0.01923077 0.02325581

## Warning in prop.test(c(d1[2], d2[2]), c(sum(d1), sum(d2))): Chi-squared
## approximation may be incorrect

##
## 2-sample test for equality of proportions with continuity correction
##
## data:  c(d1[2], d2[2]) out of c(sum(d1), sum(d2))
## X-squared = 0.013068, df = 1, p-value = 0.909
## alternative hypothesis: two.sided
## 95 percent confidence interval:
## -0.02052528 0.04274750
## sample estimates:
##      prop 1      prop 2
## 0.01111111 0.00000000

```

```
## Warning in prop.test(c(d1[2], d2[2]), c(sum(d1), sum(d2))): Chi-squared
## approximation may be incorrect

##
## 2-sample test for equality of proportions without continuity correction
##
## data: c(d1[2], d2[2]) out of c(sum(d1), sum(d2))
## X-squared = NaN, df = 1, p-value = NA
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0 0
## sample estimates:
## prop 1 prop 2
##      0      0
```

- e) Ahora realice nuevamente los puntos a MENOS d bajo un escenario con dos lotes, pero de proporciones de enfermos diferentes ( $p_1=0.1$  y  $p_2=0.15$ ), es decir, el tratamiento del lote 1 si presentó un mejor desempeño reduciendo en un 5% el porcentaje de enfermos. Bajo este nuevo escenario compare la distribución de estas diferencias ( $p_1^{\wedge}$  MENOS  $p_2^{\wedge}$ ) con las observadas bajo igualdad de condiciones en los lotes. ¿Qué puede concluir? ¿Existen puntos en los cuales es posible que se observen diferencias de  $p_1$  MENOS  $p_2$  bajo ambos escenarios (escenario 1: sin diferencias entre  $p_1$  y  $p_2$ , escenario 2: diferencia de 5%)?

```
## planta
##    0    1
## 908  92
```

```
## planta
##    0    1
## 853 147
```