



# Métodos y Simulación estadística

Success

Solution

Business S

Innovation  
Branding  
Solution  
Marketing  
Analysis  
Ideas  
Success  
Management

Innovation  
Branding  
Solution  
Marketing  
Analysis  
Ideas  
Success  
Management

SOCIAL NET



## - Taller de simulación en R -




# Inferencia estadística y simulación

**1.** El Teorema del Límite Central es uno de los más importantes en la inferencia estadística y habla sobre la convergencia de los estimadores como la proporción muestral a la distribución normal. Algunos autores afirman que esta aproximación es bastante buena a partir del umbral  $n > 30$ .

- a.** Realice una simulación en la cual genere una población de  $N=1000$  (Lote) y además que el porcentaje de individuos (plantas) enfermas sea del 50%.
- b.** Genere una función que permita obtener una muestra aleatoria de la población y calcule el estimador de la proporción muestral para un tamaño de muestra dado  $n$ .
- c.** Repita el escenario anterior (b) 500 veces y analice los resultados en cuanto al comportamiento de los 500 estimadores. ¿Qué tan simétricos son los datos?, ¿Son sesgados y qué pasa en cuanto a variabilidad?
- d.** Realice los ejercicios completos b y c para tamaños de muestra  $n=5, 10, 15, 20, 30, 50, 60, 100, 200, 500$ . Y compare los resultados de los estimadores en cuanto a la normalidad. Investigue y utilice pruebas de bondad y ajuste (shapiro wilks) y métodos gráficos (grafico qq de normalidad).
- e.** Repita toda la simulación (puntos a – d) pero ahora con lotes con 10% y 90% de plantas enfermas. Concluya todo el ejercicio.

**2.** La comparación de tratamientos es una práctica fundamental en las ciencias agropecuarias y para esto a nivel estadístico se cuenta con algunas herramientas para apoyar el proceso de toma de decisiones y lograr concluir con algún grado de confianza que los resultados observados en una muestra son representativos y se pueden asociar a los tratamientos y no se deben únicamente al azar. Por medio una simulación validemos algunos de estos resultados.

- a.** Suponga un escenario en el cual usted aplicó tratamientos diferentes a dos lotes y desea analizar si alguno de los dos 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 que menor % 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  $P_2$ ). Realice una simulación en la cual genere dos poblaciones de  $N_1=1000$  (Lote1) y  $N_2=1500$  (Lote2), además asuma que el porcentaje de individuos (plantas) enfermas en ambos lotes sea la misma 10% (es decir, sin diferencias entre los tratamientos).

- b.** Genere una función que permita obtener una muestra aleatoria de los lotes y calcule el estimador de la proporción muestral para cada lote ( $p_1$  y  $p_2$ ) para un tamaño de muestra dado  $n_1=n_2$ . Calcule la diferencia entre los estimadores  $p_1-p_2$ .
  - c.** Repita el escenario anterior (b) 500 veces y analice los resultados en cuanto al comportamiento de los 500 estimadores (diferencias  $p_1-p_2$ ). ¿Qué tan simétricos son los datos?, ¿Son siempre cero las diferencias?
  - d.** Realice los puntos b y c para tamaños de muestra  $n_1=n_2=5, 10, 15, 20, 30, 50, 60, 100, 200, 500$ . Y compare los resultados de los estimadores ( $p_1-p_2$ ) 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?
  - e.** Ahora realice nuevamente los puntos a-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 ( $p_1-p_2$ ) 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  $p_1-p_2$  bajo ambos escenarios (escenario 1: sin diferencias entre  $P_1$  y  $P_2$ , escenario 2: diferencia de 5%)?
- 3.** Con base a los artículos "Statistical Errors: P values, the gold standard of statistical validity, are not as reliable as many scientists assume" & "Statisticians issue warning on P values: Statement aims to halt missteps in the quest for certainty" escriba un resumen (máximo 2 páginas) sobre ambos artículos e incluya en este sus opiniones en cuanto al uso del valor p como criterio de decisión en inferencia estadística.
- 