

# Problema 2 - Validación de resultados

Carlos Sierra Guzman, Camilo Vega Rámirez

## Contenido

Introducción . . . . .	2
Problema 2 . . . . .	2
Punto A . . . . .	2
Metodología Punto A . . . . .	2
Resultado A . . . . .	2
Punto B . . . . .	3
Metodología Punto B . . . . .	3
Resultado B . . . . .	3
Punto C . . . . .	3
Metodología Punto C . . . . .	3
Resultado C . . . . .	4
Punto D . . . . .	4
Metodología Punto D . . . . .	4
Resultado D . . . . .	5
Punto E . . . . .	6
Metodología Punto E . . . . .	6
Resultado E . . . . .	6
Conclusión . . . . .	9
Anexos . . . . .	9
Código Librerías . . . . .	9
Código Punto A . . . . .	9
Código Punto B . . . . .	10
Código Punto C . . . . .	10
Código Punto D . . . . .	11
Código Punto E . . . . .	12

## Introducción

Este documento tiene como propósito ofrecer la solución al problema 2 de la Unidad 2 del curso Métodos y Simulación Estadística. Cada sección aborda un punto específico que se debe resolver, y se presenta la metodología utilizada y los resultados obtenidos.

Adicionalmente, al final del documento se incluyen anexos con los códigos utilizados en la elaboración de las metodologías para cada sección. Esto permitirá al lector revisar con detalle cada paso del proceso seguido para llegar a las conclusiones presentadas en el documento. Para facilitar la navegación, el documento cuenta con enlaces que permiten acceder a cada uno de los códigos desde la correspondiente sección.

## Problema 2

### *Validación 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).*

---

## Punto A

*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).*

## Metodología Punto A

Se crea la función `sim_plantas_enfermas` para simular la una proporción de plantas enfermas dada una población.

Se generan simulaciones de  $N = 1000$  con 10% de plantas enfermas y  $N = 1500$  con 10% de plantas enfermas.

Se crea tabla para comprobar que las cantidades sean las correctas.

[Ir a código sección a](#)

## Resultado A

grupo	no_enferma	enferma
grupo_1	900	100

---

grupo	no_enferma	eneferma
grupo_2	1350	150

---

## Punto B

*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 ( $\hat{p}_1 - \hat{p}_2$ ).*

### Metodología Punto B

Se crea la función `sample_prop_multi` para extraer `n` muestras de los vectores `x` y `y`, luego calcular por separado sus estimadores de proporción muestral y finalmente resta al estimador de `x` el estimador de `y`.

Se verifica el funcionamiento de la función para un `n = 500` sobre las poblaciones simuladas.

[Ir a código sección b](#)

### Resultado B

```
## [1] "Diferencia estimadores de prueba = 0.016"
```

---

## Punto C

*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?*

### Metodología Punto C

Se crea la función `rep_sample_prop_multi` que nos permite repetir la función `sample_prop_multi` un número `rep` de veces.

Se realiza la simulación de 500 veces las diferencia de los estimadores con muestras de `n = 500` sobre las poblaciones simuladas.

Se crea la función `gg_rain_cloud`, que toma un vector y genera un gráfico de rain cloud.

Se crea la función `medidas_resumen`, que toma un vector y muestra en forma de tabla medidas de resumen respecto a simetría, sesgo y variabilidad.

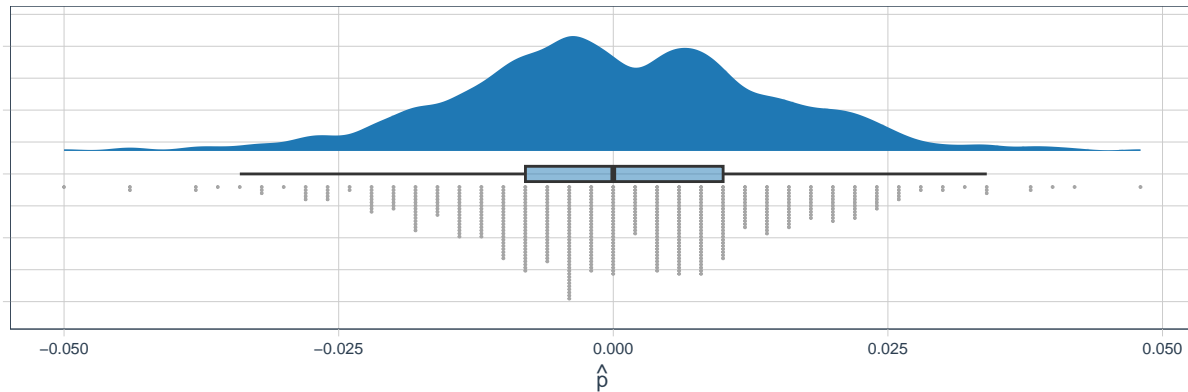
Se usan `gg_rain_cloud` y `medidas_resumen` sobre las 500 simulaciones de las diferencias en los estimadores para su análisis.

[Ir a código sección c](#)

## Resultado C

### Distribución de diferencia de estimadores para 500 repeticiones con $n = 500$

$N = 1000$  y 10% plantas enfermas vs.  $N = 1500$  y 10% plantas enfermas



mean	median	sd	min	max	skewness	kurtosis
0.0002	0	0.0144	-0.05	0.048	-0.0346	0.4444

Para un tamaño de muestra de  $n = 500$  y 500 repeticiones, se encontraron indicadores de skewness y kurtosis bajos en las diferencias de los estimadores. Además, al analizar la gráfica, se puede observar que los datos presentan una distribución simétrica. Tanto la mediana como el promedio del estimador se acercan a cero, lo que sugiere que la distribución de las diferencias de los estimadores es in sesgada.

Aunque las diferencias no siempre son cero, se observa que en general se aproximan a este valor. Sin embargo, se detectaron algunos outliers con diferencias del  $\pm 5\%$

## Punto D

*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?*

## Metodología Punto D

Se realiza la simulación de 500 veces el cálculo de las diferencias de los estimadores con múltiples tamaños de muestra  $n$  (5, 10, 15, 20, 30, 50, 60, 100, 200 y 500) sobre las poblaciones simuladas, y se colocan en un data frame.

Se crea la función `medidas_resumen_multiple`, que toma un data frame y muestra en forma de tabla medidas de resumen y test de normalidad de una columna seleccionada agrupados por otra columna seleccionada.

Se crea la función `gg_qq_plot`, que toma un data frame y realiza gráficos de normalidad tipo `qqnorm` de una columna seleccionada, agrupados por otra columna seleccionada.

Se usan `medidas_resumen_multiple` y `gg_qq_plot` sobre el data frame con las 500 simulaciones de diferencias de estimadores para distintos tamaños de muestra `n` para su análisis.

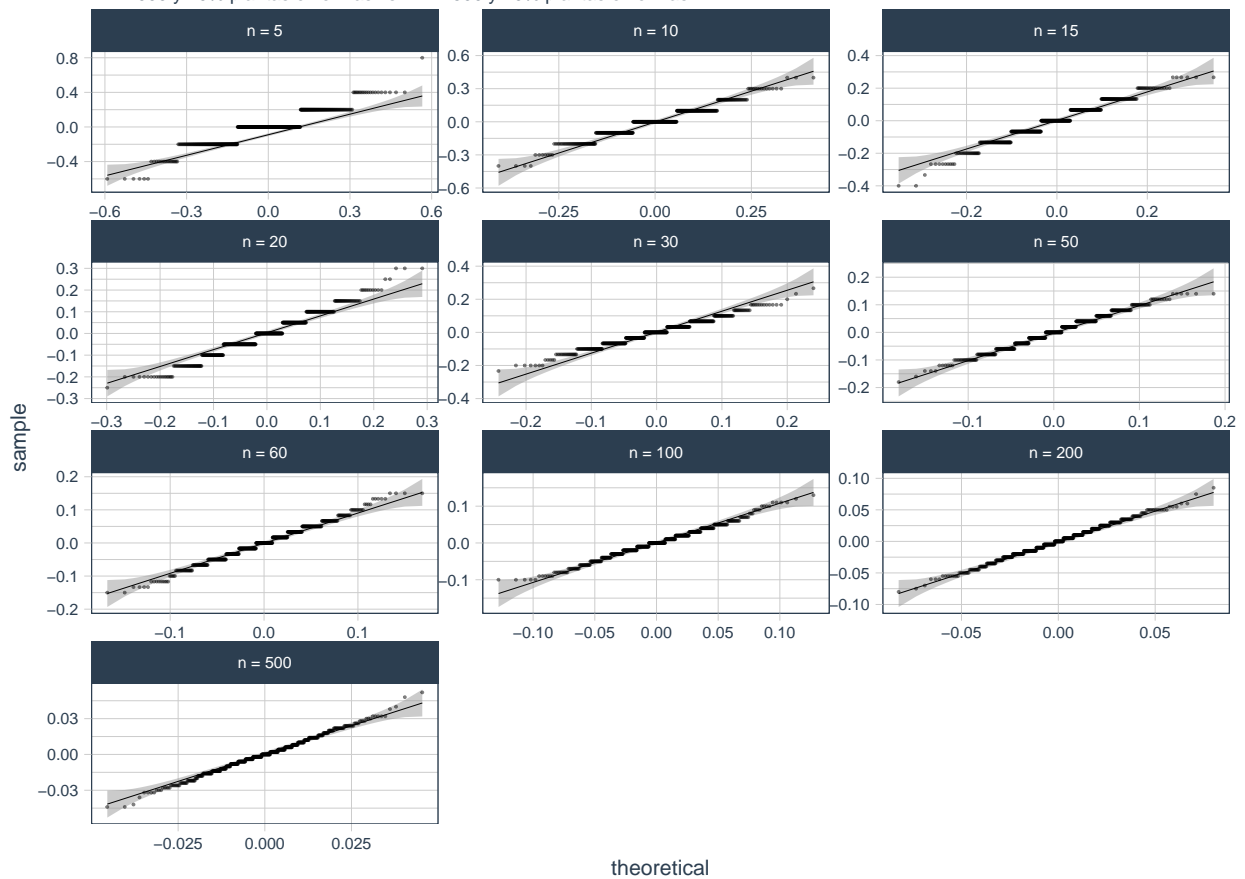
[Ir a código sección d](#)

## Resultado D

n	mean	median	sd	Shapiro-Wilk test P-Value
5	-0.0128	0	0.1874	0.0000
10	0.0024	0	0.1325	0.0000
15	-0.0017	0	0.1129	0.0000
20	-0.0040	0	0.0955	0.0000
30	-0.0009	0	0.0781	0.0000
50	0.0027	0	0.0596	0.0001
60	0.0008	0	0.0544	0.0012
100	-0.0004	0	0.0414	0.0024
200	-0.0011	0	0.0263	0.1494
500	0.0000	0	0.0147	0.1694

### qqplot de la diferencia de estimadores para 500 repeticiones con `n` múltiples

N = 1000 y 10% plantas enfermas vs. N = 1500 y 10% plantas enfermas



Se puede observar que a medida que aumentan los tamaños de muestra, el promedio de las diferencias de los estimadores se acerca cada vez más a cero. Sin embargo, esta reducción no es constante y presenta

variaciones por encima o por debajo de este valor. Incluso, en algunos casos, el promedio aumenta, como se puede ver en  $n = 200$ , donde su promedio (en valores absolutos) está por encima de la diferencia de  $n = 60$ . Por otro lado, la desviación estándar disminuye constantemente a medida que aumenta  $n$ .

Además, se puede observar en las gráficas de qqnorm que a medida que aumenta  $n$ , la distribución de las diferencias de los estimadores se asemeja más a una distribución normal. Esto se confirma con el test de Shapiro-Wilk, el cual arroja resultados positivos para normalidad a partir de  $n = 200$ .

Se observa que a medida que aumenta el tamaño de muestra, disminuye la probabilidad de concluir que hay diferencias entre la comparación de proporciones para los tratamientos, ya que los promedios se aproximan a cero. El principal factor para determinar esto son las desviaciones, las cuales son cada vez menores, pasando de desviaciones muy altas cercanas al 19% para  $n = 5$ , a desviaciones de solo alrededor del 1.5% para  $n = 500$ .

## Punto E

*e. Ahora realice nuevamente los puntos a-d bajo un escenario con dos lotes, pero de proporciones de enfermos diferentes ( $\hat{p}_1 = 0.1$  y  $\hat{p}_2 = 0.15$ ), 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 ( $\hat{p}_1 - \hat{p}_2$ ) 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  $\hat{p}_1 - \hat{p}_2$  bajo ambos escenarios (escenario 1: sin diferencias entre  $\hat{p}_1$  y  $\hat{p}_2$ , escenario 2: diferencia de 5%)?*

## Metodología Punto E

Se crea simulación de  $N = 1000$  con 15% de plantas enfermas, uniéndola con la simulación ya generada de  $N = 1000$  con 10% de plantas enfermas y se crea tabla para comprobar que las cantidades sean las correctas.

Se realiza la simulación de 500 veces el cálculo de las diferencias de los estimadores, con una muestra de  $n = 500$  sobre las poblaciones simuladas.

Se usan `gg_rain_cloud` y `medidas_resumen` sobre las 500 simulaciones de diferencias de estimadores para su análisis.

Se ejecuta simulación de 500 veces el cálculo de las diferencias de los estimadores, con múltiples tamaños de muestra  $n$  (5, 10, 15, 20, 30, 50, 60, 100, 200 y 500) sobre las poblaciones simuladas, y se colocan en un data frame.

Se usan `medidas_resumen_multiple` y `gg_qq_plot` sobre el data frame con las 500 simulaciones de las diferencias de los estimadores para distintos tamaños de muestra  $n$  sobre las poblaciones simuladas para su análisis.

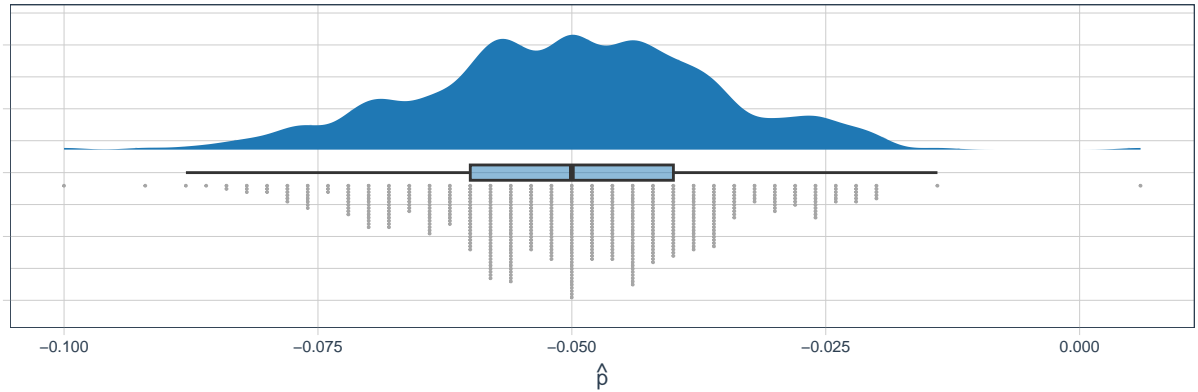
[Ir a código sección e](#)

## Resultado E

grupo	no_enferma	eneferma
grupo_1	900	100
grupo_3	850	150

### Distribución de diferencia de estimadores para 500 repeticiones con $n = 500$

$N = 1000$  y 10% plantas enfermas vs.  $N = 1000$  y 15% plantas enfermas



mean	median	sd	min	max	skewness	kurtosis
-0.0503	-0.05	0.0144	-0.1	0.006	-0.0735	0.2432

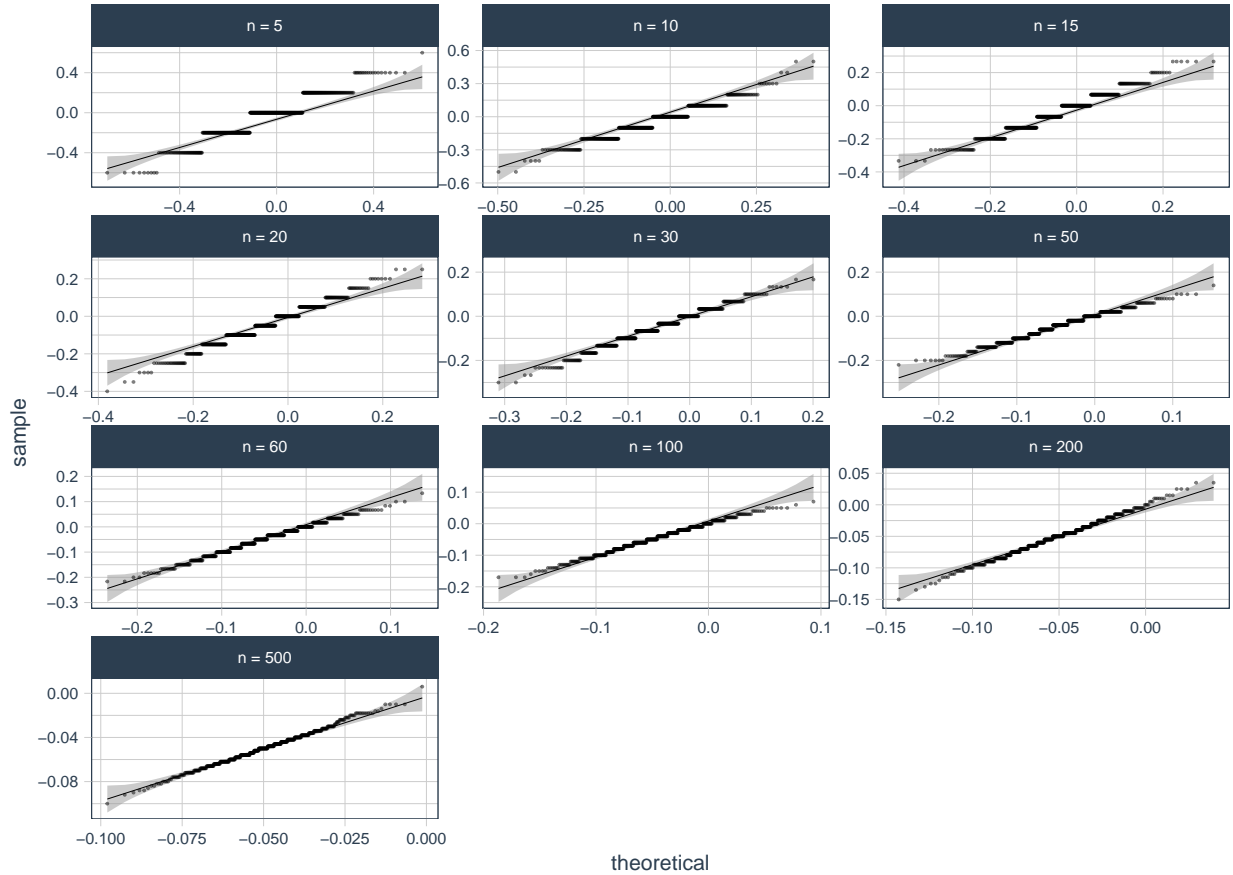
En la comparación de plantas enfermas al 10% vs 15%, con un tamaño de muestra  $n = 500$  y 500 repeticiones, los indicadores de skewness y kurtosis de las diferencias de los estimadores presentan valores bajos, lo que indica que los datos son simétricos. Además, tanto la mediana como el promedio del estimador se aproximan a -5%, lo que sugiere que la distribución de las diferencias de los estimadores es in-sesgada.

En cuanto a la normalidad de la distribución, se obtuvo un resultado similar al de la comparación de plantas enfermas al 10% para diferentes tamaños de población. En ambos casos, se encontró que a partir de un tamaño de muestra  $n = 50$ , las distribuciones se comienzan a asemejar a la normalidad.

n	mean	median	sd	Shapiro-Wilk test P-Value
5	-0.0488	0.0000	0.2104	0.0000
10	-0.0412	0.0000	0.1479	0.0000
15	-0.0476	-0.0667	0.1182	0.0000
20	-0.0489	-0.0500	0.1076	0.0000
30	-0.0546	-0.0667	0.0827	0.0000
50	-0.0496	-0.0400	0.0654	0.0001
60	-0.0490	-0.0500	0.0604	0.0030
100	-0.0466	-0.0500	0.0454	0.0088
200	-0.0517	-0.0500	0.0295	0.2128
500	-0.0496	-0.0500	0.0157	0.2803

### qqplot de la diferencia de estimadores para 500 repeticiones con n múltiples

N = 1000 y 10% plantas enfermas vs. N = 1000 y 15% plantas enfermas



En la comparación de estimadores de plantas enfermas al 10% vs 15%, se observa que a medida que aumenta el tamaño de muestra, el promedio de las diferencias de los estimadores se acerca cada vez más a -5%. Sin embargo, nunca se llega al valor exacto, a diferencia del escenario de diferencia de dos poblaciones con igual número de plantas enfermas, donde se obtuvo una diferencia estimada igual a la poblacional en  $n=500$ . También vemos que la desviación estándar disminuye a medida que aumenta  $n$ , pero sus valores son muy elevados en tamaños de muestra altos, llegando hasta 21% en  $n=5$ .

En las gráficas de  $qqnorm$  se puede observar que, a mayor  $n$ , la distribución de las diferencias de los estimadores se asemeja más a una distribución normal, lo cual se comprueba con el test de Shapiro-Wilk, que es positivo para normalidad a partir de  $n=200$ .

Al comparar los dos escenarios *escenario 1: sin diferencias entre  $\hat{p}_1$  y  $\hat{p}_2$* , *escenario 2: diferencia de 5%*, se ve que, a tamaños de muestra bajos, las desviaciones son muy altas, a pesar de que los valores medios de las diferencias son cercanos al real. Esto podría llevar a errores en la interpretación de los tratamientos, indicando una diferencia en el escenario 1 o sobrestimando el tamaño de la diferencia en el escenario 2, dependiendo de las muestras. Sin embargo, estas altas desviaciones van disminuyendo a medida que aumenta el tamaño de muestra, dando como resultado que no hay diferencia entre los tratamientos en el escenario 1 y una diferencia aproximada del 5% en los tratamientos en el escenario 2, siendo el tratamiento de la Población 1 mejor.



## Conclusión

En conclusión, es recomendable contar con métodos rigurosos al comparar tratamientos entre dos poblaciones utilizando la diferencia entre las proporciones de éxitos. Esto implica tener un tamaño de muestra adecuado y un número suficiente de repeticiones del experimento para garantizar la precisión y veracidad en las conclusiones obtenidas.

Es importante tener en cuenta que a medida que se aumenta el tamaño de muestra, la distribución de las diferencias de los estimadores se asemeja cada vez más a una distribución normal, lo que indica una mejor aproximación a la realidad. Además, se debe tener precaución en el análisis de resultados con tamaños de muestra bajos, ya que las desviaciones pueden ser muy altas y llevar a conclusiones erróneas.

---

## Anexos

### Código Librerías

```
library(tidyverse)      # Transformación de datos
library(normtest)       # Pruebas de normalidad
library(knitr)          # Renderizar tablas
library(ggdist)         # Expansión de graficas de ggplot
library(tidyquant)      # Tema de graficas de ggplot
library(nortest)        # Pruebas de normalidad
library(rapportools)    # Pruebas de normalidad
library(qqplotr)        # QQplot usando ggplot
options(scipen = 999)   # Anulación de notación científica
```

### Código Punto A

[Volver a metodología sección a](#)

```
# Función para generar población n, con una proporción prop de plantas enfermas
sim_plantas_enfermas <- function(n, prop){
  p <- round(n*prop)
  q <- n-p
  c(rep(TRUE,p), rep(FALSE,q))
}

# Simulando para N = 1000 y 0.1 de plantas enfermas
plantas_enfermas_10_1 <- sim_plantas_enfermas(1000, 0.1)

# Simulando para N = 1500 y 0.1 de plantas enfermas
plantas_enfermas_10_2 <- sim_plantas_enfermas(1500, 0.1)

# Tabla para visualizar simulación
grupo1 <- table(plantas_enfermas_10_1) %>%
  as_tibble() %>%
  mutate(grupo = "grupo_1") %>%
  rename(enferma = 1)

grupo2 <- table(plantas_enfermas_10_2) %>%
```

```

as_tibble() %>%
mutate(grupo = "grupo_2") %>%
rename(enferma = 1)

bind_rows(grupo1, grupo2) %>%
pivot_wider(names_from = enferma, values_from = n) %>%
set_names(c("grupo", "no_enferma", "eneferma")) %>%
kable()

```

## Código Punto B

[Volver a metodología sección b](#)

```

# Función para tomar una muestra de tamaño n para los vectores x,y calcular la
# proporción del estimador para cada uno y restar al estimador de x el estimador
# de y
sample_prop_multi <- function(x, y, n){
  x_sample <- sample(x, n) %>%
    sum()/n
  y_sample <- sample(y, n) %>%
    sum()/n
  x_sample-y_sample
}

# Reproducibilidad
set.seed(4321)

# Test de función con n = 500, sobre los vectores plantas_enfermas_10_1 y
# plantas_enfermas_10_2

str_c("Diferencia estimadores de prueba = ",
sample_prop_multi(plantas_enfermas_10_1, plantas_enfermas_10_2,500))

```

## Código Punto C

[Volver a metodología sección c](#)

```

# Función para repetir la función rep_sample_prop_multi un rep número de veces
rep_sample_prop_multi <- function(x, y, n, rep){
  map_dbl(1:rep, ~ sample_prop_multi(x, y, n))
}

# Reproducibilidad
set.seed(4321)

# Creación de 500 simulaciones de diferencias de estimadores, para un n = 500
# de muestras de plantas_enfermas_10_1 y plantas_enfermas_10_2
muestra_repetida_10_multi <- rep_sample_prop_multi(plantas_enfermas_10_1,
                                                    plantas_enfermas_10_2,500, 500)

# Función para crear grafico de rain cloud sobre un vector
gg_rain_cloud <- function(x, title, subtitle){

```

```

ggplot(x %>% as_tibble() , aes( y = value)) +
stat_halfeye(adjust = 0.5, justification = -0.2, .width = 0, point_colour = NA,
             fill = "#1F78B4") +
geom_boxplot(width = 0.12, outlier.color = NA, alpha = 0.5, fill = "#1F78B4") +
stat_dots(side = "left", justification = 1.1, fill = "#1F78B4") +
coord_flip() +
theme_tq() +
scale_fill_tq(theme = "light") +
theme(axis.text.y = element_blank(),
      text = element_text(size = 8)) +
xlab("") +
ylab(expression(hat("p"))) +
ggtitle(label = title,
        subtitle = subtitle)
}

# Función para crear tabla con medidas de resumen sobre un vector
medidas_resumen <- function(x){
  x %>%
  as_tibble() %>%
  summarise(mean = mean(value), median = median(value), sd = sd(value),
            min = min(value), max = max(value),
            skewness = skewness(value), kurtosis= kurtosis(value)) %>%
  mutate(across(where(is.numeric), ~ round(.,4))) %>%
  kable()
}

# Creación rain cloud y tabla de resumen sobre las 500 repeticiones con muestra
# n = 500, sobre las diferencias de estimadores simuladas.
gg_rain_cloud(
  muestra_repetida_10_multi,
  "Distribución de diferencia de estimadores para 500 repeticiones con n = 500",
  "N = 1000 y 10% plantas enfermas vs. N = 1500 y 10% plantas enfermas")
medidas_resumen(muestra_repetida_10_multi)

```

## Código Punto D

[Volver a metodología sección d](#)

```

# Reproducibilidad
set.seed(4321)

# Creación de 500 estimadores, para múltiples n de muestras de plantas_enfermas_10_1
# y plantas_enfermas_10_2
muestra_repetida_multiple_10 <- map_df(c(5, 10, 15, 20, 30, 50, 60, 100, 200, 500),
~ tibble(
  diff_p_hat = rep_sample_prop_multi(
    plantas_enfermas_10_1,
    plantas_enfermas_10_2,
    ., 500),
  n = as_factor(.)))

# Función para crear tabla con medidas de resumen sobre un data frame, para una

```

```

# columna agrupada por otra columna
medidas_resumen_multiple <- function(df, value, group){
  value <- enquo(value)
  group <- enquo(group)
  df %>%
    group_by(!!group) %>%
    summarise(mean = mean(!!value), median = median(!!value),
              sd = sd(!!value),
              `Shapiro-Wilk test P-Value` = shapiro.test(!!value)$p.value) %>%
    ungroup() %>%
    mutate(across(where(is.numeric), ~ round(.,4))) %>%
    kable()
}

# Función para crear grafico de qqplot sobre un data frame, para una columna
# agrupada por otra columna
gg_qq_plot <- function(df, value, group, title, subtitle){
  ggplot(df, aes(sample = {{value}})) +
    stat_qq_band(alpha = 0.5) +
    stat_qq_line(linewidth = 0.1) +
    stat_qq_point(alpha = 0.5, size = 0) +
    facet_wrap(vars(factor(str_c("n = ",{{group}}),
                              levels = c(str_c("n = ",c(5, 10, 15, 20, 30, 50, 60,
                                                    100, 200, 500))))),
              nrow = 4, scales = "free") +
    theme_tq() +
    theme(panel.spacing = unit(0, "lines"),
          text = element_text(size = 8)) +
    ggtitle(label = title,
            subtitle = subtitle)
}

# Creación de tabla de resumen y qqplots sobre las 500 repeticiones con muestra
# n múltiples, sobre la población simulada
medidas_resumen_multiple(muestra_repetida_multiple_10, diff_p_hat, n)
gg_qq_plot(muestra_repetida_multiple_10, diff_p_hat, n,
           "qqplot de la diferencia de estimadores para 500 repeticiones con n múltiples",
           "N = 1000 y 10% plantas enfermas vs. N = 1500 y 10% plantas enfermas")

```

## Código Punto E

[Volver a metodología sección e](#)

```

#Simulando para N = 1000 y 0.15 de plantas enfermas
plantas_enfermas_15_1 <- sim_plantas_enfermas(1000, 0.15)

# Tabla para visualizar simulación
grupo3 <- table(plantas_enfermas_15_1) %>%
  as_tibble() %>%
  mutate(grupo = "grupo_3") %>%
  rename(enferma = 1)

```

```

bind_rows(grupo1, grupo3) %>%
  pivot_wider(names_from = enferma, values_from = n) %>%
  set_names(c("grupo", "no_enferma", "eneferma")) %>%
  kable()

# Reproducibilidad
set.seed(4321)

# Creación de 500 estimadores, para un n = 500 de muestras de plantas_enfermas_10_1
# y plantas_enfermas_15_1
muestra_repetida_10_15_multi <- rep_sample_prop_multi(plantas_enfermas_10_1,
  plantas_enfermas_15_1, 500, 500)

# Creación rain cloud y tabla de resumen sobre las 500 repeticiones con muestra
# n = 500, sobre la población simulada al 10% de plantas enfermas
gg_rain_cloud(
  muestra_repetida_10_15_multi,
  "Distribución de diferencia de estimadores para 500 repeticiones con n = 500",
  "N = 1000 y 10% plantas enfermas vs. N = 1000 y 15% plantas enfermas")
medidas_resumen(muestra_repetida_10_15_multi)

# Reproducibilidad
set.seed(4321)

# Creación de 500 estimadores, para múltiples n de muestras de plantas_enfermas_10_1
# y plantas_enfermas_15_1
muestra_repetida_multiple_10_15 <- map_df(c(5, 10, 15, 20, 30, 50, 60, 100, 200, 500),
  ~ tibble(
    diff_p_hat = rep_sample_prop_multi(
      plantas_enfermas_10_1,
      plantas_enfermas_15_1,
      ., 500),
    n = as_factor(.)))

# Creación de tabla de resumen y qqplots sobre las 500 repeticiones con muestra
# n múltiples, sobre las poblaciones simuladas
medidas_resumen_multiple(muestra_repetida_multiple_10_15, diff_p_hat, n)
gg_qq_plot(muestra_repetida_multiple_10_15, diff_p_hat, n,
  "qqplot de la diferencia de estimadores para 500 repeticiones con n múltiples",
  "N = 1000 y 10% plantas enfermas vs. N = 1000 y 15% plantas enfermas")

```