

Actividad 2

Jairo Enrique Alba

21/2/2023

```
#Para que funcione como pdf se deben ejecutar estas dos lineas  
#install.packages('tinytex')  
#tinytex::install_tinytex()
```

Teorema del Límite Central

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 continuación se presentan los siguientes pasos para su verificación:

Ejercicio 1.

- Realice una simulación en la cual genere una población de $N = 1000$ (Lote), donde el porcentaje de individuos (supongamos plantas) enfermas sea del 50%.

```
#Generar la funcion  
GenerarPob = function(Tamaño = 1000, Enfermos = 0.5, Sanos = 0.5) {  
  E = round(Tamaño * Enfermos, 0)#redondear a cero decimales  
  S = round(Tamaño * Sanos, 0)#redondear a cero decimales  
  Poblacion = c(rep("Enfermo", E), rep("Sana",S))#creo un vector con Enfermo E veces y con Sana S veces  
  print(paste("Lote Generado de tamaño: ", Tamaño, " Enfermos: ",E,"Sanos: ",S))#combinan varias entradas  
  return(Poblacion)#R devolverá el resultado de cualquier comando que se ejecute en la última línea de  
}  
  
Lote = GenerarPob(Tamaño = 1000, Enfermos = 0.5, Sanos = 0.5) #mostar el tamaño de población y cuantos individuos  
  
## [1] "Lote Generado de tamaño: 1000 Enfermos: 500 Sanos: 500"
```

- Genere una función que permita:
 - Obtener una muestra aleatoria de la población y
 - Calcule el estimador de la proporción muestral \hat{p} para un tamaño de muestra dado n .

```

#Funcion para generar la muestra
GenerarMuestra = function(Poblacion, Tamaño, Condicion) {
  if (Tamaño<=length(Poblacion)) { #condición muestra mas pequeña de la población
    muestra = sample(x = Poblacion, size = Tamaño)#toma la muestra
    P_muestra = sum(muestra == Condicion)/Tamaño # para calcular la proporción
    return(P_muestra)#devuelve el resultado
  }
  else { #de lo contrario si el tamaño de la muestra supera la población
    print("Error: el tamaño de la muestra supera el tamaño de la población.")
    return(-1)
  }
}

#Generacion de la muestra de determinado tamaño

n = 200 #tamaño de muestra
muestra = GenerarMuestra (Poblacion = Lote, Tamaño = n, Condicion = "Enfermo") #generación de la muestra

print(paste("Para una muestra de tamaño:", n, "se obtuvo un ^P =",muestra)) #imprimir los resultados

```

```
## [1] "Para una muestra de tamaño: 200 se obtuvo un ^P = 0.495"
```

- Repita el escenario anterior (b) $n = 500$ veces y analice los resultados en cuanto al comportamiento de los 500 resultados del estimador \hat{p} .
 - ¿Qué tan simétricos o sesgados son los resultados obtenidos? y
 - ¿qué se puede observar en cuanto a la variabilidad?.
 - Realice en su informe un comentario sobre los resultados obtenidos.

```

#generación de 500 muestras de la población y mirar las proporciones muestrales de enfermos en cada una
GenerarMuestras_Iter = function(Poblacion, Tamaño, Condicion, Iteraciones) { #condiciones de la funcion
  if (Tamaño<=length(Poblacion)) {
    muestras_iter = array(NA,Iteraciones)#creación del array con NA, de dimensión el numero de iteraciones
    for (i in 1:Iteraciones) {
      muestras_iter[i] = GenerarMuestra (Poblacion, Tamaño, Condicion)
    }
    return(muestras_iter)
  }
  else {
    print("Error: el tamaño de la muestra supera el tamaño de la población.")
    return(-1)
  }
}

iter = 500
P_muestras = GenerarMuestras_Iter (Poblacion = Lote, Tamaño = n,
                                   Condicion = "Enfermo",Iteraciones = iter)

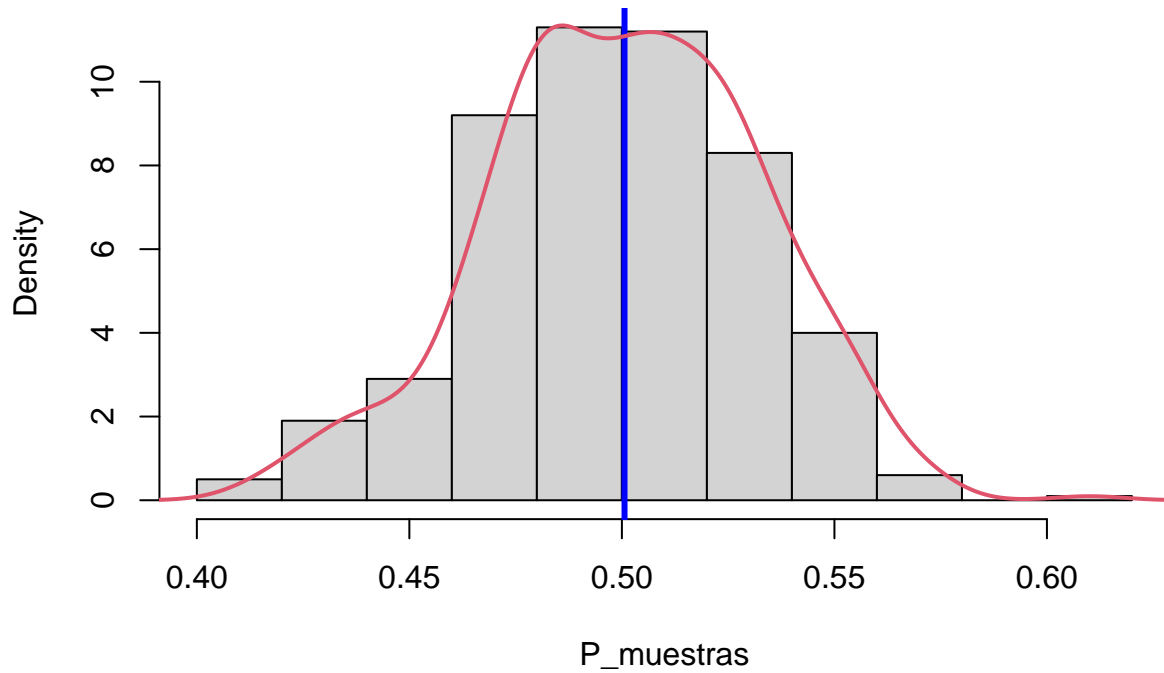
```

```

hist(P_muestras, prob = TRUE, main = "Proporción enfermos en las 500 muestras")
line = mean(P_muestras)
abline(v=line, col="blue", lwd=3)
lines(density(P_muestras), col = 2, lwd = 2)

```

Proporción enfermos en las 500 muestras

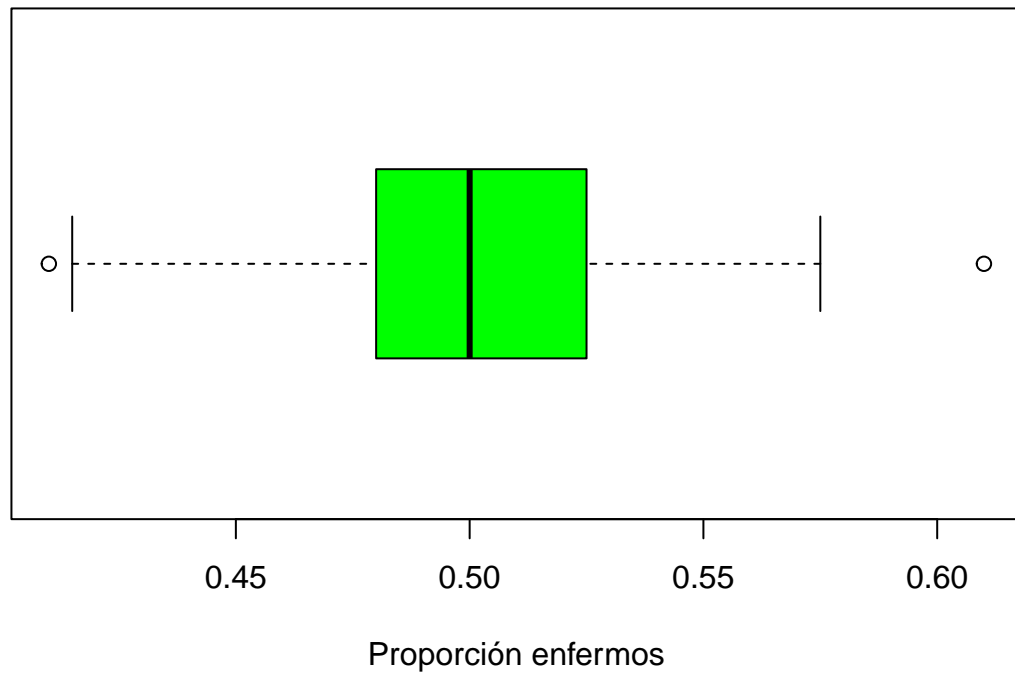


```
library(dplyr)
```

```
##  
## Attaching package: 'dplyr'  
  
## The following objects are masked from 'package:stats':  
##  
##   filter, lag  
  
## The following objects are masked from 'package:base':  
##  
##   intersect, setdiff, setequal, union
```

```
# Diagrama de caja  
boxplot(P_muestras, main = "Proporción enfermos en las 500 muestras",  
        xlab = "Proporción enfermos", col = "green", horizontal=T)
```

Proporción enfermos en las 500 muestras



```
#install.packages("moments") solo una vez  
library(moments)  
mean(P_muestras)
```

```
## [1] 0.50062
```

```
sd(P_muestras)
```

```
## [1] 0.03198847
```

```
min(P_muestras)
```

```
## [1] 0.41
```

```
median(P_muestras)
```

```
## [1] 0.5
```

```
max(P_muestras)
```

```
## [1] 0.61
```

```
skewness(P_muestras)
```

```
## [1] -0.1225454
```

```
kurtosis(P_muestras)
```

```
## [1] 2.911925
```

- Repita los puntos b y c para tamaños de muestra $n=5, 10, 15, 20, 30, 50, 60, 100, 200, 500$. Compare los resultados obtenidos para los diferentes tamaños de muestra en cuanto a la normalidad. Utilice pruebas de bondad y ajuste (shapiro wilks :shapiro.test()) y métodos gráficos (grafico de normalidad: qqnorm()). Comente en su informe los resultados obtenidos.

```
t_Muestras = c(10,15,20,30,50,60,100,200,500)
iter = 500
```

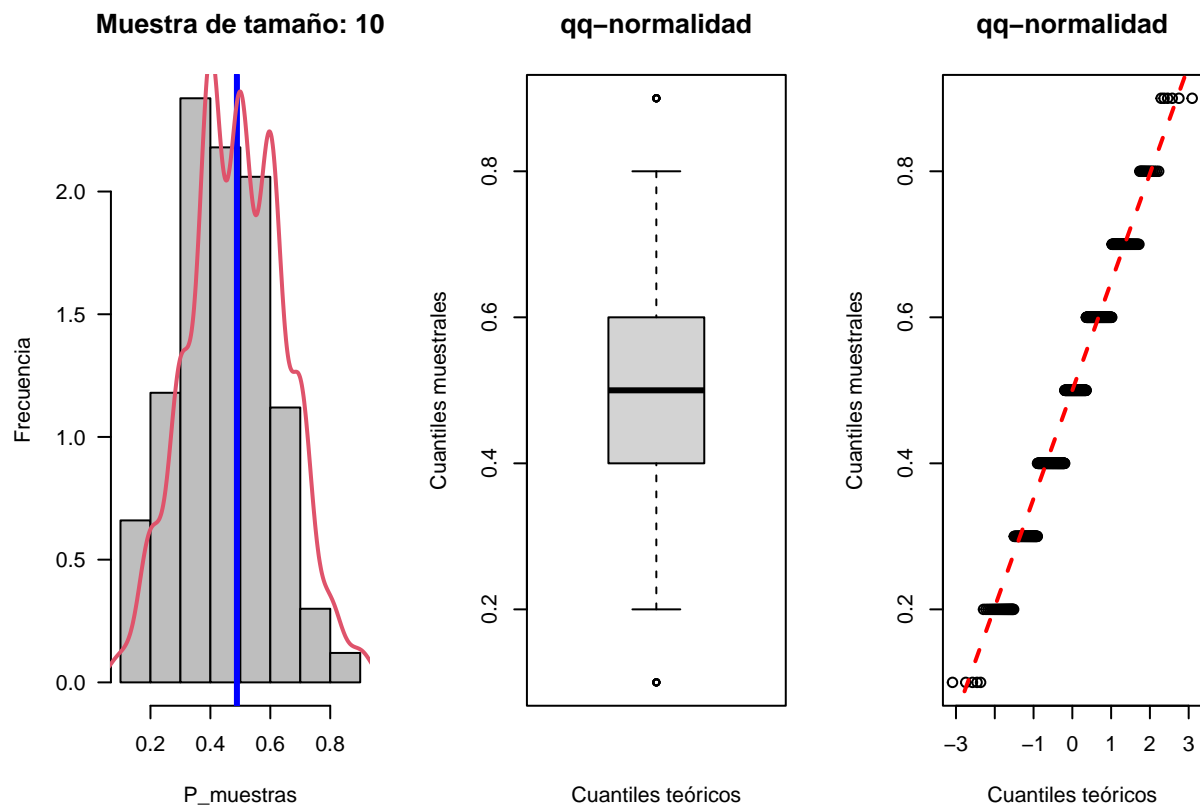
```
for (i in 1:length(t_Muestras)) {
  P_muestras = GenerarMuestras_Iter (Poblacion = Lote, Tamaño = t_Muestras[i], Condicion = "Enfermo"
  par(mfrow=c(1,3))
  hist(P_muestras, las=1, ylab = "Frecuencia", title = t_Muestras[i], main = paste("Muestra de tamaño", t_Muestras[i]))
  abline(v=mean(P_muestras), col="blue", lwd=3)
  lines(density(P_muestras), col = 2, lwd = 2)
  boxplot(P_muestras, xlab="Cuantiles teóricos", ylab="Cuantiles muestrales",main="qq-normalidad")
  qqnorm(P_muestras, xlab="Cuantiles teóricos", ylab="Cuantiles muestrales",main="qq-normalidad")
  qqline(P_muestras,col = 'red', lwd = 2, lty = 2)
  print(paste('Para una muestra de tamaño: ', t_Muestras[i]))
  print(shapiro.test(P_muestras))
  x=data.frame("ID"=i,"Tamaño_muestra"=t_Muestras[i],
               "Media"=mean(P_muestras),
               "Mediana"=median(P_muestras),
               "Desvest"=sd(P_muestras),
               "Varianza"=var(P_muestras),
               "Mín."=min(P_muestras),
               "Máx"=max(P_muestras),
               "Asimetría"=skewness(P_muestras),
               "Curtosis"= kurtosis(P_muestras)
               )
  Resultado = rbind(x)
}
```

```
## Warning in plot.window(xlim, ylim, "", ...): "title" is not a graphical
## parameter
```

```
## Warning in title(main = main, sub = sub, xlab = xlab, ylab = ylab, ...):
## "title" is not a graphical parameter
```

```
## Warning in axis(1, ...): "title" is not a graphical parameter
```

```
## Warning in axis(2, at = yt, ...): "title" is not a graphical parameter
```



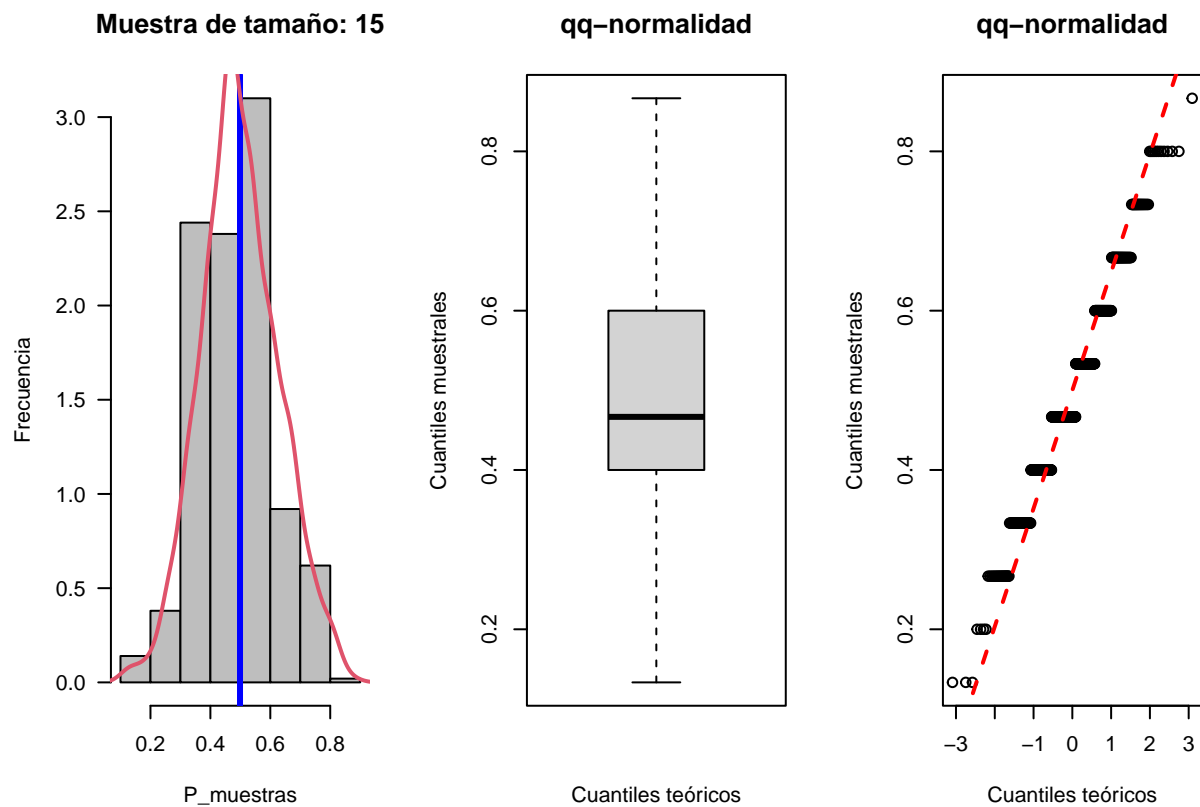
```
## [1] "Para una muestra de tamaño: 10"
##
## Shapiro-Wilk normality test
##
## data: P_muestras
## W = 0.96308, p-value = 6.974e-10

## Warning in plot.window(xlim, ylim, "", ...): "title" is not a graphical
## parameter

## Warning in title(main = main, sub = sub, xlab = xlab, ylab = ylab, ...):
## "title" is not a graphical parameter

## Warning in axis(1, ...): "title" is not a graphical parameter

## Warning in axis(2, at = yt, ...): "title" is not a graphical parameter
```



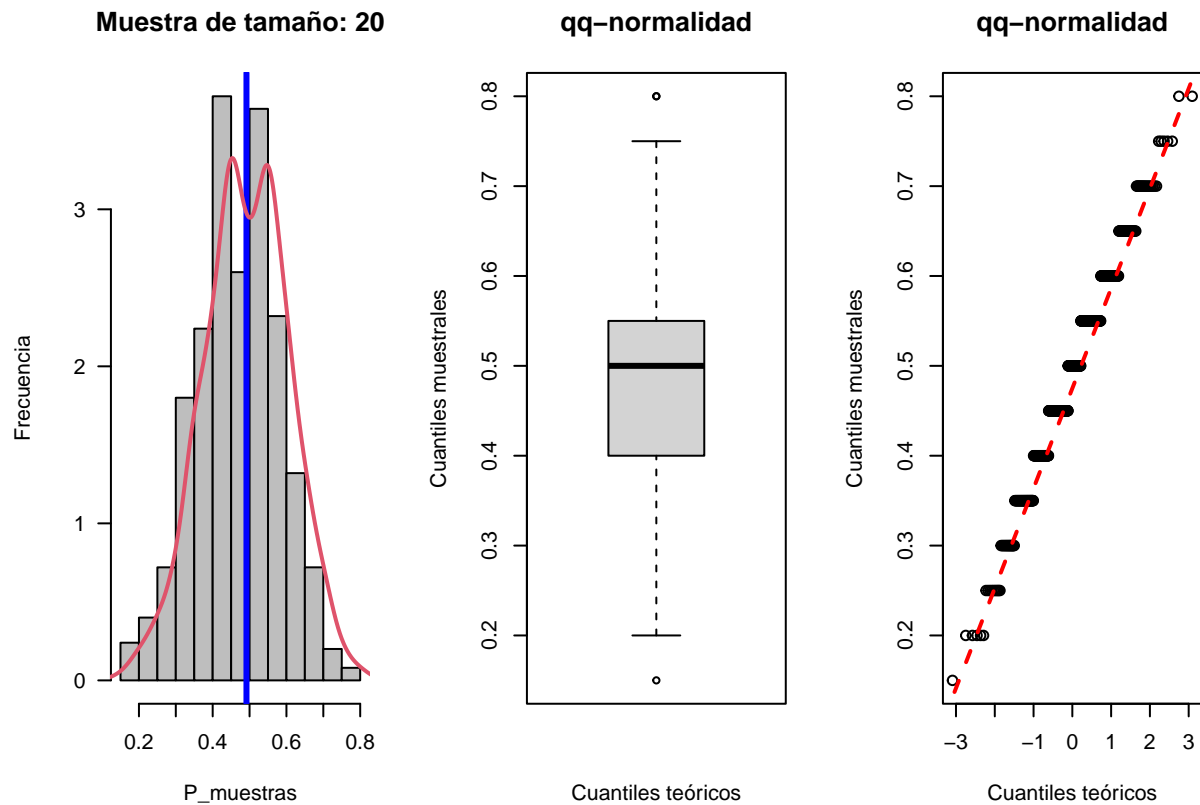
```
## [1] "Para una muestra de tamaño: 15"
##
## Shapiro-Wilk normality test
##
## data: P_muestras
## W = 0.97352, p-value = 7.328e-08

## Warning in plot.window(xlim, ylim, "", ...): "title" is not a graphical
## parameter

## Warning in title(main = main, sub = sub, xlab = xlab, ylab = ylab, ...):
## "title" is not a graphical parameter

## Warning in axis(1, ...): "title" is not a graphical parameter

## Warning in axis(2, at = yt, ...): "title" is not a graphical parameter
```



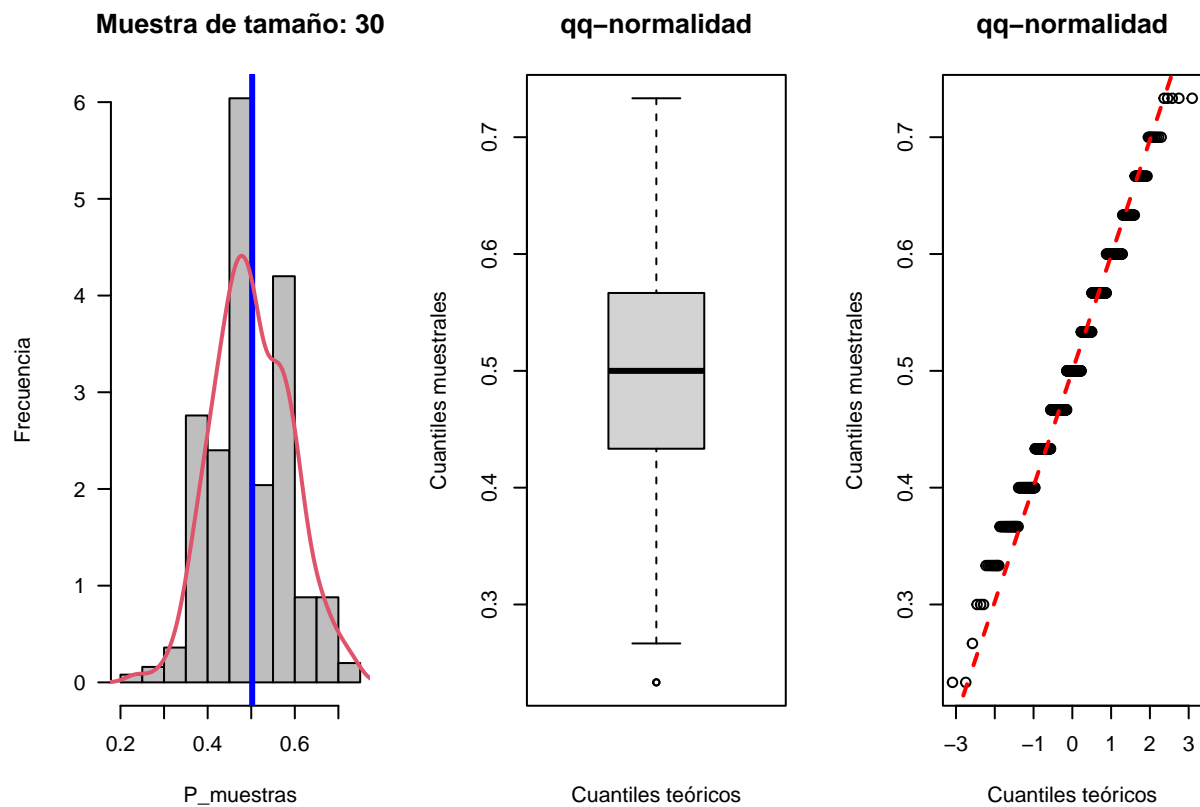
```
## [1] "Para una muestra de tamaño: 20"
##
## Shapiro-Wilk normality test
##
## data: P_muestras
## W = 0.98117, p-value = 4.547e-06

## Warning in plot.window(xlim, ylim, "", ...): "title" is not a graphical
## parameter

## Warning in title(main = main, sub = sub, xlab = xlab, ylab = ylab, ...):
## "title" is not a graphical parameter

## Warning in axis(1, ...): "title" is not a graphical parameter

## Warning in axis(2, at = yt, ...): "title" is not a graphical parameter
```

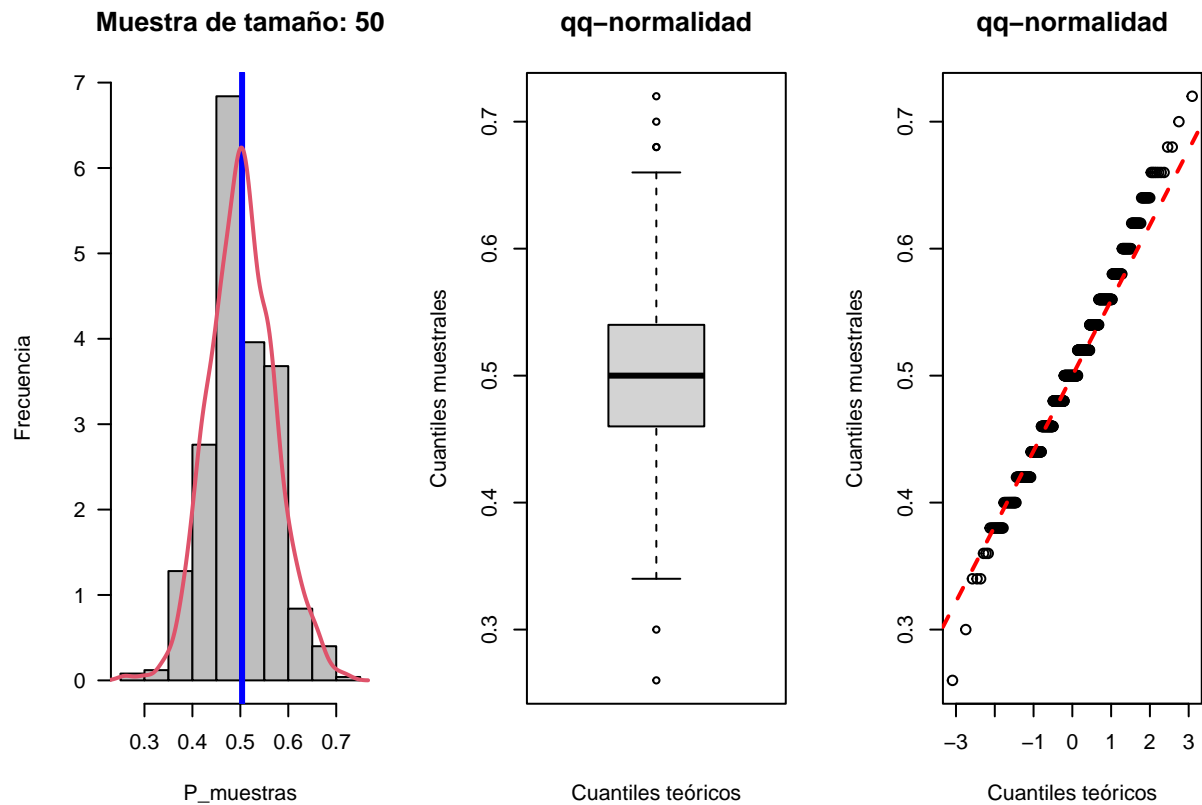
```
## [1] "Para una muestra de tamaño: 30"
##
## Shapiro-Wilk normality test
##
## data: P_muestras
## W = 0.98396, p-value = 2.537e-05

## Warning in plot.window(xlim, ylim, "", ...): "title" is not a graphical
## parameter

## Warning in title(main = main, sub = sub, xlab = xlab, ylab = ylab, ...):
## "title" is not a graphical parameter

## Warning in axis(1, ...): "title" is not a graphical parameter

## Warning in axis(2, at = yt, ...): "title" is not a graphical parameter
```



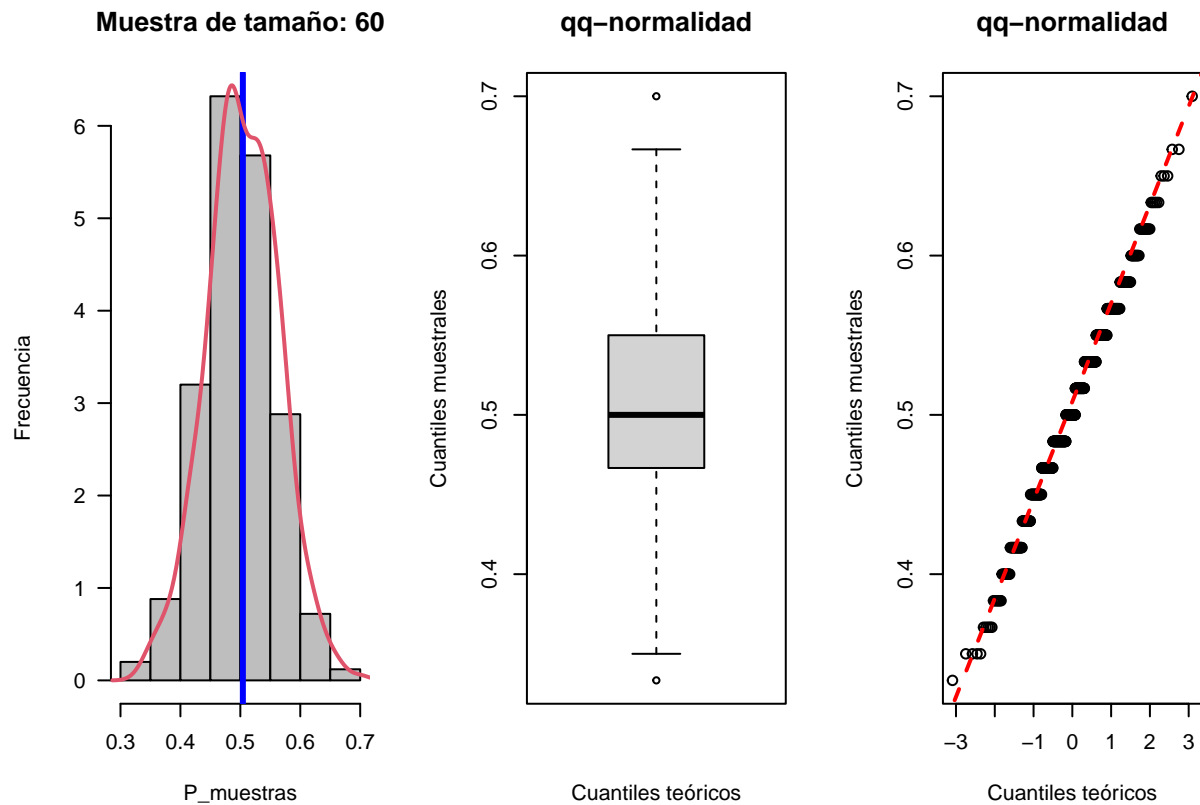
```
## [1] "Para una muestra de tamaño: 50"
##
##  Shapiro-Wilk normality test
##
## data:  P_muestras
## W = 0.99022, p-value = 0.002105

## Warning in plot.window(xlim, ylim, "", ...): "title" is not a graphical
## parameter

## Warning in title(main = main, sub = sub, xlab = xlab, ylab = ylab, ...):
## "title" is not a graphical parameter

## Warning in axis(1, ...): "title" is not a graphical parameter

## Warning in axis(2, at = yt, ...): "title" is not a graphical parameter
```



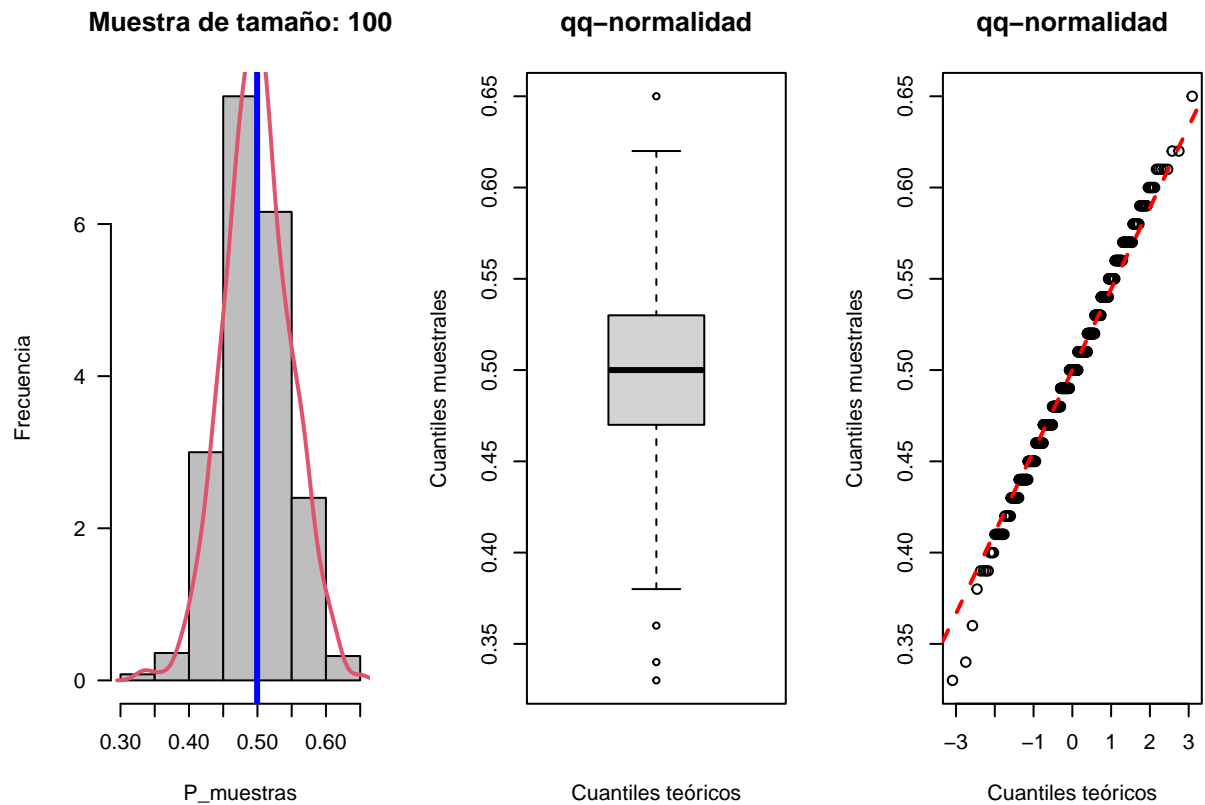
```
## [1] "Para una muestra de tamaño: 60"
##
## Shapiro-Wilk normality test
##
## data: P_muestras
## W = 0.9922, p-value = 0.01022

## Warning in plot.window(xlim, ylim, "", ...): "title" is not a graphical
## parameter

## Warning in title(main = main, sub = sub, xlab = xlab, ylab = ylab, ...):
## "title" is not a graphical parameter

## Warning in axis(1, ...): "title" is not a graphical parameter

## Warning in axis(2, at = yt, ...): "title" is not a graphical parameter
```



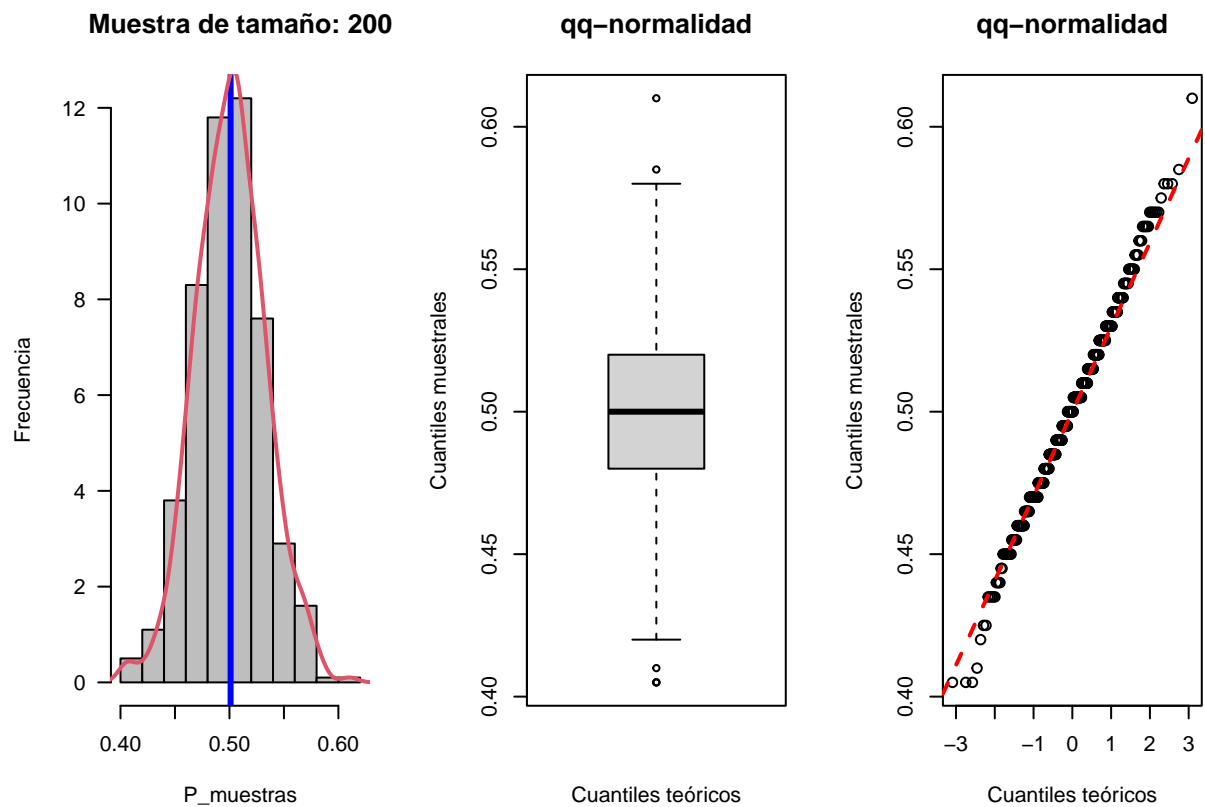
```
## [1] "Para una muestra de tamaño: 100"
##
## Shapiro-Wilk normality test
##
## data: P_muestras
## W = 0.99414, p-value = 0.05155

## Warning in plot.window(xlim, ylim, "", ...): "title" is not a graphical
## parameter

## Warning in title(main = main, sub = sub, xlab = xlab, ylab = ylab, ...):
## "title" is not a graphical parameter

## Warning in axis(1, ...): "title" is not a graphical parameter

## Warning in axis(2, at = yt, ...): "title" is not a graphical parameter
```



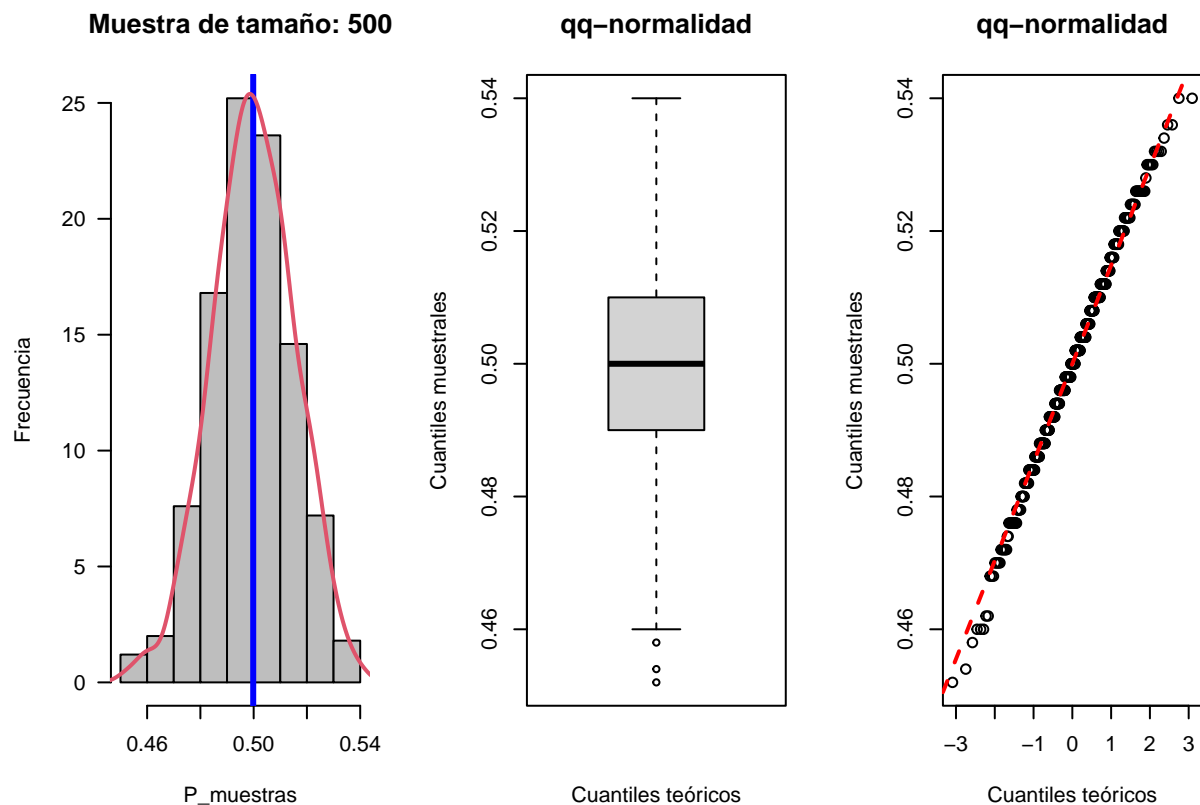
```
## [1] "Para una muestra de tamaño: 200"
##
## Shapiro-Wilk normality test
##
## data: P_muestras
## W = 0.99555, p-value = 0.1667

## Warning in plot.window(xlim, ylim, "", ...): "title" is not a graphical
## parameter

## Warning in title(main = main, sub = sub, xlab = xlab, ylab = ylab, ...):
## "title" is not a graphical parameter

## Warning in axis(1, ...): "title" is not a graphical parameter

## Warning in axis(2, at = yt, ...): "title" is not a graphical parameter
```



```
## [1] "Para una muestra de tamaño: 500"
##
## Shapiro-Wilk normality test
##
## data: P_muestras
## W = 0.99613, p-value = 0.2634
```

- Repita toda la simulación (puntos a – d), pero ahora para lotes con 10% de plantas enfermas y de nuevo para lotes con un 90% de plantas enfermas. Concluya sobre los resultados del ejercicio.

Para el 90% de plantas enfermas

```
Lote3 = GenerarPob(Tamaño = 1000, Enfermos = 0.9, Sanos = 0.1)
```

```
## [1] "Lote Generado de tamaño: 1000 Enfermos: 900 Sanos: 100"
```

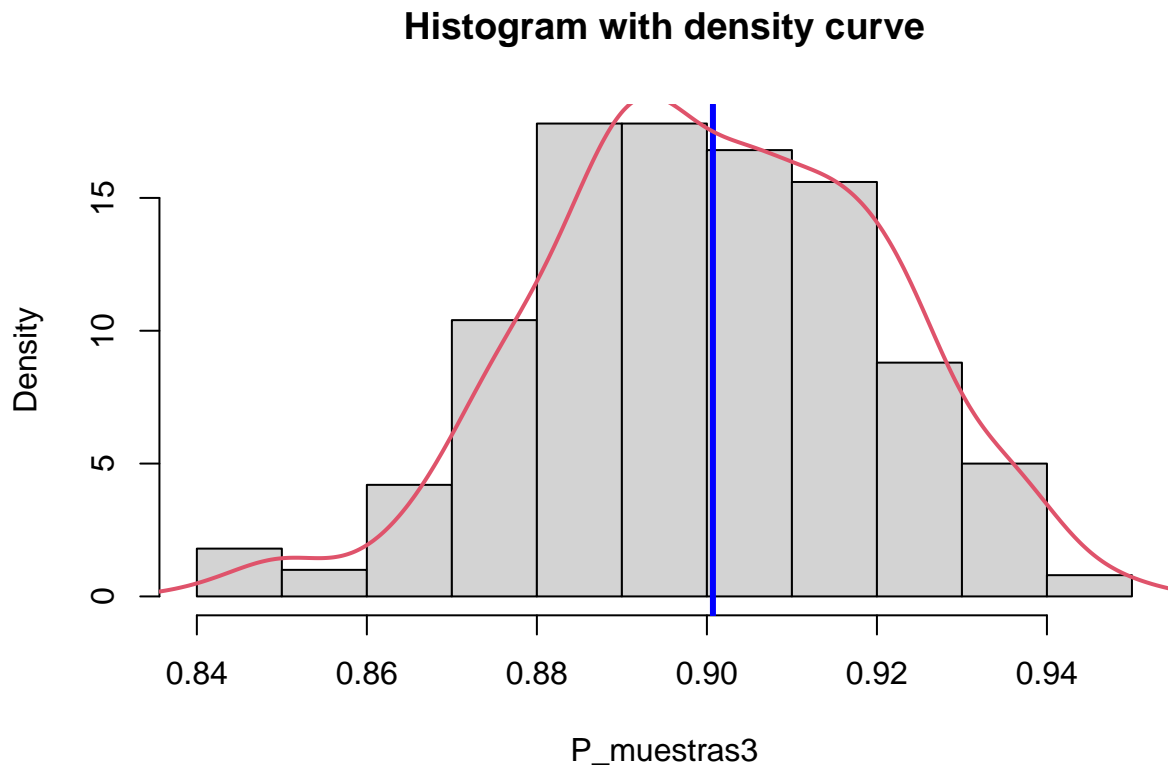
```
n3 = 300
muestra3 = GenerarMuestra (Poblacion = Lote3, Tamaño = n, Condicion = "Enfermo")

print(paste("Para una muestra de tamaño:", n3, "se obtuvo un ^P =", muestra3))
```

```
## [1] "Para una muestra de tamaño: 300 se obtuvo un ^P = 0.905"
```

```
P_muestras3 = GenerarMuestras_Iter (Poblacion = Lote3, Tamaño = n, Condicion = "Enfermo", Iteraciones = 1000)

hist(P_muestras3, prob = TRUE, main = "Histogram with density curve")
line = mean(P_muestras3)
abline(v=line, col="blue", lwd=3)
lines(density(P_muestras3), col = 2, lwd = 2)
```



```
Resultado3=data.frame("ID"=0, "Tamaño_muestra"=n3, "Media"=mean(P_muestras3), "Mediana"=median(P_muestras3))

Resultado3
```

```
##   ID Tamaño_muestra  Media Mediana  Desvest      Varianza Mín.  Máx
## 1   0             300 0.90071    0.9 0.01992335 0.0003969398 0.84 0.95
##   Asimetría Curtosis
## 1 -0.1611229 2.827129
```

```
for (i in 1:length(t_Muestras)) {
  P_muestras = GenerarMuestras_Iter (Poblacion = Lote3, Tamaño = t_Muestras[i], Condicion = "Enfermo", Iteraciones = 1000)
  par(mfrow=c(1,2))
  hist(P_muestras, las=1, ylab = "Frecuencia", title = t_Muestras[i], main = paste("Muestra de tamaño", t_Muestras[i]))
  abline(v=mean(P_muestras), col="blue", lwd=3)
  lines(density(P_muestras), col = 2, lwd = 2)
  qqnorm(P_muestras, xlab="Cuantiles teóricos", ylab="Cuantiles muestrales", main="qq-normalidad")
  qqline(P_muestras, col = 'red', lwd = 2, lty = 2)
```

```

print(paste('Para una muestra de tamaño: ', t_Muestras[i]))
print(shapiro.test(P_muestras))
x=data.frame("ID"=i,"Tamaño_muestra"=t_Muestras[i],
             "Media"=mean(P_muestras),
             "Mediana"=median(P_muestras),
             "Desvest"=sd(P_muestras),
             "Varianza"=var(P_muestras),
             "Mín."=min(P_muestras),
             "Máx"=max(P_muestras),
             "Asimetría"=skewness(P_muestras),
             "Curtosis"= kurtosis(P_muestras)
            )
Resultado3=rbind(Resultado3,x)
}

```

```

## Warning in plot.window(xlim, ylim, "", ...): "title" is not a graphical
## parameter

```

```

## Warning in title(main = main, sub = sub, xlab = xlab, ylab = ylab, ...):
## "title" is not a graphical parameter

```

```

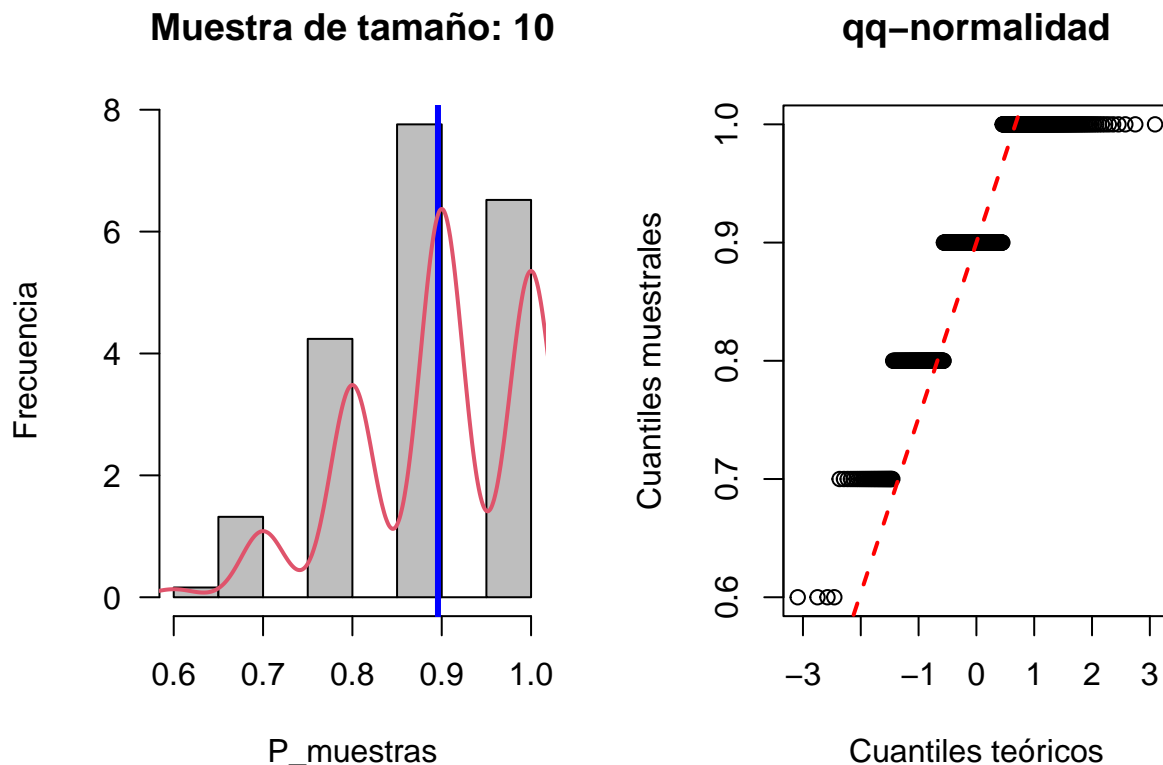
## Warning in axis(1, ...): "title" is not a graphical parameter

```

```

## Warning in axis(2, at = yt, ...): "title" is not a graphical parameter

```



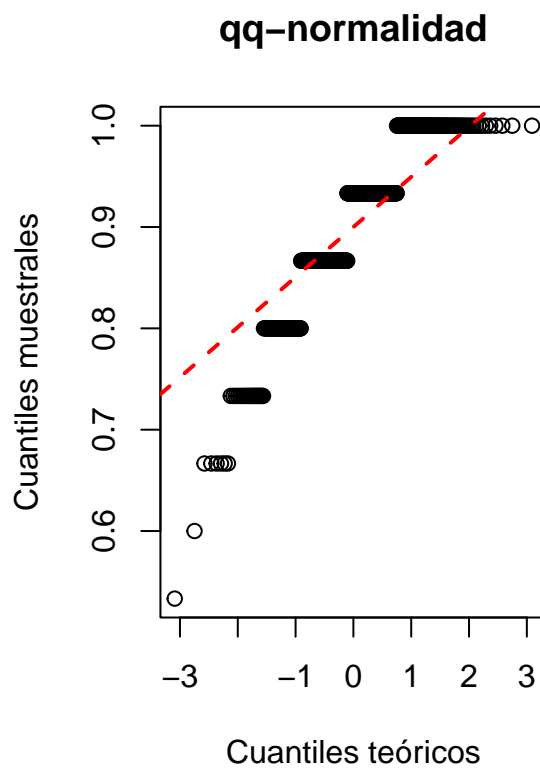
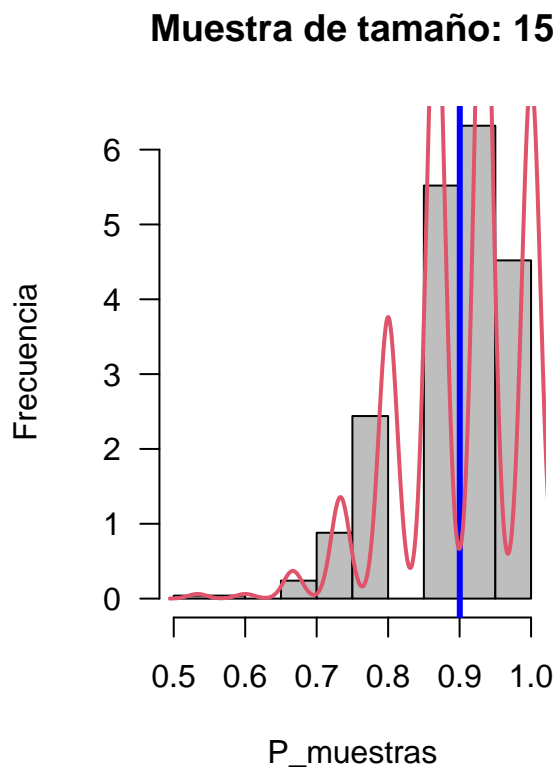

```
## [1] "Para una muestra de tamaño: 10"
##
## Shapiro-Wilk normality test
##
## data: P_muestras
## W = 0.8528, p-value < 2.2e-16

## Warning in plot.window(xlim, ylim, "", ...): "title" is not a graphical
## parameter

## Warning in title(main = main, sub = sub, xlab = xlab, ylab = ylab, ...):
## "title" is not a graphical parameter

## Warning in axis(1, ...): "title" is not a graphical parameter

## Warning in axis(2, at = yt, ...): "title" is not a graphical parameter
```



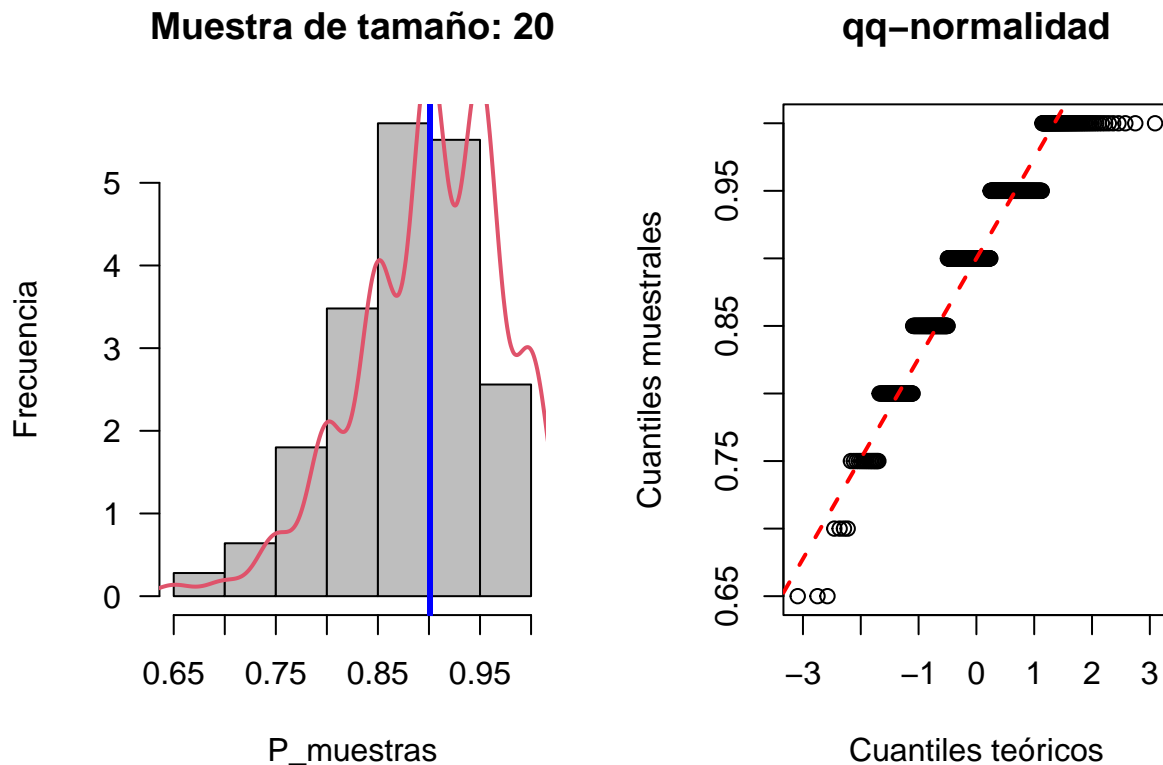
```
## [1] "Para una muestra de tamaño: 15"
##
## Shapiro-Wilk normality test
##
## data: P_muestras
## W = 0.89254, p-value < 2.2e-16
```

```
## Warning in plot.window(xlim, ylim, "", ...): "title" is not a graphical
## parameter
```

```
## Warning in title(main = main, sub = sub, xlab = xlab, ylab = ylab, ...):
## "title" is not a graphical parameter
```

```
## Warning in axis(1, ...): "title" is not a graphical parameter
```

```
## Warning in axis(2, at = yt, ...): "title" is not a graphical parameter
```



```
## [1] "Para una muestra de tamaño: 20"
```

```
##
```

```
## Shapiro-Wilk normality test
```

```
##
```

```
## data: P_muestras
```

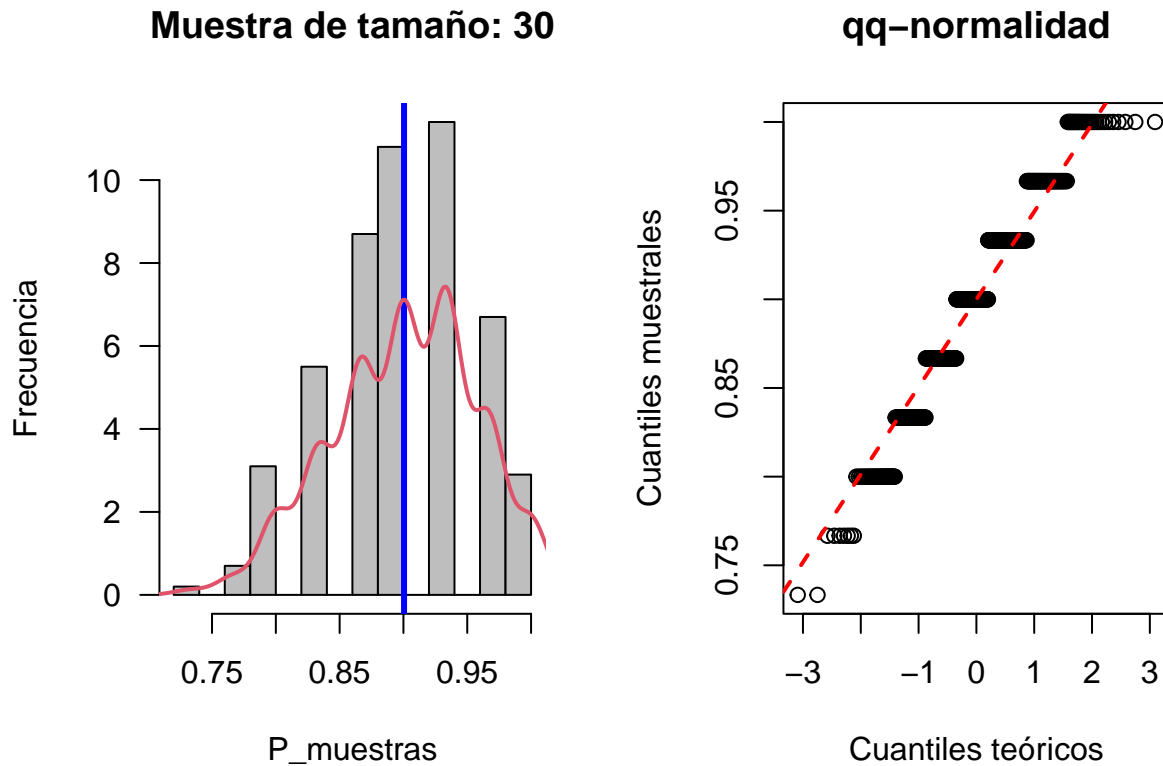
```
## W = 0.91863, p-value = 8.911e-16
```

```
## Warning in plot.window(xlim, ylim, "", ...): "title" is not a graphical
## parameter
```

```
## Warning in title(main = main, sub = sub, xlab = xlab, ylab = ylab, ...):
## "title" is not a graphical parameter
```

```
## Warning in axis(1, ...): "title" is not a graphical parameter
```

```
## Warning in axis(2, at = yt, ...): "title" is not a graphical parameter
```



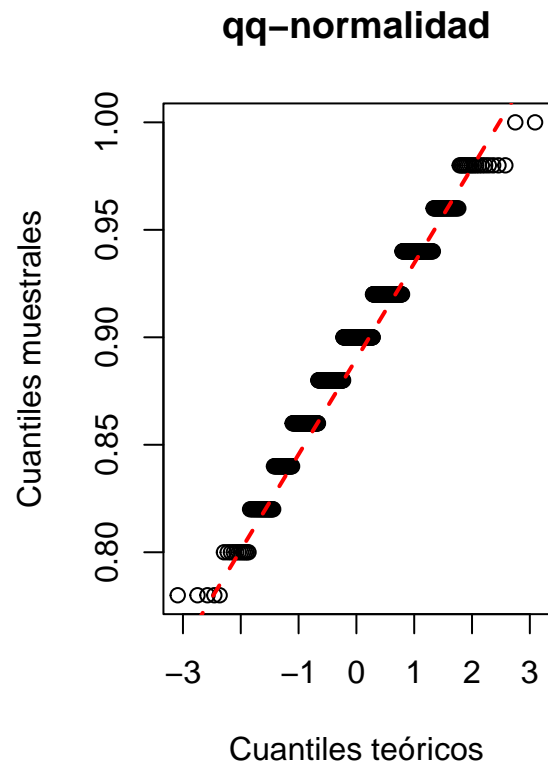
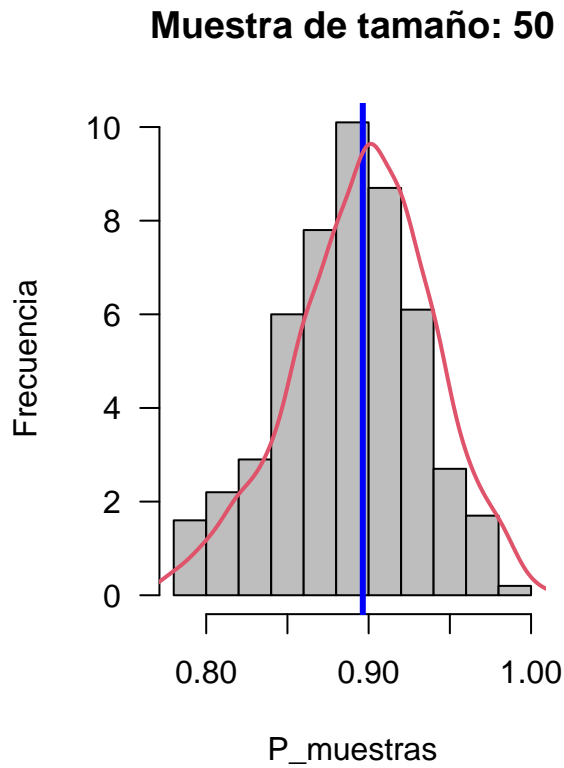
```
## [1] "Para una muestra de tamaño: 30"
##
## Shapiro-Wilk normality test
##
## data: P_muestras
## W = 0.95706, p-value = 6.861e-11
```

```
## Warning in plot.window(xlim, ylim, "", ...): "title" is not a graphical
## parameter
```

```
## Warning in title(main = main, sub = sub, xlab = xlab, ylab = ylab, ...):
## "title" is not a graphical parameter
```

```
## Warning in axis(1, ...): "title" is not a graphical parameter
```

```
## Warning in axis(2, at = yt, ...): "title" is not a graphical parameter
```



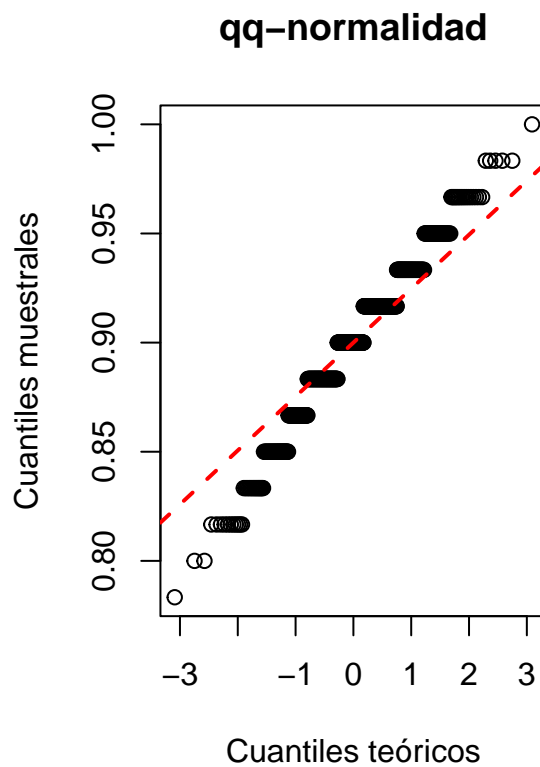
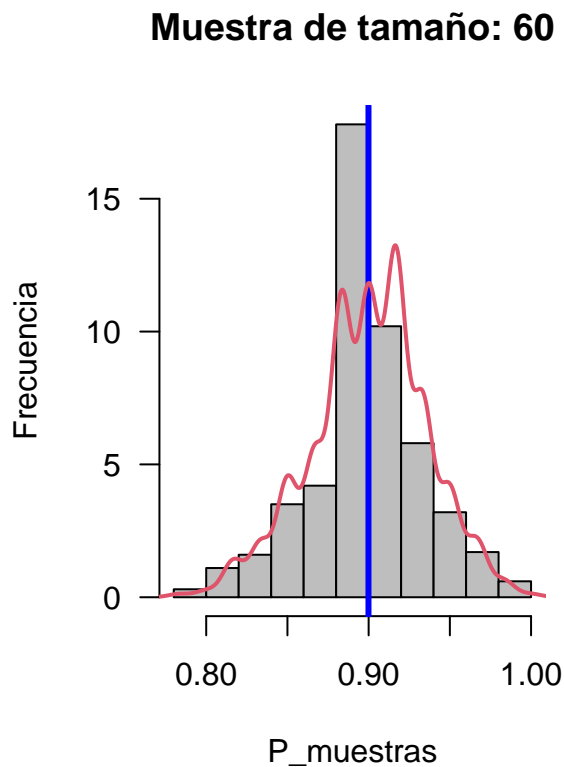
```
## [1] "Para una muestra de tamaño: 50"
##
## Shapiro-Wilk normality test
##
## data: P_muestras
## W = 0.9738, p-value = 8.4e-08

## Warning in plot.window(xlim, ylim, "", ...): "title" is not a graphical
## parameter

## Warning in title(main = main, sub = sub, xlab = xlab, ylab = ylab, ...):
## "title" is not a graphical parameter

## Warning in axis(1, ...): "title" is not a graphical parameter

## Warning in axis(2, at = yt, ...): "title" is not a graphical parameter
```



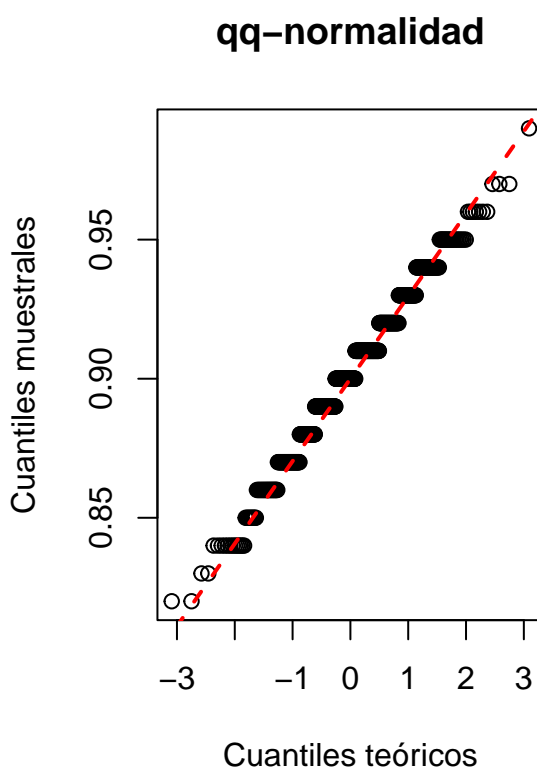
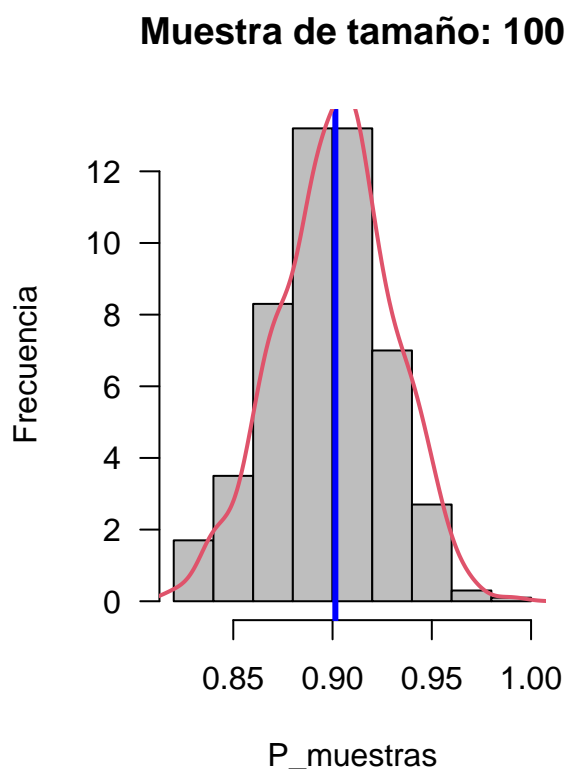
```
## [1] "Para una muestra de tamaño: 60"
##
## Shapiro-Wilk normality test
##
## data: P_muestras
## W = 0.97647, p-value = 3.296e-07

## Warning in plot.window(xlim, ylim, "", ...): "title" is not a graphical
## parameter

## Warning in title(main = main, sub = sub, xlab = xlab, ylab = ylab, ...):
## "title" is not a graphical parameter

## Warning in axis(1, ...): "title" is not a graphical parameter

## Warning in axis(2, at = yt, ...): "title" is not a graphical parameter
```



```
## [1] "Para una muestra de tamaño: 100"
##
## Shapiro-Wilk normality test
##
## data: P_muestras
## W = 0.98708, p-value = 0.0002066

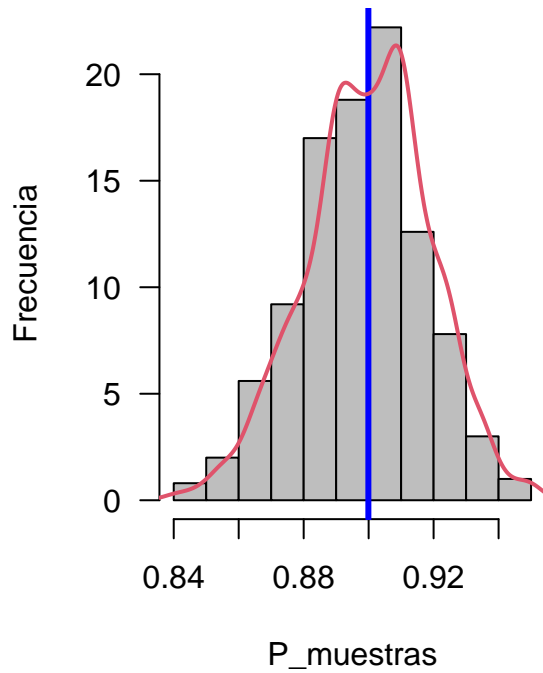
## Warning in plot.window(xlim, ylim, "", ...): "title" is not a graphical
## parameter

## Warning in title(main = main, sub = sub, xlab = xlab, ylab = ylab, ...):
## "title" is not a graphical parameter

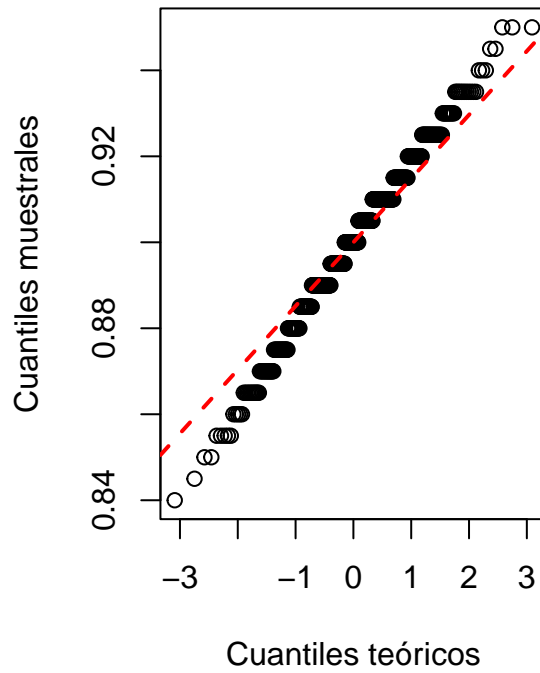
## Warning in axis(1, ...): "title" is not a graphical parameter

## Warning in axis(2, at = yt, ...): "title" is not a graphical parameter
```

Muestra de tamaño: 200



qq-normalidad



```
## [1] "Para una muestra de tamaño: 200"
##
## Shapiro-Wilk normality test
##
## data: P_muestras
## W = 0.99086, p-value = 0.003457

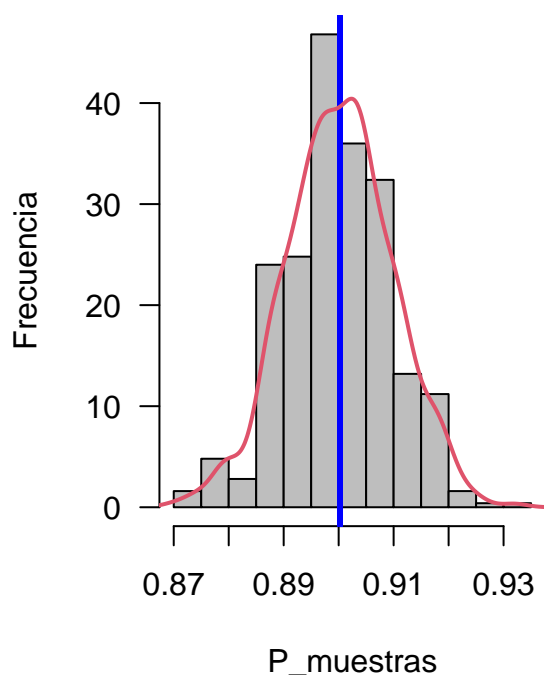
## Warning in plot.window(xlim, ylim, "", ...): "title" is not a graphical
## parameter

## Warning in title(main = main, sub = sub, xlab = xlab, ylab = ylab, ...):
## "title" is not a graphical parameter

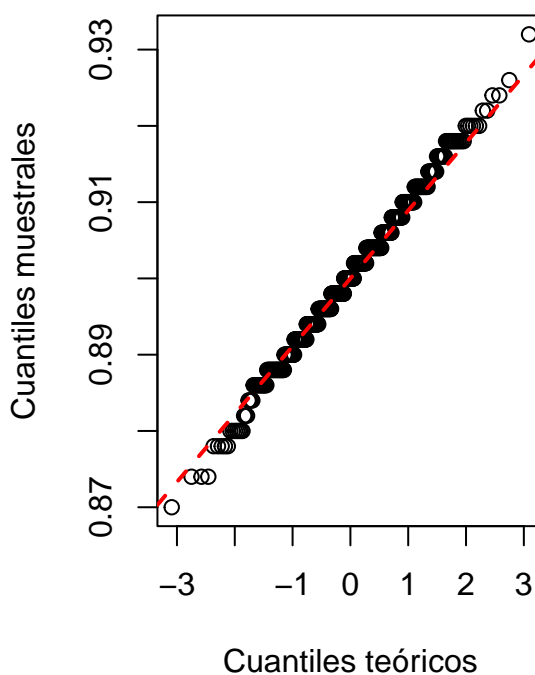
## Warning in axis(1, ...): "title" is not a graphical parameter

## Warning in axis(2, at = yt, ...): "title" is not a graphical parameter
```

Muestra de tamaño: 500



qq-normalidad



```
## [1] "Para una muestra de tamaño: 500"
##
## Shapiro-Wilk normality test
##
## data: P_muestras
## W = 0.99489, p-value = 0.09643
```

Resultado3

ID	Tamaño_muestra	Media	Mediana	Desvest	Varianza	Mín.
1	0	300	0.9007100	0.9000000	0.019923348	3.969398e-04
2	1	10	0.8958000	0.9000000	0.093487110	8.739840e-03
3	2	15	0.9002667	0.9333333	0.080468014	6.475101e-03
4	3	20	0.9010000	0.9000000	0.068763776	4.728457e-03
5	4	30	0.9004667	0.9000000	0.055651483	3.097088e-03
6	5	50	0.8964000	0.9000000	0.042617749	1.816273e-03
7	6	60	0.8999000	0.9000000	0.035617950	1.268638e-03
8	7	100	0.9014200	0.9000000	0.028562479	8.158152e-04
9	8	200	0.8999200	0.9000000	0.019026761	3.620176e-04
10	9	500	0.9002440	0.9000000	0.009676763	9.363974e-05

	Máx	Asimetría	Curtosis
1	0.950	-0.16112291	2.827129
2	1.000	-0.63837008	2.819944
3	1.000	-0.78623234	3.779637
4	1.000	-0.70690338	3.499863


```
## 5  1.000 -0.29632064 2.608244
## 6  1.000 -0.27689394 2.904585
## 7  1.000 -0.23512100 3.073598
## 8  0.990 -0.10301749 2.842682
## 9  0.950 -0.17463587 2.951167
## 10 0.932 -0.01908068 3.053227
```

Para el 10% de plantas enfermas