Tarea 4 Inferencia Estadística

Code

Hairo Ulises Miranda Belmonte

2 de Octubre de 2018

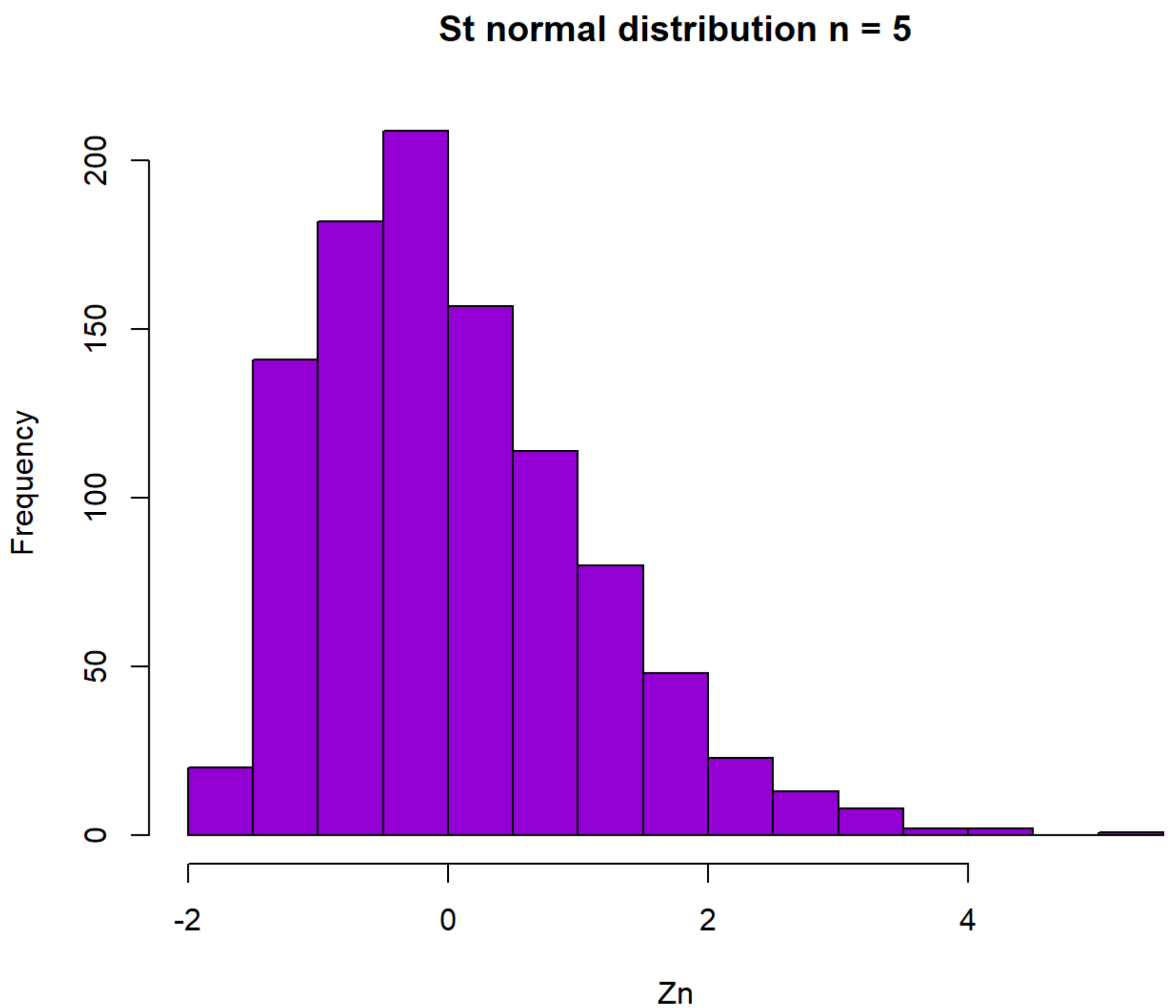
EJERCICIO 2

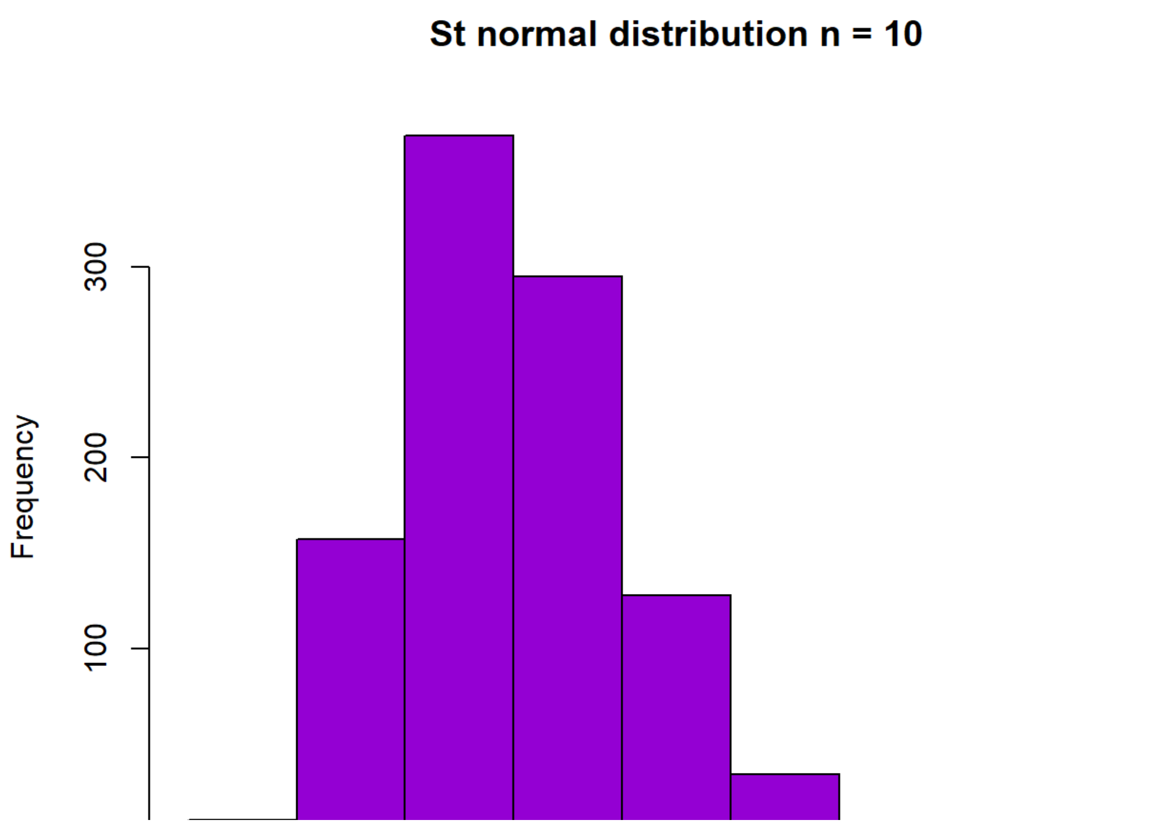
a. Escriba la siguiente funciónn en R. Simule una muestra de tamañoo n de una variable aleatoria Exponencial(λ) y calcule el estadístico $Z_n = \sqrt{n}(X_n - \lambda^{-1})/\lambda^{-1}$. Repita lo anterior m veces. La función deberá tomar como parámetros n, m y λ y regresar un vector de tamaño n conteniendo la muestra de Zn.

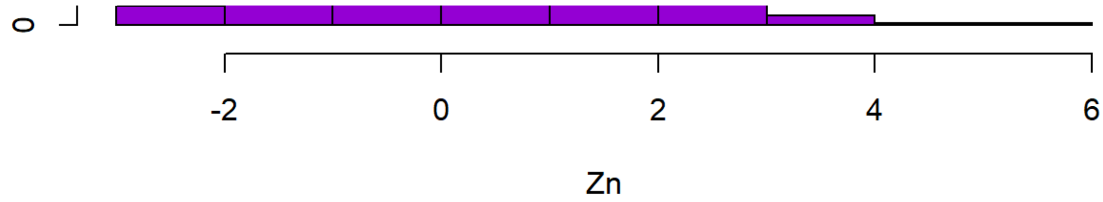
Code

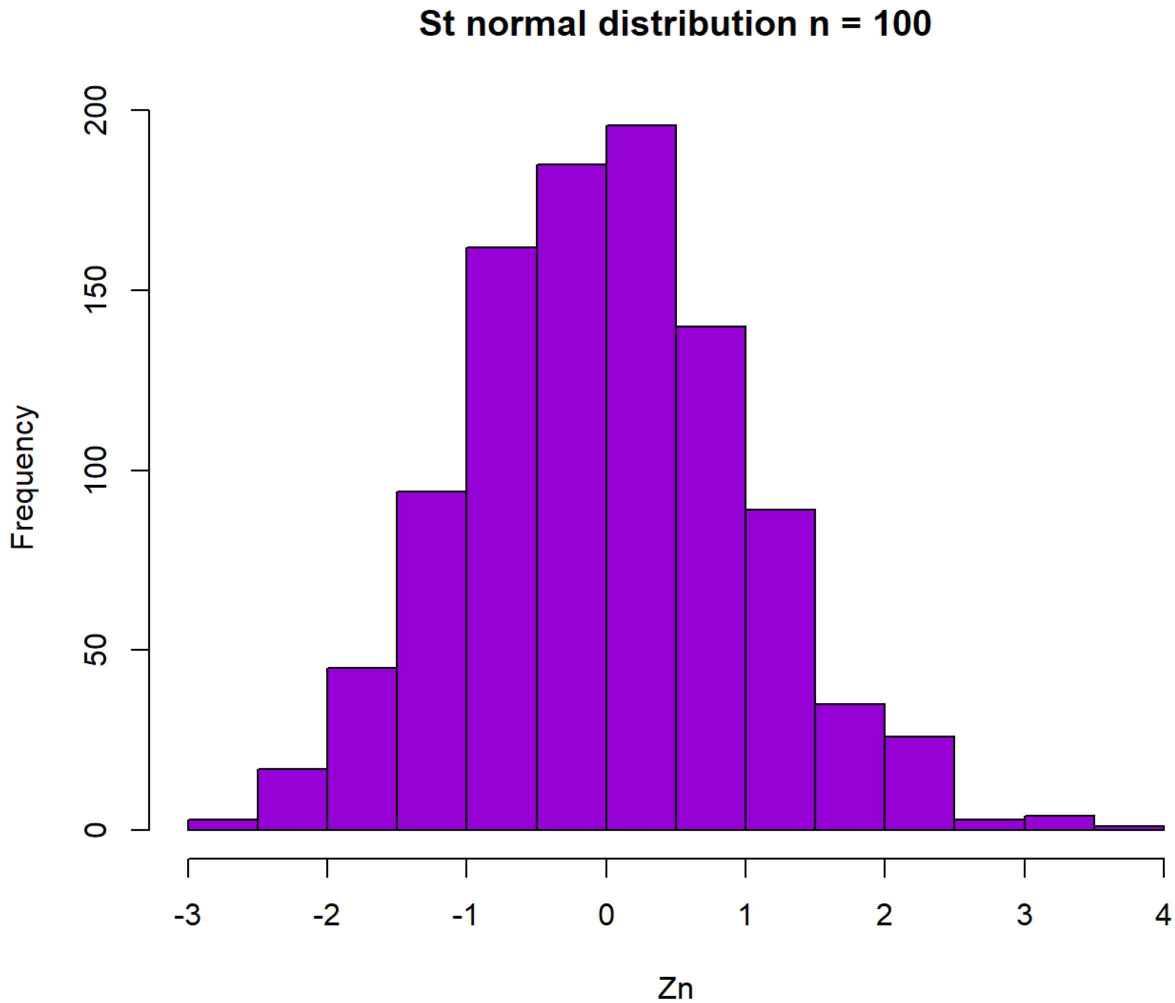
b. Para $n = 5, 10, 100, 500, 1000, 10000$, $m = 1000$ y $\lambda = 1$, utilice la función del inciso anterior para obtener muestras de Zn. Grafíque las muestras anteriores en un histograma (un histograma para cada n). ¿Qué observa? ¿Qu´e tiene que ver su resultado con el TCLC?

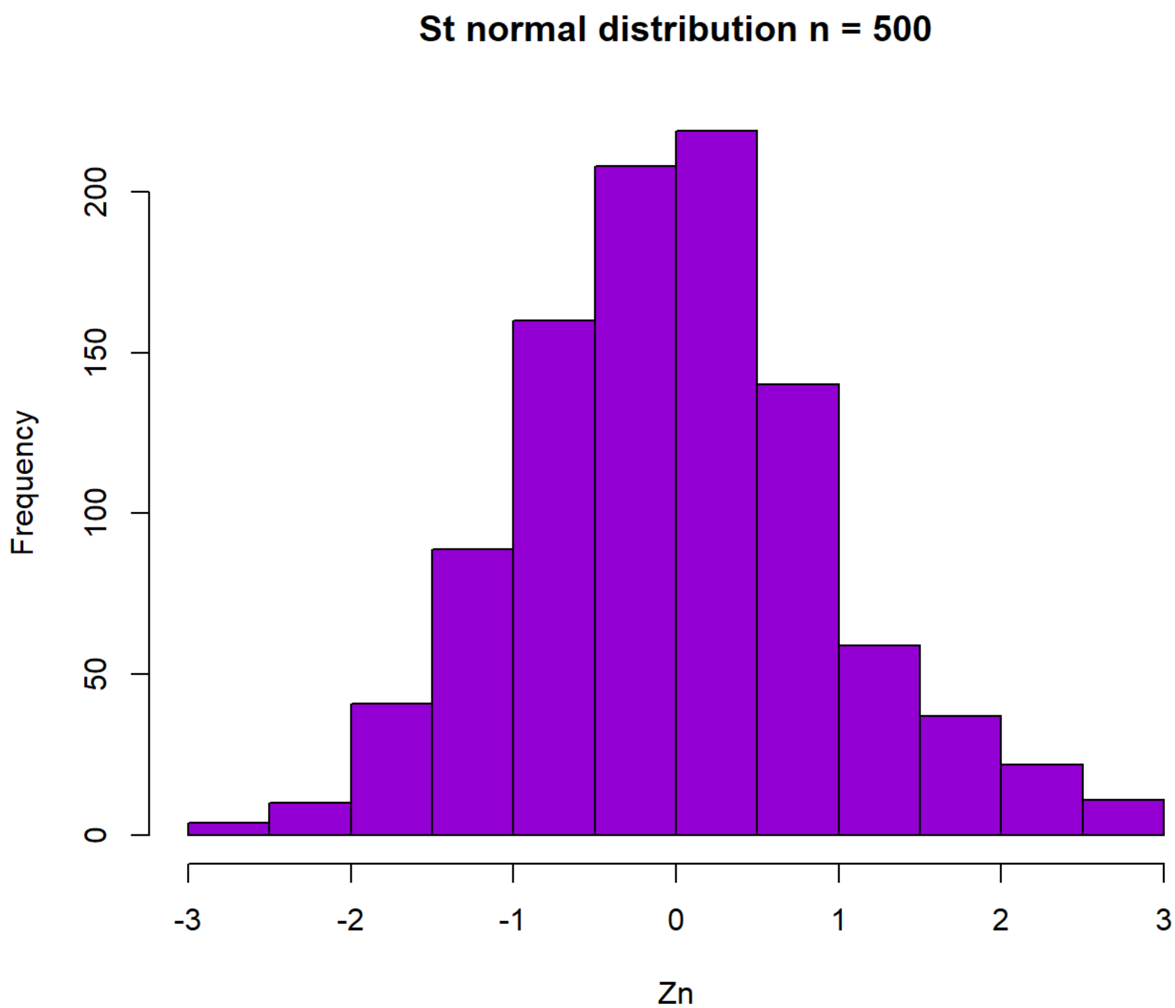
Code

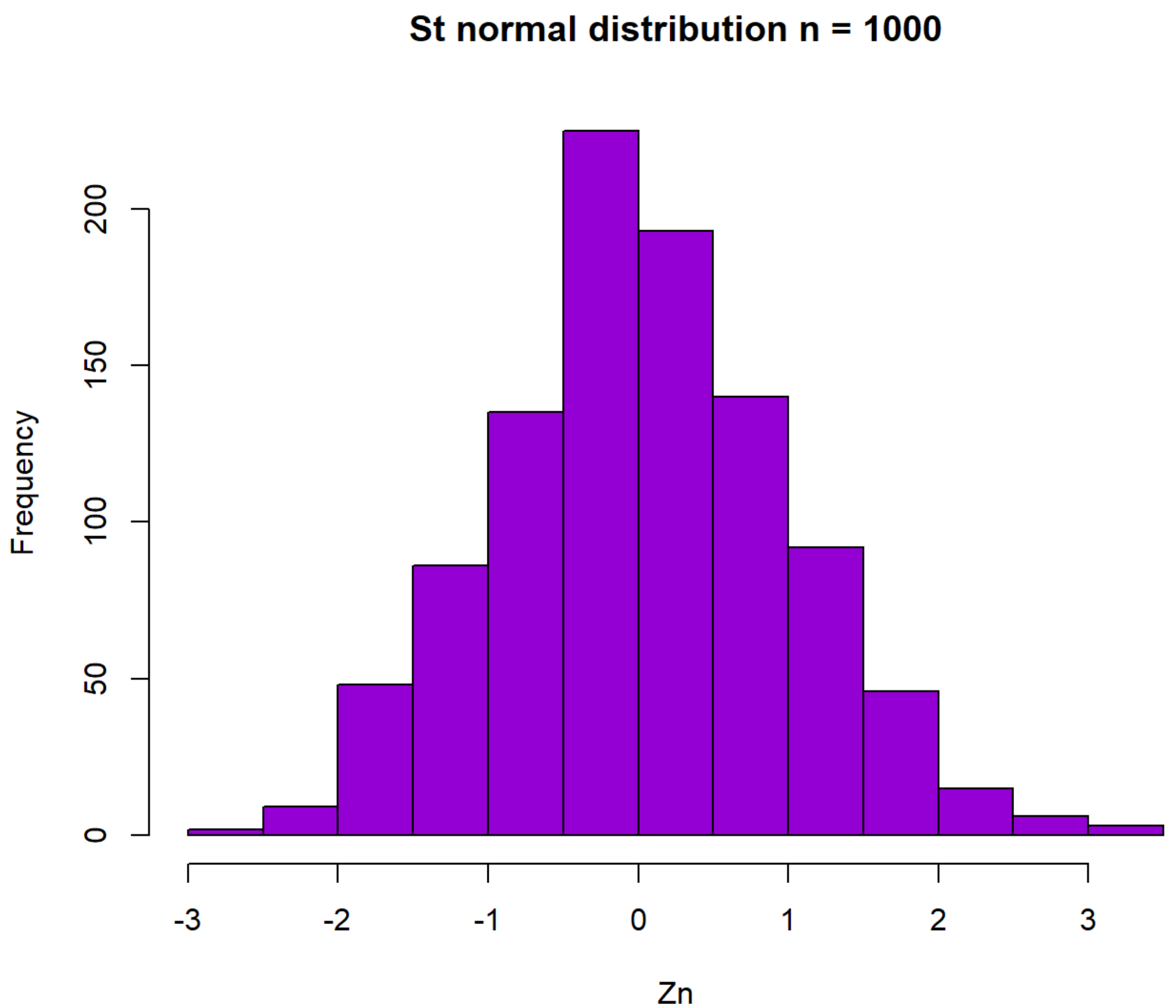

St normal distribution n = 5
Frequency
0
50
100
150
200
-2
0
2
4
Zn

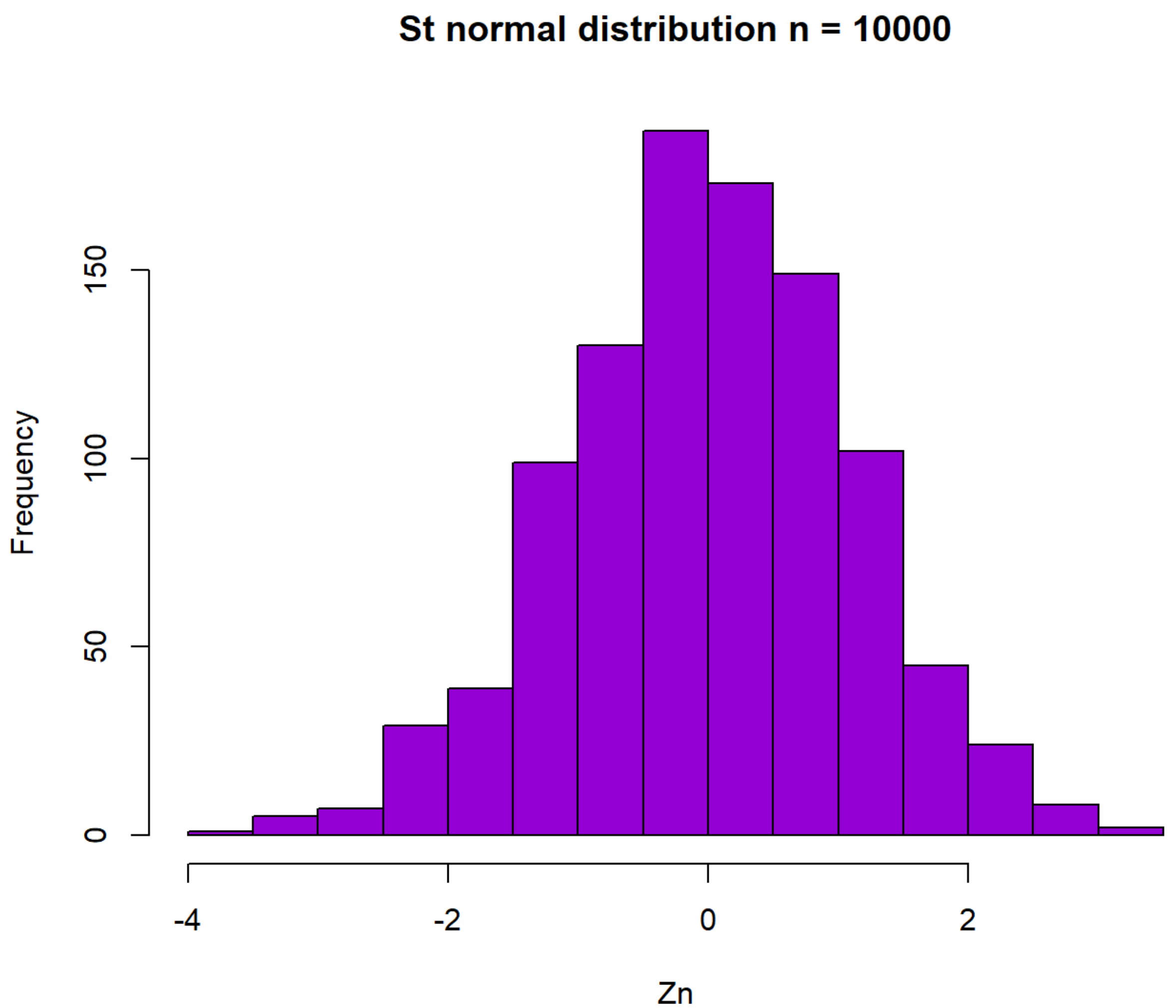

St normal distribution n = 10
Frequency
100
200
300


0
-2
0
2
4
6
Zn


St normal distribution n = 100
Frequency
0
50
100
150
200
-3
-2
-1
0
1
2
3
4
Zn


St normal distribution n = 500
Frequency
0
50
100
150
200
-3
-2
-1
0
1
2
3
Zn


St normal distribution n = 1000
Frequency
0
50
100
150
200
-3
-2
-1
0
1
2
3
Zn

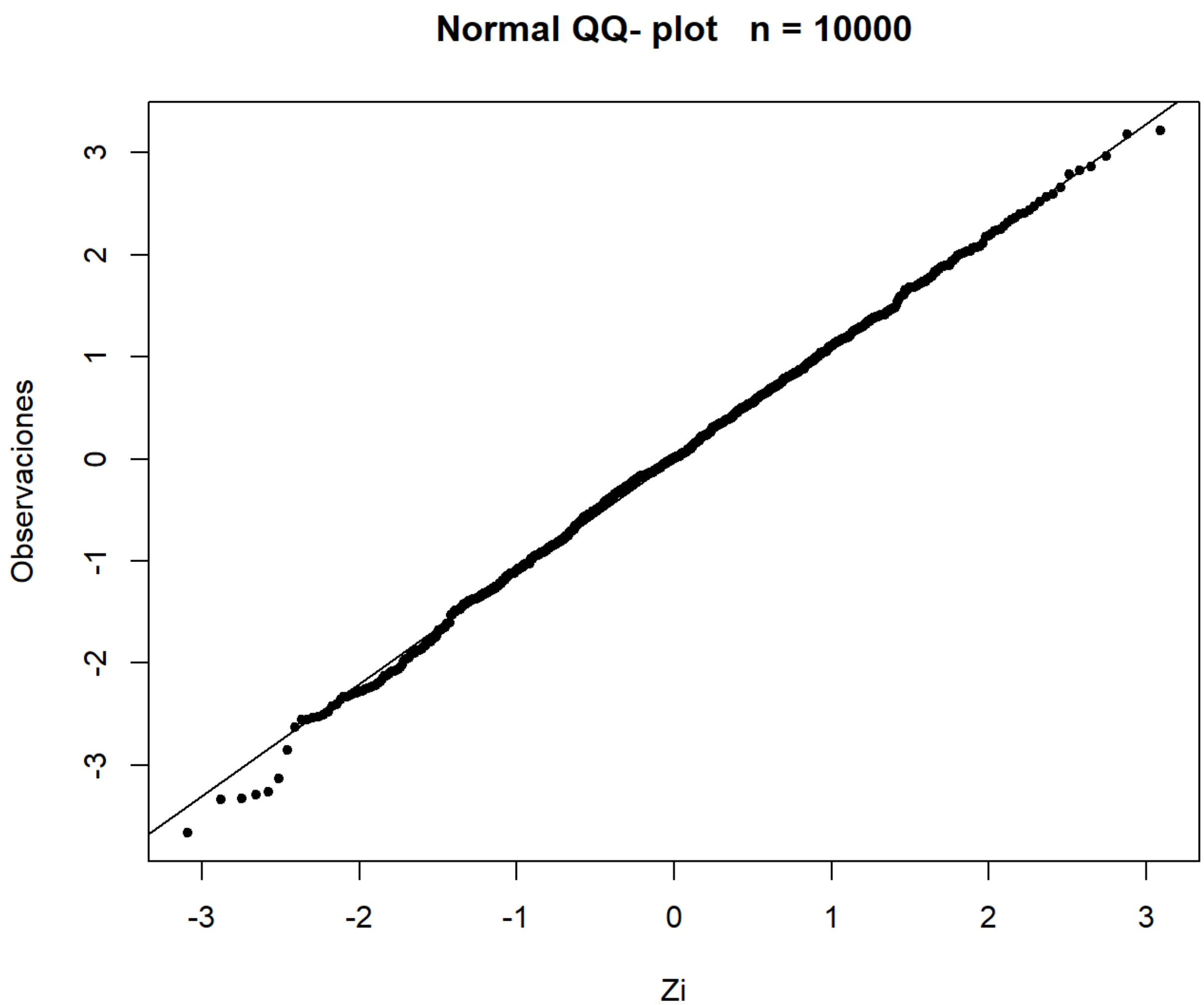


¿Qué observa? Se observa que al repetir el calculo del estadístico 1000 veces, y el valor de n se vuelve cada vez más grande, la variable aleatoria Zn, presenta una distribución limite parecida a la normal estandar.

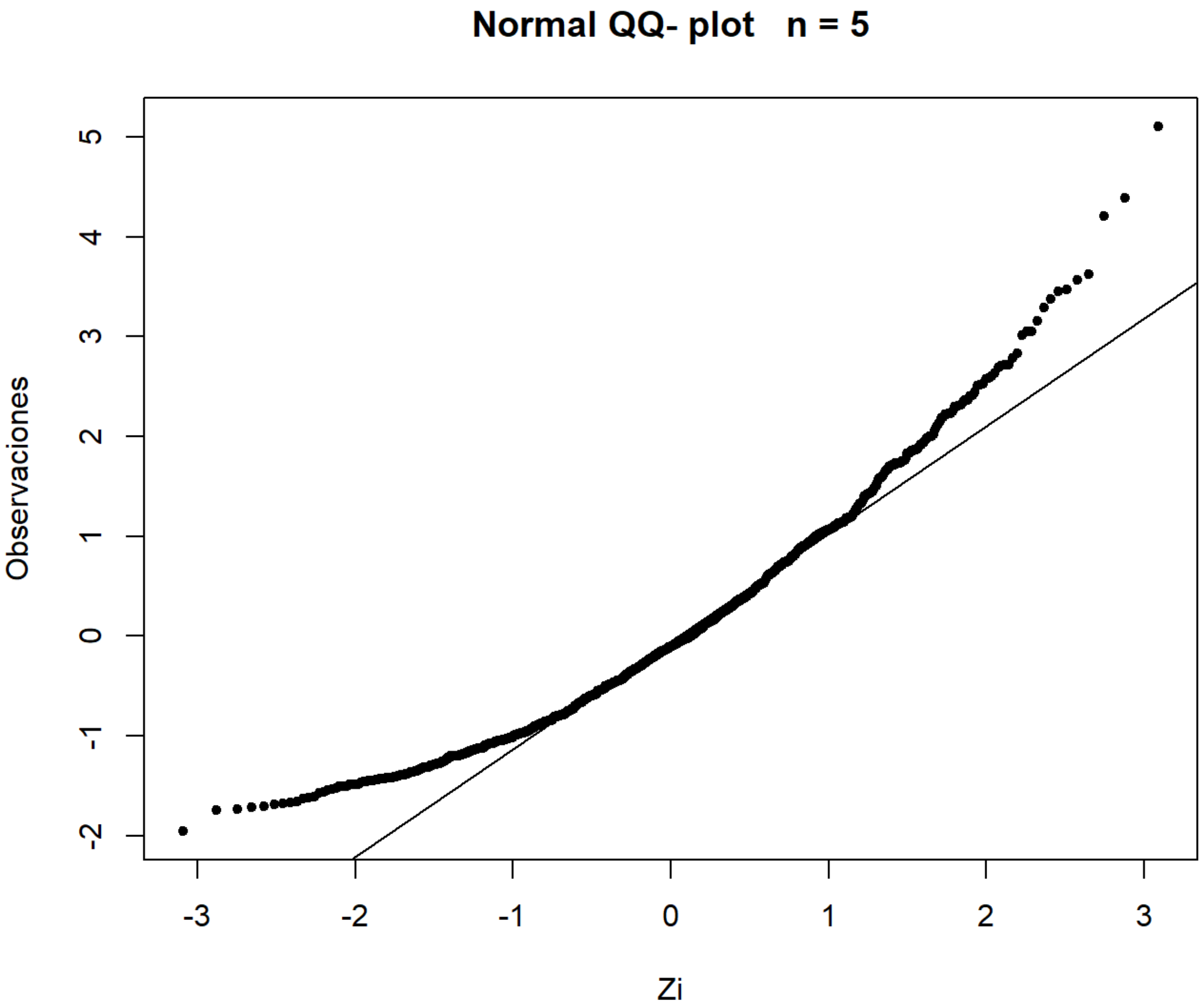
¿Qué tiene que ver su resultado con TCL? El teorema nos dice que una secuencia de variables aleatorias independientes e indenticamentes distribuidas - en nuestro caso las exponenciales con parámetro lambda-, cuya generadora de momentos exista, y con ello media y varianza sea finita. Entonces, al normalizar la media de la secuencia de variables aleatorias e incrementar n, la variable aleatoria estandar tiende a una distribución límite, la cual es una normal estandar. Por lo tanto, al normalizar cualquier variable aleatoria , pero que cumpla con lo necesarios previamente mencionado, entonces, dicha sucesión de variables estadarizadas convergerá en distribución a una normal estandar.

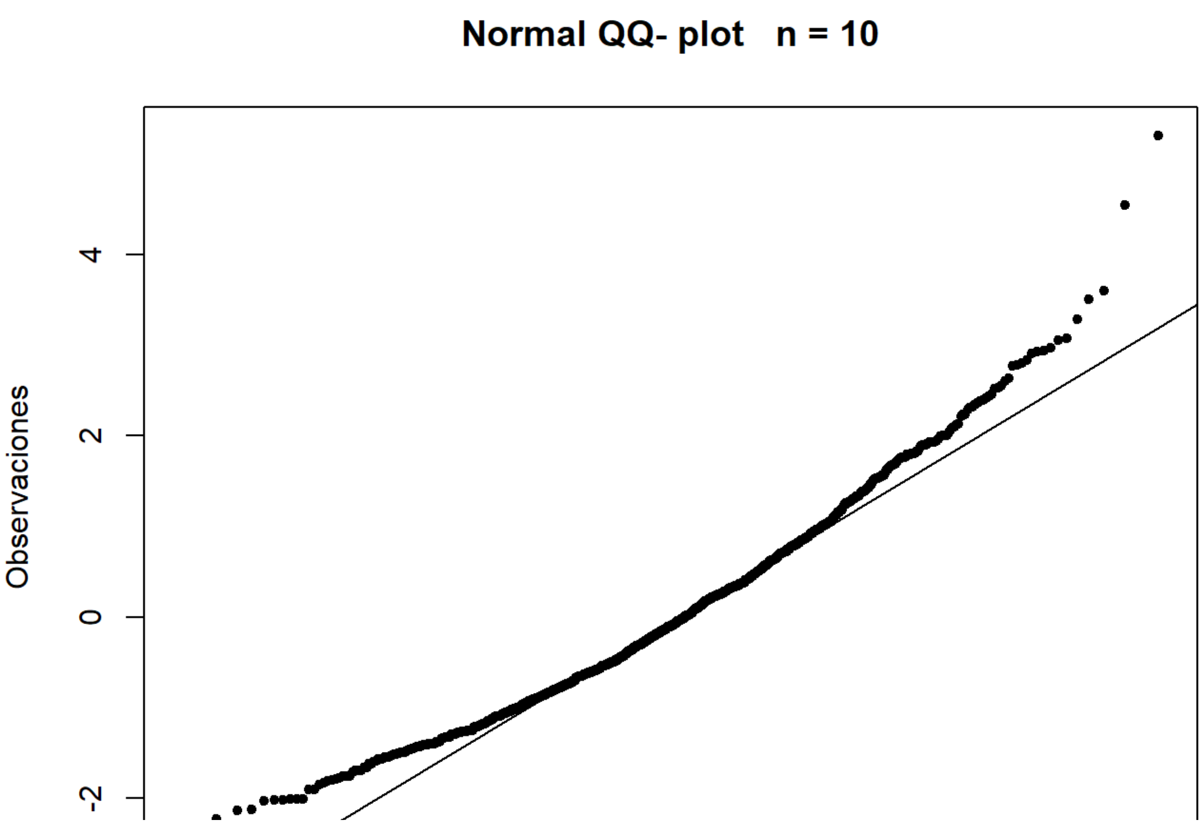
c. Para cada una de las muestras generadas en el inciso anterior, encuentre el Q-Q plot y el P-P plot normales. Comente sus resultados.

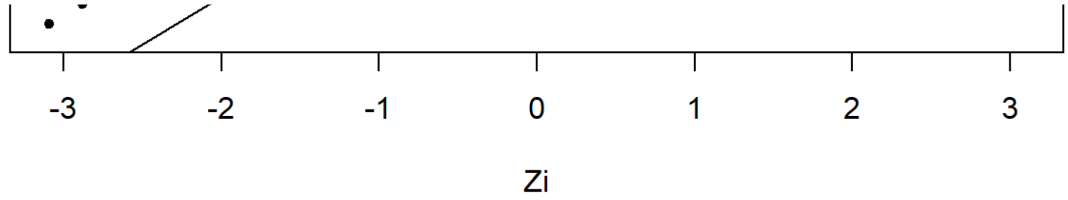
Code



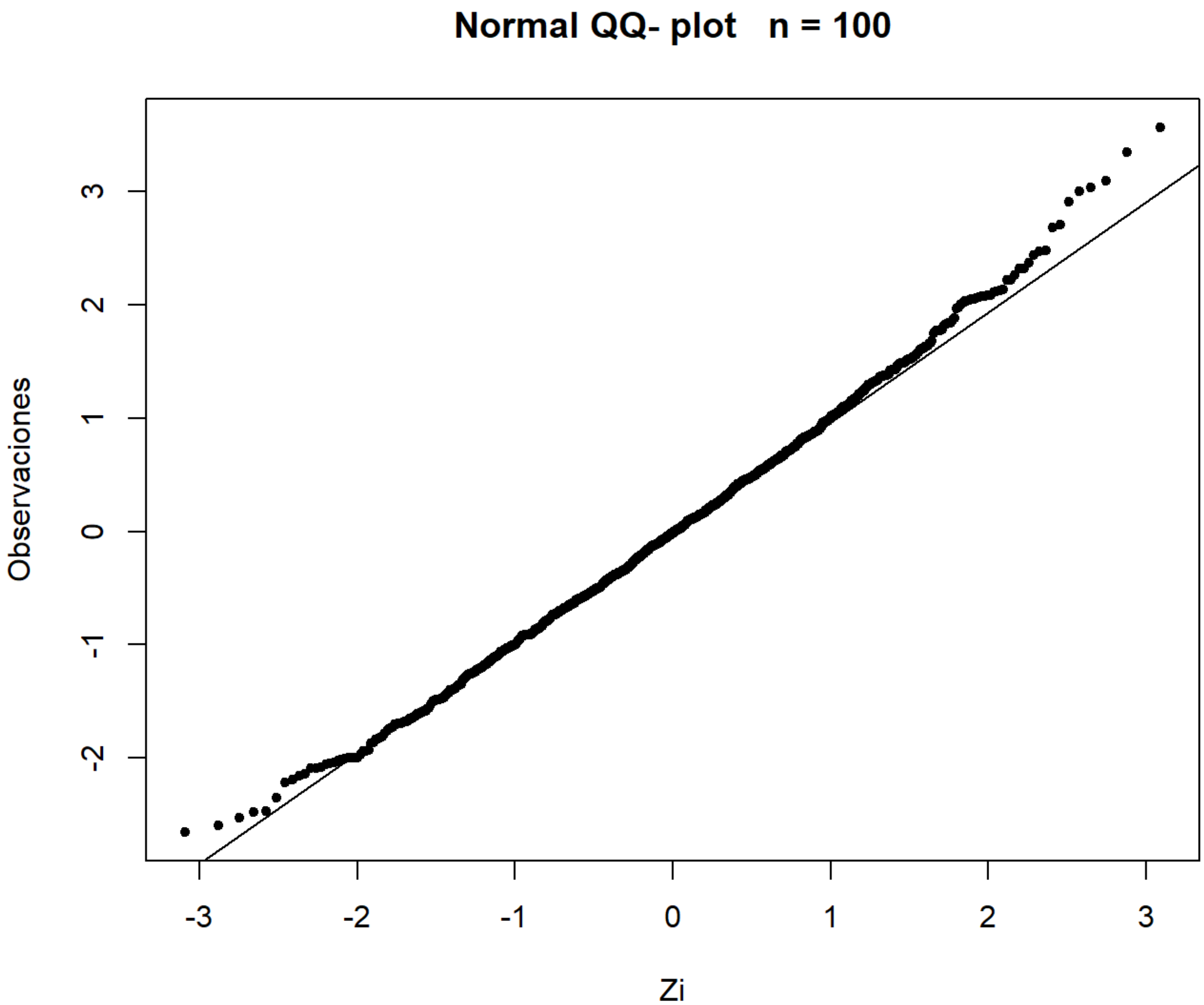
Code


Normal QQ- plot n = 5
Observaciones
Zi
5
4
3
2
1
0
-1
-2
-3
-2
-1
0
1
2
3


Normal QQ- plot n = 10
Observaciones
4
2
0
-2



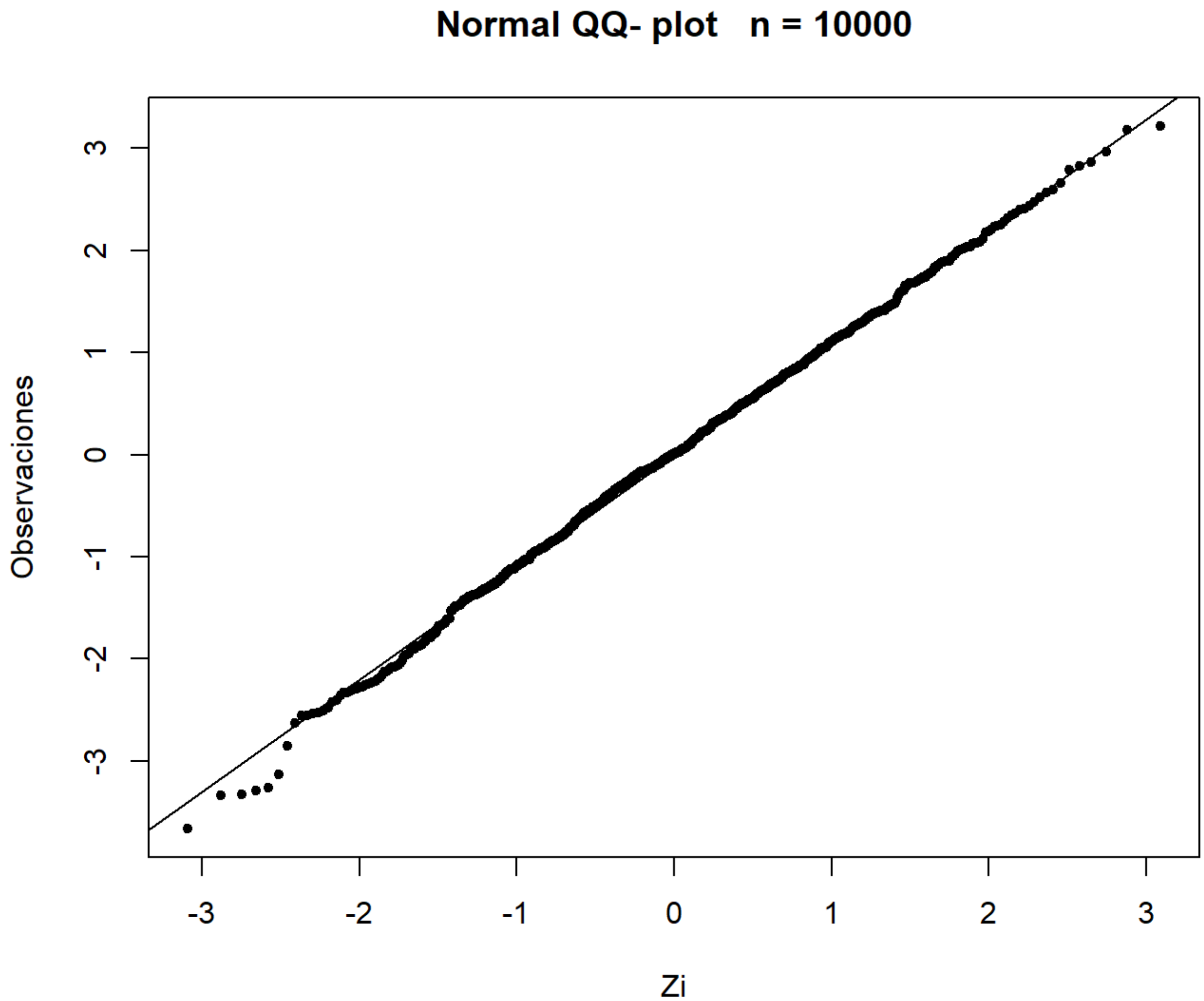
-3
-2
-1
0
1
2
3
Zi



Normal QQ- plot n = 100
Observaciones
-2
-1
0
1
2
3
-3
-2
-1
0
1
2
3
Zi


Normal QQ- plot n = 500
Observaciones
Zi
3
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3

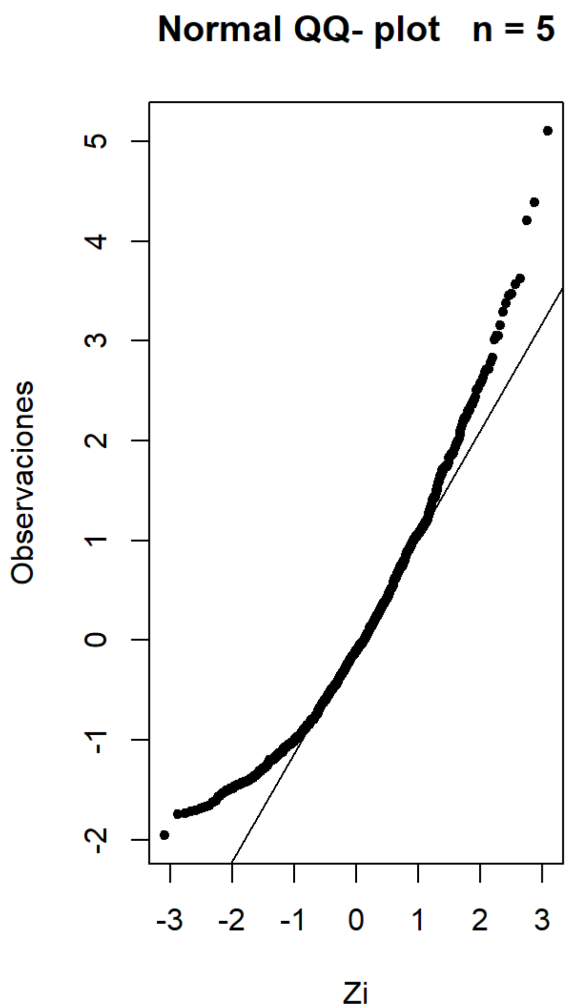

Normal QQ- plot n = 1000
Observaciones
Zi
3
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3

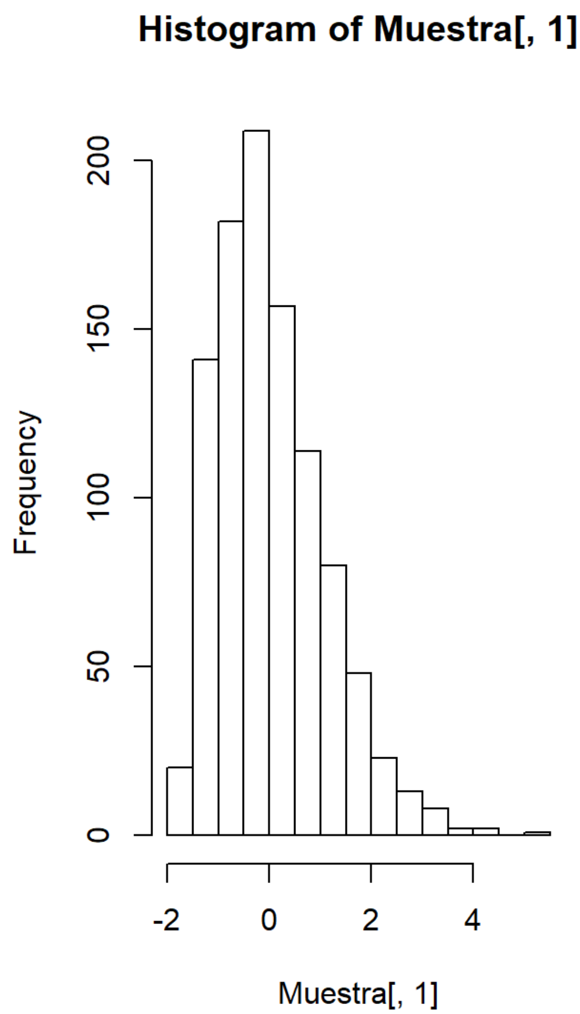


Cuando la muestra es de n = 5, claramente se observa que presenta sesgo a la derecha (izquierda para el que observa). y por ende, no aproxima correctamente a una normal. Dicho sesgo, se va corrigiendo. Para una muestra de 10000 observaciones, la distribución mmuestral converge en distribución a una normal. Lo cual, se ve representado por el qq-plot (los puntos sobre la gráfica se encuntran sobre la recta).

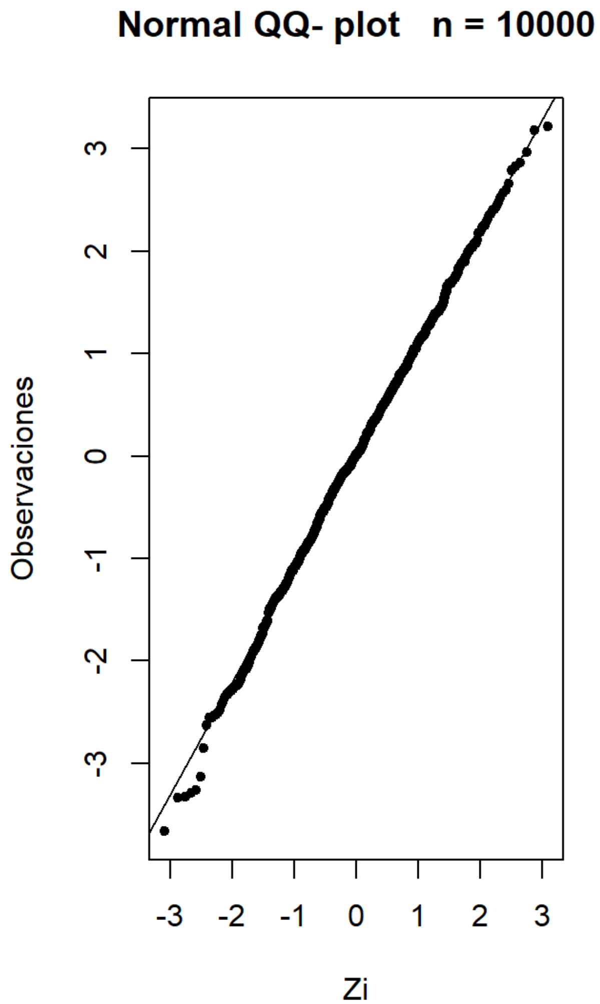
En conclusión, bajo ciertas condiciones, al estandarizar las variables aleatorias cuya distribución es una exponencial, cuando la muestra (n) tiende a infinito (sea grande), entonces, la variable aleatoria estandarizada tiende a converger en distribución a una distribución límite normal estandar.

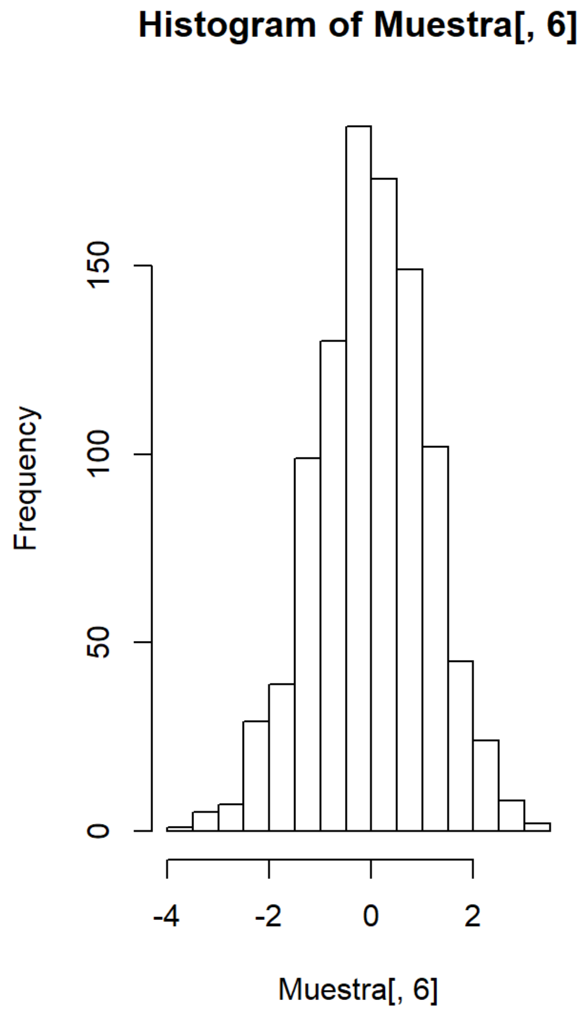
Code


Normal QQ- plot n = 5
Observaciones
Zi
-3
-2
-1
0
1
2
3
4
5

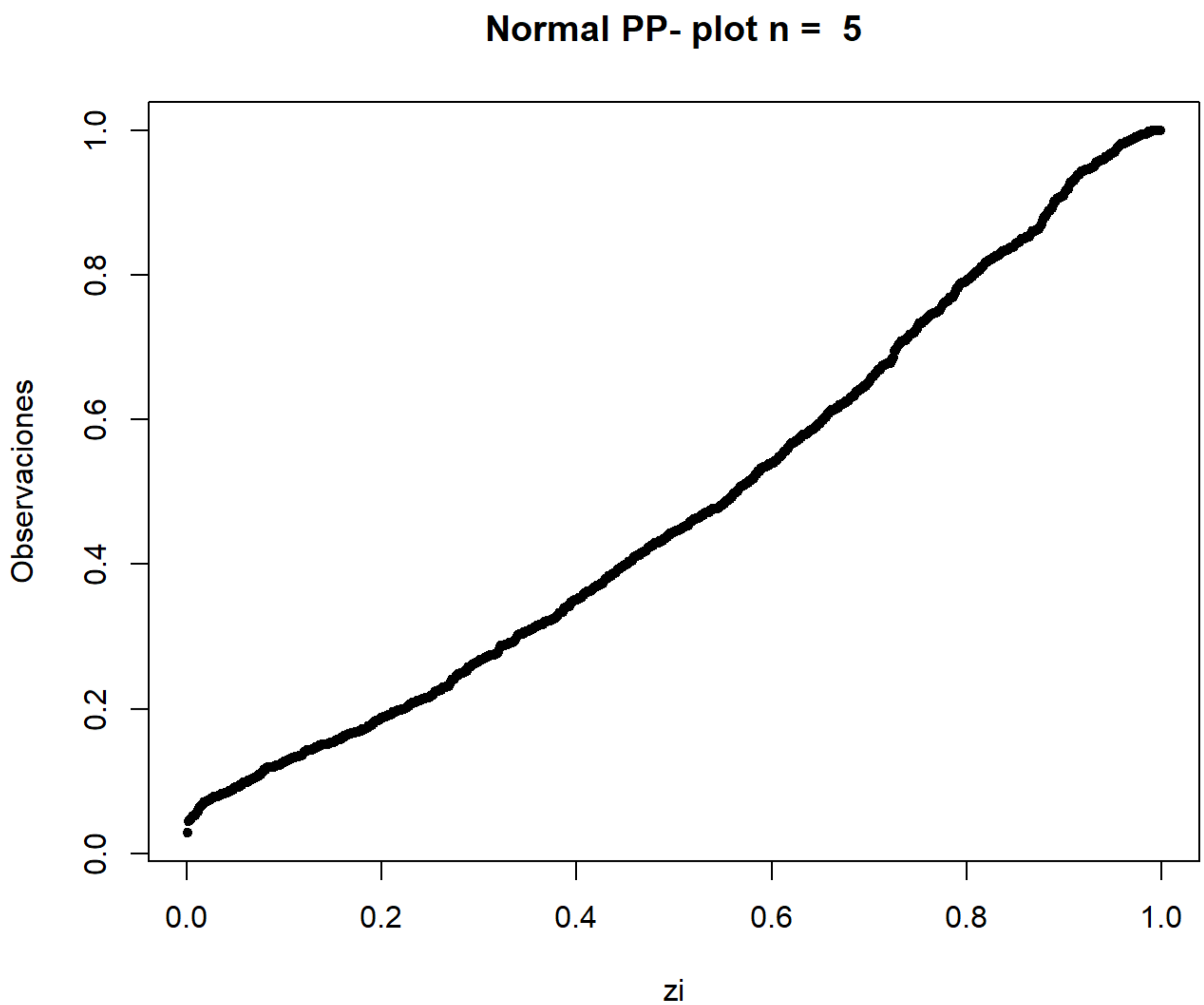

Histogram of Muestra[, 1]
Frequency
Muestra[, 1]
0
50
100
150
200
-2
0
2
4

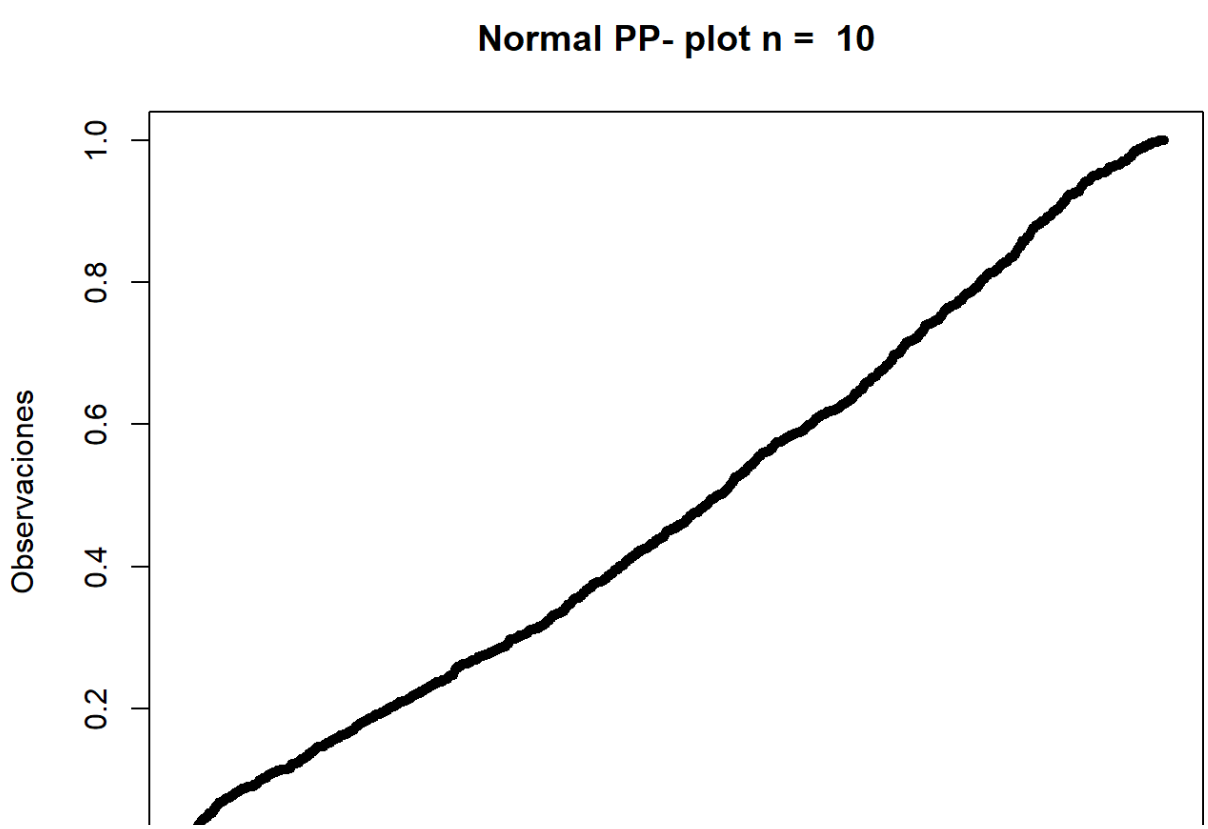
Code

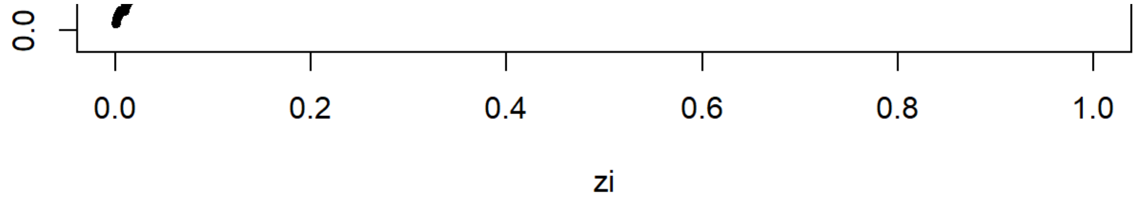

Normal QQ- plot n = 10000
Observaciones
3
2
1
0
-1
-2
-3
-3 -2 -1 0 1 2 3
Zi

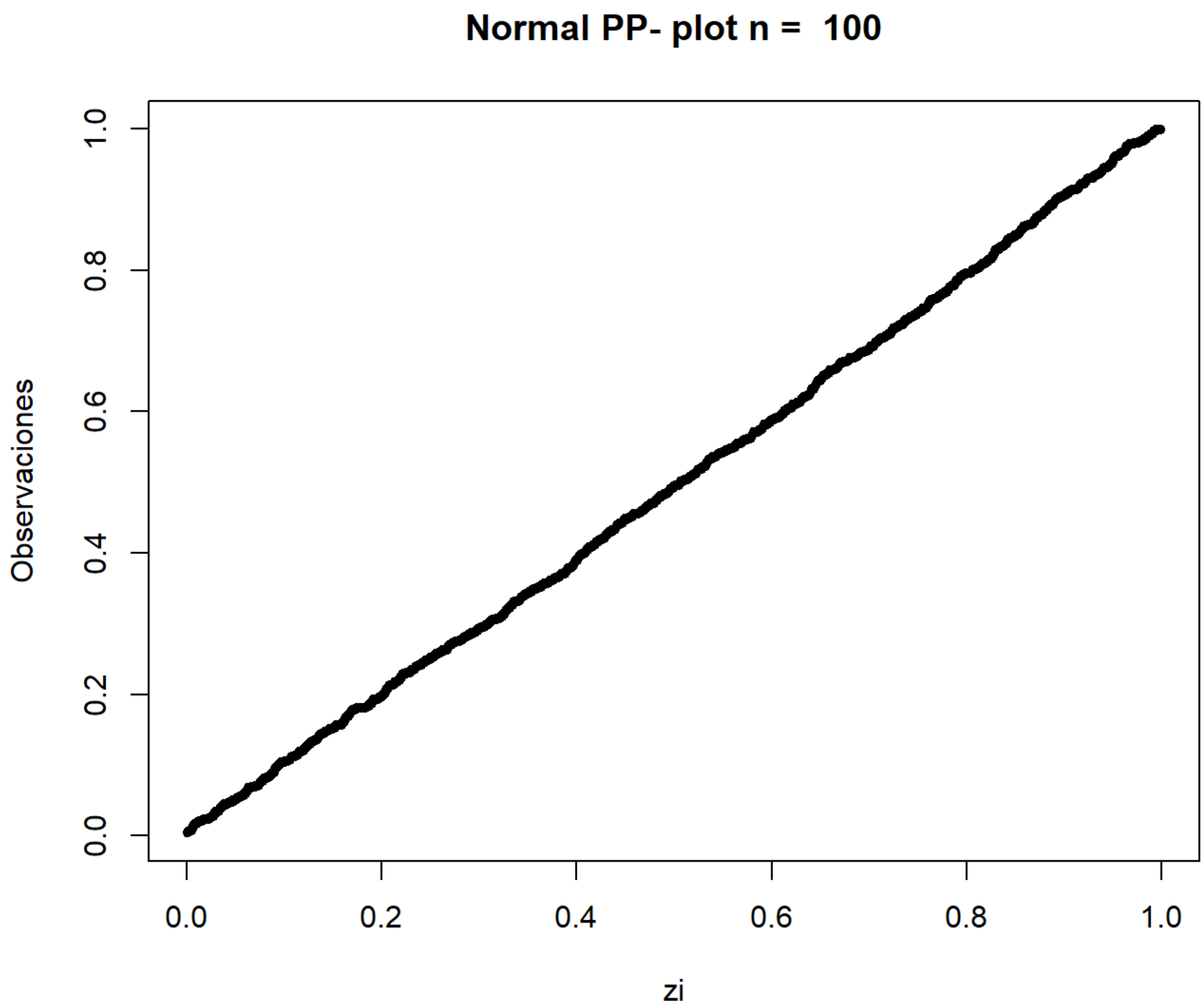

Histogram of Muestra[, 6]
Frequency
150
100
50
0
-4 -2 0 2
Muestra[, 6]

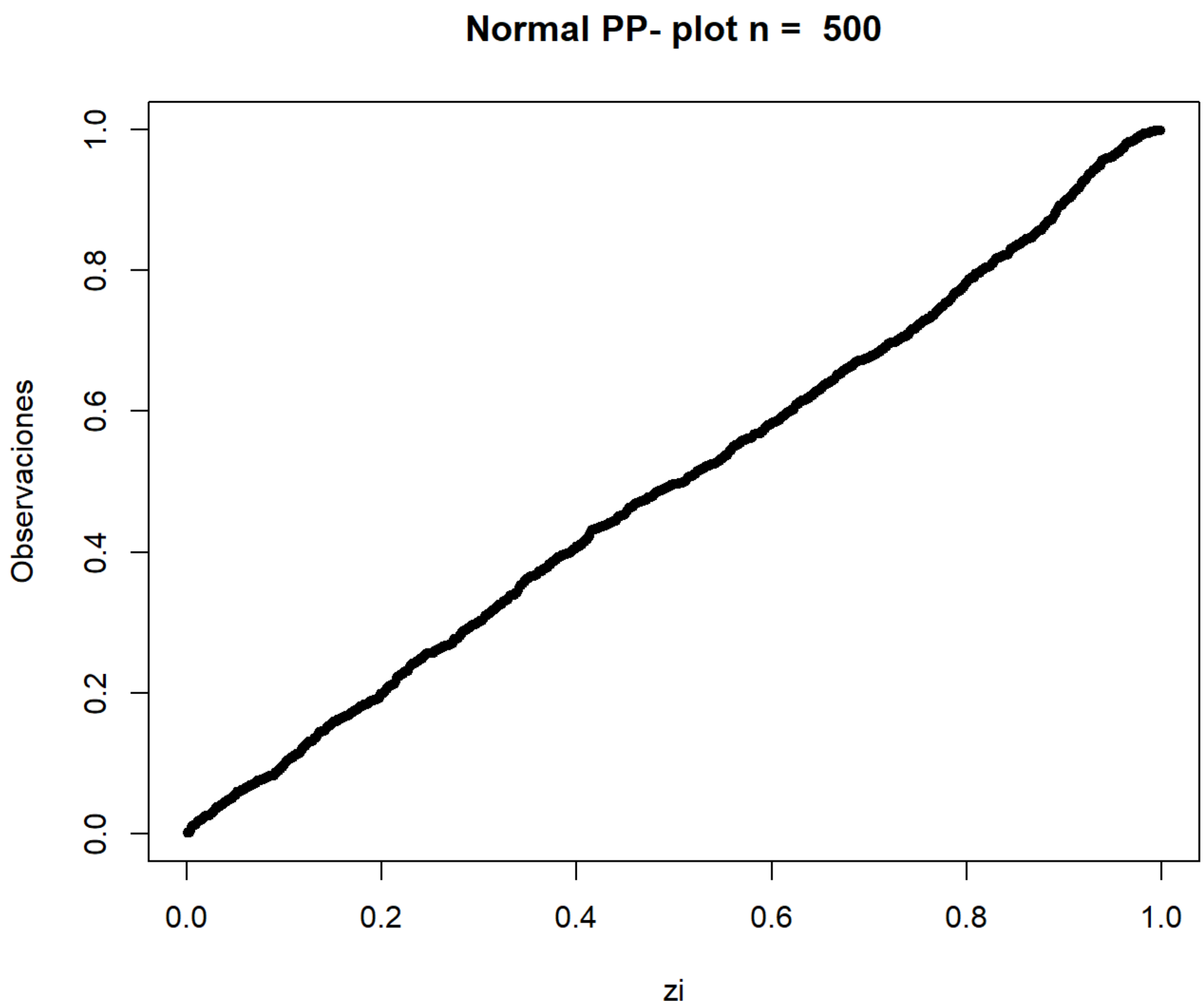
Code

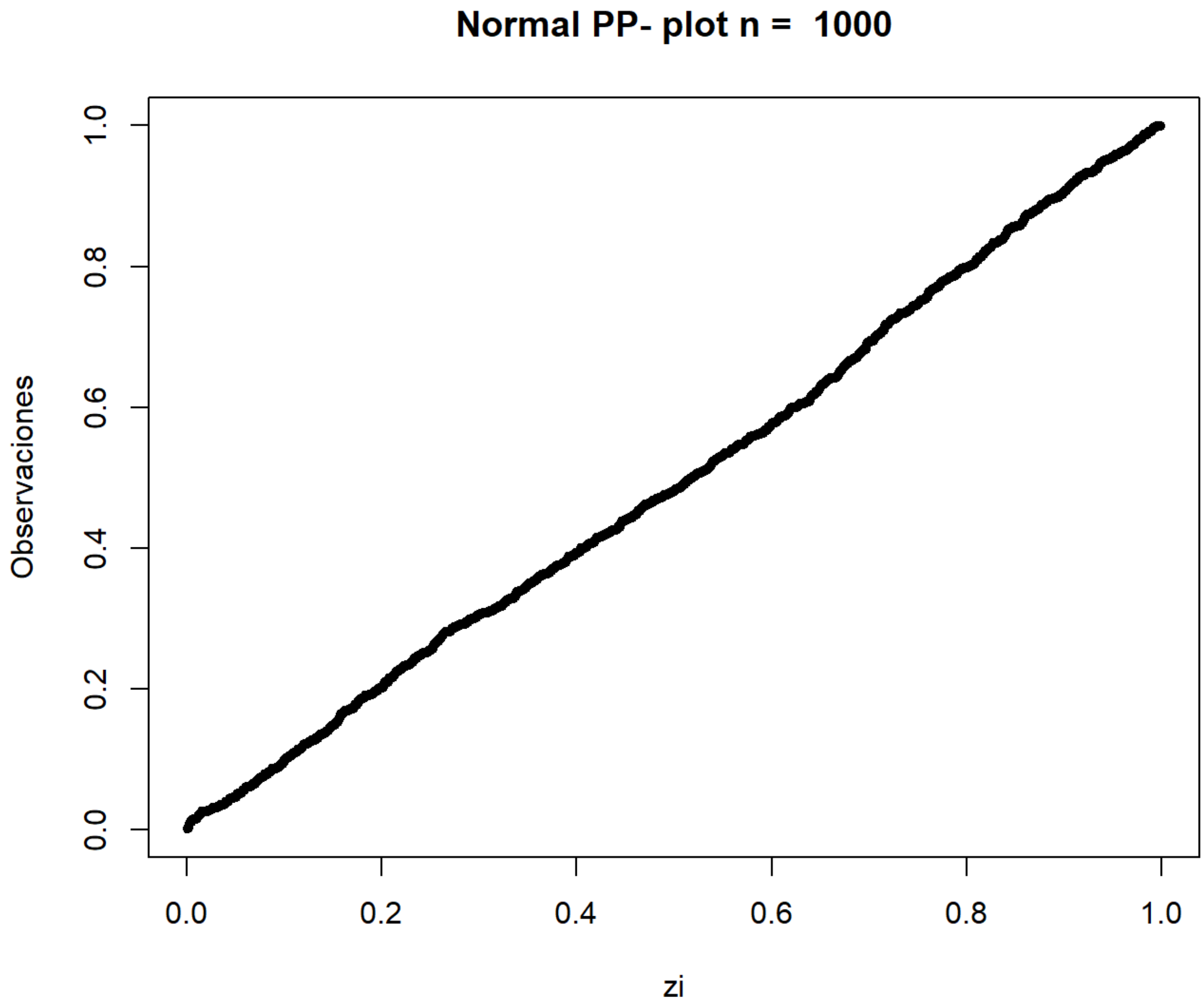

Normal PP- plot n = 5
Observaciones
zi
0.0
0.2
0.4
0.6
0.8
1.0

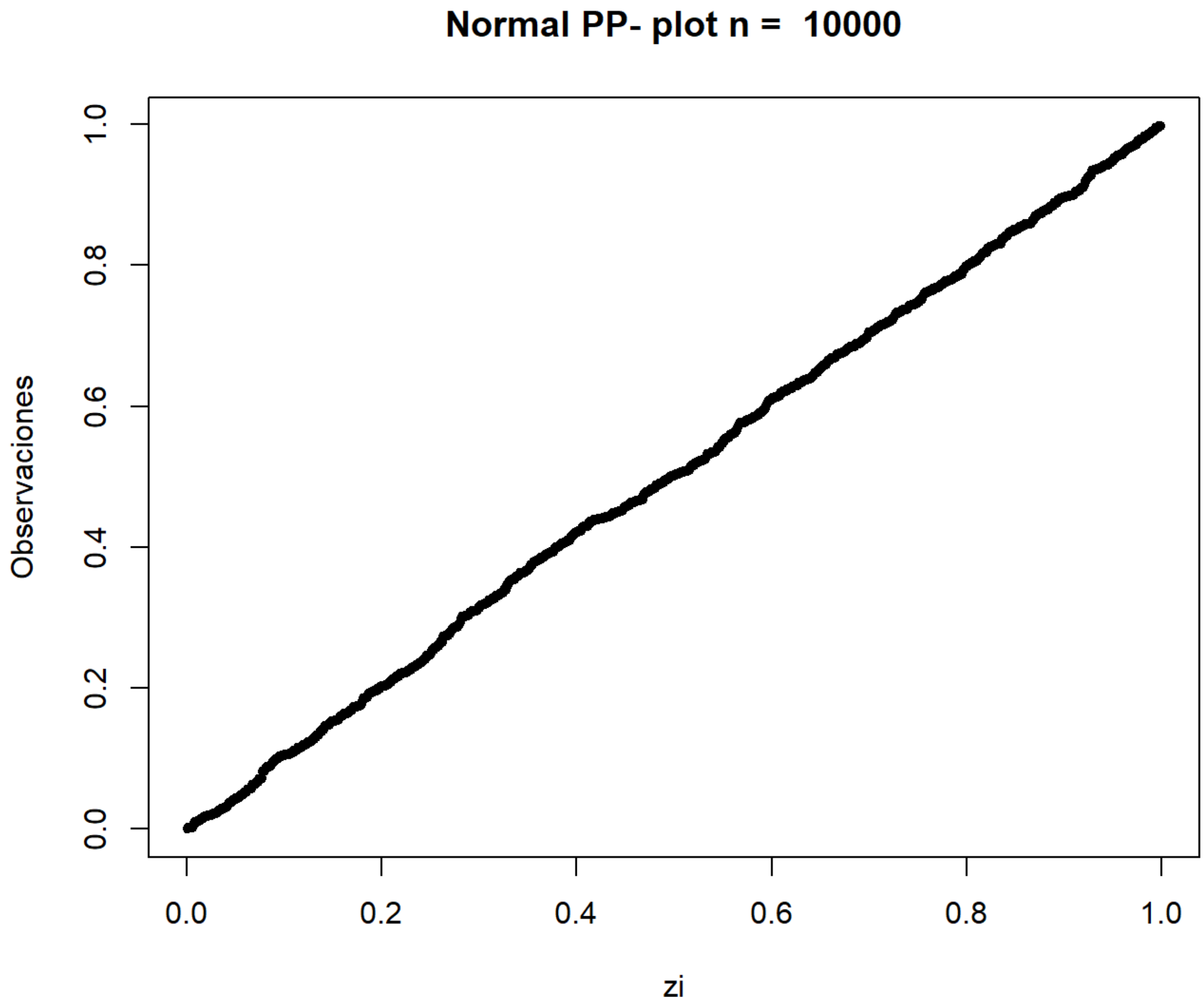

Normal PP- plot n = 10
Observaciones
0.2
0.4
0.6
0.8
1.0


0.0
0.0
0.2
0.4
0.6
0.8
1.0
zi


Normal PP- plot n = 100
Observaciones
0.0
0.2
0.4
0.6
0.8
1.0
zi


Normal PP- plot n = 500
Observaciones
zi
0.0
0.2
0.4
0.6
0.8
1.0


Normal PP- plot n = 1000
Observaciones
zi
0.0
0.2
0.4
0.6
0.8
1.0



Al igual que en el qq-plot, los resultados y conclusiones con el pp-plot son los mismos (revisar lo anterior escrito).

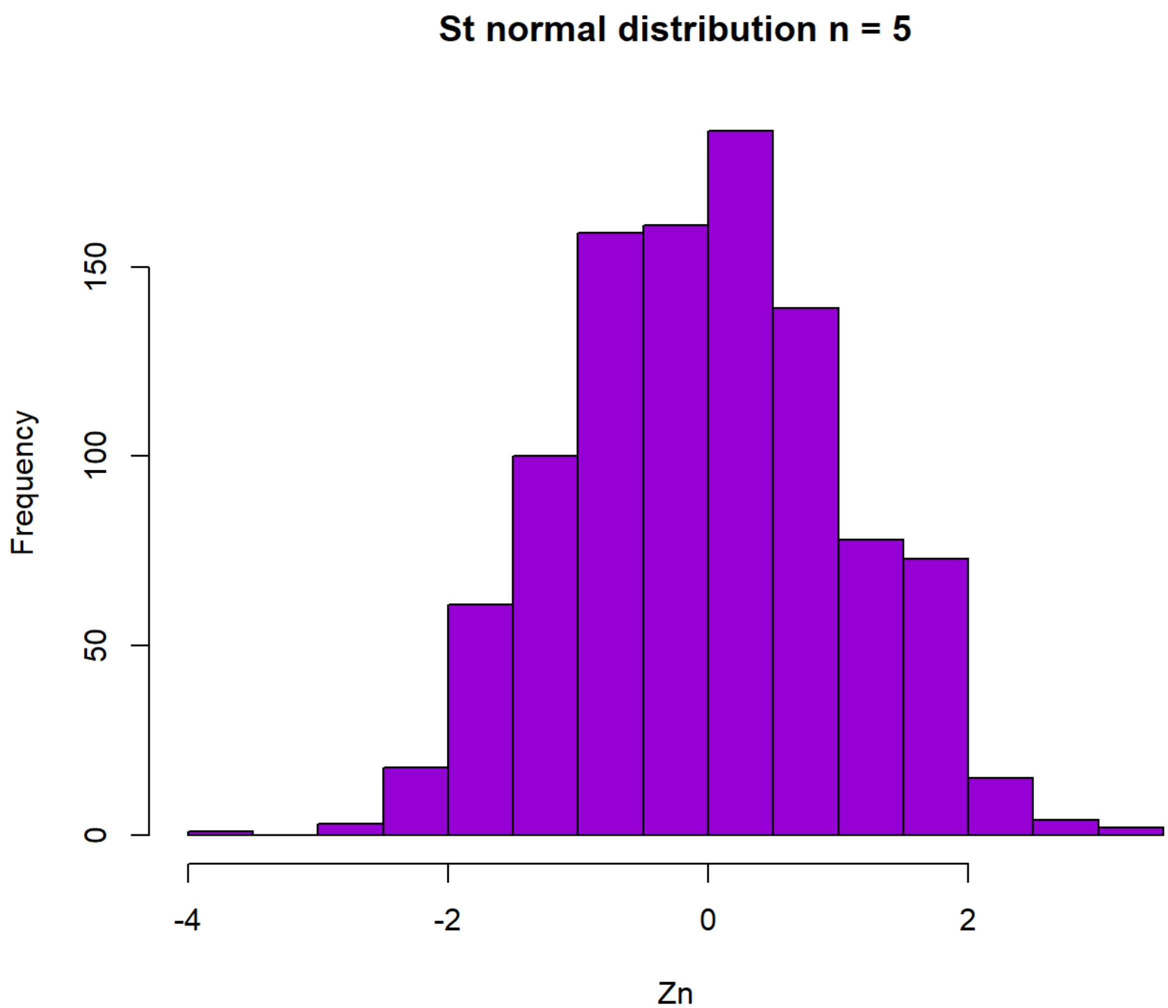
EJERCICIO 3

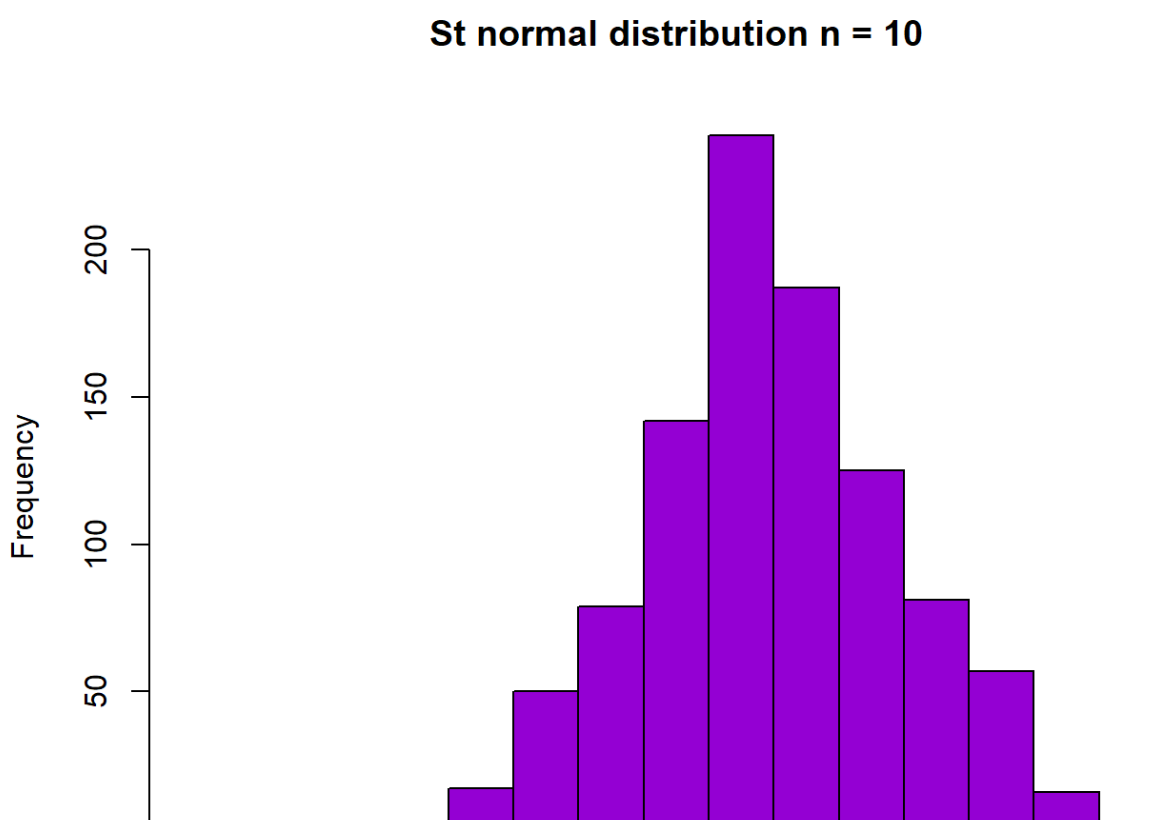
En este ejercicio volver´a a trabajar con el TCLC. a) Escriba una función análoga a la pedida en el inciso 2a) para una distribuciónn Binomial(p, N). La funciónn deber´a tomar los mismos parámetros a los pedidos en el inciso 2a), con excepción al parámetro λ que tendrá que ser sustituido p y N.

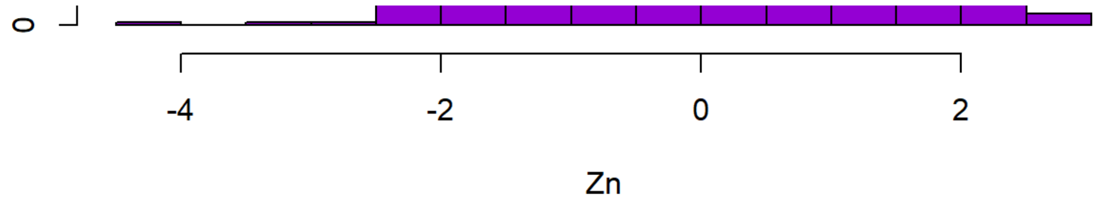
Code

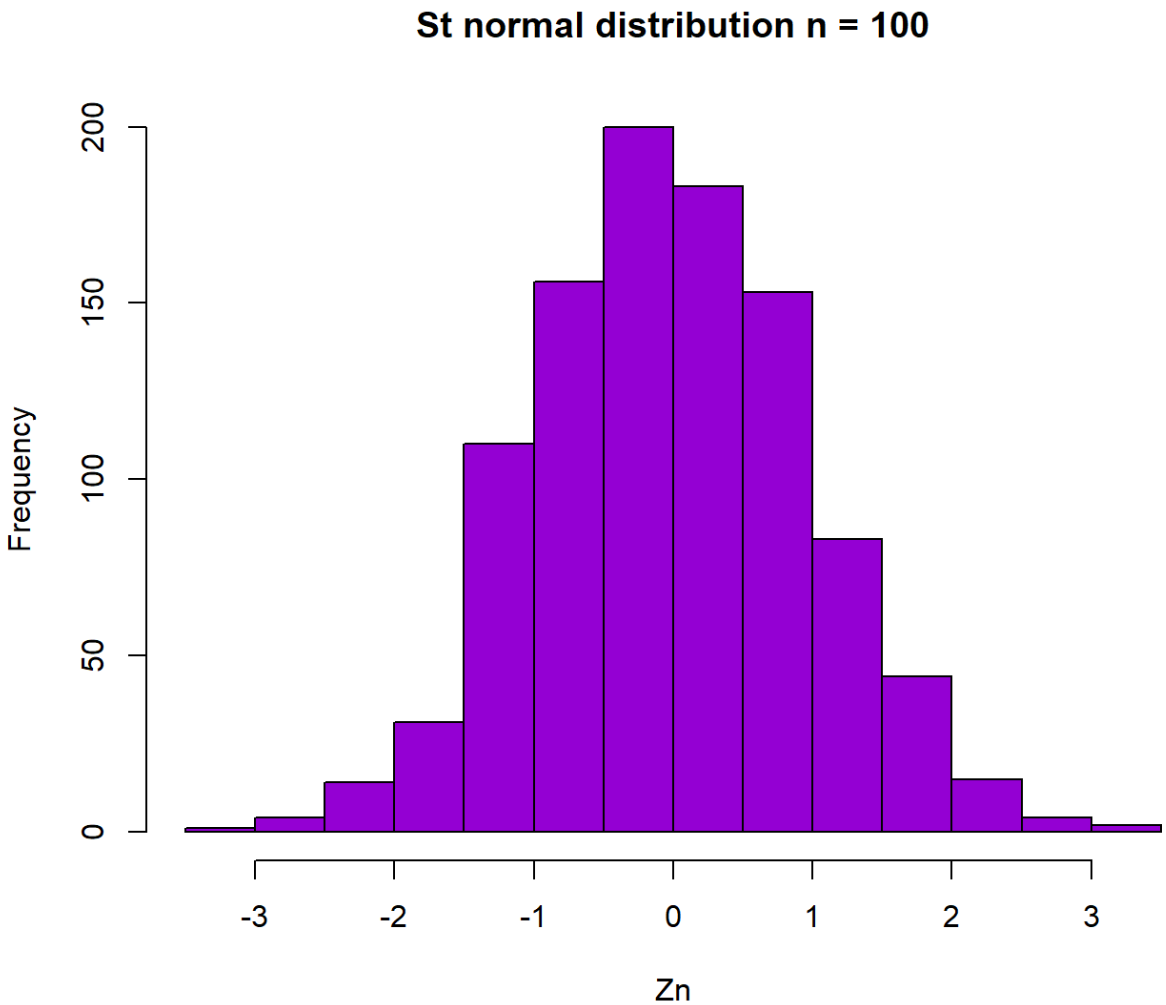
b. Para p = 1/2 y N = 15, repita los incisos 2b) y 2c) para el caso Binomial de este ejercicio.

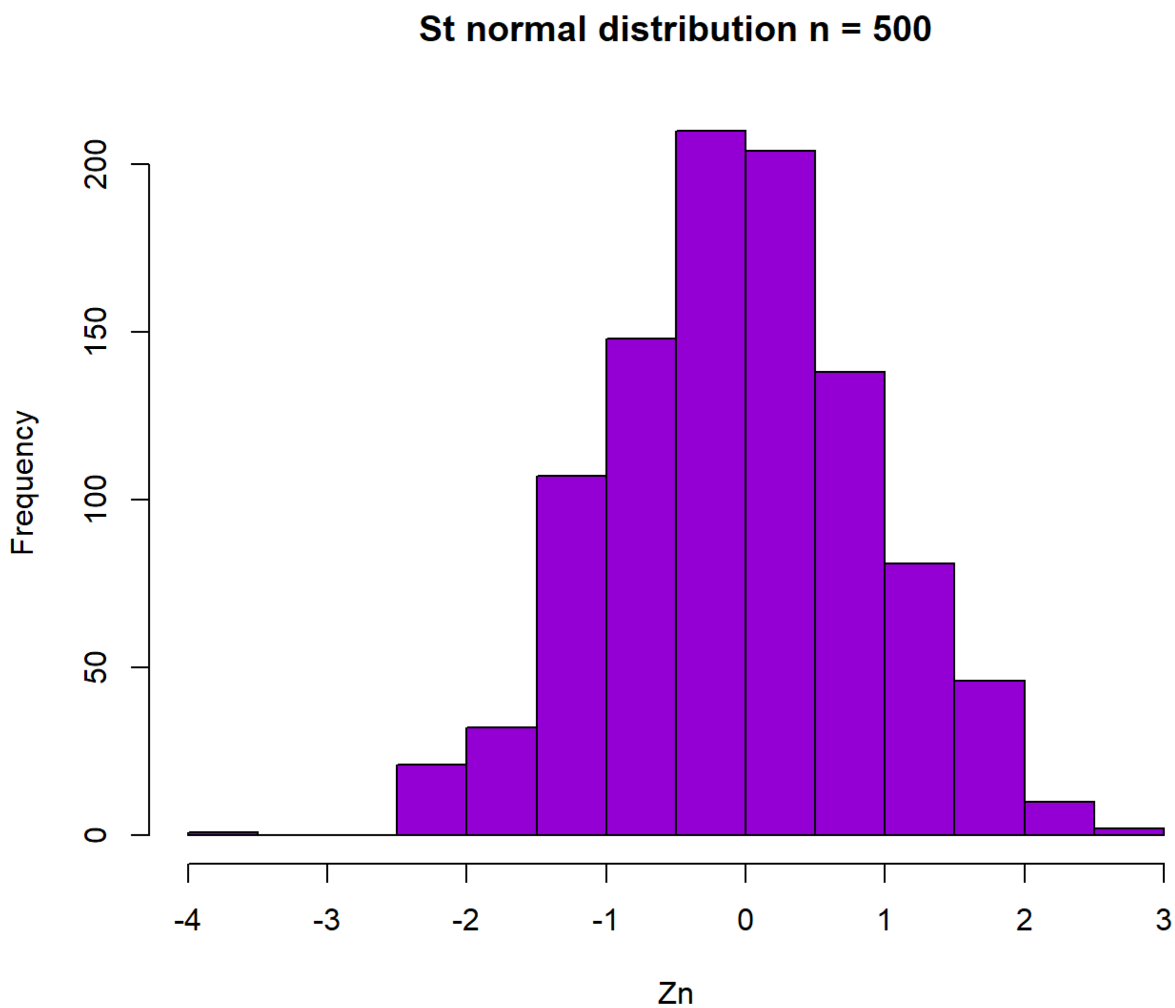
Code

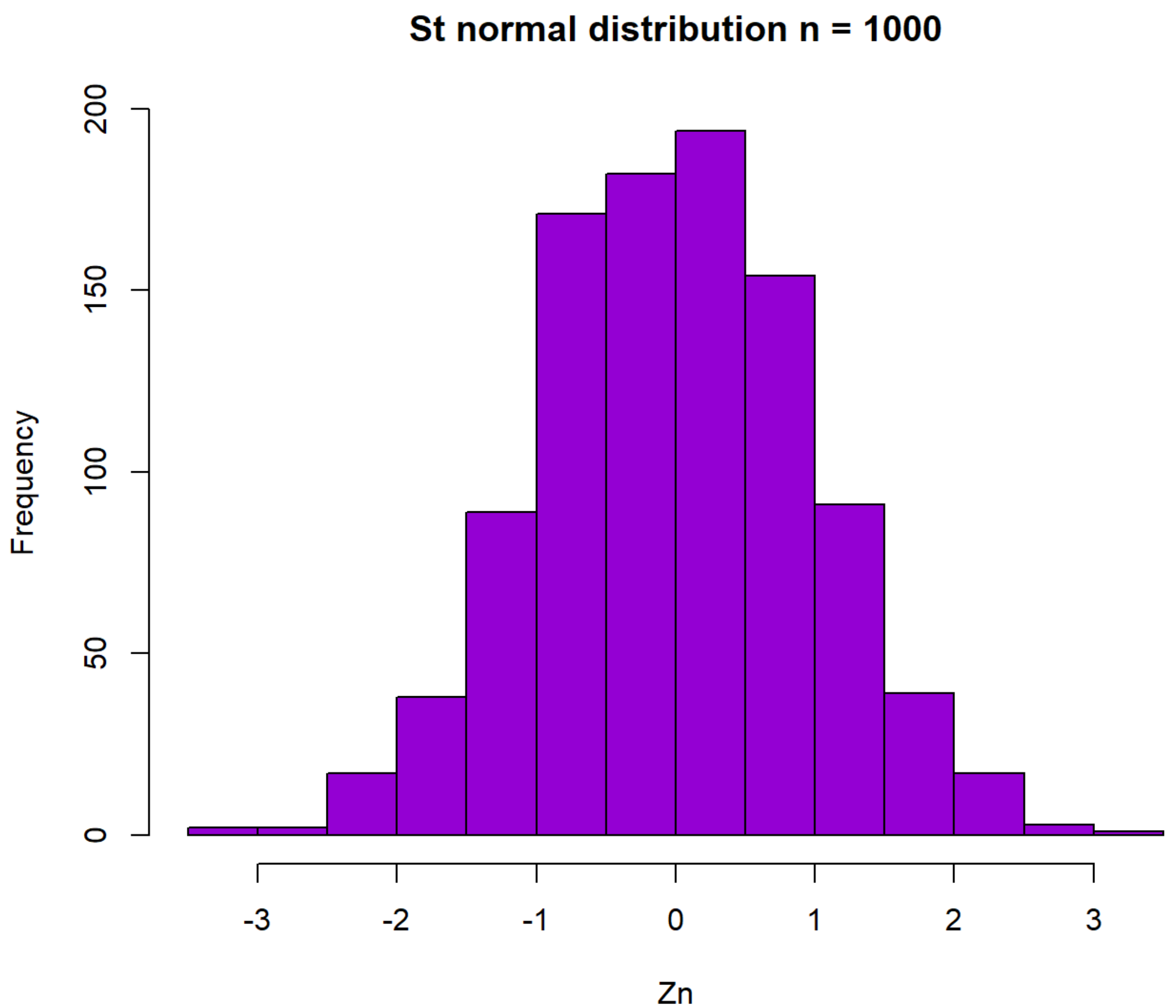

St normal distribution n = 5
Frequency
0
50
100
150
-4
-2
0
2
Zn


St normal distribution n = 10
Frequency
50
100
150
200

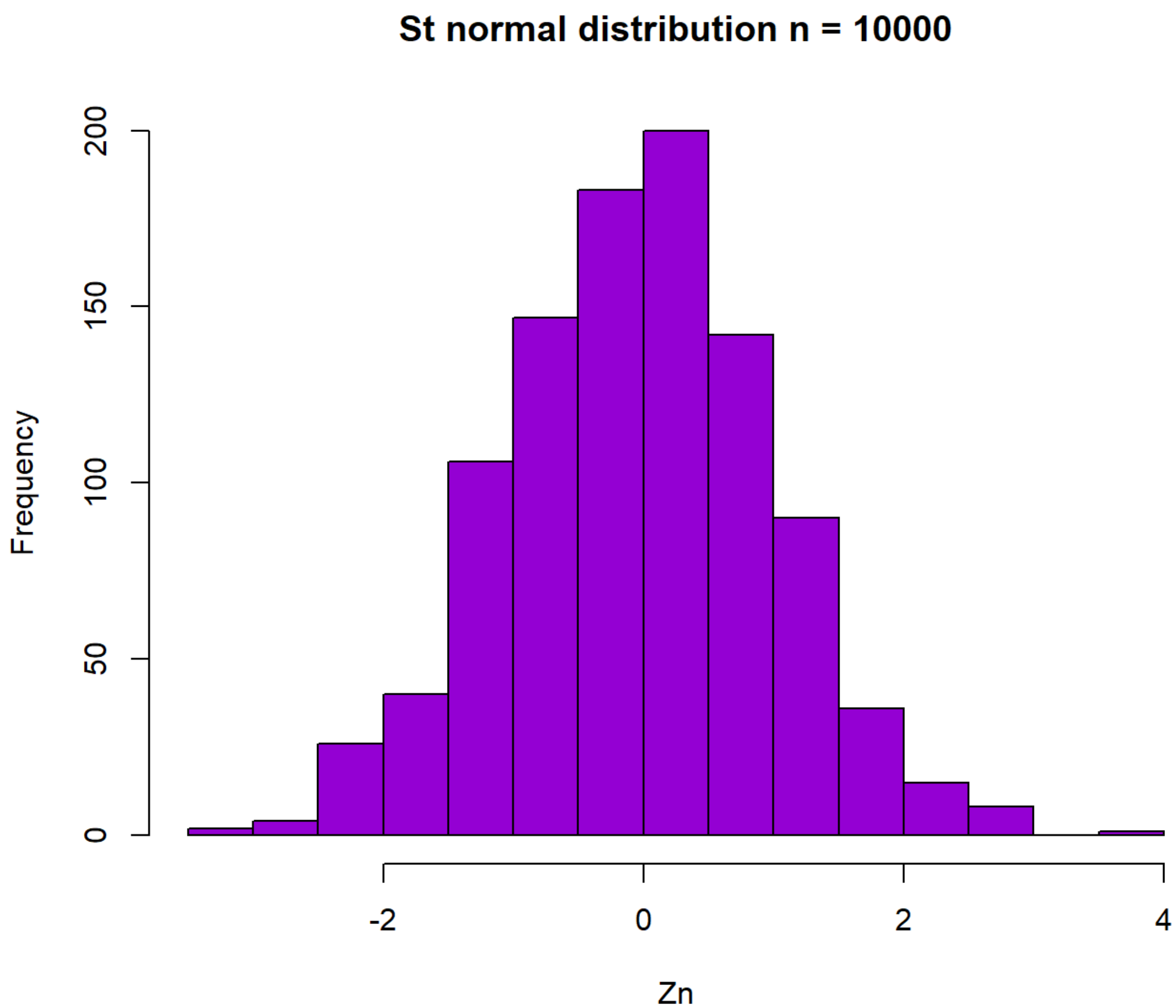

0
-4
-2
0
2
Zn


St normal distribution n = 100
Frequency
0
50
100
150
200
-3
-2
-1
0
1
2
3
Zn

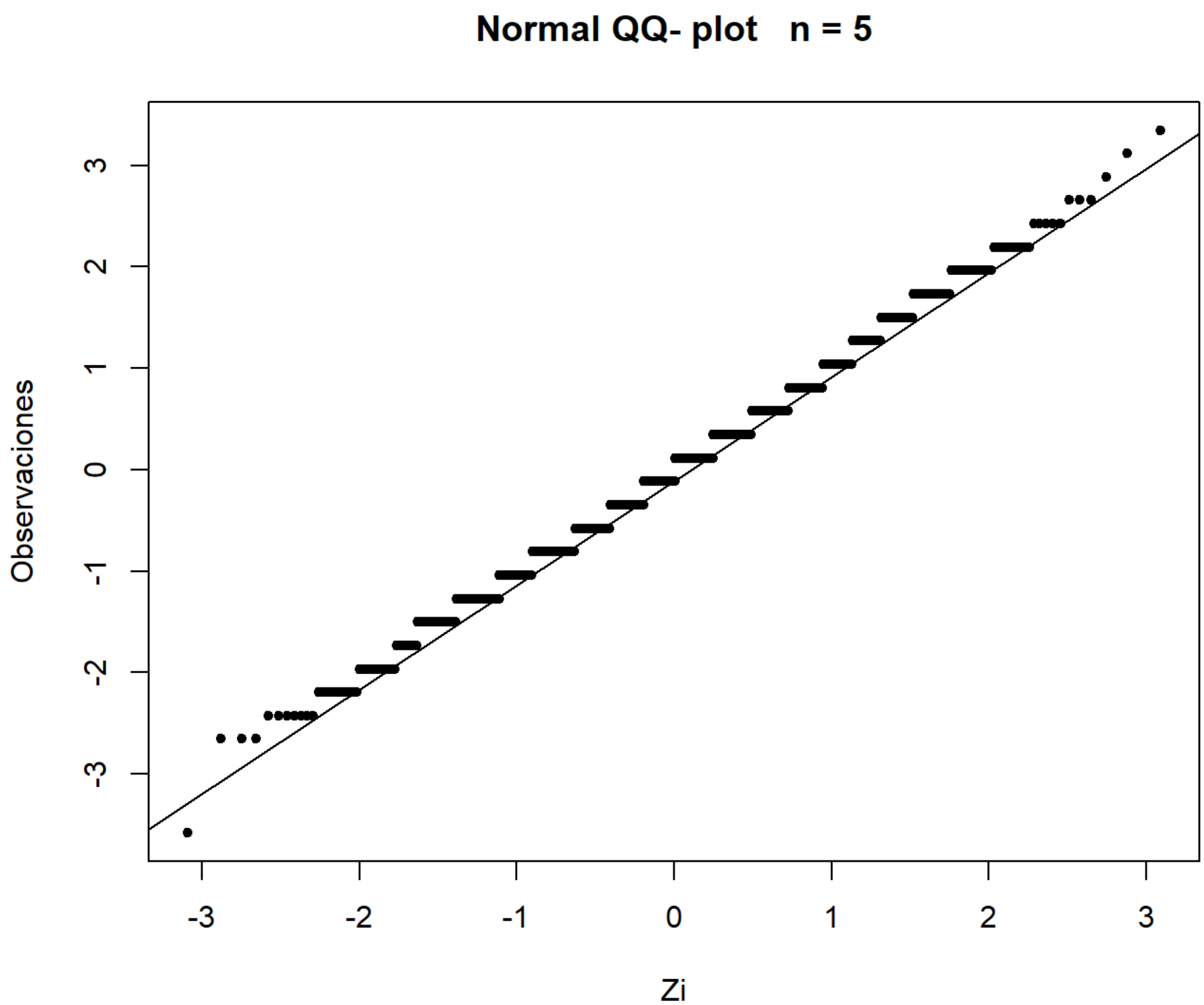

St normal distribution n = 500
Frequency
0
50
100
150
200
-4
-3
-2
-1
0
1
2
3
Zn

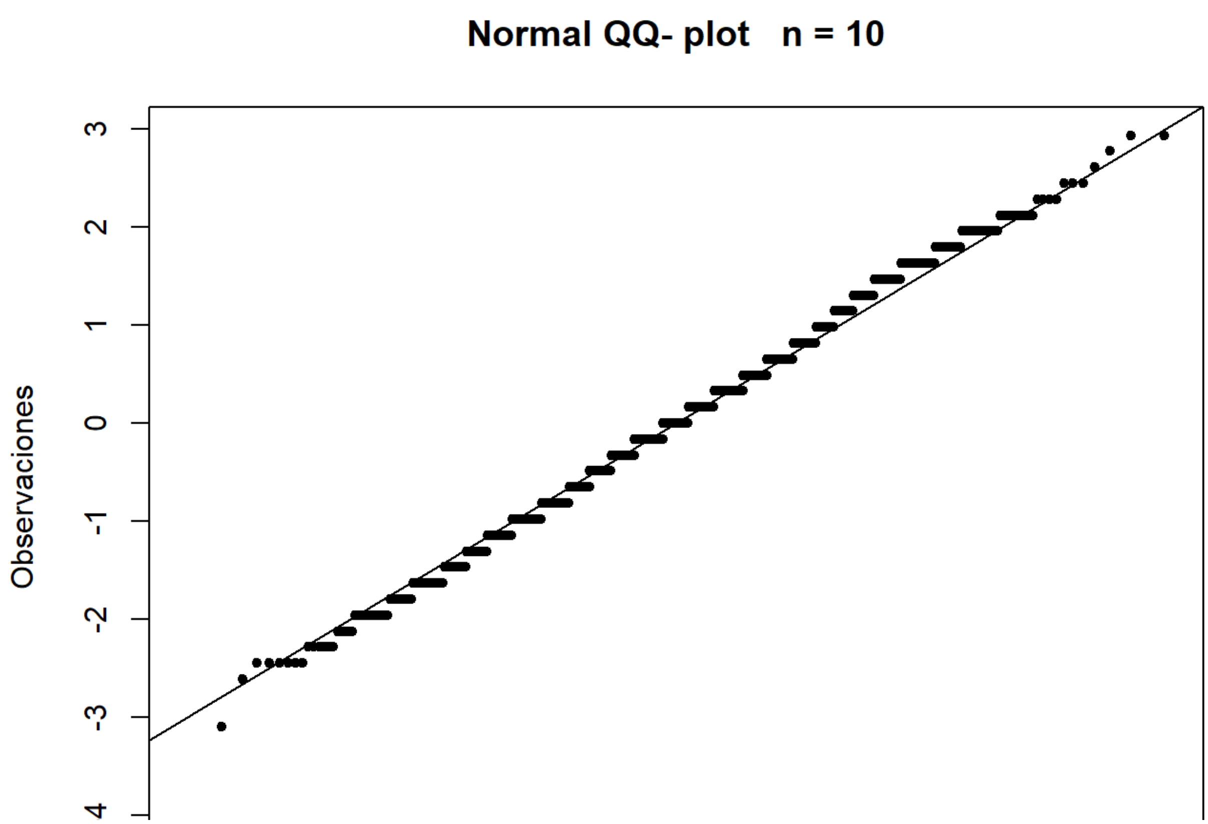


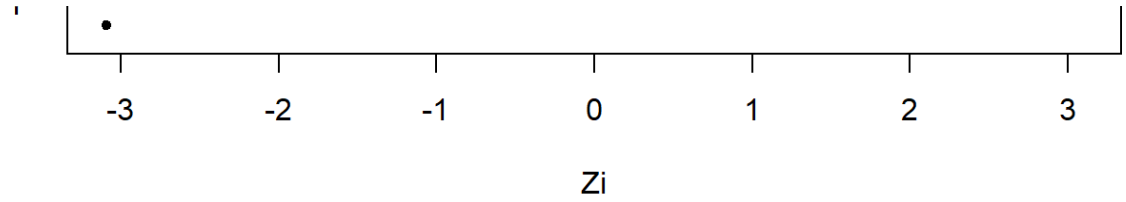
St normal distribution n = 1000
Frequency
0
50
100
150
200
-3
-2
-1
0
1
2
3
Zn

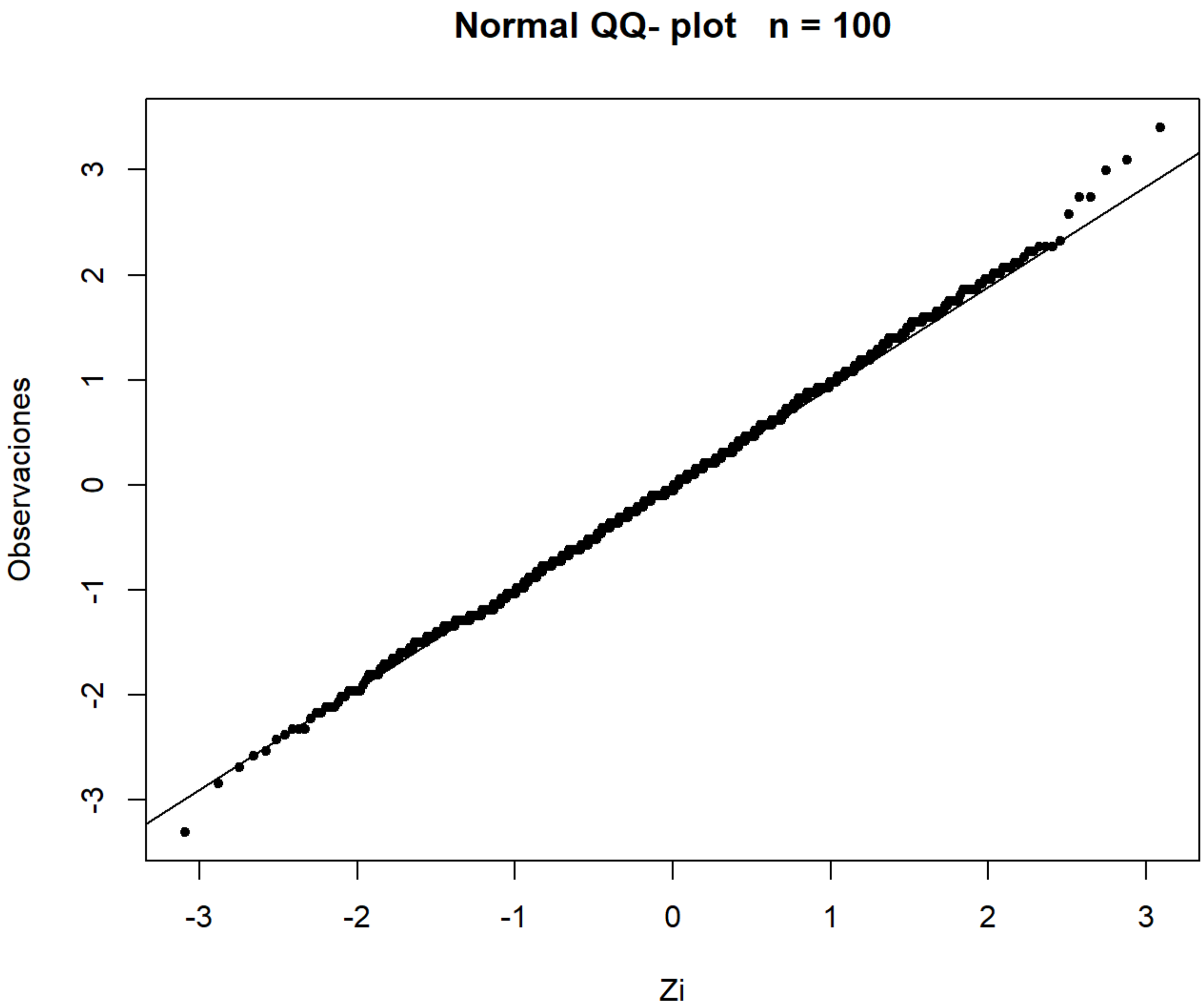

St normal distribution n = 10000
Frequency
0
50
100
150
200
-2
0
2
4
Zn

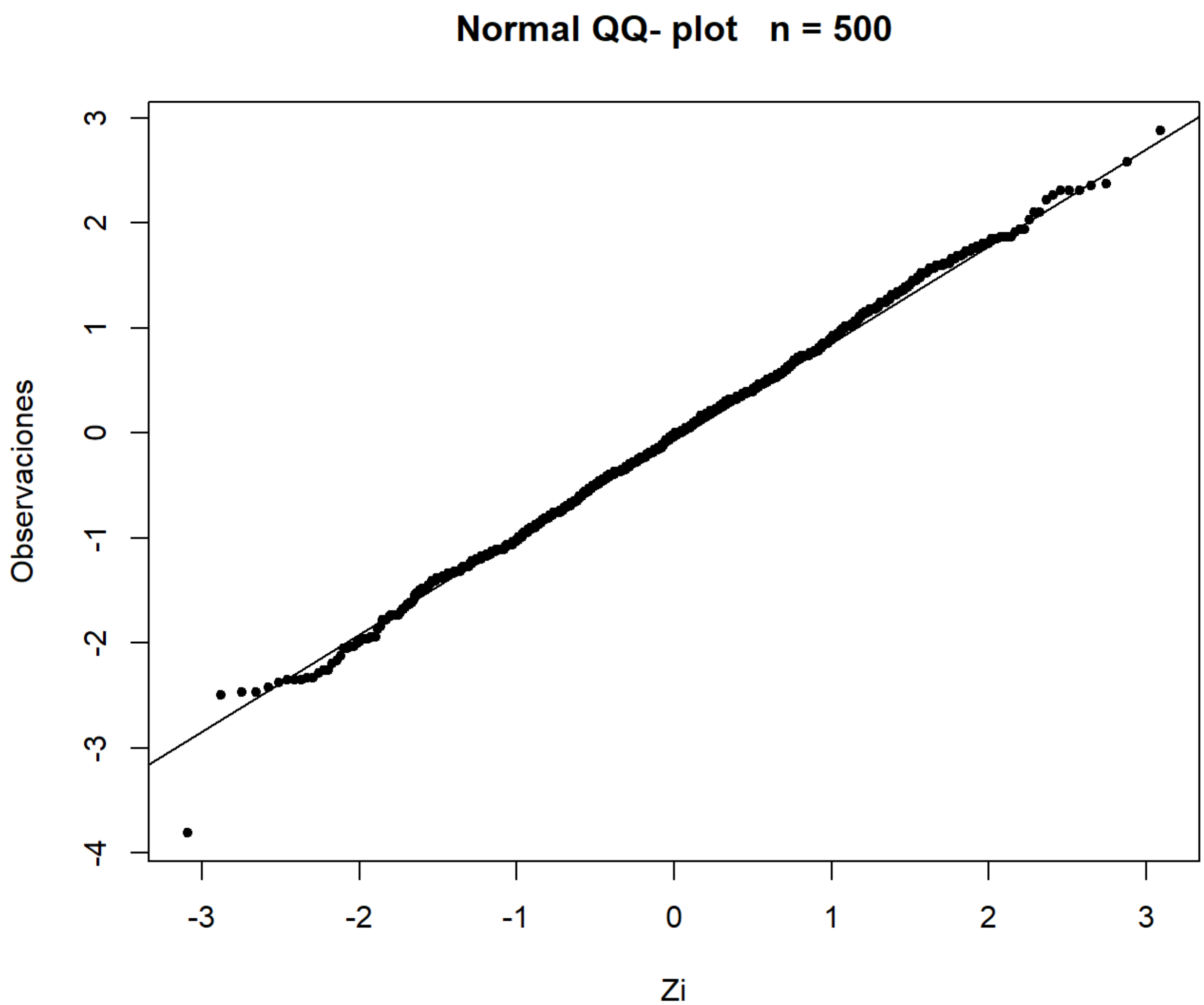
Code

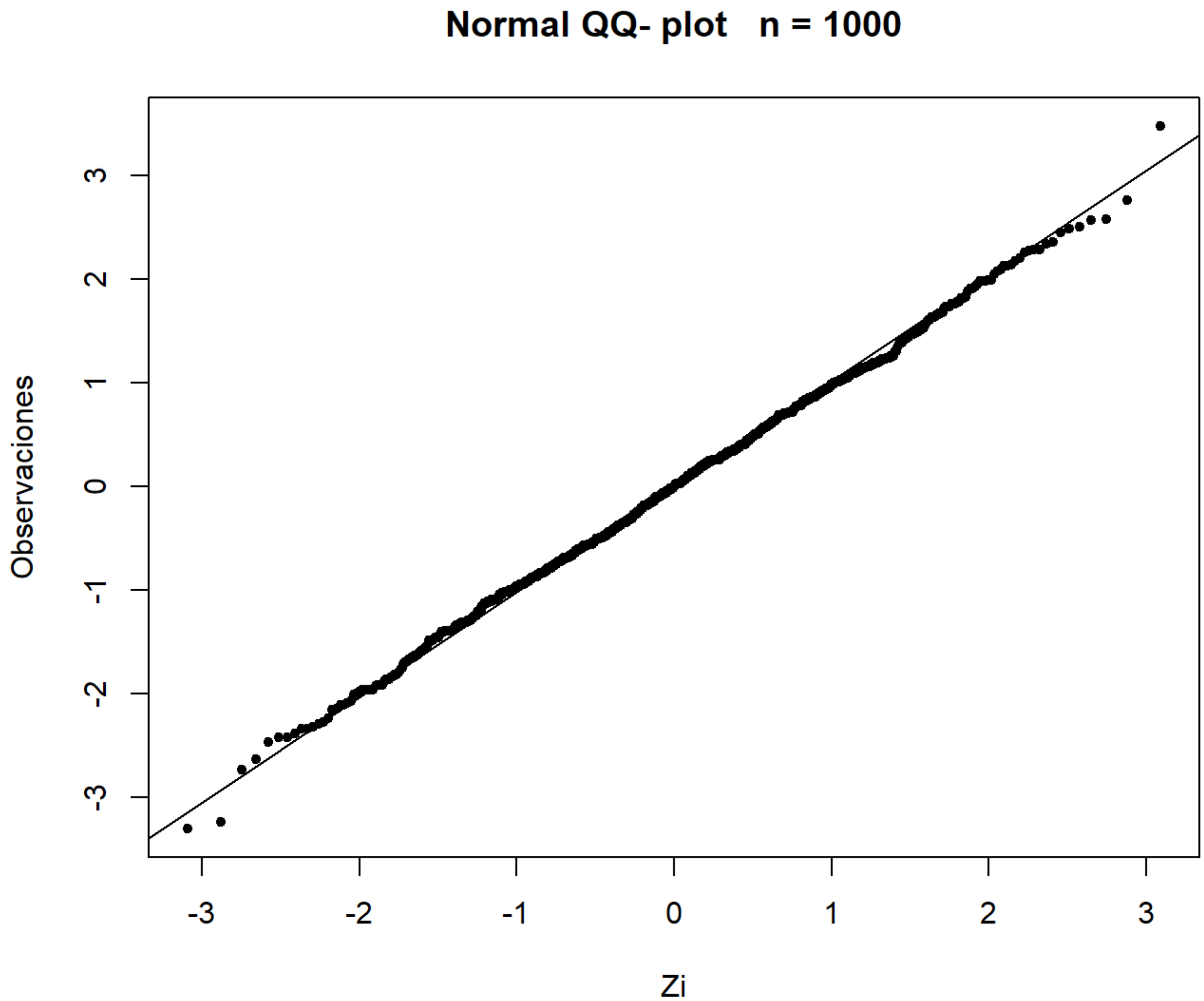

Normal QQ- plot n = 5
Observaciones
3
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3
Zi


Normal QQ- plot n = 10
Observaciones
3
2
1
0
-1
-2
-3
-4


-3
-2
-1
0
1
2
3
Zi

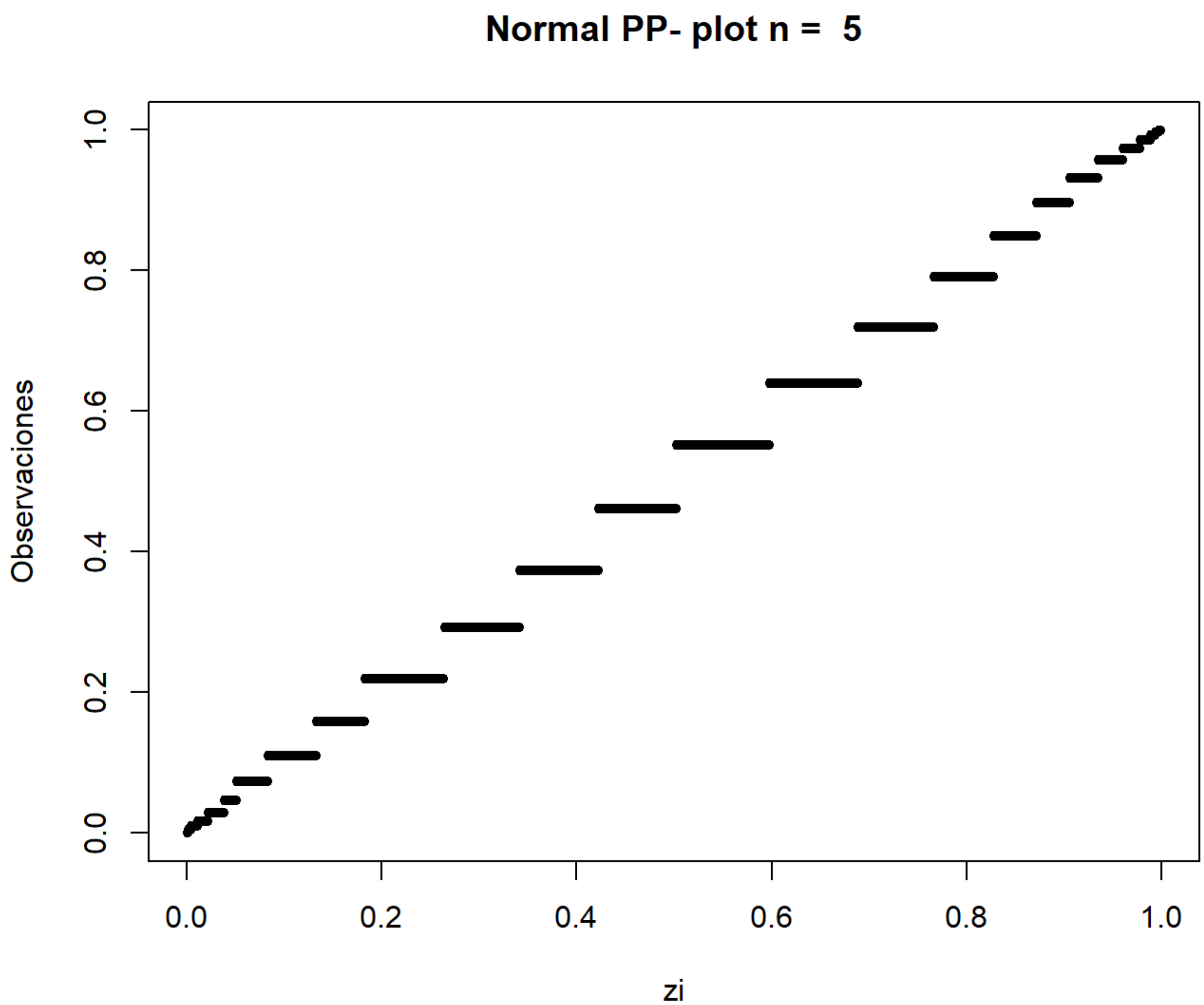

Normal QQ- plot n = 100
Observaciones
3
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3
Zi

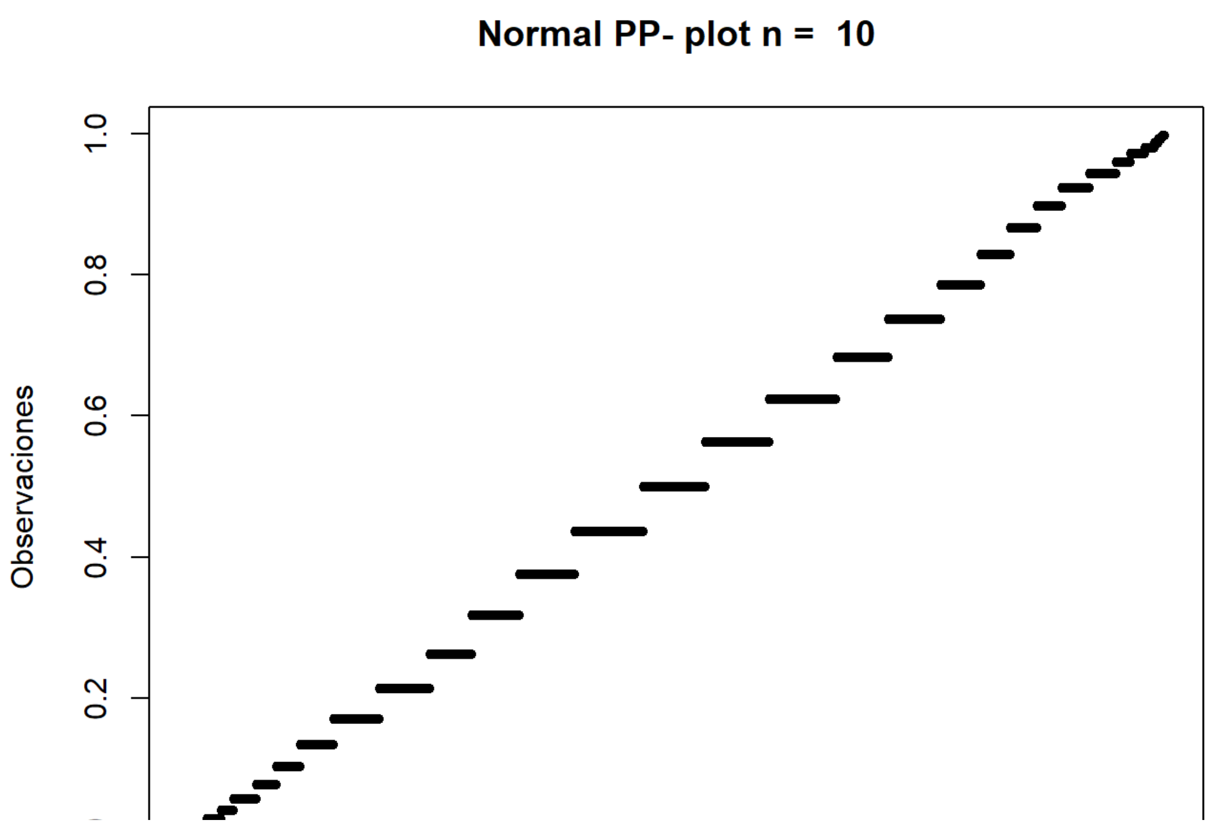

Normal QQ- plot n = 500
Observaciones
Zi
3
2
1
0
-1
-2
-3
-4

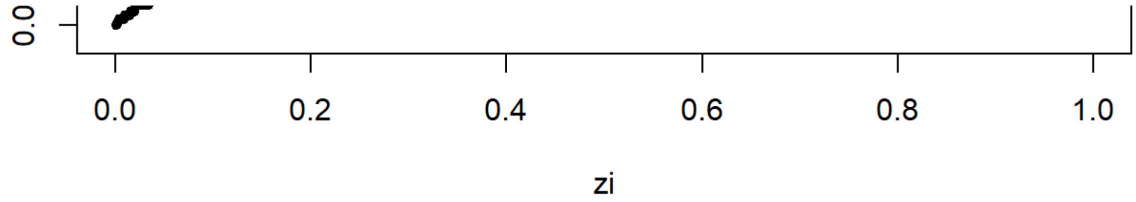

Normal QQ- plot n = 1000
Observaciones
Zi
-3
-2
-1
0
1
2
3

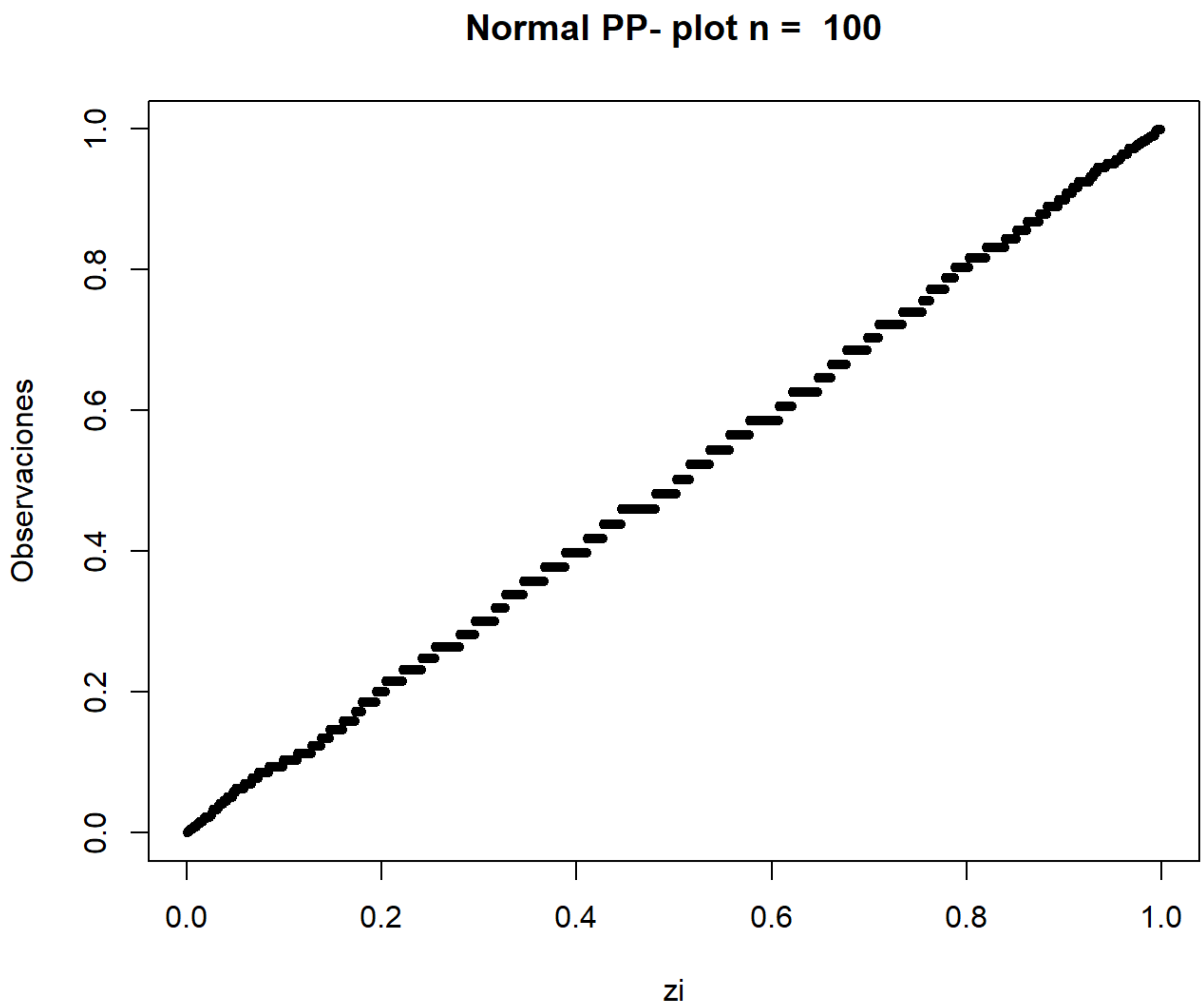
Al realizar el QQ-plot de la distribución binomial, se observa que con un tamaño de muestra pequeño (n = 5), la distribución estimada se aproxima al centro de la teoríca. Una vez incrementando el número de muestra, se puede ver que el efecto de escalonamiento - debido a que es una variable aleatoria discreta- pasa a haacer algo más apriximado a lo continuo, dejando más apreciable como el teoréma del límite central se cumple, haciendo que la sucesión de binomiales al estandarizarlas y tomar límite llevando la muesra al infinito converge en distribución a una normal estandar.

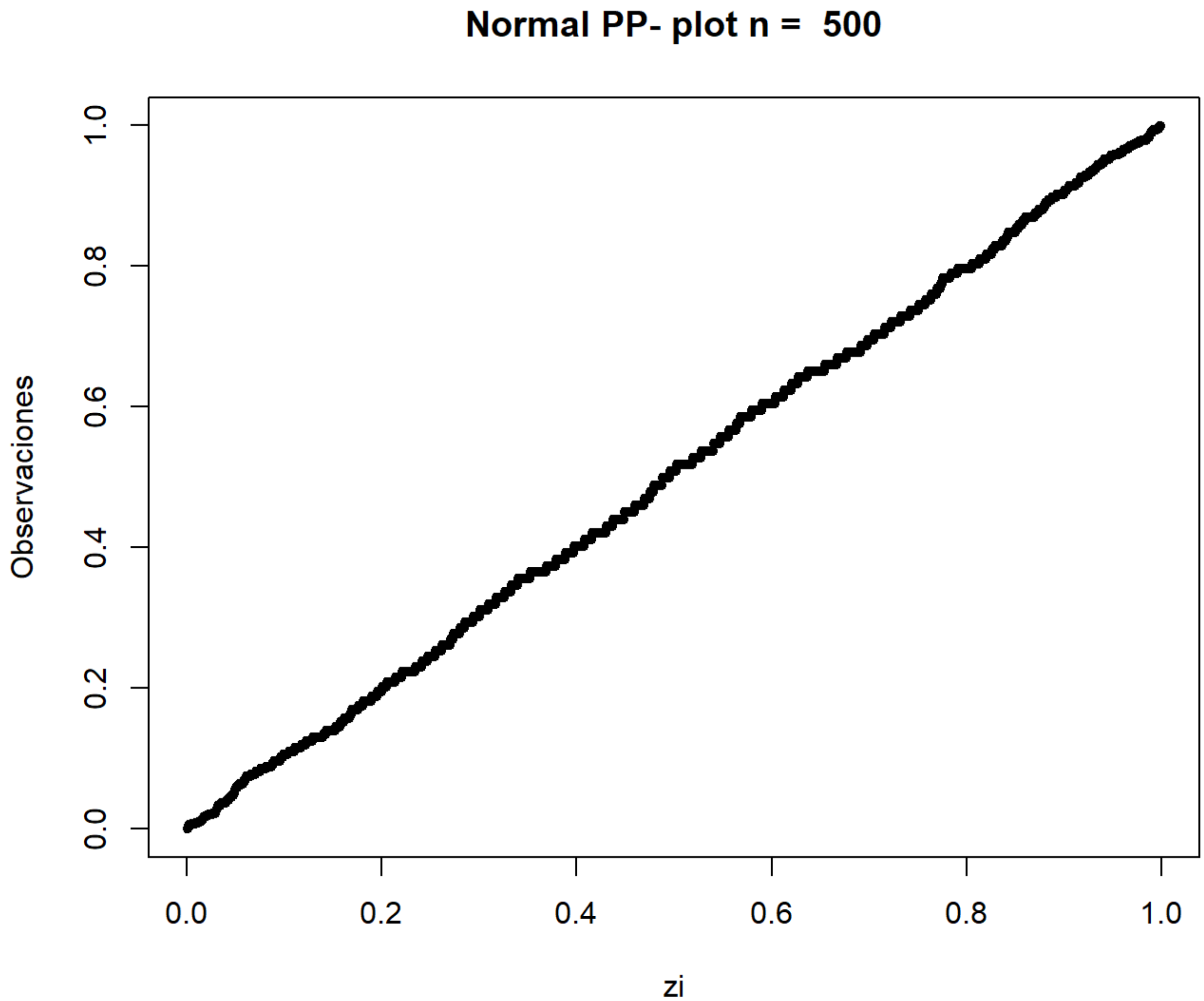
Code

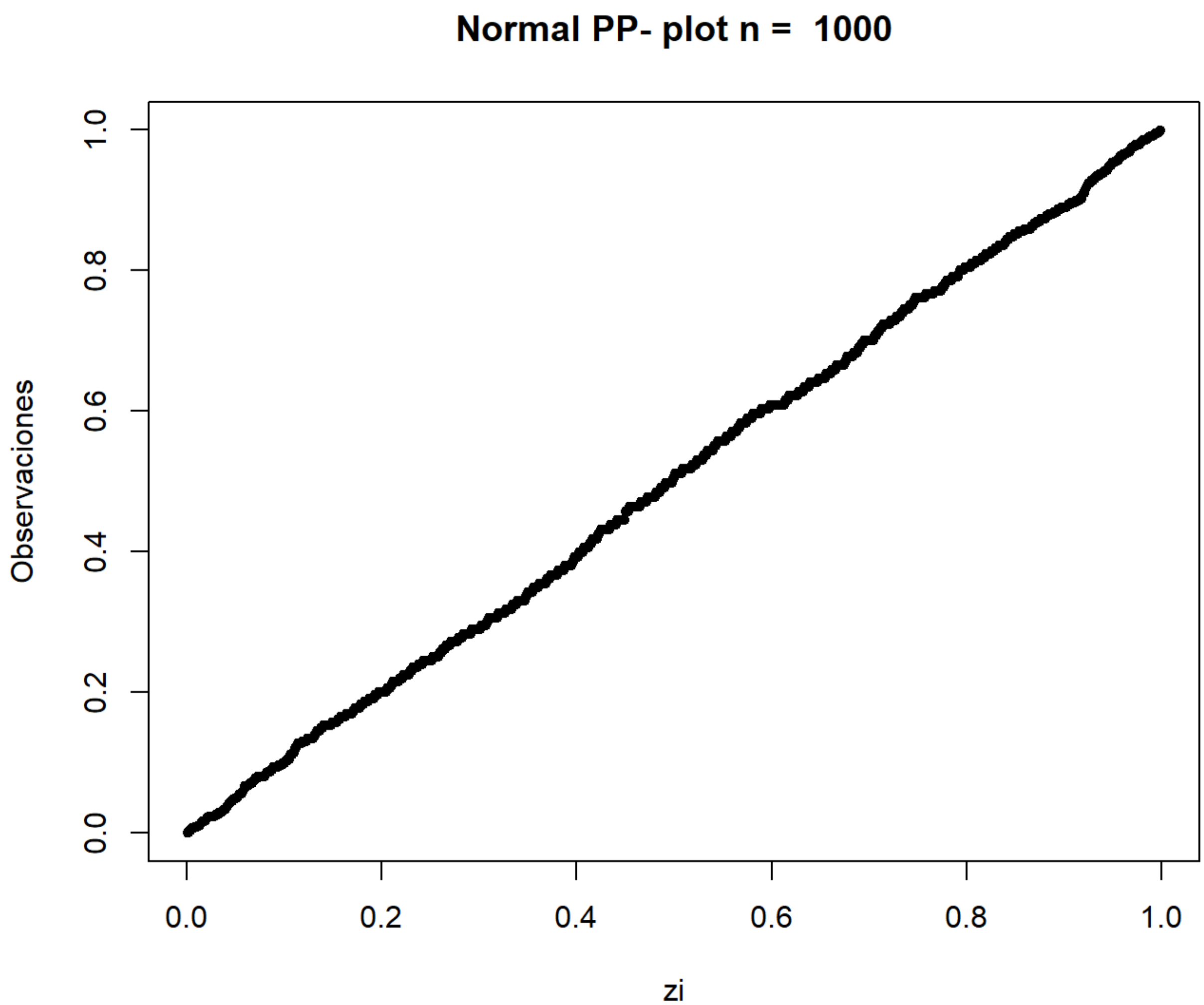

Normal PP- plot n = 5
Observaciones
zi
0.0
0.2
0.4
0.6
0.8
1.0

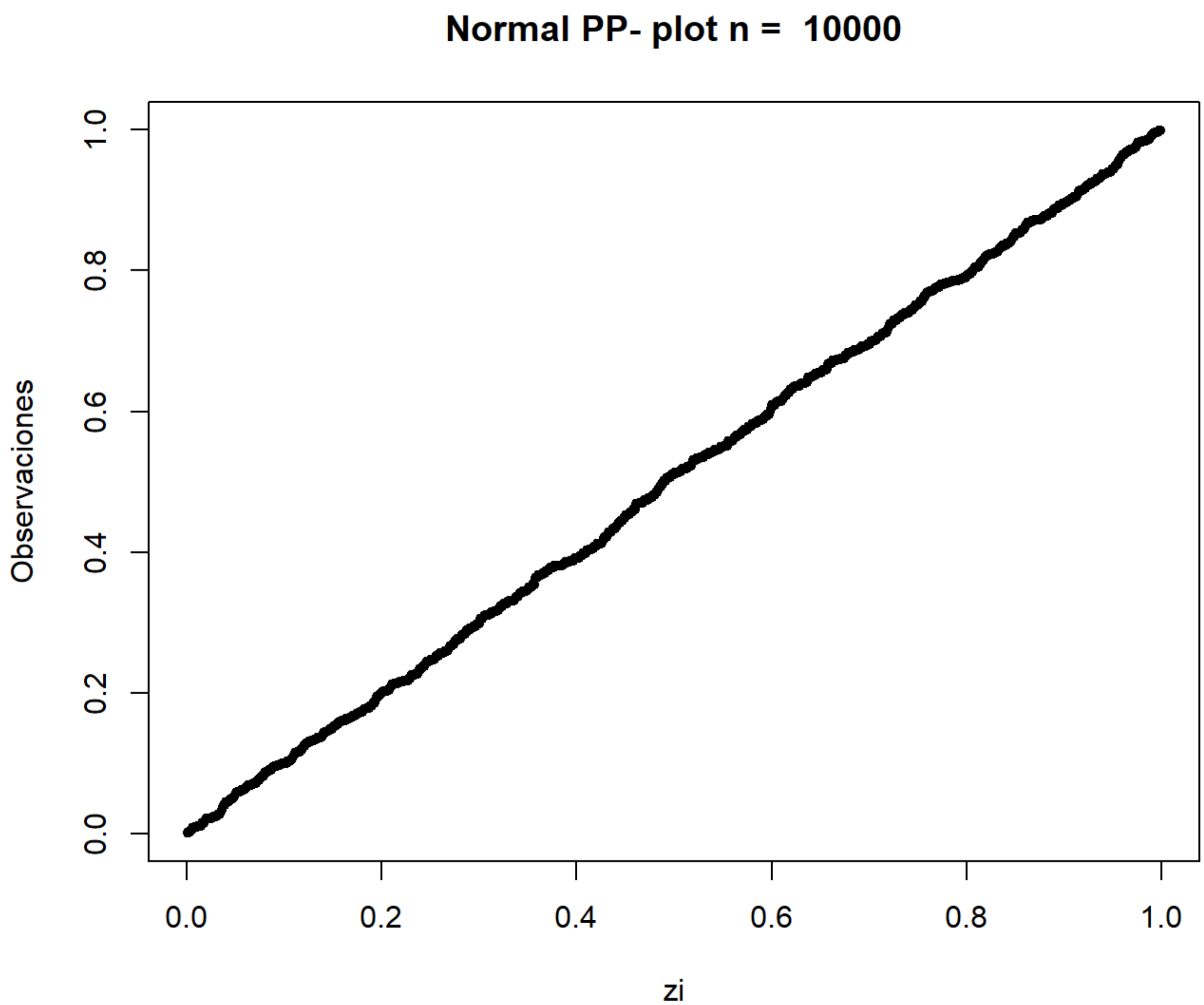

Normal PP- plot n = 10
Observaciones
0.2
0.4
0.6
0.8
1.0


0.0
0.0
0.2
0.4
0.6
0.8
1.0
zi


Normal PP- plot n = 100
Observaciones
0.0
0.2
0.4
0.6
0.8
1.0
zi


Normal PP- plot n = 500
Observaciones
zi
0.0
0.2
0.4
0.6
0.8
1.0


Normal PP- plot n = 1000
Observaciones
zi
0.0
0.2
0.4
0.6
0.8
1.0

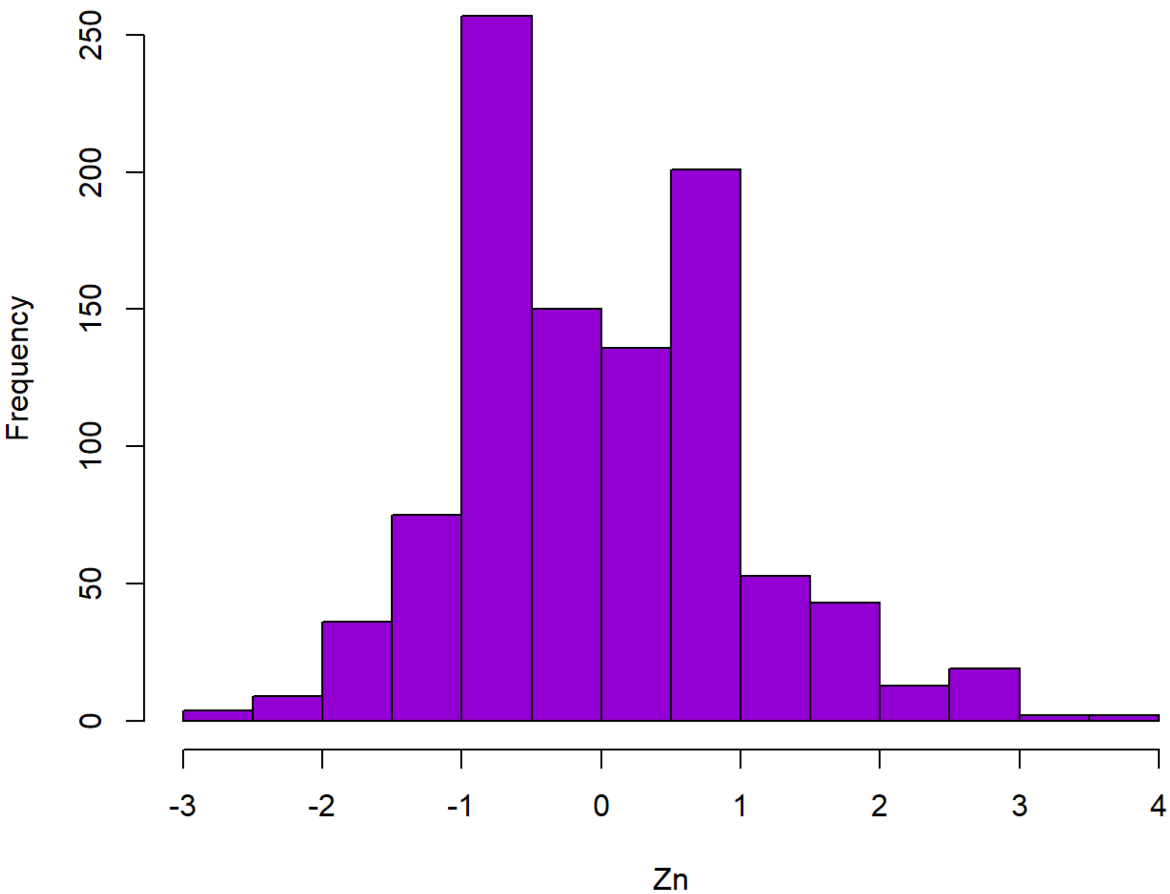


En el caso del PP- plot, presenta un comportamiento similar a lo presentado con el gráfico de QQ-plot. Estó es, que bajo muestras pequeñas se observa que el contraste de las probabilidades empiricas y teorícas rondan sobre la normal, peros es hasta que el número de muestra incrementa cuando se aprecia lo anterior mencionado.

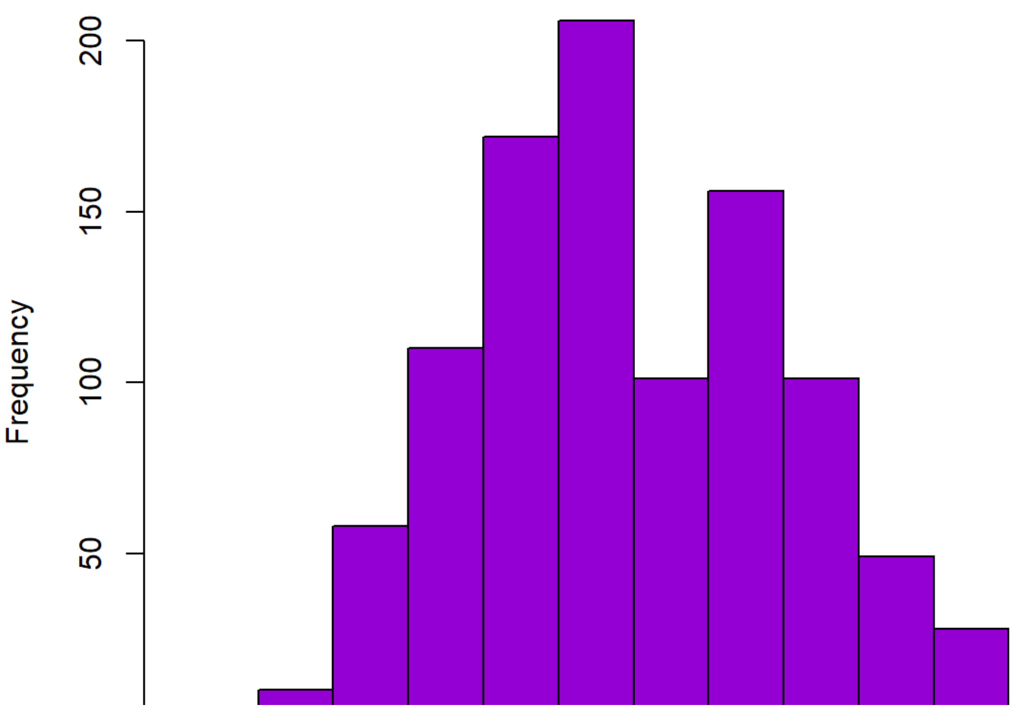
c. Para p = 0.1, N = 15, n = 5; 10; 20; 100 y m = 1000, genere muestras de Zn y grafíque estás muestras en un histograma (un histograma para cada n). ¿Qué observa? Explíque.

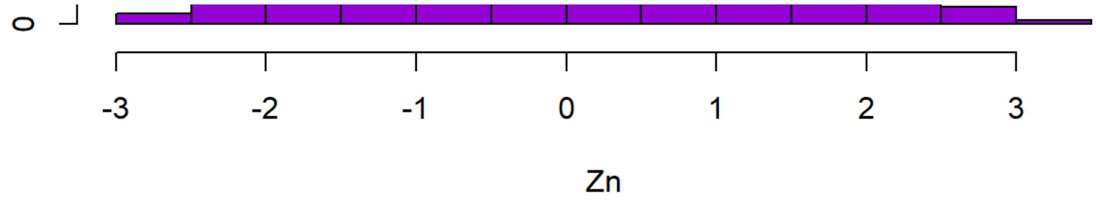
Code

St normal distribution n = 5

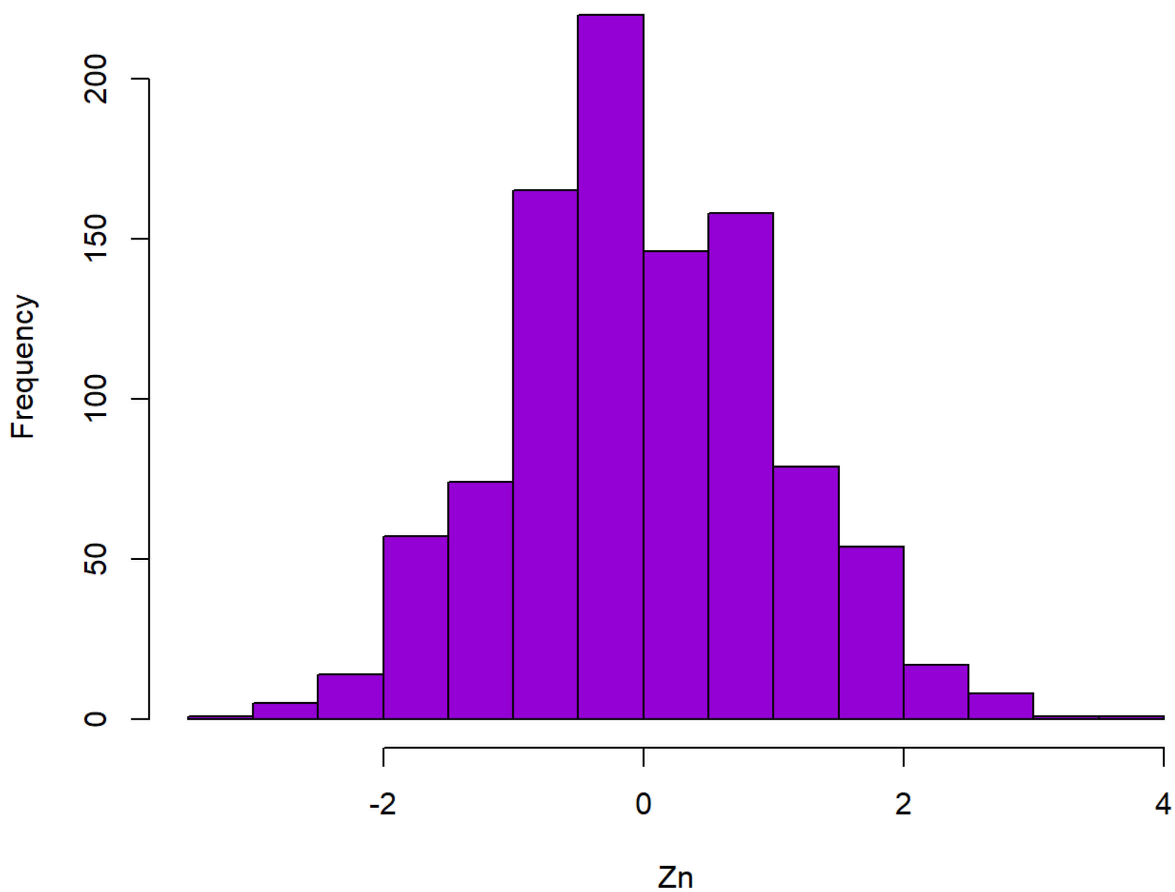


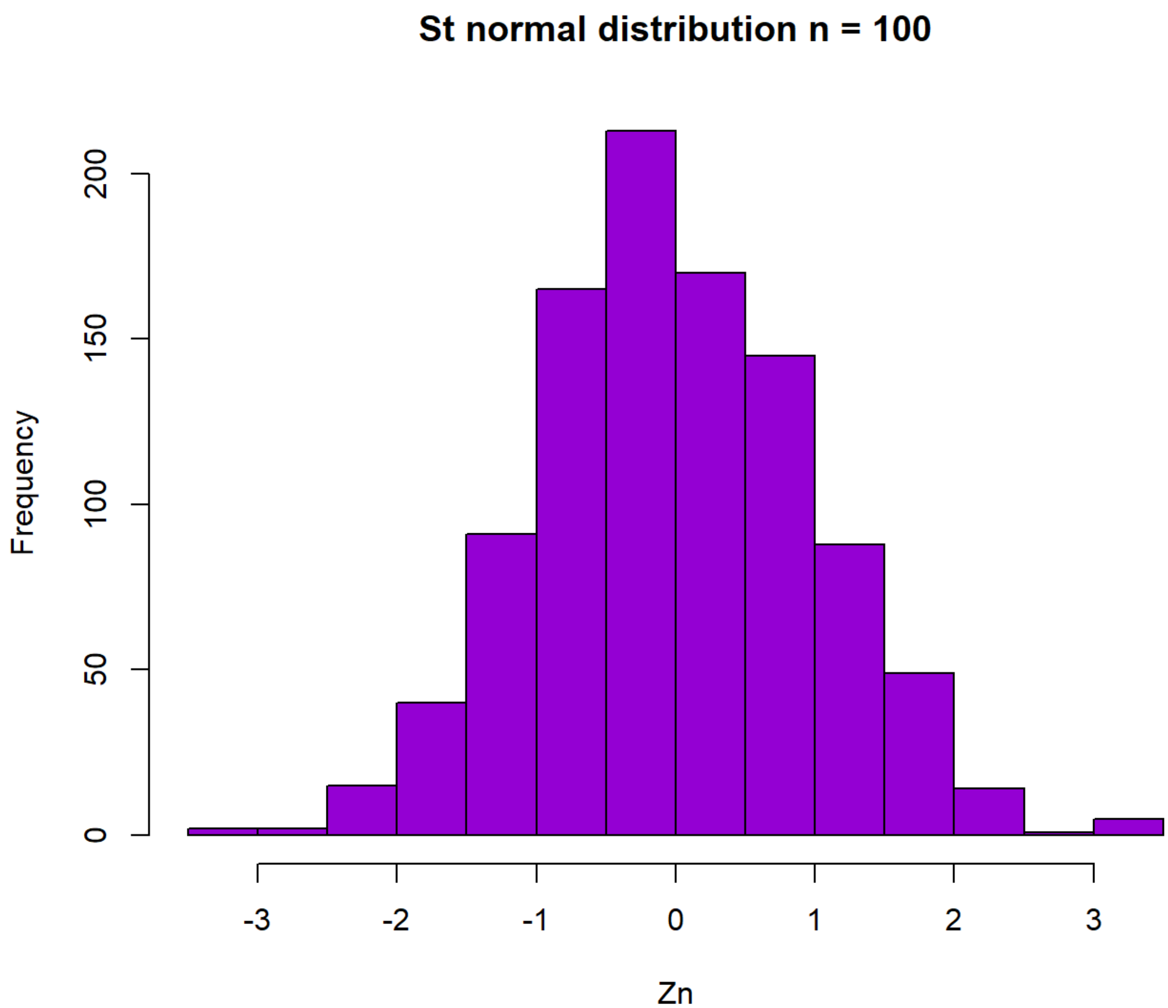
St normal distribution n = 10



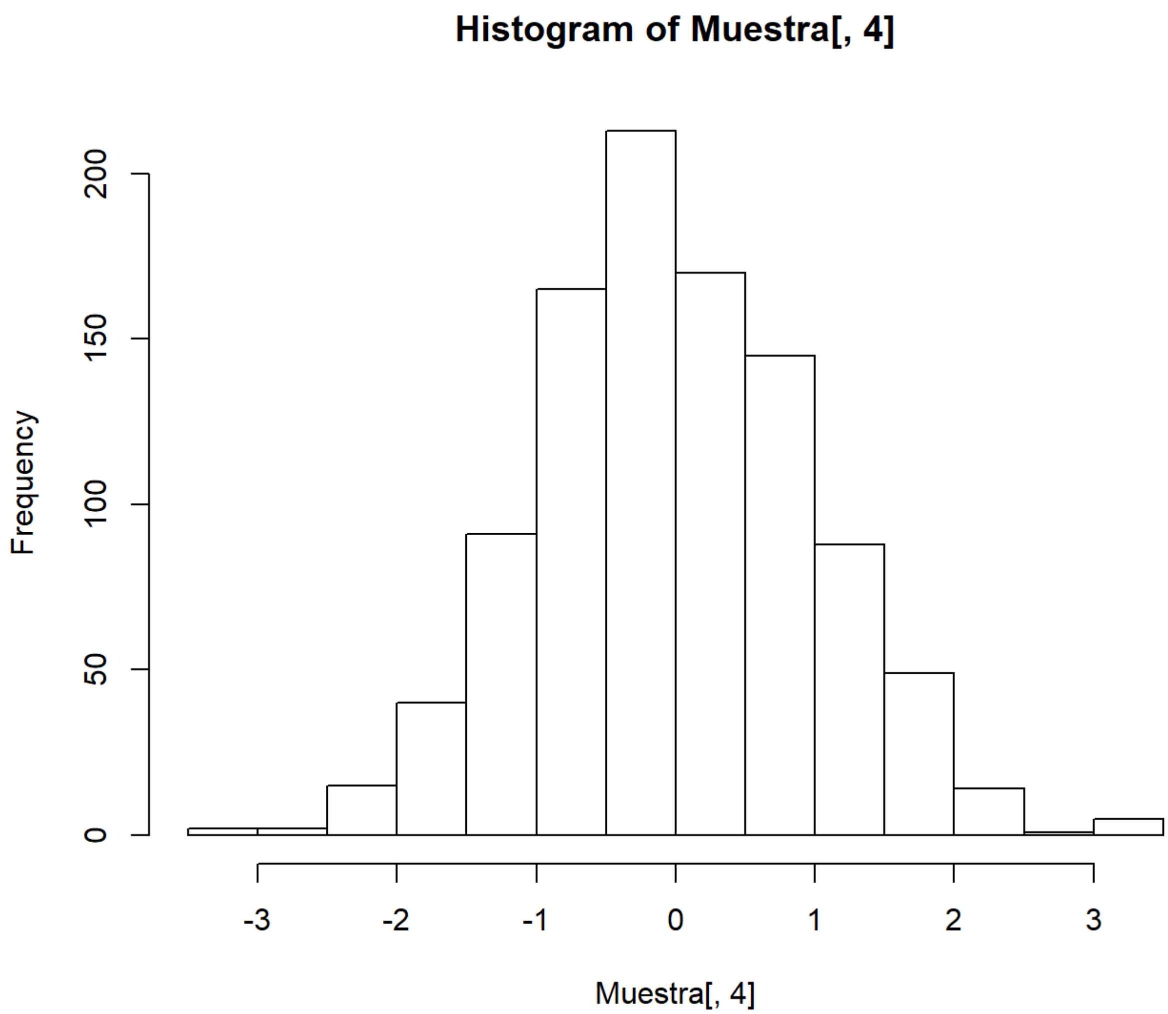


St normal distribution n = 20




St normal distribution n = 100
Frequency
0
50
100
150
200
-3
-2
-1
0
1
2
3
Zn

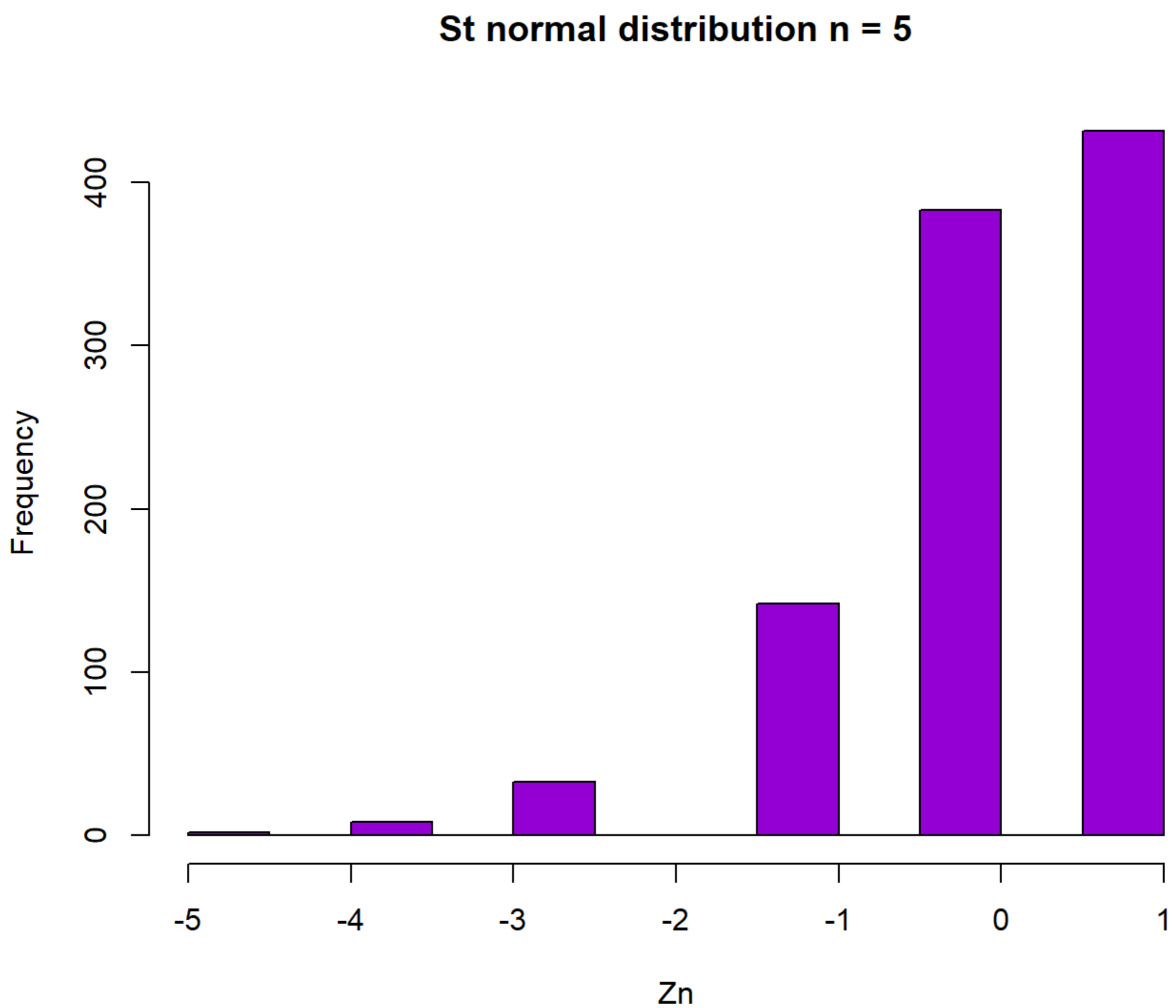
Code

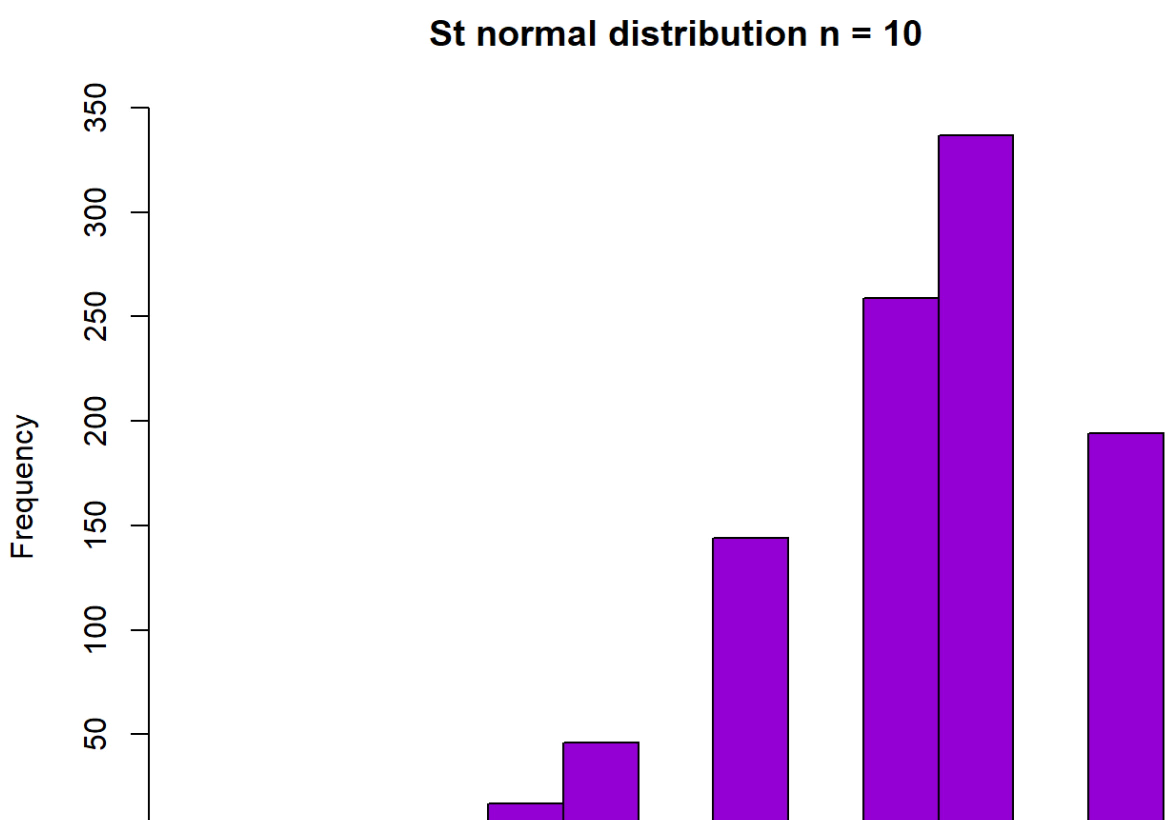


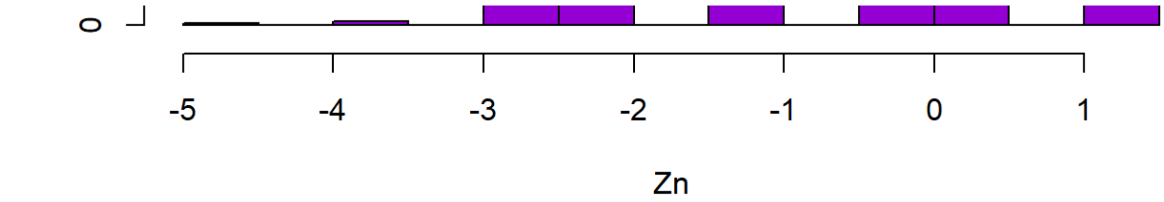
¿Qué observa? Explique Cuando el número de mmuestra es de 5 y diez observaciones, la función de densidad se aproxima a la de una distribución normal estandar, más sin embargo sigue sin ser la función de densidad de una normal estandar. Esto debido a que sigue presentando inconsistencias (i.e quiebres) en el centro de la distribución. Se aproxima cada vez más en las colas, pero no en el centro. La aproximación en el centro de la suceción estandarizada de binomiales mejora conforme el tamaño de la muestra crece. Para un tamaño de muestra de cien, se puede apreciar como la distribución presentada se aproxima en tanto colas y centro de densidad de una distribución a una normal estandar. En conclusión, bajo las condiciones de que la suceción de binomiales tiene momentos finitos, y dado a que la esperanza es finita existe una generadora de momentos, entonces, aplicando límite central, se observa como la media de la sucesión de variables aleatorias estandarizada converge en distribucion a una distribución límite normal estandar

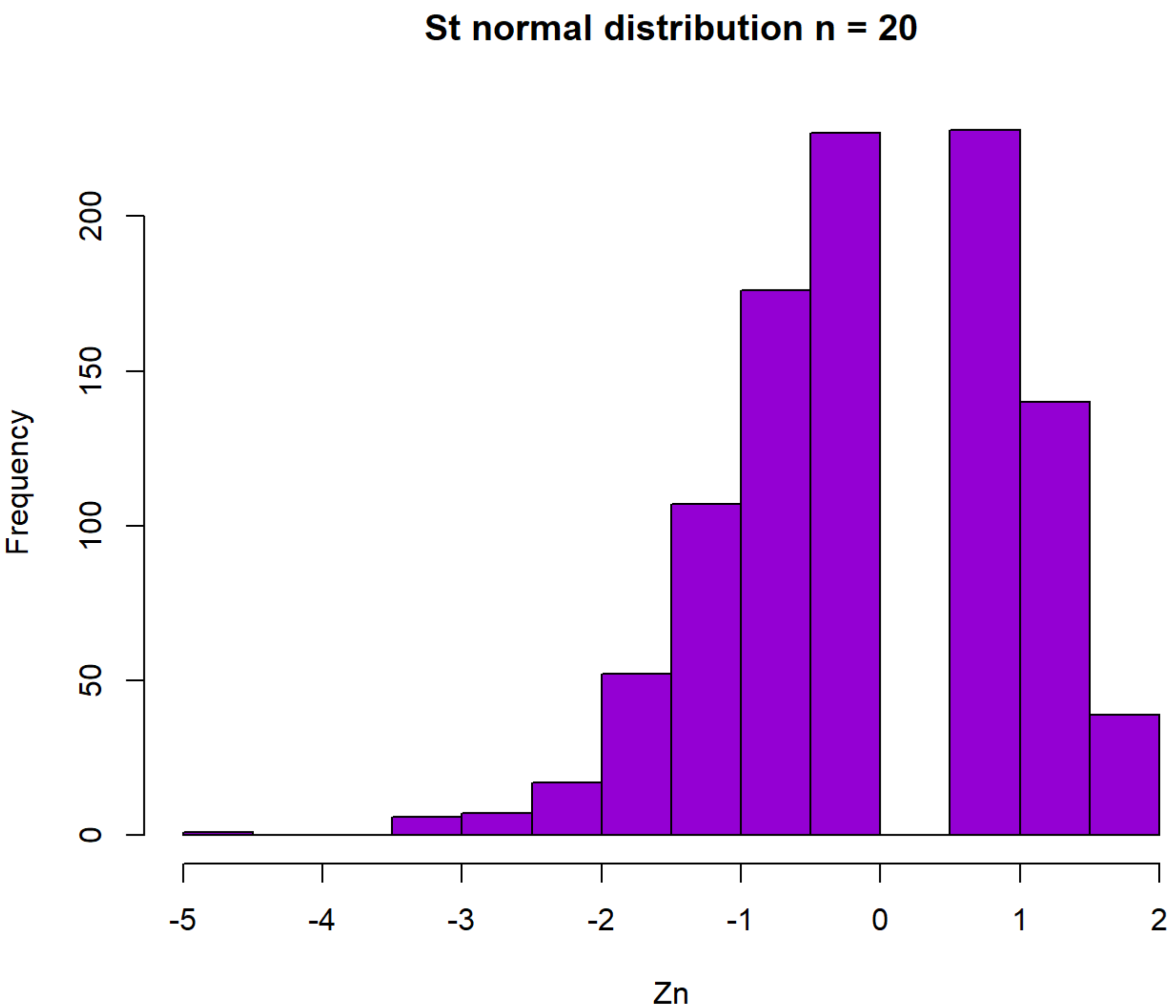
d. Repita el inciso anterior para p = 0:99. Compare su resultado con lo obtenido en el inciso anterior.

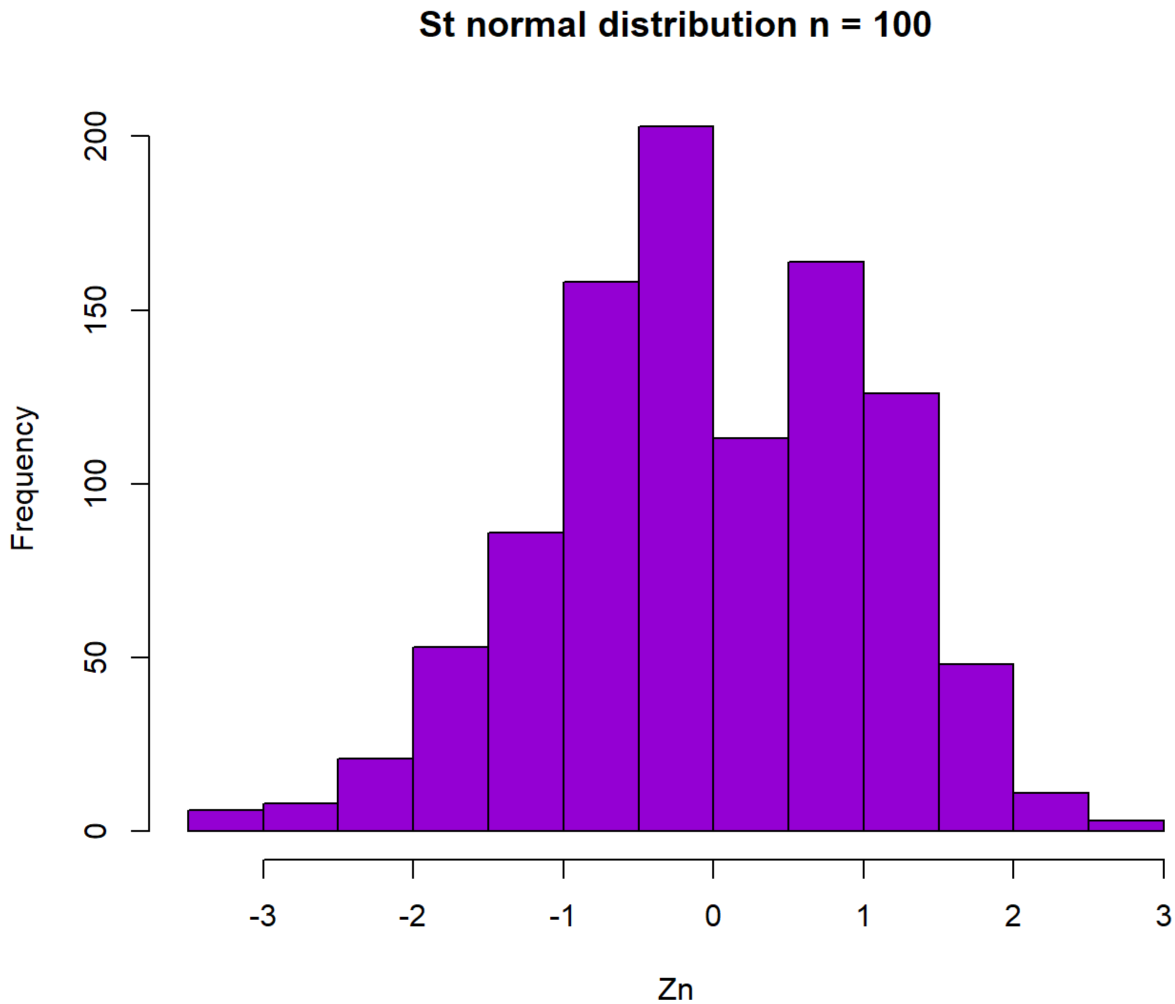
Code


St normal distribution n = 5
Frequency
0
100
200
300
400
-5
-4
-3
-2
-1
0
1
Zn


St normal distribution n = 10
Frequency
50
100
150
200
250
300
350


0
-5
-4
-3
-2
-1
0
1
Zn


St normal distribution n = 20
Frequency
0
50
100
150
200
-5
-4
-3
-2
-1
0
1
2
Zn



Con una probabilidad de 0.99, de una distribución binomial, y con un tamaño de muestra pequeño (i.e de 5 a 20 observaciones), su función de probabilidad presenta sesgo a la izquierda (derecha para el quien lo observa), es decir, la forma de los datos se cargan hacia la izquierda (a la derecha de quien lo ve) entendiendo que la función de probabilidad no es simpetrica (asímetrica). No obstante, cuando el tamaño de la muestra incrementa se puede observar como la distribución tiene a reducir dicho sesgo. Por lo tanto, bajo el teorema del límite central, bajo ciertas condiciones al incrementar el tamaño de muestra y estandarizar los eventos binomiales, dicha distribución converge a la función de densidad de una distribución normal estandar. El sesgo desaparece en su grán mayoria, pero aún las observaciones del centro no logran aproxximar del todo bien a la normal estandar.

EJERCICIO 4

b. Simule una sucesión de n = 1000 v.a. como arriba y calcule S_{1000} para p = 0.4. Repita este proceso 100 veces y grafíque la distribución empirica de S_{1000} que se obtiene de la simulacion y empalmela con la distribución asintótica teórica que obtuvo. Comente sus resultados.

Code

```
## Rlab 2.15.1 attached.
```

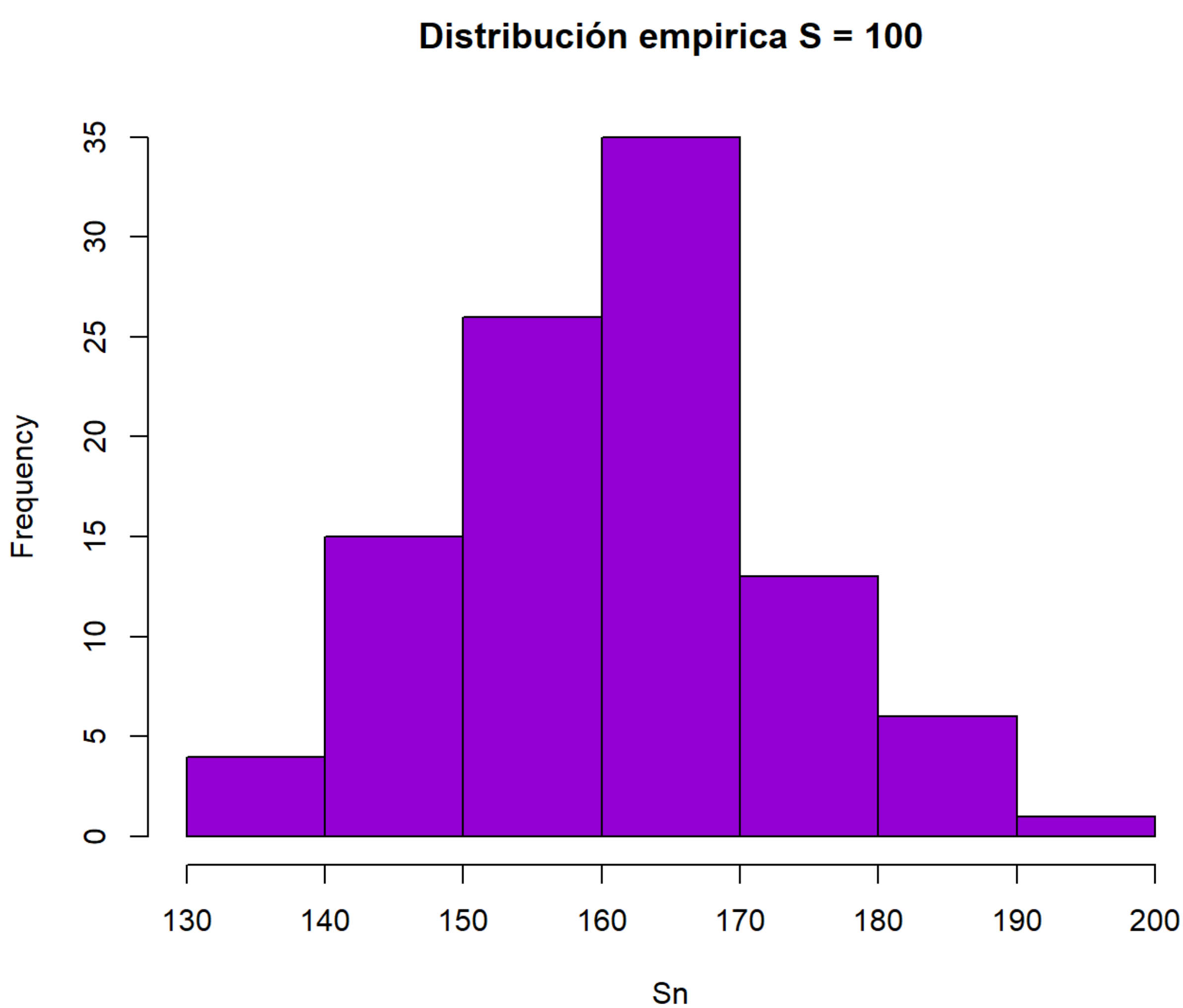
```
##
## Attaching package: 'Rlab'
```

```
## The following objects are masked from 'package:stats':
##
##     dexp, dgamma, dweibull, pexp, pgamma, pweibull, qexp, qgamma,
##     qweibull, rexp, rgamma, rweibull
```

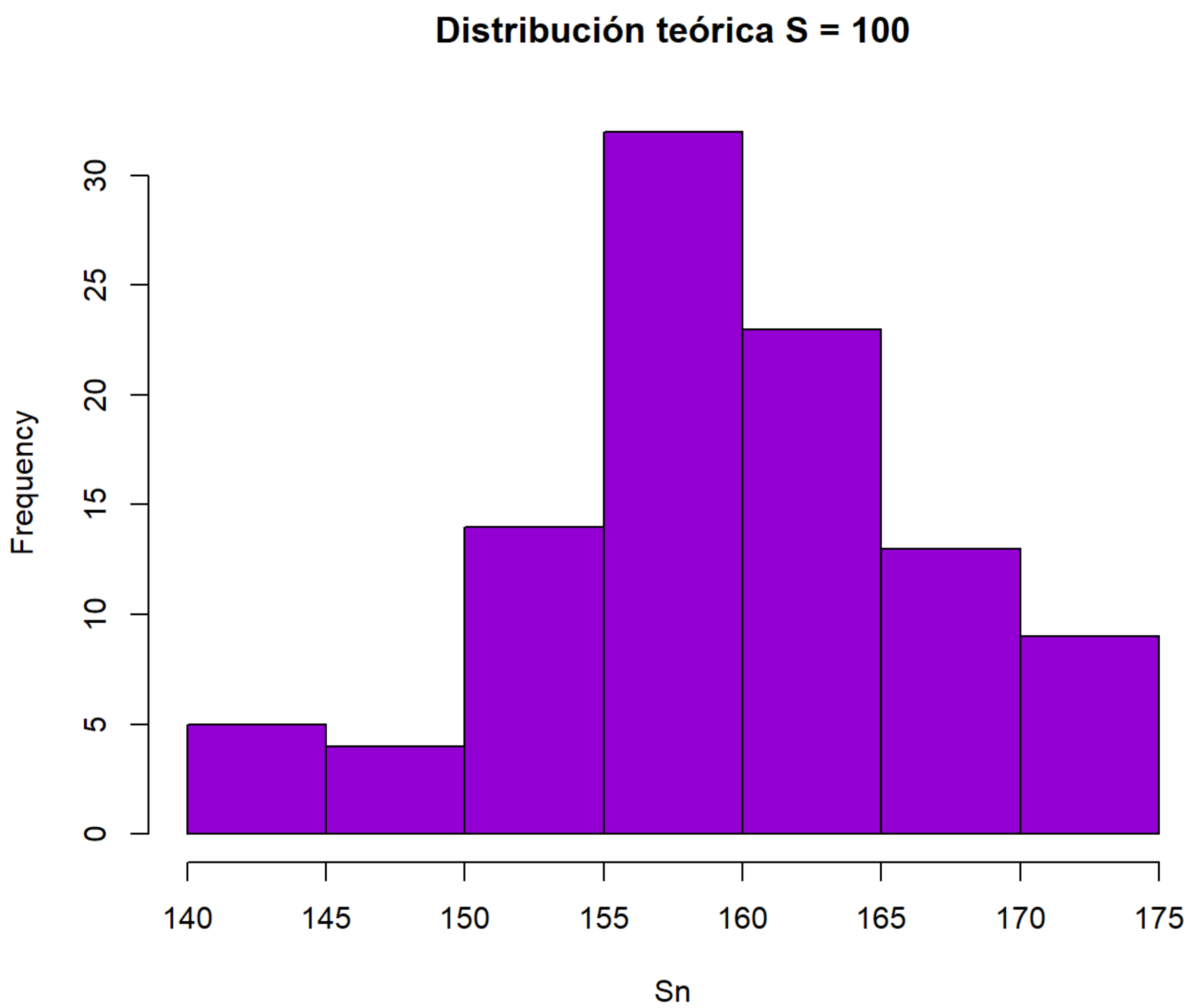
```
## The following object is masked from 'package:datasets':
##
##     precip
```

Code

Code

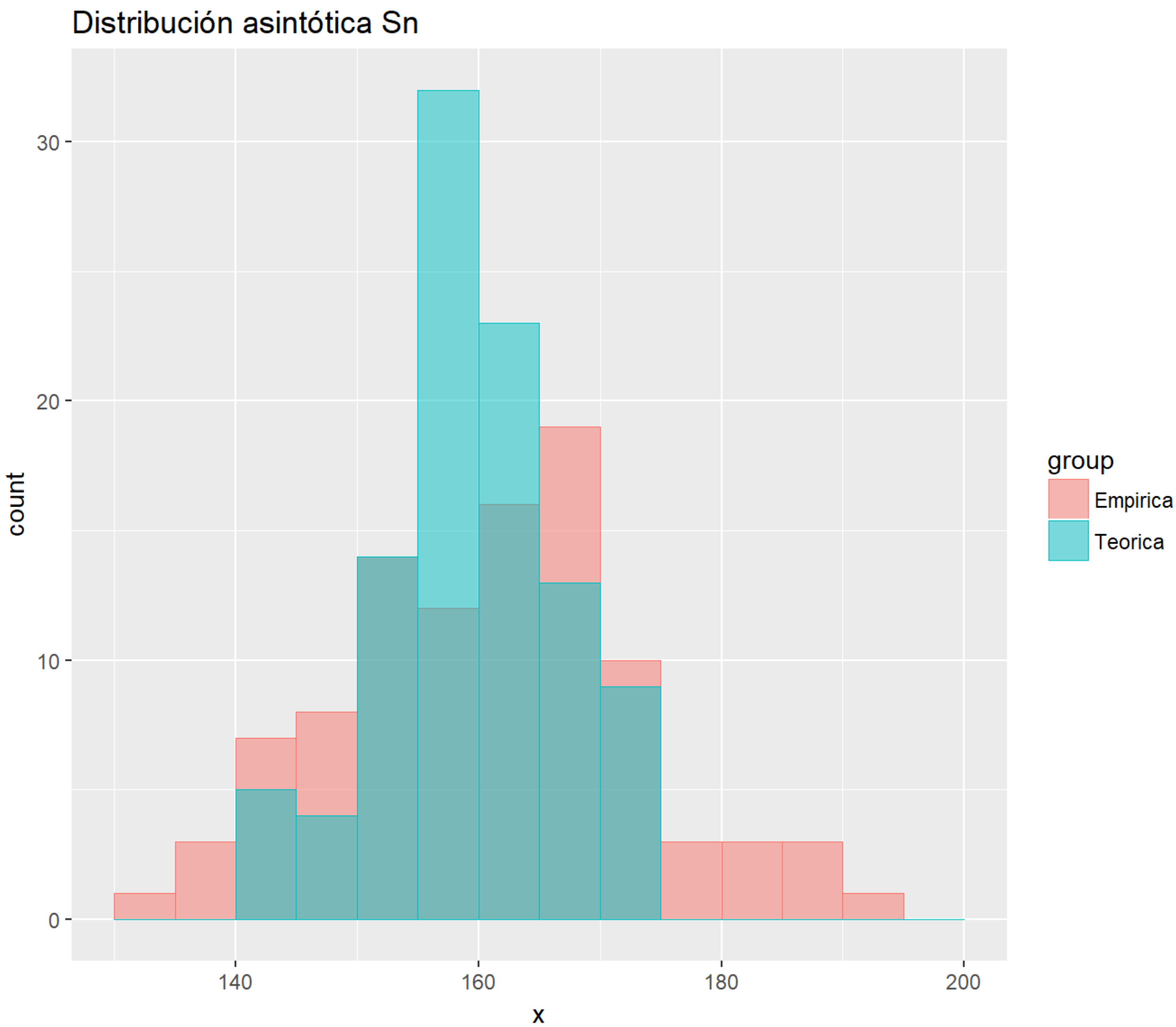


Code



Distribución teórica S = 100
Frequency
0
5
10
15
20
25
30
140
145
150
155
160
165
170
175
Sn

Code


Distribución asintótica Sn
count
30
20
10
0
140
160
180
200
x
group
Empirica
Teorica