Trabajo Práctico de Probabilidad y Estadística: Ley de los Grandes Números y el Teorema Central del Límite

Buceta Diego, Springhart Gonzalo, Tasat Dylan

November 19, 2017

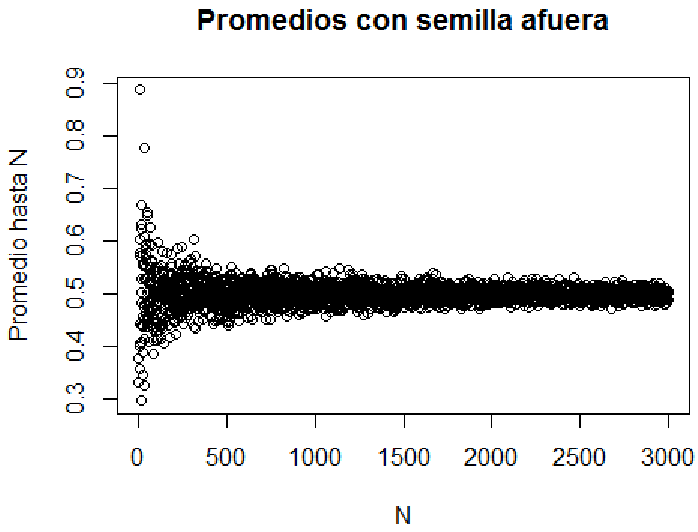
1 Preámbulo

La Ley de los Grandes Números y el Teorema Central del Límite son temas fundamentales de la probabilidad y la estadística, sin embargo entender ambos no es trivial. Mediante este trabajo buscamos entender los conceptos de la Ley de los Grandes Números y el Teorema Central del Límite mediante ejercicios que ponen en evidencia el cumplimiento de ambos.

2 Primer Ejercicio

En este ejercicio, trabajamos con una variables aleatorias exponenciales todas con parámetro $\lambda = 2$. Con estas variables generamos 3000 experimentos de n observaciones cada uno, es decir el primer experimento tiene una observación, el segundo dos y así sucesivamente.

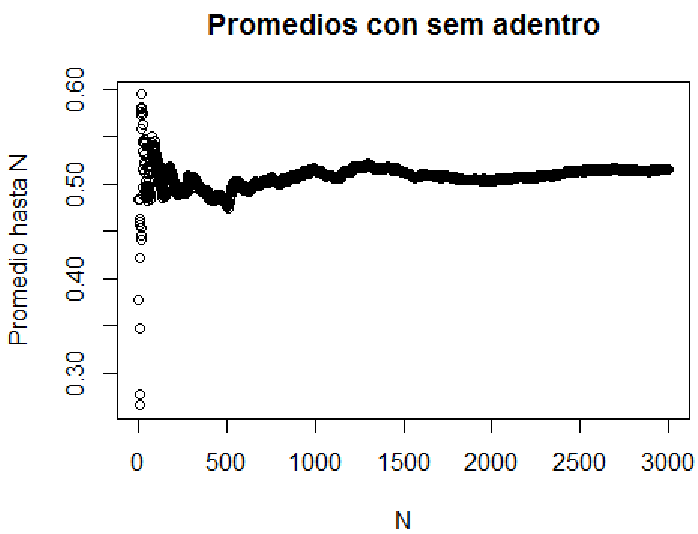
Si tomamos la media de cada uno de los experimentos y la graficamos podermos ver que queda el siguiente gráfico:



Podemos observar en el gráfico como ocurre la Ley de los Grandes Números, a medida que crece la cantidad de observaciones, el resultado del experimento se acerca a la esperanza de una variable exponencial con parámetro $\lambda = 2$.

2.1 Diferencias de Gráficos

Como el título indíca, el gráfico anterior fué computado en R seteando primero una semilla afuera de un ciclo for. Si a esa semilla la colocamos dentro del ciclo el gráfico mostrado cambia al siguiente

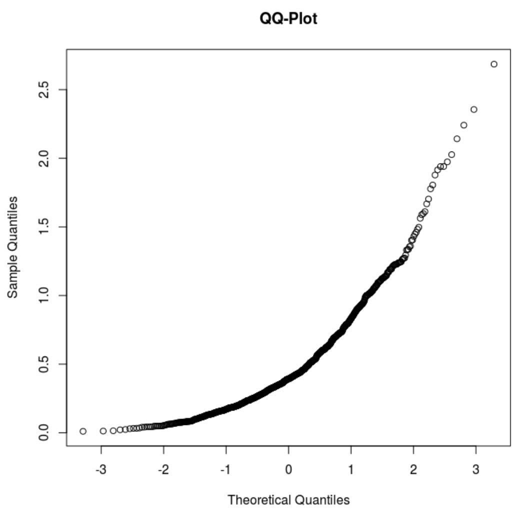


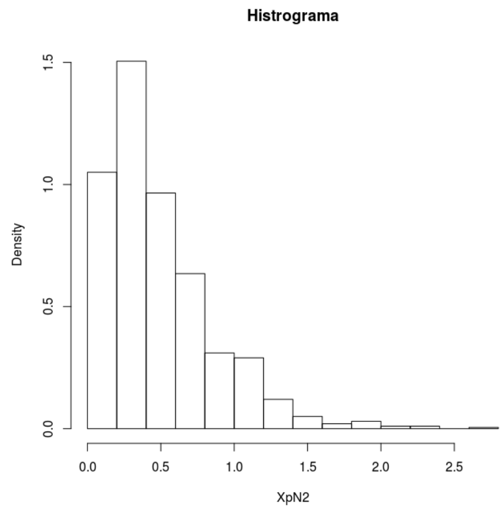
Como se puede ver, los dos gráficos son diferentes, ¿A qué se debe ésta diferencia? Se debe al modo en que R trabaja con las semillas al momento de generar datos al azar.

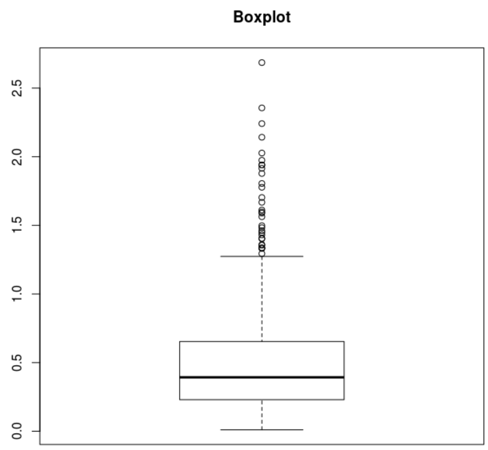
En R al pedir que se genere un valor al azar, la única forma de que ese valor siempre sea el mismo es setear la semilla antes de generarlo. Si se pide que se generen más observaciones de las que se generaron antes, la única observación distinta va a ser la final, es decir, si primero genero una variable con n observaciones y después genero una nueva variable con $n + 1$ observaciones, las primeras n observaciones de esa nueva variable van a ser iguales a las n de la anterior si seteo la misma semilla antes de generarla. En el segundo gráfico se grafican las medias calculadas de la misma forma que en el primer gráfico, pero en este caso, como la semilla esta dentro del ciclo for, las observaciones de cada experimento son las mismas, la única diferencia es la cantidad de observaciones que tiene cada experimento. En este segundo gráfico también se puede ver como las medias se empiezan a aproximar a la esperanza de la variable aleatoria exponencial cuando n es grande.

3 Segundo Ejercicio

En éste ejercicio vamos a poder apreciar la Ley de los Grandes Números, empezaremos guardando la media de dos observaciones de variables exponenciales 1000 veces, resultando en los siguientes gráficos:

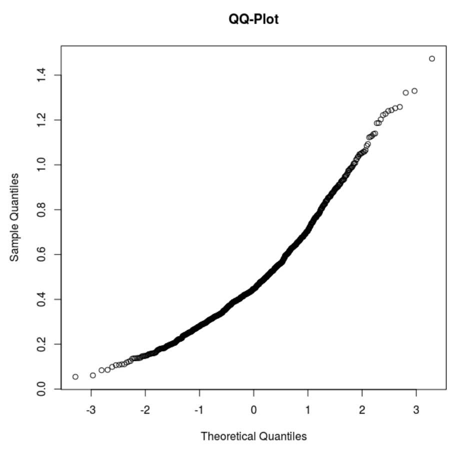


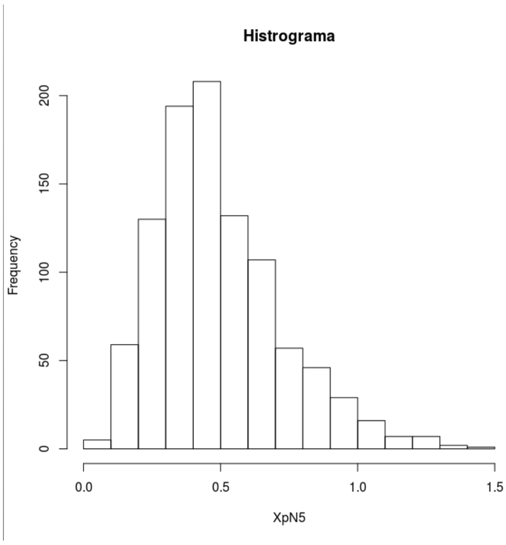


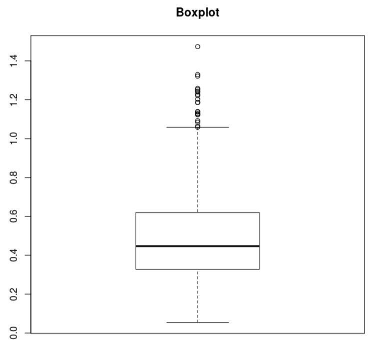


De los gráficos podemos inferir que las medias no poseen una distribución normal, ya que el QQ-plot no se asemeja a una recta, en el histograma se ve que hay mayor concentraciń cerca de la media (aunque no necesariamente en la media), y también podemos ver que hay una cantidad muy grande de "outliers" en el boxplot.

Si aumentamos la cantidad de observaciones por muestra a 5 se obtienen estos gráficos:

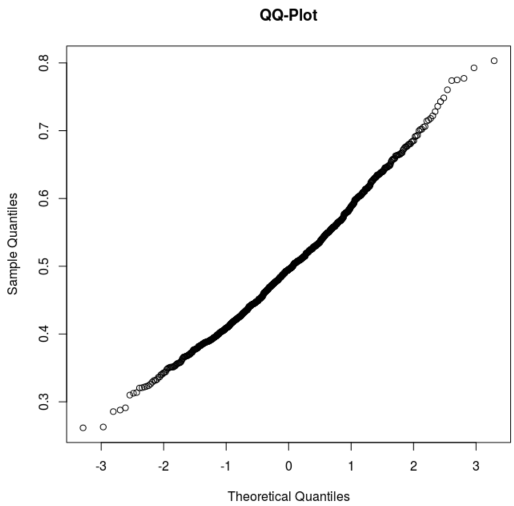


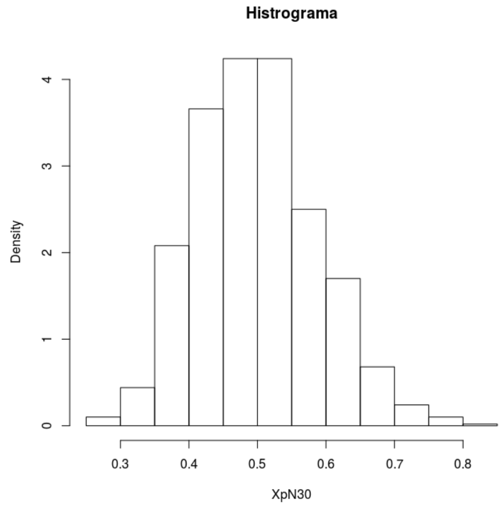


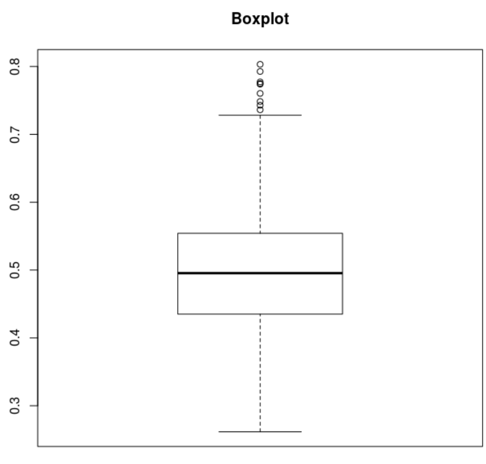


Se empiezan a notar los cambios de los gráficos, en el QQ-plot vemos que los puntos se empiezan a parecer a una recta, en el histograma se ve que la concentración se acerca más a la media, además también se ve que en el boxplot se redujeron la cantidad de "outliers" que caen fuera del gráfico.

Ahora si aumentamos la cantidad de observaciones a 30, los cambios se asentúan aún más, resultando en los siguientes gráficos:

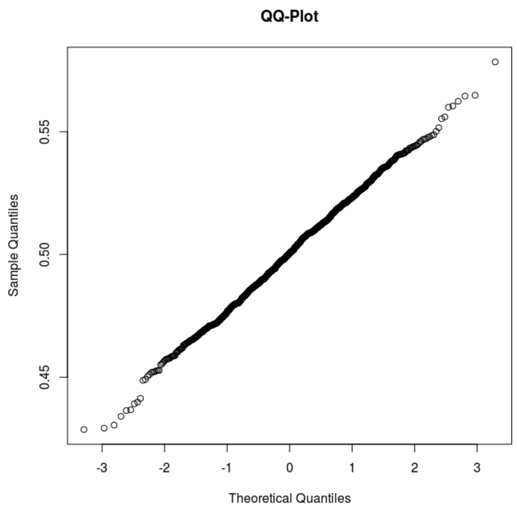


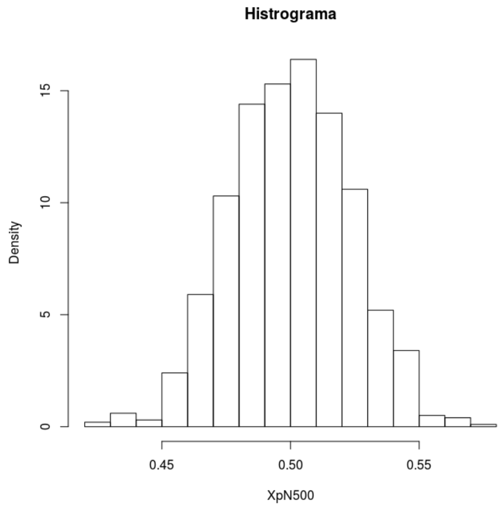


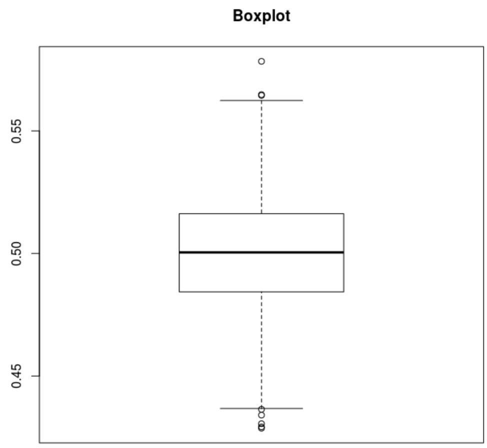


Se puede notar claramente en el QQ-plot que el gráfico se parece a una recta, indicando que la distribución de las medias puede ser normal, a su vez ahora el histograma tiene forma similar a un gráfico de una distribución normal y que el boxplot no sólo tubo otra disminución en la cantidad de "outliers" sino que también esta más centrado, de lo que podemos deducir mayor simetría en la distribución de las medias.

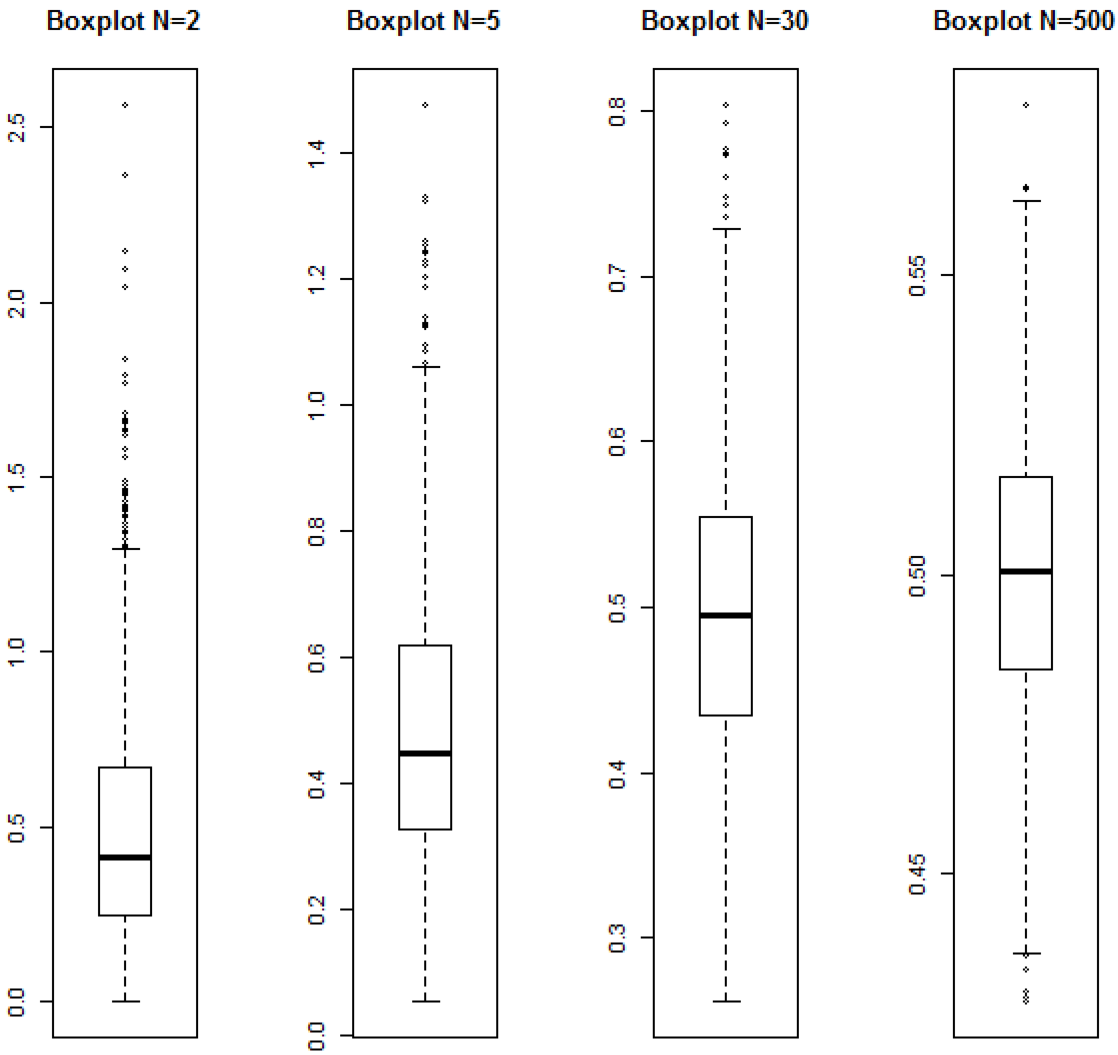
Finalmente si aumentamos las observaciones a 500 por muestra, se obtienen estos gráficos:







En estos gráficos adems de ver aún más acentuación de las características mencionadas anteriormente, podemos ver que la mayoría de las medias cae cerca de la esperanza de las variables. Para marcar ese progreso podemos ver en conjunto los boxplot de los experimentos



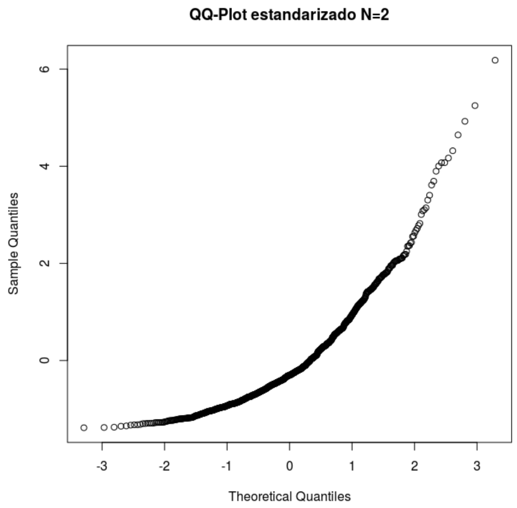
Luego de ver estos datos podemos inferir lo siguiente, a medida que aumentamos la cantidad de observaciones en cada muestra, estas muestras se acercan cada vez más a la esperanza de la variable, podemos decir entonces que si tomaramos una infinita cantidad de observaciones entonces las muestras deberían converger a la esperanza. De esto mismo habla la Ley de los Grandes Números, que dice que la media de las observaciones de n variables aleatorias i.i.d. con igual esperanza μ finita se aproxima en probabilidad a μ.

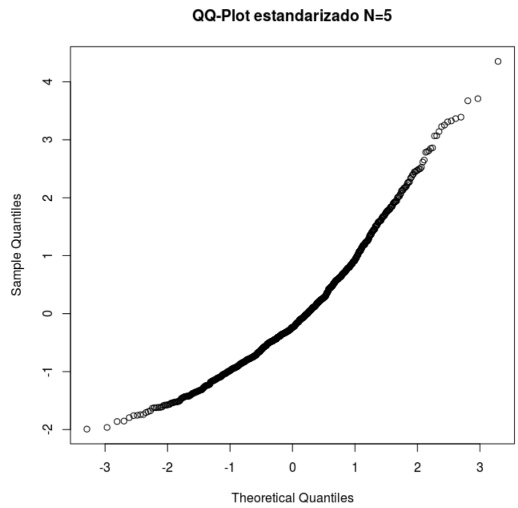
Además, medida que aumentamos las observaciones de las muestras, pudemos ver que los gráficos la distribución de las medias se iba acercando al gráfico que le corresponde a una variable aleatoria normal (de parámetros desconocidos), entonces si tomáramos una infinita cantidad de observaciones podríamos decir que la distribución de las medias es aproximadamente la de una normal (de parámetros desconocidos).

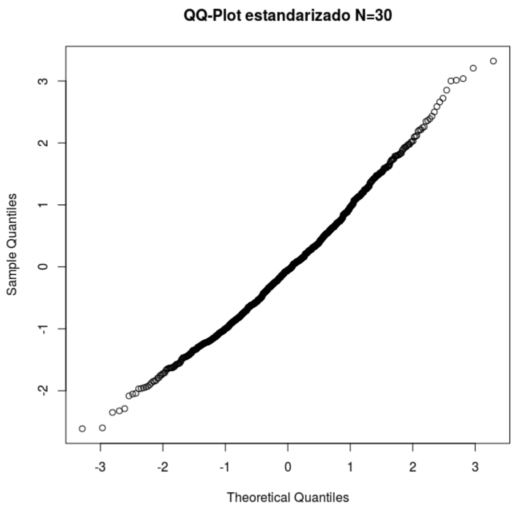
4 Tercer Ejercicio

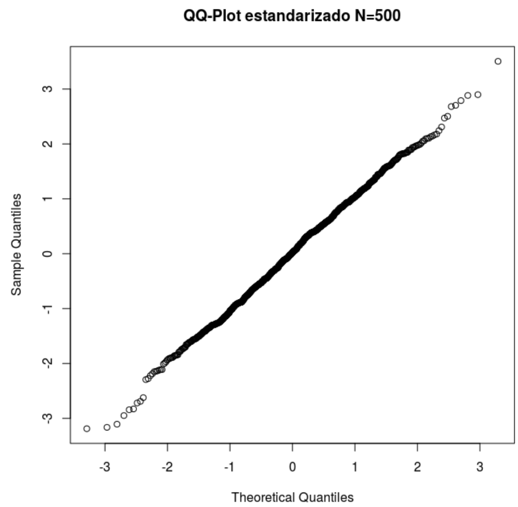
Tomando $X_1, ..., X_n$ variables aleatorias i.i.d. con distribución exponencial de parámetro $\lambda = 2$, vamos a comprobar que la distribución de la variable aleatoria $\frac{\bar{X_n} - E(X_1)}{\sqrt{\frac{Var(X_1)}{n}}}$ estandarizada se aproxima a la de una normal estandarizada cuando n es grande. Primero veamos $E(X_1)$ y $Var(X_1)$, como sabemos que $\lambda = 2$ entonces $E(X_1) = \frac{1}{\lambda} = \frac{1}{2}$ y $Var(X_1) = \frac{1}{\lambda^2} = \frac{1}{4}$.

Ahora si estandarizamos los conjuntos de variables aleatorias del ejercicio anterior, obtenemos como resultando los siguientes gráficos:

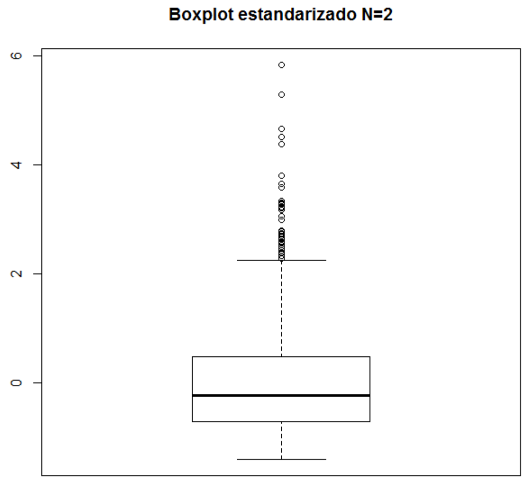


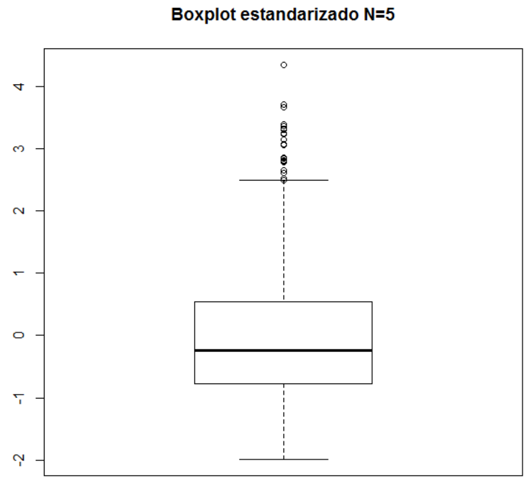


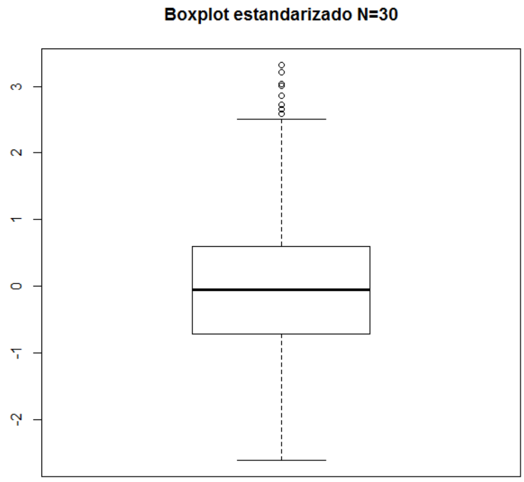


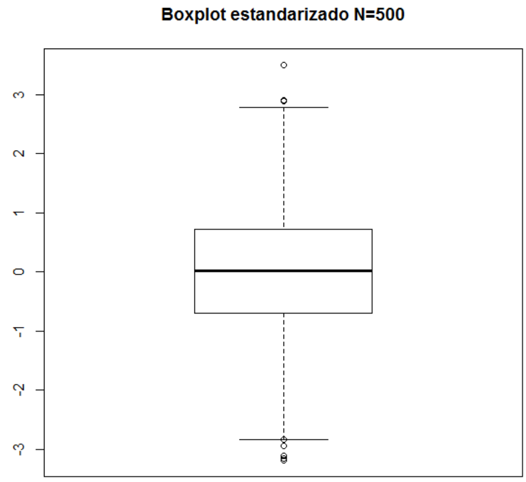


En estos QQ-plots se puede ver que a medida que n se hace más grande, la distribución de la variable estandarizada se parece más a la de una normal estándar de parámetros $\mu = 0$ y $\sigma^2 = 1$. Observando los siguientes boxplots también podemos ver que la distribución se vuelve más simétrica a medida que aumenta n:

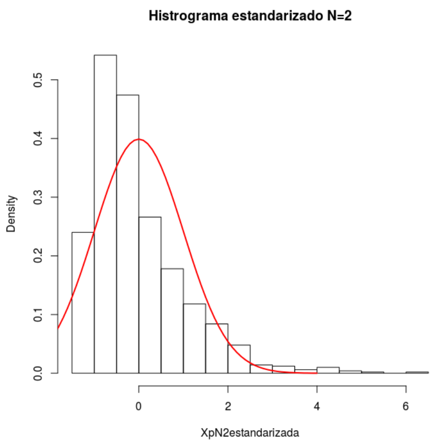


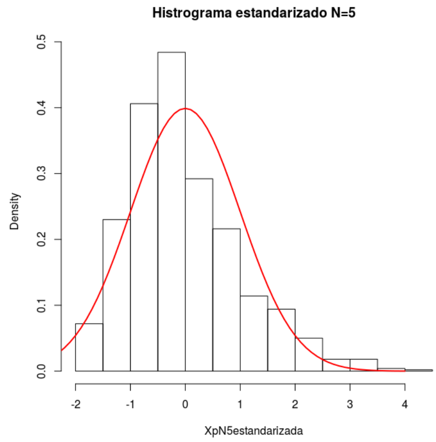


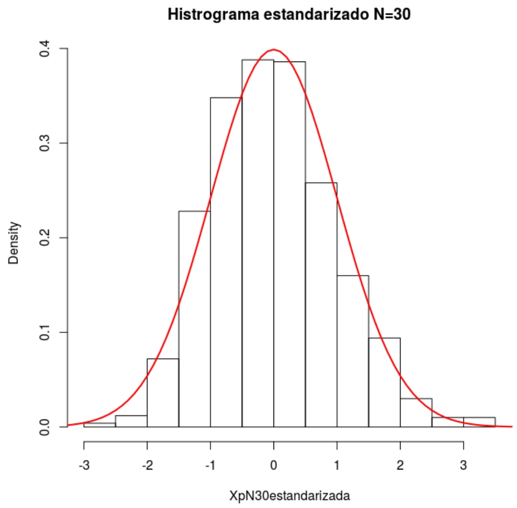


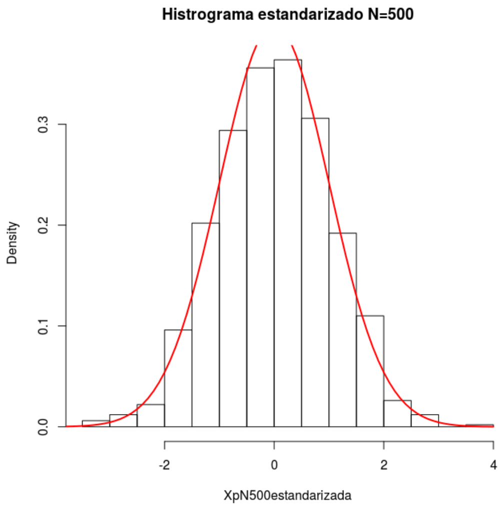


Finalmente si vemos los histogramas con una función normal estándar superpuesta encima, se puede apreciar como el gráfico se asemeja cada vez más a medida que crece la n:






Histrograma estandarizado N=30
Density
0.0
0.1
0.2
0.3
0.4
-3
-2
-1
0
1
2
3
XpN30estandarizada

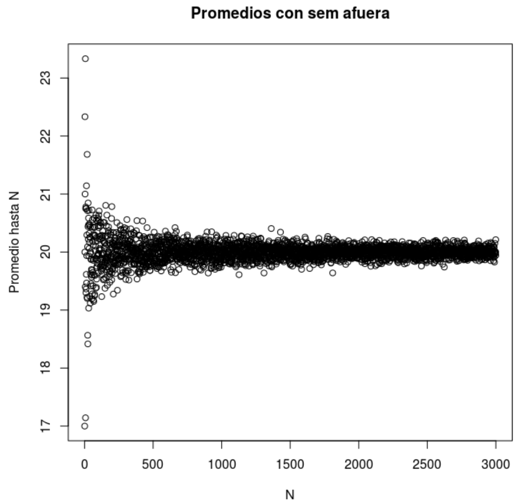

Histrograma estandarizado N=500
Density
0.0
0.1
0.2
0.3
-2
0
2
4
XpN500estandarizada

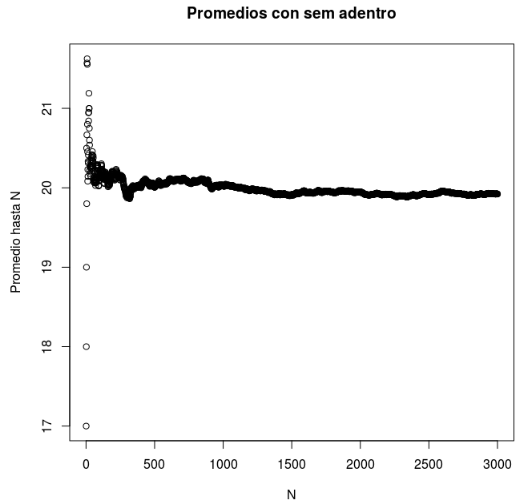
5 Cuarto Ejercicio

Para mostrar que la Ley de los Grandes Números y el Teorema Central del Límite no necesariamente son únicos de las variables continuas, vamos a realizar los ejercicios anteriores cambiando la distribución de las variables aleatorias usadas con una diferente y veremos que los resultados son similares. La nueva distribución de las variables aleatorias va a ser una binomial de parámetros $n = 6$ y $p = \frac{1}{9}$.

5.1 Primer Ejercicio Modificado

Realizando el mismo experimento que el primer ejercicio terminamos con los siguientes gráficos:

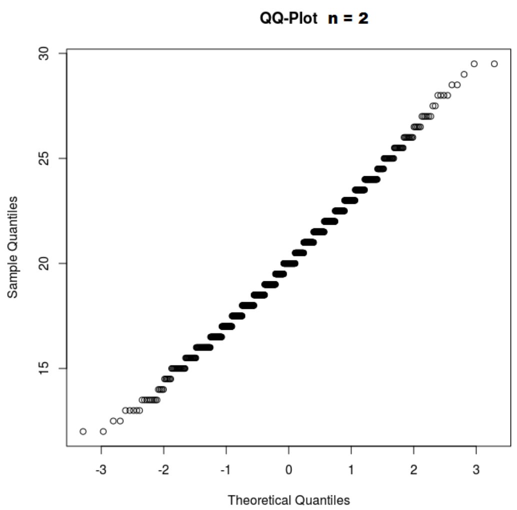


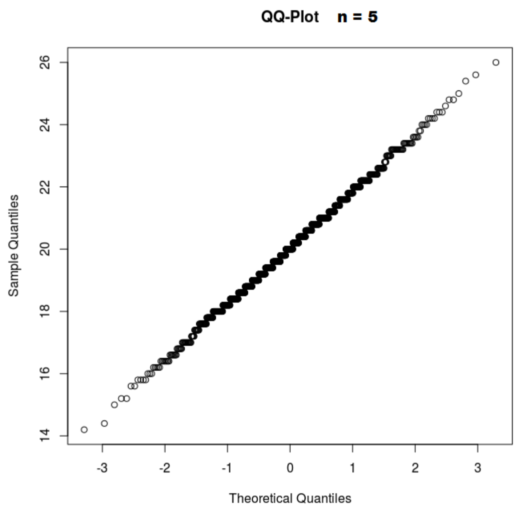


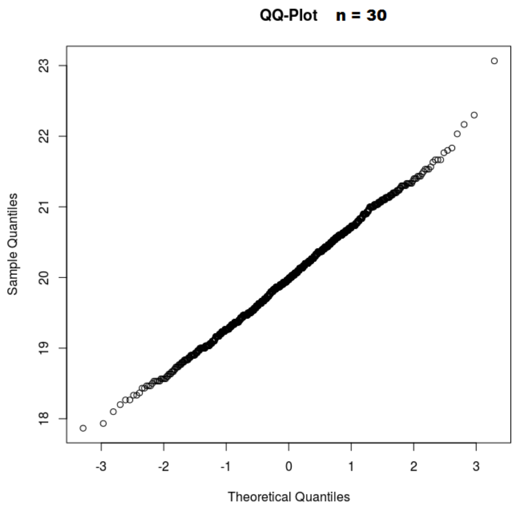
Podemos ver que pese a que cambiamos la distribución dela variable aleatoria utilizada por una discreta, terminamos con gráficos que tienen características similares a las del primer ejercicio original. El gráfico que tiene la semilla afuera converge a la esperanza de una variable aleatoria binomial de parámetros $n = 6$ y $p = \frac{1}{9}$ y el gráfico con la semilla afuera funciona de forma idéntica al del ejercicio original.

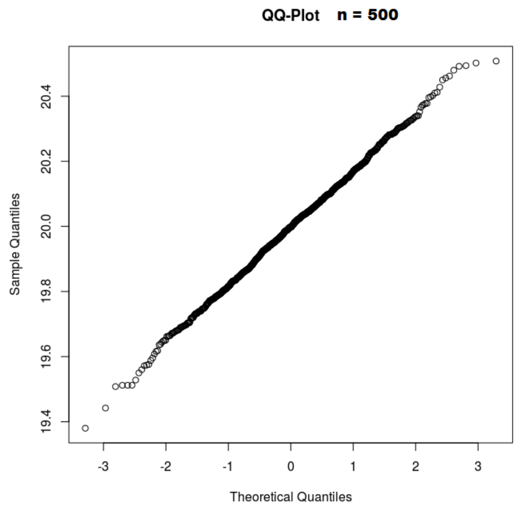
5.2 Segundo Ejercicio Modificado

Cambiando la distribución de la variable en este experimento, algunos gráficos son un poco distintos pero aún asi se comportan de forma similar. Veamos los QQ-plot primero:



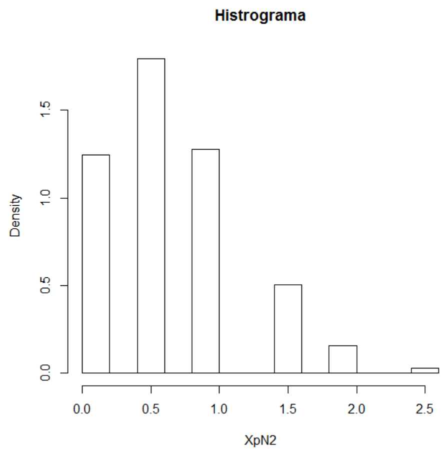


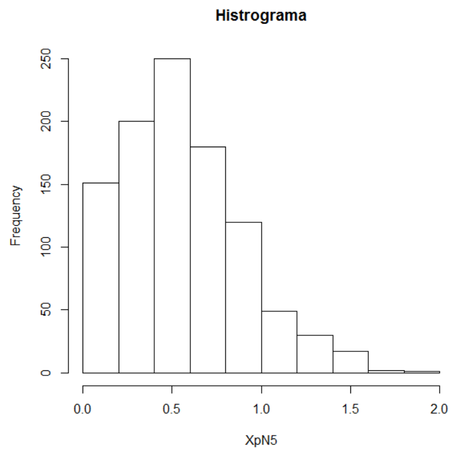


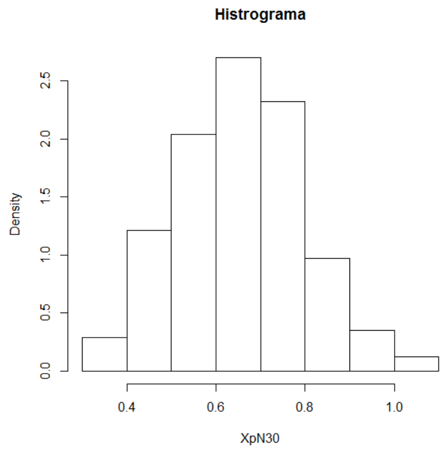


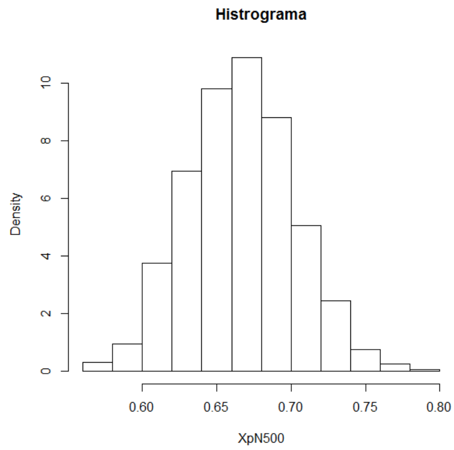
Podemos ver que al aumentar el n, la distribución se va pareciendo más a una normal como sucedió el segundo ejercicio.

Si vemos los histogramas también se nota un comportamiento similar al del segundo ejercicio:



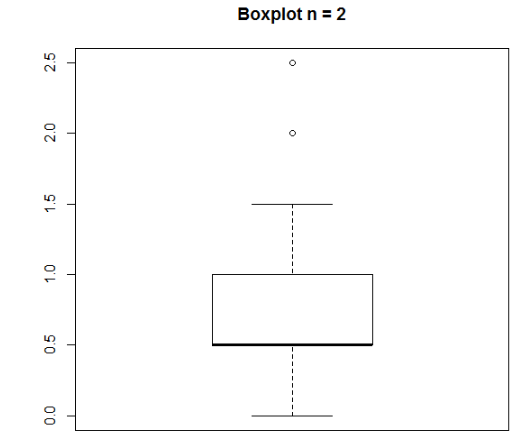


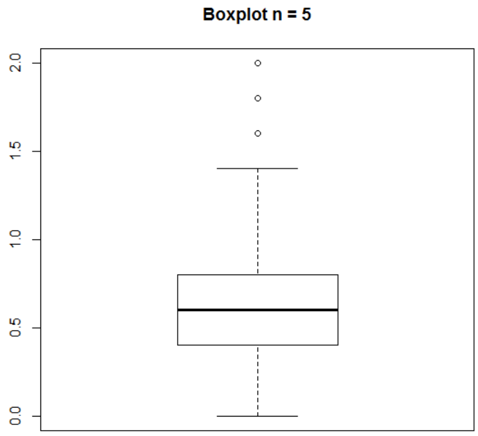


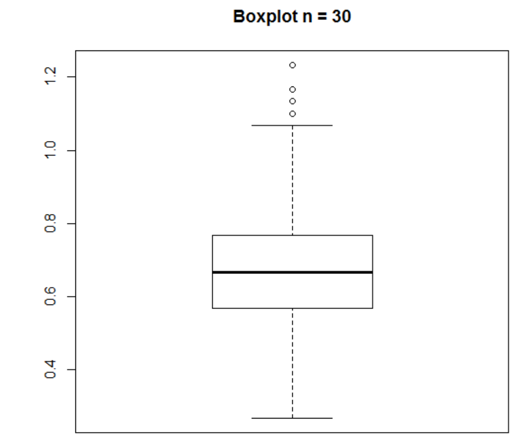


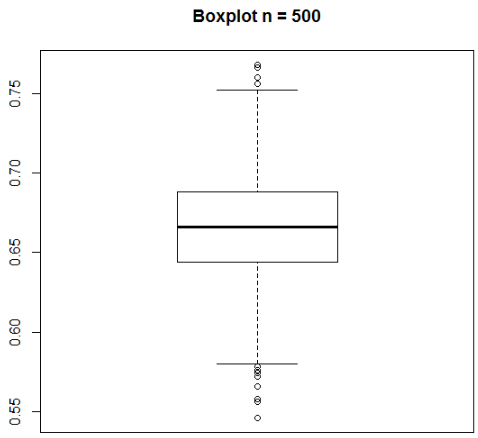
Vemos que los gráficos se parecen a los de una normal a medida que aumentamos la n, también vemos que cada ves que aumenta n, las muestras se acercan al valor de la esperanza.

Finalmente viendo los boxplots, vemos que también se centran a medida que el n se vuelve más grande, pareciéndose al boxplot de una normal.



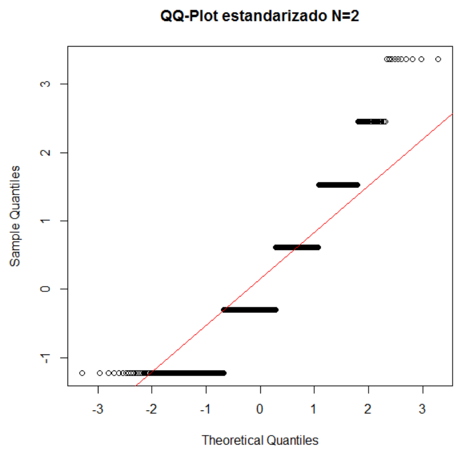


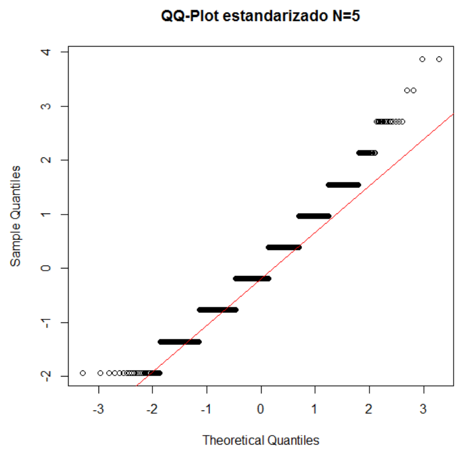


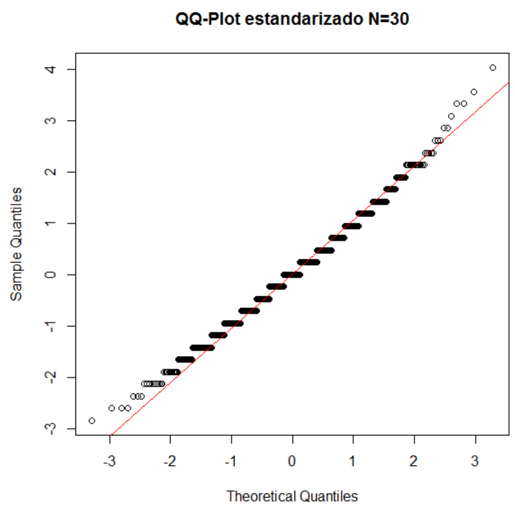


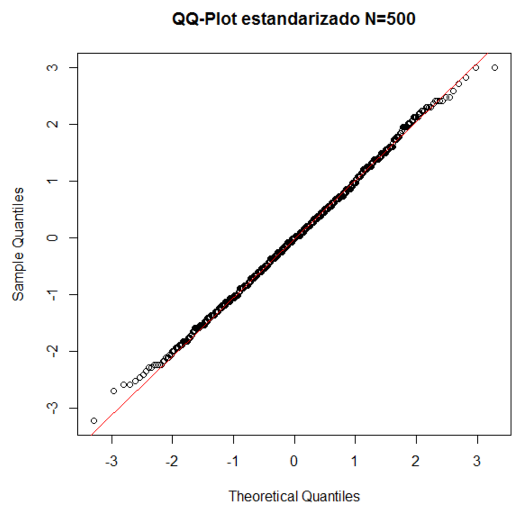
5.3 Tercer Ejercicio Modificado

Si estandarizamos las variables con distribución binomial de la misma forma que hicimos en el tercer ejercico original obtenemos los siguientes gráficos:



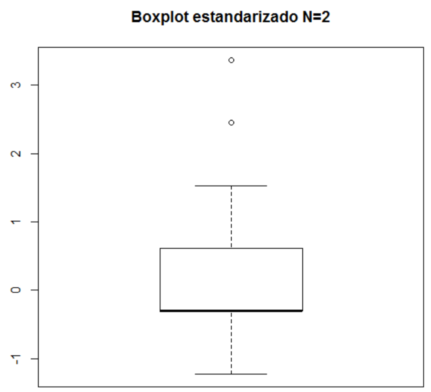


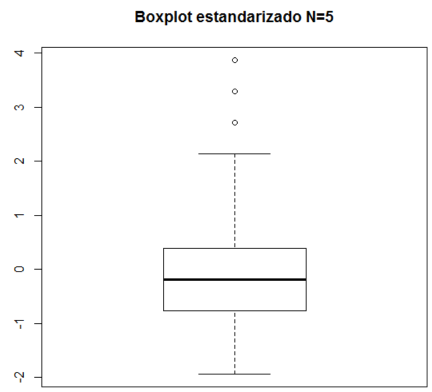


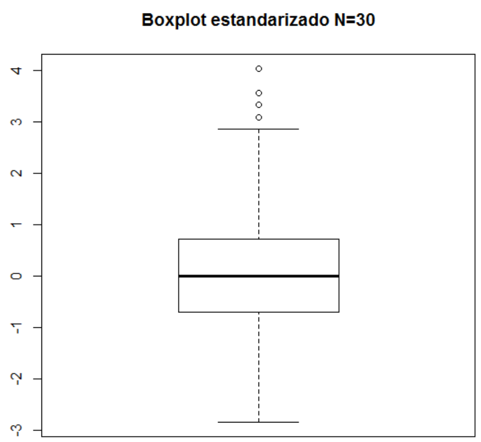


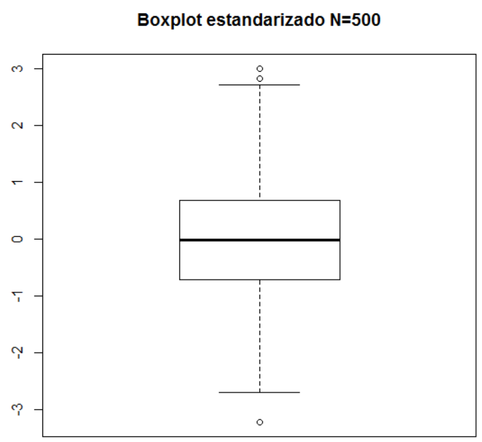
Se puede ver que mientras ms aumenta la n, el gráfico se parece cada ves más a la recta que seguiría una normal de parámetros $\mu = 0$ y $\sigma^2 = 1$.

Los boxplots de los experimentos se ven de la siguiente forma:



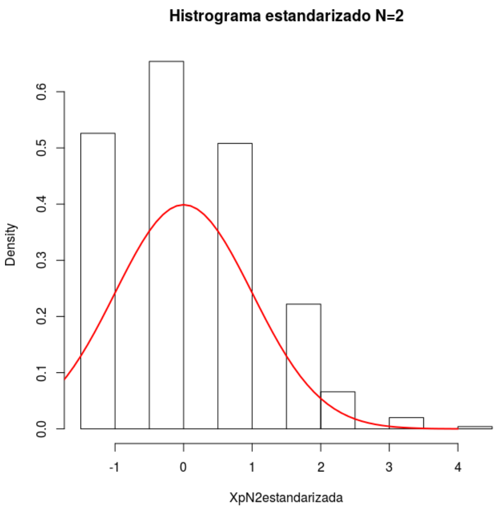


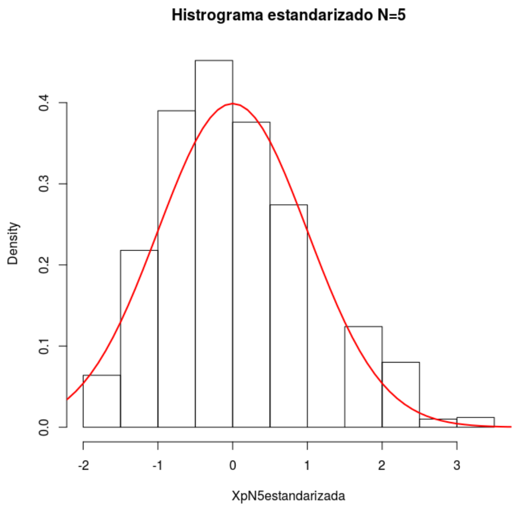


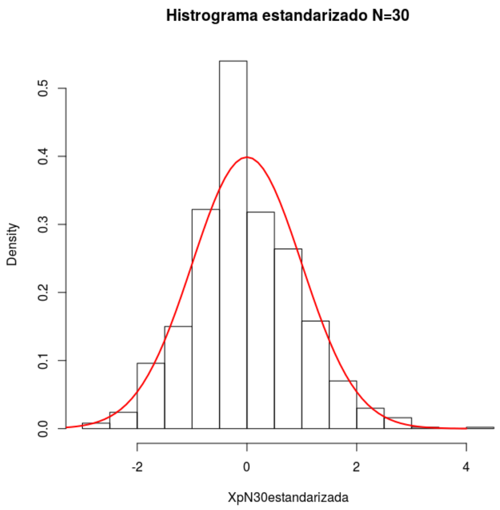


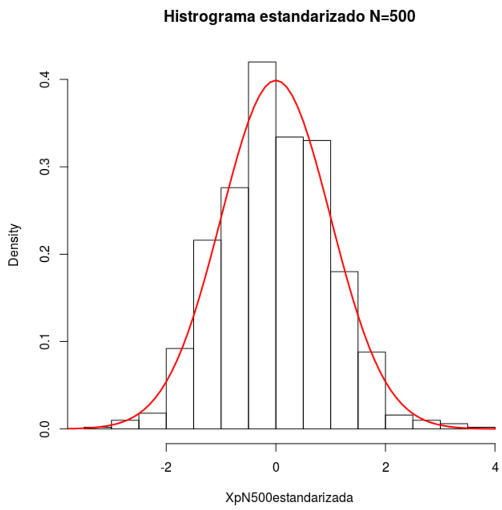
Podemos ver que los boxplots también se parecen a los de una normal de parámetros $\mu = 0$ y $\sigma^2 = 1$ a medida que n aumenta.

Finalmente viendo los histogramas:









Podemos ver que a medida que aumenta n, el gráfico se parece cada vez más el de una normal de parámetros $\mu = 0$ y $\sigma^2 = 1$.