TP – Probabilidad y Estadística

**Parte 1: Simulación**

1. Utilizando únicamente la función *random* de su lenguaje (la función que genera un número aleatorio uniforme entre 0 y 1),

implemente una función que genere un número distribuido Bernoulli con probabilidad *p*.

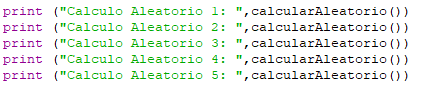
Generación del número aleatorio uniforme:

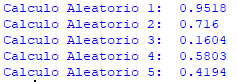
La librería dice que es uniforme y no contempla el 1 - Random.random()

Con la función random se genera el número aleatorio y se limita a 4 decimales.

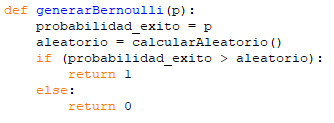


Ejemplo: Ejecuto la función 5 veces





Con la función Bernoulli, utilizamos la función anterior para generarla, recibe la probabilidad como parámetro:



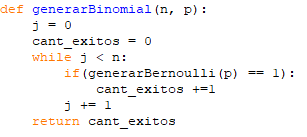
Ejemplo:





2. Utilizando la función del punto anterior, implemente otra que genere un número binomial con los parámetros *n*,*p*.

Función creada a partir de la formula binomial. Recibe N = cantidad de muestra y P = probabilidad. Se utiliza la función Bernoulli para saber si es éxito o fracaso.



Ejemplo:

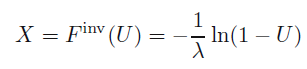


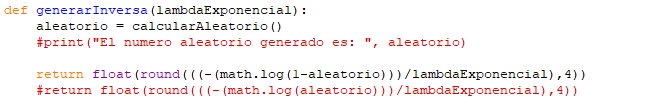




3. Utilizando el procedimiento descrito en el capítulo 6 del Dekking (método de la función inversa o de Monte Carlo), implementar una función que permita generar un número aleatorio con distribución *Exp*(*¸*).

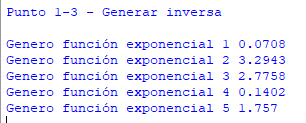
Utilizamos la siguiente función: Tal como se demuestra en el Dekking, invertimos la función exponencial, quedando:





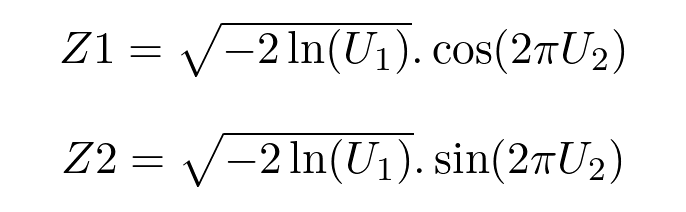
Ejemplo:



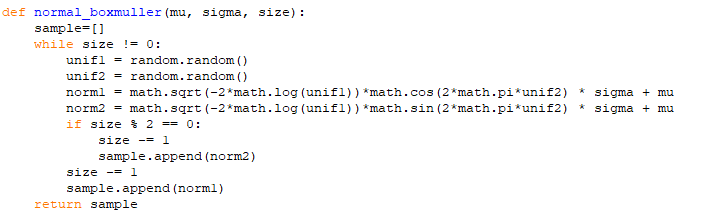


4. Investigar como generar números aleatorios con distribución normal. Implementarlo.

Fórmula utilizada: METODO DE BOX MULLER – Encontramos que existe el método de box muller para generar muestras normales a partir de números uniformes aleatorios (las Z1 y Z2).



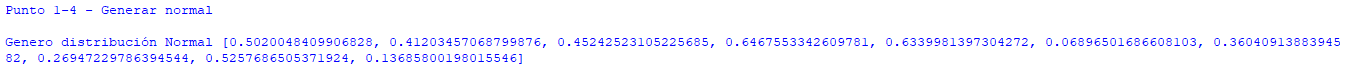
Luego, la estandarizamos con una media y varianza deseada para generar las muestras (si se aplica con 0 media y 1 varianza obtenemos una normal estándar)



Ejemplo:

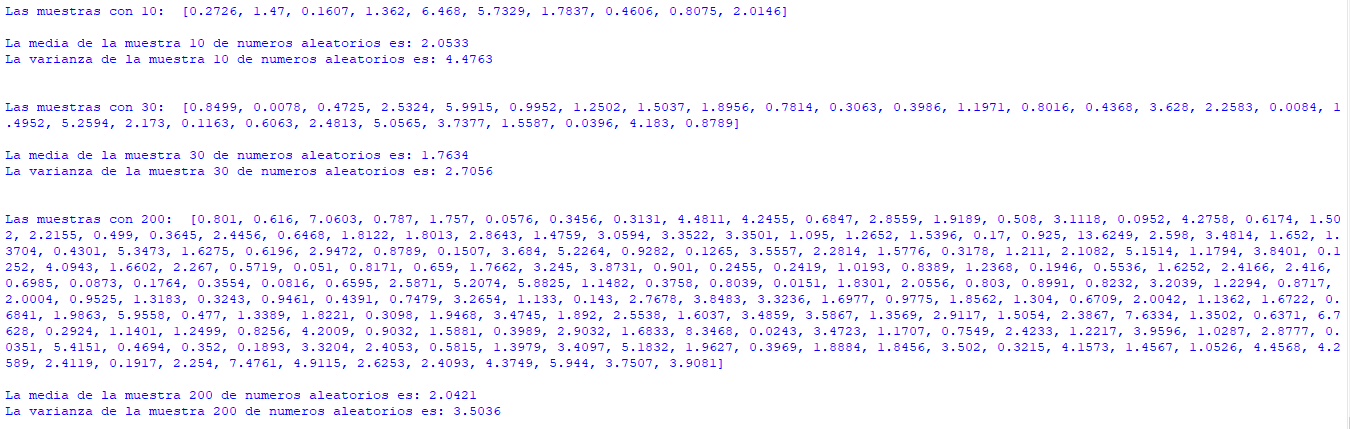






**Parte 2: Estadística descriptiva**

1. Generar tres muestras de números aleatorios *Exp*(0,5) de tamaño *n =* 10, *n* = 30 y *n* = 200. Para cada una, computar la media y varianza muestral. ¿Qué observa?





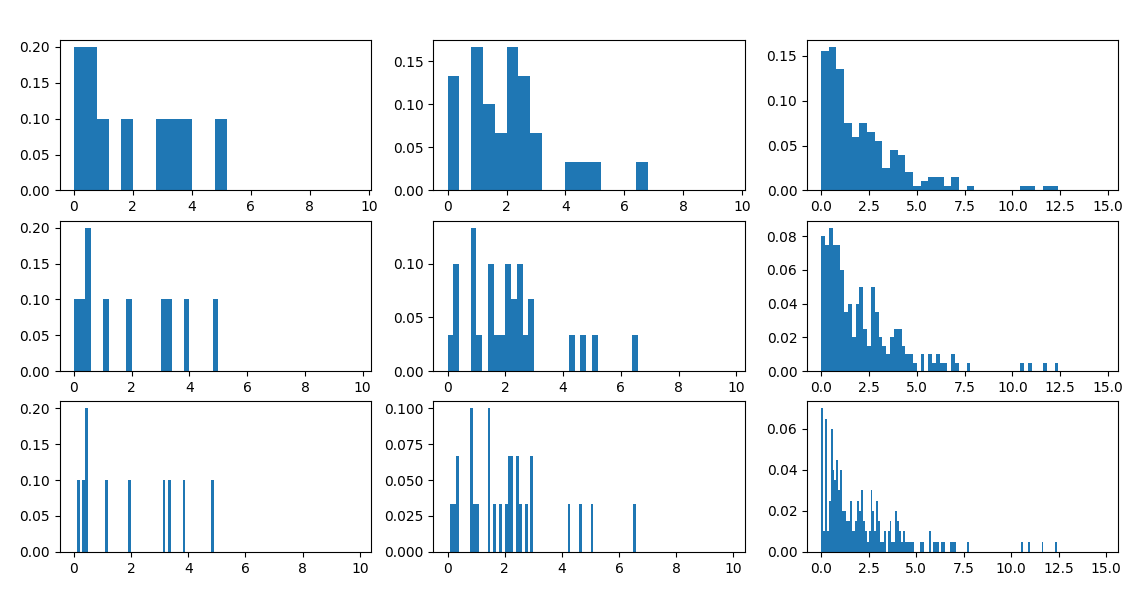




¿Qué observa? Observamos que la media y la varianza a medida que va creciendo N es masomenos la misma. No cambia porque todos las muestras rondan sobre los mismos valores.

2. Para las tres muestras anteriores, graficar los histogramas de frecuencias relativas con anchos de banda 0,4, 0,2 y 0,1; es decir, un total de 9 histogramas. ¿Qué conclusiones puede obtener?

10 30 200



0.4

0.2

0.1

10 30 200

Nos permite verificar que nuestras muestras efectivamente se aproximan

a la distribución exponencial, siendo más evidentes cuando mayor es el tamaño de la muestra.

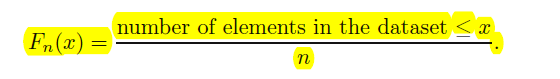
Cuando el ancho de banda es menor, se puede notar mejor las diferencias entre las muestras.

3. Generar una muestra de números *Bin*(10, 0.3) de tamaño *n* = 50. Construir la función de

distribución empírica de dicha muestra.

La muestra binomial es: [2, 4, 1, 1, 2, 2, 2, 3, 0, 3, 4, 1, 4, 3, 3, 1, 2, 3, 1, 2, 4, 4, 4, 4, 3, 1, 3, 5, 4, 3, 3, 8, 4, 3, 0, 4, 4, 6, 2, 3, 4, 0, 1, 3, 6, 4, 2, 5, 2, 4]

Función de distribución empirica:



[0.06, 0.2, 0.38, 0.62, 0.9, 0.94, 0.98, 1.0]

Para x > 0

F(x) es: 0.06

Para x > 1

F(x) es: 0.2

Para x > 2

F(x) es: 0.38

Para x > 3

F(x) es: 0.62

Para x > 4

F(x) es: 0.9

Para x > 5

F(x) es: 0.94

Para x > 6

F(x) es: 0.98

Para x > 8

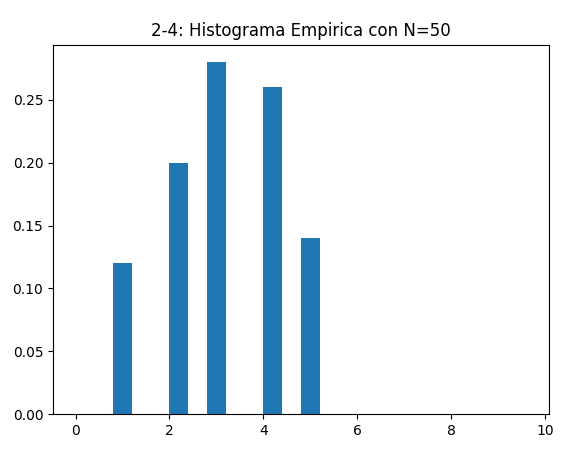
F(x) es: 1.0

F(x) = 0 en otro caso

4. A partir de la función de distribución empírica del punto anterior, generar una nueva muestra de números aleatorios utilizando el método de simulación de la primera parte. Computar la media y varianza muestral y graficar el histograma.

La media de la muestra empirica de numeros aleatorios es: 2.92

La varianza de la muestra empirica de numeros aleatorios es: 2.1936

 Grafico:

#Primero generamos la empirica a partir de la acumulada de las muestras binomiales. Luego para generar muestras aleatorias a partir

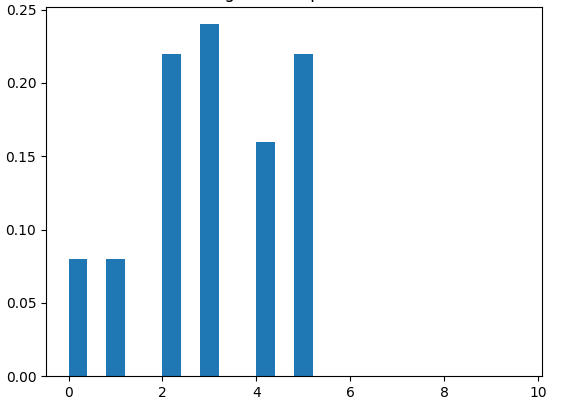
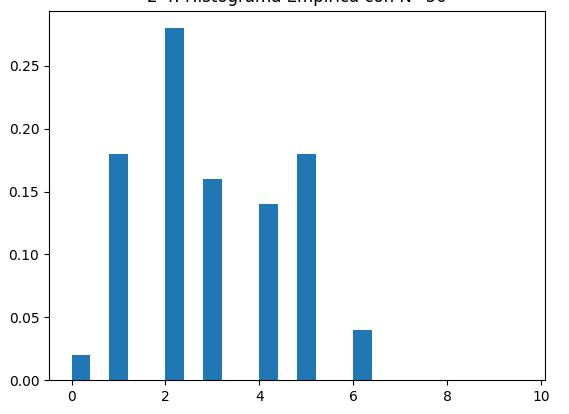
#de la empirica, se realiza una inversa. Genero un numero aleatorio uniforme, y lo comparo con la muesta empirica en la posicion.

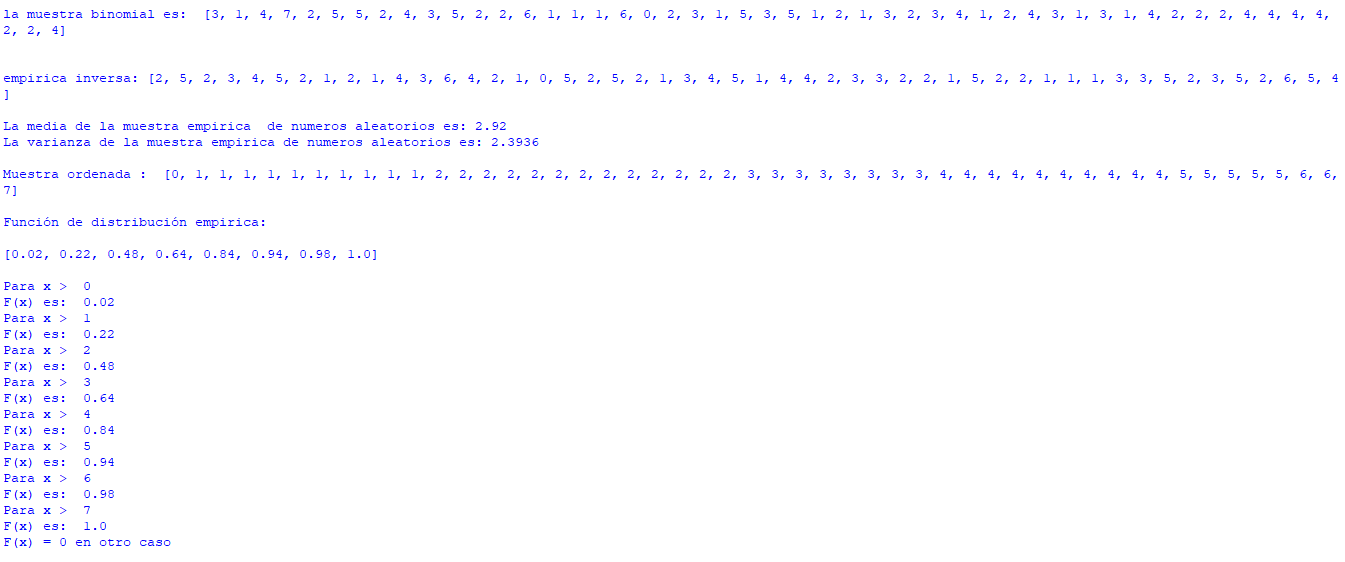
#si la empirica es mayor o igual al aleatorio, devuelvo la muestra binomial en la posicion CONTADOR

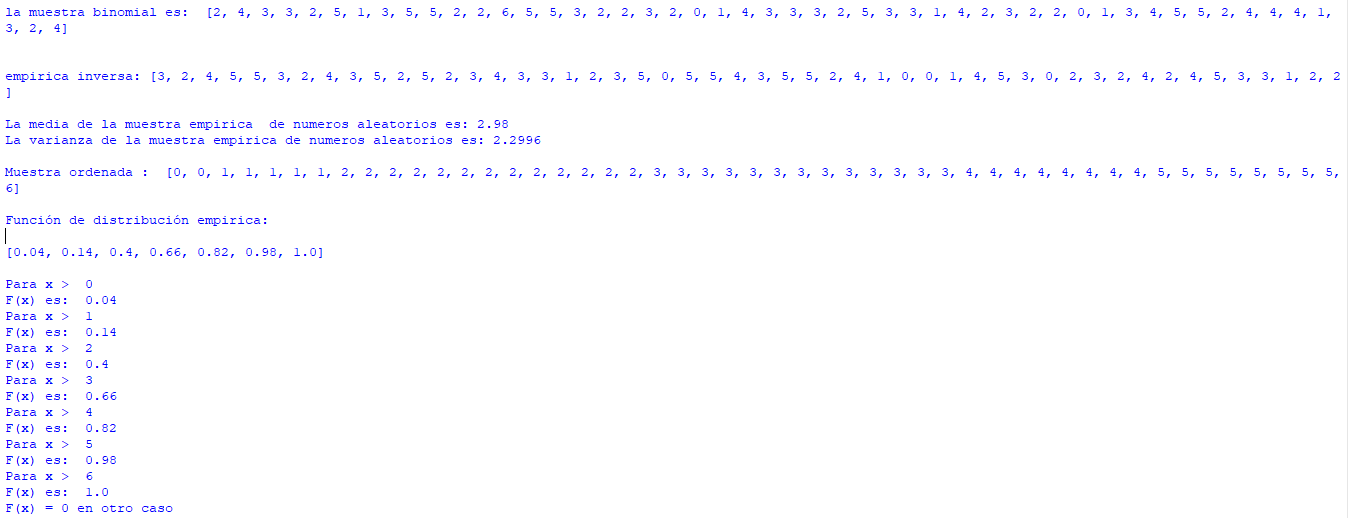
#si es menor, sumo el contador y comparo la siguiente muestra empirica.

5. Repetir el experimento de los dos puntos anteriores con dos muestras aleatorias más generadas con los mismos parámetros.

¿Qué conclusión saca?



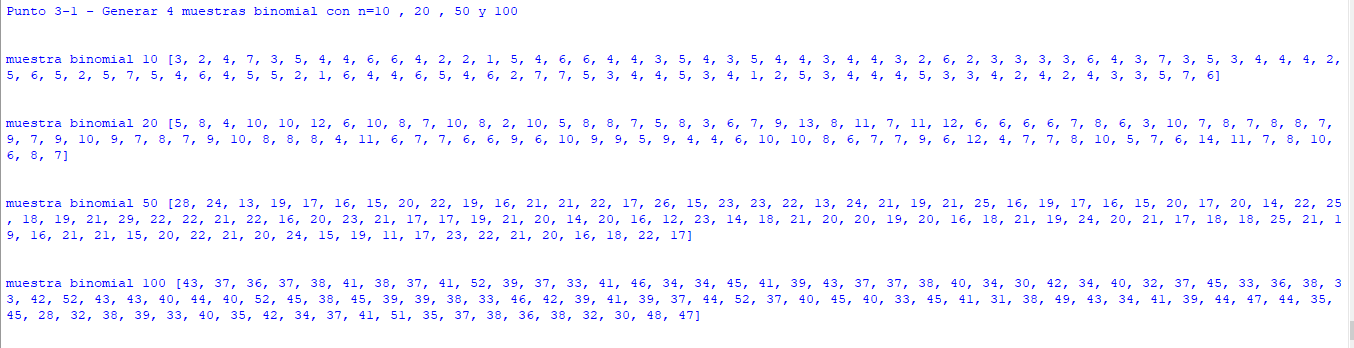




**Parte 3: Convergencia**

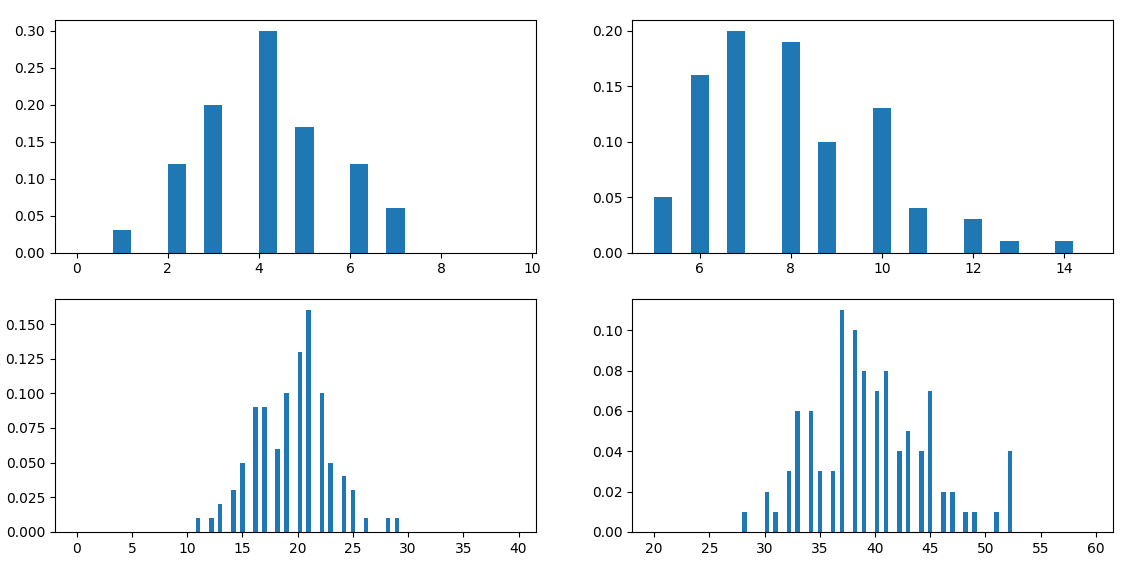
El propósito de esta sección es ver en forma práctica los resultados de los teoremas de convergencia.

1. Generar cuatro muestras de números aleatorios de tamaño 100, todas con distribución binomial con *p* = 0,40 y *n* = 10, *n* = 20, *n* = 50 y *n* = 100 respectivamente. Graficar sus histogramas. ¿Qué observa?

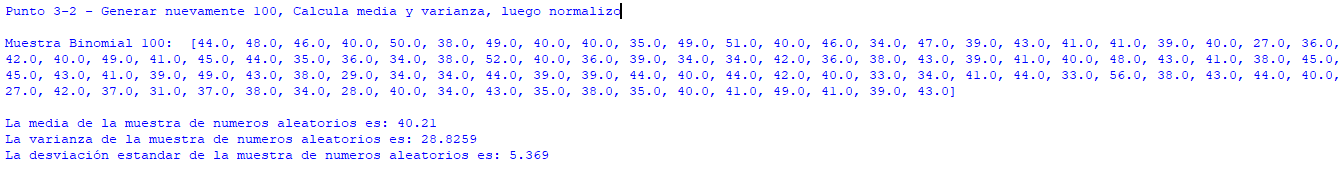


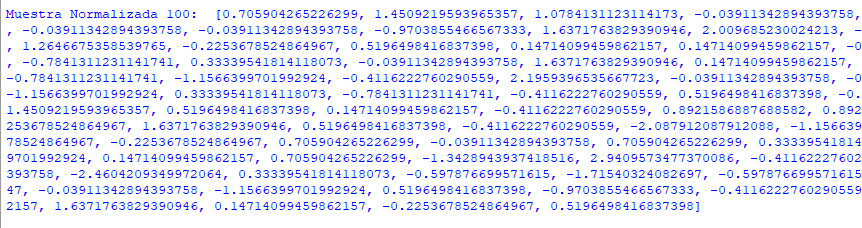
Graficos: la muestra empieza a tener la forma de una campana de Gauss y que se

aproxima a una muestra de distribucion normal.

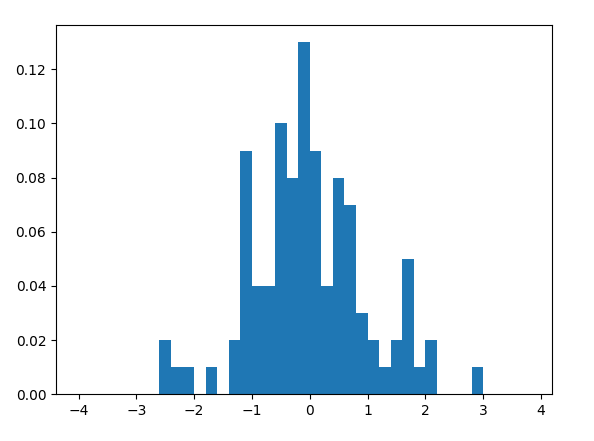


2. Elija la muestra de tamaño 200 (en realidad es 100) y calcule la media y desviación estándar muestral. Luego, normalice cada dato de la muestra y grafique el histograma de la muestra normalizada. Justifique lo que observa.

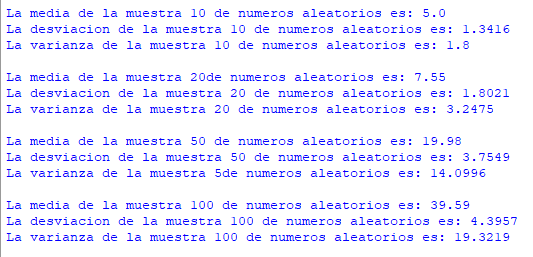




HISTOGRAMA: La forma de la campana caracteristica de una muestra con distribucion normal es mas marcada



3. Para cada una de las muestras anteriores, calcule la media muestral. Justifique lo que observa.



La media y la varianza aumenta a mayor cantidad de experimentos Bernoulli de cada muestra y se mantiene el parámetro de probabilidad de éxito, el error relativo tiende a disminuir

**Parte 4: Estadística inferencial**

Para terminar, vamos a hacer inferencia con las muestras que generamos y obtener así información sobre sus distribuciones.

1. Generar dos muestras *N*(100, 5), una de tamaño *n* = 10 y otra de tamaño *n* = 30. Obtener estimaciones puntuales de su media y varianza.

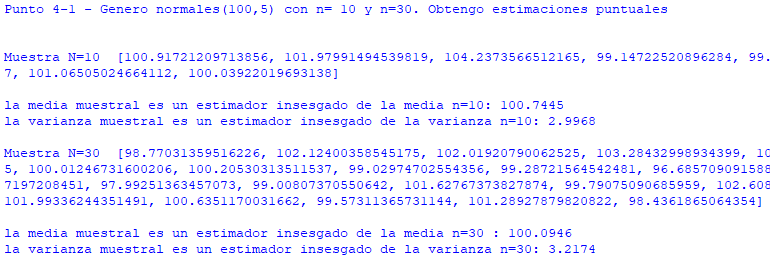
Como se vio a lo largo del curso la media muestral (o ¯ x) es un estimador insesgado de la media

por lo que es el estimador que utilizaremos.

La varianza muestral(o s2) es un estimador insesgado de la varianza por lo que es el estimador

que utilizaremos.

ESTIMADOR INSESGADO:

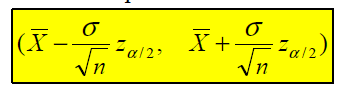


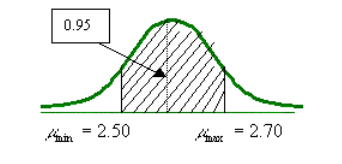
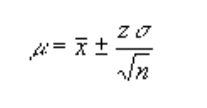
Como es una distribución normal, la media y la varianza muestral se mantienen al aumentar las muestras (N)

2. Suponga que ya conoce el dato de que la distribución tiene varianza 5. Obtener intervalos de confianza del 95% y 98% para la media de ambas muestras.

Como tenemos varianza conocida, se utiliza T de Student.

Utilizamos la tabla AREA BAJO LA CURVA NORMAL.





95% - nos sobra 5% = 0,05 = lo que nos deja en cada cola: 0,0250

Buscamos por table ese valor y sacamos el valor de Z:

En este caso es 1,96

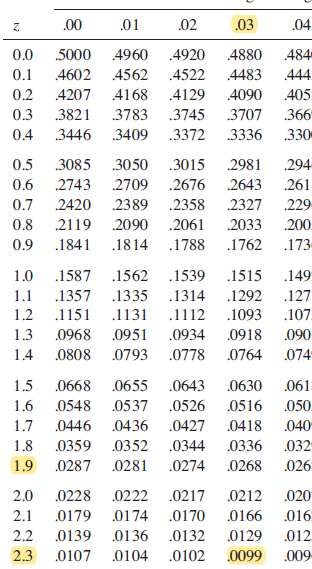
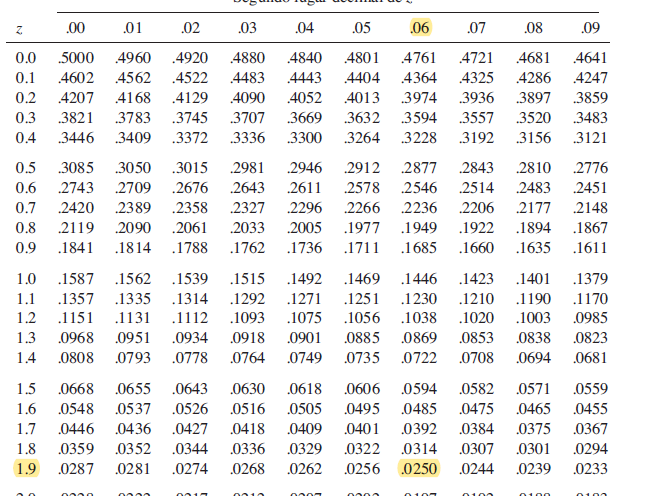
98% - nos sobra 2% = 0,02 = lo que nos deja en cada cola: 0,01

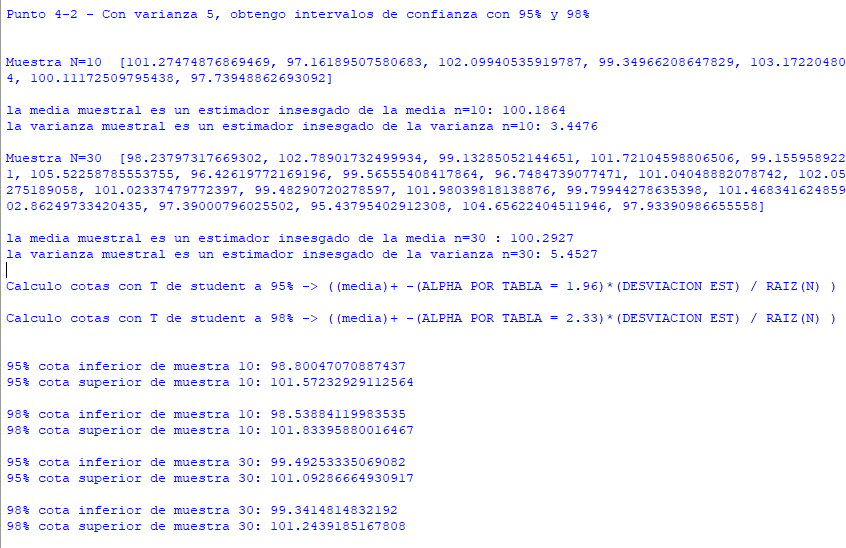
Buscamos por table ese valor y sacamos el valor de Z:

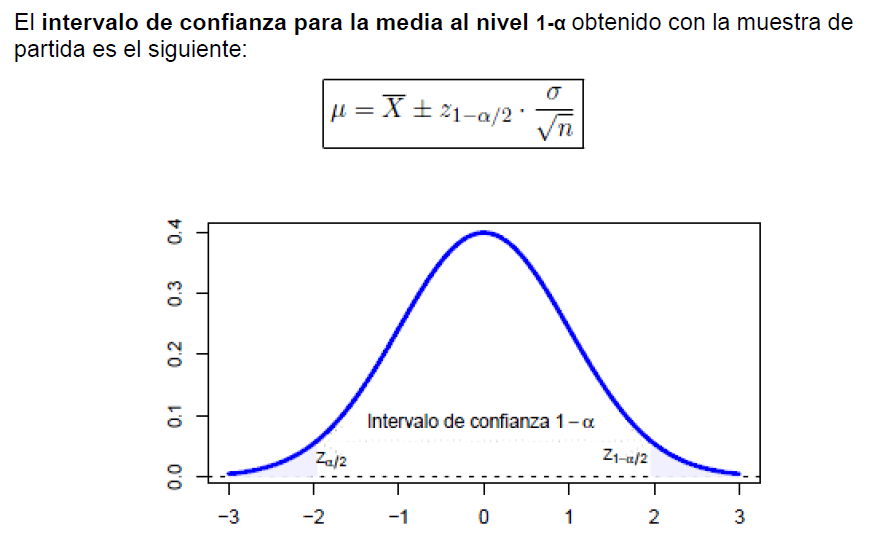
En este caso es

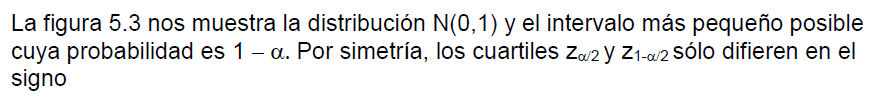
Calculo cotas con T de student a 95% -> ((media)+ -(ALPHA POR TABLA = 1.96)\*(DESVIACION EST) / RAIZ(N) )

Nuestras cotas serían: (99.208 , 100.808)



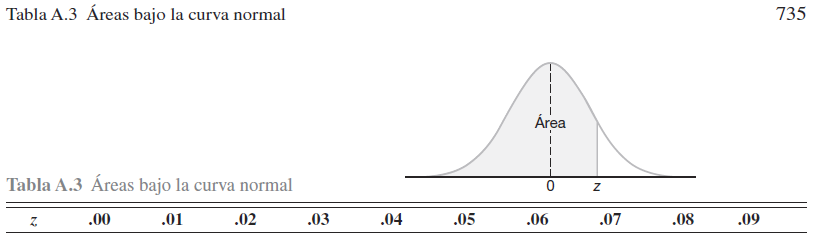


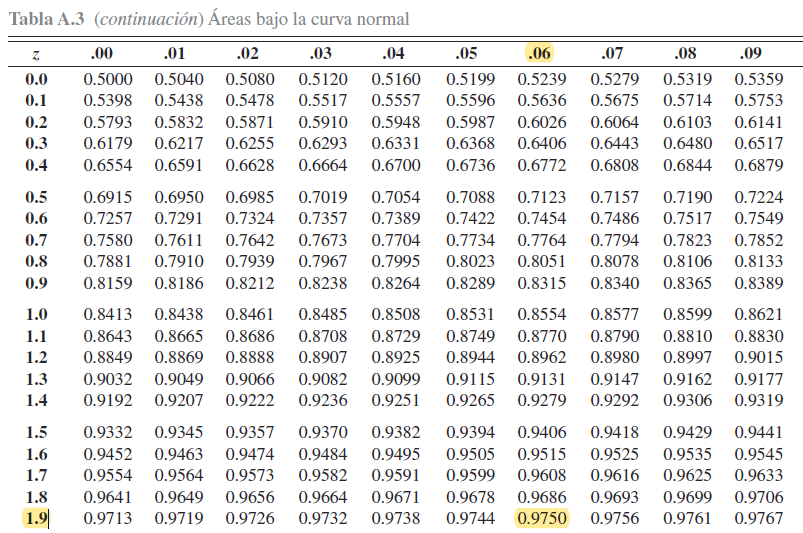




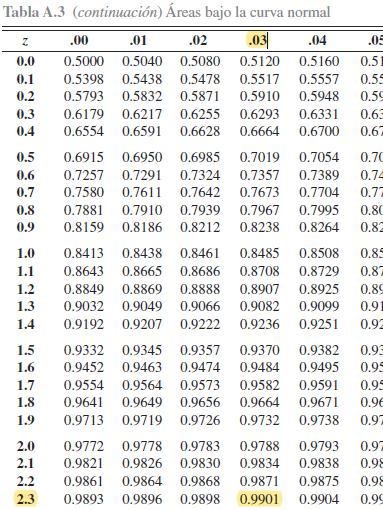
*Solución: 1-α = 0.95 entonces α = 0.05, α/2 = 0.025, zα/2 = 1.96*

En la tabla se encuentra que Z0.025=1,96 con 10 (N) o 30(N) de aquí, el intervalo de confianza de 95% para es:

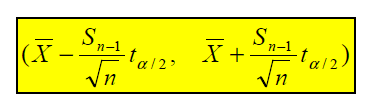


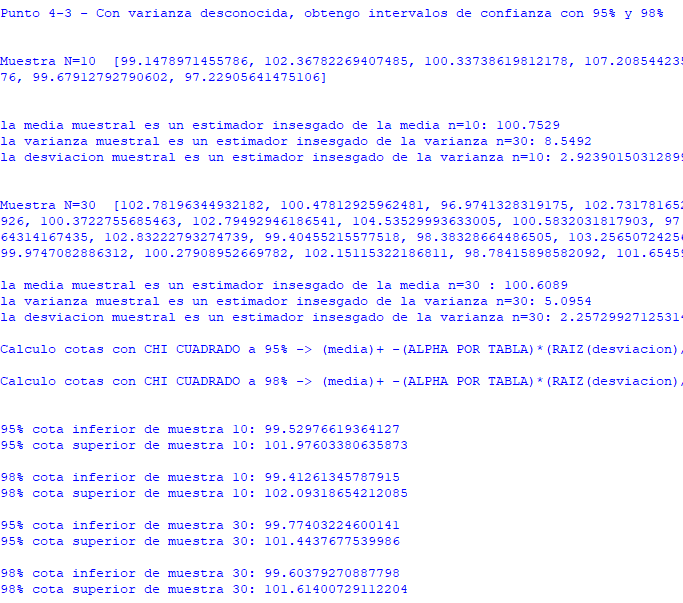


En la tabla se encuentra que Z0.01=2,33 con 10 (N) o 30(N) de aquí, el intervalo de confianza de 98% para es

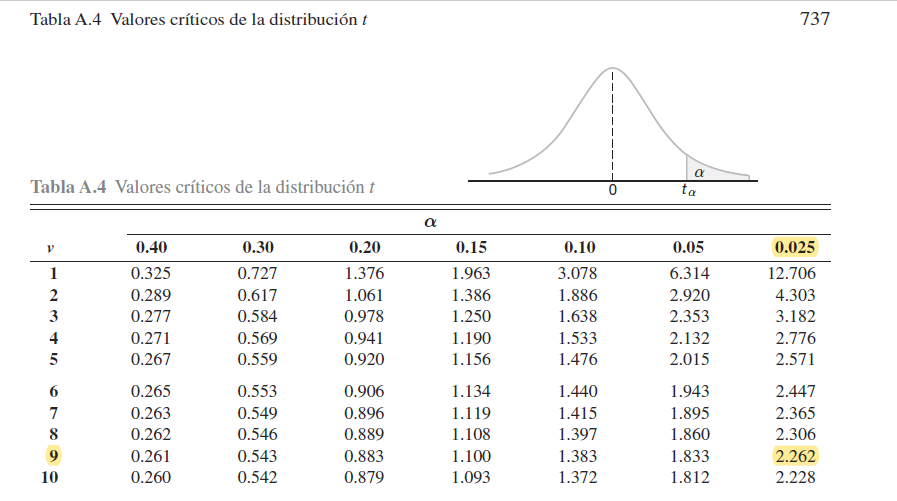


3. Repita el punto anterior, pero usando la varianza estimada *s*2, para la muestra de tamaño adecuado.

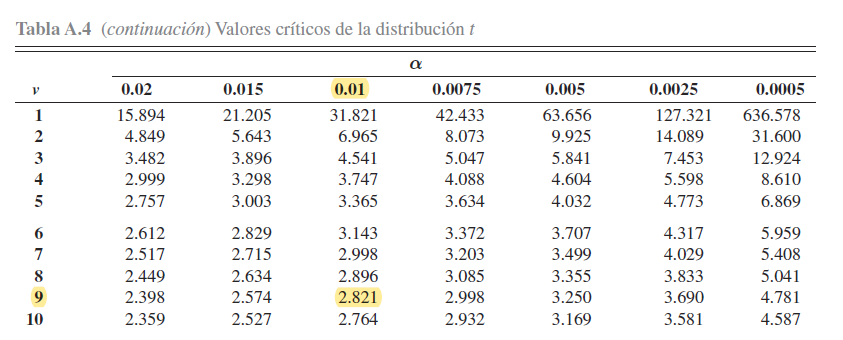




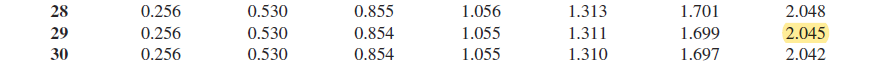
En la tabla se encuentra que t0.025=2.262 con 9 (N-1) grados de libertad, de aquí, el intervalo de confianza de 95% para es:



En la tabla se encuentra que t0.01=2.821 con 9 (N-1) grados de libertad, de aquí, el intervalo de confianza de 98% para es:



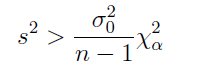
En la tabla se encuentra que t0.025=2.045 con 29 (N-1) grados de libertad, de aquí, el intervalo de confianza de 95% para es:



En la tabla se encuentra que t0.01=2.462 con 29 (N-1) grados de libertad, de aquí, el intervalo de confianza de 98% para es:



4. Probar a nivel 0,99 la hipótesis de que la varianza sea *VARIANZA* > 5. Calcular la probabilidad de cometer error tipo II para la hipótesis

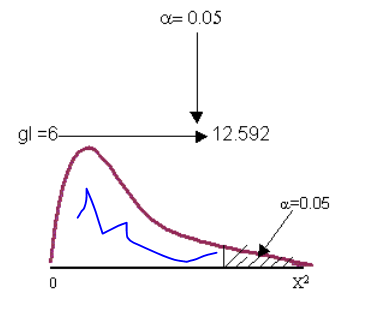
alternativa *VARIANZA* = 6.

Estadístico de prueba

Nosotros lo que debemos buscar es la varianza hipotética: 

A partir de la formula despejando la varianza hipotetica:

Primero debemos calcular CHI2 por table. Como pide 0.99 nivel de significancia, la cola serían 0,01. Por tabla da 49,58



Luego se calcula la varianza muestral, que se saca por función de la muestra N=30.

En nuestro caso da 2,74

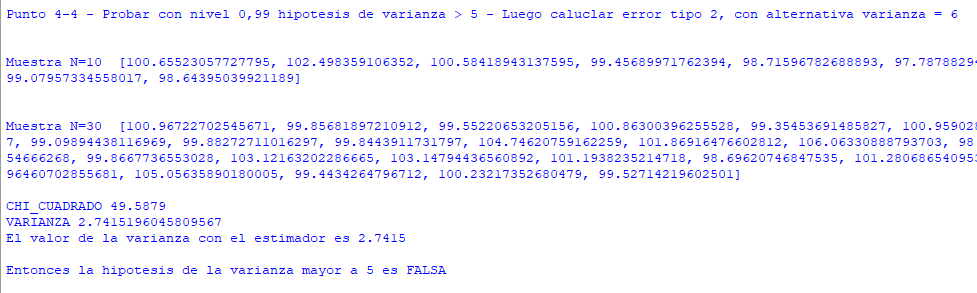
Ya con todos los datos, calculamos la varianza hipotética:

(Varianza\_muestral\*(N-1)) / CHI2 = NOS DA 2,74

Como el ejercicio pide que la varianza hipotética sea MAYOR a 5, comparamos ese resultado para ver si lo es:

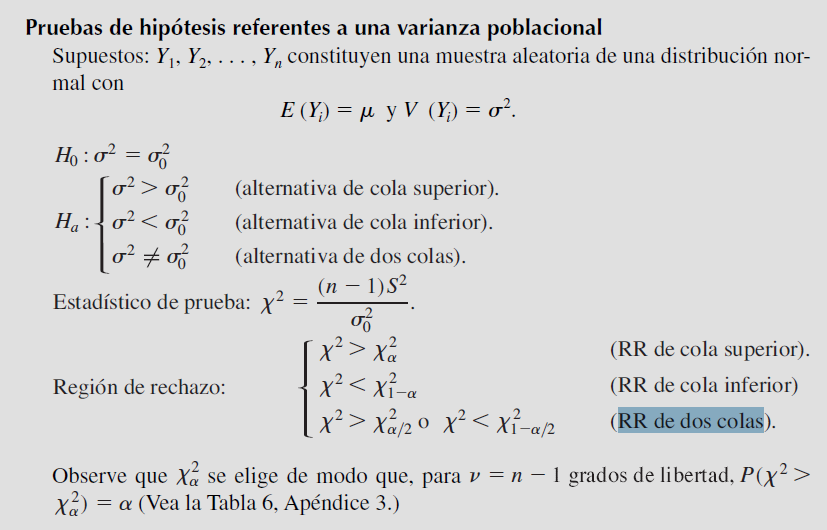
Si no cumple -> hipótesis FALSA

Si cumple -> Hipotesis VERDADERA



ERROR TIPO 2

Calcular la probabilidad de cometer error tipo II para la hipótesis

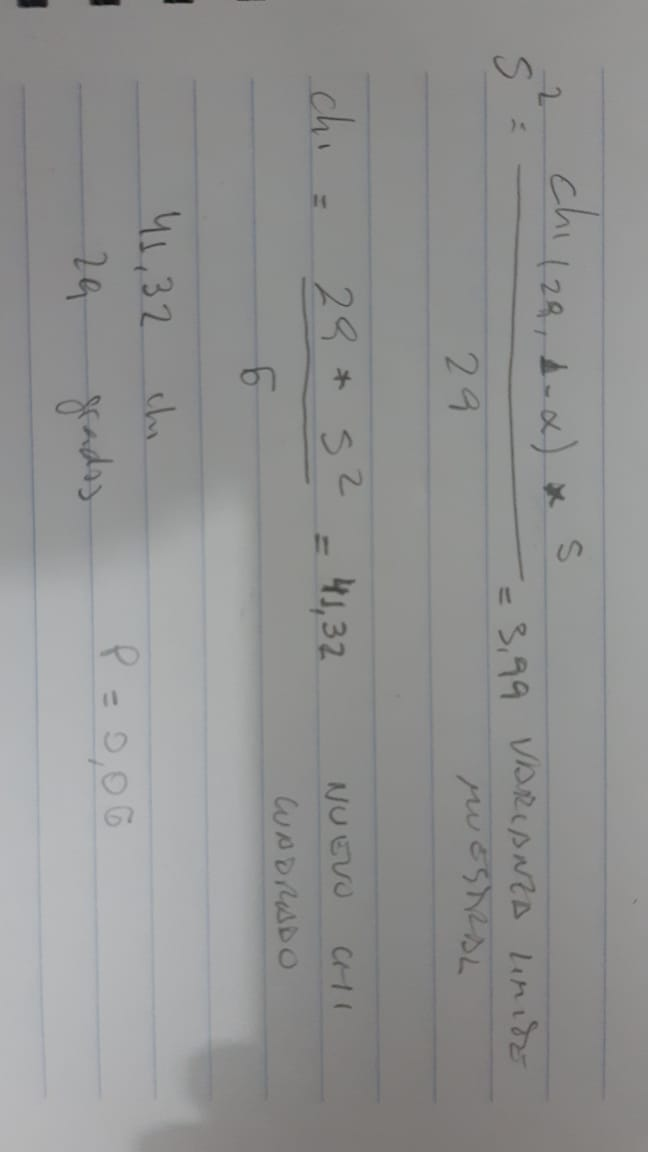


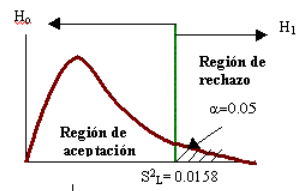
H0 (nula) => varianza = 6

H1(alternativa) => varianza < 6

H0 (nula) => varianza σ² ≠ 6-> decimos que es ≠ 6 -----rechazarla es = 6

H1(alternativa) => varianza σ² = 6-> error tipo 2 No rechazar la hipotesis nula cuando es falsa





Primero calculo la varianza muestral (que va a ser nuestra varianza limite) con la hipótesis nula (varianza hipotética = 5)

## S^2 = chi(0.99,n-1)\*(5->varianza hipotética) / (N-1)

Una vez que tengo mi nueva varianza limite, calculo cuanto debería ser mi chi2 suponiendo que tengo varianza hipotética = 6 (hipótesis alternativa).

chi(0.99,n-1)= S^2(lo que dio anterior \* (N-1) / (6 -> varianza hipotética de la alternativa)

Con esa chi2, la probabilidad la buscamos realizando la operación inversa en TABLA CHI2.

En nuestro la sacamos por una función que la arroja directamente.

Ese resultado (que vendría a ser el nivel de significancia) es la PROBABILIDAD!

5. Agrupando los datos en subgrupos de longitud 0,5, probar a nivel 0,99 la hipótesis de que la muestra proviene de una distribución

normal.

