UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS

DEPARTAMENTO DE FÍSICA

MÉTODOS NUMÉRICOS PARA LA CIENCIA E INGENIERÍA

TAREA 9

"Ajuste de curvas e Intervalos de confianza"

Braulio Sánchez Ibáñez

16.880.977-8

Profesor: Valentino González

Auxiliar: Felipe Pesce

Santiago, Chile

PREGUNTA 1

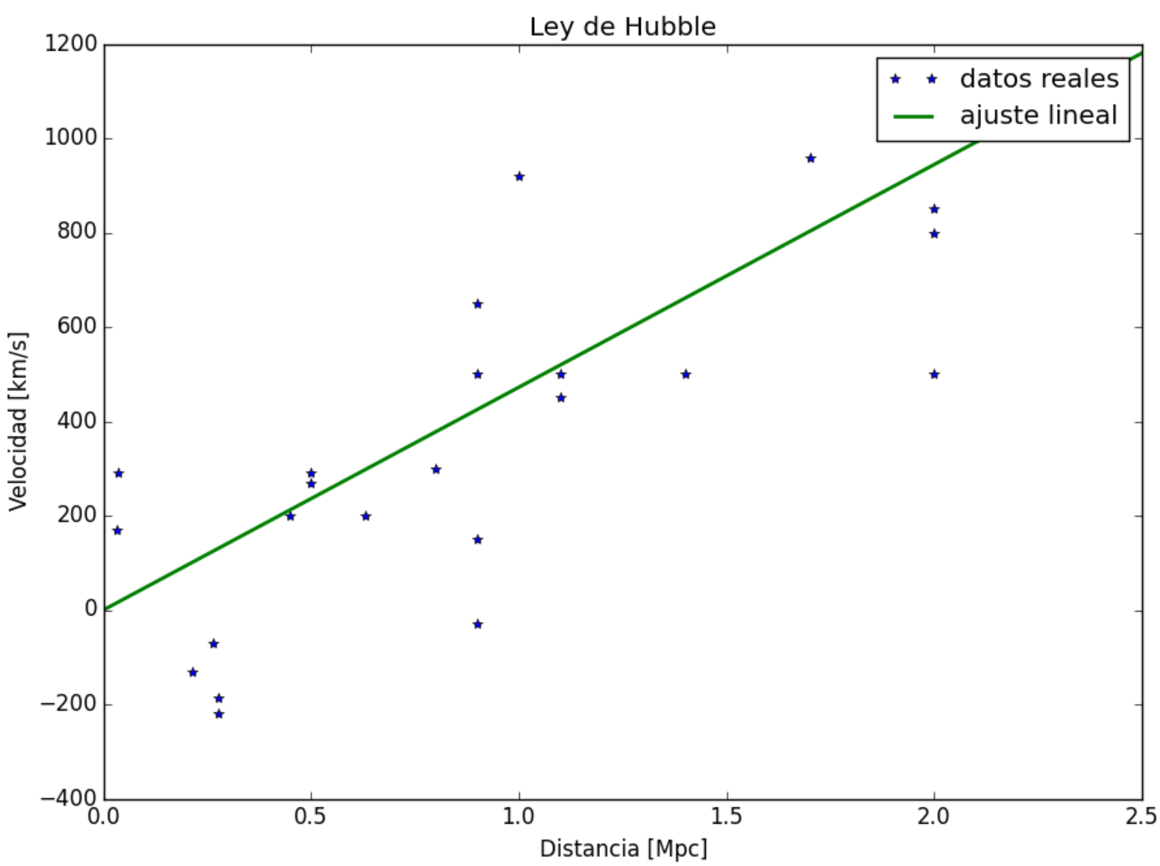
Se dispone de los datos tomados por Edwin Hubble en 1929 de la velocidad v de recesión de galaxias y su distancia d a la Tierra, que serán ajustados a las leyes $v = H_0 d$ y $d = \frac{1}{H_0} v$, donde H_0 es la constante de Hubble a determinar. Los dos ajustes dan valores distintos para H_0, por lo que el valor final será el promedio entre ambos.

Los datos se encuentran en el archivo **`Hubble.txt`**, y el ajuste se hace mediante el módulo **`scipy.optimize.curve_fit`**, que arroja el resultado

$$H_0 = 472.141 \left[\frac{km/s}{Mpc}\right]$$

El siguiente gráfico muestra los datos tomados por Hubble y la línea recta de ajuste, cuya pendiente es la constante de Hubble recién calculada.

Figura 1: Ley de Hubble con datos de 1929



El valor calculado para H_0 es bastante cercano al que estimó Hubble aquella vez, que fue un valor cercano a 500 [km/s/Mpc].

Para determinar el intervalo de confianza se realizará una simulación Bootstrap, que consiste en seleccionar una muestra aleatoria (del mismo tamaño de la base de datos original) de los datos tomados por Hubble y estimar con ella la constante de Hubble con el método usado anteriormente. Se itera este procedimiento un número N=100.000 veces y se obtiene una lista con N valores para H_0. Esta lista se ordena de menor a mayor y se calcula el intervalo de valores de H_0 que contiene al 95% de los datos, es decir, desde el valor ubicado en la posición 0.025N hasta 0.975N.

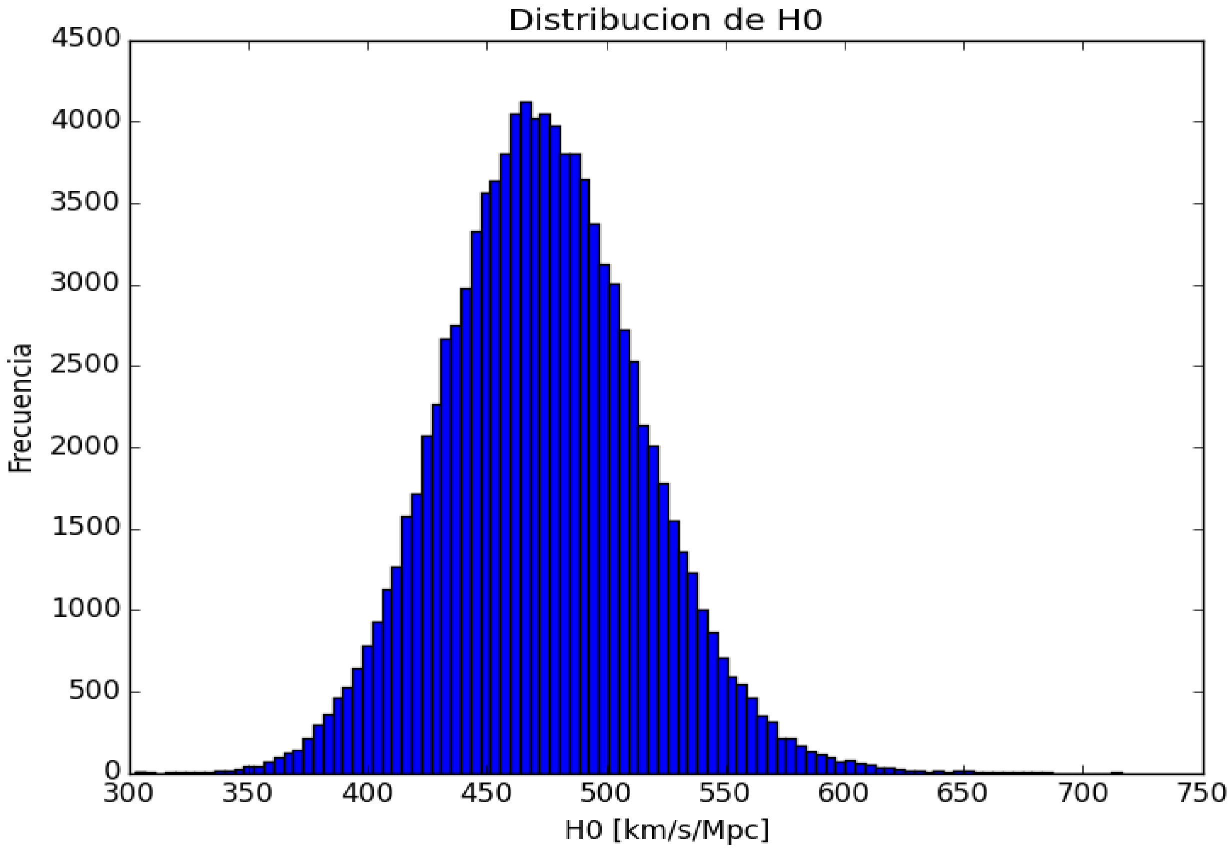
El resultado de este procedimiento arroja el siguiente intervalo de confianza

$$I = [394.697, \quad 559.310]$$

El intervalo es amplio, dando una gran incertidumbre para el valor de la constante, debido al escaso número de datos tomados y a su elevada dispersión en torno al ajuste.

El siguiente gráfico muestra la distribución de los datos generados con los cuales se calculó el intervalo de confianza.

Figura 2: Histograma muestra de H0



<u>PREGUNTA 2</u>

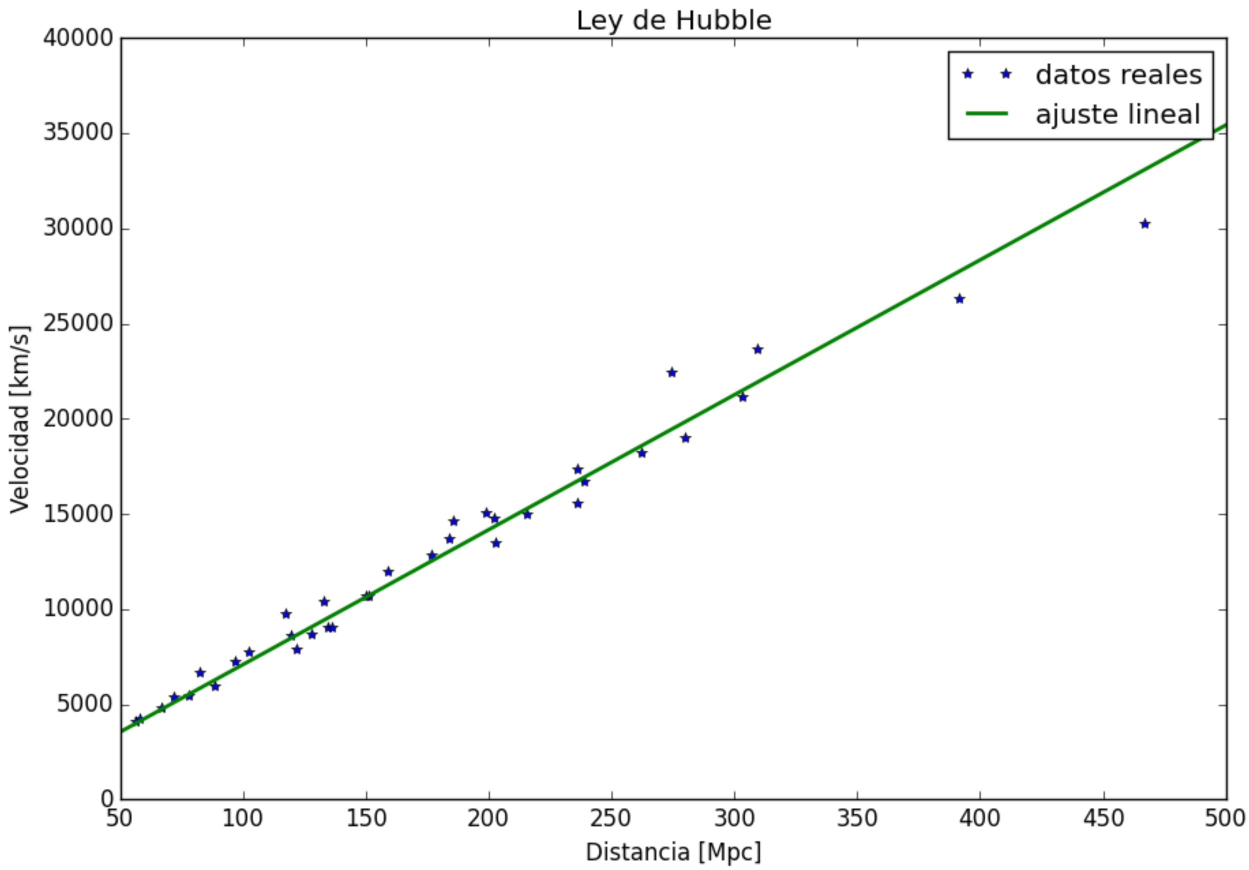
Se repite el mismo procedimiento anterior, pero para los datos tomados en el año 2000 por Friedman, que ahora calcula la distancia a las galaxias utilizando Super Novas tipo I. El valor de la constante de Hubble estimado es

$$H_0 = 70.841 \left[\frac{km/s}{Mpc}\right]$$

Este valor es mucho más cercano al aceptado actualmente, que es aproximadamente 67 [km/s/Mpc]

El siguiente gráfico muestra los datos recopilados por Freedman y la línea de ajuste.

Figura 3: Ley de Hubble con datos del 2000



Se aprecia una mayor cantidad de datos en comparación a los tomados por Hubble, lo que permite mejorar la estimación de la constante.

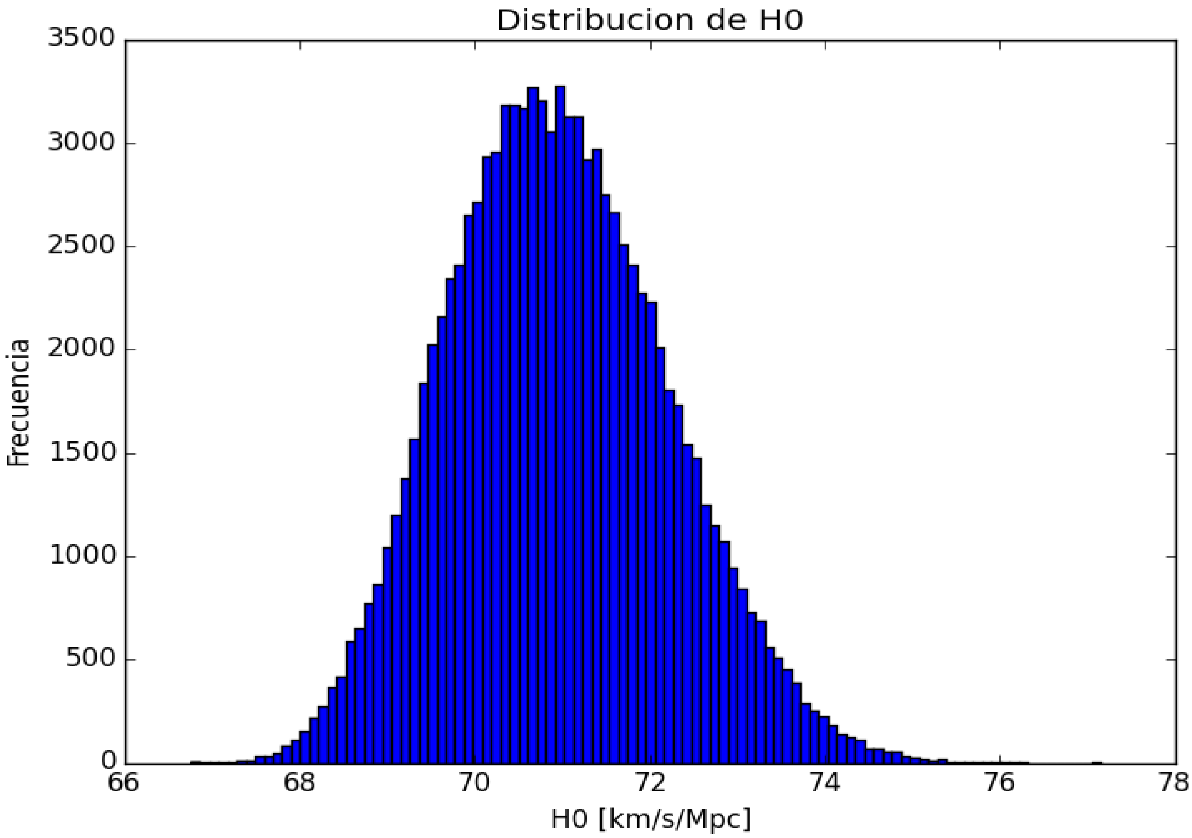
El intervalo de confianza calculado mediante Bootstrap es

$$I = [68.669,\ \ 73.546]$$

Este intervalo es mucho más acotado, aumentando la certeza del resultado.

El histograma con la distribución de los datos se presenta a continuación

Figura 4: Distribución de la muestra de H0



PREGUNTA 3

Se hará un ajuste lineal a la columna 81 y 83 de los datos del archivo **DR9Q.dat,** que corresponden a los flujos de las bandas i y z del catálogo de cuásares Data Release 9 del Sloan Digital Sky Survey. Los datos vienen en unidades *nmaggies,* que corresponden a 10^{-6} Jansky [Jy], donde 1 [Jy] equivale, en el sistema internacional, a $10^{-26}\left[\frac{W}{m^2}/Hz\right]$.

Se multiplicarán todos los valores por el factor 3.631, de modo que queden expresados en unidades de $10^{-6}[Jy]$.

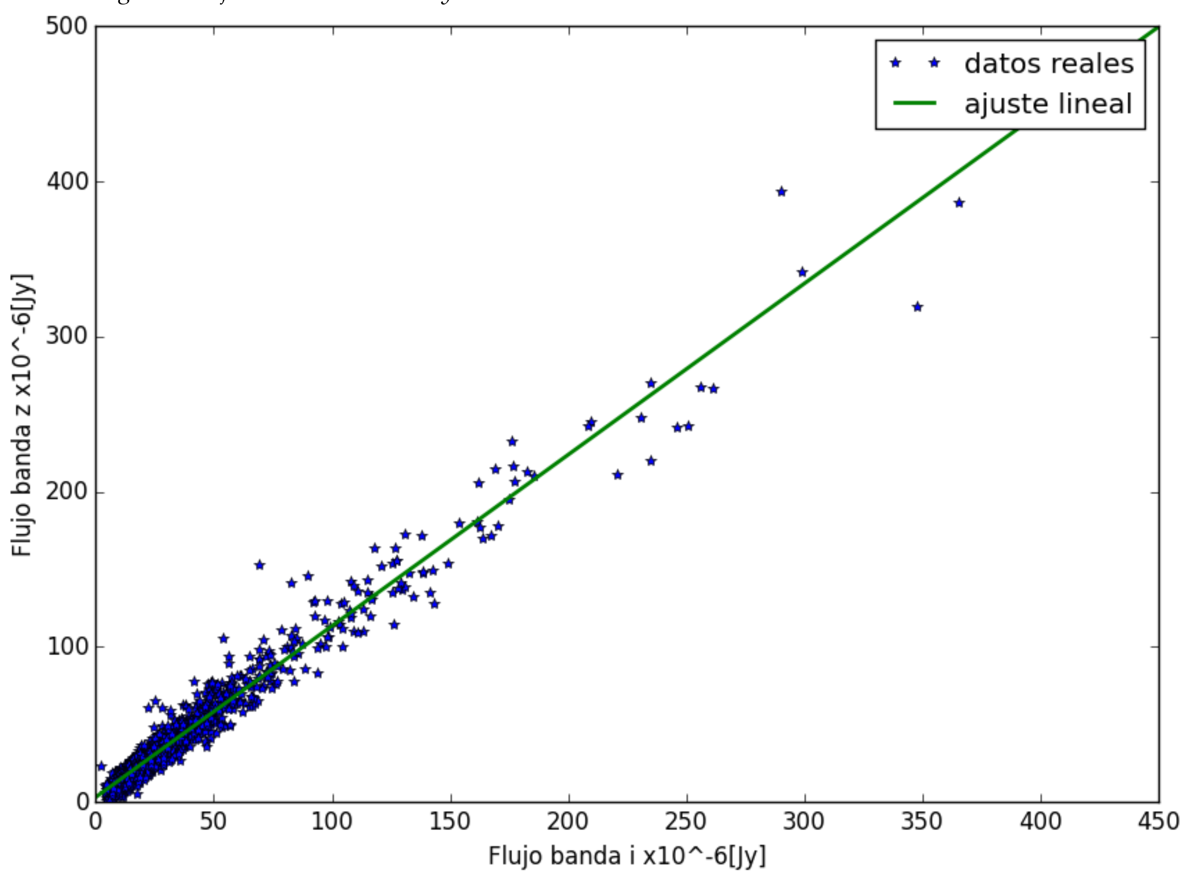
El ajuste se hace con el módulo **`scipy.optimize.curve_fit`** a orden 1, es decir, se ajustan los datos de los flujos a la curva $y = mx + n$, donde m y n son los coeficientes a determinar.

Los valores calculados para las constantes son

$$m = 1.103, \;\; n = 3.149$$

Por lo tanto, la relación es $F_z = 1.103F_i + 3.149$, donde F representa el flujo de la banda respectiva. El siguiente gráfico muestra los datos medidos y la línea de ajuste anterior.

Figura 5: Ajuste lineal bandas i y z



Los intervalos de confianza se calculan mediante una simulación de Montecarlo, donde, dados los errores asociados a una determinada medición, se define una nueva muestra de la forma

$$y_i = x_i + \varepsilon_i r$$

Donde y corresponde a la nueva muestra, x es el dato medido, ε es el error asociado y r es una variable aleatoria con distribución gaussiana. En este caso se deben generar muestras para los flujos de la banda i (columna 81) usando su error (columna 82) y para los flujos de la banda z (columna 83) con su respectivo error (columna 84). Sobre estas muestras se realiza un fiteo lineal y se generan valores para m y n. Se repite esto un número N=100.000 veces y se generan listas con 100.000 valores para m y n que permitirán calcular el intervalo de confianza como aquel que contenga el 95% de los valores.

Los intervalos de confianza para m y n son

$$I_m = [0.949, \ \ 1.142]$$

$$I_n = [2.268, \ \ 7.757\ ]$$

Las siguientes figuras representan los histogramas de las muestras generadas para m y n.

Figura 6: Histograma muestra m

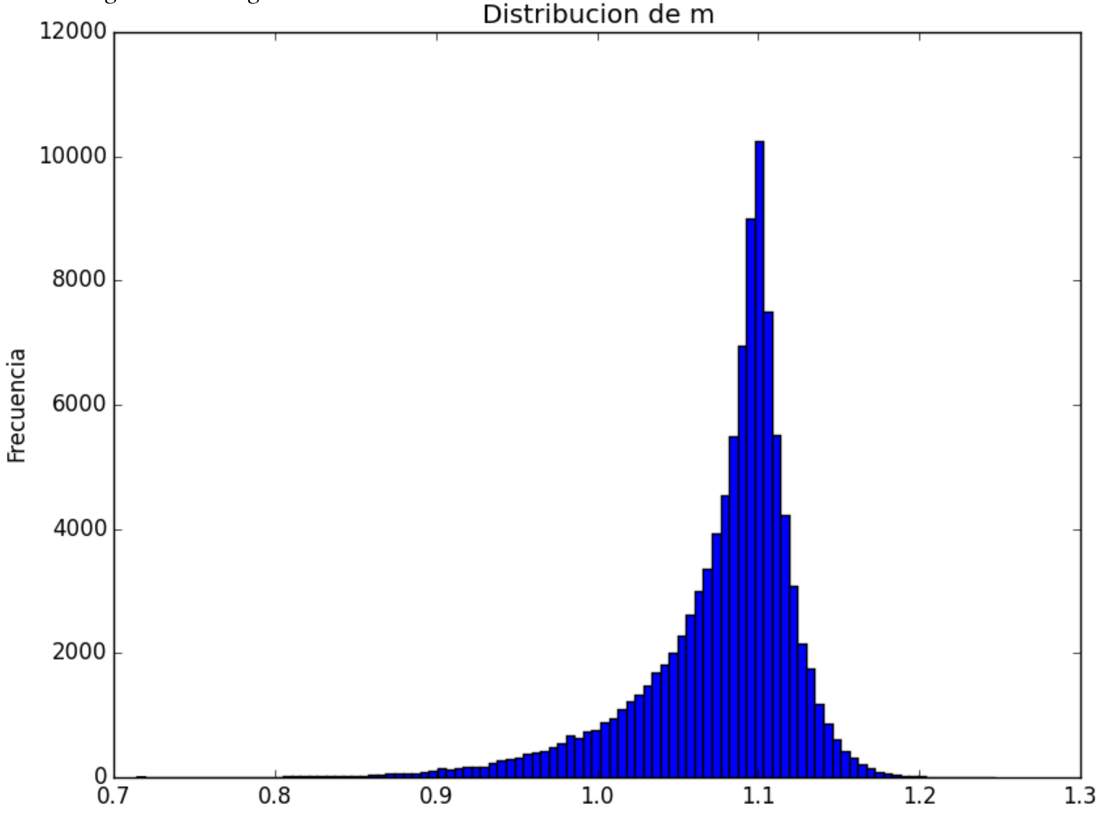


Figura 7: Histograma muestra n