

Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas

FI-3104-1 Métodos numéricos para Ciencia e Ingeniería

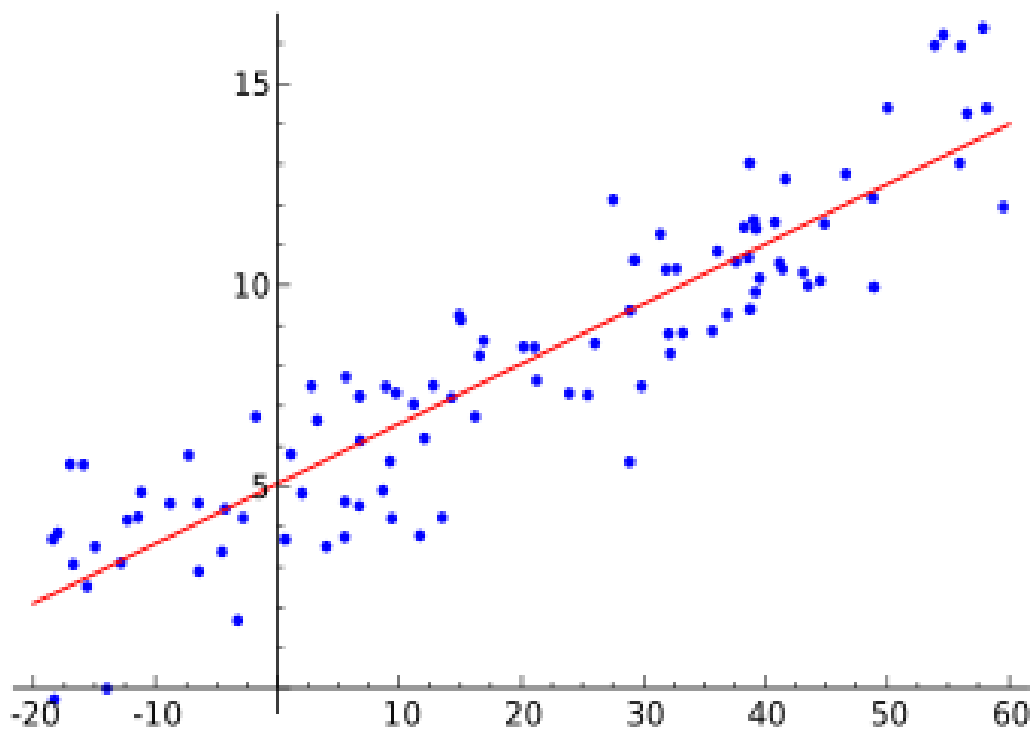
Tarea n°9

“Ajustes lineales fit para modelos científicos”

Benjamín Venegas Ríos

28-11-2015

Santiago, Chile



Resumen

En este trabajo se buscaba encontrar los mejores ajustes a distintos modelos científicos asociados a la expansión del universo, con el fin de poder determinar la constante Hubble H_0 , valor el cual fue muy discutido por varios científicos a lo largo del siglo XX. Para eso se usaron librerías de ipython, con funciones como `curve_fit` y `polyfit`, además de los Métodos de Bootstrap y Monte Carlo, de tal modo que las aproximaciones a los datos medidos, tanto por Hubble como por Freedman y para el caso de los flujos en distintas bandas se logaran ajustar de buena manera a lo experimental y así, poder deducir valores importantes para los parámetros físicos involucrados.

Pregunta 1

1.1 Introducción

En 1929 Edwin Hubble comparó la velocidad de recesión de las Nebulosas (la idea de galaxias lejanas era aún reciente así que se les llamaba nebulosas) con las distancias entre estas Nebulosas y la Tierra. Las distancias fueron medidas usando el método de las Cefeidas, que son estrellas de luminosidad variables cuyo período y luminosidad están fuertemente correlacionados. La llamada relación período–luminosidad había sido recientemente calibrada.

Hubble no fue el primero en calcular esta relación pero su trabajo fue de los más influyentes de la época, convenciendo al mundo de la expansión del Universo.

El archivo ubicado en data/hubble_original.dat contenía las mediciones originales que Hubble utilizó en este trabajo. Su modelo era bastante simple:

$$v=H_0*d \quad (1)$$

Donde H_0 es la constante de Hubble y, generalmente, se expresa en unidades de $\text{km} / \text{s} / \text{Mpc}$ (Mpc : Mega parsec, $3.086e24 \text{ cm}$).

En esta pregunta se pedía utilizar los datos originales para derivar la constante de Hubble incluyendo su intervalo de confianza al 95%.

1.2 Procedimiento

*Para hacer el ajuste lineal a las mediciones obtenidas por Edwin Hubble(archivo data/hubble_original.dat) se procedió a crear la función $v=H_0*d$ y $d=v/H_0$, de tal forma que se obtuviesen dos valores que se acercasen al valor esperado de H_0 . Sin embargo, para hacerlo más preciso, se promediaron y, así se obtuvo una buena estimación del valor de H_0 según las mediciones obtenidas por Hubble. Así, se pudo graficar la velocidad de recesión de las nebulosas versus la distancia para los datos de Hubble con un ajuste lineal determinado por un H_0 que se obtuvo promediando los H calculados mediante `curve_fit` de las funciones mencionadas previamente y el gráfico en forma de puntos (sin ajuste lineal) de lo que había medido Hubble (figura 1)*

Luego, dado que se buscaba obtener un intervalo de confianza de 95% para H_0 , es decir, que con probabilidad de 95% se encontrara a H_0 en un intervalo específico, se inició la iteración bajo el método de Bootstrap, el cual consistía

básicamente, en tomar una simple muestra de las mediciones de Hubble (del mismo tamaño de los datos originales de Hubble), y a partir de ellas, estimar el valor de H_0 . Se iteró $N=100.000$ veces para obtener finalmente un arreglo de 100.000 valores para H_0 . De esta manera, se lograría generar un histograma para H_0 y encontrar el intervalo de confianza 95%, simplemente tomando del arreglo de H_0 , los índices $N*0.025$ y $N*0.975$. Con esto en la figura 1, se agregó una zona de confianza para el ajuste lineal con líneas punteadas y además un gráfico de Histograma, en donde se muestra claramente la zona de confianza en un 95% en donde se encuentra H_0 con mayor probabilidad indicando su intervalo con líneas verticales.

En la figura 1 y 2 anexadas se muestran los valores obtenidos para H_0 y el intervalo de confianza.

Pregunta 2

2.1 Introducción

A pesar de lo influyente de su trabajo, Hubble cometió un gran error en su estimación de H_0 . El error deriva de haber utilizado una calibración equivocada de la relación período–luminosidad (entre otras cosas).

Una estimación más reciente de la constante de Hubble se obtiene con los datos ubicados en el archivo: data/SNIa.dat (Freedman et al. 2000) que utiliza Super Novas tipo I para estimar distancias para una muestra de galaxias. Entre otras ventajas, el método permite estimar distancias muy superiores a las que se pueden medir con el método de las Cefeidas.

En esta pregunta se pedía volver a estimar la constante de Hubble para estos datos, incluyendo su intervalo de confianza del 95% y comentar.

2.2 Procedimiento

Análogo a la pregunta 1, sólo que cambiando de archivo a data/SNIa.dat(Freedman et al. 2000).

En la figura 3 y 4 anexadas se muestran los valores obtenidos para H_0 y el intervalo de confianza.

Pregunta 3

3.1 Introducción

Un Cuásar es una fuente astronómica de energía electromagnética, que incluye radiofrecuencias y luz visible. Se caracterizan esencialmente porque pueden fácilmente liberar energía a niveles iguales que la combinación de cientos de galaxias medianas y, además, por encontrarse demasiado lejos como para ser vistos por telescopios pequeños a excepción del cuásar 3C273 con $z=0.16$.

El archivo data/DR9Q.dat es una sección recortada del catálogo de cuásares del Data Release 9 del Sloan Digital Sky Survey (SDSS).

Esta pregunta pedía encontrar la línea recta que mejor modela la relación entre el flujo en la banda i y la banda z, incluyendo los intervalos de confianza al 95%

para los parámetros de la línea recta.

3.2 Procedimiento

Para resolver esta parte, se abrió el archivo data/DR9Q.dat y recuperaron las columnas 80,81,82,83 para tener los flujos y errores de las respectivas bandas.

Aquí surgió un problema, ya que dado que se desconocía cual de las 2 variables era la independiente , se identificaron dos ecuaciones y las intersectamos, de tal forma que la ecuación final considerara este problema e hiciera simétrico el ajuste lineal.

A continuación se muestran las ecuaciones a analizar:

$$z(i)=a*i +b \quad (1)$$

$$i(z)=c*z +d \quad (2)$$

Además, intersectando estas dos ecuaciones para generar un ajuste simétrico y evitar el problema de la independencia de las variables se obtiene:

$$z_c=a*i_c + b$$

$$\text{donde } i_c=(bc+d)/(1-ac)$$

y como se buscaba una expresion de la forma $z-z_c=m(i-i_c)$ se calcula que las

pendientes de las rectas son a y $1/c$ (ocupando las expresiones 1 y 2), y por definición de pendiente y generando nuevamente un estimación se tiene que:

$$m = \tan(\arctg(a) + \arctg(1/c)) / 2$$

y, además:

$$z(i) = m \cdot i + z_c - m \cdot i_c \quad (3)$$

que será la ecuación del ajuste simétrico lineal realizado, donde m será la pendiente y $(z_c - m \cdot i_c = p)$ será el coeficiente de posición de la recta.

De este modo, teniendo la ecuación de la recta se creó una función que retornara justamente los valores m (pendiente y p (coeficiente de posición). La función tenía como argumento los datos de Freedman y aplicaba polyfit a las 2 funciones (1) y (2) y definía los coeficientes de las ecuaciones (a, b, c, d) con tal formar la pendiente y el coeficiente de posición.

Así, teniendo el ajuste lineal para los datos de Freedman, con m y p , y las mediciones con sus errores, se procedió a graficar en un mismo gráfico estos dos plots.

Además, se requería buscar intervalos de confianza para los parámetros de la ecuación de la recta mencionada en 3, m y p . Para hacer esta parte, en primer lugar se recurrió al Método de Monte Carlo, con un total de $N=10000$ simulaciones y se crearon dos arreglos por los 2 parámetros a determinar (m y n) gracias a la función que retornaba m y p mencionada en el párrafo anterior, los

cuales eran de tamaño N con tal de poder generar un histograma y encontrar el intervalo de confianza para 95% al igual que la parte 1. Así con esta última parte se añade en la figura x la zona de confianza para el ajuste lineal de los flujos, y además, 2 histogramas, uno por cada parámetro, indicando el valor más probable para cada uno de ellos.

Nota: Las unidades de los flujo eran nmaggies, unidad inventada por SDSS (Sloan Digital Sky Survey) tal que $1 \text{ nmaggie} \sim 3.631 \times 10^{-6} \text{ Jy}$. Entonces, se transformó todo a unidades de 10^{-6} Jy multiplicando por 3.631. Además , $1 \text{ Jy} = 10^{-26} \text{ Watts/m}^2 \cdot \text{Hz}$,que justamente son unidades de flujo por unidad de frecuencia. La unidad Jy lleva el nombre del radio astrónomo pionero estadounidense Karl Guthe Jansky.

En las figuras 5, 6 y 7 anexadas se muestran los valores obtenidos para m y p , además del ajuste lineal para la función de flujos.

Resultados

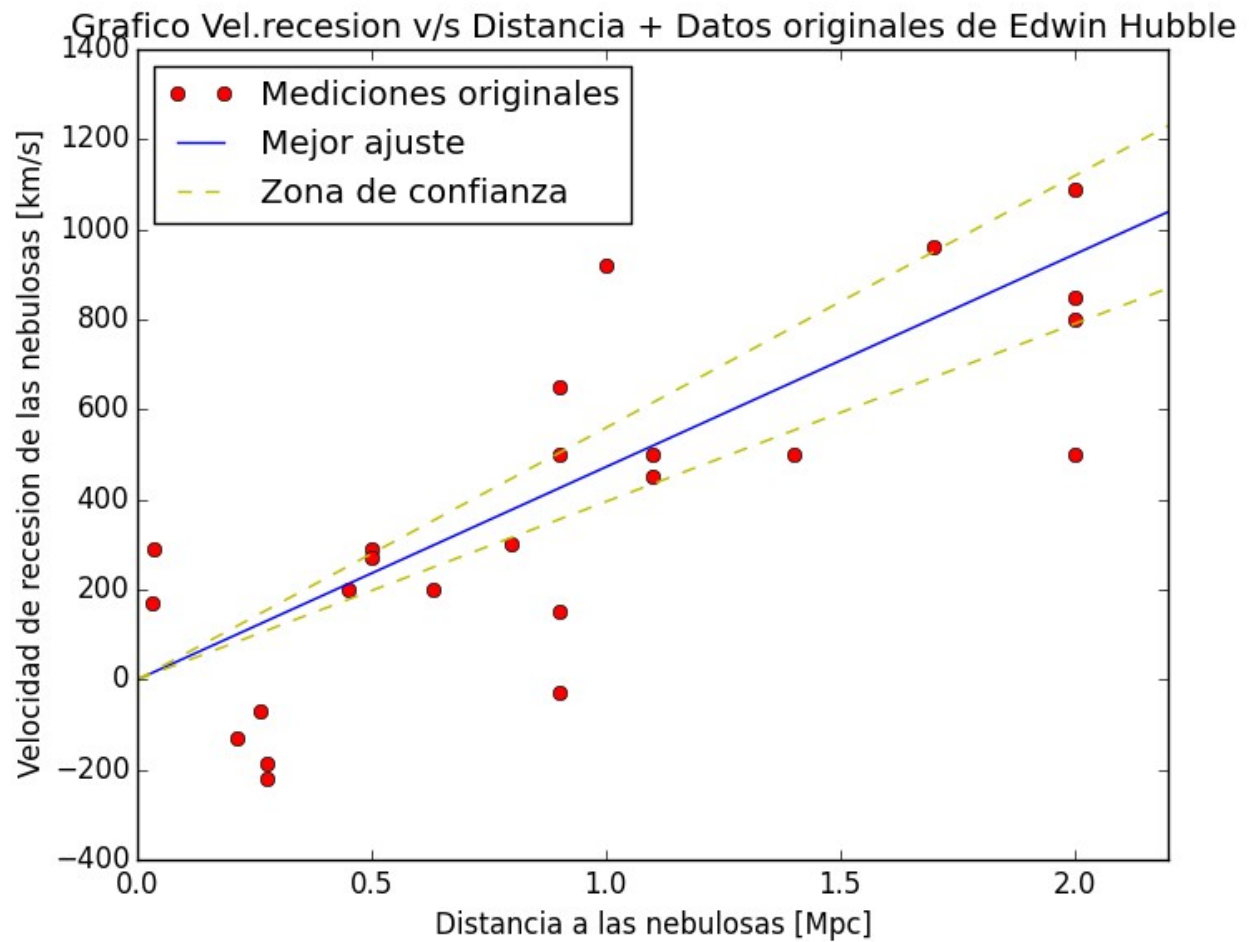


figura1. Gráfico de velocidad de recesión v/s la distancia a las nebulosas usando las mediciones de Hubble y añadiendo el ajuste lineal junto a la zona de confianza de 95% de H_0 . Se obtuvo un valor para H_0 de **472.14 km/s/Mpc**.

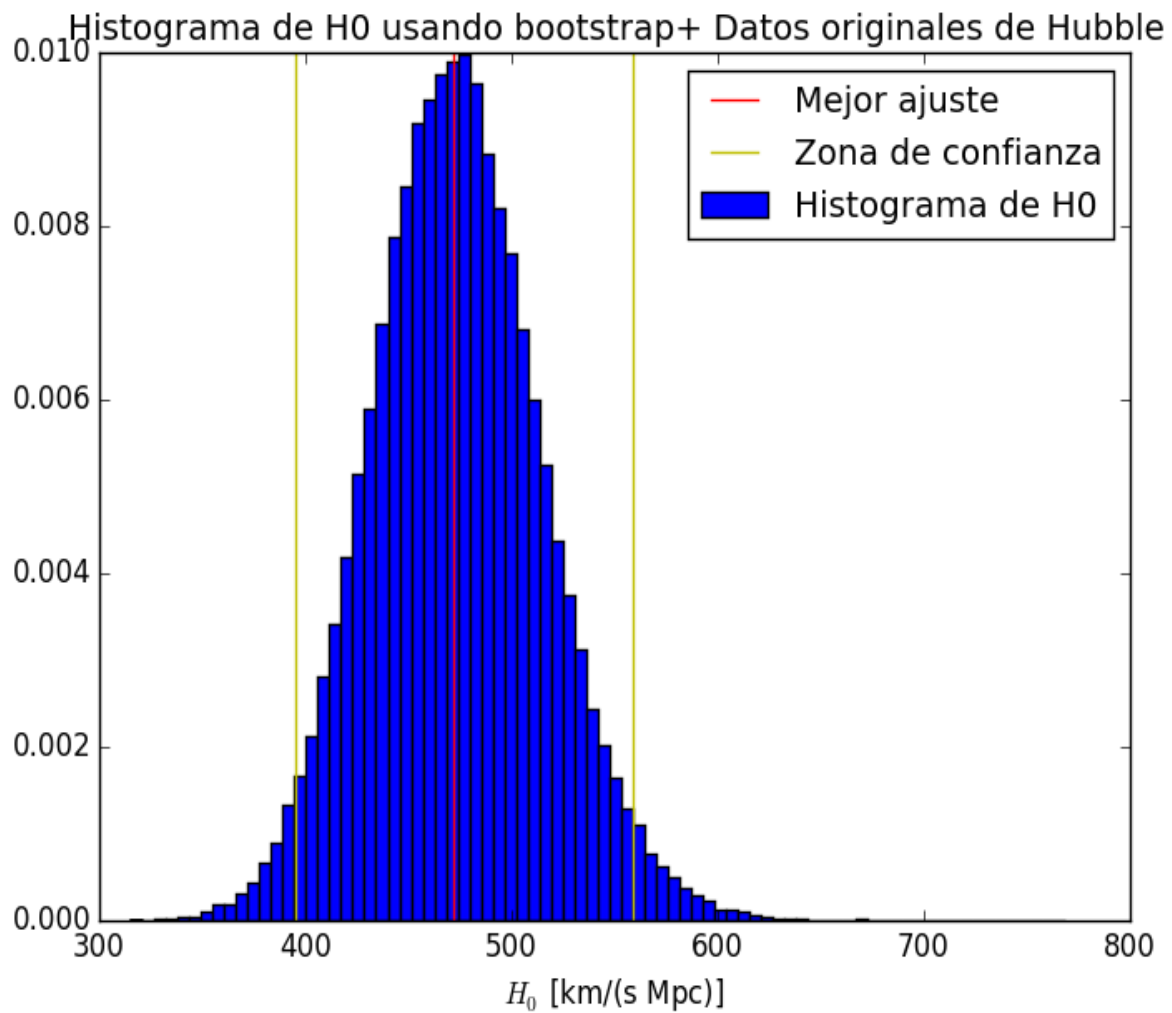


figura2. Histograma de H_0 ocupando los datos de Hubble. Se puede apreciar la zona de confianza para el valor de H_0 . El intervalo de confianza de 95% fue de **(395.08, 559.15) km/s/Mpc.** ($H_0 = 472.14$ km/s/Mpc)

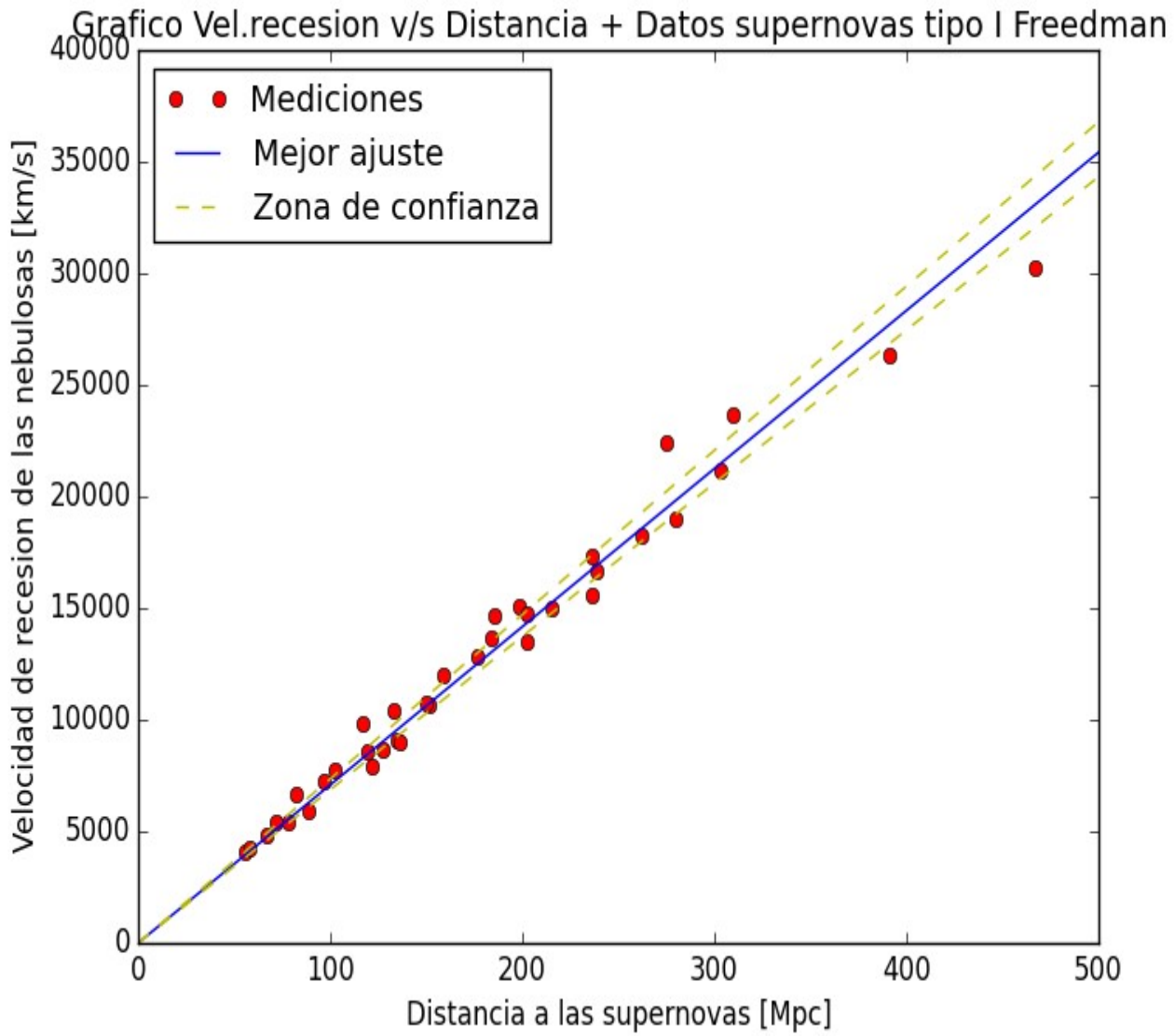


figura3. Gráfico de velocidad de recesión v/s la distancia a las nebulosas usando las mediciones de Freedman y añadiendo el ajuste lineal junto a la zona de confianza de 95% H_0 . Se obtuvo un valor para H_0 de **70.84 km/s/Mpc**.

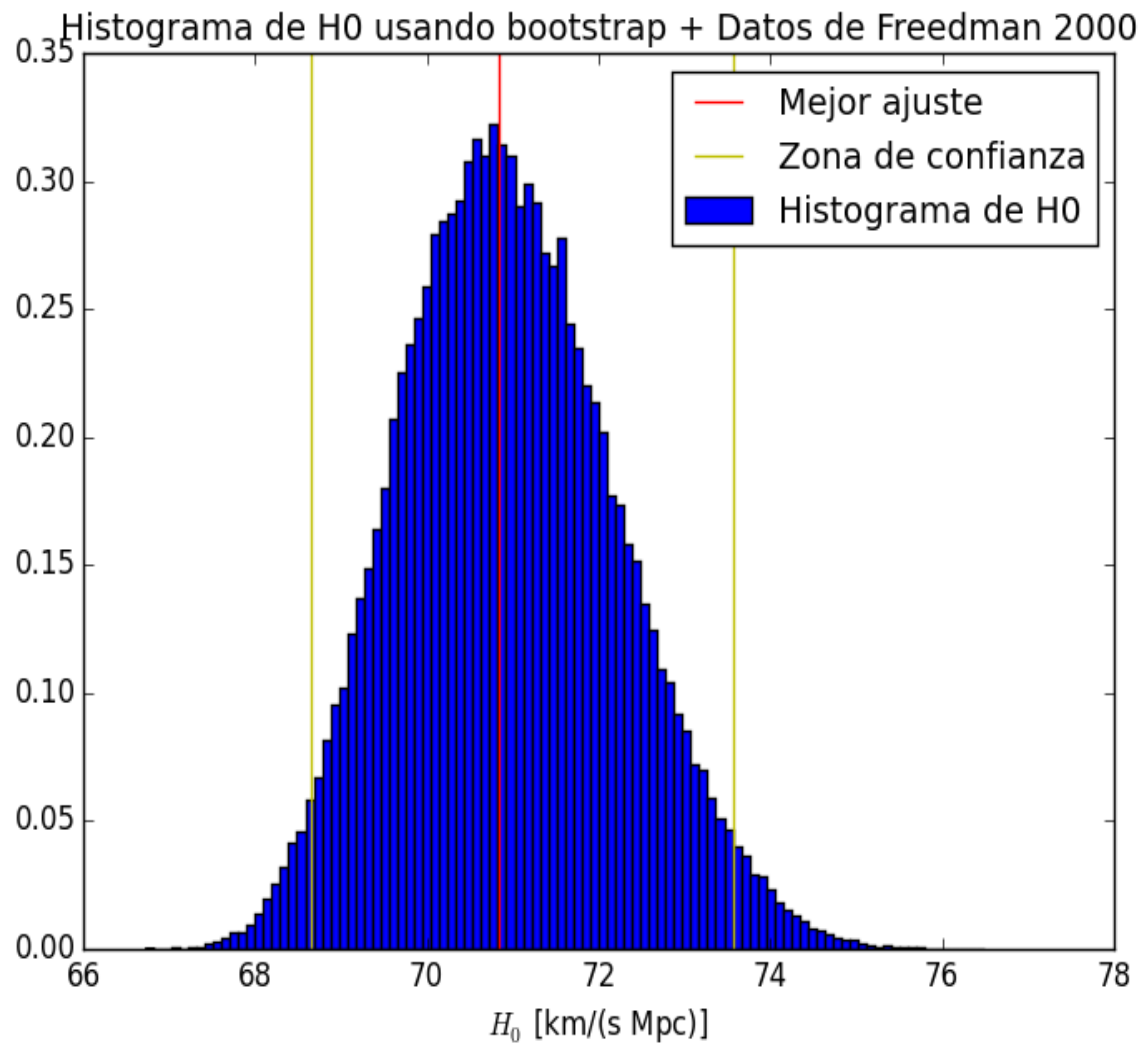


figura4. Histograma de H_0 ocupando los datos de Freedman . Se puede apreciar la zona de confianza para el valor de H_0 . El intervalo de confianza de 95% fue $(68.68, 73.55) \text{ km/s/Mpc}$. ($H_0 = 70.84 \text{ km/s/Mpc}$)

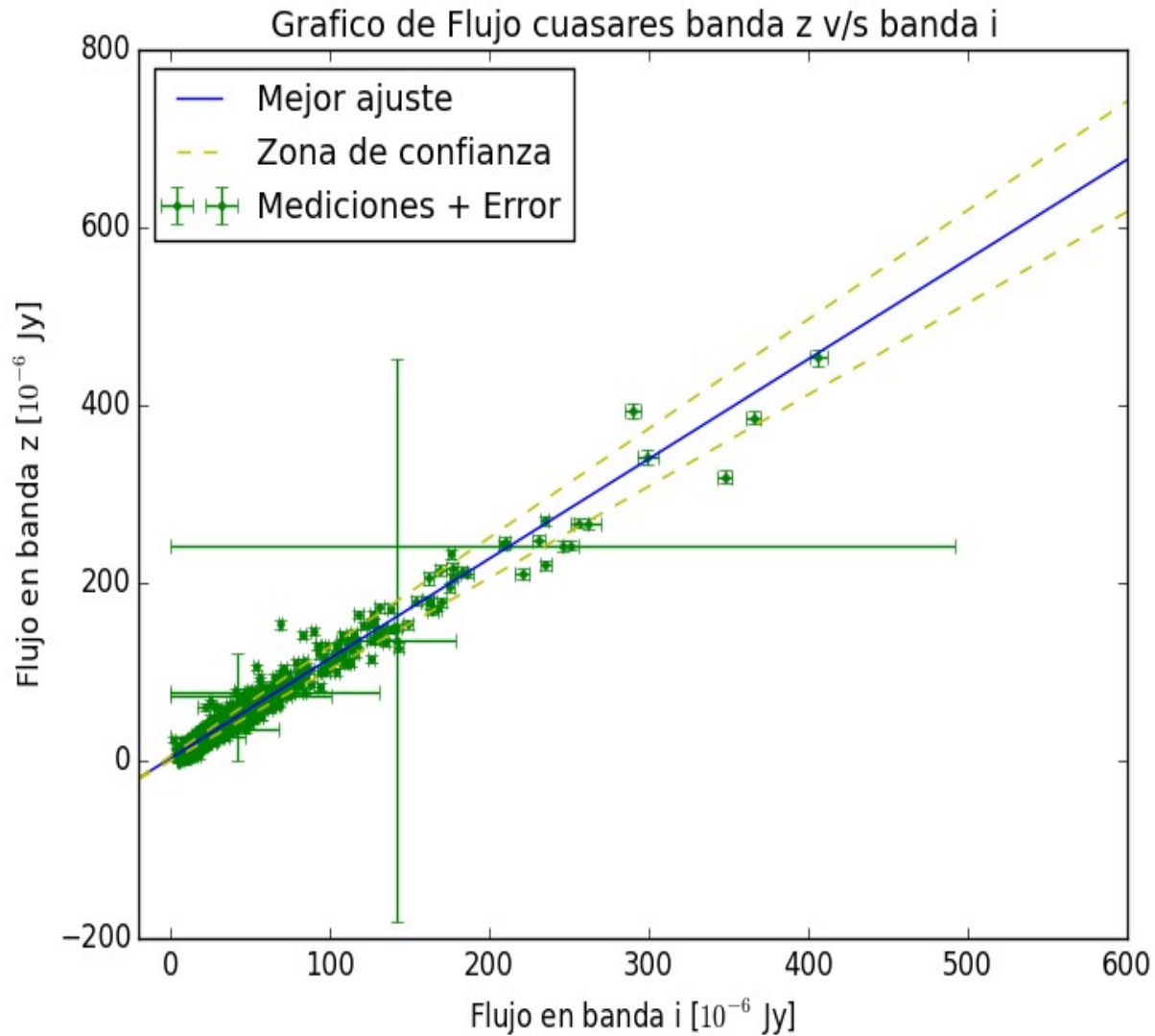


figura5. Gráfico de flujo banda Z v/s flujo banda I de los cuásares usando las mediciones de data/DR9Q.dat y añadiendo el ajuste lineal junto a la zona de confianza de 95 % (ambos dados por m y p) . En el ajuste lineal se ocupó el valor calculado con los datos : $m=1.12$ y $p= 2.50[10^{-6} \text{ Jy}]$.

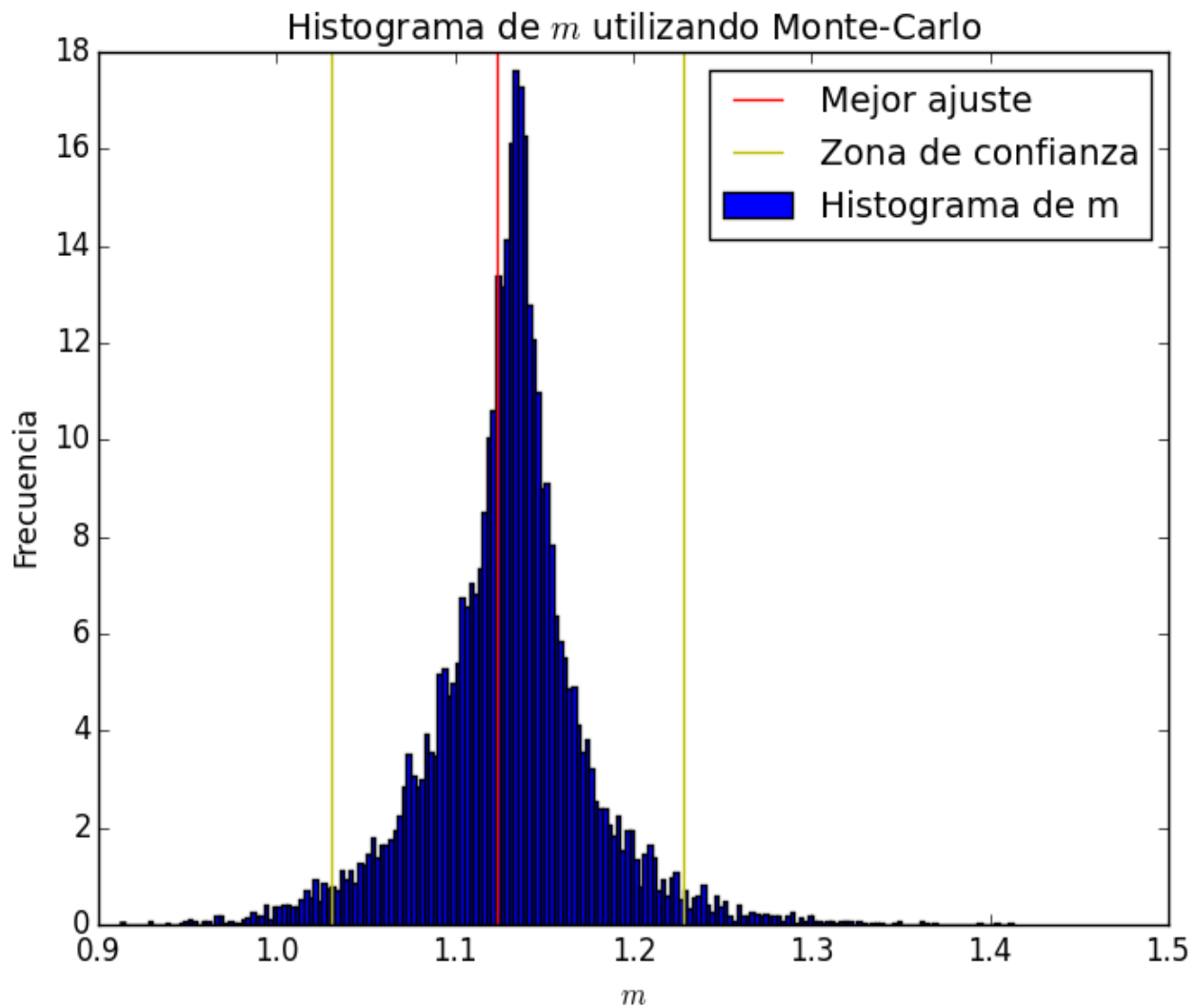


figura6. Histograma de m ocupando los datos de data/DR9Q.dat .Se puede apreciar la zona de confianza para el valor de m . El intervalo de confianza de 95% para m fue de (1.03, 1.23).(m=1.12)

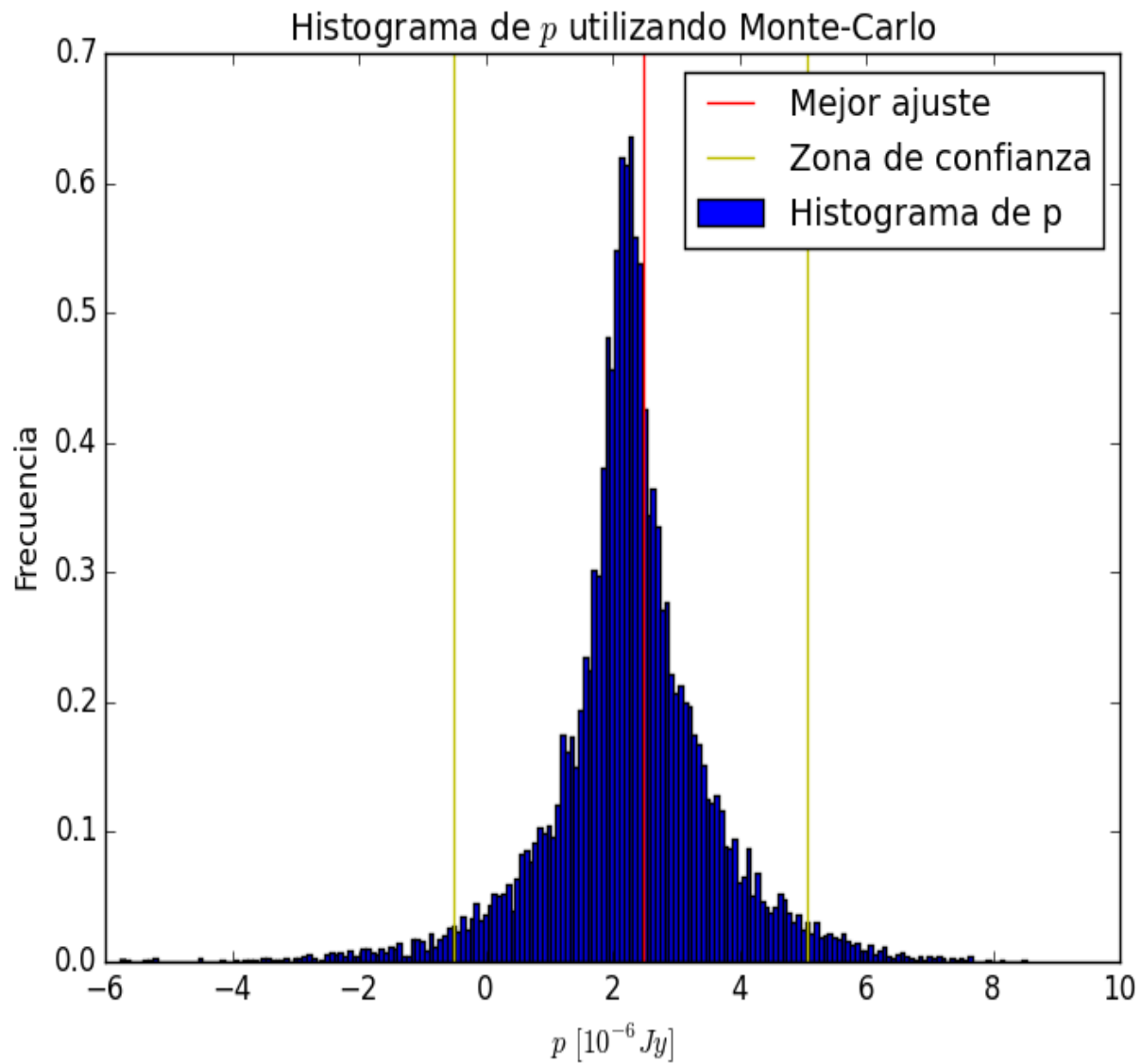


figura7. Histograma de p ocupando los datos de data/DR9Q.dat .Se puede apreciar la zona de confianza para el valor de p . El intervalo de confianza de 95% para p fue de $(-0.51, 5.06) [10^{-6} \text{ Jy}]$.($p=2.50[10^{-6} \text{ Jy}]$)

Conclusiones

Se puede concluir que en el primer caso, del astrónomo Edwin Hubble, la medida de H_0 calculada mediante el uso de `curve_fit` fue ajustada y precisa a los datos que había utilizado Hubble. No obstante, el valor de $H_0=472$ km/s/Mpc se aleja mucho del valor actual que se considera para la constante (unos 70 km/s/Mpc según WMAP el año 2006 y 68 km/s/Mpc según la misión espacial Planck del año 2014). Se afirma que el error de Hubble fue el haber medido mal las distancias en los años 20' por limitaciones instrumentales y el no haber considerado las velocidades peculiares entre las galaxias (llamadas “nebulosas” en ese entonces), las cuales causaban la gran dispersión de las medidas. Otro punto a recalcar es que Hubble consideró una muestra pequeña (En este caso se trabajó con 24 galaxias, pero él había trabajado con 46. De todas formas el error asociado era importante).

Pasando a la parte más matemática, tanto `curve_fit` como el Método de Bootstrap fueron métodos numéricos que sirvieron para ajustarse de gran manera a los datos obtenidos. De hecho, si se analiza el intervalo de confianza de 95 % para H_0 de Hubble, se puede observar que (395.08, 559.15) km/s/Mpc indica que con probabilidad bastante alta (de 95%) es posible que se encuentre H_0 entre esos valores. Lo cual, si se compara con el valor obtenido (472 km/s/Mpc), se puede ver que efectivamente era probable que se aproximara a ese valor. Pero, tal como se mencionó antes. Si bien la matemática es correcta, la muestra y los errores instrumentales mermaron las medidas de Hubble y dificultaron el poder obtener el valor real, actualmente conocido para la constante.

Para el caso de Freedman , con sus datos del año 2000 para las supernovas tipo I (objetos muy brillantes que se pueden estudiar fácilmente a mayores distancias que las consideradas por Hubble) y el haber considerado una muestra más grande, de unas 36 supernovas y menos dispersión de datos (despreciando a grandes distancias las velocidades peculiares), el ajuste lineal obtenido de igual forma que en la parte 1 se ve que es una muy buena aproximación dado que casi todos los datos están contenidos en la recta. Cuando se compara el valor de H_0 obtenido (70.84 km/s/Mpc) con el histograma y el intervalo de confianza (68.68, 73.55)km/s/Mpc, es evidente que al ser más acotado este intervalo que el calculado con los datos de Hubble es mucho más preciso para obtener el valor adecuado de H_0 . (con probabilidad de 95% se encuentra H_0 entre 68.8 y 70.84, eso es análogo, a decir que con 5% de probabilidad el valor de H_0 caería fuera de este rango, un suceso mucho menos probable!!!!)

De esta forma , el haber usado Bootstrap y curve_fit con los datos de Freedman aplican de igual forma que para el caso de Hubble, sin embargo, las mediciones de Freedman fueron mucho mejor, en el sentido de que no hubo tantos inconvenientes instrumentales como Hubble y la tecnología permitió obtener errores mucho más pequeños. Estas mediciones coinciden con lo que actualmente se sabe del Universo!! Que posee unos 14 mil millones de años!

Finalmente, ahora analizando la parte de los flujos de distintos Cuásares en bandas Z y bandas I, se puede ver que el ajuste lineal se aproxima con mucha cercanía a los datos medidos,es decir, el haber aplicado el ajuste para la simetría de las variables de flujo en banda z y en banda i, y el haber ocupado polyfit para calcular los coeficientes de las funciones mencionadas anteriormente, junto con la

pendiente y el coeficiente de posición de la recta fue una buena elección.

Ahora, para poder identificar la región de confianza de 95%, efectivamente se tenían valores en los histogramas para m de (1.03, 1.23) y para p de (-0.51, 5.06) [10^{-6} Jy], habiendo obtenido $m=1.12$ y $p=2.50$. Esto indica que el valor calculado por el ajuste lineal de los datos para m y p eran bastante probables según el Método de Monte Carlo. Además, al ver la zona de confianza en el gráfico de velocidad v/s distancia se puede deducir, que dado ese intervalo de confianza y el ajuste lineal, el modelo sigue teniendo validez matemática (a pesar de que en un par de datos los errores de los flujos se dispararon debido a que posiblemente la distancia a los cuásares era considerable y justo en ese instante su flujo bajó por recorrido de materia oscura).

Para finalizar, los Métodos de Monte Carlo, Bootstrap, curve-fit y polyfit, demostraron que tienen la capacidad de ajustarse mejor o no dependiendo del cierto grado de dispersión que pueda contener alguna medida. Pero, en general, despreciando este inconveniente, y considerando que hoy en día, la tecnología ha avanzado a pasos agigantados, la precisión en las mediciones y estos métodos son una herramienta computacional muy válida para deducir leyes y valores de parámetros, en este caso la constante de Hubble, pero más adelante quizá que otra teoría importante se podría comprobar.