

# Trabajo de Investigación sobre The Sloan Digital Sky Survey

Dan , Fabian

*Departamento de Física y Astronomía, Universidad de Valparaíso*

18 de diciembre de 2022

## Resumen

Se estudiaron los datos contenidos en la Data Release 5 de Sloan Digital Observatory Sky Survey, para los cuales se da una caracterización detallada de cada variable en cuestión, se incluyen datos estadísticos, más una descripción física de la variable en estudio, además se hicieron pruebas para encontrar correlaciones entre pares de variables y el estudio de la correlación entre el brillo en la banda cercana al infrarrojo (*i\_mag*) y brillo en la banda rojo profundo (*z\_mag*) de las cuales se encontró una relación la cual no puede ser relacionada con causa y efecto, pues son dos bandas que se encuentran muy cerca, finalmente se describió que permite conocer y que no este análisis, en los casos donde el análisis se volvió más profundo se cita a otros artículos donde se continúa.

## 1. Introducción

Se utilizó el Data release 5, de Sloan Digital Observatory Sky Survey (SDSS de aquí en adelante), data release anteriores se han utilizado para el estudio de supernovas tipo 1a gracias a su gran poder espectrométrico en varias bandas de color Sako et al., 2018, en el presente artículo el lector se encontrará primero con un análisis estadístico descriptivo sobre las variables incluidas en la data release 5 y seguidamente un análisis estadístico para poder estimar sobre la población entera de estrellas.

En este data release 5 utilizado existen, 77429 objetos entre los cuales son candidatos de quasars o clasificados por quasars, notar que no es una muestra estadística de quasars.

Para el análisis computacional el lector puede encontrar las referencias en el libro Newman, 2012 y el código explicado y comentado en Python dentro del github Trigo, s.f.

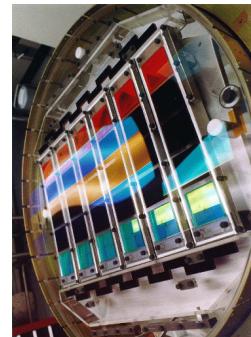


Figura 1: Cámara en SDSS,  
<https://www.sdss.org/wp-content/uploads/2014/11/faceplat2.jpg>

### 1.1. Sloan Data Base

La descripción técnica sobre estos datos se realizaron inspiradas en el paper original York et al., 2000 y Fukugita et al., 1996.

El telescopio del que se generan los datos es un telescopio de 2.5 metros, posee como mucho un error en su lectura de magnitud de  $\pm 0.02\ mag$  (de acuerdo a Fukugita et al., 1996)

El sistema se compone de 5 bandas de color desde 300 nanómetros (nm) a 1100 nm, siendo los límites capaces de medirse entre el ultravioleta que permite la atmósfera y la sensibilidad límite del silicio respectivamente. La siguiente imagen muestra la curva de los filtros:

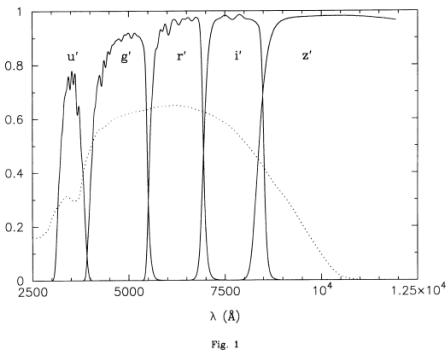


Figura 2: Curvas de cada filtro en amstrongs, se muestra la sensibilidad de cada filtro a distintas longitudes de ondas, imagen del paper Fukugita et al., 1996

el telescopio es un pipeline de datos, que toma el flujo monocromático  $f_\nu$ , el cual tiene unidades:

$$f_\nu \text{ ergs}^{-1} \text{ cm}^{-2} \text{ Hz}^{-1}$$

y entrega la magnitud

## 1.2. Magnitud

Es importante tener una sección exclusiva para explicar la magnitud y su background físico, pues todos los datos están expresados en magnitud, a esta magnitud registrada por el aparato se le llamará intercambiablemente 'brillo'.

La magnitud aparente de una estrella es la intensidad lumínica con la que se observa en el cielo nocturno, esta es función de la magnitud absoluta del objeto, la distancia y la extinción (extinción se refiere a la dispersión de la luz producto de gas entre el emisor y observador, provocando disminución de esta). Mientras que la magnitud absoluta mide el brillo a una distancia fija, registrando un parámetro intrínseco de la estrella, esta es la luminosidad de una estrella; tengase en cuenta que la magnitud absoluta es una medición indirecta, pues se convierte de magnitud aparente a absoluta mediante la distancia astronómica del objeto, cosa que a día de hoy resulta un problema en astronomía.

O sea que para este caso, estamos midiendo la magnitud aparente, pues es lo observado por el telescopio

La magnitud se define de manera que cada unidad de magnitud es 2.512 veces más brillante que la magnitud 1 siguiente. O sea magnitud -2 es 2.512 veces más brillante que magnitud -1 la magnitud de venus es -5, entonces será  $2.512^5 = 100$  veces más brillante que la magnitud 0 que es Vega (estrella)

### 1.2.1. Cálculo de Magnitud Clásico

La magnitud se calcula con

$$m_x = -2.5 \log_{10}\left(\frac{F_x}{F_{x,0}}\right)$$

Donde  $F_x$  es la irradiancia utilizando el filtro  $x$ , mientras que  $F_{x,0}$  es la irradiancia de referencia, el punto cero para ese filtro  $x$ .

### 1.2.2. Cálculo de Magnitud para SDSS

En lugar de usar el logaritmo, se utiliza el seno hiperbólico, el artículo Lupton et al., 1999 entra en detalle sobre la lógica de incorporar este sistema.

$$m = -2.5[\sinh^{-1}\left(\frac{F}{2b F_0}\right) + \ln b]$$

Donde  $b$  es el 'softening parameter', es relativa a  $F_0$  y no posee dimensiones, aquí una tabla para los distintos filtros de la data release 5, Sdss.org, s.f.-b:

Cuadro 1: Tabla de puntos de referencia para flujo 0 y softening parameters  $b$

Banda	$b$	flujo cero [ $m(f/f_0 = 0)$ ]
$u$	$1.4 \times 10^{-10}$	24.63
$g$	$0.9 \times 10^{-10}$	25.11
$r$	$1.2 \times 10^{-10}$	24.80
$i$	$1.8 \times 10^{-10}$	24.36
$z$	$7.4 \times 10^{-10}$	22.83

### 1.2.3. La estadística tras el softening parameter

De acuerdo al análisis de Lupton et al., 1999, primero se comparan dos funciones, donde  $x = F/F_0$ ,  $a = 2.5 \log_{10} 0e$

la magnitud aparente de una estrella

$$m(x) = -a \ln x$$

la magnitud de SDSS:

$$\mu(x) = -a[\sinh^{-1}\left(\frac{x}{2b}\right) + \ln b]$$

en los casos límites:

$$\lim_{x \rightarrow \infty} \mu(x) = -a \ln x = m(x)$$

$$\lim_{x \rightarrow 0} \mu(x) = -a\left[\frac{x}{2b} + \ln b\right]$$

La varianza, donde las flechas representan el comportamiento asintótico para  $x = 0$ , notar que  $V(x) = \sigma^2$

$$V(m) = \frac{a^2 \sigma^2}{x^2} \rightarrow \frac{a^2 \sigma^2}{x^2}$$

$$V(\mu) = \frac{a^2 \sigma^2}{x^2 + 4b^2} \rightarrow \frac{a^2 \sigma^2}{4b^2}$$

y la diferencia:

$$|m(x) - \mu(x)| = a \ln \left[ \frac{1 + \sqrt{1 + 4b^2/x^2}}{2} \right] \rightarrow a \ln \left( \frac{|x|}{b} \right)$$

mediante esta diferencia, se observa que disminuye mientras  $b$  es pequeño, sin embargo, la varianza  $V(\mu)$  es proporcional a  $1/4b^2$ , por tanto, aumenta con valores de  $b$  pequeños, existe así un equilibrio que debe de cumplirse, el algoritmo de búsqueda de este parámetro se encuentra fuera de los objetivos de este artículo, pero puede encontrarse en Lupton et al., 1999

#### 1.2.4. La Física y el Brillo

El brillo lo definimos como la magnitud aparente observada por el telescopio de SDSS, en la ecuación

$$m = -2.5 / \ln 10 * \left[ (2b/F_0) + \ln b \right]$$

$F$  es la energía radiante por unidad de superficie, es un flujo de esta energía electromagnética, lo que se presenta como el vector de Poynting  $\vec{S}$ :

$$\vec{S} = \mu_0 (\vec{E} \times \vec{B})$$

$$F \propto \vec{S} \cdot d$$

#### 1.2.5. Otros sistemas de Magnitud

El sistema AB la magnitud monocromática (sistema AB primero introducido por Oke, 1974) utilizando :

$$m_{AB} = -2.5 \log_{10} f_\nu - 48.60$$

popularmente utilizada, es posible pasar del sistema de magnitud de Sloan al AB.

### 1.3. Definición del Brillo

Cada variable indicada se refiere a la magnitud (indicada como brillo) en cierto filtro, esta posee un rango de sensibilidad que puede observarse en la figura 2

## 2. Análisis y Pipeline

### 2.1. Outliers

Utilizando el rango intercuartil se creó un intervalo para utilizar cada variable (ignorando a los datos relacionados a los outliers, los cuales simplemente habrían de seguir la misma pipeline pero como datos separados), el intervalo en cuestión que actuó como máscara:

$$[Q_2 - 1.5IQR; Q_2 + 1.5IQR] \quad (1)$$

se puede ver como actúa en distintas variables en los siguientes histogramas, el lector puede encontrar el notebook en la sección de referencias Trigo, s.f.

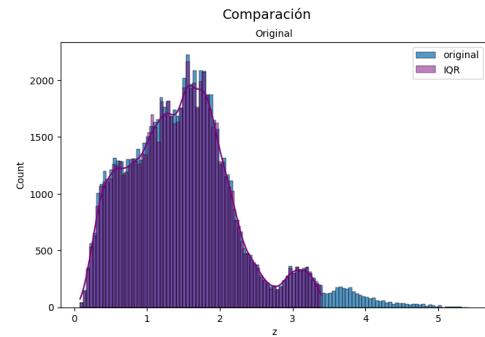


Figura 3: Comparación entre histograma de datos filtrados con la regla de outliers para variable  $z$

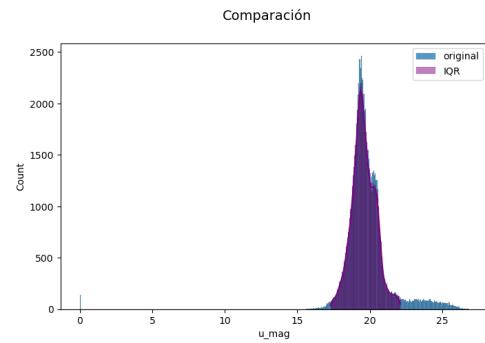


Figura 4: Comparación entre histograma de datos filtrados con la regla de outliers para variable  $u\_mag$

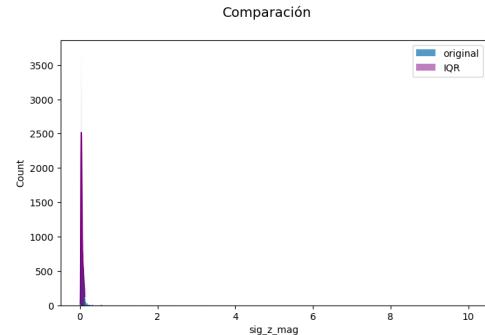


Figura 5: Comparación entre histograma de datos filtrados con la regla de outliers para variable  $sig\_z\_mag$ ; el tema con las desviaciones fue más evidente

ejecutando la limpieza en todas las variables, se construyeron los plots para cada variable:

## 2.2. Intervalos de Confianza

El intervalo de confianza es una manera de entregar una estimación puntual y un margen que estoy dispuesto a equivocarme, una confianza muy grande implica que ese intervalo es muy grande al punto de decir muy poco sobre la estimación.

La confianza utilizada en este informe corresponde a un 95 %, donde la confianza es  $\gamma$  y así definimos  $\alpha$ :

$$\gamma = 1 - \alpha$$

debido a que no se conoce la varianza poblacional, el intervalo de confianza para la media:

$$IC_{1-\alpha}(\mu) = [\hat{X} \mp t_{1-\alpha/2}(n-1) \frac{s}{\sqrt{n}}]$$

Debido a que se tiene una enorme cantidad de datos, la distribución t-student tiende a la distribución normal, por tanto, el intervalo de confianza para la media:

$$IC_{1-\alpha}(\mu) = [\hat{X} \mp Z_{1-\alpha/2} \frac{s}{\sqrt{n}}]$$

donde  $Z$  se refiere a la distribución normal estandar.

El intervalo de confianza para la varianza poblacional:

$$IC_{1-\alpha}(\sigma^2) = \left[ \frac{(n-1)S_{n-1}^2}{\chi^2_{1-\alpha/2}(n-1)}, \frac{(n-1)S_{n-1}^2}{\chi^2_{\alpha/2}(n-1)} \right]$$

Donde  $\chi^2$  se refiere a la distribución chi cuadrado.

## 2.3. Descripción de variables.

A continuación se presenta la descripción de cada variable almacenada en la base de datos junto con un resumen estadístico de esta y un histograma, donde se incluyen los intervalos de confianza para la media y varianza con un valor de significancia del 0.05 junto con los valores de: total de datos, promedio, media, moda, desviación estándar, valor mínimo, primer, segundo y tercer percentil, y valor máximo, en ese orden.

Téngase en cuenta que el análisis es incluyendo los outliers mientras que los graficos se eliminaron los outliers. De esta manera se maximiza visualización y objetividad.

### 2.3.1. z

El  $z$  se refiere al redshift, el cual le da a los astrónomos la posibilidad de calcular distancias, no confundir con  $z\_mag$  que viene mas adelante.

$IC(\mu)_{0,975}$	[ 1.364 , 1.3742 ]
$IC(\sigma^2)_{0,975}$	[ 0.5907 , 0.605 ]

count	mean	med	mode	std
53601	1.369	1.299	1.894	0.598
min	25 %	50 %	75 %	max
0.165	0.903	1.393	1.807	3.379

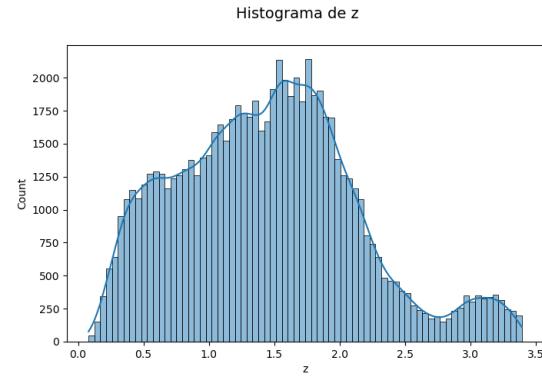


Figura 6: Histograma de datos filtrados con la regla de outliers para variable  $z$

### 2.3.2. u\_mag

Brillo en la banda  $u$  con centro en 365[nm] (ultravioleta) todo primer valor descrito proviene de Bessell, 2005, sin embargo en la pagina de Sloan para el DR5 Sdss.org, s.f.-a el valor difiere ligeramente: centro 354[nm] por lo que se incluiran los dos para transparencia al lector.

$IC(\mu)_{0,975}$	[ 19.4023 , 19.4125 ]
$IC(\sigma^2)_{0,975}$	[ 0.5967 , 0.6112 ]
count	mean
53601	19.407
min	25 %
17.605	19.016
med	50 %
	18.861
mode	75 %
	19.31
std	max
	0.604
	21.342

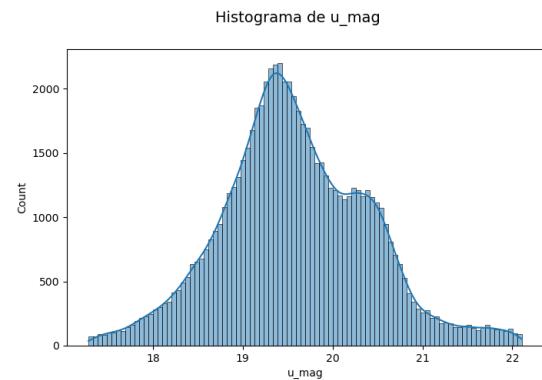


Figura 7: Histograma de datos filtrados con la regla de outliers para variable  $u\_mag$

### 2.3.3. sig\_u\_mag

Patrón de desviación para la banda ultravioleta, el equipo SDSS determina los errores de medición para cada magnitud a partir del conocimiento de las condiciones de observación y otras consideraciones técnicas.

$IC(\mu)_{0,975}$	[ 0.0352 , 0.0354 ]
$IC(\sigma^2)_{0,975}$	[ 0.0118 , 0.0121 ]

count	mean	med	mode	std
53601	0.035	0.028	0.029	0.012
min	25 %	50 %	75 %	max
0.012	0.027	0.033	0.041	0.08

count	mean	med	mode	std
53601	0.022	0.015	0.019	0.005
min	25 %	50 %	75 %	max
0.009	0.018	0.021	0.025	0.037

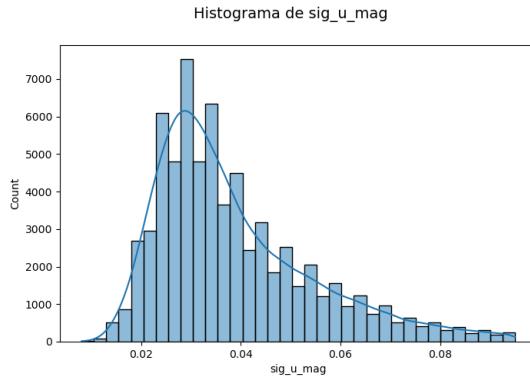


Figura 8: Histograma de datos filtrados con la regla de outliers para variable sig\_u\_mag

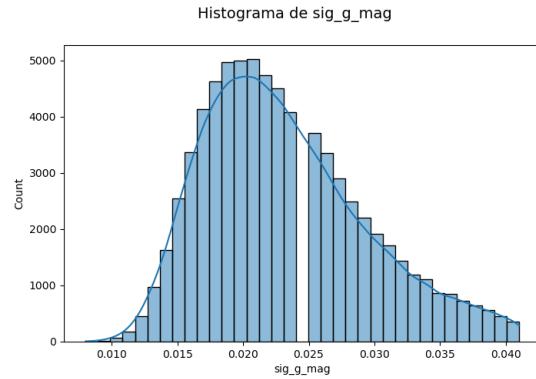


Figura 10: Histograma de datos filtrados con la regla de outliers para variable sig\_g\_mag

### 2.3.4. g\_mag

Brillo en la banda verde 464 [nm], de acuerdo Sdss.org, s.f.-a: 477[nm]

$IC(\mu)_{0,975}$	[ 19.1377 , 19.1472 ]			
$IC(\sigma^2)_{0,975}$	[ 0.5526 , 0.566 ]			
count	mean	med	mode	std
53601	19.142	18.763	19.33	0.559

count	mean	med	mode	std
53601	19.142	18.763	19.33	0.559
min	25 %	50 %	75 %	max
17.511	18.787	19.163	19.498	20.881

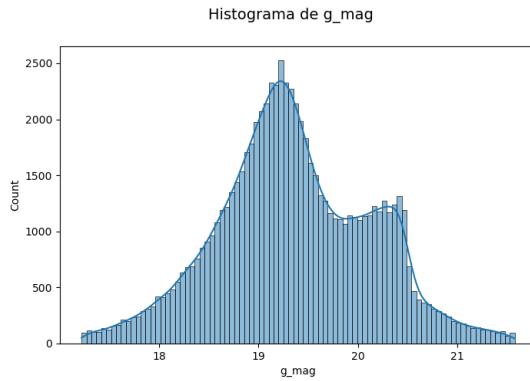


Figura 9: Histograma de datos filtrados con la regla de outliers para variable g\_mag

### 2.3.6. r\_mag

Brillo en la banda roja 658 [nm], de acuerdo Sdss.org, s.f.-a: 623[nm].

$IC(\mu)_{0,975}$	[ 18.9751 , 18.9844 ]			
$IC(\sigma^2)_{0,975}$	[ 0.5412 , 0.5543 ]			
count	mean	med	mode	std
53601	18.98	18.291	19.115	0.548

count	mean	med	mode	std
53601	18.98	18.291	19.115	0.548
min	25 %	50 %	75 %	max
17.451	18.633	19.003	19.31	20.663

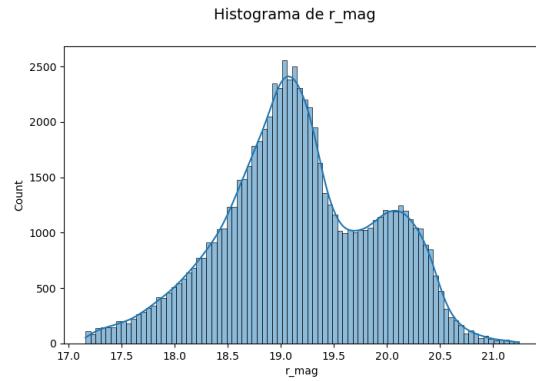


Figura 11: Histograma de datos filtrados con la regla de outliers para variable r\_mag

### 2.3.5. sig\_g\_mag

Patrón de desviación para la banda verde.

$IC(\mu)_{0,975}$	[ 0.0216 , 0.0217 ]
$IC(\sigma^2)_{0,975}$	[ 0.0051 , 0.0052 ]

### 2.3.7. sig\_r\_mag

Patrón de desviación para la banda roja.

$IC(\mu)_{0,975}$	[ 0.0194 , 0.0194 ]
$IC(\sigma^2)_{0,975}$	[ 0.0044 , 0.0045 ]

count	mean	med	mode	std
53601	0.019	0.017	0.018	0.004
min	25 %	50 %	75 %	max
0.009	0.016	0.019	0.022	0.034

count	mean	med	mode	std
53601	0.021	0.019	0.019	0.005
min	25 %	50 %	75 %	max
0.009	0.018	0.02	0.024	0.035

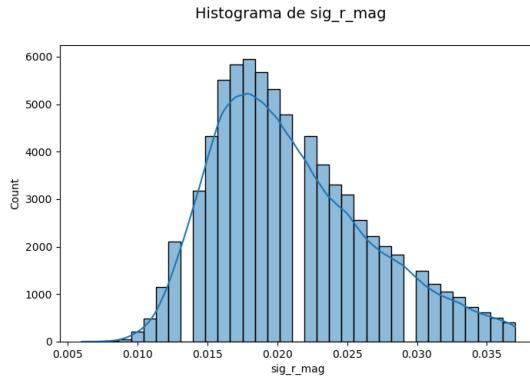


Figura 12: Histograma de datos filtrados con la regla de outliers para variable sig\_r\_mag

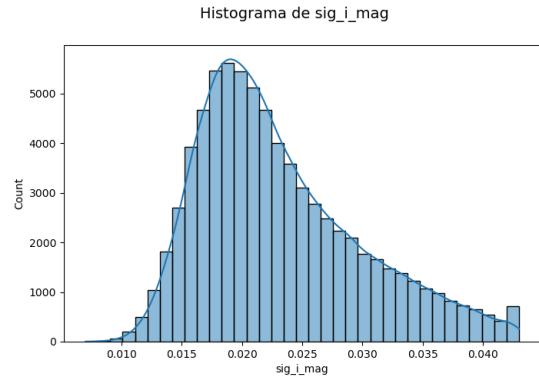


Figura 14: Figura 14: Histograma de datos filtrados con la regla de outliers para variable sig\_i\_mag

### 2.3.8. i\_mag

Brillo en la banda cercana al infrarrojo 806 [nm], de acuerdo Sdss.org, s.f.-a: 762[nm].

$IC(\mu)_{0,975}$	[ 18.8435 , 18.8524 ]			
$IC(\sigma^2)_{0,975}$	[ 0.5205 , 0.5332 ]			
count	mean	med	mode	std
53601	18.848	18.088	19.122	0.527
min	25 %	50 %	75 %	max
17.483	18.522	18.887	19.131	20.321

### 2.3.10. z\_mag

Brillo en la banda rojo profundo, infrarroja 900 [nm], de acuerdo Sdss.org, s.f.-a: 913[nm].

$IC(\mu)_{0,975}$	[ 18.7597 , 18.7687 ]			
$IC(\sigma^2)_{0,975}$	[ 0.5295 , 0.5423 ]			
count	mean	med	mode	std
53601	18.764	18.013	18.889	0.536
min	25 %	50 %	75 %	max
17.361	18.424	18.799	19.094	20.223

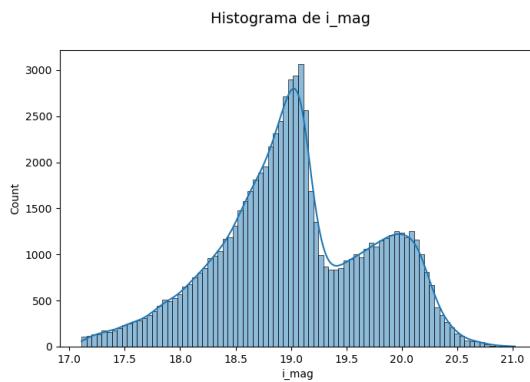


Figura 13: Histograma de datos filtrados con la regla de outliers para variable i\_mag

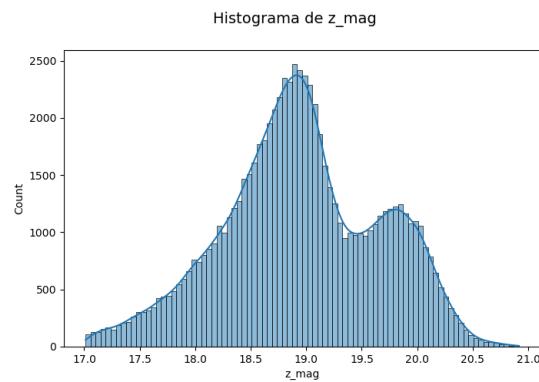


Figura 15: Histograma de datos filtrados con la regla de outliers para variable z\_mag

### 2.3.9. sig\_i\_mag

Patrón de desviación para la banda cercana al infrarrojo.

$IC(\mu)_{0,975}$	[ 0.021 , 0.0211 ]
$IC(\sigma^2)_{0,975}$	[ 0.0046 , 0.0047 ]

### 2.3.11. sig\_z\_mag

Patrón de desviación para la banda rojo profundo.

$IC(\mu)_{0,975}$	[ 0.044 , 0.0443 ]
$IC(\sigma^2)_{0,975}$	[ 0.0158 , 0.0162 ]

count	mean	med	mode	std
53601	0.044	0.026	0.036	0.016
min	25 %	50 %	75 %	max
0.016	0.032	0.041	0.053	0.09

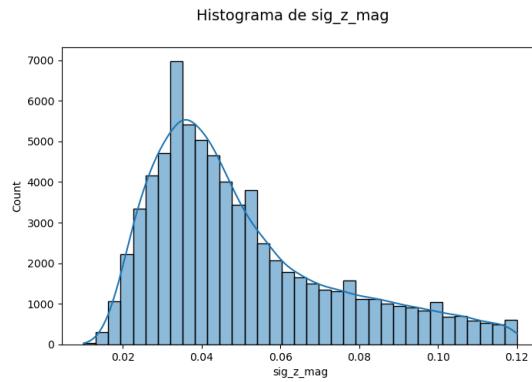


Figura 16: Histograma de datos filtrados con la regla de outliers para variable sig\_z\_mag

### 2.3.12. FIRST

Brillo en la banda de radio.

$IC(\mu)_{0,975}$	[ 0.9067 , 0.9692 ]			
$IC(\sigma^2)_{0,975}$	[ 3.6444 , 3.7327 ]			
count	mean	med	mode	std
53601	0.938	0.0	0.0	3.688
min	25 %	50 %	75 %	max
-1.0	0.0	0.0	0.0	16.4

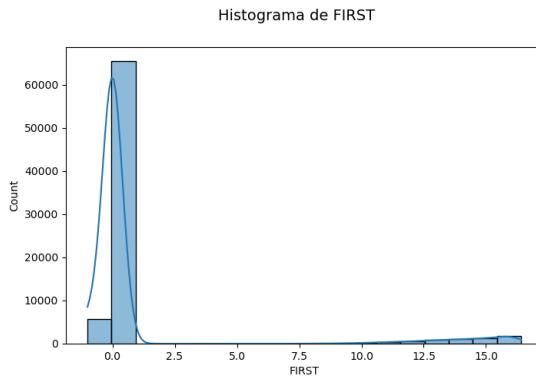


Figura 17: Histograma de datos filtrados con la regla de outliers para variable FIRST

### 2.3.13. Mp

Magnitud absoluta en la banda i, esto quiere decir que se compensa la distancia a la que están los objetos, corrigiendo su magnitud como si todas estuviesen a una distancia fija.

$$M = m - 5 \log_{10} d + 5$$

donde  $M$  es la magnitud absoluta,  $m$  la magnitud aparente a la que observamos el objeto,  $d$  la

distancia del objeto en parescs.

$IC(\mu)_{0,975}$	[ -25.4343 , -25.4113 ]
$IC(\sigma^2)_{0,975}$	[ 1.2889 , 1.3215 ]

count	mean	med	mode	std
53601	-25.423	-24.644	-26.117	1.305
min	25 %	50 %	75 %	max
-28.758	-26.359	-25.682	-24.678	-22.082

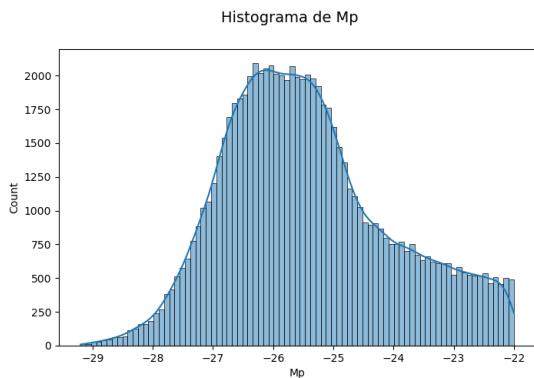


Figura 18: Figura 18: Histograma de datos filtrados con la regla de outliers para variable Mp

## 2.4. Correlación entre variables.

Con los datos filtrados se genero una matriz de covarianzas donde se cruzaron todas las variables en estudio de la cual luego se generaron los coeficientes  $r$  de Pearson para los mismos cruces. Todos los valores absolutos de  $r$  superiores a 0.7 se consideraron como correlaciones y los superiores a 0.9 fueron considerados como correlaciones fuertes, siendo los valores negativos correlaciones inversas, como se ve en la siguiente tabla.

Variables	$ r  > 0.7$
z Mp	-0.887
u_mag sig_u_mag	0.822
u_mag g_mag	0.925
u_mag r_mag	0.873
u_mag i_mag	0.835
u_mag z_mag	0.77
sig_u_mag g_mag	0.737
g_mag r_mag	0.957
g_mag i_mag	0.922
g_mag z_mag	0.878
g_mag sig_z_mag	0.774
r_mag i_mag	0.966
r_mag z_mag	0.925
r_mag sig_z_mag	0.814
i_mag z_mag	0.959
i_mag sig_z_mag	0.844
z_mag sig_z_mag	0.873

Se observó que de las correlaciones fuertes corresponden en su totalidad a pares entre las variables

de u\_mag, g\_mag, i\_mag, z\_mag, siendo estas distintas longitudes de onda para cada observación. Se prestó especial interés en la correlación entre la variable correspondiente al intervalo espectral del infrarrojo, i\_mag y z\_mag siendo esta la segunda correlación más fuerte con un valor de  $r = 0.959$ , a la cual se le aplicó una regresión lineal:

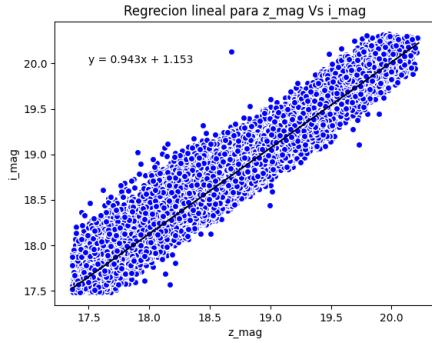


Figura 19: Regresión lineal para los datos filtrados de las variables z\_mag vs i\_mag.

Para la cual se encontró el intercepto  $B_0 = 1.153$  y la pendiente  $B_1 = 0.943$ . Téngase en cuenta que una correlación no implica causa y efecto, pues variables en un rango cercano al infrarrojo e infrarrojo profundo pueden encontrarse relacionadas debido a una galaxia ( posible candidata a quasar) posee gran cantidad de polvo, léase Hernán-Caballero et al., 2016 y Kobayashi et al., 1993 para mayor información sobre el estudio de polvo.

Sin embargo, no se puede omitir que los errores heterocedasticos los cuales producen un error en el cálculo de estimación de los mínimos cuadrados, se ha de aplicar .<sup>a1</sup> al realizar estimaciones con modelos de regresión, debe verificarse el cumplimiento de los supuestos... Si estos no se cumplen (homos-tasticidad) se requieren estrategias remédiales para continuar con el análisis estadístico "Aguilar et al., s.f.

Sobre la relación entre filtro ultravioleta y cercano al infrarrojo: la siguiente imagen 20, la luz ultravioleta (filtro u) está colorizada violeta, la luz cercana al infrarrojo (filtro i) está colorizada blanca La galaxia mas grande de las 2, es una galaxia 'Seyfert', la cual forma un 10% de todas las galaxias, siendo una de las galaxias mas activas con un AGN (active galaxy nucleus).

De acuerdo a Kaviraj, 2009 existe una relación entre la cantidad de polvo en una galaxia y la emisión cercana al infrarrojo; así como también la hay para las galaxias activas, las cuales poseen estrellas en formación.



Figura 20: This new NASA Hubble Space Telescope image of IC 4271, also known as Arp 40, is a curious pair of spiral galaxies some 800 million light-years away. The smaller galaxy is superimposed on the larger one, which is a type of active galaxy called a Seyfert galaxy. CREDIT: NASA, ESA, and B. Holwerda (University of Louisville Research Foundation, Inc.); Image processing: G. Kober (NASA Goddard/Catholic University of America)

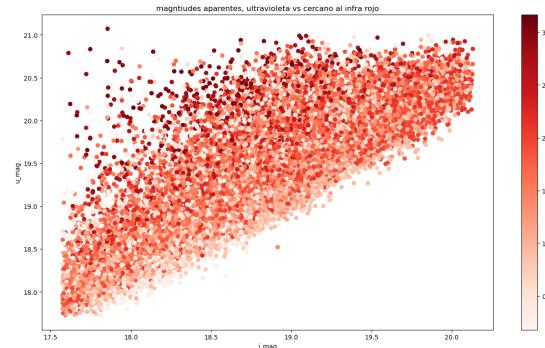


Figura 21: magnitud aparente ultra violeta vs cercana al infro rojo, con  $r = 0.835$

La emisión ultravioleta y cercana al infrarrojo se utiliza para calcular parámetros como la metalicidad (O/H), ratio de formación estelar (SFR); sin embargo, se ha de tener en cuenta que el análisis se realiza dentro de un cuerpo celeste de manera de calcular sus propiedades como la metalicidad y distancia; no como un análisis de un conjunto general repartido en distintas clases desconocidas

## 2.5. Explorando con esta información y futuras investigaciones

Gracias a tener la magnitud absoluta (el brillo de un objeto a 10 parsecs de distancia) y la aparente, se relacionan mediante

$$M = m - \log_{10}(d) - K(z) - E(z)$$

donde  $K$  es un K corrector para redshift,  $E(z)$  es un corrector para evolución

$$E(z) = 1.6(z - 0.1)$$

osea que  $M_p$  y  $i\_mag$  están relacionados mediante  $z$  el redshift, véase el siguiente gráfico para explorar esta relación:

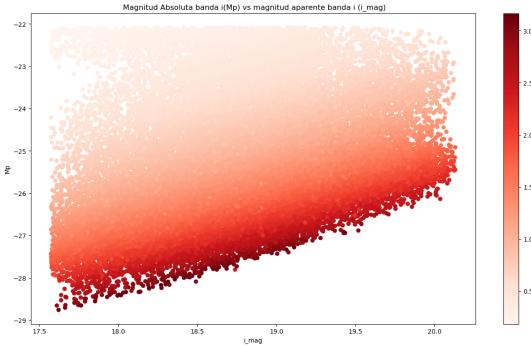


Figura 22: Como se relaciona el brillo aparente y el absoluto en la banda infrarroja, usando el redshift para observar como afecta

Se ha de tener en cuenta que este es un modelo simplificado que no toma en cuenta la extinción, que se refiere a la disminución del brillo producto de nuestra propia atmósfera.

Para calcular la distancia entonces, sin tener en cuenta evolución y redshift:

$$d = 10^{\frac{m-M-5}{5}}$$

donde  $d$  es la distancia en parsecs, he aquí la distribución

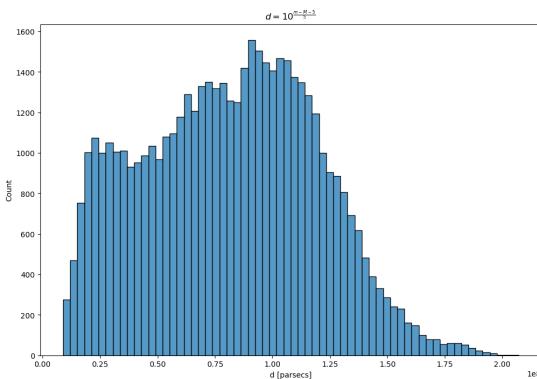


Figura 23: Distribución de las distancias aproximadas de los objetos, notar que son  $10^8$  parsecs

Luego la metalicidad y otros factores de evolución utilizan  $z\_mag$  y el  $M_p$ , lo cual escapa a la complejidad de este estudio pero se presenta el gráfico:

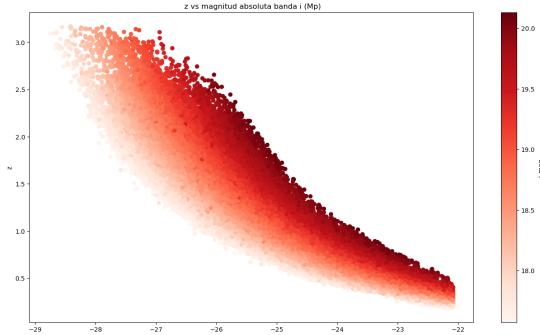


Figura 24: Brillo infrarrojo profundo vs Brillo absoluto de la banda cercana al infrarrojo

### 3. Conclusión

En el estudio de una gran base de datos se tienen muchas posibilidades para relacionar y jugar con las variables; este análisis exploratorio dio lugar a leer una gran variedad de artículos de los cuales cada uno utiliza diferentes modelos dependiendo de la precisión de sus datos.

Como un constante en toda investigación leída se observó una previa limpieza de los datos, pero la cual requiere de un conocimiento profundo de las variables en estudio, así como el objeto en cuestión que se estudia, en estos datos no se tuvo acceso a las posiciones astronómicas de los objetos por lo que se quedó con el no tan pequeño trabajo de analizar las distribuciones de las variables y sus posibles correlaciones.

Se realizó este análisis dentro de intervalos que eliminaron los outliers de acuerdo a la definición entregada en el curso de estadística Alvarado, 2022

#### 3.1. Sobre las distribuciones

Ninguna de las distribuciones para las magnitudes aparentes posee una campana gaussiana definida, todas estas poseen dos montes definidos y separados: 15, 7, 9, 12, 13; se pueden entregar varias interpretaciones, así como que el telescopio este observando a dos fuentes principales a distintas distancias cada una con una distribución aproximada a la normal, sin embargo sería irresponsable dar esta aseveración pues no se tiene acceso a las coordenadas de los datos.

Si nos separamos ese problema y redefinimos nuestra variable de estudio como la distribución de brillos para una pequeña muestra de objetos en áreas definidas del universo analizadas por SDSS en su Data Release 5, observadas desde el planeta Tierra, entonces hablaremos sobre las estimaciones de promedios y dispersión, con sus intervalos de confianza.

El redshift  $z$  que se registró en promedio se encuentra entre 1.361 y 1.3742, algunos redshift son

dinamicos y se refieren al movimiento de los cuerpos por ejemplo que orbiten algun punto; el otro tipo de redshift es el ligado a la expansion del universo. Para galaxias cercanas ( $z < 0.001$ ) la expansion del universo (o parte cosmologica del redshift) puede ser ignorada; para galaxias lejanas ( $z > 0.01$ ) la parte dinamica del redshift es menor que la parte cosmologica. Lo que observamos en estos datos se debe pri

Para las magnitudes aparentes, sin el uso de la magnitud absoluta, no podemos obtener mucha informacion sobre la muestra de cuerpos celestes, pues estas magnitudes dependenderan de la composicion de los cuerpos (spectrografia) y la distancia a la que se encuentran, solo tomando ciertas areas, corrigiendo para la distancia y extincion esta informacion se vuelve util para calcular datos como la metalicidad, edad de la galaxia y ratio de formacion estelar.

Los sigmas de cada magnitud no incluyen el 0 en su intervalo y se observan similares a una campana pero no simetrica, para la inclusion de estos errores en el analisis se requiere de un tratamiento especial para errores heterocedasticos Aguilar et al., s.f.

Para el caso de FIRST 17, que posee -1 como una no observacion no podemos darle a a la media un significado, por ello analizamos la mediana, la mediana nos entregó 0, osea que en su mayoria no identifico quasars.

Para Mp, que se refiere a la magnitud absoluta en la banda cercana al infra rojo es mas util, su distribucion es mejor comportada 18, posee un rango intercuartil del 50 % de los datos entre -26.359 y -24.678. Posee una desviacion estandar mayor que las otras magnitudes, pues estamos observando ahora una distribucion sobre la emision de varios objetos.

### 3.2. Palabras finales

A lo largo de la investigacion, mientras el conocimiento y el codigo crecian con el tiempo y la nueva experiencia, el analisis adquirio robustez en cada paso. Dado una cantidad de tiempo mayor, seria posible tener en cuenta los errores heterocedasticos estregados por SSDS (los correspondientes a 'sig\_filtro') teniendo en cuenta la informacion y tecnicas de otros investigadores.

Gracias a la investigacion es posible diferenciar tipos de correlaciones: al conocer los fenomenos fisicos que ocurren por detrás, incluido el fenomeno de la medición. De esta manera tener el cuidado de usar la estadistica correctamente, pues los calculos resultan algo trivial; sin embargo, está en su interpretación la parte más importante y que puede resultar más difícil.

## Referencias

- Oke, J. B. (1974). Absolute Spectral Energy Distributions for White Dwarfs., 27, 21. <https://doi.org/10.1086/190287>
- Kobayashi, Y., Sato, S., Yamashita, T., Shiba, H., & Takami, H. (1993). An infrared study of hot dust in quasars using prism spectrophotometry. *The Astrophysical Journal*, 404, 94-99.
- Fukugita, M., Ichikawa, T., Gunn, J., Doi, M., Shimasaku, K., & Schneider, D. (1996). The Sloan Digital Sky Survey photometric system [Copyright: Copyright 2005 Elsevier B.V., All rights reserved.]. *Astronomical Journal*, 111(4), 1748-1756. <https://doi.org/10.1086/117915>
- Lupton, R. H., Gunn, J. E., & Szalay, A. S. (1999). A Modified Magnitude System that Produces Well-Behaved Magnitudes, Colors, and Errors Even for Low Signal-to-Noise Ratio Measurements., 118(3), 1406-1410. <https://doi.org/10.1086/301004>
- York, D. G., Adelman, J., Anderson, J. E., Jr., Anderson, S. F., Annis, J., Bahcall, N. A., Bakken, J. A., Barkhouse, R., Bastian, S., Berman, E., & et al. (2000). The Sloan digital sky survey: Technical summary. *The Astronomical Journal*, 120(3), 1579-1587. <https://doi.org/10.1086/301513>
- Bessell, M. S. (2005). Standard Photometric Systems. *Annual Review of Astronomy and Astrophysics*, 43(1), 293-336. <https://doi.org/10.1146/annurev.astro.41.082801.100251>
- Kaviraj, S. (2009). An ultraviolet study of nearby luminous infrared galaxies: star formation histories and the role of AGN. *Monthly Notices of the Royal Astronomical Society*, 394(3), 1167-1181. <https://doi.org/10.1111/j.1365-2966.2009.14399.x>
- Newman, M. (2012). *Computational Physics*. CreateSpace Independent Publishing Platform.
- Hernán-Caballero, A., Hatziminaoglou, E., Alonso-Herrero, A., & Mateos, S. (2016). The near-to-mid infrared spectrum of quasars. *Monthly Notices of the Royal Astronomical Society*, 463(2), 2064-2078. <https://doi.org/10.1093/mnras/stw2107>
- Sako, M., Bassett, B., Becker, A. C., Brown, P. J., Campbell, H., Wolf, R., Cinabro, D., D'Andrea, C. B., Dawson, K. S., DeJongh, F., & et al. (2018). The data release of the Sloan Digital Sky Survey-II supernova survey. *Publications of the Astronomical Society of the Pacific*, 130(988), 064002. <https://doi.org/10.1088/1538-3873/aab4e0>

- Alvarado, E. (2022). *Repositorio curso LFIS 325*. [https://github.com/ealvnrz/LFIS325\\_2022\\_02](https://github.com/ealvnrz/LFIS325_2022_02)
- Aguilar, C. L., Araya, I. U., & Barahona, K. D. (s.f.). Comportamiento de la potencia para la prueba de igualdad de medias de área de liquenes foliosos entre la Sede Rodrigo Facio y el Monte de la Cruz, considerando distintos tamaños de muestra y en presencia de homocedasticidad y heterocedasticidad modelada con minimos cuadrados ponderados.
- Sdss.org. (s.f.-a). *Descripcion Filtros*. <http://cas.sdss.org/dr5/en/proj/advanced/color/sdssfilters.asp>
- Sdss.org. (s.f.-b). *EDR paper table 21*. <https://classic.sdss.org/dr5/algorithms/edr.tb21.html>
- Trigo, F. (s.f.). *Analisis en Python Jupyter Notebook a Sloan Digital Sky Survey Data*. <https://bit.ly/trigostats>