

Trabajo de Investigación sobre The Sloan Digital Sky Survey

Dan , Fabian

Departamento de Física y Astronomía, Universidad de Valparaíso

15 de diciembre de 2022

Resumen

Insertar aqui el abstract o resumen del trabajo una vez terminado

1. Introducción

Se utilizó el Data release 5, de Sloan Digital Observatory Sky Survey (SDSS de aqui en adelante), data release anteriores se han utilizado para el estudio de supernovas tipo 1a gracias a su gran poder espetrometrico 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 el data release 5 y seguidamente un análisis estadístico para poder estimar sobre la población entera de estrellas.

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.

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 $\pm 0.02 \text{ mag}$ (de acuerdo a Fukugita et al., 1996)

el sistema se compone de 5 bandas de color desde 300 nanometros (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:

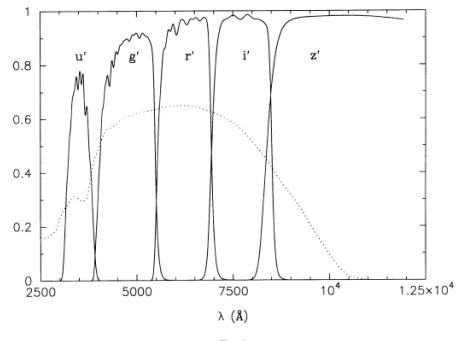


Figura 2: Curvas de cada filtro en amstrongs, se muestra la sensibilidad de cada filtro a distintas longitudes de ondas, imagen del paper Fukugita et al., 1996

el telescopio es un pipeline de datos, que toma el flujo monocromático $\{\nu\}$ 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.

la magnitud aparente de una estrella es la intensidad luminosa 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). Mientras que la magnitud



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

absoluta mide la a una distancia fija, midiendo un parametro intrínseco de la estrella, esta es la luminosidad de una estrella.

Osea 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 mas brillante que la magnitud 1 siguiente. Osea magnitud -2 es 2.512 veces mas brillantes que magnitud -1 la magnitud de venus es -5, osea que es $2.512^5 = 100$ veces mas brillantes que la magnitud 0 que es Vega (estrella)

1.2.1. Calculo de Magnitud Clasico

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. Calculo de Magnitud para SDSS

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

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

donde b es el 'softening parameter', es relativa a F_0 y no posee dimensiones, aqui una tabla para los distintos filtros del 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	1.4×10^{-10}	24.63
g	0.9×10^{-10}	25.11
r	1.2×10^{-10}	24.80
i	1.8×10^{-10}	24.36
z	7.4×10^{-10}	22.83

1.2.3. la estadistica tras el softening parameter

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

la magnitud aparente de una estrella

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

la magnitud de SDSS:

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

en los casos limites:

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

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

La varianza, donde las flechas representan el comportamiento asintotico 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

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

F es la energía radiante por unidad de superficie, es un flujo de esta energía, lo que se presenta como el vector de Poynting:

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

1.2.5. Otros sistemas de Magnitud

El sistema AB la magnitud monocromatica (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.

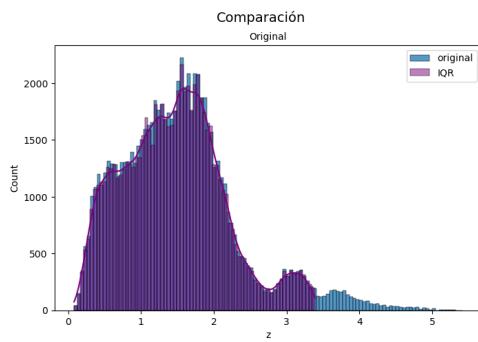


Figura 3: Comparación entre histograma de datos filtrados con la regla de outliers para variable z

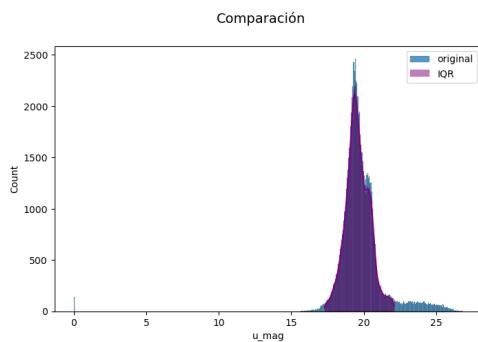


Figura 4: Comparación entre histograma de datos filtrados con la regla de outliers para variable u_mag

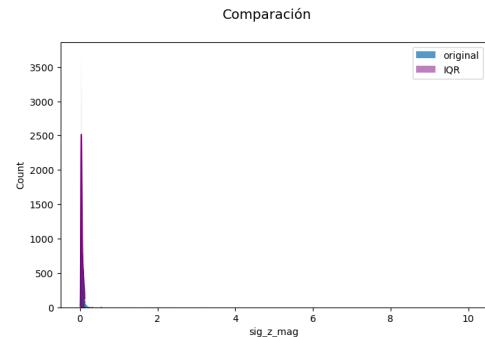


Figura 5: Comparación entre histograma de datos filtrados con la regla de outliers para variable sig z mag; el tema con las desviaciones fue más evidente

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

$$\hat{\mu} = [\hat{x} \mp]$$

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:

2.3. Descripción de variables.

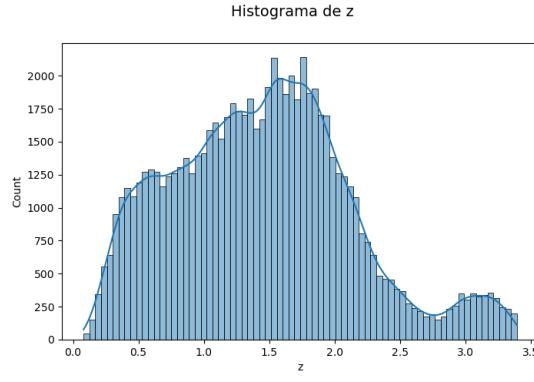
A continuación se presenta la descripción de cada variable almacenada en la base de datos después de haberles eliminado los outliers, junto con un resumen estadístico de esta y un histograma, donde se incluyen los intervalos de confianza para la media y varianza con un valor de significancia del 0.05 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.

2.3.1. z

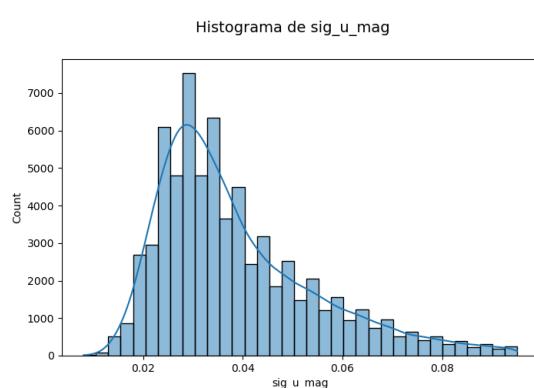
El z se refiere al redshift, el cual le da a los astrónomos la posibilidad de calcular distancias, no confundir con z_mag que viene más adelante.

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

	count	mean	med	mode	std
	53601.0	1.369	1.299	1.894	0.598
min	0.165	25 %	50 %	75 %	max



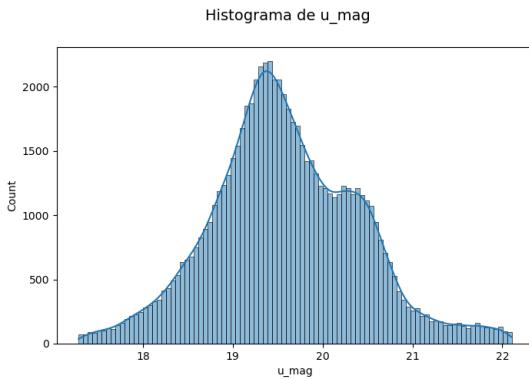
	count	mean	med	mode	std
	53601.0	0.035	0.028	0.029	0.012
min	0.012	25 %	50 %	75 %	max



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.

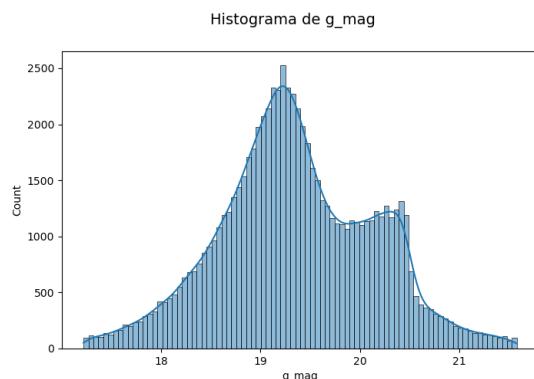
$IC(\mu)_{0,975}$	[19.4023 , 19.4125]
$IC(\sigma^2)_{0,975}$	[0.5967 , 0.6112]
	count
53601.0	19.407
min	25 %
17.605	19.016
	50 %
	19.405
	75 %
	19.809
	max
	21.342



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
53601.0	19.142
min	25 %
17.511	18.787
	50 %
	19.163
	75 %
	19.498
	max
	20.881



2.3.3. sig_u_mag

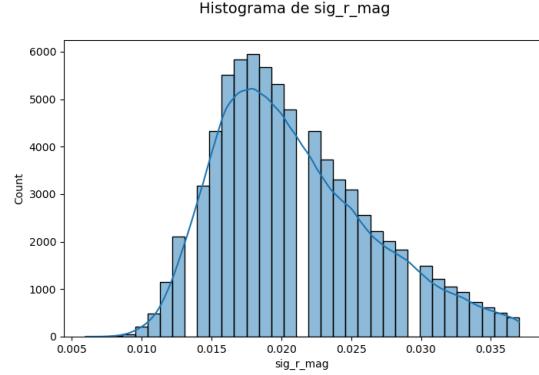
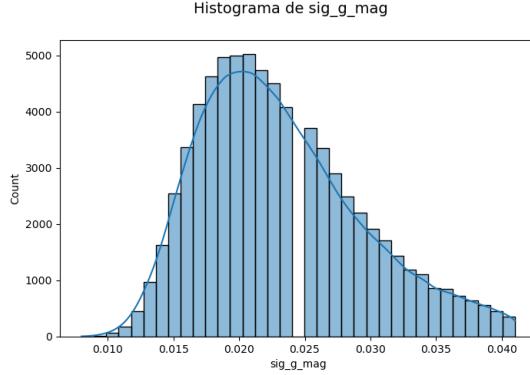
Patrón de desviación para la banda ultravioleta, el equipo SDSS determina los errores de medición para cada magnitud a partir del conocimiento de las condiciones de observación y otras consideraciones técnicas.

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

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]
	count
53601.0	0.022
min	25 %
0.009	50 %
	0.018
	75 %
	0.021
	max
	0.025
	0.037



2.3.6. r_mag

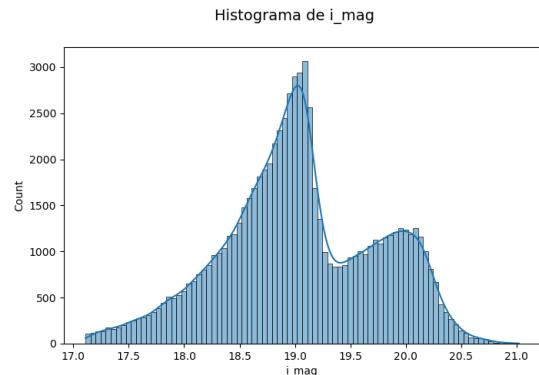
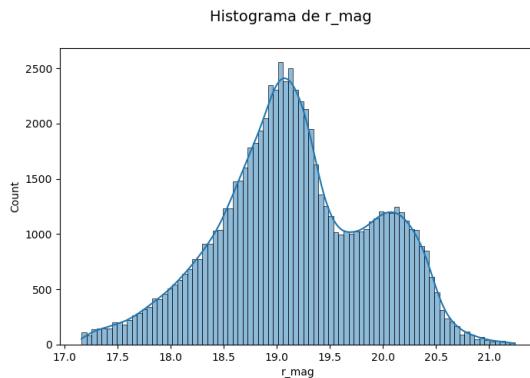
Brillo en la banda roja 658 [nm], de acuerdo Sdss.org, s.f.-a: 623[nm].

$IC(\mu)_{0,975}$	[18.9751 , 18.9844]
$IC(\sigma^2)_{0,975}$	[0.5412 , 0.5543]
count	53601.0
mean	18.98
med	18.291
mode	19.115
std	0.548
min	25 %
17.451	18.633
25 %	19.003
50 %	19.31
75 %	20.663
max	

2.3.8. i_mag

Brillo en la banda cercana al infra rojo 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	53601.0
mean	18.848
med	18.088
mode	19.122
std	0.527
min	25 %
17.483	18.522
50 %	18.887
75 %	19.131
max	20.321



2.3.7. sig_r_mag

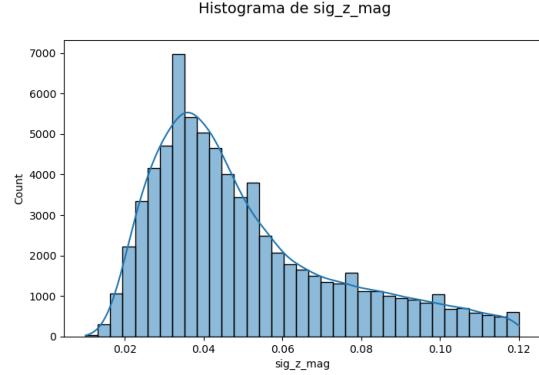
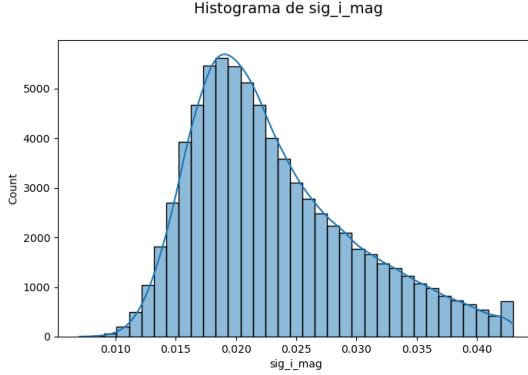
Patrón de desviación para la banda roja.

$IC(\mu)_{0,975}$	[0.0194 , 0.0194]
$IC(\sigma^2)_{0,975}$	[0.0044 , 0.0045]
count	53601.0
mean	0.019
med	0.017
mode	0.018
std	0.004
min	25 %
0.009	0.016
50 %	0.019
75 %	0.022
max	0.034

2.3.9. sig_i_mag

Patrón de desviación para la banda cercana al infra rojo.

$IC(\mu)_{0,975}$	[0.021 , 0.0211]
$IC(\sigma^2)_{0,975}$	[0.0046 , 0.0047]
count	53601.0
mean	0.021
med	0.019
mode	0.019
std	0.005
min	25 %
0.009	0.018
50 %	0.02
75 %	0.024
max	0.035



2.3.10. z_mag

Brillo en la banda rojo profundo, infra roja 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
53601.0	18.764	18.013	18.889	0.536

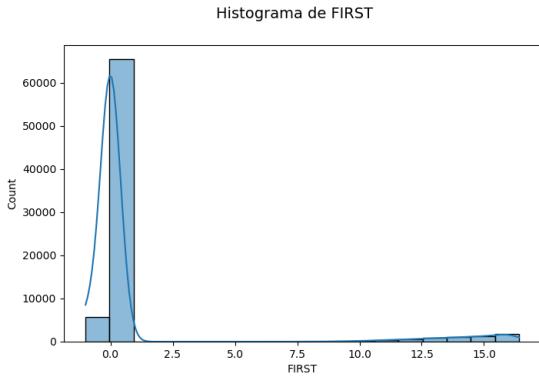
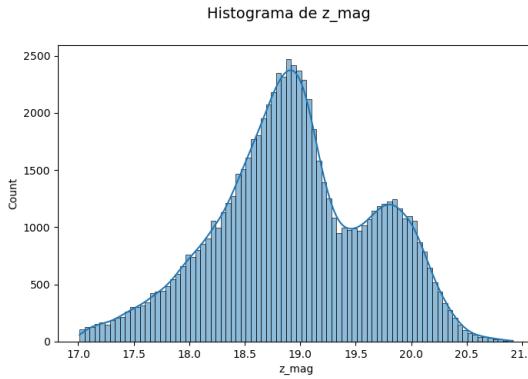
min	25 %	50 %	75 %	max
17.361	18.424	18.799	19.094	20.223

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	0.938	0.0	0.0	3.688

min	25 %	50 %	75 %	max
-1.0	0.0	0.0	0.0	16.4



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	0.044	0.026	0.036	0.016

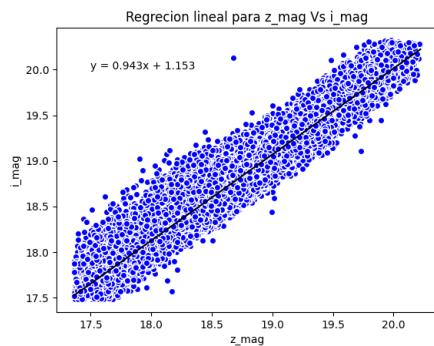
min	25 %	50 %	75 %	max
0.016	0.032	0.041	0.053	0.09

2.4. Correlación entre variables.

Con los datos filtrados se generó 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 observo 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 presta especial interés en la correlación entre la variables correspondientes al intervalo espectral del infrarrojo, i_mag y z_mag siendo esta la segunda correlación mas fuerte con un valor de $r = 0.959$, a la cual se le aplica una regresión lineal:



Para la cual se encontro el intercepto $B_0 = 1.153$ y la pendiente $B_1 = 0.943$.

Vease la siguiente imagen 6, la luz ultravioleta (filtro u) esta colorizada violeta, la luz cercana al infra rojo (filtro i) esta 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 acivas con un AGN (active galaxi nucleus).

De acuerdo a Kaviraj, 2009 existe una relación entre la cantidad de polvo en una galaxia y la emision cercana al infra rojo; así como tambien la hay para las galaxias activas las cuales poseen estrellas en formacio. La emision ultravioleta y cercana al infra rojo se utiliza para calcular parametros como la metalicidad (O/H), ratio de formacion estelar (SFR)

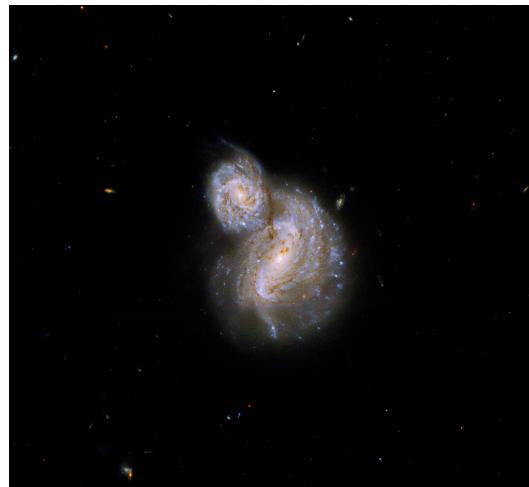


Figura 6: This new NASA Hubble Space Telescope image of IC 4271, also known as Arp 40, is a curious pair of spiral galaxies some 800 million light-years away. The smaller galaxy is superimposed on the larger one, which is a type of active galaxy called a Seyfert galaxy. CREDIT: NASA, ESA, and B. Holwerda (University of Louisville Research Foundation, Inc.); Image processing: G. Kober (NASA Goddard/Catholic University of America)

Referencias

- Oke, J. B. (1974). Absolute Spectral Energy Distributions for White Dwarfs., 27, 21. <https://doi.org/10.1086/190287>
- 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.
- 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>
- 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>