

Trabajo Final LFIS 325 - Estadística para Ciencias Físicas

Para el trabajo final del curso, utilizaremos un conjunto de datos puesto a disposición por el proyecto de sondeo astronómico *Sloan Digital Sky Survey (SDSS)*.

El SDSS ha sido un estudio del cielo extraordinariamente productivo en luz visible. Usando un telescopio de tamaño *modesto* especialmente construido en *Apache Peak Observatory (APO)* en Nuevo México, más de un tercio de la esfera celeste ha sido fotografiada usando una cámara de campo amplio en cinco bandas espectrales generando mediciones fotométricas precisas de cientos de millones de objetos. Además, se han obtenido espectros de más de un millón de objetos utilizando un espectrógrafo de fibra acoplada multiplexada. Si bien el SDSS no es el telescopio más sensible o de mayor resolución, el tamaño, la uniformidad y la confiabilidad de sus productos de datos han llevado a que varios miles de publicaciones generen casi 150000 citas desde 1996.

Los análisis deberán ser realizados utilizando algún recurso computacional de libre elección, pero se recomienda el uso de **R** o **Python**, debido a los repositorios disponibles que les podrán ser útiles.

1. Descripción del conjunto de datos:

El conjunto de datos tiene 77429 observaciones junto a 15 variables, sin datos faltantes. Los datos corresponden a la 5ta versión de datos publicados por SDSS. Proporciona la designación SDSS (que da la ubicación celeste), corrimiento al rojo (una medida de distancia), magnitudes en cinco bandas fotométricas (con errores de medición heteroscedásticos), medidas o indicadores de emisión de radio y rayos X, y magnitud absoluta (una medida de luminosidad).

Publicaciones relevantes: *York, D. G., Adelman, J., Anderson, J. E., Anderson, S. F., Annis, J., and 140 others (2000) The Sloan Digital Sky Survey: Technical Summary, Astronomical Journal, 120, 1579-1587.*

Las variables en el estudio son:

- **u_mag** y **sig_u_mag**: Brillo en la banda u (ultravioleta) en magnitudes con patrón desviación (mag). El equipo SDSS determina los errores de medición heteroscedásticos para cada magnitud a partir del conocimiento de las condiciones de observación, el fondo del detector y otras consideraciones técnicas.
- **g_mag** y **sig_g_mag**: Brillo en la banda g (verde) con desviación estándar (mag).
- **r_mag** y **sig_r_mag**: Brillo en la banda r (rojo) con desviación estándar (mag).
- **i_mag** y **sig_i_mag**: Brillo en la banda i (más rojo) con desviación estándar (mag).
- **z_mag** y **sig_z_mag**: Brillo en la banda z (más rojo) con desviación estándar (mag).
- **FIRST**: Brillo en la banda de radio, en magnitudes escaladas a partir de la densidad de flujo medida en el sondeo NRAO FIRST a 20 m. 0 indica que FIRST no detectó el cuásar, mientras que -1 indica que FIRST no lo observó.
- **ROSAT**: Brillo en la banda de rayos X, en log (tasa de conteo) del ROSAT All-Sky Survey (RASS) en la banda de 0,2 a 2,4 keV. -9 indica no detectado por RASS.
- **Mp**: Magnitud absoluta de la banda i.

2. Descripción del trabajo de investigación:

El informe a entregar, consiste en el análisis exploratorio del conjunto de datos, esto es una descripción detallada de las variables: medidas de tendencia central, variabilidad, histogramas, etc. En particular, el documento debe contener (a lo menos):

- Introducción
- Descripción física del conjunto de datos (ver artículo)
- Análisis exploratorio de datos, incluido intervalos de confianza para las variables en estudio e identificación de posibles datos atípicos
- Análisis correlacional entre las variables (recordar presentación de **correlaciones**)
- Análisis de regresión lineal simple entre la variable **z_mag** e **i_mag**
- Interpretación física de los resultados obtenidos/descritos
- Conclusión.

Máximo de integrantes por trabajo: 2

El informe deberá ser entregado a lo más el **15/12** mediante correo electrónico a **eloy.alvarado@uv.cl**, en conjunto con los códigos utilizados cualquiera sea el software utilizado. La nota final del informe será determinada según la siguiente rúbrica:

Criterios	Puntaje	Sobresaliente (5.5-7)	Suficiente (4.0 - 5.4)	Deficiente (< 4.0)
Estructura	1 - 7	El trabajo presenta claramente todos los contenidos solicitados.	El trabajo presenta claramente la mayoría de los contenidos solicitados.	El trabajo presenta menos de la mitad de los contenidos solicitados.
Metodología y resultados	1 - 7	La metodología usada es la correcta y los resultados son presentados claramente.	La metodología usada es la correcta y los resultados son presentados claramente con errores menores.	La metodología usada es incorrecta y los resultados son erróneos.
Conclusiones	1 -7	Las conclusiones son presentadas claramente y demuestran una clara comprensión del material teórico.	Las conclusiones son presentadas claramente y demuestran cierta comprensión del material teórico.	Las conclusiones son presentadas y no demuestran comprensión del material teórico.