**Aplicaciones de Data Mining en Ciencia y Tecnología**

# **TP1. Aplicaciones en astronomía: Las *Hyades***

# Informe Pre-Entrega – 30 de Septiembre 2018

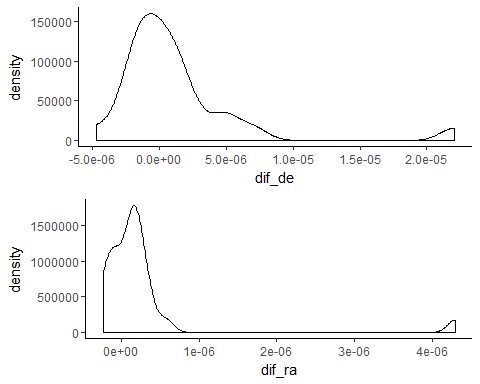
# Maestría de Explotación de Datos y Descubrimiento del Conocimiento

# Fernando Menéndez y Juan Ignacio Etcheberry Mason

El objetivo del siguiente pre-informe será analizar los errores de medición que existen entre las estrellas Hyades del catálogo Symbad, con respecto a las que ya han sido identificadas como Hyades dentro del catálogo Hipparcos. A partir de estas diferencias, se procederá a establecer algún tipo de criterio que permita determinar que otras Hyades de Symbad se podrían encontrar dentro de Hipparcos a partir de distancias similares de Ascención Recta (RA) y Declinación (DE).

Para esto se utilizan las ids cruzadas de Hipparcos y de Symbad para hallar las diferencias que existen entre ambas mediciones de RA y DE entre ambos catálogos para una misma estrella y, a partir de estas, se estudia la distribución de sus errores.

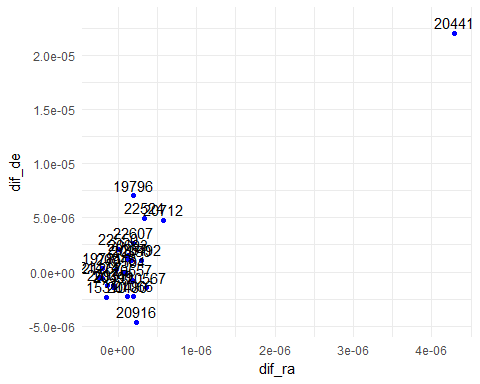
1. **Gráfico de Densidad de la Distribución de los Errores de Medición de RA y DE.**



A partir del siguiente gráfico se puede observar que las diferencias que presentan ambos catálogos son demasiado pequeñas y recién perceptibles a partir del 6to decimal. Ambas distribuciones parecen ser normales y en ambos casos se detecta la presencia de un valor atípico.

En el siguiente gráfico se pueden observar los mismos errores de medición entre ambas misiones, pero esta vez representados en un plano.

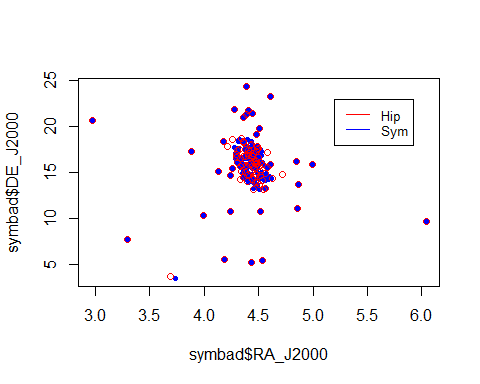
1. **Gráfico de la relación entre los errores de medición.**



En el siguiente gráfico se puede observar como en general el error de posicionamiento de la estrella en ambos catálogos se encuentra mayormente concentrado en el 6to decimal de la medición, excepto para la estrella 20441 que tiene su error en el 5to decimal.

A partir de un script en R, se explora dentro del catálogo Symbad la estrella del catálogo Hipparcos más cercana utilizando una medida de distancia euclídea, a partir las variables RA y DE con cada una de las estrellas del catálogo. De esa exploración inicial se obtiene el siguiente cruce:

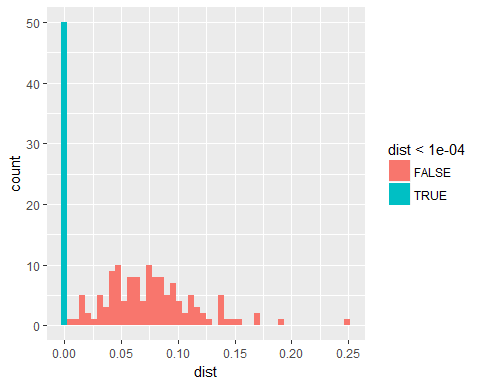
1. **Gráfico de la posición de las estrellas de Symbad y su estrella Hipparcos más cercana**



Del gráfico se puede observar que si bien existe muchas estrellas de ambos catálogos que aparecerían coincidir, existen otras donde la estrella más cercana se encuentra bastante alejada de la estrella identificada en Symbad.

A continuación, se plantea un punto de corte para la distancia entra las Symbad y su estrella más cercana en Hipparcos <0.0001. En el gráfico a continuación se puede observar como este punto de corte separa a la perfección las estrellas que deberían ser las mismas a pesar del error de medición de las que no.

1. **Gráfico de la distribución de las Distancias euclídeas**



En el siguiente gráfico, se puede observar que las potenciales Hyades en el catálogo de Hipparcos poseen distancias demasiado pequeñas con respecto a su estrella candidata dentro del catálogo Symbad. Al separar los grupos en función de su distancia, teniendo en cuenta solamente distancias inferiores al 4 decimal como “verdaderas” Hyades, se puede observar que la población celeste se concentra en su totalidad dentro del 1er bin de la distribución.

Mediante el siguiente script en R, se filtra el listado final de Hyades tomando como referencia las distancias inferiores al 4to decimal.

symbad.aumentado2 %>%  
 select(.,identifier , cercanaHip, idcruzHip, dist, dif\_ra, dif\_de) %>%  
 group\_by(cercanaHip) %>%  
 filter(dist < 0.0001)

El resultado es un Data Frame con 50 estrellas Hyades identificadas dentro del catálogo de Hipparcos.