

Sala de Entrenamiento 5: Data Cuantitativa

Análisis Estadístico con R
Semana 5
Diplomados Online

Julio 2019

En esta sala de entrenamiento tendremos una serie de ejercicios de laboratorio que nos permitan poner en práctica el conocimiento que hemos adquirido hasta ahora.

- 1) (Teorema del límite central con datos simulados). Observe el ejemplo de los datos generados por el algoritmo RANDU, donde calculamos las medias de la muestra para cada fila del dataframe randu. Repetir el análisis, pero en lugar de randu, crear una matriz de números aleatorios usando a *runif*.
- 2) (Teorema del límite central, continuación). En el contexto del ejercicio anterior, repita el análisis, pero en lugar de que el tamaño de la muestra sea 3, genere una 400 de tamaño 10. Comparar el histograma de la muestra de tamaño 3 con la de tamaño 10. ¿Qué nos dice el Teorema del Límite Central de la distribución de la media a medida que aumenta el tamaño de la muestra?
- 3) (Lotería de Vietnam de 1970). ¿Cuáles son algunas explicaciones posibles para los aparentes patrones no aleatorios en los números de la lotería de 1970?
- 4) (Cluster mamíferos). Repita el análisis de conglomerados utilizando el método de varianza mínima de Ward en lugar del método completo. El método de Ward se implementa en *hclust* con `method="ward"` cuando el primer argumento es la matriz de distancia cuadrada. Muestre un dendrograma y compare el resultado con el dendrograma del método completo.
- 5) (Identificación de grupos o conglomerados). Después del análisis de cluster, a menudo uno está interesado en identificar grupos o conglomerados en los datos. En un análisis de cluster jerárquico como el del ejemplo, esto corresponde a cortar el dendrograma a un nivel dado. La función de árbol de corte es una forma fácil de encontrar los grupos correspondientes. Por ejemplo, en el ejemplo, guardamos el resultado de nuestra cluster con el método completo, en un objeto *h*. Para cortar el árbol usamos *cutree* con `k = 5`:

$$g = \text{cutree}(h, 5)$$

Muestre *g* para ver las etiquetas de cada observación. Resumir los tamaños de grupo utilizando *table(g)*. Hay tres grupos que tienen un solo mamífero. Usar *mammals[g > 2]* para identificar cuáles son los tres mamíferos que son grupos de un solo animal.

Las respuestas a estas preguntas estarán disponibles en la próxima pildora.