

## Segundo Examen Parcial.

- Este examen es individual.
- No se pueden hacer consultas entre ustedes.
- Cualquier duda usar por favor el grupo de discusión del examen en Canvas.
- La entrega es el día Miércoles 1 de mayo antes del mediodía, vía Canvas.

1. El conjunto de datos `Harman23.cor` en el paquete `datasets` (es uno de los paquetes base, no hay que instalarlo) es una matriz de correlación de 8 medidas físicas tomadas en 305 niñas con edades entre 7 y 17 años.
  - a. Realiza un análisis de factores de estos datos usando `factanal(factors = m, covmat = Harman23.cor)` donde `m` es el número de factores.
  - b. Varía el número de factores para encontrar un ajuste adecuado del modelo e interpretar las cargas de los factores resultantes.
  - c. ¿El análisis de componentes principales produce diferentes conclusiones cuando la opción de la matriz de correlación (`cor = TRUE`) se usa? ¿Qué análisis se debería preferir?
2. El proyecto SAPA (Synthetic Aperture Personality Assessment) es una colección de datos psicológicos en la web. Un subconjunto de los datos está disponible en el paquete `psych` como `bfi`. Este subconjunto contiene datos en tres variables demográficas y 25 atributos de personalidad de 2800 voluntarios. Como ejemplos de estos atributos, se tiene:
  - Yo sé cómo confortar a otros
  - Perdí mi tiempo
  - Hago amigos fácilmente

Cada atributo se califica del 1 al 6 dependiendo de si el respondente siente que está de acuerdo con la afirmación en la escala de Likert de completamente de acuerdo (1) o completamente en desacuerdo (6). Ver `help(bfi)` para más detalles.

- a. Usar el comando `complete.cases()` para quitar a los individuos que tienen algún dato faltante en `bfi`
  - b. Usar análisis factorial para agrupar juntos atributos de naturaleza similar. Trata de interpretar la naturaleza de los atributos que se agrupan juntos.
3. En tiempos de estrés, el cuerpo pasa por un proceso llamado *vaso-constricción* en el cual las venas se cierran en las extremidades, forzando la sangre a los órganos centrales. La vaso-constricción puede ocurrir después de tomar un respiro profundo. El conjunto de datos `vaso` en el paquete `robustbase` resume la vaso-constricción (o no) de dedos de sujetos a lo largo de sus volúmenes y tasas de respiración.
  - a. Grafica las tasas de respiración y los volúmenes usando diferentes colores para aquellos con y sin vaso-constricción. Aplica LDA para ver cómo se distinguen los dos grupos.
    - Repetir el análisis de la parte (a) pero primero toma logaritmos de las variables explicativas. ¿Qué análisis se prefiere?
4. El conjunto de datos `musk` en el paquete `kernlab` es un ejemplo de un problema de discriminación en un conjunto de datos grande. En este conjunto de datos hay 476 moléculas: 207 se clasifican como almizcle y las restantes 269 no son almizcle. El objetivo es identificar una clasificación de los almizcles de los no-almizcles usando cualquier combinación de las 166 variables explicativas. Estas variables explicativas miden distancias moleculares en diferentes direcciones. Los datos originales están en [https://archive.ics.uci.edu/ml/datasets/Musk+\(Version+1\)](https://archive.ics.uci.edu/ml/datasets/Musk+(Version+1)) con información adicional que incluye una descripción completa de las variables explicativas.
5. El conjunto de datos `mtcars` tiene 11 diferentes atributos sobre autos. Estos atributos son o características diseñadas por ingenieros en la fábrica o atributos empíricos experimentados por el conductor.
  - Los atributos de diseño son: `cyl`, `disp`, `carb`, `drat`, `gear`, `vs`
  - Los atributos de experiencia son: `mpg`, `hp`, `wt`, `qsec`, `am`
  - a. ¿Cómo se pueden combinar los atributos de diseño y mostrar que se relacionan con una combinación de los atributos de experiencia?
  - b. ¿Cuál es la importancia relativa de cada atributo de diseño en crear la experiencia del conductor?
  - c. ¿Cuál experiencia del conductor domina o resume a las otras?
  - d. ¿Cuáles son los autos que aparecen al principio o al final de estas escalas?