



**UNIVERSIDAD DE COSTA RICA  
SISTEMAS DE ESTUDIO DE  
POSTGRADO EN ESTADÍSTICA**



**Examen parcial n° 1**

**Este debe ser entregado en formato .PDF, además de entregar el archivo .RMD, el archivo Excel, y el link de Tableau.**

Nombre: \_\_\_\_\_

Carné:

**I. RESPUESTAS BREVES. (40 PUNTOS / 5 PTS CU).**

1. ¿Por qué es recomendable realizar estadísticas o gráficos multivariados antes de llevar a cabo cualquier tipo de análisis? ¿Cuál cree que es el más adecuado para conocer las asociaciones entre las variables de una matriz  $p * p$  con solo variables cuantitativas? (5pts)
2. ¿Cuál es el propósito o el objetivo que busca el análisis por componentes principales? ¿Qué es lo que se busca representar? (5pts)
3. ¿A qué es igual la suma de todas las raíces características en el análisis de reducción de variables? Explique por qué cree que se da de tal forma. (5pts)
4. ¿Cuáles son principales distinciones entre el análisis por componentes principales y el análisis factorial? ¿Qué se busca mediante las rotaciones factoriales? (5pts)
5. ¿Qué se conoce como la ecuación de Fisher o la regla de decisión en el análisis discriminante? Exponga su fórmula matemática de forma generalizada, luego para la forma lineal y la cuadrática. (5pts)
6. Cuál es la riqueza que permite el uso de los dashboards que no se puede apreciar al utilizar gráficos univariados o multivariados. (5pts)
7. Por qué en el análisis discriminante se suele partir las muestras en 2 partes. Además, cuáles otros métodos existen para validar la regla de decisión. (5pts)
8. En qué se diferencian en su composición matemática los análisis de reducción de variables (PCA o FA), y el análisis de conglomerados. Explique lo anterior. (5pts)
9. ¿Qué se necesita para poder aplicar un análisis de clasificación? ¿Por qué el análisis de discriminación se considera un análisis de clasificación?
10. ¿Por qué no se puede / no se debe utilizar una variable categórica en el análisis discriminante? Mencione otro análisis de clasificación donde se puede llevar a cabo con tanto variables continuas como categóricas.

## II. PREGUNTAS PRÁCTICAS (50 pts / 10 PTS CU). + 5 PTS EXTRAS

1. A partir del archivo de “mtcars” ubicado en la librería stats, realice lo siguiente para el análisis de componentes principales, utilizando únicamente las variables cuantitativas (10 PTS).
  - a. Hacer estadísticas descriptivas de todas las variables numéricas. Describa brevemente.
  - b. Determine el número de componentes principales
  - c. Obtengan los planes de las variables.
  - d. Obtenga el plano de los sujetos.
  - e. Obtenga plano de variable y sujetos.
  - f. ¿Qué se puede concluir del presente análisis?
2. A partir del archivo de datos “attitude” ubicado en la librería de datasets, realice lo siguiente para el análisis de factores (10 PTS):
  - a. Análisis descriptivos de todas las variables numéricas. Describa brevemente.
  - b. Determinar el número de factores.
  - c. Segunda aproximación de la composición de los datos.
  - d. Ajustar 2 tipos de rotación (elija 2 de su gusto).
  - e. Determinada la mejor solución en el agrupamiento de variables.
  - f. ¿Qué se puede concluir del presente análisis?

Referencias sobre la base de attitude:

<https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/attitude.html>

3. A partir del archivo de datos “wine” de la librería car, realice el siguiente análisis discriminante, sabiendo que la variable de clase es “Type”, y las otras con las predictoras, todas continuas (10 PTS).
  - a. Análisis descriptivos de todas las variables numéricas. Describa brevemente.
  - b. Parte en 70% para entrenamiento y 30% para validación
  - c. Estime tanto la regla lineal como la cuadrática. Escriba la ecuación de la regla lineal.
  - d. Evalúe las dos reglas de decisión.
  - e. Introduzca un nuevo sujeto y comente la validez de la regla.
  - f. ¿Qué se puede concluir del presente análisis?

4. A partir del archivo “Medicina empresarial --- IAM”, se le pide que realice un Dashboard en Excel, bajo las siguientes indicaciones (10 PTS + 5 PTS EXTRAS).
  - a. Se quiere resaltar el monto promedio de la consulta y de las incapacidades promedios en días, estos como datos que deben sobresalir.
  - b. Se quiere estadísticas descriptivas para conocer la información demográfica de la población (edad, sexo, tipo de trabajador).
  - c. Interesa conocer el área de trabajo, y en esta saber ante todo por área cuál es el % de grasa, como se constituye por área las citas, y la condición % de grasa
  - d. Se quiere conocer de forma general, para el mes de junio, cuáles fueron los primeros 5 padecimientos crónicos, los primeros 5 diagnósticos micro y macro.
  - e. Finalmente se quiere conocer la relación entre la instancia de atención, y si el cliente fue entonces referido a Laboratorio, rayos “X” o consulta adicional.
  - f. Interesa tomar como segmentadores, asociadas a todos los resultados, el tipo de trabajador, el sexo, el área de trabajo.
  - g. Haga una conclusión de no más de 10 líneas sobre la situación para el mes de Junio.
5. Se desea estudiar el abandono de los estudios secundarios reportados por la UNICEF, y para esto se le pide que analice los datos en Tableau. Utilice el archivo de Excel “Out-of-School-Rate”. Interesa el análisis a nivel total, por sexo, y en las regiones urbanas y rurales. Primeramente, se quiere un estudio por regiones del mundo, luego para el continente de américa, y finalmente un análisis en Costa Rica. Proceda de la siguiente forma (10 PTS):
  - a. Realice un análisis, únicamente del total, para todas las regiones del mundo (cuadro al casi final del excel).
  - b. Luego, realice un análisis para América Latina. Tome ahora un análisis integral del total, y por sexo. Tome como países 3 de América del norte, 3 de américa del sur (Brasil, Argentina y Chile), 3 del Caribe (Jamaica, Trinidad y Tobago y Bahamas), 3 de Centro América (Panamá, Costa Rica y el Salvador)
  - c. Haga un análisis integral para el caso de Costa Rica (Total, por sexo, y por tipo de región).
  - d. Haga una pequeña conclusión del análisis (no mayor a 10 líneas).

NOTA: tome en cuenta que este archivo no es como los normales, acá ya se presentan los resultados con estadísticas descriptivas finales. Es posible un pequeño acomodo para poder realizar el ingreso de datos en Dato al definir lo que realmente debe contener el análisis.

¡BUENA SUERTE!