



ICO 187 ANÁLISIS DE DATOS

EXAMEN

Año: 2021

Profesor: Sebastián Egaña

1. Instrucciones

Debe subir su archivo DOCX a la intranet de la Universidad, en la tarea asignada.

Son 95 puntos totales, Parte 1.1 - 10 puntos, Parte 1.2 - 15 puntos, Parte 2.1 - 30 puntos y Parte 2.2 - 40 puntos.

2. Selección múltiple (25 puntos)

2.1. Generales (10 puntos).

1. Para generar tanto un modelo en Power Pivot o en Power Bi, ¿qué se tipo de llave se necesita tener para unir dos set de datos? (2 puntos)
 - ☐ Llave primerisa - Llave foránea
 - ☐ Llave primeria - Llave secundaria
 - ☐ Llave primaria - Llave foránea
 - ☐ Llave extranjera - Llave foránea
2. Selecciones el tipo de join a continuación (2 puntos):

A	B
1	NULL
2	NULL
3	3
4	4
NULL	5
NULL	6

- ☐ Ultra join
 - ☐ Outer join
 - ☐ Outer join - Left Join
 - ☐ Outer join - Right Join
3. ¿Qué frase define mejor lo que es un segmentador de datos en Excel? (2 puntos)
 - ☐ Es una herramienta de segmentación.
 - ☐ Permite aplicar filtros de manera visual e interactiva.
 - ☐ Permite determinar los segmentos de un set de datos.

- ☐ Permite segmentar.
4. ¿Cuáles son los procesos dentro de la dinámica de Power BI para el análisis de datos? (2 puntos)
- ☐ Importar datos - Visualizar datos - Generar reporte
- ☐ Obtener data - Crear modelo - Analizar los datos - Crear y publicar reporte.
- ☐ Analizar datos - Generar reporte
- ☐ Obtener data - Crear modelo - Analizar los datos - Enviar reporte por mail.
5. Dentro del entorno de Power BI, vimos dos tipos de uniones de datos, seleccione la alternativa correcta (2 puntos):
- ☐ Import y join.
- ☐ Join y multi join.
- ☐ Append y join.
- ☐ Left join y right join.

2.2. Programación en R (15 puntos).

1. ¿Cuál es el paquete utilizado para graficar en R? (3 puntos).
- ☐ ggplot
- ☐ ggplot2
- ☐ plotly
- ☐ plot
2. Elija la respuesta correcta considerando el código a continuación (3 puntos):

```
a <- 1
b <- 2
c <- a + b
a <- 10

print(paste("La suma de a + b es igual", c))
```

- ☐ La suma de a + b es igual 3
- ☐ La suma de a + b es igual 1
- ☐ La suma de a + b es igual 12
- ☐ La suma de a + b es igual 2
3. Considere la siguiente función en Excel =SI(A2 = 1; 1; 0); el segundo operador "=", por lo tanto dentro de la función SI con que operador se relaciona en R (3 puntos):.
- ☐ =
- ☐ <-
- ☐ ->
- ☐ ==
4. Considere la siguiente descripción verbal de un vector: "Se tiene un vector fila de largo n". ¿Cuál es la dimensionalidad de dicho vector? (3 puntos).
- ☐ 1:n
- ☐ n:1
- ☐ 1:\$

☐ 1:inf

5. Se tiene un conjunto de datos denominado “datos_uno”, que posee las variables “nombre”, “rut” y “edad”, ¿qué código sirve para seleccionar la variable “rut” (3 puntos):

☐ datos\$rut

☐ datos_uno

☐ datos_uno\$rut

☐ select(rut)

3. Preguntas abiertas (70 puntos)

3.1. Generales (30 puntos)

1. ¿Cuál es el procedimiento de análisis de datos visto en clases? Explique cada una de sus partes (10 puntos).
2. ¿Cuál es el objetivo del análisis de datos? (10 puntos).
3. Sobre las visualizaciones, vimos que un gráfico puede ser Feo, malo o malísimo. Explique cada categorización. (5 puntos).
4. En el contexto de la construcción de bases de datos, defina “Modelo relacional”. (5 puntos).

3.2. Programación en R (40 puntos)

1. Considere el siguiente set de datos (15 puntos):

```
mtcars[1:10,]
```

```
##           mpg  cyl  disp  hp drat   wt  qsec vs am gear carb
## Mazda RX4      21.0   6 160.0 110 3.90 2.620 16.46 0  1    4    4
## Mazda RX4 Wag  21.0   6 160.0 110 3.90 2.875 17.02 0  1    4    4
## Datsun 710      22.8   4 108.0  93 3.85 2.320 18.61 1  1    4    1
## Hornet 4 Drive  21.4   6 258.0 110 3.08 3.215 19.44 1  0    3    1
## Hornet Sportabout 18.7   8 360.0 175 3.15 3.440 17.02 0  0    3    2
## Valiant         18.1   6 225.0 105 2.76 3.460 20.22 1  0    3    1
## Duster 360      14.3   8 360.0 245 3.21 3.570 15.84 0  0    3    4
## Merc 240D       24.4   4 146.7  62 3.69 3.190 20.00 1  0    4    2
## Merc 230        22.8   4 140.8  95 3.92 3.150 22.90 1  0    4    2
## Merc 280        19.2   6 167.6 123 3.92 3.440 18.30 1  0    4    4
```

Describa dicho set de datos seleccionado:

- (a) (5 points) ¿Cuál es la unidad de observación del set de datos?
 - (b) (5 points) ¿Cuántas observaciones tiene?
 - (c) (5 points) ¿Cuántas variables tiene?
2. En R existe la opción de generar documentos utilizando el formato Rmarkdown que incluye los cálculos y gráficos realizados en dicho entorno de programación. Compare dicha forma de trabajo, con la clásica forma de trabajar usando Excel y Word. Refierase al concepto “flujo de trabajo reproducible” (15 puntos).
 3. ¿Cuáles son los pasos vistos en clase para graficar utilizando ggplot2? (10 puntos).