

## AID - Trabajo Práctico Nro. I

### Ejercicio 1: Transformaciones de datos

Seis candidatas son evaluadas para el puesto de recepcionista en una empresa, para ello pasan por dos entrevistas. En la primera las evalúa el responsable de recursos humanos de la empresa (juez 1) y en la segunda el responsable del área de la cual van a depender. La asignación de puntajes es en cordialidad, presencia y manejo de idiomas. Los puntajes asignados independientemente por estos jueces se encuentran en el archivo **recepcionistas.xls**.

- a- Calcule en promedio por juez de cada una de las aspirantes. Cual le parece que seleccionaría cada uno de ellos. Existe coincidencia?.
- b- Calcule el promedio de cada una de las aspirantes tomando en cuenta todos los rubros y ambos jueces.
- c- Transformar las puntuaciones observadas de modo tal que cada una seis variables tenga media 0 y dispersión 1. Cuál sería el objetivo de esta transformación?
- d- Transformar las puntuaciones de modo tal que cada candidata tenga para cada juez media 0 y dispersión 1. Cuál sería el objetivo de esta transformación?.
- e- Grafique los perfiles multivariados de cada una de las candidatas para ambas transformaciones. Qué observa?

### Ejercicio 2: Conjunto de datos: **Internet. 2013**

Se han registrado sobre 1500 individuos las variables siguientes:

**ID:** número de identificación del registro.

**Nacionalidad**

**Edad:** cumplida en años

**Sexo:** Masculino (1) Femenino (2)

**Estatura:** en m.

**Sitio:** sitio preferido al que se conecta; 1- chat\ 2 - correo electrónico\ 3- buscadores\ 4 – software\ 5 – música\ 6 – deportes \ 7 - otros

**Uso:** Tiempo de uso promedio por día en minutos

**Temperatura:** media anual de la zona de residencia

**Autos:** cantidad de autos en la manzana donde reside

**Cigarrillos:** Cantidad de cigarrillos consumidos mientras utiliza Internet

- a- Clasificar las variables de la base.
- b- Construir la tabla de frecuencia para la variable sexo. Hay algún valor que llame la atención? ¿Qué tipo de error considera que es?
- c- Ordenar los datos por la variable Edad. Encontró algún valor extraño?. Qué tipo de error puede ser?
- d- Construir la tabla de frecuencias de la variable Sitio. Encuentra algún valor que le llame la atención. ¿Qué tipo de error puede ser?

- e- Proceda de forma similar para las variables Temperatura, Autos y Cigarrillos.
- f- Elimine de la base los valores que no son posibles y que seguramente corresponde a un error de tipeo. Detalle valores/registros que le hayan llamado la atención pero no deban ser eliminados necesariamente.
- g- Para cuales de las variables tiene sentido calcular la media? Y la mediana?
- h- En algún caso utilizaría la media  $\alpha$ -podada? Con qué valor de  $\alpha$ ? Justifique.
- i- ¿Cuáles de las variables le parecen simétricas a partir de estos resúmenes? Confirme estas observaciones mediante un boxplot.
- j- Calcular la desviación intercuartil y detectar presencia de valores salvajes moderados y severos.

### **Ejercicio 3:** Base de datos: Gorrriones.xls

Se han registrado para 49 gorrriones las siguientes variables zoo métricas:

- Largo total
- Extensión alar
- Largo del pico y cabeza
- Largo del húmero
- Largo de la quilla del esternón
- Sobrevida (1) Si, (-1) No

- a- Indicar en cada caso de que tipo de variable se trata.
- b- Realizar una tabla de frecuencias y un resumen\* para cada una de ellas
- c- Realizar en el caso que corresponda un histograma. Ensayar el número de intervalos que conviene en cada variable.
- d- Realizar un boxplot comparativo para cada una de estas variables particionando por el grupo definido por la supervivencia. Le parece que alguna de estas variables está relacionada con la supervivencia, es decir que toma valores muy distintos en ambos grupos. Analizar en todos los casos la presencia de outliers.
- e- Construir gráficos bivariados para las todas las variables, particionando por el grupo de supervivencia. (un color para cada grupo). Observa alguna regularidad que pueda explicar la supervivencia.
- f- Calcular la matriz de varianzas y covarianzas muestrales y la matriz de correlación muestral: para el grupo general y para los subgrupos definidos por la sobrevida. Comparar estas matrices con la del grupo general y entre sí.
- g- Construir la matriz de diagramas de dispersión. Considera que algún par de estas medidas están relacionadas?. Estudiar si la asociación de algunas de estas medidas es diferente en alguno de los grupos.

Resumen \* para cada una de las variables

N (número de observaciones de la variable)

media - mediana - ds (desvío standard)

Q1 (primer cuartil)- Q3 (tercer cuartil)- DI(desvío intercuartil)

**Ejercicio 4:** Base de datos **razaperros.xls**

Se han registrado respecto de 27 razas de perros las siguientes características:

**Nombre de la raza**

**Tamaño:** 1- pequeño, 2- mediano y 3- grande

**Peso:** 1- liviano, 2- medio y 3- pesado

**Velocidad:** 1- lento, 2- mediano y 3- rápido

**Inteligencia:** niveles 1 a 3

**Afectividad:** niveles 1 a 3

**Agresividad:** niveles 1 a 3

**Función:** con tres categorías caza, utilitario y compañía.

- a- Realizar un gráfico de estrellas por raza(utilizando las variables tamaño, peso, velocidad, inteligencia y afectividad)
- b- Idem por función.
- c- Idem por agresividad.
- d- En el primer gráfico encuentra estrellas similares. Le parece que las razas son parecidas?.