

## Escuela de Ingeniería Informática

### Métodos Estadísticos, Curso 2019-20

#### Ejercicios Prácticos Lab3

**Ejercicio 1:** Leer el Data Frame que se encuentra en el fichero **“Empleo.txt”**. El fichero contiene datos de un estudio sobre la duración media en semanas de los contratos de empleo en la Unión Europea. Con los datos en él incluidos.

- a) Ordenar el data frame alfabéticamente por países
- b) Calcular la media, mediana y cuantiles de la duración del trabajo en semanas.
- c) Evaluar los parámetros de dispersión de la duración.
- d) Ordenar los países por las semanas de trabajo acumuladas en un año
- e) Visualizar las diferencias con un diagrama de caja y distinguir los valores singulares. Explicar los campos de datos resultado del uso de la función **boxplot()**
- f) Mostrar gráficamente la situación de España en e)
- g) Visualizar gráficamente las variaciones entre países de la UE (ordenados por duración y sin ordenar) y señalar en el gráfico los valores que corresponde a España.
- h) (**Opcional**) El fichero **“H\_T\_A\_UE\_2017.txt”** contiene los datos para 22 países de la UE del total del número de horas medio por trabajador en un año. Con sus datos, calcular la duración media semanal de la jornada laboral en Grecia, España y Alemania. Explicar las conclusiones.

**Ejercicio 2:** Leer el data frame que se encuentra en el fichero **“Puromicina.txt”** El fichero contiene datos de un estudio sobre la velocidad de reacción enzimática (en número de cuentas por minuto) en función de la concentración de sustrato (en partes por millón – ppm-) en experimentos donde se trataba la enzima con Puromicina (**“treated”**) o no se trataba con esta (**“untreated”**). Se pide:

- a) Calcular, las medias de la velocidad de reacción en función del empleo o no de Puromicina.
- b) Evaluar los parámetros de dispersión de la velocidad de reacción.
- c) Visualizar si la concentración del sustrato influye en la velocidad de reacción en los casos en que se trata o no con Puromicina.
- d) Ordenar por velocidad de reacción.
- e) Analizar los efectos del uso de la Puromicina y de la concentración del sustrato en la velocidad de reacción. Realizar algunas predicciones
- f) Analizar el fichero **“Puromicina\_NA.txt”** que contiene **NAs** y utilizar las funciones **na.omit()** o **complete.cases()** para evaluar el apartado a). Estudiar, y en su caso aplicar, las funciones de la librería **DMwR2** para estos casos. ¿Qué conclusiones se pueden sacar? ¿cómo afectaría el resultado si se sustituyen los **NAs** por ceros?

**Ejercicio 3:** El 2% de los equipos de un cierto fabricante de ordenadores tienen un fallo por mes de utilización y ningún ordenador tiene más de una avería por mes. El Departamento de Informática de la ULPG decide adquirir 150 de estos equipos. Se pide:

- Analizar el tipo de función de probabilidad subyacente y explicar sus características.
- Calcular la probabilidad de que el número de averías sea de 5.
- Encontrar la probabilidad de que el número de averías sea mayor o igual a 3.
- ¿Qué valor de la variable deja por debajo de sí el 75% de la probabilidad?
- Encontrar el número mínimo  $n$  tal que la probabilidad de que el número de averías sea superior a 0.99
- Calcular el percentil 95% de la distribución.
- Obtener una muestra de tamaño 1000 de esta distribución.
- Representar gráficamente la muestra de g) mediante un diagrama de barras y comparar éste con las frecuencias esperadas según el modelo que genera los datos.

**Ejercicio 4:** Consideremos una variable aleatoria que sigue una distribución  $P(x; 3)$ . Se pide:

- Calcular la probabilidad de que sea mayor o igual que 5.5.
- Calcular la probabilidad de sus valores mayores o iguales a 1 y menores o iguales a 6.
- Obtener el percentil 75 de la distribución.
- ¿Qué valor es el que deja por debajo de sí el 5% de los valores más bajos de la variable?
- Obtener una muestra de tamaño 500 de la distribución, representarla gráficamente mediante un diagrama de barras y comparar éste con las frecuencias esperadas según el modelo que genera los datos.
- Explicar la influencia del parámetro  $\lambda$  en la distribución y visualizar los diferentes resultados superpuestos.

**Ejercicio 5:** Consideremos una variable aleatoria  $W$  con distribución  $N(200, 25)$ . Se pide:

- $P[150 < W \leq 250]$
- $P[W \geq 255]$ .
- Si queremos desechar el 5% de valores más altos de la distribución y el 5% de valores más bajos, ¿con qué intervalo de valores nos quedaremos?
- Obtener una muestra de tamaño 1000 de la distribución, representar la función de densidad de esta distribución y compararla con el histograma de la muestra obtenida.
- Obtener y visualizar la función de distribución acumulada y situar sobre ella los resultados de a) y b)
- Calcular los coeficientes que definen los factores de forma de la distribución (Curtosis y Asimetría). Razonar las respuestas.