

# Tarea sobre los Temas 3 a 5

Grado en Estadística. Universidad de Sevilla

Marta Venegas Pardo

## Contents

<b>1</b>	<b>Pregunta 1</b>	<b>1</b>
1.1	Realizar una estimación no paramétrica de la función de densidad por el método del núcleo. .	1
1.2	Realizar una estimación no paramétrica de la función de densidad por el método de los logsplines.	1
1.3	Estimar $P[\text{precip} > 42]$ y el cuantil 0.90. . . . .	1
<b>2</b>	<b>Pregunta 2</b>	<b>1</b>
2.1	Fijar $M$ (número de ofertas, se recomienda al menos 1000). . . . .	2
2.2	Definir una matriz $M \times 4$ donde se irán almacenando los valores generados. . . . .	2
2.3	Repetir $M$ veces: . . . . .	2
<b>3</b>	<b>Pregunta 3</b>	<b>2</b>
<b>4</b>	<b>Pregunta 4</b>	<b>2</b>

## 1 Pregunta 1

En el dataset `datasets::precip` se recoge la cantidad media de precipitaciones de 70 ciudades de Estados Unidos (unidad = inches).

- 1.1 Realizar una estimación no paramétrica de la función de densidad por el método del núcleo.**
- 1.2 Realizar una estimación no paramétrica de la función de densidad por el método de los logsplines.**
- 1.3 Estimar  $P[\text{precip} > 42]$  y el cuantil 0.90.**

## 2 Pregunta 2

Cuando cierta empresa recibe una invitación para optar a un contrato, la oferta no se puede completar hasta que sea revisada por cuatro departamentos: Ingeniería, Personal, Legal y Contabilidad. Los departamentos empiezan a trabajar al mismo tiempo, pero lo hacen de forma independiente. El tiempo en semanas que emplean en completar la revisión es una variable aleatoria con las siguientes distribuciones. Ingeniería: Exponencial con media 3 semanas; Personal: Normal con media 4 y desviación típica 1; Legal: 2 o 4 semanas, siendo ambos valores equiprobables; Contabilidad: Uniforme continua en el intervalo (1,5). Se trata de simular el tiempo  $W$  que tarda la empresa en preparar una oferta. Para ello se pueden implementar los siguientes pasos:

**2.1 Fijar M (número de ofertas, se recomienda al menos 1000).**

**2.2 Definir una matriz Mx4 donde se irán almacenando los valores generados.**

**2.3 Repetir M veces:**

- Generar de forma independiente los cuatro tiempos según las cuatro distribuciones.
- Guardar esos cuatro tiempos en una fila de la matriz.
- Calcular W como el máximo de los cuatro tiempos.

Se pide:

- Estudiar gráficamente la distribución de la variable aleatoria W “tiempo que transcurre hasta completar la oferta”. Estimar su media y su mediana.
- Estimar la probabilidad de que W supere las 6 semanas.
- ¿Cuál es el departamento que suele tardar más en completar la revisión?
- ¿Cuál es la ordenación más frecuente de los cuatro tiempos?

### 3 Pregunta 3

Generar aleatoriamente un conjunto de datos donde tengo sentido construir un modelo de clasificación o de predicción. Ajustar el modelo y estimar su capacidad de generalización mediante Jackknife y mediante Validación Cruzada (K=10).

### 4 Pregunta 4

Bootstrap. Implementar una función que calcule el estadístico de Fisher de comparación de coeficientes de correlación lineal:

$$T = \frac{Z_1 - Z_2}{\sqrt{\frac{1}{n_1-3} + \frac{1}{n_2-3}}}$$

Donde:

- Z1 y Z2 representan la transformación de Fisher de los respectivos coeficientes de correlación lineal para dos grupos
- n1 y n2 son las frecuencias absolutas respectivas de los grupos.

Sobre un conjunto de datos apropiado (que puede ser generado), utilizando el anterior estadístico, realizar e interpretar un test bootstrap bilateral de comparación de los coeficientes de correlación lineal (B=1999).