

## EXAMEN

Este examen consiste en modelar un fenómeno a partir de las distribuciones vistas en clases, aplicando análisis y ajuste a un set de datos entregado. Se le entregará un archivo: El set de datos (dataSet01.csv).

Este archivo contiene 10.000 edades extraídas de una base de datos a la cual se quiere realiza un UPDATE (actualización de datos), aumentando las edades en 15 unidades, sin embargo el parámetro de la base de datos es de tipo NUMERIC de 2 dígitos, por lo tanto, todas aquellas edades mayores de 85 lanzaran un error a la base de datos.

Este examen se debe realizar en forma individual, y teniendo 4 horas para su desarrollo desde la hora de inicio. Debe utilizar Matlab, R, Octave, Python, Jupyter o similar (no se permite usar excel ni otro procesador de hojas de cálculo similar).

### 1. Instrucciones

#### 1.1. Estadígrafos y análisis

Grafique los datos, Establezca los intervalos que considere correspondiente a la situación planteada y calcule los estadígrafos de posición y dispersión. Use los gráficos y los estadígrafos para comentar.

#### 1.2. Distribución

Se pide ver que distribución se ajusta de mejor manera a los datos entregados, pruebe con las distribuciones vistas en clases considerando el contexto de los datos.

#### 1.3. Goodness of fit

Para ver cual distribución usar, realice un ajuste grafico con respecto a la distribución identificando aquella que minimice el error de los datos reales respecto a los teóricos.

#### 1.4. Problema

Se hace el siguiente experimento: Se tomarán 10 edades del archivo adjuntado, Se decide que ninguna marque error. En base a la distribución encontrada, calcule la probabilidad que esto suceda.

### 2. Informe

Su informe y código debe contener al menos lo siguiente:

#### 2.1. Portada

Nombre del alumno, logo(s), nombre del curso, nombre del examen, nombre del profesor y ayudantes.

## **2.2. Introducción**

Descripción de lo solicitado, descripción del problema, descripción del set de datos, etc.

## **2.3. Gráficos**

Gráficos de los datos según corresponda, junto con gráficos de las distribuciones probadas. Debe haber un gráfico que muestre tanto los datos como la distribución simultáneamente, de forma de poder ver el ajuste de forma visual. Entregar también parámetros de los ajustes, explicando su significado y usando éstos para argumentar el ajuste realizado.

## **2.4. Test de hipótesis**

Aplicar un test de hipótesis (por ejemplo el test k-s) al ajuste utilizado para argumentar su validez o significancia. Explicar el o los tests usados, sus parámetros, sus resultados al aplicarlos a su ajuste. Use un nivel de significancia de 5 %. ¿La distribución pasa el test? ¿Que implicancias tiene esto?

## **2.5. Solución de problema**

Teniendo ya definido el modelo de probabilidad que mejor se ajusta a sus datos, debe usarlo para calcular y resolver el problema planteado en la sección 1.

## **2.6. Código**

Todos los códigos usados para generar los gráficos, hacer los ajustes y los tests de hipótesis, con breves explicaciones o comentarios cuando sea pertinente para mayor claridad.

## **2.7. Conclusiones**

Conclusiones finales respecto al trabajo realizado, explicando las dificultades, y resumiendo el proceso completo realizado.