

Examen

1. Enunciado

Este examen consta de la aplicación de la probabilidad y estadística computacional, en este caso las distribuciones de probabilidad vistas en clase. Para esto se entrega un archivo csv (dataSet01.csv) que 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.

Se le pide graficar un histograma de los datos dados, a partir de su análisis respectivo, compararlo con las distribuciones vistas en clases con parámetros dados por usted, recuerde graficar igualmente la PDF (“Función de densidad de probabilidad”) y hacer la comparación dentro de un mismo grafico. Posteriormente se debe establecer la distribución que mas se acomode a los datos, realizando un ajuste respectivo a esta distribución definiendo los parámetros correspondientes.

Ya teniendo la distribución escogida, 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. Condiciones de entrega

Se tienen 4 horas como plazo para la entrega, La entrega es individual en canvas adjuntando gráficos, códigos y desarrollo. Puede utilizar Matlab, R, Octave, Python, Jupyter o similar (no se permite usar excel ni otro procesador de hojas de cálculo similar).

2.1. Estadígrafos y análisis

Grafique los datos, en base a los parámetros que definidos. Calcule los estadígrafos de posición y dispersión, y use los gráficos y los estadígrafos para comentar.

2.2. Goodness of fit

Para ver cual distribución usar, use aquella que minimice el error (distancia) cuadrática de los datos reales respecto a los teóricos.

3. Informe

Su informe y código debe contener al menos lo siguiente:

3.1. Portada

Integrantes, logo(s), nombre del curso, nombre del profesor y ayudantes.

3.2. Introducción

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

3.3. 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.

3.4. Análisis

3.5. Test de hipótesis

Aplicar un test de hipótesis (por ejemplo el test k-s) a los ajustes 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 %. ¿Cual distribución pasa el test? ¿Que implicancias tiene esto?

3.6. 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.

3.7. 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.

3.8. Conclusiones

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