

Curso EL 4702.

Proyecto #1. Estadística Descriptiva

Escuela de Ingeniería Electrónica
Licenciatura en Ingeniería Electrónica

I parte: Generalidades

Con este trabajo se busca evaluar la comprensión del estudiante mediante la aplicación en la práctica de los temas referentes a estadística descriptiva y presentación de datos vistos en el curso.

A cada equipo de trabajo (ver sección 4) se le entregará un conjunto de datos recolectados en experimentos de muestreo de la vida de real.

Para estos datos, cada equipo deberá realizar un análisis estadístico básico y presentar sus descubrimientos por medio de un reporte en formato tipo artículo científico (IEEE) de no más de 4 páginas.

1 Conjunto de datos proporcionado

El conjunto de datos proporcionado corresponde a mediciones reales del consumo de energía, temperatura y humedad en un edificio inteligente en consumo de energía¹. Para estas mediciones se buscaba estudiar la relación entre el consumo de energía y las variables de temperatura y humedad dentro de la casa, mediante un proceso conocido como regresión lineal (fuera del alcance del curso, pero muy interesante).

Sin embargo, para esta tarea solamente se analizarán las variables de humedad y temperatura para diferentes partes de la casa, una variable en específico para cada grupo de trabajo.

Mediante una red inalámbrica de sensores (WSN por sus siglas en inglés), ahora más conocido como sistemas de Internet of Things (IoT), se midieron la temperatura y humedad alrededor de cada 3.3 minutos en cada sensor. Los datos medidos se promedian cada 10 minutos (3 valores), creando un único valor para cada variable (humedad o temperatura) en cada sensor, que es el que se guarda en el archivo proporcionado.

Los datos completos se encontrarán en el tec-digital en un archivo csv, junto con un archivo README que describe la información dentro del mismo.

Cada equipo de trabajo deberá enfocarse en la descripción y análisis de **solo una de las variables** medidas por los sensores, asignada al equipo por la profesora luego de realizar el registro del equipo (ver sección 4).

2 Ambiente de programación

Para la solución de este trabajo se deberá utilizar el lenguaje de programación **Python**, y sus bibliotecas para tratamiento de datos, análisis matemático

¹Datos obtenidos de <https://archive.ics.uci.edu/ml/datasets/Appliances+energy+prediction>

y estadístico, y presentación de datos como **Numpy, Pandas, Scipy y/o Matplotlib**.

Existen lenguajes de programación (como R) o herramientas computacionales específicamente para análisis estadístico (Minitab, SPSS, Spotfire, JMP, etc.). Sin embargo, lenguajes específicos son raramente usados en ingeniería fuera del análisis estadístico o sus derivados, mientras las herramientas especializadas no son de libre acceso. Por ende, se escoge Python por ser una herramienta de libre acceso ya conocida por el estudiante, y cuyo uso será recurrente en el futuro.

3 Instrucciones

Cada equipo de trabajo deberá realizar:

1. **El cálculo estadístico de las medidas de tendencia y dispersión vistas en el curso para el conjunto proporcionado.** Para dichos cálculos es posible crear las funciones desde cero, o realizar una búsqueda y utilizar las funciones incorporadas en las bibliotecas de Python. Para el último caso, se deberá justificar el uso de las funciones escogidas y los atributos utilizados para las mismas. Se reducirán puntos si se utilizan las funciones como “cajas negras” sin justificación, aún cuando el resultado sea correcto.
2. **Presentación de los datos por medio de la(s) forma(s) de presentación de datos que el equipo considere adecuada(s).** La(s) forma(s) de presentación escogida(s) tendrá(n) el objetivo de apoyar el análisis de los datos de manera estadística y, acorde a su uso, permitirle al equipo explicar características importantes encontradas. La escogencia del método (o métodos) utilizados deberá ser justificada en base a lo que se busca presentar y posteriormente analizar.
3. **Un análisis básico estadístico de los datos.** En este se debe tratar de comprender y explicar el comportamiento detrás de la variable asignada, a partir de los números que se obtienen en las medidas de tendencia central y/o dispersión y con ayuda de la(s) representación(es) gráfica(s) escogida(s). Recuerde que los datos dados provienen de un fenómeno real, y con el análisis se busca encontrar características particulares o interesantes dentro del mismo. **Se debe** hacer énfasis en términos como sesgo, varianza de los datos, presencia de puntos atípicos, y confiabilidad de las medidas de tendencia central presentadas. Además, estos conceptos se deben relacionar con el problema y experimento real que se tiene.

Recuerde: estadística **NO es solo** el cálculo de valores, sino su interpretación.

4 Procedimiento para la solución

La resolución de esta tarea se deberá realizar en **parejas o tríos**. Grupos de trabajo de mayor tamaño no son permitidos bajo ninguna circunstancia. Personas trabajando individualmente solo serán permitidas con autorización explícita de la profesora y solo se aceptarán en casos de necesidad. Trabajos individuales presentados sin previa autorización no serán revisados.

Todos los equipos de trabajo deberán registrarse antes del 17 de febrero a las 7pm en el siguiente enlace: <https://forms.gle/kkPfkS7tdHpxxVQm9>. Este paso es imprescindible pues luego del mismo se asignará la variable para análisis que cada equipo utilizará. **No se permite escoger variables de trabajo al azar y por decisión propia.** Trabajos de equipos que así lo hagan no serán revisados y obtendrán nota nula.

Todos los archivos entregables (ver sub-sección 1 en Evaluación) deberán ser subidos al TEC-Digital **antes de la fecha de entrega**, en la sección correspondiente a Proyecto 1.

Para facilitar el inicio del trabajo, se facilita un código de arranque que permite leer las partes correspondientes a las variables desde el archivo csv. Este código es solo una ayuda, y no necesariamente debe ser usado por el grupo si se desea utilizar otra función o forma de lectura. Cualquier cambio o extensión que se decida hacer al código proporcionado correrá por cuenta del grupo de trabajo.

II parte: Evaluación

Esta tarea tiene una porcentaje del **4 %** de la nota total. La fecha de entrega para la misma será **el 3 de marzo a las 12 medio día**.

Los equipos podrán presentar el trabajo de manera tardía, pero con penalidad. Así, por cada día de atraso se impondrá un 20 % de penalidad a la nota final obtenida.

Es decir, si el trabajo se entrega el 3 de marzo luego de las 12m.d. la nota final se basará sobre un 80 %, si se entrega el 4 de marzo luego de las 12m.d. la nota final se basará sobre un 60 %, y así sucesivamente. Esta penalización incluye fines de semana, pues es solo una prórroga condicionada.

1 Entregables

Cada equipo de trabajo deberá presentar:

1. Un único archivo con el código fuente para el cálculo y creación de los gráficos utilizados. Dicho archivo puede ser extensiones .py o un Jupyter Notebook.
2. Reporte corto (no más de 4 páginas) en formato tipo artículo científico IEEE. El template para dicho reporte puede ser encontrado en el TEC digital (Word o LaTeX).

Importante: Este trabajo NO es un reporte de código/algorithm. El objetivo del reporte corto es la presentación y descripción de los resultados obtenidos por medio del cálculo de medidas y gráficos, y el análisis de los mismos. Además, la justificación de los gráficos y forma de cálculo de las medidas escogidas. Por lo tanto, no se deben incluir *pantallazos* del código en el reporte, pues para eso se entregará el mismo por aparte.

2 Desglose de evaluación

El trabajo será calificado en base al siguiente desglose:

- Demostración de manejo de los conceptos básicos del curso que se evalúan (40 %).
- Evaluación de métodos de presentación de datos y escogencia del más adecuado en base a los datos (20 %).
- Capacidad de utilizar herramientas de estadística descriptiva para realizar análisis simples de un conjunto de datos (40 %).

Toda conducta fraudulenta será tratada según lo estipulado en el artículo 75 del RREA del TEC. **Con estas conductas no engañan al profesor(a), se engañan ustedes mismos.**

III parte: Consulta

Recuerde que durante todo el período de la tarea, puede contar con la ayuda de la profesora durante las horas de consulta.

Alternativamente, también puede realizar preguntas puntuales por medio del correo, siempre y cuando las mismas sean dentro de horario laboral.