

Curso EL 4702.

## **Proyecto #3. Inferencia estadística**

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 inferencia estadística: prueba de hipótesis, estimación de parámetros, y análisis de datos para la obtención de conclusiones válidas.

A cada equipo de trabajo (ver sección 4) se le entregará un conjunto de datos que se deberá usar para realizar un análisis de inferencia estadística y presentar sus descubrimientos y conclusiones por medio de un reporte en formato tipo artículo científico (IEEE).

### 1 Conjunto de datos proporcionado (caso hipotético)

En una empresa se tiene una máquina manufacturera que dejan a su cuidado como ingeniero/a recién graduado. El rendimiento promedio de la máquina se considera aceptable cuando iguala o supera un 70 % de rendimiento medio. La máquina se mantiene funcionando 24 horas, los 7 días de la semana. Cada día a la misma hora, el rendimiento diario (últimas 24 horas) de la máquina es medido y todas las mediciones diarias son promediadas cada 3 meses para obtener un rendimiento medio.

Hace un par de meses, un colega ingeniero realizó cambios en la configuración de la máquina para mejorar su rendimiento promedio. Se asegura que dichos cambios incrementaron el rendimiento en un 5 % el rendimiento medio, con respecto al promedio requerido. Esta conclusión se basa en las mediciones diarias de 1 semana, donde se obtuvieron las mediciones en la Tabla 1, con un rendimiento promedio de 75 %.

Cuadro 1: Mediciones de rendimiento diario para máquina manufacturera luego de cambio realizado por colega. Valores en porcentajes (%)

Día	Lun.	Mar.	Mie.	Jue.	Vie.	Sab.	Dom.
Rendimiento	65	88	90	74	64	63	81

Usted recuerda lo aprendido en sus clases de probabilidad y estadística, y no está completamente de acuerdo con la conclusión de su colega. Sin embargo, antes de presentar sus dudas a la empresa debe primero tener pruebas sólidas. Además, con la experiencia que le dejó el TEC, usted cree tener una solución aún mejor que la ya implementada y desea probarla.

Para corroborar sus dudas, usted explica a su superior inmediato lo que pretende realizar y recibe permiso para obtener dos mediciones:

1. Tres meses de mediciones diarias (tamaño de muestra igual a 90 días) utilizando la configuración presentada por su colega. Este conjunto de datos se llamará *primer cambio*.

2. Tres meses de mediciones diarias (tamaño de muestra igual a 90 días) utilizando la configuración que usted propone. Este conjunto de datos se llamará *segundo\_cambio*.

Además, en un servidor de la empresa usted encuentra un conjunto de mediciones diarias de tres meses (tamaño de muestra igual a 90 días) para la misma máquina antes de la aplicación de la nueva configuración hecha por su colega. Este conjunto se llamará *inicial*.

## 2 Instrucciones

En el tec-digital se encuentran los tres conjuntos de datos descritos en la sección previa. Para los mismos, cada equipo de trabajo deberá realizar:

### 2.1. Pruebas de hipótesis

Primeramente, se deberá comprobar que todas las configuraciones presentan un valor de rendimiento medio poblacional igual o mayor (significativo) al límite de 70 % considerado como aceptable.

Además, se deberán realizar **comparaciones** estadísticamente correctas entre los datos para las 3 configuraciones, con el objetivo de encontrar cual de ellas representa la mejor solución para la empresa.

En base a estos dos objetivos (comprobación y comparación), cada equipo de trabajo deberá escoger la prueba o pruebas de hipótesis que más se ajusten a lo necesitado, realizar dichas pruebas en el ambiente de programación especificado y analizar los datos obtenidos (ver sección 2.3).

### 2.2. Estimación de parámetro de desviación estándar

Luego de un análisis adicional, se logra determinar que todas las configuraciones generan un comportamiento normal en el rendimiento de la máquina.

Utilizando el método de máxima verosimilitud (MLE), cada equipo de trabajo deberá encontrar **de manera empírica/práctica** el valor de desviación estándar (y así la varianza) de cada distribución. Se **debe suponer** que el valor de la media para esta distribución es conocido, y corresponde a la media muestral (estimación analítica).

Además, el valor de varianza obtenido empíricamente se debe comparar objetivamente (ejemplo, con porcentaje de error) con el valor obtenido por el medio analítico para una distribución normal.

### 2.3. Inferencia estadística

Basado en los resultados obtenidos para los puntos anteriores, cada equipo de trabajo deberá realizar un análisis de inferencia estadística que le permita llegar a conclusiones válidas basadas en experimentación.

Note que también puede hacer uso de cualquier herramienta estudiada sobre estadística descriptiva (histogramas, cajas, tablas, etc.) para describir y explicar mejor sus descubrimientos.

Como mínimo, en dicha inferencia estadística se debe contestar:

1. ¿Son todas las configuraciones para la máquina aceptables?
2. ¿Es correcta la conclusión del colega ingeniero en el caso hipotético respecto a la primera mejora? ¿Si, no, casi, y por qué?
3. ¿Cuál de las configuraciones es la que más beneficia a la empresa?
4. ¿Poseen las tres configuraciones la misma varianza? ¿En qué afecta al resultado de comparación entre configuraciones posibles diferencias en las varianzas de las distribuciones?

Todas sus respuestas deben de justificarse **basado en sus resultados** de los puntos anteriores.

## 3 Ambiente de programación y bibliotecas sugeridas

Para la solución de este trabajo se deberá utilizar el lenguaje de programación **Python**, y sus bibliotecas para tratamiento de datos, análisis matemático y estadístico, y presentación de datos como **Numpy, Pandas, Scipy y/o Matplotlib**.

Se recomienda la investigación y el uso de funciones como *Scipy stats.ttest\_ind*, *Scipy stats.ttest\_1samp* y *Scipy stats.norm.fit* (con esta última se deben tomar precauciones, pues como default la función estima **TODOS los parámetros**. Se debe ajustar a los requerimientos de la asignación por medio de sus atributos). **No se recomienda** la creación desde cero de algoritmos para las pruebas de hipótesis ni el uso de niveles de significancia fijos, pues esto limitará el análisis.

Se deja a discreción del equipo de trabajo si se usa un algoritmo propio para el MLE o si se utiliza una biblioteca de Python (*Scipy stats.norm.fit*), aunque se recomienda la segunda alternativa. Además, se recomienda tener cuidado con bibliotecas para el algoritmo MLE, pues estas son en su mayoría orientadas a trabajos en aprendizaje de máquina. Por lo tanto, se debe leer con cuidado y entender todos los argumentos y sub-funciones para las mismas, así como los estadísticos que se optimizan y aquellos que no.

## 4 Procedimiento para la solución

La resolución de esta tarea podrá ser realizada de manera grupal, en **parejas** o **tríos**. Grupos de trabajo de mayor tamaño no son permitidos bajo ninguna circunstancia.

Debido a la cantidad de trabajo que se requerirá, personas trabajando individualmente no se aceptarán a menos de que haya limitaciones de peso para esta alternativa. La situación debe ser discutida con la profesora previamente. Trabajos presentados de manera individual sin el consentimiento previo de la profesora no serán recibidos ni evaluados.

**IMPORTANTE:** Todos los equipos deberán registrarse para control en el siguiente link: <https://forms.gle/HojZPqNNsp78H8nY7>. Dicho registro **se debe hacer a más tardar el 19 de mayo del 2023**. Equipos sin registrarse no serán evaluados.

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 la Proyecto #3.

**Cualquier evidencia de actividad fraudulenta será tratada acorde a lo estipulado en el artículo 75 del RREA.**

## II parte: Evaluación

Esta tarea tiene un porcentaje del **10 %** de la nota total. La fecha de entrega para la misma será **el 7 de Junio del 2023 a las 12 medio día**.

**IMPORTANTE:** esta fecha es definitiva debido a la cercanía con el segundo parcial. Los equipos podrán presentar el trabajo de manera tardía con penalidad. Así, por cada día de atraso se impondrá un 20 % de penalidad a la nota final obtenida. Esta penalización incluye fines de semana, pues es solo una prórroga condicionada.

### 1 Entregables

Cada equipo de trabajo deberá presentar:

1. Archivo **UNICO** con el código fuente para los algoritmos, pruebas y estimaciones realizadas. Dicho archivo puede ser extensión .py o un Jupyter Notebook. De presentarse más de un archivo, se revisará solo el primero y se ignorarán el resto. Esto porque se busca que los equipos trabajen de manera conjunta y no *juntando* el trabajo al final.
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).

## 2 Rubricas de evaluación

El trabajo será calificado en base a las siguientes rúbricas:

- Demostración de manejo de los conceptos básicos para el uso e implementación de pruebas de hipótesis (30 %).
- Demostración de manejo de los conceptos básicos para el uso e implementación de un algoritmo de máxima verosimilitud (MLE) para estimación de parámetros (30 %).
- Capacidad de utilizar herramientas de inferencia estadística para realizar análisis y comparaciones de datos, y obtener conclusiones válidas (40 %).

## 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.