

Juan Fernando Jaramillo

Santiago del Campo

Juan David Carvajal

German Carvajal

**Diseño de Experimento**

* Etapa de planeación y realización.

1. **Entender y delimitar el problema u objeto de estudio:**

Con base en la previa investigación y con la información suministrada por parte de la empresa Allers group, se han identificado los siguientes síntomas y necesidades, que ayudan a entender el problema que se está tratando para posteriormente poder delimitarlo, dejando claro el alcance que va a tener.

**Síntomas:**

**-**Gran cantidad de datos sin aprovechar, pues la información existe, pero se encuentra almacenada en una base de datos sin usos actual.

-Datos de propósito general, que contienen información que puede o no ser útil para la empresa.

-Expansión de la empresa, lo que genera que se produzcan aún más datos.

-Falta de mecanismos para análisis de datos.

**Necesidades**:

**-**Mecanismos de análisis de datos, para tratar toda la información nueva y vieja que posee la empresa, y darle un fin.

-Depuración de los datos a analizar, pues analizar datos no relevantes es una pérdida de recursos computacionales y tiempo.

-Capacidad de predecir comportamiento de compra de los clientes, necesidad establecida por el encargado del manejo de los datos.

**Definición del problema y su alcance:**

**Introducción**

Los diseños de experimentos desempeñan un papel muy importante en ingeniería porque sirve de apoyo en las fases de validación de un servicio, modelo o desarrollo de un componente software o hardware de un sistema en condiciones más cercanas a los escenarios de trabajo, o cuando se quiere estudiar el comportamiento de un sistema.

Un componente importante en la arquitectura de un computador es el sistema de memoria, ya que se convierte en el cuello de botella de los tiempos de ejecución. Entonces la determinación y medición del tiempo de acceso es crucial para tomar decisiones de configuraciones correctas o adecuadas y determinar y explicar desde la perspectiva del programador como diferentes programas o algoritmos tener diferencias significativas a la hora de su ejecución en un computador con una configuración determinada y especialmente en su interacción de con el sistema de memoria.

Un interés seria realizar un diseño experimental en el que se comparan los tiempos de los algoritmos utilizados en el software realizado, y como diferentes elementos de software y hardware como el procesador y el sistema operativo impactan en el desempeño, para realizar un diagnóstico y recomendación sobre una posible maquina adecuada para ejecutar nuestro programa.

**Definición**

La información actual e histórica que posee la empresa Allers Group sobre las transacciones de sus clientes, productos y proveedores no está siendo utilizada, imposibilitando así la predicción de futuras ventas y transacciones.

* Solo se va a abordar el problema con los datos suministrados por la empresa, que fueron limitados por seguridad de la misma.
* Los datos no van a ser actualizados para este análisis.
* No será necesario la compra de servidores o servicios en la nube para alojar datos en grandes cantidades.
* Solo se utilizaran algunas técnicas de minería de datos para el análisis respectivo.
* Estos datos se trataran como la unidad experimental.

1. **Elegir la(s) variable(s) de respuesta que será medida en cada punto del diseño y verificar que se mide de manera confiable:**

Las variables de respuesta en esta problemática son:

* Tiempo de ejecución en los métodos de análisis y como afectan los diferentes componentes estudiados el resultado.

1. **Determinar cuáles factores deben estudiarse o investigarse, de acuerdo a la supuesta influencia que tienen sobre la respuesta:**

Factores controlables:

* Métodos a usar.
* Cantidad de datos.
* Datos estudiados.
* Sistema operativo
* Tipo de procesador

Factores no controlables:

* Desbordamientos o errores por las características de la máquina.
* Lentitud de visual studio.
* Cantidad de programas y paginas abiertas
* Cantidad de RAM ocupada

Factores estudiados:

* Cantidad de item sets.
* Sistema operativo
* Tipo de procesador
* tiempo

1. **Seleccionar los niveles de cada factor, así como el diseño experimental adecuado a los factores que se tienen y al objetivo del experimento:**

* Pruebas de los métodos con pocos datos (de 5 a 20 datos), o con todos los datos (más de 1000).
* Dos niveles (Windows y macOS) para el sistema operativo
* Dos niveles(Intel core i5 5 y 4 gen) para el procesador
* Dos repeticiones o tratamientos en total

1. **Planear y organizar el trabajo experimental:**

Este trabajo será realizado y analizado en el programa Minitab18, en el cual se realizaran pruebas de ANOVA con los datos recolectados, así como análisis de gráficos de interacción entre variables, diagramas de Pareto, diferencias de medias, entre otros. Como primeras medidas, con algunos métodos de análisis se planeó usar un experimento piloto, el cual utiliza un número pequeño de observaciones. El objetivo de su ejecución es ayudar a completar y chequear la lista de acciones a realizar. Las ventajas que proporciona la realización de un pequeño experimento piloto son las siguientes:

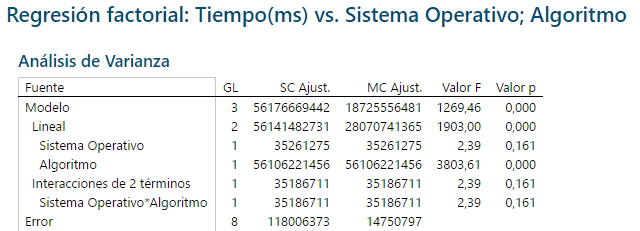
— permite practicar la técnica experimental elegida e identificar problemas no esperados en el proceso de recogida de datos,

— si el experimento piloto tiene un tamaño suficientemente grande puede ayudar a seleccionar un modelo adecuado al experimento principal,

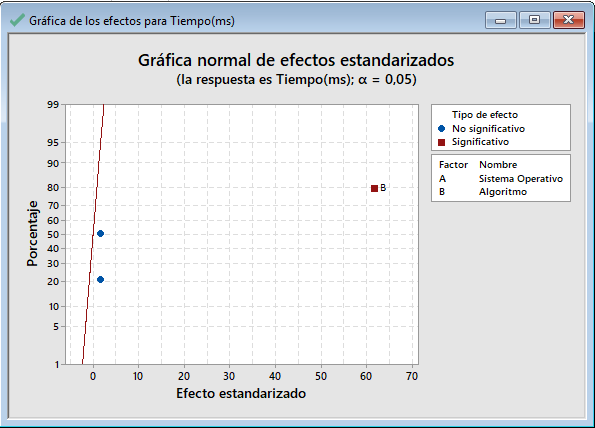
— los errores experimentales observados en el experimento piloto pueden ayudar a calcular el número de observaciones que se precisan en el experimento principal.

* Ayuda a saber qué tipo de maquina podría ser óptima para ejecutar nuestro software.

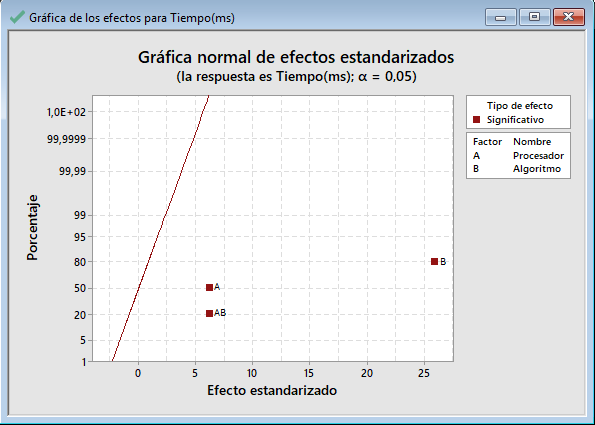
Realización:



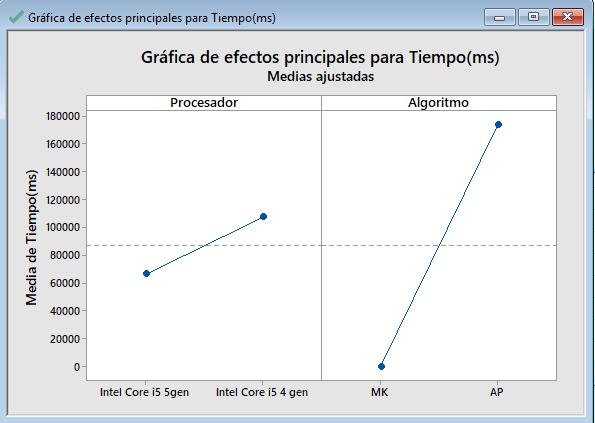
Realizando el análisis del modelo, en la tabla de varianza, podemos analizar que valores son significativos estadísticamente, y esto es cuando su valor p es menor a el nivel de significancia 0,05. En este caso el tipo de algoritmo es relevante para la respuesta, pero el sistema operativo no. Como se evidencia en la siguiente grafica normal:

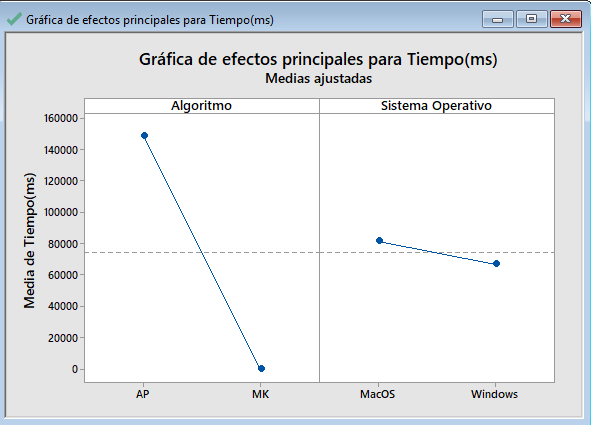


Por el contrario, cuando se realiza la comparación entre el procesador y el tipo de algoritmo utilizado, se evidencia que ambos son significativos en la respuesta y podría afectar los resultados.



Finalmente se procede a sacar la gráfica de efectos principales para el tiempo, en la cual se puede analizar y concluir según las medias, cuales son los factores que afectaron el programa con un mayor o menor tiempo de ejecución.





Con ayuda de las anteriores graficas en las que cada extremo del segmento indica la media del nivel, y la línea punteada horizontal indica la media general de todas las corridas, podemos concluir que las características óptimas con ambas comparaciones podrían ser: el uso del algoritmo MK ya que su media de tiempo de ejecución es mucho menor a la del AP, usar un Intel Core i5 5gen, y en cuanto al sistema operativo Windows presento menos media cuando se tomaron las muestras y por ende debería ser seleccionado.

Cabe recalcar que hay muchos factores no controlables que pueden afectar la respuesta, arrojando valores y resultados diferentes o variables en cada experimento.