

ISE-Hector-T4.pdf



mmr19



Ingeniería de Servidores



3º Grado en Ingeniería Informática



Escuela Técnica Superior de Ingenierías Informática y de Telecomunicación
Universidad de Granada



MÁSTER EN

Inteligencia Artificial & Data Management

MADRID

Formamos
talento para un futuro
Sostenible

saber más



Esto no son apuntes pero **tiene un 10 asegurado** (y lo vas a disfrutar igual).

Abre la Cuenta NoCuenta con el código **WUOLAH10**, haz tu primer pago y llévate 10 €.

Me interesa

1/6

Este número es indicativo del riesgo del producto, siendo 1/6 indicativo de menor riesgo y 6/6 de mayor riesgo.

ING BANK NV se encuentra adherido al Sistema de Garantía de Depósitos Holandés con una garantía de hasta 100.000 euros por depositante. Consulta más información en [ing.es](https://www.ing.es)



Ingeniería de Servidores

Apuntes de Teoría

INGENIERÍA INFORMÁTICA

2024

Consulta
condiciones aquí



do your thing

WUOLAH

ÍNDICE GENERAL

4	Capítulo 4	
	Análisis comparativo de rendimiento.	
4.1	Referenciación (benchmarking)	3
4.1.1	Eligiendo una medida	3
4.1.2	Carga real	4
4.1.3	Clasificación benchmark	5
4.2	Análisis de los resultados de un test de rendimiento	7
4.2.1	Medias	7
4.2.2	Conclusión	7
4.3	Comparación de prestaciones en presencia de aleatoriedad	8
4.3.1	Método 1	8
4.3.2	Método 2	9
4.3.3	Metodo 3	9
4.3.4	Ejercicio con tabla y como interpretarla	9
4.3.5	Estimar \bar{d}_{real}	10
4.4	Diseño de experimentos de comparación de rendimiento	10
4.4.1	Tipos de diseños experimentales	11
4.4.2	Diseños con un solo factor	11

1/6

Este número es indicativo del riesgo del producto, siendo 1/6 indicativo de menor riesgo y 6/6 de mayor riesgo.

ING BANK NV se encuentra adherido al Sistema de Garantía de Depósitos Holandés con una garantía de hasta 100.000 euros por depositante. Consulta más información en ing.es

Que te den **10 € para gastar**
es una fantasía.
ING lo hace realidad.

Abre la **Cuenta NoCuenta** con el código
WUOLAH10, haz tu primer pago y llévate 10 €.

Quiero el cash

[Consulta condiciones aquí](#)



do your thing

Ingeniería de Servidores



Comparte estos flyers en tu clase y consigue más dinero y recompensas



Banco de apuntes de la

WUOLAH

- 1** Imprime esta hoja
- 2** Recorta por la mitad
- 3** Coloca en un lugar visible para que tus compis puedan escanar y acceder a apuntes

- 4** Llévate dinero por cada descarga de los documentos descargados a través de tu QR



ANÁLISIS COMPARATIVO DE RENDIMIENTO.

4.1

Referenciación (benchmarking)

Características de un buen índice de rendimiento

- **Repetibilidad:** Siempre que se mida el índice en las **mismas condiciones**, el **valor** de éste debe ser **el mismo**.
- **Representatividad y fiabilidad:** Si un sistema A siempre presenta un **índice de rendimiento mejor** que el sistema B, es porque siempre **el rendimiento** real de A **es mejor** que el de B.
- **Consistencia:** El índice se debe poder medir en **cualquier sistema informático**, independientemente de su arquitectura o de su S.O.
- **Facilidad de medición:** Sería deseable que las **medidas** necesarias para obtener el índice sean **fáciles de tomar**.
- **Linealidad:** Sería deseable que, **si el índice de rendimiento aumenta**, el **rendimiento real** del sistema **aumente** en la misma proporción.

4.1.1. Eligiendo una medida

Tiempo de ejecución, frecuencia del reloj y CPI

No nos sirven porque hay sistemas con frecuencia peor que tienen mayores prestaciones. Las CPI y el tiempo de ejecución dependen del programa que se ejecute.

Esto no son apuntes pero **tiene un 10 asegurado** (y lo vas a disfrutar igual).

Abre la Cuenta NoCuenta con el código **WUOLAH10**, haz tu primer pago y llévate 10 €.

Me interesa

1/6

Este número es indicativo del riesgo del producto, siendo 1/6 indicativo de menor riesgo y 6/6 de mayor riesgo.

ING BANK NV se encuentra adherido al Sistema de Garantía de Depósitos Holandés con una garantía de hasta 100.000 euros por depositante. Consulta más información en [ing.es](https://www.ing.es)



MIPS

Los MIPS varían incluso en diferentes programas con el mismo ordenador y depende del juego de instrucciones del procesador (RISC, CISC).

MFLOPS

El formato de números en coma flotante varía entre arquitecturas y no representan una medida relevante para los servidores (potencialmente).

Elección final

Elegimos el tiempo de ejecución **teniendo en cuenta que dependerá de la carga que se use para la comparación.**

4.1.2. Carga real

Elegir la carga para comparar servidores es difícil porque:

- 1 Varía con el tiempo.
- 2 Es difícil reproducirla
- 3 Puede verse afectada por la velocidad del servidor.

Por eso usamos un modelo de carga (*test workload*). Este modelo tiene que ser **representativo de la carga real** y lo más **simple** posible.

Estrategias para modelos de carga

- **Personalizado** a partir de monitorización del sistema (caracterización de carga).
- **Genérico**: Cargas de prueba similares a lo que se quiere reproducir (referenciación o benchmarking).

Caracterización de la carga

Pasos

- 1 Elegimos **recursos** de la carga (CPU, RAM, disco...)
- 2 **Parámetros** de los recursos (lectura/escritura en RAM, ciclos de CPU...)
- 3 **Medir** valor de parámetros con monitores (**muestreo**)
- 4 **Analizar** datos
- 5 **Generar modelo** de carga seleccionando **representantes** + info estadística del tiempo.

Es un modelo **muy representativo** de la carga real, pero es bastante **tedioso** y **no conocemos** el **rendimiento** de **otros servidores** con nuestra carga.

Consulta condiciones aquí



do your thing

WUOLAH

Referenciación (benchmarking)

Definición 4.1.1 Benchmark program

Programa/s estándar para compara rendimiento entre sistemas informáticos.

Características

- Carga de prueba específica
- Conjunto de reglas para la correcta ejecución, obtención y validación de resultados.
- Índice de rendimiento (métrica) para realizar comparaciones.

Ventajas

- 1 Probable **encontrar** uno que reproduzca condiciones **parecidas** a las de nuestro servidor.
- 2 **Comparaciones** de rendimiento **justas**, porque siguen las reglas del benchmark.
- 3 Muchos permiten **ajustar la carga** para **medir la escalabilidad** de nuestro servidor.
- 4 Conocemos tanto el rendimiento para un benchmark de diferentes servidores como su diseño y configuración, obteniendo **info para mejorar** los nuestros.

4.1.3. Clasificación benchmark

Según estrategia de medida

- Miden tiempo ejecución de cantidad de tareas preestablecida. (La mayoría)
 - Miden cantidad de tareas para in tiempo preestablecido
- 1 **SLALOM**: Exactitud de solución para 1 min de ejecución.
 - 2 **TPC-C**: Consultas/seg de media en un servidor de base de datos.

Según generalidad del test

- **Microbenchmark** (para **componentes**): Para componentes o agrupaciones de componentes (CPU, caché, RAM)
- **Macrobenchmark** (para sistema **completo** o de **aplicación real**): Imita situaciones reales de algún área (comercio, servidor web, navegación...)

Ejemplos de microbenchmark

- *Whetstone*: Rendimiento de operaciones en coma flotante con aplicaciones científicas de sumas, multiplicaciones y funciones.
- *Linpark*: Rendimiento de operaciones en coma flotante con un algoritmo para resolver ecuaciones lineales.
- *Dhrystone*: Rendimiento de operaciones con enteros con copia y comparación de cadenas de caracteres.

- *Stream*: Ancho de banda de DRAM y SRAM.
- *IOzone*: Rendimiento de sistema de ficheros. También *HD Tune*, *fio* o *hdparm*.
- *Iperf*: Rendimiento TCP y UDP (cliente y servidor). *Pchar*, ancho de banda por salto IP.
- Paquetes de microbenchmarks:
 - AIDA64
 - Sandra
 - Phoronix Test Suite: open source, genera informes y test de rendimientos automáticamente. Recopila info de sensores del sistema en ejecución, permite visualizar el resultado de forma gráfica y exportarlo.

SPEC CPU

Fórmula 4.1.1 Para calcular SPEC

$$\sqrt[n]{\frac{t_1^{Ref}}{t_1^{base}} \times \frac{t_2^{Ref}}{t_2^{base}} \times \dots \times \frac{t_n^{Ref}}{t_n^{base}}} \quad (4.1)$$

Es un **conjunto de pruebas de rendimiento para procesadores** diseñadas por el Standard Performance Evaluation Corporation (**SPEC**). Se utiliza para **evaluar la velocidad y eficiencia** energética de los procesadores. Se distribuye como **imagen ISO** con el código, datasets, herramientas y documentación. La componen 4 conjuntos de microbenchmarks:

- SPECspeed@2017 Integer (**tiempo de respuesta** en aritmética **entera**)
- SPECspeed@2017 Floating Point (**tiempo de respuesta** en coma flotante)
- SPECrate@2017 Integer (**productividad** (cuántos programas/tiempo) en aritmética entera)
- SPECrate@2017 Floating Point (**productividad** en coma flotante)

Utiliza **programas reales** que tengan **gran portabilidad** entre arquitecturas (gcc, xz, lbm...).

Índices de prestaciones de SPECspeed Hay dos tipos:

- 1 **Base**: Todos los programas escritos en el mismo lenguaje usan las mismas opciones de compilación.
- 2 **Peak**: Permite que cada uno escoja las opciones de compilación óptimas.

Cada programa se ejecuta 3 veces y se escoge el medio (ni el mejor ni el peor). El **índice SPEC** es la **media geométrica** de las **ganancias en velocidad** con respecto a una **máquina de referencia**.

TPC

TPC es una organización sin ánimo de lucro especializada en benchmarks para procesar datos. Tiene varios tipos de benchmark:

Esto no son apuntes pero **tiene un 10 asegurado** (y lo vas a disfrutar igual).

Abre la Cuenta NoCuenta con el código **WUOLAH10**, haz tu primer pago y llévate 10 €.

Me interesa



1/6

Este número es indicativo del riesgo del producto, siendo 1/6 indicativo de menor riesgo y 6/6 de mayor riesgo.

ING BANK NV se encuentra adherido al Sistema de Garantía de Depósitos Holandés con una garantía de hasta 100.000 euros por depositante. Consulta más información en [ing.es](https://www.ing.es)

- **OLTP**: transacciones con requisitos de tiempo de respuesta.
- **DS**: Consultas complejas.
- **IoT**: Procesar muchos datos de muchos dispositivos.
- **Big Data**: Procesar mucha información usando clusters y apache hadoop.
- **Artificial Intelligence**: Ejecutar algoritmos de aprendizaje de machine learning.
- **Virtualization**: Ejecutar los benchmarks previos en entorno virtual.

Por ejemplo TPC-C es de tipo OLTP y hace peticiones que acceden a las bases de datos y ejecutan consultas. Saca transacciones que se procesan por segundo de media. Otro ejemplo es TPC-H, que es de tipo DS, exige examinar muchos datos para responder preguntas complejas.

4.2

Análisis de los resultados de un test de rendimiento

4.2.1. Medias

Necesitamos mostrar nuestras medidas en un **único valor**, para esto se suelen utilizar **medias**.

Media aritmética

Podemos hacerla normal (sumar todo y dividirlo por el número de muestras) o **ponderada** dándole peso a las medidas. Para usarla, simplemente, **la máquina que ejecuta los programas del benchmark en menor tiempo es la mejor**. Si la usamos **ponderada**, depende mucho de la máquina y su habilidad en **hacer bien las operaciones de los programas con mayor peso**.

Media geométrica

Vas multiplicando las ganancias y luego le haces la raíz de n, siendo n las medidas tomadas. Se utiliza en SPEC y SYSmark25. Lo haces **con respecto a la máquina de referencia**. Claro, según esto, lo que más importa es tener una mega ganancia respecto a los tiempos de la máquina de referencia. Mejorando **un solo aspecto**, tu índice speedup **es mayor que la competición**. Ver la imagen 4.1, índice de D es mejor que todos los demás siendo peor en general en otras medidas.

4.2.2. Conclusión

- 1 **La aritmética** es fácilmente interpretable, el menor valor indica que ejecuta el **conjunto de programas** en un tiempo menor.
- 2 **La aritmética ponderada** nos permite **priorizar algunos programas**, debe usarse **según** lo que **necesite** cada **usuario**.

Consulta condiciones aquí



do your thing

WUOLAH

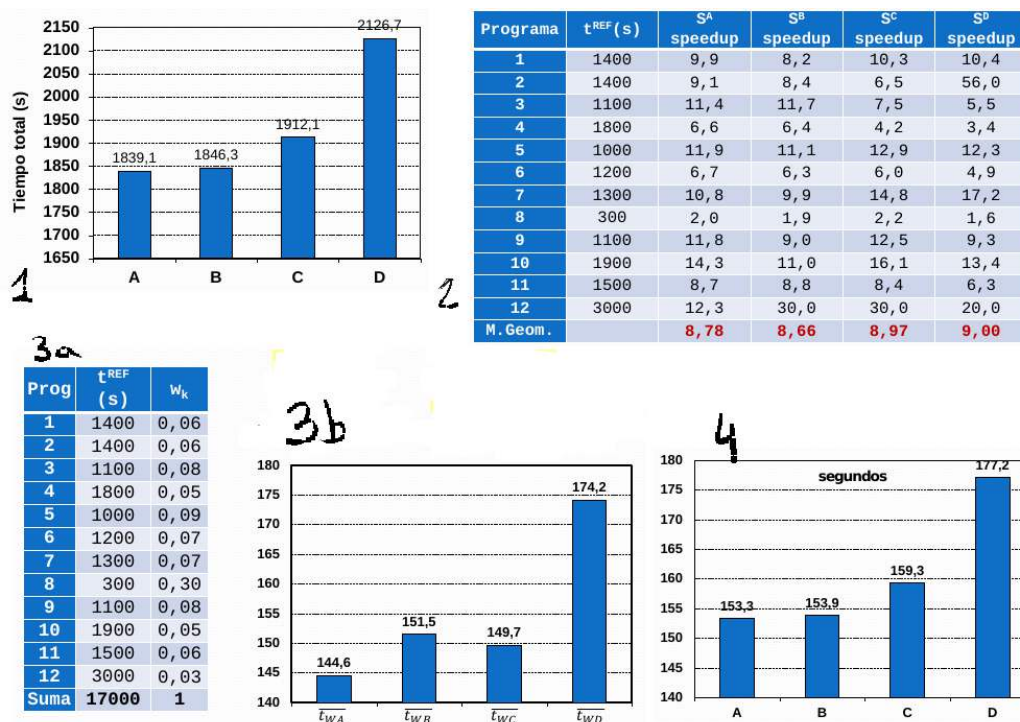


Figura 4.1: 1) Tiempo total de ejecución de 4 máquinas. 2) Índice speedup de las 4. 3a) Pesos de cada medida de referencia. 3b) Tiempos ponderados. 4) Media aritmética.

- 3 La media geométrica favorece mejoras sustanciales respecto a una máquina de referencia y no castiga al mismo nivel empeoramientos.

4.3

Comparación de prestaciones en presencia de aleatoriedad

Partimos de una tabla en la que tenemos una diferencia en tiempos de ejecución. Queremos saber si una es mejor que la otra, pero los tiempos de ejecución se ven afectados por **factores aleatorios**, así que necesitamos un método diferente a la media. Partimos de la **hipótesis nula** de que el rendimiento de las dos máquinas es equivalente, así que **las diferencias no deben ser significativas**.

4.3.1. Método 1

Tenemos un grado de significatividad α , que también te lo pueden dar en %. Si te dicen, por ejemplo, 95 %, entonces α es $1 - (95/100) = 0,05$. **Es 0.05 si no te dicen nada**. Para este método, nos dan un valor que se llama **p-value**. Si **p-value** $< \alpha$, rechazamos hipótesis nula (son diferentes los servidores y uno es mejor que otro).

4.3.2. Método 2

Imaginemos que **no sé el pvalor**. Me voy a la tabla. Busco el numerito ($t_{\frac{\alpha}{2}, n-1}$) de α y $df=n-1$, siendo **n el número de valores** que he medido. Si el tiempo experimental t_{exp} no pertenece a un intervalo de [-numerito, numerito], entonces rechazo hipótesis nula.

4.3.3. Metodo 3

Imaginemos que no tenemos pvalor ni tiempo experimental, pero tenemos un **intervalo de confianza** para \bar{d}_{real} . Este intervalo es:

$$\left[\bar{d} - \frac{s}{\sqrt{n}} \times t_{\frac{\alpha}{2}, n-1}, \bar{d} + \frac{s}{\sqrt{n}} \times t_{\frac{\alpha}{2}, n-1} \right] \quad (4.2)$$

Si 0 no pertenece a este intervalo, entonces rechazamos hipótesis nula

Fórmula 4.3.1 Tiempo experimental

$$t_{exp} = \frac{\bar{d} - \bar{d}_{real}}{s/\sqrt{n}} \sim T_{n-1} \quad (4.3)$$

Siendo \bar{d} :

$$\bar{d} = \frac{\sum_{i=1}^n d_i}{n} \quad (4.4)$$

Siendo s :

$$s = \sqrt{\frac{\sum_{i=1}^n (d_i - \bar{d})^2}{n-1}} \quad (4.5)$$

Siendo s/\sqrt{n} el error estándar

Nota ¿Me estudio las fórmulas o no?

*Esto no es un curso de estadística, evidentemente, así que nos facilitan muchos de los valores que necesitamos. No he visto ejercicios de examen que usen todas las formulas, pero sí usan la de SPEC y los 3 métodos que he descrito antes. Lo más importante es: **fórmula de SPEC, métodos de rechazar hipótesis nula y saber leer los valores que necesitamos en las imágenes que nos den**. En caso de que sí nos exigieran las fórmulas en el examen, no merece la pena, porque el temario es muy denso y este tema no vale tanto como el 2 o el 5. Invierte tu tiempo en estudiar las fórmulas del 5 y si te sobra tiempo después de todo, ven aquí y estúdiate esto.*

4.3.4. Ejercicio con tabla y como interpretarla

Si nos dieran un ejercicio con una tabla ya hecha, tenemos que tener en cuenta los métodos de antes pero sin necesitar las fórmulas. En la imagen vemos que pvalue para la primera comparación es menor que α , pero que en la segunda no. Así que en la primera, son diferentes servidores y

Esto no son apuntes pero **tiene un 10 asegurado** (y lo vas a disfrutar igual).

Abre la Cuenta NoCuenta con el código **WUOLAH10**, haz tu primer pago y llévate 10 €.

Me interesa

1/6

Este número es indicativo del riesgo del producto, siendo 1/6 indicativo de menor riesgo y 6/6 de mayor riesgo.

ING BANK NV se encuentra adherido al Sistema de Garantía de Depósitos Holandes con una garantía de hasta 100.000 euros por depositante. Consulta más información en [ing.es](https://www.ing.es)



Paired Samples T-Test		t_{exp}	degrees of freedom (n-1)		p-value	\bar{d}	Standard Error $\frac{s}{\sqrt{n}}$	95% CI for Mean Difference	
Measure 1	Measure 2	t	df	p	Mean Difference	SE Difference		Lower	Upper
EJ1_A	- EJ1_B	2.991	5	0.030	24.333	8.135		3.422	45.245
EJ2_A	- EJ2_B	1.163	4	0.310	2.400	2.064		-3.331	8.131

Intervalo de confianza (95%) para \bar{d}_{real} : $\bar{d} \pm \frac{s}{\sqrt{n}} \times t_{\frac{\alpha}{2}, n-1}$

Figura 4.2: Diferentes medidas que nos pueden dar hechas y su significado.

uno es mejor que el otro pero en la segunda no es significativo. Entonces, en la primera, vemos la diferencia. El primero menos el segundo da un valor positivo, así que el primer servidor tarda más tiempo que el segundo de media. Por lo tanto, el segundo servidor (Ej1_B) es mejor que el EJ1_A.

4.3.5. Estimar \bar{d}_{real}

Para estimar la media real y no depender de la muestral, podemos utilizar un intervalo de confianza que nos dice con una probabilidad de $(1 - \alpha) \cdot 100\%$ donde podría estar. El intervalo es:

$$\bar{d} \pm \frac{s}{\sqrt{n}} t_{\alpha/2, n-1} \quad (4.6)$$

4.4

Diseño de experimentos de comparación de rendimiento

Definición 4.4.1 Variable respuesta o dependiente

Valor cuya influencia en el resto queremos ver.

Definición 4.4.2 Factor

Variables que pueden afectar a la variable respuesta (ej. Sistema Operativo).

Definición 4.4.3 Nivel

Valores que puede tener un factor (ej. Windows, Linux).

Definición 4.4.4 Interacción

Efecto de un factor cambia para diferentes niveles de otro factor (Usar Windows gasta mas DRAM).

Consulta condiciones aquí



do your thing

WUOLAH

4.4.1. Tipos de diseños experimentales

- Diseños con **un solo factor**: Configuración base y un solo factor a la vez. Solo válida si no hay interacción entre factores.
- Diseños **multi-factoriales completos**: Se prueba cada combinación de nivel para todos los factores.
- Diseños **multi-factoriales fraccionados**: No todas las interacciones se reflejan, solo las más probables.

4.4.2. Diseños con un solo factor

Para poder analizar si cambiar un componente hardware afecta a nuestro rendimiento, hacemos un experimento con todas las posibles opciones. Para poder sacar si los cambios son significativos tendríamos que hacer un Test t para todos los cambios con todos (por ejemplo, si tenemos las interfaces de disco A, B y C tendríamos que hacer AB AC BC). Esto se vuelve cada vez más tedioso si queremos comparar muchas interfaces. Por eso, existe un test que nos **indica si el factor que estamos probando realmente importa**: el test **ANOVA** de un factor.

Tiene la **hipótesis nula** siguiente: **El factor no influye en el rendimiento**. Para saberlo, revisa si un valor pertenece a una determinada distribución (no hace falta saber cuál). Con su p-value, vemos si es menor que α . Si no lo es, el factor no influye a un $(1 - \alpha) \cdot 100\%$ de confianza. **Si influye**, comparamos las medidas entre sí con test t. A esto se le llama **prueba de múltiples rangos/comparaciones múltiples**.

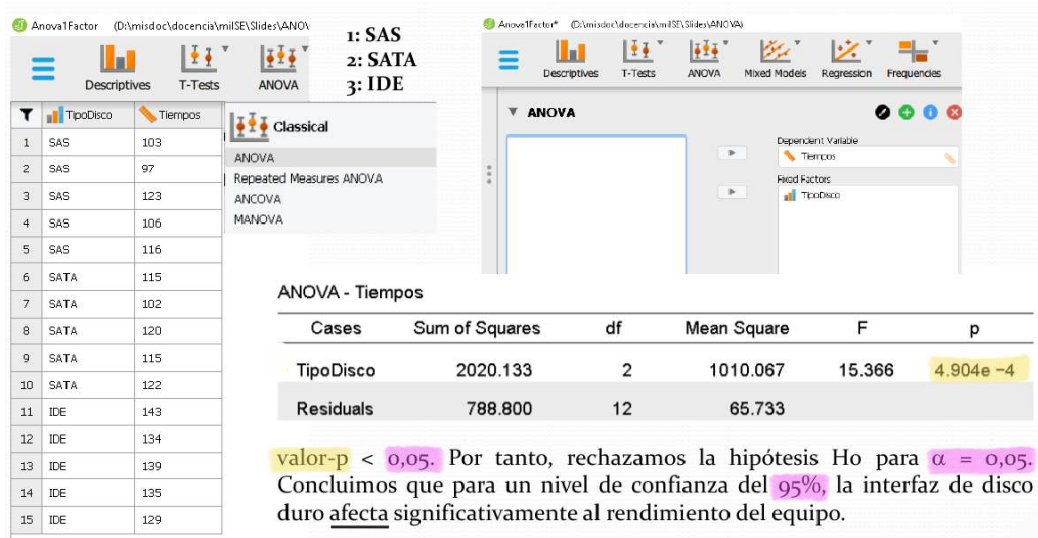


Figura 4.3: Muestra de cómo se podría saber la significancia de un factor con ANOVA y jasp.

$\alpha = 0,05$

Intervalo de confianza (95%) para \overline{d}_{real}

p-value de cada test t realizado

		95% CI for Mean Difference			SE	t	Ptukey
		Mean Difference	Lower	Upper			
SAS	SATA	-5.800	-19.480	7.880	5.128	-1.131	0.514
	IDE	-27.000	-40.680	-13.320	5.128	-5.266	5.399e-4
SATA	IDE	-21.200	-34.880	-7.520	5.128	-4.134	0.004

Figura 4.4: Un ejemplo de cómo haríamos el test de múltiples rangos. SAS y SATA tienen una pvalue mayor a alfa, así que no podemos rechazar la hipótesis de que tienen rendimientos equivalentes. Sin embargo, IDE tiene un pvalue mucho menor en las comparaciones restantes. Viendo la diferencia entre SATA y IDE, vemos que IDE tiene el peor rendimiento. En definitiva, podemos usar SAS o SATA.