
TEMA 3

Análisis comparativo del rendimiento (benchmarking)

PROBLEMA 3.1 En la tabla siguiente se muestra el tiempo de ejecución (expresado en segundos) y el número de instrucciones ejecutadas en el computador Cleopatra para cinco programas distintos.

Programa	Tiempo (s)	Instrucciones ($\times 10^6$)
asterix	56	543
obelix	59	346
panoramix	113	415
idefix	132	256
abracurcix	120	235

1. Calcule el número medio de MIPS de este computador.
2. Determine el número medio de ciclos por instrucción (CPI) obtenidos por este computador. Considere para ello que las instrucciones ejecutadas por los tres primeros programas duran 3 ciclos de media mientras que las del resto duran 5 ciclos.

SOLUCIÓN:

1. El computador obtiene 3,74 MIPS.
 2. El número medio de CPI es 3,55.
-

PROBLEMA 3.2 La tabla siguiente muestra el tipo y número de las operaciones de coma flotante ejecutadas por un programa de prueba en el computador MATES; la última columna representa el coste computacional en operaciones normalizadas.

Operación	Cantidad ($\times 10^9$)	Operaciones normalizadas
add.s, sub.s	456	1
div.s, mul.s	340	3
sqrt.s	180	12
sqrt.d	70	15
log.d	30	18

Se sabe que el programa tarda una hora en ejecutarse. Indique el rendimiento de este computador mediante el uso de MFLOPS y MFLOPS normalizados. ¿Existe mucha diferencia entre ambos valores?

SOLUCIÓN: El programa obtiene 299 MFLOPS y 1452 MFLOPS normalizados.

PROBLEMA 3.3 Considere la información (incompleta) obtenida por la orden siguiente en un computador sin más carga que la ejecución de esta orden y sin operaciones de E/S:

```
$ time simulador
  real 0m130s
  user -----s
  sys 0m5s
```

Se sabe que el número de instrucciones ejecutadas es de 32×10^9 ; de estas últimas, el 60 % se ejecuta en dos ciclos, mientras que el resto lo hace en cinco ciclos. Calcule el número medio de ciclos por instrucción (CPI) obtenidos por el programa, la frecuencia de funcionamiento del procesador y los MIPS alcanzados por el procesador.

SOLUCIÓN: El programa obtiene un CPI de 3,2 y 246,2 MIPS. La frecuencia del procesador es de 0,788 GHz. ■

PROBLEMA 3.4 La tabla siguiente muestra los tiempos de ejecución en segundos de tres programas de prueba en tres máquinas A, B y C. Aplíquense al menos dos técnicas de análisis que permitan extraer conclusiones contradictorias respecto del rendimiento de las máquinas.

Programa	A	B	C
mafalda	185	164	126
felipe	161	163	143
miguelito	182	110	295

SOLUCIÓN: La máquina más rápida es la B. Para obtener una conclusión diferente bastaría con calcular un promedio ponderado dando más peso a la máquina que se quisiera beneficiar o bien normalizando los valores respecto de esta misma máquina. ■

PROBLEMA 3.5 La tabla que se muestra a continuación refleja los tiempos de ejecución, en segundos, de los 14 programas de prueba que integran un determinado benchmark empleado para el cálculo del rendimiento en aritmética de coma flotante. En particular, los tiempos corresponden a la máquina de referencia y a una máquina que denominaremos A (columnas “Base” y “Peak”, con el mismo significado que usa SPEC para sus comparaciones).

Programa	Referencia	A-Base	A-Peak
168.wupwise	1600	419	300
171.swim	3100	562	562
172.mgrid	1800	607	607
173.applu	2100	658	605
177.mesa	1400	273	242
178.galgel	2900	571	571
179.art	2600	1040	1038
183.quake	1300	501	387
187.facerec	1900	434	434
188.amp	2200	705	697
189.lucas	2000	784	758
191.fma3d	2100	534	534
200.sixtrack	1100	395	336
301.apsi	2600	866	866

1. Calcúlese los índices SPECfp_base y SPECfp de la máquina A según el criterio de SPEC.
2. Para la columna A-Base, si se considera el tiempo total de ejecución, ¿cuántas veces es más rápida la máquina A que la máquina de referencia?
3. ¿Qué mejora del rendimiento se obtiene utilizando las opciones de optimización que ofrece el compilador?

SOLUCIÓN:

1. SPECfp_base = 3,48 y SPECfp = 3,74.
2. La máquina A es 3,44 veces más rápida que la de referencia.
3. La optimización del compilador permite mejorar 1,05 veces el tiempo de ejecución (un 5%).

PROBLEMA 3.6 Considere los tiempos de ejecución, en segundos, obtenidos en los computadores R (referencia), A y B para un conjunto de cinco programas de prueba:

Programa	R (s)	A (s)	B (s)
tinky-winky	2600	503	539
dipsy	2100	654	762
laa-laa	9800	707	716
po	2300	748	760
noo-noo	1800	363	235

1. Compare el rendimiento de A y B utilizando el tiempo total de ejecución.
2. Calcule, a la manera de SPEC, un índice de rendimiento para A y B, y compare el rendimiento de ambas máquinas con este índice. ¿Obtiene los mismos resultados que en el apartado anterior?

SOLUCIÓN:

1. Según el tiempo total de ejecución, la máquina A es 1.012 veces más rápida que B.
2. Los índices calculados a la manera de SPEC para las máquinas A y B son, respectivamente, 5.12 y 5.31; en consecuencia, según este índice la máquina B es 1,04 veces mejor que la A.

PROBLEMA 3.7 La siguiente tabla muestra los tiempos de ejecución expresados en minutos de una serie de programas de prueba en dos sistemas informáticos SI1 y SI2. La última columna muestra el número de instrucciones ejecutadas por cada programa.

Programa	SI1	SI2	Instrucciones
charlie brown	35	70	$3,55 \times 10^8$
lucy	101	78	$7,78 \times 10^{13}$
linus	57	55	$9,12 \times 10^7$
patty	76	83	$2,94 \times 10^{10}$

1. Suponiendo que todos los programas tienen la misma importancia en este estudio de evaluación, compárense las prestaciones de estos dos sistemas según:
 - a) Media aritmética de los tiempos de ejecución.
 - b) MIPS (millones de instrucciones ejecutadas por segundo).

- Repítase la primera parte del estudio suponiendo que los pesos atribuidos a cada programa de prueba son, respectivamente: 0,5, 0,1, 0,1, y 0,3.

SOLUCIÓN:

- El sistema SI1 es 1,06 más rápido que SI2 atendiendo al tiempo de ejecución. Los MIPS obtenidos por ambos sistemas son, respectivamente, 4822,2 y 4535,5.
- En este nuevo escenario el sistema SI1 es 1,3 veces más rápido que SI2 atendiendo al tiempo de ejecución.

PROBLEMA 3.8 A continuación se muestran los resultados obtenidos tras la ejecución de tres programas de prueba en un computador que dispone de un procesador con un reloj de 2 GHz:

Programa	Instrucciones	Ciclos por instrucción	Operaciones coma flotante
1	150×10^9	3,5	50×10^6
2	35×10^9	2,8	20×10^6
3	250×10^9	5,2	175×10^6

Indique, a partir de datos anteriores, los siguientes índices de prestaciones de este computador: MIPS, MFLOPS y CPI.

SOLUCIÓN: 452,42 MIPS, 0,25 MFLOPS y 4,4 CPI.

PROBLEMA 3.9 La página oficial de SPEC muestra los siguientes resultados de rendimiento para dos sistemas informáticos de la casa comercial ACER obtenidos mediante el conocido benchmark CPU2006:

Sistema	Modelo	SPECint_base2006	SPECint2006
A	Altos G5350 (AMD Opteron 246)	13,47	14,38
B	Altos G5350 (AMD Opteron 254)	17,88	19,18

- ¿Cuál de los dos sistemas presenta mejor rendimiento? Cuantifique numéricamente la mejora.
- A la vista de los resultados anteriores, ¿afecta al rendimiento de ambos sistemas la optimización llevada a cabo por el compilador en las pruebas de evaluación?
- ¿En qué medida se reflejará en los resultados anteriores una mejora importante en la unidad de coma flotante (FPU, floating point unit) del procesador?
- ¿Cuál de los dos sistemas ejecutará el benchmark *Whetstone* más rápidamente?

SOLUCIÓN:

- El sistema B obtiene un mayor rendimiento. En particular, las mejoras sobre el sistema A en los índices SPECint_base y SPECint son, respectivamente, 1,33 y 1,33 (se obtiene la misma ganancia en ambos casos).

2. La optimización de la compilación permite obtener una mejora, en ambos casos, de 1,07.
3. En teoría no afectará porque los índices mostrados afectan únicamente a la aritmética entera.
4. No se puede saber porque este benchmark es de aritmética en coma flotante.

■

PROBLEMA 3.10 Responda brevemente a las siguientes cuestiones sobre el benchmark CPU2006 que ha desarrollado el consorcio SPEC:

1. ¿Qué componentes del sistema informático evalúa?
2. ¿Cuáles son los lenguajes en que están programados los diferentes programas que lo integran?
3. ¿Cuál es la diferencia entre los índices SPECint2006 y SPECint_base2006?
4. Indique cómo se calcula el índice SPECfp2006. El método de cálculo empleado, ¿satisface todas las exigencias de un buen índice de prestaciones? Razone la respuesta.

SOLUCIÓN:

1. El procesador, el sistema de memoria y el compilador.
2. C, C++ y Fortran.
3. El primero se obtiene con los programas compilados, cada uno de ellos, con parámetros que optimizan la ejecución del código en la máquina que se evalúa con el objetivo de conseguir el menor tiempo de ejecución posible (rendimiento pico). El segundo utiliza opciones de compilación genéricas y comunes a todos los programas.
4. Se usa la media geométrica de los ratios obtenidos dividiendo los tiempos de ejecución en la máquina de referencia con los de la máquina que se evalúa, es decir, la media geométrica de las ganancias en velocidad con respecto a una máquina de referencia. Este método de cálculo no satisface las exigencias de un buen índice de prestaciones ya que no refleja de manera correcta la comparación basada en los tiempos de ejecución.

■

PROBLEMA 3.11 En un computador se ha llevado a cabo un estudio para determinar si el tipo de memoria principal es un factor importante en su rendimiento. Para ello se ha medido el tiempo de ejecución de seis programas con dos tipos de memoria: MA (más rápida y más cara) y MB (más lenta y más barata). Las medidas de los tiempos de ejecución (en segundos) de los programas son los siguientes:

Programa	MA	MB
lucho	45	48
lupita	32	35
lulila	51	56
lurdo	43	49
lutecio	48	51

Calcule si las diferencias observadas son significativas al 95% de confianza y, en caso afirmativo, determine la mejora conseguida en el rendimiento debido al uso del tipo de memoria más rápida. DATO: $|t_{0.025, 4}| = 2.78$.

SOLUCIÓN: Las diferencias son significativas. La memoria MA permite obtener una mejora del 9% en el rendimiento sobre MB.

■

PROBLEMA 3.12 La empresa Facebook está estudiando dos grandes propuestas con el objetivo de actualizar los computadores personales de su oficina principal en Menlo Park, California. El precio de cada computador es de 1850€ para el Modelo A y 2200€ para el Modelo B. Los responsables informáticos de la empresa han ejecutado los ocho programas que utilizan habitualmente en un computador de cada propuesta, y han obtenido los tiempos de ejecución, expresados en segundos, que se muestran a continuación:

Programa	Modelo A	Modelo B
1	23,6	24,0
2	33,7	41,6
3	10,1	8,7
4	12,9	13,5
5	67,8	66,4
6	9,3	15,2
7	47,4	50,5
8	54,9	52,3

Determinése si existen diferencias significativas en el rendimiento de los computadores personales de las dos propuestas y qué opción sería mejor. DATO: $|t_{0,025, 7}| = 2.365$.

SOLUCIÓN: El intervalo de confianza es $[-4,696, 1,576]$. Como este intervalo incluye el 0 podemos afirmar que las diferencias observadas en los tiempos de ejecución no son significativas. En consecuencia, la mejor opción para actualizar los computadores de la empresa es la opción A, ya que resulta menos costosa.

■

PROBLEMA 3.13 A continuación se muestran los tiempos de ejecución (en segundos) medidos en tres computadores, A, B y R, para un conjunto de cinco programas de prueba:

Programa	A	B	R
1	96,2	95,3	103,9
2	13,1	10,2	53,8
3	79,6	67,4	156,3
4	45,2	51,8	98,1
5	88,3	89,3	238,5

Calcúlese el índice de prestaciones de las máquinas A y B según se hace en el benchmark SPEC_CPU, tomando como referencia la máquina R. Compárese el rendimiento de estas máquinas atendiendo tanto a este índice como al tiempo total de ejecución. ¿Hay diferencias significativas con un grado de confianza del 95%? DATO: $|t_{0,025, 4}| = 2.78$.

SOLUCIÓN: La media geométrica de los tiempos de ejecución normalizados respecto de la máquina R son 2,20 y 2,32 para A y B respectivamente. En cualquier caso, estos datos permiten concluir que B rinde 1,06 veces más que A. En cambio la suma de los tiempos de ejecución son 322,4s y 314,0s respectivamente, lo que rebaja la mejora conseguida a 1,03. Respecto a las diferencias, éstas no son significativas al 0.05 de significatividad porque el intervalo de confianza incluye el 0.

PROBLEMA 3.14 Una gran empresa de seguros está estudiando dos propuestas con el objetivo de actualizar los computadores de su instalación informática. El precio de cada computador es de 1300€ para los de tipo A y 1450€ para los de tipo B. Se estima que el número de computadores a reemplazar es de 75. El ingeniero informático jefe de la empresa ha mandado ejecutar cinco de los programas que utilizan habitualmente en un computador de cada propuesta y ha obtenido los tiempos de ejecución, expresados en segundos, que se muestran a continuación:

Programa	Propuesta A	Propuesta B
1	23.6	24.5
2	33.7	41.6
3	10.1	6.6
4	12.9	13.7
5	67.8	66.4

- Determine si existen diferencias significativas (para un nivel de confianza del 95%) en el rendimiento de los computadores de las dos propuestas y qué opción sería la que compraría según esa información. Justifique la respuesta. DATO: $|t_{0.025, 4}| = 2.78$.
- Calcúlese el índice de prestaciones de las máquinas A y B según se hace en el benchmark SPEC_CPU, tomando como referencia la máquina A. Según ese índice, ¿qué opción es la más rápida? ¿qué opción sería la que compraría?

SOLUCIÓN: a) Las diferencias en los tiempos de ejecución entre la propuesta A y la B ($t_A - t_B$) están en el intervalo $[-6.3, 4.4]$ por lo que no son significativas al 95%. De este modo, deberíamos comprar la más barata (propuesta A). b) $SPEC_A=1$, $SPEC_B=1.027$. Según ese índice, la opción más rápida es la B. Mirando la relación prestaciones/coste, siendo las prestaciones el valor del SPEC_CPU, la opción que deberíamos comprar es la A ($7.69E-4 \text{ €}^{-1}$ frente a $7.08E-4 \text{ €}^{-1}$).

PROBLEMA 3.15 En la empresa KANDOR GRAPHICS están intentando mejorar la técnica de distribución de carga de su servidor principal de streaming de vídeo. Para ello, han realizado cinco medidas de la productividad media del servidor durante un número determinado de horas para las 2 configuraciones principales de distribución de carga: *First Alive* y *Round Robin*. Los resultados, expresados como número medio de MB transmitidos por segundo son los que aparecen a continuación:

Nº Experimento	FIRST_ALIVE (MB/s)	ROUND_ROBIN (MB/s)
1	157	165
2	125	123
3	172	185
4	152	158
5	165	172

Para poder estar seguros de la decisión, los ingenieros informáticos de la empresa realizaron un test t de Student sobre estos datos, obteniéndose los siguientes resultados:

Paired Samples Test

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	90% Confidence Interval of the Difference				
				Lower	Upper			
FIRST_ALIVE - ROUND_ROBIN	-6,40	5,413	2,421	-11,56	-1,24	-2,64	4	,057

Paired Samples Test

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	99% Confidence Interval of the Difference				
				Lower	Upper			
FIRST_ALIVE - ROUND_ROBIN	-6,40	5,413	2,421	-17,55	4,75	-2,64	4	,057

- A la vista de los resultados y para un 90% de confianza. ¿Qué método utilizaría y por qué? Determine la mejora conseguida en la productividad del mejor método como “número de veces mayor que” y como “% mayor”.
- ¿Qué conclusiones podríamos extraer para un 99% de confianza? Razone de forma general por qué y en qué medida puede afectar el % de confianza a la decisión a tomar.

SOLUCIÓN: a) Al 90% de nivel de confianza las diferencias sí son significativas. Deberíamos elegir el de mayor productividad media. En este caso, el método de Round Robin con 160.6 MB/s de productividad media frente a los 154.2 MB/s del First Alive, por lo que es 1.04 veces mayor (4% mayor). b) Al 99% ya no podríamos asegurar que las diferencias son significativas por lo que consideraríamos ambos métodos iguales. Al aumentar el % de confianza lo que realmente estoy haciendo es aumentar el ancho de los intervalos de confianza de tal forma que cada vez sea más difícil descartar la hipótesis nula de que ambos métodos sean iguales. Se debe hacer notar que de cara a extraer conclusiones acerca de la significatividad de las diferencias, sobra una de las tablas ya que el valor-p (P-value) de 0.057 habría bastado para deducir que con el 90% de nivel de confianza las diferencias son significativas pero con el 99% ya no lo serían. ■

PROBLEMA 3.16 En KINGSTON están intentando comprobar la mejora en la latencia de los módulos de memoria RAM que introduce la nueva técnica HyperX. Para ello, han realizado 100 experimentos para calcular las latencias medias en múltiples diferentes contextos. Finalmente, para comprobar que las diferencias en las latencias entre el método tradicional y la técnica HyperX no se deben a efectos aleatorios, han realizado un test t, cuyos resultados son los que aparecen en la siguiente tabla (las latencias han sido medidas en ns). A partir de dicha tabla, indique:

Paired Samples Test									
		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	NORMAL - HYPERX	,33230	,72479	,07248	,18849	,47611	4,585	99	,000

- Mirando únicamente valores medios, ¿qué técnica parece ser mejor? Justifique la respuesta.
- Al 95% de nivel de confianza, ¿son significativas esas diferencias? Justifique la respuesta.

SOLUCIÓN:

- La diferencia entre las latencias medias de ambas técnicas es 0.33 ns, medida como latencia_Normal – latencia_HyperX. Por tanto, la latencia media usando el módulo de memoria Normal es 0.33ns mayor que la del módulo HyperX. Como una latencia mayor supone un rendimiento peor, la técnica mejor en este caso es la HyperX.
- Las diferencias sí son significativas al 95%. Podemos verlo fácilmente comprobando que el 0 no está incluido en el intervalo de confianza del 95% de las diferencias entre las latencias de ambas técnicas. También se puede comprobar viendo que el p-value (celda Sig. (2-tailed)), que mide la probabilidad de que ambas técnicas sean iguales es 0,000, claramente inferior al valor umbral que marca el nivel de significatividad (0.05).



PROBLEMA 3.17 En la empresa SERENDIPITY S.L. están intentando mejorar el servidor web que alberga las páginas de la Universidad de Granada. Para ello, han ejecutado un conocido benchmark de servidores web para 5 configuraciones distintas del S.O. actualmente en uso. Como la fuente de variabilidad es alta debido a que las pruebas han tenido que realizarlas en el equipo ya actualmente en uso (se ha elegido el intervalo entre las 4 y las 5 de la mañana en días sucesivos) los experimentos se han realizado 10 veces. Los resultados del número medio de páginas servidas por segundo y las correspondientes tablas ANOVA son los que aparecen a continuación:

Exp.	Conf. 1	Conf. 2	Conf. 3	Conf. 4	Conf. 5
1	15,2	15,5	17,8	16,2	17,8
2	16,2	15,2	18,5	15,7	17,9
3	16,5	16,3	17,9	15,3	18,1
4	15,9	16,2	18,9	15,8	18,2
5	14,8	15,4	18,5	16,2	18,9
6	15,2	15,2	18,1	15,8	18,3
7	15,6	15,8	19,5	15,2	18,8
8	16,0	16,0	18,5	14,9	17,8
9	16,3	15,2	19,4	14,9	18,2
10	15,3	15,5	19,7	15,0	18,1

- a) Si atendiéramos exclusivamente a la media aritmética de los resultados, ¿qué configuración parecería la mejor?
- b) Para un nivel de confianza del 95%, ¿afecta la configuración del S.O. al rendimiento del equipo?
- c) Para un nivel de confianza del 95%, agrupe las configuraciones que afectan estadísticamente por igual. ¿Cuáles serían, en ese caso, las mejores configuraciones? ¿Y para un nivel de confianza del 99%? Explique razonadamente los resultados.

ANOVA

Tiempos

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	97,757	4	24,439	90,695	,000
Within Groups	12,126	45	,269		
Total	109,883	49			

Multiple Comparisons

Dependent Variable: Tiempos

LSD

(I) Grupo	(J) Grupo	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	,0700	,2321	,764	-,398	,538
	3	-2,9800*	,2321	,000	-3,448	-2,512
	4	,2000	,2321	,394	-,268	,668
	5	-2,5100*	,2321	,000	-2,978	-2,042
2	1	-,0700	,2321	,764	-,538	,398
	3	-3,0500*	,2321	,000	-3,518	-2,582
	4	,1300	,2321	,578	-,338	,598
	5	-2,5800*	,2321	,000	-3,048	-2,112
3	1	2,9800*	,2321	,000	2,512	3,448
	2	3,0500*	,2321	,000	2,582	3,518
	4	3,1800*	,2321	,000	2,712	3,648
	5	,4700*	,2321	,049	,002	,938
4	1	-,2000	,2321	,394	-,668	,268
	2	-,1300	,2321	,578	-,598	,338
	3	-3,1800*	,2321	,000	-3,648	-2,712
	5	-2,7100*	,2321	,000	-3,178	-2,242
5	1	2,5100*	,2321	,000	2,042	2,978
	2	2,5800*	,2321	,000	2,112	3,048
	3	-,4700*	,2321	,049	-,938	-,002
	4	2,7100*	,2321	,000	2,242	3,178

*. The mean difference is significant at the 0.05 level.

Multiple Comparisons

Dependent Variable: Tiempos

LSD

(I) Grupo	(J) Grupo	Mean Difference (I-J)	Std. Error	Sig.	99% Confidence Interval	
					Lower Bound	Upper Bound
1	2	,0700	,2321	,764	-,554	,694
	3	-2,9800*	,2321	,000	-3,604	-2,356
	4	,2000	,2321	,394	-,424	,824
	5	-2,5100*	,2321	,000	-3,134	-1,886
2	1	-,0700	,2321	,764	-,694	,554
	3	-3,0500*	,2321	,000	-3,674	-2,426
	4	,1300	,2321	,578	-,494	,754
	5	-2,5800*	,2321	,000	-3,204	-1,956
3	1	2,9800*	,2321	,000	2,356	3,604
	2	3,0500*	,2321	,000	2,426	3,674

4	4	3,1800*	,2321	,000	2,556	3,804
	5	,4700	,2321	,049	-,154	1,094
	1	-,2000	,2321	,394	-,824	,424
	2	-,1300	,2321	,578	-,754	,494
	3	-3,1800*	,2321	,000	-3,804	-2,556
5	5	-2,7100*	,2321	,000	-3,334	-2,086
	1	2,5100*	,2321	,000	1,886	3,134
	2	2,5800*	,2321	,000	1,956	3,204
	3	-,4700	,2321	,049	-1,094	,154
	4	2,7100*	,2321	,000	2,086	3,334

*. The mean difference is significant at the 0.01 level.

SOLUCIÓN: a) La mejor configuración sería la número 3 con una media de 18,68 páginas servidas por segundo. b) Tras un análisis de ANOVA de 1 factor, el valor del estadístico F es de 90,7, con una probabilidad mucho menor de 0.05 de que pertenezca a una distribución F con 4 y 45 grados de libertad. Por tanto, debemos descartar que el S.O. no tenga influencia sobre el rendimiento del servidor web --> sí afecta. c) Al 95%, un test múltiple de comparación de medias nos indica que las configuraciones 1, 2 y 4 afectan de igual forma al rendimiento. Igualmente, nos indica que las configuraciones 3 y 5 afectan de forma diferente, siendo la número 3 la mejor, seguida por la número 5. Si quisiéramos tener un 99% de confianza, sin embargo, no podríamos descartar que las configuraciones 3 y 5 afecten de forma diferente. En ese caso, ambas serían igualmente las mejores.

