

Entrega de ejercicios. **Tema 4**

Análisis comparativo del rendimiento (*benchmarking*)

Miguel Ángel Fernández Gutiérrez

Problema 4.1

Problema. 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.

1. Calcule el número medio de MIPS de este computador al ejecutar los 5 programas.
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.

Programa	Tiempo (s)	Instrucciones ($\times 10^6$)	Ciclos/instrucción	Ciclos ($\times 10^6$)
asterix	56	543	3	1629
obelix	59	346	3	1038
panoramix	113	415	3	1245
idefix	132	256	5	1280
abraracurcix	120	235	5	1175
	480	1795		6367

Solución

Apartado 1. Basta calcular:

$$\text{MIPS} = \frac{\text{número total de instrucciones } (\times 10^6)}{\text{tiempo total de ejecución (s)}} = \frac{1795}{480} = 3.74 \text{ MIPS}$$

Apartado 2. Basta calcular:

$$\text{CPI} = \frac{\text{número total de ciclos } (\times 10^6)}{\text{número total de instrucciones } (\times 10^6)} = \frac{6367}{1795} = 3.55 \text{ CPI}$$

mianfg

Problema 4.5

Problema. 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).

1. Calcúlense 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?

Programa	Referencia	A-Base	A-Peak	Ref./A-Base	Ref./A-Peak
168.wupwise	1600	419	300	3.8186	5.3333
171.swim	3100	562	562	5.5160	5.5160
172.mgrid	1800	607	607	2.9654	2.9654
173.applu	2100	658	605	3.1915	3.4711
177.mesa	1400	273	242	5.1282	5.7851
178.galgel	2900	571	571	5.0788	5.0788
179.art	2600	1040	1038	2.5000	2.5048
183.equake	1300	501	387	2.5948	3.3592
187.facerec	1900	434	434	4.3779	4.3779
188.amp	2200	705	697	3.1206	3.1564
189.lucas	2000	784	758	2.5510	2.6385
191.fma3d	2100	534	534	3.9326	3.9326
200.sixtrack	1100	395	336	2.7848	3.2738
301.apsi	2600	866	866	3.0023	3.0023
	(suma)	(suma)		(media geométrica)	
	28700	8349	7937	3.4825	3.7419

Solución

Apartado 1. El parámetro SPECfp_base se calcula haciendo la media geométrica de las ganancias respecto al sistema de referencia de las medidas *base*, que es (ver tabla) 3.4825. Del mismo modo, el parámetro SPECfp_peak se calcula realizándolo con las medidas *peak*, que es (ver tabla) 3.7419.

Apartado 2. Considerando el tiempo total de ejecución, vemos que la máquina A es más rápida que la de referencia unas 3.44 veces:

$$\text{ganancia}_{\text{Ref/A}} = \frac{28700}{8349} = 3.4375$$

Apartado 3. Obtenemos una mejora de rendimiento de 1.05 veces el tiempo de ejecución, es decir, de un 5 %:

$$\text{ganancia}_{\text{A-Base/A-Peak}} = \frac{8349}{7937} = 1.0519$$

Problema 4.6

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

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?

Programa	R (s)	A (s)	B (s)	R/A	R/B
tinky-winky	2600	503	539	5.1690	4.8237
dipsy	2100	654	762	3.2110	2.7559
laa-laa	9800	707	716	13.8614	13.6872
po	2300	748	760	3.0749	3.0263
noo-noo	1800	363	235	4.9587	7.6596
	(suma)	(suma)		(media geométrica)	
	18600	2975	3012	5.1169	5.3090

Solución

Apartado 1. De acuerdo al tiempo total de ejecución, vemos que la máquina A es $\frac{3012}{2975} = 1.0124$ veces más rápida que B.

Apartado 2. Del mismo modo que en el problema anterior, calculamos la media geométrica de los cocientes de las medidas respecto al ordenador de referencia. De este modo, vemos que el rendimiento de la máquina A es de 5.1169 y el de la máquina B es 5.3090. Vemos que la máquina B, de acuerdo a este índice calculado, es $\frac{5.3090}{5.1169} = 1.0375$ veces mejor que la máquina A. Queda evidenciado, por tanto, que no obtenemos los mismos resultados.

Problema 4.11

Problema. 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 las que aparecen a continuación.

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: $|t_{0.025,4}| = 2.78$

Programa	t_A	t_B	$d = t_B - t_A$
lucho	45	48	3
lupita	32	35	3
lulila	51	56	5
lurdo	43	49	6
lutecio	48	51	3
	(suma)		$\bar{d} = 4$
	219	239	$s = 1.4142$

Solución

Para determinar si las diferencias son significativas al 95 % de confianza ($\alpha = 0.05$), calculamos t_{exp} , teniendo en cuenta que

$$\bar{d} = 4 \quad s = 1.4142 \quad n = 5 \quad \Rightarrow \quad t_{exp} = \frac{\bar{d}}{s/\sqrt{n}} = \frac{4}{1.4142/\sqrt{5}} = 6.3246$$

Vemos que t_{exp} no está en el intervalo de confianza $[-|t_{\frac{\alpha}{2}, n-1}|, |t_{\frac{\alpha}{2}, n-1}|]$, pues:

$$6.3246 \notin [-2.78, 2.78]$$

De este modo, vemos que la media real, \bar{d}_{real} , no puede ser cero, teniendo las memorias diferencias significativas en su rendimiento al 95 % de confianza.

A continuación, determinamos la mejora del rendimiento mediante el tiempo total (nótese que, de hecho, no tenemos unos datos de referencia para calcular el rendimiento del modo SPEC_CPU), teniendo la memoria A un rendimiento $\frac{239}{219} = 1.0913$ veces superior al de la memoria B, es decir, que la memoria MA permite obtener una mejora del 9 % respecto de la memoria MB.

Problema 4.13

Problema. 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.

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: $|t_{0.025,4}| = 2.78$

Programa	t_R	t_A	t_B	t_R/t_A	t_R/t_B	$d = t_A - t_B$
1	103.9000	96.2000	95.3000	1.0800	1.0902	0.9000
2	53.8000	13.1000	10.2000	4.1069	5.2745	2.9000
3	156.3000	79.6000	67.4000	1.9636	2.3190	12.2000
4	98.1000	45.2000	51.8000	2.1704	1.8938	-6.6000
5	238.5000	88.3000	89.3000	2.7010	2.6708	-1.0000
		(suma)		(media geométrica)		$\bar{d} = 1.6800$
		322.4000	314.0000	2.1959	2.3216	$s = 6.8649$

Solución

En la tabla aparecen calculadas las medias geométricas de los cocientes entre los tiempos de ejecución de los programas para la máquina de referencia y las máquinas A y B. Haciendo las medias geométricas de éstos, vemos que la media geométrica de los tiempos de ejecución normalizados respecto a R son 2.1959 y 2.3216 para A y para B, respectivamente. Por tanto, de acuerdo a esto la máquina B rinde mejor que la máquina A $\frac{2.3216}{2.1959} = 1.0573$ veces. Sin embargo, si atendemos al tiempo de ejecución total en cada máquina, que es de 322.4 s para A y 314 s para B, veríamos que la mejora es de $\frac{314}{322.4} = 1.0268$ veces, menor que la que hemos obtenido antes.

Veremos a continuación si ambas máquinas tienen rendimientos equivalentes. Para que fuese así al 95 % de confianza ($\alpha = 0.05$) debería ser que

$$t_{exp} = \frac{\bar{d} - \bar{d}_{real}}{s/\sqrt{n}} \in [-|t_{\frac{\alpha}{2}, n-1}|, |t_{\frac{\alpha}{2}, n-1}|] \iff \bar{d}_{real} \in \left[\bar{d} - \frac{s}{\sqrt{n}} \cdot |t_{\frac{\alpha}{2}, n-1}|, \bar{d} + \frac{s}{\sqrt{n}} \cdot |t_{\frac{\alpha}{2}, n-1}| \right]$$

Y nuestra hipótesis será cierta si puede ser $\bar{d}_{real} = 0$, es decir, si el intervalo anterior incluye al cero. De la tabla y de los datos del enunciado podemos obtener los valores necesarios:

$$\bar{d} = 1.6800 \quad s = 6.8649 \quad n = 5 \quad |t_{0.025,4}| = 2.78 \quad \Rightarrow \quad \bar{d}_{real} \in [-6.8548, 10.2148]$$

Y como el intervalo de confianza contiene al 0, tenemos que las diferencias no son significativas (los rendimientos son equivalentes) al 95 % de confianza.

Problema 4.17

Problema. 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.

- Si atendiéramos exclusivamente a la media aritmética de los resultados, ¿qué configuración parecería la mejor?
- Para un nivel de confianza del 95 %, ¿afecta la configuración del S.O. al rendimiento del equipo?
- 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.

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
	15.7	15.63	18.68	15.5	18.21

Solución

Apartado a). Atendiendo exclusivamente a las medias aritméticas, parece que la configuración 3 es la mejor, por tener una mayor media de 18.68 páginas servidas por segundo.

Apartado b). De las tablas podemos ver que $F_{exp} = 90.695$. Suponiendo la hipótesis que el factor no influye en el rendimiento, la probabilidad de que esta hipótesis pueda ser cierta se identifica con la probabilidad que hay de que la muestra 90.695 o superior se haya extraído de una distribución $F_{4,45}$ (donde $4 = n_{niv} - 1 = 5 - 1$ y $45 = n_{niv}(n_{rep} - 1) = 5 \cdot (10 - 1)$), equivalentemente, el P-valor, que es $P(F \leq 90.695; 4, 45) = 0.000$ de acuerdo a la tabla ANOVA. Como $0.05 = \alpha > \text{P-valor} = 0.000$, descartamos la hipótesis. En resumen, podemos decir que la configuración afecta al rendimiento con un 95 % de confianza.

Apartado c). Viendo la tabla de múltiples comparaciones (*multiple comparisons*), vemos que, al 95 % de confianza, las configuraciones afectarán al mismo modo al rendimiento si su P-valor es menor que α . De este modo, podemos encontrar tres grupos de configuraciones que afectan igualmente al rendimiento: por una parte, las configuraciones 1, 2 y 4; por otra parte, la configuración 3; y por otra parte, la configuración 5.

Claramente, las mejores configuraciones (con un mayor rendimiento) son la 3 en primer lugar, seguido de la configuración 5.

Para un nivel de confianza del 99 %, las agrupaciones de configuraciones con el mismo rendimiento cambiarían, pues en este caso, al ser $\alpha = 0.01$, vemos que podríamos agrupar a las configuraciones 3 y 5 por afectar igualmente al rendimiento.