

Comparació entre dues targetes gràfiques

Probabilitat i estadística, grup 11

Desembre 2016

Resum

Objectiu:

Comparar la potència de dues targetes gràfiques de la mateixa gamma amb una generació de diferència per veure si compleixen les estadístiques que va presentar l'empresa NVIDIA durant l'anunci de la més recent.

Hipòtesis:

Partim de la base que ens creiem la informació proporcionada per l'empresa americana. Per tant plantejem la nostra hipòtesis així: "La targeta gràfica NVIDIA GTX 1080 és un 50% més potent que la targeta NVIDIA GTX 980".

Mètodes:

Hem sotmès les dues gràfiques que NVIDIA va comparar -GTX 980 i GTX 1080- a un mateix Benchmark -proporcionat pel videojoc Rise of the Tomb Raider- i utilitzant els mateixos components d'ordinador en els dos casos. En total, aquesta prova s'ha repetit 40 cops -20 cops amb la 980 i 20 amb la 1080- i s'han mesurat els fps -imatges per segon; traducció quantificable de la potència d'una gràfica- durant la seva execució. Finalment, amb l'objectiu de poder contrastar la nostra hipòtesis, hem enfrontat les dades recollides.

Resultats:

incluir sd !!

La mitjana de imatges per segon per la GTX 980 ha estat de 78.189 i per la GTX 1080 ha estat 142.104. Amb això podem rebutjar la H₀ -1080 és un 50% més potent que la 980- i observem que la 1080 és significativament més potent que un 50%.

Conclusió:

explicar de qué es este IC

Els nostres resultats ens han permès rebutjar la nostra hipòtesi nul·la, amb un IC = [-26.47561 -23.16574] i un p-valor $< 2.2e-16$ intuïm que la 1080 és més de un 50% més potent que la 980. Tot i que han estat condicionats a que només els hem fet en un sol videojoc. Llavors en un futur seria millor agafar noves mostres d'alguns altres amb requisits diferents.

Introducció

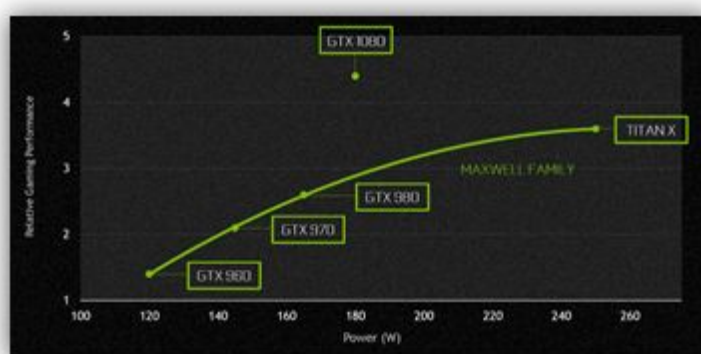
Al Maig d'aquest any, l'empresa americana NVIDIA va presentar els seus nous models de targetes gràfiques -la sèrie 10-, aquest esdeveniment va ser més important que els anuncis previs de la companyia pel fet que per primera vegada en 3 generacions el canvi de sèrie és acompanyat d'un canvi en l'arquitectura interna de les targetes -de Maxwell a Pascal-.

De les dues targetes presentades durant l'esdeveniment, nosaltres -tal com va fer NVIDIA- ens centrarem en la més poderosa, la NVIDIA GTX 1080. Aquesta gràfica consta de 9 TFLOPS de potència, 8 GB GDDR5X, arquitectura Pascal, menor consum energètic i els últims busos de dades.

La companyia americana va afirmar que la 1080 era un 50% més poderosa que el model equivalent de la generació anterior -la 980- i equivalent a un doble SLI de aquestes últimes.

Aquesta informació va anar acompanyada per dades i gràfics que reafirmaven les paraules dels executius de l'empresa. Entre aquesta

informació va destacar la demostració en directe del videojoc DOOM (2016) a 4K i configuració gràfica alta que mantenia uns sorprenents 60 fps -imatges per segon- corrent en una 1080.



Tot i que NVIDIA va presentar moltes dades que defensen les seves paraules, es sabut la facilitat amb que algunes companyes utilitzen la hipèrbole –només cal veure el que va passar amb la generació anterior, la sèrie 9-. Ens agradaria poder contrastar per nosaltres mateixos les afirmacions de la companyia americana.

Objectiu

Utilitzant una NVIDIA GTX 980 i una NVIDIA GTX 1080 com a mostra, volem veure si -tal com afirma la companyia americana- la segona presenta una potència un 50% més gran que la primera si únicament variem la gràfica i mantenim la resta de components del PC.

Material i mètodes

Per mesurar el rendiment de les dues targetes gràfiques -la GTX980 i la GTX1080- hem fet servir un mateix benchmark 20 cops per cada gràfica -en total s'ha executat 40 cops- deixant mínim 5 minuts entre cada iteració -evitem que les mostres estiguin condicionades per les anteriors-. Les dades van ser recollides els dies 15, 17 i 18 de Desembre.

Per poder aconseguir unes bones mostres hem definit una serie de variables que mantindrem constants durant la mesura de dades. S'ha utilitzat un únic ordinador amb Windows 10 com a sistema operatiu, un Intel Core i7-4770K, 16GB de RAM DDR3 1886 MHz i no s'han actualitzat SO ni drivers.

Hem fet servir el benchmark inclòs en el videojoc Rise of the Tomb Raider creat per la companyia Square Enix. Aquest test consisteix en renderitzar tres escenes del joc, aleshores hem fet servir el programa MSI Afterburner per a veure a quants fps veiem les escenes a cada determinat instant. Per a tots els càlculs necessaris també hem usat el programa R 2.12.1.

Pasos per mesurar

1. Configuració de gràfics i pantalla
2. Configuració del programa per mesurar les mostres (MSI Afterburner)
3. Execució de la funció de benchmark del joc (Rise of the Tomb Raider)
4. Visualització dels fps amb el programa Atom i recollida de dades.

No queda bien explicado cuál es el mecanismo que hace que los datos queden apareados. Debería haber 20 escenas distintas que fueran tratadas cada una 3 por las dos tarjetas. ¿Es eso lo que se ha hecho? ¿Cuál es la unidad experimental?

1. Premisses

- ❑ Normalitat per a cadascuna de les variables
- ❑ Mostres aleatòries independents
- ❑ Variàncies iguals

Apareadas?

2. Prova d'hipòtesis -bilateral-

- ❑ $H0 : \mu d = \mu(1080) - 1.5 * \mu(980) = 0$
- ❑ $H1 : \mu d \neq 0$

3. Càlcul de l'estadístic

$$\hat{t} = \frac{(\bar{D} - \mu_D)}{S_D / \sqrt{n}} = \frac{\bar{D}}{S_D / \sqrt{n}}$$

X = conjunt 1080

Y = conjunt 980

On $D = X - Y$ i S_D = variància de D

4. Estudi del p-valor ($\alpha = 0.05$)

Si el p-valor $< \alpha$ podem rebutjar la hipòtesis nul·la, amb una confiança del 95%.

5. Interval de confiança

$$IC(\mu_T - \mu_O, 1 - \alpha) = \left[\bar{x}_T - \bar{x}_O \pm t_{n_T + n_O - 2, 0.975} \cdot S_p \sqrt{\frac{1}{n_T} + \frac{1}{n_O}} \right]$$

¿?

Nuevamente, se refiere al IC de muestras independientes!

Resultats

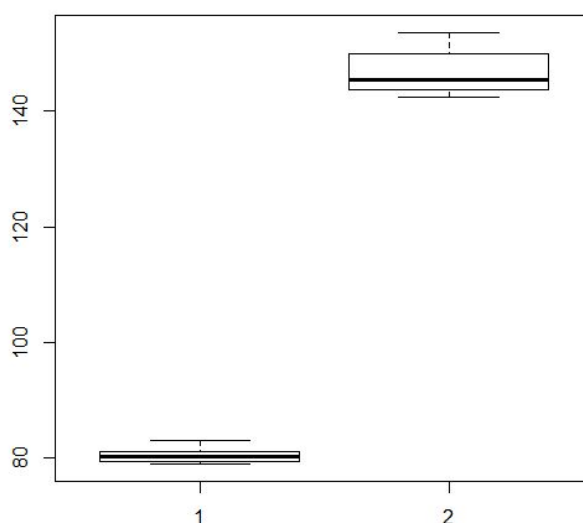
Amb les taules de dades en que tenim els fps per instant de 20 benchmarks per cada targeta gràfica, hem fet la mitjana dels instants i hem calculat els següents indicadors utilitzant les mitjanes de cada test.

GTX 980

summary(gtx980)					
Mínim	Màxim	Mitjana	Mediana	1st Qu.	3rd Qu.
79.084	93.521	81.134	80.493	79.53	81.26

GTX 1080

summary(gtx1080)					
Mínim	Màxim	Mitjana	Mediana	1st Qu.	3rd Qu.
142.406	153.544	146.561	145.383	143.7	149.7



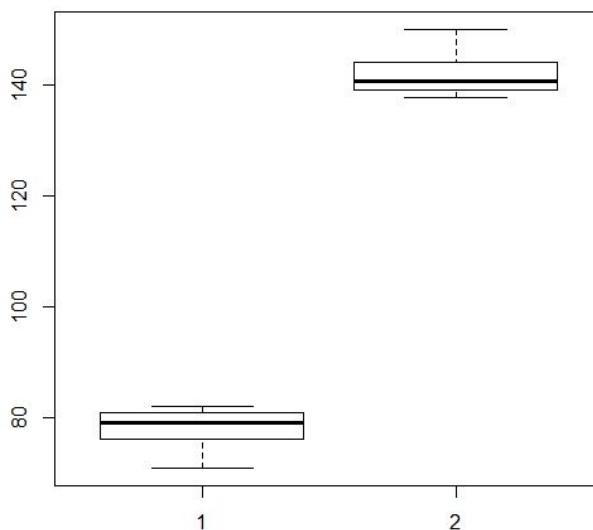
Quan vam fer els benchmarks, en un moment donat d'aquest la pantalla es posava negra i canviava l'escena, durant aquest moment no té sentit mesurar els fps i per això ens ha sortit una mostra amb molta variància -observar el boxplot superior-, per aquest motiu, hem decidit eliminar les mostres que corresponen a aquest interval i hem tornat a fer els càlculs i els gràfics -aquest cop reduint el soroll-:

GTX 980

summary(gtx980)					
Mínim	Màxim	Mitjana	Mediana	1st Qu.	3rd Qu.
70.896	82.082	78.189	79.040	76.29	80.72

GTX 1080

summary(gtx1080)					
Mínim	Màxim	Mitjana	Mediana	1st Qu.	3rd Qu.
137.694	149.974	142.104	140.762	139.0	143.9



Mejor si visualizamos además con la 1080 la variable 980 con el factor 1.50 (para ver si son muy distintas)

Validesa i normalitat de les dades

Desviacions	sd(gtx980)	sd(gtx1080)
	3.31147	3.905442

var.test(gtx980, gtx1080)

F test to compare two variances

data: gtx980 and gtx1080

F = 0.71895, num df = 19, denom df = 19, p-value = 0.4789

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

0.2845707 1.8164021

sample estimates:

ratio of variances

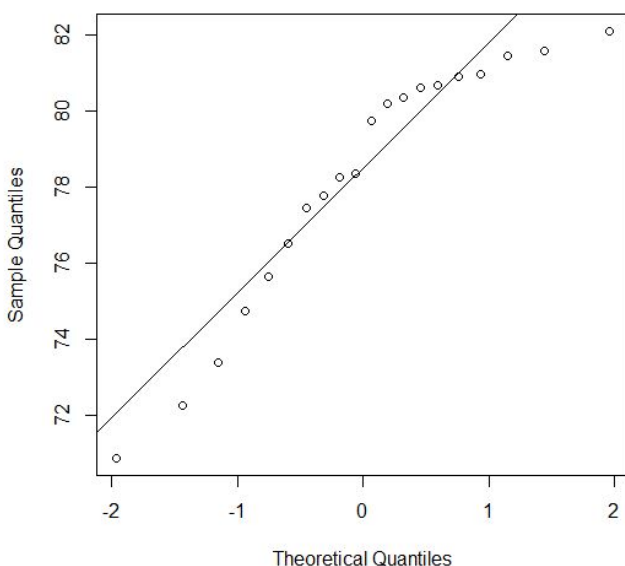
0.718954

si están apareados, este test no hace falta. Pero si hiciera falta, tendría que haber sido entre 1080 y 1.50×980 , que es lo que se compara.

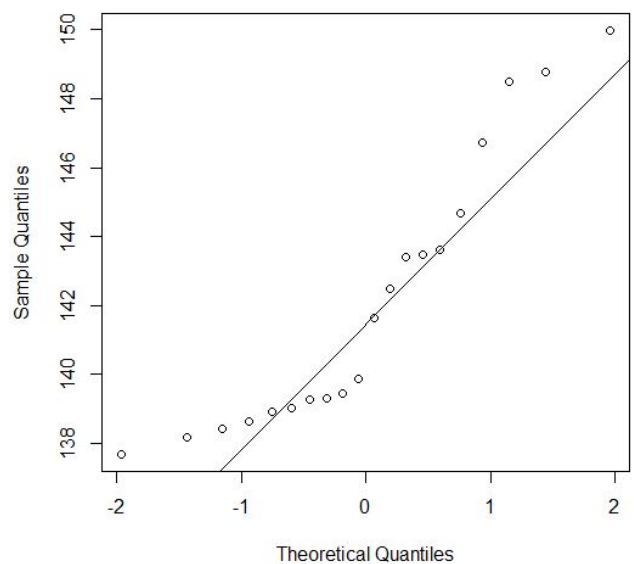
Veiem que les variàncies i les desviacions són bastant semblants, això es deu a que les dues targetes gràfiques es comporten similar davant del benchmark ja que es tractava del mateix per a tots els experiments, fet que dóna força a la validesa de les nostres dades i al nostre mètode de recollida de dades.

Per veure la normalitat de les dades fem un gràfic qqnorm amb R i veiem que mes o menys es distribueixen al voltant de la recta i per tant podem acceptar-ne la seva normalitat.

GTX 980



GTX 1080



Proves d'hipòtesis

Per poder rebutjar la hipòtesis nul·la fem un càlcul del p-valor amb la funció t.test de R:

```
t.test(gtx980*1.5,gtx1080, paired = TRUE, var.equal = TRUE)
```

Paired t-test

data: gtx980 * 1.5 and gtx1080

t = -31.391, df = 19, p-value < 2.2e-16

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-26.47561 -23.16574

sample estimates:

mean of the differences

-24.82067

sd de la diferencia = 3.536

sd 1080 = 3.905

sd 1.5*980 = 4.967

si disminuye la sd es que realmente están apareados

solo se utiliza cuando son muestras indeps.

Com que el p-valor és molt més petit que 0,05 i l'interval de confiança no inclou 0 podem refutar la hipòtesis nul·la que diu que la GTX 1080 és exactament un 50% més potent que la GTX 980 amb un 95% de confiança.

Per poder afirmar la hipòtesis alternativa fem un càlcul del p-valor amb la funció t.test de R:

```
t.test(gtx980*1.5,gtx1080, paired = TRUE, alternative = "less")
```

Paired t-test

data: gtx980 * 1.5 and gtx1080

t = -31.391, df = 19, p-value < 2.2e-16

alternative hypothesis: true difference in means is less than 0

95 percent confidence interval:

-Inf -23.45346

sample estimates:

mean of the differences

-24.82067

Com que el p-valor molt més petit que 0,05 i l'interval de confiança és tot negatiu podem afirmar la hipòtesis alternativa que diu que la GTX 1080 com a mínim un 50% més potent que la GTX 980 amb un 95% de confiança.

La interpretación sería que la 1080 muestra al menos 23.4 fps más que el 150% de lo que hace la 980. Es un poco complicado.

Discussió

Mejor si se hubiera trabajado con el ratio de fps, y tendríamos una estimación directa.

Conclusió

Els resultats obtinguts ens demostren que la targeta gràfica 1080 es un 50% més potent que la seva antecessora, la 980, que es justament el que volíem provar amb el nostre estudi. Aquesta conclusió, ha sigut obtinguda a partir de les resultats adquirits amb la $p(t)$? i el p-valor. A més, durant l'estudi dut a terme no només hem vist que la 1080 es millor que la 980, també que en alguns casos la 1080 es més d'un 50% més potent que la 980.

Limitacions

L'estudi al haver estat fet amb un sol videojoc, ha estat limitat. Això ha provocat que les nostres mostres s'han basat només en la obtenció a partir d'aquest únic videojoc això ha provocat que no siguin tan arbitraria com haguéssim volgut, hi han estat bastant condicionades.

Per tant amb aquestes limitacions en les que ens hem vist, no ha sigut possible poder saber si aquests resultats també serien els correctes en circumstàncies més generals.

Futur

Es podria fer un segon estudi, on compréssim no només amb un sol videojoc, sinó que també agaféssim alguns altre on les seves necessitats siguin diferents del que hem estat usant en aquest estudi.

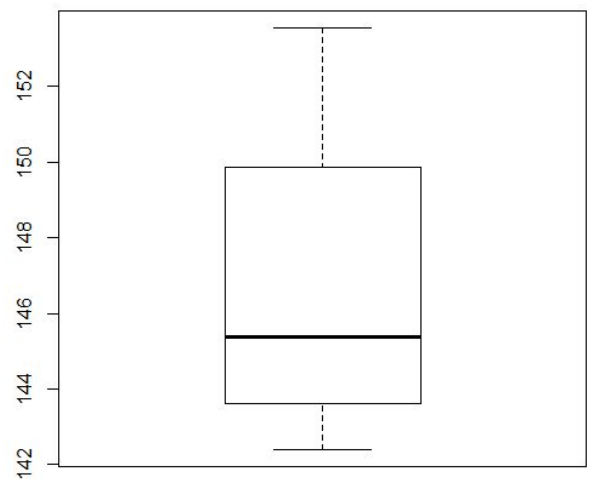
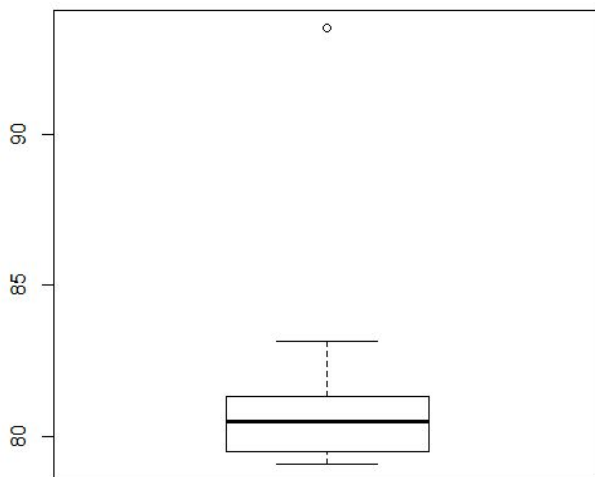
També seria un bon punt a favor d'aquest nou estudi agafar un numero més gran de mostres per tal d'aconseguir uns resultats més precisos.

y quizás utilizar otros indicadores de potencia, no solo los fps (si es que hay otros indicadores relevantes)

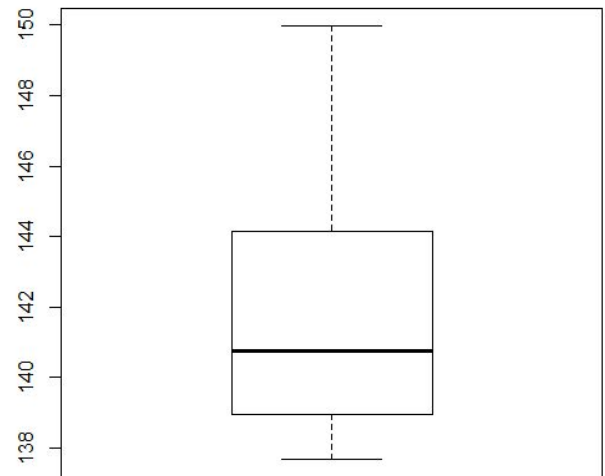
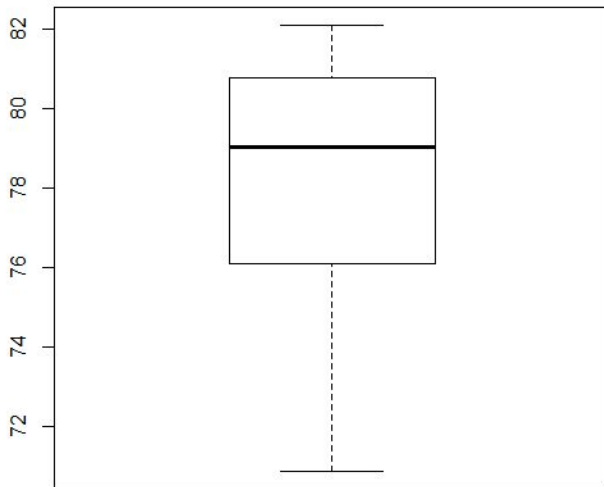
Annex



Podem veure que les mostres varien bastant però es mantenen en un interval de fins a 15 fps de diferencia.



Boxplots de les mostres abans d'eliminar els errors que vam veure quan fèiem el benchmark.



Boxplots un cop eliminats els errors.

Configuracions benchmark:

Configuració de pantalla:

- Pantalla completa exclusiva.
- No habilitar DirectX 12 - noves llibreries exclusives de windows -
- No 3D
- Resolució 1920x1080
- Freqüència d'actualització 60Hz
- Sincronització vertical desactivada
- Suavitzat FXAA -antialiasing-

Configuració gràfics (I):

- Alta
- Qualitat de textures Alt
- Filtre anisotròpic -Anisotròpic 4X-
- Qualitat de les ombres Alt

Configuració gràfics (II):

- Ombres solars suaus si
- Profunditat de camp si
- Nivell de detall Alt
- Oclusió ambiental si
- PureHair si
- Calidad de reflejo especular Normal
- Desenfoque de viñeta
- Desenfoque de movimiento
- Bloom
- Teselació
- Reflejos espaciales de pantalla
- Destellos en la lente
- efectos de pantalla
- no granulado

Dades recollides:

Veure fulls de càlcul adjunts al treball.