

USB3.0

VS

USB2.0

Cuadrimestre otoño 2016

B7 Probabilidad y Estadística

Grupo 11

Introducción

Es interesante observar cómo la gente tiende a considerar que el USB 3.0 es mejor que el 2.0 sólo por pura cuestión de términos y no por razones más racionales. En este escueto trabajo vamos a aportar una razón científica a fin apoyar la superioridad del USB 3.0.

Objetivo

Demostrar estadísticamente mediante pruebas de hipótesis cuántas veces más es mejor la versión 3.0 que la 2.0 del *Universal Serial Bus*.

Material y metodología

Vamos a comparar un USB de cada tipo, en concreto compararemos los dos siguientes:

- Pendrive 8GB - Toshiba TransMemory U202, USB 2.0
- USB 3.0 de 8 GB Store 'n' Go Verbatim

Primero deberíamos saber que no hay diferencia entre fabricantes. Así no

Estos dos dispositivos representan nuestras variables cualitativas. Aunque los dos USB pertenezcan a fabricantes distintos este hecho no debe interferir en nuestro estudio ya que nuestro objetivo es comparar sus versiones.

podemos saber si la diferencia es entre 2.0 y 3.0, o por el fabricante

La herramienta que usaremos para recoger muestras será un ordenador portátil Lenovo G50 con sistema operativo Ubuntu 15.10.

Los datos que queremos obtener para el estudio es el tiempo medio de transferencia de ficheros de determinados tamaños para cada USB. De aquí sacaremos dos variables cuantitativas: los tiempos medios de transferencia (en segundos) del USB 2.0 y del USB 3.0 para cada fichero. Cada tiempo medio se obtendrá de la media de 10 muestras independientes. Para entender mejor como sacaremos los datos, por favor, consultar la tabla “Tabla de datos” que adjuntaremos al anexo.

Finalmente, una vez tengamos todas las medias estas formarán una muestra apareada, por lo tanto crearemos una nueva variable que llamaremos D que representará la diferencia de logaritmos entre las medias del USB 2.0 y 3.0 con el objetivo de poder medir el efecto multiplicativo entre los tiempos de estos dos dispositivos.

Plan de recogida

Primero crearemos ficheros de tamaños 1, 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048 mega bytes. Luego, para cada USB recogeremos 10 muestras que consistirán en el tiempo que tarda en transferirse uno de estos ficheros, habiendo formateado el USB previamente para que cada muestra sea independiente y se haya recogido en igualdad de condiciones.

El tiempo de transferencia lo mediremos con el comando del sistema Linux *time* que mide el tiempo que tarda en acabar un proceso.

Plan estadístico

$$A = \text{medias USB 2.0} \quad B = \text{medias USB 3.0}$$

$$D = \log(A) - \log(B) = \log\left(\frac{A}{B}\right)$$

-Pruebas de hipótesis:

$$\begin{cases} H_0: uD = 0 \\ H_1: uD > 1,5 \end{cases} \quad \text{Debería ser } \mu_D > 0$$

Donde uD representa la media poblacional de D .

-Premisas:

1. Las medias de cada USB se han obtenido de muestras independientes que presentan normalidad.
2. Las medias de los USB 2.0 y 3.0 son aparejadas.
3. Prueba unilateral.
4. Damos por hecho que el USB 3.0 es más rápido que el 2.0 por lo cual descartamos la opción de $uD < 1$ y la prueba se hace unilateralmente.
5. Seguimos una distribución de *T-student*.

-Estadístico:

$$\text{Para } H_0: t' = \frac{mD}{sD/\sqrt{n}}$$

mD - log(1.5), en todo caso

$$\text{Para } H_1: t' = \frac{\log(1,5) - mD}{sD/\sqrt{n}}$$

¿? El estadístico es único

-Intervalo de confianza con $\alpha = 5\%$:

$$IC(uD, 1 - \alpha) = mD \pm t_{(0,975,n-1)} * sD * \sqrt{1/n}$$

Cálculo estadístico

Variable A:

MB	1	2	4	8	16	32	64	128	256	512	1024	2048
A	0,095	0,108	0,135	0,148	0,305	1,661	4,175	9,573	20,41	41,398	83,858	171,358

Variable B:

MB	1	2	4	8	16	32	64	128	256	512	1024	2048
B	0,099	0,113	0,116	0,114	0,16	1,135	2,718	7,162	13,86	27,791	59,685	127,744

Variable D = log(A) – log(B):

MB	1	2	4	8	16	32	64	128	256	512	1024	2048
D	-0,0412	-0,0452	0,1516	0,169	0,6451	0,3807	0,429	0,2901	0,387	0,3985	0,34	0,2937

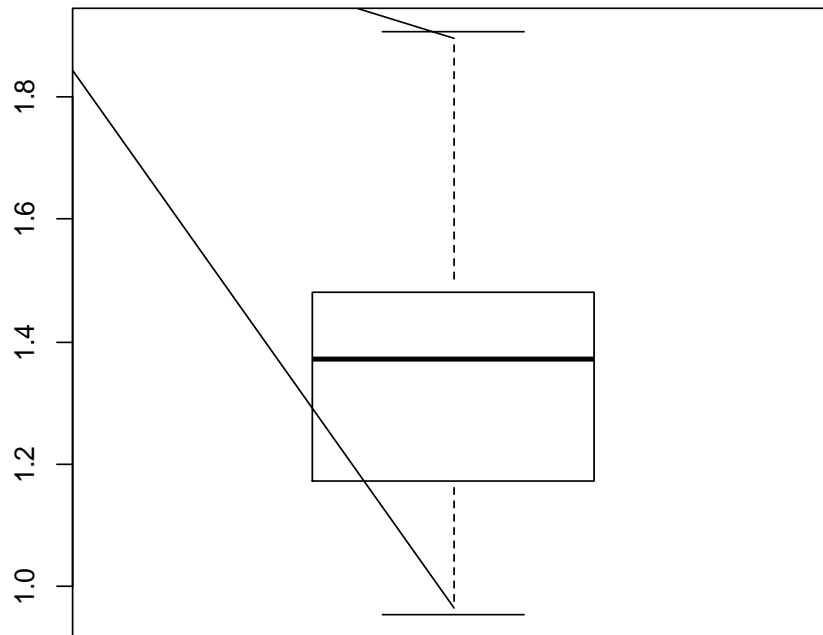
A continuación presentamos un resumen de los datos descriptivos de D:

```
> summary(D)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-0.04526  0.16460  0.31690  0.28320  0.38990  0.64510
```

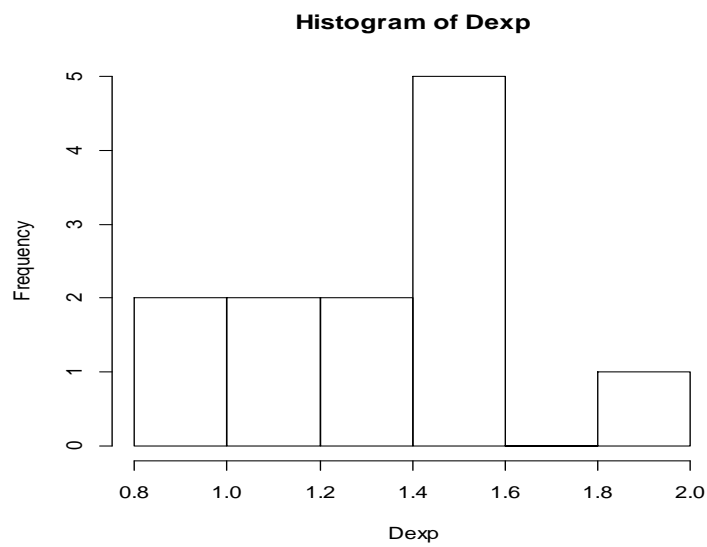
Desviación típica $sD = 0.1983613$

Con el fin de ver estos descriptivos más claramente, hemos creado una variable auxiliar que llamaremos *Dexp* que consiste en los datos logarítmicos de *D* traducidos a exponenciales. En el siguiente boxplot se puede apreciar gráficamente el “summary” de arriba:

falta qqplot de D para evaluar Normalidad



En el histograma de a continuación hemos tomado los valores de *Dexp* para que se pueda apreciar que en términos generales el USB 3.0 es un 50% más rápido que el 2.0.



Ahora procederemos a ejecutar la prueba de hipótesis con esta variable apareada D .

$$\begin{cases} H_0: uD = 0 \\ H_1: uD > 1,5 \end{cases}$$

Para ello calculamos el estadístico para H_0 usando una t-student:

$$t' = 4.946118$$

Nos da un estadístico exageradamente grande para la hipótesis nula que se aleja de nuestros límites de aceptación para el 95% de confianza. En esta prueba bilateral estos límites son $(-2.200985, 2.200985)$.

Si hacemos la prueba con el t.test de R obtenemos lo mismo:

One Sample t-test

```
data: D
t = 4.9461, df = 11, p-value = 0.0004383
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 0.1571917 0.4092572
sample estimates:
mean of x
0.2832244
```

importante

Ahora, vamos a hacer lo mismo pero para la H_1 .

Calculamos su estadístico:

$$t' = 2.134762 \quad \text{deberia ser negativo}$$

Como la prueba es unilateral, solo tenemos que contemplar el límite superior de aceptación con una confianza del 95%, que en este caso es 1.795885.

El intervalo de confianza con los datos recogido es:

$$IC = (0.1571917, 0.4092572)$$

Que traduciéndolo a exponencial queda $(1.17022, 1.505699)$.

como -2.13 no queda por encima de 1.796, no se puede aceptar que la media esté por encima de 1.5

OK

el ratio medio puede estar entre 1.17 y 1.51 (muy pobre ventaja del USB 3.0)

Quizá el fabricante también cuenta

Discusión

-Conclusiones:

Concluimos que es inverosímil que el efecto multiplicativo que se observa al comparar un USB 2.0 con uno 3.0 sea nulo porque el estadístico de la prueba de hipótesis se aleja mucho de nuestros límites de aceptación. Este hecho y nuestras premisas nos indican claramente que sí que existe un factor multiplicativo mayor que cero entre el USB 3.0 y el 2.0.

Asimismo, tampoco podemos aceptar nuestra hipótesis alternativa que quiere demostrar que el USB 3.0 es un 50% más rápido que el 2.0. El estadístico de H_1 , aunque entra en la región de aceptación de una prueba de hipótesis bilateral, no lo hace para una unilateral.

Finalmente sí que podemos afirmar que el factor multiplicativo entre estas dos generaciones de USB está entre el 17% y el 50% así como lo muestra nuestro intervalo de confianza (que es bilateral y no hay que relacionarlo con la hipótesis alternativa).

-Consideraciones de nuestro trabajo:

Una vez hecho este trabajo creemos que la metodología usada para este estudio ha sido errónea. Creemos **no deberíamos haber sacado muestras de ficheros con tamaños tan pequeños** ya que complica medir su tiempo de transferencia. Por ejemplo, deberíamos de haber establecido un intervalo de (100, 1524) MB, que son tamaños de ficheros con los que se trabaja habitualmente y hacen más fiables la medición de tiempos, al ser lo suficientemente grandes para tener precisión y lo suficientemente pequeños para que su tiempo de transferencia no se dispare.

A parte de eso consideramos que deberíamos haber establecido críticamente el número de muestras para obtener las medias y haber comprobado que efectivamente estas son independientes entre sí y se asemejan a una Normal.

Puede ser;
estos
ratios
deberían
tener
menos
influencia

Anexo

U.S.B	Tamaño del fichero en Mbytes	Tiempo de transferencia (s)										Media
		1	2	3	4	5	6	7	8	9	10	
2.0	1	0,19	0,02	0,1	0,05	0,1	0,03	0,1	0,14	0,1	0,12	0,095
3.0	1	0,1	0,01	0,01	0,05	0,1	0,01	0,1	0,5	0,1	0,01	0,099
2.0	2	0,16	0,03	0,1	0,12	0,1	0,1	0,12	0,02	0,18	0,15	0,108
3.0	2	0,13	0,1	0,1	0,09	0,1	0,1	0,15	0,1	0,16	0,1	0,113
2.0	4	0,15	0,2	0,1	0,05	0,16	0,08	0,15	0,1	0,21	0,15	0,135
3.0	4	0,1	0,11	0,11	0,1	0,1	0,1	0,2	0,07	0,14	0,13	0,116
2.0	8	0,24	0,12	0,1	0,12	0,21	0,12	0,14	0,15	0,12	0,16	0,148
3.0	8	0,11	0,11	0,1	0,09	0,14	0,2	0,1	0,08	0,11	0,1	0,114
2.0	16	0,46	0,2	0,25	0,11	0,27	0,17	0,53	0,55	0,37	0,14	0,305
3.0	16	0,12	0,13	0,3	0,17	0,13	0,12	0,12	0,26	0,13	0,12	0,16
2.0	32	2,04	1,37	1,94	1,39	1,84	1,43	2,01	1,4	1,38	1,81	1,661
3.0	32	0,95	0,97	0,98	0,98	1,22	0,99	1,3	1,67	0,99	1,3	1,135
2.0	64	4,87	4,21	4,24	4,18	3,96	3,95	3,91	4,26	4,18	3,99	4,175
3.0	64	2,7	2,63	3,02	3,12	2,78	2,55	2,45	2,73	2,3	2,9	2,718
2.0	128	10,3	9,54	9,57	9,4	9,4	9,37	9,39	9,54	9,63	9,59	9,573
3.0	128	6,84	6,12	6,88	7,6	7,64	7,44	6,98	6,9	7,86	7,36	7,162
2.0	256	21,05	20,27	20,31	20,29	20,32	20,46	20,33	20,51	20,27	20,29	20,41
3.0	256	14,15	14,22	14,27	14,26	13,95	13,04	13,69	13,74	13,55	13,73	13,86
2.0	512	42,49	41,49	41,5	41,3	41,17	41,25	41,13	41,21	41,19	41,25	41,398
3.0	512	27,43	27,56	27,71	27,79	28,25	27,66	27,95	27,86	27,86	27,84	27,791
2.0	1024	85,35	84,06	83,73	83,66	83,74	83,48	83,88	83,86	83,49	83,33	83,858
3.0	1024	57,09	59,02	61,63	61,95	60,6	60,4	59,13	59,56	58,78	58,69	59,685
2.0	2048	170,58	178,46	176,62	172,56	170,28	168,45	169,48	169	169,05	169,1	171,358
3.0	2048	137,75	140,58	132,06	124,4	125,26	124,5	120,77	124,15	120,17	127,8	127,744

1. Tabla de datos

Script usado para hacer los cálculos en R:

```

usb2 <- c(0.095,0.108,0.135,0.148,0.305,1.661,4.175,9.573,20.41,41.398,83.858,171.358)
usb3 <- c(0.099,0.113,0.116,0.125,0.16,1.135,2.718,7.162,13.86,27.791,59.685,127.744)

#VARIABLE D
log2 = log(usb2); log2
log3 = log(usb3); log3
D = log2-log3; D
#cuantas veces mas veloz es usb3 de usb2
#como usb3 es un tiempo menor esta en el denominador

#MITJANA Y DESVIACIÓ
mD = mean(D);mD
sdD = sd(D);sdD
n = length(D);n

#PROVA D'HIPOTESI
#la hipotesi es que D > log 1.5, o sigui que el temps de usb2 es 1.5 vegades major que el
#temps de usb3
# H0: uD = 0
# H1: uD > log 1.5

#ESTADÍSTIC t
#para H0
t1 = (mD)/(sdD/sqrt(n));t1
#para H1
t2 = (mD-log(1.5))/(sdD/sqrt(n));t2

#region de aceptacion
#bilateral
qt(0.975,n-1)
#unilateral
qt(0.95,n-1)

#IC 95%
#por debajo
mD-qt(0.975,11)*sdD*sqrt(1/n)
#por arriba
mD+qt(0.975,11)*sdD*sqrt(1/n)

t.test(D);

```