

MAnáDINe-TIplE

Examen

October 25, 2016

Aviso: Hay que justificar todas las respuestas, detallando el razonamiento o haciendo referencia a los apuntes o a cualquier fuente de internet.

Elabora un informe con tus respuestas del examen y envíalo, junto a los anexos que consideres oportunos, a `carleos@uniovi.es`

1. (1'2 puntos) ¿Qué lenguajes serían recomendables para...
 - (a) ...programar un algoritmo que tiene que ir muy rápido?
 - (b) ...programar a alto nivel un servidor de macrodatos (hadoop, spark, flink...)?
 - (c) ...obtener la derivada simbólica de una función matemática?
 - (d) ...acceder a un banco de datos genómico?
 - (e) ...programar a alto nivel procedimientos matemáticos, pero que vayan muy rápido?
 - (f) ...usar los últimos algoritmos estadísticos?
 - (g) ...operar con matrices cómodamente?
 - (h) ...hallar estadísticos descriptivos de un banco de datos que quepa en la memoria RAM?
 - (i) ...hallar estadísticos descriptivos de un banco de datos que no quepa en la memoria RAM?
 - (j) ...buscar ficheros en un disco duro que cumplan ciertas condiciones?
 - (k) ...filtrar (quedarse con ciertas filas) de un fichero muy grande?
 - (l) ...quedarse con los caracteres del décimo al vigésimo de cada renglón de un fichero muy grande?
2. (1'5 puntos) Crea un fichero `listado.txt` a partir de la salida del comando de Bash siguiente:

```
find /usr/share/doc -ls
```

La columna segunda contiene la cantidad de bloques ocupados en KiB (quibioctetos).

La columna sétima contiene el tamaño de los ficheros en octetos.

- (a) Mediante `cut`, obtén un fichero `octetos.txt` que contenga sólo las dos columnas mencionadas arriba.
Si no están bien alineadas las columnas en el fichero original, prueba a entubarlo a través del comando `column -t`

- (b) Haz lo mismo que en el anterior, pero mediante `awk`.
 - (c) Mediante `awk`, obtén la media y la desviación típica de la segunda columna.
 - (d) Obtén el máximo de cada columna.
 - (e) Mediante `awk`, obtén un fichero `octetos+.txt` que contenga, además, una columna con la diferencia entre octetos ocupados y octetos reales de cada fichero.
3. (0'6) En el fichero `países.ods` aparece la columna `DENSIDAD`, medida en habitantes por km^2 .
 - (a) Calcula la densidad global.
 - (b) Calcula una nueva columna que sea la densidad en habitantes por milla cuadrada de cada país, sabiendo que cada milla es igual (más o menos) a 1'6 km.
 4. (0'5) ¿Cuántos ficheros hay en `/usr/share/doc` cuyo tamaño supere los 10.000 octetos?
 5. (0'3) Considera los datos de R llamados `mtcars`. Di cómo puedes calcular coeficiente de determinación R^2 entre el peso `wt` y la «aceleración» `qsec` y usarlo en un programa.
 6. (1'5) Considera los datos de R llamados `AirPassengers`. Es un objeto de clase `ts` (serie temporal) que puede parecer una matriz pero, en este caso, es un vector con algún atributo adicional. Los elementos del vector son los totales mensuales de viajeros (en miles). ¿Cómo puedes obtener un vector (más corto) de totales anuales de viajeros?
 7. (0'3) ¿Cómo encuentras en el sistema todos los ficheros cuyo nombre contiene la cadena `BSD`?
 8. (0'5) ¿Cuáles son las cinco palabras más frecuentes del fichero `GFDL`?
 9. (1 punto) Visita la página del I.N.E.
 - Pincha a la izquierda en «INEbase, la información estadística».
 - Pincha en «censos de población y viviendas 2011».
 - Pincha a la izquierda en «resultados detallados».
 - Pincha abajo en «ficheros de microdatos».
 - Personas y hogares: Pincha en «fichero nacional» y guarda el archivo y descomprímelo.
 - Personas y hogares: Pincha en «diseño de registro y valores válidos de las variables».
 - Haz un diagrama de barras o histograma de la distribución de edades.
 10. (1 punto) Define una función en R que calcule la desviación mediana (la mediana de los valores absolutos de las diferencias respecto a la mediana) y aplícala a todas las variables de `pisa2012`, sólo para los datos de España.
 11. (0'6 puntos) ¿Cuáles son los tres años con más datos y los tres años con menos datos en el fichero `vuelos.csv`?
 12. (1 punto) Haz un gráfico de barras de los años de nacimiento de las `vacas.csv` que no tienen dato de genotipo.