

Fundamentos de Programación - ST0242

Práctica Número 02

Grupos: 033, 035, 036

Juan Francisco Cardona Mc'Cormick
Mauricio Álvarez Henao
Helmuth Trefftz Gómez

Noviembre, 2019

1. Introducción

La estadística descriptiva es un área de la estadística en la cual se obtienen ciertas medidas respecto a un conjunto de datos, para entenderlos mejor. El promedio y la desviación estándar son ejemplos de ese tipo de medidas. La estadística inferencial va un paso más allá, buscando llegar a conclusiones sobre la totalidad de una población partiendo de información que se extrae de una muestra.

2. Objetivo General

Con esta práctica se busca que los estudiantes del curso de Fundamentos de Programación pongan en práctica sus habilidades de programación, implementando algunas medidas de estadística descriptiva sobre un conjunto de datos determinado.

3. Objetivos Específicos

Con la práctica se busca que los estudiantes demuestren las siguientes habilidades:

- Leer datos de un archivo tipo texto y guardarlos en memoria en una estructura adecuada para su procesamiento
- Implementar el código en Java para calcular algunas medidas de estadística descriptiva sobre un conjunto de datos.
- Documentar el código utilizando JavaDoc

4. Descripción del Proyecto

4.1. Datos de entrada

El programa debe leer los datos de un archivo tipo texto. En cada línea se leen los datos de una persona. Para cada persona se proveen los siguientes datos:

- Número de documento
- Nombre
- Apellido
- Género (*m* es mujer y *h* es hombre)
- Edad
- Peso (en Kilogramos)
- Estatura (en centímetros)
- Ritmo cardíaco en reposo (latidos por minuto)
- Presión sistólica (en milímetros de mercurio)
- Presión diastólica (en milímetros de mercurio)
- Ha sufrido por lo menos un infarto (*n* es no y *s* es si)
- Nivel de azucar glucosa en sangre (en miligramos por decilitro)

Un ejemplo de 4 registros en el archivo sería el siguiente:

```
72514202,Juan Felipe,Barco,h,60,65,166,68,126,69,n,68
43151370,Sofía,Cardeño,m,16,35,143,63,122,75,n,78
70740339,John Jairo,Cardeño,h,47,90,173,64,127,67,s,90
43458250,Luisa Fernanda,Díaz,m,32,59,156,64,121,85,n,89
```

Los datos se deben guardar en una estructura de datos de la siguiente manera:

Se tiene una matriz donde cada columna es un rango de edad, la primera fila está relacionada con las mujeres y la segunda fila está relacionada con los hombres. Cada entrada de la matriz es un ArrayList en el cual se almacenan los datos de las personas que se corresponden con el género y el rango de edad de esa entrada en particular.

Los rangos de edad que se van a considerar son:

- 10 a 20 años
- 21 a 30 años
- 31 a 40 años

- 41 a 50 años
- 51 a 60 años
- 61 a 70 años
- 71 a 80 años
- 81 a 90 años
- mayor de 90 años

4.2. Línea de comandos

Para manejar los diferentes reportes (ver ??) que aquí se describen, es importante utilizar la línea de comandos que está disponible con los programas escritos en Java. Los programas comienzan con bibliotecas estáticas¹ a través del método estático `main` si se observa la firma² esta tiene el siguiente formato:

```
public static void main(String[] args) {
    ...
}
```

donde el `args` es un arreglo de cadenas de caracteres (*Strings*). El arreglo `args` está relacionado con una forma de invocar cada programa en java desde la línea de comando (*terminal*). En este proyecto vamos a utilizar la línea de comandos para pasar la lista de argumentos, con el siguiente es el formato:

```
$ java <ClaseConmainMetodo> [-r n] ... ficheroEntrada
```

Lo anterior indica que se van a recibir varias opciones `-r` que indica el número `n` de reporte este es `n` valor 1 ... 12 que indica el número de reporte. Supongamos que tiene una clase llamada `Ppal` que contiene un método `main`. Queremos obtener los reportes 1, 5, 9, 10 de un fichero de datos llamado `datos01.in`, invocamos el programa de la siguiente manera:

```
$ java Ppal -r 1 -r 5 -r 9 -r 10 datos01.in
```

En el siguiente ejemplo queremos obtener *todos* los reportes del mismo fichero de datos `datos01.in`:

```
$ java Ppal datos01.in
```

Si los reportes se imprimen en el mismo orden en el que se recibe, si hay duplicados se ignoran:

```
$ java Ppal -r 12 -r 1 -r 3 -r 12 -r 1 -r 9 datos01.in
```

Se imprimirán los reportes 12, 1, 3, 9 en este orden, observe que los reportes repetidos 12 y 1 se imprimirán una sola vez.

¹Clase que tiene únicamente métodos estáticos `static`.

²El nombre de la método: en conjunto con el valor de retorno y los parámetros

4.3. Reportes

El programa debe producir los siguientes reportes:

1. Índice de masa corporal promedio de todos los hombres vs. el promedio de todas de las mujeres ³.
2. Matriz con el índice de masa corporal promedio por género y rango de edad.
3. Posibles parejas formadas por una mujer y un hombre que estén en el mismo rango de edad, ambos con una presión diastólica menor que 80. Se produce una lista con los nombres de los integrantes de las parejas para el rango de 21 a 30 años y otra para el rango de 61 a 70 años. Si no hay hombres o mujeres que cumplan la condición, en un rango determinado de edad, no es posible generar parejas, y en este caso se indica con la frase **No es posible formar parejas**.
4. Presión sistólica promedio de los hombres que han sufrido un infarto vs. los que no han sufrido un infarto.
5. Mismo informe para las mujeres.
6. Promedio, mediana y desviación estándar ⁴ de la presión sistólica para todas las mujeres de la muestra.
7. Promedio, mediana y desviación estándar de la presión sistólica para todos los hombres de la muestra.
8. Matriz por género y rango de edad con ocurrencias de infarto.
9. Matriz por género y rango de edad con el valor de la mediana de la medida de glucosa en sangre por género y rango de edad.
10. Coeficiente de correlación de Pearson entre el ritmo cardíaco en reposo y la presión sistólica para los hombres de la muestra ⁵.
11. Mismo reporte para las mujeres de la muestra.
12. Matriz por género y rango de edad con los coeficientes de correlación entre el ritmo cardíaco y la presión diastólica para todas las personas de la muestra.

³ <https://cuidateplus.marca.com/alimentacion/diccionario/indice-masa-corporal-imc.html>

⁴ <https://es.khanacademy.org/math/probability/data-distributions-a1/summarizing-spread-distributions/a/calculating-standard-deviation-step-by-step>

⁵ https://es.wikipedia.org/wiki/Coeficiente_de_correlacion_de_Pearson

5. Sustentación

El estudiante debe dar cuenta de todo el código que presenta. Adicionalmente, el profesor podrá solicitar que el estudiante genere un nuevo informe.

6. Detalles de la entrega

1. La práctica es estrictamente individual.
2. La entrega del proyecto se hará utilizando el control de versiones que viene con BlueJ, este proyecto se llamará **Proyecto02**.
3. El código se debe subir al repositorio a más tardar el día 20 de noviembre.
4. La sustentación se realizará en los días siguientes al 20 de noviembre.
5. Los profesores del curso proveerán el conjunto de datos para probar el programa.