**Informe Datos**

*Clara Cerrato*

*Cristina Martín Bris*

*Cristina Galván Prieto*

*Gloria Labory Pesquer*

**“Data Analytics”**

*Marzo 2018*



Índice

1. Ficheros recibidos
2. Estudio de las variables
3. Gráficas por variable
4. Ficheros recibidos

Se nos han facilitado dos ficheros:

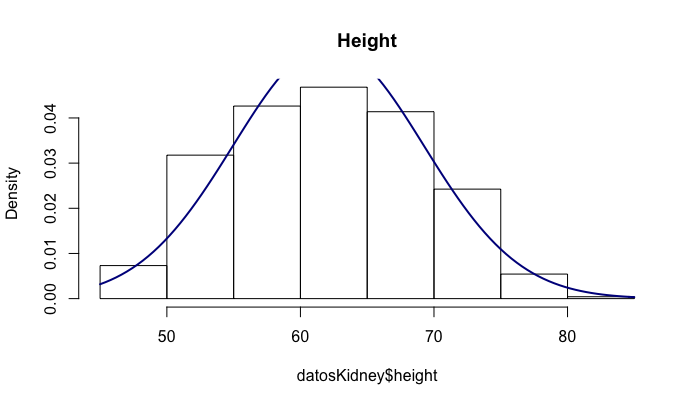
* kidney\_fail\_dataset.csv → Fichero formato CSV con todos los datos de los pacientes.  
  El fichero tiene 957 entradas y en cada entrada 20 atributos.
* codebook.docx → Fichero formato texto con descripción de los campos del fichero kidney\_fail\_dataset.csv.

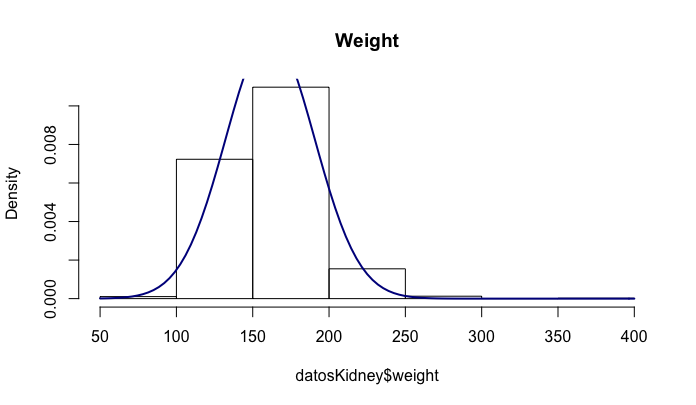
1. Estudio de las variables

Tenemos 20 variables analizadas en la siguiente tabla:

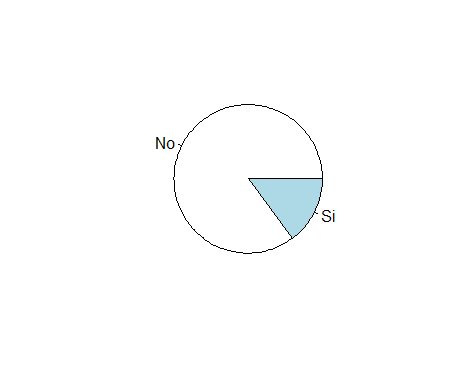
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Nombre de la variable** | **Descripción** | **Tipo de variable** | **Valor mínimo** | **Valor máximo** | **Nº de nulos** |
| patient id | Id del paciente. Será un único número entero para cada paciente | Auto numérico discreto | 1 | 957 | 0 |
| height | Altura del paciente medida en pulgadas | Numérico continuo | 45,3 | 81,1 | 0 |
| weight | Peso del paciente medido en pulgadas | Numérico continuo | 79,7 | 396,3 | 0 |
| kidney absortion test | Campo que nos indica si existe absorción en el riñón | Booleano  - Valor 1 si existe absorción  - Valor 0 si no existe | 0 | 1 | 0 |
| urea | Nivel de urea | Numérico continuo | 16,5 | 227,7 | 121 |
| monocytes | Nivel de monocitos | Numérico continuo | 3.150 | 19.005 | 125 |
| granulocytes | Nivel de granulocitos | Numérico continuo | 23.84 | 85.05 | 125 |
| kidney enzyme test | Campo que nos indica si existe una determinada enzima en el riñón del paciente | Booleano  - Valor 1 si existe enzima  - Valor 0 si no existe | 0 | 1 | 0 |
| eosinophils | Numero de eosinófilos en sangre | Numérico continuo | 0 | 19.055 | 125 |
| basophils | Número de basófilos en sangre | Numérico continuo | 0 | 2.08 | 125 |
| Glucose | Nivel de glucosa en sangre | Numérico continuo | 68.90 | 284.08 | 121 |
| kidney suffering test | Campo que nos indica si el paciente sufre de riñón | Booleano  - Valor 1 si padece de riñón  - Valor 0 si no padece | 0 | 1 | 0 |
| platelets | Nivel de plaquetas del paciente | Numérico continuo | 61.36 | 600.08 | 124 |
| trgld | Varible desconocida | Numérico continuo | 206.7 | 852.2 | 121 |
| tflr | Varible desconocida | Numérico continuo | 5.35 | 6135.38 | 285 |
| kidney genetic test (gene a3hc) | Campo que indica si existe mutación del gen a3hc | Booleano  - Valor 1 si existe mutación  - Valor 0 si no exite | 0 | 1 | 0 |
| kidney genetic test (gene 6a3cp) | Campo que indica si existe mutación del gen 6a3cp | Booleano  - Valor 1 si existe mutación  - Valor 0 si no exite | 0 | 1 | 0 |
| mean platelet volume | Volumen medio de plaquetas en sangre | Numérico continuo | 6.222 | 15.708 | 124 |
| leukocytes | Número de leucocitos | Numérico continuo | 3.432 | 18.824 | 124 |

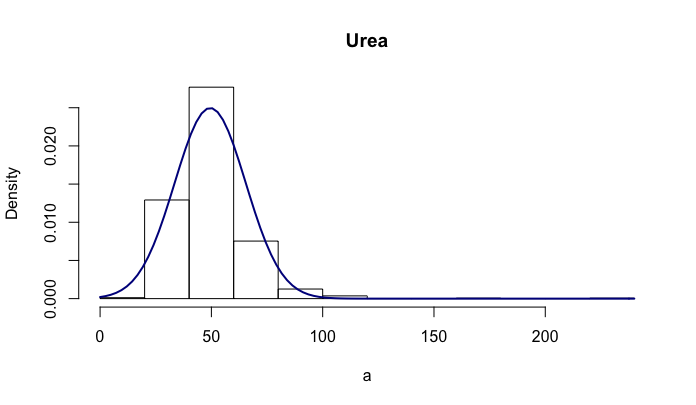
1. Gráficas

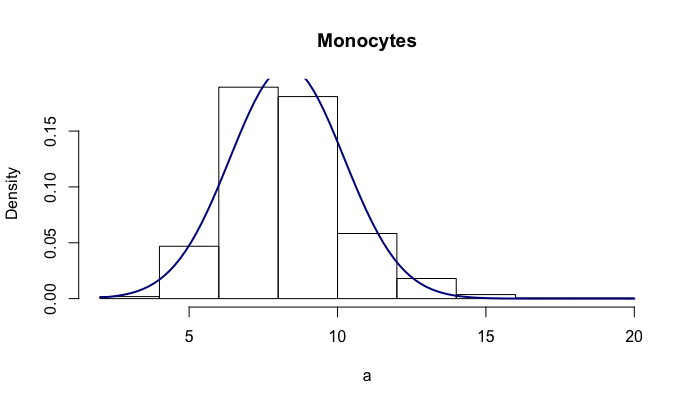


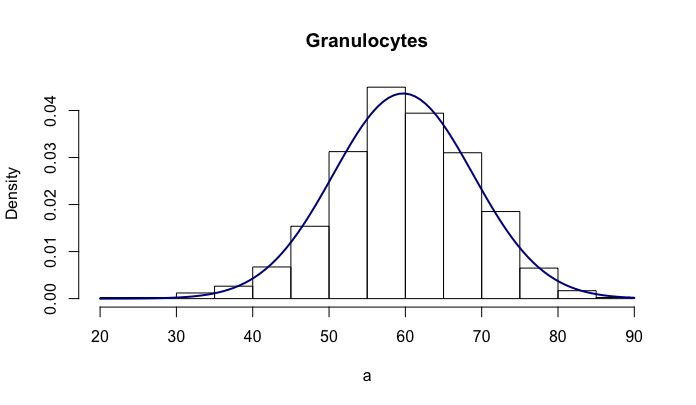


**kidney absortion test**

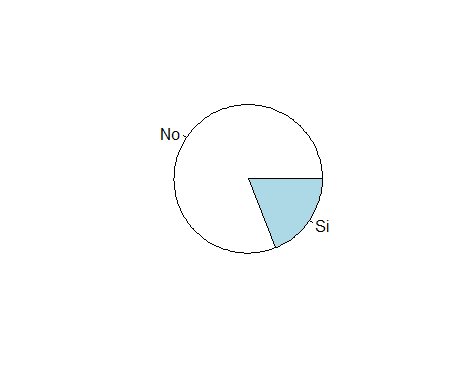


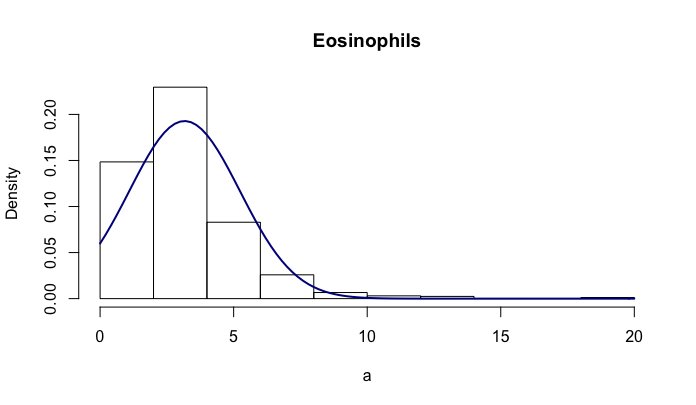


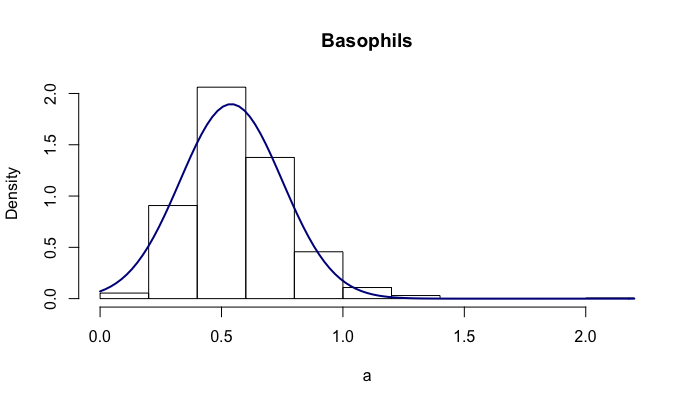


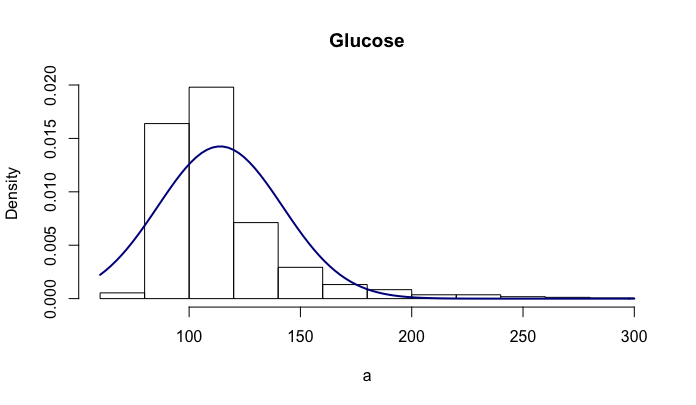


**kidney enzyme test**

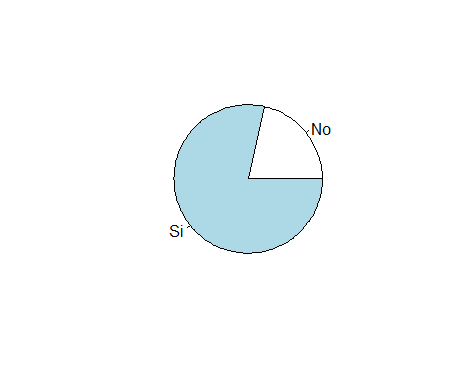


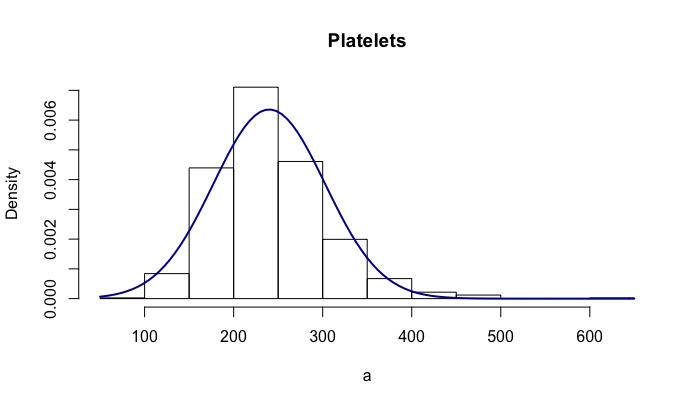


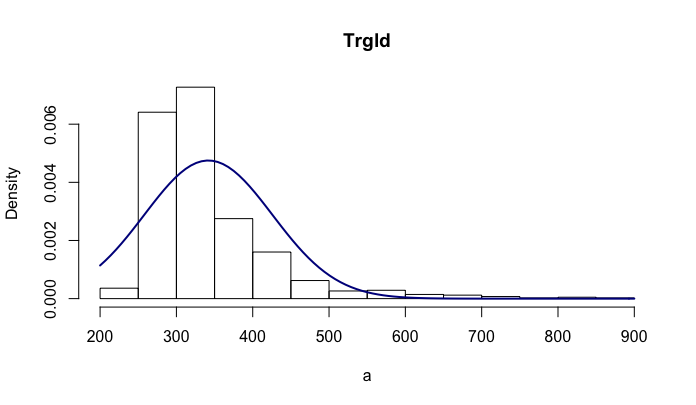


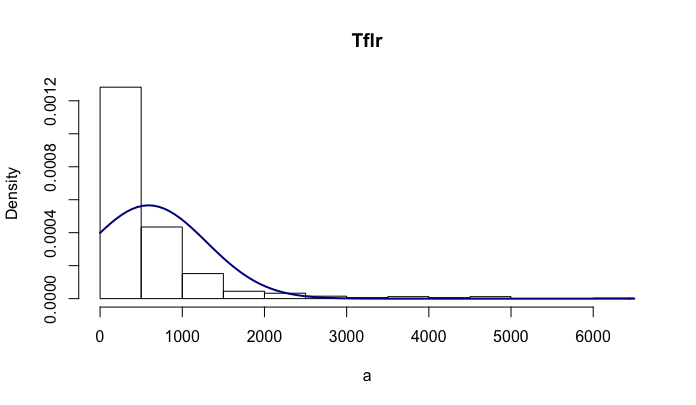


**kidney suffering test**

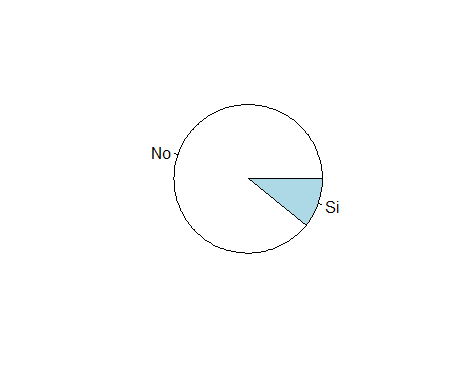








**kidney genetic test (gene a3hc)**



**kidney genetic test (gene 6a3cp)**

