

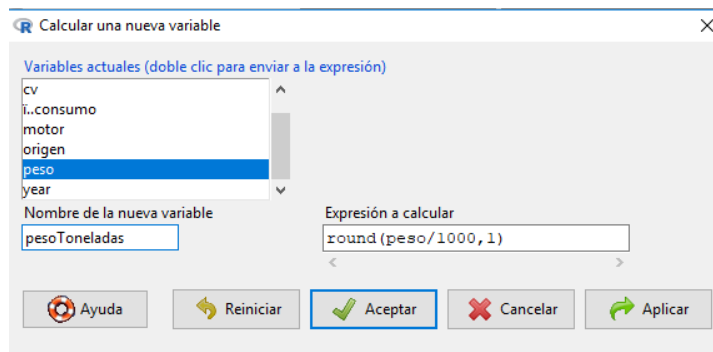
# Práctica 2 R-Commander

NOMBRE: Jaime Osés Azcona

## 1. Convertir la variable peso a toneladas con un decimal.

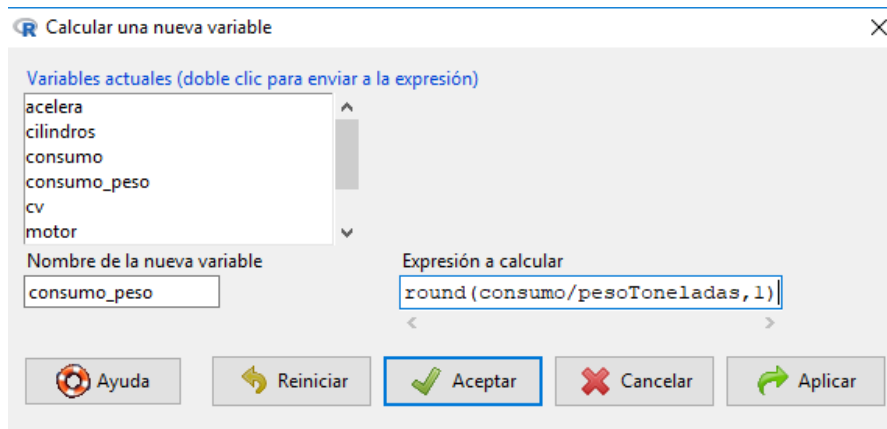
Para calcular esta nueva variable, operaremos con las diferentes columnas que ya tenemos. Para ello tendremos que ir a “**Datos-Modificar variables...-Calcular una nueva variable...**” y tendremos que definir las operaciones que queremos realizar. En este caso, como queremos que nos salga un número con un solo decimal, usaremos la función round (operación, n.º decimales).

Tendremos que dar al factor un nombre diferente a las columnas anteriores para que no nos pise ninguno de ellos, ya que estamos realizando un cálculo a partir de dichas columnas, las cuales necesitamos.



## 2. Crear una variable de ratio de consumo por peso: consumo (l/100km) por tonelada.

Para calcular esta nueva variable, operaremos con las diferentes columnas que ya tenemos. Para ello tendremos que ir a “**Datos-Modificar variables...-Calcular una nueva variable...**” y tendremos que definir las operaciones que queremos realizar. En este caso, como tenemos que hallar el consumo por cada tonelada, dividiremos el consumo entre el peso que hay en toneladas, dato que hemos hallado anteriormente, y así hallaremos el peso por 1 tonelada, que es lo que se pide.



### 3. Crear una nueva variable factor que recoja la información de la variable consumo en tres tramos, donde cada tramo tenga el mismo numero de automóviles.

Para dividir una variable en partes iguales tenemos que ir a “**Datos-Modificar variables...-Segmentar variable numérica...**” y deberemos seleccionar las opciones de segmentación que queramos. En este caso dividiremos en 3 partes con el mismo número de automóviles, lo que haremos al seleccionar la opción segmentos de igual cantidad.

En nuestro caso los segmentos no tendrán el mismo número de automóviles ya que hay una gran cantidad de ellos con los mismos datos, lo que a la hora de agrupar hace que todos ellos vayan al mismo grupo, dejando los otros grupos con una menor cantidad.

Segmentar una variable numérica

Variable a segmentar (elegir una):  
acelera  
cilindros  
**consumo**  
consumo\_peso  
cv  
motor

Nombre de la nueva variable:  
f\_consumo3partes

Número de clases: 3

Nombres de niveles:  
☐ Especificar nombres  
☒ **Números**  
☐ Rangos

Método de segmentación:  
☐ Segmentos equidistantes  
☒ **Segmentos de igual cantidad**  
☐ Segmentos naturales (mediante agrupación por K-medias)

Ayuda Reiniciar Aceptar Cancelar

#### 4. Elegir dos variables cualitativas y analizarlas. Introduce algún cambio en los gráficos.

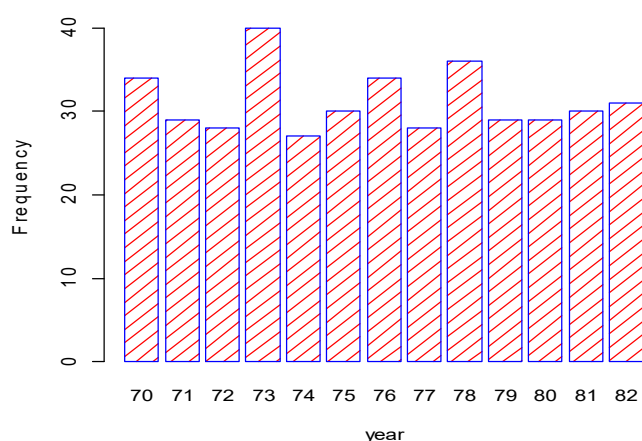
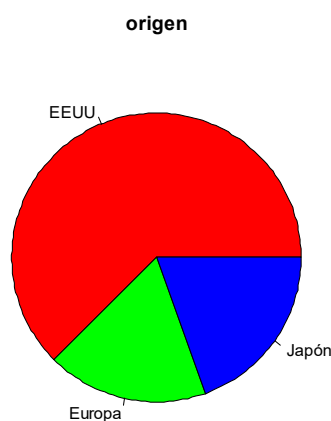
---

Como podemos observar en el gráfico de sectores, la variable origen tiene 3 grupos diferenciados: los coches que provienen de EEUU, los de Europa y los de Japón. La mayor parte de los coches (aproximadamente un 62%) provienen de EEUU y los restantes, los dos en cantidades parecidas (aproximadamente un 18% cada uno), provienen de Europa y Japón.

Como podemos observar en la grafica de barras, la variable year tiene 13 grupos diferenciados que se corresponden con los años de modelo de los coches. Casi todos los años de modelo tienen un número de coches muy parecidos que varía en un intervalo aproximadamente entre 25 y 30. Solo 4 años de modelo destacan entre los demás superando esa frecuencia de 30. Más en concreto, el año de modelo que mas coches tiene es el 73, con una cantidad de 40 coches.

Para realizar cambios en los gráficos, deberemos modificar las fórmulas desde la ventana R-Script.

- El cambio que hemos realizado en la gráfica de sectores es ponerle una paleta de colores ya existente. Para ello deberemos poner en la fórmula de R-Script **col=rainbow(3)**, indicando en el paréntesis el número de sectores que hay.
- El cambio que hemos realizado en la gráfica de barras es ponerle a las barras un borde del color que queramos y rellenar dichas barras en vez de con colores, con una especie de rayado con la densidad que queramos. Para ello deberemos poner en la fórmula de R-Script **border="blue"**, **density=10** y para indicar el color del rayado **col="red"**.

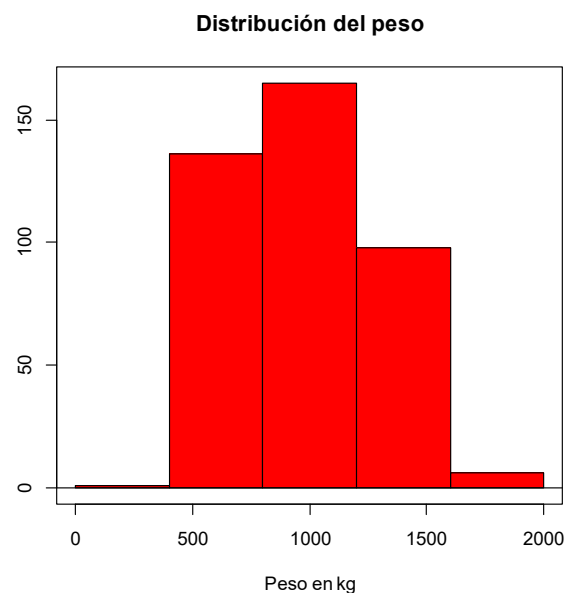
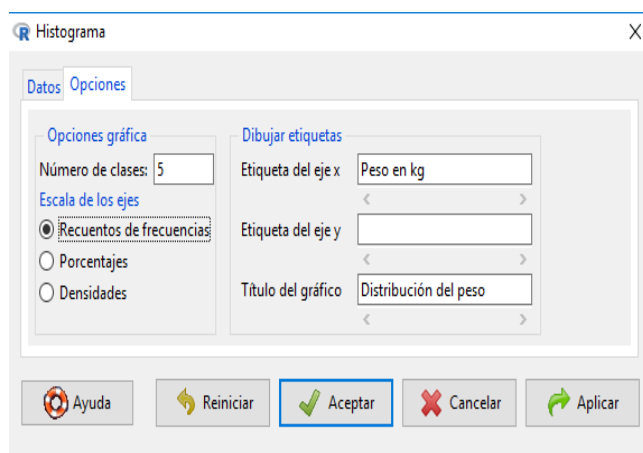


**5. Construye el histograma de la variable peso en kg en color rojo, con 5 intervalos, un título principal (Distribución del peso), otro en el eje x (Peso en kg) y sin título en el eje y.**

Para construir un histograma tendremos que ir a **Gráficas-Histograma...** y deberemos seleccionar la variable de la que se quiere realizar. Para poner las características que se especifican, deberemos ir al apartado opciones y hacer los cambios que queramos.

Hay algunos cambios que no se pueden hacer en el apartado opciones y deberemos realizarlos directamente en la fórmula de R-Script. Uno de ellos es el color, que deberemos modificarlo poniendo **col="(color que queramos)"**, en este caso **col="red"**.

Otra de ellas es los intervalos en que lo queramos dividir. En este caso desde el menú opciones lo que interpreta es una recomendación, pero no hace el número de intervalos que deseamos, lo que deberemos solucionar poniendo en la fórmula **breaks=c(0,400,800,1200,1600,2000)**.



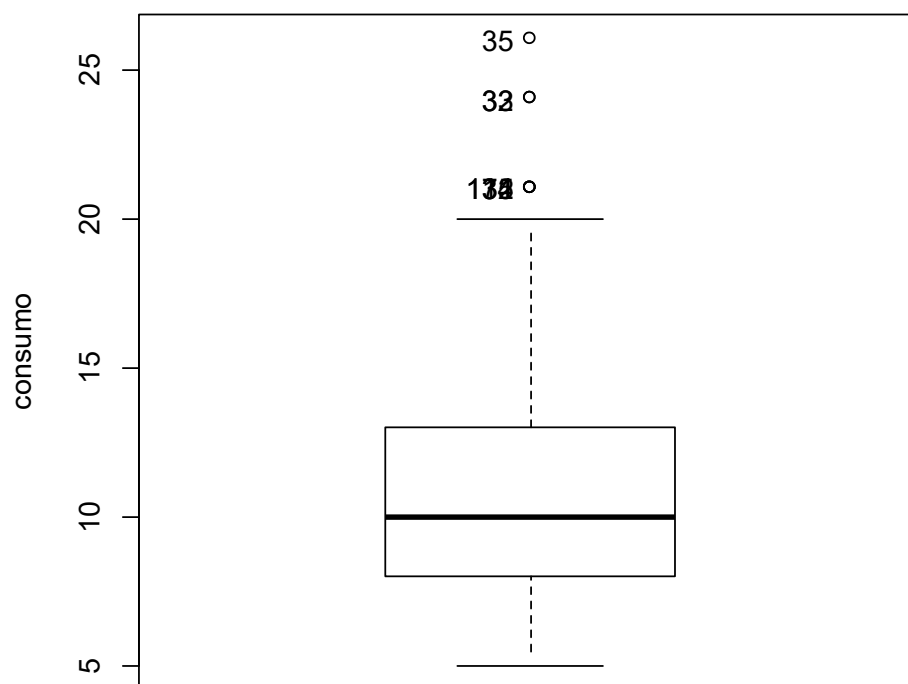
**6. Calcula los cuartiles de la variable consumo e interprétalos. Construye el diagrama de cajas de esta variable.**

Para calcular los cuartiles de una variable tendremos que ir a **Estadísticos-Resúmenes-Resúmenes numéricos...** y en el apartado Estadísticos seleccionar la opción cuartiles e indicar los que queramos realizar.

| 0% | 25% | 50% | 75% | 100% | n   | NA |
|----|-----|-----|-----|------|-----|----|
| 5  | 8   | 10  | 13  | 26   | 398 | 8  |

Esto nos indica los valores de la variable en ciertos intervalos, en este caso el 0, 20, 50, 75 y 100 % de la variable. Podemos deducir entonces que el menor consumo es 5 ya que este es el valor del  $P_0$ . También podemos deducir que hay menos coches con un consumo mayor de 13, ya que de  $P_{75}$  a  $P_{100}$  hay un cambio mas grande que en los demás y en un intervalo de la misma amplitud.

Para realizar el diagrama de cajas tendremos que ir a **Gráficas-Diagrama de caja...** y seleccionar la variable de la cual queremos realizarlo.



## 7. Analiza las variables peso total y aceleración según la variable factor creada en el apartado 3. Obtén las medidas mas relevantes y los diagramas de cajas asociados a cada variable según el grupo.

Como podemos ver en la gráfica de cajas correspondientes, la variable peso es dependiente del factor consumo creado en el apartado 3. Esto lo podemos notar ya que, al hacer una reordenación de variable adecuada desde **“Datos-Modificar variables...-Reordenar niveles de factor...”**, las cajas están en escalera. Esto significa que la variable peso tiene valores diferentes según en grupo de consumo en el que se encuentren.

En el caso de la variable acelera, las cajas son muy parecidas y están casi alineadas, excepto la última, que esta un poco para abajo. Esto significa que la variable acelera tiene valores muy parecidos en todos los grupos de consumo, por lo que podemos deducir que es una variable independiente del consumo.

Haciendo esto con números y más exacto, podemos afirmar lo mismo:

- En la variable peso, la media de sus valores por cada grupo de consumo va ascendiendo contra mayor sea el grupo. Por tanto, podemos afirmar que es una variable dependiente.

|       | mean      | sd       | se(mean) | IQR    | cv        | skewness   | kurtosis   | 0%  | 25%    | 50%    | 75%     | 100% | peso:n |
|-------|-----------|----------|----------|--------|-----------|------------|------------|-----|--------|--------|---------|------|--------|
| Bajo  | 757.2959  | 123.4077 | 9.49290  | 182.00 | 0.1629584 | 1.0712497  | 1.3749238  | 537 | 665.0  | 729.0  | 847.00  | 1241 | 169    |
| Medio | 961.0000  | 138.3934 | 13.97985 | 221.25 | 0.1440098 | 0.1569341  | -0.7282358 | 702 | 851.0  | 960.5  | 1072.25 | 1300 | 98     |
| Alto  | 1301.1221 | 206.0014 | 17.99842 | 275.50 | 0.1583259 | -1.1668236 | 4.3976866  | 244 | 1170.5 | 1318.0 | 1446.00 | 1713 | 131    |

- En la variable acelera, la media de sus valores es muy parecida en todos los grupos de consumo. Por tanto, podemos afirmar que es una variable independiente.

|       | mean     | sd       | se(mean)  | IQR  | cv        | skewness  | kurtosis  | 0%   | 25%    | 50%   | 75%    | 100% | acelera:n |
|-------|----------|----------|-----------|------|-----------|-----------|-----------|------|--------|-------|--------|------|-----------|
| Bajo  | 16.53609 | 2.445367 | 0.1881052 | 3.30 | 0.1478806 | 0.6973730 | 0.8186263 | 11.3 | 14.700 | 16.40 | 18.000 | 24.8 | 169       |
| Medio | 16.24082 | 2.225014 | 0.2247604 | 2.55 | 0.1370014 | 0.8355378 | 0.9046317 | 12.5 | 14.825 | 15.95 | 17.375 | 23.5 | 98        |
| Alto  | 13.73969 | 2.681867 | 0.2343158 | 3.30 | 0.1951912 | 0.4954292 | 0.2904263 | 8.0  | 12.000 | 13.50 | 15.300 | 21.0 | 131       |

