

PRÁCTICA 2**2.2. ESTADÍSTICOS DESCRIPTIVOS: MEDIDAS DE CENTRALIZACIÓN, LOCALIZACIÓN, DISPERSIÓN Y FORMA.**

- Para las variables cuantitativas: mínimo (**Min.**), primer cuartil (**1st Qu.**), mediana ó segundo cuartil (**Median**), media (**Mean**), tercer cuartil (**3rd Qu.**) y máximo (**Max.**).
- Para las variables cualitativas (de tipo factor), se muestran los nombres de las modalidades de la variable así como su frecuencia absoluta.

Ejemplo 1: Obtener e interpretar el resumen de los datos contenidos en el fichero *RCars*. Determinar la moda de las variables origen y cilindros.

Estadísticos → Resúmenes → Conjunto de datos activos

consumo	motor	cv	peso	acel	año	origen
Min. : 5.00	Min. : 66	Min. : 46.00	Min. : 244.0	Min. : 8.00	Min. : 70.00	EE.UU.:253
1st Qu.: 8.00	1st Qu.:1708	1st Qu.: 75.75	1st Qu.: 741.2	1st Qu.:13.62	1st Qu.:73.00	Europa: 73
Median :10.00	Median :2434	Median : 95.00	Median : 936.5	Median :15.50	Median :76.00	Japón : 79
Mean :11.23	Mean :3180	Mean :104.83	Mean : 989.5	Mean :15.50	Mean :75.94	NA's : 1
3rd Qu.:13.00	3rd Qu.:4806	3rd Qu.:129.25	3rd Qu.:1203.8	3rd Qu.:17.07	3rd Qu.:79.00	
Max. :26.00	Max. :7456	Max. :230.00	Max. :1713.0	Max. :24.80	Max. :82.00	
NA's :8		NA's :6			NA's :1	
cilindr						
Min. :3.000						
1st Qu.:4.000						
Median :4.000						
Mean :5.469						
3rd Qu.:8.000						
Max. :8.000						
NA's :1						

Para la variable Consumo, obtenemos que el valor mínimo es 5 y el máximo 26 lo que significa que, en los coches de la muestra, el mínimo consumo observado es de 5L/100Km. Y el mayor consumo es de 26L/100Km. Asimismo, observamos que el consumo medio es de 11.23L/100Km.

Tendremos que convertir la variable de tipo numérico a factor, como es el caso de la variable cilindr.

Datos → Modificar variables del conjunto de datos activos → Convertir variable numérica en factor.

Ejemplo 2: Determinar la media y el percentil 30 de la potencia en c.v. así como el peso de los coches para cada uno de los orígenes definidos en el fichero. Repetir el estudio anterior para los grupos de coches definidos por el número de cilindros.

Primero calcularemos la potencia (cv) y el peso (peso) según el origen

Estadísticos → Resúmenes → Resúmenes numéricos

Resúmenes numéricos

Datos Estadísticos

Variables (seleccione una o más)

- año
- cilindr
- consumo
- cv
- motor
- peso

Resumir según: origen

Ayuda Reiniciar Aceptar Cancelar Aplicar

Añadimos 0.30 (Percentil 30)

Resúmenes numéricos

Datos Estadísticos

☒ Media
☐ Desviación típica
☐ Error típico de la media
☐ Rango Intercuartílico
☐ Coeficiente de variación
☐ Recuentos binarios de frecuencias
☐ Asimetría ☒ Tipo 1
☐ Apuntamiento ☐ Tipo 2
☐ Tipo 3
☒ Cuantiles: 0, .25, .5, .75, 1, .30

Ayuda Reiniciar Aceptar Cancelar Aplicar

```
Variable: cv
      mean 0% 25% 50% 75% 100% 30% n NA
EE.UU. 119.60643 52 88.0 105 150.0 230 90 249 4
Europa 81.00000 46 69.5 77 90.5 133 71 71 2
Japón 79.83544 52 67.0 75 95.0 132 67 79 0

Variable: peso
      mean 0% 25% 50% 75% 100% 30% n NA
EE.UU. 1122.1146 600 906 1126 1351.0 1713 953.4 253 0
Europa 810.1233 608 688 748 933.0 1273 705.0 73 0
Japón 740.0506 537 661 718 803.5 976 669.4 79 0
```

Ahora calcularemos lo mismo, pero según sus cilindr.

Resúmenes numéricos

Datos Estadísticos

Variables (seleccione una o más)

- año
- cilindr
- consumo
- cv
- motor
- peso

Resumir según: cilindr_fac

Ayuda Reiniciar Aceptar Cancelar Aplicar

Resúmenes numéricos

Datos Estadísticos

☒ Media
☐ Desviación típica
☐ Error típico de la media
☐ Rango Intercuartílico
☐ Coeficiente de variación
☐ Recuentos binarios de frecuencias
☐ Asimetría ☒ Tipo 1
☐ Apuntamiento ☐ Tipo 2
☐ Tipo 3
☒ Cuantiles: 0, .25, .5, .75, 1, .30

Ayuda Reiniciar Aceptar Cancelar Aplicar

Variable: cv

	mean	0%	25%	50%	75%	100%	30%	n	NA
3	99.25000	90	95.25	98.5	102.50	110	96.3	4	0
4	78.47030	46	68.00	78.0	88.75	115	70.0	202	5
5	82.33333	67	72.00	77.0	90.00	103	73.0	3	0
6	101.50602	72	92.50	100.0	110.00	165	95.0	83	1
8	158.13084	90	140.00	150.0	175.00	230	145.0	107	0

Variable: peso

	mean	0%	25%	50%	75%	100%	30%	n	NA
3	799.0000	708	759.00	791	831.00	906	769.2	4	0
4	770.5507	537	681.50	744	857.50	1090	701.6	207	0
5	1034.0000	943	963.00	983	1079.50	1176	967.0	3	0
6	1065.7500	824	979.75	1067	1143.25	1302	993.4	84	0
8	1366.1028	1028	1266.00	1378	1460.00	1713	1288.6	107	0

Ejemplo 3: Calcular la desviación típica y la varianza de los datos correspondientes a la aceleración de los vehículos.

Estadísticos → Resúmenes → Resúmenes numéricos

SOLUCIÓN: DESVIACIÓN TÍPICA = 2.820984

VARIANZA: (DESVIACIÓN TÍPICA)² = 7,957951

Ejemplo 4: Calcular la desviación típica y el coeficiente de variación de las variables *consumo* y *peso*. Comparar los resultados obtenidos.

Estadísticos → Resúmenes → Resúmenes numéricos

Resúmenes numéricos

Datos Estadísticos

Variables (seleccione una o más)

- acel
- año
- cilindr
- consumo
- cv
- motor

Resumir por grupos...

Ayuda Reiniciar Aceptar Cancelar Aplicar

Resúmenes numéricos

Datos Estadísticos

☐ Media
☒ Desviación típica
☐ Error típico de la media
☐ Rango Intercuartílico
☒ Coeficiente de variación
☐ Recuentos binarios de frecuencias
☐ Asimetría ☒ Tipo 1
☐ Apuntamiento ☐ Tipo 2
☐ Tipo 3
☐ Cuantiles: 0, .25, .5, .75, 1, .30

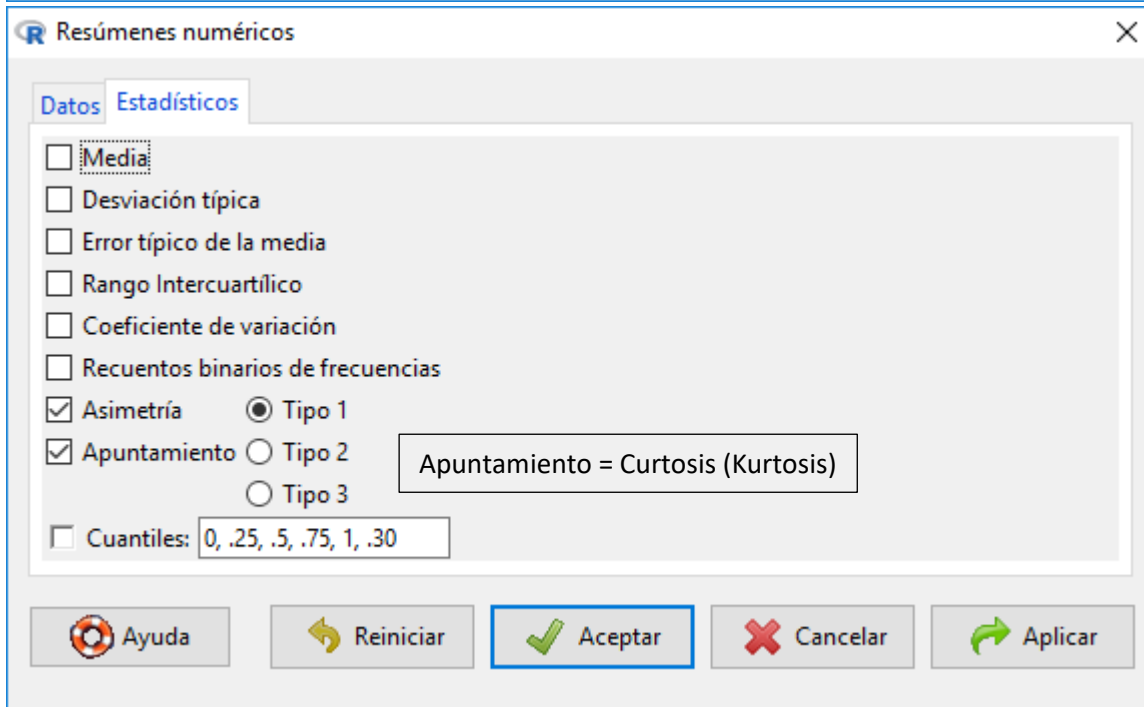
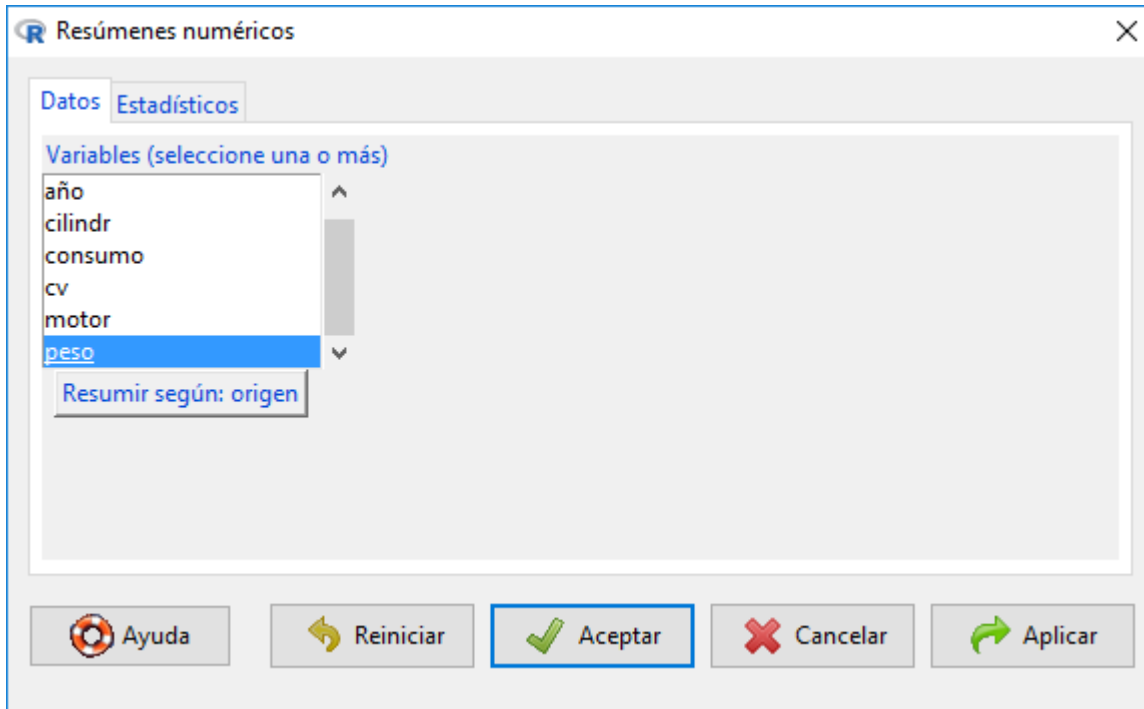
Ayuda Reiniciar Aceptar Cancelar Aplicar

	sd	cv	n	NA
consumo	3.946172	0.3514380	398	8
peso	283.277160	0.2862796	406	0

Siendo **sd** la desviación típica, y **cv** el coeficiente de variación

Ejemplo 5: Determinar e interpretar los coeficientes de asimetría y de apuntamiento de la variable *peso*, para cada uno de los orígenes.

Estadísticos → Resúmenes → Resúmenes numéricos



```

      skewness  kurtosis peso:n
EE.UU. 0.05525765 -0.9577358 253
Europa 0.75713056 -0.3857040 73
Japón 0.48769387 -0.4305033 79

```

Siendo **skewness** la Asimetría, y **kurtosis** el Apuntamiento

Ejemplo 6: Determinar la media, desviación típica y mediana del consumo de los coches según su origen y número de cilindros.

Estadísticos → Resúmenes → Tabla de estadísticas

Primero calcularemos la Media

Tabla de estadísticas

Factores (elegir uno o más)

cilindr_fact
origen

Variables explicadas (elegir una o más)

acel
año
cilindr
consumo
cv
motor

Estadístico

☒ Media
☐ Mediana
☐ Desviación típica
☐ Rango intercuartílico
☐ Otro (especificar)

Ayuda Reiniciar Aceptar Cancelar Aplicar

El espacio **Factores**, sirve para elegir las variables tipo factor que se usarán para hacer la agrupación.

El espacio **Variables explicadas**, elegiremos la/s variable/s para las que se desea calcular el *Estadístico*.

	EE.UU.	Europa	Japón
3	NA	NA	11.500000
4	8.680556	8.634921	7.666667
5	NA	9.000000	NA
6	12.243243	12.750000	10.166667
8	16.215686	NA	NA

Cuando aparece **NA**, es que no hay ninguna observación. En este caso, aparece NA en 3, por que no hay ningún dato de coches de 3 cilindros con tal media de consumo en EE.UU ni en Europa.

Esto pasa en algunos casos más, como 5 y 8 cilindros.

Ahora, la Mediana

Tabla de estadísticas

Factores (elegir uno o más)

cilindr_fact
origen

Variables explicadas (elegir una o más)

acel
año
cilindr
consumo
cv
motor

Estadístico

☐ Media
☒ Mediana
☐ Desviación típica
☐ Rango intercuartílico
☐ Otro (especificar)

Ayuda Reiniciar Aceptar Cancelar Aplicar

	EE.UU.	Europa	Japón
3	NA	NA	11.5
4	9.0	9	7.0
5	NA	9	NA
6	12.0	14	10.5
8	16.5	NA	NA

Y por último, la Desviación típica

Tabla de estadísticas

Factores (elegir uno o más): cilindr_fact, origen

Variables explicadas (elegir una o más): acel, año, cilindr, consumo, cv, motor

Estadístico:

☐ Media

☐ Mediana

☒ Desviación típica

☐ Rango intercuartílico

☐ Otro (especificar):

Ayuda Reiniciar Aceptar Cancelar Aplicar

	EE.UU.	Europa	Japón
3	NA	NA	1.290994
4	1.351331	1.834276	1.491808
5	NA	3.000000	NA
6	1.788937	3.201562	1.940790
8	2.738170	NA	NA

Ejercicio 7: Determinar la media del consumo de los coches en función de su origen, del número de cilindros y del año de fabricación (nótese que para ello es necesario categorizar previamente la variable *Año*).

PENDIENTE DE HACER

2.3. DISTRUBUCIONES DE FRECUENCIAS. TABLAS DE DOBLE ENTRADA

Ejemplo 8: Determinar cuantos coches de cada origen hay en la muestra. Determinar también qué porcentaje, del total de casos estudiados, representa cada origen.

Estadísticos → Resúmenes → Distribución de frecuencias

Distribuciones de frecuencias

Variables (seleccione una o más): año_fact, cilindr_fact, origen

☐ Test Chi-cuadrado de bondad de ajuste (sólo para una variable)

Ayuda Reiniciar Aceptar Cancelar Aplicar


```
counts:
origen
EE.UU. Europa  Japón
    253    73    79

percentages:
origen
EE.UU. Europa  Japón
    62.47  18.02  19.51
```

Siendo **counts**, las frecuencias absolutas para las modalidades de la variable origen.

Y siendo **percentages**, los porcentajes de coches que hay en dichos orígenes.

También es posible calcular frecuencias absolutas y porcentajes de las modalidades definidas, de manera conjunta, por dos o más factores. Para ello:

Estadísticos → Tablas de Contingencia → Tabla de doble entrada

Calcula frecuencias mediante tablas en las que la agrupación viene definida

Y también del menú:

Estadísticos → Tablas de Contingencia → Tabla de entradas múltiples

Permite agrupar utilizando tres o más factores.

Ejemplo 9: Responder a las siguientes cuestiones:

a) ¿Cuántos coches europeos tienen 4 cilindros?

Estadísticos → Tablas de contingencia → Tabla de doble entrada

Tabla de doble entrada

Datos Estadísticos

Variable de fila (elegir una)

año_fact
cilindr_fact
origen

Variable de columna (elegir una)

año_fact
cilindr_fact
origen

Expresión de selección

<todos los casos válidos>

Ayuda Reiniciar Aceptar Cancelar Aplicar

Tabla de doble entrada

Datos Estadísticos

Calcular porcentajes

☐ Porcentajes por filas

☐ Porcentajes por columnas

☐ Porcentajes totales

☒ Sin porcentajes

Test de hipótesis

☒ Test de independencia Chi-cuadrado

☐ Componentes del estadístico Chi-cuadrado

☐ Imprimir las frecuencias esperadas

☐ Test exacto de Fisher

Ayuda Reiniciar Aceptar Cancelar Aplicar

Frequency table:

```

      origen
cilindr_fact EE.UU. Europa  Japón
      3         0         0       4
      4        72        66      69
      5         0         3       0
      6        74         4       6
      8       107         0       0
  
```

Como podemos observar, obtenemos que hay 66 casos de coches que son Europeos y tienen 4 cilindros.

Pearson's Chi-squared test

data: .Table

X-squared = 185.79, df = 8, p-value < 2.2e-16

b) ¿Cuál es el número de cilindros más frecuente en los coches de EE.UU.?

Frequency table:

```

      origen
cilindr_fact EE.UU. Europa  Japón
      3         0         0       4
      4        72        66      69
      5         0         3       0
      6        74         4       6
      8       107         0       0
  
```

Como podemos observar, el número de cilindros más frecuente en EE.UU es de 8, ya que obtenemos 107 casos.

Pearson's Chi-squared test

data: .Table

X-squared = 185.79, df = 8, p-value < 2.2e-16

c) ¿Qué porcentaje de los coches observados tienen 6 cilindros?

Estadísticos → Tablas de contingencia → Tabla de doble entrada

Tabla de doble entrada

Datos Estadísticos

Variable de fila (elegir una) Variable de columna (elegir una)

año_fact
cilindr_fact
origen

año_fact
cilindr_fact
origen

Expresión de selección

< todos los casos válidos

Ayuda Reiniciar Aceptar Cancelar Aplicar

Tabla de doble entrada

Datos Estadísticos

Calcular porcentajes

☐ Porcentajes por filas

☐ Porcentajes por columnas

☒ Porcentajes totales

☐ Sin porcentajes

Test de hipótesis

☒ Test de independencia Chi-cuadrado

☐ Componentes del estadístico Chi-cuadrado

☐ Imprimir las frecuencias esperadas

☐ Test exacto de Fisher

Ayuda Reiniciar Aceptar Cancelar Aplicar

Total percentages:

	EE.UU.	Europa	Japón	Total
3	0.0	0.0	1.0	1.0
4	17.8	16.3	17.0	51.1
5	0.0	0.7	0.0	0.7
6	18.3	1.0	1.5	20.7
8	26.4	0.0	0.0	26.4
Total	62.5	18.0	19.5	100.0

Pearson's Chi-squared test

data: .Table

X-squared = 185.79, df = 8, p-value < 2.2e-16

La respuesta a esta pregunta la encontramos en la fila de los 6 cilindros, y en la columna Total con respecto a dicha fila. Por lo tanto, la **SOLUCIÓN: 20.7%**

d) ¿Qué porcentaje de los coches observados son europeos?

Total percentages:

	EE.UU.	Europa	Japón	Total
3	0.0	0.0	1.0	1.0
4	17.8	16.3	17.0	51.1
5	0.0	0.7	0.0	0.7
6	18.3	1.0	1.5	20.7
8	26.4	0.0	0.0	26.4
Total	62.5	18.0	19.5	100.0

Pearson's Chi-squared test

data: .Table

X-squared = 185.79, df = 8, p-value < 2.2e-16

La respuesta a esta pregunta la encontramos en la fila Total, y en la columna Europa con respecto a dicha fila. Por lo tanto, la **SOLUCIÓN: 18%**

e) ¿Qué porcentaje de coches son europeos y tienen 6 cilindros?

Total percentages:

	EE.UU.	Europa	Japón	Total
3	0.0	0.0	1.0	1.0
4	17.8	16.3	17.0	51.1
5	0.0	0.7	0.0	0.7
6	18.3	1.0	1.5	20.7
8	26.4	0.0	0.0	26.4
Total	62.5	18.0	19.5	100.0

Pearson's Chi-squared test

data: .Table

X-squared = 185.79, df = 8, p-value < 2.2e-16

La respuesta a esta pregunta la encontramos en la fila de los 6 cilindros, y en la columna Europa con respecto a dicha fila. Por lo tanto, la **SOLUCIÓN: 1%**

f) ¿En qué origen hay mayor porcentaje de coches con 4 cilindros?

Estadísticos → Tablas de contingencia → Tabla de doble entrada

Tabla de doble entrada

Datos

Estadísticos

Variable de fila (elegir una)

Variable de columna (elegir una)

año_fact

cilindr_fact

origen

año_fact

cilindr_fact

origen

Expresión de selección

< todos los casos válidos

<

>

Ayuda

Reiniciar

Aceptar

Cancelar

Aplicar

Tabla de doble entrada






Datos Estadísticos

Calcular porcentajes

☐ Porcentajes por filas
☒ Porcentajes por columnas
☐ Porcentajes totales
☐ Sin porcentajes

Test de hipótesis

☒ Test de independencia Chi-cuadrado
☐ Componentes del estadístico Chi-cuadrado
☐ Imprimir las frecuencias esperadas
☐ Test exacto de Fisher

 Ayuda
  Reiniciar
  Aceptar
  Cancelar
  Aplicar

```
Column percentages:
      origen
cilindr_fact EE.UU. Europa  Japón
3           0.0    0.0    5.1
4          28.5   90.4   87.3
5           0.0    4.1    0.0
6          29.2    5.5    7.6
8          42.3    0.0    0.0
Total     100.0  100.0  100.0
Count     253.0   73.0   79.0

Pearson's Chi-squared test
```

```
data: .Table
X-squared = 185.79, df = 8, p-value < 2.2e-16
```

Nos encontramos que:

EE.UU: 28.5% Europa: 90.4% Japón: 87.3%

SOLUCIÓN: Europa (90.4%)

g) De los coches de 4 cilindros, ¿cuál es el porcentaje de coches japoneses?

Estadísticos → Tablas de contingencia → Tabla de doble entrada

Tabla de doble entrada

Datos Estadísticos

Variable de fila (elegir una) Variable de columna (elegir una)

año_fact
cilindr_fact
origen

año_fact
cilindr_fact
origen

Expresión de selección

< todos los casos válidos

Ayuda Reiniciar Aceptar Cancelar Aplicar

Tabla de doble entrada

Datos Estadísticos

Calcular porcentajes

☒ Porcentajes por filas

☐ Porcentajes por columnas

☐ Porcentajes totales

☐ Sin porcentajes

Test de hipótesis

☒ Test de independencia Chi-cuadrado

☐ Componentes del estadístico Chi-cuadrado

☐ Imprimir las frecuencias esperadas

☐ Test exacto de Fisher

Ayuda Reiniciar Aceptar Cancelar Aplicar

Row percentages:

```

      origen
cilindr_fact EE.UU. Europa  Japón  Total  Count
3           0.0    0.0 100.0    100     4
4          34.8   31.9  33.3    100    207
5           0.0  100.0   0.0    100     3
6          88.1    4.8   7.1    100    84
8         100.0    0.0   0.0    100   107
  
```

Pearson's Chi-squared test

data: .Table

X-squared = 185.79, df = 8, p-value < 2.2e-16

El porcentaje de coches japoneses con 4 cilindros, es **SOLUCIÓN: 33.3%**