Capítulo 3: Describir datos con estadísticas

### **Encontrar la mediana**

La *mediana* de un conjunto de números es otro tipo de media. Para hallar la mediana, ordenamos los números en orden ascendente. Si la longitud de la lista de números es impar, el número situado en el centro de la lista es la mediana. Si la longitud de la lista de números es par, obtenemos la mediana tomando la media de los dos números centrales. Busquemos la mediana de la lista anterior de donaciones: 100, 60, 70, 900, 100, 200, 500, 500, 503, 600, 1000 y 1200.

Después de ordenar de menor a mayor, la lista de números se convierte en 60, 70, 100, 100, 200, 500, 500, 503, 600, 900, 1000 y 1200. Tenemos un número par de elementos en la lista (12), así que para obtener la mediana, tenemos que tomar la media de los dos números centrales. En este caso, los números centrales son el sexto y el séptimo -500 y 500- y la media de estos dos números es (500 + 500)/2, que da 500. Esto significa que la mediana es 500. Eso significa que la mediana es 500.

Ahora supongamos -sólo para este ejemplo- que tenemos otro total de donaciones para el 13º día, de modo que la lista tiene ahora este aspecto: 100, 60, 70, 900, 100, 200, 500, 500, 503, 600, 1000, 1200 y 800.

Una vez más, tenemos que ordenar la lista, que pasa a ser 60, 70, 100, 100, 200, 500, 500, 503, 600, 800, 900, 1000 y 1200. En esta lista hay 13 números (un número impar), por lo que la mediana de esta lista es simplemente el número del medio. En este caso, es el séptimo número, que es 500.

Antes de escribir un programa para hallar la mediana de una lista de números, pensemos cómo podríamos calcular automáticamente los elementos medios de una lista en cualquier caso. Si la longitud de una lista*(N*) es impar, el número del medio es el que está en la posición*(N* + 1)/2. Si *N* es par, los dos elementos del medio son *N/2* y*(N/2*) + 1. En nuestro primer ejemplo de esta sección, *N* = 12, por lo que los dos elementos del medio eran los elementos 12/2 (sexto) y 12/2 + 1 (séptimo). En el segundo ejemplo, *N* = 13, por lo que el séptimo elemento,*(N* + 1)/2, era el elemento medio.

Para escribir una función que calcule la mediana, también necesitaremos ordenar una lista en orden ascendente. Por suerte, el método sort() hace precisamente eso:

>>> samplelist = [4, 1, 3]  
>>> samplelist.sort()  
>>> samplelist  
[1, 3, 4]

Ahora podemos escribir nuestro siguiente programa, que halla la mediana de una lista de números:

'''  
Calculating the median  
'''  
  
def calculate\_median(numbers):  
➊ N = len(numbers)  
➋ numbers.sort()  
  
# Find the median  
if N % 2 == 0:  
# if N is even  
m1 = N/2  
m2 = (N/2) + 1  
# Convert to integer, match position  
➌ m1 = int(m1) - 1  
➍ m2 = int(m2) - 1  
➎ median = (numbers[m1] + numbers[m2])/2  
else:  
➏ m = (N+1)/2  
# Convert to integer, match position  
m = int(m) - 1  
median = numbers[m]  
  
return median  
  
if \_\_name\_\_ == '\_\_main\_\_':  
donations = [100, 60, 70, 900, 100, 200, 500, 500, 503, 600, 1000, 1200]  
  
median = calculate\_median(donations)  
N = len(donations)  
print('Median donation over the last {0} days is {1}'.format(N, median))

La estructura general del programa es similar a la del programa anterior que calcula la media. La función calculate\_median() acepta una lista de números y devuelve la mediana. En ➊, calculamos la longitud de la lista y creamos una etiqueta, N, para referirnos a ella. A continuación, en ➋, ordenamos la lista utilizando el método sort().

A continuación, comprobamos si N es par. Si es así, encontramos los elementos del medio, m1 y m2, que son los números que ocupan las posiciones N/2 y (N/2) + 1 en la lista ordenada. Las dos sentencias siguientes(➌ y ➍) ajustan m1 y m2 de dos formas. En primer lugar, utilizamos la función int() para convertir m1 y m2 en enteros. Esto se debe a que los resultados del operador de división siempre se devuelven como números en coma flotante, aunque el resultado sea equivalente a un entero. Por ejemplo:

>>> 6/2  
3.0

No podemos utilizar un número de coma flotante como índice en una lista, así que utilizamos int() para convertir ese resultado en un entero. También restamos 1 tanto a m1 como a m2 porque las posiciones en una lista empiezan por 0 en Python. Esto significa que para obtener el sexto y el séptimo número de la lista, tenemos que preguntar por los números del índice 5 y del índice 6. En ➎, calculamos la mediana tomando la media de los dos números de las posiciones intermedias.

A partir de ➏, el programa encuentra la mediana si hay un número impar de elementos en la lista, utilizando de nuevo int() y restando 1 para encontrar el índice adecuado. Por último, el programa calcula la mediana de la lista de donaciones y la devuelve. Cuando ejecutas el programa, éste calcula que la mediana es 500:

Median donation over the last 12 days is 500.0

Como puedes ver, la media (477,75) y la mediana (500) están bastante próximas en esta lista concreta, pero la mediana es un poco más alta.

[anterior](ch03_2.html)[Subtema 3 de 10: (Ver todo)](ch03.html)[siguiente](ch03_4.html)