

Hoja de Actividades

Matemática y lingüística: La ley de Zipf

La siguiente tabla muestra las diez primeras palabras de la Base de datos de la RAE:

Base de datos de la Real Academia Española		
Ranking (n)	Palabra	Frecuencia (f)
1	de	9 999 518
2	la	6 277 560
3	que	4 681 839
4	el	4 569 652
5	en	4 234 281
6	y	4 180 279
7	a	3 260 939
8	los	2 618 657
9	se	2 022 514
10	del	1 857 225

¿Será posible estimar la frecuencia de una palabra dada su posición en el ranking?

Actividad 1

1. Grafica la frecuencia de las primeras 10 palabras y luego responde la siguiente pregunta: ¿Qué se puede observar respecto a la forma en que disminuye la frecuencia de las palabras a medida que se avanza en el ranking?
2. Responde las siguientes preguntas:
 - a) ¿Cómo varía la frecuencia a medida que se avanza en el ranking?
 - b) A partir del gráfico, ¿es posible apreciar la frecuencia de la palabra que está en la segunda posición del ranking? ¿Y en la posición 200?
 - c) ¿Qué dificultad presenta este gráfico para visualizar los datos?
3. Completa la tabla con el logaritmo de la frecuencia, aproximado los valores a dos cifras decimales, y luego responde: ¿Cómo cambia la variable frecuencia f al aplicar logaritmo?
4. ¿Qué tipo de función podría modelar el gráfico $\log(n)$ versus $\log(f)$?
5. Responde las siguientes preguntas:
 - a) ¿La pendiente del modelo es positiva o negativa? ¿Dónde cruza el eje Y?
 - b) Observando los datos y la recta, ¿para qué rango de valores se ajusta mejor la recta?

Actividad 2

1. Completa la siguiente tabla:

Ranking	Palabra	f real	f con el modelo	Diferencia positiva
20	al	951 054	6,67	
453	grupos	27 863	6,03	