



APÉNDICE 1E: Medias bayesianas en encuestas abiertas

En encuestas abiertas de preferencias, elección de "rankings" y "tops", es común no usar la media aritmética como resultado final de las preferencias de los participantes. Esto es así porque debemos considerar más ricas y exactas las apreciaciones a una opción que tenga muchas entradas que a otra que tenga muy pocas.

Por ejemplo, si 1000 personas valoran la opción A con una media de 8,5 y sólo 10 personas valoran la opción B con una media de 9 no sería muy lógico ni justo colocar la opción B por encima de la A en el ranking final por que el universo que ha valorado a A es mucho más grande y las valoraciones a B pueden estar "viciadas" por fenómenos fan, posiciones extremas, etc.

En tales casos de número de valoraciones distintas para cada opción se usa lo que se denomina "media bayesiana" cuya fórmula podría ser la siguiente

$$\text{Media Bayesiana} = (v / (v+m)) \times R + (m / (v+m)) \times C$$

Donde

- ✓ **R = media aritmética de un elemento**
- ✓ **v = número de votos de un elemento**
- ✓ **m = mínimo número de votos requeridos para entrar en la lista**
- ✓ **C = la media de todos los votos de todos los elementos**

Al fin y al cabo, el resultado es que se da más importancia a las valoraciones que están soportadas por un gran número de opiniones que a las que están basadas en pocas opiniones y pueden tener un sesgo más acusado. (en el ejemplo de más arriba, suponiendo nota media total un 6 y mínimo de 10 valoraciones para entrar en lista, la opción A tendría 8,48 puntos y la opción B quedaría atrás con una nota de 7,66 puntos. Esta es la fórmula que utiliza, por ejemplo, el sitio web imdb.com (Internet Movie DataBase) para calcular la puntuación de las películas a partir de las valoraciones de los internautas

Existen otras medidas, también con el objetivo de minimizar la influencia de sesgos y posiciones extremas, que podrían efectuarse paralelamente, como son:

- Eliminar un porcentaje de valores extremos tanto por arriba como por abajo (media acotada).
- Usar la mediana en vez de la media para eliminar el posible sesgo de valores atípicos y/o extremos.