

Valores Faltantes

Carlos Malanche

6 de marzo de 2018

Qué es lo que debe hacerse cuando falta información en una serie?

El problema principal viene cuando tenemos dos series de datos al menos, y una de ellas tiene menos datos que la otra. Dependiendo de diversas circunstancias, uno puede hacer cosas para remediar la situación

0.1. Tirar la información

Este es el primer instinto, y muchas veces el más atinado. Para no inventar patrones o tendencias, lo mejor es no inventar la información.

0.2. Reemplazarla

En algunos casos, se prefiere reemplazar la información por un valor adecuado. Supongamos pues que tenemos dos series X y Y , ordenadas de modo que los primeros n_1 elementos de Y corresponden a los n_1 elementos en total que contiene X . Además de estos n_1 valores, Y contiene n_2 elementos que no tienen correspondencia en X . Como primer paso, veamos que ocurre al sustituir con el estimador de la media.

0.2.1. Estimador de la media

Tendremos que la serie X se ve como

$$X = \{x_1, x_2, \dots, x_{n_1-1}, x_{n_1}, \underbrace{\hat{\mu}_x, \dots, \hat{\mu}_x}_{n_2 \text{ elementos}}\} \quad (1)$$

Qué es lo que ha ocurrido en la media?

$$\hat{\mu} = \frac{1}{n_1 + n_2} \sum_{i=1}^{n_1+n_2} x_i = \frac{1}{n_1 + n_2} \left(\sum_{i=1}^{n_1} x_i + \sum_{i=n_1+1}^{n_1+n_2} \hat{\mu}_x \right) \quad (2)$$