by ROMERO MOLINA, DESIRÉ - Thursday, 10 November 2022, 12:52 PM

En vista de que los datos perdidos están generando muchas dudas voy a aclarar algunas cosas:

En primer lugar igual os debería haber indicado que esa base que os he pasado es una base real, es decir, no son datos preparados para que todo funcione perfecto sino un ejemplo de lo que uno se puede encontrar en un estudio real. Por lo tanto, como es normal, siempre hay datos faltantes, NA.

Los NA podéis tratarlos como queráis. Si queréis ser muy finos, lo correcto para ser realistas sería:

* quitar los individuos que tienen todas las variables NA porque son casos donde hubo un aborto y el feto no llegó a la semana 34 de gestación ni evidentemente tenemos medidas suyas al nacimiento.
* También quitaría los casos donde el feto si tiene medidas al nacer pero no en la semana 34 porque esos son casos de bebés prematuros.
* Igualmente quitaría los casos donde no hay medidas al nacer aunque sí en la semana 34, porque fueron casos donde nació en otro hospital y por eso no se tiene esas medidas.
* El resto de individuos que tienen NA salteados en algunas variables no los quitaría, porque si lo hacéis el n baja mucho. Esos los imputaría usando el resto de la información que sí tengo sobre el feto. Así depuro la base pero pierdo el menor número de casos posible.

De todas formas, aunque trabajéis con la base como está, eso no os afecta a la matriz de correlaciones, que al estar formada por correlaciones bivariadas ya tiene en cuenta para ello solo a los individuos que tienen esos datos. Por lo tanto, para el cálculo de las componentes y las correlaciones canónicas, no influye. Para los test es cierto que sí afecta porque al usar distribuciones aproximadas donde se tiene en cuenta el n en el cálculo, realmente se está considerando un n mayor que el real.

No voy a ser estricta con esto. Podéis trabajar con la base tal cual está o ser tan fino como queráis. Lo que no podéis hacer es quitar en cada variables sus NA porque entonces, como ha dicho Guillermo, provocáis disparidad en el n. Si quitáis un NA debéis quitar todas las variables de ese individuo.

Espero haber aclarado esta cuestión.

Saludos.