

Análisis de Datos Longitudinales: cálculo de tamaño de muestra

Grado de Estadística, UAB

Juan R González

Guía para la clase no presencial

Una vez conocemos los principales métodos para analizar datos longitudinales continuos, estamos en disposición de poder diseñar un estudio para contestar una pregunta científica que requiera recoger datos a lo largo del tiempo. Como en otros tipos de diseños (cross-over, ensayos clínicos, ...) determinar el tamaño muestral puede resultar crucial para encontrar diferencias que alcancen la significación estadística deseada al menor coste posible. El artículo que tenéis en el material docente (Barrera-Gómez et al (2013), Stat Med) describe cómo llevar a cabo el cálculo de tamaño muestral en estudios observacionales longitudinales teniendo en cuenta el número de medidas repetidas que maximiza el poder de detectar un efecto dado, para un coste determinado. Los autores han creado una librería en R (`optimalAllocation`) que implementa la metodología descrita. El paquete puede instalarse mediante el .zip file que tenéis en el material docente. Allí también podréis encontrar una *vignette* que explica, paso a paso, como llevar a cabo este tipo de cálculos. Lee el artículo, y luego sigue los ejemplos que están en la vignette intentando reproducirlos en tu ordenador. Tras esto, realiza el siguiente ejercicio comentando brevemente los principales resultados.

Ejercicio

Se quiere diseñar un estudio para estudiar la relación entre los valores de tensión arterial sistólica (variable continua) y estar expuesto a estrés (sí/no). Calcula el diseño óptimo (número de participantes y número de medidas repetidas) para minimizar el coste del estudio y tener una potencia de por lo menos 90% utilizando el paquete de R “`optimalAllocation`” bajo los siguientes supuestos:

- a) Ningún participante cambia su nivel de estrés durante el estudio.
- b) Los participantes pueden tener estrés en algunos periodos y en otros no. La variación intra-sujeto de la variable estrés está caracterizada por $\rho_e=0.8$.
- c) Con el diseño óptimo obtenido en b), dibuja la distribución esperada del número de periodos con estrés de los participantes del estudio.

Realiza todos los cálculos para cada pregunta usando estos datos adicionales:

- Se asume que las diferencias entre los que tienen estrés y los que no se mantiene constante en el tiempo.
- El investigador está dispuesto a tomar un máximo de 10 medidas por sujeto.
- La media de tensión arterial sistólica es de 127, con una varianza de 18 y una correlación entre medidas repetidas de 0.7.
- Se espera detectar una diferencia de tensión arterial de 3 unidades entre los participantes con y sin estrés con un nivel de significación del 5%.
- El porcentaje de estrés en la población es de un 20%. Este porcentaje se asume constante en el tiempo.
- Se espera que al final del estudio habrá un 20% de pérdidas de seguimiento.
- La primera medida de cada sujeto tiene un coste 3 veces superior al resto, porque incluye los costes de reclutamiento de sujetos. El coste de la primera medida es 100 €.