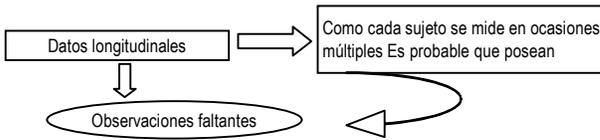


METODOS DE ESTIMACION PARA DATOS LONGITUDINALES CON INFORMACION FALTANTE

García, María del Carmen Blaiconá, María Teresa
 Instituto de Investigaciones Teóricas y Aplicadas, de la Escuela de Estadística

1..INTRODUCCION



4. RESULTADOS

Tabla 1 Estimación de los parámetros del conjunto de datos completos e incompleto con distintos porcentaje de pérdidas del tipo MCAR y MAR para el aglomerado Rosario.

MODELO	COMPLETO	METODO	PORCENTAJE DE PERDIDA						
			MCAR		MAR				
			6%	20%	6%	20%			
ORDENADA 3.0612 0.8788		GENMOD	BCS	3.05099	3.01205	3.108	3.1194		
			SEBC	0.8812	0.89006	0.8775	0.8741		
			BI	3.05238	3.01334	3.108	3.1194		
			SEBI	0.88081	0.89426	0.8776	0.8749		
		RRZ	BCS	3.05069	2.98622	3.0639	3.0679		
			SEBC	0.90028	1.00172	0.8779	0.8806		
			BI	3.05069	2.98622	3.065	3.0681		
			SEBI	0.90028	1.00172	0.8779	0.886		
		EDADV 0.1068 0.0507		GENMOD	BCS	0.1072	0.10893	0.1041	0.103
					SEBC	0.05076	0.05103	0.0506	0.0504
					BI	0.10723	0.10888	0.1041	0.1037
					SEBI	0.05073	0.05127	0.0506	0.0505
RRZ	BCS	0.10773	0.1104	0.1065	0.1064				
	SEBC	0.05159	0.05626	0.0508	0.051				
	BI	0.10773	0.1104	0.1065	0.1064				
	SEBI	0.05159	0.05626	0.0508	0.051				
EDADV2V -0.011 0.0006		GENMOD	BCS	-0.00106	-0.00108	-0.00106	-0.001		
			SEBC	0.00065	0.00065	0.0006	0.0006		
			BI	-0.00107	-0.00108	-0.00106	-0.001		
			SEBI	0.00065	0.00065	0.0006	0.0006		
		RRZ	BCS	-0.00108	-0.00111	-0.00103	-0.0011		
			SEBC	0.00065	0.00071	0.00065	0.0007		
			BI	-0.00108	-0.00111	-0.00103	-0.0011		
			SEBI	0.00065	0.00071	0.00065	0.0007		
		MEDESCOV 0.0961 0.017		GENMOD	BCS	0.09585	0.09584	0.09605	0.0957
					SEBC	0.01703	0.0171	0.01702	0.017
					BI	0.09581	0.09599	0.09606	0.0957
					SEBI	0.01704	0.01717	0.01702	0.017
RRZ	BCS			0.09525	0.09554	0.09605	0.096		
	SEB			0.0178	0.01928	0.01712	0.0172		
	BI			0.09525	0.09554	0.09605	0.0961		
	SEBI			0.0178	0.01928	0.01712	0.0172		

Referencias BCS: Promedio Beta con Simetría compuesta BI: Promedio Beta con Independencia SEBCS: Variancia empírica de los coeficientes de regresión con Simetría compuesta SEBI: Variancia empírica de los coeficientes de regresión con Independencia

2. METODOLOGIA

Para datos longitudinales gaussianos y no gaussianos, Robins, Rotnitzky y Zhao (1995) proponen método estimación cuando existen datos faltantes, modelando la probabilidad de no respuesta. Las ecuaciones de estimación son

$$\sum_{i=1}^n \left[\frac{I(R_i=1)}{\pi_i(1)} d(X_i, \beta) (Y_i - d(X_i; \beta)) \right] = 0$$

siendo, $I(R_i=1) = 1$ si $R_i=1$, es decir si el sujeto i se observa en el tiempo t , y 0 en otro caso,

$\pi_i(1) = P(R_i=1 | W_i)$: probabilidad de observar el vector completo de datos W_i (vector de observaciones y covariables), para el sujeto i .
 Los otros elementos son los mismos que las de las ecuaciones de estimación generalizadas(GEE)

Estas ecuaciones usan los datos sólo para los sujetos con información completa, es decir, sujetos con $R_i=1$, para los cuales se dispone $\pi_i(1)$. Cada sujeto con información completa aporta un término igual a su contribución a las ecuaciones de estimación que se utilizan cuando no hay observaciones faltantes ponderada por la probabilidad inversa de observar el vector de resultados completo Y_i dado W_i .

3.APLICACION

Se modela el ingreso de la fuente laboral de los varones jefe de hogar registrado en la EPH para 4 períodos consecutivos, en función de ciertas variables de interés.
 Variable respuesta: ingreso neto (Y)
 Variables explicativas: edad (edadv), edad cuadrado (edad2v) y escolaridad (medescov)

Procedimiento utilizado :

- 1.- Se estima el modelo para el conjunto completo de datos, utilizando el procedimiento GENMOD (SAS) para dos matrices de covariancias de trabajo distintas.
- 2.- Se realiza un estudio de simulación produciendo, en el conjunto datos completos, distintos porcentajes de pérdidas del tipo MCAR y MAR. Para cada ley de pérdida se estiman los parámetros de cada conjunto incompleto usando:
 - a.- el procedimiento GENMOD para dos matrices de covariancia de trabajo: independencia (I) y simetría compuesta (CS). Este procedimiento utiliza como matriz covariancias trabajo la matriz covariancias empírica calculada con los "casos disponibles" (usa los valores no perdidos). Al estimar los parámetros la contribución de cada unidad se calcula omitiendo los elementos que corresponden a los datos faltantes;
 - b.- el método de RRZ, para las matrices de covariancia de trabajo mencionadas, modela la probabilidad de respuesta mediante un modelo logístico saturado e introduce como pesos en la GEE la inversa de esa probabilidad de pérdida.

5. CONSIDERACIONES FINALES

- Los métodos de estimación convencionales(GEE) sólo permiten pérdidas MCAR
- Cuando los datos perdidos del tipo MAR se puede utilizar el método propuesto por Robins, Rotnitzky y Zhao (1995) (RRZ).
- En este estudio se observa que cuando se generan pérdidas
 - MCAR**
 - GEE y RRZ producen estimadores de magnitud similares
 - MAR**
 - GEE produce estimaciones ligeramente sesgadas
 - RRZ elimina el sesgo pero el error estándar del estimador es más grande que el verdadero

La escasa diferencia que se pueden apreciar en los resultados obtenidos, se puede deber a que los ingresos de cada sujeto no varían mucho entre las ondas.

ESTUDIO DE MONTECARLO

- 1.- Se estiman los coeficientes de regresión y los errores estándares para cada caso.
- 2.- Se realizan 100 repeticiones del proceso.
- 3.- Se calcula el promedio y la variancia empírica de los estimadores.
- 4.- Estos valores se comparan con los obtenidos para los datos completos.