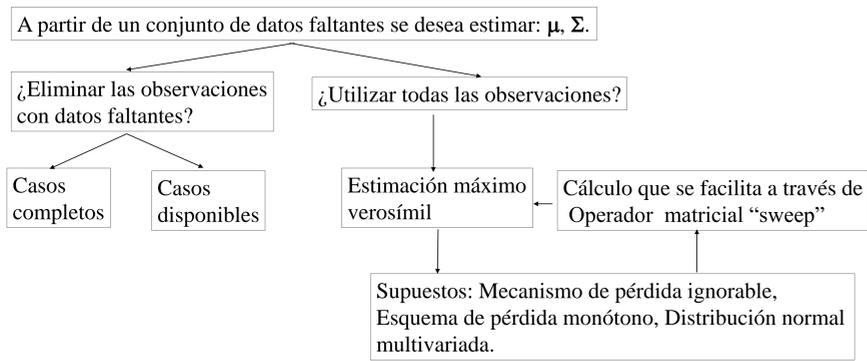


# LA ESTIMACIÓN MÁXIMO VEROSÍMIL COMO ALTERNATIVA PARA EL TRATAMIENTO DE LA FALTA DE INFORMACIÓN EN ENCUESTAS

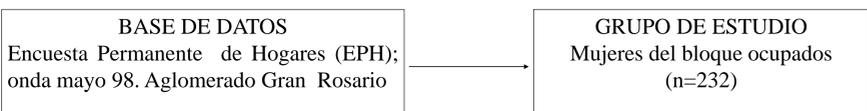
Badler, C.; Alsina, S.; Beltrán, C.; Bussi, J.; Puigsubirá, C.; Vitelleschi, M.S.

Instituto de Investigaciones Teóricas y Aplicadas, Escuela de Estadística

## INTRODUCCIÓN



## MATERIAL



### VARIABLES :

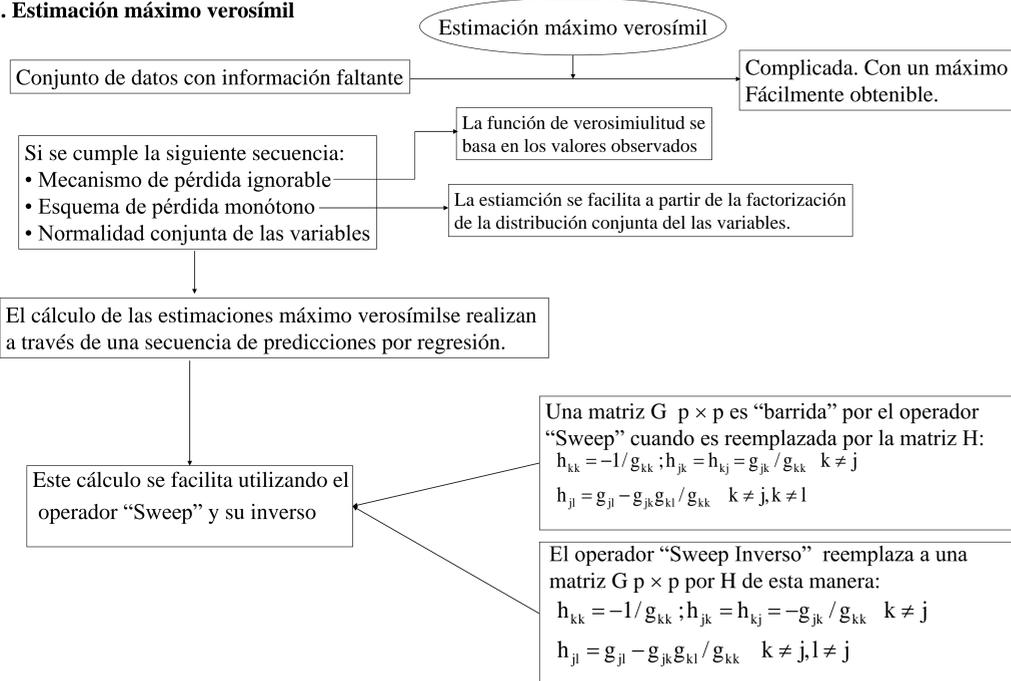
- Edad ( $Y_1$ )
- Total de Horas Trabajadas en la Semana de Referencia ( $Y_2$ )
- Años de Escolaridad ( $Y_3$ )
- Tiempo que Lleva en su Ocupación Principal, en meses ( $Y_4$ )
- Ingreso Horario de la Ocupación Principal ( $Y_5$ )

### VARIABLES TRANSFORMADAS:

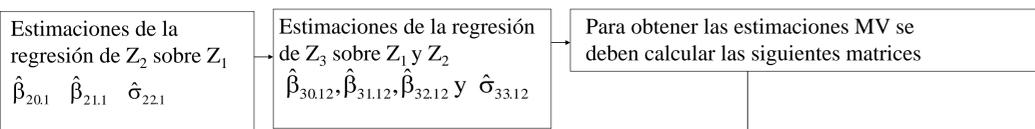
- Raíz Cúbica del Tiempo que Lleva en su Ocupación Principal ( $Y_4^*$ )
- Logaritmo Natural del Ingreso Horario de la Ocupación Principal ( $Y_5^*$ )

## METODOLOGÍA

### 1. Estimación máximo verosímil



Se presenta el cómputo para tres bloques de variables donde las variables  $Z_1$  están completamente observadas



$$A = \text{SWP}[1] \begin{bmatrix} -1 & \hat{\mu}_1^T \\ \hat{\mu}_1 & \hat{\Sigma}_{11} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \quad B = \text{SWP}[2] \begin{bmatrix} a_{11} & A_{1.23} & \hat{\beta}_{20.1} \\ A_{1.23}^T & A_{23.23} & \hat{\beta}_{21.1} \\ \hat{\beta}_{20.1} & \hat{\beta}_{21.1}^T & \hat{\sigma}_{22.1} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} \\ c_{21} & c_{22} & c_{23} & c_{24} \\ c_{31} & c_{32} & c_{33} & c_{34} \\ c_{41} & c_{42} & c_{43} & c_{44} \end{bmatrix}$$

siendo:

$$\begin{bmatrix} -1 & \hat{\mu}_1^T \\ \hat{\mu}_1 & \hat{\Sigma} \end{bmatrix} = \text{SWI}[1,2] \begin{bmatrix} c_{11} & C_{1.23} & c_{14} & \hat{\beta}_{30.12} \\ C_{1.23}^T & C_{23.23} & C_{23.4} & \hat{\beta}_{31.12} \\ c_{41} & C_{23.4}^T & c_{34} & \hat{\beta}_{32.12} \\ \hat{\beta}_{30.12} & \hat{\beta}_{31.12}^T & \hat{\beta}_{32.12} & \hat{\sigma}_{33.12} \end{bmatrix} \quad A_{1.23} = [a_{12} \ a_{13}] \quad A_{23.23} = \begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix}$$

$$C_{1.23} = [c_{12} \ c_{13}] \quad C_{23.23} = \begin{bmatrix} c_{22} & c_{23} \\ c_{32} & c_{33} \end{bmatrix} \quad C_{23.4} = \begin{bmatrix} c_{24} \\ c_{34} \end{bmatrix}$$

### 2. Estimación a través de casos disponibles y casos completos

- Casos completos: se calcula utilizando exclusivamente aquellas unidades en donde todas las variables han sido observadas ( $m_j$ ).
- Casos disponibles: se utilizan todas las observaciones disponibles para cada variable o par de variables.

### 3. Medidas de comparación de las estimaciones obtenidas

#### i) Medias y Desvíos Estándares

##### MV vs. Casos Completos

$$A = \frac{(\bar{y}_j^{(m_j)} - \hat{\mu}_j)}{\hat{\sigma}_j} \cdot 100$$

$$B = \frac{(s_j^{(m_j)} - \hat{\sigma}_j)}{\hat{\sigma}_j} \cdot 100$$

##### MV vs. Casos Disponibles

$$C = \frac{(\bar{y}^{(j)} - \hat{\mu}_j)}{\hat{\sigma}_j} \cdot 100$$

$$D = \frac{(s_j^{(j)} - \hat{\sigma}_j)}{\hat{\sigma}_j} \cdot 100$$

donde:

$\hat{\mu}_j$  y  $\hat{\sigma}_j$  MV  
 $\bar{y}_j^{(m_j)}$  y  $s_j^{(m_j)}$  casos completos  
 $\bar{y}^{(j)}$  y  $s_j^{(j)}$  casos disponibles

#### ii) Covariancias

##### MV vs. Casos Completos

$$CC = \frac{(s_{jk}^{(m_j)} - \hat{\sigma}_{jk})}{\hat{\sigma}_{jk}} \cdot 100$$

##### MV vs. Casos Disponibles

$$CD = \frac{(s_{jk}^{(j)} - \hat{\sigma}_{jk})}{\hat{\sigma}_{jk}} \cdot 100$$

donde:

$\hat{\sigma}_{jk}$  MV  
 $s_{jk}^{(m_j)}$  casos completos  
 $s_{jk}^{(j)}$  casos disponibles

## RESULTADOS

- Selección de variables según cumplimiento del supuesto de normalidad conjunta:  $Y_1, Y_2, Y_4^*$  e  $Y_5^*$ .
- Obtención esquema monótono de pérdida del nuevo conjunto de variables, descartando dos observaciones de  $Y_5^*$ :  $Y_1$  e  $Y_2$  constituyen un bloque de variables completamente observadas e  $Y_4^*$  es más observada que  $Y_5^*$
- Cantidad de observaciones por variable en esquema monótono

$Y_1$	$Y_2$	$Y_4^*$	$Y_5^*$
232	232	226	199

- Consideración de mecanismo de pérdida ignorable a partir de la aplicación del test de Little.
- Estimaciones máximo verosímiles mediante operador "Sweep", con programa desarrollado en S-PLUS.

Tabla 1: Estimaciones máximo verosímiles de la media y el desvío estándar (DS) y comparaciones con estimaciones a partir de casos completos y casos disponibles.

Variable	Max. Verosimil		Casos Completos		Casos Disponibles					
	Media	DS	Media	DS	A	B	Media	DS	C	D
$Y_1$	41.736	10.064	41.058	10.092	-6.737	0.278	41.736	10.086	0.000	0.219
$Y_2$	35.028	21.710	35.462	20.530	1.999	-5.435	35.028	21.757	0.000	0.216
$Y_4^*$	3.991	1.576	3.963	1.556	-1.777	-1.269	3.987	1.580	-0.254	0.254
$Y_5^*$	1.123	0.752	1.113	0.734	-1.330	-2.394	1.113	0.734	-1.330	-2.394

Se observan diferencias entre las estimaciones a partir de casos completos y casos disponibles respecto a máxima verosimilitud.

Tabla 2: Comparación de las estimaciones de las covariancias.

Covariancia entre	CC	CD
$Y_1-Y_2$	808.416	1.856
$Y_1-Y_5^*$	341.935	341.935
$Y_1-Y_4^*$	-5.657	1.326
$Y_2-Y_5^*$	-11.361	-11.361
$Y_2-Y_4^*$	49.328	-35.317
$Y_5^*-Y_4^*$	-4.305	-4.305

Las diferencias en las estimaciones de las covariancias a partir de casos completos con respecto a máxima verosimilitud, son iguales o mayores que las que surgen de casos disponibles. Se destacan dos situaciones extremas mayores al 100%.

## DISCUSIÓN

- La estimación máximo verosímil del vector de medias y la matriz de covariancias se ve facilitada cuando el mecanismo de pérdida es ignorable y el esquema es monótono.
- Si los datos se distribuyen conjuntamente normal se puede utilizar el operador "Sweep" para dichas estimaciones.
- La estimaciones obtenidas por máxima verosimilitud difieren de las obtenidas en casos completos y casos disponibles.

Es preferible aplicar el método de máxima verosimilitud ante la presencia de un esquema monótono de pérdida ya que utiliza mayor cantidad de información que los métodos de casos disponibles y casos completos y su aplicación es relativamente sencilla a través del operador "Sweep".