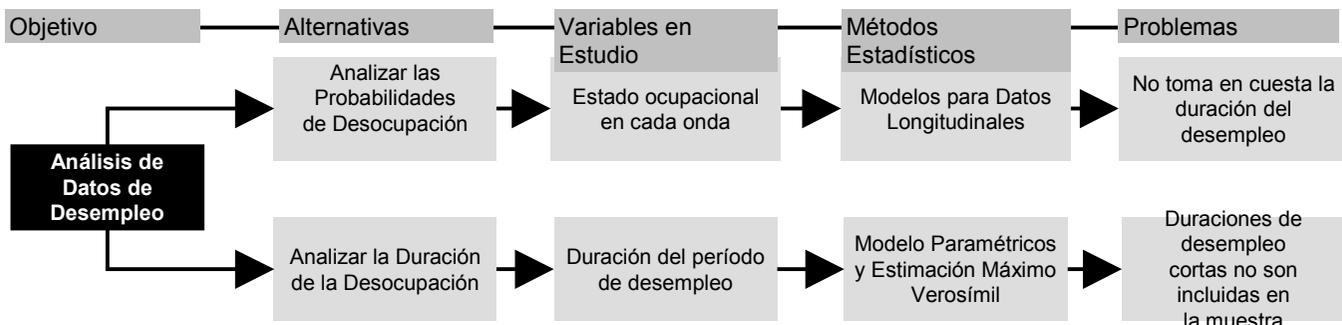


SESGO INDUCIDO POR EL METODO DE MUESTREO EN EL ANALISIS DE DATOS DE DESOCUPACION

Elsa C. Servy, Daniel Wojdyla

Instituto de Investigaciones Teóricas y Aplicadas en Estadística, Escuela de Estadística, Facultad de Ciencias Económicas y Estadística, Universidad Nacional de Rosario



TRUNCAMIENTO

Una condición que "selecciona" ciertos individuos y que provoca que el investigador no conozca su existencia. Al tener datos truncados, solamente los individuos que experimenta un determinado evento son observados por el investigador.

En la EPH, solamente los individuos que están desocupados cuando ingresan en la encuesta son considerados.

CENSURA

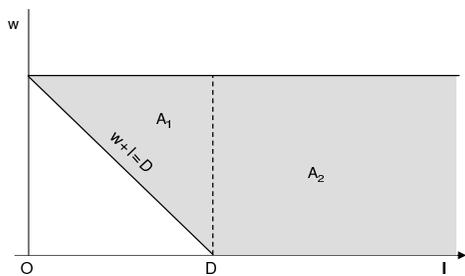
Una observación en una variable T está *censurada por derecha* si todo lo que se conoce es que T es mayor que un valor c, donde T es el tiempo de ocurrencia de un evento.

En la Encuesta Permanente de Hogares, la variable T es el tiempo hasta conseguir empleo y la censura se produce cuando los individuos dejan de participar en la muestra.

OBJETIVO

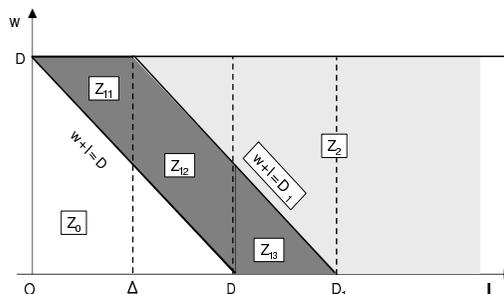
Presentar un modelo teórico que permite describir el truncamiento y la censura de los datos y contruye estimadores de los parámetros de la distribución de las duraciones que corrige los efectos indeseados de la censura y truncamiento.

CASO 1: TRUNCAMIENTO SOLAMENTE



D: momento de la encuesta (comienzo del seguimiento)

CASO 2: TRUNCAMIENTO Y CENSURA



D: momento de la encuesta
 D₁: momento en que el individuo sale de la encuesta
 Δ: ventana de observación

EJEMPLO: DISTRIBUCION EXPONENCIAL con PARAMETRO θ

$$\log L = 2n \log \theta - \theta \sum_{i=1}^n j_i - n \log(1 - \exp(-\theta D))$$

Simulación:

5000 valores de una distribución exponencial con $\theta = 0.04$ y D=7

645 observaciones truncadas

4355 observaciones no truncadas

Estimador ignorando truncamiento: 0.0352

Estimador corregido por truncamiento: 0.0399

$$\log L = (n+n_1) \log \theta - \theta \left(\sum_{i=1}^{n_1} j_i + \sum_{i=n_1+1}^n j_i^* \right) - n \log(1 - \exp(-\theta D))$$

Simulación:

1000 repeticiones para distintos tamaños muestrales (100, 500, y 5000) de una distribución exponencial con $\theta = 0.04$ y D = 7

Tamaño Muestral	$\lambda = (1/\theta)$	Desvío Estándar
100	27.05	8.80
500	25.38	3.32
1000	25.25	2.18
5000	25.08	0.94