

Modelos de Mercadotecnia

Encuestas y sesgo de selección

Jorge de la Vega Góngora

Maestría de Mercadotecnia,
Instituto Tecnológico Autónomo de México

Sesión 10



Introducción

- En cualquier tipo de investigación donde la opinión o percepción de personas es importante, la aplicación de encuestas es muy útil.
- Una encuesta es un sistema para obtener información primaria cualitativa y cuantitativa de una población.
- Las encuestas son empleadas por las empresas para conocer:
 - la aceptación de sus productos y/o de los servicios que ofrecen,
 - determinar las posibilidades de éxito de un nuevo producto o servicio y las características de los consumidores que les pueden resultar más relevantes.

- Cambio tecnológico:
 - Prácticamente las encuestas telefónicas han desaparecido o son muy difíciles de administrar
 - El *junkmail* hace difícil las encuestas por e-mail.
 - El costo de las encuestas cara a cara son muy altas.
- Las encuestas en línea ofrecen nuevas oportunidades y retos.
- Sin embargo, es dudoso que las encuestas en línea reemplacen los métodos tradicionales siempre será una alternativa o complemento.
- Las encuestas en línea no son apropiadas para todos los proyectos. Factores a tomar en cuenta:
 - Tamaño de muestra y distribución geográfica
 - Restricciones temporales
 - Anonimato
 - Target
 - Marco muestral: las respuestas online son más probables cuando se contacta por email a los participantes.
 - Tasas de respuesta: usualmente las encuestas en línea tienen tasas de respuesta promedio de 30 %-50 %.

Comparación de metodologías de encuestas

Tipo	Ventajas	Desventajas
Correo	<ul style="list-style-type: none">● Bajo Costo● Amplia cobertura geográfica● No sesgo de entrevistador● Anonimato permite preguntas sensibles	<ul style="list-style-type: none">● Baja tasa de respuesta● Periodos de respuesta largos● Preguntas contingentes efectivas● No se sabe quién responde la encuesta
Teléfono	<ul style="list-style-type: none">● Se limita el sesgo de la cobertura● Respuestas rápidas● Se pueden hacer preguntas complejas● Alcance geográfico amplio	<ul style="list-style-type: none">● Confusión con llamadas de ventas● Intrusivas● No soporte visual● Filtrado de llamadas (desde el teléfono)
Entrevistas cara a cara	<ul style="list-style-type: none">● Buenas tasas de respuesta● Se pueden hacer preguntas complejas● Se toleran entrevistas largas	<ul style="list-style-type: none">● Alcance geográfico limitado● Consumen mucho tiempo● Muy caras● Sesgo de entrevistador● Tópicos sensibles son difíciles de explorar
Online	<ul style="list-style-type: none">● Bajo costo● Rápidas● Eficientes● Preguntas sensibles● Captura directa de datos● Amplio alcance geográfico	<ul style="list-style-type: none">● Sesgo de cobertura● Depende del software● No se sabe quién responde a la encuesta

- Cuando se comparan las encuestas vía e-mail con las de correo:
 - Las tasas de respuesta en línea son menores (posiblemente por el diseño de la encuesta)
 - La velocidad de respuesta es mayor
 - Se obtiene misma o mejor calidad de datos.
- En México, las encuestas por correo ordinario no son muy factibles.

Sesgo de selección

- La selección de muestra es un problema genérico en la investigación social que surge cuando un investigador no observa una muestra aleatoria de la población de interés.
- Cuando las observaciones se seleccionan de tal manera que no son independientes de las variables resultado del estudio, esta selección de muestra lleva a inferencias sesgadas.
- Muchos de la investigación en mercadotecnia y en general en las ciencias sociales presentan sesgos de selección de muestra.
- La selección es amplia y resulta natural en estudios de conducta humana (Roy 1951, Heckman, 1974)

- Queremos estimar el efecto de la educación sobre el ingreso de una muestra de personas con ingresos menores a \$15,000.
- Los individuos se muestrean en tres niveles educativos: bajo (B), medio (M) y alto (A).

Definición

- Cuando las observaciones con valores de la variable dependiente (ingreso) son mayores a una cierta cota son excluidos, la muestra resultante es *truncada*
- Otro tipo de selección explícita se da cuando la muestra incluye a personas con ingresos mayores a \$15,000 o más, pero todo lo que se conoce de tales personas es que pueden tener un ingreso mayor a un cierto umbral. Cuando la variable dependiente está fuera de cierto rango pero no se conoce el valor exacto, se obtienen *muestras censuradas*. La censura puede ser por arriba o por abajo.
- El censurado y truncamiento puede ocurrir cuando se reporta una variable dependiente hasta cierto nivel
- En ambos casos, de truncamiento o censura, El modelo estima de manera sesgada los parámetros.

- En un modelo de regresión lineal, la selección ocurre cuando los datos sobre y se 'pierden' de manera no aleatoria, condicional en las variables independientes.
- En este caso, los métodos estadísticos elementales usualmente dan estimaciones sesgadas e inconsistentes de los efectos de las variables independientes.

Ejemplo

Si se usan mínimos cuadrados ordinarios para estimar un modelo de regresión donde los valores grandes de la variable dependiente están subrepresentados en la muestra, las estimaciones de los coeficientes de las variables dependientes podrían ser sesgadas.

- Los datos corresponden al tiempo en meses hasta que un recluso es arrestado después de ser liberado de la prisión. Algunos reclusos participaron en un programa de trabajo mientras estuvieron en la cárcel. Se consideran también variables demográficas así como de historia de la prisión y criminal. De un total de 1,445 reclusos, 893 no han sido arrestados durante el periodo de seguimiento; por lo tanto, estas observaciones son censuradas. El tiempo de censura difiere entre reclusos y va de 70 a 81 meses.

```
library(survival)
data(recid, package = 'wooldridge')

# Estima el modelo de regresión
regOLS <- lm(log(durat) ~ workprg + priors + tserve + felon + alcohol + drugs + black + married + educ + age, data = recid)

# define la variable dummy para las observaciones no censuradas.
recid$uncensored <- recid$cens == 0

# estima el modelo de regresión censurado:
regCensor <- survreg(Surv(log(durat), uncensored, type = "right") ~ workprg + priors + tserve + felon + alcohol + drugs + black + married + educ + age, data = recid,
                     dist = "gaussian")
```

Ejemplo II

```
library(stargazer) # paquete para mostrar los resultados de la regresión
stargazer(regOLS, regCensor, type = "text", keep.stat = "n")
```

=====		
	Dependent variable:	

	log(durat)	log(durat)
	OLS	Tobit
	(1)	(2)

workprg	0.009 (0.049)	-0.063 (0.120)
priors	-0.059*** (0.009)	-0.137*** (0.021)
tserve	-0.009*** (0.001)	-0.019*** (0.003)
felon	0.179*** (0.058)	0.444*** (0.145)
alcohol	-0.263*** (0.060)	-0.635*** (0.144)
drugs	-0.091* (0.055)	-0.298** (0.133)
black	-0.179*** (0.047)	-0.543*** (0.117)
married	0.134** (0.055)	0.341** (0.140)
educ	0.005 (0.010)	0.023 (0.025)









Estimación tipo Heckmann





