CI53I/CI73A Demanda de Transporte
Semestre Otoño 2005
Profesores: Marcela Munizaga y Pedro Donoso
1 hora 30 minutos

Control 1

Pregunta 1

Suponga que se ha realizado una encuesta a una muestra aleatoria simple de personas de una comuna, preguntándoles el número de viajes a caminata que realizaron en un determinado día de la semana. Juzgue la veracidad o falsedad de las siguientes preguntas, justificando su respuesta

1. Si una persona seleccionada en la muestra se negase a responder la encuesta, entonces, su reemplazo por otra que contesta provoca un cambio en los errores muestrales y no muestrales de toda la encuesta.
   ***Verdadero**:*
   El reemplazo de una persona en una muestra provoca un cambio en el error muestral porque cambia el método de selección de la muestra y con ello, el conjunto de todas las muestras posibles. La nueva muestra ya no es Aleatoria Simple, sino que fue elegida en función de una muestra anterior y bajo la condición que una persona conteste, aspecto que debe tenerse presente para el cálculo de su probabilidad de selección. Esto provoca que la varianza del estimador (una medida del cuadrado del error muestral) cambie.
   Por otra parte, el reemplazo de una persona en una muestra que no contesta elimina el error no muestral de no-respuesta, y por lo tanto, cambian los errores no muestrales. Si la nueva respuesta fuese incorrecta aparecería otro error no muestral, situación que se desconoce.
   Por lo tanto, los errores muestrales y no muestrales cambian, pero no necesariamente son mayores o menores.

2. El promedio simple del número de viajes a caminata en la muestra de personas es un estimador insesgado del promedio simple de todas las personas de la comuna
   *Esta afirmación es **verdadera bajo las siguientes condiciones**:*
   - *Solamente si no existen errores no muestrales y*
   - *Si la muestra fue elegida por Muestreo Aleatorio Simple.*

   En este caso, según el enunciado se cumple la segunda condición, pero no necesariamente la primera.
   Por ejemplo, si una persona omite algún viaje está entregando una respuesta errónea que desvía al promedio en la muestra del promedio en la población.

3. Solamente una capacitación adecuada de los encuestadores contribuye a reducir los errores no muestrales
   Falso:
   Si bien una capacitación adecuada a los encuestadores contribuye a disminuir los errores no muestrales, por ejemplo, al instruirlos para que formulen completamente las

preguntas con el objeto de que ellas se comprendan adecuadamente por el entrevistado, esta capacitación no impide la existencia de entrevistados que podrían, por ejemplo, no desear contestar la encuesta (error no muestral de no-respuesta) o responder erróneamente por equivocación si la pregunta es compleja (error no muestral de reporte incorrecto). Por estas razones, es conveniente investigar a los entrevistados sobre posibles motivaciones para no contestar y las razones para contestar erróneamente, ya sea intencional o inconscientemente. Para ello se pueden realizar, grupos focales, encuestas preparatorias de formularios y encuestas de profundidad.

4. Las encuestas piloto permiten detectar los errores no muestrales
   ***Falso**:*
   Las encuestas piloto corresponden a una prueba a una escala reducida de todas las actividades de terreno y de oficina de la encuesta. En consecuencia, no permiten, en general, detectar respuestas con información incorrecta, sólo se pueden detectar errores por no-respuesta.

5. En la medida que el tamaño de la muestra aumenta disminuye el error muestral
   ***Falso**:*
   Si el estimador utilizado es sesgado, es decir, su valor esperado no coincide con el número de viajes de caminata de todas las personas de la comuna, aunque se aumente el tamaño de la muestra, su error no necesariamente disminuirá.

6. La varianza del promedio simple del número de viajes en la muestra mide la dispersión de los valores reportados por los entrevistados
   ***Falso**:*
   La varianza de un estimador insesgado mide el valor esperado de la variabilidad del estimador con respecto al parámetro poblacional a lo largo de todas las muestras posibles.

Pregunta 2 (2/3)

a) Suponga que usted es contraparte de un estudio de demanda de transporte, en que el consultor calibró modelos lineales de generación de viajes. Los modelos entregados tienen muy buen ajuste, y tienen la particularidad de que en los casos en que había una zona que generaba mucho más viajes que las otras, se agrega una dummy aditiva que se hace cargo de esa varianza. ¿Qué opinión le merece ese modelo? ¿Qué observaciones le haría al consultor?

Ese modelo tiene un problema, ya que al incluir una variable dummy aditiva asociada a la zona de mayor generación, se está anulando la información proveniente de esa zona, y el ajuste es "engañosamente bueno". Le diría al consultor que debe hacerse cargo de modelar correctamente para esa zona, buscando tal vez especificaciones no lineales, o incorporando aquellas variables no incorporadas en la formulación original que puedan ser la causa de una diferencia tan grande en la generación de viajes.

b) Suponga que requiere modelar el efecto de la variable ingreso en un modelo de generación de viajes por hogar, pero en las discusiones de equipo se ha decidido intentar una especificación en que el efecto de la variable ingreso sea distinto para los distintos tramos de edad. Diga cómo podría modelar ese efecto. Sea muy explícito(a) en la definición de las variables involucradas y escriba el modelo algebraicamente.

Para esto, es necesario primero que nada definir una estratificación de acuerdo a edad. Si se está modelando personas, puede ser directamente un rango (0-18, 19-25,...), pero en este caso, que se está modelando hogares, es algo más complejo, se puede usar, por ejemplo: (hogares sólo adultos – hogares con niños; segmentación según rangos de edad promedio del hogar, ...). Luego, esta variable debe incorporarse a la modelación multiplicando la variable Ingreso. Suponiendo que se generó N-1 variables dummy asociadas a las categorías definidas (D_n=1 si hogar pertenece a esa categoría de edad, y cero en caso contrario).

$$Y_h = \ldots + (b I_1 \times D_1 + b_2 I \times D_2 + b_3 I \times D_3 + \ldots) \times Ingreso_h$$

c) ¿Qué nos dice el enfoque Manheim sobre la generación de viajes? ¿Qué influencia tiene eso en la modelación?

El enfoque Manheim nos dice que la generación de viajes depende de la interacción entre el sistema de transporte (T) y el sistema de actividades (A). Esto implica que cuando se pretende modelar la generación de viajes, se debe incluir variables que representen esa relación: variables que representen el nivel de actividad de las zonas (población, actividad comercial, educacional, etc.) y variables que representen la facilidad/dificultad de moverse (accesibilidad)..

d) Proponga un modelo de generación de viajes que permita incorporar la estructura de "quiebres en la vida" vista en clases. Indique: qué tipo de modelo usaría, qué datos necesitaría, cómo lo calibraría, cómo lo usaría en modalidad predictiva.

Lo más natural es proponer un modelo de análisis por categorías o análisis de clasificación múltiple. La estructura de "quiebres en la vida" vista en clases correspondía a una clasificación de hogares que genera conjuntos disjuntos (pareja que vive sóla, la llegada del primer hijo, el primer hijo va al colegio, ...) y en este caso el supuesto del modelo ACM es sustentable (que los distintos tipos de hogar generan distinta cantidad de viajes). Se puede incorporar además otras variables, como ingreso y/o tasa de motorización. Los datos que requeriría para calibrar el modelo son: generación de viajes a nivel de hogar, variables que me permitan categorizar los hogares según la etapa de la vida en que se encuentran y según las otras variables que decida incorporar (ingreso, tasa de motorización). La calibración se puede hacer usando el método visto en clases de calcular la media global y las desviaciones asociadas a la media según cada categoría de cada variable. Para usarlo en modalidad predictiva necesito predecir la distribución futura de las distintas categorías de hogares en la población.

e) En los modelos alternativos que se presenta a continuación, indique, en cada caso, cuál es el supuesto que se plantea en cada una de las alternativas (analice comparativamente A y B) sobre la forma funcional del modelo. Indique en cada caso qué prueba estadística se puede usar para saber cuál es la especificación correcta desde ese punto de vista.

e.1.: (A) $Y = \beta 0 + \beta_1$·Número de estudiantes + β_2·Número de trabajadores

(B) $Y = \beta 0 + \beta_1$·(Número de estudiantes + Número de trabajadores)

Aquí hay un modelo (A) que plantea que la variable Y depende linealmente del número de estudiantes y del número de trabajadores. La especificación (B) plantea que el efecto marginal de ambas variables es el mismo (b único). La especificación (B) es una versión restringida de (A) que plantea la hipótesis lineal $b_1=b_2$, la cual se puede probar estadísticamente usando el test F.

e.2.: (A) $Y = \beta 0 + \beta_1$·Número de estudiantes + β_2·(Número de estudiantes)^2

(B) $Y = \beta 0 + \beta_1$·Número de estudiantes

El modelo (A) plantea que la variable Número de estudiantes afecta a la variable dependiente Y según un polinomio en que hay un término lineal y uno cuadrático. La especificación (B) plantea que no hay efecto de segundo orden ($b_2=0$). También en este caso (B) es una versión restringida de (A) que plantea la hipótesis lineal $b_2=0$, la cual se puede probar estadísticamente usando el test F y usando el test t.

e.3.: (A) $Y = \beta 0 + \beta_1$·Número de estudiantes + β_2·Número de trabajadores

(B) $Y = \beta 0 + \beta_1$·Número de trabajadores

El modelo (A) plantea que las variables Número de estudiantes y Número de trabajadores afectan a la variable dependiente Y en forma lineal. La especificación (B) plantea que la variable Número de estudiantes no tiene influencia en explicar la varianza de Y. Nuevamente en este caso (B) es una versión restringida de (A) que plantea la hipótesis lineal $b_1=0$, la cual se puede probar estadísticamente usando el test F y usando el test t.