

TAREA 1
IN540 “MÉTODOS ESTADÍSTICOS PARA LA ECONOMÍA Y GESTIÓN”

PROFESOR: Cristóbal Huneus
PROF. AUXILIARES: Tadashi Takaoka
SEMESTRE: Primavera 2006

FECHA ENTREGA: Martes 4 de Abril

1. Sean X_1 y X_2 dos variables aleatorias e independientes. Cada una de ellas puede tomar dos valores: 0 y 1. El valor 1 lo toma con probabilidad p y el valor 0 con probabilidad $1 - p$. Definimos dos estimadores de $\theta = p(1 - p)$ basados en X_1 y X_2 como sigue:

$$\hat{\theta}_1 = (X_1 + X_2 - 2X_1X_2)/2 \quad (1)$$

$$\hat{\theta}_2 = X_1(1 - X_2) \quad (2)$$

Cuál estimador prefiere? Justifique tu respuesta claramente.

2. Las variables aleatorias de (X, Y) toman dos valores, 0 y 1. La distribución de (X, Y) es: $P(X = 1, Y = 1) = 2/8$, $P(X = 1, Y = 0) = 2/8$, $P(X = 0, Y = 1) = 1/8$, $P(X = 0, Y = 0) = 3/8$.

(a) Muestra que $X + Y$ y $X - (20/19)Y$ no están correlacionados.

(b) Son $X + Y$ y $X - (20/19)Y$ independientes?

3. Dos variables aleatorias, X e Y son independientes con $E(X) = 1$ y $V(X) = 1$, $E(Y) = 2$ y $V(Y) = 1$. Sea $Z = X + Y$ y $W = XY$. Calcula $Cov(Z, W)$.

4. Suponga que $y_t = y_t^* + u_t$ y $x_t = x_t^* + v_t$, $t = 1 \dots T$. Donde $\{y_t^*\}$ y $\{x_t^*\}$ son constantes desconocidas, $\{y_t\}$ y $\{x_t\}$ son variables aleatorias que se observan y $\{u_t\}$ y $\{v_t\}$ son variables aleatorias que no se observan. Suponga que (u_t, v_t) se distribuye como $N(0, 0, \sigma_u^2, \sigma_v^2, \sigma_{uv})$. El problema consiste en estimar β en la relación $y_t^* = \beta x_t^*$ $t = 1, 2, \dots, T$ sobre la base de observar sólo y_t y x_t .

(a) Obtenga el límite en probabilidades de

$$\hat{\beta} = \frac{\sum_{t=1}^T y_t x_t}{\sum_{t=1}^T x_t^2}$$

Para ello asuma que:

$$\lim_{T \rightarrow \infty} \sum_{t=1}^T (x_t^*)^2 = c$$

(b) Obtenga el limite en probabilidades de

$$\hat{\beta} = \frac{\sum_{t=1}^T y_t}{\sum_{t=1}^T x_t}$$

Para ello asuma que:

$$\lim_{T \rightarrow \infty} \sum_{t=1}^T x_t^* = d \neq 0$$

(c) Cuál estimador es mejor para estimar β , el estimador de la parte (a) o el de la parte (b)? Justifique tu respuesta claramente.

5. **[Empírico]** Usando la base de datos tarea1.dta conteste las siguientes preguntas.¹ La base de datos corresponde a los ingresos de hombres de la encuesta CASEN 1992-2003. Suponga que el modelo es el siguiente:

$$\ln(y_{i,t}) = X\beta + u_{i,t}$$

donde $y_{i,t}$ es el ingreso de la persona i en el año t , X son características observables de la persona y $u_{i,t}$ son factores no observables que influyen en el salario. Supondremos que $E(u_{i,t}|X) = 0$.

(a) Suponga que X contiene una constante, edad y edad al cuadrado. Un alumno dice que para encontrar β_1 en la siguiente ecuación da lo mismo estimar:

$$\ln g = \alpha + \beta_1 \text{Edad} + \beta_2 \text{Edad}^2 + u_{i,t}$$

que estimar la siguiente ecuación:

$$\text{Edad} = -\frac{\alpha}{\beta_1} + \frac{1}{\beta_1} \ln g - \frac{\beta_2}{\beta_1} \text{Edad}^2 - \frac{1}{\beta_1} u_{i,t}$$

donde $\ln g$ es el logaritmo del ingreso. Estime el modelo de las dos formas. Indique cuál es el valor de β_1 y β_2 según ambos métodos. De acuerdo a los supuestos del modelo cuál es la manera correcta de estimar el modelo.

(b) Estima el siguiente modelo (muestre el resultado de la regresión):

$$\begin{aligned} \ln g &= \alpha + \beta_1 \text{Edad} + \beta_2 \text{Edad}^2 + \beta_3 \text{Casado} + \beta_4 \text{Soltero} + \beta_5 \text{Propia} + \beta_6 \text{Pagandose} \\ &+ \beta_7 \text{informal} + \beta_8 \text{Ded2} + \beta_9 \text{Ded3} + \theta_1 \text{Dyr}^* + \theta_2 \text{Dr}^* + u_{i,t} \end{aligned} \quad (3)$$

donde Ded2 y Ded3 son dummies de la variable neduc cuando tiene el valor 2 y 3 respectivamente. Dyr^* son dummies de la variable ano y Dr^* son dummies de la variable r (región).

Estima el siguiente modelo (muestre el resultado de la regresión):

$$\ln g^* = \alpha + \beta_1 \text{Edad}^* + u_{i,t} \quad (4)$$

donde $\ln g^*$ y Edad^* son las variables construidas a partir de $\ln g$ y Edad respectivamente, después de haber removido toda variación de las demás variables que aparecen en la ecuación 3. Compare los valores de β_1 que encuentro cuando estimo la ecuación 3 y la ecuación 4. Por qué son diferentes/iguales? Qué puede decir sobre los errores standard de β_1 de ambas regresiones?

¹Esta es la misma base de datos que discutimos en clases el Jueves.

- (c) Cuál es el efecto de un año adicional sobre el ingreso de acuerdo a los resultados de la ecuación 3.
- (d) Existe alguna edad a la cuál deja de crecer el ingreso? Cuál es esa edad? Para calcular la edad use el promedio de *ling*.
- (e) Estima la siguientes ecuación:

$$ling = \alpha + \theta Casado + v_{i,t} \quad (5)$$

donde Casado es una dummy igual a 1 si la persona esta casada y 0 sino. Esta definida en la base de datos. Cuál es la diferencia entre el ingreso de los casados y no casados?

- (f) Estima la siguientes ecuación:

$$ling = \alpha + \theta Casado + \gamma_1 Ded2 + \gamma_2 Ded3 + v_{i,t} \quad (6)$$

donde *Ded2* y *Ded3* son dummies de la variable neduc cuando tiene el valor 2 y 3 respectivamente. Cuál es la diferencia entre el ingreso de los hombres casados y no casados? Existe alguna diferencia en los estimadores de θ entre el resultado de la ecuación 5 y 7? A qué cree usted que se puede deber esa diferencia? Le puede servir para responder esta pregunta estimar la siguiente regresión:

$$Casado = \theta_0 + \lambda_1 Ded2 + \lambda_2 Ded3 + \omega_{i,t} \quad (7)$$

Fijate en los signos de λ_1 y λ_2 y piensa que dicen sobre la relación entre la educación y estar casado.