

Instituto Tecnológico Autónomo de México
Maestría en Economía Aplicada

Econometría Aplicada I

Taller

Taller 007

Descargue el archivo de datos **CollegeDistance** que contiene datos de una muestra aleatoria de alumnos de último año de secundaria entrevistados en 1980 y vuellos a entrevistar en 1986. En este ejercicio, se utilizarán estos datos con el fin de investigar la relación entre el número de años de educación completados por adultos jóvenes y la distancia de la escuela secundaria de cada estudiante a la universidad más cercana. (La proximidad a la universidad reduce el coste de la educación, por lo que los estudiantes que viven más cerca de una universidad deberían, en promedio, completar más años de educación superior). La descripción detallada de las variables de las que se dispone aparece en el archivo **CollegeDistance-Description**.

- 1) Realice una regresión de la variable años de educación completados (ED) sobre la variable de la distancia a la universidad más cercana (Dist). ¿Cuál es la pendiente estimada? Interprete este coeficiente y su significancia.
- 2) Ahora estime una regresión de la variable ED sobre la variable Dist, pero incluyendo algunos regresores adicionales de control sobre las características del estudiante, la familia del estudiante, y el mercado laboral local. Específicamente, incluya como regresores adicionales las variables Bytest, Female, Black, Hispanic, Incomehi, Ownhome, DadColl, Cue80 y Stwmfg80. ¿Cuál es el efecto estimado de la variable Dist sobre la variable ED?
- 3) ¿Es sustancialmente distinto el efecto estimado de la variable Dist sobre la variable ED en la regresión de (2) de la regresión en (1)? En base a esto, ¿considera que la regresión (1) presenta un sesgo de variable omitida importante?

- 4) Compare el ajuste de la regresión de (1) y de (2) utilizando los errores estándar de regresión, R^2 y \bar{R}^2 . ¿Por qué R^2 y \bar{R}^2 son tan similares en la regresión (2)?
- 5) El valor del coeficiente de la variable DadColl es positivo. ¿Qué mide este coeficiente? Interprete su resultado a partir de un intervalo al 95 % de confianza.
- 6) Explique por qué las variables Cue80 y Swmfg80 aparecen en la regresión. ¿Cuáles cree que deban ser los signos de sus coeficientes (+ o -)? Interprete la magnitud de estos coeficientes y verifique su signo a partir de la prueba de hipótesis o su intervalo al 95 % de confianza.
- 7) Bob es un hombre negro. Su escuela secundaria estaba a 20 millas de la universidad más cercana. Su calificación en la prueba (Bytest) fue de 58. Su ingreso familiar en 1980 fue de \$26,000 y su familia poseía una casa. Su madre acudió a la universidad, pero su padre no. La tasa de desempleo en su condado era del 7.5 %, y el promedio del salario por hora manufacturero en su estado era de \$9.75. Estime el número de años completados de estudio por Bob utilizando la regresión de (2).
- 8) Jim tiene las mismas características que Bob, salvo que su escuela secundaria estaba a 40 millas de la universidad más cercana. Estime los años completados de estudio por Jim utilizando la regresión de (2).
- 9) Se ha argumentado que, controlando por otros factores, los negros y los hispanos completan más cursos universitarios que los blancos. ¿Es esto congruente con los resultados estimados por la regresión (2)?
- 10) Realice una regresión de la variable ED sobre las variables Dist, Female, Bytest, Tuition, Black, Hispanic, Incomehi, Ownhome, DadColl, MomColl, Cue80 y Stwmfg80. Si la variable Dist aumenta de 2 a 3 (es decir, de 20 a 30 millas), ¿cuánto se espera que cambien los años de educación? Si la variable Dist aumenta desde 6 hasta 7 (es decir, desde 60 hasta 70 millas), ¿cuánto se espera que cambien los años de educación?
- 11) Realice una regresión del $\ln(ED)$ sobre las variables Dist, Female, Bytest, Tuition, Black, Hispanic, Incomehi, Ownhome, DadColl, MomColl, Cue80 y Stwmfg80. Si la variable Dist aumenta de 2 a 3 (de 20 a 30 millas), ¿cuánto se espera que cambien los años de educación? Si la variable Dist aumenta de 6 a 7 (es decir, de 60 a 70 millas), ¿cuánto se espera que cambien los años de educación?

- 12) Realice una regresión de la variable ED sobre las variables Dist, Dist^2 , Female, Bytest, Tuition, Black, Hispanic, Incomehi, Ownhome, DadColl, MomColl, Cue80 y Stwmfg80. Si la variable Dist aumenta de 2 a 3 (de 20 a 30 millas), ¿cuánto se espera que cambien los años de educación? Si Dist aumenta de 6 a 7 (es decir, de 60 a 70 millas), ¿cuánto se espera que cambien los años de educación?
- 13) Añada el término de interacción $\text{DadColl} \times \text{MomColl}$ a la regresión de (12). ¿Qué mide el coeficiente del término de interacción?
- 14) Mary, Jane, Alexis y Bonnie tienen los mismos valores para las variables Dist, Bytest, Tuition, Female, Black, Hispanic, Fincome, Ownhome, Cue80 y Stwmfg80. Ninguno de los padres de Mary fue a la universidad. El padre de Jane asistió a la universidad, pero su madre no lo hizo. La madre de Alexis asistió a la universidad, pero su padre no lo hizo. Tanto el padre como la madre de Bonnie fueron a la universidad. Utilizando las regresiones de (12).
- ¿Qué predice la regresión acerca de la diferencia entre los años de educación de Jane y Mary?
 - ¿Qué predice la regresión acerca de la diferencia entre los años de educación de Alexis y Mary?
 - ¿Qué predice la regresión acerca de la diferencia entre los años de educación de Bonnie y Mary?
- 15) Después de realizar todas estas regresiones (y cualquier otra que quiera agregar), resuma y concluya sobre el efecto que la variable Dist tiene sobre los años de educación.

Funciones recomendadas en R

- Utilice la función `confint()` para estimar un intervalo de confianza para los parámetros del modelo (predeterminado al 95 %). Ejemplo:

```
modelo<-lm(Y~X1+X2+X3)
confint(modelo)
confint(modelo,level = 0.90)
```

- Para realizar una predicción a partir de un modelo estimado, utilice el comando `predict()`. Sólo se necesita definir un valor específico para cada variable independiente. Ejemplo:

```
modelo<-lm(Y~X1+X2+X3)
valores<- data.frame(X1=8, X2=0.5, X3=1.2)
predict(modelo,valores)
predict(modelo,valores,interval='confidence',level=0.95)
```

Puede definir varios valores a evaluar utilizando el comando `c()`.