

Taller 007

Taller de Econometría Aplicada I

Julieta Pérez, Esteban Degetau

1. En la Tabla 1 mostramos la relación lineal entre distancia al colegio y años de educación. Encontramos que por cada 10 millas de distancia, los años de educación disminuyen en promedio 0.07 años. La relación inversa confirma la distancia de la universidad es un costo para las personas estudiantes. El coeficiente es estadísticamente significativo al 1%.

```
college |>
  feols(ed ~ dist) |>
  etable(tex = T)
```

2. Al incluir controles como en la Tabla 2, efecto estimado de la distancia a la universidad en años de educación disminuyó respecto de la estimación sin controles en la Tabla 1. En particular, obtenemos que por cada 10 millas adicionales, se estudiará solo 0.03 años menos. Adicionalmente, a pesar de que el coeficiente sigue siendo estadísticamente significativo, lo es solo al 5%.

```
reg_control <- college |>
  feols(ed ~ dist + bytest + female + black + hispanic
        + incomehi + ownhome + dadcoll + cue80 + stwmfg80)

reg_control |>
  etable(tex = T, ci = 0.95, coefstat = "confint")
```

3. Sí puede ser que la regresión en la Tabla 1 esté sesgada, puesto que no se incluyen controles que pueden estar correlacionados con la distancia a la universidad y con los años de educación. Por ejemplo, puede ser que las familias afroamericanas vivan más lejos de las universidades y que también estudien menos años en promedio por razones distintas a la distancia. En ese caso, no incluir la variable `black` puede sesgar nuestros estimadores.
4. La regresión en la Tabla 2 tiene un mejor ajuste a los datos. Presenta un \bar{R}^2 de 0.277 mientras que la regresión en la Tabla 1 tiene un \bar{R}^2 de 0.007. En la regresión con controles,

Tabla 1: Relación entre educación y distancia a la universidad

Dependent Variable:	ed
Model:	(1)
<i>Variables</i>	
Constant	13.96*** (0.0377)
dist	-0.0734*** (0.0138)
<i>Fit statistics</i>	
Observations	3,796
R ²	0.00745
Adjusted R ²	0.00719
<i>IID standard-errors in parentheses</i>	
<i>Signif. Codes: ***: 0.01, **: 0.05, *: 0.1</i>	

en la Tabla 2, los estimadores R^2 y \bar{R}^2 se asemejan mucho porque las variables de control que incluimos agregan mucha información adicional al modelo. Es decir que las variables explicativas no están muy correlacionadas entre sí. El error estándar asociado a la variable `dist` no cambió mucho entre estimaciones, lo que sí cambió fue la magnitud de los coeficientes, que a su vez modificó la significancia estadística de los mismos.

5. El coeficiente asociado a la variable `DadCol` tiene signo positivo y es estadísticamente distinto de cero. Podemos interpretar el intervalo de confianza asociado a él como que 95 por ciento de las estimaciones caerán en el intervalo [0.5617, 0.8309]. Esto quiere decir que, en promedio, los hijos de padres con educación universitaria estudiarán entre 0.56 y 0.83 años más que los hijos de padres sin educación universitaria.
6. Las variables `cue80` y `swmfg80` representan el costo de oportunidad de estudiar. En particular, la tasa de desempleo del condado en 1980 `cue80` indica la facilidad con la que un egresado de preparatoria podrá encontrar un empleo. Si ésta es alta, entonces el costo de oportunidad de estudiar baja porque será más difícil encontrar un empleo, y entonces en mayor medida preferirá estudiar (tendrá signo positivo). Por el lado de `swmfg80`, el salario promedio en manufacturas indica el salario que podría ganar un egresado de preparatoria si no estudiara. Si este salario es alto, entonces el costo de oportunidad de estudiar es alto, y entonces en mayor medida preferirá no estudiar (tendrá signo negativo).

Ambas variables tienen el signo esperado y son estadísticamente distintas de cero. En particular, por cada punto porcentual que aumente la tasa de desempleo, los años de

Tabla 2: Relación entre educación y distancia a la universidad con controles

Dependent Variable:	ed
Model:	(1)
<i>Variables</i>	
Constant	8.828*** [8.337; 9.318]
dist	-0.0315** [-0.0558; -0.0073]
bytest	0.0938*** [0.0876; 0.1000]
female	0.1454*** [0.0462; 0.2446]
black	0.3680*** [0.2281; 0.5079]
hispanic	0.3985*** [0.2525; 0.5445]
incomehi	0.3952*** [0.2765; 0.5139]
ownhome	0.1521** [0.0211; 0.2831]
dadcoll	0.6961*** [0.5614; 0.8309]
cue80	0.0232** [0.0043; 0.0421]
stwmfg80	-0.0518*** [-0.0907; -0.0129]
<i>Fit statistics</i>	
Observations	3,796
R ²	0.27884
Adjusted R ²	0.27693

IID co-variance matrix, 95% confidence intervals in brackets
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Tabla 3: Predicción de años de educación para Bob

fit	se.fit	ci_low	ci_high
14.22	0.25	13.73	14.72

educación aumentarán en 0.023 años. Por cada dólar que aumente el salario por hora promedio en manufacturas, los años de educación disminuirán en 0.052 años.

7. En la Tabla 3 mostramos la estimación de años de educación para Bob. Estudiará 14.22 años en promedio, con un intervalo de confianza de 95% entre 13.73 y 14.72 años.

```
bob <- tibble(
  female = 0,
  black = 1,
  hispanic = 0,
  bytest = 58,
  dadcoll = 0,
  incomehi = 1,
  ownhome = 1,
  cue80 = 7.5,
  stwmfg80 = 9.75,
  dist = 20
)

predict(reg_control, bob, ci = T, interval = "confidence") |>
  kable(digits = 2, booktabs = T)
```

8. En la Tabla 4 mostramos la estimación de años de educación para Jim. Estudiará 13.59 años en promedio, con un intervalo de confianza de 95% entre 12.63 y 14.56 años.

```
jim <- tibble(
  female = 0,
  black = 1,
  hispanic = 0,
  bytest = 58,
  dadcoll = 0,
  incomehi = 1,
  ownhome = 1,
  cue80 = 7.5,
  stwmfg80 = 9.75,
  dist = 40
)
```

Tabla 4: Predicción de años de educación para Jim

fit	se.fit	ci_low	ci_high
13.59	0.49	12.63	14.56

)

```
predict(reg_control, jim, ci = T, interval = "confidence") |>
  kable(digits = 2, booktabs = T)
```

9. En la regresión en la Tabla 2 no podemos hacer aceveraciones sobre la cantidad de cursos que se estudiaron en la universidad. Sin embargo, sí podemos decir que, todo lo demás constante, las personas afroamericanas estudian 0.34 años más que las personas blancas, y las hispanas estudian 0.39 años más que las personas blancas.
10. Como estimamos un modelo lineal, el cambio en los años de educación asociado a un cambio en la distancia a la universidad es constante. Es decir, no depende de la distancia a la universidad. Por lo tanto, un cambio de 60 a 70 millas esta dado por la pendiente en el coeficiente de `dist`. En particular, por cada 10 millas adicionales, los años de educación disminuyen en 0.037 años. Por lo tanto, un cambio de 60 a 70 millas implica una disminución de 0.37 años en los años de educación.

```
reg_10 <- college |>
  feols(ed ~ dist + female + bytest + tuition + black + hispanic + incomehi + ownhome +
  reg_10 |>
  etable(tex = T, ci = 0.95, coefstat = "confint")

reg_11 <- college |>
  feols(log(ed) ~ dist + female + bytest + tuition + black + hispanic + incomehi + ownho
  reg_11 |>
  etable(tex = T, ci = 0.95, coefstat = "confint")
```

11. Al pasar de 20 a 30 millas de distancia de una universidad se pierden 0.0363 años de educación. Al pasar de 60 a 70 millas, se pierden 0.0359 años.
12. Incluyendo distancia como término cuadrático, encontramos que al pasar de 20 a 30 millas de distancia de una universidad se pierden 0.058 años de educación. Al pasar de 60 a 70 millas, se pierden 0.0208 años.

Tabla 5: Relación entre educación y distancia a la universidad con controles adicionales

Dependent Variable:	ed
Model:	(1)
<i>Variables</i>	
Constant	8.921*** [8.428; 9.414]
dist	-0.0367*** [-0.0616; -0.0117]
female	0.1430*** [0.0441; 0.2419]
bytest	0.0930*** [0.0868; 0.0993]
tuition	-0.1911* [-0.3888; 0.0067]
black	0.3506*** [0.2110; 0.4903]
hispanic	0.3618*** [0.2103; 0.5133]
incomehi	0.3718*** [0.2528; 0.4909]
ownhome	0.1385** [0.0077; 0.2694]
dadcoll	0.5710*** [0.4265; 0.7155]
momcoll	0.3778*** [0.2180; 0.5376]
cue80	0.0287*** [0.0093; 0.0480]
stwmfg80	-0.0425** [-0.0821; -0.0029]
<i>Fit statistics</i>	
Observations	3,796
R ²	0.28361
Adjusted R ²	0.28134

IID co-variance matrix, 95% confidence intervals in brackets
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Tabla 6: Relación entre logaritmo de educación y distancia a la universidad con controles adicionales

Dependent Variable:	log(ed)
Model:	(1)
<i>Variables</i>	
Constant	2.266*** [2.231; 2.301]
dist	-0.0026*** [-0.0044; -0.0008]
female	0.0103*** [0.0033; 0.0173]
bytest	0.0067*** [0.0062; 0.0071]
tuition	-0.0139* [-0.0280; 0.0001]
black	0.0262*** [0.0163; 0.0361]
hispanic	0.0260*** [0.0152; 0.0367]
incomehi	0.0265*** [0.0181; 0.0350]
ownhome	0.0098** [0.0005; 0.0191]
dadcoll	0.0405*** [0.0303; 0.0508]
momcoll	0.0266*** [0.0152; 0.0379]
cue80	0.0020*** [0.0007; 0.0034]
stwmfg80	-0.0029** [-0.0057; -5.13×10^{-5}]
<i>Fit statistics</i>	
Observations	3,796
R ²	0.28531
Adjusted R ²	0.28305

*IID co-variance matrix, 95% confidence intervals in brackets
Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

```
reg_13 <- college |>
feols(
  ed ~ dist + dist^2 + female + bytest + tuition + black + hispanic + incomehi + ownhome
)

etable(reg_13, tex = T)
```

13. El coeficiente del término de interacción en la Tabla 7 es significativamente negativo e indica que, todo lo demás constante, tener ambos padres que asistieron a universidad disminuye los años de educación en 0.36 años. A pesar de ser contraintuitivo, porque las variables `dadcoll` y `momcoll` están altamente correlacionadas, su interacción resta a la suma de cada uno de los coeficientes.

```
candidates <- tibble(
  candidato = c("Mary", "Jane", "Alexis", "Bonnie"),
  dist =      c(rep(1, 4)),
  female =    c(rep(1, 4)),
  bytest =    c(rep(58, 4)),
  tuition =   c(rep(2.5, 4)),
  black =     c(rep(0, 4)),
  hispanic =  c(rep(0, 4)),
  incomehi =  c(rep(1, 4)),
  ownhome =   c(rep(1, 4)),
  dadcoll =   c(0, 1, 0, 1),
  momcoll =   c(0, 0, 1, 1),
  cue80 =     c(rep(7.5, 4)),
  stwmfg80 =  c(rep(9.75, 4))
)

predict(reg_13, candidates, ci = T, interval = "confidence") |>
bind_cols(candidato = candidates$candidato) |>
kable(digits = 2, booktabs = T)
```

14. La Tabla 8 muestra resultados consistentes con los coeficientes en la Tabla 7. De hecho, Bonnie que tiene dos padres con educación universitaria es la que más años estudiará (todo lo demás constante). Jane estudiará 0.66 años más que Mary. Alexis estudiará 0.57 años más que Mary. Bonnie estudiará 0.86 años más que Mary.
15. Para concluir, encontramos que la distancia de la universidad al hogar de los candidatos tendrá un efecto negativo pero con tasas decrecientes. Para aumentar los años de educación, los datos sugieren proponer (i) vales de renta para subsidiar la habitación

Tabla 7: Relación entre educación y distancia a la universidad con interacción entre educación de los padres y controles adicionales

Dependent Variable:	ed
Model:	(1)
<i>Variables</i>	
Constant	9.002*** (0.2555)
dist	-0.0810*** (0.0255)
dist square	0.0047** (0.0023)
female	0.1406*** (0.0504)
bytest	0.0926*** (0.0032)
tuition	-0.1940* (0.1008)
black	0.3306*** (0.0717)
hispanic	0.3297*** (0.0785)
incomehi	0.3623*** (0.0608)
ownhome	0.1412** (0.0667)
dadcoll	0.6538*** (0.0843)
momcoll	0.5694*** (0.1173)
cue80	0.0258*** (0.0099)
stwmfg80	-0.0415** (0.0202)
dadcoll \times momcoll	-0.3665** (0.1613)
<i>Fit statistics</i>	
Observations	3,796
R ²	0.28535
Adjusted R ²	0.28271

IID standard-errors in parentheses

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Tabla 8: Predicción de años de educación para candidatos

fit	se.fit	ci_low	ci_high	candidato
14.24	0.16	13.92	14.56	Mary
14.90	0.17	14.56	15.23	Jane
14.81	0.19	14.43	15.19	Alexis
15.10	0.18	14.75	15.44	Bonnie

cerca de las universidades. (ii) Mejorar las vías de transporte a las universidades, para disminuir el costo de asistir. Y (iii) construir más universidades en lugares inhóspitos para proveer educación a personas que no tienen acceso a ella.