Métodos Cuantitativos I

Estudiantes: Fabián Carrillo y Yasna Lobos
R.U.T.: 19.907.537-3 y 21.263.055-1

Tarea 1

Primavera 2025

Pregunta 1. Importación y descriptivos básicos

La base contiene **17 870** observaciones. La **altura promedio** de la muestra es **66.96 pulgadas**. La diferencia de altura por sexo es **5.59** pulgadas (Hombre − Mujer) y es estadísticamente significativa al 1 por ciento.

Cuadro 1: Tabla 1: Estadística descriptiva por sexo

	N	Altura promedio (pulgadas)
Hombre	7896	70,08
Mujer	9974	64,49
Total	17 870	66,96

Notas: La tabla reporta el número de observaciones y la altura promedio por sexo. Cifras redondeadas a dos decimales; alturas en pulgadas. Diferencia de medias Hombre − Mujer: 5,59 pulgadas; t significativa al 1 % con $p < 0{,}001$.

Pregunta 2. Altos vs. bajos

Definimos una variable dicotómica `alto` que toma valor 1 si la altura del trabajador es mayor a 67 pulgadas y 0 en caso contrario. El **ingreso promedio** de los trabajadores **bajos** ($\leq$ 67 in) es **44 mil US$**, mientras que el de los **altos** ($>$ 67 in) es **50 mil US$**. En promedio, los más altos ganan más que los más bajos. La diferencia de medias es **5.5 mil US$** con IC 95 % [**4.7, 6.3**] y t = **13.59**, $p < 0.001$.

Cuadro 2: Tabla 2: Ingresos promedio por grupo de altura

	N	Ingreso anual (miles de US$)	Desviación estándar (miles de US$)
Bajos ($\leq$ 67 pulgadas)	10 114	44	27
Altos ($>$ 67 pulgadas)	7756	50	27
Total	17 870	47	27

Notas: Se define indicador `alto` como 1 si la altura $>$ 67 pulgadas y 0 en caso contrario. Ingreso anual expresado en miles de US$ y redondeado a unidades. Test de medias de Welch. Diferencia Altos − Bajos: 5,5 miles de US$, $IC_{95\,\%} = [4{,}7,\ 6{,}3]$, $t = 13{,}59$, $p < 0{,}001$.

Pregunta 3. Mecanismos causales

Hipótesis central. La estatura en la adultez resume condiciones del *desarrollo temprano* (nutrición, salud, estimulación). Mejores condiciones incrementan **habilidad cognitiva** y ciertos **rasgos no cognitivos** relevantes en el mercado laboral, lo que eleva la productividad y por ende los salarios.[1]
Canales plausibles

- *Capital humano cognitivo*: mejor nutrición y menor carga de enfermedades en la infancia aumentan estatura y capacidad cognitiva. Esta última se remunera en el mercado laboral.
- *Rasgos no cognitivos y estatus*: mayor estatura puede asociarse a liderazgo, confianza y negociación, que afectan ascensos y asignaciones.
- *Salud en la adultez*: mejor salud reduce ausentismo y eleva productividad.

[1] Ver Case y Paxson (Case & Paxson, 2008).

- *Selección ocupacional*: personas más altas pueden concentrarse en ocupaciones que premian fuerza o presencia física, con salarios distintos.

Implicación empírica. Si la estatura capta condiciones tempranas que elevan habilidad, al controlar por *proxies* de habilidad como *educación* y por *ocupación* el coeficiente de `height` debería **disminuir en magnitud**. Esto es coherente con la idea de que parte de la asociación altura–ingresos opera vía esos canales.
Modelo ilustrativo

$$\text{earnings}_i = \alpha + \beta\,\text{height}_i + \gamma\,\text{ability}_i + \delta\,\text{education}_i + \varepsilon_i,$$

con $\gamma > 0$ y $\text{Cov}(\text{height}, \text{ability}) > 0$. Omitir *ability* induce sesgo en MCO hacia un coeficiente de `height` más alejado de cero, por lo que añadir educación u otras medidas de capital humano debería atenuarlo.

Pregunta 4. Gráfico ingresos vs. altura

La Figura 1 muestra el diagrama de dispersión entre el **ingreso anual** (eje Y, expresado en **miles de US$**) y la **altura** (eje X, en pulgadas). Se incluye la *recta OLS* como guía visual de la tendencia promedio. Las observaciones se agrupan en **líneas horizontales** porque el ingreso está registrado en valores *discretos* o redondeados (por ejemplo, múltiplos de mil dólares). Muchas personas comparten exactamente el mismo monto y los puntos se apilan en esos niveles; para mejorar la visibilidad se añadió un leve *jitter* vertical.

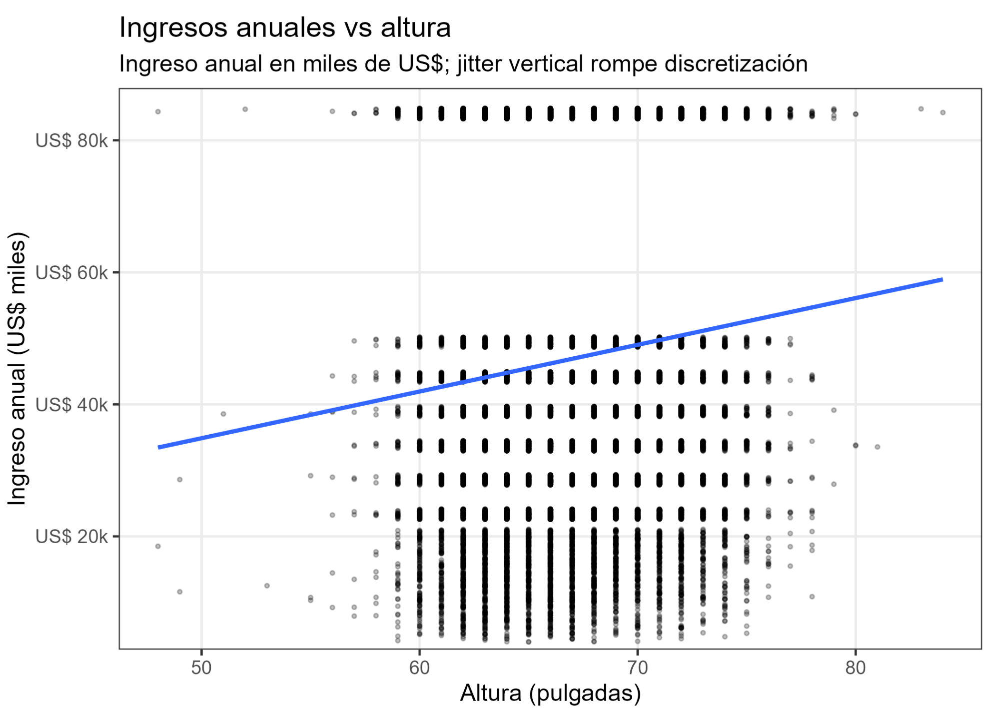

Figura 1: Ingresos anuales vs. altura

Notas: ingreso anual en miles de US$; puntos con *jitter* vertical para atenuar la discretización de los montos; línea ajustada por OLS.

Pregunta 5. Regresión con constante y predicciones

Estimamos por MCO la ecuación $earnings_i = \alpha + \beta\,height_i + u_i$. El coeficiente de *height* es **707.67** US$ por pulgada (EE robusto HC1 $= 50{,}40$, $p < 0{,}001$). Con esta especificación el poder explicativo es bajo pero típico en salarios con un solo regresor ($R^2 = 0{,}011,\ N = 17\,870$). Interpretación: una pulgada adicional de altura se asocia, en promedio, con $\approx$ **708 US$** más de ingreso anual, manteniendo la constante en el modelo.

Cuadro 3: Tabla 3: OLS de ingresos sobre altura

Variable	Coeficiente
Constante	-512.73 (3379.86)
Altura (pulgadas)	**707.67***** (50.40)

Notas: Variable dependiente: Ingreso anual (US$). Errores estándar robustos a la heterocedasticidad (HC1) entre paréntesis. $*\ p < 0{,}10$, $**\ p < 0{,}05$, $***\ p < 0{,}01$. $N = 17\,870$. $R^2 = 0{,}011$.

Cuadro 4: Tabla 4: Ingresos predichos por altura

Altura (pulgadas)	Ingreso predicho (US$)
65	45 486
67	46 901
70	49 024

Notas: Predicciones obtenidas a partir de la regresión OLS con constante de la Tabla 3.

Pregunta 6. Regresión sin constante

Estimamos la regresión de ingresos contra altura *sin* constante y comparamos con la especificación *con* constante. El coeficiente de *height* cambia de **0.71** US$ miles por pulgada (EE robusto HC1 = 0.05) en el modelo con constante a **0.70** en el modelo sin constante. La diferencia es mínima. El intercepto del modelo con constante no es estadísticamente distinto de cero ($p = 0{,}879$). Aunque el modelo sin constante puede acercarse visualmente a pasar por el origen, el R^2 que reporta es el *no centrado* y no es comparable con el R^2 usual; por convención y para interpretación económica, se prefiere mantener la constante.

Cuadro 5: Tabla 4: OLS ingresos vs. altura, con y sin constante

Variable	Con constante	Sin constante
Constante	-0.51 (3.38)	N/A
Altura (pulgadas)	**0.71***** (0.05)	**0.70***** (0.00)

Notas: Variable dependiente: Ingreso anual (US$ miles). Errores estándar robustos a la heterocedasticidad (HC1) entre paréntesis. $*\ p < 0{,}10$, $**\ p < 0{,}05$, $***\ p < 0{,}01$. $N = 17\,870$. R^2 (con constante) $= 0{,}011$; R^2 (sin constante, no centrado) $= 0{,}755$.

Discusión. Quitar la constante impone que el ingreso sea exactamente cero cuando la altura es cero, restricción sin sentido económico en este contexto. Además, altera la métrica del ajuste al usar R^2 no centrado. Dado que el intercepto no es significativo y la pendiente prácticamente no cambia, la especificación con constante es la elección adecuada para reportar e interpretar.

Pregunta 7. Cambio de unidades: pulgadas a centímetros

Definimos `height_cm` $= 2.54 \times$ `height` y estimamos por MCO:

$$earnings_i = \alpha + \beta\, height_i + u_i \quad \text{y} \quad earnings_i = \alpha + \tilde{\beta}\, height\_cm_i + u_i.$$

El resultado clave es que la **pendiente en centímetros** vale $\tilde{\beta} = \beta/2.54$; el **intercepto** y el $\mathbf{R^2}$ **no cambian**. Numéricamente: $\hat{\beta} =$ **707.67** US\$ por pulgada (EE = 50.40) y $\hat{\tilde{\beta}} =$ **278.61** US\$ por centímetro (EE = 19.84); $\hat{\alpha} = -512.73$ US\$ en ambos; $R^2 = 0.011$; $N = 17\,870$.[2]

Cuadro 6: Tabla 5: OLS de ingresos con altura en pulgadas vs. centímetros

	Modelo con pulgadas	Modelo con centímetros
Constante	−512,73 (3379,86)	−512,73 (3379,86)
Altura (pulgadas)	**707.67***** (50,40)	—
Altura (centímetros)	—	**278.61***** (19,84)
R^2	0.011	0.011
N	17 870	17 870

Notas: Variable dependiente: ingreso anual (US\$). Errores estándar robustos (HC1) entre paréntesis. $*\ p < 0{,}10$, $**\ p < 0{,}05$, $***\ p < 0{,}01$.

Interpretación y explicación de los cambios.

- *Pendiente*: al cambiar la unidad de medida de la variable explicativa, la pendiente se *reescala* en el mismo factor: 1 pulg = 2.54 cm implica $\hat{\tilde{\beta}} = \hat{\beta}/2.54$. El estadístico t y los p-valores no cambian porque el error estándar también se reescala por $1/2.54$.
- *Intercepto*: permanece igual. Solo reescalamos la variable explicativa, no la dependiente.
- $\mathbf{R^2}$, predicciones y residuos: son idénticos en ambos modelos. Cambiar unidades en un regresor es una transformación lineal que no altera los valores ajustados ni la varianza explicada.

Pregunta 8. Interpretación causal

No es razonable interpretar los coeficientes estimados como efectos causales con el modelo usado. La base es observacional y es muy difícil sostener la exogeneidad de la altura respecto de todos los determinantes del salario.

Supuestos mínimos para una interpretación causal en MCO En la regresión $earnings_i = \alpha + \beta\, height_i + u_i$ (o con controles X_i), se requiere:

1. **Exogeneidad condicional**: $E[u_i \mid height_i, X_i] = 0$. No deben quedar omitidos factores que afecten salarios y estén correlacionados con la altura.
2. **Muestreo aleatorio de la población de interés** y **SUTVA**. La altura de un individuo no debe afectar los salarios potenciales de otros y el “tratamiento” está bien definido.
3. **Medición adecuada**. Altura medida sin error sistemático; errores de medición relevantes en el regresor sesgan $\hat{\beta}$ hacia cero.
4. **Forma funcional correcta** y **apoyo común**. La relación lineal debe ser un buen resumen y debe existir variación de altura en todos los subgrupos relevantes.
5. **Ausencia de sesgo de selección por empleo**. Si el salario solo se observa para ocupados y la probabilidad de empleo depende de altura y de determinantes no observados del salario, aparece sesgo.

[2] Cálculos y tabla de coeficientes en los archivos de salida del proyecto.

¿Se cumplen aquí?

- **Variables omitidas**: la *habilidad cognitiva y no cognitiva*, el estatus socioeconómico del hogar, la nutrición y salud en la niñez y las redes laborales están correlacionadas con altura y con salarios. Esto viola la exogeneidad condicional.
- **Mediación**: altura puede afectar educación y ocupación, que a su vez afectan salarios. Sin un marco causal explícito, el coeficiente mezcla canales directos con efectos mediados.
- **Selección ocupacional y por empleo**: la altura puede influir el tipo de ocupación o la probabilidad de estar empleado en el sector formal; estimar solo en ocupados puede sesgar $\hat{\beta}$.
- **Otros aspectos**: la discretización de ingresos no induce causalidad; solo aumenta la varianza y motiva ajustes de visualización.

Conclusión Con la especificación utilizada, los coeficientes describen **asociaciones promedio** entre altura e ingresos, no efectos causales. Para avanzar hacia una interpretación causal se requeriría, por ejemplo, un diseño con variables instrumentales válidas, controles ricos que capturen condiciones tempranas y habilidad, o datos de panel con estrategias que eliminen confundentes no observados.

Preguntа 9. Diferencias por sexo en el efecto de la altura

Para testear si el efecto de la altura en salarios es igual para hombres y mujeres, estimamos:

$$\textit{earnings\_thou}_i \;=\; \alpha \;+\; \gamma\,\text{Mujer}_i \;+\; \beta\,\textit{height}_i \;+\; \delta\,[\textit{height}_i \times \text{Mujer}_i] \;+\; u_i,$$

donde *earnings_thou* es el ingreso anual en **miles de US$**. La hipótesis nula es $\text{H}_0 : \delta = 0$ (misma pendiente para ambos sexos). Rechazar H_0 indica diferencias en el efecto de la altura por sexo.

Cuadro 7: Tabla 6: Ingreso anual (US$ miles) y altura, con interacción por sexo

Variable	Coeficiente
Constante	$-43{,}13{*}{*}{*}$ (6,92)
Mujer	$55{,}78{*}{*}{*}$ (9,36)
Altura (pulgadas)	$1{,}31{*}{*}{*}$ (0,10)
Altura × Mujer	$-0{,}80{*}{*}{*}$ (0,14)
N	17 870
R^2	0.013

Notas: Errores estándar robustos (HC1) entre paréntesis. $*\ p < 0{,}10$, $**\ p < 0{,}05$, $***\ p < 0{,}01$.

Las pendientes implícitas por sexo (derivadas del modelo con interacción) son:

$$\hat{\beta}_{\text{Hombres}} = \beta = \mathbf{1{,}307}\text{ US\$ miles por pulgada (EE 0,099)}, \qquad \hat{\beta}_{\text{Mujeres}} = \beta + \delta = \mathbf{0{,}511}\text{ (EE 0,098)}.$$

Para la prueba formal, comparamos el modelo con interacción contra el mismo sin interacción ($\textit{earnings\_thou} \sim \textit{height} + \text{Mujer}$) mediante un test de Wald robusto (HC1): $\mathbf{F = 32{,}78}$, $p = \mathbf{1{,}05 \times 10^{-8}}$. Se **rechaza** H_0 al 1 % y concluimos que la pendiente de altura **difiere por sexo**. En esta muestra, el premio por altura es positivo y significativo en ambos grupos, pero **más grande en hombres**.

Interpretación. Un aumento de una pulgada en la altura se asocia, en promedio, con un incremento de aproximadamente \$1,307 para hombres y \$511 para mujeres en ingreso anual, manteniendo constante el resto de la especificación. Estos resultados describen *asociaciones*; no deben leerse como efectos causales sin supuestos adicionales.

Pregunta 10. Ocupación y fuerza

Estrategia. Para explorar si la relación altura–salario se explica (en parte) por selección hacia ocupaciones que premian fuerza/estatura, comparamos:

1. $earnings\_thou \sim height + \text{sex}$ (sin ocupar efectos fijos de ocupación), y
2. $earnings\_thou \sim height + \text{sex} + \text{occupation}$ (con *efectos fijos* de ocupación).

Nota: Incluimos un control por sexo para evitar sesgo por composición, dado que sexo está correlacionado con altura, ocupación e ingresos.

En ambos casos *earnings_thou* es el ingreso anual en *US\$ miles* y se usan EE robustos (HC1). Si al controlar por ocupación la pendiente de *height* cae en magnitud, sugiere que parte de la asociación opera vía composición ocupacional.

Cuadro 8: Tabla 7: Altura e ingresos controlando por ocupación

Modelo	Pendiente altura (US\$ miles/pulgada)	R^2
Sin ocupación (controla sexo)	**0.90***** (0.07)	0.012
Con efectos fijos de ocupación	**0.55***** (0.06)	0.159

Notas: EE robustos (HC1) entre paréntesis. $*\,p < 0{,}10$, $**\,p < 0{,}05$, $***\,p < 0{,}01$. $N = 17\,870$.
La pendiente de altura se reduce en **38.4 %** al añadir ocupación.

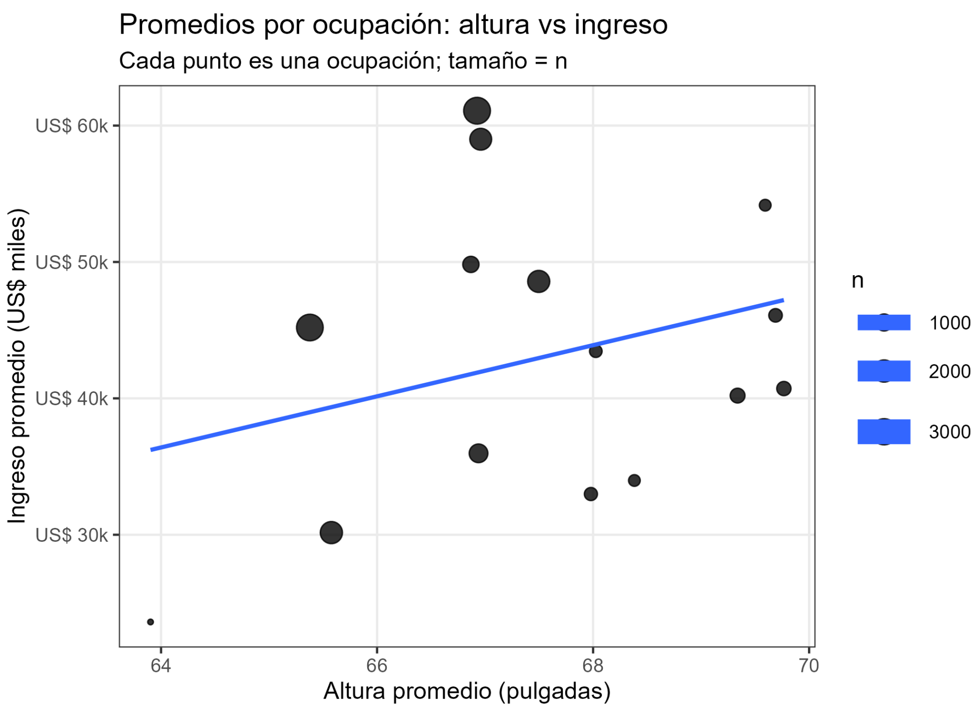

Figura 2: Promedios por ocupación: altura vs. ingreso

Cada punto es una ocupación; el tamaño representa el número de observaciones. Línea: ajuste OLS entre altura promedio y salario promedio por ocupación.

Conclusión. Al introducir efectos fijos de ocupación, la pendiente de altura cae de $\approx 0{,}90$ a $\approx 0{,}55$ mil US\$ por pulgada y el R^2 aumenta de 0.012 a 0.159. Esto respalda que **una parte** de la asociación altura–salario se explica por la **selección ocupacional**: ocupaciones con mayor remuneración concentran, en promedio, trabajadores más altos. Sin embargo, la pendiente sigue siendo positiva y significativa *dentro* de ocupación, por lo que la composición ocupacional **no agota** el vínculo.

Pregunta 11. Dirección del sesgo por variable omitida

Considere el modelo "verdadero"

$$earnings_i = \alpha + \beta\, height_i + \gamma\, ability_i + \varepsilon_i,$$

donde la *habilidad cognitiva* (*ability*) eleva los salarios ($\gamma > 0$). Si la estimación por MCO omite *ability*, el estimador cumple

$$\widehat{\beta}_{\text{MCO}} \;=\; \beta \;+\; \underbrace{\gamma\,\frac{\text{Cov}(height, ability)}{\text{Var}(height)}}_{\text{sesgo por omitida}}.$$

Bajo la hipótesis de Case y Paxson, peor nutrición temprana reduce tanto altura como desarrollo cognitivo, por lo que $\text{Cov}(height, ability) > 0$. Dado que además $\gamma > 0$, el sesgo es **positivo**.
Conclusión. La MCO *sobreestima* el efecto de la altura: si el efecto causal verdadero $\beta > 0$, entonces $\widehat{\beta}_{\text{MCO}} > \beta$; si $\beta < 0$, la estimación será menos negativa (más cercana a cero). Por ello, al controlar por un proxy de habilidad como la educación, el coeficiente de altura debería disminuir en magnitud.

Pregunta 12. Educación como proxy de habilidad (hombres)

(a) Especificaciones

Estimamos las siguientes regresiones por MCO, con errores estándar robustos (HC1):

1. (1) $earnings\_thou_i = \alpha + \beta\, height_i + u_i$
2. (2) $earnings\_thou_i = \alpha + \beta\, height_i + \theta_1 \text{LT\_HS}_i + \theta_2 \text{HS}_i + \theta_3 \text{Some\_Col}_i + u_i$

donde *earnings_thou* es ingreso anual en **US$ miles** y la categoría base es **College** (educ ≥ 16).

(b) Comparación del coeficiente de *height* e interpretación

El coeficiente de *height* cae de **1.31** a **0.74** US$ miles por pulgada al añadir dummies de educación, una reducción de **43 %**. Esto es coherente con la hipótesis de Case y Paxson: parte de la asociación altura–salario está mediada u omitida vía *habilidad/educación*, por lo que al controlar educación la pendiente se atenúa pero permanece positiva y significativa (interpretación como *efecto directo neto* de educación).

(c) Por qué se omite *College* en (2)

Con constante, incluir las cuatro dummies genera *trampa de las dummies* (multicolinealidad perfecta), pues $\text{College} = 1 - \text{LT\_HS} - \text{HS} - \text{Some\_Col}$. Se omite *College* y sus efectos quedan absorbidos en la constante; los demás coeficientes se leen como diferencias respecto a *College*.

(d) Test conjunto de educación

Probamos $H_0 : \theta_1 = \theta_2 = \theta_3 = 0$ mediante un test de **Wald robusto (HC1)** comparando (1) vs. (2).
Resultado: se **rechaza** H_0 con amplio margen; la educación es conjuntamente significativa y aporta sustancial poder explicativo.

Cuadro 9: Tabla 8: Hombres. OLS del ingreso anual con y sin indicadores de educación

Variable dependiente: Ingreso anual (US$ miles)		
	(1) Altura	(2) Altura + educación
Constante	$-43{,}13^{***}$	
	(6,93)	9,86
		(6,54)
Altura (pulgadas)	$1{,}31^{***}$	
	(0,10)	$0{,}74^{***}$
		(0,09)
Menos que secundaria (educ < 12)		$-31{,}40^{***}$
		(0,87)
Secundaria completa (educ = 12)		$-20{,}35^{***}$
		(0,70)
Terciaria incompleta (12 < educ < 16)		$-12{,}61^{***}$
		(0,80)
R^2	0.021	0.166
N	7 896	7 896
Efectos fijos	No	No

Notas: Errores estándar robustos (HC1) entre paréntesis. Significancia: $^{*}p < 0{,}10$, $^{**}p < 0{,}05$, $^{***}p < 0{,}01$. Categoría base para educación: College (educ $\geq$ 16).

Cuadro 10: Tabla 9: Test conjunto de dummies de educación (hombres)

Restringido	Irrestricto	Hipótesis	F	gl (restr.)	gl (resid.)	p	N
$earnings\_thou \sim height$	$earnings\_thou \sim height + \text{LT\_HS} + \text{HS} + \text{Some\_Col}$	$\theta_1 = \theta_2 = \theta_3 = 0$	500.92	3	7891	$5{,}64 \times 10^{-298}$	7896

Notas: R^2 pasa de 0.021 (rest.) a 0.166 (irrest.). EE robustos (HC1).

(e) Interpretación de los coeficientes de educación

En (2), los coeficientes de LT_HS, HS y Some_Col se interpretan como **diferencias de ingreso** respecto a *College*, **manteniendo constante la altura**. Valores negativos indican menores ingresos que el grupo base.

Cuadro 11: Tabla 10: Efectos de educación y niveles predichos a la altura media (hombres)

Variable dependiente: Ingreso anual (US$ miles)		
Categoría educativa	Efecto vs base (US$ miles)	Ingreso predicho a altura media (US$ miles)
Menos que secundaria (educ < 12)	-31^{***}	
	(1)	31
Secundaria completa (educ $= 12$)	-20^{***}	
	(1)	42
Terciaria incompleta ($12 <$ educ < 16)	-13^{***}	
	(1)	49
College o más (educ ≥ 16)	**base**	**62**

Notas: Errores estándar robustos (HC1) entre paréntesis bajo cada coeficiente. Significancia: $^{*}p < 0{,}10$, $^{**}p < 0{,}05$, $^{***}p < 0{,}01$. Predicciones evaluadas a la altura promedio de los hombres de la muestra.

Síntesis. La educación explica una fracción importante de la variación salarial y, al controlarla, el coeficiente de altura disminuye en magnitud pero sigue siendo positivo. Esto es consistente con la hipótesis de Case y Paxson de que condiciones tempranas mejores elevan altura y *habilidad* (proxied por educación), lo que se traduce en mayores salarios.

Referencias

Case, A., & Paxson, C. (2008). Stature and Status: Height, Ability, and Labor Market Outcomes. *Journal of Political Economy*, *116*(3), 499-532.