

Midiendo el desarrollo entre hombres y mujeres

*María Fernanda Puddy Erik Angel Andrés Bermúdez Daniel Sandoval
Gabriel A. Fuentes*

Abril, 2022

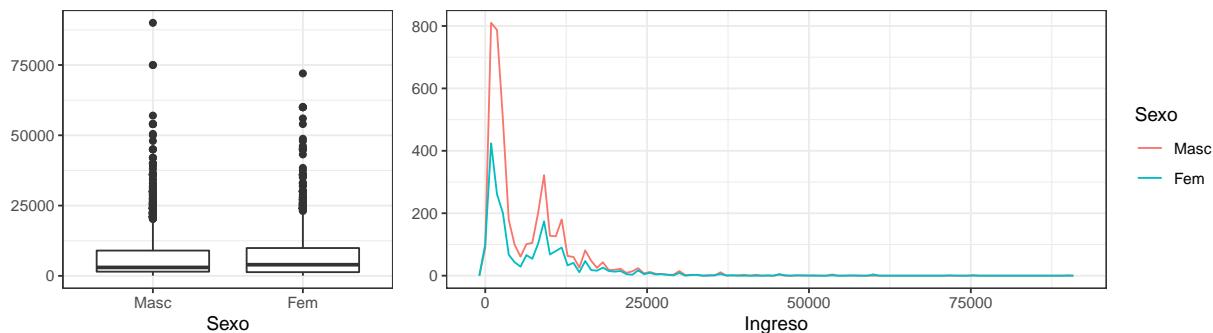
- a) Compara la media de ingreso entre hombres y mujeres y comenta acerca de las diferencias que encontra (presenta los valores y agrega como máximo dos párrafos explicativos del por qué consideras que existe la diferencia, utiliza soporte académico).

El presente trabajo define a sus sujetos de investigación como todos aquellos individuos que cuentan con, por lo menos, la edad legal mínima para trabajar, que permanecen ocupados en uno o más puestos de trabajo y que percibieron un salario durante el periodo previo a la realización de la encuesta. Además, el ingreso de las personas se define como la sumatoria de su sueldo base más los ingresos generados por horas laborales extras, comisiones, ingresos por trabajar durante algún periodo vacacional, bono 14, aguinaldo, bono vacacional, décimo quinto sueldo y bono de productividad, así como cobertura de viáticos por motivo de alimentación, vivienda y transporte.

Cuadro 1: Estadísticas descriptivas de ingresos entre hombres y mujeres

Sexo	Media	Mediana	Desv. Est.	Mínimo	Máximo	Asimetría	Curtosis	Obs.
Masc	6217.32	3000.0	6911.70	75	90000	2.84	16.01	4236
Fem	6919.04	3967.5	7689.67	70	72000	2.54	10.93	2086

Figura 1: Boxplot e histograma de frecuencias poligonales de ingresos entre hombres y mujeres



El cuadro 1 presenta las estadísticas descriptivas de esta variable cuando la muestra es segregada por sexo. Como es posible apreciar, el ingreso medio es más elevado para el género femenino que para el masculino por aproximadamente 700 quetzales. Sin embargo, la desviación estándar de esta medida también es más alta para el género femenino, sugiriendo que el mismo se distribuye con una mayor dispersión entre mujeres. Además es importante considerar que existen el doble de observaciones para los hombres que para la población femenina, por lo que este desbalance de clases podría estar sesgando a las estadísticas descriptivas.

De hecho, Gamboa (2021)—quien también utiliza los datos de la Encuesta Nacional de Empleo e Ingresos (2019)—encuentra que en realidad existe una marcada brecha salarial por género en el país que se explica debido a que la participación de la mayor parte de la fuerza laboral femenina se concentra dentro de ocupaciones caracterizadas por tener bajas y medias remuneraciones salariales.

- b) Compara ahora las medias y determina si estadísticamente son distintas (el hecho de que los valores absolutos sean distintos no necesariamente significa que estadísticamente lo sean). Presenta tu cálculo y la evidencia para afirmar tu respuesta.

Una prueba t de Welch es utilizada en este caso debido a su mayor robustez durante situaciones donde tanto los tamaños como las varianzas de ambas muestras difieren entre sí. El estadístico t de las medias muestrales se computa como:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{s_{\bar{X}_1}^2 - s_{\bar{X}_2}^2}} \quad (1)$$

donde \bar{X}_i y $s_{\bar{X}_i}$ representan la media y error estándar de la muestra i , respectivamente. El cuadro 2 presenta los resultados de esta prueba, los cuales confirman la significancia estadística en la diferencia de las medias de ambos grupos a un nivel de confianza del 99%.

Cuadro 2: Prueba t de dos muestras de Welch

Estadístico t	Valor p
-3.525169	0.0004282

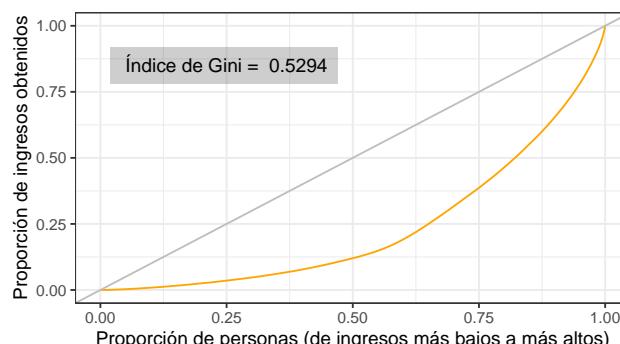
- c) Calcula el coeficiente de Gini de toda la muestra contenida en la base de datos y comenta cuál es tu apreciación del coeficiente de Gini encontrado (un párrafo como máximo).

El coeficiente de Gini, una medida que demuestra el grado de desigualdad en una distribución de ingresos/riqueza, se define como:

$$Gini = \frac{1}{2n^2\mu} \sum_{i=1}^n \sum_{j=1}^n |y_i - y_j|, \quad (2)$$

donde μ es el ingreso medio, n el número de individuos en la muestra, mientras que y_i y y_j representan el ingreso del individuo i y j , respectivamente. Debido a que valores elevados en el índice corresponden a un mayor nivel de desigualdad, es posible apreciar que la desigualdad en ingresos es considerablemente alta en la muestra general (53%).

Figura 2: Curva de Lorentz para la muestra general



- d) Calcula el coeficiente de Gini únicamente para la población masculina y para la femenina, compara los valores absolutos y determina si estadísticamente son distintos (analiza este punto con dos párrafos explicativos máximo).

Con el objeto de comparar el valor de los índices de Gini que existen al segregar a la base de datos entre mujeres y hombres y, particularmente, determinar si la diferencia entre estos coeficientes es estadísticamente significativa, se realiza un procedimiento de remuestreo aleatorio con remplazo (*bootstrapping*) en el que 1000 muestras conformadas por 1000 observaciones son extraídas de ambos grupos con el propósito de obtener el error estándar de cada Gini. Dado a que tanto el tamaño como la varianza de las muestras son iguales, una prueba *t* de Student es utilizada para comprobar la significancia estadística de la diferencia entre ambos coeficientes.

Cuadro 3: Bootstrapping del coeficiente de Gini por género

Grupo	Media	Desv. Est.	Mínimo	Máximo	Obs.		Valor
Hombres	0.5248	0.0097	0.4952	0.5550	1000	Estadístico t	-19.23295
Mujeres	0.5335	0.0106	0.5025	0.5701	1000	Valor p	0.00000

El cuadro 3 muestra los resultados de este procedimiento. Como es posible apreciar, el coeficiente de Gini es superior para mujeres (0.53) que para hombres (0.52), demostrando una ligera diferencia en términos de desigualdad de ingresos con respecto al sexo que favorece al género masculino. Además, los resultados de la prueba *t* confirman la significancia estadística en la diferencia de las medias de ambos grupos a un nivel de confianza del 99%.

- e) Calcula el anterior inciso (d) pero ahora segmentando a los hombres indígenas de los no indígenas y a las mujeres indígenas de las no indígenas. Comenta las diferencias de las medias de ingreso y de los Gini de los cuatro grupos segmentados.

De la misma forma que en el apartado anterior, se realiza un procedimiento de *bootstrapping* con el propósito de comprobar si las diferencias entre los coeficientes de Gini de hombres y mujeres segmentados por autodenominación étnica son estadísticamente significativas. El grupo “indígena” incluye a todas las personas que se autodenominan como mayas, xincas y garifunas, mientras que la categoría “no indígena” recoge a los individuos que se autodenominan como ladinos y extranjeros. Los resultados de este ejercicio se presentan en los cuadros 4 y 5.

El cuadro 4 demuestra que existe una distribución de ingresos menos equitativa dentro del grupo de personas que se autodenominan mayas, xincas o garífunas (0.55) que en la categoría “no indígena” (0.49). Además, es importante observar que el rango del índice de Gini para los hombres indígenas comienza en un mínimo mayor que el de los hombres no indígenas y también finaliza en un máximo más alto. Una prueba *t* de Student demuestra que, en efecto, ambos coeficientes pueden considerarse como estadísticamente distintos.

Cuadro 4: Bootstrapping del coeficiente de Gini entre hombres por autodenominación étnica

Grupo	Media	Desv. Est.	Mínimo	Máximo	Obs.		Valor
Indígena	0.5480	0.0107	0.5148	0.5815	1000	Estadístico t	133.053
No indígena	0.4865	0.0099	0.4558	0.5244	1000	Valor p	0.000

Mientras tanto, el cuadro 5 presenta los coeficientes de Gini para mujeres con la misma clasificación por autodenominación étnica. Los resultados de este ejercicio son muy similares a los del párrafo anterior: Las personas que pertenecen al grupo de mujeres no indígenas cuentan con una distribución de ingresos relativamente más equitativa que aquellas quienes pertenecen al grupo de mujeres indígenas. Sin embargo, de la misma manera que en el inciso (d), los valores absolutos de ambos grupos son considerablemente mayores a los de los mismos grupos del género masculino. La diferencia entre ambos coeficientes también es estadísticamente significativa a un elevado nivel de confianza.

Cuadro 5: Bootstrapping del coeficiente de Gini entre mujeres por autodenominación étnica

Grupo	Media	Desv. Est.	Mínimo	Máximo	Obs.		Valor
Indígena	0.6365	0.0095	0.6048	0.6661	1000	Estadístico t	311.3844
No indígena	0.4969	0.0105	0.4604	0.5296	1000	Valor p	0.0000

- f) Con base a la Teoría del Capital Humano (puedes revisar el estudio de Becker disponible en el portal y la presentación), arma una ecuación minceriana para evaluar los efectos de cada variable del Capital Humano en la muestra total.

La ecuación minceriana relaciona al logaritmo de los ingresos de un individuo con sus años de escolaridad y sus años de experiencia laboral, de tal modo que:

$$\ln w_i = f(s_i, x_i) = w_0 + \rho s_i + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i, \quad (3)$$

donde w representa el ingreso mensual de la persona i , s_i sus años de educación y x_i sus años de experiencia laboral potencial (edad menos años de educación menos seis). Una de las razones primordiales que explica la popularidad de esta metodología reside en que el parámetro ρ recoge la relación que existe entre los años de escolaridad y los ingresos de un individuo, por lo que este coeficiente es normalmente interpretado como la tasa de retorno de la educación—un concepto central de la teoría del capital humano que, tal y como Heckman et al. (2006) observan, fue introducido por primera vez por Becker (1964).

La especificación de la ecuación minceriana asume que la tasa de retorno de la educación es la misma para todos los niveles de escolaridad e independiente de los años de experiencia (ya que no se toma en cuenta la capacitación y educación que las personas reciben al estar empleadas). Mientras tanto, se espera una relación positiva entre la variable dependiente y las dos primeras variables: los años de educación y los años experiencia laboral. Por último, se espera una relación inversa entre el logaritmo de los ingresos y la experiencia elevada al cuadrado, debido a que cada año adicional de experiencia tiene un rendimiento marginal menor sobre los ingresos.

Cuadro 6: Función de ingresos de Mincer para la muestra completa

Coeficiente	Valor estimado	Error estándar	Estadístico t	Valor p
Intercepto	6.0979728	0.0596687	102.19711	< 2e-16
Educación	0.2022493	0.0068617	29.47531	< 2e-16
Experiencia	0.0905056	0.0037307	24.25997	< 2e-16
Experiencia acumulada	-0.0012363	0.0000596	-20.74159	< 2e-16

$$R^2 \text{ ajustado: } 0.1837375 \mid \text{Estadístico } F: 475.277557 (< 2.2e-16)$$

Una estimación de la ecuación minceriana por medio de mínimos cuadrados ordinarios en la muestra completa (cuadro 6) devela que los años de educación y experiencia tienen una relación positiva con el logaritmo del ingreso mensual de los individuos y que ambos coeficientes son estadísticamente significativos. La tasa de rentabilidad de la educación es del 20.22%, lo que indica que el incremento de un año de educación aumenta en dicha cuantía al logaritmo de los ingresos mensuales de los individuos. Sin embargo, esta especificación únicamente explica el 18% de la variación del logaritmo de los ingresos percibidos.

- g) Procede a evaluar el inciso f) realizando una ecuación para hombres y otra para mujeres, asegurándote de incluir en tu modelo variables extra (además de las variables propias del Capital Humano) que permitan evaluar el efecto de pertenecer al sector formal o informal; de estar en el área urbana o rural; y de ser indígena o no indígena (presenta tus cuadros de salida para ambas regresiones y analiza estadísticamente los resultados que obtuviste).

Los cuadros 7 y 8 presentan los resultados de regresiones mincerianas en las que la muestra general es segregada por sexo. Esta vez, tres variables dicotómicas son tomadas en consideración: una para diferenciar si el individuo reside en el área urbana o rural, otra para distinguir si este participa dentro del sector formal o informal y la última para considerar su autodenominación étnica, clasificada entre “indígena” y “no indígena”, como en apartados anteriores. En ambos modelos, todas las variables son estadísticamente significativas, tanto individualmente como en conjunto. Además, ambas regresiones cuentan con magnitudes similares en el coeficiente de determinación, sugiriendo que esta especificación explica en aproximadamente un 64% la variación en el logaritmo de los ingresos.

Cuadro 7: Función de ingresos de Mincer para hombres (controlando por área, sector y autodenominación étnica)

Coeficiente	Valor estimado	Error estándar	Estadístico t	Valor p
Intercepto	6.3506518	0.0479666	132.39739	< 2e-16
Educación	0.0741099	0.0055554	13.34027	< 2e-16
Experiencia	0.0439350	0.0029478	14.90440	< 2e-16
Experiencia acumulada	-0.0006122	0.0000465	-13.17523	< 2e-16
Área (urbana)	0.2878691	0.0223220	12.89620	< 2e-16
Sector (formal)	1.4509827	0.0229813	63.13748	< 2e-16
Autodenominación (no indígena)	0.2781599	0.0244000	11.40001	< 2e-16

$$R^2 \text{ ajustado: } 0.6366784 \mid \text{Estadístico } F: 1237.8901792 (< 2.2e-16)$$

Los resultados evidencian que la tasa de retorno de la educación es mayor para las mujeres que para los hombres, 10.2% y 7.4%, respectivamente. Esta diferencia entre las tasas de retorno puede deberse a que, en Guatemala, las mujeres tienen un menor acceso a educación que los hombres, por lo que un mayor nivel educativo para ellas representa actualmente una alta probabilidad de generar flujos de ingresos más elevados. Por otra parte, el coeficiente que mide la relación entre la experiencia laboral y el logaritmo de los ingresos tiene un peso ligeramente mayor en el grupo de hombres (4.4%) que en el de mujeres (4.1%).

En lo que respecta a las variables dicotómicas consideradas dentro del modelo, todas mantienen una relación positiva con los ingresos. Sin embargo, la variable dicotómica que divide a las muestras entre participantes del sector formal y aquellos individuos que trabajan en el sector informal presenta el coeficiente con mayor peso en las regresiones, principalmente en la regresión que considera a la población femenina que participa dentro de la fuerza laboral. Por lo tanto, es posible observar que aquellos individuos que residen en áreas rurales, participan en el sector informal y se autodenominan como maya, xinca o garífuna perciben, en promedio, salarios más bajos.

Cuadro 8: Función de ingresos de Mincer para mujeres (controlando por área, sector y autodenominación étnica)

Coeficiente	Valor estimado	Error estándar	Estadístico t	Valor p
Intercepto	6.0310946	0.0803504	75.059881	< 2e-16
Educación	0.1023565	0.0092319	11.087273	< 2e-16
Experiencia	0.0407729	0.0049369	8.258746	2.59e-16
Experiencia acumulada	-0.0005951	0.0000793	-7.504822	9.07e-14
Área (urbana)	0.2754603	0.0411450	6.694872	2.77e-11
Sector (formal)	1.6259108	0.0359074	45.280669	< 2e-16
Autodenominación (no indígena)	0.3585666	0.0437176	8.201878	4.10e-16

$$R^2 \text{ ajustado: } 0.6390504 \mid \text{Estadístico } F: 616.2383552 (< 2.2e-16)$$

h) Con base a los cálculos que realizaste en los incisos previos, utiliza dos párrafos por ítem para comentar lo siguiente:

— ¿Es Guatemala un país con alta desigualdad?

En cuanto a los resultados del coeficiente de Gini se obtuvo un coeficiente de 0.68 para toda la muestra. Y nuevamente considerando que este coeficiente toma un valor entre 0 y 1, en la cual el 0 representa la máxima equidad y el 1 la máxima desigualdad, se afirma que en el país se cuenta con una alta desigualdad en la distribución de la riqueza. Sin embargo, además de considerar solamente la magnitud del Gini, también sería importante para poder determinar si existe una alta desigualdad el poder comparar con el promedio mundial y el de la región a la cuál pertenece el país, que en este caso se consideran como América Latina y Centroamérica.

Para poder realizar esta comparación, se obtuvieron datos del Gini del año 2019, a partir del Banco Mundial (2020), los datos obtenidos y que son los que se consideran sobre este coeficiente son los siguientes: promedio mundial: 0.354, América Latina, 0.462 y Centroamérica 0.4575. Ahora bien, ya que se tienen los datos sobre los cuáles se realizará la comparación del Gini del país, se puede decir que, comparando el coeficiente de Gini de Guatemala con el promedio del mundo, América Latina y Centroamérica, Guatemala es un país con alta desigualdad. El Gini obtenido se encuentra por encima en 0.326 con el promedio mundial, en 0.218 con América Latina y en 0.22 con Centroamérica. Esta diferencia se considera significativa ya que Guatemala en 8 años, según los datos disponibles por el Banco Mundial (2020) solamente pudo reducir el Gini en 0.063 cosa que podría indicar la alta magnitud de la diferencia del Gini que presenta Guatemala junto con los demás países del mundo y su región.

— ¿Existen diferencias significativas por sexo en el mercado laboral?

En cuanto al mercado laboral si existe una diferencia significativa, principalmente entre las medias de ambos grupos, así como también en las tasas de retorno percibidas de la educación, experiencia y pertenencia a ciertos grupos a través de la función de Mincer y por último en la comparación de los Gini de ambos sexos. Primeramente, se observa que en las regresiones obtenidas a partir de la función de Mincer, los hombres perciben una menor tasa de retorno respecto a la educación, sin embargo, en cuanto a otros elementos como la pertenencia al sector formal, la pertenencia al área urbana y la autodenominación étnica no indígena, esta es mayor, demostrando que las personas que residen en áreas rurales, pertenecen al sector informal y se autodenominan como indígenas, perciben menor ingreso, pero esta es aún más evidente cuando los sujetos son mujeres.

Por otra parte, es importante mencionar que, tomando las medias como una medida para inicial para saber la diferencia entre los salarios de ambos sexos, se realizó una prueba de Welch y se obtuvo el rechazo de la hipótesis nula, demostrando que si existe una diferencia estadísticamente significativa entre las medias. Asimismo, comparando los datos obtenidos en el inciso d, sobre el cuál se evalúa la diferencia estadística entre el Gini de hombres y mujeres, se puede determinar que si existe una diferencia estadística, el índice de Gini obtenido para las mujeres es mayor y la diferencia entre las medias de ambos grupos es estadísticamente significativa. Afirmando que si existe una diferencia significativa en el mercado laboral, tanto por la comparación de la desigualdad por sexo a través del Gini como por los retornos percibidos en los ingresos por ciertas características.

— ¿Es más alta la desigualdad entre hombres o entre mujeres y por qué?

Ahora bien, para poder determinar si es más alta la desigualdad entre hombres o entre mujeres es importante considerar los incisos e y f, en los cuáles se obtiene un bootstrapping de los ginis por género, pero con distinción entre la autodenominación étnica. Estos podrían ayudar a determinar, según la magnitud de la media de los Gini y la diferencia estadística que existe entre ellos si hay una diferencia significativa entre los grupos. Considerando los resultados obtenidos, se obtuvo que la media del índice de Gini de los hombres indígenas es de 0.7255 y la de los no indígenas es de 0.6075, comparando ambas medias se obtiene que existe una diferencia significativa entre estas medias y que la diferencia entre las medias de ambos grupos es de 0.118. Ahora bien, las medias de los Gini de las mujeres indígenas son de 0.8340 y de 0.6737 para las mujeres no indígenas, realizando una prueba t nuevamente se obtuvo que si existe una diferencia significativa entre

ambas medias y realizando una diferencia entre ambas medias se puede determinar que existe una diferencia de 0.1603, siendo esta mayor a la que evidencian los hombres por autodenominación étnica. A partir de ello, se afirma que existe más alta desigualdad entre las mujeres que entre los hombres.

— ¿Consideras que todas las diferencias de ingreso que encontraste en tu análisis son producto de la discriminación?

— ¿Cuáles son tus sugerencias para reducir las brechas de ingreso entre hombres y mujeres y fortalecer el camino al desarrollo?

Referencias

- Björklund, A., & Kjellström, C. (2002). Estimating the return to investments in education: How useful is the standard Mincer equation? *Economics of Education Review*, 21(3), 195–210. [https://doi.org/10.1016/S0272-7757\(01\)00003-6](https://doi.org/10.1016/S0272-7757(01)00003-6)
- Giles, D. E. A. (2004). Calculating a Standard Error for the Gini Coefficient: Some Further Results*. *Oxford Bulletin of Economics and Statistics*, 66(3), 425–433. <https://doi.org/10.1111/j.1468-0084.2004.00086.x>
- Heckman, J. J., Lochner, L. J., & Todd, P. E. (2008). Earnings functions and rates of return. *Journal of human capital*, 2(1), 1-31.
- Hoover, G. & Yaya, M. (2010). *Racial/Ethnic Differences in Income Inequality across US Regions*. Retrieved April 23, 2022, from <https://journals.sagepub.com/doi/abs/10.1007/s12114-010-9057-z>
- Mirzaei, S., Borzadaran, G. R. M., Amini, M., & Jabbari, H. (2017). A comparative study of the Gini coefficient estimators based on the regression approach. *Communications for Statistical Applications and Methods*, 24(4), 339–351. <https://doi.org/10.5351/CSAM.2017.24.4.339>
- Gamboa, D. (2021). *Diferencia salarial entre hombres y mujeres: Un estudio de Guatemala desde la perspectiva de las ocupaciones en el 2019*. *Revista Académica ECO*, (24), 21-37.