

# Econometría Aplicada I

Taller

## *Instrucciones del Examen Final*

En el ámbito de la economía laboral, uno de los principales problemas estudiados es el análisis de la **discriminación salarial** no sólo con el fin de identificar su presencia, sino para cuantificar su magnitud e identificar sus orígenes que permitan implementar las políticas públicas adecuadas para corregirla.

Entenderemos el concepto de **discriminación salarial** cuando personas semejantes a otras de un cierto grupo en particular reciben un trato salarial diferente sólo por ser parte de cierto grupo, lo que los pone en desventaja con el resto. Generalmente este grupo corresponde al que es etiquetado como minoritario con respecto al resto de la población y puede asociarse a diversos atributos del individuo como: género, raza, condición étnica, religión, región, estado civil, estatus sindical, entre otros.

Es por ello que desde años atrás se han desarrollado diferentes **métodos econométricos** que permiten estimar y cuantificar la diferencia salarial entre grupos para conocer sus causas, haciendo énfasis en explicar su origen.

De manera independiente pero casi simultánea, Blinder (1973) y Oaxaca (1973) utilizaron las técnicas de la regresión lineal para analizar la descomposición salarial y cuantificar el origen de las brechas salariales entre grupos para identificar sus causas al rededor de la media.

El **objetivo** de su examen final es realizar un **análisis estadístico de regresión lineal** que permita entender y cuantificar la discriminación salarial que pueda existir en el mercado laboral de la entidad federativa que le fue asignada, utilizando los datos para el cuarto trimestre de 2019 de la Encuesta Nacional de Ocupación y Empleo (ENOE) del INEGI (para más información puede visitar el sitio

<https://www.inegi.org.mx/programas/enoe/15ymas/>).

Para ello deberá aplicar la técnica econométrica de la **descomposición de Blinder-Oaxaca** a la muestra de su entidad asignada. Esta descomposición asume que bajo ausencia de discriminación, el efecto de los atributos del individuo sobre su salario son los mismos entre grupos. La discriminación salarial se identifica cuando existe alguna diferencia significativa entre los efectos. Esta diferencia no sólo es captada mediante la inclusión de variables dicotómicas como comúnmente se realiza, los autores también incluyen **variaciones en la estimación de la pendiente** de los parámetros estimados.

De manera general, la metodología consiste en lo siguiente: suponga datos disponibles para el logaritmo del salario de los individuos de dos grupos, aquellos con alguna ventaja (denotados con un \* en superíndice) y los que muestran alguna desventaja (denotados con un \* en subíndice). Utilizando estos datos de logaritmo de salarios y de sus posibles atributos (educación, experiencia, entre otros) primero se estima por mínimos cuadrados ordinarios la regresión para ambos grupos:

$$\begin{aligned} \ln y^* &= X^* \beta^* + u^* \\ \ln y_* &= X_* \beta_* + u_* \end{aligned}$$

donde  $y$  es el salario del individuo,  $X$  la matriz de variables con los atributos determinantes del salario del individuo y  $u$  el término estocástico del modelo, para cada subgrupo.

Dado que  $Cov [\hat{\beta}^*, \hat{\beta}_*] = 0$ ,  $Var [\hat{\beta}^* - \hat{\beta}_*] = Var [\hat{\beta}^*] + Var [\hat{\beta}_*]$ . Además, como la recta de regresión lineal pasa siempre por las medias muestrales sabemos que:

$$\begin{aligned} \overline{\ln y^*} &= \overline{X^*} \hat{\beta}^* \\ \overline{\ln y_*} &= \overline{X_*} \hat{\beta}_* \end{aligned}$$

lo que quiere decir que:

$$\overline{\ln y^*} - \overline{\ln y_*} = \overline{X^*} \hat{\beta}^* - \overline{X_*} \hat{\beta}_*$$

Si definimos  $\Delta \hat{\beta} = \hat{\beta}^* - \hat{\beta}_*$ , entonces  $\hat{\beta}_* = \hat{\beta}^* - \Delta \hat{\beta}$ , al sustituir y reacomodar estos resultados se obtiene que:

$$\overline{\ln y^*} - \overline{\ln y_*} = \hat{\beta}^* (\overline{X^*} - \overline{X_*}) + \overline{X_*} \Delta \hat{\beta}$$

Esta última ecuación implica que la diferencia en el logaritmo de los salarios entre ambos grupos puede **descomponerse** en los efectos debidos a la diferencia entre sus atributos (primer término) y los efectos de la discriminación (segundo término). Hay que notar que aquí la diferencia promedio entre grupos está ponderada con el coeficiente estimado del grupo con ventaja, mientras que la diferencia en los coeficientes está ponderada con las características promedio del grupo con desventaja.

Una expresión alternativa puede derivarse haciendo la sustitución para  $\hat{\beta}^*$  en vez de  $\hat{\beta}_*$ , obteniendo:

$$\overline{\ln y^*} - \overline{\ln y_*} = \hat{\beta}_* (\overline{X^*} - \overline{X_*}) + \overline{X^*} \Delta \hat{\beta}$$

donde la ponderación ahora se encuentra al revés. La literatura estima ambas especificaciones de manera simultánea lo que permite una mejor caracterización de la discriminación salarial existente.

## a) Instrucciones Generales

1. Realice un **análisis estadístico aplicado** donde se ilustre el uso de varias de las herramientas del análisis de regresión lineal revisadas en el curso y su implementación en algún *software* estadístico.
2. Elija una sola entidad federativa del país y un tipo de discriminación salarial que desee analizar, así como las variables que considere más relevantes para su estimación. No se limite a analizar la clásica discriminación salarial por género, también puede explorar y analizar la discriminación salarial en varias otras dimensiones como: *sociodemográfica, geográfica, categoría de trabajo, etc.*
3. Para evitar la distorsión en el mercado laboral debido a la Covid-19, la base de datos a utilizar será la del último trimestre del 2019. El archivo se llama **Enoe4t19.csv**, disponible en el sitio **Canvas** del curso. La descripción de las variables aparece en el archivo **Descripción de Variables.pdf**. No es necesario utilizar otra fuente de datos.
4. Tome en cuenta que en la base de datos aparece la información a nivel nacional, por lo que lo primero que deberá hacer es filtrar la información para sólo trabajar con los datos de su entidad asignada. Así mismo, contemple que la base considera también menores de edad y observaciones incompletas o **NA**, que deberá omitir.
5. Por tratarse de una encuesta a partir de un muestreo probabilístico, la ENOE cuenta con un factor de expansión (variable **FAC**). Omita también el uso de este factor y utilice sólo de las observaciones de la muestra de la que dispone para su estado, sin su ponderación.
6. Aplique la técnica de la descomposición de Blinder-Oaxaca para cuantificar y verificar la presencia de la discriminación salarial elegida. No olvide incluir:
  - Resultados de la estimación
  - Representación gráfica de los resultados
  - Interpretación de todos sus resultados en el contexto del problema
7. No es necesario utilizar todas las variables de la base o limitarse sólo a las que a continuación enlisto. Cada quién puede elegir aquellas variables que considere más adecuadas de acuerdo al contexto de su problema. Las variables recomendadas son:
  - Como dependiente, puede utilizar: el logaritmo de **INGOCUP** (Ingreso mensual) o de **ING\_X\_HRS** (Promedio de ingreso por hora trabajada)
  - Como independientes, puede utilizar: **SEX** (Sexo), **EDA** (Edad), **CS\_P12** (Sabe leer y escribir), **CS\_P13\_1** (Nivel educativo), **CS\_P13\_2** (Años aprobados), **N\_HIJ** (Número de hijos), **E\_CON** (Estado conyugal), **POS\_OCU** (Posición en la ocupación), **RAMA** (Sector de actividad económica), **C\_OCU11C** (Condición de ocupación), **AMBITO1** (Tamaño de la unidad económica), **TIP\_CON** (Tipo de contrato), **ANIOS\_ESC** (Años de escolaridad), **HRSOCUP** (Horas trabajadas en la semana), **P14APOYOS** (Si recibe apoyos económicos).
8. Pueden consultar y hacer uso de cualquier **material complementario** compartido como referencia para su realización o bien cualquier otro material adicional que encuentren de utilidad para ello. Tanto para **R** como para **STATA** ya existen rutinas

elaboradas con las funciones que les permiten hacer esta estimación. Elija trabajar en el *software* que considere más conveniente para cada uno y utilice los paquetes ya disponibles. El material complementario que explica la implementación de cada paquete está disponible en el sitio Canvas del curso.

9. Escriba una **conclusión** del análisis del mercado laboral en el contexto del problema.
10. El documento deberá ser remitido en el vínculo habilitado para ello en el sitio Canvas del curso. Será necesario entregar el documento PDF con sus resultados y el *script* en R o R-Markdown que lo genera o el *do file* de STATA que hayan utilizado para realizar sus estimaciones. La ponderación que este examen tiene sobre la calificación final es de 30 %