



“DISCRIMINACIÓN POR SEXO EN EL MERCADO LABORAL”

Región Huánuco

Trabajo Grupal

Curso: Fundamentos de Econometría

Profesor: Luis García Nuñez

Fecha de entrega: 17 julio 2023

Integrantes:

- Del Carpio Cuenca, Gabriel Sebastián (20191565)
- Ayquipa Ku, Steven Alberto (20185685)
- Enrique Alonso Serrano Salas (20156790)

2023

1. Introducción:

A lo largo de las últimas décadas para los economistas ha sido de gran interés poder estudiar y comprender cómo los ingresos monetarios, salario, era afectado por variables como los años de escolaridad, la edad de la persona, y su sexo.

Uno de los modelos más conocidos para poder encontrar esta relación ha sido el modelo de Mincer, el cual establece que el logaritmo natural (tasa de crecimiento) del salario depende de los años de educación y la experiencia laboral: $\ln(W) = \beta_0 + \beta_1 \text{Edu} + \beta_2 \text{Exp} + \varepsilon$, siendo el último término el error. Otro modelo es el modelo de Heckman, el cual establece una relación similar a la anterior, con la diferencia de que establece una probabilidad de que el individuo pueda conseguir empleo, corrigiendo de esta forma el sesgo de selección en los datos, el modelo se puede mostrar bajo la siguiente expresión $\ln(W) = \beta_0 + \beta_1 \text{Edu} + \beta_2 \text{Exp} + \gamma Z + \varepsilon$, donde Z, representa las variables que afectan al individuo y su posibilidades de conseguir empleo. (CEPAL, 1990), de esta forma es posible ver los esfuerzos de los economistas por conseguir estimar de forma fehaciente un modelo que pueda recoger todas las posibilidades relevantes para entender aquello que afecta y modifica los ingresos salariales.

Para el presente trabajo, se hace uso de una variante a la ecuación de Mincer, $\ln(W_i) = \beta_1 + \beta_2 \text{Educi} + \beta_3 \text{Edadi} + \beta_4 \text{Edadi}^2 + \beta_5 \text{Sexoi} + \beta_6 \text{Sexo} \times \text{Educi} + \text{otras variables} + u_i$, en esta ecuación podemos encontrar a diferencia de la ecuación original la variable cuadrática de educación lo cual indica una forma de encontrar la relación no lineal entre la educación y salario, ya que el salario puede incrementarse al comienzo, pero disminuir al momento de la jubilación, ello será corroborado en el modelo con los datos. Además se agrega una variable dummy de Sexo, la cual ayuda a identificar si existe una brecha de género entre los ingresos masculinos y femeninos, finalmente “otras variables”, las cuales pueden incluir características relevantes como raza, ubicación, etc.

En relación a lo planteado, el objetivo principal es encontrar si existe evidencia de discriminación por sexo en el mercado laboral, la hipótesis se podrá corroborar haciendo la estimación de β_5 y β_6 , usando los datos obtenidos de la Región de Huánuco para poder utilizar el modelo de Mincer modificado. Para ello se hará uso de la base de datos obtenida por la Encuesta Nacional de Hogares (ENAHOG), del año 2022, usando el software especializado de Stata, el análisis será delimitado por un rango de edad de 24 a 65 años, siendo el foco de estudio las personas trabajadoras dependientes, en el apartado de otras variables, se hará uso de las siguientes categorías: “tamaño de la empresa, ocupación, tipo de escuela, discriminación y lengua”.

2. Marco Teórico

Históricamente, ha existido un claro sesgo laboral al momento de comparar las condiciones laborales entre hombres y mujeres, lo cual responde a una estructura social inherente en la mayor parte de las sociedades occidentales, de esta forma esta disparidad engloba dentro del ámbito laboral, todos los aspectos, así lo define Escot: “la literatura especializada en el análisis de la desigualdad entre mujeres y hombres en el mercado de trabajo se ha centrado en el análisis de cuatro tipos de desigualdad: desigualdad en la participación (tasas de actividad masculinas y femeninas); desigualdad (segregación) ocupacional y sectorial; desigualdad salarial; y desigualdad en la promoción profesional” (Escot, 2014).

Bajo este enfoque podemos relacionar diversas variables, por ejemplo, la desigualdad en la participación hace referencia a la tasa de contratación, en el modelo planteado para el actual trabajo, es posible generará este enlace a través de “otras variables”, la segregación ocupacional hace referencia a qué tipos de ocupación suelen tener las mujeres, el modelo puede ayudarnos a entender ello de forma indirecta, finalmente, la desigualdad salarial, el cual es el eje central del modelo de Mincer modificado.

Sin embargo, es muy simplista asumir que únicamente por el hecho de ser mujer u hombre una persona obtendrá un mejor salario. Para ello, es necesario ver “...el problema desde un enfoque multidimensional y multicausal; es decir, a pesar de la variedad de factores que podrían favorecer la persistencia de la brecha salarial, estos se intensifican estableciendo una relación de círculo vicioso. De acuerdo a la American Association of University Women (2018), los niveles de educación, etnicidad, edad, discapacidad, orientación sexual e identidad de género son factores importantes que afectan la brecha salarial en las mujeres.” (Women, 2018). Por lo que podemos inferir que es posible generar una estimación con mayor precisión a partir de añadir variables como las propuestas anteriormente, de esta forma englobar las distintas posibilidades y los efectos de cada una de ellas sobre el salario.

Finalmente, de acuerdo a la literatura, se puede contrastar que sí existen barreras significativas de acuerdo al sexo, para el acceso a empleos de calidad altamente correlacionados con el nivel de estudios, “Respecto al acceso a empleos de calidad, las mujeres tienen menores probabilidades de acceso que los hombres... Factores que contrarrestan estos efectos negativos están en manos de las mujeres: estar en el tercio superior durante sus estudios superiores, poder titularse, contar con habilidades TIC o la misma selección de carrera.” (Barrantes, 2019). Podemos plantear a partir de la revisión de literatura la hipótesis que sí existe discriminación por sexo en el mercado laboral, lo cual será corroborado a partir de los datos de ENAHO, para la región Huánuco en la siguiente sección.

3. Análisis de datos:

Las fuentes de información que se usaron para la presente investigación fueron en base a la Encuesta Nacional de Hogares (también conocida como ENAHO), y que fue elaborada por el INEI. En esta se emplearon las bases de datos de: las características de los miembros del Hogar, Educación y Empleo e Ingresos. En particular se emplearon las bases: *enaho01a-2022-200*, *enaho01a-2022-300* y *enaho01a-2022-500*.

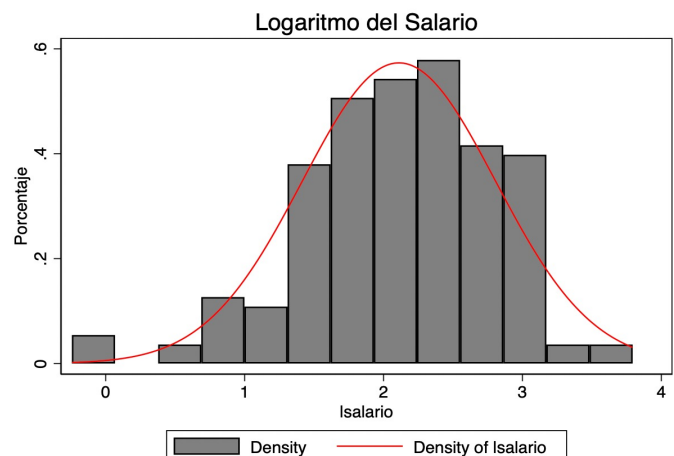
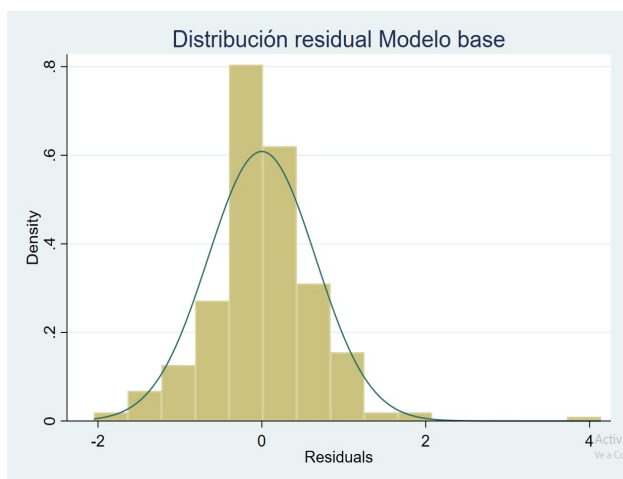
Las bases de datos fueron trabajadas en el programa de STATA, aplicando cambios según las variables de interés del presente estudio. Después, se filtró solo a la población de Huánuco, quedando como total 178 observaciones, lo cual, junto al diseño muestral, se ve modificado según el factor de expansión; sin embargo, se optó por no analizar el factor a fin de mayor énfasis en otros puntos. Es importante tomar en cuenta que solo se trató variables que cumplan con los requisitos de trabajador dependiente y entre el intervalo de edad de 24 a 65 años, cuyos ingresos anuales deflactados solo sean de su ocupación principal.

La muestra total consta de un total de 36348 viviendas particulares, repartidas en 24256 al área urbana y 12952 en el área rural. En Huánuco la muestra de conglomerados es 183 y 1324 en viviendas. En tal sentido, se hizo un análisis mediante el método de Mínimos Cuadrados Ordinarios (MCO). Se seleccionaron las variables relevantes para la muestra y se combinaron en una sola base de datos que contenga a todas ellas. Por lo tanto, se trató diferentes variables posiblemente relevantes como la actividad económica de la empresa en la operan los individuos, su categoría ocupacional, el tamaño de empresa, la lengua materna, el estado civil, entre otras variables. Para ello se crearon dummies interactivas con la variable sexo para captar el impacto del género de la persona en su salario, para luego hallar posteriormente su coeficiente y la significancia de dicho coeficiente. Las dummies están resumidas de la siguiente forma:

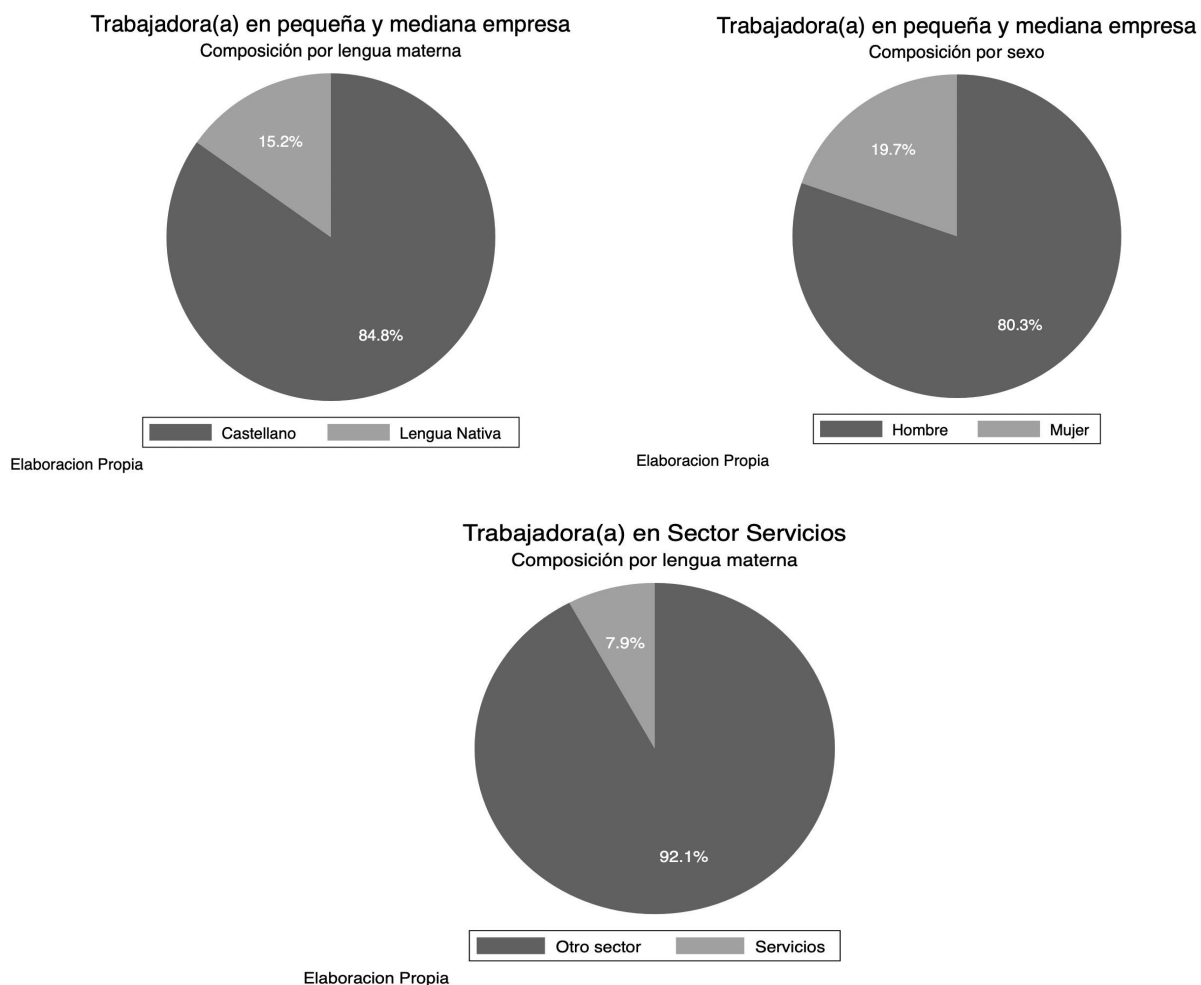
- ocupacion: (1, trabajador del hogar; 0, empleado u obrero)
- actividad1: (1, sector agropecuario, pesca y extractivo; 0, c.o.)
- actividad2: (1, sector manufactura, construcción, comercio, transporte y telecomunicaciones; 0, c.o.)
- actividad3: (1, sector servicios u otros; 0, c.o.)
- tamaño: (1, pequeña y mediana empresa; 0, gran empresa)
- lengmat: (1, lengua nativa; 0, castellano)
- est1_casado: (1, casado(a); 0, c.o.)
- est1_soltero: (1, soltero(a); 0, c.o.)
- est1_otro: (1, viudo(a) o divorciado(a); 0, c.o.)
- lsalario: logaritmo natural de la variable salario
- sxact1: (1, mujer trabajadora en actividad1; 0, c.o.)
- sxact2: (1, mujer trabajadora en actividad2; 0, c.o.)
- sxact3: (1, mujer trabajadora en actividad3; 0, c.o.)
- sxtam: (1, mujer trabajadora en pequeña y mediana empresa; 0, c.o.)

- sxleng: (1, mujer con lengua nativa; 0, c.o.)
- sxocup: (1, mujer trabajadora del hogar u otro; 0, c.o.)
- sxmarry: (1, mujer casada; 0, c.o.)
- sxsing: (1, mujer soltera; 0, c.o.)
- lxeduc: (1, lengua nativa y educación; 0, castellano y educación)
- lxxact1: (1, lengua nativa con actividad1; 0, c.o.)
- lxxact3: (1, lengua nativa con actividad3; 0, c.o.)
- latam: (1, trabajador en pequeña y mediana empresa con lengua nativa; 0, trabajador en pequeña y mediana empresa con lengua castellano)
- lxocup: (1, trabajador del hogar con lengua nativa; 0, c.o.)
- lxmarry (1, persona con lengua nativa casada(o); 0, c.o.)
- lxsing (1, persona con lengua nativa soltera(o); 0, c.o.)

Cabe señalar que se eliminó una categoría de cada dummy para evitar problemas de multicolinealidad, por lo que se toma como base a actividad2, sxact2, es1_otro, lengmat castellano, ocupacion = empleado u obrero, tamaño = gran empresa, entre otras bases.



Analizando la variable dependiente, la normalidad de los residuos se cumple, así como la normalidad de la distribución del logaritmo del salario, por lo que los test de White, Breusch-Pagan y White robusto, entre otros, lo que ayuda a realizar pruebas de hipótesis conjunta a pesar de desconocer las distribuciones de las variables a tratar.



Es importante considerar las anteriores imágenes de pie sobre algunos regresores dummies en la muestra, especialmente por la relevancia que tendrá este hecho al momento de las estimaciones econométricas.

sxtam	mean(lsala~o)	med(lsala~o)	sd(lsalarío)	min(lsala~o)	max(lsala~o)
Hombre	2.27	2.29	0.58	-0.06	3.79
Mujer	1.45	1.44	0.75	-0.24	2.80

lxtam	mean(lsala~o)	med(lsala~o)	sd(lsalarío)	min(lsala~o)	max(lsala~o)
Castellano	2.17	2.19	0.66	-0.20	3.79
Lengua Nativa	1.77	1.83	0.80	-0.24	2.80

Aquí se observa más detalladamente tablas de doble entrada según el logaritmo del salario y diversas dummies de interés. Por ejemplo, se observa en primera instancia que dependiendo del sector económico y el tamaño de la empresa, considerando el sexo y la lengua materna

como punto de partida, se presenta mayorsalario con menor variabilidad en los hombres que hablen castellano y que trabajen en otro sector que no sea servicios. Estos datos se probarán en la siguiente sección.

lxact3	mean(lsal~o)	med(lsala~o)	sd(lsalario)	min(lsala~o)	max(lsala~o)
Otro sector	2.13	2.16	0.68	-0.24	3.79
Servicios	1.89	2.03	0.86	-0.06	2.99

Para el análisis de la discriminación por sexo en el mercado laboral partiremos a partir de cinco modelos, incluyendo poco a poco posibles variables relevantes. Pasaremos del modelo modelo base (restringido), al modelo 1 (inclusión de la actividad económica), modelo 2 (inclusión del tamaño de la empresa), modelo 3 (inclusión de la lengua materna), modelo 4 (inclusión de la categoría ocupacional) y modelo 5 (inclusión del estado civil). Posteriormente, se eliminan variables que no mejoren la estimación, y se compara finalmente con el modelo base, a fin de encontrar conclusiones sobre el signo y significancia de los coeficientes β_5 y β_6 , correspondientes a la variable sexo y dummy interactiva de sexo con educación.

Adicionalmente, se trabajará un posible modelo de discriminación por lengua materna, para el cual se pasará del modelo base (irrestringido) al modelo 1 (inclusión de actividad económica), al modelo 2 (inclusión del tamaño de la empresa) y al modelo 3 (inclusión del estado civil). La dummy de sexo con lengua materna será hallada y evaluada para su tratamiento e inclusión en este modelo posteriormente, y con respecto a la inclusión de la categoría ocupacional, no se opta por esta dummy puesto que la base de datos no tiene información sobre los individuos con lenguaje nativo y ocupación de trabajadores del hogar u otros, por lo tanto es una variable irrelevante en la estimación.

4. Estimaciones Econométricas

DISCRIMINACIÓN POR SEXO EN EL MERCADO LABORAL

Modelo base: *reg lsalario educ edad edad2 sexo sxeduc*

Este es el modelo base el cuál se regresa la ecuación del Mincer donde las variables principales son la educación, la edad, la edad al cuadrado, el sexo y la dummy “sexo x educación”. En esta se analizan 178 observaciones y un R cuadrado del 0.3378. En esta vemos que la distribución residual alcanza su pico máximo cuando el residuo es 0.

Modelo 1: *reg lsalario educ edad edad2 sexo sxeduc sxact1 sxact3*

En este modelo incluimos la actividad económica, y generamos la variable *sxact1* y *sxact3* las cuáles representan el sexo de cada persona multiplicado por las actividades 1 y 3 que representan a los sectores “Agropecuaria, pesca y extractivo” y “Servicios y diversos”. Es por ello que se regresa el modelo base junto a esas nuevas variables. Además, hay indicios de multicolinealidad pues los VIF individuales son mayores a 10 y el promedio es mayor a 1. Por otro lado, podemos decir que hay heterocedasticidad pues los valores que nos arrojan el test de White y B-P son 0.0476 y 0.442.

Modelo 2: *reg lsalario educ edad edad2 sexo sxeduc sxact1 sxact3 sxtam*

En este modelo se incluye el tamaño de la empresa (se toma como base a gran empresa), por lo que generamos una variable llamada *sxtam* que representa el sexo multiplicado tamaño. Hay indicios de multicolinealidad pues los VIF individuales son mayores a 10 y el promedio es mayor a 1. Además, a partir de los resultados no se muestra que haya heterocedasticidad pues los resultados que arrojan son menores a 0.05.

Modelo 3: *reg lsalario educ edad edad2 sexo sxeduc sxact1 sxact3 sxtam sxleng*

En este caso incluimos la lengua materna y tomamos como base al castellano. Generamos la variable *sxleng* que representa el sexo por la lengua materna. Hay indicios de multicolinealidad pues los VIF individuales son mayores a 10 y el promedio es mayor a 1. Por otra parte, no hay heterocedasticidad pues los valores son mayores a 0.05. Vemos que el la lengua materna no aporta a la estimación ya que el R se mantiene y el R cuadrado disminuye.

Modelo 4: *reg lsalario educ edad edad2 sexo sxeduc sxact1 sxact3 sxtam sxleng sxocup*

Para este modelo incluimos la categoría ocupacional y tomamos como base a “empleado”. Se genera la variable *sexocup* que representa el sexo por ocupación. Hay indicios de

multicolinealidad pues los VIF individuales son mayores a 10 y el promedio es mayor a 1. Por otra parte, no hay heterocedasticidad pues los valores son mayores a 0.05. No hay certeza que sea relevante esta variable pues al ver demasiadas variables disminuye su peso.

Modelo 5: *reg lsalario educ edad edad2 sexo sxeduc sxact1 sxact3 sxtam sxleng sxocup sxmarry sxsing*

Por último generamos la variable *sxmarry* y *sxsing* que representan la inclusión del estado civil y que se multiplican la variable *sexo* por su estado civil (soltero y casado). Hay indicios de multicolinealidad pues los VIF individuales son mayores a 10 y el promedio es mayor a 1. Por otra parte, no hay heterocedasticidad pues los valores son mayores a 0.05.

VARIABLES	Modelo base	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5
<i>educ</i>	0.0698*** (0.0160)	0.0693*** (0.0160)	0.0690*** (0.0148)	0.0690*** (0.0148)	0.0692*** (0.0148)	0.0692*** (0.0149)
<i>edad</i>	0.0576* (0.0301)	0.0652** (0.0304)	0.0438 (0.0283)	0.0431 (0.0285)	0.0401 (0.0286)	0.0371 (0.0304)
<i>edad2</i>	-0.000511 (0.000347)	-0.000611* (0.000350)	-0.000401 (0.000326)	-0.000392 (0.000328)	-0.000353 (0.000330)	-0.000323 (0.000347)
<i>sexo</i>	-0.444 (0.281)	-0.369 (0.345)	0.158 (0.333)	0.149 (0.336)	0.172 (0.336)	0.161 (0.367)
<i>sxeduc</i>	0.0236 (0.0206)	0.0147 (0.0212)	-0.0135 (0.0203)	-0.0154 (0.0212)	-0.0163 (0.0212)	-0.0151 (0.0217)
<i>sxact1</i>		-0.177 (0.228)	-0.148 (0.211)	-0.130 (0.219)	-0.146 (0.219)	-0.148 (0.222)
<i>sxact3</i>		0.121 (0.189)	0.264 (0.177)	0.272 (0.179)	0.277 (0.179)	0.282 (0.182)
<i>sxtam</i>			-0.693*** (0.127)	-0.692*** (0.127)	-0.674*** (0.128)	-0.663*** (0.136)
<i>sxleng</i>				0.0439 (0.134)	0.0253 (0.135)	0.0241 (0.136)
<i>sxocup</i>					-0.420 (0.398)	-0.413 (0.401)
<i>sxmarry</i>						0.000330 (0.146)
<i>sxsing</i>						-0.0500 (0.195)
Constant	-0.130 (0.649)	-0.259 (0.651)	0.238 (0.609)	0.253 (0.612)	0.305 (0.614)	0.371 (0.655)
Observations	178	178	178	178	178	178
R-squared	0.338	0.351	0.449	0.449	0.453	0.453
R-adjusted	0.319	0.325	0.423	0.420	0.420	0.413
Standard errors in parentheses						
*** p<0.01, ** p<0.05, * p<0.1						

Se observa como aumenta el R^2 y el R^2 ajustado conforme se van agregando variables, pero es a partir del modelo 3 que se pierde significancia en el modelo, posteriormente se

estanca en el modelo 4 y el R^2 ajustado se reduce para el modelo 5, lo que nos indicaría inclusión de variables irrelevantes. Así mismo son los coeficientes de las variables educ y sxtam las que tienen mayor significancia, y sobre los coeficientes β_5 y β_6 , se muestra que, conforme se agregan más regresores (algunos más significativos) la variable sexo no explica movimientos en el salario, y de hecho cambia de ser negativa a tener signo positivo, pero igualmente poco significativa. De manera similar ocurre con la dummy sxeduc, la cual cambia de positivo a negativo mientras se agregan regresores, pero permaneciendo su irrelevancia en la estimación.

Construyendo un nuevo modelo aproximado al modelo más relevante en base a los anteriores resultados, se obtiene la comparación entre este nuevo modelo irrestricto contra el restringido:

VARIABLES	1 Modelo base	2 Modelo aprox
educ	0.0698*** (0.0160)	0.0692*** (0.0148)
edad	0.0576* (0.0301)	0.0404 (0.0285)
edad2	-0.000511 (0.000347)	-0.000358 (0.000328)
sexo	-0.444 (0.281)	0.178 (0.334)
sxeduc	0.0236 (0.0206)	-0.0153 (0.0204)
sxact1		-0.157 (0.211)
sxact3		0.272 (0.177)
sxtam		-0.674*** (0.128)
sxocup		-0.430 (0.393)
Constant	-0.130 (0.649)	0.297 (0.611)
Observations	178	178
R-squared	0.338	0.453
R-adjust	0.319	0.423
Standard errors in parentheses		
*** p<0.01, ** p<0.05, * p<0.1		

lsalario	Coef.	Std. Err.	t	P> t
educ	.0692015	.0147642	4.69	0.000
edad	.040422	.0284897	1.42	0.158
edad2	-.0003576	.0003281	-1.09	0.277
sexo	.1781913	.3337889	0.53	0.594
sxeduc	-.015271	.0203656	-0.75	0.454
sxact1	-.1568318	.2109889	-0.74	0.458
sxact3	.2721041	.1766097	1.54	0.125
sxtam	-.6737465	.1279116	-5.27	0.000
sxocup	-.429799	.3931114	-1.09	0.276
_cons	.2973956	.6107746	0.49	0.627

lsalario	Coef.	Robust Std. Err.	t	P> t
educ	.0692015	.0172511	4.01	0.000
edad	.040422	.0255713	1.58	0.116
edad2	-.0003576	.0002885	-1.24	0.217
sexo	.1781913	.3754597	0.47	0.636
sxeduc	-.015271	.0210066	-0.73	0.468
sxact1	-.1568318	.2586376	-0.61	0.545
sxact3	.2721041	.228532	1.19	0.235
sxtam	-.6737465	.132731	-5.08	0.000
sxocup	-.429799	.3120139	-1.38	0.170
_cons	.2973956	.5954489	0.50	0.618

Si realizamos el test robusto de White, se observa que aún si usamos las variables más relevantes que son educ y sxtam, haciendo el test de multicolinealidad observamos F en la tabla, y parece no haber multicolinealidad. Es probable que para el caso de Huánuco, ni el hecho de ser mujer ni la cantidad de años de estudio de las mujeres ni la actividad económica a la que su empleador trabaje, ni su vejez ni qué puesto ocupa en su trabajo afectan

significativamente en su salario. Esto quiere decir que los principales determinantes del salario en Huánuco son si se trabaja en pequeñas y medianas empresas y sus años de educación en general.

Detección de multicolinealidad del modelo irrestricto: 5 modelos

Se evidencia que en los 5 modelos planteados, los coeficientes son significativos por lo menos a un 10% y el ajuste del modelo a la variable endógena se incrementa ligeramente al agregar variables explicativas. Se detecta multicolinealidad ya que el VIF promedio es mayor a 10.

DISCRIMINACIÓN POR LENGUA MATERNA EN EL MERCADO LABORAL

Modelo base: *reg lsalario educ edad edad2 lengmat lxeduc*

En este modelo buscará hallar los coeficientes de lengmat y lxeduc a partir de la base de datos anteriormente delimitada.

Modelo 1: *reg lsalario educ edad edad2 lengmat lxeduc lxact1 lxact3*

Test de Breush Pagan arroja 33.52 con p value cercano a cero, y test de White arroja 35.15, lo que indicaría que debemos rechazar la hipótesis nula de homocedasticidad, ello con 1% de significancia y 5% de significancia respectivamente. El test de vif da como resultado 24.16 de media, habiendo clara presencia de multicolinealidad.

Modelo 2: *reg lsalario educ edad edad2 lengmat lxeduc lxact1 lxact3 lxtam*

Al agregar lxtam la correlación con la variable dependiente aumenta, pues el R^2 y R^2 ajustado aumentan simultáneamente, y algunos regresores se vuelven más significativos. El test de BP arroja 36.35 y el test de White arroja 57.39 ambos con un nivel de bondad del 1%, además la prueba vif revela aún todavía presencia de multicolinealidad, pero esta ya se va disipando conforme aumentamos regresores, lo que indicaría que son las primera variables las que presentan este problema. Se rechaza hipótesis nula de homocedasticidad, lo que se contrasta con prueba F conjunta un poco más alta (de 13.59 a 14.23)

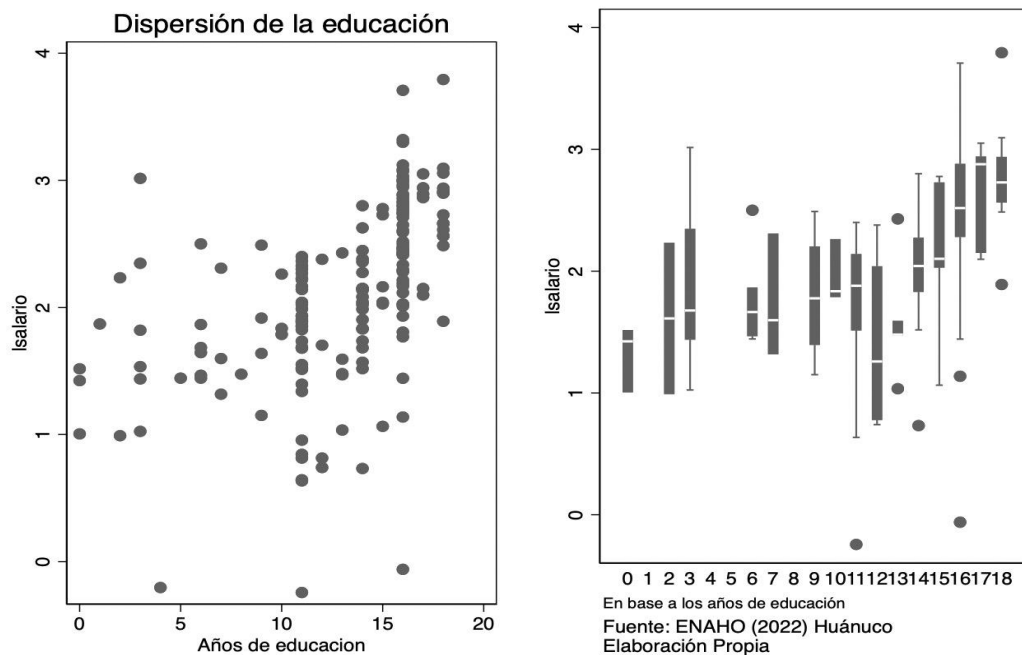
Modelo 3: *reg lsalario educ edad edad2 lengmat lxeduc lxact1 lxact3 lxtam lxmarry lxsing*

Finalmente los resultados del modelo analizado se ven mucho mejor comparando el modelo base restringido con el modelo final, además de realizar modificaciones y excluir variables irrelevantes, obtenemos:

VARIABLES	1 Modelo base	2 Modelo 1	3 Modelo 2	4 Modelo 3
educ	0.0790*** (0.0129)	0.0783*** (0.0127)	0.0767*** (0.0123)	0.0765*** (0.0124)
edad	0.0539* (0.0304)	0.0450 (0.0300)	0.0349 (0.0292)	0.0339 (0.0310)
edad2	-0.000476 (0.000349)	-0.000384 (0.000344)	-0.000300 (0.000334)	-0.000294 (0.000351)
lengmat	-0.360 (0.393)	-0.394 (0.394)	0.258 (0.425)	0.223 (0.435)
lxeduc	0.0258 (0.0272)	0.0337 (0.0271)	0.00221 (0.0277)	0.000989 (0.0297)
lxact1		0.0312 (0.230)	-0.0407 (0.224)	-0.0695 (0.230)
lxact3		-0.491*** (0.170)	-0.513*** (0.164)	-0.512*** (0.166)
lxtam			-0.510*** (0.145)	-0.520*** (0.155)
lxmarry				0.0990 (0.172)
lxsing				0.0370 (0.203)
Constant	-0.248 (0.658)	-0.0433 (0.649)	0.231 (0.633)	0.259 (0.676)
Observations	178	178	178	178
R-squared	0.326	0.359	0.403	0.404
R-adjust	0.307	0.332	0.374	0.369
Standard errors in parentheses				
*** p<0.01, ** p<0.05, * p<0.1				

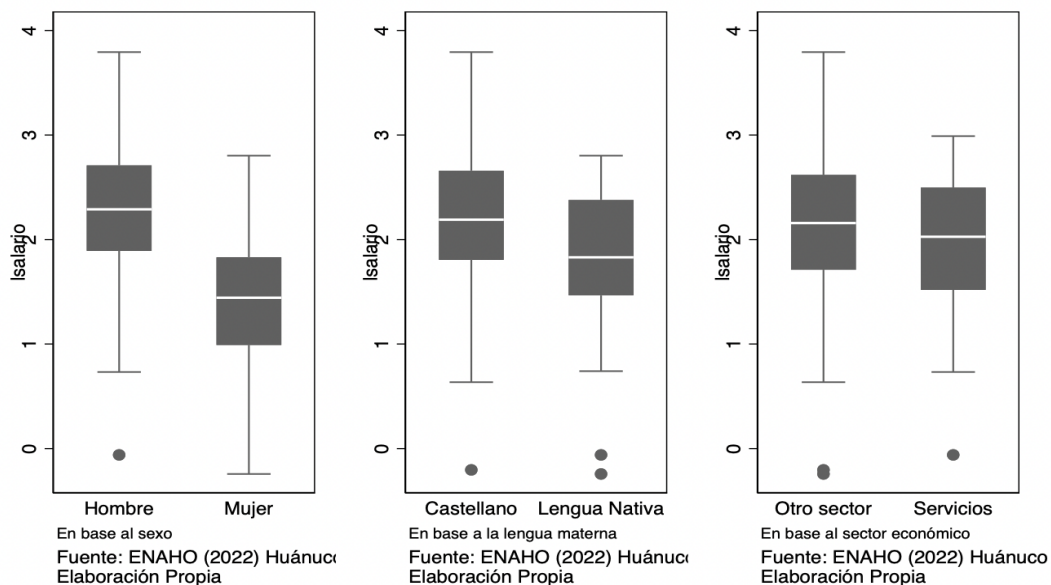
Según el cuadro, hasta el modelo 2 se incluyen variables relevantes observando el comportamiento del R cuadrado y el R cuadrado ajustado. Además, esas variables relevantes son la variable educ, lxact3 y lxtam, todas ellas con un nivel de significación del 1%. En ese sentido, observamos que el signo y significancia de los coeficientes de interés de la variable lengmat y lxeduc se vuelven de negativo a positivo y se mantiene positivo, respectivamente, además de que poseen baja significancia.

En resumen, tanto en el modelo de discriminación de sexo y de lengua materna en el mercado laboral poseen alta correlación con el regresor de la educación. En este último caso, podemos observar la relevancia de cuando el individuo tiene lenguaje nativo y trabaja en el sector de servicios, y cuando tiene lenguaje nativo y trabaja en una pequeña y micro empresa.



En este cuadro observamos como es la dispersión del logaritmo del salario, así como el gráfico de cajas, en base a los años de educación, evidenciando así su alta correlación y significancia en el modelo.

Gráfico de cajas según sexo, lengua materna y actividad económica



Estos gráficos de cajas evidencian la diferencia salarial entre dos tipos de categorías de diferentes variables dummy, de tal forma se observa que, en cierto sentido, ser mujer o tener lenguaje nativo puede influir en una reducción porcentual del salario del individuo. Sin

embargo, como se ha mostrado en las regresiones anteriores, ello tiene que venir acompañado de ciertos factores y variables extra.

5. Conclusiones

Podemos encontrar a partir del modelo econométrico planteado, una serie de conclusiones las cuales serán divididas en tres partes con relación a las etapas de las estimaciones econométricas del modelo planteado, y los datos obtenidos de la muestra.

En primer lugar, se hizo la regresión de “reg lsalario educ edad edad2 sexo sxeduc”, ya que en esta parte se relaciona el salario, educación, edad, edad no lineal, sexo y la interacción del sexo junto a la educación. Una vez obtenida la regresión, se obtuvo que el sexo es la variable que más impacto produce con un valor absoluto, su beta es de -0.36, la cual implica que el sexo, dado que como base la dummy de sexo cuando es hombre, se concluye que cuando se trata de una mujer las remuneraciones son menores.

En segundo lugar, se incluyeron nuevas variables como procedencia, categoría ocupacional, lengua materna, etc. Las cuales sirven para poder encontrar una mayor idea de las relaciones que existen en el modelo, al añadir variables. En este caso se encuentra a partir de la regresión, que la variable preponderante, en valor absoluto, es “sxtam”, siendo que esta implica el sexo y el tamaño de empresa. Esto nos demuestra que en Huánuco, si una persona es mujer y trabaja en una pequeña empresa se obtendrá una relación negativa de -0.69, lo cual es un índice bastante alto y parece corroborar las hipótesis previas, añadiendo la importancia del tamaño de la empresa lo cual resulta de la evaluación del presente trabajo.

En tercer lugar, agrupamos, las nuevas variables, del modelo 3, modelo 4 y modelo 5, ya que en cada una de ellas iremos agregando nuevas variables de las cuales se pudo obtener resultados importantes. Las variables que han sido añadidas son la lengua materna, categoría ocupacional y el estado civil. Ello es porque se considera a partir de la revisión de literatura que culturalmente dichos parámetros pueden generar un cambio importante sobre los salarios. Teniendo el modelo conjunto, se puede encontrar una que las variables que influyen mayor manera en forma absoluta son “sxact3” y “sxtam”, los cuales hace referencia a sexo multiplicado por actividad 3 (servicios), y el segundo a sexo.tamaño de empresa. De acuerdo a esto, se puede concluir que una mujer (recordar que la base dummy es hombre), en una pequeña empresa gana menos que un hombre, pero ello mejora si una mujer trabaja en una empresa pequeña pero de servicios, esto matemáticamente se explica con los valores de los beta, 0.28 y -0.66 respectivamente. Finalmente, podemos inferir que las variables más relevantes son “educ” y “sxtam” los cuales indican la variable de educación y la variable de sexo multiplicado por tamaño, posterior a ello, se realizó test de multicolinealidad observando F en la tabla, lo cual arroja como resultado “no multicolinealidad”. Este resultado se relaciona de forma directa con las hipótesis previas establecidas en el marco teórico, las cuales se sintetizan en que sí existe una relación entre la educación, sexo e ingresos salariales, una de las nuevas variables que se podían incluir en otros aspectos ha resultado ser el tamaño de empresa, la cual ayuda a explicar la estimación principal.

En ese caso, se puede obtener que para, para el caso de Huánuco, si las mujeres trabajan en pequeñas y medianas empresas, obtendrán, en promedio, una reducción porcentual en salarios (del 67.4%).

Por otro lado, aquellos individuos con lengua materna nativa que trabajen en el sector servicios dentro de una pequeña y micro empresa con mucha probabilidad (significancia) tendrán una reducción salarial respecto a la muestra (del 51% y 52% respectivamente).

En conclusión, la discriminación por sexo en el mercado laboral no se presenta únicamente, en este caso, por ser mujer, sino que hay mucha probabilidad de que si exista si es una mujer trabajadora en una pequeña y mediana empresa; además, la discriminación por lengua materna en el mercado laboral, únicamente es muy probable que ocurre cuando el individuo tienen lengua materna y trabaja en una pequeña y mediana empresa o/y su empleador se desempeña en el sector servicios. No hay evidencia suficiente de que exista discriminación como tal, sino más bien si esas características vienen acompañas con otras.

6. Bibliografía:

Barrantes, R. (2019). EN CAPILLA "Desigualdades en la inserción laboral de mujeres jóvenes". IEP. Obtenido de <http://repositorio.iep.org.pe/handle/IEP/9>

CEPAL. (1990). Aplicación de los modelos econométricos cualitativos. Montevideo, Uruguay. Obtenido de https://www.cepal.org/sites/default/files/publication/files/28673/S3314D568A_es.pdf

Escot, L. (2014). DISCRIMINACIÓN SALARIAL DE GÉNERO EN EL MERCADO. Madrid, España: Universidad Complutense de Madrid. Obtenido de <https://www.ucm.es/data/cont/docs/85-2014-01-20-Discriminaci%C3%B3n%20salarial%20g%C3%A9nero%20Comunidades%20Aut%C3%B3nomas.pdf>

Women, A. A. (2018). The simple truth about the gender pay gap. Obtenido de <https://www.aauw.org/app/uploads/2020/02/AAUW-2018-SimpleTruth-nsa.pdf>

ANEXO

global trabajo "/Users/gabriel/Documents/ARCHIVOS CATO/2023-1/Fundamentos de econometría/Trabajo final"

***** PARTE 1 *****

```
use "$trabajo/enaho01-2022-200.dta", clear
keep conglome vivienda hogar codperso ubigeo dominio estrato p203 p207 p208a p209 facpob07
merge 1:1 conglome vivienda hogar codperso using "$trabajo/enaho01a-2022-300.dta" ,
keepus(p300a p301a p301b p301c p301d p301a0)
drop if _merge==1
drop _merge
merge 1:1 conglome vivienda hogar codperso using "$trabajo/enaho01a-2022-500.dta", keepus(p513t
i524e1 i518 i513t i530a i538e1 i541a ocu500 ocupinf)
drop if _merge==1
drop _merge
tab p301a
```

```
recode p301a (1/4=0) (5/6=6) (7/10=11) (11=16) (.=.), gen(año_educ)
egen suma = rowtotal(p301b p301c)
gen educ = suma + año_educ
```

```
*Como son trabajadores dependientes, solo utilizamos algunas variables
gen horas_totales = p513*48
replace horas_totales=. if horas_totales == 0
```

```
gen Salario = i524e1/horas_totales
replace Salario=. if Salario == 0
```

```
*Generamos variable dependiente
gen lsalario = ln(Salario)
```

```
*Generamos variable edad
drop if p208a < 24
drop if p208a > 65
rename p208a edad
gen edad2 = edad^2
```

```
*Generamos dummy de sexo, tomando como base a dummy de hombre
gen sexo = 0 if p207 == 1
replace sexo = 1 if p207 == 2
label define lsexo 0 "hombre" 1 "mujer"
label values sexo lsexo
```

```
*Generamos dummy interactiva de edad y sexo
```

```
gen sxeduc = sexo*educ
```

```
*Regresionamos para observar primero resultados  
reg lsalario educ edad edad2 sexo sxeduc
```

```
* En primera instancia observamos que la variable sexo es la mas significativa, ///
```

```
* dado que tomamos como base la dummy de sexo cuando es hombre, observamos que ///
```

```
* cuando se es mujer la remuneracion es mas baja.
```

```
keep conglome vivienda hogar codperso ubigeo dominio estrato p203 p209 p300a ocu500 ocupinf  
Salario lsalario edad edad2 sexo educ sxeduc facpob07  
save "$trabajo/Datos1", replace
```

```
***** PARTE 2 *****
```

```
* Ahora procederemos a incluir otras posibles variables de interes, entre ellas ///
```

```
* la de region de procedencia, categoria ocupacional, lengua materna, entre otros
```

```
use "$trabajo/enaho01-2022-200.dta", clear
```

```
keep conglome vivienda hogar codperso ubigeo dominio estrato p203 p207 p208a p209 facpob07
```

```
merge 1:1 conglome vivienda hogar codperso using "$trabajo/enaho01a-2022-300.dta" ,
```

```
keepus(p300a p301a p301b p301c p301d p301a0)
```

```
drop if _merge==1
```

```
drop _merge
```

```
merge 1:1 conglome vivienda hogar codperso using "$trabajo/enaho01a-2022-500.dta"
```

```
tab dominio /// region de procedencia
```

```
tab p507 /// categoria ocupacional
```

```
tab p516r4 /// actividad de la empresa
```

```
tab p517d1 /// tamaño de la empresa
```

```
tab p209 /// estado civil
```

```
tab p300a /// lengua materna
```

```
***
```

```
drop if p507 == 1 /// eliminamos estas categorias dado que nos enfocamos en trabajadores  
dependientes
```

```
drop if p507 == 2
```

```
***** CATEGORIA OCUPACIONAL *****
```

```
* Definimo primero las categorias existentes, luego las agruparemos para formar dummy
```

```
gen cat_ocup = 1 if p507 == 3
```

```
replace cat_ocup = 2 if p507 == 4
```

```
replace cat_ocup = 3 if p507 == 5
```

```
replace cat_ocup = 4 if p507 == 6
```

```
replace cat_ocup = 5 if p507 == 7
```

```
label define locup 1 "empleado" 2 "obrero" 3 "familiar no remunerado" 4 "trabajador del hogar" 5  
"otro"
```

```
label values cat_ocup locup
```

```
label var cat_ocup "Categoria ocupacional"
```

*Formamos dummy de categoria ocupacional

```
recode cat_ocup (1/2=0 "Empleado u obrero") (3/5=1 "Trabajador(a) del hogar u otro"),  
gen(ocupacion)
```

***** ACTIVIDAD DE LA EMPRESA *****

```
codebook p516r4
```

```
tab p516r4, nol
```

```
tab p516r4
```

```
*tab etiquetas, nol
```

```
recode p516r4
```

```
(111=1)(112=1)(113=1)(114=1)(116=1)(119=1)(121=1)(122=1)(123=1)(124=1)(125=1)(126=1)(127  
=1)(128=1)(129=1)(130=1)(141=1)(142=1)(143=1)(144=1)(145=1)(146=1)(149=1)(150=1)(161=1)(1  
70=1)(220=1)(230=1)(311=2)(312=2)(322=2)(729=3)(810=3)(899=3)(1010=4)(1050=4)(1061=4)(10  
71=4)(1073=4)(1079=4)(1102=4)(1104=4)(1311=4)(1313=4)(1392=4)(1393=4)(1394=4)(1410=4)(14  
30=4)(1512=4)(1520=4)(1622=4)(1629=4)(1811=4)(2011=4)(2100=4)(2310=4)(2392=4)(2393=4)(23  
95=4)(2396=4)(2511=4)(2592=4)(2593=4)(2599=4)(3100=4)(3211=4)(3212=4)(3240=4)(3250=4)(32  
90=4)(3311=5)(3312=5)(3319=5)(3510=5)(3600=5)(3811=5)(4100=6)(4210=6)(4220=6)(4321=6)(43  
22=6)(4330=6)(4390=6)(4520=7)(4540=7)(4620=7)(4630=7)(4641=7)(4649=7)(4651=7)(4659=7)(46  
61=7)(4663=7)(4669=7)(4721=7)(4722=7)(4730=7)(4741=7)(4752=7)(4759=7)(4761=7)(4764=7)(47  
71=7)(4772=7)(4773=7)(4774=7)(4781=7)(4782=7)(4789=7)(4791=7)(4799=7)(4921=8)(4922=8)(49  
23=8)(5021=8)(5221=8)(5224=8)(5229=8)(5310=9)(5320=9)(5510=9)(5610=9)(5621=9)(5629=9)(56  
30=9)(5813=9)(5911=9)(6010=9)(6110=9)(6120=9)(6130=9)(6190=9)(6202=9)(6311=9)(6312=9)(el  
se=10), gen(actividad)
```

* Observamos actividad economica desagregada, con finalidad de simplificar el analisis, se agrupara posteriormente algunas actividades para crear solo 3 categorias

```
label define etiquetas2 1 "Agropecuaria" 2 "Pesca" 3 "Extractiva" 4 "Manufactura" 5 "Mantenimiento"  
6 "Construcción" 7 "Comercio" 8 "Transporte" 9 "Telecomunicaciones" 10 "Servicios y diversos"
```

```
label values actividad etiquetas2
```

```
tab actividad
```

* Definimos dummy de tres categorias de actividad económica

```
recode actividad (1/3=1) (4/9=2) (10/10=3), gen(Actividad)
```

```
label define etiquetas3 1 "Agropecuario,pesca y extractivo" 2 "Manufactura,  
construcción,comercio,transporte y telecomunicaciones" 3 "Servicios y diversos"
```

```
label value Actividad etiquetas3
```

* Generamos dummy de actividad economica de la empresa

```
gen actividad1 = (Actividad == 1) // dummy de agro pesca y extractivo
```

```
gen actividad2 = (Actividad == 2) // dummy de manu,cons,com, trans y tele
```

```
gen actividad3 = (Actividad == 3) // dummy de servicios y otros
```

```
label var actividad1 "Actividad economica 1 (agro,pesca y extrac)"
```

```
label var actividad2 "Actividad economica 2 (manu,cons,com,trans y tele)"
```

```
label var actividad3 "Actividad economica 3 (servs, otros)"
```

***** TAMAÑO DE LA EMPRESA *****

recode p512a (1/3=1) (4/5=0), gen(tamaño)

label define etiquetas4 1 "Pequeña y Mediana empresa" 0 "Gran empresa"

label value tamaño etiquetas4

***** LENGUA MATERNA *****

recode p300a (1/3=0) (8/14=0) (4/4=1) (5/7=.), gen(lengmat)

drop if lengmat == 15 /// eliminamos valores perdidos, y consideramos solo castellano y lengua nativa

label define etiquetas8 0 "Castellano" 1 "Lengua Nativa"

label value lengmat etiquetas8

***** ESTADO CIVIL *****

recode p209 (1/2=1 "Conviviente o Casado") (3/5=2 "Otro") (6/6=3 "Soltero"), gen(estado_civil)

gen est1_casado = (estado_civil == 1)

gen est1_soltero = (estado_civil == 3)

gen est1_otro = (estado_civil == 2)

keep conglome vivienda hogar codperso ubigeo estrato p207 p203 p208a p300a lengmat p512a
tamaño p507 ocupacion dominio procedencia1 p1_* procedencia2 p2_* p516r4 Actividad activi*
p209 estado_civil est1_* facpob07

save "\$trabajo/Datos2", replace

* Finalmente mantenemos todas las nuevas variables creadas y hacemos un merge con la regresion inicial

merge 1:1 conglome vivienda hogar codperso using "\$trabajo/Datos1", keep(matched)

keep conglome vivienda hogar codperso ubigeo estrato Salario lsalario sexo edad edad2 educ sxeduc
p300a lengmat p512a tamaño p507 ocupacion dominio procedencia1 p1_* procedencia2 p2_* p516r4
Actividad activi* p209 estado_civil est1_* facpob07

* Al incluir todas las variables de interes, mas de la mitad de observaciones se reduce, lo que indicaria que al agregar mas variables exogenas, si bien aumenta la capacidad para explicar el logaritmo del salario, la significancia de los regresores disminuye

* Para el analisis de las regresiones ya nos enfocamos en la area de estudio: Huanuco

keep if (_n >= 9835 & _n <= 10266)

drop procedencia* p1_* p2_*

drop dominio

save "\$trabajo/Base de datos", replace

twoway histogram lsalario || normal lsalario , fcolor(ltblueishgray) lcolor(red) title("Logaritmo del Salario") ytitle("Porcentaje")

save grafico0, replace

*observamos que los primero dos modelos tendran mas oservaciones, entonces eliminaremos la diferencia para que haya mejor estimacion

predict ehat, resid

drop if ehat==.

drop ehat

***** PARTE 3 *****

*Estimando discriminacion por sexo en el mercado laboral

*Mopdelo base

reg lsalario educ edad edad2 sexo sxeduc

outreg2 using "\$trabajo/tabla2.xls", excel append ctitle("Modelo base")

predict ehat, resid

hist ehat, normal title("Distribución residual Modelo base")

save grafico1, replace

vif

estat hettest, rhs

estat imtest, white

*Modelo 1: inclusion de actividad economica. Tomamos como base a actividad2

gen sxact1 = sexo*actividad1

gen sxact3 = sexo*actividad3

reg lsalario educ edad edad2 sexo sxeduc sxact1 sxact3

outreg2 using "\$trabajo/tabla2.xls", excel append ctitle("Modelo 1")

vif

estat hettest, rhs

estat imtest, white

*Modelo 2: inclusion de tamaño de la empresa. Tomamos como base a Gran empresa

gen sxtam = sexo*tamaño

reg lsalario educ edad edad2 sexo sxeduc sxact1 sxact3 sxtam

outreg2 using "\$trabajo/tabla2.xls", excel append ctitle("Modelo 2")

vif

estat hettest, rhs

estat imtest, white

// parece ser una variable relevante

*Modelo 3: inclusión de lengua materna. Tomamos como base a Castellano

gen sxleng = sexo*lengmat

reg lsalario educ edad edad2 sexo sxeduc sxact1 sxact3 sxtam sxleng

outreg2 using "\$trabajo/tabla2.xls", excel append ctitle("Modelo 3")

vif

estat hettest, rhs

estat imtest, white

// definitivamente lengua materna no aporta a la estimacion (R igual y R cuadrado disminuye)

*Modelo 4: inclusion de categoria ocupacional. Tomamos como base a empleado

gen sxocup = sexo*ocupacion

reg lsalario educ edad edad2 sexo sxeduc sxact1 sxact3 sxtam sxleng sxocup

outreg2 using "\$trabajo/tabla2.xls", excel append ctitle("Modelo 4")

vif

estat hettest, rhs

estat imtest, white

// probablemente sea algo significativo, pero muchas variables disminuye su peso

*Modelo 5: inclusion de estado civil. Tomamos como base a "otros"

gen sxmarry = sexo*est1_casado

gen sxsing = sexo*est1_soltero

reg lsalario educ edad edad2 sexo sxeduc sxact1 sxact3 sxtam sxleng sxocup sxmarry sxsing

vif

estat hettest, rhs

estat imtest, white

test (educ=edad=edad2=sexo=sxeduc=sxact1=sxact3=sxtam=sxleng=sxocup=sxmarry=sxsing=0)

// definitivamente estado civil no aumenta la estimacion, R y R cuadrado no suben nada

*Ahora corregimos toda las anteriore estimaciones y retiramos variables irrelevantes o que brindan informacion redundante

reg lsalario educ edad edad2 sexo sxeduc

outreg2 using "\$trabajo/mejor estimacion.xls", excel append ctitle("Modelo base")

reg lsalario educ edad edad2 sexo sxeduc sxact1 sxact3 sxtam sxocup

outreg2 using "\$trabajo/mejor estimacion.xls", excel append ctitle("Modelo aprox")

vif

estat hettest, rhs // rechazamos homocedasticidad --> hay heterocedasticidad

estat imtest, white // no rechazamos homocedasticidad --> no es concluyente

*Solucionando no rechazo de homocedasticidad

reg lsalario educ edad edad2 sexo sxeduc sxact1 sxact3 sxtam sxocup, vce(robust)

*Observamos que con aun con los mejores regresores, las variables mas ///

*relevantes son educ y sxtam, hacemos test de multicolinealidad observando F ///

*en la tabla, pero parece no haber multicolinealidad. Probablemente, para el ///

*caso de Huanuco, ni el hecho de ser mujer ni la cantidad de años de estudio ///

*de las mujeres ni el tipo de actividad economica a la que su empleador trabaje ///

*ni la vejez ni la categoria ocupaccional influyen significativamente en su//

*salario. En cambio, si las mujeres trabajan en pequeñas y medianas empresas y ///

*los años de educación en general, independienemente de si sean mujeres, son ///

*los principales determinantes del salario en Huanuco

***** PARTE 4 *****

* Estimando discriminación por lengua materna en mercado laboral

* Debemos crear nuevas dummies interactivas para el estudio

```
gen lxeduc = lengmat*educ      // dummy interactiva educ
gen lxact1 = lengmat*actividad1 // dummy interactiva agro, pesca y extrac
gen lxact3 = lengmat*actividad2 // dummy interactiva servicios y otros
gen lxtam = lengmat*tamaño     // dummy interactiva tamaño de la empresa
gen lxocup = lengmat*ocupacion // dummy interactiva categoria ocupacional
gen lxmarry = lengmat*est1_casado // dummy interactiva casado
gen lxsing = lengmat*est1_soltero // dummy interactiva soltero
```

*Se toma como base a actividad de manufactura, construcción, comercio, transporte y telecomunicaciones, base a gran empresa, base a empleado u obrero y de base a viudo, divorciado u otros.

* Modelo base

```
reg lsalario educ edad edad2 lengmat lxeduc
outreg2 using "$trabajo/tabla3.xls", excel append ctitle("Modelo base")
vif
estat hettest, rhs
estat imtest, white
```

* Modelo 1: inclusión actividad económica

```
reg lsalario educ edad edad2 lengmat lxeduc lxact1 lxact3
outreg2 using "$trabajo/tabla3.xls", excel append ctitle("Modelo 1")
vif
estat hettest, rhs
estat imtest, white
```

* Modelo 2: inclusión de tamaño de la empresa

```
reg lsalario educ edad edad2 lengmat lxeduc lxact1 lxact3 lxtam
outreg2 using "$trabajo/tabla3.xls", excel append ctitle("Modelo 2")
vif
estat hettest, rhs
estat imtest, white
```

*No se incluye variable de sexo interactiva porque se mantiene misma conclusión///

*anterior de que el hecho de ser mujer con lengua materna nativa no influye///

*significativamente en el salario

* Inclusión de categoría ocupacional

```
reg lsalario educ edad edad2 lengmat lxeduc lxact1 lxact3 lxtam lxocup
```

*encontramos que en la muestra todas las personas con lengua materna nativa no tienen la ocupación de trabajador(a) del hogar u otros, por lo que es una variable irrelevante

* Modelo 3: inclusion de estado civil

```
reg lsalario educ edad edad2 lengmat lxeduc lxact1 lxact3 lxtam lxmarry lxsing
outreg2 using "$trabajo/tabla3.xls", excel append ctitle("Modelo 3")
vif
estat hettest, rhs
estat imtest, white
```

* Mejorando estimacion y retirando variables irrelevantes

```
reg lsalario educ edad edad2 lengmat lxeduc
outreg2 using "$trabajo/mejor estimacion2.xls", excel append ctitle("Modelo base")
```

```
reg lsalario educ edad edad2 lengmat lxeduc lxact1 lxact3 lxtam
outreg2 using "$trabajo/mejor estimacion2.xls", excel append ctitle("Modelo aprox")
vif
estat hettest, rhs // rechazamos homocedasticidad --> hay heterocedasticidad
estat imtest, white // tambien rechazamos homocedasticidad --> hay heterocedasticidad
```

*Reafirmando heterocedasticidad

```
reg lsalario educ edad edad2 lengmat lxeduc lxact1 lxact3 lxtam, vce(robust)
```

*En esta estimacion para analizar la discriminacion por lengua materna, ///

*encontramos que los años de educacion, cuando se tiene idioma nativo y se ///

*trabaja en pequeñas y medianas empresas; y, cuando se tiene idioma nativo ///

*y la actividad economica de la empresa es en servicios u otros, son ///

*variables significativas que influyen en el cambio porcentual del salario.///

*En cambio, la tendencia a la desactualizacion (edad2); y, ni la lengua nativa ///

*por si misma, ni los con lengua nativa y con mayor educacion ni el hecho de ///

*tener lengua nativa y trabajan en el sector agropecuario, pesca y/o ///

*extractivo, son variables determinantes del salario.

```
label define llxact3 0 "Otro sector" 1 "Servicios"
```

```
label values lxact3 llxact3
```

*Graficos de dispersion, pie y cajas modelo de discriminacion por sexo

```
scatter lsalario educ, title("Dispersión de la educación") xtitle("Años de educacion") name(grafi1,
replace)
```

```
label define lsxtam 0 "Hombre" 1 "Mujer"
```

```
label values sxtam lsxtam
```

```
graph pie, over(sxtam) plabel(_all percent, color(white) format(%9.1f) size(medsmall))
```

```
plotregion(lstyle(none)) title("Trabajadora(a) en pequeña y mediana empresa")///
```

```
subtitle("Composición por sexo") caption("Elaboracion Propia") name(grafi2, replace)
```

```
graph box lsalario, over(educ) graphregion(color(white)) ylabel(, nogrid) note("En base a los años de
educación") caption("Fuente: ENAHO (2022) Huánuco" "Elaboración Propia") name(grafi3, replace)
```

```
graph box lsalario, over(sxtam) graphregion(color(white)) ylabel(, nogrid) note("En base al sexo")
caption("Fuente: ENAHO (2022) Huánuco" "Elaboración Propia")///
name(grafi4, replace)
```

*Graficos de dispersion, pie y cajas modelo de discriminacion por lengua materna

```
label define lllxtam 0 "Castellano" 1 "Lengua Nativa"
label values lxtam lllxtam
graph pie, over(lxtam) plabel(_all percent, color(white) format(%9.1f) ///
size(medsmall)) plotregion(lstyle(none)) title("Trabajadora(a) en pequeña y mediana empresa")
subtitle("Composición por lengua materna") ///
caption("Elaboracion Propia") name(grafi5, replace)
```

```
graph box lsalario, over(lxtam) graphregion(color(white)) ylabel(, nogrid) note("En base a la lengua
materna") caption("Fuente: ENAHO (2022) Huánuco" "Elaboración Propia") name(grafi6, replace)
```

```
graph box lsalario, over(lxact3) graphregion(color(white)) ylabel(, nogrid) note("En base al sector
económico") caption("Fuente: ENAHO (2022) Huánuco" "Elaboración Propia") name(grafi7, replace)
```

```
graph pie, over(lxact3) plabel(_all percent, color(white) format(%9.1f) ///
size(medsmall)) plotregion(lstyle(none)) title("Trabajadora(a) en Sector
Servicios") subtitle("Composición por lengua materna") ///
caption("Elaboracion Propia") name(grafi8, replace)
```

```
graph combine grafi1 grafi3
graph combine grafi4 grafi6 grafi7, title("Gráfico de cajas según sexo y lengua materna" "y actividad
económica") c(3)
```

* Realizando tablas bidimensionales

```
table sxtam, c(mean lsalario median lsalario sd lsalario min lsalario max lsalario) format(%6.2f) col
```

```
table lxtam, c(mean lsalario median lsalario sd lsalario min lsalario max lsalario) format(%6.2f) col
```

```
table lxact3, c(mean lsalario median lsalario sd lsalario min lsalario max lsalario)format(%6.2f) col
```