

Ejercicio: Discriminación salarial

Alfonso Novales
Departamento de Economía Cuantitativa
Universidad Complutense*

Noviembre 2008

1. Discriminación salarial¹

Este ejemplo tiene como objetivo describir la utilización de variables ficticias para contrastar la estabilidad del modelo de regresión entre submuestras. En la primera parte del ejercicio, utilizamos un modelo de determinación de salarios en función del nivel educativo y la experiencia laboral del trabajador, y se examinan las posibles diferencias en el modelo estimado entre las submuestras de hombres y mujeres. Tras detectar evidencia consistente con la existencia de discriminación salarial en contra de las mujeres, se profundiza en analizar si la discriminación se debe a una menor valoración del nivel educativo, la experiencia laboral, o de ambos factores. En la segunda parte explicamos diversas maneras en que el uso de variables ficticias, convenientemente definidas e introducidas en el modelo econométrico, permite contrastar la homogeneidad salarial entre trabajadores de distintas características. En esta segunda parte nos centramos en caracterizar el posible impacto diferencial que sobre el salario tengan la experiencia laboral y el nivel educativo.

*Versión muy preliminar. No citar sin permiso del autor. Las sucesivas versiones de este trabajo irán estando disponibles en <http://www.ucm.es/info/ecocuan/anc>, o pueden solicitarse en anovales@ccee.ucm.es

¹Fichero de trabajo: Bwages.wf1. La base de datos Bwages.txt está tomada de los archivos que acompañan a Kuleuven

1.1. Descripción de los datos

El archivo *Bwages.wfl* contiene datos relativos a 1.472 personas encuestadas en Bélgica en 1994, como parte del European Community Household Panel, a las que se ha preguntado por: *a*) su salario en Bef. (40 Bef. equivalen aproximadamente a 1 euro), *b*) su nivel de educación, indicando uno entre cinco niveles posibles, y *c*) su experiencia laboral, en términos del número de años que el encuestado ha trabajado hasta el momento. El salario es una variable de naturaleza continua, si bien toma únicamente valores positivos; la experiencia profesional es una variable discreta, observable numéricamente, por lo que no se trata de una variable ficticia. Por el contrario, el nivel educativo es una variable cualitativa, para la que hay que definir un variable ficticia. Esto podría hacerse de muchas formas distintas, pero lo más natural es utilizar los cinco primeros números enteros, asignándolos a cada nivel educativo, en orden creciente. Hay que entender, sin embargo, que cualquier otra asignación numérica sería asimismo posible.

A priori, es lógico considerar que ambas variables, experiencia y educación, pueden incidir positivamente sobre el nivel salarial del trabajador. Un aspecto a tener en cuenta en la interpretación del modelo es que, a diferencia de la experiencia laboral, cuyos valores numéricos sucesivos están separados siempre por un año más de experiencia, los valores sucesivos de la variable educación recogen distintos niveles educativos, no siendo en absoluto evidente que la diferencia entre dos cualesquiera de dichos niveles sucesivos haya de tener siempre un mismo efecto sobre el salario.

Por último, disponemos asimismo de información acerca de si el encuestado es hombre o mujer; ésta es también una variable cualitativa, para la que el investigador debe construir, por tanto, una variable ficticia. Dicha variable, con el nombre *male*, ya está incluida en el fichero de trabajo. En la descripción del archivo (*Bwages.txt*) se nos dice que el valor $male = 1$ corresponde a varones; en la base de datos, la variable solo toma valores 0 ó 1, por lo que se ha asociado $male = 0$ a las mujeres encuestadas.

1.2. Estadísticos descriptivos

El cálculo de estadísticos de la variable *male*, en la tabla de clasificación proporcionada por Eviews,² revela que hay en la encuesta 579 individuos con valor cero de esta variable (mujeres), siendo varones el resto, 893. Podemos calcular la

²Tras marcar la variable *male*, marcar las pestañas: View_Descriptive Statistics_ Stats by classification, y en la ventana Series/Group for classify, indicar: *male*.

media muestral (0,606) y la mediana (1,000) de esta variable pero, evidentemente, estos estadísticos carecen de interés, salvo que quisiéramos calcular, de modo muy indirecto, utilizando la media muestral, el porcentaje de personas de uno y otro sexo dentro de la muestra. La mediana únicamente nos dice en este caso que hay más individuos en la muestra con valor $male = 1$ que con valor $male = 0$; la media nos dice que el 60,6% de los encuestados son varones.

Los estadísticos muestrales, recogidos en *HIST_WAGES*, *HIST_EDUC*, *HIST_EXPER* nos revelan asimismo que los niveles educativos observados en la muestra son 1, 2, 3, 4, 5, con media muestral de 2,38 y mediana de 3,0, mientras que los niveles de experiencia oscilan entre 0 y 47, siendo únicamente números enteros, con una media de 17,22 años y una mediana de 16,5 años.

Vamos a utilizar estos datos para efectuar un doble análisis: por un lado, caracterizar los determinantes del salario que recibe un trabajador; por otro, contrastar si existe discriminación salarial entre hombres y mujeres en el mercado de trabajo de donde provienen los datos.

Para analizar los determinantes salariales, comenzamos calculando el salario medio por nivel educativo, obteniendo, respectivamente, 340,03; 371,74; 411,60; 461,13; 563,20, para los cinco sucesivos niveles educativos.³ Que dichos salarios medios sean crecientes sugiere que el nivel educativo es un posible determinante del salario. La evidencia similar para la experiencia profesional es más difícil de obtener, entre otras cosas, porque en este caso hay 48 niveles de experiencia diferentes, en muchos de los cuales hay un número reducido de encuestados, lo que hace que el salario medio de dicho grupo se mida con poca precisión. Estrictamente hablando, el salario medio no es creciente con cada año de experiencia profesional⁴, aunque se observa una clara tendencia a aumentar con dicha variable; en todo caso, la evidencia muestral acerca de que el salario aumenta con la experiencia profesional es más tenue que respecto al nivel educativo. El coeficiente de correlación entre salario y experiencia es de 0,307, sugiriendo que existe cierta relación entre estas variables. La nube de puntos (*nube_wage_exper*) que representa el salario en función de la experiencia arroja asimismo una cierta imagen de dependencia, aunque débil, especialmente para los trabajadores de mayor experiencia.

Para obtener una primera evidencia acerca de una posible discriminación salar-

³Para ello, utilizamos la variable *educ* como clasificador al calcular los estadísticos descriptivos de la variable *wage*, obteniendo *WAGE_BY_EDUC*.

⁴Como se puede ver clasificando la variable *WAGE* utilizando *EXPER* como clasificador, lo que da lugar a la tabla *WAGE_BY_EXPER*.

ial, calculamos el salario medio por sexos, obteniendo 413,95 Bef. para las mujeres, frente a 466,42 Bef. para los hombres, y un salario medio global, para todos ellos, de 445,78 Bef. que está, lógicamente, comprendido entre el promedio de uno y otro grupo. Este dato inicial ya es favorable a la posible existencia de discriminación salarial, por cuanto que el salario medio de los hombres es superior al de las mujeres, supuesto que todos ellos desempeñan tareas laborales comparables. Otro aspecto interesante es que el histograma de la variable salarios revela una fuerte curtosis, que conduce al rechazo de la hipótesis de Normalidad de acuerdo con el test de Bera-Jarque. Esto significa que no es muy razonable creer que dicha variable quede determinada mediante la suma de un conjunto de variables deterministas y un término estocástico con distribución Normal, como se supone en el modelo de regresión lineal habitualmente.

Antes de afirmar categóricamente la existencia de discriminación salarial, sin embargo, debemos considerar la posibilidad de que la diferencia entre los salarios medios de ambos grupos pueda obedecer a que las características de los hombres y mujeres encuestados, en términos de educación y experiencia, sean diferentes: que el salario medio de los varones fuera más elevado no indicaría necesariamente discriminación salarial, si su nivel educativo y experiencia profesional fuesen más altos. La experiencia profesional media es de 15,2 años entre mujeres y de 18,5 años entre hombres⁵, con una media global de 17,2 años. Los valores mediana son 14 y 18, respectivamente, para ambos grupos. El nivel educativo medio es de 3,59 entre mujeres, y de 3,24 entre hombres, con una media muestral global de 3,38. Los niveles educativos mediana son de 4 y 3, respectivamente, para mujeres y hombres. Así, en esta muestra los hombres tienen, en media, mayor experiencia laboral pero menor nivel educativo que las mujeres. Al no ser ambas variables más elevadas entre hombres que entre mujeres, la diferencia en salarios entre ambos grupos puede estar indicando, efectivamente, discriminación salarial contra las mujeres.

1.3. Análisis de regresión

Comenzamos el análisis de regresión utilizando separadamente el nivel educativo y la experiencia laboral como variables explicativas del salario,

⁵Como se puede ver clasificando la variable *EXPER* utilizando *MALE* como clasificador. View_Descriptive statistics_Stats by classification, escribiendo *male* en la ventana de Series/Group for Classify.

$$\begin{aligned} \text{Salario}_i = 249,51 + 58,10 \text{ Educación}_i + u_i, \quad \bar{R}^2 = 0,151, \quad \hat{\sigma}_u = 165,4. \quad (1.1) \\ (3,58) \\ (16,22) \end{aligned}$$

Las desviaciones típicas y los estadísticos t aparecen entre paréntesis, bajo el valor estimado de la pendiente de la recta de regresión. Como la desviación típica muestral de la variable salario es 179,53, el *Ratio* $1 - \frac{\hat{\sigma}_u}{\sigma_y} = 0,08$. El salario parece aumentar, en media, en algo más de 58 Bef. por año de educación; por el modo en que se ha definido la variable *Educación*, la regresión estimada sugiere, además, que el incremento salarial por año de educación adicional es siempre el citado, con independencia de que se trate del paso del nivel educativo 1 al 2, o del nivel 4 al nivel 5. La regresión,

$$\begin{aligned} \text{Salario}_i = 352,36 + 5,43 \text{ Experiencia}_i + u_i, \quad \bar{R}^2 = 0,094, \quad \hat{\sigma}_u = 170,9. \quad (1.2) \\ (0,44) \\ (12,38) \end{aligned}$$

proporciona una estimación de que el salario aumenta en unos 5,4 Bef. por año de experiencia laboral. Nuevamente, la regresión estimada restringe a que el incremento salarial por año de experiencia adicional sea el mismo tanto si el aumento se produce a un nivel reducido como a un nivel elevado de experiencia.

Ninguna de las dos variables consideradas parecen tener, por sí solas, gran capacidad explicativa sobre el salario. Si estimamos una regresión con ambas, tenemos un R^2 ajustado de 0,344, con $\hat{\sigma}_u = 145,4$, y ratio de ajuste $1 - \frac{\hat{\sigma}_u}{\sigma_y} = 0,19$.

$$\begin{aligned} \text{Salario}_i = 43,31 + 77,87 \text{ Educación}_i + 8,10 \text{ Experiencia}_i + u_i, \quad (1.3) \\ (3,29) \qquad \qquad \qquad (0,39) \\ (23,7) \qquad \qquad \qquad (20,8) \end{aligned}$$

$$R^2 = 0,345, \quad \bar{R}^2 = 0,344, \quad \hat{\sigma}_u = 145,45, \quad SR = 31079583;$$

Es interesante que los dos coeficientes estimados difieren apreciablemente de los obtenidos en las regresiones individuales, siendo en ambos casos superiores a los estimados en dichas regresiones simples. Ahora, estimamos que el salario aumenta en casi 78 Bef. por año de educación, y en más de 8 Bef. por año de experiencia.

Que los coeficientes hayan aumentado respecto de la regresión simple, sugiere que ambas variables estén negativamente correlacionadas; en efecto, si así fuera, un año más de educación vendría generalmente asociado con una menor experiencia laboral, como por otra parte parece razonable. De hecho, el coeficiente de correlación habitual entre ambas variables es -0,29, reflejando tal correlación negativa si bien, tratándose de variables cualitativas, el uso de dicho estadístico es cuestionable. Los niveles de experiencia medios son de 25,9, 20,7, 17,0, 14,9 y 14,5 años para los niveles educativos de 1 a 5, respectivamente, con una media global de 17,2 años. La ordenación decreciente de dichos promedios sugiere asimismo la correlación negativa entre ambas variables.

1.4. Contraste de discriminación salarial mediante variables ficticias

Los residuos de la regresión combinada tienen, por supuesto, una media muestral igual a cero. Sin embargo, su media es de -31,8 entre las mujeres (observaciones con $male = 0$) y de 20,6 entre hombres (observaciones con $male = 1$), sugiriendo claramente un diferente comportamiento de los salarios entre ambos grupos. Ello significa que dados un mismo nivel de educación y de experiencia, el salario es 52,4 Bef. inferior para las mujeres que para los hombres; esta observación constituye evidencia más clara a favor de discriminación salarial contra las mujeres. Cuando incluimos en la regresión anterior la variable ficticia $MALE$, obtenemos un R^2 ajustado de 0,364, con $\hat{\sigma}_u = 143,1$, y el ratio de ajuste aumenta a 0,20. La regresión estimada es,

$$\begin{array}{cccc}
 \text{Salario} = & 8,62 + & 7,76 & \text{Experiencia} + & 80,11 & \text{Educación} + & 54,30 & \text{Male} & (1.4) \\
 & \text{\scriptsize (15,6)} & \text{\scriptsize (0,39)} & & \text{\scriptsize (3,25)} & & \text{\scriptsize (7,77)} & & \\
 & & \text{\scriptsize (20,1)} & & \text{\scriptsize (24,6)} & & \text{\scriptsize (7,0)} & &
 \end{array}$$

que sugiere que tanto el nivel educativo como la experiencia profesional explican el salario de un trabajador, y también que existen diferencias entre los salarios que reciben mujeres y hombres que tienen un mismo nivel educativo e igual experiencia laboral.

Puesto que la variable $Male$ toma el valor 0 para las mujeres, y el valor 1 para los hombres, la regresión anterior equivale al par de regresiones,

$$\begin{array}{l}
 \text{Salario} = 62,92 + 7,76\text{Experiencia} + 80,11\text{Educación}, \text{ para los hombres} \\
 \text{Salario} = 8,62 + 7,76\text{Experiencia} + 80,11\text{Educación}, \text{ para las mujeres}
 \end{array}$$

Como ejemplo, nuestra estimación sugiere que un hombre de 10 años de experiencia laboral y 4 años de educación recibiría un salario de $62.92 + (7.76)10 + (80.11)4 = 460.96$, mientras que una mujer de igual cualificación recibiría un salario igual a $8.62 + (7.76)10 + (80.11)4 = 406.66$.

Todo ello proporciona evidencia clara acerca de discriminación salarial en el sentido antes descrito: a igualdad de experiencia y educación, un hombre recibe un salario superior en 54,3 unidades monetarias al de una mujer⁶. No parece preciso contrastar explícitamente tal hipótesis. Además, el histograma de frecuencias de la variable salarios se desvia claramente respecto de una distribución Normal por lo que la teoría estadística habitual basada, entre otros, en el supuesto de Normalidad de la perturbación del modelo y el carácter determinista de las variables explicativas, no es estrictamente aplicable en este caso. Afortunadamente, como hemos dicho, tampoco parece necesaria su aplicación.

Las dos regresiones anteriores se diferencian tan sólo en la estimación de la constante, por lo que gráficamente pueden visualizarse como dos líneas de regresión paralelas, con igual pendiente, pero con mayor ordenada en el origen para la regresión de hombres que para la correspondiente a las mujeres. Es decir, la regresión de hombres está sistemáticamente por encima de la de las mujeres, lo que hace que para cada posible combinación de nivel educativo y experiencia, el salario de los hombres (la variable dependiente del modelo) sea mayor que el de las mujeres⁷.

1.5. Aporta la variable Experiencia información acerca de la determinación salarial, adicional a la que continen el nivel educativo y el sexo del trabajador?

Aunque los resultados parecen claros, un investigador podría dudar de la verdadera relevancia de la información aportada por la variable experiencia sobre los salarios, pues la evidencia proporcionada en este sentido por el modelo de regresión simple, así como por los estadísticos descriptivos podría sugerirlo. Para ello, estimaríamos el modelo omitiendo la experiencia laboral y, utilizando, por tanto, el nivel educativo y la variable ficticia para explicar los salarios,

⁶No es casualidad que esta diferencia coincide con la disparidad antes mencionada entre las medias muestrales de los residuos en ambos grupos de trabajadores: $20,6 - (-31,8) = 52,4$

⁷Si bien este argumento no es estrictamente válido porque con dos variables explicativas, experiencia y educación, no tenemos rectas de regresión, sino planos de regresión. Sin embargo, la idea intuitiva es la misma.

$$\text{Salario} = \beta_0 + \beta_1 \text{Experiencia} + \beta_2 \text{Educación} + \beta_3 \text{Male}$$

y compararíamos los residuos de esta ecuación restringida, con los de (??). Además del claro descenso en los estadísticos de ajuste (*reg_w_edu_male*), la nube de puntos (*SIGNIFICA_EXPER*) que representa ambos conjuntos de residuos indica claramente las diferencias entre ellos o, lo que es lo mismo, que la experiencia profesional es una variable significativa para explicar la determinación de los salarios en este grupo de trabajadores. El coeficiente de correlación entre los residuos de ambos modelos, incluyendo y excluyendo la variable *Experiencia* es 0,88, claramente por debajo de 1,0, en coherencia con la interpretación de que la Experiencia aporta contenido informativo que no está incorporado en la variable

1.6. Aspectos concretos de discriminación salarial

Una vez obtenida la evidencia anterior acerca de la existencia de discriminación salarial, podríamos profundizar algo más, en el sentido de preguntarnos si la discriminación salarial en contra de la mujer tiene carácter general, que es lo que hemos supuesto hasta ahora, o alguna forma específica. Concretamente, con la información disponible, podríamos contrastar si la mujer trabajadora es discriminada al reconocer en términos salariales, bien la experiencia profesional, o bien el nivel educativo del trabajador. Para ello necesitamos definir nuevas variables, mediante el producto de la variable ficticia por cada una de las mencionadas. Por ejemplo, en la regresión,

$$\text{Salario}_i = \beta_0 + \beta_1 \text{Educación}_i + \beta_2 (\text{Educación}_i \cdot \text{Male}_i) + u_i \quad (1.5)$$

la variable producto *Educación_i · Male_i* toma un valor igual a cero para las mujeres incluidas en la muestra, mientras que coincide con la variable *Educación_i* en el caso de los hombres. Por tanto el modelo anterior equivale a los dos modelos,

$$\text{Salario}_i = \beta_0 + \beta_1 \text{Educación}_i + u_i \text{ para las mujeres}$$

$$\text{Salario}_i = \beta_0 + (\beta_1 + \beta_2) \text{Educación}_i + u_i \text{ para los hombres}$$

en los que si $\beta_2 = 0$ ambos modelos coinciden, lo que significaría que el salario recoge el nivel de educación del trabajador en igual manera en hombres

que en mujeres, no habiendo discriminación salarial en este sentido. Así, el contraste de significación del coeficiente β_2 en el modelo (1.5) equivale a un contraste de discriminación en el reconocimiento del nivel educativo del trabajador. Este modelo, al igual que el modelo que se obtiene aplicando un tratamiento análogo a la experiencia laboral, aparecen estimados en el fichero de trabajo (*REG_W_CROSSEXPER*, *REG_W_CROSSEDU*). Estimar (1.5) equivale a considerar dos rectas de regresión con igual ordenada en el origen, β_0 , pero con una pendiente diferente para hombres y para mujeres. Será mayor la primera si $\beta_2 > 0$, siendo menor si $\beta_2 < 0$. Sin embargo, dada la evidencia ya obtenida acerca de la posible discriminación salarial en contra de las mujeres, en las regresiones citadas se ha incluido asimismo explícitamente la variable *Male*, con el objeto de captar cualquier posible evidencia de discriminación sistemática. De este modo, las regresiones estimadas tienen distinta ordenada en el origen y distinta pendiente para hombres y mujeres.

Es interesante preguntarse en cuál de los aspectos, experiencia o educación, se ve más discriminada la mujer. En ambas regresiones, la variable ficticia *MALE* y los efectos cruzados, representados por las variables producto, tienen estadísticos *t* inferiores a 2,0 en valor absoluto. Estas son situaciones que suelen producirse en el análisis aplicado, generando muchas dudas en el investigador, que podría comenzar a cuestionarse si realmente hay diferencias salariales entre hombres y mujeres. Sin embargo, no hay razón para ello: desde que hemos estimado el modelo (??), sabemos que las dos regresiones que ahora consideramos están mal especificadas, pues falta un indicador en cada una de ellas. En consecuencia, la omisión de variables explicativas relevantes hace que tanto las estimaciones numéricas de los coeficientes, como de sus desviaciones típicas, sean sesgadas.

Si, a pesar de ello, nos atenemos a las estimaciones obtenidas, estas regresiones muestran que cada año de experiencia se valora a los hombres un 37,7% más que a las mujeres (1,57/4,16), mientras que el paso de un nivel educativo al siguiente se valora en los hombres un 9,8% más que en las mujeres (5,71/58,4). Por tanto, parece haber mayor evidencia de discriminación en el reconocimiento de la experiencia profesional que en el reconocimiento del nivel educativo. En los dos casos estimamos una recta con mayor ordenada en el origen y mayor pendiente para los salarios de hombres que para los de mujeres. Esto es evidencia clara sugiriendo discriminación en contra de las mujeres.

Aunque este análisis ha sido ilustrativo, no queremos que la posible detección de evidencia sugiriendo una valoración inferior de la educación en mujeres que en hombres pueda deberse a una mala especificación de los posibles modos de

discriminación. Para ello, incluimos ahora los dos indicadores, experiencia y nivel educativo en el modelo de salarios, permitiendo que ambos coeficientes, así como la ordenada en el origen, difieran para hombres y mujeres. Así, necesitamos estimar una regresión,

$$\begin{aligned}
 \text{Salario}_i = & 42,48 + 75,54 \text{ Educación}_i + 6,77 (\text{Educación}_i \cdot \text{Male}_i) + & (1.6) \\
 & (5,72) & (6,95) \\
 & (13,2) & (0,97) \\
 & + 6,61 \text{ Experiencia}_i + 1,80 (\text{Experiencia}_i \cdot \text{Male}_i) + 1,30 \text{ Male} + u_i \\
 & (0,64) & (0,80) & (31,5) \\
 & (10,3) & (2,24) & (0,04)
 \end{aligned}$$

con los resultados que se incluyen en el archivo de trabajo (*REG_W_CROSSDOBLE*). El R^2 ajustado de 0,366, con $\hat{\sigma}_u = 142,99$, y ratio de ajuste $1 - \frac{\hat{\sigma}_u}{\sigma_y} = 0,19$. Nuevamente, estimamos que la experiencia profesional se valora en los hombres un $1,80/6,61 = 27,2\%$ más que en las mujeres, y la educación en un $6,77/75,54 = 8,9\%$ más en hombres que en mujeres. La variable ficticia *MALE* tiene una contribución reducida, como indica su coeficiente estimado, pero ello es sólo aparente, pues sus posibles efectos están recogidos asimismo a través de la variables de interacción *Experiencia_i·Male_i* y *Educación_i·Male_i*.

De acuerdo con esta estimación, un varón recibe por cada año de experiencia profesional 8,41 Bef., mientras una mujer recibe tan sólo 6,61 Bef.. Por cada salto en el nivel educativo, un hombre ve incrementado su salario en 82,31 Bef., mientras que dicho incremento es de 75,54 para la mujer. La discriminación salarial estimada entre trabajadores de distinto sexo, pero de igual nivel educativo y experiencia laboral es de 1,30 Bef., más 1,80 Bef. por el número de años de experiencia, más 6,77 Bef. por el número asignado a su nivel educativo común.

Como alternativa, si hubiéramos optado por aceptar la restricción $\beta_2 = 0$ (coeficiente de *Educación_i·Male_i*) como razonable, habríamos estimado el modelo,

$$\begin{aligned}
 \text{Salario}_i = & 23,78 + 80,12 \text{ Educación}_i + 6,76 \text{ Experiencia}_i + \\
 & (17,26) & (3,25) & (0,62) \\
 & + 1,58 (\text{Experiencia}_i \cdot \text{Male}_i) + 28,41 \text{ Male} + u_i \\
 & (0,77) & (14,85)
 \end{aligned}$$

que sugiere que hay una evidencia sistemática de discriminación que hace que, a igual nivel educativo, una mujer sin experiencia laboral reciba 28,41 Bef. menos

que un trabajador varón que asimismo carezca de experiencia laboral. Además, un varón recibe 8,34 Bef. por cada año de experiencia profesional (la suma de 8,76 y 1,58 Bef.), mientras que una mujer recibe tan sólo 6,76 Bef.. Por tanto, la discriminación entre trabajadores de igual nivel educativo pero de distinto sexo se estima en 28,41 Bef. *más* 1,58 Bef. por el número de años de experiencia profesional de ambos trabajadores. Según este último modelo, el nivel educativo tiene un efecto igual sobre el salario de hombres y de mujeres, por lo que incorpora el supuesto de que no hay discriminación en la remuneración del mismo. Sin embargo, incorpora la idea de que la Experiencia se remunera de manera diferente a hombres y a mujeres.

1.7. ¿Existe evidencia de desigual remuneración de la educación entre hombres y mujeres?

El lector debe apreciar las similitudes y diferencias entre las conclusiones numéricas alcanzadas en los dos últimos modelos: la primera, que permite la posibilidad de que la remuneración salarial del nivel educativo sea distinta en hombres y mujeres, y la segunda, que impone la restricción de que dicha remuneración es igual entre ambos grupos de trabajadores. De acuerdo con el primero de los modelos, un aumento de nivel educativo incrementa el salario de hombres en 82,31 Bef., y el de las mujeres en 75,54; según el modelo que acabamos de estimar, el incremento es de 80,12 Bef., común a hombres y mujeres que, por supuesto, está entre los dos valores que estimamos con el modelo que incorporaba este tipo de discriminación. Ambos modelos implican discriminación salarial por razón de sexo, y también que la experiencia laboral se remunera de manera diferenciada en hombres y en mujeres. El primer modelo afirma lo mismo acerca del nivel educativo, mientras que el último modelo impone igual remuneración salarial por nivel educativo en hombres que en mujeres.

Es difícil decidir cuál de los dos modelos es preferible. La última regresión tiene prácticamente el mismo R^2 y la misma desviación típica residual que la anterior. En consecuencia, la aplicación de los contrastes estadísticos habituales, basados en Normalidad del término de error, variables explicativas deterministas, etc., no permiten distinguir entre ambos modelos. En esta situación, parece preferible escoger el modelo más sencillo, y concluir que no hay evidencia en la muestra de trabajadores disponible acerca de diferencias en el reconocimiento salarial del nivel educativo entre trabajadores de ambos sexos.

Dada la similitud de estadísticos, es frecuente que el investigador concluya que

ambos modelos son idénticos. Sin embargo, esto no es completamente exacto. Una interpretación alternativa del reducido estadístico t del producto *Educación*Male* es que, aunque el nivel educativo recibe distinta valoración salarial en hombres que en mujeres, las diferencias no se miden con suficiente precisión con los datos disponibles. Esta apreciación se basaría en el hecho de que el efecto discriminatorio estimado en (??) es de un 9%, que no parece que pueda considerarse despreciable. El problema es que la desviación típica con que se estima el coeficiente es prácticamente de igual tamaño que éste, revelando que es un problema de reducida precisión (alta varianza) en las estimación, lo que conduce a un estadístico t reducido, en torno a 1,0. En definitiva, el primero de los dos modelos permite más variedad salarial y puede considerarse, en tal sentido, más informativo.

El peligro es que, por estimar tal parámetro con baja precisión, las inferencias numéricas que se obtengan sobre los salarios estén poco justificadas. Dichas estimaciones son, en algunos casos particulares,

<i>SALARIOS</i>	<i>Modelo restringido</i>			<i>Modelo no restringido</i>		
	<i>Mujeres</i>	<i>Hombres</i>	<i>Ratio</i>	<i>Mujeres</i>	<i>Hombres</i>	<i>Ratio</i>
<i>Educ = 1; Exper = 3</i>	124,2	157,3	78,9%	137,9	151,3	91,1%
<i>Educ = 1; Exper = 17</i>	218,9	274,1	79,8%	230,4	269,1	85,6%
<i>Educ = 1; Exper = 30</i>	306,7	382,5	80,2%	316,3	378,4	83,6%
<i>Educ = 4; Exper = 3</i>	364,5	397,7	91,7%	364,5	398,3	91,5%
<i>Educ = 4; Exper = 17</i>	459,2	514,5	89,3%	457,0	516,0	88,6%
<i>Educ = 4; Exper = 30</i>	547,1	622,9	87,8%	542,9	625,3	86,8%

Para niveles educativos bajos, el modelo restringido implica diferencias salariales entre hombres y mujeres bastante mayores que el modelo no restringido. Lo contrario ocurre para niveles educativos altos, en los que el modelo restringido genera menores diferencias salariales entre hombres y mujeres. Es decir, el modelo que incluye explícitamente una valoración diferente para el nivel educativo de hombres y mujeres produce una estimación de la discriminación salarial más uniforme, sin que dependa del nivel educativo de los trabajadores que se comparen. Esta característica podría hacerlo preferible, pero ha de ser en última instancia la creencia del investigador acerca de si el nivel educativo se valora igual en ambos sexos o no, lo que debe llevarle a escoger uno u otro modelo.

Otra forma de analizar esta cuestión se basa en examinar los residuos del modelo restringido, el que estimamos en último lugar. Si la remuneración de la educación fuese sistemáticamente mayor en hombres que en mujeres, esperaríamos

ver residuos mayores en hombres que en mujeres, dentro de cada nivel educativo. Ello se debe a que, al no permitir diferencias por sexo, nuestra estimación de la remuneración a la educación estaría comprendida entre los niveles percibidos por hombres y mujeres; de este modo, estaríamos infravalorando la remuneración a la educación percibida por los hombres, y sobrevalorando la que perciben las mujeres. En consecuencia, los residuos correspondientes a los varones deberían ser superiores a los de las mujeres en cada nivel educativo. Si examinamos los residuos del modelo para cada nivel educativo, obtenemos medias aritméticas de 8,6 y 3,1 para hombres y mujeres en el primer nivel educativo, 4,5 y 0,0 en el segundo, -1,6 y 11,3 en el tercero, -12,9 y -5,5 en el cuarto, y 17,9 y -6,3 en el superior. Por tanto, en este sentido no surge evidencia sistemática de discriminación en la remuneración del nivel educativo, y el modelo restringido parecería suficiente.

1.8. Discriminación salarial como cambio estructural

Antes hemos planteado el contraste de discriminación a través del contraste de significación de un determinado coeficiente o conjunto de coeficientes del modelo. Otra manera de plantearlo sería a través de la *estabilidad* del modelo de determinación de salarios entre hombres y mujeres. Al igual que cuando examinamos la estabilidad temporal, se trataría, en definitiva, de dividir la muestra en dos submuestras, y comparar las estimaciones obtenidas en cada submuestra, tanto entre ellas, como con la estimación obtenida con la muestra completa. Si hay alguna variación entre los modelos de salarios estimados para hombres y mujeres, diremos también que hay cambio estructural en el mecanismo de determinación salarial, puesto que las respuestas a los determinantes del salario serían en tal caso distintos en ambos grupos de trabajadores.

Así, limitándonos por simplicidad al análisis de discriminación sistemática, podríamos estimar el modelo utilizando la submuestra de hombres en un caso⁸, y la submuestra de mujeres, en otro, obteniendo:

$$\begin{aligned} \text{Salario}_i &= 42,48 + 75,54\text{Educación}_i + 6,61\text{Experiencia}_i + u_i, \text{ para mujeres} \\ R^2 &= 0,365, \bar{R}^2 = 0,363, \hat{\sigma}_u = 153,03, SR = 20840842; \end{aligned}$$

$$\begin{aligned} \text{Salario}_i &= 43,78 + 82,31\text{Educación}_i + 8,41\text{Experiencia}_i + u_i, \text{ para hombres} \\ R^2 &= 0,331, \bar{R}^2 = 0,328, \hat{\sigma}_u = 125,93, SR = 9134157; \end{aligned}$$

⁸Introducir "1 1472 IF MALE=0" en la ventana "Sample" para estimar con observaciones de mujeres y "1 1472 IF MALE=1" para estimar con datos de trabajadores varones.

Por supuesto, que estas regresiones son comparables a la estimación del modelo (1.6). De hecho, el lector debe comprobar que de dicho modelo se deducen dos relaciones, una válida para hombres y otra para mujeres, y que *coinciden exactamente* con las dos regresiones que acabamos de estimar.

El contraste de cambio estructural se basa en la comparación de las Sumas Residuales de los modelos restringido y sin restringir. Las restricciones en este caso consisten en el supuesto de que los coeficientes del modelo de salarios son iguales para hombres y mujeres; en tal caso, el modelo sería *estable* y concluiríamos que no hay evidencia de cambio estructural. El Modelo Sin Restringir está formado por las dos regresiones anteriores, mientras que el Modelo Restringido es (1.3). El estadístico tipo- F se construye, en este caso,

$$F_{q,gdl_{MSR}} = \frac{(SRR - SRS)/q}{SRS/gdl_{MSR}} = \frac{(31079583 - (20840842 + 9134157))/3}{(20840842 + 9134157)/(N_h + N_m - 6)} = 18,0$$

donde hemos utilizado que el número de restricciones es 3, el número de coeficientes que se supone igual en ambas submuestras. El modelo restringido impone la igualdad de coeficientes para hombres y mujeres, por lo que consiste en estimar una única regresión con todos los datos; es la ecuación (1.3) y genera, por tanto, una suma residual restringida $SRR = 31079583$. El modelo sin restringir permite distintos coeficientes para hombres y mujeres; consiste en tratar las observaciones de ambas submuestras como independientes, estimando una regresión para cada una de ellas, como hemos hecho en (??) y (??). La Suma Residual de dicho modelo es el agregado de las Sumas Residuales de cada una de las dos regresiones, para hombres y mujeres. El número de grados de libertad de dicho modelo es igual a la suma de los grados de libertad de las dos regresiones: número de observaciones correspondientes a hombres, menos 3, más el número de observaciones correspondientes a mujeres, menos 3, $N_h + N_m - 6$.

El valor numérico del estadístico F está claramente por encima de los valores críticos de la distribución de probabilidad $F_{3,1466}$ a los niveles de significación habituales, 1%, 5%, 10%, por lo que rechazamos la hipótesis nula a cualquiera de dichos niveles. La hipótesis nula especifica la igualdad de coeficientes entre los modelos de salarios de hombres y mujeres, $H_0 : \beta_h = \beta_m$, por lo que concluiríamos que los modelos de salarios son diferentes. Hay que notar, sin embargo, que el estadístico utilizado sólo tendría distribución F si el término de error del modelo de salarios tuviera distribución Normal, lo que ya hemos comentado que parece altamente improbable, dado el histograma de frecuencias de los salarios.

Por sí sólo, este contraste no dice nada acerca del sentido en que se producen las diferencias, por lo que sería difícil concluir de él nada relativo a la discriminación salarial. Sin embargo, el hecho de que los coeficientes asociados tanto a nivel de educación como a la experiencia laboral sean mayores para los hombres que para las mujeres sugiere que las diferencias son en perjuicio de las mujeres. Como los términos constantes estimados son muy similares, es fácil ver que entre dos trabajadores de distinto sexo, pero de igual nivel educativo y experiencia laboral, el hombre recibe, generalmente, un salario superior al de la mujer.

2. Especificaciones con variables ficticias: contrastes de homogeneidad salarial entre grupos de trabajadores

En esta segunda parte del ejercicio, vamos a ilustrar el modo en que pueden utilizarse variables ficticias para proponer distintos grados de homogeneidad en el mecanismo de determinación salarial. Trabajando con la misma base de datos, continuamos utilizando el nivel educativo y el grado de experiencia laboral como posibles determinantes salariales. La hipótesis que ahora consideramos es que la experiencia laboral se remunera de igual modo en todos los niveles educativos. Dada la evidencia ya presentada acerca de la existencia de discriminación salarial por razón de sexo, utilizamos inicialmente las observaciones procedentes de trabajadores varones, para centrarnos exclusivamente en analizar las diferencias que puedan provenir de los dos factores citados.

2.1. Homogeneidad del modelo de salarios para distintos niveles educativos

Comentábamos al inicio de este ejercicio cómo las diferencias entre cada dos niveles sucesivos de educación pueden ser muy distintas, dependiendo de los niveles educativos que se comparen. Esto no ha sido recogido en nuestro análisis hasta ahora, porque la definición que se ha hecho de la variable educación conduce a que estimemos un incremento salarial con cada cambio de nivel educativo, con independencia de los niveles en los que se produzca. Para analizar esta cuestión en más detalle, estimamos por separado la contribución media de cada nivel educativo a la retribución salarial. Como ya sabemos que existe discriminación salarial por sexos, *vamos a utilizar únicamente las observaciones correspondientes a los hombres.*

Para ello, estimamos cinco regresiones del tipo (1.2), utilizando en cada caso datos de varones de un mismo nivel educativo⁹. Los resultados son,

$$\begin{aligned}
 \textit{nivel 1} & : \textit{Salario}_i = 318,03 + 1,67\textit{Experiencia}_i + \hat{u}_i, & (2.1) \\
 N_i & = 76, R^2 = 0,043, \hat{\sigma}_u = 67,14, SR = 333565.4; \\
 \textit{nivel 2} & : \textit{Salario}_i = 275,19 + 5,48\textit{Experiencia}_i + \hat{u}_i, \\
 N_i & = 195, R^2 = 0,212, \hat{\sigma}_u = 107,83,14, SR = 2244014; \\
 \textit{nivel 3} & : \textit{Salario}_i = 312,10 + 6,63\textit{Experiencia}_i + \hat{u}_i, \\
 N_i & = 258, R^2 = 0,218, \hat{\sigma}_u = 125,16, SR = 4010484; \\
 \textit{nivel 4} & : \textit{Salario}_i = 323,86 + 10,56\textit{Experiencia}_i + \hat{u}_i, \\
 N_i & = 164, R^2 = 0,374, \hat{\sigma}_u = 133,71, SR = 2896499; \\
 \textit{nivel 5} & : \textit{Salario}_i = 389,43 + 13,46\textit{Experiencia}_i + \hat{u}_i, \\
 N_i & = 200, R^2 = 0,257, \hat{\sigma}_u = 226,56, SR = 10162534;
 \end{aligned}$$

en los que se aprecia un aumento en el coeficiente estimado para la variable Experiencia, según aumenta el nivel educativo. Ello sugiere que el reconocimiento salarial de la experiencia profesional entre varones es mayor cuanto más alto sea su nivel educativo, quizá por ser entonces la experiencia laboral de mayor calidad, un resultado sin duda interesante.

Otro resultado que surge de este modelo estimado es que el salario para trabajadores sin experiencia laboral es creciente con el nivel educativo, excepto entre los dos primeros niveles. Esta comparación no es, sin embargo, la más interesante, por cuanto que apenas hay trabajadores sin experiencia laboral. Otra manera de interpretar el modelo consiste en acudir al promedio de la experiencia laboral, que es de 17,22 años para toda la muestra, pero es¹⁰ de 26,57 años para los hombres de nivel educativo 1, siendo de 20,42 años, 18,28, 16,19, y 15,85 años para los restantes niveles educativos. Por tanto, el salario medio para los trabajadores del primer nivel educativo se estima en $318.03 + (1.67)(26.57) = 362.4$, siendo para los sucesivos niveles: $275.19 + (5.48)(20.42) = 387.09$, $312.10 + (6.63)(18.28) = 433.3$, $323.86 + (10.56)(16.19) = 494.83$, $389.43 + (13.46)(15.85) = 602.77$.

⁹Al estimar la regresión, introducir en la ventana "Sample", el mensaje "1 1472 IF MALE=1 AND EDUC=1", e ir variando el código asignado a EDUC, de 1 a 5.

¹⁰Tras marcar la variable Experiencia en el archivo de trabajo, entrar en "View/Descriptive Statistics/Statistics by Classification" y escribir en la ventana "Series/Group for Classify": EDUC*MALE.

Así, en promedio, un trabajador varón del nivel educativo 2 recibe un salario superior en 24.7 Bef. al del nivel educativo 1. Las remuneraciones promedio asignadas a los cambios sucesivos en nivel educativo son: 46,2 Bef. entre los trabajadores de niveles educativos 2 y 3; 61,5 Bef. para el salto de niveles educativos 3 a 4, y 108 Bef. para el paso de nivel educativo 4 a nivel 5.

Como se ve, estimamos una remuneración creciente para el salto entre cada par de niveles educativos sucesivos, valorándose más un aumento de nivel educativo cuanto más alto sea el nivel educativo de partida. Esto hace que el modelo (??) sea excesivamente restringido; en él, estimábamos en 82,31 Bef. la valoración de cada nivel educativo adicional, con independencia del nivel de partida. Tal estimación debe verse como un promedio de las cuatro remuneraciones que calculamos a partir de (2.1), pero es inapropiada, dados los resultados de este último modelo.

Un modelo algo menos resringido que (??) sería,

$$Salario_i = \beta_0 + \beta_1 Educaci3n_i + \beta_2 Experiencia_i + \beta_3 (Educaci3n_i \cdot Experiencia_i) + u_i, \text{ para hombres} \quad (2.2)$$

Este modelo genera, para los distintos niveles educativos,

$$\begin{aligned} \text{Nivel educativo 1} & : \quad Salario_i = (\beta_0 + \beta_1) + (\beta_2 + \beta_3) \cdot Experiencia_i + u_i, \\ \text{Nivel educativo 2} & : \quad Salario_i = (\beta_0 + 2\beta_1) + (\beta_2 + 2\beta_3) \cdot Experiencia_i + u_i, \\ \text{Nivel educativo 3} & : \quad Salario_i = (\beta_0 + 3\beta_1) + (\beta_2 + 3\beta_3) \cdot Experiencia_i + u_i, \\ \text{Nivel educativo 4} & : \quad Salario_i = (\beta_0 + 4\beta_1) + (\beta_2 + 4\beta_3) \cdot Experiencia_i + u_i, \\ \text{Nivel educativo 5} & : \quad Salario_i = (\beta_0 + 5\beta_1) + (\beta_2 + 5\beta_3) \cdot Experiencia_i + u_i, \end{aligned}$$

que impone sobre (2.1) dos tipos de restricciones: *a*) que la diferencia en la remuneración que recibe cada año de experiencia en trabajadores de dos niveles educativos sucesivos es la misma, β_3 , independiente de los niveles de educación considerados, y *b*) que la diferencia salarial entre trabajadores de igual experiencia y niveles de educación consecutivos es siempre la misma, β_1 .

Como consecuencia, este modelo implica que para caracterizar las diferencias salariales entre trabajadores de igual experiencia laboral sólo importa la diferencia que exista entre sus niveles educativos, pero no cuáles sean estos. Si k denota la diferencia entre los niveles educativos de dos trabajadores de igual experiencia, donde k podría ser igual a 0, 1, 2, 3 ó 4, la diferencia entre sus salarios sería: $k\beta_1 + k\beta_3 Experiencia$, siendo $Experiencia$ el número de años de experiencia de

ambos trabajadores.¹¹ El lector debe asegurarse de que entiende que en (2.1) no se ha impuesto ninguna de estas dos restricciones.

Por supuesto que estas restricciones pueden contrastarse conjuntamente utilizando los estadísticos habituales, sin más que considerar a (2.1) como Modelo Sin Restringir, y a (2.2) como Modelo Restringido.

Un modelo más restrictivo consideraría que la remuneración salarial a cada año de experiencia laboral del trabajador es independiente de su nivel educativo. Dicho modelo sería,

$$Salario_i = \beta_0 + \delta_2 D_{2i} + \delta_3 D_{3i} + \delta_4 D_{4i} + \delta_5 D_{5i} + \beta_1 Experiencia_i + u_i, \quad (2.3)$$

donde la variable ficticia D_{2i} se define mediante $D_{2i} = 1$ si la observación i -ésima se refiere a un trabajador varón en el segundo nivel educativo, y $D_{2i} = 0$ en todos los demás casos. El resto de las variables ficticias se define de manera análoga. Una vez estimado este modelo tendríamos para los varones del primer nivel educativo, $Salario_i = \hat{\beta}_0 + \hat{\beta}_1 Experiencia_i + \hat{u}_i$, para los del segundo nivel educativo: $Salario_i = (\hat{\beta}_0 + \hat{\delta}_2) + \hat{\beta}_1 Experiencia_i + \hat{u}_i$, y así sucesivamente; por ejemplo, para los trabajadores varones del más alto nivel educativo, tendríamos, $Salario_i = (\hat{\beta}_0 + \hat{\delta}_5) + \hat{\beta}_1 Experiencia_i + \hat{u}_i$. Al estimar δ_2 obtenemos el diferencial salarial que reciben los trabajadores varones del segundo nivel educativo respecto de los del primero, con independencia de su experiencia laboral. Los restantes coeficientes $\delta_3, \delta_4, \delta_5$, se interpretan de manera análoga, por lo que esperaríamos que fueran todos ellos positivos.

La estimación del modelo conduce a,

$$Salario_i = 133,19 + 73,12 D_{2i} + 142,27 D_{3i} + 208,54 D_{4i} + 313,62 D_{5i} + 8,00 Experiencia_i + u_i, \quad (2.4)$$

A diferencia de los modelos (2.2) y (2.1), la diferencia salarial entre trabajadores de distinto nivel educativo pero que tienen igual experiencia, se supone ahora independiente de dicho nivel de experiencia. Por tanto, (2.3) es un modelo más restringido que los dos anteriores. En (2.3) tenemos cinco regresiones paralelas, con distinta ordenada en el origen pero igual pendiente. Por el contrario, (2.1) genera cinco rectas de regresión con distinta ordenada en el origen y diferente pendiente, es decir, cinco rectas completamente distintas.

¹¹¿Cuál sería la diferencia en salarios si no tuvieran el mismo grado de experiencia?.

En (2.2) permitimos que la remuneración a la experiencia varíe con el nivel educativo, lo cual es más general que (2.3). Es algo más restrictivo en cuando que hace que las diferencias en la ordenada en el origen sean iguales entre niveles educativos. Los modelos (2.3) y (2.2) no son directamente comparables, pues uno no puede obtenerse imponiendo restricciones sobre el otro.

El modelo alternativo,

$$\text{Salario}_i = \beta_0 + \beta_2 \text{Experiencia}_i + \beta_3 (\text{Educación}_i \cdot \text{Experiencia}_i) + u_i,$$

no es muy interesante, pues si se piensa que puede haber distinta remuneración salarial a la experiencia dependiendo del nivel educativo, es aún más probable que haya diferencias entre trabajadores de igual experiencia, pero distinto nivel educativo. En consecuencia, los modelos (2.3) y (2.2) son generalmente preferibles.

Un modelo aún más restrictivo impondría coeficientes comunes a todos los niveles educativos,

$$\begin{aligned} \text{Todos los varones} \quad : \quad \text{Salario}_i &= 360,25 + 5,73 \text{Experiencia}_i + \hat{u}_i, & (2.5) \\ N_i &= 893, R^2 = 0,093, \hat{\sigma}_u = 182,65, SR = 29723552; \end{aligned}$$

Este modelo equivale a imponer las restricciones $H_0 : \delta_2 = \delta_3 = \delta_4 = \delta_5 = 0$ en (2.3), o bien $H_0 : \beta_1 = \beta_3 = 0$ en (2.2), o $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5; \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5$, en (2.1), si denotamos por α_i los términos independientes y por β_i las pendientes de cada ecuación en este último modelo. Este último es un conjunto de 8 restricciones, el número de igualdades que se incluyen en la hipótesis nula. Cada uno de estos conjuntos de hipótesis puede contrastarse comparando las Sumas Residuales de los Modelos Restringido y Sin Restringir en cada caso, utilizando en el cálculo del estadístico tipo F el número de restricciones¹² y el número de grados de libertad del Modelo Sin Restringir: número de observaciones utilizadas en la estimación, menos número de coeficientes estimado en dicho modelo.

El modelo (??) queda en un terreno intermedio entre los anteriores: se obtiene a partir de (2.1) imponiendo las 7 restricciones $H_0 : \alpha_2 - \alpha_1 = \alpha_3 - \alpha_2 = \alpha_4 - \alpha_3 = \alpha_5 - \alpha_4; \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5$. Se obtiene asimismo a partir de (2.2), imponiendo la restricción $H_0 : \beta_3 = 0$, a partir de (2.3) imponiendo la restricción $H_0 : \delta_2 = \delta_3 = \delta_4 = \delta_5 = 0$. Por otra parte, el modelo (2.5) se obtiene a partir de

¹²Es decir, el número de igualdades utilizada para caracterizar la hipótesis nula.

(??) imponiendo la restricción $H_0 : \beta_1 = 0$. Por tanto, todas estos conjuntos de restricciones pueden contrastarse comparando las Sumas Residuales apropiadas, mediante el habitual estadístico tipo F en el que habrá que utilizar asimismo la información relativa al número de restricciones que se contrastan y el número de grados de libertad del Modelo Sin Restringir.

A modo de ejemplo, consideremos el contraste de las restricciones $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5; \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5$, sobre el modelo (2.1), que nos llevaría al modelo (2.5) como versión restringida del anterior. El coeficiente estimado para la variable Experiencia en este último modelo es un promedio de los obtenidos en las cinco regresiones que configuran el modelo (2.1), y lo mismo ocurre con la ordenada en el origen de la recta de regresión. El Modelo Sin Restringir es el conjunto de las cinco regresiones de (2.1), por lo que la Suma Residual Sin restringir es el agregado de las sumas residuales en ellas, mientras que el Modelo Restringido es el constituido por la última regresión, teniendo, por tanto el estadístico F ,

$$\begin{aligned} F_{q,gdl_{MSR}} &= \frac{(SRR - SRS) / q}{SRS / gdl_{MSR}} = \\ &= \frac{(29723552 - (333565.4 + 2244014 + 4010484 + 2896499 + 10162534)) / 8}{(333565.4 + 2244014 + 4010484 + 2896499 + 10162534) / (893 - 10)} = 56,6 \end{aligned}$$

siendo 10 el número de coeficientes estimados en el Modelo Sin Restringir: dos coeficientes en cada una de las cinco regresiones. Se contrastan 8 restricciones, pues los dos coeficientes del modelo se hacen iguales en cuatro de las regresiones, a lo que ocurra en una de ellas. El estadístico F conduce a un rechazo tan claro de la hipótesis nula de igualdad de la regresión de salarios para los distintos niveles educativos, que, incluso si la distribución de probabilidad del término de error se desvía de la Normal, la evidencia obtenida contra la hipótesis de estabilidad de la regresión ha de juzgarse como muy clara. En consecuencia, el modelo (2.5) es inapropiado, por ser excesivamente restringido.

2.2. Variables ficticias y colinealidad perfecta

Alternativamente al modelo (??), podríamos haber especificado,

$$Salario_i = \delta_1 D_{1i} + \delta_2 D_{2i} + \delta_3 D_{3i} + \delta_4 D_{4i} + \delta_5 D_{5i} + \beta_1 Experiencia_i + u_i, \quad (2.6)$$

tras definir una variable ficticia D_{1i} del mismo modo que definimos las restantes. En este caso, para los varones del primer nivel educativo tendríamos, $Salario_i = \hat{\delta}_1 + \hat{\beta}_1 Experiencia_i + \hat{u}_i$, para los del segundo nivel educativo: $Salario_i = \hat{\delta}_2 + \hat{\beta}_1 Experiencia_i + \hat{u}_i$, y así sucesivamente; para los trabajadores varones del más alto nivel educativo, tendríamos, $Salario_i = \hat{\delta}_5 + \hat{\beta}_1 Experiencia_i + \hat{u}_i$. Por supuesto, los valores numéricos de los coeficientes δ serían diferentes ahora que en el modelo anterior. Esperaríamos que las estimaciones numéricas de los coeficientes δ fuesen crecientes para los distintos niveles educativos.

La suma de las cinco variables ficticias incluidas en el modelo (2.6) es igual a uno para todas las observaciones, pues sólo una de ellas es igual a uno en cada observación, siendo las restantes iguales a cero, y esto ocurre para todas las observaciones disponibles. Por tanto, su suma es igual al valor de la variable que acompaña al término constante, por lo que éste no puede incluirse en la regresión, pues tendríamos colinealidad perfecta, no pudiendo estimarse dicho modelo. En el caso de (2.3) las cuatro variables ficticias suman uno para todas las observaciones, excepto las del primer nivel educativo, para el que suman cero; por tanto, su suma no coincide con el valor numérico de la variable que acompaña al término constante, y el modelo puede estimarse. En dicho modelo, podíamos haber optado por incluir D_{1i} y excluir otra cualquiera de las variables ficticias, y la interpretación de los coeficientes estimados sería análoga a la que propusimos para el modelo (2.3).

3. Ejercicios

Exercise 3.1. *Estimar el modelo de salarios con la única variable continua, la variable salario, en logaritmos. Interpretar los coeficientes estimados.*

Exercise 3.2. *Estimar el modelo (2.3) e interpretar los resultados obtenidos. Representar gráficamente las rectas de regresión que se deducen para cada nivel educativo. Contrastar la hipótesis de igualdad de la regresión de salarios para los cinco niveles educativos. Estimar una variante de (2.3) incluyendo D_{1i} y excluyendo otra de las variables ficticias, y probar que los resultados son idénticos a los anteriores.*

Exercise 3.3. *Estimar el modelo (2.6) e interpretar los resultados obtenidos. Representar gráficamente las rectas de regresión que se deducen para cada nivel educativo. Contrastar la hipótesis de igualdad de la regresión de salarios para los cinco niveles educativos.*

Exercise 3.4. *Estimar el modelo*

$$\text{Salario}_i = \beta_0 + \beta_1 \text{Educación}_i + \beta_2 \text{Experiencia}_i + u_i,$$

con datos de trabajadores varones, y contrastar la hipótesis de que el incremento salarial entre trabajadores de niveles educativos sucesivos es el mismo con independencia de los niveles educativos comparados, pero manteniendo la hipótesis de que la valoración de la experiencia laboral es la misma para todos los niveles educativos.

Exercise 3.5. *Repetir el Ejercicio (3.2) con datos de mujeres trabajadoras.*

Exercise 3.6. *Repetir el Ejercicio (3.3) con datos de mujeres trabajadoras.*

Exercise 3.7. *Repetir el Ejercicio (3.4) con datos de mujeres trabajadoras.*

Exercise 3.8. *Estime el modelo (1.1), obteniendo el incremento salarial estimado al pasar de cada nivel educativo al nivel inmediatamente superior. Deduzca de ello el incremento porcentual en salario estimado al pasar de un nivel educativo al superior. Vuelva a estimar el modelo, con la variable salario en logaritmos, obteniendo el incremento porcentual, constante, que se produce en el salario al pasar de un nivel educativo al superior. Deduzca de ello el incremento absoluto en el salario estimado al pasar de un nivel educativo al superior.*