



1. En el archivo Salarios.csv se encuentran los años de educación (EDU), el salario por hora (SAL, en dólares), los años de experiencia (EXP), la habilidad (HAB), la educación de la madre (EDM), la educación del padre (EDP) y el número de hermanos (HER) de cada uno de 2178 trabajadores, y de cada uno se incluye un número de identificación (ID). Se quiere ver cómo influyen todos esos factores dentro del salario del trabajador.

a) Utilizando Microsoft EXCEL:

- 1) Encuentra los valores de los parámetros correspondientes a un modelo de regresión múltiple e interprétalos.

Solución: Utilizando ESTIMACION.LINEAL() en Microsoft EXCEL obtuvimos los siguientes parámetros del modelo clásico

β_6	β_5	β_4	β_3	β_2	β_1	β_0
-0.004867019	0.003678772	0.04669123	0.13838842	0.277985734	0.700834955	-2.078530524

Del coeficiente β_1 que corresponde a los años de educación podemos decir que por cada año de estudio el salario por hora de los empleados aumenta 0.7 dolares, β_2 que corresponde a los años de experiencia podemos decir que por cada año de experiencia el salario por hora de los empleados aumenta 0.27 dolares, β_3 que corresponde a la habilidad podemos decir que por cada unidad de habilidad el salario por hora de los empleados aumenta 0.13 dolares, β_4 que corresponde a los años de educación de la madre podemos decir que por cada año de educación de la madre el salario por hora de los empleados aumenta 0.046 dolares, β_5 que corresponde a los años de educación del padre podemos decir que por cada año de educación del padre el salario por hora de los empleados aumenta 0.0036 dolares y β_6 que corresponde al número de hermanos podemos decir que por cada hermano el salario por hora de los empleados disminuye -0.0049 dolares.

□

- 2) ¿Puede decirse que la regresión anterior es significativa, a un nivel de significancia de 0.0125?

Solución: La función ESTIMACION.LINEAL() también nos da el valor del estadístico F el cual es 41.47304516 ahora, el valor del umbral con un nivel de significancia 0.0125 lo obtenemos utilizando la función INV.F(0.9875,6,2171) con lo que obtenemos 2.715066188 y por lo tanto, la prueba es significativa.

□

- 3) Construye un intervalo de confianza del 94 % para el verdadero valor del parámetro asociado a los años de experiencia.

Solución: Sabemos que

$$\frac{\hat{\beta}_j - \beta_j}{\sqrt{s^2 (\mathbf{X}^T \mathbf{X})_{jj}^{-1}}} \sim t(n - h)$$

entonces un intervalo de confianza para β es de la forma

$$\hat{\beta}_j - t(n - h) \cdot \sqrt{s^2 (\mathbf{X}^T \mathbf{X})_{jj}^{-1}} \leq \beta \leq \hat{\beta}_j + t(n - h) \cdot \sqrt{s^2 (\mathbf{X}^T \mathbf{X})_{jj}^{-1}}$$

por lo tanto usando INV.T(0.03,2171) y la desviación estándar que obtuvimos en Excel tenemos que un intervalo de confianza del 94 % para los años de experiencia es

$$0.19976158 \leq \beta \leq 0.356209887$$

□

- 4) Construye un intervalo de confianza del 98 % para el verdadero valor de la varianza de la perturbación asociada al modelo.

Solución: De Excel tenemos que el $SSE = 43733.2751$ entonces sabemos que $s^2 = \frac{\varepsilon^T \varepsilon}{2171} = 20.14429991$

$$\frac{2171 \cdot s^2}{\sigma^2} = \frac{SSE}{\sigma^2} \sim \chi^2(2171)$$

entonces un intervalo de confianza para la varianza es de la forma

$$\frac{SSE}{\chi^2(2171)} \leq \sigma^2 \leq \frac{SSE}{\chi^2(2171)}$$

por lo que el intervalo de confianza queda de la forma

$$18.79201237 \leq \sigma^2 \leq 21.64311598$$

□

- 5) Encuentra el salario por hora esperado de un trabajador con 2 hermanos, 10 años de experiencia, factor de habilidad de 1.95, 21 años de estudio, cuyo padre sólo tiene 6 años de estudios y con una madre con 15 años de educación.

Solución: Usando los coeficientes de la regresión tenemos

$$\begin{aligned}\hat{y} &= -0.004867019 \cdot 2 + 0.003678772 \cdot 6 + 0.04669123 \cdot 15 + 0.13838842 \cdot 1.95 + 0.277985734 \cdot 10 + 0.700834955 \cdot 21 \\ &\quad - 2.078530524 \\ &= 16.42089341\end{aligned}$$

por lo que se espera que una persona con estas características tenga un salario de 16.421 dolares por hora.

□

b) Repetir todos los incisos anteriores usando Stata.

- 1) *Solución:* Usando `regress sal edu exp hab edm edp her` en Stata obtuvimos los siguientes parámetros del modelo clásico

```
. regress sal edu exp hab edm edp her
```

Source	SS	df	MS	Number of obs	=	2,178
Model	5017.25748	6	836.20958	F(6, 2171)	=	41.47
Residual	43773.2751	2,171	20.1627246	Prob > F	=	0.0000
				R-squared	=	0.1028
				Adj R-squared	=	0.1004
Total	48790.5326	2,177	22.4118202	Root MSE	=	4.4903

sal	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
edu	.700835	.0578005	12.13	0.000	.5874848 .8141851
exp	.2779857	.0415693	6.69	0.000	.1964659 .3595055
hab	.1383884	.1292814	1.07	0.285	-.1151398 .3919166
edm	.0466912	.044342	1.05	0.292	-.040266 .1336485
edp	.0036788	.0356946	0.10	0.918	-.0663203 .0736779
her	-.004867	.0480496	-0.10	0.919	-.099095 .089361
_cons	-2.078531	.8338848	-2.49	0.013	-3.713826 -.4432347

Figura 1: Datos que arroja Stata con el comando `regress sal edu exp hab edm edp her`

Del coeficiente EDU que corresponde a los años de educación podemos decir que por cada año de estudio el salario por hora de los empleados aumenta 0.7 dolares, EXP que corresponde a los años de experiencia podemos decir que por cada año de experiencia el salario por hora de los empleados aumenta 0.27 dolares, HAB que corresponde a la habilidad podemos decir que por cada unidad de habilidad el salario por hora de los empleados aumenta 0.13 dolares, EDM que corresponde a los años de educación de la madre podemos decir que por cada año de educación de la madre el salario por hora de los empleados aumenta 0.046 dolares, EDP que corresponde a los años de educación del padre podemos decir que por cada año de educación del padre el salario por hora de los empleados aumenta 0.0036 dolares y HER que corresponde al número de hermanos podemos decir que por cada hermano el salario por hora de los empleados disminuye -0.0049 dolares.

□

- 2) *Solución:* Como podemos ver en la captura anterior, vemos que la $\text{Prob}>F = 0.000$ esto nos indica que la regresión es en efecto significativa ya que el valor p es menor a 0.0125.
- 3) *Solución:* Utilizando el comando `regress sal edu exp hab edm edp her, level(94)` obtenemos una regresión lineal con un nivel de confianza del 94 % lo que a sus vez nos proporciona intervalos de confianza para cada uno de los coeficientes de la regresión

□

```
. regress sal edu exp hab edm edp her, level(94)
```

Source	SS	df	MS	Number of obs	=	2,178
Model	5017.25748	6	836.20958	F(6, 2171)	=	41.47
Residual	43773.2751	2,171	20.1627246	Prob > F	=	0.0000
				R-squared	=	0.1028
				Adj R-squared	=	0.1004
Total	48790.5326	2,177	22.4118202	Root MSE	=	4.4903

sal	Coef.	Std. Err.	t	P> t	[94% Conf. Interval]
edu	.700835	.0578005	12.13	0.000	.5920673 .8096026
exp	.2779857	.0415693	6.69	0.000	.1997616 .3562099
hab	.1383884	.1292814	1.07	0.285	-.1048903 .3816671
edm	.0466912	.044342	1.05	0.292	-.0367506 .1301331
edp	.0036788	.0356946	0.10	0.918	-.0634905 .070848
her	-.004867	.0480496	-0.10	0.919	-.0952856 .0855516
_cons	-2.078531	.8338848	-2.49	0.013	-3.647716 -.5093455

Figura 2: Datos que arroja Stata con el comando `regress sal edu exp hab edm edp her, level(94)`

por lo tanto tenemos que un intervalo de confianza para el parámetro asociado a los años de experiencia es

$$0.1997616 \leq \beta_1 \leq 0.3562099$$

□

4)

5)

c) Repetir todos los incisos anteriores usando R.

- 1) *Solución:* Usando `rl<-lm(SAL ~EDU+EXP+HAB+EDM+EDP+HER,data = salarios)` en R y despues `summary(rl)` obtuvimos los siguientes parámetros del modelo clásico

```
Call:
lm(formula = SAL ~ EDU + EXP + HAB + EDM + EDP + HER, data = salarios)

Residuals:
    Min       1Q   Median       3Q      Max
-9.179 -2.408 -0.728  1.597  72.259

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.078531   0.833885  -2.493   0.0128 *
EDU           0.700835   0.057801  12.125 < 2e-16 ***
EXP           0.277986   0.041569   6.687 2.88e-11 ***
HAB           0.138388   0.129281   1.070   0.2845
EDM           0.046691   0.044342   1.053   0.2925
EDP           0.003679   0.035695   0.103   0.9179
HER          -0.004867   0.048050  -0.101   0.9193
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.49 on 2171 degrees of freedom
Multiple R-squared:  0.1028,    Adjusted R-squared:  0.1004
F-statistic: 41.47 on 6 and 2171 DF,  p-value: < 2.2e-16
```

Figura 3: Datos que arroja R con el comando `summary(rl)`

Del coeficiente EDU que corresponde a los años de educación podemos decir que por cada año de estudio el salario por hora de los empleados aumenta 0.7 dolares, EXP que corresponde a los años de experiencia podemos decir que por cada año de experiencia el salario por hora de los empleados aumenta 0.27 dolares, HAB que corresponde a la habilidad podemos decir que por cada unidad de habilidad el salario por hora de los empleados aumenta 0.13 dolares, EDM que corresponde a los años de educación de la madre podemos decir que por cada año de educación de la madre el salario por hora de los empleados aumenta 0.046 dolares, EDP que corresponde a los años de educación del padre podemos decir que por cada año de educación del padre el salario por hora de los empleados aumenta 0.0036 dolares y HER que corresponde al número de hermanos podemos decir que por cada hermano el salario por hora de los empleados disminuye -0.0049 dolares.

□

- 2) *Solución:* Como podemos ver en la captura anterior, vemos que $p\text{-value} < 2.2e-16$ por lo tanto a un nivel de significancia del 0.0125 tenemos que el valor p es menor, por lo tanto la prueba es significativa.
- 3) *Solución:* Utilizando el comando `confint(rl,level = 0.94)` obtenemos los intervalos de confianza del 94 % para cada uno de los coeficientes de la regresión

```
> confint(rl,level = 0.94)
              3 %      97 %
(Intercept) -3.64771553 -0.50934552
EDU           0.59206728  0.80960263
EXP           0.19976158  0.35620989
HAB          -0.10489029  0.38166713
EDM          -0.03675059  0.13013305
EDP          -0.06349046  0.07084801
HER          -0.09528564  0.08555160
> |
```

Figura 4: Datos que arroja R con el comando `confint(rl,level = 0.94)`

por lo tanto tenemos que un intervalo de confianza para el parámetro asociado a los años de experiencia es

$$0.19976158 \leq \beta_1 \leq 0.35620989$$

□

- 4) *Solución:* Análogamente, como hicimos en Excel calcularemos los intervalos para la varianza. Tenemos que con el comando `sum(rl$residuals^2)` obtenemos s^2 por lo tanto si utilizamos `qchisq()` para obtener lo cuantiles por lo tanto para el intervalo inferior tenemos que usando `sum(rl$residuals^2)/qchisq(0.99,2171)` arroja 18.8092 y para el superior `sum(rl$residuals^2)/qchisq(0.01,2171)` arroja 21.66291 por lo tanto

$$18.8092 \leq \sigma^2 \leq 21.66291$$

☐

5) *Solución:* Utilizando `sum(r1$coefficients*c(1,21,10,1.95,15,6,2))` obtenemos el valor del salario por hora esperado el cual es 16.40143.

☐