1) Ejercicio 5 práctica 4: Coeficientes de regresión estandarizados.

Haremos unos cálculos en el pizarrón y analizaremos las siguientes tablas

Ecuación de regresión

La tabla de coeficientes de regresión parcial (ver tabla 18.6) contiene toda la información necesaria para construir la ecuación de regresión mínimo-cuadrática.

Tabla 18.6. Coeficientes de regresión parcial.

Modelo: 1

	Coeficientes no estandarizados		Coeficientes estandarizados		
	B	Error típ.	Beta	t	Sig.
(Constante)	-3661,517	1935,490		-1,892	,059
Salario inicial	1,749	,060	,806	29,198	,000
Experiencia previa (meses)	-16,730	3,605	-,102	-4,641	,000
Nivel educativo (años)	735,956	168,689	,124	4,363	,000

En la columna encabezada *Coeficientes no estandarizados* se encuentran los coeficientes (B_k) que forman parte de la ecuación en puntuaciones directas:

$$\textit{Pronóstico en salario} = $$
$$= -3.661{,}517 + 1{,}749\ \textit{salini} - 16{,}730\ \textit{expprev} + 735{,}956\ \textit{educ}$$

Estos coeficientes no estandarizados se interpretan en los términos ya conocidos. Por ejemplo, el coeficiente correspondiente a la variable *salini*, que vale 1,749, indica que, si el resto de variables se mantienen constantes, a un aumento de una unidad (un dólar) en *salini* le corresponde, en promedio, un aumento de 1,749 dólares en *salario*.

Es necesario señalar que estos coeficientes no son independientes entre sí. De hecho, reciben el nombre de coeficientes de regresión *parcial* porque el valor concreto estimado para cada coeficiente se ajusta teniendo en cuenta la presencia del resto de variables independientes. Conviene, por tanto, interpretarlos con cautela.

El signo del coeficiente de regresión parcial de una variable puede no ser el mismo que el del coeficiente de correlación simple entre esa variable y la dependiente. Esto es debido a los ajustes que se llevan a cabo para poder obtener la mejor ecuación posible. Aunque existen diferentes explicaciones para justificar el cambio de signo de un coeficiente de regresión, una de las que deben ser más seriamente consideradas es la que se refiere a la presencia de un alto grado de asociación entre algunas de las variables independientes (colinealidad). Trataremos esta cuestión más adelante.

Coeficientes de regresión estandarizados

Los coeficientes *Beta* están basados en las puntuaciones típicas y, por tanto, son directamente comparables entre sí. Indican la cantidad de cambio, en puntuaciones típicas, que se producirá en la variable dependiente por cada cambio de una unidad en la correspondiente variable independiente (manteniendo constantes el resto de variables independientes).

Estos coeficientes proporcionan una pista muy útil sobre la importancia relativa de cada variable independiente en la ecuación de regresión. En general, una variable tiene tanto más peso (importancia) en la ecuación de regresión cuanto mayor (en valor absoluto) es su coeficiente de regresión estandarizado. Observando los coeficientes *Beta* de la tabla 18.6 vemos que la variable *salini* es la más importante; después, *educ*; por último, *expprev*. Lo ya dicho sobre la no independencia de los coeficientes de regresión parcial no estandarizados también vale aquí.

Otro ejemplo:

Luego se corre el programa. A continuación se presentan los resultados del ajuste de la variable dependiente M1_sa_d:

VARIABLE	COEFICIENTES ESTIMADOS	
	SIN ESTANDARIZAR	ESTANDARIZADOS
TBAS	-1790.13	-0.206929
PIB_SA	2.04	0.798669
Constante	-23664.82	0.000000

Los coeficientes estandarizados o coeficientes beta indican el peso relativo de cada variable, sin importar la unidad de medida en que se encuentren expresadas. Por ejemplo, si el PIB desestacionalizado se divide por 1000 (PIB_SA_1), el coeficiente sin estandarizar se ve afectado, mientras que el coeficiente estandarizado no se modifica, como se aprecia en la siguiente tabla:

VARIABLE	COEFICIENTES ESTIMADOS	
	SIN ESTANDARIZAR	ESTANDARIZADOS
TBAS	-1790.13	-0.206929
PIB_SA_1	2039.47	0.798669
Constante	-23664.82	0

2) Ej 7 práctica 4. Pesos.

Terminaremos de discutir el tema de los pesos. ¿Qué pesos podrían andar en ese ejercicio? ¿qué habría que poner en R?

3) Puntos influyentes.

Estudiar sobre un conjunto de datos todas las medidas vistan en la teórica: LEVERAGE, DISTANCIA DE COOK, DFFIT, DFBETA.

4)

Generar 40 observaciones $(Y_i, x_{1i}, x_{2i}, x_{3i}, x_{4i})$ de forma tal que

$$Y_i = 5 - 2x_{1i} + 3x_{2i} + x_{3i} + x_{4i} + \epsilon_i$$

donde $\epsilon_i \sim N(0, 0{,}01)$ independientes las covariables cumplen

- x_1, x_2 son independientes.
- $x_3 = 2x_1 + \delta$ con $\delta \sim N_{40}(0, 0{,}001)$
- $x_3 = -x_2 + \varsigma$ con $\varsigma \sim N_{40}(0, 0{,}001)$

1. Haga el ajuste de Y en las covariables. ¿Le dan bien los estimadores de los parámetros? Mire los test t para los coeficientes ¿Cuáles son los coeficientes significativos? ¿Es la regresión significativa? ¿Hay alguna incoherencia en la salida?
   Enumere algunos indicios de colinealidad.
2. Calcule los factores de Inflación de la Varianza (VIF) ¿Qué le sugieren?
3. Calcule los índices de condición para X_s la matriz de diseño escalada (es decir, luego de dividir a cada columna de la matriz de diseño por su norma). ¿Qué le sugieren?
4. Describir los criterios de selección de variables basados en el Cp de Mallows y en el R^2 ajustado. ¿Qué subconjunto de covariables eligiría para modelar la media de Y según cada criterio?
5. Describa los método Forward, Backward y Stepwise y diga con qué subconjunto de variables se quedaría mirando la salida del R.