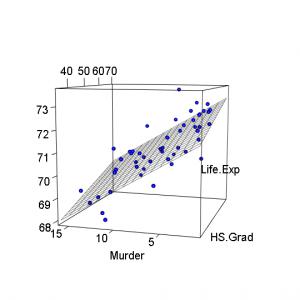
**Multiple** [**Regression**](https://www.r-exercises.com/2017/01/15/multiple-regression-part-1/)

En los siguientes ejercicios vamos a cubrir el análisis de regresión múltiple en R.



Utilizaremos el set de datos state.x77, disponible en R: help(state.x77).

**Exercise 1**

a. Cargar el set de datos state.x77.

b. Convertir el set de datos state.x77 a dataframe.

c. Re-nombrar las variables **Life Exp** a **Life.Exp**, y **HS Grad** a **HS.Grad**.

d. Despliegue una estadística descriptiva básica para cada variable.

**Exercise 2**

Utilizar el método de las distancias Cook's para detectar outliers. Ayuda: en general, distancias Cook's mayores que 4 pueden influenciar significativamente el modelo.

**Exercise 3**

Utilizar la función **nearZeroVar()** del paquete **caret** para encontrar las variables que tengan “varianza casi cero”. Descartar aquellas variables con NZV=0.

**Exercise 4**

Calcular la matriz de correlación. Grafique una matriz de los scatterplots.

**Exercise 5**

Supongamos que queremos ingresar todas las variables en un modelo de regresión lineal de primer orden con esperanza de vida como variable dependiente. Ajustar este modelo.

**Exercise 6**

Suponga que estamos interesados en remover las variables Income, Illiteracy, y Área del modelo anterior. Use la función ***update()*** para ajustar este modelo.

**Exercise 7**

Supongamos que nos hemos basado en un modelo que tiene HS.Grad y Murder como predictores. Ajustar este modelo.

**Exercise 8**

Incluya en el modelo anterior un término de interacción; por ejemplo, **HS.Grad:Murder**. Genere tres posibles iteraciones.

**Exercise 9**

Para este y los remanentes ejercicios utilizaremos el modelo del ejercicio 4. Calcular un intervalo de confianza del 95% para los coeficientes de las dos variables predictoras.

**Exercise 10**

Realizar una predicción de la expectativa de vida para un estado donde el 55% de la población son graduados de la secundaria y la rata de asesinatos es 8 por cada 100.000 habitantes.

**Exercise 11**

Calcuar un intervalo de confianza del 98% para la expectativa de vida promedio en un estado donde el 55% de la población son graduados de la secundaria, y la rata de asesinatos es 8 por cada 100.000 habitantes.

**Exercise 12**

Calcular un intervalo de confianza del 98% para la expectativa de vida de una persona viviendo en un estado donde el 55% de la población son graduados de la secundaria, y la rata de asesinatos es 8 por cada 100.000 habitantes.

**Exercise 13**

Como nuestro modelo solo tiene dos variables predictoras, podemos generar un gráfico 3D de nuestros datos y el plano de regresión ajustado.

**Exercise 14**

Utilizar el método paso a paso o stepwise; el cual emplea criterios matemáticos para decidir qué predictores contribuyen significativamente al modelo y en qué orden se introducen.

**Exercise 15 (k-fold Cross Validation)**

Evaluar el rendimiento del modelo; para eso, dividir los datos de entrada en 'k' porciones de muestras aleatorias mutuamente excluyentes. Manteniendo cada porción como datos de prueba, construimos el modelo final sugerido sobre los datos restantes (parte k-1) y calculamos el error cuadrático medio de las predicciones (MSE). Finalmente, calcular el promedio de estos errores cuadráticos medios (para las porciones 'k'), y su equivalente RMSE (Root Mean Square Error or Deviation).