***Ejercicio 1. Regresión lineal simple***

Obtener la ecuación de la recta o correlación lineal entre el la concentración de Cr (variable independiente) y la concentración de Ni (variable dependiente) a partir de los datos obtenidos en el Ejercicio 1 de las prácticas de [Matriz de correlación y gráficos de dispersión](https://web.ua.es/es/lpa/docencia/analisis-estadistico-de-datos-geoquimicos-con-r/matriz-de-correlacion-y-graficos-de-dispersion.html).

Este ejercicio se puede realizar cambiando los encabezados Cr por X1 y Ni por Y y usar la sintaxis genérica explicada anteriormente:

datos<-read.table("datos.txt",header=T,blank.lines.skip=F)  reg<-lm(Y~X1,data=datos)Y\_teor<-reg$fitted.valuesplot(datos$Y,Y\_teor)abline(lm(datos$Y~Y\_teor),col="blue")

 O bien dejar los encabezados originales y cambiar la sintaxis:

datos<-read.table("datos.txt",header=T,blank.lines.skip=F)  reg<-lm(Ni~Cr,data=datos)Y\_teor<-reg$fitted.valuesplot(datos$Ni,Y\_teor)abline(lm(datos$Ni~Y\_teor),col="blue")

***Ejercicio 2. Regresión lineal multiple***

Se ha estudiado el comportamiento frente a la cristalización de las sales 10 rocas porosas utilizadas como material de construcción. La durabilidad de estas rocas se ha estimado mediante la pérdida de masa después de cristalización de sales. Además se caracterizaron diferentes propiedades de la roca, como la resistencia a la compresión simple, la densidad de conjunto (o aparente), y la porosidad efectiva y el tamaño medio de poro, ambas obtenidas con porosimetría de intrusión de mercurio.

El objetivo de este ejercicio es seleccionar el mejor modelo lineal múltiple para predecir la durabilidad de las rocas (variable dependiente) en función de las otras variables (independientes). Este ejemplo muestra cómo se pueden transformar variables independientes para que tengan una influencia lineal sobre la variable dependiente (Datos obtenidos en: Benavente et al., (2007): The influence of petrophysical properties on the salt weathering of porous building rocks. Environ Geol 52:197–206).

El primer paso es realizar la matriz de correlación y el grafico de dispersión. De esta forma podemos evaluar qué variables se pueden transformar o eliminar por no ser significativas. Los gráficos de dispersión muestran una relación inversa entre la pérdida de masa y el tamaño de poro y la resistencia mecánica. Para ello, en la matriz de datos.txt añadir una columna con la inversa del tamaño del poro (r-1) y otra con la inversa de la resistencia a la compresión simple (RCS-1).

Observando los p-valores (Pr(>|t|)) de los contrastes de significatividad individual, vemos que las variables no significativas (p-valor > 0.05) son la porosidad y la densidad de conjunto. Eliminando estas variables (en realidad teniendo en cuenta todas las variables menos éstas: reg<-lm(LWD~r-1+RCS-1,data=datos)), volvemos a realizar la regresión lineal y obtenemos la siguiente regresión lineal múltiple:

LWD [%] = -2.99 + 1.05/r[micras] + 145.39/RCS [MPa].

En este último modelo todas las variables son significativas.

Los resultados obtenidos y el procedimiento seguido en esta práctica no son tan sencillos cuando se aplican en muchos problemas geológicos/geoquímicos, los cuales pueden llegar a ser mucho más complejos que el mostrado aquí. En la mayoría de los problemas geológicos/geoquímicos se eliminan las variables que menos están relacionadas (lo que el análisis factorial será de gran utilidad) y se intenta buscar un modelo que contenga el mayor número de variables con el mayor valor del coeficiente de correlación.

***Ejercicio 3. Regresión lineal multiple***

Obtener una expresión lineal múltiple que relaciones variables de la cuenca, X1, ...,X6, con la magnitud de la cuenca, Y, definida como el número de cursos fluviales (ríos, arroyos, afluentes, etc.). Las variables (independientes) de la cuenca son X1: elevación de la desembocadura de la cuenca (pies, ft); X2: relieve (profundidad) de la cuenca (pies,ft); X3: área de cuenca (milla2); X4: longitud total del cursos fluviales (ríos, arroyos, afluentes, etc.) (milla); X5: densidad de drenaje, definido como X4 / X3 ·100; X6: factor de forma de la cuenca, la cual cuantifica su desviación  con respecto a una sección circular. El objetivo de esta práctica es predecir el valor de la magnitud de la cuenca, Y, a partir de todas las variables de la cuenca debido a que cada una de ellas son importantes y necesarias para el modelo físico. Para evaluar la fiabilidad del ajuste o estimación de Y a partir de la expresión lineal múltiple, representar el valor real de Y frente al valor estimado por la función (Davis, 2002).