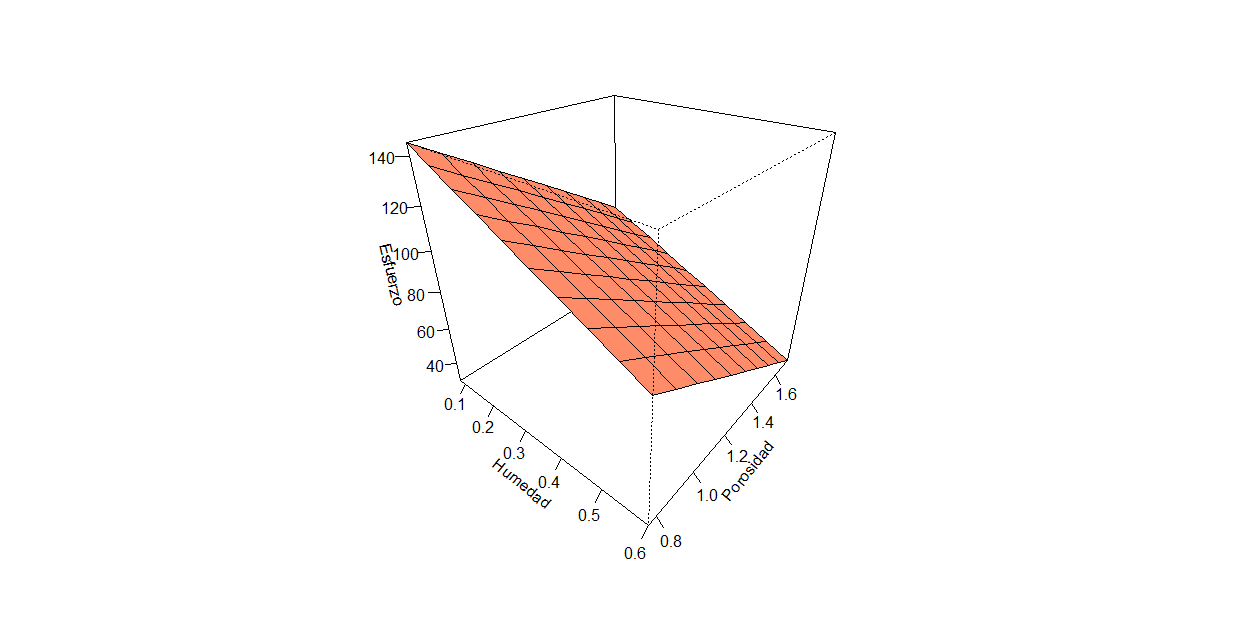
# Ejercicios sugeridos

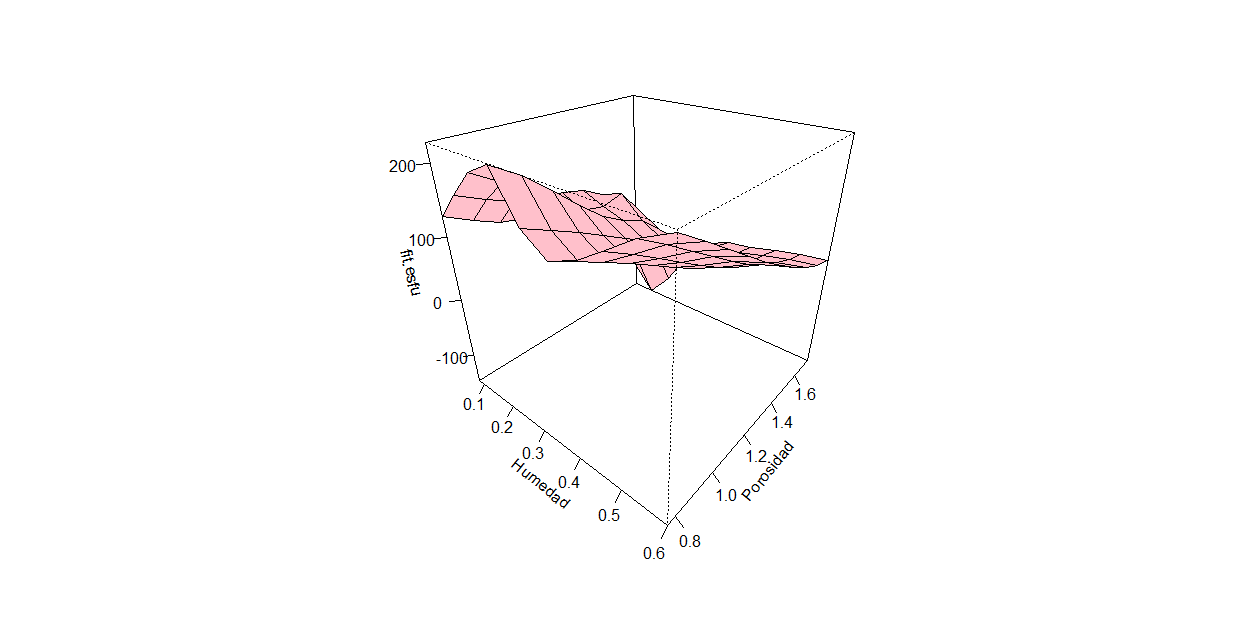
1. En la Revista Dyna de la UNAL se publicó un artículo titulado “Correlación de la resistencia a compresión uniaxial con la humedad y porosidad eficaz en rocas” en donde se presenta un estudio de cómo influye el contenido de humedad y porosidad sobre el esfuerzo de compresión de un tipo de rocas. De clic [sobre este enlace](http://mr.crossref.org/iPage?doi=10.15446%2Fdyna.v83n198.53151) para consultar el artículo y luego úselo para responder las siguientes preguntas.
   1. ¿En qué año fue publicado el artículo?
   2. Según lo que se menciona en el resumen, ¿de cuál región son las rocas usadas en el estudio?
   3. Observando la tabla 3 del artículo, ¿cuántas observaciones usaron los investigadores para el estudio?
   4. Observe con detalle la figura 4. ¿Qué variables de la tabla 3 se usaron para construir la figura?
   5. Obtenga de la tabla 3 las variables apropiadas y reconstruya la figura 4. ¿Qué conclusiones se pueden obtener al observar con detalle la figura 4 suya?
   6. Observe la tabla 9, ¿cuál fue el modelo estadístico que los autores ajustaron?
   7. Vuelva a ajustar usted el modelo de la tabla 9 y llámelo mod1 así:

mod1 <- lm(esfuerzo ~ humedad + porosidad)

* 1. Obtenga la tabla de resumen para el modelo mod1 y compárelo con los resultados de la tabla 9. ¿Qué conclusión puede usted obtener hasta este momento?
  2. ¿Qué programa usaron los autores para ajustar el modelo?
  3. ¿Cuál es el del mod1? Compárelo con el resultado mostrado en la expresión 9 del artículo. ¿Qué conclusión puede obtener?
  4. Abajo se muestra la superficie de respuesta para el modelo mod1 construida en R usando una rejilla de 10x10. Reconstruya usted mismo la figura de abajo y compárela con la figura 5 del artículo. ¿Qué conclusión puede obtener?



* 1. Construya el modelo mod2 en el cual las covariables humedad y porosidad expliquen el esfuerzo usando la función loess.
  2. Dibuje la superficie de respuesta para mod2, el resultado se debe parecer a la figura mostrada a continuación.



* 1. Para ambos modelos ajustados hasta ahora use la función predict de R para estimar el esfuerzo promedio de una roca con humedad de 0.4 y porosidad de 1.0. Respuesta: 96.92127 con mod1 y 104.3993 con mod2.
  2. Use la función names aplicada sobre ambos modelos mod1 y mod2 para identificar dónde se almacenan los valores ajustados o predichos para la variable respuesta. Respuesta: $fitted.values y $fitted.
  3. Calcule el coeficiente de correlación entre la variable respuesta y los valores ajustados obtenidos con ambos modelos. Respuesta: 0.7183326 con mod1 y 0.9288519 con mod2.
  4. ¿Qué significan los coeficientes de correlación obtenidos en el punto anterior? ¿Cómo se pueden interpretar?
  5. ¿Qué modelo le recomendaría a Manolo Galván e Iván Restrepo los autores del artículo?
  6. Escriba dos conclusiones de este ejercicio.

1. En el artículo titulado “Biosorption of chromium using factorial experimental desing” se presentan los resultados de un estudio llevado a cabo para ver el efecto del pH, temperatura y concentración de metal sobre el porcentaje de remoción de y . De clic [sobre este enlace](https://www.dropbox.com/s/pnno5s5qce560op/Carmona%2C%20et.al%20%282005%29.pdf?dl=0) para consultar el artículo y luego úselo para responder las siguientes preguntas.
   1. ¿En qué revista se publicó el artículo?
   2. ¿En qué año se publicó el artículo? ¿De qué instituciones son los autores del artículo?
   3. En la tabla 2 del artículo se muestran los datos experimentales. Construya vectores para almacenar las variables (codificadas como -1 y 1) temperatura, concentración, pH y las dos variables respuesta. ¿Cuántas observaciones tiene la base de datos? Respuesta: 16 observaciones en cada vector.
   4. En las expresiones 3 y 4 del artículo se muestran dos ecuaciones, ¿qué les falta a esas ecuaciones para estar correctamente escritas? Respuesta: falta el símbolo ^ sobre la variable.
   5. Ajuste nuevamente con R los modelos reportados en la tabla 3, llame mod1 al modelo que tiene como variable respuesta el y al otro llámelo mod2. ¿Cuáles son los de ambos modelos? ¿Fueron reportados estos dos valores en el artículo? Respuesta: 97.96% para mod1 y 98.72% para mod2.
   6. A un nivel de significancia del 5%, ¿cuáles términos NO son significativos en los modelos mod1 y mod2? Respuesta: en mod1 no es significativa la interacción triple y en el mod2 no es significativa la interacción triple y la interacción entre temperatura y pH.
   7. Vuela a ajustar los modelos mod1 y mod2 eliminando los términos no significativos. ¿Cuáles son los nuevos valores de ? Respuesta: 98.03% y 98.73%.
   8. Escriba las expresiones estimadas para el porcentaje de remoción de y con los modelos del literal anterior.
2. Regresión lineal con términos polinomiales.
   1. Simule 500 observaciones con el siguiente modelo y construya el diagrama de dispersión de versus .
   2. Ajuste a los datos anteriores el modelo mod1 de regresión lineal con la estructura que se muestra a continuación.

,

* 1. Dibuje los gráficos de residuales para el modelo mod1 e interprete los resultados. Use el código mostrado a continuación para crear los tres gráficos.

par(mfrow=c(1, 3)) # Para dividir la ventana grafica

plot(mod1, which=1:3) # Para dibujar los tres primeros

* 1. Como los gráficos anteriores muestran patrones cuadráticos e indican que el modelo no se ajusta bien, es necesario intentar un nuevo modelo. Ajuste el modelo mod2 mostrado a continuación.

mod2 <- lm(y ~ x + I(x^2))

par(mfrow=c(1, 3))

plot(mod2, which=1:3)

Observe que para incluir una potencia de una variable se usa la función I() y dentro de ella la potencia deseada.

¿Cómo se comportan ahora los residuales? ¿Cuál de los dos modelos se ajusta mejor a los datos?

* 1. Vuelva a dibujar el diagrama de dispersión inicial e incluya la curva de media estimada con los resultados del modelo mod2. ¿Qué tan bien se ajusta la curva a la nube de puntos?

1. Replicar el ejemplo de optimización de una superficie de respuesta disponible [en este enlace](http://virgo.unive.it/paolop/papers/GMP.pdf).
2. En clase se mostró cómo usar la función nlmimb de R para encontrar los valores que maximizan el Rendimiento dada la Temperatura y Concentración. Use la función optim para realizar el mismo ejemplo y compare los resultados obtenidos.
3. En [este enlace](https://stackoverflow.com/questions/15337777/fit-a-line-with-loess-in-r) está la pregunta que el usuario Leo publicó en Stackoverflow. Lea la pregunta y responda lo siguiente:
   1. ¿Cuántas covariables tiene el modelo lw1 que ajustó Leo? ¿Qué función usó para ajustar el modelo? Respuestas: una y loess.
   2. Corra el código de R que Leo suministró para hacer la pregunta. ¿Cuál es el problema con la curva de color azul?
   3. Rob Hyndman amablemente le respondió a Leo y le menciona el error que él está cometiendo. ¿Cuál es el error de Leo?
4. Consulte la ayuda de la función loess de R para responder las siguientes preguntas:
   1. ¿Qué significa que degree=2?
   2. ¿Puede degree ser igual a 3?
   3. ¿Qué valores puede tomar el parámetro span?
5. ¿Qué diferencia hay entre un spline básico y un spline natural? Rta: que el spline natural debe ser en los bordes una línea recta.
6. Suponga que queremos crear un spline para una variable X y deseamos que R nos coloque automáticamente 4 nodos. ¿Qué valor debemos darle al parámetro df para conseguir esos cuatro nodos? Rta: df=7.
7. Dentro del paquete MASS está la base de datos Cars93 con información de 93 autos. Consulte la ayuda de la base de datos para conocer sus variables.
   1. Ajuste los siguientes tres modelos:

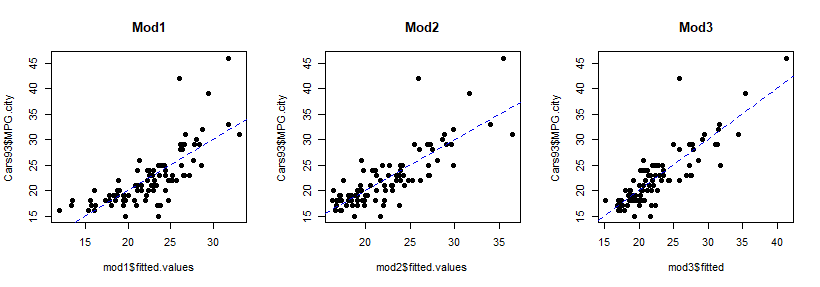
Mod1:

Mod2:

Mod3:

En el Mod2 le sugiero que use poly para escribir el polinomio y raw=FALSE para que obtenga polinomios ortogonales. En el Mod3 el término lo debe obtener con loess usando los parámetros por defecto.

* 1. Para los modelos anteriores construya la siguiente figura e incluya una línea a 45° en cada panel. ¿Qué vectores van en cada uno de los ejes?



* 1. ¿Qué conclusión se puede obtener de la figura anterior?
  2. Calcule el coeficiente de correlación lineal de Pearson para cada uno de los paneles anteriores. Rta: 0.760248, 0.8082836 y 0.8528208 respectivamente.
  3. ¿Cuál modelo usted recomendaría? Rta: el modelo 3 sin duda alguna.
  4. Como el mejor modelo es Mod3 se construyó usando span=0.75 y degree=2, la pregunta que surge es ¿cuál debería ser el valor óptimo para el parámetro span? Para responder esto usted debe construir una función llamada aux, esta función debe recibir un valor admisible para span y debe entregar la correlación entre e . Use el siguiente código como ejemplo y complete los espacios.

aux <- function(x) {

????

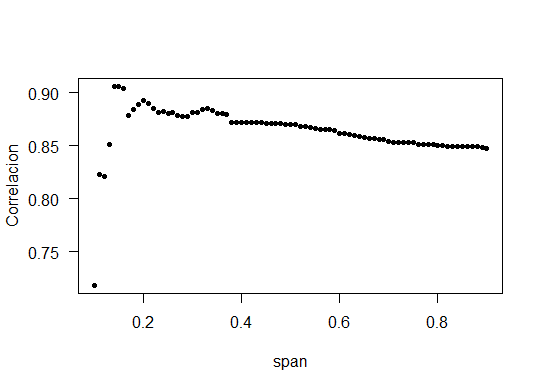
}

aux <- Vectorize(???)

* 1. Use el código mostrado a continuación para crear la figura siguiente.

span <- seq(from=0.1, to=0.9, by=0.01)

plot(x=span, y=aux(span), las=1, pch=20, ylab='Correlacion')



* 1. Mirando la figura anterior, ¿cuál debería ser el valor de span para maximizar la correlación? Rta: se debe tener buena vista para decir algo.
  2. Use la función optimize de R para obtener el valor exacto de span maximiza la correlación? Rta: el valor exacto es 0.1954733 para obtener una correlación máxima de 0.892692.
  3. Otro parámetro de la función loess que se puede modificar es degree. Obtenga la combinación de span y degree que maximizan la correlación. Rta: la mejor combinación es span=0.1105937 con degree=1 para una correlación máxima de 0.9231195.

1. [En este enlace](http://r-statistics.co/Loess-Regression-With-R.html) se encuentra una publicación de un blog en la cual se muestra como elegir el valor de span apropiado en loess para minimizar la Suma de Cuadrados del Error. Use el procedimiento mostrado para encontrar la mejor combinación de span y degree en el ejemplo anterior.
2. Considere el mismo ejercicios sobre la base de datos Cars93 y ajuste los siguientes modelos:

Mod1:

Mod2:

* 1. Encontrar los valores de y para maximizar la correlación entre e para ambos modelos.
  2. Encontrar los valores de y para minimizar la suma de error cuadrático medio de ambos mod.
  3. Elegir el mejor modelo y dibujar la superficie en 3d que lo representa.

1. ¿Qué significan las letras de las funciones bs y ns?
2. ¿Cómo se incluyen splines básico y splines naturales en gamlss?