

## Regresion Ridge

### Problema:

- En el metodo de minimos cuadrados el requisito de que  $\beta_B$  (beta sombrero) sea un estimador insesgado asegura que tiene una varianza minima en la clase de los estimadores lineales insesgados, pero no hay garantia que esa varianza sea pequena. Esto implica que los intervalos para los valores posibles de B es amplio, por lo tanto, B es muy inestable.

### Solucion:

- Una posible solucion es eliminar el requisito de que el estimador sea insesgado y determinar un estimador sesgado, que tenga menor varianza que B sombrero.

Resumen: Problema de multicolinealidad implica que los estimadores B sea muy inestables porque tiene un intervalo amplio de varianza, para poder solucionarlo, se propone que el estimador no sea insesgado e intentar obtener menos varianza en la estimacion de los parametros.

**Regresion Lineal Multiple:**  $B \text{ sombrero} = (X'X)^{-1} X'y$

**Regresion Ridge:**  $B \text{ sombrero ridge} = (X'X + kI)^{-1} X'y$

donde  $k \geq 0$  es una constante que selecciona el analista.

Si  $k = 0$  entonces estamos en una RLM. A  $k$  se la suele llamar "parametro de sesgo".

Si  $k > 0$  el sesgo de B sombrero ridge aumenta al aumentar  $k$ , sin embargo la varianza disminuye al aumentar  $k$ . (ver formula)

### Objetivo:

Seria bueno escoger un  $k$  tal que la reduccion en la varianza sea mayor que el aumento en el sesgo al cuadrado

Los parametros estimados de ridge pueden dar como resultado una ecuacion que funcione mejor para predecir observaciones futuras.

**Traza de Ridge:** es un grafico que muestra los parametros B sombrero  $r$  en funcion de  $k$ . y

los valores de  $k$  suelen estar entre 0 y 1.

A medida que  $k$  aumenta los valores de  $B$  se irán estabilizando, sin variar en forma drástica. El objetivo es encontrar un valor de  $k$  lo razonablemente pequeño, en el cual los estimadores de ridge sean estables. Así se puede llegar a obtener parámetros con menos MSE que el obtenido por mínimos cuadrados.

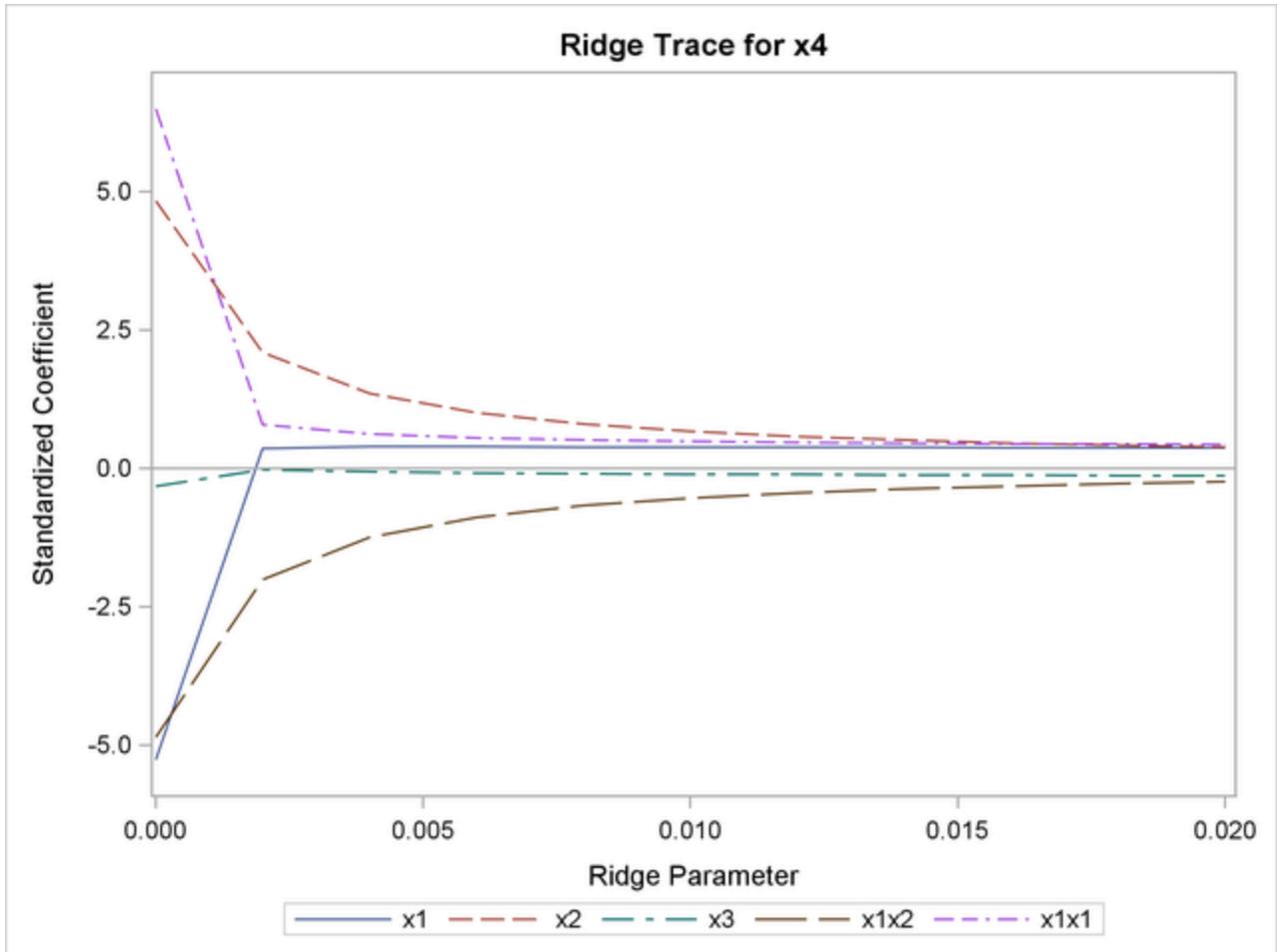
### **Datos del acetileno:**

Explicar de qué se tratan los datos, y para obtener la regresión de ridge se deben resolver las ecuaciones normales (fórmula de ridge) con varios valores de  $0 \leq k \leq 1$ .

Mostrar traza de ridge para estos datos.

Se requiere juicio para interpretar la traza de ridge y seleccionar un valor adecuado de  $k$ . Se debe seleccionar un  $k$  lo bastante grande como para producir coeficientes estables, pero que no sea innecesariamente grande porque se introduce sesgo y aumenta el cuadrado medio residual.

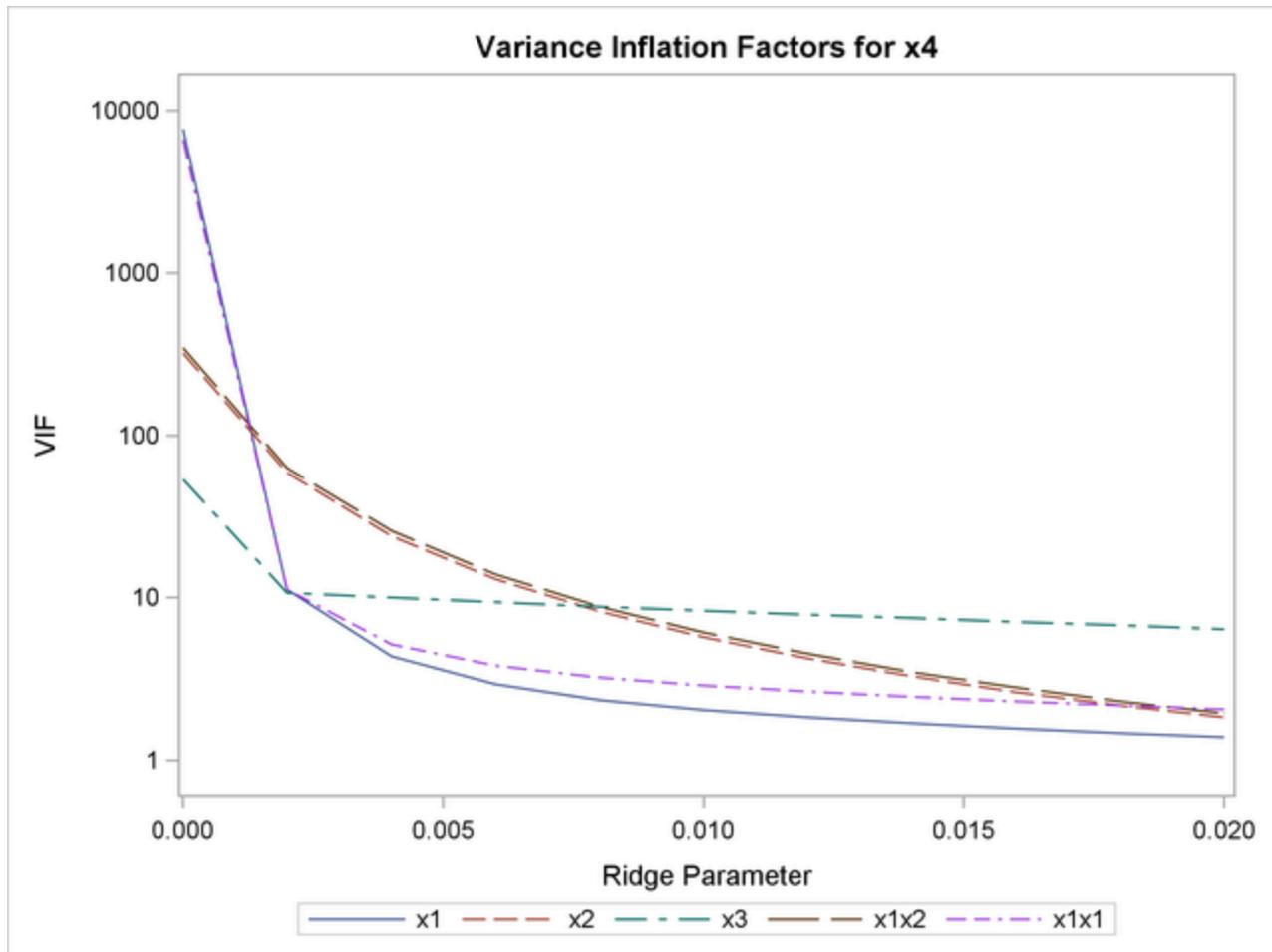
Mostrar gráficos de la salida de SAS.



### Como seleccionar el k?

La eleccion de k por inspeccion de la traza de ridge es un procedimiento subjetivo que requiere criterio por parte del analista.

Una opcion es analizar el grafico de VIF de cada estimador en funcion de k y elegir el k talque los valores de VIF esten cerca de 1.



Existen metodos analiticos para elegir el k:

- *Procedimiento iterativo* de estimacion basado en la ecuacion:

El cambio relativo de  $kj_j$  se usa para determinar el procedimiento. si (ver formula) el algoritmo debe continuar.

- *Ck Mallows*. Si el modelo se va usar para prediccion, sera mas adecuado tener en cuenta criterios orientados a predicciones, para elegir k. Ver Mallows, modifiko el cp de Mallows, para formar un ck

- *PRESS*

### Regresion de Ridge y Seleccion de variables

Al utilizar estimadores sesgados y obtener regresores ortogonales se puede realizar seleccion

de variables, y utilizar la Traza de Ridge como guía para la selección.

### Comparación y evaluación de estimadores insesgados

Si bien no hay un solo procedimiento que sea el mejor en general, hay bastante evidencia de que indica la superioridad de la estimación sesgada respecto a los mínimos cuadrados si hay multicolinealidad.

Pero existen distintas opiniones de diversos autores acerca del tema, comentamos algunos:

- **Montgomery-Peck-Vining**: prefieren la regresión ridge ordinaria seleccionando el  $k$  por inspección de la **traza de ridge**, dicen que el procedimiento es directo, fácil de implementar y el analista puede aprender con mucha rapidez a interpretar la traza de ridge. También sugieren que es útil obtener el  $k$  óptimo mediante el proceso iterativo definido por Hoerl, Kennard y Baldwin.

- **Conniffe y Stone**: critican el uso de la traza de ridge para seleccionar el parámetro de sesgo, ya que  $\widehat{\beta}_r$  **cambia con lentitud**, y al final se estabiliza al aumentar  $k$ . También dicen que si los datos no son adecuados para respaldar un análisis adecuado por mínimos cuadrados, no es probable que ayude la regresión de ridge, porque los parámetros estimados no tendrán sentido.

- **Varios autores** notaron que si bien se puede demostrar que existe un  $k$  tal que MSE de ridge es menor que el MSE por mínimos cuadrados, **no hay seguridad** que la traza de ridge o algún otro método estocástico puede **producir el  $k$  óptimo**.

- **Draper y Van Nostrand**: critican los estimadores sesgados, hacen notar que la regresión ridge en realidad solo es **adecuada** en casos en que **se agrega información externa** a un problema de mínimos cuadrados.

-**Smith y Campbell**: sugieren el uso de análisis bayesiano explícito, o la estimación mixta, para resolver problemas de multicolinealidad. Rechazan lo **metodos de ridge por debiles e imprecisos**, porque solo incorporan muy holgadamente ideas de información anteriores en el análisis.

-**Marquardt y Snee**: los métodos sesgados de estimación si se comparan muy favorablemente con otros métodos de manejo de multicolinealidad, como la eliminación de las variables. Con frecuencia **es mejor usar algo de información de todos los regresores**, como lo hace la regresión ridge, **que usar toda la información de algunos de los regresores**, como lo hace la eliminación de variables.

Finalmente, no recomendamos el uso automático o mecánico de la regresión de ridge sin un estudio concienzudo de los datos, y un análisis cuidadoso de la adecuación del modelo final. Cuando se usan en forma adecuada, los métodos de estimación sesgada son herramientas valiosas entre los recursos de análisis de datos.