IN4402 - Aplicaciones de Probabilidades y Estadística en Gestión
Profesor: Benjamín Villena
Profesores auxiliares: Ronald Leblebici, Angelo Muñoz

INGENIERIA INDUSTRIAL
UNIVERSIDAD DE CHILE

Clase auxiliar # 7

Selección de modelos: método LASSO y test Reset de Ramsey

Método LASSO: este método busca minimizar la suma de los residuos al cuadrado (como ya se vio antes), pero penalizada por un factor que permite controlar la cantidad de variables.

$$\min_{\hat{\beta}} \quad \sum_{n=1}^{N}(Y_n - \hat{\beta}_0 - \sum_{k=1}^{K}\hat{\beta}_k X_{nk})^2 + \lambda \sum_{k=1}^{K}|\hat{\beta}_k|$$

Test Reset de Ramsey:

1. Especificar $Y = X\beta + U$ y estimar $\hat{\beta}$ e $\hat{Y}$.
2. A partir de la estimación $\hat{Y}$, crear las variables $\hat{Y}^p$ para $p \in \{1, ..., P\}$
3. Regresionar $Y = X\beta + \gamma_2\hat{Y}^2 + ... + \gamma_P\hat{Y}^P + V$
4. Testear $H_0 : \gamma_2 = ... = \gamma_P = 0$ (restricciones múltiples).

1. Método LASSO

Tras fracasar en la Liga Kalos, nuestro héroe Ash Ketchum de Pueblo Paleta, ha decidido dedicarse a su verdadera vocación: la política. Afortunadamente y contra todo pronóstico, Ash ha logrado la elección como presidente de su país. Los analístas políticos buscan una explicación y lo contratan a usted para seleccionar aquellas variables que **no** deberían considerar en su análisis.

Para llevar a cabo esta tarea, se le ha entregado la base de datos almacenada en el archivo `Auxliar 07 Excel LASSO.xlsx` (almacenado en Material Docente de U-Cursos). Dicha base de datos tiene información de distintas ciudades y pueblos, como el promedio de escolaridad, desempleo, ingreso, nivel de los pokemones de la zona, un índice de ideología política, la popularidad de Pikachu y una serie de "variables basura".

1. Elija un valor de λ y ejecute el método LASSO.

   Respuesta: Ver archivo `Auxiliar 07 Excel LASSO.xlsx`.

2. ¿Qué variables descartaría en primera instancia? ¿Por qué?

   Respuesta: Descartaría aquellas variables cuyo coeficiente estimado se hace cero[1] tras la optimización del método LASSO. Dichas variables coinciden con aquellas que denominamos "variables basura" en el enunciado.

[1] Al hacer la optimización en *Excel*, los coeficientes no son exactamente cero, pero se acercan bastante.

3. Discuta: ¿cuál es la importancia del parámetro λ en el problema de optimización?

   Respuesta: La importancia del parámetro λ radica en que este representa la magnitud de una penalización marginal en la función objetivo sobre el término $\sum_{k=1}^{K} |\hat{\beta}_k|$. Cuando este término crece en una unidad, la función objetivo –que queremos minimizar– crece λ.

   A mayores valores de λ, la función objetivo se verá más castigada por el exceso de variables en el modelo y la optimización será más rigurosa al momento de seleccionar variables. Por lo tanto, λ reflejará lo selectivos que estamos siendo al controlar la cantidad de variables del modelo.

Comente:

4. "El método LASSO sirve para seleccionar variables o forma funcional correcta dentro de un modelo, incluso si hay un número mayor de variables que observaciones".

   Respuesta: **Verdadero**. El método LASSO sirve para seleccionar un número pequeño de variables dentro de un grupo de K regresores donde K puede ser mayor al tamaño muestral N.

5. "No es posible aproximar un modelo cuya forma funcional no sea polinómica mediante el método LASSO".

   Respuesta: **Falso**. Cualquier forma funcional puede ser aproximada por polinomios y productos de polinomios de las variables del modelo, por lo cual es posible aproximar la forma funcional también por LASSO, al incluir polinomios y producto de polinomios en la lista de K regresores entre los cuales el método escogería aquellos que minimicen su función objetivo.

2. Test Reset de Ramsey

Utilice la base de datos `Auxiliar 06 Test Reset Ramsey BD.dta` y considere el modelo $y_i = \beta_0 + \beta_1 z_i + \beta_2 x_i + u_i$.

1. ¿Qué modelo plantearía para realizar el test Reset de Ramsey con potencias $p = 2, 3$?

   Respuesta: Plantearía el modelo $y_i = \beta_0 + \beta_1 z_i + \beta_2 x_i + \gamma_2 \hat{y}_i^2 + \gamma_3 \hat{y}_i^3 + v_i$, donde $\hat{y}_i$ sería la estimación por MCO del modelo $y_i = \beta_0 + \beta_1 z_i + \beta_2 x_i + u_i$.

2. Plantee un test de hipótesis (hipótesis nula, hipótesis alternativa, estadístico y región de rechazo) para dicho test.

   Respuesta:

   - H_0: $\gamma_2 = \gamma_3 = 0$.
   - H_1: $\gamma_2 \neq 0 \vee \gamma_3 \neq 0$.
   - $F = \frac{(R_L^2 - R_R^2)/r}{(1-R_L^2)/(N-K)} \sim \mathcal{F}(r, N-K)$.

- $RR = \{F > F_{\text{crítico}}(r, N-K)\}$

3. Realice dicho test en *Stata*. ¿Qué puede concluir?

Respuesta: Ver archivo `Auxiliar 07 Código.do`. Como se puede ver en la imagen, se puede rechazar la hipótesis nula con una significancia del 5 % (e incluso menor). Lo anterior permite rechazar que γ_2 y γ_3 sean cero, lo que es señal de que al menos un regresor de segundo y/o tercer grado de x y z podría ser estadísticamente significativo al momento de explicar y.

```
. test (y_estimado_2=0) (y_estimado_3=0)

 ( 1)  y_estimado_2 = 0
 ( 2)  y_estimado_3 = 0

       F(  2, 97921) = 3564,37
            Prob > F =    0,0000
```

3. Omisión de variables relevantes

1. Demuestre que al omitir variables relevantes $\hat{\beta}_{MCO}$ es sesgado.

Respuesta:

Modelo estimado: $Y = X_1\beta_1 + U$

Modelo verdadero: $Y = X_1\beta_1 + X_2\beta_2 + U$

$$\begin{aligned}
\hat{\beta}_1 &= (X_1^T X_1)^{-1} X_1^T Y \\
&= (X_1^T X_1)^{-1} X_1^T (X_1\beta_1 + X_2\beta_2 + U) \\
&= \underbrace{(X_1^T X_1)^{-1} X_1^T X_1}_{Id_{K_1 \times K_1}} \beta_1 + (X_1^T X_1)^{-1} X_1^T X_2 \beta_2 + (X_1^T X_1)^{-1} X_1^T U \\
\Rightarrow \mathbb{E}(\hat{\beta}) &= \underbrace{\mathbb{E}(\beta_1)}_{\beta_1} + \underbrace{\mathbb{E}[(X_1^T X_1)^{-1} X_1^T X_2 \beta_2]}_{\neq 0 \text{ (sesgo)}} + \underbrace{\mathbb{E}[(X_1^T X_1)^{-1} X_1^T U]}_{=0} \neq \beta_1
\end{aligned}$$

Comentario: Es importante entender que bajo omisión de variables relevantes ocurren otros dos fenómenos:

- $\hat{\beta}$ es un estimador **inconsistente** de β, pues como el sesgo **no** converge en probabilidad a cero, $\hat{\beta}$ **no** converge en probabilidad a β.

- Las variables del modelo "absorben" el efecto de variables omitidas, por lo que al estimar omitiendo variables relevantes creemos tener mayor información de la que en realidad tenemos.

  Por lo tanto la varianza estimada $\hat{\mathbb{V}}(\hat{\beta}_1)$ es **erróneamente más eficiente** que la verdadera varianza $\mathbb{V}(\hat{\beta}_1)$.

log_sal	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
esc	,0974079	,0005637	172,80	0,000	,0963031	,0985128
_cons	7,818565	,0068344	1144,01	0,000	7,80517	7,831961

log_sal	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
esc	,114268	,0006729	169,82	0,000	,1129492	,1155869
exp	,0115993	,0004689	24,74	0,000	,0106803	,0125183
exp2	-,000063	8,17e-06	-7,71	0,000	-,000079	-,000047
_cons	7,38576	,0110628	667,62	0,000	7,364077	7,407443

2. Explique a qué se deben las diferencias de $\hat{\beta}_{esc}$ y $\hat{\mathbb{V}}(\hat{\beta}_{esc})$ entre las regresiones de las tablas que se muestran a continuación.

   Respuesta: Como hemos explicado, al omitir variables $\hat{\beta}_{esc}$ se vuelve sesgado (e inconsistente, el sesgo no tiende a cero con muchas muestras), lo que explica la diferencia entre la estimación con y sin omisión de variables relevantes. Por otro lado, se observa que al omitir estas variables $\hat{\mathbb{V}}(\hat{\beta}_{esc})$ disminuye artificialmente porque $\hat{\beta}_{esc}$ está abarcando más información de la que le corresponde.