Profesor: Benjamín Villena

Profesores auxiliares: Ronald Leblebici, Angelo Muñoz



Clase auxiliar # 7

Selección de modelos (método LASSO y test de Ramsey) y omisión de variables relevantes

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 auxiliar5.xlsx (almacenado en Material Docente de U-Cursos). Dicha 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.
- 2. ¿Qué variables descartaría en primera instancia? ¿Por qué?
- 3. Discuta: ¿cuál es la importancia del parámetro λ en el problema de optimización?

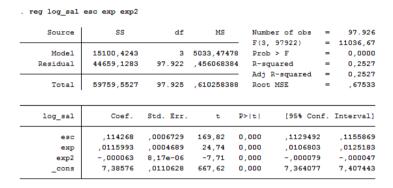
Comente:

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

2. Test Reset de Ramsey

Utilise la base de datos Auxiliar 07 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?
- 2. Plantee un test de hipótesis (hipótesis nula, hipótesis alternativa, estadístico y región de rechazo) para dicho test.
- 3. Realice dicho test en Stata. ¿Qué puede concluir?



3. Omisión de variables relevantes

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

[95% Conf. Interval]	P> t	t	Std. Err.	Coef.	log_sal
,0963031 ,0985128 7,80517 7,831961	0,000 0,000	172,80 1144,01	,0005637 ,0068344	,0974079 7,818565	esc _cons
[95% Conf. Interval]	P> t	t	Std. Err.	Coef.	log_sal
,1129492 ,1155869 ,0106803 ,0125183 -,000079 -,000047 7,364077 7,407443	0,000 0,000 0,000 0,000	169,82 24,74 -7,71 667,62	,0006729 ,0004689 8,17e-06 ,0110628	,114268 ,0115993 -,000063 7,38576	esc exp exp2 _cons