

Clase auxiliar # 3

Coefficiente de bondad de ajuste (R^2): es el porcentaje de varianza que se explica gracias a los regresores.

No es un criterio recomendable para elegir los regresores a incorporar en el modelo. $R^2 = 1 - \frac{\sum_{i=1}^N (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^N (Y_i - \bar{Y})^2}$. Este coeficiente siempre aumenta a medida que se agregan regresores, por lo que existe una versión "ajustada" que penaliza la incorporación de estos $R_a^2 = 1 - \frac{SCR}{SCT} \cdot \frac{N-1}{N-K}$.

Sesgo: sea $\hat{\beta}$ el estimador del parámetro poblacional β . El sesgo de $\hat{\beta}$ es la diferencia entre el valor esperado del estimador y el valor efectivo del parámetro. Es decir, $B(\hat{\beta}) = \mathbb{E}(\hat{\beta}) - \beta$.

Consistencia: un estimador consistente converge al valor del parámetro poblacional que busca estimar cuando el número de datos de la muestra tiende a infinito. Es decir, $\mathbb{P}(|\hat{\beta}_N - \beta| > \varepsilon) = 0$.

Linealidad: sea Y un vector de observaciones. Un estimador $\hat{\beta}$ es lineal en Y si se puede escribir como una combinación lineal de sus componentes. Es decir, $\hat{\beta} = \sum_n \omega_n Y_n$.

Teorema de Gauss-Markov: si se cumplen los supuestos de MCO, entonces el estimador $\hat{\beta}_{MCO}$ es el mejor estimador lineal insesgado (MELI) de los parámetros poblacionales β . ("Mejor" = "de mínima varianza").

Estimador de varianza de estimadores MCO: $\hat{V}(\hat{\beta}_A) = \frac{V(Y)(1-R^2)}{(N-K)V(X_A)} \cdot \frac{1}{1-R_{A,-A}^2}$

Descomposición del error cuadrático medio: $ECM(\hat{\beta}_k) = B(\hat{\beta}_k)^2 + V(\hat{\beta}_k)$

1. Teorema de Gauss-Markov

1. **[P4, Control 1, 2017]** Comente la siguiente afirmación: "Si los supuestos del teorema de Gauss-Markov no se cumplen, entonces el modelo de regresión múltiple no es una buena aproximación a la expectativa condicional de Y dado X , $\mathbb{E}(Y|X)$ ".
2. **[P4, Control 1, 2014]** Suponga el modelo de regresión lineal simple $Y_n = \beta_1 + \beta_2 X_n + U_n$, y considere el siguiente estimador para el parámetro β_2 : $\hat{\beta}_2 = \sum_n Y_n / \sum_n X_n$.
 - a. Evalúe si el estimador $\hat{\beta}_2$ es lineal e insesgado. Explícite cualquier supuesto que deba realizar para demostrar o refutar esta propiedad.
 - b. Dados los supuestos realizados, ¿podría proponer usted un mejor estimador que $\hat{\beta}_2$? ¿Qué estimador sería este? ¿En qué fundamenta esta propuesta?
 - c. Estime la varianza del estimador $\hat{\beta}_2$ y obtenga una expresión numérica para el nivel relativo de ineficiencia de dicho estimador respecto al MCO, es decir $V(\hat{\beta}_2)/V(\hat{\beta}_2)_{MCO}$. ¿Hay alguna circunstancia en la que $\hat{\beta}_2$ sea tan bueno como $\hat{\beta}_2$?

2. Teorema Frisch-Waugh / Multicolinealidad

[P5, Control 1, 2014] Un empresario productor de leche desea estudiar si dos tipos de forraje (alimento) X y Z logran elevar la producción de leche Y de 250 vacas que tiene en su predio.

- Suponga que a usted le pidieran diseñar el experimento para estudiar el efecto de los dos tipos de forraje X y Z sobre la producción de leche. Explique cómo asignaría cantidades de X y Z a las vacas para determinar los efectos de X y Z sobre la producción de leche con la mayor precisión.
- El empresario siguió su diseño experimental, pero encarga a un veterinario amigo que genere los resultados. El veterinario, solo sabe de regresión simple, por lo cual obtiene los siguientes resultados de dos regresiones donde las desviaciones estándar de los coeficientes están en el paréntesis bajo los respectivos coeficientes estimados. ¿Qué puede concluir sobre el verdadero impacto de X y Z sobre la producción de leche?

$\hat{Y}_n = 5,61 + 0,51X_n, R^2 = 0,15$ <div style="display: flex; justify-content: space-around; font-size: small;"> (0,21) (0,03) </div>	$\hat{Y}_n = 5,55 + 0,22Z_n, R^2 = 0,08$ <div style="display: flex; justify-content: space-around; font-size: small;"> (0,18) (0,12) </div>
---	---

- Suponga que en lugar de hacerle caso a ud., el empresario sigue el consejo de su amigo veterinario para diseñar el experimento. El veterinario sugiere que se asignen cantidades de X y Z aleatoriamente a cada vaca, completando exactamente 20 kilos de forraje en total. (Sugerencia: escriba la matriz de regresores). ¿Sería posible determinar los efectos individuales de X y Z bajo este diseño experimental? ¿Cómo cambia su respuesta si la cantidad de X más Z suman 20 con error aleatorio?

3. Bondad de ajuste y varianza de los estimadores

Usaremos la base de datos Auxiliar 03 - BD.xls subida a material docente. Usted está desarrollando su práctica profesional en una prestigiosa institución dedicada a la investigación económica. Su jefe le solicita regresionar el gasto en educación de los hogares en función de la edad del jefe de hogar (y su cuadrado). Sin embargo no está seguro si incluir el ingreso total de los hogares a la regresión debido al efecto que puede tener esto sobre la eficiencia de los estimadores de los otros regresores.

- Ejecute ambas regresiones en *Stata* (con y sin el ingreso total del hogar) y analice el efecto de incluir esta variable sobre las desviaciones estándar de los estimadores y los coeficientes R^2 y R_a^2 .
- Es su último día de práctica usted ha llegado a casa feliz pues le fue bastante bien. Además, como usted es un(a) alumno(a) responsable empezó a hacer su informe de práctica con anticipación, por lo que está casi listo(a) para salir de vacaciones. Solo le falta reportar un par de resultados que tiene impresos en su mochila. Sin pensarlo dos veces usted la abre y se da cuenta que se le reventó un lapiz cuya tinta no le permite ver algunos datos. Usted no tiene *Stata* instalado en su computador, sin embargo sabe utilizar las fórmulas necesarias y cuenta con *Matlab*. Obtenga los datos faltantes para poder salir de vacaciones.

Source	SS	df	MS	Number of obs = 15.727
Model	1,4932e+14	3	4,9772e+13	F(3, 15723) = 1723,59
Residual	4,5403e+14	15.723	2,8877e+10	Prob > F = 0,0000
Total	6,0335e+14	15.726	3,8366e+10	R-squared = [redacted]
				Adj R-squared = [redacted]
				Root MSE = 1,7e+05

gastoeduc	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
ejh	[redacted]	1216,075	2,57	0,010	739,3141 5506,609
ejh2	[redacted]	[redacted]	-1,58	0,114	-46,67254 4,992148
ytoth	[redacted]	,0008345	71,01	0,000	,0576259 ,0608973
_cons	[redacted]	27298,32	-1,38	0,166	-91294,03 15721,67