Efecto de la multicolinealidad en la variabilidad de los coeficientes de un modelo de regresión lineal

Karen Acuña Poveda[1], Elizabeth Araya Murillo[1], Andrés Padilla Loria[1]

Karen.acunapoveda@ucr.ac.cr, elizabeth.arayamurillo@ucr.ac.cr, andres.padillaloria@ucr.ac.cr

RESUMEN

En este estudio, se investigó la relación entre los diversos niveles de correlación y la posible presencia de multicolinealidad en modelos de regresión lineal, así como cómo esto podría haber afectado a la variabilidad de los coeficientes. Por ello, se llevó a cabo el ajuste de modelos de regresión y el análisis de coeficientes con base en errores estándar y el Factor de Inflación de la Varianza (VIF). Se consideraron diferentes casos que involucraban la interacción de coeficientes para variables correlacionadas, variaciones en el tamaño de la muestra y niveles de correlación. Entre los hallazgos, se puede destacar que la multicolinealidad perjudicó significativamente los coeficientes, y la intensidad de esta afectación dependió del tamaño de la muestra y del nivel de correlación. Este estudio tuvo como objetivo examinar el impacto pasado de la multicolinealidad en la variabilidad de los coeficientes en modelos de regresión lineal, utilizando técnicas como el análisis del nivel de correlación.

PALABRAS CLAVES: correlación, simulación, factor de inflación de la varianza, error estándar, tamaño de muestra

INTRODUCCIÓN

La regresión lineal es una de las técnicas más utilizadas en el análisis estadístico debido a su gran utilidad para estudiar la relación entre una variable dependiente y una o varias variables independientes, así como para la predicción, estimación y selección de variables apropiadas para el modelo, según lo mencionado por Paul (2006). A pesar de su amplia aplicabilidad, debe tomarse en cuenta que puede presentar limitaciones y desafíos, especialmente cuando se trabaja con datos complejos y heterogéneos. Uno de estos desafíos es la multicolinealidad, conocida como la dependencia lineal entre variables independientes en un modelo de regresión múltiple, o bien en palabras simples, se refiere a la información en una o más de estas variables es redundante.

Como señala Villegas (2017), la interrelación entre variables predictoras complica la capacidad de medir con exactitud el efecto individual de cada variable respuesta, lo que determina que las varianzas de los estimadores sean elevadas generando altas variabilidades en los estimadores. Por ello, cuando se ve una relación casi lineal entre las variables predictoras, se detecta la existencia de multicolinealidad, aunque esta no sea total.

Este estudio busca examinar el impacto de la multicolinealidad en la variabilidad de los coeficientes en modelos de regresión lineal, utilizando técnicas como el análisis del nivel de correlación. Se abordan tres casos específicos: en el primero, se explora una muestra pequeña, siguiendo la observación de Salmerón y Blanco (2016) acerca de cómo tamaños de muestra

[1] Estudiantes de Estadística de la Universidad de Costa Rica

ligeramente superiores a cero pueden aumentar la varianza del coeficiente, incrementando así la probabilidad de no rechazar la hipótesis nula en los contrastes de significancia individual en presencia de multicolinealidad grave. La identificación y comprensión de la multicolinealidad son esenciales para asegurar análisis de regresión confiables. En el segundo caso, se analiza una muestra mediana como punto intermedio para explorar los efectos que la multicolinealidad pueda generar, proporcionando información valiosa sobre cómo este fenómeno afecta los coeficientes en un escenario entre el primer y tercer caso. El tercer caso examina una muestra grande, ya que según Yu, Jiang y Land (2015), esto puede ayudar a reducir la inflación de errores estándar y disminuir la multicolinealidad. En última instancia, es crucial subrayar que el objetivo principal es analizar cómo la multicolinealidad afecta la variabilidad de los coeficientes en un modelo de regresión lineal mediante la simulación de diversos escenarios

METODOLOGÍA

Se llevó a cabo una simulación para explorar los efectos de la multicolinealidad en diferentes escenarios, la cual abarca la generación de datos, el ajuste de modelos de regresión y el análisis de los coeficientes, errores estándar y el Factor de Inflación de la Varianza (VIF, siglas en inglés). Se consideran tres casos distintos con diferentes tamaños de muestra: 30, 100 y 450 observaciones.

Para cada caso se evaluaron diversos escenarios que incluyeron variaciones en los niveles de correlación entre las variables. Este enfoque permitió comparar la variabilidad de los coeficientes, los errores estándar y el VIF de los distintos modelos, analizando así el efecto de la multicolinealidad. Se construyeron en total siete modelos diferentes, tres con una sola variable predictora, tres con combinaciones de dos de las variables independientes. Estos seis modelos de regresión se derivaron del siguiente modelo completo:

$$\mu_{Y|X_1,X_2,X_3} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$$

Las dos primeras variables (X_1, X_2) fueron construidas para presentar una alta correlación a través de una distribución normal bivariada. Sus parámetros fueron definidos con una media de 0, una desviación estándar de 1, un tamaño de muestra y un nivel de correlación que variaba según el escenario, mientras que la tercera variable (X_3) se generó a partir de una distribución normal estándar. La variable respuesta se construyó considerando los mismos coeficientes, un error distribuido normalmente y las variables predictoras, las cuales tenían una misma media y varianza, como se ilustra a continuación:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + e$$

Se llevó a cabo la simulación de los datos y se estimaron los coeficientes, así como los errores estándar de los diferentes modelos. Además, se realizó el cálculo del VIF. Se llevaron a cabo mil iteraciones para cada escenario, y sus resultados se almacenaron en matrices por separado. Posteriormente, se promediaron todos los coeficientes obtenidos para compararlos y observar cuánto variaron con respecto a los coeficientes originales. En cuanto a los errores estándar y los VIF, también se promediaron para analizar la inflación que sufrían debido a la presencia de multicolinealidad.

Se diseñó una función que requiere la introducción de varios parámetros para configurar un modelo según las preferencias del usuario. Esta función presenta los resultados de siete modelos distintos utilizando los datos generados. El primer modelo incorpora todas las variables, seguido por tres modelos (uno para cada variable). Luego, se presentan tres modelos adicionales con combinaciones que incluyen una de las dos variables altamente correlacionadas con la variable restante. Los parámetros necesarios para ejecutar la función se detallan en el Cuadro 1.

Cuadro 1

Parámetros para la función de multicolinealidad

Designación	Utilidad
n	Número de observaciones
k	Número de iteraciones
b0	Intercepto del modelo
b1	Coeficiente de X1
b2	Coeficiente de X2
b3	Coeficiente de X3
mux	Vector de medias de x
varx	Vector de variancias de x
corr	Correlación entre X1 y X2
vary	Variancia de Y

Se realizaron 9 diferentes escenarios, donde la correlación entre X_1 y X_2 variaba entre alta (0.99), media (0.45) y nula (0). Los primeros tres escenarios tenían 30 observaciones, luego en los siguientes tres escenarios se utilizaron 100 observaciones, por último los tres escenarios restantes se realizaron utilizando 450 observaciones. El sentido de esto era analizar la influencia que la cantidad de datos podría tener para corregir la multicolinealidad. El valor de los parámetros utilizados sin contar las observaciones, iteraciones y la correlación son los expuestos en el Cuadro 2. Cabe resaltar que los coeficientes fueron establecidos mediante un muestreo aleatorio siguiendo una distribución uniforme.

Cuadro 2

Parámetros utilizados

Coeficiente	Valor
mux	(0,0,0)
varx	(1,1,1)
vary	1
b_0	1.1
b_1	1.23
b_2	4.84
b_3	3.25

Finalmente, es relevante señalar que todos los análisis fueron realizados utilizando el lenguaje de programación especializado en Estadística R en su versión 4.3.1 (R Core Team, 2023), a través de la interfaz RStudio (RStudio Team, 2020). Asimismo, se emplearon los paquetes lattice (Sarkar D, 2008) y car (Fox, J., Weisberg, S., 2019).

RESULTADOS

Los resultados obtenidos en la simulación proporcionan información clave sobre cómo la multicolinealidad puede afectar a los coeficientes estimados, los errores estándar y el VIF en modelos de regresión lineal, los cuales se explicarán a continuación.

En cuanto a los coeficientes promedios obtenidos, se logra evidenciar un gran efecto causado por la presencia de multicolinealidad. Como se observa en el Cuadro 3, en los modelos donde están presentes todos los coeficientes, se observa como el promedio de coeficientes estimados son similares a las impuestas en la función. Cuando la correlación es media o alta, y si solo está presente uno de los coeficientes correlacionados, tienden a aumentar su tamaño con respecto al de los parámetros establecidos originalmente, especialmente en correlaciones altas.

Cuadro 3

Coeficientes promedios estimados para una muestra de 100 observaciones

CORRELACIÓN	COEFICIENTES	MOD	MOD1	MOD2	MOD3	MOD12	MOD13	MOD23
0.99	b_0	1,10	1,10	1,10	1,10	1,10	1,09	1,10
	b_1	1,28	6,03	-	-	1,28	6,02	-
	b_2	4,79	-	6,07	-	4,8	-	6,06
	b_3	3,25	-	-	3,27	-	3,25	3,25
0.45	b_0	1,09	1,09	1,11	1,08	1,10	1,08	1,10
	b_1	1,23	3,37	-	-	1,22	3,38	-
	b_2	4,84	-	5,38	-	4,84	-	5,39
	b_3	3,25	-	-	3,23	-	3,24	3,24
0.0	b_0	1,10	1,08	1,11	1,09	1,11	1,08	1,10
	b_1	1,23	1,25	-	-	1,23	1,25	-
	b_2	4,85	-	4,85	-	4,85	-	4,85
	b_3	3,25	-	-	3,25	-	3,25	3,24

Asimismo, al comparar estos coeficientes promedios obtenidos con los que se encuentran en los Cuadro 7 y 8 (ver en Anexos), conforme decrece o aumenta el tamaño de la muestra se puede observar un comportamiento similar. Por ende, no se visualiza una diferencia notable entre los modelos que tienen diferentes tamaños de muestra.

Con respecto al error estándar promedio de los coeficientes, se observa en los Cuadros 4, 5 y 9 (ver en Anexos) que, en el modelo completo y en el modelo donde solo están presentes las variables correlacionadas, cuando el nivel de correlación es alto, el error estándar de los coeficientes correlacionados tiende a ser mayores que al resto errores estándar de dicho modelo. Cuando el nivel de correlación es medio, este comportamiento se sigue dando, pero no difieren significativamente entre ellos; y cuando no hay correlación, los errores estándares tienen el mismo valor.

Asimismo, este comportamiento siempre está presente sin importar el tamaño de las muestras. No obstante, al comparar los datos proporcionados en los Cuadro 4 y 5, se evidencia que los valores del error estándar promedio obtenidos con correlación media y alta, tienden a ser mucho mayores cuando la muestra es pequeña, y decrecen cuando la muestra es grande. Por lo tanto, se demuestra que los errores estándar promedio de los coeficientes tienen una correlación inversa con el tamaño de la muestra.

Cuadro 4

Error estándar promedio estimado para una muestra de 30 observaciones

CORRELACIÓN	COEFICIENTES	MOD	MOD1	MOD2	MOD3	MOD12	MOD13	MOD23
0.99	b_0	0,19	0,65	0,63	1,14	0,64	0,23	0,19
	b_1	1,39	0,66	-	-	4,66	0,23	-
	b_2	1,39	-	0,64	-	4,66	-	0,19
	b_3	0,20	-	-	1,16	-	0,23	0,19
0.45	b_0	0,19	1,01	0,66	1,02	0,64	0,83	0,28
	b_1	0,22	1,04	-	-	0,73	0,86	
	b_2	0,22	-	0,68	-	0,74	-	0,29
	b_3	0,20	-	-	1,05	-	0,86	0,29
0.0	b_0	0,19	1,09	0,66	0,94	0,64	0,93	0,30
	b_1	0,20	1,12	-	-	0,65	0,96	-
	b_2	0,20	-	0,68	-	0,65	-	0,31
	b_3	0,20	-	-	0,97	-	0,96	0,31

Cuadro 5

Errores estándar estimados para una muestra de 450 observaciones

CORRELACIÓN	COEFICIENTES	MOD	MOD1	MOD2	MOD3	MOD12	MOD13	MOD23
0.99	b_0	0,05	0,16	0,16	0,29	0,16	0,06	0,05
	b_1	0,34	0,16	-	-	1,14	0,06	-
	b_2	0,34	-	0,16	-	1,14	-	0,05
	b_3	0,05	-	-	0,29	-	0,06	0,05
0.45	b_0	0,05	0,26	0,17	0,26	0,16	0,21	0,07
	b_1	0,05	0,26	-	-	0,18	0,21	-
	b_2	0,05	-	0,17	-	0,18	-	0,07
	b_3	0,05	-	-	0,27	-	0,21	0,07
0.0	b_0	0,05	0,28	0,17	0,24	0,16	0,23	0,07
	b_1	0,05	0,28	-	-	0,16	0,23	-
	b_2	0,05	-	0,17	-	0,16	-	0,07
	b_3	0,05	-	-	0,24	-	0,23	0,07

Con respecto al Factor de Inflación de la Varianza (VIF), según se puede observar en el Cuadro 6 que contiene los resultados obtenidos con una muestra de 100 observaciones, cuando se presenta una correlación alta, tiende a tener un valor muy elevado, tanto en el modelo completo como en el modelo donde solo están las variables correlacionadas. Sin embargo, en los modelos donde solo está una de las variables correlacionadas, el valor del VIF tiende a ser cercano o igual a 1.

Cuadro 6

VIF estimado para una muestra de 100 observaciones

CORRELACIÓN	VARIABLE	MOD	MOD12	MOD13	MOD23
0.99	X_1	52,28	51,70	1,01	-
	X_2	52,28	51,70	-	1,01
	X_3	1,02	-	1,01	1,01
0.45	X_1	1,28	1,27	1,01	-
	X_2	1,28	1,27	-	1,01
	X_3	1,02	-	1,01	1,01
0.0	X_1	1,02	1,01	1,01	-
	X_2	1,02	1,01	-	1,01
	X_3	1,02	-	1,01	1,01

Si comparamos por nivel de correlación estos resultados con los de las muestras de 30 y 450 observaciones (ver Cuadros 10 y 11 en Anexos), se puede observar una leve tendencia a que los valores promedio del VIF disminuyan conforme aumenta la muestra, pero la diferencia es muy pequeña.

Estos resultados son importantes, ya que permitieron mejorar la comprensión sobre el efecto que puede tener la multicolinealidad sobre la variabilidad de los coeficientes en un modelo de regresión lineal. También evidencian la relevancia de mitigar la multicolinealidad en análisis de este tipo.

CONCLUSIONES

En resumen, los resultados detallados en este estudio revelaron la sensibilidad de los modelos de regresión lineal ante la presencia de multicolinealidad. Se observó un impacto significativo en los coeficientes, errores estándar y el Factor de Inflación de la Varianza (VIF) cuando existía una correlación alta entre las variables predictoras.

La investigación reveló que la presencia de multicolinealidad influye de manera perceptible en los coeficientes de los modelos de regresión lineal. Aunque se notó un leve efecto en los coeficientes promedios, la falta de una tendencia clara en su comportamiento, incluso al variar el tamaño de la muestra, sugirió que la multicolinealidad puede afectar de manera compleja la interpretación de estos parámetros.

Respecto al error estándar, se evidenció que cuando había una correlación alta, se sufría una afectación en los coeficientes. También hubo una clara tendencia a decrecer su valor promedio conforme disminuía el nivel de correlación y aumentaba el tamaño de la muestra. Sin embargo, es importante tener en cuenta que en los casos donde el valor del error estándar es elevado causa que el intervalo de confianza para tales coeficientes aumente, si el error estándar supera de tamaño al valor absoluto del coeficiente, se invalida su variable y su utilidad en el modelo.

Es importante destacar que cuando el nivel de correlación era alto, el valor del VIF tendía a ser elevado, y cuando la muestra era grande, el VIF tendía a ser menor que cuando la muestra era pequeña. En el caso de los modelos con solo una variable correlacionada, el VIF tendió a ser cercano a 1. Conforme a Nahhas (2023), un valor cercano a 1 sería el ideal, indicando que la varianza no está inflada; no obstante, cuando el VIF es mayor a 10, puede señalar problemas con la varianza.

Según Villegas (2017) una de las posibles soluciones que se pueden implementar en casos como los expuestos en este trabajo, donde la multicolinealidad tiene una gran influencia en las estimaciones, es la eliminación de variables, esto con el fin de suprimir aquellas que están altamente correlacionadas. No obstante, en casos de muestras pequeñas donde la multicolinealidad, como se demostró, tiene una mayor influencia, eliminar variables conlleva la pérdida de información, por lo cual es mejor recurrir a otros métodos como el de análisis de componentes parciales.

En última instancia, estos hallazgos resaltan la importancia crítica de abordar la multicolinealidad en el análisis de regresión lineal, ya que su presencia puede no solo alterar la magnitud de los coeficientes, sino también impactar la interpretación y validez de los resultados. Estas conclusiones ofrecen un fundamento valioso para futuras investigaciones y refuerzan la necesidad de desarrollar estrategias efectivas para mitigar los efectos adversos de la multicolinealidad en el análisis de datos.

BIBLIOGRAFÍA

Fox J., Weisberg S. (2019). *An R Companion to Applied Regression*, tercera edición. Salvia, Thousand Oaks CA. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Nahhas, R. W. (2023, 21 noviembre). *5.19 Collinearity | Introduction to Regression Methods for Public Health using R*. https://www.bookdown.org/rwnahhas/RMPH/mlr-collinearity.html#generalized-vifs-when-at-least-one-predictor-is-categorical

Paul, R. K. (2006). Multicollinearity: Causes, effects and remedies. IASRI, New Delhi, 1(1), 58-65. http://apps.iasri.res.in/seminar/AS-299/ebooks/2005-2006/Msc/trim2/3.%20Multicollinearity-%20Causes,Effects%20and%20Remedies-Ranjit.pdf

R Core Team (2023). *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.* Versión: 4.3.1. https://www.R-project.org/

RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL. Versión: 4.3.1. http://www.rstudio.com/.

Salmerón, R. y Blanco, V. (2016). El problema de un tamaño muestral pequeño en la regresión lineal: micronumerosidad. Revista Electrónica de Comunicaciones y Trabajos de ASEPUMA, 17(2), 167-177. https://dialnet.unirioja.es/servlet/articulo?codigo=6466471

Sarkar D (2008). *Lattice: Multivariate Data Visualization with R. Springer, New York. ISBN 978-0-387-75968-5.* Versión: 0.21.8. http://lmdvr.r-forge.r-project.org

Villegas, D.A. (2017). Una contribución al estudio de la multicolinealidad en modelos de regresión lineal múltiple usando distribuciones de contorno elíptico [Tesis de Doctorado, Universidad Central de Venezuela]. http://saber.ucv.ve/handle/10872/18412

Yu, H., Jiang, S., & Land, K. C. (2015). Multicollinearity in hierarchical linear models. Social Science Research, 53, 118–136. https://doi.org/10.1016/j.ssresearch.2015.04.008

ANEXOS

Cuadro 7

Coeficientes promedios estimados para una muestra de 30 observaciones

CORRELACIÓN	COEFICIENTES	MOD	MOD1	MOD2	MOD3	MOD12	MOD13	MOD23
0.99	b_0	1,09	1,11	1,11	1,09	1,11	1,09	1,09
	b_1	1,17	6,00	-	-	0,90	6,03	-
	b_2	4,91	-	6,04	-	5,15	-	6,06
	b_3	3,26	-	-	3,22	-	3,27	3,26
0.45	b_0	1,12	1,11	1,13	1,10	1,13	1,09	1,11
	b_1	1,23	3,40	-	-	1,27	3,38	-
	b_2	4,85	-	5,39	-	4,83	-	5,40
	b_3	3,26	-	-	3,26	-		3,27
0.0	b_0	1,09	1,14	1,11	1,13	1,10	1,13	1,10
	b_1	1,23	1,18	-	-	1,23	1,17	-
	b_2	4,84	-	4,85	-	4,86	-	4,83
	b_3	3,25	-	-	3,27	-	3,27	3,25

Cuadro 8

Coeficientes promedios estimados para una muestra de 450 observaciones

CORRELACIÓN	COEFICIENTES	MOD	MOD1	MOD2	MOD3	MOD12	MOD13	MOD23
0.99	b_0	1,10	1,10	1,10	1,08	1,10	1,10	1,10
	b_1	1,23	6,02	-	-	1,21	6,02	-
	b_2	4,84	-	6,06	-	4,86	-	6,06
	b_3	3,25	-	-	3,25	-	3,25	3,25
0.45	b_0	1,10	1,10	1,10	1,09	1,10	1,10	1,10
	b_1	1,23	3,42	-	-	1,23	3,42	-
	b_2	4,84	-	5,40	-	4,84	-	5,40
	b_3	3,25	-	-	3,25	-	3,25	3,25
0.0	b_0	1,10	1,11	1,11	1,11	1,11	1,10	1,10
	b_1	1,24	1,24	-	-	1,23	1,24	-
	b_2	4,84	-	4,84	-	4,84	-	4,84
	b_3	3,25	-	-	3,25	-	3,25	3,25

Cuadro 9

Errores estándar estimados para una muestra de 100 observaciones

CORRELACIÓN	COEFICIENTES	MOD	MOD1	MOD2	MOD3	MOD12	MOD13	MOD23
0.99	b_0	0,10	0,35	0,34	0,62	0,34	0,12	0,10
	b_1	0,72	0,35	-	-	2,44	0,12	-
	b_2	0,72	-	0,34	-	2,44	-	0,10
	b_3	0,10	-	-	0,62	-	0,12	0,10
0.45	b_0	0,10	0,55	0,36	0,56	0,34	0,45	0,15
	b_1	0,11	0,55	-	-	0,38	0,45	-
	b_2	0,11	-	0,36	-	0,39	-	0,15
	b_3	0,10	-	-	0,56	-	0,45	0,15
0.0	b_0	0,10	0,59	0,36	0,51	0,34	0,50	0,16
	b_1	0,10	0,60	-	-	0,34	0,50	-
	b_2	0,10	-	0,36	-	0,34	-	0,16
	b_3	0,10	-	-	0,52	-	0,50	0,16

Cuadro 10

VIF estimado para una muestra de 30 observaciones

CORRELACIÓN	VARIABLE	MOD	MOD12	MOD13	MOD23
0.99	X_1	58,44	56,16	1,04	-
	X_2	58,41	55,16	-	1,04
	X_3	1,08	-	1,04	1,04
0.45	X_1	1,36	1,31	1,04	
	X_2	1,37	1,31	-	1,04
	X_3	1,08	-	1,04	1,04
0.0	X_1	1,08	1,04	1,04	-
	X_2	1,08	1,04	-	1,04
	X_3	1,08	-	1,04	1,04

Cuadro 11

VIF estimado para una muestra de 450 observaciones

CORRELACIÓN	VARIABLE	MOD	MOD12	MOD13	MOD23
	X_1	50,68	50,57	1,00	-
0.99	X_2	50,67	50,57	-	1,00
	X_3	1,00	-	1,00	1,00
	X_1	1,26	1,26	1,00	-
0.45	X_2	1,26	1,26	-	1,00
	X_3	1,00	-	1,00	1,00
	X_1	1,00	1,00	1,00	-
0.0	X_2	1,00	1	-	1
	X_3	1,00	-	1	1

Código de simulación

```
library(lattice),library(car), library(corrgram),
library(lmtest)
```

Librerías utilizadas en el código

```
Simmult <- function(n, k, b0, b1, b2, b3, mux, varx,
corr, vary) {
cov_matrix <- matrix(c(sqrt(varx[1])^2, corr *
sqrt(varx[1]) * sqrt(varx[2]), corr * sqrt(varx[1]) *
sqrt(varx[2]), sqrt(varx[2])^2), nrow = 2)
```

Definición de función para poner a prueba diferentes casos
Matriz de varianzas y covariancias para las variables normales bivariadas.

```
almacen_=matrix(nrow=k,ncol=4, );
colnames(almacen_)=c("bo","b1","b2","b3")
almavif=matrix(nrow=k,ncol=3);colnames(almavif)
=c("X1","X2","X3")
```

Creación almacenes de datos resultante de la todas las posibles combinaciones de coeficientes para diferentes modelos:
coeficientes y errores estándar
Almacén VIF

```
for (i in 1:k) {
data_bivariate <- MASS::mvrnorm(n = n, mu =
mux[1:2], Sigma = cov_matrix)
X1 <- data_bivariate[, 1]
X2 <- data_bivariate[, 2]
X3 <- rnorm(n, mean=mux[3], sqrt(varx[3]))
```

Generar datos bivariados con una distribución normal bivariada y una variable con distribución normal

```
Y=b0+b1*X1+b2*X2+b3*X3+rnorm(n,0,vary)
mod=lm(Y~X1+X2+X3)

almacoef[i,]=mod$coef
almaee[i,]=summary(mod)$coef[,2]
almavif[i,]=vif(mod)
}
```

Generar valores de y (basado en coef y ee)
Ajuste modelo completo para diferentes modelos
Almacenamiento de coeficientes de los diferentes modelos, desviación estándar y VIF

```
_prom=apply(alma_,2,mean)
```

Promediar coeficientes, errores estándar y VIF, usando cada almacén creado

```
 result=array(dim = c(3,7,7),dimnames =
list(c("Coef","E.E.","VIF"),
c("bo","b1","b2","b3","X1","X2","X3"),
c("MOD","MOD1","MOD2","MOD3","MOD12","M
OD13","MOD23")))
result[,,]=_prom
  result=round(result,2)
  result=ifelse(is.na(result),0,result)
  return(result)
}
```

Ordenar los resultados obtenidos al asignar coeficientes, errores estándar y VIF promedios a posiciones específicas dentro de result