

L U M B R E



Regresión múltiple y multicolinealidad

Grado: 4

Materia: Regresión y Modelado Estadístico · Probabilidad y Estadística

11/10/2026

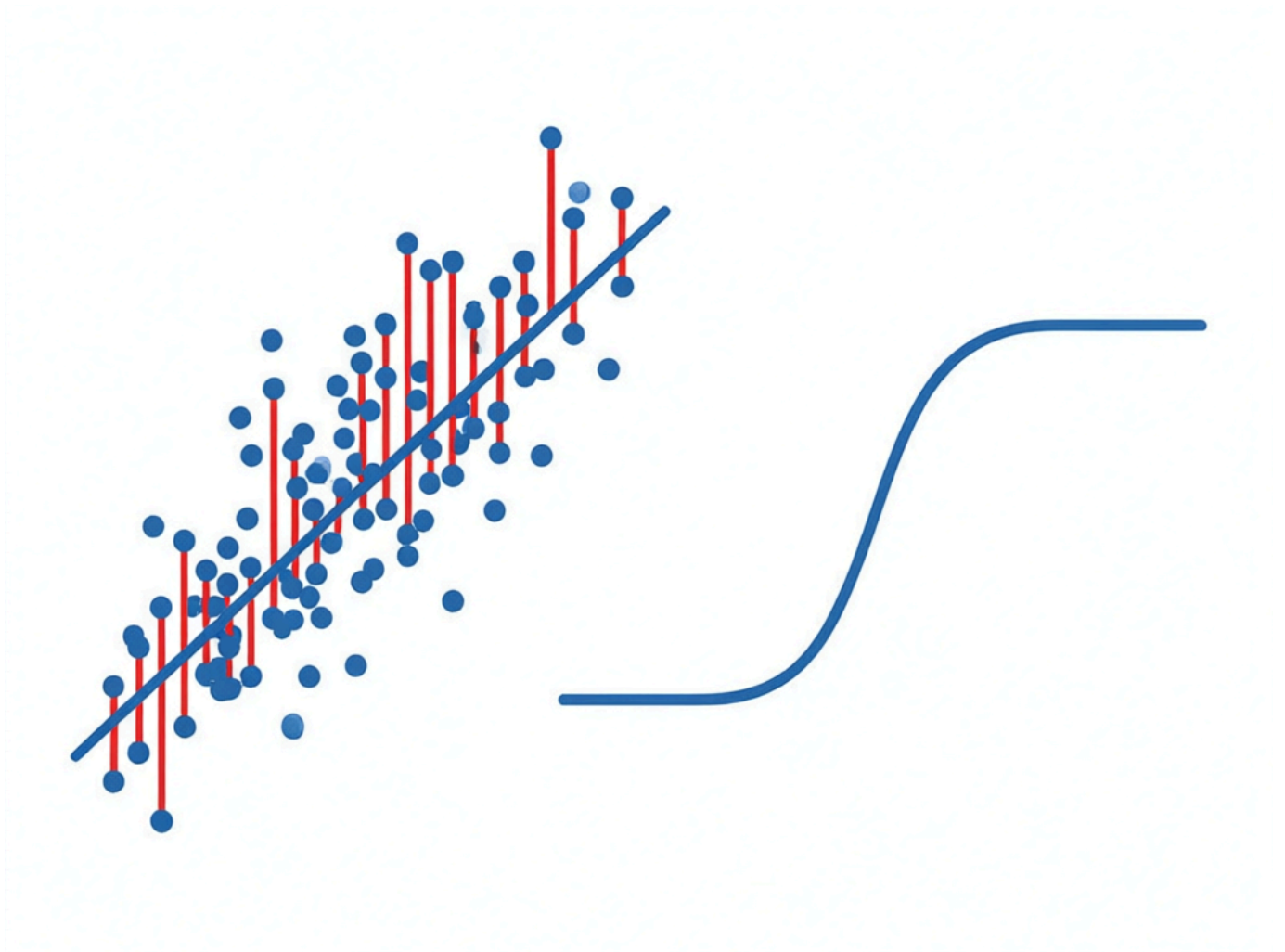
Antes de empezar

Añadir predictores convierte la recta en un plano y la pregunta en «el efecto de X_1 manteniendo fijos los demás». Ese «manteniendo fijos» es la potencia y a la vez el talón de Aquiles del modelo múltiple: cuando los predictores se parecen, los coeficientes se vuelven locos.

QUÉ APRENDERÁS

- Ampliar la regresión a varios predictores, interpretar coeficientes parciales y diagnosticar y tratar la multicolinealidad que los vuelve inestables.

Regresión múltiple y multicolinealidad



Regresión y modelado estadístico: lineal, múltiple y logístico.

Mapa conceptual

- **Regresión**
 - Lineal simple

- Beta 0 y 1
- R-cuadrado
- Residuos
- Supuestos
 - Normalidad
 - Homocedasticidad
 - Independencia
- Multiple
 - Multicolinealidad
 - Ajuste R²
 - AIC/BIC
- Logística
 - Odds ratio
 - Sigmoide
 - Regularización

El mapa mental interactivo es interactivo en la versión online.



El poder y el precio de controlar

La regresión múltiple permite afirmar el efecto de una variable «descontando» a las demás, que es justo lo que la hace valiosa para observar datos donde no se puede experimentar. Pero descuenta solo si los predictores no se solapan hasta hacer la operación imposible.

De la recta al plano

Con varios predictores, el modelo lineal deja de ser una recta y pasa a un hiperplano: una combinación lineal de todos ellos más un intercepto. Cada coeficiente mide la dirección e intensidad con que ese predictor contribuye a la respuesta.

El coeficiente parcial

El coeficiente de cada predictor es el cambio medio de la respuesta al subirlo una unidad manteniendo constantes los demás. Es un efecto «neto» o «controlado», la forma estadística de aislar la aportación propia de cada variable dentro del conjunto.

Interpretar «ceteris paribus»

Ese «manteniendo los demás fijos» es a la vez la gran virtud y una fuente de malentendidos. Un coeficiente puede cambiar de signo o tamaño al añadir o quitar predictores, porque lo que «controla» depende de con quién se le compare.

R cuadrada ajustada

La R cuadrada siempre sube al añadir predictores, incluso irrelevantes, por puro sobreajuste. La R cuadrada ajustada penaliza el número de variables, premiando solo las que aportan más de lo que cuestan en libertad perdida. Es la métrica honesta para comparar modelos de distinto tamaño.

El problema de la multicolinealidad

Cuando dos o más predictores están muy correlacionados entre sí, el modelo no sabe a cuál atribuir el mérito: la variación compartida podría explicarse por cualquiera. Los coeficientes individuales se vuelven inestables y sus errores estándar se inflan.

Por qué se inflan los errores

Estimar el efecto neto de un predictor exige separarlo de los que le acompañan. Si casi siempre van juntos, hay poca variación «independiente» que observe, y la incertidumbre sobre su papel propio se dispara, aunque la predicción conjunta siga siendo buena.

El factor de inflación de varianza

El VIF de un predictor mide cuánto se infla la varianza de su coeficiente por su correlación con los demás. Valores altos advierten de multicolinealidad. Pero ojo: el VIF señala un problema de interpretación, no necesariamente de predicción.

Colinealidad perfecta

Si un predictor es combinación lineal exacta de otros (una columna duplicada, o variables indicadoras que suman uno junto al intercepto), la matriz no se puede invertir y el modelo se rompe: la solución no es única. Es la forma extrema, y fácil de evitar con un diseño atento.

Consecuencias de la multicolinealidad

- Coeficientes grandes y cambiantes ante pequeñas modificaciones del modelo.
- Errores estándar inflados: variables relevantes salen «no significativas».
- Signos contraintuitivos que invierten la dirección esperada del efecto.

- Inestabilidad que arruina la interpretabilidad, aunque la predicción aguante.

Predecir frente a explicar

La multicolinealidad destruye la inferencia sobre coeficientes individuales, pero apenas daña la predicción: si los predictores viajan juntos también en datos nuevos, el modelo predice bien aunque no sepa atribuir. La pregunta que persigues decide cuánto te importa.

Diagnosticar la colinealidad

- Matriz de correlaciones entre predictores: pares casi perfectos saltan a la vista.
- VIF por variable: valores altos resumen la redundancia global.
- Observar coeficientes inestables o de signo raro al cambiar la especificación.
- Índice de condición: delata combinaciones de varias variables casi dependientes.

Remedios

- Eliminar predictores redundantes cuando uno es claramente superfluo.
- Combinar variables correlacionadas en un índice o mediante componentes principales.
- Regularizar (ridge) para estabilizar coeficientes inflados.
- Recolectar más o mejor variabilidad que desempate los predictores.

El peligro de eliminar por VIF

Borrar todo predictor con VIF alto no siempre es correcto: si las variables miden facetas de un mismo constructo, eliminar una distorsiona. Primero entender la estructura; la estadística señala, el dominio decide.

Variables de confusión controladas

La potencia real de la regresión múltiple en datos observacionales es incluir confusores para «restarlos» del efecto de interés. Pero controlar es delicado: meter una consecuencia de la respuesta en lugar de una causa empeora el cuadro. Elegir a controlar es tan crucial como el modelo.

Términos de interacción

Cuando el efecto de un predictor depende del nivel de otro, se añade un término producto que captura esa interacción. Sin él, el modelo impone que los efectos se suman, una restricción a menudo falsa que enmascara estructuras reales.

Errores frecuentes

- Interpretar coeficientes individuales sin mirar el VIF ni la estabilidad.
- Añadir todos los predictores posibles para maximizar la R cuadrada, sobreajustando.
- Controlar por variables que son consecuencia de la respuesta, no causa.

1. La multicolinealidad alta entre predictores afecta sobre todo a:

- ☐ La capacidad predictiva global del modelo
- ☐ La estabilidad e interpretación de los coeficientes individuales
- ☐ El cumplimiento de la linealidad
- ☐ La normalidad de los residuos

2. La R cuadrada nunca disminuye al añadir un predictor nuevo al modelo.

- ☐ Verdadero
- ☐ Falso

Estandarizar para comparar

Para comparar la aportación relativa de predictores en escalas distintas, conviene tipificarlos antes de entrar al modelo: los coeficientes quedan en unidades adimensionales comparables, sin que la escala bruta dicte quién parece más importante.

Menos a veces es más

Un modelo parsimonioso, con pocas variables bien elegidas, suele ser más robusto, interpretable y generalizable que uno atiborrado. La selección de predictores es un acto de criterio que equilibra bondad de ajuste contra complejidad.

ACTIVIDAD INTERACTIVA DISPONIBLE EN LUMBRE

3. Regresión múltiple



4. Ordena el abordaje de la multicolinealidad:

Elementos:

- diseñar predictores que no se solapen
- medir VIF y correlaciones
- juzgar según interés explicar o predecir
- eliminar, combinar o regularizar

Prever	. . .
Detectar	. . .
Decidir	. . .
Tratar	. . .

Interpretable ML <https://christophm.github.io/interpretable-ml-book/> Interpretación de coeficientes y problemas de colinealidad.

VIDEO

Video complementario

Recurso audiovisual para reforzar los conceptos.



Escanea para verlo en Lumbre

<https://www.youtube.com/watch?v=1wK1K2M7tXI>

5. Dos predictores muy correlacionados: tu jefe quiere saber «cuál importa más». ¿Qué le explicas sobre la multicolinealidad y qué harías?

NIST: Regression <https://www.itl.nist.gov/div898/handbook/pmd/section1/pmd16.htm>
Referencia completa de metodos de regresion.

Referencia curricular

Grado:	4
Materia:	Regresión y Modelado Estadístico · Probabilidad y Estadística

Versión interactiva completa (con actividades, videos y autoevaluación):



<https://univ.lumbre.lat/object/cdia-reg-003-regresion-multiple-y-multicolinealidad>

Material educativo generado desde Lumbre · lumbre.lat