Notas de Econometría

Pavel Solís
2026

3 Análisis de Regresión Múltiple: Estimación

Sabemos cómo explicar una variable dependiente y en función de una variable independiente x con modelo RLS

- Pero es difícil hacer conclusiones ceteris paribus de cómo x afecta a y
- Supuesto principal (RLS.4) es poco realista
  - RLS.4: Todos los otros factores que afectan a y no se correlacionan con x

El modelo de regresión lineal múltiple (RLM) nos permite hacer análisis ceteris paribus porque podemos controlar explícitamente por otros factores que afectan simultáneamente a la variable dependiente
RLM permite usar datos no experimentales para:

- Verificar teorías económicas
- Evaluar políticas públicas

Modelo RLM permite incluir muchas variables explicativas que pueden estar correlacionadas

- Podemos aspirar a inferir causalidad en casos en que una RLS no es apropiada
- Permite explicar más de la variabilidad en y
  - Mejores modelos para predecir y
- Permite considerar formas funcionales más generales (flexibilidad)
- Vehículo más usado para el análisis empírico en ciencias sociales

¿Qué sigue?

- Estimación y propiedades estadísticas de MCO (falta de sesgo, eficiencia) para RLM

3.1 Motivación y definición

RLM se puede usar para resolver problemas que no se pueden resolver con RLS

- Factores diferentes a la variable de interés se pueden incluir en el modelo de regresión
- RLM permite generalizar las relaciones funcionales entre las variables

3.1.1 Modelo con 2 variables independientes

y es determinada por 2 variables independientes y por otros factores no observados contenidos en u

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$

- β_0 es el parámetro del intercepto o la constante
- β_1 y β_2 son los parámetros de las pendientes
  - Aún estamos interesados en β_1
- Sacamos a x_2 de u y la ponemos explícitamente en la ecuación
  - β_1 mide el efecto de x_1 en y, manteniendo x_2 fija (ceteris paribus)
  - β_2 mide el efecto ceteris paribus de x_2 en y
- Antes, x_2 no debía estar correlacionada con x_1 (o β_1 estaría sesgado)

Ejemplos.

$$salario = \beta_0 + \beta_1 educ + \beta_2 exper + u$$

$$consumo = \beta_0 + \beta_1 ingreso + \beta_2 ingreso^2 + u$$

- Ambos ejemplos se pueden expresar en la forma general para 2 variables independientes
  - Pero la interpretación cambia
- El segundo modelo tiene una sola variable independiente pero aparece de dos formas distintas (3 parámetros)

$$\frac{\Delta consumo}{\Delta ingreso} \approx \beta_1 + 2\beta_2 \, ingreso$$

  - Efecto marginal del ingreso sobre el consumo
  - Depende de β_1, β_2 y el ingreso

Necesitamos supuestos de cómo u se relaciona con x_1 y x_2

- Supuesto principal en el modelo de 2 variables: $\mathbb{E}(u \mid x_1, x_2) = 0$
  - Para cualesquier valores de x_1 y x_2 en la población, el promedio de los factores no observados es cero

Ejemplos.

$$\mathbb{E}(u \mid educ, exper) = 0$$

- Promedio de habilidad (en u) es el mismo en todas las combinaciones de educación y experiencia

$$\mathbb{E}\left(u \mid ingreso, ingreso^2\right) = 0$$

- Es lo mismo que $\mathbb{E}(u \mid ingreso) = 0$ porque $ingreso^2$ se conoce al conocer *ingreso*

3.1.2 Modelo con k variables independientes

RLM permite considerar varios factores que afectan a y

- Ej. Para analizar el salario, podemos considerar además: capacitación, antigüedad, variables demográficas

El modelo de RLM o **modelo de regresión múltiple** se escribe

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u$$

- y es la variable dependiente, explicada, etc.
- $x_1, x_2, \ldots, x_k$ son las variables explicativas, independientes, etc.
- Contiene $k+1$ parámetros poblacionales no conocidos
  - β_0 es el intercepto
  - $\beta_1, \beta_2, \ldots, \beta_k$ son los parámetros de las pendientes asociadas con las x
- u es el término de error, contiene factores que afectan a y distintos a $x_1, x_2, \ldots, x_k$
  - Siempre habrá factores que no se puedan incluir

La interpretación de los parámetros es importante

Ejemplo.

$$\log(salario) = \beta_0 + \beta_1 \log(ventas) + \beta_2 ceoantig + \beta_3 ceoantig^2 + u$$

- $k = 3$
- β_1: elasticidad del salario con respecto a las ventas ceteris paribus
- Si $\beta_3 = 0$, $(100 \times \beta_2)$ es aproximadamente el incremento porcentual ceteris paribus en el salario cuando *ceoantig* aumenta 1 año
- Si $\beta_3 \neq 0$, el efecto de *ceoantig* en el salario es

$$\frac{\Delta \log(salario)}{\Delta ceoantig} \approx \beta_2 + 2\beta_3\, ceoantig$$

Recordatorio: Lineal en RLM significa lineal en los parámetros β_j

- Muchas aplicaciones de RLM involucran relaciones no lineales entre las variables

Supuesto clave para el modelo RLM:

$$\mathbb{E}\left(u \mid x_1, x_2, \ldots, x_k\right) = 0$$

- Clave para probar falta de sesgo en MCO

- Todos los factores en u no están correlacionados con las variables explicativas
- Cualquier problema que cause correlación entre u y cualquier variable independiente hace que el supuesto no se cumpla
- La **función de regresión poblacional** (FRP) es:

  $\mathbb{E}(y \mid x_1, x_2, \ldots, x_k) = \mathbb{E}(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + u \mid \vec{x}) = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k$

3.2 Mecánica e interpretación de MCO

En RLM:

- Tenemos k variables independientes
- Buscamos $k+1$ estimadores $\widehat{\beta}_0, \widehat{\beta}_1, \ldots, \widehat{\beta}_k$ para los parámetros $\beta_0, \beta_1, \ldots, \beta_k$

3.2.1 Obtención de los estimadores MCO

La ecuación MCO estimada, la línea de regresión MCO o la **función de regresión muestral** (FRM) es

$$\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x_1 + \widehat{\beta}_2 x_2 + \ldots + \widehat{\beta}_k x_k$$

- $\widehat{\beta}_0$ es el estimador MCO para el intercepto
- $\widehat{\beta}_1, \widehat{\beta}_2, \ldots, \widehat{\beta}_k$ son los estimadores MCO para las pendientes

Problema de minimización

El método MCO escoge los estimadores $(\widehat{\beta})$ que minimizan la suma de los residuales al cuadrado dada una muestra de n observaciones $\{(y_i, x_{i1}, x_{i2}, \ldots, x_{ik}) \mid i = 1, 2, \ldots, n\}$

$$\min \sum_{i=1}^{n} \widehat{u}_i^2 = \min \sum_{i=1}^{n} (y_i - \widehat{y})^2 = \min \sum_{i=1}^{n} \left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_{i1} - \ldots - \widehat{\beta}_k x_{ik}\right)^2$$

- En x_{ij}, i hace referencia a la observación y j a la variable independiente
  - Ej. $educ_i, exper_i$
- La suma es sobre todas las observaciones

Este problema de minimización se resuelve con cálculo de varias variables para obtener las *condiciones de primer orden* de MCO

$$\sum_{i=1}^{n} \widehat{u}_i = \sum_{i=1}^{n} \left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_{i1} - \ldots - \widehat{\beta}_k x_{ik}\right) = 0$$

$$\sum_{i=1}^{n} x_{ij} \widehat{u}_i = \sum_{i=1}^{n} x_{ij} \left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_{i1} - \ldots - \widehat{\beta}_k x_{ik}\right) = 0 \quad \forall\, j = 1, 2, \ldots, k$$

- Se aplica la derivada parcial con respecto a cada $\widehat{\beta}$
  - Se obtienen $k+1$ ecuaciones lineales con $k+1$ incógnitas $\widehat{\beta}_0, \widehat{\beta}_1, \ldots, \widehat{\beta}_k$

- También se pueden obtener como momentos muestrales de $\mathbb{E}(u) = 0$ y $\mathbb{E}(x_j u) = 0$

Solución

$$\widehat{\beta} = (\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'y$$

- Formas de obtener las $\widehat{\beta}$'s: a mano, con algebra lineal, con paquete estadístico
- Es más importante saber cómo interpretarlas
- Requiere suponer que las ecuaciones se pueden resolver de forma única
- Terminología: Corremos una regresión MCO (o regresamos o proyectamos) y sobre $x_1, x_2, \ldots, x_k$
  - Indica que el método MCO se utilizó para obtener la línea de regresión
- Generalmente siempre estimamos el intercepto junto con las pendientes

3.2.2 Interpretación de la ecuación de regresión MCO

La línea de regresión MCO expresada en cambios es

$$\Delta\widehat{y} = \widehat{\beta}_1 \Delta x_1 + \widehat{\beta}_2 \Delta x_2 + \ldots + \widehat{\beta}_k \Delta x_k$$

- Indica la predicción del cambio en y dados cambios en las x's
- $\widehat{\beta}_0$ es la predicción de y cuando $x_1 = x_2 = \ldots = x_k = 0$
  - No siempre tiene sentido pero es necesaria

Si $\Delta x_2 = \Delta x_3 = \ldots = \Delta x_k = 0$, entonces

$$\Delta\widehat{y} = \widehat{\beta}_1 \Delta x_1$$

- $\widehat{\beta}_1$ mide el cambio en y cuando x_1 aumenta en 1 unidad, manteniendo todas las otras variables fijas
- Al incluir $x_2, x_3, \ldots, x_k$ en el modelo, obtenemos un coeficiente para x_1, con una interpretación ceteris paribus
- **Controlamos** por las variables $x_2, x_3, \ldots, x_k$ cuando estimamos el efecto de x_1 en y
- Misma interpretación para las otras $\widehat{\beta}_j$'s

Todas las $\widehat{\beta}_j$'s tienen una interpretación de efecto parcial o ceteris paribus

- Por eso RLM es tan útil

Ejemplo. Queremos entender las calificaciones universitarias en términos de las calificaciones de preparatoria (en escala de 4 puntos) y el examen de admisión

$$\widehat{caluni} = 1.29 + 0.453 calprepa + 0.0094 examadmi, \quad n = 141$$

- 1.29 es la predicción de *caluni* si $calprepa = examadmi = 0$
  - No es realista
- Relación positiva entre *caluni* y *calprepa*, manteniendo *examadmi* fijo
- Interpretación de 0.453:
  - Un punto adicional en *calprepa* está asociado con 0.453 de punto en *caluni*
  - Es la mejor predicción para el caso de dos alumnos con mismo *examadmi* pero $calprepaA = calprepaB + 1$, $caluniA = caluniB + 0.453$
- Interpretación de 0.0094:
  - Manteniendo *calprepa* fijo, si *examadmi* aumenta 10 puntos, *caluni* aumenta 0.09 puntos
    * No significativo porque para *examadmi* media = 36 y desvest = 24

Ejemplo. Queremos analizar el efecto de la educación en el salario

$$\log(\widehat{salario}) = 0.284 + 0.092educ + 0.0041exper + 0.022antig, \quad n = 526$$

Interpretación de coeficientes en porcentaje y como ceteris paribus

- Interpretación de 0.092:
  - Dejando *exper* y *antig* fijos, 1 año extra de *educ*, aumenta $\log(salario)$ en 0.092 y el *salario* en 9.2%
  - Diferencia proporcional en la predicción del salario con misma *exper* y *antig* pero con diferencia en *educ* de 1 año
  - Estimado mantiene 2 factores de productividad fijos
- Para saber si 0.092 es un buen estimado del rendimiento ceteris paribus de 1 año más de *educ*, necesitamos propiedades estadísticas de MCO

3.2.3 Significado de "mantener otros factores fijos" en RLM

$\widehat{\beta}_j$'s miden el cambio esperado en la variable dependiente $\Delta\widehat{y}$ manteniendo fijas las otras variables independientes

- Los datos vienen de una muestra aleatoria sin restricciones en los valores que pueden tomar las variables independientes
- Al dar a $\widehat{\beta}_j$ una interpretación de efecto parcial, es como si hubiéramos obtenido una muestra aleatoria con los mismos valores para las otras variables independientes y solo hubiera variación en la variable independiente de interés
  - En cuyo caso, hubiéramos corrido una RLS entre la variable dependiente y la variable independiente de interés

El poder de RLM es que nos permite:

- Hacer en ambientes no experimentales, lo que los científicos pueden hacer en un ambiente controlado de laboratorio (mantener otros factores fijos)
- Imitar esa situación sin restringir los valores de ninguna variable independiente
- Dar una interpretación ceteris paribus aunque los datos no se hayan recolectado de esa forma

3.2.4 Cambiar más de una variable independiente al mismo tiempo

Si cambia más de una variable a la vez, podemos utilizar la línea de regresión MCO expresada en cambios

$$\Delta\widehat{y} = \widehat{\beta}_1\Delta x_1 + \widehat{\beta}_2\Delta x_2 + \ldots + \widehat{\beta}_k\Delta x_k$$

Ejemplo. Para estimar el efecto en el salario de una persona que se queda en la empresa 1 año más (más experiencia y antigüedad, misma educación)

$$\Delta\log(\widehat{salario}) = 0.0041\Delta exper + 0.022\Delta antig = 0.0041 + 0.022 = 0.0261 \text{ o } 2.61\%$$

3.2.5 Valores ajustados y residuales MCO

Valor ajustado o predicho para la observación i:

$$\widehat{y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_{i1} + \ldots + \widehat{\beta}_k x_{ik}$$

- Usamos los valores de las variables independientes para cada observación i en la línea de regresión MCO
- En general, $y_i \neq \widehat{y}_i$ porque MCO minimiza el promedio del error de predicción al cuadrado
  - No dice nada del error de predicción individual

Residual para la observación i (como en RLS): $\widehat{u}_i = y_i - \widehat{y}_i \quad \forall\, i$

- Si $\widehat{u}_i > 0$, la línea de regresión subestima y_i
- Si $\widehat{u}_i < 0$, la línea de regresión sobreestima y_i

Propiedades de los valores ajustados y residuales (vienen de las condiciones de primer orden y son extensiones de RLS)

1. El promedio muestral de los residuales es cero
   - $\frac{1}{n}\sum_{i=1}^{n}\widehat{u}_i = 0 \implies \bar{y}_i = \bar{\widehat{y}}_i$
2. La covarianza muestral entre cada variable independiente y los residuales MCO es cero
   - $\text{Cov}\,(x_{ij}, \widehat{u}_i) = 0 \implies \text{Cov}\,(\widehat{y}_i, \widehat{u}_i) = 0$
3. El punto $(\bar{y}, \bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k)$ siempre está sobre la línea de regresión MCO
   - $\bar{y} = \widehat{\beta}_0 + \widehat{\beta}_1\bar{x}_1 + \widehat{\beta}_2\bar{x}_2 + \ldots + \widehat{\beta}_k\bar{x}_k \implies \bar{y}_i = \bar{\widehat{y}}_i$

3.2.6 Interpretación de efecto parcial en RLM

Para algunas derivaciones, necesitamos fórmulas explícitas para las $\widehat{\beta}_j$'s

- También ayudan a entender el funcionamiento de MCO

Modelo con 2 variables independientes
Cuando $k = 2$, $\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x_1 + \widehat{\beta}_2 x_2$ con interés en $\widehat{\beta}_1$

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n} \widehat{r}_{i1} y_i}{\sum_{i=1}^{n} \widehat{r}_{i1}^2}$$

- $\widehat{r}_{i1}$ son los residuales MCO de una RLS de x_1 sobre x_2 (y no se ocupa)
- Después, corremos una RLS de y sobre $\widehat{r}_{i1}$
  - $\widehat{\beta}_1$ aquí es igual que en RLS porque $\bar{\widehat{r}}_{i1} = 0$
    * Revisa regresión por el origen en RLS

Permite interpretar el efecto parcial de $\widehat{\beta}_1$ de otra forma

- Los residuales $\widehat{r}_{i1}$ son la parte de x_{i1} que no esta correlacionada con x_{i2}
  - $\widehat{r}_{i1}$ es x_{i1} después de depurar los efectos de x_{i2}
- Entonces, $\widehat{\beta}_1$ mide la relación en la muestra entre y y x_1 después de depurar x_2
  - $\widehat{\beta}_1$ mide el cambio en y cuando x_1 aumenta en 1 unidad manteniendo x_2 fija
- Esto no existe en RLS porque no se incluyen otras variables independientes en el modelo

Modelo con k variables independientes
Cuando hay k variables independientes, $\widehat{r}_{i1}$ viene de la regresión de x_1 sobre $x_2, x_3, \ldots, x_k$ y $\widehat{\beta}_1$ se escribe igual que para el caso $k = 2$

- $\widehat{\beta}_1$ mide el efecto de x_1 sobre y después de depurar los efectos de $x_2, x_3, \ldots, x_k$

3.2.7 Comparación de los estimados de regresión simple y múltiple

¿Cuándo se obtiene el mismo estimado para el efecto de x_1 sobre y con RLS y RLM?
Modelo con 2 variables independientes

- RLS: $\widetilde{y} = \widetilde{\beta}_0 + \widetilde{\beta}_1 x_1$
- RLM: $\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x_1 + \widehat{\beta}_2 x_2$

La relación entre $\widetilde{\beta}_1$ y $\widehat{\beta}_1$ está dada por:

$$\widetilde{\beta}_1 = \widehat{\beta}_1 + \widehat{\beta}_2 \widetilde{\delta}_1$$

- $\widetilde{\beta}_1$ no nos da el efecto parcial de x_1 sobre $\widehat{y}$
- $\widetilde{\delta}_1$ es el coeficiente de la pendiente en la RLS de x_{i2} sobre x_{i1}, $i = 1, 2, \ldots, n$

- $x_2 = \delta_0 + \delta_1 x_1 + \varepsilon$, de donde se obtienen $\widetilde{\delta}_0$ y $\widetilde{\delta}_1$
- $\widetilde{\delta}_1 = \text{Cov}(x_1, x_2) / \text{Var}(x_1)$

- $\widehat{\beta}_2 \widetilde{\delta}_1$ se conoce como el término distorsionador
  - Efecto parcial de x_2 sobre $\widehat{y}$ por la pendiente de la RLS de x_2 sobre x_1

Entonces, $\widetilde{\beta}_1 = \widehat{\beta}_1$ en dos casos:

1. Cuando $\widehat{\beta}_2 = 0$: Efecto parcial de x_2 sobre $\widehat{y}$ es cero en la muestra
2. Cuando $\widetilde{\delta}_1 = 0$: x_1 y x_2 no están correlacionadas en la muestra

En general, $\widetilde{\beta}_1 \neq \widehat{\beta}_1$, pero la relación nos sirve para caracterizar cuando son similares (o muy diferentes)

- Si $\widehat{\beta}_2 \approx 0$ o si $\widetilde{\delta}_1 \approx 0$, esperaríamos que $\widetilde{\beta}_1 \approx \widehat{\beta}_1$

Modelo con k variables independientes
Los estimados de β_1 son idénticos en el caso de k variables independientes si [poco probable]

- Los coeficientes MCO para $x_2, \ldots, x_k$ son todos cero

$$\widehat{\beta}_2 = \ldots = \widehat{\beta}_k = 0$$

- x_1 no esta correlacionada con ninguna de las variables $x_2, \ldots, x_k$

$$\widetilde{\delta}_2 = \ldots = \widetilde{\delta}_k = 0$$

Pero $\widetilde{\beta}_1$ (de RLS) y $\widehat{\beta}_1$ (de RLM) pueden ser similares si

- Los coeficientes para $x_2, \ldots, x_k$ son pequeños
- Las correlaciones entre x_1 y las otras variables independientes son pequeñas

3.2.8 Bondad de ajuste

Como en RLS, definimos la variación de cada una de las partes:

- Suma de cuadrados total: $\text{SCT} = \sum_{i=1}^{n} (y_i - \bar{y})^2$
  - Variación de y_i (suponemos que no es constante, $y_i \neq c \quad \forall\, i$)
- Suma de cuadrados explicada: $\text{SCE} = \sum_{i=1}^{n} (\widehat{y}_i - \bar{y})^2$
  - Variación de $\widehat{y}_i$
- Suma de cuadrados de los residuales: $\text{SCR} = \sum_{i=1}^{n} \widehat{u}_i^2$
  - Variación de $\widehat{u}_i$

La variación total en y siempre se puede expresar como: $\text{SCT} = \text{SCE} + \text{SCR}$
Si $\text{SCT} \neq 0$,

$$1 = \frac{\text{SCE}}{\text{SCT}} + \frac{\text{SCR}}{\text{SCT}} \implies \text{R}^2 = \frac{\text{SCE}}{\text{SCT}} = 1 - \frac{\text{SCR}}{\text{SCT}}$$

- Interpretación de R^2: Proporción de la variación muestral en y_i que es explicada por la línea de regresión MCO
- $0 \leq R^2 \leq 1$ por definición
- R^2 es igual al coeficiente de correlación entre y_i y $\widehat{y}_i$ elevado al cuadrado

$$R^2 = [\text{Corr}\,(y_i, \widehat{y}_i)]^2 = \frac{[\sum(y_i - \bar{y})(\widehat{y}_i - \bar{\widehat{y}})]^2}{[\sum(y_i - \bar{y})^2][\sum(\widehat{y}_i - \bar{\widehat{y}})^2]}$$

- R^2 nunca decrece (generalmente aumenta) cuando se agrega cualquier variable independiente a la regresión (SCR nunca aumenta)
  - Mala herramienta para decidir si agregar variables al modelo
  - Una variable independiente va en el modelo si tiene un efecto parcial diferente de cero sobre y en la población
- R^2 además de medida de bondad de ajuste sirve para probar si un grupo de variables es importante para explicar y

Ejemplo.

$$\widehat{caluni} = 1.29 + 0.453 calprepa + 0.0094 examadmi, \quad n = 141, \quad R^2 = 0.176$$

- Significa que *calprepa* y *examadmi* explican alrededor de 17.6% de la variación en *caluni* en esta muestra
- Hay muchos otros factores (personalidad, calidad de educación en prepa, etc.) que contribuyen a *caluni*

Una R^2 baja:

- No significa que la ecuación no sirva
  - $\widehat{\beta}_j$'s pueden ser estimados confiables de los efectos ceteris paribus de cada variable independiente sobre y, eso no depende del valor de R^2
- Significa que es difícil predecir valores individuales de y con exactitud
  - En ciencias sociales, es difícil predecir comportamientos individuales

3.2.9 Regresión por el origen

Teoría económica o sentido común pueden sugerir que $\beta_0 = 0$,

$$\widetilde{y} = \widetilde{\beta}_1 x_1 + \widetilde{\beta}_2 x_2 + \ldots + \widetilde{\beta}_k x_k$$

- Si $x_2 = x_3 = \ldots = x_k = 0$, la predicción para $\widetilde{y}$ es 0
- $\widetilde{\beta}_j$'s son los estimados MCO de la regresión de y sobre $x_1, x_2, \ldots, x_k$ por el origen
- $\widetilde{\beta}_j$'s minimizan SCR pero fijando $\beta_0 = 0$

Las propiedades de MCO anteriores no se cumplen en regresión por el origen

1. Los residuales MCO ya no generan un promedio muestral cero
   - $\frac{1}{n}\sum_{i=1}^{n}\widehat{u}_i \neq 0,\ \bar{y}_i \neq \bar{\widetilde{y}}_i$
2. $\mathrm{R}^2 < 0$ es posible
   - Significado: $\bar{y}$ explica más de la variación en y que las variables explicativas
   - Alternativa 1: Incluir un intercepto
   - Alternativa 2: Concluir que las x's explican mal a y
3. Si se utiliza R^2 como $[\mathrm{Corr}\,(y_i, \widetilde{y}_i)]^2$, $\mathrm{R}^2 \geq 0$
   - No hay regla sobre cómo reportar R^2 en regresión por el origen

Si $\beta_0 \neq 0$ en la población, las $\widetilde{\beta}_j$'s estarán sesgadas

- Costo de estimar β_0 cuando $\beta_0 = 0$: Varianzas de las $\widetilde{\beta}_j$'s serán más grandes

3.3 Valor esperado de los estimadores MCO

Ahora estudiamos las propiedades estadísticas de los estimadores MCO

- No se refiere al caso de una muestra particular
- Para parámetros del modelo poblacional
- Supuesto: Podemos tomar muestras aleatorias de forma repetida

Las propiedades de interés son:

- Valor esperado
  - 4 supuestos (extensiones de RLS) permiten mostrar falta de sesgo
- Varianza
  - 1 supuesto adicional permite determinar eficiencia

Partimos de supuestos ideales para obtener propiedades deseables

3.3.1 Falta de sesgo en MCO

Supuesto RLM.1. El modelo es lineal en parámetros (define el modelo)

El modelo poblacional se puede escribir como

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u$$

- $\beta_0, \beta_1, \ldots, \beta_k$ son parámetros (constantes) de interés no conocidos
- u es un error aleatorio no observado

Observaciones sobre el supuesto RLM.1:

- Se conoce como el modelo poblacional o modelo verdadero
  - Incertidumbre: Podemos estimar un modelo diferente
- Clave: Modelo es lineal en los parámetros
  - Flexible porque permite capturar relaciones no lineales
  - x y y pueden ser funciones de variables de interés

Supuesto RLM.2. La muestra es aleatoria

Tenemos una muestra aleatoria de n observaciones $\{(y_i, x_{i1}, x_{i2}, \ldots, x_{ik}) \mid i = 1, 2, \ldots, n\}$ del modelo poblacional en RLM.1
Observaciones sobre el supuesto RLM.2:

- Para una observación particular sacada de la población:

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik} + u_i \quad \forall\, i$$

  - Primer subíndice indica el número de la observación y el segundo subíndice denota el número de la variable

Ejemplo.

$$\log(salario_i) = \beta_0 + \beta_1 \log(ventas_i) + \beta_2 ceoantig_i + \beta_3 ceoantig_i^2 + u_i$$

- u_i contiene factores no observados para el CEO i que afectan su salario

- En aplicaciones, el modelo se escribe generalmente en forma poblacional
  - Enfatiza que queremos estimar una relación poblacional
- Aquí $\widehat{\beta}_0, \widehat{\beta}_1, \ldots, \widehat{\beta}_k$ son estimadores de $\beta_0, \beta_1, \ldots, \beta_k$
  - Antes, para una muestra particular $\widehat{\beta}_j$'s eran escogidos tal que $\sum_{i=1}^n \widehat{u}_i = 0$ y $\sum_{i=1}^n x_{ij}\widehat{u}_i = 0 \quad \forall\, j \qquad [\mathrm{Cov}\,(x_{ij}, \widehat{u}_i) = 0 \quad \forall\, j]$
  - Pero no incluimos condiciones para que los estimadores MCO estuvieran bien definidos para una muestra

Supuesto RLM.3. No colinealidad perfecta (para poder calcular MCO)

En la muestra (y por tanto en la población),

- Ninguna de las variables independientes es constante, y
- No hay relaciones lineales exactas entre las variables independientes

Observaciones sobre el supuesto RLM.3:

- Es más complicado que en RLS porque debemos evaluar la relación entre todas las variables independientes
- Si alguna variable independiente es una combinación lineal exacta de las otras variables independientes, el modelo sufre de **colinealidad perfecta**
  - En ese caso, no se puede estimar por MCO porque $(\mathrm{X}'\mathrm{X})^{-1}$ no existe
- RLM.3 permite que las variables independientes estén correlacionadas, solo no pueden estar perfectamente correlacionadas
  - Se puede verificar revisando que $\mathrm{Corr}(x_s, x_j) \neq 1 \quad \forall\, s, j$
  - Si no se permitiera ninguna correlación, RLM sería poco útil
  - De hecho, la razón de incluir una variable independiente es que puede estar correlacionada con la de interés y queremos fijar su efecto
- RLM.3 solo descarta correlación perfecta
  - Caso poco probable con una muestra aleatoria

Ejemplo. RLM3 no se cumple si una variable es múltiplo de otra

$$y = consumo,\ x_1 = ingreso \text{ en pesos},\ x_2 = ingreso \text{ en miles de pesos}$$

- En ese caso, no se puede realizar un análisis ceteris paribus
- Funciones no lineales de la misma variable se permiten

$$x_1 = ingreso,\ x_2 = ingreso^2$$

Preguntas. ¿Cuál es el problema de estimar la siguiente ecuación?

$$\log(consumo) = \beta_0 + \beta_1 \log(ingreso) + \beta_2 \log(ingreso^2) + u$$

- ¿Qué solución se podría implementar?

Preguntas. En una campaña electoral con 2 candidatos (A y B), ¿cuál sería el problema de estimar la siguiente ecuación?

$$votoA = \beta_0 + \beta_1 gastoA + \beta_2 gastoB + \beta_3 gastototal + u$$

- ¿Tiene sentido un análisis ceteris paribus?
- ¿Qué solución se podría implementar?

Supuesto RLM.3 puede no cumplirse si

- No somos cuidadosos al especificar el modelo
- Tamaño de muestra es pequeño respecto al número de parámetros a estimar ($n < k + 1$)

  - Necesitamos al menos $k+1$ observaciones para estimar $k+1$ parámetros
  - Entre más observaciones, mejor

Si el modelo se especifica correctamente y $n \geq k+1$, RLM.3 puede fallar por coincidencia al recolectar los datos

- Caso poco probable, a menos que la muestra sea muy pequeña

Supuesto RLM.4. Media condicional cero (el más importante para probar falta de sesgo)

$$\mathbb{E}\left(u|x_1, x_2, \ldots, x_k\right) = 0$$

El error u tiene un valor esperado de cero dado cualquier valor de las variables independientes
Observaciones sobre el supuesto RLM.4:

- RLM.4 no se cumple y los estimadores $\widehat{\beta}_j$ estarán sesgados cuando
  - Hay una mala especificación de la forma funcional
    * Ej. No incluir $ingreso^2$ en $consumo = \beta_0 + \beta_1 ingreso + \beta_2 ingreso^2 + u$
    * Ej. Usar *salario* cuando el modelo verdadero es $\log(salario)$
  - Se omite un factor importante que está correlacionado con $x_1, x_2, \ldots, x_k$
    * Variables omitidas son menos problema en RLM que en RLS
    * Pero puede haber limitación en datos o desconocimiento
  - Hay otras formas en que $\text{Corr}\left(u, x_j\right) \neq 0$
    * Error de medición en las x's
    * y y x's determinadas conjuntamente (ej. precios y cantidades)
- Si se cumple RLM.4, decimos que tenemos variables independientes **exógenas**
- Si x_j se correlaciona con u, se dice que x_j es una variable independiente **endógena**
- Diferencia entre RLM.3 y RLM.4:
  - RLM.3 descarta ciertas relaciones entre las x's y no tiene nada que ver con u
    * Se puede probar
  - RLM.4 restringe relación entre las x's y u
    * Nunca sabremos, pero es un supuesto crítico

Teorema. Falta de Sesgo en MCO
Bajo supuestos RLM.1 a RLM.4,

$$\mathbb{E}\left(\widehat{\beta}_j\right) = \beta_j, \quad j = 0, 1, 2, \ldots, k,$$

para *cualquier* valor de β_j

MCO genera estimadores insesgados para los parámetros poblacionales

- Propiedad de los estimadores bajo muestras aleatorias repetidas

3.3.2 Inclusión de variables irrelevantes

Sobre-especificación del modelo:

- Incluir una o más variables independientes en el modelo que no tienen efecto parcial en la población [$\beta_s = 0$]

Supongamos que el siguiente modelo satisface RLM.1 a RLM.4 pero $\beta_3 = 0$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

- x_3 puede estar correlacionada con x_1 o x_2
  - Pero una vez que se controla por x_1 y x_2, x_3 no afecta a y
  - Usando esperanza condicional:

$$\mathbb{E}\left(y \mid x_1, x_2, x_3\right) = \mathbb{E}\left(y \mid x_1, x_2\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

- No sabemos que $\beta_3 = 0$ por lo que estimamos

$$\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x_1 + \widehat{\beta}_2 x_2 + \widehat{\beta}_3 x_3$$

- ¿Qué efecto tiene incluir una variable (irrelevante) x_3 cuando $\beta_3 = 0$?
  - No tiene efecto en términos de falta de sesgo para $\widehat{\beta}_1$ y $\widehat{\beta}_2$
    * Teorema aplica para cualquier valor de β_j
    * Entonces $\mathbb{E}\left(\widehat{\beta}_0\right) = \beta_0$, $\mathbb{E}\left(\widehat{\beta}_1\right) = \beta_1$, $\mathbb{E}\left(\widehat{\beta}_2\right) = \beta_2$, $\mathbb{E}\left(\widehat{\beta}_3\right) = 0$
  - El efecto está en las varianzas de $\widehat{\beta}_1$ y $\widehat{\beta}_2$

3.3.3 Sesgo por variable omitida

Sub-especificación del modelo:

- Excluir una variable independiente relevante
  - La variable pertenece al modelo poblacional
- Este problema generalmente causa sesgo en los estimadores MCO
  - Podemos derivar la dirección y el tamaño del sesgo
- Se conoce como **análisis de mala especificación**

Modelo con 2 variables independientes
Empezamos con el modelo verdadero y suponemos que satisface los supuestos RLM.1 a RLM.4

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$

- Ej. $y = \textit{salario}$ (por hora), $x_1 = \textit{educ}$, $x_2 = \textit{habilidad}$ (innata)
- Estamos interesados en el efecto parcial de x_1 sobre y

- Para tener $\mathbb{E}\left(\widehat{\beta}_j\right) = \beta_j$, debemos correr una regresión de y sobre x_1 y x_2

$$\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x_1 + \widehat{\beta}_2 x_2$$

- Pero por desconocimiento o falta de datos, estimamos el modelo sin incluir x_2

$$\widetilde{y} = \widetilde{\beta}_0 + \widetilde{\beta}_1 x_1$$

  - La tilde enfatiza que $\widetilde{\beta}_1$ viene de un modelo sub-especificado

Ejemplo. Modelo verdadero:

$$salario = \beta_0 + \beta_1 educ + \beta_2 habilidad + u$$

- Estimamos

$$salario = \beta_0 + \beta_1 educ + \mathrm{v}$$

- $\mathrm{v} = \beta_2 habilidad + u$ porque habilidad no se observa
- Obtenemos $\widetilde{\beta}_1$

Sabemos que

$$\widetilde{\beta}_1 = \widehat{\beta}_1 + \widehat{\beta}_2 \widetilde{\delta}_1$$

- $\widehat{\beta}_1$ y $\widehat{\beta}_2$ son estimadores (imaginarios) de la regresión de y_i sobre x_{i1} y x_{i2} $\forall\, i$
- $\widetilde{\delta}_1$ es la pendiente en la RLS de x_{i2} sobre x_{i1} $\forall\, i$
  - Solo depende de las variables independientes

Como el modelo satisface RLM.1 a RLM.4, $\mathbb{E}\left(\widehat{\beta}_1\right) = \beta_1$ y $\mathbb{E}\left(\widehat{\beta}_2\right) = \beta_2$, entonces

$$\mathbb{E}\left(\widetilde{\beta}_1\right) = \mathbb{E}\left(\widehat{\beta}_1 + \widehat{\beta}_2 \widetilde{\delta}_1\right) = \mathbb{E}\left(\widehat{\beta}_1\right) + \mathbb{E}\left(\widehat{\beta}_2\right) \widetilde{\delta}_1 = \beta_1 + \beta_2 \widetilde{\delta}_1$$

$$\text{Sesgo}\left(\widetilde{\beta}_1\right) = \mathbb{E}\left(\widetilde{\beta}_1\right) - \beta_1 = \beta_2 \widetilde{\delta}_1$$

- Se conoce como el **sesgo por variable omitida** (por excluir x_2)

Dos casos en que $\widetilde{\beta}_1$ es insesgada ($\mathbb{E}\left(\widetilde{\beta}_1\right) = \beta_1$)

- $\beta_2 = 0$: x_2 no aparece el modelo verdadero
- $\widetilde{\delta}_1 = 0$: x_1 y x_2 no están correlacionadas (porque $\widetilde{\delta}_1 = \text{Cov}\left(x_1, x_2\right) / \text{Var}\left(x_1\right)$)

Si x_1 y x_2 están correlacionadas, $\widetilde{\delta}_1$ adopta el signo de $\text{Corr}\left(x_1, x_2\right)$

- $\widetilde{\delta}_1 > 0$ si $\text{Corr}\left(x_1, x_2\right) > 0$
- $\widetilde{\delta}_1 < 0$ si $\text{Corr}\left(x_1, x_2\right) < 0$

El *signo* o *dirección* del sesgo depende de los signos de β_2 y $\widetilde{\delta}_1$

	$\text{Corr}(x_1, x_2) > 0$	$\text{Corr}(x_1, x_2) < 0$
$\beta_2 > 0$	Sesgo+	Sesgo−
$\beta_2 < 0$	Sesgo−	Sesgo+

El *tamaño* del sesgo está determinado por el tamaño de β_2 y $\widetilde{\delta}_1$

- Ej. 0.14% vs 3% relativo a 8.6%

En la práctica, β_2 es un parámetro poblacional no conocido pero generalmente

- Tenemos idea de la dirección del efecto parcial de x_2
- Podemos conjeturar el signo de la correlación entre x_1 y x_2

Ejemplo. A mayor habilidad, mayor productividad y mayor salario, entonces $\beta_2 > 0$. Si $\text{Corr}(educ, \ habilidad) > 0$, $\widetilde{\beta}_1$ de la siguiente regresión en promedio será grande

$$salario = \beta_0 + \beta_1 educ + \text{v}$$

Si $\log(salario) = \beta_0 + \beta_1 educ + \beta_2 habilidad + u$ satisface RLM.1 a RLM.4 pero estimamos

$$\widetilde{\log(salario)} = 0.584 + 0.083 educ, \quad n = 526, \quad \text{R}^2 = 0.186$$

- No sabemos si $0.083 > \beta_1$ (porque β_1 puede ser mayor o menor a 0.083)
- Pero en promedio $\mathbb{E}\left(\widetilde{\beta}_1\right) > \beta_1$

Preguntas. Supongamos que en escuelas primarias

$$califprom = \beta_0 + \beta_1 gasto + \beta_2 tpobreza + u$$

pero estimamos β_1 con una regresión de $califprom$ sobre $gasto$

- ¿Cuál podría ser el signo para β_2 y para $\text{Corr}(x_1, x_2)$?
- ¿Cuál es el posible sesgo en $\widetilde{\beta}_1$?
- ¿Cuál sería la implicación si $\beta_1 = 0$ y $\widetilde{\beta}_1 > 0$?

Terminología al omitir una variable independientes relevante

- Sin importar el signo de β_1, $\widetilde{\beta}_1$ tiene:
  - **Sesgo hacia arriba** si $\mathbb{E}\left(\widetilde{\beta}_1\right) > \beta_1$
  - **Sesgo hacia abajo** si $\mathbb{E}\left(\widetilde{\beta}_1\right) < \beta_1$
- $\widetilde{\beta}_1$ está **sesgada hacia cero** si $\mathbb{E}\left(\widetilde{\beta}_1\right)$ está más cercano a cero que β_1
  - Si $\beta_1 > 0$, $\widetilde{\beta}_1$ está sesgada hacia cero si tiene sesgo hacia abajo

- Si $\beta_1 < 0$, $\widetilde{\beta}_1$ está sesgada hacia cero si tiene sesgo hacia arriba
- En general, si $|\mathbb{E}\left(\widetilde{\beta}_1\right)| < |\beta_1|$

Modelo con 3 variables independientes
Derivar el signo del sesgo por variable omitida es más complicado cuando hay varios regresores porque las x_j's están correlacionadas en pares

- Si $\text{Corr}\,(x_s, u) \neq 0$, generalmente todos los estimadores MCO $\widehat{\beta}_j$'s estarán sesgados

Supongamos que el siguiente modelo poblacional satisface RLM.1 a RLM.4

$$y = \widehat{\beta}_0 + \widehat{\beta}_1 x_1 + \widehat{\beta}_2 x_2 + \widehat{\beta}_3 x_3$$

pero omitimos x_3 y estimamos

$$\widetilde{y} = \widetilde{\beta}_0 + \widetilde{\beta}_1 x_1 + \widetilde{\beta}_2 x_2$$

Si $\text{Corr}\,(x_1, x_3) \neq 0$ y $\text{Corr}\,(x_2,\, x_3) = 0$

- No es correcto decir que $\widetilde{\beta}_1$ está sesgada y $\widetilde{\beta}_2$ no está sesgada
- Generalmente, $\widetilde{\beta}_1$ y $\widetilde{\beta}_2$ estarán sesgadas
  - Excepto si $\text{Corr}\,(x_1,\, x_2) = 0$

Para aproximar la dirección del sesgo en $\widetilde{\beta}_1$, suponemos que x_2 no esta en la población ni estimación ($\text{Corr}\,(x_1,\, x_2) = 0$), entonces

$$\mathbb{E}\left(\widetilde{\beta}_1\right) = \beta_1 + \beta_3 \widetilde{\delta}_{31}$$

- Y Sesgo $\left(\widetilde{\beta}_1\right) > 0$ si $\beta_3 > 0$ y $\text{Corr}\,(x_1,\, x_3) > 0$, etc.

Ejemplo.

$$salario = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 habilidad + u$$

- Si omitimos *habilidad*, los estimadores de β_1 y β_2 estarán sesgados
  - Aún si $\text{Corr}\,(exper,\, habilidad) = 0$
- Si estamos interesados en el rendimiento de *educ* (β_1), quisiéramos saber si $\widetilde{\beta}_1$ está sesgada hacia arriba o hacia abajo

Para aproximar, suuponemos que $\text{Corr}\,(exper,\, habilidad) = 0$ y $\text{Corr}\,(educ,\, exper) = 0$

- Como $\beta_3 > 0$ y $\text{Corr}\,(educ,\, habilidad) > 0$, $\widetilde{\beta}_1$ tiene sesgo hacia arriba

En la práctica, seguimos un razonamiento similar para tener una idea del sesgo

- Hay que evaluar la relación entre la variable de interés y la variable omitida

3.4 Varianza de los estimadores MCO

De la distribución muestral de las $\widehat{\beta}_j$'s

- Sabemos las tendencias centrales (supuestos RLM.1 a RLM.4)
- Queremos saber la dispersión para analizar (bajo homocedasticidad)
  - Varianza
  - Eficiencia

Supuesto RLM.5. Homocedasticidad
El error u tiene la misma varianza dados cualesquier valores de las variables independientes

$$\operatorname{Var}(u \mid x_1, x_2, \ldots, x_k) = \sigma^2$$

La varianza es la misma para todas las posibles combinaciones de valores de las variables independientes

Ejemplo. Si el modelo poblacional es

$$salario = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 antig + u,$$

entonces

$$\operatorname{Var}(u \mid educ, exper, antig) = \sigma^2$$

no depende de *educ*, *exper*, *antig*

Observaciones sobre el supuesto RLM.5:

- Si RLM.5 no se cumple, hay **heterocedasticidad**
- Los supuestos RLM.1 a RLM.5 se conocen como los supuestos de Gauss-Markov (para corte transversal)
- Si escribimos $\vec{x} = (x_1, x_2, \ldots, x_k)$, podemos expresar
  - RLM.1 a RLM.4 como

    $$\mathbb{E}(y \mid \vec{x}) = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k$$

    * Lineal en parámetros, depende de las variables independientes
  - RLM.5 como

    $$\operatorname{Var}(y \mid \vec{x}) = \sigma^2$$

    * No depende de las variables independientes

Teorema. Varianzas Muestrales de los Estimadores MCO (para las pendientes)
Bajo los supuestos RLM.1 a RLM.5, y condicional en los valores muestrales de las

variables independientes,

$$\mathrm{Var}\left(\widehat{\beta}_j\right) = \frac{\sigma^2}{\mathrm{SCT}_j(1-\mathrm{R}_j^2)} \quad \forall\, j = 1, 2, \ldots, k$$

- $\mathrm{SCT}_j = \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2$ es la variación muestral total en x_j
- R_j^2 es la R^2 de la regresión de x_j sobre las otras variables independientes (con intercepto)
  - Proporción de la variación total en x_j explicada por las otras variables independientes

El tamaño de $\mathrm{Var}\left(\widehat{\beta}_j\right)$ es importante

- Una varianza grande significa un estimador menos preciso e implica
  - Intervalos de confianza más amplios
  - Pruebas de hipótesis menos precisas

3.4.1 Componentes de las varianzas de MCO: Multicolinealidad

La varianza de cada $\widehat{\beta}_j$ depende de 3 factores: σ^2, SCT_j, R_j^2

1. Varianza del error (σ^2):
   - Si sube σ^2, sube $\mathrm{Var}\left(\widehat{\beta}_j\right)$ porque hay más ruido en la ecuación
     - Hace más difícil estimar el efecto parcial de cualquier variable independiente sobre y
   - σ^2 es una característica de la población
     - No tiene que ver con el tamaño de la muestra
     - Es desconocida por lo que queremos estimarla
   - Dada y, la única forma en que σ^2 puede bajar es agregando más variables independientes ('sacarlas' de u)
     - No siempre es posible
2. Variación muestral total en x_j (SCT_j):
   - Si sube SCT_j, baja $\mathrm{Var}\left(\widehat{\beta}_j\right)$
     - Ceteris paribus queremos mayor variación en x_j
     - Al aumentar el tamaño de la muestra, aumenta la variación en cada x_j
   - Componente de $\mathrm{Var}\left(\widehat{\beta}_j\right)$ que depende sistemáticamente de la muestra
   - Si $\mathrm{SCT}_j \approx 0$, aumenta $\mathrm{Var}\left(\widehat{\beta}_j\right)$ pero no viola RLM.3
     - $\mathrm{SCT}_j = 0$ no está permitido por RLM.3
     - n pequeño puede causar varianzas muestrales grandes

3. Relaciones lineales entre las variables independientes(R_j^2):

   - Si sube R_j^2, sube $\mathrm{Var}\left(\widehat{\beta}_j\right)$
   - R_j^2 no aparece en RLS porque solo hay una x_j y $\mathrm{R}_j^2 \neq \mathrm{R}^2$
     - R_j^2 solo involucra variables independientes
   - Si $k = 2$, $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$ y $\mathrm{Var}\left(\widehat{\beta}_1\right) = \frac{\sigma^2}{\mathrm{SCT}_1(1-\mathrm{R}_1^2)}$ con R_1^2 de la regresión de x_1 sobre x_2 (con intercepto)
     - Como R^2 mide la bondad de ajuste, $\mathrm{R}_1^2 \approx 1$ implica que x_2 explica mucha de la variación en x_1 por lo que $\mathrm{Corr}\left(x_1, x_2\right)$ es alta
     - Conforme $\mathrm{R}_1^2 \to 1$, sube $\mathrm{Var}\left(\widehat{\beta}_j\right)$, entonces una alta relación lineal entre x_1 y x_2 implica que $\mathrm{Var}\left(\widehat{\beta}_1\right)$ y $\mathrm{Var}\left(\widehat{\beta}_2\right)$ son altas
   - Dados σ^2 y SCT_j, $\mathrm{Var}\left(\widehat{\beta}_j\right)$ es más baja cuando $\mathrm{R}_j^2 = 0 \iff \mathrm{Corr}\left(x_s, x_j\right) = 0 \quad \forall s$
     - Pero $\mathrm{R}_j^2 = 0$ es raro
   - Caso $\mathrm{R}_j^2 = 1$ está descartado por RLM.3
     - x_j sería una combinación lineal perfecta de las otras variables independientes
   - Más interesante es el caso $\mathrm{R}_j^2 \approx 1$ porque cuando $\mathrm{R}_j^2 \to 1$, $\mathrm{Var}\left(\widehat{\beta}_j\right) \to \infty$
     - Pero $\mathrm{R}_j^2 \approx 1$ no viola RLM.3
   - **Multicolinealidad** (MC): Se refiere a una correlación alta (pero no perfecta) entre 2 o más variables independientes
     - Problema de MC no está bien definido porque no viola supuestos
     - Efecto de $\mathrm{R}_j^2 \approx 1$ en $\mathrm{Var}\left(\widehat{\beta}_j\right)$ depende de σ y SCT_j
     - Lo que importa para inferencia estadística es $\widehat{\beta}_j$ relativo a $\mathrm{desvest}\left(\widehat{\beta}_j\right)$
   - Alta correlación entre variables independientes y n pequeño, implica que $\mathrm{Var}\left(\widehat{\beta}_j\right)$ aumenta
   - Ceteris paribus es mejor tener una correlación baja entre x_j y otras variables independientes
   - Quitar variables independientes del modelo podría reducir MC, pero puede generar sesgo si las variables pertenecen al modelo poblacional
     - Ej. En lugar de estimar efectos separados de categorías de gasto altamente correlacionadas, sería mejor juntarlas
   - La correlación entre variables independientes puede ser irrelevante para otra variable independiente (de interés)
     - Ej. $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$ con alta $\mathrm{Corr}\left(x_2,\, x_3\right)$ y altas $\mathrm{Var}\left(\widehat{\beta}_2\right)$ y $\mathrm{Var}\left(\widehat{\beta}_3\right)$ pero sin efecto en $\mathrm{Var}\left(\widehat{\beta}_1\right)$

- $*$ Si $\text{Corr}(x_1, x_2) = 0$ y $\text{Corr}(x_1, x_3) = 0$, entonces $\text{R}_j^2 = 0$ y $\text{Var}\left(\widehat{\beta}_1\right) = \frac{\sigma^2}{\text{SCT}_j}$ sin importar $\text{Corr}(x_2, x_3)$

  - Esto es importante porque generalmente se incluyen varios controles para aislar efecto causal de una variable
    - $*$ Alta correlación entre controles no dificulta determinar efectos

- Se pueden utilizar estadísticos para diagnosticar MC
  - Pero no distinguen si involucra controles que no importan
- Son más útiles estadísticos para coeficientes individuales, el más común es el **factor de inflación de la varianza**

$$\text{FIV}_j = \frac{1}{1 - \text{R}_j^2} \implies \text{Var}\left(\widehat{\beta}_1\right) = \frac{\sigma^2}{\text{SCT}_j}\text{FIV}_j$$

  - FIV_j es el factor por el que $\text{Var}\left(\widehat{\beta}_j\right)$ es más alta relativo a cuando x_j no se correlaciona con otras variables independientes
  - Ceteris paribus quisiéramos un FIV_j bajo pero rara vez podemos escoger
  - Un FIV_j “alto” para los controles, no debería afectar decisión de incluirlos
    - $*$ Ignoramos FIV_j de controles
  - Fijar un valor límite para FIV_j y declarar que probablemente hay MC para $\widehat{\beta}_j$ es arbitrario
    - $*$ Pero a veces se usa 10 ($\text{R}_j^2 > 0.9$)
    - $*$ $\text{FIV}_j > 10 \;\not\Longrightarrow$ desvest $\left(\widehat{\beta}_j\right)$ sea alta porque desvest $\left(\widehat{\beta}_j\right)$ también depende de σ y SCT_j (puede crecer si n crece)
  - En conclusión, FIV_j tiene un uso limitado

3.4.2 Varianzas en modelos mal especificados

Relación inversa entre sesgo y varianza:

- Puede ayudar a decidir sobre si incluir una variable en un modelo de regresión

Analicemos la varianza de los estimadores MCO cuando se omite una variable relevante

- Supongamos que el siguiente modelo poblacional satisface los supuestos G-M

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$

  y consideremos 2 estimadores para β_1:

$$\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x_1 + \widehat{\beta}_2 x_2$$
$$\widetilde{y} = \widetilde{\beta}_0 + \widetilde{\beta}_1 x_1$$

- Si $\beta_2 \neq 0$ y $\text{Corr}(x_1, x_2) \neq 0$, $\widetilde{\beta}_1$ estará sesgado ($\mathbb{E}\left(\widetilde{\beta}_1\right) \neq \beta_1$)
- Mientras que $\mathbb{E}\left(\widehat{\beta}_1\right) = \beta_1 \quad \forall\, \beta_2$, incluido $\beta_2 = 0$

- En este sentido, $\widehat{\beta}_1$ parece mejor a $\widetilde{\beta}_1$, pero

$$\mathrm{Var}\left(\widetilde{\beta}_1\right) = \frac{\sigma^2}{\mathrm{SCT}_1} < \frac{\sigma^2}{\mathrm{SCT}_1(1-\mathrm{R}_1^2)} = \mathrm{Var}\left(\widehat{\beta}_1\right)$$

Conclusiones si $\mathrm{Corr}\,(x_1,\, x_2) \neq 0$

1. Si $\beta_2 \neq 0$,
$$\mathbb{E}\left(\widetilde{\beta}_1\right) \neq \beta_1,\ \mathbb{E}\left(\widehat{\beta}_1\right) = \beta_1 \text{ y } \mathrm{Var}\left(\widetilde{\beta}_1\right) < \mathrm{Var}\left(\widehat{\beta}_1\right)$$
   - Se prefiere $\widehat{\beta}_1$ cuando n es grande
     - MC inducida por incluir x_2 se vuelve menos importante
     - Ambas varianzas disminuyen cuando $n \to \infty$
   - Si x_2 se excluye, x_2 estaría en u, σ^2 subiría y $\mathrm{Var}\left(\widetilde{\beta}_1\right)$ subiría
2. Si $\beta_2 = 0$,
$$\mathbb{E}\left(\widetilde{\beta}_1\right) = \beta_1,\ \mathbb{E}\left(\widehat{\beta}_1\right) = \beta_1 \text{ y } \mathrm{Var}\left(\widetilde{\beta}_1\right) < \mathrm{Var}\left(\widehat{\beta}_1\right)$$
   - Se prefiere $\widetilde{\beta}_1$
   - Incluir x_2 cuando no tiene efecto en y, agrava el problema de MC
     - El estimador de β_1 es menos eficiente
   - Una varianza más grande para el estimador de β_1 es el costo de incluir una variable irrelevante en el modelo

3.4.3 Estimación de σ^2

Conocemos $\mathrm{Var}\left(\widehat{\beta}_1\right)$, ¿por qué no podemos calcularla?

Al obtener un estimador insesgado de σ^2, tendríamos un estimador insesgado de $\mathrm{Var}\left(\widehat{\beta}_1\right)$

- Dado que $\sigma^2 = \mathbb{E}\,(u^2)$, un 'estimador' insesgado de σ^2 sería $\widetilde{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} u_i^2$, pero no observamos los errores u_i porque no conocemos las β_j's
$$u_i = y_i - \beta_0 - \beta_1 x_{i1} - \ldots - \beta_k x_{ik}$$
- Si reemplazamos las β_j's con $\widehat{\beta}_j$'s obtenemos los residuales MCO
$$\widehat{u}_i = y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_{i1} - \ldots - \widehat{\beta}_k x_{ik}$$
- Sin embargo, <u>no</u> reemplazamos $\widehat{u}_i$ por u_i para obtener $\breve{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} \widehat{u}_i^2$ porque generaría un estimador sesgado de σ^2
  - No toma en cuenta los grados de libertad

El estimador insesgado de σ^2 en RLM es

$$\widehat{\sigma}^2 = \frac{1}{n-k-1}\sum_{i=1}^{n} \widehat{u}_i^2 = \frac{\mathrm{SCR}}{n-k-1}$$

- $n-k-1$ son los grados de libertad (g.l.)
  - Número de observaciones (n) menos parametros estimados ($k+1$)
  - En RLS, $k=1$
- Intuitivamente, el ajuste por g.l. se necesita por las condiciones de primer orden

$$\sum_{i=1}^{n} \widehat{u}_i \quad \text{y} \quad \sum_{i=1}^{n} x_{ij}\widehat{u}_i \quad \forall\, j = 1, 2, \ldots, k$$

  - Hay $k+1$ resricciones impuestas sobre los residuales
  - Dados $n-(k+1)$ de los residuales, los $k+1$ residuales restantes se conocen
  - Solo hay $n-(k+1)$ g.l. en los residuales
- $n > n-k-1 \implies \frac{1}{n-k-1} > \frac{1}{n} \implies \frac{1}{n-k-1}\sum_{i=1}^{n} \widehat{u}_i^2 > \frac{1}{n}\sum_{i=1}^{n} \widehat{u}_i^2$

> Teorema. Estimación Insesgada de σ^2
> Bajo supuestos RLM.1 a RLM.5,
>
> $$\mathbb{E}\left(\widehat{\sigma}^2\right) = \sigma^2$$

$\widehat{\sigma} = +\sqrt{\widehat{\sigma}^2}$ se llama el **error estándar de la regresión** (EER)

- Estimador de la desviación estándar de u
- También se conoce como
  - Error estándar del estimado
  - Raíz del error cuadrático medio

Si agregamos una variable independientes a la regresión, $\widehat{\sigma}$ puede subir o bajar

- SCR podría bajar (numerador)
- $n-(k+1)$ podría bajar en 1 (denominador)
- No podemos saber de antemano que efecto va a dominar

3.4.4 Errores estándar de los estimadores MCO

Para construir IC y hacer PH, necesitamos estimar la desviación estándar de $\widehat{\beta}_j$

$$\operatorname{Var}\left(\widehat{\beta}_j\right) = \frac{\sigma^2}{\mathrm{SCT}_j(1-\mathrm{R}_j^2)} \implies \operatorname{desvest}\left(\widehat{\beta}_j\right) = \frac{\sigma}{\sqrt{\mathrm{SCT}_j(1-\mathrm{R}_j^2)}}$$

Como σ no es conocida, usamos su estimador $\widehat{\sigma}$ y obtenemos el **error estándar de** $\widehat{\beta}_j$

$$\operatorname{errest}\left(\widehat{\beta}_j\right) = \frac{\widehat{\sigma}}{\sqrt{\mathrm{SCT}_j(1-\mathrm{R}_j^2)}}$$

- Los estimados MCO y el error estándar se pueden calcular para cualquier muestra aleatoria
- errest $\left(\widehat{\beta}_j\right)$ depende de $\widehat{\sigma}$ por lo que tiene una distribución muestral
  - Se utiliza para hacer inferencia
- errest $\left(\widehat{\beta}_j\right)$ depende de RLM.5
  - No es válido para estimar desvest $\left(\widehat{\beta}_j\right)$ cuando los errores exhiben heterocedasticidad
  - Heterocedasticidad no genera sesgo en $\widehat{\beta}_j$ pero sí en Var $\left(\widehat{\beta}_j\right)$
    * Inválida errest $\left(\widehat{\beta}_j\right)$
  - Importante: Software reporta errest $\left(\widehat{\beta}_j\right)$ bajo RLM.5 por default
  - Si sospechamos de heterocedasticidad, los errores estándar 'usuales' no son válidos y tenemos que tomar medidas para corregirlos
- Si definimos desvest $(x_j) = \sqrt{\mathrm{SCT}_j/n}$, podemos re-expresar

$$\text{errest}\left(\widehat{\beta}_j\right) = \frac{\widehat{\sigma}}{\sqrt{n}\text{desvest}\left(x_j\right)\sqrt{1-\mathrm{R}_j^2}}$$

  - Muestra cómo n afecta directamente al error estándar
  - Otros 3 términos ($\widehat{\sigma}$, desvest (x_j) , R_j^2) cambian con diferentes muestras aleatorias pero convergen a constantes conforme $n \to \infty$
  - errest $\to 0$ a una tasa $1/\sqrt{n}$
- Valor de tener más datos:
  - La precisión de $\widehat{\beta}_j$ aumenta conforme $n \to \infty$
- En contraste, falta de sesgo se cumple para cualquier n en tanto se puedan calcular los estimadores

3.5 Eficiencia de MCO: Teorema Gauss-Markov

Sabemos que bajo RLM.1 a RLM.4, MCO es insesgado

- Pero hay muchos estimadores insesgados

¿Hay estimadores con varianza menor que MCO?

- Si limitamos apropiadamente el conjunto de estimadores potenciales, se puede probar que MCO es el mejor en ese conjunto

> Teorema Gauss-Markov
> Bajo supuestos RLM.1 a RLM.5, $\widehat{\beta}_0, \widehat{\beta}_1, \ldots, \widehat{\beta}_k$ son los mejores estimadores lineales

insesgados (MELI) de $\beta_0, \beta_1, \ldots, \beta_k$, respectivamente

Significado de MELI:

- Estimador: Regla que se puede aplicar a los datos para producir un estimado
- Estimador insesgado: $\mathbb{E}\left(\widetilde{\beta}_j\right) = \beta_j \quad \forall\, j = 0, 1, \ldots, k$
- Estimador lineal: Si se puede expresar como función lineal de la variable dependiente

$$\widetilde{\beta}_j = \sum_{i=1}^{n} w_{ij} y_i$$

  - Cada w_{ij} es una función de los valores de la muestra de todas las x_j's
  - Los estimadores MCO son lineales,

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n} \widehat{r}_{i1} y_i}{\sum_{i=1}^{n} \widehat{r}_{i1}^2}$$

- Mejor estimador: Tiene la menor varianza
  - Dados 2 estimadores, preferimos el que tenga la menor varianza

Anotaciones sobre el teorema G-M:

- El teorema justifica el uso del método MCO para estimar modelos de regresión múltiple
  - Para cada estimador $\widetilde{\beta}_j$ que sea lineal e insesgado, $\mathrm{Var}\left(\widehat{\beta}_j\right) \leq \mathrm{Var}\left(\widetilde{\beta}_j\right)$
    - Generalmente con desigualdad estricta
  - Cuando los supuestos G-M se cumplen, ningún estimador lineal e insesgado será mejor que MCO
  - Si nos dan un estimador lineal e insesgado, sabemos que su varianza es al menos igual que la varianza MCO
- Si queremos estimar una función lineal de las β_j's, la combinación correspondiente de estimadores MCO tiene la menor varianza entre todos los estimadores lineales insesgados
- Los supuestos RLM.1 a RLM.5 se conocen como supuestos G-M (para corte transversal) por este teorema
- Si alguno de los supuestos no se cumple, el teorema ya no se cumple
  - Sin RLM.4, MCO es sesgado
  - Sin RLM.5, MCO ya no tiene varianza mínima entre los estimadores lineales e insesgados
    - En ese caso, MC ponderados son mejores que MCO

3.6 Sobre el lenguaje de RLM

¿Cuál de los siguientes enunciados es correcto?

- Estimar un modelo MCO
- Estimar un modelo lineal por MCO

Diferencia entre modelo y método de estimación

- MCO es un método de estimación, no un modelo
  - Un modelo describe la población y depende de parámetros desconocidos
- El modelo lineal visto se puede escribir en la población como:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u$$

  - β_j son parámetros
  - Podemos hablar del significado de β_j sin tener datos
  - La interpretación viene del modelo
- Con una muestra aleatoria podemos estimar los parámetros
  - Hasta ahora, con MCO bajo los supuestos G-M
  - Bajo diferentes supuestos, se prefieren otros métodos de estimación
    - WLS, LAD, IV