Regresión Lineal Múltiple

Víctor Aceña - Isaac Martín

DSLAB

2025-09-11

El Modelo de Regresión Lineal Múltiple

El modelo de regresión lineal múltiple constituye la **extensión natural y más potente** del modelo simple.

Diferencias clave:

Regresión Simple:

- Una variable respuesta
- Un único predictor
- Relación bivariada

Regresión Múltiple:

- Una variable respuesta
- **Múltiples predictores**
- Relación multivariada

Capacidades únicas:

- **Modelar simultáneamente** el efecto de múltiples variables predictoras
- **Interpretación de coeficientes** en presencia de otros predictores
- **Diagnóstico específico** del modelo múltiple
- Manejo de la **multicolinealidad**

Objetivos de Aprendizaje

1. **Formular y estimar** modelos de regresión lineal múltiple, comprendiendo las diferencias clave respecto al caso simple
2. **Interpretar coeficientes** en el contexto multivariante, entendiendo el concepto de *ceteris paribus*
3. **Realizar inferencia estadística** construyendo intervalos de confianza y contrastes de hipótesis
4. **Evaluar la calidad del ajuste** usando medidas como R^2, R^2 ajustado y descomposición ANOVA
5. **Diagnosticar el modelo múltiple** aplicando técnicas específicas como gráficos CPR
6. **Identificar y tratar la multicolinealidad** usando el VIF como herramienta de diagnóstico
7. **Realizar predicciones** distinguiendo entre intervalos de confianza e intervalos de predicción

Formulación del Modelo Poblacional

Para n observaciones y p variables predictoras, el **modelo poblacional** postula:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip} + \varepsilon_i, \quad i = 1, \ldots, n$$

Componentes:

- Y_i: i-ésima variable respuesta aleatoria
- X_{ij}: i-ésima variable predictora aleatoria del j-ésimo predictor
- ε_i: término de error aleatorio
- $\beta_0, \beta_1, \ldots, \beta_p$: coeficientes poblacionales verdaderos pero desconocidos

Características clave:

- Relación **lineal en los parámetros**
- Los errores son **no observables**
- Los parámetros son **constantes poblacionales**

Formulación del Modelo Muestral

En la práctica, trabajamos con **datos observados** y estimamos el modelo:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \cdots + \hat{\beta}_p x_{ip}, \quad i = 1, \ldots, n$$

Componentes:

- $\hat{y}_i$: i-ésima predicción
- x_{ij}: i-ésima observación del j-ésimo predictor
- $\hat{\beta}_j$: coeficientes estimados

Interpretación clave de $\hat{\beta}_j$:

El cambio estimado en la media de Y ante un cambio de una unidad en X_j, **manteniendo constantes todas las demás variables predictoras**.

Este principio se conoce como ***ceteris paribus*** ("lo demás constante")

Notación Matricial: Modelo Poblacional

Modelo poblacional:

$$\mathbf{Y} = \tilde{X}\beta + \varepsilon$$

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \quad \tilde{X} = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1p} \\ 1 & X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{np} \end{bmatrix}$$

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Nota: $\tilde{X}$ contiene variables aleatorias (mayúsculas X_{ij})

Modelo muestral:

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\beta}$$

$$\hat{\mathbf{y}} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}, \quad \hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_p \end{bmatrix}$$

Observaciones:

- $\mathbf{X}$ contiene datos observados (minúsculas x_{ij})
- $\mathbf{X}$ y $\tilde{X}$ son matrices de dimensión $n \times (p+1)$
- La primera columna de unos corresponde al intercepto β_0

Supuestos del Modelo Lineal Múltiple

Condiciones de Gauss-Markov:

1. **Linealidad en los parámetros:** El modelo $E[\mathbf{Y}|\tilde{X}] = \tilde{X}\beta$ está bien especificado
2. **Exogeneidad:** Los errores tienen media cero: $E[\varepsilon|\tilde{X}] = \mathbf{0}$
3. **Homocedasticidad e independencia:** $\text{Var}(\varepsilon|\tilde{X}) = \sigma^2\mathbf{I}_n$
4. **Ausencia de multicolinealidad perfecta:** $\mathbf{X}$ tiene rango completo $(p+1)$
5. **Normalidad (para inferencia):** $\varepsilon \sim N(\mathbf{0}, \sigma^2\mathbf{I}_n)$

Implicación: Estos supuestos garantizan que los estimadores MCO sean **insesgados, consistentes y eficientes**

Estimación por Mínimos Cuadrados: Función Objetivo

Principio: Minimizar la discrepancia entre valores observados y predichos

Función objetivo:

$$S(\beta) = \sum_{i=1}^{n} e_i^2 = \mathbf{e}^T\mathbf{e} = (\mathbf{y} - \mathbf{X}\beta)^T(\mathbf{y} - \mathbf{X}\beta)$$

¿Por qué cuadrados?

- Los residuos positivos y negativos no se cancelan
- Se penalizan más fuertemente los errores grandes
- Facilita el tratamiento matemático

Resultado: MCO minimiza la **Suma de los Cuadrados de los Residuos** (SSR)

Derivación de las Ecuaciones Normales

Expandiendo la función objetivo:

$$S(\beta) = \mathbf{y}^T\mathbf{y} - 2\beta^T\mathbf{X}^T\mathbf{y} + \beta^T(\mathbf{X}^T\mathbf{X})\beta$$

Derivando respecto a β:

$$\frac{\partial S(\beta)}{\partial \beta} = -2\mathbf{X}^T\mathbf{y} + 2(\mathbf{X}^T\mathbf{X})\beta$$

Igualando a cero:

$$-2\mathbf{X}^T\mathbf{y} + 2(\mathbf{X}^T\mathbf{X})\hat{\beta} = \mathbf{0}$$

Ecuaciones Normales:

$$(\mathbf{X}^T\mathbf{X})\hat{\beta} = \mathbf{X}^T\mathbf{y}$$

Solución MCO y Condición de Invertibilidad

Solución única:

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

Condición necesaria: La matriz $(\mathbf{X}^T\mathbf{X})$ debe ser invertible

¿Cuándo es invertible?

- Cuando $\mathbf{X}$ tiene rango completo $(p+1)$
- Cuando las columnas de $\mathbf{X}$ son linealmente independientes
- Cuando no hay multicolinealidad perfecta

Propiedades de $(\mathbf{X}^T\mathbf{X})$:

- Dimensión: $(p+1) \times (p+1)$
- Simétrica
- Definida positiva (si es invertible)

Propiedades de los Estimadores MCO

Bajo los supuestos de Gauss-Markov:

1. **Insesgados:** $E[\hat{\beta}] = \beta$
2. **Eficientes:** Varianza mínima entre todos los estimadores lineales insesgados
3. **Consistentes:** $\hat{\beta} \xrightarrow{p} \beta$ cuando $n \to \infty$

Matriz de varianza-covarianza:

$$\text{Var}(\hat{\beta}) = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}$$

Bajo normalidad adicional:

$$\hat{\beta} \sim N\left(\beta, \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}\right)$$

Estimación de la Varianza del Error

Estimador insesgado de σ^2:

$$\hat{\sigma}^2 = \frac{\mathsf{SSE}}{n-p-1} = \frac{\sum_{i=1}^{n} e_i^2}{n-p-1} = \frac{\mathbf{e}^T\mathbf{e}}{n-p-1}$$

Grados de libertad: $n-p-1$

- n: número de observaciones
- $p+1$: número de parámetros estimados

Distribución:

$$\frac{(n-p-1)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-p-1}$$

Error estándar de los coeficientes:

$$\hat{\sigma}_{\beta_j} = \hat{\sigma}\sqrt{(\mathbf{X}^T\mathbf{X})^{-1}_{jj}}$$

Ejemplo: Modelo de Precios de Viviendas

Datos: Precios de viviendas basados en características

Variables predictoras:

- `superficie`: Metros cuadrados
- `habitaciones`: Número de habitaciones
- `antiguedad`: Años de antigüedad
- `distancia_centro`: Distancia al centro (km)
- `garaje`: Presencia de garaje (Sí/No)

Interpretación de los Coeficientes

Coeficiente de regresión parcial:

$$\beta_j = \frac{\partial E[Y|\tilde{X}]}{\partial X_j}$$

Interpretación: β_j representa el cambio esperado en Y por una unidad de cambio en X_j, **manteniendo todas las demás variables constantes**

Diferencia crucial:

Regresión Simple:

- Efecto **total** (directo + indirecto)
- Puede estar **confundido**
- $\hat{\beta}_j$ captura toda la asociación

Regresión Múltiple:

- Efecto **puro** o **parcial**
- **Controla** por otras variables
- Interpretación más **causal**

Concepto clave: El coeficiente proviene de una regresión entre residuos

Ejemplo: Interpretación *Ceteris Paribus*

```
                 Estimate Std. Error
(Intercept)     53750.9705 6666.70624
superficie       1171.7780   47.28087
habitaciones    15072.3104 1303.41715
antiguedad       -744.5896   75.42075
distancia_centro -2028.2715  164.87756
garajeSí        25829.4317 2349.44285
```

Interpretación *ceteris paribus*:

- **Superficie** (+1,172 €/m^2): Cada m^2 adicional incrementa el precio
- **Habitaciones** (+15,072 €): Cada habitación adicional aumenta el precio
- **Antigüedad** (-745 €/año): Cada año de antigüedad reduce el precio
- **Distancia centro** (-2,028 €/km): Cada km más lejos del centro reduce el precio
- **Garaje** (+25,829 €): Tener garaje incrementa el precio

Evaluación del Modelo: Descomposición de la Varianza

Descomposición ANOVA:

$$\text{SST} = \text{SSR} + \text{SSE}$$

Donde:

- **SST** (Sum of Squares Total): $\sum_{i=1}^{n}(y_i - \bar{y})^2$
- **SSR** (Sum of Squares Regression): $\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$
- **SSE** (Sum of Squares Error): $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$

Interpretación:

- **SST**: Variabilidad total en los datos
- **SSR**: Variabilidad explicada por el modelo
- **SSE**: Variabilidad no explicada (residual)

Coeficiente de Determinación Múltiple

R-cuadrado:

$$R^2 = \frac{\mathsf{SSR}}{\mathsf{SST}} = 1 - \frac{\mathsf{SSE}}{\mathsf{SST}}$$

Interpretación:

- Proporción de la variabilidad en Y explicada por el modelo
- Rango: $0 \leq R^2 \leq 1$
- $R^2 = 0$: El modelo no explica nada
- $R^2 = 1$: El modelo explica toda la variabilidad

Problema: R^2 siempre aumenta al añadir variables (incluso irrelevantes)

En regresión múltiple: R^2 es el cuadrado de la correlación entre $\mathbf{y}$ y $\hat{\mathbf{y}}$

El Coeficiente de Determinación Ajustado

R-cuadrado ajustado:

$$R^2_{\text{adj}} = 1 - \frac{\text{SSE}/(n-p-1)}{\text{SST}/(n-1)} = 1 - (1 - R^2)\frac{n-1}{n-p-1}$$

Ventajas:

- **Penaliza** la inclusión de variables irrelevantes
- **Puede decrecer** si una variable no aporta información suficiente
- Mejor para **comparar modelos** con diferente número de predictores

Criterio de decisión:

- Si R^2_{adj} aumenta al añadir una variable $\rightarrow$ la variable es útil
- Si R^2_{adj} disminuye $\rightarrow$ la variable no aporta información suficiente

Inferencia: Contraste de Hipótesis Individual

Hipótesis sobre un coeficiente:

$$H_0 : \beta_j = 0 \quad \text{vs} \quad H_1 : \beta_j \neq 0$$

Estadístico de contraste:

$$t = \frac{\hat{\beta}_j - 0}{\hat{\sigma}_{\beta_j}} = \frac{\hat{\beta}_j}{\hat{\sigma}\sqrt{(\mathbf{X}^T\mathbf{X})^{-1}_{jj}}} \sim t_{n-p-1}$$

Interpretación:

- **Rechazar** H_0: La variable X_j es estadísticamente significativa
- **No rechazar** H_0: No hay evidencia de efecto lineal de X_j sobre Y

Valor p: Probabilidad de observar un estadístico t tan extremo o más, bajo H_0

Intervalo de Confianza para los Coeficientes

Intervalo de confianza al $(1-\alpha)\%$**:**

$$\hat{\beta}_j \pm t_{\alpha/2,n-p-1} \cdot \hat{\sigma}_{\beta_j}$$

Interpretación:

- Con $(1-\alpha)\%$ de confianza, el verdadero valor de β_j está en este intervalo
- Si el intervalo **no contiene cero** $\rightarrow$ β_j es significativo
- Si el intervalo **contiene cero** $\rightarrow$ β_j no es significativo

Relación con el test de hipótesis:

- Intervalo de confianza del 95% $\equiv$ Test de hipótesis con $\alpha = 0.05$
- Si 0 está en el IC del 95% $\rightarrow$ No se rechaza H_0 al 5%

Inferencia Global: Test F

Hipótesis global:

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0 \quad \text{vs} \quad H_1 : \text{Al menos un } \beta_j \neq 0$$

Estadístico F:

$$F = \frac{\text{SSR}/p}{\text{SSE}/(n-p-1)} = \frac{R^2/p}{(1-R^2)/(n-p-1)} \sim F_{p,n-p-1}$$

Interpretación:

- **Rechazar** H_0: El modelo es globalmente significativo
- **No rechazar** H_0: El modelo no explica variabilidad significativa

Relación con R^2**:** El test F evalúa si R^2 es significativamente diferente de cero

Ejemplo: Inferencia en el Modelo de Viviendas

```
Call:
lm(formula = precio ~ superficie + habitaciones + antiguedad +
    distancia_centro + garaje, data = viviendas)

Residuals:
   Min     1Q Median     3Q    Max
-38847 -11074    867   9898  38486

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      53750.97    6666.71   8.063 7.53e-14 ***
superficie        1171.78      47.28  24.783  < 2e-16 ***
habitaciones     15072.31    1303.42  11.564  < 2e-16 ***
antiguedad        -744.59      75.42  -9.872  < 2e-16 ***
distancia_centro -2028.27     164.88 -12.302  < 2e-16 ***
garajeSí         25829.43    2349.44  10.994  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 15950 on 194 degrees of freedom
Multiple R-squared:  0.9094,    Adjusted R-squared:  0.9071
F-statistic: 389.4 on 5 and 194 DF,  p-value: < 2.2e-16
```

Predicción con el Modelo Múltiple

Predicción puntual: Para un nuevo vector $\mathbf{x}_0$:

$$\hat{y}_0 = \mathbf{x}_0^T \hat{\beta}$$

Dos tipos de intervalos:

Intervalo de Confianza:

- Para la **respuesta media** $E[Y|\mathbf{x}_0]$
- Incertidumbre en la estimación
- Más estrecho

Intervalo de Predicción:

- Para una **observación individual** Y_0
- Incertidumbre $+$ variabilidad natural
- Más amplio

Fórmulas: Ambos dependen de $\hat{\sigma}^2$ y de la matriz $(\mathbf{X}^T\mathbf{X})^{-1}$

Intervalos de Confianza y Predicción

Intervalo de confianza para la respuesta media:

$$\hat{y}_0 \pm t_{\alpha/2,n-p-1} \cdot \hat{\sigma}\sqrt{\mathbf{x}_0^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_0}$$

Intervalo de predicción para una observación individual:

$$\hat{y}_0 \pm t_{\alpha/2,n-p-1} \cdot \hat{\sigma}\sqrt{1 + \mathbf{x}_0^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_0}$$

Diferencia clave: El "+1" en el intervalo de predicción refleja la variabilidad adicional de una observación individual

Amplitud: Intervalo de predicción $>$ Intervalo de confianza

Diagnóstico del Modelo Múltiple

Una vez ajustado el modelo, es **fundamental realizar un diagnóstico exhaustivo** para verificar que los supuestos se cumplen.

Base del diagnóstico: Análisis de los residuos - nuestra ventana a los errores teóricos no observables

Supuestos a verificar:

1. **Normalidad**
   - Gráfico Q-Q de residuos
   - Test de Shapiro-Wilk
2. **Independencia**
   - Residuos vs tiempo
   - Test de Durbin-Watson
3. **Homocedasticidad**
   - Gráfico Scale-Location
   - Test de Breusch-Pagan
4. **Linealidad**
   - Residuos vs valores ajustados
   - **Gráficos CPR** (específicos de múltiple)

Gráficos de Componente más Residuo

Problema: El gráfico residuos vs ajustados puede ocultar una relación no lineal con **una variable específica**

Solución: Gráficos CPR para cada predictor X_j:

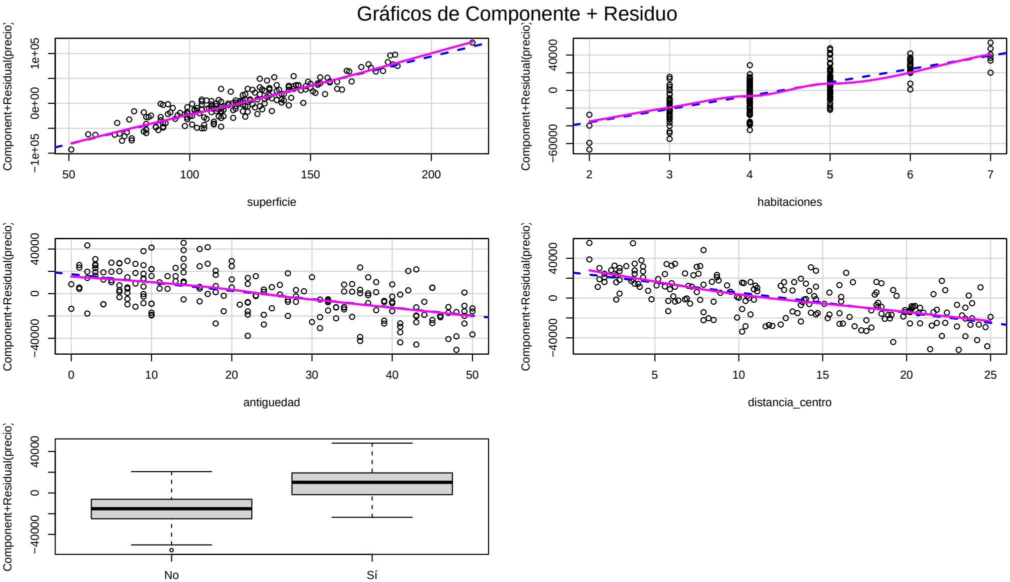
$$\text{Residuo Parcial} = e_i + \hat{\beta}_j x_{ij} \quad \text{vs.} \quad x_{ij}$$

Interpretación:

- **Línea sólida**: Relación lineal esperada (pendiente $= \hat{\beta}_j$)
- **Línea punteada**: Suavizado no paramétrico
- **Coincidencia**: Linealidad adecuada
- **Divergencia**: Posible no-linealidad $\rightarrow$ necesita transformación

Ventaja: Permite detectar no-linealidades específicas de cada variable

Ejemplo: Diagnóstico con Gráficos CPR



Interpretación de los Gráficos CPR

¿Qué observamos en las 5 variables?

Superficie y Habitaciones:

- Líneas sólida y punteada coinciden
- **Conclusión:** Relación lineal adecuada

Garaje:

- Separación clara entre grupos (No/Sí)
- **Conclusión:** Efecto categórico apropiado

Antigüedad y Distancia:

- Líneas coinciden bien
- **Conclusión:** Linealidad confirmada

Interpretación general:

- Relaciones lineales apropiadas
- No se necesitan transformaciones

Clave: Si las líneas divergen significativamente → considerar transformaciones

Multicolinealidad

¿Qué es? Correlación alta entre variables predictoras

Consecuencias:

1. **Varianza inflada**: Errores estándar muy grandes
2. **Inestabilidad**: Pequeños cambios en datos $\rightarrow$ grandes cambios en coeficientes
3. **Contradicciones**: Modelo globalmente significativo pero ningún predictor individual significativo

Nota importante: La multicolinealidad **NO viola** los supuestos de Gauss-Markov, pero **arruina la interpretación práctica**

Detección:

- **Matriz de correlaciones**: Correlaciones > 0.8 son señal de alerta
- **VIF**: Herramienta definitiva de diagnóstico

Factor de Inflación de la Varianza

Proceso de cálculo del VIF para X_j:

1. Regresar X_j sobre **todas las demás variables predictoras**
2. Obtener el R_j^2 de este modelo auxiliar
3. Calcular:

$$VIF_j = \frac{1}{1 - R_j^2}$$

Interpretación: Factor por el cual se infla la varianza de $\hat{\beta}_j$ debido a multicolinealidad

Reglas prácticas:

- **VIF $= 1$**: Ausencia de colinealidad (ideal)
- **VIF > 5**: Valores preocupantes que requieren atención
- **VIF > 10**: Multicolinealidad seria que debe ser tratada

Ejemplo: Diagnóstico de Multicolinealidad

```
Caso 1: Sin problemas de multicolinealidad

  superficie habitaciones   antiguedad
        1.40         1.40         1.01
distancia_centro           garaje
            1.01             1.01


Caso 2: Con multicolinealidad problemática

  superficie_sim habitaciones_sim metros_cuadrados
            86.4              8.5             81.2


Correlación superficie-metros_cuadrados: 0.994
```

Soluciones a la Multicolinealidad

La estrategia depende del objetivo del análisis:

1. **No hacer nada**
   - Si el objetivo es **predicción**
   - Si variables colineales no son de interés
2. **Eliminar variables**
   - Quitar la menos relevante teóricamente
   - Mantener la más correlacionada con Y
3. **Combinar variables**
   - Crear índices compuestos
   - Análisis de Componentes Principales
4. **Métodos alternativos**
   - **Ridge regression**: Reduce varianza añadiendo sesgo
   - **Lasso/Elastic Net**: Regresión penalizada
5. **Aumentar muestra**
   - Más datos pueden reducir correlaciones
   - No siempre factible

Observaciones Influyentes en Regresión Múltiple

Conceptos básicos (como en regresión simple):

- **Outlier**: Residuo grande
- **Leverage**: Valor atípico en predictores
- **Influencia**: Impacto en el modelo

Herramientas específicas de regresión múltiple:

DFBETAS: Influencia sobre coeficientes individuales

$$\text{DFBETA}_{j,i} = \frac{\hat{\beta}_j - \hat{\beta}_{j(-i)}}{\text{se}(\hat{\beta}_{j(-i)})}$$

Criterio: $|extDFBETA_{j,i}| > \frac{2}{\sqrt{n}}$ es problemático

Ventaja: Permite identificar qué observaciones afectan a qué coeficientes específicos

Gráficos de Regresión Parcial

Objetivo: Visualizar la relación entre Y y X_j **después de eliminar el efecto lineal de todos los demás predictores**

Construcción:

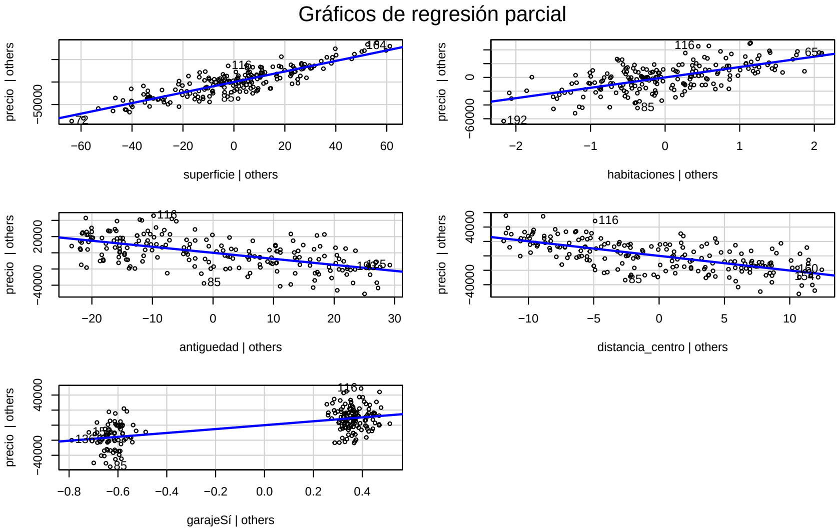
1. Residuos de Y regresado sobre todos los predictores excepto X_j: $e_{Y|X_{-j}}$
2. Residuos de X_j regresado sobre todos los demás predictores: $e_{X_j|X_{-j}}$
3. Graficar: $e_{Y|X_{-j}}$ vs $e_{X_j|X_{-j}}$

Propiedad mágica: La pendiente de la línea ajustada es **exactamente** $\hat{\beta}_j$

Utilidades:

- Visualizar magnitud y significancia del efecto "ajustado"
- Detectar no-linealidades en relaciones parciales
- Identificar observaciones influyentes para coeficientes específicos

Ejemplo: Gráficos de Regresión Parcial



Interpretación de los Gráficos de Regresión Parcial

¿Qué vemos en cada gráfico?

- **Eje X**: Residuos de X_j vs. todos los demás predictores
- **Eje Y**: Residuos de Y vs. todos los demás predictores (excepto X_j)
- **Pendiente**: Es exactamente el coeficiente $\hat{\beta}_j$ del modelo múltiple

Interpretación por variable:

- **Superficie**: Relación lineal clara, pendiente positiva
- **Habitaciones**: Relación positiva, algunos puntos influyentes
- **Antigüedad**: Relación negativa evidente
- **Distancia**: Relación negativa clara
- **Garaje**: Separación clara entre grupos (No/Sí)

Resumen y Conceptos Clave

La regresión múltiple permite:

1. **Efectos parciales:** Aislar el impacto de cada variable predictora
2. **Control de confusores:** Reducir sesgos por variables omitidas
3. **Mejores predicciones:** Incorporar múltiples fuentes de información
4. **Relaciones complejas:** Modelar fenómenos multifactoriales

Aspectos críticos:

- **Interpretación condicional:** Los coeficientes son efectos parciales (*ceteris paribus*)
- **Notación matricial:** Fundamental para la comprensión y computación
- **Supuestos:** Base para las propiedades de los estimadores
- R^2 **ajustado:** Mejor que R^2 para comparar modelos
- **Inferencia:** Tests individuales (t) y global (F)

Próximo paso: Diagnóstico del modelo y tratamiento de problemas específicos