Machine Learning

Hernán Aguirre

Universidad San Francisco de Quito

Universidad de Shinshu, Japón

Guía

- ¿Qué estudiaríamos en este curso?
    - Contenido del Curso
    - Objetivos
- Libros y materiales
- Preparación del ambiente de estudio

Contenido del Curso

1. **Introducción**
2. **Aprendizaje estadístico**
3. **Regresión lineal**
4. **Clasificación**
5. **Métodos de remuestreo**
6. **Selección y regularización de modelos lineales**
7. **Más allá de la linealidad**
8. **Métodos basados en árboles**
9. **Máquinas de vectores de soporte**
10. Aprendizaje profundo
11. Análisis de supervivencia y datos censurados
12. Aprendizaje sin supervisión
13. Pruebas múltiples

3 Regresión Lineal

1. Regresión Lineal Simple
2. Regresión Lineal Múltiple
3. Otras Consideraciones en el Modelo de Regresión
4. El Plan de Marketing
5. Comparación de Regresión Lineal con K-Vecinos más Cercanos
6. Laboratorio: Regresión Lineal
7. Ejercicios

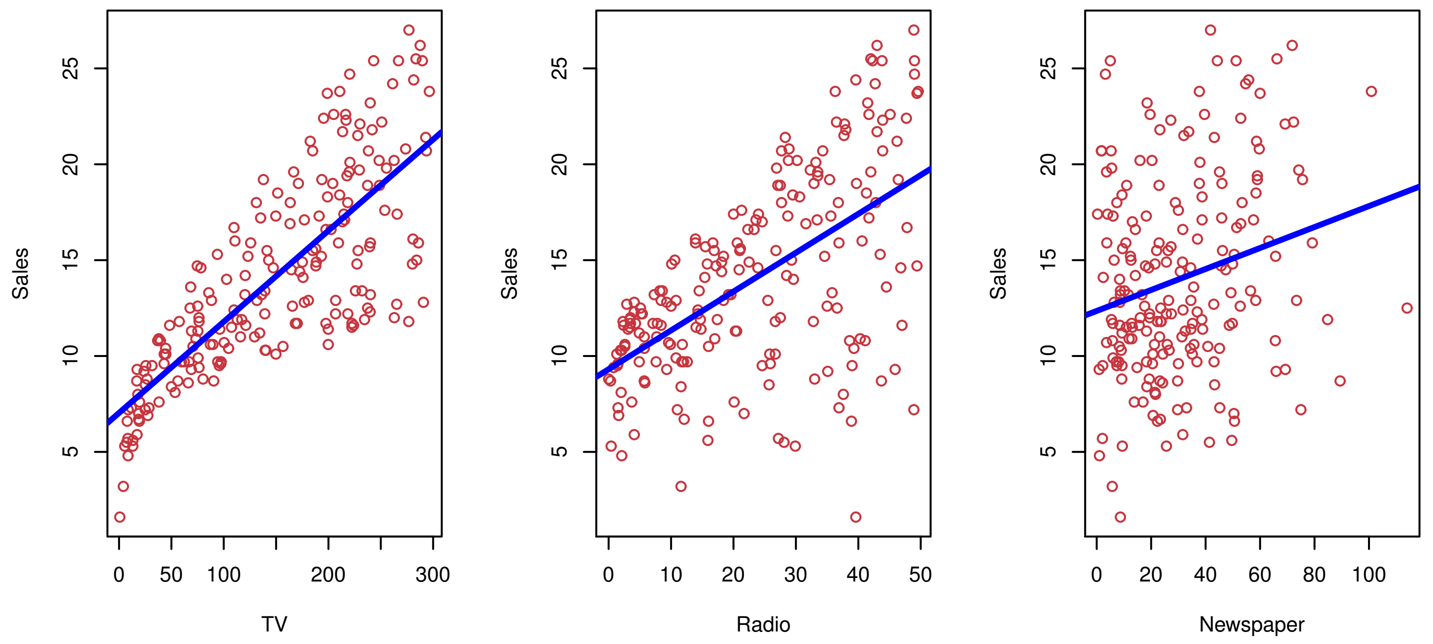
Contexto

- Enfoque muy simple
  - Aprendizaje supervisado
  - Util para predecir una respuesta cuantitativa
- Existe desde hace mucho tiempo y es muy estudiado
- Aburrido en comparación con enfoques recientes?
  - RL sigue siendo útil y ampliamente utilizado
  - Enfoques sofisticados de aprendizaje → generalizaciones o extensiones de RL
  - Importante tener una buena comprensión de RL antes de estudiar métodos más complejos

Objectivo

- Revisar algunas de las ideas clave que subyacen al modelo de regresión lineal,
- Así como el enfoque de mínimos cuadrados que se usa más comúnmente para ajustar este modelo

Publicidad y Ventas



- Ventas en función de los presupuestos de TV, radio y periódicos para 200 mercados diferentes (en miles de dólares)
- Ajuste de *mínimos cuadrados* de las ventas a cada una de las variables
- Línea azul representa un modelo simple que se puede utilizar para **predecir** las ventas en base al presupuesto en TV, radio y periódicos.

Plan de Marketing

- Se nos pide que recomendemos, sobre la base de estos datos, un plan de marketing para el próximo año que resulte en altas ventas de productos

- ¿Qué información sería útil para hacer tal recomendación?

Preguntas Importantes

1. ¿Existe alguna relación entre el presupuesto publicitario y las ventas?
2. ¿Qué tan fuerte es la relación?
3. ¿Qué medios están asociados con las ventas?
4. ¿Qué tan grande es la asociación entre cada medio y las ventas?
5. ¿Con qué precisión podemos predecir ventas futuras?
6. ¿Es la relación lineal?
7. ¿Existe sinergia entre los medios publicitarios?

Regresión Lineal es últil para responder estas preguntas

Regresión Lineal Simple

- Enfoque muy sencillo para predecir una respuesta cuantitativa Y sobre la base de una única variable predictiva X

- El modelo supone que existe aproximadamente una *relación lineal* entre X e Y

$$Y \approx \beta_0 + \beta_1 X$$

regresión de Y sobre X

Ejemplo RLS

- X puede representar *publicidad en TV*
- Y puede representar *ventas*.
- Podemos hacer una regresión de las *ventas* a la *publicidad en TV* ajustando el modelo

$$ventas \approx \beta_0 + \beta_1 TV$$

- β_0 y β_1 constantes desconocidas
  - Coeficientes o parámetros del modelo
  - Intersección y pendiente en el modelo lineal

Ejemplo RLS (2)

- Usamos los datos de entrenamiento para estimar valores de los coeficientes $\beta_0 \Rightarrow \hat{\beta}_0$ y $\beta_1 \Rightarrow \hat{\beta}_1$

- Podemos predecir las ventas futuras sobre la base de un valor particular de la publicidad en TV calculando

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

- $\hat{y}$ indica una predicción de Y sobre la base de $X = x$

Estimación los Coeficientes

- Suma residual de cuadrados (RSS)
- Minimos cuadrados → Minimiza RSS
    - Ecuaciones precisas para calcular los coeficientes β_0, β_1
    - Métodos de optimización que estimar los coeficientes (para múltiples predictores)

Suma Residual de Cuadrados (RSS)

Datos $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$

TV	Ventas	Learning	Predicción $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$	Error
x_1	y_1		$\hat{y}_1$	$e_1 = y_1 - \hat{y}_1$
x_2	y_2	$y_i \approx \hat{\beta}_0 + x\hat{\beta}_1 x_i$	$\hat{y}_2$	$e_2 = y_2 - \hat{y}_2$
$\vdots$	$\vdots$		$\vdots$	$\vdots$
x_n	y_n		$\hat{y}_n$	$e_n = y_n - \hat{y}_n$

Mínimos Cuadrados → Minimiza RSS

$$RSS = e_1^2 + e_2^2 \cdots + e_n^2$$

$$RSS = (y_1 - \hat{\beta}_0 - \hat{\beta}_1 x_1)^2 + (y_2 - \hat{\beta}_0 - \hat{\beta}_1 x_2)^2 + \cdots + (y_n - \hat{\beta}_0 - \hat{\beta}_1 x_n)^2$$

Mínimos Cuadrados

- Elige valores de $\beta_0 \ y \ \beta_1$ que minimizan la suma de residuos cuadrados RSS

pendiente $$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

intersecccción $$\hat{\beta}_0 = \bar{y} - \beta_1 \bar{x}$$

- $\bar{x} \ y \ \bar{y}$ promedios de la muestra

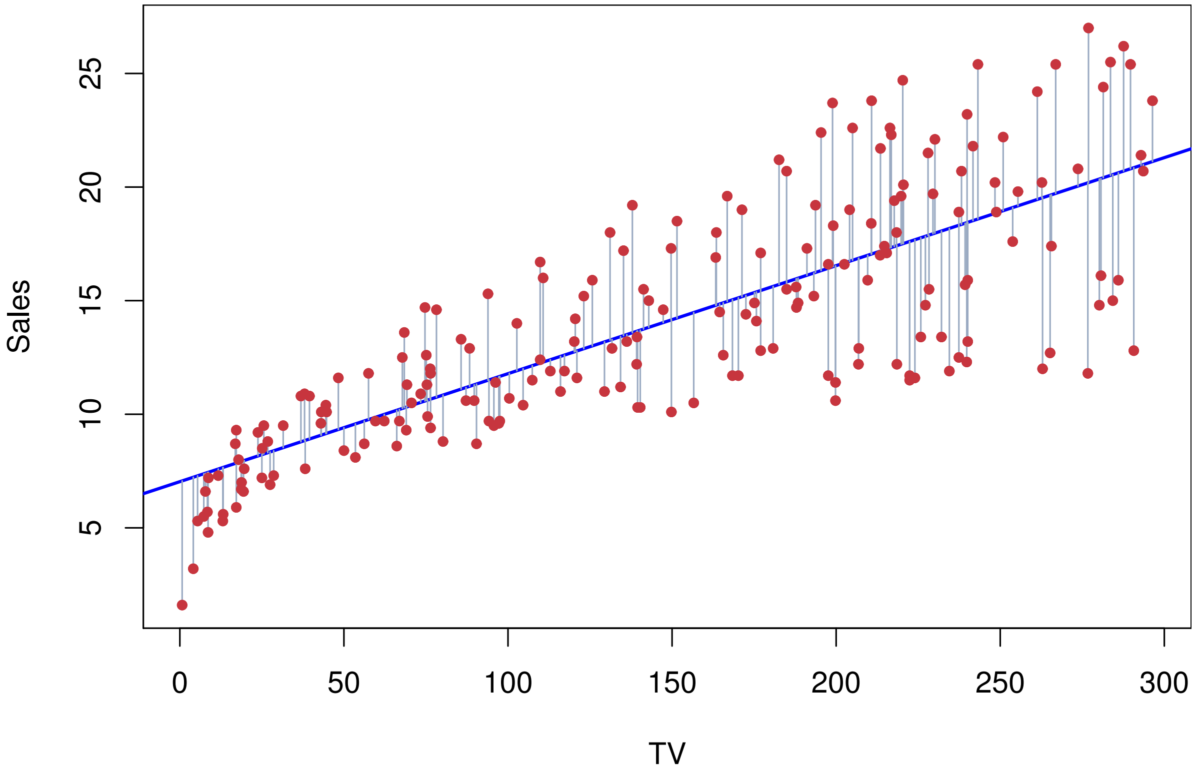
- Ecuaciones precisas para una y dos variables

$$ventas \approx \hat{\beta}_0 + \hat{\beta}_1 TV$$

$\hat{\beta}_0 = 7{,}03$

$\hat{\beta}_1 = 0{,}0475$

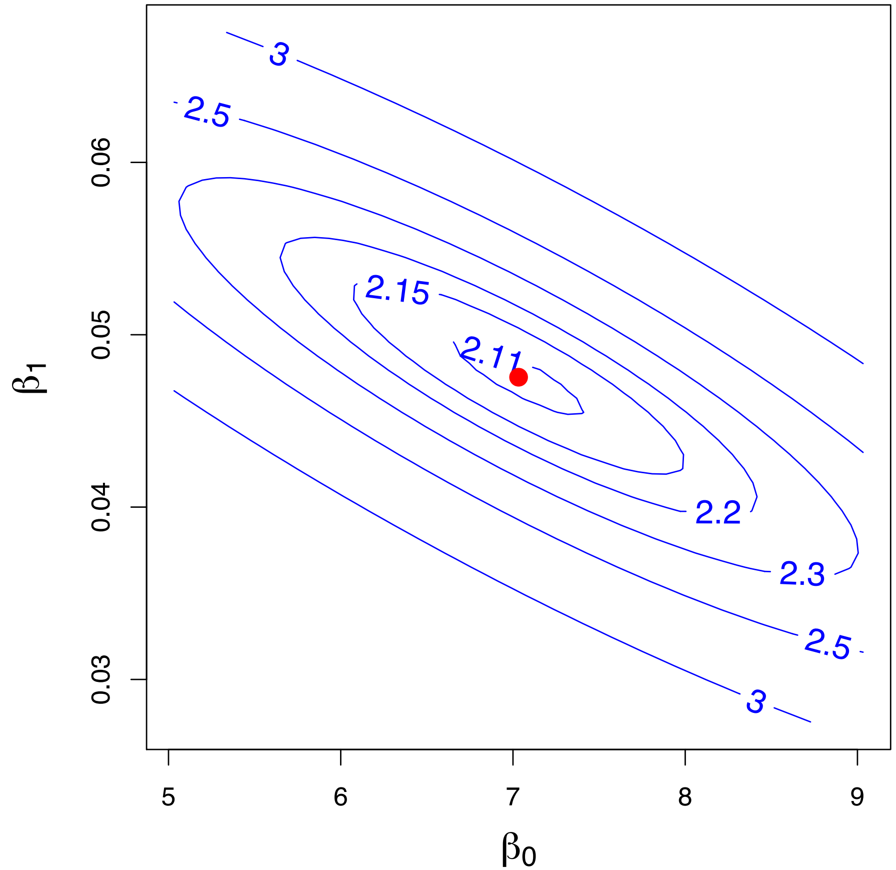
1.000 dólares adicionales en TV se asocian con la venta de 47,5 unidades adicionales

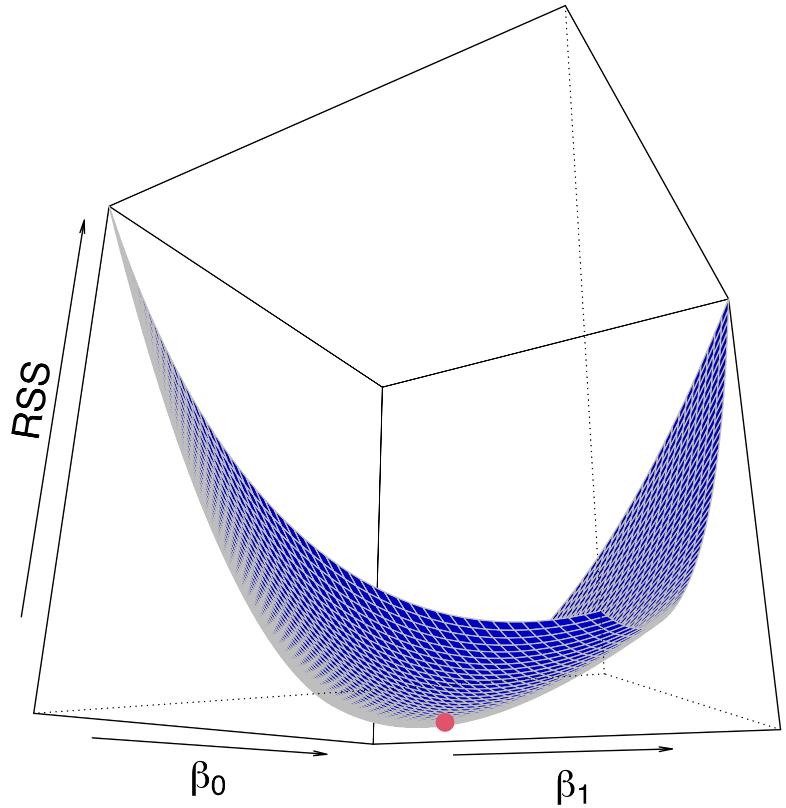


- Mínimos cuadrados ajustados para la regresión de las ventas en TV
- Ajuste minimiza la suma residual de cuadrados
- Cada segmento de línea vertical gris representa un residual
- Un ajuste lineal captura la esencia de la relación
  - ✦ Sobreestima la tendencia en el lado izquierdo del gráfico

RSS : varios $\hat{\beta}_0$ y $\hat{\beta}_1$

$$\mathit{ventas} \approx \hat{\beta}_0 + \hat{\beta}_1 \mathit{TV}$$





$$RSS = (y_1 - \hat{\beta}_0 - \hat{\beta}_1 x_1)^2 + (y_2 - \hat{\beta}_0 - \hat{\beta}_1 x_2)^2 + \cdots + (y_n - \hat{\beta}_0 - \hat{\beta}_1 x_n)^2$$

3.1.2 Determinando la Exactitud de las Estimaciones de los Coeficientes

- Suponemos que la verdadera relación toma la forma

$$Y = f(X) + \epsilon$$

  - f, función desconocida
  - ϵ, término de error aleatorio de media cero

- Si f debe aproximarse mediante una función lineal

$$Y = \beta_0 + \beta_1 X + \epsilon$$

  - β_0 término de intercepción, valor de Y cuando $X = 0$
  - β_1 pendiente, el aumento promedio en Y asociado con un aumento de una unidad en X

ϵ

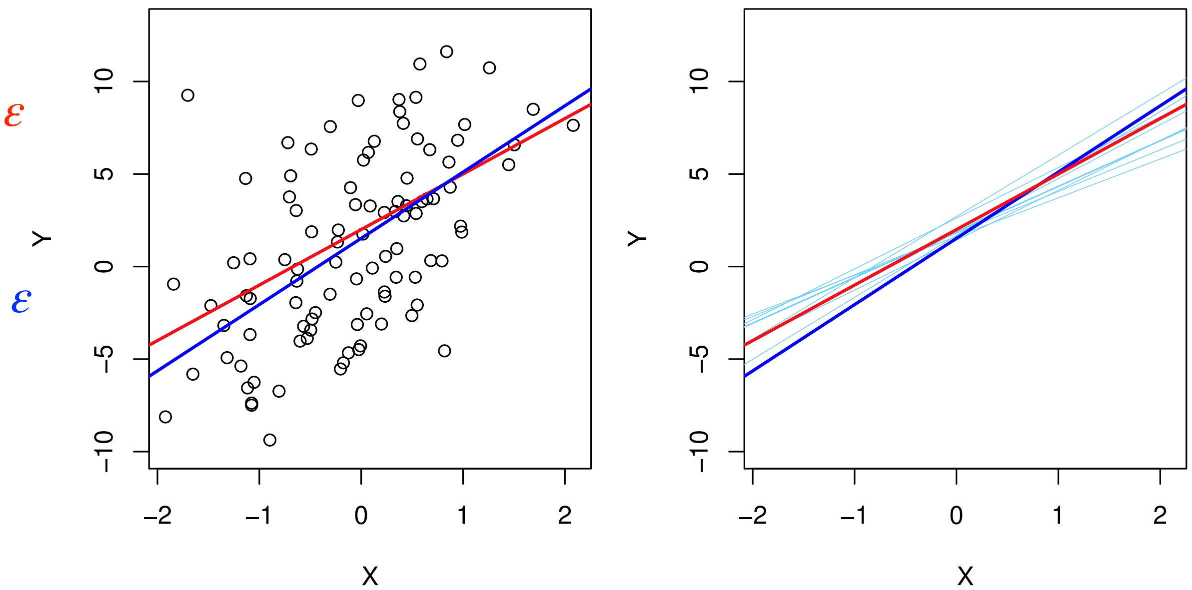
- El término de error abarca todo lo que se nos escapa en este modelo simple:
  - la verdadera relación probablemente no sea lineal
  - puede haber otras variables que causen variación en Y
  - puede haber un error de medición
  - Normalmente suponemos que el término de error es independiente de X

Lineas de Regresión Poblacional y Mínimos Cuadrados

$Y = \beta_0 + \beta_1 X + \varepsilon$

$Y = 2 + 3X + \varepsilon$

$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X + \varepsilon$



Diferentes conjuntos de datos dan como resultado *líneas de mínimos cuadrados* diferentes

- Modelo: *línea de regresión poblacional*, la mejor aproximación lineal
- Se generan 100 X aleatorias y calculan 100 Y a partir del modelo
- Estimaciones de $\hat{\beta}_0, \hat{\beta}_1$ por mínimos cuadrados: *línea de mínimos cuadrados*

Promedio $\hat{\beta}_k$

- El promedio de los coeficcientes estimados sobre muchos conjuntos de datos (m) será muy próximo al valor del coeficiente β_k del modelo ideal

$$\overline{\hat{\beta}_k} = \frac{\sum_{j=1}^{m} \hat{\beta}_{k.j}}{m} \sim \beta_k$$

- Pero una sola estimación $\hat{\beta}_{k,j}$ puede substancialmente sobreestimar o subestimar β_k

Error Estándar

- Que tan alejado está un único estimado $\hat{\beta}_{k,j}$?
  - Error estándar de $\hat{\beta}_k$, $SE(\hat{\beta}_k)$
  - $Var(\hat{\beta}_k) = SE(\hat{\beta}_k)^2 = \dfrac{\sigma^2}{n}$
  - σ es la desviación estándard de cada una de las realizaciones de $y_i \in Y$
- A groso modo, el error estándar nos dice el valor promedio que el valor estimado $\hat{\beta}_k$ difiere del valor real β_k del coeficiente

Error Estándar (2)

$$\sigma^2 = Var(\epsilon)$$

- Formulas para una variable (bajo ciertas suposiciones)

$$SE(\hat{\beta}_0) = \sigma^2[\frac{1}{n} + \frac{\bar{x}^2}{\sum_1^n (x_i - \bar{x})^2}], \quad SE(\hat{\beta}_1) = \frac{\sigma^2}{\sum_1^n (x_i - \bar{x})^2}$$

- $SE(\hat{\beta}_1)$ es menor cuando los datos x_i están más esparcidos → más *leverage* para calcular la pendiente

- $SE(\hat{\beta}_0)$ sería el mismo que $SE(\hat{u})$ si $\overline{x}$ fuera 0, en cuyo caso $\hat{\beta}_0$ sería $\overline{y}$

Error Stándar Residual

- En general σ^2 es desconocida pero puede ser estimada a partir de los datos

- El estimado de σ se conoce como el *error stándar residual* dado por la fórmula

- $RSE = \sqrt{\dfrac{RSS}{n-2}}$

- $RSS = (y_1 - \hat{\beta}_0 - \hat{\beta}_1 x_1)^2 + \cdots + (y_n - \hat{\beta}_0 - \hat{\beta}_1 x_n)^2$

Error Stándar Residual (2)

- Intervalo de confianza
- Pruebas de hipótesis sobre los coeficientes

Intervalos de Confianza

- Los errores estándar se usan para calcular *intervalos de confianza*
- IC de 95% se define como un rango de valores tales que con probabilidad del 95% el rango contiene el valor verdadero desconocido del parámetro.
- Para regresión lineal el IC del 95% para β_1 toma la siguiente fórmula aproximada
  - $\hat{\beta}_1 \pm \ 2 \cdot SE(\hat{\beta}_1)$
  - $[\ \hat{\beta}_1 - \ 2 \cdot SE(\hat{\beta}_1), \quad \hat{\beta}_1 + \ 2 \cdot SE(\hat{\beta}_1)\ ]$
  - Probabilidad del 95% de que este intervalo contenga el valor verdadero de β_1

IC : Datos de Publicidad

- Intervalo de confianza del 95%
  - $\hat{\beta}_0$ [6.130, 7.935]
  - $\hat{\beta}_1$ [0.042, 0.053]

$$ventas \approx \hat{\beta}_0 + \hat{\beta}_1 TV$$

- En ausencia de publicidad, las ventas serán, en promedio, entre 6.130 y 7.935 unidades.
- Por cada aumento de 1.000 dólares en publicidad televisiva, habrá un aumento medio en las ventas de entre 42 y 53 unidades.

Prueba de Hipótesis

- Los errores estándar también se pueden utilizar para realizar pruebas de hipótesis sobre los coeficientes del modelo

$$Y = \beta_0 + \beta_1 X + \epsilon$$

- La prueba de hipótesis más común implica probar la hipótesis nula
  - H_0: No existe relación entre X e Y $\rightarrow H_0 \;:\; \beta_1 = 0$
- versus la hipótesis alternativa
  - H_a: Existe alguna relación entre X e Y $\rightarrow H_a \;:\; \beta_1 \neq 0$

Prueba de Hipótesis (2)

- Para probar H_0: determinar si $\hat{\beta}_1$, nuestra estimación de β_1, está lo suficientemente lejos de cero
- ¿Qué tan lejos es lo suficientemente lejos?
  - Depende de la precisión de $\hat{\beta}_1$, i.e. $SE(\hat{\beta}_1)$
- $SE(\hat{\beta}_1)$ es pequeño
  - valores relativamente pequeños de $\hat{\beta}_1$ pueden proporcionar evidencia sólida de que $\beta_1 \neq 0$ y existe una relación entre X e Y
- $SE(\hat{\beta}_1)$ es grande
  - $\hat{\beta}_1$ debe ser grande en valor absoluto para que podamos rechazar la hipótesis nula.

Estadística t

- En la práctica, calculamos la estadística t

$$t = \frac{\hat{\beta}_1 - 0}{SE(\hat{\beta}_1)}$$

  - mide el número de desviaciones estándar que $\hat{\beta}_1$está alejado de 0

- Calcular la probabilidad de observar cualquier número $\geq |t|$, asumiendo $\beta_1 = 0$ → *valor-p (p-value)*

- Si no hay una relación enre X e Y se espera una *estadística t* con $n - 2$ grados de libertad

p-value

- Si p-value es pequeño
  - Si no existe una asociación verdadera entre el X e Y, es improbable que se observe una asociasión substancial por azar
  - *Rechazamos la hipótesis nula* $H_0: \ \beta_1 = 0$
  - Podemos inferir que existe una relación entre X e Y
- Valores típicos para rechazar $H_0: \ \beta_1 = 0$ son 5% o 1%
  - corresponden a una *estadística-t* de aprox. 2 y 2.75
  - $n = 30$ grados de libertad,
  - la estadística-t se asemeja a una distribución normal para $n \geq 30$

Detalles de la Regresión

$$ventas \approx \beta_0 + \beta_1 TV$$

	Coefficient	Std. error	t-statistic	p-value
`Intercept`	7.0325	0.4578	15.36	< 0.0001
`TV`	0.0475	0.0027	17.67	< 0.0001

- Coeficientes $\hat{\beta}_0$ y $\hat{\beta}_1$ son bastante grandes relativos a su error estándard SE
- Consecuentemente, las estadísticas t son grandes
- Las probabilidades (*p-value*) de observar tales valores si H_0 es verdad son practicamente 0
- Podemos concluir que $\hat{\beta}_0 \neq 0$ y $\hat{\beta}_1 \neq 0$

3.1.3

Determinando la Exactitud del Modelo

- Error residual estándar RSE
- R^2

Error Stándar Residual

- Una medida de falta de ajuste del modelo a los datos

$$RSE = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} = \sqrt{\frac{1}{n-2}RSS}$$

$$Y = \beta_0 + \beta_1 X + \epsilon$$

- RSE es un estimado de la desviación estándard de ϵ

- La cantidad promedio que la respuesta se desviará de la línea de regresión verdadera

RSE : Datos de Publicidad

- RSE = 3.26
- Las ventas reales en cada mercado se desvían de la línea de regresión verdadera aprox. 3,260 unidades en promedio
- Aun si el modelo es correcto, si β_0 y β_1 verdaderos se conocieran con exactitud, cualquier predicción en ventas con base en los anuncios en TV estarían equivocados por 3,260 unidades en promedio
- ¿Es 3,260 un error de predicción aceptable? Depende en el contexto del problema
- El valor medio de las ventas es 14,000 unidades aprox.
    - Error porcentual: 3,260 / 14,000 = 23%

Estadística R^2

- RSE provee una medida absoluta sobre la falta de ajuste del modelo a los datos

- Se mide en las unidades de Y, no es siempre claro que valor de RSE se considera bueno

- R^2 provee una medida alternativa del ajuste

- Es la proporción de la varianza explicada $R^2 \in [0,1]$

- Independiente de la escala de Y

Estadística R^2 (2)

$$R^2 = \frac{TSS - RSS}{TSS} = 1 - \frac{RSS}{TSS}$$

Suma total de cuadrados $TSS = \sum_{i=1}^{n} (y_i - \bar{y})^2$

Suma de residuos cuadrados $RSS = \sum_{i=1}^{n} (y_i - \hat{y})^2$

Estadística R^2 (3)

- TSS mide la varianza total en la respuesta Y
  - Variabilidad inherente en la respuesta antes de efectuar la regresión
- RSS mide la cantidad de variabilidad que no es explicada después de efectuar la regresión
- TSS - RSS mide la cantidad de variabilidad en la respuesta que es explicada (o removida) al efectuar la regresión

Estadística R^2 (4)

- R^2 mide la proporción de variabilidad en Y que puede ser explicada usando X

- $R^2 \sim 1$ la regresión explica una gran proporción de la variabilidad en la respuesta

- $R^2 \sim 0$ la regresión no explica mucho de la variabilidad en la respuesta

    - El modelo lineal está equivocado

    - La varianza del error σ^2 es alta

R^2 Ventas y TV

$$ventas \approx \beta_0 + \beta_1 TV$$

Quantity	Value
Residual standard error	3.26
R^2	0.612
F-statistic	312.1

- Dos tercios de la variabilidad en ventas se explican por la regresión en TV

R^2 y Correlación

- R^2 es una medida de la relación lineal entre X y Y
- Correlación también es una medida de la relación lineal entre X y Y

$$Cor(X,Y) = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}}$$

- En la regresión lineal simple $R^2 = Cor(X,Y)$
- La correlación de predictores y respuesta no se extiende al caso de regresión con múltiples predictores. R^2, sí!

Regresión Lineal Múltiple

- En la práctica tenemos más de un predictor
    - TV
    - Radio
    - Periódico
- ¿Cómo extender el análisis para acomodar más de un predictor?

Múltiples Regresiones Lineales Simples Separadas

Publicidad

Simple regression of sales on TV

$ventas = \hat{\beta}_0 + \hat{\beta}_1 TV$

	Coefficient	Std. error	t-statistic	p-value
`Intercept`	7.0325	0.4578	15.36	< 0.0001
`TV`	0.0475	0.0027	17.67	< 0.0001

Simple regression of `sales` on `radio`

$ventas = \hat{\beta}_0 + \hat{\beta}_1 radio$

	Coefficient	Std. error	t-statistic	p-value
`Intercept`	9.312	0.563	16.54	< 0.0001
`radio`	0.203	0.020	9.92	< 0.0001

Simple regression of `sales` on `newspaper`

$ventas = \hat{\beta}_0 + \hat{\beta}_1 periodico$

	Coefficient	Std. error	t-statistic	p-value
`Intercept`	12.351	0.621	19.88	< 0.0001
`newspaper`	0.055	0.017	3.30	0.00115

Múltiples Regresiones Lineales Simples (2)

- Ajustar una regresión lineal simple por predictor no es satisfactorio
- Cada uno de los presupuestos está asociado con una ecuación diferente → No es claro cómo hacer una predicción de ventas dados los tres presupuestos de medios
- Cada una de las tres ecuaciones de regresión ignora los otros medios para calcular las estimaciones de los coeficientes
- Si los presupuestos de medios están correlacionados, esto puede producir coeficientes erróneos de la asociación entre medios y ventas

Una Regresión con Múltiples Predictores

- Ampliar el modelo de regresión lineal simple para que pueda acomodar directamente múltiples predictores
- Dar a cada predictor un coeficiente de pendiente separado en un solo modelo
- Modelo de regresión lineal múltiple con p predictores distintos

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p$$

- X_j es el j-ésimo predictor y β_j cuantifica la asociación entre esa variable y la respuesta Y
- β_j es el efecto promedio sobre Y de un aumento de una unidad en X_j, *manteniendo fijos todos los demás predictores*

En el Ejemplo Publicidad

$$ventas = \beta_0 + \beta_1 TV \; + \; \beta_2 radio \; + \; \beta_3 periodico$$

3.2.1

Estimación de los Coeficientes de Regresión

- $\beta_0, \beta_1, \beta_2, \cdots, \beta_p$ son desconocidos y deben ser estimados

- Dadas las estimaciones $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \cdots, \hat{\beta}_p$ se pueden hacer predicciones

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_p x_p$$

Suma Residual de Cuadrados (RSS)

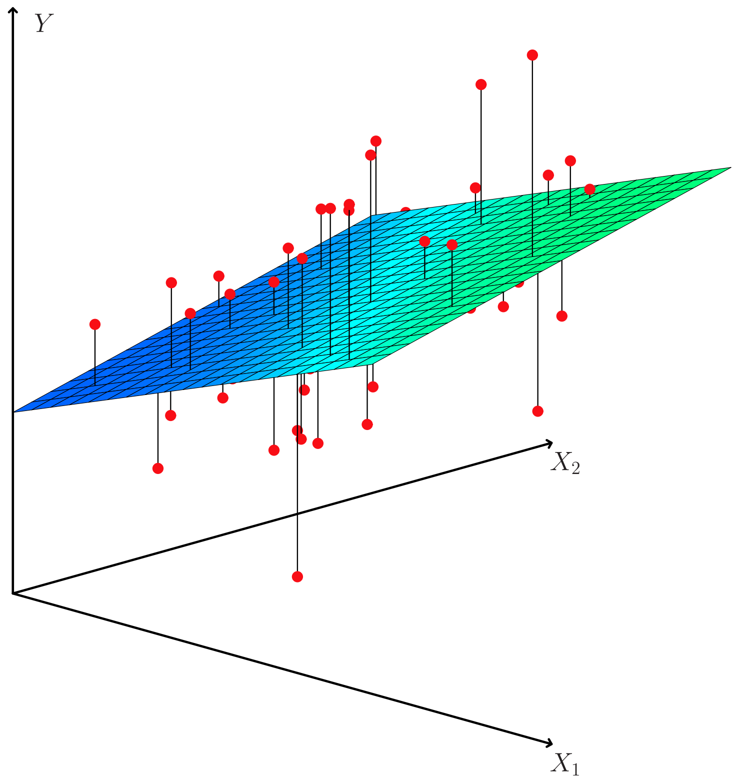
- Se seleccionan los coeficientes que minimizan la suma de residuos cuadrados

$$RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$RSS = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \cdots - \hat{\beta}_p x_{ip})^2$$

- Los valores $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \cdots, \hat{\beta}_p$ que minimizan RSS son los estimados de los múltiples coeficientes de la regresión por *mínimos cuadrados*

Dos Predictores



Con dos predictores y una respuesta, la regresión de mínimos cuadrados se convierte en un plano. El plano escogido minimiza la suma del cuadrado de las distancias verticales entre cada observación y el plano

Detalles Regresión Múltiple

Publicidad

$$ventas = \hat{\beta}_0 + \hat{\beta}_1\ TV + \hat{\beta}_2\ radio + \hat{\beta}_3\ periodico$$

	Coefficient	Std. error	t-statistic	p-value
Intercept	2.939	0.3119	9.42	< 0.0001
TV	0.046	0.0014	32.81	< 0.0001
radio	0.189	0.0086	21.89	< 0.0001
newspaper	-0.001	0.0059	-0.18	0.8599

- Interpretación:

  Para un nivel de dado de publicidad en TV y periódico, $1000 adicionales de gasto de publicidad en radio resulta en 189 unidades adicionales de ventas.

Múltiples Regresiones Lineales Simples Separadas

Publicidad

Simple regression of sales on TV

$ventas = \hat{\beta}_0 + \hat{\beta}_1 TV$

	Coefficient	Std. error	t-statistic	p-value
Intercept	7.0325	0.4578	15.36	< 0.0001
TV	0.0475	0.0027	17.67	< 0.0001

Simple regression of sales on radio

$ventas = \hat{\beta}_0 + \hat{\beta}_1 radio$

	Coefficient	Std. error	t-statistic	p-value
Intercept	9.312	0.563	16.54	< 0.0001
radio	0.203	0.020	9.92	< 0.0001

Simple regression of sales on newspaper

$ventas = \hat{\beta}_0 + \hat{\beta}_1 periodico$

	Coefficient	Std. error	t-statistic	p-value
Intercept	12.351	0.621	19.88	< 0.0001
newspaper	0.055	0.017	3.30	0.00115

Múltiples Regresiones Simples vs Regresión con Múltiples Predictores

Simple regr

	Coefficient
Intercept	7.0325
TV	0.0475

Simple regres

	Coefficient
Intercept	9.312
radio	0.203

Simple regressi

	Coefficient	Std. error	t-statistic	p-value
Intercept	12.351	0.621	19.88	< 0.0001
newspaper	0.055	0.017	3.30	0.00115

	Coefficient
Intercept	2.939
TV	0.046
radio	0.189
newspaper	-0.001

Observaciones

- Coeficientes para TV y radio de la regresión múltiple son similares a los coeficientes de las regresiones simples
- Coeficiente para periódico
  - Regresión simple: diferente de cero (significativamente)
  - Regresión múltiple: cercano de cero y el *p-value* no es significativo 0.86

Observaciones (2)

- El *coeficiente de periódico* representa el incremento promedio en ventas de producto asociado con un incremento de $1,000 en gasto de publicidad en periódico,
  - Regresión simple
    $$ventas = \hat{\beta}_0 + \hat{\beta}_1 \; periodico$$
    - ignorando otros predictores como TV y radio
  - Regresión múltiple
    $$ventas = \hat{\beta}_0 + \hat{\beta}_1 TV + \hat{\beta}_2 radio + \hat{\beta}_3 \; periodico$$
    - mientras se mantiene fijo el gasto en TV y radio

¿Hay una relación entre periódico y ventas?

- Regresión múltiple → no
- Regresión simple → sí

Tiene sentido ?

Matriz de correlación predictores y respuesta

	TV	radio	newspaper	sales
TV	1.0000	0.0548	0.0567	0.7822
radio		1.0000	0.3541	0.5762
newspaper			1.0000	0.2283
sales				1.0000

Mercados con alta publicidad en radio también tienen alta publicidad en periódico

- Suponga regresión múltiple es correcta:
  - publicidad en periódico no está asociado a ventas, pero publicidad en radio sí
- Más gasto en radio → las ventas tenderán a subir
- Periódico se lleva crédito de la asociación entre ventas y radio por su correlación con radio

Advertencia

- Contraintuitivo, pero muy común

- Otro ejemplo: Ataques de tiburones vs venta de helados en la playa en el verano

Preguntas Importantes

A. Es al menos uno de los predictores $X_1, X_2, \cdots, X_p$ útil para la predicción de la respuesta?

B. Todos los predictores ayudan a explicar X, o son útiles solamente un subconjunto de predictores?

C. Que tan bien se ajusta el modelo a los datos?

D. Dados un conjunto de valores de los predictores, que valor de respuestas deberíamos predecir, y que tan exacta es nuestra predicción?

A

Existe una Relación entre Respuesta y Predictores?

$$H_0 = \beta_1 = \beta_2 = \cdots = \beta_p = 0$$

$$H_a = \text{al menos un } \beta_j \text{ es no cero}$$

- Prueba de la hipótesis se efectúa usando la estadística F

$$F = \frac{(TSS - RSS)/p}{RSS/(n-p-1)}$$

$$TSS = \sum_{i=1}^{n} (y_i - \bar{y})^2 \qquad RSS = \sum_{i=1}^{n} (y_i - \hat{y})^2$$

Hipótesis y F

$$F = \frac{(TSS - RSS)/p}{RSS/(n - p - 1)}$$

- Si las suposiciones del modelo lineal son correctas

$$E\{RSS/(n - p - 1)\} = \sigma^2$$

- Si H_0 es verdad, **no hay** relación entre predictores y respuesta

$$E\{(TSS - RSS)/p\} = \sigma^2 \quad \rightarrow \quad F \sim 1$$

- Si H_a es verdad, almenos un predictor tiene relación

$$E\{(TSS - RSS)/p\} > \sigma^2 \quad \rightarrow \quad F > 1$$

Estadística F : Publicidad

$$ventas = \hat{\beta}_0 + \hat{\beta}_1 TV + \hat{\beta}_2 radio + \hat{\beta}_3 \ periodico$$

Quantity	Value
Residual standard error	1.69
R^2	0.897
F-statistic	570

- $F >> 1$ evidencia substancial en contral de H_0
- Al menos uno de los medios está relacionado a ventas

- evidencia en contra de H_0
  - n es largo → F un poco mayor a 1
  - n es pequeño → un F grande se requiere

B

Variables Importantes

- Es posible que todos los predictores estén asociados con la respuesta

- Pero lo más frecuente es que la respuesta solo esté asociada con un subconjunto de los predictores.

- Selección de variables: determinar qué predictores están asociados con la respuesta

Como Seleccionar Variables?

- Idealmente: seleccionar variables probando muchos modelos diferentes, cada uno de los cuales contenga un subconjunto diferente de predictores.

  - 2^p modelos, demasiados
  - Selección hacia adelante
  - Selección hacia atrás
  - Selección mixta

Calidad del Modelo

- Seleccionar el mejor modelo de todos los considerados
- ¿Cómo determinamos qué modelo es mejor?
- Se pueden utilizar varias estadísticas para juzgar la calidad de un modelo.
  - Cp de Mallow
  - Criterio de información de Akaike (AIC)
  - Criterio de información bayesiano (BIC)
  - R^2 ajustado

C

Ajuste del Modelo

- RSE
- R^2, la fracción de la varianza explicada
  - $R^2 = Cor(Y, \hat{Y})^2$
- Gráficos

Estadística R^2

$$R^2 = \frac{TSS - RSS}{TSS} = 1 - \frac{RSS}{TSS}$$

Suma total de cuadrados $$TSS = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

Suma de residuos cuadrados $$RSS = \sum_{i=1}^{n} (y_i - \hat{y})^2$$

Variables Adicionales y R^2

Modelo	R^2
$ventas = \hat{\beta}_0 + \hat{\beta}_1 TV$	0.61
$ventas = \hat{\beta}_0 + \hat{\beta}_1 TV + \hat{\beta}_2 radio$	0.89719
$ventas = \hat{\beta}_0 + \hat{\beta}_1 TV + \hat{\beta}_2 radio + \hat{\beta}_3\ periodico$	0.8972

- R^2 aumenta cuando se agreguen más variables, incluso si están sólo débilmente asociadas con la respuesta
- Agregar otra variable disminuye el RSS en los datos de entrenamiento (no necesariamente en los datos de prueba)

Error Stándar Residual

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_p x_p$$

$$RSS = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \cdots - \hat{\beta}_p x_{ip})^2$$

$$RSE = \sqrt{\frac{1}{n - p - 1} RSS}$$

Variables Adicionales y RSE

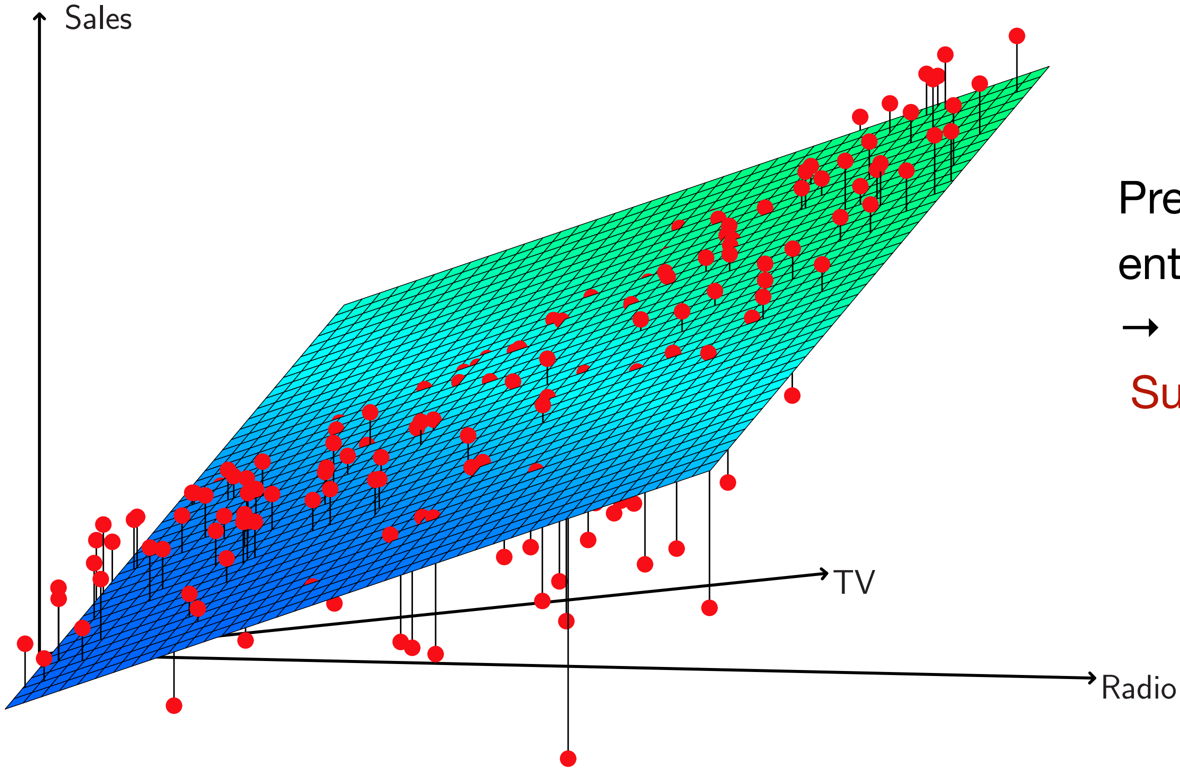
Modelo	RSE
$ventas = \hat{\beta}_0 + \hat{\beta}_1 TV$	3.26
$ventas = \hat{\beta}_0 + \hat{\beta}_1 TV + \hat{\beta}_2 radio$	1.681
$ventas = \hat{\beta}_0 + \hat{\beta}_1 TV + \hat{\beta}_2 radio + \hat{\beta}_3\ periodico$	1.686

- Modelos con más variables pueden tener un RSE más alto si la disminución en RSS es pequeña en relación con el aumento en p

Resúmenes Gráficos

$$ventas = \hat{\beta}_0 + \hat{\beta}_1 TV + \hat{\beta}_2 radio$$

La mayor parte del presupuesto se gastó en televisión o radio
→
Sobreestima las ventas



Presupuesto dividido entre los dos medios
→
Subestima las ventas

Patrón no lineal pronunciado sugiere una sinergia o efecto de interacción entre los medios publicitarios → la combinación de los medios da como resultado un mayor impulso a las ventas que el uso de un solo medio

D

Predicción

- Una vez que hemos ajustado el modelo de regresión múltiple, es sencillo predecir la respuesta Y con base en un conjunto de valores para los predictores $X_1, X_2, \cdots, X_p$

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \cdots + \hat{\beta}_p X_p$$

- Tres fuentes de incertidumbre

Incertidumbre en los Coeficientes

- Los coeficientes $\hat{\beta}_0, \hat{\beta}_1, \cdots, \hat{\beta}_p$ son estimaciones
  - *plano de mínimos cuadrados*

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \cdots + \hat{\beta}_p X_p$$

  - es una estimación del *verdadero plano de regresión poblacional*

$$f(X) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$$

- La inexactitud en las estimaciones de los coeficientes está relacionada con el *error reducible*
- Podemos calcular un *intervalo de confianza* para determinar qué tan cerca estará $\hat{Y}$ de $f(X)$

Incertidumbre en el Modelo

- En la práctica asumir un modelo lineal para $f(X)$ es casi siempre una aproximación de la realidad,
- Por lo que existe una fuente adicional de *error potencialmente reducible* que llamamos *sesgo del modelo*.

Error aleatorio en el Modelo

- Incluso si supiéramos $f(X) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$, es decir, los valores verdaderos de los coeficientes, el valor de la respuesta no se puede predecir perfectamente debido al error aleatorio ϵ en el modelo → *Error irreducible*
- ¿Cuánto variará Y respecto de $\hat{Y}$?
- *Intervalos de predicción* para responder esta pregunta
  - más amplios que los *intervalos de confianza*, incorporan
    - el error en la estimación de $f(X)$ (el error reducible)
    - la incertidumbre sobre cuánto diferirá un punto individual del *plano de regresión poblacional* (el error irreducible)

Intervalos de Confianza y de Predicción

- Utilizamos un *intervalo de confianza* para cuantificar la incertidumbre que rodea a las ventas promedio en un gran número de ciudades.

- Por otro lado, se puede utilizar un *intervalo de predicción* para cuantificar la incertidumbre que rodea las ventas de una ciudad en particular.

3.3

Otras Consideraciones en el Modelo de Regresión

- Predictores cualitativos
- Extensiones del modelo lineal
- Problemas potenciales

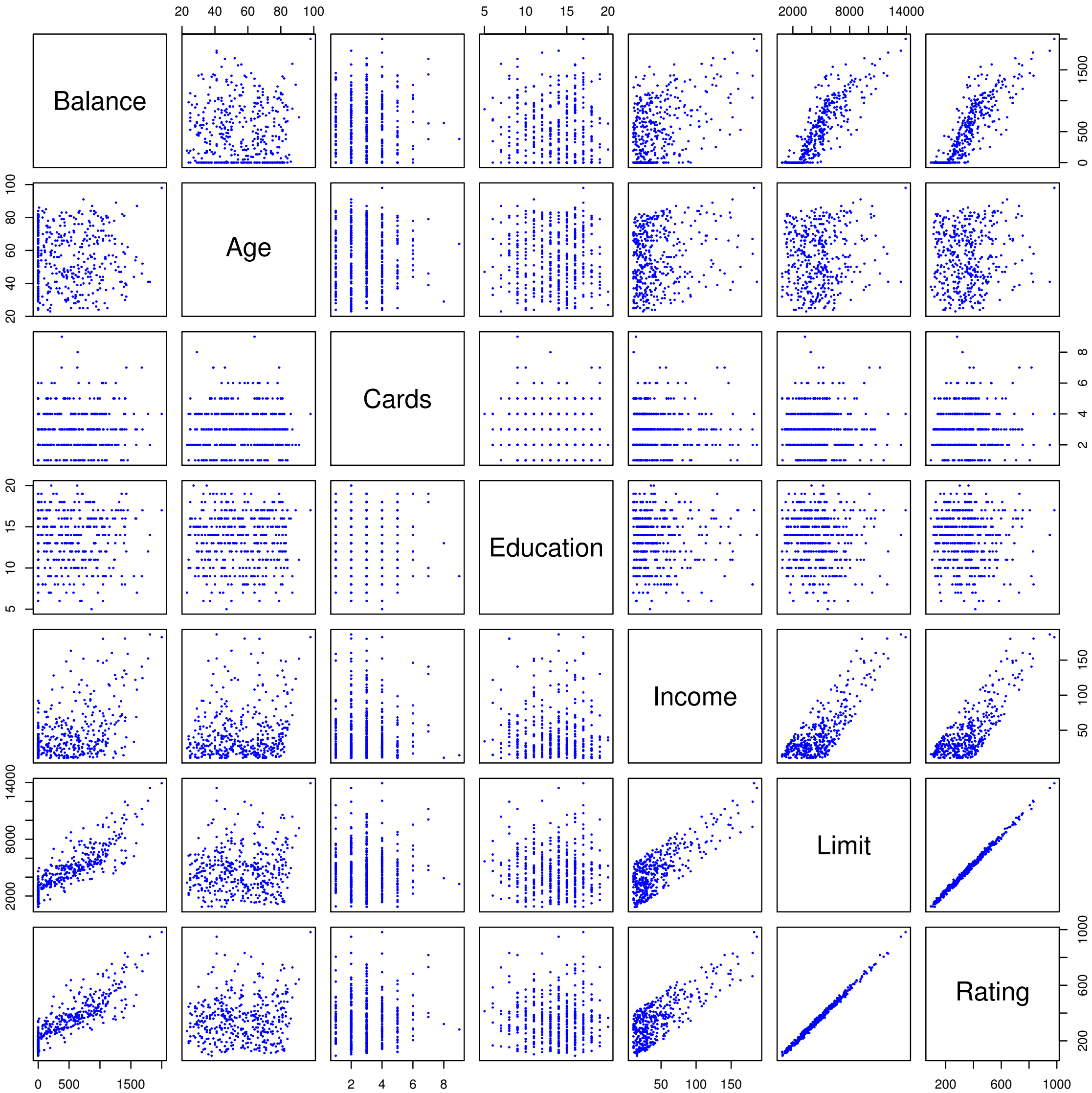
3.3.1

Predictores Cualitativos

- Predictores con sólo dos niveles
- Predictores cualitativos con más de dos niveles

Datos de Crédito

Deuda
Edad
Número de Tarjetas
Años de educación
Ingreso
Límite de crédito
Rating de crédito



Predictores Cualitativos con 2 Niveles

- Investigar la diferencia en la deuda de la tarjeta de crédito entre aquellos que son dueños de una casa y los que no

- Crear un indicador o *variable dummy*

$$x_i = \begin{cases} 1, & \textit{si la i-ésima persona es dueña de una casa} \\ 0, & \textit{caso contrario} \end{cases}$$

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

2 Niveles (2)

$$x_i = \begin{cases} 1, & \textit{dueño} \\ 0, & \textit{caso contrario} \end{cases}$$

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i = \begin{cases} \beta_0 + \beta_1 + \epsilon_i, & \textit{dueño} \\ \beta_0 + \epsilon_i, & \textit{caso contratio} \end{cases}$$

$$balance = \beta_0 + \beta_1 own$$

β_0 balance promedio no dueños

$\beta_0 + \beta_1$ balance promedio dueños

β_1 diferencia promedio en balance entre dueños y no dueños

	Coefficient	Std. error	t-statistic	p-value
Intercept	509.80	33.13	15.389	< 0.0001
own[Yes]	19.73	46.05	0.429	0.6690

2 Niveles (3)

$$x_i = \begin{cases} 1, & \textit{dueño} \\ -1, & \textit{caso contrario} \end{cases}$$

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i = \begin{cases} \beta_0 + \beta_1 + \epsilon_i, & \textit{dueño} \\ \beta_0 - \beta_1 + \epsilon_i, & \textit{caso contratio} \end{cases}$$

$$balance = \beta_0 + \beta_1 own$$

β_0 balance promedio (sin considerar el efecto de propiedad)

β_1 cantidad que dueños o no dueños están por arriba o por debajo del promedio

$$x_i = \begin{cases} 1, & \textit{dueño} \\ 0, & \textit{caso contrario} \end{cases}$$

β_0 = \$519.665

β_0 = \$9.865

	Coefficient
`Intercept`	509.80
`own[Yes]`	19.73

2 Niveles (4)

- La predicción final para los balances de la tarjeta de crédito será exactamente la misma sin importar la codificación usada

$$x_i = \begin{cases} 1, & \text{dueño} \\ 0, & \text{caso contrario} \end{cases} \qquad x_i = \begin{cases} 1, & \text{dueño} \\ -1, & \text{caso contrario} \end{cases}$$

Predictores Cualitativos con más de 2 Niveles

- Una sola variable *dummy* no puede representar todas los valores posibles
- Se crean variables *dummy* adicionales
- Por ejemplo, deseo investigar el balance en la tarjeta de crédito respecto a la región donde viven los individuos
  - Norte, Sur, Este, Oeste
  - Creo dos variables

Más de 2 Niveles (2)

$$x_{i1} = \begin{cases} 1, & \textit{del Sur} \\ 0, & \textit{no del Sur} \end{cases} \qquad x_{i2} = \begin{cases} 1, & \textit{del Oeste} \\ 0, & \textit{no del Oeste} \end{cases}$$

$$y_i = \beta_0 + \beta_1 x_{i1} + + \beta_2 x_{i2} \epsilon_i = \begin{cases} \beta_0 + \beta_1 + \epsilon_i, & \textit{del Sur} \\ \beta_0 + \beta_2 + \epsilon_i, & \textit{del Oeste} \\ \beta_0 + \epsilon_i, & \textit{del Este} \end{cases}$$

	Coefficient	Std. error	t-statistic	p-value
`Intercept`	531.00	46.32	11.464	< 0.0001
`region[South]`	−12.50	56.68	−0.221	0.8260
`region[West]`	−18.69	65.02	−0.287	0.7740

3.3.2

Extensiones del Modelo Lineal

- Eliminación del supuesto aditivo
- Relaciones no lineales

Dos Supocisiones Importantes

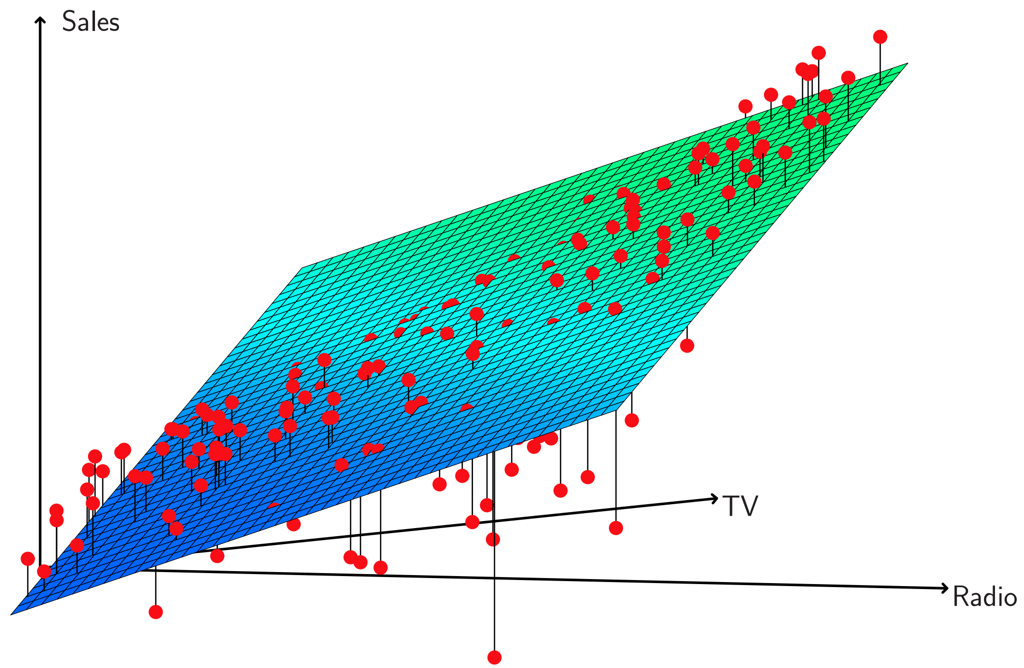
- Relaciones entre los predictores y la respuesta son aditivas y lineales
- Suposición de Aditividad
  - La relación entre el predictor X_j y la respuesta Y no depende de los valores de otros predictores
- Suposición de Linealidad
  - Cambios en la respuesta Y asociados con el cambio de una unidad en X_j es constante, sin importar el valor de X_j
- Hay métodos sofisticados que relajan estas suposiciones
- Veamos aproximaciones clásicas para extender el modelo

Removiendo la Suposición de Aditividad

- Datos de Publicidad $ventas = \beta_0 + \beta_1 TV + \beta_2 radio + \beta_3 periodico$
  - TV y radio parecen estar asociados con ventas
  - El modelo asume que los efectos en ventas al incrementar el gasto en un medio de publicidad es independiente de la cantidad gastada en otros medios
  - El incremento promedio en ventas asociado con el incremento en una unidad en TV es siempre β_1, sin importar la cantidad gastada en radio.
  - Este modelo simple puede ser incorrecto

Publicidad

- Supongamos : gastar dinero en radio incrementa la efectividad en TV
- Dado un presupuesto fijo, gastar la mitad en TV y la mitad en radio puede incrementar ventas más que asignando el presupuesto total a TV o radio
- Efecto de interacción (sinergía, epistasis)
- La figura sugiere que tal efecto podría estar presente en los datos de publicidad



- Niveles de TV o radio bajos, las ventas reales son inferiores que las predecidas
- La publicidad se reparte entre los dos medios, el modelo subestima ventas

Regresión Lineal Estándar

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

- Un incremento en una unidad en X_1 está asociado con un incremento promedio en Y Y de β_1 unidades
- La presencia de X_2 no altera esto
- $Y = 10 + 3X_1 + 5X_2$

Y	X_1	X_2
50	10	2
53	11	2
65	10	5
68	11	5

Añadiendo un Término de Interacción

- Producto de X_1 y X_2

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \epsilon$$

- ¿Cómo la inclusión de este término de interacción relaja el supuesto aditivo?

$$Y = \beta_0 + (\beta_1 + \beta_3 X_2) X_1 + \beta_2 X_2 + \epsilon$$

$$Y = \beta_0 + \tilde{\beta}_1 X_1 + \beta_2 X_2 + \epsilon \quad \rightarrow \quad \tilde{\beta}_1 = \beta_1 + \beta_3 X_2$$

- La asociación entre X_1 y Y deja de ser constante
- Un cambio en el valor de X_2 cambia la asociación entre X_1 y Y
- Un argumento similar existe para la asociación entre X_2 y Y, un cambio en X_1 cambia la asociación en tre X_2 y Y

Publicidad: Interacción entre TV y radio

$$Y = \beta_0 + \beta_1 \times TV + \beta_2 \times radio + \beta_3 \times TV \times radio + \epsilon$$

$$= \beta_0 + (\beta_1 + \beta_3 \times radio) \times TV + \beta_2 \times radio + \epsilon$$

	Coefficient	Std. error	t-statistic	p-value
`Intercept`	6.7502	0.248	27.23	< 0.0001
`TV`	0.0191	0.002	12.70	< 0.0001
`radio`	0.0289	0.009	3.24	0.0014
`TV×radio`	0.0011	0.000	20.73	< 0.0001

- Modelo que incluye el *término de interacción* parece superior al que contiene solo los *efectos principales*
- *Principio jerárquico*: Si incluimos una interacción en un modelo, también debemos incluir los efectos principales, aun si los *p-values* asociados con sus coeficientes no son significativos

Interacción y R^2

Modelo	R^2
$ventas = \hat{\beta}_0 + \hat{\beta}_1 TV$	0.61
$ventas = \hat{\beta}_0 + \hat{\beta}_1 TV + \hat{\beta}_2 radio$	0.89719
$ventas = \hat{\beta}_0 + \hat{\beta}_1 TV + \hat{\beta}_2 radio + \hat{\beta}_3\ periodico$	0.8972
$ventas = \hat{\beta}_0 + (\hat{\beta}_1 + \hat{\beta}_3 radio) TV + \hat{\beta}_2 radio$	0.968

(98.6-89.7)/(100-89.7) = 69%

69% de la variabilidad en ventas que queda después de ajustar el modelo aditivo ha sido explicada por el término de interacción

Interacciones con Variables Cualitativas

- Interacciones también se aplican a variables cualitativas o a una combinación de variables cuantitativas y cualitativas

Relaciones no Lineales

- Una manera simple de extender el modelo lineal para *incluir relaciones no lineales* entre el predictor y la respuesta es la *regresión polinomial*

- Permite incluir versiones transformas de los predictores

- Existen otras formas más complejas

Autos

$$mpg = \beta_0 + \beta_1 \times horsepower + \beta_2 \times horsepower^2$$

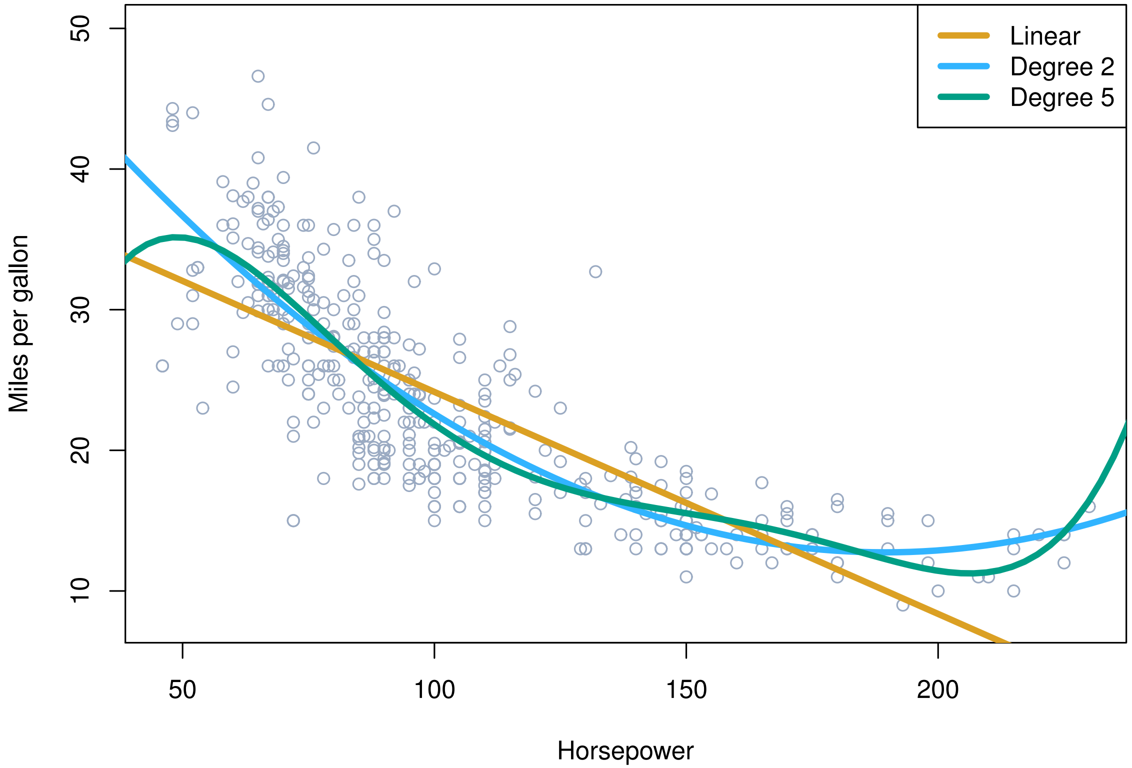
- Permite usar una función no lineal de *horsepower*

- Pero sigue siendo un modelo lineal.

- Más precisamente, un modelo de regresión lineal múltiple con $X_1 = horsepower$ y $X_2 = horsepower^2$

- → Se puede usar algoritmos de regresión lineal para estimar β_0, β_1 y $\beta_2$de tal forma de producir un ajuste no-lineal!

Autos



$$mpg = \beta_0 + \beta_1 \times horsepower + \beta_2 \times horsepower^2$$

	Coefficient	Std. error	t-statistic	p-value
`Intercept`	56.9001	1.8004	31.6	< 0.0001
`horsepower`	-0.4662	0.0311	-15.0	< 0.0001
`horsepower`^{2}	0.0012	0.0001	10.1	< 0.0001

3.3.3

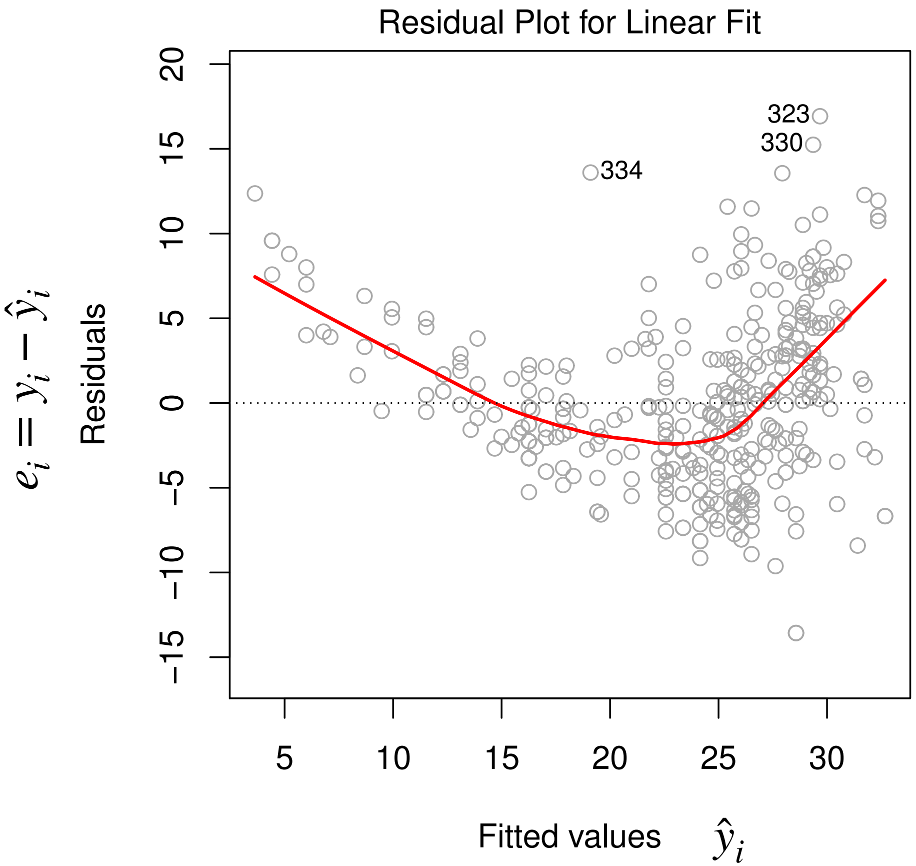
Problemas Potenciales

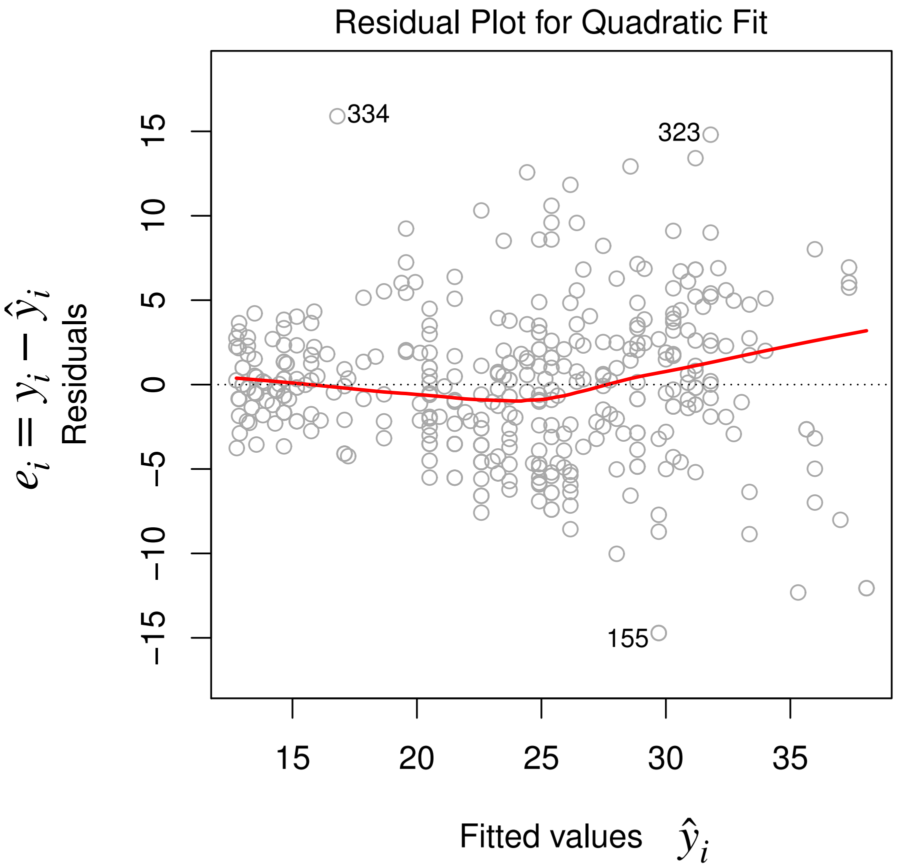
- No linealidad de las relaciones respuesta-predictor.
- Correlación de términos de error.
- Varianza no constante de los términos de error.
- Valores atípicos.
- Puntos de alto apalancamiento.
- Colinealidad.

No Linearidad de los Datos

- Gráficos residuales son una herramienta útil para identificar no linealidad
- Regresión simple : $e_i = y_i - \hat{y}_i$ versus el predictor x_i
- Regresión múltiple : $e_i = y_i - \hat{y}_i$ versus los valores predecidos $\hat{y}_i$
- Idealmente, el gráfico residual no mostrará un patrón discernible. La presencia de un patrón indica un problema con algún aspecto del modelo lineal.
- Si el gráfico residual indica que existe alguna asociación no lineal en los datos → usar transformaciones no lineales de los predictores, tales como $\log X, \sqrt{X}$ y X^2 en el modelo de regresión.
- Otras aproximaciones más sofisticadas para tratar no linealidad

Residuos vs Ajustes





$$mpg = \beta_0 + \beta_1 \times hp$$

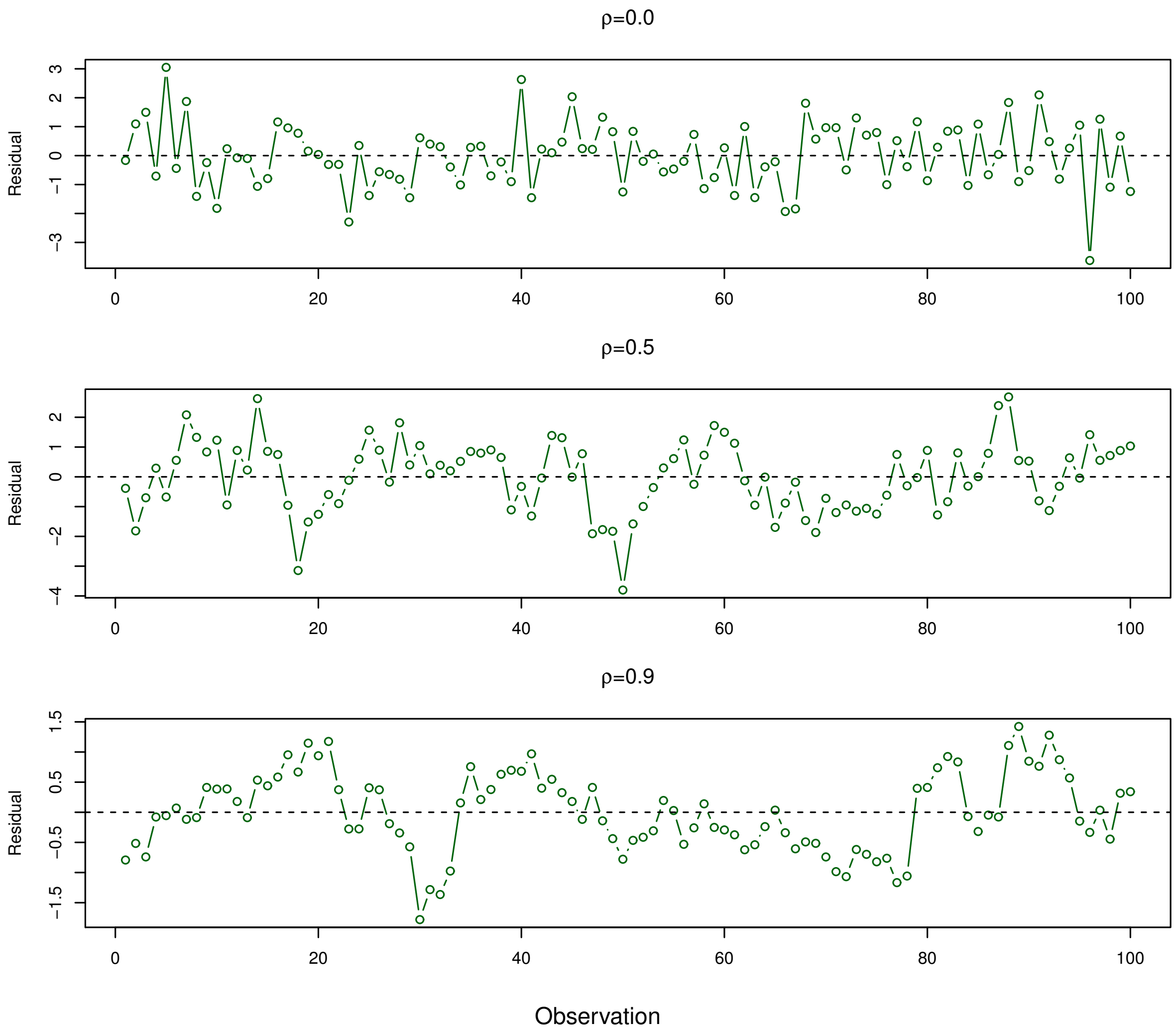
$$mpg = \beta_0 + \beta_1 \times hp + \beta_2 \times hp^2$$

Correlación de los Términos de Error

$$e_i = y_i - \hat{y}_i$$

$$e_i = y_i - \hat{y}_i$$

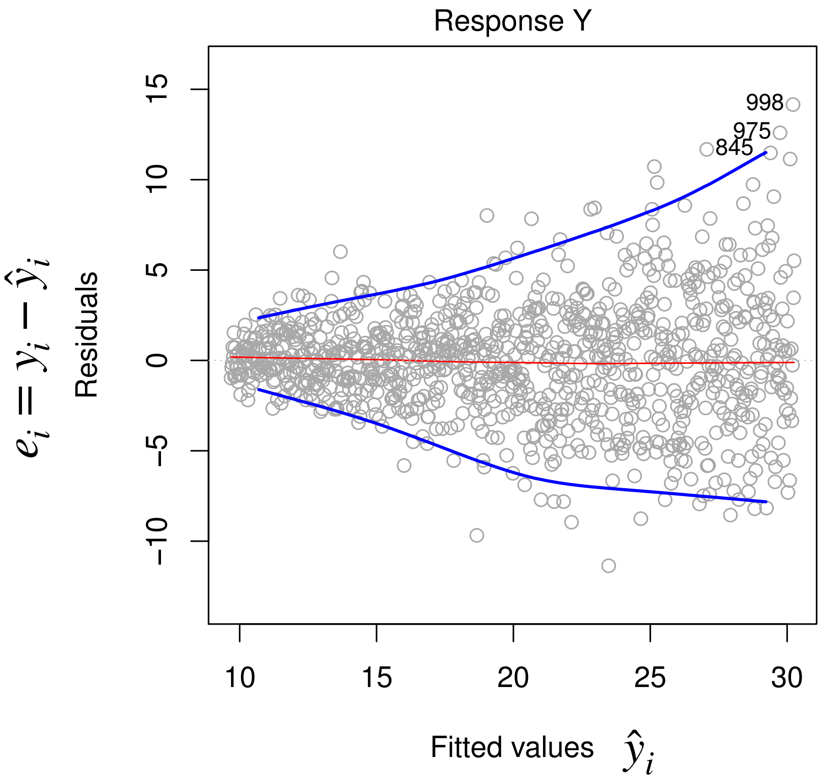
$$e_i = y_i - \hat{y}_i$$

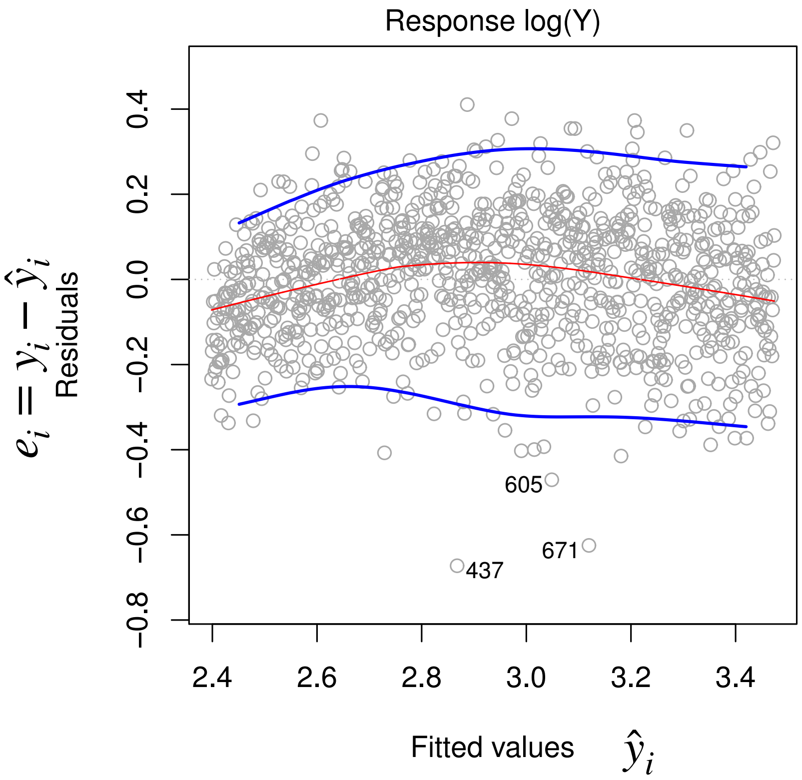


Errores no Correlacionados y Regresión Lineal

- La suposición de errores no correlacionados es extremadamente importante para la regresión lineal, así como para otros métodos estadísticos.
- Por lo tanto, un buen diseño de experimentos es crucial para mitigar el riesgo de tales correlaciones.

Varianza no Constante de los Términos de Error



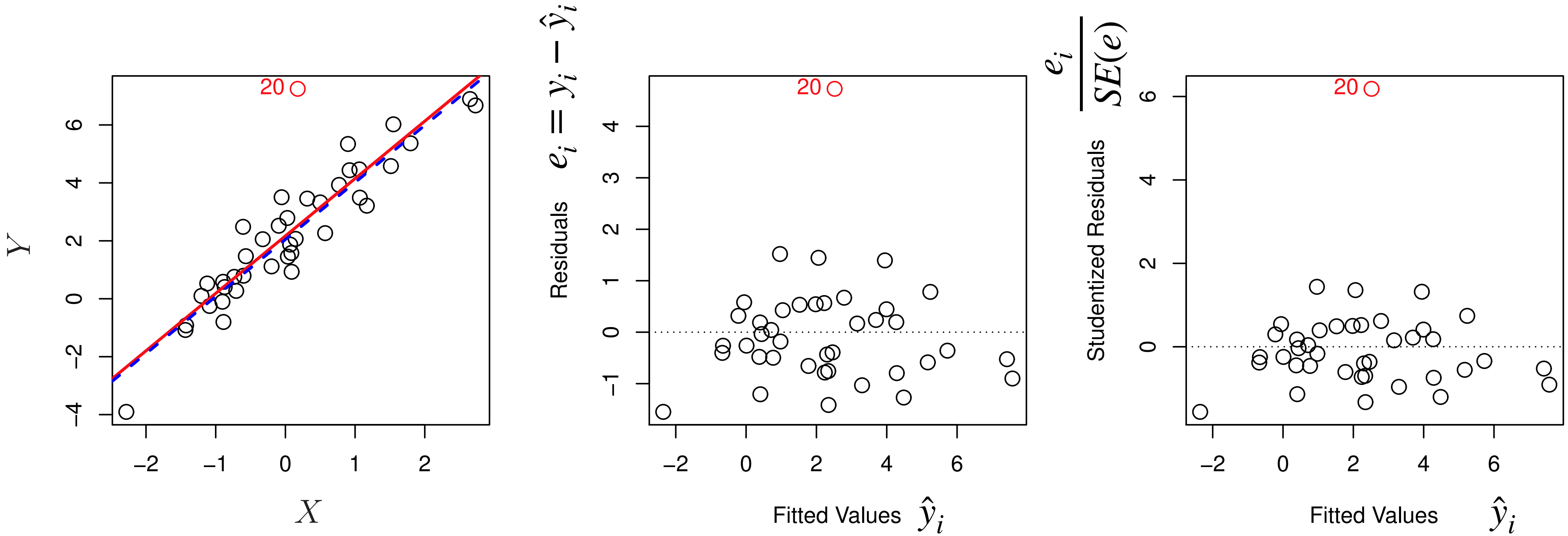


- Heteroscedasticidad, varianza no constante en los errores.

- Transformar la respuesta Y usando una función cóncava como $\log Y$ o $\sqrt{Y}$
- Encoge las respuestas largas, la heteroscedasticidad reduzca.

Valor Atípicos (Outliers)

- Un valor atípico (outlier) es un punto para el cual y_i está alejado del valor predecido $\hat{y}_i$ por el modelo.



Línea de regresión de mínimos cuadrados
Línea de regresión después de eliminar el valor atípico

Efecto del Valor Atípicos

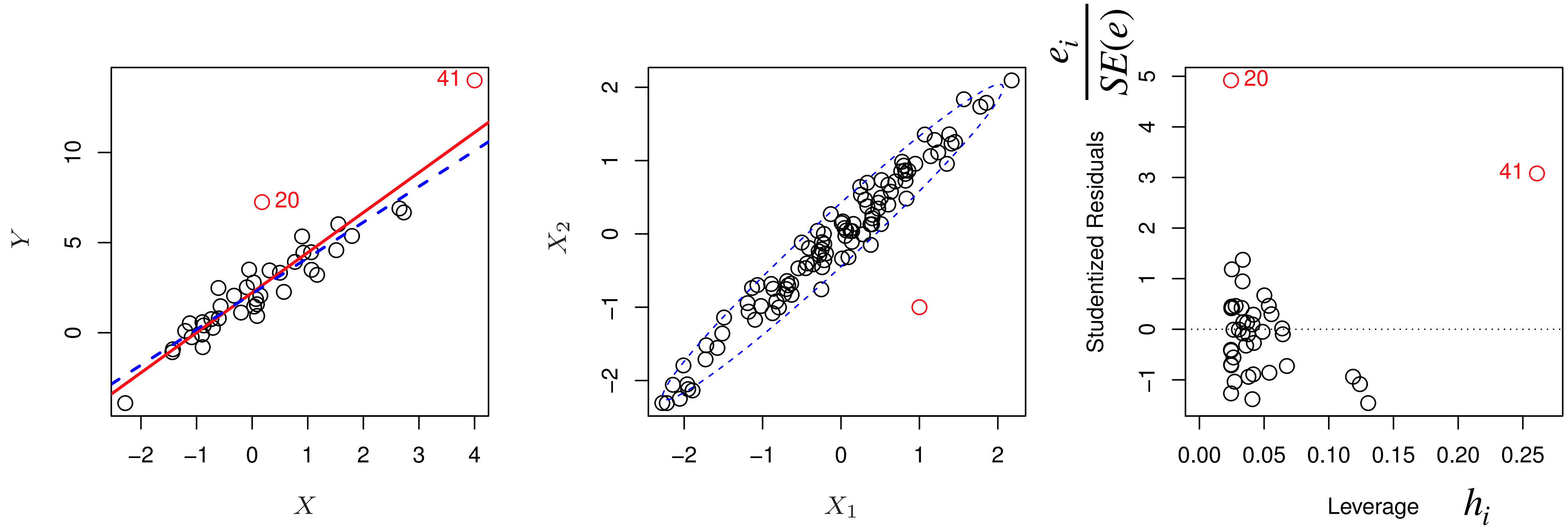
- Línea de regresión
  - Puede cambiar (no mucho impacto en el ejemplo)
- RSE (cambia de 1.09 a 0.77 en el ejemplo)
  - RSE se usa para calcuar *intervalos de confianza* y *p-values* → un solo valor atípico puede tener implicaciones para la interpretación del ajuste
- R^2 declina sustancialmente si se incluye el outlier (0.892 - 0.805)

Graficar Residuales

- Gráficos de residuales pueden usarse para identificar valores atípicos
    - ¿Que tan grande debe ser un residuo para considerarse un atípico?
- Dibujar los *residual estudentizado*, calculados dividiendo cada residuo e_i por su error estándar estimado
    - Observaciones cuyo *residual estudentizado > 3* en valor absoluto son posiblemente puntos atípicos
- Si sospechamos que un *valor atípico* ha ocurrido debido a un error en la recopilación de datos → solución es simplemente remover la observación
- Sin embargo, se debe tener cuidado ya que un *valor atípico* puede indicar una deficiencia con el modelo, tal como la omisión de un predictor

Puntos de Alto Apalancamiento (High Leverage Points)

- Observaciones con alto apalancamiento tienen un valor inusual para x_i



Línea de regresión de mínimos cuadrados
Línea de regresión después de eliminar el valor atípico

Alto Apalancamiento (2)

- Eliminar la observación de alto apalancamiento tiene un impacto mayor en la línea de mínimos cuadrados que eliminar el valor atípico

- Es motivo de preocupación si la línea de mínimos cuadrados se ve muy afectada por sólo un par de observaciones, porque cualquier problema con estos puntos puede invalidar todo el ajuste

- Es importante identificar observaciones de alto apalancamiento

Alto Apalancamiento (3)

- Regresión lineal simple
  - las observaciones de alto apalancamiento son bastante fáciles de identificar
  - buscar observaciones cuyo valor predictivo esté fuera del rango normal de las observaciones
- En una regresión lineal múltiple con muchos predictores,
  - es posible tener una observación que esté dentro del rango de los valores de cada predictor individual, pero que sea inusual en términos del conjunto completo de predictores.

Estadística de Apalancamiento (leverage statistics)

- Problema es más pronunciado en entornos de regresión múltiple con más de dos predictores
  - no existe una forma sencilla de trazar todas las dimensiones de los datos simultáneamente
- Para cuantificar el apalancamiento de una observación: *estadística de apalancamiento* (*leverage statistic*)

$$h_i = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{i'=1}^{n} (x_{i'} - \bar{x})^2} \qquad \frac{1}{n} \leq h_i \leq 1 \qquad \bar{h}_i = \frac{(p+1)}{n}$$

h_i aumenta con la distancia de x_i a $\bar{x}$

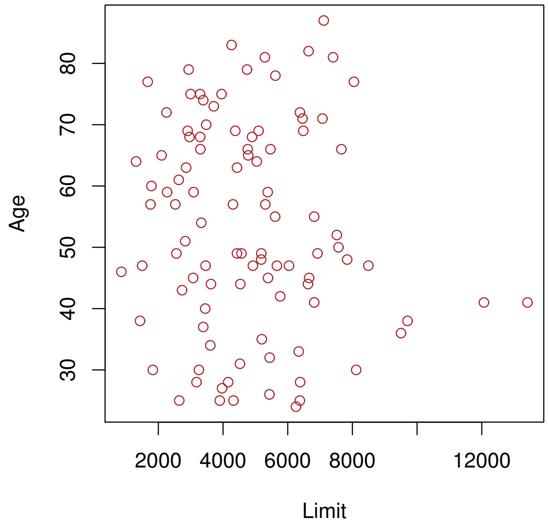
Si $\bar{h}_i \gg \frac{(p+1)}{n}$ el punto tiene un apalancamiento alto

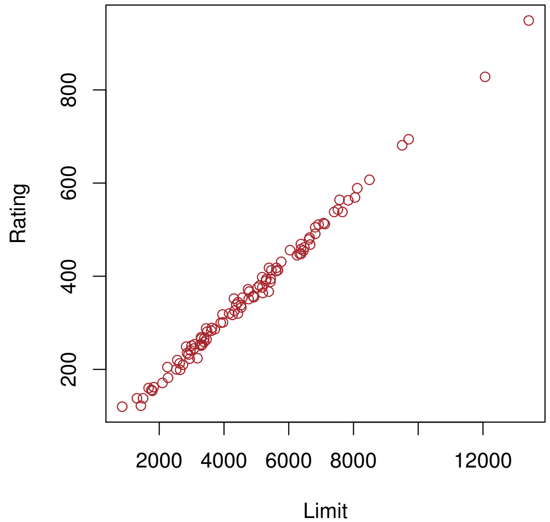
Extensión simple de h_i para múltiples predictores

Colinealidad

- La colinealidad se refiere a la situación en la que dos o más variables predictoras están estrechamente relacionadas entre sí.

ninguna relación obvia





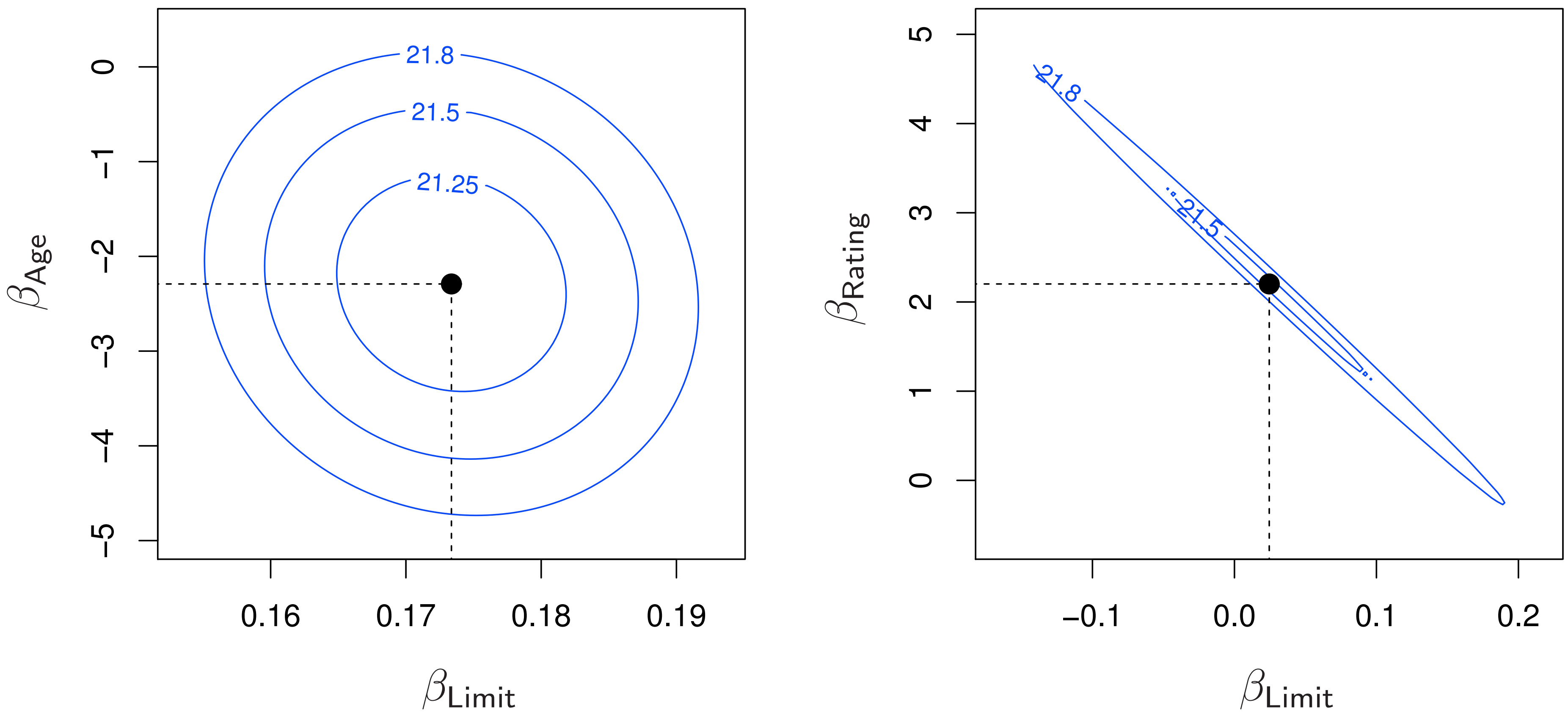
Los predictores están muy correlacionados entre sí

Conjunto de datos de crédito

La presencia de colinealidad puede plantear problemas en el contexto de la regresión → resulta difíci separar los efectos individuales de las variables colineales en la respuesta.

Colinealidad y RSS

$balance = \beta_0 + \beta_{Age}\ age + \beta_{Limit}\ limit$ $\qquad$ $balance = \beta_0 + \beta_{Rating}\ rating + \beta_{Limit}\ limit$



Debido a la colinealidad, existen muchos pares (β_{Limit} , β_{Rating}) con un valor similar para RSS

→ incertidumbre en los coeficientes obtenidos

Efectos en las Estadísticas

		Coefficient	Std. error	t-statistic	p-value
Model 1	`Intercept`	-173.411	43.828	-3.957	< 0.0001
	`age`	-2.292	0.672	-3.407	0.0007
	`limit`	0.173	0.005	34.496	< 0.0001
Model 2	`Intercept`	-377.537	45.254	-8.343	< 0.0001
	`rating`	2.202	0.952	2.312	0.0213
	`limit`	0.025	0.064	0.384	0.7012