Estadística Aplicada III

Regresión lineal múltiple

Jorge de la Vega Góngora

Departamento de Estadística,
Instituto Tecnológico Autónomo de México

Semanas 4 a 6

Introducción

Origen

- El método de regresión lineal fue mencionado por primera vez por Francis Galton entre 1866 y 1899. utilizando datos sobre estaturas de familiares (el tema del artículo era la regresión hacia la mediocridad en la estatura hereditaria. EL término *regresión* se entiende aquí como retroceso, que no es el uso que se le da a la palabra en la mayoría de los contextos.

REGRESSION *towards* MEDIOCRITY *in* HEREDITARY STATURE.
By FRANCIS GALTON, F.R.S., &c.

[WITH PLATES IX AND X.]

THIS memoir contains the data upon which the remarks on the Law of Regression were founded, that I made in my Presidential Address to Section H, at Aberdeen. That address, which will appear in due course in the Journal of the British Association has already

- Galton creó muchos conceptos en Estadística, además de la regresión: correlación, cuartíl y percentíl. Pero también se le critica ser eugenista y favorecer el concepto de raza superior. Curiosamente, era primo de Charles Darwin.
- El análisis de regresión es una de las metodologías estadísticas más usadas actualmente (incluso se la adueña ML y DS...)

Alcance del tema

- El tema de regresión es muy extenso, los temas a cubrir pueden abarcar con mucha facilidad un curso entero. En términos generales, los temas que deberían cubrirse son:
  - Modelado
  - Estimación
  - Inferencia
  - Variaciones a los supuestos
  - Diagnósticos
  - Aplicaciones
- Aquí sólo podremos ver un poco de estos temas. Sin embargo, mi recomendación es que completen un curso completo de regresion como técnica de modelado.

Modelado

Otros temas ligados a RLM I

Consideramos el siguiente escenario:

- y es una variable de respuesta.
- $\mathbf{x} = (X_1, X_2, \ldots, X_p)$ son p *predictores*, relacionados con, o que explican aspectos de la variable de respuesta. Entonces tenemos n puntos en $\mathbb{R}^{p+1}$:

$$(y_i, \mathbf{x}_i) \qquad i = 1, \ldots, n.$$

- Queremos explicar la relación entre la respuesta y los predictores a través de un modelo para la media condicional:

$$E[y|\mathbf{x}] = f(\mathbf{x}|\boldsymbol{\beta})$$

  donde f es una función *conocida* de los predictores y que está parametrizada a traves de $\boldsymbol{\theta}$ de algún modo.
- En el caso de la regresión lineal, asumimos que la dependencia de la función es *lineal en* los parámteros $\boldsymbol{\beta}$.

- Lo anterior es lo mismo que suponer un modelo de la forma:

$$y|\mathbf{x} = f(\mathbf{x}|\boldsymbol{\beta}) + \epsilon$$

donde ϵ es un error $\epsilon \sim \mathcal{N}\left(0, \sigma^2\right)$ y se supone además que los errores son independientes para diferentes observaciones.

Elementos del modelo de Regresión Lineal Múltiple (RLM)

- Los modelos de regresión múltiple son más comunes porque son muy versátiles y nos permiten modelar muchas situaciones.
- En la práctica, muchas veces no se conoce de antemano a la función f.
- A veces, la forma de f es postulada por una teoría científica, y en otras, los investigadores la suponen *a priori*, o incluso se puede estimar de manera *no paramétrica* a partir de diferentes modelos[1] (eg: *loess*, *splines*, *modelos generalizados aditivos*, etc.).
- Una de las actividades más importantes de un estadístico es tratar de encontrar una relación adecuada entre los datos, y el modelo que se postula para explicar la relación.

[1] Estos modelos no paramétricos usualmente se estudian en un curso de estadística no paramétrica

Modelo de regresión lineal múltiple

- El modelo de regresión lineal supone una función f lineal (en los parámetros) tal que la variable de respuesta y se relaciona con predictores $\mathbf{X} = (\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_p)$ a través de la siguiente relación:

$$\mathbf{y}_{n\times 1}|\mathbf{X} = \underset{n\times(p+1)}{\mathbf{X}}\underset{(p+1)\times 1}{\boldsymbol{\beta}} + \underset{n\times 1}{\boldsymbol{\epsilon}}$$

  donde $\boldsymbol{\epsilon} \sim \mathcal{N}_n\left(\mathbf{0}, \sigma^2\mathbf{I}\right)$.

- $\mathbf{x}_0 = \mathbf{1}$ es el término para incluir la constante en el modelo.
- En general, cuando los errores son normales, podemos escribir el modelo como:

$$\mathbf{y}|\mathbf{X} \sim \mathcal{N}_p\left(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}\right)$$

  Usualmente la matriz $\mathbf{X}$ es conocida como *matriz de diseño*.

- El problema consiste en estimar $\boldsymbol{\beta}$ y σ^2 a partir de una muestra de n puntos $(y_i, x_{0i}, x_{1i}, \ldots, x_{pi}) \in \mathbb{R}^{p+1}$.

Tipos de Predictores en RLM

- Una de las grandes ventajas de la RLM es la posibilidad de incorporar aspectos *no lineales* en los predictores a través de columnas de la matriz de diseño $\mathbf{X}$ como *términos*. Ejemplos de términos incluyen los siguientes:
    1. **Ordenada al orígen**: una constante β_0, para indicar una media global de los datos. Usualmente la primera columna de la matriz de diseño es un vector unitario $\mathbf{1}$.
    2. **Predictores**: variables simples numéricas, continuas o discretas.
    3. **Transformaciones de predictores**: por ejemplo, puede incluir una variable x y $\log(x)$
    4. **Polinomios** en alguna de las variables incluyendo sus productos cruzados o *interacciones*: x_1, x_1^2, x_2, x_1x_2, etc.
    5. **Variables dummies y factores** (predictores categóricos nominales). Por ejemplo, una variable `religión`,
    6. **combinaciones lineates de transformaciones de predictores**, como componentes principales o factores: combinación de predictores procesados previamente.

Ejemplo. [1. Consumo de combustibles]

Una forma de tratar de entender cómo se consume el consumo es entender las características de un estado. Las variables a considerar son las siguientes:

- El consumo de combustibles (y), como función de

$$
\begin{aligned}
X_1 &= \text{Ingreso per cápita en el estado} \\
X_2 &= \text{Número de licencias de conductor emitidas} \\
X_3 &= \text{Población mayor a 18 años} \\
X_4 &= \text{Tasa de impuesto al combustibles}
\end{aligned}
$$

- Una primera aproximación supone la siguiente relación:

$$y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \epsilon_i$$

  donde i es el índice para diferentes estados de un país.

Ejemplo. [2. Modelo cuadrático]

Supongamos que $y_i =$ Tiempo del ganador del maratón de la Ciudad de México el año i, y sea $\texttt{temp}_i =$Temperatura del medio ambiente promedio de la prueba en el año i; entonces los datos muestran que un modelo apropiado es una función cuadrática:

$$y_i = \beta_0 + \beta_1 \texttt{temp}_i + \beta_2 \texttt{temp}_i^2 + \epsilon_i$$

para los años disponibles $i = 1, 2, \ldots, n$.

- Un problema interesante es determinar cuál es la temperatura óptima para obtener el mejor tiempo ganador para la prueba.
- En este ejemplo hay sólo *predictor*, $\texttt{temp}$ y tres *términos* en el modelo:1, $\texttt{temp}$ y $\texttt{temp}^2$.

Ejemplo. [3. Modelo polinomial]

En los modelos polinomiales podemos también incorporar el impacto de predictores cruzados.

- Polinomios de grado k:
$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \ldots + \beta_k x^k + \epsilon$$
Hay un sólo predictor y $k+1$ términos.
- Interacciones:
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \epsilon$$
En este modelo hay 2 predictores y 4 términos.

Ejemplo. [4. Modelos de diseño de experimentos (ANOVA)]

Cuando el modelo sólo tiene variables *dummy*, usualmente es un modelo que caracteríza un *análisis de varianza* (ANOVA). Por ejemplo: si $z_{ij} = I[\text{obs } i \in \text{ Población } j]$ es una función indicadora de población, y se tienen 3 poblaciones o tratamientos, se puede escribir

$$y_j = \beta_0 + \beta_1 z_{1j} + \beta_2 z_{2j} + \beta_3 z_{3j} + \epsilon_j$$

Entonces el *efecto* para la población 1 será $\mu + \beta_1$, para la población 2 será $\mu + \beta_2$ y para la población 3 será $\mu + \beta_3$.
Así que ANOVA es un caso particular de la regresión lineal.

Estimación

Estimación I

- Básicamente hay dos formas de estimar un modelo de regresión lineal múltiple:
    1. Utilizando algún procedimiento de optimización matemática a través de alguna restricción, y
    2. Utilizando el método de máxima verosimilitud a partir del supuesto de alguna distribución para la muestra.
- Cuando se utiliza en (1) como criterio de optimización la suma de cuadrados de las desviaciones y en (2) la distribución normal, ambos métodos dan la misma solución.

Estimación II

Solución de Mínimos cuadrados/máxima verosimilitud

El método de mínimos cuadrados busca encontrar el valor de **b** que minimiza la siguiente función (suma de cuadrados residuales):

$$RSS(\mathbf{b}) = \boldsymbol{\epsilon}'\boldsymbol{\epsilon} = (\mathbf{y} - \mathbf{Xb})'(\mathbf{y} - \mathbf{Xb}) = \sum_{i=1}^{n} \epsilon_i^2$$

La solución al problema de optimización está dada por:

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

y

$$s^2 = \hat{\sigma}^2 = \frac{\mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y}}{n - k}$$

donde $k = p + 1$ y $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{H}$ se conoce como la *matriz sombrero*.

Estimación III

- Noten que $\hat{\mathbf{y}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{H}\mathbf{y}$. La matriz sombrero juega un papel muy importante para el diagnóstico del modelo, como veremos más adelante.
- Los residuales estimados se pueden escribir como funciones de la matriz sombrero,

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{H}\mathbf{y} = (\mathbf{I} - \mathbf{H})\mathbf{y}$$

Algunas observaciones relevantes I

- Propiedades de la matriz sombrero
  - $\mathbf{H}$ es idempotente: $\mathbf{H}^2 = \mathbf{H}$
  - Es simétrica: $\mathbf{H}' = \mathbf{H}$
  - También $\mathbf{I} - \mathbf{H}$ es simétrica e idempotente.
- Propiedades del estimador $\hat{\beta}$
  - Como el estimador $\hat{\beta}$ también es el de máxima verosimilitud, hereda sus propiedades (consistencia, eficiencia, normalidad asintótica, etc.).
  - $\hat{\boldsymbol{\beta}}$ es insesgado: $\mathrm{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$.
  - $\mathrm{Var}(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$
  - $\mathrm{cov}\left(\hat{\boldsymbol{\beta}}, e\right) = \mathbf{0}$
  - $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ es una combinación lineal de la variable de respuesta $\mathbf{y}$.
  - Teorema de Gauss-Markov: $\hat{\beta}$ es BLUE, el mejor estimador (mínima varianza) lineal insesgado, sobre la clase de todos los estimadores lineales insesgados.
- Propiedades de los residuales $\hat{\epsilon} = \mathbf{e}$:
  - $\mathrm{E}(\mathbf{e}) = \mathbf{0}$, $\mathrm{Var}(\mathbf{e}) = \sigma^2(\mathbf{I} - \mathbf{H})$
  - Si $s^2 = \frac{\mathbf{y}'(\mathbf{I}-\mathbf{H})\mathbf{y}}{n-k}$, entonces $\mathrm{E}(s^2) = \sigma^2$.

Ejemplo I

- Para una matriz de datos $\mathbf{X}_{5\times 3}$ y un vector de respuestas $\mathbf{y}$ dado,

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 2 & 3 & 4 & 1 & 5 \\ 1 & 2 & 5 & 2 & 6 \end{pmatrix} \begin{pmatrix} 1 & 2 & 1 \\ 1 & 3 & 2 \\ 1 & 4 & 5 \\ 1 & 1 & 2 \\ 1 & 5 & 6 \end{pmatrix} = \begin{pmatrix} 5 & 15 & 16 \\ 15 & 55 & 60 \\ 16 & 60 & 70 \end{pmatrix}$$

-

$$\mathbf{X}'\mathbf{y} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 2 & 3 & 4 & 1 & 5 \\ 1 & 2 & 5 & 2 & 6 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{pmatrix} = \begin{pmatrix} 15 \\ 49 \\ 58 \end{pmatrix}$$

- $$(\mathbf{X}'\mathbf{X})^{-1} = \begin{pmatrix} 1.14 & -0.41 & 0.09 \\ -0.41 & 0.43 & -0.27 \\ 0.09 & -0.27 & 0.23 \end{pmatrix}$$

- $$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \begin{pmatrix} 2.23 \\ -0.74 \\ 1.46 \end{pmatrix}$$

- Calculamos **H** y $\hat{\sigma}^2$:

Ejemplo III

```r
X <- matrix(c(rep(1,5),2,3,4,1,5,1,2,5,2,6),ncol=3)
y <- 1:5
H <- X %*% solve(t(X) %*% X) %*% t(X)
H
```

```
            [,1]        [,2]        [,3]        [,4]        [,5]
[1,]  0.52727273  0.47272727 -0.03636364  0.12727273 -0.09090909
[2,]  0.47272727  0.52727273  0.03636364 -0.12727273  0.09090909
[3,] -0.03636364  0.03636364  0.38181818  0.16363636  0.45454545
[4,]  0.12727273 -0.12727273  0.16363636  0.92727273 -0.09090909
[5,] -0.09090909  0.09090909  0.45454545 -0.09090909  0.63636364
```

```r
H %*% H
```

```
            [,1]        [,2]        [,3]        [,4]        [,5]
[1,]  0.52727273  0.47272727 -0.03636364  0.12727273 -0.09090909
[2,]  0.47272727  0.52727273  0.03636364 -0.12727273  0.09090909
[3,] -0.03636364  0.03636364  0.38181818  0.16363636  0.45454545
[4,]  0.12727273 -0.12727273  0.16363636  0.92727273 -0.09090909
[5,] -0.09090909  0.09090909  0.45454545 -0.09090909  0.63636364
```

```r
sigma2hat <- t(y) %*% (diag(1,5)- H) %*% y
sigma2hat
```

```
         [,1]
[1,] 2.254545
```

Más dimensiones, nuevos problemas

- ¿Cómo podemos “ver” si el modelo que estamos ajustando es adecuado?
- ¿Cómo vemos si los residuales son grandes?
- ¿Cuál es la distribución conjunta del vector $\hat{\beta}$?

Ver los datos

- Una primera aproximación para “ver” los datos es una matriz de gráficas de dispersión
  - Se pueden graficar las *respuestas parciales* $\{y_i, x_{ji}\}$
  - Se pueden ver las relaciones entre los predictores en pares $\{x_{mi}, x_{ji}\}$, $m \neq j$.
- Interpretar con cuidado: si cada $\{y_i, x_{ji}\}$ muestra lineas, *no quiere decir* que la relación de y con todas las $x's$ sea lineal.
- Las gráficas de los predictores ayudan a identificar variables redundantes.

Ejemplos

Ejemplos

- Consideraremos ahora algunos ejemplos prácticos en `R`; veremos cómo hacer pruebas de hipótesis para combinaciones lineales y veremos una interpretación diferente del coeficiente de determinación.
- Muchos de los ejemplos considerados aquí toman los datos del paquete `alr4` (creado para el libro: *Applied Linear Regression 4ed.* de Sandford Weisberg (mi asesor principal en el doctorado).

E1: costos de transacción

- Consideramos un problema de costos de transacción en un banco australiano.
- Los datos se llaman `Transact`. Las transacciones pueden ser de dos tipos. Hay 261 sucursales de un banco en las que se midieron las siguientes variables: `time` = total de minutos gastados en transacciones, `T1` = número de transacciones del tipo 1 y `T2` = número de transacciones de tipo 2. Los datos son del año 1985.
- El objetivo es explicar `time` como función de `T1` y `T2`. El costo es un múltiplo del tiempo ocupado en hacer transacciones.
- Suponemos además que todas las transacciones son independientes.

E1: postulando un modelo

- Si β_i denota el tiempo promedio ocupado en hacer una transacción de tipo i, entonces se espera que el tiempo total es

$$\mathsf{E}[\mathtt{Time}|\mathtt{T1},\mathtt{T2}] = \beta_0 + \beta_1\mathtt{T1} + \beta_2\mathtt{T2}$$

- β_0 representa un costo fijo de hacer transacciones en cualquier sucursal.
- En este modelo hay 2 predictores y 3 términos.
- Suponemos que la varianza $\mathsf{Var}(\mathtt{time}|\mathtt{T1},\mathtt{T2})$ es constante.
- ¿Cuáles son las unidades de los respectivos coeficientes?

E1: análisis de datos

- ¿Cómo hay que proceder para comenzar el análisis? Primero considerando cada variable en forma individual, después en forma conjunta.
  1. Podemos ganar información obteniendo estadísticas individuales de las variables, haciendo histogramas, boxplots, verificando normalidad, etc.
  2. Graficar los datos vía una matriz de gráficas de dispersión, si es posible, en una gráfica 3D.
  3. Si un modelo no ha sido propuesto, hay que proponer un modelo.
  4. Ajustar el modelo propuesto.
  5. Analizar y verificar el ajuste del modelo. Si el modelo no representa adecuadamente a los datos, se requiere buscar transformaciones de los predictores y/o de la respuesta, cambiar el modelo y volver a iterar.
  6. Una vez que se llega a un modelo adecuado, usarlo para los fines que sean apropiados: descripción, predicción, optimización, etc.

E1: Estadísticas sumarias

¿Qué se puede decir en términos generales de cada variable?

```
summary(Transact)
```

```
       t1               t2            time
 Min.   :   0.0   Min.   : 148   Min.   :  487
 1st Qu.:  85.0   1st Qu.:1516   1st Qu.: 3618
 Median : 214.0   Median :2192   Median : 5583
 Mean   : 281.2   Mean   :2422   Mean   : 6607
 3rd Qu.: 437.0   3rd Qu.:3175   3rd Qu.: 8712
 Max.   :1450.0   Max.   :5791   Max.   :20741
```

```
apply(Transact,2,sd)
```

```
       t1        t2      time
 257.0844 1180.7314 3774.0476
```

```
cor(Transact)
```

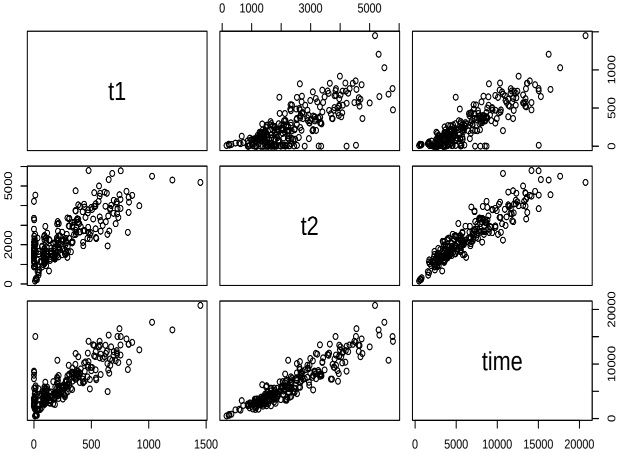
```
            t1        t2      time
t1   1.0000000 0.7715669 0.8631874
t2   0.7715669 1.0000000 0.9235965
time 0.8631874 0.9235965 1.0000000
```

Cuando los rangos de las variables son muy amplios, una transformación logarítmica podría ser apropiada.

E1: scatterplot

¿Qué se puede decir al obsevar la siguiente gráfica?

```r
pairs(Transact)
```



E1: Ajustando un modelo lineal I

```
m1 <- lm(time ~ t1 + t2, data=Transact)
summary(m1)
```

```
Call:
lm(formula = time ~ t1 + t2, data = Transact)

Residuals:
    Min      1Q  Median      3Q     Max 
-4652.4  -601.3     2.4   455.7  5607.4 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 144.36944  170.54410   0.847    0.398    
t1            5.46206    0.43327  12.607   <2e-16 ***
t2            2.03455    0.09434  21.567   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1143 on 258 degrees of freedom
Multiple R-squared:  0.9091,Adjusted R-squared:  0.9083 
F-statistic:  1289 on 2 and 258 DF,  p-value: < 2.2e-16
```

E1: interpretación de los resultados

- De acuerdo al ejemplo: $\hat{\boldsymbol{\beta}} = \begin{pmatrix} 144.369 \\ 5.46206 \\ 2.03455 \end{pmatrix}$
- La ecuación del modelo es $\hat{\texttt{Time}} = 144.4 + 5.5\texttt{T1} + 2\texttt{T2}$.
- Hay $n-k =$ 261-3 =258 grados de libertad.
- Para calcular $\hat{\sigma}^2 = \frac{RSS}{n-k}$, tomamos el `Residual standard error`: $1143 = \sqrt{\frac{RSS}{n-k}}$ y se calcula: $\hat{\sigma}^2 = 1143^2 = 1.306449 \times 10^6$.
- Noten que la desviación estándar del tiempo dados `T1` y `T2` (1143) es *menor* que la desviación estándar del tiempo ignorando la información provista por los tipos de las transacciones (3774).

Inferencia

- Bajo el supuesto de normalidad de los errores, como $\hat{\beta}$ es una combinación lineal de las variables de respuesta:
$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$
se sigue que, como $\mathrm{E}(\mathbf{y}) = \mathbf{X}\beta$ y $\mathrm{Var}(\mathbf{y}) = \sigma^2\mathbf{I}$, entonces
$$\mathrm{E}(\hat{\beta}) = \mathrm{E}((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathrm{E}(\mathbf{y}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\beta = \beta$$
y
$$\mathrm{Var}(\hat{\beta}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathrm{Var}(\mathbf{y})((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')' = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{I}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$
Por lo tanto:
$$\hat{\beta} \sim \mathcal{N}_k\left(\beta, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\right)$$
recordando que $k = p + 1$, y además:
$$(n-k)\frac{\hat{\sigma}^2}{\sigma^2} = \frac{\mathbf{e}'\mathbf{e}}{\sigma^2} = \frac{\mathbf{y}'(\mathbf{I}-\mathbf{H})\mathbf{y}}{\sigma^2} \sim \chi^2_{n-k}$$

Inferencia en regresión II

- Entonces podemos aplicar lo que vimos para la normal multivariada en relación a intervalos de confianza y pruebas de hipótesis. El siguiente resultado resume lo más importante:

Regiones de confianza para $\hat{\beta}$

A. Una región de $100(1-\alpha)$ % de confianza para $\boldsymbol{\beta}$ está dada por:

$$(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'\mathbf{X}'\mathbf{X}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \leq ks^2 F_{k,n-k,\alpha}$$

B. Los intervalos simultáneos de $100(1-\alpha)$ % para β_i están dados por:

$$\hat{\beta}_i \pm \sqrt{\widehat{\mathrm{Var}}(\hat{\beta}_i)}\sqrt{kF_{k,n-k,\alpha}} \quad i = 0, 1, \ldots, k$$

C. Los intervalos marginales para cada β_i están dados por:

$$\hat{\beta}_i \pm t_{n-k,\alpha/2}\sqrt{\widehat{\mathrm{Var}}(\hat{\beta}_i)}$$

E1: regiones de confianza

- Continuando con el ejemplo de transacciones, podemos hallar los intervalos.
- La matriz de varianzas y covarianzas de $\hat{\beta}$ es de la forma $\mathbf{V} = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$. Esta matriz se puede obtener usando la función `vcov` para el modelo ajustado:

```
vcov(m1)

             (Intercept)          t1           t2
(Intercept) 29085.29123 23.58169479 -12.683293995
t1             23.58169  0.18772109  -0.031536343
t2            -12.68329 -0.03153634   0.008899435
```

- Los intervalos marginales al 95 % para cada β_i están dados por:

```
int_marginales <- matrix(numeric(),nrow=3,ncol=2)
for(i in 1:3) int_marginales[i,]<- m1$coef[i] + c(-1,1)*qt(0.025,258,lower.tail=F)*sqrt(diag(vcov(m1))[i])
a <- cbind(m1$coef,int_marginales); colnames(a) <- c("betahat", "lim_inf", "lim_sup"); a

               betahat     lim_inf    lim_sup
(Intercept) 144.369443 -191.466242 480.205128
t1            5.462057    4.608865   6.315248
t2            2.034549    1.848781   2.220317
```

- Los intervalos simultáneos al 95 % para β_i:

```
int_simultaneos <- matrix(numeric(),nrow=3,ncol=2)
for(i in 1:3) int_simultaneos[i,]<- m1$coef[i] + c(-1,1)*sqrt(3*qf(0.05,3,258,lower.tail=F))*sqrt(diag(vcov(m1))[i])
a <- cbind(m1$coef,int_simultaneos); colnames(a) <- c("betahat","lim_inf","lim_sup"); a

               betahat     lim_inf    lim_sup
(Intercept) 144.369443 -335.546905 624.285790
t1            5.462057    4.242827   6.681286
t2            2.034549    1.769082   2.300015
```

Combinaciones lineales en general

- Algunos problemas de regresión requieren calcular intervalos de confianza para combinaciones lineales de los coeficientes del modelo (como en el caso de las transacciones).
- Una combinación lineal es de la forma $L = \mathbf{a}'\hat{\boldsymbol{\beta}} = \sum_{i=1}^{k-1} a_j \hat{\beta}_j$. La combinación lineal L es una variable aleatoria.
- La esperanza de L se obtiene:

$$\mathsf{E}(L) = \mathsf{E}(\mathbf{a}'\hat{\boldsymbol{\beta}}) = \mathbf{a}'\mathsf{E}(\hat{\boldsymbol{\beta}}) = \mathbf{a}'\boldsymbol{\beta}$$

- La varianza de L requiere conocer las covarianzas de los estimadores, y ya sabemos que las podemos obtener de la matriz $\mathbf{V} = \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$:

$$\mathsf{Var}(L) = \mathbf{a}'\mathbf{V}\mathbf{a}$$

- La matriz $\mathbf{V} = \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$ la podemos obtener directamente de R, ya vimos cómo.

Descomposición de suma de cuadrados

- En regresión se cumplen las siguientes restricciones:
  1. $\hat{\mathbf{y}}'\boldsymbol{\epsilon} = 0$
  2. De la condición anterior,
  $$\mathbf{y}'\mathbf{y} = (\hat{\mathbf{y}} + \mathbf{y} - \hat{\mathbf{y}})'(\hat{\mathbf{y}} + \mathbf{y} - \hat{\mathbf{y}}) = (\hat{\mathbf{y}} + \mathbf{e})'(\hat{\mathbf{y}} + \mathbf{e}) = \hat{\mathbf{y}}'\hat{\mathbf{y}} + \mathbf{e}'\mathbf{e}$$
  3. Debido a que la primera columna de **X** es **1**, la condición $\mathbf{X}'\mathbf{e} = 0$ incluye el requerimiento de que:
  $$0 = \mathbf{1}'\mathbf{e} = \sum_{j=1}^{n} e_j = \sum_{j=1}^{n} (y_j - \hat{y}_j),$$
  por lo que $\bar{y} = \bar{\hat{y}}$. Si restamos de ambos lados de la descomposición en (2), se obtiene:
  $$\mathbf{y}'\mathbf{y} - n\bar{y}^2 = \hat{\mathbf{y}}'\hat{\mathbf{y}} - n(\bar{\hat{y}})^2 + \mathbf{e}'\mathbf{e}$$
  o bien
  $$\sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} (\hat{y}_j - \bar{y})^2 + \sum_{i=1}^{n} e_j^2$$
  En palabras:
  $$\left(\begin{array}{c}\text{Suma de cuadrados}\\ \text{total alrededor}\\ \text{de la media}\end{array}\right) = \left(\begin{array}{c}\text{Suma de cuadrados}\\ \text{debida a la regresión}\end{array}\right) + \left(\begin{array}{c}\text{Suma de cuadrados}\\ \text{residual}\end{array}\right)$$

Cálculo de R^2 y R^2_{aj} I

- Si denotamos $S_{yy} = \mathbf{y}'\mathbf{y} - n\bar{y}^2$, $SS_{reg} = \hat{\mathbf{y}}'\hat{\mathbf{y}} - n(\bar{\hat{y}})^2$ y $RSS = \mathbf{e}'\mathbf{e}$, entonces definimos el *coeficiente de determinación* como la proporción de la variación total de la respuesta explicada por los predictores:

$$R^2 = \frac{S_{yy} - RSS}{S_{yy}} = \frac{SS_{reg}}{S_{yy}}$$

  Recuerden que la suma de cuadrados del error del modelo con sólo una constante:

$$E(y|\mathbf{x}) = \alpha_0$$

  es S_{yy}. Por esto, cuando el modelo no tiene ordenada al orígen, no tiene sentido calcular R^2.
- La ecuación se interpreta como la fracción de variabilidad en la respuesta que se explica por agregar términos en el modelo.
- Recuerden también que R^2 es el cuadrado del coeficiente de correlación entre y y $\hat{y}$, así que la R^2 se puede ver también como una comparación de modelos.

Cálculo de R^2 y R^2_{aj} II

- Es importante notar que R^2 siempre crece si se agregan más y más términos al modelo, aunque no tengan nada que ver con el problema.
- Para corregir este problema se usa una R^2 "ajustada" por el número de predictores:

$$R^2_{aj} = 1 - \frac{n-1}{n-k}(1 - R^2)$$

- Este coeficiente ajustado puede hacerse más pequeño cuando se introducen más términos en el modelo, ya que $n - k$ puede anular el incremento de R^2.
- En nuestro ejemplo de transacciones:
- $R^2_{aj} = 1 - \frac{260}{258}(1 - 0.909053) = 0.908348$.
- Usen R^2_{aj} cuando tengan muchos predictores y/o términos o estén comparando muchos modelos.

ANOVA y comparación de modelos en RLM I

- El análisis de varianza (ANOVA) es una forma de agrupar información para comparar modelos.
- El caso más inmediato es la tabla de ANOVA que se obtiene de un modelo de regresión lineal múltiple que mide la utilidad del modelo y que prueba la significancia de todos los parámetros simultáneamente:

$$H_0 : E(y|\mathbf{x}) = \alpha_0 \quad \text{vs.} \quad H_a : E(y|\mathbf{x}) = \alpha_0 + \boldsymbol{\alpha}'\mathbf{x}$$

- La hipótesis anterior es equivalente a suponer que *simultáneamente*, todos los coeficientes $\alpha_1, \alpha_2, \ldots, \alpha_p$ son 0, es decir, y es independiente de *todos* los predictores.
- A esta prueba se le conoce como *prueba de independencia*, *prueba de significancia* de la regresión o *prueba de utilidad del modelo*.
- El último renglón en la salida de regresión realiza esta prueba. La prueba de F=1289, que es el cuantíl de una distribución F con 2 y 258 grados de libertad. El p-value es prácticamente 0, por lo que se concluye que el modelo lineal es útil, o significativo.

- En general, es común querer probar que algunos coeficientes son cero, es decir, la hipótesis nula de que el modelo es un *subconjunto* del modelo *completo*:

$$H_0 : \mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\epsilon} \quad \text{vs.} \quad H_a : \mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}$$

o bien:

$$H_0 : \text{Modelo chico} \quad \text{vs.} \quad H_a : \text{Modelo grande}$$

Entonces la estadística de prueba general de esta hipótesis es de la forma:

$$F = \frac{(RSS_{H_0} - RSS_{H_a})/(gl_{H_0} - gl_{H_a})}{\hat{\sigma}^2_{H_a}} \sim F_{(gl_{H_0} - gl_{H_a}), gl_{H_a}}$$

Ejemplo 2: Comparando Salarios I

- Los datos en `salary` del paquete `alr4` fueron obtenidos para probar en una corte que existía discriminación salarial por sexo entre los profesores de una universidad americana. La respuesta es `Salary` (`Sal`) y hay tres predictores: `Rank` con tres niveles o categorías y `Sex` que tiene dos niveles, y `Ysdeg` (`Ys`) que son los años desde que un profesor se graduó.

```
#library(alr4) #cargar si no lo tienen cargado
data(salary)
str(salary) #noten que algunas variables ya son factores
```

```
'data.frame': 52 obs. of  6 variables:
 $ degree: Factor w/ 2 levels "Masters","PhD": 1 1 1 1 2 1 2 1 2 2 ...
 $ rank  : Factor w/ 3 levels "Asst","Assoc",..: 3 3 3 3 3 3 3 3 3 3 ...
 $ sex   : Factor w/ 2 levels "Male","Female": 1 1 1 2 1 1 2 1 1 1 ...
 $ year  : int  25 13 10 7 19 16 0 16 13 13 ...
 $ ysdeg : int  35 22 23 27 30 21 32 18 30 31 ...
 $ salary: int  36350 35350 28200 26775 33696 28516 24900 31909 31850 32850 ...
```

Ejemplo 2: Comparando Salarios II

- Notamos que `degree`, `rank` y `sex` ya son factores. Por ejemplo, la variable `rank` se ve así:

```
salary$rank
```

```
 [1] Prof  Prof  Prof  Prof  Prof  Prof  Prof  Prof  Prof  Prof  Prof  Assoc
[13] Prof  Assoc Prof  Prof  Prof  Assoc Assoc Prof  Asst  Assoc Prof  Prof
[25] Assoc Prof  Assoc Prof  Assoc Assoc Asst  Assoc Asst  Assoc Assoc Assoc
[37] Asst  Asst  Asst  Asst  Asst  Asst  Assoc Asst  Asst  Asst  Asst  Asst
[49] Asst  Asst  Asst  Asst
Levels: Asst Assoc Prof
```

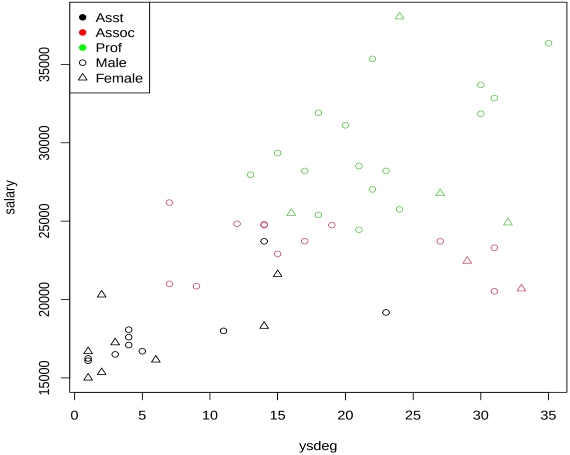
- El objetivo del estudio es entender cuál es la dependencia del salario al rango académico y a la experiencia, medida por los años desde la graduación. Consideren primero la regresión de `salary` respecto a `rank` y `ysdeg` o en nuestra notación,

$$\texttt{salary}|(\texttt{rank}, \texttt{sex}, \texttt{ysdeg}).$$

  Una gráfica de estos datos se muestra a continuación. Noten que los factore `rank` y `sex` se agregan como variables para marcar, con color y símbolo respectivamente.

```
with(salary, plot(ysdeg, salary, col = rank, pch = as.numeric(sex)))
with(salary,legend("topleft",legend = c(levels(rank),levels(sex)),
     col=c("black","red","green","black","black"),pch=c(19,19,19,1,2)))
```

Ejemplo 2: Comparando Salarios III



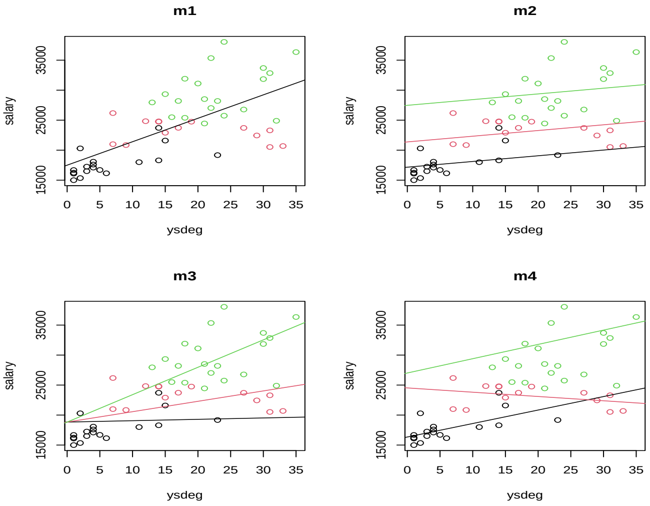
Ejemplo 2: Comparando Salarios IV

- Sin considerar el sexo, hay cuatro casos a considerar:
    1. un modelo de regresión lineal simple para todos los rangos (básicamente ignorando el rango),
    2. líneas paralelas (misma pendiente, diferentes ordenadas),
    3. líneas con una ordenada al origen común (pendientes diferentes, misma ordenada al origen), y
    4. líneas diferentes para cada rango (una línea por categoría).
- Los diferentes modelos se pueden visualizar

Ejemplo 2: Comparando Salarios V

```
m1 <- lm(salary ~ ysdeg,data=salary)
m2 <- lm(salary ~ ysdeg + rank, data=salary)
m3 <- lm(salary ~ ysdeg + ysdeg:rank, data=salary)
m4 <- lm(salary ~ ysdeg*rank, data=salary)
layout(matrix(1:4,nrow=2,byrow=T))
with(salary,plot(ysdeg,salary,col=rank,main="m1"))
        abline(m1)
with(salary,plot(ysdeg,salary,col=rank,main="m2"))
        abline(a=m2$coef[1],b=m2$coef[2])
        abline(a=m2$coef[1]+m2$coef[3],b=m2$coef[2],col=2)
        abline(a=m2$coef[1]+m2$coef[4],b=m2$coef[2],col=3)
with(salary,plot(ysdeg,salary,col=rank,main="m3"))
        abline(a=m3$coef[1], b = m3$coef["ysdeg"])
        abline(a=m3$coef[1], b = m3$coef["ysdeg"]+m3$coef["ysdeg:rankAssoc"],col=2)
        abline(a=m3$coef[1], b = m3$coef["ysdeg"]+m3$coef["ysdeg:rankProf"],col=3)
with(salary,plot(ysdeg,salary,col=rank,main="m4"))
        abline(a=m4$coef[1], b = m4$coef["ysdeg"])
        abline(a=m4$coef[1] + m4$coef["rankAssoc"], b = m4$coef["ysdeg"]+m4$coef["ysdeg:rankAssoc"],col=2)
        abline(a=m3$coef[1] + m4$coef["rankAssoc"], b = m4$coef["ysdeg"]+m4$coef["ysdeg:rankProf"],col=3)
```

Ejemplo 2: Comparando Salarios VI



Comparando salarios I

El modelo más general es el caso 4: 1,2,3 son submodelos del caso 4. El caso 1 es un submodelo de los casos 2 y 3, y los casos 2 y 3 no están relacionados.

```
summary(m1)


Call:
lm(formula = salary ~ ysdeg, data = salary)

Residuals:
    Min      1Q  Median      3Q     Max 
-9703.5 -2319.5  -437.1  2631.8 11167.3 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 17502.26    1149.70  15.223  < 2e-16 ***
ysdeg         390.65      60.41   6.466  4.1e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4410 on 50 degrees of freedom
Multiple R-squared:  0.4554,Adjusted R-squared:  0.4445 
F-statistic: 41.82 on 1 and 50 DF,  p-value: 4.102e-08
```

```
summary(m2)


Call:
lm(formula = salary ~ ysdeg + rank, data = salary)

Residuals:
    Min      1Q  Median      3Q     Max 
-5619.5 -1494.5  -341.6  1810.3  8286.2 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 17166.46     785.40  21.857  < 2e-16 ***
ysdeg          95.08      58.15   1.635  0.10855    
rankAssoc    4209.65    1279.20   3.291  0.00188 ** 
rankProf    10310.30    1359.39   7.585  9.4e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2943 on 48 degrees of freedom
Multiple R-squared:  0.7672,Adjusted R-squared:  0.7526 
F-statistic: 52.72 on 3 and 48 DF,  p-value: 3.188e-15
```

Comparando salarios II

```r
summary(m3)
```

```
Call:
lm(formula = salary ~ ysdeg + ysdeg:rank, data = salary)

Residuals:
    Min      1Q  Median      3Q     Max
-8547.9 -2190.6  -662.8  2066.6  8250.5

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     18834.25     936.89  20.103  < 2e-16 ***
ysdeg              22.91     116.94   0.196    0.845
ysdeg:rankAssoc   150.19     104.80   1.433    0.158
ysdeg:rankProf    433.76     101.58   4.270 9.17e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3334 on 48 degrees of freedom
Multiple R-squared:  0.7013,Adjusted R-squared:  0.6826
F-statistic: 37.56 on 3 and 48 DF,  p-value: 1.21e-12
```

```r
summary(m4)
```

```
Call:
lm(formula = salary ~ ysdeg * rank, data = salary)

Residuals:
    Min      1Q  Median      3Q     Max
-6915.3 -1497.9   -24.5  1272.7  8135.9

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     16345.62     942.08  17.351  < 2e-16 ***
ysdeg             224.69     106.59   2.108 0.040516 *
rankAssoc        8179.94    1963.07   4.167 0.000135 ***
rankProf         7845.14    2634.04   2.978 0.004613 **
ysdeg:rankAssoc  -295.99     134.50  -2.201 0.032822 *
ysdeg:rankProf     13.57     148.70   0.091 0.927665
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2788 on 46 degrees of freedom
Multiple R-squared:  0.7998,Adjusted R-squared:  0.778
F-statistic: 36.75 on 5 and 46 DF,  p-value: 5.6e-15
```

Comparando submodelos I

- El enfoque general para comparar submodelos es usar la prueba de F que resulta de dividir las sumas de cuadrados de los errores o residuales de un modelo chico y un modelo grande. Supongan que la hipótesis nula es un submodelo del de la hipótesis alternativa en el sentido de que cada término que aparece en la nula también aparece en la alternativa (el modelo chico es la nula y el grande la alternativa). Por ejemplo, para comparar los modelos 2 y 4, se establecen las hipótesis del siguiente modo:

$$
\begin{aligned}
NH: &\quad \texttt{salary} = \beta_0 + \beta_1 \texttt{ysdeg} + \beta_2 \texttt{rank} \\
AH: &\quad \texttt{salary} = \beta_0 + \beta_1 \texttt{ysdeg} + \beta_2 \texttt{rank} + \beta_3 \texttt{ysdeg}:\texttt{rank}
\end{aligned}
$$

- Esto es equivalente a la hipótesis de que $NH: \beta_3 = 0$. Calculamos la estadística de prueba F como

$$
F = \frac{\frac{RSS_{NH} - RSS_{AH}}{gl_{NH} - gl_{AH}}}{\hat{\sigma}^2_{AH}}
$$

Comparando submodelos II

- Hay que recordar que $RSS_{AH}/gl_{AH} = \hat{\sigma}^2_{AH}$. Esta estadística tiene distribución F con $(gl_{NH} - gl_{AH})$ en el numerador y gl_{AH} en el denominador.
- Se pueden obtener estos números usando `anova` para los modelos 2 y 4:

```
anova(m2,m4)

Analysis of Variance Table

Model 1: salary ~ ysdeg + rank
Model 2: salary ~ ysdeg * rank
  Res.Df       RSS Df Sum of Sq      F Pr(>F)
1     48 415783967
2     46 357499755  2  58284211 3.7498  0.031 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Entonces: $RSS_{NH} = 415,783,967$, $RSS_{AH} = 357,499,755$, $gl_{NH} = 48$, $gl_{AH} = 46$, y $\hat{\sigma}^2_{AH} = 357,499,755/46 = 7,771,734$. De este modo

$$F = \frac{(415,783,967 - 357,499,755)/(48-46)}{7,771,734} = \frac{58,284,211/2}{7,771,734} = 3.7498$$

con 2 y 46 grados de libertad. Esto da un $pvalue = 0.031$, lo que dice que el modelo general (modelo 4) explica mejor que el modelo 2 que no tiene el término de interación.

- Podemos hacer todas las comparaciones necesarias de estos modelos de la misma forma que mostramos aquí, siempre y cuando estén *anidados*, es decir, que $H_0 \subseteq H_a$. Por ejemplo, comparando modelo 1 contra el modelo 2, que es lo mismo que probar la hipótesis:

$$H_0 : \texttt{salary} = \beta_0 + \beta_1 \texttt{ysdeg} \text{ vs. } H_a : \texttt{salary} = \beta_0 + \beta_1 \texttt{ysdeg} + \beta_{21}[\texttt{r}]\texttt{Assoc} + \beta_{22}[\texttt{r}]\texttt{Prof}$$

```
anova(m1,m2)
```

```
Analysis of Variance Table

Model 1: salary ~ ysdeg
Model 2: salary ~ ysdeg + rank
  Res.Df       RSS Df Sum of Sq      F    Pr(>F)
1     50 972458240
2     48 415783967  2 556674273 32.133 1.393e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Predicción

Predicción I

- En predicción, se pueden considerar dos problemas:
  1. Estimar la función de regresión $\widehat{\mathrm{E}(y|\mathbf{x}_0)}$ en el vector de predictores $\mathbf{x}_0$.
  2. Estimar la respuesta $y|\mathbf{x}_0$ en $\mathbf{x}_0$
- Para el caso 1, dado $\mathbf{x}_h = (1, x_{h1}, \ldots, x_{hp})$, el valor ajustado es $\hat{y}_h = \mathbf{x}_h'\hat{\boldsymbol{\beta}}$. Un intervalo de 100(1-$\alpha$) % de confianza para $\mathrm{E}(y|\mathbf{x}_0) = \mathbf{x}_0'\boldsymbol{\beta}$ está dado por:

$$\mathbf{x}_0\hat{\boldsymbol{\beta}} \pm t_{(n-k),\alpha/2}\sqrt{s^2(\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0)}$$

- Para el caso 2, hay que considerar las dos fuentes de variación (lo que hace la estimación más incierta): la que corresponde a la estimación de la función de regresión, y la que proviene de estimar el error ϵ_0 correspondiente. Entonces un intervalo de confianza de 100(1-α) % para $y|\mathbf{x}_0$ está dado por

$$\mathbf{x}_0\hat{\boldsymbol{\beta}} \pm t_{(n-k),\alpha/2}\sqrt{s^2(1 + \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0)}$$

Predicción: ejemplo I

- Supongan que se desea predecir el salario de una profesora con rango asociada, y que tiene dos años de graduada, así como el de un profesor asistente con 10 años de graduado Consideremos el modelo más general que considera el sexo también:

```
(m5 <- lm(salary ~ ysdeg*rank*sex, data=salary)) #modelo completo: lineas ind para cada sexo y rango

Call:
lm(formula = salary ~ ysdeg * rank * sex, data = salary)

Coefficients:
            (Intercept)                    ysdeg
               16439.21                   211.48
              rankAssoc                 rankProf
                7708.66                  5926.82
              sexFemale          ysdeg:rankAssoc
                -231.20                  -253.12
         ysdeg:rankProf          ysdeg:sexFemale
                 122.13                    37.97
      rankAssoc:sexFemale       rankProf:sexFemale
               11293.33                  8670.37
ysdeg:rankAssoc:sexFemale  ysdeg:rankProf:sexFemale
                -436.34                  -452.40
```

Predicción: ejemplo II

- ¿Cómo podemos hacer una predicción de este modelo? Basta con especificar la lista de valores de las variables predictivas para el nuevo caso:

```
newx <- data.frame(rank = c("Assoc","Asst"), sex = c("Female","Male"), ysdeg = c(2,10))
predict(m5, newx, interval = "confidence", se.fit = T)
```

```
$fit
       fit       lwr      upr
1 34330.00 -25062.93 93722.93
2 18554.05  16548.37 20559.74

$se.fit
         1          2
29386.7971   992.3857

$df
[1] 40

$residual.scale
[1] 2859.36
```

Diagnósticos

Propiedades de los estimadores de RLM I

¿Cuáles son los supuestos que hicimos para los modelos de regresión hasta ahora? Estos son los supuestos que tenemos que verificar una vez que tenemos un modelo ajustado.

- Los errores tienen media 0: $\mathrm{E}(\epsilon) = 0$,
- los errores tienen varianza constante: $\mathrm{Var}(\epsilon) = \sigma^2$,
- los errores siguen una distribución normal: $\epsilon_i \sim \mathcal{N}\left(0, \sigma^2\right)$
- Los predictores son independientes de los errores
- Los predictores no son colineales (es decir, podemos invertir la matriz $\mathbf{X}'\mathbf{X}$).

Como vemos, los principales supuestos están asociados a los errores. Adicional a los supuestos anteriores, debemos considerar los siguientes temas importantes:

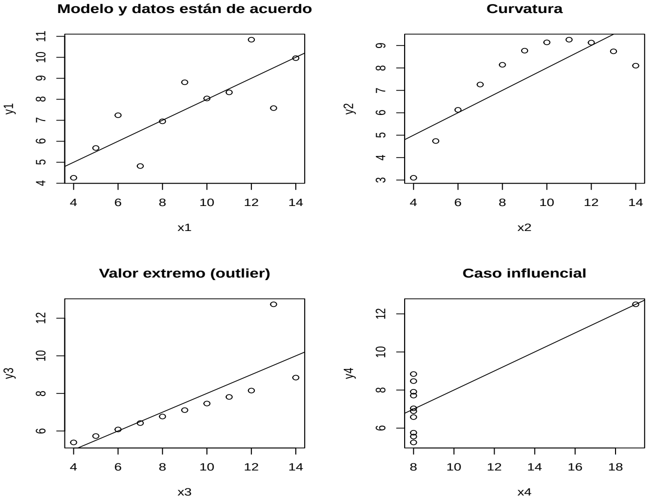
- ¿Cuál es la importancia relativa de cada observación en la estimación de los parámetros?

Los *métodos de diagnóstico* se usan para ayudar a decidir si tenemos información que contradiga los supuestos del modelo.

Diagnósticos I

- La necesidad de los diagnósticos se hace aparente considerando un ejemplo como el siguiente.
- En las siguientes gráficas, cada conjunto de datos consiste de 11 pares de puntos y cada uno produce los mismos coeficientes de regresión, dando $\hat{\beta}_0 = 3.0$, $\hat{\beta}_1 = 0.5$, $\hat{\sigma}^2 = 1.53$ y $R^2 = 0.667$.
- Este conjunto de gráficas se conoce como el *cuarteto de Anscombe*.

Diagnósticos II



Diagnósticos III

- ¿Podemos decir que el modelo lineal es apropiado para cada conjunto de datos?

Posibles fallas en un modelo de regresión I

Los problemas típicos que se pueden presentar en los modelos de regresión son los siguientes:

- Varianza no constante (Heteroscedasticidad).
- Errores correlacionados.
- No linealidad en los parámetros.
- Datos faltantes
- Valores extremos o atípicos.

Usualmente primero se tienen que detectar estos problemas y posteriormente, buscar transformaciones de los predictores o de la variable de respuesta, para tratar de reducir el problema

Diferencias entre errores y residuales I

- Como ya se ha visto, los errores poblacionales ϵ_i y los errores muestrales e_i son estimados por los residuales $\hat{e}_i = y_i - \hat{y}_i$.
- Los errores **e** son variables aleatorias que no son observables, y que deben tener $\mathsf{E}(\mathbf{e}|\mathbf{X}) = 0$ y $\mathsf{Var}(\mathbf{e}|\mathbf{X}) = \sigma^2\mathbf{I}$.
- Por otra parte, los residuales $\hat{\mathbf{e}}$ son cantidades que podemos calcular, y que tienen media y varianza dadas por:

$$
\begin{aligned}
\mathsf{E}(\hat{\mathbf{e}}|\mathbf{X}) &= \mathbf{0} \\
\mathsf{Var}(\hat{\mathbf{e}}|\mathbf{X}) &= \sigma^2(\mathbf{I} - \mathbf{H})
\end{aligned}
$$

donde $\mathbf{H}$ es la matriz sombrero $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Para cada observación, el elementos h_{ii} se conoce como el *apalancamiento* de la observación i.

Diferencias entre errores y residuales II

- De manera individual se tiene que la varianza del residual $\hat{e}_i$ es

$$\text{Var}(\hat{e}_i) = \sigma^2(1 - h_{ii})$$

donde h_{ii} es el elemento i de la diagonal de $\mathbf{H}$. Por lo tanto, los residuales **no** tienen varianza constante. Podemos considerar entonces los *residuales estandarizados*.

Residuales estandarizados

Los residuales estandarizados se definen como

$$\hat{e}_{std,i} = \frac{\hat{e}_i}{\hat{\sigma}\sqrt{1 - h_{ii}}}$$

- Consideren el siguiente ejemplo: Cielito Querido (CQ) quiere diseñar programas de entrega de insumos en sus cafeterías del centro histórico, en donde tiene 10 cafeterías, así que CQ quiere estimar el tiempo de viaje total de sus choferes. Tomando una muestra de 10 entregas se obtienen los siguientes datos: Sea $X_1 =$ km recorridos, $X_2 =$ Número de envíos, y $Y =$horas viajadas en hrs.

Gráficas de residuales estandarizados II

```
datos <- data.frame(Y = c(9.3, 4.8, 8.9, 6.5, 4.2, 6.2, 7.4, 6.0, 7.6, 6.1),
X1 = c(100, 50, 100, 100, 50, 80, 75, 65, 90, 90),
X2 = c(4, 3, 4, 2, 2, 2, 3, 4, 3, 2))
modelo <- lm(Y ~ X1 +X2, data=datos)
summary(modelo)


Call:
lm(formula = Y ~ X1 + X2, data = datos)

Residuals:
     Min       1Q   Median       3Q      Max
-0.79875 -0.32477  0.06333  0.29739  0.91333

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.868701   0.951548  -0.913 0.391634
X1           0.061135   0.009888   6.182 0.000453 ***
X2           0.923425   0.221113   4.176 0.004157 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5731 on 7 degrees of freedom
Multiple R-squared:  0.9038,Adjusted R-squared:  0.8763
F-statistic: 32.88 on 2 and 7 DF,  p-value: 0.0002762
```

Gráficas de residuales estandarizados III

- A continuación se muestra la gráfica de los residuales estandarizados. ¿Qué pueden decir?

```
modelo$residuals  # obten los residuales del modelo
```

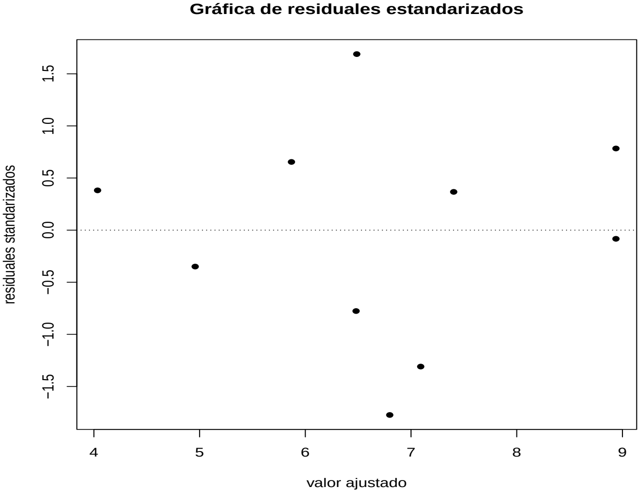
```
          1           2           3           4           5           6 
 0.36154012 -0.15830457 -0.03845988 -0.59160915  0.16512079  0.33108283 
          7           8           9          10 
 0.91333046 -0.79874892  0.19631148 -0.38026316 
```

```
rstandard(modelo) # esta función devuelve los residuales estandarizados, para graficar
```

```
          1           2           3           4           5           6 
 0.78344317 -0.34961582 -0.08334104 -1.30928723  0.38166807  0.65430764 
          7           8           9          10 
 1.68916740 -1.77371906  0.36702765 -0.77639406 
```

```
plot(modelo$fit,rstandard(modelo), pch=16,
main = "Gráfica de residuales estandarizados",
xlab = "valor ajustado",
ylab = "residuales standarizados")
abline(h=0, lty=3)
```

Gráficas de residuales estandarizados IV



Gráficas de residuales estandarizados V

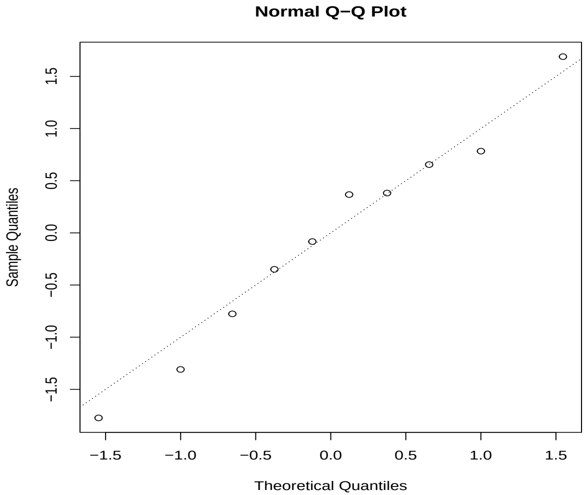
- No se ven anormalidades. Los residuales están entre -2 y 2, por lo que no hay razón para cuestionar los supuestos de que los errores son normales.

Análisis de residuales I

- Por otra parte, podemos hacer una gráfica llamada *Q-Q-plot* para ver si parece o no que los residuales tengan una distribución normal. Si los datos quedan sobre una linea recta, entonces podemos suponer que los datos son normales

```
par(pty="s")
qqnorm(rstandard(modelo))
abline(a=0,b=1, lty=3)
```

Análisis de residuales II



Análisis de residuales III

- La gráfica anterior muestra que los datos tienen una ligera desviación de la recta, por lo que no parecen ser normales.

Residuales y Valores extremos I

Valores extremos y valores influenciales

- Un *valor extremo* o *outlier* es una observación que es atípica en comparación con los otros datos, y no sigue el mismo patrón que el resto de los datos.
- Un *valor influencial* es un punto que cambia drásticamente el modelo si se elimina del conjunto de datos. Se pueden encontrar estudiando los valores de la diagonal de **H** que se llaman *apalancamientos*, y los valores de la muestra que tienen $h_{ii} > 0.5$ son los que mayor peso tienen en la estimación.

- Una regla simple para detectar un valor extremo es cuando el valor de su residual estandarizado es $|\hat{e}_{std,i}| > 2$. Pero esta regla puede fallar si el error estándar del residual es muy grande.
- Para resolver este problema, se introduce otra definición más, la de *residual estudentizado*

Residual estudentizado I

- Supongamos que la observación i se borra del conjunto de datos y se estima de nuevo la ecuación de regresión, con los $n-1$ puntos restantes.
- También calculamos el error estándar $s_{(i)} = \hat{\sigma}_{(i)}$ del estimado considerando el conjunto de datos sin la observación i.

Residual studentizado

Si calculamos la desviación estándar del residual i usando $s_{i(i)}$ en lugar de s, se obtiene el residual estudentizado:

$$\hat{e}_{stu,i} = \frac{\hat{e}_i}{s_{(i)}\sqrt{1-h_{ii}}}$$

- Si la observación i es un valor extremo, seguro $s_{(i)} < s$, y por lo tanto $|\hat{e}_{stu,i}| > |\hat{e}_{std,i}|$
- Así que para detectar outliers, es mejor usar los residuales estudentizados.

Residual estudentizado II

- La siguiente tabla muestra todos los tipos de residuales para el ejemplo de CQ:

```
library(MASS)  #para obtener los residuales estudentizados
residuales <- data.frame(residuales = modelo$residuals,
res_std = rstandard(modelo),
res_stu = studres(modelo))
residuales
```

```
    residuales     res_std     res_stu
1   0.36154012  0.78344317  0.75938374
2  -0.15830457 -0.34961582 -0.32654491
3  -0.03845988 -0.08334104 -0.07719712
4  -0.59160915 -1.30928723 -1.39494328
5   0.16512079  0.38166807  0.35709105
6   0.33108283  0.65430764  0.62519106
7   0.91333046  1.68916740  2.03187180
8  -0.79874892 -1.77371906 -2.21314094
9   0.19631148  0.36702765  0.34311914
10 -0.38026316 -0.77639406 -0.75190409
```

- Con los residuales estudentizados, podemos decir que una observación es un outlier al α % de confianza si $|\hat{e}_{stu,i}| > t_{(n-p-2,1-\alpha/2)}$, donde p es el números de predictores. En nuestro ejemplo, $n = 10$, $p = 2$ y $\alpha = 5$ %, así que

```
qt(.975,10-2-2)
```

```
[1] 2.446912
```

así que en nuestro ejemplo no tenemos outliers.

Detección de puntos influenciales I

- Para determinar qué puntos pueden ser influenciales en la regresión, se utiliza una metrica conocida como la *distancia de Cook*.

Distancias de Cook

La distancia de Cook se define como:

$$D_i = \frac{\hat{e}_i^2}{(p+1)\hat{\sigma}^2}\left[\frac{h_{ii}}{(1-h_{ii})^2}\right]$$

- El valor de la distancia de Cook será grande e indicará una observación influencial si el residual o el apalancamiento son grandes. Como regla de dedo, si $D_i > 1$ la obsevación i es influencial y debe estudiarse con más cuidado.

Detección de puntos influenciales II

- Para obtener las distancias de Cook en R, usamos la instrucción:

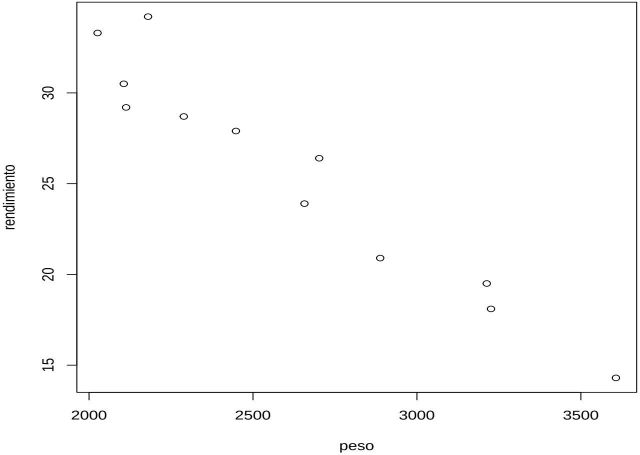
```
cooks.distance(modelo)

          1           2           3           4           5           6 
0.110993567 0.024536364 0.001256032 0.347923173 0.036663491 0.040381085 
          7           8           9          10 
0.117561490 0.650028500 0.006656224 0.074217109 
```

La distancia más grande es 0.65, por lo tanto, en este conjunto de datos no se tienen observaciones influenciales.

- Cuando la varianza no es constante, es conveniente considerar transformar la variable dependiente.
- En el siguiente conjunto de datos, se quiere relacionar el rendimiento de un auto (medido como km por litro) en función de su peso (en kl).

```
peso <- c(2289, 2113, 2180, 2448, 2026, 2702, 2657, 2106, 3226, 3213, 3607, 2888)
rendimiento <- c(28.7, 29.2, 34.2, 27.9, 33.3, 26.4, 23.9, 30.5, 18.1, 19.5, 14.3, 20.9)
plot(peso, rendimiento)
```

Problemas de varianza no constante (heteroscedasticidad) II



Problemas de varianza no constante (heteroscedasticidad) III

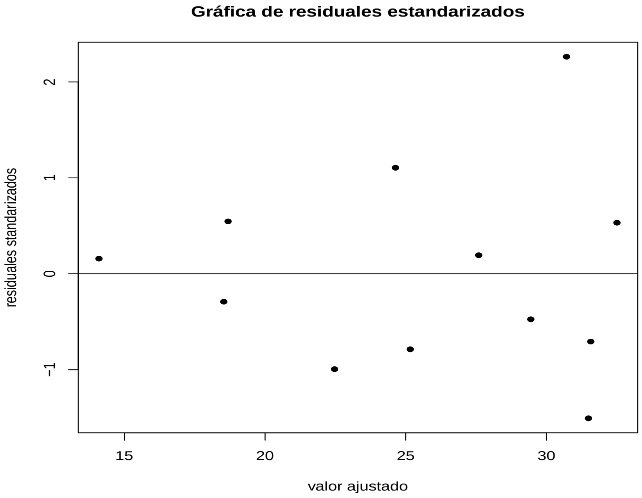
```
modelo2 <- lm(rendimiento ~ peso)
summary(modelo2)


Call:
lm(formula = rendimiento ~ peso)

Residuals:
    Min      1Q  Median      3Q     Max
-2.2928 -1.1204 -0.1155  0.7994  3.4873

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 56.0956800  2.5821354   21.73 9.55e-10 ***
peso        -0.0116436  0.0009677  -12.03 2.85e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.671 on 10 degrees of freedom
Multiple R-squared:  0.9354,Adjusted R-squared:  0.9289
F-statistic: 144.8 on 1 and 10 DF,  p-value: 2.85e-07
```

- Si vemos una gráfica de los residuales estandarizados, podemos ver que presentan un patron, mostrando que a valores de pesos más grandes, la varianza de los errores es mayor.

Problemas de varianza no constante (heteroscedasticidad) IV

```r
plot(modelo2$fit, rstandard(modelo2), pch=16,
main = "Gráfica de residuales estandarizados",
xlab = "valor ajustado",
ylab = "residuales standarizados")
abline(h=0)   # línea de referencia del origen
```

Problemas de varianza no constante (heteroscedasticidad) V



Problemas de varianza no constante (heteroscedasticidad) VI

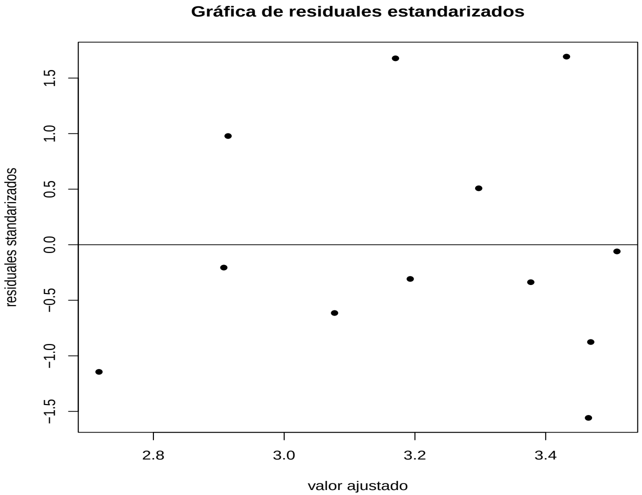
- En este ejemplo, se puede elegir una transformación a alguna potencia o al logaritmo. Por ejemplo:

Problemas de varianza no constante (heteroscedasticidad) VII

```
modelo3 <- lm(log(rendimiento) ~ peso)
summary(modelo3)


Call:
lm(formula = log(rendimiento) ~ peso)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.09125 -0.04079 -0.01536  0.03736  0.10310 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  4.524e+00  9.932e-02   45.55 6.26e-13 ***
peso        -5.011e-04  3.722e-05  -13.46 9.84e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.06425 on 10 degrees of freedom
Multiple R-squared:  0.9477,Adjusted R-squared:  0.9425 
F-statistic: 181.2 on 1 and 10 DF,  p-value: 9.842e-08
```

```
plot(modelo3$fit, rstandard(modelo3), pch=16,
main = "Gráfica de residuales estandarizados",
xlab = "valor ajustado",
ylab = "residuales standarizados")
abline(h=0)   # línea de referencia del origen
```

Problemas de varianza no constante (heteroscedasticidad) VIII



Problemas de varianza no constante I

- En el caso de los modelos lineales, los errores estándar de los parámetros serán erróneos si hay heteroscedasticidad (varianza no constante) en los errores.
- En estos casos, conviene tener estimadores más robustos de los errores estándar, ya sea a través de métodos no paramétricos como el *bootstrap* outilizando propiedades teóricas de los estimadores.
- En el caso especial de regresión lineal, se pueden usar los estimadores de varianza tipo *sandwich*. Recordando que en general $\text{Var}(\mathbf{a}\epsilon) = \mathbf{a}'\text{Var}(\epsilon)\mathbf{a}$, para un vector fijo **a** se tiene que:

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

es el estimador máximo verosímil, podemos ver que su varianza se puede escribir como

$$\text{Var}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\text{Var}(\mathbf{y})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\Sigma\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

entonces la varianza del error Σ está en sandwich con la proyección $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$

Problemas de varianza no constante II

- Para cuantificar la incertidumbre de los estimadores en un escenario heteroscedástico donde cada error tiene varianza diferente, por ejemplo $\mathbf{\Sigma} = diag(\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2)$, usamos un estimador de los errores que es consistente con heteroscedasticidad (HC), reemplazando $\mathbf{\Sigma}$ por un estimador que permite esas varianzas diferentes.

```
library(AER)
vcovHC(modelo)  # estimador robusto
```

```
             (Intercept)            X1           X2
(Intercept)  0.345264938 -0.0024485563 -0.042511012
X1          -0.002448556  0.0001221214 -0.002635846
X2          -0.042511012 -0.0026358462  0.092807101
```

```
vcov(modelo)    # estimador usual
```

```
             (Intercept)            X1            X2
(Intercept)  0.905443072 -6.795163e-03 -0.1134416683
X1          -0.006795163  9.778232e-05 -0.0003542838
X2          -0.113441668 -3.542838e-04  0.0488911625
```

- Por ejemplo, el error estándar para $\hat{\beta}_1$ es

```
sqrt(vcovHC(modelo)[2,2])
```

```
[1] 0.01105085
```

mucho más grande que el que da el modelo bajo el supuesto de varianza constante:

Problemas de varianza no constante III

```
sqrt(vcov(modelo)[2,2])
```

```
[1] 0.009888495
```

Regresión multivariada múltiple

Regresión multivariada múltiple I

- Consideramos el problema de modelar m variables de respuesta $Y_1, \ldots, Y_m$ en los mismos p predictores. Esto es equivalente a tener m *regresiones simultáneas*:

$$
\begin{aligned}
Y_1 &= \beta_{01} + \beta_{11}x_1 + \cdots + \beta_{p1}x_p + \epsilon_1 \\
Y_2 &= \beta_{02} + \beta_{12}x_1 + \cdots + \beta_{p2}x_p + \epsilon_2 \\
\vdots &= \vdots \\
Y_m &= \beta_{0m} + \beta_{1m}x_1 + \cdots + \beta_{pm}x_p + \epsilon_m
\end{aligned}
$$

  Los errores ϵ son ahora una matriz de n renglones con m columnas, con

$$
\mathrm{E}(\epsilon_{(i)}) = 0 \text{ y } \mathrm{cov}\left(\boldsymbol{\epsilon}_{(i)}, \boldsymbol{\epsilon}_{(j)}\right) = \sigma_{ij}\mathbf{I}, \quad i, j = 1, 2, \ldots, m
$$

- La simultaneidad puede ocasionar que los errores estén correlacionados. Los errores para diferentes respuestas en el mismo caso pueden estar correlacionados.

Regresión multivariada múltiple II

- Para este caso, veremos cómo resolver un ejemplo con `R`, en lugar de revisar toda la notación que implica este modelo. No hay muchas diferencias conceptuales excepto en la manera en la que se hacen hipótesis para probar los parámetros de la regresión y los intervalos de confianza para predicción.

Ejemplo con m=2 respuestas I

Consideremos el ejemplo 7.25 que involucra 17 sobredosis de la droga amitriptylina (antidepresivo). Hay dos respuestas: `TOT`: es el nivel de plasma TCAD y `AMI`:es la cantidad de amitriptylina presente en el nivel de plasma TCAD. Los predictores son:

- `GEN`, sexo (H=0,M=1)
- `AMT`, cantidad de droga tomada al tiempo de la sobredosis
- `PR`, medida de onda PR
- `DIAP`, presión de la sangre diastólica
- `QRS`, medida de onda QRS

En este ejemplo, $p = 5$, $n = 17$, $m = 2$.

```
datos <- read.csv("~/Dropbox/Academia/ITAM/2022-I/EA3_S22_I/data/J&W/T7-6.DAT",header=F,sep="")
colnames(datos) <- c("TOT","AMI","GEN","AMT","PR","DIAP","QRS")
head(datos)
```

```
   TOT  AMI GEN  AMT  PR DIAP QRS
1 3389 3149   1 7500 220    0 140
2 1101  653   1 1975 200    0 100
3 1131  810   0 3600 205   60 111
4  596  448   1  675 160   60 120
5  896  844   1  750 185   70  83
6 1767 1450   1 2500 180   60  80
```

Entonces el modelo multivariado se puede estimar como se ve en la siguiente lámina

Regresión multivariada múltiple

```
rmm1 <- lm(cbind(TOT, AMI) ~ GEN + AMT + PR + DIAP + QRS, data = datos)
summary(rmm1)

Response TOT :

Call:
lm(formula = TOT ~ GEN + AMT + PR + DIAP + QRS, data = datos)

Residuals:
   Min     1Q Median     3Q    Max 
-399.2 -180.1    4.5  164.1  366.8 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -2.879e+03  8.933e+02  -3.224 0.008108 ** 
GEN          6.757e+02  1.621e+02   4.169 0.001565 ** 
AMT          2.848e-01  6.091e-02   4.677 0.000675 ***
PR           1.027e+01  4.255e+00   2.414 0.034358 *  
DIAP         7.251e+00  3.225e+00   2.248 0.046026 *  
QRS          7.598e+00  3.849e+00   1.974 0.074006 .  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 281.2 on 11 degrees of freedom
Multiple R-squared:  0.8871,Adjusted R-squared:  0.8358 
F-statistic: 17.29 on 5 and 11 DF,  p-value: 6.983e-05


Response AMI :

Call:
lm(formula = AMI ~ GEN + AMT + PR + DIAP + QRS, data = datos)

Residuals:
    Min      1Q  Median      3Q     Max 
-373.85 -247.29  -83.74  217.13  462.72 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -2.729e+03  9.288e+02  -2.938 0.013502 *  
GEN          7.630e+02  1.685e+02   4.528 0.000861 ***
AMT          3.064e-01  6.334e-02   4.837 0.000521 ***
PR           8.896e+00  4.424e+00   2.011 0.069515 .  
DIAP         7.206e+00  3.354e+00   2.149 0.054782 .  
QRS          4.987e+00  4.002e+00   1.246 0.238622    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 292.4 on 11 degrees of freedom
Multiple R-squared:  0.8764,Adjusted R-squared:  0.8202 
F-statistic:  15.6 on 5 and 11 DF,  p-value: 0.0001132
```

Regresión multivariada múltiple I

Estas dos regresiones son las mismas que si hubiéramos hecho dos regresiones separadas. Pero ahora tenemos dos conjuntos de residuales, y prácticamente de todos los parámetros:

```
head(resid(rmm1))

         TOT        AMI
1  132.82172  161.52769
2  -72.00392 -264.35329
3 -399.24769 -373.85244
4 -382.84730 -247.29456
5 -152.39129   15.78777
6  366.78644  217.13206

coef(rmm1)

                   TOT           AMI
(Intercept) -2879.4782461 -2728.7085444
GEN           675.6507805   763.0297617
AMT             0.2848511     0.3063734
PR             10.2721328     8.8961977
DIAP            7.2511714     7.2055597
QRS             7.5982397     4.9870508

sigma(rmm1)

     TOT      AMI
281.2324 292.4363
```

Donde las cosas comienzan a ser diferentes es cuando obtenemos las covarianzas de los estimadores. Los coeficientes de los dos modelos estan correlacionados, y su covarianza tiene que ser tomada en cuenta para determinar cuánto contribuye cada predictor a los modelos.

```
options(width = 150, digits = 2, scipen = 10)
vcov(rmm1)
```

```
                TOT:(Intercept)  TOT:GEN TOT:AMT    TOT:PR  TOT:DIAP   TOT:QRS AMI:(Intercept)  AMI:GEN AMI:AMT    AMI:PR  AMI:DIAP   AMI:QRS
TOT:(Intercept)        797914.0 -61055.4 11.2369 -3157.872 -1625.349 -1414.943        702227.8 -53733.6  9.8894 -2779.179 -1430.437 -1245.262
TOT:GEN                -61055.4  26262.0  1.4171   150.023   162.904    49.084        -53733.6  23112.7  1.2471   132.032   143.369    43.197
TOT:AMT                    11.2      1.4  0.0037    -0.110     0.082    -0.054             9.9      1.2  0.0033    -0.097     0.072    -0.048
TOT:PR                  -3157.9    150.0 -0.1097    18.104     3.132    -0.428         -2779.2    132.0 -0.0965    15.933     2.756    -0.377
TOT:DIAP                -1625.3    162.9  0.0817     3.132    10.402     2.536         -1430.4    143.4  0.0719     2.756     9.154     2.232
TOT:QRS                 -1414.9     49.1 -0.0542    -0.428     2.536    14.814         -1245.3     43.2 -0.0477    -0.377     2.232    13.037
AMI:(Intercept)        702227.8 -53733.6  9.8894 -2779.179 -1430.437 -1245.262        862755.9 -66017.0 12.1501 -3414.495 -1757.432 -1529.927
AMI:GEN                -53733.6  23112.7  1.2471   132.032   143.369    43.197        -66017.0  28396.2  1.5322   162.214   176.143    53.072
AMI:AMT                     9.9      1.2  0.0033    -0.097     0.072    -0.048            12.2      1.5  0.0040    -0.119     0.088    -0.059
AMI:PR                  -2779.2    132.0 -0.0965    15.933     2.756    -0.377         -3414.5    162.2 -0.1186    19.575     3.386    -0.463
AMI:DIAP                -1430.4    143.4  0.0719     2.756     9.154     2.232         -1757.4    176.1  0.0884     3.386    11.247     2.742
AMI:QRS                 -1245.3     43.2 -0.0477    -0.377     2.232    13.037         -1529.9     53.1 -0.0586    -0.463     2.742    16.018
```

Regresión multivariada múltiple III

Para determinar la significancia de los coeficientes, se requieren pruebas multivariadas formales, ya que podemos llegar a interpretaciones contradictorias en los dos modelos.
Para determinar si se incluyen o no predictores en una regresión múltiple multivariada, se requieren de estadísticas multivariadas.
La función `Anova` del paquete `car` puede ayudar a estas pruebas:

```
library(car)
Anova(rmm1)
```

```
Type II MANOVA Tests: Pillai test statistic
     Df test stat approx F num Df den Df Pr(>F)
GEN   1     0.655     9.50      2     10 0.0049 **
AMT   1     0.691    11.18      2     10 0.0028 **
PR    1     0.346     2.65      2     10 0.1192
DIAP  1     0.324     2.39      2     10 0.1414
QRS   1     0.292     2.06      2     10 0.1781
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Regresión multivariada múltiple IV

Como `rmm1` es múltiple multivariada, se ajusta automáticamente un MANOVA (múltiple ANOVA). Se tienen en este caso sumas de cuadrados de tipo II, que se interpreta en el sentido de que los predictores se prueban considerando que están ya en el modelo. En este caso se puede ver que conjuntamente `PR` y `DIAP` no son significativos a pesar de lo que dicen los modelos marginales. Podemos actualizar el modelo eliminando esos predictores y `QRS` que tampoco es significativo:

```
rmm2 <- update(rmm1, . ~ . - PR - DIAP - QRS)
anova(rmm1, rmm2) #compara H0: modelo chico vs Ha: modelo grande

Analysis of Variance Table

Model 1: cbind(TOT, AMI) ~ GEN + AMT + PR + DIAP + QRS
Model 2: cbind(TOT, AMI) ~ GEN + AMT
  Res.Df Df Gen.var. Pillai approx F num Df den Df Pr(>F)
1     11       43803
2     14  3    51856  0.604     1.59      6     22    0.2
```