Estadística Aplicada III

Regresión Lineal Múltiple

Jorge de la Vega Góngora

Departamento de Estadística,
Instituto Tecnológico Autónomo de México

Semana 7: 26/28 de septiembre de 2018

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

- 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...)

Modelo

Elementos del modelo de Regresión Lineal Múltiple (RLM) I

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

Elementos del modelo de Regresión Lineal Múltiple (RLM) II

- 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 RLM

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:

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

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

- 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 consisten 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 la matriz de diseño $\mathbf{X}$. Ejemplos de predictores 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 $\mathbf{1}_n$.
  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. Componentes principales o factores: combinación de predictores procesados previamente.

Ejemplos I

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.

Ejemplos II

□

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 $i = 1, 2, \dots, n$.

- 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*, `temp` y tres *términos* en el modelo:1, `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.

Ejemplos III

- 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.

$\square$

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 $z_i = I[\text{obs } i \in \text{ Población } i]$ es una función indicadora de población, y se tienen 3 poblaciones, 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.

$\square$

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 $\mathbf{b}$ 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*.

- 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.

Estimación III

- 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 $\hat{\boldsymbol{\beta}}$
  - Como el estimador $\hat{\boldsymbol{\beta}}$ 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: $\hat{\boldsymbol{\beta}}$ es BLUE, el mejor estimador (mínima varianza) lineal insesgado, sobre la clase de todos los estimadores lineales insesgados.
- Propiedades de los residuales $\hat{\boldsymbol{\epsilon}} = \mathbf{e}$:
  - $\mathrm{E}(\mathbf{e}) = \mathbf{0}$, $\mathrm{Var}(\mathbf{e}) = \sigma^2(\mathbf{I} - \mathbf{H})$
  - 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 $\mathbf{X}$ y un vector de respuestas $\mathbf{y}$,

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

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

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

Ejemplo II

- $$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}\mathbf{y} = \begin{pmatrix} 2.23 \\ -0.74 \\ 1.46 \end{pmatrix}$$
- Calculamos $\mathbf{H}$ y $\hat{\sigma}^2$:

```
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

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

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 $\hat{\boldsymbol{\beta}}$?

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

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

- 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 $\mathrm{Var}(\texttt{time}|\texttt{T1},\texttt{T2})$ 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?

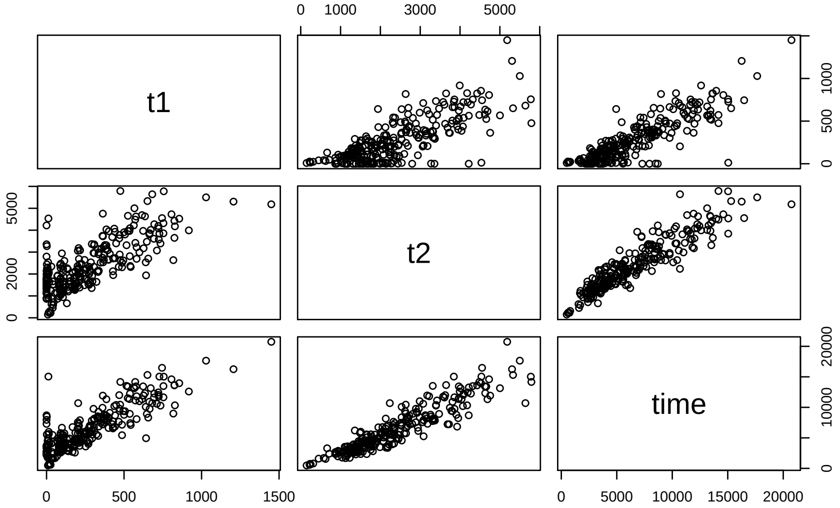
```
summary(Transact)

       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 $\widehat{\texttt{Time}} = 144.4 + 5.5\texttt{T1} + 2\texttt{T2}$.
- Hay $n - k = 261 - 3 = 258$ grados de libertad.
- Para calcular $\hat{\sigma}^2 = \frac{RSS}{n-k}$, tomamos el `Residual standard error`: 1143 y se calcula: $RSS = 1143^2 = 1.306449 \times 10^6$.
- Noten que la desviación estándar del tiempo dados `T1` y `T2` (1143) es *menor* que la desviación estándar del tiempo ignorando la información (3774).

Inferencia

Inferencia en regresión I

- Bajo el supuesto de normalidad de los errores, como $\hat{\boldsymbol{\beta}}$ es una combinación lineal de los errores, se sigue que

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

  recordando que $k = p + 1$, y además:

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

- 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:

Inferencia en regresión II

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

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

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

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

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

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

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

E1: regiones de confianza

- Continuando con el ejemplo de transacciones, podemos hallar los intervalos.
- La matriz de varianzas y covarianzas de $\hat{\boldsymbol{\beta}}$ 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$coef[i] + c(-1,1)*pt(258,.025,lower.tail=F)*sqrt(diag(vcov(m1))[i])
a <- cbind(m1$coef,int_marginales); colnames(a) <- c("betahat","lim_inf","lim_sup"); a

               betahat   lim_inf    lim_sup
(Intercept) 144.369443 74.694865 214.044020
t1            5.462057  5.285048   5.639065
t2            2.034549  1.996008   2.073089
```

- 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$coef[i] + c(-1,1)*sqrt(pf(3,258,.05,lower.tail=F))*sqrt(diag(vcov(m1))[i]
a <- cbind(m1$coef,int_simultaneos); colnames(a) <- c("betahat","lim_inf","lim_sup"); a

               betahat    lim_inf    lim_sup
(Intercept) 144.369443 -17.366141 306.105027
t1            5.462057   5.051167   5.872946
t2            2.034549   1.945084   2.124013
```

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:

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

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

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

- 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 $\mathbf{X}$ es $\mathbf{1}$, 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:
  $$\begin{pmatrix} \text{Suma de cuadrados} \\ \text{total alrededor} \\ \text{de la media} \end{pmatrix} = \begin{pmatrix} \text{Suma de cuadrados} \\ \text{debida a la regresión} \end{pmatrix} + \begin{pmatrix} \text{Suma de cuadrados} \\ \text{residual} \end{pmatrix}$$

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

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

Cálculo de R^2 y R^2_{aj} II

- 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

- ANOVA es una forma de agrupar información para comparar modelos. la tabla de ANOVA que se obtiene de un modelo de regresión lineal múltiple sirve para comparar los siguientes modelos:

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

  donde $\mathbf{u}$ es el vector de términos del modelo.
- 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.

ANOVA y comparación de modelos en RLM II

- Pruebas parciales para cada coeficiente se pueden obtener. Por ejemplo, para los datos de transacciones:

```
summary(aov(m1))

             Df    Sum Sq   Mean Sq F value Pr(>F)
t1            1 2.759e+09 2.759e+09  2113.7 <2e-16 ***
t2            1 6.072e+08 6.072e+08   465.1 <2e-16 ***
Residuals   258 3.368e+08 1.305e+06
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- 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:

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

- La prueba anterior es equivalente a la prueba de hipótesis: $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$.

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 ...
```

- Necesitamos convertir `Rank` a un factor eliminando el primer nivel; `R` hace esto de manera automática cuando convierte la variable a factor, creando variables indicadoras `rankAssoc` y `rankProf`. La variable `Sex` también puede ser considerado un factor o no.

Ejemplo 2: Comparando Salarios II

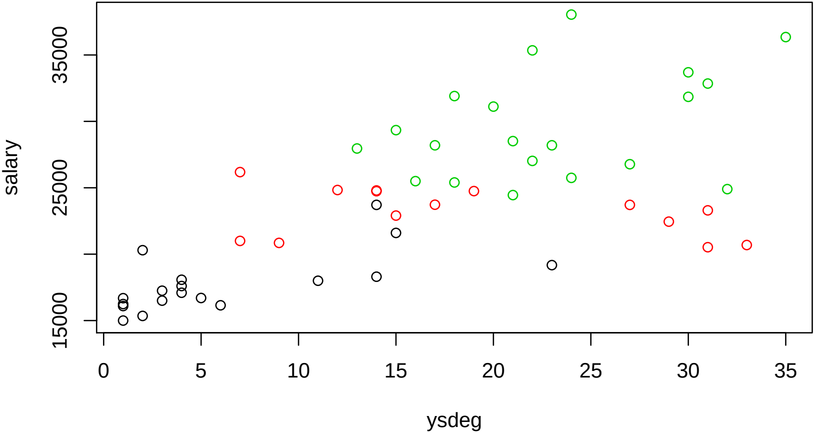
- El objetivo del estudio es entender cuál es la dependencia del salario al rango académico y a la experiencia, medida por los años desde la graduación. Consideren primero la regresión de `salary` respecto a `rank` y `ysdeg` o en nuestra notación,

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

  Una gráfica de estos datos se muestra a continuación. Noten que el factor `Rank` se agrega como una variable para marcar.

```
with(salary,plot(ysdeg,salary,col=rank))
```

Ejemplo 2: Comparando Salarios III



- Hay cuatro casos a considerar:

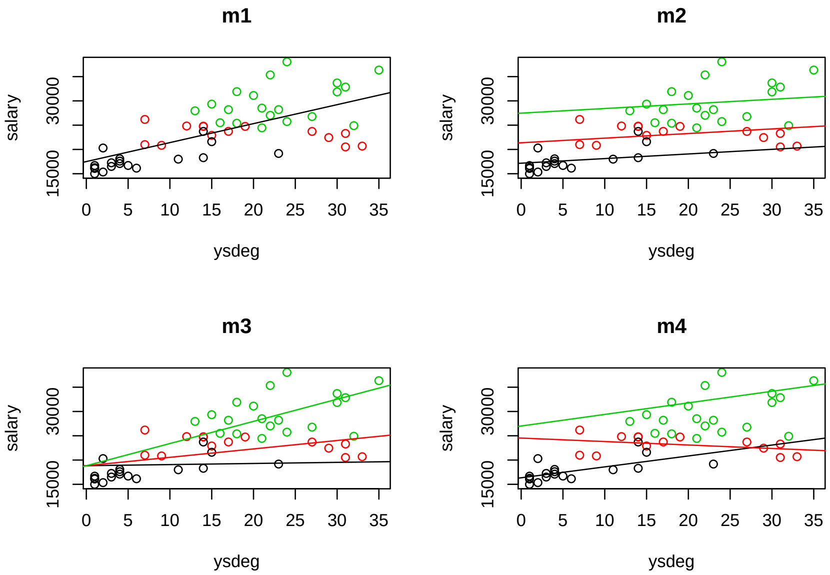
Ejemplo 2: Comparando Salarios IV

1. un modelo de regresión lineal simple para todos los rangos (básicamente ignirando 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

```r
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 V



Comparando salarios I

Comparando salarios II

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
```

```
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
```

```
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 3 y 4, se establecen las hipótesis del siguiente modo:

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

- Esto es equivalente a la hipótesis de que $NH: \beta_1 = \beta_2 = 0$. 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 3 y 4: $RSS_{NH} = 533,456,012$, $RSS_{AH} = 357,499,755$, $gl_{NH} = 48$, $gl_{AH} = 46$, y $\hat{\sigma}^2_{AH} = 7,771,734$. De este modo

$$F = \frac{(533,456,012 - 357,499,755)/(48 - 46)}{7,771,734} = 11.3203$$

  con 2 y 46 grados de libertad. Esto da un p$-value = 0.000100497$, lo que dice que el modelo general (modelo 4) explica mejor que el modelo 3 que tiene una ordenada al origen común.

Regresión multivariada múltiple

Regresión multivariada múltiple I

- Consideramos el problema de modelar m respuestas en los mismos p predictores. Esto es equivalente a tener m regresiones simultáneas.
- La simultaneidad puede ocasionar que los errores estén correlacionados.
- 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.

Ejercicio 7.25: Dos respuestas.
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

Regresión multivariada múltiple II

```
datos <- read.csv("../data/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 escribir 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.

Regresión multivariada múltiple II

```r
vcov(rmm1)
```

```
                TOT:(Intercept)       TOT:GEN      TOT:AMT        TOT:PR       TOT:DIAP        TOT:QRS AMI:(Intercept)
TOT:(Intercept)   797914.009706 -61055.389981 11.236944164 -3.157872e+03 -1.625349e+03 -1.414943e+03   702227.760065
TOT:GEN           -61055.389981  26262.026594  1.417050110  1.500225e+02  1.629044e+02  4.908351e+01   -53733.596885
TOT:AMT               11.236944      1.417050  0.003710048 -1.096946e-01  8.172903e-02 -5.416651e-02        9.889404
TOT:PR             -3157.872390    150.022536 -0.109694600  1.810410e+01  3.131923e+00 -4.282235e-01    -2779.178744
TOT:DIAP           -1625.349406    162.904357  0.081729026  3.131923e+00  1.040163e+01  2.535578e+00    -1430.436687
TOT:QRS            -1414.943019     49.083508 -0.054166512 -4.282235e-01  2.535578e+00  1.481381e+01    -1245.262340
AMI:(Intercept)   702227.760065 -53733.596885  9.889404164 -2.779179e+03 -1.430437e+03 -1.245262e+03   862755.907989
AMI:GEN           -53733.596885  23112.671147  1.247116747  1.320318e+02  1.433688e+02  4.319739e+01   -66017.011582
AMI:AMT                9.889404      1.247117  0.003265137 -9.653997e-02  7.192804e-02 -4.767084e-02       12.150106
AMI:PR             -2779.178744    132.031758 -0.096539968  1.593304e+01  2.756341e+00 -3.768708e-01    -3414.494580
AMI:DIAP           -1430.436687    143.368784  0.071928040  2.756341e+00  9.154259e+00  2.231510e+00    -1757.432236
AMI:QRS            -1245.262340     43.197389 -0.047670836 -3.768708e-01  2.231510e+00  1.303733e+01    -1529.927329
                      AMI:GEN      AMI:AMT        AMI:PR      AMI:DIAP       AMI:QRS
TOT:(Intercept) -53733.596885  9.889404164 -2.779179e+03 -1.430437e+03 -1.245262e+03
TOT:GEN          23112.671147  1.247116747  1.320318e+02  1.433688e+02  4.319739e+01
TOT:AMT              1.247117  0.003265137 -9.653997e-02  7.192804e-02 -4.767084e-02
TOT:PR             132.031758 -0.096539968  1.593304e+01  2.756341e+00 -3.768708e-01
TOT:DIAP           143.368784  0.071928040  2.756341e+00  9.154259e+00  2.231510e+00
TOT:QRS             43.197389 -0.047670836 -3.768708e-01  2.231510e+00  1.303733e+01
AMI:(Intercept) -66017.011582 12.150106211 -3.414495e+03 -1.757432e+03 -1.529927e+03
AMI:GEN          28396.190973  1.532205650  1.622140e+02  1.761427e+02  5.307224e+01
AMI:AMT              1.532206  0.004011543 -1.186089e-01  8.837068e-02 -5.856831e-02
AMI:PR             162.214008 -0.118608851  1.957531e+01  3.386437e+00 -4.630228e-01
AMI:DIAP           176.142660  0.088370675  3.386437e+00  1.124691e+01  2.741630e+00
AMI:QRS             53.072244 -0.058568314 -4.630228e-01  2.741630e+00  1.601764e+01
```

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   0.65521   9.5015      2     10 0.004873 **
AMT   1   0.69097  11.1795      2     10 0.002819 **
PR    1   0.34649   2.6509      2     10 0.119200
DIAP  1   0.32381   2.3944      2     10 0.141361
QRS   1   0.29184   2.0606      2     10 0.178092
---
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)

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 0.60386   1.5859      6     22 0.1983
```