Apunte de Regresión Lineal

María Eugenia Szretter Noste
Carrera de Especialización en Estadística
para Ciencias de la Salud
Facultad de Ciencias Exactas y Naturales,
Universidad de Buenos Aires

Agosto - Octubre de 2017

4.3. Modelo de Regresión Lineal Múltiple

El modelo de regresión lineal múltiple es un modelo para la variable aleatoria Y cuando se conocen $X_1, X_2, \ldots, X_{p-1}$ las variables regresoras. El modelo es

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_{p-1} X_{ip-1} + \varepsilon_i, \tag{44}$$

donde $\beta_0, \beta_1, \ldots, \beta_{p-1}$ son parámetros (es decir, números) desconocidos, X_{i1}, X_{i2}, $\ldots, X_{ip-1}$ son los valores de las variables predictoras medidas en el i-ésimo individuo (o i-ésima repetición del experimento o i-ésima unidad experimental, según el caso) con $1 \leq i \leq n$, n es el tamaño de muestra, Y_i es la variable respuesta medida en el i-ésimo individuo (observado) y ε_i es el error para el individuo i-ésimo, que no es observable. Haremos supuestos sobre ellos:

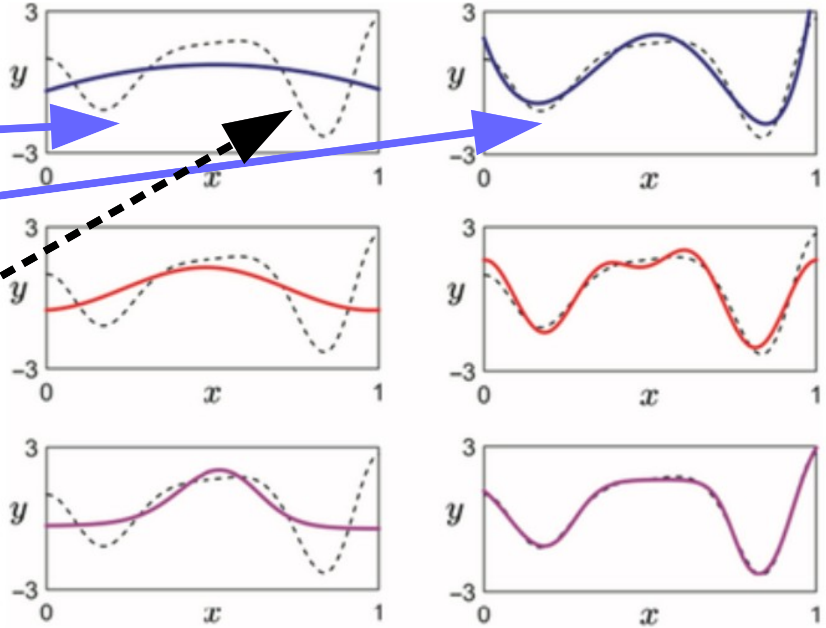
$$\varepsilon_i \sim N\left(0, \sigma^2\right), 1 \leq i \leq n, \quad \text{independientes entre sí.} \tag{45}$$

Weierstrass Approximation Theorem. Suppose f is a continuous real-valued function defined on the real interval $[a, b]$. For every $\varepsilon > 0$, there exists a polynomial p such that for all x in $[a, b]$, we have $|f(x) - p(x)| < \varepsilon$, or equivalently, the supremum norm $\|f - p\| < \varepsilon$.

Polinomio gtado 2

Polinomio gtado 6

Función a aproximar



From left to right, approximation of a continuous function (shown by the dashed black curve) over [0, 1], using $M = 2$ and $M = 6$ elements of (top row) polynomial, (middle row) Fourier, and (bottom row) single hidden layer neural network bases, respectively. While all three bases could approximate this function as finely as desired by increasing M, the neural network basis (with its adjustable internal parameters) approximates the underlying function more closely using the same number of basis elements compared to both fixed bases.

Polynomial Regression as an Alternative to Neural Nets

Xi Cheng
Department of Computer Science
University of California, Davis
Davis, CA 95616, USA
xicheng0821@gmail.com

Bohdan Khomtchouk
Department of Biology
Stanford University
Stanford, CA 94305, USA
bohdan@stanford.edu

Norman Matloff
Department of Computer Science
University of California, Davis
Davis, CA 95616, USA
matloff@cs.ucdavis.edu

Pete Mohanty
Department of Statistics
Stanford University
Stanford, CA 94305, USA
pmohanty@stanford.edu

April 11, 2019

Reza Borhani, Ph.D. Computer Science

4.4. Modelo de Regresión Lineal en notación matricial

$$
\begin{aligned}
Y_i &= \beta_0 X_{i0} + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_{p-1} X_{ip-1} + \varepsilon_i \\
&= \sum_{j=0}^{p-1} \beta_j X_{ij} + \varepsilon_i
\end{aligned}
$$

$$\underset{n\times 1}{\mathbf{Y}} = \underset{n\times p}{\mathbf{X}}\,\underset{p\times 1}{\boldsymbol{\beta}} + \underset{n\times 1}{\boldsymbol{\varepsilon}}$$

$$
\underset{n\times 1}{\mathbf{Y}} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \qquad
\underset{n\times p}{\mathbf{X}} = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1,p-1} \\ 1 & X_{21} & X_{22} & \cdots & X_{2,p-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{n,p-1} \end{bmatrix} \tag{46}
$$

$$
\underset{p\times 1}{\boldsymbol{\beta}} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix} \qquad
\underset{n\times 1}{\boldsymbol{\varepsilon}} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}
$$

Entonces tomando a las variables equis como fijas, o, lo que es lo mismo, condicional a las variables equis, la esperanza de $\mathbf{Y}$ resulta ser

$$E(\mathbf{Y} \mid \mathbf{X}) = \mathbf{X}\boldsymbol{\beta}$$

y la matriz de covarianza de las $\mathbf{Y}$ resulta ser la misma que la de $\boldsymbol{\varepsilon}$

$$Var(\mathbf{Y} \mid \mathbf{X}) = \sigma^2\mathbf{I}.$$

Al igual que hicimos con el modelo de regresión simple, muchas veces omitiremos la condicionalidad a las equis en la notación, es decir, como es bastante habitual en la literatura, escribiremos $E(\mathbf{Y})$ en vez de $E(\mathbf{Y} \mid \mathbf{X})$.

$$Var(\boldsymbol{\varepsilon}) = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2\mathbf{I}.$$

4.5. Estimación de los Parámetros (Ajuste del modelo)

Usamos el método de mínimos cuadrados para ajustar el modelo. O sea, definimos la siguiente función

$$g\left(b_0, b_1, \ldots, b_{p-1}\right) = \sum_{i=1}^{n} \left(Y_i - b_0 X_{i0} - b_1 X_{i1} - b_2 X_{i2} - \cdots - b_{p-1} X_{ip-1}\right)^2 \qquad (47)$$

Las ecuaciones de mínimos cuadrados normales para el modelo de regresión lineal general son

$$\mathbf{X}^t \mathbf{X} \widehat{\boldsymbol{\beta}} = \mathbf{X}^t \mathbf{Y}$$

donde $\mathbf{X}^t$ quiere decir la matriz traspuesta. Algunos autores lo notan $\mathbf{X}'$ (recordemos que la matriz traspuesta es aquella matriz $p \times n$ que tiene por filas a las columnas de $\mathbf{X}$). Los estimadores de mínimos cuadrados son

$$\underset{p \times 1}{\widehat{\boldsymbol{\beta}}} = \begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \\ \vdots \\ \widehat{\beta}_{p-1} \end{bmatrix}$$

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}^t \mathbf{X}\right)^{-1} \mathbf{X}^t \mathbf{Y}$$

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2},$$

$$\widehat{\beta}_0 = \overline{Y} - \widehat{\beta}_1 \overline{X}.$$

Los valores ajustados se calculan del siguiente modo

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}^t\mathbf{X}\right)^{-1}\mathbf{X}^t\mathbf{Y} \qquad \widehat{\mathbf{Y}} = \mathbf{X}\widehat{\boldsymbol{\beta}} = \mathbf{X}\left(\mathbf{X}^t\mathbf{X}\right)^{-1}\mathbf{X}^t\mathbf{Y}$$

que son los valores que están en la *superficie de respuesta ajustada* (o sea, en el plano ajustado en el caso $p = 3$). Los residuos se escriben matricialmente como

$$\begin{aligned} \mathbf{e} &= \mathbf{Y} - \widehat{\mathbf{Y}} = \mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}} \\ &= \mathbf{Y} - \mathbf{X}\left(\mathbf{X}^t\mathbf{X}\right)^{-1}\mathbf{X}^t\mathbf{Y} \\ &= \left(\mathbf{I} - \mathbf{X}\left(\mathbf{X}^t\mathbf{X}\right)^{-1}\mathbf{X}^t\right)\mathbf{Y} \end{aligned}$$

$$\widehat{Y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 X_i = \sum_{k=1}^{n} h_{ik} Y_k$$

$$h_{ik} = \frac{1}{n} + \frac{\left(X_i - \overline{X}\right)\left(X_k - \overline{X}\right)}{S_{XX}}$$

Llamando

$$\mathbf{H} = \mathbf{X}\left(\mathbf{X}^t\mathbf{X}\right)^{-1}\mathbf{X}^t \in \mathbb{R}^{n\times n}$$

a la "*hat matrix*" (la matriz que "sombrerea") tenemos que

$$\widehat{\mathbf{Y}} = \mathbf{H}\mathbf{Y} \qquad \mathbf{e} = \left(\mathbf{I} - \mathbf{H}\right)\mathbf{Y}.$$

4.8.2. Coeficiente de Determinación Múltiple (R^2 y R^2 ajustado)

$$R_a^2 = 1 - \frac{\frac{SSRes}{n-p}}{\frac{SSTo}{n-1}} = 1 - \left(\frac{n-1}{n-p}\right)\frac{SSRes}{SSTo}$$

$$|r| = \sqrt{R^2}$$

$$r = \frac{\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)\left(\widehat{Y}_i - \overline{\widehat{Y}}\right)}{\sqrt{\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2 \sum_{i=1}^{n}\left(\widehat{Y}_i - \overline{\widehat{Y}}\right)^2}}$$

4.8.3. Test F

Tabla 21: Tabla de ANOVA para el modelo de Regresión Lineal General

Fuente de variación	SS	g.l.	MS
Regresión	$\mathrm{SSReg} = \sum_{i=1}^{n} \left(\widehat{Y}_i - \overline{Y}\right)^2$	$p-1$	$\mathrm{MSReg} = \frac{\mathrm{SSReg}}{p-1}$
Residuos	$\mathrm{SSRes} = \sum_{i=1}^{n} \left(Y_i - \widehat{Y}_i\right)^2$	$n-p$	$\mathrm{MSRes} = \frac{\mathrm{SSRes}}{n-p}$
Total	$\mathrm{SSTo} = \sum_{i=1}^{n} \left(Y_i - \overline{Y}\right)^2$	$n-1$	

$H_0 : \beta_1 = \beta_2 = \cdots = \beta_{p-1} = 0$

$H_1 :$ no todos los β_k $(k = 1, 2, \ldots, p-1)$ son iguales a 0

$$F = \frac{\mathrm{MSReg}}{\mathrm{MSRes}} = \frac{\frac{\mathrm{SSReg}}{p-1}}{\frac{\mathrm{SSRes}}{n-p}} = \frac{\mathrm{SSReg}}{\mathrm{SSRes}} \frac{(n-p)}{(p-1)}. \qquad \sim \qquad F_{p-1,n-p,1-\alpha}$$

$$p - valor = P\left(F_{p-1,n-p} > F_{obs}\right)$$

4.9. Inferencias sobre los parámetros de la regresión

$$\boxed{E\left(\widehat{\beta}_k\right) = \beta_k.}$$

$$\boxed{Var\left(\widehat{\boldsymbol{\beta}}\right) = \sigma^2 \left(X^t X\right)^{-1}.}$$

$$\begin{aligned}
\widehat{\sigma}^2 &= \frac{1}{n-p}\sum_{i=1}^{n}\left(e_i - \overline{e}\right)^2 = \frac{1}{n-p}\sum_{i=1}^{n}\left(e_i\right)^2 \\
&= \frac{1}{n-p}\sum_{i=1}^{n}\left(Y_i - \widehat{Y}_i\right)^2 = \frac{\text{SSRes}}{n-p} \\
&= \text{MSRes}.
\end{aligned}$$

$$\widehat{Var}\left(\widehat{\boldsymbol{\beta}}\right) = \widehat{\sigma}^2 \left(X^t X\right)^{-1} = \text{MSRes}\left(X^t X\right)^{-1}.$$

$$\boxed{\frac{\widehat{\beta}_k - \beta_k}{\sqrt{\widehat{Var}\left(\widehat{\beta}_k\right)}} \sim t_{n-p} \text{ para } k = 0, 1, \ldots, p-1.}$$

Multi-colinealidad

$$Var\left(\widehat{\boldsymbol{\beta}}\right) = \sigma^2 \left(X^t X\right)^{-1}$$

$$C = (X'X)^{-1}$$

$$C_{jj} = \frac{1}{1 - R_j^2}.$$

Coeficiente de Determinación de X_j en función del resto de las covariables

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

¿ Qué nos dice esto ?

- En los métodos de Aprendizaje Supervisado (en particular en la Regresión Lineal Múltiple), las relaciones “aprendidas” entre la target (Y) y cualquier covariable (atributo) X está intermediada por las relaciones que existen entre las covariables en los datos.

$$\boxed{Var\left(\widehat{\beta}\right) = \sigma^2 \left(X^t X\right)^{-1}.}$$

4.10. Estimación de la Respuesta Media

$$E\left(Y_h\right) = E\left(Y_h \mid \mathbf{X}_h\right) = \mathbf{X}_h^t \boldsymbol{\beta}.$$

$$\widehat{Y}_h = \mathbf{X}_h^t \widehat{\boldsymbol{\beta}} = \widehat{\beta}_0 + \widehat{\beta}_1 X_{h1} + \widehat{\beta}_2 X_{h2} + \cdots + \widehat{\beta}_{p-1} X_{h,p-1}.$$

$$E\left(\widehat{Y}_h\right) = \mathbf{X}_h^t \boldsymbol{\beta} = E\left(Y_h\right)$$

$$Var\left(\widehat{Y}_h\right) = \sigma^2 \mathbf{X}_h^t \left(\mathbf{X}^t \mathbf{X}\right)^{-1} \mathbf{X}_h = \mathbf{X}_h^t Var\left(\widehat{\boldsymbol{\beta}}\right) \mathbf{X}_h.$$

$$\widehat{Var}\left(\widehat{Y}_h\right) = \mathrm{MSRes} \cdot \mathbf{X}_h^t \left(\mathbf{X}^t \mathbf{X}\right)^{-1} \mathbf{X}_h = \mathbf{X}_h^t \widehat{Var}\left(\widehat{\boldsymbol{\beta}}\right) \mathbf{X}_h.$$

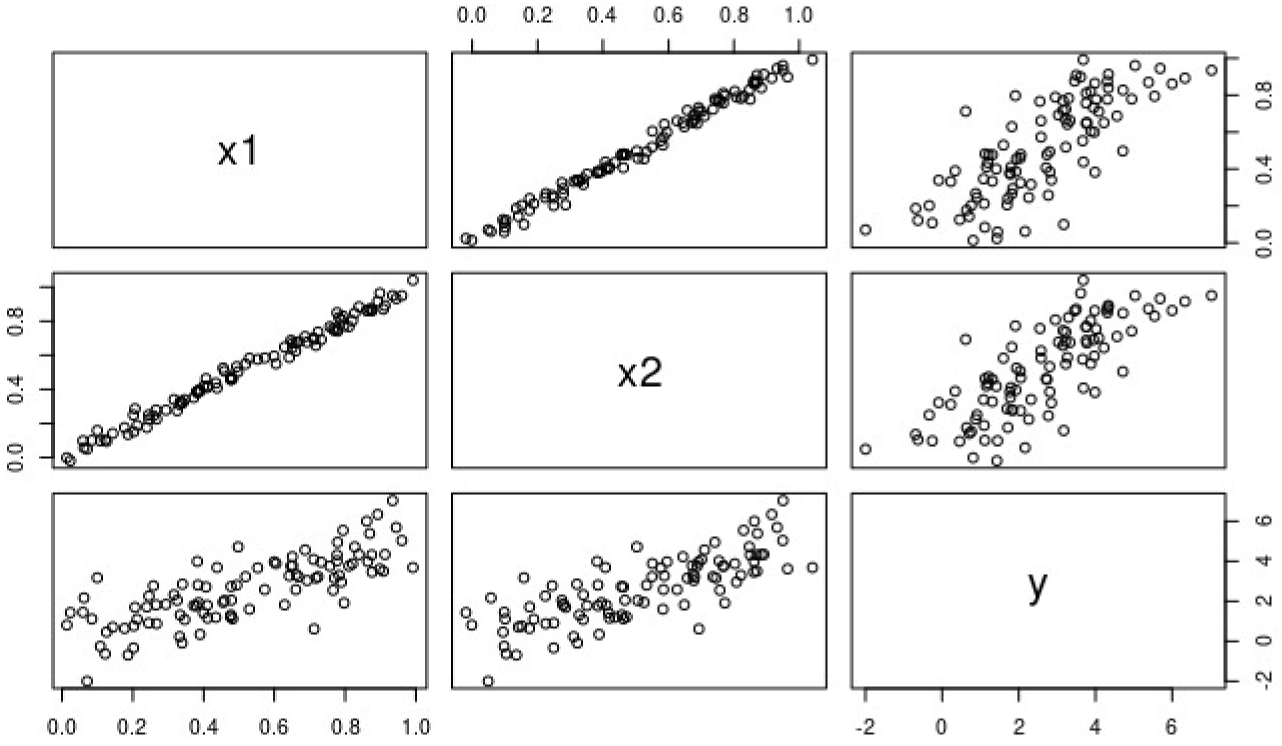
¿ Qué nos dice esto ?

Podemos encontrar situaciones (Datos) en los que la relación entre la target (Y) y un atributo esté mal estimada, pero la predicción de Y sea buena.

$$\boxed{Var\left(\widehat{Y}_h\right) = \sigma^2 \mathbf{X}_h^t \left(\mathbf{X}^t\mathbf{X}\right)^{-1} \mathbf{X}_h}$$

Ejemplo Sencillo

Sea Y dependiente de X1 y X2, ambas muy correlacionadas



4.10.2. Región de Confianza para la Superficie de Regresión

La región de confianza para toda la superficie de regresión es una extensión de la banda de confianza de Hotelling o Working-Hotelling para una recta de regresión (cuando hay una sola variable predictora). Los puntos de la frontera de la región de confianza en $\mathbf{X}_h$, se obtienen a partir de

$$\widehat{Y}_h \pm W \cdot \sqrt{\widehat{Var}\left(\widehat{Y}_h\right)}.$$

donde

$$W^2 = pF_{p,n-p;1-\alpha}. \tag{58}$$

Puede probarse que eligiendo este percentil, la región resultante cubrirá a la superficie de regresión **para todas las combinaciones posibles de las variables X** (dentro de los límites observados), con nivel $1-\alpha$. Es por eso que esta región de confianza tiene nivel simultáneo o global $1-\alpha$, como discutimos en la Sección 4.9.3.

4.11. Intervalos de Predicción para una Nueva Observación $Y_{h(\mathbf{nueva})}$

$$\begin{aligned} s^2(\text{pred}) &= \text{MSRes} + \widehat{Var}\left(\widehat{Y}_h\right) \\ &= \text{MSRes} \cdot \left(1 + \mathbf{X}_h' \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}_h\right), \end{aligned}$$

$$\widehat{Y}_h \pm t_{n-p,1-\alpha/2} \cdot s\left(\text{pred}\right)$$

$$\widehat{Y}_h \pm t_{n-p,1-\alpha/2} \cdot \sqrt{\text{MSRes} \cdot \left(1 + \mathbf{X}_h' \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}_h\right)}$$

4.16. Modelos con interacción entre variables cuantitativas y cualitativas

Como ya dijimos, cuando proponemos un modelo de regresión lineal múltiple del estilo de

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i, \tag{76}$$

$\beta_{1:2} = \beta_3$

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} \cdot X_{i2} + \varepsilon_i. \tag{77}$$

El modelo (77) es un caso particular del modelo de regresión lineal múltiple. Sea $X_{i3} = X_{i1} \cdot X_{i2}$ el producto entre las variables X_1 y X_2 medidas en el iésimo individuo, entonces el modelo (77) puede escribirse de la forma

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i,$$

$$
\begin{aligned}
Y &= \textit{pulso luego de correr una milla } (\texttt{Pulso2}) \\
X_1 &= \textit{pulso en reposo } (\texttt{Pulso1}) \\
X_2 &= \begin{cases} 1 & \textit{si la persona es mujer} \\ \\ 0 & \textit{en caso contrario} \end{cases}
\end{aligned}
$$

$$
\begin{aligned}
E\left(Y \mid X_1, X_2 = 1\right) &= \beta_0 + \beta_1 X_1 + \beta_2 \\
&= (\beta_0 + \beta_2) + \beta_1 X_1
\end{aligned}
$$

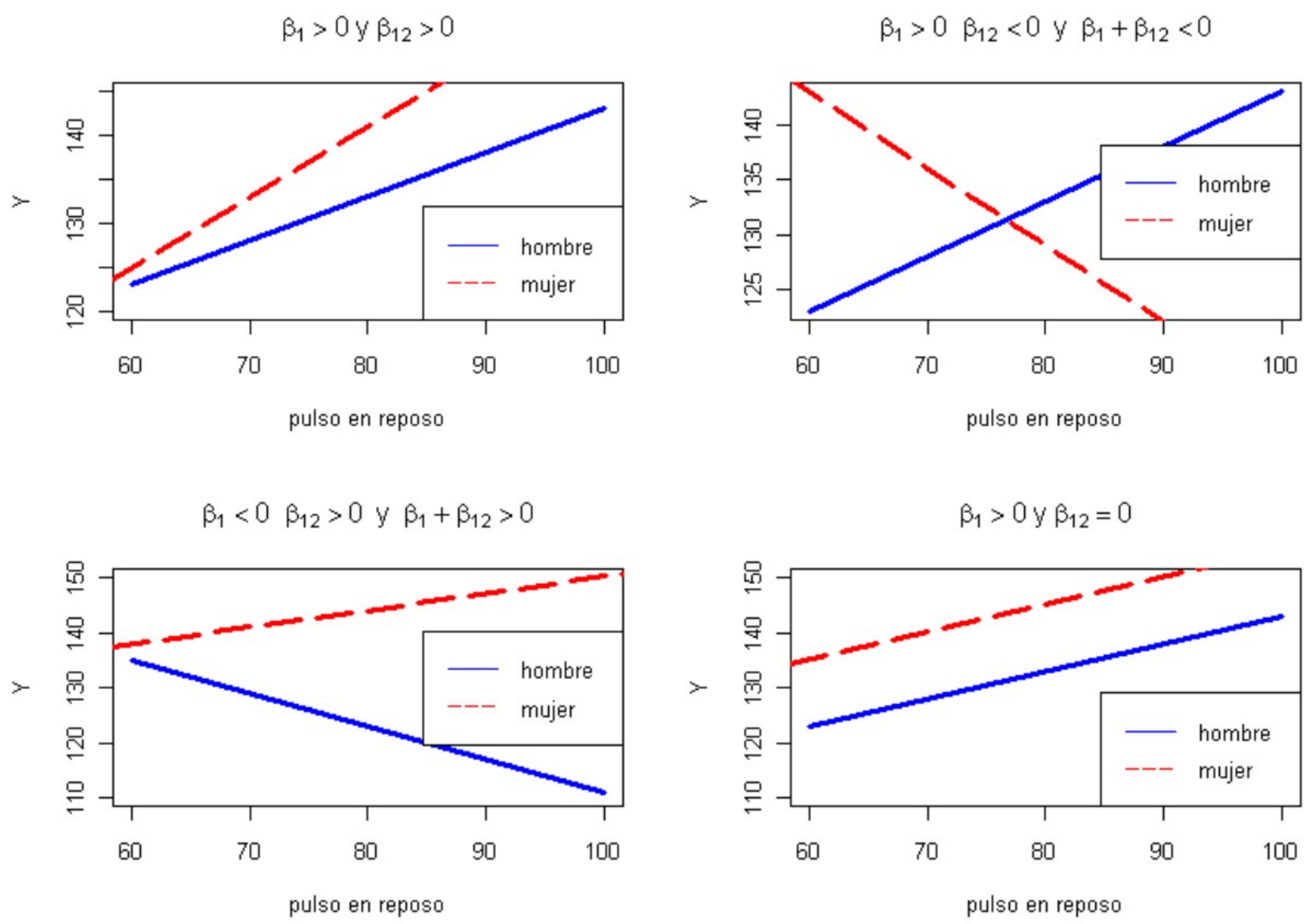
$$
Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i,
$$

$$
E\left(Y \mid X_1, X_2 = 0\right) = \beta_0 + \beta_1 X_1.
$$

$$
\begin{aligned}
E\left(Y \mid X\right) &= \beta_0 + \beta_1 X_1 + \beta_2 1 + \beta_{1:2} X_1 \cdot 1 \\
&= (\beta_0 + \beta_2) + (\beta_1 + \beta_{1:2}) X_1 \quad \text{mujeres}
\end{aligned}
$$

$$
Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_{1:2} X_{i1} \cdot X_{i2} + \varepsilon_i
$$

$$
\begin{aligned}
E\left(Y \mid X\right) &= \beta_0 + \beta_1 X_1 + \beta_2 0 + \beta_{1:2} X_1 \cdot 0 \\
&= \beta_0 + \beta_1 X_1 \quad \text{hombres}
\end{aligned}
$$

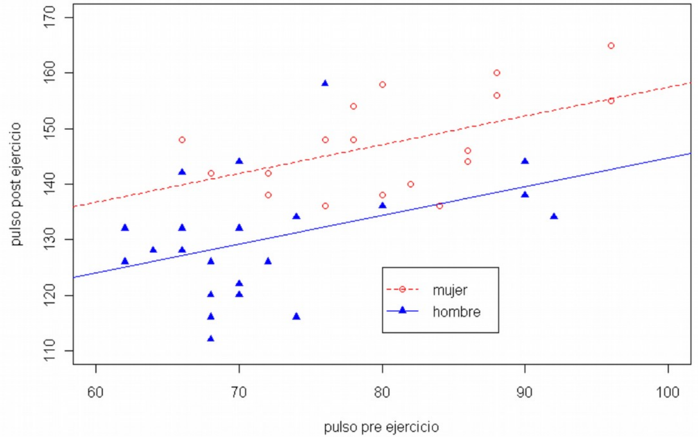

β1 > 0 y β12 > 0
β1 > 0 β12 < 0 y β1 + β12 < 0
β1 < 0 β12 > 0 y β1 + β12 > 0
β1 > 0 y β12 = 0
Y
pulso en reposo
hombre
mujer
60
70
80
90
100
110
120
125
130
135
140
150

```
> ajuste1<-lm(Pulso2~ Pulso1+mujer)
> summary(ajuste1)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  93.0970    12.5157   7.438 7.44e-09
Pulso1        0.5157     0.1715   3.007 0.004725
mujer        12.7494     3.2468   3.927 0.000361

Residual standard error: 9.107 on 37 degrees of freedom
Multiple R-squared: 0.5445, Adjusted R-squared: 0.5199
F-statistic: 22.12 on 2 and 37 DF,  p-value: 4.803e-07
```

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i,$$

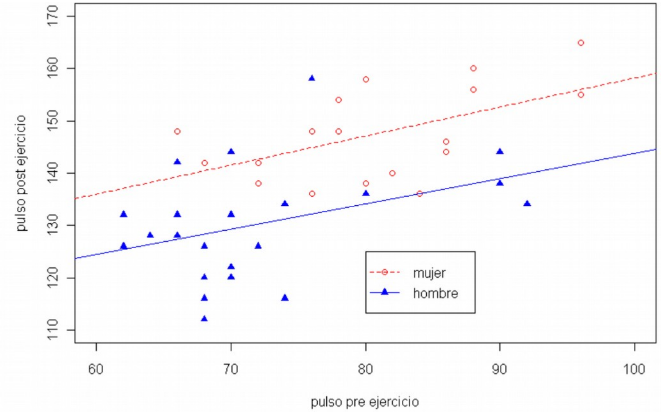



```
> ajuste2<-lm(Pulso2~ Pulso1 * mujer)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 95.42838   16.80929   5.677 1.88e-06
Pulso1       0.48334    0.23157   2.087    0.044
mujer        7.05575   27.14749   0.260    0.796
Pulso1:mujer 0.07402    0.35033   0.211    0.834

Residual standard error: 9.227 on 36 degrees of freedom
Multiple R-squared: 0.5451, Adjusted R-squared: 0.5072
F-statistic: 14.38 on 3 and 36 DF,  p-value: 2.565e-06
```

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_{1:2} X_{i1} \cdot X_{i2} + \varepsilon_i$$



5.4. Selección de modelos

Ya hemos observado que cuando tenemos K covariables disponibles y una variable a explicar Y, pueden, en principio, ajustarse 2^K modelos distintos. Decimos en principio, pues este total de modelos no incluye aquellos que tienen interacciones. En esta sección estamos pensando que si uno quiere evaluar ciertas interacciones, las debe incluir en esas K covariables iniciales. Lo mismo si uno quisiera evaluar algunas potencias de las covariables originales, o algunas transformaciones más complicadas de ellas. De este modo, cuando K es un número grande, la cantidad de modelos posibles crece exponencialmente, y evaluarlos uno por uno puede ser inmanejable. Por ejemplo, para $K = 8$, hay $2^8 = 256$ modelos posibles: hay un modelo sin covariables, 8 modelos de regresión lineal simple, cada uno con una sola covariable, $\binom{8}{2} = 28$ modelos con dos covariables $\{X_1, X_2\}$, $\{X_1, X_3\}$, $\{X_1, X_4\}$, $\{X_2, X_3\}$, etc. , $\binom{8}{3} = 56$ modelos con tres covariables, etcétera.
Lo que se denomina selección de modelos corresponde a la tarea de elegir el mejor modelo para nuestros datos.

5.4.1. Criterios para comparar modelos

- R_p^2 o $SS\text{Res}_p$: Un primer criterio para comparar modelos es mirar el R^2 obtenido con cada uno de ellos y elegir aquél con mayor R^2. Usamos el subíndice p para indicar la cantidad de parámetros $\beta's$ hay en el modelo (es decir, $p-1$ covariables). Como tenemos que

$$R_p^2 = 1 - \frac{SS\text{Res}_p}{SSTotal},$$

- $R_{a,p}^2$ o MSE_p : Como el R_p^2 no toma en cuenta el número de parámetros en el modelo de regresión, un criterio de decisión mucho más objetivo y automatizable es calcular y comparar modelos por medio del R_a^2. Lo subindicaremos como $R_{a,p}^2$ para indicar la cantidad de coeficientes $\beta's$ presentes en el modelo. Recordemos que

$$R_{a,p}^2 = 1 - \left(\frac{n-1}{n-p}\right)\frac{SS\text{Res}_p}{SSTotal} = 1 - \frac{MS\text{Res}_p}{\frac{SSTotal}{n-1}}.$$

- C_p *de Mallows*: Para utilizar esta medida hay que asumir que en el modelo con el total de las K covariables (el más grande posible) están todas las covariables importantes de modo que en ese modelo completo, la estimación de la varianza del error, σ^2, es insesgada. El valor del C_p se define por

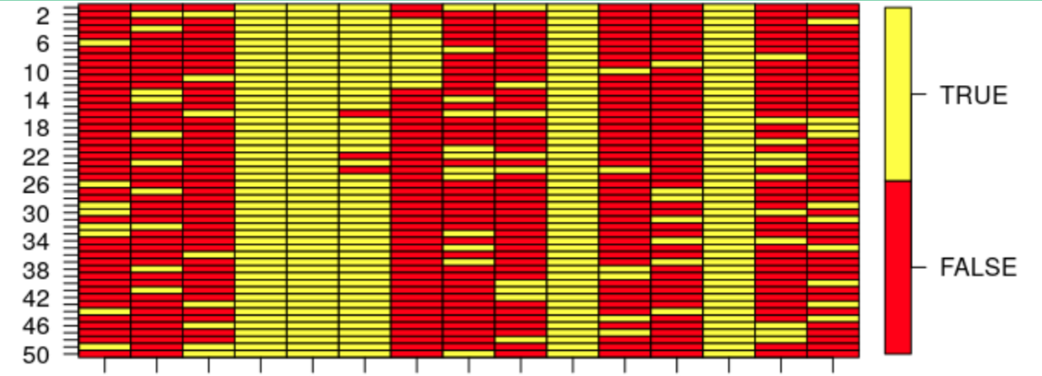
$$C_p = \frac{SS\text{Res}_p}{MS\text{Res}\,(X_1,\ldots,X_K)} - (n-2p)$$

- AIC_p, o el *Criterio de Akaike* y SBC_p o el *Criterio Bayesiano de Schwartz*, son otros dos criterios que, al igual que el C_p de Mallows, penalizan a los modelos con muchas covariables. Se buscan los modelos que tienen valores pequeños de AIC_p o SBC_p, donde estas cantidades están dadas por

$$\begin{aligned} AIC_p &= n\ln\left(SS\text{Res}_p\right) - n\ln\left(n\right) + 2p \\ SBC_p &= n\ln\left(SS\text{Res}_p\right) - n\ln\left(n\right) + p\ln\left(n\right) \end{aligned}$$

5.4.3. Selección automática de modelos

1. Todos los subconjuntos posibles (*Best subset*).
2. Eliminación *backward* (hacia atrás).
3. Selección *forward* (incorporando variables).
4. *Stepwise regression* (regresión de a pasos).



$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_{p-1} X_{ip-1} + \varepsilon_i,$$

Enfoques Alternativos

$$\hat{\bar{\beta}} = \frac{\sum_{i=1}^{R} w_i \hat{\beta}_i}{\sum_{i=1}^{R} w_i}$$

Multimodel Inference

where $\widehat{\beta}_i$ is the estimate for the predictor in a given model *i*, and w_i is the Akaike weight of that model. In this instance, $\widehat{\beta}_i$ is averaged only over the models for which the variate of interest appears. For instance, in the junco example in Table 2, $\widehat{\beta}_i$ and w_i values for PF would only be

Bayesian Model Averaging

$$\Pr(\Delta|D) = \sum_{k=1}^{K} \Pr(\Delta|M_k, D) \Pr(M_k|D)$$