
Introducción

Lo que sigue contiene una introducción muy concisa al análisis de regresión, concebida como apoyo de las clases. Hay varios niveles de lectura: en un primer nivel, las Observaciones que jalonan el texto pueden en su mayoría omitirse, sin pérdida de continuidad. Ello proporciona una lectura bastante lineal.

Si se desea una lectura más detallada, con digresiones que, no siendo imprescindibles, pueden mejorar la comprensión del conjunto, conviene leer tanto las observaciones como las secciones de COMPLEMENTOS Y EJERCICIOS al fin de cada capítulo: son parte integrante del texto a este segundo nivel y completan muchos detalles.

A lo largo del texto, tanto en demostraciones como en ejercicios o complementos se ha hecho uso abundante del símbolo de “giro peligroso” representado en el margen, popularizado por la obra clásica Knuth (1986). Se trata de fragmentos que corresponderían a un tercer nivel, con detalles de interés, extensiones de alguna idea, referencias a la literatura o ejercicios y demostraciones de mayor dificultad. La flecha vertical \uparrow remite a algún ejercicio, observación o ejemplo que son requisito previo.



Hay un mundo de diferencia entre saber *cómo se hacen* las cosas y *saber hacerlas*. Querriamos que los alumnos supieran *hacerlas*. La experiencia sugiere que lo que resulta de más ayuda al lector es ver ejemplos de aplicación detallados, que pueda reproducir o modificar para resolver sus propios problemas. Intercalados entre la teoría hay fragmentos en **R**, que el lector puede ejecutar o tomar como modelo. Todos se han ejecutado con **R** versión 2.8.1.

No se ha buscado el código más terso ni la forma más rápida o elegante de hacer las cosas, sino la que ilustra mejor la teoría.

Capítulo 1

El modelo de regresión lineal.

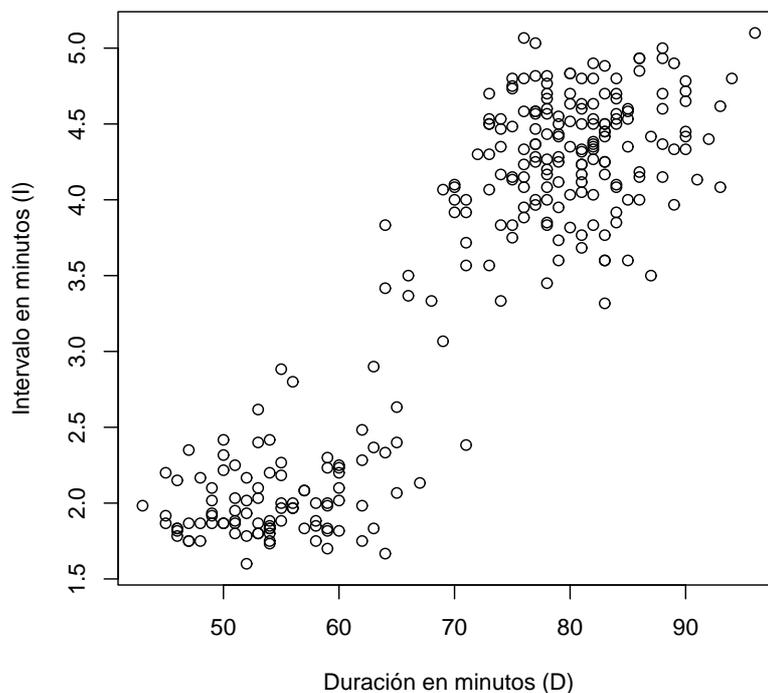
1.1. Planteamiento del problema.

Son frecuentes en la práctica situaciones en las que se cuenta con observaciones de diversas variables, y es razonable pensar en una relación entre ellas. El poder determinar si existe esta relación —y, en su caso, una forma funcional para la misma— es de sumo interés. Por una parte, ello permitiría, conocidos los valores de algunas variables, efectuar predicciones sobre los valores previsibles de otra. Podríamos también responder con criterio estadístico a cuestiones acerca de la relación de una variable sobre otra.

Ejemplo 1.1 La Figura 2.1 (pág. 4), muestra una gráfica recogiendo datos correspondientes a 272 erupciones del *geyser* Old Faithfull, en el Parque Nacional de Yellowstone (los datos proceden de Cook and Weisberg (1982)). En abscisas se representa la duración de las erupciones. En ordenadas, el intervalo de tiempo transcurrido hasta la siguiente erupción.

A la vista del gráfico, parece evidente que existe una relación entre ambas variables —erupciones de duración D corta son seguidas de otras tras un intervalo de tiempo I más reducido que en el caso de erupciones largas—. Podría interesarnos contrastar con criterio estadístico si tal relación existe (en el caso presente, la relación es tan nítida que el plantearse el contraste de hipótesis correspondiente no tendría demasiado sentido). Más interesante, en el caso presente, sería llegar a una expresión del tipo $I = f(D)$ relacionando el intervalo con

Figura 1.1: Old Faithful Geyser: datos de 272 erupciones.



la duración (ello nos permitiría anticipar en qué momento se presentará la siguiente erupción, conocida la duración D que se ha observado en la anterior).

Es claro que la relación $I = f(D)$ no puede ser exacta —es difícil pensar en una función que pase precisamente por cada uno de los 272 puntos en la Figura 2.1—. Habremos de considerar más bien funciones del tipo $I = f(D) + \epsilon$, en que el valor de I es una cierta función (desconocida) de D más una cantidad aleatoria inobservable ϵ . Decimos que $f(D)$ es una *función de regresión* de I sobre D , y nuestro objetivo es especificar su forma. Habitualmente realizamos para ello supuestos simplificadores, como el de que $f(D)$ es una función lineal.

FIN DEL EJEMPLO ■

Es de interés señalar que el ajuste de un modelo de regresión no se limita a analizar la relación entre dos variables; en general, buscaremos relaciones del tipo

$$Y = f(X_0, X_1, \dots, X_{p-1}) + \epsilon,$$

relacionando de manera aproximada los valores de Y con los que toman otras variables, X_0, \dots, X_{p-1} . Por simplicidad, limitaremos por el momento nuestra atención a funciones $f(X_0, \dots, X_{p-1})$ lineales; el modelo resultante es el modelo de regresión lineal, que se examina en la Sección 2.2 a continuación.

Señalemos, finalmente, que el hecho de aislar una variable Y al lado izquierdo y escribirla como función de otras más una perturbación aleatoria ϵ no prejuzga ninguna relación de causalidad en ningún sentido; sólo postulamos la existencia de una relación cuya forma y alcance queremos investigar. En el Ejemplo 2.1, el ajuste de un modelo del tipo $I = f(D) + \epsilon$ no implica que consideremos que la duración D *causa* el subsiguiente intervalo I hasta la próxima erupción, sino sólo que parece existir una relación entre ambas variables.

1.2. Notación

Consideramos una variable aleatoria Y (*regresando, respuesta, o variable endógena*) de la que suponemos que se genera así:

$$Y = \beta_0 X_0 + \beta_1 X_1 + \dots + \beta_{p-1} X_{p-1} + \epsilon, \quad (1.1)$$

siendo:

1. $\beta_0, \dots, \beta_{p-1}$, parámetros fijos desconocidos.
2. X_0, \dots, X_{p-1} , variables explicativas no estocásticas, *regresores*, cuyos valores son fijados por el experimentador. Frecuentemente X_0 toma el valor constante “uno”.
3. ϵ una variable aleatoria inobservable.

La ecuación (2.1) indica que la variable aleatoria Y se genera como combinación lineal de las variables explicativas, salvo en una perturbación aleatoria ϵ . En el Ejemplo 2.1, Y sería la variable I , y el único regresor sería la variable D . Si decidimos ajustar un modelo con término constante β_0 , tendríamos como regresores D y $X_0 = \text{“uno”}$. La función que aparece en (2.1) sería entonces $f(D) = \beta_0 + \beta_1 D$.

El problema que abordamos es el de estimar los parámetros desconocidos $\beta_0, \dots, \beta_{p-1}$. Para ello contamos con una muestra de N observaciones de

la variable aleatoria Y , y de los correspondientes valores de las variables explicativas X . Como se ha dicho, ϵ es inobservable. La muestra nos permitirá escribir N igualdades similares a (2.1):

$$\begin{aligned} y_1 &= \beta_0 x_{1,0} + \beta_1 x_{1,1} + \cdots + \beta_{p-1} x_{1,p-1} + \epsilon_1 \\ y_2 &= \beta_0 x_{2,0} + \beta_1 x_{2,1} + \cdots + \beta_{p-1} x_{2,p-1} + \epsilon_2 \\ &\vdots \\ y_N &= \beta_0 x_{N,0} + \beta_1 x_{N,1} + \cdots + \beta_{p-1} x_{N,p-1} + \epsilon_N. \end{aligned}$$

En forma matricial, escribiremos dichas N igualdades así:

$$\vec{y} = X\vec{\beta} + \vec{\epsilon}, \quad (1.2)$$

siendo:

- \vec{y} el vector $N \times 1$ de observaciones de la variable aleatoria Y ,
- X la matriz $N \times p$ de valores de las variables explicativas. Su elemento x_{ij} denota el valor que la j -ésima variable explicativa toma en la i -ésima observación,
- $\vec{\beta}$ el vector de parámetros $(\beta_0, \dots, \beta_{p-1})'$,
- $\vec{\epsilon}$ el vector $N \times 1$ de valores de la perturbación aleatoria ϵ .

Denotaremos mediante $\hat{\beta}$ al vector de estimadores de los parámetros, y por $\hat{\epsilon}$ al vector $N \times 1$ de residuos, definido por $\hat{\epsilon} = \vec{y} - X\hat{\beta}$; es decir, los residuos recogen la diferencia entre los valores muestrales observados y ajustados de la variable aleatoria Y .

Utilizamos minúsculas para designar valores muestrales y mayúsculas para las correspondientes variables aleatorias (así por ejemplo, \vec{y} denota el vector de valores observados de la variable aleatoria Y en una determinada experimentación). El contexto aclarará, por otra parte, cuando $\hat{\beta}$ y $\hat{\epsilon}$ son variables aleatorias o valores muestrales.

Adoptaremos para la estimación el *criterio mínimo cuadrático ordinario (MCO)*. Por consiguiente, diremos que $\hat{\beta}$ es óptimo si $\|\vec{y} - X\hat{\beta}\|^2$ es mínimo, denotando $\|\cdot\|$ la norma euclídea ordinaria:

$$\|\vec{y}\|^2 \stackrel{\text{def}}{=} \sum_i y_i^2$$

(ver Definición A.2, pág. 229).

Observación 1.1 El suponer que los valores de los regresores pueden ser fijados por el analista (apartado 2, al comienzo de esta Sección) nos coloca en una situación de *diseño experimental*. De ahí que a la matriz X se la denomine *matriz de diseño*.

Muchas veces (notablemente en Ciencias Sociales) no es posible fijar los valores de X , sino tan solo recolectar una muestra. Decimos entonces que estamos ante una *situación observacional* (en oposición a un diseño experimental). Ello no afecta a la teoría que sigue; la inferencia sobre los parámetros $\vec{\beta}$, etc. es entonces condicional a los valores observados de X .

Observación 1.2 El criterio de seleccionar como estimadores de $\vec{\beta}$ el vector $\hat{\beta}$ minimizando $\|\vec{y} - X\hat{\beta}\|^2$ es totalmente arbitrario. En lugar de minimizar la norma euclídea ordinaria, podríamos minimizar $\|\vec{y} - X\hat{\beta}\|_{L1}$ (suma de los valores absolutos de los errores de aproximación, también llamada *norma L1*), o cualquier otra cosa. Si se emplea la norma euclídea es por conveniencia matemática y por ser un criterio “razonable” desde diversos puntos de vista.

Observación 1.3 \diamond ¿Por qué introducir la norma euclídea y no limitarnos a proponer como criterio la minimización de

$$\sum_i \left(y_i - \hat{\beta}_0 x_{i0} - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_{p-1} x_{i,p-1} \right)^2?$$

Si realizamos las demostraciones en términos de normas, servirán *sea cual fuere la norma que adoptemos*. Muchos resultados serán así “todo terreno”, trasladables de inmediato a problemas con supuestos diferentes a los realizados en la Sección 2.3 a continuación. Veremos en breve (Observación 3.1, pág. 18) ventajas adicionales de plantear y resolver el problema en términos de aproximación vectorial, minimizando una norma.

1.3. Supuestos.

Además de suponer que $\vec{Y} = X\vec{\beta} + \vec{\epsilon}$ y que la matriz X es no aleatoria, requeriremos lo siguiente:

1. $E[\vec{\epsilon}] = \vec{0}$.
2. $E[\vec{\epsilon} \vec{\epsilon}'] = \sigma^2 I$.
3. $\text{rango}(X) = p < N$.

Nos referiremos a 1)–3) en lo sucesivo como los *supuestos habituales*.

El supuesto 1) no implica pérdida de generalidad ni supone ninguna restricción, al menos en el caso en que X tiene entre sus columnas una cuyos valores sean constantes (y ésto suele suceder; típicamente, la primera columna está formada por “unos”). En efecto, es claro que si:

$$\vec{Y} = \beta_0 \vec{1} + \beta_1 \vec{x}_1 + \cdots + \beta_{p-1} \vec{x}_{p-1} + \vec{\epsilon} \quad (1.3)$$

y el vector de perturbaciones verifica $E[\vec{\epsilon}] = \vec{\mu}$, entonces (2.3) puede reescribirse equivalentemente como:

$$\vec{Y} = (\beta_0 \vec{1} + \vec{\mu}) + \beta_1 \vec{x}_1 + \cdots + \beta_{p-1} \vec{x}_{p-1} + (\vec{\epsilon} - \vec{\mu}), \quad (1.4)$$

y (2.4) incorpora un vector de perturbaciones $(\vec{\epsilon} - \vec{\mu})$ verificando el primero de nuestros supuestos.

El supuesto 2), bastante más restrictivo, requiere que las perturbaciones sean incorrelacionadas (covarianzas cero) y homoscedásticas (de idéntica varianza).

El supuesto 3) simplemente fuerza la independencia lineal entre las (p) columnas de X . El requerimiento $N > p$ excluye de nuestra consideración el caso $N = p$, pues entonces $\vec{y} = X\hat{\beta}$ es un sistema de ecuaciones lineales determinado, y tiene siempre solución para algún vector $\hat{\beta}$ que hace los residuos nulos. Las estimaciones del vector $\vec{\beta}$ se obtendrían entonces resolviendo dicho sistema. Veremos en lo que sigue que este caso particular carece de interés (se dice que no tiene “grados de libertad”).

Algunos de los supuestos anteriores serán relajados, y las consecuencias que de ello se derivan estudiadas.

Observación 1.4 Nada impide que los regresores sean transformaciones adecuadas de las variables originales. Por ejemplo, si pensamos que la variable aleatoria Y depende del cuadrado de X_k y de otras variables, podríamos especificar un modelo de regresión así:

$$Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k^2 + \cdots + \beta_{p-1} x_{p-1} + \epsilon.$$

Análogamente, si pensáramos que la variable aleatoria W se genera del siguiente modo:

$$W = k z_1^{\beta_1} z_2^{\beta_2} \nu,$$

siendo ν una perturbación aleatoria no negativa (por ejemplo, con distribución logarítmico normal), nada impediría que tomáramos logaritmos para obtener

$$Y = \log(W) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon,$$

en que $x_i = \log(z_i)$, $\beta_0 = \log(k)$ y $\epsilon = \log(\nu)$. Lo que realmente se requiere es que la expresión de la variable endógena o regresando Y sea lineal *en los parámetros*.

1.4. La estimación mínimo cuadrática como problema de aproximación vectorial.

La ecuación matricial $\vec{y} = X\hat{\beta} + \hat{\epsilon}$ puede reescribirse así:

$$\vec{y} = \hat{\beta}_0 \vec{x}_0 + \cdots + \hat{\beta}_{p-1} \vec{x}_{p-1} + \hat{\epsilon}, \quad (1.5)$$

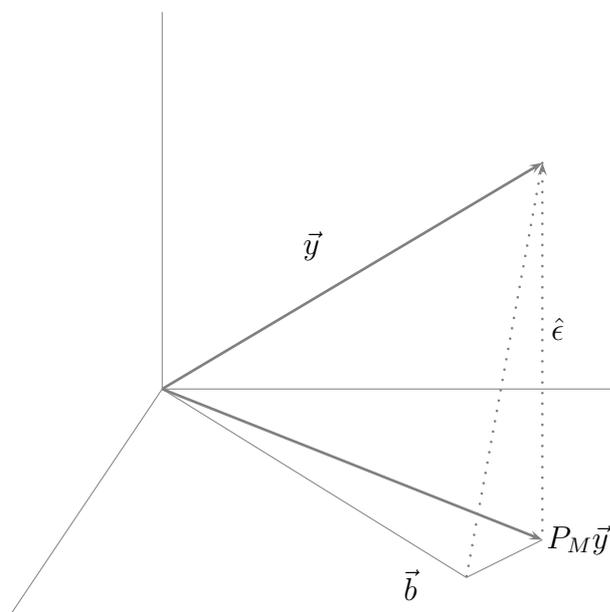
donde $\vec{x}_0, \dots, \vec{x}_{p-1}$ denotan los vectores columna de la matriz X (\vec{x}_0 será en general una columna de “unos”, como se ha indicado). Hay diferentes posibilidades en cuanto a criterio de estimación de los β . Si adoptamos el criterio MCO propuesto más arriba, consistente en minimizar $\|\hat{\epsilon}\|^2$, la ecuación (2.5) muestra que el problema puede reformularse así: ¿Cuales son los coeficientes $\hat{\beta}_0, \dots, \hat{\beta}_{p-1}$ que hacen que la combinación lineal $\hat{\beta}_0 \vec{x}_0 + \cdots + \hat{\beta}_{p-1} \vec{x}_{p-1}$ aproxime óptimamente (en sentido mínimo cuadrático) el vector \vec{y} ? Veremos inmediatamente que esta combinación lineal es lo que llamaremos *proyección* de \vec{y} sobre el subespacio generado por las columnas $\vec{x}_0, \dots, \vec{x}_{p-1}$.

1.5. Proyecciones.

Aunque en lo que sigue se hace un tratamiento generalizable, implícitamente consideramos productos internos (véase Definición A.1, pág. 229) real-valorados, lo que simplifica algunas fórmulas. Hacemos también un uso bastante tosco del lenguaje y notación, identificando vectores con matrices columna, operadores lineales y matrices asociadas a ellos, etc. Lo inadecuado del formalismo puede ser fácilmente suplido por el lector, y evita notación que podría hacerse agobiante.

Definición 1.1 *Sea H un espacio vectorial. Sea $M \subseteq H$ un subespacio del mismo, e $\vec{y} \in H$ un vector cualquiera. Decimos que \vec{u} es proyección de \vec{y} sobre M (y lo denotamos por $\vec{u} = P_M \vec{y}$) si:*

1. $\vec{u} \in M$,
2. $\vec{u} = \vec{y}$ si $\vec{y} \in M$,
3. $(\vec{y} - \vec{u}) \perp M$ si $\vec{y} \notin M$.

Figura 1.2: El vector $P_M\vec{y}$ es la proyección de \vec{y} sobre M (plano horizontal).

Siempre existe (y es única) la proyección de un vector en H sobre el subespacio M , tal como establece el teorema siguiente¹.

Teorema 1.1 *Sea H un espacio vectorial, y M un subespacio del mismo. Para cualquier vector $\vec{y} \in H$ existe siempre un único vector $\vec{u} = P_M\vec{y}$, proyección de \vec{y} sobre M . Se verifica que:*

$$\|\vec{y} - \vec{u}\|^2 = \min_{\vec{z} \in M} \|\vec{y} - \vec{z}\|^2. \quad (1.6)$$

La Fig. 2.2 ilustra en tres dimensiones la noción de proyección, y hace intuitivamente evidente el Teorema 2.1. En dicha figura se ha considerado $H = R^3$ y un subespacio M de dimensión dos representado como el plano horizontal. Consideremos $P_M\vec{y}$: podríamos describirlo como el obtenido al dejar caer una plomada desde el extremo de \vec{y} hasta hacer contacto con M .

Es claro que $\hat{\epsilon} = \vec{y} - P_M\vec{y}$ es ortogonal a M . Como consecuencia, para cualquier vector $\vec{b} \neq P_M\vec{y}$ en M , $\vec{y} - \vec{b}$ es la hipotenusa de un triángulo

¹Estrictamente incorrecto. El Teorema E.1, pág. 254 es una versión más elaborada del Teorema 2.1.

rectángulo, cuyos catetos son $\hat{\epsilon}$ y el segmento $\vec{b} - P_M\vec{y}$. Por tanto,

$$\|\vec{y} - \vec{b}\|^2 = \|\hat{\epsilon}\|^2 + \|\vec{b} - P_M\vec{y}\|^2 > \|\hat{\epsilon}\|^2$$

lo que demuestra la propiedad de $P_M\vec{y}$ de ser la mejor aproximación de \vec{y} en M . (Una demostración formal que va más allá de esta incompleta argumentación puede encontrarse en la Sección E.1, pág. 254.)

1.6. Lectura recomendada.

Sobre la teoría. Puede leerse como complemento a este capítulo Faraway (2005), Cap. 1 y Cap. 2, Sección 1 a 3, o los capítulos introductorios de la miríada de buenos textos que existe sobre regresión lineal: Seber (1977), Stapleton (1995), Arnold (1981), Draper and Smith (1998), Peña (2002), Myers (1990), Searle (1971), Ryan (1997) o Trocóniz (1987a) son algunos de ellos.

Sobre la utilización de R. El primero de los libros citados, Faraway (2005), ilustra también el modo de emplear R para hacer regresión; pero es demasiado escueto para servir de introducción al lenguaje. R es una implementación de fuente libre del lenguaje estadístico y gráfico S (ver por ejemplo Becker et al. (1988), Chambers and Hastie (1992) o Chambers (1998)). Los textos introductorios sobre S son por ello utilizables con R. Buenos manuales incluyen Venables and Ripley (1999a) (con su complemento específico para R, Venables and Ripley (1999b)), Dalgaard (2002), o Ugarte et al. (2008). Hay documentos con extensión de libro disponibles en Internet, como Maindonald (2000) o Kuhnert and Venables (2005).

COMPLEMENTOS Y EJERCICIOS

Algunos de los ejercicios que siguen requieren hacer uso de un ordenador y un programa especializado, tal como R. En la Sección 2.6, pág. 11, se proporcionan referencias.

1.1 En R para asignar un valor a una variable podemos colocarla a la izquierda del operador `<-`. Por ejemplo,

```
x <- 5
```

El valor de la variable puede ser utilizado en cálculos subsiguientes; tecleando

```
x + 5
```

obtendríamos “10”.

1.2 En R para crear un vector y asignarlo a la variable `x` haremos:

```
x <- c(1,3,4)
```

1.3 Para efectuar multitud de cálculos en R empleamos funciones. Por ejemplo, para sumar varios números y asignar el resultado a `x` podríamos escribir:

```
x <- 5 + 7 + 12
```

o también

```
x <- sum(c(5,7,12))
```

que hace uso de la función `sum`.

1.4 El producto interno euclídeo de dos vectores `x` e `y` puede calcularse así:

```
sum(x * y)
```

o alternativamente:

```
x %% y
```

1.5 En R rige la “regla del reciclado”, que permite operar con operandos disimilares. Por ejemplo, si:

```
a <- c(1,2,3)
b <- 5
```

entonces, tecleando

```
a + b
```

obtendríamos el vector $(6 \ 7 \ 8)'$. El argumento más corto, **b**, se ha usado repetidamente para construir un operando que pueda sumarse a **a**.

1.6 En R es muy fácil acceder a elementos aislados de un vector. Por ejemplo, si:

```
a <- c(6,7,8)
```

entonces, tecleando las expresiones que aparece a la izquierda obtendríamos los resultados que se indican a la derecha:

a	produce:	6 7 8
a[1]	produce:	6
a[1:2]	produce:	6 7
a[c(1,2)]	produce:	6 7
a[-1]	produce:	7 8
a[-(1:2)]	produce:	8
a[c(F,F,T)]	produce:	8
a[a>6]	produce:	7 8

Los subíndices se ponen entre corchetes, `[]`. Un subíndice negativo se interpreta como omitir el correspondiente valor. Además de subíndices numéricos, podemos emplear subíndices lógicos: **F** (falso) y **T** (cierto). Podemos incluso, como en la última línea, emplear expresiones que den como valor un vector lógico: `a > 6` produce el vector **F T T**, que empleado como subíndices retorna los elementos de **a** mayores que 6.

1.7 La función `help` permite interrogar a R sobre el modo de empleo de cualquier función. Por ejemplo, para obtener la descripción de `sum` podríamos teclear:

```
help(sum)
```

Empléese la función `help` para averiguar el cometido de las siguientes funciones de R: `t`, `cbind`, `rbind`, `solve`, `scan`, `read.table`, `list`, `nrow`, `ncol`. Obsérvese que tecleando

```
example(scan)
```

podemos ejecutar los ejemplos que aparecen en la documentación *on line* sin necesidad de retectarlos. Obsérvese también que el mandato `help.start()` abre una ventana de ayuda en un navegador —si es que hay alguno instalado en la máquina que empleamos—, lo que permite navegar cómodamente por la documentación.

1.8 Cuando escribimos expresiones como

```
sum(x * y)
```

estamos empleando funciones predefinidas (en este caso, `sum`). En R no necesitamos limitarnos a ellas; el lenguaje es extensible por el usuario. Podríamos definir una función `eucl` para realizar el producto interno así:

```
eucl <- function(x,y) { sum(x*y) }
```

que asigna a `eucl` la función especificada en el lado derecho. Para invocarla con los vectores `u` y `v`, teclearíamos: `eucl(u,v)`.

Una función puede emplearse como bloque constructivo de otras, y esto hasta el nivel de complejidad que se desee. La norma euclídea podría calcularse mediante una función definida así:

```
norma.eucl <- function(x) {
  sqrt(eucl(x,x)) }
```

que hace uso de `eucl` definida anteriormente. Tras esta definición, podemos calcular la norma euclídea de un vector `x` tecleando simplemente:

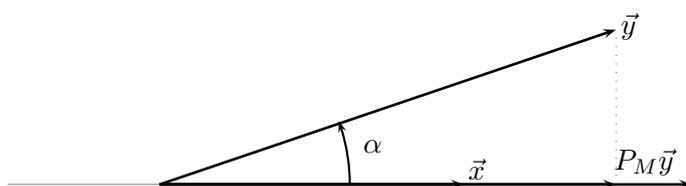
```
norma.eucl(x)
```

En realidad, la definición de una función como `eucl` es innecesaria: en R podemos emplear `x%%x` (o alternativamente `crossprod(x)`) que cumplen análogo cometido.

1.9 Recordemos que el producto euclídeo (o *escalar*) de dos vectores \vec{x}, \vec{y} en R^3 verifica:

$$\langle \vec{x}, \vec{y} \rangle = \|\vec{x}\| \|\vec{y}\| \cos(\alpha)$$

siendo α el ángulo que ambos vectores forman. Esta igualdad se extiende a R^N definiendo $\cos(\alpha)$ convenientemente (véase Definición A.3, pág. 229). Sea $P_M \vec{y}$ la proyección de \vec{y} sobre el subespacio M . Si $\|\vec{x}\| = 1$, del esquema a continuación inmediatamente se deduce que $\langle \vec{x}, \vec{y} \rangle = \|P_M \vec{y}\|$, siendo M el subespacio generado por \vec{x} .



Dedúzcase que, en el caso general en que $\|\vec{x}\| \neq 1$, se verifica:

$$P_M \vec{y} = \frac{\langle \vec{x}, \vec{y} \rangle}{\langle \vec{x}, \vec{x} \rangle} \vec{x}$$

1.10 Escribese una función que, dados dos vectores arbitrarios \vec{x} e \vec{y} , obtenga el vector proyección del segundo sobre el espacio (unidimensional) generado por el primero. Compruébese que el vector \vec{z} resultante es efectivamente la proyección buscada, para lo cual es preciso ver: i) Que \vec{z} es colineal con \vec{x} , y ii) Que $(\vec{y} - \vec{z}) \perp \vec{x}$.

1.11 Demuéstrese que los siguientes cuatro vectores de R^3 son un sistema generador de dicho espacio, pero no base.

$$\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$$

1.12 (\uparrow 2.11) Selecciónese, de entre los cuatro vectores indicados en el Problema 2.11, tres que formen base de R^3 .

1.13 (\uparrow 2.10) Los siguientes dos vectores generan un subespacio 2-dimensional de R^3 . Encuentrese —por ejemplo, mediante el procedimiento de Gram-Schmidt— una base ortonormal de dicho subespacio.

$$\begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 3 \\ 0 \end{pmatrix}$$

1.14 Demuéstrese que la correspondencia $P_M: \vec{x} \longrightarrow \vec{y} = P_M \vec{x}$ es una aplicación lineal.

1.15  La estimación de un modelo de regresión lineal realiza una aproximación del vector respuesta \vec{Y} similar a la que llevaría a cabo una red neuronal compuesta por una única neurona. “Similar” porque en el caso de una red neuronal la “estimación” (*entrenamiento o aprendizaje*) se realiza de ordinario mediante un proceso iterativo, cuyo resultado no necesariamente ha de coincidir exactamente con la estimación MCO. Un excelente manual sobre redes neuronales es Haykin (1998). Textos que tratan redes neuronales desde una perspectiva estadística son Ripley (1996) y Bishop (1996).

1.16  Hay alternativas a la regresión lineal: regresión no lineal y regresión no paramétrica (en que se considera una relación entre regresores y regresando que no está constreñida a ser lineal ni de ninguna otra forma funcional prefijada). En regresión no paramétrica se emplean principalmente tres métodos: *kernels*, vecinos más próximos y *splines*. Pueden consultarse, por ejemplo, Hastie et al. (2001) y Eubank (1988).

1.17  Como se ha indicado en la Observación 2.2, pág. 7, hay alternativas al criterio MCO. En lugar de minimizar la suma de cuadrados de los residuos, podríamos minimizar la suma de sus valores absolutos: $\sum_{i=1}^N |\hat{\epsilon}_i|$ (norma L1 del vector de residuos). Uno de sus atractivos es que los resultados resultan menos afectados por observaciones con residuo muy grande; pero es computacionalmente mucho más costosa.