2.3 La recta de mínims quadrats. Marques d'atletisme

Donats dos punts qualssevol del pla (x_1, y_1) i (x_2, y_2) sempre hi ha una recta $ax + by + c = 0$ que passa per ells. A més, si $x_1 \neq x_2$, l'equació de la recta es pot escriure com $y = mx + n$. Ara bé, és molt difícil, quan es pren una col·lecció de k punts al pla

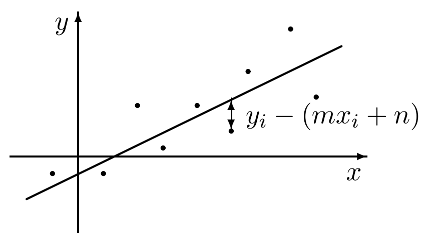
$$(x_1, y_1), (x_2, y_2), \ldots, (x_k, y_k),$$

que hi hagi una recta que passi per tots ells. En aquest context apareix la pregunta següent: quina és la recta $y = mx + n$ que aproxima millor tots els punts? Aquesta pregunta no està ben formulada matemàticament, ja que "millor"és un concepte no gaire clar.

Encara que no és l'única possibilitat, direm que la recta $y = \widehat{m}x + \widehat{n}$ és la que millor aproxima el núvol de punts (x_i, y_i), $i = 1, 2, \ldots, k$, si $\widehat{m}$ i $\widehat{n}$ són tals que la funció de dues variables

$$F(m,n) = (y_1 - (mx_1 + n))^2 + (y_2 - (mx_2 + n))^2 + \cdots$$
$$\cdots + (y_k - (mx_k + n))^2 = \sum_{i=1}^{k} (y_i - (mx_i + n))^2 \quad (*)$$

pren el seu valor més petit quan $m = \widehat{m}$ i $n = \widehat{n}$. A la figura següent s'il·lustra gràficament quins són els valors $y_i - (mx_i + n)$ que s'han d'elevar al quadrat i sumar per obtenir $F(m, n)$.



Una manera d'obtenir aquests valors $\widehat{m}$ i $\widehat{n}$ s'obté imposant condicions per tal que la funció F tingui un extrem. Ens limitarem aquí a donar el mètode per a calcular-los. El resultat és el següent:

Si prenem $\widehat{m}$ i $\widehat{n}$ com la solució del següent sistema lineal de dues equacions amb dues incògnites,

$$\begin{pmatrix} k & \sum_{i=1}^{k} x_i \\ \sum_{i=1}^{k} x_i & \sum_{i=1}^{k} x_i^2 \end{pmatrix} \begin{pmatrix} \widehat{n} \\ \widehat{m} \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{k} y_i \\ \sum_{i=1}^{k} x_i y_i \end{pmatrix},$$

aleshores la recta $y = \widehat{m}x + \widehat{n}$ és la recta que millor aproxima el conjunt de punts (x_i, y_i), $i = 1, 2, \ldots, k$ en el sentit que la funció $F(m, n)$ donada a $(*)$ pren el valor més petit possible quan $m = \widehat{m}$, $n = \widehat{n}$.

La recta $y = \widehat{m}x + \widehat{n}$ s'anomena recta de mínims quadrats o aproximació mínima quadràtica del conjunt de punts (x_i, y_i), $i = 1, \ldots, k$ o de la taula

x_1	x_2	x_3	$\ldots$	x_k
y_1	y_2	y_3	$\ldots$	y_k

La recta de mínims quadrats té múltiples utilitats a les ciències experimentals: física, química, biologia,... Nosaltres donarem aquí una aplicació a l'estudi de les marques d'atletisme. A la taula següent figuren els rècords mundials de la prova de 1500 m des del 1912 fins al 1995.

Any	min	s	Any	min	s	Any	min	s
1912	3	55.8	1943	3	45	1967	3	33.1
1917	3	54.7	1944	3	43	1974	3	32.2
1924	3	52.6	1947	3	43	1979	3	32.11
1926	3	51	1952	3	43	1980	3	31.36
1930	3	49.2	1954	3	41.8	1983	3	30.77
1933	3	49	1955	3	40.8	1985	3	29.46
1934	3	48.8	1956	3	40.5	1992	3	28.86
1936	3	47.8	1957	3	38.1	1995	3	27.37
1941	3	47.6	1958	3	36			
1942	3	45.8	1960	3	35.6			

Ens preguntem si, a partir d'aquestes dades, podem tenir una idea de quin podria ser el rècord mundial de la prova l'any 1999. Per a això fem els càlculs següents. Busquem quina és la recta $y = \widehat{m}x + \widehat{n}$ que millor aproxima els punts de la taula. Per tal de fer menys càlculs pendrem com a variables x_i els anys $12, 17, 24, \ldots, 95$, i com a marques només els segons $55.8, 54.7, 52.6, \ldots, 27.37$. Aleshores, els valors $\widehat{m}$ i $\widehat{n}$ seran la solució del sistema d'equacions:

$$\begin{pmatrix} 28 & 1476 \\ 1476 & 91768 \end{pmatrix} \begin{pmatrix} \widehat{n} \\ \widehat{m} \end{pmatrix} = \begin{pmatrix} 1153.53 \\ 55701.07 \end{pmatrix};$$

és a dir, $\widehat{m} = -0.3657$, $\widehat{n} = 60.4775$. Per tant, podem considerar com un possible rècord del món de 1999 la marca: 3 minuts més $-0.3657(99) + 60.4775$ segons, és a dir 3 min 24.27 s. Ara bé, si busquem quin és l'actual rècord del món, trobem que es va fer el juliol del 1998 i és de 3 min 26s. Per tant, hem obtingut una previsió que és bastant bona (tenint en compte, a més, que el rècord és de 1998 i, si haguéssim avaluat a $x = 98$, hauríem obtingut 3 min 24.64 s). Quina previsió de rècord del món podeu fer per a l'any 2010?

Per a acabar aquesta secció aplicarem la mateixa tècnica en una situació lleugerament diferent. A partir dels rècords del món actuals de 100 m, 200 m, 400 m, 800 m, 1500 m, 3000 m i 5000 m, intentarem deduir el de 10000 m. Fent els mateixos càlculs que a l'exemple anterior, a partir de les set primeres dades de la taula següent, obtenim que una aproximació del rècord del món de 10000 m fóra 25 min 22.44 s. Com podeu observar, el rècord real és de 26 min 22.75 s, és a dir que fem un error d'aproximadament 1 minut (de prop d'un 4%.) Cal comentar que no sembla gaire bon model el fet de pensar que el rècord depèn linealment de la longitud de la prova, i és clar que no seria gens bona idea usar aquestes dades per a predir el rècord del món de la marató, per exemple.

Prova	min	s
100 m	0	9.79
200 m	0	19.32
400 m	0	43.18
800 m	1	41.11
1500 m	3	26.00
3000 m	7	20.67
5000 m	12	39.36
10000 m	26	22.75

Representeu gràficament els resultats de la taula anterior. A partir de la seva gràfica potser s'us acudirà intentar aproximar-los per una corba diferent a la línia recta.

Si esteu interessats a tenir dades d'altres proves, podeu visitar l'adreça d'Internet

`http://www.algonet.se/~pela2/mtrack.htm`,

on trobareu les millors marques mundials de moltes proves d'atletisme.