6.2 Aproximaciones por mínimos cuadrados

En múltiples problemas de las ciencias biológicas, físicas y sociales resulta útil describir la relación entre las variables de los mismos por medio de una expresión matemática. Así, por ejemplo, se puede describir la relación entre el costo, el ingreso y la ganancia con la fórmula sencilla

$$P = R - C$$

En un contexto distinto, se puede representar la relación entre la aceleración debida a la gravedad, el tiempo que un objeto ha caído y la altura a la que estaba mediante la ley física

$$s = s_0 - v_0 t - \frac{1}{2} g t^2$$

donde s_0 es la altura inicial del objeto y v_0 es la velocidad inicial.

Por desgracia, no es fácil obtener fórmulas como las anteriores. Muy a menudo los científicos o los economistas tienen que trabajar con grandes cantidades de datos para encontrar relaciones entre las variables de un problema. Una manera común de hacer esto es ajustar una curva entre los distintos puntos de datos. Esta curva puede ser recta o cuadrática o cúbica, y así sucesivamente. El objetivo es encontrar la curva del tipo específico que se ajuste "mejor" a los datos dados. En esta sección se muestra cómo lograr esto cuando se tienen dos variables en el problema. En cada caso se supone que existen n puntos de datos $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$.

En la figura 6.3 se indican tres de las curvas que se pueden utilizar para ajustar datos.

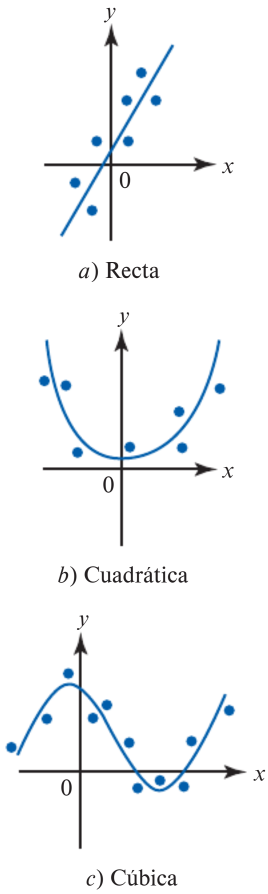

Figura 6.3
Tres curvas en el plano *xy*.

Aproximación por una recta

Antes de continuar, debe aclararse qué quiere decir "mejor ajuste". Suponga que se busca la recta de la forma $y = b + mx$ que mejor represente a los n datos $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$.

La figura 6.4 ilustra lo que ocurre (utilizando tres datos). En esta figura se ve que si se supone que las variables x y y están relacionadas por la fórmula $y = b + mx$, entonces, por ejemplo, para $x = x_1$ el valor correspondiente de y es $b + mx_1$. Esto es diferente del valor "real", $y = y_1$.

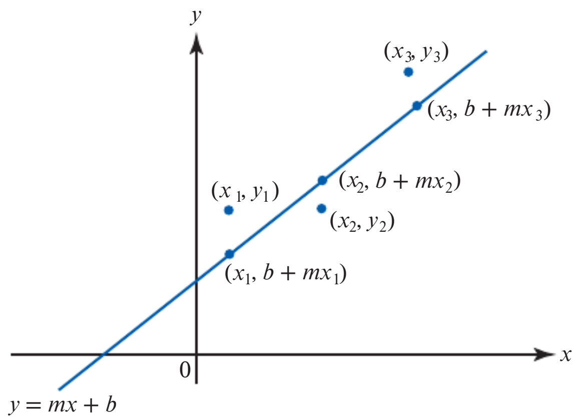

Figura 6.4
Los puntos sobre la recta tienen coordenadas $(x, b + mx)$.

En $\mathbb{R}^2$ la distancia entre los puntos (a_1, b_1) y (a_2, b_2) está dada por $d = \sqrt{(a_1 - a_2)^2 + (b_1 - b_2)^2}$. Por tanto, al determinar la manera de elegir la recta $y = b + mx$ que mejor se aproxima a los datos dados,