**65. Least squares and regression lines** When we try to fit a line $y = mx + b$ to a set of numerical data points $(x_1, y_1)$, $(x_2, y_2), \ldots, (x_n, y_n)$, we usually choose the line that minimizes the sum of the squares of the vertical distances from the points to the line. In theory, this means finding the values of $m$ and $b$ that minimize the value of the function

$$w = (mx_1 + b - y_1)^2 + \cdots + (mx_n + b - y_n)^2. \qquad (1)$$

(See the accompanying figure.) Show that the values of $m$ and $b$ that do this are

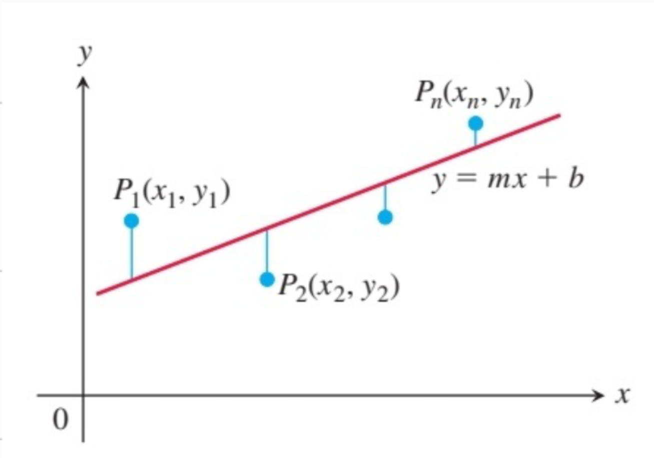
$$m = \frac{\left(\sum x_k\right)\left(\sum y_k\right) - n\sum x_k y_k}{\left(\sum x_k\right)^2 - n\sum x_k{}^2}, \qquad (2)$$

$$b = \frac{1}{n}\left(\sum y_k - m\sum x_k\right), \qquad (3)$$

with all sums running from $k = 1$ to $k = n$. Many scientific calculators have these formulas built in, enabling you to find $m$ and $b$ with only a few keystrokes after you have entered the data.

The line $y = mx + b$ determined by these values of $m$ and $b$ is called the **least squares line**, **regression line**, or **trend line** for the data under study. Finding a least squares line lets you

1. summarize data with a simple expression,
2. predict values of $y$ for other, experimentally untried values of $x$,
3. handle data analytically.




<pf> Observe that

(i) $\frac{\partial w}{\partial m} = \frac{\partial}{\partial m}\sum(mx_i+b-y_i)^2$

$= \sum 2x_i(mx_i+b-y_i)$

(ii) $\frac{\partial w}{\partial b} = \frac{\partial}{\partial b}\sum(mx_i+b-y_i)^2$

$= \sum 2(mx_i+b-y_i)$, so

if (i) = (ii) = 0, we have $b = \frac{1}{n}\sum y_i - mx_i$

and $\sum m x_i^2 + x_i b - x_i y_i = 0$

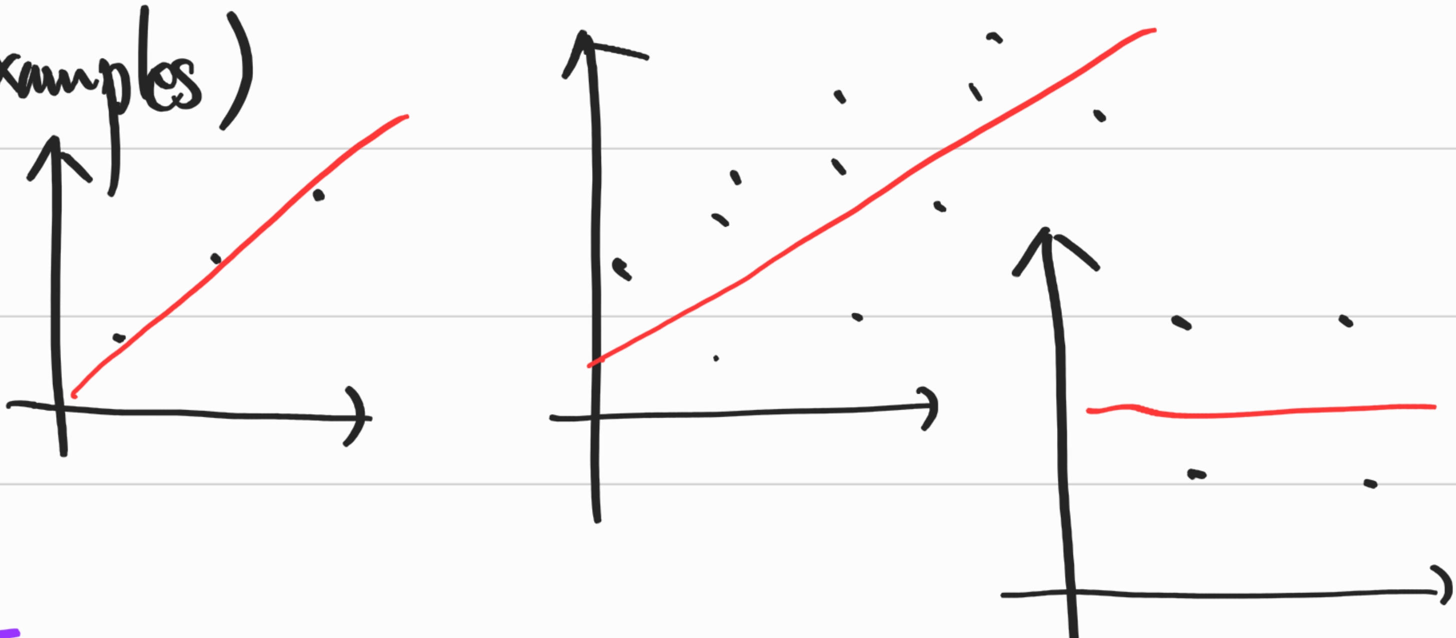
$$= m\sum x_i^2 + \frac{1}{n}\left[(\sum x_i)(\sum y_i) - m(\sum x_i)^2\right] - \sum x_i y_i = 0$$

$$\boxed{b = \frac{1}{n}(\sum y_i - m\sum x_i)}$$

$$\Rightarrow m\left(\sum x_i^2 - \frac{(\sum x_i)^2}{n}\right) = \sum x_i y_i - \frac{1}{n}(\sum x_i)(\sum y_i)$$

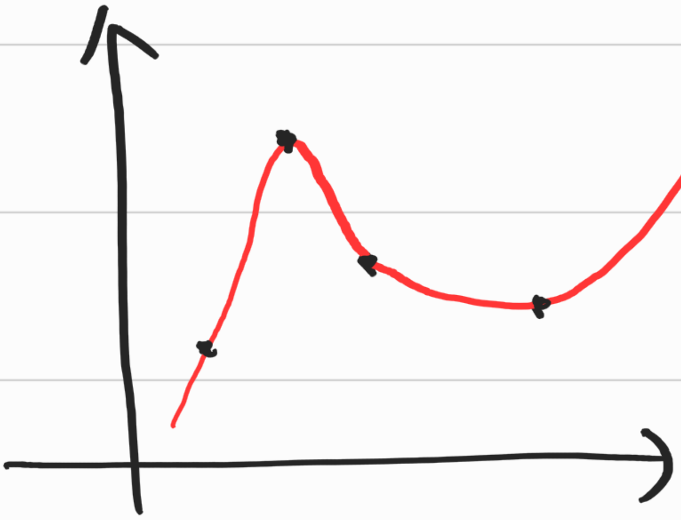
$$\Rightarrow m = \frac{\sum x_i \sum y_i - n\sum x_i y_i}{(\sum x_i)^2 - n\sum x_i^2},$$ as desired.

Examples)



The lines can predict some results.

<Remark> In another way, there is a polynomial interpolation. For example,



[Note that interpolation graph should pass through all given points.]

**48. Sum of products** Let $a_1, a_2, \ldots, a_n$ be $n$ positive numbers. Find the maximum of $\Sigma_{i=1}^{n} a_i x_i$ subject to the constraint $\Sigma_{i=1}^{n} x_i^2 = 1$.

<pf> Let $f(x_1, \cdots, x_n) = \sum a_i x_i$

$g(x_1, \cdots, x_n) = \sum x_i^2 - 1.$

Suppose $\nabla f = \lambda \nabla g.$

$\Downarrow$

$(a_1, \cdots, a_n) = \lambda(2x_1, \cdots, 2x_n)$

$\Leftrightarrow \quad x_i = \dfrac{a_i}{2\lambda} \quad i = 1, 2, \cdots, n.$

(note that if $\lambda = 0$, then $a_i$ would be 0)

Since $g\left(\frac{a_1}{2\lambda}, \cdots, \frac{a_n}{2\lambda}\right) = 0$, we have $\lambda = \pm\sqrt{\Sigma a_i^2}$ and hence $f\left(\frac{a_1}{\pm 2\sqrt{\Sigma a_i^2}}, \cdots, \frac{a_n}{\pm 2\sqrt{\Sigma a_i^2}}\right) = \pm\sqrt{\Sigma a_i^2}$.

Therefore, $\sqrt{\Sigma a_i^2}$ may be the maximum value of $f(x_1, \cdots, x_n)$

<Remark> $|\Sigma a_i x_i| \leq \sqrt{\Sigma a_i^2}\sqrt{\Sigma x_i^2}$

$= \sqrt{\Sigma a_i^2}$

Cauchy-Schwarz inequality.

**41. Converting to polar integrals**

**a.** The usual way to evaluate the improper integral $I = \int_0^\infty e^{-x^2}\,dx$ is first to calculate its square:

$$I^2 = \left(\int_0^\infty e^{-x^2}\,dx\right)\left(\int_0^\infty e^{-y^2}\,dy\right) = \int_0^\infty \int_0^\infty e^{-(x^2+y^2)}\,dx\,dy.$$

Evaluate the last integral using polar coordinates and solve the resulting equation for $I$.

**b.** Evaluate

$$\lim_{x\to\infty} \operatorname{erf}(x) = \lim_{x\to\infty} \int_0^x \frac{2e^{-t^2}}{\sqrt{\pi}}\,dt.$$

<pf> $\int_0^\infty \int_0^\infty e^{-(x^2+y^2)}\,dx\,dy$

$\overset{\uparrow}{=} \int_0^\infty \int_0^{\pi/2} e^{-r^2}\, r\, d\theta\, dr$

$x = r\cos\theta$

$y = r\sin\theta$

$dx\,dy = r\,dr\,d\theta$

$= \pi/2 \int_0^\infty r e^{-r^2}\, dr$

$\overset{\uparrow}{=} \pi/4 \int_0^\infty e^{-u}\, du$

$= \pi/4$

$r^2 = u$

$2r\,dr = du$

$\Rightarrow I = \int_0^\infty e^{-x^2}\, dx = \sqrt{\pi}/2.$

<Remark> So a normal distribution is in fact a probability distribution, namely, $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx = 2\int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \overset{\uparrow}{=} \frac{2}{\sqrt{\pi}} \int_0^\infty e^{-u^2}\, du = 1.$

$x = \sqrt{2}u$

$dx = \sqrt{2}\,du$

**11.** Evaluate the integral

$$\int_0^{\infty} \frac{e^{-ax} - e^{-bx}}{x} dx.$$

<pf> $\displaystyle\int_0^{\infty} \frac{e^{-ax} - e^{-bx}}{x} dx$

$\displaystyle = \int_0^{\infty} \int_a^{b} e^{-xy} \, dy \, dx$

$\displaystyle = \int_a^{b} \int_0^{\infty} e^{-xy} \, dx \, dy$

$\displaystyle = \int_a^{b} -\frac{1}{y} e^{-xy} \Big|_0^{\infty} dy$

$\displaystyle = \int_a^{b} \frac{1}{y} dy = \ln b - \ln a.$

**23. The value of $\Gamma(1/2)$** The gamma function,

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t}\, dt,$$

extends the factorial function from the nonnegative integers to other real values. Of particular interest in the theory of differential equations is the number

$$\Gamma\left(\frac{1}{2}\right) = \int_0^\infty t^{(1/2)-1} e^{-t}\, dt = \int_0^\infty \frac{e^{-t}}{\sqrt{t}}\, dt. \tag{2}$$

**a.** If you have not yet done Exercise 41 in Section 15.4, do it now to show that

$$I = \int_0^\infty e^{-y^2}\, dy = \frac{\sqrt{\pi}}{2}.$$

**b.** Substitute $y = \sqrt{t}$ in Equation (2) to show that $\Gamma(1/2) = 2I = \sqrt{\pi}$.

<pf> $\int_0^\infty e^{-t}/\sqrt{t}\, dt \underset{\uparrow}{=} \int_0^\infty e^{-y^2}/y \cdot 2y\, dy$

$y = \sqrt{t}$

$dy = \dfrac{dt}{2\sqrt{t}}$

$\Rightarrow 2y\, dy = dt$

$= 2\int_0^\infty e^{-y^2} dy = 2I = \sqrt{\pi}.$

Hence $\Gamma(1/2) = \sqrt{\pi}.$