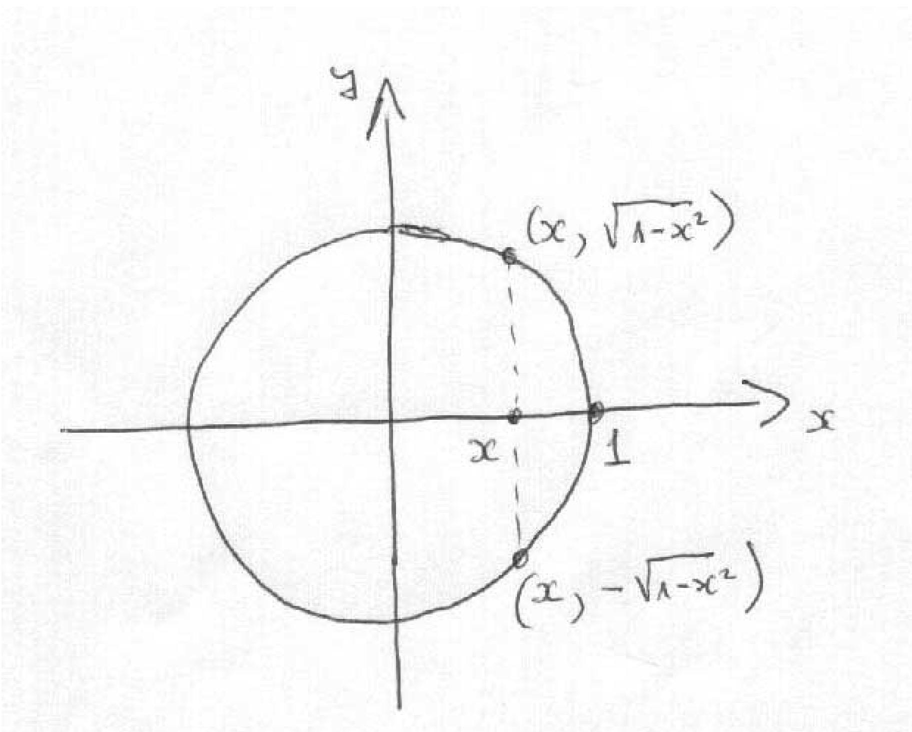
Metoda najmanjih kvadrata

Inženjerski problem.
Predpostavimo da imamo dvije veličine: x i y (na primjer, masa i temperatura, tlak i obujam, prva i druga koordinata točke u ravnini itd.).
Tada postoje dvije mogućnosti. Prva je da su te veličine (u određenim uvjetima) nezavisne, a druga je da su zavisne.
Nezavisnost znači da uz svaku očitanu vrijednost veličine x teoretski možemo očitati bilo koju vrijednost veličine y (takve su, na primjer, masa zlata i temperatura zlata pri uobičajenim masama i temperaturama; očitana masa zlata ne daje nam nikakvu informaciju o temperaturi).
Zavisnost pak znači da očitana vrijednost veličine x potpuno određuje (ili ograničuje) vrijednost veličine y.
Na primjer, ako se točka giba po kružnici u koordinatnoj ravnini, onda očitana vrijednost varijable x općenito uvjetuje dvije moguće vrijednosti varijable y.
Da bi bili još konkretniji zamislimo da se točka giba po jediničnoj kružnici sa središtem u ishodištu. Predpostavimo da smo očitali prvu koordinatu i da smo dobili x=0.8.
Tada postoje dvije mogućnosti za vrijednost veličine y: 0.6 i -0.6.
Općenito, u ovim okolnostima veličina x može poprimiti bilo koju vrijednost između -1 i 1 (uključujući i njih). Isto je s veličinom y. Međutim, te su dvije veličine povezane relacijom:

$x^2 + y^2 = 1$ (jednadžba kružnice).

Zato je $y = \pm\sqrt{1 - x^2}$ pa svakoj vrijednosti veličine x odgovaraju dvije vrijednosti veličine y (osim ako je x=1 ili x=-1; tada je y=0).



U inženjerstvu je važan slučaj kad **svaka vrijednost veličine x, uvjetuje točno jednu vrijednost veličine y**. To se kraće zapisuje pomoću funkcija:

$y = f(x)$,

gdje je f pravilo prema kojemu y ovisi o x.

Najjednostavnija pravila zavisnosti (funkcije)

1. linearna zavisnost y = ax + b (to je f(x):= ax+b, a grafički je prikaz te zavisnosti pravac)

2. kvadratna zavisnost y = ax^2+bx+c (grafički prikaz je parabola) .

3. kubna zavisnost y = ax^3+bx^2+cx+d

4. recipročna zavisnost (obrnuta proporcionalnost) y = a/x (grafički prikaz je hiperbola)

4. eksponencijalna zavisnost $y = ae^{bx}$

5. potencijska zavisnost $y=a\cdot b^x$

itd.

Uočite da ima više linearnih zavisnosti (kvadratnih zavisnosti i sl.). Linearna je zavisnost poznata (određena) onda ako znamo realne brojeve a,b (koeficijente) itd. Da naglasimo kako linearna funkcija zavisi o svojim koeficijentima pišemo:
f(x,a,b):= ax + b.

Slično, za kvadratnu funkciju:
f(x,a,b,c):=ax^2+bx+c.

Brojeve a,b odnosno a,b,c zovemo i **parametrima**. To je i općenito, a ne samo za linearne ili kvadratne veze. Na primjer,
$f(x,a,b) := a \cdot e^{bx}$
je familija funkcijskih veza (eksponencijalnog tipa) ovisna o parametrima a,b.

Zamislimo pokus u kojemu mijenjamo po volji **obujam** plina, a pri svakoj konkretnoj vrijednosti obujma očitavamo **tlak**. Tu je uobičajeno da veličina x bude obujam (nezavisna veličina, veličina koju po volji mijenjamo) a da veličina y bude tlak (zavisna veličina, veličina koju dobijemo mjerenjem).

Primjer 1. Predpostavimo da smo mijenjali veličinu x i pri tom mjerili odgovarajuće vrijednosti veličine y. Rezultate zapišimo u obliku tablice.

x	1	2	3	4	5	6
y	2.1	5.1	8.2	11.0	14.5	17.4

Iz prvog pogleda na tablicu uočavamo da se povećavanjem veličine x, povećava i veličina y. Nas zanima pravilo prema kojem se to događa. Brzo uočavamo da smo veličinu x povećavali za jednu mjernu jedinicu. Pri tom se veličina y povećavala redom za:

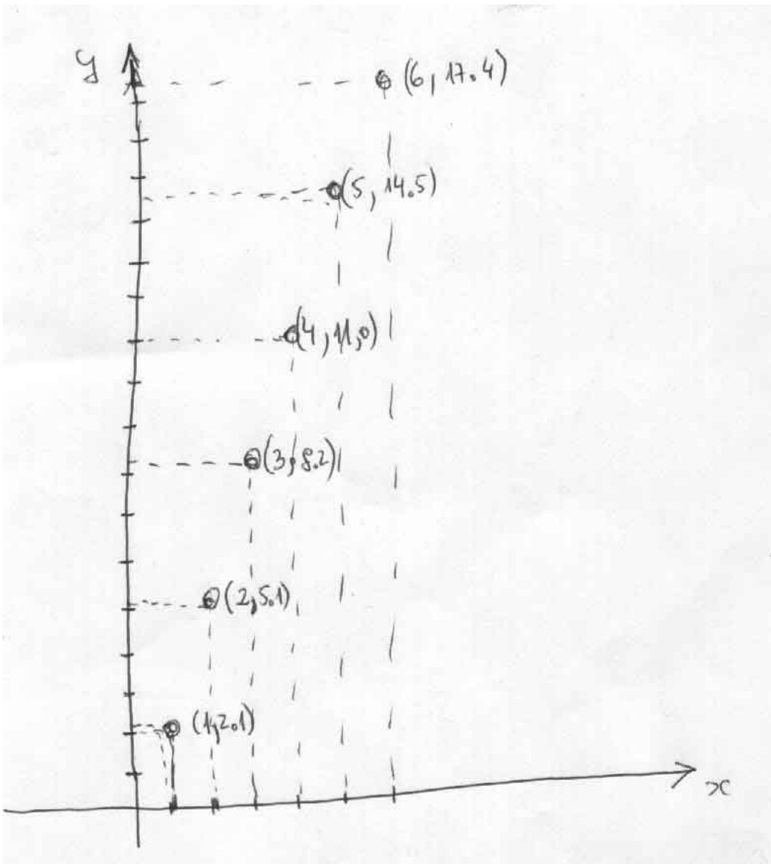
3.0 pa za 3.1 pa za 2.8 pa za 2.5 pa za 2.9.
(vidimo da su te promjene, iako različite, ipak prilično bliske).

Kad god pri jednakim promjenama jedne veličine uočimo odprilike jednake promjene druge veličine, moramo posumljati u linearnu vezu među njima. Dakle:

y = ax + b .

Drugi je način uočavanja možebitne linearne veze grafički. U tu svrhu podatke zapišimo u obliku uređenih parova i ucrtajmo ih u koordinatni sustav kojemu je x horizontalna os, a y vertikalna:

(1; 2.1) , (2; 5.1), (3; 8.2), (4; 11.0), (5; 14.5), (6; 17.4)



Uočavamo da su točke odprilike na pravcu. Grafički uvid u oblik zavisnosti možemo provesti i onda ako razmaci među vrijednostima veličine x nisu jednaki (nisu ekvidistantni), dok je takav uvid računanjem promjene veličine y otežan (iako se i on može provesti).

Sjetite se da su za crtanje pravca potrebne samo dvije točke, a mi ih imamo 5. Lako se vidi da ne postoji pravac koji prolazi kroz sve te točke. Zato od svih pravaca treba izabrati onaj koji *najbolje* prolazi pokraj tih točaka. Kriterij odabiranja tog pravca, tj. pripadajućih koeficijenata a,b daje *metoda najmanjih kvadrata*. Tako je i općenito, samo što je onda kad uočena zavisnost odudara od linearne, otežano biranje tipa zavisnosti. Često to i nije problem. Naime tip veze je u mnogim primjerima predviđen pripadajućim teorijama. Na primjer, veza između obujma i tlaka plina pri fiksiranoj temperaturi predviđena je u obliku Van-der- Wallsove jednadžbe, Peng-Robinsonove jednadžbe i sl. (a mi samo moramo odrediti parametre koji se pojavljuju u tim jednadžbama).

Općenito imamo n vrijednosti veličine x i n vrijednosti zavisne veličine y. To zadajemo tablicom:

x| x_1 x_2 .... x_n

y| y_1 y_2 ..... y_n

Načelo na kojemu se zasniva metoda najmanjih kvadrata.

Metoda najmanjih kvadrata zasniva se na načelu da su najbolji oni pararametri a,b za koje je suma kvadrata razlika između mjerenih vrijednosti y_i, i=1,2,...,n i izračunatih vrijednosti $f(x_i, a, b)$ minimalna.

Ima više matematičkih razloga za prihvaćanje ovog načela, a tu ih nećemo spominjati. Napomenimo da nije dobro razmatrati zbroj razlika eksperimentalnih i teoretskih podataka jer se pozitivne i negativne razlike (**odstupanja**) poništavaju. Da bi uzeli u obzir i pozitivna i negativna odstupanja, matematičari su na početku razmatrali apsolutne vrijednosti razlika i tražili da njihova suma bude minimalna. To nije loš kriterij, ali su apsolutne vrijednosti nepogodne jer se ne mogu općenito derivirati. Taj, ali i neki drugi razlozi, prevagnuli su u korist sume kvadrata.

Postupak odredjivanja parametara metodom najmanjih kvadrata.

Označimo i-to odstupanje kao $D_i := y_i - f(x_i, a, b)$

To je razlika izmedju mjerene (eksperimantalne) vrijednosti y_i i teoretske vrijednosti $f(x_i, a, b)$, tj. vrijednosti funkcije f(x,a,b) za $x=x_i$.

Prema metodi najmanjih kvadrata, parametre odredjujemo tako da suma

$$D_1^2 + D_2^2 + \ldots + D_n^2$$

bude minimalna.

Taj izraz ovisi o nepoznatim parametrima a, b. Zato pišemo:

$$F(a,b) := [y_1 - f(x_1, a, b)]^2 + [y_2 - f(x_2, a, b)]^2 + \ldots + [y_n - f(x_n, a, b)]^2$$

ili, kraće

$$F(a,b) := \sum_{i=1}^{n} [y_i - f(x_i, a, b)]^2$$

(*F* ovisi i o vrijednostima x_i, y_i, za i=1,2,...n, međutim te su vrijednosti poznate). Funkcija F često se naziva **funkcija cilja** (općenito, funkcija cilja može biti i neka druga pogodna funkcija, na primjer, odstupanja se mogu množiti nekim težinama).

Treba odrediti parametre a,b u kojima funkcija cilja *F* postiže minimum.

Uvjeti lokalnog ekstrema (pomoću parcijalnih derivacija) za funkciju F jesu:

$$\frac{\partial F}{\partial a}=0 \quad \text{i} \quad \frac{\partial F}{\partial b}=0$$

Dakle

$$\frac{\partial(\sum_{i=1}^{n}[y_i - f(x_i,a,b)]^2)}{\partial a}=0 \quad \text{i} \quad \frac{\partial(\sum_{i=1}^{n}[y_i - f(x_i,a,b)]^2)}{\partial b}=0$$

Koristeći svojstva derivacija (derivacija zbroja je zbroj derivacija) dobijemo:

$$\sum_{i=1}^{n}2[y_i - f(x_i,a,b)]\cdot(-\frac{\partial f(x_i,a,b)}{\partial a})=0 \quad \text{i} \quad \sum_{i}2[y_i - f(x_i,a,b)]\cdot(-\frac{\partial f(x_i,a,b)}{\partial a})=0$$

Nakon sređivanja dobijemo sustav dviju jednadžba s dvjema nepoznanicama a,b.

$$\sum_{i=1}^{n}[y_i - f(x_i,a,b)]\cdot\frac{\partial f(x_i,a,b)}{\partial a}=0$$

$$\sum_{i=1}^{n}[y_i - f(x_i,a,b)]\cdot\frac{\partial f(x_i,a,b)}{\partial b}=0 \quad (*)$$

Iz tog sustava odredjujemo nepoznate parametre a,b . Općenito sustav može imati više rješenja. Takodjer može se dogoditi da neka rješenja odgovaraju maksimumu ili sedlastoj točki, a ne minimumu. Može se dogoditi i to da neka rješenja nemaju fizikalna značenja. Na sreću, u najvažnijem slučaju, slučaju linearnih veza i njima srodnih, rješenje tog sustava je jedinstveno, tj. parametri se mogu odrediti jednoznačno.

Linearna regresija.
Određivanje parametara a,b za linearnu vezu (određivanje regresijskog pravca).

Tu je f(x,a,b):=ax+b, pa je

f(x_i,a,b) = ax_i+b

$$\frac{\partial f(x_i,a,b)}{\partial a}=x_i \quad \text{i} \quad \frac{\partial f(x_i,a,b)}{\partial b}=1$$

Ako to uvrstimo u sustav (*), dobijemo

$$\sum_{i=1}^{n}(y_i - ax_i - b)\cdot x_i = 0$$

$$\sum_{i=1}^{n}(y_i - ax_i - b)\cdot 1 = 0$$

nakon raspisivanja dobijemo sustav dviju linearnih jednadžba s dvjema nepoznanicama:

$$(\sum_{i=1}^{n}x_i^2)\cdot a + (\sum_{i=1}^{n}x_i)\cdot b = \sum_{i=1}^{n}x_i\, y_i$$

$$(\sum_{i=1}^{n}x_i)\cdot a + n\cdot b = \sum_{i=1}^{n}y_i$$

U tom su sustavu parametri a,b nepoznanice, a brojevi
$\sum_{i=1}^{n} x_i^2$, $\sum_{i=1}^{n} x_i$, n, $\sum_{i=1}^{n} x_i\, y_i$, $\sum_{i=1}^{n} x_i$ **koeficijenti sustava** (oni su poznati, jer se dobiju iz mjernih podataka).
Rješavanjem tog linearnog sustava dobijemo konačne formule (indekse ispuštamo!):

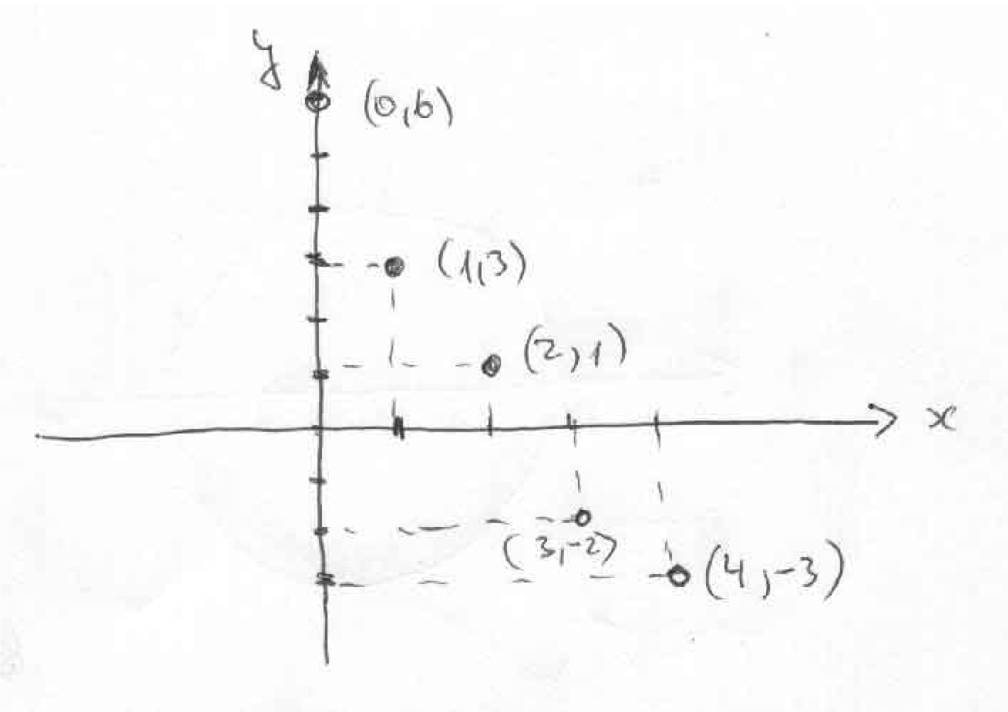
$$a = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\cdot\sum x_i^2 - (\sum x_i)^2}, \quad b = \frac{\sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i}{n\cdot\sum x_i^2 - (\sum x_i)^2} \quad (**)$$

Dobiveni pravac s jednadžbom y = ax+b zove se **regresijski pravac**.

Primjer 2. Procijenimo oblik veze, odredimo parametre, jednadžbu regresijskog pravca, odstupanja i vrijednost funkcije cilja ako je:

x_i	0	1	2	3	4
y_i	6	3	1	-2	-3

Točke (0,6), (1,3), (2,1), (3,-2), (4,-3) predočimo u koordinatnom sustavu.



Vidimo da su točke približno na pravcu, pa tražimo linearnu vezu, tj. vezu oblika y=ax+b.
Unaprijed vidimo da je a<0 i da je $b\approx 6$. Da bismo lakše računali a,b iz (**), izradimo ovakvu tablicu (u posljednjem su stupci zbrojevi elemenata u odgovarajućem redku):

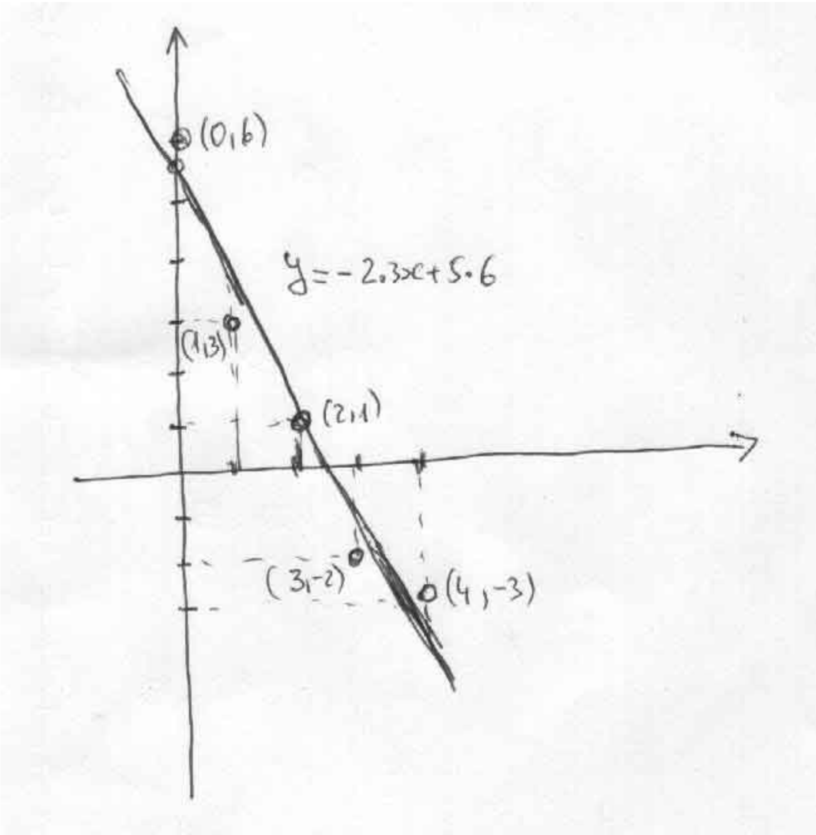
x_i	0	1	2	3	4	10
y_i	6	3	1	-2	-3	5
x_i^2	0	1	4	9	16	30
$x_i y_i$	0	3	2	-6	-12	-13

Tu je jos n=5, pa iz (**) dobijemo:

$$a = \frac{5 \cdot (-13) - 10 \cdot 5}{5 \cdot 30 - 10^2} = \frac{-115}{50} = -2.3$$

$$b = \frac{30 \cdot 5 - 10 \cdot (-13)}{5 \cdot 30 - 10^2} = \frac{280}{50} = 5.6$$

Dakle, tražena je linearna veza: y = -2.3 x+ 5.6 (to je jednadžba regresijskog pravca).



Da bismo odredili odstupanja i vrijednost funkcije cilja, izračunajmo najprije vrijednosti funkcije f(x):=-2.3x+5.6, redom za x=0,1,2,3,4.
f(0)=5.6
f(1)=3.3
f(2)=1.0
f(3)=-1.3
f(4)=-3.6
Vidimo da su teoretski rezultati bliski eksperimentalnim podatcima, što pokazuje i tablica (u posljednjem su redku **odstupanja** $D_i := y_i - f(x_i)$).

x_i	0	1	2	3	4
y_i	6	3	1	-2	-3
$f(x_i)$	5.6	3.3	1.0	-1.3	-3.6
D_i	0.4	-0.3	0.0	-0.7	0.6

Uočite da je **zbroj odstupanja jednak nuli** (to vrijedi općenito: $\sum D_i = 0$).
Minimalna vrijednost funkcije cilja jednaka je zbroju kvadrata odstupanja:
$\sum D_i^2 =$ 0.4^2 + (-0.3)^2 + 0.0^2 + (-0.7)^2+0.6^2 = 1.10 .

Napomena. Za veliki broj podataka provođenje metode najmanjih kvadrata može biti mukotrpno. Zato je dobro naučiti primjenu grafičkog kalkulatora ili nekog dostupnog kompjutorskog programa. Na primjer, naredba LinReg na grafičkom kalkulatoru, za podatke iz Primjera 1. daje nam (zaokruženo na tri decimale) a=3.071 i b = -1.033.

Dobivena linearna veza može nam poslužiti za procjenu (približno određivanje) vrijednosti veličine y za vrijednosti x unutar područja mjerenja (**interpolacija**) ili izvan njega (**ekstrapolacija**)

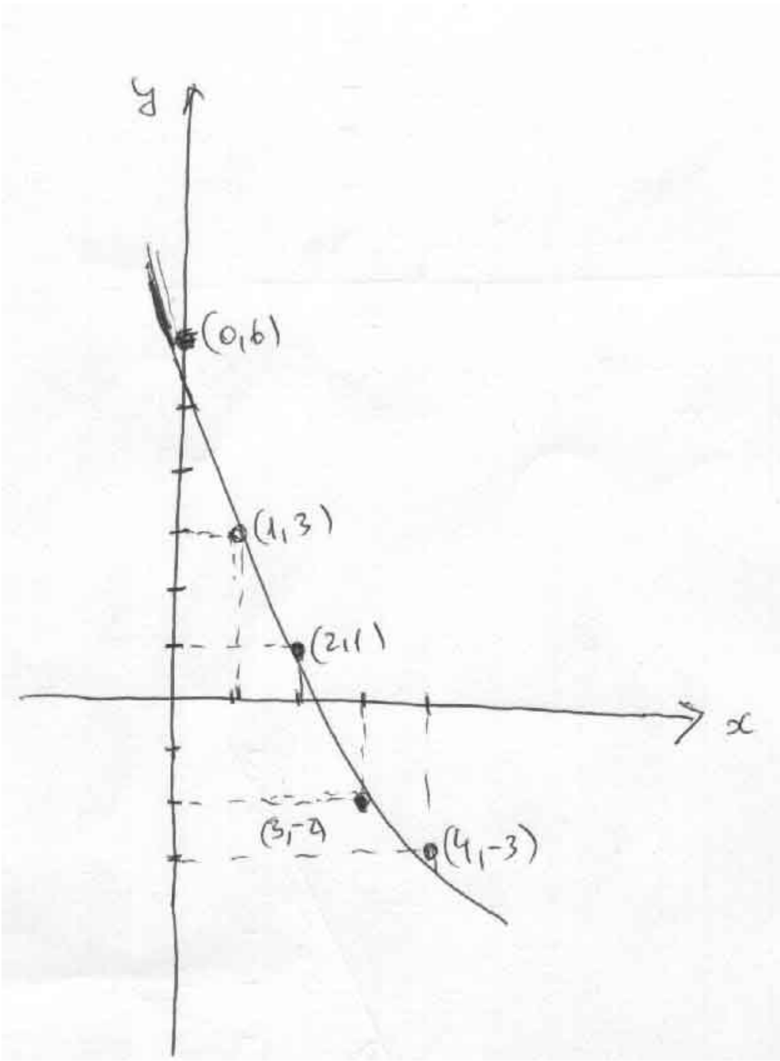
Primjer 3. Procijenimo vrijednost veličine x iz predhodnog primjera za x=2.5 i x=5.
Iz f(x):=-2.3x+5.6, dobijemo f(2.5)=-2.3·2.5+5.6 = -0.15.
Dakle interpolacijom smo dobili da vrijednosti x=2.5 približno odgovara vrijednost y= - 0.15 (koju, inače, nismo mjerili). Uočite da se ta vrijednost razlikuje od srednje vrijednosti u x=2 i x=3 (koja je jednaka $\frac{2+(-3)}{2}$ =-0.5).
Takodjer se dobije f(5)= -2.3·5+5.6= -5.9, pa je y=-5.9 približna vrijednost veličine y za x=5 (kažemo da smo je dobili ekstrapolacijom).

Uočimo da u Primjeru 2. dobiveni regresijski pravac prolazi podatkom (2,1). Općenito, on ne mora prolaziti ni kroz jednu točku. Jedan od razloga zašto se tu tako dogodilo jest taj što je 2 aritmetička sredina vrijednosti x veličine, a 1 aritmetička sredina vrijednosti y veličine (provjerite; to je slučaj, sredina skupa podataka općenito nije podatak). Naime, regresijski pravac uvijek prolazi točkom $(\bar{x}, \bar{y})$, gdje je

$$\bar{x} := \frac{x_1 + x_2 + ... + x_n}{n}, \quad \bar{y} := \frac{y_1 + y_2 + ... + y_n}{n}.$$ Dakle $\bar{y} = a\bar{x} + b$.

Odredjivanje primjerene krivulje.

U prijašnjim primjerima sve se vrtjelo oko regresijskog pravca za zadani skup podataka. Inače, ako teoretski ili neki drugi razlozi upućuju na linearnu vezu među veličinama, provodit ćemu linearnu regresiju čak ako podatci znatnije odudaraju od pravca. Sasvim su drukčije okolnosti ako unaprijed nemamo nikakvih zahtijeva na vrstu veze, već nas zanima krivulja s relativno jednostavnom jednadžbom, koja će dobro odgovarati podatcima. Na primjer, letimičan pogled na podatke iz Primjera 2 upućuju na malu zakrivljenost (na to upućuju i izračunata odstupanja D_i).



To nam nameće ideju da pokušamo tražiti najbolju kvadratnu vezu medju podatcima, tj. vezu oblika y = ax^2+bx+c.

Primjer 4. Metodom najmanjih kvadrata odredimo najbolju kvadratnu vezu medju podatcima iz Primjera 2.

Koeficijente a,b,c mogli bismo izračunati koristeći se formulama (*), medjutim mi ćemo se poslužiti gotovim programom. Na primjer, naredba QuadReg na grafičkom kalkulatoru daje nam (uz zaokruživanje na tri decimale)
a = 0.214
b = -3.157
c = 6.029
pa dobijemo, najbolju kvadratnu aproksimaciju
$y = 0.214x^2 -3.157x + 6.029$
Takodjer, za vrijednost funkcije cilja dobijemo
$\sum D_i^2 = 0.457$
što je za više od 50% manje nego kod linearne regresije. Da dobivena parabola bolje prolazi zadanim točkama od regresijskog pravca vidi se i iz crteža.

Veze koje se svode na linearne.

U primjeni se često pojavljuju nelinearne veze medju dvjema veličinama koje se, promjenom mjerne skale, svode na linearne, na primjer:

1. $y = a \cdot x^b$,

je nelinearna veza, a svodi se na linearnu logaritmiranjem:

log y= b ·log x+ log a

(tu su log x i log y linearno povezane),

2. $y = a \cdot b^x$,

svodi se na linearnu

log y= log b · x+ log a

(tu su x i log y linearno povezane),

3. $y = \frac{a}{b+x}$,

svodi se na linearni preko $\frac{1}{y} = \frac{b+x}{a}$, tj. $\frac{1}{y} = \frac{1}{a} \cdot x + \frac{b}{a}$

(tu su $\frac{1}{y}$ i x linearno povezani), itd.

Primjer 5. Odredimo vezu oblika $y = a \cdot b^x$ ako je:

x_i	-1	0	1	2	4
y_i	1.6	2.9	5.9	11.8	48.0

Logaritmiranjem dobijemo log y= log b · x+ log a.

Nakon zamjena Y:=log y, A:=log b, B:=log a, dobijemo linearnu vezu

Y=A·x+B.

Sad imamo ovakvu tablicu (vrijednosti zaokružujemo na 4 decimale).

x_i	-1	0	1	2	4	6
Y_i	0.2041	0.4624	0.7729	1.0719	1.6812	4.1905
x_i^2	1	0	1	4	16	22
x_iY_i	-0.2041	0.0000	0.7709	2.1438	6.7250	9.4355

Tu je još n=5, pa iz formula (**) dobijemo

$$A = \frac{5 \cdot 9.4355 - 6 \cdot 4.1095}{5 \cdot 22 - 6^2} = 0.2978$$

$$B = \frac{22 \cdot 4.1095 - 6 \cdot 9.4355}{5 \cdot 22 - 6^2} = 0.4808$$

To znači da je Y=0.2978·x+04808

najbolja linearna veza, prema metodi najmanjih kvadrata, izmedju Y i x. Međutim, mi trebamo vezu između y i x, tj. trebaju nam parametri a,b. Njih određujemo pomoću parametara A,B. Naime (uz zaokruživanje na 3 decimale),

iz A=log b, dobijemo $b=10^A=1.985$,

a iz $B=\log a$, dobijemo $a=10^B=3.025$.
Dakle,
$y=3.025\cdot 1.985^x$,
je tražena veza između x i y. Vidimo da tu vezu možemo pisati i kao
$y=3\cdot 2^x$

Izračunajmo, radi provjere, vrijednosti funkcije $f(x):= 3\cdot 2^x$, redom za x = -1,0,1,2,4.
$f(-1)= 1.5$
$f(0)=3.0$
$f(1)=6.0$
$f(2)=12.0$
$f(4)=48.0$
Vidimo da se teoretski rezultati jako dobro slažu s eksperimentalnim podatcima.
Za vrijednost funkcije cilja, za $f(x):= 3\cdot 2^x$, dobijemo
$\sum D_i^2 = 0.07$

Napomena. Zadatak smo mogli riješiti pomoću grafičkog kalkulatora naredbom *ExpReg*.