# METODA NAJMANJŠIH KVADRATOV

Raziskovalna naloga srednješolcev s področja matematike


Avtorica: Manca LEPOŠA

Mentor: Marko KOSELJ, prof.


Strahinj, 2022

## ZAHVALA

Zahvaljujem se svojemu mentorju prof. Marku Koselju, ker je z mano delil ogromno znanja in mi omogočil izvedbo te projektne naloge.

## I KAZALO (VSEBINE)

## II. KAZALO SLIK IN GRAFOV

## Povzetek

Nekega dne sem se pri uri matematike vprašala, kako bi točkam, ki ne ležijo na isti premici, izračunali najbolje prilegajočo premico. Tako sem se lotila te projektne naloge. Prvi korak je bil izpeljava formul za izračun vrednosti $k$ in $n$, da se bo naša linearna funkcija $y = kx + n$ najbolje prilegala danim točkam.

Nato sem se lotila izračunov. Ugotovila sem, da metodo najmanjših kvadratov uporablja veliko različnih aplikacij, s pomočjo katerih lahko izračunamo vrednosti $k$ in $n$ glede na vnesene točke. Prvi primer sem izračunala ročno. Nato sem izračun naredila še s pomočjo računala, v programih Excel in PSPP ter napisala program za izračun vrednosti $k$ in $n$ še v programskem jeziku Python. Za vsak izračun sem uporabila nove vrednosti.

Odkrila sem, da linearna funkcija ni edina funkcija, ki jo lahko izračunamo z metodo najmanjših kvadratov. Nekatere preostale funkcije se da linearizirati in tako lahko namesto premice dobimo še ostale najbolje prilegajoče funkcije, kot so eksponentna funkcija, logaritemska funkcija in potenčna funkcija. Za vsako sem tudi naredila posamezen primer.

**Ključne besede:** linearna regresija, metoda najmanjših kvadratov

## Abstract

One day at my math class, I wondered how to calculate the best fitting line for points that do not lie on the same line. So, I tackled this project task. The first step was to derive the formulas to calculate the values of $k$ and $n$ so that our linear function $y = kx + n$ will best fit the given points.

Then I started the calculations. I found out that the least squares method is used by many different applications that can be used to calculate the values of $k$ and $n$ based on the points entered. I calculated the first case manually. Then I did the other calculations with the help of a calculator, in Excel and PSPP, and also wrote a program for calculating the values of $k$ and $n$ in the Python programming language. I used new values for each calculation.

I discovered that the linear function is not the only function that can be calculated by the least squares method. Some of the other functions can be linearized, so that instead of a linear one we can get other best-fitting functions such as exponential function, logarithmic function and power function. I also made an individual case for each.

**Key words:** linear regression, least squares method

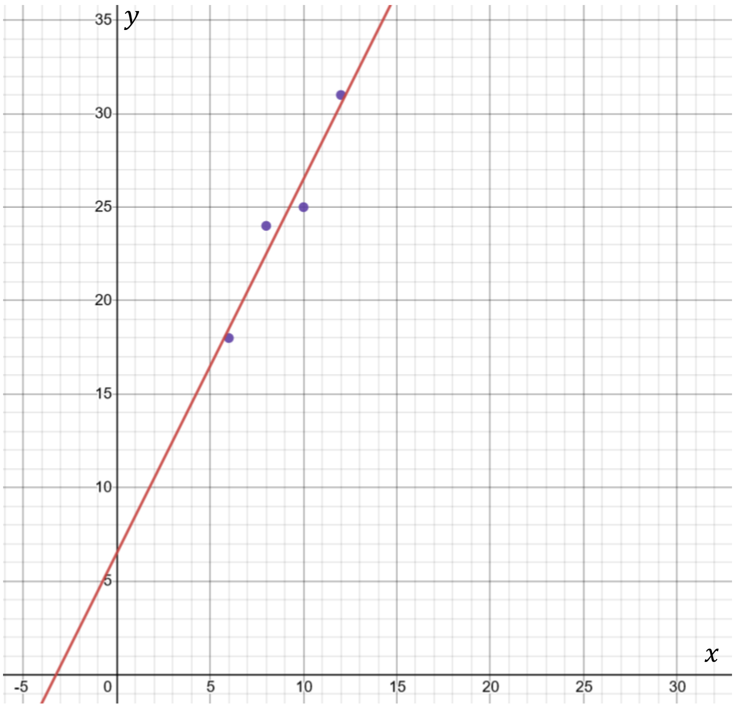
# 1 UVOD

V raziskovalni nalogi želim spoznati metodo najmanjših kvadratov, ki danim točkam v koordinatnem sistemu poišče premico, ki se tem točkam najbolj prilega. Želim spoznati tudi programska orodja za uporabo te metode. In po možnosti še to, ali bom ob pomoči mentorja uspela izpeljati obrazce za rešitev te naloge.

Začnimo s konkretnim problemom. Snežilo je. Ob 6. uri je bila višina snežne odeje 18 cm. Ob 8. uri je bila višina snežne odeje 24 cm. Ob 10. uri je bila višina snežne odeje 25 cm. Ob 12. uri je bila višina snežne odeje 31 cm. Zberimo podatke še v tabeli in narišimo ustrezne točke v koordinatnem sistemu.

| | | | | | |
|---|---|---|---|---|---|
| X | Čas (h) | 6 | 8 | 10 | 12 |
| Y | Višina (cm) | 18 | 24 | 25 | 31 |

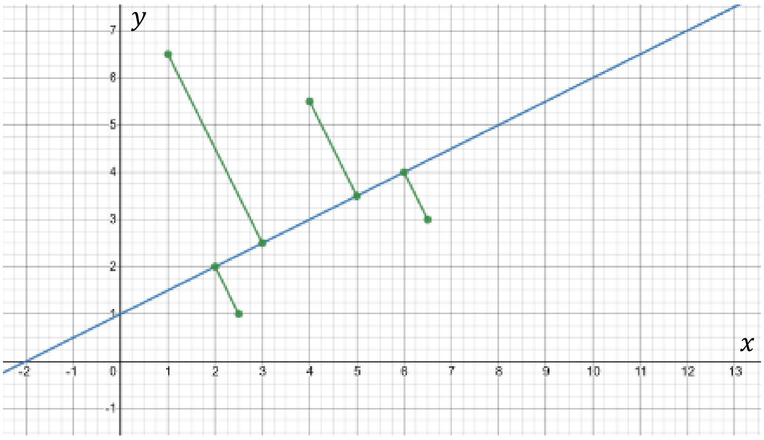
Zanima nas, katera funkcija se najbolj prilega tem točkam. Vemo sicer, da skozi štiri točke lahko položimo graf polinoma tretje stopnje, ki se bo popolnoma prilegal točkam, ali pa graf kakšne druge funkcije, ki bi se točkam popolnoma prilegal. Toda take rešitve v našem primeru niso zaželene, ker predpostavljamo, da ves čas sneži približno enako intenzivno. Naše vprašanje zato lahko zastavimo konkretneje. Ali znamo poiskati premico, ki se najbolj prilega danim točkam?



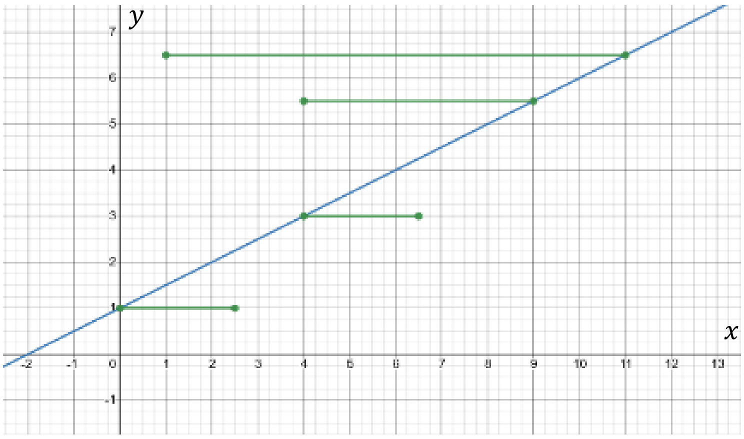

Slika 1: Premica, ki se najbolje prilega meritvam snežne odeje

Rezultat take naloge bi bil uporaben. Če poznamo premico, ki se najbolj prilega danim točkam, znamo z njeno pomočjo ugotoviti, kolikšna bo višina snežne odeje čez nadaljnje tri ali štiri ure. To pa je lahko uporaben podatek, na primer za organizatorja zimske snežne službe, da se lahko pravilneje odloči, koliko snežnih plugov naj pošlje na mestne ulice.

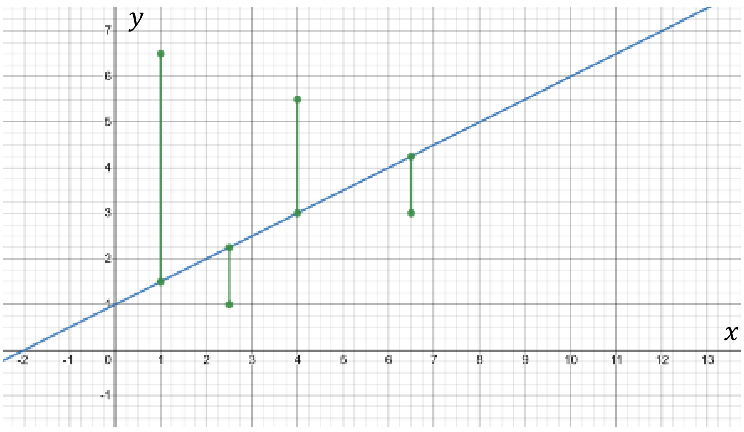
Opredeliti moramo še, kako bomo merili, katera premica se bolje prilega danim točkam in katera se slabše. Zagotovo je smiselno opazovati razdalje točk od premice. Razdaljo lahko merimo na več načinov. Lahko v običajnem smislu, ko je razdalja točke od premice enaka razdalji točke do najbližje točke na premici (slika 2), ali pa tako, da opazujemo razdaljo točke od premice v x-smeri (slika 3). Lahko pa tudi opazujemo razdaljo točke od premice v y-smeri (slika 4).




Slika 2: Razdalja od premice v običajnem smislu




Slika 3: Odmiki od premice v x-smeri




Slika 4: Odmiki od premice v y-smeri

Zadnji način pa je za obravnavo najbolj enostaven, zato bomo odmike točk od premice merili v navpični smeri, kot je prikazano na sliki 4.

Odmiki so lahko pozitivni ali negativni. Ena možnost za merjenje, koliko se premica prilega točkam, je, da opazujemo vsoto

$$f = |d_1| + |d_2| + |d_3| + |d_4|. \quad (1)$$

Tista premica, pri kateri je vsota f najmanjša, se najbolje prilega.

Druga možnost pa je, da opazujemo vsoto

$$f = d_1^2 + d_2^2 + d_3^2 + d_4^2. \quad (2)$$

Tudi pri tem načinu je poskrbljeno, da so negativni odmiki enako obravnavani kot pozitivni. Ima pa ta način to prednost, da imajo veliki odmiki na vsoto f še večji vpliv, saj so kvadrirani in imajo v primerjavi s kvadriranimi majhni odmiki še večje razmerje. V tem primeru bo vsota f še bolj skrbela za to, da bo za najbolje prilegajočo premico izbrana tista, ki ima čim manj točk z velikimi odmiki. Seveda pa je tudi, kot že prej omenjeno, lažji za obravnavo.

Zato se odločimo za drugi način (2). Vprašanje naše naloge o naraščanju snežne odeje je torej, katera premica je tista, pri kateri je vsota kvadratov odmikov točk od premice (2) čim manjša.

Zastavimo si nalogo še v splošnem. Imamo podanih N točk v dvorazsežnem koordinatnem sistemu: $(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots (x_N, y_N)$. Določimo tisto premico $y = kx + n$, pri kateri je vsota kvadratov odmikov

$$f(k,n) = (y_1 - (kx_1 + n))^2 + (y_2 - (kx_2 + n))^2 + \cdots + (y_N - (kx_N + n))^2 \quad (3)$$

najmanjša med vsemi. Iščemo torej smerni koeficient k in začetno vrednost n tiste premice, pri kateri je vsota f najmanjša možna.

S tem problemom se je intenzivno ukvarjal Carl Friedrich Gauβ (1777–1855) in ga tudi uspešno rešil. Verjetno tudi v kakšni še bolj splošni obliki. Ker želimo minimizirati vsoto kvadratov odmikov, se postopek za rešitev te naloge imenuje metoda najmanjših kvadratov (least squares method). Ker želimo najti premico oziroma linearno funkcijo, ki se najbolj prilega danim točkam, v literaturi pri statistikih najdemo obravnavo te naloge tudi pod imenom linearna regresija (linear regression). Praktiki pa pogosto reševanje te naloge imenujejo prilagajanje krivulj (curve fitting).

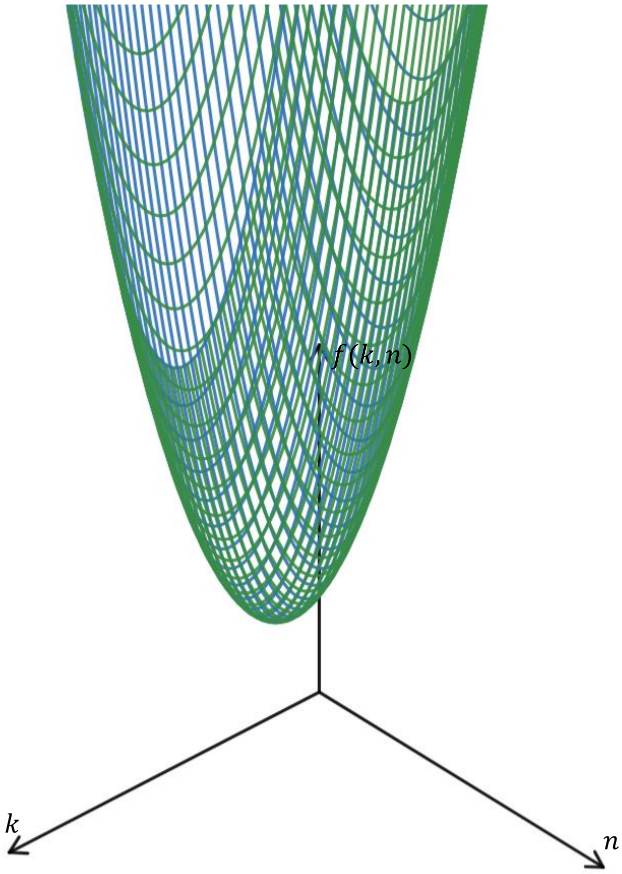
Vrnimo se k našemu primeru, ko nas zanima, kako se s časom spreminja višina snežne odeje. Če v enačbo (3) vstavimo podatke naših štirih točk, dobimo

$$f(k,n) = (18 - (k6 + n))^2 + (24 - (k8 + n))^2 + (25 - (k10 + n))^2 + (31 - (k12 + n))^2$$

Če izraz poenostavimo, dobimo

$$f(k,n) = 344k^2 + 72kn + 4n^2 - 1844k - 196n + 2486.$$

Zanima nas, pri katerem k in n ima funkcija f minimum oziroma najnižjo vrednost. Naloga je zahtevna, saj moramo poiskati minimum funkcije dveh spremenljivk. Grafično problem izgleda približno takole:




**Slika 5: Graf funkcije dveh spremenljivk z minimumom**

Graf funkcije dveh spremenljivk f(k, n) je predstavljen kot ploskev v trirazsežnem koordinatnem sistemu. Zanima nas, pri katerem k in n ima ta ploskev najnižjo točko oziroma predstavlja najnižjo vrednost funkcije.

# 2 IZPELJAVA OBRAZCEV

Funkcijo $f(k,n)$, ki nastopa v (3) in katere minimum bi radi našli, bi radi zapisali v prikladnejši obliki. Ker bomo imeli veliko opravka z vsotami, vpeljimo oznako za vsoto:

$$\sum_{i=1}^{N} x_i = x_1 + x_2 + x_3 + x_4 + \ldots + x_N \,.$$

Hitro se da pokazati, da ima vsota dve lepi lastnosti, ki ju lahko uporabljamo v nadaljevanju:

$$\sum_{i=1}^{N} (x_i + y_i) = \sum_{i=1}^{N} x_i + \sum_{i=1}^{N} y_i \quad ,$$

$$\sum_{i=1}^{N} (\lambda x_i) = \lambda \cdot \sum_{i=1}^{N} x_i \quad .$$

S pomočjo oznake za vsoto in obeh lastnosti vsote lahko funkcijo $f(k,n)$ zapišemo takole:

$$f(k,n) = \big(y_1 - (kx_1 + n)\big)^2 + \big(y_2 - (kx_2 + n)\big)^2 + \ \ldots + \big(y_N - (kx_N + n)\big)^2$$

$$= \sum_{i=1}^{N} \big(y_i - (kx_i + n)\big)^2$$

$$= \sum_{i=1}^{N} (y_i - kx_i - n)^2$$

$$= \sum_{i=1}^{N} (y_i^2 + k^2 x_i^2 + n^2 - 2kx_i y_i - 2ny_i + 2knx_i)$$

$$= \sum_{i=1}^{N} y_i^2 + \sum_{i=1}^{N} k^2 x_i^2 + \sum_{i=1}^{N} n^2 + \sum_{i=1}^{N} (-2kx_i y_i) + \sum_{i=1}^{N} (-2ny_i) + \sum_{i=1}^{N} 2knx_i$$

$$= \sum_{i=1}^{N} y_i^2 + k^2 \sum_{i=1}^{N} x_i^2 + n^2 \sum_{i=1}^{N} 1 - 2k \sum_{i=1}^{N} x_i y_i - 2n \sum_{i=1}^{N} y_i + 2kn \sum_{i=1}^{N} x_i$$

$$= \sum_{i=1}^{N} y_i^2 + k^2 \sum_{i=1}^{N} x_i^2 + n^2 N - 2k \sum_{i=1}^{N} x_i y_i - 2n \sum_{i=1}^{N} y_i + 2kn \sum_{i=1}^{N} x_i$$

$$= k^2 \sum_{i=1}^{N} x_i^2 + 2kn \sum_{i=1}^{N} x_i + Nn^2 - 2k \sum_{i=1}^{N} x_i y_i - 2n \sum_{i=1}^{N} y_i + \sum_{i=1}^{N} y_i^2 \quad .$$

Vsote, ki tu nastopajo, vedno lahko izračunamo iz danih podatkov. Za izračunane vsote vpeljimo nove oznake:

$$S_x = \sum_{i=1}^{N} x_i = x_1 + x_2 + x_3 + \ldots + x_N \,,$$

$$S_y = \sum_{i=1}^{N} y_i = y_1 + y_2 + y_3 + \ldots + y_N \,,$$

$$S_{xx} = \sum_{i=1}^{N} x_i^2 = x_1^2 + x_2^2 + x_3^2 + \ldots + x_N^2 \,,$$

$$S_{yy} = \sum_{i=1}^{N} y_i^2 = y_1^2 + y_2^2 + y_3^2 + \ldots + y_N^2 \,,$$

$$S_{xy} = \sum_{i=1}^{N} x_i y_i = x_1 y_1 + x_2 y_2 + \ \ldots + x_N y_N \,.$$

S temi oznakami lahko našo funkcijo zapišemo

$$f(k,n) = k^2 S_{xx} + 2knS_x + Nn^2 - 2kS_{xy} - 2nS_y + S_{yy} \quad . \tag{4}$$

Radi bi poiskali minimum te funkcije dveh spremenljivk. V literaturi vidimo, da je minimum mogoče hitro najti s pomočjo višješolske matematike. Minimum je tam, kjer je tangentna ravnina na graf funkcije vodoravna. To je tam, kjer sta parcialna odvoda funkcije $f(k,n)$ po $k$ in po $n$ enaka $0$. Ko imamo napisana ta dva pogoja, je treba rešiti le še sistem dveh enačb z dvema neznankama. In dobimo

$$k \ = \ \frac{NS_{xy} - S_x S_y}{NS_{xx} - (S_x)^2} \,,$$

$$n \ = \ \frac{S_{xx} S_y - S_{xy} S_x}{NS_{xx} - (S_x)^2} \ = \ \frac{S_y - k \cdot S_x}{N} \quad .$$

To rešitev najdemo v vseh višješolskih učbenikih, ki obravnavajo to tematiko. Izziv pa je, ali se znamo dokopati do te rešitve le s srednješolsko matematiko.

Spomnimo se, da v drugem letniku srednje šole znamo splošno obliko kvadratne funkcije ene spremenljivke pretvoriti v temensko obliko. Iz te oblike pa takoj vidimo, kje je teme oziroma kje ima funkcija minimum ali maksimum. Ker je kvadratna funkcija precej enostavna, jo znamo s pomočjo dopolnjevanja do popolnega kvadrata pretvoriti v temensko obliko. Če tega ne bi znali, bi si lahko pomagali z ustreznim nastavkom.

Splošna oblika kvadratne funkcije je

$$y = ax^2 + bx + c.$$

Temenska oblika je

$$y = A \cdot (x - P)^2 + Q \ .$$

Ta izraz lahko razčlenimo:

$$y = A\,(x^2 - 2Px + P^2) + Q$$

$$= Ax^2 - 2APx + AP^2 + Q \ .$$

Če želimo, da sta polinom splošne oblike in polinom, dobljen iz temenske oblike, enaka, se morata ujemati v vseh koeficientih:

$$A = a \ ,$$

$$-2AP = b \ ,$$

$$AP^2 + Q = c \ .$$

Od tod dobimo:

$$A = a, \qquad P = -\frac{b}{2a}, \qquad Q = \frac{-(b^2 - 4ac)}{4a}.$$

Na ta način, s pomočjo nastavka z nedoločenimi koeficienti $A,\ P,\ Q$, znamo pretvoriti splošno obliko kvadratne funkcije v temensko obliko, ne da bi poznali postopek dopolnjevanja do popolnega kvadrata. Od tod pa takoj dobimo teme oziroma minimum ali maksimum funkcije. Če je $A$ pozitiven, potem vemo, da ima funkcija v točki (P, Q) minimum, saj je vrednost izraza $A\,(x - P)^2$ najmanjša takrat, ko ima $x$ vrednost $P$ in je zato vrednost tega izraza

$$A\,(x - P)^2 = A\,(P - P)^2 = a \cdot 0^2 = 0 \ .$$

Nižje vrednosti ta izraz ne more imeti, saj je njegova vrednost vedno nenegativna.

Ta ideja nam pride na misel tudi tu, kjer želimo poiskati minimum funkcije dveh spremenljivk. Z dopolnjevanjem do popolnega kvadrata oziroma do dveh popolnih kvadratov bo to najbrž prezahtevno. S primernim nastavkom z nekimi nedoločenimi koeficienti kot zgoraj za funkcijo ene spremenljivke pa bo mogoče uspelo.

Poskusimo funkcijo (4) zapisati s primernim nastavkom. Najenostavneje bi bilo, če bi jo lahko zapisali v obliki

$$f(k,n) = A\,(k - B)^2 + C(n - D)^2 + E\ , \qquad\qquad A, C\ \geq 0.$$

Iz take oblike bi takoj dobili, da ima funkcija minimum pri $k = B$ in $n = D$. Ko zgornji izraz razčlenimo, dobimo

$$f(k,n) = A(k^2 - 2Bk + B^2) + C(n^2 - 2Dn + D^2) + E$$

$$= Ak^2 - 2ABk + AB^2 + Cn^2 - 2CDn + CD^2 + E$$

$$= Ak^2 + Cn^2 - 2ABk - 2CDn + AB^2 + CD^2 + E\ .$$

Če to primerjamo z našo funkcijo (4)

$$f(k,n) = k^2S_{xx} + 2knS_x + Nn^2 - 2kS_{xy} - 2nS_y + S_{yy} \quad ,$$

opazimo, da se koeficientov ne bo dalo izenačiti. Ker v našem nastavku manjka mešani člen, ki bi bil večkratnik izraza $kn$, si s tem nastavkom ne moremo pomagati.

Poskusimo še z razširjenim nastavkom

$$f(k,n) = A\,(k-B)^2 + C(n - Fk - D)^2 + E\,, \qquad A, C\ \geq 0. \qquad (5)$$

Če se nam posreči funkcijo $f(k,n)$ zapisati v taki obliki in je res $A \geq 0$ in $C \geq 0$, potem ima funkcija $f(k,n)$ najmanjšo vrednost oziroma minimum tam, kjer je vrednost obeh izrazov v oklepajih enaka $0$. Torej mora biti

$$k - B = 0\,,$$

$$n - Fk - D = 0$$

oziroma

$$k = B\,,$$

$$n = D + Fk\,.$$

In v tem primeru bi bila naša naloga rešena.

Naš nastavek (5) razčlenimo in uredimo:

$$f(k,n) = \ A(k^2 - 2Bk + B^2) + C(n^2 + F^2k^2 + D^2 - 2Fkn - 2Dn + 2DFk) + E$$

$$= Ak^2 - 2ABk + AB^2 + Cn^2 + CF^2k^2 + CD^2 - 2CFkn - 2CDn + 2CDFk + E$$

$$= \ k^2(A + CF^2) - 2CFkn + Cn^2 + k(-2AB + 2CDF) - 2CDn + AB^2 + CD^2 + E\,.$$

Izenačimo koeficiente s koeficienti funkcije, zapisane v (4):

$$A + CF^2 = S_{xx}\,,$$

$$-2CF = 2S_x\,,$$

$$C = N\,,$$

$$-2AB + 2CDF = -2S_{xy}\,,$$

$$-2CD = -2S_y\,,$$

$$AB^2 + CD^2 + E = S_{yy}\,.$$

To je sistem šestih enačb s šestimi neznankami $A$, $B$, $C$, $D$, $E$, $F$. Hitro izluščimo:

$$C = N\,,$$

$$D = \frac{-2S_y}{-2C} = \frac{S_y}{N}\,,$$

$$F = \frac{2S_x}{-2C} = -\frac{S_x}{N}\,,$$

$$A = S_{xx} - CF^2 = S_{xx} - N\left(-\frac{S_x}{N}\right)^2 = S_{xx} - \frac{NS_x^2}{N^2} = S_{xx} - \frac{S_x^2}{N} = \frac{NS_{xx} - S_x^2}{N}\,,$$

$$B = \frac{-2S_{xy} - 2CDF}{-2A} = \frac{-2(S_{xy} + CDF)}{-2A} = \frac{S_{xy} + CDF}{A} = \frac{S_{xy} + N \cdot \frac{S_y}{N} \cdot \left(-\frac{S_x}{N}\right)}{\frac{NS_{xx} - S_x^2}{N}} =$$

$$= \frac{NS_{xy} - S_xS_y}{N} \cdot \frac{N}{NS_{xx} - S_x^2} = \frac{NS_{xy} - S_xS_y}{NS_{xx} - S_x^2}\,,$$

$$E = S_{yy} - AB^2 - CD^2 = S_{yy} - \frac{NS_{xx} - S_x^2}{N} \cdot \left(\frac{NS_{xy} - S_xS_y}{NS_{xx} - S_x^2}\right)^2 - N \cdot \left(\frac{S_y}{N}\right)^2$$

$$= S_{yy} - \frac{\left(NS_{xy} - S_xS_y\right)^2}{N \cdot (NS_{xx} - S_x^2)} - \frac{S_y^2}{N}\,.$$

Preveriti je treba le še to, da sta koeficienta $A$ in $C$ nenegativna. $C = N$, vemo pa, da je število točk $N$ naravno število in je zato koeficient $C$ res nenegativen. Poglejmo si še koeficient $A = \frac{NS_{xx} - S_x^2}{N}$. Imenovalec $N$ je naravno število. Preveriti je treba le še, da je števec $NS_{xx} - S_x^2$ nenegativen:

$$NS_{xx} - S_x^2 = N \cdot (x_1^2 + x_2^2 + x_3^2 + \ldots + x_N^2) - (x_1 + x_2 + x_3 + \ldots + x_N)^2$$

$$= Nx_1^2 + Nx_2^2 + \ldots + Nx_N^2 - (x_1^2 + x_2^2 + \ldots + x_N^2 + 2x_1x_2 + 2x_1x_3 + \ldots + 2x_{N-1}x_N)$$

$$= Nx_1^2 + Nx_2^2 + \ldots + Nx_N^2 - x_1^2 - x_2^2 - \ldots - x_N^2 - 2x_1x_2 - 2x_1x_3 - \ldots - 2x_{N-1}x_N$$

$$= (N-1)x_1^2 + (N-1)x_2^2 + \ldots + (N-1)x_N^2 - 2x_1x_2 - 2x_1x_3 - \ldots - 2x_{N-1}x_N$$

$$\begin{aligned} &= (x_1 - x_2)^2 + (x_1 - x_3)^2 + (x_1 - x_4)^2 + \ldots + (x_1 - x_N)^2 \\ &\quad + (x_2 - x_3)^2 + (x_2 - x_4)^2 + \ldots + (x_2 - x_N)^2 \\ &\quad + (x_3 - x_4)^2 + \ldots + (x_3 - x_N)^2 \\ &\quad + \ldots + (x_{N-1} - x_N)^2\,. \end{aligned} \tag{6}$$

Ta izraz pa je res nenegativen, saj je vsota kvadratov realnih števil.

Naša funkcija (5) ima zato res minimum pri

$$k - B = 0\,,$$

$$n - Fk - D = 0\,.$$

To pomeni

$$k = B = \frac{NS_{xy} - S_x S_y}{NS_{xx} - S_x^2}\,,$$

$$n = D + Fk = \frac{S_y}{N} + \left(-\frac{S_x}{N}\right) \cdot k = \frac{S_y - kS_x}{N}\,.$$

Če izraz za $n$ izrazimo še naprej, dobimo

$$n = \frac{S_y - \frac{NS_{xy} - S_x S_y}{NS_{xx} - S_x^2} \cdot S_x}{N} = \frac{NS_{xx}S_y - S_x^2 S_y - NS_{xy}S_x + S_x^2 S_y}{NS_{xx} - S_x^2} \cdot \frac{1}{N}$$

$$= \frac{NS_{xx}S_y - NS_{xy}S_x}{NS_{xx} - S_x^2} \cdot \frac{1}{N} = \frac{N \cdot \left(S_{xx}S_y - S_{xy}S_x\right)}{(NS_{xx} - S_x^2) \cdot N} = \frac{S_{xx}S_y - S_{xy}S_x}{NS_{xx} - S_x^2}\,.$$

V izrazih za $B, E, k$ in $n$ nastopa v imenovalcu izraz $NS_{xx} - S_x^2$. Skrbi nas, kaj se zgodi, če ima ta izraz vrednost $0$. Ta izraz imamo v drugačni obliki zapisan v (6). Opazimo, da v primeru, ko sta vsaj dva $x$-a naših točk različna, je ta izraz pozitiven in do deljenja z $0$ ne more priti. Postavimo torej zahtevo, da naše točke ne smejo vse imeti iste $x$-koordinate. V tem primeru je možno izračunati vse prej naštete koeficiente in tudi $k$ in $n$.

Obrazca metode najmanjših kvadratov nam je uspelo izpeljati le z uporabo srednješolske matematike.

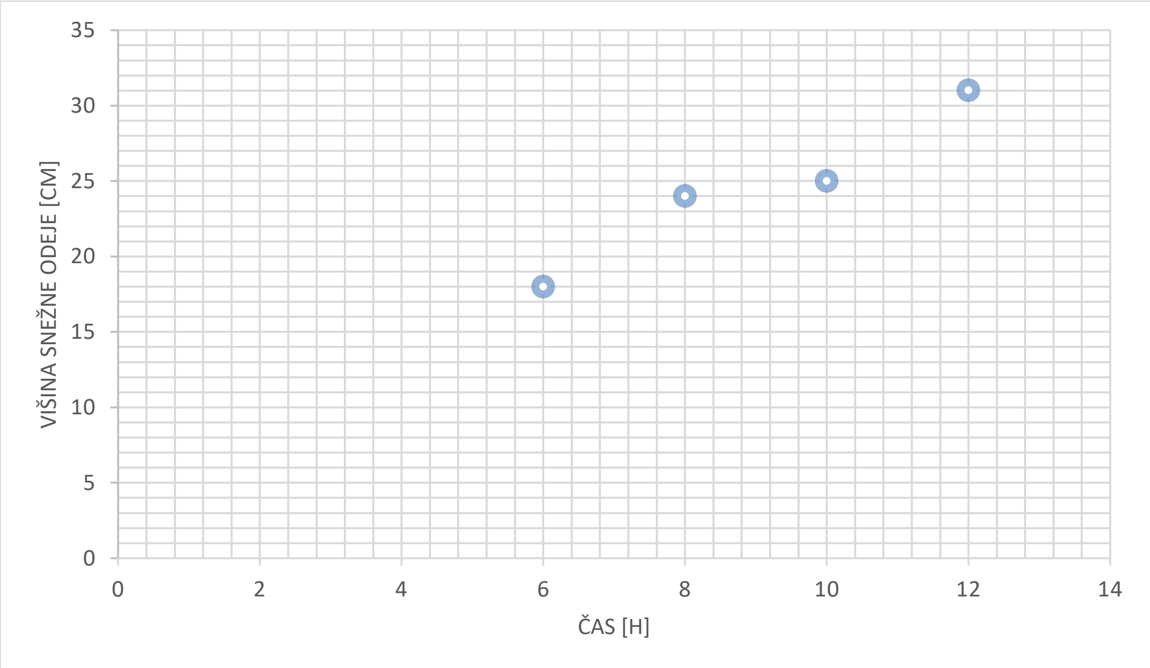
# 3 IZRAČUN

## 3.1 Ročni izračun

Imamo torej 4 točke. Te lahko nazorno ponazorimo v tabeli.

| | Ura (h) | Višina snežne odeje (cm) |
|---|---|---|
| 1. | 6 | 18 |
| 2. | 8 | 24 |
| 3. | 10 | 25 |
| 4. | 12 | 31 |

Če ponazorim te točke na grafu, je iz njega razvidno, da premica, ki bi šla skozi vse te točke, ne obstaja. Sami pa bi zlahka predvideli približno rešitev. Po občutku bi narisali premico, ki se najbolje prilega točkam. Vendar bomo z računanjem poiskali zares najboljšo rešitev.




Graf 1: Odvisnost višine snežne odeje od časa

Računanja se lotimo s pridobivanjem vrednosti $Sx, Sy, Sxx, Sxy$:

$$Sx = 6 + 8 + 10 + 12 = 36,$$
$$Sy = 18 + 24 + 25 + 31 = 98,$$
$$Sxx = 36 + 64 + 100 + 144 = 344,$$
$$Sxy = 108 + 192 + 250 + 372 = 922.$$

Ko so vse vrednosti izračunane, jih moramo le še vstaviti v formuli:

$$k = \frac{NSxy - SxSy}{NSxx - SxSx},$$

$$n = \frac{SxxSy - SxySx}{NSxx - SxSx} = \frac{Sy - kSx}{N}.$$

Za naše točke to izgleda takole:

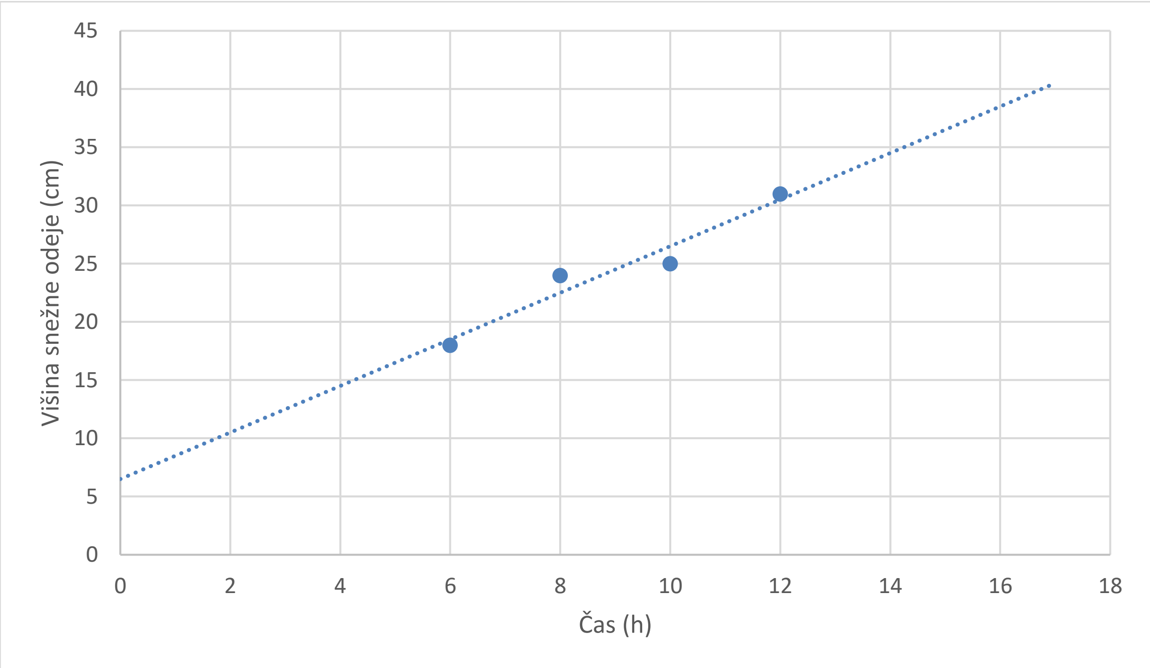
$$k = \frac{4*922-36*98}{4*344-36*36} = 2,$$

$$n = \frac{344*98-922*36}{4*344-36*36} = 6{,}5.$$

Tako torej izvemo, da se našim točkam najbolje prilegajoča linearna funkcija glasi:

$$f(x) = 2x + 6{,}5.$$

Oglejmo si to premico skupaj z danimi točkami:




Graf 2: Odvisnost višine snežne odeje od časa

## 3.2 Izračun s pomočjo žepnega računala

Opisan račun je bil izveden na Sharp računalu, saj so ti eni najbolj popularnih računal. V roke primemo računalo in ga vklopimo. Nato na desni strani kliknemo gumb, na katerem piše MODE in nato še 1. Tako vklopimo način za statistiko. S pritiskom na gumb 1 izberemo LINE MODE, ki se uporablja za izračunavanje linearne regresije.

Naslednji korak je vnašanje podatkov. Najprej vtipkamo vrednost x-koordinate prve točke in nato pritisnemo tipko z oznako $(x, y)$. Nadaljujemo z vnašanjem vrednosti y-koordinate. Nato pritisnemo gumb, na katerem piše DATA. S tem smo vnesli točko. Če hočemo vnesti več točk, ponavljamo ta proces, dokler nismo vnesli vseh. Vnesemo jih lahko do 100. Če se nam nekatere vrednosti ponavljajo, nam jih ni potrebno večkrat vpisovati, saj obstajajo krajši načini. Če se nam določena točka ponovi večkrat, lahko po njenem vnosu (po pritisku gumba $(x, y)$ ob koncu vnosa y-koordinate) vnesemo število vseh ponovitev točke med podatki in šele nato pritisnemo DATA. To nam bo shranilo toliko identičnih točk, kolikor je velika zadnja podana vrednost. Na primer, če imamo dani enaki točki (3,5) in (3,5), potem to vnesemo z ukazi:

$$3\ (x,y)\ 5\ (x,y)\ 2\ DATA.$$

Ko končamo z vnašanjem točk, se lahko preusmerimo na izračun $k$-ja in $n$-ja. Dobimo ju s tem, da pritisnemo gumb RCL ter nato gumb $a$ ali $b$. Gumb $a$ stoji za vrednost $n$-ja, medtem ko $b$ predstavlja vrednost $k$-ja.
Poleg vrednosti $k$ in $n$ nam računalo s pomočjo drugih tipk nudi še druge statistične parametre o porazdelitvi vrednosti x-koordinat in y-koordinat naših točk.

Primer:
Majina mama se je odločila pregledati stara oblačila, ki jih njena hči ne nosi več. Spomnila se je trenutkov, ko je Maja te obleke še nosila. Nato se je začela spraševati, kako hitro je njena hčerka pravzaprav zrastla. Tukaj so podatki o Majinem letu starosti in njeni višini pri tisti starosti:

| Majina starost | Višina |
|---|---|
| 2 | 84,3cm |
| 3 | 96,5cm |
| 5 | 110,1cm |
| 7 | 120, 3cm |
| 9 | 129,9cm |
| 12 | 155,2cm |
| 13 | 157,4cm |
| 15 | 159,7cm |
| 17 | 167,1cm |

Po zgornjem postopku vnesemo podatke. Računalo nam vrne vrednosti $n = 79{,}8$ in $k = 5{,}57$. V grobem lahko trdimo, da je telesna višina Maje v centimetrih naraščala z njeno starostjo v letih po formuli

$$višina = 6 * leta + 80.$$

Ta odvisnost nam da grobe rezultate za obdobje od 2. do 17. leta Majine starosti.

Lahko tudi odtipkamo na primer $14\ 2ndF\ y'$. To nam da odgovor, da je bila Maja pri 14 letih visoka približno 157,8 cm. Če pa odtipkamo na primer $141\ 2ndF\ x'$, dobimo odgovor, da je to višino imela Maja pri 11 letih.

## 3.3 Izračun s pomočjo programa Microsoft Excel

Izmerili smo porabo bencina pri vožnji z avtom na relaciji z določeno razdaljo. Podatki so v tabeli:

| Pot (km) | Poraba (litri) |
|---|---|
| 153 | 11,3 |
| 212 | 14,5 |
| 342 | 22,9 |
| 417 | 29,6 |

Iščemo premico, ki se najbolj prilega tem točkam. V Excelov delovni list vnesemo zgornje podatke, na primer tako, da je prvi stolpec v poljih A2-A5 in drugi stolpec v poljih B2-B5.

Za izračun koeficientov premice v Excelu uporabljamo funkcijo LINEST. Ker ta funkcija vrne matriko vrednosti, jo moramo vnesti kot formulo matrike. To storimo tako, da z miško označimo dve prazni sosednji polji v isti vrstici, vnesemo formulo

$$= LINEST(B2{:}\,B5, A2{:}\,A5, TRUE, FALSE)$$

in pritisnemo Ctrl + Shift + Enter. Takrat se v označenih poljih prikažeta izračunana koeficienta premice $k$ in $n$, in sicer 0,068622 in 0,29213482. Enačba premice je torej

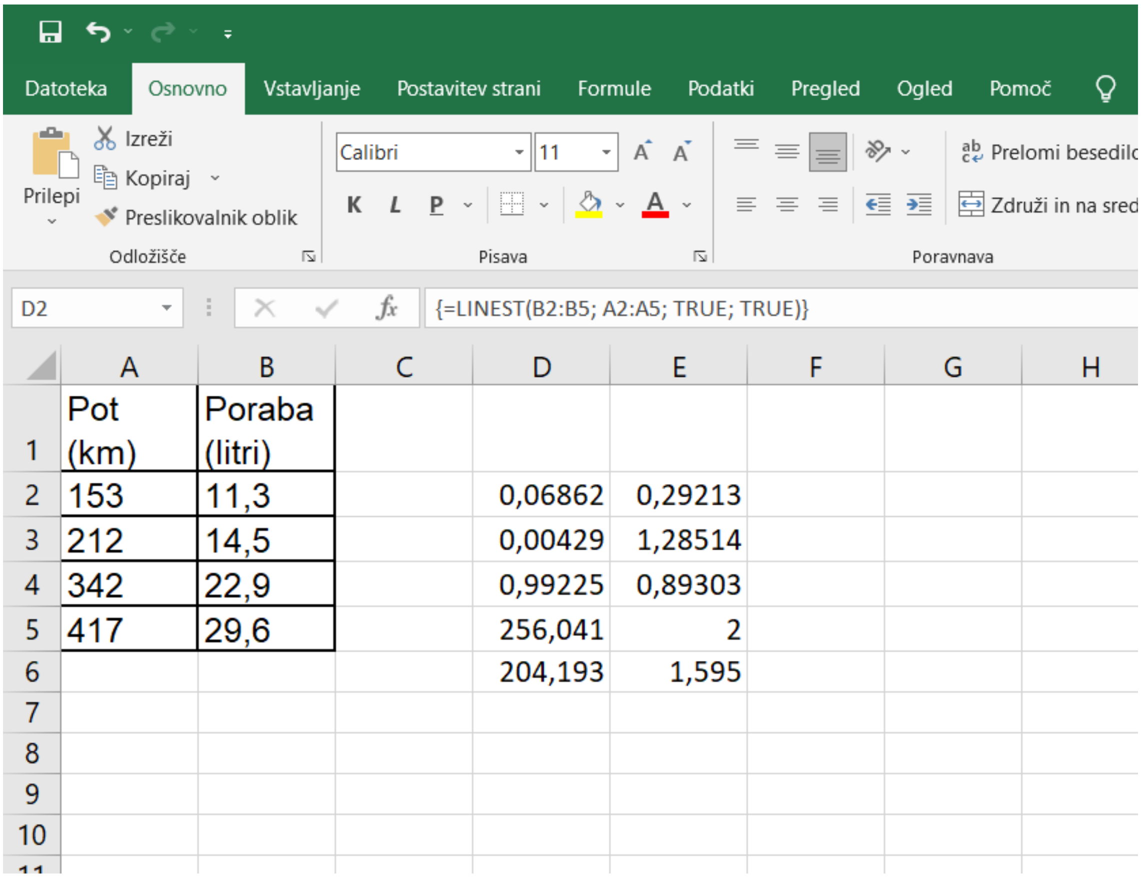
$$y = 0{,}068622 * x + 0{,}29213482.$$

To nam pove, da lahko pričakujemo, da bo naš avto pri vožnji porabil 0,292 litra bencina v vsakem primeru, potem pa še 0,0686 litra za vsak prevožen kilometer.

V funkciji LINEST moramo vnesti štiri parametre. Prvi je obseg celic z y-koordinatami, drugi je obseg celic z x-koordinatami, tretji je TRUE, če iščemo enačbo premice v obliki $y = kx + n$, in FALSE, če iščemo enačbo premice v obliki $y = kx$, in četrti parameter je TRUE, če želimo poleg $k$ in $n$ pridobiti še razne druge statistične vrednosti, in FALSE v nasprotnem primeru. Če se res odločimo za pridobitev dodatnih statističnih vrednosti, moramo v zgornjem primeru z miško označiti nekoliko večje območje praznih sosednjih polj, in sicer vsaj dva stolpca in pet vrstic. Ko odtipkamo

$$= LINEST(B2{:}\,B5, A2{:}\,A5, TRUE, TRUE)$$

in pritisnemo tipke Ctrl + Shift + Enter, dobimo še nekaj dodatnih statističnih parametrov:




| | A | B | C | D | E |
|---|---|---|---|---|---|
| 1 | Pot (km) | Poraba (litri) | | | |
| 2 | 153 | 11,3 | | 0,06862 | 0,29213 |
| 3 | 212 | 14,5 | | 0,00429 | 1,28514 |
| 4 | 342 | 22,9 | | 0,99225 | 0,89303 |
| 5 | 417 | 29,6 | | 256,041 | 2 |
| 6 | | | | 204,193 | 1,595 |

Slika 6: Izgled Excelove tabele z izračunanimi parametri

Zanimivi so koeficient določnosti $r^2$ v tretji vrstici prvega stolpca ter regresivna vsota kvadratov $ssreg$ in residualna vsota kvadratov $ssresid$, ki sta v peti vrstici. Koren iz koeficienta določnosti $r = \sqrt{r^2}$, ki se imenuje koeficient množične korelacije, nam pove, kako dobro naša dobljena enačba premice pojasni odnos med neodvisno in odvisno spremenljivko. Vendar so to že bolj zahtevni statistični pojmi. Povejmo le še to, kako je $r^2$ izračunan. Prej omenjeni vsoti sta:

$$ssreg = \sum_{i=1}^{N} (y(x_i) - \bar{y})^2,$$

$$ssresid = \sum_{i=1}^{N} (y_i - y(x_i))^2.$$

Iz njiju dobimo skupno vsoto kvadratov:

$$sstotal = ssreg + ssresid.$$

Potem pa lahko izračunamo:

$$r^2 = \frac{ssreg}{sstotal} \text{ oziroma } r^2 = 1 - \frac{ssresid}{sstotal}.$$

Možno pa je tudi uporabiti posamični funkciji za izračun $k$ in $n$. V zgornjem primeru bi s funkcijo

$$= SLOPE(B2{:}B5, A2{:}A5)$$

dobili le smerni koeficient $k$. S funkcijo

$$= INTERCEPT(B2{:}B5, A2{:}A5)$$

pa bi dobili le začetno vrednost $n$.

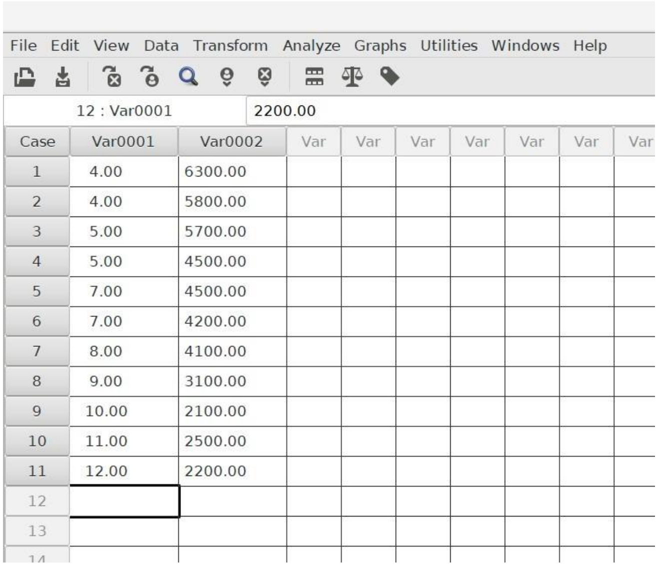
## 3.4 Izračun s pomočjo programa PSPP

PSPP je brezplačna aplikacija, namenjena analizi podatkov. Služi kot brezplačna alternativa za IBM SPSS Statistics, ki je eden vodilnih programov za statistične analize. Tokrat preučujemo razmerje med starostjo in ceno rabljenih avtomobilov, ki jih je v zadnjem letu prodala avtohiša.

Tukaj je tabela s podatki:

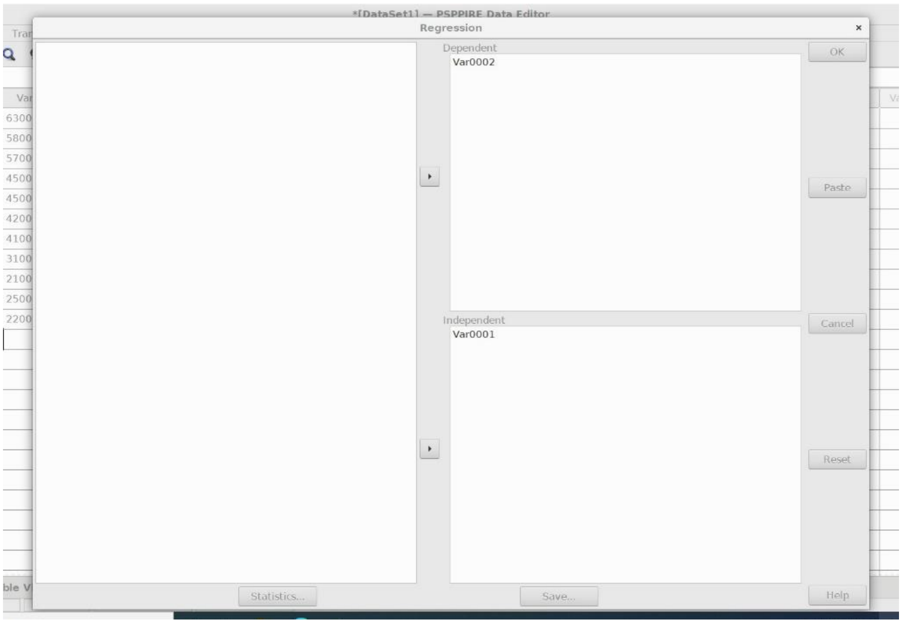
| Starost avta v letih | Cena avta v dolarjih |
|---|---|
| 4 | 6300 |
| 4 | 5800 |
| 5 | 5700 |
| 5 | 4500 |
| 7 | 4500 |
| 7 | 4200 |
| 8 | 4100 |
| 9 | 3100 |
| 10 | 2100 |
| 11 | 2500 |
| 12 | 2200 |

Podatke vnesemo v program. Iz delovnega lista »Variable View« se preselimo na delovni list »Data View« in tam vpišemo podatke:



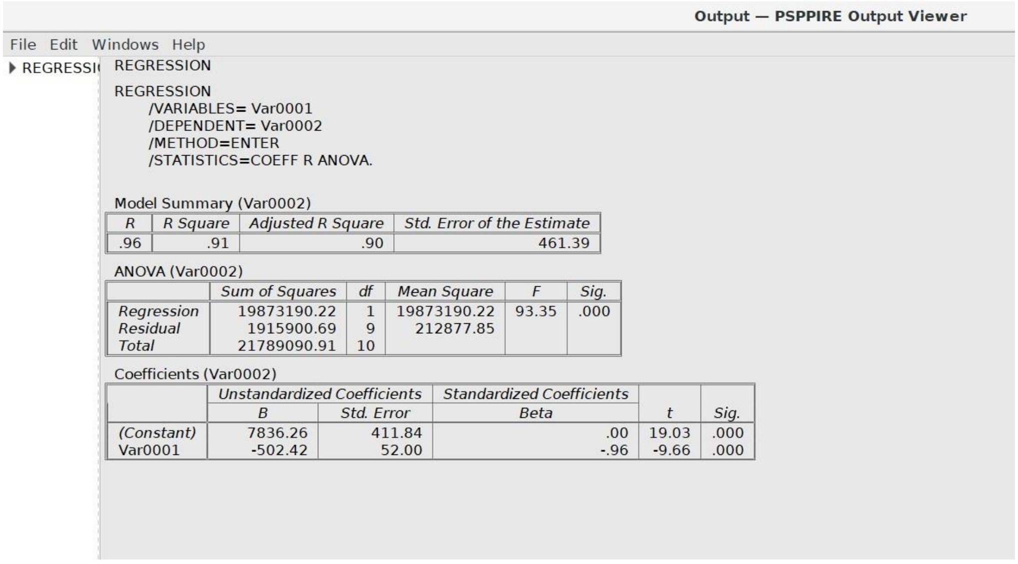
Slika 7: Točke vpisane v PSPP

Iz menija izberemo »Analyse/Regression/Linear«. V oknu, ki se prikaže, moramo opredeliti, katera je neodvisna in katera odvisna spremenljivka. To storimo tako, da kliknemo na posamezno spremenljivko in nato s puščicami urejamo, ali je neodvisna ali odvisna spremenljivka.



Slika 8: Razvrščanje spremenljivk pri iskanju linearne odvisnosti v programu PSPP

Ko imamo spremenljivki razvrščeni, kliknemo »OK«. Pojavi se novo okno z rezultati statistične analize.



Output — PSPPIRE Output Viewer

File Edit Windows Help

REGRESSION

```
REGRESSION
      /VARIABLES= Var0001
      /DEPENDENT= Var0002
      /METHOD=ENTER
      /STATISTICS=COEFF R ANOVA.
```

Model Summary (Var0002)

| R | R Square | Adjusted R Square | Std. Error of the Estimate |
|---|---|---|---|
| .96 | .91 | .90 | 461.39 |

ANOVA (Var0002)

| | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Regression | 19873190.22 | 1 | 19873190.22 | 93.35 | .000 |
| Residual | 1915900.69 | 9 | 212877.85 | | |
| Total | 21789090.91 | 10 | | | |

Coefficients (Var0002)

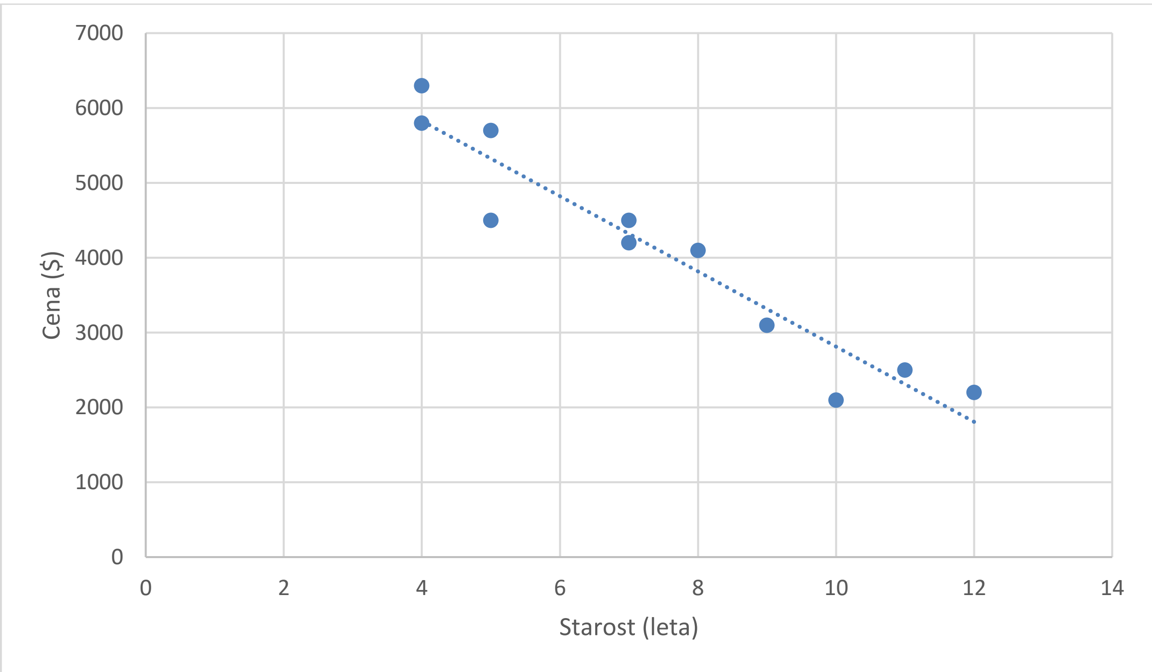
| | Unstandardized Coefficients | | Standardized Coefficients | | |
|---|---|---|---|---|---|
| | B | Std. Error | Beta | t | Sig. |
| (Constant) | 7836.26 | 411.84 | .00 | 19.03 | .000 |
| Var0001 | -502.42 | 52.00 | -.96 | -9.66 | .000 |

**Slika 9: Rezultati analize v PSPP**

V prvem stolpcu z oznako »B« najbolj spodnje tabele zagledamo koeficienta $n$ (zgoraj) in $k$ (spodaj). Opazimo, da je $n = 7836{,}26$ in $k = -502{,}42$. To pomeni, da je naša odvisnost

$$y = -502{,}42 * x + 7836{,}26.$$

Oglejmo si graf te funkcije skupaj z danimi točkami:




**Graf 3: Cena avtomobila v odvisnosti od starosti**

Iz te odvisnosti vidimo, da se cena rabljenemu avtu vsako leto zniža približno za 500 dolarjev.

### 3.5 Izračun s pomočjo programskega jezika Python

Python je eden novejših programskih jezikov in se večinoma uporablja za podatkovne znanosti in statistične problematike. Njegova koda je povprečno krajša od kode ostalih programskih jezikov in zato lahko berljiva, zaradi česar je Python eden najprimernejših jezikov za začetnike; je eden najbolj priljubljenih programskih jezikov v tem času in ima veliko dobrih knjižnic za obravnavo različnih podatkov.

Za pisanje tega programa moramo imeti na našem računalniku nameščen Python ali pa enega od ostalih pythonovih IDE-jev (integrated development environment). Seveda lahko uporabimo tudi tiste, ki jih najdemo na spletu (recimo Programiz na strani: https://www.programiz.com/python-programming/online-compiler/).
Postopek zagona programa, ki sem ga napisala in je naveden v nadaljevanju, je naslednji:

1. korak: Odpremo naš IDE (če imamo naložen Python, v iskalni vrstici poiščemo program imenovan »IDLE« in ga odpremo) in shranimo datoteko pod nekim izbranim imenom.
2. korak: Če prepišemo ali prekopiramo kodo, napisano spodaj, bomo dobili delujoči program. Spodnja koda je obrazložena in nagovarjamo vas, da se poskusite poigrati z njo.
3. korak: Kodo moramo seveda zagnati (v IDLE-ju to naredimo tako, da kliknemo na oddelek imenovan »Run« in nato na tipko »Run module«).

```python
def linear_regression():
    try: #zaženemo določeno kodo, če uporabnik ne vnese števila temveč npr. besedo
        num = int(input("Koliko točk imaš? ")) #spremenljivka num vsebuje naš N

    except ValueError:
        print("Napaka! Vnesti morate celo število!")

#spremenljivke potrebne za izračun
    SX = 0
    SY = 0
    SXX = 0
    SXY = 0

    i = 0

    while i < num: #v tej zanki uporabnik izpiše x in y koordinate točk
        x = int(input("vpiši x od" + str(i + 1) + ". točke: "))
        SX += x #SX-u prišteje x
        SXX += x * x #SXX-u prišteje kvadrat x-a

        y = int(input("vpiši y od" + str(i + 1) + "točke: "))
        SY += y #SY-u prišteje y
        SXY += x * y #SXY-u prišteje zmnožek x*y
        i += 1

# formule za izračun k-ja in n-ja
    k = (num * SXY - SX * SY) / (num * SXX - SX * SX)
    n = (SY - k * SX) / num

    return k, n


print("k, n =", linear_regression())
```

Primer: Poiskati želimo najboljšo linearno zvezo med višino in maso učencev nekega oddelka. Naši podatki:

| Višina | Masa |
|---|---|
| 165 | 60 |
| 179 | 72 |
| 170 | 57 |
| 145 | 37 |
| 154 | 42 |
| 162 | 67 |
| 167 | 51 |
| 172 | 56 |
| 181 | 60 |
| 164 | 44 |
| 159 | 50 |
| 176 | 65 |
| 172 | 53 |
| 158 | 48 |
| 169 | 50 |

Program v Python IDLE lahko zaženemo tako, da na zgornji vrstici kliknemo gumb Run in nato v stolpcu najdemo Run Module ali pa pritisnemo gumb f5 na tipkovnici. Ko zaženemo program, nas program povpraša po številu naših točk. Odgovorimo s celim številom. V našem primeru je to 15. Še naprej sledimo navodilom programa. Posamično vpisujemo x in y koordinate točk. Na koncu nam program vrne vrednosti k in n.

*Python 3.7.4 Shell*

File Edit Shell Debug Options Window Help

```
Python 3.7.4 (tags/v3.7.4:e09359112e, Jul  8 2019, 19:29:22) [MSC v.1916 32 bit (Intel)] on win32
Type "help", "copyright", "credits" or "license()" for more information.
>>>
= RESTART: C:\Users\mancy\OneDrive\Namizje\Programming\linear_regression.py =
Koliko točk imaš?15
vpiši x od1. točke: 165
vpiši y od1točke: 60
vpiši x od2. točke: 179
vpiši y od2točke: 72
vpiši x od3. točke: 170
vpiši y od3točke: 57
vpiši x od4. točke: 145
vpiši y od4točke: 37
vpiši x od5. točke: 154
vpiši y od5točke: 42
vpiši x od6. točke: 162
vpiši y od6točke: 67
vpiši x od7. točke: 167
vpiši y od7točke: 51
vpiši x od8. točke: 172
vpiši y od8točke: 56
vpiši x od9. točke: 181
vpiši y od9točke: 60
vpiši x od10. točke: 164
vpiši y od10točke: 44
vpiši x od11. točke: 159
vpiši y od11točke: 50
vpiši x od12. točke: 176
vpiši y od12točke: 65
vpiši x od13. točke: 172
vpiši y od13točke: 53
vpiši x od14. točke: 158
vpiši y od14točke: 48
vpiši x od15. točke: 169
vpiši y od15točke: 50
k, n = (0.7445054945054945, -69.60347985347987)
>>>
```

Slika 10: Vnašanje koordinat točk in izpis koeficientov našega programa

Zveza, ki predstavlja odvisnost telesne mase od telesne višine, je torej

$$y = 0{,}745 * x - 69{,}6.$$

# 4 PRILAGODITEV METODE ZA DRUGAČNE FUNKCIJE

Metodo se da razširiti v več pogledih:

1. Namesto premice, ki se najbolje prilega točkam v 2-razsežnem prostoru iščemo ravnino, ki se najbolje prilega točkam v 3-razsežnem prostoru ali pa celo hiperravnino v večdimenzionalnem prostoru. Na primer v obliki:
$$v = k_1x + k_2y + k_3z + k_4u + n.$$

2. Namesto premice, ki se najbolje prilega točkam, lahko iščemo kvadratno parabolo, ki se najbolje prilega točkam ali celo polinom višje stopnje. Na primer v obliki:
$$y = ax^3 + bx^2 + cx + d.$$

3. Namesto premice, ki se najbolje prilega točkam, lahko iščemo tudi funkcijo povsem druge oblike. V tem primeru poskusimo funkcijo linearizirati.
   a. Eksponentna funkcija (Primer: Temperatura juhe, ki se ohlaja v odvisnosti od časa)
      $y = a * e^{bx}$
      Lineariziramo jo tako, da zvezo med x in y logaritmiramo.
      $$y = a * e^{bx},$$
      $$\ln y = \ln(a * e^{bx}),$$
      $$\ln y = \ln a + \ln(e^{bx}),$$
      $$\ln y = \ln a + b * x,$$
      $$Y = N + K * X.$$
      Povezave med spremenljivkami:
      $$X = x,$$
      $$Y = \ln y,$$
      $$K = b,$$
      $$N = \ln a.$$

      Pogledali si bomo primer ohlajanja juhe. Razlika med temperaturo juhe in sobno temperaturo bi praviloma morala imeti eksponentno odvisnost od časa. Različni zunanji vplivi (veter, nihanje temperature okolja) pa povzročajo nihanje temperature juhe. To so naši podatki:

| Čas x = X | Temperatura juhe | Sobna temperatura | Razlika temperatur |
|---|---|---|---|
| t (h) | T (°C) | $T_o$ (°C) | $\Delta T$ (°C) |
| 0 | 100 | 20 | 80 |
| 1 | 87 | 20 | 67 |

| | | | |
|---|---|---|---|
| 2 | 73 | 20 | 53 |
| 5 | 50 | 20 | 30 |
| 8 | 40 | 20 | 20 |
| 20 | 21 | 20 | 1 |

Sedaj spremenimo vsako koordinato $y$ naših danih točk v $\ln y$. Te vrednosti nam bodo služile pri izračunu najbolje prilegajoče eksponentne funkcije.

| x = X = t (h) | y = ΔT (°C) | Y = ln y |
|---|---|---|
| 0 | 80 | 4,3820 |
| 1 | 67 | 4,2046 |
| 2 | 53 | 3,9702 |
| 5 | 30 | 3,4011 |
| 8 | 20 | 2,9957 |
| 20 | 1 | 0 |

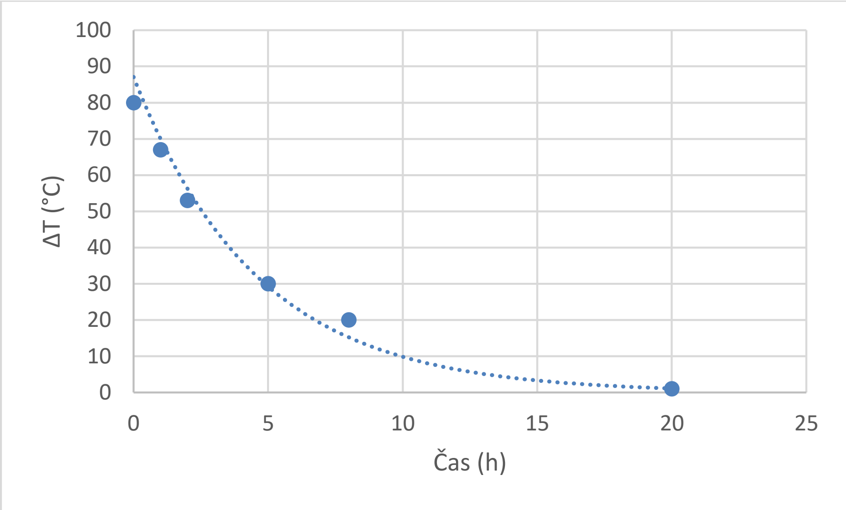
Sedaj smo funkcijo linearizirali. Vse, kar moramo storiti zdaj, je namesto naših vrednosti $y$ uporabiti vrednosti $\ln y$, tako da bomo sedaj uporabljali $X$ in $Y$. Metoda najmanjših kvadratov nam za spremenljivki $X$ in $Y$ vrne

$$K = -0{,}220401,$$
$$N = 4{,}499763.$$

Ko dobimo $K$ in $N$ pa to še nista končni vrednosti naše eksponentne funkcije. Preden vrednosti vnesemo nazaj v enačbo, jih moramo spremeniti iz $K$ v $b$ in iz $N$ v $a$. Vrednost $K$ je sicer že enaka vrednosti $b$, zato nadaljnje določanje/računanje ni potrebno. Iz $N$-ja pa dobimo $a$ tako, da obrnemo enačbo $N = \ln a$. Tako dobimo $a = e^N$ oziroma

$$b = -0{,}220401,$$
$$a = e^N = e^{4{,}499763} = 89{,}995822.$$

Odvisnost spremembe temperature od časa je torej $y = 90 * e^{-0{,}22x}$. Oglejmo si graf te funkcije skupaj z danimi točkami.




**Graf 4:Razlika med temperaturo juhe in sobno temperaturo v odvisnosti od časa**

b. Logaritemska funkcija (Primer: Rast deklic v odvisnosti od starosti)

$y = a + b * \ln x$

Dovolj je, da namesto ln x vpeljemo novo spremenljivko $X = \ln x$ in tako dobimo linearno odvisnost.

$$y = b * X + a$$

Pogledali si bomo rast deklic od rojstva do 6. meseca.

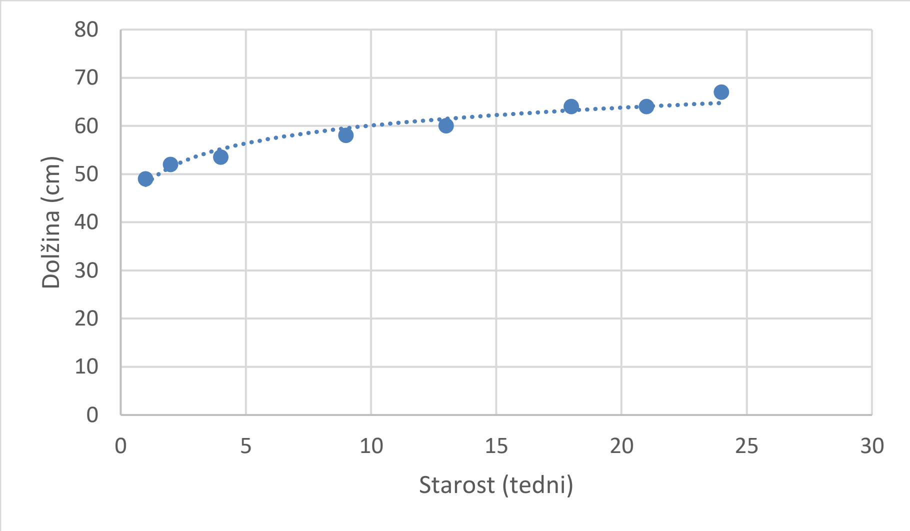
To so naši podatki:

| x = Starost deklice (tedni) | y = Y = Dolžina deklice (cm) | X = ln x |
|---|---|---|
| 1 | 49 | 0 |
| 2 | 52 | 0,693147 |
| 4 | 53,5 | 1,386294 |
| 9 | 58 | 2,197225 |
| 13 | 60 | 2,564949 |
| 18 | 64 | 2,890372 |
| 21 | 64 | 3,044522 |
| 24 | 67 | 3,178054 |

Metoda najmanjših kvadratov nam za spremenljivki $X$ in $Y$ vrne povezavo $Y = 5{,}35 * X + 47{,}8$. Odvisnost dolžine deklic od starosti je torej

$$y = 5{,}35 * \ln x + 47{,}8.$$

Oglejmo si graf te funkcije skupaj z danimi točkami:




**Graf 5: Dolžina deklic v odvisnosti od starosti**

c. Potenčna funkcija

$y = a * x^b$

Lineariziramo jo tako, da to zvezo med y in x logaritmiramo:

$$y = a * x^b,$$
$$\log y = \log(a * x^b),$$
$$\log y = \log a + \log x^b,$$
$$\log y = \log a + b \log x,$$
$$Y = N + K * X.$$

Nove spremenljivke:

$$X = \log x,$$
$$Y = \log y,$$
$$K = b,$$
$$N = \log a.$$

Merili smo odvisnost mase jabolk v gramih od premera jabolk v centimetrih. V tabeli so meritve in logaritmirane vrednosti:

| Premer (x) | Masa (y) | X = log x | Y = log y |
|---|---|---|---|
| 7,1 | 161 | 0,8512 | 2,2068 |

| | | | |
|---|---|---|---|
| 5,6 | 113 | 0,7481 | 2,0530 |
| 6,2 | 113 | 0,7923 | 2,0530 |
| 7,2 | 142 | 0,8573 | 2,1522 |
| 6,4 | 136 | 0,8061 | 2,1335 |
| 6,0 | 107 | 0,7781 | 2,0293 |
| 7,0 | 161 | 0,8450 | 2,2068 |

Po metodi najmanjših kvadratov dobimo

$$K = 1{,}5799,$$
$$N = 0{,}8377$$

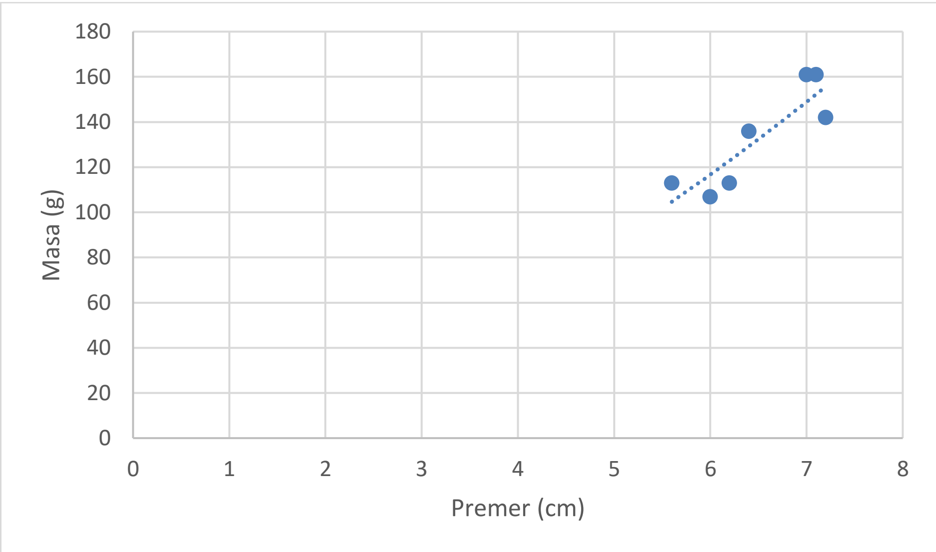
in zvezo med $Y$ in $X$, in sicer

$$Y = 1{,}5799 * X + 0{,}8377.$$

Od tod dobimo $b = K = 1{,}58$ in $a = 10^N = 6{,}88$.

Torej imamo v našem primeru odvisnost $y = 6{,}88 * x^{1{,}58}$.

Oglejmo si še graf te funkcije skupaj z danimi točkami:




**Graf 6: Masa jabolka v odvisnosti od premera**

Pričakovali smo sicer kubično odvisnost, saj volumen jabolk narašča s kubom premera. Odvisnost je najbrž prišla nekoliko drugačna, ker so bila v vzorcu jabolka med sabo zelo podobne velikosti in ker po obliki jabolka niso bila enaka.

Sodobna računala imajo že vgrajeno linearizacijo eksponentne, logaritemske in potenčne funkcije, in sicer tako, da nam že kar vrnejo parametra $a$ in $b$ od eksponentne oziroma logaritemske oziroma potenčne funkcije. V računalu Sharp EL-531TH dosežemo te vgrajene funkcije z ukazom MODE 1 3 oziroma MODE 1 4 oziroma MODE 1 5. V teh načinih delovanja nam ukaza RCL a in RCL b že kar vrneta parametra eksponentne, logaritemske oziroma potenčne funkcije in ni potrebno nobeno ročno preračunavanje.

# 5 ZAKLJUČEK

Vesela sem, da sem spoznala metodo najmanjših kvadratov, ki danim točkam v koordinatnem sistemu poišče premico $y = k * x + n$, ki se točkam najbolje prilega. Sedaj znam ročno izračunati $k$ in $n$ po obrazcih in znam za izračun uporabiti tudi vgrajeno funkcijo računala. Te vrste izračunov sem spoznala tudi z uporabo tehnologije, ki je zelo uporabna za statistične izračune. To so funkcije v programu Microsoft Excel in funkcije v programu PSPP, ki je ekvivalenten enemu najbolj izpopolnjenih programov za statistične obdelave IBM SPSS Statistics. Uspelo mi je tudi napisati program oziroma kodo za izračun koeficientov regresijske premice v programskem jeziku Python.

Še posebej zadovoljna pa sem, da nama je z mentorjem uspelo izpeljati formuli za $k$ in $n$ s pomočjo srednješolske matematike, saj v literaturi povsod najdemo izpeljavo le s pomočjo parcialnih odvodov funkcije dveh spremenljivk, kar pa je zame še prezahtevno, saj to spada že v visokošolsko matematiko. Naloga je preizkusila najino izvirnost in bila poseben izziv, saj reševanja tega problema na tak način nisva opazila nikjer drugje. Prav zaradi tega se bom te naloge zapomnila kot ene najpomembnejših, saj so bili to prvi poskusi matematike brez raznih virov, na katere bi se lahko opirala.

# 6 VIRI IN LITERATURA

Bohte Z. (1987). *Numerične metode* (3. natis). Ljubljana: Zveza organizacij za tehnično kulturo Slovenije.

Željko M. (2011). *Matematične metode* (Izpis: 19. avgust 2011). Ljubljana: BF-Biotehnologija. Na http://zeljko.dmfa.si/. Pridobljeno 3. 12. 2021 s http://zeljko.dmfa.si/lectures/2011/BioT-Matematika.pdf.

*LINEST function.* (2022). Na https://support.microsoft.com/. Pridobljeno 7. 1. 2022 s https://support.microsoft.com/en-us/office/linest-function-84d7d0d9-6e50-4101-977a-fa7abf772b6d.

*PSPP User's Guide* (GNU PSPP Statistical Analysis Software Release 1.4.1). (6. 9. 2020). Free Software Foundation Inc. Na https://www.gnu.org/. Pridobljeno 3. 12. 2021 s https://www.gnu.org/software/pspp/manual/pspp.pdf.

*Regression Analysis.* (28. 2. 2022). Na https://en.wikipedia.org. Pridobljeno 28. 2. 2022 s https://en.wikipedia.org/wiki/Regression_analysis.

*Sharp Scientific Calculator EL-531TH Operation Manual*. Sharp Corporation.

*Length-for-age GIRLS: Birth to 6 months (z-scores).* Na https://www.who.int/. Pridobljeno 7. 1. 2022 s https://www.who.int/tools/child-growth-standards/standards/length-height-for-age.