



2004-2006 m. Bendrojo programavimo dokumento 2 prioriteto „Žmogiškųjų išteklių plėtra“
4 priemonė „Mokymosi visą gyvenimą sąlygų plėtra“

Projekto sutarties numeris: **ESF/2004/2.4.0-K01-160/SUT-261**

Projekto pavadinimas: **Inovatyvūs mokymo(si) metodai ir naujausios technologijos gamtos mokslų bakalauro rengimui**

APL 416. APLINKOS DUOMENŲ ANALIZĖ IR MODELIAVIMAS

2 laboratorinis darbas

Netiesinio ryšio regresinių modelių taikymas

Darbo tikslas:

- Naudodami netiesinio ryšio regresijos metodus, nustatykite populiacijos tankumo įtaką pomidorų (*Lycopersicon esculentum* Mill.) diametro ir aukščio kitimui eksperimento eigoje.

Teorija.

Norint suprasti procesų ir reiškinių esmę, reikia iširti jų ryšius su kitais procesais ir reiškiniais. Pirmiausia iškyla klausimas apie priežastis, sąlygojančias konkretų procesų ar reiškinių vyksmą. Pavyzdžiui, derlius priklauso nuo dirvos struktūros, trąšų kiekio, meteorologinių sąlygų bei augalų tarpusavio sąveikos (piktžolių, tankumo ir t.t.). Kokia derliaus priklausomybė nuo pateiktų faktorių? – tai pagrindiniai klausimai, į kuriuos ir turėtų padėti atsakyti koreliacinė ir regresinė analizė. Regresija – tai vienas pusė statistinė priklausomybė, kuri išreiškiama vadinamąja regresijos funkcija.

Labai dažnai atsako priklausomybė nuo jį veikiančių faktorių yra netiesinė. Daugiausia naudojamos šios netiesinės priklausomybės:

- a) Augimo kreivė, išreiškianti prieaugio (*Growth*) priklausomybę nuo amžiaus (*Age*):

$Growth = \exp(-b_1 * Age) + \varepsilon$ (augimo modelis su adityviaja paklaida). Augimo kreivę su multiplikatyviaja paklaida $Growth = \exp(-b_1 * Age) * \varepsilon$ galima ištiesinti logaritmuojant tos kreivės lygtį: $\ln(Growth) = (-b_1 * Age) + \ln \varepsilon$, o koeficientą b_1 vertinti, naudojant tiesinės regresijos modulį. Prieaugio

priklausomybei nuo amžiaus vertinti naudojama bendresnė augimo kreivė: $\text{Growth} = b_0 + \exp(-b_1 * \text{Age}) + \varepsilon$.

b) Atsako (y) į vaistų dozę (x) kreivė $y = b_0 - b_0 / (1 + (x/b_2)^{b_1}) + \varepsilon$.

c) Logistiniai ir probit modeliai.

Netiesinės regresijos pagalba modeliuojami daugelis individe vykstančių biocheminių procesų. Netiesinės regresijos modelis apibrėžiamas taip:

$$y_i = f(\mathbf{x}_i) + \varepsilon_i, \quad i = 1, 2, \dots, n,$$

čia $\mathbf{x}_i = (x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(k)})$, $y_i - i$ – tojo individo daugiamačio faktoriaus ir atsako reikšmė, $f(\mathbf{x})$ – netiesinė bendru atveju daugiamačio argumento funkcija, ε_i – nepriklausomi ats. d., $\varepsilon_i \sim N(0, \sigma^2)$.

Regresijos funkcija $f(\mathbf{x})$ parenkama taip, kad atspindėtų biologinį ryšį, esantį tarp faktorių ir atsako. Pasirinkus netiesinę regresijos funkciją $f(\mathbf{x})$, sekantis modelio sudarymo etapas – įvertinti nežinomus $f(\mathbf{x})$ parametrus.

Programai „Statistica“ pateikus mums regresinę kreivę bei jos parametrus, kyla klausimas, ar gerai regresinė kreivė atitinka eksperimentinius duomenis. Vienas iš svarbiausių tinkamumo matų yra vadinamas *determinacijos koeficientas*. Determinacijos koeficientas žymimas R^2 ir apibrėžiamas santykiu: regresinio nuokrypio kvadratų suma (KSR)/ bendro nuokrypio kvadratų suma (KSB). Determinacijos koeficientas kinta nuo 0 iki 1. Kuo R^2 arčiau vieneto, tuo regresinė kreivė geriau tinka eksperimentiniams duomenims. Pavyzdžiui, kai $R^2=0,7$, galime teigti, kad 70% y variacijos yra paaiškinama nepriklausomu kintamuoju x .

Regresinė analizė glaudžiai susijusi su koreliaciniais analizės metodais. Koreliacijos tarp kintamųjų stiprumo matas yra *koreliacijos koeficientas*. Jis žymimas R ir apibrėžiamas kaip kvadratinė šaknis iš determinacijos koeficiento R^2 . Kuo arčiau 1 ar -1 yra R , tuo stipresnis koreliacinis ryšys sieja nagrinėjamus kintamuosius.

Duomenys analizei.

Duotas duomenų failas “**PomAugDinam.STA**”, kuriame suvesti duomenys apie pomidorų daigų aukštį ir diametrą eksperimento metu.

Trumpas eksperimento aprašymas:

Pomidorai su pirmuoju tikruoju lapeliu buvo pikuojami į vegetacinius indus skirtingu tankumu. Priklausomai nuo pasirinkto tankumo varianto, į vegetacinį indą buvo sodinama nuo 1 iki 60 daigų, tai atitiko nuo 6 iki 340 vnt./m² augimo tankumus. Eksperimento metu, kuris truko 67 paras (duomenų faile kintamasis – trukme), kas savaitė buvo matuojami pomidorų diametrai (D) ir aukščiai (H). Duomenų faile pateikta kiekvienos savaitės pomidorų augusių skirtinguose tankumuose vidurkiai.

Pomidorų tankumo variantai:

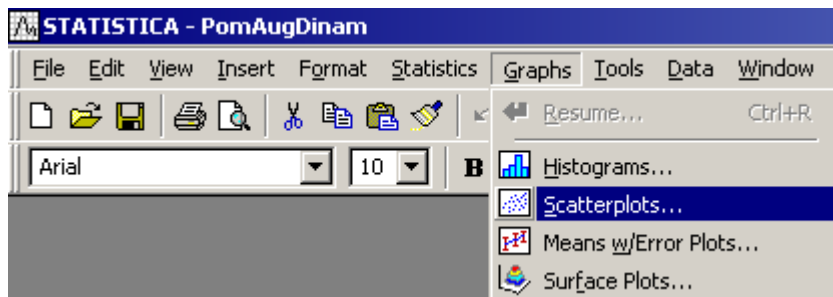
- I) 6 indiv./m² (duomenų faile diametras D_6 ir aukštis H_6)
- II) 35 indiv./m² (D_35 ir H_35)
- III) 50 indiv./m² (D_50 ir H_50)
- IV) 85 indiv./m² (D_85 ir H_85)
- V) 160 indiv./m² (D_160 ir H_160)
- VI) 340 indiv./m² (D_340 ir H_340)

Darbo eiga:

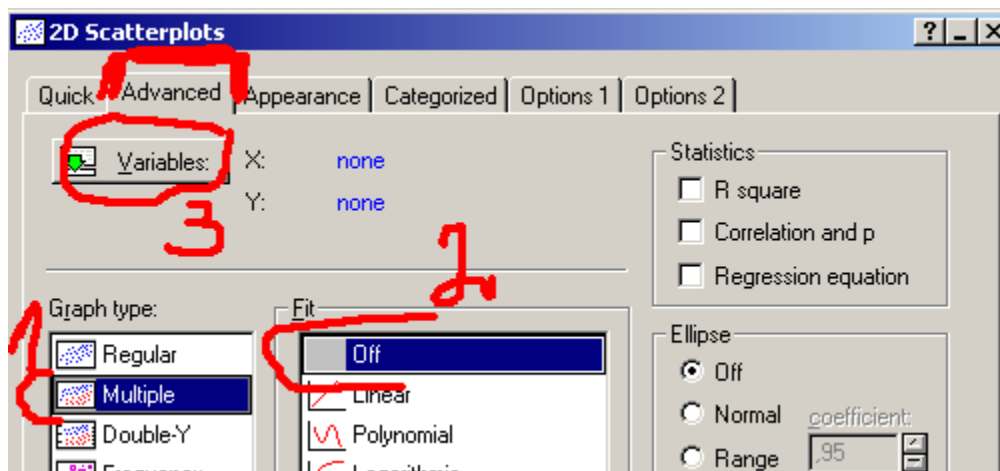
1 užduotis.

Pateikite grafiškai pomidorų diametro ir aukščio vidurkius eksperimento eigoje, esant skirtingam jų augimo tankumui.

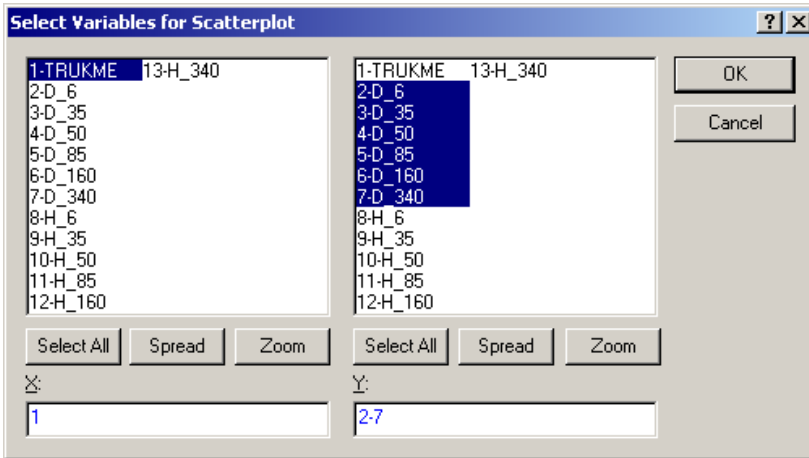
Programos „*Statistica*“ pagrindiniame meniu renkamės komandą „*Graphs*“, toliau „*Scatterplots...*“



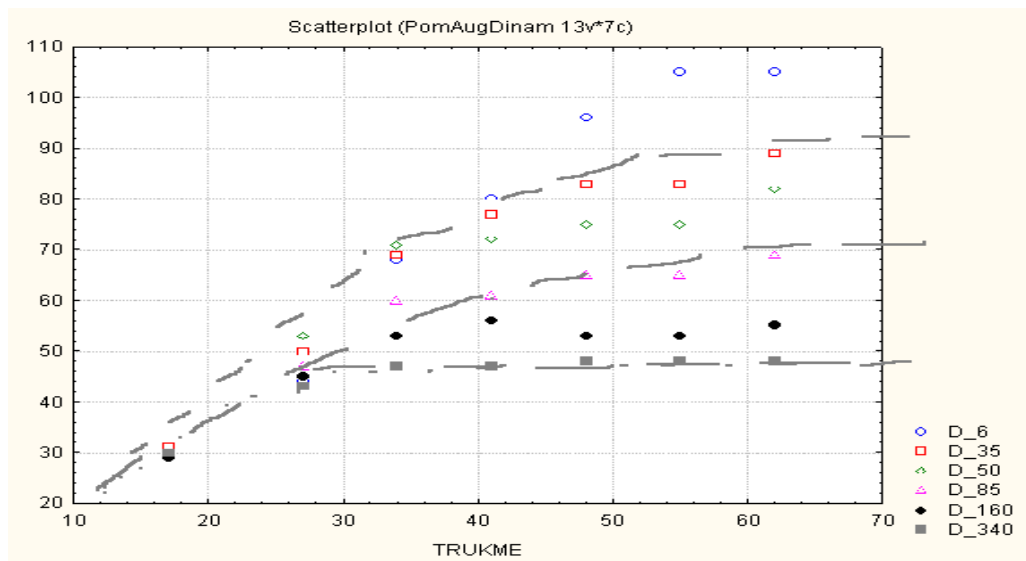
„*2D Scatterplots*“ lange pasirenkame „*Advanced*“ lapą, kuriame renkamės grafiko tipą (*Multiple*), bei kintamuosius „*Variables*“



Kintamuosius žymime sekanciai:



Atsakymas pateikiamas rezultatų lange:



Kaip matote, eksperimento metu sukaupti duomenys apie diametro storį (D) gerai aproksimuotūsi (eksponentine) Micherlich'o augimo kreive, kuri matematiškai aprašoma sekančiai:

$$Y = A (1 - e^{-bT})^c$$

čia:

Y- morfometrinis augalo rodiklis;

T- augalo amžius;

e- natūrinis logaritmas

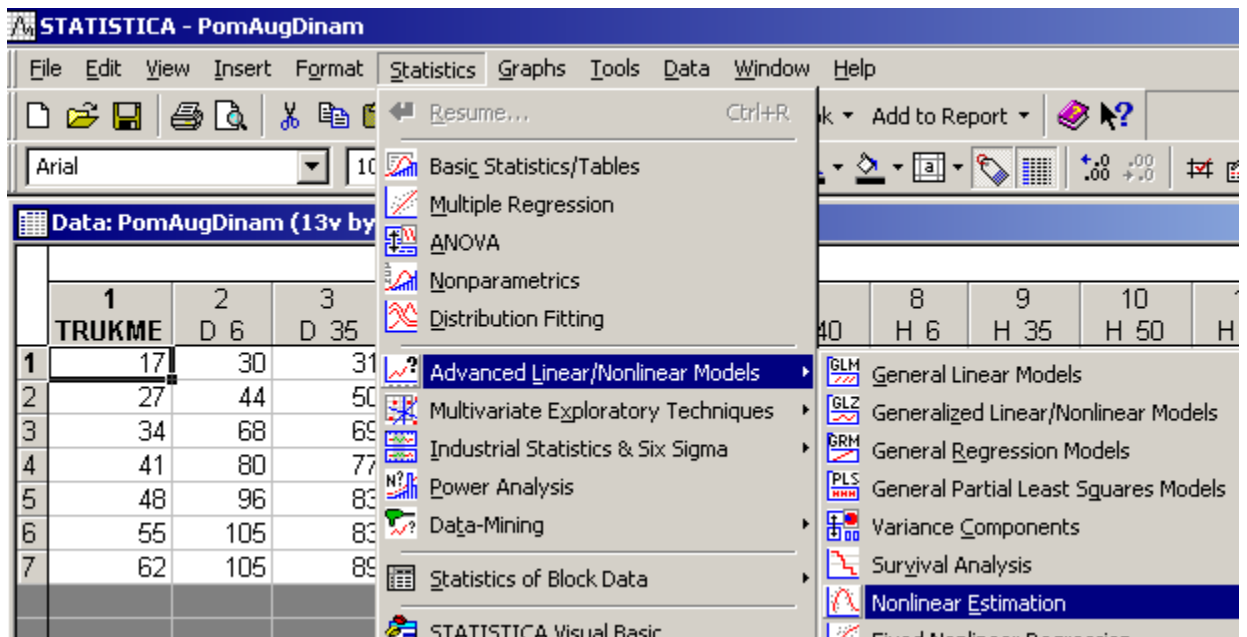
A, b, c – augimo funkcijos koeficientai.

2 uždutis.

Naudodami “*Nonlinear Estimation*” analizės modulį pritaikykite Micherlich'o augimo tipo kreivę pateiktiems eksperimento duomenims:

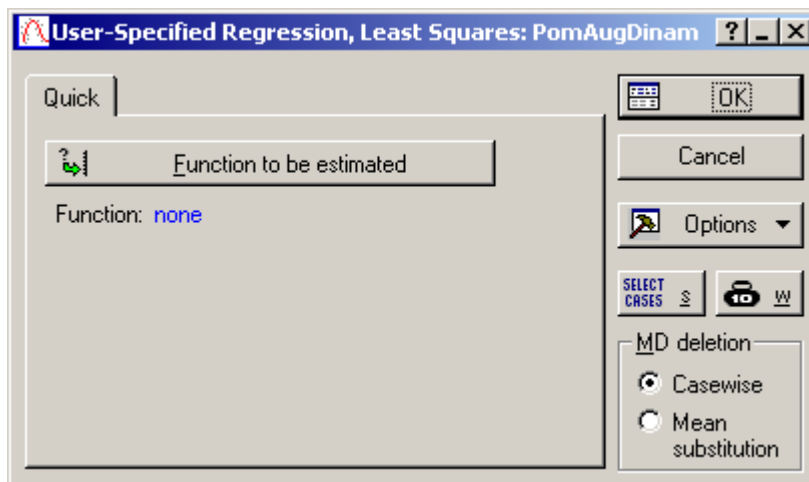
a) 6 indiv./m² tankumu augusių pomidorų diametriui.

Programoje „Statistika“ pasirenkame „*Statistics*“ „*Advanced Linear/ Nonlinear Models*“ „*Nonlinear Estimation*“:

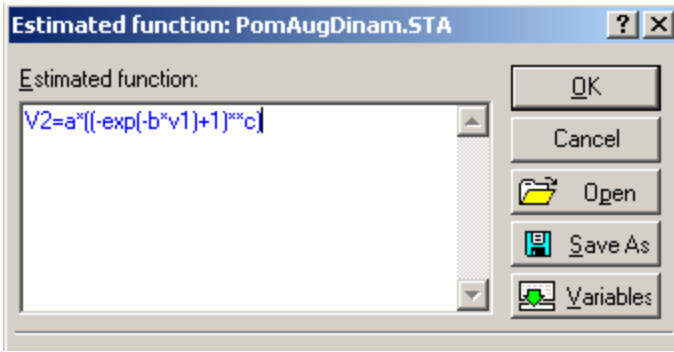


„*Nonlinear Estimation*“ analizės lange renkamės „*User-specified regression, least squares*“ „OK“

Naujai atsidariusiame lange renkamės „*Function to be estimated*“



Micherlich'o augimo kreivę parašome sekančiai: $V2=a*((-\exp(-b*v1)+1)**c)$



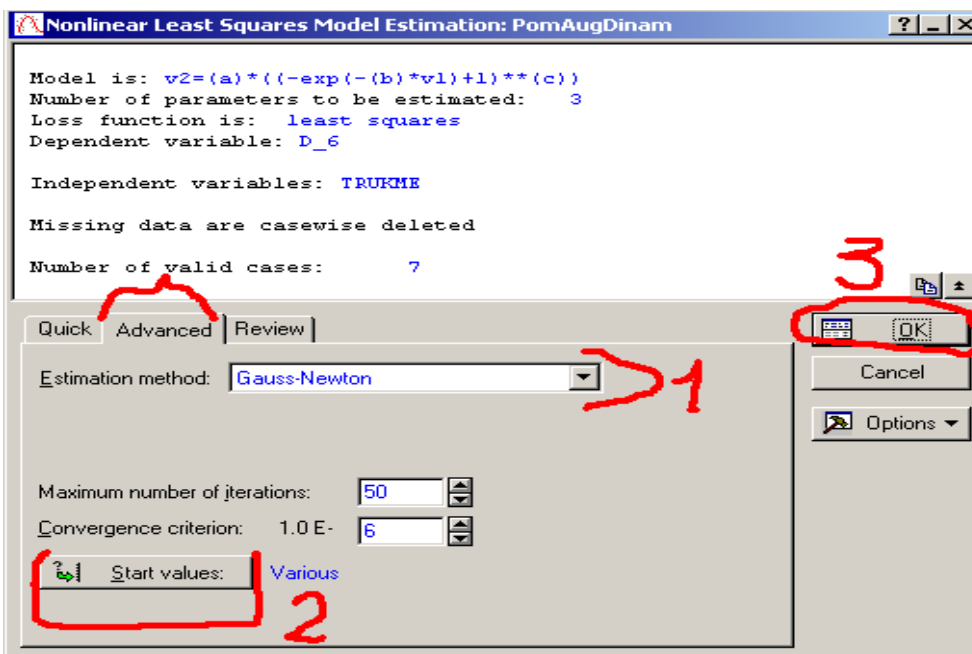
čia:

v2 (arba galima rašyti D_6) – 6 indiv./m² diametro vidurkių kaita eksperimento eigoje, tai „Y“ kintamasis.

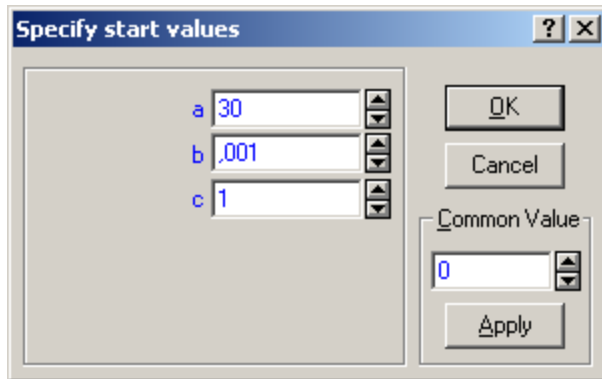
a, b ir c – A, b ir c augimo funkcijos koeficientai.

v1 (TRUKME) – eksperimento trukmė paromis, šiuo atveju „x“ kintamasis.

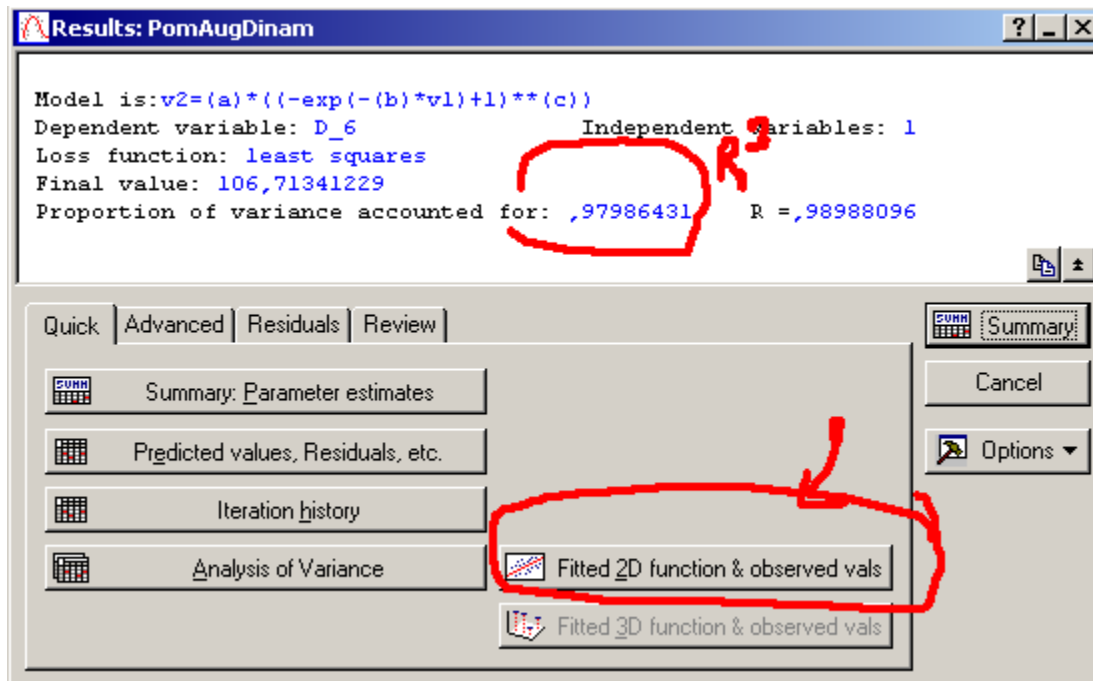
Suvedę formulę spaudžiame „OK“. Toliau analizės lange atsidarome „Advanced“ lapą ir pasirenkame kaip nurodyta žemiau:



Pasirinkus „Start values“ komandą, reikia rašyti tokias koeficientų skaičiavimo startines reikšmes:

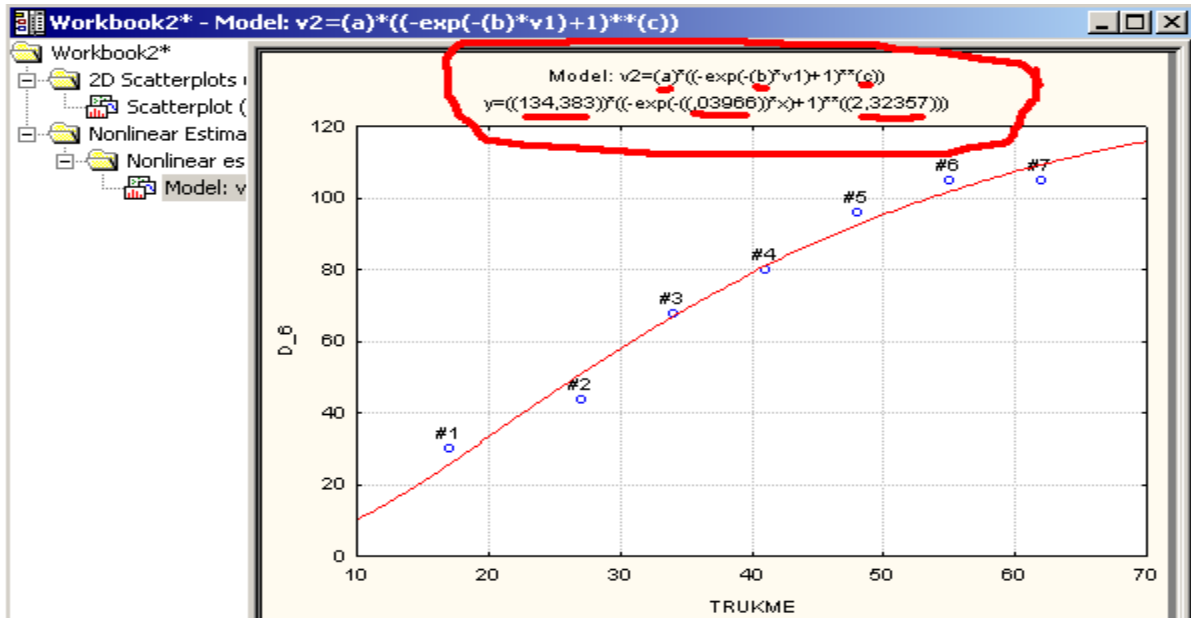


Spaudžiame „OK“.



Skaičiavimų lange jau pateikiamas modulio determinacijos koeficientas R^2 ir koreliacijos koeficientas R . Pastarasis $R=0.9899$ rodo, kad yra stiprus koreliacinis ryšys tarp pomidorų amžiaus (eksperimento trukmė) ir diametro storio. O R^2 šiuo atveju lygus 0,97986. Koeficiento reikšmė nurodo, kad apie 98% eksperimento duomenų (6 indiv./m² tankumu augusių pomidorų diametrai) kito pagal Micherlich'o augimo kreivę, t.y. šio tipo kreivė statistiškai patikimai aproksimuoja pomidorų diametro kitimą laiko eigoje.

Toliau pasirinkus komandą „*Fitted 2D function & observed vals*“, atsakymų lange pateikiamas aproksimuotos kreivės grafikas ir koeficientų reikšmės (pažymėta raudonai).

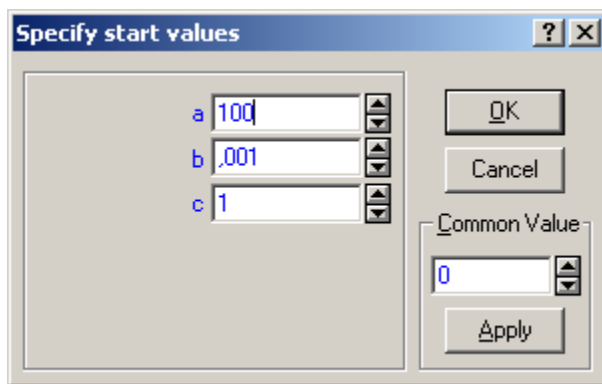


b) Remdamiesi aukščiau pateiktu aprašu, pritaikykite Micherlich'o augimo tipo kreivę ir likusių tankumų (35, 50, 85, 160, 340 indiv./m²) pomidorų diametrams bei pasižymėkite sudarytų kreivių determinacijos koeficientų R² reikšmes.

c) Analogiškai pritaikykite Micherlich'o augimo tipo kreivę visų tankumų pomidorų aukščiui.

Skaičiuojant aukščio kitimo kreivių koeficientus yra vienas skirtumas!!!!!!

Pasirinkus „*Start values*“ komandą, reikia rašyti kitas koeficientų skaičiavimo startines reikšmes:

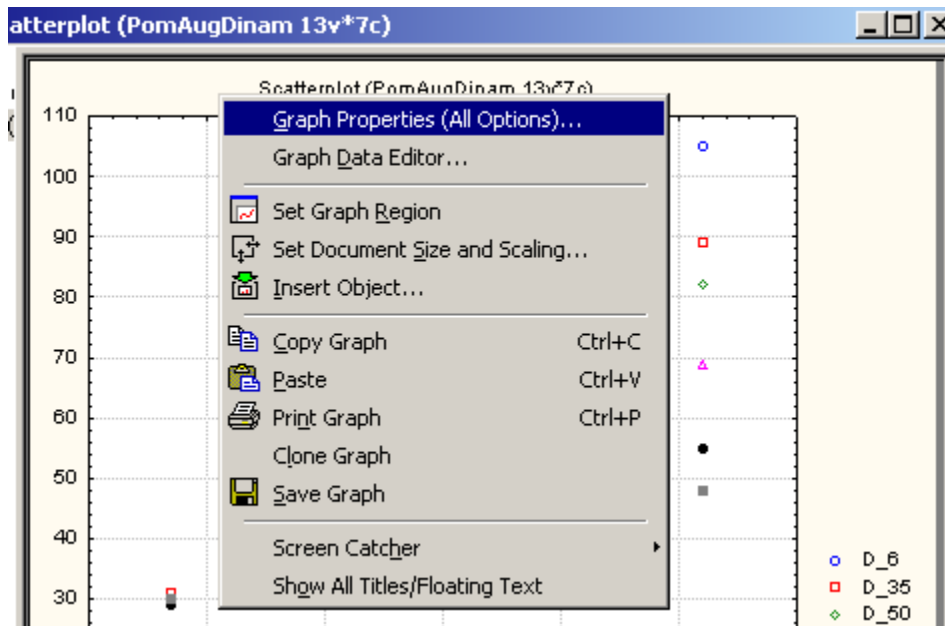


Visa kita darbo eiga nekinta.

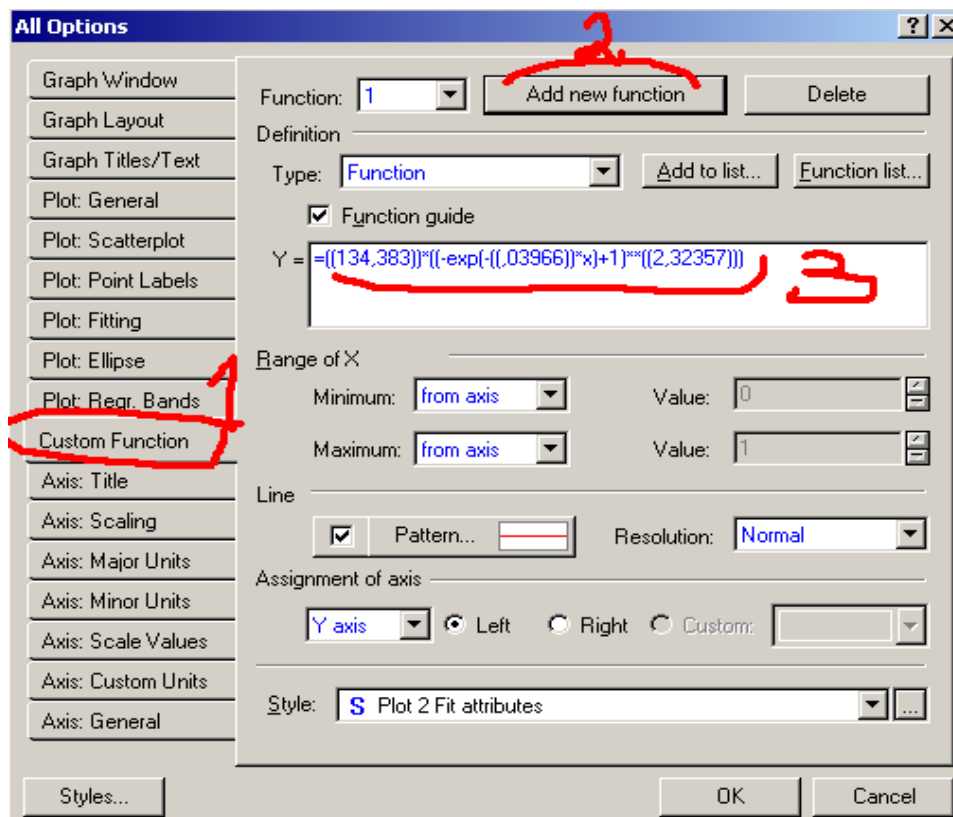
3 uždutis.

Gautas funkcijas ir jų determinacijos koeficiento R² reikšmes įkelkite į pirmoje užduotyje gautą grafiką.

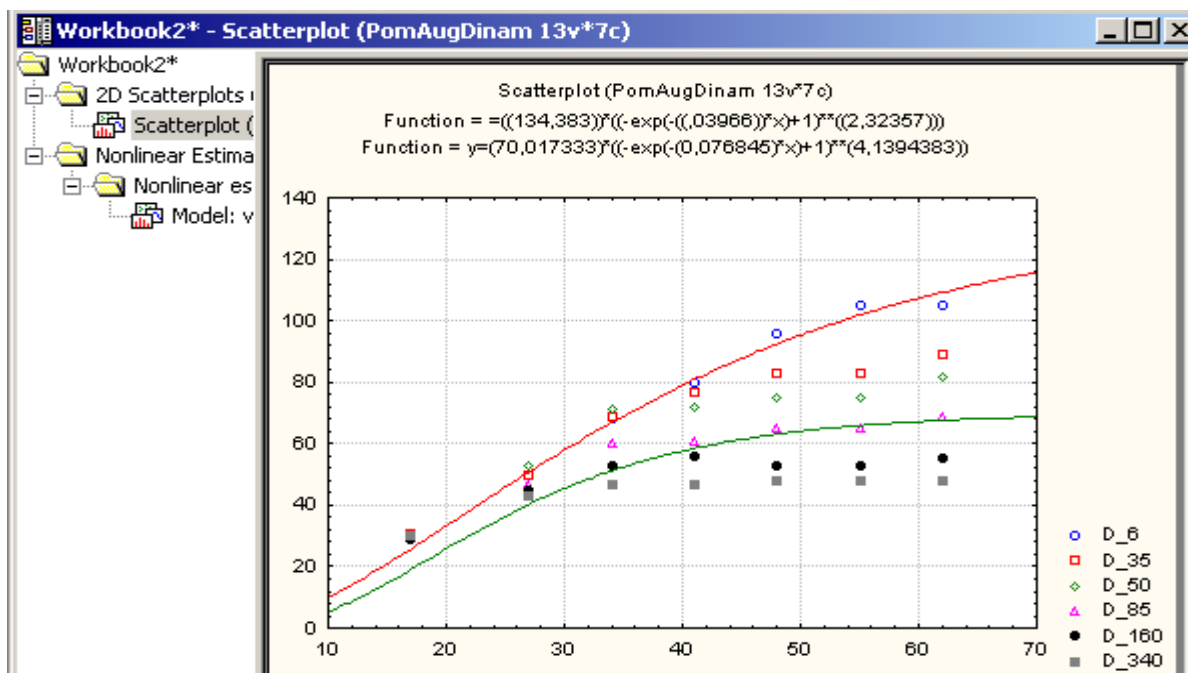
Pasirenkame pirmoje užduotyje nubražytą grafiką. Spaudžiame pelės dešinį klavišą ir pasirenkame komandą „*Graph Properties (All Options)...*“



Toliau „*All Options*“ lange renkamės „*Custom Function*“ lapą, kuriame spaudžiame „*Add new function*“. Funkcijos lange įkeliamo D_6 tankumo diametro augimo lygtį, kurią gavome atliekant 2 užduoties a) dalį. Spaudžiame „*OK*“.



Darbą kartokite kol sukelsite visas funkcijas.



4 uždutis.

Atsakykite šiuos klausimus:

Kokie yra gautų pomidorų diametro ir aukščio augimo kreivių determinacijos koeficientai?

Tiek procentų eksperimento duomenų atitinka Micherlich'o augimo kreivės kitimą?

Kaip kito pomidorų parametrai eksperimento eigoje, esant skirtingiems tankumams?

Klausimai savikontrolei:

Kaip suprantate netiesinį ryšį?

Kokių tipų gali būti netiesinio ryšio regresijų?

Kas yra determinacijos koeficientas? Kokios gali būti jo reikšmės?

Literatūra:

Vencloviene J. 2008. Statistiniai metodai aplinkotyroje. Kaunas. 225p.

Sakalauskas V. 2003. Duomenų analizė su STATISTICA. Vilnius. 235 p.

Parengė: APL 416. APLINKOS DUOMENŲ ANALIZĖS IR MODELIAVIMO praktinių darbų dėstytoja dr. I. Januškaitienė.

2008 m.

Kaunas, VDU, Gamtos mokslų fakultetas.