

VILNIAUS UNIVERSITETAS
EKONOMIKOS FAKULTETAS

EKONOMETRIJA
Paskaitų konspektas

Vita Karpuškienė

Vilnius
2018

ĮVADAS

Nūdienos mokslo raidai būdinga tai, kad matematika vis labiau skverbiasi į įvairias mokslo šakas, pvz., fizika, medicina, gamtos mokslai ir kt. Socialiniai mokslai, tame tarpe ir ekonomika, taip pat nėra išimtis.

Matematinis metodas ekonominių procesų analizėje pradėta taikyti gana senai. Pavyzdžiui, anglų ekonomistas Viljamas Peti savo knygoje “Politinė aritmetika” (1676) siūlė pereiti prie griežtesnio minčių reiškimo skaičiais, svoriais, matais.

Pirmąjį pasaulyje ūkio modelį sukūrė prancūzų mokslininkas Fransua Kenė. 1756 m. paskelbtame darbe “Ekonominė lentelė” autorius ūkio reprodukcijos procesus nagrinėjo pasitelkdamas matematinę lygtį.

Prancūzų mokslininko Ogiusto Kurno darbas “Turto teorijos matematinų principų tyrimas”, išleistas 1838 metais, faktiškai suformavo klasikinę matematikos mokyklą ekonomikoje, kuri padarė labai didelę įtaką ekonomikos mokslo raidai. XX amžių galima būtų pavadinti ekonomikos matematizavimo amžiumi.

Ekonomikos ir matematikos sąveika nėra vienpusė. Ne tik matematiniai metodai skverbiasi į ekonomikos mokslą, bet yra ir priešingas poveikis. Ekonominiai uždaviniai ir problemos skatina specializuotų matematikos šakų, tokių kaip matematinis programavimas, lošimų teorija, masinio aptarnavimo teorija, aktuarijų (draudimo) matematika, atsiradimą. Glaudžioje ekonomikos ir matematikos mokslų sąveikoje susiformavo ir ekonometrija. **Ekonometrija - tai ekonomikos teorijos ir matematinės statistikos junginys.** Statistika ekonominių duomenų analizei pradėta taikyti gana senai. 1699 m. Charles Davenant* paskelbė pirmąją “empirinę” paklausos duomenų lentelę 1707 m. italų statistikas Rudolfo Enini* atliko labai novatorišką paklausos statistinę analizę. Visgi esminis ekonometrijos, kaip mokslo šakos atsiradimo etapas buvo norvegų ekonomisto ir matematiko Ragnaro Frišo darbų paskelbimas, kuriuose 1926 m. pirmą kartą paminėtas ir pats “ekonometrijos” terminas. 1930 m. susibūrė ekonometrijos draugija, o 1933 m. sausį pradėtas leisti mokslinis žurnalas “Econometrica”, kuris yra leidžiamas iki šiol.

Ekonometrijos apibrėžimai yra gana įvairūs, pradedant nuo labai plačių, įtraukiančių įvairiausių ekonominius matematinis metodus, iki gana siaurų, kai apsiribojama keliais matematinės statistikos metodais.

* S. J. Stigler, Essays in the History of Economics, Chicago: University of Chicago Press, 1965.

Garsių rusų ekonomistų ir matematikų S. A. Aivazian ir V S. Mkhitariano vadovėlyje**

pateikiamas toks apibrėžimas:

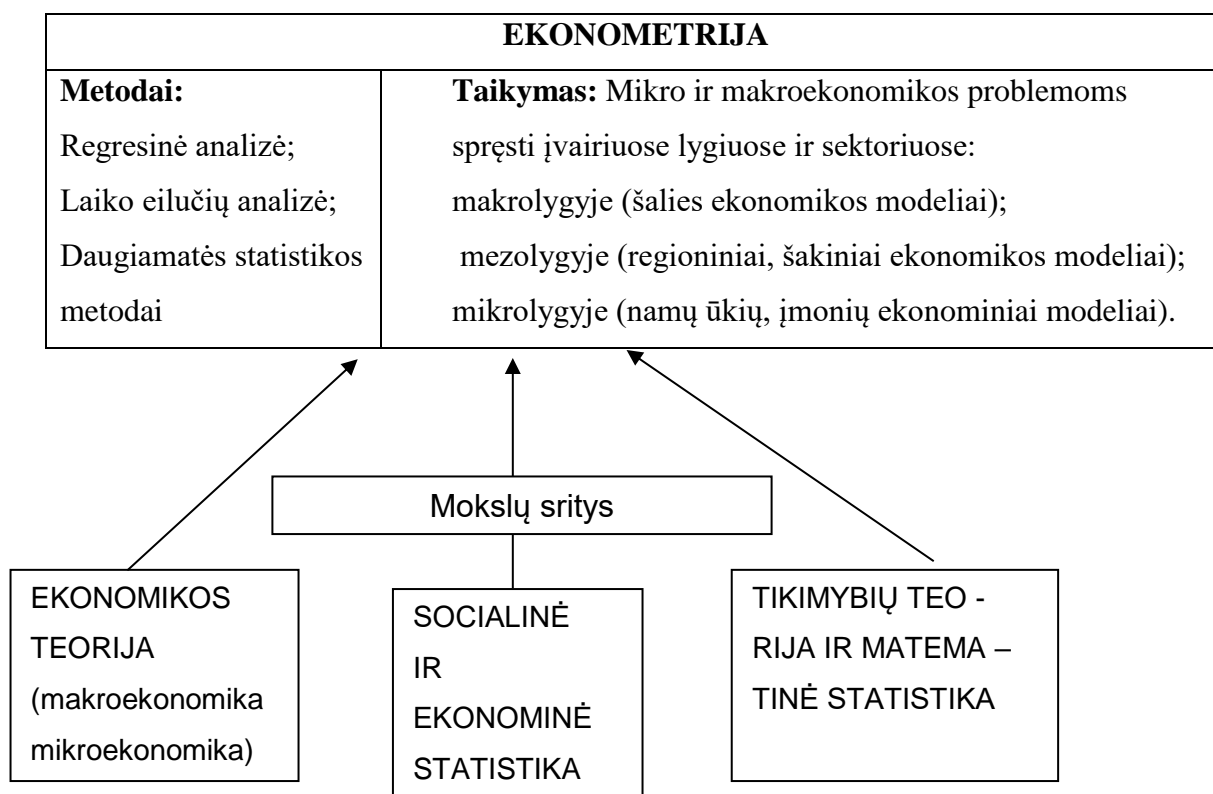
Ekonometrija - tai mokslas, apimantis ekonominių procesų kiekybinę analizę, atliekant ją ekonomikos teorijos, ekonominės bei matematinės statistikos metodais.

D.Gujaraty vadovėlyje "Basic econometrics" pateikia siauresnį apibrėžimą.

Ekonometrija - tai atskira disciplina , kuri apjungia ekonomikos teoriją ir statistinius metodus, siekiant suteikti skaitines reikšmes ekonominiams procesams.

Dažniausiai ekonometrijos vadovėlių turinį sudaro du metodai: regresinė analizė ir laiko eilučių tyrimo metodai, prieš tai siauriau ar plačiau supažindinant su klasikinės matematikos, tikimybių teorijos bei matematinės statistikos pagrindais.

Apibendrinus, ekonometrijos mokslo turinį ir taikymo sritis galima pavaizduoti tokia schema.



1 pav.

Pateikiamas paskaitų konspektas taip pat susideda iš dviejų pagrindinių metodų - regresinės analizės ir laiko eilučių nagrinėjimo. Regresinė analizė padės studentams atsakyti į klausimus, kokie veiksniai ir kaip veikia konkretų ekonominį reiškinį.

** (S.A. Aivazian, V.S. Mkhitarian, Prikladnaja statistika I osnovy ekonometriki. Moskva, 1998)

Laiko eilučių nagrinėjimas leis nustatyti ekonominių reiškinių pagrindines tendencijas, ciklinius bei sezoninius svyravimus ir kitus svarbius aspektus.

Pagrindiniai lietuviški ekonometrijos vadovėliai yra šie:

1. Martišius S. Ekonometrija ir prognozavimas. - V.:VU I-kla, 2000
2. Martišius S. Regresinės ir koreliacinės analizės metodai. – V.: VU I-kla, 1992.
3. Čekanavičius V., Gedinimas Murauskas. Statistika ir jos taikymas. II.-Vilnius TEV, 2004
4. Stankus E. Ekonometrija V: Vilniaus vadybos kolegija, 2002
5. Boguslauskas V., Ekonometrija, Kaunas: Technologija, 1999,
6. Kruopis J. Matematinė statistika. – V.: Mokslo ir enciklopedijų I-kla, 1993.
7. Mišeikis F. , Statistika ir ekonometrija, Vilnius, “Technika”, 1997,


Vadovėliai anglų kalba

G.S. Maddala, Introduction to Econometrics, second edition, Prentice Hall International, Inc., 1992,

Hill C, Grffiths W., Judge G., Undergraduate Econometrics. John Wiley&Sons, Inc.1997.

Gujarati D.N. Basic Econometrics. – New York, etc.: McGraw-Hill, 1995

Minėtais vadovėliais remtasi ruošiant šį paskaitų konspektą.

Studentai , skaitydami paskaitų konspektą pastebės ženklą  . Šiuo sunkumų kilnotojo ženklu išskirtos tos teksto vietos, kurios yra skirtos studentams, norintiems daugiau sužinoti apie dėstomą medžiagą.

Tačiau nebus įtrauktos į egzamino klausimus.

REGRESINĖ ANALIZĖ IR PROGNOZAVIMAS

Regresinė analizė ir prognozavimas (I)

Paskaitos dalys:

1. Regresija – ryšio analizės priemonė.
2. Porinė tiesinė regresija.

Pagrindiniai terminai:

Regresija – ryšio analizės priemonė

- Regresija
- Priklausomas / nepriklausomas kintamasis
- Tiesinė / netiesinė regresija
- Porinė / dauginė regresija
- Parametras
- Įvertis
- Nepaslinktas įvertis
- Efektyvus įvertis

Porinė tiesinė regresija

- Mažiausių kvadratų metodas
- Paklaida
- Paklaidos įvertis
- Sąlyginė prognozė

1, Regresija – ryšio analizės priemonė

Regresinės analizės samprata

Kiekvieną ekonominį reiškinį veikia bent keli veiksniai. Pvz. obuolių kainai ūkininkų turguje turės įtakos, obuolių auginimo ir sandėliavimo kaštai, importuojamų obuolių kainos, kitų vaisių, kurie gal būti obuolių pakaitalais, pvz. kriaušių, kainos. Atsitiktiniai veiksniai, tokie, kaip oro sąlygos taip pat turės įtakos obuolių derliui. Priimant sprendimus dažnai neužtenka vien tik išvardinti nagrinėjamą ekonominį reiškinį sąlygojančius veiksnius, bet reikia jų poveikį įvertinti kiekybiškai. Šiam tikslui yra naudojama regresinė analizė, kurios pagalba veiksnių įtaką nagrinėjamam ekonominiam reiškiniui galima užrašyti matematinės lygties pagalba. Turint matematinę priklausomybės lygtį, galima:

- (1) Parodyti ekonominio reiškinio susiformavimo mechanizmą,
- (2) Matematiškai aprašyti nagrinėjamo ekonominio reiškinio priklausomybę nuo jį sąlygojančių veiksnių.
- (3) nustatyti reikšmingus ir nereikšmingus veiksnius;
- (4) prognozuoti nagrinėjamo ekonominio reiškinio variantus

- (5) modeliuoti įvairias situacijas ir stebėti, kaip kinta nagrinėjamas reiškinys, kintant veiksniams.
- (6) Konkreti regresijos ryšio matematinė forma leidžia gauti ekonominės analizės išvadoms naudingus rodiklius: ryšio ženklą ir pobūdį, nagrinėjamo reiškinio elastingumą kiekvienam iš veiksnių arba visų veiksnių poveikiui bendrai.

Regresinė analizė yra statistinis metodas, kai naudojant matematinės procedūras, gaunama lygtis arba jų sistema, įvertinanti vieno ar daugiau veiksnių įtaką nagrinėjamas reiškiniui. Gauta matematinė lygtis yra vadinama **regresijos lygtimi** arba tiesiog **regresija**. Bendras regresinės lygties pavidalas atrodo taip:

$$y_i = f(x_{1i}, \dots, x_{ki}) + \varepsilon_i \quad 1,1$$

kur y_i – nagrinėjamas ekonominis reiškinys, x_{1i}, \dots, x_{ki} – jį sąlygojantys veiksniai, ε_i – regresijos paklaida, i – stebėjimo numeris.

Kiekviena regresija susideda iš dviejų dalių: **sisteminės** ir **atsitiktinės**. Sisteminė dalis lygtyje 1.1 yra $f(x_{1i}, \dots, x_{ki})$, o ε_i – atsitiktinė

Svarbu įsidėmėti, kad regresinė analizė taikoma ekonominių reiškinų analizei, kuriuose veiksnių tarpusavio ryšiai yra tikimybiniai ir netaikoma esant griežtai determinuotam veiksnių sąryšiui. Pvz., prognozuojant pramogų centrų plėtrą, regresinė analizė padės išsiaiškinti paslaugų paklausą, t.y., kiek vidutiniškai įvairios gyventojų grupės pvz., studentai, šeimos su vaikais ar kokio nors miesto gyventojai ir t.t pramogoms išleidžia pinigų, jeigu jų pajamos yra 400, 500 ar 1000 litų per mėnesį. Sąryšis tarp šių dviejų veiksnių yra tikimybinis, nes vienas gyventojas pramogoms išleidžia daugiau, kitas mažiau turėdami net ir vienodas pajamas. Tyrinėtojas gali tik vidutiniškai pasakyti, kiek pinigų per mėnesį išleis gyventojų grupė. Regresinė analizė nebus reikalinga, jeigu žinosime, visų gyventojų mėnesio išlaidas pramogoms prognozuojamu periodu. Šiuo atveju atsakymą gausime susumavę visų gyventojų išlaidas. Tačiau tokios informacijos nėra ir vargu ar gali būti.

Regresinio modelio sudarymo etapai ir žingsniai.

Galima skirti tris regresinio modelio sudarymo etapus

1. Ekonominis modelis
2. Statistinis modelis
3. Ekonometrinis modelis

Kiekviename etape atliekami tam tikri žingsniai ir daromos atitinkamos išvados. Regresinio modelio etapus papildžius žingsniais regresinio modelio sudarymo procedūra atrodo taip:

1 etapas: EKONOMINIS MODELIS

Pirmas žingsnis: Ekonominės problemos formulavimas

Rezultatas: Įvardinamas nagrinėjamas reiškinys ir jį įtakojuantys veiksniai

Antras žingsnis: Ekonominių hipotezių iškėlimas

Rezultatas: Užduodami reikalavimai matematinei modelio išraiškai

Trečias žingsnis: Duomenų rinkimas

Rezultatas: Sudaromos nagrinėjama reiškinį ir įtakojančius veiksnius apibūdinančios duomenų lentelės.

2 etapas: STATISTINIS MODELIS

Ketvirtas žingsnis: Grafinė duomenų analizė

Rezultatas: Nagrinėjamo reiškinio priklausomybės nuo atskirų veiksnių grafikai. Tai svarbi informacija modelio matematinės formos parinkimui

Penktas žingsnis: Modelio matematinės išraiškos užrašymas

Rezultatas: Užrašoma matematinė modelio lygtis (lygtys)

Šeštasis žingsnis: Parametrų įverčių skaičiavimas

Rezultatas. Užrašomas modelis su skaitiniais koeficientais

Septintasis žingsnis: Veiksnių statistinio reikšmingumo analizė

Rezultatas. Atsakoma į klausimą, kurie veiksniai statistiškai reikšmingai įtakoja nagrinėjamą reiškinį.

Aštuntasis žingsnis: Viso modelio patikimumo tikrinimas

Rezultatas. Atsakoma į klausimą ar modelį galima taikyti ekonominei analizei ir prognozavimui.

3 etapas EKONOMETRINIS MODELIS

Devintas žingsnis: Ekonominės problemos analizė naudojant apskaičiuotas modelio parametrų įverčius ir kt. skaitines charakteristikas

Rezultatas. Modelio pagalba daromos ekonominės išvados, kurių negalima būtų gauti, betarpiško ekonominio reiškinio stebėjimo ar kitu būdu.

Dešimtas žingsnis: Ekonominių scenarijų kūrimas, prognozavimas.

Pavyzdys

Privati gyventojų įdarbinimo bendrovė "Turbo karjera" turi savo atstovybes 6 miestuose. Kiekvienos atstovybės biuro išlaikymo sąnaudos priklauso nuo daugelio veiksnių: Vienas iš svarbiausių -

tai darbuotojų skaičius, nuo kurio priklauso išlaidos darbo užmokesčiui ir įvairūs mokesčiai valstybei. Greta darbuotojų skaičiaus biuro kaštai apima ir kitas sąnaudas - tai išlaidos patalpų nuomai, šildymui, ryšio ir kt. paslaugoms, kurių dydis priklauso nuo jų įkainių.

Įdarbinimo bendrovės biurų išlaikymo kaštai, darbuotojų skaičius ir kitų veiksnių tarifai pateikti 1.1 lentelėje.

1.1 lentelė Įdarbinimo bendrovės duomenys

Veiksniai	Atstovybės					
	Vilniuje	Vilniuje	Kaune	Klaipėdoje	Šiauliuose	Panevėžyje
Biuro išlaikymo sąnaudos tūkst. Eurų metams (y_i)	100	90	140	130	120	100
Darbuotojų skaičius (x_1)	10	8	14	12	11	7
Patalpų nuomos tarifai (eur/m ²) (x_2)	45	45	20	28	18	22
Ryšio paslaugų tarifai (cent/min.) (x_3)	65	50	54	65	56	71

Regresinė analizė padės atsakyti į klausimą, kaip vidutiniai keičiasi biuro išlaikymo kaštai, kintant darbuotojų skaičiui, ryšių paslaugų ir patalpų nuomos tarifams. Pažymėkime y_i - biuro išlaikymo kaštus i -oje atstovybėje, X_{1i} - i -os atstovybės darbuotojų skaičių, X_{2i} - patalpų nuomos tarifus i -oje atstovybėje,

$$y_i = f(x_{1i}, x_{2i}, x_{3i}) + \varepsilon_i$$

X_{3i} - vidutinius ryšio paslaugų tarifus i -oje atstovybėje, ε_i - i -ojo biuro kaštų paklaida. Bendras regresijos lygties pavidalas bus toks:

Pagrindinės sąvokos ir simboliai

Priklausomi ir nepriklausomi kintamieji

Apibrėžiant regresijos sampratą naudojome sąvokas "nagrinėjamas ekonominis reiškinys ir jį sąlygojantys veiksniai. Toliau pereisime prie trumpesnių ir griežtesnių terminų.

Priklausomas kintamasis - tai kairėje regresijos lygties pusėje esantis kintamasis (y_i lygtyje 1.1.), kurio vidutinių reikšmių pokyčius stengiamasi paaiškinti kitų - dešinėje lygties pusėje esančių kintamųjų pokyčiais.

Nagrinėjamame pavyzdyje priklausomas kintamasis - tai biuro išlaikymo kaštai.

Nepriklausomi kintamieji (x_{ji}) - tai dešinėje lygties pusėje esantys kintamieji, kurie veikia priklausomąjį kintamąjį (y_i) Nepriklausomųjų kintamųjų reikšmės gali laisvai kisti, t.y. jų neįtakoja priklausomojo kintamojo reikšmės. Pavyzdyje nepriklausomi kintamieji - tai darbuotojų skaičius, patalpų nuomos, ryšių ir šildymo paslaugų tarifai..

Ų Regresinės funkcijos kintamiesiems apibrėžti naudojamos ir sąvokos *endogeninis* bei *egzogeninis* kintamasis. Dažniausiai jų turinys yra tapatus sąvokoms priklausomas ir nepriklausomas kintamasis. Tačiau jis skiriasi tuo atveju, kai nagrinėjama lygčių sistema su vėluojančiomis regresijomis, kuriose endogeninis kintamasis gali būti dešinėje lygties pusėje greta nepriklausomų kintamųjų

$y_i = f(x_{1i}, \dots, x_{ki}, y_{i-1}, \dots, y_{i-p}, \varepsilon_i)$, kur y_i – endogeninis kintamasis, x_{1i}, \dots, x_{ki} – egzogeniniai kintamieji, y_{i-1}, \dots, y_{i-p} – vėluojantys priklausomi kintamieji, p – vėlavimo periodų skaičius, ε_i – regresijos paklaida, i – stebėjimo numeris¹. Ų

Grafinis priklausomybės pateikimas

Pieš atliekant regresinę analizę visada patartina grafiškai pavaizduoti kintamųjų tarpusavio sąryšius. Tai leidžia daryti preliminarias išvadas apie kintamųjų tarpusavio ryšio matematinės priklausomybės formą. Žinia, kad paprasčiausia nubrėžti grafiką dvimatėje erdvėje, kai priklausomas kintamasis yra susietas tik su vienu kintamuoju. Jeigu nagrinėjamas reiškinys priklauso nuo kelių kintamųjų, tuomet rekomenduotina nubrėžti priklausomojo kintamojo grafikus su kiekvienu iš nepriklausomų kintamųjų. Šiuo atveju, dvimačių grafikų būtų tiek, kiek yra nepriklausomų kintamųjų.

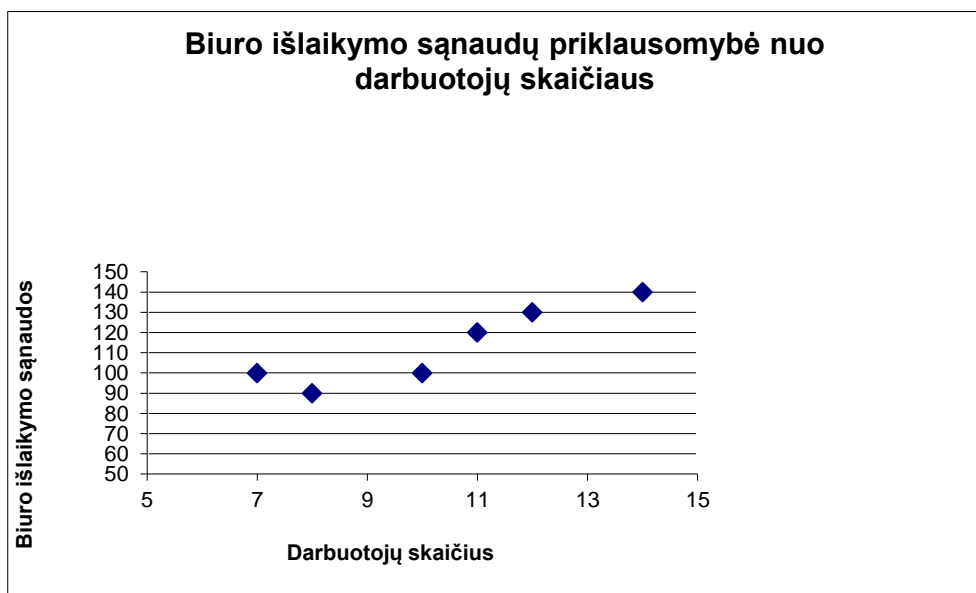
Pavyzdys

Grįžkime prie įdarbinimo bendrovės "Turbo karjera" pavyzdžio. Paprastumo dėlei pradžioje apsiribokime biuro kaštų priklausomybės analize tik nuo vieno veiksnio - darbuotojų skaičiaus.

Veiksnių priklausomybę galima pavaizduoti grafiškai, kaip parodyta 1,1 paveiksle, kuriame dvimatės koordinatinių sistemos vertikalioje ašyje yra atidėtos priklausomojo kintamojo (y_i) - biuro išlaikymo sąnaudų, o horizontalioje - nepriklausomojo kintamojo (x_{1i}) - darbuotojų skaičiaus reikšmės. Grafinis taškų išsidėstymas suteikia naudingos informacijos nustatant matematinę veiksnių priklausomybės formą, pvz., stebime, kaip keičiasi (didėja/mažėja) priklausomojo kintamojo (y_i) reikšmės, didėjant nepriklausomojo kintamojo (x_i) reikšmėms; ar yra lūžio taškai, kurie priklausomojo kintamojo didėjimo (mažėjimo) intervalą pakeičia mažėjimo (didėjimo) intervalu,

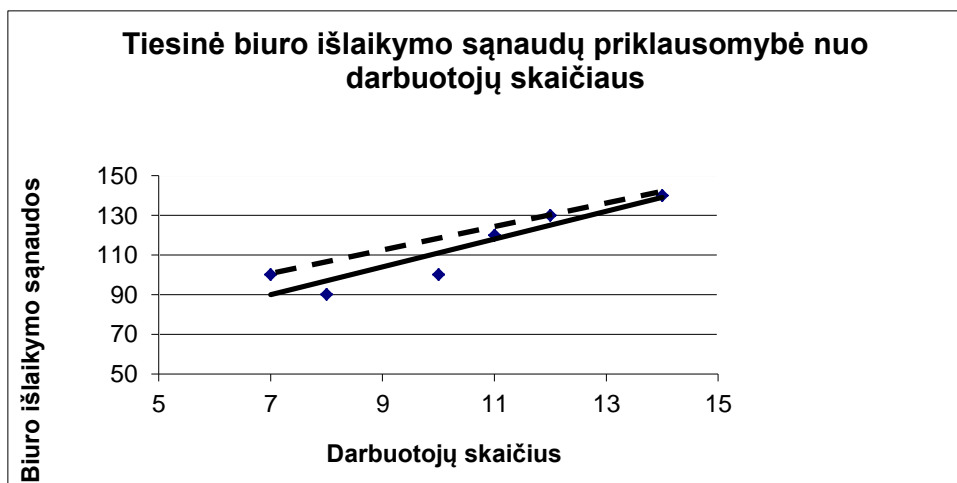
Nagrinėjamame pavyzdyje biuro išlaikymo kaštai sumažėja, darbuotojų skaičiui išaugus nuo 7 iki 8, tačiau toliau didėjant darbuotojų skaičiui, biuro kaštai taip pat didėja.

¹ Jei tiriama ne skerspjū, o *laiko eilučių* regresija, tai indeksas žymimas ne raide i , o raide t .



1,1 pav.

Per atidėtus taškus galima nubrėžti linija, kuri atvaizduoja biuro kaštų ir darbuotojų skaičiaus priklausomybę. Pati paprasčiausia matematinės priklausomybės forma - tiesinė. Prisiminkime iš mokyklos 5 klasės kurso tiesės lygtį: $y=b_0 + b_1x$. Per atidėtus taškus tiesę galima brėžti įvairiais būdais. 1,2 paveiksle nubrėžtos dvi tiesės: punktyrinė - jungia arčiausiai ir toliausiai nuo koordinatinių pradžių nutolusius taškus; o ištisinė, nubrėžta taip, kad atstumai nuo pažymėto taško iki tiesės būtų mažiausi.



1.2 pav.

Porinė ir dauginė regresija

Porinė regresija yra tokia regresija, kai vertinamas dviejų kintamųjų tarpusavio ryšys. Nagrinėjamame pavyzdyje porinė regresija parodo biuro kaštų priklausomybę nuo vieno veiksnio pvz., darbuotojų skaičiaus.

1.2 paveiksle nubrėžtos tiesinės priklausomybės matematinė regresijos išraiška atrodytų taip:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad 1.2$$

kur y_i - tai i -ojo biuro išlaikymo kaštai, kai $i = \overline{1,6}$, x_i - darbuotojų skaičius tame biure, β_0 ir β_1 - koeficientai (literatūroje dažnai naudojamas terminas - parametrai), kuriuos reikia apskaičiuoti, norint surasti regresijos lygtį, ε_i - regresijos paklaida (i -ojo taško nukrypimas nuo regresijos tiesės) $(\beta_0 + \beta_1 x_i)$ - tai sisteminė regresinės lygties dalis, kuri aprašo biuro kaštų priklausomybę nuo darbuotojų skaičiaus.

ε_i - tai atsitiktinė dalis, kurioje atsispindi įvairių atsitiktinių veiksnių įtaka kaštams.

Dauginė regresija tai regresija, kurioje nepriklausomų kintamųjų yra daugiau nei vienas. Dauginės regresijos pagalba galima tirti daugelio veiksnių bendrą įtaką nagrinėjamam reiškiniui. Bendrą įtaką suformuoja visų veiksnių poveikių suma. Atskiro veiksnio įtaka yra vadinama daline ir nustatoma, darant prielaidą, kad kitų nepriklausomų kintamųjų reikšmės yra pastovios.

Įdarbinimo bendrovės pavyzdyje dauginė regresija įvertins biuro kaštų priklausomybę nuo darbuotojų skaičiaus, patalpų nuomos ir ryšių paslaugų tarifų. Tiesinė trijų kintamųjų regresijos lygties matematinė išraiška atrodys taip.

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i \quad 1.3$$

kur y_i - i -ojo biuro išlaikymo kaštai, kai $i = \overline{1,6}$; x_{1i} - darbuotojų skaičius i -ame biure; x_{2i} - patalpų nuomos tarifai i -ame biure; x_{3i} - ryšių paslaugų tarifai i -ame biure; $\beta_0, \beta_1, \beta_2, \beta_3$ - parametrai, kuriuos reikia apskaičiuoti, norint surasti dauginės regresijos lygtį, ε_i - regresijos paklaida.

Tiesinė ir netiesinė regresija

Tiek porinės, tiek dauginės regresijos matematinė išraiška kintamųjų y_i ir x_{ji} atžvilgiu gali būti ne tik tiesinė, bet ir netiesinė. 1.2. lentelėje pateiktos dažniausiai naudojamos regresijos lygties matematinės funkcijos. Visas šias lygtis nesunkiai galima suvesti į tiesinę išraišką matematinų procedūrų pagalba

(logaritmuojant ir prilyginant regresijoje esančių netiesinių kintamųjų reikšmes naujai įvestų kintamųjų reikšmėms, pvz. $v_i = \ln(x_i)$).

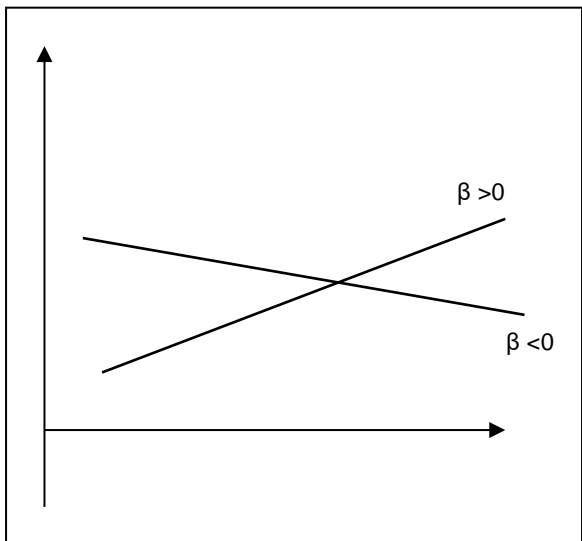
1.2 lentelė. Dažniausiai naudojamos regresijos funkcijos (porinė regresija)

Regresijos lygties forma	Matematinė išraiška
Tiesinė	$y_i = \beta_0 + \beta_1 x_i$
Eksponentinė	$y_i = \beta_0 (e^{\beta x_i})$
Hiperbolinė	$y_i = \beta_0 + \beta_1 (1/x_i)$
Logaritminė	$y_i = \beta_0 + \beta_1 (\ln x_i)$
Kvadratinė	$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2$
Rodiklinė	$y_i = \beta_0 (x_i^\beta)$

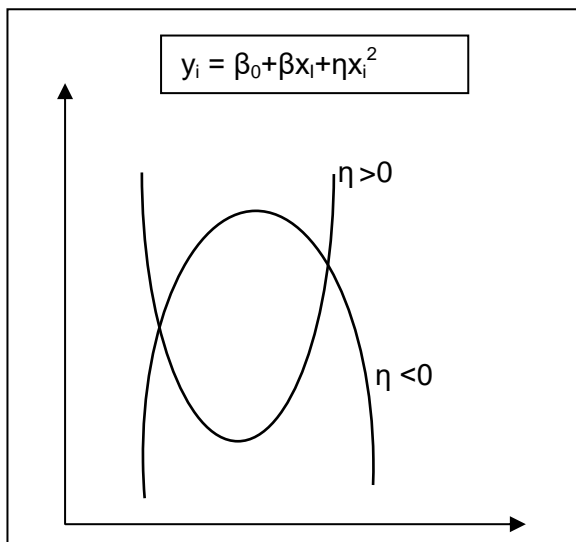
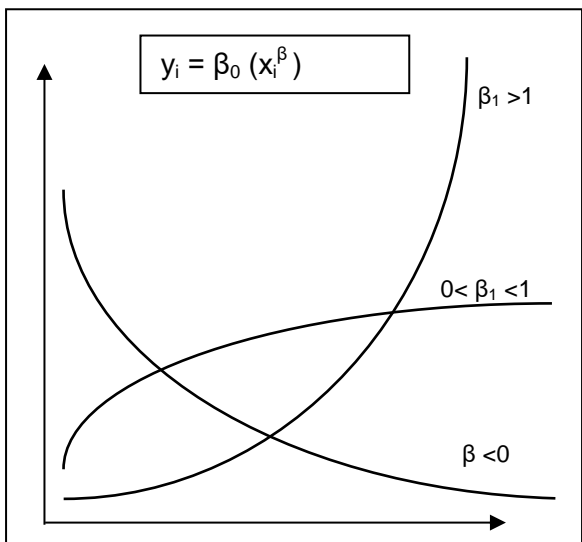
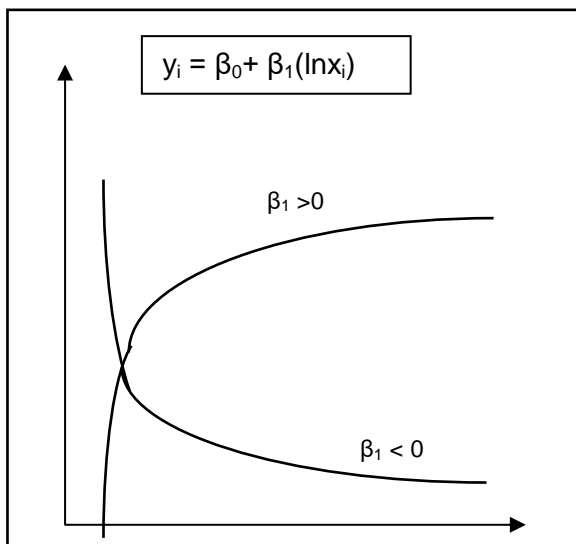
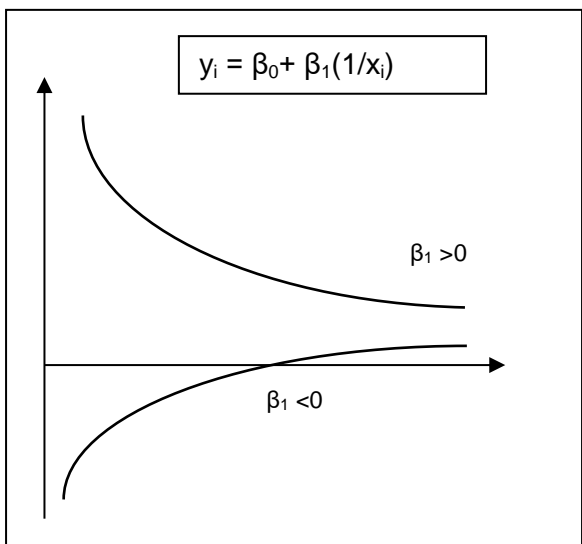
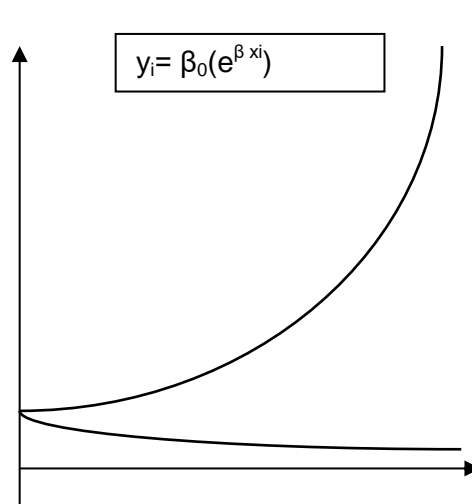
Funkcijų grafinė forma pateikta kitame puslapyje.

¶ Kvadratinė funkcija yra antros eilės polinomo funkcija. Regresinei analizei galima naudoti ir aukštesnių eilių polinominės funkcijas. Įsidėmėtina, kad didėjant polinomo eilei, jo funkcija vis tiksliau aprašo stebėjimus, pagal kuriuos įvertinti polinomo parametrai. Tačiau didesnės eilės polinomas yra visiškai netinkama funkcija prognozuojant. Praktiškai regresinei analizei ir prognozei taikytinas tik antros eilės polinomas, t.y. kvadratinė funkcija. ¶

Regresinės lygties matematinų išraiškų grafinė forma



$y_i = \beta_0 + \beta_1(\ln x_i) + \varepsilon_i$



Regresijos ryšio parametrai ir jų įverčiai

Pagrindinis uždavinys regresinėje analizėje - teisingai įvertinti regresijos lygties koeficientus kurie ir yra veiksnių sąryšio matai. Mes vartojame žodį "Įvertinti", o ne "surasti" arba "apskaičiuoti" kadangi labai dažnai apskaičiuoti tikrąsias koeficientų reikšmes yra neįmanoma. Regresinės lygties koeficientai yra vadinami parametrais arba parametru įverčiais

Tikrosios parametru reikšmės dažnai nėra žinomos ir negali būti tiksliai nustatytos, nes visada esama ribojančių objektyvių ir subjektyvių kliūčių, su kuriomis susiduria analitikas sudarydamas regresinę lygtį. Prie objektyvių kliūčių galima priskirti faktą, kad dažnai neįmanoma, o neretai ir netikslinga įtraukti visus nagrinėjamą reiškinį veikiančius veiksnius. Neretai juos labai sunku kiekybiškai išmatuoti. Be to, dažnai tenka reikalingus analizei duomenis pakeisti statistinėse ataskaitose fiksuojamais mažiau tiksliais duomenimis ir apsiriboti, tik tam tikra jų apimtimi. Prie subjektyvių kliūčių priskirtini netikslumai parenkant veiksnius, nustatant priklausomybės matematinę išraišką, klaidas suvedant informaciją ir kt. Todėl ekonometrijoje yra sakoma, kad skaičiuojant regresines lygtis gaunami nagrinėjamų duomenų pagrindų apskaičiuoti įverčiai, o ne tikrosios parametru reikšmės. Siekiant pabrėžti, kada yra kalbama apie tikrąsias parametru reikšmes, o kada apie parametru įverčius, parametrus ir įverčiams yra naudojami skirtingi žymėjimai. Tikrąsias parametru reikšmes žymime graikiškomis raidėmis β , ε , o jų įverčius lotyniškais: b ir e .

Parametru įverčių tikslumas

Parametru įverčiai tuo tikslesni, kuo jie artimesni tikrosioms parametru reikšmėms. Įverčiai bus įmanomai artimi, jeigu skaičiuojant regresinę lygtį, bus tenkinami trys pagrindiniai reikalavimai:

1. *Naudojamas tinkamas įverčių radimo metodas,*
2. *Regresinė lygtis tenkina klasikinės regresinės analizės prielaidas*
3. Duomenų pakankamumas

1. Įverčių radimo metodai

Regresinėje analizėje dažniausiai naudojami du parametru įverčių nustatymo metodai: mažiausių kvadratų metodas (MKM) ir maksimalaus tikėtimumo metodas (MTM). Mažiausių kvadratų metodas bus aptartas kitoje paskaitoje. MTM yra sudėtingesnis, todėl šiame paskaitų konspekte nėra nagrinėjamas.

2. Klasikinės regresinės analizės prielaidos

Įverčiai bus netikslūs, jeigu apskaičiuota regresijos lygtis netenkina klasikinių regresijos prielaidų, kurios pateiktos 1,3 lentelėje.

1.3. lentelė Klasikinės regresijos prielaidos

Prielaida	Prielaidos simbolinė išraiška
I. Regresijos funkcija koeficientų ir paklaidų atžvilgiu yra tiesinė (tiesiškumas)	$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_n x_{in} + \varepsilon_i$
II. Paklaidų vidurkis lygus nuliui (nulinis vidurkis)	$E(\varepsilon_i) = 0$
III. Paklaidos neautokoreliuoja (likučių ne autokoreliacijos), t.y. paklaidos tarpusavyje nėra susijusios ir nestebimi sklaidos dėsninčiai.	$Cov(\varepsilon_i, \varepsilon_j) = 0, \forall i, j / i \neq j$
IV. Paklaidų dispersija yra pastovi (ne heteroskedastiškumas) Didėjant nepriklausomų kintamųjų reikšmėms, priklausomojo kintamojo sklaidos intervalas išlieka pastovus.	$\sigma^2(\varepsilon_i) = \text{const.}$
V. Nepriklausomi kintamieji nėra tiesiškai tarpusavyje susiję, t.y. nėra tiesinės vieni kitų tiesinės kombinacijos (ne multikolinearumas, neinterkoreliacija)	$x_i \neq \gamma + \theta_j x_j, \forall i, j / i \neq j$
VI. Paklaidos pasiskirsčiusios pagal normalųjį skirstinį (normalumas).	$\varepsilon_i \sim N(0, \sigma^2)$

Jeigu skaičiuojant parametrų įverčius yra tenkinami pirmieji du viršuje paminėti reikalavimai, tuomet turime taip vadinamus "geriausius" parametrų įverčius, kurie pasižymi trimis savybėmis: yra **nepaslinkti, efektyvūs ir suderinti**.

Įverčių nepaslinktumas reiškia, jog, apskaičiavus tą pačią regresijos lygtį su skirtingomis duomenų imtimis, gauname iverčius, kurių vidurkis yra lygus tikrajai parametro reikšmei.

Įverčių efektyvumas. Įverčiai yra efektyvūs tada, kai jų dispersija yra minimali. Ši savybė reiškia, kad skirtingoms imtims apskaičiuoti regresijos lygties įverčiai įmanomai arti išsibarstę aplink tikrąsias parametro reikšmes.

Suderinti įverčiai reiškia, kad **didinant stebėjimų skaičių imtyje**, t.y., stebėjimų skaičiui artėjant prie begalybės įverčio reikšmė artėja prie tikrosios parametro reikšmės.

3. Duomenų pakankamumas

Regresinei analizei atlikti reikia turėti kuo daugiau stebėjimų. Taip didėja tikimybė taikant tinkamą regresijos parametrų įverčių radimo metodą nustatyti tikriesiems parametrams artimas įverčių reikšmes.

2. Porinė tiesinė regresija

Pati paprasčiausia matematinė veiksnių priklausomybės išraiška - tai tiesinė porinės regresijos lygtis, kuri priklausomąjį kintamąjį tiesiškai susieja tik su vienu nepriklausomu kintamuoju. Kai priklausomybės forma yra nesudėtinga lengviau suprasti ir įrodyti regresinės analizės teiginius ir išvadas, kurių daugumą, atlikus tam tikrus papildymus, nesunkiai galima pritaikyti sudėtingesnėms regresijos lygtims. Be to porinę tiesinę regresiją nesunku pavaizduoti grafiškai, ją lengviau interpretuoti.

Šioje paskaitos dalyje bus aptarti tokie klausimai :

- porinės tiesinės regresijos išraiška,
- parametrų įverčių nustatymas mažiausių kvadratų metodu MKM,
- regresijos prielaidų tikrinimas.

Porinės tiesinės regresijos išraiška

Pagrindinis porinės tiesinės regresijos analizės tikslas - nustatyti tiesinę priklausomybę tarp dviejų veiksnių ir apskaičiuoti jos parametrų įverčius. Matematiškai tiesinė porinė priklausomybė užrašoma tokiu būdu:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad 1.4$$

kur y_i – nagrinėjamas veiksnys- lygtyje aiškinamasis kintamasis; x – įtakojantis veiksnys - nepriklausomas kintamasis; β_0 ir β_1 - parametrai, ε_i - paklaida (likutis), parodanti priklausomojo kintamojo faktinių ir apskaičiuotų, remiantis porine tiesine regresija reikšmių skirtumus. Kadangi tikrosios parametrų reikšmės nežinomos, tai analizės metu pagal kintamųjų stebėjimus yra nustatomi dviejų tiesinės funkcijos parametrų β_0 ir β_1 bei likučių ε įverčiai. Primename, kad tikrąsias parametrų reikšmes ir paklaidas žymime graikiškomis raidėmis $\beta_0, \beta_1, \varepsilon_i$, kaip parodyta 1,4 lygtyje, o apskaičiuotas - lotyniškais raidėmis: b_0, b_1 , ir e_i : (1,5 lygtyje)

$$y_i = b_0 + b_1 x_i + e_i \quad 1.5$$

Parametrų įverčių nustatymas mažiausių kvadratų metodu

Nustatant regresijos parametrų įverčius priklausomo ir nepriklausomų kintamųjų duomenys turi būti suderinti laiko, vietos ir periodiškumo atžvilgiu. Pats populiariausias įverčių radimo metodas - tai mažiausių kvadratų metodas (MKM).

MKM tikslas - nustatyti tokius regresijos parametrų įverčius, kurie minimizuoja skirtumų tarp faktinių (y_i) ir apskaičiuotų (\hat{y}_i) pagal pasirinktą regresijos lygtį priklausomojo kintamojo reikšmių kvadratų sumą. Matematiškai mažiausių kvadratų metodas užrašomas taip:

$$\sum (y_i - \hat{y}_i)^2 = \sum e_i^2 \rightarrow \min \quad 1.6$$

kur y_i - faktinės priklausomojo kintamojo reikšmės, o \hat{y}_i - apskaičiuotos priklausomojo kintamojo reikšmės. Tiesinės porinės regresijos atveju \hat{y}_i reikšmės bus lygios:

Jei regresija tenkina klasikines prielaidas, tuomet MKM pagalba nustatyti parametrų įverčiai yra geriausi, t.y., nepaslinkti, efektyvūs ir suderinti. Toliau pateikiama skaičiavimų procedūra.

Užduotis 1. Apskaičiuojamos funkcijos $\Sigma(y_i - (b_0 + b_1 x_i))^2$ pirmosios dalinės išvestinės ir prilyginamos nuliui²:

$$\partial [\Sigma(y_i - b_0 - b_1 x_i)^2] / \partial b_0 = \Sigma[-2(y_i - b_0 - b_1 x_i)] = 0$$

$$\partial [\Sigma(y_i - b_0 - b_1 x_i)^2] / \partial b_1 = \Sigma[-2(y_i - b_0 - b_1 x_i) * x_i] = 0$$

2. Toliau sprendžiama lygčių sistema:

$$\begin{cases} \Sigma[-2(y_i - b_0 - b_1 x_i)] = 0 \\ \Sigma[-2(y_i - b_0 - b_1 x_i)x_i] = 0 \end{cases} \rightarrow \begin{cases} \Sigma(y_i - b_0 - b_1 x_i) = 0 \\ \Sigma(y_i - b_0 - b_1 x_i) * x_i = 0 \end{cases} \rightarrow \begin{cases} \Sigma y_i = n b_0 + b_1 \Sigma x_i \\ \Sigma x_i y_i = b_0 \Sigma x_i + b_1 \Sigma x_i^2 \end{cases}$$

Išsprendus lygčių sistemą gaunamos porinės tiesinės regresijos lygties parametrų įverčių nustatymo formulės:

$$b_1 = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2} \quad 1.7$$

$$b_0 = \frac{(\sum y_i - b_1 \sum x_i)}{n} = \bar{y} - b_1 \bar{x} \quad 1.8$$

kur \bar{y} - priklausomojo kintamojo faktinių reikšmių vidurkis, o \bar{x} - nepriklausomojo kintamojo reikšmių vidurkis. n- stebėjimų skaičius

Pavyzdys

Pasinaudosime pavyzdžiu apie įdarbinimo bendrovės biurų išlaikymo kaštus. Kadangi nagrinėjame porinę tiesinę regresiją - tai susiesime bendrovės atstovybių biurų kaštus (priklausomas kintamasis y_i) su vienu veiksnium - darbuotojų skaičiumi (nepriklausomas kintamasis - x_i). Duomenys apie 6 biurų išlaikymo kaštus ir darbuotojų skaičių pateikti 1,1 lentelėje, o grafiškai pavaizduoti 1,1 paveiksle. Tiesinė priklausomybės matematinė išraiška tarp dviejų kintamųjų, kaip jau buvo užrašyta atrodo taip:

Mūsų užduotis apskaičiuoti MKM pagalba parametrų įverčius b_0 ir b_1 . Pasinaudosime formulėmis 1.7 ir 1.8.

$$y_i = b_0 + b_1 x_i + e_i$$

Skaičiuoti bus patogiau, jeigu sudarysime pradinių duomenų ir atliktų skaičiavimų lentelę 1.4

² Antros turi būti teigiamos

Lentelė 1.4

i- ojo biuro išlaikymo kaštai y_i	Dirbančiųjų skaičius i ame biure x_i	x_i^2	$y_i x_i$
100	10	100	1000
90	8	64	720
140	14	196	1960
130	12	144	1560
120	11	121	1320
100	7	49	700
$\Sigma y_i =$ 680	$\Sigma x_i =$ 62	$\Sigma x_i^2 =$ 674	$\Sigma x_i y_i =$ 7260

Įstatę į 1.7 formulę apskaičiuotas reikšmes gauname parametro b_1 įvertį:

$$b_1 = \frac{6 * 7260 - 62 * 680}{6 * 674 - (62)^2} = 7$$

Parametro įverčiui a apskaičiuoti pasinaudojame formulę 1.8.

$$b_0 = \frac{(680 - 7 * 62)}{6} = 41$$

Apskaičiavę a ir b parametrų įverčius galime užrašyti įdarbinimo bendrovės biurų išlaikymo kaštų priklausomybės nuo darbuotojų skaičiaus matematinę lygtį:

$$y_i = 41 + 7x_i + e_i$$

Regresinės lygties parametrų ekonominė interpretacija

Apskaičiuoti ekonominio reiškinių regresinės lygties koeficientai tyrinėtoji suteikia daug svarbios informacijos apie nagrinėjamą reiškinį. Parametro įvertis b_0 parodo priklausomojo kintamojo vidutinę reikšmę, kai nepriklausomas kintamasis lygus nuliui. Kitais žodžiais tariant, b_0 parodo, kokią vidutinę reikšmę įgytų nagrinėjamas reiškinys y , jeigu jo neveiktų pasirinktas veiksnys x .

Parametro įvertis b_1 parodo vidutinį priklausomojo kintamojo y_i reikšmės pasikeitimą, kai nepriklausomas kintamasis x_i pasikeičia vienu vienetu. Šis parametras kiekybiškai įvertina pasirinkto veiksnio x įtaką nagrinėjamam reiškiniui.

Turint parametrų įverčių skaitines reikšmes, galima apskaičiuoti ir elastingumą. Elastingumas (E) parodo, kiek procentų vidutiniškai pakis priklausomas kintamasis y , jeigu nepriklausomas kintamasis x pasikeis vienu procentu. Elastingumas skaičiuojamas pagal tokia formulę:

$$E = \frac{\text{procentinis pokytis } y}{\text{procentinis pokytis } x} = \frac{\Delta y}{\Delta x} * \frac{x}{y} = b_1 \frac{x}{y} \quad 1.9 \text{ Tiesin\text{e}}$$

funkcijos atveju elastingumo reikšmės priklauso nuo kintamųjų y_1 ir x_1 reikšmių. Todėl dažnai yra skaičiuojamas vidutinis elastingumas, kuris yra lygus įverčiui b_1 , padaugintam iš kintamųjų reikšmių vidurkių santykio :

$$E = b_1 \frac{\bar{x}}{\bar{y}}$$

Pavyzdys

Pavyzdyje apie įdarbinimo bendrovę parametro β_0 įvertis =41, o parametro β_1 įvertis $b_1=7$. Šiuos skaičius galima interpretuoti tokiu būdu. Įvertį b_0 šiuo konkrečiu atveju galima suprasti, kaip pastovius biuro išlaikymo kaštus, kurie nepriklauso nuo darbuotojų skaičiaus. Taigi įdarbinimo bendrovės viduriniai pastovūs biuro išlaikymo kaštai yra lygūs 41 tūkst. Eurų per metus.

Parametro įvertis b_1 parodo, kad padidėjus vienu darbuotoju, biuro išlaikymo kaštai padidėja eurų 7 tūkst. eurų per metus.

Biuro išlaikymo kaštų elastingumas darbuotojų skaičiaus pokyčiams yra lygus: $E=7(10,33/113,33)=0,64$, t.y. darbuotojų skaičiui padidėjus vienu proc., biurų išlaikymo kaštai vidutiniškai padidėja 0,64 proc.

Regresijos prielaidų tikrinimas

Atliekant porinę tiesinę regresijos analizę turi būti užtikrintos klasikinės regresijos prielaidos. Pirmoji prielaida – tiesiškumas - automatiškai patenkinama, kadangi aptariamas modelis yra *tiesinė* porinė regresija. Šeštoji prielaida – ne multikolinearumas – porinėje regresijoje neaktuali, kadangi yra tik vienas aiškinantis veiksnys. Likusios prielaidos yra susijusios su likučiais (paklaidomis): šie turi tenkinti nulinio vidurkio, pastovios dispersijos (homoskedastiškumo - ne heteroskedastiškumo), neautokoreliacijos, ir normalumo prielaidas. Kadangi tikrųjų likučių reikšmių neįmanoma žinoti, tai prielaidų tenkinimas tikrinamas pagal likučių įverčius. Jie apskaičiuojami iš faktinių stebėjimo dydžių atimant įvertintą teorinę regresijos reikšmę:

$$e_i = y_i - \hat{y}_i = y_i - (b_0 + b_1 x_i).$$

Taigi 1.3. lentelėje pateiktų prielaidų tenkinimą reikia tikrinti vietoje paklaidos ε įstatant paklaidos įvertį e .

Prognozavimas regresijos pagalba

Nustačius parametrų įverčius ir įsitikinus, kad tenkinamos klasikinės regresijos prielaidos, toliau galima skaičiuoti bei prognozuoti priklausomojo kintamojo reikšmes. Priklausomojo kintamojo reikšmė

y_p yra vadinama *sąlygine prognoze*, ir gaunama, į regresijos lygtį įstačius, prognozinę nepriklausomojo kintamojo (x_p) reikšmę. darant prielaidą, kad ji yra žinoma iš anksto arba yra apskaičiuota kitame modelyje.

$$y_p = b_0 + b_1 x_p.$$

Prisiminkime Įdarbinimo bendrovės pavyzdį. Tarkim, kad bendrovės savininkai nori padidinti darbuotojų skaičių, iki $x_p = 15$. Tuomet įstatę į 1,9 lygtį šią reikšmę gauname, kad vidutiniai biuro išlaikymo kaštai bus 146 tūkst. eurų per metus:

$$y_p = 41 + 7(15) = 146$$

Regresinė analizė ir prognozavimas (II)

Paskaitos dalys:

1. Taškiniai ir intervaliniai įverčiai.
2. Iškeltų hipotezių tikrinimas. Prognozavimas regresija.

Pagrindiniai terminai:

Taškiniai ir intervaliniai įverčiai

- Taškinis parametro įvertis
- Intervalinis parametro įvertis
- Koeficientas
- Įverčio standartinė paklaida
- Laisvės laipsnių skaičius

Iškeltų hipotezių tikrinimas, prognozavimas regresija

- Hipotezė
- Nulinė hipotezė
- Alternatyvi hipotezė
- t statistika
- Reikšmingumo lygmuo
- Patikimumo (pasikliovimo) lygmuo
- Vienpusė hipotezė
- Dvipusė hipotezė
- Standartinė prognozės paklaida

1. Taškiniai ir intervaliniai įverčiai

Naudojant taškinius įverčius atliekami konkretūs regresinės analizės skaičiavimai pvz., prognozuojami įvairūs nagrinėjamo ekonominio reiškinio variantai, modeliuojamos įvairios ekonominės situacijos, skaičiuojami išvestiniai ekonominės analizės rodikliai, tokie kaip elastingumas, pakeičiamumo normos, masto ekonomijos efektai ir kt

Intervaliniai įverčiai būtini įvairių hipotezių apie tikrąsias parametrų reikšmes patvirtinimui arba atmetimui. Pvz., ar galima atmesti hipotezę, kad pasirinktas veiksnys neturi įtakos nagrinėjamam reiškiniai, arba su kokia tikimybe galima teigti, kad nagrinėjamas reiškinys įgis vienokią ar kitokią reikšmę.

Turinys

- Taškiniai parametrų įverčiai,
- Standartinė taškinio įverčio paklaida
- Intervaliniai parametrų įverčiai,
- Paklaidų dispersijos įverčiai.

Taškiniai parametru įverčiai

Taškinis įvertis yra toks regresijos parametro įvertis, kuris įgyja konkrečią skaitinę reikšmę. Pavyzdžiui, nustatyta, y reiškinį tiesiškai veikia k veiksnių grupė. Regresijos lygtis užrašoma taip:

$$y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki} + \varepsilon_i \quad 2.1$$

Tarkim, kad MKM yra apskaičiuoti regresijos parametru įverčiai. Į regresijos lygtį įrašome jų reikšmes ir užrašome regresinę lygtį su skaitinėmis reikšmėmis.

$$y_t = b_0 + b_1 x_{1i} + \dots + b_k x_{ki} + e_i \quad 2.2$$

Taškiniai įverčiai yra prieš nepriklausomus kintamuosius (x) esančios parametru įverčių reikšmės (b_0, b_1, \dots, b_k). Šios reikšmės sudaro atskirų parametru taškinų įverčių vektorių: $B' = [b_1, \dots, b_k]$.

Taškinis įvertis yra traktuojamas kaip "geriausias" parametru atspindintis rodiklis ir naudojamas konkrečiuose ekonominiuose skaičiavimuose. Tačiau egzistuoja tikimybė, jog įverčių ir tikroji parametro reikšmė gali būti skirtingos. Su tam tikra tikimybe tikroji parametro reikšmė gali būti lygi 0 ($P\{b_i = \beta_i\} = 0, \forall i$), o tai reiškia, kad veiksnys esantis prie šio parametro nedaro įtakos priklausomam kintamajam. Todėl turėdami tik taškinį įvertį jokių išvadų apie tikrąją parametro reikšmę daryti negalime - ji gali būti tiek didesnė, tiek mažesnė už parametro įverčio reikšmę. Norėdami tikrinti įvairias hipotezes apie tikrąsias parametru reikšmes turime nagrinėti intervalinius parametru įverčius.

Pavyzdys

Prisiminkime privačią gyventojų įdarbinimo bendrovę. Nagrinėjome porinę tiesinę regresiją, kurioje bendrovės atstovybių biurų kaštus susiejome su vienu veiksnium - darbuotojų skaičiumi. Apskaičiavome dviejų kintamųjų lygties $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, parametru β_0, β_1 įverčius MKM metodu, kurie lygūs $b_0 = 41, b_1 = 7$.

Į 1,4 regresijos lygtį įsistačius apskaičiuotus parametru įverčius gauname tokią lygtį:

$$y_i = (41 + 7x_i) + e_i.$$

Šiame pavyzdyje taškiniai parametru įverčiai yra 41 ir 7.

Taškinio įverčio standartinė paklaida

Taškinio įverčio patikimumui matuoti yra skaičiuojama standartinė įverčio paklaida SE_{b_j} . Porinės tiesinės regresijos lygties laisvajam nariui standartinė parametro įverčio paklaida apskaičiuojama pagal formulę:

$$SE_{b_0} = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-2} \frac{\sum x_i^2}{n \sum (x_i - \bar{x})^2}} \quad 2.3$$

Dauginės regresijos atveju standartinė laisvojo nario paklaida skaičiuojama pagal formulę:

$$SE_{b_0} = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2 \sum x_i^2}{n-k-1 \sum (x_i - \bar{x})^2}} \quad 2.3$$

kur k yra nepriklausomų kintamųjų ir atitinkamai vertinamų koeficientų prie jų skaičius. Porinės regresijos atveju jis yra lygus -1 . Jeigu regresijoje yra laisvasis narys, tuomet bendras vertinamų koeficientų skaičius yra vienu didesnis negu nepriklausomų kintamųjų skaičius, t.y., porinės regresijos atveju yra 2 , o dauginės $k+1$. i – stebėjimo numeris, y_i - faktinė i -ojo stebėjimo priklausomojo kintamojo reikšmė, \hat{y}_i - teorinė t.y. apskaičiuota pagal regresijos lygtį, i - ojo stebėjimo priklausomojo kintamojo reikšmė, x_i i -ojo nepriklausomojo kintamojo reikšmė, \bar{x} –nepriklausomojo kintamojo vidurkis.

Įverčio b_1 standartinė paklaida porinės ir dauginės regresijos atveju apskaičiuojama pagal formules:

$$SE_{b_1} = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-2 \sum (x_i - \bar{x})^2}} \quad SE_{b_1} = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-k-1 \sum (x_i - \bar{x})^2}} \quad 2.4$$

Intervaliniai parametrų įverčiai

Intervalinis parametro įvertis yra reikšmių intervalas, į kurį su tam tikra tikimybe gali pakliūti parametro tikroji reikšmė. Klasikinės regresijos atveju intervalinio įverčio vidurys yra taškinis parametro įvertis. Intervalo apatinis rėžis nustatomas iš taškinio parametro įverčio atimant, o viršutinis - pridant koreguojantį dydį, kuris yra lygus $- t_{n-k-1, \alpha/2} SE_{\beta_i}$

Taigi parametro b_i intervalinis įvertis užrašomas taip:

$$b_i \in [b_j \pm t_{n-k-1, \alpha/2} SE_{b_j}], \quad 2.5$$

kur apatinis b_i įverčio rėžis yra lygus: $b_j - t_{n-k-1, \alpha/2} SE_{b_j}$, o viršutinis b_i rėžis: $b_j + t_{n-k-1, \alpha/2} SE_{b_j}$,

SE_{β_i} - standartinė i - įverčio paklaida, apskaičiuota pagal 2.3 arba 2.4 formulę.

$t_{n-k-1, \alpha/2}$ - teorinė Student'o skirstinio statistika su $n-k-1$ laisvės laipsnių skaičiumi³ ir α reikšmingumo lygmeniu, Skaitinė $t_{n-k, \alpha/2}$ reikšmė randamas iš Studento statistinių skirstinių lentelių. (žr. priedą). Reikšmingumo lygmenį α pasirenka tyrinėtojas. Dažniausiai sutinkamas $\alpha= 0,05$, kuris reiškia, kad išvados yra daromos su 5 proc. tikimybe suklysti arba, $1-0,05=0,95$, t.y 95 proc. pasikliovimo lygmeniu. n –stebėjimų, skaičius, $k+1$ – vertinamų koeficientų skaičius,

Atlikus visus skaičiavimus gaunamas intervalas į kurį su tikimybe $1- \alpha$ patenka tikroji parametro reikšmė. Kuo šis intervalas yra siauresnis, tuo tikslesnis yra parametro įvertis.

Iš intervalinio iverčio formulės 2,5 matyti, kad įverčiai bus tuo tikslesni, kuo mažesnės bus įverčių paklaidos ir didesnis laisvės laipsnių skaičius. Taigi norint gauti tikslesnius įverčius, reikia siekti dviejų dalykų: pirma, kad į analizę įtrauktų stebėjimų skaičius n būtų kiek galima didesnis, ir, antra, kad įtakojančių veiksnių (nepriklausomų kintamųjų) reikšmės būtų kuo įvairesnės.

Pavyzdys

Nagrinėkime aukščiau paminėtą pavyzdį apie įdarbinimo bendrovę. Apskaičiuota porinė tiesinė regresijos lygtis: $y_i = (41 + 7x_i) + e_i$. Parametrų taškiniai įverčiai yra 41 ir 7,

Kaip jau buvo rašyta, intervalinio įverčio apatinis ir viršutinis rėžiai nustatomi atimant ir pridodant prie taškinio parametro įverčio koreguojantį dydį - $t_{n-k, \alpha/2} * SE_{\beta_i}$.

Mūsų užduotis rasti šį dydį. Studento pasiskirstymo $t_{n-k-1, \alpha/2}$ reikšmė randama Studento pasiskirstymo lentelėse. Mūsų pavyzdyje, pasirinkime įprastą reikšmingumo lygmenį $\alpha=0,05$. Laisvės laipsnių skaičius lygus $n-k-1=6-1-1=4$. kadangi stebėjimų skaičius $n=6$ (6 biurai), o vertinamų parametrų skaičius 2 (b_0 ir b_1 įverčiai). Iš lentelių randame , $t_{4, 0,05/2}=2.776$ (žr. priedą

Standartinės parametrų įverčių paklaidos apskaičiuojamos pagal 2.3 ir 2.4 formules. Kad būtų patogiau, sudarysime pradinių duomenų ir atliktų skaičiavimų lentelę 2.1.

2.1. lentelė

i-ojo biuro išlaikymo kaštai y_i	Dirbančiųjų skaičius i-ame biure x_i	x_i^2	$(x_i - \bar{x})^2$	$e_i^2 = (y_i - \hat{y}_i)^2$
100	10	100	0.11	121
90	8	64	5.44	49
140	14	196	13.44	1

³ Regresinėje analizėje laisvės laipsniai – tai stebėjimų skaičius, kurių neapriboja vertinami parametrai. Jei, pavyzdžiui, turime 100 stebėjimų ir vertiname 10 parametrų, tai turėsime 90 laisvų stebėjimų, Patogumo dėlei vertinamų parametrų skaičių galime sutapatinti su į regresiją įtrauktų nepriklausomų kintamųjų (įtakojančių veiksnių skaičiumi). Tik svarbu atkreipti dėmesį, ar regresinė lygtis turi laisvąjį narį, nes tokiu atveju k padidėja vienu vienetu, žyminčiu laisvąjį narį. . Kuo didesnis laisvės laipsnių skaičius, tuo tikslesnis gaunamas įvertis.

	130	12	144	2.78	25	
	120	11	121	0.44	4	
	100	7	49	11.11	100	
$\sum y_i =$	680	$\sum x_i =$	62	$\sum x_i^2 = 674$	$\sum (x_i - \bar{x})^2 = 33.33$	$\sum (y_i - \hat{y}_i)^2 = 300$

Įstatę į formules apskaičiuotas reikšmes gauname $SE_{b_0} = 15.94$, $SE_{b_1} = 1.5$. Sudauginę atitinkamus skaičius iš $t_{4,0,05/2} = 2.776$ gauname koreguojantį dydį ir intervalinių įverčių režius. Koreguojantis dydis parametru b_0 yra lygus 44.25, parametru b_1 – 4.16.

2.2 lentelėje Parametrų taškiniai ir intervaliniai įverčiai.

Parametras	Parametro įvertis	Koreguojantis dydis	Parametro intervalinis įvertis
β_0	$b_0 = 41$	44,25	[-3,25; 85,25]
β_1	$b_1 = 7$	4,16	[2,84; 11,16]

Gautus skaičiavimų rezultatus galima interpretuoti tokiu būdu; Su 95% tikimybe galima teigti, kad parametro β_1 tikroji reikšmė yra intervale $2,84 < \beta_1 < 11,16$ t.y., padidėjus vienu darbuotoju biuro išlaikymo kaštai padidės nuo 2,84 iki 11,16 tūkst. litų per ketvirtį. Laisvojo nario β_0 intervaliniai įverčiai paprastai nėra interpretuojami, Jie reikalingi hipotezių tikrinimui.

2. Iškeltų hipotezių tikrinimas

Hipotezių tikrinimas regresinėje analizėje leidžia atsakyti į tokius klausimus;

- ar galime teigti, kad veiksnys daro poveikį nagrinėjamam reiškiniui (nepriklausomojo kintamojo poveikis priklausomajam kintamajam yra statistiškai reikšmingas),
- kokia yra veiksnio įtaka, t.y., didina ar mažina nagrinėjamą reiškinį (koks poveikio ženklas: teigiamas ar neigiamas),

Šioje paskaitos dalyje bus išaiškintas tokių sąvokų turinys:

- Nulinė ir alternatyvi hipotezė
- Hipotezės apie parametrus tikrinimas,
- Vienpusė ir dvipusė hipotezė,
- Prognozavimas regresija.

Hipotezės samprata

Hipotezė yra iš anksto griežtai suformuluotas ir tam tikra analize tikrinamas teiginys. Regresinės analizės metu hipotezė gali būti arba *statistiškai* patvirtinama, arba atmetama, tačiau ji niekada nėra įrodoma.

Regresinėje analizėje hipotezių tikrinimo procedūroje būtini keturi elementai:

Nulinė hipotezė H_0

Alternatyvi hipotezė H_A ,

Testo statistika,

Hipotezės paneigimo taisyklė.

Nulinė hipotezė (H_0) – tai teiginys arba prielaida, kurią reikia statistiškai patvirtinti, remiantis nagrinėjamais stebėjimais.

Alternatyvi hipotezė (H_A) - tai teiginys arba prielaida, kuris yra teisingas, kai nulinė hipotezė statistiškai atmetama.

Statistiškai patvirtinti hipotezę reiškia, jog nagrinėjami duomenys teiginį patvirtins su didesne tikimybe nei pasirinktas patikimumo lygmuo. Jei patvirtinama nulinė hipotezė, tai alternatyvi hipotezė atmetama. Jei nepatvirtinama nulinė hipotezė, tai priimama alternatyvi. Taigi statistinę hipotezę sudaro dviejų alternatyvų visuma

2.3. lentelėje pateiktos dažniausiai tikrinamos hipotezės apie regresinės lygties

$$y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki} + \varepsilon_i$$

parametrus.

2.3. lentelė Pagrindinės regresinėje analizėje tikrinamos hipotezės

Tikrinamas teiginys	Nulinė (H_0) ir alternatyvi (H_A) hipotezė	Pastabos
Regresijos lygties laisvasis narys (atkarpos parametras) nereikšmingas (lygus nuliui)	$H_0: \beta_0=0$ $H_A: \beta_0 \neq 0$	
Nepriklausomas kintamasis neturi įtakos priklausomam kintamajam	$H_0: \beta_i=0, i=1..k$ $H_A: \beta_i \neq 0, i=1..k$	
Nepriklausomojo kintamojo parametras yra lygus tam tikrai reikšmei	$H_0: \beta_i=c, H_A: \beta_i \neq c, c$ – tikrinama parametro reikšmė	Galima pastebėti, jog ankstesnės hipotezė $H_0: \beta_i=0$, yra atskiras hipotezės $H_0: \beta_i=c$ atvejis, kai $c=0$.
Įvertinta regresija yra statistiškai nereikšminga	$H_0: \beta_i=0, \forall i, i=1..k$ $H_A: \beta_i \neq 0, \forall i, i=1..k$	Tai jungtinė hipotezė, jog visi parametrai kartu paėmus nėra reikšmingi.

Vienpusė ir dvipusė hipotezė

Prieš taikant regresinės analizės rezultatus hipotezių tikrinimui, turi būti nustatyta, ar tikrinama dvipusė ar vienpusė hipotezė. Nepriklausomų kintamųjų statistinio reikšmingumo tikrinimas - tai dvipusės hipotezės pavyzdys. *Dvipusės hipotezės* atveju alternatyvi hipotezė formuluojama su nelygybės ženklu, pvz.:

a) Kintamųjų poveikio statistinio reikšmingumo tikrinimas:

$$H_0: \beta_i = 0; \quad H_a: \beta_i \neq 0$$

b) Hipotezė, apie konkrečią (pvz.; $c=3$) parametro reikšmę:

$$H_0: \beta_i = 3; \quad H_a: \beta_i \neq 3.$$

Vienpusė yra tokia *hipotezė*, kurioje tikrinamo parametro reikšmės gali būti tik didesnės arba tik mažesnės už hipoteze tikrinamą dydį. Parametro konkretaus ženklo tikrinimas yra tokios hipotezės pavyzdys. Vienpusės hipotezės pavyzdžiai:

a) Parametro ženklo tikrinimas: $H_0: \beta_i = 0; \quad H_a: \beta_i > 0$ (hipotezė: didėjant kintamajam x_j , priklausomas kintamasis y_j irgi didėja)

b) $H_0: \beta_i = 0; \quad H_a: \beta_i < 0$ (hipotezė: didėjant kintamajam x_j , priklausomas kintamasis y_j mažėja)

Vienpusės ir dvipusės hipotezių atmetimo procedūra skiriasi. Kai nagrinėjama vienpusė hipotezė, išvados apie nulines hipotezės atmetimą daromos lyginant apskaičiuotą testo statistikos t reikšmę su α reikšmingumo t -skirstinio teorine reikšme. Dvipusės hipotezės atveju apskaičiuota testo statistikos t reikšmė lyginama ne su α , o su $\alpha/2$ reikšmingumo t -skirstinio teorine reikšme.

Hipotezių apie įverčių reikšmingumą tikrinimas

Regresinėje analizėje svarbu nustatyti ar konkretus veiksnys daro įtaką nagrinėjamam reiškiniui ar ne. Veiksnių poveikio statistinis reikšmingumas tikrinamas tokiu būdu: iškeliami hipotezė, kad veiksnio poveikis yra nereikšmingas. Matematiškai tai užrašoma taip: parametras $\beta_i = 0$, t.y. parametras prie veiksnio su tikimybe $1-\alpha$ gali būti lygus 0. Paprastai šis teiginys yra įvardijamas nuline hipoteze. t.y., $H_0: \beta_j = 0$

Regresinės analizės įverčių reikšmingumo tikrinimo procedūrą sudaro keturi žingsniai:

1. žingsnis. *Formuluojamos hipotezės:*

$H_0: \beta_i = 0$ (nepriklausomas veiksnys (x_i) nedaro įtakos priklausomam kintamajam t.y., koeficientas prie veiksnio gali būti lygus 0)

$H_1: \beta_i \neq 0$ (x_i poveikis reikšmingas - regresijos koeficientas prie veiksnio nelygus 0)

2. žingsnis. *Apskaičiuojama testo statistika.* Veiksnių reikšmingumui tikrinti dažniausiai naudojama t statistika, kuri yra apskaičiuojama pagal formulę

$$t = \frac{b_j - \beta_j}{SE_{b_j}} = \frac{b_j - 0}{SE_{b_j}} \quad 2.6$$

Dydis t yra pasiskirstęs pagal Stjudento t-skirstinį su $\alpha/2$ reikšmingumo lygmeniu ir n-k-1 laisvės laipsniais. t.y. $t \sim t_{\alpha/2}(n-k-1)$

3 žingsnis Apskaičiuota t statistikos reikšmė lyginama su teorine t-skirstinio $t_{\alpha/2}(n-k-1)$ reikšme.

4 žingsnis. *Daromos išvados* Jei apskaičiuota t reikšmė modulis yra **didesnis** už teorinę t-skirstinio reikšmę, tuomet **nulinė hipotezė atmetama** ir priimama alternatyvi hipotezė. Su $1-\alpha$ tikimybe (pvz., $\alpha=0,05$, t.y., 95 proc. tikimybe) galime tvirtinti, kad j-ojo veiksnio poveikis yra statistiškai reikšmingas. Priešingu atveju, kai t apskaičiuotos reikšmės modulis yra **mažesnis** už teorinę reikšmę $t_{\alpha/2}(n-k-1)$, negalime atmesti nulinės hipotezės, o tai reiškia, kad negalime tvirtinti, kad j veiksnio poveikis yra statistiškai reikšmingas.

Ÿ Hipotezės paprastai tikrinamos taikant 95% ar 99% pasiklovimo lygmenį ($1-\alpha$). Jis rodo tikimybę, jog tikroji parametro reikšmė gali būti lygi tikrinamajai reikšmei (c). Reikšmingumo lygmuo α rodo nulinės hipotezės atmetimo tikimybę, kai ji iš tiesų yra teisinga. Tokia klaida vadinama α arba I tipo klaida. β arba II tipo klaidos (neatmesti nulinės hipotezės, kai ji iš tiesų neteisinga) tikimybę.

Lygiavertę išvadą apie tikrinamą hipotezę galima gauti lyginant pagal apskaičiuotą t reikšmę nustatytą faktinį reikšmingumo lygmenį p (t.y., tikimybę, kad apskaičiuotas parametras įgaus nulio reikšmę) su analitiko pasirinktu α - reikšmingumo lygmeniu. Jei faktinis yra mažesnis nei pasirinktas, tai nulinė hipotezė atmetama, jei didesnis – priimama. Ÿ

Veiksnių reikšmingumo tikrinimo pavyzdys

Kiekvienos regresinės analizės metu būtina patikrinti parametrų įverčių reikšmingumą.

Prisiminkime privačios įdarbinimo bendrovės pavyzdį. Pagal 6 stebėjimų duomenis įvertinta regresija:

$$y_I = 41 + 7x_i + e_i \quad (15.94) \quad (1.5)$$

Skliaustuose pateiktos standartinės paklaidos reikšmės SE_{b_0} ir SE_{b_1} .

1. žingsnis. Tikrinama parametrų lygybės nuliui hipotezė

$H_0: \beta_j=0$, (parametras prie nepriklausomojo kintamojo x gali būti lygus 0, t.y., x darbuotojų skaičiaus kitimas nedaro įtakos bendrovės kaštams)

$H_A: \beta_j \neq 0$. (parametras prie nepriklausomojo kintamojo x nelygus 0 t.y., x darbuotojų skaičiaus kitimas daro poveikį kaštų kitimui)

2 žingsnis Apskaičiuojama testo statistika:
$$t = \frac{b_1 - 0}{SE_{b_1}} = \frac{7 - 0}{1.5} = 4,67$$

3 žingsnis Apskaičiuota faktinė t reikšmė lyginama su $t_{\alpha/2, n-k-1}$ teorine reikšme. Iš t pasiskirstymo lentelių (žr. priedus) randame $t_{0,025,4} = 2,776$ (laisvės laipsnių skaičius: $n-k-1=6-2=4$, ir $\alpha/2=0,05/2=0,025$ randame $t_{0,025,4} = 2,776$ Palyginę apskaičiuotą t reikšmę su teorine $t_{0,025,4}$, matome $4,667 > 2,776$

4 žingsnis Išvada. Su 95% tikimybe atmetame nulinę hipotezę $H_0: \beta_1=0$ ir priimama alternatyvi hipotezė $H_A: \beta_1 \neq 0$, kuri reiškia, kad veiksnio x - darbuotojų skaičiaus poveikis yra statistiškai reikšmingas.

Analogiškai tikriname ir laisvojo nario reikšmingumą. Skaičiavimai pateikti lentelėje 2,4. Atlikus visus

žingsnius matome, kad apskaičiuota testo statistikos $t = \frac{b_0 - 0}{SE_{b_0}} = \frac{41 - 0}{15,94} = 2,57$ reikšmė yra mažesnė už

teorinę reikšmę - $t_{0,025,4} = 2,776$ t.y., $2,57 < 2,776$. Todėl darome išvadą, kad nulinės hipotezės $H_0: b_0=0$ atmesti negalime. Tai reiškia, kad su 95 proc. tikimybe galime teigti, kad biuro išlaikymo pastovūs kaštai, kurie nėra susiję su darbuotojų skaičiumi gali būti lygūs 0. Tai yra teisinga, kai darbuotojų skaičius $x=0$. Logiška išvada, kadangi beprasmiška laikyti biurą, jeigu jame nėra darbuotojų.

2.4. lentelė Koeficientų reikšmingumo nustatymo rodikliai

Parametro įvertis	Skaitinė įverčio reikšmė	Standartinė įverčio paklaida (SE)	Apskaičiuota t statistikos reikšmė	$t_{0,025,4}$ iš T - pasiskirstymo lentelės	Išvada apie koeficiento reikšmingumą
b_0	41	15.94	2.57	2.776	nereikšmingas
b_i	7	1.5	4.67	2.776	reikšmingas

Kai yra padaroma išvada, kad priklausomas kintamasis arba laisvasis narys yra statistikai nereikšmingas, tuomet iš naujo reikia skaičiuoti regresijos lygtį be to nepriklausomo kintamojo arba laisvojo nario. Įsidėmėtina, jog išmetus iš regresijos lygties nereikšmingą veiksnį, likusių nepriklausomų kintamųjų parametru įverčiai ir standartinės paklaidos keičiasi, todėl turi būti iš naujo įvertinti parametrai ir patikrintas jų reikšmingumas.

Prognozavimas regresija

Prognozavimas yra viena iš labai svarbių regresinės analizės galimybių. Prognozės gali būti taškinės arba intervalinės. Taškinė prognozė y_0 yra apskaičiuojama į regresijos lygtį įstatant prognozinės nepriklausomų kintamųjų reikšmes- x_{i0} . Porinės tiesinės regresijos atveju taškinė prognozės reikšmė bus lygi: $y_0=b_0+b_1x_0$

Intervalinės prognozės apskaičiavimas skiriasi⁴. Intervalines priklausomojo kintamojo prognozės reikšmės (y_0) apskaičiuojamos pagal formules 2,8 ir 2,9.

$$y_0 \in (y_0 \mp t_{\alpha/2, n-k-1} se(f)) \quad 2,8$$

Kur $se(f)$ yra prognozės paklaida, kuri apskaičiuojama pagal tokią formulę:

$$se(f) = \sqrt{\frac{\sum e_i^2}{n-k-1} \left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right]} \quad 2,9$$

Iš prognozės paklaidos formulės matyti, kad prognozės yra tikslesnės, kai

- regresinės lygties paklaidos yra mažesnės (e_i),
- į skaičiavimus įtraukta daugiau stebėjimų (n)
- kai prognozinė reikšmė x_0 yra artima šio kintamojo vidurkiui.

Pavyzdys

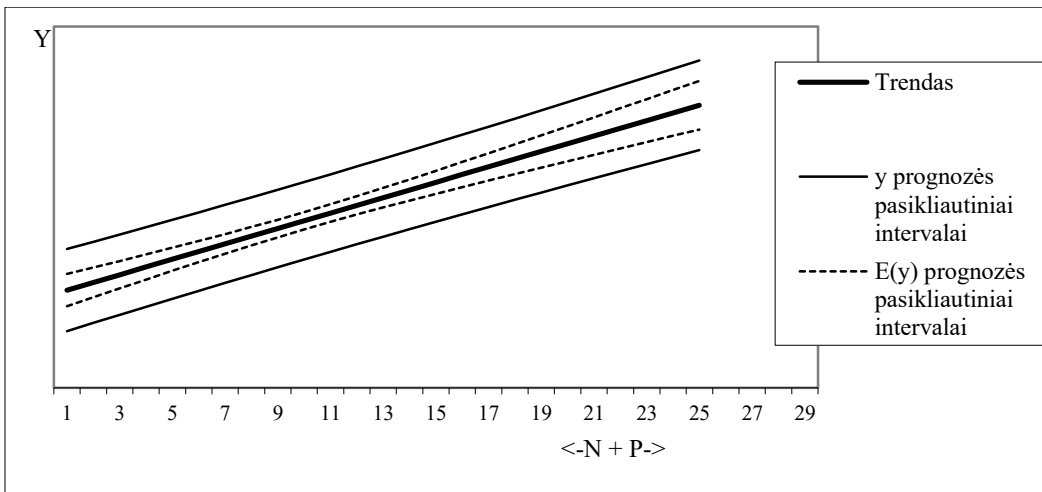
Prisiminkime privačią įdarbinimo bendrovę ir nustatykite priklausomojo kintamojo reikšmės intervalinį prognozės įvertį. Tarkime, norime sužinoti, kokie bus biuro išlaikymo kaštai, jeigu biure dirbs $x_0=16$ darbuotojų. Atlikę skaičiavimus bei pasinaudoję 2.1 lentelės duomenimis, turime $y_0=41+7*16=153$,

$t_{0.025,4} = 2.776$, $se(f) = \sqrt{\frac{300}{6-2} \left[1 + \frac{1}{6} + \frac{16-10.33}{33.33} \right]} = 10.01$ Įsistatę gautas reikšmes į 2.8 formulę,

gauname: $y_0 \in (153 \pm 2,776 * 10,01) = (153 \pm 27,79)$. Taigi prognozuojami bendrovės, kurioje dirbs 16 darbuotojų, kaštai bus intervale $[125,21; 180,79]$ tūkst.eurų per metus .

9.1. pav Priklausomojo kintamojo taškinės (y) ir intervalinės prognozės

⁴ Pateikiamos tik porinės tiesinės regresijos intervalinių prognozių nustatymo formulės.



Regresinė analizė (III)

Paskaitos dalys:

3. Fiktyvūs kintamieji, vėluojančio poveikio įvertinimas.
4. Regresijos ryšio determinuotumas.

Pagrindiniai terminai:

Regresijos ryšio determinuotumas

- Determinacijos koeficientas
- Koreguotas determinacijos koeficientas
- Regresijos reikšmingumas
- F statistika

Fiktyvūs kintamieji, vėluojančio poveikio įvertinimas

- Fiktyvus kintamasis
- Vėluojantis poveikis

1. Fiktyvūs kintamieji, vėluojančio poveikio įvertinimas

Tarp aiškinamąjį kintamąjį įtakojančių veiksnių gali būti tokių, kurie yra kokybiniai, o ne kiekybiniai. Pavyzdžiui, lyties ar išsilavinimo požymiai nagrinėjant gyventojų pajamas ir išlaidas. Regresinėje analizėje į tokius veiksnus atsižvelgiama įtraukiant fiktyvius kintamuosius. Fiktyvūs kintamieji taip pat naudingi, kai reikia eliminuoti vieno ar kelių netipinių stebėjimų įtaką regresijai. Čia fiktyvūs kintamieji labai naudingi įvertinant ekonominių šokų ar tam tikros politikos intervencijos į ekonominius procesus poveikį, pavyzdžiui užimtumo lygiui, bendrajam vidiniam produktui ir pan. Trečia sritis, kurioje plačiai naudojami fiktyvūs kintamieji, – sezoniškumo tyrimai.

Vėluojanti regresija padeda įvertinti kitą dažnai sutinkamą veiksnių sąveikos atvejį, - kai aiškinančiojo veiksnio reikšmės pokytis įtakoja aiškinamąjį kintamąjį tik po tam tikro laiko (pavyzdžiui, investicijos – bendrojo vidaus produkto augimą). Tokią regresiją patogu naudoti prognozavimui, kadangi aiškinančiųjų kintamųjų reikšmės yra dabartiniu metu žinomi dydžiai.

Šiame skyriuje aptariamas fiktyvių kintamųjų taikymas tinka ne tik regresinei, bet ir laiko eilučių analizėje.

Turinys

- Fiktyvių kintamųjų samprata,
- fiktyvių kintamųjų naudojimo atvejai ir tipai,
- vėluojanti regresija,
- vėluojančios regresijos variantai.

Fiktyvių kintamųjų samprata

Fiktyvus kintamasis – tai į regresijos lygtį įtraukiamas veiksnys, įgyjantys ne tikrąsias, o pagal tam tikrus požymius suformuotas fiktyvias reikšmes. Pavyzdžiui, analizuojant įmonių pelningumą, kokybiniu veiksmiu gali būti užsienio kapitalo buvimas įmonės kapitale. Šiuo atveju galima įvairiai organizuoti tyrimą. Pavyzdžiui, į regresiją įtraukiant fiktyvų kintamąjį, kuris įgauna 0 reikšmę, jei įmonės kapitalas suformuotas iš vietinių lėšų ir 1 reikšmę, jei yra pritrauktas užsienio (ar ir užsienio) kapitalas. Tačiau tikėtina, jog tikslingiau tokį tyrimą organizuoti išskiriant daugiau grupių; pvz. vietinio kapitalo, užsienio kapitalo ir mišraus kapitalo įmones. Šiuo atveju reikėtų jau dviejų fiktyvių kintamųjų - vieno, kuris įgautų 1 reikšmę, jei tai yra užsienio kapitalo įmonė, ir dar vieno, kuris įgautų 1 reikšmę, jei įmonė yra mišraus kapitalo; trečias atvejis – kai įmonės kapitalas vietinis – gaunamas kai abiejų minėtų fiktyvių kintamųjų reikšmės lygios nuliui. Aptartas pavyzdys formalizuotas žemiau darant prielaidą, kad be fiktyvių kintamųjų buvo nagrinėjami kiti du įmonių pelningumą aiškinantys veiksniai (pvz., vidutinis įmonės darbo užmokestis ir skolinto bei nuosavo kapitalo santykis):

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 D_{1i} + \beta_4 D_{2i} + \varepsilon_i,$$

kur D_1 ir D_2 yra fiktyvūs kintamieji, įgyjantys 1 arba 0 reikšmes. $D_1=1$, jei įmonė yra užsienio kapitalo, kitu atveju $D_1=0$; $D_2=1$, jei įmonė – mišraus kapitalo, kitu atveju $D_2=0$.

Įvertinus parametrus gauta tokia lygtis⁵:

$$Y_i = 10000 - 140 * x_{1i} + 4500 * x_{2i} + 37000 * D_{1i} + 13000 * D_{2i}$$

Šio hipotetinio pavyzdžio atveju galima daryti išvadą, jog vidutiniškai, kai įmonės kapitalas yra mišrus, metinis pelnas yra didesnis apie 13000 eurų, o kai įmonė yra užsienio kapitalo – 37000 eurų. Vietinio kapitalo įmonei vidutinis pelno dydis būtų apskaičiuojamas be paskutiniųjų dviejų regresijos kintamųjų, t.y. kai $D_{1i}=0$ ir $D_{2i}=0$.

Fiktyvių kintamųjų naudojimo atvejai ir tipai

Aukščiau aptartas vienas fiktyvių kintamųjų naudojimo atvejis, kai reikia įvertinti konkrečių kokybinių veiksnių įtaką rezultatiniam kintamajam. Toks naudojimas yra paplitęs analizuojant pjūvinius duomenis. Regresija tiriant laiko eilutes dažniau reikia įvertinti konkretaus išsiskiriančio stebėjimo ar grupės išskirtinių stebėjimų poveikį. Fiktyvių kintamųjų naudojimas leidžia eliminuoti įprastinių sąlygų ryšiams nebūdingus tarpsnius. Pavyzdžiui, bankų krizė gali sąlygoti pinigų apyvartos greičio priklausomybės nuo palūkanų normų regresijos ryšio pokytį. Todėl į regresijos lygtį yra įtraukiamas fiktyvus kintamasis, įgaunantis 1 reikšmę bankinės krizės įtakos metu. Šiuo atveju gaunama dviguba

⁵ Nagrinėjamas grynai hipotetinis atvejis: veiksniai, ženklai pasirenkami atsitiktinai. Taip pat daroma prielaida jog pasirinkti tinkami veiksniai ir visi koeficientai reikšmingi

nauda: pirmiausia, nustatomas neiškreiptas nuolatinis ryšys (kadangi eliminuojamas netipinis krizės poveikis); antra, netipinė įtaka įvertinama kiekybiškai. Tačiau lieka klausimas, kaip nustatyti, kada prasidėjo ir kada baigėsi išskirtinės sąlygos? Paprasčiausias, tačiau mažiausiai patikimas būdas – nustatyti sąlygų pasikeitimo momentus ekspertiniu būdu. Tiksliau šį klausimą spręsti ekonometriškai (šiam kurse tai nenagrinėjama⁶).

Regresinėje analizėje fiktyviais kintamaisiais galima eliminuoti ar nustatyti sezoniškumo įtaką⁷. Čia pasirenkamas vienas bazinis sezonas, kuriam fiktyvus kintamasis nepriskiriamas. Pavyzdžiui ketvirtinio periodiškumo duomenims būtų taikoma tokia regresija, kurioje bazinis yra ketvirtas sezoniškumo periodas:

$$y_i = \beta_0 + \beta_1 x_{1i} + \gamma_1 D_{1i} + \gamma_2 D_{2i} + \gamma_3 D_{3i} + \varepsilon_i,$$

kur $D_{1i}=1$ pirmą ketvirtį, kitaip $D_{1i}=0$, antrą ketvirtį $D_{2i}=1$, trečią - $D_{3i}=1$.

Ši regresija interpretuojama taip: ketvirto ketvirčio kintamųjų ryšys ($y_i = \beta_0 + \beta_1 x_{1i}$) gaunamas tada, kai D_{1i}, D_{2i}, D_{3i} lygūs nuliui; pirmo ketvirčio ryšys ($y_i = \beta_0 + \beta_1 x_{1i} + \gamma_1$) gaunamas, kai $D_{1i}=1$ ir t.t.

Fiktyvūs kintamieji yra labai naudingi, norint atsakyti į klausimą, ar tam tikros sąlygos, priemonės ir kt. turėjo įtakos nagrinėjam reiškiniui (pvz., reklama arba akcijos prekių pardavimo apimtims). Šiuo atveju regresijos sudarymas nesiskiria nuo ankstesniųjų regresijų, tik joje atsiranda fiktyvus kintamasis, kuris įgauna 1 reikšmę, tiems stebėjimams, kai priemonė buvo taikyta ir 0, kai netaikyta. Atsakymą į klausimą apie poveikį gausime, patikrinę fiktyvaus kintamojo statistinio reikšmingumo hipotezę.

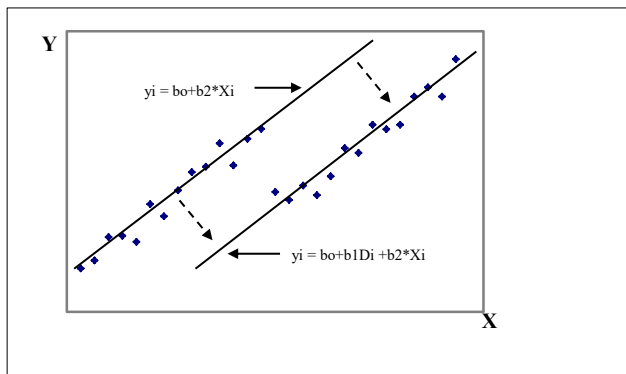
Fiktyvių kintamųjų naudojimas taip pat leidžia vertinti išskirtinių sąlygų poveikio tipą, t.y. kaip kinta regresija apibūdinamas ryšys: fiksuotu dydžiu padidėja/sumažėja reakcijos dydis ar keičiasi jautrumas ir pan. Visa tai leidžia suprasti tiriamo proceso ypatumus ir, susidarant panašioms situacijoms, formuoti atitinkamus atsakomųjų veiksnių scenarijus.

Patys fiktyvūs kintamieji gali būti dviejų tipų: poslinkio ir posūkio. Pirmu atveju nenulinis fiktyvus kintamasis įtakoja regresijos tiesės pasislinkimą ordinačių ašies atžvilgiu (žr. 10.1. paveikslą). Antru atveju – regresijos tiesės nuolydžio kampo pasikeitimą (žr. 10.2. paveikslą).

⁶ Apie įvairaus ryšio grįžimo į pradinę padėtį variantus žr. [4], o struktūrinių lūžių momentų tyrimus [3].

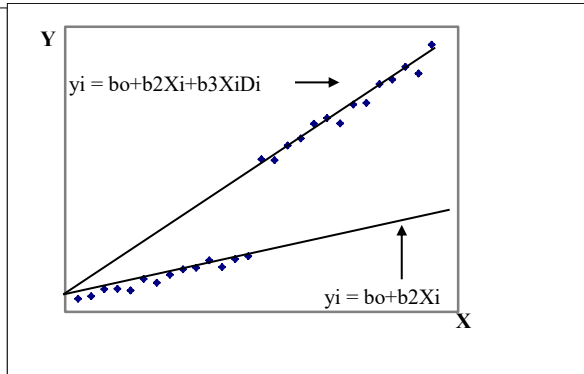
⁷ Tačiau dažniausiai prieš regresine analize tiriant sezonines laiko eilutes, iš jų sezoniškumas pašalinamas, pvz., klasikinio išskaidymo būdu.

Poslinkio fiktyvus kintamasis



4.1. pav.

Posūčio fiktyvus kintamasis



4.2. pav.

Norint įvertinti fiktyvaus kintamojo poslinkio poveikį, naudojama tokia regresijos išraiška:

$$y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 x_{1i} + \varepsilon_i.$$

Kai $D_{1i}=0$, tai regresijos laisvasis (atkarpos) koeficientas lygus β_0 , o kai $D_{1i}=1$ – laisvojo koeficiento bendras dydis apskaičiuojamas kaip β_0 ir β_1 koeficientų suma.

Jei norima įvertinti įtaką regresijos tiesės nuolydžiui, tai naudojama tokia regresijos forma

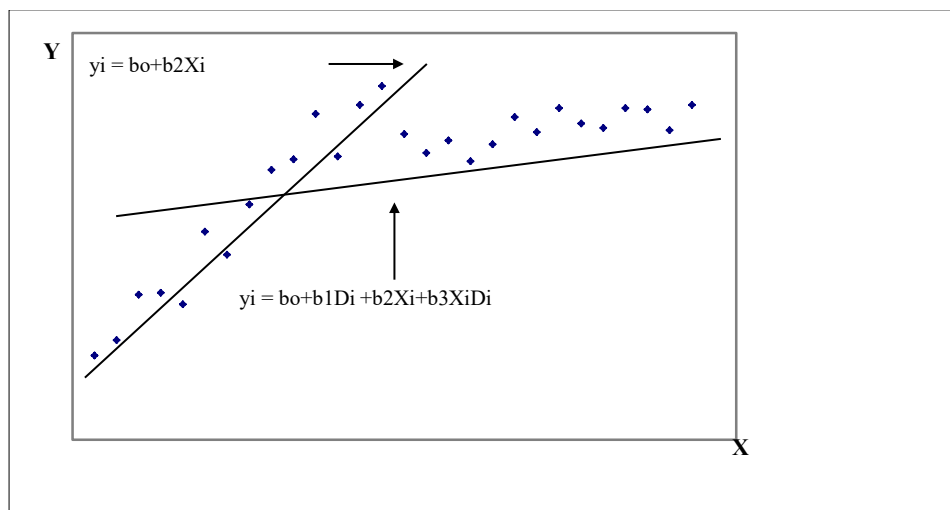
$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{1i} D_{1i} + \varepsilon_i.$$

Šiuo atveju, kai $D_{1i}=0$, regresijos nuolydžio kampo prieš kintamąjį x_{1i} koeficiento dydis lygus β_1 , o kai $D_{1i}=1 - x_{1i}$ koeficiento bendras dydis apskaičiuojamas kaip β_1 ir β_2 koeficientų suma.

Žinoma, dažniausias yra abiejų atvejų derinys (žr.4.3. paveikslą), kurį galima įvertinti apjungiant poslinkio bei nuolydžio lygtis:

$$y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 x_{1i} + \beta_3 x_{1i} D_{1i} + \varepsilon_i.$$

Poslinkio ir posūčio efektas



4.3. pav.

Vėluojanti regresija

Vėlavimas yra neišvengiamas visų socialinių ir ekonominių procesų sąveikos bruožas. Įgyvendintos ekonominės priemonės ar veiksmai duota rezultatus po tam tikro laiko, pvz. investicijos, kvalifikacijos kėlimas ir kt. Kitas dalykas, ar analizuojamo periodiškumo duomenys leidžia tą vėlavimą pastebėti. Pavyzdžiui, jei analizuojami metinio periodiškumo duomenys, tai gamintojų pasiūlos reakcijos į vidutinius kainų pokyčius vėlavimo tikriausiai nenustatysime, tačiau, analizuojant mėnesinio periodiškumo duomenis, toks vėlavimas tikrai bus pastebimas.

Vėluojanti regresija yra tokia regresija, kurios nepriklausomų kintamųjų reikšmės yra atsilikusios per tam tikrą periodų skaičių nuo priklausomojo kintamojo reikšmių (pavyzdžiui, investicijų reakcija į palūkanų normos pokyčius). Vėluojanti regresija galima tik laiko eilučių regresijoje:

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{1t-1} + \dots + \beta_{n+1} x_{1t-k} + \varepsilon_t,$$

kur x_{it} yra nepriklausomas kintamasis, kurio visas poveikis rezultatiniam kintamajam Y_t pasireiškia per k periodų.

Vėluojančioje regresijoje susiduriama su problema, jog didėjant vėlavimui, mažėja bendrų aiškinamojo ir nepriklausomųjų kintamųjų stebėjimų skaičius⁸ (žr. 4.1. lentelę). Kadangi parametrus galima vertinti tik iš tų stebėjimų, kuriuose turimos y_t ir x_{t-k} reikšmės, tai jų skaičius k periodų vėluojančioje regresijoje sumažėja iki $n-k$.

4.1. lentelė. Vėluojančios regresijos kintamieji

y_t	X_{1t-1}	x_{1t-2}	..
		1369.7	..
	1369.7	2837.7	
1369.7	2837.7	5596.4	
2837.7	5596.4	6865.9	
5596.4	6865.9	7288.4	
6865.9	7288.4	7896.1	
7288.4	7896.1	...	
7896.1	...		

⁸ vėlavimui padidėjant kiekvienu papildomu periodu, prarandami 2 stebėjimai.

Vėluojančios regresijos privalumas – labai paprastas prognozavimas: į regresijos lygybę tiesiog įsistatome žinomas dabartines nepriklausomų kintamųjų reikšmes.

Vėluojančios regresijos variantai

Aukščiau aptartas vėluojančios regresijos variantas, kai ankstesnė nepriklausomojo kintamojo reikšmė (x_{t-k}) įtakoja priklausomojo kintamojo vėlesnę reikšmę (y_t). Lygiai taip pat priklausomas kintamasis gali būti aiškinamas ir savo paties ankstesnių laikotarpių reikšmėmis:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \dots + \beta_k y_{t-p} + \varepsilon_t.$$

Regresija, kuri rezultatinį kintamąjį aiškina ankstesnėmis to paties kintamojo reikšmėmis yra vadinama *autoregresija* ir žymima AR(p). Čia p yra vėlavimo periodų skaičius autoregresijoje. Tinkama AR eilė gali būti nustatyta tiriant dinaminės eilutės y_t autokoreliacijos grafiką.

Galimas ir mišrus variantas, t.y. kai rezultatinis kintamasis aiškinamas per savo paties bei egzogeninio kintamojo ankstesnių laikotarpių reikšmes:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \dots + \beta_k y_{t-k} + \beta_{k+1} x_{t-1} + \beta_{k+2} x_{t-1} + \dots + \beta_{k+n+1} x_{t-k} + \varepsilon_t.$$

Veiksnių, kurie rezultatinį kintamąjį įtakoja su tam tikru vėlavimu, taip pat gali būti ne vienas. Kadangi didėjant veiksnių ir vėlavimo laikotarpiui sparčiai mažėja laisvės laipsnių skaičius, tai daugelio veiksnių vėluojanti regresija gali būti taikoma tik labai ilgoms laiko eilutėms.. Stebėjimų ir laisvės laipsnių skaičiaus mažėjimą iliustruoja 4,1 lentelė, kurioje tamsesniu šriftu išskirta duomenų rinkinys, kurio pagrindu galima suskaičiuoti parametrų įverčius vėluojančioje regresijoje ..

Regresinė analizė (III)

. Regresijos ryšio determinuotumas

Nagrinėjant ekonominių reiškinių priklausomybę, svarbu žinoti, sąryšio tarp jų tamprumą, t.y atsakyti į klausimą: kaip stipriai pasirinkti veiksniai įtakoja nagrinėjamą reiškinį. Šiam tikslui yra nagrinėjamas regresijos ryšio determinuotumas ir skaičiuojamas determinacijos koeficientas $-R^2$.

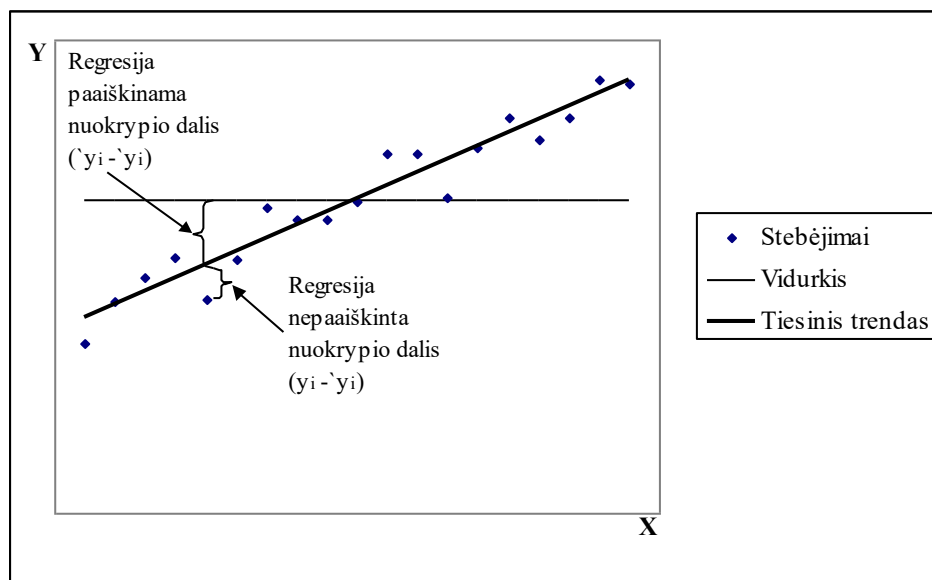
Turinys

- Ryšio determinuotumo samprata,
- Determinacijos koeficientas
- Regresijos statistinio patikimumo tikrinimas

Ryšio determinuotumo samprata

Įvertinti atskirus regresijos parametrus, nustatyti jų reikšmingumą yra tik pradinis regresinės analizės etapas. Toliau reikia nagrinėti, kokia dalimi regresinė funkcija paaiškina priklausomojo kintamojo reikšmių išsibarstymą apie vidurkį. Kuo regresinė lygtis tiksliau aprašo priklausomojo kintamojo pokyčius, tuo ryšys yra labiau determinuotas. Ryšio determinuotumas nustatomas tarpusavyje lyginant regresija ir vidurkiu paaiškinamą stebėjimų išsibarstymą (žr.5,1. paveikslą) - kuo didesnę stebėjimo nuokrypio nuo vidurkio dalį paaiškina regresija, tuo regresinis ryšys yra geriau determinuotas.

Regresija paaiškinta ir nepaaiškinta stebėjimo dalis



5.1. pav.

Nagrinėjant atskirą stebėjimą galima matyti, jog stebėjimo nuokrypis nuo vidurkio suskyla į dvi dalis:

$$(Y_i - \bar{y}) = (\hat{Y}_i - \bar{y}) + (Y_i - \hat{Y}_i),$$

\uparrow \uparrow \uparrow
 Bendras Regresija Nepaaiškinta
 nuokrypis paaiškinta dalis
 dalis

Atitinkamas visų stebėjimų dispersijos išskaidymas:

$$\Sigma(y_i - \bar{y})^2 = \Sigma(\hat{Y}_i - \bar{y})^2 + \Sigma(y_i - \hat{Y}_i)^2$$

\uparrow \uparrow \uparrow
 Visa Regresijos Regresijos
 dispersija reikšmių dispersija likučių dispersija
 (paaiškinta dalis) (nepaaiškinta dalis)

\hat{Y}_i – i-tojo stebėjimo apskaičiuota pagal regresijos funkciją reikšmė, \bar{y} stebėjimų vidurkis, y_i – i-asis stebėjimas.

Regresijos ryšio determinuotumas vertinamas lyginant paaiškintos dispersijos dalies santykį su

$$R^2 = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2} \quad 5.1$$

visa dispersija:

, kur $\Sigma(y_i - \bar{y})^2$ – faktinių y_i reikšmių nuokrypių nuo vidurkio kvadratų suma, $\Sigma(\hat{Y}_i - \bar{y})^2$ pagal regresijos lygtį apskaičiuotų y_i reikšmių nuokrypių nuo vidurkio kvadratų suma.

Šis rodiklis yra vadinamas *determinacijos koeficientu*. Kai regresija paaiškina visą faktinių priklausomojo kintamojo reikšmių išsibarstymą apie vidurkį, determinacijos koeficientas įgyja vieneto reikšmę ($R^2=1$). Kuo mažesnę stebėjimų nuokrypių nuo vidurkio dalį regresinis ryšys paaiškina, tuo determinacijos koeficiento reikšmė artimesnė nuliui (visiško nepaaiškinimo atveju $R^2=0$). Taigi determinacijos koeficiento reikšmė gali būti tarp 0 ir 1. Kuo arčiau vieneto yra determinacijos koeficiento reikšmė, tuo nepriklausomi kintamieji stipriau įtakoja priklausomą kintamąjį. Pvz. determinacijos koeficiento reikšmė $R^2=0,75$, rodo, kad 75 proc. priklausomojo kintamojo pokyčių sąlygoja nepriklausomų veiksnių kitimas.

Tiesinės regresijos atveju determinacijos koeficiento kvadratinė šaknis lygi koreliacijos koeficientui ($\sqrt{R^2} = r$). Netiesinio ryšio atveju ši lygybė negalioja.

Determinacijos koeficientas dažnai naudojamas, norint parinkti tinkamiausią regresijos lygtį. Tačiau jis turi vieną trūkumą - daugėjant regresijoje priklausomų veiksnių skaičiui, determinacijos koeficientas visuomet didėja. Nesvarbu, ar naujai įtrauktas veiksnys yra statistiškai reikšmingas, ar ne. Norint išvengti šio trūkumo, yra skaičiuojamas *koreguotas determinacijos koeficientas*. Šis koeficientas

$$\bar{R}^2 = 1 - \left[\frac{n-1}{n-k-1} (1-R^2) \right] \quad 5.2$$

žymimas \bar{R}^2 ir skaičiuojamas pagal formulę:

kur n – stebėjimų skaičius, k – regresijos nepriklausomų kintamųjų skaičius.

Regresijos lygties statistinio reikšmingumo įvertinimas

Taikant regresinę analizę neužtenka skaitiškai įvertinti, kiek priklausomojo kintamojo kitimo paaiškina nepriklausomų kintamųjų veikimas. Logiška, kad atlikus mažai stebėjimų daryti išvadas apie daugelio veiksnių įtaką, net ir turint aukštą determinacijos koeficientą nėra patikima. Pvz., turint tik dviejų biurų metinius kaštus, padarytos išvados apie darbuotojų skaičiaus, ryšio tarifų ir nuomos kainų pokyčių įtaką įdarbinimo bendrovių biuro išlaikymo kaštams bus mažai vertingos. Norint žinoti, ar galima pasikliauti apskaičiuota regresija yra atliekama regresijos statistinio reikšmingumo tikrinimo procedūra, naudojant Fišerio testą. Tuo tikslu naudojama F statistika, kur α - pasirinktas reikšmingumo lygmuo, k ir $n-k-1$ yra atitinkami laisvės laipsnių skaičiai.

$$F_{k,n-k-1} = \frac{\sum (\hat{y}_i - \bar{y})^2 / (k)}{\sum (y_i - \hat{y}_i)^2 / (n-k-1)} \quad 5.3$$

Jei pagal regresiją nustatyta F statistika yra didesnė už pasirinkto reikšmingumo lygmens teorinę $F_{k-1,n-k}$ skirstinio reikšmę, tai apskaičiuota regresija yra statistiškai reikšminga.

Hipotezės tikrinimo procedūra tradiciškai susideda iš keturių žingsnių:

1. žingsnis. Iškeliame hipotezes:

$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$, (visi parametrai prie nepriklausomų kintamųjų yra lygūs 0 t.y., regresija yra nereikšminga, nes nė vienas veiksnys neįtakoja priklausomojo kintamojo)

H_A : bent vienas iš parametru β_j nėra lygus 0 (regresija statistiškai reikšminga, nes yra bent vienas veiksnys, kuris įtakoja priklausomą kintamąjį)

2 žingsnis Apskaičiuojama pagal formulę 3,3 statistikos reikšmė ir laisvės laipsnių skaičius k , ir $n-k-1$.

3 žingsnis Apskaičiuotą faktinę F reikšmę lyginame su pasirinkto reikšmingumo, pvz., 5 proc. ($\alpha=0,05$), teorine $F_{k,n-k-1}$ reikšme iš F -skirstinio lentelių (žr. priedus)

4 žingsnis Išvada. Jeigu $F_{\text{apskaičiuota}} > F_{k,n-k-1}$, tuomet su 95% pasiklovimo lygmeniu atmetame nulinę hipotezę, jog regresija yra statistiškai nereikšminga ir priimame alternatyvią, kad bent vienas nepriklausomas kintamasis įtakoja nagrinėjamą kintamąjį. Jeigu yra priešingai, t.y., $F_{\text{apskaičiuota}} < F_{k,n-k-1}$, tuomet negalime atmesti H_0 hipotezės, kad kintamųjų priklausomybė yra statistiškai nereikšminga.

Pavyzdys

Grįžkime prie įdarbinimo bendrovės pavyzdžio. Suskaičiuosime determinacijos koeficientą ir patikrinsime hipotezę, ar gauta regresija yra statistiškai reikšminga. Šiam tikslui pradinių duomenų ir skaičiavimų lentelę 2.1 papildykime biuro išlaikymo kaštų variacijos skaičiavimais, kurie pateikti 5 ir 6 stulpelyje.

5.1. lentelė

i-ojo biuro išlaikymo kaštai y_i	Dirbančiųjų skaičius x_i	x_i^2 (3)	Apskaičiuotos pagal regresiją \hat{y}_i reikšmės (4)	$(\hat{y}_i - \bar{y})^2$ (5)	$(y_i - \bar{y})^2$ (6)	$(y_i - \hat{y}_i)^2$
100	10	100	111	5,3	177,8	121
90	8	64	97	265,7	542,9	49
140	14	196	131	660,5	712,9	1
130	12	144	125	136,9	278,9	25
120	11	121	118	22,1	44,9	4
100	7	49	90	542,9	176,8	100
$\sum y_i = 680$	$\sum x_i = 62$	$\sum x_i^2 = 674$		$\sum (\hat{y}_i - \bar{y})^2 = 1633,3$	$\sum (y_i - \bar{y})^2 = 1933,3$	$\sum (y_i - \hat{y}_i)^2 = 300$

Determinacijos koeficiento skaičiavimas

Pirmiausia apskaičiuojame determinacijos koeficiento vardiklį t.y., biuro išlaikymo kaštų kvadratinių nuokrypių nuo vidurkio sumą, kuri yra lygi: $\sum (y_i - \bar{y})^2 = 1933,3$ (skaičiavimai pateikti 6 stulpelyje)

.Determinacijos koeficiento skaitikliui surasti apskaičiuojame \hat{y}_i , įsistatydami į regresijos lygtį

$\hat{y}_i = 41 + 7 x_i$ nepriklausomojo kintamojo reikšmes. (skaičiavimai 4 stulpelyje).

Po to apskaičiuojame \hat{y}_i reikšmių skirtumų nuo \bar{y} vidurkio kvadratų sumą., kuri yra lygi $\sum (\hat{y}_i - \bar{y})^2 = 1633,3$. (skaičiavimai pateikti 5 stulpelyje). Taigi determinacijos koeficientas $R^2 = 1633,3/1933,3 = 0.8448$, kuris parodo, kad darbuotojų skaičiaus pokyčiai sąlygoja 84 proc. biuro išlaikymo kaštų kitimo.

Regresijos reikšmingumo tikrinimas

Atliekame keturis hipotezių tikrinimo procedūros žingsnius.

1. *žingsnis*. Tikriname hipotezę, ar regresinis ryšys statistiškai yra reikšmingas:

H_0 : visi $\beta_1 = 0$, (Mūsų pavyzdyje turime tik vieną nepriklausomą kintamąjį x -darbuotojų skaičių)

H_A : β_1 nėra lygus 0 (regresija statistiškai reikšminga)

2 *žingsnis* Apskaičiuojame pagal formulę 5,3 statistikos $F_{\text{apskaičiuota}}$ reikšmę, ir laisvės laipsnius : $k-1=2-1=1$ ir $n-k=6-2=4$ laisvės laipsniai:

$$F_{\text{apskaičiuota}} = \frac{1633.3/1}{300/4} = 21.777$$

3 *žingsnis* $F_{\text{apskaičiuota}}$ reikšmę lyginame su 5 proc. ($\alpha=0,05$) reikšmingumo teorine $F_{1,4}$ reikšme iš F-skirstinio lentelių (žr. priedus). Matome $F = 21.777 > F_{1,4} = 7.71$

4 *žingsnis Išvada*. Su 95% pasiklovimo lygmeniu atmetame nulinę hipotezę, jog abu įverčiai gali būti lygūs 0 ir priimame alternatyvią hipotezę, kad bent vienas veiksnys reikšmingai įtakoja biuro išlaikymo kaštus. Kaip parodė ankstesnė analizė, darbuotojų skaičius yra reikšmingas veiksnys, įtakojantis biuro kaštus

LAIKO EILUČIŲ ANALIZĖ

Paskaitos dalys:

5. Tendencijų, ciklų, sezoninių svyravimų nustatymas.
6. Prognozavimas pagal laiko eilutes.

Pagrindiniai terminai:

Tendencijų, ciklų, sezoninių svyravimų nustatymas

- Klasikinis išskaidymas
- Trendas
- Ciklai
- Sezoniškumas
- Atsitiktiniai svyravimai (paklaidos)
- Adityvus išskaidymas (sumos)
- Multiplikatyvus išskaidymas (sandaugos)

Prognozavimas pagal laiko eilutes

- Globalus trendas
- Lokalus trendas
- Slenkantys vidurkiai
- Svertiniai slenkantys vidurkiai
- Centruoti slenkantys vidurkiai
- Eksponentinis išlyginimas

1. Tendencijų, ciklų, sezoninių svyravimų nustatymas

Laiko eilučių analizė –tai ekonometrinis tyrimo metodas, kai naudojama tik nagrinėjamą reiškinį apibūdinantys įvairių laikotarpių duomenys. Neturint papildomos informacijos apie proceso dinamikos priežastis ir sąlygas, Analizuojant laiko eilutes galima išsiaiškinti pagrindines tendencijas ir kitus procesus, būdingus duomenų laiko eilutei. Ši analizė padeda suprasti praėjusio laikotarpio nagrinėjamo reiškinio kitimo dėsningumus ir tendencijas; yra naudinga prognozuojant galimus variantus bei priimant sprendimus.

Laiko eilučių analizė apima labai platų problemų spektrą. Šiame kurse nagrinėjamos tik dvi elementariosios analizės: (1) kaip nustatyti laiko eilučių sudėtines dalis (trendo, cikliškumo, sezoniškumo ir atsitiktinių svyravimų komponentus), (2) kaip nustatyti pagrindinę tendenciją ir prognozuoti elementariaisiais metodais.

Žemiau aptariami tokie klausimai:

- laiko eilutės samprata,
- laiko eilutės sudėtinės dalys,

- laiko eilučių išskaidymo sumos ir sandaugos būdai,
- klasikinis laiko eilučių išskaidymas.

Laiko eilutės samprata

Laiko eilutė⁹ – periodiškų stebėjimų visuma, kurių duomenis gali sudaryti tam tikrais laiko momentais fiksuoti stebėjimų dydžiai, arba per periodą stebimų dydžių suma. Laiko eilutės gali būti suformuotos iš įvairaus dažnumo, tačiau vienodo periodiškumo duomenų: valandinių, kasdienių, savaitinių, mėnesinių, metinių ir pan. Išskiriamos dviejų tipų laiko eilutės: intervalinės ir momentinės. Intervalinių laiko eilučių reikšmės yra per laiko periodą agreguoti duomenys. Tokių laiko eilučių stambesnio periodiškumo eilutės gali būti gaunamos apjungiant smulkesnio periodiškumo laiko eilutės stebėjimus. Momentinių laiko eilučių duomenys yra kintamųjų stebėjimai konkrečiomis datomis. Jų stambesnio periodiškumo laiko eilutės gaunamos ne agreguojant smulkesnio periodiškumo laiko eilučių duomenis, o priklausomai nuo eilutės ir analizuojamo kintamojo tipo (pvz., kaip vidurkiai nedarbo lygio rodikliui, kaip svertiniai vidurkiai - palūkanų normoms, ar tiesiog praleidžiant atitinkamus smulkesnio periodiškumo stebėjimus ir fiksuojant tik stambesnio periodiškumo stebėjimus – indeksams bei pan).

6.1. lentelė Laiko eilučių pavyzdžiai

Intervalinė laiko eilutė				Momentinė laiko eilutė			
Lietuvos bendrojo vidaus produkto ketvirtinio dažnumo laiko eilutė ¹⁰ . Šalia - iš jos gauta metinio dažnumo eilutė.				Ketvirtinė nedarbo lygio rodiklio reikšmių seka ¹¹ , iš kurios metinio periodiškumo duomenys gaunami kaip aritmetiniai vidurkiai.			
Ketvirčiai	BVP, mln. LEur.	Metai	BVP, mln. Lt	Ketvirčiai	Nedarbo lygis %	Metai	Nedarbo lygis %
2000 I	10404,7	2004 I	13366,6	2000 I		2004 I	13,0
2000 II	11622,2	2004 II	15346,2	2000 II	15,9	2004 II	11,3
2000 III	11899,2	2004 III	16660	2000 III		2004 III	10,6
2000 IV	11922,4	2004 IV	17067,3	2000 IV	16,9	2004 IV	10,6
2001 I	10888,7	2005 I	14734,5	2001 I		2005 I	10,2
2001 II	12238,2	2005 II	17638,6	2001 II	16,8	2005 II	8,5

⁹ Terminui *laiko eilutė* sinonimiškai gali būti naudojama sąvoka *laiko seka*.

¹⁰ Šaltinis: Lietuvos banko puslapis Internet'e: <http://www.lbank.lt>

¹¹ Šaltinis: Lietuvos ekonominė ir socialinė raida 2006

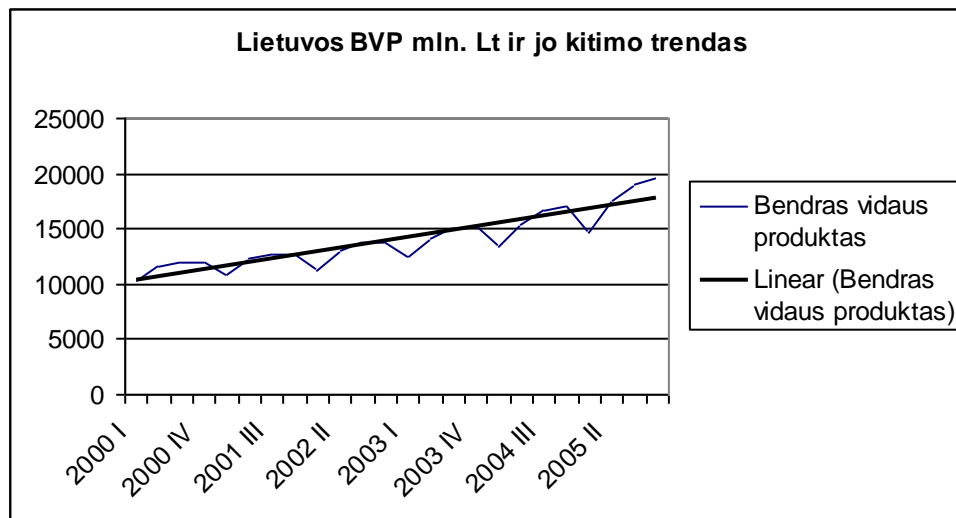
2001 III	12691,3	2005 III	19070,1	2001 III		2005 III	7,2
2001 IV	12744,7	2005 IV	19640,7	2001 IV	17,9	2005 IV	7,1
2002 I	11318,9			2002 I	17,1		
2002 II	13084,7			2002 II	13,0		
2002 III	13810,4			2002 III	11,9		
2002 IV	13734,5			2002 IV	13,0		
2003 I	12442,5			2003 I	13,6		
2003 II	14120,6			2003 II	12,9		
2003 III	14956,4			2003 III	11,6		
2003 IV	15252,4			2003 IV	11,6		

Laiko eilutės sudėtinės dalys

Elementarioje dinamikos eilučių analizėje nagrinėjami įvairūs rodikliai, apibūdinantys stebimų dydžių pokyčius (absoliutinis padidėjimas, didėjimo ir padidėjimo tempai ir t.t.). Sudėtingesnėje laiko eilučių analizėje vienas svarbiausių tyrimo būdu yra jos sudėtinių dalių nustatymas. Sudėtinėmis dalimis dažniausiai yra skiriami tokie laiko eilutės komponentai: (1) trendas, (2) cikliniai svyravimai, (3) sezoniniai svyravimai, ir (4) atsitiktiniai svyravimai.

1) *Trendas*. Laiko eilutės stebėjimų ilgalaikis kitimo pobūdis yra vadinamas laiko eilutės kitimo tendencija. Dažnai ši tendencija yra išreiškiama tam tikra matematine funkcija (žr. 6.1. paveikslą), kuri vadinama trendu. Tokį ilgalaikį kitimo pobūdį gali nulemti įvairūs nagrinėjamą procesą nuolat veikiantys veiksniai, pvz., gyventojų demografinės sudėties, technologijų, vartotojų preferencijų ir paklausos bei pan. pokyčiai. Atkreiptinas dėmesys į tai, jog nagrinėjant tik laiko eilutės duomenis šie veiksniai nėra statistiškai identifikuojami, tačiau ekonominis socialinis pagrindimas yra būtinas analitikui pasirenkant tinkamą funkcinę trendo formą ir ypač prognozuojant. Trendo forma gali būti įvairi: tiesinė, eksponentinė, logaritminė, hiperbolinė, rodiklinė ir pan. Jos pasirinkimas priklauso nuo konkrečių analizuojamų duomenų dinamikos ypatumų ir numanomų proceso raidos priežasčių.

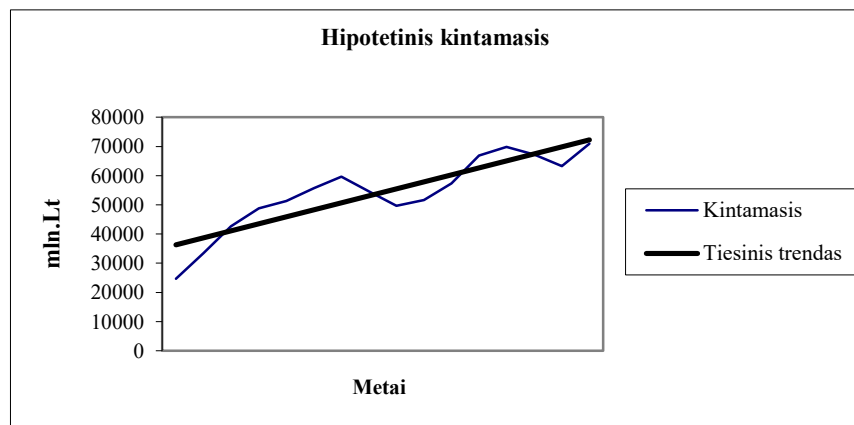
Laiko eilutės pirminių duomenų kitimas



6.1. pav.

2) *Cikliniai svyravimai.* Trendas parodo pagrindinę proceso tendenciją, nuo kurios faktiniai duomenys gali periodiškai nukrypti (grafiškai tai atsispindi faktinių laiko eilutės stebėjimų svyravimuose apie trendą, žr. 6.2. paveikslą). Ciklinius svyravimus paprastai galima pastebėti per ilgesnį laikotarpį, dažniausiai jiems būdingas ne staigus, o daugiau ar mažiau tolygūs perėjimas iš vienos fazės į kitą. Tokių svyravimų pavyzdys – ūkio vystymosi ciklai, kuriuos apibūdina cikliniai bendrojo vidaus produkto, infliacijos, nedarbo ir kitų rodiklių svyravimai.

Cikliniai svyravimai



6.2. pav.

3) *Sezoniniai svyravimai.* Sezoniniais svyravimais vadinami reguliarūs svyravimai, vykstantys metų laikotarpyje. Kartais sezoniniais svyravimais vadinami ir per mažesnę nei metų laikotarpį vykstantys

reguliarūs kitimai, pvz., per savaitę įvykstantys kasdieniai ar per mėnesį įvykstantys savaitiniai ir pan. dinamikos atsikartojimai.

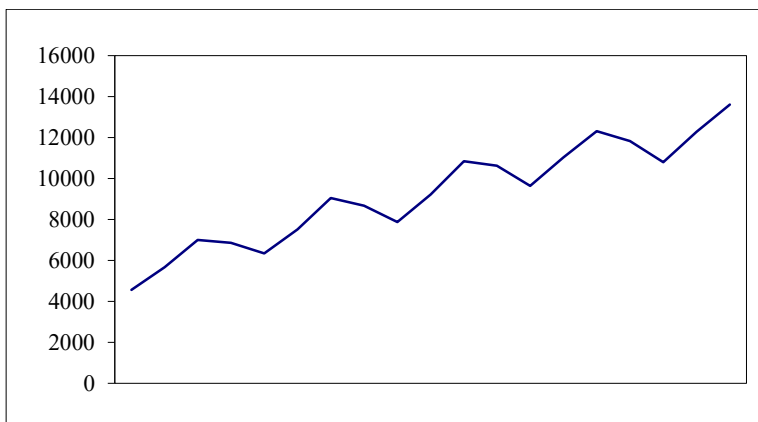
4) *Atsitiktiniai svyravimai*. Išskirti trendo, cikliškumo ir sezoniškumo komponentai tik išskirtiniais atvejais visiškai paaiškina faktinius duomenis. Jais nepaaiškinta laiko eilutės dalis yra vadinama atsitiktiniais svyravimais. Juos sąlygoja trumpalaikiai, netikėti ir nereguliarūs veiksniai. Analitikui būtina atkreipti dėmesį į tai, jog atsitiktinio svyravimo dydį gali įtakoti ir netinkamai parinktas trendas bei analizės procedūra.

Laiko eilučių išskaidymo sumos ir sandaugos būdai

Laiko eilutės gali būti išskaidomos į trendą (T), ciklinius (C), sezoninius (S) ir atsitiktinius (A) svyravimus dviem būdais - sumos ir sandaugos.

Jei ciklinių ar sezoninių svyravimų amplitudės mažai priklauso nuo trendo reikšmių dydžio, t.y. kintant trendo absoliutinėms reikšmėms, svyravimų amplitudė mažai kinta (žr. 6.3. paveikslą), tokiu atveju nustatant laiko eilutės komponentus tikslinga taikyti sumos (arba kitaip adityvinį) būdą:

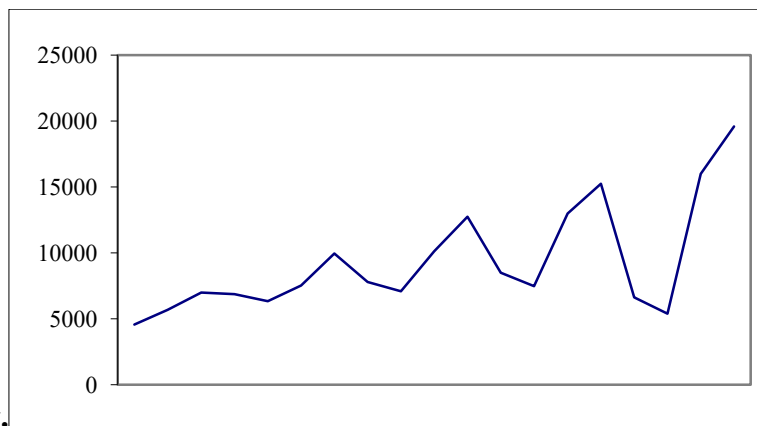
$Y_t = T_t + C_t + S_t + A_t$, kur t – laiko eilutės stebėjimo indeksas, o Y_t – laiko eilutės stebėjimai laiko momentu t .



6.3. pav.

Jei ciklinių ar sezoninių svyravimų amplitudės priklauso nuo trendo reikšmių dydžio, t.y. esant didesnei trendo reikšmei, taip pat didėja arba mažėja ir ciklinių bei sezoninių svyravimų amplitudė (pvz., žr. 6.4. paveikslą), tuomet išskaidant laiko eilutę tikslinga taikyti sandaugos (arba kitaip multiplikatyvinį) būdą:

$$Y_t = T_t \cdot C_t \cdot S_t \cdot A_t.$$



6.4. pav.

Įsidėmėtina, kad nagrinėjant metinius duomenis sezoniškumo komponentas sumos būde lygus nuliui, o sandaugos – vienetui.

Klasikinis laiko eilutės išskaidymas

Pasirinkus laiko eilutės išskaidymo būdą - sumą arba sandaugą - toliau laiko eilutė skaidoma į trendą (T), ciklinių (C), sezoninių (S) ir atsitiktinių (A) svyravimų komponentus. Čia pateikiama klasikinio laiko eilutės išskaidymo procedūra, naudojant sandaugos būdą (sumos skaidymo atveju atliekami analogiški veiksmai, tik vietoje sandaugos ženklo reikia naudoti sumos, o vietoje dalybos – atimties ženklus):

Skaičiavimus pailiustruosime pasinaudodami Lietuvos bendrojo vidaus produkto ketvirtinio periodiškumo duomenimis, pateiktais 6,2 lentelėje. Kiekvienas žingsnis lentelėje atskirtas stulpeliu nurodant procedūros eiliškumo numerį.

1. Pirmiausiai slenkančių vidurkių metodu (apie jį plačiau žr. šio skyriaus antrame skirsnyje Prognozavimas pagal laiko eilutes) išlyginami pirminiai stebėjimai (Y_t). Tokiu būdu nustatomas trendo ir ciklinių svyravimų komponentų bendras dydis ($T_t \cdot C_t$). Sezoniškumo ir atsitiktinumo komponentų ($S_t \cdot A_t$) įtaką slenkantys vidurkiai eliminuoja. Priklausomai nuo to, kokio periodiškumo laiko eilutė analizuojama, pasirenkamas atitinkamas slenkančio vidurkio dėmenų skaičius. Pavyzdžiui lyginant laiko eilutę trijų dėmenų slenkančiu vidurkiu, pirmiausia yra suskaičiuojama pirmų trijų laiko eilutės stebėjimų reikšmių suma, kuri padalinama iš trijų. Taip gaunama pirmoji laiko eilutės slenkančio vidurkio reikšmė, Gautoji reikšmė yra priskiriama slenkančio vidurkio dėmenų viduriniam stebėjimui, t.y. jeigu buvo susumuotos $t=1,2$, ir 3 stebėjimo reikšmės, tuomet gautoji slenkančio vidurkio reikšmė yra

priskiriama antrajam stebėjimui $t=2$. Nagrinėjant ketvirtinio ar mėnesinio periodiškumo laiko eilutę, taikomas atitinkamai keturių arba dvylikos dėmenų slenkantis vidurkis, tuomet turime dvi slenkančio vidurkio periodo vidurio reikšmės t.y., antras ir trečias stebėjimas. Šiuo atveju slenkantį vidurkį reikia papildomai centruoti, t.y. iš dviejų gretimų slenkančio vidurkio reikšmių apskaičiuoti slenkantį vidurkį. .

Pailiustruokime slenkančių vidurkių skaičiavimą, pasinaudodami 6.2 lentelės duomenimis.

Lietuvos BVP duomenys pateikti ketvirčiais, todėl taikysime keturių dėmenų slenkantį vidurkį.

Pirmoji slenkančio vidurkio reikšmė yra lygi keturių pirmų eilutės dėmenų (2000 I, 2000 II, 2000 III, 2000 IV) BVP reikšmių vidurkiui t.y.:

$$\frac{10404,7 + 11622,2 + 11899,2 + 11922,4}{4} = 11462,1$$

Gautą slenkančio vidurkio reikšmę negalime priskirti apskaičiuoto slenkančio vidurkio viduriui, nes dėmenų skaičius lygus keturiems, t.y. yra du periodo vidurio stebėjimai: antras ir trečias. Laikinau priskiriam gautą reikšmę 2000 III ketvirčiui arba trečiajam periodui $t=3$ (žiūrėti 1 stulpelį).

Antroji slenkančio vidurkio reikšmė yra lygi praleidus pirmąjį stebėjimą sekančių keturių BVP eilutės dėmenų (2000 II, 2000 III, 2000 IV , 2001 I) stebėjimų vidurkiui t.y.:

$$\frac{11622,2 + 11899,2 + 11922,4 + 10888,7}{4} = 11583,1$$

Ši reikšmė priskiriama 2000 IV ketvirčiui. Taip, praleidus po vieną dėmenį, yra skaičiuojamas slenkantis vidurkis iki paskutinės galimos keturių dėmenų vidurkio reikšmės. Šie duomenys surašyti lentelės 6,2 3stulpelyje (1 žingsnis T·C stulpelyje).

Keturių dėmenų slenkančio vidurkio reikšmes reikia centruoti, nes įtraukiamų duomenų skaičius yra lyginis t.y., 4.

Pirmoji centruotos eilutės (duomenys 6,2 lentelėje 4 stulpelis Centruoti duomenys T·C) reikšmė yra lygi:

$$\frac{11462,1 + 11583,1}{2} = 11522,6 \text{ ir t.t.}$$

2. Iš faktinių laiko eilutes reikšmių Y_t eliminuojant trendo ir ciklinių svyravimų bendrą komponentę T·C, apskaičiuota slenkančių vidurkių metodu, nustatomas bendras sezoniškumo ir atsitiktinumo komponentių ($S_t \cdot A_t$) dydis. Pateiktame pavyzdyje ši procedūra atliekama,

padalinant BVP faktines reikšmes (2 stulpelis) iš centruotų slenkančių vidurkių metodu išlygintų reikšmių ((4 stulpelis)

$$S_t A_t = \frac{Y_t}{T_t C_{centruotas}}$$

Antrojo žingsnio skaičiavimo duomenys yra pateikti 6,2 lentelės 5 stulpelyje $S_t \cdot A_t$. Pvz.

2000 III ir 2000 IV periodo duomenys:

$$\frac{11899,2}{11522,6} = 1,03, \quad \frac{11922,4}{11660,1} = 1,02 \text{ ir t.t.}$$

3. Siekiant iš bendro sezoniškumo ir atsitiktinių svyravimų dydžio ($S_t \cdot A_t$) išskirti tik sezoniškumo komponentę, reikia apskaičiuoti visų metų kiekvieno sezono (pvz., visų metų I, II, III, ir IV ketvirčių) sandaugų $S_t \cdot A_t$ vidurkius. Apskaičiuotos sezono vidurkių reikšmės yra atitinkamo periodo sezoniškumo indeksas.

Pvz. trečiojo metų ketvirčio sezoniškumo indeksas yra lygus 2000 III, 2001 III, 2002 III ir 2003 III, 2004 III, 2005 III ketvirčių $S_t \cdot A_t$ reikšmių vidurkiui t.y.:

$$\frac{1,03 + 1,04 + 1,05 + 1,05 + 1,06}{5} = 1,05$$

Tokiu pat būdu suskaičiuojame I, II ir IV ketvirčių sezoniškumo indeksus; 0,90; 1,01; 1,04. Gauti keturių ketvirčių sezoniškumo indeksai surašyti prie atitinkamų ketvirčių 6,2 lentelės 6 stulpelyje.

Jei metų periodų (pvz. vieno metų I, II, III, IV ketvirčių) sezoniškumo indeksų vidurkis nėra lygus vienetui, tai kiekvieno periodo sezoniškumo indeksą reikia pakoreguoti padalinant iš metų sezoniškumo indeksų vidurkio.

Pateiktame pavyzdyje sezoniškumo indeksų vidurkis yra lygus:

$$\frac{0,90 + 1,01 + 1,05 + 1,04}{4} = 0,998$$

Gauta reikšmė yra labai artima 1, todėl sezoniškumo indeksų koregavimas, praktiškai nepakeitė sezoniškumo indekso reikšmių. Koreguotos sezoniškumo indeksų reikšmės, kurios po apvalinimo yra lygios pradiniam sezoniškumo indeksams, yra pateiktos 6,2 lentelės 7 stulpelyje.

4. Nustačius sezoniškumo komponentę, iš laiko eilučių pirminių duomenų eliminuojama jo įtaka ir gaunamos laiko eilutės stebėjimų reikšmės be sezoninių svyravimų, t.y., desezonizuoti duomenys:

$$T_t * C_t A_t = \frac{Y_t}{S_{tkoreg}}$$

Pvz.: 2000 I ketvirčio desezonizuoti duomenys apskaičiuojami taip:

$$\frac{10404,7}{0,902} = 11535$$

Šie duomenys surašyti 6,2 lentelės 8 stulpelyje.

5. Eliminavus sezoniškumo įtaką laiko eilutės stebėjimų reikšmėms, apskaičiuojamas desezonizuotų laiko eilutės duomenų trendas, dažniausiai taikant mažiausių kvadratų metodą (MKM). Tiesinio trendo matematinė išraiška atrodo taip:

$$y_t (\text{desezonizuoti}) = \beta_0 + \beta_1 t + \varepsilon,$$

kur y_t – laiko eilutės t -stebėjimo reikšmė, kur t - tai periodo numeris, t.y., $t = 1, 2, 3, 4$, it t.t. atitinkantis laiko eilutės stebėjimo laikotarpį, pvz. 2000 I - tai pirmasis stebėjimas, t.y. $t=1$. 2000 II tai antrasis stebėjimas, $t=2$, ir t.t. Periodų numeriai yra pateikti 6,2 lentelės pirmame stulpelyje.

Mūsų pavyzdyje Excel programos pagalba apskaičiuotas tiesinis BVP 2000-2005 metų trendas yra lygus:

$$y_t (\text{desezonizuoti}) = 10178,0 + 306,48 * t,$$

Šios lygties koeficientai parodo, kad 2000-2005 metais Lietuvos BVP apimtys per ketvirtį vidutiniškai padidėdavo po 306,48 mln. Lt. (prisiminti tiesinės porinės regresijos parametru ekonominę interpretaciją iš šio paskaitų konspekto regresinės analizės dalies)

Taigi 6,2 lentelės 9 stulpelyje pateiktos BVP tiesinio trendo reikšmės. Pirmoji reikšmė yra gaunama:

$$10484 = 10178,0 + 306,48 \cdot 1, \text{ ir t.t.}$$

6. Ciklinių svyravimų komponentė nustatoma kaip trendo ir ciklinių svyravimų komponentių bendro dydžio ($T_t * C_t$) bei trendo reikšmių santykis:

$$C_t = \frac{T_t C_t}{T_t} \quad \text{Mūsų pavyzdyje ciklinių svyravimų duomenys yra pateikti 10 stulpelyje, kurio}$$

reikšmės yra gaunamos padalinus 4,2 lentelės slenkančių vidurkių reikšmes, esančias stulpelyje 4 iš tiesinio trendo reikšmių, esančių 9 stulpelyje.

6.2. lentelė

Lietuvos bendrojo vidaus produkto laiko eilutės klasikiniu išskaidymas sandaugos metodu

		Pirminiai duomenys	Pirmas žingsnis		Antras žingsnis	Trečias žingsnis		Ketvirtas žingsnis	Penktas žingsnis	Šeštas žingsnis	Septintas žingsnis
periodas	Periodo numeris t	BVP mln.Lt. Y_t	Slenkančių vidurkiu išlyginta eilutė $T_t \cdot C_t$	Centruoti duomenys $T_t \cdot C_t$	$S_t \cdot A_t$	Sezoniškumo indeksas S_t	Koreguotas sezoniškumo indeksas $S_{koreguotas}$	Desezonizuoti duomenys $T_t \cdot C_t \cdot A_t$	Trendas tiesinis (MKM) T_t	Cikliniai svyravimai C_t	Atsitiktinė komponentė A_t
	1	2	3	4	5	6	7	8	9	10	11
2000 I	1	10404,7				0,90	0,902	11537	10484		
2000 II	2	11622,2				1,01	1,013	11478	10791		
2000 III	3	11899,2	11462,1	11522,6	1,03	1,05	1,048	11357	11097	1,03	0,99
2000 IV	4	11922,4	11583,1	11660,1	1,02	1,04	1,038	11488	11404	1,02	0,99
2001 I	5	10888,7	11737,1	11836,1	0,92	0,90	0,902	12073	11710	1,00	1,03
2001 II	6	12238,2	11935,2	12037,9	1,02	1,01	1,013	12087	12017	0,99	1,01
2001 III	7	12691,3	12140,7	12194,5	1,04	1,05	1,048	12113	12323	0,99	1,00
2001 IV	8	12744,7	12248,3	12354,1	1,03	1,04	1,038	12281	12630	0,97	1,00
2002 I	9	11318,9	12459,9	12599,8	0,90	0,90	0,902	12550	12936	0,96	1,01
2002 II	10	13084,7	12739,7	12863,4	1,02	1,01	1,013	12923	13243	0,96	1,01
2002 III	11	13810,4	12987,1	13127,6	1,05	1,05	1,048	13181	13549	0,96	1,01
2002 IV	12	13734,5	13268,0	13397,5	1,03	1,04	1,038	13234	13856	0,96	1,00
2003 I	13	12442,5	13527,0	13670,3	0,91	0,90	0,902	13796	14162	0,96	1,02
2003 II	14	14120,6	13813,5	14003,2	1,01	1,01	1,013	13946	14469	0,95	1,01
2003 III	15	14956,4	14193,0	14308,5	1,05	1,05	1,048	14274	14775	0,96	1,01
2003 IV	16	15252,4	14424,0	14577,2	1,05	1,04	1,038	14697	15082	0,96	1,02
2004 I	17	13366,6	14730,4	14943,4	0,89	0,90	0,902	14821	15388	0,96	1,01
2004 II	18	15346,2	15156,3	15383,2	1,00	1,01	1,013	15156	15695	0,97	1,00
2004 III	19	16660	15610,0	15781,0	1,06	1,05	1,048	15900	16001	0,98	1,02
2004 IV	20	17067,3	15952,0	16238,6	1,05	1,04	1,038	16446	16308	0,98	1,03
2005 I	21	14734,5	16525,1	16826,4	0,88	0,90	0,902	16338	16614	0,99	0,99
2005 II	22	17638,6	17127,6	17449,3	1,01	1,01	1,013	17420	16921	1,01	1,02
2005 III	23	19070,1	17771,0			1,05	1,048	18200	17227		
2005 IV	24	19640,7				1,04	1,038	18925	17534		

Pavyzdžiui, 2000 III ketvirtį matyti toks komponentių išskaidymas: $Y_t = 11899$ mln. Lt; $T_t = 11097$ mln.Lt; $C_t = 1.03$; $S_t = 1.05$; $A_t = Y_t / T_t \cdot C_t \cdot S_t = 0.99$.

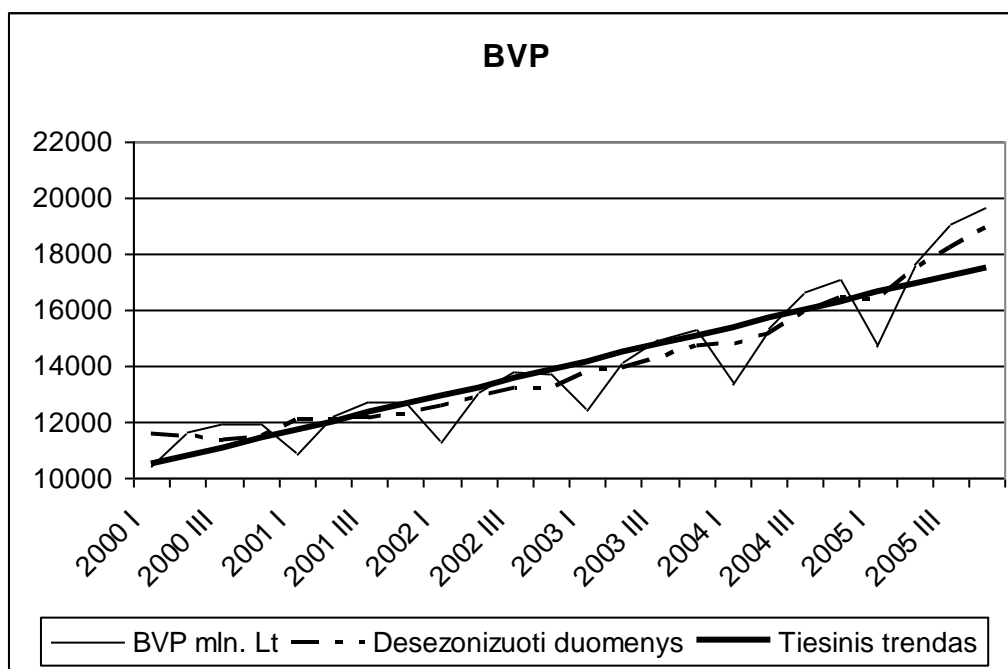
Atlikę BVP duomenų klasikinį išskaidymą, galime daryti tokias išvadas.

- Visų duomenų kontekste išryškėja trečiaisiais ketvirčiais pasireiškiantis žymus sezoniškumo pikas (sezoniškumo indeksas 1,05), t.y BVP šiuo metų laikotarpiu vidutiniškai 5 proc. didesnis, tuo tarpu pirmasis metų ketvirtis rodo vidutinį 10 proc. BVP nuosmukį.
- Tiesinis trendas rodo, kad BVP per ketvirtį 2000-2005 metų laikotarpyje vidutiniškai padidėdavo po 306,48 mln. Lt.

- Taip pat galima konstatuoti tą faktą, jog nagrinėjamas 2000-2005 laikotarpis yra ekonomikos ciklinių svyravimų persilaužymo fazėje. Ekonominiai procesai, eliminavus sezoniškumo ir ilgalaikės tendencijos įtaką, juda iš ciklinio nuosmukio 2000 I -2003 II į ciklinį pakilimą nuo 2003 III –iki nagrinėjamo laikotarpio pabaigos.

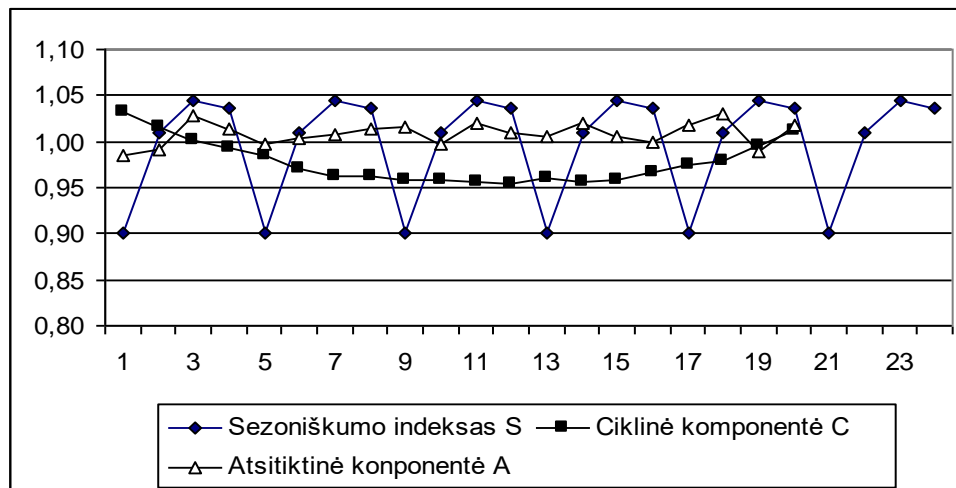
Analizuoti bei daryti išvadas padeda grafinis laiko eilutės sudėtinių dalių vaizdas. 4,4 paveiksle pavaizduotos pirminės reikšmės (BVP), stebėjimai, eliminavus sezoniškumą ($T_t \cdot C_t \cdot A_t = Y_t / S_t$), bei pastariesiems nustatytas trendas (tiesinis $T_t \cdot C_t \cdot A_t$ trendas). 4,5 paveiksle pateikiamos ciklinių, sezoninių ir atsitiktinių svyravimų komponentės.

Pirminių reikšmių, dydžių be sezoniškumo dinamika bei trendas



4,4 pav.

Ciklinių, sezoninių ir atsitiktinių svyravimų komponentai



6.5 pav.

Eilučių analizė neturėtų riboti aptartu atskirų laiko eilučių komponentų nagrinėjimu. Remiantis analitiko disponuojama platesne socioekonominė informacija, turėtų būti nagrinėjamos proceso raidą lemiančios priežastys ir prognozuojama tolimesnis vyksmas.

2. Prognozavimas pagal laiko eilutes

Svarbiausias prognozavimo pagal laiko eilutes privalumas yra tas, jog pakanka turėti tik nagrinėjamojo kintamojo duomenis, taigi prognozavimui reikia minimalių laiko ir finansų sąnaudų.

Prognozėms pagal laiko eilutes gali būti pasitelkiami įvairūs metodai. Iš kurių šiame kurse aptariami tik paprasčiausi – prognozavimas trendu bei laiko eilutės išlyginimas slenkanciais vidurkiais. Toliau aptariami tokie klausimai:

- prognozavimas trendu,
- prognozavimas naudojant išlyginimo metodus
- slenkantys vidurkiai,
- eksponentinis išlyginimas.

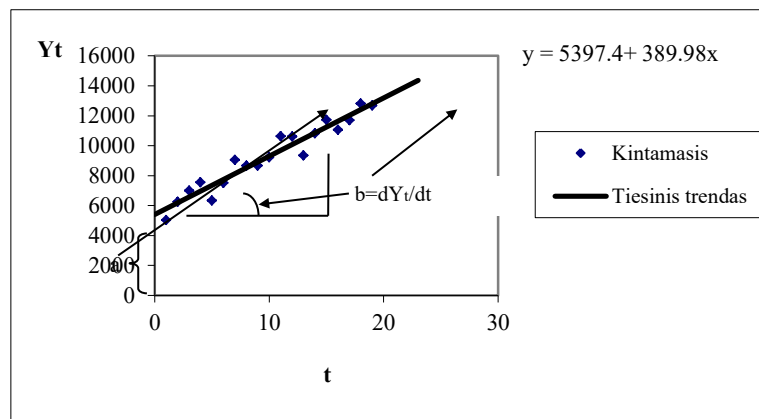
Prognozavimas trendu

Kai laiko eilutei būdinga pastovi kitimo tendencija, prognozei tikslinga naudoti trendą. Paprasčiausia trendo funkcija yra tiesinė:

$$\hat{y}_t = a + bt$$

kur \hat{y}_t – trendo reikšmė, a – laisvasis koeficientas, lemiantis trendo tiesės poslinkį lygiagrečiai t (abscisių) ašiai, b – trendo tiesės nuolydžio kampo koeficientas, t - laikas.

Tiesinio trendo koeficientų skaitinė ir grafinė prasmė



4,6pav.

Dažniausiai trendo koeficientai randami taikant mažiausių kvadratų metodą (MKM) (koeficientai apskaičiuojami analogiškai, kaip ir porinės tiesinės regresijos atveju), kai minimizuojama faktinių

reikšmių nuokrypių nuo teorinių (trendo) reikšmių kvadratų suma, t.y. siekiama rasti tokius a ir b koeficientus, kurie

$$\sum_{t=1}^n (y_t - (a + bt))^2 \rightarrow \min$$

kur n – stebėjimų skaičius, o t –periodo numetis.

Tokie tiesinio trendo koeficientai nustatomi pagal šias formules:

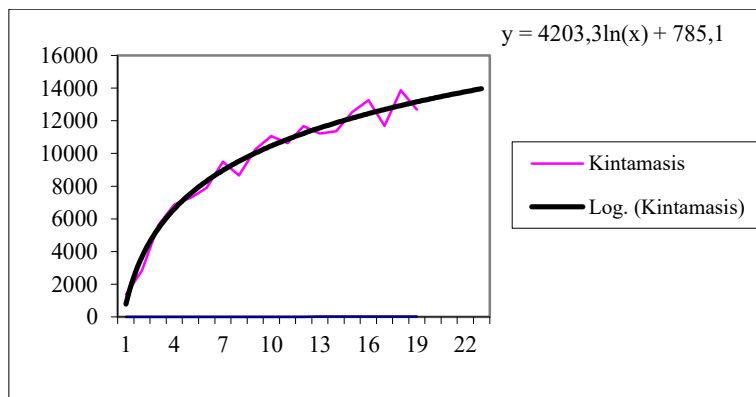
$$b = \frac{n \sum t y_t - \sum t \sum y_t}{n \sum t^2 - (\sum t)^2} \quad a = \frac{1}{n} (\sum y_t - b \sum t)$$

Pavaizdavus procesą grafiškai, neretai galima pastebėti, jog tikslingiau būtų taikyti netiesinę trendo funkciją (pvz., žr.4,7. paveikslą, kur tinkamiausia yra logaritminė funkcija, galinti aprašyti vis mažėjančio prieaugio procesą). Praktikoje dažniausiai pakanka tokių netiesinių funkcijų, kurias logaritmvimu ar pakeitimu galima nesunkiai transformuoti į tiesinę trendo funkciją. Toliau pateikiama dažniausiai naudojamų netiesinių trendo funkcijų pavyzdžių lentelė.

6.3. lentelė Netiesinės funkcijos ir jų transformavimas į tiesinę formą

Pavadinimas	Pirminė išraiška	Į tiesinę formą transformuota išraiška
Eksponentinė	$\hat{y}_t = a \cdot e^{bt}$	$\ln(\hat{y}_t) = \ln(a) + b \cdot t$
Hiperbolinė	$\hat{y}_t = a + b/t$	$\hat{y}_t = a + b \cdot T$, kur $T = 1/t$
Logaritminė	$\hat{y}_t = a + b \cdot \ln t$	$\hat{y}_t = a + b \cdot T$, kur $T = \ln(t)$
Rodiklinė	$\hat{y}_t = a \cdot t^b$	$\ln(\hat{y}_t) = \ln(a) + b \cdot \ln(t)$

Netiesinis trendas



4,7pav

Pirminės išraiškos koeficientų reikšmės nustatomos į tiesinę formą transformuotos lygties koeficientus antilogaritmuojant arba išreiškiant per skaičiavimuose naudotą pažymėjimą.

Nustačius trendo funkcijos parametrus, prognozuojant reikia į trendo funkcinę išraišką (vietoje t) įrašyti norimo prognozės periodo atitinkamą numerį. Numeris sudaromas iš faktinių stebėjimų ir norimų prognozuoti laikotarpių sumos. Pavyzdžiui, jei laiko eilutės stebėjimų skaičius buvo 20, ir norima gauti prognozę vienam periodu į priekį, į trendo funkciją įrašomas 21. Norint gauti prognozę dviem laikotarpiais į priekį, reikėtų įrašyti 22 ir t.t. Pavyzdžiui, jei parametrai $a=500$, $b=0.2$, tai vieno periodo prognozė lygi:

$$\hat{y}_{21} = 500 + 0.2 \cdot 21 = 504.2.$$

Prognozavimas naudojant išlyginimo metodus

Iki šiol aptartas trendas yra vadinamas *globaliu trendu*. Šis trendas ne visada yra tinkamas prognozavimui, kadangi analizuojamas procesas gali būti nevienalytis (pavyzdžiui, situacija gali iš esmės pasikeisti dėl šokų ir struktūrinių lūžių) arba turimas per trumpas stebėjimų skaičius neleidžia įvertinti ilgalaikės proceso tendencijos. Tokiais atvejais prognozei gali būti naudojami *lokalūs trendai*, parodantys atskirų stebėjimo atkarpų tendencijas. Prognozavimo išlyginimo metodu tikslas – “išvalyti” tiriamą procesą nuo atsitiktinės komponentės naudojant tam tikrą išlyginimo (suvidurkinimo) procedūrą. Pagrindiniai naudojami metodai – slenkantys vidurkiai ir eksponentinis išlyginimas (pastarasis taip pat yra slenkančių vidurkių tam tikra variacija).

Slenkantys vidurkiai

Slenkančių vidurkių esmė - laiko eilutės transformavimas į naują laiko eilutę, kurioje atskiri ankstesnės eilutės stebėjimai pakeičiami kelių (pvz., trijų, penkių ir pan.) stebėjimų vidutiniu dydžiu.

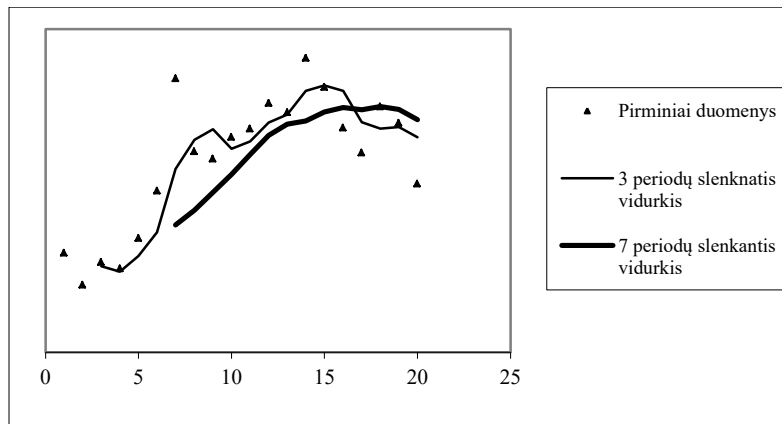
Sąvoka “slenkantys” nusako, jog naujas vidurkis skaičiuojamas iš naujo stebėjimų rinkinio: naujas stebėjimas yra įtraukiamas, o vienas seniausias stebėjimas - pašalinamas. Prognozuojant, paskutinioji slenkančio vidurkio reikšmė yra laikoma proceso prognozuojamu dydžiu. Praktikoje naudojami įvairūs slenkantys vidurkiai, pagrindiniai kurių apibūdinti lentelėje.

6.4 lentelė Slenkančio vidurkio rūšys

Slenkančio vidurkio tipas	Formulė	Pastabos
Paprastas	$MA_t = 1/n \sum_{i=1}^n y_{t-n+i} = 1/n (y_t + y_{t-1} + \dots + y_{t-n+1})$	Vienodai pasveria visus slenkantį vidurkį sudarančius n stebėjimus (dėmenis)
Svertinis	Pvz.: $WMA_t = \frac{\sum(i \cdot y_{t-n+i})}{\sum i} = \frac{[n y_t + (n-1)y_{t-1} + \dots + 1 y_{t-n+1}]}{[n + (n-1) + \dots + 1]}$	Čia stebėjimai pasveriami nevienodai. Sviurių priskyrimo schemas gali būti įvairios. Pavyzdyje pateikiamas variantas, kai senesniai stebėjimai priskiriami vis mažesnis svoris.
Centruotas	$CMA_t = 1/n \sum_{i=1}^n y_i = 1/n [y_{t+(n-1)/2} + y_{t+(n-1)/2-1} + \dots + y_{t-(n-1)/2}]$	Čia išlyginimas realizuojamas pakeičiant pirminės laiko eilutės reikšmes vidurkiu, suformuotu iš vienodo skaičiaus prieš stebėjimą ir po stebėjimo esančių stebėjimų. Šio slenkančio vidurkio trūkumas - vėliausiems periodams slenkantys vidurkiai nėra apskaičiuojami. Be to, siekiant, kad slenkančių vidurkių reikšmės būtų gaunamos laikotarpio pabaigai ir palyginamos su pirminėmis reikšmėmis, slenkančio vidurkio dėmenų skaičius (n) turi būti nelyginis.

Įsidėmėtina, jog didėjant slenkančių vidurkių dėmenų, t.y. stebėjimų, iš kurių skaičiuojamas vidurkis, skaičiui, vis laiko eilutė vis labiau išlyginama, t.y. vis tolydesnė. Palyginimui pateikiama laiko eilutė ir jos 3 bei 7 dėmenų slenkantys vidurkiai (žr. 6.9. paveikslą).

Pirminių duomenų išlyginimas paprastu slenkančiu vidurkiu



4,8. pav

Prognozuojama slenkančiais vidurkiais darant prielaidą, kad paskutinis slenkantis vidurkis yra proceso tolimesnio vystymosi reikšmė.

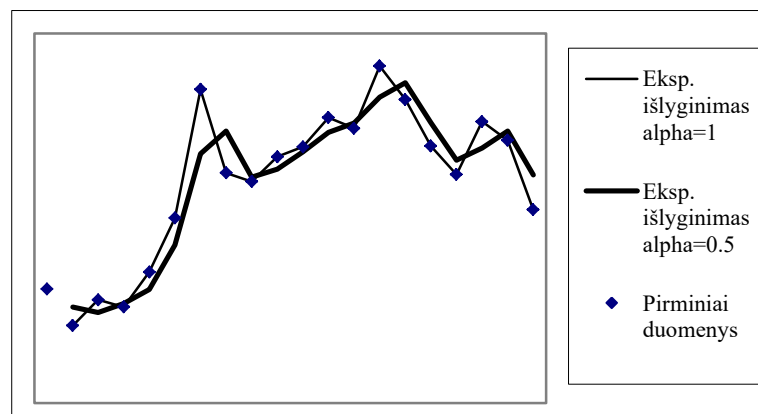
Eksponentinis išlyginimas

Eksponentinis išlyginimas praktikoje naudojamas dažniau nei aukščiau apibūdinti slenkantys vidurkiai. Eksponentinis išlyginimas taip pat yra svertinis vidurkis, kurio praeities stebėjimams priskiriami eksponentiškai mažėjantys svoriai. Eksponentinio išlyginimo matematinė išraiška:

$EWMA_t = w y_t + (1-w) EWMA_{t-1}$, kur w yra analitiko nuožiūra parenkamas išlyginimo faktorius.

Išlyginimo faktoriaus w reikšmė gali kisti nuo nulio iki vieneto. Tačiau paprastai jos reikšmė yra nuo 0,05 iki 0,3. Kuo ji didesnė, tuo didesnis svoris priskiriamas paskutiniam stebėjimui ir mažiau kreipiamas dėmesys į praėjusių laikotarpių išlygintas reikšmes (kraštutiniu atveju, kai $w=1$, $EWMA_t = y_t$, t.y. paskutinio stebėjimo reikšmei); kuo w mažesnė – tuo didesnę įtaką prognozuojamam dydžiui turi ankstesnės išlygintos reikšmės (žr. 6.10 paveikslą).

Pirminių duomenų eksponentinis išlyginimas



4.9. pav.

Nors išlyginimo faktoriaus reikšmės pasirinkimas yra laisvas ir priklauso nuo analitiko sprendimo, tačiau konkrečiai laiko eilutei gali būti nustatyta tam tikra optimali w reikšmė (pvz., tokia, kuri minimizuoja faktinių ir išlygintų reikšmių skirtumų kvadratų sumą).

Čia aptartas tik paprastas eksponentinis išlyginimas, kuris paprastai taikomas palyginti didelių svyravimų neaiškios tendencijos laiko eilutei.

Priedai

Paskaitų konspekte naudojami matematiniai simboliai :

y_i - priklausomojo kintamojo i-ojo stebėjimo faktinė reikšmė ($i=1 \dots n$)

\hat{y}_i - pagal regresijos lygtį apskaičiuota priklausomojo kintamojo i-ojo stebėjimo reikšmė ($i=1 \dots n$)

\bar{y} - priklausomojo kintamojo reikšmių vidurkis

x_{ji} - j nepriklausomojo kintamojo i-ojo stebėjimo reikšmė ($i=1 \dots n$)

\bar{x}_j - j nepriklausomojo kintamojo reikšmių vidurkis

n- stebėjimų skaičius

k nepriklausomų kintamųjų skaičius

β_0 - regresijos lygties laisvojo nario parametras

β_j - regresijos lygties parametras prie j- ojo nepriklausomojo kintamojo

b_0 - regresijos lygties laisvojo nario parametro įvertis

b_j - regresijos lygties parametro prie j ojo nepriklausomojo kintamojo įvertis

σ^2 - regresijos lygties dispersija

ε_i - regresijos modelio paklaida

e_1 - regresijos modelio paklaidos įvertis

SE - regresijos lygties standartinė paklaida

SE_{b_j} - regresijos j parametro įverčio standartinė paklaida

PRAKTINĖ UŽDUOTIS

(Priedas prie ekonometrijos paskaitų konspekto)

Nagrinėkime tokį pavyzdį:

Privati gyventojų įdarbinimo bendrovė turi savo atstovybes 6 miestuose. Kiekvienos atstovybės biuro išlaikymo sąnaudos priklauso nuo daugelio veiksnių: Vienas iš svarbiausių - tai darbuotojų skaičius, nuo kurio priklauso išlaidos darbo užmokesčiui ir įvairūs mokesčiai valstybei. Greta darbuotojų skaičiaus biuro kaštai apima ir kitas sąnaudas - tai išlaidos patalpų nuomai, šildymui, ryšio ir kt. paslaugoms, kurių dydis priklauso nuo jų įkainių.

Įdarbinimo bendrovės biurų išlaikymo kaštai, darbuotojų skaičius ir kitų veiksnių tarifai pateikti 1.1 lentelėje.

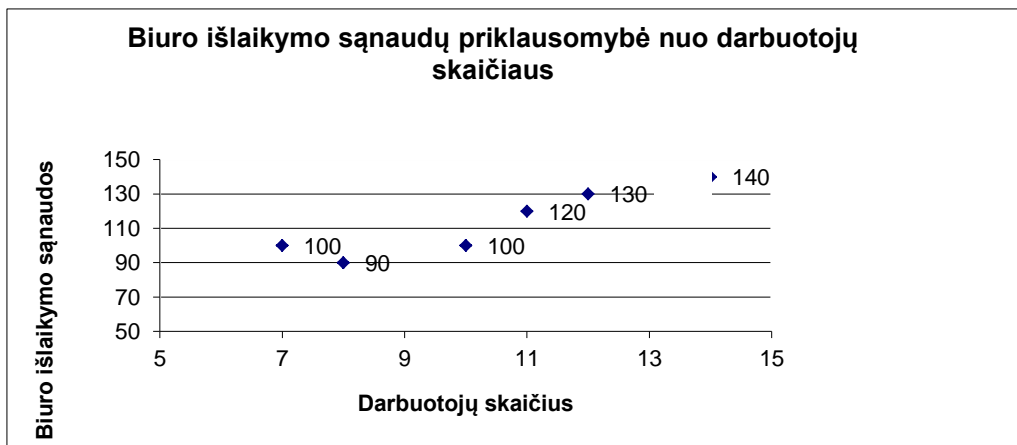
Lentelė 1.1

Veiksniai	Atstovybės					
	Vilniuje	Vilniuje	Kaune	Klaipėdoje	Šiauliuose	Panevėžyje
Biuro išlaikymo sąnaudos tūkst. eur. metus (y_1)	100	90	140	130	120	100
Darbuotojų skaičius (x_1)	10	8	14	12	11	7
Patalpų nuomos tarifai (eur./10m ²) (x_2)	45	45	20	28	18	22
Ryšio paslaugų tarifai (cnt/min.)	65	50	54	65	56	71

Paprastumo dėlei pradžioje apsiribokime biuro kaštų priklausomybės analize tik nuo vieno veiksnio - darbuotojų skaičiaus. Veiksnių priklausomybę galima pateikti grafiškai.

Užduotis 1. Grafinis priklausomybės vaizdavimas

Paveiksle 1.1 biuro išlaikymo sąnaudų ir darbuotojų skaičiaus priklausomybė pavaizduota, kaip taškų sklaida dvimatėje sistemoje. Pasinaudodami pradinių duomenų lentele 1.1, pažymėkite grafike trūkstantą



tašką.

1.1 Pav.

Užduotis 2. Parametrų įverčių skaičiavimas MKM

Tiesinė priklausomybės matematinė išraiška tarp dviejų kintamųjų atrodo taip:

$$y_i = (b_0 + b_1 x_i) + e_i$$

Parametrų įverčiai a ir b apskaičiuojami pagal šias formules:

$$b_0 = \frac{(\sum y_i - b_1 \sum x_i)}{n} = \bar{y} - b_1 \bar{x}; \quad b_1 = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2}$$

Pasinaudodami pradinių duomenų ir atliktų skaičiavimų lentele 1.4, į tuščias vietas įrašykite trūkstamus skaičius.

Lentelė 1.4

i- ojo biuro išlaikymo kaštai y_i	Dirbančiųjų skaičius I ame biure x_i	X_i^2	$y_i * x_i$
100	10	100	1000
90	8	64	720
140	14	196	1960
130	12	144	1560
120	11	121	1320
100	7	49	700
$\sum y_i =$ 680	$\sum x_i =$ 62	$\sum x_i^2 =$ 674	$\sum x_i * y_i =$ 7260

$$b_1 = \frac{6 * 7260 - 62 * \square}{6 * 674 - (62)^2} = 7; \quad b_0 = \frac{(680 - 7 * 62)}{\square} = 41$$

Apskaičiavę b_0 ir b_1 parametrų įverčius galime užrašyti įdarbinimo bendrovės biurų išlaikymo kaštų priklausomybės nuo darbuotojų skaičiaus matematinę lygtį:

$$y_i = (41 + \square x_i) + e_i$$

Užduotis 3. Standartinių parametru įverčių paklaidų skaičiavimas

$$SE_{b_0} = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2 \sum x_i^2}{n-k-1 \sum (x_i - \bar{x})^2}} \quad SE_{b_1} = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{\sum (x_i - \bar{x})^2}}$$

Standartinės parametru įverčių paklaidos skaičiuojamos pagal šias formules:

Pasinaudodami pradinių duomenų ir atliktų skaičiavimų lentelę 2.1, į tuščias vietas įrašykite trūkstamus skaičius.

2.1. lentelė

i-ojo biuro išlaikymo kaštai y_i	Dirbančiųjų skaičius i-ame biure x_i	x_i^2	$(x_i - \bar{x})^2$	$(y_i - \hat{y}_i)^2$
100	10	100	0.11	121
90	8	64	5.44	49
140	14	196	13.44	1
130	12	144	2.78	25
120	11	121	0.44	4
100	7	49	11.11	100
$\sum y_i = 680$	$\sum x_i = 62$	$\sum x_i^2 = 674$	$\sum (x_i - \bar{x})^2 = 33.33$	$\sum (y_i - \hat{y}_i)^2 = 300$

$$SE_{b_0} = \sqrt{\frac{300 * \boxed{}}{6-2 \cdot \boxed{}}} = 15.94 \quad SE_{b_1} = \sqrt{\frac{\boxed{300}}{6-2}} = 1.5$$

Užduotis 4. Koeficientų reikšmingumo tikrinimas

Į tuščias vietas įrašykite trūkstamus skaičius.

Pagal 6 stebėjimų duomenis įvertinta regresija yra :

$$y_I = 41 + 7x_i + e_I \quad (15.94) \quad (1.5)$$

Skliaustuose pateiktos standartinės paklaidos reikšmės SE_a ir SE_b .

1. žingsnis. Tikrinama parametru lygybės nuliui hipotezė

$H_0: \beta_1=0$, (Įvertis prie nepriklausomojo kintamojo x gali būti lygus 0, t.y., x veiksnys yra nereikšmingas)

$H_A: \beta_1 \neq 0$. (Įvertis prie nepriklausomojo kintamojo x nelygus 0 t.y., x yra reikšmingas)

2 žingsnis Apskaičiuojama statistikos $t=b/SE_b$ reikšmė, kuri yra lygi $t=7/1,5 = \boxed{}7$.

3 žingsnis Apskaičiuota faktinė t reikšmė lyginama su Student'o pasiskirstymo $t_{\alpha/2, n-k}$ teorine reikšme. Iš t pasiskirstymo lentelių (žr. priedus) randame $t_{0,025,4} = 2,779$ (laisvės laipsnių skaičius: $n-k=6-2=4$, ir $\alpha/2=0,05/2=0.025$). Matome $t=4,667 > t_{0,025,4} = 2,779$.

4 žingsnis Išvada. Išnagrinėjus pavyzdį su 95% pasiklovimo lygmeniu galima daryti išvadą, jog nulinė hipotezė $H_0: \beta=0$ atmetama ir priimama alternatyvi hipotezė $H_A: \beta \neq 0$, kuri reiškia, kad veiksnys x - darbuotojų skaičius yra statistiškai reikšmingas.