|  |
| --- |
| **Įvadas……………………………………………………………….….……..3** |
| 1. **Regresinė analizė: samprata ir turinys……………………….....….3** |
| 1. **Ekonominis regresinio modelio sudarymo etapas…………..…..…7** 2. **Statistinis regresinio modelio sudarymo etapas…………….….…14** 3. **Verslo situacijos ekonometrinės analizės etapas…………….……33**   **Priedai……………………………………………………………….……….36** |

**Turinys**

**1.1 Trumpas įvadas**

**Ekonometrija - tai ekonomikos teorijos ir matematinės statistikos junginys. E**konometrijos, kaip mokslo šakos atsiradimas siejamas su norvegų ekonomisto ir matematiko Ragnaro Frišo darbų paskelbimu, 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 įvairiausius ekonominius matematinius metodus, iki gana siaurų, kai apsiribojama keliais matematinės statistikos metodais.

*Ekonometrija - tai atskira disciplina , kuri apjungia ekonomikos teoriją ir statistinius metodus, siekiant suteikti skaitinius įvertinimus ekonominiams procesams.*

Dažniausiai ekonometrijos vadovėlių turinį sudaro du metodai: regresinė analizė ir laiko eilučių tyrimo metodai. Šiame paskaitų konspekte yra pateikiami regresinės analizės taikymo verslo vadyboje pagrinai.

# 1,1 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 verslo situaciją 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.

Regresinė analizė yra statistinis metodas, kai naudojant matematines procedūras, gaunama lygtis arba jų sistema, įvertinanti vieno ar daugiau veiksnių įtaką nagrinėjamas reiškiniui.

Turint matematinę priklausomybės lygtį, galima:

1. Matematiškai aprašyti nagrinėjamo reiškinio priklausomybę nuo jį sąlygojančių veiksnių. Konkreti regresijos ryšio matematinė lygtis leidžia gauti ekonominės analizės išvadoms naudingus rodiklius: ryšio ženklą ir įtakos pobūdį, nagrinėjamo reiškinio elastingumą kiekvienam iš veiksnių arba visų veiksnių poveikiui bendrai, veiksnių pakeičiamumo santykius
2. ustatyti nagrinėjamą situaciją veikiančius ir nedarančius jai esminio poveikio veiksnius;
3. Prognozuoti nagrinėjamo reiškinio variantus
4. Modeliuoti įvairias verslo situacijų alternatyvas ir lyginti jas tarpusavyje.

Šiam konspekte nagrinėjami tik vienos lygties regresiniai modeliai. Gauta matematinė lygtis yra vadinama **regresiniu modeliu, kurio b**endras pavidalas atrodo taip:

1.1.1.

kur Y – nagrinėjamas ekonominis reiškinys, išreikštas tamtikru rodukliu (pvz. Pardavimų apimtys, kaina) X1,..., Xk – jį sąlygojantys veiksniai , kurie yra sunumeruoti nuo 1 ikri k. Bendru atveju įtakojantį veiksnį žymėsime kur j ε - modelio paklaida.

Vienos lygties regresinis modelis susideda iš dviejų dalių: **sisteminės** ir **atsitiktinės**. Sisteminė dalis lygtyje 1.11 yra o atsitiktinė

Svarbu įsidėmėti, kad regresinė analizė taikoma ekonominių reiškinių analizei, kuriuose veiksnių tarpusavio ryšiai yra tikimybiniai ir netaikoma esant griežtai determinuotam veiksnių sąryšiui. Veiksnys Y ir X ryšys yra tikimybinis, kai esant tai pačiai X veiksnio reikšmei, Y gali įgauti nevienodas reikšmes. Pvz. dažniausiai produktų kainų ir kaštų ryšys yra tikimybinis. Ūkininkas prekiaujantis turguje gali parduoti savo išaugintus obuolius, kurių auginimo ir sandėliavimo kaštai vienodi, skirtingomis kainomis net ir tą pačią prekybos dieną. Todėl obuolių kaina turi tikimybinį pobūdį ir todėl yra atsitiktinis dydis. 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. išleidžia pramogoms pinigų, jeigu jų pajamos yra 500 ar 1000 litų per mėnesį. Sąryšis tarp gyventojų pajamų ir išlaidų pramogoms irgi yra tikimybinis, nes du gyventojai turintys vienodas pajamas išleis skirtingas sumas pramogoms, o be to skirsis ir to paties gyventojo išlaidos pramogoms skirtingais mėnesiais. 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, kurios bus visais periodais vienodos. Šiuo atveju atsakymą gausime susumavę visų gyventojų išlaidas. Tačiau tokios informacijos nėra ir vargu ar gali būti.

Apibrėžiant regresijos sampratą naudojome sąvokas "nagrinėjama verslo situacija arba ekonominis reiškinys ir jį sąlygojantys veiksniai. Toliau pereisime prie trumpesnių ir griežtesnių terminų.

*Priklausomas kintamasis - tai kairėje regresijos lygties pusėje esantis kintamasis (yi lygtyje 1.1.1), kurio vidutinių reikšmių pokyčius stengiamasi paaiškinti kitų - dešinėje lygties pusėje esančių kintamųjų pokyčiais.*

Nepriklausomi kintamieji (Xj) - tai dešinėje lygties pusėje esantys kintamieji, kurie veikia priklausomąjį kintamąjį () Nepriklausomųjų kintamųjų reikšmės gali laisvai kisti, t.y. jų neįtakoja priklausomojo kintamojo reikšmės.

## Regresinis modelis gali būti porinis arba dauginis.

***Porinis regresinis modelis*** yra tokia matematinė lygtis, kurioje vertinamas dviejų kintamųjų ryšys, t.y., kai regresijos lygtyje vertinamas tik vieno nepriklausomojo kintamaojo X įtaka priklausomam kintamajam.

***Dauginis regresinis modelis*** tai priklausomybės lygtis, kurioje nepriklausomų kintamųjų yra daugiau nei vienas. Tokio modelio 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.

**Tiesinis ir netiesinis regresinis modelis**

Pats paprasčiausias ekonometrinis modelis – tai porinis tiesinis regresinis modelis, kurio matematinė išraiška yra pateikta apačioje:

Kaip matome šis modelis susideda iš trijų komponentų: kintamųjų Y ir X; koeficientų ; ir paklaidos

Tiesiškumas arba netiesiškumas gali būti visų trijų komponentų atžvilgiu, Pateikta \_\_\_lygtis yra tiesinė kintamųjų, koeficientų bei paklaidos atžvilgiu. Žemiau pateikiami netiesinių modelių pvz.

netiesiškumas kintamųjų atžvilgiu

netiesiškumas koeficientų atžvilgiu

**Regresinio modelio koeficientai, parametrai ir įverčiai**

Kaip buvo minėta, regresinio modelio tikslas yra kiekybiškai įvertinti pasirinktų veiksnių įtaką nagrinėjamam reiškiniui. Šią įtaką atspindi regresinio modelio koeficientai, kurie yra apskaičiuojami modelio sudarymo eigoje. Šie koeficientai gali būti dviejų tipų: parametrai arba įverčiai.

Parametrai, tai tikrieji sąryšio koeficientai, kurie yra žymimi graikiškomis raidėmis . Tikrosios parametrų reikšmės labai 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, tarp kurių dalis yra nežinomi arba kiekybiškai neišmatuojami (pvz. gyventojų skoniai arba vienos ar kitos prekės vertinimo kriterijai). Be to, tenka reikalingus analizei duomenis pakeisti statistinėse ataskaitose fiksuojamais mažiau tiksliais duomenimis ir apsiriboti, tik tam tikra jų imtimi. Prie subjektyvių kliūčių priskirtini tyrėjo padaryti netikslumai parenkant veiksnius, nustatant priklausomybės matematinę išraišką, klaidas suvedant informaciją ir kt. Todėl ekonometrijoje yra sakoma, kad skaičiuojant regresines lygties koeficientus gaunami nagrinėjamų duomenų pagrindų apskaičiuoti įverčiai, o ne tikrosios parametrų reikšmės. Siekiant pabrėžti, kada yra kalbama apie tikrąsias parametrų reikšmes, o kada apie parametrų įverčius, parametrams ir įverčiams yra naudojami skirtingi žymėjimai. Tikrąsias parametrų reikšmes žymime graikiškomis β, ε, o jų įverčius lotyniškomis raidėmis: b ir e.

**Regresinio modelio sudarymo etapai ir žingsniai .**

**Galima skirti tris**  regresinio modelio sudarymo etapus

1. Ekonominis
2. Statistinis
3. Ekonometrinės analizės

Kiekviename etape atliekami tam tikri žingsniai ir daromos atitinkamos išvados. Regresinio modelio etapus papildžius žingsniais regresinio modelio sudarymo procedūra atrodo taip: 1

**etapas: EKONOMINIS**

Pirmas *žingsnis***: Ekonominės problemos formulavimas.**

Analizuojamas verslo situacijos esminė problema ir tikslas. Sudaromas galomai įtakojančių veisknių sąrašas.

*Rezultatas*: Įvardinamas nagrinėjamas reiškinys ir jį įtakojantys veiksniai, kurie yra būsimo modelio kintamieji.

Antras *žingsnis***: Ekonominių hipotezių iškėlimas**

*Analizuojamas kiekvieno iš veiksnių sąveikos su nagrinėjamu reiškiniu, kryptis ir pobūdis.*

*Rezultatas*: Užduodami reikalavimai matematinei modelio išraiškai

Trečias*žingsnis***: Duomenų rinkimas**

*Rezultatas*: Sudaromos nagrinėjamą reiškinį ir įtakojančius veiksnius apibūdinančios duomenų lentelės.

**2 etapas: STATISTINIS**

*Ketvirtas žingsnis***: Grafinė duomenų analizė.**

Braižomos linijinės bei sklaidos diagramos

*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)*

*Šeštas žingsnis***: Parametrų įverčių skaičiavimas**

Excel skaičiuokle arba kitomis specialiomis programomis modelio koeficientai

*Rezultatas. Užrašomas modelis su skaitiniais koeficientais*

*Septintas žingsnis***: Veiksnių statistinio reikšmingumo analizė**

Taikomos hipotezių tikrinimo procedūros.

*Rezultatas****.*** *Atsakoma į klausimą, kurie veiksniai statistiškai reikšmingai veikia nagrinėjamą reiškinį****.*** *o kurių įtaka nėra statistiškai reikšminga*

*Aštuntas žingsnis***: Viso modelio patikimumo tikrinimas**

Tikrinamas modelio determinuotumas ir klasikinių modelio prielaidų tenkinimas

*Rezultatas. Atsakoma į klausimą ar modelis sudarytas korektiškai ir jį galima taikyti ekonominei analizei ir prognozavimui.*

**3 etapas EKONOMETRINĖS ANALIZĖS**

Devintas *žingsnis***: analizei taikomos apskaičiuotos modelio rodiklių skaitinės reikšmės**

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

Dešimtas *žingsnis***: Ekonominių scenarijų kūrimas, prognozavimas.**

## Tolesnė paskaitų konspekto medžiaga bus pateikiama nuo pirmojo iki dešimtojo žingsnio nuosekliai.

## Todėl dėstona medžiaga sudalinta pagal etapus ir žingsnius. Pateikiamas pavyzdys, kurio pagrindu bus aiškinamas kiekvieno žingsnio turinys. Visi žingsniai pradedami nuo jų tikslo ir paskirties, po to pateikiant ekonometrinę teoriją, bei Excel skaičiuoklės komandas. Žingsnio aprašymas baigiamas skaičiavimo rezultatų išaiškinimu ir įžvalgomis.

2. Ekonominis regresinio modelio sudarymo etapas.

Kepyklėlės „Grūdas prie grūdo“ pardavimų vadybininkas, planuodamas ateities pardavimo problemas ir investicijų pagrįstumą turi numatyti būsimas duonos pardavimo kainas. Išnagrinėjęs mėnesines finansines ataskaitas nuo 2009 metų suvokė, kad būtina tam tikra analizės sistema ir procedūra. Prisiminęs ekonometrijos kursą bakalauro studijoje nusprendė įvertinti kokią įtaką kepamos ruginės duonos kainai turės brangstančios duonos kepimo sąnaudos. Pradėjo nuo regresinio modelio sudarymo pirmojo žingsnio

Pirmas žingsnis**: Nagrinėjamos verslo situacijos detalizavimas. Veiksnių įvardinimas**

Vadybininkas, nusprendė nustatyti pagrindinių duonos kepimo sąnaudų: rugių, cukraus, elektros energijos ir dyzelino kainų bei darbuotojų algų poveikį kepamos ruginės duonos kainai. Pirmiausia įtraukė pagrindines juodos duonos žaliavas: rugiius ir cukrų, Dyzelino kaina buvo pasirinkta kaip veiksnys, nuo kurio priklauso pagrindinės transportavimo išlaidos, susijusios su pagamintos produkcijos ir žaliavų, reikalingų produkcijos gamybai, pervežimu. Tuo tarpu elektros energija bei darbo jėga yra naudojama gaminant produkciją, todėl jų kaina (darbo jėgos kaina – vidutinis darbo užmokestis) įeina į gamybos kaštus ir veikia galutinio produkto kainą. 2009 m. rugsėjį padidintas PVM tarifas taip pat turėjo pakelti duonos kainą. Taigi vadybininko tikslas sudaryti regresinį modelį, kuris parodys, kokį poveikį turės kiekvienas iš šių veiksnių duonos kainai.

*Pirmojo žingsnio rezultatas*: Įvardinti modelio kintamieji: priklausomas kintamasis yduonos kaina –tai nagrinėjamas reiškinys – ruginės duonos kaina, nepriklausomi kintamieji: duonos kepimo produktų: rugių (xrugių kaina ) ir cukraus (xcukraus kaiana ), kaina bei energetinių: resursų kainos: elektra (xelektros kaina ), dyzelinas (xdyzelino kaina ), taip pat, darbo užmokestis (xdarb užm ).

Antrasis žingsnis: **Ekonominių hipotezių iškėlimas.**

Kiekvieno iš išvardintų veiksnių kainos padidėjimas (kitiems veiksniams nekintant) didina duonos savikainą, o tai neišvengiamai turi atsiliepti ir duonos kainai. Bet kurio vieno iš šešių nepriklausomų veiksnių reikšmės sumažėjimas (kitiems veiksniams nekintant) mažina duonos sąnaudų vertę , o tai , taip pat, gali atsilieti gamybos kaštams, bet nebūtinai duonos kainai, nes ne visuomet ji yra mažinama, jeigu pirkėjų perkamoji galia ir paklausa nėra sumažėjusi. Tačiau duonos kainos reakcijos jautrumas arba elastingumas į veiksnių kainos kitimą gali skirtis. Tikėtina, kad rugių kainai pakilus, proporcingai augs ir duonos kaina, nes tai yra pagrindinė duonos žaliava. Tačiau degalų arba elektros kainos augimas gali turėti ir silpnėjantį poveikį duonos kainai, jeigu kepyklėlė pradės naudoti kurui ir energijai taupesnes mašinas ir technologijas. Apibendrinant galima pasakyti, kad visų veiksnių kainų padidėjimas iššaukia ir duonos kainos padidėjimą. Schematiškai galima pavaizduoti šiuos sąryšius tarp veiksnių tokiu būdu

**X1↑→Y↑ ; X2 ↑→Y↑; X3 ↑→Y↑; X4 ↑→Y↑; X5 ↑→Y↑; X6↑→Y↑**

Rodyklės rodo, kap keičiasi duonos kaina, didėjant sąnaudų kainoms. Jeigu rodyklės yra tos pačios krypties, tuomet, įvertinus modelį, prie šio kintamojo turėtų būti teigiamas koeficientas, Jeigu rodyklių kryptis yra priešinga, tuomet modelyje bus neigiamas koeficientas

jeigu nėra panaudota atvirkštinė kita mažėjanti funkcija.

**Trečias žingsnis – tai duomenų rinkimas.**

Regresiniai modeliai yra sudaromi naudojant dviejų tipų duomenis: laiko arba skerspjūvio duomenis, t.y., to paties veiksnio praeities faktinius duomenis arba dabarties skirtingų objektų pvz., skirtingų kepyklų, skirtingų rajonų arba miestų ir kt. duomenis. Šiuose modeliuose nėra būtina, kad visų kintamieji: Y ir X1 , X2 , ...X6 , būtų matuojami vienodais matais.

*Regresinių modelių duomenys dažniausiai būna skerspjūvio stebėjimų arba laiko eilutės.*

*Skerspjūvio (erdvės) duomenys - tai informacija apie stebėjimo vienetų:pvz.darbuotojų, įmonės padalinių, įmonių, vartotojų, namų ūkių, regionų, miestų ir kt. būklę konkrečiu laiko momentu. Žymėjimas pvz.:(yi ; xij). Kai i=1,2…N, kur N- skaičius, parodantis, kiek tyrime įtraukta stebimų objektų.*

*Laiko duomenys (laiko eilutės) - tai informacija apie stebėjimo vieneto : pvz.vieno darbuotojo, įmonės padalinio arba vienos įmonės kokio nors rodiklio kitimą laike. Žymėjimas pvz.:(yt; xjt). kur t=1,2…T, kur T- skaičius, parodantis, kiek turime stebimų objektų.*

*Duomenys gali būti skirtingo pobūdžio:* ***pradiniai-absoliutūs, santykiniai, intervaliniai, ranginai.***

***Pradiniai- absoliutūs duomenys. Pvz.: parduotų prekių vienetais arba verte.*** *Pateiktame pavyzdyje mėnesinis darbo užmokestis litais yra tokių duomenų pavyzdys. Pradiniai absoliutūs duomenys, išreikšti piniginiais matais gali būti realūs arba nominalūs priklausomai nuo kainų indekso. Pasirinkimo problema dažniau iškyla nagrinėjant laiko eilutes. Patartina naudoti realius duomenis. Dėl kainų poveikio tarp dviejų nominalių kintamųjų gali atsirasti melaginga koreliacija. Norint įtraukti realius duomenis reikia juos koreguoti pagal infliacijos indeksą.*

***Santykiniai duomenys*** *– tai duomenys pateikti ne absoliučiu dydžiu, bet tam santykiu, pvz.: BVP/1 vienam gyventojui, pelnas/ produkcijos vienetui, ir kt. Rodiklio pokyčiai( duonos kainos pokytis per tam tikrą laiką arba lyginant su per laikotarpį) bei augimo tempai (duonos kainos pokytis lyginant su ataskaitiniu laikotarpio reikšme) taip pat priklauso santykiniams rodikliams. Pradinius duomenis galima įtraukti ir logaritmene forma (pvz. imamas ne darbo užmokesčio dydis, o jo logaritmuota reikšmė). Toks kintamojo pakeitimas atspindi procentinį reiškės pokytį.*

Pavyzdyje apie duonos kainą surinkti tokie duomenys: duonos kaina, išreikšta litais už kg.; dyzelino kaina, išreikšta litais už litrą; -rugių kaina, išreikšta litais už toną; elektros energijos kaina, išreikšta centais už kilovatvalandę; vidutinis darbo užmokestis, išreikštas litais per mėnesį, tokiose ekonominėse veiklose kaip žemės ūkis, maisto produktų gamyba, didmeninė ir mažmeninė prekyba; cukraus kaina, išreikšta litais už kilogramą. Įsidėmėtina, kad regresinėje analizėje labai svarbų žinoti kintamųjų matavimo vienetus, kurie gali būti labai skirtingi, kaip yra pateiktame pavyzdyje.

Duomenų lentelė atrodo taip:

1.Lentelė. Duonos kainos priklausomybės modelio duomenys

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Metai/Mėnuo | Duonos kaina, Lt/kg | Dyzelino kaina, Lt/ltr | Rugių kaina, Lt/t | Elektros kaina, ct/kWh | Vid. Darbo Užmokestis, Lt/mėn. | Cukraus kaina, Lt/kg |
| 2009/03 | 4,50 | 3,22 | 238,87 | 23,54 | 1884,43 | 3,18 |
| 2009/04 | 4,51 | 3,12 | 244,12 | 23,54 | 1884,43 | 3,21 |
| ir toliau | Duomenys paimti iš : <http://www.stat.gov.lt/lt/> bei <http://www.lesto.lt/> | | | | | |

Iš viso į tyrimą buvo įtraukti 36 stebėjimai. Duonos kainos buvo fiksuojamos nuo 2009 m. kovo mėn. iki 2012 m. vasario mėn. imtinai. Kadangi praktiškai nepriklausomų veiksnių reikšmės negali įtakoti to paties mėnesio priklausomo veiksnio reikšmę, buvo fiksuojama kaip vieno mėnesio nepriklausomi veiksniai įtakos priklausomą veiksnį po dviejų mėnesių, pvz.: kaip 2009 m. balandžio mėn. dyzelino, rugių, elektros, darbo jėgos ir cukraus kainos įtakos 2009 m. birželio mėn. duonos kainą.

***Regresinio modelio duomenys***

*Regresinių modelių duomenys dažniausiai būna skerspjūvio arba laiko.*

*Pjūvio (erdvės) duomenys - tai informacija apie stebėjimo vienetų:pvz.darbuotojų, įmonės padalinių, įmonių, vartotojų, namų ūkių, regionų, miestų ir kt. būklę konkrečiu laiko momentu. Žymėjimas pvz.:(Yi ; Xij). Kai i=1,2…N, kur N- skaičius, parodantis, kiek tyrime įtraukta stebimų objektų.*

*Laiko duomenys (laiko eilutės) - tai informacija apie stebėjimo vieneto : pvz.vieno darbuotojo, įmonės padalinio arba vienos įmonės kokio nors rodiklio kitimą laike. Žymėjimas pvz.:(Yt; Xjt). kur t=1,2…T, kur T- skaičius, parodantis, kiek turime stebimų objektų.*

*Duomenys gali būti skirtingo pobūdžio:* ***pradiniai-absoliutūs, santykiniai,***

***Pradiniai- absoliutūs duomenys.*** *Pvz.: parduotų prekių vienetais arba verte. Pateiktame pavyzdyje mėnesinis darbo užmokestis litais yra tokių duomenų pavyzdys. Pradiniai absoliutūs duomenys, išreikšti piniginiais matais gali būti realūs arba nominalūs priklausomai nuo kainų indekso. Realių ir nominalių duomenų pasirinkimo problema dažniau iškyla nagrinėjant laiko eilutes. Patartina naudoti realius duomenis. Dėl kainų poveikio tarp dviejų nominalių kintamųjų gali atsirasti melaginga koreliacija. Norint įtraukti realius duomenis reikia juos koreguoti pagal infliacijos indeksą.*

***Santykiniai duomenys*** *– tai duomenys pateikti ne absoliučiu dydžiu, bet tam santykiu, pvz.: BVP/1 vienam gyventojui, pelnas/ produkcijos vienetui, ir kt. Rodiklio pokyčiai( duonos kainos pokytis per tam tikrą laiką) bei augimo tempai (duonos kainos pokytis lyginant su ataskaitiniu laikotarpio reikšme) taip pat priklauso santykiniams rodikliams. Pradinius duomenis galima įtraukti ir logaritmene forma (pvz. imamas ne darbo užmokesčio dydis, o jo logaritmuota reikšmė). Toks kintamojo pakeitimas atspindi procentinį reiškės pokytį.*

Nagrinėjamą verslo situaciją gali veikti ne vien tik veiksniai, kuriuos galima išreikšti kiekybiniais matais, bet ir kokybiniai veiksniai, pvz., krizės paveikio duonos, kainai, PVM mokesčio tarifo pakitimai. Kokybinius veiksnius į regresinį modelį galima įtraukti pasinaudojant fiktyviais (arba pseudo) kintamaisiais. Pateiktame pavyzdyje galima būtų traukti tokius kokybinius kintamuosius: PVM mokesčio tarifo pasikeitimas nuo 18 iki 21proc. 2008 metų krizės poveikis, bei sezoniškumo įtaka rugių arba duonos kainai.

*Fiktyvūs kintamieji,*

*Fiktyvus kintamasis – tai į regresijos lygtį įtraukiamas veiksnys, įgyjantys ne tikrąsias, o pagal tam tikrus požymius suformuotas fiktyvias reikšmes. Pavyzdžiui, analizuojant įmonių pelningumą, kokybiniu veiksniu 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):*

*yi = β0+ β 1X1i +β2X2i+β3D1i+β4D2i+ε i,*

*kur D1 ir D2 yra fiktyvūs kintamieji, įgyjantys 1 arba 0 reikšmes. D1=1, jei įmonė yra užsienio kapitalo, kitu atveju D1=0; D2=1, jei įmonė – mišraus kapitalo, kitu atveju D2=0.*

*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, finansinė krizė gali sąlygoti pinigų apyvartos greičio priklausomybės nuo palūkanų normų regresijos ryšio pokytį. Todėl į regresijos lygtį yra įtraukiamas fiktyvus kintamasis, įgaunantis 1 reikšmę krizės įtakos metu. Šiuo atveju gaunama dviguba 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. Regresinėje analizėje fiktyviais kintamaisiais galima eliminuoti ar nustatyti sezoniškumo įtaką. Čia pasirenkamas vienas bazinis sezonas, kuriam fiktyvus kintamasis nepriskiriamas. Pavyzdžiui ketvirtinio periodiškumo duomenims būtų taikoma tokia regresija, kurioje bazinis yra ketvirtas sezoniškumo periodas:*

*yi = β0+ β1x1i+δ1D1i+ δ2D2i+ δ3D3i+ε i,*

*kur D1i=1 pirmą ketvirtį, kitu atveju D1i=0, antrą ketvirtį D2i=1, trečią - D3i=1.*

*Ši regresija interpretuojama taip: ketvirto ketvirčio kintamųjų ryšys (Yi = β0+ β1x1i) gaunamas tada, kai D1i, D2i, D3i lygūs nuliui; pirmo ketvirčio ryšys (Yi = β0+ β1x1i+ δ1) gaunamas, kai D1i=1 ir t.t.*

*Svarbu įsidėmėti, kad jeigu kokybinis veiksnys gali įgauti kelias būsenas, (pvz. ketvirtinių duomenų sezoniškumas turi 4 sezonus) tai įtraukiamų į modelį fiktyvių kintamųjų turi būti vienu mažaiu, negu būsenų, (t.y., įtraukiame 3 fiktyvius kintamuosius, o vieną būseną-sezoną pavadiniame bazine, kurios atžvilgiu bus įvertintas kitų trijų sezonų įtaka.Įtraukus keturis sezonų fiktyvius kintamuosius, negalima bus apskaičiuoti koeficientų prie fiktyvių kintamųjų, nes tokiu atveju susiduriame su fiktyvių kintamųjų spąstais, kai skaičiuojant koeficientus atsiranda dalyba iš 0.*

*Fiktyvūs kintamieji yra labai naudingi, norint atsakyti į klausimą, arkokybinis veiksnys t.y., 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, priimti atitinkamus sprendimus*

*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).*

### Poslinkio fiktyvus kintamasis Posūkio fiktyvus kintamasis

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

Yi = βo+β1Xi+δ1D1+εi

Kai D1=0, tai regresijos laisvasis (atkarpos) koeficientas lygus β0, o kai D1=0, – laisvojo koeficiento bendras dydis apskaičiuojamas kaip β0 ir δ1 koeficientų suma.

Jei norima įvertinti kokybinio veiksnio poveikį modelio nepriklausomam kintamajam t.y., nagrinėjamą reiškinį įtakojančiam veiksnui, tuomet naudojama tokia regresijos forma

Yi = βo+β1Xi+δ1D1Xi +εi

Šiuo atveju, kai, D1=0, regresijos nuolydžio kampo prieš kintamąjį X1 koeficiento dydis lygus β1, o kai D1=1, – X1 koeficiento bendras dydis apskaičiuojamas kaip β1 ir δ1 koeficientų suma.

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

Yi = β0+δ1D1i + β1Xi+δ2XiD1i+εi.

## Poslinkio ir posūkio efektas

Pateiktame pavyzdyje aosiribota tik vieno fiktyvaus kintamojo: PVM mokesčio tarifo pakeitimo poveikio duonos kainai vertinimu. Šis veiksnys –tai poslinkio fiktyvuss kintamasis – kurių įtaka pasireiškia tiesiogiai nagrinėjamam reiškiniui. Pavyzdyje keliama hipotezė, kad padidintas pridėtinės vertės mokestis gali pakelti parduodamo produkto kainą. Todėl į modelį įtrauktas kintamasis D,- PVM pakeitimas, kuris įgauna reikšmes D1=0 iki 2009 spalio mėnesio, t.y., kai buvo 19 proc. PVM tarifas, o. nuo 2009 lapkričio mėn, D1=1 kai įsigaliojo 21 proc. tarifas .

Regresinio modelio ekonominio ekonominis etapas baigiamas sudarius duomenų lentelę. Nagrinėjamo tyrimo duomenų lentelė yra pateikta ( žr. lentelę priede 1)

2. STATISTINIS EKONOMETRINIO modelio sudarymo etapas.

**Ketvirtas žingsnis: Grafinė duomenų analizė**

Pieš pradedant skaičiuoti regresinio modelio statistikas, patartina nubraižyti priklausomojo kintamojo reikšmių išsibarstymo diagramą ir 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 nepriklausomu kintamuoju. Diagromose vertikalioje ašyje atidedamos priklausomojo kintamojo, o horizontalioje ašyje įtakojančių veiksnių reikšmės. Jeigu nagrinėjamas reiškinys priklauso nuo kelių kintamųjų, tuomet rekomenduotina nubrėžti priklausomojo kintamojo grafikus su kiekybiniais nepriklausomais kintamaisiais Šiuo atveju, dvimačių diagramų būtų tiek, kiek yra nepriklausomų kiekybinių kintamųjų. Pavyzdyje apie duonos kainos priklausomybę turime penkis veiksnius, žemiau pateikiama priklausomojo kintamojo ir penkios sklaidos diagramos.

|  |  |
| --- | --- |
| 1.pav. Duonos kainos kitimas | 2 pav. Duonos kainos priklausomybė nuo dizelino kainos |

|  |  |
| --- | --- |
| 3.pav. Duonos kainos priklausomybė nuo rugių kainos | 4 pav. Duonos kainos priklausomybė elektros kainos |

|  |
| --- |
| 5.pav. Duonos kainos priklausomybė nuo darbo užmokesčio |

Atlikta grafinė duomenų analizė leidžia pastebėti tokias įžvalgas:

* 1 paveiksle matomas duonos kainos augimas viso laikotarpio metu. Visgi kainos didėjimo tendencijoje galima išskirti 2010-06 iki 2011-06 laikotarpį, kai duonos kaina augo gana sparčiai. Kiti du periodai: 2009/03 - 2010-09 ir 2011/09-2012/02 stabilios kainos laikotarpiai.
* 2 pav. pavaizduotoje diagramoje, stebime gana plačią duonos kainų sklaidą, esant dyzelino kainai intervale nuo 3Lt. iki 3,50Lt. o, kai kaina artėja prie 4 Lt. Ir ją viršija, tuomet stebimas didėjančios dyzelino kainos poveikis duonos kainai.
* 3 pav. matome priklausomybę tarp duonos ir rugių kainos. Kuo didesnė rugių kaina, tuo aukščiau kyla ir duonos kaina.
* 4. diagramoje pavaizduota elektros kainos įtaka duonos kainai. Iš diagramos matyti, kad nagrinėjamu laikotarpiu buvo tik trys elektros kilovatvalandės tarifai: 22,18 ct/kW; 23,54ct/kW ir 29,31ct/kw. Esant trečiajam elektros tarifui, stebimos labai įvairios duonos kainos. Tokia situacija leidžia spėti, kad elektros kainos pokyčiai neturės labai reikšmingos įtakos duonos kainos kitimui.
* 5 diagramoje pateikiama duonos kainos priklausomybė nuo darbo užmokesčio kitimo. Šioje diagramoje nėra lengva įžvelgti akivaizdžios priklausomybės.

**Penktas žingsnis: Matematinės modelio formos parinkimas**

**Tiesinė ir netiesinė regresija**

Tiek porinės, tiek dauginės regresijos matematinė išraiška kintamųjų yI ir xji atžvilgiu gali būti ne tik tiesinė, bet ir netiesinė Lentelėje apačioje pateiktos dažniausiai naudojamos regresijos lygties matematinės funkcijos. Visas šias lygtis matematinių procedūrų pagalba (logaritmuojant ir prilyginant regresijoje esančių netiesinių kintamųjų reikšmes naujai įvestų kintamųjų reikšmėms, pvz. VI = ln(XI) nesunkiai galima pakeisti į tiesinę formą

**Dažniausiai naudojamos regresijos funkcijos (porinė regresija)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Regresijos lygties forma** | **Matematinė regresijos lygties išraiška** | **Pakeitimai ir pažymėjimai** | **Tiesinė modelio išraiška** | **Modelio pavadinimas** |
| Tiesinė | Yi=β0 + β1Xi | - | Yi=β0 + β1Xi | Lin- modelis |
| Eksponentinė | Yi= β0(eβ xi)  Logatitmuojame | ln(Y ) = β0+β1Xi  ln(Y) =Z | Z = β0+β1 Xi | Log-lin |
| Logaritminė | yi = β0+ β1ln(Xi) | ln(Xi)=Vi | Yi = β0+ β1Vi | Lin-log |
| Hiperbolinė | yi = β0+ β1(1/xi) | 1/Xi=Vi | Yi = β0+ β1Vi | Atvirkštinė |
| Kvadratinė | yi = β0+β1xI+ β2 xi2 | Xi2=Vi | Yi=β0+β1XI+β2Vi2 | Antro laipsnio daugianarė |
| Rodiklinė | Yi = β0 (Xiβ)  Logaritmuojame | ln(Yi)=ln(β0)+β1ln(Xi)  ln(Yi)=Z; ln(β0)= β‘0 ln(Xi)=V | Zi=β‘0+ β1Vi | Log-log |

|  |  |
| --- | --- |
|  | *Tiesinė regresija* Tai - paprasčiausias sąryšis tarp įtakojančio ir priklausomojo veiksnio. Sąryšio pobūdis išlieka pastovus, esant tiek mažoms tiek ir didelėms x reikšmėms  Tiesinė matematinė išraiška dažnai naudojama produkcijos ir kaštų sąryšiui aprašyti. Pvz.: turime modelį: Yi=β0 + β1Xi ,. kuriame Yi – tai produkto kaina; β0 – fiksuoti produkcijos gamybos kaštai, o sandauga- β1Xi – tai kintami kaštai , kurioje β1parodo X resurso (pvz. darbo jėgos) sąnaudas produkcijos vienetui. |
|  | *Eksponentinė regresija* – Tai x veiksnio poveikio priklausomam kintamajam kintančio poveikio funkcija. Kurios ypatybė yra ta, kad esant x nedidelėms reikšmėms y gana lėtai auga, jeigu koeficientas β1>0, tačiau spartėjančiu tempu. Todėl esant x didesnėm reikšmėm augimas tampa vis spartesnis. Pvz. gyventojų skaičiaus augimas pasaulyje arba užkrečiamų ligų plitimas (. gripo) tam tikroje teritorijoje  Esant β1<0,x veiksnio poveikis sparčiai mažina y reikšmes, tačiau silpnėjančių tempu, kol pasiekia tokį lygį, kai x kitimas daro labai nežymų poveikį. Pvz. Išmetamų teršalų mažėjimas priėmus įstatymą apie leidimą parduoti tik aukštos kokybės degalus. Po nutarimo įsigaliojimo vis daugiau mašinų pradės naudoti švarius degalus, ir todėl išmetimai mažės gana sparčiai, tačiau, kai dauguma pereis prie naujų degalų teršalų išmetimai statilizuosis prie tam tikros ribos. |
|  |  |

X

|  |  |
| --- | --- |
|  | *Logaritminė regresija* (lin-log)Tai – kintančio poveikio kreivė, kuriai būdingas silpnėjanti x veiksnio įtaka priklausom kintamajam y. Pvz. mažėjantis žemės derlingumasmetams bėgant, jeigu žemė nėra tręšiama.  Arba mažėjantis darbo našumas dėl nuovargio, didėjant darbo valandų skaičiui Pvz.: turime modelį: Yi=β0 + β1ln(Xi) ,. kuriame yi – tai surinktų braškių kiekis (kg), o x darbo valandų skaičius per dieną. Tikėtina, kad po 8-9 darbo valandų našumas pradės kristi  Gali būti ir neigiamas silpėjantis x poveikis kintamajam y. Pvz. mažėjanti mažėjanti šeimos išlaidų dalis maisto produktams, didėjant šeimos pajamoms, jeigu šeimos narių skaičius nekinta. |
|  | *Hiperbolės regresija (atvirkštinė)* – Tai nepastovaus X veiksnio poveikio priklausomam kintamajam funkcija ,kuriai yra būdingas atvirkštinis ryšys tarp y ir x kintamųjų t.y. x didėjant y mažėja, tačiau y mažėjimas turi neperžengiamą ribą, žemiau, kurios y reikšmės nenukrenta. Atvirkšinės kreivės forma turi Filipso kreivė makroekonomikos teorijoje, kuria remiantis aprašomas sąryšis tarp infliacijos ir nedarbo lygio, darant prielaidą, kad infliacijai didėjant nedarbo lygis mažėja, tačiau neperžiangia natūralaus nedarbo lygio ribos. |
|  | Kvadratinė regresija (2 laipsnio daugianarė funkcija) išsiskiria tuo, kad turi lūžio tašką, kuris dalina kreivę į augimo ir smukimo periodus (kai β2 koeficientas prie X2 yra neigiamas) ir atvirkščiai kai mažėjimo ir didėjimo periodus, kai β2 teigiamas). Makroekonomikoje kvadratine funkcija yra išreikšta Laferio kreivė, kurios pagalba nustatomas ryšys tarp mokestinių pajamų surinkimo į biudžetą ir mokesčio tarifo reikšmės. T.y., surenkamos mokestinės pajamos į šalies biudžetą didėja, didinant mokesčio tarifą, tačiau tik iki tam tikro lygio, kurį peržengus žmonės praranda motyvaciją dirbti ar pradeda slėpti pajamas, todėl surenkamos pajamos pradeda mažėja. Kvadratinė funkcija yra antros eilės polinomo funkcija. Regresinei analizei galima naudoti ir aukštesnių eilių polinomines 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. |
|  | Rodiklinė (log-log) regresija gali įgauti labai įvairias formas, todėl ji gana dažnai yra naudojama. Mikroekonomikoje Kobo-Duglaso funkcija yra būtent log-log funkcija,susieti pagamintos produkcijos apimtis su gamybos ištekliais pvz. darbo jėga ir kapitalu. Tuomet apskaičiuotas laipsnio rodiklis parodo sąnaudų elastingumą. |

Įvairių matematinių išraiškų panaudojimas regresiniuose modeliuose labai praplečia ekonometrinio modeliavimo taikymo galimybes, tačiau sukelia tyrėjui klausimų, o kokią būtent formą parinkti analizuojamai verslo situacijai tirti. Atsakyti į šį klausimą gali padėti, iškeltos antrame žingsnyje hipotezės apie veiksnių sąryšį, skaidos diagramos bei determinuotumo rodikliai, apie kuriuos bus kalbama kitame skyrelyje.

Norint sudaryti log-log modelį pateiktame pavyzdyje pradžioje reikėtų atlikti duomenų pakeitimus, kaip parodyta lentelėje, t.y., visus pradinius duomenis pakeisti jų logaritmais. Lentelėje apačioje pateikiami visi 1 lentelės duomenys, bet ne absoliučia, o logaritmine forma,

2 .Lentelė. Duonos kainos priklausomybės nuo sąnaudų kainų logaritmuoti duomenys

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Metai/Mėnuo  ***Excel skaičiuoklės pagalba duomenys logarittmuojami, nurodant “=ln(pele nurodytas langelis su pirmuoju pradinių duomenų stebėjimu” Po to pele nutempiant šio langelio įrašą per visą stulpelį pvz.: =LN(S3) Taip pakeičiama visa duomenų lentelė*** | Ln (Duonos kaina, Lt/kg) | Ln(Dyzelino kaina, Lt/ltr) | Ln(Rugių kaina, Lt/t) | Ln(Elektros kaina, ct/kWh) | Ln(Vid. Darbo Užmokestis, Lt/mėn). |
| 2009/03 | 1,50 | 1,25 | 5,48 | 3,16 | 7,54 |
| 2009/04 | 1,51 | 1,25 | 5,50 | 3,16 | 7,54 |

Atlikus grafinę analizę galima pereiti prie modelio matematinės formulės užrašymo. Esminei reikalavimai matematinei formai yra pateikti antrame žingsnyje formuluojant veiksnių sąryšio hipotezes, pagal kurias veiksnių poveikis yra tiesioginis, t.y., veiksnio kainai didėjant, duonos kaina taip pat didėja, tačiau šis didėjimas gali būti ir lėtėjančiais tempais. Todėl galime manyti, kad teisingas gali būti tiesinis arba logaritminis ryšys. Šiam teiginiui neprieštarauja ir nubraižytos diagramos .

Fiktyvūs kintamieji modelyje gali būti įtraukiami tik teisine forma, nes matematiškai nėra įmanoma apskaičiuoti logaritminės arba atvirkštinės funkcijo, kai yra nulinės stebėjimų reikšmės. Todėl duonos kainos pavyzdyje fiktyvus kintamasis PVM tarifas yra įtrauktas į modelį tiesine forma.

Baigdamas grafinę analizę tyrėjas gali pasirinkti ne vieną, o kelias modelio alternatyvas. Kepyklėlės pavyzdyje pasirinktos modelio dvi alternatyvas: pirmoji TS- tiesinis modelis ir antroji LN –logaritminis modelis: Abiejų modelių matematinė išraiška pateikta

|  |  |
| --- | --- |
| TS –tiesinis modelis | Yduonos kaina =β0 +β1Xdyz kaina+β2Xrugių kaina+β3Xelek kaina+β4Xdarbu užm +β5DPVM  + ε |
| LN –logaritminis modelis | ln(Yduonos kaina )=β0 +β1ln(Xdyz kaina)+β2(Xrugių kaina)+β3 (Xelek kaina)+β4ln(Xdarbu užm )+β5DPVM + ε |

**Šeštas žingsnis: Parametrų įverčių skaičiavimas**

Pagrindinis regresinės analizės uždavinys - 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 parametrų įverčiais

Atlikus šį žingsnį penktame žingsnyje užrašytas modelio matematines lygtis užrašysime su koeficientais, kurie turės skaitines reikšmes.

## Parametrų įverčių nustatymas mažiausių kvadratų metodu

Pats populiariausias ir geriausiai ištyrinėtas koeficientų skaičiavimo būdas – tai mažiausių kvadratų metodas (MKM). Prieš pradedant vertinti regresijos parametrų įverčius, priklausomo ir nepriklausomų kintamųjų duomenys turi būti atsakingai paruošti, t.y., suderinti laiko, vietos ir periodiškumo atžvilgiu.

**MKM** tikslas - nustatyti tokius regresijos parametrų įverčius, kurie minimizuoja skirtumų tarp faktinių (YI)ir apskaičiuotų ( ) pagal pasirinktą regresijos lygtį priklausomojo kintamojo reikšmių kvadratų sumą. Matematiškai mažiausių kvadratų metodas užrašomas taip:



kur Yi -faktinės priklausomojo kintamojo reikšmės, o  - apskaičiuotos priklausomojo kintamojo reikšmės. MKM Įverčių skaičiavimo formules pagrindimas iliustruotas porinės regresijos pavyzdžiu Tiesinės porinės regresijos atveju  reikšmės bus lygios: 



Apskaičiuojamos funkcijos Σ(yi – (b0 +b1x i))2 pirmosios dalinės išvestinės ir prilyginamos nuliui:

Toliau sprendžiama lygčių sistema:





Toliau reikia sudaryti lygčių sistemą iš dešinėje lygybės pusėje esančių reiškinių:

 → ΣY i = n b0+b1 ΣX i

 → Σ X iY i = b0ΣX i +b1ΣX i2

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

 

kur =‾Yi - priklausomojo kintamojo faktinių reikšmių vidurkis, o‾Xi .- nepriklausomojo kintamojo reikšmių vidurkis. n- stebėjimų skaičius

**Parametrų įverčių tikslumas**

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

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

*1. Įverčių radimo metodai*

Regresinėje analizėje dažniausiai naudojami du parametrų įverčių nustatymo metodai: mažiausių kvadratų metodas (MKM) ir maksimalaus tikėtinumo metodas (MTM).. MTM yra sudėtingesnis, todėl šiame paskaitų konspekte nėra nagrinėjamas.

*2. Klasikinės regresinės analizės prielaidos*

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

\_\_\_ lentelė Klasikinės regresijos prielaidos

|  |  |
| --- | --- |
| **Prielaida** | **Prielaidos simbolinė išraiška** |
| 1. Regresijos funkcija koeficientų ir paklaidų atžvilgiu yra tiesinė (tiesiškumas) | yi =+β0+β1x1i+…+βnxni+εi |
| 1. Paklaidų vidurkis lygus nuliui (nulinis vidurkis) | E(εi) = 0 |
| 1. Paklaidos neautokoreliuoja (likučių ne autokoreliacijos) , t.y, paklaidos tarpusavyje nėra susijusios ir nestebimi sklaidos dėsningumai. | Cov(εi εj) = 0, ∀i,j / i≠j |
| 1. Paklaidų dispersija yra pastovi (ne heteroskedastiškumas)   Didėjant nepriklausomų kintamųjų reikšmėms, priklausomojo kintamojo sklaidos intervalas išlieka pastovus. | σ2(εi) = konstanta |
| 1. Nepriklausomi kintamieji nėra tiesiškai tarpusavyje susiję, t.y. nėra tiesinės vieni kitų tiesinės kombinacijos (ne multikolinearumas, neinterkoreliacija ) | xi ≠ γ+θjxj, ∀i,j / i≠j |
| 1. Paklaidos pasiskirsčiusios pagal normalųjį skirstinį (normalumas). | εi ~ N (0, σ2) |

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.

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

## 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. pav. apačioje) - kuo didesnę stebėjimo nuokrypio nuo vidurkio dalį paaiškina regresija, tuo regresinis ryšys yra geriau determinuotas.

Regresija paaiškinta ir nepaaiškinta stebėjimo dalis



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



Bendras Regresija Nepaaiškinta

nuokrypis paaiškinta dalis dalis

kur,– i-tojo stebėjimo apskaičiuota pagal regresijos lygtį reikšmė ‾y stebėjimų vidurkis,

yi –faktinė priklausomojo kintamojo i-ojo stebėjimo reikšmė.

Pakėlus kvadratu visus tris viršuje esančios lygybės narius bei susumavus visų stebėjimų reikšmes gauname tokią lygybę:



Bendrieji svyravimai regresija paaiškinta nepaaiškinta dalis

(TSS) dalis (ESS) (RSS)

Regresijos ryšio determinuotumas skaičiuojamas, lyginant regresija paaiškintos dalies ir bendrų svyravimų santykiu:



kur Σ( yi -‾y)2  faktinių yi reikšmių nuokrypių nuo vidurkio kvadratų suma, Σ( -‾y)2  pagal regresijos lygtį apskaičiuotų reikšmių nuokrypių nuo vidurkio kvadratų suma.

Šis rodiklis yra vadinamas *determinacijos koeficientu, jis parodo, kurią dalį priklausomojo kintamojo svyravimų apie vidurkį , galima paaiškinti į regresiją įtrauktų kintamųjų svyravimais.*

Kai regresija paaiškina visą faktinių priklausomojo kintamojo reikšmių išsibarstymą apie vidurkį, determinacijos koeficientas įgyja vieneto reikšmę (R2=1). Kuo mažesnę stebėjimų nuokrypių nuo vidurkio dalį regresinis ryšys paaiškina, tuo determinacijos koeficiento reikšmė artimesnė nuliui (visiško nepaaiškinimo atveju R2=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 R2=0,75, rodo, kad 75 proc. priklausomojo kintamojo pokyčių sąlygoja nepriklausomų veiksnių kitimas.

*Tiesinės regresijos atveju* determinacijos koeficiento kvadratinė šaknis lygi koreliacijos koeficientui (√R2 = r). Netiesinio ryšio atveju ši lygybė negalioja.

Determinacijos koeficientas dažnai naudojamas, norint parinkti tinkamiausią regresijos lygtį. Tačiau jis turi vieną trūkumą - daugėjant regresijoje nepriklausomų 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 žymimas ir skaičiuojamas pagal formulę*: (√R2 = r). Netiesinio ryšio atveju ši lygybė negalioja.

Determinacijos koeficientas dažnai naudojamas, norint parinkti tinkamiausią regresijos lygtį. Tačiau jis turi vieną trūkumą - daugėjant regresijoje nepriklausomų 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 žymimas ir skaičiuojamas pagal formulę*:



kur n – stebėjimų skaičius, k – regresijos nepriklausomų kintamųjų skaičius. Todėl koreguotas determinacijos koeficientas yra taikomas patikrinti ar papildomai įtraukus veiksnį į regresinį modelio jo determinuotumas padidėjo, t.y., ar naujos lygties yra didesnis už pradinės lygties.

*Excel skaičiuoklė visus tris determinacijos koeficientus pateikia kartu su kitais regresijos skaičiavimais. Šie rodikliai yra pateikti* ***Regression*** *skaičiavimo išklotinės lentelėje* ***Regression statistics.*** *Dauginės koreliacijos koeficientas***Multiple R**, *determinacijos koeficientas* **R Square, o** *koreguotas determinacijos koeficientas* **Adjusted R.**

Pavyzdyje apie duonos kainų priklausomybę yra tokie determinacijos rodikliai

|  |  |  |
| --- | --- | --- |
|  | TS- Tiesinis modelis | LN-ogaritmins modelis |
| *Dauginės koreliacijos koeficientas***Multiple R**, | 0,98 | 0,98 |
| *determinacijos koeficientas* **R Square** | 0,97 | 0,97 |
| *koreguotas determinacijos koeficientas* **Adjusted R.** | 0,96 | 0,96 |

Abiejų modelių determinuotumo rodikliai yra aukšti. t.y.,R2 =0,97 reikšmė rodo, kad 97 proc. visų duonos kainų svyravimų apie vidutinę nagrinėjamo periodo reikšmę galima paaiškinti įtrauktų veiksnių: rugių ir dyzelino kainų bei darbo užmokesio kitimu. Dauginės koreliacijos koeficientas yra labai artimas vienetui, o tai patvirtina išvadą apie stiprų sąryšį. Koreguotą determinacijos koeficientą galima buvo panaudoti atmetant regresijoje nereikšmingus veiksnius.

Pateiktame kainos priklausomybės pavyzdyje yra akivaizdu, kad TS- tiesinis ir LS logaritminis modeliai yra gerai determinuoti. Tačiau praktikoje dažnai apskaičiavus modelį, pasitaiko determinacijos koeficiento reikšmės 0,5 arba 0,33. Kyla klausimas, kokią išvadą turėtų daryti tyrėjas apie tokį modelio determinuotumą. Išvados gali būti skirtingos Jeigu modelyje yra įtraukta mažai stebėjimų, tuomet tokio determinuotumo, žinoma, nepakanka, o jeigu stebėjimų daug, tuomet tokio determinuotumo gali ir pakakti. Konkretų atsakymą į šį klausimą galima gauti atlikus regresijos bendrojo statistinio reikšmingumo testą,

## Regresijos lygties statistinio reikšmingumo įvertinimas

Taikant regresinę analizę neužtenka 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 pusmečio, t.y., 6 stebėjimų duonos kainos priklausomybę nuo išteklių kainų regresijos vertinimai nėra patikimi. Norint žinoti, ar galima pasikliauti apskaičiuota regresiniu modeliu, yra atliekama regresijos statistinio reikšmingumo tikrinimo procedūra, naudojant Fišerio testą. Tuo tikslu skaičiuojama F statistika,



kur α- pasirinktas reikšmingumo lygmuo, k ir n-k-1 yra atitinkami laisvės laipsnių skaičiai F-statistikos skaitiklyje ir vardiklyje.

Jei pagal regresiją apskaičiuota F statistika yra didesnė už pasirinkto reikšmingumo

lygmens teorinę Fk, ,n-k-1 skirstinio reikšmę, tai apskaičiuota regresija yra statistiškai reikšminga.

Hipotezės tikrinimo procedūr tradiciškai susideda iš keturių žingsnių:

*1. žingsnis*. Iškeliam hipotezes:

H0: β1 =β2 =… =βk = 0, (visi parametrai prie nepriklausomų kintamųjų yra lygūs 0 t.y., regresija yra nereikšminga, nes nė vienas veiksnys neįtakoja priklausomojo kintamojo)

HA: bent vienas iš parametrų βj nėra lygus 0 (regresija statistiškai reikšminga, nes yra bent vienas veiksnys, kuris įtakoja priklausomą kintamąjį)

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

*3 žingsnis* Apskaičiuotą faktinę F reikšmę lyginame su pasirinkto reikšmingumo, pvz., 5 proc. (α=0,05), teorine Fk,n-k-1 reikšme iš F-skirstinio lentelių (žr. priedus 8)

*4 žingsnis Išvada*. Jeigu Fapskaičiuota > Fk,n-k-1 , tuomet su 95% pasikliovimo lygmeniu atmetame nulinę hipotezę, jog regresija yra statistiškai nereikšminga ir priimame alternatyvią, kad bent vienas nepriklausomas kintamasis įtakoją nagrinėjamą kintamąjį. Jeigu yra priešingai ,t.y., Fapskaičiuota < Fk,n-k-1 , tuomet negalime atmesti H0 hipotezės, kad kintamųjų priklausomybė yra statistiškai nereikšminga.

Visa reikiamus regresijos statistinio reikšmingumo tikrinimui reikalingus rodiklius galima matyti Excel –Regression skaičiavimų išklotinės (priedai 4,5) antroje lentelėje ANOVA

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ANOVA |  |  |  |  |  |
|  | *Df-laisvės laipsniai* | *SS-kvadratinių nuokrypių sumos* | *MS (Stulpelis SS/df-laisvės laipsnių)* | *F – apskaičiuota* | *Reikšmingumo lygmuo F* |
| Regression (E) | 3 | 2,409 | 0,803 | 225,772 | 0,000 |
| Residual ( R) | 32 | 0,114 | 0,004 |  |  |
| Total (T) | 35 | 2,523 |  |  |  |

Determinacijos koeficientas yra apskaičiuojamas iš SS stulpelio duomenų

R2=ESS/TSS2,409/2,523=0,955, o FapskaičiuotaEMS/RMS=0.804/0.004225.772

Regresijos statistinio reikšmingumo hipotezei patikrinti atliekami tokie žigsniai:

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

H0: visi parametrai βj =0, prie kintamųjų (mūsų pavyzdyje turime 3 nepriklausomus kintamuosius xrugių kaina, xdyzelino kaina ir xdarbo užm ) yra statisiškai nereikšmingi

HA: bent vienas iš parametrų βj, nėra lygus 0 (regresija statistiškai reikšminga)

*2 žingsnis* Apskaičiuojame pagal formulę statistikos Fapskaičiuota reikšmę, ir laisvės laipsnius :k=3 ir n-k-1=36-3-1=32 laisvės laipsniai:



*3 žingsnis* Fapskaičiuota reikšmę lyginame su 5 proc. (α=0,05) reikmingumo teorine

F3,32 =2,92 reikšme iš F-skirstinio lentelių (žr. priedus8). Matome, kad

Fapskaičiuota =225,772> F3,32= 2,92

*4 žingsnis Išvada*. Su 95% pasikliovimo lygmeniu atmetame nulinę hipotezę ir priimame alternatyvią hipotezę, kad bent vienas veiksnys reikšmingai įtakoja duonos kainą, t.y., modelis yra statistiškai reikšmingas

**Septintas žingsnis: Veiksnių statistinio reikšmingumo analizė**

Šeštame žingsnyje suskaičiuoti regresinio modelio koeficientai vadinami taškiniais parametrų įverčiais. Jeigu apskaičiuoti įverčiai yra patikimi ir tenkina regresijos klasikines prielaidas, t.y., jie turi būtinas įverčių savybes –yra tiesiniai, nepaslinkti, suderinti ir efektyvūs. Modelio sudarytojas prieš taikydamas šiuos koeficientus verslo situacijos analizei **BŪTINAI** turi įsitikinti įverčių patikimumu, nes Excel skaičiuoklė visuomet pateiks įverčių skaitines reikšmes. Įverčių patikimumo tikrinimas apima tokias procedūras:1) įverčių statistinio reikšmingumo patikrinimas, apskaičiuojant koeficientų paklaidas, sudarant pasikliautinus intervalus ir patikrinant veiksnių statistinio reikšmingumo hipotezes. 2) Patikrinti veiksnių koreliuotumo galimybę, t.y., (penktoji klasikinė regresijos prielaida).

**Taškiniai ir intervaliniai įverčiai. Standartinės paklaidos**

*Taškinis įvertis* yra toks regresijos parametro įvertis, kuris įgyja konkrečią skaitinę reikšmę. Taškiniai įverčiai yra prieš nepriklausomus kintamuosius (x) esančios parametrų įverčių reikšmės (b0,b1,...,bk)

Pvz. tyrėjas surinkęs n stebėjimų duomenų eilutę ketina sudaryti k veiksnių įtakos regresinį modelį

yi = β0+β1x1i+...+βkxki+εi,

Šios imties pagrindu apskaičiavo įverčius b0, b1 ir bk. ir įrašė į regresijos lygtį jų reikšmes:

yt = b0+b1x1i+...+bkxki+ei

Taškinis įvertis yra traktuojamas kaip “geriausias” parametrą atspindintis rodiklis ir naudojamas konkrečiuose ekonominiuose skaičiavimuose. Tačiau tikėtina, jog įverčio ir tikroji parametro reikšmė gali būti skirtingos. Su tam tikra tikimybe tikroji parametro reikšmė gali būti lygi nuliui, 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 turėti daugiau informacijos apie tikrąsias parametrų reikšmes turime nagrinėti intervalinius parametrų įverčius. Intervaliniai įverčiai būtini įvairių hipotezių apie tikrąsias parametrų reikšmes patvirtinimui arba atmetimui.

*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 - pridedant koreguojantį dydį. Intervalinį įvertį galima užršyti taip:

bj ± koreguojantis dydis

Intervalinio parametro įverčių koreguojantis dydis priklauso nuo dviejų dalykų: pirma, apskaičiuoto taškinio įverčio tikslumo ir, antra, tikimybės su kuria tyrėjas nori sudaryti parametrų pasiklautinus intervalus. Įverčio tikslumą apibūdina standartinė įverčio paklaida, o tikimybes įtraukiai remdamiesi tikimybių tankio funkcija.

**Taškinio įverčio standartinė paklaida**

Taškinio įverčio tikslumui matuoti yra skaičiuojama jo standartinė paklaida SEbj .

Dauginės regresijos atveju standartinė laisvojo nario paklaida skaičiuojama pagal formule:



kur i – stebėjimo numeris, yi -faktinėi-ojo stebėjimo priklausomojo kintamojo reikšmė, - teorinė t.y. apskaičiuota pagal regresijos lygtį, i- ojo stebėjimo priklausomojo kintamojo reikšmė, xI i-oji nepriklausomojo kintamojo reikšmė, ‾x –nepriklausomojo kintamojo vidurkis.

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 apskaičiuojami du modelio koeficientai, o dauginės regresijos atveju k+1

**1.**

Įverčio b1 standartinė paklaida apskaičiuojama pagal formules:



Taigi parametro βj intervalinis įvertis užrašomas taip:

βj∈ [bj ± tn-k-1,α/2 SEbj]., arba P[bi - tn-k-1,α/2 SEbi ≤ βj ≤ bi + tn-k-1,α/2 SEbi] =1-α.,

kur apatinis βj įverčio rėžis yra lygus: bj - tn-k-1,α/2 SEbj., o viršutinis bI rėžis: bi + tn-k-1,α/2 SEbi.,

SEβI - standartinė i- įverčio paklaida, apskaičiuota pagal viršuje patektas formules

tn-k-1,α/2 - teorinė Student’o skirstinio statistika su n-k-1 laisvės laipsnių skaičiumi ir α reikšmingumo lygmeniu, Regresinėje analizėje laisvės laipsniai – tai stebėjimų skaičiaus ir įtrauktų į modelį koeficientų skaičiaus skirtumas. 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 atvejų k padidėja vienu vienetu, žyminčiu laisvąjį narį. Kuo didesnis laisvės laipsnių skaičius, tuo tikslesnis gaunamas įvertis.

Skaitinė tn-k,α/2 reikšmė randama iš Studento statistinių skirstinių lentelių. (žr. priedą 7). Reikšmingumo lygmenį α pasirenka tyrinėtojas. Dažniausiai sutinkamas α= 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 yra stebėjimų, skaičius, k+1 – vertinamų parametrų skaičius,

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

Iš intervalinio iverčio formulės 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.

Excel skaičiuoklės išklotinėje (žr. priedus 2) yra pateikiamos įverčių standartinių paklaidų bei intervalinių įverčių apatinio ir viršutinio rėžių reikšmės .

TS –tiesinio modelio įverčių skaičiavimo rezultatai (Excel išklotinė)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | *Koeficientai* | *Standartinė paklaida* | *t Stat* | *P-reišmė* | *Apatinis rėžis 95%* | *Viršutinis rėžis 95%* |
| Laisvasis narys | 0,418 | 0,802 | 0,521 | 0,606 | -1,222 | 2,059 |
| Dyzelino kaina, Lt/ltr | 0,065 | 0,092 | 0,702 | 0,488 | -0,123 | 0,252 |
| Rugiai, Lt/t | 0,001 | 0,000 | 6,265 | 0,000 | 0,0006 | 0,0012 |
| Elektros kaina, ct/kWh | 0,005 | 0,008 | 0,630 | 0,533 | -0,011 | 0,021 |
| Vid. Darbo Užmokestis, Lt/mėn. | 0,001 | 0,000 | 3,546 | 0,001 | 0,001 | 0,002 |
| Cukraus kaina, Lt/kg | 0,231 | 0,065 | 3,543 | 0,001 | 0,097 | 0,364 |
| PVM pakeitimas | 0,011 | 0,032 | 0,359 | 0,722 | -0,053 | 0,076 |

Pateiktoje lentelėje yra apskaičiuotos koeficientų reikšmės, pateiktos stulpelyje Koeficientai, jų standartinės paklaidos ir pasikliautini intervalai. Pastarieji parodo, kad 95 proc. tikimybę galima teigti, kad tikroji pz. laisvojo nario β0 parametro reikšmė yra intervale [-1,222; 2,059], o parametro prie rugių kainos , kurio įvertis brugių kaina= 0,001 pasikliautinas intervalas yra gerokai siauresnis [0,0006;0,0012]. Kuo mažesnė standartinė paklaida to paties koeficiento, tuo bus tikslesni įverčiai ir siauresni pasikliautini intervalai.

Šioje dalyje prasminga turėti griežtesnį kriterijų, kurio pagalba galima būtų tiksliau atsakyti į klausimą, kokią įtaką daro konkretus veiksnys duonos kainai. Tokiu kriterijumi gali būti

hipotezių tikrinimo procedūra.

**Hipotezės samprata**

*Hipotezė* yra iš anksto griežtai suformuluotas ir tam tikra analize tikrinamas teiginys. Regresinėje analizėje taikant statistinių išvadų metodą hipotezė gali būti *statistiškai* atmetama arba neatmetama tačiau ji niekada nėra įrodoma.

Regresinėje analizėje hipotezių tikrinimo procedūroje būtini keturi elementai:

Nulinė hipotezė Ho

Alternatyvi hipotezė Ha,

Testo statistika,

Hipotezės paneigimo taisyklė.

*Nulinė hipotezė* (H0) – tai teiginys arba prielaida, kuri statistiškai patvirtinama arba ne, remiantis nagrinėjamais stebėjimais.

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

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

Pateiktoje apačioje lentelėje pateiktos dažniausiai tikrinamos hipotezės apie regresinės lygties

yi = β0+β1x1i+...+βkxki+εI parametrus.

**Pagrindinės regresinėje analizėje tikrinamos hipotezės**

|  |  |  |
| --- | --- | --- |
| **Tikrinamas teiginys** | **Nulinė (H0) ir alternatyvi (HA) hipotezė** | **Pastabos** |
| Regresijos lygties laisvasis narys nereikšmingas (lygus nuliui) | H0: β0=0  HA: β0≠0 |  |
| Nepriklausomas kintamasis neturi įtakos priklausomam kintamajam | H0: βi=0, i=1,..k  HA: βi≠0, i=1...k |  |
| Nepriklausomojo kintamojo parametras yra lygus tam tikrai reikšmei | H0: βi=c,  HA: βi≠c, c – tikrinama parametro reikšmė | Galima pastebėti, jog ankstesnės hipotezė H0:βi=0, yra atskiras hipotezės H0: βi=c atvejis, kai c=0. |
| Įvertinta regresija yra statistiškai nereikšminga | H0: βi=0, ∀ i, i=1..k  HA: βi≠0, ∀ i, i=1..k | Tai jungtinė hipotezė, jog visi parametrai kartu paėmus nėra reikšmingi. |

**Vienpusė ir dvipusė hipotezė**

Hipotezės gali būti vienpusės arba dvipusės. *Dvipusės hipotezės* atveju alternatyvi hipotezė formuluojama su nelygybės ženklu, pvz.:

a) Kintamųjų poveikio statistinio reikšmingumo tikrinimas (pvz. rugių kaina nedaro įtakos duonos kainai): H0: βi = 0; Ha: βi ≠ 0

*Vienpusės hipotezės atveju t*ikrinamo parametro reikšmės gali būti tik didesnės arba tik mažesnės už hipoteze tikrinamą dydį. Parametro konkretaus ženklo tikrinimas yra tokios hipotezės pavyzdys

Vienpusės hipotezės pavyzdžiai:

1. Parametro ženklo tikrinimas: tiesioginis poveikis (hipotezė: didėjant rugių kainai duonos kaina taip pat didėja)

H0: βi = 0; Ha: βi > 0.

1. Arba priešingas poveikis (didėjant kintamajam xj , priklausomas kintamasis yj mažėja)

H0: βi = 0; Ha: βi < 0

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

**Hipotezių apie įverčių reikšmingumą tikrinimas**

Regresinėje analizėje svarbu nustatyti ar konkretus veiksnys daro įtaką nagrinėjamam reiškiniui ar ne. Veiksnio poveikio statistinis reikšmingumas tikrinamas tokiu būdu: iškeliama hipotezė, kad veiksnio poveikis yra nereikšmingas. Matematiškai tai užrašoma taip: parametras **βj =**0 t,y parametras prie j- veiksnio su tikimybe 1-**α** gali yra lygus 0. Paprastai šis teiginys yra įvardijamas nuline hipoteze. t.y., H0 **βj =0**

Regresinės analizės įverčių reikšmingumo tikrinimo procedūrą sudaro keturi žingsniai:

**1.žingsnis**. *Formuluojamos hipotezės:*

H0  βi = 0 (nepriklausomas veiksnys (xi) nedaro įtakos priklausomam kintamajam t.y., koeficientas prie veiksnio gali būti lygus 0)

H1 βi ≠ 0 (xi poveikis reikšmingas - regresijos koeficientas prie veiksnio nelygus 0)

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



Dydis t yra pasiskirstęs pagal Stjudento t-skirstinį su α/2 reikšmingumo lygmeniu ir n-k-1 laisvės laipsniais. t.y t~ tα/2(n-k-1)

**3 žingsnis** Apskaičiuota t statistikos reikšmė lyginama su teorine t-skirstino tα/2(n-k-1) reikšme.

**4 žingsnis**. *Daromos išvados* Jei apskaičiuotos |t| reikšmės modulis yra **didesnis** už teorinę t-skirstinio reikšmę, tuomet **nulinė hipotezė atmetama** ir priimama alternatyvi hipotezė. Su 1-α tikimybe (pvz., α= 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 tα/2(n-k-1), negalime atmesti nulinės hipotezės, o tai reiškia, kad negalime tvirtinti, kad j veiksnio poveikis yra statistiškai reikšmingas.

Hipotezės paprastai tikrinamos, taikant 90% 95% ar 99% pasikliovimo lygmenį (1-α), kas yra tas pats, kaip reikšmingumo lygmuo α: 0,1%; 0,05%; 0,001%. Reikšmingumo lygmuo α rodo toleruojamą hipotezių tikrinimo klaidos tikimybę atmesti nulinę hipotezę, kai ji yra teisinga. Pvz. kai tikrinama hipotezė H0 βelekt. kaina=0 (elektros kainos pokyčiai nedaro statistiškai reikšmingo poveikio duonos kainai) su reikšmingumo lygmeniu α= 0,05, tai rodo kad tyrėjas 95% tikimybe teisingai atmeta H0 hipotezę, kad „elektros kainos poveikis yra statistiškai nereikšmingas“,. ir toleruoja 5% tikimybę klaidingai atmesti hipotezę nors tai netiesa.

Lygiavertę išvadą apie tikrinamą hipotezę galima gauti lyginant apskaičiuotąją t reikšmę atitinkantį faktinį reikšmingumo lygmenį p su analitiko pasirinktuoju α - reikšmingumo lygmeniu. Jei apskaičiuotas reikšmingumo lygmuo α yra mažesnis nei tyrėjo pasirinktas, tai nulinė hipotezė atmetama, jei didesnis –nulinės hipotezės atmesti negalima.

**Veiksnių reikšmingumo tikrinimo pavyzdys**

Kiekvienos regresinės analizės metu būtina patikrinti parametrų įverčių reikšmingumą. Pavyzdyje apie rugių kainos priklausomybę patikrinsime, kurių veiksnių kainų pokyčiai daro poveikį duonos kainą, o kurie neturi reikšmingos įtakos. Taigi, tikrinsime modelio koeficientų statistinio reikšmingumo hipotezę.

*1. žingsnis*. Tikrinama parametrų lygybės nuliui hipotezė

H0: βj=0, (parametras prie nepriklausomojo kintamojo xj yra lygus 0, t.y., konkretaus veiksnio kainos pokyčiai nedaro statistiškai reikšmingo poveikio duonos kainai)

HA: βj≠0. (parametras prie nepriklausomojo kintamojo x nelygus 0 t.y., xj nėra lygus 0, t.y., konkretaus veiksnio kainos pokyčiai daro statistiškai reikšmingą poveikį duonos kainai

*2 žingsnis* Apskaičiuojama t-stjudento testo statistikos visiens koeficientams.: tb0 apskaičiuota=ir t.t. Šios statistikos yra pateiktos Excel išklotinėe pateiktoje lentelėje viršuje ir priede 2 lentelės stulpelyje t-stat.

*3 žingsnis* Visų koeficientų tapskaičiuota - reikšmė lyginama su tα/2,n-k-1 teorine reikšme. Iš t pasiskirstymo lentelių (žr. priedas 7 ) randame **t0.05/2;29** = 2,045 (laisvės laipsnių skaičius: n-k-1=36 -6-1=29), ir α/2=0,05/2=0.025 Palyginę apskaičiuotą t reikšmę su teorine **t0.025,29,** matome, kad tapskaičiuota =0,521 yra mažesnė už **t0.025,29** =2,045, o kintamojo rugių kaina

tapskaičiuota =6,265> **t0.025,29**=2,045

*4 žingsnis Išvada*. Su 95% tikimybe atmetama nulinė hipotezė H0: βj=0 trims kintamiesiems: rugių kainai, darbo užmokesčiui ir cukraus kainai ir priimama alternatyvi hipotezė HA: β1≠0, kuri reiškia, kad šių veiksnių kainų pokyčiai statistiškai reikšmingai veikia duonos kainą.

Kitų gi kintamųjų: dyzelino, elektros kainų kitimas ir PVM mokesčio tarifo pakeitimo koeficientams atmesti H0 βj=0 hipotezės, kad jų poveikis yra statistiškai nereikšmingas, negalime, nes tapskaičiuotos statistikos modulis yra mažesnis už **t0.05/2;29** = 2,045.

Analogiškai tikriname ir laisvojo nario statistinį reikšmingumą, kurio t\_statistika yra mažesnė už **t0.05/2;29** = 2,045. Vadinasi ir laisvasis narys nėra statistiškai reikšmingas.

Kai yra daroma 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 parametrų įverčiai ir standartinės paklaidos keičiasi, todėl turi būti iš naujo įvertinti parametrai ir patikrintas jų reikšmingumas.

Atlikus statistinio veiksnių reikšmingumo testus ir suradus, kad modelyje yra nereikšmingų kintamųjų, reikia išsiaiškinti nereikšmingumo priežastis ir koreguoti modelį. Kintamieji gali būti statistiškai nereikšmingi dėl šių priežasčių:

1. Veiksniai iš tiesų nedaro įtakos nagrinėjamas reiškiniui. Jeigu tai yra tiesa, tuomet ši statistinė išvada yra dėsninga, kuri suteikia mums svarbios informacijos, kad parinktas veiksnys reikšmingos įtakos nagrinėjamam reiškiniui neturi.
2. Tyrimui surinkta per mažai duomenų ir dėl laisvės laipsnių trūkumo, gaunamos didelės įverčių paklaidos. Šiuo atveju reikėtų papildyti duomenų eilutes naujais stebėjimais. Vertėtų prisiminti ekonometrijoje taikomą „Nykščio taisyklę“ kuri sako, kad, norint išvengti kintamųjų nereikšmingumo dėl per mažo laisvės laipsnių skaičiaus, į modelį įtrauktų stebėjimų skaičius bent 6 kartus turi būti didesnis už nepriklausomų kintamųjų skaičių.
3. Tam tikri veiksniai netenkina interkoreliacijos klasikinės prielaidos, todėl gavome modelį su nereikšmingais kintamaisiais. Kaip patikrinti šią prielaidą, bus paaiškinta žemiau.
4. Neteisingai parinkta modelio matematinė forma. Šiuo atveju galima pabandyti kintamąjį įtraukti į modelį kita forma, pvz. logaritmine arba kvadratine.

Suskaičiavus regresinį modelį ir pastebėjus, kad yra statistiškai nereikšmingų veiksnių, vertėtų patikrinti penktąją klasikinę regresijos prielaidą, apie nepriklausomų kintamųjų multikolinearumą.

***Multikolinearumo problema***

Multikolinearumas regresiniame modelyje atsiranda tuomet, kai yra stipri tiesinė priklausomybė tarp nepriklausomų kintamųjų, t.y. xi galima išreikšti kaip tiesinę kombinacijąlikusių nepriklausomų kintamųjų.Du terminai: multikolinearumas ir interkoreliacija dažnai vartojami, kaip sinonimai. Vis gi interkoreliacijos terminas daugiau tinka, kai kalbama apie dviejų nepriklausomų kintamųjų, o multikolinearumas – daugiau negu dviejų nepriklausomų kintamųjų tarpusavio sąryšį. Esant multikolinearumui iškyla tam tikros problemos:

1. apskaičiuoti įverčiai (koeficientai) yra labai nestabilūs. Tai reiškia, kad jų reikšmės ir gali ženklai pasikeisti, jei įtrauksime papildomą veiksnį arba kelis naujus stebėjimus.
2. Tikrinant hipotezes įverčių t statistikos taip pat labai nestabilios, iš reikšmingų gali tapti nereikšmingomis, jei bus įtrauktas papildomas veiksnys.
3. Modelio skaičiavimai rodo, kad jis yra gerai determinuotas, bet nėra reikšmingų nepriklausomų kintamųjų.

Multikolinearumą nustatyti galima skaičiuojant *porinių koreliacijų matricą* tarp visų modelio kintamųjų. *Porinių koreliacijų matrica yra sudaryta iš koreliacijos koeficientų rij, kurių reikšmės kinta nuo -1 iki +1. Nulinė koeficiento reikšmė rodo, kad tiesinio ryšio tarp dviejų kintamųjų nėra*, o reikšmės artimos +1 arba -1 rodo stiprų tiesioginį arba atvirkštinį ryšį. Excel pagalba surandama koreliacijos koeficientų matrica, kurioje pirmasis stulpelis rodo koreliacijos koeficientus rij tarp y ir visų xj. Likę stulpeliai rodo koreliacijos koeficientus tarp xi ir xj. Vertinant tuos koeficientus (t.y. matrica be pirmojo stulpelio), galima vadovautis Nykščio taisykle. Ji teigia: jei porinės koreliacijos koeficientas moduliu yra didesnis už 0,8, tuomet regresinis modelis pasižymi multikolinearumu tarp nepriklausomų xi ir xj veiksnių.

Multikolinearumo problemą galima išspręsti taikant šiuos būdus:

* Vieno ar kelių stipriai koreliuojančių veiksnių pašalinimas. Tyrėjas pirmiausia pats nusprendžia, kurio kintamojo atsisakyti. Statistiškai tikslingiau būtų atsisakyti to veiksnio, kurio koreliacijos koeficientas su priklausomu kintamuoju (y), yra mažesnis.
* Papildomų stebėjimų įtraukimas. Tikėtina, kad didesnis stebėjimų skaičius sumažintų įtakojančių veiksnių tarpusavio sąryšį.
* Duomenų koregavimas. Dažniausiai naudojamas koregavimo būdas yra logaritmuoti kintamąjį (kelių koreliuojančių kintamųjų logaritmuoti nereikėtų, nes išliktų tarp jų tokia pati koereliacija). Jei duomenys yra laiko eilutės (skerspjūvio duomenims tai netinka ), galima kintamojo pradinių reikšmių paimti skirtumus: . Taip pat galima vieną iš koreliuotų veiksnių apibūdinti kitaip, pavyzdžiui, nedarbo lygį pakeisti į užimtumo lygį. Galimi ir kitokie duomenų koregavimo būdai.

Septintame žingsnyje patikrinę veiksnių statistinio reikšmingumo hipotezes radome, duonos kainai statistiškai reikšmingą įtaką daro rugių kaina, darbo užmokestis ir cukraus kaina, o dyzelino, elektros kainų kitimas, PVM mokesčio tarifo pakeitimas ir laisvasis narys yra statistikškai nereikšmingi veiksniai. Gavus tokius rezultatus verta patikrinti ar veiksnių nereikšmingumas nėra atsiradęs dėl multikolinearumo. Šiuo tikslu suskaičiuojame porinių koreliacijų matricą Excel skaičiuoklės pagalba

Apačioje pateikiama porinių koreliacijų koeficientų lentelė



*Koreliacijos matricos apskaičiavimui reikia atverti* ***Data Analysis*** *langą, kuriame matysite duomenų analizės priemonių sąrašą. Pasirinkite* ***Correlation*** *ir paspauskite mygtuką OK, Ekrane pasirodys lentelė* ***Correlation,*** *kurioje langeliuose:* ***Input Y range*** *reikia pažymėti atitinkamų duomenų pirmojo ir paskutinio stebėjimų langelių koordinates arba, atsistojus įrašo langelyje, su pele apibrėžti Y ir X kintamųjų duomenų lentelę. Pažymint duomenų lentelę, galima įtraukti ir kintamųjų pavadinimų langelius. Tuomet reikėtų varnelę uždėti langelyje ties* ***Labels in first row****, G****rouped b****y srityje reikia pažymėti* ***Columns****. Srityje* ***Output*** *reikia nurodyti vietą, kurioje bus patalpinti skaičiavimo rezultatai. Patogiausia -talpinti tos pačios bylos naujame puslapyje* ***New Worksheet Ply*** *ir nurodyti pavadinimą, pvz Koreliacija*

Pagrindinėje lentelės įstrižainėje matome koreliacijos koeficientus lygius vienetui, t.y., visišką duomenų sutapimą, ir tai yra dėsninga, nes tai ryšis tarp to paties kintamojo duomenų. Paryškintame lentelės stulpelyje yra pateikiami porinės koreliacijos koeficientai tarp Y ir visų X, t.y., duonos kainos ir įtrauktų X veiksnių. DU koeficientai šiame stulpelyje yra artimi 1 ir rodo stiprų sąryšį tarp duonos kainos ir dyzelino bei rugių kainos, kiti šio stulpelio rodo vidutinį ryšio stiprumą. Pirmame stulpelyje esantys koeficientai nesukelia multikolinearumo problemos. Blogai yra tuomet, kai likusioje, be pažymėto stulpelio, matricos dalyje yra didesnių negu |0,8| koreliacijos koeficientų reikšmių. Tokios koeficientų reikšmės yra tarp dviejų porų kintamųjų: dyzelino bei rugių kainų susikirtime(0,89) ir cukraus bei dizelino (0,87) Taigi dyzelino, rugių ir cukraus kainos turi labai panašias kitimo tendencijas, todėl skaičiuojant koeficientus MKM metodu jų reikšmės gali būti paslinktos t.y., nutolusios nuo tikrųjų parametro reikšmių. Rugių ir dizelino kainų pokyčiai yra svarbūs tyrime todėl abu kintamuosius tikslinga išsaugoti modelyje. Problemą išspręsime pakeisdami dyzelino duomenis į dirbtinį kintamąjį, kuris įgauna tik tris reikšmes 3,5; 4,0; 4,5 Lt/ltr. t.y., dizelino kainoms iki 3,5 Lt, suteikiama reikšmė 3,5 Lt.; kainoms, didesnėms už 3,5 bet mažesnėms už 4,0 Lt priskiriama 4,0 Lt. reikšmė, o kainoms didesnėms už 4 Lt. bet mažesnėms už 4,5 Lt priskiriama kaina 4,5 Lt. Padarius tokį pakeitimą, koreliaciją tarp dyzelino ir rugių kainų sumažėjo iki 0,72. Todėl tolesniuose skaičiavimuose dyzelino kaina bus įtraukta dirbtiniu pavidalu. Tačiau toks dyzelino kainos pakeitimas neišsprendė interkoreliacijos problemos tarp cukraus ir dyzelino kainų. Kadangi cukraus duonos sudėtyje yra labai mažai, o transportavimo kaštai gana žymus, tai atsiradusią stiprią koreliaciją tarp šių veiksnių galima eliminuoti, atsisakant vieno iš veiksnių. Šiuo atveju –cukraus kainos. Todėl tolesnėje tyrimo eigoje cukraus kintamojo į modelį nebeįtrauksime. Atlikus šiuos pakeitimus galima teigti, kad sudarytas modelis tenkina penktąją klasikinę regresinio modelio prielaidą apie multikolinearumo nebuvimą.

**Statišktiškai nereikšmingų veiksnių atsisakymas**

Išsprendus multikolinearumo problemą, suskaičiuojame naują regresijos lygtį su pakeistais dyzelino duomenimis ir be kintamojo cukraus kaina. Žemiau pateikiama naujos regresijos duomenų lentelė

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | *Koefiientai* | *Standarinės paklaidos* | *t Stat* | *P-reikšmė* | *Viršutinis 95%* | *Apatinisr 95%* |
| Intercept | 0,483 | 0,896 | 0,539 | 0,594 | -1,346 | 2,312 |
| Rugiai, Lt/t | 0,001 | 0,000 | 8,120 | 0,000 | 0,001 | 0,001 |
| Vid. Darbo Užmokestis, Lt/mėn. | 0,001 | 0,000 | 3,055 | 0,005 | 0,000 | 0,002 |
| PVM pakeitimas | 0,005 | 0,035 | 0,142 | 0,888 | -0,066 | 0,076 |
| Elektros kaina, ct/kWh | 0,003 | 0,007 | 0,527 | 0,602 | -0,010 | 0,017 |
| Dyzelino kainos dirbt dydis | 0,269 | 0,044 | 6,118 | 0,000 | 0,179 | 0,358 |

Pakeistame modelyje vis dar lieka du statistiškai nereikšmingi veiksniai: PVM tarifo pakeitimas ir elektros kaina tstat=0,527 bei laisvasis narys b0. tstat=0,539 Labai realu, kad pirmieji du kintamieji labai silpnai veika 1 kg. duonos kainą. Todėl tikėtina, kad šis statistinis nereikšminumas atspindi faktą, kad nei PVM pakeitimas nei elektros kainų didėjimas duonos kainai reikšmingos įtakos neturėjo. Įprasta, kad regresiniame modelyje neturėtų būti statistiškai nereikšmingų veiksnių, todėl xelektros tarif ir DPVM kintamuosius reikėtų pašalinti iš modelio. Kai yra daroma išvada, kad priklausomas kintamasis arba laisvasis narys yra statistikai nereikšmingi, tuomet iš naujo reikia skaičiuoti regresijos lygtį be nereikšmingo nepriklausomo kintamojo arba laisvojo nario. Įsidėmėtina, jog išmetus iš regresijos lygties nereikšmingą veiksnį, likusių nepriklausomų kintamųjų parametrų įverčiai ir standartinės paklaidos keičiasi, todėl turi būti iš naujo įvertinti parametrai ir patikrintas jų reikšmingumas.

Paprastai nereikšmingi kintamieji yra šalinami po vieną, pradedant nuo to veiksnio, kurio tapskaičiuot statistika moduliu yra mažiausia. Mūsų pavyzdyje fiktyvaus kintamojo PVM mokesčio pakeitimų statistika yra mažiausia. Todėl apskaičiuojame dar vieną modelį be kintamojo xPVM Atlikti skaičiavimai parodė, kad išmetus veiksnį xPVM ir perskaičiavus regresiją, elektros kainos reikšmingumas nepasikeitė ir naujame modelyje, todėl atsisakome ir kintamojo elektros kainos ir liekame prie modelio su trimis kintamaisiais: rugių kaina, dyzelino kaina ir darbo užmokesčio kitimas. Pašalinus du veiksnius, modelyje liko nereikšmingas laisvasis narys b0. Galima suskaičiuoti modelį ir be laisvojo nario. Norint tai padaryti Excel skaičiuoklėje, išsikvietus **Data Analysis \_ Regression**, reikėtų uždėti varnele ties **Constant is Zero.**  Visgi, nereikėtų skubėti atsisakyti laisvojo nario, nes jo buvimas modelyje užtikrina antrosios klasikinės regresijos prielaidos tenkinimą, kuri reikalauja, kad modelio paklaidų vidurkis turi būti lygus nuliui. Jeigu modelyje yra laisvasis narys, tuomet modelio paklaidų vidurkis visuomet bus lygus nuliui.

Aprašytus veiksmus veiksmus galima atlikti su logaritmuotais duomenimis

Apibendrinant septintojo žingsnio rezultatus, galima užrašyti galutines regresijos lygtis su reikšmingais kintamaisiais:

|  |  |
| --- | --- |
| TS | Yduonos kaina =0,921+ 0,001Xrugių kaina+0,001Xdarb užm +0,26XDyzelinas+ e |
| LN | ln(Yduon kaia )= -2,971β0 ++0,081(Xrugių kaina)+0,496 ln(Xdarbu užm)+,0219ln(Xdyz\_kaina)+e |

Duonos kainų priklausomybės tyrime turime du modelius: TS ir LN ir kiekviename iš jų po šešis koeficientus, kuriuos reikia apskaičiuoti taikant MKM. Šiam tikslui labai patogu naudotis Microsoft Excel skaičiuokle duomenų analizės Data Analysis moduliu. Prieš kviesdami **Data Analysis** turime sutvarkyti duomenų lentelę, kurioje pirmame stulpelyje būtų surašyti priklausomojo kintamojo y stebėjimai. Greta lentelėje stulpeliais, nepaliekant tuščių langelių, stulpelių ir eilučių turi būti surašyti įtakojančių veiksnių x stebėjimų duomenys. *(priedai 1)*

*Jeigu Jūsų kompiuteryje nėra* ***Data Analysis*** *modulio, jį reikia aktyvinti. Tai atliekame tokiu būdu. Windows XP ar kitoje operacinėje sistemoje paleidžiame veikti Excel programą, paspaudę pagrindinio meniu mygtuką* ***File*** *pasirenkame Excel pasirinktys ( Options). Šiame lange pasirenkame* ***Priedai (Add-Isn) ir*** *aktyviname pažymėdami* ***Analizės įrankų paketą ir Analizės įrankių paketą VBA.*** *Lango apačioje paspaudžiame mygtuką* ***Vykdyti*** *(Go). Pasirodo lentelė kurioje reikia pažymėti varneles ties* ***Analizės įrankų paketas ir Analizės įrankių paketas VBA*** *ir paspausti mygtuką* ***OK.*** *Po šių veiksmų pagrindinio meniu juostoje esančioje grupėje* ***Duomenys (DATA)*** *dešiniame viršutiniame kampe atsiras modulis* ***Data Analysis,*** *kuris ir išliks Jūsų kompiuteryje.*

*Regresinio modelio apskaičiavimui reikia atverti* ***Data Analysis*** *langą, kuriame matysite duomenų analizės priemonių sąrašą. Pasirinkite* ***Regression*** *ir paspauskite mygtuką OK, Ekrane pasirodys lentelė* ***Regression,*** *kurioje langeliuose:* ***Input Y range*** *ir* ***Input X range*** *reikia pažymėti atitinkamų duomenų pirmojo ir paskutinio stebėjimų langelių koordinates arba su pele apibrėžti y stulpelį ir x kintamųjų duomenų lentelę. Pažymint duomenų lentelę galima įtraukti ir kintamųjų pavadinimų langelius. Labai svarbu, kad kintamųjų x ir y stebėjimų skaičius būtų vienodas. Po to, žemiau esančiuose trijuose langeliuose programa prašo nurodyti ir pažymėti varneles* ***Labels*** *langelyje, jeigu, žymėdami duomenis, įtraukėte ir pavadinimo langelius, Langelyje* ***Constant is Zero****, reikia pažymėti varnelę jeigu norite modelį suskaičiuoti be laisvojo nario. Langelyje* ***Confidence Level*** *galima įrašyti kitokią intervalinių įverčių pasikliovimo tikimybę, pvz. 0,9. Programa tuomet skaičiuoja pasikliautinus intervalus su 0,95 ir 0,9 proc.tikimybe. Toliau* ***Regression*** *lentelės* ***Output*** *srityje reikia nurodyti vietą, kuioje bus patalpinti skaičiavimo rezultatai. Patogiausia juos patalpinti tos pačios bylos naujame puslapyje* ***New Worksheet Ply*** *ir nurodyti pavadinimą, pvz skaičiuojant tiesinį modelį – TS, o logaritminį LN .*

*Paskutinėje* ***Regression*** *lentelės skiltyje* ***Residuals*** *programa prašo pažymėti , kuriuos regresinio modelio skaičiavimų rezultaus pateikti išklotinėje. Galima pažymėti visus langelius, tačiau svarbiausia pažymėti* ***Residuals*** *ir* ***Standartized residuals,*** *t.y.,kad būtų suskaičiuotos modelio paklaidos ir standatizuotos paklaidos.*

*Atlikę visas šias operacijas gauname TS ir LN modelių koeficientų reikšmes, kurie yra pateikti skaičiavimų išklotinėse, esančiose prieduose 2 ir 3 trečiosios lentelės skiltyje* ***Coefficients***

Atlikę visas Excel komandas galima surašyti du pasirinktus regresinius modelius su skaitinėmis reikšmėmis.

|  |  |
| --- | --- |
| TS | Yduonos kaina =0,4180+0,065+Xdyz kaina+0,001Xrugių kaina+0,005Xelek kaina+0,001Xdarbu užm +0,231XCukrus+0,011DPVM+ ε |
| LN | ln(Yduon kaia )= -2,311β0 +0,022ln(Xdyz\_kaina+0,09(Xrugių kaina)+-0,03ln(Xelek kaina)+0,419 ln(Xdarbu užm )+ 0,189lnXCukrus +0,010DPVM+ ε |

**Aštuntas žingsnis. Modelio patikimumo tikrinimas Paklaidų analizė.**

## 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. pav. apačioje) - kuo didesnę stebėjimo nuokrypio nuo vidurkio dalį paaiškina regresija, tuo regresinis ryšys yra geriau determinuotas.

Regresija paaiškinta ir nepaaiškinta stebėjimo dalis



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



Bendras Regresija Nepaaiškinta

nuokrypis paaiškinta dalis dalis

kur,– i-tojo stebėjimo apskaičiuota pagal regresijos lygtį reikšmė ‾y stebėjimų vidurkis,

yi –faktinė priklausomojo kintamojo i-ojo stebėjimo reikšmė.

Pakėlus kvadratu visus tris viršuje esančios lygybės narius bei susumavus visų stebėjimų reikšmes gauname tokią lygybę:



Bendrieji svyravimai regresija paaiškinta nepaaiškinta dalis

(TSS) dalis (ESS) (RSS)

Regresijos ryšio determinuotumas skaičiuojamas, lyginant regresija paaiškintos dalies ir bendrų svyravimų santykiu:



kur Σ( yi -‾y)2  faktinių yi reikšmių nuokrypių nuo vidurkio kvadratų suma, Σ( -‾y)2  pagal regresijos lygtį apskaičiuotų reikšmių nuokrypių nuo vidurkio kvadratų suma.

Šis rodiklis yra vadinamas *determinacijos koeficientu, jis parodo, kurią dalį priklausomojo kintamojo svyravimų apie vidurkį , galima paaiškinti į regresiją įtrauktų kintamųjų svyravimais.*

Kai regresija paaiškina visą faktinių priklausomojo kintamojo reikšmių išsibarstymą apie vidurkį, determinacijos koeficientas įgyja vieneto reikšmę (R2=1). Kuo mažesnę stebėjimų nuokrypių nuo vidurkio dalį regresinis ryšys paaiškina, tuo determinacijos koeficiento reikšmė artimesnė nuliui (visiško nepaaiškinimo atveju R2=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 R2=0,75, rodo, kad 75 proc. priklausomojo kintamojo pokyčių sąlygoja nepriklausomų veiksnių kitimas.

*Tiesinės regresijos atveju* determinacijos koeficiento kvadratinė šaknis lygi koreliacijos koeficientui (√R2 = r). Netiesinio ryšio atveju ši lygybė negalioja.

Determinacijos koeficientas dažnai naudojamas, norint parinkti tinkamiausią regresijos lygtį. Tačiau jis turi vieną trūkumą - daugėjant regresijoje nepriklausomų 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 žymimas ir skaičiuojamas pagal formulę*: (√R2 = r). Netiesinio ryšio atveju ši lygybė negalioja.

Determinacijos koeficientas dažnai naudojamas, norint parinkti tinkamiausią regresijos lygtį. Tačiau jis turi vieną trūkumą - daugėjant regresijoje nepriklausomų 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 žymimas ir skaičiuojamas pagal formulę*:



kur n – stebėjimų skaičius, k – regresijos nepriklausomų kintamųjų skaičius. Todėl koreguotas determinacijos koeficientas yra taikomas patikrinti ar papildomai įtraukus veiksnį į regresinį modelio jo determinuotumas padidėjo, t.y., ar naujos lygties yra didesnis už pradinės lygties.

*Excel skaičiuoklė visus tris determinacijos koeficientus pateikia kartu su kitais regresijos skaičiavimais. Šie rodikliai yra pateikti* ***Regression*** *skaičiavimo išklotinės lentelėje* ***Regression statistics.*** *Dauginės koreliacijos koeficientas***Multiple R**, *determinacijos koeficientas* **R Square, o** *koreguotas determinacijos koeficientas* **Adjusted R.**

Pavyzdyje apie duonos kainų priklausomybę yra tokie determinacijos rodikliai

|  |  |  |
| --- | --- | --- |
|  | TS- Tiesinis modelis | LN-ogaritmins modelis |
| *Dauginės koreliacijos koeficientas***Multiple R**, | 0,98 | 0,98 |
| *determinacijos koeficientas* **R Square** | 0,97 | 0,97 |
| *koreguotas determinacijos koeficientas* **Adjusted R.** | 0,96 | 0,96 |

Abiejų modelių determinuotumo rodikliai yra aukšti. t.y.,R2 =0,97 reikšmė rodo, kad 97 proc. visų duonos kainų svyravimų apie vidutinę nagrinėjamo periodo reikšmę galima paaiškinti įtrauktų veiksnių: rugių ir dyzelino kainų bei darbo užmokesio kitimu. Dauginės koreliacijos koeficientas yra labai artimas vienetui, o tai patvirtina išvadą apie stiprų sąryšį. Koreguotą determinacijos koeficientą galima buvo panaudoti atmetant regresijoje nereikšmingus veiksnius.

Pateiktame kainos priklausomybės pavyzdyje yra akivaizdu, kad TS- tiesinis ir LS logaritminis modeliai yra gerai determinuoti. Tačiau praktikoje dažnai apskaičiavus modelį, pasitaiko determinacijos koeficiento reikšmės 0,5 arba 0,33. Kyla klausimas, kokią išvadą turėtų daryti tyrėjas apie tokį modelio determinuotumą. Išvados gali būti skirtingos Jeigu modelyje yra įtraukta mažai stebėjimų, tuomet tokio determinuotumo, žinoma, nepakanka, o jeigu stebėjimų daug, tuomet tokio determinuotumo gali ir pakakti. Konkretų atsakymą į šį klausimą galima gauti atlikus regresijos bendrojo statistinio reikšmingumo testą,

## Regresijos lygties statistinio reikšmingumo įvertinimas

Taikant regresinę analizę neužtenka 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 pusmečio, t.y., 6 stebėjimų duonos kainos priklausomybę nuo išteklių kainų regresijos vertinimai nėra patikimi. Norint žinoti, ar galima pasikliauti apskaičiuota regresiniu modeliu, yra atliekama regresijos statistinio reikšmingumo tikrinimo procedūra, naudojant Fišerio testą. Tuo tikslu skaičiuojama F statistika,



kur α- pasirinktas reikšmingumo lygmuo, k ir n-k-1 yra atitinkami laisvės laipsnių skaičiai F-statistikos skaitiklyje ir vardiklyje.

Jei pagal regresiją apskaičiuota F statistika yra didesnė už pasirinkto reikšmingumo

lygmens teorinę Fk, ,n-k-1 skirstinio reikšmę, tai apskaičiuota regresija yra statistiškai reikšminga.

Hipotezės tikrinimo procedūr tradiciškai susideda iš keturių žingsnių:

*1. žingsnis*. Iškeliam hipotezes:

H0: β1 =β2 =… =βk = 0, (visi parametrai prie nepriklausomų kintamųjų yra lygūs 0 t.y., regresija yra nereikšminga, nes nė vienas veiksnys neįtakoja priklausomojo kintamojo)

HA: bent vienas iš parametrų βj nėra lygus 0 (regresija statistiškai reikšminga, nes yra bent vienas veiksnys, kuris įtakoja priklausomą kintamąjį)

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

*3 žingsnis* Apskaičiuotą faktinę F reikšmę lyginame su pasirinkto reikšmingumo, pvz., 5 proc. (α=0,05), teorine Fk,n-k-1 reikšme iš F-skirstinio lentelių (žr. priedus 8)

*4 žingsnis Išvada*. Jeigu Fapskaičiuota > Fk,n-k-1 , tuomet su 95% pasikliovimo lygmeniu atmetame nulinę hipotezę, jog regresija yra statistiškai nereikšminga ir priimame alternatyvią, kad bent vienas nepriklausomas kintamasis įtakoją nagrinėjamą kintamąjį. Jeigu yra priešingai ,t.y., Fapskaičiuota < Fk,n-k-1 , tuomet negalime atmesti H0 hipotezės, kad kintamųjų priklausomybė yra statistiškai nereikšminga.

Visa reikiamus regresijos statistinio reikšmingumo tikrinimui reikalingus rodiklius galima matyti Excel –Regression skaičiavimų išklotinės (priedai 4,5) antroje lentelėje ANOVA

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ANOVA |  |  |  |  |  |
|  | *Df-laisvės laipsniai* | *SS-kvadratinių nuokrypių sumos* | *MS (Stulpelis SS/df-laisvės laipsnių)* | *F – apskaičiuota* | *Reikšmingumo lygmuo F* |
| Regression (E) | 3 | 2,409 | 0,803 | 225,772 | 0,000 |
| Residual ( R) | 32 | 0,114 | 0,004 |  |  |
| Total (T) | 35 | 2,523 |  |  |  |

Determinacijos koeficientas yra apskaičiuojamas iš SS stulpelio duomenų

R2=ESS/TSS2,409/2,523=0,955, o FapskaičiuotaEMS/RMS=0.804/0.004225.772

Regresijos statistinio reikšmingumo hipotezei patikrinti atliekami tokie žigsniai:

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

H0: visi parametrai βj =0, prie kintamųjų (mūsų pavyzdyje turime 3 nepriklausomus kintamuosius xrugių kaina, xdyzelino kaina ir xdarbo užm ) yra statisiškai nereikšmingi

HA: bent vienas iš parametrų βj, nėra lygus 0 (regresija statistiškai reikšminga)

*2 žingsnis* Apskaičiuojame pagal formulę statistikos Fapskaičiuota reikšmę, ir laisvės laipsnius :k=3 ir n-k-1=36-3-1=32 laisvės laipsniai:



*3 žingsnis* Fapskaičiuota reikšmę lyginame su 5 proc. (α=0,05) reikmingumo teorine

F3,32 =2,92 reikšme iš F-skirstinio lentelių (žr. priedus8). Matome, kad

Fapskaičiuota =225,772> F3,32= 2,92

*4 žingsnis Išvada*. Su 95% pasikliovimo lygmeniu atmetame nulinę hipotezę ir priimame alternatyvią hipotezę, kad bent vienas veiksnys reikšmingai įtakoja duonos kainą, t.y., modelis yra statistiškai reikšmingas

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Prisimename, kad regresinį modelį sudaro regresijos lygtis – tai sisteminė modelio dalis ir paklaidos -atsitiktinė modelio dalis. t.y.,  Yi =f(X1; X2; …X1k) + ɛi  Sisteminė dalis Atsitiktinė dalis    Kaip ir buvo minėta, adekvatus ir patikimas modelis bus tuomet, kai modelyje dominuos sisteminė dalis, t.y., kai ji bus statistiškai reikšminga. Taip pat modelyje yra labai svarbu patikrinti ir modelio paklaidas. Sudarytas modelis bus adekvatus ir patikimas tuomet, kai o paklaidos bus atsitiktinis dydis, kurio kitimas nepriklauso nuo jokių dėsningumų. Kitais žodžiais tariant modelio paklaidos turi tenkinti klasikines modelio prielaidas, kurios yra nurodytos konspekto 19psl.o, šioje lentelėje pateikiamos tik tos, kurios susijusios su modelio paklaidomis.   |  |  | | --- | --- | | **Prielaida** | **Prielaidos simbolinė išraiška** | | II. Paklaidų vidurkis lygus nuliui (nulinis vidurkis) | E(εi) = 0 | | III. Paklaidos neautokoreliuoja (likučių ne autokoreliacijos) , t.y, paklaidos tarpusavyje nėra susijusios ir nestebimi sklaidos dėsningumai. | Cov(εi εj) = 0, ∀i,j / i≠j | | IV. Paklaidų dispersija yra pastovi (ne heteroskedastiškumas) Didėjant nepriklausomų kintamųjų reikšmėms, priklausomojo kintamojo sklaidos intervalas išlieka pastovus. | σ2(εi) = konstanta | | VI. Paklaidos pasiskirsčiusios pagal normalųjį skirstinį (normalumas). | εi ~ N (0, σ2) |   II prielaida yra visuomet tenkinama, jeigu į modelį yra įtrauktas laisvasis narys β0. Šio nario įtraukimas į modelį užtikrina, kad bus tenkinama II modelio prielaida. Todėl atradus, kad regresijos lygtyje nereikšmingas laisvasis narys, būtina įsitinti, ar jį pašalinus, bus tenkinama II prielaida.  Tai galima padaryti, suskaičiavus EXCEL skaičiuokle regresinį modelį ir apskaičiavus išklotinės 4 lentelėje pateiktų paklaidų vidurkį. Jeigu modelyje be laisvojo nario paklaidų vidurkio reikšmė nėra lygi 0, tuomet verta palikti modelyje laisvąjį narį, nors jis statistiškai ir nėra reikšmingas.  III prielaida reikalauja, kad paklaidos nebūtų koreliuotos, t.y., nepriklausytų vienos nuo kitų. Šios prielaidos netenkinimas nurodo, kad paklaidos nėra atsitiktinis dydis, o jose galima įžvelgti dėsningumus, o tai reiškia, kad modelis nėra sudarytas adekvačiai ir jo paklaidose atsispindi sisteminės dalies elementas.  **Paklaidų autokoreliacija**  Autokoreliacijos problema yra tada, kai modelio paklaidos yra susijusios tarpusavyje, t.y. jų koreliacijos koeficientas nėra lygi 0.  **Kodėl blogai?** Kai sudarytame modelyje yra autokoreliacija, tai:   1. Mažiausių kvadratų metodu (MKM) apskaičiuotas determinacijos koeficientas R2 yra didesnis už tikrąjį. 2. Mažiausių kvadratų metodu (MKM) apskaičiuotas standartinės paklaidos SEbj yra mažesnės. 3. Tikrinant hipotezes negalima naudoti nei t-stjudento nei F kriterijaus, nes gaunamos didesnės negu iš tikrųjų statistikų reikšmės ir galima suklysti, darant išvadas apie veiksnių statistinį reikšmingumą.     **Kokiu būdu nustatyti?** Paklaidų autokoreliacijai nustatyti yra žinomi specialūs statistiniai testai, iš kurių šiame paskaitų konspekte bus pateikti du:   1. Durbin-Watson testas 2. Ženklų sekos kriterijus.   Durbin-Watson testas – tai autokoreliacijos nebuvimo modelio paklaidose hipotezės tikrinimo procedūra:  H0: autokoreliacijos nėra. H1: autokoreliacija yra.  Šiam testui atlikti reikia suskaičiuoti DW d statistiką: kurioje ei – tai paklaida, o ei-1 – tai ankstesnio stebėjimo arba vėluojanti paklaida.  ,  Apskaičiuota statistika d gali įgauti reikšmes nuo 0 iki 4. Autokoreliacijos nėra, kai d statistika lygi arba artima 2. Norint tiksliai atsakyti į klausimą, ar yra ar nėra paklaidų autokoreliacija, reikia Durbin\_Watson d reikšmių lentelėse (pateiktos dėst. tnklalapyje) rasti kritines :apatinio rėžio - dL irviršutinio rėžio dU reikšmes, priklausomai nuo stebėjimų skaičiaus n (lentelės pirmas stulpelis) ir į modelų įtrauktų veiksnių skaičiaus k (viršutinė lentelės eilutė)   * Autokoreliacijos nėra arba H0 neatmetama , kaidU ≤ d ≤ 4 - dU * Autokoreliacija yra, kai d ≤ dL (teigiama autokoreliacija) arba d ≥ 4 - dL (neigiama autokoreliacija). Šiuo atveju hipotezė H0 yra atmetama.⇒ H1 * Šiame teste gali būti dvi neapibrėžtumo sritys, dL ≤ d ≤ dU arba 4- dU ≤ d ≤ 4 - dL kuriose griežto atsakymo apie paklaidų autokoreliacija nėra.   Grafinis testo rezultatų pavidalas pateiktas žemiau.      Ženklų sekų kriterijus.  Šiam kriterijui reikalinga turėti modelio paklaidas, kurias galima rasti Excel Regression skaičiuoklės Residual Output lentelėje. Šalia paklaidų reikia surašyti ženklus: jei paklaida didesnė už 0, tai “+”, ir jei paklaida mažesnė už 0, tai “–”. Tada reikia suskaičiuoti sekas, o viena seką sudaro vienodi ženklai. Kai ženklas pasikeičia, tai prasideda antra seka. Reikia suskaičiuoti sekų vidurkį ir dispersiją. :  ir ,  kur n- stebėjimų skaičius, n1 – paklaidų su ”+” ženklu skaičius, n2 –paklaidų su ”-”, skaičius k- sekų skaičius. Tada reikia atlikti hipotezių tikrinimo procedūrą:  H0:  Autokoreliacijos nėra. Sekų skaičius k atsitiktinis, nepriklausomas ir pasiskirstęs pagal normalųjį skirstinį dydis.  HA:  Autokoreliacija yra. Sekų skaičius k nėra atsitiktinis, nepriklausomas ir pasiskirstęs pagal normalųjį skirstinį,  Jei apskaičiuota k (sekų skaičiaus reikšmė) patenka į intervalą:  ,  tai H0 atmesti nėra pagrindo, tad su 95 % tikimybe galima teigti, jog autokoreliacijos nėra.  **Kaip išspręsti?** Autokoreliacijos sprendimo būdai:   * Įtraukti naujus veiksnius. arba laiko veiksnį, kuris rodo stebėjimo periodo numerį (1,2,3..it t.t.) Šie veiksniai gali išimti inerciją, jeigu ji yra būdinga nagrinėjamam reiškiniui. * Į modelį įtraukti vėluojantį priklausomąjį knitamąjį yt-1, , bet šiuo atveju būtina patikrinti ar neatsirado stipri koreliacija tarp yt-1 ir modelio paklaidų ei. * Peržiūrėti modelio matematinę išraišką. Gal būt šį reiškinį gali geriau atspindėti ne tiesinė, o kita, pvz, logaritminė, eksponentinė ar kita, regresijos lygties forma. * Transformuoti duomenis. Pavyzdžiui, skaičiuoti ne absoliučius dydžių, o pokyčių regresiją:   **Autokoreliacijos tikrinimas duonos kainos modelyje**  Pirmiausia patikrinsime ar egzistuoja paklaidų autokoreliacija Durbin Watson testu. Šiam tikslui mums reikia turėti modelio paklaidas. Jos yra pateikiamos EXCEL\_Regression išklotinės ketvirtoje lentelėje RESIDUALS OUTPUT. Pateiktoje lentelėje apačioje trys pirmieji stulpeliai – tai tiesinio modelio Residuals Output išklotinės fragmentas.   |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | Autokoreliacijos tikrinimas DW -testu | | | | | | Ženklų kriterijus | | |  | | *Stebėjimo nr.* | *Y-apskaičiuota Duonos kaina, Lt/kg* | *Paklaidos ei (Residuals)* | *vėluojančios paklaidos ei-1* | (ei -e i-1)^2 | (ei )^2 |  | *Paklaidos ei (Residuals)* | Ženklai |  | | 1 | 4,4288 | 0,0712 |  |  |  |  | 0,0712 | + |  | | 2 | 4,4346 | 0,0754 | 0,0712 | 0,0000 | 0,0057 |  | 0,0754 | + |  | | 3 | 4,4065 | 0,0735 | 0,0754 | 0,0000 | 0,0054 |  | 0,0735 | + |  | | 4 | 4,4201 | -0,0201 | 0,0735 | 0,0088 | 0,0004 |  | -0,0201 | - |  | | 5 | 4,4163 | -0,0863 | -0,0201 | 0,0044 | 0,0075 |  | -0,0863 | - |  | | 6 | 4,4018 | -0,0918 | -0,0863 | 0,0000 | 0,0084 |  | -0,0918 | - |  | | 7 | 4,4042 | -0,0442 | -0,0918 | 0,0023 | 0,0020 |  | -0,0442 | - |  | | 8 | 4,3660 | -0,0260 | -0,0442 | 0,0003 | 0,0007 |  | -0,0260 | - |  | | 9 | 4,3461 | 0,0339 | -0,0260 | 0,0036 | 0,0012 |  | 0,0339 | + |  | | 10 | 4,3226 | -0,0026 | 0,0339 | 0,0013 | 0,0000 |  | -0,0026 | - |  | | 11 | 4,3458 | -0,0058 | -0,0026 | 0,0000 | 0,0000 |  | -0,0058 | - |  | | 12 | 4,3568 | -0,0268 | -0,0058 | 0,0004 | 0,0007 |  | -0,0268 | - |  | | 13 | 4,2788 | -0,0088 | -0,0268 | 0,0003 | 0,0001 |  | -0,0088 | - |  | | 14 | 4,2984 | -0,0084 | -0,0088 | 0,0000 | 0,0001 |  | -0,0084 | - |  | | 15 | 4,3022 | 0,0678 | -0,0084 | 0,0058 | 0,0046 |  | 0,0678 | + |  | | 16 | 4,3295 | 0,0305 | 0,0678 | 0,0014 | 0,0009 |  | 0,0305 | + |  | | 17 | 4,3344 | -0,0544 | 0,0305 | 0,0072 | 0,0030 |  | -0,0544 | - |  | | 18 | 4,3384 | 0,0216 | -0,0544 | 0,0058 | 0,0005 |  | 0,0216 | + |  | | 19 | 4,4533 | -0,0133 | 0,0216 | 0,0012 | 0,0002 |  | -0,0133 | - |  | | 20 | 4,5346 | -0,0046 | -0,0133 | 0,0001 | 0,0000 |  | -0,0046 | - |  | | 21 | 4,5932 | -0,0632 | -0,0046 | 0,0034 | 0,0040 |  | -0,0632 | - |  | | 22 | 4,5974 | 0,0226 | -0,0632 | 0,0074 | 0,0005 |  | 0,0226 | + |  | | 23 | 4,6077 | 0,0523 | 0,0226 | 0,0009 | 0,0027 |  | 0,0523 | + |  | | 24 | 4,8126 | -0,0626 | 0,0523 | 0,0132 | 0,0039 |  | -0,0626 | - |  | | 25 | 4,8742 | -0,1142 | -0,0626 | 0,0027 | 0,0130 |  | -0,1142 | - |  | | 26 | 4,9314 | -0,0814 | -0,1142 | 0,0011 | 0,0066 |  | -0,0814 | - |  | | 27 | 4,8981 | 0,0719 | -0,0814 | 0,0235 | 0,0052 |  | 0,0719 | + |  | | 28 | 4,8792 | 0,1208 | 0,0719 | 0,0024 | 0,0146 |  | 0,1208 | + |  | | 29 | 4,8342 | 0,1058 | 0,1208 | 0,0002 | 0,0112 |  | 0,1058 | + |  | | 30 | 4,9549 | -0,0249 | 0,1058 | 0,0171 | 0,0006 |  | -0,0249 | - |  | | 31 | 4,9483 | 0,0117 | -0,0249 | 0,0013 | 0,0001 |  | 0,0117 | + |  | | 32 | 4,9316 | -0,0516 | 0,0117 | 0,0040 | 0,0027 |  | -0,0516 | - |  | | 33 | 4,9358 | 0,0242 | -0,0516 | 0,0057 | 0,0006 |  | 0,0242 | + |  | | 34 | 4,9757 | 0,0043 | 0,0242 | 0,0004 | 0,0000 |  | 0,0043 | + |  | | 35 | 4,9687 | 0,0313 | 0,0043 | 0,0007 | 0,0010 |  | 0,0313 | + |  | | 36 | 4,9675 | -0,0275 | 0,0313 | 0,0035 | 0,0008 |  | -0,0275 | - |  | |  |  |  | Suma | 0,1305 | 0,1087 |  |  |  |  | |  |  |  | **DW-d 0,1305/0,1087** | | | | | | | |  |  | **dL** | **1,098** |  |  |  |  |  |  | |  |  | **dU** | **1,442** | **Išvada apie autokoreliaciją : Rezultatas neapibržtas** | | | | | |   Atliekamas DW testas. **Pirmas žingsnis**: Formuluojamos hipotezės:  H0: autokoreliacijos nėra. H1: autokoreliacija yra.  **Antras žingsnis:** Skaičiuojama DW-d statistika pagal formulę    Šiam tikslui reikia turėti modelio paklaidas ei ir vėluojančias vienu periodu paklaidas ei-1. Paklaidos ei yra pateikiamos trečiajame lentelės stulpelyje. Vėluojančios paklaidos , tai tos pačios paklaidos tik paslinktos viena eilute žemiau, kaip parodyta ketvirtajame lentelės stulpelyje. Tuomet yra skaičiuojamas DW-d statistikos skaitikliui reikalingos paklaidų ir vėluojančių paklaidų skirtumo kvadratų reikšmės, t.y., (ei- ei-1)2 reikšmės, kurios pateiktos penktajame lentelės stulpelyje. Apskaičiavus reikšmes, jos susumuojamos ir gaunama formulės skaitiklyje esanti suma lygi 0,1305. Po to skaičiuojama vardiklyje esanti paklaidų kvadratų reikšmė, pradedant nuo antrosios paklaidos reikšmės. Šios reikšmės yra pateiktos lentelės šeštajame stulpelyje. Taip pat suskaičiuojama jų suma, kurios reikšmė yra 0,1087. Tuomet belieka pirmąją sumą padalinti iš antrosios ir gauname DW –d reikšmę 1,199817.  Trečias žingsnis:  Palyginti apskaičiuotą DW-d statistikos reikšmę su teorinėmis d reikšmėmis, kurios yra pateikiamos specialiose lentelėse. Tokia lentelė yra pateikta dėstytojos tinklalapyje (V.Karpuskiene rubrikoje KVST ) Kritines reikšmes rasime tokiu būdu: Pirmame DW statistinių lentelių stulpelyje yra nurodoma, kiek sudarytame modelyje yra įtraukta stebėjimų. Mūsų pavyzdyje yra n36 stebėjimai. Tuomet eilutėje surandame k=3. Šis skaičius k parodo, kiek yra į modelį įtraukta nepriklausomų kintamųjų. Mūsų pavyzdyje yra 3 kintamieji : rugiai, dyzelinas ir darbo užmokestis.  Eilutės su 36 stebėjimais ir stulpelių ksusikirtime randame dL ir du reikšmes . Jos yra lygios  dL ir dU  Apskaičiuota statistika d gali įgauti reikšmes nuo iš intervalo nuo 0 iki 4, kuris padalintas į:   * Autokoreliacijos nėra arba H0 neatmetama , kaidU ≤ d ≤ 4 - du,.t.y.,1,442≤d ≤2,558 * Autokoreliacija yra, kai d ≤ dL (teigiama autokoreliacija) arba d ≥ 4 - dL (neigiama autokoreliacija). Šiuo atveju hipotezė H0 yra atmetama.⇒ * Šiame teste gali būti dvi neapibrėžtumo sritys, dL ≤ d ≤ dU arba 4- dU ≤ d ≤ 4 - dL, kuriose griežto atsakymo apie paklaidų autokoreliacija nėra. Mūsų pavyzdyje apskaičiuota d reikšmė patenka į neapibrėžtumo intervalą: 1,098≤ d ≤ 1,442. Todėl išvada , kad DW testas tiksliai atsakyti apie autokoreliacija negali.   Kai gaunamas neapibrėžtas atsakymas, tuomet verta atlikti Ženklų sekų testą.  Ženklų sekų kriterijus.  Šiam kriterijui reikalinga Residual Output lentelė, kurioje būtų paklaidų reikšmės. Šalia paklaidų reikia surašyti ženklus: jei paklaida > 0, tai “+”, ir jei paklaida < 0, tai “–”.  Po to, reikia suskaičiuoti ženklų sekas, o viena seką sudaro vienodi ženklai. Tai pavaizduota lentelėje viršuje 8 ir 9 stulpelyje. Kaip matyti pirmos trys paklaidos yra teigiamos, todėl prie jų rašome “+” ženklą. Tai pirmoji ženklų seka. Ketvirtoji ir dar kelios paklaidos žemiau yra neigiamos, todėl prie jų pažymime “-“ ženklą. Tai yra antroji ženklų seka. Kai ženklas pasikeičia, tai prasideda kita seka. Mūsų pavyzdyje yra 16 sekų. Reikia suskaičiuoti kiek yra n1 - teigiamų su ”+” paklaidų : Mūsų pavyzdyje n1 16 ir kiek neigiamų su ”-” paklaidų n220  Po to suskaičiuojame    ir =8.52  H0:  Autokoreliacijos nėra.  HA:  Autokoreliacijos yra.  Jei apskaičiuota k (sekų skaičiaus reikšmė) patenka į intervalą:  ,  gauname kad apskaičiuotas sekų skaičius patenka į intervalą: 13,07<k=16<24.52  H0 atmesti nėra pagrindo, tad su 95 % tikimybe galima teigti, jog autokoreliacijos nėra.  Taigi pagal šį kriterijų negalime atmesti H0, kuri teigia, kadpaklaidų autokoreliacijos nėra. Galima konstatuoti faktą, kad abu sudaryti modeliai analizei yra tinkami. |

PRIEDAS 1

Pradinių tyrimo duomenų lentelė

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Metai/Mėnuo | Duonos kaina, Lt/kg | Dyzelino kaina, Lt/ltr | Rugių kaina, Lt/t | Elektros kaina, ct/kWh | Vid. Darbo Užmokestis, Lt/mėn. | Cukraus kaina, Lt/kg | PVM pakeitimas |
| 2009/03 | 4,50 | 3,22 | 238,87 | 23,54 | 1884,43 | 3,18 | 0 |
| 2009/04 | 4,51 | 3,12 | 244,12 | 23,54 | 1884,43 | 3,21 | 0 |
| 2009/05 | 4,48 | 2,96 | 218,93 | 23,54 | 1884,43 | 3,19 | 0 |
| 2009/06 | 4,40 | 2,98 | 235,31 | 23,54 | 1880,63 | 3,18 | 0 |
| 2009/07 | 4,33 | 3,07 | 231,95 | 23,54 | 1880,63 | 3,20 | 0 |
| 2009/08 | 4,31 | 3,23 | 218,93 | 23,54 | 1880,63 | 3,17 | 0 |
| 2009/09 | 4,36 | 3,16 | 215,37 | 22,18 | 1885,77 | 3,17 | 0 |
| 2009/10 | 4,34 | 3,09 | 181,15 | 22,18 | 1885,77 | 3,20 | 0 |
| 2009/11 | 4,38 | 3,01 | 163,31 | 22,18 | 1885,77 | 3,21 | 1 |
| 2009/12 | 4,32 | 3,00 | 161,21 | 22,18 | 1868,70 | 3,18 | 1 |
| 2010/01 | 4,34 | 3,13 | 181,99 | 22,18 | 1868,70 | 3,17 | 1 |
| 2010/02 | 4,33 | 3,09 | 191,86 | 22,18 | 1868,70 | 3,13 | 1 |
| 2010/03 | 4,27 | 3,29 | 209,91 | 29,57 | 1789,33 | 3,11 | 1 |
| 2010/04 | 4,29 | 3,33 | 227,54 | 29,57 | 1789,33 | 3,07 | 1 |
| 2010/05 | 4,37 | 3,40 | 230,90 | 29,57 | 1789,33 | 3,09 | 1 |
| 2010/06 | 4,36 | 3,52 | 238,04 | 29,57 | 1805,00 | 3,08 | 1 |
| 2010/07 | 4,28 | 3,57 | 242,44 | 29,57 | 1805,00 | 3,09 | 1 |
| 2010/08 | 4,36 | 3,59 | 246,01 | 29,57 | 1805,00 | 2,98 | 1 |
| 2010/09 | 4,44 | 3,62 | 271,62 | 29,31 | 1874,77 | 2,93 | 1 |
| 2010/10 | 4,53 | 3,57 | 344,46 | 29,31 | 1874,77 | 2,95 | 1 |
| 2010/11 | 4,53 | 3,56 | 396,94 | 29,31 | 1874,77 | 2,97 | 1 |
| 2010/12 | 4,62 | 3,57 | 407,85 | 29,31 | 1868,33 | 3,02 | 1 |
| 2011/01 | 4,66 | 3,58 | 417,09 | 29,31 | 1868,33 | 3,06 | 1 |
| 2011/02 | 4,75 | 3,85 | 484,26 | 29,31 | 1868,33 | 3,05 | 1 |
| 2011/03 | 4,76 | 4,02 | 566,75 | 29,31 | 1843,73 | 3,14 | 1 |
| 2011/04 | 4,85 | 4,10 | 617,97 | 29,31 | 1843,73 | 3,32 | 1 |
| 2011/05 | 4,97 | 4,31 | 588,16 | 29,31 | 1843,73 | 3,55 | 1 |
| 2011/06 | 5,00 | 4,41 | 539,04 | 29,31 | 1872,77 | 3,72 | 1 |
| 2011/07 | 4,94 | 4,29 | 498,74 | 29,31 | 1872,77 | 3,84 | 1 |
| 2011/08 | 4,93 | 4,25 | 606,84 | 29,31 | 1872,77 | 3,86 | 1 |
| 2011/09 | 4,96 | 4,31 | 460,96 | 29,31 | 1872,77 | 3,88 | 1 |
| 2011/10 | 4,88 | 4,27 | 446,05 | 29,31 | 1872,77 | 3,92 | 1 |
| 2011/11 | 4,96 | 4,25 | 449,83 | 29,31 | 1872,77 | 3,92 | 1 |
| 2011/12 | 4,98 | 4,33 | 462,22 | 29,31 | 1915,07 | 3,96 | 1 |
| 2012/01 | 5,00 | 4,42 | 455,92 | 29,31 | 1915,07 | 3,96 | 1 |
| 2012/02 | 4,94 | 4,43 | 454,87 | 29,31 | 1915,07 | 4,01 | 1 |



PRIEDAS 2 TS –Tiesinis modelis

PRIEDAS 3 LN modelis –Tiesinis modelis



PRIEDAS 4 TS –tiesinio modelio galutinė regresijos lygtis

PRIEDAS 5 LS – logaritminio modelio galutinė regresijos lygtis

