

PARDAVIMŲ PROGNOZAVIMAS NAUDOJANT REGRESINĮ MODELIAVIMĄ

Aistė Kazakevičiūtė, vadovas Virgilijus Sakalauskas
Kauno kolegija

Šiais laikais prognozavimas atlieką svarbų vaidmenį mūsų gyvenime. Vieni prognozuoja savo finansus, kiti būsimas rinkos tendencijas, tačiau dažniausiai su prognozavimu susiduriame apie tai net nepagalvoję. Šiame straipsnyje naudojantis regresine analize ir programine įranga R, bus sukurtas modelis, kuris padės prognozuoti akcijose dalyvaujančios prekės A būsimų pardavimų prognozes ir patikimumą.

Įvadas

Prognozavimas visais laikais buvo neatsiejama žmogaus veiklos dalis. Prognozavimas nuo spėjimo skiriasi tuo, kad spėjant remiamasi tik subjektyviu mąstymu, intuicija, o prognozuojant būsimus įvykius remiamasi dažniausiai istoriniais duomenimis, kurie gali atskleisti pasikartojimus, struktūrą ir pan.. Nuo tikslesnių prognozių priklausydavo ekonominė gerovė – laiku pasėtas ar nuimtas derlius galėdavo reikšti turtą arba nepritekliaus nebuvimą.

Temos aktualumas: šiandien prognozavimas yra reikalingas kiekvienai įmonei, norinčiai pasiekti geresnių finansinių rezultatų. Prognozė padeda organizuoti tolimesnę įmonės veiklą, sudaro plėtros galimybes ir t.t..

Tyrimo metodika: pardavimų prognozavime dažnai naudojami metodai: koreliacinė ir regresinė analizė, dispersinė analizė ir laiko eilutės. Modelio sukūrimui naudojama daugianarė regresinė analizė. Daugianarės regresijos lygtis sudaro galimybę numatyti, kaip keisis pardavimų apimtys, esant įvairiems dalyvaujančių veiksmų ar reikšmių variantams [1]. Pagrindinis regresijos tikslas - priklausomo kintamojo prognozavimas, atsižvelgiant į regresorių įtaką priklausomam kintamajam. Teigiama, jog įsitikinus, kad duomenys korektiški ir koreliuojantys, bei apskaičiavus rodiklius, rezultatai atspindi tikrus ryšius tarp nagrinėjamų reiškinių. **Temos problematika:** viena didžiausių prognozavimo problemų - prognozės tikslumas. Veiksmingas pardavimų prognozavimas atlieka svarbų vaidmenį įmonės valdyje, tai padeda parodyti įmonės augimo galimybes, išanalizuoti pardavimus, efektyviai valdyti centrinio sandėlio atsargas, stebėti marketingo pokyčius. **Darbo tikslas:** akcijinių prekių pardavimų prognozavimas naudojant regresinį modeliavimą. **Darbo uždaviniai:**

1. Išanalizuoti regresinės analizės galimybes prognozuojant prekių pardavimus;
2. Pagal atrinktus faktorius ir įmonės surinktus istorinius duomenis sudaryti regresinį pardavimų prognozavimo modelį;
3. Patikrinti modelio tinkamumą vienos prekės akcijiniams pardavimams.

Teorinis tyrimo pagrindimas

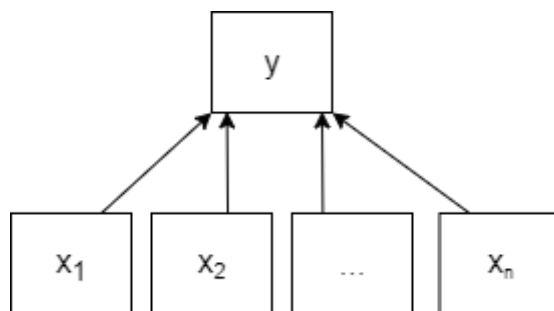
Žinome, kad tam tikrus reiškinius nusako vienas ar keli veiksniai. Priimant sprendimus dažnu atveju neužtenka išvardinti nagrinėjamą situaciją sąlygojančius veiksnius, jų poveikį įvertinti kiekybiškai. Šiam tikslui naudojama regresinė analizė, kurios pagalba veiksmų įtaką reiškiniui galima užrašyti matematine lygtimi.

Regresinė analizė - statistinis metodas, kai naudojant matematinės procedūras, gaunama lygtis arba jų sistema, įvertinanti vieno ar daugiau veiksmų įtaką nagrinėjamam reiškiniui [3].

Turint matematinę priklausomybės lygtį, galima:

- matematiškai aprašyti nagrinėjamo reiškinio priklausomybę nuo jį sąlygojančių veiksmų. Konkreti regresijos ryšio matematinė lygtis leidžia gauti ekonominės analizės išvados naudingus rodiklius: ryšio ženklą ir įtakos pobūdį, nagrinėjamo reiškinio elastingumą kiekvienam iš veiksmų arba visų veiksmų poveikiui bendrai, veiksmų pakeičiamumo santykius;
- nustatyti nagrinėjamą situaciją veikiančius ir nedarančius jai esminio poveikio veiksnius;
- prognozuoti nagrinėjamo reiškinio variantus;
- modeliuoti įvairias verslo situacijų alternatyvas ir lyginti jas tarpusavyje.

Kintamasis y - priklausomas kintamasis, kintamieji x_1, x_2, x_3 - aiškinamieji kintamieji, dar žinomi kaip regresoriai. Visos spėjamos priklausomybės yra tiesinės. Priklausomam kintamajam daro įtaką vienas ar keli nepriklausomi kintamieji. Bendroji modelio schema:



1 pav.: Bendroji modelio schema

Formalus regresinis modelis atrodo taip:

$$y = a_0 + a_1x_1 + a_2x_2 + \dots + e$$

Čia e žymi liekamąją paklaidą, t.y. viską įvertiname ir nuo ko dar gali priklausyti y . Nei koeficientai a_1, a_2, a_3 , nei liekamoji paklaida nėra žinomi. Įverčiai $\hat{a}_1, \hat{a}_2, \hat{a}_3$ gaunami panaudojant imties duomenis. Gauname regresijos lygtis apytikslei \hat{y} reikšmei:

$$\hat{y} = \hat{a}_0 + \hat{a}_1x_1 + \hat{a}_2x_2 + \dots \quad (1)$$

Lygtis (1) naudojama kokybinei bei kiekybinei kintamųjų priklausomybių analizei.

Koeficiento ženklas nurodo ar regresorius didėjant \hat{y} didės, ar mažės.

- Jei $\hat{a}_1x_1 > 0$, tai x didėjant \hat{y} didėja.
- Jei $\hat{a}_1x_1 < 0$, tai x didėjant \hat{y} mažėja.

Koeficientas \hat{a}_1x_1 parodo, kiek pasikeis \hat{y} reikšmė, vienu vienetu padidėjus x ir fiksavus visų kitų regresorių reikšmes. Prognozuojant pasirinktos konkrečios regresorių reikšmės būna įstatomos į gautą regresijos lygtį (formulė (1)). Reikia nepamiršti, jog prognozės daromos tik a, x reikšmėms, kurios mažai skiriasi nuo turimų imties reikšmių. Visi duomenys turi būti skaitiniai. Kategorinių kintamųjų reikšmės yra koduojamos skaičiais. Taip pat:

- priklausomas kintamasis y ir regresoriai (tik ne dvireikšmiai) vadinami intervaliniais kintamaisiais;
- klasikiniame modelyje, teigiama, jog regresoriai matuojami be paklaidų ir yra neatsitiktiniai;
- protarpiais į modelį įtraukiami ir kategoriniai kintamieji - *pseudokintamieji*. Visus pseudokintamuosius reikia perkoduoti taip, kad įgytų tik dvi reikšmes - 0 arba 1;
- skirtingų stebinių liekamoji paklaida e negali koreliuoti;
- regresoriai negali stipriai koreliuoti;
- duomenyse negali būti išskirčių;
- duomenys privalo būti homoskedastiški. Reikalaujama, jog laikomoji paklaidos dispersija nepriklaustų nuo regresorių reikšmių.

Taigi regresinė analizė - ryšys tarp dviejų ar daugiau kintamųjų, o jos ryšio stiprumo nustatymas - koreliacijos analizė. Daugianarė koreliacinė regresija leidžia įvertinti vieno iš veiksnių ryšį su visais kitais kaip visumą [2].

Prognozuojant modelį dauguma matavimų yra susiję su tam tikru netikslumu.

Absoliučioji paklaida (žym. Δa) - dydžio apytikslės ir tikslios reikšmių skirtumo modulis. Kitaip tariant, tikslaus skaičiaus a ir jam artimo apytikslio skaičiaus x skirtumo modulis.

$$\Delta a = |x - a|$$

Ši paklaida, o kartu ir matavimo tikslumas, susijęs su matavimo prietaiso netobulumu, teorinio modelio ir taikomo metodo matavimu. Atsitiktinių paklaidų priežastys yra atsitiktiniai ir eksperimento metu nekontroliuojami trikdžiai, kurių įtakos matavimo rezultatui tiesiogiai įvertinti negalima dėl to, kad trikdžių yra daug, jų prigimtis įvairi ir poveikis skirtingas.

Santykinė paklaida (žym. δa) - absoliučiosios paklaidos ir apytikslės reikšmės santykis.

$$\delta a = \frac{\Delta a}{x} = \frac{|x - a|}{x}$$

Santykinė paklaida gali būti apskaičiuojama ir **procentine išraiška**:

$$\delta a_{\%} = \frac{|x - a|}{x} \cdot 100\%$$

Paklaidos tikslumas - vertinamas, jeigu santykinė procentinė paklaida yra didesnė arba lygi 100%, tikslumas 0%. Jeigu santykinė procentinė paklaida mažiau 100%, skaičiuojama pagal formulę:

$$100 - \delta a_{\%} = 100 - \frac{|x - a|}{x} \cdot 100\%$$

Tyrimo rezultatai ir jų analizė

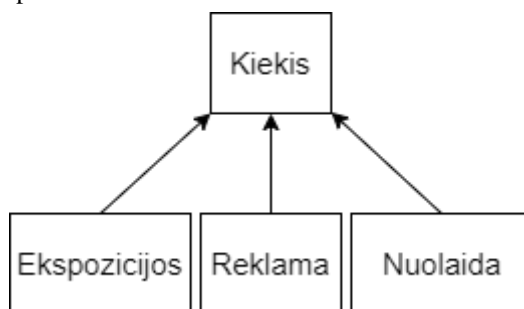
Šiame poskyryje bus sukurtas regresijos modelis vienai prekei. Analizuojama prekė A. Šios prekės kiekis centriniame sandėlyje palaikomas automatiškai. Istoriniuose duomenyse matoma, jog seniau prekė dalyvavo akcijose, kurių trukmė įvairi. Dėl šios priežasties žinodami, jog būsima akcija truks daugiau (arba mažiau) dienų, parduotą kiekį išlyginsime proporcijos pagalba. Taip pat turime faktorius: nuolaida, reklama ir ekspozicijos tipas. Reklamos ir ekspozicijos tipai duomenyse užfiksuoti kaip kategoriniai kintamieji.

Prieš kuriant modelį iš duomenų išmetami tam tikrų mėnesių pardavimai. Gruodžio mėnesio - dėl šventinio laikotarpio, kovo ir balandžio mėnesiai - dėl karantino (fizinės parduotuvės tuo metu nedirbo).

Eliminavus nekorektiškus stebinius duomenyse, liko 16 stebinių. Prekė A nuo 2017 metų reklamoje dalyvavo vieną kartą, ekspozicijoje keturis kartus. Nuolaida daugiausia kartų buvo taikoma 19%. Pastebima, kaip prekė dalyvavo reklamoje ir ekspozicijoje, parduotas kiekis buvo 1018 vnt.. Kai nedalyvavo reklamoje, tik ekspozicijoje, pardavimai buvo ~700 vnt.. Nuolaida abu kartus buvo vienoda - 42%. Tai puikus pavyzdys, kokią įtaką turi reklamos faktorius.

Toliau bus sudaryta regresijos tiesės lygtis ir suprognozuotas vienos prekės kiekis, atsižvelgiant į būsimą nuolaidą ir kitus faktorius. Nagrinėjamuose duomenyse nepriklausomi kintamieji: nuolaida, reklama, ekspozicija, priklausomas kintamasis - kiekis.

Bus sudarytas modelis, kuris padės prognozuoti akcijos kiekius ir prognozę automatizuoti. Faktoriniai, kuriuos išskirsime yra pateikti 2 paveiksle.



2 pav.: Modelio schema, atvaizduojanti įtakos pardavimams turinčius faktorius

Pasinaudojus programine įranga „R“ ir panaudojus istorinius duomenis, buvo apskaičiuoti įverčiai:

$$\hat{a}_0 = -29,371; \hat{a}_1 = 12,879; \hat{a}_2 = 212,575; \hat{a}_3 = 263,876.$$

1 lentelė: Modelio koeficientai ir jų reikšmingumas

| | Įvertis | Standartinė paklaida | T | Pr(> t) |
|-------------|----------|----------------------|-------|----------|
| Konstanta | -29,3710 | 102,5059 | -0,29 | 0,7794 |
| Nuolaida | 12,8790 | 5,1664 | 2,49 | 0,0283 |
| Reklama | 212,5753 | 55,2153 | 3,85 | 0,0023 |
| Ekspozicija | 263,8763 | 109,3376 | 2,41 | 0,0327 |

Daugianarės regresinės analizės koreliacijos koeficiento patikimumas nustatomas pagal Studento kriterijaus t reikšmę. Lentelėje 1 paskutiniame stulpelyje pastebima, kad nuolaida, reklama, ekspozicija yra statistiškai reikšmingi. Regresorius *Konstanta* yra nereikšmingas, nes $0,7794 > 0,05$, todėl šią konstantą reikia pašalinti iš modelio.

Pagal gautus rezultatus susidarome regresinę lygtį, kuri atrodo taip:

$$\hat{y} = 12,88x_1 + 212,56x_2 + 263,88x_3 \quad (2)$$

arba:

$$\widehat{Kiekis} = 12,88 \cdot Nuolaida + 212,56 \cdot Reklama + 263,88 \cdot Ekspozicija$$

Patikrinus modelio tikslumą (lentelė 2) gauta, jog determinacijos koeficiento reikšmė $R^2 \sim 0,98$, Tai parodo, kad šiai prekei prognozės modelis yra ganėtinai tikslus.

2 lentelė: Determinacijos koeficiento R² ir reikšmingumo lygmens reikšmės

| | |
|----------------|-----------|
| R ² | 0,9762 |
| p reikšmė | 5,329e-10 |

Progozė: žinome, jog rugpjūčio mėnesį prekei A, bus taikoma 35% nuolaida, tačiau nebus nei reklamos, nei ekspozicijos. Kokį kiekį reikia užsakyti į centrinį sandėlį? Pasinaudojus regresijos lygtimi, kurią sudarėme pagal 2 formulę, įsistatome turimus duomenis ir gauname:

$$\hat{y} = 12,88 \cdot 35 + 212,56 \cdot 0 + 263,88 \cdot 0 = 450,8 \sim 451$$

Taigi, suprognuozuotos sąlygos: nuolaida 35%, ekspozicijos ir reklamos nėra, reikia užsakyti 451 vnt., prekės, Toliau bus atliktas prognozės patikrinimas, kuriame apskaičiuota absoliučioji paklaida (AP), santykinė paklaida (SP), santykinė procentinė paklaida (SP%) ir prognozės tikslumas (PT).

Prieš nustatant prekės tikslumą ir paklaidas reikia įsitikinti, kad prekė turi tenkinti sąlygą, jog teisingai buvo suprognuozuotas kiekis, t.y. akcijos pirmą dieną dalis užsakyto kiekio yra parduotuvėse, kita dalis centriniam sandėlyje. Nutarta, atlikus prieš akcijinį paskirstymą, likutis sandėlyje turi būti didesnis už 40% prognozuojamo kiekio, parduotuvės likutis didesnis už 60% prognozuojamo kiekio.

Gauti rezultatai pavaizduoti 3 lentelėje, Prognozės tikslumas yra 82%, santykinė paklaida procentais - 18%, absoliučioji paklaida - 82 vnt., Tai parodo, kad modelis yra pakankamai tikslus ir patikimas.

3 lentelė: Prognozės paklaidos ir tikslumas

| Kiekis | Prognozuojamas kiekis | R ² | AP | SP | SP% | PT |
|--------|-----------------------|----------------|----|------|-----|----|
| 369 | 451 | 0,98 | 82 | 0,18 | 18 | 82 |

Įmonė naudoja modelį, sukurtą 2013 metais. Šis modelis atsižvelgia tik į vieną faktorių – nuolaidą. Naudotas įmonės modelis buvo suprognuozavęs 615 vnt., kiekį akcijinio pardavimo, Absoliučioji paklaida gauta 246 vnt., santykinė procentinė paklaida - 40%, o prognozės tikslumas - 60%. Lyginant naują modelį ir senąjį modelį, nesunku pastebėti, kad naujas modelis yra su mažesne santykinė procentine paklaida ir didesniu prognozės tikslumu.

Išvados ir rezultatai:

1. Darbe išanalizavus turimus istorinius duomenis, jų struktūrą ir kaupimo laiko periodą, pasirinkta naudoti regresinę analizę.
2. Naudojant programinę įrangą R, buvo sukurtas modelis, paremtas regresine analize. Kiekvienai prekei kuriamas atskiras, tik jos duomenis atspindintis regresinis modelis.
3. Buvo patikrintas modelio tikslumas, apskaičiuota absoliučioji, santykinė, santykinė procentinė paklaida bei prognozės tikslumas. Naujos prognozės tikslumas buvo palygintas su senuoju modeliu. Naujo modelio santykinė procentinė paklaida gauta 18%, prognozės tikslumas prekei A - 82%. Prieš tai įmonėje naudoto modelio santykinė procentinė paklaida - 40%, prognozės tikslumas - 60%. Naujas modelis daro mažesnę santykinę paklaidą ir didesnę prognozės tikslumą.

Literatūra

1. Pabedinskaitė. *Kiekybiniai sprendimų metodai I dalis*. Vilnius: VGTU. 2009, 102p.
2. V. Bartosevičienė. *Ekonominės statistikos pagrindai*. KTU. 2011, 112p.
3. V. Karpuškienė. *Ekonometrija*. Vilnius. 2018.

SALE FORECASTING USING REGRESSION MODELS

Aiste Kazakeviciute, supervisor Virgilijus Sakalauskas
Kaunas University of Applied Sciences

Nowadays forecasting plays a big role in our lives. Some forecast their finances, some - upcoming market tendencies, however we most often forecast without even thinking, i.e. weather forecasting. In this article using regression analysis and software R we will create a model that will help determine future forecast and reliability of a product A that is a subject of various campaigns.