

VILNIAUS UNIVERSITETAS
MATEMATIKOS IR INFORMATIKOS FAKULTETAS
MATEMATINĖS STATISTIKOS KATEDRA

Daiva Lukaševičiūtė

**REGRESINIAI MODELIAI IŠGYVENAMUMO ANALIZĖJE
IR JŲ TAIKYMAS LIGONIŲ, SERGANČIŲ REUMATOIDINIŲ
ARTRITU, MIRTINGUMO ANALIZEI**

Magistro baigiamasis darbas

Vilnius, 2009

Darbo vadovas:

Profesorius Habil. dr. Vilijandas Bagdonavičius

Recenzentas:

Docentas Dr. Julius Jonas Kruopis

Registracijos Nr.:

Darbo gynimo data:

TURINYS

ĮVADAS.....	4
IŠGYVENAMUMO ANALIZĖ	5
1. IŠGYVENIMO TRUKMĖ.....	5
2. CENZŪRUOTI DUOMENYS.....	5
2.1. I tipo cenzūravimas.....	5
2.2. II tipo cenzūravimas.....	6
2.3. III tipo cenzūravimas	7
2.4. Cenzūravimas iš dešinės	9
2.5. Nupjovimas iš kairės.....	10
2.6. Nupjovimas iš kairės ir cenzūravimas iš dešinės	11
3. FUNKCIJOS.....	11
3.1. Išgyvenamumo (survivorship) funkcija	12
3.2. Tikimybinio tankio (density) funkcija.....	13
3.3. Nesėkmių intensyvumo (hazard) funkcija.....	15
4. FUNKCIJŲ SĄRYŠIAI	17
5. COX'Ų MODELIS	20
5.1. Dalinė tikėtinumų funkcija	20
5.2. Reikšmingų kovariančių nustatymas.....	23
6. PROPORCINGŲ MIRIMŲ INTENSYVUMŲ IR PERSIKERTANČIŲ MIRIMŲ INTENSYVUMŲ MODELIAI.....	24
7. KRITERIJUS COX'Ų MODELIO ADEKVATUMUI TIKRINTI, NAUDOJANT NUPJAUTUS IŠ KAIRĖS IR CENZŪRUOTUS IŠ DEŠINĖS DUOMENIS	27
REUMATOIDINIS ARTRITAS (RA).....	33
1. LIGONIŲ, SERGANČIŲ RA, SVEIKATOS BŪKLĖS TYRIMAS.....	33
2. COX'Ų MODELIO TAIKYMAS, TIRIANT LIGONIŲ, SERGANČIŲ REUMATOIDINIŲ ARTRITU, MIRTINGUMĄ	35
2.1. Ligonų mirtingumas kaip laiko nuo susirgimo funkcija.....	35
2.2. Ligonų mirtingumas kaip amžiaus funkcija	48
CHEMOTERAPIJA	57
1. PERSIKERTANČIŲ MIRIMŲ INTENSYVUMŲ MODELIO TAIKYMAS VĖŽIU SERGANTIEMS LIGONIAMS	57
2. KRITERIAUS COX'Ų MODELIO ADEKVATUMUI TIKRINTI TAIKYMAS REUMATOIDINIŲ ARTRITU SERGANTIEMS LIGONIAMS	60
IŠVADOS	61
UŽSIENIO KALBA	62
LITERATŪRA IR ŠALTINIAI	63
PRIEDAS NR. 1.....	64
PRIEDAS NR. 2.....	66
PRIEDAS NR. 3.....	68
PRIEDAS NR. 4.....	69
PRIEDAS NR. 5.....	70
PRIEDAS NR. 6.....	72
PRIEDAS NR. 7.....	74
PRIEDAS NR. 8.....	77
PRIEDAS NR. 9.....	85

IVADAS

Statistiniai išgyvenamumo analizės metodai labiausiai buvo išplėtoti per paskutinius du dešimtmečius. Šie metodai taikomi ne tik vėžio gydyme, patikimumo teorijoje, bet ir verslo, kriminologijos, epidemiologijos, socialinių mokslų srityse. Įvairūs statistiniai metodai taikomi analizuojant išgyvenamumo duomenis, atliekant laboratorinius tyrimus su gyvūnais, klinikinius ir epidemiologinius tyrimus su žmonėmis bei kitus panašius tyrimus.

Išgyvenamumo analizės tikslas yra nustatyti (nuspėti) pasveikimo, mirties tikimybes. Lyginami išgyvenimo skirstiniai, eksperimentuojant su gyvūnais ir stebint ligonius žmones. Bandoma nustatyti susirgimo riziką, reakciją į gydymą, išgyvenimą bei ligos vystymąsi lemiančius prognozuojančius veiksnius.

Praktikoje dažnai taikomi aprašomieji ir analiziniai statistiniai modeliai cenzūruotiems išgyvenamumo duomenims tirti, naudojant skaitinius ir grafinius metodus.

Dažniausiai naudojami išgyvenamumo skirstiniai yra eksponentinis, Veibulo, lognormalusis ir gama. Išvados priklauso nuo pasirinkto skirstinio.

Išgyvenamumo analizėje dažnai naudojami ne tik parametriniai, bet ir semiparametriniai bei neparametriniai tyrimų metodai.

Atliekant analizę, ypatingai svarbu parinkti tinkamą modelį arba skirstinį. Parinkus tinkamą statistinį metodą ir suradus parametrų įverčius, jau galima prognozuoti išgyvenamumo trukmę, pritaikyti tinkamiausią gydymą, planuoti tolesnius klinikinius ar laboratorinius tyrimus. Grafiniai metodai yra paprastas ir neformalus būdas statistinių modelių parinkimui ir parametrų įvertinimui.

Šiame darbe nagrinėjami duomenys apie Vilniaus miesto ligonius, sergančius reumatoidiniu artritu. Duomenų specifika yra ta, kad jie yra cenzūruoti iš dešinės ir nupjauti iš kairės.

Vienas iš darbo tikslų yra nustatyti ligonių mirtingumą lemiančius veiksnius, naudojant Cox'o modelį.

Kitai duomenų imčiai, t.y. vėžiu sergančių ligonių duomenų aibei, buvo taikomas Persikertančių mirimų intensyvumų modelis, t.y. tikrinta hipotezė, ar duomenims galioja Cox'o modelis.

Pagrindinis darbo tikslas buvo sukonstruoti kriterijų Cox'o modelio adekvatumui tikrinti, naudojant nupjautus iš kairės ir cenzūruotus iš dešinės duomenis, sudaryti programą kriterijui realizuoti.

Išgyvenamumo analizė

1. Išgyvenimo trukmė

Išgyvenimo trukmė plačiąja prasme gali būti apibrėžta kaip laikas iki tam tikro įvykio pasirodymo. Tas įvykis gali būti recidyvas, susirgimas (auglio išsivystymas), remisija, reakcija į gydymą arba mirtis.

2. Cenzūruoti duomenys

Dauguma mokslininkų išgyvenamumo analizę laiko paprasčiausiu trijų tradicinių statistinių metodų - parametrinio, semiparametrinio ir neparametrinio taikymu tam tikriems uždaviniams, problemoms spręsti. Parametrinis statistinis metodas taikomas tada, kai išgyvenimo trukmės skirstinys yra parenkamas iš kokios nors skirstinių šeimos, pavyzdžiui, normalaus, Veibulo ar kitų, o neparametrinis – kai išgyvenimo trukmės skirstinys yra nežinomas.

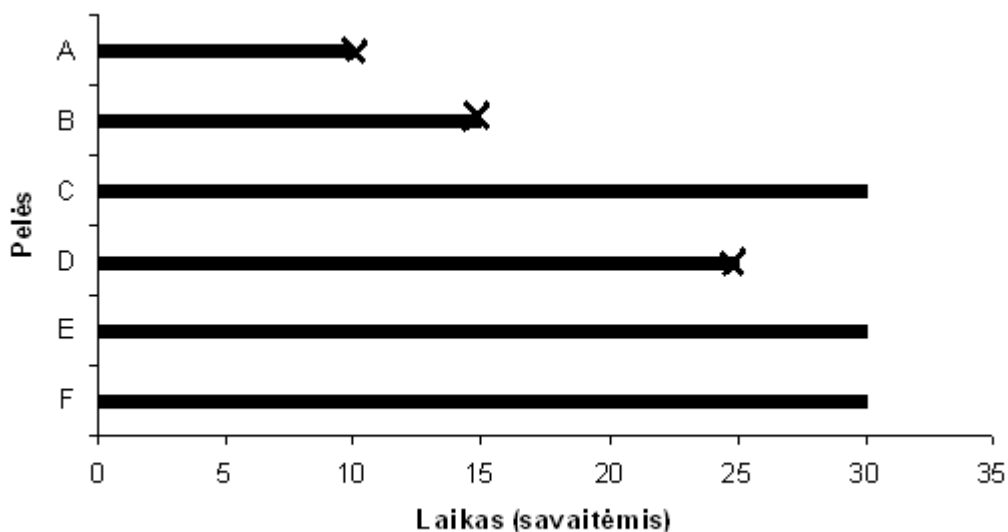
Reikia atkreipti dėmesį į tai, kad kai kuriems objektams dominantis įvykis neįvyksta iki tyrimo ar analizei skirto laiko pabaigos. Pavyzdžiui, kai kurie ligoniai vis dar gali likti gyvi arba nesusirgti, tyrimui pasibaigus. Tokių ligonių tiksli išgyvenimo trukmė lieka nežinoma ir stebėjimai vadinami cenzūruotais. Tokiu atveju imtis vadinama cenzūruota. Kai minėtų atvejų nėra, imtis vadinama pilna.

Cenzūravimas gali būti trijų tipų.

2.1. I tipo cenzūravimas

Tyrimai su gyvūnais dažniausiai pradedami fiksuojant gyvūnų, kuriems taikomas koks nors gydymas, skaičių. Dėl laiko ar/ir lėšų stygiaus, mokslininkai dažnai tiesiog negali tęsti tyrimo tol, kol mirs visi gyvūnai. Vienas iš būdų šiai problemai spręsti yra vykdyti tyrimą nustatytą, fiksuotą laiko tarpą, tarkim šešis mėnesius, kuriems praėjus, išgyvenę gyvūnai toliau nebestebimi. Išgyvenimo trukmė tų gyvūnų, kurie mirė per šešis mėnesius, yra žinoma. Tokie stebėjimai vadinami necenzūruotais. Tuo tarpu gyvūnų, kurie nemirė per šešis mėnesius, išgyvenimo trukmė nėra tiksliai žinoma. Tokie stebėjimai vadinami cenzūruotais. Be to, kai kurie gyvūnai gali atsitiktinai nugaišti, prieš pasirodant augliui, tada jų išgyvenimo trukmė nuo eksperimento pradžios iki auglio pasirodymo taip pat bus cenzūruotas stebėjimas. Taigi esant I tipo cenzūravimui, kai nėra atsitiktinių mirčių, cenzūruotų stebėjimų trukmė lygi tyrimo trukmei.

Pavyzdžiui, šešioms pelėms nebuvo panaudoti kancerogenai prieš išvirkščiant auglių ląsteles į jų pėdas. Šiuo atveju tiriamas auglio pasirodymo laikas. Tyrinėtojai nusprendžia baigti eksperimentą po trisdešimties savaičių. 2.1. paveikslėlis vaizduoja auglio išsivystymo laiką:



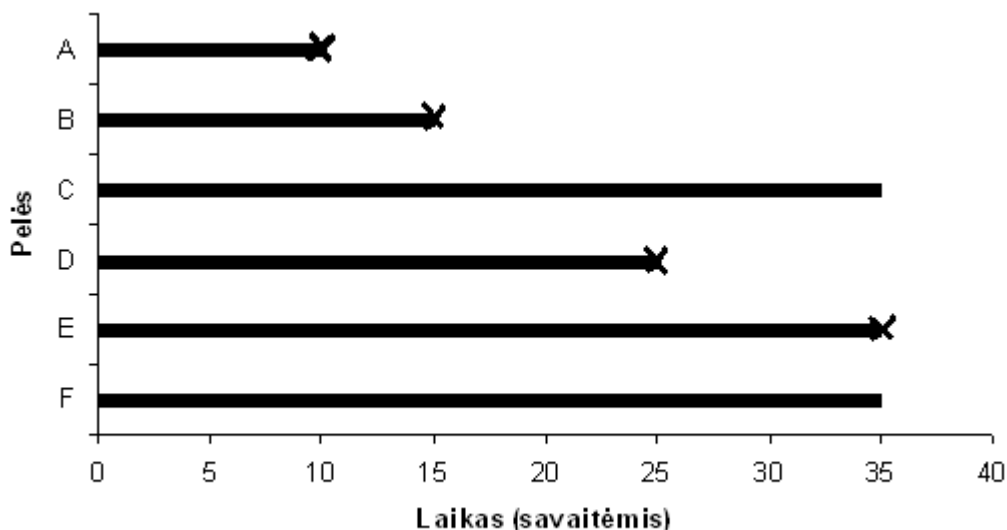
2.1.pav. I tipo cenzūravimas

Pelėms A, B ir D auglys išsivystė po 10, 15 ir 25 savaičių atitinkamai. C, E ir F pelėms auglys neišsivystė iki tyrimo pabaigos, todėl dažnai yra žymima 30+, o tai reiškia, kad joms auglys vystėsi ilgiau nei trisdešimt savaičių. Tada tiriamų pelių išgyvenamumo duomenys yra tokie: 10, 15, 30+, 25, 30+, 30+ savaičių.

2.2. II tipo cenzūravimas

Tyrinėjant gyvūnus galima stebėti juos tol, kol nugaiš tam tikras, fiksuotas gyvūnų skaičius, tarkime 80 iš 100, o išgyvenę gyvūnai toliau nebebus stebimi. Tokiu atveju turėsime II tipo imties cenzūravimą. Cenzūruotų stebėjimų skaičius bus lygus 20 šiuo atveju.

Pavyzdžiui, atliekant aprašytą tyrimą su šešiomis pelėmis, mokslininkai gali nuspręsti baigti (nutraukti) stebėjimus, kai auglys išsivystys keturioms iš šešių pelių. Tuomet išgyvenimo trukmė arba laikas iki auglio išsivystymo toms šešioms pelėms bus tokie: 10, 15, 35+, 25, 35, 35+ savaičių. Aprašytą atvejį iliustruoja 2.2. paveikslėlis:

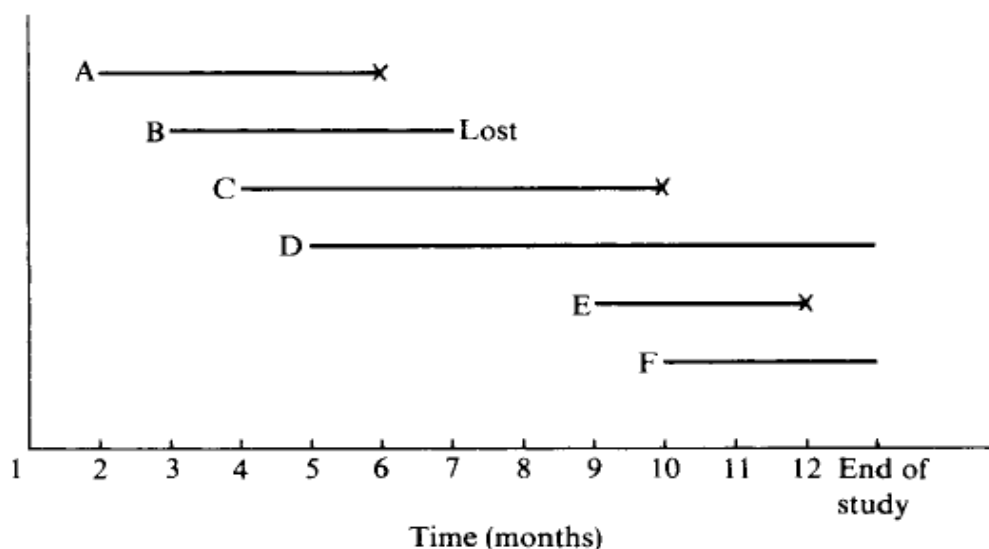


2.2.pav. II tipo cenzūravimas

2.3. III tipo cenzūravimas

Daugumoje klinikinių ir epidemiologinių tyrimų stebėjimų periodo trukmė yra fiksuota, o ligoniai įtraukiami į tyrimą skirtingu laiku. Kai kurie ligoniai gali mirti dar nepasibaigus tyrimams, taigi jų išgyvenimo trukmė bus žinoma. Kiti ligoniai gali išvykti, pasitraukti iš tyrimo ir tęsti jų stebėjimo jau bus nebeįmanoma. Taip pat gali būti ir tokių ligonių, kurie lieka gyvi tyrimo laikui pasibaigus. „Prarastų“ ligonių išgyvenimo trukmė trunka nuo jų įėjimo į tyrimą iki paskutinio kontakto su jais. Išgyvenusių ligonių išgyvenimo trukmė yra laikomas laikas nuo jų įsitraukimo į tyrimą iki tyrimo pabaigos. Pastarosios dvi stebėjimų rūšys yra cenzūruoti stebėjimai. Kadangi ligoniai gali įsitraukti į tyrimą skirtingu laiku, jų cenzūravimo momentai taip pat yra skirtingi. Toks cenzūravimas vadinamas III tipo cenzūravimu.

Pavyzdžiui, sakykime, ūminiu kraujo vėžiu (leukemija) sergantys ligoniai gali įsitraukti į klinikinį tyrimą per visą jo vykdymo trukmę, t.y. metus. Tarkime, kad visi šeši ligoniai teigiamai reaguoja į gydymą ir pasveiksta. Aprašytą situaciją iliustruoja 2.3. paveikslėlis:



2.3.pav. III tipo cenzūravimas

Ligoniai A, C ir E pradėjo sveikti nuo antro, ketvirto ir devinto mėnesių pradžios ir išgijo po keturių, šešių ir trijų mėnesių atitinkamai. Ligonis B pradėjo sveikti nuo trečio mėnesio, bet po keturių mėnesių jo stebėjimai nutrūko, todėl jo gydymo trukmė yra bent jau keturi mėnesiai. Ligoniai D ir F pradėjo sveikti po penkto ir dešimto mėnesio atitinkamai, tačiau šis procesas tęsėsi tyrimui pasibaigus. Jų gydymo trukmė yra mažiausiai aštuoni ir trys mėnesiai atitinkamai. Taigi kiekvieno iš tirtų šešių ligonių sveikimo trukmė yra atitinkamai 4, 4+, 6, 8+, 3, 3+ mėnesių.

I ir II tipų cenzūruoti stebėjimai dar yra vadinami vienpusio cenzūravimo duomenimis, o III tipo – progresuojančiais (tebesitęsiančiais) cenzūruotais duomenimis pagal Cohen (1965 m.). Dažnai III tipo cenzūravimas dar vadinamas atsitiktiniu cenzūravimu. Visi minėti trijų tipų cenzūravimai vadinami cenzūravimu iš dešinės.

Dar pasitaiko ir imties cenzūravimas iš kairės bei intervalinis cenzūravimas. Cenzūravimas iš kairės būna tada, kai yra žinoma, jog tiriamas įvykis prasideda prieš tam tikrą laiko momentą t , bet tikslus to įvykio pražios laikas nežinomas. Pavyzdžiui, epidemiologai nori sužinoti, kokiam amžiuje žmonės suserga cukriniu diabetu. Tyrimo pradžioje, penkiasdešimtmečiams tyrimo dalyviams jau buvo nustatyta cukraligė, bet įrašų jų medicininėse kortelėse apie tai, kada jie susirgo šia liga, nėra. Taigi tokiame tyrime 50 m. amžius yra cenzūruotas iš kairės. Tai reiškia, kad tokiems ligoniams ligos diagnozavimo pradžia yra vėliausiai 50 m. amžius.

Intervalinis imties cenzūravimas būna tada, kai yra žinoma, kad tiriamas įvykis įvyksta tarp laiko momentų a ir b . Pavyzdžiui, medicininiai įrašai rodo, kad 45 m. amžiaus ligonis, aprašytas aukščiau, nesirgo cukralige. Vadinasi amžius, kuriame diagnozuojama minėta liga, yra amžius tarp 45 – 50 m.

2.4. Cenzūravimas iš dešinės

Kaip jau buvo minėta, visi aukščiau aprašyti trijų tipų cenzūravimai vadinami cenzūravimu iš dešinės. Matematinis tokio duomenų cenzūravimo aprašymas aprašomas taip:

Tarkime, kad eksperimento metu stebima n objektų. Šių objektų mirčių momentai (angl. failure times) modeliuojami kaip nepriklausomi, vienodai pasiskirstę (n. v. p.) atsitiktiniai dydžiai T_1, T_2, \dots, T_n . Sąvoka „mirties momentas“ suprantama labai bendrai – tai gali būti gyvo organizmo gyvenimo trukmė nuo gimimo, susirgimo ar operacijos iki mirties; gaminio funkcionavimo trukmė nuo jo pagaminimo arba pardavimo iki gedimo arba kokio nors defekto atsiradimo; laikas nuo bedarbio užsiregistravimo darbo biržoje iki konkretaus darbo pasiūlymo ir pan.

Sakykime, kad eksperimento metu mirties momento T_i reikšmė sužinoma tik tada, kai $T_i \leq C_i$, čia $C_i \geq 0$ yra atsitiktinis dydis arba konstanta, vadinama cenzūravimo iš dešinės momentu. Priešingu atveju a.d. T_i reikšmė nežinoma, bet tada žinoma, kad ji yra didesnė už žinomą cenzūravimo momento C_i reikšmę.

Pažymime $X_i = T_i \wedge C_i$, $\delta_i = 1_{\{T_i \leq C_i\}}$.

Atsitiktinis dydis X_i yra i-tojo objekto stebėjimo pabaigos momentas. Atsitiktinis dydis δ_i yra įvykio $T_i \leq C_i$ indikatorius, t.y. nurodo, ar X_i yra mirties momentas T_i ($\delta_i = 1$), ar cenzūravimo momentas C_i ($\delta_i = 0$).

Tada atsitiktinių vektorių rinkinys

$$(X_1, \delta_1), \dots, (X_n, \delta_n)$$

vadinamas cenzūruotąja iš dešinės imtimi.

Pavyzdžiui, stebimi n ligonių, kuriems momentais t_1, \dots, t_n atlikta tam tikra operacija. Statistinės analizės tikslas – momentu $t > \max t_i$ padaryti išvadas apie ligonių gyvenimo trukmę po operacijos. Jei i-tasis ligonis miršta prieš momentą t , tai žinoma jo gyvenimo trukmė nuo operacijos iki mirties T_i . Jei jis dar gyvas momentu t , tai jo mirties momentas nežinomas, bet žinoma, kad ligonis po operacijos išgyveno laiką, ilgesnį už žinomą konstantą $C_i = t - t_i$. Taigi šie duomenys aprašomi cenzūruotąja iš dešinės imtimi.

Cenzūruotosios iš dešinės imtys klasifikuojamos taip:

1. Jei $C_i = t, i = 1, \dots, n$, tai imtis vadinama I tipo cenzūruotąja iš dešinės imtimi. Taigi ši imtis gaunama, kai visų dar gyvų (funkcionuojančių) objektų stebėjimas momentu t nutraukiamas.

2. Jei $C_i = T_{(r)}, i = 1, \dots, n, r \leq n$, kur $T_{(r)}$ - r-oji pozicinė statistika, tai imtis vadinama II tipo cenzūruotąja iš dešinės imtimi. Tokios imtys dažniausiai naudojamos atliekant patikimumo eksperimentus, kai bandoma n gaminių tol, kol r iš jų sugenda.
3. Jei $C_1, \dots, C_n, T_1, \dots, T_n$ yra nepriklausomi a.d., tai imtis vadinama atsitiktinai cenzūruotąja imtimi. Atskiru atveju C_1, \dots, C_n gali būti konstantos. Pavyzdžiui, ligoniams operacijos gali būti atliktos (arba gaminiai gali būti pradėti bandyti) momentais t_1, \dots, t_n , o mirtingumo (gedimų) duomenys gali būti pradėti analizuoti momentu $t, t > \max t_i$. Tada cenzūravimo momentai $C_i = t - t_i$ yra neatsitiktiniai. Kartais toks cenzūravimas dar vadinamas daugkartiniu arba progresuojančiu. I tipo cenzūravimas taip pat yra atskiras atsitiktinio cenzūravimo atvejis.

2.5. Nupjovimas iš kairės

Tarkime, kad tiriamas sergančiųjų tam tikra liga vilniečių mirtingumas po susirgimo. Laiko nuo susirgimo pradžios iki mirties T pasiskirstymo funkciją pažymėkime $F(t)$.

Kalendoriniu momentu t pradedami stebėti sergantys tam tikra liga vilniečiai. Pažymėkime D_i i-ojo ligonio ligos trukmę momentu t . Tie vilniečiai, kurie mirė iki momento t , apskritai neįtraukiami į imtį. Bėgant laikui fiksuojami į imtį įjungtų ligonių mirčių momentai T_1, \dots, T_n . Atsitiktinio dydžio T_i skirstinys žinant D_i sutampa su sąlyginiu T skirstiniu, žinant, kad $T > D_i$, t.y. a.d. T_i sąlyginė pasiskirstymo funkcija yra $F(t | T > D_i)$.

Kitas pavyzdys. Tiriamas skirstinys atstumo T, kurį nuvažiuoja 2006 metais pagaminti tam tikro modelio automobiliai, kol stabdžių kaladėlė nusidėvi iki kritinio lygio. Gamintojas atsitiktinai parenka partiją automobilių, kurių stabdžių kaladėlės dar nepakeistos, t.y. jų susidėvėjimas neviršija kritinio lygio. Fiksuojamas atstumas D_i , kurį jau nuvažiavo automobilis ir toliau eksperimentas atliekamas, kol kaladėlė nusidėvi. Taigi į partiją nepatenka automobiliai, kurių stabdžių kolonėlės jau susidėvėjo. Į imtį įtraukto i-ojo automobilio stabdžių kaladėlės funkcionavimo trukmę žymėsime T_i . Atsitiktinio dydžio T_i skirstinys, žinant D_i , sutampa su sąlyginiu T skirstiniu, žinant, kad $T > D_i$, t.y. a.d. T_i sąlyginė pasiskirstymo funkcija yra $F(t | T > D_i)$.

Mirčių momentų T_i (mirčių momentai antrame pavyzdyje yra stabdžių kaladėlių susidėvėjimo momentai) ir amžiaus D_i eksperimento pradžios momentu t (pirmame eksperimente tai laikas nuo susirgimo, antrame – pravažiuotas atstumas iki momento t) poros

$$(T_1, D_1), \dots, (T_n, D_n)$$

sudaro nupjautąją iš kairės imtį.

A.d. T_i sąlyginė pasiskirstymo funkcija žinant D_i yra $F(t|T > D_i)$.

Nupjovimas iš kairės skiriasi nuo cenzūravimo iš kairės tuo, kad nupjovimo atveju objektai, kuriems $T \leq D_i$, apskritai neįjungiami į imtį, o cenzūravimo atveju apie šiuos objektus žinoma, kad jų mirtis įvyksta prieš žinomą momentą E_i .

2.6. Nupjovimas iš kairės ir cenzūravimas iš dešinės

Tarkime, tęsiamas vieno iš aukščiau pateiktų pavyzdžių nagrinėjimas. Jei kalendoriniu momentu t pradėjus tirti tam tikra liga sergančius vilniečius, kalendoriniu momentu $t_i > t$ reikia padaryti išvadas apie jų mirtingumą, tai iki šio momento gauti duomenys ne tik nupjauti iš kairės, bet ir cenzūruoti iš dešinės. i -ojo imties nario cenzūravimo momentas (jis skaičiuojamas nuo ligos pradžios) $C_i = t_i - t + D_i$. Taigi nupjautoji iš kairės ir cenzūruotoji iš dešinės imtis turi pavidalą

$$(X_1, D_1, \delta_1), \dots, (X_n, D_n, \delta_n)$$

kur $X_i = T_i \wedge C_i$, $\delta_i = 1_{\{T_i \leq C_i\}}$. Atsitiktinio dydžio T_i sąlyginė pasiskirstymo funkcija žinant D_i yra $F(t|T > D_i)$, kur F yra mirties momento T besąlyginė pasiskirstymo funkcija.

3. Funkcijos

Išgyvenimo trukmė matuojama kaip laikas iki tam tikro tiriamo įvykio, pavyzdžiui, iki kokios nors nesėkmės, mirties, reakcijos į gydymą, recidyvo (relapse). Taip pat tas laikas gali būti kokios nors ligos vystymosi periodas, kalėjimo laikas iki ligtinio paleidimo ar vedybų trukmė iki skyrybų.

Išgyvenimo trukmės skirstinys dažniausiai yra aprašomas ir charakterizuojamas tokiomis trimis funkcijomis:

- 1) Išgyvenamumo funkcija (the survivorship or survival function);
- 2) Tikimybinio tankio funkcija (the probability density function);
- 3) Nesėkmių rizikos funkcija (the hazard function).

Šios trys funkcijos yra ekvivalenčios, t.y. jei duota tik viena iš jų, kitas galima lengvai surasti.

Praktikoje šios trys funkcijos gali būti naudojamos pavaizduoti skirtingus duomenų aspektus (bruožus, charakteristikas). Esminė išgyvenamumo analizės problema yra įvertinti vieną ar daugiau tiriamos imties funkcijų ir padaryti išvadas apie visos populiacijos išgyvenamumą.

3.1. Išgyvenamumo (survivorship) funkcija

Tarkime, kad T reiškia išgyvenimo trukmę.

Išgyvenamumo funkcija, žymima $S(t)$, reiškia tikimybę, kad individas išgyvena ilgiau nei laiką t :

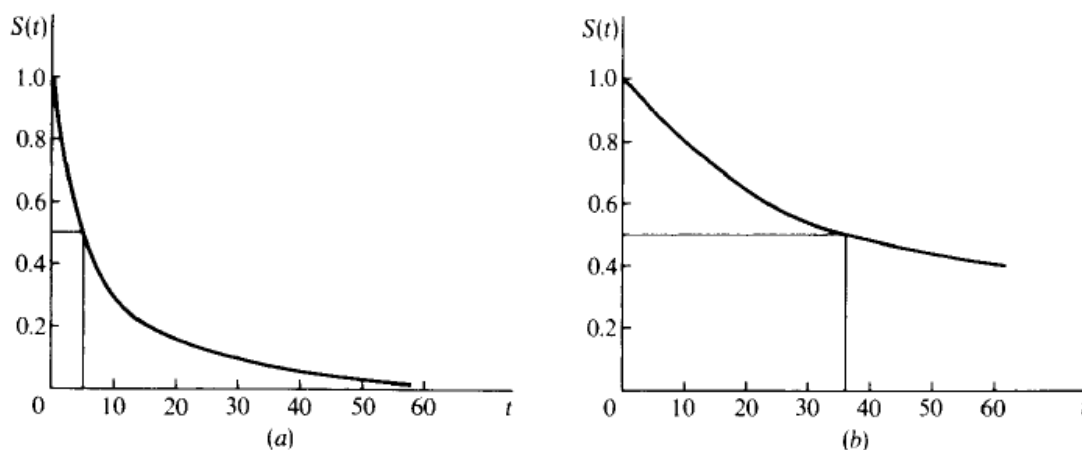
$$S(t) = P(\text{individas išgyvena ilgiau nei } t) = P(T > t) \quad (3.1.1)$$

Iš pasiskirstymo funkcijos $F(t) = P(T < t)$ apibrėžimo seka, kad

$$S(t) = 1 - P(\text{individas mirs iki laiko momento } t) = 1 - F(t) \quad (3.1.2)$$

Išgyvenamumo tikimybė lygi vienetui, kai laikas lygus nuliui, ir lygi nuliui, kai laikas t artėja į begalybę.

Berkson (1942 m.) pasiūlė grafinę funkcijos $S(t)$ interpretaciją, kuri vadinama išgyvenamumo kreive. Funkciją $S(t)$ iliustruoja 3.1.1. paveikslėlis:



3.1.1.pav. Išgyvenamumo kreivė

Beveik stati $S(t)$ dalis paveikslėlyje (a) vaizduoja trumpą išgyvenimo trukmę. Plokštesnė išgyvenamumo kreivė paveikslėlyje (b) vaizduoja ilgesnę išgyvenimo trukmę.

Išgyvenamumo funkcija arba išgyvenamumo kreivė gali būti panaudojamos apskaičiuoti 50-tą procentilį (medianą) ir kitus procentilius, pvz., 25-tą, 75-tą procentilius, kurie atitinkamai vadinami pirmu ir trečiu kvartiliais. Taip pat šios dvi charakteristikos naudojamos dviejų ar daugiau grupių skirstinių palyginimui. Paveikslėlyje 3.1.1.(a) mediana yra apytiksliai 5-asis laiko

vienetas, o paveikslėlyje (b) – apytiksliai 36-tas laiko vienetas. Vidurkis dažniausiai naudojamas nusakyti centrinę skirstinio tendenciją, tačiau išgyvenamumo analizėje mediana dažniau naudojama nei vidurkis, nes dėl mažo individų su itin ilga ar trumpa išgyvenimo trukme skaičiaus, išgyvenimo trukmės vidurkis tampa neproporcingai didelis arba mažas.

Praktikoje, kai nėra cenzūruotų stebėjimų, išgyvenamumo funkcija yra įvertinama kaip ligonių, išgyvenusių ilgiau nei iki laiko momento t , santykis su visų ligonių skaičiumi:

$$\hat{S}(t) = \frac{l}{n}, \text{ kur} \quad (3.1.3)$$

l - ligonių, išgyvenusių ilgiau nei iki laiko momento t ,
 n - visų ligonių skaičius.

$\hat{S}(t)$ yra išgyvenamumo funkcijos įvertis.

Kai imtyje pasitaiko cenzūruotų stebėjimų, formulė (3.1.3) jau netinka įverčiui $\hat{S}(t)$ rasti.

Pavyzdžiui, tarkime, kad turime tokią imtį: 4, 6, 6+, 10+, 15, 20. Pasinaudoję (3.1.3) formule, galime apskaičiuoti, pavyzdžiui, $\hat{S}(5) = \frac{5}{6} \approx 0.8333$. Tačiau negalime įvertinti $\hat{S}(11)$, nes tikslus ligonių, išgyvenusių ilgiau nei iki laiko momento $t = 11$, skaičius nėra žinomas. Tik galima spėti, kad trečiasis ir ketvirtasis stebėti ligoniai (6+ ir 10+ atitinkamai) galėjo išgyventi ilgiau arba trumpiau nei iki laiko momento $t=11$.

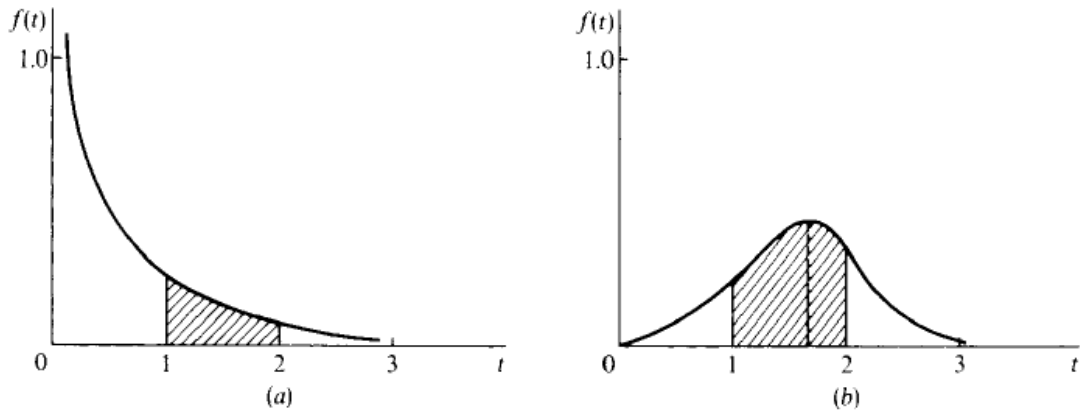
Taigi kai imtyje pasitaiko cenzūruotų stebėjimų, formulė (3.1.3) jau netinkama, norint rasti $S(t)$ įvertį. Tokiu atveju $S(t)$ įverčiu imamas Kaplano – Mejerio įvertis.

3.2. Tikimybinio tankio (density) funkcija

Kaip ir bet koks atsitiktinis dydis išgyvenimo trukmė T turi tikimybinio tankio funkciją. Ji apibrėžiama kaip tikimybės individui mirti trumpame laiko intervale $(t, t + \Delta t]$ ir intervalo ilgio Δt santykis. Arba trumpiau – tai nesėkmių tikimybės trumpame laiko intervale ir to intervalo ilgio santykio riba, t.y.:

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{P[t < T \leq t + \Delta t]}{\Delta t} \quad (3.2.1)$$

Funkcija $f(t)$ vadinama tikimybinio tankiu. Tokių kreivių pavyzdžiai pateikiami 3.2.1. paveiksle:



3.2.1.pav. Tikimybinio tankio kreivės

Tankio funkcija turi tokias savybes:

- 1) $f(t)$ yra neneigiama funkcija, t.y. $f(t) \geq 0$, kai $t \geq 0$ ir $f(t) = 0$, kai $t < 0$.
- 2) Plotas, ribojamas tankio kreivės ir t ašies, lygus vienetui.

Praktikoje, kai imtyje nėra cenzūruotų stebėjimų, tikimybinio tankio funkcija $f(t)$ įvertinama kaip mirusių ligonių skaičiaus intervale $[t, t + \Delta t]$ santykis su to intervalo ilgiu Δt , t.y.:

$$f(t) = \frac{d}{n \cdot \Delta t}, \text{ kur} \quad (3.2.2)$$

d - ligonių, mirusių intervale, kurio pradžia t , skaičius,

n - visų ligonių skaičius,

Δt - intervalo ilgis.

Panašiai, kaip ir skaičiuojant $S(t)$ įvertį, kai imtyje pasitaiko cenzūruotų stebėjimų, formulės (3.2.2) negalima taikyti tikimybinio tankio $f(t)$ įverčio radimui.

Ligonių, kurie miršta bet kokiame laiko intervale, skaičius gali būti randamas, naudojant tankio funkciją. Kreivė 3.2.1.pav.(a) vaizduoja aukštą nesėkmių dažnį tyrimų pradžioje ir mažėjantį nesėkmių dažnį, bėgant laikui. To paties paveikslėlio (b) kreivės vaizduojamas aukščiausias nesėkmių dažnis yra apytiksliai 1,7 - tuoju laiko momentu. Santykis individų, kurie miršta tarp 1-o ir 2-o laiko momentų, yra lygus užbrūkšniuotam plotui, kurį riboja tankio kreivė ir ašis t . Tankio funkcija dar vadinama besąlyginiu mirčių (nesėkmių) intensyvumu.

3.3. Nesėkmių intensyvumo (hazard) funkcija

Ši funkcija $h(t)$ proporcinga nesėkmių tikimybei labai trumpame laiko intervale tarp individų, kurie yra dar gyvi intervalo pradžioje. Arba ji suprantama kaip riba santykio tikimybės kad individas miršta arba patiria nesėkmę labai trumpame laiko intervale $(t, t + \Delta t]$, kai yra žinoma, kad jis išgyveno iki laiko momento t , bei intervalo ilgio:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P[t < T \leq t + \Delta t | T > t]}{\Delta t} \quad (3.3.1)$$

Funkcija $h(t)$ taip pat priklauso nuo pasiskirstymo $F(t)$ bei tikimybinio tankio $f(t)$ funkcijų:

$$h(t) = \frac{f(t)}{1 - F(t)} \quad (3.3.2)$$

Ši funkcija dažnai dar vadinama momentiniu nesėkmių dažniu, mirtingumo galia, sąlyginiu mirtingumo dažniu arba nuo amžiaus priklausančiu nesėkmių dažniu. Pavyzdžiui, jeigu formulėje (3.3.1) t yra amžius, tai ta formulė traktuojama kaip polinkio į nesėkmę matas, t.y. kaip individo amžiaus funkcija, laikant, kad dydis $\Delta t \cdot h(t)$ yra tikėtina individų, kurie mirs trumpame laiko intervale $(t, t + \Delta t]$, dalis. Vadinasi nesėkmių intensyvumo funkcija charakterizuoja nesėkmių (mirties) riziką. Ši funkcija yra labai svarbi išgyvenamumo analizėje.

Praktikoje, kai nėra cenzūruotų stebėjimų, nesėkmių intensyvumo funkcija $h(t)$ įvertinama kaip santykis ligonių, mirstančių tam tikrame laiko intervale, kai yra žinoma, kad jie išgyveno iki to laiko intervalo pradžios:

$$\hat{h}(t) = \frac{d}{m \cdot \Delta t}, \text{ kur} \quad (3.3.3)$$

d - mirusių ligonių skaičius intervale, prasidedančiame laiko momentu t .

m - išgyvenusių ligonių skaičius iki laiko momento t .

Δt - intervalo ilgis.

Statistikos specialistai dažnai naudoja nesėkmių dažnio vidurkį intervale, kuriame mirusių ligonių skaičius per laiko intervalą yra suskirstytas į mažesnius intervalus pagal išgyvenusių ligonių skaičių intervalo vidurio taške:

$$\hat{h}(t) = \frac{d}{m - \frac{c}{2}}, \text{ kur} \quad (3.3.4)$$

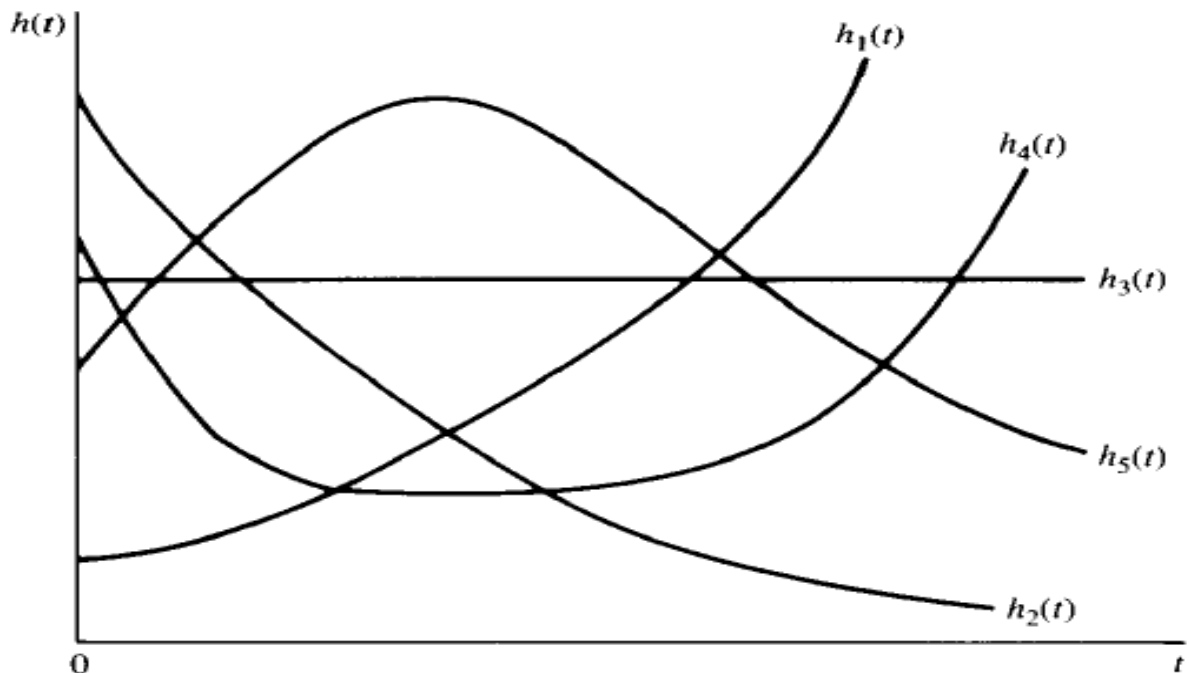
d - mirusių ligonių skaičius per laiko intervalą.

m - išgyvenusių ligonių skaičius iki laiko momento t .

c - cenzūruotų duomenų skaičius intervale.

Statistinis nesėkmių dažnio įvertis, apskaičiuotas, naudojantis formule (3.3.4), yra didesnis nei apskaičiuotas, naudojantis formule (3.3.3).

Funkcija $h(t)$ gali didėti, mažėti, būti pastovi arba vaizduoti kokį nors sudėtingesnę procesą. Tokių funkcijų pavyzdžiai pateikiami 3.3.1. paveiksle:



3.3.1.pav. Nesėkmių funkcijų pavyzdžiai

Pavyzdžiui, ligoniai, sergantys ūmia leukemija (kraujo vėžiu) ir kuriems gydymas neturi jokio poveikio (įtakos), turi didėjantį nesėkmių dažnį. Funkcijos $h_1(t)$, $h_2(t)$ yra mažėjančios ir vaizduoja, pavyzdžiui, išgyvenamumą kareivių, kurie buvo sužeisti ir vėliau išoperuoti. Pagrindinis pavojus yra operacija, tačiau tas pavojus mažėja, kai operacija pavyksta. Pastovios nesėkmių funkcijos pavyzdys yra funkcija $h_3(t)$, kuri vaizduoja, pavyzdžiui, sveikų 18 – 40 metų amžiaus asmenų pavojų mirti nelaimingų atsitikimų atvejais. „Vonios“ pavidalo funkcija $h_5(t)$ vaizduoja žmonių gyvenimo procesą. Pradiniu gyvenimo periodu nesėkmių rizika yra didelė (didelis kūdikių mirtingumas). Vėliau funkcija $h(t)$ tampa apytiksliai pastovi iki tam tikro laiko, kai ji pradeda didėti dėl nesėkmių rizikos didėjimo. Kitas pavyzdys - ligonių, sergančių tuberkulioze, nesėkmių rizika iš pradžių didėja, bet vėliau, po gydymo, mažėja. Tokia didėjanti, o vėliau mažėjanti nesėkmių rizikos funkcija yra funkcija $h_5(t)$.

Suminė (suvestinė) nesėkmių rizikos funkcija apibūdinama taip:

$$H(t) = \int_0^t h(x) dx \quad (3.3.5)$$

Vėliau bus parodyta, kad

$$H(t) = -\ln S(t) \quad (3.3.6)$$

Tarkime, kad $t=0$, tada $S(t)=1$, o $H(t)=0$. O kai $t=\infty$, tada $S(t)=0$, $H(t)=\infty$. Taigi funkcija $H(t)$ gali įgyti bet kokią reikšmę iš intervalo $[0; \infty]$.

4. Funkcijų sąryšiai

Trys funkcijos, apibrėžtos trečiame skyriuje, yra matematiškai ekvivalenčios, t.y. jei yra duota viena iš minėtų funkcijų, kitos dvi randamos labai paprastai. Pavyzdžiui:

- 1) Iš formulės (3.1.2) $S(t) = 1 - F(t)$ ir (3.3.2) $h(t) = \frac{f(t)}{1 - F(t)}$, gauname, kad

$$h(t) = \frac{f(t)}{S(t)} \quad (4.1)$$

Šis santykis taip pat gaunamas iš (3.3.1) formulės, pritaikius sąlyginės tikimybės apibrėžimą.

- 2) Kadangi tankio funkcija yra pasiskirstymo funkcijos išvestinė, be to, $F(t) = 1 - S(t)$ iš formulės (3.1.2), tai

$$f(t) = \frac{d}{dt}[1 - S(t)] = -S'(t) \quad (4.2)$$

- 3) Įstatę formulės (4.2) išraišką į formulę (4.1), gauname

$$h(t) = -\frac{S'(t)}{S(t)} = -\frac{d}{dt} \ln S(t) \quad (4.3)$$

- 4) Suintegravę funkciją $h(t)$ iš formulės (4.3) nuo 0 iki t , bei laikydami, kad $S(0) = 1$, turime

$$-\int_0^t h(x) dx = \ln S(t)$$

arba

$$H(t) = -\ln S(t)$$

t.y.
$$S(t) = \exp\{-H(t)\} = \exp\left\{-\int_0^t h(x) dx\right\} \quad (4.4)$$

5) Iš formulių (4.1) ir (4.4) gauname

$$f(t) = h(t) \cdot \exp\{-H(t)\} \quad (4.5)$$

Be to, jei yra duota funkcija $f(t)$, išgyvenamumo funkcija $S(t)$ gaunama iš pagrindinio sąryšio tarp funkcijų $f(t)$ ir $F(t)$ bei iš formulės (3.1.2). O funkcija $h(t)$ šiuo atveju gaunama iš (4.1) formulės.

Kai yra žinoma funkcija $S(t)$, $f(t)$ ir $h(t)$ gaunamos iš formulių (4.2) ir (4.1) atitinkamai. Arba $h(t)$ iš pradžių galima gauti iš (4.3) formulės, o tada $f(t)$ gauti iš (4.1).

Jei žinome funkciją $h(t)$, $S(t)$ ir $f(t)$ gaunamos atitinkamai iš formulių (4.4) ir (4.5).

Taigi, kaip jau buvo minėta anksčiau, jei turima viena iš trijų išgyvenamumo funkcijų, kitos dvi gaunamos labai paprastai. Ši teiginį iliustruoja 4.1. pavyzdys:

4.1. Pavyzdys. Tarkime, kad populiacijos išgyvenamumo laikas turi tokią tankio funkciją:

$$f(t) = \lambda e^{-\lambda t}, \quad t \geq 0.$$

Pasinaudoję pasiskirstymo funkcijos apibrėžimu, gauname

$$F(t) = \int_0^t f(x) dx = \lambda \int_0^t e^{-\lambda x} dx = -e^{-\lambda x} \Big|_0^t = 1 - e^{-\lambda t}$$

Iš formulės (3.1.2) $S(t) = 1 - F(t)$, gauname

$$S(t) = 1 - F(t) = 1 - 1 + e^{-\lambda t} = e^{-\lambda t}$$

Tada funkcija $h(t)$ gali būti gauta iš formulės (4.1):

$$h(t) = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda$$

4.2. Pavyzdys. Tarkime, duota: $h(t) = c$. Reikia rasti funkcijas $S(t)$ ir $f(t)$.

Pasinaudoję trečiame ir ketvirtame skyriuose išvestomis formulėmis, gauname

$$S(t) = \exp\{-H(t)\} = \exp\left\{-\int_0^t c dx\right\} = \exp\{-c \cdot t + c \cdot 0\} = \exp\{-ct\} = e^{-ct},$$

$$f(t) = h(t) \cdot \exp\{-H(t)\} = c \cdot \exp\{-ct\} = c \cdot e^{-ct}.$$

4.3. Pavyzdys. Tarkime, yra duota tokia išgyvenamumo funkcija $S(t) = \exp\left\{-\left(\frac{t}{\theta}\right)^\gamma\right\}$.

Rasime $f(t)$ ir $h(t)$ funkcijas.

Vėl pasinaudoję formulėmis iš trečio ir ketvirto skyrių, gauname

$$f(t) = -S'(t) = -\exp\left\{-\left(\frac{t}{\theta}\right)^\gamma\right\} \cdot (-\gamma t^{\gamma-1}) = \gamma t^{\gamma-1} \cdot \exp\left\{-\left(\frac{t}{\theta}\right)^\gamma\right\},$$

$$h(t) = \frac{f(t)}{S(t)} = \frac{\gamma t^{\gamma-1}}{\theta^\gamma}.$$

Parametrinė informantinė funkcija.

Parametrinė informantinė funkcija gaunama Didžiausiojo tikėtinumo (DT) metodu. Cenzūravimo iš dešinės atveju mirčių momentų tankių sandauga vadinama tikėtinumo funkcija, kuri turi tokią formą:

$$L(\theta) = \prod_{i=1}^n f^{\delta_i}(X_i, \theta) S^{1-\delta_i}(X_i, \theta).$$

Pasinaudojus sąryšiu $f(t, \theta) = \lambda(t, \theta)S(t, \theta)$, tikėtinumo funkcija užrašoma taip:

$$L(\theta) = \prod_{i=1}^n \lambda^{\delta_i}(X_i, \theta) S(X_i, \theta)$$

Įvertinys $\hat{\theta}$, maksimizuojantis tikėtinumo funkciją $L(\theta)$, vadinamas didžiausiojo tikėtinumo įvertiniu.

Tikėtinumo funkcijos logaritmas $l(\theta) = \sum_{i=1}^n \delta_i \ln \lambda(X_i, \theta) + \sum_{i=1}^n \ln S(X_i, \theta)$ maksimizuojamas tame pačiame taške kaip ir tikėtinumo funkcija.

Jei funkcija $\lambda(u, \theta)$ pakankamai glodi, tai parametro θ DT įvertinys tenkina tokią lygtį:

$$\dot{l}(\hat{\theta}) = 0,$$

kur \dot{l} yra toks informančių vektorius:

$$\dot{l}(\hat{\theta}) = \left(\frac{\partial}{\partial \theta_1} l(\theta), \dots, \frac{\partial}{\partial \theta_m} l(\theta) \right)^T$$

Nupjautos iš kairės ir cenzūruotos iš dešinės imties atveju tikėtinumo funkcija turės tokį pavidalą:

$$L(\theta) = \prod_{i=1}^n \left(\frac{f(X_i, \theta)}{S(D_i, \theta)} \right)^{\delta_i} \left(\frac{S(X_i, \theta)}{S(D_i, \theta)} \right)^{1-\delta_i} = \prod_{i=1}^n \lambda^{\delta_i}(X_i, \theta) \frac{S(X_i, \theta)}{S(D_i, \theta)}.$$

5. COX'o modelis

Išgyvenamumo analizės praktikoje dažnai pasitaiko atveju, kai tikimybinis skirstinys yra nežinomas, taigi tokiais atvejais neįmanoma pritaikyti atitinkamo modelio turimų duomenų analizei. Be to, parametrinių metodų naudojimas nustatant prognozuojančius veiksnius taip pat dažnai yra ribotas dėl tam tikrų sąlygų.

Tokiais atvejais dažniausiai naudojamas Cox'o (1972 m.) santykinis nesėkmių intensyvumo modelis. Taikant šį modelį, nebūtina žinoti imties pasiskirstymo funkcijos formos. Nesėkmių intensyvumo funkcija šiame modelyje gali turėti bet koki pavidalą. Laikoma, kad individų, stebimų prie skirtingų kovariančių, nesėkmių intensyvumo funkcijos yra proporcingos ir jų santykis nepriklauso nuo laiko. Šiame modelyje įprastinė tikėtinumo funkcija pakeičiama daline. Svarbu pastebėti, kad statistinės išvados, padarytos, remiantis daline tikėtinumo funkcija, yra labai panašios į išvadas, padarytas, remiantis įprastine tikėtinumo funkcija.

Taigi visiškai neatsitiktinai Cox'o regresijos modelis tapo dažniausiai naudojamu modeliu, atliekant išgyvenamumo duomenų regresinę analizę. Kaip jau buvo minėta, taikant šį modelį, nebūtina žinoti išgyvenamumo laiko pasiskirstymo funkcijos, taip pat modelis gali būti taikomas turint laike kintančias kovariantes. Modelis taikomas ne tik esant diskrečioms, bet taip pat ir tolydžioms kovariantėms. Taip pat galimas ir duomenų nupjovimas iš kairės.

5.1. Dalinė tikėtinumo funkcija

Kaip jau buvo minėta, Cox'o modelyje atskirų individų nesėkmių intensyvumo funkcijos yra proporcingos, pavyzdžiui, $\frac{h(t | x_1)}{h(t | x_2)} = c$. Šis funkcijų santykis dviejų individų, turinčių pastovius laike prognozuojančius veiksnius arba kovariantes $x_1 = (x_{11}, x_{21}, \dots, x_{p1})'$ ir $x_2 = (x_{12}, x_{22}, \dots, x_{p2})'$, yra pastovus, t.y. nepriklauso nuo laiko momento t . Tai reiškia, kad santykis tarp dviejų individų, esant skirtingoms kovariantėms, rizikos mirti yra toks pat ir nesvarbu, kaip ilgai jie išgyvena. Ši savybė reiškia, kad, kokį laiko momentą beimtume, nesėkmių intensyvumo funkcija, pavaizduota kaip kovariančių $x = (x_1, x_2, \dots, x_p)'$ ir laiko t funkcija, gali būti užrašyta kaip sandauga bazinės $h_0(t)$ nesėkmių intensyvumo funkcijos ir kovariančių funkcijos $g(x_1, \dots, x_p)$, t.y.

$$h(t | x_1, \dots, x_p) = h_0(t) \cdot g(x_1, \dots, x_p) \quad \text{arba} \quad h(t | x) = h_0(t) \cdot g(x) \quad (5.1.1)$$

Funkcija $g(x)$ parodo kovariančių vaidmenį (įtaką) nesėkmių intensyvumui. Funkcija $h_0(t)$ gali būti interpretuojama kaip nesėkmių intensyvumo funkcija, kai ignoruojamos visos kovariantės arba kai $g(x) = 1$. Todėl ji ir vadinama bazine nesėkmių intensyvumo funkcija. Dviejų individų, turinčių skirtingas kovariantes x_1 ir x_2 , nesėkmių santykis yra:

$$\frac{h(t | x_1)}{h(t | x_2)} = \frac{h_0(t) \cdot g(x_1)}{h_0(t) \cdot g(x_2)} = \frac{g(x_1)}{g(x_2)} \quad (5.1.2)$$

Šis santykis yra pastovus, nepriklausomai nuo laiko.

Cox'o proporcingų nesėkmių intensyvumų modelyje laikoma, kad $g(x)$ iš formulės (5.1.1) yra eksponentinė kovariančių funkcija, t.y.

$$g(x) = \exp \left\{ \sum_{j=1}^p b_j \cdot x_j \right\} = \exp \{ b' \cdot x \}$$

o nesėkmių intensyvumo funkcija tada yra tokia:

$$h(t | x) = h_0(t) \cdot \exp \left\{ \sum_{j=1}^p b_j \cdot x_j \right\} = h_0(t) \cdot \exp \{ b' \cdot x \} \quad (5.1.3)$$

kur $b = (b_1, \dots, b_p)^T$ yra kovariančių koeficientai. Šiuos koeficientus galima įvertinti, turint konkrečius duomenis. Jie parodo atitinkamų kovariančių svarbą. Pavyzdžiui, tarkime, kad turima tik viena kovariantė. Apibrėžkime: $x_1 = 0$, jei ligonis buvo gydomas Placebo (vaistu, kuris neturi jokie poveikio – nei teigiamo, nei neigiamo) ir $x_1 = 1$, jei ligonis buvo gydomas eksperimentiniais vaistais. Tada nesėkmių santykis ligonio, gydyto eksperimentiniais vaistais, ir ligonio, gydyto Placebo, naudojantis (5.1.2) ir (5.1.3) formulėmis, yra

$$\frac{h(t | x_1 = 1)}{h(t | x_1 = 0)} = \exp \{ b_1 \}$$

Beje, gydymų, taikant skirtingus vaistus, poveikis yra vienodas jei $b_1 = 0$ ir eksperimentiniai vaistai turi mažesnę (didesnę) pavojų numirti nei Placebo, jei $b_1 < 0$ ($b_1 > 0$).

Formulė (5.1.3) yra ekvivalenti tokiai formulei:

$$S(t | x) = [S_0(t)]^{\exp \left(\sum_{j=1}^p b_j x_j \right)} = [S_0(t)]^{\exp(b' \cdot x)}, \text{ kur} \quad (5.1.4)$$

$$S_0(t) = \exp \left\{ - \int_0^t h_0(u) du \right\}$$

Taigi kovariantės gali būti įtraukiamos į išgyvenamumo funkciją. Formulės (5.1.3) panaudojimo pavyzdžiai gali būti tokie:

- 1) Dviejų imčių uždaviniai. Tarkime, kad $p = 1$, t.y. turime tik vieną kovariantę x_1 , kuri yra tokia funkcija, t.y. indikatorius:

$$x_{1i} = \begin{cases} 0, & \text{jei } i\text{-tasis individas priklauso grupei 0,} \\ 1, & \text{jei } i\text{-tasis individas priklauso grupei 1.} \end{cases}$$

Tada iš formulės (5.1.3) grupių 0 ir 1 nesėkmių intensyvumo funkcijos atitinkamai yra $h_0(t)$ ir $h_1(t) = h_0(t) \cdot \exp(b_1)$. Kaip matome, grupės 1 nesėkmių intensyvumo funkcija lygi grupės 0 nesėkmių intensyvumo funkcijai, padaugintai iš konstantos $\exp(b_1)$, t.y. šios dvi funkcijos yra proporcingos. Pasinaudoję išgyvenamumo funkcija, gauname, kad

$$S(t) = [S_0(t)]^c, \text{ kur } c = \exp(b_1).$$

Norint patikrinti hipotezę apie grupių išgyvenamumo funkcijų lygybę, reikia patikrinti hipotezę $H : c = 1$, kuri yra ekvivalenti hipotezei $H : b_1 = 0$.

- 2) Dviejų imčių uždaviniai su kovariantėmis. Kovariantės iš (5.1.3) formulės gali būti indikatoriniai kintamieji kaip x_1 iš 1) atvejo bei prognozuojantys veiksniai. Kai turime vieną ar daugiau kovariančių, kurios yra prognozuojantys veiksniai formulėje (5.1.1), galime nagrinėti santykį tarp dviejų grupių.

Be kovariantės x_1 yra ir kitų kovariančių, pavyzdžiui, x_2, x_3, \dots, x_p .

$S(t | x) = [S_0(t)]^{\exp\{b_1 x_1 + b_2 x_2 + b_3 x_3 + \dots + b_p x_p\}}$. Faktorius (kovariantė) x_1 neturi įtakos išgyvenamumui tada ir tik tada (t.t.t.), kai $b_1 = 0$.

- 3) Regresiniai uždaviniai. Abi formulės (5.1.3) puses padalinę iš $h_0(t)$ bei logaritmavę lygybę, gauname

$$\ln \frac{h_i(t)}{h_0(t)} = b_1 x_{1i} + b_2 x_{2i} + \dots + b_p x_{pi} = \sum_{j=1}^p b_j \cdot x_{ji} = b' \cdot x_i \quad (5.1.5)$$

kur x_i yra i -tojo individo kovariantė. Kairioji (5.1.5) formulės pusė yra nesėkmių intensyvumo santykis (arba santykinė rizika), o dešinioji pusė yra tiesinė kovariančių ir jų koeficientų funkcija.

Kaip buvo minėta anksčiau, $h_0(t)$ yra nesėkmių intensyvumo funkcija, kai ignoruojamos visos kovariantės. Jei kovariantės yra standartizuotos (normuotos) pagal vidurkį, tada naudojamas toks modelis

$$\ln \frac{h_i(t)}{h_0(t)} = b_1 (x_{1i} - \bar{x}_1) + b_2 (x_{2i} - \bar{x}_2) + \dots + b_p (x_{pi} - \bar{x}_p) = b' (x_i - \bar{x}) \quad (5.1.6)$$

kur $\bar{x}' = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_p)$, o \bar{x}_j yra visų ligonių j -tosios kovariantės vidurkis.

Naudojant formulę (5.1.5), svarbu nustatyti lemiančius prognozuojančius veiksnus. T.y. iš turimų p kovariančių reikia išskirti poaibį tų kovariančių, kurios labiau įtakoja nesėkmių intensyvumą, ir galiausiai, reikia nustatyti ligonių išgyvenimo trukmę. Čia reikia remtis regresijos koeficientais. Jei $b_i = 0$, tai atitinkama kovariantė neturi įtakos išgyvenamumui. Jei $b_i \neq 0$, tai šis koeficientas parodo kovariantės x_i įtakos rizikai svarbą, kai kitos kovariantės yra fiksuotos.

Koeficientų b_1, b_2, \dots, b_p įvertinimui Cox'as (1972 m.) pasiūlė naudoti dalinę tikėtinumo funkciją, remiantis sąlyginėmis nesėkmių tikimybėmis, laikant, kad nėra pasikartojančių išgyvenamumo trukmės reikšmių. Tačiau praktikoje dažnai pasitaiko pasikartojančių išgyvenamumo trukmės reikšmių, todėl Cox'o pasiūlyta dalinė tikėtinumo funkcija buvo modifikuota, kad apimtų pasikartojančių reikšmių atvejį. Todėl įdomu apskaičiuoti įverčius, atsižvelgiant į pasikartojančias reikšmes ir neatsižvelgiant į jas, bei gautus rezultatus palyginti.

5.2. Reikšmingų kovariančių nustatymas

Svarbiausias dalykas tyrimuose yra reikšmingų prognozuojančių veiksnių arba kovariančių nustatymas. Tam naudojami hipotezių tikrinimo testai ir kovariančių atrankos procesai. Cox'o modelis nuo kitų parametrinių metodų skiriasi tuo, kad šis modelis naudoja dalinę tikėtinumo funkciją, kurioje vieninteliai parametrai yra kovariančių koeficientai. Statistinės išvados, padarytos remiantis daline tikėtinumo funkcija, yra panašios į išvadas, padarytas remiantis įprasta tikėtinumo funkcija. Taigi šis modelio supaprastinimas didelės įtakos rezultatams neturi, t.y. neiškreipia rezultatų.

6. Proporcingų mirimų intensyvumų ir Persikertančių mirimų intensyvumų modeliai

Tarkime $S_x(t)$, $\lambda_x(t)$ ir $\Lambda_x(t)$ yra išgyvenamumo, nesėkmių intensyvumo ir sukaupto nesėkmių intensyvumo funkcijos su m -mačiu paaiškinančiu kintamuoju x .

Tikrinama tokia hipotezė:

$$H_0 : \lambda_x(t) = e^{\beta^T x} \lambda(t)$$

Proporcingų mirimų intensyvumų (angl. Generalized proportional hazards (GPH)) ir Persikertančių mirimų intensyvumų (angl. Simple cross-effects (SCE)) modeliai remiasi proporcingų nesėkmių intensyvumų (angl. Proportional hazards) modeliu.

GPH modelio nesėkmių intensyvumo funkcija yra tokia

$$\lambda_x(t) = e^{\beta^T x} (1 + \Lambda_x(t))^{-\gamma+1} \lambda(t) = e^{\beta^T x} \left(1 + \gamma e^{\beta^T x} \Lambda(t)\right)^{\frac{1}{\gamma}-1} \lambda(t), \quad (6.1)$$

o SCE modelio atveju nesėkmių intensyvumo funkcija turi tokį pavidalą

$$\lambda_x(t) = e^{\beta^T x} \{1 + \Lambda_x(t)\}^{1-e^{\gamma T x}} \lambda(t) = e^{\beta^T x} \left\{1 + e^{(\beta+\gamma)^T x} \Lambda(t)\right\}^{e-\gamma^T x-1} \lambda(t), \quad \text{kur} \quad (6.2)$$

$$\Lambda(t) = \int_0^t \lambda(u) du.$$

Parametras γ yra vienmatis GPH modelyje ir m -matis SCE modelyje, parametras δ yra m -matis abiejuose modeliuose.

PH modelis yra atskiras atvejis GPH ir SCE modelių atvejis, kai $\gamma=1$ (GPH), $\gamma=0$ (SCE).

Būdingos minėtų modelių savybės yra tokios:

GPH – nesėkmių rizikos santykio dažnis didėja, mažėja arba yra pastovus, nesėkmių rizikos dažnis ir išgyvenamumo funkcija nekerta intervalo $(0, \infty)$.

SCE – nesėkmių rizikos santykio dažnis didėja, mažėja arba yra pastovus, nesėkmių rizikos dažnis ir išgyvenamumo funkcija nekerta arba tik kartą kerta intervalą $(0, \infty)$.

Nagrinėjama cenzūruotoji iš dešinės imtis. Kaip jau buvo rašyta 2.4. skyrelyje, cenzūruotoji iš dešinės imtis žymima taip:

$$(X_1, \delta_1, x_1), \dots, (X_n, \delta_n, x_n), \quad \text{kur}$$

$$X_i = T_i \wedge C_i, \delta_i = 1_{\{T_i \leq C_i\}}, i = 1, \dots, n$$

T_i yra mirčių momentai, C_i - cenzūravimo momentai.

Cenzūruoti iš dešinės duomenys gali būti užrašyti ir taip:

$$(N_1(t), Y_1(t), x_1, t \geq 0), \dots, (N_n(t), Y_n(t), x_n, t \geq 0), \quad \text{kur}$$

$$N_i(t) = 1_{\{x_i \leq t, \delta_i = 1\}}, Y_i(t) = 1_{\{x_i \geq t\}}$$

1_A žymi įvykio A indikatorių. Taip pat

$$N(t) = \sum_{i=1}^n N_i(t), Y(t) = \sum_{i=1}^n Y_i(t)$$

Tarkime, kad išgyvenamumo skirstiniai visų n objektų žinant x_i yra absoliučiai tolydūs ir $S_i(t)$ yra išgyvenamumo funkcijos, $\lambda_i(t)$ - mirimų intensyvumo funkcijos.

Taip pat tariame, kad teisingas multiplikatyvių intensyvumų modelis: kompensatoriai skaičiuojančio proceso N_i atžvilgiu istorijos yra $\int Y_i \lambda_i du$.

Žemiau aprašomo sukaupto mirimų intensyvumo Λ įverčio išraiškoje naudojama funkcija

$$S^{(0)}(t, \beta) = \sum_{i=1}^n Y_i(t) e^{\beta^T x_i}$$

Su žinomu λ nežinomas baigtinio matavimo parametras θ apima parametrus β, γ arba β, γ, δ (pagal 6.1 ir 6.2 lygybes). Parametro θ maksimalaus tikėtimumo įvertis $\hat{\theta}$ tikrinamas šia lygybe:

$$U(\hat{\theta}) = 0, \text{ kur}$$

U yra tokia taškų funkcija:

$$U(\theta) = \sum_{i=1}^n \int_0^{\infty} \frac{\partial}{\partial \theta} \ln \lambda_i(t, \theta) \{dN_i(t) - Y_i(t) \lambda_i(t, \theta) dt\},$$

t.y. nežinomas parametras θ vertinamas DT metodu, sprendžiant lygčių sistemą $\dot{l}(\hat{\theta}) = 0$ β, γ parametrų atžvilgiu, kur \dot{l} yra informantinė funkcija.

Minėtuose GPH ir SCE modeliuose parametrinė taškų (score) funkcija U pakeičiama

sukaupto nesėkmių intensyvumo Λ įverčiu pagal Breslow $\hat{\Lambda} = \sum_{i: \delta_i = 1, X_i \leq t} \frac{1}{S^{(0)}(X_i, \beta)}$,

regresijos parametras β pakeičiamas parametro įverčiu $\hat{\beta}$, parametrai γ ir δ atitinkamai keičiami į $\gamma = 1$ (GPH modelyje) arba $\gamma = 0$ (SCE modelyje). Taip gaunamos $(m+1)$ -o matavimo ir $2m$ -o matavimo statistikos atitinkamai. Pirmosios m šių statistikų komponentės yra lygios nuliui. Taigi testams apskaičiuoti yra naudojama vienmatė (GPH atveju) arba m -matė (SCE atveju) \bar{U} statistika, kuri turi tokį pavidalą

$$\bar{U} = \sum_{i=1}^n \int_0^{\infty} \{h(x_i, t, \hat{\beta}) - E_*(t, \hat{\beta})\} dN_i(t), \text{ kur}$$

$$h(x_i, t, \beta) = \begin{cases} -\ln \left(1 + e^{\beta^T x_i} \Lambda(t) \right) & (\text{GPH atveju}); \\ -x_i \ln \left(1 + e^{\beta^T x_i} \Lambda(t) \right) & (\text{SCE atveju}); \end{cases} \quad (6.3)$$

$$E_*(t, \beta) = \frac{S_*^{(1)}(t, \beta)}{S^{(0)}(t, \beta)} \quad (6.4)$$

$$S_*^{(1)}(t, \beta) = \begin{cases} -\sum_{l=1}^n Y_l(t) e^{\beta^T x_l} \ln \left(1 + e^{\beta^T x_l} \Lambda(t) \right) & (\text{GPH atveju}); \\ -\sum_{l=1}^n x_l Y_l(t) e^{\beta^T x_l} \ln \left(1 + e^{\beta^T x_l} \Lambda(t) \right) & (\text{SCE atveju}). \end{cases}$$

Gali būti parodyta, kad

$$T = n^{-1} \bar{U}^T \bar{D}^{-1} \bar{U} \xrightarrow{D} \chi^2(k),$$

(k=1 GPH atveju ir k=m SCE atveju), kur

$$\bar{D} = \hat{\Sigma}^{**} - \hat{\Sigma}^{*T} \hat{\Sigma}_0^{-1} \hat{\Sigma}^* \quad (6.5)$$

$$\hat{\Sigma}^* = n^{-1} \int_0^\infty V_*(u, \beta) dN(u), \quad \hat{\Sigma}^{**} = n^{-1} \int_0^\infty V_{**}(u, \beta) dN(u),$$

kai j=0, $\hat{\Sigma}_0$ gaunama iš lygybės

$$\hat{\Sigma}_j = n^{-1} \sum_{i: \delta_i=1} V(X_i, \beta) \Lambda^j(X_i), \quad (j=0, 1, 2, \dots) \quad (6.6)$$

$$V_*(u, \beta) = \frac{S_*^{(2)}(u, \beta)}{S^{(0)}(u, \beta)} - E(u, \beta) E_*^T(u, \beta), \quad (6.7)$$

$$V_{**}(u, \beta) = \frac{S_{**}^{(2)}(u, \beta)}{S^{(0)}(u, \beta)} - E_*^{\otimes 2}(u, \beta) \quad (6.8)$$

$$S_*^{(2)}(u, \beta) = \begin{cases} -\sum_{l=1}^n x_l Y_l(t) e^{\beta^T x_l} \ln \left(1 + e^{\beta^T x_l} \Lambda(t) \right) & (\text{GPH atveju}); \\ -\sum_{l=1}^n x_l^{\otimes 2} Y_l(t) e^{\beta^T x_l} \ln \left(1 + e^{\beta^T x_l} \Lambda(t) \right) & (\text{SCE atveju}); \end{cases}$$

$$S_{**}^{(2)}(u, \beta) = \begin{cases} -\sum_{l=1}^n x_l Y_l(t) e^{\beta^T x_l} \ln^2(1 + e^{\beta^T x_l} \Lambda(t)) & \text{(GPH atveju);} \\ -\sum_{l=1}^n x_l^{\otimes 2} Y_l(t) e^{\beta^T x_l} \ln^2(1 + e^{\beta^T x_l} \Lambda(t)) & \text{(SCE atveju).} \end{cases}$$

Testo statistika užrašoma tokia forma, patogiai skaičiavimams:

$$T = n^{-1} \bar{U}^T \bar{D}^{-1} \bar{U},$$

$$\bar{U} = \sum_{i:\delta_i=1} \left\{ h(x_i, X_i, \beta) - E_*(X_i, \beta) \right\}, \text{ kur}$$

h ir E_* randami iš (6.3) ir (6.4) lygybių atitinkamai, \bar{D} randama iš (6.5), kur $\hat{\Sigma}_0$ gaunama iš (6.6), kai $j=0$, ir

$$\hat{\Sigma}_* = n^{-1} \sum_{i:\delta_i=1} V_*(X_i, \beta), \quad \hat{\Sigma}_{**} = n^{-1} \sum_{i:\delta_i=1} V_{**}(X_i, \beta), \text{ kur}$$

V_* ir V_{**} yra išreikštos (6.7) ir (6.8) lygybėmis.

Nulinė hipotezė yra atmetama su reikšmingumo lygmeniu α jeigu $T > \chi_{1-\alpha}^2(k)$, ($k=1$ GPH modelio alternatyvos atveju ir $k=m$ SCE modelio alternatyvos atveju).

7. Kriterijus Cox'o modelio adekvatumui tikrinti, naudojant nupjautus iš kairės ir cenzūruotus iš dešinės duomenis

Kaip jau buvo minėta 2.6 skyrelyje, nupjovimo iš kairės ir cenzūravimo iš dešinės atveju nagrinėjama tokia imtis:

$$(\tilde{X}_i, D_i, \delta_i, x^{(i)}), i = 1, \dots, n,$$

kur $x^{(i)}$ yra m-matės kovariantės, o

$$\tilde{X}_i = X_i \wedge C_i, \quad \delta_i = 1\{X_i \leq C_i\}, \quad P\{X_i \leq t | D_i, x^{(i)}\} = F_T(t | T > D_i, x^{(i)}).$$

$$\checkmark \text{žymėsime } N_i(t) = 1\{\tilde{X}_i \leq t, \delta_i = 1\}, \quad N = \sum_{i=1}^n N_i(t), \quad Y_i(t) = 1\{D_i < t \leq \tilde{X}_i\}, \quad Y = \sum_{i=1}^n Y_i(t).$$

Mirimų intensyvumas, žinant $x^{(i)}$, persikertančių mirimų intensyvumų (SCE) modelio atveju yra

$$\lambda_i(t, \theta) = e^{\beta^T x^{(i)}} \{1 + e^{(\beta+\gamma)^T x^{(i)}} \Lambda(t)\} e^{\gamma^T x^{(i)} - 1} \Lambda(t), \quad \theta = (\beta, \gamma)^T, \quad \Lambda(t) = \int_0^t \lambda(u) du. \quad (*)$$

Jei Λ žinomas, tai parametrinė informacinė funkcija yra

$$\dot{l}_\Lambda(\theta) = \sum_{i=1}^n \int_0^\infty \frac{\partial}{\partial \theta} \ln \lambda_i(t, \theta) \{dN_i(t) - Y_i(t) \lambda_i(t, \theta) dt\} \quad (1)$$

Reikia patikrinti hipotezę, kad galioja Cox'o modelis: $H : \lambda_i(t; \theta) = e^{\beta^T x^{(i)}} \lambda(t)$. Ši hipotezė yra ekvivalenti hipotezei: $H : \gamma = 0$ SCE modelio (*) ribose.

Cox'o modelio atveju parametras β vertinamas maksimizuojant dalinio tikėtinumo funkciją, iš kurios gaunama informantinė funkcija (t.y. tikėtinumo funkcijos logaritmo išvestinė):

$$\dot{l}(\beta) = \sum_{i=1}^n \int_0^{\infty} [x^{(i)} - E(t, \beta)] dN_i(t) \quad (2)$$

$$\text{kur } E(t; \beta) = \frac{S^{(1)}(t, \beta)}{S^{(0)}(t, \beta)}, \quad S^{(0)}(t, \beta) = \sum_{i=1}^n Y_i(t) e^{\beta^T x^{(i)}}, \quad S^{(1)}(t, \beta) = \sum_{i=1}^n x^{(i)} e^{\beta^T x^{(i)}} Y_i(t).$$

Radus įvertį $\hat{\beta}$, randamas nežinomas funkcijos $\Lambda(t)$ Breslou įvertinys

$$\hat{\Lambda}(t) = \hat{\Lambda}(t, \hat{\beta}) \quad (3)$$

$$\hat{\Lambda}(t, \hat{\beta}) = \int_0^t \frac{dN(u)}{S^{(0)}(u, \hat{\beta})} \quad (4)$$

Įvertinio $\hat{\beta}$ asimptotinės savybės yra:

$$\text{Žymėkime } M_i = N_i(t) - \int_0^t \tau_i(u) \lambda_i(u, \beta) du.$$

$$\text{Tada } \frac{1}{\sqrt{n}} \dot{l}(\beta) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^{\infty} (x^{(i)} - E(t, \beta)) dM_i(t),$$

$$\sqrt{n}(\hat{\beta} - \beta) = \left(\frac{1}{n} \int_0^{\infty} V(t, \beta) dN(t) \right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^{\infty} (x^{(i)} - E(t, \beta)) dM_i(t) + o_p(1), \text{ čia}$$

$$V(t, \beta) = \frac{S^{(2)}(t, \beta)}{S^{(0)}(t, \beta)} - E^{\otimes 2}(t, \beta), \quad S^{(2)}(t, \beta) = \sum_{i=1}^n (x^{(i)})^{\otimes 2} e^{\beta^T x^{(i)}} Y_i(t).$$

$$\text{Pažymėję } \hat{\Sigma}_o = \frac{1}{n} \int_0^{\infty} V(t, \beta) dN(t), \text{ gauname}$$

$$\sqrt{n}(\hat{\beta} - \beta) = (\hat{\Sigma}_o)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^{\infty} (x^{(i)} - E(t, \beta)) dM_i(t) + o_p(1) \quad (5)$$

Kriterijaus konstravimo idėja yra tokia:

(1) išraiškoje $\theta = (\beta, \gamma)^T$ keičiamas į $(\hat{\beta}, 0)^T$. Čia $\hat{\beta}$ - parametro β įvertinys Cox'o modelyje, o Λ keičiamas į Breslou įvertinį (3).

Taigi gauname

$$\ln \lambda_i(t; \theta) = \beta^T x^{(i)} + (e^{-\gamma^T x^{(i)}} - 1) \ln \{1 + e^{(\beta + \gamma)^T x^{(i)}} \Lambda(t)\}$$

$$\begin{cases} \frac{\partial}{\partial \beta} \ln \lambda_i(t; \theta) = x^{(i)} + (e^{-\gamma^T x^{(i)}} - 1) \frac{x^{(i)} e^{(\beta+\gamma)^T x^{(i)}} \Lambda(t)}{1 + e^{(\beta+\gamma)^T x^{(i)}} \Lambda(t)}, \\ \frac{\partial}{\partial \gamma} \ln \lambda_i(t; \theta) = -x^{(i)} e^{-\gamma^T x^{(i)}} \ln\{1 + e^{(\beta+\gamma)^T x^{(i)}} \Lambda(t)\} + (e^{-\gamma^T x^{(i)}} - 1) \frac{x^{(i)} e^{(\beta+\gamma)^T x^{(i)}} \Lambda(t)}{1 + e^{(\beta+\gamma)^T x^{(i)}} \Lambda(t)}, \\ \frac{\partial}{\partial \beta} \ln \lambda_i(t; \theta)_{|\theta=(\hat{\beta}, o)^T, \Lambda=\hat{\Lambda}} = x^{(i)}, \\ \frac{\partial}{\partial \gamma} \ln \lambda_i(t; \theta)_{|\theta=(\hat{\beta}, o)^T, \Lambda=\hat{\Lambda}} = -x^{(i)} \ln\{1 + e^{\hat{\beta}^T x^{(i)}} \hat{\Lambda}(t)\} \end{cases},$$

$\lambda_i(t; \theta) dt_{|\theta=(\hat{\beta}, o)^T, \Lambda=\hat{\Lambda}} = e^{\hat{\beta}^T x^{(i)}} d\hat{\Lambda}(t)$, todėl įstatę visas išraiškas į (1), gauname

$$\dot{\tilde{l}}_{\hat{\Lambda}}(\hat{\beta}, 0) = \left(\sum_{i=1}^n \int_0^{\infty} (x^{(i)})^T \{dN_i(t) - Y_i(t) e^{\hat{\beta}^T x^{(i)}} d\hat{\Lambda}(t)\}, - \sum_{i=1}^n \int_0^{\infty} (x^{(i)})^T \ln\{1 + e^{\hat{\beta}^T x^{(i)}} \hat{\Lambda}(t)\} \{dN_i(t) - Y_i(t) e^{\hat{\beta}^T x^{(i)}} d\hat{\Lambda}(t)\} \right)^T$$

Pastebėtina, kad pirmosios $\dot{\tilde{l}}_{\hat{\Lambda}}(\hat{\beta}, 0)$ komponentės vektorius yra lygus nuliniam vektoriui:

$$\begin{aligned} \sum_{i=1}^n \int_0^{\infty} x^{(i)} \{dN_i(t) - Y_i(t) e^{\hat{\beta}^T x^{(i)}} d\hat{\Lambda}(t)\} &= \sum_{i=1}^n \int_0^{\infty} x^{(i)} dN_i(t) - \int_0^{\infty} \frac{\sum_{i=1}^n x^{(i)} Y_i(t) e^{\hat{\beta}^T x^{(i)}}}{S^{(0)}(t, \hat{\beta})} dN(t) = \\ &= \sum_{i=1}^n \int_0^{\infty} x^{(i)} dN_i(t) - \int_0^{\infty} E(t, \hat{\beta}) dN(t) = \sum_{i=1}^n \int_0^{\infty} [x^{(i)} - E(t, \hat{\beta})] dN_i(t) = \dot{\tilde{l}}(\hat{\beta}) = 0. \end{aligned}$$

Antrosios $\dot{\tilde{l}}_{\hat{\Lambda}}(\hat{\beta}, 0)$ komponentės vektorių žymėsime \hat{U} , be to dar žymėsime

$$h(x^{(i)}; t, \hat{\beta}) = -x^{(i)} \ln[1 + e^{\hat{\beta}^T x^{(i)}} \hat{\Lambda}(t)].$$

Taigi gauname $\hat{U} = \sum_{i=1}^n \int_0^{\infty} h(x^{(i)}; t, \hat{\beta}) dN_i(t) - \int_0^{\infty} \frac{\sum_{i=1}^n h(x^{(i)}; t, \hat{\beta}) Y_i(t) e^{\hat{\beta}^T x^{(i)}}}{S^{(0)}(t, \hat{\beta})} dN(t)$, t.y.

$$\hat{U} = \sum_{i=1}^n \int_0^{\infty} \{h(x^{(i)}; t, \hat{\beta}) - E_*(t, \hat{\beta})\} dN_i(t), \text{ kur } E_*(t, \hat{\beta}) = \frac{S_*^{(1)}(t, \hat{\beta})}{S^{(0)}(t, \hat{\beta})},$$

$$S_*^{(1)} = \sum_{i=1}^n h(x^{(i)}; t, \hat{\beta}) Y_i(t) e^{\hat{\beta}^T x^{(i)}}.$$

Panagrinėkime statistikos \hat{U} asimptotinį elgesį, kai $n \rightarrow \infty$ ir teisingas Cox'o modelis.

Cox'o modelio atveju $dN_i(t) = e^{\beta^T x^{(i)}} Y_i(t) d\Lambda(t) + dM_i(t)$, kur M_i yra martingalas atžvilgiu

filtracijos, generuotos Y_i ir N_i , taigi

$$\hat{U} = \sum_{i=1}^n \int_0^{\infty} \{h(x^{(i)}; t, \hat{\beta}) - E_*(t, \hat{\beta})\} e^{\beta^T x^{(i)}} Y_i(t) d\Lambda(t) + \sum_{i=1}^n \int_0^{\infty} \{h(x^{(i)}; t, \hat{\beta}) - E_*(t, \hat{\beta})\} e^{\beta^T x^{(i)}} Y_i(t) dM_i(t) =$$

$$\begin{aligned}
&= \sum_{i=1}^n \int_0^{\infty} \{h(x^{(i)}; t, \hat{\beta}) - E_*(t, \hat{\beta})\} e^{\hat{\beta}^T x^{(i)}} Y_i(t) d\Lambda(t) + \sum_{i=1}^n \int_0^{\infty} \{h(x^{(i)}; t, \hat{\beta}) - E_*(t, \hat{\beta})\} \{e^{\hat{\beta}^T x^{(i)}} - e^{\beta^T x^{(i)}}\} Y_i(t) d\Lambda(t) + \\
&+ \sum_{i=1}^n \int_0^{\infty} \{h(x^{(i)}; t, \hat{\beta}) - E_*(t, \hat{\beta})\} dM_i(t) = - \sum_{i=1}^n \int_0^{\infty} \{h(x^{(i)}; t, \hat{\beta}) - E_*(t, \hat{\beta})\} \{e^{\hat{\beta}^T x^{(i)}} - e^{\beta^T x^{(i)}}\} Y_i(t) d\Lambda(t) + \\
&+ \sum_{i=1}^n \int_0^{\infty} \{h(x^{(i)}; t, \hat{\beta}) - E_*(t, \hat{\beta})\} dM_i(t).
\end{aligned}$$

$$(6) \frac{1}{\sqrt{n}} \hat{U} = - \frac{1}{n} \sum_{i=1}^n \int_0^{\infty} \{h(x^{(i)}; t, \hat{\beta}) - E_*(t, \hat{\beta})\} e^{\hat{\beta}^T x^{(i)}} Y_i(t) (x^{(i)})^T d\Lambda(t) \sqrt{n} (\hat{\beta} - \beta) +$$

$$+ \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^{\infty} \{h(x^{(i)}; t, \hat{\beta}) - E_*(t, \hat{\beta})\} dM_i(t) + o_p(1), \text{ nes}$$

$$\begin{aligned}
\sqrt{n} (e^{\hat{\beta}^T x^{(i)}} - e^{\beta^T x^{(i)}}) &= e^{\beta^T x^{(i)}} \sqrt{n} (\hat{\beta} - \beta)^T x^{(i)} + o_p(1) = e^{\beta^T x^{(i)}} (x^{(i)})^T \sqrt{n} (\hat{\beta} - \beta)^T + o_p(1) = \\
&= e^{\hat{\beta}^T x^{(i)}} (x^{(i)})^T \sqrt{n} (\hat{\beta} - \beta) + o_p(1).
\end{aligned}$$

Pirmojo nario integralą užrašome taip:

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^n \int_0^{\infty} \{h(x^{(i)}; t, \hat{\beta}) - E_*(t, \hat{\beta})\} e^{\hat{\beta}^T x^{(i)}} Y_i(t) (x^{(i)})^T d\Lambda(t) &= \frac{1}{n} \int_0^{\infty} \left\{ S_*^{(2)}(t, \hat{\beta}) - E_*(t, \hat{\beta}) [S^{(1)}(t, \hat{\beta})]^T \right\} d\Lambda(t) = \\
&= \frac{1}{n} \int_0^{\infty} \left[\frac{S_*^{(2)}(t, \hat{\beta})}{S^{(0)}(t, \hat{\beta})} - E_*(t, \hat{\beta}) E^T(t, \hat{\beta}) \right] S^{(0)}(t, \hat{\beta}) d\Lambda(t) = \frac{1}{n} \int_0^{\infty} V_*(t, \hat{\beta}) dN(t) + o_p(1),
\end{aligned}$$

$$\text{čia } V_*(t, \hat{\beta}) = \frac{S_*^{(2)}(t, \hat{\beta})}{S^{(0)}(t, \hat{\beta})} - E_*(t, \hat{\beta}) E^T(t, \hat{\beta}),$$

$$S_*^{(2)} = \sum_{i=1}^n h(x^{(i)}; t, \hat{\beta}) (x^{(i)})^T Y_i(t) e^{\hat{\beta}^T x^{(i)}} = - \sum_{i=1}^n (x^{(i)})^{\otimes 2} \ln[1 + e^{\hat{\beta}^T x^{(i)}} \Lambda(t)] Y_i(t) e^{\hat{\beta}^T x^{(i)}}.$$

$$\text{Pažymėkime } \hat{\Sigma}_* = \frac{1}{n} \int_0^{\infty} V_*(t, \hat{\beta}) dN(t).$$

Naudojamės (5) išraišką:

$$\sqrt{n} (\hat{\beta} - \beta) = \hat{\Sigma}_0^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^{\infty} (x^{(i)} - E(t, \hat{\beta})) dM_i(t) + o_p(1).$$

Tada (6) išraišką užrašome tokiu pavidalu:

$$\frac{1}{\sqrt{n}} \hat{U} = - \hat{\Sigma}_* \hat{\Sigma}_0^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^{\infty} (x^{(i)} - E(t, \hat{\beta})) dM_i(t) + \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^{\infty} \{h(x^{(i)}; t, \hat{\beta}) - E_*(t, \hat{\beta})\} dM_i(t) + o_p(1).$$

$$\begin{aligned}
\text{Numatoma variacija } \left\langle \frac{1}{\sqrt{n}} \hat{U} \right\rangle &= \hat{\Sigma}_* \hat{\Sigma}_0^{-1} \frac{1}{n} \sum_{i=1}^n \int_0^\infty [x^{(i)} - E(t, \hat{\beta})]^{\otimes 2} e^{\beta^T x^{(i)}} Y_i(t) d\Lambda(t) \hat{\Sigma}_0^{-1} \hat{\Sigma}_*^T - \\
&- \hat{\Sigma}_* \hat{\Sigma}_0^{-1} \frac{1}{n} \sum_{i=1}^n \int_0^\infty [x^{(i)} - E(t, \hat{\beta})] [h(x^{(i)}; t, \hat{\beta}) - E_*(t, \hat{\beta})]^T e^{\beta^T x^{(i)}} Y_i(t) d\Lambda(t) + \\
&+ \frac{1}{n} \sum_{i=1}^n \int_0^\infty \{h(x^{(i)}; t, \hat{\beta}) - E_*(t, \hat{\beta})\} \{x^{(i)} - E(t, \hat{\beta})\}^T e^{\beta^T x^{(i)}} Y_i(t) d\Lambda(t) \hat{\Sigma}_0^{-1} \hat{\Sigma}_*^T + \\
&+ \frac{1}{n} \sum_{i=1}^n \int_0^\infty \{h(x^{(i)}; t, \hat{\beta}) - E_*(t, \hat{\beta})\}^{\otimes 2} e^{\beta^T x^{(i)}} Y_i(t) d\Lambda(t) + o_p(1).
\end{aligned}$$

$$\begin{aligned}
\text{Kadangi } \frac{1}{n} \sum_{i=1}^n \int_0^\infty \{h(x^{(i)}; t, \hat{\beta}) - E_*(t, \hat{\beta})\}^{\otimes 2} e^{\beta^T x^{(i)}} Y_i(t) d\Lambda(t) &= \\
&= \frac{1}{n} \int_0^\infty \left(S^{(2)}(t, \hat{\beta}) - S^{(1)}(t, \hat{\beta}) E^T(t, \hat{\beta}) - E(t, \hat{\beta}) (S^{(1)}(t, \hat{\beta}))^T + E(t, \hat{\beta}) E^T(t, \hat{\beta}) S^{(0)}(t, \hat{\beta}) \right) d\Lambda(t) = \\
&= \frac{1}{n} \int_0^\infty \left(\frac{S^{(2)}(t, \hat{\beta})}{S^{(0)}(t, \hat{\beta})} - E(t, \hat{\beta}) E^T(t, \hat{\beta}) \right) S^{(0)}(t, \hat{\beta}) d\Lambda(t) = \hat{\Sigma}_0 + o_p(1), \\
\frac{1}{n} \sum_{i=1}^n \int_0^\infty (x^{(i)} - E(t, \hat{\beta})) (h(x^{(i)}; t, \hat{\beta}) - E_*(t, \hat{\beta}))^T e^{\beta^T x^{(i)}} Y_i(t) d\Lambda(t) &= \\
&= \frac{1}{n} \int_0^\infty \left((S_*^{(2)}(t, \hat{\beta}))^T - S^{(1)}(t, \hat{\beta}) E_*^T(t, \hat{\beta}) - E(t, \hat{\beta}) (S_*^{(1)}(t, \hat{\beta}))^T + E(t, \hat{\beta}) E_*^T(t, \hat{\beta}) S^{(0)}(t, \hat{\beta}) \right) d\Lambda(t) = \\
&= \frac{1}{n} \int_0^\infty \left(\frac{S_*^{(2)}(t, \hat{\beta})}{S^{(0)}(t, \hat{\beta})} - E(t, \hat{\beta}) E_*^T(t, \hat{\beta}) \right) dN + o_p(1) = \frac{1}{n} \int_0^\infty (V_*(t, \hat{\beta}))^T dN(t) + o_p(1) = \Sigma_* + o_p(1), \\
\frac{1}{n} \sum_{i=1}^n \int_0^\infty \{h(x^{(i)}; t, \hat{\beta}) - E_*(t, \hat{\beta})\}^{\otimes 2} e^{\beta^T x^{(i)}} Y_i(t) d\Lambda(t) &= \\
&= \frac{1}{n} \int_0^\infty \left(S_{**}^{(2)}(t, \hat{\beta}) - S_*^{(1)}(t, \hat{\beta}) E_*^T(t, \hat{\beta}) - E_*(t, \hat{\beta}) (S_*^{(1)}(t, \hat{\beta}))^T + E_*(t, \hat{\beta}) E_*^T(t, \hat{\beta}) S^{(0)}(t, \hat{\beta}) \right) d\Lambda(t) = \\
&= \frac{1}{n} \int_0^\infty \left(\frac{S_{**}^{(2)}(t, \hat{\beta})}{S^{(0)}(t, \hat{\beta})} - E_*(t, \hat{\beta}) E_*^T(t, \hat{\beta}) \right) S^{(0)}(t, \hat{\beta}) d\Lambda(t) = \frac{1}{n} \int_0^\infty V_{**}(t, \hat{\beta}) dN + o_p(1) = \hat{\Sigma}_{**} + o_p(1), \text{ kur} \\
V_{**}(t, \hat{\beta}) &= \frac{S_{**}^{(2)}(t, \hat{\beta})}{S^{(0)}(t, \hat{\beta})} - E_*(t, \hat{\beta}) E_*^T(t, \hat{\beta}),
\end{aligned}$$

$$S_{**}^{(2)}(t, \hat{\beta}) = \sum_{i=1}^n \left(h(x^{(i)}; t, \hat{\beta}) \right)^{\otimes 2} e^{\beta^T x^{(i)}} Y_i(t) + o_p(1), \quad \hat{\Sigma}_{**} = \frac{1}{n} \int_0^\infty V_{**}(t, \hat{\beta}) dN.$$

Taigi

$$\left\langle \frac{1}{\sqrt{n}} \hat{U} \right\rangle = \hat{\Sigma}_* \hat{\Sigma}_0^{-1} \hat{\Sigma}_0 \hat{\Sigma}_0^{-1} \hat{\Sigma}_*^T - \hat{\Sigma}_* \hat{\Sigma}_0^{-1} \hat{\Sigma}_*^T - \hat{\Sigma}_* \hat{\Sigma}_0^{-1} \hat{\Sigma}_*^T + \hat{\Sigma}_{**} + o_p(1) = \hat{\Sigma}_{**} - \hat{\Sigma}_* \hat{\Sigma}_0^{-1} \hat{\Sigma}_*^T + o_p(1) = \hat{D} + o_p(1),$$

$$\text{kur } \hat{D} = \hat{\Sigma}_{**} - \hat{\Sigma}_* \hat{\Sigma}_0^{-1} \hat{\Sigma}_*^T.$$

$$\text{Tada statistika } T = \left(\frac{1}{\sqrt{n}} \hat{U} \right)^T (\hat{D})^{-1} \left(\frac{1}{\sqrt{n}} \hat{U} \right) = \frac{1}{n} \hat{U}^T \hat{D}^{-1} \hat{U} \xrightarrow{d} \chi^2(m), \text{ kai } n \rightarrow \infty.$$

$\left\langle \frac{1}{\sqrt{n}} \hat{U} \right\rangle \xrightarrow{d} D$, nes iš centrinės ribinės teoremos (CRT) martingalams išplaukia, kad

$$\frac{1}{\sqrt{n}} \hat{U} \xrightarrow{d} N(0, D) \text{ ir } \hat{D} \xrightarrow{p} D.$$

Hipotezė atmetama su reikšmingumo lygmeniu α , jei $T > \chi_\alpha^2(m)$

Reumatoidinis artritas (RA)

1. Ligonių, sergančių RA, sveikatos būklės tyrimas

Tyrimas buvo atliktas Vilniaus universiteto Eksperimentinės ir klinikinės medicinos institute. Darbo vadovė doc. Jolanta Dadonienė, atsakinga vykdytoja Edita Gražulevičiūtė.

Bendra apžvalga. Reumatoidinis artritas – tai lėtinis, nuolat progresuojantis autoimuninės kilmės erozinis destruktivinis poliartritas. Ši liga gali pakenkti daugeliui kūno sąnarių, periartikuliniais (aplink sąnarį esantiems) audiniams, vidaus organams, akims, kraujagyslėms. Reumatoidiniu artritu serga apie 0,5 – 1 proc. suaugusių žmonių. Moterys serga du - tris kartus dažniau nei vyrai. Lietuvoje užregistruota apie dvylika tūkstančių suaugusių asmenų, besigydančių reumatoidinį artritą, o kasmet šia liga suserga apie 1 200 asmenų. Reumatoidiniu artritu gali sirgti įvairaus amžiaus žmonės, taip pat ir vaikai. Vis dėlto dažniausiai serga 50-70 metų žmonės. Reumatoidinis artritas - tai liga su atoslūgiais ir paūmėjimais, ilgai progresuojanti, sukelianti ne tik žymių sveikatos pokyčių, bet ir turinti rimtų socialinių pasekmių. Sveikatos ir socialinės problemos yra didžiulė našta ligoniui, jo šeimai ir visuomenei.

RA yra lėtinė nepagydoma liga. Šia liga sergantys ligoniai vaistus vartoja labai ilgai. Be vaistų pacientai gyventi negali, nes juos vargina labai stiprūs skausmai. Vaistai turi pakankamai daug pašalinių poveikių, kurie gali skatinti kitų ligų progresavimą.

Tyrimo tikslas. Išanalizuoti Vilniaus miesto reumatoidiniu artritu sergančių ligonių ligos charakteristikas, sveikatos, darbingumo būklę, gydymo taktiką, remiantis Vilniaus miesto reumatoidinio artrito registro duomenimis, palyginti šiuos duomenis su kitų šalių reumatoidinio artrito registru duomenimis.

Medžiaga ir metodai. Vilniuje reumatoidiniu artritu sergančiųjų analizė atlikta naudojant Vilniaus reumatoidinio artrito registro, kuris įkurtas 1998 m. pabaigoje ir sudarytas iš 1018 reumatoidiniu artritu sergančių pacientų (iš 486506 suaugusių, vyresnių kaip 16 metų, Vilniaus miesto gyventojų), duomenis.

Rezultatai. 85.9 proc. ligonių buvo moterys, t.y. 456 moterys, ir 75 vyrai, t.y. 14.1 proc.. Ligonių amžiaus vidurkis – 58 metai. Tirtų ligonių vidutinė ligos trukmė – 10.89 metų. 26.7 proc. sergančiųjų, t.y. 142 ligonių, buvo darbingi. Ligonių amžiaus vidurkis, kai jiems buvo diagnozuotas reumatoidinis artritas, yra 48.53 metų. Nesteroidinius vaistus nuo uždegimo vartojo beveik visi (98.3 proc.) ligoniai, sergantys reumatoidiniu artritu. Gliukokortikoidus vartojo 74,3 proc. sergančiųjų, 53.5 proc. ligonių anksčiau skirtos gliukokortikoidų injekcijos į sąnarius, dažniausiai į kelių sąnarius. Ligą modifikuojančiais vaistais gydyti 85.6 proc. reumatoidiniu artritu sergančių ligonių, dažniausiai skirta šios grupės vaisto - sulfasalazino. Sąnarių chirurginis gydymas taikytas

13.4 proc. ligonių, sinovektomija – 8.7 proc., sąnarių protezavimo operacijos - 4.7 proc. ligonių. Protezuojama praėjus vidutiniškai 16 metų nuo ligos pradžios.

Po penkerių metų iš tiriamo 531-o ligonio, mirė 32, t.y. 6 proc. Iš jų 8 vyrai, t.y. 25 proc. mirusių, ir 24 moterys, t.y. 75 proc. mirusių. Mirusių vyrų amžiaus vidurkis – 63.9 metų, o mirusių moterų – 69.2.

Taigi matome, kad moterys dažniau serga reumatoidiniu artritu, dažniau miršta juo sirgdamos, bet miršta vyresniame amžiuje nei vyrai. Nors reumatoidiniu artritu dažniau serga moterys, bet ankstesniame amžiuje dažniau miršta vyrai. Galima spėti, kad susirgusiems vyrams RA turi didesnę poveikį nei moterims. Galbūt todėl, kad vyrai apskritai gyvena trumpiau nei moterys, todėl liga dar labiau sumažina išgyvenimo trukmę.

Išvados. Reumatoidiniu artritu dažniau serga vyresnio amžiaus moterys. Ligą modifikuojančių vaistų gerokai dažniau vartoja trumpai sergantys ligoniai, dažniausiai sulfasalaziną. Lietuvos žmonės, sergantys reumatoidiniu artritu, dažniau neturi darbo, didesnę jų dalis turi invalidumo grupes, jų yra didesnė funkcinė negalia.

Statistinio paketo SAS užklauso, naudotos aprašant šiuos duomenis, pateiktos 2 priede.

2. Cox'o modelio taikymas, tiriant ligonių, sergančių reumatoidiniu artritu, mirtingumą

2.1. Ligonų mirtingumas kaip laiko nuo susirgimo funkcija

Atliekant minėtą tyrimą, ligoniams, sergantiems reumatoidiniu artritu (RA), buvo pateiktos anketos. Tyrimas pradėtas 1999 m. lapkričio mėnesį. Ligonai pildė anketas, kuriose buvo įvairūs klausimai apie ligonių amžių, sveikatos būklę, vartojamus vaistus, ligos trukmę ir pan. Po penkerių metų tiems patiems ligoniams vėl buvo duotos kitos anketos. Užpildyta, perduota institutui ir įvesta į duomenų bazę buvo 531 anketa. Šios anketos ir buvo panaudotos tyrimui.

Turimi duomenys yra tokie kaip ir dauguma išgyvenamumo analizei naudojamų duomenų, t.y. ne visų ligonių amžius yra vienodas ar bent jau panašus, jų ligos trukmės laikas taip pat yra skirtingas. Taip pat kai kurie ligoniai mirė, kiti – liko gyvi.

Tyrimas atliekamas, naudojantis SAS programiniu paketu. Taikomas Cox'o modelis.

Tyrimo metu 499 pacientai išgyveno, 32 mirė. Cenzūravimo kintamasis VSTATUS turi dvi reikšmes – 0 ir 1, kurios reiškia, kad ligonis išgyveno arba mirė atitinkamai iki tyrimo pabaigos.

Kintamieji, susiję su ligonių išgyvenamumu, yra:

- LYTKOD – lyties kodas;
- AMZDGN – ligonio amžius, kai jam buvo diagnozuotas RA;
- RA_GIM – ar ligonio giminėje yra sergančių ar sirgusių RA;
- RANK_DAR – darbo pobūdis (rankų darbas);
- GYD_DL_P – taikytas gydymas Delagiliu / Plakveniliu;
- GYD_AUKS – taikytas gydymas Aukso tabletėmis arba injekcijomis;
- GYD_PEN – taikytas gydymas Penicilaminu;
- GYD_MTX – taikytas gydymas Metotreksatu;
- AZA_IMUR – taikytas gydymas Azatriopinu / Imuranu;
- GYD_CYC – taikytas gydymas CYC / Endoxanu;
- GYD_SSL - taikytas gydymas SSL (Sulfasalazinu);
- GYD_CSP - taikytas gydymas Ciklosporinu;
- PO_ST – ar ligonis vartojo peroralinius steroidus;
- GYD_BAZ – kiek bazinių preparatų buvo taikyta;
- SINOVEKT – ar ligoniui buvo atlikta sąnarių chirurgijos operacija, Sinovektomija;
- KT_OP – ar ligoniui buvo atliktos kitos sąnarių operacijos;
- LIG_TR – ligos trukmė pilnais metais.

Po penkerių metų mirė nevisi ligoniai, todėl turima imtis yra cenzūruota iš dešinės. Kadangi į imtį buvo įtraukti tik tie ligoniai, kurie 1999 m. lapkričio mėn. buvo gyvi, tai praėjusio laiko nuo susirgimo iki mirties X sąlyginis skirstinys su sąlyga $X > V$, kur V yra laikas nuo susirgimo iki 1999 m. lapkričio mėn. Šiuo atveju nėra žinoma, kiek buvo ligonių iki 1999 m. lapkričio mėn. Taigi turimi duomenys taip pat yra ir nupjauti iš kairės. Ligonų išgyvenamumas tiriamas kaip laiko funkcija, t.y. ligos trukmės funkcija.

Taigi reikia nustatyti svarbius tiriamų ligonių išgyvenamumą lemiančius prognozuojančius veiksnius.

Kintamųjų RA_GIM, RANK_DAR, GYD_DL_P, GYD_AUKS, GYD_PEN, GYD_MTX, AZA_IMUR, KT_OP, GYD_SSL, GYD_CSP, PO_ST, SINOVEKT, GYD_CYC reikšmės duomenų bazėje buvo ne skaitinės. O procedūra **proc phreg** naudoja skaitines kintamųjų reikšmes. Tam reikėjo pakeisti minėtų kintamųjų reikšmes.

Programos SAS kodas, kuris buvo parašytas analizuojant turimus duomenis, pateikiamas 1 priede.

Duomenims buvo taikyta „pažingsninė“ (stepwise) regresija. Jos kodas yra:

```
proc phreg data=mano.analize_mag;  
  model LIG_TR*VSTATUS(0)= LYTKOD RA_GIM RANK_DAR AMZDGN GYD_DL_P  
    KT_OP GYD_PEN GYD_MTX AZA_IMUR GYD_CYC GYD_SSL GYD_CSP  
    PO_ST SINOVEKT GYD_AUKS  
/  
  selection=stepwise slentry=0.25  
  slstay=0.15 details;  
run;
```

Naudojamos regresijos „pažingsninis“ (fazinis) atrankos procesas susideda iš serijos besikaitaliojančių step-up („aukštyn“) ir step-down („žemyn“) fazių. Vieni kintamieji įtraukiami į modelį, kiti vėliau išeliminuojami iš modelio. Teorinis pažingsninės regresijos modelis pateikiamas 5 priede.

Pažingsninė regresinė analizė turi specialų žymėjimą modelio sakinyje: **SELECTION=STEPWISE**. Žymėjimas **SLENTRY=0.25** reiškia, kad kintamasis turi turėti ne didesnę kaip 0.25 reikšmingumo lygmenį, prieš jo įtraukimą į modelį. Opcija **SLSTAY=0.15** pažymi, kad kintamasis modelyje turi turėti ne didesnę 0.15 reikšmingumo lygmenį tam, kad liktų modelyje.

Priklausomas kintamasis yra LIG_TR – ligos trukmė pilnais metais, t.y. domina ligonių išgyvenamumas nuo susirgimo momento.

Įvykdę programą, gauname tokius rezultatus:

Model Information		
Data Set	MANO.ANALIZE_MAG	
Dependent Variable	LIG_TR	ligos trukme pilnais metais
Censoring Variable	vstatus	
Censoring Value(s)	0	
Ties Handling	BRESLOW	

Summary of the Number of Event and Censored Values			
Total	Event	Censored	Percent Censored
531	32	499	93.97

Taškų testo rezultatai kiekvienam kintamajam:

Analysis of Variables Not in the Model			
Variable	Score Chi-Square	Pr > ChiSq	Label
LYTKOD	5.7512	0.0165	lyties kodavimas
RA_GIM	0.5943	0.4407	
RANK_DAR	4.8693	0.0273	
AMZDGN	50.3818	<.0001	amzius kai diagnozavo
GYD_DL_P	5.6720	0.0172	
KT_OP	1.2163	0.2701	
GYD_PEN	0.8282	0.3628	
GYD_MTX	0.2674	0.6051	
AZA_IMUR	0.7879	0.3747	
GYD_CYC	1.0040	0.3163	
GYD_SSL	0.3497	0.5543	
GYD_CSP	0.3766	0.5394	
PO_ST	0.5801	0.4463	
SINOVEKT	0.3593	0.5489	

Analysis of Variables Not in the Model			
Variable	Score Chi-Square	Pr > ChiSq	Label
GYD_AUKS	2.8190	0.0932	

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
60.4374	15	<.0001

Taškų testas naudojamas tam, kad nustatytume, kuris iš aprašytų kintamųjų bus pirmasis įtrauktas į modelį. Šiuo atveju taškų testas kiekvienam iš kintamųjų yra globalus testas modeliui, kuris traktuoja kintamąjį kaip vienintelį aprašytą kintamąjį. χ^2 statistika yra lyginama su $\chi^2(1)$ skirstiniu.

Chi-kvadrat statistikos formulė pateikiama 3 priede, taškų testo statistikos išraiška – 4 priede.

Rezultatų lentelė „Analysis of Variables Not in the Model“ parodo kintamųjų χ^2 statistiką ir atitinkamą p-reikšmę. Kintamasis AMZDGN turi didžiausią χ^2 reikšmę (50.3818) ir yra svarbiausias, esant SLENTY=0.25 reikšmingumo lygmeniui, p-reikšmė yra **$p < 0.0001$** . Taigi kintamasis (kovariantė) AMZDGN yra įtraukiamas į modelį.

Pirmasis fazinio atrankos proceso modelis, kuriam priklauso viena kovariantė AMZDGN:

Step 1. Variable AMZDGN is entered. The model contains the following explanatory variables:

AMZDGN

Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Without Covariates	With Covariates
-2 LOG L	327.422	268.430
AIC	327.422	270.430
SBC	327.422	271.896

Testing Global Null Hypothesis: BETA=0
--

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	58.9913	1	<.0001
Score	50.3818	1	<.0001
Wald	46.0919	1	<.0001

Analysis of Maximum Likelihood Estimates							
Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Variable Label
AMZDGN	1	0.12861	0.01894	46.0919	<.0001	1.137	amzius kai diagnozavo

Aukščiau pateikta lentelė vaizduoja modelio rezultatus. Kadangi Wald χ^2 statistika turi SLSTAY=0.15 reikšmingumo lygmenį, tai AMZDGN lieka modelyje, nes jo p-reikšmė **$p < 0.0001$** .

Taškų testo panaudojimas kovariantei AMZDGN:

Analysis of Variables Not in the Model			
Variable	Score Chi-Square	Pr > ChiSq	Label
LYTKOD	4.8689	0.0273	lyties kodavimas
RA_GIM	1.1370	0.2863	
RANK_DAR	1.9631	0.1612	
GXD_DL_P	0.0295	0.8636	
KT_OP	0.2173	0.6411	
GXD_PEN	0.4160	0.5189	
GXD_MTX	3.2461	0.0716	
AZA_IMUR	2.6990	0.1004	
GXD_CYC	0.9479	0.3303	
GXD_SSL	0.2801	0.5967	
GXD_CSP	0.1635	0.6859	
PO_ST	0.3657	0.5453	
SINOVEKT	0.0565	0.8121	
GXD_AUKS	0.0420	0.8376	

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
15.3185	14	0.3567

Šiame žingsnyje atrenkamas kitas kintamasis ir įtraukiamas į modelį.

Lentelė „Analysis of Variables Not in the Model“ vaizduoja χ^2 statistiką ir p-reiškmes taškų testo, pritaikyto kintamajam AMZDGN ir likusiems keturiolikai kintamųjų. Score χ^2 duotam kintamajam yra tikėtinumo taškų testo reikšmė, tiriant kintamojo reikšmingumą ir modelyje jau esant kintamajam AMZDGN. Kintamasis LYTKOD yra atrenkamas, nes turi didžiausią χ^2 reikšmę (4.8689), esant SLENTY=0.25 reikšmingumo lygmeniui (p-reiškė **p=0.0273**).

Antrasis fazinio atrankos proceso modelis jau yra iš dviejų paaiškinančių kovariančių – LYTKOD, AMZDGN:

Step 2. Variable LYTKOD is entered. The model contains the following explanatory variables:

LYTKOD AMZDGN

Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Without Covariates	With Covariates
-2 LOG L	327.422	264.476
AIC	327.422	268.476
SBC	327.422	271.408

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	62.9453	2	<.0001
Score	53.5037	2	<.0001
Wald	47.7380	2	<.0001

Pritaikomas modelis abiem kintamiesiems - AMZDGN ir LYTKOD. Remiantis Wald statistika, nei AMZDGN, nei LYTKOD nepašalinami iš modelio.

Analysis of Maximum Likelihood Estimates							
Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Variable Label
LYTKOD	1	0.90085	0.42176	4.5623	0.0327	2.462	lyties kodavimas
AMZDGN	1	0.13129	0.01949	45.3695	<.0001	1.140	amzius kai diagnozavo

Analysis of Variables Not in the Model			
Variable	Score Chi-Square	Pr > ChiSq	Label
RA_GIM	0.6454	0.4218	
RANK_DAR	1.5348	0.2154	
GYD_DL_P	0.0025	0.9602	
KT_OP	0.1296	0.7188	
GYD_PEN	0.6093	0.4351	
GYD_MTX	3.1490	0.0760	
AZA_IMUR	3.1012	0.0782	
GYD_CYC	0.7960	0.3723	
GYD_SSL	0.1397	0.7086	
GYD_CSP	0.2148	0.6431	
PO_ST	0.3348	0.5628	
SINOVEKT	0.1219	0.7270	
GYD_AUKS	0.2068	0.6493	

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
10.8972	13	0.6194

Trečiasis fazinės regresijos modelis jau su trimis paaiškinančiomis kovariantėmis:

Step 3. Variable GYD_MTX is entered. The model contains the following explanatory variables:
LYTKOD AMZDGN GYD_MTX

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Without Covariates	With Covariates
-2 LOG L	327.422	261.665
AIC	327.422	267.665
SBC	327.422	272.062

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	65.7571	3	<.0001
Score	54.9250	3	<.0001
Wald	49.5226	3	<.0001

Analysis of Maximum Likelihood Estimates							
Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Variable Label
LYTKOD	1	0.89150	0.42254	4.4516	0.0349	2.439	lyties kodavimas
AMZDGN	1	0.13693	0.02015	46.1719	<.0001	1.147	amzius kai diagnozavo

Ir taip toliau, kaip aukščiau buvo aprašyti pirmi procedūros žingsniai, tęsiamas kintamųjų įtraukimas į modelį ar išeliminavimas iš jo.

Po kintamojo RANK_DAR pašalinimo fazinis atrankos procesas sustoja, nes pagal toliau skaičiuojamą Wald statistiką ir reikšmingumo lygmenį kintamąjį RANK_DAR vėl galima būtų įtraukti į modelį, tačiau to daryti negalima.

Step 6. Variable RANK_DAR is removed. The model contains the following explanatory variables:
LYTKOD AMZDGN GYD_MTX AZA_IMUR

Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	327.422	259.549
AIC	327.422	267.549
SBC	327.422	273.412

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	67.8728	4	<.0001
Score	55.9943	4	<.0001
Wald	50.1526	4	<.0001

RA 2.1.1.lent.Hipotezės tikrinimas

Analysis of Maximum Likelihood Estimates							
Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Variable Label
LYTKOD	1	0.94870	0.42486	4.9863	0.0255	2.582	lyties kodavimas
AMZDGN	1	0.13994	0.02048	46.6840	<.0001	1.150	amzius kai diagnozavo
GYD_MTX	1	0.63307	0.41154	2.3664	0.1240	1.883	
AZA_IMUR	1	0.62677	0.41393	2.2928	0.1300	1.872	

RA 2.1.2.lent. Maksimalaus tikėtinoumo įverčiai

Aukščiau pateiktos lentelės nesėkmių rizikos santykis (Hazard ratio) parodo, kad vyro rizika numirti yra 2.582 karto didesnė nei moters (nes 0-moteris, 1-vyras). Vyrai serga rečiau, bet susirgus jų mirtingumas didesnis. Žinoma, priežastis gali būti nebūtinai reumatoidinis artritas.

Kuo vyresniame amžiuje diagnozuota liga, tuo didesnis ligonių mirtingumas nuo ligos pradžios iki mirties. Metais vyresnių pacientų rizika numirti yra 15 procentų didesnė.

Gydymas metotreksatu pacientų mirtingumą padidina 88.3 procento arba 1.883 karto, o gydymas Azatriopinu / Imuranu pacientų mirtingumą padidina 87.2 procento arba 1.872 karto. Tokie rezultatai yra natūralūs, nes vaistai, skirti gydyti reumatoidiniam artritui, šiek tiek palengvina sąnarių skausmus, bet stipriai pakenkia kitiems organams (kepenims, inkstams), todėl organizmas silpsta.

Taigi po visų proceso žingsnių gaunama tokia modelio atrankos proceso santrauka:

Summary of Stepwise Selection							
Step	Variable		Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq	Variable Label
	Entered	Removed					
1	AMZDGN		1	50.3818	.	<.0001	amzius kai diagnozavo
2	LYTKOD		2	4.8689	.	0.0273	lyties kodavimas
3	GYD_MTX		3	3.1490	.	0.0760	
4	AZA_IMUR		4	2.3595	.	0.1245	
5	RANK_DAR		5	1.7438	.	0.1867	
6		RANK_DAR	4	.	1.7133	0.1906	

Iš gautų rezultatų matome, kad svarbiausi prognozuojantys veiksniai yra AMZDGN, LYTKOD, GYD_MTX ir AZA_IMUR t.y. šie veiksniai turi didžiausios įtakos ligonių, sergančių reumatoidiniu artritu, išgyvenamumui.

Rezultatų RA 2.1.1. lentelė tikrina hipotezę $H : b' = 0$. Ji yra atmetama, nes p-reikšmės mažesnės už 0.0001. O tai reiškia, kad kai kurios kovariantės yra reikšmingos, t.y. modelio koeficientai yra reikšmingi.

Lentelėje RA 2.1.2. „Analysis of Maximum Likelihood Estimates“ spausdinami:

- Parameter Estimate – koeficientų b_i maksimalaus tikėtumo įverčiai;
- Standard error – gautų parametrų įverčių standartinių nuokrypių iverčiai;
- Chi-square – Wald χ^2 statistika, kuri yra lygi parametro įverčio kvadratui, padalintam iš standartinio nuokrypio iverčio kvadrato;
- P-reikšmė ;
- Hazard ratio – nesėkmių rizikos santykis, t.y. exponentė, laipsnyje b_i , t.y. $\exp\{b_i\}$; Taip

gaunama iš santykio, kai, pavyzdžiui, turime dichotominę kovariantę $x_1 = \begin{cases} 1, \\ 0. \end{cases}$, o kitas

kovariantes fiksuojame, tada
$$\frac{\lambda(t | 1, x_2, x_3, \dots, x_p)}{\lambda(t | 0, x_2, x_3, \dots, x_p)} = \frac{\exp\{b_1 \cdot 1 + b_2 \cdot x_2 + \dots + b_p \cdot x_p\}}{\exp\{b_1 \cdot 0 + b_2 \cdot x_2 + \dots + b_p \cdot x_p\}} = \exp\{b_1\};$$

- Variable lable – kintamojo etiketė, t.y. komentaras apie tai, ką reiškia kintamojo žymėjimas. Pavyzdžiui, AMZDGN kintamojo etiketė, t.y. jo reikšmė, yra „ligonio amžius, kai jam buvo diagnozuotas reumatoidinis artritas“. Turimuose duomenyse ne visiems kintamiesiems yra priskirtos etiketės.

Cox'o modelio pavidalas yra toks: $h(t; X) = h_0(t) \cdot \exp\{b' X\}$

Taigi jei pažymėtume x_1 – LYTKOD, x_2 – AMZDGN, x_3 – GYD_MTX, x_4 – AZA_IMUR, tada modelį galima būtų užrašyti taip:

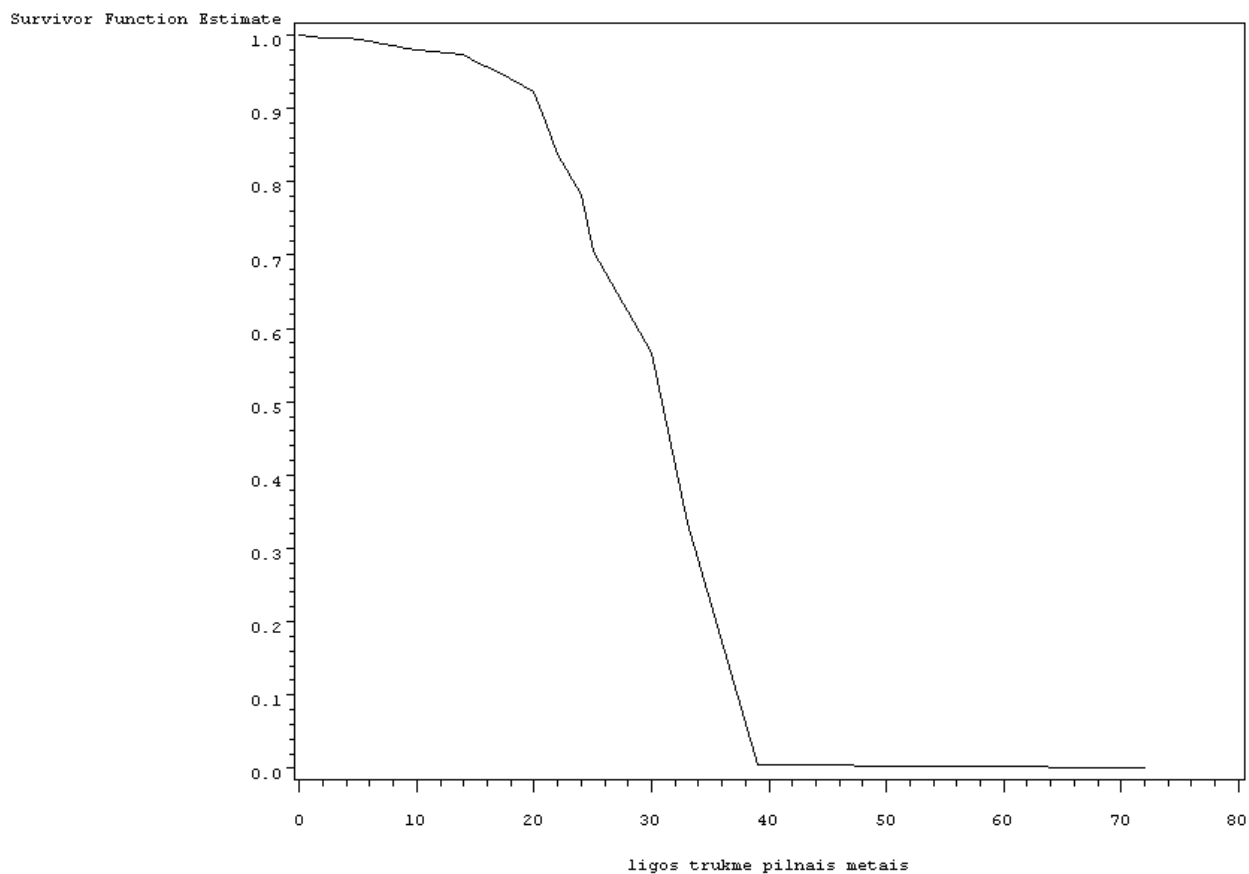
$$h(t | x) = h_0(t) \cdot \exp\left\{\sum_{j=1}^p b_j \cdot x_j\right\} =$$
$$= h_0(t) \cdot \exp\{0,94870 \cdot x_1 + 0,13994 \cdot x_2 + 0,63307 \cdot x_3 + 0,62677 \cdot x_4\}, \text{ arba}$$
$$S(t | x) = [S_0(t)]^{\exp\{0,94870 \cdot x_1 + 0,13994 \cdot x_2 + 0,63307 \cdot x_3 + 0,62677 \cdot x_4\}}$$

Šiuo atveju $p = 4$, t.y. $X = (x_1; x_2; x_3; x_4)$, o funkcija $h_0(t)$ - nesėkmių intensyvumo funkcija, kai ignoruojamos visos kovariantės, t.y. imamos nulinės kovariančių reikšmės.

Bazinės išgyvenamumo funkcijos įterčiai buvo apskaičiuoti, į procedūrą **proc phreg** įvedus keletą pakeitimų. O grafikas nubraižytas, naudojantis SAS procedūra **proc gplot**:

```
proc phreg data=mano.analize_mag;  
  model LIG_TR*VSTATUS(0)= RA_GIM RANK_DAR GYD_DL_P LYTKOD AMZDGN  
    GYD_BAZ KT_OP GYD_PEN GYD_MTX AZA_IMUR  
    GYD_CYC GYD_SSL GYD_CSP PO_ST SINOVEKT  
    GYD_AUKS;  
baseline out=a survival=s logsurv=ls loglogs=lls;  
run;  
  
proc print data=a;  
run;  
  
data grafikas;  
set a;  
ls=-ls;  
run;  
  
proc gplot data=grafikas;  
symbol1 value=none interpol=join;  
plot s*LIG_TR;run;
```

Bazinės išgyvenamumo kreivės grafikas atrodo taip:

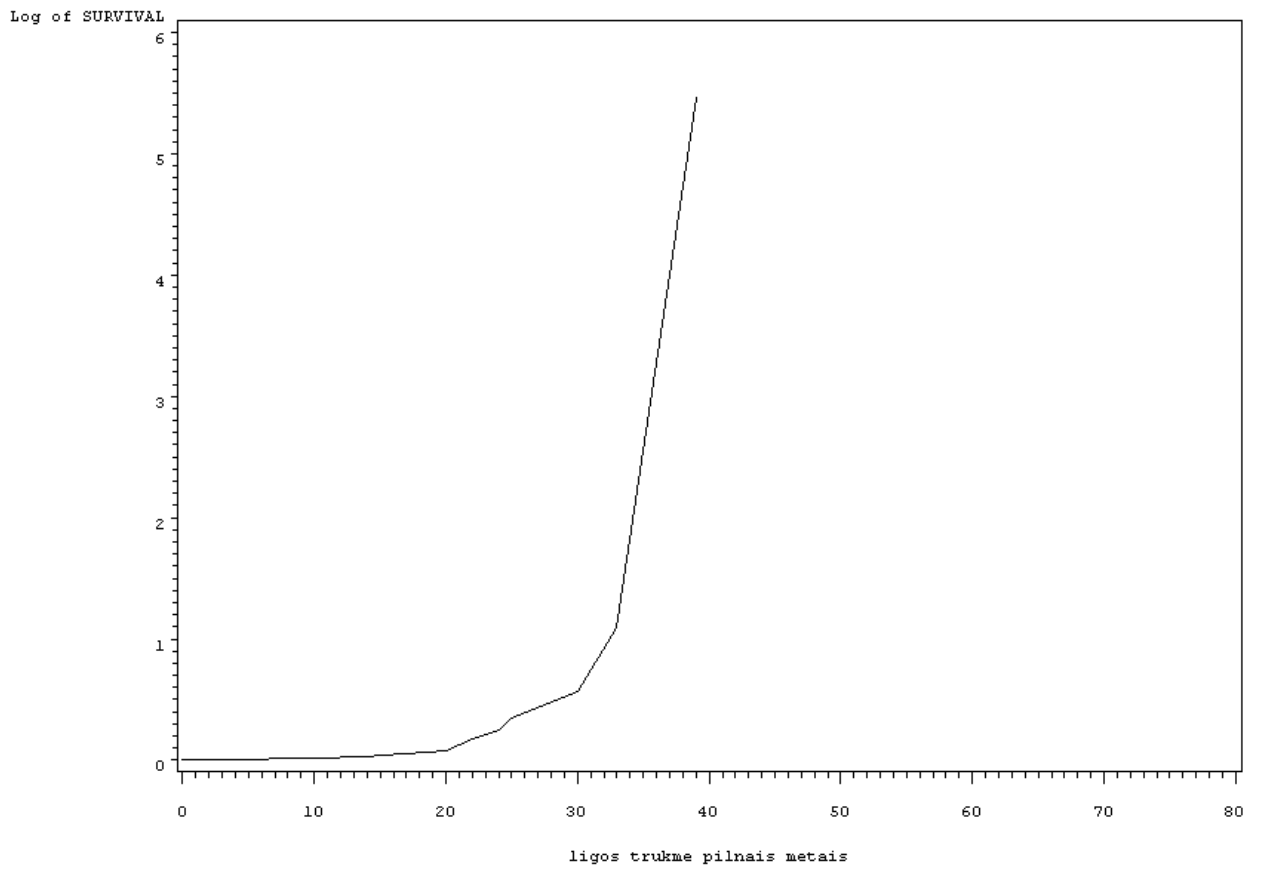


RA 2.1.pav. Bazinė išgyvenamumo kreivė

2.1.paveikslėlis parodo, kad kuo ilgesnė ligos trukmė, tuo mažesnė tikimybė išgyventi. Iš grafiko matome, kad sergant reumatoidiniu artritu ilgiau nei 40 metų, tikimybė išgyventi artėja prie nulio.

Norint rasti išgyvenamumo funkciją prie konkrečių kovariančių reikšmių, imamas šios funkcijos laipsnis, kurį apsprendžia kovariančių reikšmės.

Išgyvenamumo kreivės logaritmo grafikas yra toks:



RA 2.2.pav. Išgyvenamumo kreivės logaritmas

2.2. Ligoniu mirtingumas kaip amžiaus funkcija

Ankstesniame skyriuje buvo tirtas ligoniu, sergančiu reumatoidiniu artritu, mirtingumas kaip laiko nuo susirgimo funkcija.

Šiame skyriuje nagrinėjamas šių ligoniu mirtingumas kaip ligoniu amžiaus funkcija.

Tam programoje buvo įvesti keli nauji kintamieji:

- *ageacct* – tik pakeistas kovariantės AMZDGN pavadinimas, o ši kovariantė reiškia ligoniu amžių ligos diagnozavimo metu;
- *ageacct_2* – ligonio amžius tyrimo pradžioje;
- *agels* – amžius ligonio, kada jis paskutinį kartą buvo matytas, t.y. stebėjimo cenzūravimas arba ligonio mirtis;
- (*ageacct,agels*) – ligonio amžiaus intervalas, kurio metu jis buvo rizikos grupėje, t.y. galėjo mirti.

Taigi šiuo atveju taip pat taikoma pažingsninė regresija. SAS kodas šiek tiek pakeičiamas taip:

```
proc phreg data=mano.analyze_mag;  
  model (ageacct,agels)*VSTATUS(0)= LYTKOD RA_GIM RANK_DAR LIG_TR GYD_DL_P  
    KT_OP GYD_PEN GYD_MTX AZA_IMUR GYD_CYC GYD_SSL GYD_CSP  
    PO_ST SINOVEKT GYD_AUKS  
  /  
  selection=stepwise slentry=0.25  
  slstay=0.15 details;  
run;
```

Įvykdę programą, gauname tokius rezultatus:

The PHREG Procedure

Model Information	
Data Set	MANO.ANALIZE_MAG
Dependent Variable	ageacct
Dependent Variable	agels
Censoring Variable	vstatus
Censoring Value(s)	0
Ties Handling	BRESLOW

Nagrinėjant duomenis šiuo aspektu, jau yra du priklausomi kintamieji. Tai ligoniu amžius ligos diagnozavimo metu ir ligoniu amžius, kada jie paskutinį kartą buvo matyti, t.y. mirė arba stebėjimas buvo cenzūruotas. Cenzūravimo kintamasis lieka toks pats:

$$vstatus = \begin{cases} 1, & \text{jei ligonis mirė,} \\ 0, & \text{kitu atveju.} \end{cases}$$

Summary of the Number of Event and Censored Values			
Total	Event	Censored	Percent Censored
531	32	499	93.97

Taškų testo rezultatai kiekvienam kintamajam:

Analysis of Variables Not in the Model			
Variable	Score Chi-Square	Pr > ChiSq	Label
LYTKOD	2.5279	0.1119	lyties kodavimas
RA_GIM	0.9458	0.3308	
RANK_DAR	1.9019	0.1679	
LIG_TR	0.0437	0.8345	ligos trukme pilnais metais
GYD_DL_P	0.0208	0.8852	
KT_OP	0.0316	0.8589	
GYD_PEN	0.4304	0.5118	
GYD_MTX	2.6711	0.1022	
AZA_IMUR	2.9574	0.0855	
GYD_CYC	0.8695	0.3511	
GYD_SSL	0.2483	0.6183	
GYD_CSP	0.1792	0.6721	
PO_ST	0.6569	0.4177	
SINOVEKT	0.0238	0.8774	
GYD_AUKS	0.3222	0.5703	

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
12.7860	15	0.6188

Taškų testas naudojamas tam, kad nustatytume, kuris iš aprašytų kintamųjų bus pirmasis įtrauktas į modelį. Šiuo atveju taškų testas kiekvienam iš kintamųjų yra globalus testas modeliui, kuris traktuoja kintamąjį kaip vienintelį aprašytą kintamąjį. χ^2 statistika yra lyginama su $\chi^2(1)$ skirstiniu.

Rezultatų lentelė „Analysis of Variables Not in the Model“ parodo kintamųjų χ^2 statistiką ir atitinkamą p-reikšmę. Kintamasis AZA_IMUR turi didžiausią χ^2 reikšmę (2.9574) ir yra svarbiausias, esant SLENTY=0.25 reikšmingumo lygmeniui, p-reikšmė yra **p=0.0855**. Taigi kintamasis (kovariantė) AZA_IMUR yra įtraukiamas į modelį.

Pirmasis fazinio atrankos proceso modelis, kuriam priklauso viena kovariantė AZA_IMUR yra toks:

Step 1. Variable AZA_IMUR is entered. The model contains the following explanatory variables:

AZA_IMUR

Convergence Status	
Convergence criterion (GCONV=1E-8) satisfied.	

Model Fit Statistics		
Criterion	Without Covariates	With Covariates
-2 LOG L	286.242	283.659
AIC	286.242	285.659
SBC	286.242	287.125

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	2.5831	1	0.1080
Score	2.9574	1	0.0855
Wald	2.8515	1	0.0913

Analysis of Maximum Likelihood Estimates							
Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Variable Label
AZA_IMUR	1	0.68022	0.40282	2.8515	0.0913	1.974	

Aukščiau pateikta lentelė vaizduoja modelio rezultatus. Kadangi Wald χ^2 statistika turi SLSTAY=0.15 reikšmingumo lygmenį, tai AZA_IMUR lieka modelyje, nes jo p-reikšmė $p=0.0913$, t.y. nėra didesnė už 0.15.

Taškų testo panaudojimas kovariantei AZA_IMUR:

Analysis of Variables Not in the Model			
Variable	Score Chi-Square	Pr > ChiSq	Label
LYTKOD	2.6472	0.1037	lyties kodavimas
RA_GIM	1.0664	0.3018	
RANK_DAR	2.2352	0.1349	
LIG_TR	0.0002	0.9894	ligos trukme pilnais metais
GYD_DL_P	0.0071	0.9331	
KT_OP	0.0832	0.7730	
GYD_PEN	0.6354	0.4254	
GYD_MTX	1.8394	0.1750	
GYD_CYC	0.7888	0.3745	
GYD_SSL	0.2975	0.5855	
GYD_CSP	0.2506	0.6167	
PO_ST	0.2109	0.6461	
SINOVEKT	0.0190	0.8904	
GYD_AUKS	0.1596	0.6895	

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
9.7086	14	0.7832

Šiame žingsnyje atrenkamas kitas kintamasis ir įtraukiamas į modelį.

Lentelė „Analysis of Variables Not in the Model“ vaizduoja χ^2 statistiką ir p-reikšmes taškų testo, pritaikyto kintamajam AZA_IMUR, likusiems keturiolikai kintamųjų. Score χ^2 duotam kintamajam tikėtinumo yra taškų testo reikšmė, tiriant kintamojo reikšmingumą ir modelyje jau

esant kintamajam AZA_IMUR . Kintamasis LYTKOD yra atrenkamas, nes turi didžiausią χ^2 reikšmę (2.6472), esant SLENTY=0.25 reikšmingumo lygmeniui (p-reikšmė $p=0.1037$).

Antrasis fazinio atrankos proceso modelis iš dviejų paaiškinančių kovariančių – LYTKOD, AZA_IMUR:

Step 2. Variable LYTKOD is entered. The model contains the following explanatory variables:

LYTKOD AZA_IMUR

Convergence Status	
Convergence criterion (GCONV=1E-8) satisfied.	

Model Fit Statistics		
Criterion	Without Covariates	With Covariates
-2 LOG L	286.242	281.379
AIC	286.242	285.379
SBC	286.242	288.310

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	4.8633	2	0.0879
Score	5.5699	2	0.0617
Wald	5.3626	2	0.0685

Analysis of Maximum Likelihood Estimates							
Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Variable Label
LYTKOD	1	0.65838	0.41193	2.5545	0.1100	1.932	lyties kodavimas
AZA_IMUR	1	0.69456	0.40375	2.9593	0.0854	2.003	

Pritaikomas modelis abiemis kintamiesiems LYTKOD ir AZA_IMUR. Abu kintamieji modelyje lieka, nes jų p-reikšmės neviršija pasilikimo modelyje sąlygos SLSTAY=0.15.

Tokiu pat principu, tikrinant sąlygas, kintamieji įtraukiami į modelį arba eliminuojami iš jo.

Po kintamojo RANK_DAR pašalinimo fazinis atrankos procesas sustoja, nes pagal toliau skaičiuojamą Wald statistiką ir reikšmingumo lygmenį kintamąjį RANK_DAR vėl galima būtų įtraukti į modelį, tačiau to daryti negalima.

Step 4. Variable RANK_DAR is removed. The model contains the following explanatory variables:
LYTKOD AZA_IMUR

Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Without Covariates	With Covariates
-2 LOG L	286.242	281.379
AIC	286.242	285.379
SBC	286.242	288.310

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	4.8633	2	0.0879
Score	5.5699	2	0.0617
Wald	5.3626	2	0.0685

RA 2.2.1.lent. Hipotezės tikrinimas

Analysis of Maximum Likelihood Estimates							
Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Variable Label
LYTKOD	1	0.65838	0.41193	2.5545	0.1100	1.932	lyties kodavimas
AZA_IMUR	1	0.69456	0.40375	2.9593	0.0854	2.003	

RA 2.2.2.lent. Maksimalaus tikėtinumo įverčiai

Tirdami reumatoidinio artrito ligonių mirtingumą kaip amžiaus funkciją, gauname jog ligonių vyrų rizika numirti yra 1.932 karto didesnė nei moterų. Vyrų serga rečiau, bet susirgus, jų mirtingumas didesnis. Aišku, priežastis gali būti nebūtinai reumatoidinis artritas.

Gydymas Azatriopinu / Imuranu pacientų mirtingumą padidina 2.003 karto. Toks rezultatas yra natūralus, nes vaistai, skirti gydyti reumatoidiniam artritui, šiek tiek palengvina sąnarių

skausmus, bet stipriai pakenkia kitiems organams (kepenims, inkstams), todėl organizmas silpsta.

Analysis of Variables Not in the Model			
Variable	Score Chi-Square	Pr > ChiSq	Label
RA_GIM	0.7519	0.3859	
RANK_DAR	1.9852	0.1588	
LIG_TR	0.0122	0.9119	ligos trukme pilnais metais
GYD_DL_P	0.0439	0.8340	
KT_OP	0.0304	0.8616	
GYD_PEN	0.9144	0.3389	
GYD_MTX	1.9458	0.1630	
GYD_CYC	0.6918	0.4055	
GYD_SSL	0.1953	0.6585	
GYD_CSP	0.3470	0.5558	
PO_ST	0.1334	0.7149	
SINOVEKT	0.0744	0.7850	
GYD_AUKS	0.3972	0.5285	

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
6.7861	13	0.9128

Model building terminates because the variable to be entered is the variable that was removed in the last step.

Procesui sustojus, gaunama tokia modelio atrankos proceso santrauka:

Summary of Stepwise Selection							
Step	Variable		Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq	Variable Label
	Entered	Removed					
1	AZA_IMUR		1	2.9574	.	0.0855	

Summary of Stepwise Selection							
Step	Variable		Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq	Variable Label
	Entered	Removed					
2	LYTKOD		2	2.6472	.	0.1037	Iyties kodavimas
3	RANK_DAR		3	1.9852	.	0.1588	
4		RANK_DAR	2	.	1.9450	0.1631	

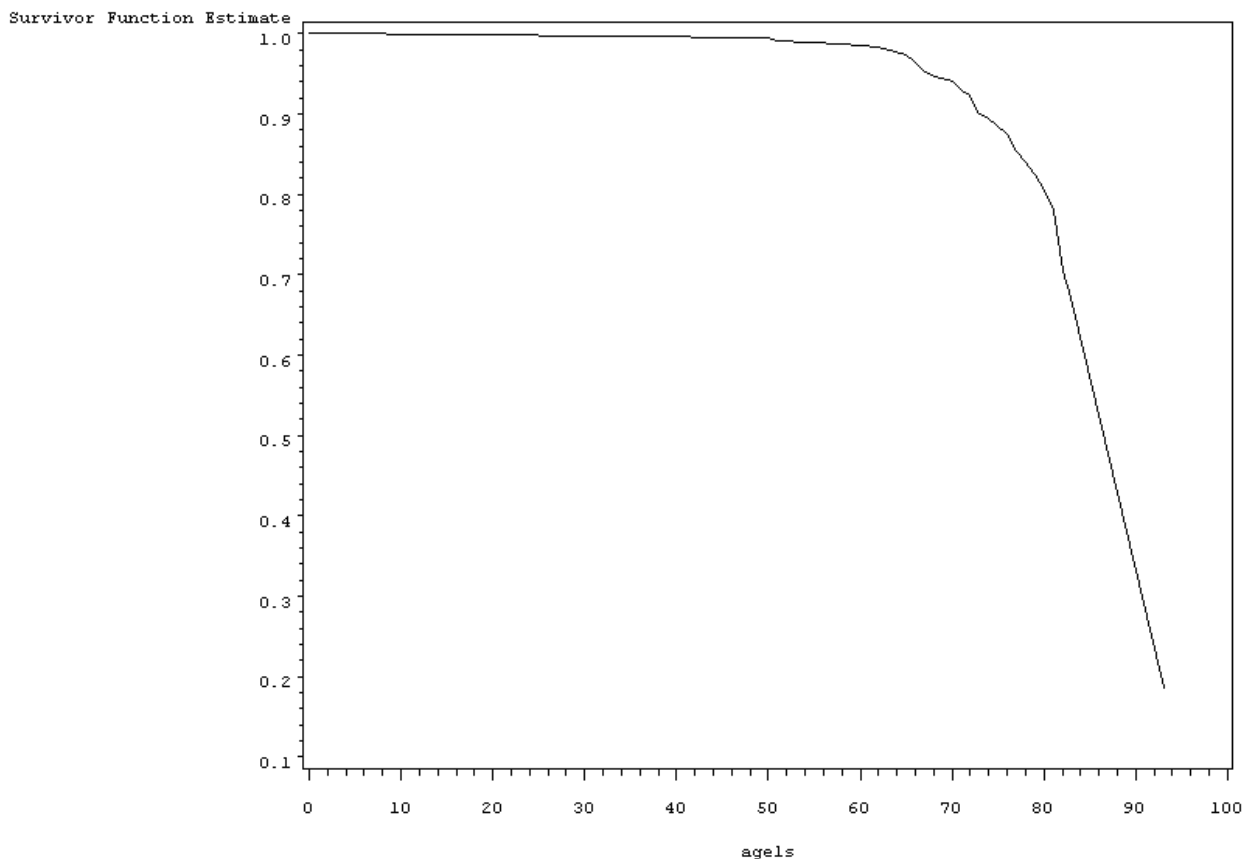
Iš gautų rezultatų matome, kad svarbiausias prognozuojantys veiksniai yra LYTKOD ir AZA_IMUR, t.y. šie veiksniai turi didžiausios įtakos ligonių, sergančių reumatoidiniu artritu, išgyvenamumui, tiriant išgyvenamumą kaip amžiaus funkciją.

Lentelėje RA 2.2.2. modelio kintamojo LYTKOD koeficiento įvertis yra 0.65838, kintamojo AZA_IMUR koeficiento įvertis yra 0.69456, tada jei pažymėtume LYTKOD – x_1 , o AZA_IMUR - x_2 modelio pavidalas būtų toks:

$$h(t | x) = h_0(t) \cdot \exp \left\{ \sum_{j=1}^p b_j \cdot x_j \right\} = h_0(t) \cdot \exp \{ 0.65838 \cdot x_1 + 0.69456 \cdot x_2 \}, \text{ arba}$$

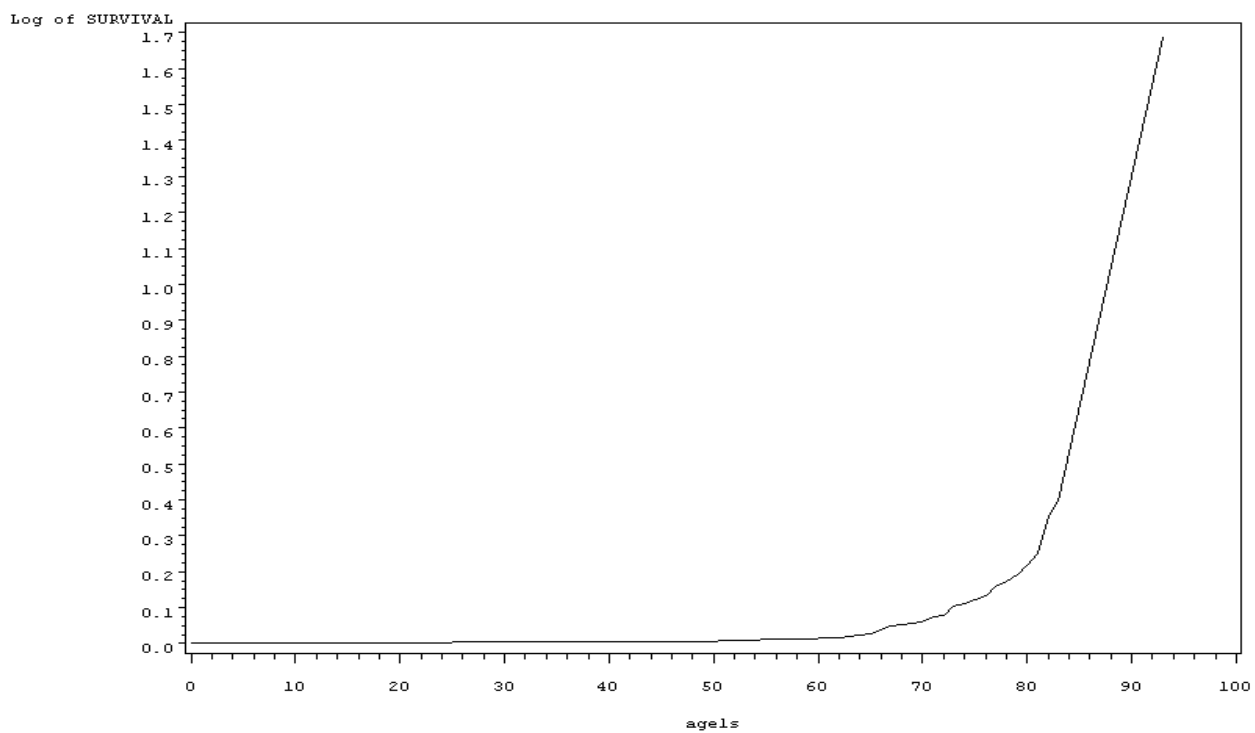
$$S(t | x) = [S_0(t)]^{\exp \{ 0.65838 \cdot x_1 + 0.69456 \cdot x_2 \}}$$

Išgyvenamumo kreivės kaip amžiaus funkcijos grafikas yra toks:



RA 2.2.1.pav. Išgyvenamumo kreivė

RA 2.2.1.paveikslėlis parodo, jog jau 50-ties metų amžiaus ir vyresnių ligonių išgyvenamumas sparčiai mažėja.



RA 2.2.2.pav. Išgyvenamumo kreivės logaritmas

Chemoterapija

Šiame skyrelyje taikomas pirmos darbo dalies šeštame skyrelyje aprašyto Paprastų persidengiančių kovariančių (SCE) modelio taikymas.

1. Persikerančių mirimų intensyvumų modelio taikymas vėžiu sergantiems ligoniams

Minėto modelio taikymui iš pradžių pasirinkta vėžiu sirgusių 90 ligonių duomenų aibė. Ligoniams buvo taikyti du gydymo būdai – chemoterapija arba radioterapija. Duomenų aibė ir SAS programos kodas pridamas 6 priede.

Chemoterapija - priešvėžinių vaistų (citotoksinų) naudojimas siekiant sunaikinti vėžines ląsteles (įskaitant leukemijas ir limfomas). Tai terapija, kai žmogui duodami vaistai, naikinantys ląsteles.

Yra daugiau nei 50 rūšių skirtingų chemoterapinių vaistų, dažnai vartojami keli vaistai vienu metu (kombinuota chemoterapija).

Kai kurioms vėžio rūšims užtenka vien chemoterapijos, kitais atvejais chemoterapija derinama su kitais gydymo būdais – chirurgine operacija, radioterapija (švitinimu), hormonų terapija, imunoterapija ar iškart su keliais būdais.

Chemoterapijos veikimas. Chemoterapijoje naudojami vaistai užkerta kelią vėžinėms ląstelėms dalintis ir daugintis. Vaistai krauju išnešiojami po visus organizmo audinius. Vaistai taip pat paveikia kai kurias sveikas ląsteles – burnos gleivinę, kaulų čiulpus (kurie gamina kraują), plaukų folikulus, virškinimo sistemą. Sveikosios ląstelės po chemoterapijos atsistato, tačiau vėžinės ląstelės negali atsistatyti ir žūna. Didelė chemoterapinio poveikio dalis susijusi su itin sparčia vėžinių ląstelių medžiagų apykaita: vėžinė ląstelė per tą patį laiką įsisavina bent kelis kartus daugiau toksinių medžiagų. Dėl šios priežasties anksčiau chemoterapijai būdavo naudojama nemažai paprastų nuodingųjų preparatų, pavyzdžiui, arseno junginių.

Skirtingi vaistai skirtingai veikia vėžines ląsteles. Jei naudojami keli vaistai, jie parenkami taip, kad papildytų vienas kitą skirtingu poveikiu.

Chemoterapija kruopščiai planuojama taip, kad kiekvieno kurso metu sunaikintų kuo daugiau vėžinių ląstelių, bet nesunaikintų sveikų audinių.

Skirtingi chemoterapiniai vaistai skirtingai veikia greitai besidauginančias ląsteles:

- blokuoja ląstelės receptorių, priimančius signalus, verčiančius ląstelę dalintis;

- veikia ląstelės DNR ir neleidžia genetinei medžiagai dalintis;
- veikia pagalbines ląstelės dalis (pavyzdžiui, branduolį), reikalingas ląstelės dauginimuisi.

Radioterapija yra piktybinių navikų gydymo būdas naudojant aukštos energijos jonizuojančiąją spinduliuotę – rentgeno, gama spindulius, fotonų, elektronų, neutronų srautus ir kita. Jonizuojančiųjų spindulių poveikiui jautriausios yra greitai besidauginančios ląstelės – kaip tik tokios ir yra piktybinės ląstelės. Jonizuojantieji spinduliai, veikdami jas, sutrikdo jų dauginimosi procesus – navikas nustoja augti, sunyksta. Radioterapija kitaip dar vadinama švitinimu, gydymu spinduliais, spinduline terapija. Tai vienas iš pagrindinių piktybinių navikų gydymo būdų.

Radioterapija – lokalus (vietinis) piktybinių navikų gydymo būdas, t.y. siekiama jonizuojančiais spinduliais veikti patį naviką ar pažeistus limfmazgius.

Kai radioterapija taikoma siekiant sunaikinti piktybinį naviką ir išgydyti pacientą, ji vadinama radikaliąja radioterapija.

Radioterapija gali būti naudojama piktybiniam procesui progresuojant, siekiant sušvelninti ar numalšinti pacientą varginančius simptomus, pavyzdžiui, skausmą. Tai yra paliatyvioji radioterapija. Šiais atvejais spindulių dozės mažesnės, trumpesnė gydymo trukmė nei radikalesios radioterapijos atveju.

Kartais spindulinis gydymas taikomas derinant jį su kitais metodais – prieš ar po chirurginės operacijos, prieš chemoterapiją ar po jos. Kai radioterapija taikoma su chemoterapija, tai vadinama chemospinduliniu gydymu.

Atliekant piktybinių navikų radioterapiją, neišvengiamai jonizuojančiais spinduliais paveikiamos ir sveikosios ląstelės, pakliūnančios į švitinimo lauką. Tačiau jos daug mažiau jautrios spindulių poveikiui, todėl netrukus atsigauna ir atsinaujina.

Radioterapijos tikslas – sunaikinti piktybinį naviką. Piktybiniam navikams gydyti vis dažniau pasitelkiamos šiuolaikinės modernios spindulinio gydymo technologijos. Tai ženkliai sumažina radioterapijos šalutinius poveikius, tačiau visiškai jų išvengti kol kas neįmanoma.

Tirtą imtį sudaro 90 – ties ligonių duomenys. 8 iš jų buvo cenzūruoti.

Iš pradžių šiems duomenims buvo pritaikytas Cox'o modelis. Gauti rezultatai yra tokie:

Model Information	
Data Set	WORK.CHEMOTERAPIJA_2
Dependent Variable	X
Censoring Variable	delta_i

Model Information	
Censoring Value(s)	0
Ties Handling	BRESLOW

Summary of the Number of Event and Censored Values			
Total	Event	Censored	Percent Censored
90	82	8	8.89

Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Without Covariates	With Covariates
-2 LOG L	615.171	614.946
AIC	615.171	616.946
SBC	615.171	619.353

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	0.2243	1	0.6358
Score	0.2250	1	0.6353
Wald	0.2248	1	0.6354

Analysis of Maximum Likelihood Estimates						
Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
x_terapija	1	0.10589	0.22335	0.2248	0.6354	1.112

Iš gautų rezultatų matome, kad ligonių, sirgusių vėžiu ir jiems taikius chemoterapiją ir radioterapiją kartu, mirtingumas buvo 1.112 karto didesnis nei tų ligonių, kuriems buvo taikyta vien tik chemoterapija.

Kintamojo x_{terapija} koeficiento įvertis yra 0.10589. Tada jei pažymėtume $x_{\text{terapija}} - x_1$, modelio pavidalas būtų toks:

$$h(t|x) = h_0(t) \cdot \exp\left\{\sum_{j=1}^p b_j \cdot x_j\right\} = h_0(t) \cdot \exp\{0.10589 \cdot x_1\}, \text{ arba } S(t|x) = [S_0(t)]^{\exp\{0.10589 \cdot x_1\}}$$

Gavus koeficiento $\beta = 0.10589$ įvertį, buvo skaičiuota T statistikos reikšmė SCE modelio atveju.

SAS programos kodas statistikos skaičiavimui pateikiamas 7 priede.

SCE modelio atveju gauta $T = 13.108959 > \chi_{0,05}^2(1) = 3.84146$. Vadinasi nulinė hipotezė akivaizdžiai atmetama, nes Chemo-radio terapijos duomenų atveju Cox'o modelio nėra.

2. Sukonstruoto kriterijaus Cox'o modelio adekvatumui tikrinti taikymas reumatoidiniu artritu sergantiems ligoniams

Pirmos darbo dalies 7 skyrelyje buvo sukonstruotas kriterijus Cox'o modelio adekvatumui tikrinti nupjautų iš kairės ir cenzūruotų iš dešinės duomenų atveju.

Reumatoidiniu artritu sergančių ligonių duomenų specifika ir yra tokia, kad jie yra nupjauti iš kairės ir cenzūruoti iš dešinės, todėl šiai duomenų imčiai ir buvo patikrintas minėtas kriterijus.

Tiriant ligonių mirtingumą kaip amžiaus funkciją, t.y. taikant Cox'o modelį, prognozuojančių kovariančių koeficientų įverčiai buvo gauti tokie: $\hat{\beta}_1 = 0.65838$, $\hat{\beta}_2 = 0.69456$.

Naudojantis statistiniu paketu SAS, gauta statistika $T = 2.4690499 < \chi_{0,05}^2(2) = 5.99146$. Vadinasi hipotezė dėl Cox'o modelio adekvatumo duomenims neatmetama.

IŠVADOS

Darbo metu buvo išnagrinėta įvairių faktorių (kovariančių) įtaka reumatoidiniu artritu sergančio 531 ligonio mirtingumui. Buvo taikomas vienas iš regresinių išgyvenamumo modelių – Cox'o modelis.

Iš minėtos 531 ligonio imties mirę buvo 32 ligoniai.

Iš pradžių buvo tiriama ligonių imtis laiko nuo ligos pradžios aspektu. Šiuo atveju prognozuojantys veiksniai buvo amžius, kada liga buvo diagnozuota (AMZDGN), lytis (LYTKOD), gydymas Metotreksatu (GYD_MTX) ir gydymas Azatriopinu/Imuranu (AZA_IMUR).

Vėliau, tiriant ligonių mirtingumą kaip amžiaus funkciją, nustatyti svarbiausi lemiantys veiksniai buvo šie: ligonių lytis (LYTKOD) ir gydymas Azatriopinu/Imuranu (AZA_IMUR).

Gauti rezultatai, t.y. ligonių išgyvenamumą lemiančios kovariantės (veiksniai), beveik visiškai sutampa su gydytojų nurodytais. Tai dar kartą patvirtina matematinių statistinių modelių, šiuo atveju nagrinėjamo Cox'o modelio, taikymo realiame gyvenime, svarbą.

Kitai duomenų imčiai, t.y. vėžiu sergančių ligonių duomenų aibei, buvo taikomas Persikertančių mirimų intensyvumų (SCE) modelis, t.y. tikrinama Cox'o modelio adekvatumo duomenims hipotezė. Hipotezė buvo atmesta, nes minėtiems duomenims Cox'o modelis negalioja.

Pagrindinis darbo rezultatas yra šis: gautas kriterijus Cox'o modelio adekvatumui tikrinti, naudojant nupjautus iš kairės ir cenzūruotus iš dešinės duomenis, sudarytos programos kriterijui realizuoti. Reumatoidinio artrito ligonių duomenų aibei, t.y. nupjautiems iš kairės ir cenzūruotiems iš dešinės duomenims, hipotezė buvo neatmesta, t.y. Cox'o modelio adekvatumas duomenims neatmestas.

UŽSIENIO KALBA

In this work the Cox proportional hazards model was applied to investigate the influence of various factors (covariates) to mortality of rheumatoid arthritis patients of Vilnius.

In the first case, the sample of 531 patients was analysed.

Analysing survival of patients of the sample as function of time from the beginning of the disease, the prognostic factors were LYTKOD (the sex of patients), AMZDGN (patients' age, when the rheumatoid arthritis was diagnosed), GYD_MTX (treatment with metotrexat) and AZA_IMUR (treatment with Azatriopin/Imuran). When survival was analysed as function of age then the prognostic factor were LYTKOD (the sex of patients) and AZA_IMUR (treatment with Azatriopin/Imuran).

The results are almost identical to those, which doctors suggested. This fact confirms the importance of using mathematical statistical models to solve the problems of the real life. In this case, the importance of using the Cox model.

On the other hand, Simple cross-effects (SCE) model was applied for the sample of cancer patients. In the case of this model the hypothesis of Cox model fitting for cancer patients' data was rejected.

The most important result of this work is that the criterion of Cox model fitting to left truncated and right censored data was constructed. Also a program of SAS for the criterion was created. The the hypothesis of Cox model fitting for the rheumatoid arthritis patients wasn't rejected, because Cox model fit for these data.

LITERATŪRA IR ŠALTINIAI

1. Elisa T.Lee, John Wenyu Wang. STATISTICAL METHODS FOR SURVIVAL DATA ANALYSIS. Third Edition. Department of Biostatistics and Epidemiology and Center for America Indian Health Research, College of Public Health, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma. 2003 m.
2. Interneto tinklapis http://en.wikipedia.org/wiki/Score_test.
3. Interneto tinklapis
http://www.ats.ucla.edu/STAT/SAS/faq/survival_repeated_events.htm
4. Interneto tinklapis
http://www.chfpatients.com/text/VO2max_to_predict_CHFer_survival.txt (Tyrimo su Stanford'o širdies implantų duomenimis aprašymas).
5. Interneto tinklapis <http://www.sas.com> .
6. Interneto tinklapis <http://www.siluteszinios.lt> (Straipsnis apie reumatoidinį artritą, jo ypatumus ir gydymą).
7. J. Klein, M.Moeschberger. SURVIVAL ANALYSIS, p. 64-70.
8. J. Kruopis. Matematinė statistika. Vilnius: Mokslas, 1993 m.
9. MEDICINA 2003 m., 39 tomas, Nr.5, p. 505 – 510.
10. P.K. Andersen, O.Borgan, R.D.Gill, N.Keiding. STATISTICAL MODELS BASE DON COUNTING PROCESSES, 1993 m., p. 135 – 361, 476 – 539.
11. Paul D. Allison. SURVIVAL ANALYSIS USING SAS: A PRACTICAL GUIDE, 2005 m.
12. Prof. Dr. Y. LI. SURVIVAL DATA ANALYSIS, 2005 m.
13. Rūta Levulienė. Statistika su SAS, Vilnius, 2007 m.
14. V. Bagdonavičius Cenzūruotųjų imčių analizė. Išgyvenamumo analizės konceptas I statistikos magistrantūros kursui, 2007 m.
15. V. Bagdonavičius, J. Kruopis. Matematinė statistika, I dalis. Vilnius: TEV 2007m.
16. V. Bagdonavičius, M. Nikulin straipsnis „On goodness-of-fit for homogeneity and proportional hazards“. Žurnalas Applied stochastic models in business and industry 2006 m., 22 tomas, Nr. 4.
17. V. Čekanavičius, G. Murauskas. Statistika ir jos taikymai II. Vilnius: TEV, 2002 m.

Priedas Nr. 1

SAS programos kodas

```
/* Programa */
libname mano 'C:\DAIVA\MAGISTRANTURA\Magistrinis';
data mano.analize_mag;
format mirt_date date9.;
set mano.ra_study (where=(RANK_DAR_old ne ''))
    rename=(RA_GIM=RA_GIM_old
            RANK_DAR=RANK_DAR_old
            GYD_DL_P=GYD_DL_P_old
            GYD_AUKS=GYD_AUKS_old
            GYD_PEN=GYD_PEN_old
            GYD_MTX=GYD_MTX_old
            AZA_IMUR=AZA_IMUR_old
            GYD_CYC=GYD_CYC_old
            GYD_SSL=GYD_SSL_old
            GYD_CSP=GYD_CSP_old
            PO_ST=PO_ST_old
            SINOVEKT=SINOVEKT_old
            KT_OP=KT_OP_old
            )
);
if MIRE='1' then vstatus=0;
else vstatus=1;
if LYTKOD=1 then LYTKOD=0; /* moteris */
else if LYTKOD=2 then LYTKOD=1; /* vyras */
if missing(KL_STRA) then KL_STRA=0;
if upcase(trim(compress(RA_GIM_old)))='NE' then RA_GIM=0;
else RA_GIM=1;
if upcase(trim(compress(RANK_DAR_old)))='NERANKU' then RANK_DAR=0;
else RANK_DAR=1;
if upcase(trim(compress(GYD_DL_P_old)))='NE' then GYD_DL_P=0;
else GYD_DL_P=1;
if upcase(trim(compress(GYD_AUKS_old)))='NE' then GYD_AUKS=0;
else GYD_AUKS=1;
if upcase(trim(compress(GYD_PEN_old)))='NE' then GYD_PEN=0;
else GYD_PEN=1;
if upcase(trim(compress(GYD_MTX_old)))='NE' then GYD_MTX=0;
else GYD_MTX=1;
if upcase(trim(compress(AZA_IMUR_old)))='NE' then AZA_IMUR=0;
else AZA_IMUR=1;
if upcase(trim(compress(GYD_CYC_old)))='NE' then GYD_CYC=0;
else GYD_CYC=1;
if upcase(trim(compress(GYD_SSL_old)))='NE' then GYD_SSL=0;
else GYD_SSL=1;
if upcase(trim(compress(GYD_CSP_old)))='NE' then GYD_CSP=0;
else GYD_CSP=1;
if upcase(trim(compress(PO_ST_old)))='NE' then PO_ST=0;
else PO_ST=1;
if upcase(trim(compress(SINOVEKT_old)))='NE' then SINOVEKT=0;
else SINOVEKT=1;
if upcase(trim(compress(KT_OP_old)))='NE' then KT_OP=0;
else KT_OP=1;
if upcase(trim(compress(FIZ_AKT_old)))='1' then FIZ_AKT= 1 ;
else if upcase(trim(compress(FIZ_AKT_old)))='2' then FIZ_AKT= 2 ;

else if upcase(trim(compress(FIZ_AKT_old)))='3' then FIZ_AKT= 3 ;

else if upcase(trim(compress(FIZ_AKT_old)))='4' then FIZ_AKT= 4 ;
```



```

else FIZ_AKT= 5 ;
ageacct = AMZDGN; /* ageacct-ligonio amzius, diagnozavus liga */
mirt_date=input(mirt_dat,yymmdd10.);
ageacct_2 = 1999 - GIMIMO_M; /* ageacct_2 -ligonio amzius tyrimo pradžioje */

if not missing(MIRT_DATE) then
do;
agels=year(MIRT_DATE)- GIMIMO_M;
/*agels-amzius paciento,kada jis paskutini karta buvo matytas-mirtis arba cenzuravimas*/
end;
else agels=2005- GIMIMO_M;
run;

proc phreg data=analyze_mag;
model LIG_TR*VSTATUS(0)= LYTKOD RA_GIM RANK_DAR AMZDGN GYD_DL_P
KT_OP GYD_PEN GYD_MTX AZA_IMUR GYD_CYC GYD_SSL GYD_CSP
PO_ST SINOVEKT GYD_AUKS FIZ_AKT
/
selection=stepwise slentry=0.25
slstay=0.15 details
;
/*run;*/
baseline out=a survival=s logsurv=ls loglogs=lls; /*Isgyvenamumo funkcijos iverciai*/
run;

proc print data=a;
run;
data grafikas;
set a;
ls=-ls;
run;

proc gplot data=grafikas;
symbol1 value=none interpol=join;
plot s*LIG_TR;
run;

proc phreg data= analyze_mag;
model (ageacct,agels)*VSTATUS(0)= LYTKOD RA_GIM RANK_DAR LIG_TR GYD_DL_P
KT_OP GYD_PEN GYD_MTX AZA_IMUR GYD_CYC GYD_SSL GYD_CSP
PO_ST SINOVEKT GYD_AUKS
/
selection=stepwise slentry=0.25
slstay=0.15 details;
run;
baseline out=a survival=s logsurv=ls loglogs=lls; /*Isgyvenamumo funkcijos iverciai*/
run;

proc print data=a;
run;
data grafikas;
set a;
ls=-ls;
run;

proc gplot data=grafikas;
symbol1 value=none interpol=join;
plot ls*agels;
run;

```

Priedas Nr. 2

SAS užklauso

```
/*UZKLAUSOS*/
```

```
/* 0. MIRUSIU SKAICIUS*/
```

```
proc sql;  
select count(LYTKOD) as visi_mire from mano.analize_mag where VSTATUS=1;  
run;
```

```
/* 1. VYRAI */
```

```
proc sql;  
select LYTKOD, VSTATUS from mano.analize_mag where LYTKOD = 1;  
run;
```

```
/* 2. KIEK VYRU DALYVAVO TYRIME */
```

```
proc sql;  
select count(LYTKOD) as vyrai from mano.analize_mag where LYTKOD=1;  
run;
```

```
/* 3. KIEK VYRU MIRE TYRIMO METU */
```

```
proc sql;  
select count(LYTKOD) as vyrai from mano.analize_mag where LYTKOD=1  
and VSTATUS=1;  
run;
```

```
/* 4. KIEK MOTERU DALYVAVO TYRIME */
```

```
proc sql;  
select count(LYTKOD) as moterys from mano.analize_mag where LYTKOD=0;  
run;
```

```
/* 5. KIEK MOTERU MIRE TYRIMO METU */
```

```
proc sql;  
select count(LYTKOD) as moterys from mano.analize_mag where LYTKOD=0  
and VSTATUS=1;  
run;
```

```
/* 6. MIRUSIU VYRU AMZIUS */
```

```
proc sql;  
select LYTKOD, AMZIUS, RA_GIM, RANK_DAR from mano.analize_mag  
where LYTKOD=1  
and VSTATUS=1;  
run;
```

```
/* 7. MIRUSIU MOTERU AMZIUS */
```

```
proc sql;  
select LYTKOD, AMZIUS, RA_GIM, RANK_DAR from mano.analize_mag  
where LYTKOD=0  
and VSTATUS=1;  
run;
```

```
/* 8. LIGONIU AMZIAUS VIDURKIS */
```

```
proc sql;  
select (sum(amzius))/531 as amz_vid from mano.analize_mag;  
run;
```

```
/* 9. LIGONIU VIDUTINE LIGOS TRUKME */
```

```
proc sql;  
select (sum(LIG_TR))/531 as vid_lig_tr from mano.analize_mag;
```

```

run;

/* 10. KIEK LIGONIU BUVO DARBINGI?*/
proc sql;
select count(DIRB_NE) as darbingi from mano.analize_mag
where DIRB_NE='dirbate';
run;

/* 11. MIRUSIU LIGONIU AMZIAUS VIDURKIS*/
proc sql;
select (sum(AMZIUS))/19 as vid_amz_mir from mano.analize_mag
where vstatus=1;
run;

/* 12. AMZIAUS, KURIAME LIGA BUVO DIAGNOZUOTA, VIDURKIS (t.y. AMZDGN VIDURKIS)*/
proc sql;
select (sum(AMZDGN))/531 as AMZDGN_VID from mano.analize_mag;
run;

/* 13. MIRUSIU VYRU AMZIAUS VIDURKIS*/
proc sql;
select (sum(AMZIUS))/8 as vyru_vid_amz_mir from mano.analize_mag
where vstatus=1 and LYTKOD=1;
run;

/* 14. MIRUSIU MOTERU AMZIAUS VIDURKIS*/
proc sql;
select (sum(AMZIUS))/24 as moteru_vid_amz_mir from mano.analize_mag
where vstatus=1 and LYTKOD=0;
run;

```

Priedas Nr. 3

χ^2 kriterijus

χ^2 kriterijus yra universalus. Jis nustato, ar yra ryšys tarp kintamųjų, bet nesuteikia informacijos apie to ryšio stiprumą. χ^2 kriterijus naudojamas *suderinamumo, homogeniškumo ir nepriklausomumo* hipotezėms tikrinti.

Paprastosios suderinamumo hipotezės tikrinimas.

Tarkime, kad $X = (x_1, x_2, x_3, x_4, \dots, x_n)$ yra atsitiktinio dydžio X imtis. Tegu a.d. X skirstinys priklauso šeimai $P = \{F, F \in \mathcal{F}\}$.

Tikriname paprastąją hipotezę $H_0 : F(x) = F_0(x), x \in R$,

čia $F(x)$ yra X pasiskirstymo funkcija, $F_0(x)$ – pilnai nusakyta šeimos F pasiskirstymo funkcija.

Tarkime, kad atliekame bandymą, kurio metu įvyksta vienas iš nesutaikomų įvykių A_1, A_2, \dots, A_k , sudarančių pilnąją įvykių grupę. Kiekviename bandyme tikimybė, kad įvyks įvykis A_i lygi p_i ($i=1, \dots, k$, $p_1 + \dots + p_k = 1$) ir nepriklauso nuo kitų bandymų rezultatų. Atlikus n bandymų, įvykusių įvykių A_1, A_2, \dots, A_k skaičių pažymėkime V_1, V_2, \dots, V_k . Tada atsitiktinių dydžių V_1, V_2, \dots, V_k ($V_1 + V_2 + \dots + V_k = n$) tikimybinius skirstinys yra polinominis:

$$P\{V_1 = n_1, V_2 = n_2, \dots, V_k = n_k\} = \frac{n!}{n_1! n_2! \dots n_k!} p_1^{n_1} p_2^{n_2} \dots p_k^{n_k}.$$

Hipotezė $H_0 : F(x) = F_0(x), x \in R$, keičiama hipoteze $H_0' : p_i = p_i^0, i = 1, \dots, k$.

Alternatyva: $H_1' : p_i \neq p_i^0$, bent vienam $i = 1, \dots, k$.

Jeigu hipotezė H_0' teisinga, tai statistika $X^2 = \sum_{i=1}^k \frac{(V_i - np_i^0)^2}{np_i^0} = \sum_{i=1}^k \frac{V_i^2}{np_i^0} - n$ asimptotiškai

(t.y. kai $n \rightarrow \infty$) pasiskirsčiusi pagal chi-kvadrato skirstinį su $k-1$ laisvės laipsniu. [12,13]

Priedas Nr. 4

Taškų testas (Score test)

Taškų testas yra statistinis testas, naudojamas paprastų hipotezių tikrinimui, pavyzdžiui, paprastos nulinės hipotezės, kad tiriamas parametras lygus kokiam nors reikšmei, t.y. $H_0 : \theta = \theta_0$:

$$\left(\frac{d \log L(\theta | x)}{d\theta} \right)_{\theta=\theta_0} \geq C,$$

kur L – tikėtinumo funkcija, θ_0 – dominančio parametro reikšmė pagal nulinę hipotezę, C – konstantų aibė, priklausanti nuo testo reikšmingumo, t.y. tikimybė atmesti nulinę hipotezę, kai ji yra teisinga (I rūšies klaida).

Taškų testas yra labai reikšmingas, tinkamiausias mažiems nuokrypiams nuo nulinės hipotezės. Įsitikinimui patikrinkime $\begin{cases} H_0 : \theta = \theta_0, \\ H_1 : \theta = \theta_0 + h. \end{cases}$. Remiantis Niumano – Pirsono lema,

reikšmingiausio testo forma yra: $\frac{L(\theta_0 + h | x)}{L(\theta_0 | x)} \geq K$. Logaritmuojant abi nelygybės puses, gauname

$$\log L(\theta_0 + h | x) - \log L(\theta_0 | x) \geq \log K.$$

Yra galima tokia taškų testo aproksimacija

$$\log L(\theta_0 + h | x) \approx \log L(\theta_0 | x) + h \cdot \left(\frac{d \log L(\theta | x)}{d\theta} \right)_{\theta=\theta_0},$$

vietoj C imant $\log(K)$.

Daugelio parametrų atvejis.

Bendresnis taškų testas gaunamas, kai skirstinys turi keletą parametrų. Tarkime, kad $\bar{\theta}_0$ yra parametro θ maksimalaus tikėtinumo įvertis, kai teisinga nulinė hipotezė. Tada $U'(\bar{\theta}_0) I^{-1}(\bar{\theta}_0) U(\bar{\theta}_0) \sim \chi_k^2$, kai teisinga nulinė hipotezė. Čia k – laisvės laipsnių skaičius,

$$U(\bar{\theta}_0) = \frac{d \log L(\bar{\theta}_0 | x)}{d\theta}, \quad I(\bar{\theta}_0) = -\frac{d^2 \log L(\bar{\theta}_0 | x)}{d\theta d\theta'}. \quad [9]$$

Priedas Nr. 5

Pažingsninė regresija (Stepwise regression)

Apskritai, regresinėje analizėje sprendžiami vienu kintamųjų (nepriklausomų) prognozavimo, remiantis kitais kintamaisiais (priklausomais) uždaviniai. Prognozavimas būtinas sprendžiant praktines problemas, uždavinius. Pavyzdžiui, meteorologas prognozuoja orą ateičiai, remdamasis tam tikrų atmosferos požymių matavimais praeityje. Taip pat gydytojas prognozuoja ligos eigą, remdamasis tam tikrų ligos simptomų matavimais ir pan.

Dažnai, taikydamas regresinę analizę, tyrėjas neturi pakankamai informacijos apie tai, kurie iš kintamųjų X_1, X_2, \dots, X_k labiau tinka a.d. Y prognozuoti. Be to, galutinai parinktas modelis yra tuo vertingesnis, kuo mažesniu skaičiumi parametrų jis nusakomas. Suprantama, stengiamasi, kad kintamųjų skaičiaus X_1, X_2, \dots, X_k sumažinimas iš esmės nepablogintų prognozės tikslumo.

Matematinėje statistikoje taikoma daug įvairių metodų, leidžiančių sumažinti vektoriaus (X_1, X_2, \dots, X_k) matavimo dydį iš esmės nepabloginant prognozės tikslumo. Vienas iš paprastesnių ir pakankamai dažnai taikomų metodų tokiai problemai spręsti ir yra *Pažingsninė regresija*.

Tarkime, vektorius $(Y, X_1, X_2, \dots, X_k)$ yra pasiskirstęs pagal $(k+1)$ -matį normalųjį dėsnį. Stebint šį atsitiktinį vektorių, gauta tūrio n imtis $(Y_i, X_{1i}, X_{2i}, \dots, X_{ki})$, $i = 1, 2, \dots, n$. Kadangi tiesinio prognozavimo tikslumo matas yra suvestinis koreliacijos koeficientas, tai būtų natūralu imti tokį kintamųjų X_1, X_2, \dots, X_k poaibį, kad suvestinio koreliacijos koeficiento įvertis būtų didžiausias. Tačiau tokio metodo negalime realizuoti praktiškai, jeigu kintamųjų skaičius k yra pakankamai didelis. Pavyzdžiui, jeigu $k = 10$, tai galima sudaryti $2^{10} - 1 = 1023$ skirtingų tiesinės regresijos lygčių. Kaip jau buvo minėta, vienas paprasčiausių ir dažniausiai praktikoje taikomų metodų yra *Pažingsninė regresija*.

Nulinis žingsnis. Parenkame du reikšmingumo lygmenis P' ir P , taip, kad būtų $P' > P$, randame koreliacijos koeficientų tarp a.d. Y ir X_i įverčius $r(Y, X_i) = r_{YX_i} = r_{0i}$, $i = 1, 2, \dots, k$.

Pirmasis žingsnis. Išrenkame maksimalų ivertį $\max_{1 \leq i \leq k} (r_{0i}) = r_{0i}$. Kintamasis X_{i_1} įtraukiamas į kintamųjų Y prognozuoti sąrašą, jeigu teisinga nelygybė $F_{0i_1} = r_{0i_1}^2 (n-2) / (1 - r_{0i_1}^2) > F_p(1, n-2)$ (1). Priešingu atveju daroma išvada, kad Y prognozavimas pagal X_1, X_2, \dots, X_k negalimas.

Antrasis žingsnis. Randame dalinių koreliacijos koeficientų iverčius $r_{YX_i \cdot (X_{i_1})} = r_{0i \cdot (i_1)}$, $i = 1, 2, \dots, k$, $i \neq i_1$, ir iš jų išrenkame maksimalų: $\max_{i \neq i_1} (r_{0i \cdot (i_1)}) = r_{0i_2 \cdot (i_1)}$. Kintamasis X_{i_2} įtraukiamas į sąrašą, jeigu teisinga nelygybė $r_{0i_2 \cdot (i_1)}^2 (n-3) / (1 - r_{0i_2 \cdot (i_1)}^2) > F_p(1, n-3)$ (2). Priešingu atveju sąraše paliekamas tik kintamasis X_{i_1} . Jei ką tik užrašyta (2) nelygybė teisinga, tai tikrinama, ar nereikia išbraukti iš sąrašo X_{i_1} , kai įrašytas kintamasis X_{i_2} . Kintamasis X_{i_1} išbraukiamas, jeigu teisinga nelygybė $r_{0i_1 \cdot (i_2)}^2 (n-3) / (1 - r_{0i_1 \cdot (i_2)}^2) < F_p(1, n-3)$.

Taigi po antrojo žingsnio a.d. Y prognozuoti gali likti tik kintamasis X_{i_1} (jei (1) nelygybė neteisinga); tik kintamasis X_{i_2} (jeigu (1), (2) nelygybės tenkinamos); vektorius (X_{i_1}, X_{i_2}) (jei (1) nelygybė teisinga, o (2) neteisinga). Pirmuoju atveju procesas užbaigiamas; antruoju ir trečiuoju atvejais pereinama prie trečiojo žingsnio.

Trečiasis, ketvirtasis, ... žingsniai. Rekurentiškai kartojamas antrasis žingsnis. Tiksliau, remiantis dalinių koreliacijos iverčių reikšmėmis, pagal taisykles, analogiškas (1), daroma išvada, ar reikia papildyti kintamųjų sąrašą dar vienu kintamuoju. Irašius naują kintamąjį, tikrinama, ar nereikia kurio nors kintamojo iš sąrašo išbraukti. Procesas užbaigiamas, kai nelygybės, atitinkančios (1) ir (2), netenkinamos. [12,13]

Priedas Nr. 6

Vėžiu sirgusių ligonių duomenys ir SAS programos kodas Cox'o modeliui.

```
/*SAS programa Cox'o modeliui*/
```

```
data chemoterapija_2 ;      /* delta_i : 0 - cenzuruota, 1 - necenzuruota */
input X delta_i x_terapija ; /* x_terapija : 0 – chemo, 1 – chemo+radio */
datalines ;
1 1 0
63 1 0
105 1 0
129 1 0
182 1 0
216 1 0
250 1 0
262 1 0
301 1 0
301 1 0
342 1 0
354 1 0
356 1 0
358 1 0
380 1 0
383 1 0
383 1 0
388 1 0
394 1 0
408 1 0
460 1 0
489 1 0
499 1 0
523 1 0
524 1 0
535 1 0
562 1 0
569 1 0
675 1 0
676 1 0
748 1 0
778 1 0
786 1 0
797 1 0
955 1 0
968 1 0
1000 1 0
1245 1 0
1271 1 0
1420 1 0
1551 1 0
1694 1 0
2363 1 0
2754 0 0
2950 0 0
17 1 1
42 1 1
44 1 1
48 1 1
60 1 1
72 1 1
74 1 1
```



```
95 1 1
103 1 1
108 1 1
122 1 1
144 1 1
167 1 1
170 1 1
183 1 1
185 1 1
193 1 1
195 1 1
197 1 1
208 1 1
234 1 1
235 1 1
254 1 1
307 1 1
315 1 1
401 1 1
445 1 1
464 1 1
484 1 1
528 1 1
542 1 1
567 1 1
577 1 1
580 1 1
795 1 1
855 1 1
1366 1 1
1577 1 1
2060 1 1
2412 0 1
2486 0 1
2796 0 1
2802 0 1
2934 0 1
2988 0 1
```

```
;
```

```
run ;
```

```
proc phreg data=chemoterapija;
```

```
  model X *delta_i(0)= x_terapija;
```

```
  baseline out=a survival=s logsurv=ls loglogs=lls; /*Isgyvenamumo funkcijos iverciai*/
```

```
run;
```

Priedas Nr. 7

Statistikos T skaičiavimas SCE modelio atveju Chemoterapijos duomenims.

```
/* SCE modelis */

%let n = 90;
%let beta = 0.10589;

proc sort data = chemoterapija_2;
by descending X ;
run;

data modelis_SCE;
set chemoterapija_2;
beta_x = &beta * x_terapija;
e_beta_x = exp(beta_x);
run;

data modelis_SCE;
set modelis_SCE;
retain S_0(0);
S_0 = S_0 + e_beta_x;
run;

proc sort data = modelis_SCE;
by X;
run;

data modelis_SCE;
set modelis_SCE;
S_0_1 = 1 / S_0;
run;

data modelis_SCE;
set modelis_SCE;
retain Lambda_iv(0);
if delta_i = 1 then
Lambda_iv = Lambda_iv + S_0_1;
run;

data modelis_SCE;
set modelis_SCE;
daugyba = e_beta_x * Lambda_iv;
argumentas = 1 + daugyba;
logaritmas = log(argumentas);
h = - x_terapija * logaritmas;
run;

proc sort data = modelis_SCE;
by descending X;
run;

proc iml;
use modelis_sce;
read all var{x_terapija e_beta_x Lambda_iv} into M;
S_1_zv = j(&n,1,0);
S_2_zv = j(&n,1,0);
S_2_zv_zv = j(&n,1,0);
do i = 1 to &n ;
do l = 1 to i;
```

```

S_1_zv[i] = S_1_zv[i] + ( M[l,1] * M[l,2] * log(1 + M[l,2]*M[i,3]) );
S_2_zv[i] = S_2_zv[i] + ( M[l,1]*M[l,1] * M[l,2] * log(1 + M[l,2]*M[i,3]) );
S_2_zv_zv[i] = S_2_zv_zv[i] + ( M[l,1]*M[l,1] * M[l,2] * (log(1 + M[l,2]*M[i,3]))*(log(1 + M[l,2]*M[i,3])) );
end;
end;
create mano_S_1_zv from S_1_zv[colname='S_1_zv'];
append from S_1_zv;
create mano_S_2_zv from S_2_zv[colname='S_2_zv'];
append from S_2_zv;
create mano_S_2_zv_zv from S_2_zv_zv[colname='S_2_zv_zv'];
append from S_2_zv_zv;
run;

proc iml;
create mano_S_1_zv from S_1_zv[colname='S_1_zv'];
append from S_1_zv;
run;

proc iml;
create mano_S_2_zv from S_2_zv[colname='S_2_zv'];
append from S_2_zv;
run;

proc iml;
create mano_S_2_zv_zv from S_2_zv_zv[colname='S_2_zv_zv'];
append from S_2_zv_zv;
run;

data sujungimas;
merge modelis_sce mano_S_1_zv mano_S_2_zv mano_S_2_zv_zv;
run;

data sujungimas;
set sujungimas;
S_1_zv_ = - S_1_zv;
E_zv = S_1_zv_ / S_0;
run;

data sujungimas;
set sujungimas;
retain S_1(0);
S_1 = S_1 + (x_terapija * e_beta_x);
run;

data sujungimas;
set sujungimas;
E = S_1 / S_0;
run;

data sujungimas;
set sujungimas;
S_2_zv_ = - S_2_zv;
V_zv = (S_2_zv_ / S_0) - (E * E_zv);
V_zv_zv = (S_2_zv_zv / S_0) - (E_zv)**2;
run;

data sujungimas;
set sujungimas;
retain suma_zv(0);
if delta_j = 1 then
suma_zv = suma_zv + V_zv;
run;

```

```

data sujungimas;
set sujungimas;
Sigma_zv = (1/&n)* suma_zv ;
run;

```

```

data sujungimas;
set sujungimas;
retain suma_zv_zv(0);
if delta_i = 1 then
suma_zv_zv = suma_zv_zv + V_zv_zv;
run;

```

```

data sujungimas;
set sujungimas;
Sigma_zv_zv = (1/&n) * suma_zv_zv ;
run;

```

```

data sujungimas;
set sujungimas;
retain S_2(0);
S_2 = S_2 + (x_terapija**2 * e_beta_x);
run;

```

```

data sujungimas;
set sujungimas;
V = (S_2 / S_0) - E**2 ;
run;

```

```

data sujungimas;
set sujungimas;
retain suma(0);
if delta_i = 1 then
suma = suma + V;
run;

```

```

data sujungimas;
set sujungimas;
Sigma_0 = (1/&n) * suma;
run;

```

```

data sujungimas;
set sujungimas;
retain U(0);
if delta_i = 1 then
U = U + (h - E_zv);
run;

```

```

proc iml;
use sujungimas;
read point {&n} var {Sigma_zv} into Sigma_zv_iv;
print Sigma_zv_iv;
read point {&n} var {Sigma_zv_zv} into Sigma_zv_zv_iv;
print Sigma_zv_zv_iv;
read point {&n} var {Sigma_0} into Sigma_0_iv;
print Sigma_0_iv;
read point {&n} var {U} into U_iv;
print U_iv;
D = Sigma_zv_zv_iv - (Sigma_zv_iv**2 / Sigma_0_iv);
print D;
T = (1/&n) * (U_iv**2 / D);
print T;
quit;
run;

```

Priedas Nr. 8

Statistikos T skaičiavimas SCE modelio atveju Reumatoidinio artrito ligonių duomenų imčiai.

```
/* SCE modelis */
```

```
data mano.nupjovimas_sce;  
set mano.nupjovimas_sce;  
keep LIG_TR VSTATUS LYTKOD AMZDGN AZA_IMUR ;  
run;
```

```
data mano.nupjovimas_sce;  
set mano.nupjovimas_sce;  
rename AMZDGN = X  
VSTATUS = delta_i  
LIG_TR = D_i  
LYTKOD = kint1  
AZA_IMUR = kint2 ;  
run;
```

```
proc sort data = mano.nupjovimas_sce;  
by descending X;  
run;
```

```
%let n = 531;  
%let beta_1=0.65838; /* LYTKOD */  
%let beta_2=0.69456; /* AZA_IMUR */
```

```
data mano.nupjovimas_sce;  
set mano.nupjovimas_sce;  
beta_x = &beta_1*kint1 + &beta_2*kint2;  
e_beta_x = exp(beta_x);  
run;
```

```
proc sort data = mano.nupjovimas_sce;  
by descending X ;  
run;
```

```
data mano.nupjovimas_sce;  
set mano.nupjovimas_sce;  
retain S_0(0);  
if D_i < X then  
S_0 = S_0 + e_beta_x;  
run;
```

```
proc sort data = mano.nupjovimas_sce;  
by X;  
run;
```

```
data mano.nupjovimas_sce;  
set mano.nupjovimas_sce;  
S_0_1 = 1 / S_0;  
run;
```

```
data mano.nupjovimas_sce;  
set mano.nupjovimas_sce;  
retain Lambda_iv(0);  
if delta_i = 1 then  
Lambda_iv = Lambda_iv + S_0_1;  
run;
```

```
data mano.nupjovimas_sce;
```

```

set mano.nupjovimas_sce;
daugyba = e_beta_x * Lambda_iv;
argumentas = 1 + daugyba;
logaritmas = log(argumentas);
run;

proc sort data = mano.nupjovimas_sce;
by descending X;
run;

data mano.nupjovimas_sce;
set mano.nupjovimas_sce;
h_1 = -kint1 * logaritmas;
h_2 = -kint2 * logaritmas;
run;

proc iml;
use mano.nupjovimas_sce;
read all var{h_1 h_2} into h;
run;

data mano.nupjovimas_sce;
set mano.nupjovimas_sce;
retain S_1_zv_1(0);
if D_i < X then
S_1_zv_1 = S_1_zv_1 + (kint1 * e_beta_x * logaritmas);
run;

data mano.nupjovimas_sce;
set mano.nupjovimas_sce;
retain S_1_zv_2(0);
if D_i < X then
S_1_zv_2 = S_1_zv_2 + (kint2 * e_beta_x * logaritmas);
run;

data mano.nupjovimas_sce;
set mano.nupjovimas_sce;
S_1_zv_1_m = - S_1_zv_1;
S_1_zv_2_m = - S_1_zv_2;
run;

proc iml;
use mano.nupjovimas_sce;
read all var{S_1_zv_1_m S_1_zv_2_m} into S_1_zv;
run;

data mano.nupjovimas_sce;
set mano.nupjovimas_sce;
E_zv_1 = S_1_zv_1_m / S_0;
E_zv_2 = S_1_zv_2_m / S_0;
run;

proc iml;
use mano.nupjovimas_sce;
read all var{E_zv_1 E_zv_2} into E_zv;
run;

data mano.nupjovimas_sce;
set mano.nupjovimas_sce;
retain S_1_1(0);
if D_i < X then
S_1_1 = S_1_1 + (kint1 * e_beta_x);
run;

```

```

data mano.nupjovimas_sce;
set mano.nupjovimas_sce;
retain S_1_2(0);
if D_i < X then
S_1_2 = S_1_2 + (kint2 * e_beta_x);
run;

```

```

proc iml;
use mano.nupjovimas_sce;
read all var{S_1_1 S_1_2} into S_1;
run;

```

```

data mano.nupjovimas_sce;
set mano.nupjovimas_sce;
E_1 = S_1_1 / S_0;
E_2 = S_1_2 / S_0;
run;

```

```

proc iml;
use mano.nupjovimas_sce;
read all var{E_1 E_2} into E;
run;

```

```

data mano.nupjovimas_sce;
set mano.nupjovimas_sce;
retain S_2_zv_1_1(0);
if D_i < X then
S_2_zv_1_1 = S_2_zv_1_1 + (kint1**2 * e_beta_x * logaritmas);
run;

```

```

data mano.nupjovimas_sce;
set mano.nupjovimas_sce;
retain S_2_zv_1_2(0);
if D_i < X then
S_2_zv_1_2 = S_2_zv_1_2 + (kint1*kint2 * e_beta_x * logaritmas);
run;

```

```

data mano.nupjovimas_sce;
set mano.nupjovimas_sce;
retain S_2_zv_2_1(0);
if D_i < X then
S_2_zv_2_1 = S_2_zv_2_1 + (kint1*kint2 * e_beta_x * logaritmas);
run;

```

```

data mano.nupjovimas_sce;
set mano.nupjovimas_sce;
retain S_2_zv_2_2(0);
if D_i < X then
S_2_zv_2_2 = S_2_zv_2_2 + (kint2**2 * e_beta_x * logaritmas);
run;

```

```

data mano.nupjovimas_sce;
set mano.nupjovimas_sce;
S_2_zv_1_1_m = - S_2_zv_1_1;
S_2_zv_1_2_m = - S_2_zv_1_2;
S_2_zv_2_1_m = - S_2_zv_2_1;
S_2_zv_2_2_m = - S_2_zv_2_2;
run;

```

```

proc iml;
use mano.nupjovimas_sce;
read all var{S_2_zv_1_1_m S_2_zv_1_2_m S_2_zv_2_1_m S_2_zv_2_2_m} into S_2_zv;

```

```

run;

data mano.nupjovimas_sce;
set mano.nupjovimas_sce;
retain S_2_zv_zv_1_1(0);
if D_i < X then
S_2_zv_zv_1_1 = S_2_zv_zv_1_1 + (kint1**2 * e_beta_x * logaritmas**2);
run;

data mano.nupjovimas_sce;
set mano.nupjovimas_sce;
retain S_2_zv_zv_1_2(0);
if D_i < X then
S_2_zv_zv_1_2 = S_2_zv_zv_1_2 + (kint1*kint2 * e_beta_x * logaritmas**2);
run;

data mano.nupjovimas_sce;
set mano.nupjovimas_sce;
retain S_2_zv_zv_2_1(0);
if D_i < X then
S_2_zv_zv_2_1 = S_2_zv_zv_2_1 + (kint1*kint2 * e_beta_x * logaritmas**2);
run;

data mano.nupjovimas_sce;
set mano.nupjovimas_sce;
retain S_2_zv_zv_2_2(0);
if D_i < X then
S_2_zv_zv_2_2 = S_2_zv_zv_2_2 + (kint2**2 * e_beta_x * logaritmas**2);
run;

proc iml;
use mano.nupjovimas_sce;
read all var{S_2_zv_zv_1_1 S_2_zv_zv_1_2 S_2_zv_zv_2_1 S_2_zv_zv_2_2} into S_2_zv_zv;
run;

data mano.nupjovimas_sce;
set mano.nupjovimas_sce;
V_zv_1_1 = (S_2_zv_1_1_m / S_0) - (E_zv_1 * E_1);
V_zv_1_2 = (S_2_zv_1_2_m / S_0) - (E_zv_1 * E_2);
V_zv_2_1 = (S_2_zv_2_1_m / S_0) - (E_zv_2 * E_1);
V_zv_2_2 = (S_2_zv_2_2_m / S_0) - (E_zv_2 * E_2);

V_zv_zv_1_1 = (S_2_zv_zv_1_1 / S_0) - (E_zv_1)**2;
V_zv_zv_1_2 = (S_2_zv_zv_1_2 / S_0) - (E_zv_1 * E_zv_2);
V_zv_zv_2_1 = (S_2_zv_zv_2_1 / S_0) - (E_zv_1 * E_zv_2);
V_zv_zv_2_2 = (S_2_zv_zv_2_2 / S_0) - (E_zv_2)**2;
run;

proc iml;
use mano.nupjovimas_sce;
read all var{V_zv_1_1 V_zv_1_2 V_zv_2_1 V_zv_2_2} into V_zv;
run;

proc iml;
use mano.nupjovimas_sce;
read all var{V_zv_zv_1_1 V_zv_zv_1_2 V_zv_zv_2_1 V_zv_zv_2_2} into V_zv_zv;
run;

data mano.nupjovimas_sce;
set mano.nupjovimas_sce;
retain suma_zv_1_1(0);
if delta_i = 1 then
suma_zv_1_1 = suma_zv_1_1 + V_zv_1_1;

```


run;

```
data mano.nupjovimas_sce;  
set mano.nupjovimas_sce;  
retain suma_zv_1_2(0);  
if delta_i = 1 then  
suma_zv_1_2 = suma_zv_1_2 + V_zv_1_2;  
run;
```

```
data mano.nupjovimas_sce;  
set mano.nupjovimas_sce;  
retain suma_zv_2_1(0);  
if delta_i = 1 then  
suma_zv_2_1 = suma_zv_2_1 + V_zv_2_1;  
run;
```

```
data mano.nupjovimas_sce;  
set mano.nupjovimas_sce;  
retain suma_zv_2_2(0);  
if delta_i = 1 then  
suma_zv_2_2 = suma_zv_2_2 + V_zv_2_2;  
run;
```

```
data mano.nupjovimas_sce;  
set mano.nupjovimas_sce;  
Sigma_zv_1_1 = (1/&n) * suma_zv_1_1;  
Sigma_zv_1_2 = (1/&n) * suma_zv_1_2;  
Sigma_zv_2_1 = (1/&n) * suma_zv_2_1;  
Sigma_zv_2_2 = (1/&n) * suma_zv_2_2;  
run;
```

```
proc iml;  
use mano.nupjovimas_sce;  
read all var{Sigma_zv_1_1 Sigma_zv_1_2 Sigma_zv_2_1 Sigma_zv_2_2} into Sigma_zv;  
run;
```

```
data mano.nupjovimas_sce;  
set mano.nupjovimas_sce;  
retain suma_zv_zv_1_1(0);  
if delta_i = 1 then  
suma_zv_zv_1_1 = suma_zv_zv_1_1 + V_zv_zv_1_1;  
run;
```

```
data mano.nupjovimas_sce;  
set mano.nupjovimas_sce;  
retain suma_zv_zv_1_2(0);  
if delta_i = 1 then  
suma_zv_zv_1_2 = suma_zv_zv_1_2 + V_zv_zv_1_2;  
run;
```

```
data mano.nupjovimas_sce;  
set mano.nupjovimas_sce;  
retain suma_zv_zv_2_1(0);  
if delta_i = 1 then  
suma_zv_zv_2_1 = suma_zv_zv_2_1 + V_zv_zv_2_1;  
run;
```

```
data mano.nupjovimas_sce;  
set mano.nupjovimas_sce;  
retain suma_zv_zv_2_2(0);  
if delta_i = 1 then  
suma_zv_zv_2_2 = suma_zv_zv_2_2 + V_zv_zv_2_2;  
run;
```

```

data mano.nupjovimas_sce;
set mano.nupjovimas_sce;
Sigma_zv_zv_1_1 = (1/&n) * suma_zv_zv_1_1;
Sigma_zv_zv_1_2 = (1/&n) * suma_zv_zv_1_2;
Sigma_zv_zv_2_1 = (1/&n) * suma_zv_zv_2_1;
Sigma_zv_zv_2_2 = (1/&n) * suma_zv_zv_2_2;
run;

proc iml;
use mano.nupjovimas_sce;
read all var{Sigma_zv_zv_1_1 Sigma_zv_zv_1_2 Sigma_zv_zv_2_1 Sigma_zv_zv_2_2} into Sigma_zv_zv;
run;

data mano.nupjovimas_sce;
set mano.nupjovimas_sce;
retain S_2_1_1(0);
if D_i < X then
S_2_1_1 = S_2_1_1 + (kint1**2 * e_beta_x);
run;

data mano.nupjovimas_sce;
set mano.nupjovimas_sce;
retain S_2_1_2(0);
if D_i < X then
S_2_1_2 = S_2_1_2 + (kint1*kint2 * e_beta_x);
run;

data mano.nupjovimas_sce;
set mano.nupjovimas_sce;
retain S_2_2_1(0);
if D_i < X then
S_2_2_1 = S_2_2_1 + (kint1*kint2 * e_beta_x);
run;

data mano.nupjovimas_sce;
set mano.nupjovimas_sce;
retain S_2_2_2(0);
if D_i < X then
S_2_2_2 = S_2_2_2 + (kint2**2 * e_beta_x);
run;

proc iml;
use mano.nupjovimas_sce;
read all var{S_2_1_1 S_2_1_2 S_2_2_1 S_2_2_2} into S_2;
run;

data mano.nupjovimas_sce;
set mano.nupjovimas_sce;
V_1_1 = (S_2_1_1 / S_0) - E_1**2 ;
V_1_2 = (S_2_1_2 / S_0) - E_1*E_2 ;
V_2_1 = (S_2_2_1 / S_0) - E_1*E_2 ;
V_2_2 = (S_2_2_2 / S_0) - E_2**2 ;
run;

proc iml;
use mano.nupjovimas_sce;
read all var{V_1_1 V_1_2 V_2_1 V_2_2} into V;
run;

data mano.nupjovimas_sce;
set mano.nupjovimas_sce;
retain suma_1_1(0);

```

```

if delta_i = 1 then
suma_1_1 = suma_1_1 + V_1_1;
run;

data mano.nupjovimas_sce;
set mano.nupjovimas_sce;
retain suma_1_2(0);
if delta_i = 1 then
suma_1_2 = suma_1_2 + V_1_2;
run;

data mano.nupjovimas_sce;
set mano.nupjovimas_sce;
retain suma_2_1(0);
if delta_i = 1 then
suma_2_1 = suma_2_1 + V_2_1;
run;

data mano.nupjovimas_sce;
set mano.nupjovimas_sce;
retain suma_2_2(0);
if delta_i = 1 then
suma_2_2 = suma_2_2 + V_2_2;
run;

data mano.nupjovimas_sce;
set mano.nupjovimas_sce;
Sigma_0_1_1 = (1/&n) * suma_1_1;
Sigma_0_1_2 = (1/&n) * suma_1_2;
Sigma_0_2_1 = (1/&n) * suma_2_1;
Sigma_0_2_2 = (1/&n) * suma_2_2;
run;

proc iml;
use mano.nupjovimas_sce;
read all var{Sigma_0_1_1 Sigma_0_1_2 Sigma_0_2_1 Sigma_0_2_2} into Sigma_0;
run;

data mano.nupjovimas_sce;
set mano.nupjovimas_sce;
retain U_1(0);
if delta_i = 1 then
U_1 = U_1 + (h_1 - E_zv_1);
run;

data mano.nupjovimas_sce;
set mano.nupjovimas_sce;
retain U_2(0);
if delta_i = 1 then
U_2 = U_2 + (h_2 - E_zv_2);
run;

proc iml;
use mano.nupjovimas_sce;
read all var{U_1 U_2} into U;
run;

proc iml;
use mano.nupjovimas_sce;
read point {&n} var {Sigma_zv_1_1 Sigma_zv_1_2} into Sigma_zv_iv_1;
print Sigma_zv_iv_1;
read point {&n} var {Sigma_zv_2_1 Sigma_zv_2_2} into Sigma_zv_iv_2;
print Sigma_zv_iv_2;

```

```

Sigma_zv_iv = { -0.047888 0.0019762, 0.0017193 -0.057803};
print Sigma_zv_iv;
read point {&n} var {Sigma_zv_zv_1_1 Sigma_zv_zv_1_2} into Sigma_zv_zv_iv_1;
print Sigma_zv_zv_iv_1;
read point {&n} var {Sigma_zv_zv_2_1 Sigma_zv_zv_2_2} into Sigma_zv_zv_iv_2;
print Sigma_zv_zv_iv_2;
Sigma_zv_zv_iv = { 0.0376728 -0.002102, -0.002102 0.0448256};
print Sigma_zv_zv_iv;
read point {&n} var {Sigma_0_1_1 Sigma_0_1_2} into Sigma_0_iv_1;
print Sigma_0_iv_1;
read point {&n} var {Sigma_0_2_1 Sigma_0_2_2} into Sigma_0_iv_2;
print Sigma_0_iv_2;
Sigma_0_iv = { 0.0744863 0.0011083, 0.0011083 0.0955894};
print Sigma_0_iv;
read point {&n} var {U_1 U_2} into U_iv;
print U_iv; /* U_iv = { 0.4560317 -3.484931} */
D = Sigma_zv_zv_iv - (Sigma_zv_iv` * inv(Sigma_0_iv) * Sigma_zv_iv);
print D; /*D = { 0.0068232 0.0006401
0.0006401 0.009778}*/
T = (1/&n) # (U_iv * inv(D) * U_iv`);
print T;
run;

```

Priedas Nr. 9

Atsitiktinių dydžių $X = X(t)$, $t \geq 0$, apibrėžtų toje pačioje tikimybinėje erdvėje (Ω, F, P) , rinkinys vadinamas atsitiktiniu procesu.

σ algebrų rinkinys $F = \{F_t, t > 0\}$, toks, kad $F_s \subset F_t \subset F$ su visais $0 < s < t$, $s, t \in T \subset R$, vadinamas filtracija (aibėje T). Cenzūruotųjų imčių analizėje aibė T dažniausiai yra intervalas $[0, \tau)$ arba $[0, \tau]$, $\tau < \infty$.

Atsitiktinis procesas X suderintas su filtracija F, jei su bet kuriuo fiksuotu $t \geq 0$ atsitiktinis dydis $X(t)$ yra F_t matas, t.y. su bet kokia Borelio aibe $B \subset R$ įvykis $\{X(t) \in B\} \in F_t$.

F-suderintas procesas $\{X(t), t \geq 0\}$ vadinamas skaičiuojančiu, jei $X(0) = 0$, $X(t) < \infty$ b.t. su visais $t > 0$ ir jo trajektorijos yra tolydžios iš dešinės laiptuotos funkcijos su vienetiniais šuoliais. Pvz., atsitiktiniai procesai $N_i(t) = 1\{X_i \leq t, \delta_i = 1\} = 1\{T_i \leq t, T_i \leq C_i\}$ ir jų sumos yra skaičiuojantys procesai.

Atsitiktinis procesas $M(t)$, $t \in T$ vadinamas martingalu (submartingalu) filtracijos F atžvilgiu, jei

1. M yra F-suderintas;
2. $E|M(t)| < \infty$ su visais $t \in T$;
3. $E(M(t) | F_s) = M(s)$ ($E(M(t) | F_s) \geq M(s)$) b.t. su visais $t \geq s, s, t \in T$

F submartingalo M^2 kompensatorius $\langle M, M \rangle$ vadinamas numatoma F martingalo M variacija.

Atsitiktinis procesas $\langle M_1, M_2 \rangle = \frac{1}{4} (\langle M_1 + M_2, M_1 + M_2 \rangle - \langle M_1 - M_2, M_1 - M_2 \rangle)$ vadinamas martingalu M_1 ir M_2 numatoma kovariacija.

Dūbo-Mejerio išdėstymas.

Tarkime, kad X yra tolydus iš dešinės neneigiamas F submartingalas. Tada egzistuoja tolydus iš dešinės F martingalas M ir nemažėjantis tolydus iš dešinės numatomas procesas A, toks, kad $E(A(t)) < \infty$ ir

$$X(t) = M(t) + A(t) \text{ b.t.}$$

su visais $t \geq 0$. Jei $\Lambda(0) = 0$ b.t., tai šis išdėstymas b.t. vienintelis, t.y. jei $X^*(t) = M^*(t) + A^*(t)$ su visais $t \geq 0$, $A^*(0) = 0$ su visais $t \geq 0$,

$$P\{M^*(t) \neq M(t)\} = P\{A^*(t) \neq A(t)\}.$$

Atsitiktinis procesas A vadinamas submartingalo X kompensatoriumi.

Centrinė ribinė teorema (CRT) martingalams.

Nagrinėjamos sumos

$$U_j^{(n)}(t) = \sum_{i=1}^{k_n} \int_0^t H_{ij}^{(n)}(v) dM_i(v), j = 1, \dots, m \quad (*)$$

Čia $k_n = n$, arba $k_n = k \in N$, k yra fiksuotas skaičius, $M_i = N_i - A_i$, o A_i yra skaičiuojančių procesų N_i F kompensatoriai, H_{ij} - tolydūs iš kairės F suderinti procesai.

Integralų, turinčių pavidalą (*), ribos dažnai randamos naudojant CRT martingalamis. Ši teorema grindžiama Gauso procesų su nepriklausomais prieaugliais savybe.

Tarkime, kad su visais $t \in [0, \tau]$, $t < \infty$, neatsitiktinė matrica $\Sigma(t) = (\sigma_{jj'}(t))_{m \times m}$ yra teigiamai apibrėžta, $\sigma_{jj'}(0) = 0$.

Gauso procesų su nepriklausomais prieaugliais charakterizavimas.

m -matis atsitiktinis procesas $V = (V_1, \dots, V_m)$ yra Gauso procesas su komponentėmis, turinčiomis nepriklausomus prieauglius ir

$$E(V_j(t)) = 0 \quad \text{Cov}(V_j(s), V_{j'}(t)) = \sigma_{jj'}(s \wedge t) \quad (s, t \geq 0)$$

t. ir t. t., kai V_i yra integruojamo kvadrato martingalai su tolydžiomis trajektorijomis ir tokie, kad su visais $t \in [0, \tau]$

$$V_j(0) = 0, \quad \langle V_j, V_{j'} \rangle(t) = \sigma_{jj'}(t) \quad (j, j' = 1, \dots, m)$$

Tokiu būdu, jei martingalų $U_j^{(n)}$ šuoliai intervale $[0, \tau]$ artėja į nulį (Lindebergo sąlyga) ir

$$\langle U_j^{(n)}, U_{j'}^{(n)} \rangle(t) \xrightarrow{P} \sigma_{jj'}(t) \quad \text{su visais } t \in [0, \tau], \text{ kai } n \rightarrow \infty, \text{ tai galima tikėtis tokio}$$

konvergavimo:

$$U^{(n)} = (U_j^{(n)}, \dots, U_m^{(n)})^T \xrightarrow{D} V = (V_1, \dots, V_m)^T \text{ erdvėje } (D[0, \tau])^m.$$

Pažymėkime

$$U_{j\varepsilon}^{(n)}(t) = \sum_{i=1}^{k_n} \int_0^t H_{ij}^{(n)}(v) I\{|H_{ij}^{(n)}(v)| \geq \varepsilon\} dM_i(v), j = 1, \dots, m$$

Fiksuokime momentą τ .

Teorema. Tarkime, kad

- skaičiuojančių procesų N_i kompensatoriai A_i tolydūs intervale $[0, \tau]$;
- $H_{ij}^{(n)}(v)$ yra tolydūs iš kairės su baigtinėmis ribomis iš dešinės suderinti procesai intervale $[0, \tau]$;
- $\langle U_j^{(n)}, U_{j'}^{(n)} \rangle(t) \xrightarrow{P} \sigma_{jj'}(t)$ su visais $t \in [0, \tau]$, kai $n \rightarrow \infty$;
- $\langle U_{j\varepsilon}^{(n)}, U_{j\varepsilon}^{(n)} \rangle(t) \xrightarrow{P} 0$ su visais $t \in [0, \tau]$, kai $n \rightarrow \infty$;

e) Matrica $\Sigma(t) = (\sigma_{jj'}(t))_{m \times m}$ teigiamai apibrėžta su visais $t \in [0, \tau]$.

Tada

$$U^{(n)} = (U_1^{(n)}, \dots, U_m^{(n)})^T \xrightarrow{D} V = (V_1, \dots, V_m)^T \text{ erdvėje } (D[0, \tau])^m;$$

Kur V yra m -matis Gauso martingalas, kurio komponentės turi nepriklausomus prieauglius, $V_j(0) = 0$ b.t. ir su visais $0 \leq s \leq t \leq \tau$:

$$\text{cov}(V_j(s), V_{j'}(t)) = \sigma_{jj'}(s).$$