Survoon ohjelmoitu SURVIVAL-moduli

Kim Huuhko

SURVIVAL-moduli on ohjelmoimani elinaika-analyysin menetelmiä tarjoava laajennusohjelma SURVO-nimiseen käyttöympäristöön. SURVIVAL-ohjelma keskittyy nimenomaan elinaika-analyysin ei-parametristen menetelmien käyttöön. Tämä lisämoduli tarjoaa välineet muun muassa eloonjäämisfunktion ei-parametriseen estimointiin ja sen graafiseen tarkasteluun. Monet tilastolliset ohjelmistot tarjoavat myös välineet arvioida mahdollisten selittävien tekijöiden vaikutusta eloonjäämiseen. Nämä menetelmät käsittävät usein muutamia parametrisia malleja kuten eksponentti- ja Weibull-jakaumat sekä semiparametrisen niin sanotun Coxin verrannollisten hasardien mallin. Täysin ei-parametrisia regressiomenetelmiä ei kuitenkaan ole laajalti tarjolla varsinkaan yleisemmissä tilastollisissa ohjelmistoversioissa ja niinpä SURVIVAL-modulissa keskitytään juuri niihin. SURVIVAL-ohjelma tarjoaa käyttäjälle täysin ei-parametrisen niin sanotun eloonjäämispuumallin. Menetelmä tarkastelee hieman eri tavalla selittävien muuttujien vaikutusta eloonjäämiseen kuin perinteiset parametriset mallit. Yleensä näitä menetelmiä kannattaa käyttää rinnakkain toisiaan täydentäen, jolloin ne tarjoavat monipuolisemman ja tarkemman selvityksen eloonjäämiseen vaikuttavista tekijöistä. Toinen syy miksi SURVIVAL-moduli keskittyy nimenomaan puumallin käyttöön on puumallin yksinkertainen ja helposti tulkittava rakenne, mikä on mahdollistanut sen suhteellisen helpon teknisen toteutuksen.

3.1 Survo-käyttöympäristö

Survo on eri osista koostuva ohjelma, jolla voidaan suorittaa tilastollista analysointia, laskemista, piirtämistä ja tulosten muokkaamista. Sen kaikki toiminnot perustuvat editoriaaliseen lähestymistapaan. Survon keskeinen toimintaympäristö on toimituskenttä. Survolla työskennellään kirjoittamalla tekstiä tähän toimituskenttään sekä aktivoimalla operaatioita ja käskyjä, jotka voidaan kirjoittaa suoraan tekstin sekaan. Myös aineisto sekä sovellusten tulokset voidaan tulostaa tähän samaan toimituskenttään. Editoriaalinen käyttöliittymä mahdollistaa joustavan työskentely-ympäristön, kun sekä kirjoitus että laskenta tapahtuvat samassa toimituskentässä. Tästä syystä myös tulosten muokkaaminen on helppoa. Survon etuja on nimenomaan

se, että se ei hävitä laskuissa ja toiminnoissa käytettyjä alkuarvoja, mikä on ensiarvoisen tärkeää työn dokumentointia ajatellen. Survossa on olemassa myös oma tiedostomuotonsa datatiedostoille ja tulostaulukoille. Myös näitä Survon omia havaintotiedostoja voidaan helposti tutkia ja muokata saman toimituskentän kautta. (Mustonen 1992)

Ohjelmakokoelmana Survo on avoin käyttäjien tekemille lisämoduleille, joita voidaan käyttää heti, kun ne on ohjelmoitu ja käännetty. Jokainen Survon ohjelmamoduli on itsenäinen ohjelma, minkä vuoksi Survon laajentaminen ja omien ohjelmien luominen on yksinkertaista. Käyttäjällä on esimerkiksi mahdollisuus kirjoittaa ja kääntää C-ohjelmia, linkittää ne Survo-operaatioiksi sekä testata ja käyttää niitä samassa toimituskentässä saman istunnon aikana. Tämän lisäksi Survo-ohjelmoinnin perusapuvälineet löytyvät valmiina c-kielisinä kirjastoina. Varsinaisia rajoituksia omien ohjelmien sisällölle ei ole, mutta käyttäjän mukavuuden vuoksi olisi suotavaa, että ne toimisivat samaan tapaan kuin muutkin Survo-sovellukset. (Mustonen 1992)

Survon modulit voidaan ajaa aliohjelman tavoin, kun pääohjelma on läsnä ja välittää tarvittavat tiedot modulille. Modulin käytön jälkeen se häviää muistista ja toiminta palaa pääohjelmaan. Modulista riippuen sen palauttamat tiedot joko tulostetaan kuvaruudulle pääohjelman toimituskenttään ja/tai talletetaan Survo-tiedostoiksi. Survossa toimituskenttä on tärkeä linkki pääohjelmien ja modulien välillä. Se muun muassa välittää olennaiset tiedot aliohjelmille. Pääohjelma valitsee tarvittavat modulit käyttäjän aktivoimien komentojen mukaan. Yleensä tällä toimituskentässä aktivoidulla rivillä ovat myös aliohjelman tarvitsemat parametrit. (Mustonen 1992)

3.2 SURVIVAL-moduli

SURVIVAL-modulilla voidaan estimoida eloonjäämisfunktioita sekä tulorajaestimattorin että elinajantaulu -menetelmän avulla. Ohjelma tarjoaa myös mahdollisuuden tutkia näiden kuvaajien ominaisuuksia graafisesti sekä vertailla eri ryhmien estimaattoreita tilastollisten merkitsevyystestien avulla. Tämän lisäksi ohjelma sisältää mahdollisuuden luoda tutkittavalle aineistolle täysin ei-parametrisia puumalleja, jotka koostuvat näiden estimaattorien ja niitä vertailevien testisuureiden tehokkaasta käytöstä. SURVIVAL-

modulin eloonjäämispuumalli on ohjelmoitu pitkälti LeBlancin ja Crowleyn kehittelemän eloonjäämispuumallin pohjalta. SURVIVAL-modulin ohjelmalistaus, luokkakaaviot ja niiden selitykset löytyvät liitteistä 1-3. Ohjelmaa rakennettaessa apuna on käytetty muun muassa Seppo Mustosen ’Programming SURVO 84 in C’ –teosta, josta löytyvät Survo-ohjelmoinnin perusapuvälineet C-kielisinä kirjastoina. Lisäksi estimaattorien eri funktioiden muodostuksessa on käytetty apuna internetistä löytyviä SAS-ohjelmistossa käytettyjä elinaika-analyysin laskukaavoja (SAS online manual).

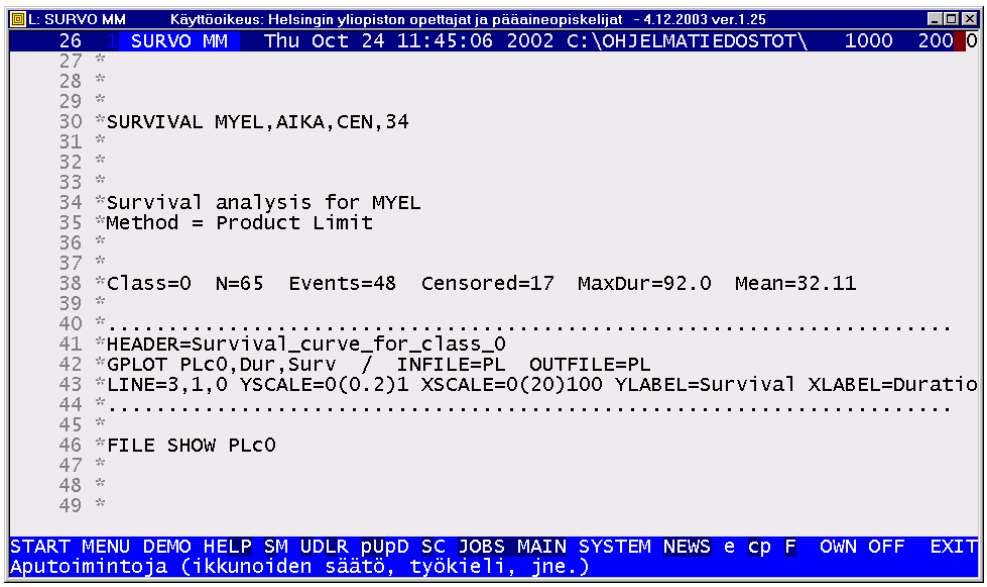
3.2.1 Eloonjäämisfunktioiden estimointi SURVIVAL-modulilla

SURVIVAL-moduli käynnistyy perustapauksessa aktivoimalla komento

SURVIVAL AINEISTO,X,Y,RIVI ,

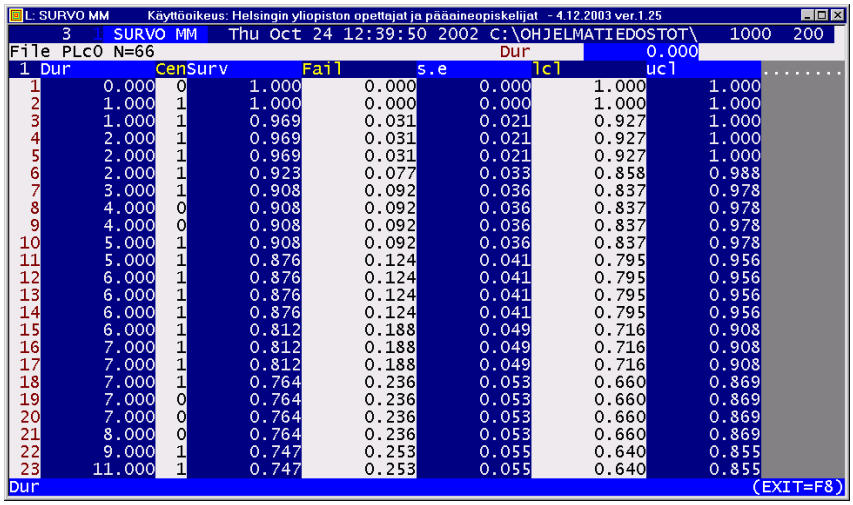
jossa AINEISTO on sen aineistotiedoston nimi, jota halutaan tutkia. X on tässä tapauksessa sen muuttujan nimi, joka kertoo kunkin havainnon seurannassa viettämän ajan. Y on puolestaan muuttuja, joka kertoo päättyikö tämä aika tapahtumaan vai sensurointiin, eli onko kyseisen havainnon ensimmäisessä muuttujassa määritelty aika tapahtuma- vai sensurointiaika. Näiden muuttujien ei tarvitse olla millään tavalla etukäteen järjestettyjä, vaan moduli hoitaa itse niiden järjestämisen tarvitsemaansa järjestykseen. Aikamuuttujan täytyy luonnollisesti olla ei-negatiivinen suhdeluku-asteikollinen muuttuja. Sensurointimuuttujaa ohjelma käsittelee dikotomisena muuttujana siten, että 1 merkitsee aina tapahtumaa ja kaikki muut arvot tulkitaan sensuroiduiksi havainnoiksi. RIVI on sen rivin numero Survo:n toimituskentässä, josta eteenpäin tulokset halutaan näkyviin.

Perustapauksessa SURVIVAL-moduli laskee sille määrätystä aineistosta ei-parametrisen estimaattorin eloonjäämisfunktiolle. Tämä estimaattori on käyttäjän komennoista riippuen joko tulorajaestimaattori tai elinajantaulu. Oletusarvona on tulorajaestimaattori. Alla on näyte Survon toimituskentästä tulorajaestimaatorin luomisen jälkeen.



Rivillä 30 on komento, jonka aktivoimalla tulokset saatiin näkyviin annetulle riville 34. Ohjelma kertoo, mikä oli kyseessä olevan aineiston nimi ja käytetty menetelmä sekä tulostaa ruudulle keskeisimmät aineiston tunnusluvut, kuten havaintojen kokonaismäärän (65), tapahtumien (48) ja sensuroitujen havaintojen (17) määrät sekä pisimmän seuranta-ajan (92) ja tapahtuma-aikojen keskiarvon (32.11).

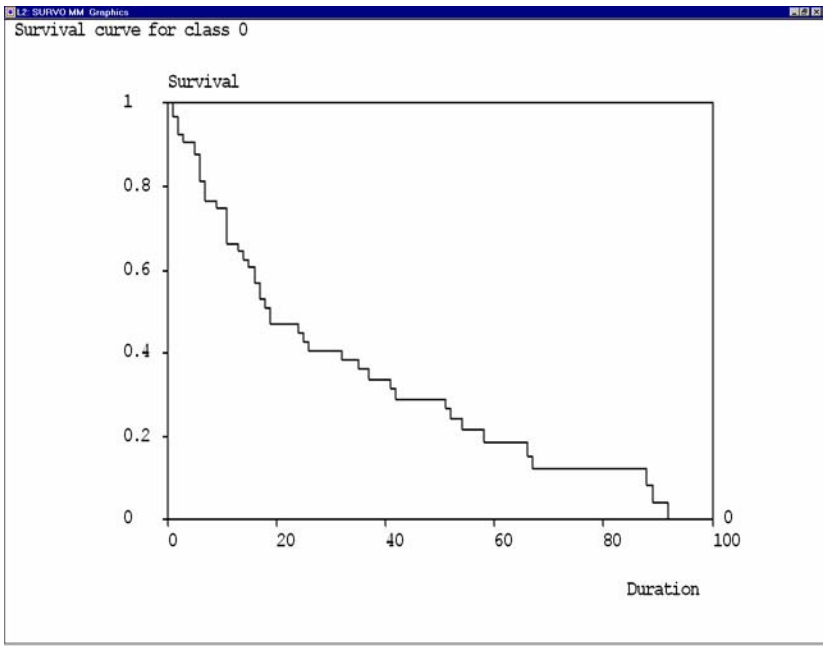
Tulorajaestimaattorin tapauksessa ohjelma tallettaa saadut tulokset uuteen tiedostoon, jonka nimi on PLc0.SVO. Riville 46 ohjelma tulosti FILE SHOW -komentorivin, jonka aktivoimalla käyttäjä pääsee tutkimaan tämän tiedoston sisältöä.



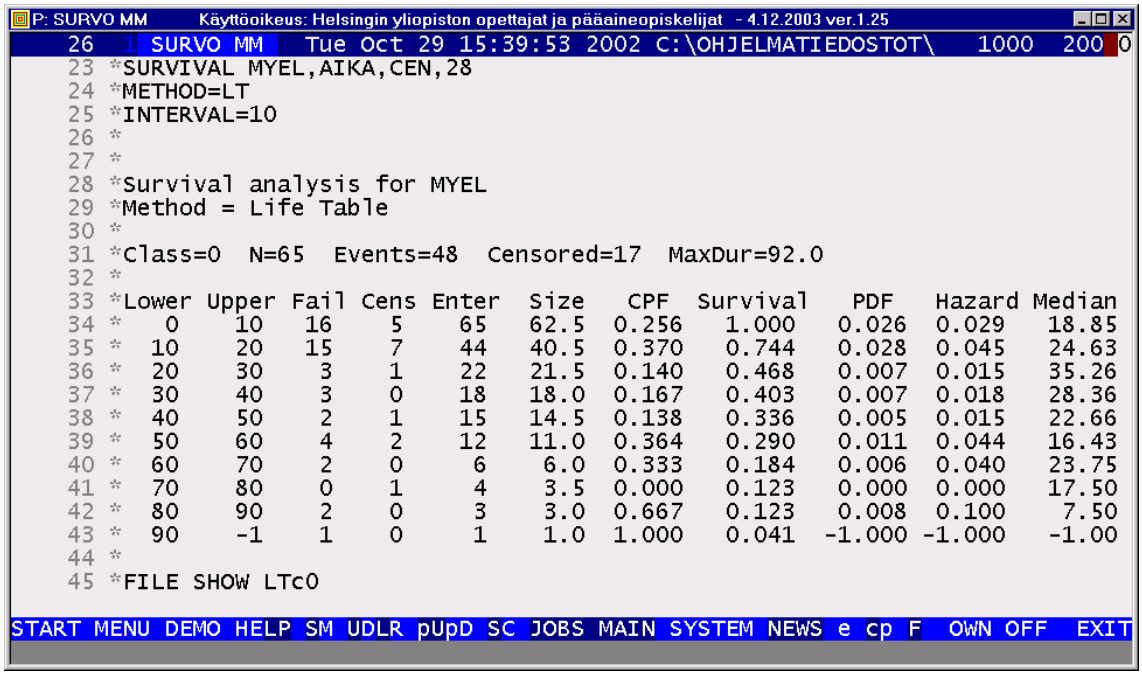
Kyseisen tiedoston ensimmäinen rivi kuvaa aina alkutilannetta, jolloin aika on 0 ja estimaattorin arvo on 1, eli yhdellekään havainnolle ei ole vielä tapahtunut tapahtumaa. Tiedoston jokainen muu rivi kuvaa yhtä alkuperäisen aineiston havaintoa eli tiedostossa on aina n+1 riviä. Tämän tiedoston ensimmäiset sarakkeet kuvaavat kunkin havainnon tapahtuma-ajan (Dur) ja sen, päättyikö tapahtuma-aika tapahtumaan vai sensurointiin (Cen). Tiedoston tärkein muuttuja on sarake Surv, joka kertoo tulorajaestimaattorin arvon kullakin erillisellä ajanhetkellä. Tulorajaestimaattorin arvon aina tietyn ajanhetken jälkeen näkee tämän ajanhetken viimeisen havainnon riviltä. Esimerkiksi kyseisellä aineistolla estimaattorin arvo 6 (viikon) jälkeen on 0.812 ja 7 (viikon) jälkeen 0.764. Seuraavassa sarakkeessa Fail on kuvattu tulorajaestimaattorin komplementti, eli todennäköisyys, että yksilölle on tapahtunut tutkittava tapahtuma tämän ajanhetken loppuun mennessä. Viimeiset kuvassa näkyvät sarakkeet ovat tulorajaestimaattorin keskihajonta (s.e) sekä 95% luottamusvälin ala- (lcl) ja ylärajat (ucl).

SURVIVAL-moduli tulosti toimituskentän riveille 41-43 myös GPLOT-kaavion, jonka aktivoimalla (rivi 42) käyttäjä saa näkyviin eloonjäämisfunktion kuvaajan erilliseen grafiikkaikkunaan. Kuviossa 5 on kuvattu tämä koko aineistolle piirretty tulorajaestimaattori. Tarpeen vaatiessa käyttäjä voi piirtää samaan kuvaan myös tämän estimaattorin luottamusvälit, jotka antavat arvion sen stabiilisuudesta.

Kuvio 5. Tulorajaestimaattori esimerkkiaineistolle MYEL



Jos käyttäjä lisää alkuperäisen SURVIVAL-komentorivin perään määritteen METHOD=LT, ohjelma muodostaa aineistosta elinajantaulun. Tässä tapauksessa ohjelma tallettaa saadut tulokset uuteen tiedostoon, jonka nimi on LIFEc0.SVO. Ohjelma tulostaa tärkeimmät näistä tulossarakkeista myös suoraan näytölle. Elinajantaulun tapauksessa kannattaa käyttää myös lisämääritettä INTERVAL, jonka avulla käyttäjä voi jakaa aineiston elinajat halutun kokoisiin aikaväleihin. Tämä on suotavaa varsinkin niissä tapauksissa, joissa havaintoja on paljon. INTERVAL-määritteellä käyttäjä pystyy näin säätelemään tulostaulukon kokoa ja pitämään sen helposti hallittavan kokoisena. Alla on näkymä Survon toimituskentästä, johon ohjelma on tulostanut elinajantaulun rivistä 28 eteenpäin.



Taulukosta nähdään muun muassa aineiston tapahtumien (48) ja sensuroitujen havaintojen (17) määrät sekä aineiston pisin seuranta-aika (92.0). Taulukon ensimmäiset sarakkeet kuvaavat kunkin intervallin ala- ja ylärajaa. Seuraavaksi on esitetty kussakin intervallissa tapahtuneet tapahtumat (Fail) sekä sensuroidut (Cens) havainnot. Seuraavat sarakkeet kuvaavat kunkin intervalliin alkuun asti selvinneiden havaintojen määrän (Enter) sekä potentiaalisen riskiryhmän uusintaleikkaukselle tässä intervallissa (Size). CPF-sarakkeessa on tapahtuman todennäköisyys kyseisessä intervallissa ehdolla, että sitä ei ole tapahtunut jo aiemmin. PDF on tapahtumien tiheysfunktion ja Hazard-hasardifunktion kuvaajat. Näiden sarakkeiden luvut ovat

käytetyn aikamittarin tiheydestä riippuen usein hyvin pieniä lukuja, jolloin ne voivat pyöristyksestä johtuen näkyä ruudulla epätarkasti. Näiden muuttujien arvoja voidaan kuitenkin tutkia myös piirtämällä ne erilliseen kuvaikkunaan tai kertomalla ne uusiin survo-tiedoston muuttujiin. Luku -1.000 viittaa näissä taulukoissa siihen, että kyseistä arvoa ei voida laskea. Aineiston viimeisen intervallin yläraja on aina määrittämätön ja tämän vuoksi ei myöskään tiheysfunktion arvoa, hasardia tai mediaania voida laskea viimeiselle intervallille. Mediaani on kuhunkin intervalliin selvinneiden havaintojen vielä odotettavissa oleva "elinikä". Mediaanielinikää ei luonnollisesti voida määrittää intervalleille, joiden alkuun asti selvinneistä havainnoista yli puolet ovat sensuroituja. Survival on kyseisen taulun sarakkeista olennaisin. Siihen on kuvattu aineistoa vastaavan eloonjäämiskäyrän arvo kunkin intervallin alussa. Tämä eloonjäämisfunktio voidaan piirtää tulorajaestimaattorin tavoin erilliseen kuvaikkunaan ohjelman näytölle tulostamien GPLOT-kaavioiden avulla. GPLOT-kaavioita tulostuu elinajantaulun tapauksessa kolme, joista ensimmäinen on eloonjäämisfunktion kuvaaja, toinen on tämän eloonjäämisfunktion 95%:n luottamusvälit ja kolmas on hasardifunktion kuvaaja.

Eloonjäämisfunktion kuvaajasta tutkija voi silmämääräisesti tarkastella kyseisen aineiston eloonjäämistodennäköisyyksien erityispiirteitä ajan suhteen. Hasardifunktio taas kertoo aineiston yksilöiden hetkellisen tapahtumariskin ajan suhteen, yhtä aikayksikköä kohti. Tulostaulukoista tutkija voi puolestaan kerätä eloonjäämistä kuvaavat mielenkiintoiset tunnusluvut kuten eloonjäämisfunktion mediaanit, kvartiilit ja keskiarvot.

Alkuperäisen komentorivin perään käyttäjä voi lisätä myös CLASS-määritteen. Tällä määritteellä voidaan ohjelmalle antaa luokittelumuuttujan nimi, jonka mukaan aineisto halutaan jakaa osiin. Ohjelma tekee tällöin oman erillisen tulorajaestimaattorin tai elinajantaulun kaikille tämän luokittelumuuttujan eri luokille. Uusien tiedostojen nimet määräytyvät siten, että esimerkiksi uusi tulorajaestimaattoritiedosto kyseisen luokittelu-muuttujan arvolle 4 on nimeltään PRODc4.SVO ja niin edelleen. Ohjelma tulostaa luonnollisesti kullekin näistä luokista myös omat GPLOT-kaavionsa. Piirtämällä nämä eri luokkien eloonjäämisfunktiot samaan grafiikkaikkunaan, voi käyttäjä silmämääräisesti vertailla näiden eri osajoukkojen eloonjäämistä ajan suhteen. Tulostaulukoiden ja kaavioiden perään ohjelma tulostaa myös testisuureita, joilla käyttäjä voi vahvistaa silmämääräisten arvioiden johtopäätöksiä. Testien

nollahypoteesina on, että kaikkien näiden eri luokkien eloonjäämisfunktiot ovat samankaltaisia. Jos testisuureita vastaavat p-arvot jäävät pieniksi (alle 0.05), niin testi kertoo tilastollisesti merkittävästä erosta kyseisten ryhmien eloonjäämisen välillä. Testisuureita on kolme: log-rank, Wilcoxon ja Tarone–Ware. Nämä testit mittaavat eloonjäämisen eroja painottamalla hieman eri kohtia eloonjäämisfunktioissa. Niinpä osoitukseksi erilaisista eloonjäämishistorioista riittää, että yksikin testi osoittaa tilastollisesti merkittävää eroa.

3.2.2 Eloonjäämispuiden muodostaminen SURVIVAL-modulilla

Näiden komentojen avulla käyttäjä voi estimoida eloonjäämisfunktioita ja niiden eroja eri ryhmissä. Usein elinaika-analyysin mielenkiinnon kohteena on kuitenkin myös muiden, selittävien muuttujien, vaikutus näihin eloonjäämisfunktioihin. SURVIVAL-modulin avulla tämä onnistuu merkitsemällä mahdolliset selittävät muuttujat MASK-komennolla arvolla X ja antamalla CLASS määritteeseen komento TREE. MASK-määritteen avulla valituille mahdollisille selittäville muuttujille on olemassa tiettyjä rajoituksia. Ohjelma kykenee tutkimaan vain suhdeluku-, intervalli- tai ordinaaliasteikollisia sekä binäärisiä muuttujia. Se ei siis osaa tutkia useampiluokkaisten luokittelumuuttujien kaikkia mahdollisia kombinaatioita. Käyttäjä voi toki muodostaa tällaisista muuttujista binäärisiä muuttujia ja sisällyttää ne näin analyysiinsä.

Jos SURVIVAL-modulille on annettu määrite CLASS=TREE, niin käyttäjän aktivoidessa alkuperäisen komentorivin, ohjelma ei muodostakaan aineistosta tulorajaestimaattoria tai elinajantaulua vaan niin sanotun puumallin. Käytännössä tämä tarkoittaa sitä, että ohjelma käy läpi kaikkien annettujen selittävien muuttujien kaikki mahdolliset katkaisupisteet ja muodostaa kullekin tällaiselle binääriselle jaolle näiden ryhmien eloonjäämisfunktioiden eroja mittaavan testisuureen. Ohjelma valitsee näistä jaoista sen, joka parhaiten erottelee aineiston kahteen erilliseen osajoukkoon ja jakaa aineiston sen mukaan. Tämän jälkeen ohjelma suorittaa samat laskutoimitukset ja jaot kummallekin näin muodostuneelle osajoukolle erikseen ja niin edelleen. Näin siis muodostuu ikään kuin binäärinen valintapuu, joka jakaa aineiston osajoukkoihin, joiden havainnoilla on keskenään samankaltainen eloonjäämishistoria. Menetelmä siis kertoo, mitkä tekijät kulloinkin vaikuttavat eniten eloonjäämiseen ja miten.

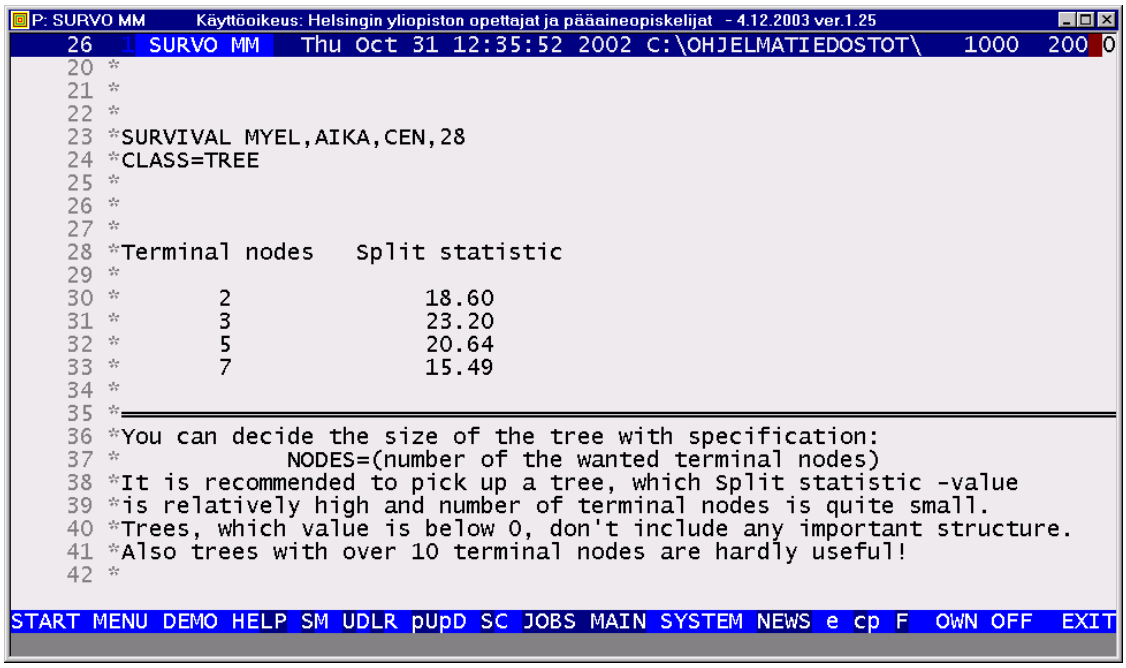
Puun jakamista jatketaan niin kauan kun se on mahdollista. Tämän jälkeen moduli laskee kullekin syntyneen puun solmulle sen merkitystä vastaavan arvon. Tämä arvo muodostetaan kyseisen solmun parhaalle jaolle saadusta testisuureen chi^2-arvosta. Kunkin yksittäisen solmun arvo on tällöin sen oman ja kaikkien sen muodostaman puunhaaran jaettujen solmujen arvojen summa. Tuloksena on puun juuresta lähtien solmuja pitkin alaspäin mentäessä monotonisesti laskeva arvomittari. Tämän jälkeen kunkin solmun arvosta vähennetään sen muodostaman puunhaaran monimutkaisuutta kuvaava arvo, joka on haittaparametri α * (tämän puunhaaran jaettujen solmujen määrä). Haittaparametrin α oletusarvo on 4, mutta käyttäjä voi itse asettaa sen arvoksi mitä tahansa 2 ja 4 väliltä määritteellä PENALTY=(tämä haluttu arvo)[1]. Näiden kullekin solmulle määriteltyjen suhteellisten arvojen joukosta etsitään pienin arvo ja katkaistaan puu niin, että tästä solmusta muodostuu uuden karsitun puun päätesolmu. Tämän jälkeen jokaiselle tämän karsitun puun solmulle lasketaan uudet suhteelliset arvot ja karsitaan tämä puu jälleen pienimmän suhteellisen arvon kohdalta. Tätä niin sanottua "heikoimman lenkin karsimista" toistetaan niin kauan kunnes jäljellä on enää puun juurisolmu. Näin saadaan muodostettua sarja sisäkkäisiä, optimaalisesti karsittuja osapuita, joiden joukosta lopullinen puunvalinta on mahdollista suorittaa.

Näiden optimaalisesti karsittujen osapuiden arvo on kaikkien niiden sisältämien solmujen suhteellisten arvojen summa. Tämä arvo kuvaa siis tavallaan puun sisältämää tilastollisesti merkittävää rakennetta suhteessa sen monimutkaisuuteen. Jos puun monimutkaisuus on suurempi kuin puun sisältämä merkittävä rakenne, niin puun arvo jää alle nollan. SURVIVAL-moduli tulostaa kuvaruudulle näiden karsittujen osapuiden "Split statistic" -arvot sekä näiden puiden päätesolmujen määrän. Lopullinen puunvalinta tulisi olla jonkinlainen kompromissi näistä luvuista. Periaatteessa suurempi "Split statistic" -arvo viittaa aina tilastollisesti informatiivisempaan rakenteeseen, mutta käytännössä puun koon pitäminen suhteellisen pienenä on kuitenkin näistä tekijöistä tärkeämpi. Syntyneistä puista voidaan aina poimia mielenkiintoiset osajoukot erilleen ja tehdä näille osajoukoille omat puunsa, jolloin myöskään koko aineiston suurempi monimutkaisuus ei pääse häiritsemään näiden osajoukkojen sisäisten rakenteiden havaitsemista. Yli 10 päätesolmua sisältävät puut ovat usein kyseenalaisia ja harvoin tarpeellisia.

[1]) Haittaparametrin arvo 4 vastaa suurin piirtein testisuureiden p-arvoa 0.05, eli silloin puuta ei käytännössä voida katkaista kohdista, jossa paras löydetty jako oli tätä huonompi. Parametrin arvo 2 on huomattavasti sallivampi ja mahdollistaa myös heikompien jakojen muodostumisen ja näin ollen suuremman puun syntymisen.

Syntyneiden puiden kokoa voidaan säädellä myös OBSLIMIT -määritteen avulla. Tämän määritteen arvo asettaa puiden sallituille päätesolmuille niiden sisältämän havaintomäärän alarajan. Mitä pienempi tämä alaraja on sitä enemmän aineistossa on mahdollisia katkaisupisteitä ja sitä suurempi alkuperäisestä puusta tulee. Jos tämä alaraja asetetaan ykköseksi, niin puu jakaa aineistoa niin kauan kuin se on mahdollista eli tuloksena saattaa olla jopa puu, jossa jokainen havainto muodostaa oman päätesolmunsa. Käytännössä tämä ei kuitenkaan vähänkään suuremmilla aineistoilla ole järkevää, koska silloin puun sisäinen rakenne häviää tämän monimutkaisuuden varjoon. Pienillä aineistolla tätäkin alarajaa voi käyttää, jos haluaa etsiä aineistosta yksittäisiä poikkeavia havaintoja. Oletusarvona sallitun havaintomäärän alarajana on n/10+1, eli noin kymmenesosa koko aineiston havaintomäärästä. Oletusarvo on melko korkea, koska usein aineistosta on tarkoitus erotella vain noin 2-8 erillistä osajoukkoa, joissa kaikissa on merkittävä määrä havaintoja. Syntyneille osajoukoille voi aina tarvittaessa rakentaa omat puunsa ja näin muodostaa suurempi puu, haara kerrallaan. Usein on kuitenkin hyvä kokeilla puunrakennusta myös alhaisemmilla OBSLIMIT:in arvoilla, jotta mahdollisten pienempien poikkeavien osajoukkojen vaikutukset eivät jäisi havaitsematta.

Seuraavassa kuvassa on MYEL-aineistolle muodostetun puumallin optimaalisesti karsittujen osapuiden päätesolmujen määrät ja Split Statistic -arvot. Aineistolle voitaisiin hyvin rakentaa 5 päätesolmunkin puu, mutta koska aineisto on melko pieni (48 tapahtumaa), niin tässä tapauksessa on parempi valita "rakenteeltaan" paras, vain 3 päätesolmua sisältävä puu.

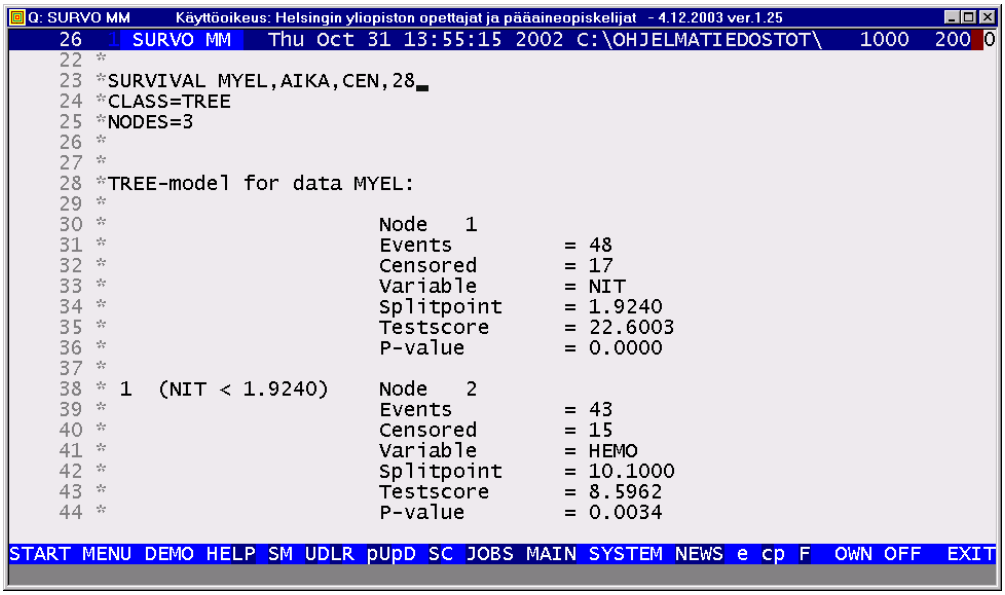


Tämä päätesolmujen määrä kertoo siis, kuinka moneen eri osajoukkoon alkuperäinen aineisto on kussakin tapauksessa jaettu. Valittuaan haluamansa päätesolmujen määrän käyttäjä voi lisätä, NODES -määritteen yhteydessä, tämän luvun alkuperäisen komentorivin perään. Kun alkuperäinen komentorivi aktivoidaan, ohjelma rakentaa puun uudestaan, mutta tekee siitä tällä kertaa vain halutun kokoisen. Tämän jälkeen ohjelma tulostaa näytölle kyseistä puuta koskevan informaation. Käyttäjä voisi toki määritellä ohjelmalle halutun puun koon jo ensimmäisellä kerralla, mutta tällöin puun koon valinnalla ei olisi mitään tilastollista tukea.

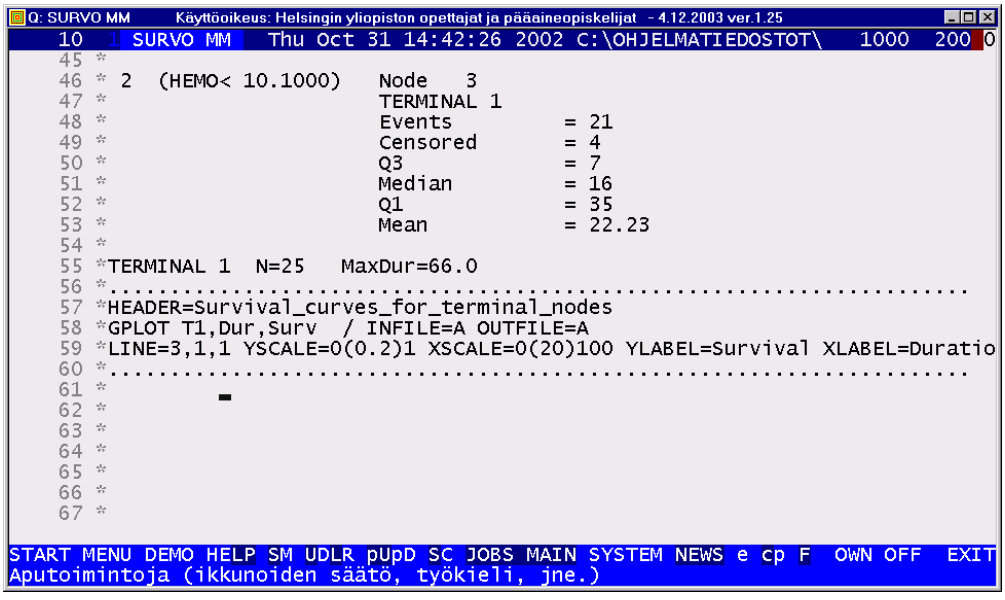
Puumallin hahmottaminen syntyneestä tulostuksesta onnistuu solmujen numeroiden avulla. Kunkin solmun (juurisolmua lukuun ottamatta) vasemmalla puolella on ensin kerrottu, mikä oli tämän solmun isäntäsolmun numero. Tämän jälkeen on tulostettu minkä muuttujan minkä katkaisupisteen mukaan kyseinen isäntäsolmu oli jaettu ja kummasta syntyneestä osajoukosta on nyt kyse. Puumalli on ainakin ensimmäisillä käyttökerroilla selvyyden vuoksi syytä hahmotella myös paperille. Tällöin kannattaa merkitä vasemmaksi haaraksi aina ne havainnot, joiden arvo oli pienempi kuin kyseinen katkaisupiste.

Kustakin jaetusta solmusta on ensin kerrottu kyseisen solmun numero sekä kuinka monta tapahtumaa ja kuinka monta sensuroitua havaintoa tähän osajoukkoon kuuluu.

Tämän lisäksi on kerrottu, mikä oli tämän solmun havaintoja parhaiten erotteleva muuttuja ja mikä oli tämän muuttujan kyseinen paras katkaisupiste. Sen lisäksi solmussa on kerrottu parhaalle jaolle saadun testisuureen arvo sekä tätä testisuuretta vastaava p-arvo.



Kullekin päätesolmulle on puolestaan kerrottu solmun numeron lisäksi myös se, kuinka mones päätesolmu on kyseessä. Näiden päätesolmujen numeroiden avulla voidaan muun muassa hahmottaa kullekin osajoukolle muodostetun eloonjäämisfunktion kuvaaja kuvaikkunasta, johon on piirretty useiden eri osajoukkojen kuvaajat. Päätesolmuista on kerrottu tapahtumien ja sensuroitujen havaintojen määrän lisäksi muutamia olennaisia tunnuslukuja. Näitä tunnuslukuja ovat mm. kyseiselle osajoukolle muodostetun eloonjäämisfunktion mediaani, kvartiilit ja keskiarvo.



Kullekin päätesolmulle ohjelma muodostaa uuden SVO-tiedoston sekä GPLOT-kaavion. Tiedoston nimeksi tulee esimerkiksi ensimmäiselle päätesolmulle T1.SVO ja niin edelleen. FILE SHOW -komennolla voidaan näistä tiedostoista katsoa kyseisen osajoukon tulorajaestimaattorin arvot kunakin ajanhetkenä. Aktivoimalla GPLOT-kaavion (rivi 66) käyttäjä voi arvioida tämän estimaattorin piirteitä myös silmämääräisesti. Piirtämällä eri osajoukkojen eloonjäämisfunktiot samaan ikkunaan tutkija voi vertailla näiden ryhmien eloonjäämishistorioiden eroja. Yhdessä jakavien solmujen sisältämän informaation kanssa nämä eloonjäämisfunktiot tarjoavat kätevän ja helposti tulkittavan välineen arvioida selittävien muuttujien vaikutusta itse eloonjäämiseen.

Menetelmä saattaa jakaa aineiston useita kertoja jonkun saman jatkuvan muuttujan eri kohdista, jolloin saattavat paljastua myös tämän muuttujan funktionaalisesti monimutkaisemmatkin vaikutukset selviytymiseen. Merkittävin etu kyseisenlaisessa puumallitarkastelussa on kuitenkin se, että se kykenee löytämään vuorovaikutussuhteet, joilla on merkitystä ainoastaan osajoukoissa. Jos OBSLIMIT on asetettu tarpeeksi matalaksi, menetelmä jakaa poikkeavat havainnot omiksi erillisiksi pieniksi päätesolmuikseen, jolloin niiden havaitseminen, tutkiminen ja mahdollisesti poistaminen on helppoa. Käyttäjä voi esimerkiksi rajoittaa aineistosta luettavien havaintojen määrää normaalin SURVO-modulin tavoin käyttämällä IND- ja CASES-määritteitä.

Käyttäjä voi myös itse valita, mitä testisuuretta käyttäen puu muodostetaan. Valinta suoritetaan lisäämällä määrite TEST ja asettamalla sen arvoksi joko 1 (log-rank), 2 (Wilcoxon) tai 3 (Tarone–Ware). Oletusarvona käytetään log-rank -testisuuretta. Käytännössä testisuureen valinnalla ei useinkaan ole merkitystä syntyneen puun topologiaan, mutta joissain tapauksissa on syytä muodostaa puumalli käyttäen kutakin eri testisuuretta ja valita niistä se, joka on parhaiten tulkittavissa.

Puumallien rakentaminen on ennen kaikkea kokeellinen menetelmä eikä se tarjoa nopeaa ja tyhjentävää ratkaisua yhtä puuta rakentamalla. Puiden koko kannattaa pitää usein pienenä ja rakentaa erillisiä puita aineiston mielenkiintoisimmille osajoukoille.

Myöskin puiden rakennus eri haittaparametrin arvolla ja pienemmällä solmujen havaintomäärän alarajalla saattavat paljastaa aineistosta piirteitä, jotka muuten jäisivät havaitsematta. Erittäin tärkeää on myös puiden rakentaminen mukana olevien selittävien muuttujien eri kombinaatioilla. Usein kaikkein tärkeimmän selittävän muuttujan jättäminen pois puuta rakennettaessa paljastaa myös muiden muuttujien merkityksen aineistossa. Joskus on hyödyllistä myös tutkia muuttujia erikseen ja etsiä niistä parhaita katkaisupisteitä ja näin syntyneiden osajoukkojen erityispiirteitä. Puumallit tarjoavat nimenomaan tehokkaan välineen aineiston pilkkomiseen ja tarkasteluun useammasta eri näkökulmasta pikemminkin kuin yksiselitteisen ja valmiin ratkaisun.

Yksittäiset eloonjäämisfunktion estimaattorit muodostuvat suhteellisen yksinkertaisten laskujen tuloksena, mutta koska puumalleja rakennettaessa näitä laskutoimituksia saatetaan suurilla aineistoilla tehdä jopa miljoonia, on puumallin muodostaminen laskennallisesti melko raskas menetelmä. Puumallin rakentamiseen saattaa kulua runsaastikin aikaa, jos havaintoja on paljon ja useat selittävistä muuttujista ovat jatkuvia.