



Pääsiäinen on aina sunnuntaina

Kimmo Vehkalahti

Sosiaalitieteiden laitos, tilastotiede
Helsingin yliopisto

Innostuin (otsikkoa myöten!) *Matti Lehtisen* jutusta ”Jouluaatto on harvemmin sunnuntaina” (Solmu 3/2011). Juhlapäivien sijoittuminen kalenteriin on kiehtova aihe, ja sitä voi lähestyä matemaattisesti monelta suunnalta. Keskitän huomioni pääasiassa pääsiäiseen, joka on laskennallisesti erityisen mielenkiintoinen.

Tarkastelen lisäksi eräitä muita juhlapäiviä, joista osa on pääsiäisen tapaan kirkollisia (kuten pyhäinpäivä) ja osa maallisia (kuten vappu). Yhteiskunnallisesti kiinnostavia juhlapäivistä tekevät niiden aiheuttamat poikkeukset arkirutiineihin, esimerkiksi kauppojen aukioloaikoihin. Juhlapäivät ja etenkin niitä edeltävät aattopäivät asettavat myös yllättäviä haasteita muun muassa myyntiennusteiden laatimiseen. Tarkastelujeni taustalla onkin pari vuotta sitten eräälle yritykselle räätälöity myynnin analysointijärjestelmä, jota ohjelmoidessa sai perehtyä juhlapäiviin perusteellisesti.

Uskonnollisia juhlia kuten uskontojakin on maailmassa lukuisia, mutta rajaan tarkastelut Suomessa vietettäviin, kristillisen kirkkovuoden juhliin. Yhdessä maallisten juhlien kanssa ne muodostavat sarjan, joka toistuu suomalaisessa kalenterissa vuodesta toiseen. Sarja on silti aina hieman erilainen riippuen siitä, mihin kohtiin tietyt juhlakalenterit sijoittuvat. Erikoisen mielenkiintoisia tilanteita voi syntyä kirkollisten ja maallisten juhlien ”törmätessä”. Laskennallisilla keinoilla tällaisia ilmiöitä voi mukavasti tutkia minkä tahansa vuoden osalta. Soveltavana tilastotieteilijänä lähestyn aihetta tilastollisen tietojenkäsittelyn näkökulmasta.

Muste leviää ja jättää jälkiä

Kirjoitan tätä Musteella, joka on uusi, R-ohjelmistolle kehitetty, vapaasti saataville tuleva avoimen lähdekoodin toteutus *Survo*-ohjelmistosta. *Survo* on professori *Seppo Mustosen* elämäntyö, joka alkoi jo 1960-luvulla ja jatkuu edelleen. Musteen päävastaullinen kehittäjä on *Reijo Sund*, joka 2009 esitti idean Musteesta sekä sen toteuttamisesta R:n laajennuspaketina.

R on *Ross Ihakan* ja *Robert Gentlemanin* alullepanema toteutus professori *John M. Chambersin* johdolla kehitetystä tilastollisesta S-ohjelmointikielestä. Nykyisin R on laajan kansainvälisen käyttäjäyhteisön kehittämä, vapaasti saatava ohjelmisto, joka on noussut 2000-luvun aikana suosioon tilastotieteen, matematiikan ja tietojenkäsittelyn piirissä, mutta myös yhä useammalla näitä menetelmätieteitä hyödyntävillä tieteenaloilla.

Survo ja R ovat hioutuneet vuosikymmenien aikana ja ne sisältävät monia kekseliäitä tapoja data-analyysiin, matriisilaskentaan, kuvien piirtoon ja raportointiin. Tärkein *Survon* innovaatioista on *editoriaalinen käyttö-tapa*, jolla työskentelystä jää talteen hyödyllisiä jälkiä. *Muste* liittyy tämän uniikin käyttötavan osaksi R:ää.

Aloittaessani tätä juttua (5.2.2012) Musteen versio-numero oli 0.4.80. Tämänhetkisen versionumeron voi tarkistaa sivulta www.survo.fi/muste, josta Musteen voi myös ladata käyttöönsä. *Muste* leviää kaikkialle (Windows, Mac, Linux).

Juhlallista laskentaa

Pääsiäinen on joulun ohella kirkkovuoden suurimpia juhlia. Siihen kuuluu peräti kolme pyhäpäivää: *pitkäperjantai*, joka nimensä mukaisesti on perjantai, *pääsiäispäivä*, joka on aina sunnuntai sekä *toinen pääsiäispäivä*, joka on maanantai. Pääsiäinen hallitsee kirkollista juhlavuotta, sillä se määrittää useiden muiden juhlapäivien, kuten helatorstain, ajankohdan. Laskennallisesti kiinnostavaksi pääsiäisen tekee sen ajankohta, joka riippuu muun muassa kuun kierrosta. Tuo ajankohta vaihtelee vuosittain yli kuukauden pituisen ajanjakson puitteissa: aikaisin mahdollinen pääsiäissunnuntai on jo 22. maaliskuuta ja myöhäisin vasta 25. huhtikuuta.

Pääsiäissunnuntain päivämäärä on osattu selvittää laskennallisesti jo keskiajalla. Tätä aikoinaan erittäin tärkeällä sijalla ollutta toimintaa kutsuttiin latinankielisellä sanalla *computus*, joka on sittemmin alkanut yleisemminkin tarkoittaa laskentaa. Englanniksi sanan johdoksia ovat muun muassa *computation* ja *computer*. Ei niistä pääsiäinen ihan ensimmäiseksi tule mieleen!

Ensimmäisen matemaattisen algoritmin pääsiäissunnuntain päivämäärän laskemiseen kehitti lukemattomista muistakin yhteyksistä kuuluisa matemaatikko *Carl Friedrich Gauss* 1800-luvun alussa. Tuolloin oli jo monissa maissa siirrytty vanhasta, epätarkasta juliaaniseen kalenterista tarkempaan, nykyisinkin käytössä olevaan, gregoriaaniseen kalenteriin. Suomessa siirtyminen tapahtui Ruotsi–Suomen aikana vuonna 1753, eräissä muissa Euroopan maissa vasta 1900-luvulla.

Gaussin algoritmi ja muut vastaavat menetelmät perustuvat melko monimutkaisiin, taivaankappaleisiin liittyviin tekijöihin, ja niissä on otettava huomioon koko joukko kalenteria koskevia sääntöjä ja poikkeuksia. Matemaattisesti algoritmit ovat yksinkertaisia, sillä ne perustuvat pelkkiin peruslaskutoimituksiin, ennen kaikkea jakolaskuihin ja jakojäännöksiin.

Yksi suoraviivaisimmista pääsiäisalgoritmeista julkaistiin *Nature*-lehdessä vuonna 1876. Sen laatija jäi hämärän peittoon, kun tekijäksi mainittiin vain lehden ”New Yorkin kirjeenvaihtaja”, mutta algoritmista tuli pian suosittu, kun sitä kopioitiin useisiin aikakauden kalenterijulkaisuihin.

Sovelsin kyseistä algoritmia alussa mainitsemassani analysointijärjestelmässä, jonka toteutin *Survolla*. Sivutuotteena tulín tehneeksi pienen pääsiäisaiheisen animaation, johon voi tutustua verkossa osoitteessa www.survo.fi/demos/index.html#ex19.

Seuraavassa sama algoritmi näkyy *Musteen toimituskenttään* kirjoitettuna *laskentakaaviona*, jossa funktio $\text{mod}(x, y)$ tarkoittaa jakojäännöstä, kun x jaetaan y :llä, ja $\text{int}(x)$ annetun luvun tai lausekkeen x kokonaisuosa. Symbolit a, b, \dots, n ovat algoritmin mukaisia lausekkeita, joiden yksityiskohtia ei ryhdytä avaamaan.

Selvitetään pääsiäissunnuntain ajankohta. Algoritmi: Anonymous (1876). To find easter. *Nature*, 13, 487.

Luku $n=(h+1-7*m+114)$ lasketaan 12 muun luvun avulla:

$$\begin{aligned} a &= \text{mod}(\text{vuosi}, 19) & b &= \text{int}(\text{vuosi}/100) & d &= \text{int}(b/4) & i &= \text{int}(c/4) \\ f &= \text{int}((b+8)/25) & c &= \text{mod}(\text{vuosi}, 100) & e &= \text{mod}(b, 4) & k &= \text{mod}(c, 4) \\ g &= \text{int}((b-f+1)/3) & & & h &= \text{mod}((19*a+b-d-g+15), 30) \\ m &= \text{int}((a+11*h+22*1)/451) & & & l &= \text{mod}((32+2*e+2*i-h-k), 7) \end{aligned}$$

Sanalliset ilmaisut lausekkeiden lomassa eivät häiritse laskentaa. Kaavion voi kirjoittaa vapaasti; tässä se on sovitettu palstakoon sallimaan tilaan.

Kaavion jatko-osassa on määritelty n :stä (ja sen myötä kaikista muista) riippuvat symbolit *kuukausi* ja *päivä*. Tarkastelun kohteena oleva vuosi on myös ohimennen ilmaistu osana virkettä:

Kaavion avulla selviää minkä tahansa pääsiäissunnuntain $\text{kuukausi}=\text{int}(n/31)$ ja $\text{päivä}=\text{mod}(n,31)+1$ - esimerkiksi, jos $\text{vuosi}=2012$, niin $\text{kuukausi}=\text{ ja päivä}=\text{$

Mitään laskentaa ei kuitenkaan toistaiseksi tapahdu: kaavio on vain kokoelma tekstiä, lukuja ja numeroita.

Mitä taitoja koulussa aktivoidaan?

Elämä sujuu, kun osaa **lukea**, **laskea** ja **kirjoittaa**. Myös **Muste** ja **Surv** osaavat nämä arvokkaat taidot, aivan omalla tavallaan.

Aktivoimalla lauseke (joko hiiren kaksoisklikkauksella tai **esc**-napilla merkin = vierestä) edellä oleva kaavio sähköistyy hetkeksi: toimituskenttää hallinnoiva editoriohjelma **lukee** lausekkeen ympäriltä tarvitsemansa tiedot, **laskee** vaaditut laskut ja **kirjoittaa** tulokset. Kaikki tapahtuu yhdessä silmänräpäyksessä.

Kun siis kaavion alimmalla rivillä *kuukausi* ja *päivä* aktivoidaan, rivi muuttuu muotoon

`jos vuosi=2012, niin kuukausi=4 ja päivä=8`

eli pääsiäissunnuntai on tänä vuonna 8. huhtikuuta. Muuttamalla vuodeksi 2008 ja aktivoimalla uudelleen saadaan:

`jos vuosi=2008, niin kuukausi=3 ja päivä=23`

Kokeillaan vielä kerran ja kurkataan tulevaisuuteen:

`jos vuosi=2222, niin kuukausi=3 ja päivä=31`

Musteen (tai *Survon*) laskentakaavio toimii siis samaan tapaan kuin mikä tahansa tietokoneohjelma. Erona on se, että laskentakaavion voi kirjoittaa huomattavasti vapaamuotoisemmin, tarvitsematta niin paljoa välittä ohjelmointikielille ominaisista muutosäännöistä.

Tuhat vuotta ja yhdeksän pyhää

Siirrytään tutkimaan suomalaisen kalenterin juhlapäiviä tarkemmin. Keskeisiä, yhteiskunnallisesti merkittäviä juhlia on yhdeksän: uusivuosi, loppiainen, pääsiäinen, helatorstai, vappu, juhannus, pyhäinpäivä, itsenäisyyspäivä ja joulukuu. Näitä juhlia on tapana kutsua myös ”juhlapyhiksi” tai ”pyhäpäiviksi” riippumatta siitä, ovatko ne kirkollisia vai maallisia.

Laskennan kannalta on hyödyllisempää jakaa ”pyhät” kahteen ryhmään sen mukaan, *kumpi vaihtelee vuosittain: viikonpäivä vai päivämäärä*. Juhlat, joiden viikonpäivä vaihtelee, ovat itsenäisyyspäivä, joulukuu, uusivuosi, loppiainen ja vappu. Juhlat, joiden päivämäärä vaihtelee, mutta viikonpäivä ei, ovat pyhäinpäivä, pääsiäinen, helatorstai ja juhannus.

Osa juhlista koostuu pääsiäisen tapaan useista pyhäpäivistä, mutta laskelmissani jokaista juhlaa edustaa täsmälleen yksi päivä: uudenvuodenpäivä, loppiaispäivä, pääsiäissunnuntai, helatorstai, vapunpäivä, juhannuspäivä, pyhäinpäivä, itsenäisyyspäivä ja jouluaatto.

Kalenterissa on useita muitakin yleisiä juhlapäiviä, mutta niillä ei ole vastaavaa merkitystä, koska ne eivät eroa tavallisista arki- tai sunnuntaipäivistä. Esimerkiksi helluntai, jonka vuodesta 1705 suomeksi ilmestynyt *Yliopiston almanakka* sisällyttää virallisiin juhlapäiviin, on seitsemäs pääsiäisen jälkeinen sunnuntai.

Toisinaan juhlapäivien määrittelyt voivat muuttuakin. Helatorstai on viimeiset 20 vuotta ollut perinteisellä paikallaan, torstaina, tasan 40 päivää pääsiäislauantain jälkeen, mutta vuosina 1973–1991 sitä vietettiin jo edeltävän viikon lauantaina nimellä Kristuksen tai vaaseenastumisen päivä.

Aloitin jutun kirjoittamisen kynttiläpäivänä, joka sattui olemaan myös Runebergin päivä. Edellinen on kirkollinen, jälkimmäinen maallinen. Nämäkin juhlat voidaan luokitella mainittuihin ryhmiin: Runebergin päivä voi olla minä viikonpäivänä tahansa, kun taas kynttiläpäivä on aina sunnuntaina. Molemmilla on merkitystä lähinnä niille, jotka kyseisiä päiviä juhlistavat.

Tarkastellaan seuraavaksi kalenteria ja juhlia tilastollisen tietojenkäsittelyn keinoin rakentamalla data eli tilastoaineisto, joka ulottuu vuosisatojen päähän sekä menneisyyteen että varsinkin tulevaisuuteen.

Valitsin tarkasteltavaksi **tuhannen vuoden** pituisen ajanjakson 1.1.1800–31.12.2799. Todellisuudessa osaa juhlia ei ole vietetty vielä sataakaan kertaa, mutta laskennallisia kokeiluja se ei mitenkään estä.

Nyt jutun toimituskentästä poimituissa näkymissä on mukana myös eräitä Musteen ja Survon editoriaalisen käyttötavan hallintaan kuuluvia elementtejä. Niistä näkyvimpiä ovat vasemman reunan *kontrollisarake* ja toimituskenttää osiin jakavat *rajarivit*.

Kontrollisarake koostuu oletuksena tähdistä *, mutta jokainen rivi voidaan varustaa millä tahansa muullakin kontrollimerkillä. Niihin voi kytkeä erilaisia toimintoja, joista alla olevissa näkymissä on joitain esimerkkejä. Selkeyden vuoksi näkyvissä ovat vain välttämättömät kontrollimerkit; todellisuudessa tähdet ynnä muut kontrollimerkit tuikkivat himmeämmin erottuen siten varsinaisesta tekstinkirjoitusalueesta.

Rajarivi erottaa laskentakaavioita tai komentoja ohjaavia *täsmennyksiä* eli merkillä = varustettuja sanoja. Sen muodostaa vähintään kymmenen peräkkäistä pistettä tähtirivillä. Usein rajarivit vedetään täyteen leveyteen:

```
*.....
Tämän aktivointi aktivoi kaikki '+'-lla merkityt rivit:
/ACTIVATE +

+TIME COUNT START / mitataan rakentamiseen kuluva aika

+FILE CREATE KALJU / luodaan tyhjä datatiedosto (.svo)
KALJU: KAlenteri ja JUhlat / KV 15.2.2010 (19.2.2012)
FIELDS:
 1 S 10 pvm      päivämäärä muodossa vvvv-kk-pp
 2 N 4 pvJulian päivän järjestysno (1 = 1.1.1800)
 3 N 1 pvViikko viikonpäivä (1-7)
 4 S 2 pvNimi   viikonpäivä [ma,ti,ke,to,pe,la,su]
 5 N 2 pvVuosi  päivän järjestysnumero (1-366)
 6 N 1 vkoVuosi viikon järjestysnumero (1-53)
 7 S 3 Juhla    [Juh,Pyh,Its,Jou,Uus,Lop,Paa,Vap,Het]
 8 S 4 vvvv     pvm: vuosi (1800-2799)
 9 S 2 kk       pvm: kuukausi [01,02,...,12]
10 S 2 pp       pvm: päivä/kk [01,02,...,31]
END
```

Koska vuosia on tasan tuhat, on päiviä melko tarkkaan 365000. Lisäksi tulee pieni määrä karkauspäiviä, joilla gregoriaaninen kalenteri pysyy synkronissa maan kiertoliikkeiden kanssa. Karkauspäivien idea on selostettu tarkemmin muun muassa Matti Lehtisen alussa mainitussa artikkelissa. Karkeasti karkauspäivä on joka neljäs vuosi, joten niitä osuu jaksolle vajaat 250.

Lasketaan päivien määrä tarkasti DATE-komennolla ja rakennetaan sen jälkeen vaiheittain koko data:

```
*.....
DATE 1.1.2800 - 1.1.1800 / Difference=365242

+FILE INIT KALJU,365242 / ajellaan data kaljuksi
+VAR pvJulian=ORDER TO KALJU / ja numeroidaan päivät

Mitoitetaan päivien numerot tähän dataan sopivammiksi:
DATE 1.1.1800,Julian / Jan 01 1800 Julian_day=2378497
JULIAN_DAY0=2378496 (31.12.1799) -> 1, 2, ..., 365242
ja muodostetaan DATE-komennolla pvm ym. muuttujia:

+DATE KALJU / ODATE=YYYYMMDD ODEL=- MASK=di-ajw----
Vähän selostusta yllä käytetyistä täsmennyksistä:
ODATE ("output date") ja ODEL ("output delimiter")
ohjaavat päivämäärät haluttuun muotoon "2012-02-19".
MASK määrää kullekin valitulle muuttujalle roolin,
esim. d ("date"), i ("input") ja w ("week number").
*.....
```

Suomennetaan ja numeroidaan päivien lyhennetyt nimet:

```
+CLASSIFY KALJU,Viikko1,pvNimi,pvNimi
CLASSIFICATION Viikko1
Mo: ma
Tu: ti
We: ke
Th: to
Fr: pe
Sa: la
Su: su
END
```

```
+CLASSIFY KALJU,Viikko2,pvNimi,pvViikko
CLASSIFICATION Viikko2
ma: 1
ti: 2
ke: 3
to: 4
pe: 5
la: 6
su: 7
END
```

*.....

Erotellaan pvm:stä muuttujat vuosi, kuukausi ja päivä:
IDATE=YYYYMMDD IDEL=- (kuten edellä, mutta nyt "input")

```
+DATE KALJU / ODATE=YYYY VARS=pvm(D),vvvv(d)
+DATE KALJU / ODATE=MM VARS=pvm(D),kk(d)
+DATE KALJU / ODATE=DD VARS=pvm(D),pp(d)
```

*.....

Datan alku- ja loppupää näyttävät nyt tällaisilta:

```
#FILE LOAD -KALJU CUR+1 / IND=ORDER,1,10
1800-01-01 1 3 ke 1 1 - 1800 01 01
1800-01-02 2 4 to 2 1 - 1800 01 02
1800-01-03 3 5 pe 3 1 - 1800 01 03
1800-01-04 4 6 la 4 1 - 1800 01 04
1800-01-05 5 7 su 5 1 - 1800 01 05
1800-01-06 6 1 ma 6 2 - 1800 01 06
1800-01-07 7 2 ti 7 2 - 1800 01 07
1800-01-08 8 3 ke 8 2 - 1800 01 08
1800-01-09 9 4 to 9 2 - 1800 01 09
1800-01-10 10 5 pe 10 2 - 1800 01 10
```

```
#FILE LOAD -KALJU CUR+1 / IND=ORDER,365233,365242
2799-12-22 365233 3 ke 356 51 - 2799 12 22
2799-12-23 365234 4 to 357 51 - 2799 12 23
2799-12-24 365235 5 pe 358 51 - 2799 12 24
2799-12-25 365236 6 la 359 51 - 2799 12 25
2799-12-26 365237 7 su 360 51 - 2799 12 26
2799-12-27 365238 1 ma 361 52 - 2799 12 27
2799-12-28 365239 2 ti 362 52 - 2799 12 28
2799-12-29 365240 3 ke 363 52 - 2799 12 29
2799-12-30 365241 4 to 364 52 - 2799 12 30
2799-12-31 365242 5 pe 365 52 - 2799 12 31
```

Data on valmis! Vain juhlapäivien tiedot puuttuvat (-).

Edellä on käytössä *risuaitatekniikka*, jolla Muste lisää itse toimituskenttään tarvitsemansa tilan tulostuksille (tässä datan listaukselle). Rivitunnus CUR viittaa komentoriviin; vastaavasti END kentän viimeiseen eityhjään riviin. Ilman risuaitoja # tilan hallinnasta vastaa käyttäjä, jolloin END+2 on suositeltavampi valinta tulostuksen aloitusriviksi kuin yllä sovellettu CUR+1.

Yhdeksän tuhatta juhlaa

Täydennetään data merkkiaamalla juhlapäivät muuttu- jaan **Juhla**. Juhlista käytetään kolmikirjaimisia lyhen- teitä, jotka näkyvät FILE CREATE -kaaviosta.

Aloitetaan helpoimmista eli niistä juhlista, joiden si- jainti kalenterissa määräytyy päivämäärän perusteella, mutta viikonpäivä vaihtelee vuosittain. Näitä ovat siis itsenäisyyspäivä, joulukuusi, uusivuosi, loppiainen ja vappu.

Kun VAR-komento aktivoidaan, se kirjoittaa kyseisen juhlan lyhenteen dataan kaikkien ehdot täyttävien päi- vien kohdalle. Jokainen VAR-komento käy siis läpi koko datan ja päivittää tuhat tietuetta, yhden per vuosi:

*.....

Ehtolauseke SELECT=K*P on yhteinen ja yksinkertainen: sen mukaan K:n ja P:n tulee molempien toteutua, jotta koko SELECT-ehto toteutuu. Käytetyt alkeisehdot K ja P määrittelevät juhlat täsmällisesti:

1) Itsenäisyyspäivä: 6.12. K=kk,12 P=pp,6
+VAR str(Juhla)="Its" TO KALJU

2) Joulukuusi: 24.12.
+VAR str(Juhla)="Jou" / K=kk,12 P=pp,24

3) Uusivuosi: 1.1.
+VAR str(Juhla)="Uus" / P=kk,1 K=pp,1

4) Loppiainen: 6.1.
+VAR str(Juhla)="Lop" / K=kk,1 P=pp,6

5) Vappu: 1.5.
+VAR str(Juhla)="Vap" / K=kk,5 P=pp,1

*.....

Äskeisessä (kohdat 1-5) ei tarvita rajarivejä erotta- maan VAR-komentoja, sillä itse komentorivi on etusi- jalla, kun Muste lukee täsmennyksiä. Näin SELECT vai- kuttaa kaikkiin viiteen VAR-komentoon, mutta alkeiseh- dot K ja P määritellään tilannekohtaisesti. Kohdassa 1 käyttöön tulevat tässä rajarivien välisessä *osakentässä* ensimmäisenä annetut määrittelyt.

Seuraavassa (kohdat 6-7) rajariviä ei myöskään tarvita, kun SELECT-ehto on komentokohtainen, ja alkeisehdot ovat viikonpäivää lukuunottamatta eri nimisiä:

*.....

6) Juhannus: lauantai 20.-26.6.

K=kk,6 P=pp,20,26 V=pvNimi:la

+VAR str(Juhla)="Juh" / SELECT=K*P*V

7) Pyhäinpäivä: lauantai 31.10.-6.11.

K1=kk,10 P1=pp,31 (toisinaan lokakuun puolella!)
K2=kk,11 P2=pp,1,6

+VAR str(Juhla)="Pyh" / SELECT=(K1*P1+K2*P2)*V

*.....

Vuorossa (kohta 8) on juhlista vaativin. Pääsiäinen on aina sunnuntaina, mutta se tieto ei riitä: tarvitaan taas alussa esitettyä algoritmia. Jotta päästään soveltamaan sitä, poimitaan erilleen yksi päivä jokaisesta vuodesta. Muuttujista mukaan riittää ottaa päivämäärän kolme komponenttia sekä juhlapäivän ilmaisin:

```
*.....
      8) Pääsiäinen: sunnuntai 22.3.-25.4.
+FILE AGGR KALJU BY vvvv TO KALJUPAA / (!)
VARIABLES:
vvvv  LAST  vvvv  / Jokaisesta tuhannesta vuodesta
kk    LAST  kk    / poimitaan sen viimeinen päivä
pp    LAST  pp    / (LAST) ja kootaan näistä uusi,
Juhla LAST  Juhla / vain tuhannen havainnon data.
END
*.....
```

Edellä esitetty algoritmi on havainnollisuuden vuoksi toistettu tässä ilman ylimääräisiä kommentteja, vain hieman eri tavalla aseteltuna. Nyt vuositieto haetaan uuden datan muuttujasta vvvv, ja VAR-aktivointi päivittää samaisen datan muuttujien pp ja kk sisällöt:

```
Anonymous (1876). To find easter. Nature, 13, 487.
a=mod(vvvv,19)  f=int((b+8)/25)  d=int(b/4)
b=int(vvvv/100) g=int((b-f+1)/3)  e=mod(b,4)
c=mod(vvvv,100) n=(h+1-7*m+114)  i=int(c/4)
h=mod((19*a+b-d-g+15),30)        k=mod(c,4)
l=mod((32+2*e+2*i-h-k),7)        kk=int(n/31)
m=int((a+11*h+22*l)/451)         pp=mod(n,31)+1

+VAR pp,kk TO KALJUPAA
+VAR str(Juhla)="Paa" / (ei ehtoja: lisätään kaikkiin)
*.....

Isoon dataan kopiointia varten lisätään etunollat, niin
datojen täsmäytys MATCH-täsmennyksellä onnistuu:

+VAR str(pp,1,1)="0" / IND=pp,1,9
+VAR str(kk,1,1)="0" / IND=kk,3,4 (vältetään rajarivi)
*.....

Kopioidaan pääsiäiset omasta datastaan isoon dataan:

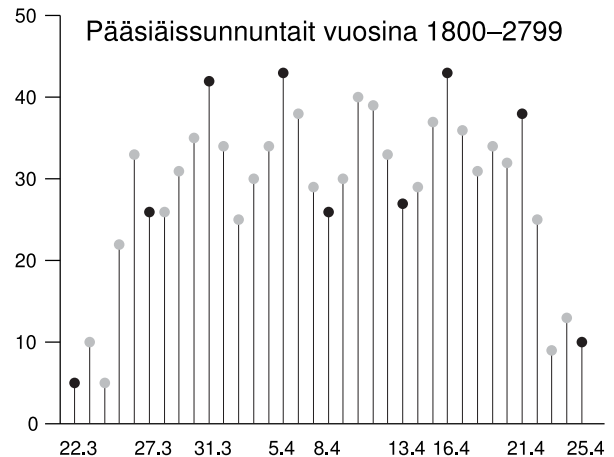
MATCH=vvvv,kk,pp (kaikkien pitää täsmätä!)
+FILE COPY KALJUPAA TO KALJU / VARS=Juhla
*.....

Nyt kaikki tuhat pääsiäistä ovat oikeilla paikoillaan.
Enää on jäljellä helatorstai. Koska sen sijainti riippuu
täysin pääsiäisestä, kohta 9 hoituu helposti:

*.....

      9) Helatorstai: 39. päivä pääsiäissunnuntaista
+VAR str(Juhla[+39])="Het" TO KALJU / CASES=Juhla:Paa
+TIME COUNT END      60.032

Aikaa kului siis noin minuutti: 60.032(s:hms)=00:01:00
*.....
```



Pääsiäisten kuvailua

Kuvion perusteella varhaiset ja myöhäiset pääsiäiset esiintyvät harvimmin. Tämä pätee yleisesti, vieläpä niin, että kaikkein varhaisin on kaikkein harvinaisin.

Tuhat vuotta saattaa tuntua pitkältä ajalta, mutta se on lopulta vain lyhyt jakso, sillä pääsiäissunnuntain päivämäärien kiertokulku toistuu samanlaisena vasta **5,7 miljoonan vuoden** jälkeen. Jakson täydellinen kuva löytyy muun muassa pääsiäisen laskennallisia kysymyksiä käsittelevältä englanninkieliseltä Wikipedia-sivulta.

Kuvanpiirtoa varten edellä (kohdassa 8) syntynyt pääsiäisdata on järjestetty kuukauden ja päivän mukaan ja talletettu lajitteluavain uudeksi muuttujaksi kkpp. Sen avulla on muodostettu uusi data, joka sisältää vain 35 havaintoa; yhden jokaisesta mahdollisesta pääsiäissunnuntain päivämäärästä. Muuttujaan n tulevat kuvan pystyakselilla esitetyt lukumäärät:

```
FILE SORT KALJUPAA BY kk,pp TO PAA1000 / KEY_SAVED=kkpp
```

```
FILE AGGR PAA1000 BY kkpp TO PAA1000A
VARIABLES:
i:S2  MISSING  -
kkpp  FIRST    kkpp / aggregointitunniste
v1    ORDER(1) vvvv / 1. vuosi
v2    ORDER(2) vvvv / 2. vuosi
v3    ORDER(3) vvvv / 3. vuosi
vj    ORDER(-3) vvvv / j. vuosi (j=n-3+1)
vk    ORDER(-2) vvvv / k. vuosi (k=n-2+1)
vn    ORDER(-1) vvvv / n. vuosi (n=n-1+1)
n:S2  N        - / vuosien lukumäärä
color:1 MISSING -
END
*.....
```

Muuttujiin v1, v2, ..., vn tallentuu kolme ensimmäistä ja kolme viimeistä vuotta, jolloin pääsiäinen on (tai on ollut) kyseisen havainnon ilmaisemana päivänä.

Dataa on aina hyvä selailla, jotta pysyy selvillä siitä, mitä on tekemässä. Nyt sen alku maaliskuun päivien osalta näyttää tällaiselta:


```
FILE LOAD PAA1000A CUR+2 / CASES_WILD*** CASES=kkpp:03*
```

```
*DATA PAA1000A*,A,B,C
```

```
C i kkpp v1 v2 v3 vj vk vn n col
A - 0322 1818 2285 2353 2353 2437 2505 5 -
  - 0323 1845 1856 1913 2532 2600 2752 10 -
  - 0324 1940 2391 2475 2475 2543 2695 5 -
  - 0325 1883 1894 1951 2638 2779 2790 22 -
  - 0326 1815 1826 1837 2711 2722 2733 33 -
  - 0327 1842 1853 1864 2738 2749 2760 26 -
  - 0328 1869 1875 1880 2624 2771 2776 26 -
  - 0329 1807 1812 1891 2714 2787 2798 31 -
  - 0330 1823 1834 1902 2719 2730 2741 35 -
B - 0331 1839 1850 1861 2746 2757 2768 42 -
```

```
*.....
```

Listauksesta näkyy, kuinka harvinainen pääsiäisen esiintyminen 22. maaliskuuta oikein on: viime kerrasta on jo 2012-1818=194 vuotta. Seuraavaa saadaan odottaa vielä 2285-2012=273 vuotta.

Vilkaistaan toisesta datasta, miltä maaliskuun pääsiäisten tilanne näyttää seuraavien 20 vuoden aikana:

```
CASES=kk:03 IND=vvvv,2013,2032 MASK=AAA--
FILE LOAD -PAA1000 CUR+2
```

```
2016 03 27
2027 03 28
2032 03 28
2013 03 31
2024 03 31
```

```
*.....
```

Näemmä pääsiäissunnuntai osuu maaliskuun puolelle heti ensi vuonna. Kuvan perusteella 31. maaliskuuta onkin varsin tyypillinen pääsiäissunnuntain ajankohta.

Juhlat törmäyskurssilla

Tarkastellaan lopuksi, millaisia tilanteita voi syntyä kirkollisten ja maallisten juhlapäivien kohdatessa. Taas tapahtumien keskipisteessä on pääsiäinen.

Jos pääsiäinen on myöhään, se sijoittuu vapun lähelle. Väliin jää vähimmillään vain viisi päivää, joista yksi on pääsiäismaanantai. Kaupallisissa sovelluksissa, joissa saatetaan olla kiinnostuneita sekä juhlia edeltävistä että seuraavista myyntipäivistä, on tällöin vaikea erottaa näiden kahden juhlan vaikutusalueita toisistaan.

Jos pääsiäinen on aikaisin, sijoittuu helatorstai vapun lähelle. Tämä onkin kaikista törmäyskurseista mielenkiintoisin. Pääsiäisen liikkuvuudesta seuraa, että helatorstai, joka on yleensä toukokuussa, on toisinaan vasta kesäkuun alussa. Aikaisimmillaan helatorstai voi olla jopa huhtikuussa, samana päivänä kuin vappuaatto:

```
Haetaan päivät isosta datasta: CASES=Juhla:Het IND=kk,4
FILE LOAD -KALJU CUR+2 / VARS=Juhla,pvm
```

```
Het 1818-04-30
Het 2285-04-30
Het 2353-04-30
Het 2437-04-30
Het 2505-04-30
```

```
*.....
```

Kyseiset viisi esiintymää ovat täsmälleen niinä vuosina, jolloin pääsiäinen on aikaisimmillaan, 22. maaliskuuta:

```
CASES=Juhla:Paa IND=vvvv:1818,2285,2353,2437,2505
FILE LOAD -KALJU CUR+2 / VARS=Juhla,pvm
```

```
Paa 1818-03-22
Paa 2285-03-22
Paa 2353-03-22
Paa 2437-03-22
Paa 2505-03-22
```

```
*.....
```

R ja kadonneiden juhlien metsästys

Tehdään tässä välissä pieni vierailu R-puolelle, joka on kokoajan käytettävissä Musteen välityksellä. Ensin muunnetaan Musteen KALJU-data R-dataksi kalju:

```
FILE LOAD KALJU TO R>kalju
```

Raaputetaan kaljusta pari tietoa R-funktioilla, joita Musteen toimituskentästä aktivoidaan joko komennolla R tai napilla Ctrl-R:

```
dim(kalju) # tarkistetaan datan dimensiot
```

R-funktion dim tulos kirjoittuu oletuksena Musteen rinnalla toimivaan R:n omaan komentoikkunaan, mutta minkä tahansa R-funktion tulokset saa myös halutesaan suoraan Musteen toimituskenttään jatkokäsittelyä varten. Nyt tulos sisältää ainoastaan kaksi lukua:

```
[1] 365242 10
```

R-funktiolla summary tiivistetään tiedot kevätjuhlista:

```
summary(as.factor(kalju$Juhla))[c('Paa','Het','Vap')]
```

```
Paa Het Vap
1000 1000 990
```

Mihin on kadonnut 10 vappua? Kun juhlapäivät edellä merkattiin dataan, oli selvää, että kaikkia on täsmälleen tuhat kappaletta. Siinä yhteydessä ei kuitenkaan huomattu, että *juhlat voivat osua samallekin päivälle*. Aina kun tekee datalle jotain, sitä olisi syytä katsoa.

Katsotaan nyt tarkemmin Musteen STAT-komennolla:

STAT KALJU CUR+2 / VARS=Juhla

Basic statistics: KALJU N=365242

Variable: Juhla [Juh,Pyh,Its,Jou,Uus,Lop,Paa,Vap,Het]
N(missing)=356252

Juhla	f	%	*=32 obs.
Het	1000	11.1	*****
Its	1000	11.1	*****
Jou	1000	11.1	*****
Juh	1000	11.1	*****
Lop	1000	11.1	*****
Paa	1000	11.1	*****
Pyh	1000	11.1	*****
Uus	1000	11.1	*****
Vap	990	11.0	*****

*.....

Nuo 10 vapunpäivää ovat näemmä ainoat kadonneet. Ne törmäävät helatorstaihin, joka juhlapäiviä merkattaessa (kohdassa 9) ajaa aiemmin (kohdassa 5) merkattun vapun yli. Törmäyskurssi on sangen harvinainen, mutta sattumalta koettu aivan hiljattain, vuonna 2008:

```
helatorstai=Juhla:Het SELECT=helatorstai*toukokuun*1
toukokuun=kk,5 1=pp,1
FILE LOAD KALJU CUR+2 / VARS=vvvv FORMAT=LIST
```

```
DATA vvvv: 1845 1856 1913 2008 2160 2228 2380 2532 2600
2752 END
```

*.....

Näemmä edellisen kerran näin tapahtui 99 vuotta sitten, ennen ensimmäistä maailmansotaa. Jutun alussa mainitussa analysointijärjestelmässä, joka käsittelee aikasarja-aineistoja vuodesta 2000 lähtien, vuoden 2008 törmäyksellä oli työllistävä vaikutus: jouduin otamaan sen erikseen huomioon yli 20 kohdassa ohjelman lähdekoodia. Kuten edeltä näkyy, seuraava törmäys odottaa – mutta vasta noin 150 vuoden päässä.

Lopuksi

Pääsiäisaiheisen jutun myötä olen tullut käyneeksi läpi muutamia esimerkkejä Musteen ja Survon käytöstä laskelmien teossa, datan muokkauksessa ja kuvien piirroksa. Nämä (ja monet muut) tehtävät sujuvat mukavasti editoriaalisella käyttötavalla. Siitä tuli Survon käyttöliittymä vuonna 1979, jolloin se nopeasti syrjäytti valikopohjaisen käyttötavan silloisessa SURVO 76:ssa (ks. www.survo.fi/julkaisut erit. vuosilta 1979–1980).

Lyhyesti lähteistä

Heikki Ojan *Aikakirja* on varsinainen kalentereiden ja juhlapäivien runsaudensarvi. Jean Meeusin teos on erinomainen johdatus astronomisten algoritmien lähteille. Reinhold Bien valottaa historian hämää.

Jatkot verkossa!

Artikkelin työkaaviot ynnä muuta täydentävää materiaalia löytyy sivulta www.survo.fi/juhlat. Ottamalla käyttöön Musteen (sivulta www.survo.fi/muste) voi toistaa kaikki jutussa esitetyt datan käsittelyt, piirtää edellä olevan kuvan tai vaikka laajentaa kokeiluja vielä kauemmas historiaan tai tulevaisuuteen.

Kiitokset

Seppo Mustonen ja Reijo Sund esittivät eri vaiheissa lukuisia hyödyllisiä kommentteja. Kiitos kaikista!

Viitteet

- [1] <http://en.wikipedia.org/wiki/Computus>.
- [2] Anonymous. To find easter. *Nature*, 13:487, 1876.
- [3] Reinhold Bien. Gauss and beyond: The making of easter algorithms. *Archive for History of Exact Sciences*, 58:439–452, 2004.
- [4] Ross Ihaka and Robert Gentleman. R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5:299–314, 1996.
- [5] Matti Lehtinen. Jouluaatto on harvemmin sunnuntaina. *Solmu*, (3):17–18, 2011.
- [6] Jean Meeus. *Astronomical Algorithms*. Willmann-Bell, 2nd edition, 1998.
- [7] Seppo Mustonen. SURVO MM: käyttöympäristö tekstin ja numeerisen tiedon luovaan käsittelyyn. 2001–2012, <http://www.survo.fi>.
- [8] Heikki Oja. *Aikakirja*. Helsingin yliopiston almanakkatoimisto, Helsingin yliopisto, 2007. <http://almanakka.helsinki.fi/aikakirja/Aikakirja2007kokonaan.pdf>.
- [9] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2012. <http://www.R-project.org>.
- [10] Reijo Sund. Muste – Multiplatform Survo Type Editorial Environment for Data Analysis. 2009–2012, <http://www.survo.fi/muste>.