TAMPEREEN YLIOPISTO
Pro gradu -tutkielma

Jussi Alho

# Tarkistusnumeroiden matematiikkaa

Matematiikan ja tilastotieteen laitos
Matematiikka
Maaliskuu 2008

Tampereen yliopisto

Matematiikan ja tilastotieteen laitos

ALHO, JUSSI: Tarkistusnumeroiden matematiikkaa

Pro gradu -tutkielma, 24 s.

Matematiikka

Maaliskuu 2008

---

# Tiivistelmä

Tutkielmassa käsitellään erilaisiin tunnistenumeroihin liittyviä tarkistusnumeroita. Tutkielmassa esitellään useita tarkistusnumerojärjestelmiä sekä kerrotaan niihin liittyvistä virhemahdollisuuksista. Luvussa 1 esitellään tutkielmassa tarvittavia lukuteorian perusominaisuuksia, kuten jaollisuus, suurin yhteinen tekijä ja kongruenssi. Luvussa 2 käsitellään tutkielman varsinaista aihetta, eli erilaisia tarkistusnumerojärjestelmiä. Tarkistusnumeroita löytyy esimerkiksi tuotteiden, passien, ajokorttien, kirjojen ja henkilötunnuksien tunnistenumeroista, joista tutkielmassa on useita esimerkkejä. Tutkielman viimeisessä luvussa kerrotaan, millaisia virheitä tunnistenumeroissa voi esiintyä ja miten ne voidaan havaita tarkistusnumeroiden avulla. Tutkielman lähdeteoksina on käytetty useita kirjoja ja artikkeleita, joista keskeisimmiksi nousevat Thomas Koshyn *Elementary Number Theory with Applications* sekä Joseph Kirtlandin *Identification Numbers and Check Digit Schemes*. Lukijalta edellytetään kokonaisluvun käsitteen tuntemista sekä kokonaislukuihin liittyvien peruslaskutoimitusten hallitsemista.

# Sisältö

# Johdanto

Kyky säilyttää, hakea ja siirtää tietoa täsmällisesti on erittäin keskeisessä asemassa tämän päivän yhteiskunnassa. Jotta se tapahtuisi tehokkaasti, käytetään tunnistenumeroita tiedon esittämiseen ja koodaamiseen esimerkiksi erilaisissa tuotteissa, dokumenteissa, asiakastileissä sekä ihmisten yksilöimisessä. Tunnistenumerot mahdollistavat yksinkertaisen ja täsmällisen tietojen siirtämisen, joten niitä kirjataan dokumentteihin, tallennetaan tietokoneille, lähetetään internetin välityksellä tai siirretään jollain muulla tavalla miljoonia kertoja päivässä. Koska tällaisia tapahtumia esiintyy niin usein, tapahtuu varmasti myös virheitä. Virheen sattuessa onkin ratkaisevaa tietää, ovatko tunnistenumerot siirtyneet oikeassa muodossaan. Tunnistenumeron vastaanottajan täytyy siis jollain tavalla varmistaa numeron oikeellisuus. Aina ei ole kuitenkaan mahdollista tavoittaa viestin lähettäjää, joten on kehitetty menetelmiä, joiden avulla vastaanottaja voi tarkistaa, onko numerossa virheitä. Tämän tutkielman tarkoituksena on esitellä joitain matemaattisia menetelmiä, eli tarkistusnumerojärjestelmiä, jotka mahdollistavat tunnistenumeroiden tarkistamisen.

Tutkielman luvussa 1 esitellään myöhemmin tarvittavia lukuteorian perusominaisuuksia. Alaluvussa 1.1 määritellään jaollisuuden käsite ja esitetään joitain sen perusominaisuuksia. Lisäksi määritellään suurin yhteinen tekijä. Alaluvussa 1.2 puolesteen määritellään kongruenssi, joka on erittäin keskeisessä asemassa tässä tutkielmassa. Kongruenssista esitetään myös joitain keskeisiä tuloksia.

Luvussa 2 käsitellään tutkielman varsinaista aihetta, tarkistusnumerojärjestelmiä. Alaluvussa 2.1 esitellään binäärikoodit sekä määritellään tutkielman kannalta olennainen pistetulon käsite. Alaluku 2.2 käsittelee kauppojen tuotteista löytyviä UPC-koodeja. Kaikista kirjallisista teoksista löytyviä ISBN-numeroita käsitellään alaluvussa 2.3. Alaluvussa 2.4 puolestaan esitellään Yhdysvalloissa käytössä olevaa postinumerojärjestelmää ja siihen liittyviä ominaisuuksia. Alaluku 2.5 käsittelee eri puolilla maailmaa käytössä olevia ajokortti- ja passinumerojärjestelmiä. Kaikilla suomen kansalaisilla on henkilötunnus, johon perehdytään hieman tarkemmin alaluvussa 2.6. Viimeinen luku 2.7 käsittelee tärkeää asiaa, eli virheitä, joita tunnistenumeroissa voi esiintyä.

Tämän tutkielman lukijalta edellytetään kokonaisluvun käsitteen tuntemista sekä kokonaislukuihin liittyvien peruslaskutoimitusten hallitsemista. Lisäksi summamerkinnän käyttö sekä kertoman ja itseisarvon käsitteet oletetaan tunnetuiksi. Tutkielman pääasiallisina lähteinä on käytetty Thomas Koshyn kirjaa *Elementary Number Theory with Applications* sekä Kenneth H. Rosenin teosta *Elementary Number Theory and its Applications*. Muina lähteinä on käytetty Joseph Kirtlandin kirjaa *Identification Numbers and Check Digit Schemes*, Joseph A. Gallianin artikkelia *The Mathematics of Identification Numbers* sekä Suomen Väestörekisterikeskuksen verkkosivuja.

# 1 Esitietoja

Tässä luvussa esitellään käsitteitä ja tuloksia, joita tullaan tarvitsemaan myöhemmin tässä tutkielmassa. Ensimmäisessä alaluvussa määritellään jaollisuuden käsite ja esitellään siihen liittyviä tuloksia, jotka ovat olennaisia tämän tutkielman kannalta. Lisäksi määritellään suurimman yhteisen tekijän käsite. Toinen alaluku käsittelee kongruenssia, josta niin ikään esitellään sellaisia tuloksia, joita tarvitaan tämän tutkielman myöhemmissä vaiheissa. Luvun lähdeteoksena on käytetty Kenneth H. Rosenin kirjaa *Elementary Number Theory and its Applications*.

## 1.1 Jaollisuus

Kun kokonaisluku jaetaan toisella kokonaisluvulla ($\neq 0$), osamäärä ei välttämättä ole aina kokonaisluku. Esimerkiksi $32/8 = 4$ on kokonaisluku, kun taas $19/5 = 3.8$ ei ole kokonaisluku. Tämä havainto johtaa seuraavaan määritelmään.

**Määritelmä 1.1.** Olkoot $a$ ja $b$ kokonaislukuja ja olkoon lisäksi $a$ nollasta eroava. Luku $a$ *jakaa* luvun $b$, eli luku $a$ on luvun $b$ *tekijä*, jos on olemassa sellainen kokonaisluku $c$, että $b = ac$.

Jos luku $a$ jakaa luvun $b$, niin merkitään $a \mid b$. Muussa tapauksessa merkitään $a \nmid b$.

**Esimerkki 1.1.** Luku 3 on luvun 12 tekijä, sillä $12 = 3 \cdot 4$. Siispä $3 \mid 12$. Toisaalta, luku 3 ei ole luvun 10 tekijä, sillä ei ole olemassa kokonaislukua $c$ siten, että $10 = 3 \cdot c$. Näin ollen $3 \nmid 10$.

**Esimerkki 1.2.** Luvun 6 tekijät ovat $\pm 1$, $\pm 2$, $\pm 3$ ja $\pm 6$. Luvun 5 tekijät ovat $\pm 1$ ja $\pm 5$.

Seuraavissa luvuissa tarvitaan joitain jaollisuuden perusominaisuuksia, jotka nyt osoitetaan.

**Lause 1.1.** *Olkoot $a$, $b$ ja $c$ kokonaislukuja. Tällöin, jos $a \mid b$ ja $b \mid c$, niin $a \mid c$.*

*Todistus* (vrt. [4, s. 37]). Koska $a \mid b$ ja $b \mid c$, niin on olemassa sellaiset kokonaisluvut $e$ ja $f$, että $ae = b$ ja $bf = c$. Siten,

$$c = bf = (ae)f = a(ef),$$

missä $ef$ on kokonaisluku. Siis määritelmän perusteella $a \mid c$. □

**Esimerkki 1.3.** Koska $8 \mid 48$ ja $48 \mid 144$, niin lauseen 1.1 mukaan $8 \mid 144$.

**Lause 1.2.** *Olkoot $a$, $b$, $m$ ja $n$ kokonaislukuja. Tällöin, jos $c \mid a$ ja $c \mid b$, niin $c \mid (ma + nb)$.*

*Todistus* (vrt. [4, s. 37]). Koska $c \mid a$ ja $c \mid b$, niin on olemassa sellaiset kokonaisluvut $e$ ja $f$, että $a = ce$ ja $b = cf$. Siten,

$$ma + nb = mce + ncf = c(me + nf),$$

missä $me + nf$ on kokonaisluku. Siis määritelmän perusteella
$c \mid (ma + nb)$. □

**Esimerkki 1.4.** Koska $4 \mid 16$ ja $4 \mid 28$, niin lauseen 1.2 mukaan $4 \mid 12$, koska $6 \cdot 16 - 3 \cdot 28 = 96 - 84 = 12$.

Määritellään vielä suurimman yhteisen tekijän käsite, jota tullaan myös tarvitsemaan myöhemmin tässä tutkielmassa.

Olkoot $a$ ja $b$ sellaisia lukuja, että ainakin toinen on $\neq 0$. Tällöin lukujen $a$ ja $b$ yhteisten tekijöiden joukko on äärellinen joukko lukuja, joka sisältää aina luvut 1 ja $-1$. Nyt ollaan kiinnostuneita suurimmasta luvusta, joka löytyy kahden luvun yhteisten tekijöiden joukosta.

**Määritelmä 1.2.** Kokonaislukujen $a$ ja $b$, joista ainakin toinen on $\neq 0$, *suurin yhteinen tekijä* on suurin luku, joka jakaa sekä luvun $a$ että luvun $b$.

Lukujen $a$ ja $b$ suurinta yhteistä tekijää merkitään symbolilla $(a, b)$. Määritellään lisäksi, että $(0, 0) = 0$.

**Esimerkki 1.5.** Lukujen 24 ja 60 yhteiset tekijät ovat $\pm 1$, $\pm 2$, $\pm 3$, $\pm 4$, $\pm 6$ ja $\pm 12$. Näin ollen, $(24, 60) = 12$. Vastaavasti, $(12, 81) = 3$, $(200, 5) = 5$, $(19, 25) = 1$, $(0, 36) = 36$, $(-9, -15) = 3$ ja $(-13, 221) = 13$.

## 1.2 Kongruenssi

**Määritelmä 1.3.** Olkoon $m$ positiivinen kokonaisluku. Sanotaan, että kokonaisluku $a$ on *kongruentti* luvun $b$ kanssa *modulo* $m$, jos

$$m \mid (a - b).$$

Jos luku $a$ on kongruentti luvun $b$ kanssa modulo $m$, niin merkitään

$$a \equiv b \pmod{m}.$$

Jos $m \nmid (a - b)$, niin merkitään $a \not\equiv b \pmod{m}$ ja sanotaan, että luku $a$ on epäkongruentti luvun $b$ kanssa modulo $m$. Kokonaislukua $m$ sanotaan *moduliksi*.

**Esimerkki 1.6.** Luku $26 \equiv 5 \pmod{7}$, sillä $7 \mid (26 - 5) = 21$. Vastaavasti $18 \not\equiv 2 \pmod{5}$, sillä $5 \nmid (18 - 2) = 16$.

**Lause 1.3.** *Olkoot $a$ ja $b$ kokonaislukuja. Tällöin $a \equiv b \pmod{m}$, jos ja vain jos on olemassa sellainen kokonaisluku $k$, että $a = b + km$.*

*Todistus* (vrt. [4, s. 120]). Jos $a \equiv b \pmod{m}$, niin kongruenssin määritelmän mukaan $m \mid (a-b)$. Tämä tarkoittaa, että on olemassa kokonaisluku $k$ siten, että $km = a - b$, joten $a = km + b$.

Toisaalta, jos on olemassa kokonaisluku $k$ siten, että $a = b + km$, niin $km = a - b$. Siis jaollisuuden määritelmän mukaan $m \mid (a-b)$ ja edelleen $a \equiv b \pmod{m}$. □

**Esimerkki 1.7.** Luku $29 \equiv 5 \pmod{6}$ ja $29 = 5 + 4 \cdot 6$.

**Lause 1.4.** *Olkoon $m$ positiivinen kokonaisluku. Tällöin kongruensseille modulo $m$ pätevät seuraavat ominaisuudet:*

(1) *(Refleksiivisyys) Jos $a$ on kokonaisluku, niin $a \equiv a \pmod{m}$.*

(2) *(Symmetrisyys) Jos $a$ ja $b$ ovat sellaisia kokonaislukuja, että $a \equiv b \pmod{m}$, niin $b \equiv a \pmod{m}$.*

(3) *(Transitiivisuus) Jos $a, b$ ja $c$ ovat sellaisia kokonaislukuja, että $a \equiv b \pmod{m}$ ja $b \equiv c \pmod{m}$, niin $a \equiv c \pmod{m}$.*

*Todistus* (vrt. [4, s. 121]).

(1) Koska $m \mid (a-a) = 0$, niin kongruenssin määritelmän mukaan $a \equiv a \pmod{m}$.

(2) Oletetaan, että $a \equiv b \pmod{m}$. Silloin $m \mid (a-b)$, joten on olemassa kokonaisluku $k$ siten, että $km = a - b$. Tämän perusteella $(-k)m = b - a$, joten $m \mid (b-a)$. Siis $b \equiv a \pmod{m}$.

(3) Oletetaan, että $a \equiv b \pmod{m}$ ja $b \equiv c \pmod{m}$. Silloin $m \mid (a-b)$ ja $m \mid (b-c)$, joten on olemassa sellaiset kokonaisluvut $k$ ja $l$, että $km = a-b$ ja $lm = b-c$. Siis $a-c = (a-b)+(b-c) = km+lm = m(k+l)$, josta seuraa, että $m \mid (a-c)$ ja edelleen $a \equiv c \pmod{m}$.

□

**Lause 1.5.** *Olkoot $a, b$ ja $c$ kokonaislukuja ja olkoon $m$ positiivinen kokonaisluku. Jos $a \equiv b \pmod{m}$, niin*

(1) $a + c \equiv b + c \pmod{m}$,

(2) $a - c \equiv b - c \pmod{m}$ ja

(3) $ac \equiv bc \pmod{m}$.

*Todistus* (vrt. [4, s. 122]). Olkoot $a$, $b$ ja $c$ kokonaislukuja ja olkoon $m$ positiivinen kokonaisluku. Oletetaan, että $a \equiv b \pmod{m}$, joten $m \mid (a-b)$.

(1) Kirjoitetaan $(a-b)$ muodossa $(a+c)-(b+c) = a-b$. Siis $m \mid ((a+c)-(b+c))$, joten $a + c \equiv b + c \pmod{m}$.

(2) Todistetaan vastaavasti kuin kohta (1).

(3) Koska $m \mid (a-b)$, niin $m \mid c(a-b) = ac - bc$. Siis $m \mid ac - bc$, joten $ac \equiv bc \pmod{m}$.

□

**Esimerkki 1.8.** Koska $17 \equiv 3 \pmod{7}$, niin lauseen 1.5 perusteella $19 = 17 + 2 \equiv 3 + 2 = 5 \pmod{7}$, $13 = 17 - 4 \equiv 3 - 4 = -1 \pmod{7}$ ja $34 = 17 \cdot 2 \equiv 3 \cdot 2 = 6 \pmod{7}$.

# 2 Tarkistusnumerot

Koodausteoria on matematiikan ala, joka käsittelee virheiden havaitsemista ja korjaamista erilaisissa koodeissa. Seuraavissa alaluvuissa tarkastellaan, kuinka kongruenssia käytetään virheiden havaitsemisessa ja korjaamisessa tunnistenumeroiden yhteydessä. Kaikkien alalukujen lähdeteoksena on käytetty Thomas Koshyn teosta *Elementary Number Theory with Applications*, jollei toisin mainita.

## 2.1 Binäärikoodit

Tässä alaluvussa tarkastellaan *binäärikoodeja*, eli viestejä, jotka ovat muunnettu biteiksi ja siirretty jotain kanavaa, esimerkiksi puhelinlinjaa, pitkin. Viestin vastaanottaja yrittää löytää viestin alkuperäisen sanoman tulkkaamalla vastaanotettua viestiä. Viestin vastaanottajan täytyy havaita ja korjata kaikki mahdolliset vastaanotetussa viestissä esiintyvät virheet, jotka voivat syntyä esimerkiksi viestikanavassa olevista häiriöistä. Seuraavalla menetelmällä on tärkeä osa binäärikoodeissa esiintyvien virheiden havaitsemisessa ja korjaamisessa.

Ennen viestin lähettämistä liitetään jokaiseen binäärijonoon $x_1x_2 \ldots x_n$ *pariteettitarkistusbitti*, joka määritellään kaavalla

$$x_{n+1} \equiv x_1 + x_2 + \cdots + x_n \pmod{2}.$$

Pariteettitarkistusbitti on siis 1, jos binäärikoodissa on pariton määrä ykkösiä, ja muulloin 0. Tämä menetelmä säilyttää ykkösten lukumäärän binäärikoodissa aina parillisena.

**Esimerkki 2.1.** Tarkastellaan kymmenbittistä merkkijonoa 1011010111. Tällöin $x_{11} \equiv 1 + 0 + 1 + 1 + 0 + 1 + 0 + 1 + 1 + 1 \equiv 1 \pmod{2}$, joten tarkistusbitti on 1 ja lähetettävä viesti on 10110101111.

Tarkastellaan sitten bittijonoa 11010111001. Koska jonossa on pariton määrä ykkösiä, niin tiedetään, että siinä on oltava ainakin yksi virhe. Jos virheellisiä bittejä on täsmälleen yksi ja sen paikka tiedetään, oikea bittijono saadaan vaihtamalla tuossa paikassa oleva bitti.

Tarkistusnumeroita käytetään siis usein virheiden havaitsemiseksi erilaisissa lukusarjoissa. Esimerkiksi pankit, kustantamot, kirjastot ja yritykset, joiden täytyy jäljittää suuria määriä tuotteitaan, käyttävät tarkistusnumeroita omissa tunnistenumeroissaan virheiden löytämistä helpottaakseen. Seuraavat esimerkit havainnollistavat asiaa. Määritellään ensin kuitenkin pistetulon käsite, jota tullaan tarvitsemaan tarkistusnumeroiden laskemisessa.

**Määritelmä 2.1.** Olkoot $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)$ ja $\boldsymbol{y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_n)$ vektoreita. Tällöin vektoreiden $\boldsymbol{x}$ ja $\boldsymbol{y}$ *pistetulo* on

$$\boldsymbol{x} \cdot \boldsymbol{y} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n) \cdot (\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_n) = \sum_{i=1}^{n} x_i y_i.$$

Seuraavassa esimerkessä, jossa lasketaan pankkishekin tarkistusnumero, käytetään pistetuloa.

**Esimerkki 2.2.** Pankkishekissä on 8-numeroinen tunnistenumero $d_1 d_2 \ldots d_8$. Varsinaisen tunnistenumeron perässä on lisäksi tarkistusnumero $d$, joka määritellään kaavalla

$$d \equiv (d_1, d_2, \ldots, d_8) \cdot (7, 3, 9, 7, 3, 9, 7, 3) \pmod{10},$$

missä oikean puolen tulo on vektoreiden pistetulo. Tarkastellaan shekkiä, jonka tunnistenumero on 28643867. Tällöin

$$\begin{aligned} d &\equiv (2, 8, 6, 4, 3, 8, 6, 7) \cdot (7, 3, 9, 7, 3, 9, 7, 3) \pmod{10} \\ &\equiv 2 \cdot 7 + 8 \cdot 3 + 6 \cdot 9 + 4 \cdot 7 + 3 \cdot 3 + 8 \cdot 9 + 6 \cdot 7 + 7 \cdot 3 \pmod{10} \\ &\equiv 264 \equiv 4 \pmod{10}, \end{aligned}$$

joten shekin 9-numeroinen tunnistenumero on 286438674.

## 2.2 UPC-koodit

Tämä alaluku käsittelee kauppojen tuotteista löytyviä UPC-tuotekoodeja. Koshyn kirjan lisäksi lähteenä on käytetty Kirtlandin kirjaa *Identification Numbers and Check Digit Schemes*.

Kaikkien kauppojen kaikista tuotteista pitäisi löytyä UPC-koodi (Universal Product Code). UPC-koodeja käytetään vaihtelevilla tavoilla, sillä ne eivät pelkästään yksilöi tuotteita, vaan myös kertovat minkälaisesta tuotteesta on kyse ja kuka sen on valmistanut. Käyttämällä UPC-koodia jokainen kauppa määrää tuotteilleen hinnat, jotka näkyvät asiakkaille kun tuotteiden viivakoodit luetaan kassalla. Hinnoittelun lisäksi liikkeet käyttävät UPC-koodeja inventoinnissa, sillä niiden avulla voidaan helposti seurata eri tuotteiden menekkiä ja vaihtuvuutta.

UPC-järjestelmä kehitettiin Yhdysvalloissa vuonna 1973 standardiksi tuotteiden hinnoittelulle. Kolme vuotta myöhemmin kehitettiin EAN-koodi

(European Article Numbering), jota käytetään nykyisin useimmissa maissa, myös Suomessa. UPC-koodista on olemassa viisi eri versiota ja EAN-koodistakin kaksi. Eniten käytetty versio, etenkin Yhdysvalloissa, on 12-numeroinen UPC-A -koodi, josta on esimerkki kuvassa 1.


**Kuva 1.** Esimerkki UPC-A -viivakoodista.

UPC-A -koodi koostuu siis kahdestatoista numerosta $d_1, d_2, \ldots, d_{12}$, joista ensimmäinen numero $d_1$ kertoo tuotteen tyypin. Taulukosta 1 löytyy numeron $d_1$ eri arvot ja niitä vastaavat tuoteryhmät. Numerot $d_2d_3d_4d_5d_6$ puolestaan kertovat tuotteen valmistajan ja numerot $d_7d_8d_9d_{10}d_{11}$ itse tuotteen. Viimeinen numero $d_{12}$ on tarkistusnumero. Tarkistusnumeroa ei ole aina painettu tuotteisiin, mutta se löytyy kuitenkin aina viivakoodista. Yritysten kasvaessa ja tuotteiden lisääntyessä, osa yrityksistä on lyhentänyt oman valmistajatunnuksensa nelinumeroiseksi $d_2d_3d_4d_5$ ja pidentänyt tuotenumeronsa kuusinumeroisiksi $d_6d_7d_8d_9d_{10}d_{11}$. Käytännöstä riippumatta, tarkistusnumero määräytyy kuitenkin aina samalla tavalla.

**Taulukko 1.** UPC-A -koodin tuoteryhmät.

| $d_1$ | Tuoteryhmä |
|---:|---|
| 0 | Yleiset ruokatavarat |
| 2 | Lihatuotteet |
| 3 | Lääkkeet ja terveystuotteet |
| 4 | Ei-ruokatavarat |
| 5 | Kupongit |
| 6, 7 | Muut tuotteet |
| 1, 8, 9 | Varattu tulevalle käytölle |

Esimerkiksi Yhdysvalloissa Kraft General Foods Inc.:n valmistaman pikakahvin UPC-koodi on 04300079470. Tuoteryhmän, valmistajan ja itse tuotteen koodit ovat siis 0, 43000 ja 79470, toisin sanoen

| **0** | **43000** | **79470** | **-** |
|---|---|---|---|
| ↑ | ↑ | ↑ | ↑ |
| tuoteryhmä | valmistaja | tuote | tarkistusnumero. |

UPC-koodin tarkistusnumeron $d_{12}$ tulee täyttää seuraava ehto:

$$\begin{aligned}
& (d_1, d_2, \ldots, d_{12}) \cdot (3,1,3,1,3,1,3,1,3,1,3,1) \equiv 0 \pmod{10} \\
\Leftrightarrow\ & (d_1, d_2, \ldots, d_{11}) \cdot (3,1,3,1,3,1,3,1,3,1,3) + d_{12} \equiv 0 \pmod{10} \\
\Leftrightarrow\ & d_{12} \equiv -(d_1, d_2, \ldots, d_{11}) \cdot (3,1,3,1,3,1,3,1,3,1,3) \pmod{10}.
\end{aligned}$$

Seuraava esimerkki havainnollistaa UPC-koodin tarkistusnumeron laskemista.

**Esimerkki 2.3.** Lasketaan tarkistusnumero $d_{12}$ tuotteelle, jonka UPC-koodi on 04750056940. Nyt

$$\begin{aligned}
d_{12} &\equiv -(d_1, d_2, \ldots, d_{11}) \cdot (3,1,3,1,3,1,3,1,3,1,3) \pmod{10} \\
&\equiv -(0,4,7,5,0,0,5,6,9,4,0) \cdot (3,1,3,1,3,1,3,1,3,1,3) \pmod{10} \\
&\equiv -(0+4+21+5+0+0+15+6+27+4+0) \pmod{10} \\
&\equiv -82 \equiv 8 \pmod{10}.
\end{aligned}$$

Siispä tarkistusnumero on 8 ja UPC-koodi on 0-47500-56940-8.

UPC-koodin tarkistusnumerojärjestelmä on parempi kuin esimerkiksi moduli 7 - tai moduli 9 -järjestelmät. Nuo järjestelmät eivät paljasta kaikkia yhden numeron virheitä, toisin kuin moduli 10 -pohjainen UPC-järjestelmä paljastaa. Tarkastellaan UPC-koodia 5-02004-81693-4 ja seuraavanlaista yhden numeron virhettä.

| | | |
|---|---|---|
| Todellinen numero: | 5-02004-81693-**4** | 5-0200**4**-81693-4 |
| | ↓ | ↓ |
| Virheellinen numero: | 5-02004-81693-**9** | 5-0200**7**-81693-4 |

Molemmat virheet voidaan huomata. Yhden numeron virhe, jossa tarkistusnumero vaihtuu vääräksi, johtaa seuraavaan virheelliseen tulokseen:

$$\begin{aligned}
& (5,0,2,0,0,4,8,1,6,9,3,9) \cdot (3,1,3,1,3,1,3,1,3,1,3,1) \equiv 0 \pmod{10} \\
& 5 \cdot 3 + 0 \cdot 1 + 2 \cdot 3 + 0 \cdot 1 + 0 \cdot 3 + 4 \cdot 1 + 8 \cdot 3 + 1 \cdot 1 \\
& \qquad +6 \cdot 3 + 9 \cdot 1 + 3 \cdot 3 + 9 \cdot 1 \equiv 0 \pmod{10} \\
& 15 + 0 + 6 + 0 + 0 + 4 + 24 + 1 + 18 + 9 + 9 + 9 \equiv 0 \pmod{10} \\
& \qquad 95 \equiv 0 \pmod{10}.
\end{aligned}$$

Huomataan kuitenkin, että $95 \equiv 5 \pmod{10}$, joten koodissa on virhe. Vastaavasti, jos koodin kuudes numero vaihtuu vääräksi, saadaan jälleen virheellinen lauseke:

$$\begin{aligned}
& (5,0,2,0,0,7,8,1,6,9,3,9) \cdot (3,1,3,1,3,1,3,1,3,1,3,1) \equiv 0 \pmod{10} \\
& 5 \cdot 3 + 0 \cdot 1 + 2 \cdot 3 + 0 \cdot 1 + 0 \cdot 3 + 7 \cdot 1 + 8 \cdot 3 + 1 \cdot 1 \\
& \qquad +6 \cdot 3 + 9 \cdot 1 + 3 \cdot 3 + 9 \cdot 1 \equiv 0 \pmod{10} \\
& 15 + 0 + 6 + 0 + 0 + 7 + 24 + 1 + 18 + 9 + 9 + 9 \equiv 0 \pmod{10} \\
& \qquad 98 \equiv 0 \pmod{10}.
\end{aligned}$$

Nyt huomataan, että $98 \equiv 8 \pmod{10}$, joten koodi on virheellinen.

Osoitetaan seuraavaksi, että UPC-järjestelmä paljastaa kaikki yhden numeron virheet. Olkoon $a_1 \dots a_i \dots a_{12}$ UPC-koodi, jossa $1 \leq i \leq 12$. Kun esiintyy yhden numeron virhe, $a_1 \dots a_i \dots a_{12}$ tulee muotoon $a_1 \dots b_i \dots a_{12}$, missä $a_i \neq b_i$, eli luku $a_i$ korvautuu luvulla $b_i$.

Oletetaan, että tuota yhden numeron virhettä ei huomata. Tällöin sekä

$$(3, 1, \dots, 3, 1) \cdot (a_1, \dots, a_i, \dots, a_{12}) \equiv 0 \pmod{10}$$

että

$$(3, 1, \dots, 3, 1) \cdot (a_1, \dots, b_i, \dots, a_{12}) \equiv 0 \pmod{10}.$$

Kirjoitetaan kongruenssit toisessa muodossa, jolloin saadaan

$$\begin{aligned}(3, 1, \dots, 3, 1) \cdot (a_1, \dots, a_i, \dots, a_{12}) - (3, 1, \dots, 3, 1) \cdot (a_1, \dots, b_i, \dots, a_{12}) \\ \equiv 0 \pmod{10}.\end{aligned}$$

Nyt on kaksi mahdollisuutta, jotka voidaan ottaa huomioon. Laskettaessa pistetuloja $(3, 1, \dots, 3, 1) \cdot (a_1, \dots, a_i, \dots, a_{12})$ ja $(3, 1, \dots, 3, 1) \cdot (a_1, \dots, b_i, \dots, a_{12})$, sekä $a_i$ että $b_i$ kerrotaan joko luvulla 3 tai luvulla 1. Tarkastellaan seuraavaksi kumpaakin tapausta erikseen.

- **Tapaus I.** Tässä tapauksessa, sekä $a_i$ että $b_i$ kerrotaan luvulla 3. Tämä johtaa seuraavaan tulokseen:

  $$\begin{aligned}0 &\equiv (3, 1, \dots, 3, 1) \cdot (a_1, \dots, a_i, \dots, a_{12}) \\ &\quad - (3, 1, \dots, 3, 1) \cdot (a_1, \dots, b_i, \dots, a_{12}) \pmod{10} \\ &\equiv (3 \cdot a_1 + 1 \cdot a_2 + \cdots + 3 \cdot a_i + \cdots + 3 \cdot a_{11} + 1 \cdot a_{12}) \\ &\quad - (3 \cdot a_1 + 1 \cdot a_2 + \cdots + 3 \cdot b_i + \cdots + 3 \cdot a_{11} + 1 \cdot a_{12}) \pmod{10} \\ &\equiv 3 \cdot a_1 + 1 \cdot a_2 + \cdots + 3 \cdot a_i + \cdots + 3 \cdot a_{11} + 1 \cdot a_{12} \\ &\quad - 3 \cdot a_1 - 1 \cdot a_2 - \cdots - 3 \cdot b_i - \cdots - 3 \cdot a_{11} - 1 \cdot a_{12} \pmod{10} \\ &\equiv 3 \cdot a_i - 3 \cdot b_i \pmod{10} \\ &\equiv 3 \cdot (a_i - b_i) \pmod{10}.\end{aligned}$$

  Siis $3 \cdot (a_i - b_i) \equiv 0 \pmod{10}$. Koska $a_i \neq b_i$, niin $a_i - b_i \neq 0$. Lisäksi, koska $(3, 10) = 1$, niin $3 \cdot (a_i - b_i)$ ei voi koskaan olla luvun 10 monikerta. Tämä johtaa kuitenkin ristiriitaan. Siis oletus, että virhettä ei huomata, on väärä. Ollaan siis osoitettu, että yhden numeron virhe UPC-koodissa voidaan aina havaita.

- **Tapaus II.** Tässä tapauksessa, sekä $a_i$ että $b_i$ kerrotaan luvulla 1. Tämä johtaa seuraavaan tulokseen:

$$\begin{aligned}
0 &\equiv (3,1,\ldots,3,1)\cdot(a_1,\ldots,a_i,\ldots,a_{12}) \\
&\quad -(3,1,\ldots,3,1)\cdot(a_1,\ldots,b_i,\ldots,a_{12}) \pmod{10} \\
&\equiv (3\cdot a_1+1\cdot a_2+\cdots+1\cdot a_i+\cdots+3\cdot a_{11}+1\cdot a_{12}) \\
&\quad -(3\cdot a_1+1\cdot a_2+\cdots+1\cdot b_i+\cdots+3\cdot a_{11}+1\cdot a_{12}) \pmod{10} \\
&\equiv 3\cdot a_1+1\cdot a_2+\cdots+1\cdot a_i+\cdots+3\cdot a_{11}+1\cdot a_{12} \\
&\quad -3\cdot a_1-1\cdot a_2-\cdots-1\cdot b_i-\cdots-3\cdot a_{11}-1\cdot a_{12} \pmod{10} \\
&\equiv 1\cdot a_i-1\cdot b_i \pmod{10} \\
&\equiv a_i-b_i \pmod{10}.
\end{aligned}$$

  Siis $a_i - b_i \equiv 0 \pmod{10}$. Koska $a_i \neq b_i$, niin $a_i - b_i \neq 0$. Lisäksi, koska sekä $a_i$ että $b_i$ ovat lukuja lukujen 0 ja 9 väliltä, niin $a_i - b_i$ ei voi koskaan olla luvun 10 monikerta. Tämä johtaa kuitenkin ristiriitaan. Siis oletus, että virhettä ei huomata, on väärä. Ollaan siis osoitettu, että yhden numeron virhe UPC-koodissa voidaan aina havaita.

UPC-järjestelmä paljastaa siis kaikki yhden numeron virheet, mutta sekään ei paljasta kaikkia vierekkäisten numeroiden paikanvaihtovirheitä. Jos $a$ ja $b$ ovat UPC-koodin kaksi vierekkäistä numeroa, niin paikanvaihtovirhe $\cdots ab\cdots \to \cdots ba\cdots$ ei paljastu, kun $|a-b| = 5$. Tarkastellaan kahta tämäntyyppistä virhetilannetta.

| | | |
|---|---|---|
| Todellinen numero: | 5-02004-**81**693-4 | **5-0**2004-81693-4 |
| | ↓ | ↓ |
| Virheellinen numero: | 5-02004-**18**693-4 | **0-5**2004-81693-4 |

Ensimmäinen virhe, jossa $a = 8$ ja $b = 1$, voidaan havaita, sillä $|a-b| = |8-1| = |7| \neq 5$. Sen sijaan toista virhettä, jossa $a = 5$ ja $b = 0$, ei voida havaita, sillä $|a-b| = |5-0| = |5| = 5$.

## 2.3 ISBN-numerot

Tässä alaluvussa käsitellään kaikista kirjoista löytyviä ISBN-numeroita ja niiden ominaisuuksia. Lähdeteoksena on käytetty Koshyn kirjan lisäksi Rosenin teosta *Elementary Number Theory and its Applications*.

Vuodesta 1972 lähtien käytännössä jokaisella, missäpäin maailmaa tahansa, julkaistulla kirjalla on ISBN-numero (International Standard Book Number), joka on 10-numeroinen tunnistenumero. ISBN-järjestelmä mahdollistaa kirjatietojen tietokoneistetun siirtämisen ja varastoimisen. Järjestelmä sai alkunsa Iso-Britanniassa vuonna 1967 ja seuraavana vuonna se esiteltiin Yhdysvalloissa.

ISBN-numero koostuu neljästä osasta: yksinumeroisesta ryhmäkoodista, kaksinumeroisesta kustantajan koodista, kuusinumeroisesta kirjan koodista ja tarkistusnumerosta. Esimerkiksi ISBN-numeron 0-07-035471-5 luku 0 tarkoittaa, että kirja on julkaistu englanninkielisessä maassa, 07 kertoo kustantajan, 035471 on kustantajan kirjalle antama tunnistenumero ja viimeinen luku 5 on tarkistusnumero.

Olkoon nyt ISBN-numero muotoa $x_1x_2\ldots x_{10}$, missä $x_{10}$ on tarkistusnumero. Tällöin tarkistusnumero $x_{10}$, missä $0 \leq x_{10} \leq 10$ ja lukua 10 merkitään symbolilla $X$, on valittava siten, että

$$\begin{aligned}
& \sum_{i=1}^{10} ix_i \equiv 0 \pmod{11} \\
\Leftrightarrow \quad & \sum_{i=1}^{9} ix_i \equiv -10x_{10} \pmod{11} \\
& = -10x_{10} + 11k \\
& = x_{10} - 11x_{10} + 11k \\
& = x_{10} + 11(k - x_{10}) \\
& \equiv x_{10} \pmod{11} \\
\Leftrightarrow \quad & x_{10} \equiv \sum_{i=1}^{9} ix_i \pmod{11}.
\end{aligned}$$

Seuraava esimerkki havainnollistaa tätä koodausjärjestelmää.

**Esimerkki 2.4.** Lasketaan tarkistusnumero $x_{10}$ ISBN-koodille, jonka yhdeksän ensimmäistä numeroa ovat 0-07-048236. Tällöin

$$x_{10} \equiv 1\cdot 0 + 2\cdot 0 + 3\cdot 7 + 4\cdot 0 + 5\cdot 4 + 6\cdot 8 + 7\cdot 2 + 8\cdot 3 + 9\cdot 6 \equiv 5 \pmod{11}.$$

Siispä tarkistusnumero on 5 ja ISBN-koodi kokonaisuudessaan on 0-07-048236-5.

Osoitetaan seuraavaksi, että yksittäisen numeron virhe tai kahden numeron paikanvaihtovirhe voidaan havaita käyttämällä ISBN-numeron tarkistusnumeroa. Oletetaan ensin, että $x_1x_2\ldots x_{10}$ on aito ISBN-koodi, mutta tuo numero painetussa muodossa on $y_1y_2\ldots y_{10}$. Tiedetään, että

$$\sum_{i=1}^{10} ix_i \equiv 0 \pmod{11},$$

sillä $x_1x_2\ldots x_{10}$ on aito ISBN-koodi.

Oletetaan sitten, että ISBN-koodin painamisessa on tapahtunut täsmälleen yksi virhe. Tällöin jollain kokonaisluvulla $j$ pätee $y_i = x_i$, kun $i \neq j$ ja

$y_j = x_j + a$, missä $-10 \leq a \leq 10$ ja $a \neq 0$. Nyt siis $a = y_j - x_j$ on virhe $j$. numeron kohdalla. Huomataan, että

$$\sum_{i=1}^{10} iy_i = \sum_{i=1}^{10} ix_i + ja \equiv ja \not\equiv 0 \pmod{11},$$

sillä $\sum_{i=1}^{10} ix_i \equiv 0 \pmod{11}$ ja $11 \nmid ja$, koska $11 \nmid j$ ja $11 \nmid a$. Näin ollaan onnistuttu havaitsemaan virhe ja osoitettu, että $y_1y_2 \ldots y_{10}$ ei ole aito ISBN-koodi.

Oletetaan seuraavaksi, että ISBN-koodin jotkin kaksi erisuuruista numeroa ovat vaihtaneet paikkaa. Tällöin on olemassa erilliset kokonaisluvut $j$ ja $k$ siten, että $y_j = x_k$ ja $y_k = x_j$ sekä $y_i = x_i$, jos $i \neq j$ ja $i \neq k$. Tästä seuraa, että

$$\sum_{i=1}^{10} iy_i = \sum_{i=1}^{10} ix_i + (jx_k - jx_j) + (kx_j - kx_k) \equiv (j-k)(x_k - x_j) \not\equiv 0 \pmod{11},$$

sillä $\sum_{i=1}^{10} ix_i \equiv 0 \pmod{11}$ ja $11 \nmid (j-k)$ ja $11 \nmid (x_k - x_j)$. Nähdään siis, että $y_1y_2 \ldots y_{10}$ ei ole aito ISBN-koodi, joten ollaan osoitettu, että kahden numeron paikanvaihtovirhe voidaan havaita.

## 2.4 Postinumerot

Yhdysvaltojen postilaitos (The United States Postal Service) käyttää viivakoodeja postinumeroiden koodaamiseen postilähetyksissä. Viivakoodit ovat helposti ja nopeasti luettavissa edullisilla viivakoodien lajittelukoneilla.
POSTNET-viivakoodi (POSTal Numeric Encoding Technique) voi tarkoittaa joko 5-numeroista ja 32-viivaista postinumerokoodia, 9-numeroista ja 52-viivaista postinumeron ja neljän muun numeron muodostamaa koodia tai 11-numeroista ja 62-viivaista jakelupistekoodia. Koodeissa käytetään sekä binäärilukuja 0 ja 1 että tarkistusnumeroita. Viivoja on kahden mittaisia, pitkiä ja lyhyitä, kuten kuva 2 osoittaa.

**Kuva 2.** Esimerkki POSTNET-viivakoodista.

Pitkä eli kokonainen viiva vastaa numeroa 1 ja lyhyt eli puolikas viiva vastaa numeroa 0. Kaksi äärimmäistä viivaa ovat aina pitkiä ja ne voidaan jättää huomiotta. Loput viivat ovat jaettu viiden pylvään ryhmiin, joista viimeinen ryhmä kuvaa tarkistusnumeroa. Järjestelmä, jolla desimaaliluvut muunnetaan binäärisiksi, perustuu Yhdysvalloissa 1940-luvulla käytettyyn koodausjärjestelmään.

**Taulukko 2.** Lukuja 0-9 vastaavat koodit.

| Luku-arvo | Viivojen paikkojen painot: Binäärinen 74210 | Viivakoodi 74210 |
|---|---|---|
| 1 | 00011 |  |
| 2 | 00101 |  |
| 3 | 00110 |  |
| 4 | 01001 |  |
| 5 | 01010 |  |
| 6 | 01100 |  |
| 7 | 10001 |  |
| 8 | 10010 |  |
| 9 | 10100 |  |
| 0 | 11000 |  |

On olemassa $\frac{5!}{2!3!} = \frac{5\cdot4\cdot3\cdot2\cdot1}{2\cdot1\cdot3\cdot2\cdot1} = \frac{20}{2} = 10$ järjestelyä kahdesta pitkästä ja kolmesta lyhyestä viivasta. Kukin järjestely vastaa yhtä numeroa, kuten yllä oleva taulukko 2 osoittaa.

Lukuun ottamatta numeroa 0, viiden viivan yhdistelmää vastaava lukuarvo saadaan laskemalla kahden pitkän viivan painot yhteen. Vasemmalta oikealle, viivojen paikkoja vastaavat painot ovat 7, 4, 2, 1 ja 0. Ainoan poikkeuksen tälle säännölle tekee yhdistelmä  , jonka kokonaispaino olisi 11, mutta sen arvoksi on määrätty 0.

Tarkastellaan sitten 5-numeroista postinumeroa $z_1 z_2 \dots z_5$. Sen tarkistusnumero $d$ lasketaan kaavalla

$$d \equiv -\sum_{i=1}^{5} z_i \pmod{10}.$$

Esimerkiksi postinumeron 13264 tarkistusnumero on

$$\begin{aligned} d &\equiv -(1+3+2+6+4) \pmod{10} \\ &\equiv -6 \equiv 4 \pmod{10}, \end{aligned}$$

joten $d = 4$.

Tarkistusnumero lasketaan samalla tavalla myös 9-numeroisille koodeille.

Siis esimerkiksi koodin 12637-4859 tarkistusnumero on

$$\begin{aligned} d &\equiv -(1+2+6+3+7+4+8+5+9) \pmod{10} \\ &\equiv -5 \equiv 5 \pmod{10}, \end{aligned}$$

joten $d = 5$.

Jakelupisteviivakoodi (Delivery Point Barcode) kehitettiin Yhdysvaltojen postilaitoksen toimesta yli 115 miljoonan jakelupisteen yksikäsitteiseksi tunnistamiseksi. Järjestelmä helpottaa postinlajittelijoiden työtä huomattavasti ennen postinjakelua. Jakelupisteviivakoodi muodostetaan lisäämällä olemassaolevaan 9-numeroiseen koodiin 10 pylvästä. Nuo pylväät vastaavat kahta uutta numeroa, jotka normaalisti kuvaavat katuosoitteen kahta viimeistä numeroa. Esimerkiksi 12345-6789-014 on jakelupistekoodi, jossa 01 vastaa jakelupistettä ja 4 on tarkistusnumero.

## 2.5 Ajokortti- ja passinumerot

Tässä alaluvussa esitellään eri puolilla maailmaa käytössä olevia passi- ja ajokorttinumerojärjestelmiä. Lähdeteoksena Koshyn teoksen lisäksi on käytetty Rosenin teosta *Elementary Number Theory and its Applications*.

Yhdysvalloissa ajokorttien tunnistenumeroiden määrittämiseen käytettävät menetelmät vaihtelevat suuresti osavaltioiden välillä. Jotkut osavaltiot käyttävät tarkistusnumeroita ajokorttinumeroissaan havaitakseen virheitä ja ajokorttien väärennöksiä.

Esimerkiksi Utahin osavaltiossa on käytössä 8-numeroinen ajokorttinumero $d_1 d_2 \dots d_8$, johon on liitetty tarkistusnumero $d_9$, joka määritellään kaavalla

$$\begin{aligned} d_9 &\equiv \sum_{1}^{8} (10-i) d_i \pmod{10} \\ &\equiv (9,8,7,6,5,4,3,2) \cdot (d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8) \pmod{10}. \end{aligned}$$

Samaa menetelmää käytetään Yhdysvalloissa myös kemikaalien rekisteröimisessä sekä lähes identtistä menetelmää Kanadan Newfoundlandin provinssissa ajokorttinumeroissa. Seuraava esimerkki valaisee edellä kuvattua koodausjärjestelmää.

**Esimerkki 2.5.** Lasketaan tarkistusnumero $d_9$ Utahin osavaltion ajokortille, jonka tunnistenumero on 23847095. Tällöin

$$\begin{aligned} d_9 &\equiv (9,8,7,6,5,4,3,2) \cdot (2,3,8,4,7,0,9,5) \pmod{10} \\ &\equiv (18+24+56+24+35+0+27+10) \pmod{10} \\ &\equiv 4 \pmod{10}. \end{aligned}$$

Siis ajokortin tunnistenumero kokonaisuudessaan on 238470954.

Joissain Yhdysvaltojen osavaltioissa on käytössä jopa monimutkaisempia koodausjärjestelmiä ajokorttien tunnistenumeroiden muodostamiseksi. Arkansas, New Mexico ja Tennessee liittävät tarkistusnumeron $d_8$ 7-numeroiseen ajokorttinumeroon seuravaaksi kuvatulla tavalla. Olkoon

$$x \equiv -(d_1, d_2, \ldots, d_7) \cdot (2, 7, 6, 5, 4, 3, 2) \pmod{11}.$$

Tällöin

$$d_8 = \begin{cases} 1, & \text{jos } x = 0 \\ 0, & \text{jos } x = 10 \\ x & \text{muulloin.} \end{cases}$$

Vermontin osavaltio käyttää muilta osin samaa järjestelmää, paitsi kun $x = 0$, tarkistusnumeroa merkitään symbolilla $A$. Seuraava esimerkki auttaa asian havainnollistamisessa.

**Esimerkki 2.6.** Määritetään tarkistusnumero $d_8$ New Mexicon osavaltion myöntämälle ajokortille, jonka 7-numeroinen tunnistenumero on 0438629. Lasketaan ensin $x$:

$$\begin{aligned} x &\equiv -(0, 4, 3, 8, 6, 2, 9) \cdot (2, 7, 6, 5, 4, 3, 2) \pmod{11} \\ &\equiv -(0 + 28 + 18 + 40 + 24 + 6 + 18) \pmod{11} \\ &\equiv -134 \equiv 9 \pmod{11}. \end{aligned}$$

Siis, määritelmän mukaan, $d_8 = 9$ ja ajokorttinumero kokonaisuudessaan on 04386299.

Maailmalla on käytössä myös hieman eksoottisempia järjestelmiä tunnistenumeroiden laskemiseksi. Esimerkiksi Norjassa käytetään kahta tarkistusnumeroa kansalaisten rekisteröintinumeroiden muodostamisessa. Rekisteröintinumero on 11-numeroinen ja sen kaksi viimeistä numeroa ovat tarkistusnumeroita, jotka määritellään seuraavalla tavalla:

$$\begin{aligned} d_{10} &\equiv -(d_1, d_2, \ldots, d_9) \cdot (3, 7, 6, 1, 8, 9, 4, 5, 2) \pmod{11} \\ d_{11} &\equiv -(d_1, d_2, \ldots, d_{10}) \cdot (5, 4, 3, 2, 7, 6, 5, 4, 3, 2) \pmod{11}. \end{aligned}$$

Lisäksi rekisteröintinumerot, joille joko $d_{10} = 10$ tai $d_{11} = 10$, jätetään käyttämättä. Seuraava esimerkki havainnollistaa tätä menetelmää.

**Esimerkki 2.7.** Erään Norjan kansalaisen rekisteröintinumeron yhdeksän ensimmäistä numeroa ovat 054354427. Lasketaan tunnistenumeron kaksi tarkistusnumeroa $d_{10}$ ja $d_{11}$. Nyt

$$\begin{aligned} d_{10} &\equiv -(d_1, d_2, \ldots, d_9) \cdot (3, 7, 6, 1, 8, 9, 4, 5, 2) \pmod{11} \\ &\equiv -(0, 5, 4, 3, 5, 4, 4, 2, 7) \cdot (3, 7, 6, 1, 8, 9, 4, 5, 2) \pmod{11} \\ &\equiv -(0 + 35 + 24 + 3 + 40 + 36 + 16 + 10 + 14) \pmod{11} \\ &\equiv -178 \equiv 9 \pmod{11} \end{aligned}$$

ja

$$\begin{aligned} d_{11} &\equiv -(d_1, d_2, \ldots, d_{10}) \cdot (5,4,3,2,7,6,5,4,3,2) \pmod{11} \\ &\equiv -(0,5,4,3,5,4,4,2,7,9) \cdot (5,4,3,2,7,6,5,4,3,2) \pmod{11} \\ &\equiv -(0+20+12+6+35+24+20+8+21+18) \pmod{11} \\ &\equiv -164 \equiv 1 \pmod{11}. \end{aligned}$$

Siispä tarkistusnumerot ovat 9 ja 1 ja näin ollen rekisteröintinumero on 05435442791.

Tarkistusnumeroita käytetään useissa maissa myös passinumeroiden virheiden tunnistamisessa. Esimerkiksi useissa Euroopan maissa käytettävässä järjestelmässä, jos $x_1x_2x_3x_4x_5x_6x_7$ on passin tunnistenumero, niin tarkistusnumero $x_7$ lasketaan kaavalla

$$x_7 \equiv 7x_1 + 3x_2 + x_3 + 7x_4 + 3x_5 + x_6 \pmod{10}.$$

**Esimerkki 2.8.** Olkoon passin tunnistenumero 213968. Tällöin

$$x_7 \equiv 7 \cdot 2 + 3 \cdot 1 + 1 \cdot 3 + 7 \cdot 9 + 3 \cdot 6 + 1 \cdot 8 \equiv 9 \pmod{10},$$

joten tarkistusnumero on 9 ja passin tunnistenumero kokonaisuudessaan 2139689.

Yksittäinen virhe passin tunnistenumerossa on mahdollista havaita, mikäli tarkistusnumero lasketaan edellä mainitulla tavalla. Tämän huomaamiseksi oletetaan, että tehdään virhe, jossa jokin numeroista vaihtuu vääräksi, toisin sanoen $y_j = x_j + a \pmod{10}$, missä $x_j$ on oikea $j$. numero ja $y_j$ on sen korvannut väärä numero. Tarkistusnumeron määritelmästä seuraa, että luvun $x_7$ muutos on joko $7a$, $3a$ tai $a \pmod{10}$. Sen sijaan virheet, joissa kaksi numeroa vaihtaa paikkaa, voidaan havaita, jos ja vain jos näiden kahden luvun erotus ei ole 5 tai $-5$, eli jos ne eivät ole lukuja $x_i$ ja $x_j$, joille pätee $|x_i - x_j| = 5$. Tämä tarkistusnumerojärjestelmä tunnistaa lisäksi suuren joukon virheitä, joissa sekoitetaan kolme numeroa.

## 2.6 Henkilötunnukset

Tämä alaluku käsittelee kaikille suomalaisille läheistä aihetta, henkilötunnusta, ja lähteenä on käytetty Suomen Väestörekisterikeskuksen internetsivuja [5].

Henkilötunnus on yksilöimiskeino, joka yksilöi kansalaiset nimeä tarkemmin. Täysin samannimisiä ihmisiä on olemassa, mutta ei kahta henkilöä, joilla olisi sama henkilötunnus. Suomessa henkilötunnus on otettu käyttöön 1960-luvulla ja sen saa jokainen Suomen kansalainen syntymätodistuksen perusteella.

Suomen henkilötunnukset ovat muotoa 051256-312N. Kuusi ensimmäistä numeroa kertovat syntymäajan, eli päivän, kuukauden ja vuoden. Syntymäajan jälkeen oleva merkki kertoo syntymävuosisadan; 1800-luvulla syntyneillä se on plusmerkki (+), 1900-luvulla syntyneillä yhdysmerkki (-) ja 2000-luvulla syntyneillä A-kirjain. Yksilönumerolla, joka esimerkkitunnuksessa on 312, erotetaan samana päivänä syntyneet henkilöt toisistaan. Yksilönumeroksi annetaan jokin luku väliltä 002-899. Naisilla se on parillinen ja miehillä pariton.

Henkilötunnuksen viimeinen merkki on tarkistusmerkki, joka voi olla joko numero tai kirjain. Tarkistusmerkki $z$ määritetään seuraavaksi kuvatulla tavalla. Olkoon kokonaisluku

$$x \equiv d_1d_2d_3d_4d_5d_6d_7d_8d_9 \pmod{31},$$

missä $d_1d_2 \dots d_6$ on syntymäaika ja $d_7d_8d_9$ on yksilönumero. Tällöin henkilötunnuksen tarkistusmerkki $z$ on luku $x$, mikäli $0 < x < 10$. Muussa tapauksessa tarkistusmerkki määräytyy luvun $x$ perusteella taulukon 3 mukaisesti.

**Taulukko 3.** Henkilötunnuksen tarkistusmerkit.

| | | | | | | |
|---|---|---|---|---|---|---|
| 10 A | 11 B | 12 C | 13 D | 14 E | 15 F | 16 H |
| 17 J | 18 K | 19 L | 20 M | 21 N | 22 P | 23 R |
| 24 S | 25 T | 26 U | 27 V | 28 W | 29 X | 30 Y |

Eli mikäli luku $x$ saa esimerkiksi arvon 18, niin tarkistusmerkiksi tulee kirjain $J$ ja $x$:n arvolla 5 tarkistusmerkiksi tulee 5.

**Esimerkki 2.9.** Olkoon erään suomalaisen syntymäaika 18.8.2007 ja henkilötunnuksen yksilönumero 739. Tällöin henkilötunnuksen tarkistusmerkiksi saadaan

$$\begin{aligned} z &\equiv 180807739 \pmod{31} \\ &\equiv 22 \pmod{31}. \end{aligned}$$

Siispä taulukon 2 perusteella henkilötunnuksen tarkistusmerkki on P ja henkilötunnus kokonaisuudessaan on 180807A739P.

## 2.7 Virheet

Tässä alaluvussa esitellään, millaisia virheitä tunnistenumeroissa voi esiintyä, sekä käydään läpi miten virheet voidaan huomata. Lähdeteoksina on käytetty Kirtlandin kirjaa *Identification Numbers and Check Digit Schemes* sekä Gallianin artikkelia *The Mathematics of Identification Numbers*.

Erilaisia tunnistenumeroita luetellaan päivittäin puhelimen välityksellä, kirjataan dokumentteihin, kirjoitetaan ja skannataan tietokoneille, lähetetään internetin välityksellä tai siirretään jollain muulla tavalla. Aina kun

näin tapahtuu, on mahdollista, että tunnistenumeron yksi tai useampi numero muuttuu, tai numeroiden järjestys muuttuu niiden siirtyessä paikasta toiseen. Esimerkiksi jonkin tuotteen UPC-koodi 0-43000-79470-8 voisi muuttua vaikkapa tietokoneelle kirjoitettaessa muotoon 0-43000-79472-8 tai 0-43000-97470-8. Kummassakin tapauksessa tunnistenumerossa esiintyy virhe. Edellä mainittujen virheiden lisäksi varsinkin rikolliset yrittävät muuttaa esimerkiksi luottokorttien tunnistenumeroita taloudellisen hyödyn toiveessa.

On tärkeää varmistaa, että tunnistenumerot siirtyvät oikealla tavalla eikä muutoksia tapahdu. Jos esimerkiksi jonkin myytävänä olevan tuotteen UPC-koodi tallennetaan kassajärjestelmään väärässä muodossa, voidaan sen seurauksena asiakkaalta laskuttaa vahingossa väärä hinta, joko liian pieni tai liian suuri. Lisäksi varsinkin pankkiliikenteessä on erittäin tärkeää, että sähköisesti tapahtuvissa tilisiirroissa rahat liikkuvat oikeiden tilien välillä.

Virheitä tunnistenumeroiden siirrossa tapahtuu, kun viivakoodeja skannataan väärin, numeroita kirjoitetaan tai näppäillään väärässä järjestyksessä tai tapahtuu muunlaisia erehdyksiä. Seuraavaksi luetellaan tyypillisimmin esiintyviä virheitä ja niiden esiintymistodennäköisyyksiä.

- *Yhden numeron virhe* (A single-digit error) esiintyy, kun tunnisteen yksi numero vaihtuu joksikin toiseksi. Kaikista virheistä 79.1% on tällaisia virheitä.

- *Vierusnumeroiden paikanvaihto -virhe* (A transposition-of-adjacent-digits error) esiintyy, kun kaksi eri vierekkäistä numeroa vaihtavat paikkaa. 10.2% kaikista virheistä on tätä tyyppiä.

- *Hyppäys-paikanvaihto -virhe* (A jump-transposition error) esiintyy, kun kaksi eri numeroa, joiden välissä on kolmas numero, vaihtaa paikkaa. Kaikista esiintyvistä virheistä 0.8% on tällaisia.

- *Kaksosvirhe* (A twin error) esiintyy, kun kaksi samaa vierekkäistä numeroa vaihtuu toiseksi samoista numeroista koostuvaksi pariksi. Kaikista virheistä 0.5% on tällaisia virheitä.

- *Äänteellinen virhe* (A phonetic error) esiintyy, kun kaksi numeroa, jotka esitetään suullisesti, kirjataan virheellisesti. Tämä on yleistä varsinkin englannin kielessä, sillä esimerkiksi sanat ”fourteen” (14) ja ”forty” (40) kuulostavat äänen lausuttuna lähes samoilta. 0.5% kaikista esiintyvistä virheistä on tätä tyyppiä.

- *Hyppäys-kaksos -virhe* (A jump-twin error) esiintyy, kun kaksi samaa numeroa, joiden välissä on kolmas numero, vaihtuu toisiksi numeroiksi, jotka ovat keskenään samoja. Kaikista esiintyvistä virheistä 0.3% on tätä tyyppiä.

Taulukossa 4 annetaan esimerkit kustakin virhetyypistä ja siinä käytetään mallina 6-numeroista tunnistenumeroa $a_1a_2a_3a_4a_5a_6 = 181455$. Taulukko 5 puolestaan listaa kaikki virhetyypit ja niiden suhteelliset frekvenssit.

**Taulukko 4.** Esimerkit eri virhetyypeistä.

| Virhetyyppi | Todellinen numero | Siirtovirhe |
|---|---|---|
| Yksi numero | 181455 | 181955 |
| Vierusnumeroiden paikanvaihto | 181455 | 181545 |
| Hyppäys-paikanvaihto | 181455 | 185415 |
| Kaksonen | 181455 | 181433 |
| Äänteellinen | 181455 | 184055 |
| Hyppäys-kaksonen | 181455 | 383455 |

**Taulukko 5.** Virhetyyppien muodot ja suhteelliset frekvenssit.

| Virhetyyppi | Muoto | Suhteellinen frekvenssi |
|---|---|---|
| Yksi numero | $a \to b$ | 79.1% |
| Vierusnumeroiden paikanvaihto | $ab \to ba$ | 10.2% |
| Hyppäys-paikanvaihto | $abc \to cba$ | 0.8% |
| Kaksonen | $aa \to bb$ | 0.5% |
| Äänteellinen | $a0 \leftrightarrow 1a^*$ | 0.5% |
| Hyppäys-kaksonen | $aca \to bcb$ | 0.3% |

*Kun $a = 2, 3, \ldots, 9$.

Mikä tahansa taulukossa 5 esitetyistä virheistä voi esiintyä, kun tunnistenumero siirretään paikasta toiseen. Numeron vastaanottaja ei voi tietää, onko numero oikea, ennen kuin hän on yhteydessä numeron lähettäjään. Aina ei kuitenkaan ole mahdollista tavoittaa tunnistenumeron lähettäjää. Juuri tämän perusteella on kehitetty tässäkin tutkielmassa esitettyjä menetelmiä, joiden avulla tunnistenumeron vastaanottaja voi havaita, kun numero on siirretty virheellisesti.

Tarkastellaan seuraavaksi tunnistenumeroita, jotka ovat muotoa $a_1a_2 \ldots a_n$ ja täyttävät lisäksi ehdon

$$(a_1, a_2, \ldots, a_n) \cdot (w_1, w_2, \ldots, w_n) \equiv 0 \pmod{k}.$$

Tällaisille tunnistenumeroille voidaan helposti määrittää mahdolliset yhden numeron virheet sekä paikanvaihtovirheet, kuten seuraava lause osoittaa.

**Lause 2.1.** *Olkoon $a_1a_2\ldots a_n$ tunnistenumero, jolle pätee*

$$(a_1, a_2, \ldots, a_n) \cdot (w_1, w_2, \ldots, w_n) \equiv 0 \pmod{k}.$$

*Tällöin yhden numeron virhe $a_i \to a_i'$ on tunnistamaton, jos ja vain jos $(a_i - a_i')w_i \equiv 0 \pmod{k}$, ja paikanvaihtovirhe, joka vaihtaa numerot $i$. ja $j$. paikoilla $(i \neq j)$, on tunnistamaton, jos ja vain jos $(a_i - a_j)(w_i - w_j) \equiv 0 \pmod{k}$.*

*Todistus* (vrt. [1, s. 198]). Oletetaan, että tunnistenumerossa on virhe $i$. numeron kohdalla, eli $a_i$ on vaihtunut numeroon $a_i'$. Tällöin todellisesta numerosta lasketun pistetulon ero virheellisestä numerosta lasketun pistetulon kanssa on $(a_i - a_i')w_i$. Siten virhe on tunnistamaton, jos ja vain jos $(a_i - a_i')w_i \equiv 0 \pmod{k}$.

Oletetaan sitten, että on tapahtunut virhe muotoa

$$\cdots a_i a_{i+1} \cdots a_j a_{j+1} \cdots \to \cdots a_j a_{i+1} \cdots a_i a_{j+1} \cdots .$$

Tällöin todellisen ja virheellisen tunnistenumeron pistetulojen erotus on

$$(a_i w_i + a_j w_j) - (a_j w_i + a_i w_j) = (a_i - a_j)(w_i - w_j).$$

Näin ollen virhe on tunnistamaton, jos ja vain jos

$$(a_i - a_j)(w_i - w_j) \equiv 0 \pmod{k}.$$

□

Kun tarkistusnumero toteuttaa ehdon

$$(a_1, a_2, \ldots, a_n) \cdot (w_1, w_2, \ldots, w_n) \equiv 0 \pmod{k}$$

ja numerot $a_1, a_2, \ldots, a_n$ ovat rajoitettuja välille $0, 1, \ldots, k-1$, voidaan määrittää ehdot, joiden mukaan kukin virhetyyppi on tunnistettavissa. Nämä ehdot on esitetty seuraavan sivun taulukossa 6.

**Taulukko 6.** Virhetyyppien tunnistamisehdot.

| Virhetyyppi | Muoto | Ehto |
|---|---|---|
| Yksi numero | $a_i \to a_i'$ | $(w_i, k) = 1$ |
| Paikanvaihto | $\cdots a_i \cdots a_j \cdots \to \cdots a_j \cdots a_i \cdots$ | $(w_j - w_i, k) = 1$ |
| Kaksonen | $aa \to bb$ (paikat $i$ ja $i+1$) | $(w_i + w_{i+1}, k) = 1$ |
| Äänteellinen | $a0 \leftrightarrow 1a$ (paikat $i$ ja $i+1$) | $aw_{i+1} \not\equiv (a-1)w_i$* |
| Hyppäys-kaksonen | $aca \leftrightarrow bcb$ (paikat $i, i+1$ ja $i+2$) | $(w_i + w_{i+2}, k) = 1$ |

*Aina, kun $a = 2, 3, \ldots, k-1$.

Lause 2.1 ja taulukko 6 osoittavat, miksi tarkistusnumerojärjestelmät, joissa käytetään modulina pienempiä lukuja kuin 10, ovat melko harvinaisia, kun taas moduli 11 -järjestelmät ovat melko yleisiä. Tapauksessa, jossa moduli $k$ on pienempi kuin 10, kaikkia yhden numeron virheitä eikä paikanvaihtovirheitä voida tunnistaa ilman, että rajoitetaan numerot välille $0, 1, \ldots, k-1$. Esimerkiksi moduli 7 -järjestelmä, jota käyttää ainakin jotkut lentoyhtiöt, ei voi erottaa lukuja $a$ ja $a'$, kun $|a - a'| = 7$. Toisaalta, moduli 11 -järjestelmien kohdalla, edellisessä lauseessa esitetyt ehdot yhden numeron virheiden ja paikanvaihtovirheiden tunnistamiseksi voidaan toteuttaa yksinkertaisesti valitsemalla kullekin numerolle erilliset painot lukujen 0 ja 10 väliltä. Kuten tämänkin tutkielman joistain esimerkeistä voidaan huomata, moduli 11 -järjestelmien ainoa haittapuoli on joko tarve ottaa käyttöön erikoismerkkejä tai välttää tiettyjen numeroiden käyttämistä.

# Viitteet

[1] Gallian, Joseph A., *The Mathematics of Identification Numbers*, The College Mathematics Journal, Vol. 22, No. 3 (1991) 98-199.

[2] Kirtland, Joseph, *Identification Numbers and Check Digit Schemes*, Mathematical Association of America, Washington, 2001.

[3] Koshy, Thomas, *Elementary Number Theory with Applications*, Harcourt/Academic Press, San Diego, 2002.

[4] Rosen, Kenneth H., *Elementary Number Theory and its Applications*, 3rd ed., Addison-Wesley Publishing Company, Reading, Massachusetts, 1993.

[5] Väestörekisterikeskus - Palvelut kansalaisille - Henkilötunnus [Verkkodokumentti], 2007 [Viitattu 29.2.2008], URL http://www.vaestorekisterikeskus.fi/vrk/home.nsf/pages/2575C011A439C3ACC22571FB00248CA2.