Liitteet

Liite 1 Kyselykaavakkeet

Kyselykaavakkeet on lähetetty valituille matkaviestinoperaattoreille ja palveluoperaattoreille sähköisessä muodossa seuraavan lähtökohtaisen jaottelun mukaisesti. Tutkimuksen aikana kyselyitä tarkennettiin. Lopulliset sähköiset kyselylomakkeet, täydentävät kysymykset ja niihin saadut luottamukselliset vastaukset ovat Viestintäviraston hallussa.

I Verkko- ja liikennetiedot Viestintävirastolta ja operaattoreilta Viestintäviraston kautta VTT:lle (Work package Technological analysis)

Tiedot tukiasemittain valituilta postinumeroalueilta (6-7 kpl) yhtenäisessä formaatissa
- tukiaseman tyyppi (mikrosolu / makrosolu, 900 / 1800)
- sijainti (formaatti kkj tai ykj)
- antennin korkeus
- onko sisä- vai ulkotukiasema
- kanavanumerot
- TRX lukumäärä
- lähetysteho
- liikennemäärät (kiiretunti); yhteisesti sovitulla tavalla mitattuna
- esto (tukiasemittain, jos mahdollista ja/tai suunnittelussa käytetty eston raja)

Arvio eri tukiasematyyppien teknisestä kestävyydestä sekä teknisestä ja taloudellisesta käyttöiästä

Arvio verkon uudelleensuunnittelun ja -konfiguroinnin vaatimasta työmäärästä kussakin skenaariossa kullakin operaattorilla

E-GSM:ään liittyvät tiedot
- operaattorien arvio E-GSM:n käyttöönottoon liittyvän kofigurointityön laajuudesta ja kustannuksista
- operaattorien arvio E-GSM-kanavien käyttöönottoon liittyvästä uusien tavallisten ja mikrotukiasemien hankintamäärästä ja vastaavista kustannuksista
- operaattorien arvio muista E-GSM-käytöstä aiheutuvat tekniset ongelmat ja kustannukset

II Taloudelliseen analyysiin tarvittavat tekniset tiedot VTT:ltä Etlatiedolle (Work package Economic modelling)

- Operaattoreilta Viestintäviraston välityksellä saataviin tietoihin perustuva arvio siitä, miten paljon työtä ja teknisiä investointitarpeita liittyy E-GSM-kanavien käyttöön:
    - Arvio E-GSM-kanavien käyttöönottoon liittyvän konfigurointityön tarpeesta eri operaattoreiden laitekannassa

  - Arvio E-GSM-kanavien käyttöönottoon liittyvästä uusien tavallisten ja mikrotukiasemien hankintamäärätarpeesta eri operaattoreiden laitekannassa
  - Muut E-GSM-käytöstä aiheutuvat tekniset ongelmat
- Tukiasemien käyttöikä
  - Tekninen kestävyys ja käyttöikä. Operaattoreilta Viestintäviraston välityksellä saataviin tietoihin perustuva arvio siitä, onko käyttöiässä eroa uudempien ja vanhempien tukiasemien välillä tai tavallisten ja mikrotukiasemien välillä
  - Näkemys tukiasemien taloudellisesta käyttöiästä, jos se poikkeaa merkittävästi teknisestä käyttöiästä.
- Operaattoreilta Viestintäviraston kautta saatavan tiedon pohjalta laaditaan arvio verkon uudelleensuunnittelun ja -konfiguroinnin vaatimasta työmäärästä kussakin skenaariossa kullakin operaattorilla
- Mistä osista työmääräarvio koostuu
- Mikä on kunkin operaattorin hyväksymä esto% verkossaan
- Arvio eri skenaarioiden vaikutuksesta tukiasemamääriin kullakin tarkasteluun valitulla postinumeroalueella nykyisillä liikennemäärillä
  - Operaattorikohtaisesti
  - Tukiasematyypeittäin (makro/mikro, E-GSM/P-GSM/1800)
- Teknisen analyysin perusteella arvio kussakin skenaariossa kullakin operaattorilla
  - Mikä liikennemäärä / km^2 aiheuttaa sen, että tukiasemien lukumäärän määrääväksi tekijäksi tulee alueellisen peiton sijasta tarve tarjota riittävä määrä kanavia / km^2
  - Mikä liikennemäärä / km^2 aiheuttaa sen, että tukiasemasuunnittelussa siirrytään limittäisestä suunnittelusta hierarkkiseen suunnitteluun
  - Arvio, minkä muotoinen on funktio, jossa riippumattomana muuttujana on joko kanavamäärä tai liikennemäärä / km^2 ja riippuvana muuttujana tukiasemien määrä / km^2
  - Kullakin operaattorilla kussakin suunnittelutapauksessa (peitto/limittäinen/hierarkkinen) kussakin skenaariossa

  - Riippuvana muuttujana sekä tukiasemien kokonaismäärä että kukin tukiasematyyppi erikseen (makro/mikro, 900/1800)

III Taloudelliseen analyysiin tarvittavat yleistiedot Viestintävirastolta Etlatiedolle (Work packages Economic modelling and Market and competition effects, social benefits)

- Verkko-operaattorien markkinaosuudet tilaajamäärien ja liikevaihtojen suhteen 2002-2003
- Tilaajamäärät verkoittain ja verkko-operaattoreittain 2002-2003
- Liikennemäärät verkko-operaattoreittain 2002-2004

IV Taloudelliseen analyysiin tarvittavat tiedot operaattoreilta Viestintäviraston kysymänä Etlatiedolle (Work packages Economic modelling and Market and competition effects, social benefits)

- Tiedot tukiasemien ja tukiasemalokaatioiden hinnoista
  - Pelkän tukiasemauuslaiteinvestoinnin hinta (keskiarvo eri tukiasematyypeille makro/mikro, P-GSM/P+E-GSM/E-GSM/1800)
  - Muun lokaatiouusinvestoinnin hinta ilman tukiasemaa (keskiarvo eri tukiasematyypeille makro/mikro, P-GSM/P+E-GSM/E-GSM/1800)
- Tukiasemien taloudelliset käyttöiät ja operaattorin tukiasemainvestoinneissaan käyttämät laskentaperusteet
- Operaattoreiden laskelmissaan käyttämät käyttöiät eri investointierille
- Operaattoreiden laskelmissaan käyttämä korkokanta ja sen määrittämisperuste
- Operaattoreiden nykyisen tukiasema- ja tukiasemalokaatiokannan määrät, hankintahinnat sekä nykyiset kirjanpitoarvot (vuoden 2004 loppu) seuraavin jaotteluin
  - Tukiasematyypit makro/mikro, P-GSM/P+E-GSM/E-GSM/1800
  - Hankintavuosi: 1991-1995; 1996-1997; 1998-1999; 2000-2001; 2002-2003; 2004-2005

Verkko-operaattorien kustannusrakennetiedot:

- Henkilömäärä ja henkilökustannukset neljännesvuosittain vuosilta 2002-2004
- CAPEX 2003 ja 2004 ja sen muodostuminen seuraavasti jaoteltuna
    - BS
    - Muu BS-lokaation kustannus (ml. suora työ, mutta ei yleiskustannuksia)
    - BSC- ja MSC-kokonaiskustannus (ml. suora työ, mutta ei yleiskustannuksia)
    - Verkko-investointien hallinnan ja laskennan kustannus (ei yleiskustannuksia)
    - Muu GSM-verkko-operaattorin CAPEX (ml. jyvitetyt yleiskustannukset ja jos niitä käytetään, jyvitysperuste)
- OPEX 2004 ja sen muodostuminen seuraavasti jaoteltuna
    - BS-lokaatioiden vuokrat
    - Muiden GSM-verkkoinfra-lokaatioiden vuokrat
    - GSM-verkon aiheuttamat tietoliikennekustannukset (suorat; jos sisäistä laskutusta, määritysperusteet)
    - Verkon ylläpidon ja hallinnan kustannukset (ml. suora työ)
    - Liittymien suora hankinta- ja provisiointikustannus (ml. liittymien hankkijoille maksetut provisiot, SIM-korttien kustannukset ja mahdolliset päätelaitetuet)
    - Verkko-operaattorin asiakaspalvelu
    - Muut verkko-operaattorin henkilöstökustannukset
    - Muut verkko-operaattorin kustannukset (ml. jyvitetyt yleiskustannukset ja jos niitä käytetään, jyvitysperuste)

Palveluoperaattorien kustannusrakennetiedot:

- Henkilömäärä ja henkilökustannukset neljännesvuosittain vuosilta 2002-2004
- CAPEX 2003 ja 2004

- OPEX 2004
- Kuluttaja-asiakkaiden liikennemäärät kuukausittain vuosilta 2001-2005/1 ja neljännesvuosittain vuosilta 1997-2000
- Yritysasiakkaiden liikennemäärät kuukausittain vuosilta 2001-2005/1 ja neljännesvuosittain vuosilta 1997-2000
- Kuluttajaliittymien lukumäärä ja palvelulaskutus kuukausittain vuosilta 2001-2005/1 ja neljännesvuosittain vuosilta 1997-2000
- Yritysliittymien lukumäärä ja palvelulaskutus kuukausittain vuosilta 2001-2005/1 ja neljännesvuosittain vuosilta 1997-2000

Liite 2 Alueiden kertoimet

Alueiden kertoimet[1]

Postino	Nimi	Pinta-ala	Asukasluku	Asukas-tiheys	Liikennemäärä-ennuste	Taajuudet	Alue-luokka	Edustavuus
00100	Helsinki keskusta	2,9	15640	5393,103	2,021316	1	9	1,870968
00120	Punavuori	0,5	6039	12078	2,321877	1	10	2,333333
00130	Kaartinkaupunki	0,5	1375	2750	2,186459	1	10	2,333333
00140	Kaivopuisto	0,9	7516	8351,111	1,899223	1	9	1,870968
00150	Eira	1,5	8439	5626	1,828836	1	9	1,870968
00160	Katajanokka	0,7	4130	5900	2,028721	1	9	1,870968
00170	Kruunuhaka	0,9	6711	7456,667	2,063992	1	9	1,870968
00180	Ruoholahti	2,8	12032	4297,143	1,783756	1	8	2,624107
00190	Suomenlinna	1,1	833	757,2727	1,209656	1	6	2,001057
00200	Lauttasaari	3,9	14170	3633,333	1,43555	1	7	1,651497
00210	Vattuniemi	1,6	4953	3095,625	1,729155	1	8	2,624107
00240	Länsi-Pasila	2,2	4551	2068,636	1,805028	1	8	2,624107
00250	Taka-Töölö	3,6	10400	2888,889	1,561617	1	8	2,624107
00260	Keski-Töölö	0,7	4997	7138,571	2,00914	1	9	1,870968
00270	Pohjois-Meilahti	0,9	7327	8141,111	1,888402	1	9	1,870968
00280	Ruskeasuo	0,9	2674	2971,111	1,549396	1	7	1,651497
00290	Meilahden sairaala-alue	0,3	175	583,3333	1,532457	1	7	1,651497
00300	Pikku Huopalahti	1,4	5600	4000	1,670263	1	8	2,624107
00310	Kivihaka	0,5	891	1782	1,720043	1	8	2,624107
00320	Etelä-Haaga	2	8899	4449,5	1,573067	1	8	2,624107

[1] Demografiatiedon mukainen ennuste kapasiteetin tarpeesta määrää postinumeroalueiden jakamisen kymmeneen luokkaan. Se, kuinka suurta osaa koko Suomesta kukin analysoitu alue edustaa, määräytyy laskemalla kunkin luokan kaikkien postinumeroalueiden pinta-alat yhteen ja jakamalla tämä luokan tutkittujen alueiden pinta-aloilla. Kaikki saman luokan alueen muutokset verkkoelementeissä kerrotaan saadulla osamäärällä (edustavuuskerroin).

00330	Munkkiniemi	2,2	8353	3796,818	1,556196	1	8	2,624107
00340	Kuusisaari-Lehtisaari	1,3	1614	1241,538	1,084845	1	6	2,001057
00350	Munkkivuori	1,9	8880	4673,684	1,623797	1	8	2,624107
00360	Pajamäki	0,8	1886	2357,5	1,328949	1	7	1,651497
00370	Reimarla	2,2	5815	2643,182	1,481148	1	7	1,651497
00380	Pitäjänmäki	1,1	3408	3098,182	1,950792	1	9	1,870968
00390	Konala	3,3	4481	1357,879	1,389298	1	7	1,651497
00400	Pohjois-Haaga	1,6	8392	5245	1,668054	1	8	2,624107
00410	Malminkartano	2	8481	4240,5	1,519619	1	7	1,651497
00420	Kannelmäki	2,8	13330	4760,714	1,560206	1	8	2,624107
00430	Maununneva	3,7	4759	1286,216	0,951472	1	5	1,483984
00440	Lassila	1,3	4438	3413,846	1,692032	1	8	2,624107
00500	Sörnäinen	0,5	11749	23498	2,367491	1	10	2,333333
00510	Etu-Vallila	1,1	8165	7422,727	2,095017	1	9	1,870968
00520	Itä-Pasila	1,3	5425	4173,077	1,980872	1	9	1,870968
00530	Kallio	1,5	17368	11578,67	2,094686	1	9	1,870968
00550	Vallila	1,1	9299	8453,636	1,902084	1	9	1,870968
00560	Toukola-Vanhakaupunki	6,9	4416	640	1,098481	1	6	2,001057
00570	Kulosaari	2,2	3780	1718,182	1,174335	1	6	2,001057
00580	Hermanni teollisuusalue	2,8	293	104,6429	1,338172	1	7	1,651497
00600	Koskela	1,1	3994	3630,909	1,510329	1	7	1,651497
00610	Käpylä	2,2	7959	3617,727	1,459776	1	7	1,651497
00620	Metsälä	1,9	2900	1526,316	1,468606	1	7	1,651497
00630	Maunula-Suursuo	2,7	7857	2910	1,271644	1	6	2,001057
00640	Oulunkylä-Patola	2,3	7522	3270,435	1,449029	1	7	1,651497
00650	Veräjämäki	1,6	4372	2732,5	1,235125	1	6	2,001057
00660	Länsi-Pakila	2,6	6224	2393,846	1,212242	1	6	2,001057
00670	Paloheinä	4,8	6139	1278,958	0,831903	1	5	1,483984
00680	Itä-Pakila	2,3	3541	1539,565	1,119038	1	6	2,001057
00690	Tuomarinkylä	5,3	2788	526,0377	0,676993	1	4	4,044304
00700	Malmi	6,4	13186	2060,313	1,369041	1	7	1,651497
00710	Pihlajamäki	2	11160	5580	1,566993	1	8	2,624107
00720	Pukinmäki-Savela	2,2	8482	3855,455	1,469795	1	7	1,651497
00730	Tapanila	3,7	10141	2740,811	1,127913	1	6	2,001057
00740	Siltamäki	3,2	10403	3250,938	1,234151	1	6	2,001057
00750	Puistola	2,6	9081	3492,692	1,342239	1	7	1,651497
00760	Suurmetsä	3,1	7963	2568,71	1,133131	1	6	2,001057
00770	Jakomäki	3,5	5706	1630,286	1,20133	1	6	2,001057
00780	Tapaninvainio	2,4	11487	4786,25	1,268279	1	6	2,001057
00800	Länsi-Herttoniemi	3,9	7125	1826,923	1,089898	1	6	2,001057
00810	Herttoniemi	3,9	10408	2668,718	1,472693	1	7	1,651497
00820	Roihuvuori	1,3	7107	5466,923	1,528416	1	7	1,651497
00830	Tammisalo	2,2	2215	1006,818	0,868228	1	5	1,483984

00840	Laajasalo	6,2	8392	1353,548	0,912968	1	5	1,483984
00850	Jollas	5,8	3010	518,9655	0,639526	1	4	4,044304
00860	Santahamina	6,8	612	90	0,779758	1	5	1,483984
00870	Etelä-Laajasalo	1,7	4554	2678,824	1,153638	1	6	2,001057
00900	Puotinharju	1,3	3712	2855,385	1,421645	1	7	1,651497
00910	Puotila	1	5819	5819	1,519145	1	7	1,651497
00920	Myllypuro	3,7	9439	2551,081	1,125672	1	6	2,001057
00930	Itäkeskus-Marjaniemi	2,4	6812	2838,333	1,517986	1	7	1,651497
00940	Kontula	6,3	25028	3972,698	1,235015	1	6	2,001057
00950	Vartioharju	4	5528	1382	0,983639	1	5	1,483984
00960	Pohjois-Vuosaari	4,7	6686	1422,553	0,904535	1	5	1,483984
00970	Mellunkylä	1,6	10031	6269,375	1,451754	1	7	1,651497
00980	Etelä-Vuosaari	12,8	22218	1735,781	0,959874	1	5	1,483984
00990	Helsinki	0,6	2041	3401,667	1,105112	1	6	2,001057
01200	Hakunila	15,5	11726	756,5161	0,75498	1	5	1,483984
01230	Vaarala	2,9	2792	962,7586	1,027318	1	6	2,001057
01260	Itä-Hakkila	8,8	4217	479,2045	0,693776	1	4	4,044304
01280	Länsimäki	4,4	9700	2204,545	1,012974	1	6	2,001057
01300	Tikkurila	6	16194	2699	1,383825	1	7	1,651497
01350	Hiekkaharju	2,9	8152	2811,034	1,136491	1	6	2,001057
01360	Koivukylä-Havukoski	6,7	11211	1673,284	0,984685	1	5	1,483984
01370	Jokiniemi	2,8	5521	1971,786	1,304507	1	7	1,651497
01380	Kuusikko-Hakkila	5,4	1909	353,5185	0,846191	1	5	1,483984
01390	Ruskeasanta-Ilola	6,4	7724	1206,875	0,811832	1	5	1,483984
01400	Rekola	5,1	7292	1429,804	0,858376	1	5	1,483984
01420	Päiväkumpu	3,5	3118	890,8571	0,675522	1	4	4,044304
01450	Korso	14	16490	1177,857	0,785584	1	5	1,483984
01480	Mikkola	2,9	5823	2007,931	1,012829	1	6	2,001057
01490	Nikinmäki	5,7	2692	472,2807	0,534999	1	4	4,044304
01510	Kirkonkylä-Veromäki	7,2	3059	424,8611	1,13557	1	6	2,001057
01520	Pakkala	2,2	2158	980,9091	0,620827	1	4	4,044304
01530	Veromiehenkylä	16,4	219	13,35366	1,009168	1	5	1,483984
01600	Myyrmäki	2,8	14750	5267,857	1,663295	1	8	2,624107
01610	Kaivoksela	2,4	3984	1660	1,336915	1	7	1,651497
01620	Martinlaakso	3	11361	3787	1,4379	1	7	1,651497
01630	Hämeenkylä	2,7	2125	787,037	0,957853	1	5	1,483984
01640	Hämevaara	3,2	2434	760,625	1,073231	1	6	2,001057
01650	Vapaala	1,9	3910	2057,895	1,185493	1	6	2,001057
01660	Varisto	1,2	2418	2015	1,217143	1	6	2,001057
01670	Vantaanlaakso	2,5	2580	1032	0,984004	1	5	1,483984
01680	Askisto	9	2077	230,7778	0,467114	1	3	4,587932
01690	Ylästö	5,9	3320	562,7119	0,524439	1	4	4,044304
01700	Kivistö	4,9	1822	371,8367	0,548683	1	4	4,044304
01710	Pähkinärinne	1	4692	4692	1,479701	1	7	1,651497
01720	Petikko	4,3	83	19,30233	1,019259	1	6	2,001057

01730	Vantaanpuisto	13,2	2087	158,1061	0,532211	1	4	4,044304
01740	Tuupakan teollisuusalue	8,5	210	24,70588	0,811703	1	5	1,483984
01750	Keimola	21,3	764	35,86854	0,082729	1	2	6,770064
01760	Seutula	30,4	2114	69,53947	0,125867	1	2	6,770064
02100	Tapiola	2,3	3735	1623,913	1,569861	1	8	2,624107
02110	Otsolahti	0,9	2662	2957,778	1,38807	1	7	1,651497
02120	Länsikorkee-Hakalehto	0,8	2759	3448,75	1,436625	1	7	1,651497
02130	Pohjois-Tapiola	2,4	5920	2466,667	1,339355	1	7	1,651497
02140	Laajalahti-Friisinmäki	3,2	4135	1292,188	0,909181	1	5	1,483984
02150	Otaniemi	2,6	3172	1220	1,756021	1	8	2,624107
02160	Westend	3,2	3090	965,625	0,928022	1	5	1,483984
02170	Haukilahti	2,3	5606	2437,391	1,244099	1	6	2,001057
02180	Mankkaa	3,7	6484	1752,432	0,992584	1	5	1,483984
02200	Niittykumpu	7,4	10060	1359,459	1,160447	1	6	2,001057
02210	Olari-Kuitinmäki	3,5	11285	3224,286	1,351538	1	7	1,651497
02230	Matinkylä	7	16967	2423,857	1,169199	1	6	2,001057
02240	Friisilä	1,3	1546	1189,231	1,275989	1	6	2,001057
02260	Iivisniemi	2,9	4060	1400	0,866632	1	5	1,483984
02270	Suomenoja	3,4	2175	639,7059	1,067384	1	6	2,001057
02280	Puolarmetsä	3,3	4261	1291,212	0,854112	1	5	1,483984
02300	Nöykkiö	1,7	2429	1428,824	0,887873	1	5	1,483984
02320	Kivenlahti	5,3	15511	2926,604	1,208988	1	6	2,001057
02330	Kattilalaakso	6,5	3475	534,6154	0,612575	1	4	4,044304
02340	Latokaski	2,7	5113	1893,704	0,919545	1	5	1,483984
02360	Espoonlahti	6,4	8998	1405,938	0,89411	1	5	1,483984
02380	Suvisaaristo	11,3	560	49,55752	0,156398	1	2	6,770064
02600	Etelä-Leppävaara	3,9	16581	4251,538	1,589909	1	8	2,624107
02610	Kilo	3,1	5084	1640	1,191379	1	6	2,001057
02620	Karakallio	3,9	6320	1620,513	0,994806	1	5	1,483984
02630	Nihtisilta	2,7	2697	998,8889	1,39919	1	7	1,651497
02660	Lintuvaara	3,3	5757	1744,545	0,961587	1	5	1,483984
02680	Uusimäki	1,5	565	376,6667	0,57352	1	4	4,044304
02700	Suur-Kauniainen	6	8522	1420,333	1,006369	1	5	1,483984
02710	Viherlaakso	2,4	5577	2323,75	1,139073	1	6	2,001057
02720	Laaksolahti-Lähderanta	2	2660	1330	1,027724	1	6	2,001057
02730	Laaksolahti	2,8	4452	1590	0,894802	1	5	1,483984
02740	Karvasmäki-Kunnarla	23,7	1806	76,20253	0,44067	1	3	4,587932
02760	Espoon keskus Itäinen	2,7	11477	4250,741	1,216429	1	6	2,001057
02770	Espoon keskus läntinen	8,7	9152	1051,954	1,092536	1	6	2,001057
02780	Kauklahti	28,5	5491	192,6667	0,392934	1	3	4,587932
02810	Kumpyöli-Karhusuo	5,3	778	146,7925	0,341271	1	3	4,587932

02820	Nupuri-Nuuksio	65,1	1330	20,43011	-0,11276	1	1	1,200483
02860	Siikajärvi	15	630	42	0,097333	1	2	6,770064
02920	Niipperi	12,9	4034	312,7132	0,658112	1	4	4,044304
02940	Lippajärvi	24,1	8772	363,9834	0,390947	1	3	4,587932
02970	Kalajärvi	23,9	3069	128,41	0,229215	1	3	4,587932
02980	Lakisto	28,5	387	13,57895	0,28841	1	3	4,587932
04200	Kerava keskus	8,7	10373	1192,299	1,001776	1	5	1,483984
04220	Ahjo-Kaskela	7,7	4050	525,974	0,661252	1	4	4,044304
04230	Kaleva	3,2	6690	2090,625	0,935601	1	5	1,483984
04250	Alikerava	10,1	4124	408,3168	0,849994	1	5	1,483984
04260	Savio	5,6	6126	1093,929	0,862822	1	5	1,483984
04300	Hyrylä	34,1	9065	265,8358	0,512219	1	4	4,044304
04310	Tuusula keskus	8,6	2382	276,9767	0,476104	1	4	4,044304
04320	Riihikallio	6,6	4376	663,0303	0,626482	1	4	4,044304
04330	Lahela	4,4	1967	447,0455	0,394478	1	3	4,587932
04340	Mallila	3	1961	653,6667	0,759486	1	5	1,483984
04350	Nahkela	18,6	506	27,2043	-0,04551	1	2	6,770064
04360	Ruotsinkylä	28	1195	42,67857	0,303051	1	3	4,587932
04370	Rusutjärvi	16,5	670	40,60606	0,011495	1	2	6,770064
04380	Tuomala	15,5	545	35,16129	-0,18627	1	1	1,200483
04390	Jäniksenlinna	14,4	393	27,29167	-0,05178	1	2	6,770064
04400	Järvenpää keskus	17,3	18397	1063,41	0,866954	1	5	1,483984
04420	Kyrölä	4,2	3678	875,7143	0,624625	1	4	4,044304
04430	Kinnari-Mikonkorpi	12,5	8230	658,4	0,610728	1	4	4,044304
04440	Jamppa	7,2	4651	645,9722	0,755439	1	5	1,483984
04460	Nummenkylä	4,6	1233	268,0435	0,412914	1	3	4,587932
04480	Haarajoki	9,8	1523	155,4082	0,290145	1	3	4,587932
04500	Kellokoski	37,4	4852	129,7326	0,133287	1	2	6,770064
05400	Jokela	31,3	5615	179,393	0,201488	1	3	4,587932
05430	Nuppulinna	17,5	927	52,97143	0,007392	1	2	6,770064
15100	Asemanseutu	2,2	5841	2655	1,564235	0	9	14,58696
15110	Keski-Lahti	1,3	4871	3746,923	1,866693	0	10	1,130435
15140	Paavola	2,5	12678	5071,2	1,720474	0	10	1,130435
15150	Möysä	24,6	4445	180,6911	0,499812	0	6	12,36008
15170	Tenava-Tonttila	3,4	4087	1202,059	1,051226	0	7	14,75216
15200	Kiveriö	2,3	3785	1645,652	1,272394	0	8	10,68027
15210	Kivistönmäki	3	4482	1494	1,187943	0	8	10,68027
15230	Kytölä	8,6	2442	283,9535	0,598883	0	6	12,36008
15240	Mukkula	12,7	8017	631,2598	0,685409	0	6	12,36008
15300	Myllypohja-Koiskiala	12,9	3987	309,0698	0,482758	0	6	12,36008
15320	Ahtiala	4,2	2297	546,9048	0,575692	0	6	12,36008
15340	Kunnas-Hiekkanummi	46,8	2646	56,53846	-0,09827	0	4	155,6789
15500	Liipola	2,7	5313	1967,778	1,209757	0	8	10,68027
15520	Saksala	3	2526	842	1,084072	0	7	14,75216

15610	Laune-Nikkilä	7,4	5134	693,7838	0,746803	0	6	12,36008
15680	Renkomäki-Ämmälä	14,1	2259	160,2128	0,271635	0	5	70,0377
15700	Hennala-Jokimaa	47,9	5485	114,5094	0,281598	0	5	70,0377
15800	Kärpänen	2,9	3450	1189,655	1,202246	0	8	10,68027
15810	Kasakkamäki	1,3	1408	1083,077	0,945872	0	7	14,75216
15820	Okeroinen	22,6	919	40,66372	0,077381	0	4	155,6789
15830	Metsäkangas	1,6	3976	2485	1,174094	0	8	10,68027
15840	Riihelä	1,7	2298	1351,765	0,835692	0	7	14,75216
15850	Likolampi	2,9	670	231,0345	0,711437	0	6	12,36008
15900	Jalkaranta	3,7	1698	458,9189	0,68018	0	6	12,36008
15950	Kiikkula	11	4130	375,4545	0,420563	0	5	70,0377
20100	Turku keskus	13,6	23045	1694,485	1,797654	1	8	2,624107
20200	Iso-Heikkilä	2,6	1922	739,2308	1,344776	1	7	1,651497
20210	Pahaniemi	3,8	4501	1184,474	1,019177	1	6	2,001057
20240	Pansio-Perno	6,8	5557	817,2059	1,014721	1	6	2,001057
20250	Pitkämäki	1,4	778	555,7143	1,070975	1	6	2,001057
20300	Pohjola-Kastu	4,5	5696	1265,778	1,073296	1	6	2,001057
20320	Länsikeskus	11,5	14309	1244,261	0,988915	1	5	1,483984
20360	Kärsämäki- Halinen	15,6	11157	715,1923	0,849723	1	5	1,483984
20380	Räntämäki-Saramäki	20,9	2704	129,378	0,613633	1	4	4,044304
20400	Moisio	13,9	2312	166,3309	0,150738	1	2	6,770064
20460	Jäkärla	19	3874	203,8947	0,267037	1	3	4,587932
20500	Vesilinna	0,9	5353	5947,778	1,85784	1	9	1,870968
20520	Kupittaa	2,1	2595	1235,714	1,492491	1	7	1,651497
20540	Nummi	11,2	18421	1644,732	0,964182	1	5	1,483984
20610	Varissuo	3,5	13342	3812	1,218536	1	6	2,001057
20660	Littoinen	24	6517	271,5417	0,285352	1	3	4,587932
20700	Vartiovuori	1,1	5163	4693,636	1,709067	1	8	2,624107
20720	Vasaramäki	4,6	6087	1323,261	0,986712	1	5	1,483984
20740	Ilpoinen-Harittu	5	10040	2008	1,066333	1	6	2,001057
20750	Huhkola-Lauste-Vaala	3	5659	1886,333	1,159719	1	6	2,001057
20760	Piispanristi	10,8	3643	337,3148	0,585458	1	4	4,044304
20780	Kaarina keskus	16,5	10402	630,4242	0,709554	1	4	4,044304
20810	Martti-Korppolaismäki	4,4	15449	3511,136	1,329993	1	7	1,651497
20880	Uittamo	1,9	4287	2256,316	1,047667	1	6	2,001057
20900	Moikoinen-Pikisaari	26,4	6430	243,5606	0,273401	1	3	4,587932
20960	Kakskerta	35,2	1383	39,28977	-0,06166	1	2	6,770064
21100	Naantali keskus	40,2	6512	161,99	0,383499	1	3	4,587932
21110	Taimo-Nuhjala-Lietsala	14,2	7049	496,4085	0,519891	1	4	4,044304
21120	Viherjäinen	7,2	1778	246,9444	0,329979	1	3	4,587932
21200	Raisio keskus	21,2	11547	544,6698	0,767715	1	5	1,483984
21210	Hakinmäki	2,8	2576	920	0,851455	1	5	1,483984
21220	Somersoja	7,2	1420	197,2222	0,327536	1	3	4,587932

21260	Ihala	3,5	3591	1026	0,812507	1	5	1,483984
21280	Kuloinen	3,2	2079	649,6875	0,896458	1	5	1,483984
21330	Paattinen	39,4	1772	44,97462	-0,01969	1	2	6,770064
21340	Tortinmäki	43,2	542	12,5463	-0,20045	0	3	130,136
21350	Ilmarinen	8,6	1609	187,093	0,250234	1	3	4,587932
21360	Lieto asemanseutu	54	1740	32,22222	-0,07994	1	1	1,200483
21410	Vanhalinna	9,4	1932	205,5319	0,296493	1	3	4,587932
21420	Lieto keskus	58,8	6216	105,7143	0,216622	1	3	4,587932
21430	Yliskulma	43,8	745	17,00913	-0,06467	0	4	155,6789
21500	Piikkiö	86,4	6064	70,18519	0,053318	1	2	6,770064
21620	Lielax	20,8	1520	73,07692	0,086504	1	2	6,770064
40100	Jyväskylä keskus	3,1	10344	3336,774	1,767737	0	10	1,130435
40200	Mannila-Taulumäki	3,6	3171	880,8333	0,996263	0	7	14,75216
40250	Ritoniemi-Lohikoski	19,8	5766	291,2121	0,365457	0	5	70,0377
40270	Pappilanrinne	77,3	9594	124,1138	0,131767	0	4	155,6789
40320	Seppälän teollisuusalue	7,7	3896	505,974	1,057614	0	7	14,75216
40340	Huhtasuo	2,9	6763	2332,069	1,161553	0	8	10,68027
40400	Halssila	3,8	3822	1005,789	0,869909	0	7	14,75216
40420	Jyskä	5,3	4436	836,9811	0,711731	0	6	12,36008
40500	Keljo-Rstonmaa	22,3	3045	136,5471	0,536097	0	6	12,36008
40520	Kuokkala-Ristikivi	19,8	14415	728,0303	0,655802	0	6	12,36008
40530	Keljonkangas	7,5	3906	520,8	0,586311	0	6	12,36008
40600	Mattilanpelto	1,3	2123	1633,077	1,187378	0	8	10,68027
40620	Keskussairaala-alue	0,4	326	815	1,445572	0	9	14,58696
40630	Kypärämäki	30,8	3961	128,6039	0,282015	0	5	70,0377
40640	Keltinmäki	12,6	6269	497,5397	0,570141	0	6	12,36008
40660	Länsi-Palokka	18,5	596	32,21622	-0,1208	0	4	155,6789
40700	Mäki-Matti	2	3917	1958,5	1,394662	0	9	14,58696
40720	Nisula	2,5	2794	1117,6	1,109423	0	8	10,68027
40740	Kortepohja	4,8	5924	1234,167	0,959073	0	7	14,75216
40800	Vaajakoski	87	7765	89,25287	0,107847	0	4	155,6789
40820	Haapaniemi	3,1	939	302,9032	0,548263	0	6	12,36008
40900	Säynätsalo	36	3435	95,41667	0,118178	0	4	155,6789
41120	Puuppola	44,8	1393	31,09375	-0,14499	0	3	130,136
41140	Kuikka	40,1	469	11,69576	-0,23398	0	3	130,136
41160	Tikkakoski	49	4545	92,7551	0,231598	0	5	70,0377
41450	Leppälahti	65,7	626	9,528158	-0,27818	0	3	130,136
41630	Oravasaari	100	354	3,54	-0,37569	0	3	130,136
41930	Kuohu	113,7	698	6,138962	-0,32845	0	3	130,136
41940	Vesanka	133,2	1178	8,843844	-0,37102	0	3	130,136
60100	Seinäjoki keskus	10,2	8792	861,9608	1,132065	0	8	10,68027
60120	Pohja	3,2	2544	795	1,040469	0	7	14,75216
60150	Kasperi	2,2	3744	1701,818	0,959169	0	7	14,75216
60200	Törnävä	80,1	7587	94,7191	0,028357	0	4	155,6789

60220	Huhtala	10,7	1792	167,4766	0,640833	0	6	12,36008
60320	Alakylä	23,7	6069	256,0759	0,405182	0	5	70,0377
60420	Impivaara	22,1	968	43,8009	-0,08983	0	4	155,6789
61100	Peräseinäjoki keskus	147,6	2223	15,06098	-0,1168	0	4	155,6789
61110	Vuolle	26,9	188	6,988848	-0,20331	0	3	130,136
61120	Louko	68,8	582	8,459302	-0,25877	0	3	130,136
61140	Pasto	12,6	108	8,571429	-0,16194	0	3	130,136
61180	Haapaluoma	70,2	251	3,575499	-0,27945	0	3	130,136
61650	Kalakoski	158,7	492	3,100189	-0,38951	0	3	130,136
90100	Oulu keskus	3,8	12213	3213,947	1,681921	1	8	2,624107
90120	Heinäpää	4,1	5304	1293,659	1,175456	1	6	2,001057
90140	Karjasilta	1,1	2264	2058,182	1,214996	1	6	2,001057
90150	Höyhtyä	1,6	3119	1949,375	1,40056	1	7	1,651497
90160	Nokela-lintula	1,7	3528	2075,294	0,999353	1	5	1,483984
90210	Oulunsuun sairaala	0,5	210	420	1,280943	1	6	2,001057
90220	Kontinkangas	0,9	739	821,1111	1,45341	1	7	1,651497
90230	Värttö-maikkula	3,6	6820	1894,444	0,964696	1	5	1,483984
90240	Iinatti	8,9	7084	795,9551	0,653657	1	4	4,044304
90250	Kaukovainio	3,8	7007	1843,947	1,093052	1	6	2,001057
90310	Madekoski	122,8	3019	24,58469	-0,34277	1	1	1,200483
90420	Mäntylä	16,9	7096	419,8817	0,332614	1	3	4,587932
90440	Kempele keskus	44,9	7741	172,4053	0,250839	1	3	4,587932
90450	Honkanen	50,4	5779	114,6627	0,048537	1	2	6,770064
90460	Oulunsalo keskus	38,6	7766	201,1917	0,229519	1	3	4,587932
90470	Varjakka	37,2	1126	30,26882	-0,19719	1	1	1,200483
90500	Tuira	6,4	8181	1278,281	1,087853	1	6	2,001057
90520	Taskila-Toppila	3,6	4979	1383,056	1,06102	1	6	2,001057
90530	Välivainio	2,2	3182	1446,364	1,257235	1	6	2,001057
90540	Kuivasjärvi	13,1	4042	308,5496	0,297988	1	3	4,587932
90550	Pyykkösjärvi	6,2	5037	812,4194	0,815482	1	5	1,483984
90560	Koskela	1,6	3670	2293,75	1,042223	1	6	2,001057
90570	Kaijonharju-Linnanmaa	6,7	7188	1072,836	1,14963	1	6	2,001057
90580	Rajakylä	3,6	4995	1387,5	0,838426	1	5	1,483984
90630	Heikinharju	32,9	5673	172,4316	0,367632	1	3	4,587932
90650	Haapalehto	197,2	11214	56,86613	-0,12415	1	1	1,200483
90800	Pateniemi	32,2	9076	281,8634	0,145169	1	2	6,770064
90810	Kiviniemi	10,3	2700	262,1359	0,160713	1	2	6,770064
90820	Kello	109,6	2893	26,39599	-0,16577	1	1	1,200483
90830	Haukipudas keskus	33,3	6505	195,3453	0,298187	1	3	4,587932
90840	Haukipudas asema	101,9	1911	18,75368	-0,24839	1	1	1,200483
90850	Martinniemi	10,6	1943	183,3019	0,178591	0	5	70,0377
90860	Halosenniemi	18	489	27,16667	-0,0895	0	4	155,6789

Liite 3 Taloudellisen tarkastelun tunnusluvut

Taulukko 3.1. Taloudelliseen tarkasteluun liittyvät tunnusluvut postinumerokohtaisesti.

Muuttuja	Tyyppi	Kommentit
Yleiset tiedot		
Analyysialue	string	
Postinumero	string	
Pinta-ala	double	km x km
Tukiasemaohjaimien lukumäärä	integer	
Tukiasemien lukumäärä	interger	
Solujen määrä	integer	
Kanavien määrä	integer	
Kuidulla liitettyjen solujen määrä	integer	
Linkillä liitettyjen solujen määrä	integer	
P-GSM900 tiedot		
Tukiasematiedot		
Mikrosolu	integer	
Makrosolu	integer	
Sisätilasolu	integer	
Solutiedot		
Mikrosolu	integer	
Makrosolu	integer	
Sisätilasolu	integer	
Kantoaaltotiedot (solut, jotka eivät sisällä E-GSM-kantoaaltoja)		
Mikrosolu	integer	
Makrosolu	integer	
Sisätilasolu	integer	
E-GSM900 tiedot		
Tukiasematiedot		
Mikrosolu	integer	
Makrosolu	integer	
Sisätilasolu	integer	
Solutiedot		
Mikrosolu	integer	
Makrosolu	integer	
Sisätilasolu	integer	
Kantoaaltotiedot (solut, joissa E-GSM-kantoaaltoja)		
Mikrosolu	integer	
Makrosolu	integer	
Sisätilasolu	integer	
GSM1800 tiedot		
Tukiasematiedot		
Mikrosolu	integer	
Makrosolu	integer	

Muuttuja	Tyyppi	Kommentit
Sisätilasolu	integer	
Solutiedot		
Mikrosolu	integer	
Makrosolu	integer	
Sisätilasolu	integer	
Kantoaaltotiedot		
Mikrosolu	integer	
Makrosolu	integer	
Sisätilasolu	integer	
Käyttöönottoon liittyvät tiedot solutasolla (tarkasteltavaa aikaväliä voidaan muuttaa)		
Käytössä (<= 1999)	integer	sisältää kaikki vuonna 1999 ja sitä ennen käyttöön otetut solut
Käytössä (<= 2000)	integer	sisältää kaikki vuonna 2000 ja sitä ennen käyttöön otetut solut
Käytössä (<= 2001)	integer	sisältää kaikki vuonna 2001 ja sitä ennen käyttöön otetut solut
Käytössä (<= 2002)	integer	sisältää kaikki vuonna 2002 ja sitä ennen käyttöön otetut solut
Käytössä (<= 2003)	integer	sisältää kaikki vuonna 2003 ja sitä ennen käyttöön otetut solut
Käytössä (<= 2004)	integer	sisältää kaikki vuonna 2004 ja sitä ennen käyttöön otetut solut
Käytössä (<= 2005)	integer	sisältää kaikki vuonna 2005 ja sitä ennen käyttöön otetut solut

Liite 4 Etenemismallit

1 1SM-malli

Peittoalueen laskennassa on tärkeää ottaa huomioon solun tyyppi, sillä se vaikuttaa oleellisesti käytettäviin laskentakaavoihin. Sisätilasoluilla käytetään 1SM-mallia (one-slope model)

$$L = L_0 + 10n \cdot \log(d) \Leftrightarrow d = 10^{\frac{L-L_0}{10n}}, \qquad (1)$$

missä L on vaimennus, L_0 on vaimennus 1 m etäisyydellä ja d on peittoalueen säde metreinä.

2 COST231 Walfisch-Ikegami -malli

Kaupunkialueilla, 800-2000 MHz:n taajuusalueella, voidaan käyttää empiiristä COST231 Walfisch-Ikegami -mallia. Malli jakaantuu kahteen eri tapaukseen sen mukaan, onko tukiaseman ja mobiilin välillä näköyhteyttä. Kun näköyhteyttä ei ole, kaava saa muodon

$$L_b = \begin{cases} L_0 + L_{rts} + L_{msd}, & L_{rts} + L_{msd} > 0 \\ L_0, & L_{rts} + L_{msd} \le 0 \end{cases}. \qquad (2)$$

Vapaan tilan vaimennus (free-space loss) L_0 saadaan kaavasta

$$L_0 = 32.4 + 20\log(d) + 20\log(f), \qquad (3)$$

missä d on peittoalueen säde kilometreinä ja f taajuus megahertseinä.

L_{rts} on katosta-kadulle diffraktio- ja sirontavaimennus (roof-top-to-street diffraction and scatter loss) ja se lasketaan kaavalla

$$L_{rts} = -8.2 - \log(w) + 10\log(f) + 20\log(h_{Roof} - h_{Mobile}) + L_{Ori}, \qquad (4)$$

missä w on tien leveys metreinä, f taajuus megahertseinä, h_{Roof} rakennusten korkeus metreinä ja h_{Mobile} mobiilin korkeus metreinä. L_{Ori} on empiirinen korjauskerroin, jonka arvot perustuvat mittaustuloksiin.

L_{msd} on monen esteen diffraktiovaimennus ja lasketaan kaavalla

$$L_{msd} = L_{bsh} + k_a + k_d \log(d) + k_f \log(f) - 9\log(b), \qquad (5)$$

missä L_{bsh} on tukiaseman korkeudesta riippuva korjaustermi, b on rakennusten välinen etäisyys metreinä ja k_a on tukiasema-antennien vaimennuksen kasvu rakennusten kattojen alapuolella. k_d ja k_f ovat monen esteen difraktiovaimennuksen korjauskertoimia, edellinen etäisyyden suhteen ja jälkimmäinen radiotaajuuden suhteen.

Siirtämällä peittoalueen säteen d sisältävä termi kaavan ulkopuolelle, saadaan

$$L_{msd} = L_{msd'} + k_d \log(d). \qquad (6)$$

Yhdistämällä kaavat saadaan säteeksi

$$d = 10^{\frac{L_b - L_{rts} - L_{msd'} - 32.4 - 20\log(f)}{20 + k_d}}. \qquad (7)$$

Kun tukiaseman ja mobiilin välillä on näköyhteys, COST231 Walfisch-Ikegami -malli on huomattavasti yksinkertaisempi

$$L_b = 42.6 + 26\log(d) + 20\log(f) \Leftrightarrow$$
$$d = 10^{\frac{L_b - 42.6 - 20\log(f)}{26}}. \qquad (8)$$

3 Okumura-Hata-malli

Taajuusalueella *f* = 150-1000 MHz etäisyys voidaan laskea Okumura-Hata-mallilla

$$L_b = 69.55 + 26.16\log(f) - 13.82\log(h_{Base}) - a(h_{Mobile}) + (44.9 - 6.55\log(h_{Base}))\log(d) \Leftrightarrow \quad \text{,} \qquad (9)$$
$$d = 10^{\frac{L_b - 69.55 - 26.16\log(f) + 13.82\log(h_{Base}) + a(h_{Mobile})}{44.9 - 6.55\log(h_{Base})}}$$

missä $a(h_{Mobile})$ on ympäristöstä riippuva korjaustermi mobiiliantennin korkeudelle. Mallia voidaan käyttää, jos tukiaseman korkeus h_{Base} on 30-200 m ja mobiilin korkeus h_{Mobile} on 1-10 m.

4 COST231-Hata-malli

Suuremmilla taajuuksilla, 1500-2000 MHz, käytetään COST231-Hata-mallia

$$L_b = 46.3 + 33.9\log(f) - 13.82\log(h_{Base}) - a(h_{Mobile}) + (44.9 - 6.55\log(h_{Base}))\log(d) + C_m \Leftrightarrow \quad \text{,} \qquad (10)$$
$$d = 10^{\frac{L_b - 46.3 - 33.9\log(f) + 13.82\log(h_{Base}) + a(h_{Mobile}) - C_m}{44.9 - 6.55\log(h_{Base})}}$$

missä C_m on 0 dB keskikokoisille kaupungeille ja esikaupunkien keskustoille ja 3 dB suurkaupunkien keskustoille. Malli soveltuu käytettäväksi kaupunkiympäristöissä, kun taajuus *f* on 1500-2000 MHz, tukiaseman korkeus h_{Base} on 30-200 m ja mobiilin korkeus h_{Mobile} on 1-10 m.

Liite 5 Tilastolliset menetelmät

Tässä liitteessä selostetaan lyhyesti käytetyt tilastolliset menetelmät, niiden ominaisuudet ja rajoitteet. Tarkemmat kuvaukset menetelmistä on esitetty tämän liitteen lähteissä olevassa materiaalissa (Flury, 1997; Greene, 2003; Grönroos, 2003; Hair et al., 1998; Krzanowski, 2000; Mustonen, 1995; Nummenmaa et al., 1997 ja Ranta et al., 2002).

1 Regressioanalyysi

Regressioanalyysin avulla tutkitaan yhden tai useamman selittävän muuttujan vaikutusta selitettävään muuttujaan. Sen avulla voidaan pyrkiä vastaamaan esimerkiksi siihen vaikuttaako suuruuteen ja jos vaikuttaa, niin kuinka voimakas tämä vaikutus on. Regressioanalyysin erityinen etu on, että siinä voidaan tutkia yhtä aikaa monen selittävän muuttujan vaikutusta selitettävään muuttujaan. Tällöin tuloksen kertovat, mikä on yksittäisen selittävän muuttujan osuus silloin kuin muiden vaikuttavien tekijöiden vaikutus selitettävään muuttujaan on otettu huomioon.

Regressioanalyysi on monipuolinen ja joustava menetelmä muuttujien välisten kausaalisuhteiden tutkimukseen. Sen edellytyksenä on, että selitettävä muuttuja on vähintään välimatka-asteikollinen. Selittävät muuttujat ovat yleensä myös vähintään välimatka-asteikollisia, mutta myös luokittelu- ja järjestysasteikollisia muuttujia voidaan sisällyttää analyysiin. Tällöin niistä tehdä ns. dummy-muuttujia.

Luvun 5.1 kuvioon 1 piirretyt viivat ovat ns. regressiokäyriä (regression line). Regressiosuora (tai yleisemmin –käyrä) osoittaa muuttujien välisen yhteyden voimakkuuden. Jos regressiosuora laskee alaspäin, on muuttujilla negatiivinen yhteys ja jos se nousee ylöspäin, on niillä positiivinen yhteys. Mitä lähempänä vaakatasoa suora on, sitä vähemmän muuttujilla on yhteyttä toisiinsa.

Regressiosuora voidaan esittää

$$y = b_0 + b_1 x_1 + \varepsilon,$$

missä y on selitettävän muuttujan arvo, b_0 ns. vakiotekijä, x selittävän muuttujan arvo, b_1 regressiokerroin (regression coefficient) ja ε virhe- tai jäännöstermi. Regressiokerroin on regressiosuoran kulmakerroin. Jos se saa negatiivisen (positiivisen) arvon, on suora laskeva (nouseva). Jos regressiokerroin on nolla, ei muuttujien välillä ole lineaarista yhteyttä. Vakiotekijä kertoo, minkä arvon selitettävä muuttuja saa silloin, kun selitettävän muuttujan x arvo on nolla, ts. missä kohtaa regressiosuora leikkaa kuvion y-akselin. Regressiokerroin kertoo, kuinka paljon selitettävä muuttuja muuttuu, kun selittävä muuttuja kasvaa yhden yksikön.

Edellisissä regressioanalyysin esimerkeissä oli vain yksi selittävä muuttuja. Regressioanalyysin etu on kuitenkin se, että siihen voi sisällyttää useita selittäviä muuttujia yhtäaikaisesti. Tällöin muuttujien regressiokertoimet kertovat, kuinka paljon selitettävän muuttujan arvo muuttuu, kun selittävän muuttujan arvo muuttuu yhdellä yksiköllä ja kaikkien muiden muuttujien arvo pysyy samana. Toisin sanoen usean muuttujan regressioanalyysissa regressiokertoimet ilmoittavat selittävän muuttujan vaikutuksen selitettävään muuttujaan niin, että muiden mallin muuttujien vaikutus on vakioitu.

Usean selittävän muuttujan regressioanalyysin kaava voidaan esittää

$$y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + \ldots b_k x_k + \varepsilon,$$

missä y on selitettävän muuttujan arvo, b_0 vakiotekijä, x_i selittävät muuttujat, b_i niiden regressiokertoimet ja k selittävien muuttujien lukumäärä. Malli voidaan esittää matriisimuodossa

$Y = Xb + \varepsilon$,

missä $Y = (y_1,y_2,y_3,...,y_n)'$ on selitettävän muuttujan havaintojen vektori, $X = (I,X_1,X_2,X_3,...X_k)$ on selittävien muuttujien arvojen matriisi, jossa ensimmäinen sarake I on ykkösten muodostama ja kukin $X_i = (x_{i1},x_{i2},x_{i3},...,x_{in})'$ on kyseisen muuttujan arvojen vektori eri havaintopisteissä, ja $b = (b_0,b_1,b_2,...,b_k)$ on regressiokertoimien vektrori.

Mallin standardioletukset ovat

(i) selittävien muuttujien X arvot kiinteitä eli ei-satunnaisia

(ii) selittäjien X välillä ei ole lineaarisia riippuvuuksia

(iii) $E(\varepsilon_j) = 0$

(iv) $Var(\varepsilon_j) = \sigma^2$ (vakiovarianssi eli homoskedastisuus)

(v) $Cor(\varepsilon_i,\varepsilon_j) = \sigma^2$

Lisäksi virhetermeistä ε tehdään tavallisesti jakauman normaalisuusoletus

(vi) $\varepsilon_j \sim N(0,\sigma^2)$.

Jäljempänä selostetaan ongelmia ja niiden ratkaisumahdollisuuksia silloin, kun nämä oletukset eivät ole voimassa.

Regressiomallin pätevyyttä voidaan arvioida sen mukaan, kuinka lähelle kuvion pisteet sijoittuvat regressiosuoraa. Mitä lähempänä suoraa ne sijaitsevat, sitä parempi on regressiomallin selitysvoima ja päinvastoin. Jos kuvion pisteen sijoittuvat hyvin lähelle suoraa, on mallilla hyvä selitysvoima, koska sen avulla voidaan hyvin tarkasti arvioida, mikä on jonkin yksittäisen maan lukutaidottomuusprosentti silloin, kun tiedetään kuinka paljon maassa sijoitetaan koulutukseen. Mitä kauempana pisteet suorasta sijaitsevat, sitä epävarmempia ovat ennusteet.

Yksittäisen havainnon arvon etäisyyttä regressiosuorasta kutsutaan havainnon virhetermiksi tai residuaaliksi (residual). Mitä suuremmat mallin virhetermit itseisarvoltaan ovat, sitä huonompi ennustearvo regressiomallilla on ja päinvastoin.

Regressiomallin tavanomaisin estimointitapa on ns pienimmän neliösumman menetelmä (pns; ordinary least squares, OLS). Pns:ssä regressiosuora asetetaan niin, että virhetermien neliöiden summa on pienin mahdollinen.

Regressioanalyysin tulosten tulkinta

Regressioanalyysin yhteydessä testataan jokaisen selittävän muuttujan osalta onko niillä vaikutusta selitettävään muuttujaan eli eroavatko ne tilastollisesti merkitsevästi nollasta. Tällaiseen tarkoitukseen sopiva testimenetelmä on t-testi. Testin tuloksena jokaiselle selittävälle muuttujalle saadaan t-arvo, jonka suuruus ratkaisee sen, voidaanko muuttujan kerrointa pitää nollaa suurempana tilastollisten kriteerien mukaan. Taulukon viimeisessä sarakkeessa on esitetty t-testien merkitsevyystasot. Jos t-testin luku on "riittävän lähellä" nollaa, se tarkoittaa, ettei muuttuja ole tilastollisesti merkitsevä, toisin sanoen ettei muuttujan vaihtelu selitä riippuvan muuttujan vaihtelua (se, onko luku riittävän lähellä nollaa, selviää t-testin taulukosta).

Tärkeimmät regressioanalyysin selitysvoimaa kuvaavat testit ovat R^2-luku ja F-testi. R^2-luku on regressiomallin selitysosuus. Se kertoo kuinka suuren osuuden selitettävän muuttujan vaihtelusta regressionanalyysin selittävät muuttujat pystyvät selittämään. R^2-luku vaihtelee nollan ja yhden välillä. Se saadaan laskemalla selitettävän muuttujan arvojen ja mallin tuottamien ennustearvojen korrelaation neliö. Jos R^2-luku on pieni regression selittävät muuttujan pystyvät selittämään vain vähän selitettävän muuttujan vaihtelusta ja päinvastoin. Esimerkiksi tutkimusta varten laaditun regressiomallin ”Kysyntäyhtälön estimointitulokset aggregaattiaineistolla” R^2 on 0.918. Tämä tarkoittaa, että malli pystyy erittäin hyvin selittämään vaihtelua puheluminuuttien vaihtelua. Regressiomallin avulla 91.8 % vaihtelusta voidaan selittää mallin avulla. Korjattua R^2-lukua (adjusted R^2) käytetään silloin, kun halutaan verrata kahden regressioanalyysin tuloksia keskenään. Korjattu R^2-luku ottaa huomioon mallin sisältämien selittävien muuttujien lukumäärän. Se on arvoltaan aina pienempi tai yhtä suuri kuin varsinainen R^2-luku. Korjaus R^2-lukuun tarvitaan sen vuoksi, että uusien selittävien muuttujien lisääminen regressioanalyysiin nostaa aina R^2-lukua, vaikka nämä lisätyt muuttujat eivät todellisuudessa pystyisikään lisäämään selityskykyä. Silloin kun tarkasteltavana on vain yksi regressiomalli, ei korjatun R^2-luvun käyttäminen ole tarpeellista, mutta regressiomalleja verratessa siitä on hyötyä.

F-testi on tilastollinen testi, joka kertoo pystytäänkö regressioanalyysissa olevilla muuttujilla ylipäänsä selittämään selitettävän muuttujan vaihtelua. Koska se on tilastollinen testi, saadaan sille myös merkitsevyystaso. Taulukossa F-testin tulos on erittäin merkitsevä. Tämä ei sinänsä ole yllätys, koska myös selittävien muuttujien regressiokertoimet ovat tilastollisesti merkitseviä.

Estimaatin keskivirhe (standard error of estimate) ilmoittaa regressiomallin virhetermien keskihajonnan. Mitä suurempi se on, sitä suurempi on virhetermien hajonta ja samalla sitä pienempi mallin selitysvoima. Estimaatin keskivirheen suuruus riippuu aina regressiomallin hyvyyden lisäksi selitettävän muuttujan mittaluokasta.

Dummy-muuttujat

Dummy-muuttujaksi kutsutaan sellaista muuttujaa, joka voi saada vain kaksi eri arvoa, jotka on koodattu nollaksi ja yhdeksi. Tyyppiesimerkki tällaisesta muuttujasta on vastaajan sukupuoli, mutta vaihtoehtoja on helppo keksiä lisää. Dummy-muuttujien avulla regressioanalyysiin voidaan helposti sisällyttää luokittelu- tai järjestysasteikollisia muuttujia.

Lineaarisuusoletus

Regressioanalyysi on joustavuudessaan erinomainen menetelmä muuttujien riippuvuussuhteiden tarkasteluun. Siihen liittyy kuitenkin rajoitteita, joista on syytä olla tietoinen. Tässä yhteydessä rajoitteet esitellään vain lyhyesti. Regressioanalyysi tarjoaa myös monia mahdollisia tapoja ottaa rajoitteet huomioon ja korjata niiden vaikutukset regressioanalyysissa.

Regressioanalyysin avulla voidaan tutkia muuttujien välisiä lineaarisia eli suoraviivaisia kausaalisuhteita. Jos regressioanalyysin tulokset osoittavat, että selittävällä muuttujalla ei ole tilastollisesti merkitsevää yhteyttä selitettävään muuttujaan, tarkoittaa tämä tarkasti ottaen ainoastaan sitä, ettei lineaarista yhteyttä esiinny. Muuttujilla voi kuitenkin olla epälineaarinen yhteys. Regressioanalyysin avulla voi kuitenkin tarkastella myös muuttujien epälineaarisia suhteita. Tämä tapahtuu muuttujien muunnosten avulla. Muunnoksen kohteena voi olla sekä selitettävä ja selittävät muuttujat tilanteen mukaan. Lievien epälineaarisuuksien korjaamiseen käytetään logaritmi- tai potenssimuunnosta.

Usein käytetty muunnos on estimoida malli log-lineaarisessa muodossa

$$\ln y = b_0 + b_1 \ln x_1 + b_2 \ln x_2.$$

Tällöin regressiokertoimet b_i merkitsevät joustoja: esimerkiksi b_1 kertoo, että jos muuttujan x_1 arvo muuttuu yhden prosentin, y:n arvo muuttuu b_1 prosenttia. Tämäkin malli on lineaari; olennaista on se, onko malli lineaari tai onko se muunnettavissa lineaariksi parametrien b suhteen.

Poikkeavat havainnot (outliers)

Joskus yksittäisillä poikkeavilla havainnoilla voi olla suuri vaikutus regressioanalyysiin tuloksiin. Tällaisia havaintoja kutsutaan niiden englanninkielisen nimen mukaan outlier-tapauksiksi.

Multikollineaarisuus

Regressioanalyysissa on aivan luonnollista, että selittävät muuttujat korreloivat keskenään. Joskus niiden keskinäinen korrelaatio voi kuitenkin olla niin suuri, että se aiheuttaa ongelmia regressioanalyysin tulosten tarkkuuden kannalta. Tällaista tilannetta kutsutaan multikollineaarisuudeksi. Yleensä multikollineaarisuusongelmia ei synny, jollei selittävien muuttujien välillä ole todella suuria riippuvuuksia (esimerkiksi korrelaatiokerroin yli 0,9). Ongelmana on, että kaikkia multikollineaarisuusongelmia ei voi havaita tarkastelemalla pelkästään selittävien muuttujien välisiä korrelaatiokertoimia. Tämän vuoksi on kehitetty erilaisia multikollineaarisuusmittareita, jotka ilmaisevat ongelman mahdollisen vakavuuden (esimerkiksi VIF-mittari).

Heteroskedastisuus

Heteroskedastisuus viittaa tilanteeseen, jossa regressiomallin virhetermien hajonta vaihtelee suuresti ja systemaattisesti x-muuttujien arvojen muuttuessa. Heteroskedastisuudella ei oikeastaan ole haitallisesta vaikutusta regressiokertoimien arvoon. Sen sijaan sillä voi olla vaikutusta niiden tilastolliseen merkitsevyyteen. Tämä voi johtaa esimerkiksi tilanteeseen, jossa tietty muuttuja ei näytä olevan tilastollisesti merkitsevä Y:n selittäjä vaikka se todellisuudessa sellainen onkin. Heteroskedastisuusongelmien havainnoimiseksi on kehitetty erilaisia testejä, joita ei kuitenkaan esitellä tässä yhteydessä. Jos testit osoittavat, että aineistossa on heteroskedastisuutta, voidaan regressioanalyysin tulosten estimointiin käyttää sellaista menetelmää, joka pystyy ottamaan huomioon nämä ongelmat.

Havaintojen aikariippuvuus

Yksi regressioanalyysin perusolettamuksista on, että havaintojen virhetermit ovat toisistaan riippumattomia. Jos analysoitavana on aikasarja-aineisto, tämä oletus ei useinkaan ole pätevä. Tämä johtuu siitä, että eri ajankohtina kerättyjen havaintojen virhetermit korreloivat keskenään. Jos analysoitavana on esimerkiksi eri vuosina, on tietyn vuoden osittain riippuvainen edellisen vuoden tasosta. Jos tätä riippuvuutta ei oteta huomioon, regressioanalyysin tulokset vääristyvät. Havaintojen aikariippuvuuden korjaamiseksi on useita eri tapoja. Käyttämiemme testien perusteella aineistossa aikariippuvuus ei ole ongelma.

Instrumenttimenetelmä

Eräs edellä selostetun yleisen linearisen mallin Y = Xb keskeinen oletus on, että mallin selittäjät $X = (X_1, X_2, X_3, \ldots X_k)$, $X_i = (x_{i1}, x_{i2}, x_{i3}, \ldots, x_{in})'$ ovat kiinteitä eli ei-satunnaisia muuttujia. Mutta jos matriisi X on satunnainen, esimerkiksi koska kyseessä on markkinadata, joka perustuu kk-keskiarvoihin, ei pns-menetelmä välttämättä tuota harhattomia tai edes tarkentuvia estimaattoreita regressiokertoimille (estimaattorit muuttuvat tarkemmiksi datan määrän kasvaessa). Näin käy esimerkiksi silloin, kun virhetermi ja selittäjät korreloivat. Jos regressiokertoimien pns-estimaattorit eivät ole harhattomia tai tarkentuvia, mallia koskevaa tavanomaista tilastollista päättelyä ei voida soveltaa. Jos yleisen lineaarisen mallin selittäjät X ovat satunnaismuuttujia, mallia koskevia standardioletuksia täytyy hieman muuttaa.

Olkoon w_{ji} (i = 1,2, …,k; j = 1,2,…,n) muuttujan w_i havaittu arvo havaintoyksikössä j. Muodostetaan havaituista arvoista w_{ji} matriisi W, jossa ensimmäinen sarake on ykkösten muodostama ja jonka dimensio on siis $n \times (k+1)$. Oletetaan, että matriisilla W on seuraavat ominaisuudet:

(i) $E(\varepsilon | W) = 0$

(ii) $r(W'X) = k + 1$

Ehdon (i) mukaan instrumenttimuuttujat $w_1, w_2, \ldots, w_k$ eivät saa korreloida mallin virhetermin ε kanssa. Ehdon (ii) mukaan instrumenttimuuttujien $w_1, w_2, \ldots, w_k$ pitää korreloida selittäjien $x_1, x_2, \ldots, x_k$ kanssa niin voimakkaasti, että matriisi $W'X$ on ns. epäsingulaarinen. Tällöin muuttujat $w_1, w_2, \ldots, w_k$ kelpaavat instrumenteiksi selittäjille $x_1, x_2, \ldots, x_k$. Muuttujia $w_1, w_2, \ldots, w_k$ kutsutaan tavallisesti instrumentti- tai välinemuuttujiksi. Instrumenttiestimaattori $b^{IV} = (W'X)^{-1}W'Y$ on silloin regressiokertoimien vektorin tarkentuva estimaattori.

Simultaanisuus ja 3SLS

Toisinaan meillä on kaksi tai useampia selitettäviä muuttujia, jotka määräytyvät yhtaikaisesti. Mallina tämä voidaan esittää

(1) $q = b_0 + b_1 p + b_2 x_2 + b_3 x_3 + ... b_k x_k$

(2) $p = b_0 + b_1 q + b_2 z_2 + b_3 z_3 + ... b_l z_l$

ts yhtälön (1) selitettävä muuttuja q on yhtälön (2) selittävänä muuttujana ja yhtälön (2) selitettävä muuttuja p on yhtälön (1) selittävänä muuttujana. Yhtälö (1) voi kuvata esimerkiksi matkapuheluiden kysyntää ja yhtälö (2) niiden tarjontaa, jolloin q on esimerkiksi asiakkaiden ostama määrä (minuutteja) ja p palvelun yksikköhinta. Markkinoilla menekki (minuutit) ja yksikköhinta määräytyvät yhtaikaisesti kysynnän, tarjonnan ja kilpailukeinojen vuorovaikutuksessa. Tällöin kunkin yhtälön estimoiminen erikseen voi johtaa ns simultaanisuusharhaan, kun esimerkiksi yhtälössä (1) q vaikuttaa p:hen, vaikka malli olettaa p:n arvon olevan annettu ja erityisesti q:stä riippumaton vakio. Simultaanimallien ongelma on, että mallin selittäjät ja virhetermit korreloivat keskenään. Jos simultaanisuusharhaa esiintyy, yhtälöt täytyy estimoida yhtäaikaisesti (simultaanisesti).

Simultaaniyhtälöiden estimoinnissa käytetään kahden tai kolmen vaiheen pns–menetelmää (2SLS, 3SLS) tai suurimman uskottavuuden menetelmää (maximum likelihood).

Hierarkkinen regressioanalyysi

Kun etukäteen ei ole tiedossa, mitkä joukosta potentiaalisia riippumattomia muuttujia vaikuttavat suoraan riippuvaan muuttujaan ja minkä muuttujien vaikutuksen nämä suoraan vaikuttavat muuttujat välittävät, käytetään hierarkkista regressioanalyysiä sellaisen mallin määrittämiseen, jossa sovitussuhteet ovat tiedossa. Analyysissä katsotaan, että muuttuja a välittää muuttuja b:n vaikutuksen, jos b:n ja riippuvan muuttujan välinen suhteen merkittävyys pienenee, kun a lisätään regressiomalliin. Kun mallissa ovat sekä a että b, a on tilastollisesti merkittävästi vaikuttamassa riippuvaan muuttujaan. Laskennallisesti välitysvaikutuksen olemassaolo mitataan Sobelin testillä, joka formalisoi yllä esitetyn periaatteen. Muuttuja, joka muuttuu tilastollisesti ei-merkittäväksi, kun välittävä muuttuja lisätään malliin, voidaan joko säilyttää kokonaismallissa tai poistaa siitä. Jos kokonaisanalyysissä käsitellään malleja, joissa ei-merkittäväksi yhdessä mallissa muuttuva muuttuja säilyy merkittävänä, on yleensä perusteltua säilyttää ei-merkittävä muuttuja mallissa kontrollimuuttujana.

2 Pääkomponenttianalyysi

Jos tutkimuksen kohteena olevaa satunnaisilmiötä kuvaa hyvin suuri joukko erilaisia muuttujia tai tekijöitä, johtopäätösten tekeminen saattaa olla hyvin hankalaa. Ongelma voidaan ratkaista, jos alkuperäiset muuttujat voidaan korvata pienellä joukolla uusia, keinotekoisia muuttujia, joiden kautta tutkimuksen kohteena oleva ilmiö on helpommin ymmärrettävissä kuin alkuperäisten muuttujien kautta. Tällöin vaatimuksena uusille muuttujille on se, että niiden pitää olla tulkittavissa ja se, että niiden on säilytettävä mahdollisimman tarkkaan alkuperäisten muuttujien olennaiset piirteet. Koska informaatio tutkimuksen kohteista sisältyy kohteita kuvaavien muuttujien arvojen vaihteluun, uusia muuttujia muodostettaessa on järkevää ottaa kriteeriksi se, että uusien muuttujien on säilytettävä mahdollisimman suuri osa alkuperäisiin muuttujiin sisältyvästä vaihtelusta.

Pääkomponenttianalyysi on tilastollinen menetelmä, jossa alkuperäiset muuttujat pyritään korvaamaan pienellä määrällä uusia, keinotekoisia muuttujia, jotka säilyttävät mahdollisimman suuren osan alkuperäisten muuttujien vaihtelusta. Pääkomponenttianalyysin takana ei ole mitään tilastollista mallia. Tämä merkitsee sitä, että pääkomponenttianalyysissa havainnoille ei määritellä todennäköisyysjakaumaa, jossa pääkomponenttianalyysissa määrättävät suureet, kuten esimerkiksi pääkomponenttimatriisin alkiot, olisivat parametreina. Pääkomponenttianalyysia on lähinnä heuristinen menetelmä, vaikka se onkin osoittautunut varsin hyödylliseksi apukeinoksi moniulotteisia aineistoja analysoitaessa.

Ajatellaan, että mielenkiinnon kohteena olevaa ilmiötä kuvaa satunnaisvektori $x = (x_1, x_2, x_3,..., x_p)$, jonka kovarianssi on $Cov(x) = \Sigma$. Etsitään muuttujien x sellainen lineaarinen yhdistely, joka sisältää mahdollisimman suuren osan satunnaismuuttujien x_1, x_2, …, x_p vaihtelusta. Siten tehtävänä on etsiä satunnaisvektorin x alkioiden lineaarikombinaatio $\beta' x$, jonka varianssi on suurin.

Tehtävä ratkaistaan maksimoimalla lineaarikombinaation $\sum_{i=1}^{p} \beta_i x_i$ varianssi $D^2(\beta' x)$ ehdolla, että $|\beta|^2 = \beta'\beta = 1$. Vektori β on painovektori, joka kertoo millä painolla kukin muuttujista x_j vaikuttaa lineaarikombinaation $\beta' x$ varianssiin. Ehto $\beta'\beta = 1$ on normeerausehto.

Voidaan osoittaa, että maksimi on

$$\beta_1 \Sigma \beta_1' = \lambda_1,$$

jossa λ_1 = kovarianssimatriisin Σ suurin ominaisarvo ja β_1 = kovarianssimatriisin Σ suurinta ominaisarvoa vastaava ominaisvektori. Merkitään $y_1 = \beta_1 x$ on ensimmäinen pääkomponentti.

Etsitään seuraavaksi satunnaismuuttujien x sellainen lineaarinen yhdistely, joka on korreloimaton 1. pääkomponentin y_1 kanssa ja joka sisältää mahdollisimman suuren osan satunnaismuuttujien x vaihtelusta. On siis etsittävä satunnaisvektorin x alkioiden lineaarikombinaatio $\beta' x$ joka on korreloimaton 1. pääkomponentin y1 kanssa ja jonka varianssi on suurin. Ratkaisu saadaan maksimoimalla lineaarikombinaation $\sum_{i=1}^{p} \beta_i x_i$ varianssi $D^2(\beta' x)$ ehdoilla $Cov(y_1, \beta' x) = 0$ ja $|\beta|^2 = \beta'\beta = 1$.

Voidaan osoittaa, että maksimi on

$$\beta_2 \Sigma \beta_2' = \lambda_2,$$

jossa λ_2 = kovarianssimatriisin Σ toiseksi suurin ominaisarvo ja β_2 = kovarianssimatriisin Σ toiseksi suurinta ominaisarvoa vastaava ominaisvektori. Merkitään $y_2 = \beta_2 x$ on toinen pääkomponentti.

Jatkamalla kuvattua menettelyä saadaan määrätyksi vektorin x alkioiden kaikki lineaarikombinaatiot, (i) joiden varianssi on suurin mahdollinen ja (ii) joka on korreloimaton aikaisemmin määrättyjen lineaarikombinaatioiden kanssa. Näin löydetyt lineaarikombinaatiot muodostavat satunnaisvektorin x pääkomponentit.

Pääkomponentit saadaan siis kovarianssimatriisin Σ ns. pääakselihajotelmasta $\Sigma = B\Lambda B'$, jossa $\Lambda = diag(\lambda_1, \lambda_2,..., \lambda_p)$ on matriisin Σ ominaisarvojen $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_p$ muodostama diagonaalimatriisi ja matriisi B on vastaavien ominaisvektoreiden (matriisin B sarakkeina) muodostama ortogonaalinen matriisi $B'B = BB' = I$. Pääkomponentit saadaan satunnaisvektorin x lineaarikombinaatioina $y_i = \beta_i' x$ = i. pääkomponentti, i = 1,2, … , p.

3 Faktorianalyysi

Sekä pääkomponentti- että faktorianalyysi ovat dimension redusointitekniikoita, jossa päämääränä on korvata alkuperäiset muuttujat pienemmällä määrällä piilomuuttujia. Pääkomponenttianalyysin takana ei ole mitään tilastollista mallia. Tämä merkitsee sitä, että pääkomponenttianalyysissa havainnoille ei määritellä todennäköisyysjakaumaa, jossa pääkomponenttianalyysissa määrättävät suureet, kuten esimerkiksi pääkomponenttimatriisin alkiot, olisivat parametreina.

Pääkomponenttianalyysia voidaankin pitää lähinnä heuristisena menetelmänä, vaikka se onkin osoittautunut varsin hyödylliseksi apukeinoksi moniulotteisia aineistoja analysoitaessa. Tarkastellaan edelleen sellaista tilannetta, jossa tutkimuksen kohteena oleva satunnaisilmiötä voidaan kuvata suurella joukolla havaittavia muuttujia, joiden taustalla oletetaan olevan pieni joukko ei-havaittavia muuttujia, jotka generoivat havaittujen muuttujien arvot. Faktorianalyysi on yritys löytää ja tulkita tällaiset ei-havaittavat, piilevät muuttujat eli faktorit.

Faktorianalyysi on esimerkki ns. piilomuuttujamalleista, joissa havaittavien muuttujien takana oletetaan olevan ei-havaittavia muuttujia. Faktorianalyysi on menetelmänä läheistä sukua pääkomponenttianalyysille, mutta eroaa siitä kuitenkin siinä, että faktorianalyysi perustuu aitoon tilastollisen malliin.

Oletetaan, että havaittu satunnaismuuttuja $x = (x_1, x_2, \ldots, x_p)$ voidaan esittää ei-havaittavien satunnaismuuttujien $f = (f_1, f_2, \ldots, f_r)$ ja $u = (u_1, u_1, \ldots, u_r)$ avulla

$$x_i = \mu_i + a_{i1}f_1 + a_{i2}f_2 + \cdots + a_{ir}f_r + u_i \text{ , } i = 1, 2, \ldots, p, r \leq p$$

Matriisimerkinnöin nämä yhtälöt voidaan esittää seuraavassa muodossa:

(1) $x = \mu + Af + u$

jossa $\mu = (\mu_1, \mu_2, \ldots, \mu_p)$ ja A on $p\times r$-matriisi, jossa $r \leq p$. Tavallisesti toivotaan, että r on paljon pienempi kuin p. Mallia (1) kutsutaan satunnaismuuttujan x faktorimalliksi.

Faktorimallista tehdään yleensä seuraavan tyyppiset oletukset:

(i) $E(x) = \mu$, $Cov(x) = \Sigma$

(ii) $E(f) = 0$, $Cov(f) = \Phi$

(iii) $E(u) = 0$, $Cov(u) = \Psi = diag(\psi_1^2, \psi_2^2, \ldots\psi_p^2)$

(iv) u ja f ovat riippumattomia.

(v) $r(A) = r \leq p$

Lisäksi tavallisesti tehdään oletus faktoreiden ortogonaalisuudesta:

(vi) $Cov(f) = I$

Faktorimallin osista käytetään seuraavia nimityksiä:

$f_j = [f]_j$ = j. yhteisfaktori, $j = 1, 2, \ldots , r$

$u_i = [u]_i$ = i. ominaisfaktori, $i = 1, 2, \ldots , p$

$ai_j = [A]_{ij}$ = i. muuttujan lataus j. faktorille, $i = 1, 2, \ldots , p$, $j = 1, 2, \ldots , r$

A on $p\times r$ faktori- tai latausmatriisi

Faktorimallin mukaan havaittavat muuttujat x_i, $i = 1, 2, \ldots , p$ riippuvat samoista, ei-havaittavista yhteisfaktoreista f_j, $j = 1, 2, \ldots , r$, jossa $r \leq p$. Lisäksi jokainen havaittava muuttuja xi riippuu myös ei-havaittavasta ominaisfaktorista u_i.

Faktorimalli voidaan rinnastaa useampia selitettäviä muuttujia sisältävään monimuuttujaiseen lineaariseen regressiomalliin seuraavalla tavalla:

x_i ~ i. mallin selitettävä muuttuja, i = 1, 2, …, p

f_j ~ j. selittäjä, j = 1, 2, …, r

a_{ij} ~ j. selittäjän regressiokerroin i. mallissa, i = 1, 2, …, p, j = 1, 2, …, r

u_i ~ i. mallin virhetermi, i = 1, 2, …, p

Faktorimallissa jokaista ”selitettävää” muuttujaa x_i selitetään samoilla ”selittäjillä” f_j. Faktorimallissa – toisin kuin regressiomallissa – sekä faktoreiden lukumäärä r, ”regressiokertoimet” a_{ij} että selittäjien f_j ”arvot” ovat tuntemattomia. Niiden määräämistä kutsutaan faktoroinniksi.

Faktorimatriisissa A pyritään tulkinnan helpottamiseksi tavallisesti ns. yksinkertaiseen rakenteeseen. Faktorimatriisin A rakenne on yksinkertainen, jos se sisältää kullakin rivillä vain muutaman nollasta ”poikkeavan” latauksen ja mahdollisimman monta ”lähellä” nollaa olevaa latausta. Yksinkertaiseen rakenteeseen pyritään rotatoimalla eli kiertämällä faktoreita sopivien ehtojen vallitessa. Rotatointi saadaan aikaan sopivasti valitulla ortogonaalisella matriisilla T.

Voidaan osoittaa, että ehtojen (i) – (v) pätiessä

$$\mathrm{Cov}(x) = \Sigma = A\Phi A' + \Psi.$$

Tätä yhtälöä kutsutaan tavallisesti faktorianalyysin perusyhtälöksi. Yhtälön mukaan kovarianssit

$$\sigma_{ij} = \mathrm{Cov}(x_i, x_j) = [\Sigma]_{ij}$$

voidaan esittää matriisien A, Φja Ψ alkioiden a_{ij}, φ_i ja ψ_i funktioina. Edelleen

$$\mathrm{Cov}(x, f) = A\Phi.$$

Jos faktoreiden ortogonaalisuusehto (6) pätee,

$$\mathrm{Cov}(x) = \Sigma = AA' + \Psi$$

ja

$$\mathrm{Cov}(x, f) = A.$$

Ehdon (vi) pätiessä latausmatriisi A sisältää siis havaittavien muuttujien x_i ja faktoreiden f_j kovarianssit. Faktorilatauksia käytetään faktoreiden tulkinnassa apuna samalla tavalla kuin pääkomponenttimatriisin alkioita pääkomponenttianalyysissa.

Havaittavan muuttujan x_i varianssi voidaan esittää seuraavassa muodossa:

$$D^2(x_i) = \sigma_{ii} = a_{i1}^2 + a_{i2}^2 + \ldots + a_{ir}^2 + \psi_i^2 = h_i^2 + \psi_i^2$$

jossa lauseketta h kutsutaan muuttujan i kommunaliteetiksi ja suuretta ψ_i^2 kutsutaan muuttujan i ominaisvarianssiksi. Kommunaliteetti kuvaa muuttujan i varianssin systemaattista, faktoreilla f_j selitettävää osuutta.

Faktorointi tehdään tavallisesti korrelaatiomatriisista. Tällöin kovarianssimatriisin Σ paikalle on yo. kaavoissa asetettava korrelaatiomatriisi P. Jos faktorointi on tehty korrelaatiomatriisista P, latausmatriisi A sisältää havaittavien muuttujien x_i ja faktoreiden f_j korrelaatiot

$$[A]_{ij} = a_{ij} = Cor(x_i, f_j),\ i = 1, 2, \ldots, p;\ j = 1, 2, \ldots, r$$

Tällöin faktorianalyysin perusyhtälö voidaan esittää muodossa

$$Cor(x) = P = AA' + \Psi$$

ja muuttujan i kommunaliteetilla h_i^2 on seuraava ominaisuus:

$$h_i^2 = 1 - \psi_i^2 < 1.$$

Satunnaismuuttujaa

$$y = x - \mu - u = Af$$

kutsutaan muuttujan x systemaattiseksi osaksi. Muuttujan x systemaattisella osalla y on seuraavat ominaisuudet:

$$E(y) = 0$$

$$Cov(y) = AA' = Cov(x, y)$$

Jos faktorointi on tehty korrelaatiomatriisista P,

$$Cor(x_i, y_i) = h_i$$

Faktorianalyysissa on tapana puhua r-ulotteisesta faktoriavaruudesta, jonka virittävät (ortogonaaliset) faktorit $f_1, f_2, \ldots, f_r$. Muuttujan x systemaattisen osan y komponentit y_i voidaan yhtälön y = Af perusteella esittää pisteinä $a_i = (a_{i1}, a_{i2}, \ldots, a_{ir})$, i = 1, 2, …, p, tässä avaruudessa. Vektorin a_i pituus saadaan yhtälöstä $|a_i|^2 = a_i'a_i = h_i^2$. Silloin vektoreiden a_i ja a_j välisen kulman φ_{ij} kosini on $\cos(\varphi_{ij}) = \cos(a_i,a_j) = Cor(a_i,a_j) = a_i'a_i/h_ih_j = \rho_{ij}/h_ih_j$, missä $\rho_{ij} = Cor(x_i,x_j)$. Oikean faktoriluvun määrääminen on olennaisen tärkeää.

Mittarin reliabiliteetti

Mittarin reliabiliteetti eli luotettavuus, käyttö- tai toimintavarmuus viittaa mittarin johdonmukaisuuteen; siihen, että se mittaa kokonaisuudessaan samaa asiaa. Jos mittari on täysin reliaabeli, siihen eivät vaikuta satunnaisvirheet eivätkä olosuhteet.

Reliabiliteetissa voidaan erottaa kaksi osatekijää: stabiliteetti ja konsistenssi. Stabiliteetissa on kysymys mittarin pysyvyydestä ajassa. Epästabiilissa mittarissa näkyvät olosuhteiden ja muiden satunnaisvirheiden (esim. vastaajan mielialan) vaikutukset helposti. Mittarin konsistenssilla eli yhtenäisyydellä tarkoitetaan sitä, että kun useista väittämistä koostuva mittari jaetaan kahteen joukkoon väittämiä, kumpikin väittämäjoukko mittaa samaa asiaa. Tällöin molempien väittämäjoukkojen kokonaispistemäärien välinen korrelaatiokerroin saa suuren arvon.

Paljon käytetty tunnusluku reliabiliteetin mittaamiseksi on Cronbachin alfa. Sillä mitataan nimenomaan mittarin konsistenssia eli sisäistä yhtenäisyyttä. Cronbachin alfa lasketaan muuttujien välisten keskimääräisten korrelaatioiden ja väittämien lukumäärän perusteella. Mitä suurempi alfan arvo on, sitä yhtenäisempi mittarin voidaan katsoa olevan. Reliaabeliutta osoittamaan voidaan laskea alfa-kerroin myös käyttäen ns. puolitusmenetelmää (Split-Half), jolloin muuttujat jaetaan kahteen ryhmään ja alfa-kertoimet lasketaan kummallekin osiolle. Cronbachin alfa soveltuu yhdistettyjen muuttujien (ns. summamuuttujien) ja useita osioita sisältävien testien sisäisen yhdenmukaisuuden tarkasteluun. Alfan standardoitu estimaatti lasketaan kaavalla alfa = k*r/(1+(k-1)*r), jossa r = väittämien välisten Pearsonin korrelaatiokertoimien keskiarvo ja k = väittämien lukumäärä. Alfa saa arvoja [0-1] -väliltä. Mitä lähempänä 1:tä kerroin on, sitä korkeampi reliabiliteetti.

Lähteet

Flury, B. (1997), *A First Course in Multivariate Statistics*, Springer-Verlag, New York.

Greene, W. (2003), *Econometric Analysis*, 5th ed., Prentice Hall.

Grönroos, M. (2003), Johdatus tilastotieteeseen - kuvailu, mallit ja päättely, Finn Lectura.

Hair, J.F., R.E. Anderson, R.L. Tatham and W. Black (1998), *Multivariate Data Analysis*, 5th ed., Prentice Hall.

Krzanowski, W.J. (2000), *Principles of Multivariate Analysis* (revised edition), Oxford University Press.

Mustonen, S. (1995), *Tilastolliset monimuuttujamenetelmät*, Survo Systems, Helsinki

Nummenmaa, T., R. Konttinen, J. Kuusinen ja E. Leskinen (1997), *Tutkimusaineiston analyysi*. WSOY.

Ranta, E., H. Rita ja J. Kouki (2002), *Biometria*, 8. painos, Yliopistopaino.

Liite 6 Ekonometrinen mallintaminen

Matkapuhelinpalveluiden kysyntää ja tarjontaa mallinnetaan ekonometrisesti seuraavalla mallilla:

$$\log(Q_{it}) = a_0 + a_1 P_{it} + a_2 L_{it} + a_3 X_{1it} + \varepsilon_{it} \qquad 5.1$$

$$\log(P_{it}) = b_0 + b_1 MC_{it} + b_2 X_{2it} + \nu_{it}, \qquad 5.2$$

jossa Q_{it} = yrityksen i matkaviestinpalveluiden kysyntä ajanhetkellä t, P_{it} = yrityksen i matkapuheluiden hinta ajanhetkellä t, L_{it} = Yrityksen tarjoamien matkaviestinpalveluiden laatu funktiona investoinneista ja allokoiduista radiotaajuuksista, MC_{it} = yrityksen i matkaviestinpalveluiden tarjonnan rajakustannukset ajanhetkellä t, X_{1it} = vektori muista matkapuheluiden kysyntään vaikuttavista tekijöistä, X_{2it} = vektori muista yrityksen tarjontaan vaikuttavista tekijöistä, ja ε_{it} ja ν_{it} ovat virhetermejä.

Yhtälöiden 5.1 ja 5.2 muodostama malli estimoidaan simultaanisesti käyttäen 3SLS-instrumenttimuuttujamenetelmää (3SLS IV –menetelmä eli three stage least squares instrumental variable method on esitetty tämän liitteen lähteenä olevassa kirjassa Ameniya, 1985), joka sallii mallin kaikkien parametrien estimoinnin yhtäaikaisesti ja lisäksi selittävien *endogeenisten* muuttujien mukaan lukemisen malliin. Tässä tutkimuksessa 3SLS IV-mallin etuna on se, että malli sallii tarjonta- ja kysyntäyhtälöiden virhetermien keskinäisen korrelaation ja hintamuuttujan kohtelemisen endogeenisena myös kysyntäyhtälössä.

Kysyntää, Q_{it}, mitataan yrityksen verkosta soitetuilla matkapuhelinminuuteilla per liittymä (muuttuja KYS) ja yrityksen matkapuheluiden hintaa, P_{it}, arvioidaan sen liittymistä ja puheluista saamilla tuloilla jaettuna sen verkosta soitetuilla minuuteilla (muuttuja HINTAT). Kysyntäyhtälössä endogeenisen hinnan instrumentteina käytetään kaikkien selittävien muuttujien lisäksi numeronsiirrettävyys-dummya (NUMSIIRT), yrityksen markkinaosuutta (MS) ja markkinoilla kilpailevien matkaviestinnän palveluyritysten lukumäärää. Hinnan instrumenttien valinnassa taustalla on ajatus, että numeron siirrettävyydellä ja kilpailulla on ollut matkapuheluiden hintaa laskeva vaikutus ja yrityksen markkinaosuus voi vaikuttaa sen suhteellisen markkinavoiman kautta yrityksen hinnoitteluun.

Kysynnän ja palveluiden laadun välistä yhteyttä testaa kertoimien $\hat{a}_3$ suuruus ja merkittävyys. Kolme erillistä matkaviestinpalveluiden laatumuuttujaa (LPGSM, LGSM1800 ja LEGSM) on rakennettu kertomalla operaattorin investoinnit niille allokoitujen radiotaajuuksien lukumäärällä (PGSM, GSM1800,EGSM) ja jakamalla tämä luku yrityksen asiakkaiden lukumäärällä. Laatumuuttujat joudutaan estimoimaan erillisissä malleissa muuttujien voimakkaan keskinäisen korrelaation takia.

Tarjontayhtälön tärkein selittävä muuttuja on yrityksen rajakustannukset eli yhden lisäminuutin tarjonnan kustannukset, jota approksimoidaan muuttujalla MC. Muuttuja MC saadaan laskemalla verkko-operaattorin radioverkon operoinnista aiheutuvat kustannukset per minuutti. Tarjontayhtälössä radiotaajuusjaon muutosten oletetaan vaikuttavan yrityksen hinnoitteluun rajakustannusten kautta.

Yhtälöiden 5.1 ja 5.2 estimoinneista saaduissa tuloksissa on keskeistä estimoitujen kertoimien $\hat{a}_1$ ja $\hat{b}_1$ suuruus ja merkittävyys. Kerroin $\hat{a}_1$ mittaa kysynnän hintajoustoa eli sitä, kuinka monta prosenttiyksikköä tuotteen tai palvelun kysyntä muuttuu, kun sen hintaa muutetaan yhden prosenttiyksikön verran. Kerroin $\hat{b}_1$ kertoo kuinka monta prosenttiyksikköä yritykset muuttavat keskimäärin hintaansa, jos niiden rajakustannukset nousevat prosentilla. Näiden estimoitujen kertoimien perusteella käyttäen apuna arvioita radiotaajuusjaon muutosten aiheuttamista muutoksista rajakustannuksissa pystytään arvioimaan radiotaajuusjakojen muutosten taloudellisia vaikutuksia tekstissä esitetyllä tavalla.

Tutkimuksessa pyritään myös arvioimaan sitä, miten radiotaajuuksien uudelleenallokointi mahdollisesti vaikuttaisi kilpailuun ja suurimpien telepalveluita tarjoavien yritysten markkinaosuuksiin. Tätä varten edellisessä kysyntä-tarjontamallissa yhtälö 5.1. korvataan seuraavalla liittymien kysyntää mallintavalla yhtälöllä:

$$\log(S_{it}) = c_0 + c_1 P_{it} + c_2 X_{3it} + \varepsilon_{3it} ,$$

jossa S_{it} on yrityksen i matkapuhelinliittymien lukumäärä ajanhetkellä t, P_{it} = yrityksen i matkapuheluiden hinta ajanhetkellä t, X_{3it} = vektori muista matkapuhelinliittymien kysyntään vaikuttavista tekijöistä ja ε_{3it} on virhetermi. Yhtälön estimoinnista saatava kerroin $\hat{c}_1$ kertoo kuinka paljon yhden prosentin hinnan muutos matkapuheluissa vaikuttaa yrityksen liittymien kysyntään.

Radiotaajuusmuutosten vaikutus hintaan saadaan yllä esitetyllä tavalla yhtälöryhmän 5.1 ja 5.2 estimointien ja näistä tehtyjen laskelmointien perusteella. Radiotaajuusmuutoksen aiheuttamat prosentuaaliset muutokset yritysten liittymämäärissä saadaan kertomalla niiden hinnan %-muutos estimaatilla $\hat{c}_1$. Tämän perusteella voidaan laskea, mitkä olisivat olleet liittymämäärät ja yritysten markkinaosuudet liittymämäärillä mitattuna, mikäli radiotaajuudet olisi allokoitu toisin vuonna 2004.

Lähteet

Ameniya, T. (1985), *Advanced Econometrics*, Harward University Press, Cambridge, Mass.

Liite 7 Kuluttaja- ja tuottajaylijäämä eri skenaarioissa, yhteenveto

	Muutos kuluttajaylijäämässä (1000 euroa)	Muutos tuottajaylijäämässä (1000 euroa)	Hyvinvointimuutos (1000 euroa)
Skenaario A1	-16 660,050	-71 036,138	-87 696,188
Skenaario A2	-16 316,110	-67 252,894	-83 569,004
Skenaario B1	-6 470,979	-12 861,092	-19 332,070
Skenaario B2	-4 886,481	-21 183,215	-26 069,696
Skenaario B3	-22 023,516	-91 195,330	-113 218,846
Skenaario A1, vain Finnet muuttaa hintoja	+226,358	-81 448,576	-81 222,217
Skenaario B1, vain Finnet muuttaa hintoja	+184,685	-17 280,884	-17 096,199
Skenaario B1, 400 000 asiakasta siirtyy Soneralta Elisalle	-24 662,154	34 629,801	9 967,647
Skenaario B2, 400 000 asiakasta siirtyy Soneralta Elisalle	-24 562,405	26 047,945	1 485,540
Skenaario B3, 400 000 asiakasta siirtyy Soneralta Elisalle	-50 345,946	-62 098,948	-112 444,893