
Pentti Manninen

Johdatus tilastolliseen data-analyysiin

Sovellus- ja atk-keskeinen
näkökulma

Opetusmoniste B 44

7. painos

Toukokuu 2004

ISBN 951-44-60197

ISSN 1456-3177

Matematiikan, tilastotieteen ja

filosofian laitos

Tampereen yliopisto

33014 Tampereen yliopisto

© PENTTI MANNINEN
Matematiikan, tilastotieteen ja
filosofian laitos
Tampereen yliopisto
33014 Tampereen yliopisto
E-mail:stpema@uta.fi

Painopaikka
Juvenes Print
Tampereen Yliopistopaino Oy

Seitsemännän painoksen esipuhe

Seitsemänteen painokseen on tehty mm. seuraavat lisäykset ja muutokset:

- Ø Multiresponse-muuttujien muodostaminen dikotomisista muuttujista (2.5)
- Ø Summatiivisten muuttujien muodostaminen (2.7)
- Ø Kahden havaintomatriisin yhdistäminen
- Ø Viitearvot (3.2)
- Ø Järjestyslukujen käyttö mittaamisessa (4.2)
- Ø Keskiarvojen vertailut viitearvoihin
- Ø Normaalijakauma (6.7)
- Ø Tärkeys&tyytyväisyys-muuttujaparit ja kuiluanalyysi (6.8)
- Ø Profiilikuvio (7.1)
- Ø Korrespondenssianalyysi (8.9)
- Ø Tixel-ohjelma (luku 9)
- Ø Pyramidikuvio (10.4)
- Ø Tulostaulukoiden sivutus (10.6)
- Ø Kuvion muotoilujen kopiointi (11.5)
- Ø Kuvioden siirto PowerPoint-ohjelmaan (11.6)
- Ø Uusia graafisia esitystapoja: kelluvat palkit ja tähtikarttakuvio

Tampereella 21.5.2004

Pentti Manninen

Sisältö

1. JOHDANTO	1
1.1. YLEISTÄ	1
1.2. TILASTOLLINEN TUTKIMUSPROSESSI.....	2
2. TILASTOYKSIKKÖ, MUUTTUJA JA HAVAINATOMATRIISI	4
2.1. HAVAINTOAINEISTON SYÖTTÖ	5
2.2. TUNNUSTIEDOT JA LUOKKARAJAT	7
2.3. MUUTTUJIIN VIITTAAMINEN.....	11
2.4. EHTOJEN KÄYTTÖ	12
2.5. MULTIRESPONSE-MUUTTUJAT.....	14
2.6. UUSIEN MUUTTUJEN MUODOSTAMINEN.....	16
2.7. SUMMATIIVISTEN MUUTTUJEN MUODOSTAMINEN	19
2.8. KAHDEN MATRIISIN YHDISTÄMINEN.....	21
2.9. ESIMERKKIAINEISTOT.....	24
3. LISÄLUOKITUKSET JA VIITEARVOT.....	29
3.1. LUOKITUSARKKI	29
3.2. VIITEARVOT	31
4. MITTAAMINEN	32
4.1. MITTA-ASTEIKKOTYYPIT	32
4.2. JÄRJESTYSLUVUN KÄYTTÖ MITTAAMISESSA.....	33
5. YKSIULOTTEINEN JAKAUMA.....	35
5.1. KVALITATIIVINEN MUUTTUJA	35
5.2. KVANTITATIIVINEN MUUTTUJA	38
5.3. TIXEL-OHJELMAN KÄYTTÖ	42
5.4. USEAN JAKAUMAN KUVAAMINEN	43
5.5. PAINOMUUTTUJA	45
5.6. MULTIRESPONSE-MUUTTUJAT.....	46
6. YHDEN MUUTTUJAN TUNNUSLUVUT.....	48
6.1. KESKILUVUT.....	48
6.2. MUUT SIJAINLUVUT	50
6.3. HAJONTALUVUT.....	50
6.4. TIXEL-OHJELMAN KÄYTTÖ	51
6.5. JAKAUMAN KESKIVARVO JA KESKIHAJONTA	58
6.6. STANDARDOINTI	59
6.7. NORMAALIJAKAUMA	60
6.8. TYYTYVÄISYYS-TÄRKEYS MUUTTUJAPARIT JA KUILUANALYYSI	61
7. RYHMÄKOHTAISET TUNNUSLUVUT	63

7.1. RYHMÄKOHTAISET TUNNUSLUVUT, YKSI RYHMITTELYMUUTTUJA.....	63
7.2. RYHMÄKOHTAISET TUNNUSLUVUT, KAKSI RYHMITTELYMUUTTUJAA.....	71
8. KAKSIULOTTEISET JAKAUMAT	74
8.1. RISTIINTAULUKOINTI	74
8.2. ABSOLUUTTISET FREKVENSIT PROSENTEIKSI.....	79
8.3. KONTINGENSSEIKERROIN JA CRAMERIN PHI.....	80
8.4. RIIPPUVUUSLUKU RISTIINTAULUKOSTA.....	83
8.5. KORRELAATIODIAGRAMMI	85
8.6. KORRELAATIOKERROIN.....	88
8.8 KORRELAATIOMATRIISI.....	90
8.9. KORRESPONDENSSEIANALYYSI.....	91
9. TIXEL-OHJELMA	93
9.1. YLEISTÄ.....	93
9.2. TIXELIN MENETELMÄT	94
9.3. TIXELIN KÄYTTÖ.....	96
10. EXCEL.....	98
10.1. YLEISTÄ.....	98
10.2. HAVAINDOMATRIISIN SYÖTTÖ TYÖARKILLE	100
10.3. HAVAINDOMATRIISIN TALLENTAMINEN JA AVAAMINEN	102
10.4. LUKUJEN MUOTOILU.....	103
10.5. LAJITTELU.....	105
10.6. TYÖARKIN SIVUTUS.....	106
11. EXCEL-GRAFIikka	107
11.1. KUVION LUONTI.....	107
11.2. KUVIOTYYPIN VALINTA	113
11.3. KUVION MUOKKAAMINEN	114
11.4. YHDISTELMÄKUVIO.....	118
11.5. KUVION MUOTOILUJEN KOPIOINTI	119
11.6. KUVIOIDEN SIIRTO POWERPOINT-OHJELMAAN	119
12. WORD.....	121
12.1. TIEDON SIIRTÄMINEN SOVELLUSOHJELMASTA TOISEEN.....	121
12.2. TAULUKOT	122
12.3. KEHYS.....	122
12.4. RAJAUS.....	123
12.5. SISÄLLYSLUETTELOJEN LISÄÄMINEN ASIAKIRJAAN	124
13. TILASTOLLINEN HYPOTEEsin TESTAUS	125
13.1. HYPOTEEsin	126
13.2. LUOTTAMUSVÄLI	127
13.3. PROSENTTILUKUTESTIT	128
13.4. KESKIARVOTESTIT	130
13.5. OTOSKOKO	133

LIITE 1. EXCELIN TÄRKEIMMÄT KUVATYYPIT	135
LIITE 2. DATAN SIIRTO EXCELIN JA SPSS:N VÄLILLÄ.....	139
EXCELISTÄ SPSS:ÄÄN	139
SPSS:STÄ EXCELIIN	140

1. Johdanto

1.1. Yleistä

Tämä kirja on tarkoitettu ensimmäiseksi tilastotieteen oppikirjaksi sovellustieteiden - lähinnä kauppatieteiden, yhteiskuntatieteiden ja terveystieteen opiskelijoille. Lähestymistapa on korostetun atk- ja sovelluskeskeinen. Kirjan nimi olisi hyvin voinut olla ”Johdatus tilastolliseen atk-analyysiin”.

Kirjan tavoitteena on antaa lukijalle välineet toteuttaa sovellustieteiden alaan kuuluvan tutkimuksen tilastotieteen osuus alusta loppuun. Tavoite pitää sisällään aineiston syötön mikrotietokoneelle, tilastollisten analyysien tekemisen mukaan lukien graafiset esitykset ja lopuksi raportoinnin tekstinkäsittelyohjelmalla.

Tilastollisista menetelmistä mukana ovat yksiulotteinen jakauma (luku 5), tunnusluvut (luku 6), ryhmäkohtaiset tunnusluvut (luku 7) sekä kaksiulotteiset jakaumat ja korrelaatiotarkastelut(luku 8).

Monille opiskelijoille ensimmäinen tilastotieteen kurssi jää valitettavasti myös viimeiseksi. Siksi olen nähnyt tarpeelliseksi käsitellä lyhyesti myös tilastollista testausta - varsinkin, kun Tixel-ohjelmassa on testejä varten kätevät syöttöpohjat.

Kirjassa on käsitelty myös nk. multiresponse-tyyppisiä muuttujia, joiden käyttö antaa lisää mahdollisuuksia tilastomenetelmien soveltamiseen. Tilastoyksikkö (esim. henkilö) voi saada multiresponse-muuttujalla useita arvoja. Kysyttäessä esimerkiksi, minkä ketjujen myymälöistä haastateltava ostaa päivittäistavaroita, voi vastauksena olla useampia myymäläketjuja.

Aineiston atk-käsittely perustuu Tixel-ohjelmaan, joka on luotu Excel-
taulukkolaskentaympäristöön. Tälle linjavalinnalle on useita syitä:

1. Excel on yleisin mikrotietokoneiden taulukkolaskentaohjelmista, ja lähes sataprosenttisesti sama ohjelma toimii sekä PC- että Mac-mikroilla.
2. Käyttäjälle riittää yksi ohjelmaympäristö, jolla pystyy tekemään taulukkolaskennan, grafiikan, tietokantojen ja tilastoanalyysin tehtävät.
3. Aitona Windows-ohjelmana Tixelin tulosteet (taulukot ja kuvat) on helppo siirtää tekstinkäsittelyohjelmiin.
4. Excel-ympäristössä on käytettävissä Visual Basic for Applications-ohjelmointikieli (VBA), jota voidaan hyödyntää tulosten mahdollisessa jatkokäsittelyssä.

Yleiskatsaus Tixel-ohjelmaan samoin kuin ohjelman käynnistys on esitetty luvussa 9. Muutoin Tixel-ohjelma on integroitu tilastomenetelmiin siten, että jokaisen menetelmän yhteydessä on esitetty myös Tixel-ohjelman käyttö kyseisen menetelmän yhteydessä.

Jos lukijalla ei ole aikaisempaa kokemusta Excel-taulukkolaskennasta, on syytä aloittaa luvusta 10 ja esimerkiksi syöttää jokin pieni tilastoaineisto työarkille ja tallentaa se.

Tixel-ohjelma luo automaattisesti tulosta kuvaavan grafiikkaesityksen. Käytännössä kuvaa pitää kuitenkin vielä viimeistellä ja sen vuoksi Excelin grafiikkaa on käsitelty hieman laajemmin luvussa 11. Tässä yhteydessä on esitetty myös kuvan luonti valmiista taulukosta.

1.2. Tilastollinen tutkimusprosessi

Tilastollisessa tutkimusprosessissa voidaan erottaa neljä vaihetta:

1. Ongelma
2. Suunnitelma ongelman ratkaisemiseksi
3. Data-analyysi
4. Johtopäätösten teko

Tarkastellaan näitä neljää vaihetta esimerkin valossa. Ajatellaan että olemme seurusteluravintolan pitäjiä ja haluamme lisätä ravintolan kannattavuutta. Johtoryhmän kokouksessa esitetään useita vaihtoehtoja, mm. ulkomaisten oluiden ottamista valikoimaan, elävää musiikkia, halvempia hintoja, henkilökuntaan pannotamista jne. Miten asiakaskunta arvostaisi näitä toimenpiteitä, on jokseenkin epävarmaa. Tämän vuoksi päätettiin tehdä *markkinatutkimus*, jolla selvitettäisiin mm., mitä seikkoja potentiaaliset asiakkaat arvostavat seurusteluravintolan valinnassa.

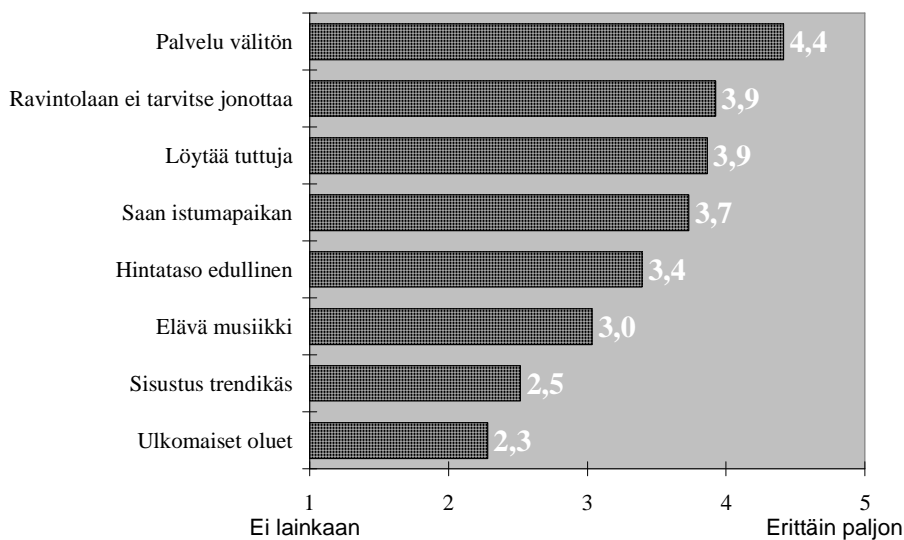
Vaikka johtoryhmän keskusteluissa ei ehkä kertaakaan mainittu sanaa *tilastotiede* tai sen johdannaisia, viittaa tämä ongelman asettelu selvästi tilastotieteeseen. Syytä on lähinnä kaksi. Ensinnäkin tutkimuksen kohderyhmä – potentiaaliset asiakkaat – on suhteellisen suuri ja toisekseen emme ole kiinnostuneita jokaisen tähän kohderyhmään kuuluvan henkilön mielipiteistä, vaan siitä miten eri markkinointivaihtoehtoihin *keskimääräisesti* suhtaudutaan. Eli halutaan asiakkaiden valintapreferenssien kymmenen kärjessä-lista.

Koska potentiaalinen asiakaskunta on kohtuullisen suuri, päätetään siitä poimia satunnaisesti osajoukko (otos), johon kuuluvilta tietoja kysytään. Päätetään myös käyttää *suljettua kyselyä*, jossa kysymykset ja vastausvaihtoehdot ovat edeltäkin määrätty. Näin saadaan tilastollinen aineisto – raakamateriaali-, joka jalostetaan data-analyysillä tilastollisiksi tunnusluvuiksi mm. prosenteiksi ja keskiarvoiksi.

Tilastollinen data-analyysi on tilastotieteen osa-alue, jossa lähtökohtana on empiirinen aineisto (ei todennäköisyysteoria), ja menetelmät ainakin tämän kirjan puitteissa suhteellisen yksinkertaisia. Tilastografiikan käytöllä on tulosten havainnollistamisessa keskeinen asema.

Seuraavassa kuviossa on pieni osa vastaavan todellisen tutkimuksen tulosta. Vaikkei olisi tilastolliseen data-analyysiin aiemmin perehtynyt, jokainen ymmärtänee tämän kuvion sisällön, sillä samantapaisia graafisia esityksiä ovat tiedotusvälineet pullollaan.

Kuvio 1.2.1. Miten voimakkaasti eri tekijät vaikuttavat seurusteluravintolan valintaan (aineisto vuodelta 1990).



2. Tilastoyksikkö, muuttuja ja havaintomatriisi

Mittauksen kohteet ovat *tilastoyksiköitä* (tutkimusyksiköitä) - esimerkkeinä henkilöt, kunnat ja maat. *Perusjoukko* eli *populaatio* on kaikkien tilastoyksiköiden muodostama joukko. Äänioikeutetut suomalaiset on puolueiden kannatuskyselyjen perusjoukko.

Perusjoukosta poimitaan yleensä satunnaisesti *otos* tai muuten valitaan osajoukko, josta mittaukset suoritetaan.

Mitattavia ominaisuuksia kutsutaan *muuttujiksi*. Esimerkkejä muuttujista ovat henkilön ikä, alkoholiin käytetty rahamäärä tietyssä viikkona ja kunnan taajama-aste. Näissä esimerkeissä tilastoyksikkö saa täsmälleen yhden arvon muuttujalla. Harvinaisempi muuttujatyyppi on nk. *multiresponse*-muuttuja. Tilastoyksikkö voi saada useampia arvoja tällaisella muuttujalla. Esimerkiksi jos kysytään, mitä mainostajia kesätapahtumaan osallistujan mieleen on jäänyt, voi hän ilmoittaa useampia mainostajia. Tarkemmin multiresponse-muuttujien käytöstä alaluvussa 2.5.

Mittatulokset esitetään taulukkomuodossa nk. *havaintomatriisina*. Siinä (vaaka)rivit vastaavat tilastoyksiköitä ja (pysty)sarakkeet muuttujia. Tiedot esitetään yleensä numeromuodossa, vaikka kyseessä olisi laadullinen muuttuja kuten sukupuoli. Numeroiden sisältö kirjoitetaan joko havaintomatriisin loppuun tai luokituskilille, joka sijaitsee havaintomatriisi-tiedostossa.

Muuttujien arvot voivat olla myös tekstitietoa, esim. kunnan nimi tai yrityksen osaston koodi.

Jos tilastoyksikön arvoa muuttujalla ei tunneta, on kyseessä *puuttuva tieto*. Sen merkitsemisessä on kaksi vaihtoehtoa:

- 1) Oletusvaihtoehto on **tyhjä solu**. Solun täytyy tällöin olla todella tyhjä. Esimerkiksi jos solussa on välilyönti (space), niin solu näyttää tyhjältä, mutta ei sitä kuitenkaan ole.
- 2) Merkitään puuttuva tieto erityisellä **puuttuvan tiedon arvolla**. Se voi olla mikä tahansa sellainen arvo, jota muuttuja ei voi todellisenä mittatuloksena saada. Puuttuvan tiedon arvo voi vaihdella muuttujakohtaisesti, mutta selkeintä on käyttää samaa puuttuvan tiedon arvoa kaikilla muuttujilla. Tyypillinen puuttuvan tiedon arvo on esimerkiksi -1 . Ohjelma lukee puuttuvan tiedon arvot havaintomatriisin toiselta riviltä

Näitä kahta tapaa merkitä puuttuva tieto voidaan myös yhdistää. Siten muuttujalla voi olla kaksi puuttuvan tiedon merkkiä: tyhjä solu ja jokin puuttuvan tiedon arvo. Tämä on kätevää silloin, jos yhdessä tilastomenetelmässä halutaan pitää

puuttuvan tiedon arvot mukana laskennassa ja toisessa tilastomenetelmässä pitää ne laskennan ulkopuolella. Siten esimerkiksi ”ei osaa sanoa”-vaihtoehto voidaan pitää tilastokäsittelyssä mukana tai jättää ulkopuolelle sen mukaan, mitä toiselle riville on kirjoitettu.

Seuraavassa taulukossa on pieni osa nk. opiskelija-aineiston havaintomatriisista. Esimerkkiaineistoja käsittelevässä alaluvussa (2.7) on esitelty tarkemmin kaikki aineistot, joihin tämän teoksen esimerkkitulosteet perustuvat.

Taulukko 2.1. Havaintomatriisin osa

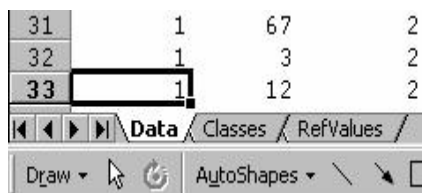
	A	B	C	D	E	F	G	H
1	Kysely-aineisto							
2		0	0	0	0	0	0	0
3	Nro	Suku puoli	Ikä luokka	Siviili säätty	Tutkinto	Koulutus	Aloitus vuosi	Luennot + harj./vk
4	1	2	3	4	14	2	82,5	13
5	2	1	4	1	20	1	78,5	25
6	3	2	4	4	34	2	77,5	10
7	4	1	2	4	38	2	82,5	29
8	5	1	1	4	38	2	82,5	26
9	6	1	5	1	3	1	81,5	20
10	7	1	3	4	20	2	82,5	30
11	8	1	3	4	3	2	79,5	20
12	9	2	2	4	14	2	81,5	25
13	10	1	3	4	38	2	82,5	26

2.1. Havaintoaineiston syöttö

2.1.1. Työarkit

Havaintomatriisin sisältävässä Excel-työkirjassa voi olla useampia työarkkeja, joita hyödynnetään tilastoanalyysissä.

Data-arkit. Yleisin tapaus on, että havaintomatriisi mahtuu yhdelle työarkille. Tälle työarkille kannattaa antaa nimi *Data*. Se tehdään merkkaamalla arkin vanha nimi ja kirjoittamalla päälle uusi nimi. Tämä uudelleen nimeäminen ei välttämättä, mutta se on suositeltavaa. Data-



arkkeja voi olla myös useampia. Esimerkiksi eri vuosien aineistot voidaan sijoittaa eri työarkeille, jolloin niitä voidaan käsitellä erikseen. Tällöin mitään työarkkia ei nimitä Data-nimellä ja Tixel käsittelee aina sitä työarkkia, joka on aktiivinen, kun tilastomenetelmä valitaan. Tilastoaineisto on myös usealla arkilla sil-

loin, kun muuttujien tai tilastoyksiköiden lukumäärät ylittävät työarkin rajat (256 saraketta, n. 65536 riviä). Tätä tapausta ei tarkastella tässä yhteydessä lähemmin. Suuri aineisto voidaan tallentaa myös MsOfficen Access-tietokantaan, jolloin näitä Excel-työarkin rajoituksia ei ole.

Luokitusarkki. Luokitusarkkiin sijoitetaan lisäluokitukset. Tällaisia ovat esimerkiksi iän vaihtoehtoiset luokitukset. Tixelissä ei tarvitse määritellä uusia muuttujia uusille ikäluokituksille, vaan riittää kun määritellään ikä-muuttujalle vaihtoehtoiset luokitukset. Luokitusarkin nimi on *CLASSES* ja luokitusarkkia käsitellään tarkemmin luvussa 3.1.

Viitearkki. Viitearkille sijoitetaan muuttujien mahdolliset viitearvot, jollaisia ovat esimerkiksi muuttujien keskiarvot aikaisemmissa tutkimuksissa. Nämä viitearvot saadaan automaattisesti esille tulostuksissa. Viitearkin nimi *RefValues* ja asiaa käsitellään tarkemmin luvussa 3.2.

Kerroinarkki. Summatiivisten muuttujien muodostamisessa tarvitaan kertoimia, jotka sijoitetaan kerroinarkille. Kerroinarkin nimi on *Kertoimet* (kts. luku 2.7).

2.1.2. Havaintomatriisin kolme ensimmäistä riviä

Havaintomatriisin kolmen ensimmäisen rivin sisältö on seuraava:

Ensimmäiselle riville soluun A1 sijoitetaan aineiston otsikko, joka automaattisesti tulostuu kaikkiin Tixelin tulosteisiin. Ensimmäiselle riville sijoitetaan myös luokituksen mahdollinen numero (tarkemmin luvussa 3) ja multiresponse-muuttujaa tarkoittava mr-merkintä.

Toiselle riville kirjoitetaan muuttujia vastaaviin sarakkeisiin kunkin muuttujan mahdollinen puuttuvan tiedon arvo. Jos havaintomatriisin solussa on puuttuvan tiedon arvo, Tixel jättää sen automaattisesti laskennasta pois. Puuttuvan tiedon oletusvaihtoehto on tyhjä solu.

Kolmannella rivillä ovat muuttujien nimet. Muuttujien niminä ei saa käyttää solun osoitetta, esim. X10. Muuttujien nimien pituudella ei ole ylärajaa ja muuttujan nimi voi muodostua useammasta sanasta, esim. *Lippuihin käytetty rahamäärä*. Jos nimi on niin pitkä, että osa siitä peittyy seuraavan solun "alle", voidaan kolmannelle riville antaa muotoilu **Format/Cells/Alignment** (Muotoile/Solut/Tasaus), jonka valintaikkunassa **Wrap**-kohtaan (Tekstin rivitys) laiteetaan rasti. Tällöin solussa oleva pitkä teksti rivitetään siten, että se näkyy kokonaan. Samalla rivin korkeus kasvaa. Tällainen määrittely pitkien muuttujien nimien kohdalla ei kuitenkaan ole pakollinen ja pitkät muuttujien nimet siirtyvät täydellisinä tulostaulukoihin.

Neljänneltä riviltä alkaen soluihin sijoitetaan muuttujien arvot tilastoyksiköillä.

Jos neljännen rivin solut muotoillaan **Format/Cells/Number**-komennolla (Muotoile/Solut/Luku), niin ohjelma käyttää samaa muotoilua tulosteissa sikäli kun se on järkevää. Siten esimerkiksi keskiarvoihin on mahdollista määrätä etukäteen näytettävien desimaalien määrä.

Kun havaintoaineisto on syötetty työarkille, tallennetaan se kovalevyille tai diskettille **File/Save**-komennolla (Tiedosto/Tallenna). Jos käytetään kovalevyä, on parasta sijoittaa tiedosto *datahakemistoon* (esim. C:\TIXEL\DATAT).

2.2. Tunnustiedot ja luokkarajat

Havaintomatriisin lisäksi tilastokäsittelyssä tarvitaan muuttujien arvojen *tunnustietoja*, esimerkiksi sukupuolen kohdalla on tiedettävä, kumpi arvo tarkoittaa naista ja kumpi miestä. Kvantitatiivisen muuttujan tapauksessa käytetään usein luokitusta eli muuttujan arvot ryhmitellään luokkiin, esimerkiksi ikä luokitellaan kolmeen luokkaan: alle 20 v., 21-40 v. ja yli 40 v.

Muuttujan tunnustietoja ja luokkarajoja kutsutaan muuttujan *määrittelytiedoiksi*. Ne voidaan määrittellä kahdella eri tavalla.

- 1) Tiedot syötetään havaintomatriisin alapuolelle kunkin muuttujan omaan sarakkeeseen.
- 2) Tiedot syötetään *luokitusarkille*, joka sijaitsee havaintomatriisi-tiedostossa.

Seuraavassa käsitellään tarkemmin ensimmäistä tapaa. Jälkimmäinen vaihtoehto esitellään kolmannessa luvussa.

Muuttujan määrittelytiedot sijoitetaan havaintomatriisin alapuolelle ko. muuttujien sarakkeisiin siten, että havaintomatriisin viimeisen rivin ja määrittelytietojen väliin jää **yksi tyhjä rivi**.

Esimerkki. Terveyskeskuksen potilastutkimuksessa oli mm. kaksi kvalitatiivista muuttujaa (viikonpäivä ja jonosyy) ja kaksi kvantitatiivista muuttujaa (ikä ja ilmoittautumisaika). Havaintomatriisin viimeisen tilastoyksikön tiedot ovat rivillä 749 ja määrittelyalue alkaa riviltä 751. Seuraavassa taulukossa on esillä kolme ensimmäistä riviä ja määrittelyalue. Taulukon näyttö on jaettu kahteen osaan **Window/Split**-komennolla (kts. tarkemmin luku 10: Excel).

Taulukko 2.2.1. Osa terveystietokeskus-aineiston havaintomatriisia

	A	B	C	D	E	
1	Jonotusaineisto					
2	-1	-1		-1		-1
3	Viikonpäivä	Ikä	Tulosyy	Ilmoittautumisaika		
747	5	54		2		8,00
748	5	55		1		8,18
749	5	18		9		7,42
750						<i>Viimeinen tilastoyks.</i>
751	ma	5	"1=Tapaturma"			"6/Yö"
752	ti	10	"4=Vatsavaiva"	"10/Aamupäivä"		
753	ke	20	"6=Rintakipu"			"14/Päivä"
754	to	40	"8=Korvakipu"			"18/Iltapäivä"
755	pe	60				"24/Iltä"
756	la	80				
757	su	100				
758						
759						
760	<i>Määrittelyalue</i>					

Muuttujan määrittelytiedoissa on viisi eri vaihtoehtoa, joista kahta ensimmäistä käytetään yleensä kvalitatiivisiin muuttujiin ja kahta seuraavaa kvantitatiivisiin muuttujiin. Viides vaihtoehto on, että muuttujalle ei anneta mitään määrittelytietoa.

1. Luokkien numerointi ykkösestä eteenpäin ykkösen välein

Kvalitatiivisen muuttujan tapauksessa on suositeltavin tapa numeroida luokkien arvot ykkösestä eteenpäin ykkösen välein. Esimerkiksi viikonpäivän tapauksessa ma=1, ti=2, ke=3 jne. Tällöin luokkatunnukset sijoitetaan luokitusalueella allekkain ilman numeroita. Ohjelma päättelee tunnusten järjestyksen perusteella, mitä arvoa kukin luokkatunnus vastaa. Tämä on helpoin tapa määrittellä kvalitatiivisen muuttujan luokkatunnukset. Taulukossa 2.2.1 viikonpäivä-muuttuja on näin määritelty. Viereisessä esimerkissä Sukupuoli-muuttujan luokat määritelty siten, että 1=Nainen ja 2=Mies.

Luokkien nimiä ei koske muu rajoitus kuin että nimi ei saa alkaa lainausmerkillä ("), koska tämä merkki varattu seuraavaan määrittelytapaan. Luokkia voi käytännössä olla

	C
1	
2	-1
3	Sukupuoli
749	2
750	
751	Nainen
752	Mies

kuinka paljon tahansa.

2. Luokkien numerointi muunlainen kuin kohdassa yksi

Esimerkiksi yrityksen osastoilla on vakiintuneet numerot 10, 11, 20, 21, 30 ja 31 ja näitä numeroita käytetään myös aineiston syö-tössä. Tällöin jokainen luokka määritellään seuraavalla tavalla:

”numero=luokkatunnus”.

Vieressä on esimerkki tämän vaihtoehdon käytöstä.

”10=Tuotanto1”
”11=Tuotanto2”
”20=Varasto”
”21=Kiinteistö”
”30=Hallinto”
”31=Markkinointi”

Taulukossa 2.2.1 on tulosyy-muuttujalla kymmenen arvoa, jotka on numeroitu 1-10. Tilastoanalyysiin halutaan ottaa kuitenkin mukaan vain neljä tulosyötä (1=tapaturma, 4=vatsavaiva, 6=rintakipu ja 8=korvakipu), ja tämä saadaan aikaan taulukon 2.2.1 esittämällä tavalla.

Samalla tavalla saadaan *ei osaa sanoa*-vaihtoehto pois viisiluokkaisesta asteikosta: 1=täysin samaa mieltä, 2=jokseenkin samaa mieltä, 3=ei osaa sanoa, 4=jokseenkin eri mieltä, 5=täysin eri mieltä. Tällöin määritellään luokkanimet arvoille 1, 2, 4 ja 5, mutta kolmas arvo jätetään määrittelemättä.

3. Luokkien ylärajat

Kvantitatiivisen muuttujan luokat määritellään luokkien ylärajojen avulla, jotka sijoitetaan muuttujan määrittelyalueelle allekkain nousevassa järjestyksessä. Tulostaulukoihin luokkien tunnuksiksi tulevat luokkavälit. Taulukossa 2.2.1 ikämuuttujan luokkarajat antavat seuraavat luokat: -5, 6-10, 11-20, 21-40 jne. On huomattava että kun viimeisen luokan yläraja on 100, niin yli satavuotiaat jäävät ikäjakauman ulkopuolelle.

Tixel olettaa luokkavälejä määrätessään, että muuttujan mittatarkkuus on yksi, jolloin luokan alaraja on yhtä kuin edellisen luokan yläraja plus yksi. Jos mittatarkkuus on pienempi kuin yksi, on käytettävä seuraavaa vaihtoehtoa, jolla voidaan vapaasti määrittellä luokkatunnukset.

4. Luokkien ylärajat ja tunnukset

Luokka määritellään ylärajan ja tunnuksen avulla seuraavasti:

”yläraja/luokkatunnus”.

”6/Yö”
”10/Aamupäivä”
”14/Päivä”
”18/Iltapäivä”
”24/Ilta”

Taulukossa 2.2.1 ilmoittautumisaika on luokiteltu viiteen vuorokaudenaikaan viereisten määrittelyjen mukaisesti. Ensimmäinen luokkatunnus *yö* vastaa aikaa klo 0-6.00, toinen luokkatunnus *aamupäivä* vastaa aikaa klo 6.01-10.00, jne. Jos olisi käytetty vaihtoehdon kolme mukaisesti vain ylärajoja, olisi tulostaulukoihin tullut vain luokkavälit - ei luokkatunnuksia.

5. Muuttujan määrittelyalue on tyhjä.

Muuttujan määrittelyalueen ollessa tyhjä ohjelma etsii kaikki muuttujalla esiintyvät arvot (tai tekstit) ja käyttää niitä luokkina. Esimerkiksi jos terveyskeskusaineiston ikä-muuttujalle ei määritellä luokitusta, käyttää ohjelma kaikkia aineistosta löytyviä ikävuosia luokkina. Yksiulotteinen jakauman alku- ja loppuosa on silloin viereisen taulukon mukainen. On huomattava, että kaikkia arvoja nollan ja suurimman arvon välillä ei esiinny. Esim. 90 on tällainen arvo.

Ikä vuosina	Lkm
0	11
1	30
2	28
3	22
4	10
...	...
...	...
88	1
89	1
91	1
92	1
Yht	745

Jos muuttujan arvot tekstitietoa, esim. asuinkunta tai yrityksen nimi, ei muuttujalle voida antaa luokkamäärittelyä, joten määrittelyalue jää tyhjäksi.

Yleensä kannattaa tallennetulle havaintomatriisille tehdä *tarkistusajo* yksiulotteisen jakauman menetelmällä (katso luku 5) siten, että muuttujat ovat ilman määrittelyä. Tällöin paljastuvat tiettytyypiset syöttövirheet. Esimerkiksi jos sukupuolen arvot ovat 1 ja 2 ja tulostaulukossa esiintyy myös arvo 11, voidaan päätellä, että syöttövirhe on tosiasia. **Edit/Find**-komennolla (Muokkaa/Etsi) etsitään, millä tilastoyksiköllä virhe esiintyi, ja tehdään korjaus vastaavaan tähän soluun.

Yksiulotteisen jakauman valintaikkunassa on rasti sitä varten, että muuttujat käsitellään ilman määrittelytietoja, joten määrittelyaluetta ei tarvitse tarkistusajoa varten konkreettisesti tyhjentää.

Luokitusten muuttaminen

Luokituksia voidaan muokata, kun Tixel-ajo on tehty, tulokset ovat Tulos.xls-työkirjassa ja järjestelmä on Excel-tilassa. Window-valikon avulla voit siirtyä takaisin havaintomatriisiin ja tehdä luokituksiin tarvittavat muutokset (kts. tarkemmin 9.2).

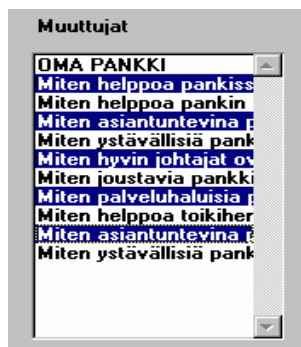
2.3. Muuttujiin viittaaminen

Eri tilastollisissa menetelmissä viitataan muuttujiin tietyillä nimillä. Esimerkiksi menetelmässä *Ryhätunnusluvut*, yksi *ryhmittelymuuttuja* käsitellään kahta muuttujaa, joista toista kutsutaan *riippuvaksi muuttujaksi* ja toista *ryhmittelymuuttujaksi*. Näiden lisäksi voidaan käyttää *ehtomuuttujia*, joiden avulla tilastokäsittely rajataan koskemaan havaintoyksiköiden tiettyä osajoukkoa.

Erityyppiset muuttujat määritellään niille varattujen ikkunoiden avulla. Seuraavassa on edellä mainitun menetelmän valintaikkuna. Riippuvaksi muuttujaksi on määritelty jonotusaika ja ryhmittelymuuttujaksi viikonpäivä eli tarkoituksena on laskea jonotusajan tunnusluvut eri viikonpäivinä.

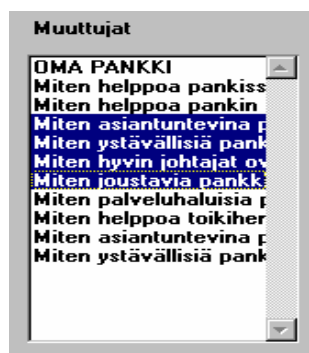
Muuttujien määrittely tehdään siten, että ikkunassa olevaa muuttujan nimeä napsautetaan hiirellä, jolloin kyseisen nimen kohdalle ilmestyy tumma (sininen) palkki. Jos muuttujan tyyppiä ilmaisevan tekstin kohdalla on monikkomuoto (esim. Riippuva(t) muuttuja(t)), on muuttujia mahdollista valita useampia. Se tehdään samalla tavalla kuin monivalinta yleisestikin Windows-ympäristössä.

Hajavalinta. Jos valittavat muuttujat ovat listassa hajallaan (ei peräkkäin), pidetään Ctrl-näppäintä alhaalla, kun muuttujan nimeä napsautetaan.



Välivalinta. Peräkkäiset muuttujat saadaan helpoimmin valittua siten, että napsautetaan ensin ensimmäisen muuttujan nimeä. Viimeinen muuttuja merkitään Shift+napsautuksella.

Peräkkäiset muuttujat voidaan valita myös hiirtä vetämällä, jolloin hiiren vasenta näppäintä pidetään alhaalla.



Jos muuttujia on niin paljon, että kaikkien nimet eivät näy valintaikkunassa, käytetään vierityspalkkia piilossa olevien muuttujien esille tuomiseksi. Palkin päissä on painikkeet, jotka tuovat yhden uuden muuttujan näkyviin. Napsauttamalla vieritysneliön eli "hissin" alapuolella olevaa aluetta saadaan näkyviin uusi ikkuna muuttujia. Hiirellä voi myös tarttua "hissin" ja vetää sitä alas- tai ylöspäin.

Kun valintaikkunan muuttujatyypit on näin määritelty, napsautetaan OK-painiketta, jolloin ohjelma siirtyy seuraavaan vaiheeseen.

2.4. Ehtojen käyttö

Ehtoja käytetään kun tilastokäsittely halutaan rajata tilastoyksiköiden tiettyyn osajoukkoon. Esimerkiksi halutaan ensin käsitellä aineistosta ne, jotka asuvat kaupungeissa.

Ehtojen lukumäärä voi enimmillään olla kolme. Jokainen ehto määritellään ehtomuuttujan ja matemaattisen lausekkeen avulla. Erilaisia ehtotyyppjeä on neljä:

- 1) Ehtomuuttujan arvo on tietty luku. Esimerkiksi sukupuoli=1.

- 2) Ehtomuuttujan arvo ei ole tietty arvo. Tällöin merkitään rasti *Eri suuri*-kohtaan. Esimerkiksi viikonpäivä # 7 eli viikonpäivä ei ole sunnuntai.
- 3) Ehtomuuttujan arvo kuuluu tiettyyn väliin. Esimerkiksi ikä on suurempi kuin 20 vuotta ja korkeintaan 50 vuotta.
- 4) Ehtomuuttujan arvo ei kuulu tiettyyn väliin. Tällöin merkitään rasti *Eri suuri*-kohtaan.

Ehtomuuttujat ja ehdon tyytit ja arvot määritellään seuraavalla valintaikkunalla.

Huom1. Huomaa että *Väli*-tyyppisessä ehdossa vain ylärajassa on yhtä suuruus.

Huom2. Pelkästään *suurempi kuin*-ehtotyyppiä ei ole, vaan ehtoon on liitettävä aina yläraja. Esimerkiksi jos tilastokäsittelyyn halutaan mukaan yli 50-vuotiaat, alarajaksi merkitään 50 ja ylärajaksi riittävän suuri luku, vaikkapa 120. Sama asia pätee *pienempi kuin*-ehtotyyppiin, jolloin alarajaksi pitää merkitä luku, joka minimiarvoa pienempi.

Jos ehtoja (ehtomuuttujia) on enemmän kuin yksi, on määrittävä, yhdistetäänkö ehdot JA- vai TAI-operaattorilla. Napsauttamalla alaspäin osoittavaa nuolta saadaan molem-



mat vaihtoehdot esille.

Esimerkki 1. Jos tilastokäsittelyyn halutaan mukaan 11-50-vuotiaat miehet, valitaan ensimmäiseksi ehtomuuttujaksi sukupuoli ja ehto on tyyppiä yksi. Toinen ehtomuuttuja on ikä ja ehto on tyyppiä kolme. Ehdot yhdistetään JA-operaattorilla.

Esimerkki 2. Jos tilastokäsittelyyn halutaan mukaan 0-10-vuotiaat ja yli 70-vuotiaat, käytetään kahta ehtoa, joissa molemmissa ehtomuuttuja on ikä. Ehdot yhdistetään TAI-operaattorilla ja molemmat ehdot ovat tyyppiä kolme. Tässä puhekielen ilmaisu johtaa helposti harhaan, koska siinä käytetään *ja*-sanaa. Yksittäinen tilastoyksikkö ei kuitenkaan voi kuulua molempiin ryhmiin. Tilastoyksikkö on joko korkeintaan 10-vuotias tai yli 70-vuotias. Tai sitten ei kumpakaan, jolloin ehdon totuusarvo on epätosi.

2.5. Multiresponse-muuttujat

Tilastokäsittelyn peruslähtökohta on, että tilastoyksikkö (esim. henkilö tai kunta) saa *yhden* arvon muuttujalla. Joskus on kuitenkin tilanteita, joissa tilastoyksikön olisi kätevä saada muuttujalla *useita* arvoja. Esimerkiksi, kun halutaan tietää, mitä oireita potilaalla on ollut, ja vastaajalla on mahdollisuus valita useita vaihtoehtoja.

Eräs tapa koodata tällaiset vastaukset on varata vastaukseen esim. kolme muuttujaa, jolloin 1. muuttuja kertoo 1. oireen, 2. muuttuja toisen oireen jne. Jos vastaajalla on vain yksi oire, saavat kaksi muuta muuttujaa puuttuvan tiedon arvon. Jos henkilön oireiden lukumäärä on kolmea suurempi, merkitään näistä oireista vain kolme.

Toinen mahdollisuus on varata jokaiselle oireelle oma kaksiluokkainen eli dikotominen muuttuja (1=on ko. oire, 0=ei ole). Näiden molempien menettelyjen epäkohtana on, että tulokset hajoavat useamman muuttujan osalle. Tällöin on hankalaa saada ominaisuuden suora jakauma ja ristiintaulukot muiden muuttujien kanssa.

Paras menettely on, että muuttuja voi saada useampia arvoja. Tällaista muuttujaa kutsutaan multiresponse- eli monivastausmuuttujaksi.

Kun yhteen soluun kirjoitetaan useita lukuja, on luvut erotettava toisistaan jollain erottimella. Tätä erotinta kutsutaan mr-erottimeksi ja sen oletusarvo plus-merkki (+).

Esimerkiksi jos henkilö on valinnut muuttujan vastausvaihtoehtoista 2, 4 ja 7, niin soluun kirjoitetaan 2+4+7. Excel ei tulkitse tätä yhteenlaskuksi, koska lauseke ei ala yhtäsuuruusmerkillä.

Multiresponse-muuttujan sarakkeen ensimmäiseen soluun merkitään mr-tunniste, jotta Tixel osaa käsitellä muuttujaa oikealla tavalla.

Mr-muuttujan luokkien nimet sijoitetaan tavalliseen tapaan syöttöarvojen alapuolelle. Viereisessä esimerkissä on muuttujan *Missä ravintoloissa käynyt puolen vuoden aikana?* kahdeksan henkilön vastaukset.

Muita esimerkkejä mr-muuttujista:

Mitä atk-ohjelmia olet käyttänyt? Mahdollisten ohjelmien luettelo käsittää useita kymmeniä ohjelmia ja vastaaja on yleensä käyttänyt useita ohjelmia.

Kysytään mitä adjektiiveja vastaaja liittää tuotteeseen. Vastaaja saa vapaasti valita useiden kymmenien adjektiivien joukosta ne, jotka parhaiten hänen mielestään kuvaavat tuotetta.

	A
1	mr
2	
3	Missä ravintoloissa käynyt puolen vuoden aikana?
4	1+2
5	1
6	2+3
7	1+2+3
8	2
9	2
10	3
11	2+3
12	
13	Ravintola A
14	Ravintola B
15	Ravintola C

Mr-muuttujan muodostaminen kaksiluokkaisista muuttujista

Tixel-ohjelmalla voidaan kaksiluokkaisista eli dikotomisista muuttujista muodostaa yksi mr-muuttuja. Tarkastellaan viereistä esimerkkiä, jossa on kolme dikotomisista muuttujaa x, y ja z.

	A	B	C	D	E
1					
2					
3	nr	x	y	z	
4	1	1			
5	2	1	1		
6	3		1	1	
7	4		1		
8	5	1	1	1	
9	6			1	

Valitse Tixel-valikosta *Havaintomatriisin aputoiminnot/Multiresponse muuttujan muodostaminen*. Saadaan viereinen määrittelyikkuna, jossa merkitään mr-muuttujan määrittelevät muuttujat. Kyllä-vastauksen arvoksi kirjoitetaan yksi.

Mr-muuttuja tulostuu havaintomatriisin loppuun oheisen esimerkin mukaisesti.

	A	B	C	D	E	F
1						
2						
3	nr	x	y	z	Mr-muuttuja	
4	1	1			1	
5	2	1	1		1+2	
6	3		1	1	2+3	
7	4		1		2	
8	5	1	1	1	1+2+3	
9	6			1	3	

2.6. Uusien muuttujien muodostaminen

Excelin avulla on helppo muodostaa uusia muuttujia havaintomatriisissa olevien muuttujien funktioina. Tällöin on mahdollista käyttää myös ehtoa siten, että funktion lauseke riippuu ehtomuuttujan arvosta. Uuden muuttujan avulla voidaan myös määritellä epäyhtenäinen luokitus.

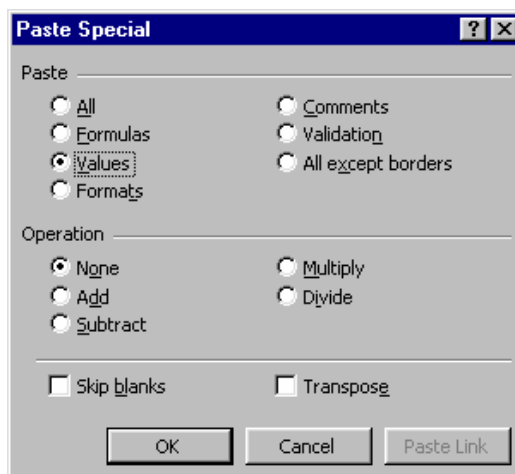
Uuden muuttujan määrittely

Tarkastellaan esimerkkiä, jossa henkilön painon ja pituuden avulla muodostetaan nk. body-indeksi. Indeksä on yhtä kuin paino jaettuna pituuden neliöllä. Oheisen havaintomatriisin soluun D4 kirjoitetaan kaava $=B4/C4^2$, joka antaa tulokseksi 24,39. Huomaa, että kaava alkaa aina yhtäsuuruus-merkillä. Potenssiin korotus eli hattu-merkki tulee näytölle näkyviin vasta, kun olet kirjoittanut seuraavan merkin.

	A	B	C	D
1				
2				
3	Sukupuoli	Paino	Pituus	
4	1	70,5	1,70	24,39
5	2	78,3	1,62	
6	1	70,0	1,77	
7	1	80,8	1,65	
8	2	69,5	1,73	
9	2	71,2	1,54	
10	1	83,5	1,82	

Seuraavaksi merkitään (“maalataan”) alue, joka alkaa kaavasta ja jatkuu alaspäin niin pitkälle kuin tilastoyksiköitä riittää. Huomaa, että kaavan solu pysyy valkoisena. Kaavan kopiointi merkityihin soluihin tehdään komennolla **Edit/Fill/Down**. Tämän jälkeen merkityn alueen soluissa on kaavat, jotka tulee muuttaa vielä pelkiksi luvuiksi, sillä havaintomatriisissa ei saa olla kaavoja.

Kaavat on mahdollista muuttaa luvuiksi **Edit/Copy-Edit/Paste Special** -komentoparilla. Molemmissa



komennossa tulee kaikki uuden muuttujan solut olla merkittyjä. Jälkimmäisen komennon valintaikkunassa napsautetaan **Values**-kohtaa.

(Paste Special on muuten varsin monipuolinen komento. Voit esimerkiksi transponoida matriisin eli vaihtaa rivit sarakkeiksi).

Anna lopuksi nimi uudelle muuttujalle.

Ehdollinen funktio

Joskus halutaan käyttää eri funktioita ehtomuuttujan eri arvoilla. Esimerkiksi sukupuolen arvolla yksi käytetään edellä mainittua body-indeksiä, kun sen sijaan sukupuolen arvolla kaksi on 0,9 kertaa paino jaettuna pituuden neliöllä. Tällainen määrittely saadaan **IF**-lauseella. Soluun D4 kirjoitetaan kaava

$$=IF(A4=1;B4/C4^2;0,9*B4/C4^2).$$

IF-lausekkeessa on kolme argumenttia, jotka on erotettu toisistaan puolipisteellä. Ensimmäinen argumentti on ehto ($A4=1$), jonka totuusarvo on tosi tai epätosi. Jos totuusarvo on tosi, käytetään toista argumenttia ($B4/C4^2$). Jos totuusarvo

on epätosi eli solussa A4 on ykkösestä eroava arvo, käytetään kolmatta argumenttia eli kaavaa $0,9*B4/C4^2$.

Edellä esitetty sopii tilanteeseen, jossa ehtomuuttujalla on kaksi arvoa. Jos arvoja on useampia, on mahdollista käyttää sisäkkäisiä IF-lausekkeita. Toinen ja ehkä selkeämpi tapa on määritellä niin monta uutta osamuuttujaa kuin ehtomuuttujalla on eri arvoja ja sen jälkeen summata nämä osamuuttujat lopulliseksi muuttujaksi.

Epäyhtenäinen luokitus

Luvussa 2.2 käsiteltiin luokitusta, jossa yhtenäiset arvojoukot muodostivat luokituksen. Tarkastellaan esimerkkiä epäyhtenäisestä luokituksesta, jossa tyyppilinen viisiluokkainen asteikko (1=täysin samaa mieltä, 2=jokseenkin samaa mieltä, 3=ei osaa sanoa, 4=jokseenkin eri mieltä, 5=täysin eri mieltä) pitää muuttaa kolmiluokkaiseksi siten, että uusi asteikko mittaa asenteen jyrkyyttä. Silloin lukuja 1 ja 5 kuvataan ykkösellä, lukuja 2 ja 4 kakkosella ja luku 3 pysyy muuttumattomana.

Havaintomatriisin viimeisen sarakkeen oikealle puolelle kirjoitetaan *vastintaulukko* siten, että viimeisen sarakkeen ja vastintaulukon väliin jää vähintään yksi tyhjä sarake. Vastintaulukon ensimmäiseen sarakkeeseen kirjoitetaan alkuperäisen muuttujan arvot allekkain pienimmästä suurimpaan. Näiden solujen oikealle puolelle kirjoitetaan vastaavat uudet arvot oheisen esimerkin mukaisesti (havaintomatriisissa on vain yksi muuttuja Asenne1).

	A	B	C	D
1			Vastintaulu	
2			1	1
3	Asenne1		2	2
4		4	3	3
5		1	4	2
6		1	5	1

Merkitään (“maalataan”) vastintaulukko ja annetaan sille esim. nimi TAU, komennolla **Insert Name Define**, jossa määrittelykenttään kirjoitetaan TAU ja painetaan OK:ta. Uusi muuttuja määritellään **VLOOKUP**-funktiolla (PHAKU-funktio). Esimerkissä kirjoitetaan soluun B4

=VLOOKUP(A4;TAU;2).

Solu B4 kopioidaan muille tilastoyksiköille kuten edellä on esitetty. Lisäksi soluista poistetaan kaavat ja tyhjennetään vastintaulukko. Tulos on viereisen taulukon mukainen.

	A	B
3	Asenne1	Asenne2
4		4
5		1
6		1
7		2

Keskiarvo-muuttuja

Useasta muuttujasta voidaan muodostaa keskiarvo- tai summafunktiolla uusi muuttuja. Seuraavassa esimerkissä on laskettu kolmen muuttujan keskiarvo-muuttuja D-sarakkeeseen. Soluun D4 kirjoitetaan keskiarvon kaava, joka kopioidaan muille tilastoyksiköille. Lopuksi poistetaan kaavat **Edit/Copy-** ja **Edit/Paste Special/Values-**komennoilla.

	A	B	C	D	E	F	G
1							
2					=AVERAGE(A4:C4)		
3	Asenne1	Asenne2	Asenne3				
4	3	2	3	2,67			
5	4	3	2	3,00			
6		4	4	4,00			
7	1	5		3,00			
8							

Puuttuvan tiedon paikalla on tyhjä, jolloin keskiarvofunktio laskee ei-puuttuvien tietojen keskiarvon. (Tyhjän paikalla voi olla myös esim. välilyönti).

2.7. Summatiivisten muuttujien muodostaminen

Kuten edellä huomattiin, Excelillä on helppo muodostaa muutaman muuttujan keskiarvoja ja summalausekkeita. Mutta jos yhteenlaskettavien muuttujien lukumäärä on kymmenkunta tai enemmänkin ja jos muuttujat eivät ole peräkkäin, on lausekkeiden muodostaminen vaivaloista ja virheherkkää. Lisäksi puuttuvan tiedon huomioon ottaminen on pelkästään Excelillä toimittaessa hankalaa. Tixelissä onkin menetelmä, jolla voidaan muodostaa summatiivisiä muuttujia.

Summatiivisella muuttujalla tarkoitetaan seuraavia neljä muuttujatyyppeä:

Muuttujien keskiarvo

Muuttujien summamuuttuja

Muuttujien painotettu keskiarvo

Muuttujien painotettu summamuuttuja

Menetelmällä voi lisäksi *kääntää* lausekkeessa olevan muuttujan asteikon eli esimerkiksi asteikon 1-5 arvot muuttuvat siten, että $1 \rightarrow 5$, $2 \rightarrow 4$, $3 \rightarrow 3$, $4 \rightarrow 2$ ja $5 \rightarrow 1$. Tällaista kääntämistä tarvitaan, jos jonkun muuttujan korrelaatiot ovat negatiiviset muihin lausekkeessa olevien muuttujiin.

Havaintomatriisiin työkirjaan lisätään *Kertoimet*-arkki, jossa ensimmäisessä sarakkeessa alkaen A2-solusta on oltavat summatiiviseen lausekkeeseen kuuluvien muuttujien nimet täsmälleen samassa kirjoitusmuodossa kuin havaintomatriisin kolmannella rivillä. Ensimmäisessä sarakkeessa voi olla myös ylimääräisten muuttujien nimiä.

Seuraaviin sarakkeisiin kirjoitetaan muuttujien kertoimet ja ensimmäiselle riville uusien muuttujien nimet. Jos muuttujan asteikko halutaan kääntää, kirjoitetaan kerroin vastalukuna eli negatiivisena.

Esimerkki Kertoimet-arkista.

	A	B	C
1		Mainontamyönteisyys	Muotitietoisuus
2	69 Kirjaan menoni/tuloni		
3	70 Luottokortilla rahoitan yllättäviä mielihaluja		
4	71 Seuraan päivittäin mainoksia	1	
5	72 Saan uusia ideoita mainoksista	1	
6	73 Mainonta on turhaa	-1	
7	74 Awaan yleensä kotiin tulevan mainospostin	1	
8	75 Seuraan muotivirtauksia		1
9	76 Pukeutumiseni huomataan		1
10	77 Ostan harvoin uusia vaatteita		-1
11	78 Pidän ajattomista vaatteista		-1
12	79 Pidän huomiotaherättävistä vaatteista		1
13	80 Tyylikkyys on tärkeintä		1

Kun Tixel-valikosta valitaan *Havaintomatriisin aputoiminnot /Summatiivisten muuttujien muodostamisen-vaihtoehto*, näytölle tulee seuraava määrittelyikkuna:

Ensimmäisellä parametrilla määrätään, halutaanko laskea keskiarvoja vai summia.

Puuttuvien tietojen sietokynnys (%) tarkoittaa prosenttilukua, joka kertoo kuinka paljon muuttujien todellisia arvoja tilastoyksiköllä pitää vähintään olla, jotta summatiivisen muuttujan arvo lasketaan. Muussa tapauksessa arvoksi tulee puuttuva tieto. Esimerkiksi jos summatiivisessa muut-

tujassa on viisi muuttujaa ja puuttuvien tietojen sietokynnys on 50 %, niin vähintään kolmella muuttujalla tulee olla todellinen (ei-puuttuva) arvo, jotta summatiivisen muuttujan arvo lasketaan. Jos muuttujien todellisia arvoja on vähemmän kuin kolme, on summatiivisen muuttujan arvo puuttuva tieto.

Asteikon ala- ja ylärajoja tarvitaan vain jos kertoimissa on negatiivisia arvoja. Tällöin oletetaan, että kaikilla muuttujilla on sama asteikko.

Tixel sijoittaa uudet summatiiviset muuttuvat havaintomatriisiin loppuun seuraavan esimerkin mukaisesti.

BQ	BR	BS	BT
2	2		
-1	-1	-1	-1
79 Pidän huomiotaherä ttävistä vaatteista	80 Tyylikkyys on tärkeintä	Mainontamyönteisyys	Muotitietoisuus
4	4	3.00	3.00
3	3	2.50	2.50
3	5	1.75	3.17
2	4	4.00	3.83
5	5	4.75	3.50
1	-1	3.25	2.40
3	-1	1.75	2.60
1	5	4.75	2.00
1	5	4.50	2.83
4	3	3.25	4.00

2.8. Kahden matriisin yhdistäminen

Kahden havaintomatriisin yhdistämisessä voidaan erottaa kolme eri tapausta:

1. Havaintomatriiseilla on (osittain) samat muuttujat ja eri tilastoyksiköt. Tällainen tilanne syntyy esimerkiksi silloin kun aineistoa tallentaa useampi tallentaja. Eri osamatriisit kopioidaan vain Excelissä allekkain ja näin saadaan yhteinen matriisi.
2. Havaintomatriiseilla on eri muuttujat ja (osittain) samat tilastoyksiköt. Tämäkin tilanne hoituu helposti silloin kun molemmissa matriiseissa on täsmälleen samat tilastoyksiköt samassa järjestyksessä. Tällöin havaintomatriisit vain kopioidaan vierekkäin. Jos tilastoyksiköt eivät ole täysin samoja ja ne ovat eri järjestyksessä, voidaan käyttää Tixelin yhdistämismenetelmää.

3. Kahden havaintomatriisin tiedot voidaan yhdistää myös silloin kun havaintomatriisien muuttujat ja tilastoyksiköt poikkeavat toisistaan lähes kokonaan. Tällainen tilanne on esimerkiksi silloin kun on erikseen henkilöaineisto ja kunta-aineisto. Jos henkilöaineistossa on kunnan nimi tai koodi, voidaan kuntien halutut tiedot liittää henkilöaineistoon Tixelin yhdistämismenetelmällä.

Valitse Tixel-valikosta *Havaintomatriisin aputoiminnot/Kahden havaintomatriisin yhdistäminen*-vaihtoehto, jolloin näytölle ilmestyy seuraava määrittelyikkuna.

Määrittelyt

1. matriisi
Matriisi1.xls
Tunnusmuuttuja
id
Mukaan otettavat muuttujat
x
y
id
z

2. matriisi
Matriisi2.xls
Tunnusmuuttuja
id
Mukaan otettavat muuttujat
xx
yy
zz
id
u

Mukaan otettavat rivit
 Kaikki rivit
 Vain yhteiset rivit

OK
Cancel

Määrittelyikkunan vasemmalla puolella ovat ensimmäisen matriisin tiedot ja oikealla puolella toisen matriisiin tiedot.

Tunnusmuuttujien avulla ohjelma löytää yhteiset tilastoyksiköt. Tunnusmuuttujien nimien ei tarvitse olla samat eikä niiden tarvitse esim. ensimmäisinä muuttujina. Tunnusmuuttujien arvojen tulee olla kuitenkin yksikäsitteiset eli usealla tilastoyksiköllä ei saa olla sama arvo.

Mukaan otettavien muuttujien avulla määrätään, mitkä muuttujat otetaan mukaan yhteiseen matriisiin.

Esimerkki

Tarkastellaan kahta havaintomatriisia: Matriisi1.xls ja Matriisi2.xls.

Matriisi1.xls

	A	B	C	D
1				
2	-11	-12	-13	-14
3	x	y	id	z
4	23	24	1	25
5	24	25	2	26
6	25	26	3	27

Matriisi2.xls

	A	B	C	D	E
1					
2	-1	-2	-3	-4	
3	xx	yy	zz	id	u
4	3	4	5	2	6
5	5	6	7	3	7
6	7	8	9	5	8

Jos yhdistämisparametriksi valitaan *Vain yhteiset rivit*, ja yhdistettyyn matriisiin otetaan mukaan ensimmäisestä matriisista muuttujat x, y ja id sekä toisesta matriisista yy, zz ja id, saadaan tulokseksi seuraava matriisi.

	A	B	C	D	E	F
1						
2	-11	-12	-13	-2	-3	-4
3	x	y	id	yy	zz	id
4	24	25	2	4	5	2
5	25	26	3	6	7	3

Jos sen sijaan yhdistämisparametriksi valitaan *Kaikki rivit*, saadaan seuraava matriisi

	A	B	C	D	E	F
1						
2	-11	-12	-13	-2	-3	-4
3	x	y	id	yy	zz	id
4	23	24	1			
5	24	25	2	4	5	2
6	25	26	3	6	7	3
7				8	9	5

2.9. Esimerkkiaineistot

Opiskelija-aineisto

Opiskelija-aineisto on kerätty vuodesta 1982 lähtien Tampereen yliopiston yhteiskuntatieteellisen tiedekunnan tilastotieteen yleisopintojaksolla. Muuttujia tässä aineistossa on parikymmentä ja tilastoyksiköitä yli 4 000.

Aineiston muuttujat ovat seuraavat.

Muuttujan nimi	Muuttujan sisältö
Nro	Tilastoyksikön järjestysnumero
Sukupuoli	1=mies, 2=nainen
Ikäluokka	1=18-19, 2=20-21, 3=22-24, 4=25-30, 5=31-
Siviilisäätö	1=avioliitossa, 2=avioliitossa, 3=leski tai eronnut, 4=naimaton
Tutkinto	Tavoitteena oleva tutkinto

Koulutus	1=vähemmän kuin yo.,2=yo. tai enemmän
Aloitusvuosi	esim. 82,0= klk 1982, 82,5=slk 1982
Luennot+ harj./vk	Luentojen ja harjoitusten tuntimäärä viikossa
Vaikeata helppoa	Odotukset tilastotieteestä asteikolla 1-5
Kuivaa kiinnostavaa	Odotukset tilastotieteestä asteikolla 1-5
Alkoholi mk/vk	Alkoholiin keskimäärin käytetyt markat viikossa
Alkoholin nautinta paikka	1=kotona, 2=kylässä, 3=kapakassa, 4=muualla
Avio/ avioliitto	1=hyväksyy vain avioliiton, 2=myös avioliiton
Uskollisuus	1=seurustelussa, avo- ja avioliitossa, 2=avo- ja avioliitossa, 3=vain avioliitossa, 4=ei ollenkaan
Presidentin otteet	1=onnistuneet, 2=melko hyvät, 3=keskinkertaiset, 4=melko huonot, 5=epäonnistuneet
Puolue	1=vasemmisto, 2=keskusta, 3=oikeisto, 4=vihreät, 5=sitoutumaton, 6=ei osaa/halua sanoa
Suhde uskontoon	1=uskovainen, 2=ei uskovainen mutta hieman uskonnollinen, 3=ei uskonnollinen mutta suhtautuu myönteisesti, 4=välipitämämön uskonnon suhteen, 5=suhtautuu uskontoon kielteisesti
Kyselyn ajankohta	1=slk 82,.....,24=klk 94

Koska kyselyjä on tehty runsaan kymmenen vuoden ajan, on alkoholiin käytetyt markat *deflatoitu* vuoden 1994 markkoiksi.

Ravintola-aineisto

Tutkimuksen tavoitteena oli selvittää tamperelaisten ravintolakäyttäytymistä. Otos poimittiin satunnaisesti täysi-ikäisistä tamperelaisista, mutta rajattiin haastatteluvaiheessa koskemaan vain niitä, jotka puolen vuoden aikana ovat käyttäneet tamperelaisia ravintolapalveluja. Muuttujia kyselylomakkeessa oli runsaat neljä sataa. Seuraavassa on niistä pieni osa.

1 Sukupuoli

2 Ikäryhmä

Alle 20 v., 20-25 v., 26-30 v., 31-35 v., 36-40 v., 41-45 v., 46-50v., 51-55 v., 56-60 v. ja yli 60 v.

4 Mihin ryhmään lähinnä katsotte kuuluvanne

Työväestö
Toimihenkilö
Johtavassa asemassa oleva
Itsenäinen yrittäjä
Eläkeläinen
Opiskelija/ koululainen
Kotiäiti, -isä
Työtön
Muu

7 Tuloluokka, johon talous kuuluu

- Alle 50 000 mk/v.
- 50 000-100 000 mk/v.
- 100 001-150 000 mk/v.
- 150 001-200 000 mk/v.
- 200 001-250 000 mk/v.
- 250 001-300 000 mk/v.
- Yli 300 000 mk/v.

8 Kuinka usein käy : Seurusteluravintoloissa

- Päivittäin tai lähes päivittäin
- 2-3 kertaa viikossa
- Kerran viikossa
- Kerran kahdessa viikossa
- 1-2 krt kuukaudessa
- Harvemmin

9 Kuinka usein käy : Ruokaravintoloissa

- Vaihtoehdot kuten edellä

10 Kuinka usein käy : Yökerhoissa/ tanssiravintoloissa

- Vaihtoehdot kuten edellä

Ravintolakäyttäytyminen

1=täysin eri mieltä, 2=jokseenkin eri mieltä, 3=ei samaa eikä eri mieltä, 4=jokseenkin samaa mieltä, 5=täysin samaa mieltä

- 21 Iltaa istuessani juon yleensä olutta
 - 22 Pidän ravintoloista, joissa on juhlapaa arvokkuutta
 - 23 Humalaiset häiritsevät minua ravintolassa
 - 24 Illan hauskuus ei mitenkään riipu juomani alkoholin määrästä
- (useita kymmeniä muuttujia)

Ravintolan valintaan vaikuttavat tekijät

1=ei lainkaan, 2=melko vähän, 3=en osaa sanoa, 4=melko paljon, 5=erittäin paljon

SEURUSTELURAVINTOLA

- 184 Ravintolaan ei tarvitse jonottaa
- 185 Hintataso edullinen
- 186 Monipuolinen ulkomaisten oluiden valikoima
- 187 Voi ruokailla
- 188 Soitetaan elävää musiikkia
- 189 Musiikki huolella valittua
- 190 Ravintola pieni ja kodikas
- 191 Sisustus trendikäs
- 192 Sisustus tunnelmallinen
- 193 Ravintola perinteikäs

- 194 Palvelu nopeaa
 195 Palvelu hyväntuulinen/välitön
 196 Tietää löytävänsä tuttuja
 197 Väkeä on paljon
 198 Muu asiakaskunta samantyyppistä kuin minä
 199 Käy kaikenlaista väkeä
 200 Voi pitää hauskaa riehakkaastikin
 201 Käy mielenkiintoisia ihmisiä
 202 Saan halutessani istumapaikan
 203 Henkilökunta on tuttua
 204 Ravintolassa voi kierrellä
 205 Portsari valvoo järjestystä
 206 Jokin muu asia

Kunta-aineisto

Tilastoyksikkönä on kunta ja muuttujina ovat suurimpien puolueiden kannatusprosentit vuoden 2000 kunnallisvaaleissa sekä eräitä sosio-ekonomisia mittareita. Seuraavassa taulukossa on havaintomatriisin alkua.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Kunta-aineisto, v. 2000															
2																
3	Kunta	Nro	Täysikäisten lkm	SDP	KESK	KOK	VAS	VIHR	ASUK KAIDE N KESKI- IKÄ	KOUL UTETT UJEN OSUU S	MAAT ALOU KSIEN OSUU S	JOHT AJAT ALOU KSIEN OSUU S	TYÖV ÄEST ÖTAL EN OSUU S	TALOUKS IEN KESKITUL OT,MK	Kunt aryh män umer o	Maak unta Num ero
4	Alahärmä	004	3792	5,6	55,6	27,9	3,1	0,0	40,6	53	11	6	21	171597	4	14
5	Alajärvi	005	6921	6,8	56,5	15,5	3,7	0,0	38,4	52	7	8	19	169103	3	14
6	Alastaro	006	2491	5,3	50,6	19,3	16,7	3,1	44	51	12	5	18	157594	4	2
7	Alavieska	009	2162	3,0	83,5	0,0	6,8	0,0	38	54	12	6	17	147214	4	17
8	Alavus	010	7597	15,4	50,5	21,8	3,4	2,5	40,4	52	7	6	20	162148	4	14
9	Anjalankoski	015	14010	43,7	29,4	15,4	6,5	1,3	41,4	54	4	4	31	188436	3	8
10	Artjärvi	015	1267	12,1	52,7	19,1	0,0	0,0	44,5	47	18	7	13	152502	4	1
11	Asikkala	016	6794	11,9	27,4	18,9	4,0	13,8	42	55	5	7	20	174087	3	7
12	Askainen	017	729	18,9	56,5	12,7	0,0	0,0	42,5	53	11	5	24	173304	4	2
13	Askola	018	3264	17,1	43,6	16,7	9,0	9,4	38	54	7	8	25	205565	4	1
14	Aura	019	2442	19,4	35,1	24,2	17,1	4,2	37,5	55	4	8	27	184583	4	2
15	Dragsfjärd	040	2938	22,1	0,0	0,0	13,8	0,0	45,9	42	2	5	26	159638	4	2
16	Elimäki	044	6493	17,8	40,6	24,8	1,7	6,8	39,3	56	8	6	21	172546	4	8
17	Eno	045	5637	39,2	24,5	10,8	9,2	0,0	42,5	48	2	4	25	148128	4	12

Terveyskeskus-aineisto

Katso alaluvun 2.2 esimerkkiaineistoa. Siinä olleiden muuttujien lisäksi mitattiin mm. sukupuoli ja jonotusaika lääkärin vastaanotolle minuutteina.

Ympäristö-aineisto

Tämä on kansainvälisen ympäristötutkimuksen Suomen aineisto vuodelta 2000. Aineistoa hyödyntävien tulee viitata siihen seuraavasti:

International social survey programme: ympäristö II, 2000: Suomen aineisto [Elektroninen aineisto]. Helsinki: Tilastokeskus [aineistonkeruu], 2000. Tampere: Yhteiskuntatieteellinen tietoarkisto [jakaja], 2001.

Lisätietoja aineistosta saa nettiosoitteesta <http://www.fsd.fi/aineistot.html>.

3. Lisäluokitukset ja viitearvot

Havaintomatriisiin liitettyjen luokitusten lisäksi voidaan määrittellä vaihtoehtoisia luokituksia **luokitusarkin** avulla. Esimerkiksi ristiintaulukoinnissa muuttujia ehkä halutaan käsitellä harvemmalla luokituksella, jotta taulukoiden koko pysyisi kohtuullisena.

Joskus taulukoihin halutaan muuttujien *viitearvoja*, jotka voivat olla esimerkiksi muuttujien keskiarvoja edellisvuotisesta tutkimuksesta tai jonkin standardin mukaisia arvoja. Tällaiset viitearvot merkitään **viitearvojen arkille**.

3.1. Luokitusarkki

Alaluvussa 2.2 esitettiin viisi tapaa määrittellä muuttujien luokkatunnukset ja luokkarajat. Nämä määrittelytiedot voidaan välittää ohjelmalle kahdella eri tavalla: joko ne sijoitetaan havaintomatriisiin viimeisen rivin alapuolelle (kts. 2.2) tai havaintomatriisin erilliselle luokitusarkille, jota käsitellään tässä luvussa.

Näitä kahta tapaa voidaan käyttää rinnakkain. Muuttujien luokkatunnukset ja perusluokitukset on syytä kirjoittaa havaintomatriisin alapuolelle ja mahdolliset *vaihtoehtoiset* määrittelyt sijoittaa luokitusarkille.

Milloin sitten vaihtoehtoisia määrittelyitä tarvitaan?

Yleisesti ottaen yksiulotteisissa jakaumissa käytetään alkuperäisiä luokituksia, kun sen sijaan ristiintaulukoinneissa luokkien määrää usein vähennetään, jotta taulukoiden koko ei paisuisi liian suureksi. Esimerkiksi iästä halutaan perusjakauma viiden vuoden luokkavälin pituudella, mutta ristiintaulukoinneissa käytetään kolmeluokkaista jakoa: nuoret, keski-ikäiset ja vanhat.

Luokitusarkille annetaan nimi CLASSES, jotta ohjelma osaa erottaa luokitusarkin data-arkista. Arkin nimi on arkin alareunassa ja



se muutetaan merkittävällä vanha nimi ja kirjoittamalla päälle uusi nimi. Luokitusarkki ei saa olla työkirjan ensimmäinen arkki, koska työkirjan tulee alkaa data-arkilla. Jos havaintomatriisi-arkin nimi on *Data*, tätä rajoitusta ei ole.

Luokitusarkilla määrittelytiedot kirjoitetaan sarakkeille alkaen A-sarakkeesta ja kolmannesta rivistä. Määrittelytiedot eivät nyt liity mihinkään tiettyyn muuttu-jaan, vaan ne (eli vastaavat sarakkeet) numeroidaan juoksevasti ykkösestä eteenpäin. Järjestysnumero sijoitetaan ensimmäiselle riville.

	A	B	C
1	1	2	3
2			
3	20	"30/Nuoret"	"1=Täysin samaa mieltä"
4	40	"60/Keski-ikäiset"	"2=Jokseenkin samaa mieltä"
5	60	"100/Vanhat"	"4=Jokseenkin eri mieltä"
6	100		"5=Täysin eri mieltä"
7			

Tässä esimerkissä on määritelty kolme luokitusta. Kahta ensimmäistä on tarkoitus soveltaa ikä-muuttujaan. Ensimmäinen luokitus tuottaa seuraavat luokat tulostaulukoihin: -20, 21-40, 41-60 ja 61-100 täsmälleen tässä muodossa. Toisessa luokituksessa tulostaulukoihin tulevat luokkatunnukset *Nuoret*, *Keski-ikäiset* ja *Vanhat*, jotka vastaavat ikäluokkia -30, 31-60 ja 61-100

Ohjelmalle välitetään tieto siitä, mitä vaihtoehtoista luokitusta muuttujaan halutaan käyttää, kirjoittamalla havaintomatriisin ensimmäiselle riville ko. muuttujaa vastaavaan sarakkeeseen luokituksen järjestysnumero. Jos tämä solu on tyhjä, käytetään muuttujan alkuperäisiä määrittelytietoja.

Taulukko 3.1.1. Esimerkki vaihtoehtoisten luokitusten käytöstä

	A	B	C	D
1		2	3	3
2				
3	Nro	Ikä vuosina	Väittä ^{mä} 1	Väittä ^{mä} 2
4	1	36	2	1
5	2	18	5	3
262	259	44	3	4
263	260	51	1	5
264				
265		20	Täysin samaa mieltä	Täysin samaa mieltä
266		30	Jokseenkin samaa mieltä	Jokseenkin samaa mieltä
267		40	Ei osaa sanoa	Ei osaa sanoa
268		50	Jokseenkin eri mieltä	Jokseenkin eri mieltä
269		60	Täysin eri mieltä	Täysin eri mieltä
270		100		
271				

Tässä esimerkissä havaintomatriisissa ovat muuttujat *Nro*, *Ikä vuosina*, *Väittä^{mä}1* ja *Väittä^{mä}2*, joiden määrittelytiedot on sijoitettu havaintomatriisin alapuolelle alkaen riviltä 265. Koska ensimmäisellä rivillä on solussa B1 luku 2, käytetään ikään luokitusta 2 eli luokat ovat *Nuoret*, *Keski-ikäiset* ja *Vanhat*. Väittä^{mä}1

miin 1 ja 2 käytetään luokitusta 3, joka jättää *Ei osaa sanoa*-vaihtoehdon pois tulostaulukoista.

3.2. Viitearvot

Muuttujan viitearvo voi olla esimerkiksi kyseisen muuttujan keskiarvo edellisen vuoden tutkimuksessa tai jokin tavoitearvo. Muuttujien viitearvoja varten havaintomatriisi-työkirjaan lisätään **RefValues**-niminen arkki, johon viitearvot sijoitetaan. Muuttujien nimet sijoitetaan A-sarakkeeseen alkaen solusta A3. Viereen B-sarakkeeseen sijoitetaan viitearvot ja soluun B1 viitteen *tunnus*, joka tulee esille tulostaulukossa ja kuviossa. Viitteen tunnus voi olla mikä tahansa teksti, joka kuvaa kyseistä asiaa, esim. v. 2000 keskiarvo, tavoite, viite jne. Viereisessä esimerkissä viitteinä ovat muuttujien keskiarvot vertailuaineistossa.

	A	B
1		Vertailuaineisto
2	Työn määrälliset vaatimukset	3.4
3	Päätöksenteon vaatimukset	3.7
4	Oppimisvaatimukset	2.8
5	Työroolin selkeys	4.2
6	Ristiriidat työssä	2.5
7	Työn haasteellisuus	4.1
8	Vaikutusmahdollisuudet päätöksenteossa	3
9	Vaikutusmahdollisuudet työtahtiin	2.9

RefValues-arkilla ei tarvitse olla kaikkien havaintomatriisissa olevien muuttujien nimiä, vaan ainoastaan niiden, joiden viitearvot halutaan ottaa mukaan tuloksiin. Toisaalta muuttujien nimien tulee RefValues-arkilla olla *täsmälleen* samassa muodossa kuin havaintomatriisissa, joten kannattaa kopioida muuttujien nimet havaintomatriisista RefValues-arkille Transpose-ominaisuuden avulla. Edellisessä taulukossa on esimerkki RefValues-arkista.

Viitearvot tulevat niiden menetelmien tulostuksissa, joissa lasketaan keskiarvoja.

4. Mittaaminen

4.1. Mitta-asteikkotyypit

Tilastoyksikön arvo muuttujalla ilmaistaan yleensä luvulla. Luvuilla on erilaisia ominaisuuksia, jotka jakavat muuttujat neljään **mitta-asteikkotyyppiin**.

1) **Luokitus**. Kokonaisluvuilla ilmaistaan mihin luokkaan tilastoyksikkö kuuluu. Esimerkkeinä

- sukupuoli (1=nainen, 2=mies),
- lääni, johon kunta kuuluu

Tässä tapauksessa lukujen käyttö ei tuo mitään "lisäarvoa", vaan yhtä hyvin pärjättäisiin kirjaimilla (esim. N ja M, läänien nimet). Pitkien tekstien syöttäminen on kuitenkin hidasta ja jotkut tilastolliset tietokoneohjelmat edellyttävät lukujen käyttöä kaikissa tapauksissa.

2) **Järjestys**. Kokonaislukujen järjestysominaisuutta hyödynnetään, kun mittaus asettaa tilastoyksiköt järjestykseen mitatun ominaisuuden suhteen. Seuraavassa kaksi tyyppistä esimerkkiä, jotka poikkeavat selvästi toisistaan:

a) Tilastoyksiköille annetaan järjestysnumerot: 1=paras, 2= toiseksi paras, jne.

b) On käytössä ennalta määrätty luokat, jotka voidaan asettaa järjestykseen. Esim. kyselyssä esitetään väite "Kaikkien alkoholijuomien vapaa myynti tulisi sallia" ja vastausvaihtoehdot ovat

1	täysin samaa mieltä
2	jokseenkin samaa mieltä
3	en osaa sanoa, yhdentekevää
4	jokseenkin eri mieltä
5	täysin eri mieltä

Kysymykseen vastanneet henkilöt asettuvat myönteisyys-kielteisyys järjestykseen. Yksittäisen luokan sisällä ei voida tehdä eroa vastaajien asenteissa.

3) **Tasavälisyys**. Kokonaisluvuilla on tasavälisyys-ominaisuus eli $2-1=3-2$ jne. Tätä ominaisuutta tarvitaan, kun mitataan kvantitatiivisia muuttujia, joiden arvot perustuvat mittayksiköiden käyttöön. Esimerkkeinä olkoot lämpötilan mittaaminen Celsius-asteina ja tuotannon 12 kuukauden muutosprosentti. Näissä molemmissa tapauksissa muuttuja voi saada myös negatiivisia arvoja eli nolla ei tarkoita absoluuttista nollakohtaa.

4) **Nollapist**e. Luvulla nolla voidaan ilmaista muuttujan teoreettinen minimiarvo. Esimerkkeinä pituus, paino, lämpötilan mittaus Kelvin-asteina.

Lukujen neljä ominaisuutta jakavat muuttujat neljään luokkaan riippuen siitä mitä ominaisuuksia mittauksessa on käytetty.

Luokitus	Laatuero- eli nominaaliasteikko
Luokitus ja järjestys	Järjestys- eli ordinaaliasteikko
Luokitus, järjestys ja tasavälisyys	Välimatka- eli intervalliasteikko
Luokitus, järjestys, tasavälisyys ja nollapiste	Suhdeluku- eli suhdeasteikko

Laatuero- ja järjestysasteikot ovat **kvalitatiivisia asteikkoja**.

Välimatka- ja suhdeasteikot ovat **kvantitatiivisia asteikkoja**.

Aritmeettiset peruslaskutoimitukset (+, -, *, /) ovat sallittuja vain kvantitatiivisille asteikoille. Käytännössä kuitenkin peruslaskutoimituksia hyödyntäviä menetelmiä sovelletaan myös järjestysasteikolle.

4.2. Järjestysluvun käyttö mittaamisessa

Järjestyslukuja voidaan käyttää eri tekijöiden asettamisessa tärkeysjärjestykseen. Seuraavassa esimerkissä on ravintolan tekemä pieni selvitys, jossa pyydetään asiakasta valitsemaan kolme tärkeintä tekijää ja asettamaan ne tärkeysjärjestykseen.

Asia	Ensimmäiseksi. tärkein	Toiseksi tärkein	Kolmanneksi tärkein.
Ruuan terveellisyys			
Ruuan maku			
Ravintolamiljö			
Palvelun nopeus			
Ruuan hinta			
Palvelu ystävällisyys			

Tällainen aineisto voidaan tallentaa kolmella tavalla havaintomatriisiin muuttujiksi (sarakkeiksi).

1. Määritellään kolme muuttujaa (1., 2. ja 3. tärkein), jolloin asiat ovat muuttujien ”arvoja” ja kyseessä on laatueroasteikollinen mittaaminen.
2. Kuusi asiaa on muuttujia ja muuttujien arvot ovat järjestyslukuja 1-3.

3. Kuten tapaus kaksi mutta käännetään järjestys siten, että tärkein asia saa arvon kolme, toiseksi tärkein arvon kaksi ja kolmanneksi tärkein arvon yksi. Kolme mainitsematonta asiaa saavat arvo nolla. Lisäksi vastaaja jättää kokonaan vastaamatta näihin kysymyksiin, niin muuttujat saavat puuttuvan tiedon arvon. Näin määritellyistä muuttujista voidaan laskea tietyin varauksin esimerkiksi keskiarvoja. Voidaan selvittää esimerkiksi, miten sukupuoli tai ikä vaikuttaa siihen, miten vastaajat asettavat eri asioita tärkeysjärjestykseen.

Aineiston syötön kannalta kolmas vaihtoehto on hankalin, koska siinä joudutaan kääntämään järjestysluvut. Tixel-ohjelmassa on kuitenkin menetelmä, joka muuntaa ensimmäisellä tai toisella tavalla tallennetun aineiston kolmannen esitystavan mukaiseksi.

5. Yksiulotteinen jakauma

Kun tilastoaineistoa (=havaintomatriisia) aletaan jalostaa, ensimmäinen kysymys on, miten muuttujien arvot vaihtelevat: mitkä ovat usein esiintyviä arvoja ja mitkä harvoin esiintyviä. Esimerkkeinä olkoot vaikkapa muuttujat *siviilisääty* ja *alkoholiin käytetty markkamäärä*. Tähän antaa vastauksen yksiulotteinen jakauma.

Viereisessä taulukossa on näiden kahden muuttujan arvot viidellätoista ensimmäisellä henkilöllä. Siviilisääty on kvalitatiivinen muuttuja ja alkoholiin käytetty markkamäärä kvantitatiivinen muuttuja. Yksiulotteisen jakauman muodostaminen näillä kahdella muuttujatyypillä eroaa toisistaan.

Siviilisääty	Alkoholi mk/vk
4	60
1	30
4	50
4	10
4	20
1	0
4	50
4	70
4	15
4	100
4	80
4	20
4	40
4	-1
4	0

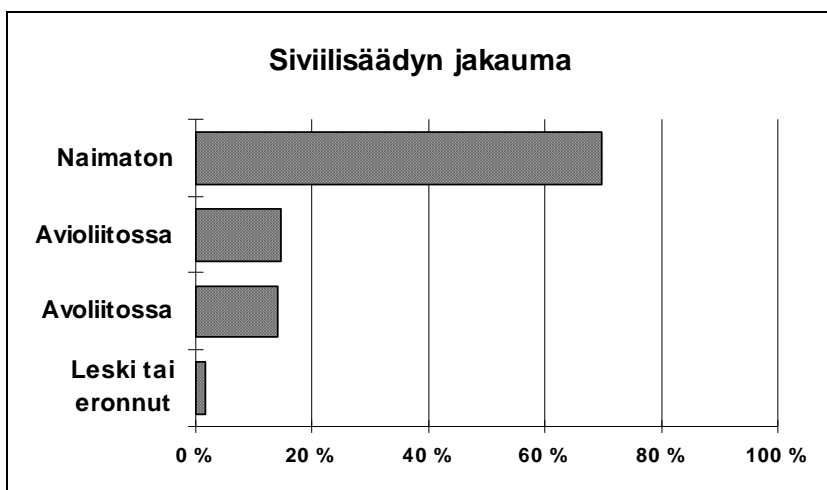
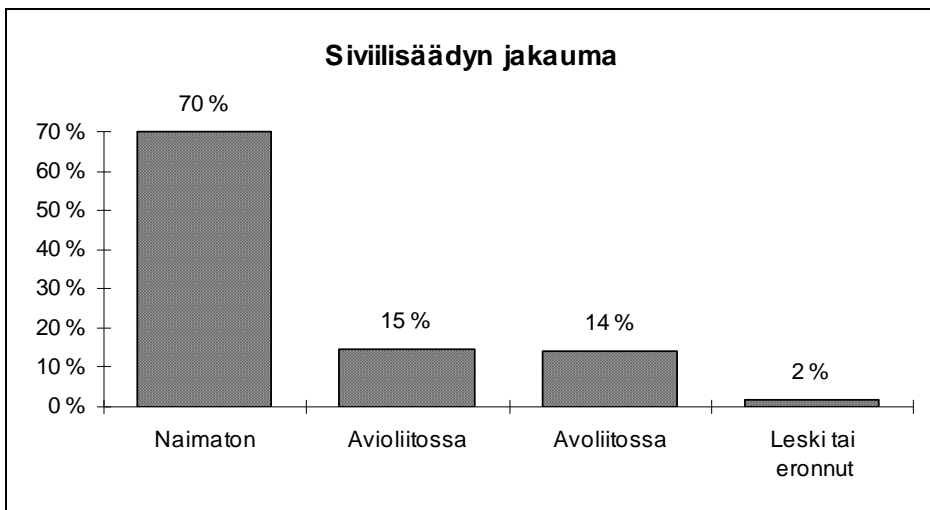
5.1. Kvalitatiivinen muuttuja

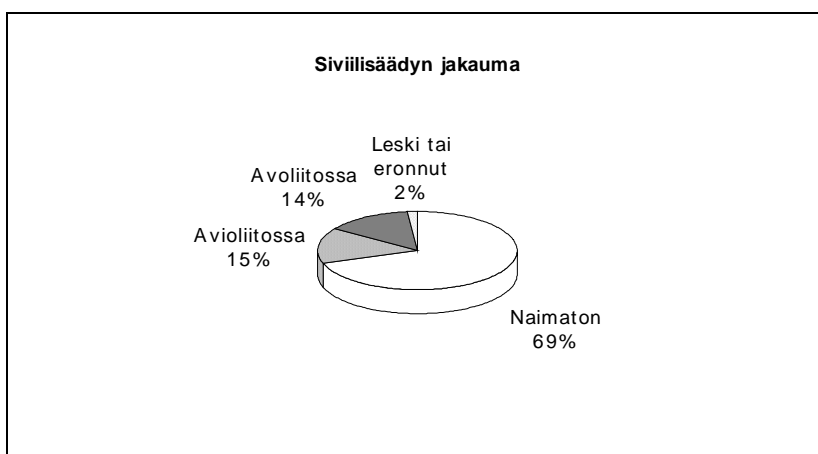
Kvalitatiivisen muuttujan kohdalla muuttujan arvot käyvät sellaisinaan jakauman *luokiksi*. Arvojen sijaan vain merkitään jakaumataulukkoon selväkieliset nimet. Näin saatua taulukkoa kutsutaan myös **suoraksi jakaumaksi**. Jokaiselle luokalle lasketaan, kuinka monta kertaa se esiintyy havaintomatriisin ko. sarakkeessa. Käytännössä laskenta tehdään tietysti tietokoneella. Seuraavassa taulukossa on opiskelija-aineistosta laskettu siviilisäädyn suora jakauma:

Siviilisääty	Lkm	%
Avioliitossa	594	15 %
Avoliitossa	568	14 %
Leski tai eronnut	63	2 %
Naimaton	2833	70 %
Yhteensä	4058	100 %

Lukumäärää kutsutaan myös luokan (*absoluuttiseksi*) frekvenssiksi. Prosentuaaliset frekvenssit saadaan, kun frekvenssit jaetaan frekvenssien summalla ja kerrotaan sadalla.

Graafisesti yksilolotteista jakaumaa voidaan kuvata usealla tavalla. Seuraavassa esitetty kolme tapaa: (pysty)pylväsdiagrammi, (vaaka)palkkidiagrammi ja ympyräesitys. Luokkien järjestys on muutettu frekvenssien suuruusjärjestyksen mukaiseksi.

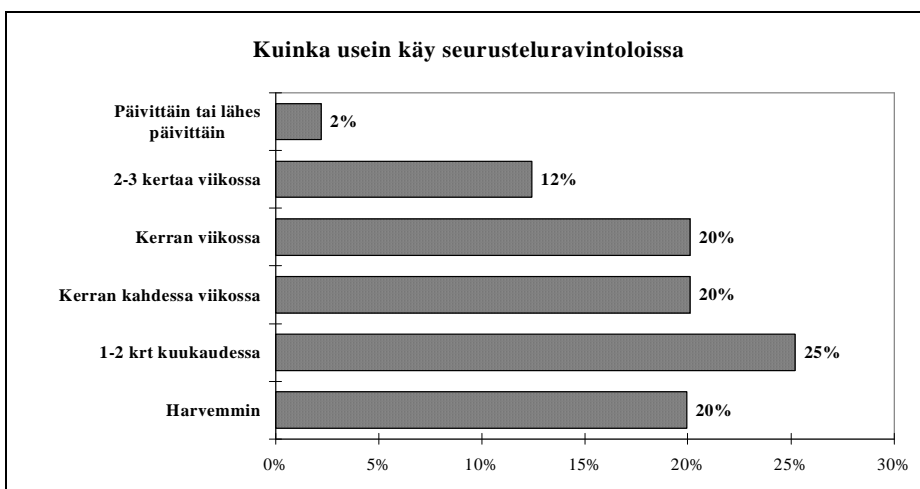




Valintaan näiden kolmen kuvatyypin välillä vaikuttavat mm. seuraavat seikat:

- 1) Jos luokkien tunnukset ovat pitkiä, on syytä käyttää palkkikuvaa.
- 2) Jos luokkia on paljon ($\geq 7-8$), ei ole syytä käyttää ympyräkuvaa, koska luokkien teksteille ei silloin jää tarpeeksi tilaa.
- 3) Ympyräkuva on havainnollinen, jos yhden luokan osuus on lähellä puolta.
- 4) Ihmissilmä pystyy erottamaan paremmin pylväiden väliset erot kuin palkkien väliset erot.

Tarkastellaan vielä yhtä esimerkkiä. Seuraavassa kuvassa on ravintola-aineistosta laskettu kysymyksen ”Kuinka usein käy seurusteluravintoloissa” suora jakauma.



Huomataan, että yleisin luokka on 1-2 kertaa kuukaudessa.

5.2. Kvantitatiivinen muuttuja

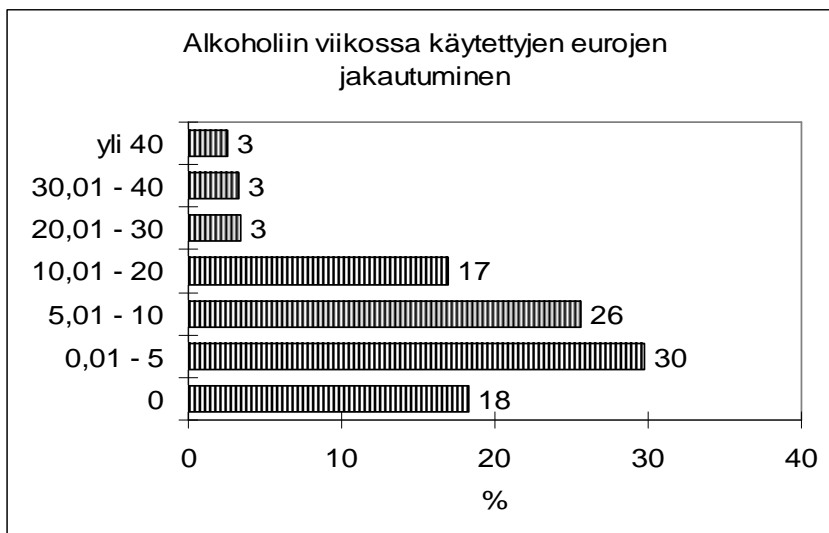
Kvantitatiivisen muuttujan kohdalla voitaisiin toki toimia samalla tavalla kuin kvalitatiivisen muuttujan tapauksessa. Kuitenkin muuttujan eri arvoja on yleensä niin paljon, että selkeän yleiskuvan saamiseksi jakaumasta on arvoja yhdisteltävä luokiksi. Luokat määritellään **luokkarajojen** avulla. Katso luokituksesta tarkemmin luvusta 2.2 (Tunnustiedot ja luokkarajat).

Seuraavassa esimerkissä on viikossa alkoholiin käytettyjen eurojen jakauma (opiskelija-aineisto).

Alkoholieurot/vk	<i>Lkm</i>	<i>%</i>	<i>Kum. lkm</i>	<i>Kum-%</i>
0	249	18	249	18
0,01 - 5	403	30	652	48
5,01 - 10	348	26	1000	74
10,01 - 20	230	17	1230	91
20,01 - 30	47	3	1277	94
30,01 - 40	45	3	1322	97
yli 40	35	3	1357	100
Yht.	1357	100	1357	100

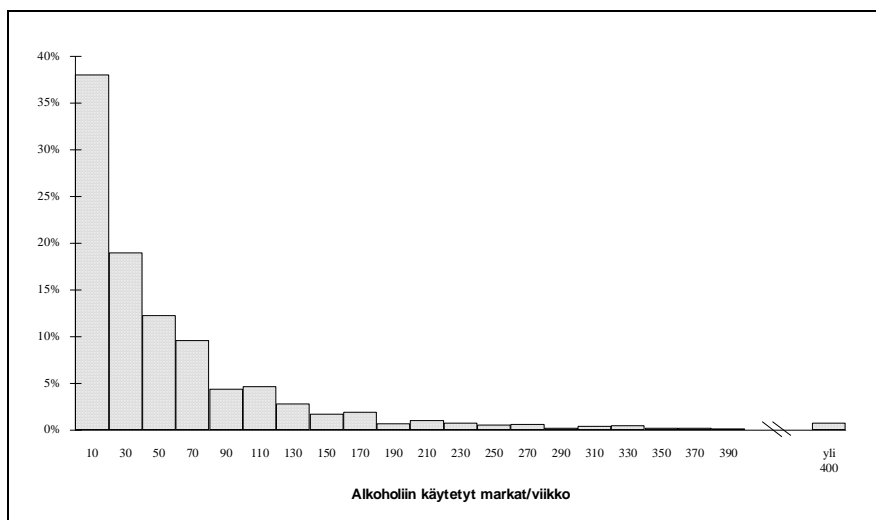
Absolutisteja on aineistosta 18 % ja yli 20 euroa viikossa alkoholiin käyttäviä 6 %.

Graafisesti tämän jakauman kuvaamisessa voidaan käyttää pylväs- tai palkkiesitystä.



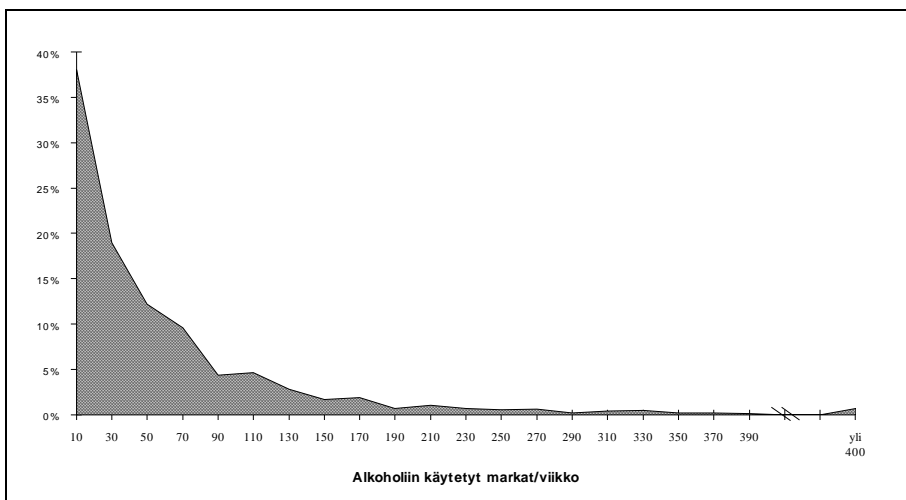
Kun verrataan tätä kuvaa edellä olleeseen taulukkoon, huomataan luokkien järjestystä koskeva ristiriita: kuvassa luokkien markkamäärät kasvavat alhaalta ylöspäin. Luokkien järjestys kuvassa on kuitenkin helposti käännettävissä (kts. 11.4: Kuvion muokkaaminen).

Tasavälisen luokituksen tapauksessa on parempi käyttää pylväskuvaa ja vielä siten, että pylväät ovat kiinni toisissaan. Seuraavassa kuvassa on edellisen muuttujan tasavälinen jakauma, kun luokkavälin pituus on 20 mk. Vaakasuoran akselin arvot ovat luokkien luokkakeskuksia.



Kuvasta käy selvästi ilmi jakauman **vinous** eli suurin osa käyttää kohtuullisen vähän rahaa alkoholiin, minkä lisäksi on harvalukuinen suurkuluttajien joukko.

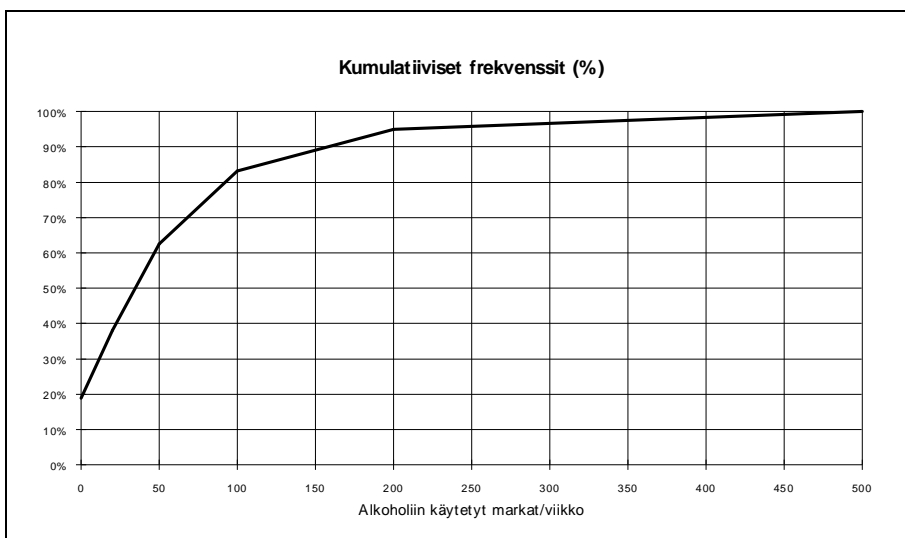
Sama kuva voidaan esittää myös aluekuvana.



Kumulatiiviset frekvenssit kertovat, kuinka monella tilastoyksiköllä on ollut pienempi tai yhtä suuri arvo kuin luokan yläraja. Esimerkiksi, kuinka moni käyttää rahaa alkoholiin 50 mk tai vähemmän. Tämän tiedon avulla saadaan luonnollisesti tietää myös niiden lukumäärä, jotka käyttävät rahaa alkoholiin enemmän kuin 50 mk.

Kumulatiiviset frekvenssit voidaan esittää myös prosentuaalisessa muodossa.

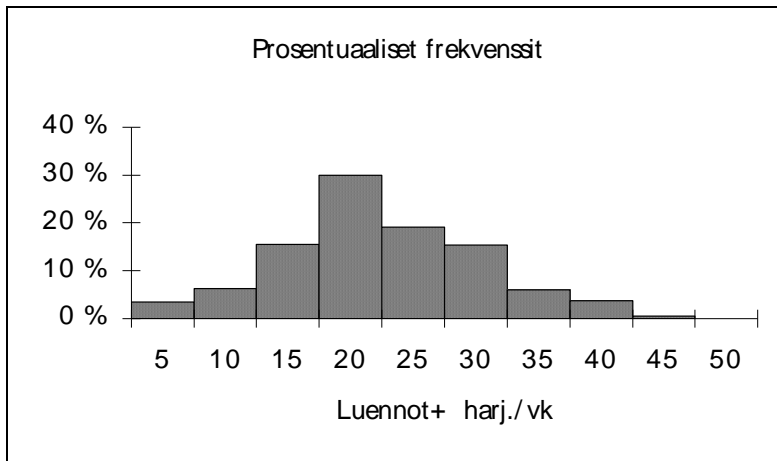
Kumulatiiviset frekvenssit kuvataan graafisesti *summakäyrän* avulla. Summakäyrästä voidaan arvioida muuttujan *prosenttipisteet eli fraktiilit*. Esimerkiksi 60 %:n prosenttipiste saadaan siten, että pystysuoran akselin 60 %:n kohdalta edetään vaakasuoraan niin pitkälle, kunnes saavutetaan summakäyrä. Tämän pisteen x-koordinaatti on likiarvo 60 %:n prosenttipiste.



Toisena esimerkkinä kvantitatiivisen muuttujan jakaumasta on luentoihin ja harjoituksiin viikossa käytetyn tuntimäärän jakauma. Luokitus on tasavälinen ja luokkavälin pituus on viisi tuntia (opiskelija-aineisto).

Luokan yläraja	Luokkakeskus	Lkm	%	Kum. lkm	%-kum.
7,5	5	140	3 %	140	3 %
12,5	10	255	6 %	395	10 %
17,5	15	627	15 %	1022	25 %
22,5	20	1214	30 %	2236	55 %
27,5	25	773	19 %	3009	74 %
32,5	30	623	15 %	3632	90 %
37,5	35	245	6 %	3877	96 %
42,5	40	150	4 %	4027	99 %
47,5	45	21	1 %	4048	100 %
52,5	50	1	0 %	4049	100 %

Graafisesti taulukon prosentteja voidaan havainnollistaa seuraavalla pylväskuvalla.



5.3. Tixel-ohjelman käyttö

Tixel-ohjelman valikosta valitaan vaihtoehto **Jakaumat/Yksiuotteinen jakauma**. Ohjelma pyytää tekemään määrittelyt seuraavalla valintaikkunalla.

Yksiuotteisen jakauman määrittely

Muuttuja(t)

Keskiarvo

%-jakaumat yhteen taulukkoon

Jakaumien summajakauma

Luokituksia ei käytetä

Nro
 [q1] Sukupuolenne?
 [q2] Syntymävuotenne?
 Ikä
 [q3] Oletteko tällä hetkellä...
 [q4] Kuinka monta henkilöä asuu samass
 [q5] Kotitaloutenne 6-17-vuotiaiden her
 [q6] Kotitaloutenne alle 6-vuotiaiden las
 [q7] Asutteko?
 [q8a] Yksityinen yritysliiketoiminta on paras t
 [q8b] Valtiovallan tehtävä on pienentää
 [q9] Mikä seuraavista on tärkein asia, jo
 [q10] Entä mikä mielestänne on toiseksi
 [q11a] Luotamme liikaa tieteeseen ja lita
 [q11b] Kaiken kaikkiaan tieteestä on ene
 [q11c] Nykyaikainen tiede ratkaisee ymp
 [q12a] Olemme liian huolissamme ympäri
 [q12b] Melkein kaikki nykyisessä elämär
 [q12c] Ihmiset ovat liian huolissaan siitä
 [q13a] Ympäristönsuojelu edellyttää Suc
 [q13b] Eläimiä voidaan käyttää lääketiet
 [q13c] Talouskasvu vahingoittaa aina yr
 [q13d] Maapallo ei yksinkertaisesti kestä
 [q14] Rengastakaa vaihtoehto, joka on
 [q15a] Maksamaan paljon nykyistä korke

Samalla ajolla voidaan laskea usean muuttujan jakaumat (usean muuttujan valinnasta kts. 2.3). Lisäksi valintaikkunassa on painikkeet tilastografiikalla, ehto-muuttujien käytölle (kts. 2.4) ja painomuuttujalla (kts. 5.5) sekä neljä muuta valintaa.

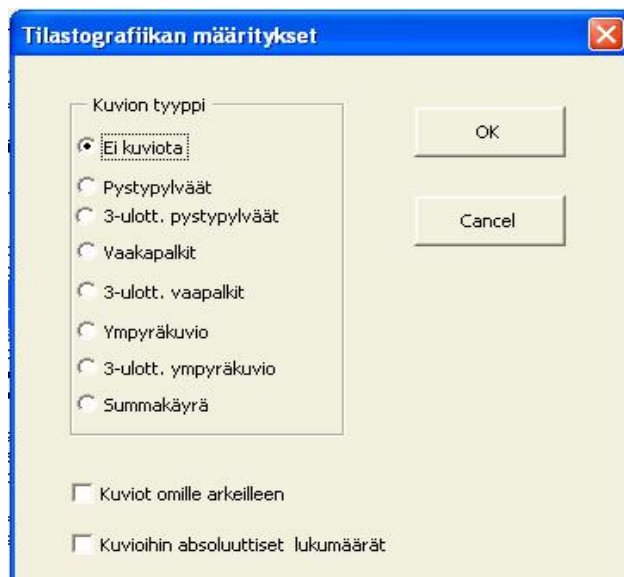
Keskiarvo. Jokaisen valitun muuttujan keskiarvo merkitään taulukon alareunaan.

%-jakaumat yhteen taulukkoon. Ohjelma sijoittaa eri muuttujien prosenttija-kaumat allekkain yhteen taulukkoon, jolloin jakaumien keskinäinen vertailu hel-pottuu (kts. 5.4). Taulukosta tehdään automaattisesti myös kuvio.

Jakaumien summajakauma. Esimerkiksi kuvatkoon kolme jakaumaa, miten pal-jon banaaneja, appelsiineja ja omenia asiakkaat ovat ostaneet. Tällöin summaja-kauma kertoo hedelmäostoksen jakauman. Kyse ei siis ole summamuuttujan ja-kaumasta.

Luokituksia ei käytetä. Ensimmäinen tarkistusajo kannattaa tehdä siten, että luokituksia ei käytetä, jolloin ohjelma käyttää alkuperäisiä arvoja luokkina.

Klikkaamalla *Tilastografiikka*-painiketta saadaan viereinen määrittelyikkuna, jolla määritellään kuvion tyyppi. Kuvio voi sijaita joko taulukon vie-ressä (oletusvaihtoehto) tai omalla grafiikka-arkilla. Kuvio voi perus-tua joko prosenttilukuihin (oletusvaihtoehto) tai absoluuttisiin lukumää-riin.



5.4. Usean jakauman kuvaaminen

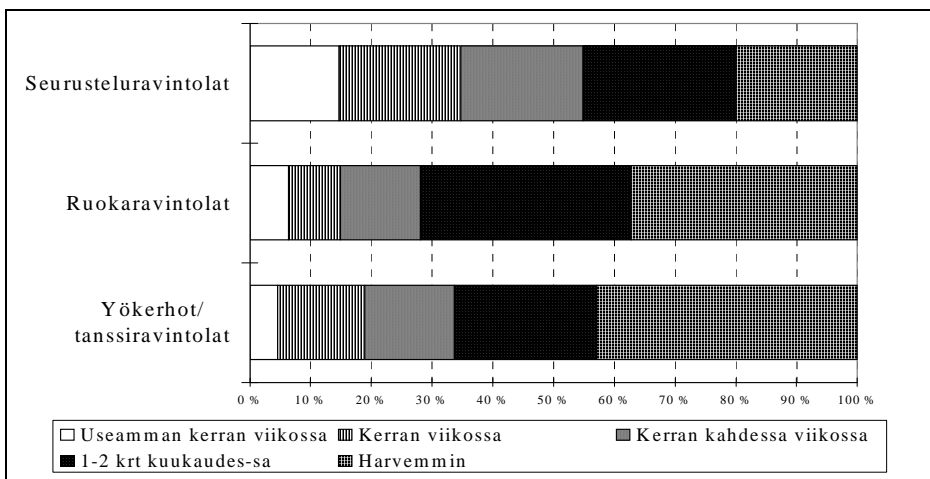
Jos usealla muuttujalla on sama luokitus, saattaa olla havainnollista kuvata muuttujien prosentuaalisia osuuksia yhden taulukon ja kuvion avulla. Tätä varten

valintaikkunassa on ”%-jakaumat yhteen taulukkoon”-valintaruutu. Yhteistaulukko ja kuvio tulevat tulostiedoston *One_table*-arkille.

Ravintola-aineistossa on kolme muuttujaa, joilla mitataan, kuinka usein henkilö käy 1) seurusteluravintoloissa, 2) ruokaravintoloissa ja 3) yökerhoissa/tanssiravintoloissa. Seuraavassa taulukossa on nämä kolme jakaumaa.

	Useamman kerran viikossa	Kerran viikossa	Kerran kahdessa viikossa	1-2 krt kuukau- dessa	Harvem- min
Seurusteluravintolat	15 %	20 %	20 %	25 %	20 %
Ruokaravintolat	6 %	8 %	13 %	35 %	37 %
Yökerhot/tanssiravintolat	5 %	14 %	15 %	23 %	43 %

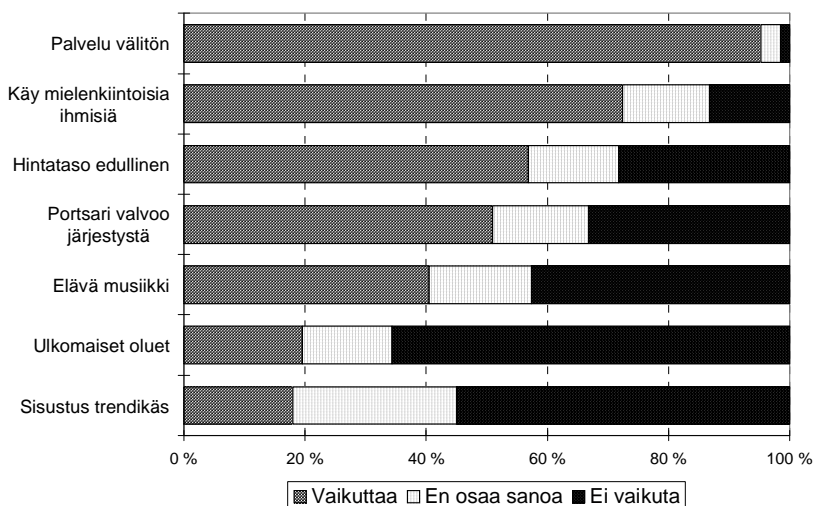
Ohjelma tuottaa taulukosta seuraavan graafisen esityksen.



Huomataan että seurusteluravintoloissa käydään selvästi useammin kuin kahdessa muussa ravintolatyypissä.

Tällaisen kuvion hahmottamista helpottaa, jos vastaavan taulukon rivit lajitellaan (“sortataan”) jonkin luokan prosenttien mukaan. Lajittelu tehdään **Data/Sort**-komennolla (kts. 10.5).

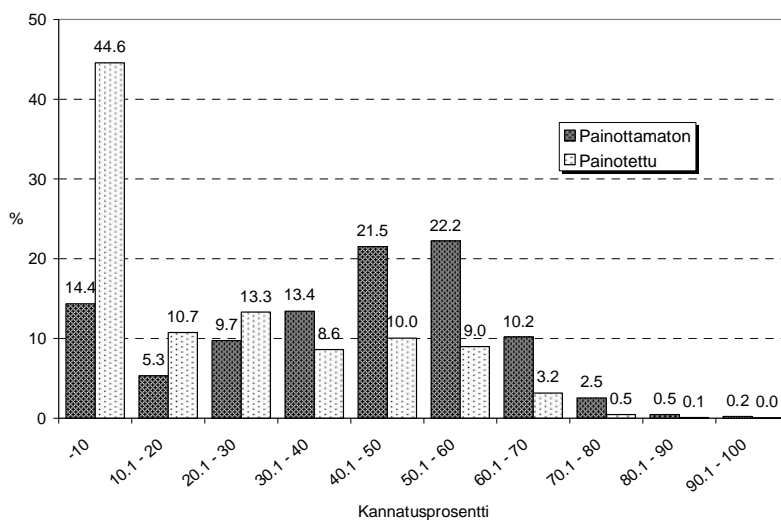
Tarkastellaan vielä toista esimerkkiä, jossa on havainnollistettu eri tekijöiden vaikutusta seurusteluravintoloiden valintaan. Alkuperäinen viisiluokkainen asteikko on Tixel-ajossa muutettu kolmiluokkaiseksi, muuttujat lajiteltu *Vaikuttaa*-luokan prosenttien mukaan ja luokkien järjestys muutettu siten, että *Vaikuttaa*-luokka on ensimmäinen ja *Ei vaikuta*-luokka viimeinen (kts 11.3).



5.5. Painomuuttuja

Painomuuttujan avulla voidaan ottaa huomioon tilastoyksikön "suuruuden" vaihtelu. Esimerkkinä tarkastellaan keskustapuolueen kannatusjakaumaa kunnissa vuoden 2000 kunnallisvaaleissa. Kun jakaumaa lasketaan painottamattomana siten, että jokainen kunta on samanarvoinen, on moodiluokka 50.1-60. Käytettäessä painomuuttujana täysikäisten lukumäärää on moodina luokka -10. Seuraavassa kuviossa jakaumat ovat vierekkäin.

Kuvio 5.5.1. Keskustapuolueen kannatusjakauma kunnissa



Jakaumien ero johtuu siitä, että Suomessa on paljon pieniä kuntia, joissa keskustapuolueen kannatus on voimakasta. Kun painomuuttujana käytetään kunnan täysikäisten lukumäärää, saadaan jakauma joka vastaa sellaista keinotekoisista tilannetta, jossa koko maan väestö on jaettu esim. sadan hengen rypäisiin ja näistä laskettu keskustapuolueen kannatusjakauma.

5.6. Multiresponse-muuttujat

Yksiulotteinen jakauma voidaan laskea myös multiresponse-muuttujista eli muuttujista, joilla tilastoyksiköt voivat saada useita arvoja (kts 2.5). Tämäntyyppisissä muuttujissa ei luokkia voida yhdistää, vaan ohjelma määrää luokiksi kaikki ne arvoyhdistelmät, mitä vastauksissa esiintyy. Tulostaulukko sijoitetaan *Multiresponse*-arkille.

Tämän tulostaulukon lisäksi Tixel muodostaa toisen taulukon, jossa luokkina ovat muuttujan alkuperäiset luokat. Tässä taulukossa prosentit on laskettu siten, että prosenttilukujen jakajana on käytetty tilastoyksiköiden lukumäärää. Tämän vuoksi prosenttilukujen summa on yleensä suurempi kuin sata. Jakajana käytetty luku on merkitty viimeiselle riville (N:).

Esimerkki 1. Seuraavassa on alaluvun 2.5 esimerkkiaineiston molemmat jakaumat.

Missä ravintoloissa käynyt puolen vuoden aikana?			Missä ravintoloissa käynyt puolen vuoden aikana?		
	<i>Lkm</i>	%		<i>Lkm</i>	%
1	1	13	Ravintola A	3	38
2	2	25	Ravintola B	6	75
3	1	13	Ravintola C	4	50
1+2	1	13	Yht	13	163
1+2+3	1	13	N:	8	
2+3	2	25			
Yht.	8	100			

Esimerkiksi luokka yksi esiintyy kolmessa yhdistelmässä, joiden lukumäärien summa on kolme. Siten oikeanpuoleisessa taulukossa luokan yksi (= Ravintola A) frekvenssi on kolme.

Esimerkki 2. Taloustutkimus Oy:n Omnibus-kyselyssä tiedustellaan mm. mitä aikakauslehtiä vastaaja lukee. Listalla on 32 lehteä. Eräässä vuoden 2000 tutkimuksessa kymmenen lehden osalta saatiin seuraava jakauma:

Taulukko 5.6.1. Viisi eniten luettua ja viisi vähiten luettua lehteä.

Aikakauslehdet		
	<i>Lkm</i>	<i>%</i>
Pirkka	490	54
Seura	233	26
Apu	224	25
7 päivää -lehti	220	24
HS Kuukausiliite	215	24
...
Vauhdin Maailma	40	4
Gloria	35	4
Kaks'Plus	30	3
Muoti +	19	2
Koti ja Keittiö	16	2
Yht.	3568	393
N:	907	

Yhteensä vastaajia oli 907 henkilöä. He lukivat 3568 lehteä. Suosituin oli ilmaisjakelulehti Pirkka, jonka peittoprosentti oli $490/3568 \cdot 100 \% = 54 \%$. Prosenttiosuuksien summa on yleensä mr-muuttujien tapauksessa yli sadan. Prosenttilukujen summasta voidaan myös päätellä, kuinka monta lehteä lukijat keskimäärin lukevat. Tässä tapauksessa lukua on 3.93.

Tulos.xls-työkirjan Multiresponse-arkille tulostuvat kaikkien yhdistelmien lukumäärät. Neljä yleisintä vähintään kahden lehden yhdistelmää on seuraavassa taulukossa:

Taulukko 5.6.2. Neljä yleisintä yhdistelmää

Nrot	Lehdet	Lkm
6+17	HS Kuukausiliite + HS/Nyt	18
2+32	Apu + Pirkka	15
26+32	Seura + Pirkka	13
2+18+32	Apu + Seura + Pirkka	6

Mielenkiintoinen seikka on että yleisin kahden yhdistelmä ei sisällä luetuinta lehteä eli Pirkkaa.

6. Yhden muuttujan tunnusluvut

Jakauma-esityksessä oleva tieto halutaan usein tiivistää muutamaaan lukuun - muuttujan *tunnuslukuihin*, jotka kuvaavat jakauman tiettyjä ominaisuuksia. Näitä ovat mm.

Sijaintia ilmaisevat	
Keskiluvut	Ilmaisevat jakauman keskimääräisen arvon, esim. keskiarvo.
Muut sijaintiluvut	Kuvaavat jakauman sijaintia, esim. prosenttipisteet.
Muotoa ilmaisevat	
Hajontaluvut	Mittaavat arvojen hajonnan suuruutta, esim. keskihajonta, kvartiilipoikkeama.
Vinousluvut	Kuvaavat jakauman vinoutta

Huom. Seuraavassa luvussa käsitellään ryhmäkohtaisia tunnuslukuja. Tällöin tunnusluvut lasketaan jonkin ryhmittelymuuttujan (esim. sukupuolen) määräämissä luokissa.

6.1. Keskiluvut

Muuttujan mitta-asteikko lähinnä määrää, mitä keskilukua on käytettävä. Sallitut keskiluvut eri asteikkotyypeillä ovat seuraavat:

Mitta-asteikko	Keskiluku
Laatueroasteikko	Moodi
Järjestysasteikko	Mediaani, moodi
Välimatka-asteikko	Keskiarvo, mediaani, moodi
Suhdeasteikko	Keskiarvo, mediaani, moodi, geometrinen keskiarvo

Ensinmainittu keskiluku kunkin mitta-asteikon kohdalla on yleisimmin käytetty. Vaikein valinta on keskiarvon ja mediaanin välillä.

Keskiarvo

Keskiarvon laskentakaava on varmaan kaikille tuttu: lasketaan muuttujan arvot yhteen ja jaetaan yhteenlaskettavien lukumäärällä eli kaavana

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Esimerkiksi alkoholiin opiskelijat käyttävät rahaa viikossa keskimäärin 9,25 euroa. Luennot ja harjoitukset vievät keskimäärin 22,7 tuntia viikossa (opiskelija-aineisto).

Keskiarvon avulla saadaan myös laskettua kokonaissummia. Esimerkiksi kuinka paljon rahaa Tampereen yliopiston opiskelijat käyttävät alkoholiin viikossa ja vuodessa. Viikkoa koskeva arvio saadaan kertomalla keskiarvo opiskelijoiden kokonaismäärällä.

Mediaani

Mediaani on muuttujan arvojen keskimäinen luku. Vaikka mediaanin määrittely on näin yksinkertainen, joudutaan sen kaavaa varten tarkastelemaan kahta eri tapausta sen mukaan, onko havaintojen lukumäärä n pariton vai parillinen.

Merkitään muuttujan x havaittuja arvoja suuruusjärjestykseen asetettuina seuraavasti:

$$x_{(1)}, x_{(2)}, \dots, x_{(n)}$$

1. n on pariton. Tällöin mediaanille on voimassa lauseke

$$Md = x_{((n+1)/2)}$$

2. n on parillinen. Tällöin ei ole yksikäsitteistä keskimäistä arvoa, vaan mediaani on kahden keskimäisen arvon keskiarvo,

$$Md = (x_{(n/2)} + x_{(n/2+1)}) / 2$$

Esimerkki. Iän arvojen ($n=9$) 39, 21, 22, 36, 20, 21, 20, 20, 30 mediaani on

$$x_{((9+1)/2)} = x_{(5)} = 21$$

Iän arvojen ($n=8$) 39, 21, 22, 36, 21, 20, 20, 30 mediaani on $(x_{(4)} + x_{(5)}) / 2 = (21 + 22) / 2 = 21,5$.

Moodi

Moodia käytetään lähinnä laatueroasteikollisten muuttujien keskilukuna. Moodi on se luokka, jonka frekvenssi on suurin. Esimerkiksi siviilisäätymuuttujan moodi on naimaton-luokka (opiskelija-aineisto).

6.2. Muut sijaintiluvut

Esimerkiksi halutaan tietää tuntimäärä c siten, että 20 % opiskelijoista käyttää luentoihin ja harjoituksiin aikaa vähemmän tai yhtä kuin c tuntia. Tällaista lukua c kutsutaan 20 %:n prosenttipisteeksi eli fraktiiliksi.

Yleisesti x -muuttujan p %:n prosenttipisteeksi kutsutaan sellaista lukua $x^{(p)}$, jota pienempiä tai yhtä suuria muuttujan havaittuja arvoja on p %. Eräillä p :n arvoilla prosenttipisteistä käytetään myös seuraavan taulukon mukaisia nimityksiä.

25 %:n fraktiili	alakvartiili
50 %:n fraktiili	mediaani
75 %:n fraktiili	yläkvartiili
10 %:n fraktiili	ensimmäinen desiili
20 %:n fraktiili, jne	toinen desiili

Seuraavassa taulukossa on luentoihin ja harjoituksiin käytetyn tuntimäärän neljä prosenttipistettä.

Muuttuja	Lkm	20 %	40 %	60 %	80 %
Luennot+ harj./vk	4054	16	20	24	30

Siten 40 % opiskelijoista käyttää aikaa luentoihin ja harjoituksiin 20 tuntia tai vähemmän. Koska 80 %:n fraktiili on 30 tuntia, niin 20 % opiskelijoista uurastaa yli 30 tuntia luentojen ja harjoitusten parissa.

6.3. Hajontaluvut

Keskiahjonta

On luonnollista perustaa hajontamitta erotuksiin $x_i - \bar{x}$. Nämä erotukset ovat sekä positiivisia että negatiivisia siten että niiden summa on nolla. Negatiiviset erotukset pitää muuttaa positiivisiksi. Ensimmäinen mieleen tuleva keino on käyttää itseisarvoa, mutta itseisarvolla ei ole hyviä tilastomatemattisia ominaisuuksia. Niinpä kaikki erotukset korotetaan neliöön ja neliöt summataan ja jaetaan havaintojen määrällä. Näin saadaan varianssi s^2

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

(Huom. Jakajana voi olla myös luku $n-1$).

Varianssi ei kuitenkaan ole varsinainen hajontaluku, koska se on toista astetta. Niinpä siitä otetaan neliöjuuri ja saadaan **keskihajontana** s tunnettu hajontaluku,

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

Muuttujien alkoholimarkat ja luennot+harj. keskihajonnat ovat vastaavasti 12,45 euroa ja 7,7 tuntia. Jos näitä verrataan muuttujien keskiarvoihin, huomataan että alkoholin käyttö vaihtelee huomattavasti enemmän kuin luennoilla ja harjoituksissa käyminen.

Muuttujan keskihajonnan suhdetta keskiarvoon kutsutaan **variaatiokertoimeksi**. Se voidaan ilmaista myös prosentuaalisena sadalla kerrottuna. Variaatiokerroin on mittayksiköstä riippumaton hajontaluku (toisin kuin keskihajonta), joten eri mittayksiköillä mitattujen muuttujien variaatiokertoimia voidaan verrata keskenään. Variaatiokertoimen käyttö edellyttää suhteasteikollista mittausta. Seuraavassa esimerkissä muuttujien mittayksiköt ovat marka ja tunti.

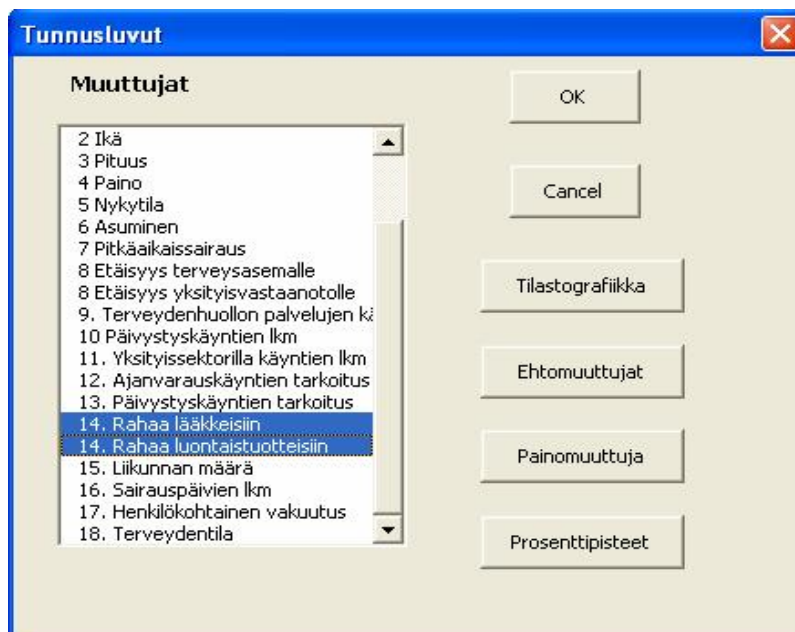
Taulukko 6.3.1. Kahden muuttujan keskiarvot, keskihajonnat ja variaatiokertoimet.

	Keskiarvo	Keskihajonta	Variaatiokerroin
Alkoholieurot	9,25	12,45	135%
Luennot+harj.	22,7	7,7	34%

Taulukon variaatiokertoimista nähdään, että alkoholieurojen vaihtelu on suhteellisesti huomattavasti suurempaa kuin luentoihin ja harjoituksiin käytetyn ajan vaihtelu.

6.4. Tixel-ohjelman käyttö

Tixel- valikosta valitaan vaihtoehto **Tunnusluvut/Tunnusluvut ilman ryhmitteilyä**. Se laskee keskiarvon, mediaanin, keskihajonnan, ala- ja yläkvartiilin sekä minimin ja maksimin. Lisäksi tulostuu myös keskiarvon 95 %:n luottamusväli. Muuttujat määrätään seuraavalla valintaikkunalla.



Seuraavassa taulukossa on lääkkeisiin ja luontaistuotteisiin kuukaudessa käytettyjen rahamäärien (euroa) tunnusluvut.

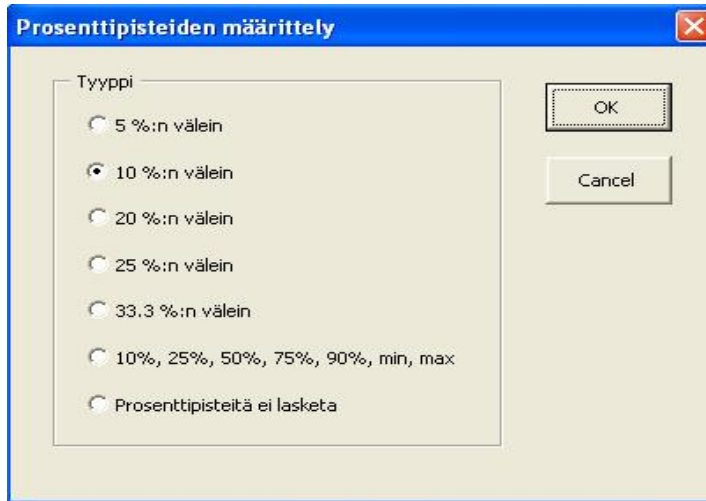
Muuttuja	Lkm	Keskiarvo	Mediaani	Keskihajonta	Minimi	Alakvartiili	Yläkvartiili	Maksimi
Rahaa lääkkeisiin	291	11,92	4,20	23,75	0,00	0,00	13,45	268,91
Rahaa luontaistuotteisiin	290	4,78	0,00	10,15	0,00	0,00	5,04	84,03

Taulukosta nähdään eräitä mielenkiintoisia seikkoja.

- 1) Lääkkeisiin käytetään huomattavasti enemmän rahaa kuin luontaistuotteisiin.
- 2) Molemmilla muuttujilla keskiarvot ovat selvästi suuremmat kuin mediaanit. Tästä seuraa, että jakaumat ovat hyvin vinoja eli pieni määrä henkilöitä käyttää lääkkeitä ja luontaistuotteita paljon (ei välttämättä samat henkilöt).
- 3) Yli puolet vastaajista ei käytä lainkaan rahaa luontaistuotteisiin. Eniten rahaa lääkkeisiin käyttävä neljännes käyttää rahaa lääkkeisiin vähintään 13,45 euroa.

Prosenttipisteet

Jos *Prosenttipisteet*-kohtaan merkitään rasti, antaa Tixel valintaikkunan, jolla määrätään prosenttipisteiden prosenttiarvot. Koska yleensä prosenttipisteet halutaan tasavälisin prosentein, on niitä valmiina viisi sarjaa: 5 %:n, 10 %:n, 20 %:n, 25 %:n ja 33,3 %:n välein sekä vielä kuudes vaihtoehto, jossa on mukana mm. kvartiilit.



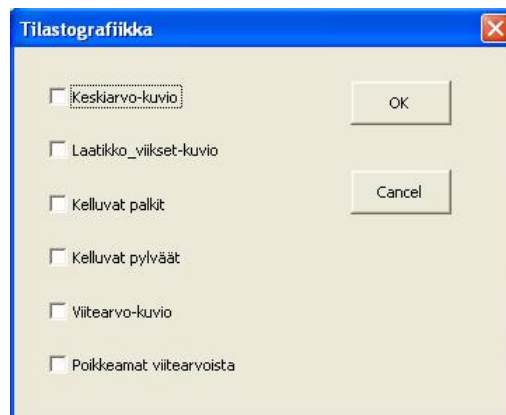
Esimerkiksi kymmenen prosentin välein saadaan lääkkeisiin ja luontaistuotteisiin käytettyjen markkamäärien seuraavat prosenttipisteet eli *desiilit*.

Muuttuja	Lkm	0	10	20	30	40	50	60	70	80	90	100
Rahaa lääkkeisiin	291	0,00	0,00	0,00	0,84	3,36	4,20	6,72	8,40	16,81	33,61	268,91
Rahaa luontaistuotteisiin	290	0,00	0,00	0,00	0,00	0,00	0,00	1,68	3,61	8,40	16,81	84,03

Tämä taulukko tarkoittaa vielä edellä saatua kuvaa. Esimerkiksi eniten lääkkeitä ostava kymmenes käyttää rahaa lääkkeisiin vähintään 33,61 euroa.

Tilastografiikka

Napsauttamalla tunnuslukujen määrittelyikkunassa *Tilastografiikka*-kohtaa, saadaan viereinen ikkuna, jolla voidaan määritellä kuusi eri kuviota. Kaksi viimeistä kuviota edellyttää, että aineistossa on ky-



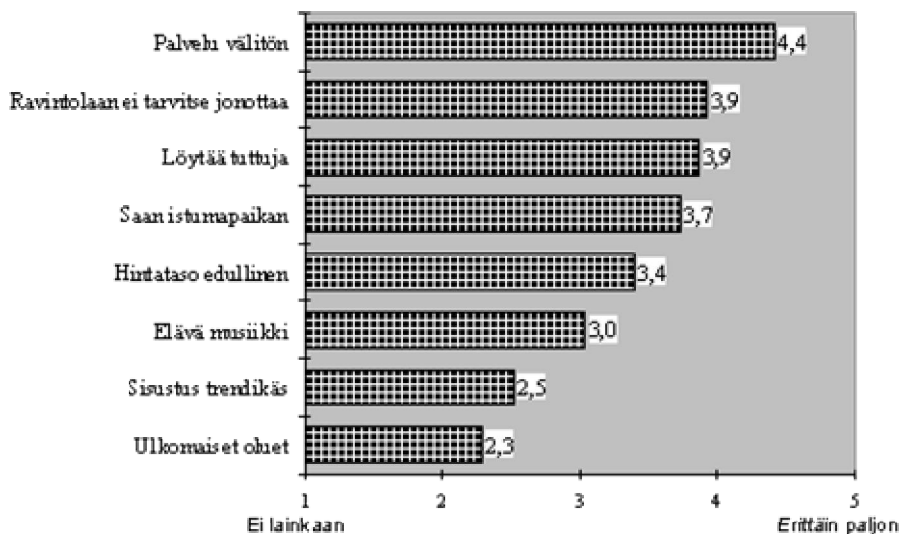
seisten muuttujien kohdalla käytetty viitearvoja.

Keskiarvokuvio. Tixel tekee muuttujien keskiarvoista pylväskuvion. Viereinen kuvio havainnollistaa lääkkeisiin ja luonnontuotteisiin käytettyä rahamäärää. Kuvio tulee omalle grafiikka-arkille.



Kuvion tyyppi voidaan muuttaa myös palkkikuvioksi. Seuraava esimerkki kuvaa, kuinka paljon eri tekijät vaikuttavat seurusteluravintolan valintaan. Kuvioon on otettu kahdeksan tekijän keskiarvot, jotka on lajiteltu suuruuden mukaiseen järjestykseen.

Kuvio 6.4.1. Ravintolan valintaan vaikuttavat tekijät



Laatikko&viikset-kuvio. Tämä kuvallinen esitystapa välittää enemmän tietoa kuin keskiarvokuvio. Siinä havainnollistetaan muuttujan kaikkia viittä järjestystunnuslukua: minimi, alakvartiili, mediaani, yläkvartiili ja maksimi. Laatikon ala-

ja yläreuna ovat ala- ja yläkvartiileja, keskellä oleva piste on mediaani sekä ala- ja yläviivojen päät ovat minimi ja maksimi. Laatikon sisään jää siten puolet havainnoista. Mitä korkeampi laatikko on, sitä enemmän muuttujassa on vaihtelua.

Viereisessä kuviossa on kahden muuttujan *odotusaika1* ja *odotusaika2* laatikko&viikset-kuvio.

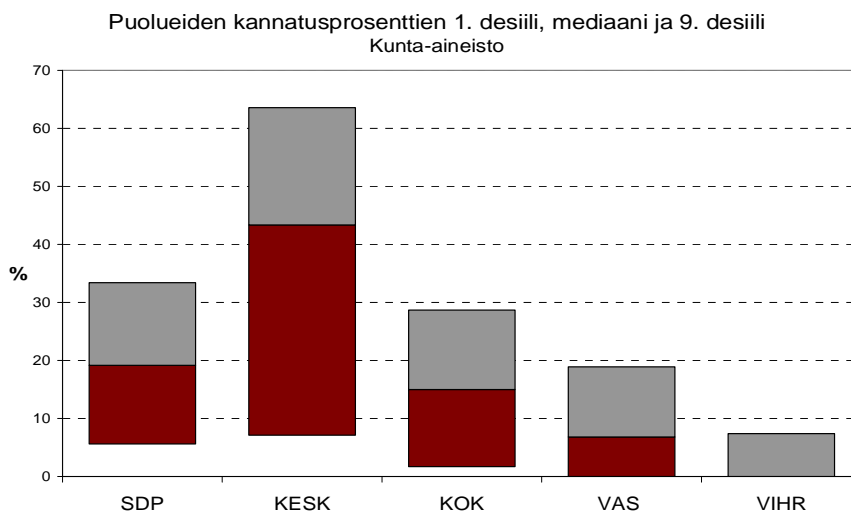
Muuttujat mittaavat teollisuusprosessin eri vaiheisiin liittyviä odotusaikoja.

Huomataan että toisen

odotusajan mediaani ja vaihtelu ovat suurempia kuin ensimmäisellä odotusajalla.

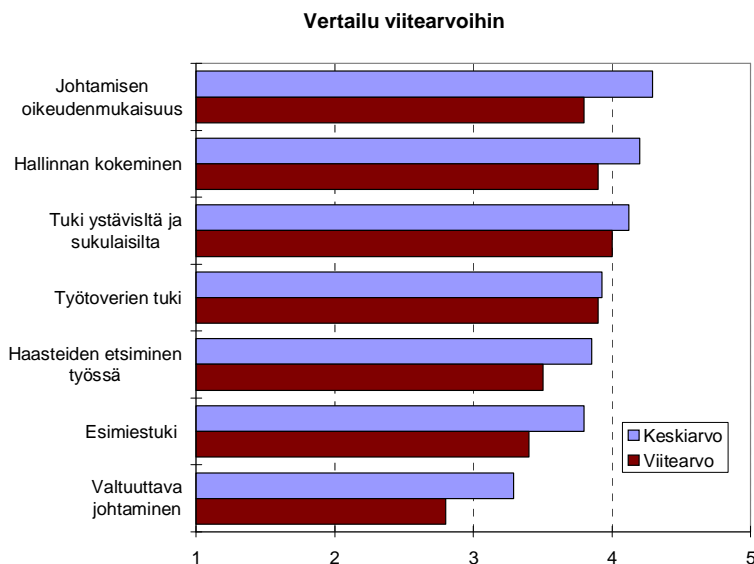
Kelluvat palkit ja pylvääät. Näillä kuvatyypeillä havainnollistetaan muuttujien ensimmäistä desiliä, mediaania ja yhdeksättä desiliä. Palkit tai pylvääät sisältävät 80 % havainnoista. Seuraavassa esimerkissä on viiden puolueen kannatusprosenttien tunnusluvut laskettuna kunta-aineistosta, jossa kunnat ovat tilastoyksikköinä.

Kuvio 6.4.2. Kelluvat pylvääät

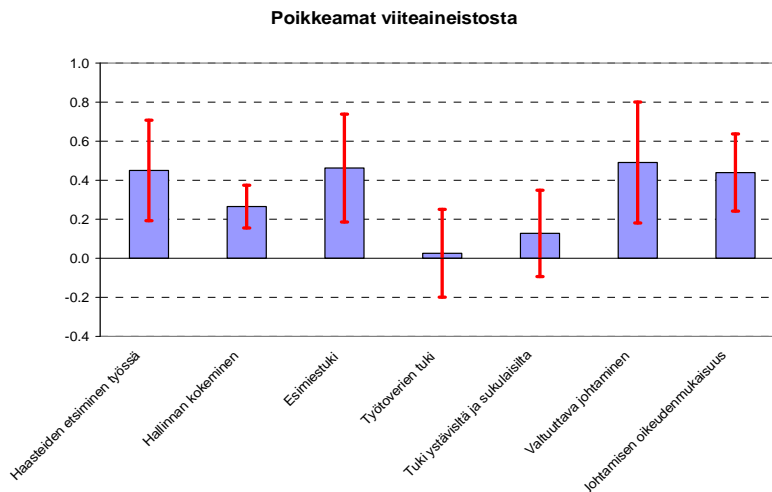


Kuviosta huomataan esimerkiksi, että keskustapuolueen kannatuksen vaihtelu on suurinta ja että yli puolet kunnista on sellaisia, joissa vihreillä ei ole kannatusta.

Viitearvo-kuvio. Kuvio esittää muuttujien keskiarvot ja viitearvot samassa kuviossa. Tämän kuvion käyttö edellyttää luonnollisesti, että aineistossa on kyseisten muuttujien kohdalla viitearvot. Ne voivat olla esimerkiksi laajemman vertailuaineiston keskiarvoja tai asetettuja tavoitearvoja. Seuraavassa esimerkissä. Seuraavassa esimerkissä on pieni osa erään yrityksen sisäisen ilmapiirikartoituksen tuloksista.



Poikkeamat viitearvoista. Tässä kuviossa kuvataan pylväillä keskiarvojen poikkeamia viitearvoista ja kuvioon lisätään vielä janoina keskiarvojen 95 %:n luottamusvälit. Jos janaan ei sisälly origoa (0-arvoa), niin erotuksen poikkeamaa nolasta eli keskiarvon poikkeamaa viitearvosta voidaan pitää ”todellisena” eikä sattumasta johtavana. (Luottamusvälit eivät kuulu tämän kirjan varsinaiseen aihepiiriin.)

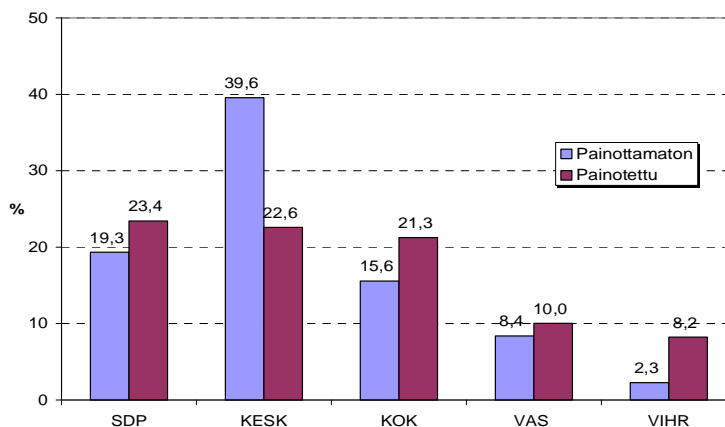


Kuviosta nähdään että viiden muuttujan poikkeamat ovat tilastollisesti merkitseviä eli janat eivät sisällä nolaa.

Painomuuttuja

Halutaan laskea suurimpien puolueiden kannatusprosenttien keskiarvot kunta-aineistosta. Kun keskiarvot lasketaan tavalliseen (painottamattomaan) tapaan, on kaikilla kunnilla – niin suurilla kuin pienilläkin – sama paino keskiarvoon. Koska pieniä kuntia on paljon, niiden vaikutus korostuu keskiarvossa. Kunnan asukasmäärä voidaan ottaa huomioon keskiarvon laskennassa painomuuttujan avulla. Seuraavassa taulukossa on puolueiden kannatusprosenttien keskiarvot laskettu painottamattomina ja täysikäisten lukumäärällä painotettuna.

Kuvio 6.4.2. Puolueiden kannatusprosentit vuoden 2000 kunnallisvaaleissa.



Huomataan että keskustapuolueen kannatusprosentin painottamaton keskiarvo on lähes 40 %, kun sen sijaan täysikäisten lukumäärällä painotettu keskiarvo on 22.6 %, mikä on varsin lähellä keskustapuolueen todellista kannatusprosenttia. Jos painomuuttujana olisi käytetty äänestäneiden lukumäärää kunnassa, olisi saatu täsmälleen kunnallisvaalien antama kannatusprosentti.

6.5. Jakauman keskiarvo ja keskihajonta

Laskentapohja laskee yksiulotteisen jakauman keskiarvon, keskihajonnan ja varianssin, kun tiedetään luokkien luokkakeskukset ja absoluuttiset tai prosentuaaliset frekvenssit. Laskentapohja saadaan käyttöön, kun heti Tixel-valikosta valitaan *Laskentapohjat/Empiirisen jakauman tunnusluvut*.

Luokkakeskus on luokan ala- ja ylärajan keskiarvo. Esimerkiksi jos luokka on määritelty välinä 1-100, on luokkakeskus 50,5.

Laskentapohja on seuraavanlainen Excel-taulukko.

Muuttuja:			
Luokkakeskus	Frekvenssi		
		Keskiarvo	#DIV/0!
		Keskihajonta	#DIV/0!
		Varianssi	#DIV/0!
		Frekvenssien summa	0
		Luokkien lkm	0

Esimerkki. Olkoon muuttujan yksiulotteinen jakauma seuraava:

Luokka	Frekvenssi
1-100	10
101-200	25
201-300	35
301-400	5

Tämän jakauman tunnusluvut käyvät ilmi seuraavasta taulukosta.

Luokkakeskus	Frekvenssi		
50,5	10	Keskiarvo	197,167
150,5	25	Keskihajonta	80,5536
250,5	35	Varianssi	6488,89
350,5	5	Frekvenssien summa	75
		Luokkien lkm	4

6.6 Standardointi

Eräs yleisimmistä muuttujan lineaarisista eli viivallisista muunnoksista on muuttujan standardointi. Siinä muuttujan mittayksiköksi otetaan tavallaan keskihajonta ja nollakohtaksi keskiarvo. Useamman muuttujan standardoituja arvoja voidaan siten verrata keskenään. Muuttujan x standardoitua esitystä merkitään symbolilla u ja sen lauseke on

$$u = \frac{x - \bar{x}}{s}$$

jossa \bar{x} ja s ovat vastaavasti muuttujan x keskiarvo ja keskihajonta. Seuraavassa taulukossa on lapsen syntymäpainon ja -pituuden arvot ja vastaavat standardoidut arvot. Koko aineiston koko on 120 lasta.

Taulukko 6.6.1. Osa muuttujien arvoista ja vastaavista standardoiduista arvoista.

Lapsen paino	Lapsen pituus	Stand. paino	Stand. pituus
3410	51	-0,296	0,186
3120	51	-0,892	0,186
3060	48	-1,015	-1,299
3500	52	-0,111	0,680
3200	49	-0,727	-0,804
4320	54	1,576	1,670
3600	52	0,095	0,680
4550	54	2,049	1,670

Standardoidut arvot vaihtelevat nollan molemmin puolin. Jos muuttuja noudattaa *normaalijakaumaa*, on itseisarvoltaan kahta suuremman standardoidun arvon esiintymistodennäköisyys noin viisi prosenttia. Eli itseisarvoltaan kahta suurempaa arvoa voidaan pitää suhteellisen harvinaisena. Taulukon standardoiduista arvoista viimeisellä rivillä oleva arvo 2,049 ylittää tämän rajan. Sitä vastaava lapsen paino on 4550.

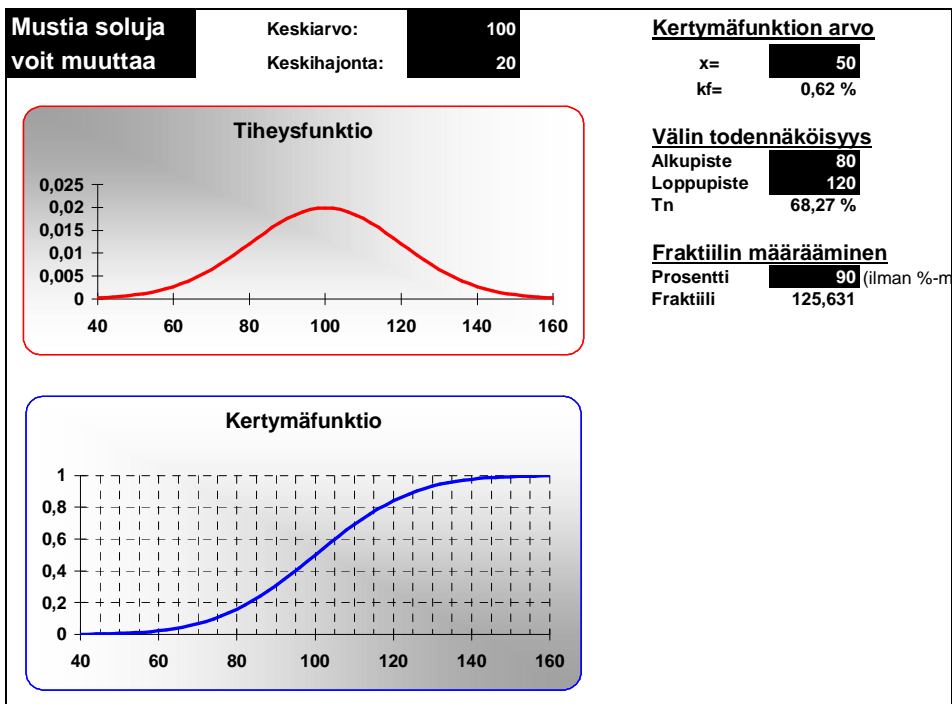
6.7. Normaalijakauma

Tässä kirjassa ei käsitellä yksityiskohtaisesti normaalijakaumaa, joka on teoreettinen jakauma vastakohtana edellä käsitellyille empiirisille jakaumille. Hyvin ylimalkaisesti ilmaistuna muuttujan jakauma on likimäärin normaalin, jos se on symmetrinen ja muuttujan arvoja on keskiarvon läheisyydessä enemmän kuin etäällä keskiarvosta. Esimerkiksi lapsen syntymäpituus ja –paino ovat likimäärin normaalisti jakautuneita, kun sen sijaan ”alkoholiin käytetyt eurot”-muuttuja ei noudata normaalijakaumaa, sillä jakauma on oikealle vino.

Tixelissä on laskentapohja, joka piirtää normaalijakauman tiheys- ja kertymäfunktioiden kuvaajat annetulla keskiarvolla ja keskihajonnalla sekä laskee eräitä todennäköisyyksiä ja prosenttipisteitä. Tiheysfunktion vastine empiiristen jakaumien puolella on prosentuaalinen frekvenssimurtoviiva ja kertymäfunktion vastine on summakäyrä.

Valitse Tixel-valikosta *Erikoiskuviot/Normaalijakauma*, niin saat seuraavan laskentapohjan. Kirjoittamalla keskiarvon ja keskihajonnan soluihin arvoja, saat vastaavan normaalijakauman kuvaajat.

Kuvio 6.7.1. Normaalijakauma



6.8. Tyytyväisyys-tärkeys muuttujaparit ja kuilu-analyysi

Asiakaspalautetutkimuksissa on tärkeää tietä paitsi vastaajien tyytyväisyyden taso eri seikkoihin niin myös se kuinka tärkeänä eri tekijöitä pidetään. Tarkastellaan seuraavaa yksinkertaista esimerkkiä, jossa ruokaravintolan asiakkailta on kysytty kuuden tekijän osalta tyytyväisyys ja tärkeys asteikolla 1-5. Tekijät ja niiden keskiarvot ovat seuraavassa taulukossa.

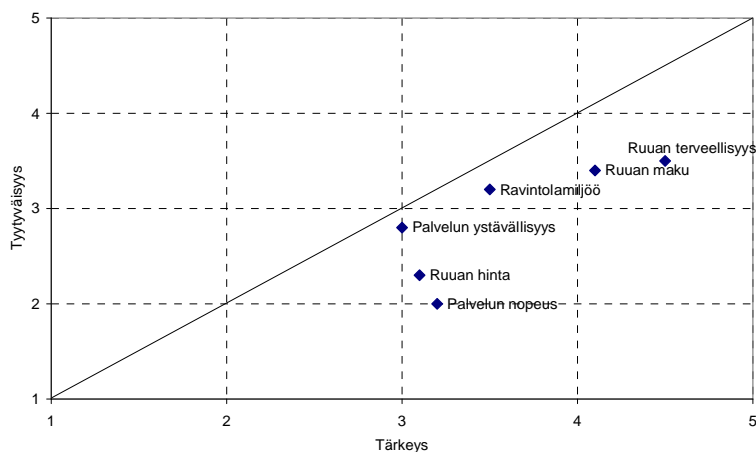
Tekijä	Tyytyväisyys	Tärkeys
Ruuan maku	3,4	4,1
Ruuan hinta	2,3	3,1
Ruuan terveellisyys	3,5	4,5
Palvelun nopeus	2,0	3,2
Palvelun ystävällisyys	2,8	3,0
Ravintolamiljö	3,2	3,5

Taulukko määrittelee kuusi muuttujaparia eli yhteensä kaksitoista muuttujaa. Muuttujien järjestykselle havaintomatriisissa on kaksi vaihtoehtoa:

- Tyytyväisyys-muuttujat ovat peräkkäin ja sen jälkeen tärkeys-muuttujat.
- Muuttujaparit ovat havaintomatriisissa peräkkäin.

Tulosten graafisessa havainnollistamisessa voidaan käyttää joko tähtitaivaskuviota tai kuilu-analyysia. Molemmat saadaan Tixel-valikosta valitsemalla vaihtoehto *Erikoiskuviot/Tärkeys-tyytyväisyys kuvio*.

Kuvio 6.8.1. Tähtitaivaskuvio



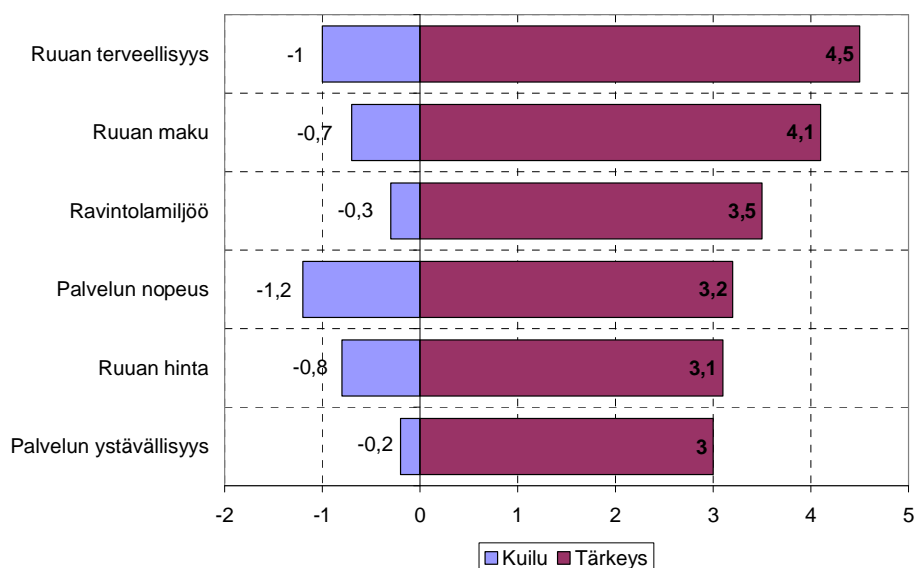
Kuvion on piirretty myös lävistäjä, jonka alapuolella olevat pisteet kertovat *alipanostuksesta* ja yläpuolella olevat pisteet *ylipanostuksesta*. Mitä etäämmällä piste on suorasta, sitä suurempi on ali- tai ylipanostus. Tässä tapauksessa kaikkien tekijöiden kohdalla on alipanostusta ja suurin se on palvelun nopeuden kohdalla.

Yli- ja alipanostuksen mittana voidaan käyttää tyytyväisyyden ja tärkeyden erotusta, jota kutsutaan *kuiluksi* eli

$$\text{kuilu} = \text{tyytyväisyys} - \text{tärkeys}.$$

Kuilukuvioon sijoitetaan yleensä kuilumuuttujan lisäksi tärkeys ja palkit lajitellaan tärkeyden mukaiseen järjestykseen. Tällöin ylimmäksi tulevat tärkeimmät tekijät ja vasemmalla ovat vastaavat kuilut.

Kuvio 6.8.2. Kuilukuvio



7. Ryhmäkohtaiset tunnusluvut

Edellä olleet tunnuslukujen kaavat ovat perustuneet siihen, että tunnusluvut lasketaan koko aineistosta. Samat kaavat pätevät luonnollisesti myös silloin, kun tunnusluvut lasketaan aineiston osajoukoista, esim. alkoholiin käytetyn rahamäärän tunnusluvut (esim. keskiarvo, ala- ja yläkvartiili) erikseen naisten ja miesten ryhmässä. Tällaisia tunnuslukuja kutsutaan *ryhmäkohtaisiksi* tai mahdollisiksi tunnusluvuiksi

Tällöin on tarkoitus yleensä verrata **riippuvan muuttujan** (esim. alkoholiin käytetty rahamäärä) tunnuslukuja **ryhmittelymuuttujan** (esim. sukupuoli) määräämissä luokissa. Tässä esimerkissä on yksi ryhmittelymuuttuja. Jos halutaan selvittää alkoholiin käytetyn rahamäärän riippuvuus yhtäaikaisesti sukupuolesta ja ikäluokasta, on ryhmittelymuuttujia kaksi (sukupuoli ja ikäluokka). Tixelin menetelmävalikossa on eri vaihtoehdot sen mukaan, onko ryhmittelymuuttujia yksi vai kaksi.

Nämä molemmat menetelmät edellyttävät, että riippuva muuttuja on kvantitatiivinen. Myös dikotominen (eli kaksiluokkainen) muuttuja käy, jolloin keskiarvot voidaan tulkita prosenttiluvuiksi (0/1-muuttujat ja %-muotoilu). Ryhmittelymuuttujalle ei aseteta mitään mittauksen tasoon liittyviä vaatimuksia. Jos ryhmittelymuuttuja on kvantitatiivinen, esim. ikä vuosina, on se kuitenkin luokiteltava esim. ikäryhmiksi.

Jos halutaan käyttää kolmea ryhmittelymuuttujaa, on kolmas ryhmittelymuuttuja määriteltävä ehtomuuttujaksi (kts, alaluku 2.4). Koska ehtomuuttujien maksimimäärä on kolme, voidaan aineisto ryhmitellä ryhmittely- ja ehtomuuttujien avulla maksimissaan viiden muuttujan suhteen.

7.1. Ryhmäkohtaiset tunnusluvut, yksi ryhmittelymuuttuja

Perustulostuksessa Tixel laskee riippuvan muuttujan (eli selitettävän muuttujan) keskiarvon, keskihajonnan ja lukumäärät ryhmittelymuuttujan määräämissä tilastoyksiköiden ryhmissä. Muuttujat määrätään seuraavanlaisella valintaikkunalla, jolla nyt lasketaan alkoholimarkkojen tunnusluvut miehille ja naisille (opiskelija-aineisto).

Kuva 7.1.1. Ryhmäkohtaisten tunnuslukujen määrittelyikkuna

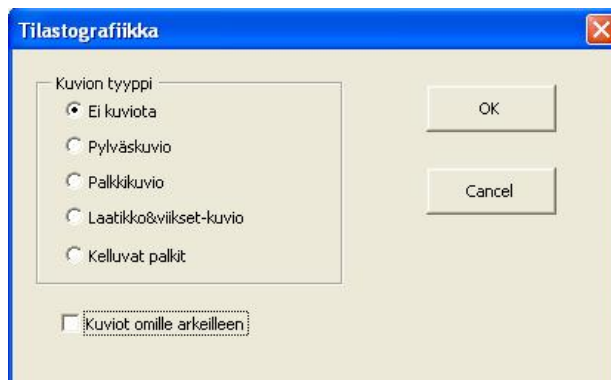
Tuloksena saadaan viereinen taulukko.

Miesopiskelijat käyttävät huomattavasti enemmän rahaa alkoholiin kuin naisopiskelijat (keskiarvot 15 € ja 6 €). Mediaanit ovat molemmissa ryhmissä selvästi pienempiä kuin keskiarvot, mikä osoittaa, että alkoholimarkkojen jakauma on miesten ja naisten ryhmissä varsin vino.

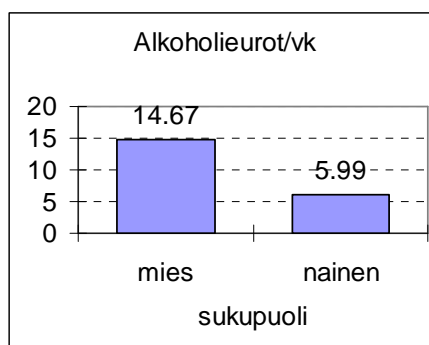
Alkoholieurot/vk			
Ryhmittelymuuttuja: sukupuoli			
	<i>mies</i>	<i>nainen</i>	<i>Yht</i>
Keskiarvo	15	6	9
Keskihajonta	15	8	11
Minimi	0	0	0
Alakvartiili	3	1	2
Mediaani	8	3	5
Yläkvartiili	17	8	10
Maksimi	84	84	84
Lkm	398	958	1356

F = 194.49 vap. asteet: 1 ja 1354
p = 0 Tilastollisesti erittäin merkitsevä

Klikkaamalla *Tilastografiikka*-kohtaa saadaan viereinen määrittelyarkki. Graafisia esitystapoja on neljä, jotka esitellään seuraavassa.

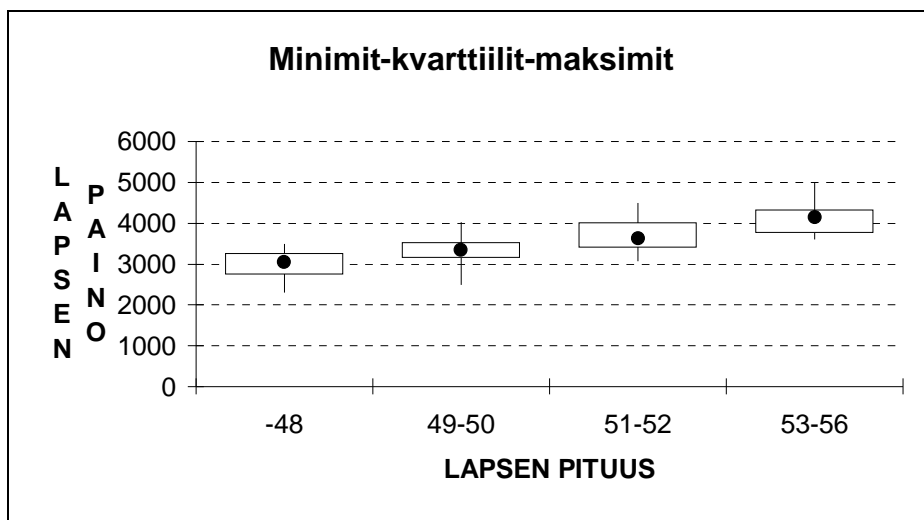


Pylväs- ja palkkikuvio. Pylväiden korkeudet ovat yhtä kuin riippuvan muuttujan keskiarvot luokissa. Palkkikuviossa palkit ovat vaakasuunnassa.



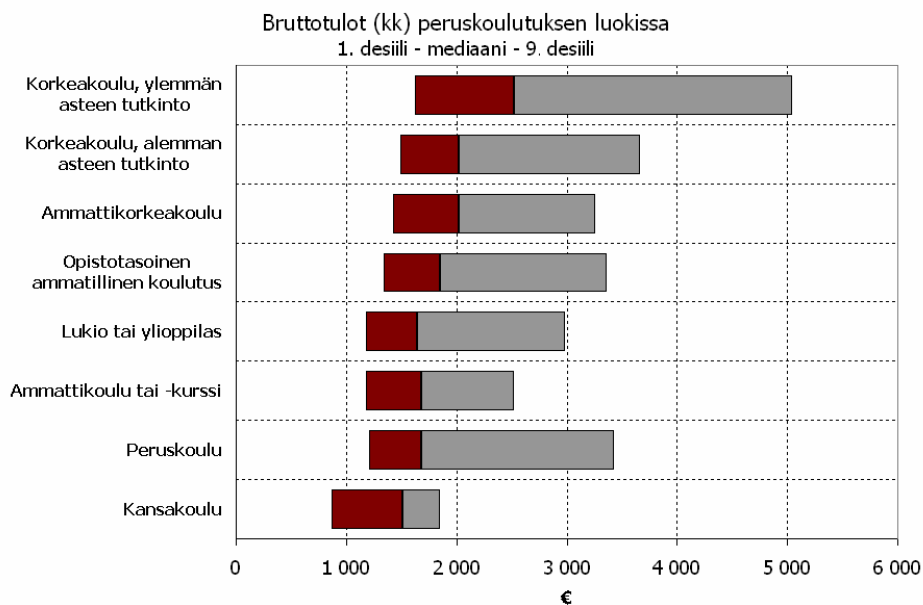
Laatikko&viikset-kuvio. Siinä havainnollistetaan riippuvan muuttujan kaikkia viittä järjestystunnuslukua: minimi, alakvartiili, mediaani, yläkvartiili ja maksimi. Laatikon ala- ja yläreuna ovat ala- ja yläkvartiileja, keskellä oleva piste on mediaani sekä ala- ja yläviivojen päät ovat minimi ja maksimi. Jokaiselle luokalle tulee oma laatikko&viikset-esitys.

Esimerkki laatikko&viikset-kuviosta on lapsen syntymäpainon riippuvuus syntymäpituudesta.



Kuviosta nähdään että lapsen paino nousee pituuden mukana.

Kelluvat palkit. Jokaiselle luokalle piirretään kaksiosainen palkki, jonka ensimmäisen osan määräävät riippuvan muuttujan ensimmäinen desiili ja mediaani sekä toisen osan mediaani ja yhdeksäs desiili. Seuraavassa kuvassa on havainnollistettu bruttopalkan riippuvuutta peruskoulutuksesta. Havaintojen kokonaismäärä on 638 ja aineisto on vuodelta 2000.



Kuviosta nähdään mm. että mediaanitulo nousee koulutuksen myötä. Suurin kasvu on ylimmän koululuokan kohdalla. Tosin peruskoulupohjaltakin eniten ansaitseva kymmenes ansaitsee vähintään 3500 euroa/kk. Suurimmat palkkaerot eli palkin kokonaispituus on ylimmällä koulutustasolla ja peruskoulutasolla.

Prosenttipisteet

Prosenttipisteet määrätään aluvussa 6.4 olleella valintaikkunalla.

Esimerkki. Alkoholimarkkojen desiilit eli 10 %:n, ... ,90 %:n prosenttipisteet halutaan laskea erikseen naisille ja miehille (opiskelija-aineisto). Ryhmittelymuuttuja on tällöin sukupuoli ja riippuva muuttuja alkoholieurot. Tuloksena saadaan seuraava taulukko.

Taulukko 7.1.1. Muuttujan prosenttipisteet sukupuolen mukaan

Alkoholieurot/vk

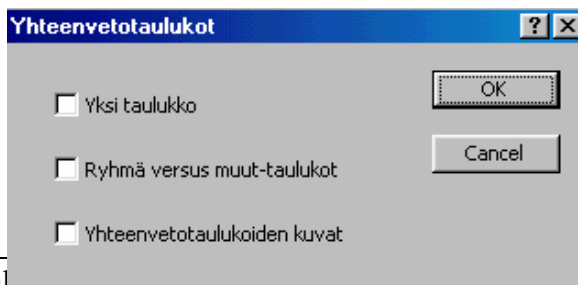
Ryhmittelymuuttuja: sukupuoli

	Lkm	K.a.	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
mies	398	14.67	0.00	0.29	2.86	5.04	8.40	8.40	13.45	16.81	25.21	33.61	84.03
nainen	958	5.99	0.00	0.00	0.00	1.68	3.33	3.36	5.04	8.40	8.40	16.81	84.03
Yht	1356	8.54	0.00	0.00	0.84	1.68	3.36	5.04	8.40	8.40	13.45	16.81	84.03

Huomataan että miesten prosenttipisteet ovat kymmenen prosentin pistettä lukuun ottamatta selvästi suurempia kuin naisten vastaavat arvot. Esimerkiksi miesten 70 %:n prosenttipiste on 16,81 € kun naisilla vastaava tunnusluku on 8,40 €

Yhteenvetotaulukot

Jos useamman riippuvan muuttujan keskiarvoja vertaillaan ryhmittelymuuttujan luokissa, kannattaa keskiarvot sijoittaa yhteen taulukkoon eli yhteenvetotaulukkoon. Tätä varten napsautetaan Yhteenvetotaulukot-painiketta valintaikkunassa 7.1.1, jolloin



Johdatus tilastolliseen data-analyysiin

saadaan viereinen määrittelyikkuna. Yhteenvetotaulukoita on kahta tyyppiä: 1) Kaikkien muuttujien ryhmäkohtaiset keskiarvot ja yhteensä-keskiarvot kootaan yhteen taulukkoon. 2) Ryhmäkohtaisille keskiarvoille lasketaan vertailukeskiarvoina muiden ryhmien keskiarvot (ryhmä versus muut-asetelma) ja parittaiset keskiarvot sijoitetaan ryhmäkohtaisiin taulukoihin. Taulukot tulostuvat Tulokset.xls-työkirjan arkeille, joiden nimet ovat vastaavasti *One_table* ja *Versus*.

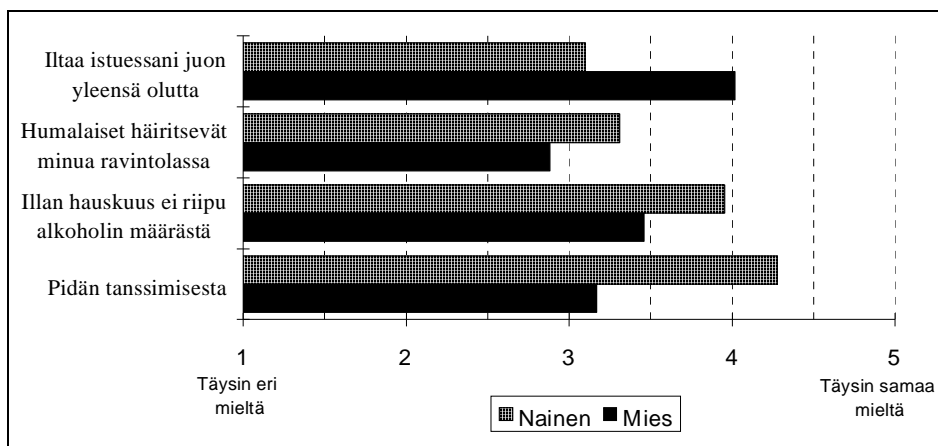
Tarkastellaan ensin esimerkkiä, jossa keskiarvot sijoitetaan yhteen taulukkoon. Neljän ravintolakäyttäytymistä kuvaavan muuttujan keskiarvot laskettiin naisille ja miehille. Asteikko on 1=täysin eri mieltä, ... 5= täysin samaa mieltä (ravintolaineisto).

Taulukko 7.1.2 Yhteenvetotaulukko (One_table)

	Nainen	Mies	Yht.
Iltaa istuessani juon yleensä olutta	3,10	4,02	3,49
Humalaiset häiritsevät minua ravintolassa	3,31	2,88	3,13
Illan hauskuus ei riipu alkoholin määrästä	3,96	3,46	3,74
Pidän tanssimisesta	4,28	3,17	3,80

Taulukon graafinen esitys on seuraava.

Kuva 7.1.2. Muuttujien keskiarvot sukupuolen mukaan



Sukupuolten väliset erot ovat odotetunlaiset (?) : miehet juovat olutta, naiset häiriintyvät humalaisista, naisilla illan hauskuus ei riipu juodun alkoholin määrästä ja naiset pitävät tanssimisesta.

Ryhmä vs muut-tyyppistä yhteenvetotaulukkoa varten lasketaan edellisen esimerkin riippuvat muuttujat kolmessa ikäryhmässä: alle 20 v., 20-30 v. ja yli 30 v. Tixel tulostaa seuraavat taulukot.

Taulukko 7.1.3. Yhteenvetotaulukko (Versus-tyyppi)

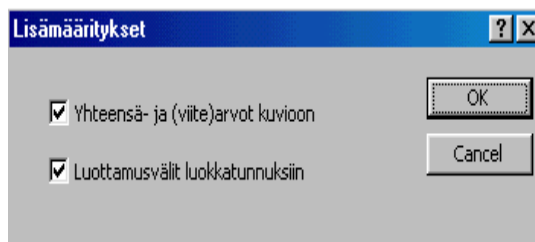
Ryhmittelymuuttuja: 2 Ikäryhmä											
Alle 20 v.		Muut	p-arvo	20-30 v.		Muut	p-arvo	Yli 30 v.		Muut	p-arvo
Iltaa istuess	3,24	3,52	0,1472	Iltaa istuess	3,61	3,33	0,0249	Iltaa istuess	3,36	3,55	0,1580
Humalaiset	2,89	3,15	0,0901	Humalaiset	3,05	3,23	0,0715	Humalaiset I	3,35	3,03	0,0021
Illan hausku	3,71	3,74	0,8452	Illan hausku	3,69	3,81	0,2662	Illan hausku	3,84	3,70	0,1872
Pidän tanss	4,31	3,74	0,0006	Pidän tanss	3,66	3,99	0,0014	Pidän tanssi	3,88	3,76	0,2769

Esimerkiksi ensimmäisessä taulukossa (alle 20 v.) ryhmä Muut tarkoittaa yli 20-vuotiaita. Taulukoissa on myös kahdesta riippumattomasta otoksesta laskettujen keskiarvojen t-testin p-arvot.

Lisämäärittelykset

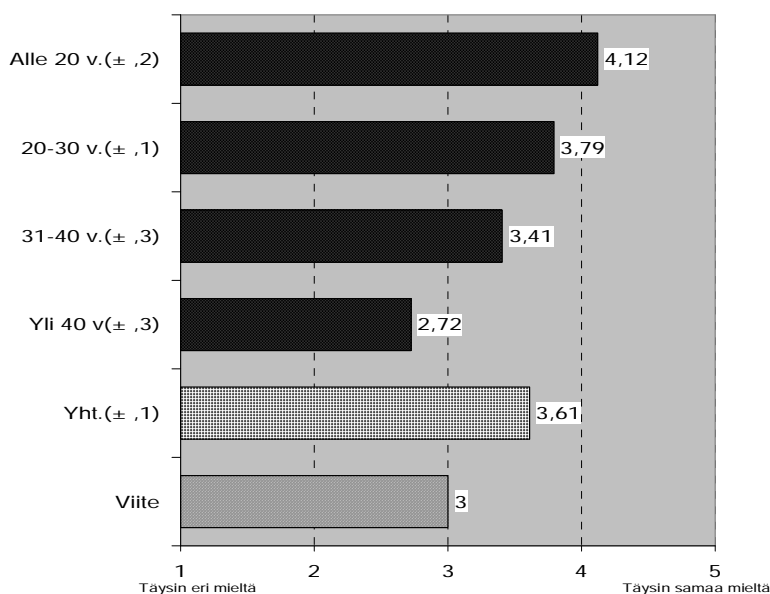
Menetelmän valintaikkunassa on Lisämäärittelykset-painike, jota napsauttamalla saadaan viereinen määrittelyarkki.

Jos ensimmäiseen vaihtoehtoon laitetaan rasti, lisää Tixel yhteensä- keskiarvon ja viitearvon kuvioon. Jos viitearvo ei ole, lisätään kuvioon vain yhteensä-keskiarvo. Jos toiseen vaihtoehtoon laitetaan rasti, Tixel lisää 95 %:n luottamusvälit taulukon luokkatunnuksiin ja myös kuvaan.



Seuraavassa kuvassa on hyödynnetty lisämääreiden molempia mahdollisuuksia.

Ostan juomani yleensä baaritiskiltä voidakseni liikusella vapaammin

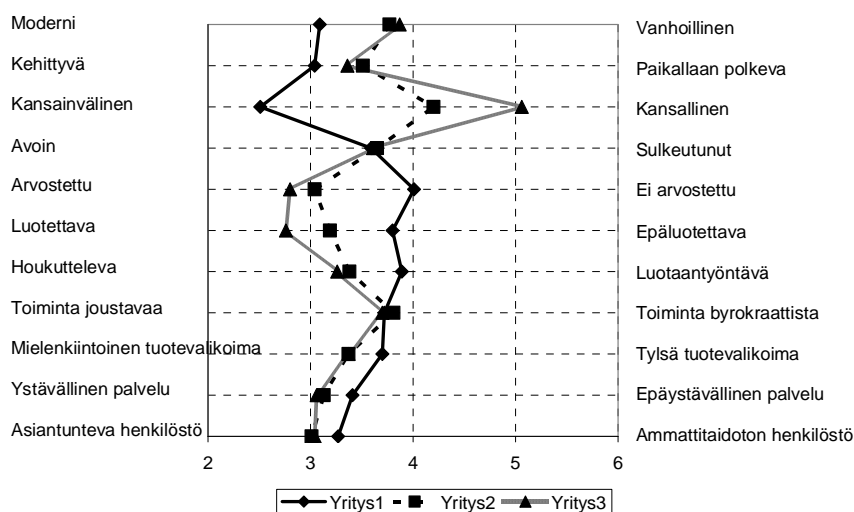


Profiilikuvio

Profiilikuvio esittää usean muuttujan ryhmäkohtaiset keskiarvot pystymurtovivoina. Tarkastellaan seuraavaa esimerkkiä, jossa muuttujat mitattu nk. Osgoodin asteikolla. Siinä vastaajalle annetaan muuttujien ääripäät sanallisesti ilmaistuina. Taulukossa ovat luokkien (yritysten) keskiarvot yhdellätoista muuttujalla.

	Yritys1	Yritys2	Yritys3	
Moderni	3.09	3.77	3.87	Vanhoillinen
Kehittyvä	3.04	3.51	3.36	Paikallaan polkeva
Kansainvälinen	2.51	4.2	5.06	Kansallinen
Avoin	3.59	3.65	3.61	Sulkeutunut
Arvostettu	4.01	3.04	2.8	Ei arvostettu
Luotettava	3.8	3.19	2.76	Epäluotettava
Houkutteleva	3.89	3.38	3.26	Luotaantyöntävä
Toiminta joustavaa	3.72	3.81	3.7	Toiminta byrokraattista
Mielenkiintoinen tuotevalikoima	3.7	3.37	3.37	Tylsä tuotevalikoima
Ystävällinen palvelu	3.41	3.13	3.06	Epäystävällinen palvelu
Asiantunteva henkilöstö	3.27	3.01	3.04	Ammattitaidoton henkilöstö

Valitse Tixel-valikosta *Erikoiskuvio/Profiilikuvio*, jolloin aukeaa syöttöpohja, johon muuttujien nimet ja keskiarvot kopioidaan. Lopuksi saadaan seuraavanlainen kuvio.



Kuviossa kullakin yrityksellä on oma profiilinsa. Oikealla olevia tekstejä ei välttämättä tarvita, vaan riittää että muuttujien nimet ovat vasemmalla. Kovin monta profiilia ei yhteen kuvaan mahdu, jos kuvio halutaan pitää selkeänä.

7.2. Ryhmäkohtaiset tunnusluvut, kaksi ryhmittelymuuttujaa

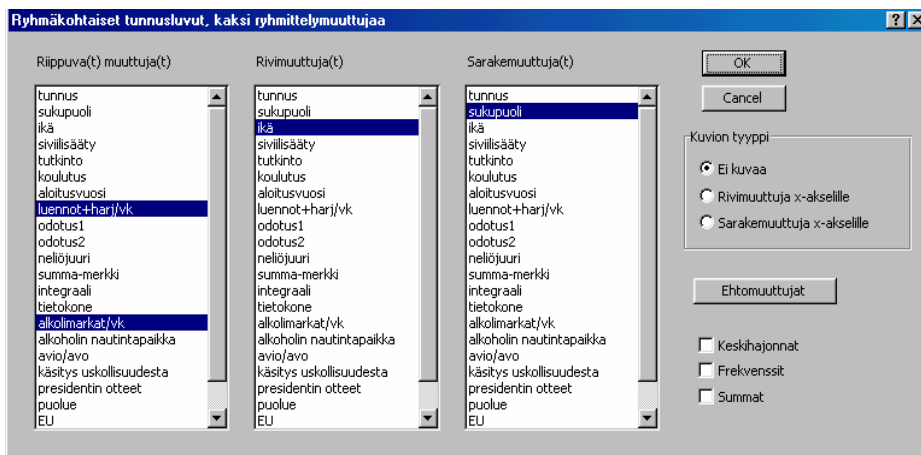
Tixel-ohjelma laskee riippuvan muuttujan keskiarvot ja keskihajonnat kahden ryhmittelymuuttujan määräämässä taulukossa. Ryhmittelymuuttujien luokkien määräämisessä voidaan käyttää joko havaintomatriisissa olevia luokkamäärittäjiä tai erillistä luokitustiedostoa.

Tutkitaan, miten opiskelijoiden alkoholin käyttö riippuu kyselyn ajankohdasta (vuodesta -82 vuoteen -94) ja sukupuolesta (opiskelija-aineisto). Alkoholin käytön kohdalla ollaan kiinnostuneita 1) absolutistien osuudesta ja 2) alkoholimarkkojen keskiarvosta.

Alkoholiin käytetyistä markoista muodostetaan uusi kaksiluokkainen muuttuja siten, että 0=ei käytä alkoholia, 1=käyttää alkoholia. Tämän muuttujan keskiarvo

on yhtä kuin absolutistien prosentuaalinen osuus, kun keskiarvojen muotoilussa käytetään prosenttiluku-formaattia.

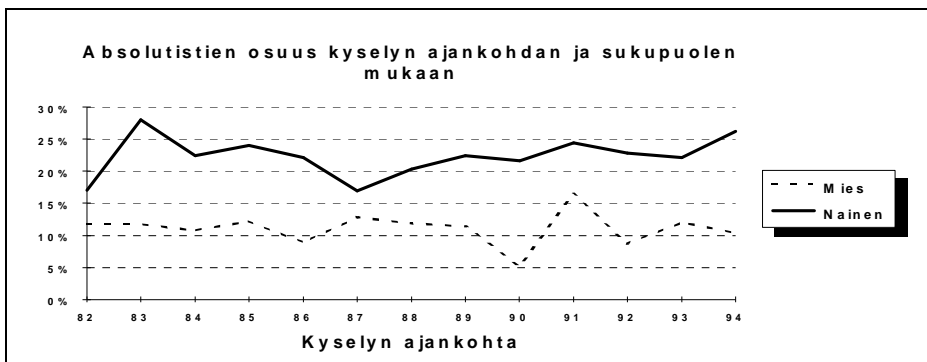
Tixel pyytää määrittelemään muuttujat seuraavalla valintaikkunalla.



Kahdesta ryhmittelymuuttujasta toinen asetetaan rivimuuttujaksi ja toinen sarakuuttujaksi. Tämä vaikuttaa vain keskiarvotaulukon muotoon - ei itse keskiarvoihin.

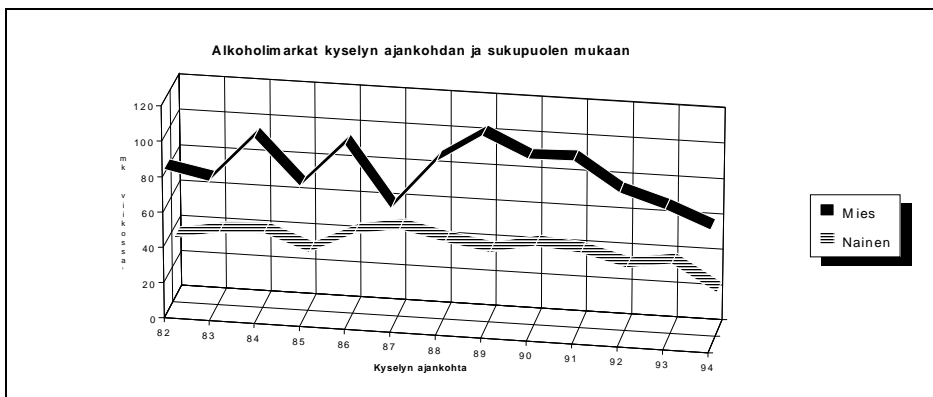
Tixelin loppuvalikossa on kaksi kuva-vaihtoehtoa sen mukaan, kumpi ryhmittelymuuttuja asetetaan x-akselille.

Kun kyselyn ajankohta asetetaan x-akselille, saadaan seuraava kuvio.



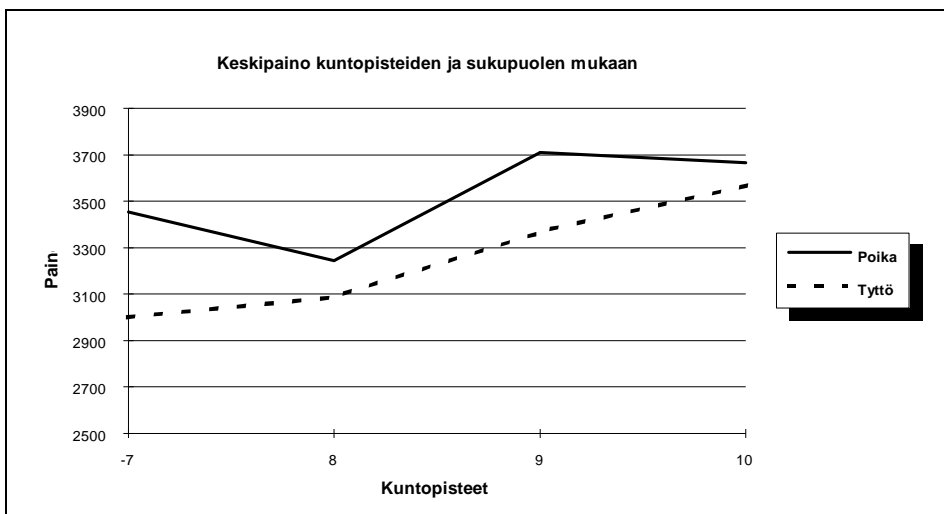
Naisopiskelijoiden joukossa absolutistien osuus on ollut noin kaksinkertainen verrattuna miesten vastaavaan osuuteen.

Jos riippuvana muuttujana on alkoholiin käytetyt markat, saadaan seuraava kuva.



Mielenkiintoista on havaita, että 80-luvun lopun nousuhuuma kasvatti mies- ja naisopiskelijoiden alkoholin käytön eroja.

Toisena esimerkkinä on lapsen syntymäpainon riippuvuus sukupuolesta ja apgar-kuntopisteistä, joiden vaihteluväli on 0-10.



Poikalasten keskipaino on kaikilla kuntopisteiden tasoilla korkeampi kuin tyttölasten keskipaino. Kuntopisteillä on lievä positiivinen vaikutus painoon.

8. Kaksiulotteiset jakaumat

Kaksiulotteisissa jakaumissa tarkastellaan kahden muuttujan yhtäaikaista jakautumista. Tällöin tavoitteena on yleensä selvittää, onko muuttujien välillä *tilastollista riippuvuutta*.

Tilastollisen riippuvuuden määritelmä: Kahden muuttujan välillä vallitsee tilastollinen riippuvuus, jos ensimmäisen muuttujan arvon tunteminen vähentää epätietoisuutta toisen muuttujan arvosta.

Itse asiassa edellisessä luvussa ryhmäkohtaisten tunnuslukujen avulla selvitettiin myös tilastollisen riippuvuuden olemassaoloa. Esimerkiksi alkoholiin käytetyllä markkamäärällä on tilastollinen riippuvuus sukupuoleen. Jos tiedämme, että kyseessä on miesopiskelija, niin silloin on todennäköistä, että hän käyttää rahaa alkoholiin keskimääräistä enemmän verrattuna siihen mitä naisopiskelijat käyttävät.

On huomattava, että edellä todettu koskee tilastollisia tunnuslukuja ja todennäköisyyksiä eikä niiden perusteella voida sanoa mitään *varmaa* yksittäisestä henkilöstä.

Tässä luvussa käsitellään kahta kaksiulotteisen jakauman esitystapaa: *ristiintaulukointia ja korrelaatiodiagrammia*. Ristiintaulukointia käytetään, kun vähintään toinen muuttujista on kvalitatiivinen ja korrelaatiodiagrammia kun molemmat muuttujat ovat kvantitatiivisia.

8.1. Ristiintaulukointi

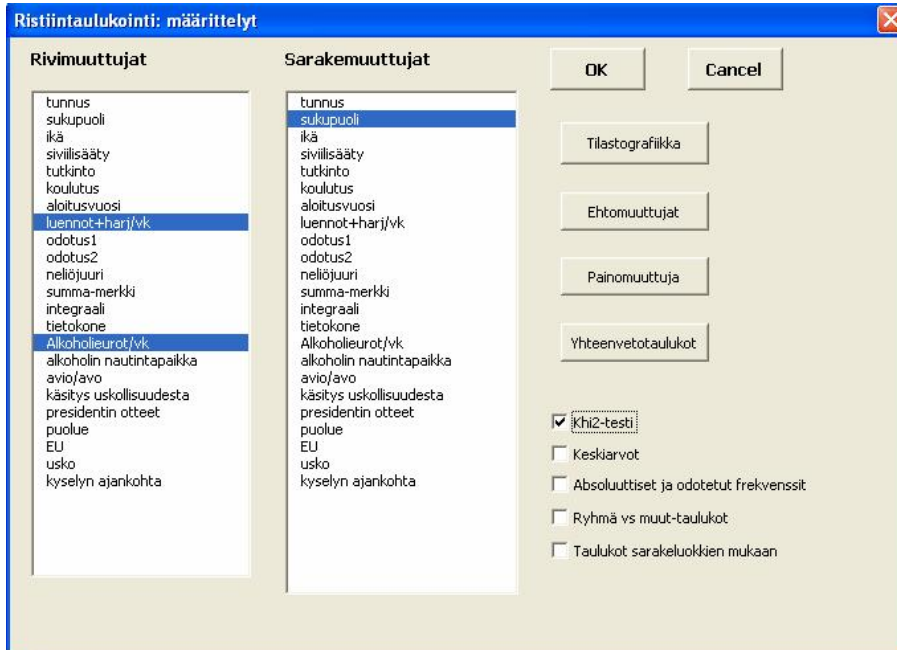
Tarkastellaan edellä ollutta sukupuolen ja alkoholinkäytön esimerkkiä nyt ristiintaulukoinnin avulla. Erityisesti halutaan selvittää absolutistien ja suurkäyttäjien osuudet miehillä ja naisilla, minkä vuoksi alkoholimarkkojen luokat ovat seuraavat: 0, 1-20, 21-50, 51-100, 101-200 ja 201-.

Ristiintaulukko on kaksiulotteinen taulukko, jonka määrittelevät rivi- ja sarakemuuttuja. *Rivimuuttujan* luokat ovat taulukon riveinä ja *sarakemuuttujan* luokat taulukon sarakkeina.

Tixelin tulostustavasta johtuen mahdollinen syy-muuttuja asetetaan sarakemuuttujaksi ja seuraus-muuttuja rivimuuttujaksi.

Tixel pyytää osoittamaan rivi- ja sarakemuuttujan seuraavalla valintaikkunalla. Molempia muuttujia voidaan valita useampia. Tässä tapauksessa lasketaan sukupuolen ristiintaulukot kahden muuttujan suhteen, jotka ovat luentoihin ja harjoituksiin käytetty aika sekä alkoholiin käytetyt markat.

Kuva 8.1.1. Ristiintaulukoinnin määrittelyikkuna



Lopputulokseksi saadaan sarakesuuntaan ja rivisuuntaan laskettujen prosenttilukujen taulukot. Rivisuuntaan laskettujen prosenttien taulukot tulostuvat omalle arkille. Optiona on mahdollisuus saada myös (absoluuttisten) frekvenssien taulukko. Siinä ovat myös reuna- eli marginaalifrekvenssit, jotka ovat yhtä kuin molempien muuttujien yksiulotteiset jakaumat.

Taulukko 8.1.1. Esimerkki ristiintaulukoinnista

Rivimuuttuja: Alkoholiin käytetyt markat/vk

Sarakemuuttuja: Sukupuoli

	<i>Mies</i>	<i>Nainen</i>	<i>Yht.</i>
0	11	22	19
1-20	12	22	19
21-50	18	27	24
51-100	27	18	21
101-200	21	8	12
201-	12	2	5
Yht.	100	100	100

	<i>Mies</i>	<i>Nainen</i>	<i>Yht.</i>
0	134	616	750
1-20	141	617	758
21-50	218	750	968
51-100	320	502	822
101-200	248	216	464
201-	142	59	201
Yht.	1203	2760	3963

Khi²=432.94

Vapausasteet=5

p=0

Kontingensikerto=0.3138
Kontingensikerto on data-analysiin

Prosentit

Sukupuolen ja alkoholimarkkojen riippuvuus tulee ilmi siinä, että absolutisteja on miesten joukossa 11 % ja naisten joukossa 22 %. Suurkuluttajia (yli 200 mk) on miehistä 12 % ja naisista vain 2 %.

Frekvenssit

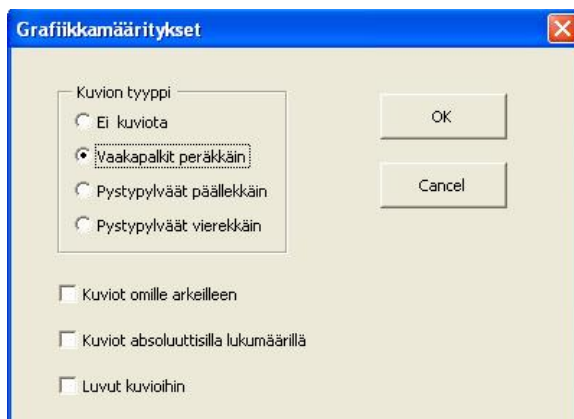
Keskiarvot. Ohjelma laskee rivimuuttujan keskiarvot sarakemuuttujan luokissa ja sijoittaa keskiarvot prosenttitaulukon alareunaan.

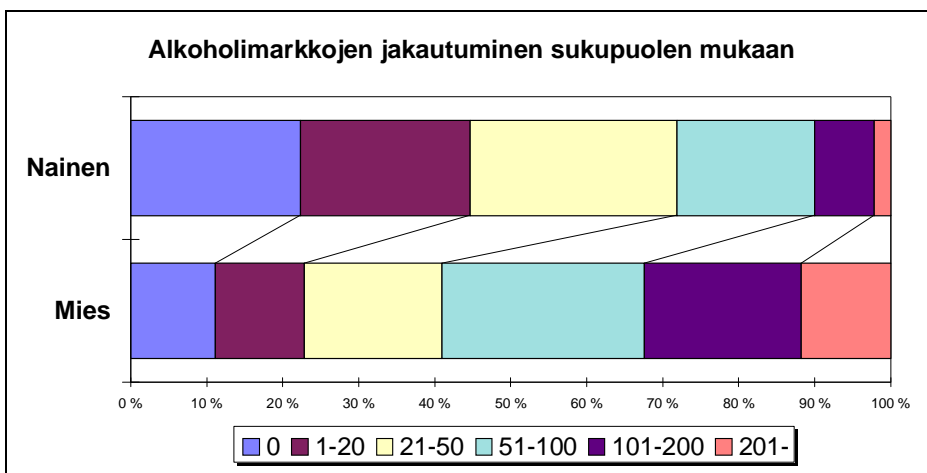
Khi2-testi. Muuttujien riippuvuutta testaavan Khiin neliön testi, siihen liittyvä p-arvo ja merkitsevyytaso sekä kontingenssikertoimen arvo (kts. 8.3) sijoitetaan prosenttitaulukon alapuolelle.

Odotetut frekvenssit. Niitä tarvitaan Khiin neliön testin laskemisessa ja käsitellään tarkemmin luvussa 8.3.

Tilastografiikka

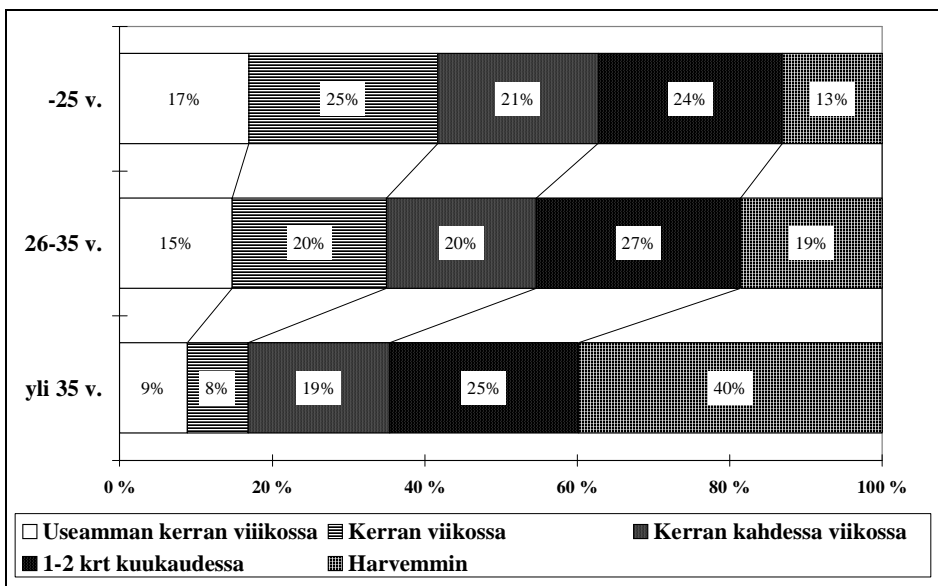
Klikkaamalla *Tilastografiikka*-painiketta saadaan viereinen määrittelyarkki. Siinä on kolme kuviotyppiä, jotka oletusarvoisesti perustuvat ensimmäisen arkin prosenttilukutaulukoihin. Kuviot on mahdollista saada lukumäärätaulukoista merkitsemällä rasti kyseiseen kohtaan. Prosentit tai lukumäärät voidaan tulostaa myös kuvioihin.





Palkkikuvan avulla voidaan arvioida myös ryhmien kumulatiiviset prosenttiosuudet. Niiden vertailun helpottamiseksi kuvaan on lisätty yhteysviivat.

Seuraavassa on iän ja seurusteluravintoloissa käymisen riippuvuutta kuvaava esitys, johon on lisätty palkkien prosenttiluvut.

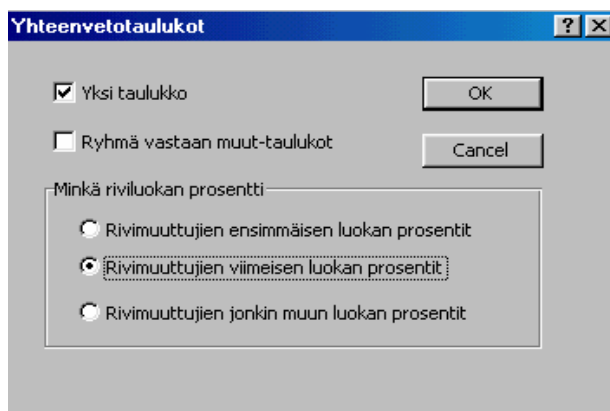


Kuvasta nähdään että iällä ja seurusteluravintoloissa käymisellä on selvä tilastollinen riippuvuus: nuoret käyvät vanhempia ahkerammin seurusteluravintoloissa.

Huom. Jos rivimuuttujia on enemmän kuin 25-30, ei taulukoiden viereen ole syytä sijoittaa kuvia, sillä kuvat vievät paljon käyttömuistia ja ohjelma saattaa pysähtyä ”Not enough memory”-ilmoitukseen.

Yhteenvetotaulukot

Samaan tapaan kuin ryhmäkohtaisissa tunnusluvuissa on ristiintaulukoinnissakin mahdollisuus yhdellä kuvalla havainnollistaa sarakemuuttujan vaikutusta useampaan rivimuuttujaan. Ristiintaulukoinnissa tilanne on kuitenkin hieman mutkikkaampi, koska nyt keskiarvon sijaan on valittava rivimuuttujien jonkin luokan prosentti. Napsauttamalla *Yhteenvetotaulukot*-painiketta saadaan edellä ollut määrittelyarkki. Yleensä yhteenvetotaulukkoon valitaan joko ensimmäisen tai viimeisen riviluokan prosentit. Kuten ryhmäkohtaisissa keskiarvoissa on nykyin tarjolla kaksi yhteenvetotaulukkotyyppiä: Toisessa kaikki prosenttiluvut sijoitetaan yhteen taulukkoon ja toisessa lasketaan jokaisen ryhmän prosenttiluvulle vertailuarvo, joka on muiden kuin tähän ryhmään kuuluvien prosenttiluku. Tulokset ovat vastaavasti *Tulos.xls*-työkirjan *One_table*-arkilla ja *Versus*-arkilla.



Seuraavassa esimerkissä on Tixel-ajossa yhdistetty viisiluokkaisen asteikon kaksi viimeistä luokkaa ”Samaa mieltä”-luokaksi ja tämän luokan neljällä rivimuuttujalla on esitetty yhdessä taulukossa.

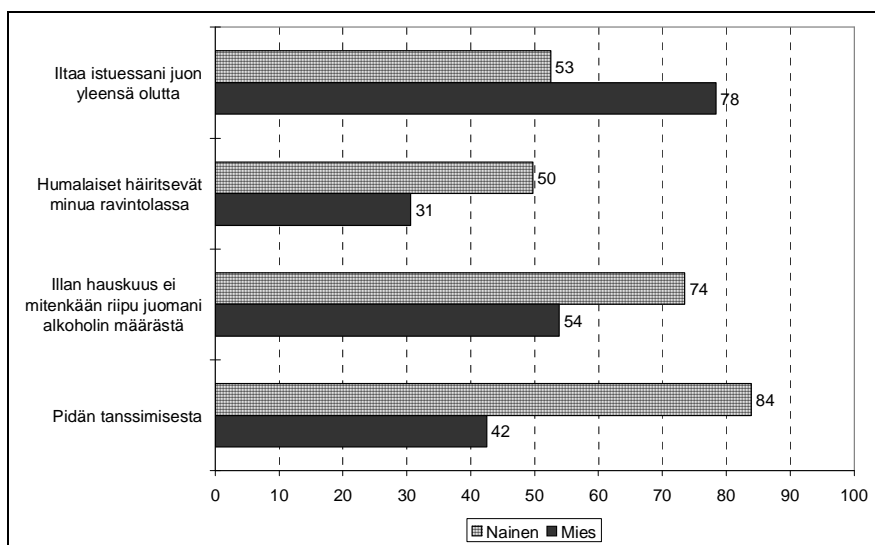
Taulukko 8.1.2. Yhteenvetotaulukko (*One_table*).

Luokan *Samaa mieltä* prosentit
Sarakemuuttuja: 1 Sukupuoli

%	<i>Nainen</i>	<i>Mies</i>	<i>Yht</i>
Iltaa istuessani juon yleensä olutta	52,5	78,4	63,7
Humalaiset häiritsevät minua ravintolassa	49,7	30,6	41,5
Illan hauskuus ei mitenkään riipu juomani a	73,5	53,9	65,1
Pidän tanssimisesta	83,9	42,5	66,0

Tixel muodostaa tämän taulukon perusteella seuraavan kuvan.

Kuvio 8.1.1. Samaa mieltä olevien prosenttiosuudet



Vertaa tätä kuviota keskiarvokuvioon 7.1.2. Kumpi on havainnollisempi?

Esimerkkinä *Ryhmä vastaan muut*-tyyppisistä yhteenvetotaulukoista tarkastellaan samoja muuttujia kuin edellä mutta nyt ryhmittelymuuttujana on ikäryhmä (kolme luokkaa). Saadaan seuraava taulukko.

Taulukko 8.1.3. Yhteenvetotaulukko (Versus-tyyppi)

Luokan *Samaa mieltä* prosentit
Sarakeuuttuja: 2 Ikäryhmä

Alle 20 v. Muut			20-30 v Muut			Yli 30 v. Muut					
	Alle 20 v.	Muut	p-arvo		Muut	p-arvo		Muut	p-arvo		
Iltaa istuessani juon yl	51,5	65,2	0,0296	Iltaa istuessani juon yl	69,4	55,6	0,0005	Iltaa istuessani juon yl	57,1	66,6	0,0254
Humalaiset häiritsevät	28,8	43,0	0,0274	Humalaiset häiritsevät	37,4	47,0	0,0179	Humalaiset häiritsevät	53,5	36,0	0,0001
Illan hauskuus ei miter	62,1	65,4	0,6032	Illan hauskuus ei miter	63,2	67,5	0,2823	Illan hauskuus ei miter	69,4	63,0	0,1339
Pidän tanssimisesta	87,7	63,3	0,0001	Pidän tanssimisesta	61,7	71,8	0,0098	Pidän tanssimisesta	66,3	65,8	0,8995

Taulukossa on prosenttilukujen lisäksi kahden riippumattoman prosenttiluvun testin p-arvo. Esimerkiksi ensimmäisestä osataulukosta nähdään, että *Pidän tanssimisesta*-väitteen kanssa samaa mieltä olevien prosenttiosuus on alle 20-vuotiaiden ryhmässä 87,7 % ja vähintään 20-vuotiaiden ryhmässä (=Muut) vastaavasti 63,3 %. Testisuureen p-arvo on 0,0001 eli näiden prosenttilukujen ero on tilastollisesti erittäin merkitsevä (kts. 13.1).

8.2. Absoluuttiset frekvenssit prosenteiksi

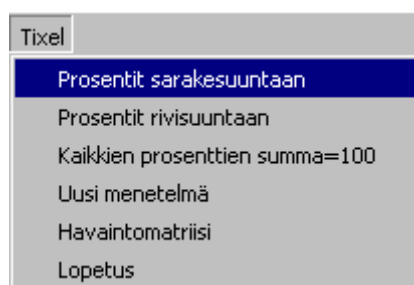
Joskus on tarpeen muuntaa ristiintaulukon absoluuttiset frekvenssit prosenttiluvuiksi. Prosenttiluvut voidaan laskea kolmella eri tavalla: 1) sarakesuuntaan, 2) rivisuuntaan ja 3) siten, että kaikkien prosenttilukujen summa on sata.

Tixelissä on laskentapohja, jolla absoluuttiset frekvenssit voidaan muuntaa prosentteiksi. Tämä pohja saadaan käyttöön valitsemalla Tixel-valikosta vaihtoehto *Laskentapohjat/Lukumäärien muuntaminen prosentteiksi*.

Laskentapohjaan syötetään absoluuttisten frekvenssien taulukko siten, että taulukon vasemman ylänurkan frekvenssi tulee soluun C8.

	A	B	C	D	E	F	G
1	Absoluuttisten frekvenssien muuntaminen prosentteiksi						
2							
3	<div style="border: 1px solid black; padding: 5px;"> <p>Ohje: Syötä absoluuttiset frekvenssit siten, että taulukon vasemman ylänurkan luku (ei luokkatunnus) on solussa C8. Jatko: Valitse Tixel-valikosta sarake-, rivi- tai yhteensä-vaihtoehto.</p> </div>						
4							
5							
6							
7							
8							

Tixel-valikko antaa viereiset vaihtoehdot.



Esimerkki. 3*2-*taulukon* frekvensseistä on ohjelma on laskenut prosenttiluvut siten, että kaikkien prosenttien summa on 100 %.

Taulukko 8.2.1. Frekvenssien muuttaminen prosentteiksi

Frekvenssit		Kaikkien prosenttien summa = 100	
10	15	8,3	12,5
20	25	16,7	20,8
30	20	25,0	16,7

8.3. Kontingenssikerroin ja Cramerin phi

Ristiintaulukon riippuvuusluvuna käytetään *kontingenssikerrointa tai Cramerin phiä*. Ne mittaavat kahden kvalitatiivisen muuttujan välistä riippuvuutta. Muuttajat voivat olla kvantitatiivisiakin, mutta silloin ne on luokiteltava (esim. ikäluokat). Molemmat riippuvuusluvut perustuvat Khiin neliön lausekkeeseen.

Tarkastellaan yleistä ristiintaulukkoa, jossa on g riviä ja h saraketta. Muuttujan x luokkaan X_i ja muuttujan y luokkaan Y_j kuuluvien tilastoyksiköiden lukumäärä on f_{ij} . Ne voidaan esittää seuraavassa taulukkomuodossa.

		x				$yht.$
		X_1	X_2	\dots	X_h	
y	Y_1	f_{11}	f_{12}	\dots	f_{1h}	f_{1*}
	Y_2	f_{21}	f_{22}	\dots	f_{2h}	f_{2*}
	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
	Y_g	f_{g1}	f_{g2}	\dots	f_{gh}	f_{g*}
$yht.$		f_{*1}	f_{*2}	\dots	f_{*h}	n

Taulukkoon on merkitty myös *reuna-* eli *marginaalifrekvenssit*, jotka ovat frekvenssien summia rivi- ja sarakesuunnassa.

Havaittuja frekvenssejä f_{ij} verrataan *odotettuihin frekvensseihin* e_{ij} , jotka ovat odotettuja siinä tilanteessa, että riippuvuutta muuttujien x ja y välillä ei esiinny. Odotetuille frekvensseille saadaan lauseke, kun oletetaan kunkin sarakkeen frekvenssien noudattavan prosentuaalista reunajakaumaa. Odotetun frekvenssin lauseke on

$$e_{ij} = \frac{f_{i*} f_{*j}}{n}$$

Alkoholimarkkojen ja sukupuolen ristiintaulukon odotetut frekvenssit ovat seuraavat (opiskelija-aineisto).

	Mies	Nainen	Yht.
0	227,7	522,3	750
1-20	230,1	527,9	758
21-50	293,8	674,2	968
51-100	249,5	572,5	822
101-200	140,9	323,1	464
201-	61,0	140,0	201
Yht.	1203	2760	3963

Esimerkiksi e_{11} :n arvoksi saadaan $750 \cdot 1203 / 3963 = 227,7$.

Mitä enemmän havaitut frekvenssit poikkeavat odotetuista frekvensseistä, sitä voimakkaampaa on muuttujien välinen riippuvuus. Eräs mittari tälle riippuvuudelle on χ^2 (khiin neliö), jonka lauseke on

$$\chi^2 = \sum_{i=1} \sum_{j=1} \frac{(f_{ij} - e_{ij})^2}{e_{ij}}$$

Tämä ei kuitenkaan ole hyvä riippuvuusluku, koska sillä ei ole mitään ylärajaa.

Ristiintaulukon riippuvuusluvut.

χ^2 :n perustuvat riippuvuusluvut ovat *kontingenssikerroin C* ja *Cramerin phi ϕ_c* . Näiden kaavat seuraavat:

$$C = \sqrt{\frac{\chi^2}{n + \chi^2}} \quad \text{ja} \quad \phi_c = \sqrt{\frac{\chi^2}{n(k-1)}} ,$$

jossa **k** on pienempi ristiintaulukon rivien tai sarakkeiden lukumääristä.

Kontingenssikertoimen huono puoli on, että sen suurin teoreettinen arvo ei ole yksi vaan luku, joka riippuu ristiintaulukon rivien ja sarakkeiden lukumääristä seuraavan kaavan mukaan:

$$C_{\max} = \sqrt{\frac{k-1}{k}} , \text{ jossa } k = \min(g, h).$$

Seuraavassa taulukossa on kontingenssikertoimen maksimiarvoja eräillä **k**:n arvoilla.

Taulukko 8.3.1. Kontingenssikertoimen maksimiarvoja (**k** = pienempi rivien ja sarakkeiden lukumääristä)

k	2	3	4	5	6
----------	----------	----------	----------	----------	----------

C_{\max}	0,71	0,82	0,87	0,89	0,91
------------	------	------	------	------	------

Edellä olleessa esimerkissä (sukupuolen ja alkoholimarkat) on χ^2 :n arvo 432,94, joten kontingenssikerroin on

$$C = \sqrt{\frac{432,94}{3963 + 432,94}} = 0,314.$$

Cramerin phiin ϕ_c vaihteluväli on nolasta ykköseen ja se on siksi suositeltavampi riippuvuusluku ristiintaulukon tapauksessa. Esimerkin tapauksessa saadaan Cramerin phiin arvoksi

$$\phi_c = \sqrt{\frac{432,94}{(2-1) 3963}} = 0,331.$$

8.4. Riippuvuusluku ristiintaulukosta

Joskus voi tulla vastaan tilanne, jossa khiin neliön testi pitää laskea taulukosta, johon liittyvä havaintomatriisi ei ole käytettävissä. Tixel-ohjelmassa on käytettävissä syöttöpohjat, joihin taulukot voidaan syöttää ja sen jälkeen ohjelma laskee testisuureen. Syöttöpohjia on kaksi: absoluuttisten frekvenssien taulukolle ja sarakesuuntaan laskettujen prosenttien taulukolle. Syöttöpohjat saadaan käyttöön valitsemalla Tixel-valikosta *Laskentapohjat/Khi2-testit*-vaihtoehto ja sen jälkeen joko absoluuttisten frekvenssien tai %-frekvenssien Khi2-riippuvuustesti (työar-kin nimet ovat vastaavasti KHI ja KHI%)

Absoluuttiset frekvenssit

Khi2-riippuvuustesti (absoluuttiset frekvenssit)

Ohje: Täytä taulukon luvut absoluuttisilla frekvensseillä. Reunafrekvenssejä ei kirjoiteta. Taulukossa molemmilla muuttujilla on vain kaksi luokkaa, mutta voit lisätä luokkia tarpeen mukaan. Samoin voit kirjoittaa kirjainten sijaan luokille selväkieliset tunnukset. Valitse lopuksi Tixel-valikosta khi2-testisuure.

		Muuttuja Y	
		a	b
Muuttuja	a		
X	b		

Esimerkki. Seuraavassa taulukossa on muuttujien ASEMA (luokat aaa ja bbb) ja SUORITUSTASO (luokat matala, keski ja korkea) ristiintaulukko sekä Tixelin antama tulostus.

		Asema		
		aaa	bbb	
Suoritus- taso	Matala	10	15	Rivien lkm: 3
	Keski	20	15	Sarakkeiden lkm: 2
	Korkea	30	15	Tapausten lkm: 105
				Khiin neliö: 4,67
				Vap.asteet: 2
				Kontingenssikerroin: 0,206
				Riskitaso: 0,097

Riskitaso $p=0,097$ osoittaa, että muuttujien välillä ei ole tilastollisesti merkittävää riippuvuutta.

Prosenttijakaumat

Testi testaa, eroavatko x-muuttujan prosenttijakaumat y-ryhmittelymuuttujan määräämissä luokissa tilastollisesti merkitsevästi toisistaan. Eli onko x- ja y-muuttujien välillä riippuvuutta. Prosenttilukujen lisäksi on tiedettävä ryhmien otoskoot. Testisuuren arvo on aivan sama kuin jos se olisi laskettu vastaavasta absoluuttisten frekvenssien taulukosta.

Khi2-riippuvuustesti (prosenttiluvut)

Ohje: Täytä taulukon solut prosenttiluvuilla (ilman %-merkkiä). Lukujen summien tulee olla sarakeittain sata. Otokoko-riville syötetään otokoot absoluuttisina lukuina (ei %-lukuina). Taulukossa molemmilla muuttujilla on vain kaksi luokkaa, mutta voit lisätä luokkia tarpeen mukaan. Samoin voit kirjoittaa kirjainten sijaan luokille selväkieliset tunnukset. Valitse lopuksi Tixel-valikosta khi2-testisuure.

		Ryhmittelymuuttuja Y	
		a	b
Muuttuja X	Otokoko:		
	a		
	b		

Esimerkki. Seuraavassa taulukossa on muuttujien ASEMA ja SUORITUSTASO ristiintaulukko siten, että sarakesummat ovat sata. Sarakeluokkien otokoot ovat 150 ja 200.

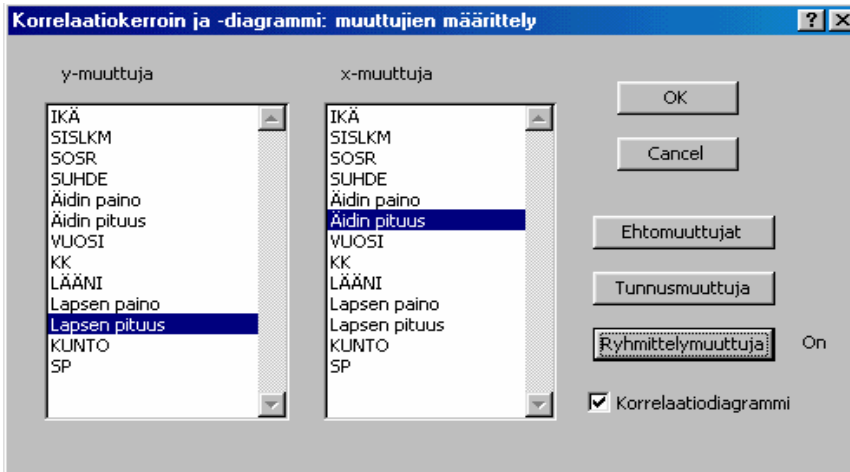
		Asema		
		aaa	bbb	
Suoritus- taso	Otokoko:	150	200	Rivien lkm: 3
	Matala	25	40	Sarakkeiden lkm: 2
	Keski	35	30	Tapausten lkm: 350
	Korkea	40	30	Khiin neliö: 8,91
				Vap.asteet: 2
				Kontingenssikerroin: 0,158
				Riskitaso: 0,012

Tulostus osoittaa, että muuttujien välinen riippuvuus on tilastollisesti merkitsevää riskitasolla $p=0,012$.

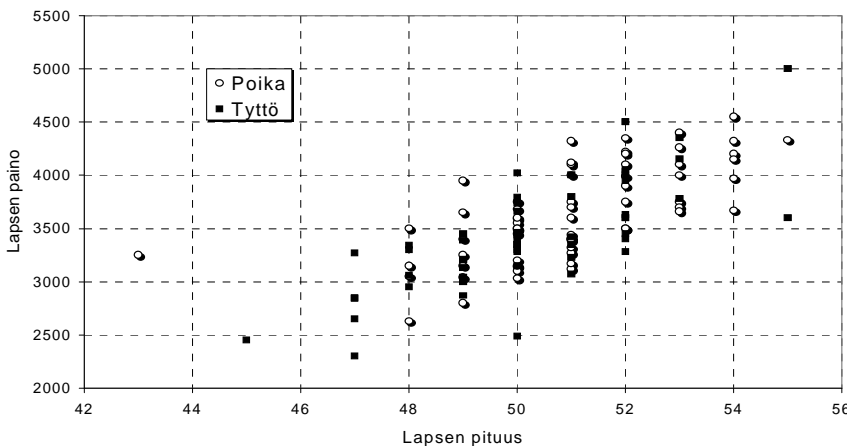
8.5. Korrelaatiodiagrammi

Korrelaatiodiagrammia eli pisteparvea käytetään kahden kvantitatiivisen muuttujan x ja y riippuvuuden kuvaamiseen. Siinä tilastoyksiköt kuvataan pisteinä xy-koordinaatistossa. Jokaisen pisteen viereen on mahdollista saada *tunnusmuuttujan* arvo tai teksti. *Ryhmittelymuuttujan* avulla eri ryhmien pisteet voidaan kuvata eri merkeillä.

Tixelin menetelmävalikosta valitaan menetelmä *Korrelaatiodiagrammi ja korrelaatiokerroin*, joka antaa seuraavan valintaikkunan muuttujien määrittämiseksi.

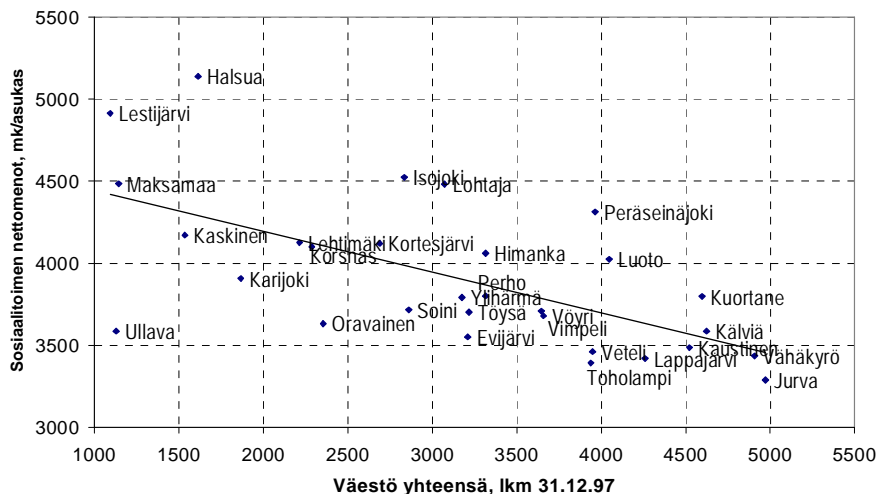


Seuraavassa kuviossa on lapsen syntymäpainon ja -pituuden korrelaatiodiagrammi, jossa ryhmittelymuuttujana on sukupuoli.

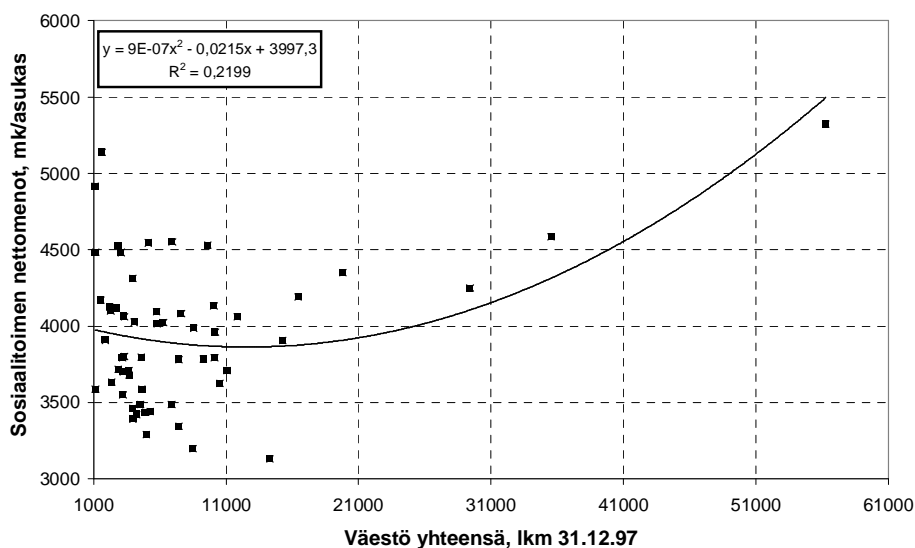


Huomataan että lapsen pituudella ja painolla on positiivinen viivallinen riippuvuus: mitä suurempi on pituus, sitä suurempi on myös keskimäärin paino. Ryhmittelymuuttujan eli sukupuolen vaikutus näkyy siinä, että poikien pisteet keskittyvät jonkin verran enemmän oikeaan ylänurkkaan ja tyttöjen vastaavasti vasempaan alanurkkaan.

Regressioanalyysin avulla saadaan korrelaatiodiagrammiin suora, joka kulkee mahdollisimman keskellä pisteparvea. Itse menetelmää ei käsitellä tässä kirjassa, mutta seuraavaan korrelaatiodiagrammiin regressiosuora on kuitenkin merkitty. Siinä tilastoyksikköinä ovat Etelä-Pohjanmaan väestömäärältään alle viiden tuhannen hengen kunnat ja muuttujina väestömäärä ja sosiaalitoimen nettomenot asukasta kohden. Tunnusmuuttujana on kunnan nimi.



Huomataan että tässä aineistossa muuttujien välillä on negatiivinen riippuvuus: mitä suurempi kunta sitä pienempiä ovat keskimäärin sosiaalitoimen nettomenot henkeä kohden. ($r = -0,627$). Jos mukaan otetaan kaikki Etelä-Pohjanmaan kunnat, niin riippuvuus muuttuu toisentyyppiseksi eli epälineaariseksi. Pienissä ja suurissa kunnissa nettomenot ovat keskimääräistä suurempia. Tätä havainnollistaa seuraava kuvio, johon piirretty toisen asteen käyrä siten, että se kulkee mahdollisimman keskellä pisteparvea.



8.6. Korrelaatiokerroin

Korrelaatiokerroin r mittaa kahden kvantitatiivisen muuttujan **viivallista eli lineaarista riippuvuutta**. Viivallinen riippuvuus voi olla positiivinen tai negatiivinen. Positiivisessa tapauksessa y :n arvo kasvaa keskimäärin, kun x :n arvo kasvaa (lapsen syntymäpituuden keskiarvo kasvaa, kun äidin pituus kasvaa). Negatiivisessa tapauksessa y :n keskimääräinen arvo pienenee, kun x :n arvo kasvaa (energiaan kulutus pienenee, kun ulkolämpötila nousee).

Korrelaatiokertoimen kaavaa varten määritellään ensin muuttujien x ja y **kovarianssi**

$$s_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n}$$

Korrelaatiokerroin r on muuttujien x ja y standardoitujen esitysten kovarianssi, jolle saadaan johdettua seuraava kaava

$$r = \frac{s_{xy}}{s_x s_y}$$

eli korrelaatiokerroin on kovarianssi jaettuna keskihajontojen tulolla.

Korrelaatiokertoimen vaihteluväli on $-1 \leq r \leq 1$.

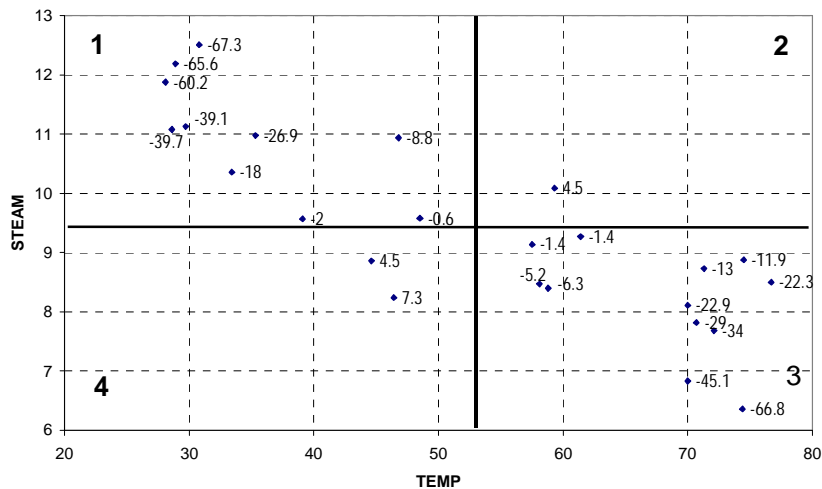
$r = 1$	Täydellinen positiivinen lineaarinen riippuvuus. Kaikki havaintopisteet sijaitsevat nousevalla suoralla.
$r = 0$	Muuttujien välillä ei ole lainkaan lineaarista riippuvuutta. Epälineaarista riippuvuutta voi kuitenkin olla.
$r = -1$	Täydellinen negatiivinen lineaarinen riippuvuus. Kaikki havaintopisteet sijaitsevat laskevalla suoralla.

Tarkastellaan tarkemmin kovarianssin lauseketta, joka on korrelaatiokertoimen ydin. Havaintoaineistona on kuukausitilasto, jossa on erään tuotantolaitoksen höyryenergian kulutus ja kuukauden keskilämpötila Fahrenheit-asteina. Kuvioon 8.6.1 on merkitty jokaisen pisteen viereen tulo

$$(x_i - \bar{x})(y_i - \bar{y})$$

arvo. Tämä tehdään siten, että lisätään havaintomatriisiin muuttujaksi kyseinen tulo ja valitaan korrelaatiodiagrammi-menetelmässä tunnusmuuttujaksi kyseinen tulomuuttuja.

Kuvio 8.6.1. Kuukauden keskilämpötila ja höyryenergian kulutus.



Lisäksi kuvio on jaettu keskiarvojen kautta kulkevilla suorilla neljään osaan, joiden numerot on merkitty ylä- ja alanurkkiin. Huomataan että 2. ja 4. osassa tulon arvo on positiivinen ja 1. ja 3. osassa negatiivinen. Kovarianssi on näiden tulojen keskiarvo. Jos 2. ja 4. osassa on selvästi enemmän pisteitä kuin 1. ja 3. osassa, on korrelaatio positiivinen. Päinvastaisessa tilanteessa korrelaatio on negatiivinen.

Kuviossa 8.6.1 korrelaatiokertoimen arvo on $-0,85$.

Pisteeseen liittyvän tulon itseisarvo on sitä suurempi mitä lähempänä piste on kuvion jotain neljää kulmaa. Jos havaintoja on vähän, saattaa yksikin poikkeava piste vaikuttaa voimakkaasti korrelaatiokertoimen arvoon. Näin on asia varsinkin jos tilastoyksiköiden koko vaihtelee paljon kuten esim. kuntien ja yritysten tapauksissa.

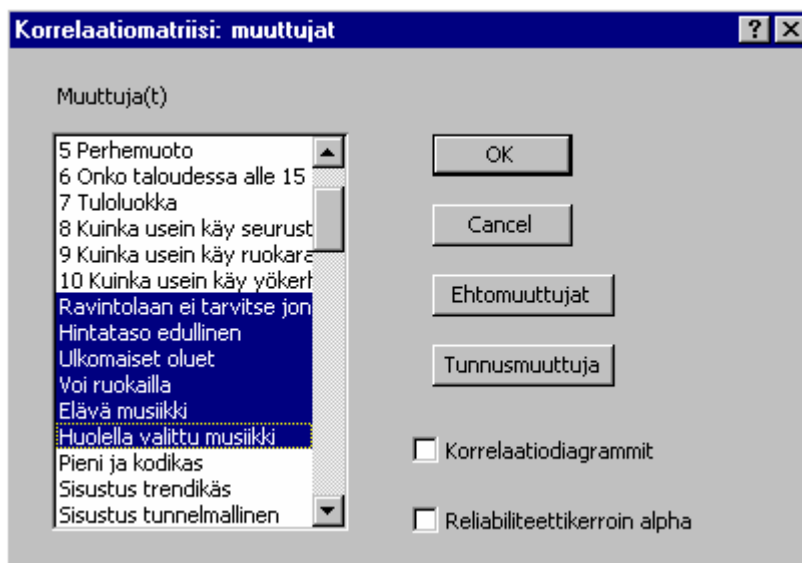
Korrelaatiokertoimen neliö kuvaa, kuinka paljon toinen muuttuja (x) selittää toisen muuttujan (y) vaihtelusta. Tätä neliötä kutsutaan *selitysasteeksi* ja ilmaistaan usein sadalla kerrottuna prosentteina. Edellisessä esimerkissä saadaan selitysasteeksi $(-0,85)^2\% = 72\%$.

8.8 Korrelaatiomatriisi

Korrelaatiomatriisi on taulukko, jossa riveinä ja sarakkeina ovat muuttujat ja taulukon alkiaina rivi- ja sarakemuuttujien korrelaatiot. Matriisin lävistäjä-alkiot ovat ykkösiä.

Tixel-ohjelman menetelmävalikosta valitaan Korrelaatiomatriisi-vaihtoehto. Muuttujat määrätään seuraavalla valintaikkunalla.

Kuva 8.8.1. Korrelaatiomatriisin määrittelyikkuna



Seuraava taulukko on seurusteluravintolan valintaan vaikuttavan kuuden muuttujan korrelaatiomatriisi (ravintola-aineisto).

	Ravintolaan ei tarvitse jonottaa	Hintataso edullinen	Ulkomaiset oluet	Voi ruokailla	Elävä musiikki	Huolella valittu musiikki
Ravintolaan ei tarvitse jonottaa	1,00	0,21	0,18	0,21	0,11	0,09
Hintataso edullinen	0,21	1,00	0,13	-0,01	0,16	0,10
Ulkomaiset oluet	0,18	0,13	1,00	0,26	0,20	0,15
Voi ruokailla	0,21	-0,01	0,26	1,00	0,28	0,28
Elävä musiikki	0,11	0,16	0,20	0,28	1,00	0,35
Huolella valittu musiikki	0,09	0,10	0,15	0,28	0,35	1,00

Suurin korrelaatio 0,35 on muuttujien ”Elävä musiikki” ja ”Huolella valittu musiikki” välillä. Lähes nolla-korrelaatio on muuttujien ”Hintataso edullinen” ja ”Voi ruokaila” välillä.

Korrelaatiokertoimet sisältävän arkin lisäksi tulostiedostossa on kaksi arkkia: toisella arkilla (n-values) on korrelaatiokertoimien laskennassa havaintoyksiköiden lukumäärät ja toisella arkilla (p-values) korrelaatiokertoimien 2-suuntaiseen testaukseen liittyvät p-arvot.

Korrelaatiokertoimien lisäksi ohjelma muodostaa muuttujaparien korrelaatiogrammeja, jos kyseinen optio valitaan. Koska graafiset diagrammit vievät käyttömuistia runsaanlaisesti, on muuttujien määrä tässä tapauksessa rajattu kahdeksaan. Kahdeksalla muuttujalla saadaan 28 korrelaatiogrammia.

8.9. Korrespondenssianalyysi

Korrespondenssianalyysin avulla voidaan ristiintaulukolle antaa vielä yksinkertaisempi graafinen kuvaus kuin edellä olivat pylväs- tai palkkikuviot. Matematiikaltaan korrespondenssianalyysi on niin monimutkainen, että siitä ei tässä kannata mainita riviäkään.

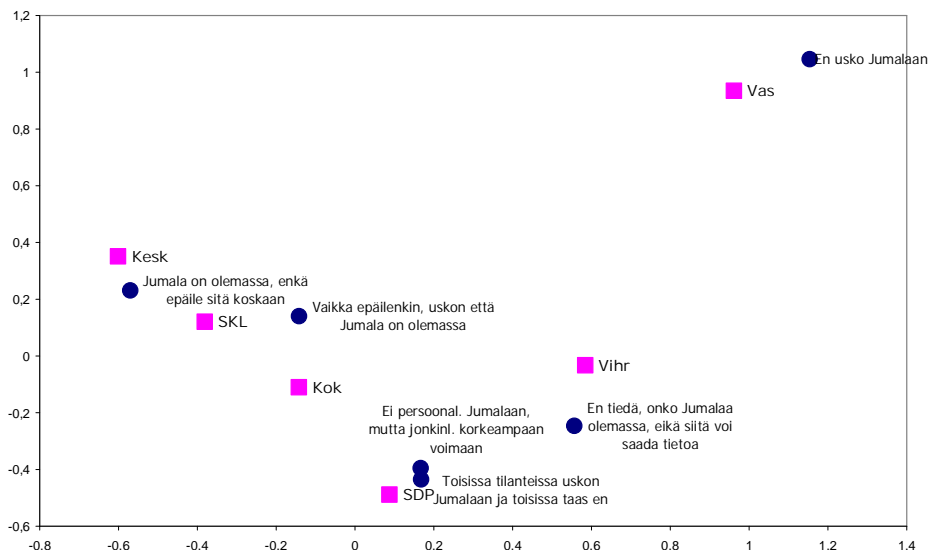
Tarkastellaan seuraavaa ympäristö-aineistosta laskettua esimerkkiä, jossa muuttujina ovat puoluekanta (kuusi puoluetta) ja suhde Jumalaan.

Taulukko 8.9.1. Suhde Jumalaan puoluekannan mukaan

%	SDP	Kesk	Kok	Vas	Vihr	SKL	Yht.
En usko Jumalaan	4	4	5	20	11	6	6
En tiedä, onko Jumalaa olemassa, eikä siitä voi saada tietoa	13	7	10	15	13	2	11
Ei persoonal. Jumalaan, mutta jonkinl. korkeampaan voimaan	30	18	26	20	28	30	25
Toisissa tilanteissa uskon Jumalaan ja toisissa taas en	9	6	8	5	9	6	8
Vaikka epäilenkin, uskon että Jumala on olemassa	30	18	26	20	28	30	25
Jumala on olemassa, enkä epäile sitä koskaan	23	39	30	20	16	34	27
Yht.	100	100	100	100	100	100	100
N	219	200	185	60	152	50	866

Valitsemalla ristiintaulukoinnin jälkeen *Tixel-muotoilut* valikosta *Korrespondenssianalyysi*, saadaan seuraava kuvio.

Kuvio 8.9.1. Muuttujien suhde Jumalaan ja puoluekannan korrespondenssianalyysi



Kuvio antaa kaksiulotteisen kuvauksen ristiintaulukosta siten, että molempien muuttujien luokat esitetään pisteinä. Myönteisin väite *Jumala on olemassa enkä epäile sitä koskaan* sijaitsee lähellä Keskustapuoluetta ja Suomen kristillistä liittoa. Vasemmistoliitto poikkeaa muista puolueista eniten ja on lähellä luokkaa *En usko Jumalaan*. Kokoomus, Sdp ja Vihreät ovat näiden kahden ääripään välimaastossa.

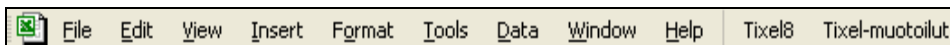
9. Tixel-ohjelma

9.1. Yleistä

Tilastollinen Tixel-ohjelma on tehty Visual Basic-ohjelmointikielellä (VBA) Excel-ympäristöön. Excel versioista käyvät 2000 tai uudempi, ja käyttöympäristö voi olla PC tai Mac. Kielen suhteen ei ole rajoituksia (suomi/englanti). Tixelin valikkojen ja tulosten kieleksi voidaan valita suomi tai englanti. Myös muut kielet ovat mahdollisia.

Tixel käynnistetään napsauttamalla Windowsin työpöydällä olevaa kuvaketta, jossa lukee *Tixelin käynnistys* tai jotain vastaavaa. Tällöin avautuu myös Excel, jos Excel ei ole jo avattu.

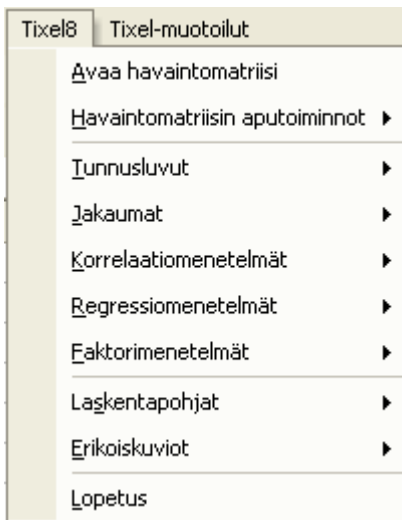
Kun Tixel avattu, valikkorivin loppuun ilmestyy kaksi uutta valikkoa:



Tixel8-valikossa ovat kaikki Tixelin perustoiminnot:

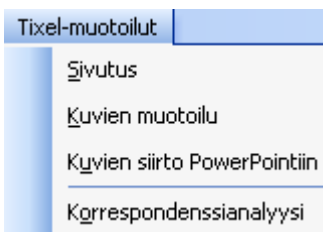
- havaintomatriisin avaaminen
- havaintomatriisin virheettömyyden tarkistaminen
- tilastollisen menetelmän valinta
- syöttöpohjiin perustuvien menetelmien valinta
- erikoiskuviot
- Tixelin lopetus

Tixel8-valikko:

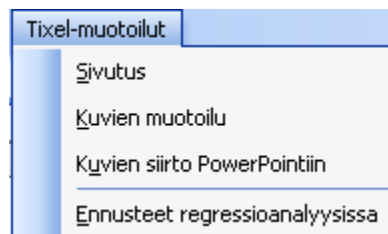


Tixel-muotoilut-valikossa on Tixel-tuloksiin liittyviä toimintoja. Vaihtoehdot riippuvat siitä mitä menetelmää on käytetty. Kolme ensimmäistä vaihtoehtoa ovat samat. Neljäs vaihtoehto tulee esiin, kun käytetty menetelmä on ristiintaulukointi tai regressioanalyysi.

Ristiintaulukointi



Regressioanalyysi



9.2. Tixelin menetelmät

Tilastolliset menetelmät on ryhmitelty seuraavasti:

<u>H</u> avaintomatriisin aputoiminnot ▶	<u>H</u> avaintomatriisin tarkistaminen <u>T</u> ekstilukujen muunto luvuiksi <u>S</u> ummatiivisten muuttujien muodostaminen <u>M</u> ultiresponse-muuttujan muodostaminen <u>K</u> ahden havaintomatriisin yhdistäminen
--	---

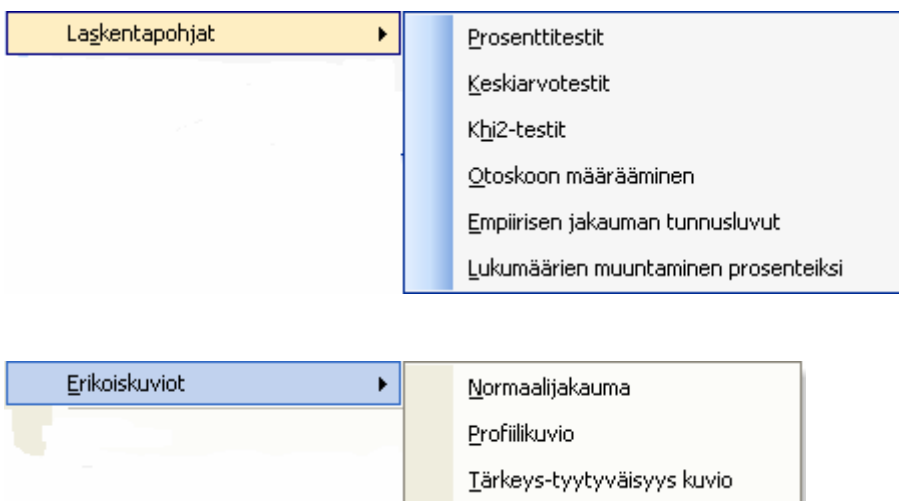
<u>T</u> unnusluvut ▶	<u>T</u> unnusluvut ilman ryhmittelyä <u>T</u> unnusluvut, yksi ryhmittelymuuttuja <u>T</u> unnusluvut, kaksi ryhmittelymuuttujaa <u>K</u> aksisuuntainen varianssianalyysi
-----------------------	--

<u>J</u> akaumat ▶	<u>Y</u> ksiulotteinen jakauma <u>R</u> istiintaulukointi
--------------------	--

<u>K</u> orrelaatiomenetelmät ▶	<u>K</u> orrelaatiokerroin ja -diagrammi <u>K</u> orrelaatiomatriisi ja reliabiliteetti
---------------------------------	--

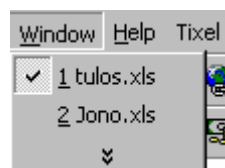
<u>R</u> egressiomenetelmät ▶	<u>R</u> egressioanalyysi <u>V</u> aiheittainen regressioanalyysi
-------------------------------	--

<u>F</u> aktorimenetelmät ▶	<u>F</u> aktorianalyysi <u>P</u> ääkomponenttianalyysi
-----------------------------	---



9.3. Tixelin käyttö

Tixel-ajon jälkeen tulokset ovat näytöllä Tulos.xls-työkirjassa. Havaintomatriisi on edelleen auki ja se saadaan näytölle valitsemalla Excelin perusvalikosta Window-vaihtoehto ja siitä edelleen havaintomatriisin nimi (viereisessä esimerkissä Jono.xls). Tämän jälkeen voidaan esim. luokituksiin tehdä muutoksia. Jos muutokset ovat vain väliaikaisia eli niiden halutaan olevan voimassa vain kyseisen Tixel-istunnon ajan, ei havaintomatriisia tarvitse tallentaa. Sen sijaan jos muutokset halutaan pysyviksi, on havaintomatriisi tallennettava.



Jos Tixel-ajon tulokset halutaan tallentaa, on se tehtävä ennen uutta ajoa käyttäen jotain muuta nimeä kuin Tulos.xls.

Havainnollisuuden vuoksi sijoitetaan vielä tulostiedosto ja havaintomatriisi vierekkäin näytölle. Tämä saadaan aikaan **Window/Arrange**-komennolla. Käytännössä näin ei tietenkään tarvitse tehdä.

Kuva 9.2.1. Tixel-tulos ja havaintomatriisi yhtaikaa näytöllä.

	A	B	C	D	E
1	Yksiuotteinen jakauma				
2	Jonotusaineisto				
3					
4	Viikonpäivä				
5		<i>Lkm</i>	<i>%</i>	<i>Kum. lkm</i>	<i>Kum-%</i>
6	ma	128	17	128	17
7	ti	103	14	231	31
8	ke	125	17	356	48
9	to	101	14	457	61
10	pe	89	12	546	73
11	la	101	14	647	87
12	su	99	13	746	100
13	Yht.	746	100	746	100

	A	B	C	D	E
1	Jonotusaineisto				
2	-1	-1	-1		-1
3	Viikon päivä	Ikä	Sukupuoli	Kotikunta	qve
4	1	64	1	1	4
5	1	5	2	2	1
6	1	34	1	1	1
7	1	10	1	1	2
8	1	55	2	2	2
9	1	35	2	2	3
10	1	18	1	1	5
11	1	5	2	2	1
12	2	40	1	1	2
13	2	11	1	1	1

Huom. Havaintoaineisto on tallennettava levyllä aina ennen Tixelin käynnistämistä.

10. Excel

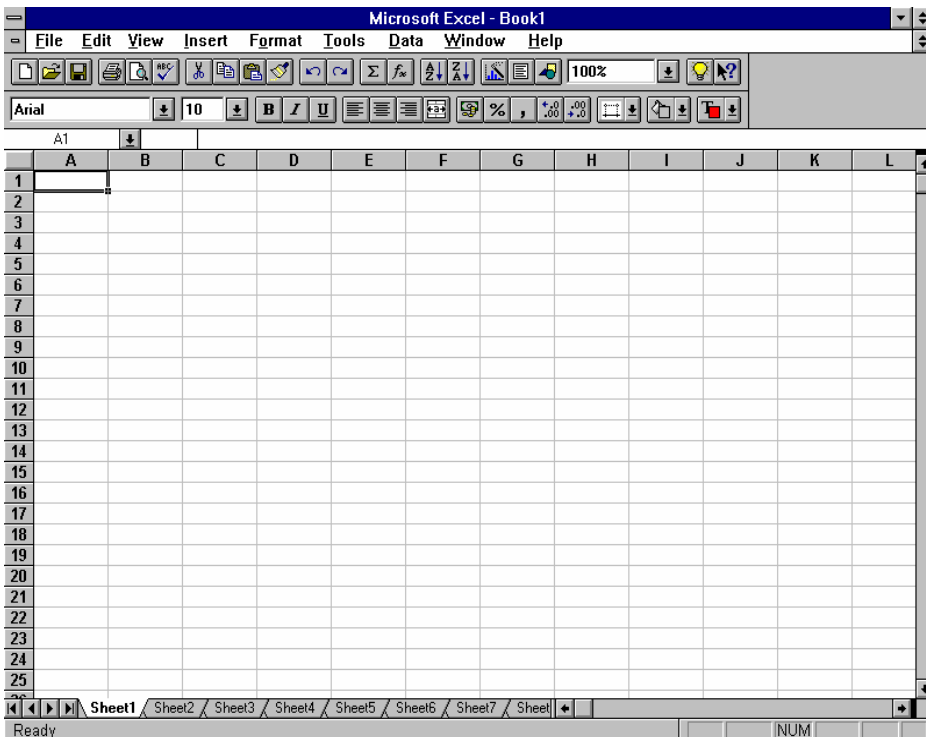
Seuraavassa käsitellään lyhyesti Tixelin käytön kannalta Excelin tärkeimpiä ominaisuuksia. Excelin grafiikkaa käsitellään luvussa 11.

10.1. Yleistä

Excel-taulukkolaskennasta on sekä PC- että Mac-versiot, jotka ovat toiminnoltaan lähes identtiset. Tässä luvussa käsitellään PC-Excel 2000:ta.

Taulukkolaskenta viittaa laskentaan. Laskennan lisäksi Exceliin sisältyy grafiikka, tietokanta (eli kortisto) ja Visual Basic-ohjelmointikieli.

Excel on Windows-ohjelma, minkä vuoksi on ensin käynnistettävä Windows-käyttöliittymä. Excel käynnistyy kaksoisnapsauttamalla Excel-kuvaketta. Jos Excel-kuvaketta ei ole näkyvässä Windowsin käynnistyttyä, ohjelma voidaan valita Start-painikkeen antamasta luettelosta. Näytölle ilmestyy seuraavanlainen **työarkki**, "elektroninen A4":



Työarkki muodostuu *riveistä* (rows) ja *sarakkeista* (columns), joiden leikkauskohdissa ovat *solut* (cells). Riveihin viitataan kokonaisluvuilla 1,2,3,...,65536 ja

sarakkeisiin kirjaimilla tai kahden kirjaimen yhdistelmillä A,B,C,...,Y,Z,AA,AB,...,IV. Sarakkeiden lukumäärä on 256. Soluihin viitataan sarake&rivi-osoitteella. Esimerkiksi vasemman ylänurkan solun osoite on A1, sen alla olevan solun osoite on A2 jne.

Yksi työarkin soluista on *aktiivinen solu*, minkä merkiksi kyseinen solu on ympäröity suorakaiteella. Kirjoitettava asia sijoittuu aina aktiiviseen soluun. Ristinmuotoinen merkki on hiiren kohdistin.

Voit siirtää aktiivista solua kolmella eri tavalla:

1. Siirrä hiiri haluamasi solun kohdalle ja napsauta.
2. Siirrä aktiivista solua nuoli-näppäinten avulla (ylös, alas, oikealle, vasemmalle)
3. Paina toimintonäppäintä <F5> ja kirjoita solun osoite.

Page Up- ja Page Down-näppäinten avulla voit selata työarkkia pystysuorassa suunnassa näyttöittäin.

Ylhäältä lukien työarkissa voidaan erottaa seuraavat osat:

Otsikkorivi, jossa on käytetyn ohjelman nimi (Microsoft Excel) ja työkirjan nimi (esim. Book1). Vasemmalla on järjestelmävalikko ja oikealla painikkeet, joilla ikkuna voidaan pienentää kuvakkeeksi ja suurentaa maksimikokoonsa.

Valikkorivi, jossa ovat kaikki komennot (File=Tiedosto; Edit=Muokata jne). Napsauttamalla hiirellä komentosanaa aukeaa ns. alavetovalikko, jossa on tarkemmin kyseiseen komentoon liittyvät toiminnot.

Työkaluvalikko sisältää usein käytettyjä toimintoja, kuten tiedoston avaaminen ja tallennus, solun sisällön lihavointi ja asemointi solun sisällä (vasemmalle, keskelle ja oikealle). Työkaluvalikkorivejä voi olla useampia.

Kaavarivillä (eli syöttörivillä) näkyy aktiivisen solun sisältö. Vasemmalla puolella on solun osoite sekä syötön hyväksymis- ja peruutuspainikkeet, jotka ovat näkyvissä kun tietoa syötetään. Solun sisältöä voit muokata paitsi kaavarivillä myös itse solussa.

Excelin **työkirja** voi sisältää suurimmillaan 256 työarkkia. Arkkien nimet ovat näytön alareunassa olevalla arkkirivillä. Tietyille arkille siirrytään napsauttamalla arkin nimeä. Vasemmassa alanurkassa olevilla neljällä nuolinäppäimellä voidaan myös siirtyä arkilta toiselle. Arkkien nimien jälkeen on tällä rivillä *vierityspalkki*, jonka avulla työarkkia voidaan selata vaakasuorassa suunnassa.

31	1	57	2
32	1	3	2
33	1	12	2

◀ ▶ ↻ 🔍 \Data / Classes / RepValues /

Työarkin oikeassa reunassa on pystyvierityspalkki, jonka avulla työarkkia voidaan selata pystysuorassa suunnassa.

Alin rivi on *tilarivi*, jossa näkyy lyhyitä ilmoituksia.

Soluihin voidaan syöttää lukuja, tekstitietoa ja kaavoja. Tixel-ohjelman yhteydessä Exceliä käytetään lähinnä havaintomatriisin tallentamiseen ja kuvien editointiin.

10.2. Havaintomatriisin syöttö työarkille

Havaintomatriisissa erotetaan neljä osaa:

1. Aineiston *otsikko* tulee soluun A1, joka voi olla myös tyhjä. Otsikko tulostuu automaattisesti kaikkien Tixelin tulosteiden alkuun. Otsikon pituudelle ja sanojen lukumäärälle ei ole mitään rajoituksia.
2. Toiselle riville kirjoitetaan *puuttuvien tietojen* merkkeinä käytetyt luvut. Silloin Tixel jättää automaattisesti puuttuvat tiedot tilastokäsittelyn ulkopuolelle. Puuttuvan tiedon oletusarvo tyhjä solu, jolloin toinen rivi jätetään tyhjäksi.
3. Kolmannelle riville kirjoitetaan *muuttujien nimet*.
4. Neljännestä rivistä eteenpäin kirjoitetaan muuttujien tilastoyksiköillä saamat *arvot*.

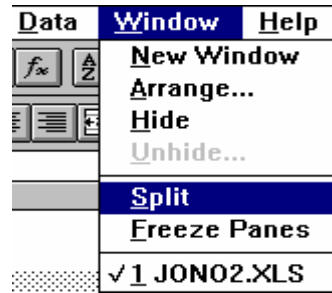
Tekstin ja lukujen kirjoittaminen työarkille on varsin yksinkertaista. Napsauta hiirellä sitä solua, johon haluat kirjoittaa ja sen jälkeen kirjoita teksti tai luku. Vaihtoehtoisesti vie nuolinäppäimillä kohdistin solun kohdalle ja ala kirjoittaa. Kun olet valmis, paina Enter- tai nuolinäppäintä tai napsauta hiirellä hyväksymispainiketta (v-merkki kaavarivillä).

Seuraavassa esimerkissä on syötetty terveyskeskuksen asiakastutkimuksen aineistosta viiden henkilön tiedot kuudesta muuttujasta (viikonpäivä, ikä, sukupuoli, kotikunta, tulosyy ja jonotusaika minuutteina). Jatkossa tätä aineistoa kutsutaan jonotusaineistoksi.

	A	B	C	D	E	F	G
1	Jonotusaineisto						
2	-1	-1	-1	-1	-1	-1	-1
3	Viikon päivä	Ika	Sukupuoli	Kotikunta	Tulosyy	Jonotus aika	
4	1	64	1	4	2	180	
5	1	5	2	1	8	155	
6	1	34	1	1	8	160	
7	1	10	1	2	1	30	
8	1	55	2	2	2	135	
9							
10							

Kun tietoja syötettäessä saavutetaan näytön alareuna, alkaa näyttö vierä ä eli yläosan rivejä häipyä pois ja alhaalta tulee uusia rivejä tilalle. Tällöin myös muuttujien nimirivi eli kolmas rivi siirtyy näkymättömiin, mikä vaikeuttaa tietojen syötöä. Työarkin ikkuna voidaan kuitenkin jakaa vaakasuorassa suunnassa kahteen osaan siten, että kolme ensimmäistä riviä ovat näytöllä koko ajan.

Tee A4 aktiiviseksi soluksi ja valitse Window-valikosta **Split**-vaihtoehto.



Seuraavassa taulukossa näkyy jonotusaineiston kolme ensimmäistä riviä sekä viimeiset rivit. Riviltä 751 alkavat luokitukset, joista on tarkemmin tietoa aluvussa 2.2.

	A	B	C	D	E	F
1	Jonotusaineisto					
2	-1	-1	-1	-1	-1	-1
3	Viikon päivä	Sukupuoli	Kotikunta		Tulosyy	Jonotusaika
747	5	54	1	4	2	95
748	5	55	2	5	1	15
749	5	18	2	1	9	75
750						
751	ma	5 nainen		Tapaturma		60
752	ti	10 mies		Heng.tietulehdus/kuume		120
753	ke	20		Virtasatieoire		180
754	to	30		Vatsavaiva		240
755	pe	40		Mielenterveys/päihdeoire		300
756	la	50		Rintakipu		360
757	su	60		Hengenahdistus		420
758		70		Korvakipu		
759		80		Niveloire/vanha oire		
760		90		Muu		
761		100				
762						

Näytön jako saadaan pois **Window/Remove Split**-komennolla.

Näyttö voidaan jakaa myös yhtäaikaisesti rivi- ja sarakesuunnassa. Tätä ominaisuutta voidaan hyödyntää, kun halutaan selata havaintomatriisia. Yleensä havaintomatriisin ensimmäiseen sarakkeeseen sijoitetaan tilastoyksikön tunnistetieto (esim. kyselylomakkeen nro tai yrityksen nimi). Kun havaintomatriisia selaillaan, on oleellista tietää, mistä tilastoyksiköstä ja muuttujasta jokaisen solun kohdalla

on kyse. Ensimmäinen sarake ja kolmas rivi saadaan näytölle pysyvästi seuraavalla tavalla.

1) Aktivoi solu B4.

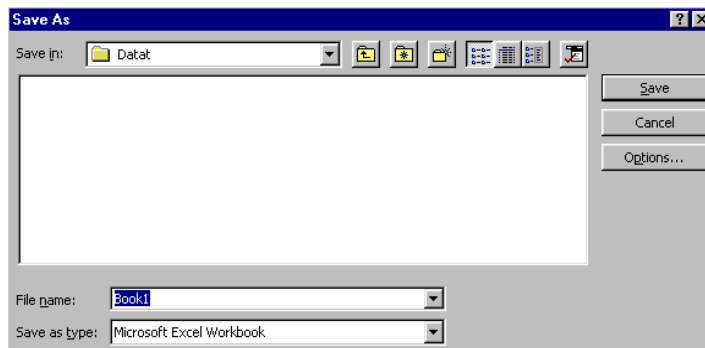
2) Valitse Window-valikosta *Split*-vaihtoehto.

Näyttö jakautuu neljään osaan. Voit nyt liikkua alemmassa oikeanpuoleisessa osassa siten, että tilastoyksikkö- ja muuttuja-tiedot ovat koko ajan esillä.

3	Kunta	Ikä0-14 %	Ikä15-64 %	Ikä65- %	Keskias- teen koulutus	Korkea- asteinen koulutus
65	Forssa	19,3	67,9	12,8	5291	963
66	Lahti	18,2	69,5	12,3	26375	5586
67	Mänttä	20,9	69,4	9,7	2354	529
68	Nokia	19,5	68,9	11,6	6120	997
69	Riihimäki	18	68,8	13,2	6654	1623
70	Tampere	17,1	69,5	13,4	49655	13370
71	Toijala	18,5	66,9	14,6	2086	392
72	Valkeakoski	19,8	68,7	11,5	6349	1313
73	Virrat	17,6	65,9	16,5	2356	333
74	Turku	17	68,8	14,2	44838	14810

10.3. Havaintomatriisin tallentaminen ja avaaminen

Kun havaintomatriisi tai sen osa syötetty työarkille, se tulee tallentaa joko kova-levylle tai disketille. Tallennus tehdään ensimmäisellä kerralla **File/Save As**-komennolla, jonka valintaikkuna on seuraavanlainen:



Määrittele oikea hakemisto (**Save in**) ja kirjoita nimikenttään tiedoston nimi. Esimerkissä Excel tarjoaa oletusnimeä Book1. Koska nimikenttä on mustennettu, voidaan uusi nimi kirjoittaa suoraan oletusnimen päälle ilman että oletusnimi ensin poistetaan. Uuden nimen tarkentimeksi tulee taulukon tapauksessa auto-

maattisesti .XLS, joten sitä ei tarvitse kirjoittaa. Napsauta lopuksi OK:ta tai paina Enteriä.

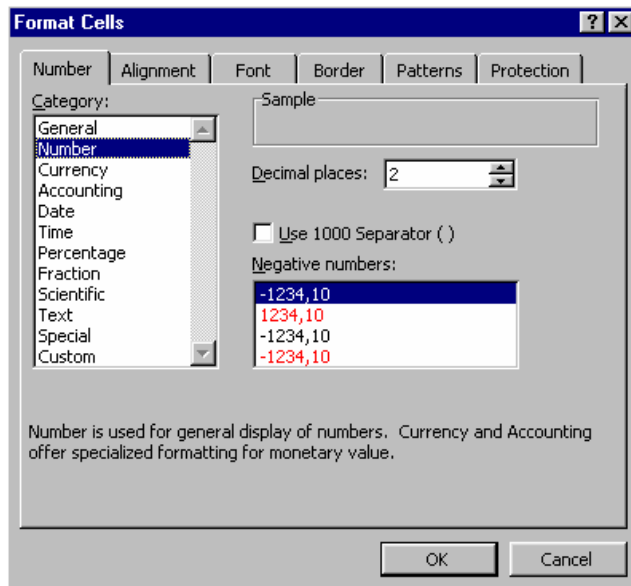
Jos olet aikaisemmin tallentanut osan havaintomatriisista ja haluat jatkaa tallentamista, avaa tiedosto komennolla **File/Open**. Kun myöhemmin tallennat havaintomatriisin, voit käyttää **File/Save**-komentoa, jolloin Excel ei kysy erikseen tiedoston nimeä, vaan käyttää alkuperäistä nimeä.

10.4. Lukujen muotoilu

Tixel-ohjelman tulostaulukoissa luvut on yleensä esitetty useamman desimaalin tarkkuudella. Tällainen tarkkuus on usein liian suuri esityksiä ajatellen, minkä vuoksi lukujen esitysmuotoa on muutettava.

Merkitse ("maalaa") ensin alue, jonka lukujen esitysmuotoa haluat muuttaa. Anna sen jälkeen komento **Format/Cells**, joka tuo näytölle kuuden arkin määrittelyikkunan. Valitse siitä *Number*-arkki.

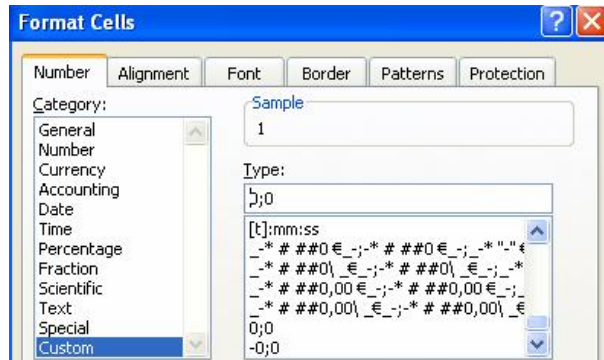
Lukujen muotoiluvaihtoehdot on jaettu *luokkiin* (Category), joista käytetyimmät ovat *Number*- ja *Percentage*-luokat. Desimaalit-ikkunaan (Decimal places) voit valita sopivan määrän desimaaleja.



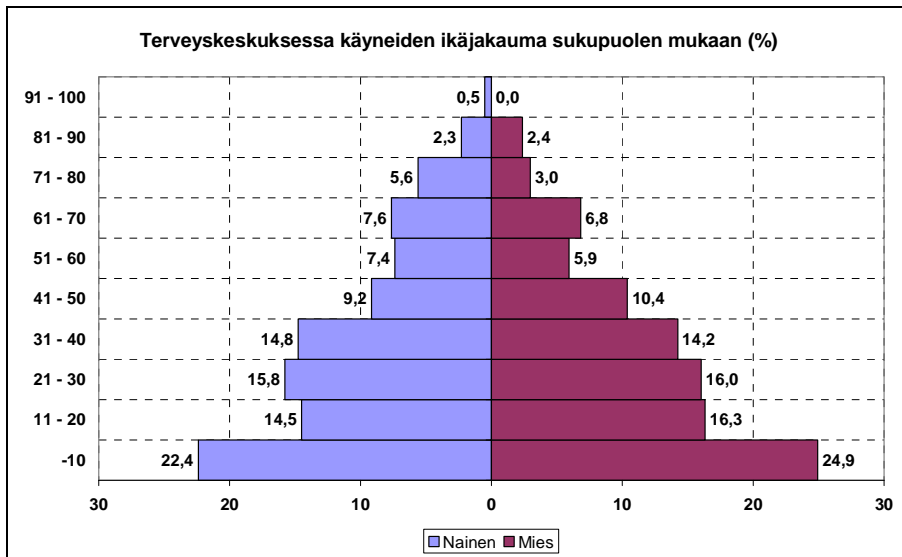
Prosenttien esitysmuodosta on huomattava, että se kertoo luvun sadalla ja lisää prosentimerkin luvun perään.

Luku:	Luku %-formaattilla:
0,4	40 %

Käyttäjä voi määritellä myös omia muotoiluja Custom-vaihtoehdon (Oma) avulla. Eräissä tilanteissa on kätevä määritellä kuvion vaakasuoran akselin negatiivinen puolisko siten, että sen luvut näkyvät positiivisina. Tämä tehdään klikkaamalla Custom-vaihtoehtoa ja täyttämällä Type-kenttä viereisen esimerkin mukaisesti. Siinä ensimmäinen nolla tarkoittaa, että *positiiviset luvut* esitetään positiivisina kokonaislukuina ja toinen nolla että *negatiiviset luvut* esitetään positiivisina kokonaislukuina. Tämän muotoilun avulla saadaan luotua seuraava ikäpyramidi.



Kuvio10.4.1. Terveyskeskuksessa käyneet ikäryhmittäin sukupuolen mukaan



10.5. Lajittelu

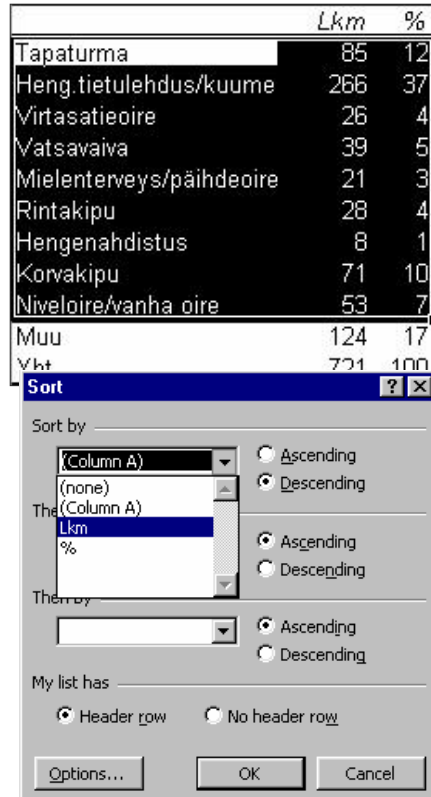
Lajittelua eli ”sorttausta” tarvitaan tilasto-analyseissa esim. silloin, kun suoran jakauman luokat halutaan lukumäärän mukaiseen järjestykseen. Viereisessä taulukossa on terveystieteiden tutkimuksen syy suora jakauma. Sen luokat halutaan lajitella niin, että ensimmäiseksi tulee yleisin syy ja viimeiseksi harvinaisin syy. Luokka ”Muu” halutaan kuitenkin pitää kaikkein viimeisimpänä luokkana.

Merkitse (”maalaa”) se osa taulukosta, jonka haluat lajitella. On huomattava, että kaikki taulukon sarakkeet on otettava merkinnän piiriin. Se solu, josta alueen merkin- tä aloitettiin, pysyy valkoisena.

Anna komento **Data/Sort**, jolloin näytölle ilmestyy viereinen valintaikkuna.

Valitse ensinnäkin vaihtoehto *My List Has Header Row*, jos se ei ole valittu. Vaihtoehto tarkoittaa, että taulukossa on otsikkorivi (Lkm, %). Ikkunassa on kolme lajitte- luavainta, jotka on rajattu suorakaiteella.

Napsauta ensimmäisen lajitte- luavaimen kenttä auki ja merkitse lajitte- luavaimiksi Lkm tai %. Lisäksi valitse joko *nouseva* (Ascending) tai *laskeva* (Descending) lajittelu. Paina lopuksi OK:ta ja tulos on viereisen taulukon mukainen.



	Lkm	%
Heng.tietul	266	37
Tapaturma	85	12
Korvakipu	71	10
Niveloire/va	53	7
Vatsavaiva	39	5
Rintakipu	28	4
Virtasatieoi	26	4
Mielenterve	21	3
Hengenahc	8	1
Muu	124	17
Yht	721	100

10.6. Työarkin sivutus

Tilastoajon tulokset saattavat ulottua useille sadoille ehkä tuhansillekin työarkin riveille. Kun tulokset liitetään esimerkiksi raportin liitteeksi, on tarpeen että sivutus ei jaa taulukoita eri sivuille. Excelissä on *View*-valikossa *Page Break Preview*-toiminto, jolla voidaan helposti asettaa sivunvaihdot haluttuihin paikkoihin. Kun toiminto on käytössä, sivunvaihdot näkyvät leveinä katkoviivoina. Jos sivutus on huonossa paikassa, niin katkoviivaa voi vetää hiirellä ylöspäin sopivaan paikkaan. Pakotettua sivunvaihtoa osoittava viiva muuttuu nyt yhtenäiseksi. Seuraavassa taulukossa on tilanne ennen ja jälkeen editoinnin.

Taulukko 10.6.1. Sivutuksen muuttaminen *Page Break Preview*-tilassa.

	A	B	C	D	E
46	Täysin eri mieltä	43	7	43	7
47	Jaksoenkin eri mieltä	73	12	116	19
48	Ei samaa eikä eri mieltä	140	23	256	43
49	Jaksoenkin samaa mieltä	168	28	424	70
50	Täysin samaa mieltä	178	30	602	100
51	Yht.	602	100	602	100
52					
53					
54	26 Suunnitnan ravintalassa paikkaan, missä vain alla tap. kerkkipirt...				
55		<i>lkm</i>	<i>%</i>	<i>Asum. Nro</i>	<i>Asum.-%</i>
56	Täysin eri mieltä	99	17	99	17
57	Jaksoenkin eri mieltä	190	32	289	48
58	Ei samaa eikä eri mieltä	132	22	421	70
59	Jaksoenkin samaa mieltä	126	21	547	91
60	Täysin samaa mieltä	52	9	599	100
61	Yht.	599	100	599	100

	A	B	C	D	E
46	Täysin eri mieltä	43	7	43	7
47	Jaksoenkin eri mieltä	73	12	116	19
48	Ei samaa eikä eri mieltä	140	23	256	43
49	Jaksoenkin samaa mieltä	168	28	424	70
50	Täysin samaa mieltä	178	30	602	100
51	Yht.	602	100	602	100
52					
53					
54	26 Suunnitnan ravintalassa paikkaan, missä vain alla tap. kerkkipirt...				
55		<i>lkm</i>	<i>%</i>	<i>Asum. Nro</i>	<i>Asum.-%</i>
56	Täysin eri mieltä	99	17	99	17
57	Jaksoenkin eri mieltä	190	32	289	48
58	Ei samaa eikä eri mieltä	132	22	421	70
59	Jaksoenkin samaa mieltä	126	21	547	91
60	Täysin samaa mieltä	52	9	599	100
61	Yht.	599	100	599	100

Jos taulukoita on paljon, on tällainen manuaalisesti tehtävä sivutus työläs. Tixelissä onkin työarkille automaattinen sivutus. Se saadaan käyttöön valitsemalla *Tixel muotoilut*-valikosta *Sivutus*-vaihtoehto. Työarkki sivutetaan siten, että sivunvaihto ei tule taulukon keskelle. Jos työarkin fonttikokoa muutetaan, ei sivutus ole enää optimaalinen. Tämä Tixelin kommento voidaan kuitenkin antaa uudelleen, jolloin se poistaa vanhan sivutuksen ja tekee uuden optimaalisen sivutuksen kyseiselle fonttikoolle.

11. Excel-grafiikka

Tixel-ohjelma muodostaa automaattisesti tilastoajon perusteella tulosta havainnollistavan kuvion. Joskus on kuitenkin tarve kuvata graafisesti esim. jonkun julkaisun taulukkoa. Tämä tehdään Excelin grafiikka-kehittimellä ilman Tixel-ohjelman apua.

11.1. Kuvion luonti

Kuvion laadinta aloitetaan siten, että taulukon luvut, mahdolliset luokkien tunnukset ja lukusarjojen selitteet syötetään työarkille. Alla olevassa esimerkissä on poliklinikan ilmoittautumisajan jakauma naisten ja miesten osalta. Taulukosta voidaan erottaa seuraavat osat:

- **luokkatunnukset** (categories labels): aikaluokat 0-6, 7-12 jne,
- **lukusarjat** (data series), joita on kaksi: naisten ja miesten frekvenssit,
- lukusarjojen **selitteet** (legend): nainen ja mies.

	A	B	C	D	E	F
1		Naiset	Miehet	←	Selitteet	
2	0-6	4	14			
3	7-12	148	133			
4	13-18	162	109			
5	19-21	51	60			
6	22-24	23	17			
7	↑		↑			
8						
9	Luokka-	Lokusarjat				
10	tunnukset					
11						

Huom1. Taulukko voidaan tulkita myös vaakasuorassa suunnassa eli luokkatunnukset ovat yhdellä rivillä ja lukusarjat vastaavasti rivisuunnassa.

Huom2. xy-tyyppisessä kuviossa ei ole lainkaan luokkatunnuksia.

Kuvion luonti aloitetaan merkitsemällä(”maalaamalla”) taulukko mukaan lukien selitteet ja luokkatunnukset. Esimerkin tapauksessa taulukko on A1:C6.

Excelin on tulkittava merkitty taulukko kuvion pohjaksi joko pysty- tai vaakasuorassa suunnassa. Pystysuora suunta tarkoittaa, että lukusarjat ovat pystysuorassa suunnassa ja vaakasuora suunta, että lukusarjat ovat vaakasuorassa suunnassa. Excel tekee automaattisesti valinnan näiden kahden vaihtoehdon välillä taulukon ulottuvuuksien mukaan. Taulukko tulkitaan siten, että luokkia on enemmän kuin lukusarjoja.

Taulukon tulkintaa voidaan kuvan luonnin yhteydessä muuttaa eli lukusarjat voidaan tulkita pysty- tai vaakasuorassa suunnassa riippumatta siitä, miten Excel on alkuaan asian tulkinnut.

Epäyhtenäiset alueet

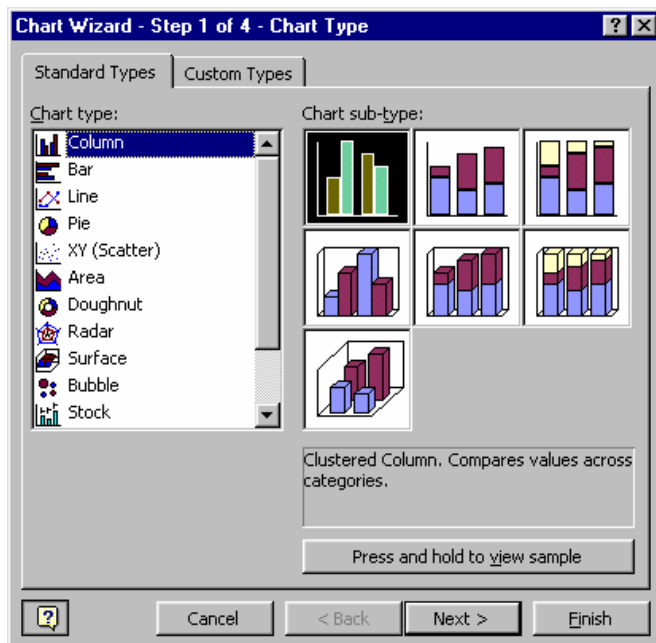
Epäyhtenäiset alueet määrätään hiirellä siten, että uusia osa-alueita merkittäessä pidetään Control-näppäintä (Ctrl) alhaalla.

Kun taulukko on merkitty, valitaan valikosta **Insert/Chart**, jolloin näytölle ilmestyy ensimmäinen vaihe kuvion luontiprosessista.

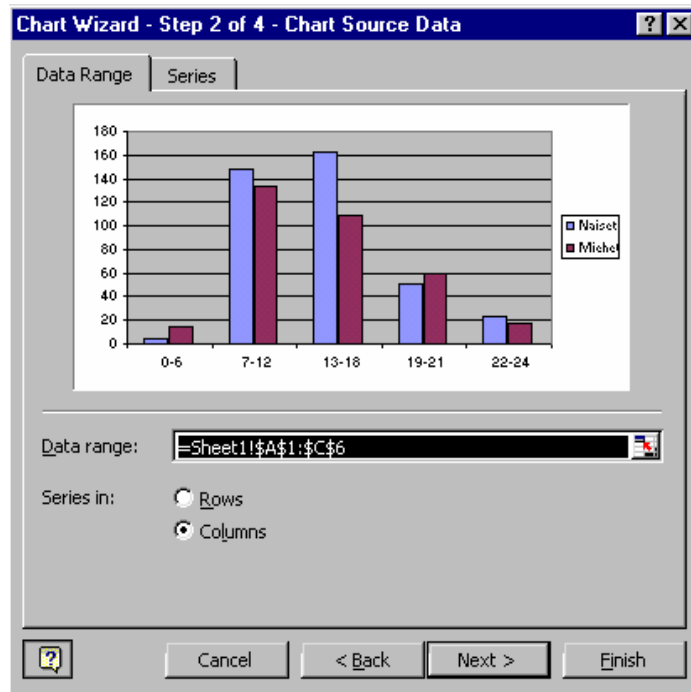
Luontiprosessi voidaan käynnistää myös napsauttamalla valikon grafiikkavelho-painiketta (ChartWizard).



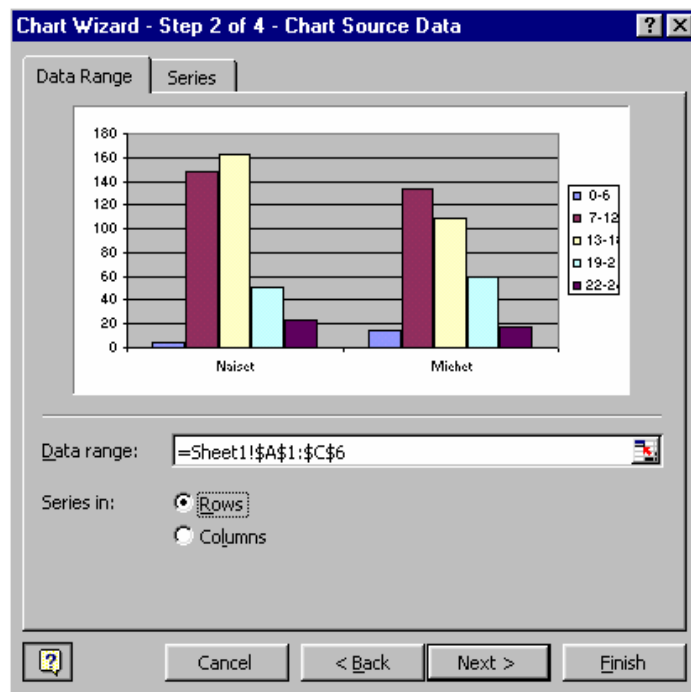
Luontiprosessin ensimmäisessä vaiheessa valitaan sopiva kuviotyyppi *Chart type* -valikosta. Oikealla olevilla kuviovaihtoehtoilla valintaa voidaan täsmentää halutun kaltaiseksi. Kun valinnat on suoritettu, jatketaan kuvion luontia valitsemalla *next*-vaihtoehto. Liite 2:ssa on esitetty tärkeimmät kuvatyypit.



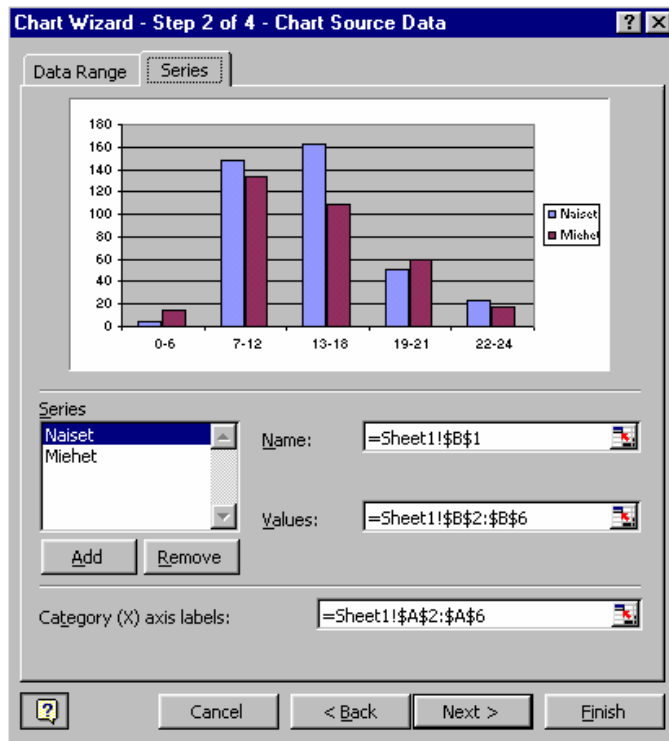
Vaiheessa kaksi näkyy alue (Data Range), joka on merkitty kuvaa varten. Luku- eli arvosarjat (Data Series) voivat olla joko riveinä (Rows) tai sarakkeina (Columns).



Viereisessä kuvassa on valittu lukusarjojen suunnaksi rivisuunta. Huomaa kuvien ero.

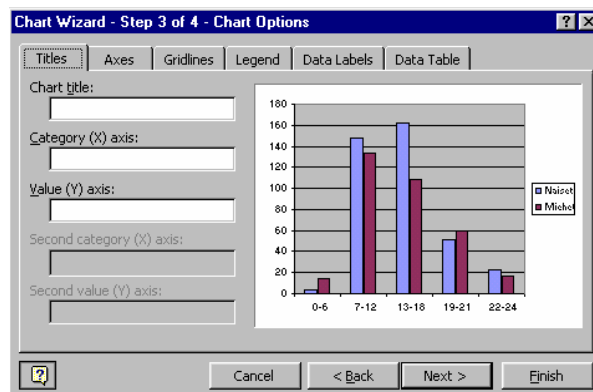


Valitsemalla lukusarja-arkin (Series) voit vielä tarkentaa lukusarja-aluetta. Yksittäisen lukusarjan voi lisätä tai poistaa. Lisäksi kirjoittamalla kaavan arvojen määrityskohtaan voit muodostaa uuden arvosarjan. Uudet lukusarjat voi nimetä ja vanhoja nimiä voi muuttaa.



Kolmannessa vaiheessa voit muotoilla kuvion asetuksia kuuden eri arkin avulla. Muutokset päivittyvät kuvioikkunaan välittömästi.

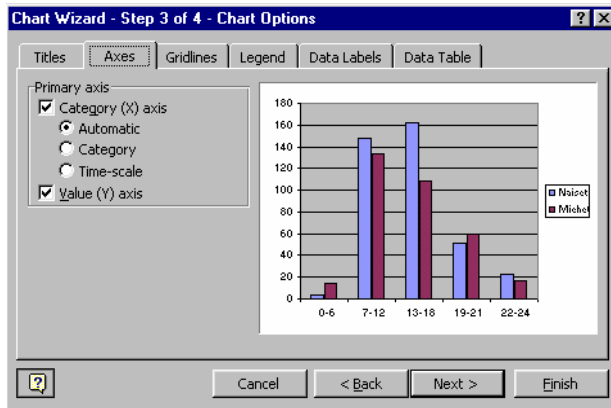
1) Otsikkoarkkiin (Titles) voit kirjoittaa kuvion otsikon (Chart title), luokka-akselin otsikon (Category X axis) sekä arvoakselin (Value Y axis) otsikon. Joissain kuvatyypeissä on kaksi x-akselia tai y-



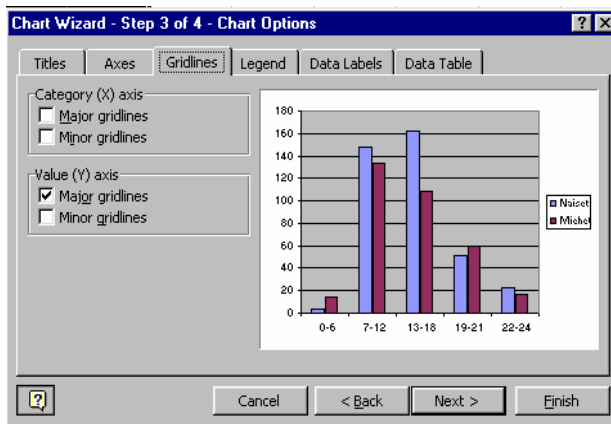
Johdatus tilastolliseen data-analyysiin

akselia. Voit nimetä myös nämä akselit.

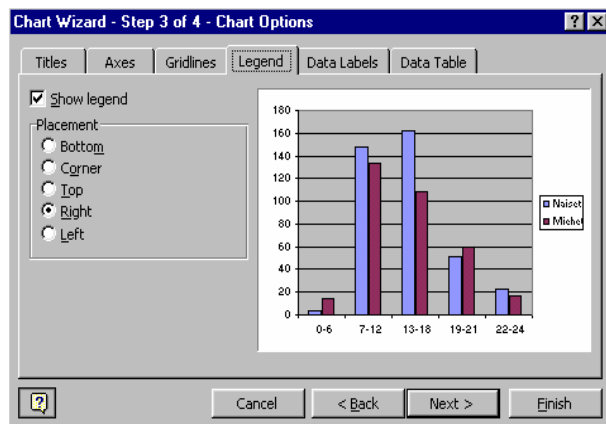
2) Akseliarkin (Axes) avulla voit määrittellä näkyvät akselien asteikot kuviossa. X-akselin asteikon muodoille on kolme vaihtoehtoa. Ensimmäisenä on Excelin oletusarvoinen vaihtoehto (Automatic). Jos luokkatunnukset ovat lukuja, on mahdollista käyttää näitä lukuja akselin asteikkona. Tällöin sopiva vaihtoehto on *Category*. Viimeisenä vaihtoehtona on aikasarja (Time Series). Jos valitset tämän vaihtoehdon, akselin asteikko näkyy päivämäärämuotoisena.



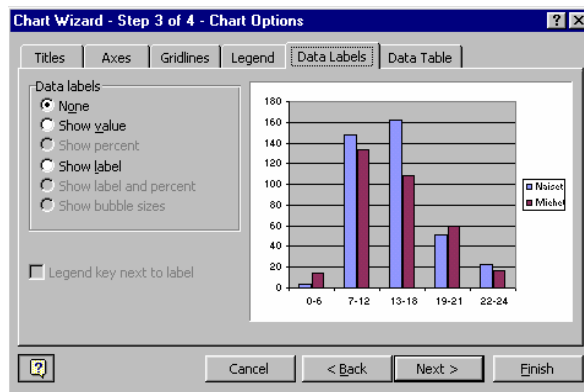
3) Voit lisätä kuvaan vaaka- tai pystysuorat taustaviivat rastittamalla valintasi taustarudukkoarkille (Gridlines).



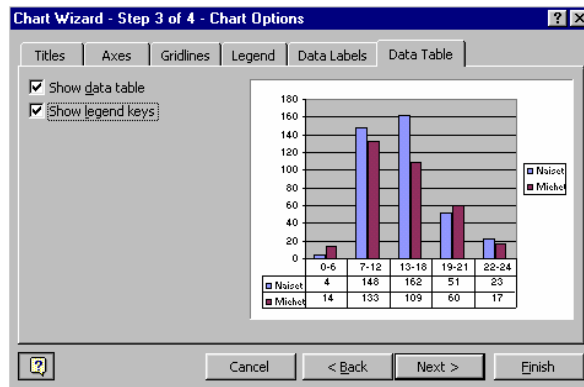
4) Selitearkilla (Legend) voit määrittellä haluatko kuvioon selitelaatikon. Vaihtoehtoisia selitelaatikon sijainteja kuviossa on useita. Selitelaatikon sijaintia voi myöhemmin kuvan muokkauksen yhteydessä määrittellä tarkemmin.



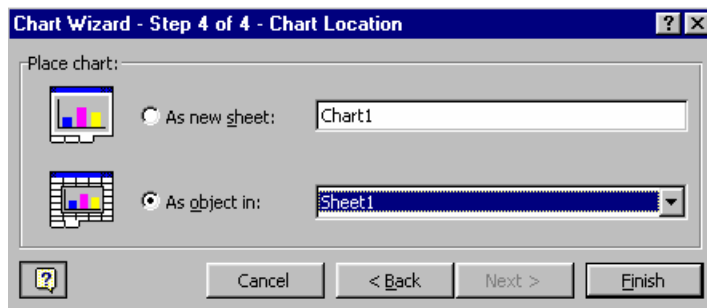
5) Arvopisteiden otsikot -arkilla (Data Labels) tuoda kuvioon lukuja, esim. pylväiden yläpuolelle voi sijoittaa pylväiden "tarkan" korkeuden sekä pylvään otsikkotekstin.



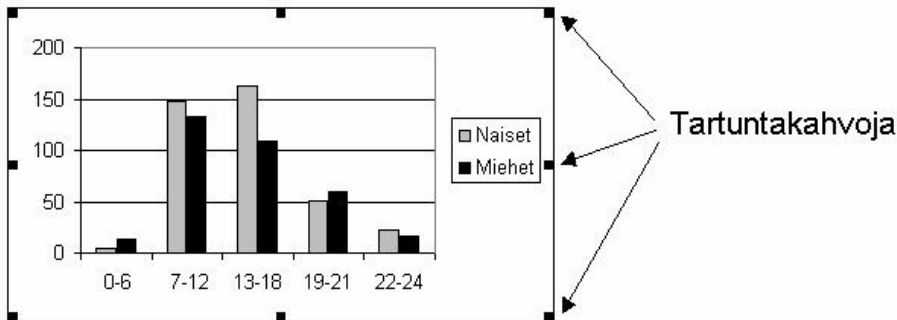
6) Viimeisenä arkkina on arvotaulukkoarkki (Data Table). Arkin avulla voit lisätä kuvion alapuolelle merkittynä alueena olleen taulukon. Rastittamalla *Show legend keys* -vaihtoehdon tuot pylväiden värikoodit näkyviin.



Viimeisessä kuvanluontiprosessin vaiheessa määritellään mihin kuvio sijoitetaan. Vaihtoehdoista ensimmäinen (As new sheet) tekee kuvasta uuden Excel-arkin. Jälkimmäinen vaihtoehto (As object in) sijoittaa kuvan halutulle arkille.



Työarkille sijoitetun kuvion kokoa voit muuttaa napsauttamalla ensin kuvaa ja vetämällä sitten tartuntakahvoista kuva sopivan kokoiseksi. Kuviossa on kahdeksan tartuntakahvaa.



Voit myös siirtää kuviota työarkilla tarttumalla hiirellä kuvan reunasta (ei tartuntakahvasta) kiinni ja vetämällä kuva sopivaan paikkaan.

11.2. Kuviotyypin valinta

1. Aikasarjojen kuvaamiseen käytetään

- a) *pystyylväitä tai aluekuviota*, kun aikasarjoja on vain yksi,
- b) *viivakuviota* kun aikasarjoja on useampia

Poikkeus: *Päällekkäisiä pylväitä* käytetään kuvaamaan usean aikasarjan summaa.

2. Jakauma-tyyppisen tiedon esittämisessä (esim. markkinaosuudet, yrityksen osakkeiden jakautuminen) käytetään

- a) *ympyräkuvaa*, kun luokkia on suhteellisen vähän ($\leq 6-7$) ja kuvataan %-jakaumaa,
- b) *vaakapalkkeja tai pystyylväitä*, kun luokkia on enemmän ($> 6-7$)

Vaakapalkit ovat käyttökelpoisia varsinkin, kun luokkien nimet ovat pitkiä.

Jos luokkien välillä ei ole luonnollista järjestystä, kannattaa luokat asettaa lukumäärän suuruuden mukaiseen järjestykseen.

3. Kahden laadullisen muuttujan ristiintaulukointi:

Sataan summaavat prosenttijakaumat kuvataan peräkkäisillä vaakapalkeilla tai vierekkäisillä pystyylväillä. Peräkkäisten palkkien avulla voidaan verrata myös summajakaumia.

4. Kolmeulotteiset kuvat

Jos käytät 3-ulotteisia kuvia huomioarvon lisäämiseksi, laita kuvion viereen taulukko, josta selviää tarkat arvot. Niitä voi olla vaikea lukea kuviosta.

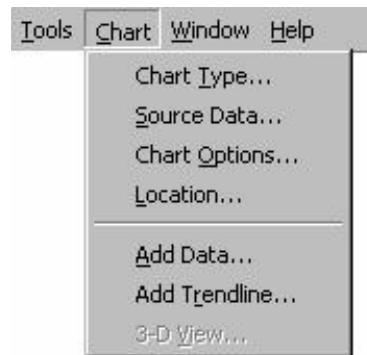
11.3. Kuvion muokkaaminen

Voit muokata kuviota samalla tavalla riippumatta siitä, onko kuvio työarkilla vai omana grafiikka-arkkina. Jos kuvio on työarkilla, napsauta kuviota, jolloin tar-
tuntakahvat ilmestyvät kuvion ympärille. Perusvalikon (File, Edit jne) nimet ovat samat kuin työarkin tapauksessa, mutta niiden alavalikkojen nimet ja sisällöt ovat muuttuneet.

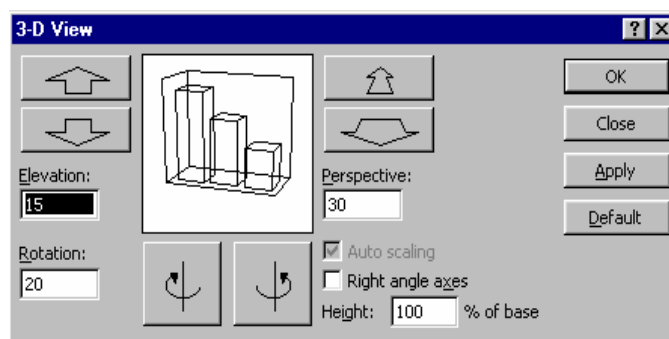
Jos kuvio on omana grafiikkaikkunana, voidaan sitä muokata suoraan. Kuvion muokkaamisessa käytetään *Chart-* ja *Format-* valikkoja.

Chart-valikko

Chart-valikolla avulla voit palata kuvanluontiprosessin neljään eri vaiheeseen (ks. 11.1 Kuvion luonti). Alasvetovalikon ensimmäinen vaihtoehto *Chart type* vastaa kuvanluontiprosessin ensimmäistä vaihetta, *Source Data* toista jne. *Chart Options* -vaihtoehdon valitsemalla voit lisätä mm. otsikoita, lukusarjojen arvoja ja taustaruudukon kuvaan.



Alasvetovalikon viimeisimmät vaihtoehdot ovat tietojen lisäys (Add Data), käyrän lisääminen pisteparveen (Add Trendline) ja kolmiulotteisen näkymän muokkaus (3-D View). Viereinen valintaikkuna saadaan, kun muunnetaan kolmiulotteista pylväskuviota.

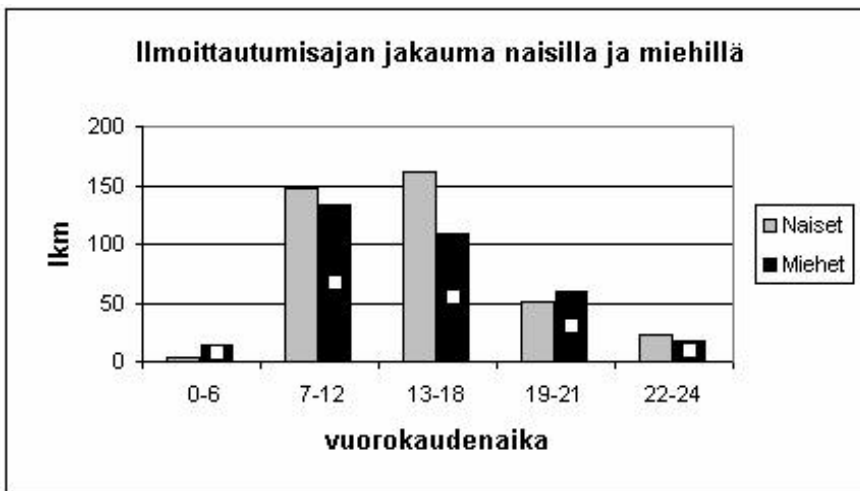


Huom. Yleensä kaksiulotteiset kuvat ovat kolmiulotteisia kuvia selkeämpiä. Epäselkeyden lisäksi kolmiulotteiset kuvat saattavat "valeyhdellä".

Format-valikko

Format-valikon avulla voidaan määrätä kuvaobjektien ominaisuudet (esim. väri, koko, sijainti jne). Ennen kuin *Format*-valikkoa voidaan käyttää, on merkittävä kuvan osa, jota halutaan muokata. Muokattavat osat voivat olla esim. yhden lukusarjan pylväät, yksittäinen pylväs, akselit niihin liittyvine teksteineen otsikot ja selitteet tai koko kuva.

Kuvan osan merkitseminen tehdään hiirellä napsauttamalla. On oltava tarkkana, että hiiren kohdistin osoittaa juuri sitä kuvan osaa, joka halutaan merkitä. Vähin-



tään kaksi mustaa (tai valkoista) neliötä osoittaa, mikä kuvan osa on tullut

merkityksi. Oheisessa kuvassa on merkitty toisen lukusarjan pylväät.

Merkityn kuvaelementin nimi on vasemman ylänurkan kentässä. Esimerkin tapauksessa siinä lukee *Series "Miehet"*.

Huom. Yksittäinen pylväs merkitään napsauttamalla hiirellä kyseistä pylvästä kaksi kertaa peräkkäin. Ensimmäisellä napsautuksella aktivoituu koko pylvässarja ja jälkimmäisellä napsautuksella haluttu pylväs.

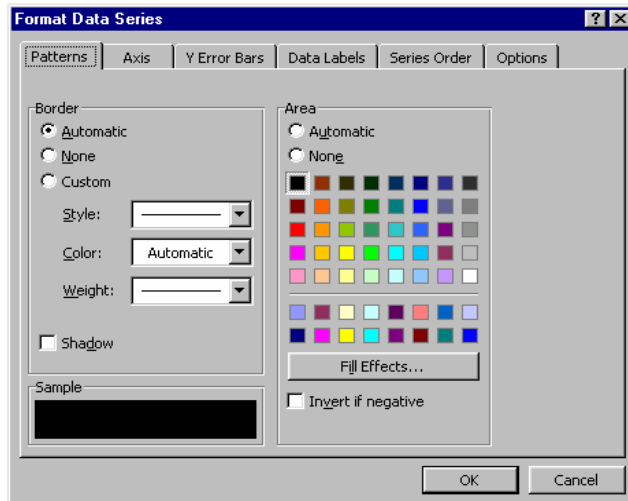
Koko kuvan merkitsemistä tarvitaan esim. jos halutaan kopioida kuva tekstinkäsittelyohjelmaan (kts. tarkemmin luku 12). Merkitseminen tehdään napsauttamalla hiirellä kuvan reuna-aluetta.

Kun kuvaobjekti on merkitty, valitaan *Format*-valikko. Sen ensimmäinen vaihtoehto kertoo, mikä kuvaobjekti on valittu (esim. *Selected Data Series* tai *Selected Axis*). Kun tämä valitaan, avautuu valitulle kuvaobjektille ominaiset määrittelyarkit. Tarkastellaan lähemmin lukusarjoihin ja akseleihin liittyviä määrittelyarkkeja.

Lukusarjojen muotoilu (Format Data Series)

Tärkeimmät vaihtoehdot ovat *Patterns*, jolla voidaan määrätä pylväiden reunaviivojen ominaisuudet (Border) ja pylväiden pintojen ominaisuudet (Area).

Border-vaihtoehdolla voidaan määrätä merkityn elementin reunan tyyli (Style), väri (Color) ja leveys (Weight). *Area*-vaihtoehdolla määrätään objektin väri ja rasterointi (Fill Effects).

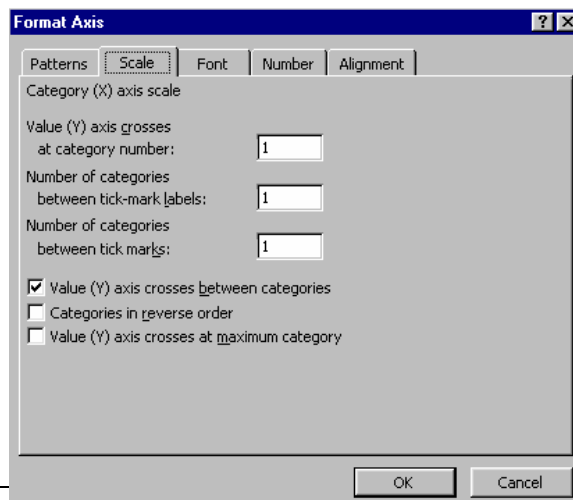


Series Order -arkin avulla voidaan vaihtaa pylväiden järjestystä; esimerkkitapauksessa siis miesten pylväät voidaan siirtää naisten pylväiden vasemmalle puolelle.

Options-arkilla on mahdollista leventää pylväitä suhteessa pylväiden väliin jättämään tyhjään tilaan. *Gap Width* -luvun arvo on pylväiden välin leveys ilmaistuna prosentteina yhden pylvään leveydestä. Excel käyttää oletusarvona 150 prosenttia, mikä on liian suuri arvo etenkin, jos kuviossa on useita pylväitä.

Akselien formatointi (Format Selected Axis)

Merkitse akseli napsauttamalla akselia tai akseliin liittyviä asteikkotekstejä. Akselin päihin ilmestyy kaksi mustaa neliötä. Valitse **Format/Selected Axis**, joka tuo esille seuraavan määrittelyikkunan. Siinä on viisi määrittelyarkkia: *Patterns*, *Scale*, *Font*, *Number* ja *Alignment*.



Käsitellään tarkemmin *Scale*-arkkia. Sen sisältö riippuu siitä, onko merkitty akseli luokitusakseli (X) tai arvoakseli (Y). Seuraavassa on luokitusakselin määrittelykortti.

Number of Categories Between Tick-Mark Labels tarkoittaa, millä poimintavälillä luokkatunnuksia otetaan mukaan kuvaan. Jos arvo on yksi, otetaan mukaan kaikki luokkatunnukset. Jos arvo on kaksi, otetaan joka toinen jne.

Number of Categories Between Tick Marks tarkoittaa samaa kuin edellinen mutta nyt sovellettuna pieniin luokkaviivoihin.

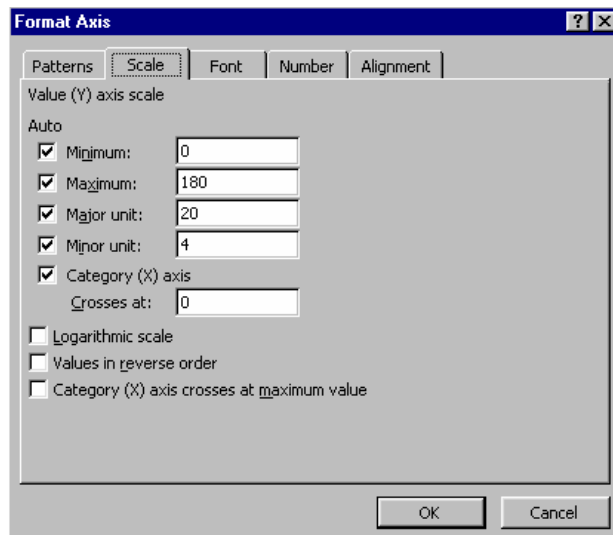
Categories in Reverse Order tarkoittaa, että luokkatunnukset ovat käänteisessä järjestyksessä. Tällä saadaan *Bar*- eli vaakapalkki-tyyppisessä kuvassa luokkien järjestys samaksi kuin mitä se on työarkin taulukossa.

Seuraavassa on arvoakselin määrittelyarkki.

Auto-kohdassa oleva rasti tarkoittaa, että kyseessä on Excelin oma valinta. Jos haluat muuttaa näitä määrittelyjä, syötä kenttiin uudet arvot.

Minimum- ja *Maximum*-kohdat tarkoittavat y-akselin alku- ja loppupäätä. Näiden avulla pystyt suuresti vaikuttamaan esim. pylväskuvan muotoon.

Major Unit ja *Minor Unit* tarkoittavat suurempaa ja pienempää asteikkoväliä. Nämä määräävät paitsi poikkiviivojen paikat myös viivoitusten sijainnin (*Gridlines*). Viivoitusten mukaanotto tehdään kuitenkin **Chart/Chart Options**-valikosta.



Patterns

Voit määrittellä akselin paksuuden, värin ja tyylin. Tällä määrittelyarkilla on myös akselin väliviivoihin liittyviä määrittelyjä.

Font

Font-vaihtoehdolla määritellään tekstin koko ja tyyliominaisuudet samaan tapaan kuin työarkilla.

Alignment

Koordinaattiakseleihin liittyvien tekstien suunta määrätään asteina vaakasuoraan suuntaan nähden. Toinen vaihtoehto on tekstin sijoittaminen niin, että kirjaimet ovat normaalissa asennossa, mutta allekkain. Jos määrittelet tekstin suunnan vaakasuoraksi, saattaa käydä niin, että tunnukset sijoittuvat vaaka-akselilla osittain päällekkäin. Tällöin *Scale*-vaihtoehdon avulla on jätettävä joka *k*. tunnus pois tai käännettävä kuva vaakapalkkikuvioksi.

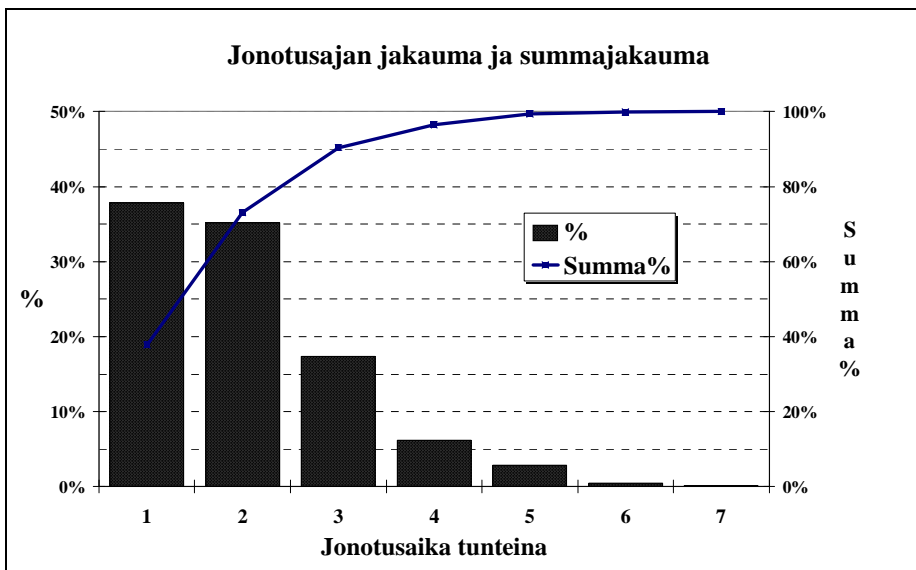
Tekstin vapaamuotoinen svöttö

Otsikkotekstin ja akseleiden tekstien lisäksi kuvaan voidaan vapaamuotoisesti sijoittaa tekstiä. Merkitse ensin kuvio. Kirjoita teksti ja paina enteriä. Teksti tulee nyt keskelle kuvaa mustien merkkiniöiden ympäröimänä. Tarttumalla hiirellä merkkiniöiden väliseen viivaan voit siirtää tekstialueen mihin tahansa. Tekstiä voit myös muotoilla **Format/Text Box/Font**-komennolla.

11.4. Yhdistelmäkuvio

Yhdistelmäkuviossa on yhdistetty kaksi eri kuvatyyppeä, esim. pylväs- ja viiva-kuvio. Yhdistelmäkuva on mahdollista tehdä, jos lukusarjoja on vähintään kaksi. Erilaisia yhdistelmäkuvaavaihtoehtoja on valittavana ensimmäisessä kuvanluontiprosessin vaiheessa Mukautetut kaaviot -arkilla (Custom Types). Myös useammasta lukusarjasta muodostettu kuva, esimerkiksi pylväskuva, voidaan muuttaa yhdistelmäkuvioksi. Tämä tehdään merkitsemällä ensin muutettavan lukusarjan kuvaobjektit (esim. pylväät) ja valitsemalla sen jälkeen **Chart/ Chart Type**. Valintaikkunalla voit määrätä lukusarjalle uuden kuvatyypin, esimerkiksi viivan.

Seuraavassa kuviossa on poliklinikan jonotusajan jakauman prosentit ja kumulatiiviset prosentit.



11.5. Kuvion muotoilujen kopiointi

Kuvioiden muotoilut voidaan kopioida toiseen kuvioon seuraavasti:

1. Merkitse se kuvio, jonka muotoiluja olet kopioimassa. Merkitseminen tehdään klikkaamalla kuvion reuna-aluetta, jolloin kuvion kulmiin ilmestyy pienet musta neliöt ja näytön vasemmassa ylänurkassa lukee *Chart Area*. Anna komento *Edit/Copy*.
2. Siirry siihen kuvioon, johon haluat muotoilut kopioitavan ja anna komento *Edit/Paste Special*, josta valitse *Formats-vaihtoehto*.

Tässä menettelyssä on vain yksi ongelma: ensimmäisen kuvion otsikko kopioituu myös toiseen kuvioon. Jos tätä ei haluta, voidaan käyttää *Tixel-muotoilut*-valikossa olevaa *Kuvien muotoilu*-vaihtoehtoa. Se kopioi työkirjan **ensimmäisen** kuvion muotoilut työkirjan kaikkiin muihin kuvioihin siten, että kuvioiden alkuperäiset otsikot säilyvät.

11.6. Kuvioiden siirto PowerPoint-ohjelmaan

PowerPointissa on paremmat kuvien esitysominaisuudet (SlideShow) kuin Excelissä. Siksi saattaa tulla tarve siirtää Excelissä tehdyt kuviot PowerPoint-

ohjelmaan. Muutaman kuvion siirtäminen on helppo tehdä ”käsini”, mutta jos kuvioita on useita kymmeniä, kannattaa käyttää Tixelissä olevaa siirto-ohjelmaa.

Aktivoi Excelin työkirja, jossa kuviot ovat. Valitse *Tixel-muotoilut* valikosta *Kuvien siirto PowerPointiin*-vaihtoehto. Tixel avaa PowerPointin, jos se ei ole jo avattu ja siirtää työkirjan kaikki kuviot PowerPointiin.

12. Word

Seuraavassa oletetaan, että lukijalla on perustiedot tekstinkäsittelystä Windows-ympäristössä. Tässä luvussa käsitellään eräitä tilastoraportoinnissa tarvittavia Word-tekstinkäsittelyohjelman ominaisuuksia. Käsiteltävät asiat ovat: tiedon (esim. taulukoiden ja kuvien) siirto sovellusohjelmasta toiseen, taulukoiden luonti ja muokkaaminen, kehyksen käyttö, rajauksen käyttö sekä sisällysluettelon lisääminen asiakirjaan.

12.1. Tiedon siirtäminen sovellusohjelmasta toiseen

Windows-ympäristössä on hyvin helppo siirtää tietoa ohjelmasta toiseen. Esimerkiksi Excel-*taulukkolaskentaohjelmalla* tai Tixel-*tilasto-ohjelmalla* tehtyjä taulukoita ja kuvia voidaan siirtää Wordiin.

Avattujen ohjelmien nimet näkyvät näytön alareunassa. Kun haluat siirtää esim. taulukon tai kuvan Excelistä (tai Tixelistä) Wordiin, toimi seuraavasti:

1. Siirry tietoa lähettävään sovellusohjelmaan (esim. Excel) ja avaa tiedosto, josta tietoa siirretään.
2. Muokkaa taulukko tai kuva valmiiksi Excelissä ennen Wordiin siirtämistä.
3. Merkitse taulukko tai kuva, jonka haluat siirtää ja anna Edit/Copy-komento.
4. Siirry Wordiin tai jos Wordia ei ole jo aikaisemmin avattu, avaa se.
5. Siirrä kohdistin paikkaan, johon haluat siirrettävän tiedon (taulukko tai kuva).
6. Anna komento **Edit/Paste tai Paste Special**. Jälkimmäinen komento antaa useita vaihtoehtoja siirrettävän tiedon muotoiluun.

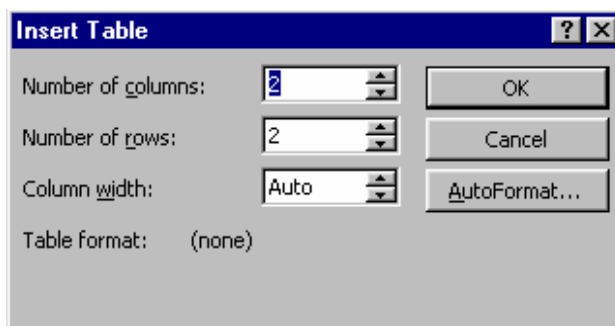
Huom. Jos olet jo avannut jonkin ohjelman, älä avaa sitä toista kertaa ohjelma-kuvakkeesta, vaan siirry ohjelmasta toiseen alapalkin avulla.

Objektina liitetyn taulukon tai kuvan eräitä ominaisuuksia voi Wordissa edelleen muuttaa. Kun haluat muuttaa asetuksia, merkitse ensin objekti anna komento **Format/Object**.

12.2. Taulukot

Taulukoiden esittämiseen on käytössä kätevä taulukkokehitin, jolla taulukon sarakkeet pysyvät "ojennuksessa". Taulukon solujen ympärillä on mustat reunaviivat, joiden tulostusasua voidaan muuttaa. Taulukon soluihin voi sijoittaa tekstiä, grafiikkaa tai muita objekteja. Siirtyminen taulukon sisällä solusta toiseen tehdään esim. nuolinäppäimillä tai hiirellä.

Uusi taulukko: Siirrä kohdistin kohtaan, johon haluat taulukon sijoittuvan. Valitse **Table/Insert Table**, jolloin näytölle tulee seuraavanlainen ikkuna.



Määrää sarakkeiden (columns) ja rivien (rows) lukumäärät.

Ylöspäin (alaspäin) olevat kolmiot lisäävät (vähentävät) lukumäärää.

Uusi rivi taulukon viimeiseksi riviksi: Paina sarkain-näppäintä, kun kohdistin on taulukon viimeisessä solussa.

Uusi rivi taulukkoon: Merkitse rivi, jonka yläpuolelle uuden rivin haluat. Valitse **Table/Insert Rows**.

Uusi sarake: Merkitse sarake, jonka vasemmalle puolelle haluat uuden sarakkeen. Valitse **Table/Insert Columns**.

Rivin tai sarakkeen poistaminen: Merkitse poistettava rivi tai sarake, ja valitse **Table/Delete Rows** tai **Table/Delete Columns**.

Solujen sisällön tyhjentäminen: Merkitse solut ja paina Delete-näppäintä.

Taulukon poistaminen: Merkitse taulukko **Table/Select Table**-komennolla ja valitse **Edit/Cut**.

Rivin korkeus: Solun sisältö (esimerkiksi kuvan koko tai tekstin pituus) määrittää yleensä rivin korkeuden. Samalla rivillä olevat solut ovat samankorkuisia. Eri rivit voivat olla eri korkuisia.

Sarakkeen leveys: Vedä hiirellä sarakkeen reunaviivoista.

12.3. Kehys

Kehyksen käyttö on vaihtoehtoinen tapa asetella tekstiä kuvan tai taulukon viereen. Määrittele sivunäyttö kehyksiä varten komennolla **View/Page Layout**.

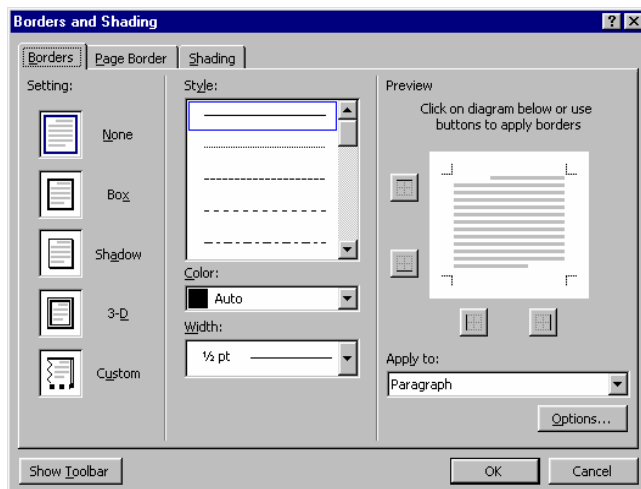
Tällöin kehyksen todellinen sijainti sivulla näkyy ja kehyksen kokoa sekä sijaintia voi hiirellä muuttaa.

Luo kehys antamalla komento **Insert/Text Box**. Rajaa ristinmuotoisella kohdistimella kehyksen alue. Kehykseen voit sijoittaa kuvan tai taulukon. Kehyksellä on samoja ominaisuuksia kuin kuva- ja taulukko-objekteilla. Kun haluat muuttaa kehyksen ominaisuuksia, anna komento **Format/Text Box**, jolloin saat samanlaisen valintaikkunan kuin objektia muokatessasi.

12.4. Rajaus

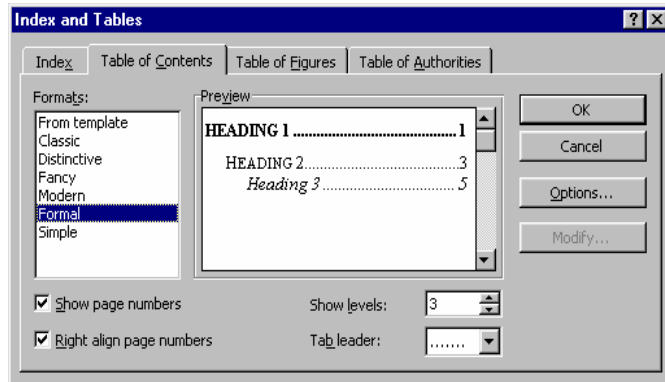
Taulukko, taulukon solu, teksti, kuva tai kehys voidaan rajata. Merkitse reunustettava alue ja anna komento **Format/Border/Setting/Box**. *Style*-valikolla voit määrätä viivan tyylin ja *Width*-valikolla viivan paksuuden. *Shading*-arkilla voit rasteroida merkityn alueen.

Jos haluat rajauksen vain esim. taulukon alareunaan, napsauta esikatseluvälirikosta reunaviivamallin alareunaa ja valitse reunaviivatyyppi. Reunaviivan poistettua saat **Setting/None-**komennolla.



12.5. Sisällysluettelon lisääminen asiakirjaan

Merkitse **heading 1**-tyylillä niiden kappaleiden otsikot, jotka haluat sisällysluetteloon, (tai **heading 2**-tyylillä, jos on kyseessä alemman tason otsikko). Vie kohdistin kohtaan, johon haluat sisällysluettelon ja anna komento **Insert/Index and Tables**. Valitse seuraavasta valintaikkunasta sisällysluetteloarkki (Table of Contents) ja siitä mieleisesi sisällysluettelon muoto.



Sisällysluettelon saat päivitettyä, kun ensin aktivoit sisällysluettelon hiirellä napsauttamalla ja sen jälkeen painat F9:ää.

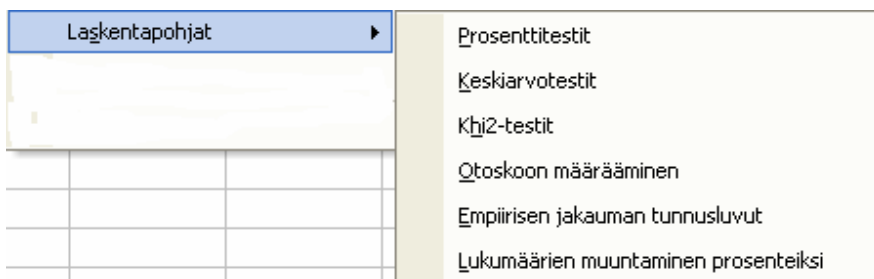
13. Tilastollinen hypoteesin testaus

Tilastollisella testauksella tehdään johtopäätöksiä *perusjoukon* eli populaation tuntemattomista tunnusluvuista eli parametreista, kun on käytettävissä perusjoukosta satunnaisesti poimittu *otos*. Esimerkiksi presidentin vaalien toisen kierroksen kaksi ehdokasta saavat kahden tuhannen hengen kyselyssä 52%:n ja 48%:n kannatukset. Voidaanko tämän perusteella päätellä, että kyselyssä suuremman kannatuksen saaneella ehdokkaalla on enemmistö myös äänioikeutettujen muodostamassa perusjoukossa.

Tixelissä on eri menetelmien tulostaulukoissa tilastolliseen testaukseen liittyviä suureita (testisuureet ja riskitasot). Tällaisia menetelmiä ovat esim. ristiintaulukointi ja ryhmäkohtaiset tunnusluvut. Näissä menetelmissä lähtökohtana on havaintomatriisi, josta Tixel on laskenut tulostaulukon ja tilastollisen testauksen suuret. Joskus esiintyy kuitenkin tilanteita, joissa tunnusluvut on jo laskettu ja halutaan tehdä niille tilastollinen testaus. Tähän tarkoitukseen on Tixelissä *laskentapohjat*, joihin tunnusluvut ja muut tarvittavat tiedot syötetään, jolloin testauksen tulos näkyy laskentapohjalla välittömästi.

Laskentapohjan avulla voidaan myös havainnollistaa, miten muutokset syöttötiedoissa, esim. keskiarvoissa, keskihajonnoissa ja otoskoissa vaikuttavat testin lopputulokseen.

Laskentapohjat saadaan käyttöön, kun Tixelin käynnistämisen jälkeen ensimmäisestä valikosta valitaan vaihtoehto ”Laskentapohjiin perustuvat menetelmät”.



Tässä luvussa käsitellään prosenttilukuihin ja keskiarvoihin liittyviä testejä. Ristiintaulukon tilastollinen testaus on käsitelty alaluvussa 8.4 (Kontingenssikerroin ristiintaulukosta).

Laskentapohjien avulla saadaan myös parametrin *luottamusväli*, johon parametri tietyllä todennäköisyydellä kuuluu. Luottamusvälejä on käsitelty alaluvussa 13.2.

Otoksen suuruus eli otoskoko liittyy läheisesti hypoteesin testaukseen. Usein kysytään, kuinka suuri otos on riittävä tietyn tarkkuuden saavuttamiseksi. Alaluvussa 13.5 on käsitelty otoskoon määräämistä.

Seuraavassa taulukossa on kaikki laskentapohjiin perustuvat menetelmät.

13.1. Hypoteesit

Hypoteesit ovat tuntemattomia parametria koskevia väittämiä. Niitä on kahdenlaisia: *nollahypoteesi* H_0 ja *vaihtoehtoinen hypoteesi* H_1 .

Nollahypoteesi ilmaisee yleensä parametrien yksinkertaisimman asetelman. Edellä olleessa esimerkissä nollahypoteesi on, että ehdokkaiden kannatusprosentit perusjoukossa θ_1 ja θ_2 ovat yhtä suuret eli

$$H_0: \theta_1 = \theta_2 = 50 \text{ \%}.$$

Vaihtoehtoinen hypoteesi H_1 voi olla joko *yksisuuntainen* tai *kaksisuuntainen*. Kaksisuuntaista vaihtoehtoista hypoteesia käytetään, kun parametreista ei ole käytettävissä mitään vankkaa etukäteisinformaatiota. Tässä esimerkissä kaksisuuntainen vaihtoehtoinen hypoteesi on

$$H_1: \theta_1 \neq 50 \text{ \% (jolloin myös } \theta_2 \neq 50 \text{ \%)}$$

Tilastollinen testaus osoittaa, kumpaa hypoteesia päädytään otosinformaation perusteella kannattamaan. Johtopäätös perustuu luonnollisesti otoksesta lasketun ensimmäisen (tai toisen) ehdokkaan kannatusprosentin p poikkeamaan nollahypoteesin ilmoittamasta arvosta 50 %. Mutta kuinka suuri pitää poikkeaman olla, jotta nollahypoteesista voidaan luopua? Vastauksen antaa *testisuureen* arvo ja siihen liittyvä *riskitodennäköisyys* α ., jota yleisesti kutsutaan myös p -arvoksi ja jollaisena se esiintyy myös Tixelin tulosteissa. Tässä esityksessä riskitodennäköisyyteen viitataan kuitenkin α :lla, koska p -kirjainta käytetään kuvaamaan prosenttilukua.

Esimerkin tapauksessa testisuureen z lauseke on

$$z = \sqrt{n} (p - 50) / 50$$

Mitä enemmän otosprosenttiluku p poikkeaa 50 %:sta, sitä suurempi testisuureen z itseisarvo on. Kun testisuureen arvo kasvaa, pienenee siihen liittyvä riskitodennäköisyyden α arvo. Eli mitä pienempi riskitodennäköisyys α on, sitä painokkaammin voidaan nollahypoteesi hylätä. Käytössä on seuraava viisiportainen asteikko.

Riskitodennäköisyys α	Riskitaso eli merkitsevyystaso
$\alpha > 0,10$	tilastollisesti ei-merkitsevä
$0,05 < \alpha \leq 0,10$	tilastollisesti oireellinen
$0,01 < \alpha \leq 0,05$	tilastollisesti melkein merkitsevä
$0,001 < \alpha \leq 0,01$	tilastollisesti merkitsevä
$\alpha \leq 0,001$	tilastollisesti erittäin merkitsevä

Usein ”tilastollisesti oireellinen”-taso unohtetaan ja tilastollisesta merkitsevyydestä puhutaan vasta kun riskitodennäköisyys on pienempi kuin 0,05.

Huom. Käytä tilastollisen testauksen yhteydessä aina sanaa merkitsevä - ei esim. sanaa merkittävä. Tulosten käytännön merkittävyys on aivan eri asia kuin tilastollinen merkitsevyys.

Tilastolliset testit ryhmitellään sen mukaan, mistä tunnusluvusta on kyse. Seuraavassa käsitellään prosenttiluku- ja keskiarvotestejä. Tixelissä on näille testeille valmiit syöttöpohjat, joihin syötetään parametrejä, otostunnuslukuja ja otoskokoja koskevat tiedot. Tuloksena saadaan testisuureen arvo ja riskitodennäköisyys yksi- ja kaksisuuntaisen vaihtohtoisen hypoteesin tapauksessa.

Yksisuuntaiselle vaihtohtoiseselle hypoteesille riskitaso lasketaan aina testisuureen itseisarvosta. Tämä ei aiheuta virheitä, jos otostestisuure on vaihtohtoisen hypoteesin osoittamalla alueella. Jos sen sijaan näin ei ole, tulee nollahypoteesi hyväksyä itsestään selvyytenä riippumatta siitä mitä riskitodennäköisyys näyttää.

13.2. Luottamusväli

Populaation tuntemattomalle tunnusluvulle eli parametrille määrätään luottamusväli otostunnuslukujen funktiona siten, että parametri halutulla todennäköisyydellä kuuluu tähän väliin. Yleisimmin käytetään 95 %:n luottamusväliä, joka vastaa hypoteesin testauksessa 5 %:n riskitasoa.. Esimerkiksi keskiarvon 95 %:n luottamusväliksi saadaan

$$\bar{x} \pm 1,96 \frac{s}{\sqrt{n}}$$

jossa \bar{x} on otoskeskiarvo, s on keskihajonta ja n on otoskoko. Kerroin 1,96 on oikea vain, jos otoskoko n on suurempi kuin 100. Sataa pienemmällä otoskoon arvoilla kyseinen kerroin katsotaan t-jakaumasta.

Esimerkiksi erään sektorin palkkatutkimuksessa saatiin kahden sadan hengen otoksesta kuukauden keskiansioksi 13400 mk ja keskihajonnaksi 1300 mk Edellä mainittu kaava antaa todellisen keskiansion 95%:n luottamusväliksi (13219 mk, 13581 mk). Eli voidaan kohtalaisella varmuudella sanoa, että sektorin todellinen keskiansio kuuluu tähän väliin.

Laskentapohjissa tulee automaattisesti näkyviin luottamusväli, kun tarvittavat tunnusluvut on syötetty arkille.

13.3. Prosenttilukutestit

Prosenttiluvuille on kolme laskentapohjaa: yhden otoksen, kahden korreloimattoman otoksen ja kahden korreloivan otoksen testit. Kahden korreloimattoman otoksen asetelma on kyseessä silloin, kun otokset on poimittu toisistaan riippumattomalla tavalla. Korreloivat otokset syntyvät esim. *ennen-jälkeen-*tutkimusasetelmasta ja *vastinparimenettelystä*.

13.3.1. Yhden otoksen prosenttilukutesti

Testi testaa poikkeako otoksesta laskettu %-luku tilastollisesti merkitsevästi nollahypoteesin %-luvusta. Näiden kahden prosenttiluvun lisäksi on tunnettava otoskoko. Tiedot syötetään oheiselle lomakkeelle. Jos halutaan laskea vain luottamusväli, jätetään nollahypoteesi tyhjäksi.

YHDEN OTOKSEN PROSENTTILUKUTESTI			
AINEISTO:			
MUUTTUJA:			
	OTOS	NOLLAHYPOTEESI	TESTISUURE
%-luku:			#DIV/0!
Otoskoko:			RISKITASO RISKITASO
			1-suunt. 2-suunt.
		p-arvo:	#DIV/0! #DIV/0!
			#DIV/0! #DIV/0!
Perusjoukon prosenttiluvun 95%:n luottamusväli:			#DIV/0! #DIV/0!
<p>Ohje: Täytä harmaalla merkityt solut (prosenttiluvut ilman %-merkkiä). Lv: ssä ei tarvita nollahypoteesia. Jatko: A vaa Tixel-valikko ja valitse vaihtoehto.</p>			

Esimerkki Vaaleissa oli kaksi ehdokasta, joiden kannatusosuudet tuhannen hengen otoksessa olivat 47 % ja 53 %. Poikkeavatko prosenttiluvut tilastollisesti merkitsevästi toisistaan? Kyseessä on yhden otoksen prosenttilukutesti, jossa otoksen prosenttiluku on 47 % (tai 53 %) ja nollahypoteesi on, että molemmilla on yhtä suuri kannatus eli prosenttiluku on 50 %.

Kaksisuuntaisen testin riskitaso 0,058

osoittaa, että prosenttiluvut eivät eroa tilastollisesti merkitsevästi toisistaan. Populaation tuntemattoman prosenttiluvun 95 %:n luottamusväliksi saadaan (43,9 %, 50,1 %).

	OTOS	NOLLAHYPOTEESI	TESTISUURE
%-luku:	47	50	-1,90
Otoskoko:	1000		RISKITASO RISKITASO
			1-suunt. 2-suunt.
		p-arvo	2,890 % 5,781 %
			Melkein merki Oireellinen
Perusjoukon prosenttiluvun 95%:n luottamusväli:			43,91 50,09

13.3.2. Kahden korreloimattoman otoksen prosenttilukutesti

Kaksi otosta on korreloimattomia, jos niillä ei ole mitään yhteyttä keskenään. Seuraavassa alaluvussa (13.1.3) käsitellään kahden korreloivan otoksen tapaus-ta.

Kahden korreloimattoman otoksen prosenttilukutesti testaa poikkeavtko kahden korreloimattoman otoksen prosenttiluvut toisistaan tilastollisesti merkitsevästi. Kahden prosenttiluvun lisäksi on tiedettävä molempien otosten otoskoot. Tiedot syötetään oheiselle lomakkeelle.

KAHDEN KORRELOIMATTOMAN OTOKSEN PROSENTTILUKUTESTI				
AINEISTO:				
MUUTTUJA:				
	1. OTOS	2. OTOS	TESTISUURE	
%-luku:				#DIV/0!
Otoskoko:			RISKITASO	RISKITASO
			1-suunt.	2-suunt.
			#DIV/0!	#DIV/0!
Populaatioiden prosenttilukujen erotuksen				
				0
		piste-estimaatti:		
		95%:n luottamusväli:	#DIV/0!	#DIV/0!

O hje: Täytä harmaalla merkityt solut (prosenttiluvut ilman %-merkkiä).
Riskitaso ilmoittaa testin tuloksen.
Jatko: A vaa Tixel-valikko ja valitse vaihtoehto.

Esimerkki. Erään ominaisuuden prosenttiluvut miesten ja naisten otoksissa olivat vastaavasti 30 % ja 40 %. Poikkeavtko prosenttiluvut tilastol-

	1. OTOS	2. OTOS	TESTISUURE	
%-luku:	30	40		-2,04
Otoskoko:	200	180	RISKITASO	RISKITASO
			1-suunt.	2-suunt.
			0,020478367	0,040956735
Populaatioiden prosenttilukujen erotuksen				
				-10
		piste-estimaatti:		
		95%:n luottamusväli:	-19,59	-0,41

lisesti merkitsevästi toisistaan, kun otoskoot olivat 200 ja 180 henkeä.

Riskitasot osoittavat, että prosenttilukujen ero on tilastollisesti melkein merkitsevä. Erotuksen luottamusväli ei sisällä nollaa, mikä on sopusoinnussa tilastollisen testin tuloksen kanssa.

13.3.3. Kahden korreloivan otoksen prosenttilukutesti

Kaksi korreloivaa otosta syntyy esimerkiksi vastinparimennettelyn kautta tai ennen-jälkeen tutkimusasetelmasta.

Vaikka tilastollinen nollahypoteesi on sama kuin edellisen alaluvun tapauksessa eli perusjoukkojen prosenttiluvut ovat yhtä suuret, on testisuure eri ja se edellyttää ristiintaulukon frekvenssien tuntemista. Frekvenssit syötetään ohaiseen laskentapohjaan.

Kahden korreloivan prosenttiluvun testi			
	1	2	%
1			#DIV/0!
2			#DIV/0!
%	#DIV/0!	#DIV/0!	
Testisuure		Riskitaso	Riskitaso
#DIV/0!		1-suunt.	2-suunt.
		#DIV/0!	#DIV/0!

O hje: Täytä harmaalla merkityt solut ristiintaulukon absoluuttisilla frekvensseillä. Riskitaso ilmoittaa testin tuloksen.
Jatko: A vaa Tixel-valikko ja valitse vaihtoehto.

Esimerkki. 148 hengen otos mitattiin ennen koulutustilaisuutta ja sen jälkeen.

Kyllä-vastausten (=1) osuus oli ennen koulutustilaisuutta 47 % ja koulutustilaisuuden jälkeen 39 %. Viereiseen taulukkoon on syötetty ristiintaulukon frekvenssit.

	1	2	%
1	42	28	47 %
2	16	62	53 %
%	39 %	61 %	
Testisuure		Riskitaso	Riskitaso
1,80907		1-suunt.	2-suunt.
		0,03524	0,07047

Kaksisuuntainen testi ei anna prosenttiluvuille tilastollisesti merkitsevää eroa. Sen sijaan yksisuuntainen testin mukaan prosenttiluvut eroavat tilastollisesti melkein merkitsevästi.

13.4. Keskiarvotestit

Keskiarvotesteistä käsitellään yhden otoksen, kahden korreloimattoman otoksen ja useamman korreloimattoman otoksen eli yksisuuntaisen varianssianalyysin testejä.

Kahden korreloivan otoksen keskiarvotesti palautuu yhden otoksen keskiarvotestiin siten, että muodostetaan erotusmuuttuja ja testataan, poikkeako erotusmuuttujan keskiarvo nolasta perusjoukon tasolla. Tällöin käytetään alaluvun 13.4.1.

(Yhden otoksen keskiarvotesti) laskentapohjaa siten, että keskihajonta lasketaan erotusmuuttujasta ja otoskoko on vastinparien lukumäärä.

13.4.1. Yhden otoksen keskiarvotesti

Yhden otoksen keskiarvotestillä testillä testataan, poikkeako otoskeskiarvo tilastollisesti merkittävästi populaation keskiarvosta. Näiden kahden keskiarvon lisäksi laskentapohjaan kirjoitetaan keskihajonta ja otoskoko.

YHDEN OTOKSEN KESKIARVOTESTI		
OTOS	NOLLAHYPOTEESI	TESTISUURE
Keskiarvo:		#DIV/0!
Keskihajonta:		VAP. ASTE
Otoskoko:		-1
		RISKITASO RISKITASO
		1-suunt. 2-suunt.
	p-arvo:	#DIV/0! #DIV/0!
		#DIV/0! #DIV/0!
Perusjoukon keskiarvon 95%:n luottamusväli:	#NUM!	#NUM!

Ohje: Täytä harmaalla merkityt solut. Jatko: Avaa Tixel-valikko ja valitse vaihtoehto.

Esimerkki. Testataan poikkeako otoskeskiarvo 98 tilastollisesti merkittävästi nollahypoteesin mukaisesti keskiarvosta 100. Kahden sadan hengen otoksesta laskettu keskihajonta on 8,5.

OTOS	NOLLAHYPOTEESI	TESTISUURE
Keskiarvo:	98	100
Keskihajonta:	8,5	-3,3276
Otoskoko:	200	VAP. ASTE
		199
		RISKITASO RISKITASO
		1-suunt. 2-suunt.
	p-arvo:	0,052 % 0,104 %
		Erittäin merk Merkitsevä
Perusjoukon keskiarvon 95%:n luottamusväli:		96,81 99,19

Kaksisuuntaisen testin riskitaso $p=0,00104$ osoittaa, että ero tilastollisesti merkittävä. Vastaavasti arvo 100 ei kuulu luottamusväliin.

13.4.2. Kahden korreloimattoman otoksen keskiarvotesti

Kaksi otosta on korreloimattomia, jos niillä ei ole mitään yhteyttä keskenään. Esimerkiksi kaupungeissa ja maaseudulla asuvien otokset ovat korreloimattomia, jos ei ole käytetty vastinparimenettelyä. Tyypillisesti korreloimattomassa tapauksessa otoskoot ovat erisuuret. Jos otoskoot ovat yhtä suuret, kannattaa ainakin tarkistaa, ettei otosten välillä ole riippuvuutta.

Kahden korreloimattoman otoksen testi testaa, poikkeavatko otoskeskiarvot tilastollisesti merkitsevästi toisistaan. Laskentapohjaan syötetään keskiarvojen lisäksi keskihajonnat ja otoskoot.

KAHDEN KORRELOIMATTOMAN OTOKSEN KESKIARVOTESTI			
	1. OTOS	2. OTOS	TESTISJURE
Keskiarvo:			#DIV/0!
Hajonta:			VAP. ASTE
Otoskoko:			-2
			RISKITASO RISKITASO
			1-suunt. 2-suunt.
			#DIV/0! #DIV/0!
Populaatioiden keskiarvojen erotuksen			0
piste-esimaatti			#NUM! #NUM!
95%:n luottamusväli:			

Ohje: Täytä harmaalla merkityt solut. Riskitaso ilmoittaa testin tuloksen. Jatko: A vaa Tixel-valikko ja valitse vaihtoehto.

Esimerkki. Testataan eroavatko kahden sektorin ostotutkimuksista lasketut keskipalkat 10200 mk ja 10500 mk tilastollisesti merkitsevästi toisistaan. Vastaavat keskihajonnat ovat 1200 mk ja 1300 mk sekä otoskoot 200 ja 150 henkeä.

	1. OTOS	2. OTOS	TESTISJURE
Keskiarvo:	10200	10500	-2,2330428
Hajonta:	1200	1300	VAP. ASTE
Otoskoko:	200	150	348
			RISKITASO RISKITASO
			1-suunt. 2-suunt.
			0,0130907 0,02618149
Populaatioiden keskiarvojen erotuksen			-300
piste-esimaatti			-564,23 -35,77
95%:n luottamusväli:			

Kaksisuuntaisen testin riskitaso osoittaa, että keskipalkoissa on tilastollisesti melkein merkitsevä ero riskitasolla $p=0,026$. Vastaavasti nolla ei kuulu luottamusväliin.

13.4.3. Yksisuuntainen varianssianalyysi

Yksisuuntaisella varianssianalyysillä testataan, poikkeavatko yhden ryhmittelymuuttujan mukaan lasketut keskiarvot tilastollisesti merkitsevästi toisistaan. Keskiarvojen tulee keskenään korreloimattomia.

Laskentapohjaan syötetään ryhmien otoskoot, keskiarvot ja keskihajonnat. Keskihajonnoja laskettaessa

Yksisuuntainen varianssianalyysi

Ohje: Täytä lukumäärä-, keskiarvo- ja keskihajonta-sarakkeet. Ryhmäsarakkeeseen voit kirjoittaa ryhmien tunnuksset (ei pakollinen). Valitse lopuksi Tixel-valikosta 1-suunt. VA-vaihtoehto.

Ryhmä	Lkm	Keskiarvo	Keskihajonta
-------	-----	-----------	--------------

tulee jakajana käyttää lauseketta $n-1$.

Seuraavassa esimerkissä on testattu, onko keskiarvojen 5, 6 ja 7 välillä tilastollisesti merkitsevää eroa, kun keskihajonnat ovat kahden molemmin puolin ja otoskoot ovat 10, 20 ja 30.

Ryhmä	Lkm	Keskiarvo	Keskihajonta
A	10	5	1,8
B	20	6	2
C	30	7	2,2

F = 3,87 vap. asteet: 2 ja 57
 p = 0,0266 Tilastollisesti melkein merkitsevä

Testin riskitodennäköisyys on ilmaistu merkinnällä "p=0,0266" eli keskiarvojen ero on tilastollisesti melkein merkitsevä.

Huom. Havaintomatriisista saadaan yksisuuntainen varianssianalyysi menetelmällä "Ryhmäkeskiarvot, yksi ryhmittelymuuttuja".

13.5. Otokoko

Usein tilastotieteilijältä kysytään, kuinka suuri tulee otoksen olla, jotta perusjoukosta saadaan edustava näyte. Tähän kysymykseen ei ole vastausta, vaan kysymys on asetettava seuraavasti:

Kuinka suuri otoskoon tulee olla, jotta perusjoukon tuntematon parametri saadaan arvioituksi tietyllä tarkkuudella.

Tarkkuus ilmaistaan luottamusvälin pituuden puolikkaana. Tätä kutsutaan myös virhemarginaaliksi. Seuraavassa on laskentapohjat otoskoon määrittämiseksi prosenttiluvun ja keskiarvon tapauksissa. Ne saadaan käyttöön valitsemalla tämän luvun alussa olevasta menetelmäluettelosta vaihtoehto "Otoskoon määrittäminen ja estimoinnin tarkkuus, prosenttiluku (tai keskiarvo)". Kuten tekstistä ilmenee voi näitä laskentapohjia käyttää paitsi otoskoon määrittämiseen myös estimoinnin tarkkuuden laskemiseen, kun otoskoko tiedetään.

13.5.1. Otoksoon määrääminen prosenttiluvun tapauksessa

Jotta otoskoko voidaan prosenttiluvun tapauksessa laskea, on tuntemattomasta parametrinä oltava jonkinlainen arvio. Esimerkiksi jos kyseessä on puolueiden kannatustutkimus, kannattaa arviona käyttää noin 20 %:ia, sillä suurimpien puolueiden kannatusosuudet ovat lähellä tätä lukua.

Otoskoon määrääminen ja estimoinnin tarkkuus
%-luvut

Perusjoukon suuruus:	4 000 000	
Arvio estimoitavasta %-luvusta:	20	Laskennan tulos:
Otoskoko:		1535
<i>tai</i>		#DIV/0!
95%:n luottamusvälin puolikas:	2	

Ohje: Täytä rasteroidut solut. Otoskoko ja 95%:n luottamusvälin pituus ovat vaihtoehtoisia tapoja määrittellä otannan tarkkuus. Jos syötät otoskoon arvon, antaa laskenta lv:n puolikkaan pituuden. Jos syötät lv:n puolikkaan arvon, antaa laskenta otoskoon arvon.

Jos tarkkuusvaatimus on kaksi prosenttiyksikköä, saadaan otoskooksi 1535.

On huomattava, että perusjoukon suuruus ei tässä tapauksessa paljonkaan vaikuta lopputulokseen. Jos perusjoukon suuruus olisi ollut 100 000 henkeä, saadaan otoskooksi 1513.

13.5.2. Otoskoon määrääminen keskiarvon tapauksessa

Määrättäessä otoskoko keskiarvon tapauksessa on oltava arvio muuttujan keskihajonnasta. Esimerkiksi jos halutaan arvioida 50 000 palkansaajan kuukausiansion keskiarvo sadan markan tarkkuudella ja tiedetään keskihajonnan olevan noin 1300 mk, saadaan otoskooksi 641.

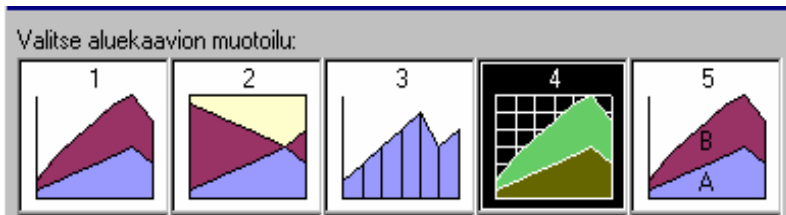
Otoskoon määrääminen ja estimoinnin tarkkuus
Keskiarvo

Perusjoukon suuruus:	50 000	
Keskihajonta:	1300	Laskennan tulos:
Otoskoko:		641
<i>tai</i>		#DIV/0!
95%:n luottamusvälin puolikas:	100	

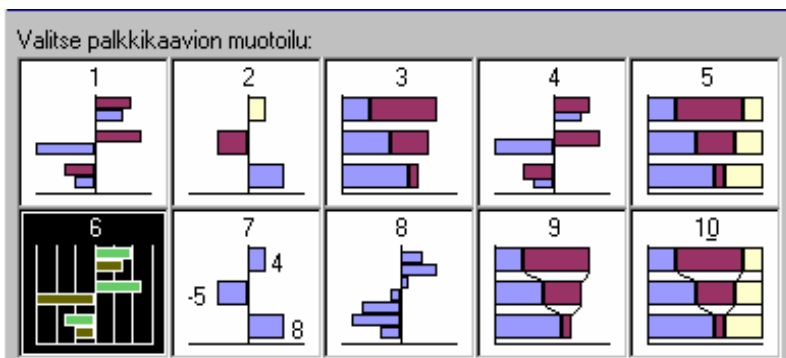
Ohje: Täytä rasteroidut solut. Otoskoko ja 95%:n luottamusvälin pituus ovat vaihtoehtoisia tapoja määrittellä otannan tarkkuus. Jos syötät otoskoon arvon, antaa laskenta lv:n puolikkaan pituuden. Jos syötät lv:n puolikkaan arvon, antaa laskenta otoskoon arvon.

Liite 1. Excelin tärkeimmät kuvatyypit

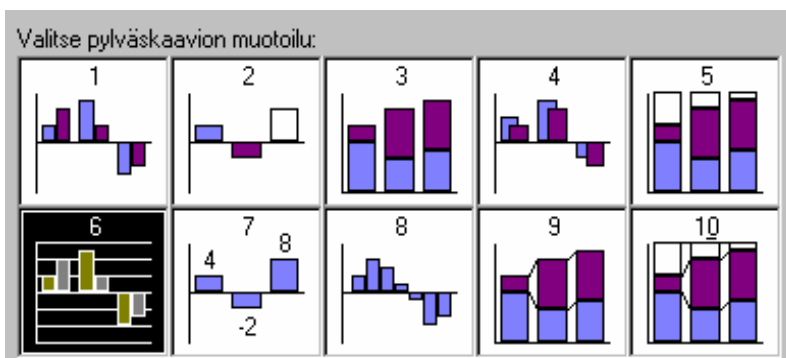
Alue-kuvat (area)



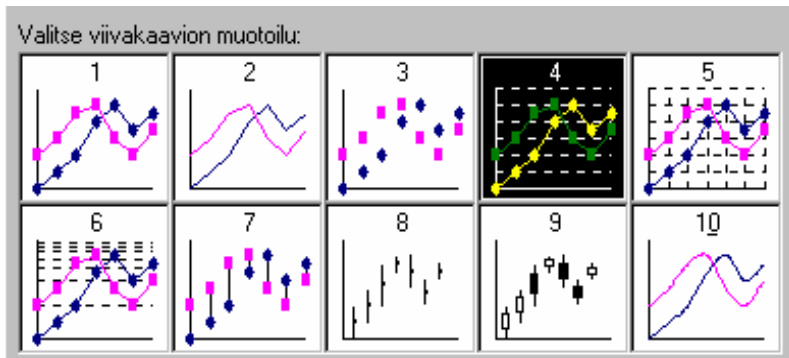
Palkkikuvat (bar)



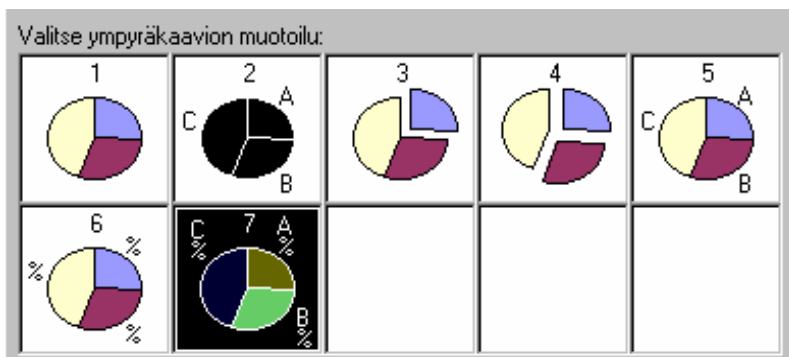
Pylväskuvat (columns)



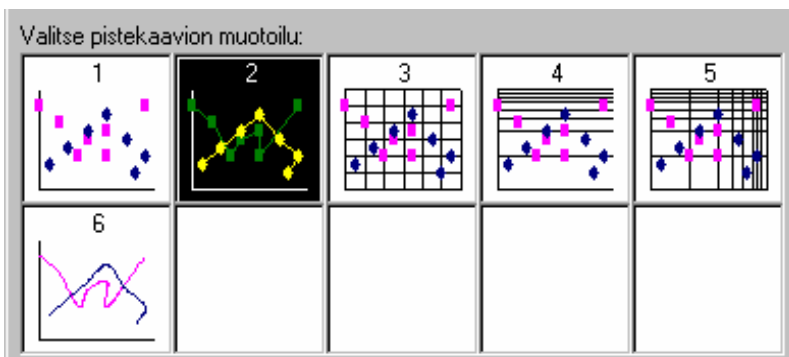
Viivakuvat (line)



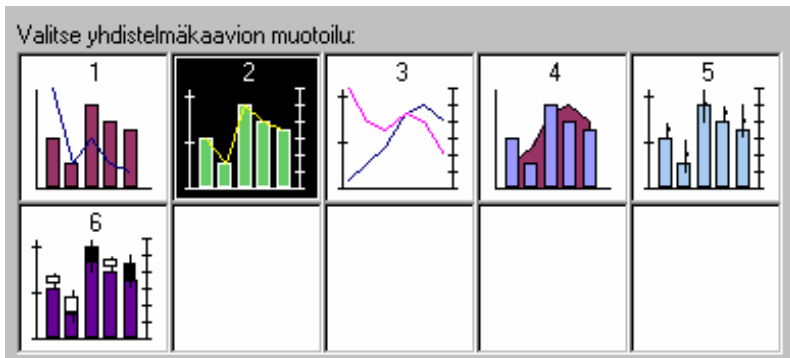
Piirakkakuvat (Pie)



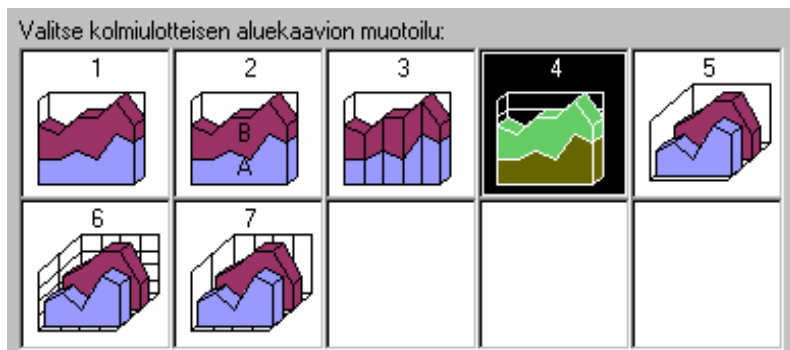
Hajontakuvat (xy)



Yhdistelmäkuvat (Combination)



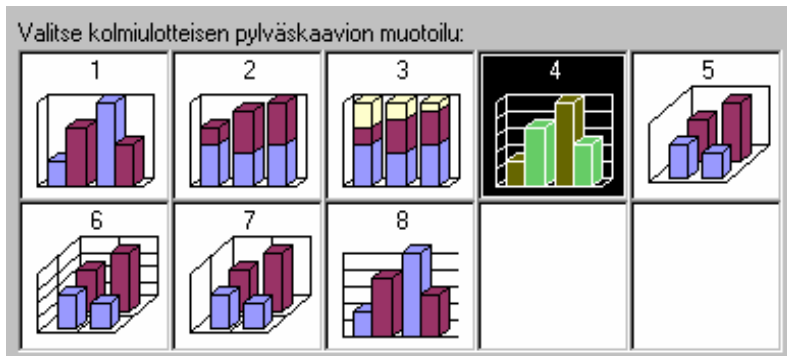
Kolmiulotteiset aluekuvat (3-D Area)



Kolmiulotteiset palkkikuvat (3-D Bar)



Kolmiulotteiset pylväskuvat (3-D Column)



Kolmiulotteiset viivakuvat (3-D Line)



Kolmiulotteiset piirakkakuvat (3-D Pie)

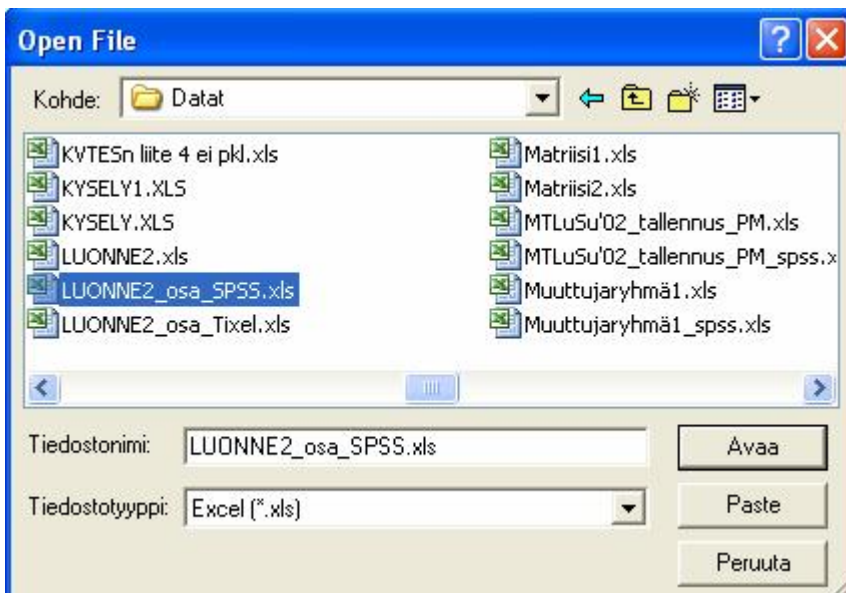


Liite 2. Datan siirto Excelin ja SPSS:n välillä

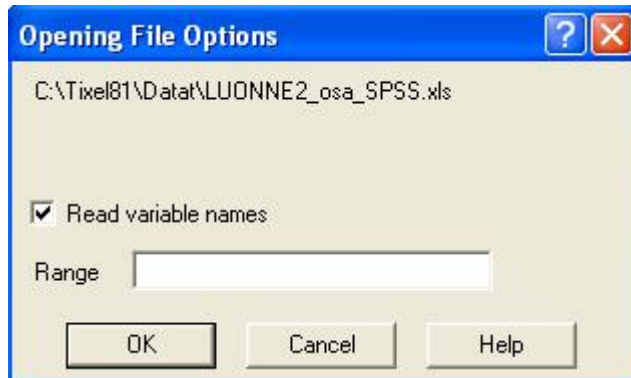
Excelistä SPSS:ään

Tixelin data-tiedosto on valmis käytettäväksi SPSS:ssä, kun seuraavat muutokset on tehty Excelin puolella.

- 1) Tixelin datan kaksi ensimmäistä riviä poistetaan **Edit/Delete**-komennolla (kaksi ensimmäistä riviä on ensin merkittävä).
- 2) Jos havaintomatriisiin on sijoitettu luokkien nimiä viimeisen varsinaisen rivin alapuolelle, poistetaan nämä luokkien nimet (kts 2.2).
- 3) Tallenna muokattu matriisi uudella nimellä (**File/Save As**-komento).
- 4) Sulje tiedosto **File/Close**-komennolla.
- 5) Siirry SPSS:ään ja anna komento **File/Open Data**. Saat seuraavanlaisen valintaikkunan.



Valitse *Tiedostotyyppi*-kentästä Excel-vaihtoehto, jolloin *Tiedostonimi*-ikkunassa näkyvät aktiivisen hakemiston .xls-tyyppiset tiedostot (voit ehkä joutua vaihtamaan hakemistoa). Merkkää Excelissä tallennettu tiedosto ja klikkaa *Avaa*-painiketta, jolloin näytölle tulee seuraava ikkuna.

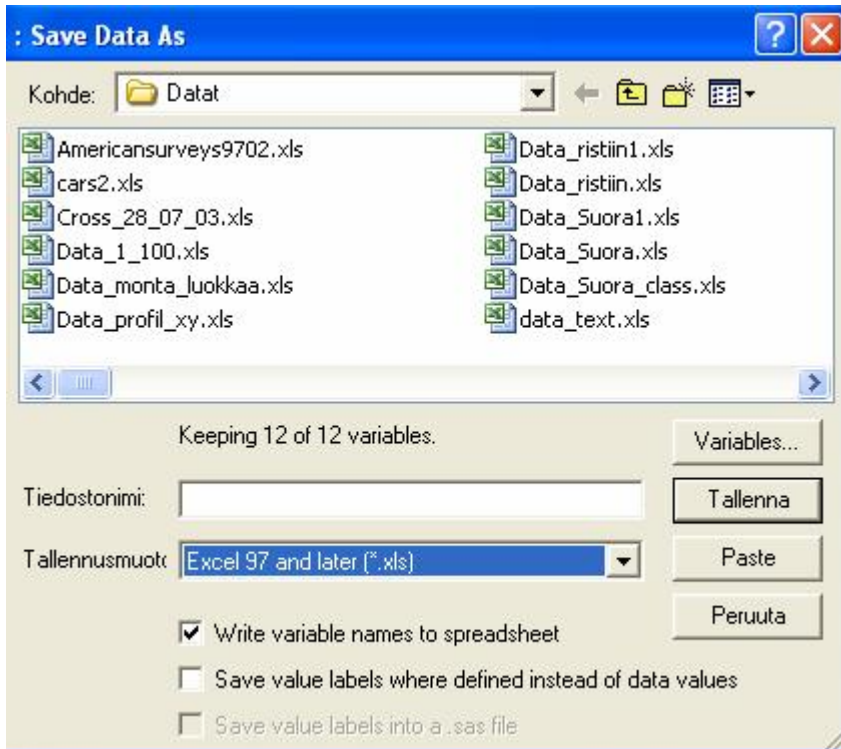


Laita rasti *Read variable names*-kohtaan. *Range*-vaihtoehdon avulla olisi mahdollista lukea Excel-matriisin jokin alue, mutta silloin pitää tietää alueen koordinaatit.

Huom. SPSS käyttää muuttujien nimistä vain kahdeksaa ensimmäistä merkkiä eikä tunnista skandinaavisia merkkejä. Muista lisäksi määritellä puuttuvan tiedon arvo. Tämä tieto ei siirry automaattisesti.

SPSS:stä Exceliin

Avaa siirrettävä data SPSS:ssä ja anna komento **File/Save As**. Saat seuraavanlaisen valintaikkunan.



Valitse *Tallennusmuoto*-kentästä Excel-vaihtoehto ja merkitse *Write variable names to spreadsheet*-kohtaan rasti. Keksi tiedostolle uusi nimi ja paina *Tallenna*-painiketta.

Siirry takaisin Exceliin ja avaa tallennettu tiedosto. Muokkaa havaintomatriisi Tixel-muotoon lisäämällä matriisin alkuun kaksi riviä (puuttuvien tietojen arvot!).

Jos SPSS:ssä on käytetty tyhjää solua puuttuvan tiedon merkinä, on Excelissä näissä soluissa merkintä DIV/0#. Muuta tämä merkintä puuttuvan tiedon arvoksi komennolla **Edit/Replace**.