

998

Toistojen määrän arviointi kokeellisessa tutkimuksessa

Stig Storfors

Tilastotieteen pro gradu -tutkielma
14. toukokuuta 1998

Jyväskylän yliopisto
tilastotieteen laitos

Tilastotoimen menetelmien maisteriohjelma

(<http://www.stat.jyu.fi/>)

(Päivitetty 16.5.1997)

Tilastotoimen menetelmien maisteriohjelman tavoite on kouluttaa opiskelija tilastotiedon keruun, jatkojalostuksen ja käytön asiantuntijaksi nykyaikaisessa tilastojärjestelmäympäristössä, jossa datat ovat survey-, koeasetelma- tai rekisteriperusteisia. Koulutus järjestetään yhteistyönä Jyväskylän yliopiston tilastotieteen laitoksen kanssa. Tilastotieteelliseltä kannalta kyse on survey-menetelmiin, biostatistiikkaan ja ekonometriaan erikoistuneiden tilastoasiantuntijoiden koulutuksesta. Ohjelman kesto päätoimisella opiskelijalla on kaksi lukuvuotta.

Tilastotoimen ytimenä ovat tilastotieteen, erityisesti tilastotoimen teorian ja tietojenkäsittelyn kurssit. Ohjelmaan voidaan sisällyttää myös koulutustavoitteita tukevia opintojaksoja lähitieteistä (esim. talous-, yhteiskunta- ja viestintätieteet) sekä yritystoimintaan perehdyttäviä kursseja. Tarjottava opetus koostuu osittain tilastotieteen laitoksen opetusohjelmasta, erityisesti suunnitelluista tilastotoimen teorian kursseista ja ostopalveluina hankituista muista erikoiskursseista. Opettajista pääosa on kotimaasta. Vierailijoina on myös ulkomaalaisia asiantuntijoita, joten opetus on osin englanninkielistä.

Ohjelman tärkeä osa on yliopiston ulkopuolisessa yhteistyötoimipaikassa suoritettu pro gradu -tutkielma ja siihen liittyvä harjoittelu. Yhteistyötoimipaikat ovat virallisesta tilastotoimesta, suuryrityksistä tai tutkimuslaitoksista. Ne osallistuvat osaltaan harjoittelun kustannuksiin. Perusajatus on, että pro gradu -tutkielma tai osa siitä toisi tutkimustulostensa osalta lisäarvoa yhteistyötoimipaikalle. Tätä tukenee se, että harjoittelu jaetaan kahteen osaan, joista ensimmäinen eli orientoiva osa on ensimmäisen opintovuoden lopussa. Sen jälkeen opiskelija palaa yliopisto-opiskeluihin yhden lukukauden ajaksi syventääkseen tietojaan siinä tilastotieteen osa-alueessa, joka on tarpeen aiotussa pro gradu -tutkielmassa. Maisteriohjelman viimeinen lukukausi muodostaa yhtenäisen tutkimusjakson, jonka aikana tehdään pro gradu -tutkielma.

Master's Programme in Statistical Systems

(<http://www.stat.jyu.fi/>)

(Updated 16.5.1997)

The main target of the Program is to educate high qualified professionals in the collection analysis, managing and dissemination of large data sets. The Program has been built mainly on the regular curriculum of the Department of Statistics in the University of Jyväskylä. In point of view of statistical sciences, the Program in Statistical Systems concentrate in the knowledge of survey methodology, biometry and econometry. Full-time students can graduate in two Academic years. The core courses cover the advanced theory in mathematical statistics, of stational systems especially the survey methodology and some courses in information technology. Still the program is flexible that a moderate amount of different kinds of studies in the neighboring sciences can be included as studies in economics, social sciences and communication. Most of the instructors come from Finland but visiting professors from outboard are eventual using English.

An important part of the Program crows from the cooperational research work created between the Department of Statistics and the research units located outside the University of Jyväskylä. Those units are research and development departments in the bodies of official statistics, big business firms and research institutes. They share the costs of practices. Basic idea is that the MS.c. Thesis wil be written from the topics given by the cooperational research units and so the results could be of some contributed value to them. As an appropriate mode of policy may be here thus that the time of practice skills for the Thesis. This is proceeded in 6 final months of the Program, in collaboration with the same cooperating research unit.

Tiivistelmä

Stig Storfors: *Toistojen määrän arviointi kokeellisessa tutkimuksessa*

Tilastotieteen pro gradu -tutkielma, Jyväskylän yliopisto, 14. toukokuuta 1998.

Sivuja 34, liitteitä 3.

Tässä työssä on pyritty vastaamaan toimeksiantajan vaatimukseen liittyen tehtävään, jonka perusteella voitaisiin arvioida tulevan kokeen toistojen määrää. Siihen pyritään sekä teoreettisella tasolla että käytännönläheisesti työn myötä valmistuvan sovelluksen kautta.

Teoriaan paneudutaan lähtemällä liikkeelle yksinkertaisimmasta mahdollisesta kyseiseen tulevasta tapauksesta eli yksisuuntaisesta varianssianalyysistä. Siitä edetään kaksisuuntaisen varianssianalyysin ja lineaarisen mallin avulla viimein regressiomallin esittämiseen lineaarisella mallilla.

Teoriasta on tehty käytäntöä valmistuneen sovelluksen myötä. Se on tehty Excel- taulukkolaskentaohjelman versiolla 5.0 . Makrokielenä on käytetty Excelin omaa Visual Basic -kieltä. Lisäksi on käytetty laajasti hyväksi erilaisia taulukkolaskentaohjelmille tyypillisiä toimintoja.

Sovelluksesta on tehty myös käyttöohje, joka on liitteessä 3. Siinä edetään sovelluksen käyttö läpi askel askeleelta. Opastus tapahtuu yhden oletetun kokeen avulla. Siinä olevat kokeen tiedot ovat täysin itse keksittyjä.

Sisällysluettelo

1 Johdanto	1
1.1 Toimeksiannosta yleisesti	1
1.2 Työn sisällön kuvaus	2
2 Tutkimuksen lähtökohdat ja tavoitteet	3
3 Yleistä varianssianalyysistä	5
3.1 Yksisuuntaisen varianssianalyysin malli	5
3.2 Parametrien estimointi	7
3.3 Yksisuuntaisen varianssianalyysin konstruointi	9
3.4 F-testisuureen johtaminen	11
3.5 Kaksisuuntainen varianssianalyysi	12
3.6 F-testisuure kaksisuuntaiselle varianssianalyysille	15
4 Yleinen lineaarinen malli	17
4.1 Lineaarista rajoitteista	18
4.2 Yleinen lineaarinen hypoteesi	21
5 Toistojen määrän arviointi käsittelyvaikutusten avulla	22
5.1 OC-käyrät	22
6 Epäkeskeinen F-jakauma	26
6.1 F-testisuure epäkeskeisyysparametrin alaisessa tilanteessa	26
6.2 Voimakkuus F-jakauman avulla	27
6.3 Epäkeskeisen F-jakauman laskeminen tavallisen F-jakauman avulla	27
7 Toistojen määrän arviointi regressiomallin avulla	30
7.1 Siirtyminen regressiomallista lineaariseen malliin	31
8 Sovellus toistojen määrän ja voimakkuuden laskemiseen	34
8.1 Yhteenveto	34
Lähteet	35
Liite 1 Malliin tulevien termien kyselyyn liittyvät funktiot ja aliohjelmat	
Liite 2 Toistojen määrän ja voimakkuuden laskemiseen käytettävät funktiot ja aliohjelmat	
Liite 3 Käyttöohje	

1 Johdanto

Pro gradu -työn aiheena on toistojen määrän arviointi kokeellisessa tutkimuksessa. Aihetta ei ainakaan kirjallisuuden mukaan ole paljonkaan tutkittu. Syynä tähän on ollut yleensä tietämättömyys tutkittavan vastemuuttujan vaihtelusta. Tätä ongelmaa on kuitenkin Cultor Oy :ssä pyritty ratkaisemaan, mikä antoi mahdollisuuden tarttua tähän aiheeseen.

1.1 Toimeksiannosta yleisesti

Perinteisesti koesuunnittelussa yksi tärkeimmistä ennakkoon suunniteltavista asioista on ollut toistojen määrän arviointi. Jos kokeen suorittajalla ei ole minkäänlaista ennakkotietoa tulevasta kokeesta, on toistojen määrän arviointi suurelta osin arvailujen varassa. Pro gradu -työn ideana on Cultor Oy :ssä aiemmin tehdyistä kokeista saatujen jäännösvariانسsien hyväksikäyttö ja sovelluksen suunnitteleminen tulevien kokeiden toistojen määrän arviointiin. Tähän antaa mahdollisuuden aiempien kokeiden suuri määrä ja niihin liittyvien tietojen suhteellisen helppo saatavuus. Työhön liittyvä toimeksiantajan esittämä kysymys voidaan siis esittää muodossa: “Kuinka paljon tällaisessa kokeessa vaaditaan toistoja näillä taustatiedoilla päästäksemme haluttuun tarkkuuteen?”

Toimeksiantona on siis tiettyjen alkutietojen ja käytössä olevien taustatietojen perusteella määrittää riittävä toistojen määrä. Tällä tarkoitetaan tässä yhteydessä määrää, jolla tutkimuksessa päästään haluttuun tarkkuuteen, kuitenkin siten, ettei toistoja ole 'liikaa'. Tällöin vältytään turhilta lisäkustannuksilta. Täten alapuolelta toistojen määrää rajoittaa annetun tarkkuuden saavuttaminen ja toisaalta yläpuolelta kokeeseen sijoitettava rahamäärä.

Pro gradu -työssä käytettävä toistojen määrän arviointi perustuu ensinnäkin aiemmista tutkimuksista saatuaan arvioon tutkittavan populaation tietyn muuttujan vaihtelusta. Tämän

avulla saadaan arvio tarvittavasta toistojen määrästä. Lisäksi toistojen määrään vaikuttaa myös käytetty ja malli.

Pro gradu -työn tuloksena valmistuva sovellus on linkki käyttäjän haluamien tulosten ja työn tekijän laatiman teorian välillä. Sen perustana on Excel-taulukkolaskentaohjelma, joka on käytössä Cultor Oy :ssä. Tämän ohjelman avulla voidaan tehdä suhteellisen monipuolisiakin matemaattisia laskelmia, joiden avulla luvussa 1 esitelty työ saadaan tehdyksi.

1.2 Työn sisällön kuvaus

Toisessa luvussa käsitellään tutkimuksen lähtökohtia ja tavoitteita. Eli käytännössä sitä, mihin työllä pyritään ja miten siihen päästään.

Kolmannessa luvussa käsitellään sekä yksi- että kaksisuuntaista varianssianalyysiä. Tämä on yksinkertaisin tapaus jäljempänä esiteltävästä ja sovelluksen kannalta tärkeästä yleisestä lineaarisesta mallista. Liikkeelle lähdetään aivan perusteista ja tavoitteena on käsitteiden ja perustoimenpiteiden hallinta.. Lisäksi johdetaan yleisesti käytössä oleva F-testisuure molemmille tapauksille.

Luvussa neljä päästään yleiseen lineaariseen malliin. Malli esitellään laajennettuna versiona varianssianalyttisille malleille. Siihen paneudutaan myös esimerkkien avulla. Lisäksi käsitellään lineaarisia rajoitteita eli kontrasteja.

Luvussa viisi käsitellään toistojen määrän arviointia käsittelyvaikutusten avulla. Menetelmiä on kirjallisuudessa (esim. Montgomery, 1984) esitelty useitakin, mutta tässä käsitellään vain arviointia OC-käyrien avulla.

Luvussa kuusi käsitellään epäkeskeistä F-jakaumaa. Tähän jakaumaan päädytään pro gradu -työn aiheena olevassa toistojen määrän arvioinnissa ja siten luku on avainroolissa tehtävään sovellukseen nähden. Luvun toinen keskeinen käsite on voimakkuuden määrittäminen.

Luvussa seitsemän tutkitaan toistojen määrän arviointia regressiomallin avulla. Tämä luku muodostaa edellisen luvun kanssa kivijalan sovelluksen tekemiseen. Regressiomallia käsitellään hieman yleisesti ja tarkemmin esimerkin avulla.

Kahdeksannessa luvussa kerrotaan yleistä asiaa tehdystä Excel-ohjelmasta.

2 Tutkimuksen lähtökohdat ja tavoitteet

Vaihtelun arvioimiseen on käytössä tyypillisesti kolme mahdollisuutta. Ensimmäinen menetelmä ei oikeastaan perustu suoraan mihinkään tieteelliseen toimintaan, vaan siinä tutkijalla on jonkinlainen näkemys toistojen määrästä, jolla päästäisiin haluttuun tarkkuuteen tulosten suhteen. Tämän hän on lähinnä kokemusperäisesti hyväksi havainnut. Toisena menetelmänä on jonkinlaisen kokeen pilottiversion tekeminen. Tämä tarkoittaa sitä, että tehdään kokeesta 'pienoismalli'. Tällöin siis tehdään vain pieni osa tulevasta kokeesta ja tämän perusteella arvioidaan tiettyjen muuttujien vaihtelua. Tämä keino saattaa antaa suhteellisen hyviä tuloksia, mutta se voi olla melko kallista, aikaa vievää tai teknisesti täysin mahdotonta. Kolmas menetelmä perustuu aiempien tutkimusten antaman aputiedon käyttöön. Pro gradu -työssä tullaan käyttämään viimeksimainittua menetelmää.

Aiempien tutkimusten antaman aputiedon käyttö perustuu Cultor Oy :n aiemmin tekemiin kokeisiin. Näiden lukumäärä on suuri ja kaikki tarvittava tieto on suhteellisen helposti saatavilla. Jokainen tähän asti suoritettu koe on talletettu sähköisessä muodossa. Vaikka mitään täysin yhtenäistä standardia kokeiden dokumentoimiseksi ei tähän saakka olekaan ollut olemassa, on jokaisesta kokeesta olemassa ainakin jonkin verran myös taustatietoja. Pro gradu -työssä ei tulla varsinaisesti tuottamaan mitään varastointimenetelmää, mutta tarkoituksena on kuitenkin antaa suuntaviivat siihen, mitä seikkoja pitäisi ottaa huomioon tietovarastoa suunniteltaessa.

Aiemmin tehtyjen kokeiden hyväksikäyttö perustuu niistä saataviin vastemuuttujan vaihtelua kuvaaviin tunnuslukuihin. Tämä tarkoittaa lähinnä varianssin tutkimista. Jos aiemmin on tehty koe, joka oleellisilta osin vastaa nyt suunnittelun alla olevaa koetta, voidaan toistojen määrälle laskea arvio vanhasta kokeesta saatavaa populaation vaihtelua hyväksikäyttäen. Arvion laatu riippuu vanhoista kokeista koostuvan tietovaraston laajuudesta siten, että mitä laajempi ja kattavampi tietovarasto on, sitä parempia arvioita saadaan. Laatua parantaa tietovaraston kumuloituva luonne. Toisin sanoen sen koko kasvaa

tehtyjen kokeiden myötä.

Toisaalta usein käynee niin, ettei täsmälleen samanlaista koetta löydy. Tällöin tietyillä kriteereillä etsitään mahdollisimman lähellä olevia kokeita. Jos sellainen koe löydetään, niin siitä saatava vaihtelu otetaan suurella painolla mukaan ja toisaalta, jos koe on vain jossakin määrin samanlainen kuin suunnitteilla oleva, otetaan siitä saatava vaihtelu pienemmällä painolla mukaan.

Tutkimuksen tavoitteena on selvittää teoria, jonka pohjalta sovellusta voidaan lähteä kehittämään. Tutkimus lähtee liikkeelle yksinkertaisesta varianssianalyysistä. Sen jälkeen käsitellään regressioanalyysiä. Tämän jälkeen laajennetaan edellisiä yleisen lineaarisen mallin suuntaan. Lopuksi linkitetään regressioanalyttinen malli lineaarisen mallin avulla sovelluksen vaatimaan muotoon.

Opittua teoriaa tullaan lopuksi soveltamaan käytännössä. Tämä tapahtuu Excel-ympäristöön valmistuvan sovelluksen kautta. Sen tarkoituksena on palvella tutkijaa kokeen suunnitteluvaiheessa, jolloin päätetään kokeeseen tuleva toistojen määrä.

3 Yleistä varianssianalyysistä

Varianssianalyysissä kiinnostuksen kohteena on tarkasteltavien tekijöiden eri tasojen keskiarvojen vertailu. Nimitys johtuu siitä, että keskiarvojen yhtäsuuruuden testaus perustuu erilaisiin varianssin kaltaisiin neliösummiin. Kun ryhmiä (joukkoja) on vain kaksi, voidaan käyttää t-testiä. Jos taas ryhmiä on kolme tai enemmän, tarkasteltavien parittaisten vertailujen määrä kasvaa nopeasti ja todennäköisyys tehdä virhepäätelmiä kasvaa. Tämän välttämiseksi pyritään usean erillisen vertailun korvaamiseen yhdellä vertailulla. Tarkasteltaessa useampaa ryhmää samanaikaisesti tehdään varianssianalyysi, jolloin päätös hypoteesin hylkäämisestä tai hyväksymisestä tehdään vain kerran. Testattavana hypoteesina voi esimerkiksi olla, että kaikkien ryhmien keskiarvot ovat yhtäsuuret ja vastahypoteesina, että ainakin yhden joukon keskiarvo poikkeaa muiden joukkojen keskiarvoista.

Yksisuuntaisessa varianssianalyysissä tarkastellaan yhden koetekijän eri tasojen vaikutusta tutkittavaan muuttujaan, kun taas kaksi- tai useampisuuntaisessa varianssianalyysissä kahden tai useamman tekijän omavaikutusten lisäksi otetaan huomioon tekijöiden yhdysvaikutus.

3.1 Yksisuuntaisen varianssianalyysin malli

Tarkastellaan tyypillistä yksisuuntaisessa varianssianalyysissä käsiteltävää dataa. Se esitetään taulukossa (3.1), jossa käsittelyiden (tasojen) määrä on a ja näissä jokaisessa toistojen määrä on n . Tällöin päästään yksinkertaisimpaan eli tasapainoiseen asetelmaan. Jos toistojen määrä vaihtelee eri tasoilla, laskukaavat muuttuvat hieman, mutta periaate säilyy ennallaan. Tulosuuttuja on y . Jatkossa merkitään havaintojen kokonaismäärää (eli

havaintojen $lkm \cdot$ käsittelyiden lkm) N :llä. Lisäksi merkitään käsittelyn i saaneen havainnon j saamaa tulosmuuttujan y arvoa y_{ij} :llä.

Taulukko 3.1 Havaintotaulukko

		Havainto			
		1	2	...	n
Käsittely	1	y_{11}	y_{12}	...	y_{1n}
	2	y_{21}	y_{22}	...	y_{2n}

	a	y_{a1}	y_{a2}	...	y_{an}

Yksisuuntaisen varianssianalyysin mukainen malli havainnolle y_{ij} voidaan esittää muodossa

$$y_{ij} = \mu + \tau_i + \epsilon_{ij}, \quad (3.1)$$

missä μ = tulosmuuttujan y odotusarvo (eli y :n keskiarvo populaatiossa)
 τ_i = käsittelyn i vaikutus y :n odotusarvoon
 ϵ_{ij} = satunnaisvirhe ja
 $i = 1, 2, \dots, a$ ja $j = 1, 2, \dots, n$.

Lisäksi merkitään

μ_i = tulosmuuttujan odotusarvo käsittelylle i , jolloin
 $\tau_i = \mu_i - \mu$.

Satunnaisvirhetermille ϵ_{ij} oletetaan:

- 1) $E(\epsilon_{ij}) = 0$ kaikille i, j
- 2) $\text{var}(\epsilon_{ij}) = \sigma^2$ kaikille i, j (niinsanottu homoskedastisuusehto)
- 3) ϵ_{ij} :t korreloimattomia keskenään.

Kuten jäljempänä käy ilmi, on parametreja $a + 1$ kappaletta, joten tarvitaan ehto, jolla malli

tulee identifioituvaksi. Se on

$$\sum_{i=1}^a \tau_i = 0 . \quad (3.2)$$

3.2 Parametrien estimointi

Parametrien estimoinnissa käytetään niisanottua PNS-menetelmää (pienimmän neliösumman menetelmä). Rajoitutaan seuraavassa tasapainoiseen asetelmaan eli oletetaan, että $n_1 = n_2 = \dots = n_a = n$.

Merkitään virheneliösummaa L :llä siten, että

$$L = \sum_{i=1}^a \sum_{j=1}^n \epsilon_{ij}^2 = \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \mu - \tau_i)^2 . \quad (3.3)$$

Tavoitteena on minimoida L μ :n ja τ_i :n suhteen. Tämä tapahtuu siten, että asetetaan L :n derivaatat sekä μ :n että τ_i :n suhteen nolaksi ja ratkaistaan näin saaduista $a+1$ yhtälöstä μ :n ja τ_i :n estimaatit. Eli merkitään

$$\begin{aligned} \frac{\partial L(\mu, \tau_i)}{\partial \mu} &= 0 \\ \frac{\partial L(\mu, \tau_i)}{\partial \tau_i} &= 0 \quad i = 1, 2, \dots, a . \end{aligned} \quad (3.4)$$

Tämä johtaa yhtälöryhmään

$$\begin{aligned} -2 \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \hat{\mu} - \hat{\tau}_i)^2 &= 0 \\ -2 \sum_{j=1}^n (y_{ij} - \hat{\mu} - \hat{\tau}_i)^2 &= 0 \quad i = 1, 2, \dots, a . \end{aligned} \quad (3.5)$$

Tässä kohden tarvitaan kaavan (3.2) mukaista identifioituvuusehtoa. Nyt muodostuu $a+1$ kappaletta yhtälöitä eli niin sanotut PNS-yhtälöt, jotka ovat muotoa

$$\hat{\epsilon}_{ij} = y_{ij} - \bar{y}_{i.}, \quad (3.10)$$

jota sanotaan jäännökseksi (residual). Se on yksilön poikkeama käsittelyryhmänsä keskiarvosta. Kaavan (3.1) mukaan havaintoarvoissa y_{ij} esiintyvä vaihtelu on siis peräisin sekä käsittelyiden välisistä eroista (τ_i -parametrien vaihtelusta) että satunnaisvaihtelusta (ϵ_{ij} -termien vaihtelusta). Nämä kaksi vaihtelulähdettä ovat toisistaan riippumattomia.

3.3 Yksisuuntaisen varianssianalyysin konstruointi

Yksinkertaisimpana hypoteesina on, että kaikkien käsittelyiden keskiarvot ovat samoja.

Toisin sanoen

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_a$$

$$H_1 : \mu_i \neq \mu_j \quad \text{joillekin } i, j.$$

Saman voi ilmaista myös käsittelyvaikutusten avulla

$$H_0 : \tau_1 = \tau_2 = \dots = \tau_a = 0$$

$$H_1 : \tau_i \neq 0 \quad \text{ainakin yhdelle } i.$$

Kokonaisvaihtelun osittamiseksi jaetaan vaihtelu komponentteihin. Kaavojen (3.7), (3.8) ja (3.10) avulla voidaan kirjoittaa

$$\begin{aligned} y_{ij} &= \hat{\mu} + \hat{\tau}_i + \hat{\epsilon}_{ij} \\ &= \bar{y}_{..} + (\bar{y}_{i.} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{i.}), \end{aligned} \quad (3.11)$$

ja edelleen

$$(y_{ij} - \bar{y}_{..}) = (\bar{y}_{i.} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{i.}). \quad (3.12)$$

Havainnon y_{ij} vaihtelu yleiskeskiarvon ympärillä voidaan osittaa ryhmäkeskiarvojen väliseen vaihteluun ja ryhmien sisäiseen vaihteluun. Korotetaan kaavasta (3.12) yhtälön

molemmat puolet neliöön ja summataan kaikkien havaintojen yli. Saadaan siis

$$\sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2 = n \sum_{i=1}^a (\bar{y}_{i.} - \bar{y}_{..})^2 + \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{i.})^2 + 2 \sum_{i=1}^a \sum_{j=1}^n (\bar{y}_{i.} - \bar{y}_{..})^2 (y_{ij} - \bar{y}_{i.})^2 . \quad (3.13)$$

Kaavan (3.13) ristitulotermin saa arvon 0, koska

$$\sum_{j=1}^n (y_{ij} - \bar{y}_{i.})^2 = y_{i.} - n \bar{y}_{i.} = y_{i.} - n(y_{i.}/n) = 0 . \quad (3.14)$$

Tällöin saadaan neliösummahajotelma

$$SS_{Tot} = SS_{Treatm} + SS_{Err} , \quad (3.15)$$

jossa

$$\begin{aligned} SS_{Tot} &= \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2 \\ SS_{Treatm} &= n \sum_{i=1}^a (\bar{y}_{i.} - \bar{y}_{..})^2 \\ SS_{Err} &= \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{i.})^2 . \end{aligned} \quad (3.16)$$

Kaavan (3.16) neliösummille saadaan seuraavat vapausasteet:

$$\begin{aligned} df_{Tot} &= N - 1 \\ df_{Treatm} &= a - 1 \\ df_{Err} &= N - a . \end{aligned}$$

Koska $df_{Tot} = df_{Treatm} + df_{Err}$, voidaan Cochranin lausetta (Montgomery, 1984; s. 50)

hyväksikäyttäen todeta, että SS_{Treatm} ja SS_{Err} ovat riippumattomia.

Keskineliösummat saadaan jakamalla kaavan (3.16) neliösummat vapausasteillaan.

Kokonaisvarianssin estimaattoriksi saadaan

$$MS_{Tot} = \frac{SS_{Tot}}{N-1} . \quad (3.17)$$

Samoin voidaan esittää ryhmäkeskiarvojen varianssiestimaattori

$$MS_{Treatm} = \frac{SS_{Treatm}}{a-1} \quad (3.18)$$

ja edelleen virhevariانسin estimaattori ($= \hat{\sigma}^2$)

$$MS_{Err} = \frac{SS_{Err}}{N-a} . \quad (3.19)$$

Keskineliöiden odotusarvoiksi saadaan (Montgomery, 1984; s. 50)

$$E(MS_{Err}) = \sigma^2 \quad (3.20)$$

$$E(MS_{Treatm}) = \sigma^2 + \frac{n \sum_{i=1}^a \tau_i^2}{a-1} . \quad (3.21)$$

Jos nollahypoteesi $H_0 : \mu_1 = \mu_2 = \dots = \mu_a$ on tosi, eli $H_0 : \tau_1 = \tau_2 = \dots = \tau_a = 0$ on voimassa, niin $E(MS_{Treatm}) = E(MS_{Err})$. Tämä on varianssianalyysin kannalta tärkeä tulos, sillä tämän suhteen avulla voidaan tarkastella nollahypoteesin voimassaoloa.

3.4 F-testisuureen johtaminen

Oletetaan, että virhetermit ϵ_{ij} :t ovat normaalisti jakautuneita ja riippumattomia odotusarvolla 0 ja varianssilla σ^2 , joten myös havainnot y_{ij} ovat normaalisti jakautuneita ja riippumattomia odotusarvolla $\mu + \tau_i$ ja varianssilla σ^2 .

Koska $df_{Tot} = df_{Treatm} + df_{Err}$, voidaan Cochranin teoreeman mukaan sanoa, että SS_{Treatm} / σ^2 ja SS_{Error} / σ^2 ovat riippumattomia χ^2 -jakautuneita muuttujia.

Todennäköisyyslaskennan mukaan kahden toisistaan riippumattoman χ^2 -jakautuneen vapausasteillaan jaetun muuttujan osamäärä on F-jakautunut. Täten voidaan esittää

F-testisuure, joka on muotoa

$$F_0 = \frac{MS_{Treatm}}{MS_{Err}} \quad (3.22)$$

ja on F-jakautunut vapausastein $a - 1$ ja $N - a$.

Jos $E(MS_{Treatm}) > \sigma^2$ eli $F_0 > 1$, niin ainakin yhdessä ryhmässä keskiarvo on erisuuri kuin muissa. Jos lisäksi pätee, että $F_0 >$ kriittinen arvo eli $F_0 > F_{a, a-1, N-a}$, niin tällöin H_0 voidaan hylätä eli havaitun eron voidaan sanoa johtuvan muusta kuin satunnaisvaihtelusta.

3.5 Kaksisuuntainen varianssianalyysi

Yksisuuntaiseen varianssianalyysiin verrattuna kaksisuuntaisessa varianssianalyysissä on jokaisen muuttujan omavaikutuksen lisäksi tutkittavana myös muuttujien yhdysvaikutus eli niinsanottu interaktio. Kun yksisuuntaisessa varianssianalyysissä tarkasteltiin yhtä F-arvoa ja sen p-arvoa, on tarkastelun kohteena nyt useita F-arvoja p-arvoineen. Tällöin voidaan tutkia kunkin tekijän omavaikutuksen lisäksi myös näiden yhdysvaikutuksen merkitsevyyttä mallissa.

Oletuksena on siis tilanne, jossa malliin kuuluu kaksi faktoria (koetekijää). Faktorilla A on p ja faktorilla B q kappaletta tasoja. Lisäksi merkitään toistojen määrää n :llä ja täten havaintojen kokonaismäärää voidaan merkitä $N = pqn$. Tällöin merkintä y_{ijk} tarkoittaa siis tulomuuttujan y arvoa A:n tasolla i ja B:n tasolla j , kun kyseessä on toisto k . Tällöin varianssianalyysin mukainen malli voidaan esittää muodossa

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk} \quad (3.23)$$

missä

- μ = tulomuuttujan y odotusarvo
- α_i = faktorin A käsittelyn i vaikutus y :n odotusarvoon
- β_j = faktorin B käsittelyn j vaikutus y :n odotusarvoon
- $(\alpha\beta)_{ij}$ = faktorin A käsittelyn i ja faktorin B käsittelyn j yhdysvaikutus y :n odotusarvoon
- ϵ_{ijk} = satunnaisvirhe

ja lisäksi $i = A:n$ tasot eli $i = 1, 2, \dots, p$
 $j = B:n$ tasot eli $j = 1, 2, \dots, q$
 $k =$ toistojen määrä eli $k = 1, 2, \dots, n$.

Lisäksi pätee

$$\sum_{i=1}^p \alpha_i = \sum_{j=1}^q \beta_j = \sum_{i=1}^p (\alpha\beta)_{ij} = \sum_{j=1}^q (\alpha\beta)_{ij} = 0 . \quad (3.24)$$

Termejä α_i ja β_j kutsutaan päävaikutustermeiksi ja termiä $(\alpha\beta)_{ij}$ vastaavasti yhdysvaikutustermiksi.

Kappaleessa 3.3 esitettyjen kaavojen (3.15) ja (3.16) idea säilyy myös kaksisuuntaisessa varianssianalyysissä. Nyt vastaavissa kaavoissa näkyy myös se, että mukaan tulee sekä A:n että B:n omavaikutukset, näiden interaktio ja lisäksi toistot. Eli neliösummahajotelmaksi saadaan (Neter, 1974; s. 572):

$$SS_{Tot} = SS_{Treatm} + SS_{Err} , \quad (3.25)$$

jossa

$$\begin{aligned} SS_{Tot} &= \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^n (y_{ijk} - \bar{y}_{...})^2 \\ SS_{Treatm} &= n \sum_{i=1}^p \sum_{j=1}^q (\bar{y}_{ij.} - \bar{y}_{...})^2 \\ SS_{Err} &= \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^n (y_{ijk} - \bar{y}_{ij.})^2 . \end{aligned} \quad (3.26)$$

Keskineliösummat termeille SS_{Treatm} ja SS_{Err} ovat

$$\begin{aligned} MS_{Treatm} &= \frac{SS_{Treatm}}{pq - 1} \\ MS_{Err} &= \frac{SS_{Err}}{pq(n - 1)} . \end{aligned} \quad (3.27)$$

Lisäksi voidaan yhtälöryhmän (3.26) keskimäinen termi jakaa vielä osiin. Eli saadaan:

$$SS_{Treatm} = SS_A + SS_B + SS_{AB} , \quad (3.28)$$

jossa

$$\begin{aligned} SS_A &= nq \sum_{i=1}^p (\bar{y}_{i..} - \bar{y}_{...})^2 \\ SS_B &= np \sum_{j=1}^q (\bar{y}_{.j.} - \bar{y}_{...})^2 \\ SS_{AB} &= n \sum_{i=1}^p \sum_{j=1}^q (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2 . \end{aligned} \quad (3.29)$$

Yhtälöryhmän (3.29) yhtälöille saadaan seuraavat keskineliösummat

$$\begin{aligned} MS_A &= \frac{SS_A}{p-1} \\ MS_B &= \frac{SS_B}{q-1} \\ MS_{AB} &= \frac{SS_{AB}}{(p-1)(q-1)} . \end{aligned} \quad (3.30)$$

Pienimmän neliösumman estimaattoreiksi kaavan (3.23) termeille saadaan (Neter, 1974; s. 571) minimoimalla

$$Q = \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^n (y_{ijk} - \mu - \alpha_i - \beta_j - (\alpha\beta)_{ij})^2 . \quad (3.31)$$

Kun otetaan huomioon kaavan (3.24) rajoitukset saadaan parametrien estimaateiksi

$$\hat{\mu} = \bar{y}_{...} \quad (3.32)$$

$$\hat{\alpha}_i = \bar{y}_{i..} - \bar{y}_{...} \quad (3.33)$$

$$\hat{\beta}_j = \bar{y}_{.j} - \bar{y}_{...} \quad (3.34)$$

$$(\hat{\alpha}\hat{\beta})_{ij} = \bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...} \quad (3.35)$$

Sijoittamalla nämä tulokset kaavaan (3.23) saadaan virhetermin ϵ_{ijk} 'estimaattoriksi'

$$\hat{\epsilon}_{ijk} = y_{ijk} - \bar{y}_{ij.}, \quad (3.36)$$

3.6 F-testisuure kaksisuuntaiselle varianssianalyysille

Kaksisuuntaisen varianssianalyysin F-testissä voidaan erikseen tarkastella kahden faktorin tapauksessa molempien faktoreiden omavaikutuksen sekä yhdysvaikutuksen merkitsevyyttä.

Tutkittaessa interaktion merkitsevyyttä saadaan aluksi seuraavat hypoteesit:

$$H_0 : (\alpha\beta)_{ij} = 0, \text{ kaikille } i, j$$

$$H_1 : (\alpha\beta)_{ij} \neq 0, \text{ joillekin } i, j.$$

F-testisuure saadaan nyt laskettua kaavalla (Neter, 1974; s. 579):

$$F_0 = \frac{MS_{AB}}{MS_{Err}}. \quad (3.37)$$

Nyt kun H_0 on voimassa, F_0 noudattaa F-jakaumaa vapausastein $(p - 1)(q - 1)$ ja $pq(n - 1)$. Nyt jos $F_0 >$ kriittinen arvo eli $F_0 > F_{\alpha; (p-1)(q-1), pq(n-1)}$, niin tällöin H_0 voidaan hylätä eli interaktion vaikutus on tilastollisesti merkitsevä.

Samoin voidaan tutkia päävaikutusten tilastollista merkitsevyyttä. Huomattavaa kuitenkin on, että päävaikutusten tulkintaa ei voida kunnolla tehdä, jos interaktion merkitsevyys on suuri. Hypoteesit päävaikutusten testaamiseen ovat faktorille A seuraavat:

$$H_0 : \mu_{1.} = \mu_{2.} = \dots = \mu_{a.}$$

$$H_1 : \mu_{i.} \neq \mu_{j.} \text{ joillekin } i, j.$$

Vaihtoehtoisesti voidaan hypoteesit kirjata myös muodossa:

$$H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_p = 0$$

$$H_1 : \alpha_i \neq \alpha_j, \text{ joillekin } i, j.$$

F-testisuure saadaan kaavalla:

$$F_0 = \frac{MS_A}{MS_{Err}} . \quad (3.38)$$

H_0 :n ollessa voimassa F_0 noudattaa F-jakaumaa vapausastein $(p - 1)$ ja $pq(n - 1)$. Nyt jos $F_0 >$ kriittinen arvo eli $F_0 > F_{\alpha; (p-1), pq(n-1)}$, niin tällöin H_0 voidaan hylätä eli faktorin A päävaikutus on tilastollisesti merkitsevä. Faktorille B voitaisiin tehdä sama tarkastelu. Ainoana erona F-testisuureen kaavaan olisi osoittajan MS_A :n vaihtuminen MS_B :ksi ja ensimmäisen vapausasteen muuttuminen $(p - 1)$:stä $(q - 1)$:ksi.

4 Yleinen lineaarinen malli

Luvussa 3 esitellyt varianssianalyttiset mallit voidaan helposti kuvata yleisenä lineaarisena mallina. Ne ovat itse asiassa erikoistapauksia lineaarisesta mallista, joka voidaan kuvata matriisimerkinnöin seuraavasti (Horton, 1986; s. 9)

$$Y = X\Theta + E, \quad (4.1)$$

jossa Y = vastemuuttuja (dimensio $n \times 1$)
 X = mallin muuttujien kertoimet (dimensio $n \times p$)
 Θ = mallin muuttujat (dimensio $p \times 1$)
 E = mallin virhetermit (dimensio $n \times 1$).

Nyt saadaan Θ :n estimaatiksi (Horton, 1986; s. 11)

$$\hat{\Theta} = (X'X)^{-1}X'Y. \quad (4.2)$$

Esimerkki 4.1 Varianssianalyttinen mallin esittäminen lineaarisena mallina

Oletetaan, että varianssianalyttinen malli on muotoa

$$y_i = \mu + \alpha_i + \epsilon_i, \quad i = 1, 2, 3,$$

jolloin se saadaan purettua kolmeksi erilliseksi yhtälöksi:

$$\begin{aligned} y_1 &= \mu + \alpha_1 + \epsilon_1 \\ y_2 &= \mu + \alpha_2 + \epsilon_2 \\ y_3 &= \mu + \alpha_3 + \epsilon_3. \end{aligned} \quad (4.3)$$

Yhtälöryhmä (4.3) saadaan kirjoitettua matriisimerkinnöin kaavan (4.1) mukaisesti, kun valitaan matriisit seuraavasti:

$$Y_{3 \times 1} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}, \quad X_{3 \times 4} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}, \quad \Theta_{4 \times 1} = \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}, \quad E_{3 \times 1} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{pmatrix}.$$

4.1 Lineaarisista rajoitteista

Lineaarisia rajoitteita käsitellään jo tässä vaiheessa lukua, jotta niihin voitaisiin jatkossa viitata. Lineaaristen rajoitteiden käyttö on yleisen lineaarisen mallin eräs erikoistapaus. Tutkijalla pitäisi olla jonkinlainen ennakkotieto siitä, mitkä tasot mahdollisesti poikkeavat toisistaan. Jos näin on, voi tutkija asettaa niinsanottuja kontrasteja, joiden avulla tutkitaan määrättyjen ryhmien keskiarvojen eroavaisuuksia. Toisaalta jos esimerkiksi varianssianalyysin mukaan tasoilla on eroja, käytetään kontrasteja näiden erojen tarkempaan tutkimiseen. Näin siis erona esimerkiksi varianssianalyysiin on se, että kontrastien avulla tutkija vertaa tiettyjä ennakkoon määrättyjä ryhmiä keskenään ja jos tehdyt analyysit puoltavat keskiarvojen eroja, voi tutkija päätellä ryhmät, joiden välillä eroja on.

Kontrasti on käsittelyiden odotusarvojen lineaarikombinaatio:

$$C = \sum_{i=1}^a c_i \mu_i, \quad (4.4)$$

missä kertoimille c_i pätee

$$\sum_{i=1}^a c_i = 0. \quad (4.5)$$

Jos merkitään $c^T = [c_1 \ c_2 \ c_3 \ \dots \ c_a]$ ja vastaavasti $\mu = [\mu_1 \ \mu_2 \ \mu_3 \ \dots \ \mu_a]$ saadaan kaava (4.4) esitettyä yksinkertaisesti muodossa $C = c^T \mu$.

Jos tutkitaan vaikkapa neljän ryhmän käyttäytymistä, saadaan esimerkiksi taulukossa 4.1 esiintyvät lineaariset rajoitteet:

Taulukko 4.1 Esimerkkejä lineaarisista rajoitteista.

LINEAARINEN RAJOITE

RYHMÄ	1	2	3	4	5	6
1	1	0	1	-1	-1	1
2	-1	0	1	-1	3	1
3	0	-1	0	2	-1	-1
4	0	1	-2	0	-1	-1

Nyt rajoite 1 vertaa ryhmien 1 ja 2 keskiarvoja toisiinsa

rajoite 2 “ ” 3 ja 4 keskiarvoja toisiinsa

rajoite 3 “ ” 1 ja 2 keskiarvoa ryhmän 4 keskiarvoon

rajoite 4 “ ” 1 ja 2 keskiarvoa ryhmän 3 keskiarvoon

rajoite 5 “ ” 1, 3 ja 4 keskiarvoa ryhmän 2 keskiarvoon

rajoite 6 “ ” 1 ja 2 keskiarvoa ryhmien 3 ja 4 keskiarvoon.

Testattaessa hypoteeseja rajoitteiden avulla saadaan esimerkiksi yllä olevasta rajoitteesta numero 4 seuraavanlainen esitystapa:

$$H_0: -\mu_1 - \mu_2 + 2\mu_3 = 0$$

$$H_1: -\mu_1 - \mu_2 + 2\mu_3 \neq 0$$

Eli tässä $c^T = [-1 \ -1 \ 2]$ ja H_0 :n mukaan $c^T \mu = 0$.

Toisessa lähestymistavassa voidaan verrata tiettyä kontrollitasoa muihin. Jos ryhmiä on a kappaletta ja valitsemalla taso $i = 1$ kontrolliryhmäksi, niin muiden vertailu tähän voidaan esittää $a - 1$ lineaarisen rajoitteen $\mu_i - \mu_1 = 0$, jossa $i = 2, 3, \dots, a$ avulla.

Kontrastin arvosta voidaan sanoa seuraavaa:

Jos $C = 0$, niin ryhmien keskiarvoilla ei ole eroa

Jos $C \neq 0$, niin ryhmien keskiarvot eroavat.

Kontrasteja voidaan testata t-testillä. Aluksi asetetaan hypoteesit:

$$H_0 : C = \sum_{i=1}^a c_i \mu_i = 0$$

$$H_1 : C = \sum_{i=1}^a c_i \mu_i \neq 0 .$$

Havaintoaineistosta lasketaan kontrastin arvo kaavalla

$$\hat{C} = \sum_{i=1}^a c_i \bar{y}_i . \quad (4.6)$$

Tämän keskivirhe saadaan kaavalla

$$s_{\hat{C}} = \sqrt{MS_{Err} \sum_{i=1}^a \frac{c_i^2}{n_i}} . \quad (4.7)$$

Tässä MS_{Err} on sama kuin kappaleessa 3.3 esitelty. Yllä olevien avulla saadaan t-testisuure, joka on

$$t = \frac{\hat{C}}{s_{\hat{C}}} \sim t(df_{Err}) . \quad (4.8)$$

F-testisuuretta varten lasketaan ensin kontrastia C vastaavasta keskiarvoerosta johtuva vaihtelu SS_C :

$$SS_C = \frac{\hat{C}^2}{\sum_{i=1}^a \frac{c_i^2}{n_i}} . \quad (4.9)$$

Vapausasteita on 1, joten

$$MS_C = \frac{SS_C}{1} = SS_C . \quad (4.10)$$

Tästä saadaan F-testisuurelle arvo

$$F = \frac{MS_C}{MS_{Err}} \sim F(1, df_{Err}) . \quad (4.11)$$

Huomattavaa on, että $F = t^2$.

4.2 Yleinen lineaarinen hypoteesi

Yleisen lineaarisen mallin nollahypoteesi voidaan esittää muodostuvan siten, että malliin asetetaan q kappaletta lineaarisia rajoitteita (Yandell, 1997; s. 202). Tällöin muodostuu $q \times p$ matriisi C , jolloin nollahypoteesi voidaan kirjoittaa muodossa

$$H_0 : C\Theta = 0 .$$

Nyt Θ on $p \times 1$ -matriisi, joka koostuu p :stä malliin tulevasta muuttujasta.

Edellä mainitulla nollahypoteesilla voidaan lyhyesti merkitä esimerkiksi yksisuuntaisen varianssianalyysin nollahypoteesia

$$H_0 : \tau_i = 0$$

siten, että valitaan

$$C_{3 \times 4} = \begin{vmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix} \quad ja \quad \Theta_{4 \times 1} = \begin{vmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \end{vmatrix} ,$$

jolloin päästään vastaavaan nollahypoteesiin

$$H_0 : C\Theta = 0 .$$

Otetaan käyttöön merkintä $R(C\Theta) = (C\hat{\Theta})^T [C(X^T X)^{-1} C^T]^{-1} C\hat{\Theta}$, jolloin (Yandell, 1997; s. 202):

$$F = \frac{R(C\Theta) / q}{\sigma^2} \sim F(q, N-a; \delta^2) , \quad (4.12)$$

jossa epäkeskeisyysparametri δ^2 on muotoa

$$\delta^2 = (C\Theta)^T [\sigma^2 C(X^T X)^{-1} C^T]^{-1} C\Theta . \quad (4.13)$$

Huomattavaa on, että epäkeskeisyysparametri δ^2 häviää, jos nollahypoteesi H_0 on voimassa.

5 Toistojen määrän arviointi käsittelyvaikutusten avulla

Jokaisessa koesuunnittelussa tärkeä vaihe on toistojen määrän valinta. Jos halutaan havaita pieniä ryhmien välisiä eroja, tarvitaan enemmän toistoja kuin jos tyydyttäisiin suurten erojen etsimiseen.

Toistojen määrään vaikuttaa yhtenä osana tutkittavan muuttujan varianssi. Lisäksi otoskokoon vaikuttaa valittu koeasetelma. Samaa muuttujaa tutkittaessa täysin samoilla tekijöillä päästään koeasetelmaa muuttamalla erilaisiin tuloksiin. Kriteerinä toistojen määrän arvioimiseen on testin voimakkuus, joka kertoo todennäköisyyden $P(\text{Hylätään } H_0 \mid H_0 \text{ ei voimassa})$. Lisäksi kiinnitetään testin merkitsevyytaso α , joka on yleensä 0.05 tai 0.01.

5.1 OC-käyrät

Toisen lajin virheeksi sanotaan tilannetta, jossa H_0 hyväksytään vaikka se on epätosi. Tämän todennäköisyyttä merkitään β :lla ja sen toivotaan olevan mahdollisimman pieni. Määritellään 2. lajin virhe formaalisti:

$$\begin{aligned}\beta &= 1 - P(\text{Hylätään } H_0 \mid H_0 \text{ ei voimassa}) \\ &= 1 - P(F > F_{\alpha}(a-1, N-a) \mid H_0 \text{ ei voimassa}) .\end{aligned}$$

Testin voimakkuudeksi sanotaan todennäköisyyttä $1 - \beta$. Voidaan osoittaa, että jos H_0 ei ole voimassa, niin testisuure $F_0 = MS_{\text{Treatm}} / MS_{\text{Err}}$ noudattaa epäkeskeistä F-jakaumaa vapausasteilla $a - 1$ ja $N - a$ sekä epäkeskeisyysparametrilla δ . Jos $\delta = 0$, niin jakaumaksi tulee tavallinen F-jakauma.

OC-käyrien (operating characteristics) avulla tehdyissä otoskokokartoissa kuvataan 2. lajin virheen todennäköisyyttä β tietyille toistojen määrälle parametria Φ vastaan. Näiden

käyrien avulla tutkija voi arvioida toistojen lukumäärää siten, että kokeen avulla saadaan esiin potentiaalisia eroja käsittelyiden välillä. Lisäksi oletuksena on, että havaintojen määrät eri ryhmissä ovat yhtäsuuret.

Määritellään

$$\Phi^2 = \frac{1}{a} \sum_{i=1}^a \tau_i^2 \cdot n \quad (5.1)$$

missä a = ryhmien lukumäärä

τ_i = käsittelynvaikutus

σ^2 = varianssi.

Φ^2 on eräänlainen signaali-kohina-suhde. Eli se kertoo käsittelyiden vaihtelun suhteen satunnaisvaihteluun. Jos siis eri käsittelyillä ei ole eroa yleiskeskisarvoon (eli $\tau_i = 0$), saadaan Φ :n arvoksi 0.

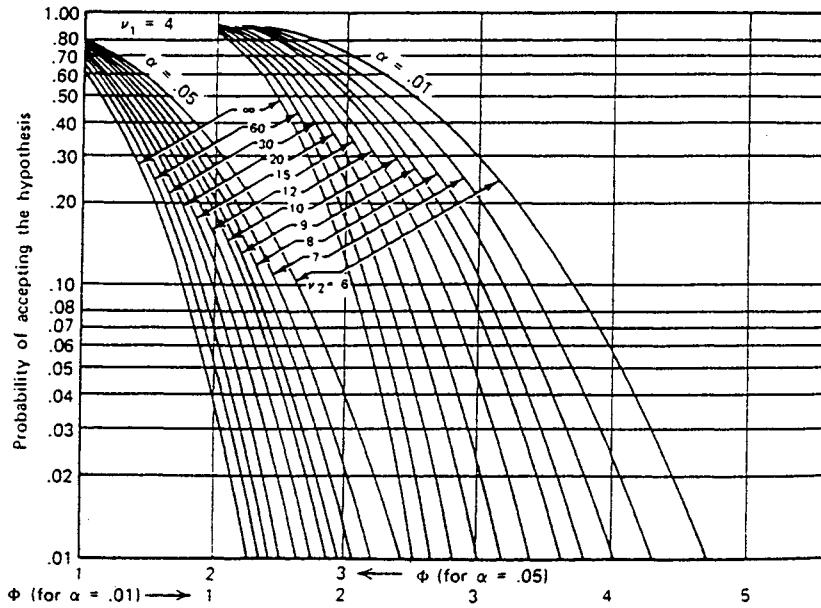
Nyt jos tiedetään tasojen lukumäärä a , tutkittavan muuttujan varianssi σ^2 ja käsittelyvaikutus τ_i^2 , päästään esimerkiksi muotoon $\Phi^2 = 1.3 \cdot n$, jolloin voidaan eri n :n arvoilla laskea Φ :lle arvoja ja katsoa OC-käyrän avulla laaditusta otoskokokartasta saavutettava voimakkuus. Tässä menetelmässä tarvitaan siis tietoa $\sum \tau_i^2$:sta ja σ^2 :stä.

Esimerkki 5.1 Toistojen määrän arvioiminen OC-käyrän avulla tehdystä otoskokokartasta
Oletetaan tilanne, jossa ryhmiä on viisi eli $a = 5$, merkitsevyystaso $\alpha = .01$ ja $\sum \tau_i^2 = 50$.
Lisäksi tutkijalla on aiempaa tietoa siitä, että keskihajonta on suurimmillaan 3 yksikköä eli $\sigma^2 = 9$. Käytetään vapausasteita $a - 1 = 5 - 1 = 4$ ja $N - a = a(n - 1) = 5(n - 1)$. Kriteerinä pidetään voimakkuuden arvoa 0.90.

Lasketaan Φ^2 seuraavasti:

$$\Phi^2 = \frac{n \sum_{i=1}^a \tau_i^2}{a \sigma^2} = \frac{n(50)}{5(3)^2} = 1.11n \quad (5.2)$$

Tämän avulla voidaan laatia taulukko 5.1 ja arvioida tarvittava toistojen määrä kuvasta 5.1.



Kuva 5.1 OC-käyrän avulla saatu otoskokokartta (Montgomery, 1984; s. 516)

Valitaan kuvasta 5.1 ne viivat, jotka kuvaavat tilannetta, jossa $\alpha = .01$. Kaavalla (5.2) saadaan esimerkiksi n :n arvolla 4 Φ :n arvoksi 2.11 . Koska $v_2 = 5(n - 1) = 5(4 - 1) = 15$, etsitään kuvasta vastaava käyrä, jossa $v_2 = 15$ ja tutkitaan käyrän arvo y -akselia vasten, kun x -akselilla on arvo 2.11 . Se on silmämääräisesti noin 0.30 eli voimakkuudeksi saataisiin $1 - 0.30 = 0.70$. Näin jatkamalla saadaan taulukko 5.1 täytettyä.

Taulukko 5.1 Arvot voimakkuuden määrittäminen

n	Φ^2	Φ	$a(n-1)$	β	$1-\beta$
4	4.44	2.11	15	.30	.70
5	5.55	2.36	20	.15	.85
6	6.66	2.58	25	.04	.96

Havaitaan, että kriteerinä annettu haluttu voimakkuus (eli tässä 0.90) saavutetaan, kun toistojen määrä on vähintään 6 .

Ongelmana edellisessä lähestymistavassa on valita käsittelyiden vaikutuksen arvo tutkittavan muuttujan odotusarvoon eli τ_i^2 , jonka mukaan Φ^2 -arvo lasketaan. Käytännössä tutkijalla täytyisi siis olla jonkinlainen ennakkotieto keskiarvojen poikkeavuuden neliöistä. Vaihtoehtona tähän on valita suurin mahdollinen minkä tahansa kahden käsittelyn välinen ero. Jos ero minkä tahansa kahden käsittelyn välillä on suurimmillaan D , niin minimissään

$$\Phi^2 = \frac{nD^2}{2a\sigma^2} . \quad (5.3)$$

Minimistä johtuen saadaan konservatiivinen $n:n$ arvo, toisin sanoen saadaan $n:n$ arvo, joka ainakin vastaa annettua tehokkuutta $1 - \beta$.

6 Epäkeskeinen F-jakauma

Epäkeskeiseen F-jakaumaan päädytään silloin, kun nollahypoteesi ei ole voimassa. Luvun tavoitteena on selvittää epäkeskeisyysparametrin olemus ja sen vaikutus lähinnä F-testiin. Koska kaikki nykyiset alaan liittyvät ohjelmistot eivät osaa käsitellä epäkeskeistä F-jakaumaa, niin luvussa esitetään myös keino approksimoida epäkeskeinen F-jakauma tavallisen F-jakauman avulla. Tällä keinolla päästään yleensä tarpeeksi tarkkoihin tuloksiin.

6.1 F-testisuure epäkeskeisyysparametrin alaisessa tilanteessa

Lähdetään liikkeelle tilanteesta, jossa $X = \sum_{i=1}^n x_i^2$ ja $Y = \sum_{j=1}^m y_j^2$. Lisäksi oletetaan, että

$x_i \sim N(0,1)$, $y_j \sim N(0,1)$ ja muuttujat x_i ja y_j ovat riippumattomia. Nyt $X \sim \chi^2(n)$ ja samoin $Y \sim \chi^2(m)$. Tällöin suurelle F pätee (Anderson, 1958; 5.4.1):

$$F = \frac{X / n}{Y / m} \sim F(n, m). \quad (6.1)$$

Laajennetaan tilannetta seuraavasti: Olkoot $Y = (y_1, y_2, \dots, y_p) \sim N(\mathbf{v}, \mathbf{I})$, jossa $\mathbf{v} = (v_1, v_2, \dots, v_p)$. Merkitään nyt $V = Y^T Y = \sum_{i=1}^p y_i^2$, jossa $i = 1, 2, \dots, p$ ja lisäksi määritellään epäkeskeisyysparametri

$$\lambda = \sum_{i=1}^p v_i^2. \quad (6.2)$$

Nyt $V \sim \chi^2(p, \lambda)$ ja jos merkitään $W \sim \chi^2(m)$ saadaan:

$$F = \frac{V / p}{W / m} \sim F(p, m; \lambda). \quad (6.3)$$

6.2 Voimakkuus F-jakauman avulla

Valitaan $H_0 : \tau_i = 0$ (, jossa τ_i on ryhmän i käsittelyvaikutus) ja $H_1 : \tau_i \neq 0$ jollekin i . Lisäksi merkitään $\alpha = P(H_0 \text{ hylätään} \mid H_0 \text{ voimassa})$, jolloin $\beta = P(H_0 \text{ hyväksytään} \mid H_0 \text{ ei voimassa})$ eli $\beta = 1 - P(H_0 \text{ hylätään} \mid H_0 \text{ ei voimassa})$. Nyt saadaan testin voimakkuudeksi (Yandell, 1997: s 83) $1 - \beta = P(H_0 \text{ hylätään} \mid H_0 \text{ ei voimassa})$. Täten $\alpha = P(F > F_\alpha(a-1, a(n-1)))$ ja tilanteessa, jossa H_0 ei ole voimassa saadaan testin voimakkuudeksi:

$$1 - \beta = P(F \geq F_\alpha(a-1, a(n-1)); \lambda), \quad (6.4)$$

jossa niin sanottu epäkeskeisyysparametri on muotoa:

$$\lambda = \frac{\sum_{i=1}^a \tau_i^2}{\sigma^2} \cdot n. \quad (6.5)$$

6.3 Epäkeskeisen F-jakauman laskeminen tavallisen F-jakauman avulla

Määritellään aluksi

$$F' = \frac{z_1/v_1}{z_2/v_2}, \quad (6.6)$$

jossa $z_1 \sim \chi^2(v_1, \lambda)$ ja $z_2 \sim \chi^2(v_2)$. Eli $F' \sim (v_1, v_2; \lambda)$.

Nyt jos $z_1 \sim \chi^2(v_1, \lambda)$, niin (Kendall, 1963 s. 867) saadaan

$$z_1 / \frac{v_1 + 2\lambda}{v_1 + \lambda} \sim \chi^2(v^*) \quad , \text{ jossa } v^* = \frac{(v_1 + \lambda)^2}{v_1 + 2\lambda} . \quad (6.7)$$

Tästä saadaan

$$F = \frac{\chi^2(v^*) / v^*}{\chi^2(v_2) / v_2} \sim F(v^*, v_2) , \quad (6.8)$$

joka voidaan kirjoittaa muodossa:

$$\begin{aligned} F &= \frac{\frac{z_1}{\left(\frac{v_1 + 2\lambda}{v_1 + \lambda}\right)} / \frac{(v_1 + \lambda)^2}{v_1 + 2\lambda}}{z_2 / v_2} = \frac{z_1 \left(\frac{v_1 + \lambda}{v_1 + 2\lambda} \cdot \frac{v_1 + 2\lambda}{(v_1 + \lambda)^2}\right)}{z_2 / v_2} \\ &= \frac{z_1 / (v_1 + \lambda)}{z_2 / v_2} = \frac{z_1 / v_1}{z_2 / v_2} \cdot \frac{v_1}{v_1 + \lambda} = F' \frac{v_1}{v_1 + \lambda} \end{aligned} \quad (6.9)$$

eli

$$F' = \frac{v_1 + \lambda}{v_1} F , \quad (6.10)$$

jossa $F \sim F(v^*, v_2)$.

Tämä approksimaatio (Kendall, 1961; s. 867) on riittävän tarkka, koska kaksi ensimmäistä momenttia ovat täysin samat oikeaan arvoon verrattuna.

Eli käytännössä epäkeskeinen F-jakauma lasketaan seuraavasti:

1.) Lasketaan tavallisesta F-jakaumasta piste $x = F_\alpha(v_1, v_2)$.

2.) Lasketaan uusi piste $x' = x \frac{v_1}{v_1 + \lambda}$.

3.) Lasketaan $\beta = F(x' ; \frac{(v_1 + \lambda)^2}{v_1 + 2\lambda}, v_2)$.

4.) Nyt voidaan laskea voimakkuus $1 - \beta$.

Huomattavaa on, että voimakkuus riippuu myös valitusta n :stä λ :n kautta. Eli toistojen määrää muuttelemalla voidaan vaikuttaa saatavaan testin voimakkuuteen.

7 Toistojen määrän arviointi regressiomallin avulla

Kuten varianssianalyttiset mallit, niin myös regressiomalli on lineaarisen mallin yksi erikoistapaus. Eli myös se voidaan kuvata yleisenä lineaarisena mallina kaavan (4.1) mukaan. Jos oletetaan, että regressiomalli olisi muotoa

$$\begin{aligned}y_1 &= \beta_0 + x_1 \cdot \beta_1 + \epsilon_1 \\y_2 &= \beta_0 + x_2 \cdot \beta_1 + \epsilon_2 ,\end{aligned}\tag{7.1}$$

voidaan se kirjoittaa lyhyemmin matriisimerkinnöin

$$Y = XB + \epsilon ,\tag{7.2}$$

missä

$$Y_{2 \times 1} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, X_{2 \times 2} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \end{bmatrix}, B_{2 \times 1} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \text{ ja } \epsilon_{2 \times 1} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix}.$$

Merkitään B:n estimaattia b:llä, jolloin saadaan (Draper, 1966; s. 53)

$$b = (X'X)^{-1} X'Y.\tag{7.3}$$

joka vastaa täsmälleen kappaleessa 4 esitettyä kaavaa (4.2).

Yleensä regressiomallissa on jatkuvia muuttujia. Cultor Oy :ssä käytetään kuitenkin regressiomalleissa muuttujia, jotka ovat luonteeltaan diskreettejä eli epäjatkuvia. Yleensä näillä muuttujilla on 1:stä n:ään kappaletta tasoja, joissa muuttujat saavat erilaisia arvoja. Edellä mainitut seikat ovat syynä siihen, että tässä luvussa esiteltävät asiat tulevat olemaan hieman poikkeavia verrattuna yleiseen regressiomallin käsittelyyn.

7.1 Siirtyminen regressiomallista lineaariseen malliin

Kahden muuttujan tapauksessa regressiomallin voidaan olettaa koostuvan seuraavasti:

$Y = \text{vakio} + \text{kerroin}_1 \cdot \text{muuttuja}_1 + \text{kerroin}_2 \cdot \text{muuttuja}_2 + \text{kerroin}_3 \cdot \text{interaktiotermi}$.

Oletetaan, että tutkittavana on tilanne, jossa malli on:

$$Y = 10 \cdot C + 1.3 \cdot A + 1.7 \cdot B + 0.9 \cdot AB, \quad (7.4)$$

jossa $Y = \text{tulosmuuttuja}$

$C = \text{vakio}$

$A = \text{malliin valittu diskreetti muuttuja, jolla on 3 tasoa}$

$B = \text{malliin valittu diskreetti muuttuja, jolla on 2 tasoa}$

$AB = \text{malliin tuleva interaktiotermi}$.

Edellä mainitun A:n 3 tasoa koostuvat siten, että

1. tason numeerinen arvo on 0,
2. tason numeerinen arvo on 5.3 ja
3. tason numeerinen arvo on 8.2 .

B:n kahdelle tasolle vastaavat numeeriset arvot ovat 0 ja 2.5 .

Merkitään seuraavassa i :llä A:n käsiteltävää tasoa ja j :llä vastaavasti B:n tasoa. Tällöin siis esimerkiksi tulosmuuttujan Y merkintä y_{12} tarkoittaa tilannetta, jossa mallissa on A:n taso 1 ja B:n taso 2. Näin ollen erilaisia kombinaatioita tulee yhteensä $2 \cdot 3 = 6$ kappaletta. Lisäksi on huomioitava, että esimerkiksi tilanteessa, jossa $i=3$ ja $j=1$ tulee malliin mukaan yhdysvaikutustermi, joka koostuu A:n tason 3 numeerisesta arvosta kerrottuna B:n tason 1 numeerisella arvolla. Tätä menetelmää käyttäen saadaan seuraava kuudesta yhtälöstä koostuva yhtälöryhmä:

$$\begin{aligned}
y_{11} &= 10 + 1.3 \cdot 0 + 1.7 \cdot 0 + 0.9 \cdot 0 \cdot 0 \\
y_{12} &= 10 + 1.3 \cdot 0 + 1.7 \cdot 2.5 + 0.9 \cdot 0 \cdot 2.5 \\
y_{21} &= 10 + 1.3 \cdot 5.3 + 1.7 \cdot 0 + 0.9 \cdot 5.3 \cdot 0 \\
y_{22} &= 10 + 1.3 \cdot 5.3 + 1.7 \cdot 2.5 + 0.9 \cdot 5.3 \cdot 2.5 \\
y_{31} &= 10 + 1.3 \cdot 8.2 + 1.7 \cdot 0 + 0.9 \cdot 8.2 \cdot 0 \\
y_{32} &= 10 + 1.3 \cdot 8.2 + 1.7 \cdot 2.5 + 0.9 \cdot 8.2 \cdot 2.50
\end{aligned} \tag{7.5}$$

Nyt voidaan siirtyä kaavan (4.1) mukaisiin merkintöihin. Valitaan matriisit X ja Θ siten, että

X = koeasetelmasta poimitut tasojen numeeriset arvot eli yhtälöryhmästä (7.5) otetut arvot

Θ = muuttujien kertoimet, jotka ovat nähtävissä kaavoissa (7.4) ja (7.5).

Eli tämän perusteella saadaan:

$$X_{6 \times 4} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 2.5 & 0 \\ 1 & 5.3 & 0 & 0 \\ 1 & 5.3 & 2.5 & 13.2 \\ 1 & 8.2 & 0 & 0 \\ 1 & 8.2 & 2.5 & 20.5 \end{pmatrix},$$

jossa ensimmäinen rivi kertoo, että tulosmuuttujaan y_{11} tulee mukaan A:n numeerinen arvo 0 eli tason 1 arvo ja vastaavasti B:n numeerinen arvo 0 eli tason 1 arvo. Ensimmäisen rivin viimeinen arvo on saatu kertomalla edelliset keskenään eli $0 \cdot 0 = 0$. Matriisi Θ on pystyvektori, joka koostuu mallissa olevien termien kertoimista. Se on

$$\Theta_{1 \times 4}^T = [10 \quad 1.3 \quad 1.7 \quad 0.9].$$

Matriisi C valitaan sen mukaan, mistä termistä tai termeistä olemme tulevan testaamisen kannalta kiinnostuneita. Eli jos haluamme tietää A:n kertoimen tilastollista merkitsevyyttä, valitsemme matriisiin C siten, että

$$C_{1 \times 4} = [0 \quad 1 \quad 0 \quad 0].$$

Tällöin matriisi C 'poimii' Θ :stä juuri A:n. Nyt voidaan laskea kaavalla (4.13) epäkeskeisyysparametri δ^2 :

$$\delta^2 = (C\Theta)^T \cdot [\sigma^2 \cdot C(X^T X)^{-1} \cdot C^T]^{-1} \cdot C\Theta .$$

Jos Θ :n dimensio on $p \times 1$, C:n $q \times p$ ja X:n $r \times p$, niin vapausasteiksi F-testiin saamme $(q, n \cdot r - p)$, jossa $n =$ toistojen määrä. Eli esimerkkitapauksessamme saataisiin, että β :n dimensio on 4×1 , C:n 1×4 ja X:n 6×4 . Jos lisäksi oletamme, että toistoja on vain yksi eli $n = 1$ ja varianssi $\sigma^2 = 4$, niin saamme epäkeskeisyysparametriksi $\delta^2 = 864.5$. Näiden avulla voidaan laskea testin voimakkuus, joka on muotoa:

$$1 - \beta = F_{\alpha}(1, 2; 864.5) .$$

Nyt voimme asettaa tietyn merkitsevyytason α ja mahdollisesti muuttella toistojen määrää eli n :ää ja katsoa minkälaisiin voimakkuuksiin päädytään.

Laskettuja arvoja voidaan tietenkin tulkita myös toisin päin eli voidaan kysyä: "Mikä on se toistojen määrä, jolla päästään tiettyyn voimakkuuteen?". Tällöin tekniikkana on laskea eri n :n arvoilla niin kauan kunnes ollaan päästy annetun kriteerin ylittävään voimakkuuteen.

8 Sovellus toistojen määrän ja voimakkuuden laskemiseen

Kuten alussa oli mainittu, kaiken edellä esitetyn perusteella on valmistunut Excel-pohjainen sovellus, joka tekee kaiken laskennallisen puolen käyttäjän antamien tietojen mukaan. Ohjelmalle on annettu nimi REPLIC. Se vaatii toimiakseen vähintään Excel 5.0 -version. Ohjelmaan voi tutustua lähemmin lukemalla läpi käyttöohjeet, jotka ovat liitteessä 3. Ohjelman koodi on liitteessä 1 ja 2. Lisäksi liitteenä 3 olevasta sovelluksen käyttöohjeesta löytyy esimerkki, jolla valaistaan sekä käyttöliittymän että itse laskennan eri vaiheita.

8.1 Yhteenveto

Aiemmissa luvuissa on kirjattu pala kerrallaan sovelluksessa käytettyä teoriaa. Lisänä on ollut muutamia esimerkkejä, jotka toivottavasti auttavat asian hahmottamisessa. Sovelluksen käyttämiseen tarvitaan aluksi tiedot tehtävästä kokeesta eli malliin tulevat muuttujat nimineen, luokkineen, arvoineen ja kertoimineen. Lisäksi sovellukseen täytyy antaa tiedot koeasetelmasta. Tutkittavan tulosmuuttujan jäännösvarianssi on tieto, jonka arvio saadaan aiempien kokeiden perusteella. Edellä mainittujen tietojen perusteella sovellus laskee tarpeen mukaan joko tarvittavan toistojen määrän tai voimakkuuden. Tästä on esimerkki liitteessä 3, joka sisältää sovelluksen käyttöoppaan.

Lähteet

- Andersson, T.W. (1958), *An Introduction to Multivariate Statistical analysis*. New York: Wiley.
- Draper, N.R. and Smith, H. (1966), *Applied Regression Analysis*. New York: Wiley.
- Horton, R.L. (1986), *The General Linear Model*. Florida: Krieger.
- Kendall, M.G. and Stuart, A. (1963), *The Advanced Theory of Statistics*. London: Griffin & Co.
- Montgomery, D.C. (1984), *Design and Analysis of Experiments*. New York: Wiley.
- Neter, J. and W. Wasserman (1974), *Applied Linear Statistical Models*. Homewood: Irwin.
- Winer, B.J. (1962), *Statistical Principles in Experimental Design*. New York: McGraw-Hill.
- Yandell, B.S (1997), *Practical Data Analysis for Designed Experiments*. London: Chapman & Hall.

Liite 1

Malliin tulevien termien kyselyyn liittyvät
funktiot ja aliohjelmat

Ask_Term_Macros

```

/ Clear_Level_Values_Dialog(EditBoxes_amount) "Clears Level_Values_Dialog"
/-----
/ Cancel_Ask_Level_Values "Cancels asking values for levels" box_amount
/-----
/ Move_Level_Values(What_Enzyme) "Moves values of levels to Level_Values-sheet
/ Clear_Level_Values_Dialog levels_of_enzyme
/-----
/ Ask_Terms "Asks terms which are involved"
/ Next_Row canc_term_selection
/ Ask_coefficients more_enzymes_to_come
/ ** SHOWS DIALOG: "Term_Selection_Dialog" **
/-----
/ Add_Terms Add_Active_Term(printing_line)
/ printing_line
/ enzyme_range
/-----
/ Add_Active_Term "Makes ':'s between the components" how_many_enzymes
/-----
/ Next_Row "Clears the old values"
/ Move_Term_To_Terms_Included
/ Update_Ready_Terms
/-----
/ Move_Term_To_Terms_Included "Move terms to Terms_Included-sheet"
/ how_many_enzymes
/ whole_term
/-----
/ Update_Ready_Terms "Types selected terms to the bottom of the dialog"
/ N_of_Terms
/-----
/ Remove_x_variable "Removes one or more comonents"
/ Do_Removing_Box return_pressed
/ *Removing_Actions canc_removing
/ Clear_Removing_Box remove_term
/ ** SHOWS DIALOG: "Remove_Dialog" **
/-----
/ Do_Remove_x_variable "Run if user presses Remove-button"
/ remove_term
/-----

```

Ask_Term_Macros

```

' Cancel_Term_Selection      "Cancels selection of terms"
'   Clear_all               canc_term_selection
'
'-----
' Clear_All                 "Clears few sheets and dialogs"
'-----
'
' * = conditional
'*****
' Some Global Variables (Name says what they do)

Dim errors_on_number As Boolean
Dim errors_on_levels As Boolean
Dim errors_on_level_values As Boolean
Dim errors_on_coefficients As Boolean
Dim return_pressed As Boolean

Dim canc_removing As Boolean
Dim canc_term_selection As Boolean
Dim canc_n_of_enz As Boolean
Dim canc_ask_levels As Boolean
Dim canc_ask_level_values As Boolean
Dim canc_coefficients As Boolean

Dim more_terms_to_come As Boolean
Dim more_enzymes_to_come As Boolean
Dim remove_term As Boolean
Dim OK_for_design As Boolean
Dim n_of_enzymes

'*****
Sub Start_macros ()
    ' "Main" Program
    Clearing_All
    Clear_Design_table
    Show_N_of_Enz
    If OK_for_design = True Then Plan_Design
    ' clears the background-sheets
    ' shows the first page
    ' if everything OK then continue
End Sub

```

Ask_Term_Macros

```

MsgBox "Your value is not numerical", vbExclamation, "Error"
errors_on_number = True
End If

If IsNumeric(DialogSheets("N_of_Var_Dialog").EditBoxes(1).Text) = True Then
n_of_enzymes = Val(DialogSheets("N_of_Var_Dialog").EditBoxes(1).Text)
If n_of_enzymes <> Int(n_of_enzymes) Then ' checks if value is integer
errors_on_number = True
MsgBox "Value must be an integer", vbExclamation, "Error"
End If
If n_of_enzymes > 50 Then ' checks if value is larger than 50
errors_on_number = True
MsgBox "Value cannot be over 50", vbExclamation, "Error"
End If
If n_of_enzymes < 1 Then ' checks if value is smaller than 1
errors_on_number = True
MsgBox "Value cannot be smaller than 1", vbExclamation, "Error"
End If
End If

End Sub

'*****

Sub Cancel_N_of_Enz()

canc_n_of_enz = True

End Sub

'*****

Sub Move_N_of_Enz()

Worksheets("Variable_Levels").Columns("A:A").ColumnWidth = 20
Worksheets("Variable_Levels").Cells(1, "A").Value = "Number of Enzymes"
Worksheets("Variable_Levels").Cells(1, "B").Value = n_of_enzymes

End Sub

'*****

```

Ask_Term_Macros

```

End Sub
'*****
Sub Cancel_Ask_Levels()
    ' canceling asking levels
    ' deleting all boxes
    For counter = 1 To n_of_enzymes
        Sheets("Levels_Dialog").DrawingObjects("Name " & counter).Delete
        Sheets("Levels_Dialog").DrawingObjects("Value " & counter).Delete
    Next counter
    Sheets("Levels_Dialog").DialogFrame.Left = 10
    Sheets("Levels_Dialog").DialogFrame.Top = 10
    Sheets("Levels_Dialog").DialogFrame.Height = 75
    Canc_ask_levels = True
End Sub
'*****
Sub Check_Names_And_Errors()
    'checking if there are errors
    'on names or levels
    errors_on_levels = False
    For counter = 1 To n_of_enzymes
        If Sheets("Levels_Dialog").DrawingObjects("Name " & counter).Text = "" Then
            errors_on_levels = True
            MsgBox "The names of the variables are required", vbExclamation, "Error"
        Exit For
    End If
    If IsNumeric(Sheets("Levels_Dialog").DrawingObjects("Value " & counter).Text) = False Then
        errors_on_levels = True
        MsgBox "Levels must be integers", vbExclamation, "Error"
    Exit For
    End If
    level = Val(Sheets("Levels_Dialog").DrawingObjects("Value " & counter).Text)
    If Int(level) <> level Then
        MsgBox "Levels must be integers", vbExclamation, "Error"
        errors_on_levels = True
    Exit For
    End If
    If level < 1 Then
        errors_on_levels = True
        MsgBox "Number of levels must be over zero", vbExclamation, "Error"
    End If

```

Ask_Term_Macros

```

Sheets("Level_Values_Dialog").DialogFrame.Height = 100
For counter = 1 To n_of_enzymes
    Do_Level_Values_Dialog (counter)
    'make Level_Value_Dialog
    If errors_on_level_values = True Then Exit Sub
    If canc_ask_level_values = True Then Exit Sub
Next counter

End Sub

/*****
Sub Do_Level_Values_Dialog(What_Enzyme)
    'make the Dialog with some labels and boxes

    label_text = "Value of Levels for: " & Worksheets("Variable_Levels").Cells(What_Enzyme + 2, "A").Value
    levels_of_enzyme = Worksheets("Variable_Levels").Cells(What_Enzyme + 2, "B").Value
    Sheets("Level_Values_Dialog").Labels.Add(20, 30, 170, 13.5).Name = "Label 1"
    Sheets("Level_Values_Dialog").DrawingObjects("Label 1").Characters.Text = label_text
    x_pos_name = 63
    x_pos_value = 115
    y_pos = 53
    height_both = 13.5
    lenght_name = 60
    lenght_value = 30
    y_more = 18

    For i = 1 To levels_of_enzyme
        'make edit_boxes to every level
        Sheets("Level_Values_Dialog").Labels.Add(x_pos_name, y_pos, lenght_name, height_both).Name = "Level " & i
        Sheets("Level_Values_Dialog").EditBoxes.Add(x_pos_value, y_pos, lenght_value, height_both).Name =
            "Level_value " & i
        Sheets("Level_Values_Dialog").DrawingObjects("Level " & i).Characters.Text = "Level " & i - 1
        Y_pos = Y_pos + Y_more
        If i > 3 Then Sheets("Level_Values_Dialog").DialogFrame.Height = _
            Sheets("Level_Values_Dialog").DialogFrame.Height + Y_more
    Next i
Do
    canc_ask_level_values = False
    Sheets("Level_Values_Dialog").Show
    If canc_ask_level_values = True Then Exit Do
    If errors_on_level_values = False Then Move_Level_Values (What_Enzyme)
Loop While errors_on_level_values = True

End Sub

```

Ask_Term_Macros

```

'*****
Sub Clear_Level_Values_Dialog(EditBoxes_amount)
    'clear every dialog
    Sheets("Level_Values_Dialog").DrawingObjects("Label 1").Delete
    For counter = 1 To EditBoxes_amount
        ' deleteing all dialogs
        Sheets("Level_Values_Dialog").DrawingObjects("Level_value " & counter).Delete
        Sheets("Level_Values_Dialog").DrawingObjects("Level " & counter).Delete
    Next counter
End Sub
'*****
Sub Cancel_Ask_Level_Values()
    'cancel asking
    Sheets("Level_Values_Dialog").DrawingObjects("Label 1").Delete
    box_amount = Sheets("Level_Values_Dialog").EditBoxes.Count
    For i = 1 To box_amount
        'and delete all boxes in Level_Value-sheet
        Sheets("Level_Values_Dialog").DrawingObjects("Level_value " & i).Delete
        Sheets("Level_Values_Dialog").DrawingObjects("Level " & i).Delete
    Next i
    canc_ask_level_values = True
End Sub
'*****
Sub Move_Level_Values(What_Enzyme)
    levels_of_enzyme = Worksheets("Variable_Levels").Cells(What_Enzyme + 2, "B").Value
    rows_so_far = 0
    If What_Enzyme > 1 Then
        For row_count = 1 To What_Enzyme - 1
            rows_so_far = rows_so_far + Worksheets("Variable_Levels").Cells(row_count + 2, "B").Value
        Next row_count
    End If
    start_line = rows_so_far + 1
    'linenumber to start putting values
    For What_Box = 1 To levels_of_enzyme
        Worksheets("Variable_Level_Values").Cells(start_line + What_Box + 1, "C").Value = _
            Sheets("Level_Values_Dialog").DrawingObjects("Level_value " & What_Box).Text
    Next What_Box
End Sub
'*****

```

Ask_Term_Macros

```

If more_enzymes_to_come = False Then Next_Row
If more_terms_to_come = False Then Ask_Coefficients
Loop While (canc_term_selection = False And more_terms_to_come = True)

End Sub

'*****
Sub Add_terms ()
    'add terms that were just chosen
    If Worksheets("extra_page").Cells(1, "B").Value = 0 Then 'checks that at least one term chosen
        MsgBox "You didn't choose a variable", vbExclamation, "Error"
        Exit Sub
    End If
    If Worksheets("extra_page").Cells(2, "B").Value = "" Then Worksheets("extra_page").Cells(2, "B").Value = 1
    printing_line = Worksheets("extra_page").Cells(3, "B").Value ' adding terms to Terms-sheet
    what_column = Worksheets("extra_page").Cells(1, "B").Value
    what_column = Worksheets("extra_page").Cells(2, "B").Value + 1
    Worksheets("Terms").Cells(printing_line, what_column).Value = _
        Worksheets("Variable_Levels").Cells(what_line + 1, "A").Value
    Add_Active_Term (printing_line)
    Worksheets("extra_page").Cells(2, "B").Value = Worksheets("extra_page").Cells(2, "B").Value + 1
End Sub

'*****
Sub Add_Active_Term(output_line)
    Worksheets("extra_page").Cells(12, "A").Value = "Interact ?" ' makes "*" :s between the components
    ' write title

    how_many_enzymes = Worksheets("extra_page").Cells(2, "B").Value
    Act_name = Worksheets("Terms").Cells(output_line, "B").Value
    If how_many_enzymes > 1 Then
        Worksheets("extra_page").Cells(12, "B").Value = "YES" ' set value true if there is at least one interacti
    on
    For what_column = 3 To how_many_enzymes + 1
        Act_name = Act_name & "*" & Worksheets("Terms").Cells(output_line, what_column).Value
        Next what_column
    End If
    Sheets("Term_Selection_Dialog").DrawingObjects("Active Enzyme").Characters.Text = Act_name

```


Ask_Term_Macros

```

N_of_Terms = Worksheets("extra_page").Cells(3, "B").Value

x_place = 30
y_place = 140
length = 147
label_height = 13.5
y_more = 12
For new_boxes = 1 To N_of_Terms
    Sheets("Term_Selection_Dialog").Labels.Add(x_place, y_place, length, label_height). _
        Name = "Term " & new_boxes
    term_name = Worksheets("Terms_Included").Cells(new_boxes, "B").Value
    Sheets("Term_Selection_Dialog").DrawingObjects("Term " & new_boxes).Text = term_name
    y_place = y_place + y_more
    If new_boxes > 3 Then
        Sheets("Term_Selection_Dialog").DialogFrame.Height = _
            y_place + 5
    End If
Next new_boxes

End Sub

*****

Sub Remove_x_variable()
    'remove one or more terms
    'make box, where the term(s) can be chosen
    ' do until removing has been made
    Do
        Sheets("Remove_Dialog").Show
        If remove_term = True Then Removing_Actions 'if remove is pressed then removing_actions
            remove_term = False
        Loop While (return_pressed = False And cancel_removeing = False)
        Clear_Removing_Box
    End Sub

    'if really want to remove a term
    Sub Do_Remove_x_variable()

```

Ask_Term_Macros

```

Worksheets("Terms").Rows(remover).Delete
Worksheets("Terms_Included").Rows(remover).Delete

Worksheets("extra_page").Cells(3, "B").Value = _
Worksheets("extra_page").Cells(3, "B").Value - 1 'make the number of the
'remover to extra_page
Update_Ready_Terms 'remove from ready terms list
Sheets("Term_Selection_Dialog").DrawingObjects("Active Enzyme").Characters.Text = ""
Clear_Removing_Box 'deleting all from box
Do_Removing_Box 'and make it again with
'right terms
Worksheets("extra_page").Cells(12, "B").Value = "" 'takes interactio "yes" away

down_row = Worksheets("extra_page").Cells(3, "B").Value
For check_row = 1 To down_row 'if there still are interactions, puts "yes" back
If Worksheets("Terms").Cells(check_row, "C").Value <> "" Then
Worksheets("extra_page").Cells(12, "B").Value = "YES"
Exit For
End If
Next check_row
End Sub

,*****

Sub Ask_Coefficients() 'ask coefficients for terms

canc_coefficients = False
Clear_Coefficients 'delete all boxes in coefficients-dialog
Initialize_Coefficients_Dialog 'and make it again with right titles
Do
errors_on_coefficients = False
Sheets("Coefficients_Dialog").Show
Loop While (errors_on_coefficients = True)

If canc_coefficients = True Then
more_terms_to_come = True
Clear_Coefficients
Exit Sub
End If

End Sub

```

Ask_Term_Macros

```

If boxes_count > 0 Then
  Sheets("Coefficients_Dialog").Labels.Delete
  Sheets("Coefficients_Dialog").EditBoxes.Delete
End If

End Sub

*****
Sub Check_Coefficients()
  max_counter = Worksheets("extra_page").Cells(3, "B").Value - 1
  For counter = 1 To max_counter
    If IsNumeric(Sheets("Coefficients_Dialog").EditBoxes(counter).Text) = False Then
      MsgBox "Coefficients must be numbers ", vbExclamation, "Error"
      errors_on_coefficients = True
    Exit Sub
  End If

  Next counter

  For moving_counter = 1 To max_counter
    Worksheets("Terms").Cells(moving_counter, "A").Value = _
    Sheets("Coefficients_Dialog").EditBoxes(moving_counter).Text
    Worksheets("Terms_Included").Cells(moving_counter, "A").Value = _
    Sheets("Coefficients_Dialog").EditBoxes(moving_counter).Text
  Next moving_counter
  OK_for_design = True
End Sub

*****
Sub Return_Pressing()
  return_pressed = True
End Sub

```

Ask_Term_Macros

```
If Worksheets("Terms_Included").Cells(1, "B").Value = " " Then 'checks that at least one term chosen
MsgBox "You didn't choose a variable", vbExclamation, "Error"
Exit Sub
End If
end adding

more_terms_to_come = False
End Sub
'*****
Sub Cancel_Term_Selection()
'cancel pressed

canc_term_selection = True
Worksheets("FrontPage").Select
Clear_All
End Sub
'clear "all" cells in sheets and dialogs
'(all is defined in next sub)
'*****
Sub Clear_All()
'clear the following dialogs and sheets

Worksheets("Levels_Dialog").EditBoxes.Delete
Worksheets("Term_Selection_Dialog").ListBoxes.Delete
Worksheets("Term_Selection_Dialog").Labels.Delete
Worksheets("Variable_Levels").Cells.ClearContents
Worksheets("Variable_Level_Values").Cells.ClearContents
Worksheets("Terms").Cells.ClearContents
Worksheets("extra_page").Cells.ClearContents
End Sub
'*****
Sub Help_N_of_x_variable()
Worksheets("Help_N_of_x_variable").Show
End Sub
'*****
```

End Sub

Liite 2

Toistojen määrän ja voimakkuuden
laskemiseen käytettävät funktiot ja
aliohjelmat

Design_Macros

```

' Show_Matrix_C
'
' Count_Non_Central
' Make_User_Table
' ** SHOWS DIALOG: "Matrix_C_Dialog" **
'
' Make_Matrix_C
'
' Clear_Matrix_C
' Make_Color(what_line)
' Do_Zero_Matrix(cross)
' Finish_Matrix_C(cross)
' Uncolor_All_Terms
'
' f Find_Name(comp_name)
'
' Do_Zero_Matrix(cross)
'
' Make_Color(what_line)
'
' Uncolor_All_Terms
'
' Clear_Matrix_C
'
' Cancel_Matrix_C
'
' Finish_Matrix_C(cross)
'
' Count_Non_Central
'
' Make_Names
'
' Make_User_Table
'
' Do_Borders ("A")
' Coloring_Area ("A")
' Centering_Area ("A")
' Set_Width ("A")
' Bold_Area ("A")
' Numbers_Width ("B")
' Insert_Boxes
' Insert_Included_Terms
' Initialize_N_Table

```

"Shows matrix C"

errors_on_matrix_c
cancel_matrix_c_dialog

"After showing the dialog, makes matrix C"

errors_on_matrix_c

"Looks for comp_name from All_Terms-sheet"

"Makes matrix C full of zeros"

"Colors the parameter line from All_Terms-sheet2"

"Uncolor All_Terms-sheet"

"Clears Matrix_C -sheet"

"Cancel-button pressed" cancel_matrix_c_dialog

"Makes Ones to matrix C"

"Makes a formula to extra_page cell B9"

"Names the matrixes as 'Global-variables'"

"Initializes User_Page-sheet"

Design_Macros

```

' Initialize_N_Table "Sets formulas, titles and so on"
' -----
' Make_Start_Values(col) "Sets few formulas to N_Table"
' -----
' Count_Power (col) "Counts Power"
' -----
' Count_N (col) "Counts Number of Replicates"
' -----
' f = function * = optional
' *****
Dim errors_on_matrix_c As Boolean
Dim cancel_matrix_c_dialog As Boolean
Dim is_singular As Boolean
'Dim terms_amount As Integer
' *****

Sub Plan_Design()
    'Main-Program for making design and so on
    Const max_lines = 300 'constant value for max_lines = height of the design-matrix
    Clear_Design_table 'clears the Design_table-sheet
    N_of_columns = Worksheets("Variable_Levels").Cells(1, "B").Value
    first_column = 2
    last_column = N_of_columns + 1
    last_column_alfa = Chr(64 + last_column)
    Worksheets("Design_table").Cells(1, "A").Value = "Type Design"
    Worksheets("Design_table").Cells(2, "A").Value = "max level"
    Worksheets("Design_table").Range("A1").Font.Bold = True
    For what_column = 2 To N_of_columns + 1 'Makes names and max_levels to Design_table-sheet
        input_line = what_column + 1
        x_variable_name = Worksheets("Variable_Levels").Cells(input_line, "A").Value
        Worksheets("Design_table").Cells(1, what_column).Value = x_variable_name
        max_levels = Worksheets("Variable_Levels").Cells(input_line, "B").Value - 1
        max_lines = 100
        max_lines = max_lines * (max_levels + 1)
        Worksheets("Design_table").Cells(2, what_column).Value = max_levels
    Next what_column

```

Design_Macros

```
Sheets("Design_table").Range("A3:A8").BorderAround Weight:=xlThin

End Sub

'*****

Sub Color_Area(last_column_alfa, max_lines) 'Colors the area in Design_table-sheet
  Sheets("Design_table").Range("A1:" & last_column_alfa & "2").Interior.ColorIndex = 22
  Sheets("Design_table").Range("B3:" & last_column_alfa & max_lines + 2).Interior.ColorIndex = 38

  area = "A3:A8"
  Sheets("Design_table").Range(area).Interior.ColorIndex = 38
End Sub

'*****

Sub Center_All(last_column_alfa, max_lines) 'Centers area in Design_table-sheet
  area = "A1:" & last_column_alfa & max_lines + 2
  Sheets("Design_table").Range(area).HorizontalAlignment = xlCenter
End Sub

'*****

Sub Make_boxes()
  'Makes new boxes to Design_table-sheet
  'Boxes: Help and Design_Ready

  Remove_Buttons
  x_place = 6
  y1_place = 31
  y2_place = 70
  lenght = 45
  height_1 = 29
  height_2 = 18

  Sheets("Design_table").Buttons.Add(x_place, y1_place, lenght, height_1).Name = "Button 1"
  Sheets("Design_table").Buttons.Add(x_place, y2_place, lenght, height_2).Name = "Button 2"

  Sheets("Design_table").DrawingObjects("Button 1").Text = "Design Ready"
  Sheets("Design_table").DrawingObjects("Button 2").Text = "Help"

  Sheets("Design_table").DrawingObjects("Button 1").OnAction = "Make_Design_Matrix"
  Sheets("Design_table").DrawingObjects("Button 2").OnAction = "HelpDesign"
End Sub
```

Design_Macros

```
numbers = 0
For what_line = 3 To max_rows
    max_value = Sheets("Design_table").Cells(2, what_column).Value
    cell_value = Sheets("Design_table").Cells(what_line, what_column).Value

    If (IsNumeric(cell_value) = False) Then
        MsgBox "Values must be integers", vbExclamation, "Error"
        Sheets("Design_table").Range(Cells(what_line, what_column), _
            Cells(what_line, what_column)).Select
        Test_Values = False
        Exit Function
    End If

    If (cell_value < 0 Or cell_value > max_value) Then
        MsgBox "Your value is out of range", vbExclamation, "Error"
        Sheets("Design_table").Range(Cells(what_line, what_column), _
            Cells(what_line, what_column)).Select
        Test_Values = False
        Exit Function
    End If

    If cell_value <> Int(cell_value) Then
        MsgBox "Values must be integers", vbExclamation, "Error"
        Sheets("Design_table").Range(Cells(what_line, what_column), _
            Cells(what_line, what_column)).Select
        Test_Values = False
        Exit Function
    End If

    If cell_value <> " " Then numbers = numbers + 1
Next what_line

If what_column = 2 Then numbers_for_first_row = numbers
total_numbers = total_numbers + numbers
'Tests if the number of values is different is some row
If numbers <> numbers_for_first_row Then
    MsgBox "The number of values differ in some rows", vbExclamation, "Error"
    Sheets("Design_table").Select
    Test_Values = False
    Exit Function
End If
Next what_column
```

Design_Macros

```

Sub Make_Full_Design()
    'Takes all terms and move them to All_Terms-sheet

    enzyme_amount = Worksheets("Variable_Levels").Cells(1, "B").Value + 1
    end_line = enzyme_amount + 1
    For from_line = 2 To end_line
        'Moves the names from Variable_Levels to All_Terms
        Worksheets("All_Terms").Cells(from_line - 1, "B").Value =
            Worksheets("Variable_Levels").Cells(from_line, "A").Value
    Next from_line

    terms_amount = Worksheets("extra_page").Cells(3, "B").Value - 1

    For item_line = 1 To enzyme_amount
        For search_line = 1 To terms_amount
            search_item = Worksheets("All_Terms").Cells(item_line, "B").Value
            target = Worksheets("Terms_Included").Cells(search_line, "B").Value
            If target = search_item Then
                Worksheets("All_Terms").Cells(item_line, "A").Value =
                    Worksheets("Terms_Included").Cells(search_line, "A").Value
                Worksheets("Terms_Included").Range("B" & search_line).Interior.ColorIndex = 3
            Exit For
        End If
    Next search_line
    'if coefficient=" then it is 0
    If Worksheets("All_Terms").Cells(item_line, "A").Value = " Then
        Worksheets("All_Terms").Cells(item_line, "A").Value = 0
        Worksheets("All_Terms").Cells(item_line, "A").Interior.ColorIndex = 15
    End If
    Next item_line

    output_line = enzyme_amount + 1
    For copy_line = 1 To terms_amount
        'if color of the cell is "3" then...
        If Worksheets("Terms_Included").Range("B" & copy_line).Interior.ColorIndex <> 3 Then
            Worksheets("All_Terms").Cells(output_line, "A").Value =
                Worksheets("Terms_Included").Cells(copy_line, "A").Value
            Worksheets("All_Terms").Cells(output_line, "B").Value =
                Worksheets("Terms_Included").Cells(copy_line, "B").Value
            Call Copy_Whole_Terms(output_line, copy_line)
            output_line = output_line + 1
        End If
    Next copy_line
    Worksheets("Terms_Included").Columns("B:B").Interior.ColorIndex = xlNone

```

Design_Macros

```

start_col = Worksheets("Variable_Levels").Cells(1, "B").Value + 2
end_col = Worksheets("extra_page").Cells(5, "C").Value + 1
For output = start_col To end_col
    Worksheets("Matrix_X").Cells(2, output).Value = _
        Worksheets("All_Terms").Cells(output, "B").Value
    Worksheets("Matrix_X").Cells(1, output).Value = output
Next output

End Sub

'*****
Sub Make_Matrix_X()
    'Makes formulas to matrix X

    If Worksheets("extra_page").Cells(12, "B").Value <> "YES" Then Exit Sub ' if there are interactions then exit sub

    start_col = Worksheets("Variable_Levels").Cells(1, "B").Value + 2
    end_col = Worksheets("extra_page").Cells(5, "C") + 1
    start_row = 3
    end_row = Worksheets("extra_page").Cells(5, "B") + 2
    For what_col = start_col To end_col
        row_for_name = Worksheets("Matrix_X").Cells(1, what_col).Value
        For what_row = start_row To end_row
            search_col = 3
            new_text = ""
            formula_text = ""
            found = 0
            Do
                name_to_search = Worksheets("All_Terms").Cells(row_for_name, search_col).Value
                For matrix_col = 1 To start_col
                    If name_to_search = Worksheets("Matrix_X").Cells(2, matrix_col).Value Then
                        col_letter = Chr(matrix_col + 64)
                        new_text = col_letter & what_row
                        If found > 0 Then new_text = "*" & new_text
                        formula_text = formula_text & new_text
                        Worksheets("Matrix_X").Cells(what_row, what_col).Formula = "=" & formula_text
                        found = found + 1
                    Exit For
                End If
            Next matrix_col
            search_col = search_col + 1
        For what_row = start_row To end_row
            search_col = 3
            new_text = ""
            formula_text = ""
            found = 0
            Do
                name_to_search = Worksheets("All_Terms").Cells(row_for_name, search_col).Value
                For matrix_col = 1 To start_col
                    If name_to_search = Worksheets("Matrix_X").Cells(2, matrix_col).Value Then
                        col_letter = Chr(matrix_col + 64)
                        new_text = col_letter & what_row
                        If found > 0 Then new_text = "*" & new_text
                        formula_text = formula_text & new_text
                        Worksheets("Matrix_X").Cells(what_row, what_col).Formula = "=" & formula_text
                        found = found + 1
                    Exit For
                End If
            Next matrix_col
            search_col = search_col + 1
        End For
    End For
End Sub

```

Design_Macros

```

End Sub
'*****
Sub Make_Constant_Col(max_row)
    'Makes column full of Ones to matrix X
    Worksheets("Matrix_X").Cells(2, "A").Value = "Constant"
    For what_row = 3 To max_row
        Worksheets("Matrix_X").Cells(what_row, "A").Value = 1
    Next what_row
End Sub
'*****
Sub Write_Dims()
    'Makes formulas and values for dimensions to extra_page
    Worksheets("extra_page").Cells(5, "C").Formula = "=count(Matrix_X!3:3)"
    Worksheets("extra_page").Cells(6, "A").Value = "Dim of Beta"
    Worksheets("extra_page").Cells(6, "B").Formula = "=count(All_Terms!A:A)"
    Worksheets("extra_page").Cells(6, "C").Value = 1
    Worksheets("extra_page").Cells(7, "A").Value = "Dim of C"
    Worksheets("extra_page").Cells(7, "B").Formula = "=count(Matrix_C!$A:$A)"
    Worksheets("extra_page").Cells(7, "C").Formula = "=count(All_Terms!$A:$A)"
    Worksheets("extra_page").Cells(8, "A").Value = "N of Terms"
    Worksheets("extra_page").Cells(8, "B").Formula = "=count(All_Terms!A:A)"
End Sub
'*****
Sub Make_Matrix_C_Dialog()
    'Makes dialog for asking values for matrix C
    Sheets("Matrix_C_Dialog").DialogFrame.Left = 10
    Sheets("Matrix_C_Dialog").DialogFrame.Top = 10
    Sheets("Matrix_C_Dialog").DialogFrame.Height = 100
    If Sheets("Matrix_C_Dialog").CheckBoxes.Count > 0 Then Sheets("Matrix_C_Dialog").CheckBoxes.Delete
    terms_amount = Worksheets("extra_page").Cells(8, "B").Value
    x_place = 22

```

Design_Macros

```

Count_Non_Central
Make_User_Table

'Makes a quite difficult matrix formula to extra_page
'Initializes User_Page-sheet

End Sub

'*****

Sub Make_Matrix_C()
    'After asking terms to deal with, makes matrix C

    Clear_Matrix_C
    Sheets("All_Terms").Columns("B:B").Interior.ColorIndex = xlNone
    box_amount = Sheets("Matrix_C_Dialog").CheckBoxes.Count
    cross = 0
    For box_now = 1 To box_amount
        If Sheets("Matrix_C_Dialog").CheckBoxes(box_now).Value = 1 Then
            comp_name = Sheets("Matrix_C_Dialog").CheckBoxes(box_now).Text
            what_line = Find_Name(comp_name) 'finds the name and returns line number
            Make_Color (what_line)        'colors the right line
            cross = cross + 1
        End If
    Next box_now
    If cross = 0 Then
        errors_on_matrix_c = True
        MsgBox "You must choose at least one term", vbExclamation, "Error"
    End If
    Exit Sub
End If
Do_Zero_Matrix (cross)
Finish_Matrix_C (cross)
Uncolor_All_Terms

End Sub

'*****

Function Find_Name(comp_name)
    'Finds the comp_name from All_Terms-sheet

    max_line = Worksheets("extra_page").Cells(8, "B").Value
    For what_line = 1 To max_line
        If Worksheets("All_Terms").Cells(what_line, "B").Value = comp_name Then
            Find_Name = what_line
            Exit Function
        End If
    End If
End Function

```

Design_Macros

```

'*****
Sub Cancel_Matrix_C()
    'if cancel pressed then run this macro

    cancel_matrix_c_dialog = True
End Sub

'*****
Sub Finish_Matrix_C(cross)
    'Makes Ones to matrix C if required

    found = 1
    what_line = 1
    Do
        'if color is "3" then that cell is One in matrix C
        'found is row-index in All_Terms but column-index in Matrix_C
        If Worksheets("All_Terms").Cells(found, "B").Interior.ColorIndex = 3 Then
            Worksheets("Matrix_C").Cells(what_line, found).Value = 1
            what_line = what_line + 1
        End If
        found = found + 1
    Loop While what_line <= cross
End Sub

'*****
Sub Count_Non_Central()
    'Sets quite long matrix formula to extra_page cell B9

    Make_Names

    Worksheets("extra_page").Cells(9, "A").Value = "NonCentPar"
    Worksheets("extra_page").Range("B9").FormulaArray = "=MMULT(MMULT(TRANSPOSE(MMULT(Mat_C,B))) & _
        ",MINVERSE(MMULT(MMULT(Mat_C,MINVERSE((MMULT(TRANSPOSE(X,X))), " & _
        "TRANSPOSE(Mat_C))), MMULT(Mat_C,B)))"
End Sub

'*****
Sub Make_Names()
    'Makes names of matrixes as Global Names

```


Design_Macros

```

End Sub

'*****
Sub Make_User_Table()
    'Initializes User_Page-sheet
    Worksheets("extra_page").Cells(13, "A").Value = "Last Column"
    Worksheets("extra_page").Cells(13, "B").Value = 2
    Worksheets("extra_page").Cells(13, "C").Formula = "=Char(64+B13)" '65="A", 66="B"
    Worksheets("User_Page").Cells.Delete Shift:=xlUp
    Worksheets("User_Page").Cells.Interior.ColorIndex = 19

    Worksheets("User_Page").Rows("3:6").RowHeight = 25

    Do_Borders ("A")
    Do_Borders ("B")
    Coloring_Area ("A")
    Coloring_Area ("B")
    Centering_Area ("A")
    Centering_Area ("B")
    Set_Width ("A")
    Bold_Area ("A")
    Numbers_Width ("B")
    Insert_Boxes
    Insert_Included_Terms
    Initialize_N_Table

    Worksheets("User_Page").Cells(1, "A").Value = "Determining the Power / Number of Replicates"
    Worksheets("User_Page").Cells(1, "A").Font.Bold = True
    Worksheets("User_Page").Cells(1, "A").Font.Size = 12

    Worksheets("User_Page").Range("A1:F1").HorizontalAlignment = xlCenterAcrossSelection

    Worksheets("User_Page").Cells(3, "A").Value = " RSD"
    Worksheets("User_Page").Cells(4, "A").Value = " Significance"
    Worksheets("User_Page").Cells(5, "A").Value = " N of Replicates"
    Worksheets("User_Page").Cells(6, "A").Value = " Power (%)"
    Worksheets("User_Page").Select
End Sub

'*****

```

Design_Macros

```
Sub Bold_Area(col)
Worksheets("User_Page").Range(col & "3:" & col & "6").Font.Bold = True
End Sub
'*****
Sub Numbers_Width(col)
'Numberformat in rows 3,4,6 is "0.00" and in row 5 it is "0"
Worksheets("User_Page").Range(col & "3:" & col & "4").NumberFormat = "0.00"
Worksheets("User_Page").Range(col & "6").NumberFormat = "0.0000"
Worksheets("User_Page").Range(col & "5").NumberFormat = "0"
End Sub
'*****
Sub Insert_Boxes()
'Makes buttons to User_Page-sheet
Clear_Buttons
Y_place = 135
b_height = 22
length = 40
Y2_place = 165
Y3_place = 195
Y4_place = 225
Y5_place = 255
count_start = Worksheets("User_Page").Columns("A").Width
count_end = Worksheets("User_Page").Columns("B").Width
'sets help-button to the middle of the first column
Sheets("User_Page").Buttons.Add((count_start - length) / 2, Y_place, length, b_height).Name = "Help"
Sheets("User_Page").DrawingObjects("Help").Text = "Help"
Sheets("User_Page").DrawingObjects("Help").OnAction = "HelpUser"
Sheets("User_Page").DrawingObjects("Help").PrintObject = True
count_place = count_start + (count_end - length) / 2 'in the middle of the second column
Sheets("User_Page").Buttons.Add(count_place, Y_place, length, b_height).Name = "Count 1"
Sheets("User_Page").DrawingObjects("Count 1").Text = "Calc"
Sheets("User_Page").DrawingObjects("Count 1").OnAction = "Count_User_Table"
Sheets("User_Page").DrawingObjects("Count 1").PrintObject = True
```

Design_Macros

```

End Sub
'*****
Sub Insert_Included_Terms()
    'Makes line where is the names of the terms
    Worksheets("User_Page").Cells(20, "A").Value = "You have chosen term(s):"
    rows_C = Worksheets("extra_page").Cells(7, "B").Value
    cols_C = Worksheets("extra_page").Cells(7, "C").Value
    If Worksheets("All_Terms").Cells(1, "A").Value = 0 _
    And Worksheets("All_Terms").Cells(1, "A").Interior.ColorIndex = 15 Then cols_C = cols_C + 1

    user_row = 21
    For what_row = 1 To rows_C
        For what_col = 1 To cols_C
            'finds the terms which are chosen
            If Worksheets("Matrix_C").Cells(what_row, what_col).Value = 1 Then
                term_name = Worksheets("All_Terms").Cells(what_col, "B").Value
                Worksheets("User_Page").Cells(user_row, "B").Value = term_name
                user_row = user_row + 1
            Exit For
        End If
    Next what_col
    Next what_row
End Sub
'*****
Sub Clear_User_Table()
    'clears column B and deletes columns C-...
    last_col = Worksheets("extra_page").Cells(13, "C").Value

    If last_col <> "B" Then Worksheets("User_Page").Range("C3:" & last_col & "6")._
        Delete Shift:=xlUp
    Worksheets("User_Page").Range("B3:B6").ClearContents
    Worksheets("extra_page").Cells(13, "B").Value = 2

    If Worksheets("User_Page").Cells(6, "B").Font.Bold = True Then Worksheets("User_Page").Cells(6, "B").Font.Bold = False
    If Worksheets("User_Page").Cells(5, "B").Font.Bold = True Then Worksheets("User_Page").Cells(5, "B").Font.Bold = False

```

Design_Macros

```

Worksheets("extra_page").Cells(14, "A").Value = "Det (X*X)"
Worksheets("extra_page").Range("B14").FormulaArray = "=MDETERM(MMULT(TRANSPOSE(X), X))"
determinant = Worksheets("extra_page").Cells(14, "B").Value
If determinant < det_limit Then
    MsgBox "Your design matrix is singular," & Chr(13) & _
        "so You Have to Change Either Model or Design", vbExclamation, "Error"
    Check_User_Table = False
    is_singular = True
    Exit Function
End If
'singular-test ends here

' Testing if non-central parameter is 0 (= if coeff of tested term is 0)
non_cent = Worksheets("extra_page").Cells(9, "B").Value
If non_cent = 0 Then
    MsgBox "The coefficient of your term is zero," & Chr(13) & _
        "so you have to change coefficients or chosen term(s)", vbExclamation, "Error"
    Check_User_Table = False
    Exit Function
End If
'end non-central
If (RSD = "" Or alfa = "") Then 'if both are empty
    MsgBox "You have to enter RSD and significance", vbExclamation, "Error"
    Check_User_Table = False
    Exit Function
End If
'if they are non-numerical
If (IsNumeric(RSD) = False Or IsNumeric(alfa) = False) Then
    MsgBox "Values must be numbers", vbExclamation, "Error"
    Check_User_Table = False
    Exit Function
End If

If (alfa >= 1 Or alfa <= 0) Then 'if alfa is out of range
    MsgBox "Significance must be between 0 and 1", vbExclamation, "Error"
    Check_User_Table = False
    Exit Function
End If

'replic = N of Replicates and Power
replic = Worksheets("User_Page").Cells(5, col).Value
pow = Worksheets("User_Page").Cells(6, col).Value
'if BOTH are empty or non-empty
If ((pow = "" And replic = "") Or (pow <> "" And replic <> "")) Then

```

Design_Macros

```
Coloring_Area (col)
Centering_Area (col)
Numbers_Width (col)

Worksheets("User_Page").Cells(3, col).Value = Worksheets("User_Page").Cells(3, last_col).Value
Worksheets("User_Page").Cells(4, col).Value = Worksheets("User_Page").Cells(4, last_col).Value

End Sub
'*****

Sub Bold_Cell_to_Count(col) 'make the font in the empty cell bold, italic and underline

If Worksheets("User_Page").Cells(5, col).Value = "" Then
    empty_row = 5
Else
    empty_row = 6
End If

Worksheets("User_Page").Cells(empty_row, col).Font.Bold = True
Worksheets("User_Page").Cells(empty_row, col).Font.Italic = True
Worksheets("User_Page").Cells(empty_row, col).Font.Underline = xlSingle

End Sub
'*****

Sub Make_N_Table(col) 'checks if user want to count power or number of replicates

Make_Start_Values (col) 'makes some formulas to N_Table
If Worksheets("User_Page").Cells(5, col).Value = "" Then
    Count_N (col) 'if Power-cell is empty -> count N
Else
    Count_Power (col) 'if N-cell is empty -> count Power
End If

End Sub
'*****

Sub Initialize_N_Table() 'Making titles, formulas and so on
```

Design_Macros

```

Sub Make_Start_Values(col)      'some formulas to User_Page-sheet
    Worksheets("N_Table").Cells(2, "B").Formula = "=User_page!" & col & "3" & "^2" 'variance
    Worksheets("N_Table").Cells(3, "B").Formula = "=User_Page!" & col & "4" 'significance
    Worksheets("N_Table").Cells(4, "B").Formula = "=extra_page!B9/(User_Page!" & col & "3" & "^2)" 'delta
End Sub

'*****
Sub Count_Power(col)          'Counts Power if N is given
    min_value_for_n = 200      ' value for checkin if something can be counted
                                ' (checks that n * rows of X > columns of X)
    n_value = Worksheets("User_Page").Cells(5, col).Value
    matrix_X_height = Worksheets("extra_page").Cells(5, "B").Value
    matrix_X_lenght = Worksheets("extra_page").Cells(5, "C").Value
    If n_value * matrix_X_height <= matrix_X_lenght Then
        For n_min = 2 To min_value_for_n
            If n_min * matrix_X_height > matrix_X_lenght Then Exit For
        Next n_min
        MsgBox "Sorry, but your design or choise of terms" & Chr(13) & _
            "is bad for calculations. This situation can " & Chr(13) & _
            "be solved if the value of N is at least " & n_min, vbExclamation, "Error"
        Worksheets("N_Table").Cells(10, "A").Value = n_min
        Worksheets("User_Page").Cells(5, col).Value = n_min
    Else
        Worksheets("N_Table").Cells(10, "A").Value = n_value      'Sets the value of N to N_Table
    End If
    power_value = Worksheets("N_Table").Cells(10, "M").Value      'Sets the value of Power too
    Worksheets("User_Page").Cells(6, col).Value = power_value * 100
End Sub

'*****
Sub Count_N(col)             'counts N if Power is given
    max_n = 250
    criteria = Worksheets("User_Page").Cells(6, col).Value      'value of given Power

```

Design_Macros

```
For search_from_line = 1 To max_line + 1
  search_name = Worksheets("All_Terms").Cells(search_from_line, "B").Value
  For target_line = 1 To max_line
    If search_name = Worksheets("Terms_Included").Cells(target_line, "B") Then
      Worksheets("All_Terms").Cells(search_from_line, "A").Value = _
        Worksheets("Terms_Included").Cells(target_line, "A").Value
    Exit For
  End If
  Next target_line
Next search_from_line
Clear_User_Table
Worksheets("User_Page").Select

End Sub
,*****

Sub Activate_User_Page()
  Worksheets("User_Page").Select
  Sheets("Coefficients_Dialog").DrawingObjects("Button 2").OnAction = "Check_Coefficients"
End Sub
,*****
```

Liite 3

Ohjelman käyttöohjeet

REPLIC

Koesuunnittelun apuna käytettävä ohjelma

toistojen määrän ja voimakkuuden

laskemiseen

KÄYTTÖOHJEET

Sisällysluettelo

1 Yleistä	1
2 Aloitus	2
3 X-muuttujien valinta	3
4 X-muuttujien nimet ja tasojen lukumäärät	4
5 Muuttujien eri tasojen numeeriset arvot	5
6 Malliin tulevien termien valinta	7
6.1 Vakio- ja päävaikutustermien lisäys	7
6.2 Yhdysvaikutustermien lisääminen malliin	9
6.3 Mallissa olevien termien poisto	11
7 Oletettujen kertoimien syöttö	13
8 Design-matriisin teko	14
9 Testattavaksi tulevien termien valinta	17
10 Toistojen määrän ja voimakkuuden laskeminen	18
10.1 Toistojen määrän arviointi voimakkuuden avulla	20
10.2 Voimakkuuden arvon laskeminen toistojen määrän avulla	21
11 Työn talletus	22

Tervetuloa REPLIC-ohjelman käyttäjäksi. Tämän käyttöohjeen tarkoituksena on tutustuttaa sinut ohjelman toimintaan siten, että saat siitä suurimman mahdollisen hyödyn tulevaa koettasi varten.

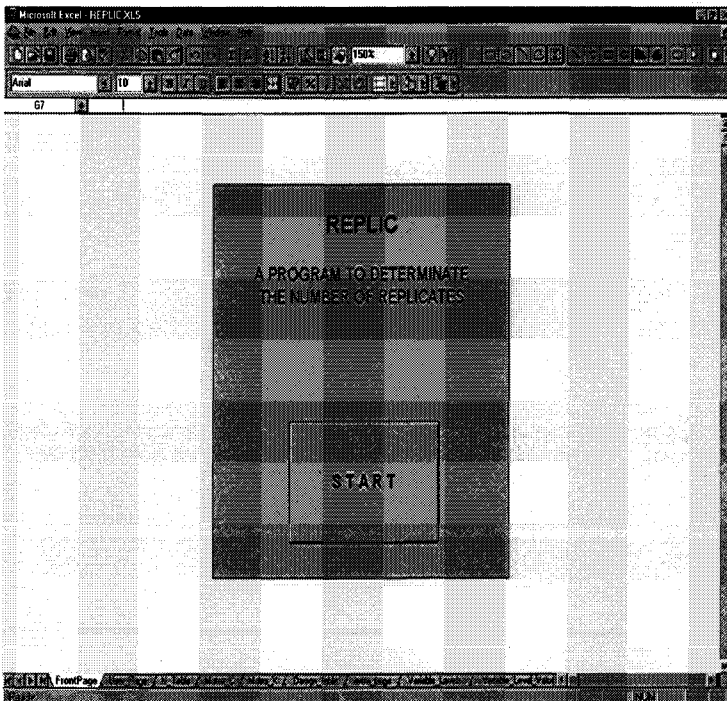
1 Yleistä

Ohjelman tarkoitus on toimia kokeen tekijän apuna koetta suunniteltaessa. Ohjelma on suunniteltu Cultor Oy:n Teknologiakeskuksen tarpeisiin. Ohjelma toimii Excel-tilukkolaskentaohjelman alaisuudessa. Lähtökohtana on regressioanalyttinen lähestyminen tutkittavaan ilmiöön. Tavoitteena on saada tietää eri muuttujien kertoimien mahdollisen nollastapoisuuden tilastollinen merkitsevyys. Vastemuuttujasta vaaditaan arvio sen varianssista. Lisäksi on päätettävä käytettävän testin merkitsevyydestä. Näillä eväillä voidaan laskea joko tarvittava toistojen määrä tietyn testin voimakkuuden saavuttamiseen tai toisaalta testin voimakkuus tietyllä toistojen määrällä.

Cancel-painikkeella päästään joka kohdassa edelliselle tasolle ja *Help*-painikkeella saadaan lisätietoja käsiteltävästä asiasta.

2 Aloitus

Koska ohjelma on tehty toimimaan Excel-ympäristössä, on aluksi avattava kyseinen taulukkolaskenta-ohjelma. Sen alaisuudesta on avattava tämä REPLIC-ohjelma, joka löytyy nimellä REPLIC.XLS. Ohjelma avataan kuten mikä tahansa Excel-työ. Avaamisen jälkeen saadaan koneen näyttöön kuvan 2.1 mukainen näkymä.

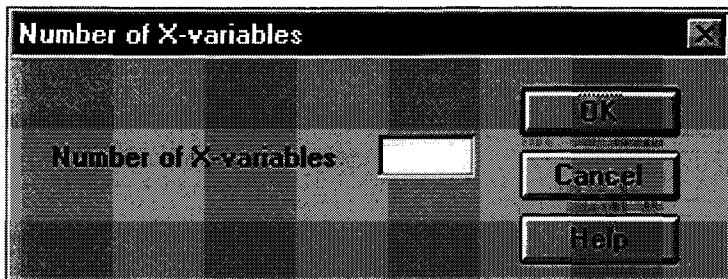


Kuva 2.1 Aloitussivu

Tästä päästään itse ohjelmaan klikkaamalla START-painiketta.

3 X-muuttujien valinta

Tämän jälkeen saadaan esiin X-muuttujien lukumäärän valintataulu. Se on esitetty kuvassa 3.1.



Kuva 3.1 X-muuttujien lukumäärän valinta

Oletetaan, että malli on muotoa

$$Y = \text{vakio} + c1 * \text{muuttuja}_1 + c2 * \text{muuttuja}_2 + c3 * \text{muuttuja}_1 * \text{muuttuja}_2$$

eli mallissa on:

- vastemuuttuja Y
- muuttuja_1 ja sillä kerroin c1
- muuttuja_2 ja sillä kerroin c2
- yhdysvaikutustermi, jossa on kahden edellisen yhdysvaikutus, eli ns. interaktio, jolla on kerroin c3.

Tällöin kohtaan *Number of X-variables* tulee mallissa olevien x-muuttujien lukumäärä eli tässä tapauksessa 2. Jos taas malli olisi muotoa

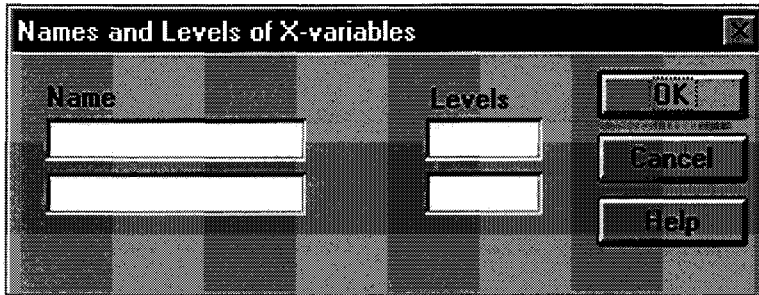
$$Y = \text{vakio} + c1 * \text{muuttuja}_1 + c2 * \text{muuttuja}_2 + c2 * \text{muuttuja}_3 + \text{'mahdolliset interaktiot'}$$

saataisiin x-muuttujien lukumääräksi kolme, joka laitettaisiin kuvassa 3.1 olevaan tilaan.

Kun lukumäärä on annettu kokonaislukuna, voidaan klikata OK-painiketta, josta päästään seuraavaan vaiheeseen.

4 X-muuttujien nimet ja tasojen lukumäärät

Seuraavassa vaiheessa asetetaan valitulle määrälle x-muuttujia nimet ja tasojen lukumäärät. Tämä tapahtuu kuvan 4.1 mukaisessa ympäristössä.



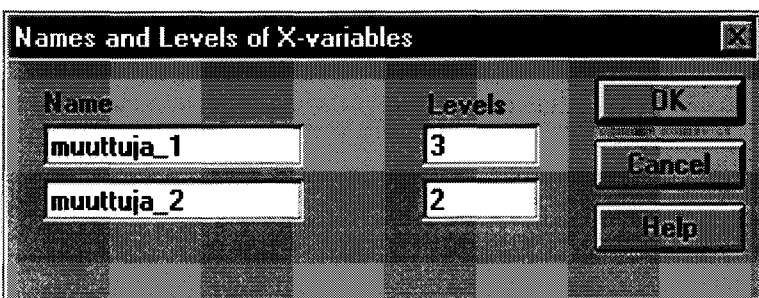
Kuva 4.1 Nimien ja tasojen määrän asetus

Kohtaan *Name* tulee kirjoittaa x-muuttujan nimi ja kohtaan *Levels* vastaavan muuttujan tasojen lukumäärä.

Oletetaan, että käytettävä malli on muotoa:

$$Y = 10 + 1.3 * \text{muuttuja}_1 + 1.7 * \text{muuttuja}_2 + 0.9 * \text{muuttuja}_1 * \text{muuttuja}_2.$$

Lisäksi oletetaan, että muuttuja_1:llä on kolme tasoa ja vastaavasti muuttuja_2:lla kaksi tasoa. Tällöin voidaan kirjoittaa kuvan 4.2 mukaisesti.

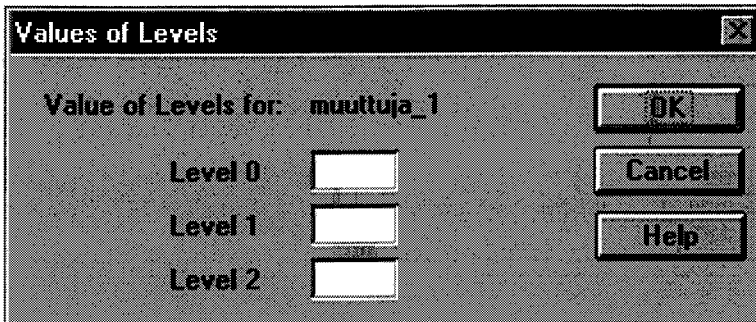


Kuva 4.2 Muuttujien nimet ja tasojen lukumäärät

Huomattavaa on, että tästä eteenpäin tasoja käsitellään siten, että esimerkiksi yllä muuttuja_1:llä on tasot 0,1 ja 2 sekä vastaavasti muuttuja_2:lla tasot 0 ja 1. OK-painiketta klikkaamalla päästään jälleen eteenpäin.

5 Muuttujien eri tasojen numeeriset arvot

Eri muuttujien tasojen lukumäärien valinnan jälkeen on eri tasoille annettava myös numeeriset arvot. Se tapahtuu kuvan 5.1 mukaisessa ympäristössä.



Kuva 5.1 Tasojen numeeristen arvojen syöttö

Oletetaan edelleen tilanne, jossa malli on muotoa:

$$Y = 10 + 1.3*muuttuja_1 + 1.7*muuttuja_2 + 0.9*muuttuja_1*muuttuja_2.$$

Lisäksi voidaan olettaa, että muuttuja_1:llä on kolme tasoa siten, että:

taso 0 = 0 <yksikkö>

taso 1 = 5.3 <yksikkö>

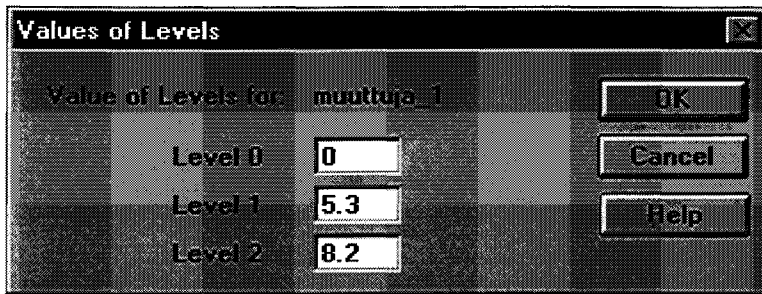
taso 2 = 8.2 <yksikkö>.

Vastaavasti muuttuja_2:lla on kaksi tasoa siten, että:

taso 0 = 0 <yksikkö>

taso 1 = 2.5 <yksikkö>.

Tällöin voidaan kirjoittaa kuvan 5.2 mukaisesti. Ohjelma sijoittaa *Value of Levels for* -kohdan perään sen muuttujan, jolle kulloisellakin kerralla annetaan tasojen arvot.



Kuva 5.2 Tasojen numeeristen arvojen syöttö

OK-painikkeen klikkauksen jälkeen siirrytään muuttuja_2:een ja annetaan tälle vastaavat arvot. Näin jatketaan, kunnes kaikilla muuttujilla on jokaista tasoa vastaavat numeeriset arvot.

6 Malliin tulevien termien valinta

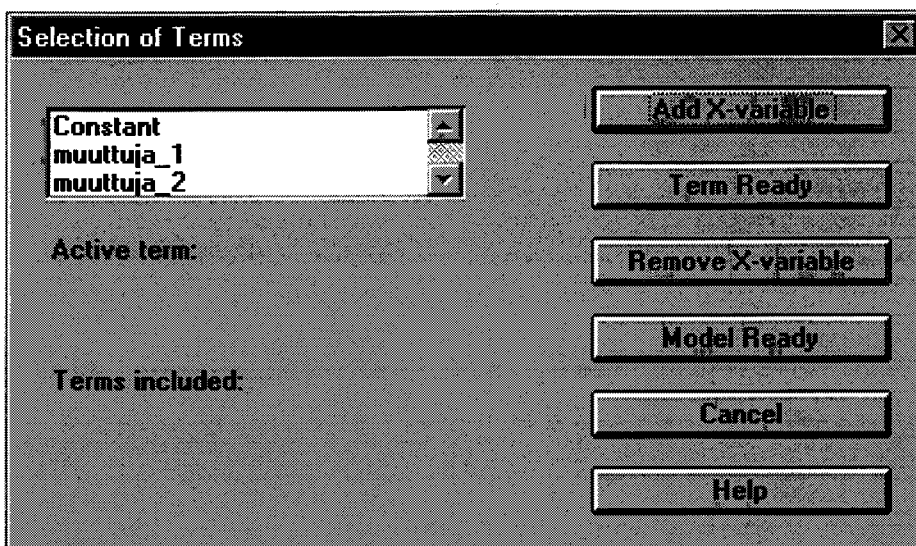
Mallin syöttö tapahtuu antamalla siihen yksitellen mukaan tulevat termit. Huomattavaa on, että seuraavissa kuvissa aiemmin käytetty 'vakio'-sana on korvattu ohjelman toimesta sanalla 'Constant'. Malliin voidaan lisätä yksittäisiä termejä eli niin sanottuja päävaikutustermejä sekä yhdysvaikutustermejä eli interaktioita. Jos vahingossa annetaan turhia termejä, saa ne poistettua helposti.

6.1 Vakio- ja päävaikutustermien lisäys

Oletetaan edelleen malli:

$$Y = 10 + 1.3*muuttuja_1 + 1.7*muuttuja_2 + 0.9*muuttuja_1*muuttuja_2.$$

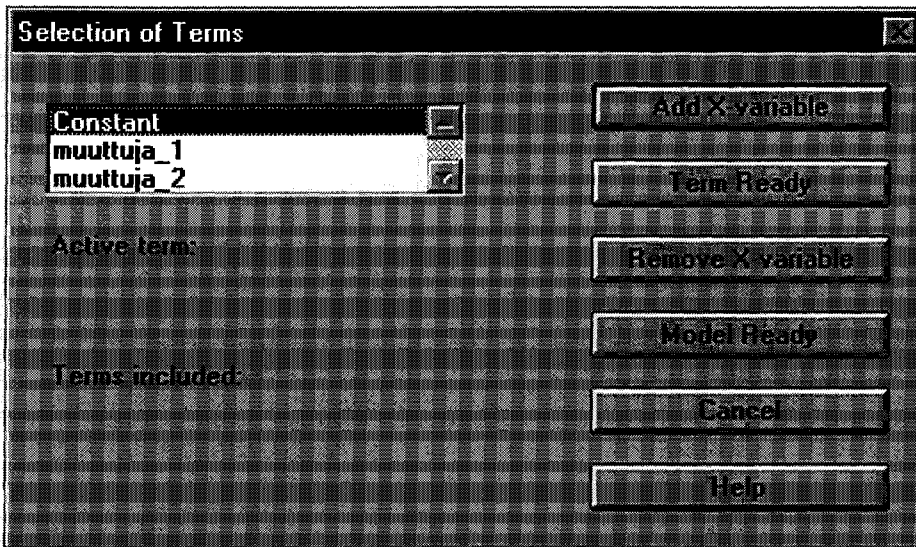
Tästä päästään kuvan 6.1.1 mukaiseen näkymään.



Kuva 6.1.1 Termien valinta

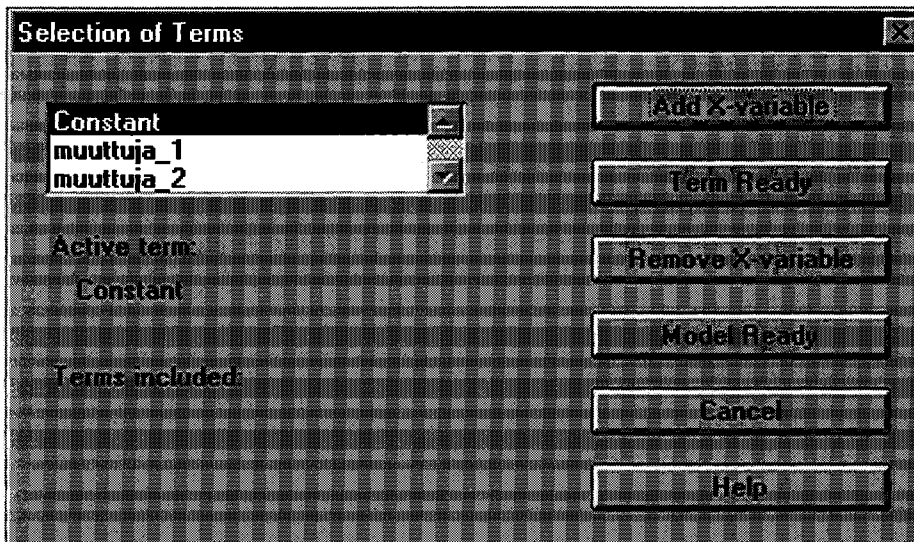
Kuvan 6.1.1 vasemmassa ylälaudassa olevassa ikkunassa näkyy kerrallaan kolme termiä. Jos termien lukumäärä on suurempi kuin tämä, saadaan ne näkyviin nuolilla, jotka ovat saman ikkunan oikeassa reunassa. Vakiotermi saadaan mukaan siten, että aluksi klikataan 'Constant'-sanaa. Tällöin sanan tausta

muuttuu kuvan 6.1.2 mukaisesti tummaksi.



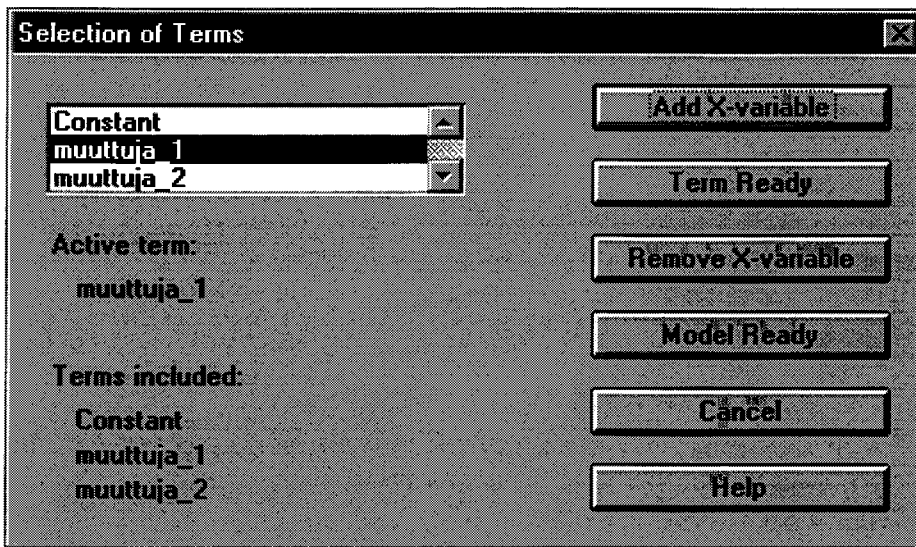
Kuva 6.1.2 Vakiotermin aktivointi

Tämän jälkeen voidaan joko kaksoisklikata tätä sanaa tai vaihtoehtoisesti painaa *Add X-variable* -painiketta, jolloin termi siirtyy kohtaan *Active term* kuvan 6.1.3 mukaisesti.



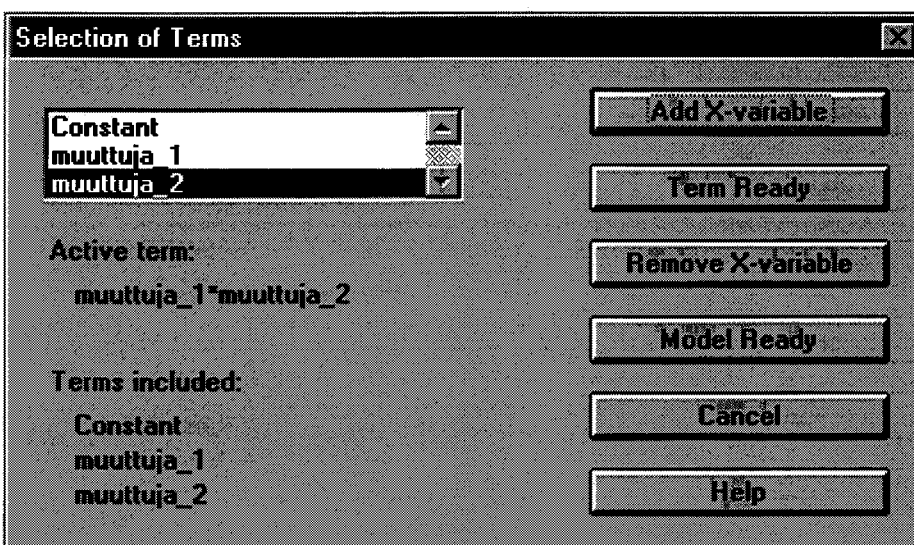
Painike vie muuttujan *Active Term* -kohtaan. Ei siis vielä itse malliin. Malliin siirto tapahtuu *Term Ready* -painikkeella.

Kuva 6.1.3 Vakiotermin siirto



Kuva 6.2.1 Ensimmäisen interaktioon tulevan termin valinta

Tämän jälkeen valitaan samaan yhdysvaikutustermiin myös toinen yksittäinen termi eli muuttuja_2. Tämä tapahtuu yksinkertaisesti siten, että valitaan muuttuja_2 joko kaksoisklikkaamalla sitä tai klikkaamalla sitä ja valitsemalla *Add X-variable*. Tämän jälkeen kohtaan *Active term* ilmestyy yhdysvaikutustermi muodossa $\text{muuttuja}_1 * \text{muuttuja}_2$. Tämä on esitetty kuvassa 6.2.2.

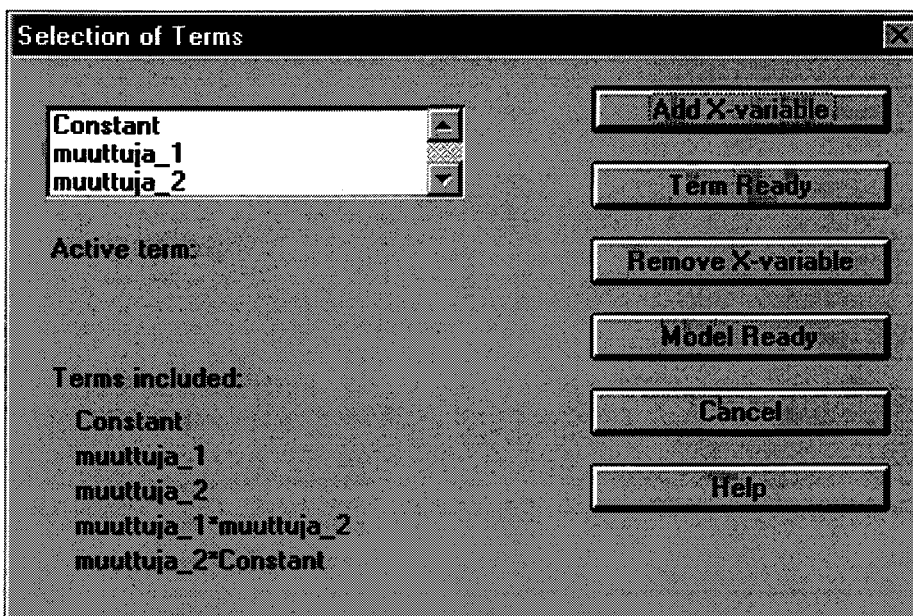


Kuva 6.2.2 Interaktiotermin valinta

Tämän jälkeen termi saadaan *Terms included* -kohtaan klikkaamalla *Term Ready* -painiketta. Näin on saatu suoritettua esimerkin mukainen termien valinta. Kun malli on valmis, klikataan *Model Ready* -painiketta.

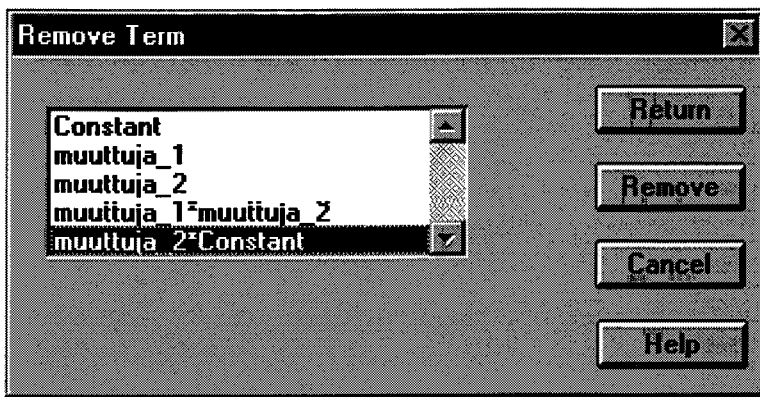
6.3 Mallissa olevien termien poisto

Kun on valittu malliin tulevat termit ja havaitaan, että joukossa on turhiakin termejä, voidaan ne helposti poistaa. Tähän tilanteeseen on päädytty kuvassa 6.3.1, jossa on vahingossa lisätty malliin interaktiotermi muuttuja_2*Constant.



Kuva 6.3.1 Turhien termien poisto

Termien poisto onnistuu *Remove X-variable* -painikkeella. Tällöin eteen avautuu *Remove Term* -ikkuna, josta aluksi valitaan termi klikkaamalla sitä ja tämän jälkeen valittu termi poistetaan klikkaamalla *Remove*-painiketta. Tämä on esitetty kuvassa 6.3.2.

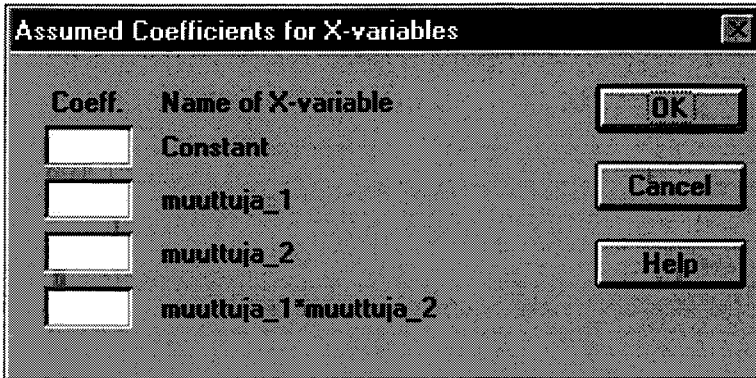


Kuva 6.3.2 Valitun termin poisto

Remove Term -ikkunasta poistutaan klikkaamalla *Return*-painiketta. Tämän jälkeen voidaan jatkaa malliin tulevien termien syöttöä tai jos malli on valmis, klikataan *Model Ready* -painiketta.

7 Oletettujen kertoimien syöttö

Regressioanalyttisessä lähestymistavassa tutkijalla on jokin oletus mallissa olevista kertoimista. Näiden kertoimien syöttö tapahtuu kuvan 7.1 mukaisesti.



Coeff.	Name of X-variable
<input type="text"/>	Constant
<input type="text"/>	muuttuja_1
<input type="text"/>	muuttuja_2
<input type="text"/>	muuttuja_1*muuttuja_2

OK
Cancel
Help

Kuva 7.1 Oletettujen kertoimien syöttö

Oletetut kertoimet kirjoitetaan *Coeff*-kohtaan. Jos jostakin syystä halutaan asettaa kertoimeksi 0, on sekin kirjoitettava numeroin eli pelkkä tyhjä ruutu ei riitä.

8 Design-matriisin teko

Design-matriisi kuvaa kokeen järjestelyä. Sen avulla voidaan määrätä, millaiset termit ovat tietyillä arvoilla yhtä aikaa mallissa mukana. Aiempana olleesta esimerkistä saataisiin vaikkapa seuraavanlainen koejärjestely:

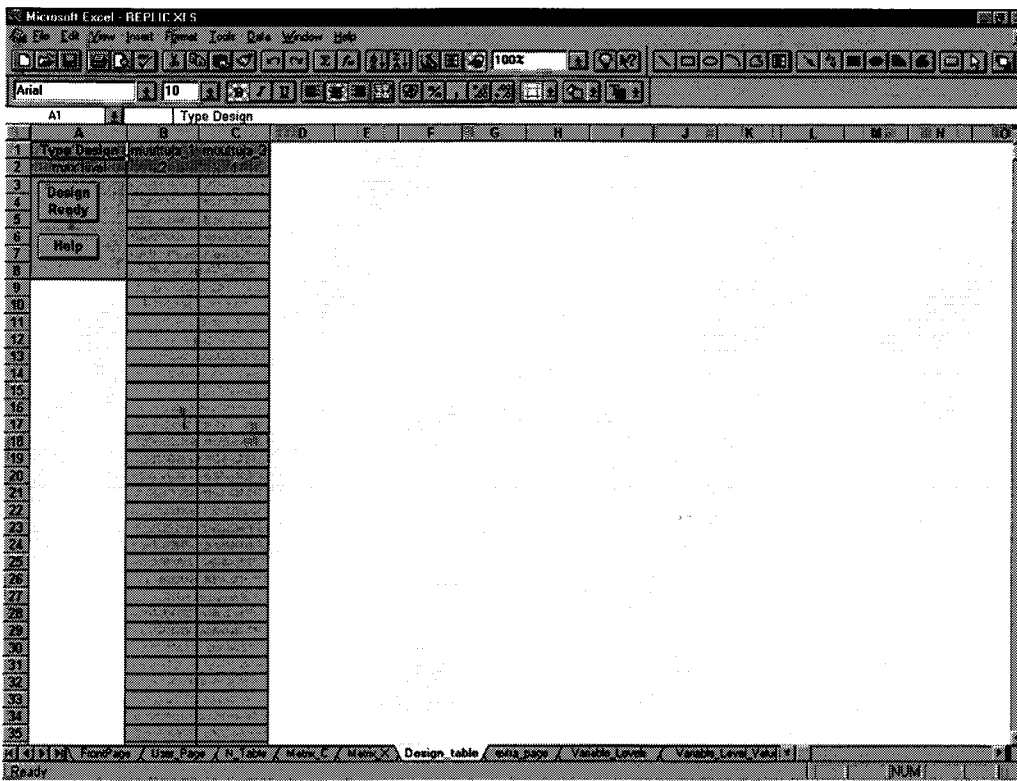
		muuttuja_1	muuttuja_2	muuttuja_1*muuttuja_2
koe nro	1.)	taso 0	taso 0	taso 0 * taso 0
- “ -	2.)	taso 0	taso 1	taso 0 * taso 1
- “ -	3.)	taso 1	taso 0	taso 1 * taso 0
- “ -	4.)	taso 1	taso 1	taso 1 * taso 1
- “ -	5.)	taso 2	taso 0	taso 2 * taso 0

Eli edellä kokeessa numero 1 otetaan muuttuja_1:stä taso 0 ja vastaavasti muuttuja_2:sta myös taso 0.

Lisäksi interaktiotermiin tulee näiden tasojen tulo. Itse tasojen numeerisista arvoista ei tässä vaiheessa tarvitse tietää mitään. Samoin esimerkiksi kokeessa numero 5 otetaan muuttuja_1:stä taso 2 ja muuttuja_2:sta taso 0 ja interaktioon näiden tulo. Yllä oleva koe ei ole täydellinen, koska siitä puuttuu sellainen yhdistelmä, jossa olisi muuttuja_1:n taso 2 yhtä aikaa muuttuja_2:n tason 1 kanssa.

Tasojen syöttö tapahtuu kuvassa 8.1 näkyvään tauluun, jossa sarakkeesta B lähtien on muuttujien nimet.

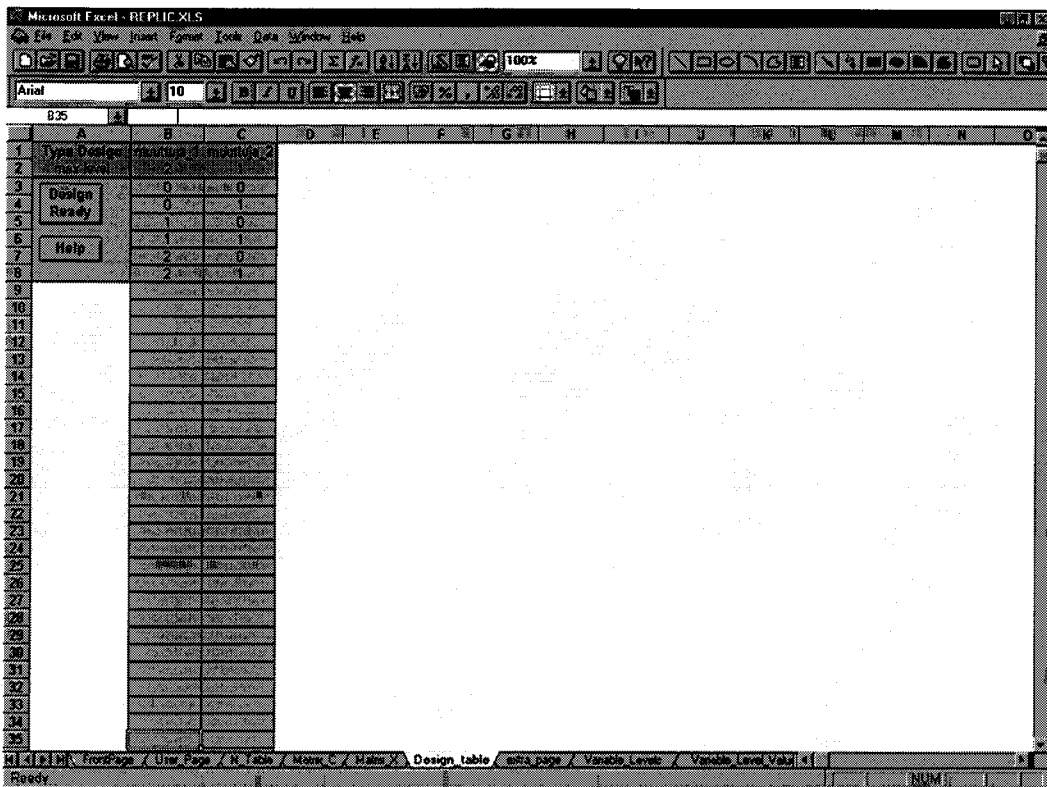
Muuttujan nimen alla on vastaavan muuttujan tason maksimiarvo (= *max level*), toisin sanoen sallitut arvot ovat kokonaislukut välillä 0 ja kyseinen maksimiarvo.



Kuva 8.1 Design-taulun runko

Design-tauluun voidaan syöttää sekä epätäydellisiä kokeita, eli joista puuttuu jokin tietty yhdistelmä, tai toisaalta 'ylisuuria' kokeita eli joissa jokin yhdistelmä esiintyy enemmän kuin kerran. Ehtoina on, että taulussa ei ole 'reikiä' eli mallit ovat järkeviä. Design-tauluun mahtuu sata eri kombinaatiota.

Kuvassa 8.2 on esitetty aiemmin olleelle esimerkkimallille koejärjestely, jossa on mukana tasojen kaikki eri kombinaatiot.

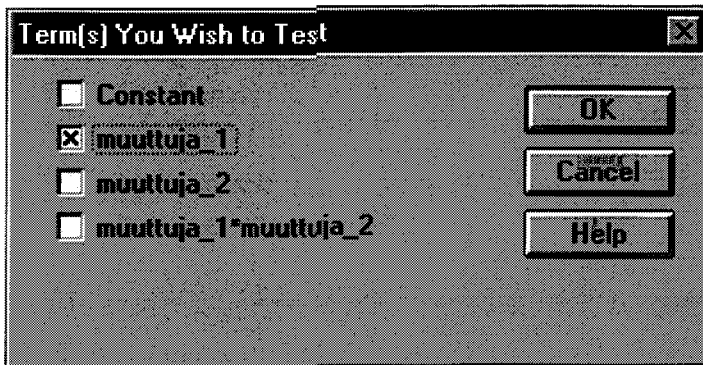


Kuva 8.2 Design-taulu esimerkkimallille

Design-taulun ollessa valmis voidaan klikata *Design Ready* -painiketta, joka muodostaa annetun design-taulun perusteella koetta kuvaavan matriisin.

9 Testattavaksi tulevien termien valinta

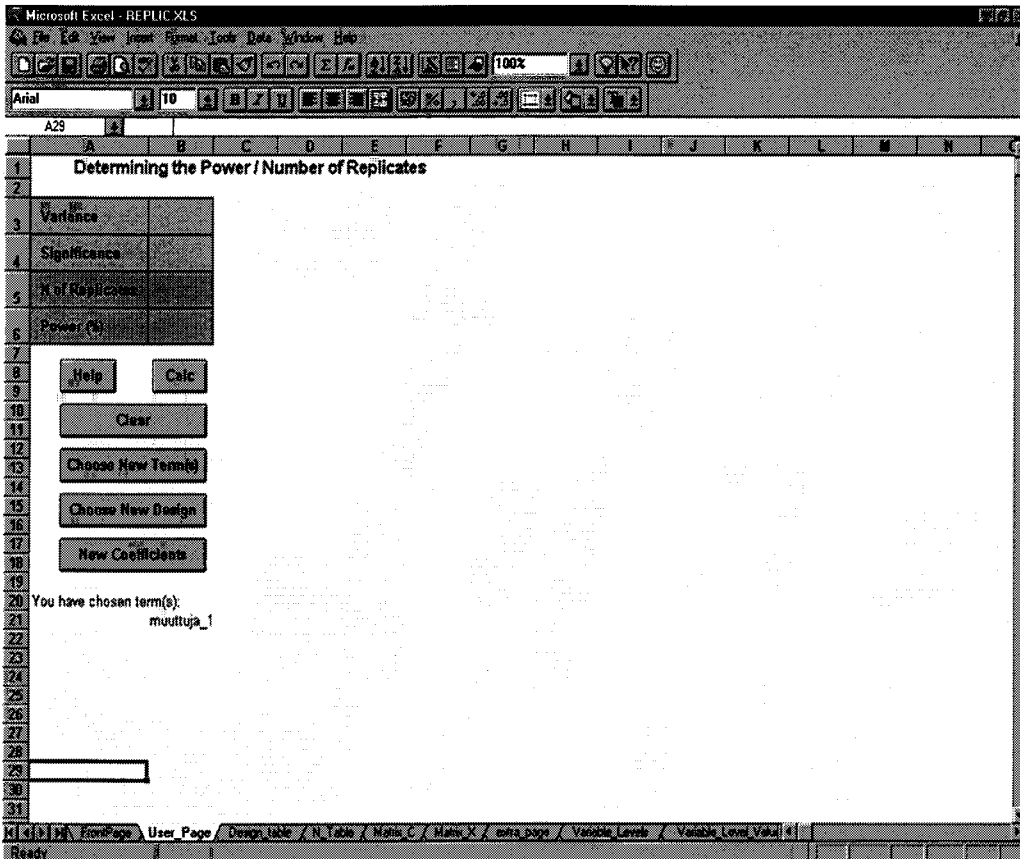
Termit, joita myöhemmässä vaiheessa halutaan testata, valitaan kuvan 9.1 mukaisesta ikkunasta. Valinta tapahtuu yksinkertaisesti klikkaamalla haluttua termiä, jolloin sen viereen ilmestyy rasti. Valinnan saa peruttua klikkaamalla uudelleen kyseistä termiä. Termejä voi valita myös useampia kuin yhden. Kuvassa 9.1 on esimerkin vuoksi valittu muuttuja_1.



Kuva 9.1 Testattavien termien valinta

10 Toistojen määrän ja voimakkuuden laskeminen

Kun tiedetään vastemuuttujan varianssi ja haluttu testin merkitsevyystaso, voidaan tällä ohjelmalla laskea joko toistojen avulla voimakkuus tai toisaalta voimakkuuden avulla toistojen määrä. Kuvassa 10.1 on esitetty sivu, jonka avulla tämä pystytään hoitamaan.



Kuva 10.1 Toistojen määrän ja voimakkuuden laskeminen

Variance-riville annetaan arvio vastemuuttujan varianssista. Kohtaan *Significance* kirjoitetaan tulevan testin merkitsevyystaso, esimerkiksi 0.01 tai 0.05.

Vaihtoehtoisesti voidaan antaa joko *N of Replicates* eli toistojen määrä tai *Power (%)* eli voimakkuus

prosentteina. Tämän jälkeen klikataan *Calc*-painiketta ja ohjelma laskee arvon tyhjäksi jääneeseen kohtaan ja selvyuden vuoksi alleviivaa ja kursivoi laskemansa arvon. Kun on laskettu yksi sarake valmiiksi, ohjelma tekee automaattisesti seuraavan sarakkeen ja kopioi siihen edellisellä kerralla käytetyn varianssin ja merkitsevyytason. Näitä saa toki muuttaa, jos haluaa. Tähän riittää, että kirjoittaa niiden päälle uuden haluamansa arvon. Samalla sivulla voi laskea sekä toistoja että voimakkuutta eli yhdessä sarakkeessa voi laskea toistot ja toisessa sarakkeessa voimakkuuden.

Painikkeiden alapuolella näkyy termi tai termit, jotka on valittu.

Kuvassa 10.1 olevat painikkeet:

<i>Calc</i>	-	lasketaan haluttu arvo
<i>Clear</i>	-	puhdistaa taulun vanhoista arvoista
<i>Choose New Term(s)</i>	-	valitaan uudet testattavat termit
<i>Choose New Design</i>	-	valitaan uusi koejärjestely
<i>New Coefficients</i>	-	valitaan uudet kertoimet

10.1 Toistojen määrän arviointi voimakkuuden avulla

Kuvassa 10.1.1 on laskettu toistojen määriä tietyillä voimakkuuden arvoilla. Huomattavaa on, että annettaessa tietty voimakkuus, ohjelma hakee sen toistojen määrän, jolla annettu arvo ylittyy. Eli lasketun toistojen määrän voidaan sanoa olevan se vähimmäismäärä, jolla kriteerinä oleva voimakkuus saavutetaan.

Determining the Power / Number of Replicates

Variance	4.00	4.00	4.00	4.00	4.00
Significance	0.05	0.05	0.05	0.05	0.05
N of Replicates	2	2	2	3	
Power (%)	40.0000	61.0000	80.0000	99.9000	

Buttons: Help, Calc, Clear, Choose New Term(s), Choose New Design, New Coefficients

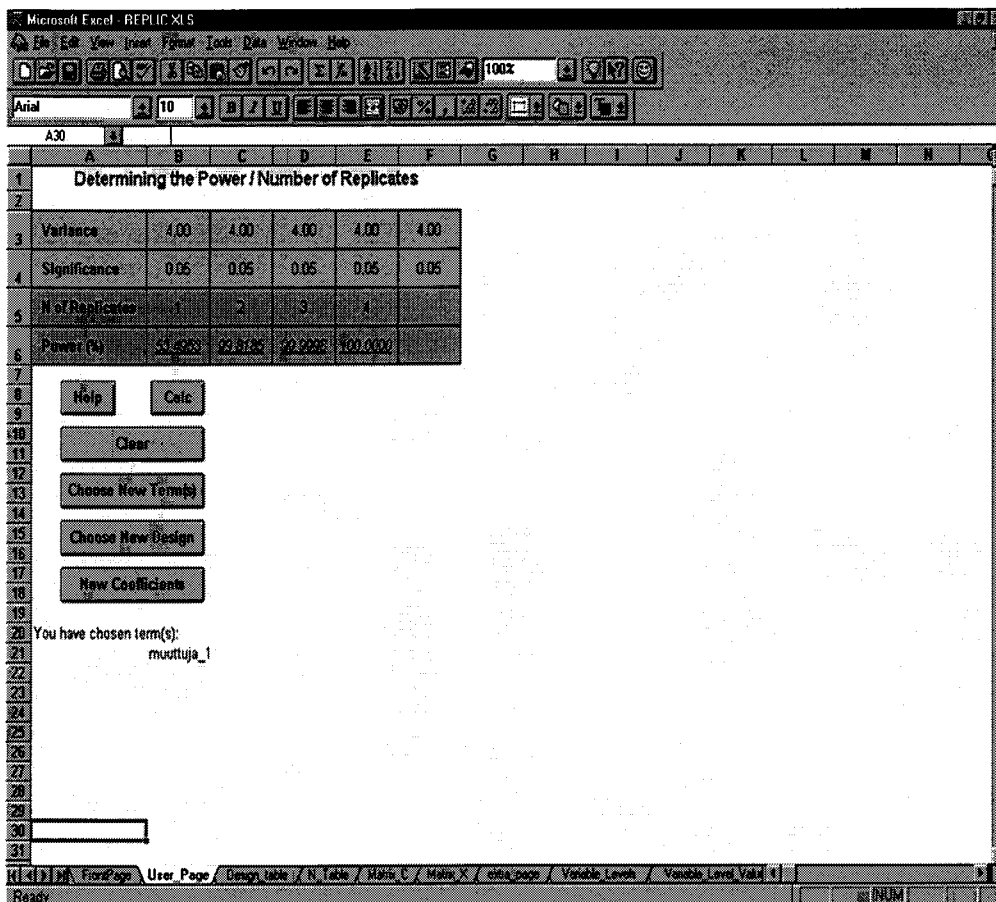
You have chosen term(s):
muuttuja_1

Kuvasta 10.1.1 havaitaan, että esimerkiksi vähintään 80 prosentoin voimakkuuteen päästään toistojen määrällä 2 ja 99.9 prosentoin voimakkuus saavutetaan 3 toistolla.

10.2 Voimakkuuden arvon laskeminen toistojen määrän avulla

Voimakkuus lasketaan samalla periaatteella kuin edellisessä toistotkin. Eli kohtaan *N of Replicates* sijoitetaan toistojen määrä ja tämän antama voimakkuus selviää *Calc*-painikkeen painamisen jälkeen *Power (%)* riviltä alleviivattuna ja kursivoituna.

Kuvassa 10.2.1 on esitetty muutamia voimakkuuden arvoja esillä olleelle mallille samalla termien valinnalla kuin kappaleessa 10.1.



Kuva 10.2.1 Voimakkuuden laskeminen toistojen avulla

Kuvasta 10.2.1 voidaan todeta, että esimerkiksi toistojen määrällä 2 päästään tällä mallilla ja valitulla design-tilillä 99.8 prosentin voimakkuuteen.

11 Työn talletus

Jos työ halutaan tallettaa, voidaan se tehdä aivan kuten talletettaisiin mikä tahansa Excel-työkirja. Ohjelma on tehty *Read-only*-periaatteella, joten talletus tapahtuu automaattisesti eri nimelle. Kun työ on talletettu, voidaan se avata tavalliseen tapaan ja käyttää esimerkiksi samaa mallia uusilla kertoimilla tai koejärjestelyillä.

Tilastotoimen menetelmien maisteriohjelman pro gradu -tutkielma sarja

1. Salmikuukka, J. (1997) Aikasarjojen perusrakennemalleista ja niiden soveltaminen Jyväskylän kaukolämmön kulutuksen analysointiin ja ennustamiseen. (76 s., 1 liite) Jyväskylän Energia Oy, Jyväskylä
2. Yrjölä, T. (1997) Lasten päivähoidon tuottavuusvertailu suurissa kaupungeissa DEA-menetelmällä. (72 s., 2 liitettä) Jyväskylän kaupungin terveystoimi, Jyväskylä
3. Ainiala, N. (1997) Helsingin osa-alueiden työvoimatilastojen estimointi pienaluetekniikalla valtakunnallisesta työvoimatutkimuksesta. (73 s., 3 liitettä) Helsingin kaupungin tietokeskus, Helsinki
4. Puhakka, E. (1997) Kiintiöpoiminnan tilastolliset ominaisuudet pk-yritysbarometritutkimuksessa. Sovelluksena 2/1996 aineisto (82 s., 2 liitettä) (salainen) Tietoykkönen, Jyväskylä
5. Salonen, R. (1997) Muutoksen ja tason estimointi ratatoivassa paneeliaineistossa eri estimaattoreiden avulla. Sovellus työvoimatutkimuksen aineistoon. (39 s, 4 liitettä) Tilastokeskus, Helsinki
6. Kunttu, S. (1997) Alueellisen teollisuustuotannon volyymindeksin estimointi Etelä-Pohjanmaalle. (103 s.) Tilastokeskus, Seinäjoki
7. Koponen, M. (1997) Epätäydelliseen otantakehikkoon perustuvan yritysaineiston estimointi. Käsiteanalyysiä ja soveluksia Tilastokeskuksen järjestäytymättömiä yrityksiä koskevaan palkkatiedusteluun. (56 s., 5 liitettä) Tilastokeskus, Helsinki
8. Kainulainen, P. (1997) Toimialatilastot sanomalehdessä: Monimuuttujaisen rekisteriaineiston havainnollistaminen ja tuotteistaminen (62 s., 6 liitettä) Sanomalehtien liitto, Helsinki
9. Storfors, S. (1998) Toistojen määrän arviointi kokeellisessa tutkimuksessa. (34 s., 3 liitettä) Cultor Oy, Helsinki