

*M a a t a l o u d e n
t u t k i m u s k e s k u k s e n
j u l k a i s u j a*

S A R J A B

16

Lauri Jauhiainen

**Virallisten lajikekokeiden
tulosten
laskentaperusteet**

Lauri Jaubainen

Virallisten lajikekokeiden tulosten laskentaperusteet

Maatalouden tutkimuskeskus

ISBN 951-729-524-3

ISSN 1238-9943

Copyright

Maatalouden tutkimuskeskus

Lauri Jauhiainen

Julkaisija

Maatalouden tutkimuskeskus, 31600 Jokioinen

Jakelu ja myynti

Maatalouden tutkimuskeskus, tietopalveluyksikkö, 31600 Jokioinen

Puh. (03) 4188 7502, telekopio (03) 418 8339

Painatus

Yliopistopaino, 1998

Sisäsivujen painopaperille on myönnetty pohjoismainen joutsenmerkki.
Kansimateriaali on 75-prosenttisesti uusiokuitua.

Tiivistelmä

Avainsanat: luottamusväli, parivertailu, suhdeluku, tilastolliset testit, voimakkuusfunktio.

Maatalouden tutkimuskeskuksen koordinoiman peltokasvien virallisen lajikekoetoinnin puitteissa suoritetaan vuosittain yli sata yksittäistä kenttäkoetta. Kokeiden pääasiallisena tarkoituksena on kerätä aineistoa laajempien koesarjakohtaisten analyysien tekemiseksi. Tällaiset koesarjakohdalliset analyysit tehdään vuosittain käyttäen kahdeksana edellisenä vuonna suoritettuja kenttäkokeita. Analyysien tulokset julkaistaan maatalouden tutkimuskeskuksen julkaisusarjassa A. Jotta analyysien tulosten oikea tulkitseminen olisi mahdollista, täytyy tulkitsijan tuntea analyysissä käytettävä tilastollinen malli ja sen sisältämät oletukset. Lisäksi tärkeää on tietää, mitä kaikki raportoidut tulokset tarkoittavat ja miten ne on laskettu. Tässä raportissa kuvataan tar-

kasti tällä hetkellä käytettävä testaustapa, joka tunnetaan lajikekoetoinnassa nimellä parivertailu. Erityistä huomiota on keskitetty sadon suhdelukuihin, joita aikaisemmin on usein tulkittu väärin perustein. Usein tulkitaan myös virheellisesti tilastollisesti merkitsevän eron tarkoittavan samaa kuin käytännön kannalta merkitsevä ero. Näiden kahden merkitsevyyden välisen eron tunteminen edesauttaa oikeanlaisten johtopäätösten tekemistä. Tässä raportissa annetaan myös neuvoja siitä, kuinka parivertailutulosten luotettavuutta tulee arvioida. Lopuksi esitetään tietopalveluyksikössä tehtävät koeaineiston laadun tarkastukset. Laaduntarkastuksella on tarkoitus varmistaa käytetyn aineiston käyttökelpoisuus ja sitä kautta taata tulosten oikeellisuus.

Sisällys

Tiivistelmä	3
1 Johdanto	5
2 Tilastollinen testaus	5
2.1 Koeaineisto ja poikkeamien estimointi	5
2.2 Poikkeamien tilastollinen testaus	7
2.3 Esimerkki	8
2.4 Testien voimakkuus	9
3 Tulosten tulkinta	12
3.1 Mallin vaikutus	12
3.2 Kokeiden lukumäärä	12
3.3 Tilastollinen merkitsevyys	12
3.4 Suhdeluku	14
3.5 Monivuotiset kasvit	15
4 Parivertailutulosten raportointi	16
4.1 Poikkeamasivut	16
4.2 Vuosittaiset vertailut	16
4.3 Muut vertailut	17
5 Parannusehdotuksia parivertailuun	17
5.1 Luottamusvälit	18
5.2 Parametrittomat vaihtoehdot	19
6 Laaduntarkastus	22
Kirjallisuus	23
Litteet	

1 Johdanto

Kasvilajikelautakunta päättää Suomessa lajikkeiden kauppakelpoisuuden. Maa- ja metsätalousministeriön päätöksen mukaisesti Maatalouden tutkimuskeskuksen tulee avustaa kasvilajikelautakuntaa lajikkeiden viljely- ja käyttöarvon tutkimisessa. Tämän velvoitteen MTT suorittaa tekemällä virallisia lajikekokeita ja valmistamalla koesarjakohtaiset analyysit. Analyysien tulokset esitetään nykyisin maatalouden tutkimuskeskuksen julkaisusarjassa A (Järvi et al. 1995). Vaikka lajikekokeet ja koesarjakohtaiset analyysit olisivat tehty kuinka hyvin tahansa, ei saatujen tulosten käyttöarvo ole kuitenkaan suuri, mikäli tuloksia ei pystytä tulkitsemaan oikein. Suurin syy väärin tulkintoihin on ollut kattavien tulkintaohjeiden puuttuminen. Tämänkaltaisen ohjeen tarpeellisuus on havaittu lajikekoetoinnin laadunkehittämishankkeen yhteydessä ja tämän raportin pääasiallinen tarkoitus onkin toimia koesarjakohtaisten analyysien tulosten tulkintaoppaana.

Lajikkeiden testaaminen on tilastollista tutkimusta. Käytettyä tilastollista menetelmää kutsutaan lajikekoetoinnassa nimellä parivertailu. Jotta saatuja parivertailutuloksia olisi mahdollista tulkita oikein, täytyy tulkitsijan tuntea tyydyttävästi parivertailussa käytetty tilastollinen malli ja sen sisältämät oletukset. Tässä raportissa on pyritty yksinkertaisesti esittämään lajikekoaineistojen tilastollinen analyysi edellyttämättä lukijalta tilastotieteen osaamista. Tulosten tulkinnan kannalta on erityisen tärkeää tietää, mitä kaikki raportoidut luvut tarkoittavat ja kuinka ne on laskettu. Tässä raportissa esitellään kaikki tärkeimmät laskukaavat, joihin lajiketestauksen raportointi perustuu.

Tulosten tulkinta ei aina ole helppoa, vaikka tuntisikin käytetyn tilastollisen mallin ja kaikki käytetyt laskukaavat. Erityisen vaikeata näyttää olevan sadon suhdelukujen tulkinta, sillä niitä tulkitaan usein aivan väärin perustein. Virheellisesti tulkitaan myös tilastollisesti merkitsevän eron tar-

koittavan samaa kuin käytännönkannalta merkitsevä ero. Näiden kahden merkitsevyyden välisen eron huomioiminen edesauttaa oikeanlaisten johtopäätösten tekemistä. Johtopäätöksiä tehtäessä on aina hyvä arvioida myös parivertailutulosten luotettavuutta. Luotettavuuden arvioinnin helpottamiseksi on tässä raportissa annettu muutamia yksityiskohtaisia neuvoja. Lajikevertailussa on myös testattavia ominaisuuksia, joiden analysointi käyttäen parivertailun testiä on erittäin epäluotettavaa. Näiden ominaisuuksien analysointiin on kuitenkin olemassa vaihtoehtoisia menetelmiä, joista muutama toimii lähes parivertailun tavoin, eikä saadun raportin ulkoasusta voi päätellä, mitä tilastollista testiä on käytetty. Nämä vaihtoehtoiset testit on esitelty lyhyesti luvussa 5.2 Parametrittomat vaihtoehdot.

Lopuksi tässä raportissa esitetään tietopalveluyksikössä tehtävät koesarjojen laaduntarkastukset. Laaduntarkastuksella on tarkoitus varmistaa käytetyn aineiston käyttökelpoisuus ja sitä kautta taata tulosten oikeellisuus. Tulosten tulkinnan kannalta ei ole tarpeellista tietää, miten koeaineiston laaduntarkistus on suoritettu, mutta tieto laatukontrollista lisää tulosten luotettavuutta. Koesarjojen laaduntarkastuksen lisäksi jokainen koepaikka suorittaa yksittäisen kokeen laadun tarkastuksen (Järvi et al. 1998).

2 Tilastollinen testaus

2.1 Koeaineisto ja poikkeamien estimointi

Lajiketestauksessa verrataan kerrallaan aina yhtä testattavaa lajiketta yhteen mittarilajikkeeseen. Koeaineistona kussakin vertailussa käytetään testi- ja mittarilajikkeiden havaintoja niistä kokeista, joissa sekä testilajike että mittarilajike ovat olleet samanaikaisesti. Tällöin saatava koeaineisto on Taulukon 1 mukainen.

Taulukko 1. Parivertailussa käytettävä havaintoaineisto.

Koe	testilajik- keen arvo	mittarilajik- keen arvo	lajikkeiden erotus
1. koe	t_1	m_1	$d_1=t_1-m_1$
2. koe	t_2	m_2	$d_2=t_2-m_2$
.	.	.	.
.	.	.	.
n. koe	t_n	m_n	$d_n=t_n-m_n$
keskiarvo:	\bar{t}	\bar{m}	\bar{d}

Eri vertailuissa käytettyjen lajikekokeiden lukumäärä (=n) vaihtelee pääsääntöisesti siksi, ettei jokaisessa yksittäisessä lajikekokeessa ole viljelty kaikkia vertailuissa mukana olevia lajikkeita. Kaikkia testattavia ominaisuuksia ei ole myöskään mitattu jokaisessa kokeessa. Kolmanneksi on mahdollista, että yksittäisiä havaintoja on joko kokeen suorituksen aikana tai havaintoaineiston laaduntarkistuksessa poistettu aineistosta. Nämä kolme seikkaa johtavat siihen, että analyysissä käytettävien kokeiden lukumäärä vaihtelee sekä testilajikkeiden että yhden testilajikkeen eri ominaisuuksien välillä.

Lajiketestauksessa kiinnostaa aina kahden lajikkeen eli testi- ja mittarilajikkeen välisen todellisen eron suuruus. Tätä tuntematonta eroa merkitään jatkossa symbolilla Δ . Lajiketestausta suoritettaessa lasketaan Taulukossa 1 esitetyistä erotuksista d_1, \dots, d_n tavallinen aritmeettinen keskiarvo, jota merkitään jatkossa symbolilla \bar{d} . Laskettu \bar{d} on harhaton estimaatti tuntemattomalle arvolle Δ . Tämä tarkoittaa käytännössä sitä, että \bar{d} kuvaa hyvin testilajikkeen ja mittarilajikkeen välistä tuntematonta eroa.

Virallisten lajikekokeiden tuloraportissa ilmoitetaan aina testilajikkeen poikkeama mittarilajikkeeseen verrattuna. Tämä raportoitu poikkeama on erotusten keskiarvo \bar{d} . Ennen vuoden 1998 kevättä julkais-

tuissa raporteissa ei sadon kohdalla ole ilmoitettu lajikkeiden keskimääräistä eroa, vaan testilajikkeen keskiarvo (= \bar{t}). Pelkän keskiarvon tulostamisesta on kuitenkin luovuttu, koska sitä ei voida yleensä verrata mihinkään arvoon ja se on näin täysin hyödytön.

Olisi tietysti mahdollista valita raporttiin keskiarvon \bar{d} tilalle jokin muu Δ :a hyvin kuvaava luku, kuten esimerkiksi erotusten mediaani tai trimmattu keskiarvo. Trimmattu keskiarvo tarkoittaa keskiarvoa Taulukon 1 erotuksista d_1, \dots, d_n siten, että erotuksista on arvoltaan suurin ja pienin poistettu (tai vaihtoehtoisesti k suurinta ja k pienintä arvoa, missä $k > 1$). Erotusten mediaani on vastaavasti sellainen erotus, jota suurempia ja pienempiä erotuksia on täsmälleen yhtä monta. Mediaani ja trimmattu keskiarvo ovat siinä mielessä tavallista keskiarvoa parempia, ettei niihin vaikuta kovinkaan paljon yhden tai useamman yksittäisen kokeen poikkeavat havainnot. Tavallinen keskiarvo on paras, jos tässä raportissa myöhemmin mainittu erotusten normaalisuusoletus on voimassa. Tällöin keskiarvo antaa keskimäärin tarkimman tuloksen, mitä käytettävissä olevalla havaintomäärällä voidaan saada aikaiseksi.

Satotuloksissa esitetään satoeron lisäksi sadonsuhdeluku (=sl). Suhdeluvun tarkempi tulkinta ja käyttökelpoisuus on selvitetty myöhemmin luvussa 3.4 Suhdeluku.

Raportoitu suhdeluku saadaan laskemalla valituista lajikekokeista ensin testilajikkeen keskiarvo $\bar{t}=(t_1+t_2+\dots+t_n)/n$ ja mittarilajikkeen keskiarvo $\bar{m}=(m_1+\dots+m_n)/n$, missä n on niiden kokeiden lukumäärä, jossa kumpikin lajike on ollut samanaikaisesti. Suhdeluku on näiden kahden keskiarvon \bar{t} ja \bar{m} osamäärä kerrottuna sadalla eli

$$s/\bar{m} = 100 \frac{\bar{t}}{\bar{m}}$$

Lajiketestausten tuloraportissa mittarilajikkeeseen liittyvät arvot lasketaan testattaviin lajikkeisiin verrattuna poikkeavalla tavalla. Jokaisen ominaisuuden kohdalla mittarilajikkeen arvo on mittarilajikkeen havaintojen keskiarvo kaikista niistä kokeista, joissa mittarilajike on ollut mukana. Jos mittarilajikkeen keskiarvoon lisätään testilajikkeen ja mittarilajikkeen välinen erotus, ei saada testilajikkeen keskiarvoa. Tämä johtuu siitä, ettei yksikään testilajike ole ollut yleensä kaikissa niissä kokeissa, joiden perusteella mittarilajikkeen keskiarvo on laskettu.

2.2 Poikkeamien tilastollinen testaus

Tilastollisten testien avulla tutkitaan kiinnostavien, lajikkeita koskevien väittämien (hypoteesien) paikkansapitävyyttä. Testauksessa on aina ennalta asetettu hypoteesi, jonka todenperäisyyttä halutaan tutkia. Tilastolliset hypoteesit kirjoitetaan parametrisissä malleissa siten, että ne ovat tuntematonta parametria koskevia väitteitä. Lajiketestauksessa tämä tuntematon parametri on Δ , mikä kuvaa kahden lajikkeen välistä todellista eroa. Yleensä esitetään kaksi toisensa poissulkevaa väitettä H_0 ja H_1 . Ensimmäistä väitettä kutsutaan nollahypoteesiksi ja jälkimmäistä väitettä kutsutaan vastahypoteesiksi tai vaihtoehdohypoteesiksi. Lajiketestauksessa käytetty nollahypoteesi on aina seuraava:

$H_0: \Delta = 0$, eli testilajike ja mittarilajike ovat testattavan ominaisuuden osalta samanlaisia.

Vastahypoteesia muodostettaessa on otettava huomioon se, millaisia poikkeamia nollahypoteesista voi ilmetä. Koska poikkeamien suunnasta ei ole välttämättä mitään ennakkotietoa, on lajiketestauksessa valittu ns. kaksisuuntainen testi, jolloin vastahypoteesi on seuraava:

$H_1: \Delta \neq 0$, eli testilajike ja mittarilajike eroavat testattavan ominaisuuden osalta.

Lajiketestauksessa käytetään erotuksen tilastollista merkitsevyyttä tutkittaessa ns. kaltaistettujen parien t -testiä. Käytetty testi perustuu oletukseen, että Taulukossa 1 lasketut erotukset d_1, \dots, d_n ovat satunnaisotos normaali-jakaumasta, jonka odotusarvo on Δ ja varianssi σ^2 (merkitään symbolein $N(\Delta, \sigma^2)$). Varianssin σ^2 estimaatiksi lajiketestauksessa on luonnollista valita erotuksista d_1, \dots, d_n laskettu otosvarienssi, jota merkitään jatkossa symbolilla σ_d^2 . Otoskeskihajonta saadaan ottamalla otosvarienssista neliöjuuri. Tätä hajontaa merkitään jatkossa symbolilla σ_d . Käytetty t -testi vaatii lisäksi, että erotukset d_1, \dots, d_n ovat keskenään riippumattomia. Tällöin myös kokeiden tulee olla keskenään riippumattomia.

Vaihtoehtoisesti voidaan myös merkitä, että $d_i = \Delta + e_i$, $i = 1, \dots, n$. Testin oletukset voidaan tällöin kirjata siten, että e_1, \dots, e_n ovat satunnaisotos normaali-jakaumasta $N(0, \sigma^2)$ ja arvot e_1, \dots, e_n ovat keskenään riippumattomia. Arvoja e_1, \dots, e_n kutsutaan tällöin jäännöksiksi tai jäännösvirheiksi.

Tilastollista testausta varten lajikekoearineisto voidaan tiivistää. Arvojen n , σ_d^2 ja \bar{d} avulla saadaan laskettua tuloraportissa esitetyt lajikkeiden väliset erot ja erojen tilastollinen merkitsevyys. Näiden kolmen arvon avulla tehdään päätös joko hyväksyä tai hylätä nollahypoteesi H_0 . Jos nollahypoteesi hylätään, uskotaan tällöin vastahypoteesin olevan tosi eli lajikkeiden välinen ero on

tällöin tilastollisessa mielessä merkitsevä. Tämä johtopäätös tehdään laskemalla t-testisuure, jonka kaava on

$$t = \frac{\bar{d}}{\sigma_d / \sqrt{n}} \quad (1)$$

ja testisuure noudattaa likimain t-jakaumaa vapausasteella n-1, kun nollahypoteesi H_0 tosi. Esimerkiksi t-jakauman taulukosta katsomalla nähdään, onko laskettu t-testisuureen arvo niin suuri, että nollahypoteesi voidaan hylätä. Kun n on suhteellisen pieni (<15), olisi käytännössä aika vaivatonta luopua t-jakaumasta ja perustaa t-arvon tulkinta testin tarkkaan hylkäysalueeseen. Hylkäysalueella tarkoitetaan niitä t-testisuureen arvoja, jotka johtavat nollahypoteesin hylkäämiseen. Tällaisen tarkan hylkäysalueen laskukaava on yksinkertainen ja sen on esittänyt mm. Manly (1997, p. 93). Kuitenkaan tämä parannus ei käytännössä toisi mitään hyötyä lajiketestaukseen.

Jos nollahypoteesi hylätään, otetaan samalla riski tehdä väärä johtopäätös. Kyseisen riskin todennäköisyyttä merkitään yleisesti α :lla, joka on samalla testin merkitsevyytaso. Edellä kuvattua väärää johtopäätöstä, jossa oikea nollahypoteesi hylätään, kutsutaan tilastotieteessä tyypin I virheeksi. Mikäli käytetään pelkkää hyväksy/hylkää-testiä, jossa on kiinteä α :n arvo, niin α on tällöin maksimaalinen tyypin I virheen riski. Lajiketestauksessa α :lla on kolme tasoa: 0.05, 0.01 ja 0.001. Tulostuksessa merki-

tään tähdillä sitä α :n tasoa, jolla nollahypoteesi on hylätty. Yksi tähti tulostuksessa merkitsee, että nollahypoteesi on hylätty tasolla $\alpha=0.05$ eli maksimaalinen riski tehdä tyypin I virhe on 0.05 (=5%). Tulostuksessa kaksi ja kolme tähteä vastaavat tilanteita, joissa α on 0.01 ja 0.001 (= 1% ja 0.1%). Jos ei käytettäisi jyrkkää hyväksy/hylkää-testiä, olisi mahdollista laskea t-testisuureta vastaava p-arvo. Tämä p-arvo on todennäköisyys saada vähintään yhtä paljon nollasta eroava arvo kuin nyt saatu erotus \bar{d} on, jos nollahypoteesi on tosi eli lajikkeiden välillä ei olisi eroa. Mikäli p-arvo on tarpeeksi pieni, voidaan päätellä, etteivät lajikekokeista saadut tulokset voi olla nollahypoteesin mukaisesta populaatiosta, vaan vastahypoteesin mukaisesta populaatiosta. Nyt päätöksen lisäksi jää myös p-arvo kuvaamaan riskiä, jolla tämä valinta tehtiin. Lajiketestauksessa ei ole nähty tarpeelliseksi raportoida p-arvoja vaan maksimaalisen tyypin I virheen taso, jossa nollahypoteesi on hylätty.

2.3 Esimerkki

Taulukossa 2 on verrattu mm. hehtolitrapainoa kahden lajikkeen A ja M osalta. Lajikkeista M on valittu mittarilajikkeeksi, jonka tunnuksena on merkintä '=C' (merkitty Taulukossa 2 symbolilla (5)). Lajike A on ollut 40:ssä lajikekokeessa viimeisen kahdeksan vuoden aikana ja lajike M vastaavasti 82 kokeessa, mutta koeaineistossa

Taulukko 2. Lajikekokeiden tulokset kahden lajikkeen (M ja A) osalta. Lajikkeista M on valittu mittarilajikkeeksi.

Lajike	Viljely-	. . .	kpl	lako-	kpl	Tj _p	kpl	H _{1p}	
	vyöhyke			%		g		kg	
M	=C	1-4	. . .	82	20	80	40.3	80	63.9 ←④
A		1-4	. . .	36	-6*	34	3.3***	36	0.8*
	↑							↑	
	⑤							①	② ③

ei ole merkintää hehtolitrapainosta kahden kokeen osalta lajikkeella M. Yhteensä lajikekokeita on tehty testattavan peltokasvin osalta 90 kappaletta näiden kahdeksan vuoden aikana. Lajikkeilla A ja M on aineistossa tulos hehtolitrapainon osalta samassa kokeessa 36 kertaa. Nyt parivertailussa käytetään koeaineistona lajikkeiden A ja M hehtolitrapainoja näistä 36 kokeesta (1). Lajikkeelta A jätetään näin käyttämättä neljän kokeen tulokset ja lajikkeelta M 44 kokeen tulokset.

Kun koeaineisto tiivistetään, saadaan laskettua seuraavat arvot: $n=36$, $\bar{d}=0.8$ kg (2) ja $\sigma_d=1.9$ kg. T-testisuureksi saadaan 2.53 em. kolmen luvun avulla käyttäen kaavaa (1). Nollahypoteesi on, että lajikkeiden A ja M välinen todellinen ero on nolla ja vastahypoteesiksi on valittu, että kyseinen todellinen ero poikkeaa nolasta. Jotta nollahypoteesi voidaan hylätä :n tasolla 0.05, 0.01 tai 0.001, tulisi t-testisuureen arvon olla suurempi kuin 2.03, 2.72 tai 3.59. Arvot on saatu t-jakauman taulukosta 35:llä vapausasteella (=36-1). Koska t-arvo on suurempi kuin 2.03, mutta pienempi kuin 2.72, niin nollahypoteesi hylätään :n tasolla 0.05 ja tulostukseen tämä johtopäätös merkitään laittamalla lajikkeiden välisen eron (=0.8 kg) perään yksi tähti (3). Jos tähtien asemesta haluttaisiin käyttää p-arvoa, olisi likimääräinen p-arvo tässä tilanteessa 0.02. Mittarilajikkeen hehtolitrapainojen keskiarvo 80 kokeessa on merkitty myös tulostukseen (4).

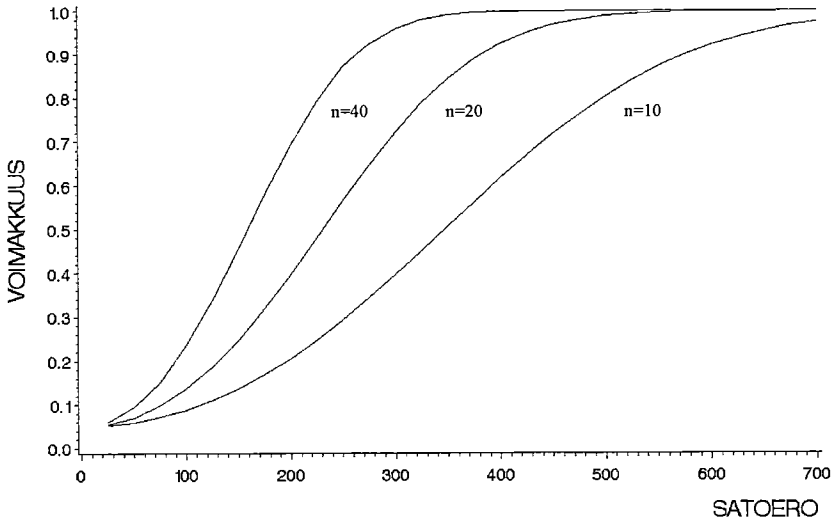
2.4 Testien voimakkuus

Tilastotieteessä tyypin II virheeksi kutsutaan sitä, että väärä nollahypoteesi jää hylkäämättä. Lajiketestauksessa tämä tarkoittaa sitä, että hyväksytään lajikkeiden olevan yhtähyviä, vaikka lajikkeet todellisuudessa eroaisivatkin toisistaan. Yleisesti todennäköisyyttä tehdä tyypin II virhe merkitään β :lla. On huomattavaa, että β :n arvo riippuu tuntemattomasta parametrin Δ . Suuretta $1-\beta$ kutsutaan testin voimakkuudeksi. Voimakkuus on aina nollan ja yhden välillä

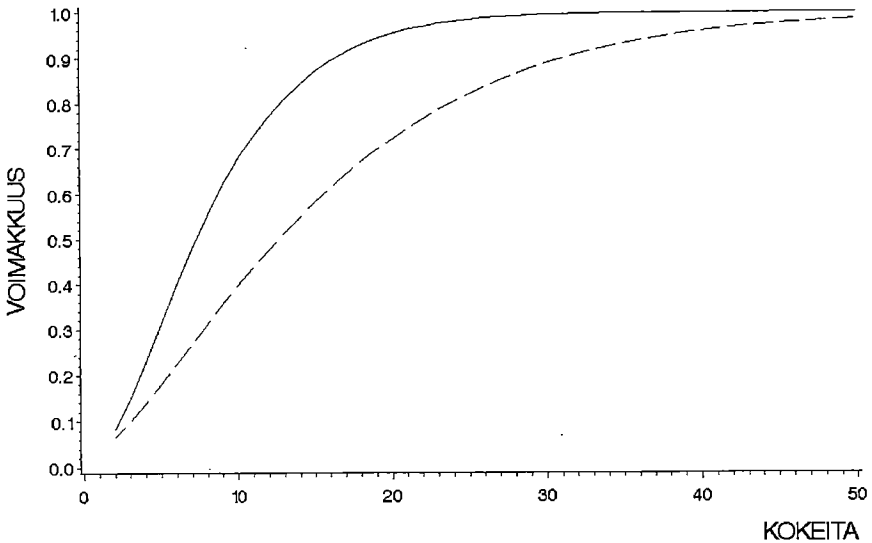
ja testi on sitä voimakkaampi, mitä lähempänä $1-\beta$ on ykköstä. Voimakkuus kuvaankin parhaiten tuntemattoman parametrin Δ funktiona (Kuva 1). Kun tuntemattoman parametrin arvoksi asetetaan nollahypoteesin mukainen arvo (eli lajiketestauksessa nolla), on tällöin voimakkuusfunktion arvo sama kuin testin merkitsevyytaso α .

Koska testin voimakkuus on Δ :n funktio, ei voida antaa vain yhtä arvoa, joka kuvaisi testin voimakkuutta. Yleensä tiedetään etukäteen pienin ero, jolla on käytännön kannalta jotain merkitystä. Laskemalla testin voimakkuus tämän eron kohdalta, tiedetään miten hyvin testi havaitsee tilastollisesti eron, jolla on käytännössä merkitystä. Esimerkiksi kauralla testi- ja mittarilajikkeen hehtaarisatojen erotusten otoskeskihajonta (= σ_d) on lajikkeista riippuen yleensä 350–500 kg/ha. Jos halutaan 200 kg/ha suuruiset erot tilastollisesti merkitseviksi ($\alpha=0.05$) siten, että tyypin II virhe olisi korkeintaan 0.2 (=20%), niin mittari- ja testilajikkeelta on oltava tulos samanaikaisesti 25–48 kokeesta. Jos vaatimustasoa hiukan lasketaan ja vähintään 300 kg/ha suuruiset erot halutaan tilastollisesti merkitseviksi, lajikekokeita tarvittaisiin 13–24. Lasketut koemäärät ovat aika korkeita, sillä lajiketta voidaan esittää lajikelistalle jo kahden vuoden testauksen jälkeen, jolloin esitettyjä koemääriä ei olla välttämättä saavutettu. Tietysti riski hyväksyä oletus, että lajikkeet ovat yhtä hyvä, vaikka lajikkeet eroaisivatkin, on laskelmissa ollut vain 20 % (testin voimakkuus on tällöin 0.80). Hyväksymällä tälle riskille suurempi todennäköisyys, saadaan vaadittavien kokeiden lukumäärää vähennettyä selvästi. Kauran kaikkien testattavien ominaisuuksien vaatimat koemäärät kahdella eri testin voimakkuuden tasolla löytyvät Taulukosta 3. Tuloksista selviää, että ainoastaan lakoprosentin analyysiin ei riitä kahden eikä edes kolmen vuoden lajikekokeet. Taulukossa esitettyjä kokeiden lukumääriä ei voi soveltaa muiden kasvien lajikekokeisiin.

Kuvassa 1 on parivertailussa käytetyn testin voimakkuusfunktio kauran hehtaari-



Kuva 1. Testinvoimakkuus kauralla hehtaarisadon todellisen eron funktiona kolmella eri havaintomäärällä ($n=10, 20$ ja 40). Kuvassa on lajikkeiden erotuksen keskihajonnaksi (σ_d) valittu 500 kg/ha . Kyseinen keskihajonta voi olla joskus hiukan pienempiäkin, jolloin tarvittavat koemäärät ovat pienempiä.



Kuva 2. Testin voimakkuus kauralla kokeiden lukumäärän funktiona, kun käytännössä merkitsevä ero on 300 kg/ha . Kuvassa on piirretty lajikkeiden erotuksen keskihajonnan arvoille 350 kg/ha (yhtenäinen viiva) ja 500 kg/ha (katkoviiva) oma käyränsä.

sadolle. Kuvasta näkee testin voimakkuuden halutulla lajikkeiden välisellä erolla kokeiden lukumäärän ollessa 10, 20 tai 40. Kuvasta voidaan helposti havaita 10 kokeen riittävän vain suurien erotusten havaitsemiseen. Kuvasta 2 pystyy näkemään valitulla testin voimakkuuden tasolla, kuinka suuri kokeiden lukumäärän on oltava, jotta 300 kg/ha suuruiset erot tulevat tilastollisesti merkitseviksi. Valitsemalla kuvassa testin voimakkuudeksi 0.5 ja 0.8, saadaan Taulukon 3 hehtaarisatoa koskevat koemäärät. Kuva osoittaa, että jo 20 koetta riittää hyvin toteamaan 300 kg/ha suuruiset lajikkeiden väliset erot. Vastaavasti 10 tai sitä pienempi koemäärä vaikuttaa liian vähäiseltä. Tällaisia koemääriä kuitenkin käytetään eri lajitteluiden mukaisissa vertailuissa.

Testien voimakkuutta ei ole esitetty lajikekokeiden tulosraportissa. Käytännössä testien voimakkuuden voi havaita helposti.

Esimerkiksi lajikkeiden välisen satoeron ollessa vähintään 300 kg/ha:lla, olisi tämän suuruisella erolla käytännön kannalta merkitystä. Jos näin suuret tai jopa suuremmatkin erot eivät kuitenkaan ole tulleet tilastollisesti merkitseviksi eli satoeron perässä ei ole yhtään tähteä, ei testi ole ollut riittävän voimakas. Tehottomuus johtuu pääasiassa siitä, että vertailua laskettaessa on ollut liian vähän kokeita käytettävissä, eikä siten 300 kg/ha suuruisen erojen tilastollista merkitsevyyttä voida todeta. Tämä ei kuitenkaan koskaan tarkoita sitä, etteikö lajikkeiden välillä olisi mitään todellista eroa. Tuloksen oikea tulkinta onkin, ettei lajikkeiden välistä eroa, eikä eron suuntaa voida tilastollisesti varmentaa vaaditulla käytännönkannalta merkitsevällä tasolla (=300kg/ha). Tarkempia ohjeita tilastollisen merkitsevyyden tulkinnasta on annettu luvussa 3.3 Tilastollinen merkitsevyys.

Taulukko 3. Tarvittava lajikekokeiden lukumäärä parivertailussa, kun merkitsevyytaso $\alpha=0.05$ ja testin voimakkuus on joko 0.80 tai 0.50. Voimakkuuksista 0.80 voidaan pitää tavoitetasona ja 0.50 minimivaatimuksena lajiketestauksessa. Kokeiden lukumäärät on laskettu kauran lajiketestauksessa käytettävälle ominaisuuksille SOLO Power Analysis-ohjelman avulla. Keskihajonnat (σ_s) on arvioitu vanhoista lajikekokeista tarkastelemalla vain niitä lajikkeita, jotka ovat olleet vähintään 30 eri kokeessa vuosina 1990-1997. Taulukossa minimierotus tarkoittaa pienintä lajikkeiden välistä erotusta, jolla on jo käytännössä merkitystä.

Testattava ominaisuus	σ_d	minimi-erotus	kokeiden lukumäärä	
			testin voimakkuus $1-\beta=0.80$	$1-\beta=0.50$
Sato	350-500 kg/ha	300 kg/ha	13-24	8-13
lako-%	14-17 %	5 %	64-93	33-47
kasvuaika	2.0-3.5 pv	2 pv	10-27	6-14
pituus	3.5-5.0 cm	5 cm	7-10	5-6
tjp	1.5-2.5 g	1.0 g	20-52	11-26
hlp	1.0-2.0 kg	1 kg	10-34	6-18
valk-%	0.6-1.0 %	0.5 %	14-34	8-18
valkuaissato	50-100 kg	40 kg	15-52	9-26
kuori-%	0.8-1.1 %	0.5 %	23-40	12-21
ydinsato	250-350 kg	250 kg	10-18	6-10

3 Tulosten tulkinta

3.1 Mallin vaikutus

Valittu tilastollinen malli vaikuttaa siihen, millaisia tulkintoja saaduista tuloksista voidaan tehdä. Käytetty malli on määrännyt myös suurelta osin tehtyjen raporttien muodon. Tulkintoja, jotka ovat käytetyn mallin oletusten kanssa ristiriidassa, ei voi hyvän tieteellisen tutkimustavan mukaan tehdä. Tuloksia tulkittaessa täytyy aina samanlaisesti huomioida kokeiden lukumäärä, lajikkeiden välinen ero ja eron tilastollinen merkitsevyys.

Lajiketestauksessa verrataan kerralla yhtä testilajiketta yhteen mittarilajikkeeseen. Tällaisesta testaustavasta johtuen eri vertailuista saatuja tuloksia ei voi vertailla keskenään.

Lajikekokeiden tulokset sisältävät monta erilaista testattavaa ominaisuutta. Kaikki ominaisuudet on kuvattu lyhyesti vuosittaisen tulosraportin alkuosassa ja lajikekokeiden suoritusoppaassa (Järvi et al., 1998). Luvussa 2.1 mainittu normaalisuusoletus tulee olla voimassa, jotta tuloksia voidaan pitää luotettavina. Lakoprosentin ja talvituhon analyseissä normaalisuus oletus ei ole erityisen hyvin voimassa, joten näihin tuloksiin kannattaa suhtautua aina hiukan varovaisesti. Vastaavanlaisia testattavia ominaisuuksia ovat perunalla ulkoisen laadun kaikki ominaisuudet. Nämä muuttujat olisi parempi analysoida käyttäen esimerkiksi parametrittomia testejä. Tällaiset testit on esitetty kappaleessa 5.2 Parametrittomat vaihtoehdot.

3.2 Kokeiden lukumäärä

Kokeiden lukumäärää laskettaessa on huomioitu vain ne kokeet, joissa testattavasta ominaisuudesta on saatu tulos sekä mittarilajikkeeseen että testattavan lajikkeeseen osalta. Kokeiden lukumäärä saattaa vaihdella kunakin lajikkeeseen eri ominaisuuksien kohdalla.

Tämä johtuu pääasiassa siitä, että ettei jokaisesta kokeesta ole tehty laboratorioanalyysyjä tai yksittäisiä havaintoja on poistettu. Poiston syynä voi olla esimerkiksi mittausvirhe, eikä oikean arvon mittaaminen ole ollut enää jälkikäteen mahdollista. Lakoisuuden testauksesta on poistettu ne kokeet, joissa lakoa ei ole ilmennyt juuri ollenkaan. Lajiketestaukseen on lisätty tarvittaessa uusia ominaisuuksia, joten niistä ei ole saatavilla havaintoja lisäyspäättöstä edeltäviltä vuosilta.

On siis selvää, että kokeiden lukumäärä vaihtelee eri ominaisuuksissa ja kokeiden lukumäärä vaikuttaa tulosten tarkkuuteen ja testin voimakkuuteen havaita käytännössä merkitseviä eroja. Tästä syystä on keväällä 1998 julkaistussa tulosraportissa ensimmäisen kerran tulostettu käytettyjen kokeiden lukumäärä kaikkien ominaisuuksien kohdalla, eikä pelkästään sadon yhteydessä. Tällä lisäinformaatiolla on pyritty vähentämään tuloksista tehtävien väärin johtopäätösten lukumäärää.

Mittarilajikkeen kohdalla kokeiden lukumäärä ilmoittaa moneenko havaintoon mittarilajikkeen keskiarvo perustuu.

3.3 Tilastollinen merkitsevyys

Lajiketestauksessa käytettyjen kokeiden lukumäärä vaikuttaa paljon saadun tuloksen luotettavuuteen. Parivertailussa käytetty erotuksen tilastollista merkitsevyyttä testaava t-testi voidaan tehdä aina, kun otoskeskihajonta σ_d saadaan laskettua aineistosta. Tämä tarkoittaa käytännössä sitä, että tulokset on oltava vähintään kahdesta lajikekokeesta siten, ettei vertailtavien lajikkeiden välinen ero ole jokaisessa kokeessa täsmälleen sama. Lajikkeiden välinen ero voidaan kuitenkin laskea aina, kun vertailtavat lajikkeet ovat olleet vähintään kerran samassa kokeessa.

Aineistosta laskettu otoskeskihajonta σ_d vaikuttaa paljon tilastollisen merkitsevyyden havaitsemiseen. Jos havaintoja on todella vähän (eli alle viisi), voi otoskeskihajonta sattumasta johtuen olla todellista

tuntematonta vaihtelua pienempi. Tämä tarkoittaa sitä, että käytetty parivertailutesti on yleensä täysin epäluotettava, jos kokeiden lukumäärä on todella pieni ($kpl < 5$). Poikkeuksen tästä muodostavat vertailut, joissa lajikkeiden välinen ero on todella suuri. Jos kokeiden lukumäärä on hieman suurempi, mutta silti pieni ($kpl < 9$) on tilastollisiin testeihin suhtauduttava yhä varovaisesti. Nämä esitetyt kokeiden lukumäärät ovat vain suuntaa antavia. Jotkut ominaisuudet, kuten kasvuston pituus ja kasvuaika ovat ominaisuuksia, joihin lajike itse vaikuttaa paljon ja siten lajikkeiden väliset erot vaihtelevat vain vähän eri kokeissa. Satoerot sitävastoin vaihtelevat paljon eri kokeiden välillä. Tämä johtuu siitä, että erilaiset kontrolloimattomat vaihtelun lähteet vaikuttavat satotuloksiin enemmän kuin esimerkiksi kasvuaikaan.

Usein tulkitaan virheellisesti tilastollisesti merkitsevän eron tarkoittavan samaa kuin käytännönkannalta merkitsevä ero. Tilastollisella merkitsevyydellä ei kuitenkaan ole mitään tekemistä käytännössä merkitsevän eron kanssa. Käytännössä merkitsevä ero tarkoittaa sellaista eroa, jolla on lajiketta valittaessa merkitystä. Hehtaarisadossa käytännössä merkitsevä ero voi olla esimerkiksi 300 kg/ha. Käytännössä merkitsevä ero riippuu tuloksia tulkitsevista henkilöstä, eikä mitään yhtä oikeaa arvoa voida välttämättä määrittää. Tilastollinen merkitsevyyden ja käytännössä merkitsevä erotus voivat toteuttaa neljä erilaista vaihtoehtoa:

1) Erotus voi olla tilastollisesti merkitsevä, muttei käytännössä merkitsevä. Esimerkiksi lajikkeiden välinen satoero 100 kg/ha on tullut tilastollisesti merkitseväksi, mutta tulosta tulkitsevan henkilön mielestä tällaisella erolla ei ole käytännön kannalta suurtakaan merkitystä. Tällöin lajikkeiden voidaan katsoa olevan yhtähyviä, vaikka lajikkeiden välinen ero poikkeaa tilastollisesti nollasta. Tällaiseen tilanteeseen päädytään

yleensä silloin, kun vertailtavat lajikkeet ovat olleet pitkään lajikekoetöinnässä mukana.

2) Erotus ei ole tilastollisesti merkitsevä, mutta käytännössä havaitulla erolla olisi merkitystä. Esimerkiksi testilajikkeeseen satoero mittarilajikkeeseen on kokeissa keskimäärin 600 kg/ha ja käytännön kannalta jo pienemmälläkin erolla olisi merkitystä. Nyt tilastollinen testi ei ole ollut riittävän voimakas (vrt. luku 2.4 Testien voimakkuus). Lajikkeiden välinen ero 600 kg/ha voi olla täysin sattumasta johtuen havaitun suuruinen, mutta todellisuudessa testilajike voi olla jopa heikompi satoero kuin mittarilajike. Mitään varmaa johtopäätöstä ei voida tehdä, sillä kokeita tarvittaisiin lisää. Joskus erotuksen 95%:n luottamusvälin esittäminen voisi tämänkaltaisessa ongelmatilanteessa antaa perusteita tehdä edes jonkunlainen johtopäätös lajikkeiden välisestä paremmuudesta. Luottamusvälin käytöstä koituvia etuja on arvioitu luvussa 5.1 Luottamusvälit.

3) Erotus on sekä tilastollisesti että käytännön kannalta merkitsevä. Nyt voidaan tehdä johtopäätös, että testilajike eroaa mittarilajikkeesta.

4) Erotus ei ole tilastollisesti merkitsevä eikä myöskään käytännössä merkitsevä. Tämä tulos ei aina tarkoita sitä, että lajikkeet ovat todellisuudessa yhtä hyviä. Mikäli vertailussa käytettyjen kokeiden lukumäärä on pieni, voi lajikkeet todellisuudessa erota toisistaan, mutta eroa ei kuitenkaan ole havaintoaineiston pienyydestä johtuen löydetty. Myös nyt luottamusväli olisi tarpeellinen, sillä luottamusvälin avulla voidaan arvioida eron suurin ja pienin mahdollinen arvo.

Tilastollinen merkitsevyys perustuu paljon käytetyn testin oletusten realistisuuteen. Jos luvussa 2.2 Poikkeamien tilastollinen testaus mainittuja oletuksia ei voida pitää realistisina, saadut tilastolliset merkitsevyydet voivat olla harhaanjohtavia. Erityistä varovaisuutta tulee noudattaa tulkittaessa lakoprosenttia, talvituhhoa ja niitä perunan testattavia ominaisuuksia, joita on mitattu kappaleissa tai perustuvat subjektiiviseen arvioon. Kaikissa edellä luetelluissa ominaisuuksissa erotuksen normaalijakumaoletus ei ole useinkaan realistinen.

3.4 Suhdeluku

Suhdeluvun laskukaava on määritelty luvussa 2.1 Koeaineisto ja poikkeamien estimointi. Suhdeluvun laskukaavaa tarkasteltaessa selviää, että se on vain tulosten perusteella saatu laskennallinen arvo. Tämä arvo kertoo montako prosenttia enemmän tai vähemmän satoa testilajike tuottaa mittarilajikkeeseen verrattuna keskimääräisissä olosuhteissa. Keskimääräiset olosuhteet riippuvat käytetyistä kokeista ja ovat näin jokaisessa vertailussa erilaiset. Tästä johtuen eri lajikkeille lasketut suhdeluvut eivät ole keskenään täysin vertailukelpoisia. Suhdeluvun eroavaisuutta arvosta 100 ei testata tilastollisesti.

Ei varmasti ole turhaa korostaa, että valittu tilastollinen malli määrää, millaisia tulkintoja saaduista tuloksista voidaan tehdä. Parivertailussa valittu tilastollinen malli tutkii siis kahden lajikkeen (eli testi- ja mittarilajikkeen) välisen erotuksen absoluuttista arvoa. Tehty analyysi perustuu oletukseen, että lajikkeiden välinen todellinen erotus on aina samansuuruinen riippumatta siitä, minkälaiset olosuhteet kokeessa ilmenivät. Tämä tarkoittaa myös sitä, että mallin oletusten perusteella suhdeluku ei voi olla sama kaikissa erilaisissa olosuhteissa, vaan suhdeluku lähestyy arvoa 100 keskimääräisen satotason kasvaessa ja erkanee arvosta 100 keskimääräisen satotason pienentyessä. Lajikkeiden ollessa täsmälleen

yhtähyviä, suhdeluku on tietysti kasvuolosuhteista riippumaton vakio (= 100). Näin ollen täysin vääriin tuloksiin voidaan päätyä, jos oletetaan suhdeluvun saavan raportissa esitetyn arvon kaikissa mahdollisissa olosuhteissa.

Edellä kuvattua vaikutusta kutsutaan additiiviseksi, koska tämä olettaa testattavan lajikkeen tuottavan keskimäärin aina saman verran enemmän (tai vähemmän) satoa kuin mittarilajike. Tämä vaikutustapa on ollut lajiketestauksessa aina. Jos testattava lajike tuottaa tosiasiallisesti erilaisissa kasvuolosuhteissa suhteessa yhtä paljon enemmän tai vähemmän satoa kuin mittarilajike, niin esiintyvää vaikutusta kutsutaan tällöin multiplikatiiviseksi. Nyt lajikkeiden välinen eroavaisuus voitaisiin kuvata esimerkiksi suhdeluvun kaltaisella kertoimella. Tämänkaltaisen analyysi olettaa, että testattavan lajikkeen ja mittarilajikkeen välinen suhdeluku pysyy samana erilaisissa kasvuolosuhteissa. Se tarkoittaisi samalla myös sitä, että näiden kahden lajikkeen välinen satoero vaihtelisi kasvuolosuhteista riippuen suhdeluvun määräämällä tavalla. Additiivisen ja multiplikatiivisen vaikutusten erilaisuus on helppo havaita Taulukosta 4.

Taulukko 4. Biologinen malli testilajikkeen ja mittarilajikkeen väliselle satoerolle kaikilla erilaisilla satotasoilla.

Vaikutus	biologinen malli
additiivinen	testilajike=mittarilajike+vakio
multiplikatiivinen	testilajike=kerroin*mittarilajike

Käytännössä parivertailu voitaisiin tehdä kummalla tahansa esitetyistä tavoista. Käytetyn mallin valintaa ei voida tehdä sen mukaan, halutaanko tulkita lajikkeiden satoeroa prosenteissa vai kiloissa. Mallin valinta täytyy perustua siihen, kumpi vaihtoehtoista malleista vaikuttaa havaintoaineiston perusteella realistisemmalta vaihtoehdolta. Omien tarkastelujeni perusteella nykyinen, additiivisuuden perustuva malli vaikuttaa yleensä realistisemmalta. Tähän tulokseen olen päässyt analysoimalla eri

kasveilla lajikkeiden väliset erot kummallakin tavalla ja tutkinut mallien sisältämien oletusten realistisuutta. Lisäksi tarkastelin graafisesti aineistosta sitä, kumpi malli näyttää järkevämmältä.

Aivan ehdottomasti additiivinen malli ei kuitenkaan ole parempi kuin multiplikatiivinen malli. Jos multiplikatiivinen malli osoittautuu jonkin kasvin kohdalla paremmaksi, voidaan tämä kasvi analysoida käyttäen multiplikatiivista mallia, jolloin lajikkeiden välinen ero kuvataan suhdeluvun kaltaisella kertoimella. Sellainen vaihtoehto, että kyseinen kasvi analysoidaisiin additiivisella mallilla, mutta tulokset tulkittaisiin suhdelukuina, on täysin väärä toimintatapa.

Additiivinen malli ei aina ole paras mahdollinen, sillä jotkut lajikkeet ovat herkempiä erilaisille kasvuolosuhteille kuin toiset. Tällaisia vaikutuksia ei kuitenkaan voida mitenkään huomioida parivertailussa, vaan on käytettävä esimerkiksi regressioanalyysiin perustuvaa analysointimenetelmää (Finlay & Wilkinson 1963). Koska kaikki lajikkeet eivät ole olleet samoissa kokeissa, niin tämä regressioanalyysiin perustuvassa menetelmässä käytetty kokeen keskimääräinen satotaso on virheellinen. Tämän virheen on pyrkinyt korjaamaan Öfversten (1997) esittämällä uuden tavan analysoida lajikkeiden kykyä tuottaa satoa erilaisilla satotasoilla.

On tavallista, että virallisten lajikekokeiden tuloksia tulkitaan sadon osalta seuraavalla tavalla: 'testilajike on 7 % mittarilajiketta satoisampi'. Tällainen tulkinta perustuu täysin suhdelukuun ja on siksi virheellinen. Koska vuodesta 1998 lähtien on

tulostettu satoero, niin tuloksia olisi parempi tulkita esimerkiksi seuraavalla tavalla: 'testilajike tuottaa 350 kg/ha enemmän satoa kuin mittarilajike, mikä tarkoittaa keskimääräisissä olosuhteissa noin 7 %:n sadonlisäystä' tai 'testilajike tuottaa 350 kg/ha enemmän satoa kuin mittarilajike'.

Vaikka tässä jatkuvasti onkin tarkasteltu pelkkää hehtaarisadon analyysiä, niin myös muidenkin ominaisuuksien analyysi voidaan tehdä käyttäen kumpaa tahansa esitettyä tapaa. Koska suhdelukuja ei ole esitetty esimerkiksi kasvuston pituuden tuloksissa, ei pituuden tuloksia voi tulkita väärin käyttämällä pelkästään suhdelukua.

3.5 Monivuotiset kasvit

Monivuotisilla kasveilla yksi koe kestää yleensä kolme vuotta. Yhdestä koeruudusta mitataan samaa ominaisuutta näin kolme eri kertaa. On selvää, että käytetyn t-testin vaatima havaintojen keskinäinen riippumattomuusoletus ei ole voimassa, koska samasta koeruudusta eri vuosina saadut tulokset korreloivat keskenään. Tästä syystä niiden tilastollisten testien luotettavuus on erittäin heikko, joissa on käytetty useita samasta ruudusta saatuja havaintoja.

Monivuotisilla kasveilla onkin järkevintä tarkastella erikseen eri ikäisiä kasvustoja (=luokittelu iän mukaan). Tällöin tehdyt t-testitkin ovat luotettavampia, koska analyysissä ei ole käytetty useampaa samasta koeruudusta saatua tulosta. Tämänkaltaisen esimerkkitulostus löytyy Taulukosta 5.

Taulukko 5. Niittynurmikan virallisten lajikekokeiden tulokset, kun koeaineisto on lajiteltu kasvuston iän mukaan.

Lajike	Nurmen ikä 1			Nurmen ikä 2			Nurmen ikä 3		
	kpl	%	sl	kpl	%	sl	kpl	%	sl
CONNI =C	21	76.8	100	24	76.7	100	20	70.9	100
AMASON	17	77.9	101	21	72.4	93 ***	18	69.0	96 ***
BARON	21	70.9	92 ***	24	69.9	91 ***	20	66.7	94 **
COCTAIL	9	75.0	98	8	79.3	98	7	72.6	98
FORTUNA	10	73.8	96	8	75.6	95	7	68.4	93 ***

4 Parivertailutulosten raportointi

4.1 Poikkeamasivut

Jos tulostuksessa ei ole otsikossa mainintaa luokittelusta tai muista rajoituksista, on analyysissä käytetty kaikki kokeet, joissa sekä testattava lajike että mittarilajike ovat olleet samanaikaisesti. Ainoana poikkeuksena ohralla ei käytetä liejusavella tehtyjä kokeita. Tällaista sivua kutsutaan poikkeamasivuksi. Liitteessä 1 on esimerkki kauran poikkeamasivusta.

Lajikekokeita ei ole suunniteltu siten, että jokaista lajiketta olisi viljelty jokaisella koepaikalla tai maalajilla. Tämän takia jokin koepaikka, vyöhyke tai maalaji voi olla yliedustettuna aineistossa. Tällöin tuloksiin pääsee vaikuttamaan jokin yksittäinen alue tai maalaji niin paljon, ettei saatuja tuloksia voida varauksetta yleistää koskemaan kaikkia niitä kasvuolosuhteita, joissa lajiketta on tarkoitettu viljeltävän. Uusien lajikkeiden kohdalla tulokset perustuvat yleensä vain muutama vuoteen, jolloin lajikkeiden tuloksiin vaikuttaa myös paljon lajikkeen sopeutuminen kyseisten vuosien sääolosuhteille. Erilaisten olosuhteiden vaikutusvaltaisuudesta saa käsityksen, kun tutkii erilaisten luokitteluiden mukaiset tulokset. Näistä tuloksista näkee aika tarkasti, monessako kokeessa lajike on ollut eri vyöhykkeissä, maalajeissa tai vuosina.

Poikkeamasivulla sadon tulokset on esitetty hiukan muista ominaisuuksista poik-

keavalla tavalla. Jokaiselle lajikkeelle on esitetty lajikkeen satotulosten keskiarvo (Taulukko 6, kohta ①) niissä kokeissa, joissa se on ollut mittarilajikkeen kanssa samanaikaisesti. Nyt on helppoa laskea mittarilajikkeen keskimääräinen sato niissä kokeissa, joissa testi- ja mittarilajike ovat olleen samanaikaisesti. Tämä onnistuu, kun testilajikkeen keskimääräisestä sadosta vähennetään satoero (②). Edelleenkin eri lajikkeiden sadon keskiarvot eivät ole keskenään vertailukelpoisia ja suhdeluku on oikein vain, kun testilajikkeen sato on sarakkeessa ① ilmoitetun suuruinen.

4.2 Vuosittaiset vertailut

Vuosittaiset vertailut tehdään aina sadosta sekä nurmilla myös talvituhosta. Koska kokeiden lukumäärä on aina pieni (yleensä alle 10), niin tilastollisten testien tuloksiin ei pidä luottaa liikaa. Tärkeintä onkin tutkia lajikekeskiarvoja ja eroja. Todellisuudessa lajikkeiden väliset erot eivät välttämättä vaihtele vuosittain, vaikka parivertailun tulosraportista sellaisen kuvan voikin saada. Tällaista väärää huomiota ei voi tehdä, jos huomioi vuosittaisissa testeissä käytetyt koemäärät.

Vuosittaisten testien tekeminen on tietysti turhaa, jos lajikkeiden ero ei vaihtele todellisuudessa vuosittain. Tätä vuoden ja lajikkeen yhdysvaikutuksen olemassaoloa ei voida parivertailussa testata, vaan olisi käytettävä esimerkiksi lineaarisia sekamalleja eli ns. suoravertailua. Luottamusvälien esittäminen antaisi enemmän tietoa oikeiden johtopäätösten tekemiselle.

Taulukko 6. Poikkeamasivun uusi hehtaarisadon tulosten esittämistapa.

Lajike	Viljely- vyöhyke	kpl	sato kg/ha	sato- ero	sl	kpl	lako- %	. . .
M	=C	1-4	82 5020	0	100	82	20	. . .
T		1-4	36 5600	450	109**	36	-6**	. . .
			↑ ①	↑ ②				

Vuosittaisten ja muidenkin luokitteluai-
neistojen kohdalla on mahdollisuus laskea
sadon (ja myös viherpeittävyuden ja talvitu-
hon) kohdalla mittarilajikkeen likimääräi-
nen keskiarvo niissä kokeissa, joissa sekä
mittarilajike että testilajike ovat olleet sa-
manaikaisesti. Tämä voidaan tehdä kaavan
(2) mukaisesti.

$$\text{mittarilajike} = \frac{\bar{d}}{(sl/100)-1} \quad (2)$$

Koska raportoidut arvot ovat pyöristet-
tyjä arvoja, kaavan (2) mukaan laskettu
mittarilajikkeen keskiarvo saattaa erota to-
dellisesta keskiarvosta useita prosentteja eli
saatu keskiarvo on vain suuntaa antava.
Testilajikkeen sadon määrän vastaavissa
kokeissa saa laskettua, kun lisää mittarila-
jikkeen satoon lajikkeiden erotukset (= \bar{d}).
Myös tämä arvo saattaa erota todellisesta
keskiarvosta useita prosentteja. Laskemalla
itse testilajikkeen ja mittarilajikkeen liki-
määräiset keskiarvot saa tarkemman kuvan
siitä, millaisia kokeita kuhunkin vertailuun
on sisältynyt.

4.3 Muut vertailut

Alueelliset vertailut ja maalajittaiset ver-
tailut on tehty valitsemalla vain tarkastel-
tavaan luokkaan sisältyneet kokeet.
Tämä johtaa usein siihen, että kussakin
vertailussa käytettyjen kokeiden luku-
määrä on liian pieni, eikä tilastollisen tes-
tin luotettavuus ole riittävä (vrt. Tauluk-
ko 3). Varsinkin kahden vuoden kokeiden
perusteella ei voida osoittaa kuin erittäin

suuret erot tilastollisesti merkitseviksi.

5 Parannusehdotuksia parivertailuun

Parivertailun raportointia voidaan kehittää
informatiivisemmaksi. Myös itse parivertai-
lua voidaan parantaa luopumalla t-testin
käytöstä silloin, kun t-testin käyttö ei ole
soveliaista aineiston kannalta. Käyttämällä
soveliaampia menetelmiä ja testejä, pysty-
tään lisäämään tulosten luotettavuutta.

Parivertailun suurin ongelma on, ettei
saman kasvin eri lajikkeita voida testata yh-
tään aikaisesti. Tällöin päädytään usein ristirii-
tisiin tuloksiin. Esimerkiksi taulukosta 7
pystytään havaitsemaan, että testilajike A
on parempi ja testilajike B huonompi kuin
mittarina käytetty Pohto. Tällöin voidaan
tietysti väittää, että Testilajike A on testila-
jike B parempi. Vaihdettaessa mittarila-
jikkeeksi Arve, lajike B osoittautuu parem-
maksi ja lajike A huonommaksi kuin mitta-
rilajike. Nyt voitaisiin väittää, että lajike B
on parempi kuin lajike A. Näin on siis pää-
dytty kahteen erilaiseen tulokseen. Yksi
mahdollisuus ratkaista lajikkeiden parem-
muus yksiselitteisesti olisi käyttää lineaari-
sia sekamalleja (ts. suoravertailua) lajike-
testauksessa. Lineaariset sekamallit pystyisivät
huomioimaan kokeiden välisen riippuvuu-
den ja samasta koeruudusta eri vuosina saa-
tujen mittaustulosten välisen korreloitumi-
sen.

Taulukko 7. Parivertailun tulokset kahden mittarilajikkeen osalta.

Lajike	satoero	sl	Lajike	satoero	sl
Pohto =C	0	100	Arve =C	0	100
lajike A	12	100	lajike A	-40	99
lajike B	-112	98	lajike B	782	113 ***

5.1 Luottamusvälit

Monesti tulkitaan joko tahattomasti tai tarkoituksella, että saadut lajikkeiden väliset erot ovat tarkkoja arvoja. Tosiasiassa saatu erotus on vain arvio todellisesta arvosta. Luottamusväli kertoo sen, millaiset todelliset erotukset voivat havaintojen perusteella olla hyvin todennäköisiä. Luottamusvälien laskukaavan parivertailun t-testiä käytettäessä on esittänyt mm. Gardner ja Altman (1989, p. 20). Lisäksi he esittävät yksityiskohtaisesti, milloin ja miksi luottamusvälien esittäminen on järkevämpää kuin pelkkinen tilastollisten merkitsevyyksien esittäminen. Erityisen mielekkäänä he pitävät luottamusvälien esittämistä silloin, kun erotuksen suuruus kiinnostaa paljon enemmän kuin erotuksen tilastollinen merkitsevyys. Juuri tästähän lajiketestauksessa on kysymys.

Virallisissa lajikekokeissa kauralla on kahden lajikkeen hehtaarisatojen erotusten keskivirhe lajikkeista riippuen yleensä 350-500 kg/ha. Tämä tarkoittaa sitä, että saatu keskimääräinen satoero voi erota todellisesta satoerosta jopa 250-350 kg/ha, kun lajikkeita on testattu samanaikaisesti yhdeksässä kokeessa. Jos kokeita on ollut enemmän kuin yhdeksän, virhe on pienempi ja vastaavasti pienemmällä koemäärällä virhe on suurempi. Jos havaittu satoero yhdeksässä kokeessa on ollut esimerkiksi 200 kg/ha, niin 95%:n luottamusväli erotukselle voi tällöin olla optimitilanteessa (-50 kg/ha, 550 kg/ha). Luottamusvälin mukaan testilajike voi tuottaa todellisuudessa keskimäärin jopa 550 kg/ha enemmän satoa kuin mittarilajike. Sama luottamusväli kertoo myös sen, että testilajike voi todellisuudessa tuottaa myös satoa 50 kg/ha mittarilajiketta vähemmän. Koska 95%:n luottamusvälin alku- ja loppupiste ovat eri merkkiset, ei lajikkeiden välinen ero ole tullut tilastollisesti merkitseväksi riskitasolla 0.05. Kuitenkin luottamusvälin perusteella voidaan päätellä, että testilajike ei voi olla oleellisesti heikkosatoisempi kuin mittarilajike, jolloin on saatu edes jonkinlainen johtopäätös aikaiseksi. Ilman luottamusväliä voidaan to-

detta vain havaintoaineiston olevan liian pieni minkäänlaisien johtopäätösten esittämiseen.

Jos satoeron luottamusväliksi saadaan esimerkiksi (-400,1400), nähdään heti havaintomäärän olleen liian pieni, sillä testilajike voi käytännössä olla selvästi mittarilajiketta parempi tai huonompi. Tällöin pelkän keskimääräisen satoeron (=500 kg/ha) julkaiseminen voi aiheuttaa sen, että testilajike tulkitaan väärin perustein mittarilajiketta paremmaksi, vaikka luvussa 3.3 Tilastollinen merkitsevyys esitetyn vaihtoehdon 2 mukaisesti näin ei kuitenkaan olisi saanut tehdä.

Myös silloin, kun lajikkeiden välinen ero ei ole käytännössä eikä tilastollisesti merkitsevä, luottamusväli voi olla todella tarpeellinen. Esimerkiksi 95%:n luottamusvälin ollessa (-500 kg/ha, 500 kg/ha) lajikkeiden välinen ero on keskimäärin ollut 0 kg/ha ja lajikkeet mielletään tällöin yhtähyviksi. Todellisuudessa lajikkeiden välinen ero voi olla 500 kg/ha kumman tahansa lajikkeen hyväksi. Luottamusväliin perustuva johtopäätös olisikin, ettei lajikkeiden erosta voida sanoa mitään, koska kokeita on tehty liian vähän. Taulukossa 8 on esitetty kauralle luottamusvälit, kun testilajikkeiden hehtaarisatoa on verrattu mittarina olleen VELI-lajikkeen satoon. Taulukosta havaitaan helposti, kuinka kokeiden lukumäärä vaikuttaa tulosten tarkkuuteen. Linjalla BOR 1407 on keskihajonta ($=\sigma_d$) sattumasta johtuen estimoitunut liian pieneksi, jolloin luottamusvälistäkin on tullut liian lyhyt.

Luvussa 4.2 Vuosittaiset vertailut oli toteama, ettei lajikkeiden väliset erot välttämättä vaihtele vuosittain, vaikka parivertailun tulosraportista sellaisen kuvan helposti saakin. Esimerkiksi Liitteessä 2 on verrattu Leilan satoa Velin satoon eri vuosina. Pelkkiä keskiarvoja ja suhdelukuvia tarkastelemalla voi arvioida Leilan sadon olleen Velin satoon verrattuna vuonna 1994 selvästi suurempi kuin muina vuosina. Kuvassa 3 on piirrettynä kuhunkin vuoteen liittyvä keskimääräinen satoero ja satoeron luottamusväli. Kuvan perusteella ei voida päätyä johtopäätökseen, että vuonna 1994 lajik-

Taulukko 8. Luottamusvälien käyttö hehtaarisadon tulosten raportoinnissa.

Lajike	kpl	satoero	satoeron 95 %:n luottamusväli	
VELI =C	115	-	-	-
PUHTI	92	109 *	25	192
YTY	75	376 ***	280	473
SISKO	52	44	-57	144
LISBETH	44	-2306 ***	-2519	-2092
AARRE	68	37	-39	112
JO 1307	20	-1847 ***	-2295	-1399
BOR 1323	33	485 ***	327	642
BOR 1357	35	153 **	51	255
ROOPE	53	604 ***	498	710
BOR 1407	9	144 *	11	276
BOR 70260	6	419 *	66	772
BOR 87712	6	557 *	153	961
BOR 88322	12	749 ***	520	978
VIRMA	52	358 ***	235	482
KATRI	68	479 ***	378	581
FREJA	46	579 ***	417	742
SW 881330	36	272 ***	127	418
BELINDA	27	1018 ***	784	1252
SALO	98	201 ***	101	300
LEILA	103	-84 *	-155	-13
KOLBU	76	303 ***	222	385

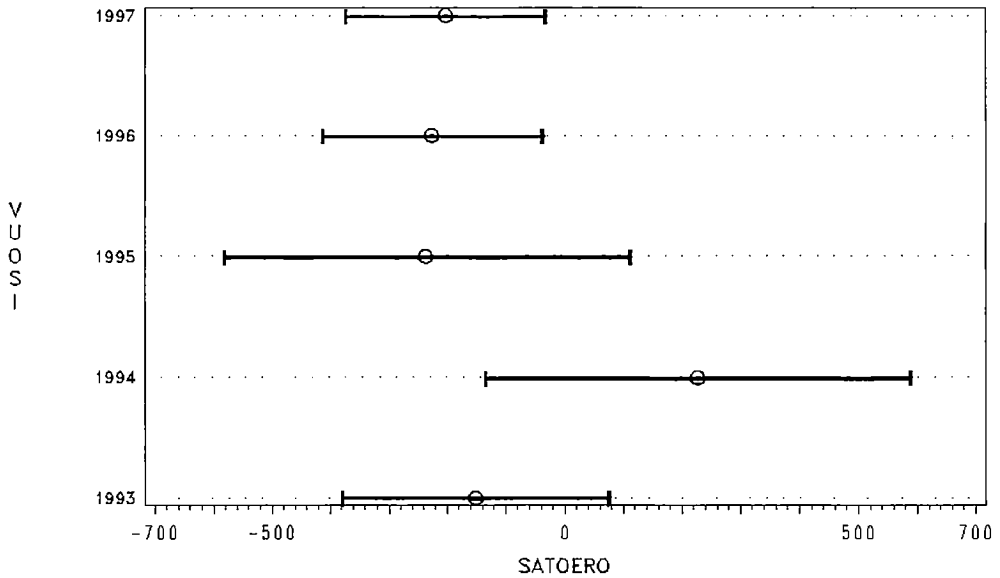
keiden välinen ero olisi selvästi erilainen kuin muina vuosina. Lisäksi vuoteen 1994 liittyvä luottamusväli sisältää kaikista vuosista lasketun ko. lajikkeiden keskimääräisen eron (= -83 kg/ha, Liite 1).

5.2 Parametrittomat vaihtoehdot

Lajikekokeista saatuun aineistoon (Taulukko 1) voidaan soveltaa kaltaistettujen parien t-testin sijasta joko yhden otoksen merkkitesti tai Wilcoxonin testi. Näitä testejä käyttäen saadaan aikaiseksi täsmälleen samantyyppiset tulokset kuin perinteisellä parivertailuilla, mutta testien vaatimukset

tulevat täytettyä jakaumaoletusten osalta paremmin.

Käytännössä ei ole järkevää käyttää parametrittomia menetelmiä, mikäli lajikkeiden välisen erotuksen jakauma on peräisin normaalijakaumasta. Jos näin silti tehtäisiin, olisi kahden lajikkeen todellista eroa kuvaava tunnusluku silti oikea, mutta testi ei olisi kuitenkaan niin tehokas kuin se voisi olla. Testin heikentyminen vaikeuttaa johtopäätösten tekoa, eikä käytännössä merkittäviä eroja voitaisi aina tilastollisesti vahvistaa. Testien suhteellisia asymptoottisista tehokkuuksista on raportoinut mm. Gibbons ja Chakraborti (1992). Wilcoxonin testin suhteellinen tehokkuus verrattuna



Kuva 3. Lajikkeiden Leila ja Veli välinen satoero ja eron luottamusväli viitenä viimeisenä vuotena. Jokainen vuosi on analysoitu erikseen, kuten parivertailussa normaalisti tehdäänkin. Ympyrällä on merkitty lajikkeiden keskimääräiset erot, jotka löytyvät myös liitteestä 2.

parivertailussa käytettyyn t-testiin on normaalijakauman tilanteessa $3/\pi (=0,955)$ ja huonoimmillaankin peräti 0,864, kun havainnot (eli lajikekokeissa lajikkeiden väliset erotukset) ovat peräisin mistä tahansa jatkuvasta ja symmetrisestä jakaumasta. Näiden tietojen valossa ei tehokkuuden menetys ole kovinkaan merkitsevä, jos t-testin tilalla käytettäisiin aina Wilcoxonin testiä. Jos taas käytetään normaalijakaumaoletukseen perustuvaa testiä havaintoaineistoon, joka ei ole peräisin normaalijakaumasta, voidaan helposti saada lajikkeiden väliselle erolle tilastollinen merkitsevyys, vaikka tosiasiassa eroa ei olisikaan. Tämä tulee hyvin ilmi pitkähäntäisissä vasteissa, joita lajikekokeissa ovat lähes kaikki prosenttivasteet.

Yksi parametrittomat menetelmien hyvistä puolista on se, etteivät ne ole niin herkkiä poikkeaville havainnoille kuin normaalisuuteen perustuvat menetelmät. Tämä johtuu pitkälti siitä, että lajikkeiden välisiä eroja laskettaessa käytetään keskiarvojen sijasta mediaania tai Hodgesin-Lehmanin estimaattia. Näistä kolmesta keskiarvo on

kaikista herkin poikkeaville havainnoille. Tällä hetkellä virallisten lajikekokeiden havaintoaineisto sisältää runsaasti poikkeavia havaintoja, joista osa on suoranaisia kirjajamismvirheitä. Virheellisten havaintojen runsaus johtunee puutteellisista poikkeavien havaintojen tunnistamismenetelmien tunnistamisesta. Poikkeavien havaintojen tunnistamisesta lajikekokeissa ovat raportoineet mm. Jauhiainen & Öfversten (1997).

Myös parametrittomat testit sisältävät aineistoa koskevia oletuksia, joiden tulee olla voimassa. Wilcoxonin testi on käyttökelpoinen, mikäli vastemuuttujan jakauma on symmetrinen, muttei välttämättä normaalin. Koska virallisissa lajikekokeissa vastemuuttuja on aina kahden lajikkeen välinen erotus, niin symmetrisyys on voimassa lähes aina. Yhden otoksen merkkitesti on Wilcoxonin testiä väljempi, eikä se sisällä oletusta vasteen jakaumasta. Koska symmetrisyys oletus on kuitenkin voimassa lähes aina, ei ole yleensä mitään syytä käyttää merkkitestiä lajikekoesarjojen analysoinnissa. Kumpikin parametriton testi perustuu

Taulukko 9. Kauran lakoisuuden analyysin tulokset käytettynä sekä t-testiä että Wilcoxonin testiä. Lakoerolle on lisäksi laskettu 95%:n luottamusväli. Testattavien lajikkeiden kohdalla lakoprosentti kertoo mittarilajikkeeseen ja testilajikkeeseen välisen erotuksen.

Lajike	kpl	parivertailun t-testi			Wilcoxonin testi		
		lako -%	lakoeron -%	95%:n luottamusväli	lako -%	lakoeron -%	95%:n luottamusväli
VELI	=C 70	28.8			25.0		
KATRI	39	-6.5 *	-12.6	-0.4	-5.3 **	-11.0	-2.0
KOLBU	43	-6.9 *	-12.4	-1.5	-5.5 ***	-8.0	-3.0
LEILA	63	1.4	-2.8	5.6	0.0	-2.5	3.5
SALO	54	-12.6 ***	-18.8	-6.4	-11.0 ***	-18.0	-5.5
ROOPE	29	-3.9	-8.5	0.8	-2.0	-5.5	0.0
AARRE	34	-3.2 *	-5.7	-0.6	-2.0 *	-5.0	-0.5
SISKO	32	-7.3 **	-12.6	-1.9	-5.0 *	-10.5	-1.0
YTY	44	7.0 **	1.9	12.1	3.5 **	1.0	8.0
PUHTI	49	3.7	-1.3	8.7	-1.0	-3.5	1.0
BELINDA	16	-11.4	-24.1	1.3	-4.5 *	-21.0	0.0
SW 881330	22	-9.1 *	-16.6	-1.6	-5.5 ***	-13.5	-3.0
BOR 70260	5	-7.4	-28.1	13.3	-4.0	-34.0	8.0
BOR 1407	8	-18.3 *	-33.0	-3.5	-16.5 *	-36.0	-2.5
VIRMA	29	-0.4	-8.2	7.5	-2.0	-6.0	2.5
FREJA	25	-1.0	-7.6	5.5	-1.5	-6.5	2.0
JO 1307	11	-9.6	-20.9	1.6	-7.0	-24.0	1.0
BOR 88322	7	-6.1	-14.3	2.1	-4.5	-14.0	1.0
BOR 87712	5	-10.8	-36.6	15.0	-2.5	-45.0	5.0
BOR 1357	21	1.7	-4.5	8.0	0.0	-2.0	2.5
BOR 1323	17	-0.6	-8.6	7.4	-3.0 *	-6.0	-0.5
LISBETH	31	4.5	-2.0	10.9	2.3	-2.5	8.5

oletukseen, että eri kokeista saadut havainnot ovat toisistaan riippumattomia. Tämä oletus ei kuitenkaan ole täysin voimassa, kuten parivertailun t-testin kohdalla on jo todettukin.

Parametrittomien menetelmien käytön kannalta ongelmallisinta ovat vasteet, joissa lajikkeiden välinen ero on usein nolla. Tosin tällaisia vasteita on vaikea analysoida käyttäen mitään muutakaan menetelmää. Eriytyisen hankalaksi tilanne muodostuu, mikäli mittarilajike sisältää lähes pelkästään nollia. Tällöin Wilcoxonin testin symmetriavaatimus ei ole voimassa, joten merkkiteesti lienee tällöin ainoa mahdollinen vaihtoehto. Tästä hyvänä esimerkkinä on ohralla noki-tähkien lukumäärä kasvustossa (10 m²),

jota ei kuitenkaan enää tänä vuonna ole sisällytetty virallisten lajikekokeiden julkaisuun. Vastaavia ominaisuuksia löytyy kuitenkin edelleen perunalta.

Taulukossa 9 on analysoituna kauran lakoisuus mittarilajikkeeseen verrattuna sekä tavallisella parivertailulla että parametrittomalla vaihtoehdolla. Linjan 'BOR 1323' tuloksissa on havaittavissa, kuinka yhden kokeen tulos vaikuttaa parivertailun tuloksiin. Kyseinen linjalla on lakoprosentti ollut 13 kokeessa pienempi kuin mittarina toimineella VELI-lajikkeella. Kahdessa kokeessa lakoprosentti on ollut yhtäsuuri ja kahdesti linjalla on ollut korkeampi lakoprosentti kuin mittarilajikkeella. Yhden kokeen suuri lakoprosentti aiheuttaa, ettei parivertailus-

sa linjan ja mittarilajikkeen ero tulee tilastollisesti merkitseväksi riskitasolla 0.05. Sitä vastoin parametrittomiin testeihin yhden kokeen poikkeava arvo ei vaikuta yhtä paljon, ja lajikkeiden välinen ero tulee tilastollisesti merkitseväksi käytettäessä Wilcoxonin testiä (Taulukko 9) tai merkkitestiä (ero = -3.0, $p < 0.01$). Kokeiden lukumäärän ollessa erittäin suuri (> 30), parivertailun ja parametrittomien testien tulokset alkavan olla lähes identtiset.

Huomioimalla kaikki parametrittomien testien edut ja haitat, voi parametrittomia vaihtoehtoja suosittelaa varauksella kaikkien ei-normaalisten ominaisuuksien testaukseen. Koska kaikkia poikkeavia havaintoja ei ole koeaineistosta havaittu, parametrittomia testejä voi suosittelaa käytettäväksi lajiketestauksessa aina.

6 Laaduntarkastus

Yksittäisen lajikekoeaineiston laaduntarkastus tapahtuu ensisijaisesti koepaikalla. Myös lajivastaavat tarkastavat koeaineistot. On kuitenkin mahdollista, että aineiston tiedonsiirrossa tietopalveluyksikköön on tapahtunut virhe tai jokin yksittäinen virheelinen havainto on jäänyt koepaikalla huomaamatta. Näiden virheiden havaitsemiseksi koeaineistojen yhdistämisen jälkeen

jokaisen analysoitavan ominaisuuden arvot tutkitaan. Tämä tarkastus suoritetaan yksinkertaisesti tulostamalla ominaisuuskerrollaan kaikkien havaintojen jakauma. Tästä jakaumasta pystyy silmämääräisesti havaitsemaan selvät virheet.

Tietopalveluyksikössä laaduntarkastuksen päämääränä on havaita yksittäiset kokeet tai kokeiden osat, jotka tuottavat muista kokeista poikkeavia tuloksia. Tämä tehdään yksinkertaisesti analysoimalla aineisto käyttäen lineaarisia sekamalleja. Tällöin riittää käyttää todella yksinkertaista mallia, jossa lajiketekijä on ainoa ns. kiinteävaikutteinen tekijä ja koetekijä on ns. satunnaisvaikutteinen tekijä. Koetekijä identifioi kokeet ja se koostuu vuodesta, koepaikasta ja koepaikalla tehtävästä koesarjasta. Tarkasteltaessa koevaikutusten estimoituja arvoja, havaitaan muista kokeista poikkeavat kokeet (Öfversten & Nikander 1996). Itseisarvoltaan suurimpia jäännöksiä (residuaaleja) tarkasteltaessa pystytään myös löytämään yksittäiset poikkeavat havainnot, joita koepaikoilla ei ole havaittu.

Ottamalla malliin useampia satunnais-tekijöitä, voidaan löytää sekä poikkeavia kokeita että yksittäisiä poikkeavia havainnot. Tämä työ vaatisi kuitenkin liian suurta panostusta verrattuna saatavaan hyötyä. Tällaista tarkempaa laaduntarkastusta tehdäänkin pääasiassa vain, jos aineistossa havaitaan jotain erikoisuutta.

Kirjallisuus

- Finlay, K.W. & Wilkinson, G.N.** 1963. The analysis of adaptation in a plant breeding programme. *Australian Journal of Agricultural Research* 14: 742–754.
- Gardner, M.J. & Altman, D.G.** 1989. *Statistics with confidence-Confidence intervals and statistical guidelines*. London: The British Medical Journal. 108 p. ISBN 0–7279–0222–9.
- Gibbons, J.D. & Chakraborti, S.** 1992. *Nonparametric statistical inference*. Third edition. New York: Marcel Dekker. 462 p. ISBN 0–8247–8661–0.
- Jauhiainen, L. & Öfversten, J.** 1997. Lajikekokeiden suunnittelu ja laadunhallinta. Maatalouden tutkimuskeskuksen julkaisuja. Sarja B 10. Jokioinen: Maatalouden tutkimuskeskus. 23 p. ISBN 951–729–500–6. ISSN 1238–9943.
- Järvi, A., Kangas, A., Mustonen, L., Salo, Y., Talvitie, H., Vuorinen, M. & Mäkelä, L.** 1995. Virallisten lajikekokeiden tuloksia 1987–1994. Maatalouden tutkimuskeskus, Tiedote 2/95. Jokioinen: Maatalouden tutkimuskeskus. 126 p. ISBN 951–729–466–2. ISSN 0359–7652.
- , **Kangas, A., Mattila, I., Mäkelä, L., Rahkonen, A., Salo, Y., Vuorinen, M. & Öfversten, J.** 1998. Virallisten lajikekokeiden suoritusohjeet. Maatalouden tutkimuskeskuksen julkaisuja. Sarja B 14. Jokioinen: Maatalouden tutkimuskeskus. 71 p. ISBN 951–729–524–3. ISSN 1238–9943.
- Manly, B. F.J.** 1997. *Randomization, Bootstrap and Monte Carlo Method in Biology*. Second Edition. London: Chapman & Hall. 399 p. ISBN 0–412–72130–9.
- Öfversten, J.** 1997. Assessing Sensitivity of Agricultural Crop Varieties. *Journal of Agricultural, Biological, and Environmental Statistics* 3 (1): 37–47.
- & **Nikander, H.** 1996. Lajikekoesarjojen analysointi. Maatalouden tutkimuskeskuksen julkaisuja. Sarja B 2. Jokioinen: Maatalouden tutkimuskeskus. 27 p. ISSN 1238–9943.

Oats, *Avena sativa* L., 1990-1991, Official variety trials

Poikkeamat mittarilajikkeen tuloksista, Differences from results of control variety

Lajike	Viljely- vyöhyke	Sato kg/ha	Sato- ero	sl	kpl	Lako- %	kpl	Kasvuaika pv	kpl	Pituus cm
Variety	Cultivation zone	Yield kg/ha	Yield difference	rel.	trials	Lodging- %	trials	Growing time d	trials	Length cm
	=C									
VELI	1-4	5323	0	100	70	28	115	97	113	102
LEILA	1-4	5280	-83	98*	63	1	103	0*	101	-14***
SALO	1-3	5530	200	104***	54	-12***	98	7***	98	-18***
AARE	1-4	5425	36	101	34	-3*	68	0	66	3***
BELINDA	1-2	6137	953	118***	11	-7	18	6***	18	-12***
FREJA	1-3	6027	579	111***	25	-1	46	7***	46	-6***
KATRI	1-3	5822	479	109***	39	-6*	68	6***	68	-4***
KOLBU	1-4	5626	303	106***	43	-6*	76	1***	74	-8***
LISBETH	1-3	3415	-2243	60***	27	5	39	2***	40	-1*
PUHTI	1-3	5489	120	102**	47	3	90	5***	90	6***
ROOPE	1-3	6067	603	111***	29	-3	53	4***	53	-3***
VIRMA	1-3	5846	358	107***	29	0	52	5***	52	-6***
YTY	1-3	5789	376	107***	44	7**	75	4***	75	2***
SISKO	1-4	5592	43	101	32	-7**	52	-1***	50	-3***
BOR 1323	1-3	5992	484	109***	17	0	33	4***	33	-6***
BOR 1357	1-3	5626	154	103*	17	2	31	0	31	-1
BOR 1407	1-4	5282	143	103*	8	-18*	9	0	9	-11***
BOR 70260	1-4	5007	419	109*	5	-7	6	6**	6	-7**
BOR 87712	1-4	5145	557	112*	5	-10	6	6**	6	-9**
BOR 88322	1-4	5844	749	115***	7	-6	12	7***	12	-8***
JO 1307	1-4	3448	-1846	65***	11	-9	20	5***	20	2*
SW 881330	1-4	5792	272	105***	22	-9*	36	3***	34	-14***

LIITE 1 (2/2)

Kaura, 1990-1997, Viralliset kokeet
Oats, *Avena sativa* L., 1990-1997, Official variety trials

Poikkeamat mittarilajikkeen tuloksista, Differences from results of control variety

Lajike	Viiljely- vyöhyke	Tjip g	Hlp kg	Valk.-		Prot.-		Valk.sato kg	Kuori- %	Ydinsato kg		
				kpl	%	kpl	%					
Variety	Cultivation zone	1000-grain weight g		Hl weight kg		Prot.yield kg		Hull content %		Groat yield kg		
		trials	trials	trials	trials	trials	trials	trials	trials			
	=C	115	36.1	115	54.5	115	13.5	612	111	22.9	111	4114
VELI	1-4	103	2.0***	103	0.3*	102	-0.1	-15	99	0.9***	99	-113***
LEILA	1-3	98	4.1***	98	-1.4***	96	-1.1**	-32***	94	-9***	94	206***
SALO	1-4	68	1.4***	68	-0.5***	68	-0.0	6	65	-1.6***	65	118***
AARRE	1-2	18	5.3***	18	-0.7	18	-1.8**	9	18	1.1**	18	653***
BELINDA												
FREJA	1-3	46	2.4***	46	0.3	46	-1.7**	-20	44	-.4*	44	464***
KATRI	1-3	68	-0.2	68	-1.1***	68	-1.0**	4	66	-1.1***	66	427***
KOLBU	1-4	76	-0.4	76	-2.0***	76	-1.1**	-19**	73	0.1	73	218***
LISBETH	1-3	40	-7.5***	40	7.2***	40	3.7**	-144***	40	-18.***	40	-1E3***
PUHTI	1-3	90	0.3	90	-2.1***	90	-0.9**	-29***	86	-1.4***	86	167***
ROOPE	1-3	53	-0.1	53	-2.0***	53	-0.9**	17*	51	-1.***	51	542***
VIRMA	1-3	52	-2.3***	52	-1.4***	51	-1.1**	-11	50	-8.***	50	315***
YTY	1-3	75	2.1***	75	-1.7***	74	-1.3**	-18	71	-1	71	310***
SISKO	1-4	52	-4.3***	52	-1.7***	51	-0.5**	-20**	50	1.0***	50	-25
BOR 1323	1-3	33	1.7***	33	-0.5**	33	-1.5**	-24	33	0.7**	33	329***
BOR 1357	1-3	30	1.6***	30	-0.9***	30	0.0	17*	30	-1.9***	30	219***
BOR 1407	1-4	9	-1.1*	9	1.1**	9	0.2	26	9	-2.**	9	211**
BOR 70260	1-4	6	3.2***	6	-0.9	6	-0.9*	11	6	-9*	6	368*
BOR 87712	1-4	6	-0.0	6	-2.5**	6	-1.5*	3	6	0.4*	6	397*
BOR 88322	1-4	12	0.1	12	-2.8***	12	-1.5**	8	12	1.4***	12	483***
JO 1307	1-4	20	-8.0***	20	6.5***	20	3.0**	-112***	20	-18.***	20	-809***
SW 881330	1-4	36	-1.6***	36	-1.0***	36	-0.6**	-4	36	1.7***	36	100

LIITE 2

Kaura, 1993-1997, Viralliset kokeet: Luokittelu vuosittain
 Oats, *Avena sativa* L., 1993-1997, Official variety trials: Classification according to years

Lajike Variety	V 1993		V 1994		V 1995		V 1996		V 1997							
	kpl	kg/ha	kpl	kg/ha	kpl	kg/ha	kpl	kg/ha	kpl	kg/ha						
	trials	kg/ha	trials	kg/ha	trials	kg/ha	trials	kg/ha	trials	kg/ha						
	=C	15	5618	100	12	4876	100	13	6054	100	12	5612	100	12	4791	100
VELI		15	-151	97	5	229	105	11	-258	96	11	-226	96*	12	-202	96*
LEILA		13	-35	99	10	144	103	11	97	102	10	-231	96	10	327	107**
SALO		15	252	104*	12	109	102	13	-104	98	11	-52	99	3	-183	97
AARRE					7	1360	125**	10	615	111**	10	615	111**	10	1181	124***
BELINDA																
FREJA		7	649	111**	8	531	110**	9	612	110*	7	-16	100	9	753	115***
KATRI					11	421	107*	6	-53	99	11	99	102	11	189	104*
KOLBU		15	364	106**	10	375	108*	13	359	106**	11	-2385	59***	11	-2757	52***
LISBETH					2	-1545	74	11	-2385	59***	11	-2757	52***	1	-2370	60
PUHTI		13	36	101	10	38	101	11	-58	99	12	-27	100	1	330	106
ROOPE		13	706	112***	10	367	108**	11	768	113**	9	407	107**	9	668	113***
VIRMA		4	406	108	2	475	110*	5	292	105	5	143	102	1	205	104
YTY		13	159	103	7	302	106	5	170	103	5	127	102	3	313	107
SISKO		15	-111	98	8	-97	98	5	128	102	1	30	101	10	647	113***
BOR 1323		1	-50	99				11	486	108*	10	340	106*	10	647	113***
BOR 1357					11	302	105*	12	100	102	12	100	102	12	69	101
BOR 1407								6	891	116**	6	891	116**	9	143	103*
BOR 70260								6	419	109*	6	419	109*	6	419	109*
BOR 87712								6	557	112*	6	557	112*	6	557	112*
BOR 88322								6	606	113**	6	606	113**	6	606	113**
JO 1307								10	-2290	59***	10	-2290	59***	10	-1403	72***
SW 881330					13	220	104	12	219	104	12	219	104	11	391	108**

	Julkaisun sarja ja numero Maatalouden tutkimuskeskuksen julkaisuja. Sarja B 16	
	Julkaisuaika (kk ja vuosi) Lokakuu 1998	
Tekijä(t) Lauri Jauhiainen	Tutkimushankkeen nimi	
	Toimeksiantaja(t) Maatalouden tutkimuskeskus	
Nimike Virallisten lajikekokeiden tulosten laskentaperusteet.		
Tiivistelmä Maatalouden tutkimuskeskuksen koordinoiman peltokasvien virallisen lajikekoetoiminnan puitteissa suoritetaan vuosittain yli sata yksittäistä kenttakoetta. Kokeiden pääasiallisena tarkoituksena on kerätä aineistoa laajempien koesarjakohtaisten analyysien tekemiseksi. Tällaiset koesarjakohtaiset analyysit tehdään vuosittain käyttäen kahdeksana edellisenä vuonna suoritettuja kenttäkokeita. Analyysien tulokset julkaistaan maatalouden tutkimuskeskuksen julkaisusarjassa A. Jotta analyysien tulosten oikea tulkitseminen olisi mahdollista, täytyy tulkitsijan tuntea analyysissä käytettävä tilastollinen malli ja sen sisältämät oletukset. Lisäksi tärkeää on tietää, mitä kaikki raportoidut tulokset tarkoittavat ja miten ne on laskettu. Tässä raportissa kuvataan tarkasti tällä hetkellä käytettävä testaus tapa, joka tunnetaan lajikekoetoiminnassa nimellä parivertailu. Erityistä huomiota on keskitetty sadon suhdelukuihin, joita aikaisemmin on usein tulkittu väärin perustein. Usein tulkitaan myös virheellisesti tilastollisesti merkitsevän eron tarkoittavan samaa kuin käytännön kannalta merkitsevä ero. Näiden kahden merkitsevyyden välisen eron tunteminen edesauttaa oikeanlaisten johtopäätösten tekemistä. Tässä raportissa annetaan myös neuvoja siitä, kuinka parivertailutulosten luotettavuutta tulee arvioida. Lopuksi esitetään tietopalveluyksikössä tehtävät koeaineiston laadun tarkastukset. Laaduntarkastuksella on tarkoitus varmistaa käytetyn aineiston käyttökelpoisuus ja sitä kautta taata tulosten oikeellisuus.		
Avainsanat luottamusväli, parivertailu, suhdeluku, tilastolliset testit, voimakkuusfunktio		
Toimintayksikkö Tietopalveluyksikkö, 31600 Jokioinen		
ISSN 1238-9943	ISBN 951-729-524-3	<input type="checkbox"/> Tuloksia voi soveltaa luomuviljelyssä
Myynti: MTT tietopalveluyksikkö, 31600 JOKIOINEN Puhelin (03) 4188 7502 Telekopio (03) 4188 339	Sivuja 23 s.+ 2 liitettä	Hinta



Yliopistopaino
Helsinki 1998
ISBN 951-729-524-3
ISSN 1238-9943