



Graderingsmetodik i ogräsförsök

Lena Haby, Johannes Forkman, Anders Larsolle och Johan Mickelåker

Sveriges lantbruksuniversitet
Fakulteten för landskapsplanering, trädgårds- och jordbruksvetenskap

Alnarp, oktober 2010

Kontaktuppgifter till författarna:

Lena Haby, SLU Alnarp, Område Jordbruk - odlingssystem, teknik och produktkvalitet
lena.haby@ltj.slu.se
Tel. 040-415151

Johannes Forkman, FältForsk, SLU Ultuna
johannes.forman@et.slu.se
Tel. 018-671410

Anders Larsolle, SLU Ultuna, Inst. för energi och teknik
anders.larsolle@et.slu.se
Tel. 018-673446

Johan Mickelåker, DataVäxt AB
jmi@datavaxt.se
Tel. 0514-650200

Graderingsmetodik i ogräsförsök

Detta projekt är finansierat av Partnerskap Alnarp, Stiftelsen Lantbruksforskning, Svenskt Växtskydd samt Jordbruksverket. Författare är Lena Haby, SLU Alnarp (projektledare); Johan Mickelåker (fd. projektledare på SLU Alnarp), DataVäxt; Johannes Forkman (ansvarig för avsnittet "metoder för statistisk bearbetning av graderingsdata"), SLU FältForsk; Anders Larsolle (ansvarig för projektdel om utveckling av bildanalysdataprogram), SLU Ultuna.

Bakgrund

Fältförsöksverksamheten i Sverige har lång erfarenhet av att utföra ogräsförsök och dokumentera effekterna av kemiska bekämpningsmedel. Resultatet har nästan undantagslöst beskrivits som procent ogräseffekt. Metoden för detta har varit att räkna och, i de flesta fall, väga ogräsen vid ett tillfälle, normalt fyra veckor efter sista bekämpningen. Denna metod har också använts i andra länder, men har efter hand i många fall övergivits för att istället ersättas av okulär ogräsbedömning. Ogräsgradering liksom räkning och vägning av ogräsen används för att visa hur snabb verkan preparaten har. Den slutgiltiga effekten av preparaten mäts i form av avkastningsmätningar i grödan. Beställande företag (främst växtskydds företag) har framfört synpunkter om att nyttan med ogräsvägning är begränsad och om att svenska försöksutförare också ska övergå till okulär bedömning. Detta skulle underlätta internationella effektjämförelser. Frågan har behandlats i FältForsk ämneskommitté Ogräs, där man beslutade att försöka genomföra detta projekt.

Fördelarna med en övergång till okulär ogräsbedömning är bl.a. följande:

- 1) En internationellt erkänd metod som finns omnämnd i EPPO guidelines (EPPO Standards PP1, 2nd Edition, Volume 4. Efficacy Evaluation of Herbicides & Plant Growth Regulators). Detta underlättar internationella försökssammanställningar,
- 2) Beställande företag föredrar metoden framför vägning,
- 3) Hela försöksparcellen bedöms,
- 4) Bedömning sker vid 2-3 tillfällen under säsongen, alltså inte bara en situationsbild så som är fallet i dagsläget,
- 5) Mer väderoberoende,
- 6) Snabbare

För att byta utvärderingssystem krävs det dock 1) underlag för att översätta tidigare försöksresultat till det nyare systemet, 2) utbildning och träning av försöksutförande personal, samt 3) anpassning av statistik- och informationshanteringsprogram.

Under 2007 avslutades två projekt om platsspecifik ogräsbekämpning i vårsäd. En viktig del i dessa projekt var att utveckla den bildanalysalgoritm som beräknar antal och storlek av ogräs. I dagsläget anser man att metoden är relativt robust och säker för vårsäd, samt att kraven på bildtagning enkelt kan uppfyllas med vanlig digitalkamera. En förutsättning för att kunna implementera platsspecifik ogräsbekämpning med bildanalys är att metoden kan jämföras med försöksdata, samtidigt som en koppling mellan försöksdata och bildanalysresultat skulle kunna användas för att utveckla, förbättra och bredda användningen av bildanalys för ogräsidentifiering.

Syfte

Det övergripande syftet med projektet var att möjliggöra övergång från nuvarande bedömningsmetod för ogräseffekt till okulär ogräsgradering. Mer specificerat var syftet att

- jämföra och studera sambandet mellan traditionell metod med okulär ogräsgradering
- utbilda försökspersonal och studera hur arbetssituationen i försöksverksamheten förändras vid övergång
- studera möjligheten att i framtiden införa bildanalys för bedömning av ogräsmängd och ogräseffekt, samt jämföra resultatet från bildanalys med ogräsräkning och okulär ogräsgradering.

Metod

Innan säsongen 2007 utarbetades instruktioner för fotografering och okulär gradering som skickades till försöksutförarna.

Inga unika fältförsök har genomförts enbart för detta projekt, utan istället har befintliga försöksserier utnyttjats. Försöksserierna har legat i södra och mellersta Sverige. De försöksplaner som använts för projektet har varit L5-400 "Ogräsbekämpning i vårsäd" (vårkorn; fem försök 2007 och sex försök 2008 ingick i studien), samt L5-300S "Örtogräs i höstvetete, vårbehandling" (fem försök 2007 och ett försök 2008 ingick i studien).

Den 25 maj 2007 hölls en utbildningsdag för försöksutförare i de serier som var tänkta att ingå i projektet. Dagen började med teoretisk genomgång av metoder, men huvuddelen var praktiska övningar, förlagda i försöksfälten på Klostergården, Linköping. Totalt 16 personer medverkade under kursdagen. Morten Nygaard, Dow Agro, och Sven-Åke Rydell, HS Östergötland, höll i övningarna. Motsvarande kurs hölls den 16 maj 2008 i Vreta Kloster (10 st deltagare från växtskyddsfirma, Hushållningssällskapet, Växtskyddscentralen samt SLU) samt den 13 juni 2008 på Hellegården i Kristianstad (9 st deltagare från bl.a. Hushållningssällskapet).

Stativ för fotografering skickades ut till samtliga försöksutförare strax efter kursdagen.

Delar av försöksdatan har samlats in genom FältForsk databas. Statistisk bearbetning har utförts av FältForsks statistiker. Tabell 1 visar en sammanställning av de okulära graderingar samt ogräsräkningar/vägningar som genomförts.

Tabell 1. Förteckning över de graderings- samt räkning/vägningsaktiviteter som utfördes under 2007 och 2008

Försöksserie	Gröda	Försöksplats	Ansvarig	Bek.datum	Räkn/Vägn.	Gradering 1	Gradering 2	Gradering 3	Foto
L5-400-2007	Vårkorn	Mellby, Laholm	Erik Ekre	18-maj	24-jul	31-jul	-	-	-
L5-400-2007	Vårkorn	Osvall, Vintrosa	Lars D.	21-maj	25-jun	13-jun	15-jul	-	-
L5-400-2007	Vårkorn	Klostergården, Vreta Kloster	Sven-Åke Rydell	16-maj	14-jun	13-jun	10-jul	15-aug	-
L5-400-2007	Vårkorn	Brunnby gård, Västerås	Lars D.	03-jun	14-jun	17-jul	18-aug	-	18-jul
L5-400-2007	Vårkorn	Viggeby 3, Enköping	Lars D.	31-maj	11-jul	19-jul	-	-	19-jul
L5-400-2007	Vårkorn	Berget, Skara	Anna-Karin	22-maj	26-jun	15-jun	20-jul	16-aug	15-jun
L5-300-2007	Höstvetete	Ölsta gård, Märra	Lars D.	Ja	Ja	Ja	Ja	Ja	21-jun
L5-300-2007	Höstvetete	Brunnby gård, Västerås	Lars D.	Ja	Ja	Ja	Ja	Ja	-
L5-300-2007	Höstvetete	Hyltinge, Borensberg	Sven-Åke Rydell	Ja	Ja	Ja	Ja	Ja	Ja
L5-300-2007	Höstvetete	Solö, Fogdö	Lars Danielsson	Ja	Ja	-	-	-	-
L5-300-2007	Höstvetete	Härlingstorp, Kvänum	Anna-Karin	Ja	Ja	Ja	Ja	-	14-jun
L5-300-2007	Höstvetete	Logården, Gråstorp	Anna-Karin	Ja	Ja	Ja	-	-	12-jun
L5-300S-2008	Höstvetete	Uddetorp, Skara	Anna-Karin Krijger	Ingick ej	Ingick ej	Ja	Ja	Ja	-
L5-400-2008	Vårkorn	Sätuna, Björklinge	Lars Danielsson	Ingick ej	Ingick ej	Ja	Ja	-	Ja
L5-400-2008	Vårkorn	Brestorp, Skänninge	Sven-Åke Rydell	Ingick ej	Ingick ej	Ja	Ja	-	Ja
L5-400-2008	Vårkorn	Byvägen 16, Halmstad	Erik Ekre	Ingick ej	Ingick ej	Ja	Ja	Ja	-
L5-400-2008	Vårkorn	Ledsberget, Vedum	Anna-Karin Krijger	Ingick ej	Ingick ej	Ja	Ja	Ja	-
L5-400-2008	Vårkorn	Osvall, Vintrosa	Lars Danielsson	Ingick ej	Ingick ej	Ja	Ja	-	-
L5-400-2008	Vårkorn	Brunnby gård, Västerås	Lars Danielsson	Ingick ej	Ingick ej	Ja	Ja	-	-

Instruktionen var att ogräsräkning samt vägning skulle utföras en gång per försök (normalt fyra veckor efter sista behandlingen) och ogräsgradering skulle göras tre gånger; fyra, åtta respektive tolv veckor efter sista behandlingstillfället. Tolv veckor efter avslutad behandling var ofta vid skörd. Vid tidigare skörd gjordes sista graderingen vid skördetillfället. Vid varje ogräsgraderingstillfälle bestämdes i det obehandlade kontrollledet vilka ogräsarter som skulle ingå i jämförelsen mellan de olika behandlingarna i försöket. I det obehandlade ledet noterades den totala täckningsgraden av ogräs, 0-100 % (hela skalan skulle användas). Därefter specificerades en till fem arter samt en grupp för övriga ogräs och hur stor andel av hela ogräsmängden som dessa utgjorde. Summan skulle vara 100 %. Vilka ogräs som skulle specificeras avgjordes av om det framgick av försöksseriens syfte samt vilka ogräsarter som i huvudsak var representerade på försöksplatsen. Arterna inom gruppen övriga ogräs noterades. Vid gradering av de behandlade leden jämfördes sedan effekten hela tiden mot kontrollrutan för de arter som det noterats täckningsgrad för i kontrollrutan. Behandlingseffekten på

ogräsförekomsten angavs som 0-100 % (hela skalan skulle användas) ogräseffekt vid gradering i förhållande till kontrollrutans ogräsmängd och ogräsen utveckling i samma block. Det är möjligt att i vissa fall få mer ogräs i en behandlad ruta än i kontrollrutan. Detta saknar praktisk betydelse eftersom detta preparat i så fall har ingen eller mycket liten effekt mot den specifika ogräsarten. Behandlingseffekten anges därför som 0 % och inte som ett negativt värde. Effekten på hela försöksrutan bedömdes. 0 % innebar att det inte var någon effekt på ogräsen i jämförelse med obehandlat led, 50 % innebar lätta skador, ogräsen växte fortfarande och 100 % ogräseffekt innebar att alla ogräs var döda. Hänsyn togs till dvärgväxt och andra skador på ogräsen. Vid sista graderingstillfället (tolv veckor efter avslutad behandling) graderades ogrästäckningen i samtliga parceller, utöver ogräseffekten i behandlade led.

Vid bedömning av ogräseffekten där man använt den traditionella ogräsräknings- och vägningsmetoden, så jämför man ogräsvikten eller ogräsantalet i den behandlade rutan med den obehandlade rutan. När man vill titta på om det finns någon korrelation mellan ogräsvikt och antal ogräs samt resultaten från ogräsgradering, så är detta endast möjligt att göra för de obehandlade kontrollrutorna. Detta eftersom effekten i de behandlade rutorna som ogräsgraderats inte är ett oberoende mätvärde, utan står i relation till ogräsförekomsten i kontrollrutan.

Fotografering av försöken för bildanalys utfördes rutvis vid ett tillfälle under säsongen 2007 i en del av försöken. Dessa gjordes i samband med något av de två första tillfällena för okulär ogräsgradering, dvs. normalt fyra respektive åtta veckor efter sista behandling. Under 2008 utfördes fotografering rutvis en gång, tidigast en vecka innan eller i samband med bekämpning.

Graderingar samt ogräsräkning och vägning har utförts i försök under både 2007 och 2008 i syfte att lära upp försökspersonal. Försöksresultaten från 2008 har inte används för någon analys, utan försöksutförandet detta år gjordes enbart i övningssyfte.

Från sammanställningen av 2007 års resultat var en slutsats att det behövdes en annan statistisk metod för behandling av datamaterialet. Detta har Johannes Forkman på FältForsk arbetat vidare med under 2008, men då fortfarande med 2007 års data. Samliga analyser gjordes i SAS, version 9, per kombination av försök, tidpunkt och art. Proceduren GLM användes för metoderna 1 och 2 och 5, proceduren GLIM för metod 3 och proceduren FREQ för metod 4. Följande statistiska metoder prövades:

1. Vanlig variansanalys (ANOVA) på graderingsdata y
2. Analys på transformerade data: (a) Logaritmisk transformation: $\ln(y)$, (b) Roten ur: $y^{1/2}$, (c) Dubbel transformation: $\ln(y^{1/2})$, (d) Arcus-sinus-transformation: $\arcsin(y^{1/2})$
3. Generaliserad linjär modell
4. Friedmans test
5. Randomiseringstest

Under 2008 utfördes även en jämförelse mellan hur olika graderare bedömde samma försök. Detta utfördes under kursdagen för försöksutförarna som hölls i Kristianstad samt Vreta Kloster.

Resultat

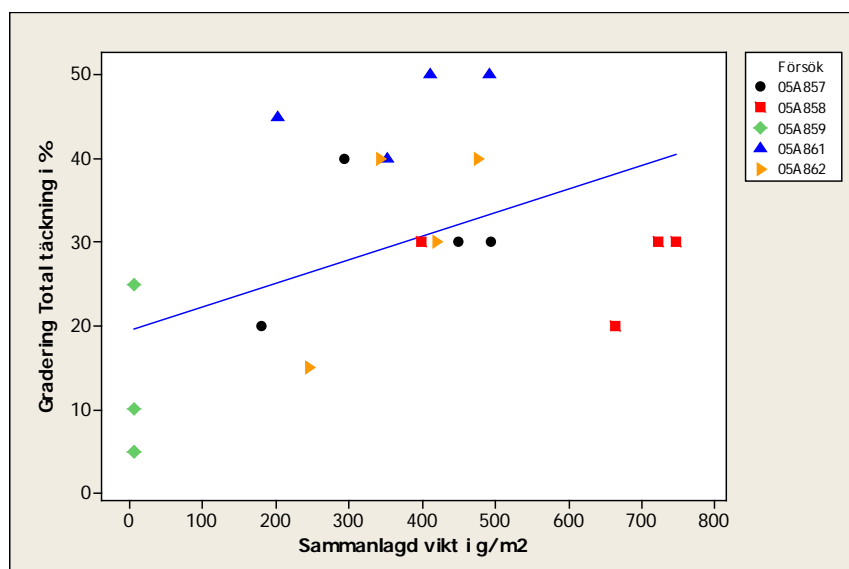
Korrelationen mellan de två metoderna för att bedöma ogräseffekt

I Figur 1-3 visas korrelationen mellan de två olika metoderna för att bedöma bekämpningseffekten av ogräsen år 2007 i försöken med vårkorn (L5-400). De jämförelser som tagit med är de jämförelser som innan projektets start ansågs mest troliga för att ge en god korrelation mellan de två metoderna. Varje punkt i figurerna utgör resultatet för en ogräsart alternativt ”övriga ogräs” i ett block i ett specifikt försök. För att korrelationen mellan metoderna ska vara god så ska punkterna ligga samlade runt den heldragna linjen. Resultaten från höstveteförsöken 2007, L5-300S, redovisas inte med figurer i denna rapport, men korrelationerna är jämförbara med de som redovisas för vårkornförsöken.

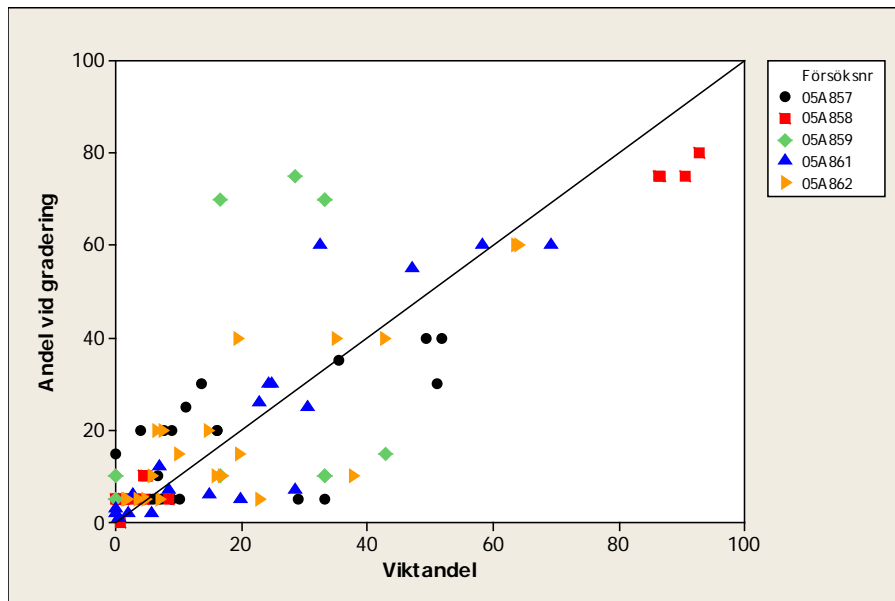
Figur 1 visar tydligt att det inte finns något tydligt samband mellan den totala täckningsgraden i kontrollrutan vid första graderingstillfället med okulär gradering och den sammanlagda ogräsvikten i kontrollrutan genom befintlig metod. Båda metoderna utfördes vid ungefär samma tillfälle (maximalt tolv dagars skillnad).

Andelen för respektive ogräsart i kontrollrutan utav totala ogräsförekomsten vid första graderingstillfället med okulär gradering stämmer ganska väl överens med viktandelen i kontrollrutan för respektive ogräsart utav totala ogräsvikten med befintlig metod (Figur 2). Men även om korrelationen mellan andel vid gradering och viktandel är relativt hög (0,807), så visar figuren att med den ”ideala” linjen som har korrelationen 1:1, att det kan vara svårt att översätta resultat mellan de olika graderingsmetoderna. Båda metoderna utfördes vid ungefär samma tillfälle (maximalt tolv dagars skillnad).

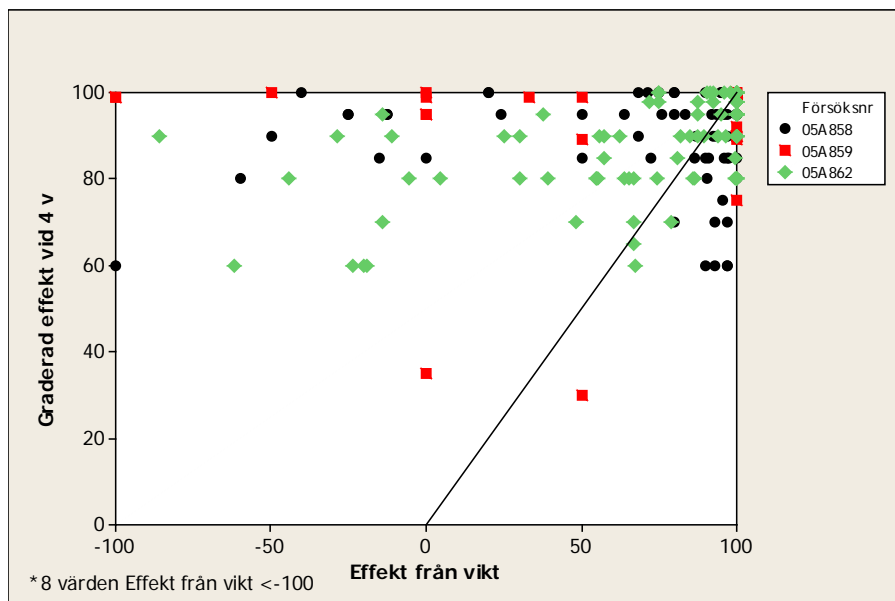
Figur 3 visar att det inte finns något klart samband mellan ogräseffekt i de behandlade parcellerna i förhållande till kontrollrutan genom okulär gradering vid första graderingstillfället och ogräsvikt i behandlade led i förhållande till kontrollrutan genom befintlig metod. Båda metoderna utfördes vid ungefär samma tillfälle (maximalt tolv dagars skillnad). Sambandet var också mycket dåligt för samma jämförelse mellan metoderna vid andra och tredje graderingstillfället. Vid dessa senare okulära graderingar gjordes fortfarande jämförelsen med befintlig metod på resultat från ogräsvägningen som utfördes i samband med första graderingstillfället ca fyra veckor efter bekämpning. Alltså jämfördes i detta fall effekterna från okulär gradering med vikteffekterna från ogräsvägning vid ett tidigare utvecklingsstadium.



Figur 1. Den totala täckningsgraden i kontrollrutan vid första graderingstillfället med okulär gradering mot sammanlagda ogräsvikten i kontrollrutan genom befintlig metod. Båda metoderna utfördes vid ungefär samma tillfälle (maximalt tolv dagars skillnad). Försök L5-400; vårkorn. Varje punkt utgör den totala ogräsförekomsten i ett utav fyra block. Den heldragna regressionslinjen är anpassad efter punkterna och beskriver endast variationen i 18,4 % av fallen.



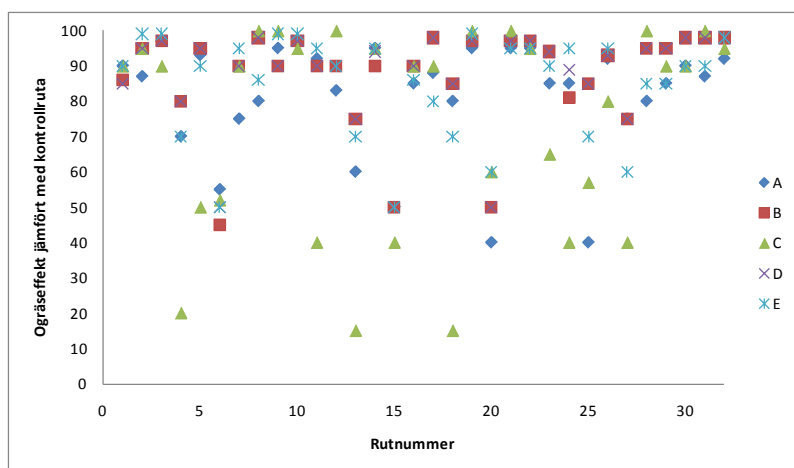
Figur 2. Andelen för respektive ogräsart i kontrollrutan utav totala ogräsförekomsten vid första graderingstillfället med okulär gradering mot viktandelen i kontrollrutan för respektive ogräsart utav totala ogräsvikten med befintlig metod. Båda metoderna utfördes vid ungefär samma tillfälle (maximalt tolv dagars skillnad). Försök L5-400; vårkorn. Korrelationen mellan de båda metoderna var 0,807. Den heldragna linjen symboliserar en perfekt korrelation mellan metoderna.



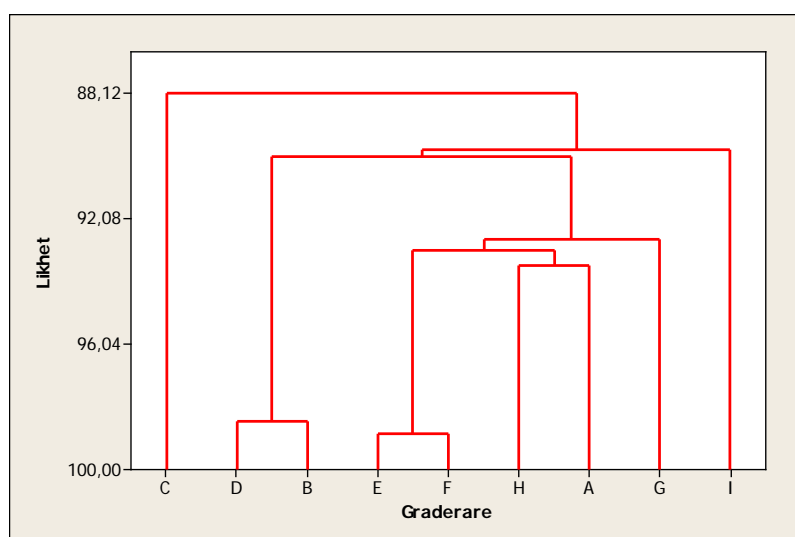
Figur 3. Ogräseffekt i de behandlade parcellerna i förhållande till kontrollrutan genom okulär gradering vid första graderingstillfället, mot ogräsvikt i behandlade led i förhållande till kontrollad med befintlig metod. Båda metoderna utfördes vid ungefär samma tillfälle (maximalt tolv dagars skillnad). En viktig skillnad mellan metoderna är att ogräseffekt som mäts som vikt kan ge negativa värden, vilket inte är möjligt för den graderade ogräseffekten. Försök L5-400; vårkorn.

Jämförelse mellan olika graderare som bedömt samma försök

Nedan redovisas hur olika graderare bedömt ogräseffekten i samma försök vid samma tillfälle genom okulär gradering i försök 05A914 i Skepparslöv (Figur 4 och 5). Graderarna är samma personer i både Figur 4 och 5.



Figur 4. Figuren visar hur fem olika graderare (A-E) har bedömt ogräseffekten av arten svinmålla i 32 olika parceller i samma försök (05A914) i förhållande till artens förekomst i en kontrollruta.



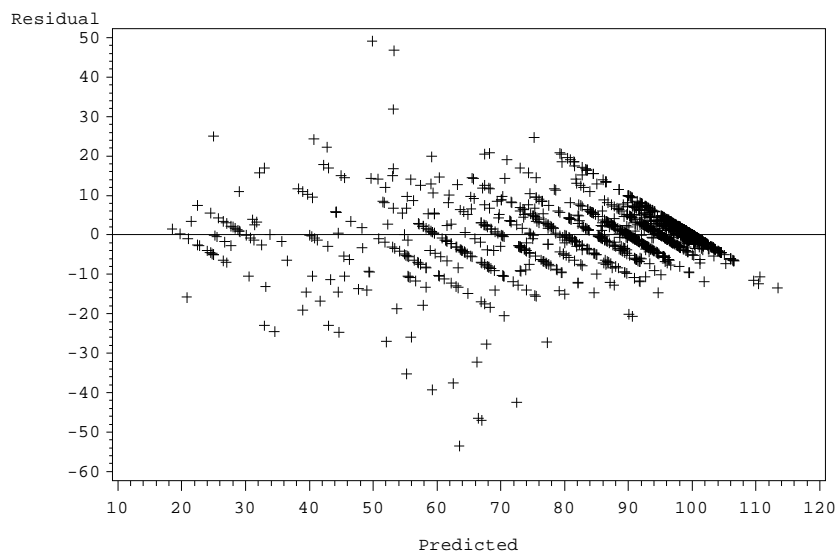
Figur 5. Figuren visade hur lika eller hur varierande olika personer graderade ogräseffekten av svinmålla i samma försök(05A914). Det är tydligt att graderare D och B gått tillsammans och diskuterat liksom graderare E och F.

Det är rimligt att olika graderare bedömer effekten på en viss ogräsart till olika procenttal, men den inbördes ordningen av effekten i olika led borde vara samma för alla graderare. Så var det för många ogräsarter, men inte för alla. Endast resultaten för svinmålla redovisas i rapporten. Sammanfattningsvis kan man konstatera att resultatet av graderingen beror på vem som graderar. Det bör dock noteras att flera av graderarna i detta test utförde okulär ogräsgradering för första gången eftersom testet utfördes under introduktionskursen till metoden.

Metoder för statistisk behandling av graderingsdata

Utredningen som presenteras under denna rubrik syftar till att hitta metoder för statistisk analys av graderingsdata från fältförsök. Utredningen redovisas i sin helhet i bilaga 1.

Figur 6 visar residualerna från en vanlig variansanalys (ANOVA) med antagande om normalfördelning och homogen varians. Om dessa antagande är korrekta så ska residualerna ligga som ett slumpmässigt moln kring noll på y-axeln, dvs. inte vara beroende av, i detta fall, effektnivån. I figur 6 finns emellertid ett tydligt mönster. För behandlingar med stor effekt kan inte enskilda rutor avvika särskilt mycket uppåt, men betydligt mer nedåt, och tvärtom. Nära 0 och nära 100 finns inte stort utrymme för variation. Graderingarna varierar mer när effekten är ca 50 än när effekten är ca 0 eller ca 100. Vanlig variansanalys baserad på normalfördelning ger också upphov till predicerade graderingar större än 100.



Figur 6. Analys utan transformation. Residualer plottade mot predicerade värden. Det diagonala linjemönstret beror på att graderare inte utnyttjar hela skalan från 0 till 100, utan föredrar jämna 5- och 10-tal.

Skalan för okulär gradering blir något märklig eftersom den i teorin är kontinuerlig, men inte riktigt i praktiken. Exempelvis är det antagligen mer troligt att en graderare sätter värdet 95 än värdet 94, även om det är fullt möjligt att använda en kontinuerlig skala (Figur 6). Det är också ett problem för analysen, och jämförelsen med räkning/vägning, att graderingarna aldrig kan överstiga 100, inte ens när ogräset är mer utbrett i den behandlade rutan än i den obehandlade. Beräknad effekt utifrån räkning/vägning kan däremot vara negativ, vilket inte är fallet med effekt utifrån okulär gradering.

Ingen av de undersökta transformationerna lyckades framgångsrikt skapa lika varians och normalfördelning. De tre första transformationerna är egentligen ämnade för situationen att variansen ökar kontinuerligt. Därför är det inte förvånande att de fungerade dåligt. Mer överraskande är att arcus-sinus-transformationen inte fungerade bättre. På grund av att graderarna föredrar tal som är jämnt delbara med 5 skapas en speciell fördelning med problem som inte försvinner med transformationer.

Den binomiala ansatsen (generaliserad linjär modell) är inte heller lämplig för graderingsdata. Man kan inte förvänta sig att metoden ska konvergera när något led eller block genomgående graderats till 100. I den undersökta datamängden är det vanligt att effekten är maximal på ett stort antal rutor.

Friedmans test fungerar dåligt när antalet block är litet, vilket i fältförsök är regel snarare än undantag. För graderingsdata är Friedmans test särskilt olämpligt eftersom det är vanligt att flera observationer är lika. När man analyserar rangtal istället för ursprungliga data går också mycket av informationen i försöket förlorad.

Randomiseringstestet ger exakta sannolikhetsvärden. Nackdelen är att det krävs ett stort antal randomiseringar, och att det inte finns någon färdig procedur i SAS som utför testet. Det är emellertid möjligt att göra 10 000 randomiseringar på bara några sekunder, så det bör inte vara något större problem att införa metoden. Tyvärr är det inte möjligt att med randomiseringstest få signifikanta resultat i parvisa jämförelser när antalet block är mindre än 6.

Det största problemet med graderingsdata, när det gäller statistisk analys, är att olika par av led har olika varians. Om vi emellertid är intresserade av ett specifikt par, t.ex. C jämfört med B, är det alltid möjligt att studera det paret separat och göra ett vanligt t-test under antagande om normalfördelning. Om varken led B eller led C är approximativt normalfördelade tenderar ändå fördelningen för skillnaden mellan medelvärdet av led C och medelvärdet av led B vara mer lik normalfördelningen.

Diskussion och slutsatser

Sambandet mellan traditionell metod och okulär ogräsgradering:

- Det saknades en korrelation mellan resultat från ogräsvägning/räkning och ogräsgradering.
- Vid gradering ska hela parcellens ogräseffekt bedömas. Detta görs genom att titta på ett par olika ställen i parcellen genom att böja undan grödan med en pinne. Om ett efterfrågat ogräs inte hittas direkt, så får graderaren titta på fler ställen. Resultatet kan därför skilja mellan olika graderare eftersom inte alla tittar på samma platser i parcellen. Även den befintliga metoden med vägning och räkning av ogräsen kan ge olika resultat för olika graderare eftersom mätytorna slumpas ut. Risken för skillnader mellan olika avläsningar är dock mindre med ogräsgradering än med befintlig räknings- och vägningsmetod eftersom en större del av parcellen bedöms.
- Resultatet av okulär ogräsgradering är väldigt beroende av vem som graderar.
- Även David Hanssons avhandling (2002) visar att det är stor skillnad mellan olika graderingspersoner vid okulär gradering av ogrästäckning. I en dos-responsstudie av ogräsbekämpning visade han att korrelationen mellan vikt och täckningsgrad är dålig vid låga och höga täckningsgrader. Med dos-responsförsök kunde han också visa att ED₉₀-värdet (den dos som krävs för att reducera friskvikten eller täckningsgraden med 90 %) var lägst för bildanalys (en automatiserad variant av okulär gradering), därefter plantvikt och högst för plantantal.

Utbildning av försökspersonal:

- Fokus i projektet låg i att utbilda försökspersonal i den nya ogräsgraderingsmetodiken, vilket också har genomförts.
- Kurserna var lyckade och hade god uppslutning.
- Det var svårt för försökspersonalen att få gjort alla graderingar.
- Under år 2008 skickades instruktioner ut för sent till försökspersonalen, vilket innebar att tiden redan hade passerat för vissa graderingstidpunkter samt fotografering.
- All försökspersonal som deltog i projektet var överens om att gemensamma graderingstillfällen var nödvändiga varje år för att få en samstämmig graderingsskala. Nya graderare skulle helst gå vid sidan om en mer erfaren graderare i början för att ”läras upp”.

Anpassning av statistik- och informationshanteringsprogram:

- Sammanställningen av 2007 års graderingsresultat visade att det behövdes en annan statistisk metod för analys av datamaterialet än vad som idag används för redovisning av ex. skörd. Den främsta anledningen till detta är att villkoret om lika stor variation i graderingarna i alla led inte är uppfyllt.
- Under 2008 fokuserade projektet på att finna en lämplig statistisk metod för behandling av graderingsresultat. Resultatet blev följande:
 - Graderingsdata är inte normalfördelade med homogen varians, varken före eller efter transformation. Därför kan vanlig variansanalys på ursprungliga data (ANOVA) eller på transformerade data ge missvisande resultat.

- Varken binomial analys (generaliserad linjär modell) eller Friedmans test fungerar tillfredsställande.
- Med randomiseringstest är det möjligt att beräkna exakta sannolikhetsvärden utan antagande om lika varians och normalfördelning.
- Parvisa t-test, baserade på antagandet om normalfördelning, kan ge ledning om vilka led som har olika effekt. Vid parvisa t-test jämförs nämligen ett par i taget, och då är det inget problem att variansen inte är lika för alla par.

Bildanalys:

- Befintlig bildanalysmetod är avsedd för att användas i tidiga stadier då man fortfarande ser grödraden. Detta motsvarar normalt tiden fram tills tidpunkten för kemisk ogräsbekämpning. Manuell gradering sker däremot senare, tidigast två veckor efter bekämpning. Vid denna tidpunkt täcker grödan raden och befintligt bildanalysprogram fungerar då inte. Den enda användningen av bildanalys i graderingssyfte är därför att använda bildanalys vid bekämpningstillfället för att dokumentera utgångsläget samt variationen i ogrästäthet mellan rutorna.
- Det var inte möjligt att utföra bildanalys på foto som tagits i samband med första eller andra ogräsgraderingstillfället, dvs. ca fyra respektive åtta veckor efter avslutad behandling. Även bekämpningstillfället visade sig vara en alltför sen tidpunkt för fotografering till bildanalys eftersom grödraderna redan då gick ihop och ogräsen till stor del var täckt av grödans blad (Figur 7).



Figur 7. Foto taget i samband med bekämpning (Foto: Lars Danielsson)

- Det visade sig vara svårt för försökspersonalen att ta foto för bildanalys vid bekämpning eftersom de hade mycket annat arbete att utföra vid denna tid på året.
- Befintligt program för bildanalys som tagits fram inom projektet "Platsspecifik ogräsbekämpning i skånsk vårsäd" kräver tillgång till ett specialprogram som är för dyrt och komplicerat för att vara av intresse för försöksutförarna. En lösning skulle kunna vara att försöksutförare skickar sina bilder för bildanalys till en resursperson med tillgång till samt kunskap om programmet. Dock saknas en organisation i försöks-Sverige för att det rutinemässigt ska vara möjligt att skicka många och tunga bildfiler för en sådan analys. Inom detta projekt gjordes en vidareutveckling av bildanalysprogrammet från projektet "Platsspecifik ogräsbekämpning i skånsk vårsäd". Målet var att skapa ett användarvänligt och licensfritt program så att försöksutförarna själv kunde analysera sina bilder och sedan lägga in resultaten i försöksdatabasen. Dessvärre var detta mer komplicerat än beräknat och gick inte att

uppnå inom ramen för detta projekt. Ny finansiering får därför sökas för att slutföra utvecklingen av programmet. Mer om utvecklingen av bildanalysprogrammet finns att läsa i en rapport av Haby och Mickelåker (2010), <http://pub-epsilon.slu.se:8080/1580/>

Publikationer och övrig resultatförmedling till näringen

Slutrapport på SLFs och Partnerskap Alnarps hemsida samt en något mer omfattande rapport på FältForsks hemsida, www.slu.se/faltforsk under fliken ”utbildning”. Publicering av faktablad i Alnarps egen faktabladsserie ”LTJ-fakultetens faktablad”.

Genom samordningen med befintlig försöksorganisation så har huvuddelen av målgruppen redan nåtts och fått ta del av resultatet. Genom ämneskommittéerna har ännu flera intressenter mötts. En kortare slutredovisning kommer också ske vid Växjö möte den 7 december 2010.

Projektet har delredovisats för FältForsks ogräskommitté i februari 2008 samt på ett möte som arrangerades av Svenskt Växtskydd i februari 2008. Svenskt Växtskydds möte handlade om hur ogräsförsöksverksamheten skulle utvecklas samt en diskussion om priser och ansvar mellan beställare (medlemsföretagen) och utförare (HS). Vid tre tillfällen under projektiden har en kursdag arrangerats för de personer som varit med och graderat försök i projektet. Kursdagarna hölls den 25 maj 2007 i Linköping (16 deltagare), den 16 maj 2008 på Vreta kloster (10 st deltagare från växtskyddsfirmor, Hushållningssällskapet, Växtskyddscentralen samt SLU) samt den 13 juni 2008 på Helgegården i Kristianstad (9 st deltagare från bl.a. Hushållningssällskapet).

Referenser

Hansson, D. (2002). Hot Water Weed Control on Hard Surface Areas. Diss. Alnarp: Sveriges lantbruksuniversitet. Kapitel: Assessment of weed control effect – development of image analysis and comparison with plant weight, plant numbers and visual assessment.

Haby, L. och Mickelåker, J. (2010). Global Navigation Satellite Systems (GNSS) i fältförsök. Rapport 2010:9. Fakulteten för landskapsplanering, trädgårdsgårds- och jordbruksvetenskap, SLU Alnarp.

Metoder för statistisk analys av graderingsdata

1. Inledning

Inom SLF-projektet Graderingsmetodik i ogräsförsök har okulär ogräsgradering utförts i fältförsök. Projektet syftar till att jämföra okulär ogräsgradering med traditionell räkning och vägning av ogräs. En viktig aspekt är möjligheten att analysera observationerna statistiskt. Vanlig variansanalys förutsätter normalfördelning och homogen varians. Dessa villkor är inte uppfyllda för graderingsdata. I detta PM utreds hur ogräsgraderingar skulle kunna analyseras statistiskt.

2. Syfte

Utredningen som presenteras i detta PM syftar till att hitta metoder för statistisk analys av graderingsdata från fältförsök. Metoderna ska ge bra

- i. skattningar av skillnader mellan led
- ii. test av skillnader mellan led
- iii. test av parvisa skillnader mellan led

3. Material och metoder

Olika analysmetoder har prövats på graderingsdata från projektet. Datamängden omfattar två serier: L5300, som omfattar 700 graderingar i höstvetete, och L5400, som omfattar 775 graderingar i vårkorn. Totalt har 11 försök utförts. Graderingar har gjorts på som mest 11 arter vid upp till 5 olika tidpunkter (4, 5, 6, 8 och 12 veckor). I varje försök jämfördes 5 behandlingsled med ett obehandlat led. Effekten av växtskyddsbehandlingarna graderades på en skala från 0 till 100 %. Alla försök var upplagda som fullständiga randomiserade blockförsök. Sammanlagt omfattar datamängden 81 kombinationer av försök, tidpunkter och arter. Varje sådan kombination innehåller graderingar från 5 behandlingsled i som mest 4 replikat.

I tabell 1 visas, som ett exempel, hur effekten på ett ogräs av behandlingarna B–F graderats i ett av försöken i datamängden. Behandling E hade stor effekt i samtliga block. Andra behandlingar var inte riktigt lika framgångsrika i försöket, men frågan är om de observerade skillnaderna är så stora att de är statistiskt signifikanta.

Tabell 1. Ogräsgraderingar i ett försök med 5 led (B–F) och 4 block (I–IV).

Led	I	II	III	IV
B	100	100	95	95
C	100	100	95	95
D	85	95	95	95
E	100	100	100	100
F	85	90	90	95

Följande statistiska metoder prövades:

1. Vanlig variansanalys (ANOVA) på graderingdata y
2. Analys på transformerade data
 - 2.1. Logaritmisk transformation: $\ln(y)$
 - 2.2. Roten ur: $y^{1/2}$
 - 2.3. Dubbel transformation: $\ln(y^{1/2})$
 - 2.4. Arcus-sinus-transformation: $\arcsin(y^{1/2})$
3. Generaliserad linjär modell
4. Friedmans test
5. Randomiseringstest

Samliga analyser gjordes i SAS, version 9, per kombination av försök, tidpunkt och art. Proceduren GLM användes för metoderna 1 och 2 och 5, proceduren GLIM för metod 3 och proceduren FREQ för metod 4.

Med metod 1 anpassades en modell med Led och Block som fixa faktorer (samtliga block var fullständiga, så det fanns ingen information mellan blocken om skillnader mellan led). Med metod 2 anpassades samma modell efter transformation av graderingarna y .

För varje analys gjordes två figurer: dels plottades residualerna mot predicerade värden, dels gjordes en så kallad qq-plot.

Med metod 3 förutsattes graderingarna vara binomialfördelade observationer med $n = 100$. Logit-länk användes. Pearsons χ^2 användes för justering för överspridning (se t.ex. Piegorsch och Bailer, 2005).

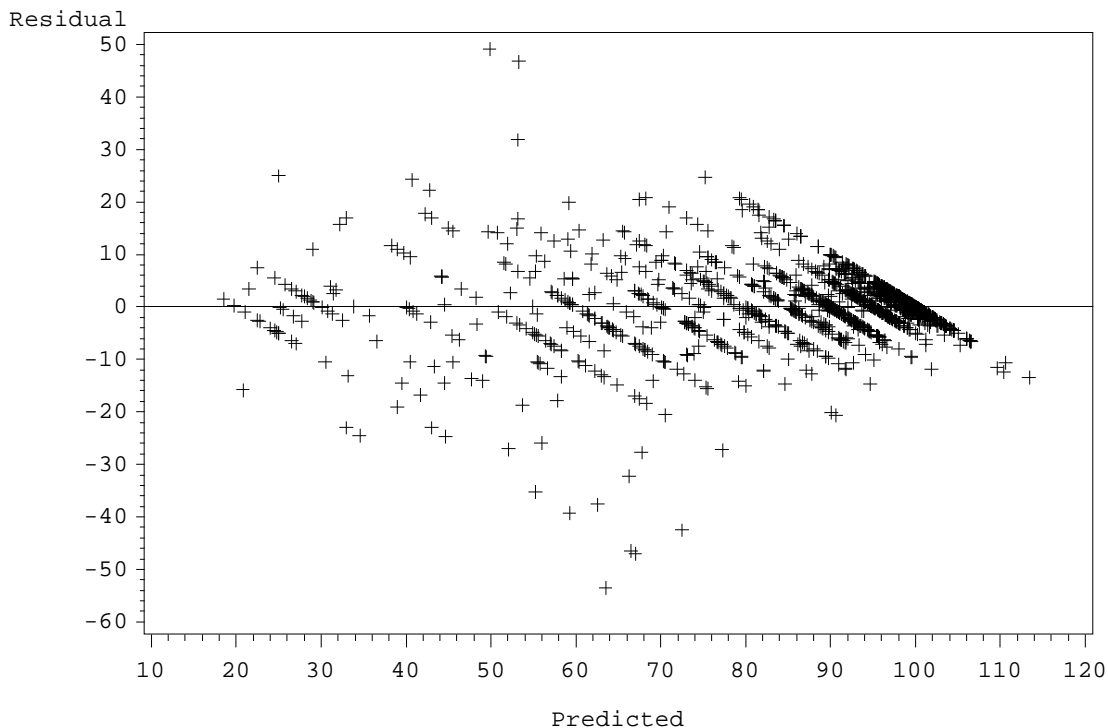
Med metod 4 rangordnades graderingarna inom block, och hypotesen att det inte finns några skillnader i rangtal mellan leden testades med χ^2 -test (Hollander och Wolfe, 1999).

Med metod 5 randomiserades leden inom block. Totalt gjordes 10 000 randomiseringar per kombination av försök, tidpunkt och art. Varje randomiserad datamängd analyserades med variansanalys som ett vanligt blockförsök. F-statistikan i den observerade datamängden jämfördes med fördelningen av F-statistikorna i de randomiserade datamängderna och sannolikhetsvärdet för hypotesen att leden är lika beräknades (Edgington, 1987).

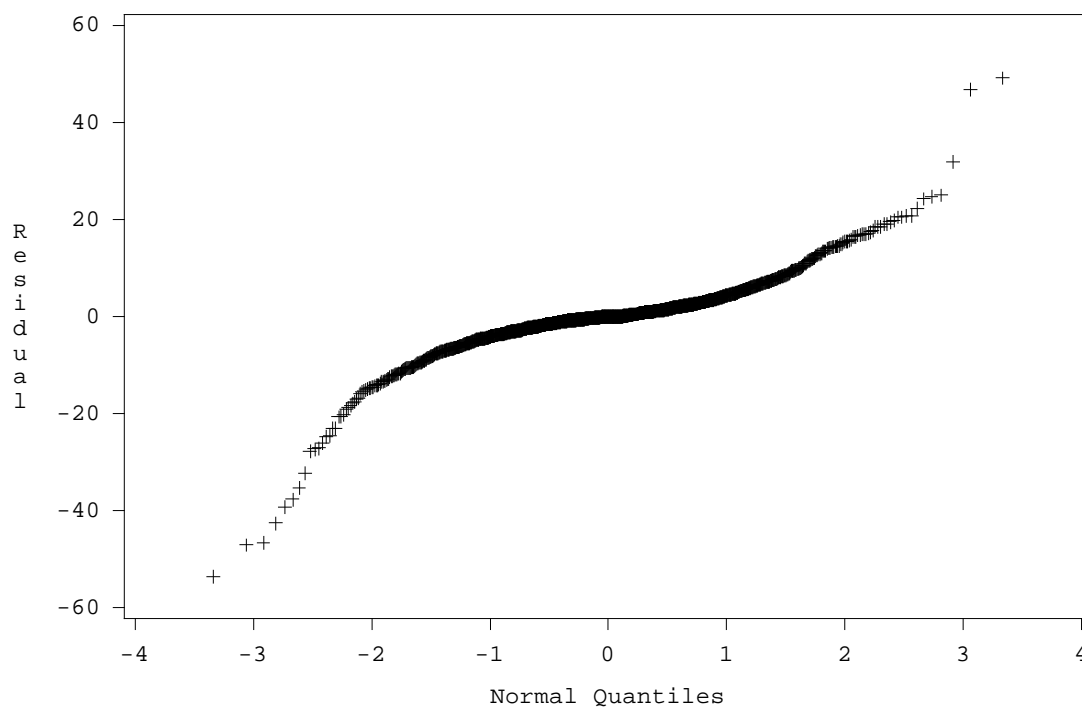
4. Resultat

4.1 Vanlig ANOVA

I figur 1 syns att residualerna är större för graderingar kring 50 % än för graderingar nära 0 % eller nära 100 %. Variansen beror på nivån. Av figur 2 framgår att residualerna inte är normalfördelade, för i så fall hade de legat på en rät linje.



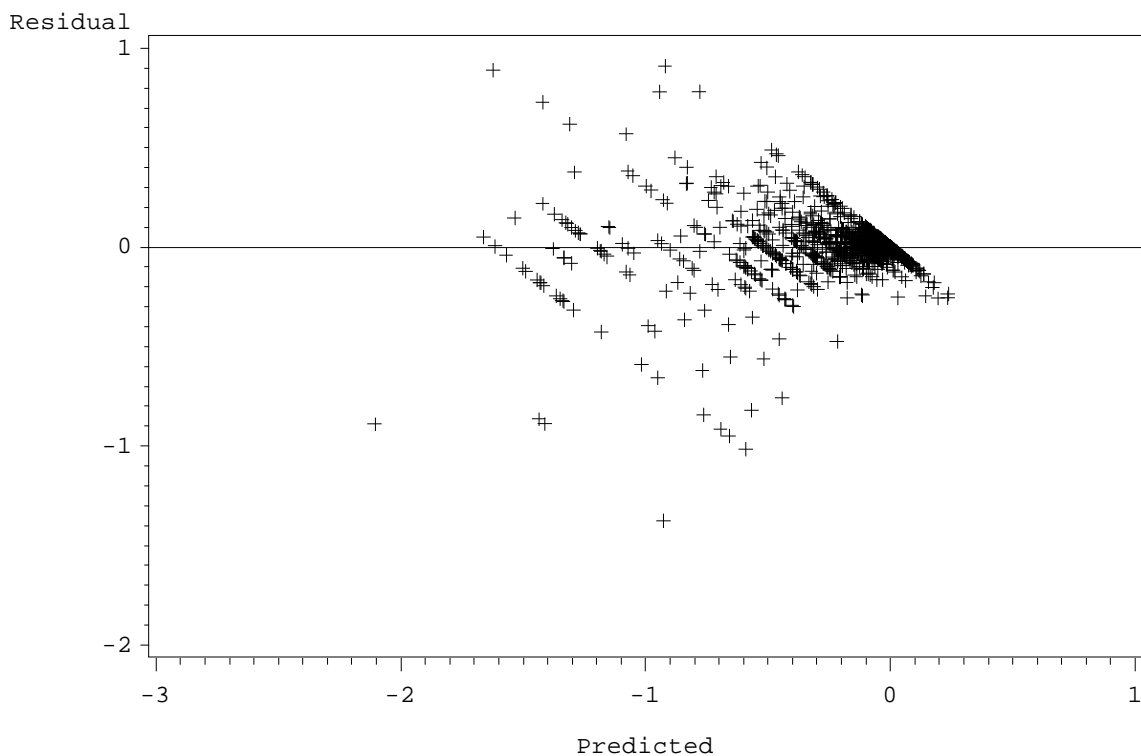
Figur 1. Analys utan transformation. Residualer plottade mot predicerade värden. Det diagonala linjemönstret beror på att graderare inte utnyttjar hela skalan från 0 till 100, utan föredrar jämna 5- och 10-tal.



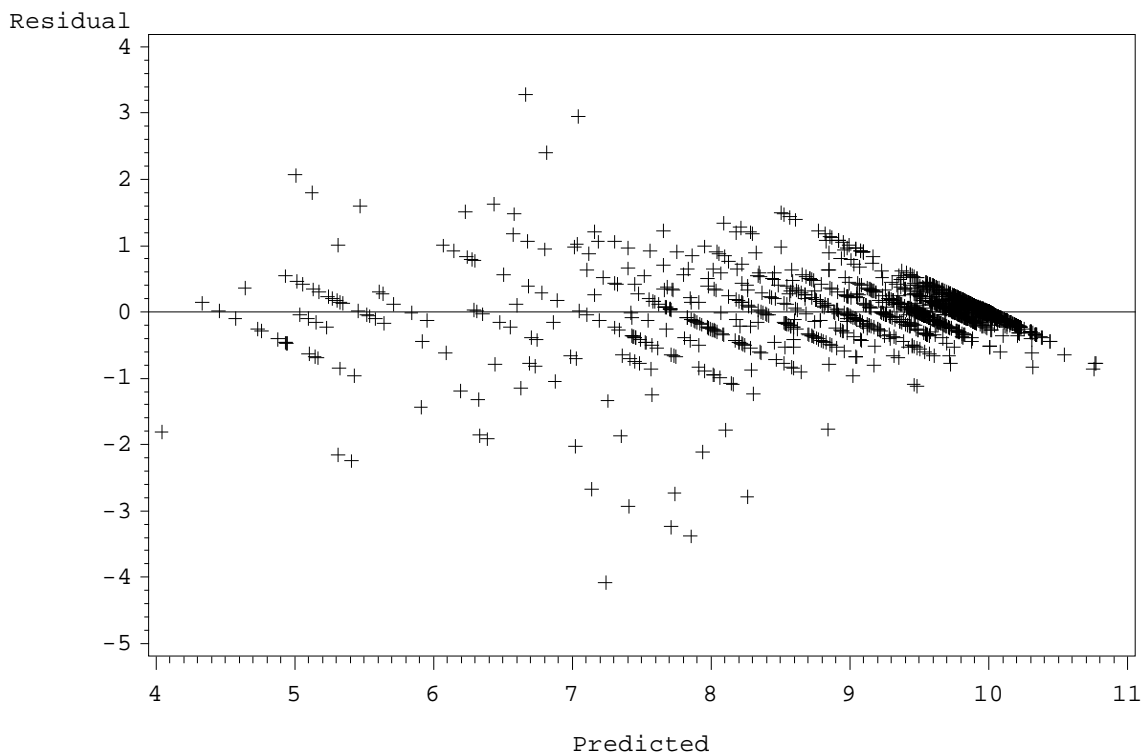
Figur 2. Analys utan transformation. Kontroll av normalfördelningsantagandet.

4.2 Analys av transformerade data

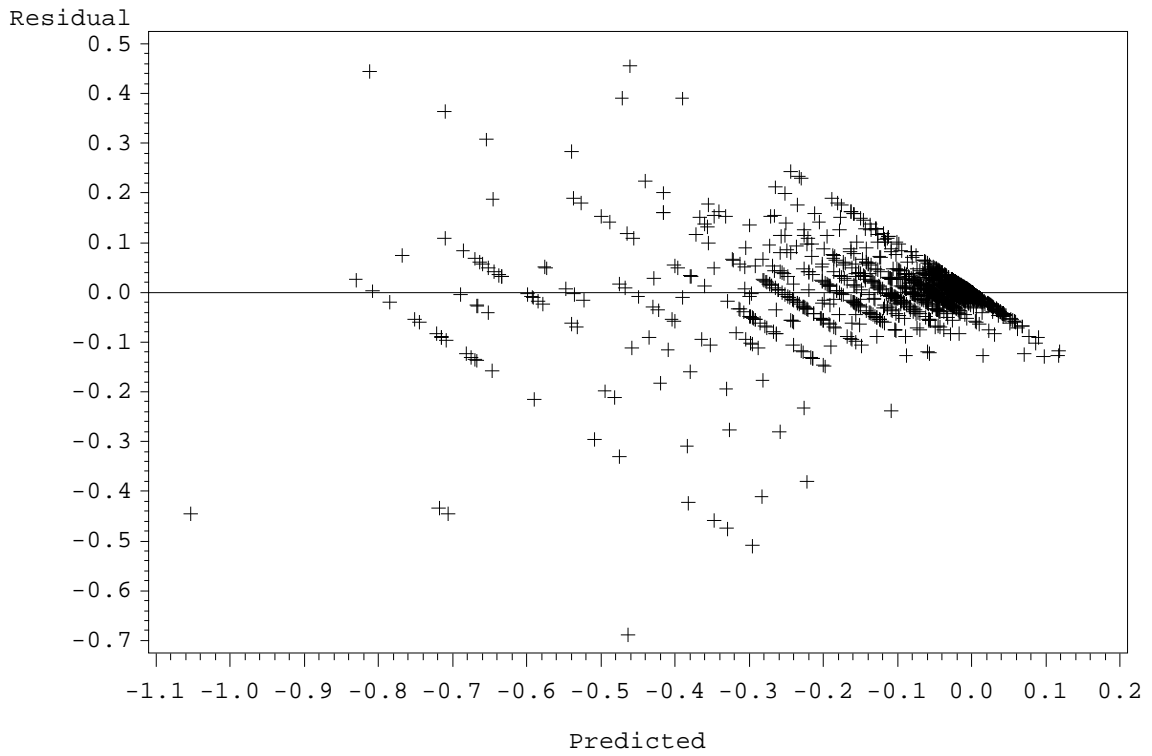
Ingen av de undersökta transformationerna lyckas tillfredsställa skapa en homogen varians. Detta framgår av figurerna 3–6.



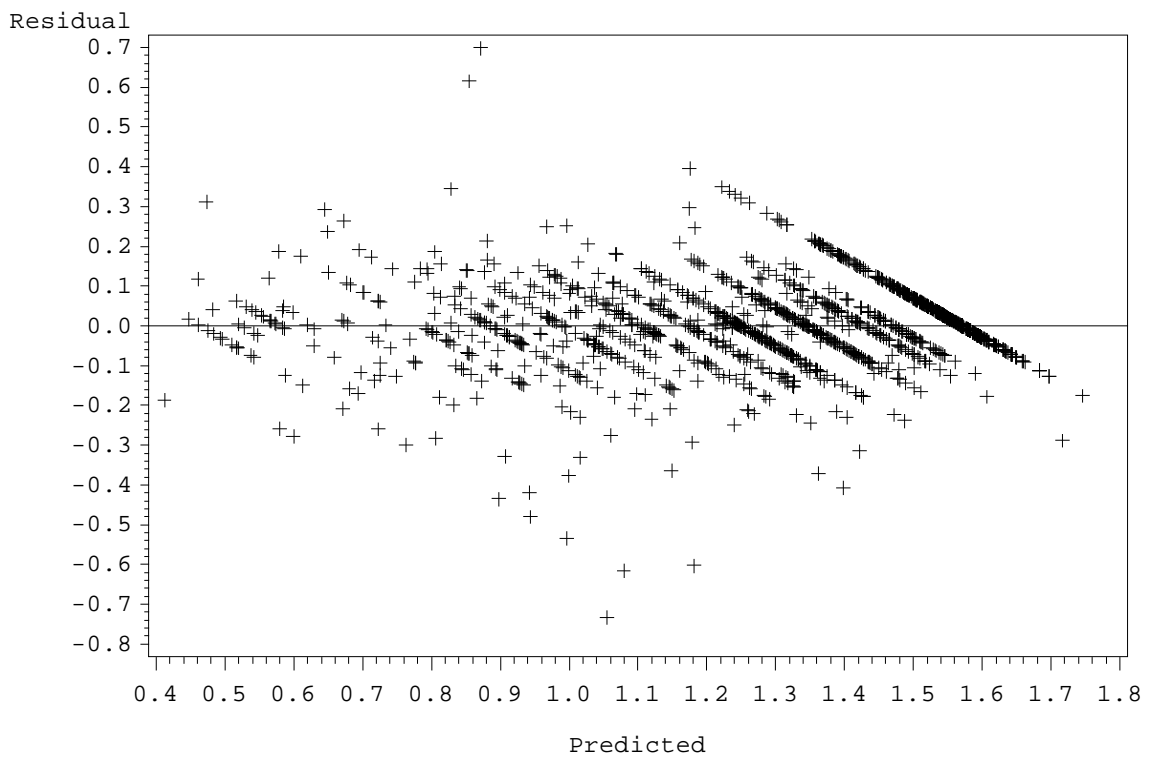
Figur 3. Analys av logaritmerade data. Residualer plottade mot predicerade värden.



Figur 4. Analys av roten ur graderingarna. Residualer plottade mot predicerade värden.

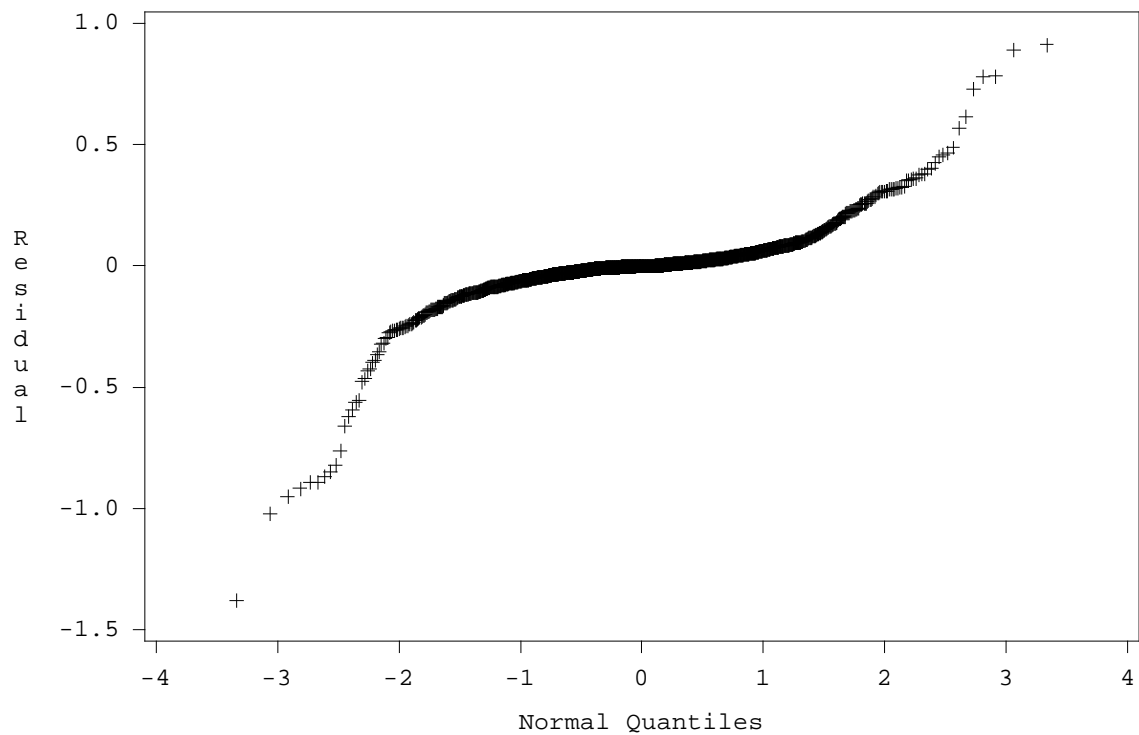


Figur 5. Analys av dubbeltransformerade data. Residualer plottade mot predicerade värden.

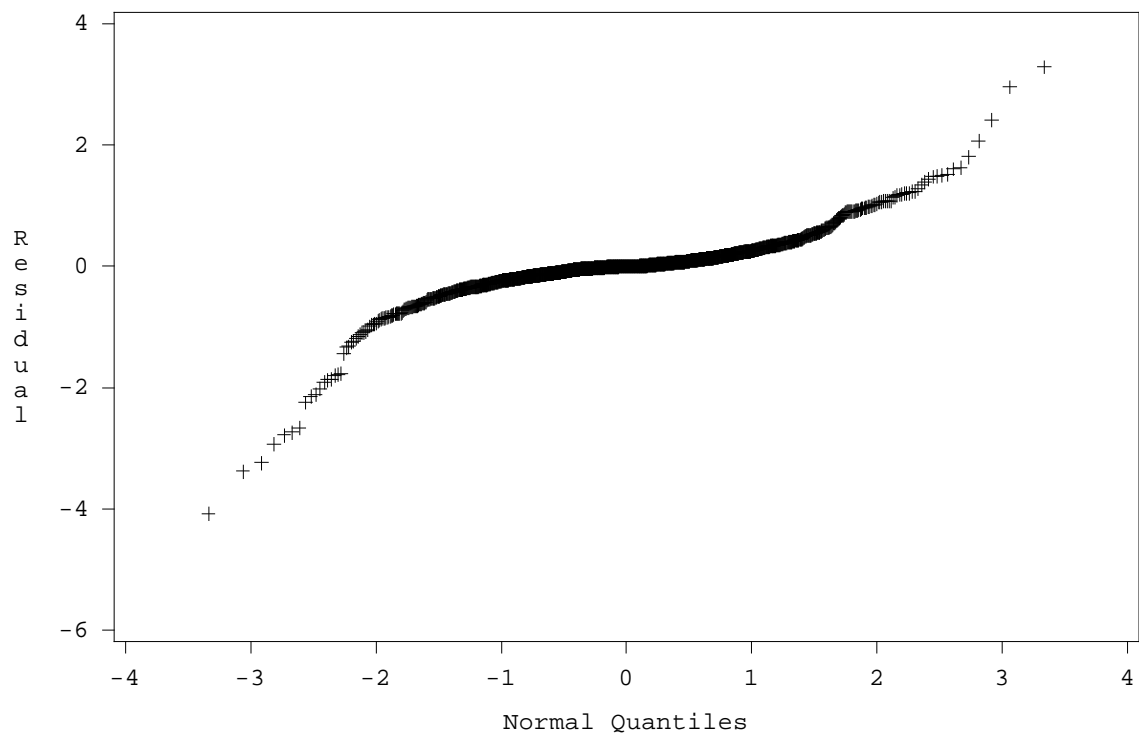


Figur 6. Analys av arcus-sinus-transformerade data. Residualer plottade mot predicerade värden.

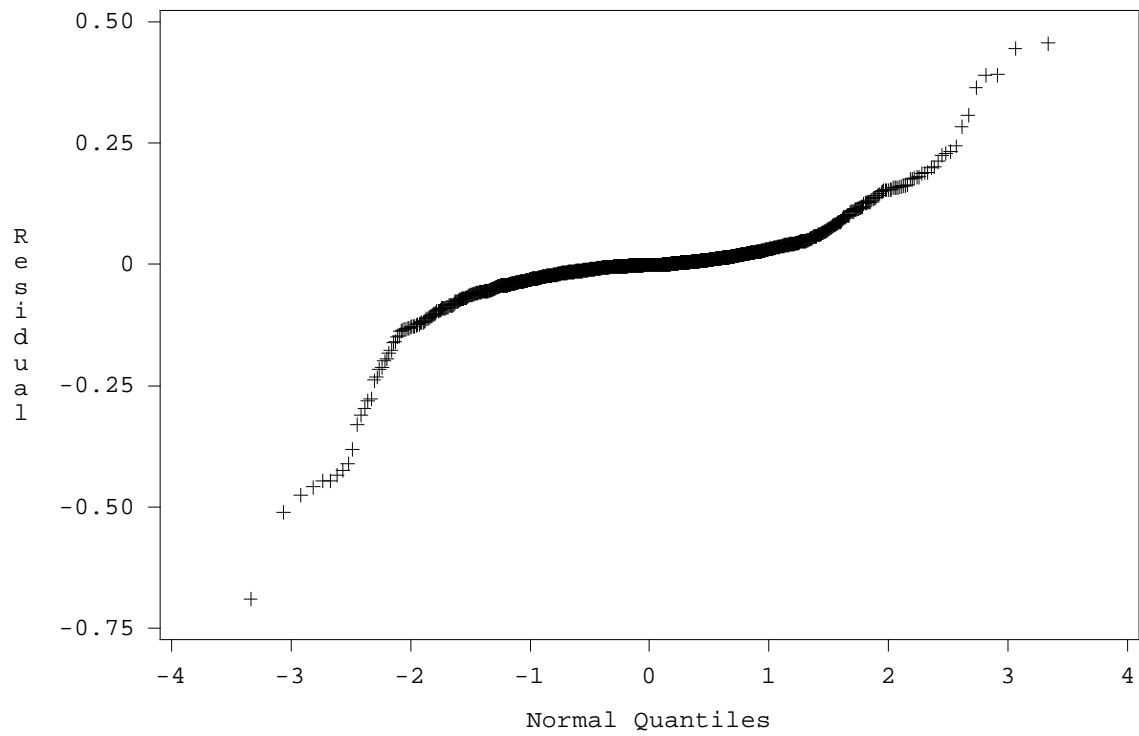
Inte heller lyckades någon transformation skapa normalfördelade residualer. I så fall hade nämligen punkterna i någon av figurerna 7–10 legat på en rät linje.



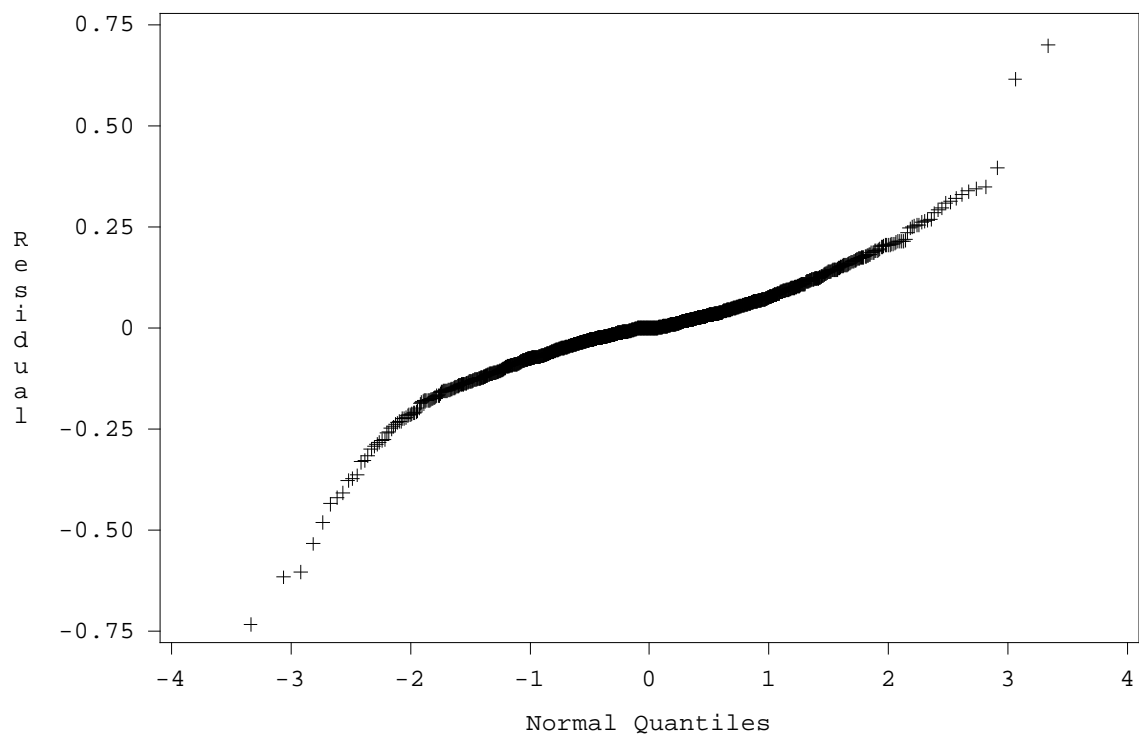
Figur 7. Analys av logitmerade data. Kontroll av normalfördelningsantagandet.



Figur 8. Analys av roten ur graderingarna. Kontroll av normalfördelningsantagandet.



Figur 9. Analys av dubbeltransformerade data. Kontroll av normalfördelningsantagandet.



Figur 10. Analys av arcus-sinus-transformerade data. Kontroll av normalfördelningsantagandet.

4.3 Generaliserad linjär modell

Generaliserade linjära modeller anpassas numeriskt. För de flesta av kombinationerna av försök, tidpunkt och art konvergerade metoden, men i några fall hittade SAS ingen lösning. Detta inträffade exempelvis för den kombination som redovisas i tabell 2.

Tabell 2. Graderingar per led (B–F) och block (I–IV)

Led	I	II	III	IV
B	100	80	100	100
C	100	100	100	100
D	100	100	100	100
E	70	100	100	100
F	100	98	100	70

Om emellertid maximala antalet n ändrades från $n = 100$ till $n = 1000$ och samtidigt graderingarna multiplicerades med 10 uppnåddes konvergens. Vid konvergens uppskattades samtliga led ha effekten 100, vilket måste anses vara egendomligt med tanke på att led F graderats till 70 i block IV (tabell 2).

4.4 Friedmans test

Friedmans test fungerade utom i några mindre intressanta fall med alltför liten variation i graderingarna. Sannolikhetsvärdena är dock tveksamma på grund av att χ^2 -testet är asymptotiskt. För att testet ska vara giltigt krävs att antalet block är stort, men i materialet förekom aldrig fler än 4 block. Testet är inte heller tillförlitligt när flera led har samma gradering i ett block. I datamängden är graderingar som är jämnt delbara med 5 över-representerade (82 % av graderingarna var jämnt delbara med 5). Därför hade flera led ofta samma gradering i ett block.

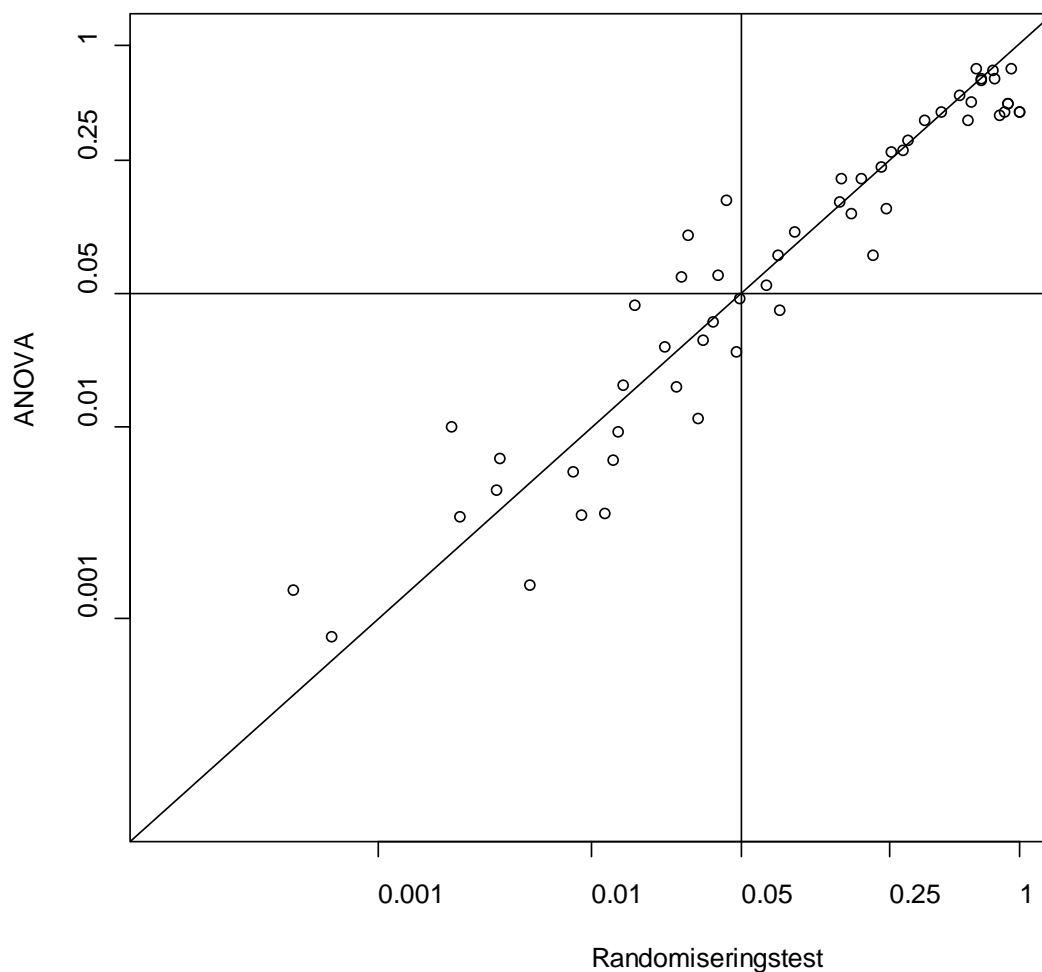
4.5 Randomiseringstest

Randomiseringstest ger korrekta sannolikhetsvärden för den övergripande hypotesen att det inte finns några skillnader mellan behandlingsleden.

I figur 11 jämförs den vanliga variansanalysens (metod 1) signifikansnivåer med randomiseringstestets (metod 2). Resultatet sammanfattas i tabell 3.

Tabell 3. Sannolikhetsvärden med ANOVA jämfört med randomiseringstest

	Randomiseringstest	
	P < 0.05	P > 0.05
ANOVA	4	30
P > 0.05		
P < 0.05	37	1



Figur 11. Sannolikhetsvärden med ANOVA jämfört med randomiseringstest

Det framgår av figur 11 och tabell 3 att den vanliga variansanalysen och randomiseringstestet för det mesta ger samma slutsats. Endast i 1 av de 38 fall variansanalysen gav ett signifikant resultat gav randomiseringstestet ett icke signifikant resultat. Endast i 4 av de 34 fall variansanalysen gav ett ickesignifikant resultat gav randomiseringstestet ett signifikant resultat.

Det är även möjligt att göra parvisa randomiseringstest. De två led som ska jämföras randomiseras inom block på de rutor som i verkligheten behandlats med de båda leden. Vid exempelvis 4 block finns $2^4 = 16$ möjliga randomiseringar. Minsta möjliga signifikansnivå vid tvåsidigt test är i så fall $2/16 = 0,125$. Det krävs minst 6 block för att det ska vara möjligt att få en signifikant skillnad mellan två givna behandlingar. I datamaterialet fanns aldrig fler än 4 block, så inga par av behandlingar var signifikant olika.

5. Diskussion

Ingen av de undersökta transformationerna lyckades framgångsrikt skapa lika varians och normalfördelning. De tre första transformationerna är egentligen ämnade för situationen att variansen ökar kontinuerligt. Därför är det inte förvånande att de fungerade dåligt. Mer överraskande är att arcus-sinus-transformationen inte fungerade bättre. På grund av att graderarna föredrar tal som är jämnt delbara med 5 skapas en speciell fördelning med problem som inte försvinner med transformationer.

Den binomiala ansatsen är inte heller lämplig för graderingsdata. Man kan inte förvänta sig att metoden ska konvergera när något led eller block genomgående graderats till 100. I den undersökta datamängden är det vanligt att effekten är maximal på ett stort antal rutor.

Friedmans test fungerar dåligt när antalet block är litet, vilket i fältförsök är regel snarare än undantag. För graderingsdata är Friedmans test särskilt olämpligt eftersom det är vanligt att flera observationer är lika. När man analyserar rangtal istället för ursprungliga data går också mycket av informationen i försöket förlorad.

Randomiseringstestet ger exakta sannolikhetsvärden. Nackdelen är att det krävs ett stort antal randomiseringar, och att det inte finns någon färdig procedur i SAS som utför testet. Det är emellertid möjligt att göra 10 000 randomiseringar på bara några sekunder, så det bör inte vara något större problem att införa metoden.

Tyvärr är det inte möjligt att med randomiseringstest få signifikanta resultat i parvisa jämförelser när antalet block är mindre än 6.

Det största problemet med graderingsdata, när det gäller statistisk analys, är att olika par av led har olika varians. Om vi emellertid är intresserade av ett specifikt par, t.ex. C jämfört med B, är det alltid möjligt att studera det paret separat och göra ett vanligt t-test under antagande om normalfördelning. Om varken led B eller led C är approximativt normalfördelade tenderar ändå fördelningen för skillnaden mellan medelvärdet av led C och medelvärdet av led B vara mer lik normalfördelningen.

6. Exempel

En traditionell variansanalys av materialet i tabell 1 i avsnitt 2 ger sannolikhetsvärdet 0,010 för testet av hypotesen att det inte finns några skillnader mellan leden. Minsta signifikanta skillnad (LSD) är 5,5. I tabell 4 redovisas medelvärdena, och med bokstäver visas vilka skillnader som är mindre än LSD.

Tabell 4. Medelvärden per led. Led med samma bokstav skiljer mindre än LSD.

Led	Medel	
B	97,5	ab
C	97,5	ab
D	92,5	bc
E	100	a
F	90	c

Ett randomiseringstest med 10 000 randomiseringar ger sannolikhetsvärdet 0,002, vilket indikerar tydliga effekter av behandlingarna. För att ta reda på vilka behandlingar som har signifikant olika effekt gjordes parvisa t-test. Resultatet redovisas med bokstäver i tabell 5.

Tabell 5. Medelvärden per led. Led med samma bokstav är inte signifikant olika med parvisa t-test.

Led	Medel	
B	97,5	ab
C	97,5	ab
D	92,5	ab
E	100	a
F	90	b

7. Slutsatser

Graderingsdata är inte normalfördelade med homogen varians, varken före eller efter transformation. Därför kan vanlig variansanalys på ursprungliga data (metod 1) eller på transformerade data (metod 2) ge missvisande resultat.

Varken binomial analys (metod 3) eller Friedmans test fungerar tillfredsställande.

Med randomiseringstest (metod 4) är det möjligt att beräkna exakta sannolikhetsvärden.

Parvisa t-test, baserade på antagandet om normalfördelning, kan ge ledning om vilka led som har olika effekt. Vid parvisa t-test jämförs nämligen ett par i taget, och då är det inget problem att variansen inte är lika för alla par.

Referenser

Edgington, E. S. (1987). *Randomization tests*. 2:a uppl. Marcel Dekker.

Hollander, M. och Wolfe, D. A. (1999). *Nonparametric statistical methods*. 2:a uppl. Wiley.

Piegorsch, W. W, och Bailer, A. J. (2005). *Analyzing environmental data*. Wiley.