

3.6 Generella statistiska samband och en modell med för sockerskörden begränsande variabler

Hans Larsson, SLU och Olof Hellgren, SLU

Inledning

En uppgift för projektet var att identifiera ett antal påverkbara variabler som begränsar sockerskörden. För att vara säkra på att både fält med hög skörd och medelskörd skulle ingå i undersökningen valdes 14 pargårdar ut där den ena gården haft hög skörd de senaste 5 åren (plusgård), medan den andra haft lägre skörd (medelgård). I det här avsnittet koncentrerar vi oss på att förklara sockerskörden oberoende av om fältet legat på en s.k. plusgård eller medelgård.

Statistiska analysmetoder

Huvudsakligen två statistiska metoder har använts för att analysera och strukturera datamaterialet. Dels har diskriminantanalys använts, dels multipel linjär regression. Dessa två metoder skiljer sig på en central punkt.

Diskriminantanalysen diskriminerar mellan grupper, dvs ger utslag för skillnader mellan angivna grupper. Denna används för att analysera om datamaterialet innehåller variabler vars data skiljer ut grupperna. Analysen anger också hur mycket av skillnaderna som kan förklaras av variablernas värden, dvs hur avgörande för skillnader mellan grupperna dessa data är. Diskriminantanalysen kan inte leda fram till en kontinuerlig, kvantitativ modell. Den kan ange ett antal variabler som sannolikt är avgörande för skillnaden mellan grupper.

Den multipla linjära regressionsanalysen är en typ av regressionsanalys där sambandet mellan ett antal oberoende variablers värden och en beroende variabel testas. Det som testas är styrkan i de oberoende variablernas förmåga att förklara den beroende variabelns värden samt ange sambandet i en linjär ekvation där varje variabel är en komponent som bidrar till ett värde för den beroende variabeln. Denna analys kan användas för att konstruera en kvantitativ modell över ett angivet giltighetsområde för komponenternas respektive värden.

Generell modell för de 14 pargårdarna 1998-2000

Den generella modellen är beräknad på medelvärden för de 14 pargårdarna under de tre åren 1998-2000 med hjälp av linjär multipel regression. Modellen är vald efter principen att med så få variabler som möjligt kunna ge en så hög förklaringsgrad för sockerskörd som möjligt. Ju fler variabler som tas med ju högre förklaringsgrad erhålls visserligen men samtidigt ökar också osäkerheten i modellen och risken är att generaliteten minskar. Den statistiska analysen baserad på fyra variabler, med skörd av utvinnbart socker i ton per ha (**UEKSH**) som funktion av variablerna, gav följande resultat:

$$\text{skörd} = 2.9 + 1.5 \text{ vertikal infiltration} + 1.12 \text{ pH i matjorden} - 0.07 \text{ sådatum} - 0.4 \text{ medelsvamp}$$

Korrelationskoefficienten i kvadrat, r^2 , för de fyra variablerna är 0.85 vilket innebär att de tillsammans kan förklara 85 % av variationen i sockerskörd. För de variabler, som funktionen består av, gäller bestämda giltighetsområden, dvs endast angivna värden kan användas i modellen. Giltiga områden är för vertikal infiltration 0.1-0.9 (cm/h), för pH i matjorden 6.4-7.6, för sådatum 1-30 (dagar i april) och för medelsvamp 0.1-3.5 (%). Modellen är linjär och tar

alltså inte hänsyn till asymptotiska värden, dvs värden där en ökning eller minskning inte längre ger något resultat. För att förklara vad som ligger bakom de fyra variablerna har korrelationer och regressioner genomförts på fältmedelvärden för de tre åren.

Sådatum (**sådat**) i antal dagar efter 1 april betyder att ju fler dagar senare man sår ju mer reduceras möjligheterna till högre skörd ($-0.07 \times$ antal dagar). Slutsatsen och erfarenheterna är att det räcker inte med ett tidigt sådatum. Ett tidigt sådatum måste också väljas efter lämplig väderlek.

Sådatum är positivt korrelerat med porositeten och negativt med skrymdensiteten vilket troligen innebär att de fält som det sås tidigt på har en tidigare upptorkning till följd av hög porositet. En stark korrelation fanns också mot friska plantor, dvs ju tidigare sådd desto friskare plantor. En porös jord skulle innebära att plantorna snabbt växer ifrån eventuella angrepp från skadegörare.

Den vertikala infiltrationen, **kvert**, betyder att dräneringen av regnvatten ska fungera bra, dvs ju bättre dränering ju högre tillägg till potentiell skörd kan göras ($+1.5 \times$ vertikal infiltration). Kvert var direkt korrelerad med dagmaskförekomsten och indirekt genom stark korrelation med kaliumhalten i matjord och i alv, vilken i sin tur var starkt korrelerad med dagmaskantalet. I linjär multipel regression med **kvert** som beroende variabel var också luftvolym och sand-grovmo i alven (**sgalv**) viktiga variabler, även om de enskilt bara var korrelerade med **kvert** på 10 %-nivå. Dessa båda var också starkt korrelerade inbördes. Fosfor i matjorden (**palmatj**) var också starkt korrelerat med **kvert**. Sammanfattningsvis hade jordar med hög bördighet och stor förekomst av dagmask hög vertikal infiltration.

Svamp som ett beräknat medelvärde av förekomst av svamppatogener (**medelsv**) betyder att ju högre svampförekomst desto större blir skördebortfallet ($-0.4 \times$ svampangrepp). Bakom svampangreppsbedömningen (**medelsv**) fanns starka korrelationer till variabeln skadebedömningen (**ds3**) och variabeln andelen friska plantor (**friska3**). Vid ökade förekomster av skadedjur, vilket gäller både *Onychiurus* och alla skadedjur tillsammans (**pestind**), ökade svampangreppet eller blev allvarligare. Sammanfattningsvis är det inte bara det procentuella svampangreppet som är direkta orsaken till skördereduktionen utan även bakomliggande skadedjursförekomster och kanske också förekomsten av svamp i fältet och dess påverkan på betplantornas tillväxt.

För sockerbeta är det väl känt att låga pH ger sämre skördar. I funktionen betyder ökat pH i matjorden (**pHmatj**) en ökning i skörd ($+1,12 \times$ pH i matjorden). pH var positivt korrelerat med porositeten (**por-**) och negativt med skrymdensiteten (**sd-**).

Svårigheten med att använda modellen är att få ett lätt tillgängligt mått på den vertikala infiltrationen och ett korrekt värde på medelsvamp.

I tabell 1 visas skillnaderna mellan gårdar med >10 ton utvinnbart socker och <10 ton för ett antal variabler.

Tabell 1. Medelvärden för fält med sockerskörd >10 ton/ha jämfört med fält med sockerskörd <10 ton/ha över åren 1998 till 2000

| | Sådd dagar in i april | Dag-grad2 | Plh 1000/ha | Friska3 % | Hjblad cm ² | Junimkt % | Slutmkt % | Sk3 ton/ha | Sk4 ton/ha | Betskörd ton/ha | Skörd ton/ha |
|------------|-----------------------|-----------|-------------|-----------|------------------------|-----------|-----------|------------|------------|-----------------|--------------|
| >10 ton/ha | 16 | 112 | 82 | 65 | 1.51 | 29 | 79 | 26 | 50 | 67 | 10.7 |
| <10 ton/ha | 24 | 115 | 80 | 51 | 1,45 | 21 | 68 | 18 | 41 | 55 | 8.7 |
| diff | -8 | -3 | 2 | 15 | 0.06 | 8 | 11 | 8 | 9 | 12 | 2.0 |

Fält med hög skörd såddes 8 dagar tidigare (**Sådd dagar in i april**), hade snabbare uppkomst (**Daggrad2**), fick ett högre plantantal (**Plh**), hade fler friska plantor (**Friska3**) med större hjärtblad (**Hjblad**) och högre junimarktäckning (**Junimkt**) samt slutmarktäckning (**Slutmkt**) än fält med skörd under 10 ton.

Skörden i juli (**Sk3**) var 44 % större och skörden i augusti (**Sk4**) var 22 % större på högskördefälten. Slutskörden i betor var också 22 % större liksom sockerskörden.

Linjär multipel regression och diskriminantanalys för enskilda år och över åren

Ordningen på variablerna är efter resultatet i den multipla regressionen. Detta innebär att den första variabeln är den viktigaste då den bidrar mest till förklaringen av resultatet (ex. skörde-resultatet). Variablerna kommer sedan i minskande betydelse. R²-värdet för variablerna finns angivet och anger hur stor procentuell andel som kan förklaras av variablerna.

Tabell 2. Linjär multipel regression och diskriminantanalys för enskilda år och över åren

| Fältnivå | junimkt | Sk3 | Sk4 | ueksh | Diskriminant-analys, ueksh |
|-----------|----------------------------------------|-----------------------------------------------|------------------------------------------------------|----------------------------------------------|----------------------------|
| 1998-2000 | Sådat Sgalv 0,39 | Friska3 Mullalv Luftvol Sgalv 0,58 | Betbagge Mullmat Kvert Medelsv 0,39 | Sådat Medelsv PHmatjord Betbagge 0,46 | Sådat medelsv |
| 1998 | Sådat Palmatj Kvert Rotbra 0,99 | Leralv Sdalv Betbagge Tusenfoting 0,94 | Friska3 Palmatj Tusenfoting pHmatj Sådat 0,93 | Friska3 Medelsv 0,66 | Friska3 Palalv Kvert |
| 1999 | Kalv Sådat Kalalv 0,92 | Betbagge Sdmatj Kvert 0,82 | Kalv Sgmatj pHalv 0,77 | Medelsvamp Kmatj 0,68 | Medelsv kvert |
| 2000 | Kalv 0,38 | Ej sign | Ej sign | Kvert 0,52 | kvert |

1998

Året var det år då plantorna hade flest skador orsakade av skadedjur och svamp vilket är tydligt i både skörd 3 (**sk3**) och 4 (**sk4**) där variablerna betbagge, tusenfoting, medelsvamp och friska plantor finns med. Även slutlig sockerskörd har bestämts av friska plantor och svampangrepp. För junimarktäckningen (**junimkt**) är sådatum den starkaste variabeln.

1999

Året hade inte lika stora skador av skadegörare som under 1998 men betbagge förklarar värdena för tidig skörd (**sk3**) och medelsvamp förklarar en del av sluskskörden. Vertikal infiltration finns som förklaring vid alla nivåer. Junimarktäckning förklaras av sådatum och vertikal infiltration i alven.

2000

Det år som har lägst förklaringsgrad. Ingen signifikant förklaring finns för skörd 3 och 4. Där emot förklarar kvert sluskskörden. Även junimarktäckningen förklaras av vertikal infiltration i alven.

1998-2000

Friska plantor var den viktigaste variabeln för förklaring av tidig tillväxt mätt vid skörd 3. Mull i alven (**mullalv**) visar på markens bördighet medan luftvolym (**luftvol**) och sand-grov-mohalt i alven (**sgalv**) ger god porositet. Vid skörd 4 var betbaggen den viktigaste variabeln. Även här kommer mullhalten med, denna gång i matjorden (**mullmatj**), och den vertikala infiltrationen. Som fjärde variabel finns svampangrepp med. Slutlig skörd förklaras i stort av samma variabler som finns i modellen vilken är uträknad på medelvärden. Som fjärde variabel finns här betbagge istället för kvert, men betbagge kan ersättas av kvert vilket endast ger en marginell sänkning av förklaringsgraden jämfört med betbagge.

Diskriminantanalysen bekräftar i stort de resultat som erhållits med linjär multipel regression.

Diskriminantanalys

1998-2000, fält

Materialet delades in efter storlek på sockerskörden. Fält med mer än eller lika med 10 ton socker/ha hamnade i en grupp. Dessa utgjorde 18 av totalt 42 fält eller 43 %.

Diskriminerande variabler var medelsv och sådat. Med dessa två variabler klassades 79 % av fälten i rätt klass med avseende på skörden.

Tabell 3.

| | Grupp 1 <10 ton socker/ha | Grupp 2 >10 ton socker/ha |
|----------------------------------|---------------------------|---------------------------|
| Diskriminerande variabler | | |
| medelsv | 2,1 | 0,9 |
| sådat | 24 | 16 |
| ueksh | 8,7 | 10,7 |

1998, fält

Materialet delades in efter storlek på sockerskörden. Fält med mer än eller lika med 9,2 ton socker/ha hamnade i en grupp. Dessa utgjorde 7 av totalt 14 fält eller 50 %.

Diskriminerande variabler var friska3, kvert och palalv. Med dessa tre variabler klassades 100% av fälten i rätt klass med avseende på skörden.

Tabell 4.

| | Grupp 1 <9,2 ton socker/ha | Grupp 2 >9,2 ton socker/ha |
|----------------------------------|--------------------------------------|--------------------------------------|
| Diskriminerande variabler | | |
| friska3 | 35 | 57 |
| palalv | 2,7 | 9,5 |
| kvert | 1,0 | 1,4 |
| ueksh | 8,1 | 9,7 |

1999, fält

Materialet delades in efter storlek på sockerskörden. Fält med mer än eller lika med 10 ton socker/ha hamnade i en grupp. Dessa utgjorde 9 av totalt 14 fält eller 64 %.

Diskriminerande variabler var medelsv och kvert. Med dessa två variabler klassades 86 % av fälten i rätt klass med avseende på skörden.

Tabell 5.

| | Grupp 1 <10 ton socker/ha | Grupp 2 >10 ton socker/ha |
|----------------------------------|-------------------------------------|-------------------------------------|
| Diskriminerande variabler | | |
| medelsv | 2,4 | 0,6 |
| kvert | 0,5 | 1,0 |
| ueksh | 9,0 | 10,6 |

2000, fält

Materialet delades in efter storlek på sockerskörden. Fält med mer än eller lika med 10 ton socker/ha hamnade i en grupp. Dessa utgjorde 7 av totalt 14 fält eller 50 %.

Diskriminerande variabel var kvert. Med denna variabel klassades 86 % av fälten i rätt klass med avseende på skörden.

Tabell 6.

| | Grupp 1 <10 ton socker/ha | Grupp 2 >10 ton socker/ha |
|----------------------------------|-------------------------------------|-------------------------------------|
| Diskriminerande variabler | | |
| kvert | 0,2 | 1,0 |
| ueksh | 8,4 | 11,0 |

Diskussion

Den generella modellen anger fyra påverkbara variabler vilka påverkat sockerskördarna på pargårdarna under tre år. De fyra ingående variablerna är alla enskilt korrelerade med sockerskörd. Det som är viktigt att betona är emellertid att **det är samspelet mellan dessa fyra som gör att modellen kan förklara variationer i sockerskörd.**

Sådatum är ju den komponent i modellen som anger det tillfälle då tillväxten sätter igång men viktigt att veta är att markfysikaliska faktorer som porositet kan påverkas genom lämplig plöjningspunkt på hösten. Med bättre porositet kan man så tidigare på våren. Sådatum är alltså en faktor som både är påverkbar och är starkt integrerad med andra faktorer.

pH i matjorden är en markkemisk faktor som begränsar tillgängligheten av näringsämnen men också gynnar svampsjukdomar om det är för lågt. Modellen beskriver den som en positiv faktor men den blir naturligtvis negativ om den blir så hög att den begränsar tillgängligheten av vissa ämnen. pH är alltså en komponent i modellen som är integrerad med bl.a. så avgörande faktorer som näringstillgänglighet och svampsjukdomar.

Den vertikala infiltrationen är en markfysikalisk faktor som begränsar transporten av vatten och näring men också rötternas markgenomträngningsförmåga. Den kan förbättras genom att dagmaskantalet ökas i jorden, men även av faktorer som minskar markpackning, ökar dränering och genom andra åtgärder som positivt kan påverka markens struktur.

Svampangrepp reducerar plantantalet på våren men reducerar också tillväxten under sommaren. Marksmitta kan minskas genom en korrekt växtföljd, och genom kalkning. Angreppet på plantan kan minskas genom betning. Tidig sådd är också ett sätt att minska påverkan av svampen. Svamp gynnas dessutom av låga pH och dålig porositet. Tillsammans med svampar påverkar skadedjur i jorden (en markbiologisk faktor) plantornas friskhet. Svampangrepp är nära korrelerat med andelen friska plantor men också med förekomsten av skadedjur i jorden. Skadedjuren kan underlätta för svampen att infektera plantorna.

Modellen visar på de här fyra variablernas betydelse och kan användas för att förbättra förutsättningarna för hög sockerskörd på den enskilda gården. Viktigt att påpeka är dock att dessa fyra variabler är en modell av flera som är möjliga att välja. Säkerheten i det datamaterial som ligger till grund för modellen är avgörande för dess generalitet och prediktionsstyrka. Modellen beskriver statistiska samband mellan valda oberoende variabler och den beroende, sockerskörd, men är inte avsedd att förklara sambandens inneboende mekanistiska struktur.