

Populärvetenskaplig sammanfattning

Varje sommar publicerar Jordbruksverket en prognos för hur stora skördarna av tretton viktiga grödor kommer att bli det året. Den låter både bönder och livsmedelsföretag planera sin verksamhet. Därför är det viktigt att prognoserna är så bra som möjligt. Hur stora skördarna blir per odlad hektar (den s.k. hektarskörden) varierar mellan grödor och beror bl.a. på jordmånen och vädret. Jordbruksverket använder sig av linjär regression för att modellera sambandet mellan hektarskörden och vädret under året. Jag visar hur denna modell kan förändras för att ge bättre prognoser. Som bäst har dessa modifierade modeller hälften så mycket fel som Jordbruksverkets.

Alla prognoser osäkra. När man ska tolka prognoser är det därför viktigt att veta hur osäkra de är. Jordbruksverkets modell ger oss ett (mer eller mindre riktigt) samband mellan årets väder och hektarskörden av en viss gröda i ett visst län. Detta låter oss simulera hur stora skördarna kommer vara om detta samband håller. Säg att vi gör 1000 sådana simulationer. Då kan vi avgöra inom vilket intervall som 95 % av de simulerade skördarna faller inom. Detta intervall är ett mått på osäkerheten inneboende i modellen.

Årets väder fångas av ett antal mått på temperatur och nederbörd. Utöver medeltemperatur och total nederbörd för växtsäsongens månader anges också sådant som antalet dagar med nederbörd under en månad och medeltemperaturen kl. 12:00 för månadens fem varmaste dagar. Ju fler mått desto mer information har modellen att utgå ifrån. Vissa mått är däremot starkt kopplade: regnar det många dagar under en månad kommer ofta den totala nederbörden vara hög, t.ex. När sådana kopplingar, s.k. kolinearitet, finns förvirrar det regressionsmodellen och den ger sämre prognoser. Därför visar det sig att skördeprognosen förbättras av att ta bort vissa vädermått och slå ihop andra till ex. medeltemperatur för hela våren. Detta är att ändra data som modellen utgår ifrån.

Kolinearitet mellan olika vädermått kan också åtgärdas genom att ändra hur själva regressionsmodellen fungerar. Vanlig linjär regression hittar det samband bland data som ger minst prognosfel. Är flera mått kolineära så är risken att detta samband inte gäller för kommande år, d.v.s. för data som modellen inte har sett tidigare. Istället väljer vi ett samband som tar så lite hänsyn till så få vädermått som möjligt, så länge som prognosfelet fortfarande blir lågt. Detta sätt att reducera modellen, som statistiker säger, kallas för regulariserad regression. I uppsatsen visar jag att modellens prognosfel kan halveras antingen genom att ändra data så att korrelationen mellan mått minskar, eller genom att använda regulariserad regression.

I uppsatsens sista del utforskar jag en väsentligt annorlunda sorts modell. Jordbruksverkets modell (och de andra modeller som jag har diskuterat) behandlar varje gröda i varje län var för sig. Tyvärr har vi för flera grödor i flera län inte statistik för så många år. Detta gör att många samband bygger på väldigt få data; dåligt underbyggda samband leder till dåliga prognoser. Jag lägger fram ett utkast till en modell som tittar på hur en gröda växer i flera län samtidigt. På så sätt bygger dess prognoser på många gånger fler data än Jordbruksverkets modell. Preliminära resultat visar att denna modell presterar bättre än någon annan som undersöks i uppsatsen. Detta är en lovande riktning för vidareutveckling av Jordbruksverkets skördeprognos.