

3.2.5 Statistisk analys

Hans Larsson, SLU

Materialet har databearbetats med programmet SPSS Base 10,0.

Deskriptiv statistik

Test av normalfördelning. Alla variablerna har de enskilda åren studerats grafiskt i histogram hur de fördelat sig kring medelvärdet. Om inte variablerna är approximativt normalfördelade kan de inte behandlas vidare med olika statistiska metoder. Även normalfördelningskurvor utan högsta och lägsta värde har konstruerats för att studera påverkan av outliers.

Medelvärde och standardavvikelse är användbara beskrivningar av data om de följer en normalfördelningskurva men kan vara vilseledande om de är snedfördelade, har outliers eller är en blandning av undergrupper.

Korrelationer, multipla regressioner och diskriminantanalys har körts på varje beroende variabel.

Linjära korrelationer

Pearsons korrelationskoefficient beskriver styrkan av ett linjärt förhållande mellan två kvantifierbara variabler. Normalfördelning förutsätts.

Spearman rank korrelationskoefficient rankar ordningen på data och förutsätter inte normalfördelning.

Linjära korrelationer beskriver sambanden mellan alla använda variabler. Signifikanser på 0.05 och 0.01-nivån markeras.

Multipel regression

Multipel regression har använts som en förklaringsmodell med flera oberoende variabler. För multipel regression används determinationskoefficienten som är kvadraten på korrelationskoefficienten R . R square varierar mellan 0 och 1 och 0,81 betyder att 81 % av variationen i den beroende variabeln förklaras av de oberoende variablerna som ingår i ekvationen.

Det viktiga med regressionsmodellerna är att begränsa antalet använda variabler. Ju färre variabler desto större chans för att modellen skall vara giltig även för nya fall.

Diskriminantanalys

Diskriminantanalys är besläktad med både multivariat analys och multipel regression. Man börjar med fall i två eller fler kända grupper och sedan används diskriminantproceduren för att identifiera en linjär kombination av variabler som bäst karakteriserar skillnaderna mellan grupperna.

Om man har två grupper som överlappar varandra i ett plotdiagram så används diskriminantanalysen för att bestämma en riktning för att projicera punkterna så att man maximerar skillnaderna mellan grupperna. Ekvationen som beskriver den nya projiceringen kallas Fisher's diskriminantfunktion.

Test av skillnaderna mellan gruppernas medelvärde, s.k. centroider, sker med Wilk's lambda - en multivariat analys som varierar mellan 0 och 1. Små värden indikerar att det är skillnader mellan grupperna. Wilk's lambda är den del av den totala variansen som inte förklaras av skillnaderna mellan grupperna. Är Wilk's lambda 0.25 så är det bara 25 % som inte förklaras av skillnaderna mellan grupperna.

Med fler än två grupper används canoniska variabler. Den första canoniska variabeln är en funktion av variabler som maximerar skillnaderna mellan medelvärdena i en dimension och den andra canoniska variabeln representerar den maximala spridningen av medelvärdena i en riktning som är vinkelrät mot den första. Ett plotdiagram med den första canoniska variabeln mot den andra ger en bra uppfattning om skillnaderna mellan grupperna.

Group statistics innehåller medelvärde och standardavvikelse för de variabler som ingår i modellen. Med denna tabell beskriver man lätt skillnader mellan grupperna.

Standardiserade canoniska diskriminantfunktionskoefficienter ger ett direkt mått på hur mycket varje variabel tillför diskriminantfunktionen, dvs hur viktiga de är.

Structure matrix ger ett mått på hur varje variabel är korrelerad med den canoniska variabeln.

Classification function koefficienter används för att räkna ut hur ett nytt fall skulle klassificeras. Koefficienterna multipliceras med varje variabel och konstanten läggs till för att få ett mätvärde för gruppen.

Casewise statistics visar hur du indelat materialet i kolumnen actual groups och hur programmet klassat grupperna i kolumnen predicted groups. Man kan också se second highest group, dvs vilken grupp som ligger närmast den första klassificeringen.

Diskriminantanalysen används för att testa multivariata skillnader mellan grupper och för att bestämma vilka variabler bland många som är bäst på att beskriva skillnader mellan grupper. Med hjälp av funktionerna kan sedan nya fall klassas.

Analysresultatet beskrivs bäst med antalet fall som klassas rätt och med så få variabler som möjligt. Med lika många variabler som fall klassas ofta alla fall rätt utan att det har någon signifikans.

Diskriminantanalys har använts för att beskriva skillnaderna mellan t ex hög och låg skörd men också för att beskriva skillnaderna mellan plus- och medelgårdar.