

Statistik 2, del 3:

I datafilen GeneticTraitsData.sav finns samma genetiska information som betraktades redan under principalkomponentanalysen, men därutöver finns det observationer på fyra kontinuerliga fenotyper (Trait1-4). Vi använder linjär regression skilt för varje fenotyp och sätter in de genetiska indikatorerna som förklarande variabler (prediktorer) och letar reda på vilka som antyds ha en koppling till fenotyperna.

Kör linjär regression med forward och backward sökningen och notera vilka variabler som blir signifikanta prediktorer. Residualanalys avslöjar att modellenpassningen är förhållandevis god. För varje fenotyp upptäcks det två eller fler prediktorer som är signifikanta på 5%-nivån. Däremot vet man (eftersom det fenotypiska datat är simulerat) att endast V21 har en riktig effekt på fenotyp Trait4, alla andra identifierade prediktorer med signifikant p-värde är falska positiva fynd. Därför bör korrektion för multipla test användas i dylika sammanhang, t ex se följande länkarna:

http://en.wikipedia.org/wiki/Bonferroni_correction

http://en.wikipedia.org/wiki/False_discovery_rate