Datorexamination

Statistisk analys av komplexa data, 7.5 hp

Moment 3 - Longitudinella data

Uppgiften löses individuellt. Det är inte tillåtet att arbeta tillsammans. Labbrapport med svar på nedanstående frågor samt relevanta **utskrifter** och **SAS-koder** (eller koder från annat program) skickas med e-mail till Linda Wänström (linda.wanstrom@liu.se) senast måndagen den 18:e december.

Datamaterialet (finns på hemsidan) som ska användas till uppgifterna består av barn (från det officiella amerikanska datamaterialet NLSY79) som är mätta under åren 1996, 1998, 2000 och 2002. Uppgifter finns om deras idnummer (ID), deras mammas idnummer (momid), etnicitet (race: 1=Hispanic, 2=Black, 3=non-Hispanic non-Black), kön (sex: 1=pojke, 2=flicka), födelseordning (birthorder), mammans ålder vid barnets födsel (agemom), födelseår (birthyear), ålder (a: i månader) för de fyra tidpunkterna, PIAT Math-poäng (m) för de fyra tidpunkterna, samt åldersstandardiserade PIAT Math-poäng (mz) för de fyra tidpunkterna (time).

Syftet, i kommande uppgifter, är att undersöka hur barnens utvecklingskurvor (i PIAT Math) ser ut. Syftet är även att undersöka om någon/några bakgrundsvariabler kan tänkas påverka barnens utvecklingskurvor.

Du får använda vilket datorprogram du vill. Bra information om SAS-kod finns i SAS-hjälpen om man söker på proceduren PROC MIXED respektive PROC CALIS. Bra information finns även i Singers artikel, Tan, Kang och Hogan’s artikel, kapitel 6 i Goldsteins bok, samt föreläsningsanteckningarna på hemsidan. Det kan vara lättare att använda PROC CALIS-kod från föreläsningsanteckningarna än från artikeln.

Det är inget krav att centrera några variabler i nedanstående uppgifter (men det kan underlätta vid tolkningar av skattade parametrar). Datamaterialet till uppgifterna är balanserat (dvs alla barn har mätningar under alla fyra tillfällen). Detta minskar riskerna med att man får problem med att skatta G-matrisen. Skulle du få problem med detta ändå kan det dock hjälpa att centrera oberoende variabler. Det kan även hjälpa att i stället för att använda PIAT Math, då använda åldersstandardiserat PIAT Math.

Obs! Under tiden du jobbar med labben kan det hända att du stöter på skattningsproblem som kanske leder till att du till slut bestämmer dig för andra variabler/andra modeller än vad du började med. Detta händer ofta i verkligheten. Var noga med att presentera alla steg du genomfört och vilka resultat de gett, dvs redovisa inte enbart dina slutgiltiga svar på respektive fråga.

1. I datamaterialet motsvarar varje rad ett barn. Transformera data så att varje rad motsvarar en tidpunkt. Printa de första 20 raderna och bifoga dessa. (1p)

(Bifoga SAS-koden du använt i anslutning till ditt svar)

1. Vad kan du säga om barnens utvecklingskurvor? Hur ser formen på utvecklingskurvorna ut? Varierar utvecklingskurvorna mellan barnen? För att svara på dessa frågor ska du skatta en multilevel-modell (random coefficient) där responsvariabeln är **Piat Math.** Välj som oberoende variabel **Time** eller **Ålder** och eventuellt någon funktion av variabeln du valt (motivera ditt val av oberoende variabel(ler). Antag ingen struktur (förutom konstant varians) i R-matrisen. (3p)

(Bifoga SAS-koden du använt samt utskrift du tolkar i anslutning till ditt svar)

1. a) Utgå från modellen i uppgift 2. Undersök om det finns någon/några variabler som har någon effekt på barnens utvecklingskurvor. Vad blir din slutsats? (2p)

(Bifoga SAS-koden du använt samt utskrift du tolkar i anskutning till ditt svar)

1. Undersök om du behöver ha någon kovariansstruktur i R-matrisen. Prova flera olika, lämpliga, kovariansstrukturer. Vilken struktur tycker du passar bäst? (Hjälp: sök i SAS-hjälpen på proc mixed efter flera lämpliga strukturer.) Vad blir din slutgiltiga modell? (Tips: Om du får problem med att skatta modellen – varning i log – kan du prova att skatta en modell med en enklare kovariansstruktur. Se även tipsen som gavs ovan i texten innan uppgifterna.) (5p)

(Bifoga SAS-koden du använt samt utskrift du tolkar i anslutning till ditt svar)

1. Skatta en ”latent growth curve model” (exempelvis i proc calis i SAS) där ett latent intercept och en linjär latent lutning skattas med hjälp av de fyra mätningarna på PIAT Math. Tolka utskriften. Hur blir anpassningen av modellen? Vad blir skattningarna för parametrarna (tolka dessa även om anpassningen är dålig, men poängtera i så fall det i svaren.). (3p)

(Bifoga SAS-koden du använt samt utskrift du tolkar i anslutning till ditt svar)

1. Lägg till en kvadratisk faktor till modellen i uppgift 5. Skatta modellen och tolka utskriften. Hur blir anpassningen av modellen? Vad blir skattningarna för parametrarna (tolka dessa även om anpassningen är dålig, men poängtera i så fall det i svaren.). Är den kvadratiska modellen (i denna uppgift) bättre än den linjära modellen (uppgift 5) för datamaterialet? Motivera. Vilken slutsats drar du angående barnens utvecklingskurvor? (3p)

(Bifoga SAS-koden du använt samt utskrift du tolkar i anslutning till ditt svar)

1. Kan du ändra modellen i uppgift 6 på något sätt för att förbättra modellens anpassning? Beskriv hur du har gått till väga samt vilka resultat du har fått. (Tips: sök i SAS-hälpen/på nätet/i litteratur för att hitta något/några sätt att förbättra modellen.) (2p)

(Bifoga SAS-koden du använt samt utskrift du tolkar i anslutning till ditt svar)

1. I denna datorexamination har du skattat ungefär samma modeller med två olika metoder. Spelar det någon roll vilken av de två metoderna man använder för att skatta modellerna i denna examination? Motivera. (1p)