Datorexamination

Statistisk analys av komplexa data, 7.5 hp

Moment 1 - Hierarkiska 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 onsdag 22/11.

Datamaterialet (finns på hemsidan) som ska användas till uppgifterna består av 25 st sjukhus (HOSPITAL). Varje sjukhus har flera avdelningar (WARDID). Avdelningarna har randomiserats till experimentgrupp eller kontrollgrupp. I experimentgruppen har alla sjuksköterskor på avdelningarna genomgått ett program i stresshantering (EXPCON = 1). I kontrollgruppen har de inte genomgått något sådant program (EXPCON=0). Deras stressnivåer har därefter uppmätts (STRESS). Vi har uppgifter om sjuksköterskornas ålder i år (AGE), deras kön (GENDER: 0 = male, 1 = female), och deras erfarenhet i år (EXPERIEN). Vi har också uppgifter om avdelningstyp (WARDTYPE; 0 = general care, 1 = special care). Slutligen har vi uppgifter om sjukhusets storlek (HOSPSIZE: 0 = litet, 1 = medel, 2 = stort).

Syftet, i kommande uppgifter, är att undersöka om programmet i stresshantering hade någon effekt, dvs om de som genomgick detta program har lägre stressnivåer än de som inte gjorde det. Samtidigt vill vi ta hänsyn till den hierarkiska strukturen i data, dvs att vi har sjukhus som består av avdelningar som i sin tur består av sjuksköterskor. Vi vill också kontrollera för olika kontrollvariabler.

Du får använda vilket datorprogram du vill. Bra information om SAS-kod finns i SAS-hjälpen om man söker på proceduren PROC MIXED. Bra information finns även i Singers artikel. Statistiska teorin hittas i kapitel 1 och 2 i Goldsteins bok på hemsidan.

Vi kommer endast ta hänsyn till två nivåer av data. Eftersom alla sjuksköterskor på samma avdelning genomgick samma program (stresshantering eller ej) så kommer vi att titta på sjukhus och sjuksköterskor (och strunta i avdelningar).

Börja med att centrera oberoende variabler (Grand mean centering). Detta gör det lättare att tolka interceptet i modellerna och kan dessutom minska problem vid skattningar av slumpmässiga effekter. Skapa även två dummyvariabler av Hospsize. Obs! Du behöver inte centrera dummyvariabler. Det är inte heller nödvändigt att centrera oberoende variabler som bara antar två värden. Men tänk på vad detta får för konsekvenser på tolkningen av parametrar.

1. a) Testa, med ett formellt test, om det finns någon variation i stressnivåer (STRESS) mellan sjukhus. Utifrån ditt resultat, bör du använda multilevelmodeller för att analysera data, eller går det bra att använda ”vanliga” regressionsmodeller? (2p)

b) Hur stor andel av variationen i stressnivå (STRESS) finns mellan sjukhus? (1p)

(Klistra in koden du använt samt utskrift du tolkar här direkt efter uppgiften)

1. a) Skriv upp ekvationen/rna för en varianskomponentmodell där STRESS är responsvariabeln och de oberoende variablerna är EXPCON, AGE, GENDER, EXPERIEN och HOSPSIZE. Var noga med att definiera vad olika subskript står för. Skriv även upp modellens antaganden. (1p)

b) Skatta modellen i a) ovan. Utför hypotestest för de fixa parametrarna. Tolka skattningarna för de parametrar som är signifikant skilda från noll. Tolka även skattningarna för kovariansparametrarna. (5p)

(Klistra in koden du använt samt utskrift du tolkar här direkt efter uppgiften)

1. a) Skriv upp ekvationen/rna för en random coefficient modell där STRESS är responsvariabeln och de oberoende variablerna är EXPCON, AGE, GENDER, EXPERIEN och HOSPSIZE. Endast interceptet och lutningen för EXPCON ska vara slumpmässiga (utöver att de är fixa). Skriv även upp modellens antaganden. (1p)

b) Skatta modellen i a) ovan. Testa nollhypotesen att variansen i lutningen för EXPCON och kovariansen mellan lutningen och interceptet är noll för modellen i a) ovan med ett test. Tolka skattningarna om minst en av dem är skild från noll enligt testet. Utför också hypotestest för de fixa parametrarna. Tolka skattningarna för de fixa parametrar som är signifikant skilda från noll. (5p)

(Klistra in koden du använt samt utskrift du tolkar här direkt efter uppgiften)

1. Utifrån dina analyser i uppgift 1) – 3), vad är dina slutsatser om huruvida programmet i stresshantering har haft effekt på sjuksköterskornas stressnivåer? Är en eventuell effekt olika på olika sjukhus? Finns det i så fall någon variabel som kan förklara varför effekten skiljer sig åt mellan sjukhus? Motivera utifrån de resultat du har fått i uppgifterna ovan. (3p)
2. I uppgifterna ovan centrerade du de oberoende variablerna genom att dra bort det totala medelvärdet från varje observation. Beskriv ett annat sätt du kunde ha använt. Hade det funnits några fördelar/nackdelar med denna typ av centrering för detta datamaterial? Hade det blivit några skillnader i hur olika parametrar tolkas? (Du behöver inte genomföra analyser med denna centrering) (2p)