



Hemtentamen 1 - Hierarkisk data

732G34 Statistiska metoder för komplex data

Hampus Beijer

7 november 2025

1 Uppgift 1

Gör i ordning dina oberoende variabler som ska vara med i analyserna (enligt frågeställningarna ovan). Skapa dummyvariabel/variabler om det är lämpligt samt motivera varför det görs/inte görs. Centrera, på lämpligt sätt, oberoende variabler. (Observera att det inte är nödvändigt/lämpligt att centrera alla oberoende variabler.) För varje oberoende variabel – motivera att du centrerar den eller inte, samt typ av centring. Förlara vilka typer av centring man kan använda här. Referera till Singers artikel i ditt svar. Klistra in SAS-kod samt de första 20 observationerna vid ditt svar.

1.1 Datamaterial

Detta datamaterial består av 25 sjukhus, där varje sjukhus består av flera avdelningar (WARDID). Enligt instruktionen har avdelningar randomiseras till en experimentgrupp eller en kontrollgrupp. Experimentgruppen identifieras som *EXPCON* = 1 och kontrollgruppen som *EXPCON* = 0. I experimentgruppen har flera sjuksköterskor deltagit i ett program för att undersöka stress, där deras stressnivåer har mätts. Följande tabell visar variablerna och dess skala:

- **HOSPITAL** – Sjukhusnummer. *Kategorisk (nivå 2)*.
- **WARDID** – Avdelningsnummer. *Kategorisk (ej använd i analysen)*.
- **EXPCON** – Gruppindelning: 1 = experimentgrupp (stresshantering), 0 = kontrollgrupp. *Kategorisk (nivå 2)*.
- **STRESS** – Sjuksköterskans stressnivå. *Kontinuerlig (nivå 1)*.
- **AGE** – Sjuksköterskans ålder (år). *Kontinuerlig (nivå 1)*.
- **GENDER** – Kön: 0 = man, 1 = kvinna. *Kategorisk (nivå 1)*.
- **EXPERIEN** – Erfarenhet i år som sjuksköterska. *Kontinuerlig (nivå 1)*.
- **WARDTYPE** – Typ av avdelning: 0 = general care, 1 = special care. *Kategorisk (nivå 1)*.
- **HOSPSIZE** – Sjukhusets storlek: 0 = litet, 1 = medel, 2 = stort. *Kategorisk (nivå 2)*.

Variablernas skala kan ses ovan. Jag beskriver de dock kort i text. Följande variabler är kategoriska *hospital*, *ward*, *wardid*, *gender*, *wardtype*, *hospsize*, och *expcon*. Följande variabler är kontinuerliga *experien* och *age*. Dock är *stress* på ordinalskala, men i denna analys hanteras den som en kontinuerlig variabel.

1.2 Tabell över datamaterial

| hospital | ward | wardid | nurse | age | gender | experien | stress | wardtype | hospsize | expcon |
|----------|------|--------|-------|-----|--------|----------|--------|----------|----------|--------|
| 1 | 1 | 11 | 1 | 36 | 0 | 11 | 7 | 0 | 2 | 1 |
| 1 | 1 | 11 | 2 | 45 | 0 | 20 | 7 | 0 | 2 | 1 |
| 1 | 1 | 11 | 3 | 32 | 0 | 7 | 7 | 0 | 2 | 1 |
| 1 | 1 | 11 | 4 | 57 | 1 | 25 | 6 | 0 | 2 | 1 |
| 1 | 1 | 11 | 5 | 46 | 1 | 22 | 6 | 0 | 2 | 1 |
| 1 | 1 | 11 | 6 | 60 | 1 | 22 | 6 | 0 | 2 | 1 |
| 1 | 1 | 11 | 7 | 23 | 1 | 13 | 6 | 0 | 2 | 1 |
| 1 | 1 | 11 | 8 | 32 | 1 | 13 | 7 | 0 | 2 | 1 |
| 1 | 1 | 11 | 9 | 60 | 0 | 17 | 7 | 0 | 2 | 1 |
| 1 | 2 | 12 | 10 | 45 | 0 | 21 | 6 | 1 | 2 | 1 |
| 1 | 2 | 12 | 11 | 57 | 1 | 24 | 6 | 1 | 2 | 1 |
| 1 | 2 | 12 | 12 | 47 | 0 | 24 | 6 | 1 | 2 | 1 |
| 1 | 2 | 12 | 13 | 32 | 1 | 14 | 6 | 1 | 2 | 1 |
| 1 | 2 | 12 | 14 | 42 | 1 | 13 | 6 | 1 | 2 | 1 |
| 1 | 2 | 12 | 15 | 42 | 1 | 17 | 6 | 1 | 2 | 1 |
| 1 | 2 | 12 | 16 | 53 | 1 | 20 | 5 | 1 | 2 | 1 |
| 1 | 2 | 12 | 17 | 60 | 0 | 28 | 5 | 1 | 2 | 1 |
| 1 | 2 | 12 | 18 | 33 | 1 | 21 | 6 | 1 | 2 | 1 |
| 1 | 3 | 13 | 19 | 64 | 1 | 22 | 5 | 0 | 2 | 0 |
| 1 | 3 | 13 | 20 | 37 | 0 | 17 | 5 | 0 | 2 | 0 |

1.3 Behandling av datamaterial i SAS

1.3.1 Dummyvariabler

I denna hemtentamen skapades inga dummyvariabler manuellt eftersom funktionen *class* i SAS kan hantera kategoriska variabler. Det vill säga, SAS skapar dummyvariabler automatiskt när funktionen *class* används.

1.3.2 Centrering av variabler

De kategoriska variablerna centreras ej på grund av de är kategoriska. De kontinuerliga variablerna, men responsvariabeln görs ej. Anledning nämns strax!

Enligt Singer (1998) kan variabler centreras på två olika sätt: grand-mean centering eller group-mean centering. Singer förklarar att vid grand-mean centering subtraheras hela stickprovets medelvärde från

variabeln, vilket innebär att interceptet representerar utfallet för en individ med ett genomsnittligt värde i hela populationen. Vid group-mean centering subtraheras i stället respektive gruppens medelvärde, vilket medför att interceptet beskriver gruppens genomsnitt och att lutningen tolkas som effekten inom grupper.

Grand-mean centering: $X_{ij} - \bar{X}$

Group-mean centering: $X_{ij} - \bar{X}_j$

I ekvationerna står i för observationen och j för gruppen. Med det sagt är skillnaden mellan dessa sätt att centrera att grand-mean centering utgår från hela stickprovets medelvärde, medan group-mean centering baseras på medelvärdet inom varje grupp (Singer, 1998, s. 334).

I denna hemtentamen centreras inte variabeln *stress* på grund av det är responsvariabeln. De kontinuerliga variablerna (*experien* och *age*) centreras kring hela stickprovets medelvärde, det vill säga grand-mean centering. Detta innebär att varje individs värde jämförs med genomsnittet i hela datamaterialet. Motivering till val av centrering är att man vill undersöka skillnader mellan sjukhus i stress (enligt uppgiftens instruktioner). Resterande varibler är kategoriska och centreras därför inte. Anledningen till detta är att man förstör variabelns kategoriska ordning.

1.4 Bifogad SAS kod för uppgift 1

1.4.1 Inladdning av datamaterial

```

1  data hospital;
2    infile "X:\hospitals.txt"
3      dlm='09'x
4      firstobs=2
5      dsd;
6
7    input hospital ward wardid nurse age gender experien stress wardtype hospsize expcon;
8  run;
9
10

```

1.4.2 Tabell över datamaterial

```

1  proc print data=hospital(obs=20) label noobs;
2    title "Observationerna i datamaterialet";
3  run;

```

1.4.3 Centrering av variabler

```

1  /*Nedan säger jag till SAS att använda proceure means, utan att prnta utskfiten.
2  Där den ska plocka ut variablene age och experien
3  dessa ska sparas i tre variabler som används efter denna procedure;
4  proc means data=hospital noprint;
5    var age experien;
6    output out=medelvarden mean=medel_age medel_experien;
7  run;

```

```

8
9 *Nu sker själva centreringen. där jag säger till SAS att ta minus medelvärdet av variabeln
10 för Age och Experien och lagra allt i en ny variabel CHospital;
11 data CHospital;
12   if _n_ = 1 then set medelvarden;
13   set hospital;
14   CAge = age - medel_age;
15   CExperien = experien - medel_experien;
16 run;
17
18 *Testar att printa för att se hur det ser ut;
19 proc print data=CHospital;
20

```

2 Uppgift 2

2.1 Uppgift a

Beräkna intraklasskorrelationen samt tolka den. Klistra in relevant SAS-utskrift och SAS-kod vid ditt svar

Intraklasskorrelation är ett mått som visar hur stor del av variansen som beror på gruppennivå. Rent tolkningsmässigt kan man säga att om det visar sig att korrelationen är noll finns ingen klusterstruktur i datamaterialet vilket innebär att man kan använda vanlig regressionsmodell (OLS) men om korrelationen är mer än 0 kan det tolkas som att det finns en klusterstruktur i datamaterialet vilket innebär att en hierarkisk modell är lämplig. Det bör noteras att intraklasskorrelationen kan aldrig bli negativ eftersom att en varians är mer eller lika med noll. Se formel nedan:

$$\rho = \frac{\sigma_{u0}^2}{\sigma_{u0}^2 + \sigma_e^2}$$

I formeln ovan är:

- ρ är intraklasskorrelationen
- σ_{u0}^2 är mellan gruppvariancen. Detta är även känt som interceptet för modellen.
- σ_e^2 är inom gruppvariancen. Denna kallas även för residualen.

Det vill säga, σ_{u0}^2 beskriver i denna hemtentamen hur mycket gruppernas genomsnitt skiljer sig åt från varandra. σ_e^2 beskriver hur mycket individerna inom en grupp skiljer sig från varandra.

För att testa detta skapas en noll modell i SAS, vilket innebär en modell med endast ett intercept och inga förklarande variabler. Följande reslutat fås:

The Mixed Procedure

Covariance Parameter Estimates

| Cov Parm | Subject | Estimate |
|-----------|----------|----------|
| Intercept | hospital | 0.2844 |
| Residual | | 0.6870 |

Fit Statistics

| | |
|--------------------------|--------|
| -2 Res Log Likelihood | 2533.0 |
| AIC (Smaller is Better) | 2537.0 |
| AICC (Smaller is Better) | 2537.0 |
| BIC (Smaller is Better) | 2539.4 |

Solution for Fixed Effects

| Effect | Estimate | Standard Error | DF | t Value | Pr > t |
|-----------|----------|----------------|----|---------|---------|
| Intercept | 4.9969 | 0.1099 | 24 | 45.48 | <.0001 |

Intraklasskorrelationen behöver beräknas manuellt, vilket blir:

$$\rho = \frac{0.2844}{0.2844 + 0.6870} = 0.293$$

Intraklasskorrelationen är 29% vilket innehåller att 29% av variationen i stress beror på skillnader mellan sjukhus. För att beräkna skillnaden inom samma sjukhus blir detta $1 - 0.29 = 0.71\%$. Det innehåller att 71% beror på skillnader inom samma sjukhus medan 29% beror på skillnader mellan dem. Med det sagt verkar det finnas en klusterstruktur i datamaterialet där individers stress liknar varandra inom samma sjukhus, och av denna anledning kan en hierarkisk modell med ett slumprässigt intercept användas.

2.2 Uppgift b

Testa, med ett likelihood ratio-test, om det finns någon variation i stressnivåer (STRESS) mellan sjukhus. Klista in relevant SAS-utskrift och SAS-kod vid ditt svar.

I denna del vill man undersöka om det finns en variation i stressnivåer mellan sjukhusen. Detta innehåller att man vill undersöka om man bör ha ett slumprässigt intercept. För Likelihood Ratio testet prövas nu två modeller där första modellen har ett slumprässigt intercept och andra modellen har ett fixed intercept. I praktiken innehåller detta att modell 1 antar att sjukhus har olika medelvärden, medan modell 2 antar att alla sjukhus har samma medelvärde.

Modell 1

| The Mixed Procedure | | | | | |
|---------------------------------|--------|---------------------|----|----|----|
| Model Information | | | | | |
| Data Set | | WORK.CHOSPITAL | | | |
| Dependent Variable | | stress | | | |
| Covariance Structure | | Variance Components | | | |
| Subject Effect | | hospital | | | |
| Estimation Method | | REML | | | |
| Residual Variance Method | | Profile | | | |
| Fixed Effects SE Method | | Model-Based | | | |
| Degrees of Freedom Method | | Containment | | | |
| Class Level Information | | | | | |
| Class | Levels | Values | | | |
| hospital | 25 | 1 | 2 | 3 | 4 |
| 5 | 6 | 7 | 8 | 9 | 10 |
| 11 | 12 | 13 | 14 | 15 | 16 |
| 17 | 18 | 19 | 20 | 21 | 22 |
| 23 | 24 | 25 | | | |
| Dimensions | | | | | |
| Covariance Parameters | | 2 | | | |
| Columns in X | | 1 | | | |
| Columns in Z per Subject | | 1 | | | |
| Subjects | | 25 | | | |
| Max Obs per Subject | | 52 | | | |
| Number of Observations | | | | | |
| Number of Observations Read | | 999 | | | |
| Number of Observations Used | | 999 | | | |
| Number of Observations Not Used | | 0 | | | |

| Iteration History | | | | | |
|--------------------------------|-------------|-----------------|------------|---------|---------|
| Iteration | Evaluations | -2 Res Log Like | Criterion | | |
| 0 | 1 | 2797.68354813 | | | |
| 1 | 2 | 2532.97051401 | 0.00000001 | | |
| Convergence criteria met. | | | | | |
| Covariance Parameter Estimates | | | | | |
| Cov Parm | Subject | Estimate | | | |
| Intercept | hospital | 0.2844 | | | |
| Residual | | 0.6870 | | | |
| Fit Statistics | | | | | |
| -2 Res Log Likelihood | | 2533.0 | | | |
| AIC (Smaller is Better) | | 2537.0 | | | |
| AICC (Smaller is Better) | | 2537.0 | | | |
| BIC (Smaller is Better) | | 2539.4 | | | |
| Solution for Fixed Effects | | | | | |
| Effect | Estimate | Standard Error | DF | t Value | Pr > t |
| Intercept | 4.9969 | 0.1099 | 24 | 45.48 | <.0001 |

Man kan notera att modellen har konvergerat vilket är bra. Modellen innehåller endast ett intercept vilket är enligt t-teset signifkanta. Notera även att funktionen *class* fungerade, där SAS lyckades automatiskt skapa en dummyvariabel. Modellens har skattats med hjälp av REML vilket står för Restricted Maximum Likelihood, vilket i detta fallet mer varianskomponenter ger väntevärdesriktiga skattningar. Om REML inte används blir skattningarna mycket små och ej väntesvärdesriktiga. I modellens utskrift kan ses -2 Res Log Likelihood vilket är 2533.0. Detta behöver användas strax. Först ska modell 2 skattas.

Modell 2

| The Mixed Procedure | | | | | | | | | |
|----------------------------------------|-----------------|----------------|-----|---------|---------|--|--|--|--|
| Model Information | | | | | | | | | |
| Data Set | | WORK.CHOSPITAL | | | | | | | |
| Dependent Variable | | stress | | | | | | | |
| Covariance Structure | | Diagonal | | | | | | | |
| Estimation Method | | REML | | | | | | | |
| Residual Variance Method | | Profile | | | | | | | |
| Fixed Effects SE Method | | Model-Based | | | | | | | |
| Degrees of Freedom Method | | Residual | | | | | | | |
| Class Level Information | | | | | | | | | |
| Class | Levels | Values | | | | | | | |
| hospital | 25 | 1 | 2 | 3 | 4 | | | | |
| | | 5 | 6 | 7 | 8 | | | | |
| | | 9 | 10 | 11 | 12 | | | | |
| | | 13 | 14 | 15 | 16 | | | | |
| | | 17 | 18 | 19 | 20 | | | | |
| | | 21 | 22 | 23 | 24 | | | | |
| | | 25 | | | | | | | |
| Dimensions | | | | | | | | | |
| Covariance Parameters | | 1 | | | | | | | |
| Columns in X | | 1 | | | | | | | |
| Columns in Z | | 0 | | | | | | | |
| Subjects | | 1 | | | | | | | |
| Max Obs per Subject | | 999 | | | | | | | |
| Number of Observations | | | | | | | | | |
| Number of Observations Read | | 999 | | | | | | | |
| Number of Observations Used | | 999 | | | | | | | |
| Number of Observations Not Used | | 0 | | | | | | | |
| Covariance Parameter Estimates | | | | | | | | | |
| Cov Parm | Estimate | | | | | | | | |
| Residual | 0.9593 | | | | | | | | |
| Fit Statistics | | | | | | | | | |
| -2 Res Log Likelihood | | 2797.7 | | | | | | | |
| AIC (Smaller is Better) | | 2799.7 | | | | | | | |
| AICC (Smaller is Better) | | 2799.7 | | | | | | | |
| BIC (Smaller is Better) | | 2804.6 | | | | | | | |
| Solution for Fixed Effects | | | | | | | | | |
| Effect | Estimate | Standard Error | DF | t Value | Pr > t | | | | |
| Intercept | 4.9760 | 0.03099 | 998 | 160.57 | <.0001 | | | | |

För modell 2 kan det ses i utskriften igen att interceptet är signifkanta. -2 Res Log Likelihood”är 2797.7. Nu kan Likelihood Ratio testet beräknas. Statistikan formuleras enligt:

$$D_{01} = -2 \ln L(\lambda_0 / \lambda_1) \sim \chi_q^2$$

I statistikan är λ_0 och λ_1 likelihood under respektive alternativhypotes. Statistikan är chi2 fördelad med q frihetsgrader, där q står för skillnaden mellan antalet skattades parametrar för modellerna. Då fås följande statistika med frihetsgrader $q = 2 - 1 = 1$ (se utskriften i Covariance parameters):

$$D_{01} = 2797.7 - 2533.0 = 264.7 \sim \chi_1^2$$

Om man kollar upp chi2 tabellen ser man att med en signifikansnivå på 5% blir det kritiska värdet 3.84. Det vill säga man kan förkasta nollhypotesen på 5 procent signifikans. Att förkasta nollhypotesen i detta fall innebär att det finns variation i stress mellan sjukhusen ($H_0 : \sigma_{u0}^2 = 0$ och $H_0 : \sigma_{u0}^2 > 0$). Man kan dra slutsatsen att ett slumpmässigt intercept bör inkluderas i modellen.

2.3 Uppgift c

Utifrån resultaten i a) och b), bör du använda multilevelmodeller för att analysera data, eller går det bra att använda ”vanliga” regressionsmodeller?

I reslutatet a för del att kunde man notera att intraklasskorrelationen var större än 0 vilket innebor att det fanns variation i stress mellan de 25 sjukhusen, och inte bara mellan de i sjukhuset. Från uppgift b visades att en modell med ett slumpmässigt intercept var signifikant bättre än ett fixed. Anledningen till varför en vanlig linjärregression inte är lämplig i detta fallet är på grund av att datamaterialet visar psuedo replikor. Detta innebär att antagandet om oberoende i datamaterialet bryts. Om en vanlig linjär modell används kommer standardfel och p-värden att bli missvisande. Eftersom en hierarkisk modell tål beroende i datamaterialet är detta en lämpligare metod.

2.4 Bifogad SAS kod för uppgift 2

2.4.1 Uppgift 2a

```

1 *skattar en nollmodell för att genomföra intraklasskorrelationen;
2 proc mixed data=CHospital method=reml;
3   class hospital;
4   model stress = / solution;
5   random intercept / subject=hospital;
6 run;
```

2.4.2 Uppgift 2b

```

1 *Skattar två modeller för att genomföra ett LRT tes;
2 *REML används för väntevärdesriktiga skatningar;
3 *Class används för att automatiskt skapa dummyvariabler;
4 proc mixed data=CHospital method=reml;
5   class hospital;
6   model stress = / solution;
7   random intercept / subject=hospital type=vc;
8 run;
```

```

9
10 proc mixed data=CHospital method=reml;
11   class hospital;
12   model stress = / solution ;
13 run;
14

```

3 Uppgift 3

Skriv upp ekvationen/rna för en random coefficient-modell där STRESS är responsvariabeln och de beroende variablerna är EXPCON, AGE, GENDER, EXPERIEN och HOSPSIZE. Skriv upp den mest komplicerade formen av modellen, dvs den formen där alla beta-parametrar (intercept och lutningar) antas variera mellan sjukhus, samt där alla individnivå-variabler finns med i ekvationen på nivå 1, samt alla sjukhusvariabler finns med i alla ekvationer på nivå 2. Var noga med att definiera vad olika subskript står för. Skriv även upp modellens antaganden.

Innan respektive nivå formuleras behöver man förstå vad som menas med nivå 1 och 2. Man vill modellera variabeln stress vilket antas ha två nivåer. Detta kan man skriva som stress_{ij} där $i = \text{nivå 1}$ och $j = \text{nivå 2}$. I detta fallet är i individer och j sjukhusen. Med det sagt vill man börja med att formulera hur stress påverkas av individnivå-variablene ålder, kön, och erfarenhet. Modellen formuleras enligt:

$$\text{stress}_{ij} = \beta_{0j} + \beta_{1j} \cdot \text{age}_{ij} + \beta_{2j} \cdot \text{gender}_{ij} + \beta_{3j} \cdot \text{experien}_{ij} + \epsilon_{ij}$$

Där β_{0j} är interceptet vilket modellerar sjukhusets medelstress". Respektive $\beta_{1j}, \beta_{2j}, \beta_{3j}$ är lutningarna för respektive variabel, alltså hur stress påverkas av varje individ-variabel. ϵ_{ij} är individens residual som är $\sim N(0, \sigma_e^2)$. Man kan se ekvationen ovan som att när modellen skattas fås ett eget intercept för varje sjukhus. Detta tar oss vidare till beskrivningen av nivå 2.

Medan nivå 1 fokuserar på hur stress beror på ålder, kön, och erfarenhet så fokuserar nivå 2 på att förklara hur dessa nivåer skiljer sig åt. Det vill säga alla sjukhus kan ha olika intercept där man vet att medelstress mellan dessa intercept kan bero på stress och sjukhusets storlek! Vi formulerar respektive lutning enligt följande ekvation:

$$\beta_{0j} = \beta_{00} + \beta_{01} \cdot \text{expcon}_j + \beta_{02} \cdot \text{hospsize}_j + u_{0j}$$

där β_{00} är alla sjukhusens medelvärde av stress. β_{01} är hur stress skiljer sig beroende på sjukhuset j . β_{02} är hur stress skiljer sig beroende på storleken av sjukhuset j . u_{0j} är den slumpmässiga avvikelsen för sjukhuset j .

Men detta är endast för interceptet (β_{0j}) i ekationsformuleringen av nivå 1. Nivå 1 består av variablene *age*, *gender*, och *experien* vilket innebär att det finns tre till intercept. Dessa formuleras exakt som för β_{0j} :

$$\beta_{1j} + \beta_{11} \cdot \text{expon}_j + \beta_{12} \cdot \text{hospsize}_j + u_{1j}$$

$$\beta_{2j} + \beta_{21} \cdot \text{expon}_j + \beta_{22} \cdot \text{hospsize}_j + u_{2j}$$

$$\beta_{3j} + \beta_{31} \cdot \text{expon}_j + \beta_{32} \cdot \text{hospsize}_j + u_{3j}$$

Här antas respektive slumpmässiga avvikelse följa en normalfördelning med väntevärde 0, och kovariansmatris \mathbf{T} . Kovariansmatrisen betecknas endast för det tar mycket tid att skriva ut. Man förstår konceptet hoppas jag:

$$\begin{bmatrix} u_{0j} \\ u_{1j} \\ u_{2j} \\ u_{3j} \end{bmatrix} \sim N(0, \mathbf{T})$$

Nu kan nivå 1 och 2 skrivas ihop genom att använda nivå 1s ekvation och sätta in formuleringarna för nivå 2s ekvation:

$$\begin{aligned} stress_{ij} = & (\beta_{00} + \beta_{01} \cdot expcon_j + \beta_{02} \cdot hospsize_j + u_{0j}) \\ & + (\beta_{10} + \beta_{11} \cdot expcon_j + \beta_{12} \cdot hospsize_j + u_{1j}) \cdot age_{ij} \\ & + (\beta_{20} + \beta_{21} \cdot expcon_j + \beta_{22} \cdot hospsize_j + u_{2j}) \cdot gender_{ij} \\ & + (\beta_{30} + \beta_{31} \cdot expcon_j + \beta_{32} \cdot hospsize_j + u_{3j}) \cdot experien_{ij} \\ & + \varepsilon_{ij} \end{aligned}$$

4 Uppgift 4

Modellen i uppgiften ovan är komplex att skatta. Därför är det lämpligt om man hittar en algoritm för i vilken ordning variabler bör läggas till modellen. T ex ska man börja med GENDER eller EXPCON. Formulera en lämplig algoritm.

4.1 Sammanfattning av min algoritim

Denna algoritmen kan användas för genomföra en analys med en multilevelmodell. Algoritmen beskriver hur man börjar med en noll modell och därefter lägger till respektive nivåer (1 och 2) där man utför kontroll att parameterarna faktiskt bidrar till att förklara responsvariabeln.

4.2 Del 1

Man börjar med att skapa en noll modell, det vill säga en modell med endast ett slumpmässigt intercept. Detta görs för att undersöka om det finns skillnader i stressnivåer (responsvariabeln) mellan sjukhusen:

$$stress_{ij} = \beta_{00} + u_{0j} + \varepsilon_{ij}$$

Modellen kan även formuleras på följande sätt, vilket visar tydligt hur modellen endast har ett intercept (se uppgift 3 där jag visar varför man kan skriva på detta sätt):

$$stress_{ij} = \beta_{0j} + \varepsilon_{ij}$$

För noll modellen bör nu intraklasskorrelationen för att undersöka om en hierarkisk modell är nödvändig:

$$\text{Intraklasskorrelation} = \frac{\sigma_{u0}^2}{\sigma_{u0}^2 + \sigma_e^2}$$

Om intraklasskorrelationen är noll (eller mycket nära noll) finns ingen hierarkisk struktur i datamaterialet, vilket medför att vanlig linjär regression kan användas. Om intraklasskorrelationen är mer än noll innebär

det att observationerna i datamaterialet är beroende, och därav är linjär regression inte lämpligt. I detta fall skulle det innebära att observationerna inom samma sjukhus är beroende och en hierarkisk modell är lämplig.

4.3 Del 2

I det andra stadiet, givet att en hierarkisk modell är lämplig början man modellera nivå 1 (se förklaring av nivåer i uppgift 3). Nu vill man lägga till de förklarande variablerna (i denna uppgift *age*, *experien*, och *age*) steg för steg för kontrollera att dessa är signifikanta, det vill säga variabeln bidrar med förklaring av responsvariabeln. Om variabeln är icke-signifikant bidrar den inte med något för förklaringen av responsvariabeln och den kan ignoreras. I detta steg vill man kontrollera om lutningarna (respektive β för den förklarande variabeln) varirerar mellan nivå 2 (j), detta innebär att man kontrollera om slumpmässiga lutningar bör inkluderas i analysen. Själva syftet med detta är att undersöka hur individer inom sjukhusen skiljer sig åt. För bättre förståelse se exempel här:

$$stress_{ij} = \beta_{0j} + \beta_{1j} \cdot age_{ij} + \epsilon_{ij}$$

Här får varje nivå j (sjukhus) en egen lutning. Man lägger nu till en slumpmässig lutning $\beta_{1j} = \beta_{10} + u_{1j}$ och vi omformulerar ekvationen för att förstå vad som händer:

$$stress_{ij} = \beta_{0j} + (\beta_{10} + u_{1j}) \cdot age_{ij} + \epsilon_{ij}$$

Vad man ser nu är att en slumpmässig lutning har lagts till i modellen. Om denna lutning är signifikant innebär det att lutningarna skiljer sig åt mellan sjukhusen.

4.4 Del 3

I det tredje stadiet kan nivå 2 av variablerna introduceras, det vill säga de olika sjukhusnivåerna *expcon* och *hopssize*. Variablerna i nivå 2 används för att undersöka skillnaderna mellan sjukhusen. Detta innebär att man nu vill lägga till så respektive förklarande variabel får sin egen ekvation beroende på *expcon* och *hopssize*. För att förstå detta demonstrerar jag hur ekvationerna utvidgas.

Man börjar alltså med modellen:

$$stress_{ij} = \beta_{0j} + \beta_{1j} \cdot age_{ij} + \beta_{2j} \cdot gender_{ij} + \beta_{3j} \cdot experien_{ij} + \epsilon_{ij}$$

Vi formulerar om respektive förklarande variabels lutning enligt nivå 2:s variabler (Notera att u_{1j}, u_{2j}, u_{3j} representerar de slumpmässiga lutningarna i modellen):

$$\beta_{1j} + \beta_{11} \cdot expcon_j + \beta_{12} \cdot hospsize_j + u_{1j}$$

$$\beta_{2j} + \beta_{21} \cdot expcon_j + \beta_{22} \cdot hospsize_j + u_{2j}$$

$$\beta_{3j} + \beta_{31} \cdot expon_j + \beta_{32} \cdot hospsize_j + u_{3j}$$

I detta stadiet vill man då finna hur nivå 2:s (j) variabler förklarar skillnaderna i responsvariabeln. Sätter man ihop ekvationerna från del 1 och 2 får den fullständiga modellen:

$$\begin{aligned}
stress_{ij} = & (\beta_{00} + \beta_{01} \cdot expcon_j + \beta_{02} \cdot hospsize_j + u_{0j}) \\
& + (\beta_{10} + \beta_{11} \cdot expcon_j + \beta_{12} \cdot hospsize_j + u_{1j}) \cdot age_{ij} \\
& + (\beta_{20} + \beta_{21} \cdot expcon_j + \beta_{22} \cdot hospsize_j + u_{2j}) \cdot gender_{ij} \\
& + (\beta_{30} + \beta_{31} \cdot expcon_j + \beta_{32} \cdot hospsize_j + u_{3j}) \cdot experien_{ij} \\
& + \varepsilon_{ij}
\end{aligned}$$

4.5 Del 4

När man skapades dessa modeller, eftersom det görs steg för steg kan man utföra Likelihood Ratio tester för att bedömma om modellen förbättras för varje del som läggs till. Notera att detta görs för respektive nivåer 1 och 2. Sedan kombinerar man dem! Till slut bör man ha en modell som ser ut som sista ekvationen i del 3.

5 Uppgift 5

I uppgiften ovan formulerade du en algoritm för hur variabler ska läggas till i modellen. Använd nu denna algoritm. Börja därför med att skatta en enkel varianskomponentmodell med endast en oberoende variabel som bara antas ha en fix effekt. Bestäm, utifrån ett hypotestest, om variabeln ska vara med som en fix effekt eller inte. Bestäm därefter, utifrån ett hypotestest, om den även ska vara med som en slumpmässig effekt. Lägg därefter till en ny oberoende variabel och fortsätt tills du har fått en modell som du är nöjd med. Redovisa, för varje steg (om tillämpbart), a) ekvationen/ekvationerna för modellen du testar, b) modellens antaganden, c) resultat från t/z-test för fixa parametrar, d) resultat från likelihood-ratio-test för slumpmässiga parametrar, e) relevanta utskrifter samt SASkod.

5.1 Min algorit i praktiken

I denna uppgift testar jag min algoritm i praktiken. Denna uppgift följer samma start som tidigare uppgift därför kommer samma kod att återanvändas vid vissa fall (t.ex med nollmodellen).

5.1.1 Del 1: noll modell

Noll modellen presenterade i tidigare avsnitt, och eftersom denna modell är samma som tidigare återvänts koden från uppgiften. Modellen med ett fixt intercept formuleras som:

$$stress_{ij} = \beta_0 + \epsilon_{ij}$$

Medan modellen med ett slumpmässigt intercept formuleras enligt:

$$stress_{ij} = \beta_0 + u_{0j} + \epsilon_{ij}$$

$$\rho = \frac{0.2844}{0.2844 + 0.6870} = 0.293$$

Intraklasskorrelationen är 29% vilket innebär att 29% av variationen i stress beror på skillnader mellan sjukhus. För att beräkna skillnaden inom samma sjukhus blir detta $1 - 0.29 = 0.71\%$. Det innebär att 71% beror på skillnader inom samma sjukhus medan 29% beror på skillnader mellan dem. Med det sagt verkar det finnas en klusterstruktur i datamaterialet där individers stress liknar varnadra inom samma sjukhus, och av denna anledning kan en hierarkisk modell med ett slumpmässigt intercept användas.

Två modeller skapas där ett har ett slumpmässigt intercept med ett -2 Res Log Likelihood på 2533.0 och en modell med ett fixt intercept som får -2 Res Log Likelihood på 2797.7. Nu beräknas Likelihood Ratio testet för att se om undersöka om en modell med ett slumpmässigt intercept är bättre än en utan. Följande statistika fås

$$D_{01} = 2797.7 - 2533.0 = 264.7 \sim \chi^2_1$$

Med en frihetsgrad blir det kritiska värdet på 5% signifikans 3.84, vilket medför att nollhypotesen kan förkastas. En mer avancerad modell med ett slumpmässigt intercept är bättre än en modell utan.

Antaganden

- Det slumpmässiga interceptet u_{0j} antas vara oberoende och normalfördelat med väntevärde 0 och varians σ_{u0}^2 , det vill säga $u_{0j} \sim N(0, \sigma_{u0}^2)$.
- Feltermen på individnivå ε_{ij} antas vara oberoende och normalfördelade med väntevärde 0 och konstant varians σ_e^2 , det vill säga $\varepsilon_{ij} \sim N(0, \sigma_e^2)$.
- Det slumpmässiga interceptet u_{0j} antas vara oberoende av feltermerna ε_{ij} .
- Observationer inom samma sjukhus får vara korrelerade via det slumpmässiga interceptet, men observationer mellan olika sjukhus antas vara oberoende.

5.1.2 Del 2: lägger till nivå 1

Age

Börjar med att lägga till *age* som en fixed effekt i modellen. Modellen formuleras enligt:

$$\text{stress}_{ij} = \beta_0 + \beta_1 \cdot \text{Age}_{ij} + u_{0j} + \epsilon_{ij}$$

För denna modell används ett t-test för att undersöka om parameterskattningen är signifikant. Det vill säga, följande hypotes testas:

$$H_0 : \beta_{age} = 0$$

$$H_a : \beta_{age} \neq 0$$

P-värdet för modellen är 0.1388 vilket medför att nollhypotesen inte kan förkastas på 5% signifikans. Med detta sagt kan man inte säkerställa att *age* har en signifikant fixed effekt på stress. Jag testar nu att lägga till *experien* som fixed effekt i nästa del.

| Iteration History | | | | |
|-------------------|-------------|-----------------|------------|--|
| Iteration | Evaluations | -2 Res Log Like | Criterion | |
| 0 | 1 | 2805.30794981 | | |
| 1 | 2 | 2541.18030363 | 0.00000001 | |

Convergence criteria met.

| Covariance Parameter Estimates | | |
|--------------------------------|----------|----------|
| Cov Parm | Subject | Estimate |
| Intercept | hospital | 0.2836 |
| Residual | | 0.6862 |

| Fit Statistics | | |
|--------------------------|--|--------|
| -2 Res Log Likelihood | | 2541.2 |
| AIC (Smaller is Better) | | 2545.2 |
| AICC (Smaller is Better) | | 2545.2 |
| BIC (Smaller is Better) | | 2547.6 |

| Solution for Fixed Effects | | | | | |
|----------------------------|----------|----------------|-----|---------|---------|
| Effect | Estimate | Standard Error | DF | t Value | Pr > t |
| Intercept | 4.9968 | 0.1097 | 24 | 45.54 | <.0001 |
| CAge | -0.00325 | 0.002196 | 973 | -1.48 | 0.1388 |

| Type 3 Tests of Fixed Effects | | | | |
|-------------------------------|--------|--------|---------|--------|
| Effect | Num DF | Den DF | F Value | Pr > F |
| CAge | 1 | 973 | 2.20 | 0.1388 |

| Iteration History | | | | |
|-------------------|-------------|-----------------|------------|--|
| Iteration | Evaluations | -2 Res Log Like | Criterion | |
| 0 | 1 | 2805.30794981 | | |
| 1 | 2 | 2541.18030363 | 0.00000001 | |

| |
|---------------------------|
| Convergence criteria met. |
|---------------------------|

| Covariance Parameter Estimates | | |
|--------------------------------|----------|----------|
| Cov Parm | Subject | Estimate |
| Intercept | hospital | 0.2836 |
| Residual | | 0.6862 |

| Fit Statistics | |
|--------------------------|--------|
| -2 Res Log Likelihood | 2541.2 |
| AIC (Smaller is Better) | 2545.2 |
| AICC (Smaller is Better) | 2545.2 |
| BIC (Smaller is Better) | 2547.6 |

| Solution for Fixed Effects | | | | | |
|----------------------------|----------|----------------|-----|---------|---------|
| Effect | Estimate | Standard Error | DF | t Value | Pr > t |
| Intercept | 4.9968 | 0.1097 | 24 | 45.54 | <.0001 |
| CAge | -0.00325 | 0.002196 | 973 | -1.48 | 0.1388 |

| Type 3 Tests of Fixed Effects | | | | |
|-------------------------------|--------|--------|---------|--------|
| Effect | Num DF | Den DF | F Value | Pr > F |
| CAge | 1 | 973 | 2.20 | 0.1388 |

Experien

Nu läggs *experien* till som en fixed effekt i modellen, medan *age* behålls som en fixed med. Modellen med både *age* och *experien* som fixa effekter formuleras enligt:

$$\text{stress}_{ij} = \beta_0 + \beta_1 \cdot \text{Age}_{ij} + \epsilon \cdot \text{Experien}_{ij} + u_{0j} + \epsilon_{ij}$$

Man ser att både *age* och *experien* är signifkanta vilket medför att *age* blir signifikant när *experien* läggs till i modellen. Mest troligtvis innehåller detta att variablerna är korrelerade. Följande hypoteser testas:

$$H_0 : \beta_{age} = 0 \quad H_0 : \beta_{experien} = 0$$

$$H_a : \beta_{age} \neq 0 \quad H_a : \beta_{experien} \neq 0$$

Ett F-test används nu för att undersöka om variablerna har en signifikant fixed effekt. F-test för *age* är 28.61 och 58.09 för *age*, med respektive p-värden 0.001 och 0.001. Med det sagt kan nollhypotesen förkastas. Variablerna har båda signifikanta fixa effekter.

| The Mixed Procedure | | |
|----------------------------------------|---------------------|-------------------------------------------------------------------|
| Model Information | | |
| Data Set | WORK.CHOSPITAL | |
| Dependent Variable | stress | |
| Covariance Structure | Variance Components | |
| Subject Effect | hospital | |
| Estimation Method | REML | |
| Residual Variance Method | Profile | |
| Fixed Effects SE Method | Model-Based | |
| Degrees of Freedom Method | Containment | |
| Class Level Information | | |
| Class | Levels | Values |
| hospital | 25 | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 |
| Dimensions | | |
| Covariance Parameters | 2 | |
| Columns in X | 3 | |
| Columns in Z per Subject | 1 | |
| Subjects | 25 | |
| Max Obs per Subject | 52 | |
| Number of Observations | | |
| Number of Observations Read | 999 | |
| Number of Observations Used | 999 | |
| Number of Observations Not Used | 0 | |

| Iteration History | | | | |
|-------------------|-------------|-----------------|------------|--|
| Iteration | Evaluations | -2 Res Log Like | Criterion | |
| 0 | 1 | 2773.43373081 | | |
| 1 | 2 | 2492.67819892 | 0.00000000 | |

| |
|---------------------------|
| Convergence criteria met. |
|---------------------------|

| Covariance Parameter Estimates | | |
|--------------------------------|----------|----------|
| Cov Parm | Subject | Estimate |
| Intercept | hospital | 0.2850 |
| Residual | | 0.6481 |

| Fit Statistics | | |
|--------------------------|--------|--|
| -2 Res Log Likelihood | 2492.7 | |
| AIC (Smaller is Better) | 2496.7 | |
| AICC (Smaller is Better) | 2496.7 | |
| BIC (Smaller is Better) | 2499.1 | |

| Solution for Fixed Effects | | | | | | |
|----------------------------|----------|----------------|-----|---------|---------|--|
| Effect | Estimate | Standard Error | DF | t Value | Pr > t | |
| Intercept | 4.9964 | 0.1098 | 24 | 45.50 | <.0001 | |
| CAge | 0.01980 | 0.003702 | 972 | 5.35 | <.0001 | |
| CExperien | -0.05628 | 0.007384 | 972 | -7.62 | <.0001 | |

| Type 3 Tests of Fixed Effects | | | | |
|-------------------------------|--------|--------|---------|--------|
| Effect | Num DF | Den DF | F Value | Pr > F |
| CAge | 1 | 972 | 28.61 | <.0001 |
| CExperien | 1 | 972 | 58.09 | <.0001 |

Nu är man intresserad av att undersöka om *age* och *experien* bör vara fixa effekter eller slumpmässiga. Därför görs en modell med *Age* som slumpmässig effekt, medan *experien* hålls som fixed. Modellen när *age* får en slumpmässig lutning men *experien* hålls fixed formuleras enligt:

$$\text{stress}_{ij} = \beta_0 + \beta_1 \cdot \text{Age}_{ij} + \beta_2 \cdot \text{Experien}_{ij} + u_{0j} + u_{1j} \cdot \text{Age}_{ij} + \epsilon_{ij}$$

Hypoteserna kommer inte skrivas upp nu eftersom de är samma som tidigare, hur som helst är F testerna signifikanta. Men nu vill man inte undersöka signifikanta parametrar utan om en modell med *age* som slumpmässig och *experien* som fixed är bättre än en modell där båda variabler hålls fixed. Detta görs med ett Likelihood Ratio test. Från utskiftena ser man att -2 Res Log Likelihood är 2492.7 för båda modellerna. Hypoteser ställs nu upp och statistikan beräknas.

$$H_0 : \sigma_{\text{age}}^2 = 0 \quad H_0 : \sigma_{\text{age}}^2 > 0$$

$$D_{01} = 2492.7 - 2492.7 = 0 \quad \sim \chi^2_1 = 3.83$$

Man kan notera att statistikan är mindre än det kritiska värdet vilket medför att nollhypotesen inte kan förkastas. Det vill säga att *age* bör inte vara en slumomässig effekt. Därför behålls nu *age* som fixed, men *experien* testas nu som slumpmässig effekt.

| The Mixed Procedure | | |
|----------------------------------------|--------|-------------------------------------------------------------------|
| Model Information | | |
| Data Set | | WORK.CHOSPITAL |
| Dependent Variable | | stress |
| Covariance Structure | | Variance Components |
| Subject Effect | | hospital |
| Estimation Method | | REML |
| Residual Variance Method | | Profile |
| Fixed Effects SE Method | | Model-Based |
| Degrees of Freedom Method | | Containment |
| Class Level Information | | |
| Class | Levels | Values |
| hospital | 25 | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 |
| Dimensions | | |
| Covariance Parameters | | 3 |
| Columns in X | | 3 |
| Columns in Z per Subject | | 2 |
| Subjects | | 25 |
| Max Obs per Subject | | 52 |
| Number of Observations | | |
| Number of Observations Read | | 999 |
| Number of Observations Used | | 999 |
| Number of Observations Not Used | | 0 |

| Iteration History | | | | |
|-------------------|-------------|-----------------|------------|--|
| Iteration | Evaluations | -2 Res Log Like | Criterion | |
| 0 | 1 | 2773.43373081 | | |
| 1 | 2 | 2492.67820730 | 0.00000003 | |
| 2 | 1 | 2492.67819854 | 0.00000000 | |

Convergence criteria met.

| Covariance Parameter Estimates | | |
|--------------------------------|----------|----------|
| Cov Parm | Subject | Estimate |
| Intercept | hospital | 0.2850 |
| CAge | hospital | 0 |
| Residual | | 0.6481 |

| Fit Statistics | |
|--------------------------|--------|
| -2 Res Log Likelihood | 2492.7 |
| AIC (Smaller is Better) | 2496.7 |
| AICC (Smaller is Better) | 2496.7 |
| BIC (Smaller is Better) | 2499.1 |

| Solution for Fixed Effects | | | | | |
|----------------------------|----------|----------------|-----|---------|---------|
| Effect | Estimate | Standard Error | DF | t Value | Pr > t |
| Intercept | 4.9964 | 0.1098 | 24 | 45.50 | <.0001 |
| CAge | 0.01980 | 0.003702 | 24 | 5.35 | <.0001 |
| CExperien | -0.05628 | 0.007384 | 948 | -7.62 | <.0001 |

| Type 3 Tests of Fixed Effects | | | | |
|-------------------------------|--------|--------|---------|--------|
| Effect | Num DF | Den DF | F Value | Pr > F |
| CAge | 1 | 24 | 28.61 | <.0001 |
| CExperien | 1 | 948 | 58.10 | <.0001 |

Nu testas *experien* som en slumpmässig effekt medan *age* hålls fixed. Modellen när *experien* får en slumpmässig lutning men *age* hålls fixed formuleras enligt:

$$\text{stress}_{ij} = \beta_0 + \beta_1 \cdot \text{age}_{ij} + \beta_2 \cdot \text{experien}_{ij} + u_{0j} + u_{1j} \cdot \text{experien}_{ij} + \epsilon_{ij}$$

Det kan noteras att modellen återigen får ett -2 Res Log Likelihood på 2492.7 vilket medför att testet statistikan kommer bli 0 (som tidigare). Detta medför att nollhypotesen, att $\sigma_{\text{experien}}^2 = 0$ inte kan förkastas på 5% signifikans. På grund av detta bör alltså *age* och *experien* hållas som fixa effekter. Nu kan man gå vidare och lägga till *gender*.

The Mixed Procedure

| Model Information | |
|---------------------------|---------------------|
| Data Set | WORK.CHOSPITAL |
| Dependent Variable | stress |
| Covariance Structure | Variance Components |
| Subject Effect | hospital |
| Estimation Method | REML |
| Residual Variance Method | Profile |
| Fixed Effects SE Method | Model-Based |
| Degrees of Freedom Method | Containment |

| Class Level Information | | |
|-------------------------|--------|-------------------------------------------------------------------|
| Class | Levels | Values |
| hospital | 25 | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 |

| Dimensions | |
|--------------------------|----|
| Covariance Parameters | 3 |
| Columns in X | 3 |
| Columns in Z per Subject | 2 |
| Subjects | 25 |
| Max Obs per Subject | 52 |

| Number of Observations | |
|---------------------------------|-----|
| Number of Observations Read | 999 |
| Number of Observations Used | 999 |
| Number of Observations Not Used | 0 |

| Iteration History | | | |
|-------------------|-------------|-----------------|------------|
| Iteration | Evaluations | -2 Res Log Like | Criterion |
| 0 | 1 | 2773.43373081 | |
| 1 | 2 | 2492.67820792 | 0.00000003 |
| 2 | 1 | 2492.67819854 | 0.00000000 |

Convergence criteria met.

| Covariance Parameter Estimates | | |
|--------------------------------|----------|----------|
| Cov Parm | Subject | Estimate |
| Intercept | hospital | 0.2850 |
| CExperien | hospital | 0 |
| Residual | | 0.6481 |

| Fit Statistics | |
|--------------------------|--------|
| -2 Res Log Likelihood | 2492.7 |
| AIC (Smaller is Better) | 2496.7 |
| AICC (Smaller is Better) | 2496.7 |
| BIC (Smaller is Better) | 2499.1 |

| Solution for Fixed Effects | | | | | |
|----------------------------|----------|----------------|-----|---------|---------|
| Effect | Estimate | Standard Error | DF | t Value | Pr > t |
| Intercept | 4.9964 | 0.1098 | 24 | 45.50 | <.0001 |
| CAge | 0.01980 | 0.003702 | 948 | 5.35 | <.0001 |
| CExperien | -0.05628 | 0.007384 | 24 | -7.62 | <.0001 |

| Type 3 Tests of Fixed Effects | | | | |
|-------------------------------|--------|--------|---------|--------|
| Effect | Num DF | Den DF | F Value | Pr > F |
| CAge | 1 | 948 | 28.61 | <.0001 |
| CExperien | 1 | 24 | 58.10 | <.0001 |

Gender

Nu läggs *gender* till i modellen som en fixed effekt, medan *age* och *experien* hålls fixa. Modellen när *gender* läggs till med en fixed lutning, medan *age* och *experien* hålls fixa formuleras enligt:

$$\text{stress}_{ij} = \beta_0 + \beta_1 \cdot \text{age}_{ij} + \beta_2 \cdot \text{experien}_{ij} + \beta_3 \cdot \text{gender}_{ij} + u_{0j} + \epsilon_{ij}$$

I utskriften kan det noteras att alla variabler är signifikanta. F-statistikan för *gender* är signifikant med ett p-värde på 0.0001 vilket medför att följande nollhypoteskan förkastas:

$$H_0 : \beta_{\text{gender}} = 0 \quad H_a : \beta_{\text{gender}} \neq 0$$

The Mixed Procedure

| Model Information | |
|---------------------------|---------------------|
| Data Set | WORK.CHOSPITAL |
| Dependent Variable | stress |
| Covariance Structure | Variance Components |
| Subject Effect | hospital |
| Estimation Method | REML |
| Residual Variance Method | Profile |
| Fixed Effects SE Method | Model-Based |
| Degrees of Freedom Method | Containment |

| Class Level Information | | |
|-------------------------|--------|-------------------------------------------------------------------|
| Class | Levels | Values |
| hospital | 25 | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 |
| gender | 2 | 0 1 |

| Dimensions | |
|--------------------------|----|
| Covariance Parameters | 2 |
| Columns in X | 5 |
| Columns in Z per Subject | 1 |
| Subjects | 25 |
| Max Obs per Subject | 52 |

| Number of Observations | |
|---------------------------------|-----|
| Number of Observations Read | 999 |
| Number of Observations Used | 999 |
| Number of Observations Not Used | 0 |

| Iteration History | | | | |
|-------------------|-------------|-----------------|------------|--|
| Iteration | Evaluations | -2 Res Log Like | Criterion | |
| 0 | 1 | 2724.17963195 | | |
| 1 | 2 | 2430.51405695 | 0.00000000 | |

| |
|---------------------------|
| Convergence criteria met. |
|---------------------------|

| Covariance Parameter Estimates | | |
|--------------------------------|----------|----------|
| Cov Parm | Subject | Estimate |
| Intercept | hospital | 0.2790 |
| Residual | | 0.6065 |

| Fit Statistics | | |
|--------------------------|--------|--|
| -2 Res Log Likelihood | 2430.5 | |
| AIC (Smaller is Better) | 2434.5 | |
| AICC (Smaller is Better) | 2434.5 | |
| BIC (Smaller is Better) | 2437.0 | |

| Solution for Fixed Effects | | | | | | |
|----------------------------|--------|----------|----------------|-----|---------|---------|
| Effect | gender | Estimate | Standard Error | DF | t Value | Pr > t |
| Intercept | | 4.8719 | 0.1095 | 24 | 44.47 | <.0001 |
| gender | 0 | 0.4668 | 0.05649 | 971 | 8.26 | <.0001 |
| gender | 1 | 0 | . | . | . | . |
| CAge | | 0.01996 | 0.003581 | 971 | 5.57 | <.0001 |
| CExperien | | -0.05783 | 0.007146 | 971 | -8.09 | <.0001 |

| Type 3 Tests of Fixed Effects | | | | |
|-------------------------------|--------|--------|---------|--------|
| Effect | Num DF | Den DF | F Value | Pr > F |
| gender | 1 | 971 | 68.27 | <.0001 |
| CAge | 1 | 971 | 31.07 | <.0001 |
| CExperien | 1 | 971 | 65.49 | <.0001 |

Nu testas det om *gender* bör läggas till som en slumpmässig effekt istället. Modellen när *gender* läggs till med en slumpmässig lutning, medan *age* och *experien* hålls fixa formuleras enligt:

$$\text{stress}_{ij} = \beta_0 + \beta_1 \cdot \text{age}_{ij} + \beta_2 \cdot \text{experien}_{ij} + \beta_3 \cdot \text{gender}_{ij} + u_{0j} + u_{1j} \cdot \text{gender}_{ij} + \epsilon_{ij}$$

Det kan noteras att från utskrifterna att -2 Res Log Likelihood är 2430.5 för båda modellerna. På grund av detta kommer statistikan att bli 0 igen, vilket medför att nollhypotesen förkastas inte $H_0 : \beta_{\text{gender}} = 0$ på 5% signifikans. Detta innebär att *gender* bör hållas som en fixed effekt istället.

The Mixed Procedure

| Model Information | |
|---------------------------|---------------------|
| Data Set | WORK.CHOSPITAL |
| Dependent Variable | stress |
| Covariance Structure | Variance Components |
| Subject Effect | hospital |
| Estimation Method | REML |
| Residual Variance Method | Profile |
| Fixed Effects SE Method | Model-Based |
| Degrees of Freedom Method | Containment |

| Class Level Information | | |
|-------------------------|--------|-------------------------------------------------------------------|
| Class | Levels | Values |
| hospital | 25 | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 |
| gender | 2 | 0 1 |

| Dimensions | |
|--------------------------|----|
| Covariance Parameters | 3 |
| Columns in X | 5 |
| Columns in Z per Subject | 3 |
| Subjects | 25 |
| Max Obs per Subject | 52 |

| Number of Observations | |
|---------------------------------|-----|
| Number of Observations Read | 999 |
| Number of Observations Used | 999 |
| Number of Observations Not Used | 0 |

| Iteration History | | | | |
|-------------------|-------------|-----------------|------------|--|
| Iteration | Evaluations | -2 Res Log Like | Criterion | |
| 0 | 1 | 2724.17963195 | | |
| 1 | 2 | 2430.51406621 | 0.00000003 | |
| 2 | 1 | 2430.51405694 | 0.00000000 | |

Convergence criteria met.

| Covariance Parameter Estimates | | |
|--------------------------------|----------|----------|
| Cov Parm | Subject | Estimate |
| Intercept | hospital | 0.2790 |
| gender | hospital | 0 |
| Residual | | 0.6065 |

| Fit Statistics | |
|--------------------------|--------|
| -2 Res Log Likelihood | 2430.5 |
| AIC (Smaller is Better) | 2434.5 |
| AICC (Smaller is Better) | 2434.5 |
| BIC (Smaller is Better) | 2437.0 |

| Solution for Fixed Effects | | | | | | |
|----------------------------|--------|----------|----------------|-----|---------|---------|
| Effect | gender | Estimate | Standard Error | DF | t Value | Pr > t |
| Intercept | | 4.8719 | 0.1095 | 24 | 44.47 | <.0001 |
| gender | 0 | 0.4668 | 0.05649 | 24 | 8.26 | <.0001 |
| gender | 1 | 0 | . | . | . | . |
| CAge | | 0.01996 | 0.003581 | 947 | 5.57 | <.0001 |
| CExperien | | -0.05783 | 0.007146 | 947 | -8.09 | <.0001 |

| Type 3 Tests of Fixed Effects | | | | |
|-------------------------------|--------|--------|---------|--------|
| Effect | Num DF | Den DF | F Value | Pr > F |
| gender | 1 | 24 | 68.27 | <.0001 |
| CAge | 1 | 947 | 31.07 | <.0001 |
| CExperien | 1 | 947 | 65.49 | <.0001 |

Antaganden

För modellerna där age , $experien$ och $gender$ läggs till som fixa effekter antas:

- Sambandet mellan stress och de förklarande variablerna på individnivå är linjärt, det vill säga att effekten av age , $experien$ och $gender$ kan beskrivas med raka linjer i modellen.
- Slumpmässiga intercept u_{0j} på sjukhusnivå antas vara oberoende och normalfördelade med väntevärde 0 och varians σ_{u0}^2 .
- Feltermerna ε_{ij} antas vara oberoende, normalfördelade med väntevärde 0 och konstant varians σ_e^2 givet de slumpmässiga effekterna.
- Det slumpmässiga interceptet u_{0j} är oberoende av feltermerna ε_{ij} .
- För modellerna där slumpmässiga lutningar testas (tillfälligt för age och $experien$) antas vektorn av slumpmässiga effekter på sjukhusnivå vara multivariat normalfördelad med väntevärde 0.

Slutgiltig modell

Den slutgiltiga modellen inkluderar *age*, *experien*, och *gender* som fixa effekter.

5.1.3 Del 3: lägger till nivå 2

Hospsize

I denna del av algoritmen börjas med att lägga till *hospsize* som en fixed effekt i modellen. Ingen av variablene i nivå 2 kommer att testat som slumpmässiga effekter eftersom de beskriver egenskaper hos sjukhusen som inte kan vara slumpmässiga. Det vill säga det är inte ett slumpmässigt urval av nivåer. När *hospsize* läggs till i modellen kommer ekvationen av förändringar. Ekationsformuleringen blir:

$$stress_{ij} = \beta_0 + \beta_1 \cdot age_{ij} + \beta_2 \cdot experien_{ij} + \beta_3 \cdot gender_{ij} + \beta_4 \cdot hospsize_j + u_{0j} + \epsilon_{ij}$$

För att se om *hospsize* har en signifikanta effekt på responsvariabeln sätts följande hypoteser upp:

$$H_0 : \beta_4 = 0 \quad H_a : \beta_4 \neq 0$$

Ett F-test används där statistikan blir 6.16 med tillhörande p-värde på 0.0022. Med sagt förkastas nollhypotesen på 5% signifikans, vilket medför att *hospsize* har en signifikant effekt på responsvariabeln (stress). Denna behålls nu i modellen. Ett likelihood ratio test kan användas för att testa om en modell med *hospsize* (mer komplex modell) är signifikant bättre än en modell med endast variablene för nivå 1. Statistikan ställs upp som tidigare.

$$D_{01} = 2430.5 - 2442.3 = 8.2 \quad \sim \chi^2_2 = 5.99$$

Statistikan och det kritiska värdet ovan visar att en mer komplex modell med variabeln *hospsize* för nivå 2 är signifikant bättre än en modell med endast variablene från nivå 1.

The Mixed Procedure

| Model Information | |
|---------------------------|---------------------|
| Data Set | WORK.CHOSPITAL |
| Dependent Variable | stress |
| Covariance Structure | Variance Components |
| Subject Effect | hospital |
| Estimation Method | REML |
| Residual Variance Method | Profile |
| Fixed Effects SE Method | Model-Based |
| Degrees of Freedom Method | Containment |

| Class Level Information | | |
|-------------------------|--------|-------------------------------------------------------------------|
| Class | Levels | Values |
| hospital | 25 | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 |
| gender | 2 | 0 1 |
| expcon | 2 | 0 1 |
| hospsize | 3 | 0 1 2 |

| Dimensions | |
|--------------------------|----|
| Covariance Parameters | 2 |
| Columns in X | 8 |
| Columns in Z per Subject | 1 |
| Subjects | 25 |
| Max Obs per Subject | 52 |

| Number of Observations | |
|---------------------------------|-----|
| Number of Observations Read | 999 |
| Number of Observations Used | 999 |
| Number of Observations Not Used | 0 |

| Convergence criteria met. | | | | | | | |
|------------------------------------------------|--------|----------|----------|----------------|-----|---------|---------|
| Covariance Parameter Estimates | | | | | | | |
| Cov Parm Subject Estimate | | | | | | | |
| Intercept hospital 0.1902 | | | | | | | |
| Residual 0.6065 | | | | | | | |
| Fit Statistics | | | | | | | |
| -2 Res Log Likelihood 2422.3 | | | | | | | |
| AIC (Smaller is Better) 2426.3 | | | | | | | |
| AICC (Smaller is Better) 2426.3 | | | | | | | |
| BIC (Smaller is Better) 2428.7 | | | | | | | |
| Solution for Fixed Effects | | | | | | | |
| Effect | gender | hospsize | Estimate | Standard Error | DF | t Value | Pr > t |
| Intercept | | | 5.3951 | 0.2278 | 22 | 23.68 | <.0001 |
| CAge | | | 0.01994 | 0.003580 | 971 | 5.57 | <.0001 |
| CExperien | | | -0.05777 | 0.007144 | 971 | -8.09 | <.0001 |
| gender | 0 | | 0.4667 | 0.05648 | 971 | 8.26 | <.0001 |
| gender | 1 | | 0 | . | . | . | . |
| hospsize | 0 | | -0.9023 | 0.2727 | 971 | -3.31 | 0.0010 |
| hospsize | 1 | | -0.4126 | 0.2622 | 971 | -1.57 | 0.1159 |
| hospsize | 2 | | 0 | . | . | . | . |
| Type 3 Tests of Fixed Effects | | | | | | | |
| Effect | Num DF | Den DF | F Value | Pr > F | | | |
| CAge | 1 | 971 | 31.01 | <.0001 | | | |
| CExperien | 1 | 971 | 65.40 | <.0001 | | | |
| gender | 1 | 971 | 68.27 | <.0001 | | | |
| hospsize | 2 | 971 | 6.16 | 0.0022 | | | |

Nu läggs *expcon* till i modellen. Detta innebär att modellekvationen förändras till:

$$\text{stress}_{ij} = \beta_0 + \beta_1 \cdot \text{age}_{ij} + \beta_2 \cdot \text{experien}_{ij} + \beta_3 \cdot \text{gender}_{ij} + \beta_4 \cdot \text{hospsize} + \beta_5 \cdot \text{expcon}_{ij} + \epsilon_{ij}$$

För se om *expcon* faktiskt bidrar signifikant med att förklara responsvariabeln ställs följande hypoteser upp:

$$H_0 : \beta_5 = 0 \quad H_a : \beta_5 \neq 0$$

Ett F-test används där statistikan blir 297.06 med tillhörande p-värde på 0.0001. Detta medför att nollhypotesen kan förkastas och *expcon* bidrar med en signifikant effekt. För att undersöka om denna modell med *expcon* är bättre än en modell utan används återigen ett likelihood ratio test. Från föregående modell var -2 Res Log Likelihood 2422.3, och för denna modell 2167.3. Det innebär att statistikan och det kritiska värdet blir:

$$D_{01} = 2422.3 - 2167.3 = 2550 \quad \sim \chi^2_1 = 3.84$$

Statistikan är mycket större än det kritiska värdet vilket medför att en mer komplex modell är bättre. Det vill säga att inkludera *expcon* i modellen är bättre än en modell utan den.

Antaganden

- Sambanden mellan stress och de förklarande variablerna (age , $experien$, $gender$, $hospsize$, $expcon$) antas vara linjära.
- $Hospsize$ och $expcon$ behandlas som fixa effekter som beskriver systematiska skillnader mellan sjukhusen, inte slumptäckta varierande nivåer.
- De slumptäckta intercepten u_{0j} på sjukhusnivå antas vara oberoende och normalfördelade med väntevärde 0 och varians σ_{u0}^2 .
- Feltermerna ε_{ij} antas vara oberoende, normalfördelade med väntevärde 0 och konstant varians σ_e^2 givet de slumptäckta effekterna.
- Slumptäckta effekter u_{0j} antas vara oberoende av feltermerna ε_{ij} , och slumptäckta effekter för olika sjukhus antas vara oberoende av varandra.

5.2 Slutgiltiga modellen

Modellen ovan är den sista modellen vilket även är den bästa modellen för detta datamaterial. Alla variabler kunde inkluderas. Det visade sig att ingen av variablerna för nivå 1 bör ha slumptäckta lutningar, alltså har dessa behållts som fixa effekter. Båda variabler för nivå 2 är inkluderas som fixa effekter (motivering: se tidigare del).

5.3 Bifogad SAS KOD

5.3.1 Del 1 kod

```
1 *skattar en nollmodell för att genomföra intraklasskorrelationen;
2 proc mixed data=CHospital method=reml;
3   class hospital;
4   model stress = / solution;
5   random intercept / subject=hospital;
6 run;
```

```
1 *Skattar två modeller för att genomföra ett LRT test;
2 *REML används för väntevärdesriktiga skattningar;
3 *Class används för att automatiskt skapa dummyvariabler;
4 proc mixed data=CHospital method=reml;
5   class hospital;
6   model stress = / solution;
7   random intercept / subject=hospital type=vc;
8 run;
9
10 proc mixed data=CHospital method=reml;
11   class hospital;
12   model stress = / solution ;
13 run;
```

5.3.2 Del 2 kod

```
1 *Gr age med fixed effect;
2 proc mixed data=CHospital method=reml;
3   class hospital;
4   model stress = Cage / solution;
5   random intercept / subject=hospital type=vc;
6 run;
7
8 *gr samma med experien;
9
10 proc mixed data=CHospital method=reml;
11   class hospital;
12   model stress = Cage Cexperien / solution;
13   random intercept / subject=hospital type=vc;
14 run;
15
16
17 *testar nu gra age till slumpmssig;
18 proc mixed data=CHospital method=reml;
19   class hospital;
20   model stress = CAge CExperien / solution;
21   random intercept CAge / subject=hospital type=vc;
22 run;
23
24 *Testar age som fix men experien som slump;
25 proc mixed data=CHospital method=reml;
26   class hospital;
27   model stress = CAge CExperien / solution;
28   random intercept CExperien / subject=hospital type=vc;
29 run;
30
31 *Testar nu lägga till gender som fixed;
32 proc mixed data=CHospital method=reml;
33   class hospital gender;
34   model stress = gender CAge CExperien / solution;
35   random intercept / subject=hospital type=vc;
36 run;
37
38 *Lägger till gender som slump;
39 proc mixed data=CHospital method=reml;
40   class hospital gender;
41   model stress = gender CAge CExperien / solution;
42   random intercept gender / subject=hospital type=vc;
43 run;
```

5.3.3 Del 3 SAS kod

```
1 *Lägger till hospsize som fixed effekt (nivå 2);
2 proc mixed data=CHospital method=reml;
3   class hospital gender expcon hospsize;
4   model stress = CAge CExperien gender hospsize / solution;
5   random intercept / subject=hospital type=vc;
6 run;
7
8 *Lägger till expcon (nivå 2) som sista variabel;
9 proc mixed data=CHospital method=reml;
10  class hospital gender expcon hospsize;
```

```

11 model stress = CAge CExperien gender hospsize expcon / solution;
12 random intercept / subject=hospital type=vc;
13 run;

```

6 Uppgift 6

Tolka alla skattade parametrar (beta-koefficienter, varianser, kovarianser) i modellen som du kom fram till i uppgiften ovan.

Innan parameterskattningarna tolkas är det viktigt att förstå vilken skala variablerna har, och om dessa är kategoriska hur många nivåerna de har. Den fullständiga modellen består av följande variabler:

- Age: kontinuerlig.
- Experien: kontinuerlig.
- Gender: kategorisk med två nivåer (0/1). Där 0 är man och 1 är kvinna.
- Hospsize: kategorisk med tre nivåer (0/1/2). Där 0 är ett litet sjukhus, 1 är ett medelstort, och 2 är ett stort.
- Expcon: kategorisk med två nivåer (0/1). Där 0 står för kontrollgruppen som inte genomgick ett program i stresshantering och 1 står för experimentgruppen som genomgick programmet.

Här är tabellen som redovisar parameterskattningarna för respektive variabel:

| Solution for Fixed Effects | | | | | | | | |
|----------------------------|--------|--------|----------|----------|----------------|-----|---------|---------|
| Effect | gender | expcon | hospsize | Estimate | Standard Error | DF | t Value | Pr > t |
| Intercept | | | | 5.0320 | 0.2293 | 22 | 21.95 | <.0001 |
| CAge | | | | 0.01921 | 0.003135 | 970 | 6.13 | <.0001 |
| CExperien | | | | -0.05786 | 0.006255 | 970 | -9.25 | <.0001 |
| gender | 0 | | | 0.4530 | 0.04946 | 970 | 9.16 | <.0001 |
| gender | 1 | | | 0 | . | . | . | . |
| hospsize | | | 0 | -0.9035 | 0.2736 | 970 | -3.30 | 0.0010 |
| hospsize | | | 1 | -0.4157 | 0.2630 | 970 | -1.58 | 0.1143 |
| hospsize | | | 2 | 0 | . | . | . | . |
| expcon | | 0 | | 0.7441 | 0.04317 | 970 | 17.24 | <.0001 |
| expcon | | 1 | | 0 | . | . | . | . |

| Covariance Parameter Estimates | | |
|--------------------------------|----------|----------|
| Cov Parm | Subject | Estimate |
| Intercept | hospital | 0.1952 |
| Residual | | 0.4647 |

Tolkning av interceptet (icke slumpmässigt)

Från tabellen kan man notera att för varje kategoriska variabel finns det en nivå som inte har några tillhörande skattningar. Detta innebär att denna nivå är referenskategorin. För *gender* är referenskategorin 1, för *hospsize* är det 2, och för *expcon* är det 1. Interceptet har en signifikant parameterskattning med en genomsnittlig stressnivå på 5.03. Mer specifikt tolkas detta som den förväntade stressnivån givet att man är en kvinnlig sjukskötterska (*gender* = 1) som arbetar på ett stort sjukhus (*hospsize* = 2) tillhörande experimentgruppen (*expcon* = 1), med en genomsnittlig ålder (centrerad *age* = 0) samt en genomsnittlig erfarenhet (centrerad *experien* = 0). Det kan noteras att *ålder* går att tolkas nu eftersom den är centrerad. Det vill säga att man får en genomsnittlig ålder. Om *age* inte hade var centrerad skulle man inte kunna tolka den, eftersom det skulle innebära en person som är noll år.

Tolkning av förklarande variablers parameterskattningar

Parameterskattningen för *age* är 0.01921 och signifikant. Detta innebär att för varje år äldre än medelåldern i urvalet ökar den förväntade nivån av stress i genomsnitt 0.019 enheter givet att resterande variabler hålls konstanta. Skattningen för *experien* är -005786 och signifikant, vilket innebär att för varje år mer erfarenhet än medelerfarenheten minskar den förväntade nivån av stress med 0.058 enheter, givet att alla andra variabler hålls konstanta. För *gender* är skattningen 0.4530 (signifikant) jämfört med referenskategorin. Detta innebär att män har i genomsnitt högre stress än kvinnor när man kontrollerar för *age*, *experien*, *hospsize*, och *expcon*. För *hospsize* finns det två parameterskattningar där nivå 2 är referenskategorin. Man kan notera att för små sjukhus (*hospsize* = 0) är skattningen -0.9035 och signifikant. För medelstora sjukhus (*hospsize* = 1) är skattningen -0.4157, men icke signifikant. Detta innebär att små sjukhus har i genomsnitt 0.9 enheter lägre enheter av stress jämfört med stora sjukhus, när man kontrollerar för *age*, *experien*, *hospsize*, och *expcon*. Tolkningen för medelstora sjukhus är densamma för små, men eftersom skattningen är icke-signifikant är den ointressant. Skattningen för *expcon* är 0.7441 med referenskategorin 1. Det vill säga de individer i kontrollgruppen, i genomsnitt har 0.74 enheter mer stress än de i experimentgruppen, när man kontrollera för de nämnda förklarande variablerna.

Tolkning av varianskomponenterna i modellen

Varianskomponenterna i modellen finner man i tabellen över *Covariance Parameter Estimates* vilket berör *interceptet* och *residualen*. Interceptet här är variansen i sjukhusens slumpmässiga intercept u_{0j} . Tänk nu på att detta är variansen $\sigma_{u0}^2 = 0.1952$. Om denna parameter hade varit noll skulle det innebära alla den genomsnittliga stressnivån bland de 25 sjukhusen är lika (fixed). På grund av att σ_{u0}^2 är större än noll tyder det på att sjukhusen i sina genomsnittliga stressnivåer skiljer sig åt. *Residual* = 0.4647 detta är σ_e^2 vilket visar variationen mellan sjukskötterskorna inom sjukhusen som modellen inte lyckas fånga.

Intraklasskorrelationen beräknas för att förstå hur bra modellen fångar verkligheten i datamaterialet. Formel som tidigare presenterades återanvänds:

$$\rho = \frac{\sigma_{u0}^2}{\sigma_{u0}^2 + \sigma_e^2} = \frac{0.1952}{0.1952 + 0.4647} = 0.296$$

Intraklasskorrelationen indikerar att 29.6% av variationen i stressnivåer berör skillnader mellan sjukhusen. För skillnaderna mellan sjukskötterskorna i sjukhusen är det $1 - 0.296 = 0.704$. Det vill säga att modellen visar att 70.4% av variationen är mellan individer inom sjukhusen.

Utförande i SAS: type = "vc" istället för "un"

I denna analys användes type = vc istället för Type = un. Det vc (variance component) gör annorlunda från un är att de slumpmässiga effekterna bara får varsin varians. Det vill säga det finns ingen kovarians

mellan effekterna, men un ger både varians och kovarians. Det visades att den fullständiga modellen endast innehåller ett slumpmässigt intercept för sjukhus vilket i sin tur innebär att ingen kovarians kan skattas. Anledningen varför ingen kovarians kan skattas för endast ett slumpmässigt intercept (σ_{u0}^2). För beräkning av kovarians krävs två slumpmässiga effekter. Jag kommer nu att bevisa att detta är sanning genom att skatta den slutgiltiga modellen med type = un istället för vc .

| The Mixed Procedure | | |
|----------------------------------------|--------|-------------------------------------------------------------------|
| Model Information | | |
| Data Set | | WORK.CHOSPITAL |
| Dependent Variable | | stress |
| Covariance Structure | | Unstructured |
| Subject Effect | | hospital |
| Estimation Method | | REML |
| Residual Variance Method | | Profile |
| Fixed Effects SE Method | | Model-Based |
| Degrees of Freedom Method | | Containment |
| Class Level Information | | |
| Class | Levels | Values |
| hospital | 25 | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 |
| gender | 2 | 0 1 |
| expcon | 2 | 0 1 |
| hospsize | 3 | 0 1 2 |
| Dimensions | | |
| Covariance Parameters | | 2 |
| Columns in X | | 10 |
| Columns in Z per Subject | | 1 |
| Subjects | | 25 |
| Max Obs per Subject | | 52 |
| Number of Observations | | |
| Number of Observations Read | | 999 |
| Number of Observations Used | | 999 |
| Number of Observations Not Used | | 0 |

| Convergence criteria met. | | | | | | | | |
|-----------------------------------------|------------|------------|----------|----------|----------------|-----|---------|---------|
| Covariance Parameter Estimates | | | | | | | | |
| Cov Parm | Subject | Estimate | | | | | | |
| UN(1,1) | hospital | 0.1952 | | | | | | |
| Residual | | 0.4647 | | | | | | |
| Fit Statistics | | | | | | | | |
| -2 Res Log Likelihood | 2167.3 | | | | | | | |
| AIC (Smaller is Better) | 2171.3 | | | | | | | |
| AICC (Smaller is Better) | 2171.4 | | | | | | | |
| BIC (Smaller is Better) | 2173.8 | | | | | | | |
| Null Model Likelihood Ratio Test | | | | | | | | |
| DF | Chi-Square | Pr > ChiSq | | | | | | |
| 1 | 244.58 | <.0001 | | | | | | |
| Solution for Fixed Effects | | | | | | | | |
| Effect | gender | expcon | hospsize | Estimate | Standard Error | DF | t Value | Pr > t |
| Intercept | | | | 5.0320 | 0.2293 | 22 | 21.95 | <.0001 |
| CAge | | | | 0.01921 | 0.003135 | 970 | 6.13 | <.0001 |
| CExperien | | | | -0.05786 | 0.006255 | 970 | -9.25 | <.0001 |
| gender | 0 | | | 0.4530 | 0.04946 | 970 | 9.16 | <.0001 |
| gender | 1 | | | 0 | . | . | . | . |
| hospsize | | 0 | | -0.9035 | 0.2736 | 970 | -3.30 | 0.0010 |
| hospsize | | 1 | | -0.4157 | 0.2630 | 970 | -1.58 | 0.1143 |
| hospsize | | 2 | | 0 | . | . | . | . |
| expcon | 0 | | | 0.7441 | 0.04317 | 970 | 17.24 | <.0001 |
| expcon | 1 | | | 0 | . | . | . | . |
| Type 3 Tests of Fixed Effects | | | | | | | | |
| Effect | Num DF | Den DF | F Value | Pr > F | | | | |
| CAge | 1 | 970 | 37.53 | <.0001 | | | | |
| CExperien | 1 | 970 | 85.57 | <.0001 | | | | |
| gender | 1 | 970 | 83.88 | <.0001 | | | | |
| hospsize | 2 | 970 | 6.12 | 0.0023 | | | | |
| expcon | 1 | 970 | 297.06 | <.0001 | | | | |

Notera i utskriften att denna modell är identisk med den som skattades med type = vc skillnaden är att det slumpmässiga interceptet, vilket är densamma som tidigare, betecknas $UN(1, 1)$. Sammanfattningsvis skattas bara ett slumpmässigt intercept för sjukhusen, i sin tur innebär detta att kovariansmatrisen inte kan beräknas eftersom den endast innehåller en effekt, medan kovariansen behöver två effekter för att beräknas. På grund av detta blir resulaten samma oavsett om type = un eller type = vc. Nu har dock ett likelihood ratio test lagts till, vilket är ointressant eftersom den jämfört med en tom modell.

SAS kod för type = "unmodell"

```

1 proc mixed data=CHospital method=reml;
2   class hospital gender expcon hospsize;
3   model stress = CAge CExperien gender hospsize expcon / solution;
4   random intercept / subject=hospital type=un;
5   run;

```

7 Uppgift 7

Utifrån dina analyser i uppgifterna ovan, 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.

Resulaten från föregående uppgit visar att stresshanteringsprogrammet har en effekt, där sjuksköterskor i kontrollgruppen har ett högre genomsnitt på stressnivån än de som deltagit i programmet. Det slumpmässiga interceptet i modellen visar att den genomsnittliga stressnivån skiljer sig mellan de tjugofem sjukhusen. Intraklasskorrelationen visar att drygt 30% av variationen i stressnivåer beror på skillnader mellan sjukhusen, och 70% av variationen beror på skillnader mellan sjuksköterskorna inom sjukhusen.

Modellen inkluderar inga slumpmässiga lutningar för variablerna i nivå 2, vilket innebär att effekten av programmet antas vara densamma för alla sjukhus. Modellen kan alltså visa att sjukhusen skiljer sig åt i sin genomsnittliga stressnivå (m.h.a det slumpmässiga interceptet) och att stresshanteringsprogrammet har en övergripande effekt, men modellen lyckas inte säga något om effekten skiljer sig åt mellan sjukhusen (t.ex om programmet fungerar bättre på sjukhus 1 än på sjukhus 2).

För att undersöka om effekten av programmen skiljer sig åt mellan sjukhusen kan man prova att lägga till en interaktion mellan nivå 2 variablerna och undersöka om dessa är signifikanta. Enligt teorin är det lämpligt för variabler inom nivå 1 att tillåtas ha slumpmässiga lutningar.