BST5220 multilevel HW1

Due Wednesday by 12:00 pm, 2/11

Miao Cai <miao.cai@slu.edu>

2/4/2019

You will first need to run the SAS code (homework1\_data.sas) to create the SAS data set. The purpose of this study is to assess how job-related stress is related to hospital size (0=small, 1=medium, 2=large) and nurse’s job experience (years). Data are from nurses working in 25 hospitals. In each hospital, a sample of about 40 nurses is selected and given a test that measures job-related stress (coded on a scale of 0-7).

1. Identify the type of the data structure (cross-sectional clustered, longitudinal, or clustered longitudinal).
2. Identify the variables at each level.
3. Graphically examine the association between job stress and nurse’s experience within each hospital. Refer to slides 8-11, lecture 2.
4. Use the model building strategies discussed in lecture 3 to select the best model for the data

dat = rio::import("data/a1.sas7bdat")  
head(dat)

## hospital hospital\_size nurse stress experience  
## 1 1 2 1 7 11  
## 2 1 2 2 7 20  
## 3 1 2 3 7 7  
## 4 1 2 4 6 25  
## 5 1 2 5 6 22  
## 6 1 2 6 6 22

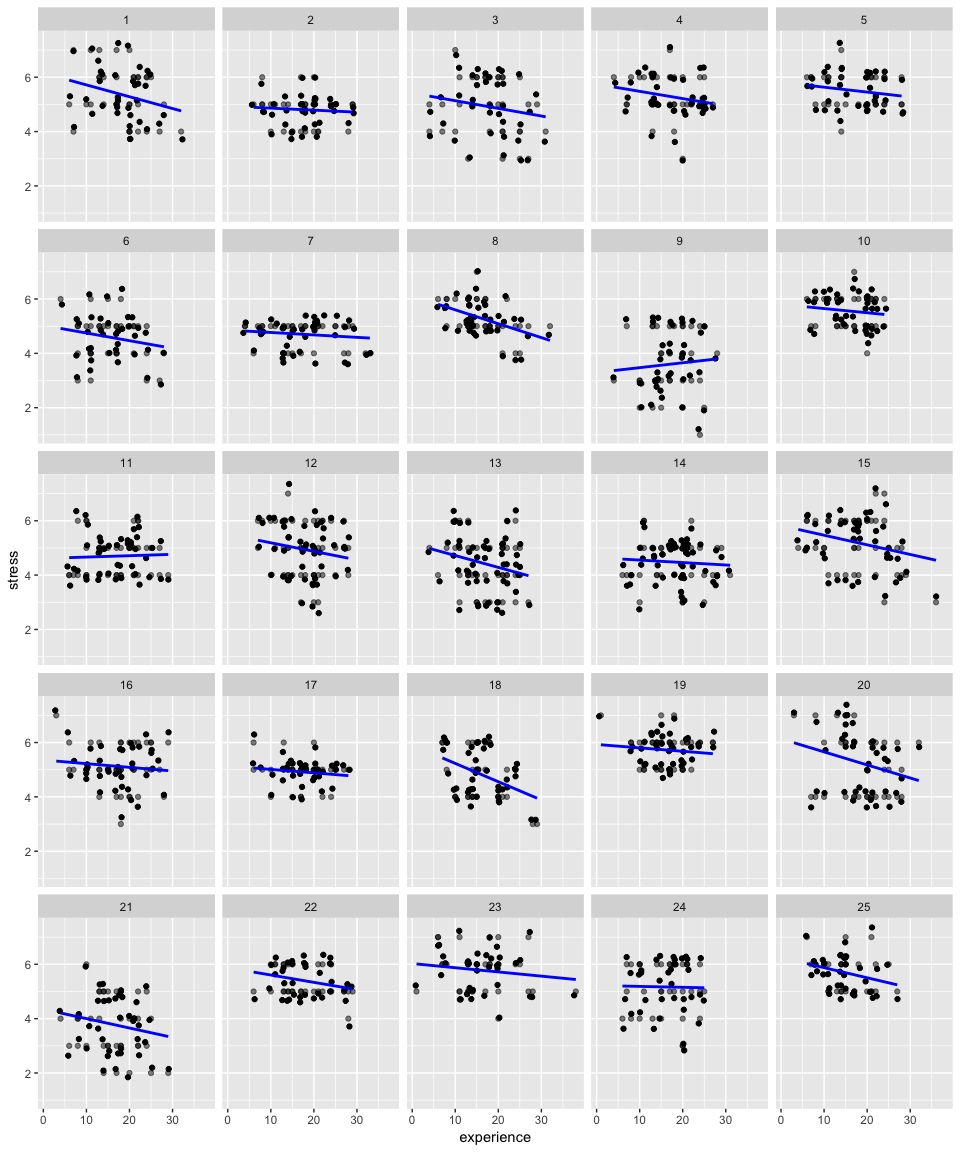
1. This is a cross-sectional clustered data struture
2. Variables at each level:

* Hospital level: hospital, hospital\_size
* Nurse level: stress, experience

1. Graphically examine the association between job stress and nurse’s experience within each hospital

pacman::p\_load(tidyverse)  
  
dat %>%   
 ggplot(aes(experience, stress,)) +

geom\_point(alpha = 0.5) + geom\_jitter() +   
 geom\_smooth(method = "lm", se=FALSE, color="blue") +   
 facet\_wrap(.~hospital, ncol = 5)



1. First, I started with a random intercept model:

/\* 0 random intercept model \*/

**proc** **mixed** data = hw1.a1 noclprint covtest noitprint method=ml;

class hospital;

model stress = / solution ddfm = bw;

random intercept / subject = hospital type = UN;

**run**;\*-2loglikelihood: 2532.0;

| **Covariance Parameter Estimates** | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Cov Parm** | | **Subject** | | **Estimate** | | **Standard Error** | | | | **Z Value** | | **Pr > Z** | |
| **UN(1,1)** | | hospital | | 0.2728 | | 0.08203 | | | | 3.33 | | 0.0004 | |
| **Residual** | |  | | 0.6863 | | 0.03109 | | | | 22.08 | | <.0001 | |
| **Solution for Fixed Effects** | | | | | | | | | | | | | |
| **Effect** | | **Estimate** | | **Standard Error** | | | | **DF** | **t Value** | | | **Pr > |t|** | |
| **Intercept** | | 4.9973 | | 0.1077 | | | | 24 | 46.38 | | | <.0001 | |
| **Fit Statistics** | | | | | | | | | | |
| **-2 Log Likelihood** | | | | | | | 2532.0 | | | |
| **AIC (Smaller is Better)** | | | | | | | 2538.0 | | | |
| **AICC (Smaller is Better)** | | | | | | | 2538.1 | | | |
| **BIC (Smaller is Better)** | | | | | | | 2541.7 | | | |

The random intercepts and between-hospital variances are significant, which indicates that two-level model is appropriate.

Then, I added the only one nurse level fixed effect variable, experience:

/\* 1 random intercept model with nurse-level fixed effects\*/

**proc** **mixed** data = hw1.a1 noclprint covtest noitprint method=ml;

class hospital;

model stress = experience/ solution ddfm = bw;

random intercept / subject = hospital type = UN;

**run**; \*-2loglikelihood: 2501.8;

\* difference in -2loglikelihood: 2532.0-2501.8 = 30.2;

| **Covariance Parameter Estimates** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Cov Parm** | **Subject** | **Estimate** | **Standard Error** | **Z Value** | **Pr > Z** |
| **UN(1,1)** | hospital | 0.2711 | 0.08140 | 3.33 | 0.0004 |
| **Residual** |  | 0.6655 | 0.03014 | 22.08 | <.0001 |

| **Fit Statistics** | | | | |
| --- | --- | --- | --- | --- |
| **-2 Log Likelihood** | | | 2501.8 | |
| **AIC (Smaller is Better)** | | | 2509.8 | |
| **AICC (Smaller is Better)** | | | 2509.8 | |
| **BIC (Smaller is Better)** | | | 2514.6 | |
| **Solution for Fixed Effects** | | | | | | | | |
| **Effect** | | **Estimate** | **Standard Error** | | **DF** | | **t Value** | **Pr > |t|** |
| **Intercept** | | 5.4046 | 0.1300 | | 24 | | 41.56 | <.0001 |
| **experience** | | -0.02391 | 0.004312 | | 974 | | -5.55 | <.0001 |

I also did a deviance test between the random intercept only model and this current model by comparing the difference of -2\*log likelihood:

**data** pvalue;

df =**1**; chisq = **30.2**;

pvalue = **1** - probchi(chisq, df);

**run**;

**proc** **print** data = pvalue noobs;**run**;

| **df** | **chisq** | **pvalue** |
| --- | --- | --- |
| 1 | 30.2 | 3.8971E-8 |

The test is significant which suggests that model 2 is more appropriate.

Then I added a hospital level fixed effect (hospital size), and performed the deviance test between model 2 and this model

/\* 2 random intercept model with hospital level effects \*/

**proc** **mixed** data = hw1.a1 noclprint covtest noitprint method=ml;

class hospital;

model stress = experience hospital\_size/ solution ddfm = bw;

random intercept / subject = hospital type = UN;

**run**; \*-2loglikelihood: 2490.6;

\* difference in -2loglikelihood: 2501.8 - 2490.6 =11.2;

**data** pvalue;

df =**1**; chisq = **11.2**;

pvalue = **1** - probchi(chisq, df);

**run**;

**proc** **print** data = pvalue noobs;**run**;

| **Covariance Parameter Estimates** | | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Cov Parm** | | | **Subject** | | | **Estimate** | | | **Standard Error** | | | | **Z Value** | | **Pr > Z** |
| **UN(1,1)** | | | hospital | | | 0.1675 | | | 0.05222 | | | | 3.21 | | 0.0007 |
| **Residual** | | |  | | | 0.6655 | | | 0.03014 | | | | 22.08 | | <.0001 |
| **Fit Statistics** | | | | | | | | | | | |
| **-2 Log Likelihood** | | | | | | | | | 2490.6 | | |
| **AIC (Smaller is Better)** | | | | | | | | | 2500.6 | | |
| **AICC (Smaller is Better)** | | | | | | | | | 2500.7 | | |
| **BIC (Smaller is Better)** | | | | | | | | | 2506.7 | | |
| **Solution for Fixed Effects** | | | | | | | | | | | | | | | | |
| **Effect** | | | | | | **Estimate** | | | **Standard Error** | | | **DF** | | **t Value** | | **Pr > |t|** |
| **Intercept** | | | | | | 5.0331 | | | 0.1503 | | | 23 | | 33.48 | | <.0001 |
| **experience** | | | | | | -0.02390 | | | 0.004310 | | | 974 | | -5.54 | | <.0001 |
| **hospital\_size** | | | | | | 0.4644 | | | 0.1240 | | | 23 | | 3.75 | | 0.0011 |
| **df** | | | | **chisq** | **pvalue** | | |
| 1 | | | | 11.2 | .000817973 | | |

The deviance test suggests this model is significantly better than model 2. This turns out to be the best fitting model.

Then I tested whether nurse level variable experience has random effects:

/\* 3 Assessing whether nurse level variable has random effect\*/

**proc** **mixed** data = hw1.a1 noclprint covtest noitprint method=ml;

class hospital;

model stress = experience hospital\_size/ solution ddfm = bw;

random intercept experience/ subject = hospital type = UN;

**run**; \*Estimated G matrix is not positive definite (variance is nearly zero). NOT ADOPTED;

The SAS system reported that “Estimated G matrix is not positive definite (variance is nearly zero)”, which indicated that there is no nurse experience random effects at hospital level. Therefore, this model is not adopted.

Lastly, I tested cross-level effects between experience and hospital size:

\* 4 adding cross-level ;

**proc** **mixed** data = hw1.a1 noclprint covtest noitprint method=ml;

class hospital;

model stress = experience|hospital\_size/ solution ddfm = bw;

random intercept/ subject = hospital type = UN;

**run**;\* -2loglikelihood: 2490.6;

\* difference in -2loglikelihood: 0, no need to use Chi-square test;

It turns out that the -2\*log likelihood did not change between the two models, so there was no need to perform deviance test.

In conclusion, the best fitting model to the data is:

Level 1: Stressij = b0j + b1\*experienceij + eij

Level 2: b0j = r00 + r01\*hospital\_size + u0j