\*Notes to lecture in KLMED8008, 10 April 2013. Written by Stian Lydersen, 3 April 2013.

\*Stata commands for examples in Rabe-Hesketh & Skrondal 3ed Chapter 4

\* page numbers refer to the same book

\*avoid that the output stops during result reporting:

set more off

use http://www.stata-press.com/data/mlmus3/gcse, clear

\*regression analysis for school number 1:

regress gcse lrt if school==1

predict p\_gcse, xb

\* xb gives linear prediction

twoway (scatter gcse lrt) (line p\_gcse lrt, sort) if school==1, xtitle(LRT) ytitle(GCSE)

\*Produce a trellis graph with one plott for each of the 65 schools:

\*Norwegian for trellis: espalier, tremmeverk, sprinkelverk, gitter(verk), tråleverk, spileverk

twoway (scatter gcse lrt) (lfit gcse lrt, sort lpatt(solid)), by(school, compact legend(off) cols(5)) xtitle(LRT) ytitle(GSCE) ysize(3) xsize(2)

\*Regression analyses for all schools with at least 4 students:

egen num =count(gcse), by(school)

statsby inter=\_b[\_cons] slope=\_b[lrt], by(school) saving(M:\diverse\kurs\KLMED8008\ols, replace): regress gcse lrt if num>4

\*In the line above, I have added the path M:\diverse … and added “replace” after “ols” in “saving”.

sort school

merge m:1 school using M:\diverse\kurs\KLMED8008\ols

drop \_merge

\*Note a misprint on Page 185, line 7 from below:

\*There are not two schools with less than five students, but one school (with two students) with

\*less than five students. This is school no 48 with two students, as seen in the trellis graph.

twoway scatter slope inter, xtitle(Intercept) ytitle(Slope)

egen pickone=tag(school)

summarize inter slope if pickone==1

correlate inter slope if pickone==1

correlate inter slope if pickone==1, covariance

\*Calculate the fitted values for each school

generate pred=inter+slope\*lrt

sort school lrt

twoway (line pred lrt, connect(ascending)), xtitle(LRT) ytitle(Fitted regression line)

\*Analyses using xtmixed are shown below.

\*If you refer to the 2nd edition of the textbook:

\*We prefer xtmixed to gllamm for these analyses using normal distributions.

\* gllamm is more general than xtmixed, and was available before xtmixed was available.

\* xtmixed has been further extended (improved) since the since the 2nd edtion of the textbook.

\*Random intercept (only) model:

xtmixed gcse lrt || school:, mle

\*To compare this with the random coefficient model, we store the estimates:

estimates store ri

\*Random coefficient model: Include random slope in addition:

\*We must specify cov(unstructured), since the (stupid!) default is correlation=0,

\*see top of page 194

xtmixed gcse lrt || school:lrt, cov(unstructured) mle

\*Store the estimates

estimates store rc

\*test if the random slope model fits better than the random intercept model:

lrtest rc ri

\*Note that the provided LR test is conservative, see page 197-198.

estimates restore rc

\*results from rc are active now

\*Maximum likelihood estimation of the random intercepts and slopes

predict fixed, xb

generate totres=gcse-fixed

statsby mli=\_b[\_cons] mls=\_b[lrt], by(school) saving(M:\diverse\kurs\KLMED8008\ols, replace): regress totres lrt

sort school

merge m:1 school using M:\diverse\kurs\KLMED8008\ols

drop \_merge

\*School 48 has only two observations:

list lrt gcse mli mls if school==48, clean noobs

\*Here, the maximum likelihood estimates look odd. Empirical Bayes prediction is preferred.

Fortsett her …

\*Empirical Bayes:

estimates restore rc

predict ebs ebi, reffects

list school mli ebi mls ebs if pickone==1 & (school<10 | school==48), noobs

twoway (scatter ebi mli if pickone==1, mlabel(school))(function y=x, range(-35 10)), xtitle(ML estimate) ytitle(EB prediction) legend(off)

\*without school 48:

twoway (scatter ebi mli if pickone==1 & school!=48, mlabel(school))(function y=x, range(-10 10)), xtitle(ML estimate) ytitle(EB prediction) legend(off)

twoway (scatter ebs mls if pickone==1, mlabel(school))(function y=x, range(-8 0.6)), xtitle(ML estimate) ytitle(EB prediction) legend(off)

\*without school 48:

twoway (scatter ebs mls if pickone==1 & school!=48, mlabel(school))(function y=x, range(-0.6 0.6)), xtitle(ML estimate) ytitle(EB prediction) legend(off)

\*Note that the above plotting commands give the random part of the intercept and slope,

\*whose mean are (per definition) zero.

\*Predicted model-implied regression lines for random intercept model, and for the random slope model:

estimates restore ri

predict muri, fitted

sort school lrt

twoway (line muri lrt, connect(ascending)), xtitle(LRT) ytitle(Empirical Bayes regression lines Model 1)

estimates restore rc

predict murc, fitted

sort school lrt

twoway (line murc lrt, connect(ascending)), xtitle(LRT) ytitle(Empirical Bayes regression lines Model 2)

\*in the two following lines, there are misprints in the 2nd edition of the book, but not in 3rd ed.

\* Errata are found in http://www.stata-press.com/books/errata/mlmus2.html

histogram ebi if pickone==1, normal xtitle(Predicted random intercepts)

histogram ebs if pickone==1, normal xtitle(Predicted random slopes)

scatter ebs ebi if pickone==1, saving(M:\diverse\kurs\KLMED8008\yx, replace) xtitle(“Random intercept”) ytitle(“Random slope”) ylabel(, nogrid)

histogram ebs if pickone==1, freq horizontal saving(M:\diverse\kurs\KLMED8008\hy, replace) normal yscale(alt) ytitle(“ “) fxsize(35) ylabel(,nogrid)

histogram ebi if pickone==1, freq saving(M:\diverse\kurs\KLMED8008\hx, replace) normal xscale(alt) xtitle(“ “) fysize(35) ylabel(,nogrid)

graph combine M:\diverse\kurs\KLMED8008\hx.gph M:\diverse\kurs\KLMED8008\yx.gph M:\diverse\kurs\KLMED8008\hy.gph, hole(2) imargin(0 0 0 0)

predict res1, residuals

histogram res1, normal xtitle(Predicted Level 1-residuals)

\*Here I get a slightly different histogram than Fig 4.13 page 207, due to other histogram binning.

\*To assess normality, I prefer QQplots to histograms. Note that the default QQ plots in Stata have \*predicted normal values on the x axis, and observed values on the y aksis, opposite to SPSS default.

qnorm ebi if pickone==1

qnorm ebs if pickone==1

qnorm res1

\*One school has possibly a large slope. To find out which one, add mlabel(school):

qnorm ebs if pickone==1, mlabel(school)

\*4.8.5 inferences for the individual schools

estimates restore rc

predict slope\_rc inter\_se, reses

gsort +ebi-pickone

generate rank=sum(pickone)

generate labpos=ebi+1.96\*inter\_se+0.5

serrbar ebi inter\_se rank if pickone==1, addplot(scatter labpos rank, mlabel(school) msymbol(none) mlabpos(0)) scale(1.96) xtitle(Rank) ytitle(Prediction) legend(off)