

Workshop 7 - Hierarchical Data

JPS

01/03/2022

Introduction

We are going to work with two surveys we have used previously, where there is a clear two-level hierarchy, the European Social Survey (ESS) and the Crown Court Sentencing Survey (CCSS). In the ESS the level-1 observations are represented by the individuals that participated in the survey, with countries as level-2 observations (or clusters). In the CCSS, cases sentenced represent the level-1 observations, with court locations as level-2 observations. We will learn the different approaches that can be used to adjust for this hierarchical structure (taking within cluster correlations as a nuisance), but we will also learn how to interpret such hierarchical structures using multilevel modelling.

Exercise 1: In the first part of this exercise we are going to assess the effect of leaving the within cluster correlations present in the ESS unadjusted (known as naive analysis). We will do so by comparing results from such a naive analysis with adjustments based on the sandwich estimator and fixed effects models. As we learn how to adjust for within cluster correlations we will also try to estimate the specific effect of cohabitation (living with other people) on self-reported happiness. In the second part of the exercise we will employ multilevel models to explore two additional research questions: Is the variability in self-reported happiness between countries larger than between individuals within the same countries? And, is the effect of cohabitation on happiness uniform across countries?

Exercise 2: Here you are requested to investigate two key research questions in sentencing research using multilevel modelling. To what extent is the sentencing practice consistent across court locations in the Crown Court? Consistency has been the driving goal of the new sentencing guidelines, developed by the Sentencing Council for England and Wales over the last decade (see BBC interview with the ex-chairman of the Council Lord Justice Leveson). Hence, by exploring this question you will be generating evidence assessing the impact of this transformative process. The second research question to be explored is: Are the aggravating and mitigating factors intoxication and remorse applied consistently across courts? This is probably an even more policy-relevant question since the Sentencing Council launched a consultation in 2019 on whether specific guideline factors need to be more tightly defined to promote consistency. I suggest focusing on these two factors (intoxication and remorse) since, as we saw in Workshop 2, these are controversial factors in the sense that researchers and practitioners have suggested they are not used uniformly. Therefore, the evidence generated from this analysis is highly relevant to the Sentencing Council to identify areas in their guidelines that could be redesigned.

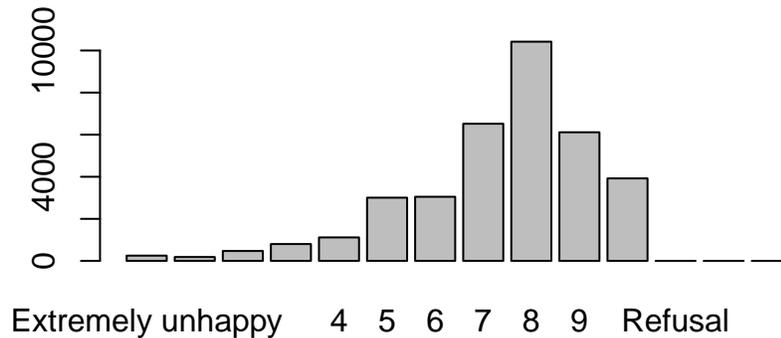
Exercise 1. Happiness in Europe

Let's access the ESS and trim it down so we keep just those variables that we will be using today. These are: gender ('gndr'), self-reported happiness ('happy'), how often meet with other people ('sclmeet'), feeling of safety walking alone in the dark ('aesfdrk'), number of people living in the same household ('hhmmb'), and the country where each interview took place ('cntry').

```
library(foreign)
ess = read.dta("ESS9e01_1.dta")
vars = c("agea", "gndr", "happy", "sclmeet", "aesfdrk", "hhmmb", "cntry")
ess = ess[vars]
```

We can move on to run a quick exploratory analysis for the three main variables of interest, 'happy', 'hhmmb', and 'cntry'. As we did in Workshop 5 we will also proceed to recode 'happy' so it can be expressed as a numeric variable.

```
plot(ess$happy)
```



```
table(ess$happy, useNA="ifany")
ess$happyrec = ifelse(ess$happy=="Extremely unhappy",0,
                      ifelse(ess$happy=="Extremely happy",10,as.numeric(as.character(ess$happy))))
table(ess$happyrec, useNA="ifany") #The recoding process went ok.
table(ess$cntry) #Looking at the distribution of level-2 clusters.
table(ess$hhmmb, useNA="ifany")
```

Now, to start exploring the effect of cohabitation on happiness we can rely on our old friend the linear regression model.

```
naive = lm(happyrec~sclmeet+aesfdrk+hhmmb+gndr, data=ess)
summary(naive)
```

```
##
## Call:
## lm(formula = happyrec ~ sclmeet + aesfdrk + hhmmb + gndr, data = ess)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.5411 -0.8615  0.2338  1.1970  5.6564
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      5.667669   0.074517  76.059 <2e-16 ***
## sclmeetLess than once a month  0.746384   0.076872   9.709 <2e-16 ***
## sclmeetOnce a month          1.386608   0.076640  18.092 <2e-16 ***
## sclmeetSeveral times a month  1.660976   0.073286  22.664 <2e-16 ***
## sclmeetOnce a week           1.814894   0.073751  24.608 <2e-16 ***
## sclmeetSeveral times a week   1.998592   0.072675  27.500 <2e-16 ***
## sclmeetEvery day             2.041786   0.075234  27.139 <2e-16 ***
## aesfdrkSafe                 -0.470864   0.022741 -20.706 <2e-16 ***
```

```
## aesfdrkUnsafe          -1.055157   0.030798 -34.261 <2e-16 ***
## aesfdrkVery unsafe    -1.476075   0.051999 -28.387 <2e-16 ***
## hhmb                   0.152026   0.007087  21.452 <2e-16 ***
## gndrFemale             0.223578   0.019770  11.309 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.802 on 35346 degrees of freedom
## (657 observations deleted due to missingness)
## Multiple R-squared:  0.1193, Adjusted R-squared:  0.119
## F-statistic: 435.4 on 11 and 35346 DF,  p-value: < 2.2e-16
```

This is an ok model since we have controlled for some relevant confounders. The effect of ‘hhmb’ is positive and significant. However, we are completely disregarding the hierarchical dimension of the data (interviewees clustered within countries), and with it the potentially non-negligible violation of the assumption of independence. We normally refer to such models, where an assumption has been flagrantly violated, as *naive* models. To assess whether that is the case we will use the sandwich estimator, which can be called using the *coefest* command within the *sandwich* package.

```
library(lmtest)
library(sandwich)
#Below we are requesting to re-estimate the model 'naive' using a variance covariance empirically determined
coefest(naive, vcov = vcovHC(naive, type="HC1"))
```

```
##
## t test of coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    5.6676692  0.1107015  51.198 < 2.2e-16 ***
## sclmeetLess than once a month 0.7463839  0.1142834   6.531 6.622e-11 ***
## sclmeetOnce a month          1.3866078  0.1128059  12.292 < 2.2e-16 ***
## sclmeetSeveral times a month 1.6609762  0.1096299  15.151 < 2.2e-16 ***
## sclmeetOnce a week           1.8148936  0.1098545  16.521 < 2.2e-16 ***
## sclmeetSeveral times a week  1.9985923  0.1090448  18.328 < 2.2e-16 ***
## sclmeetEvery day            2.0417859  0.1112326  18.356 < 2.2e-16 ***
## aesfdrkSafe                -0.4708644  0.0211905 -22.221 < 2.2e-16 ***
## aesfdrkUnsafe              -1.0551575  0.0334564 -31.538 < 2.2e-16 ***
## aesfdrkVery unsafe         -1.4760748  0.0692148 -21.326 < 2.2e-16 ***
## hhmb                       0.1520262  0.0073892  20.574 < 2.2e-16 ***
## gndrFemale                 0.2235776  0.0198490  11.264 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Notice how the size of the coefficients are identical in each model, but the standard errors across in the second model are slightly bigger. This difference is the result of the naive model assuming independence and therefore underestimating the standard errors and any other measures of uncertainty derived from them. The sandwich estimator is a handy technique to add to our toolbox. To be considered when the precision in the estimation process is of paramount importance and we suspect the assumptions of independence or heteroskedasticity do not hold.

Since there are only 19 countries we could also consider a fixed effects (FE) model introducing each level-2 unit in the model as dummy variables. This should help to partially adjust the standard errors by modelling average between cluster differences in happiness directly into the fixed part of the model. In addition, this approach can be quite useful to control for unobserved country-level differences that might be confounding (biasing) the effect of ‘hhmb’ estimated in our previous two models.

```
FE = lm(happyrec~sclmeet+aesfdrk+hhmb+gndr+cntry, data=ess)
summary(FE)
```

```
##
## Call:
## lm(formula = happyrec ~ sclmeet + aesfdrk + hhmb + gndr + cntry,
##     data = ess)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.4103 -0.8526  0.1665  1.0975  6.3178
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      5.949646   0.079618  74.727 < 2e-16 ***
## sclmeetLess than once a month  0.785394   0.074167  10.590 < 2e-16 ***
## sclmeetOnce a month      1.305565   0.074217  17.591 < 2e-16 ***
## sclmeetSeveral times a month  1.581087   0.071213  22.202 < 2e-16 ***
## sclmeetOnce a week      1.654218   0.071798  23.040 < 2e-16 ***
## sclmeetSeveral times a week  1.842953   0.070976  25.966 < 2e-16 ***
## sclmeetEvery day        2.054264   0.073516  27.943 < 2e-16 ***
## aesfdrkSafe            -0.308557   0.022609 -13.647 < 2e-16 ***
## aesfdrkUnsafe          -0.720622   0.030868 -23.345 < 2e-16 ***
## aesfdrkVery unsafe      -1.003526   0.051434 -19.511 < 2e-16 ***
## hhmb                   0.162168   0.006912  23.463 < 2e-16 ***
## gndrFemale              0.189168   0.019068   9.921 < 2e-16 ***
## cntryBE                 -0.069223   0.054516  -1.270 0.204174
## cntryBG                 -1.898169   0.052796 -35.953 < 2e-16 ***
## cntryCH                  0.193778   0.056571   3.425 0.000615 ***
## cntryCY                 -0.261757   0.071632  -3.654 0.000258 ***
## cntryCZ                 -0.672121   0.050516 -13.305 < 2e-16 ***
## cntryDE                  0.060743   0.050121   1.212 0.225542
## cntryEE                 -0.309309   0.053362  -5.796 6.83e-09 ***
## cntryFI                  0.292933   0.054216   5.403 6.59e-08 ***
## cntryFR                 -0.538295   0.052614 -10.231 < 2e-16 ***
## cntryGB                 -0.113022   0.051051  -2.214 0.026840 *
## cntryHU                 -0.617987   0.056692 -10.901 < 2e-16 ***
## cntryIE                 -0.000791   0.051281  -0.015 0.987693
## cntryIT                 -0.620261   0.048888 -12.687 < 2e-16 ***
## cntryNL                  0.067116   0.055364   1.212 0.225415
## cntryNO                 -0.150490   0.058742  -2.562 0.010415 *
## cntryPL                 -0.405425   0.058421  -6.940 4.00e-12 ***
## cntryRS                 -1.024462   0.052934 -19.354 < 2e-16 ***
## cntrySI                 -0.266168   0.059795  -4.451 8.56e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.733 on 35328 degrees of freedom
## (657 observations deleted due to missingness)
## Multiple R-squared:  0.1853, Adjusted R-squared:  0.1846
## F-statistic: 277.1 on 29 and 35328 DF,  p-value: < 2.2e-16
```

We can see that the standard errors are practically identical to the naive model, so in this case the FEs model did not really adjust them. However, we can see that the estimates of the regression coefficients are not

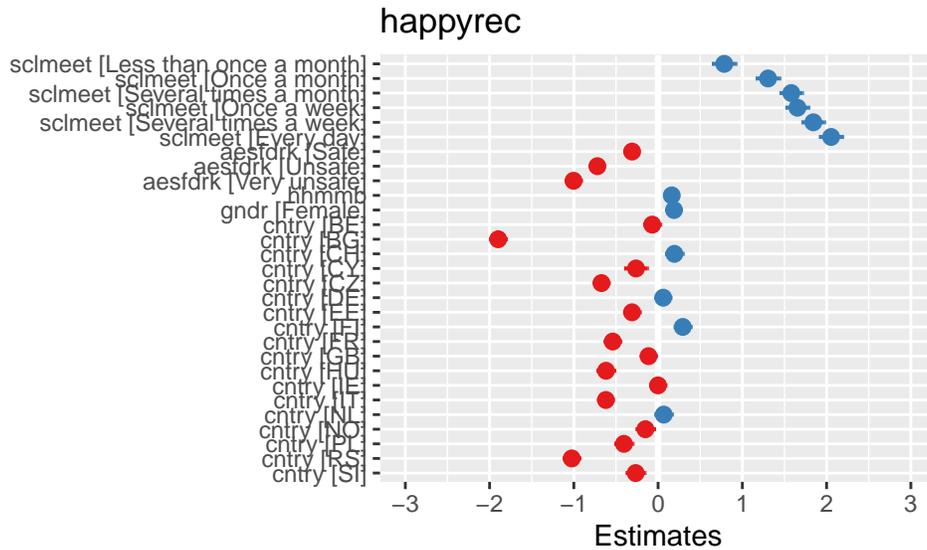
exactly the same. For the case of ‘hhmb’ the regression coefficient is around 12% bigger. This is probably due to the country-level confounding effects that we are now controlling for (e.g. differences in the size of residential properties across countries, or cultural differences, with Northern countries being generally more individualistic than Southern countries, where family bonds are more present). We could further interrogate the potential non-linear effect of ‘hhmb’ (using polynomial regression as in Workshop 4) but for the sake of simplicity we are just going to carry on assuming linearity. Similarly, we could also adjust for problems of non-response and coverage error present in the ESS using probability weights (as in Workshop 5), but we won’t do it here just to try to keep things simple.

Coming back to the comparison of the three models that we have used (naive, sandwich and fixed-effects) and considering that our aim is the estimation of the effect of ‘hhmb’ as accurately and precisely as possible, I would choose to report the FEs model. This is because:

1. The bias in FEs standard errors compared to those estimated using the sandwich estimator is almost negligible; and
2. There is a significant difference in the regression coefficient for ‘hhmb’, compared to either the naive model or the model resorting on the sandwich estimator.

Beyond, answering our first research question, the results of our FEs model are also quite illustrative of the substantial between country variability in self-reported levels of happiness. To visualise this more clearly we can use the `plot_model` command available in the `sjPlot` library.

```
library(sjPlot)
plot_model(FE)
```



People in Bulgaria (‘BG’) appear to be really miserable whereas Finnish (FI) seemed to be the happiest in Europe! Note Austria (‘AT’) is not there as it is being used a reference category. If you want to find out what labels are used for other countries you can find them in this website.

In the second part of this exercise we proceed to explore the two substantial research questions:

1. Is the variability in self-reported happiness between countries larger than between individuals within the same countries?
2. Is the effect of cohabitation on happiness uniform across countries?

The between country variability observed in the FEs model signals the relevance of these questions. However, the FEs model is only giving us country specific estimates of happiness, which is not exactly a measure of

the overall variability in happiness between countries, and it definitely is not telling us anything about any potential variability in the effect of ‘hhmmb’ on happiness between countries.

To explore those research questions we will use multilevel modelling. For example, to estimate the overall variability in happiness across countries we can expand our naive analysis by adding a random intercepts term u_j , so the model becomes, $Y_{ji} = (\beta_0 + u_j) + \beta_k X_{kji} + \epsilon_{ji}$. To estimate this multilevel model we will use the `lmer` command from the `lme4` package. The code required is identical to the standard `lm` model with the one difference that now we will need to specify the variable capturing the cluster units as follows ‘(1|cluster)’, see below.

```
library(lme4)
RI = lmer(happyrec ~ sclmeet+aesfdrk+hhmmb+gndr+(1|cntry), data=ess)
summary(RI)
```

Notice how the output from our ‘RI’ model includes a *random effects* part. Here we can find the estimated standard deviation of the level-2 and level-1 residuals, $\hat{\sigma}_u = 0.505$ and $\hat{\sigma}_\epsilon = 1.733$, which can be interpreted as the unexplained differences (after controlling for the four variables included in our model) between countries and between individuals. As we anticipated from the results of our FE model, the variability stemming at the country level is substantial. However, since the variability stemming from the individual level is more than three times bigger, we can determine that individual characteristics are more important at determining differences in happiness than country characteristics. To be more precise in answering the research question we can calculate the intraclass-correlation coefficient, $ICC = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_\epsilon^2}$.

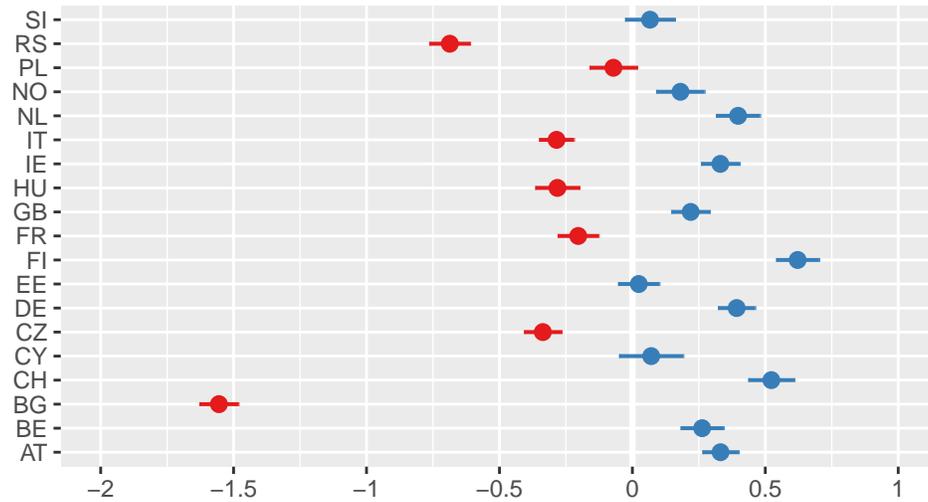
```
0.255 / (0.255+3.004)
```

So, only 7.8% of the unobserved variability stems from country differences. Another common interpretation of the ICC is as the correlation (similarity) of level-1 units within clusters, in our cases individuals within countries. In relation to our specific research question, we could interpret this findings as: although there are significant average differences in happiness across countries, in determining individual happiness, individual characteristics are more meaningful than country characteristics.

If we want to obtain the specific random intercept for each country (\hat{u}_j) we can use `ranef`. Results are very similar to those from the FE model, with Finland leading and Bulgaria lagging behind. Notice however, that unlike for the FE model these estimates are obtained from the residual (random) part of the model. As such, they are not really controlling for any potential country level confounders.

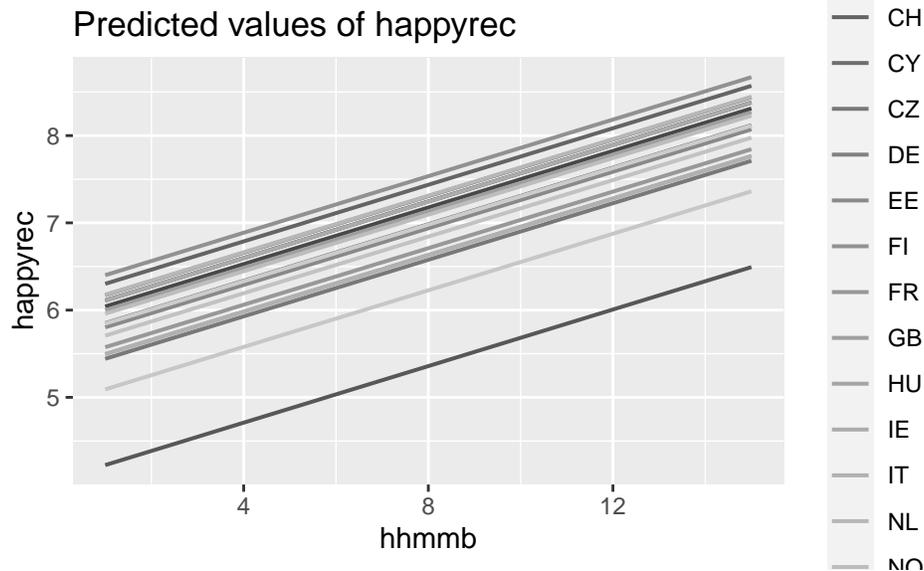
```
library(glmmTMB) #This is to be able to use the option "re" in plot_model
ranef(RI)
plot_model(RI, type="re")
```

Random effects



If you want to obtain the specific intercept for each country ($\beta_0 + u_j$) you can use `coef`. We can also create a linechart putting in context both the variability of these random intercepts and the effect of 'hhmb' using `plot_model`.

```
coef(RI)
#To drop the legend use, 'show.legend=FALSE'
plot_model(RI, type="pred", terms=c("hhmbb", "cntry"), pred.type="re", colors="gs", ci.lvl=NA)
```



Notice how the random intercepts model that we have estimated, and plotted in the last graph, assumes that the effect of 'hhmbb' on happiness is uniform across countries, what changes is the intercept, but not the slope. We can relax this assumption using a random slopes model. To tell R that we want both random intercepts and a random slopes term for 'hhmbb' we need to include 'hhmbb' in the random part of the model, as shown below.

```
RS = lmer(happyrec ~ sclmeet+aesfdrk+hhmbb+gndr+(1+hhmbb|cntry), data=ess)
summary(RS)
```

The model run and we obtained a set of estimates, however notice how we also got a warning saying that 'the

model failed to converge'. This can occur since random slopes models quickly become much more complex than standard linear models, and can make the estimation of the model unreliable. We could try to solve this problem by using a different estimation process, for example we could use Bayesian methods based on Markov chain Monte Carlo (McMC) but this is beyond the scope of today's session. What we will do instead is simplify the model by removing some of the explanatory variables. The model runs fine (achieves convergence) simply by dropping 'sclmeet'.

```
RS = lmer(happyrec ~ aesfdrk+hhmb+gndr+(1+hhmb|cntry), data=ess)
summary(RS)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: happyrec ~ aesfdrk + hhmb + gndr + (1 + hhmb | cntry)
## Data: ess
##
## REML criterion at convergence: 141333.8
##
## Scaled residuals:
##   Min       1Q   Median       3Q      Max
## -5.1856 -0.4857  0.0997  0.6359  3.1077
##
## Random effects:
## Groups Name Variance Std.Dev. Corr
## cntry (Intercept) 0.507343 0.71228
## hhmb 0.005549 0.07449 -0.85
## Residual 3.139760 1.77194
## Number of obs: 35456, groups: cntry, 19
##
## Fixed effects:
## Estimate Std. Error t value
## (Intercept) 7.22076 0.16574 43.566
## aesfdrkSafe -0.36364 0.02302 -15.799
## aesfdrkUnsafe -0.82087 0.03140 -26.139
## aesfdrkVery unsafe -1.16743 0.05234 -22.305
## hhmb 0.17250 0.01856 9.295
## gndrFemale 0.20586 0.01947 10.571
##
## Correlation of Fixed Effects:
## (Intr) asfdrS asfdrU asfdVu hhmb
## aesfdrkSafe -0.085
## aesfdrkUnsf -0.062 0.524
## asfdrkVryun -0.037 0.318 0.284
## hhmb -0.821 0.013 0.020 0.015
## gndrFemale -0.044 -0.144 -0.219 -0.160 0.003
```

Ok, notice how now in the random effects part of the output we get a standard deviation for 'hhmb', that is the random slope representing the between country variability in the effect of 'hhmb'. Notice as well that now we have a correlation term. This is indicating the correlation between the random intercepts and random slopes ($\hat{\rho}_{u_{0j}u_{1j}}$). The correlation is negative and pretty strong, which points at how, on average, in those countries where average happiness is higher, the positive effect of cohabitation is weaker, and vice versa, in countries with average low happiness, the effect of cohabitation is stronger (more positive).

However, to know for sure whether we can claim this we need to test whether the addition of this random slope (together with its correlation with the random intercepts) is significant. To test that we can run a likelihood ratio test using the command *lrtest* within the *lmtest* package. Remember that we dropped 'sclmeet' in our last random slopes model, so to ensure that the test compares truly nested models we need to estimate a new random intercepts model without 'sclmeet'.

```

RI = lmer(happyrec ~ aesfdrk+hhmmb+gndr+(1|cntry), data=ess)
library(lmtest)
lrtest(RI, RS)

## Likelihood ratio test
##
## Model 1: happyrec ~ aesfdrk + hhmmb + gndr + (1 | cntry)
## Model 2: happyrec ~ aesfdrk + hhmmb + gndr + (1 + hhmmb | cntry)
##   #Df LogLik Df  Chisq Pr(>Chisq)
## 1    8 -70715
## 2   10 -70667  2 97.065 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

The extension of the random intercepts to a random slopes model is statistically significant. Therefore, we can conclude that the effect of cohabitation on happiness is not uniform across countries. However, we need to make an additional effort at interpreting our results to ascertain not only statistical significance but also substantive significance (effect size). In our case, whether the variability in the effect of ‘hhmmb’ seems meaningful. We should always do this, but especially when using samples as big as the one we have here, since the bigger the sample the more likely that we will find results to be statistically significant.

To assess the effect size of these random slopes (\hat{u}_{1j}), we can put them in context by comparing them with the fixed effect of ‘hhmmb’ ($\hat{\beta}_{hhmmb}$). We can do this by estimating a 95% confidence interval representing the between country variability in the effect of ‘hhmmb’, ($\hat{\beta}_{hhmmb} - 1.96\hat{\sigma}_{u1}, \hat{\beta}_{hhmmb} + 1.96\hat{\sigma}_{u1}$).

```
0.172-1.96*0.074 #lower bound
```

```
## [1] 0.02696
```

```
0.172+1.96*0.074 #upper bound
```

```
## [1] 0.31704
```

According to our confidence interval the effect of cohabitation is always positive, but given the relative width of the confidence Interval (0.027, 0.317), we can see that it varies vastly across countries. We can see how the effect can be more than ten times stronger in some countries than others. So, yes, this is statistically AND substantively significant. This result illustrates very well the relevance of our research question, and the importance of context and culture when modelling behavioural and attitudes data.

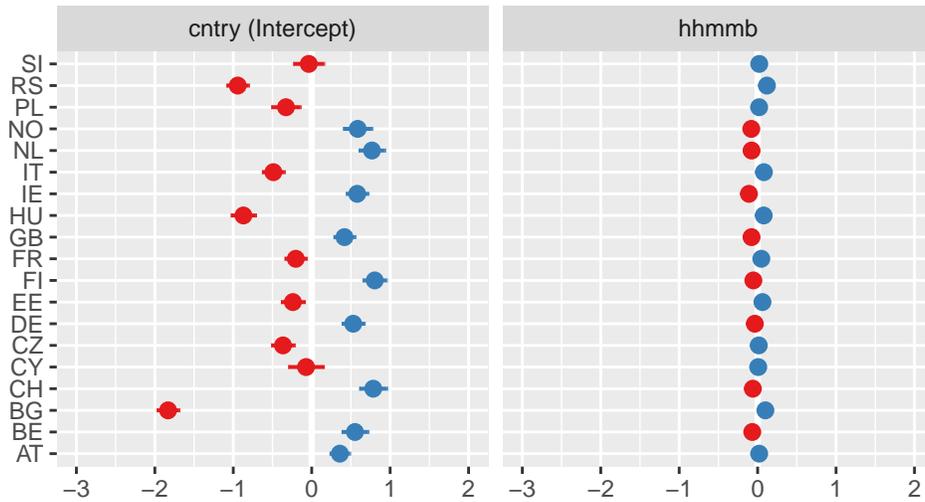
As before, we can find, and plot, the specific random slopes using *ranef*, *coef*, and *plot_model*.

```

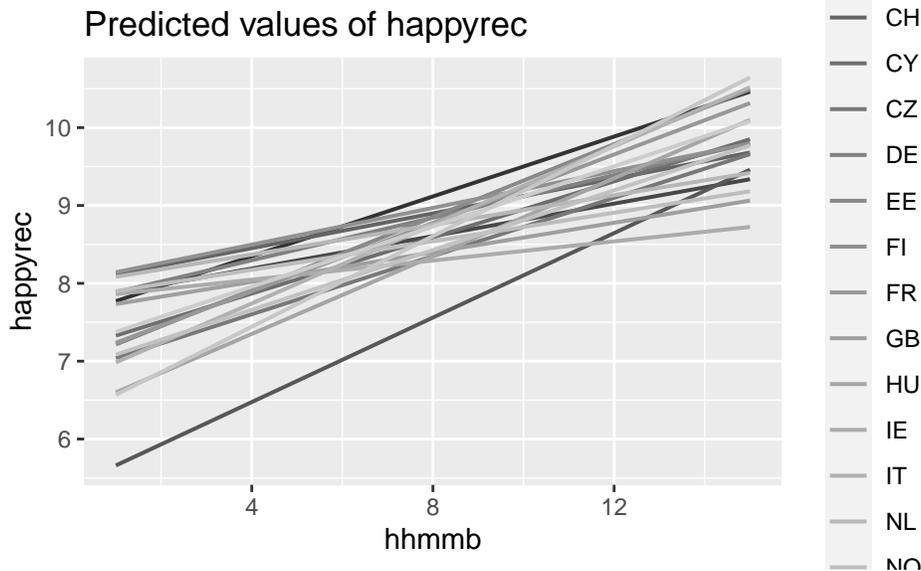
ranef(RS)
plot_model(RS, type="re")

```

Random effects



```
coef(RS)
plot_model(RS, type="pred", terms=c("hhmb", "cntry"), pred.type="re", colors="gs", ci.lvl=NA)
```



Notice how in the line charts depicting the different effects of ‘hhmb’ across countries, the between country variability at $hhmb = 0$ (the intercept) is wider than what we observe in the right-hand side part of the plot. That is, the graph seems to be ‘funneling in’. This is in accordance with the previously noted negative correlation between the random intercepts and random slopes terms ($\hat{\rho}_{u_{0j}u_{1j}} = -0.85$).

To recap, when presented with hierarchical data we can opt for one of three approaches, robust standard errors using the sandwich estimator, fixed effects models including the cluster units as dummy variables, or multilevel modelling. If we care most about precision in the estimation of the standard errors we should use the sandwich estimator. This approach is also interesting because (unlike multilevel modelling) it does not rely on additional assumptions, and because it adjusts for lack of independence in our residuals (within cluster correlation, autocorrelation, etc.), but also for heteroskedasticity. If we want to control for potential cluster level confounders, and in so doing obtain more accurate estimates of our regression coefficients we should use FEs. However, this approach might end up rendering an overfitted model if the number of cluster units is too big. The advantage of multilevel modelling lies on its capacity to model the unexplained cluster-level

variability, which as we have just seen can help us explore some important research questions. Incidentally, if properly specified, multilevel models can adjust for problems of within cluster correlation affecting the assumption of independence.

Exercise 2. Consistency in Sentencing

In this exercise you are requested to use sentencing data and multilevel modelling to explore two research questions: Is the sentencing practice consistent across courts? And, are the aggravating and mitigating factors *intoxication* and *remorse* applied consistently across courts? These two same questions were explored in Pina-Sánchez & Linacre (2013) using similar multilevel models to the ones you will employ here.

-Start by opening the data, which can be downloaded from Minerva. This is a trimmed version of the assault offences captured by the Crown Court Sentencing Survey (CCSS), with assault causing bodily harm taken as the reference category.

-Then take a quick look at the variables in it running some univariate stats. You might want to drop the missing cases in 'Sentencing.court', for which you could use the command `complete.cases` as we did in Workshop 2.

-To explore the first question you should use a random intercepts model.

-Since the only sentence outcome available in the dataset is whether a custodial sentence was imposed you are going to have to use binary logit multilevel models, which can be requested changing `lmer` with a new command, `glmer`, and adding `family="binomial"` as an option.

-When interpreting the results from your logit multilevel models notice that, just like for the case of a binary logit model, the level-1 residuals are not normally distributed. Hence, we will only be able to interpret the variability at level-2. To help you assess whether the between court variability seems meaningful or not you can use the `ranef` command, which you might first transform using `exp` so the random effects reflect the odds ratios of incarceration in each court, compared to a hypothetical average court. To do so use the following combination of commands, `exp(ranef(name.model)$Sentencing.court)`, changing 'name.model' by the name you used for your random intercepts model.

-To explore the second research question you will need to estimate random slopes models. I recommend that to keep the estimation process as simple as possible you replicate your random intercepts model by including, first, a random slope for under the influence and alcohol and drugs ('Aggravating.72'), and then, another model where that random slope is substituted by another one for remorse ('Mitigating.36').

-To test whether each of those two random slopes are statistically significant you can compare each of their model with the random intercepts model using `lrtest`.

#####To be removed from the handout#####

Opening the dataset.

```
ccss = read.csv("CCSS_trimmed.csv", stringsAsFactors = TRUE)
```

Some exploratory analysis and further trimming down the dataset to remove missing cases.

```
summary(ccss)
ccss = ccss[complete.cases(ccss),]
```

We can explore the first research question using a random intercepts model. Since we want the random intercepts model to be capturing unwarranted disparities across courts we should control for as many factors considered in the guidelines as possible.

```
RI = glmer(custody ~ common+intent+GBH+affray+disorder+Mitigating.36+Aggravating.72+
           (1|Sentencing.court), family="binomial", data=ccss)
summary(RI)
```

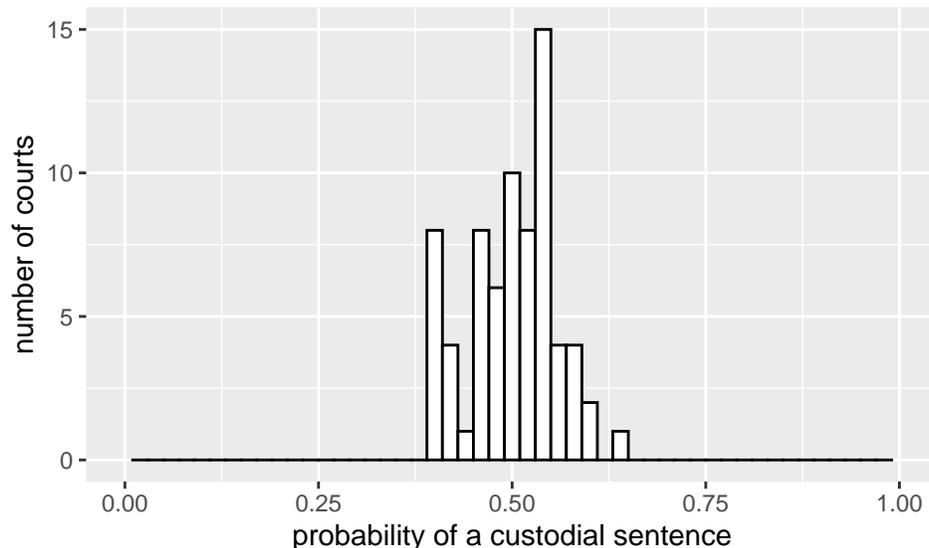
With $\hat{\sigma}_u = 0.326$ the random intercepts term looks quite relevant, pointing at substantial between court variability in determining custodial sentences for similar cases of assault. To be able to see more clearly the dimension of these between court disparities we can estimate and transform the random intercepts for each court. To do so we can use the *ranef* command and the following formula to express log(odds) as probabilities,

$$\frac{e^{(\hat{\beta}_0 + \hat{u}_j)}}{1 + e^{(\hat{\beta}_0 + \hat{u}_j)}}$$

```
exp(ranef(RI)$Sentencing.court)
prob = exp(ranef(RI)$Sentencing.court)/(1+exp(ranef(RI)$Sentencing.court))
summary(prob)
```

The average probability of custody across courts ranges from 0.390 to 0.632. This is remarkable, however, the range is not a great measure of disparities since it can be heavily affected by the presence of outliers. To get a more representative view of the between court disparities we can plot them.

```
names(prob) = "intercept"
library(ggplot2)
ggplot(prob, aes(x=intercept)) + xlim(0,1) +
  geom_histogram(binwidth=0.02, color="black", fill="white") +
  labs(x = "probability of a custodial sentence", y = "number of courts")
```



This histogram is perhaps the best way to provide an empirical answer to our first research question, and from the look of it, I would say yes, the presence of between court disparities is not negligible. The sentencing practice does not seem to be consistent across the jurisdiction of England and Wales.

Let's now look into the second research question on whether the aggravating factor 'under the influence of alcohol and drugs', and the mitigating factor 'genuine remorse' are applied uniformly across courts. There is plenty of qualitative and anecdotal evidence suggesting this is not the case. Let's explore that now using a representative sample of offences to assess whether those suspicions can be confirmed or not. We can extend the random intercepts model that we have just used with two different random slopes models, each exploring the variability in the effect of the two factors to be considered.

```
RS_remorse = glmer(custody ~ common+intent+GBH+affray+disorder+Mitigating.36+Aggravating.72+
  (1+Aggravating.72|Sentencing.court), family="binomial", data=ccss)
summary(RS_remorse)
lrtest(RI, RS_remorse)
```

```
RS_alcohol = glmer(custody ~ common+intent+GBH+affray+disorder+Mitigating.36+Aggravating.72+
(1+Mitigating.36|Sentencing.court), family="binomial", data=ccss)
summary(RS_alcohol)
lrtest(RI, RS_alcohol)
```

We do not find evidence that any of these two controversial factors are being used inconsistently across courts. It is possible that some of the concerns raised by the different sentencing scholars pointing at this problem are unfounded, perhaps due to a problem of selection bias, or ‘cherry picking’ evidence, which plagues much of the research based on anecdotal evidence.