Homework 2

Multilevel Modeling

Langdon Holmes

Question 1

```
library(haven) # read .sav file
library(lme4)
library(stargazer) # LaTeX tables
library(performance) # ICC
df <- read_sav("../data/mlm-homework-2.sav")
```

Use the "build-up" stepwise strategy of model building to construct a model for the educational data, using language test scores (LANGPOST) as the dependent variable and (potentially, depending on how it goes) percentage minority (PERCMINO) and SES (SES) as predictors. To anchor your analysis, use the null model (random effects ANOVA) as your simplest model.

```
# we can use the constant 1 as a predictor to specify a null model
up.null <- lme4::lmer(langpost ~ 1 + (1|schoolnr), data = df)
cat("An adjusted ICC of", round(icc(up.null)$ICC_adjusted, 2), "indicates that MLM is appr
```

```
An adjusted ICC of 0.23 indicates that MLM is appropriate.
```

```
up.1 <- lmer(langpost ~ 1 + percmino + (1|schoolnr), data = df)
up.2 <- lmer(langpost ~ ses + percmino + (1|schoolnr), data = df)
up.3 <- lmer(langpost ~ ses + percmino + (ses|schoolnr), data = df)
up.4 <- lmer(langpost ~ ses*percmino + (ses|schoolnr), data = df)
```

```
stargazer(up.null, up.1, up.2, up.3, up.4, header=FALSE,
          column.labels = c("Null", "+percmino", "+ses", "+ses slope", "+ses*percmino")
            )
```

Table 1

	Dependent variable:				
	langpost				
	Null	+percmino	+ses	+ses slope	+ses*percmino
	(1)	(2)	(3)	(4)	(5)
ses			0.295^{***}	0.300^{***}	0.280^{***}
			(0.017)	(0.018)	(0.021)
percmino		-0.147^{***}	-0.096^{***}	-0.087^{***}	-0.154^{***}
		(0.028)	(0.027)	(0.028)	(0.044)
ses:percmino					0.003^{**}
					(0.001)
Constant	40.362^{***}	41.467^{***}	33.038^{***}	32.833^{***}	33.379^{***}
	(0.428)	(0.441)	(0.640)	(0.728)	(0.780)
Observations	2,287	2,287	2,287	2,287	2,287
Log Likelihood	−8,126.541	−8,116.429	−7,978.204	−7,974.398	−7,978.109
Akaike Inf. Crit.	16,259.080	16,240.860	15,966.410	15,962.800	15,972.220
Bayesian Inf. Crit.	16,276.290	16,263.800	15,995.080	16,002.940	16,018.100

Note: *p<0.1; **p<0.05; ***p<0.01

I would select the third model, which includes socio-economic status of the student as a level-1 predictor and percentage minority of the school as a level-2 predictor. There is also a random intercept for school. While the interaction between socioeconomic status and percentage minority was significant, it reduced overall model fit as measured by AIC. Adding a random slope for socioeconomic status had little to no effect on model fit. Compared to model 3, the random slope resulted in a slight improvement in Log Likelihood and AIC, but a slight deterioration in BIC. Unless this slope was critical to my research question, I would prefer to use model 3, which is more parsimonious.

Question 2

Now use the "tear down" stepwise strategy. To anchor your analysis, use as the most complex model one with random intercepts, in which SES serves as a level-1 predictor with random slopes and PERCMINO as a level-2 predictor of both intercepts and slopes. Do you settle on the same model as in #1? (I'm not leading you; I really don't know!) [15]

```
down.1 <- lmer(langpost ~ ses*percmino + (ses|schoolnr), data = df)
down.2 <- lmer(langpost ~ ses + percmino + (ses|schoolnr), data = df)
down.3 <- lmer(langpost ~ ses + percmino + (1|schoolnr), data = df)
down.4 <- lmer(langpost ~ ses + (1|schoolnr), data = df)
```

```
stargazer(down.1, down.2, down.3, header=FALSE,
          column.labels = c("Max", "-ses*percmino", "-ses slope", "-percmino")
          )
```

In this case, I would again choose the model with a fixed slope because I do not think that the minor improvements in model fit would justify the inclusion of additional parameters The Bayesian information criterion supports my approach in both cases. This makes sense, since the BIC penalizes model complexity, which feels appropriate for an exploratory analysis such as this.

Question 3

Now fit the model with a random intercept and a random slope. Use the website to plot and interpret the significant cross-level interaction effect. Leave the "df" boxes blank, and remember that the web page does not understand scientific notation (i.e., if you see 0.193383E-02, enter 0.00193383 instead). If Rweb is not working, you can simply copy and paste the generated code directly into R. Include and interpret: [20]

Table 2

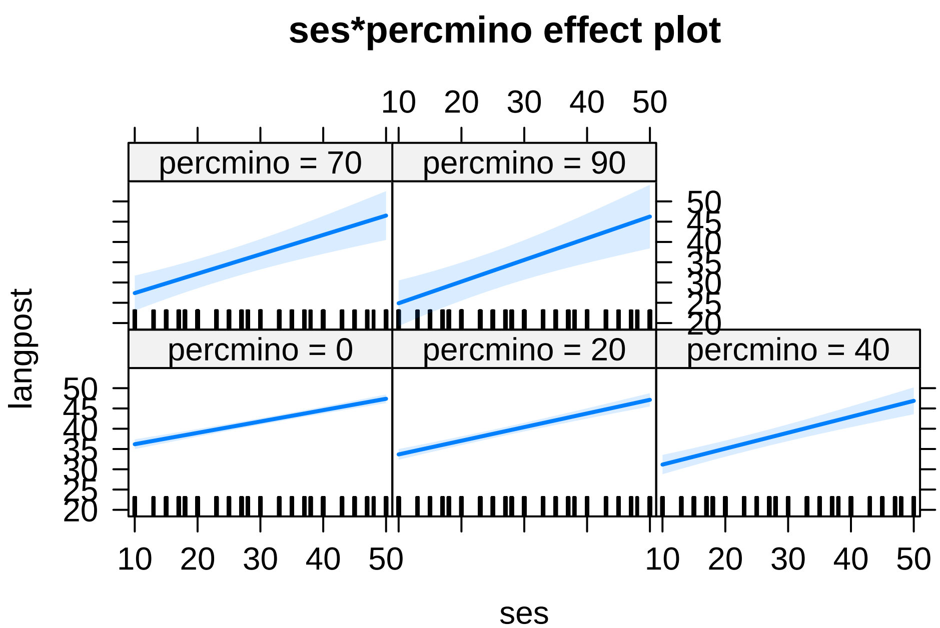
	Dependent variable:		
		langpost	
	Max	-ses*percmino	-ses slope
	(1)	(2)	(3)
ses	0.280^{***}	0.300^{***}	0.295^{***}
	(0.021)	(0.018)	(0.017)
percmino	-0.154^{***}	-0.087^{***}	-0.096^{***}
	(0.044)	(0.028)	(0.027)
ses:percmino	0.003^{**}		
	(0.001)		
Constant	33.379^{***}	32.833^{***}	33.038^{***}
	(0.780)	(0.728)	(0.640)
Observations	2,287	2,287	2,287
Log Likelihood	−7,978.109	−7,974.398	−7,978.204
Akaike Inf. Crit.	15,972.220	15,962.800	15,966.410
Bayesian Inf. Crit.	16,018.100	16,002.940	15,995.080
Note:			*p<0.1; **p<0.05; ***p<0.01

```r
library(effects)
model.1 <- lmer(langpost ~ ses*percmino + (ses | schoolnr), data = df)
```

```
Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
Model failed to converge with max|grad| = 0.0370203 (tol = 0.002, component 1)
```

```
Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, : Model is near
 - Rescale variables?
```

```r
plot(effect(c("ses*percmino"), model.1, KR=T))
```




```r
summary(model.1)
```

```
Linear mixed model fit by REML ['lmerMod']
Formula: langpost ~ ses * percmino + (ses | schoolnr)
   Data: df

REML criterion at convergence: 15956.2
```

```
Scaled residuals:
    Min      1Q  Median      3Q     Max 
-3.4429 -0.6256  0.0843  0.7177  3.0229 

Random effects:
 Groups   Name        Variance  Std.Dev. Corr 
 schoolnr (Intercept) 28.977298 5.38306       
          ses          0.005574 0.07466  -0.88
 Residual             56.521111 7.51805       
Number of obs: 2287, groups:  schoolnr, 131

Fixed effects:
              Estimate Std. Error t value
(Intercept)  33.379351   0.779925  42.798
ses           0.280045   0.020576  13.611
percmino     -0.153779   0.044438  -3.461
ses:percmino  0.002824   0.001427   1.979

Correlation of Fixed Effects:
            (Intr) ses    percmn
ses         -0.857              
percmino    -0.479  0.430       
ses:percmin  0.357 -0.474 -0.777
optimizer (nloptwrap) convergence code: 0 (OK)
Model failed to converge with max|grad| = 0.0370203 (tol = 0.002, component 1)
Model is nearly unidentifiable: very large eigenvalue
 - Rescale variables?
```

```r
# Variances for gammas can be found by squaring the standard error.
print("Covariance of intercept and slope (for tau_10):")
```

```
[1] "Covariance of intercept and slope (for tau_10):"
```

```r
as.matrix(Matrix::bdiag(VarCorr(model.1)))
```

```
            (Intercept)          ses
(Intercept)  28.9772985 -0.354289229
ses          -0.3542892  0.005574296
```

```r
print("Variance Covariance Matrix:")
```

```
[1] "Variance Covariance Matrix:"
```

```r
vcov(model.1)
```

```
4 x 4 Matrix of class "dpoMatrix"
               (Intercept)           ses      percmino   ses:percmino
(Intercept)   0.6082826014 -1.375507e-02 -1.660930e-02  3.979571e-04
ses          -0.0137550698  4.233518e-04  3.929310e-04 -1.391987e-05
percmino     -0.0166092977  3.929310e-04  1.974698e-03 -4.926732e-05
ses:percmino  0.0003979571 -1.391987e-05 -4.926732e-05  2.037613e-06
```

a

Text output (interpret only the "simple intercepts and simple slopes" and "regions of significance" sections).

```
CASE 3 TWO-WAY INTERACTION SIMPLE SLOPES OUTPUT

Your Input
========================================================
  w1(1)        = 0
  w1(2)        = 45
  w1(3)        = 90
  x1(1)        = 0
  x1(2)        = 27
  x1(3)        = 45
  Intercept    = 33.379351
  x1 Slope     = 0.280045
  w1 Slope     = -0.153779
  w1x1 Slope   = 0.002824
  alpha        = 0.05

Asymptotic (Co)variances
========================================================
  var(g00) 0.6082826
  var(g10) 0.00042335
```

```
  var(g01) 0.0019747
  var(g11) 0.00000204
  cov(g00,g01) -0.0166093
  cov(g10,g11) -0.00001392
  cov(g00,g10) -0.01375507
  cov(g01,g11) -0.00004927

Region of Significance on w (level-2 predictor)
========================================================
  w1 at lower bound of region = -11538.7709
  w1 at upper bound of region = -45.6588
  (simple slopes are significant *outside* this region.)

Simple Intercepts and Slopes at Conditional Values of w
========================================================
  At w1(1)...
    simple intercept = 33.3794(0.7799), z=42.7982, p=0
    simple slope     = 0.28(0.0206), z=13.6106, p=0
  At w1(2)...
    simple intercept = 26.4593(1.7641), z=14.9984, p=0
    simple slope     = 0.4071(0.0574), z=7.0907, p=0
  At w1(3)...
    simple intercept = 19.5392(3.6897), z=5.2957, p=0
    simple slope     = 0.5342(0.1201), z=4.4482, p=0

Simple Intercepts and Slopes at Region Boundaries for w
========================================================
  Lower Bound...
    simple intercept = 1807.8(513.1292), z=3.5231, p=0.0004
    simple slope     = -32.3054(16.4808), z=-1.9602, p=0.05
  Upper Bound...
    simple intercept = 40.4007(2.4983), z=16.171, p=0
    simple slope     = 0.1511(0.0771), z=1.9602, p=0.05

Region of Significance on x (level-1 predictor)
========================================================
  x1 at lower bound of region = 33.1067
  x1 at upper bound of region = 3327.9384
  (simple slopes are significant *outside* this region.)

Simple Intercepts and Slopes at Conditional Values of x
========================================================
  At x1(1)...
```

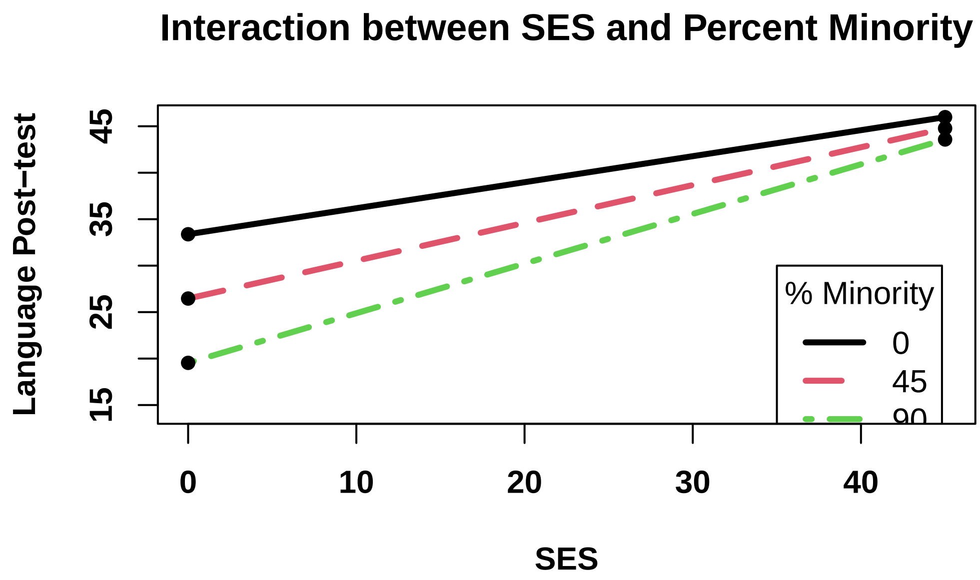
```
    simple intercept = 33.3794(0.7799), z=42.7982, p=0
    simple slope     = -0.1538(0.0444), z=-3.4606, p=0.0005
  At x1(2)...
    simple intercept = 40.9406(0.4173), z=98.1102, p=0
    simple slope     = -0.0775(0.0283), z=-2.7417, p=0.0061
  At x1(3)...
    simple intercept = 45.9814(0.4771), z=96.3791, p=0
    simple slope     = -0.0267(0.0408), z=-0.654, p=0.5132

Simple Intercepts and Slopes at Region Boundaries for x
========================================================
  Lower Bound...
    simple intercept = 42.6507(0.4019), z=106.121, p=0
    simple slope     = -0.0603(0.0308), z=-1.9602, p=0.05
  Upper Bound...
    simple intercept = 965.3519(67.8067), z=14.2368, p=0
    simple slope     = 9.2443(4.716), z=1.9602, p=0.05
```

b

A plot of the simple regression of LANGPOST on SES at three conditional values of PERCMINO: the minimum observed (0%), middle (45%), and maximum observed (90%).

```r
xx <- c(0,45)   #  <-- change to alter plot dims
yy <- c(14.2508,45.9814)   #  <-- change to alter plot dims
leg <- c(35,30)   #  <-- change to alter legend location
x <- c(0,45)   #  <-- x-coords for lines
y1 <- c(33.3794,45.9814)
y2 <- c(26.4593,44.7799)
y3 <- c(19.5392,43.5785)
plot(xx,yy,type='n',font=2,font.lab=2,xlab='SES',ylab='Language Post-test',main='Interacti
lines(x,y1,lwd=3,lty=1,col=1)
lines(x,y2,lwd=3,lty=5,col=2)
lines(x,y3,lwd=3,lty=6,col=3)
points(x,y1,col=1,pch=16)
points(x,y2,col=1,pch=16)
points(x,y3,col=1,pch=16)
legend(leg[1],leg[2],legend=c('0','45','90'),lwd=c(3,3,3),lty=c(1,5,6),col=c(1,2,3), title
```



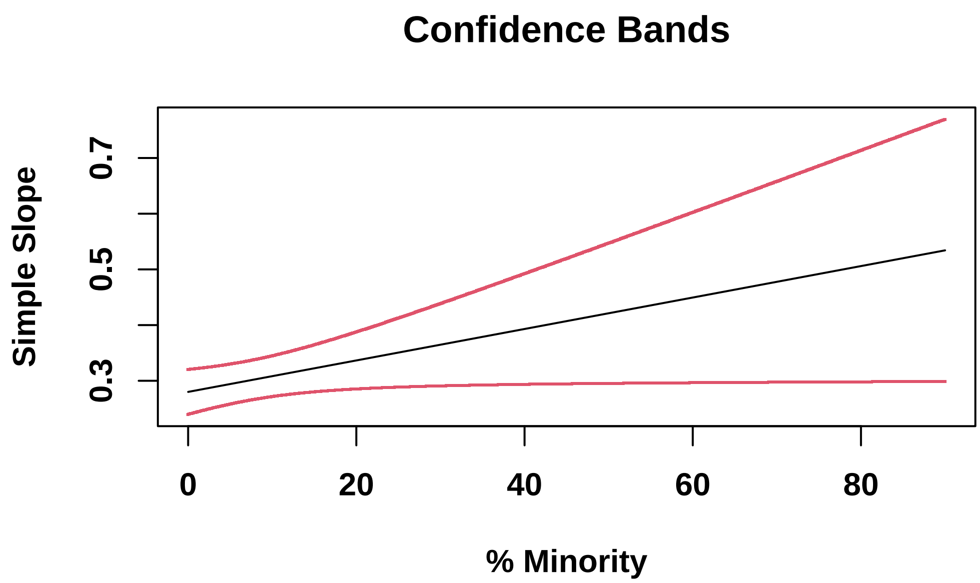
This plot illustrates how the effect of SES on post-test scores is moderated by the minority composition of the school. The socioeconomic status of students attending schools with a lower percentage of minority students has a less pronounced effect on language post-test scores. Students attending schools with a higher percentage of minority students perform less well across the board, but this effect is less pronounced for high socioeconomic status students at these schools. In practical terms, the (negative) effect of attending a high minority percentage school is more pronounced for low SES students.

c

A plot of the confidence bands around the simple slope of LANGPOST regressed on SES. The x-axis of this plot should extend from the minimum to maximum observed values of PERCMINO.

```
z1=0  #supply lower bound for w1 here
z2=90   #supply upper bound for w1 here
z <- seq(z1,z2,length=1000)
fz <- c(z,z)
y1 <- (0.280045+0.002824*z)+(1.9602*sqrt(0.0004233518+(2*z*-0.00001391987)+((z^2)*0.000002
y2 <- (0.280045+0.002824*z)-(1.9602*sqrt(0.0004233518+(2*z*-0.00001391987)+((z^2)*0.000002
fy <- c(y1,y2)
```

```
fline <- (0.280045+0.002824*z)
plot(fz,fy,type='p',pch='.',font=2,font.lab=2,col=2,xlab='% Minority',ylab='Simple Slope',
lines(z,fline)
f0 <- array(0,c(1000))
lines(z,f0,col=8)
abline(v=-11538.7709,col=4,lty=2)
abline(v=-45.6588,col=4,lty=2)
```



The confidence band indicates a 95% confidence interval for the estimate of the effect of SES on post-test scores. At lower values of percentage minority, the slope is increasing. At the highest values of percentage minority, the confidence bands extend to a zero or near-zero slope, indicating that we should be less confident about the slope estimate for schools with a high minority percentage. The confidence band is roughly symmetrical, so it is equally likely that the slope is even steeper for schools with a high minority percentage. In practical terms, we can be more confident about the effect of SES on language post-test scores for schools with a lower percentage of minority students.

Question 4

Einstein allegedly claimed that you never really know a subject until you can explain it to your grandmother. Please pretend I am your grandmother, and that I just asked you what group mean centering and grand mean centering are. Explain these concepts to the best of your ability. Assume that your "grandmother" has no quantitative training, speaks fluent English, and is genuinely curious. [10]