Stat 414 - Day 18
Level 1 Predictors (Section 4.4)

Last Time: Adding a Level 1 variable

- Creates a model with parallel lines for each level 2 unit (intercepts following a normal distribution with some shrinkage)
  - intercept is the (predicted or expected) (mean) response for the "average" level 2 unit
  - slope is the (estimated or expected) change in response holding level 2 unit fixed
- Level 1 variables (x_{ij}) are expected to explain Level 1 variation (lower σ). Can also explain Level 2 variation if the distribution of the variable changes across the Level 2 units
  - Assess statistical significance with either t-tests/partial F-tests or LRT test
  - Interpretation of adjusted intraclass correlation coefficient (is grouping still relevant after conditioning on covariate)

Example 1: Netherlands language scores cont.

```
nullmodel = lmer(langPOST ~ 1 + (1|schoolnr), data = neth, REML = FALSE)
model1 = lmer(langPOST ~ 1 + IQ_verb + (1|schoolnr), data = neth, REML=F)
```

(a) What percentage of the Level 1 variance was explained by verbal IQ? ("Pseudo-R^2")

(b) What percentage of the Level 2 variance was explained by verbal IQ?

(c) What percentage of the total variance was explained by verbal IQ?

Note: We can think of ICC as the proportion of total variance explained by the grouping variable and R^2 as the proportion explained by the fixed effects. The difference between adjusted/unadjusted ICC is whether you take into account the "variance explained by the fixed effects" in the denominator. The difference between conditional and marginal R^2 is whether you account for the random effects in the numerator (conditional - marginal = unadjusted ICC).

(d) Is the amount of variability explained statistically significant?

```
summary(model1)
anova(nullmodel, model1)
```

Example 2: Radon (note: this data file may not match the one from HW)

Radon comes from underground and can enter more easily when a house is built into the ground (i.e., has a basement). In this dataset (for 919 homes across 85 counties in MN), *floor* indicates whether the measurement was taken in the basement or the first floor, and *basement* indicates whether the house had a basement. County level data was merged with information about the houses, at both the state and county level.

```
model0 <- lmer(logradon ~ 1 + (1 | newfips), data = mergedmn); summary(model0)
```

(a) What is the estimated mean (log) radon ?

(b) What is the basemeas variable about?

(c) Do you predict higher or lower radon levels if the house has a basement and the measurement was taken in the basement? What if the house doesn't have a basement?

(d) Check your predictions

```
model1 <- lmer(logradon ~1 + basemeas + (1 | newfips), data = mergedmn)
```

(e) What do the estimated random effects $\hat{u}_j$ represent in this model?

(f) How does adding this variable to the model change the variance components?

(g) Let's explore Roseau county

Notes:

- Marginal R^2 measures the variance explained by the fixed effects as a proportion of the sum of all the variance components ($\hat{\sigma}^2 + \hat{\tau}^2 + var(X\beta)$ ("The fixed effects explain …")
- Conditional R^2 measures the variance explained by both the fixed and random effects in the model. ("The fixed and random effects explain …")
    - The unadjusted ICC is the difference between these! (the contribution of the random effects…)
- There is still some lack of agreement ("the literature does not seem to have converged onthis topic") in how to calculate R^2 values for these models as the formulas provided here can actually turn out to be negative!*