21.1 Variable Uncertainty

Growth rates of children vary. By gathering data we can estimate the variation.

$$\sigma_0^2, \text{intercept } \sigma_1^2, \text{slope}$$

Estimates are uncertain. Large sample sizes provide more accurate estimates.

$$\widehat{\sigma}_0^2 \rightarrow \sigma_0^2$$

and we get better estimates of β_0, β_1.
Larger sample size does not reduce the variance of a random effect. σ_0^2 does not go to 0.
Read about "value-added" models.

21.2 Super-Pop vs finite-Pop

b_{0j} and b_{1j} are one kid's intercept and growth rate.
σ_0^2 and σ_1^2 describe their variation in the super-population of all kids.
s_0^2 and s_1^2 describe variation in the sample of kids. (Not the same quantity.)
In JAGs we have defined parameters `sigma.0` and `sigma.1`. We could compute within sample SD with

```
s.0 <- sd(y - y.hat)
```

(JAGs) BUGs need to define residuals in a loop.

21.3 Contrasts

How to handle discrete inputs?
Classical: Use 'county' as factor, or use uranium as county level predictor, not both.
(or two-stage modeling)
Multilevel: can do both.
AP Exam scores: 1, 2, 2.5, 3, 3.5, 4, 4.5, 5
as predictor for grade in Calculus class (as GPA coded number).
Multilevel: $\alpha_j = \gamma_0 + \gamma_1 u_j + \eta_j$, $\eta_j \sim N(0, \sigma_\alpha^2)$.

Super or finite population?

only $J = 10$ groups gives wide spread to $\widehat{\gamma}_1$.

MCMC trick
go back to saved "alpha"s and for each MCMC draw: do a regression on u_j's, look at that distribution.

Similarly, we could want to look at contrasts between particular airports used for flight simulator.
Trick: add back in the average to get to original scale.

21.4 Average Predictive Comparisons

Inputs vs Linear Predictors in logistic regression:
Inputs, u: just basic columns like age, gender.
Linear predictors: age and gender, but also v could include age^2 and the interaction.
How does predicted y change for a 1 unit change in u holding v constant?
Relative change: $b_u(u^{(lo)}, u^{(hi)}, v, \theta) = \frac{E(y|u^{(hi)},v,\theta) - E(y|u^{(lo)},v,\theta)}{u^{(hi)} - u^{(lo)}}$
Avg predictive diff: $B_u(u^{(lo)}, u^{(hi)}) = \frac{1}{n}\sum_{i=1}^n b_u(u^{(lo)}, u^{(hi)}, v_i, \theta)$

Example p 467

What does it mean to change u and hold v fixed?
Change gender, but keep gender by age fixed?
Change age, but keep age by gender fixed?
No, but we can average over observed inputs to get the average effect of going from girl to boy at age $= 10$.
Avg Predictive difference is a better gauge of the shift.
Just using average of v's can be misleading.

21.5 R^2

Using V as "variance operator", for one-level:

$$R^2 = 1 = \frac{V_{i=1}^n \epsilon_i}{V_{i-1}^n y_i}$$

is explained variance. But in mixed models, R^2 is a random variable.
Split up each level to get a fit and residual: $\theta_k = \hat{\theta}_k + \epsilon_k$, with superscript m for level (drop them).

$$R^2 = 1 - \frac{E(V_{i=1}^n \epsilon_i)}{E(V_{i-1}^n theta_i)}$$

Use average over MCMC draws to get E's.
R^2 at each level gives proportion of explained variance **at this level**. Adding an individual predictor might increase R^2 at bottom level and hurt it at group level.

Classical R^2

$$R^2 = 1 - \frac{E(V_{i=1}^n \epsilon_i)}{E(V_{i-1}^n theta_i)} = 1 - \frac{y^t(I-H)y}{y^t I_c y}$$

With flat prior on $(\beta, \log(\sigma))$, this is almost R^2_{adj}.
Compute in MCMC loop by saving errors at each level:

```
## In JAGs
e.y <- y - y.hat ## in likelihood step (jags avoids loop)
E.B <- B - B.hat ## in group level loop
## save e.y and E.B
## in R
rsquared.y <- 1 - mean(apply(e.y,1,var))/var(y)
rsquared.a <- 1 - mean(apply(E.B[, , 1],1, var))/ mean(apply(a, 1, var)
rsquared.y <- 1 - mean(apply(E.B[, , 2],1, var))/ mean(apply(b, 1, var)
```

21.6 Partial Pooling

Pooling is a precision-weighted average of overall mean and $\overline{y}_j$.

$$\text{Pooling factor} = \text{Weight on 'prior mean'}, \ \omega_j = \frac{\sigma_\alpha^{-2}}{\sigma_\alpha^{-2} + n_j/\sigma_y^2}$$

Posterior variance: $\text{var}(\alpha_j) = (\sigma_\alpha^{-2} + n_j/\sigma_y^2)^{-1}$
Larger sample size gives less pooling.
Compute ω_j using MCMC draws based on $\epsilon_j = \alpha_j - \hat{\alpha}_j$ and $\omega_j = \frac{\text{SE}(\epsilon_j)^2}{\sigma_\alpha^2}$.
Save `e.a[j] <- a[j] - a.hat[j]` for each draw, and in R compute a vector of `omega[j]`'s

```
omega <- pmin( sd(e.a)^2/ sigma.a^2, 1)
```

See figure 21.8. Note average pooling at a level is similar to R^2.

$$\lambda = 1 - \frac{V_{k=1}^K E(\epsilon_k)}{E\left(V_{k=1}^K \epsilon_k\right)}$$

21.7 Another Variable Can Increase s

Radon data with no covariates has mean α_j for each county.
Estimates: $\hat{\sigma}_y = 0.80$, $\hat{\sigma}_\alpha = 0.12$.
Adding floor effects, βx_i to the model we get
Estimates: $\hat{\sigma}_y = 0.76$, $\hat{\sigma}_\alpha = 0.16$. Counties with more `floor = 0` had larger α's, masking variation.
Soil type is a confounder?

21.8 Multiple Comparisons

Meta-analysis

Randomized experiment of EMF on chick brains using a range of frequencies.

Original presented separate tests for each frequency with groups of frequencies flagged as 'more dangerous'.

Gelman combines to get a multilevel model, pooling effects toward 0.1.