

STA 610L: MODULE 2.10

RANDOM EFFECTS ANOVA (BAYESIAN ESTIMATION II)

DR. OLANREWaju MICHAEL AKANDE

CHALLENGE TO VALIDITY: HETEROGENEOUS MEANS AND VARIANCES

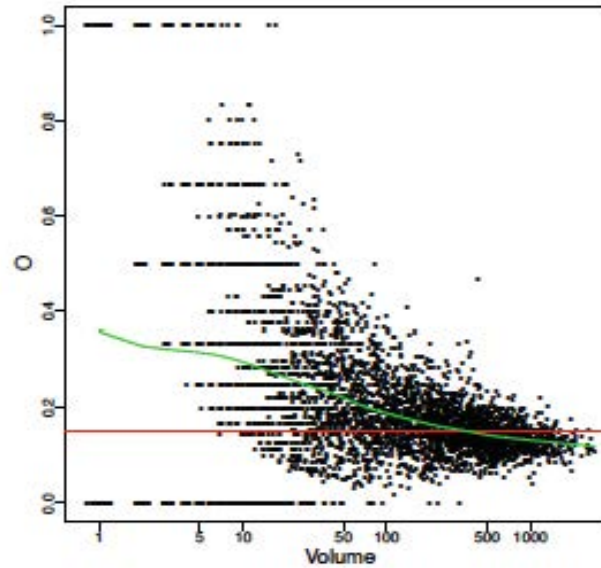
We have looked at the hierarchical normal model of the form

$$\begin{aligned} y_{ij} | \mu_j, \sigma^2 &\sim \mathcal{N}(\mu_j, \sigma^2); \quad i = 1, \dots, n_j \\ \mu_j | \mu, \tau^2 &\sim \mathcal{N}(\mu, \tau^2); \quad j = 1, \dots, J. \end{aligned}$$

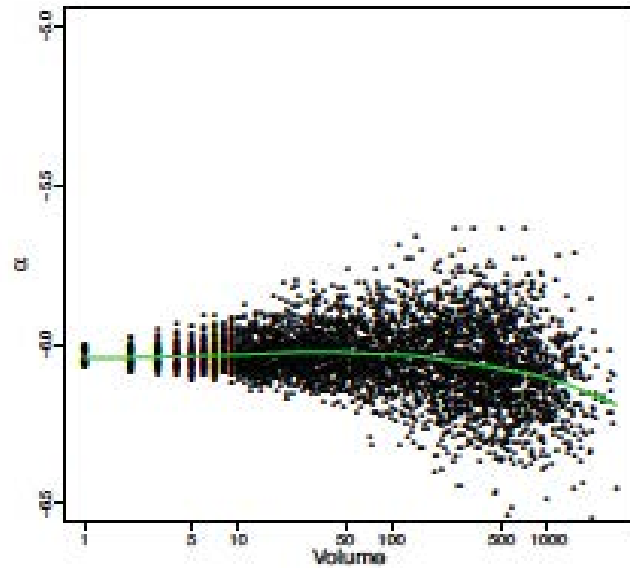
The model gives us an extra hierarchy through the prior on the means, leading to sharing of information across the groups, when estimating the group-specific means.

While many people feel that shrinkage can "do no harm," it can be quite detrimental when the shrinkage target is not correctly specified.

MORTALITY BY VOLUME



ESTIMATED RANDOM INTERCEPTS BY VOLUME



GROUP-SPECIFIC VARIANCES

How might we specify a model to avoid these problems? We could introduce predictors to model group means and or group variances.

$$\alpha_j \sim N(\mu_j(z), \tau_j^2(z))$$

Another potential challenge is that the variance of the response may not be the same for each group anyway. This could be due to a variety of factors.

One potential remedy for this issue is to allow the error variance to differ across groups. A natural extension is

$$\sigma_1^2, \dots, \sigma_J^2 | \nu_0, \sigma_0^2 \sim \mathcal{IG} \left(\frac{\nu_0}{2}, \frac{\nu_0 \sigma_0^2}{2} \right)$$

POSTERIOR INFERENCE

- The full posterior is now:

$$\begin{aligned}\pi(\mu_1, \dots, \mu_J, \sigma_1^2, \dots, \sigma_J^2, \mu, \tau^2, \nu_0, \sigma_0^2 | Y) &\propto p(y | \mu_1, \dots, \mu_J, \sigma_1^2, \dots, \sigma_J^2, \mu, \tau^2, \nu_0, \sigma_0^2) \\ &\times p(\mu_1, \dots, \mu_J | \sigma_1^2, \dots, \sigma_J^2, \mu, \tau^2, \nu_0, \sigma_0^2) \\ &\times p(\sigma_1^2, \dots, \sigma_J^2 | \mu, \tau^2, \nu_0, \sigma_0^2) \\ &\times \pi(\mu, \tau^2, \nu_0, \sigma_0^2)\end{aligned}$$

$$\begin{aligned}&= p(y | \mu_1, \dots, \mu_J, \sigma_1^2, \dots, \sigma_J^2) \\ &\times p(\mu_1, \dots, \mu_J | \mu, \tau^2) \\ &\times p(\sigma_1^2, \dots, \sigma_J^2 | \nu_0, \sigma_0^2) \\ &\times \pi(\mu) \cdot \pi(\tau^2) \cdot \pi(\nu_0) \cdot \pi(\sigma_0^2)\end{aligned}$$

$$\begin{aligned}&= \left\{ \prod_{j=1}^J \prod_{i=1}^{n_j} p(y_{ij} | \mu_j, \sigma_j^2) \right\} \\ &\times \left\{ \prod_{j=1}^J p(\mu_j | \mu, \tau^2) \right\} \\ &\times \left\{ \prod_{j=1}^J p(\sigma_j^2 | \nu_0, \sigma_0^2) \right\} \\ &\times \pi(\mu) \cdot \pi(\tau^2) \cdot \pi(\nu_0) \cdot \pi(\sigma_0^2)\end{aligned}$$

FULL CONDITIONALS

- Notice that this new factorization won't affect the full conditionals for μ and τ^2 from before, since those have nothing to do with all the new σ_j^2 's.
- That is,

$$\pi(\mu | \dots\dots\dots) = \mathcal{N}(\mu_n, \gamma_n^2) \quad \text{where}$$

$$\gamma_n^2 = \frac{1}{\frac{J}{\tau^2} + \frac{1}{\gamma_0^2}}; \quad \mu_n = \gamma_n^2 \left[\frac{J}{\tau^2} \bar{\theta} + \frac{1}{\gamma_0^2} \mu_0 \right],$$

and

$$\pi(\tau^2 | \dots\dots\dots) = \mathcal{IG} \left(\frac{\eta_n}{2}, \frac{\eta_n \tau_n^2}{2} \right) \quad \text{where}$$

$$\eta_n = \eta_0 + J; \quad \tau_n^2 = \frac{1}{\eta_n} \left[\eta_0 \tau_0^2 + \sum_{j=1}^J (\mu_j - \mu)^2 \right].$$

FULL CONDITIONALS

- The full conditional for each μ_j , we have

$$\pi(\mu_j | \mu_{-j}, \mu, \sigma_1^2, \dots, \sigma_J^2, \tau^2, Y) \propto \left\{ \prod_{i=1}^{n_j} p(y_{ij} | \mu_j, \sigma_j^2) \right\} \cdot p(\mu_j | \mu, \tau^2)$$

with the only change from before being σ_j^2 .

- That is, those terms still include a normal density for μ_j multiplied by a product of normals in which μ_j is the mean, again mirroring the previous case, so you can show that

$$\pi(\mu_j | \mu_{-j}, \mu, \sigma_1^2, \dots, \sigma_J^2, \tau^2, Y) = \mathcal{N}(\mu_j^*, \tau_j^*) \quad \text{where}$$

$$\tau_j^* = \frac{1}{\frac{n_j}{\sigma_j^2} + \frac{1}{\tau^2}}; \quad \mu_j^* = \tau_j^* \left[\frac{n_j}{\sigma_j^2} \bar{y}_j + \frac{1}{\tau^2} \mu \right]$$

HOW ABOUT WITHIN-GROUP VARIANCES?

- Before we get to the choice of the priors for ν_0 and σ_0^2 , we have enough to derive the full conditional for each σ_j^2 . This actually takes a similar form to what we had before we indexed by j , that is,

$$\pi(\sigma_j^2 | \sigma_{-j}^2, \mu_1, \dots, \mu_J, \mu, \tau^2, \nu_0, \sigma_0^2, Y) \propto \left\{ \prod_{i=1}^{n_j} p(y_{ij} | \mu_j, \sigma_j^2) \right\} \cdot \pi(\sigma_j^2 | \nu_0, \sigma_0^2)$$

- This still looks like what we had before, that is, products of normals and one inverse-gamma, so that

$$\pi(\sigma_j^2 | \sigma_{-j}^2, \mu_1, \dots, \mu_J, \mu, \tau^2, \nu_0, \sigma_0^2, Y) = \mathcal{IG} \left(\frac{\nu_j^*}{2}, \frac{\nu_j^* \sigma_j^{2(*)}}{2} \right) \quad \text{where}$$

$$\nu_j^* = \nu_0 + n_j; \quad \sigma_j^{2(*)} = \frac{1}{\nu_j^*} \left[\nu_0 \sigma_0^2 + \sum_{i=1}^{n_j} (y_{ij} - \mu_j)^2 \right].$$

REMAINING HYPER-PRIORS

- Now we can get back to priors for ν_0 and σ_0^2 . Turns out that a semi-conjugate prior for σ_0^2 (you have seen this on the homework) is a gamma distribution. That is, if we set

$$\pi(\sigma_0^2) = \mathcal{Ga}(a, b),$$

then,

$$\begin{aligned}\pi(\sigma_0^2 | \mu_1, \dots, \mu_J, \sigma_1^2, \dots, \sigma_J^2, \mu, \tau^2, \nu_0, Y) &\propto \left\{ \prod_{j=1}^J p(\sigma_j^2 | \nu_0, \sigma_0^2) \right\} \cdot \pi(\sigma_0^2) \\ &\propto \mathcal{IG}\left(\sigma_j^2; \frac{\nu_0}{2}, \frac{\nu_0 \sigma_0^2}{2}\right) \cdot \mathcal{Ga}(\sigma_0^2; a, b)\end{aligned}$$

- Recall that

- $\mathcal{Ga}(y; a, b) \equiv \frac{b^a}{\Gamma(a)} y^{a-1} e^{-by}$, and

- $\mathcal{IG}(y; a, b) \equiv \frac{b^a}{\Gamma(a)} y^{-(a+1)} e^{-\frac{b}{y}}.$

REMAINING HYPER-PRIORS

- So $\pi(\sigma_0^2 | \mu_1, \dots, \mu_J, \sigma_1^2, \dots, \sigma_J^2, \mu, \tau^2, \nu_0, Y)$

$$\begin{aligned}
 &\propto \left\{ \prod_{j=1}^J p(\sigma_j^2 | \nu_0, \sigma_0^2) \right\} \cdot \pi(\sigma_0^2) \\
 &\propto \prod_{j=1}^J \mathcal{IG} \left(\sigma_j^2; \frac{\nu_0}{2}, \frac{\nu_0 \sigma_0^2}{2} \right) \cdot \mathcal{Ga}(\sigma_0^2; a, b) \\
 &= \prod_{j=1}^J \frac{\left(\frac{\nu_0 \sigma_0^2}{2} \right)^{\left(\frac{\nu_0}{2} \right)}}{\Gamma \left(\frac{\nu_0}{2} \right)} (\sigma_j^2)^{-\left(\frac{\nu_0}{2} + 1 \right)} e^{-\frac{\nu_0 \sigma_0^2}{2(\sigma_j^2)}} \cdot \left[\frac{b^a}{\Gamma(a)} (\sigma_0^2)^{a-1} e^{-b\sigma_0^2} \right] \\
 &\propto \prod_{j=1}^J (\sigma_0^2)^{\left(\frac{\nu_0}{2} \right)} e^{-\frac{\nu_0 \sigma_0^2}{2(\sigma_j^2)}} \cdot \left[(\sigma_0^2)^{a-1} e^{-b\sigma_0^2} \right] \\
 &\propto (\sigma_0^2)^{\left(\frac{J\nu_0}{2} \right)} e^{-\sigma_0^2 \left[\frac{\nu_0}{2} \sum_{j=1}^J \frac{1}{\sigma_j^2} \right]} \cdot \left[(\sigma_0^2)^{a-1} e^{-b\sigma_0^2} \right]
 \end{aligned}$$

REMAINING HYPER-PRIORS

- That is, the full conditional is

$$\begin{aligned}\pi(\sigma_0^2 | \dots) &\propto \left[(\sigma_0^2)^{\left(\frac{J\nu_0}{2}\right)} e^{-\sigma_0^2 \left[\frac{\nu_0}{2} \sum_{j=1}^J \frac{1}{\sigma_j^2}\right]} \right] \cdot \left[(\sigma_0^2)^{a-1} e^{-b\sigma_0^2} \right] \\ &\propto \left[(\sigma_0^2)^{\left(a + \frac{J\nu_0}{2} - 1\right)} e^{-\sigma_0^2 \left[b + \frac{\nu_0}{2} \sum_{j=1}^J \frac{1}{\sigma_j^2}\right]} \right] \\ &\equiv \mathcal{Ga}(\sigma_0^2; a_n, b_n),\end{aligned}$$

where

$$a_n = a + \frac{J\nu_0}{2}; \quad b_n = b + \frac{\nu_0}{2} \sum_{j=1}^J \frac{1}{\sigma_j^2}.$$

REMAINING HYPER-PRIORS

- Ok that leaves us with one parameter to go, i.e., ν_0 . Turns out there is no simple conjugate/semi-conjugate prior for ν_0 .
- Common practice is to restrict ν_0 to be an integer (which makes sense when we think of it as being degrees of freedom, which also means it cannot be zero). With the restriction, we need a discrete distribution as the prior with support on $\nu_0 = 1, 2, 3, \dots$
- **Question: Can we use either a binomial or a Poisson prior on for ν_0 ?**
- A popular choice is the geometric distribution with pmf $p(\nu_0) = (1 - p)^{\nu_0 - 1} p$.
- However, we will rewrite the kernel as $\pi(\nu_0) \propto e^{-\alpha \nu_0}$. How did we get here from the geometric pmf and what is α ?

FINAL FULL CONDITIONAL

- With this prior, $\pi(\nu_0 | \mu_1, \dots, \mu_J, \sigma_1^2, \dots, \sigma_J^2, \mu, \tau^2, \sigma_0^2, Y)$

$$\begin{aligned}
 &\propto \left\{ \prod_{j=1}^J p(\sigma_j^2 | \nu_0, \sigma_0^2) \right\} \cdot \pi(\nu_0) \\
 &\propto \prod_{j=1}^J \mathcal{IG} \left(\sigma_j^2; \frac{\nu_0}{2}, \frac{\nu_0 \sigma_0^2}{2} \right) \cdot e^{-\alpha \nu_0} \\
 &= \left[\prod_{j=1}^J \frac{\left(\frac{\nu_0 \sigma_0^2}{2} \right)^{\left(\frac{\nu_0}{2} \right)}}{\Gamma \left(\frac{\nu_0}{2} \right)} (\sigma_j^2)^{-\left(\frac{\nu_0}{2} + 1 \right)} e^{-\frac{\nu_0 \sigma_0^2}{2(\sigma_j^2)}} \right] \cdot e^{-\alpha \nu_0} \\
 &\propto \left[\left(\frac{\left(\frac{\nu_0 \sigma_0^2}{2} \right)^{\left(\frac{\nu_0}{2} \right)}}{\Gamma \left(\frac{\nu_0}{2} \right)} \right)^J \cdot \left(\prod_{j=1}^J \frac{1}{\sigma_j^2} \right)^{\left(\frac{\nu_0}{2} + 1 \right)} \cdot e^{-\nu_0 \left[\frac{\sigma_0^2}{2} \sum_{j=1}^J \frac{1}{\sigma_j^2} \right]} \right] \cdot e^{-\alpha \nu_0}
 \end{aligned}$$

FINAL FULL CONDITIONAL

- That is, the full conditional is

$$\pi(\nu_0 | \dots) \propto \left[\frac{\left(\frac{\nu_0 \sigma_0^2}{2} \right)^{\left(\frac{\nu_0}{2} \right)^J}}{\Gamma\left(\frac{\nu_0}{2}\right)} \cdot \left(\prod_{j=1}^J \frac{1}{\sigma_j^2} \right)^{\left(\frac{\nu_0}{2} + 1 \right)} \cdot e^{-\nu_0 \left[\alpha + \frac{\sigma_0^2}{2} \sum_{j=1}^J \frac{1}{\sigma_j^2} \right]} \right],$$

which is not a known kernel and is thus unnormalized (i.e., does not integrate to 1 in its current form).

- This sure looks like a lot, but it will be relatively easy to compute in R.
- Now, technically, the support is $\nu_0 = 1, 2, 3, \dots$, however, we can compute this to compute the unnormalized distribution across a grid of ν_0 values, say, $\nu_0 = 1, 2, 3, \dots, K$ for some large K , and then sample.

FINAL FULL CONDITIONAL

- One more thing, computing these probabilities on the raw scale can be problematic particularly because of the product inside. Good idea to transform to the log scale instead.
- That is,

$$\pi(\nu_0 | \dots) \propto \left[\frac{\left(\frac{\nu_0 \sigma_0^2}{2} \right)^{\left(\frac{\nu_0}{2} \right)^J}}{\Gamma\left(\frac{\nu_0}{2}\right)} \cdot \left(\prod_{j=1}^J \frac{1}{\sigma_j^2} \right)^{\left(\frac{\nu_0}{2} - 1 \right)} \cdot e^{-\nu_0 \left[\alpha + \frac{\sigma_0^2}{2} \sum_{j=1}^J \frac{1}{\sigma_j^2} \right]} \right]$$

$$\begin{aligned} \Rightarrow \ln \pi(\nu_0 | \dots) &\propto \left(\frac{J\nu_0}{2} \right) \ln \left(\frac{\nu_0 \sigma_0^2}{2} \right) - J \ln \left[\Gamma\left(\frac{\nu_0}{2}\right) \right] \\ &\quad + \left(\frac{\nu_0}{2} + 1 \right) \left(\sum_{j=1}^J \ln \left[\frac{1}{\sigma_j^2} \right] \right) \\ &\quad - \nu_0 \left[\alpha + \frac{\sigma_0^2}{2} \sum_{j=1}^J \frac{1}{\sigma_j^2} \right] \end{aligned}$$

FULL MODEL

As a recap, the final model is therefore:

$$y_{ij}|\mu_j, \sigma_j^2 \sim \mathcal{N}(\mu_j, \sigma_j^2); \quad i = 1, \dots, n_j; \quad j = 1, \dots, J$$

$$\mu_j|\mu, \tau^2 \sim \mathcal{N}(\mu, \tau^2); \quad j = 1, \dots, J$$

$$\sigma_1^2, \dots, \sigma_J^2|\nu_0, \sigma_0^2 \sim \mathcal{IG}\left(\frac{\nu_0}{2}, \frac{\nu_0 \sigma_0^2}{2}\right); \quad j = 1, \dots, J$$

$$\mu \sim \mathcal{N}(\mu_0, \gamma_0^2)$$

$$\tau^2 \sim \mathcal{IG}\left(\frac{\eta_0}{2}, \frac{\eta_0 \tau_0^2}{2}\right).$$

$$\pi(\nu_0) \propto e^{-\alpha \nu_0}$$

$$\sigma_0^2 \sim \mathcal{Ga}(a, b).$$

WHAT'S NEXT?

MOVE ON TO THE READINGS FOR THE NEXT MODULE!