STA 610L: Module 2.10

Random effects ANOVA (Bayesian estimation II)

Dr. Olanrewaju Michael Akande

STA 610L

Challenge to validity: heterogeneous means and variances

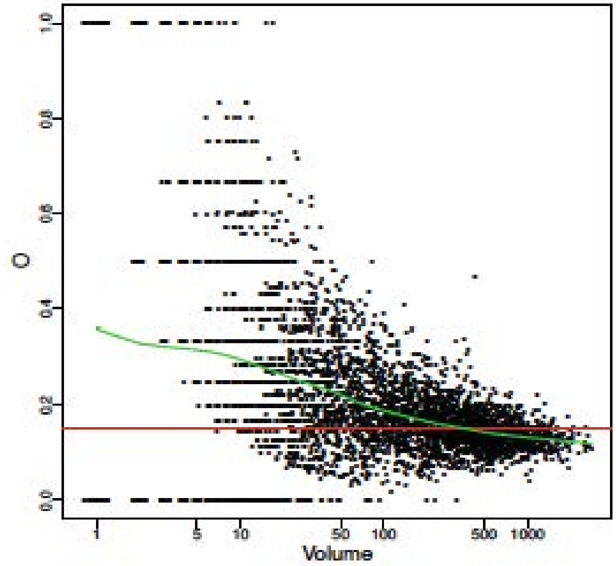
We have looked at the hierarchical normal model of the form

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

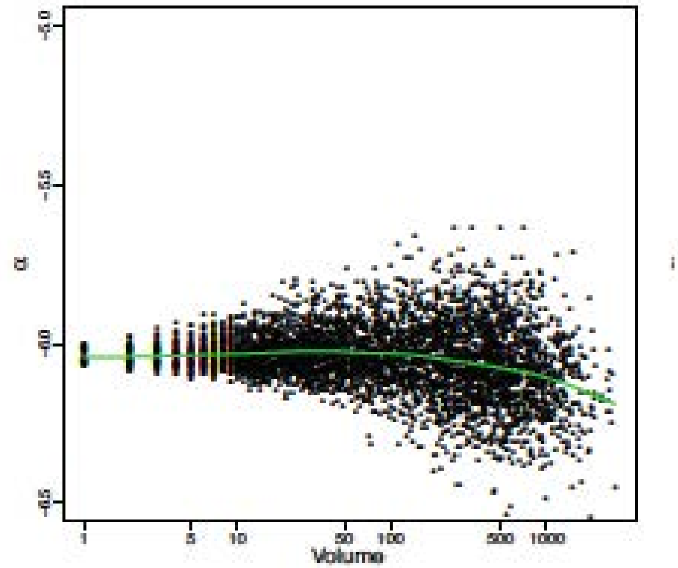
The model gives us an extra hierarchy through the prior on the means, leading to sharing of information across the groups, when estimating the group-specific means.

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

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

STA 610L

Posterior inference

- The full posterior is now:

$$
\begin{aligned}
\pi(\mu_1,\ldots,\mu_J,\sigma_1^2,\ldots,\sigma_J^2,\mu,\tau^2,\nu_0,\sigma_0^2|Y) \propto\; & p(y|\mu_1,\ldots,\mu_J,\sigma_1^2,\ldots,\sigma_J^2,\mu,\tau^2,\nu_0,\sigma_0^2)\\
& \times p(\mu_1,\ldots,\mu_J|\sigma_1^2,\ldots,\sigma_J^2,\mu,\tau^2,\nu_0,\sigma_0^2)\\
& \times p(\sigma_1^2,\ldots,\sigma_J^2|\mu,\tau^2,\nu_0,\sigma_0^2)\\
& \times \pi(\mu,\tau^2,\nu_0,\sigma_0^2)\\
\\
=\; & p(y|\mu_1,\ldots,\mu_J,\sigma_1^2,\ldots,\sigma_J^2)\\
& \times p(\mu_1,\ldots,\mu_J|\mu,\tau^2)\\
& \times p(\sigma_1^2,\ldots,\sigma_J^2|\nu_0,\sigma_0^2)\\
& \times \pi(\mu)\cdot\pi(\tau^2)\cdot\pi(\nu_0)\cdot\pi(\sigma_0^2)\\
\\
=\; & \left\{\prod_{j=1}^{J}\prod_{i=1}^{n_j} p(y_{ij}|\mu_j,\sigma_j^2)\right\}\\
& \times \left\{\prod_{j=1}^{J} p(\mu_j|\mu,\tau^2)\right\}\\
& \times \left\{\prod_{j=1}^{J} p(\sigma_j^2|\nu_0,\sigma_0^2)\right\}\\
& \times \pi(\mu)\cdot\pi(\tau^2)\cdot\pi(\nu_0)\cdot\pi(\sigma_0^2)
\end{aligned}
$$

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,

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

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

and

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

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

STA 610L

Full conditionals

- The full conditional for each μ_j, we have

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

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

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

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

STA 610L

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,

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

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

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

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

STA 610L

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

$$\pi(\sigma_0^2) = \mathcal{G}a\left(a, b\right),$$

then,

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

- Recall that
  - $\mathcal{G}a(y; a, b) \equiv \dfrac{b^a}{\Gamma(a)} y^{a-1} e^{-by}$, and
  - $\mathcal{IG}(y; a, b) \equiv \dfrac{b^a}{\Gamma(a)} y^{-(a+1)} e^{-\dfrac{b}{y}}$.

STA 610L

REMAINING HYPER-PRIORS

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

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

STA 610L

Remaining hyper-priors

- That is, the full conditional is

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

where

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

STA 610L

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 $\nu_0 = 1, 2, 3, \ldots$.
- **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?

STA 610L

Final full conditional

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

$$\propto \left\{\prod_{j=1}^{J} p(\sigma_j^2|\nu_0,\sigma_0^2)\right\}\cdot\pi(\nu_0)$$

$$\propto \prod_{j=1}^{J} \mathcal{IG}\left(\sigma_j^2;\frac{\nu_0}{2},\frac{\nu_0\sigma_0^2}{2}\right)\ \cdot\ e^{-\alpha\nu_0}$$

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

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

STA 610L

Final full conditional

- That is, the full conditional is

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

  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 $\nu_0 = 1, 2, 3, \ldots$, however, we can compute this to compute the unnormalized distribution across a grid of ν_0 values, say, $\nu_0 = 1, 2, 3, \ldots, K$ for some large K, and then sample.

STA 610L

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,

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

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

Full model

As a recap, the final model is:

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

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

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

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

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

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

$$\sigma_0^2 \sim \mathcal{G}a\left(a,b\right).$$

STA 610L

Gibbs sampler

```
#Data summaries
J <- #number of groups
ybar <- #vector of the group sample means
s_j_sq <- #vector of the group sample variances
n <- #vector of the number of observations in each group


#Hyperparameters for the priors
mu_0 <- ...
gamma_0_sq <- ...
eta_0 <- ...
tau_0_sq <- ...
alpha <- ...
a <- ...
b <- ...


#Grid values for sampling nu_0_grid
nu_0_grid <- 1:5000


#Initial values for Gibbs sampler
theta <- ybar #theta vector for all the mu_j's
sigma_sq <- s_j_sq
mu <- mean(theta)
tau_sq <- var(theta)
nu_0 <- 1
sigma_0_sq <- 100
```

STA 610L

Gibbs sampler

```r
#first set number of iterations and burn-in, then set seed
n_iter <- 10000; burn_in <- 0.3*n_iter
set.seed(1234)


#Set null matrices to save samples
SIGMA_SQ <- THETA <- matrix(nrow=n_iter, ncol=J)
OTHER_PAR <- matrix(nrow=n_iter, ncol=4)


#Now, to the Gibbs sampler
for(s in 1:(n_iter+burn_in)){

  #update the theta vector (all the mu_j's)
  tau_j_star <- 1/(n/sigma_sq + 1/tau_sq)
  mu_j_star <- tau_j_star*(ybar*n/sigma_sq + mu/tau_sq)
  theta <- rnorm(J,mu_j_star,sqrt(tau_j_star))

  #update the sigma_sq vector (all the sigma_sq_j's)
  nu_j_star <- nu_0 + n
  theta_long <- rep(theta,n)
  nu_j_star_sigma_j_sq_star <-
    nu_0*sigma_0_sq + c(by((Y[,"mathscore"] - theta_long)^2,Y[,"school"],sum))
  sigma_sq <- 1/rgamma(J,(nu_j_star/2),(nu_j_star_sigma_j_sq_star/2))

  #update mu
  gamma_n_sq <- 1/(J/tau_sq + 1/gamma_0_sq)
  mu_n <- gamma_n_sq*(J*mean(theta)/tau_sq + mu_0/gamma_0_sq)
  mu <- rnorm(1,mu_n,sqrt(gamma_n_sq))
```

STA 610L

Gibbs sampler

```
  #update tau_sq
  eta_n <- eta_0 + J
  eta_n_tau_n_sq <- eta_0*tau_0_sq + sum((theta-mu)^2)
  tau_sq <- 1/rgamma(1,eta_n/2,eta_n_tau_n_sq/2)

  #update sigma_0_sq
  sigma_0_sq <- rgamma(1,(a + J*nu_0/2),(b + nu_0*sum(1/sigma_sq)/2))

  #update nu_0
  log_prob_nu_0 <- (J*nu_0_grid/2)*log(nu_0_grid*sigma_0_sq/2) -
    J*lgamma(nu_0_grid/2) +
    (nu_0_grid/2+1)*sum(log(1/sigma_sq)) -
    nu_0_grid*(alpha + sigma_0_sq*sum(1/sigma_sq)/2)
  nu_0 <- sample(nu_0_grid,1, prob = exp(log_prob_nu_0 - max(log_prob_nu_0)) )
  #this last step substracts the maximum logarithm from all logs
  #it is a neat trick that throws away all results that are so negative
  #they will screw up the exponential
  #note that the sample function will renormalize the probabilities internally


  #save results only past burn-in
  if(s > burn_in){
    THETA[(s-burn_in),] <- theta
    SIGMA_SQ[(s-burn_in),] <- sigma_sq
    OTHER_PAR[(s-burn_in),] <- c(mu,tau_sq,sigma_0_sq,nu_0)
  }
}
colnames(OTHER_PAR) <- c("mu","tau_sq","sigma_0_sq","nu_0")
```

STA 610L

What's next?

Move on to the readings for the next module!