



Schools: ranking school examination results using multivariate hierarcical models

Goldstein et al. (1993) present an analysis of examination results from inner London schools. They use hierarchical or multilevel models to study the between-school variation, and calculate school-level residuals in an attempt to differentiate between 'good' and 'bad' schools. Here we analyse a subset of this data and show how to calculate a rank ordering of schools and obtain credible intervals on each rank.

Data

Standardized mean examination scores (Y) were available for 1978 pupils from 38 different schools. The median number of pupils per school was 48, with a range of 1--198. Pupil-level covariates included gender plus a standardized London Reading Test (LRT) score and a verbal reasoning (VR) test category (1, 2 or 3, where 1 represents the highest ability group) measured when each child was aged 11. Each school was classified by gender intake (all girls, all boys or mixed) and denomination (Church of England, Roman Catholic, State school or other); these were used as categorical school-level covariates.

Model

We consider the following model, which essentially corresponds to Goldstein et al.'s model 1.

$$\begin{aligned}
 Y_{ij} &\sim \text{Normal}(\mu_{ij}, \tau_{ij}) \\
 \mu_{ij} &= \alpha_{1j} + \alpha_{2j} \text{LRT}_{ij} + \alpha_{3j} \text{VR}_{1ij} + \beta_1 \text{LRT}_{ij}^2 + \beta_2 \text{VR}_{2ij} + \beta_3 \text{Girl}_{ij} \\
 &\quad + \beta_4 \text{Girls' school}_j + \beta_5 \text{Boys' school}_j + \beta_6 \text{CE school}_j \\
 &\quad + \beta_7 \text{RC school}_j + \beta_8 \text{other school}_j \\
 \log \tau_{ij} &= \theta + \phi \text{LRT}_{ij}
 \end{aligned}$$

where i refers to pupil and j indexes school. We wish to specify a regression model for the variance components, and here we model the logarithm of τ_{ij} (the inverse of the between-pupil variance) as a linear function of each pupil's LRT score. This differs from Goldstein et al.'s model which allows the *variance* σ^2_{ij} to depend linearly on LRT. However, such a parameterization may lead to negative estimates of σ^2_{ij} .

Prior distributions

The fixed effects β_k ($k=1, \dots, 8$), θ and ϕ were assumed to follow vague independent Normal distributions with zero mean and low precision = 0.0001. The random school-level coefficients

α_{kj} ($k = 1, 2, 3$) were assumed to arise from a multivariate normal population distribution with unknown mean γ and covariance matrix Σ . A non-informative multivariate normal prior was then specified for the population mean γ , whilst the inverse covariance matrix $\mathbf{T} = \Sigma^{-1}$ was assumed to follow a Wishart distribution. To represent vague prior knowledge, we chose the the degrees of freedom for this distribution to be as small as possible (i.e. 3, the rank of \mathbf{T}). The scale matrix \mathbf{R} was specified as

$$\begin{vmatrix} 0.1 & 0.005 & 0.005 \\ 0.005 & 0.01 & 0.005 \\ 0.005 & 0.005 & 0.01 \end{vmatrix}$$

which represents our prior guess at the order of magnitude of Σ .

The BUGS code is given below:

```
model
{
  for(p in 1 : N) {
    Y[p] ~ dnorm(mu[p], tau[p])
    mu[p] <- alpha[school[p], 1] + alpha[school[p], 2] * LRT[p]
      + alpha[school[p], 3] * VR[p, 1] + beta[1] * LRT2[p]
      + beta[2] * VR[p, 2] + beta[3] * Gender[p]
      + beta[4] * School.gender[p, 1] + beta[5] * School.gender[p, 2]
      + beta[6] * School.denom[p, 1] + beta[7] * School.denom[p, 2]
      + beta[8] * School.denom[p, 3]
    log(tau[p]) <- theta + phi * LRT[p]
    sigma2[p] <- 1 / tau[p]
    LRT2[p] <- LRT[p] * LRT[p]
  }
  min.var <- exp(-(theta + phi * (-34.6193))) # lowest LRT score = -34.6193
  max.var <- exp(-(theta + phi * (37.3807))) # highest LRT score = 37.3807

  # Priors for fixed effects:
  for (k in 1 : 8) {
    beta[k] ~ dnorm(0.0, 0.0001)
  }
  theta ~ dnorm(0.0, 0.0001)
  phi ~ dnorm(0.0, 0.0001)

  # Priors for random coefficients:
  for (j in 1 : M) {
    alpha[j, 1 : 3] ~ dmnorm(gamma[1:3], T[1:3, 1:3]);
    alpha1[j] <- alpha[j, 1]
  }

  # Hyper-priors:
```

```

gamma[1 : 3] ~ dmnorm(mn[1:3 ], prec[1:3 ,1:3 ]);
T[1 : 3, 1 : 3 ] ~ dwish(R[1:3 ,1:3 ], 3)
}

```

[Data](#) (click to open)

Note that `school` is a 1978 x 3 matrix taking value 1 for all pupils in school 1, 2 for all pupils in school 2 and so on. For computational convenience, Y , μ and τ are indexed over a single dimension $p = 1, \dots, 1978$ rather than as pupil i within school j as used in equations above. The appropriate school-level coefficients for pupil p are then selected using the school indicator in row p of the data array --- for example `alpha[school[p], 1]`.

[Inits](#) (click to open)

Results

A 1000 update burn in followed by a further 10000 updates gave the parameter estimates

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
beta[1]	2.64E-4	9.842E-5	2.429E-6	7.499E-5	2.638E-4	4.558E-4	1001	10000
beta[2]	0.4219	0.06225	0.00344	0.301	0.4206	0.5466	1001	10000
beta[3]	0.1725	0.04834	0.001427	0.07689	0.1722	0.2682	1001	10000
beta[4]	0.125	0.1377	0.006001	-0.1558	0.1276	0.3939	1001	10000
beta[5]	0.06201	0.1038	0.004941	-0.1475	0.06323	0.2624	1001	10000
beta[6]	-0.2769	0.1875	0.007158	-0.6584	-0.2728	0.08729	1001	10000
beta[7]	0.1441	0.1061	0.004271	-0.05912	0.1428	0.36	1001	10000
beta[8]	-0.1667	0.1733	0.006393	-0.4943	-0.1675	0.1846	1001	10000
gamma[1]	-0.6778	0.09568	0.005593	-0.8668	-0.6783	-0.4862	1001	10000
gamma[2]	0.03135	0.01019	1.396E-4	0.01139	0.03139	0.05137	1001	10000
gamma[3]	0.9597	0.08626	0.004849	0.7947	0.959	1.129	1001	10000
phi	-0.002605	0.002829	3.159E-5	-0.008146	-0.00259	0.002927	1001	10000
theta	0.5801	0.03205	3.518E-4	0.5163	0.5803	0.6414	1001	10000

Estimating the ranks

The school-specific intercept α_{j1} measures the 'residual effect' for school j after adjusting for pupil- and school-level covariates. This might represent an appropriate quantity by which to rank schools' performance. We compute the ranks in BUGS using the "rank" option of the "Statistics" menu, which we set for the variable `alpha` at the same time as we set the "sample monitor" option. Since the rank is a function of stochastic nodes, its value will change at every iteration. Hence we may obtain a posterior distribution for the rank of `alpha[, k]` which may be summarized by posterior histograms as shown below: