Integrating Multivariate Normal and Hierarchical Models: Bayesian Analysis of Leukemia Subtypes

Alberto Galdino Nogara (ID: 5201011)
Tommaso Biganzoli (ID: 5306102)

Section I : Dataset Introduction

In the realm of medical research, particularly in the study of leukemia, understanding the molecular and genetic differences between various subtypes is crucial for developing targeted therapies and improving patient outcomes. Leukemia, a type of blood cancer, can be classified into different subtypes based on the French-American-British (FAB) classification system. This project aims to leverage Bayesian modeling techniques to analyze a leukemia dataset with the primary objective of comparing posterior expectations between different FAB groups.

The provided dataset includes various molecular markers and their corresponding levels in patients with different subtypes of leukemia. In the above we are just considering only 4 groups (M0,M1,M2 and M4), because of the presence of multiple Na's in the others.

The main goals of this analysis are three:

1- Verify the posterior probability distribution of each parameter, i.e. we are going to take the parameters Ω , θ_0 , θ_j , τ^2 and derive their posterior from the Gibbs output. The Gibbs is made necessary by the fact that the marginal distribution of the data is not available in a closed form, so we are going to sample directly from the full conditional of each parameter.
2- Comparison of Posterior Expectation between groups, i.e. we are going to take the posterior of θ_j [vectors of expectations] with mean for every p covariate, and from these vectors of expectations we are going to extract the mean value of a protein (a covariate) in a given group and compare it to the mean of the same protein of the other three groups.
3- Shrinkage: what influence have the groups within each other regarding the values of the covariates, so how far is the the posterior expectation of θ_j from $\overline{y}_j$.

Section II : Statistical Introduction

Bayesian modelling has become an indispensable tool in modern statistics and data science due to its coherent framework for incorporating prior knowledge and updating beliefs with new

data. Among the various Bayesian models, the Multivariate Normal Hierarchical Model (MVNHM) stands out for its flexibility and applicability to a wide range of complex, real-world problems. This project extends the traditional Multivariate Normal Model, by integrating it with the classical hierarchical Normal Model, enhancing its capability to handle intricate dependencies and structures inherent in high-dimensional data.
As we know, the following are the key features of this model:

$$Let(X_1, \dots, X_p) be\ a\ random\ vector.$$

We say that
$(X_1, \dots, X_p)$ has a multivariate Normal distribution with parameters μ (mean) and Σ (covariance matrix)

$$(X_1, \dots, X_p) \mid \mu, \Sigma \sim \mathcal{N}_p(\mu, \Sigma)$$

$$\boldsymbol{\mu} = (\mu_1, \dots, \mu_p)^\top, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \cdots & \sigma_{1p} \\ \vdots & \ddots & \vdots \\ \sigma_{p1} & \cdots & \sigma_p^2 \end{bmatrix}$$

if its probability density function is given by

$$p(x \mid \mu, \Sigma) = (2\pi)^{-\frac{p}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right\}$$

While for the Hierarchical Normal Model the within and between group sampling models are both normal:

$$(Y_{1,j}, \dots, Y_{i,j}) \mid \theta_j, \Omega^{-1} \ \text{iid} \sim N_p(\theta_j, \Omega^{-1}); \quad j = 1, \dots, d$$

$$(\theta_1, \dots, \theta_d) \mid \theta_0, T_0^{-1} \ \text{iid} \sim N_p(\theta_0, T_0^{-1})$$

Priors:

$$\Omega \sim \text{Wp}(a, U)$$

$$\theta_0 \sim N_p(m_0, K_0)$$

$$T_0 \sim \text{Wp}(b, W)$$

We can represent through a diagram the structure of our model.
In the case of our four groups of interest, like those in the dataset of reference, the diagram is the following:

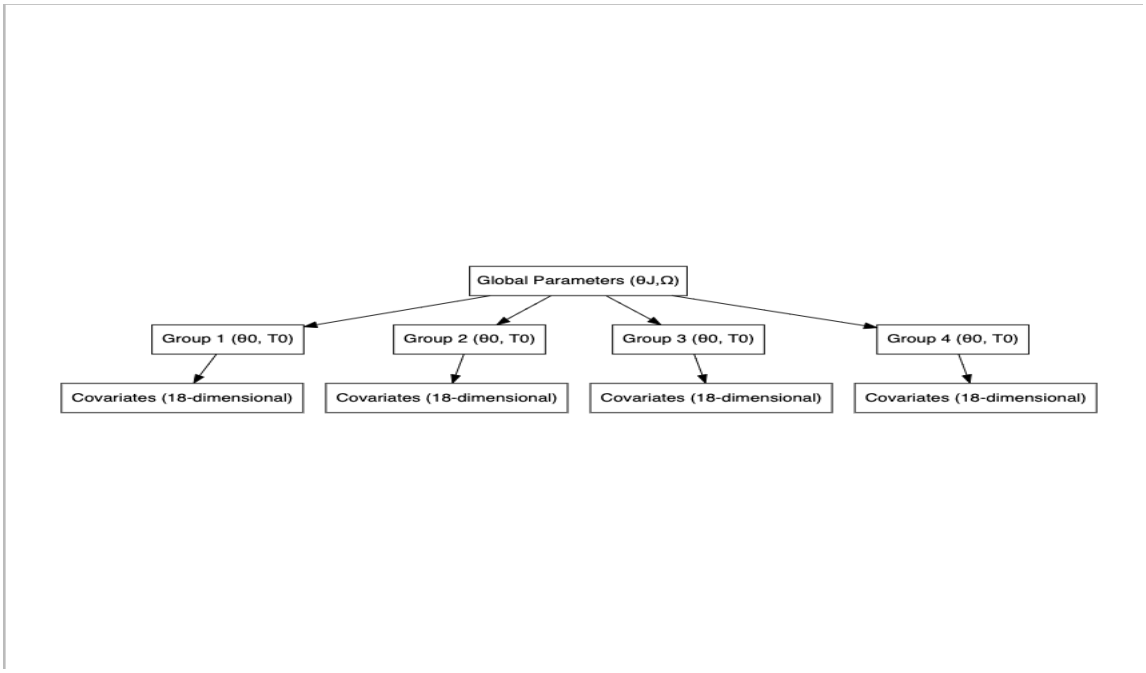

Figure 1

Having introduced the above well-known models, we can now extend them introducing the following Joint Posterior and the respective full conditionals:

Joint Posterior (Assuming Independence a priori)

$$P(\theta_1, \dots, \theta_d, \Omega, \theta_0, \mathrm{T}_0 \mid x) \propto P(\Omega)P(\theta_0)P(\mathrm{T}_0)\left\{\prod_{j=1}^{d} P\left(\theta_j \mid \theta_0, \mathrm{T}_0^{-1}\right)\right\}\left\{\prod_{j=1}^{d}\prod_{i=1}^{n_j} P\left(x_{ij} \mid \theta_j, \Omega^{-1}\right)\right\}$$

Full conditional of Ω

$$P(\Omega \mid \cdot) = P(\Omega) \prod_{j=1}^{d} \prod_{i=1}^{n_j} \left(x_{ij} \mid \theta_j, \Omega^{-1} \right)$$

$$\propto |\Omega|^{\frac{a-p-1}{2}} \exp\left\{-\frac{1}{2}\mathrm{tr}(U\Omega)\right\} |\Omega|^{\frac{n_j}{2}} \exp\left\{-\frac{1}{2}\sum_j \sum_i \left(x_{ij} - \theta_j\right)^T \Omega \left(x_{ij} - \theta_j\right)\right\}$$

Let

$$\sum_j \sum_i \left(x_{ij} - \theta_j\right)^T \Omega \left(x_{ij} - \theta_j\right) = \mathrm{tr}(S\Omega)$$

Where

$$S = \sum_j \sum_i \left(x_{ij} - \theta_j\right)\left(x_{ij} - \theta_j\right)^T$$

$$\propto |\Omega|^{\frac{a+n-p-1}{2}} \exp\left\{-\frac{1}{2}\mathrm{tr}\left((S+U)\Omega\right)\right\}$$

Let $a_n = a + n$ and $U_n = S + U$

So $\boldsymbol{P(\Omega \mid \cdot) \propto \mathbf{dWp}(\Omega; a_n, U_n)}$

Full conditional of $\boldsymbol{\theta_0}$

$$\mathrm{P}(\theta_0 \mid \cdot) \propto \prod_{\mathrm{j}=1}^{\mathrm{d}} \underbrace{P\left(\theta_j \mid \theta_0, \mathrm{T}_0^{-1}\right)}_{(1)} \underbrace{\mathrm{P}(\theta_0)}_{(2)}$$

$$(2) \quad \propto \exp\left\{-\frac{1}{2}\theta_0^T A_0 \theta_0 + \theta_0^T b_0\right\}$$

$$\text{with} \quad A_0 = k_0^{-1} \quad \text{and} \quad b_0 = k_0^{-1}\mathrm{m}_0$$

$$(1) \times (2)$$

$$\propto \exp\left\{-\frac{1}{2}\theta_0^T (d\mathrm{T}_0)\theta_0 + \theta_0^T \mathrm{T}_0 \left(\sum_{j=1}^{d} \theta_j\right)\right\} \cdot \exp\left\{-\frac{1}{2}\theta_0^T A_0 \theta_0 + \theta_0^T b_0\right\}$$

$$= \exp\left\{-\frac{1}{2}\theta_0^T (A_0 + d_0\mathrm{T}_0)\theta_0 + \theta_0^T (b_0 + \bar{\theta})\right\}$$

$$with \quad A_c = A_0 + d\mathrm{T}_0 \quad \text{and} \quad b_c = b_0 + \bar{\theta}$$

$$\text{So:} \quad \mathbf{P(\theta_0 \mid \cdot) \sim dNp(\theta_0 \mid A_c^{-1} b_c, A_c^{-1})}$$

Full Conditional of $\boldsymbol{\theta_j}$

$$P(\theta_j \mid \cdot) \propto P(\theta_j \mid \theta_0, \mathrm{T}_0^{-1}) \prod_{i=1}^{n_j} P(x_{ij} \mid \theta_j, \Omega^{-1})$$

$$\propto \exp\left\{-\frac{1}{2}(\theta_j - \theta_0)^T \mathrm{T}_0 (\theta_j - \theta_0)\right\} \cdot \exp\left\{-\frac{1}{2}\sum_i (x_i - \theta_j)^T \Omega (x_i - \theta_j)\right\}$$

$$\text{Let} \quad \mathrm{T}_0 = A_p \quad \text{and} \quad \mathrm{T}_0\theta_0 = b_p$$

$$2\exp\left\{-\frac{1}{2}\theta_j^T A_p \theta_j + \theta_j^T b_p\right\} \cdot \exp\left\{-\frac{1}{2}\sum_i \left[\theta_j^T \Omega \theta_j - 2\theta_j^T \Omega x_i\right]\right\}$$

$$\text{Let } n\Omega = A_L, \quad n\Omega\bar{x}_{\square} = b_L$$

$$\text{and} \quad A_L + A_p = A_n, \quad b_L = b_L + b_p$$

$$\textbf{So: } \boldsymbol{P(\theta_j \mid \cdot) \sim dNp\left(A_n^{-1} b_n, A_n^{-1}\right)}$$

Full conditional of $\boldsymbol{T_0}$ or $\boldsymbol{\widetilde{T_0} = T_0^{-1}}$

$$P(\widetilde{T}_0 \mid \text{rest}) \propto \underbrace{P(\underline{\theta}_1, \ldots, \underline{\theta}_d \mid \underline{\theta}_0, \widetilde{T}_0)}_{(1)} \underbrace{P(\widetilde{T}_0)}_{(2)}$$

$$(1) \propto |\widetilde{T}_0|^{\frac{d}{2}} \exp\left\{-\frac{1}{2}\underbrace{\sum_{j=1}^{d} (\underline{\theta}_j - \underline{\theta}_0)^T \widetilde{T}_0 (\underline{\theta}_j - \underline{\theta}_0)}_{(*)}\right\}$$

$$(2) \propto \left|\widetilde{T_0}\right|^{\frac{b-p-1}{2}} \exp\left\{-\frac{1}{2}\mathrm{tr}\left[W\widetilde{T_0}\right]\right\}$$

$$(*) = \mathrm{tr}\left[\tilde{S}\widetilde{T_0}\right] = \sum_{j=1}^{d}\left(\underline{\theta_j} - \underline{\theta_0}\right)\left(\underline{\theta_j} - \underline{\theta_0}\right)^T = \tilde{S}$$

Hence

$$\boldsymbol{P\left(\widetilde{T_0} \mid \mathbf{rest}\right) \propto \left|\widetilde{T_0}\right|^{\frac{d+b-p-1}{2}} \exp\left\{-\frac{1}{2}\mathrm{tr}\left[\left[W + \tilde{S}\right]\right]\widetilde{T_0}\right\}}$$
$$\boldsymbol{= \mathrm{dW}_p\left(\widetilde{T_0} \mid d + b, W + \tilde{S}\right)}$$

Section III : Gibbs sampling algorythm

We decided to use an MCMC strategy to derive the posterior distribution of our parameters, in particular we decided to implement a Gibbs algorithm that updated the values of Prior hyperparameters.
The choice falls on the fact that our joint posterior distribution is not known, but are the different full conditional, allowing us to make a draw direct from them.

Before presenting the algorithm, let's look at the fixed Prior Hyperparameter. Since we are in a Weakly informative scenario, I will have very wide variance (so very small precisions) and very small prior sample size. In particular omega will be the inverse of the variance-covariance matrix, therefore the within group variability of our data.
θ_j initialize it at $\overline{y}_j$, the arithmetic mean of the j-th group.

```
p <- ncol(y) ;n <- nrow(y);D <- 4;omega <- solve(cov(y));T0 <- solve(diag(100, p
));theta0 <- rep(0, p);U <- diag(0.01, p);a <- p;b <- p;W <- diag(0.01, p);m0 <-
rep(0, p);k0 <- diag(1,p) ;S <- 2000

#initializing thetaj as the mean in group d (ybar)
ybar <- list()
for (group in list(M0, M1, M2, M4)) {
  group <- group[, -which(names(group) == "Group")]
  media_group <- colMeans(group)
  ybar[[length(ybar) + 1]] <- media_group
}
ybar <- lapply(ybar, function(x) as.numeric(x))
```

Then the structure of our Gibbs will consist of two large for-loop, the first iterating between the S samples, and the second between the number D of groups, which we said to be 4.

Initialization of parameters:

```
omega_post = array(NA, c(p, p, S))
thetaj_post = array(NA,c(S,p,D))
theta0_post = matrix(NA,S,p)
T0_post = array(NA, c(p, p, S))
```

Main loop (for each iteration `s`):

- **Update omega**:
 - Compute parameters `an`, `Sn`, `Un`.
 - Sample `omega` using `rWishart`.

- **Update theta0**:
 - Compute `thetaJ_bar`, `A0`, `b0`, `mp`, `kp`.
 - Sample `theta0` using `rmvnorm`.

- **Update T0**:
 - Compute parameters `bn`, `S2`, `Wn`.
 - Sample `T0` using `rWishart`.

- **Update thetaj**:
 - Inner loop for each `d`:
  - Compute `An1`, `bn`.
  - Sample `thetaj[[d]]` using `rmvnorm`.
  - Store `thetaj_post`.

- **Output storage**:
 - thetaj_post[s,,d] = thetaJ_post
  omega_post[,,s] = omega
  theta0_post[s,] = theta0
  T0_post[,,s] = T0

Return results:

- List containing `thetaj_post`, `Omega_post`, `theta0_post`, `T0_post`.

Let's have a look at the posterior distribution of our parameters, resulting from the Gibbs above:

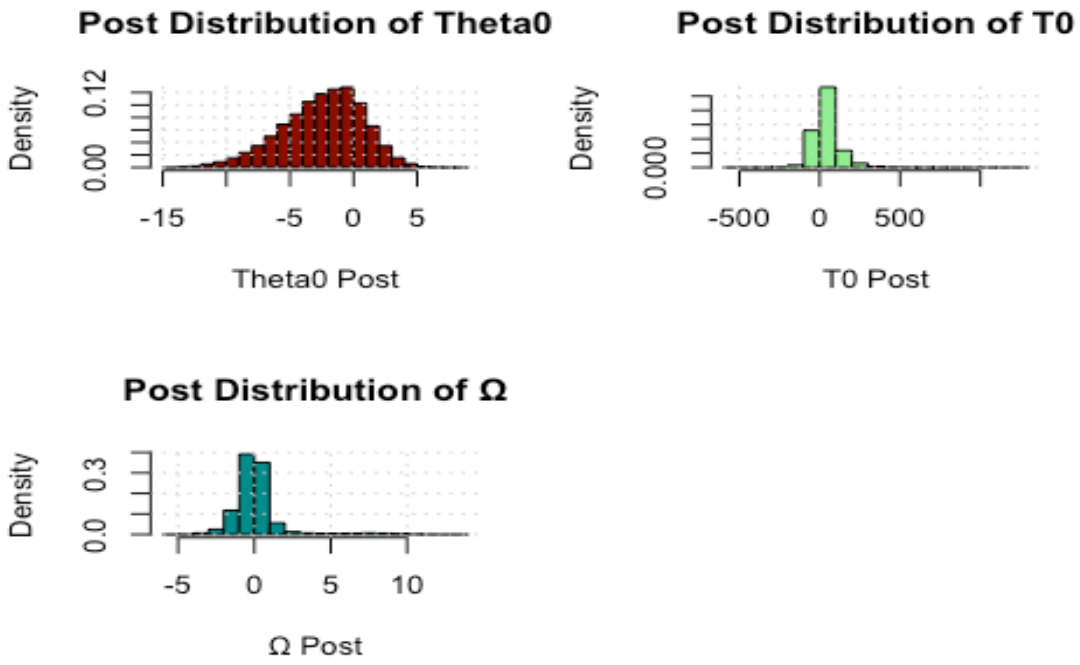

Figure 2

Above we have the histograms of the posterior distribution of our model parameters updated with the Gibbs algorithm. The first one in dark red can be brought back to normal, as it seems to have a belly shape a little shifted to the left. Instead for the posteriors of T0 and Omega, we can see that the values are mostly concentrated around 0. These results reflect the data a lot, because we remember that we have supposed some Weakly informative Priors, so the data still dominate our posterior distributions.

Section IV

We can then return to our initial problem, that of comparing the posterior expectations between the groups, extracting from each of the four thetaj_post two covariates and comparing the means.
We know that thetaj_post is an array of 4 vectors, each for each d group, of 18 averages (each for each covariate). For each group we extracted two different predictors, in the case of Figure 3 the protein AKT, and in the case of Figure 4, the protein BAD.
As we can see from both figures there is no big difference between the averages for each group. This, in our investigation is reflected in the fact that a posteriori, the averages of the concentration of proteins AKT and BAD in the blood do not differ significantly in changing from one group to another.

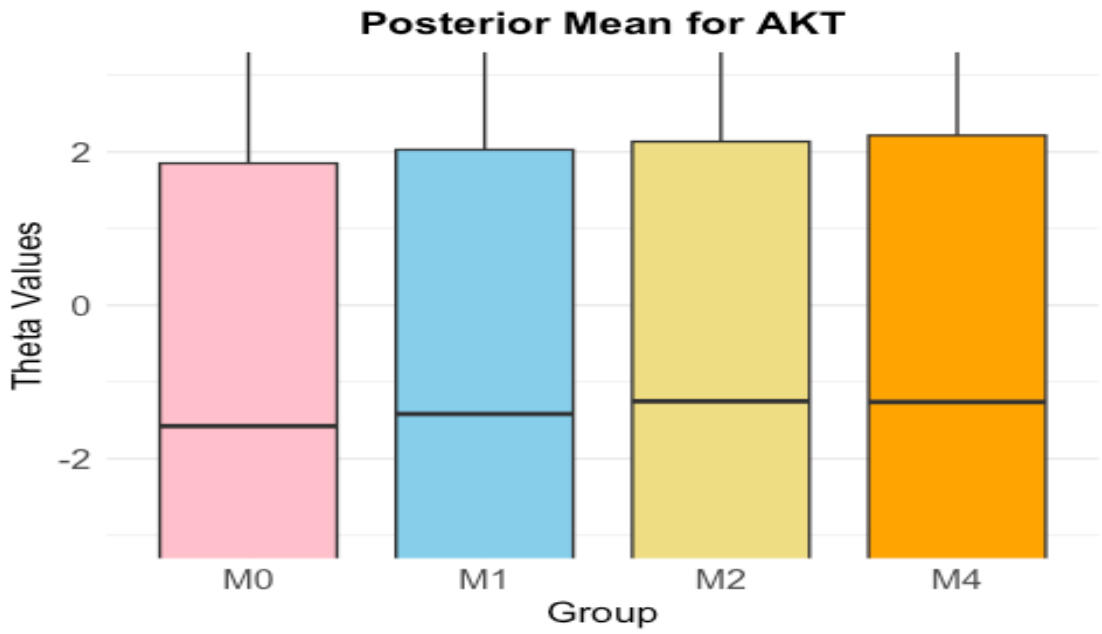

Figure 3

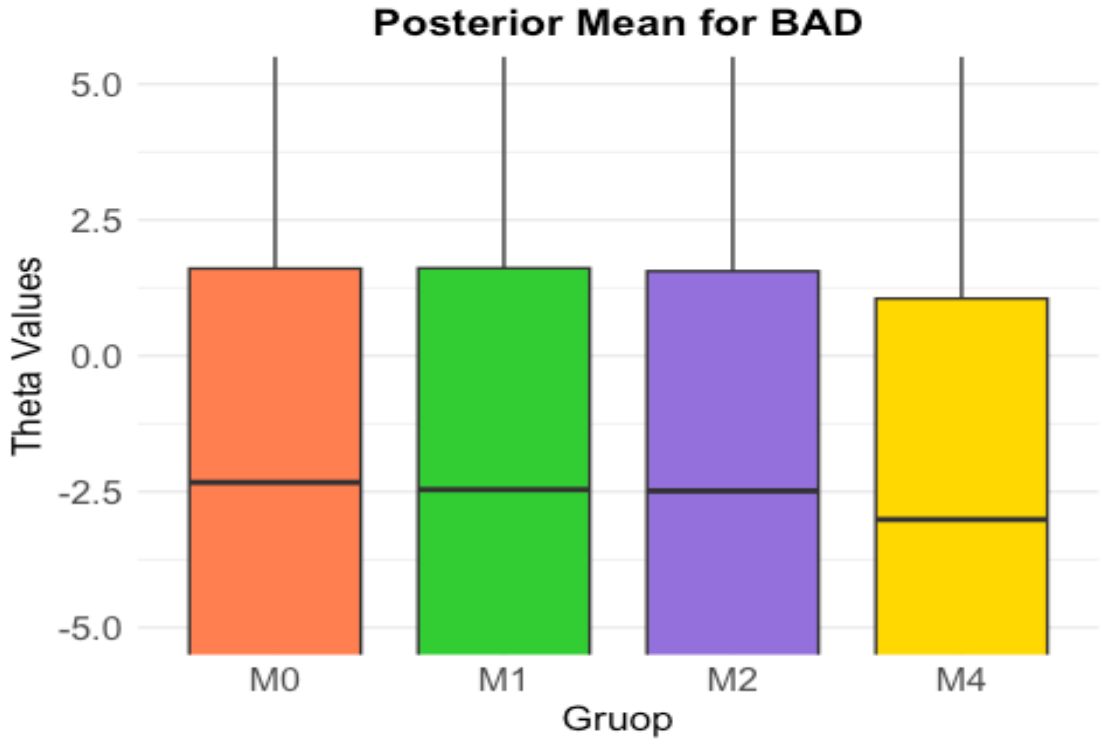

Figure 4

Shrinkage

A key concept employed in Bayesian statistics, particularly in the context of this project, is shrinkage. Shrinkage refers to the process of "pulling" estimates towards a central value or overall mean. In our case we are looking for observation (protein) that a posteriori are pretty different from their mean value a priori, not due to high prior information, but to the thetaj_post media component which also depends on the overall mean.

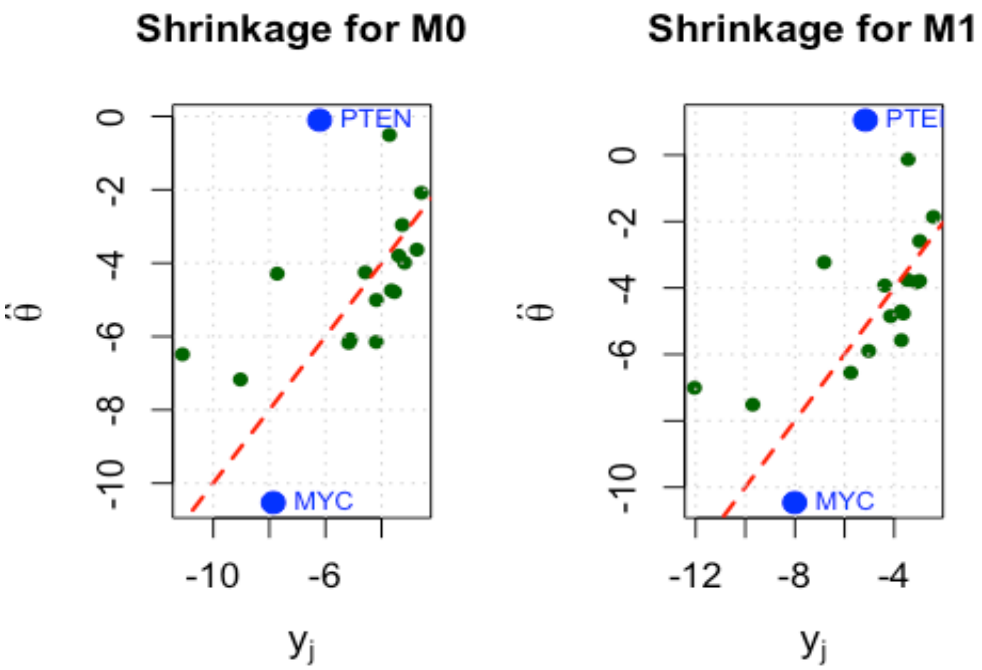

Figure 5

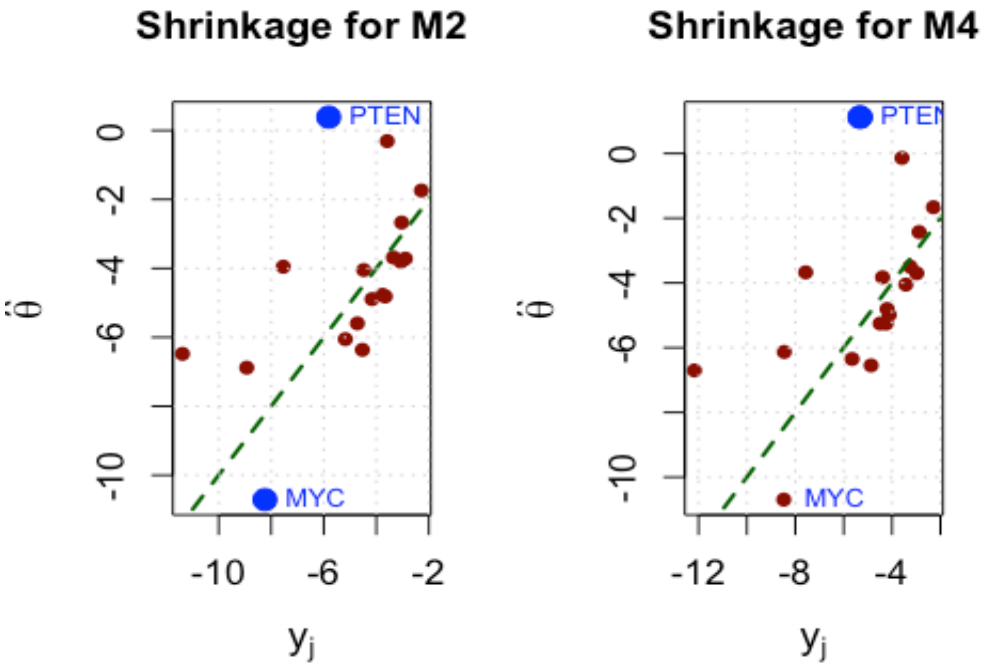

Figure 6

- Interpretation of the Shrinkage plots

If the considered protein is below the shrinkage line, it means that $\overline{y_j}$ smaller than its posterior expectation (*prior mean < posterior mean),* that is to say that the overall mean will negatively influence the value of our posterior mean.

The opposite can be said for the observations that fall over the shrinkage line.