

Bayesian hierarchical regression model for healthcare performance evaluation via complication rates

Vincenzo Roberto Sillitti

This project investigates the factors that may influence the rate of post-operative complications following cholecystectomy procedures in Italian healthcare facilities. Using data from 2023, I have used a Bayesian hierarchical regression model to assess the impact of hospital characteristics and regional differences on surgical outcomes.

1) Introduction

Dataset description

The dataset contains information on the 30-day post-operative complication rate for cholecystectomy procedures performed in various healthcare facilities across Italy. The complication rates are adjusted for sex, age, and pre-existing medical conditions. The dataset includes 336 observations referring to hospitals (including university and research hospitals) and accredited private clinics, disaggregated by Italian region (with the exception of Aosta Valley and Molise). The dataset was obtained from the PNE (National Outcomes Program) website and refers to the year 2023. Additional variables relevant to the analysis include:

Table 1: Covariates included in the model

Covariate	Description
Length of stay shorter than 3 days	Adjusted percentage of patients who were discharged within 3 days
N. interventions	Total number of cholecystectomy procedures performed in each facility
Emergency Room	Indicates whether the facility has an emergency room
Healthcare facility category	Categorical variable: University hospital/IRCCS, Private clinic, Public hospital

The percentage of patients discharged within 3 days can be considered a performance indicator, as the medical literature indicates that discharge typically occurs between 3 and 5 days after this type of intervention. The inclusion of emergency department availability is particularly meaningful, as it may indicate whether a facility handles urgent cases.

Goal of the analysis

The aim of this analysis is to assess whether the selected covariates have an effect on the rate of post-operative complications. In addition, the goal is to investigate potential regional disparities and determine whether some regions perform better than others. For this reason, I opted for a hierarchical regression model. The Bayesian approach is particularly useful and appropriate in this context, as it allows for borrowing strength across groups — especially relevant when some regions have a small number of observations. This leads to more stable and informative estimates compared to a purely frequentist model.

The analysis focuses on a common and relatively simple procedure (cholecystectomy), which is widely performed across healthcare facilities in Italy and can serve as a representative indicator of hospital performance. However, this approach could be extended to other procedures with high incidence or clinical relevance — such as hip fracture surgery in patients over the age of 65 — to assess whether the results are consistent. If they are, this would suggest that observed differences are attributable to the healthcare structure itself; if not, further clinical or systemic explanations may be needed.

Preliminary considerations

Since the response variable is expressed as a percentage, and the normal regression model assumes support on the entire real line, I applied a logit transformation to the 30-day post-operative complication rate. The logit function is particularly appropriate for bounded outcomes (i.e., variables constrained between 0 and 1), as it maps the open interval (0, 1) onto the entire real line, thereby allowing for more appropriate modeling within a linear regression framework.

```
p <- (Data$Adj_percentage_complications)/100
Data$y_logit <- log(p/(1-p))
```

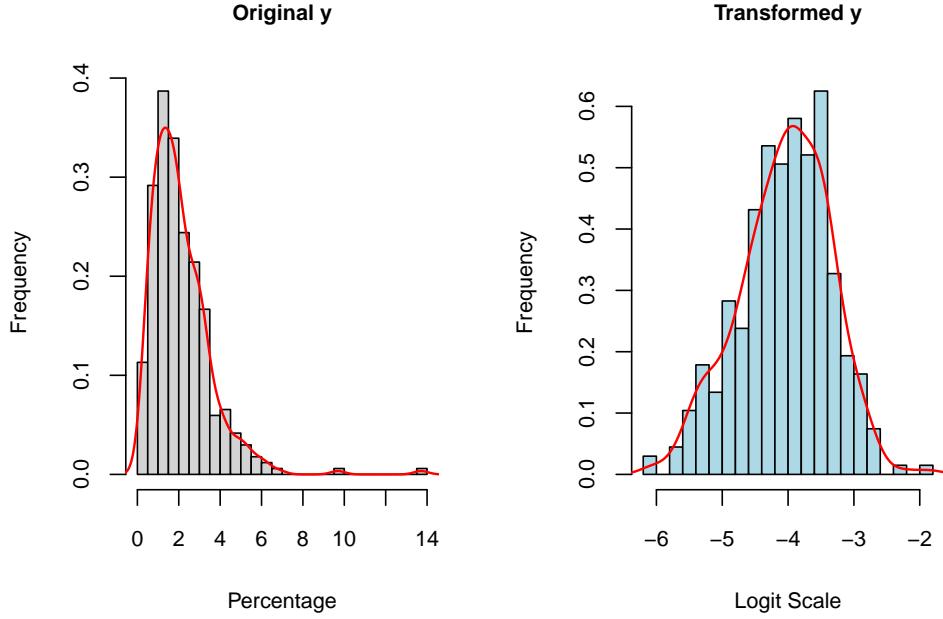


Figure 1: Comparison between original response variable and transformed response variable

The histogram on the left shows that the complication rate (in percentage) is highly right-skewed, with most values clustered between 1% and 3%. After applying a logit transformation (right plot), the distribution becomes approximately symmetric and more bell-shaped, making it more suitable for normal regression models.

Additionally, I applied a logarithmic transformation to the variable representing surgical volume (n. interventions), as its distribution was highly skewed. This transformation also helps improve the linearity of the relationship between surgical volume and the outcome.

2) Model specification

As mentioned, I opted for a two-level hierarchical regression model specified as follows:

Sampling model

$$Y_{1,j}, \dots, Y_{n_j,j} \mid \beta_j, \sigma^2 \sim \mathcal{N}(\mathbf{x}_{i,j}^\top \beta_j, \sigma^2), \quad \text{for } i = 1, \dots, n_j, \quad j = 1, \dots, m$$

where each group j has n_j observations and $\beta_j = (\beta_{1,j}, \dots, \beta_{p,j})$ is a group-specific coefficient vector. Moreover, the data within groups share a common variance σ^2 that does not depend on the group j .

The likelihood is:

$$\begin{aligned}
p(\mathbf{Y}_1, \dots, \mathbf{Y}_m \mid \boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_m, \sigma^2) &= \prod_{j=1}^m \prod_{i=1}^{n_j} p(y_{i,j} \mid \boldsymbol{\beta}_j, \sigma^2) \\
&= \prod_{j=1}^m \prod_{i=1}^{n_j} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y_{i,j} - \mathbf{x}_{i,j}^\top \boldsymbol{\beta}_j)^2\right)
\end{aligned}$$

with σ^2 that denotes the within-group variability with prior:

$$\sigma^2 \sim \text{Inverse-Gamma}\left(\frac{\nu_0}{2}, \frac{\nu_0\sigma_0^2}{2}\right)$$

Prior level 1:

$$\boldsymbol{\beta}_j \mid \boldsymbol{\theta}, \Sigma \sim \mathcal{N}_p(\boldsymbol{\theta}, \Sigma), \quad \text{for } j = 1, \dots, m$$

Prior level 2:

$$\boldsymbol{\theta} \sim \mathcal{N}_p(\boldsymbol{\mu}_0, \boldsymbol{\Lambda}_0) \quad \Sigma \sim \text{Inverse-Wishart}(\eta_0, \mathbf{S}_0^{-1})$$

where Σ denotes the between-group variability.

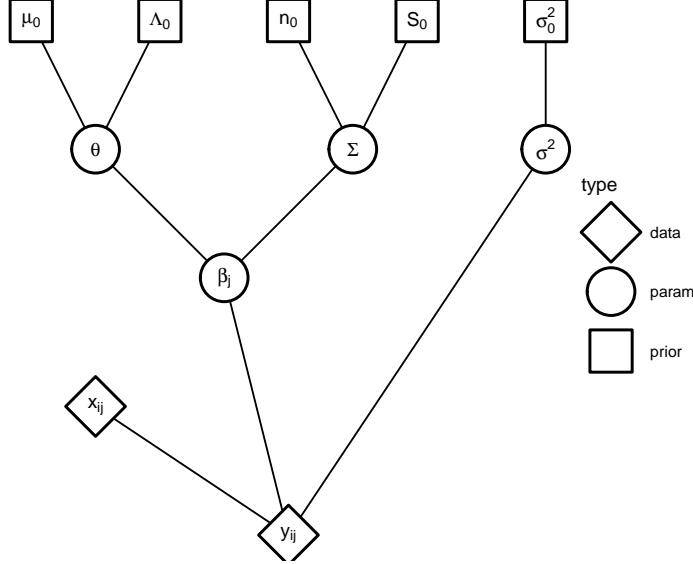


Figure 2: DAG representation of the model

Joint posterior:

$$\begin{aligned}
p(\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_m, \boldsymbol{\theta}, \Sigma, \sigma^2 \mid \mathbf{Y}) &= \prod_{j=1}^m \prod_{i=1}^{n_j} p(y_{i,j} \mid \mathbf{x}_{i,j}^\top \boldsymbol{\beta}_j, \sigma^2) \\
&\quad \times \prod_{j=1}^m p(\boldsymbol{\beta}_j \mid \boldsymbol{\theta}, \Sigma) \\
&\quad \times p(\boldsymbol{\theta}) \cdot p(\Sigma) \cdot p(\sigma^2)
\end{aligned}$$

Full conditional distributions

Since this is a semi-conjugate normal model, the posterior distribution can be approximated using a Gibbs sampler. Below is the formulation of the full conditional distributions required to implement the algorithm.

1. Full conditional for β_j

$$p(\beta_j | \text{rest}) \propto \prod_{i=1}^{n_j} \mathcal{N}(y_{ij} | \mathbf{x}_{ij}^\top \beta_j, \sigma^2) \cdot \mathcal{N}_p(\beta_j | \boldsymbol{\theta}, \boldsymbol{\Sigma})$$

$$\Rightarrow \beta_j | \text{rest} \sim \mathcal{N}_p(\tilde{\mathbf{b}}_j, \tilde{\mathbf{V}}_j)$$

with:

$$\tilde{\mathbf{V}}_j = \left(\frac{1}{\sigma^2} \mathbf{X}_j^\top \mathbf{X}_j + \boldsymbol{\Sigma}^{-1} \right)^{-1}, \quad \tilde{\mathbf{b}}_j = \tilde{\mathbf{V}}_j \left(\frac{1}{\sigma^2} \mathbf{X}_j^\top \mathbf{y}_j + \boldsymbol{\Sigma}^{-1} \boldsymbol{\theta} \right)$$

2. Full conditional for $\boldsymbol{\theta}$

$$p(\boldsymbol{\theta} | \text{rest}) \propto \prod_{j=1}^m \mathcal{N}_p(\beta_j | \boldsymbol{\theta}, \boldsymbol{\Sigma}) \cdot \mathcal{N}_p(\boldsymbol{\theta} | \boldsymbol{\mu}_0, \boldsymbol{\Lambda}_0)$$

$$\Rightarrow \boldsymbol{\theta} | \text{rest} \sim \mathcal{N}_p(\boldsymbol{\mu}_n, \boldsymbol{\Lambda}_n)$$

with:

$$\boldsymbol{\Lambda}_n = (\boldsymbol{\Lambda}_0^{-1} + m \boldsymbol{\Sigma}^{-1})^{-1}, \quad \boldsymbol{\mu}_n = \boldsymbol{\Lambda}_n (\boldsymbol{\Lambda}_0^{-1} \boldsymbol{\mu}_0 + m \boldsymbol{\Sigma}^{-1} \bar{\boldsymbol{\beta}}), \quad \text{with } \bar{\boldsymbol{\beta}} = \frac{1}{m} \sum_{j=1}^m \beta_j$$

3. Full conditional for $\boldsymbol{\Sigma}$

$$p(\boldsymbol{\Sigma} | \text{rest}) \propto \prod_{j=1}^m \mathcal{N}_p(\beta_j | \boldsymbol{\theta}, \boldsymbol{\Sigma}) \cdot \mathcal{IW}(\boldsymbol{\Sigma} | \eta_0, \mathbf{S}_0^{-1})$$

$$\Rightarrow \boldsymbol{\Sigma} | \text{rest} \sim \mathcal{IW}(\eta_n, \mathbf{S}_n^{-1})$$

with:

$$\eta_n = \eta_0 + m, \quad \mathbf{S}_n = \mathbf{S}_0 + \sum_{j=1}^m (\beta_j - \boldsymbol{\theta})(\beta_j - \boldsymbol{\theta})^\top$$

4. Full conditional for σ^2

$$p(\sigma^2 | \text{rest}) \propto \prod_{j=1}^m \prod_{i=1}^{n_j} \mathcal{N}(y_{ij} | \beta_j^\top \mathbf{x}_{ij}, \sigma^2) \cdot \mathcal{IG}\left(\sigma^2 | \frac{\nu_0}{2}, \frac{\nu_0 \sigma_0^2}{2}\right)$$

$$\Rightarrow \sigma^2 | \text{rest} \sim \mathcal{IG}\left(\frac{\nu_n}{2}, \frac{\nu_n \sigma_n^2}{2}\right)$$

with:

$$\nu_n = \nu_0 + \sum_{j=1}^m n_j, \quad \sigma_n^2 = \frac{1}{\nu_n} \left(\nu_0 \sigma_0^2 + \sum_{j=1}^m \sum_{i=1}^{n_j} (y_{ij} - \beta_j^\top \mathbf{x}_{ij})^2 \right)$$

3) Gibbs Sampler

Gibbs Sampling Scheme

Using the full conditional distributions above I have implemented the following Gibbs sampler:

Start from initial values:

$$\sigma^{2(0)}, \boldsymbol{\theta}^{(0)}, \boldsymbol{\Sigma}^{(0)}.$$

For $s = 1, \dots, S$:

1. Update β_1, \dots, β_m :
For each $j = 1, \dots, m$:

$$\beta_j^{(s)} \sim p(\beta_j | Y_{1j}, \dots, Y_{nj}, \sigma^{2(s-1)}, \boldsymbol{\theta}^{(s-1)}, \boldsymbol{\Sigma}^{(s-1)})$$

2. Update $\boldsymbol{\theta}$:

$$\boldsymbol{\theta}^{(s)} \sim p(\boldsymbol{\theta} | \beta_1^{(s)}, \dots, \beta_m^{(s)}, \boldsymbol{\Sigma}^{(s-1)})$$

3. Update $\boldsymbol{\Sigma}$:

$$\boldsymbol{\Sigma}^{(s)} \sim p(\boldsymbol{\Sigma} | \beta_1^{(s)}, \dots, \beta_m^{(s)}, \boldsymbol{\theta}^{(s)})$$

4. Update σ^2 :

$$\sigma^{2(s)} \sim p(\sigma^2 | Y_{11}, \dots, Y_{njm}, \beta_1^{(s)}, \dots, \beta_m^{(s)})$$

Implementation

```
X <- model.matrix(~ Adj_percentage_stay_under_three_days + log(N_interventions) +
  Healthcare_facility_category + ER, data = Data)

split_X <- split(as.data.frame(X), Data$Region)
X_list <- lapply(split_X, as.matrix) # This is the model matrix for each region

y_list <- split(Data$y_logit, Data$Region) # This is the response vector for each region

gibbs_sampler_hierarchical_normal_regression <- function(X_list, y_list,
  mu0, Lambda0,
  eta0, S0,
  nu0, sigma2_0,
  n_iter,
  burn_in = 0,
  thin = 1) {

  m <- length(X_list)      # number of groups (regions)
  p <- ncol(X_list[[1]])    # number of covariates
  save_iter <- floor((n_iter - burn_in) / thin)
  #Total number of saved iterations after burn-in and thinning

  # ---- INITIALIZATION ----

  samples_theta <- matrix(NA, nrow = save_iter, ncol = p)
  samples_sigma2 <- numeric(save_iter)
  samples_beta <- array(NA, dim = c(save_iter, m, p))
```

```

beta_list      <- vector("list", m)
theta   <- rep(0, p)
Sigma    <- diag(1, p)
sigma2 <- 1
save_index <- 1

# ----- ALGORITHM -----

for (t in 1:n_iter) {
  # Step 1: Sample beta_j
  for (j in 1:m) {
    Xj <- X_list[[j]]                                # model matrix for region j
    yj <- matrix(y_list[[j]], ncol = 1)              # response vector for region j
    XtX <- t(Xj) %*% Xj
    XtY <- t(Xj) %*% yj
    V_j <- solve(solve(Sigma) + XtX / sigma2)
    b_j <- V_j %*% (solve(Sigma) %*% theta + XtY / sigma2)
    beta_list[[j]] <- mvtnorm(1, b_j, V_j)}

  # Step 2: Sample theta
  beta_mat <- do.call(rbind, beta_list)
  beta_bar <- colMeans(beta_mat)
  Lambda_m <- solve(solve(Lambda0) + m * solve(Sigma))
  mu_m     <- Lambda_m %*% (solve(Lambda0) %*% mu0 + m * solve(Sigma) %*% beta_bar)
  theta    <- mvtnorm(1, mu_m, Lambda_m)

  # Step 3: Sample Sigma
  S_theta <- matrix(0, p, p)
  #S_theta is the sum of (beta_j - theta)(beta_j - theta)^T across all groups
  for (j in 1:m) {
    diff <- matrix(beta_list[[j]] - theta, ncol = 1)
    S_theta <- S_theta + diff %*% t(diff)}
  Sigma <- rinvwish(df = eta0 + m, scale = S0 + S_theta)

  # Step 4: Sample sigma^2
  SSR <- 0
  for (j in 1:m) {
    Xj <- X_list[[j]]
    yj <- matrix(y_list[[j]], ncol = 1)
    bj <- beta_list[[j]]
    resid <- yj - Xj %*% bj
    SSR <- SSR + sum(resid^2)}
  shape <- (nu0 + sum(sapply(y_list, length))) / 2
  rate  <- (nu0 * sigma2_0 + SSR) / 2
  sigma2 <- 1 / rgamma(1, shape = shape, rate = rate)

# ----- RESULTS -----

# Save samples if past burn-in and at thinning interval
if (t > burn_in && ((t - burn_in) %% thin == 0)) {
  samples_theta[save_index, ] <- theta
  samples_sigma2[save_index] <- sigma2

```

```

    for (j in 1:m) {
      samples_beta[save_index, j, ] <- beta_list[[j]]
    }
    save_index <- save_index + 1}

return(list(theta = samples_theta,
            sigma2 = samples_sigma2,
            beta = samples_beta))}
```

The Gibbs sampler was implemented using the following parameter settings:

```

results <- gibbs_sampler_hierarchical_normal_regression(
  X_list = X_list,      y_list = y_list,
  mu0     = rep(0, p),   Lambda0 = diag(1, p),
  eta0    = p + 2,       S0 = diag(1, p),
  nu0     = 3,           sigma2_0 = 0.2,
  n_iter  = 250000,     burn_in = 100000,   thin = 10)
```

The priors chosen reflect weakly informative beliefs to support convergence without imposing strong constraints. Specifically, $\mu_0 = \text{rep}(0, p)$ and $\Lambda_0 = \text{diag}(1, p)$ define a normal prior centered at 0 with unit precision for the group-level means. The hyperparameter $\eta_0 = p + 2$ is the value of the inverse-Wishart prior on the group-level covariance matrix, which is weakly informative but proper. A prior variance of $\sigma_0^2 = 0.2$ was chosen to reflect moderate variability within groups. I set the number of iterations equal to 250000, with burn in = 100000 and thin = 10 to reduce autocorrelation in the MCMC chains. These settings were selected after inspecting ACFs, which indicated slow mixing. In this way the effective sample size became better too. However, the final number of samples after burn in and thinning is 15000.

4) Convergence Diagnostics

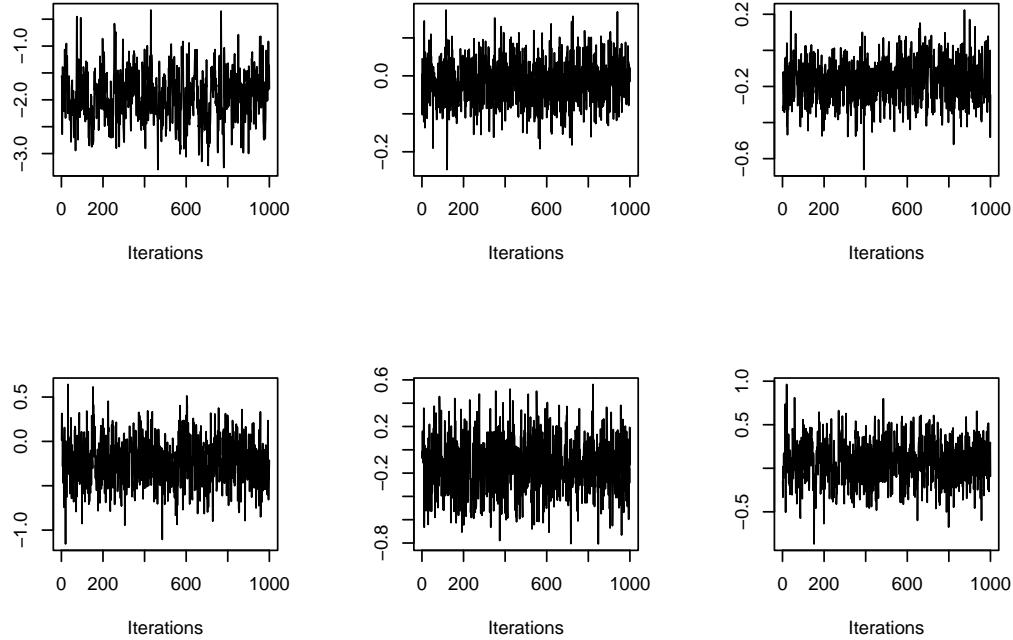


Figure 3: Trace plots for θ

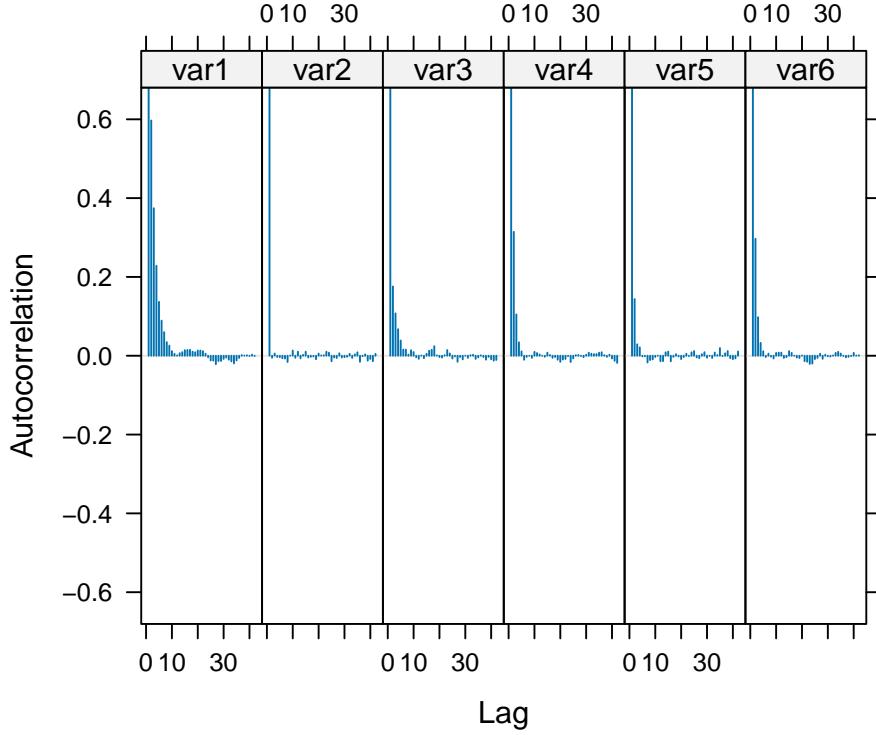


Figure 4: ACF plots for θ

To ensure the validity of posterior inference, I examined the convergence of the chains for the global parameter vector θ , which represents the population-level means of the regression coefficients. First, I visually inspected the trace plots for all components of θ (I inspected just 1000 iterations to see better if there are any issues). The chains appear to have good mixing, showing no signs of trends. The sampled values fluctuate steadily around a stable mean and cover the posterior support well, indicating that the Gibbs sampler explored the target distribution effectively. These plots suggest that convergence has been reached and that the posterior samples are reliable for inference. Secondly, I analyzed the autocorrelation functions (ACF) for all components of θ up to lag 40. As shown in the ACF plots most parameters (especially var2 and var4 to var6) exhibit rapid decay in autocorrelation, with values approaching zero after only a few lags. Some parameters (e.g., var1 and var3) show slightly slower decay, suggesting mild autocorrelation in early lags. However, no component shows problematic autocorrelation or persistent correlation patterns.

Table 2: ESS and Geweke test

Covariate	ESS	Geweke_Z
var1 Intercept	3577.6	-0.831
var2 Adj percentage stay under three days	15000.0	-2.551
var3 log(N. interventions)	8087.0	1.023
var4 Healthcare facility: Public	7821.4	-0.625
var5 Healthcare facility: University hospital / IRCCS	10682.5	-1.207
var6 ER: Yes	8134.0	0.211

I also computed the Effective Sample Size (ESS) and the Geweke diagnostic Z-scores. All parameters exhibit high ESS values, indicating that the Markov chains have mixed well and that the posterior estimates are based on a large number of effectively independent samples. However, the ESS for the intercept is relatively lower, which is modest given the total of 15,000 samples. This may reflect slightly higher autocorrelation for

the intercept compared to other parameters. The Geweke Z-scores for almost all variables lie well within the acceptable range of $[-1.96, +1.96]$, suggesting no significant difference between the means of the initial (10%) and last (50%) portions of the chains. This provides strong evidence of convergence for each parameter.

The MCMC diagnostics for the group-level regression coefficients β_j is also necessary to see if there is convergence and good mixing across all groups and covariates. For example, the plots below display the ACFs for β_j in Group 1 and Group 2.

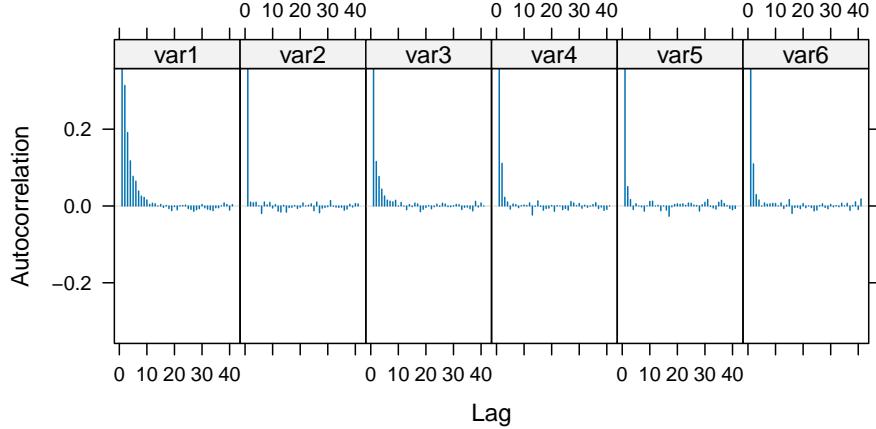


Figure 5: ACF plots of β_j for Group 1

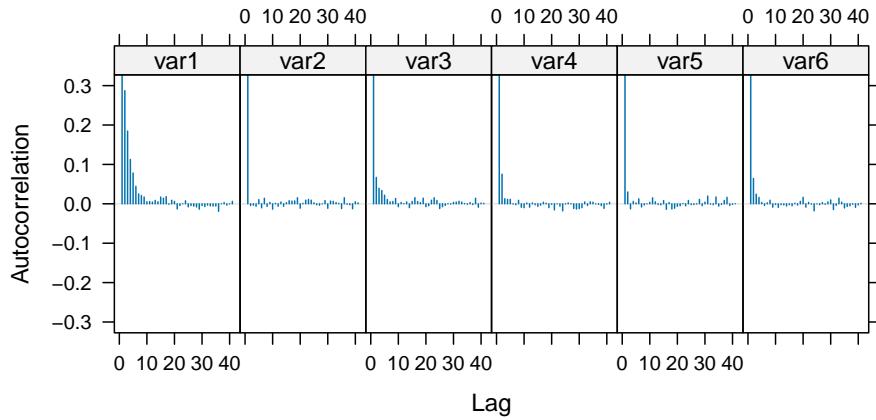


Figure 6: ACF plots of β_j for Group 2

Table 3: ESS and Geweke test for Group 1

Covariate	ESS	Geweke_Z
var1 Intercept	5462.8	-0.686
var2 Adj percentage stay under three days	15000.0	0.048
var3 log(N. interventions)	9598.8	0.582
var4 Healthcare facility: Public	12002.3	0.391
var5 Healthcare facility: University hospital / IRCCS	13182.9	-1.056
var6 ER: Yes	11616.9	-0.822

The ACF plots for both Group 1 and Group 2 show a moderate level of autocorrelation in the early lags, particularly for some covariates (the intercept, but also var 3 and 6). However, autocorrelation declines

steadily and becomes negligible beyond lag 10-15, suggesting acceptable mixing of the chains. While not optimal, the overall sampling efficiency remains adequate for posterior inference. To conclude, I also checked the trace plot, ACF and the ESS of σ^2 .

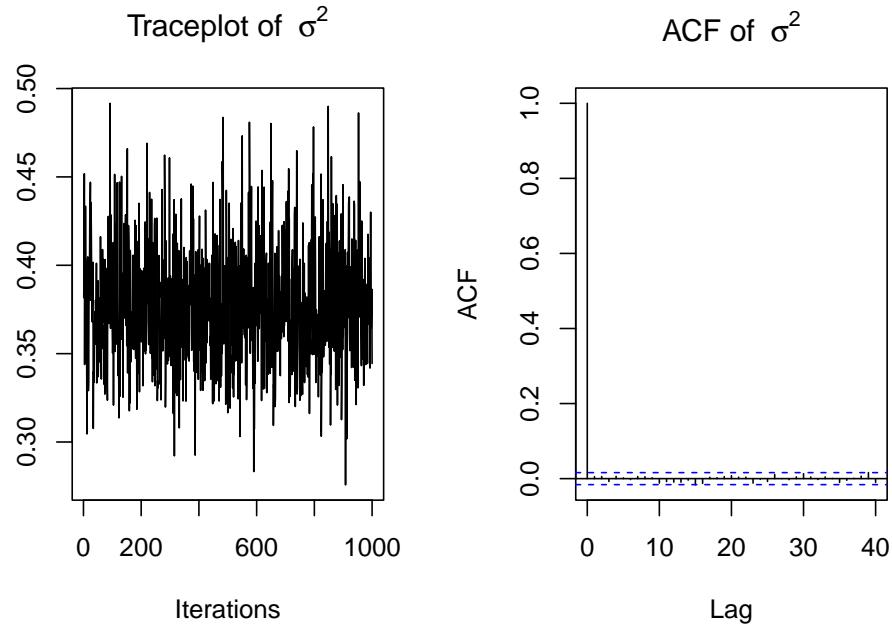


Figure 7: Trace plot and ACF plot of σ^2

```
ess_sigma2
```

```
##   var1
## 15000
```

5) Interpretation of the results

In the table below are shown the posterior mean, standard deviation and credible interval for θ for all covariates.

Table 4: Summary stastistics for θ

Covariate		Mean	SD	CI_2.5	CI_97.5
var1	Intercept	-1.9086444	0.4782540	-2.8373295	-0.9451964
var2	Adj percentage stay under three days	-0.0132164	0.0552732	-0.1208817	0.0960466
var3	log(N. interventions)	-0.1758787	0.1153955	-0.4029763	0.0506798
var4	Healthcare facility: Public	-0.2193068	0.2532998	-0.7209424	0.2839223
var5	Healthcare facility: University hospital / IRCCS	-0.1283337	0.2298507	-0.5784808	0.3245298
var6	ER: Yes	0.0683320	0.2380694	-0.3959297	0.5316339

The expected value of the logit of the adjusted percentage of complications is estimated to be about -1.91 when all predictors are set to zero. This estimate corresponds to the posterior mean $\mathbb{E}[\theta_0 | \text{data}]$. The other coefficients can be interpreted as the (positive or negative) change in the response variable. However, the 95% credible intervals for all covariates include zero, suggesting a lack of strong evidence for any global effect at the population level.

The greatest posterior uncertainty is observed for the intercept, as indicated by its larger standard deviation. Despite this, its credible interval does not include zero, pointing to a significant global baseline effect. All other covariates have wide credible intervals centered near zero, which implies substantial posterior uncertainty and limited overall influence at the global level.

The model was estimated using weakly informative priors, chosen to allow sufficient flexibility while avoiding extreme parameter values. Although these priors are reasonable, the posterior inference revealed that many credible intervals include zero, suggesting limited support for strong covariate effects.

However, since θ represents the global (population-level) effect, it is still possible for individual groups to exhibit some deviations. For example, the table below reports the same summary statistics computed for β_j in the first group (Piedmont). In this case, the second variable (adjusted percentage of patients discharged within three days after the intervention) shows a significant negative effect on the response variable, as its 95% credible interval does not include zero. So, in Piedmont, an increase in the percentage of patients discharged early is associated with a lower number of complications. The remaining covariates still have credible intervals that cover zero, indicating non-significant effects. Importantly, this result is specific to Group 1 and should not be generalized without examining the other groups individually. Similar results are observed in Group 3 (Trentino-Alto Adige), Group 7 (Emilia-Romagna), and Group 13 (Campania), where the same covariate appears to have a significant effect. Additionally, the logarithm of the total number of interventions has a significant negative influence in Group 4 (Veneto) and Group 6 (Liguria).

Table 5: Posterior summary statistics for β_j for Group 1

Covariate	Mean	SD	CI_2.5	CI_97.5
Intercept	-2.003	0.663	-3.344	-0.730
Adj percentage stay under three days	-0.025	0.009	-0.043	-0.007
log(N. interventions)	-0.020	0.184	-0.376	0.345
Healthcare facility: Public	-0.167	0.436	-1.032	0.703
Healthcare facility: University hospital / IRCCS	-0.136	0.487	-1.103	0.851
ER	0.112	0.429	-0.738	0.963

Table 6: Posterior summary for error variance σ^2

Mean	CI_2.5	CI_97.5
0.375782	0.3191647	0.4426118

To conclude, the posterior distribution of σ^2 appears well concentrated, suggesting that the data provided sufficient information for a stable estimation of variance.

6) Posterior predictive checks

Posterior predictive checks (PPCs) are used to evaluate whether the model is capable of generating data that resemble the observed data. By comparing simulated values with the observed data we can assess model fit and identify potential discrepancies.

```
y_rep_list <- vector("list", m)
for (j in seq_len(m)) {
  Xj <- X_list[[j]]
  nj <- nrow(Xj)
  y_rep_mat <- matrix(NA, nrow = S, ncol = nj)

  for (s in 1:S) {
```

```

beta_j_s <- results$beta[s, j, ]
mu_j_s <- Xj %*% beta_j_s
y_rep_mat[s, ] <- rnorm(nj, mean = mu_j_s, sd = sqrt(results$sigma2[s]))
}
y_rep_list[[j]] <- y_rep_mat

# Computation of the simulated means for each group
mean_rep_list <- lapply(y_rep_list, function(mat) apply(mat, 1, mean))
mean_obs <- sapply(y_list, mean)
# Computation of the simulated variance for each group
var_rep_list <- lapply(y_rep_list, function(mat) apply(mat, 1, var))
var_obs <- sapply(y_list, var)

```

The plot below shows the estimated density of the observed data (black line) against the posterior predictive distributions (blue lines), providing a global goodness-of-fit check.

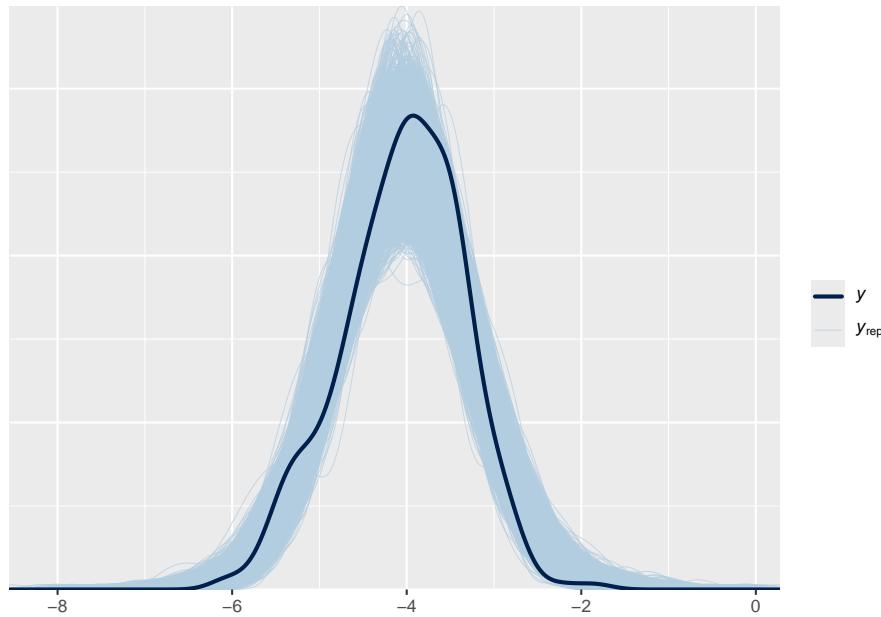


Figure 8: Observed data against posterior predictive distributions

Predictive checks by groups

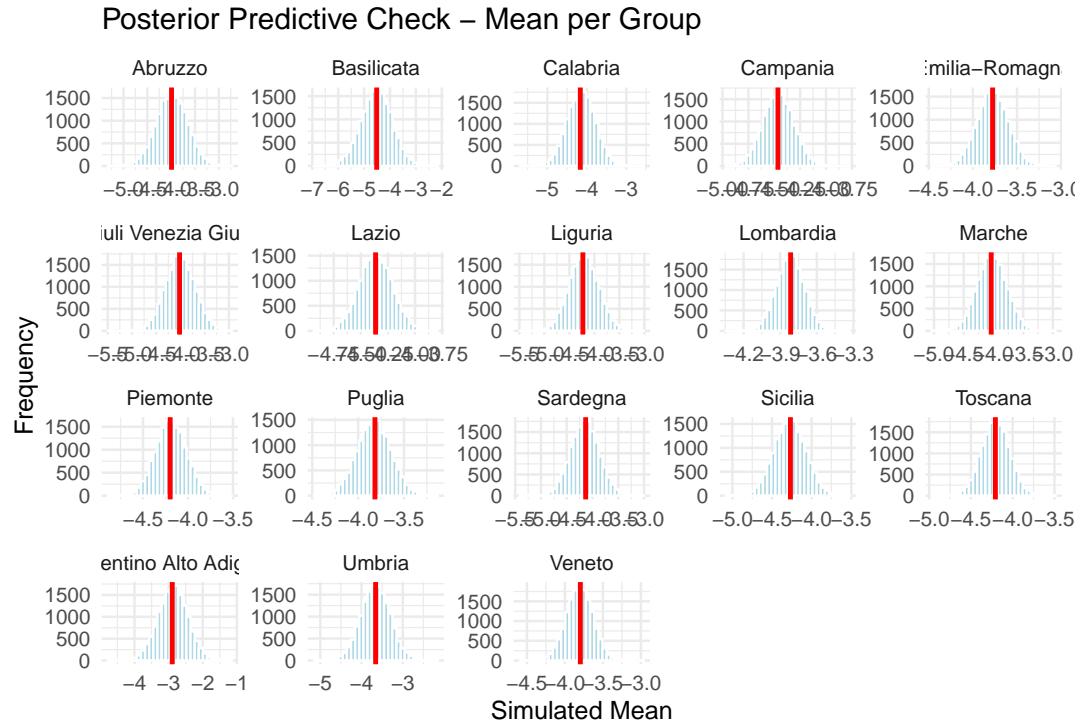


Figure 9: PPC (mean)

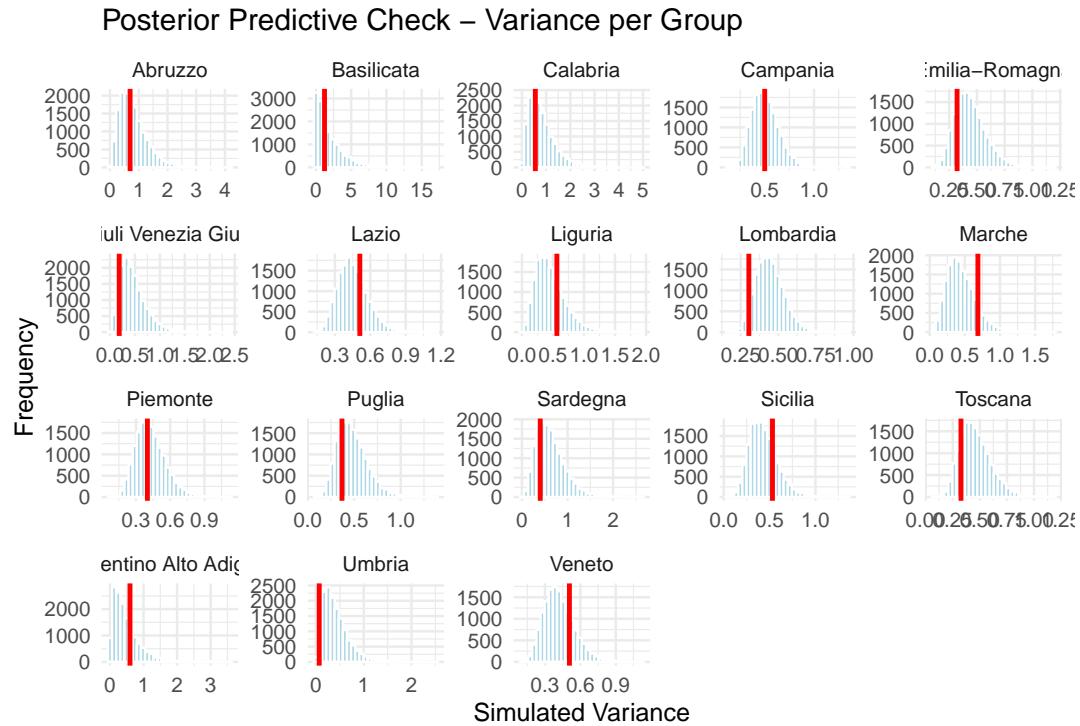


Figure 10: PPC (variance)

The two PPC plots compare the observed group-level means and variances (red lines) against the distributions of simulated values under the model (blue histograms). The observed means (red lines) fall near the center of the simulated distributions in almost all groups, suggesting a good overall model fit in terms of average outcome prediction. The simulated distributions for the mean are very similar across regions. This could imply: strong pooling induced by the hierarchical prior on $\beta_j \sim N(\theta, \Sigma)$ or limited variability in covariate effects across groups.

The posterior predictive distributions for the variance differ more across groups, capturing some heterogeneity. However, in some regions (e.g. Basilicata, Trentino Alto Adige), the variance is highly overestimated or shows very high posterior dispersion. Moreover, the long right tails in many histograms suggest that the posterior of σ^2 is skewed.

6) Conclusion

In summary, while the model reproduces the means relatively well, it struggles more with variance, particularly in certain regions. Tuning priors and improving hierarchical structure can enhance model fit and predictive accuracy, for example by adding covariates to explain some of the residual variance but also group-level covariates to better capture heterogeneity among groups. Additionally to improve inference, one possible strategy is to increase the informativeness of the prior on σ^2 , for example by centering it on empirical estimates or incorporating external knowledge. Even though the credible intervals include zero, the model appears to perform well: it predicts the mean accurately and does not perform so poorly in terms of variance. The diagnostics for all parameters are also satisfactory. However, some adjustments and refinements are needed, and to obtain more valuable results, it would be useful to replicate the analysis using different surgical interventions, as mentioned at the beginning

Possible extention of the model

Finally, it is worth mentioning that the model can be extended by allowing for group-specific error variances σ_j^2 , rather than assuming a common error variance across all groups.

In this case the full conditional distribution for the group-specific error variance σ_j^2 is given by:

$$\sigma_j^2 | rest \sim \text{Inverse-Gamma} \left(\frac{\nu_0 + n_j}{2}, \frac{\nu_0 \sigma_0^2 + \sum_{i=1}^{n_j} (y_{i,j} - \mathbf{x}_{i,j}^\top \boldsymbol{\beta}_j)^2}{2} \right)$$

independent for each group j.

It is then possible to modify the Gibbs sampler and proceed as in the standard model.

References

- AGENAS – Agenzia Nazionale per i Servizi Sanitari Regionali. (2024). Programma Nazionale Esiti – Dataset indicatori esiti [Dataset]. <https://pne.agenas.it/ospedaliera/indicatori/28?tab=italia&mode=0>
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). Bayesian Data Analysis (3rd ed.). CRC Press.
- Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1(3), 515–533. <https://sites.stat.columbia.edu/gelman/research/published/taumain.pdf>