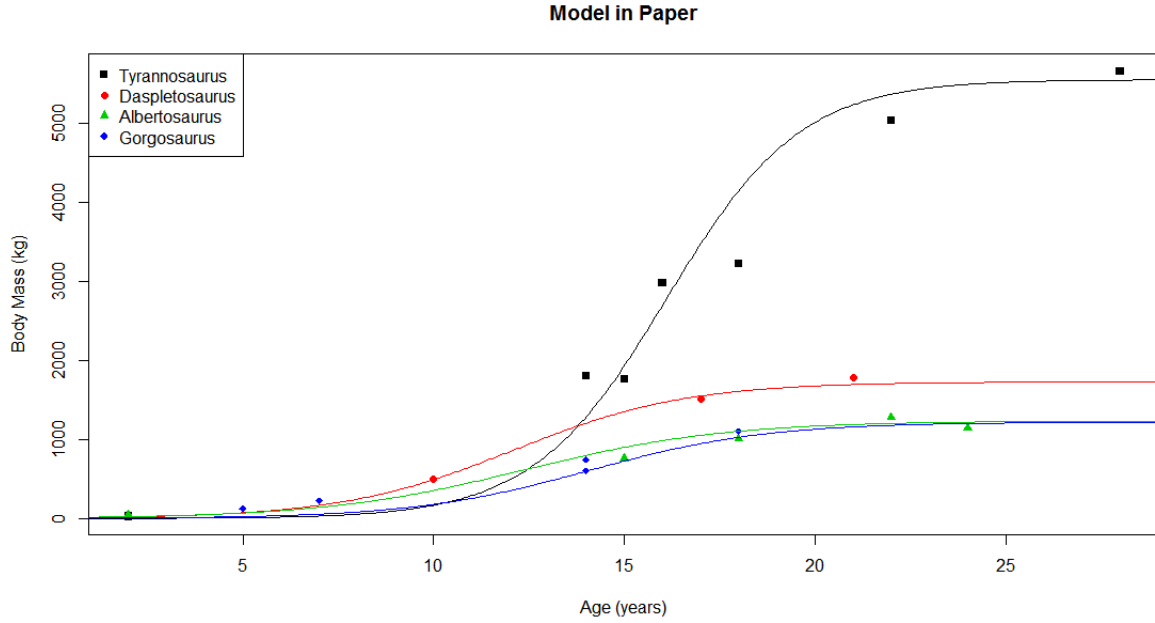


ST495/590 Assignment 9 - Solutions

- (1) Download the data from Table 1 and reproduce Figure 2 (sans cheesy dinosaurs).



- (2) Describe and fit a separate model for each species.

Let $i = 1, 2, 3, 4$ denote the four species (A, D, G, T), $j = 1, 2, \dots, n_i$ denote the j th dinosaur of the i th species, Y_{ij} denote the body mass (kg), X_{ij} denote the age. Let $IG(\alpha, \beta)$ denote an inverse gamma distribution with shape and scale parameter α and β respectively.

Separate model for each species i :

$$Y_{ij} \sim N(\mu_{ij}, \sigma_i^2), \quad \mu_{ij} = \frac{a_i}{1 + \exp(-b_i(X_{ij} - c_i))} + 5,$$

priors:

$$a_i \sim N(\mu_{ai}, \sigma_{ai}^2), \quad b_i \sim \text{Uniform}(0, 1), \quad c_i \sim N(\mu_{ci}, \sigma_{ci}^2), \quad \sigma_i^2, \sigma_{ai}^2, \sigma_{ci}^2 \sim \text{IG}(e, e),$$

where e is some small value, say 0.01, μ_{ai} and μ_{ci} are hyperparameters. In this data set, since we only have a few observations, the posterior is quite sensitive to the priors. So we use informative priors, i.e., making μ_{ai} and μ_{ci} close to the “truth”, the number

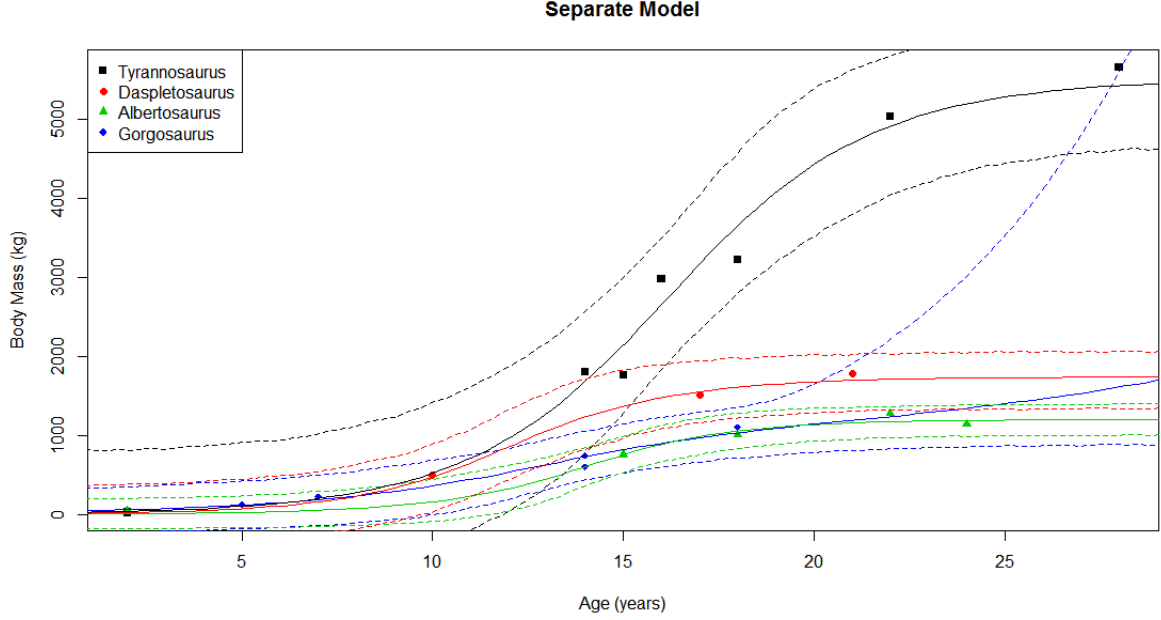


Table 1: Posterior mean, 2.5% and 97.5% prediction of body mass at age 30 for separate model.

Species	Mean	2.5%	97.5%
A	1205.082	1003.416	1405.284
D	1740.142	1018.445	2265.397
G	1206.9306	895.0475	1573.5506
T	5472.978	4661.896	6281.214

in the logistic growth curve in the paper. This makes sense, as the parameters a_i , b_i and c_i might have some scientific explanations, hence then a informative prior is reasonable.

- (3) *Describe and fit a hierarchical model for all species.*

Hierarchical model:

$$Y_{ij} \sim N(\mu_{ij}, \sigma^2), \mu_{ij} = \frac{a_i}{1 + \exp(-b_i(X_{ij} - c_i))} + 5,$$

priors:

$$a_i \sim N(\mu_a, \sigma_a^2), b_i \sim \text{Uniform}(0, 1), c_i \sim N(\mu_c, \sigma_c^2), \sigma^2, \sigma_a^2, \sigma_c^2 \sim \text{IG}(e, e).$$

Here we assume for different species, the parameters in the model, a_i , b_i and c_i are still different, but are given the same priors. (We can still assume they follow separate priors same as (2), however, for simplicity, here I use this assumption.) We assume a common variance σ^2 for all individuals regardless of species. μ_a and μ_c are the hyperparameters, and again given some informative value. Here, I just make $\mu_a = 3000$ and $\mu_c = 10$.

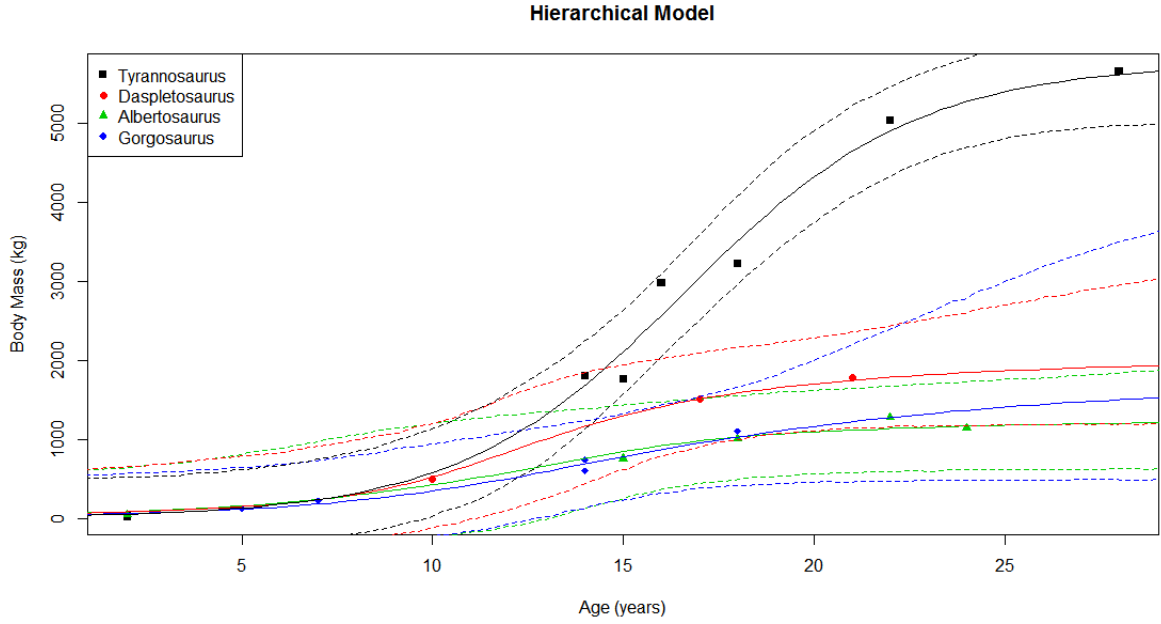


Table 2: Posterior mean, 2.5% and 97.5% prediction of body mass at age 30 for hierarchical model.

Species	Mean	2.5%	97.5%
A	1219.0259	635.9104	1903.5904
D	1951.983	1203.036	3111.414
G	1548.9581	494.0203	3768.9934
T	5690.578	5016.770	6338.995

(4) *Compare the two models.*

In the hierarchical model, we have less parameters to estimate compared to the separate model. To estimate these variance components, all data are used instead of the individuals for only one species. More observations to estimate less parameters, thus leading to a more stable model than the separate model. So the hierarchical model is less sensitive to the priors than the separate model. However, whether assuming a common variance for all the species makes sense depends on scientific explanations. The separate model leads to a narrower predictive credible interval, which might be due to a better estimate of the variance term at each species. Again, both models are quite subjective, intensely depending on the informative priors. Since there are only a few observations in this example, the estimation of neither model is quite accurate. More scientific background and more observations will be helpful.