

Analyzing the Height of Pre-adolescent Girls: A Regression Approach

Test ID: 6494¹

Department of Statistical Science, UC Santa Cruz¹

Abstract

As a hot topic in biology, scientists care about the height of children, especially girls, who may receive less attention while growing up. Children's height is usually assumed to be related to their age and genetic factor. In this report, we estimate the influence of age and genetic factors to the height of girls, using a dataset that tracing 20 pre-adolescent girls' height from their age 6 to 10. The genetic factor is measured by the group of their mother's height, forming as an ordinal categorical variable. We propose several linear models, using the girls' height as the response variable and the age as covariates. We also consider the effect of either the grouping of their mother's height or their personal condition as well. Our goal is achieved by calibrating the regression coefficients in the model. We assess the model from both the traditional frequentist approach and the Bayesian approach. Modeling checking and comparison are also performed to make sure we have obtained useful conclusions. Based on our analysis, girls' height does differ by the group of their mothers' height. Taller mothers tend to have taller children as well. However, this genetic factor does not affect the rate of growth with respect to age. Considering only the age and group factor, all girls increase 5.716cm annually in the age from 6 to 10. However, when considering the growth curve with respect to age for each child, both intercept and slope coefficients differ by individual. The rate of growth can be as fast as 5.403cm while as slow as 4.866cm annually.

KEY WORDS: Linear Regression Analysis, Random Effect Model, Hierarchical Model, Bayesian Data Analysis.

1. Introduction

The study of height of children has long been a hot topic in biology. Biologists usually assume that the height of a child is related to their age and some genetic factors, and they are interested in quantifying the influence of those factors on pre-adolescent girls' height, as it is essential in the health of our next generation. In statistics, such a problem is usually done by calibrating linear regression models. In this report, we have studied a specific problem in this field, the relationship of girls' height with respect to their age with the impact of either a factor indicating the height group of their mother or a the label of girls,

representing their personal condition. The dataset we use has been initially analyzed in [3], where simple linear regression model with age, group effect and their interactions are concerned as the covariates. Further extension of the regression analysis on this dataset by the linear mixed model can be found in [4]. Based on the previous research, we intend to study the effect of age, group and their interactions on the height of girls, based on both simple linear regression model with classic frequentist inference approach and the impact of age on the height individually, through hierarchical regression model with Bayesian approach. In both cases, we will focus on the inference of regression coefficients.

As to proceed, we start with an in-depth exploration of the dataset. We first plot the histogram of girls' height in Figure 1. The distribution of the height seems to be symmetric with respect to the centre, which indicates that no transformation is needed. The empirical distribution of girl's height has mean 128.26 cm, median 127.75 cm and standard deviation 9.50.

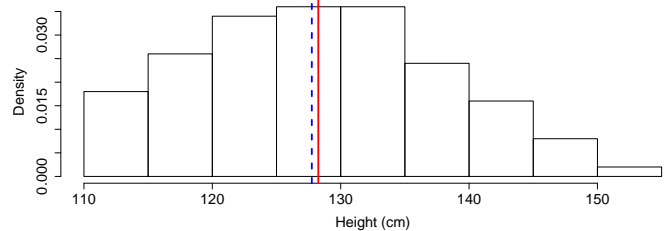


Figure 1: Empirical histogram formulated by the height of girls, with the red solid line indicating the mean height and the blue dashed line representing the median of height among all heights.

The two covariates in the dataset are the age of children and the height group of their mother. The age variable has 5 integer values, range from 6 to 10. The height group of mother is an ordinal categorical variable with 3 levels, where level 1, 2, and 3 indicates the girl's mother was categorized as short, medium and tall, respectively. A more interesting analysis would be to visualize the distribution of heights after grouping. Figure 2 shows the distribution of heights grouped by the age of girls as well as the height grouping of their mothers in box plots. This plot shows that the distribution of height after group differs much in the sense of mean. Elder children and the group of children whose mothers are taller tend to be higher.

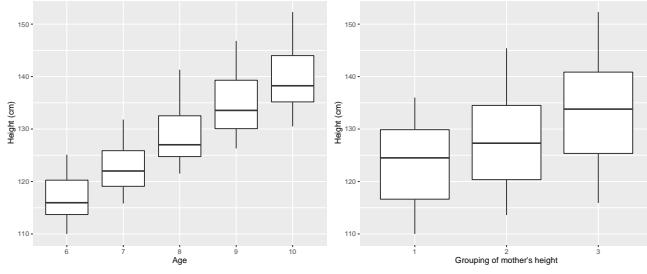


Figure 2: Box plot for the distribution of heights grouped by age (left panel) and height grouping of mothers (right panel).

Besides the distribution of single covariate with response variable, one may also be interested in the joint effect of two factors to the response variable. This information is shown by the scatter plots in Figure 3. We plot the numerical covariate, girls' age, against the girls' height. The color indicates the grouping of mothers' height. We also perform simple linear regression models for the variables on the horizontal and vertical axis and allow the intercept and slope to differ with respect to the group. The plot suggests that the height of girls has a positive correlation with age. The intercept parameter in the linear function may differ with the height grouping of mothers, while the slope parameter may remain the same, as the three regression lines are almost parallel. Noticing the common sense in biology, the rate of growth may not be impacted by genetic factor do make sense. Thus, the result suggested in Figure 3 should not be of too much surprise. However, such conclusion should be tested with more formal models.

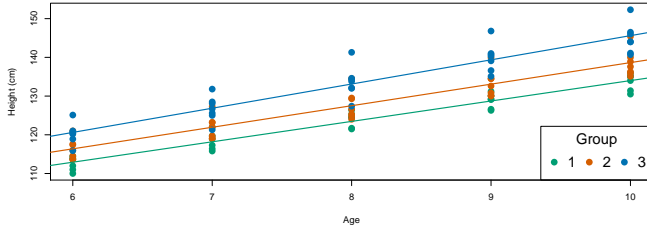


Figure 3: Scatter plot of height of girls with age. The regression line is fitted using simple linear regression model with group dependent intercept and slope parameters.

Finally, for each individual girl k , $k = 1, \dots, 20$, we plot the trace of their height for age from 6 to 10 in Figure 4. This plot suggests that although most of the girls have a growth curve that seems to have the same rate, there are also girls with a more steep growth curve. Therefore, one may be interested in considering linear regression models with coefficients differing by child.

Since the main goal is to examine the effect of age, with either grouping of mother's height or personal conditions, to the height of girls, as suggested in [1,6], linear re-

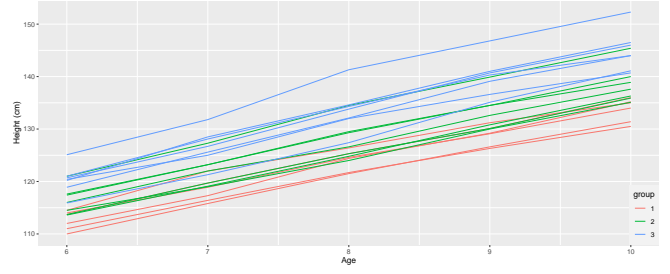


Figure 4: Trace plot of each individual girl's height with respect to their age.

gression model is appropriate. Such a model have good interpretability such that from inference of regression coefficients, we can easily draw conclusions. In assessing such models, we have two possible approaches, namely the frequentist approach and the Bayesian approach. The frequentist approach, with details in [10], is classic in addressing such a problem. The simple linear regression model is readily implementable while both point and interval estimation of model parameters have a simple closed-form solution. This modeling and inference procedure is appropriate for modeling the effect of both age and group since there are not too many parameters to be estimated. However, for a more complicated regression model, such an approach may be tedious to implement or even prohibitive. For example, when considering each individual's impact on the intercept and slope of growth curve, simple linear regression model would require to estimate 60 parameters. With only 100 data points, such an estimate has serious problems. On the other hand, Bayesian approach, although shown in the literature later, is more applicable under complicated model settings. With hierarchical modeling scheme, one can achieve flexible regression model while the inference is relatively easy. Meanwhile, Bayesian approach is also appealing in the sense that it will not only provide a point estimation but also with a full probability model to quantify the uncertainty of the estimation through posterior inference. Hence, we tend to include Bayesian regression analysis to extend the model. Bayesian analysis of linear models are also well studied in the literature, such as [1,2,7]. The decision making and model checking with respect to such models can be found in [1,5,8].

The rest of this report is organized as follows. We will start with inference based on simple linear regression model with frequentist approach in Section 2. Then, the inference will be done with the linear regression model using Bayesian approach in Section 3. Section 4 considers extension of the model using a hierarchical modeling scheme. In all three sections of analyzing models, both inference of model parameters and validation of model is performed. The extended hierarchical model is compared with the original model in Section 5. Finally, section 6 is the concluding remarks.

2. Simple Linear Regression Model with Frequentist Approach

2.1 Baseline Model

To start with, we consider the simple linear regression model with only grouping of mothers' height as the co-variate. The model is stated in (1), where y_{ij} denote the j -th height observation in group i , with $i = 1, 2, 3$, and $j = 1, \dots, n_i$, where $n_1 = 30$, $n_2 = 35$ and $n_3 = 35$.

$$\mathcal{M}_1 : y_{ij} = \mu + \delta_i + \epsilon_{ij} \quad (1)$$

The corresponding error distribution is assumed to be $\epsilon_{ij} \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$. For identifiable issue, we fix δ_1 at 0. The point and interval estimate of regression coefficients is provided in Table 1.

Table 1: Summary of inference result for model \mathcal{M}_1 .

Dependent variable:	
Height	
δ_2	4.051* (-0.209, 8.312)
δ_3	9.651*** (5.391, 13.912)
μ	123.460*** (120.334, 126.586)
R^2	0.172
Adjusted R^2	0.155
$\hat{\sigma}$	8.736 (df = 97)

Note: *p<0.1; **p<0.05; ***p<0.01

The number in the same row as the name of the parameter is the point estimate of that parameter, while the numbers in the bracket below provide a 95% confidence interval of the parameter. The annotation of stars next to the point estimation indicates the region of p -value of the corresponding t test with the null hypothesis been that the parameter is 0. Therefore, at a 95% confidence level, children in group 3 are generally higher than girls in the other two groups. We also notice that this model is not reasonable in explaining the girls' height, as the R^2 is quite small. Regardless, as a baseline model, this model is fine. It does explain more variability than the model using merely the mean of response variable.

2.2 Linear Regression with Group Based Intercept and Slope

Based on the exploratory data analysis result, a more reasonable linear regression model would be to consider both factors and their interactions in the linear regression function, as stated in (2). Here y_{ijt} is the height of girl j in group i at age $5 + t$, for $i = 1, 2, 3$, $j = 1, \dots, m_i$ and $t = 1, \dots, 5$, where $m_1 = 6$, $m_2 = 7$ and $m_3 = 7$.

$$\mathcal{M}_2 : y_{ijt} = \alpha_i + \beta_i(5 + t) + \epsilon_{ijt} \quad (2)$$

The error distribution is also assumed to be $\epsilon_{ijt} \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$.

A more clear parameterization of model \mathcal{M}_2 is stated in (3), it is also the default parameterization used in R built-in function *lm*.

$$y_{ijt} = a_0 + a_i + b_0(5 + t) + b_i(5 + t) + \epsilon_{ijt} \quad (3)$$

for $i = 1, 2, 3$, $j = 1, \dots, m_i$ and $t = 1, \dots, 5$. To make the model identifiable, we need to fix $a_1 = b_1 = 0$. By the variable transformation defined in (4), the two models are equivalent.

$$\begin{cases} \alpha_i = a_0 + a_i & i = 1, 2, 3 \\ \beta_i = b_0 + b_i & i = 1, 2, 3 \end{cases} \quad (4)$$

Fit the model in R, we have the summary statistics of re-parametrized model in Table 2. Based on that, we can further get inference on the original parameter based on (4) and the result is summarized in (5).

$$\begin{cases} \hat{\alpha}_1 = 81.300, & \hat{\alpha}_2 = 82.974, & \hat{\alpha}_3 = 83.123 \\ \hat{\beta}_1 = 5.270, & \hat{\beta}_2 = 5.567, & \hat{\beta}_3 = 6.249 \end{cases} \quad (5)$$

Based on Table 2 and (5), we have the effect of age to the height is 5.270 at mean level. It is also significant at 95% confidence level. In addition, the effect of group 1, 2 and 3 is 81.300, 82.974 and 82.123, respectively. The group factor also have effect on the rate of height growth with respect to age. If the girl is in group 2, she will grow 0.297cm more than the girl in group 1, while if the girl is in group 3, such increase is 0.979cm.

We also notice that by fitting model \mathcal{M}_2 , the difference in both the intercept and slope at the group level is not statistically significant at 95% confidence level. Thus, considering the interaction of the two factors may be too much. We tend to model \mathcal{M}_3 where only the marginal effects of the two factors are considered.

$$\mathcal{M}_3 : y_{ijt} = a_0 + a_i + b_0(5 + t) + \epsilon_{ijt} \quad (6)$$

for $i = 1, 2, 3$, $j = 1, \dots, m_i$ and $t = 1, \dots, 5$ and $\epsilon_{ijt} \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$. The model fitting result is summarized in Table 2.

Table 2: Summary of inference result for model \mathcal{M}_2 and \mathcal{M}_3 .

	Dependent variable:	
	Height	
	\mathcal{M}_∞	\mathcal{M}_3
a_2	1.674 (−6.670, 10.018)	4.051*** (2.586, 5.517)
a_3	1.823 (−6.521, 10.167)	9.651*** (8.186, 11.117)
b_0	5.270*** (4.516, 6.024)	5.716*** (5.300, 6.133)
b_2	0.297 (−0.730, 1.324)	
b_3	0.979* (−0.049, 2.006)	
a_0	81.300*** (75.177, 87.423)	77.728*** (74.226, 81.230)
R^2	0.907	0.903
Adjusted R^2	0.902	0.900
$\hat{\sigma}$	2.979 (df = 94)	3.006 (df = 96)

Note: *p<0.1, **p<0.05; ***p<0.01

The result suggests that in general, girls in group 2 will be taller than girls in group 1 by 4.051cm, while girls in group 3 will be taller than girls in group 1 by 9.651cm. Such a difference in the mean of height is statistically significant at 95% confidence level. Despite the group, girls all have the same rate of growth in height with respect to their age. Every year, their mean rate of growth is 5.761cm. Based on the inference of model \mathcal{M}_1 , \mathcal{M}_2 and \mathcal{M}_3 , we would prefer the linear regression function suggested in \mathcal{M}_3 as all the parameters are statistically significant and the R^2 of the model is satisfiable.

We complete this section with a model checking procedure with respect to model \mathcal{M}_3 , which is the best among the three. Model checking consists of two parts, checking the residuals to see if they violate the model assumption and checking for outliers. Figure 5 shows the residuals corresponding to model \mathcal{M}_3 . The left panel shows the residual versus fitted value plot while the right panel shows the Q-Q plot of residuals.

From the plot, the model assumption of homoscedasticity seems not to be violated. Even though the residuals have some pattern that is not captured by the model (as seen in the red line), it is not quite significant. Meanwhile,

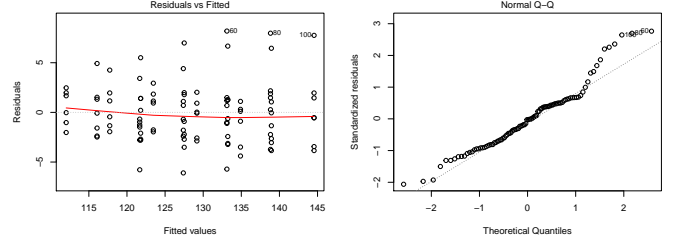


Figure 5: Residual plot of the model \mathcal{M}_3 .

the normally distributed assumption seems to be violated. The Q-Q plot suggests that the error distribution may skew right with respect to the normal distribution.

We also perform the portmanteau test statistic with respect to the sample correlations of the residuals over the first five lags. The test statistic is $\hat{Q} = 10.977$. Since the cut-off points of χ^2_5 at 0.95 is $11.071 > \hat{Q}$, we fail to reject the null hypothesis that the residuals are independent. Meanwhile, we check the normality assumption of residual using the squared correlation between the quantites in the Q-Q plot. The squared correlation is 0.958 which is is not far from 1, we fail to reject the null hypothesis and conclude that the residuals are normally distributed. In conclusion, we can say that the model assumption is not violated.

Finally, we consider high leverage points and outliers. Figure 6 shows the residuals versus leverage plot as well as the Cook's distance. Using the rule of thumb as the rule of identifying high leverage points, we find there are no high leverage points. There are three potential outliers identified by large Cook's distance. They are the 80th, 89th and 100th observation. Checking the original data, we can find that these observations are local maximizers, having much larger height than the girls similar to them.

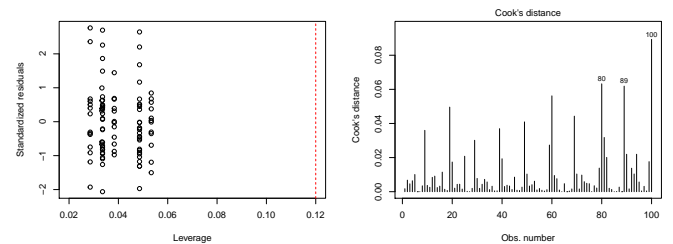


Figure 6: Checking for high leverage points and outliers. The red dashed line in the left panel is the reference line for high leverage points computed by rule of thumb.

3. Simple Linear Regression Model with Bayesian Approach

For Bayesian inference, we first consider the simplest setting where only the effect of age is considered and the error term has the same distribution assumption as

the simple linear regression model. The model \mathcal{M}_4 we considered is given in (7). The model is completed with prior on $\gamma = (\gamma_0, \gamma_1)$ and σ^2 .

$$\begin{aligned} y_{kt} &= \gamma_0 + \gamma_1(5 + t) + \epsilon_{kt}, k = 1, \dots, 20, t = 1, \dots, 5 \\ \epsilon_{kt} | \sigma^2 &\stackrel{i.i.d.}{\sim} N(0, \sigma^2) \\ \pi(\gamma, \sigma^2) &= p(\gamma | \sigma^2) \pi(\sigma^2) \end{aligned} \quad (7)$$

As specified in (8), we choose conjugate prior for the model parameters.

$$\begin{aligned} p(\gamma | \sigma^2) &\sim N(\gamma | \mu_\gamma, \sigma^2 \mathbf{V}_\gamma) \\ \pi(\sigma^2) &\sim \text{Inverse-Gamma}(\sigma^2 | a_{\sigma^2}, b_{\sigma^2}) \end{aligned} \quad (8)$$

Denote the design matrix of model \mathcal{M}_4 as X and all the response variables as \mathbf{y} . From Appendix A.1, we have the posterior distribution of γ condition on σ^2 as $N(\mu_\gamma^*, \sigma^2 \mathbf{V}_\gamma^*)$ and σ^2 as Inverse-Gamma($a_{\sigma^2}^*, b_{\sigma^2}^*$), with parameters given in (9).

$$\begin{aligned} \mu_\gamma^* &= (\mathbf{V}_\gamma^{-1} + X^T X)^{-1} (\mathbf{V}_\gamma^{-1} \mu_\gamma + X^T \mathbf{y}) \\ \mathbf{V}_\gamma^* &= ((\mathbf{V}_\gamma)^{-1} + X^T X)^{-1} \\ a_{\sigma^2}^* &= a_{\sigma^2} + \frac{n}{2} \\ b_{\sigma^2}^* &= b_{\sigma^2} + \frac{1}{2} [\mu_\gamma^T \mathbf{V}_\gamma^{-1} \mu_\gamma + \mathbf{y}^T \mathbf{y} - (\mu_\gamma^*)^T (\mathbf{V}_\gamma^*)^{-1} \mu_\gamma^*] \end{aligned} \quad (9)$$

Therefore, to sample from posterior distribution of γ and σ^2 , we can first sample σ^2 from Inverse-Gamma($a_{\sigma^2}^*, b_{\sigma^2}^*$), then sample γ from $N(\mu_\gamma^*, \mathbf{V}_\gamma^*)$.

As for the choice of hyperparameters $\mu_\gamma, \mathbf{V}_\gamma, a_{\sigma^2}$ and b_{σ^2} , we choose them according to the result of fitting the corresponding linear model with R built-in function *lm*. That is, we set $\mu_\gamma = \hat{\gamma}$ and $\mathbf{V}_\gamma = \text{Var}(\gamma)$. As for a_{σ^2} and b_{σ^2} , we set them to have the corresponding inverse Gamma distribution have mean $\hat{\sigma}^2$ and variance 10 (less informative). According to computation in R, we set the hyperparameters as (10).

$$\begin{aligned} \mu_\gamma &= \begin{pmatrix} 82.524 \\ 5.717 \end{pmatrix}, \mathbf{V}_\gamma = \begin{pmatrix} 8.088 & -0.980 \\ -0.980 & 0.123 \end{pmatrix} \\ a_{\sigma^2} &= 62.070, b_{\sigma^2} = 1496.761 \end{aligned} \quad (10)$$

Usually we also perform posterior sensitive analysis for the choice of hyperparameters. That is, we try different sets of hyperparameters, run the sampler and compare the posterior distribution. Here we also try several sets of prior hyperparameters, and the posterior inference looks similar, which suggests that our choice of hyperparameters is appropriate.

By now we have prepared all that are needed for posterior inference. Figure 7 shows the empirical posterior

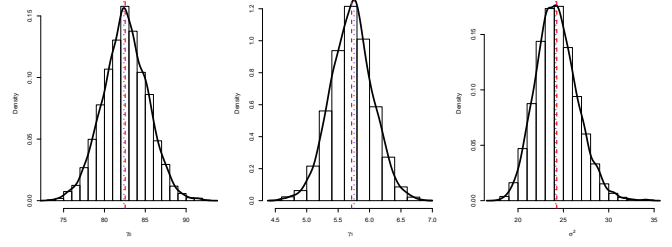


Figure 7: Empirical distribution of posterior samples of γ_0 (left panel), γ_1 (middle panel) and σ^2 (right panel). The red dashed line shows the mean of the posterior distribution while the blue dotted line is the posterior mode.

distribution of γ_0, γ_1 and σ^2 based on 5000 posterior samples.

Table 3 provides mean and 95% posterior credible interval of model parameters as well as the posterior mode. From the table we can conclude that age have positive influence on the height of girls. Girls in average will grow 5.715cm annually, with credible interval ranging from 5.049cm to 6.382cm.

Table 3: Summary of posterior inference for parameters in model \mathcal{M}_4 .

Parameter	Mean	C.I.	Mode
γ_0	82.538	(77.190, 87.898)	82.384
γ_1	5.715	(5.049, 6.382)	5.753
σ^2	24.272	(20.166, 29.062)	24.168

We complete this section with an assessment of the model's fitness to the data. Figure 8 shows the mean and interval estimation of the regression function. Most of the data points falls with in the region formed by the 95% interval estimate of regression function with only three exceptions.

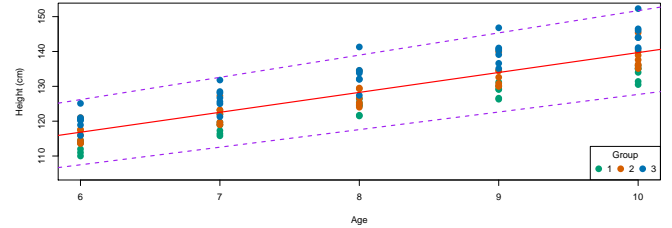


Figure 8: Mean and 95% credible interval estimate of the regression function. The red solid line corresponding to the mean regression function while the purple dashed line gives the interval estimate.

Since the posterior predictive distribution for a new data $\tilde{\mathbf{y}}$ with covariates \tilde{X} is defined in (11), we can thus draw samples from posterior predictive distribution by draw

from $N(\tilde{X}\gamma^{(s)}, (\sigma^2)^{(s)})$ where $\gamma^{(s)}$ is the s -th posterior samples of γ and same meaning applied for $(\sigma^2)^{(s)}$.

$$p(\tilde{y}|data) = \int N(\tilde{y}|\tilde{X}\gamma, \sigma^2)p(\gamma, \sigma^2|data)d\gamma d\sigma^2 \quad (11)$$

We can then obtain a 95% credible interval of the posterior predictive distribution of \tilde{y} for age taking values 6, \dots , 10. The credible interval is shown in Figure 9.

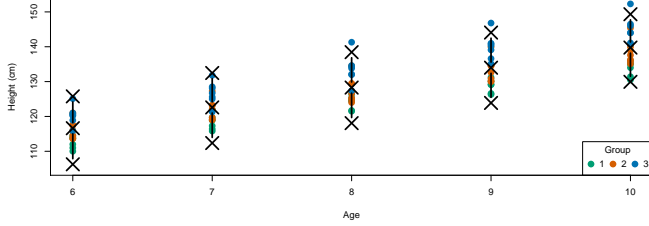


Figure 9: 95% credible interval of posterior predictive samples marked by cross symbol and solid black line.

There are still only three exceptions that falls outside the credible interval. Based on the previous analysis, we can conclude that the model fits the data relatively well.

4. Hierarchical Linear Regression Model

In general, simple nonhierarchical models may have problems in the following sense: with few parameters, they generally cannot fit the datasets accurately, whereas with many parameters, they tend to overfit the data which diminish the prediction power of the model. In contrast, hierarchical models can have enough parameters to fit the data well, while using a population distribution to structure some dependence into the parameters, thereby avoiding problems of overfitting. Therefore, we consider extending the model \mathcal{M}_4 in Section 3 by hierarchical modeling.

Based on exploratory data analysis shown in Figure 4 and model \mathcal{M}_4 , we proposed the following hierarchical model \mathcal{M}_5 stated in (12). The model is completed with prior of μ , Σ and σ^2 .

$$\begin{aligned} y_{kt} | (\gamma_{0k}, \gamma_{1k}), \sigma^2 &\stackrel{ind.}{\sim} N(\gamma_{0k} + \gamma_{1k}(5+t), \sigma^2) \\ (\gamma_{0k}, \gamma_{1k}) | \mu, \Sigma &\stackrel{i.i.d.}{\sim} N(\mu, \Sigma) \\ \pi(\mu, \Sigma, \sigma^2) &= \pi(\mu|\Sigma)\pi(\Sigma)\pi(\sigma^2) \\ \Sigma &\sim Inv - Wishart(\nu_0, \Lambda_0^{-1}), \mu|\Sigma \sim N(\mu_0, \Sigma/\kappa_0) \\ \sigma^2 &\sim Inverse - Gamma(a, b) \end{aligned} \quad (12)$$

for $k = 1, \dots, 20$ and $t = 1, \dots, 5$. We use inverse-Gamma distribution as the prior distribution for σ^2 and the normal-inverse-Wishart distribution as prior of (μ, Σ) to achieve conjugacy. The posterior samples can be

obtained according to the Gibbs sampler algorithm described in Appendix A.2.

We would like to choose the prior hyperparameters to reflect prior information. Denote $\bar{y}_t = \frac{1}{20} \sum_{k=1}^{20} y_{kt}$, we first fit the linear regression model (13) in R.

$$\bar{y}_t = \gamma_0 + \gamma_1(5+t) + \epsilon_t, \epsilon_t \stackrel{i.i.d.}{\sim} N(0, \sigma^2) \quad (13)$$

Denote the estimate of such linear model parameters as $\hat{\gamma}_*$ and $\hat{\sigma}_*^2$. We set prior hyperparameters $(\mu_0, \kappa_0, \nu_0, \Lambda_0)$ to reflect the distribution of $\hat{\gamma}_*$ while (a, b) to reflect the information in $\hat{\sigma}_*^2$. According to [7], we can set $(\mu_0, \kappa_0, \nu_0, \Lambda_0)$ as (14),

$$\begin{aligned} \mu_0 &= \hat{\gamma}_*, \kappa_0 = \nu_0 = 2 + p \\ \Lambda_0 &= (\nu_0 - p - 1)Var(\hat{\gamma}_*) \end{aligned} \quad (14)$$

where p is the dimension of γ . Then as for (a, b) , we set them to satisfies equation (15).

$$\begin{cases} E(\sigma^2) = \frac{b}{a-1} = \hat{\sigma}_*^2 \\ Var(\sigma^2) = \frac{b^2}{(a-1)^2(a-2)} = 10 \end{cases} \quad (15)$$

By now we have prepared all the expressions that are needed to implement the Gibbs sampler. We run the sampling process 6000 times with 1000 burn-in period. We first examed the posterior distribution of regression coefficients for each child. As in Figure 10, we display the 95% credible interval for posterior distributions of $(\gamma_{0k}, \gamma_{1k})$ for every k . From the plot, different child seems to have different growing pattern.

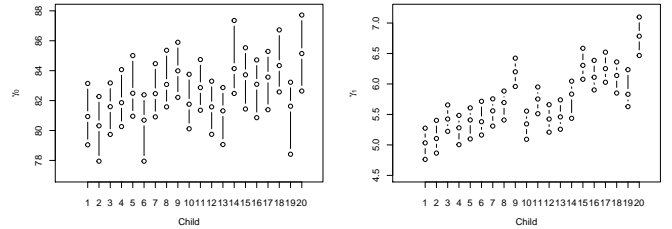


Figure 10: 95% credible interval based on posterior distributions of regression coefficients γ_0 and γ_1 for every child.

This result can be compared with Figure 4, where we plot the growth trend for each child. In that plot, girls do have different growing curves, which suggests considering fitting a regression line for each individual. That conclusion has been testified here. Based on the posterior samples of hierarchical model, we can plot the posterior mean and interval estimate of the regression line for each child in Figure 11. The model fits data pretty well, as all the data points fall around the mean regression line. Note that if we fit the same model with simple linear regression setting, we need to estimate 60 parameters, leading to

unrobust results. However, with hierarchical modeling, we achieve the same goal but with fewer parameters to be estimated.

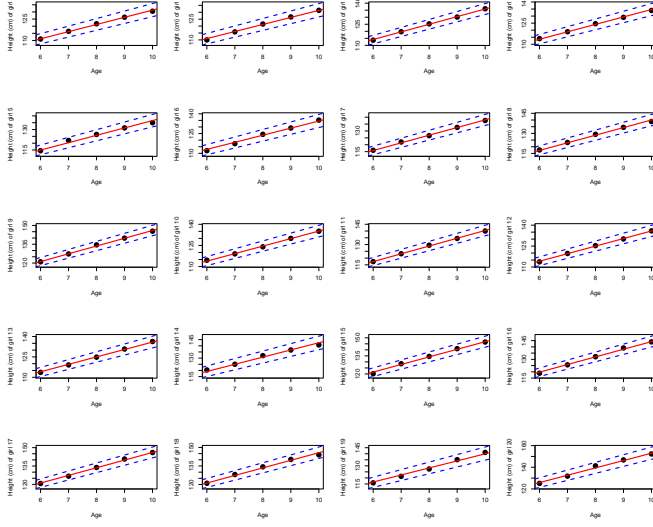


Figure 11: Mean and 95% credible interval estimate of the regression function. The red solid line corresponding to the mean regression function while the blue dashed line gives the interval estimate.

We conclude this section with Bayesian model checking using posterior predictive residuals. Based on posterior samples of model parameters, we can obtain posterior samples of replicated data y_{kt}^{rep} from $N(\gamma_{0k}^{(s)} + \gamma_{1k}^{(s)}(5 + t), (\sigma^{(s)})^2)$ where s means the s th sample from posterior distribution. Then we can compute Bayesian residual defined as $r_b = y_{kt} - y_{kt}^{rep}$. However, directly assessing the distribution of Bayesian residuals for all data point is tedious. As discussed in [1], we tend to the Bayesian p -value. Defined as in (16), Bayesian p -value can be easily obtained with posterior samples of parameters.

$$p_B = \int \int I(T(y^{rep}, \theta) \geq T(y, \theta)) p(y^{rep} | \theta) p(\theta | y) dy^{rep} d\theta \quad (16)$$

We consider four test statistics, namely the maximum, minimum, mean and standard deviation, as the criteria of comparing the posterior predictive distribution of replicate data with original data. The result is shown in Figure 12. Since all the p -values are not significant, we can conclude that this hierarchical model fits the data well, as the posterior predictive distribution of replicate data basically captures the center, range and spread of the original data. One possible drawback may be that for the extremem values in the dataset, the model dose not catches them as good as for the rest of the data. Thus, we may consider robust estimation by assuming the distribution of y_{kt} as a t distribution rather than the normal distribution.

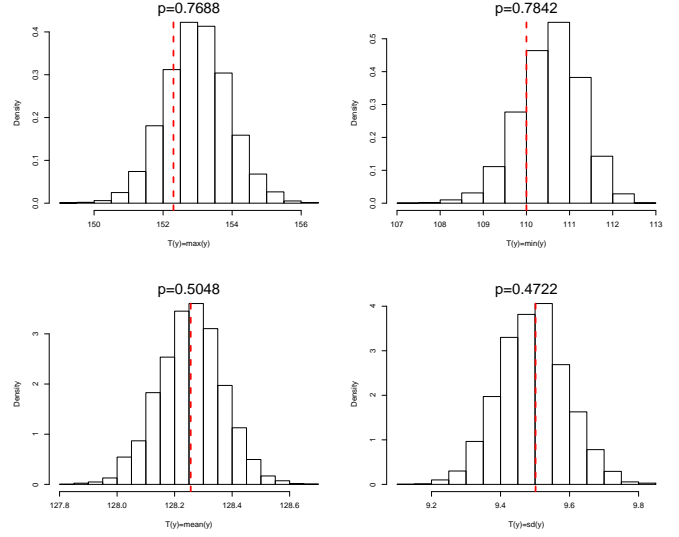


Figure 12: Posterior predictive distribution, observed result, and p -value for each of four test statistics for the model \mathcal{M}_5 .

5. Model Comparison

5.1 Model Comparison Based on Robustness of Estimation

Since our goal is to estimate the effect of factors such as age to the girls' height, we would like to have a robust model. When removing some data from the original dataset, the estimation of the effect should not differ too much. If that is the case, we are more confident in drawing conclusions. The robustness of a model can be checked by the leave-one-out analysis, which refers to holdout one data point each time and does the inference using the remaining data points. For our calibration purpose, we perform leave-one-out cross validation with respect to the inference on posterior predictive distribution. The logic is that if the model provide robust posterior predictive distribution, removing some data points does not change the overall response functional. Hence, the inference of the parameters in the model is also more trustable.

From [2], Denote $data(-i) = \{(n_{i'}, y_{i'}) : i' \neq i\}$. The cross-validated posterior predictive density is defined as (17), here θ represent all model parameters and covariates,

$$P_{CV}(z_i | data(-i)) = \int N(z_i | \theta) p(\theta | data(-i)) d\theta \quad (17)$$

and the conditional predictive ordinate (CPO) for each y_ℓ is defined as $CPO_\ell := P_{CV}(y_\ell | data(-\ell))$, $\ell = 1, \dots, 100$. To calculate CPO_ℓ , referring to [1], we have the formula in (18). It also suggests the way of calculating CPO_i based on simulation. Since we already have the posterior

distribution, the integral in (18) can be calculated using Monte Carlo method. Then CPO_ℓ can be obtained by simply inverse the integral.

$$CPO_\ell = \left(\int \frac{1}{N(y_\ell|\theta)} p(\theta|data) d\theta \right)^{-1} \quad (18)$$

Then all the CPOs are combined with the CPO log-score criteria defined in (19) and we would prefer a model with a smaller CPO log-score.

$$LS_{CPO}(\mathcal{M}) = -\frac{1}{N} \sum_{\ell=1}^N \log(CPO_\ell^{\mathcal{M}}) \quad (19)$$

The CPO log-score for two model is calculated in (20). Based on the result, we would prefer model \mathcal{M}_5 as it provides more robust estimation.

$$\begin{cases} LS_{CPO}(\mathcal{M}_4) = 3.037 \\ LS_{CPO}(\mathcal{M}_5) = 1.347 \end{cases} \quad (20)$$

5.2 Model Comparison Based on Measure of Predictive Accuracy

The assessment of the prediction power of the model should also be considered in our setting. Because we are inferencing the factors that may affect girls' height, we want the model to have good prediction power, in the sense that when obtained covariates of a new girl, we should be able to give an accurate estimation of her height. As introduced by Gelfand and Ghosh in [8], the posterior predictive loss criteria is a common used measure for predictive accuracy. Using the notation from [2], the loss D for model \mathcal{M} is defined as (21).

$$D^{\mathcal{M}} := \sum_{\ell=1}^N (y_\ell - E(z_\ell|data))^2 + \sum_{\ell=1}^N Var(z_\ell|data) \quad (21)$$

Here the first term is a measure of goodness-of-fit and the second term is a penalty term. In our setting, y_ℓ is the height of girls. For each of the two models, the corresponding loss can be calculated using posterior samples of y obtained in previous analysis. The result is summarized in Table 4. The result suggests that the hierarchical model \mathcal{M}_5 is far more better than the non-hierarchical model \mathcal{M}_4 , in the sense of a better fit to the data as well as a smaller variance.

6. Concluding Remarks

In this report, we examined the growth curve with respect to age for pre-adolescent girls. Using the linear regression model with frequentist approach, we studied whether the intercept and slope of the curve will differ by the grouping

Table 4: Comparison of model \mathcal{M}_4 and \mathcal{M}_5 based on posterior predictive loss criteria.

Model	Good-of-fit	Penalty	$D^{\mathcal{M}}$
\mathcal{M}_4	46.209	75.739	121.948
\mathcal{M}_5	2407.446	2526.864	4934.311

of their mothers. Based on hypothesis testing and model checking, we finally choose the model with only the separate effect of age and grouping factor, meaning that only the intercept of the growth curve differs by the group of mothers while the slope remains the same. The model also suggests that when considering only the effect of these two factors, girls in general increase 5.716cm annually in the age from 6 to 10. Those girls whose mother is in the "tall" group are on average 9.651cm taller than the girls whose mother is in the "short" group, and 5.600cm taller than the girls whose mother is in the "medium" group. This difference is statistically significant. Besides, we also consider the growth curve for each individual. We first fit a linear model assuming that all girls have the same growth curve, then considered a hierarchical model which lets both the intercept and slope term differ by individual. The inference for both models is done with a Bayesian approach. Based on model comparison with respect to both the robustness and prediction power, the model that assumes different intercept and slope term for each individual is much better. From posterior samples of model parameters, the growth rate with respect to girls' age can be as high as 5.403cm per year while as slow as 4.866cm per year. Although the difference may not seems to be large, accumulating it does cause a vast difference in girls' height.

We should also notice that the models analyzed in this report are just a naive try in studying this kind of problem. We can at least extend the model in the following three directions. First of all, the current model assumptions can be modified to reflect patterns showed in the data. For example, the independent and identical distribution assumption on the error term may be too strong; We may consider models under heteroskedasticity setting. Meanwhile, since the current hierarchical model does not perform well for extreme values, we may also consider models for robust estimation by changing the normal distribution assumption on the error to a t distribution. Secondly, the model can be augmented with more covariates. Based on the data we have, adding the effect of group to the hierarchical model seems reasonable. Also, if we can collect other factors for those girls, such as their living conditions and the family income, we can get a more in-depth understanding of each girl's growth curve. Last but not least, the linear shape of the growth curve is a substantial restriction. People's height seldomly has an unchanged annual growth rate. Therefore, we can also

consider release that restriction as well, which can be done with Bayesian nonparametric models. Based on DP mixture models, we can jointly model the covariates and response variable to perform density regression. Then solve the inverse problem by appropriate conditional distributions. Such a model can release both the restriction on the regression function and the error distribution, leading to a very flexible model. Regardless, this report provides a solid foundation for further research.

REFERENCES

1. Gelman, Andrew, Carlin, John B., Stern, Hal S. and Rubin, Donald B. (2014), "Bayesian Data Analysis (3rd Edition)", *Chapman and Hall/CRC*, ISBN 978-1-4398-9820-8.
2. Sansó, Bruno (2020), "Lecture Notes on Intermediate Bayesian Statistical Modeling", available at UCSC course website.
3. Goldstein, H. (1979), "The Design and Analysis of Longitudinal Studies", London: Academic Press.
4. Verbeke, G., and Molenberghs, G. (2000), "Linear Mixed Models for Longitudinal Data", New York: Springer-Verlag.
5. Albert, J. (2009), "Bayesian Computation with R (2nd Edition)", *Springer*, ISBN 978-0-387-92297-3.
6. Faraway, J. (2014). "Linear Models with R (2nd Edition)", *Chapman and Hall/CRC*, ISBN 978-1-4398-8734-9.
7. Hoff, P. (2009), "A First Course in Bayesian Statistical Methods", *Springer Texts in Statistics*, ISBN 978-0-387-92299-7.
8. Robert, Christian P. (2007), "The Bayesian Choice (2nd Edition)", *Springer Texts in Statistics*, ISBN 978-0-387-71598-8.
9. Gelfand, Alan and Ghosh, Sujit (1998), "Model choice: A minimum posterior predictive loss approach", *Biometrika*, Volume 85, Issue 1, Pages 1–11.
10. Lund, Robert (2020), "Lecture Notes on Linear Statistical Models", available at UCSC Canvas.

A. Appendix: Formula Derivation Details

A.1 Deriving Posterior Distribution for Model \mathcal{M}_4

Based on the model specification in (7) and prior given in (8), the joint posterior can be written as (22).

$$\begin{aligned}
 p(\gamma, \sigma^2 | data) &\propto p(\mathbf{y} | \gamma, \sigma^2) p(\gamma | \sigma^2) \pi(\sigma^2) \\
 &\propto \left(\frac{1}{\sigma^2}\right)^{n/2} \exp\left\{-\frac{(\mathbf{y} - X\gamma)^T (\mathbf{y} - X\gamma)}{2\sigma^2}\right\} \\
 &\times \left(\frac{1}{\sigma^2}\right)^{p/2} \exp\left\{-\frac{(\gamma - \mu_\gamma)^T \mathbf{V}_\gamma^{-1} (\gamma - \mu_\gamma)}{2\sigma^2}\right\} \\
 &\times \left(\frac{1}{\sigma^2}\right)^{a_{\sigma^2}+1} \exp(-b_{\sigma^2}/\sigma^2) \\
 &\propto \left(\frac{1}{\sigma^2}\right)^{a_{\sigma^2}^*+1} \exp\left(-\frac{b_{\sigma^2}^*}{\sigma^2}\right) \\
 &\times \left(\frac{1}{\sigma^2}\right)^{p/2} \exp\left\{-\frac{(\gamma - \mu_\gamma^*)^T (\mathbf{V}_\gamma^*)^{-1} (\gamma - \mu_\gamma^*)}{2\sigma^2}\right\} \\
 &\propto p(\sigma^2 | data) p(\gamma | \sigma^2, data)
 \end{aligned} \tag{22}$$

where

$$\begin{aligned}
 \mu_\gamma^* &= (\mathbf{V}_\gamma^{-1} + X^T X)^{-1} (\mathbf{V}_\gamma^{-1} \mu_\gamma + X^T \mathbf{y}) \\
 \mathbf{V}_\gamma^* &= ((\mathbf{V}_\gamma)^{-1} + X^T X)^{-1} \\
 a_{\sigma^2}^* &= a_{\sigma^2} + \frac{n}{2} \\
 b_{\sigma^2}^* &= b_{\sigma^2} + \frac{1}{2} [\mu_\gamma^T \mathbf{V}_\gamma^{-1} \mu_\gamma + \mathbf{y}^T \mathbf{y} - (\mu_\gamma^*)^T (\mathbf{V}_\gamma^*)^{-1} \mu_\gamma^*]
 \end{aligned} \tag{23}$$

In (22), we have implemented the multivariate completion of squares to obtain (24).

$$\begin{aligned}
 b_{\sigma^2} &+ \frac{1}{2} \{(\gamma - \mu_\gamma)^T \mathbf{V}_\gamma^{-1} (\gamma - \mu_\gamma) + (\mathbf{y} - X\gamma)^T (\mathbf{y} - X\gamma)\} \\
 &= b_{\sigma^2}^* + \frac{1}{2} (\gamma - \mu_\gamma^*)^T \mathbf{V}_\gamma^* (\gamma - \mu_\gamma^*)
 \end{aligned} \tag{24}$$

Thus, we have the posterior distribution of γ , condition on σ^2 as $N(\mu_\gamma^*, \sigma^2 \mathbf{V}_\gamma^*)$ and σ^2 as Inverse-Gamma($a_{\sigma^2}^*, b_{\sigma^2}^*$).

A.2 Gibbs Sampler for Posterior Inference of Model \mathcal{M}_5

Assume the model in (12), and introducing the following notation in (25), where \mathbf{y}_k is the height from age 6 to 10 of the k th girl, X_k is the corresponding design matrix, and γ_k is the vector of regression coefficients, for $k = 1, \dots, 20$.

$$\begin{aligned}
 \mathbf{y}_k &= (y_{k1} \quad \dots \quad y_{k5})^T \\
 X_k &= \begin{pmatrix} 1 & \dots & 1 \\ 6 & \dots & 10 \end{pmatrix}^T \\
 \gamma_k &= (\gamma_{0k}, \gamma_{1k})^T
 \end{aligned} \tag{25}$$

We can now write the posterior distribution as in (26).

$$\begin{aligned}
 f(\gamma_1, \dots, \gamma_{20}, \mu, \Sigma, \sigma^2 | data) &\propto f(y | \gamma_1, \dots, \gamma_{20}, \sigma^2) \\
 &\times f(\gamma_1, \dots, \gamma_{20} | \mu, \Sigma) \pi(\mu | \Sigma) \pi(\Sigma) \pi(\sigma^2) \\
 &\propto \prod_{k=1}^{20} \left(\frac{1}{\sigma^2}\right)^{5/2} \exp\left(-\frac{(\mathbf{y}_k - X_k \gamma_k)^T (\mathbf{y}_k - X_k \gamma_k)}{2\sigma^2}\right) \\
 &\times \prod_{k=1}^{20} |\Sigma|^{-1/2} \exp\left(-\frac{(\gamma_k - \mu)^T \Sigma^{-1} (\gamma_k - \mu)}{2}\right) \\
 &\times |\Sigma|^{-(\nu_0+2)/2+1} \exp\left(-\frac{1}{2} \text{tr}(\Lambda_0 \Sigma^{-1})\right) \\
 &- \frac{\kappa_0}{2} (\mu - \mu_0)^T \Sigma^{-1} (\mu - \mu_0) \times \left(\frac{1}{\sigma^2}\right)^{a+1} \exp\left(-\frac{b}{\sigma^2}\right)
 \end{aligned} \tag{26}$$

Now, we can compute the full conditional distributions for all model parameters.

Denote the $\cdot|\mathcal{P}$ as the full conditionals to simplify the notation. Firstly, consider the full conditional distributions for γ_k , $k = 1, \dots, 20$. We have,

$$\begin{aligned} f(\gamma_k|\mathcal{P}) &\propto \exp\left(-\frac{(\mathbf{y}_k - X_k\gamma_k)^T(\mathbf{y}_k - X_k\gamma_k)}{2\sigma^2}\right) \\ &\times \exp\left(-\frac{(\gamma_k - \boldsymbol{\mu})^T\sigma^2\Sigma^{-1}(\gamma_k - \boldsymbol{\mu})}{2\sigma^2}\right) \\ &\propto \exp\left(-\frac{(\gamma_k - \boldsymbol{\mu}_k^*)^T(\Sigma_k^*)^{-1}(\gamma_k - \boldsymbol{\mu}_k^*)}{2}\right) \end{aligned} \quad (27)$$

Thus, the full conditionals of γ_k is the bivariate normal distribution $N(\boldsymbol{\mu}_k^*, \Sigma_k^*)$ with parameters given in (28).

$$\begin{aligned} \boldsymbol{\mu}_k^* &= \Sigma_k^* = (\Sigma^{-1} + X_k^T X_k / \sigma^2)^{-1} (\Sigma^{-1} \boldsymbol{\mu} + X_k^T \mathbf{y}_k / \sigma^2) \\ \Sigma_k^* &= (\Sigma^{-1} + X_k^T X_k / \sigma^2)^{-1} \end{aligned} \quad (28)$$

As for σ^2 , we have the full conditionals of it as in (29).

$$f(\sigma^2|\mathcal{P}) \propto \left(\frac{1}{\sigma^2}\right)^{a^*+1} \exp\left(-\frac{b^*}{\sigma^2}\right) \quad (29)$$

Hence, the full conditionals of σ^2 is the inverse Gamma distribution inverse-Gamma(a^* , b^*) with the parameters given in (30).

$$\begin{aligned} a^* &= a + \frac{n}{2} \\ b^* &= b + \frac{1}{2} \sum_{k=1}^{20} (\mathbf{y}_k - X_k\gamma_k)^T (\mathbf{y}_k - X_k\gamma_k) \end{aligned} \quad (30)$$

Then for full conditionals of $\boldsymbol{\mu}$ and Σ , we have conjugacy. From the result in [1], we have the full conditionals of $\boldsymbol{\mu}$ as a normal distribution $N(\boldsymbol{\mu}_*, \Sigma/\kappa_*)$ and the full conditionals of Σ as a inverse Wishart distribution Inv-Wishart(ν_* , Λ_*^{-1}) where all the parameters are given in (31).

$$\begin{aligned} \boldsymbol{\mu}_* &= \frac{\kappa_0}{\kappa_0 + 20} \boldsymbol{\mu}_0 + \frac{20}{\kappa_0 + 20} \bar{\gamma} \\ \kappa_* &= 20 + \kappa_0, \quad \nu_* = 20 + \nu_0 \\ \Lambda_* &= \Lambda_0 + S + \frac{20\kappa_0}{\kappa_0 + 20} (\bar{\gamma} - \boldsymbol{\mu}_0)(\bar{\gamma} - \boldsymbol{\mu}_0)^T \\ S &= \sum_{k=1}^{20} (\gamma_k - \bar{\gamma})(\gamma_k - \bar{\gamma})^T \end{aligned} \quad (31)$$

Here $\bar{\gamma}$ denote the mean of $\{\gamma_k\}_{k=1}^{20}$.

Now, with all the required full conditional distributions at hand, we can apply the Gibbs sampler to obtain posterior samples of all model parameters.

B. Appendix: Code in R

Please note that to show the full code in this two-column setting, I have to use many unnecessary line breaks. It

may lead to problems when running them. The problem should be fixed by combining lines when needed. I am also willing to provide a version of code in other formats that can run fluently by directly copy and paste to R if needed.

B.1 Code for Exploratory Data Analysis in Introduction Section

```
## load data
dat=read.table("~/Desktop/UCSC/fye2020/
               data_schoolgirls.txt",
               header=TRUE,sep="",
               na.strings=".")
dat$child=as.factor(dat$child)
dat$group=as.factor(dat$group)
dat=dat[,2:5]

## histogram
hist(dat$height,freq=FALSE,xlab='Height (cm)',
     main='',cex.axis=1.5,cex.lab=1.5)
abline(v=mean(dat$height),col='red',
       lwd=3,lty=1)
abline(v=median(dat$height),col='blue',
       lwd=3,lty=2)

## boxplot
library(ggplot2)
library(tidy)
library(gridExtra)

plotdata=dat
plotdata$age=as.factor(plotdata$age)

plot1=plotdata %>%
  ggplot(mapping = aes(x=age, y=height)) +
  geom_boxplot() +
  labs(y="Height (cm)", x = "Age")

plot2=plotdata %>%
  ggplot(mapping = aes(x=group, y=height)) +
  geom_boxplot() +
  labs(y="Height (cm)",
       x = "Grouping of mother's height")

grid.arrange(plot1,plot2,ncol=2)

## scatter plot with regression line
colors.group<-c("#009E73", "#D55E00", "#0072B2")
colors<-colors.group[as.numeric(dat$group)]

## fit regression model
L1=lm(height ~ group*age,data=dat)
mu1 = L1$coefficients[1]
gp1.mean1 = mu1 + 0
```

```

gp2.mean1 = mu1 + L1$coefficients[2]
gp3.mean1 = mu1 + L1$coefficients[3]
slope1 = L1$coefficients[4]
gp1.slope1 = slope1 + 0
gp2.slope1 = slope1 + L1$coefficients[5]
gp3.slope1 = slope1 + L1$coefficients[6]

plot(x=dat$age, y=dat$height,col=colors,
     pch = 16,cex=1.6,xlab = "Age",
     ylab = "Height (cm)", main='')
legend('bottomright',legend=levels(dat$group),
      horiz = TRUE, col=colors.group,
      cex = 1.6,pch = 16,title='Group')
abline(gp1.mean1, gp1.slope1,lty=1,
      lwd=2, col=colors.group[1])
abline(gp2.mean1, gp2.slope1,lty=1,
      lwd=2, col=colors.group[2])
abline(gp3.mean1, gp3.slope1,lty=1,
      lwd=2, col=colors.group[3])

ggplot(dat, aes(age,height,
                group=child,col=group)) +
  geom_line() +
  scale_fill_manual(values=colors.group,
                    name="Height\nGroup")+
  labs(y="Height (cm)", x = "Age") +
  theme(legend.justification=c(1,0),
        legend.position=c(1,0),
        legend.background = element_rect(
          fill="#f2f2f2",
          size=0.5, linetype="solid"))

```

B.2 Code for Frequentist Linear Regression Inference in Section 2

```

## fit linear models
lm1=lm(height~group,data=dat)
L1=lm(height ~ group*age,data=dat)
lm2=lm(height~group+age,data=dat)

## plot residual
par(mfrow=c(1,2))
plot(lm2,which=1:2)

## test residual
n=nrow(dat)
sample.res.corr=function(h){
  numerator=sum(lm2$residuals[1:(n-h)]
               *lm2$residuals[(h+1):n])
  denominator=sum(lm2$residuals^2)
  return(numerator/denominator)
}

```

```

Q=sapply(seq(1,5,by=1), sample.res.corr)
Q=n*sum(Q^2)

res.vs.norm=qnorm(lm2$residuals,
                  plot.it = FALSE)
(cor(res.vs.norm$x,res.vs.norm$y))^2

## plot high leverage points, outlier
p=length(lm2$coefficients)
leverageaes.model=hatvalues(lm2)
std.res=rstandard(lm2)

par(mfrow=c(1,2))
plot(leverageaes.model,std.res,
     ylab='Standardized residuals',
     xlab='Leverage',main="",
     xlim=c(0.02,0.12))
abline(v=3*p/n,col='red',lty=2)

plot(lm2,which=4)

```

B.3 Code for Bayesian Inference in Section 3

```

## set hyperparameters
lm3=lm(height~age,data=dat)
mu.gamma=lm3$coefficient
V.gamma=unname(vcov(lm3))
a.sigma=2+(sigma(lm3))^4/10
b.sigma=(sigma(lm3))^2*(a.sigma-1)

## sample and plot
library(mvtnorm)

## calculate parameter
X=model.matrix(lm3)
y=matrix(dat$height,ncol=1)
n=nrow(dat)
V.gamma.inverse=chol2inv(chol(V.gamma))
V.gamma.star=chol2inv(
  chol(V.gamma.inverse+t(X)%*%X))
V.gamma.star.inv=chol2inv(
  chol(V.gamma.star))
mu.gamma.star=V.gamma.star%*%
  (V.gamma.inverse%*%mu.gamma+t(X)%*%y)
a.sigma.star=a.sigma+n/2
b.sigma.star=b.sigma+1/2*(t(mu.gamma)%*%
  V.gamma.inverse%*%mu.gamma+t(y)%*%y
  -t(mu.gamma.star)%*%V.gamma.star.inv
  %*%mu.gamma.star)

## save samples
gamma.sample=matrix(0,nrow=5000,ncol=2)
sigma.sample=rep(0,5000)

```

```

## sample
for (i in 1:5000) {
  set.seed(i)
  sigma.sample[i]=1/rgamma(1,
  shape=a.sigma.star,rate=b.sigma.star)
  gamma.sample[i,]=rmvnorm(1,
  mean=mu.gamma.star,sigma=sigma.sample[i]
  *V.gamma.star)
}

par(mfrow=c(1,3))
hist(gamma.sample[,1],freq=FALSE,
  xlab=expression(gamma[0]),
  ylab='Density',main='')
lines(density(gamma.sample[,1]),type='l',lwd=3)
abline(v=mean(gamma.sample[,1]),col='red',
  lty=2,lwd=2)
abline(v=density(gamma.sample[,1])$x
  [which.max(density(gamma.sample[,1])$y)],
  col='blue',lty=3,lwd=2)

hist(gamma.sample[,2],freq=FALSE,
  xlab=expression(gamma[1]),
  ylab='Density',main='')
lines(density(gamma.sample[,2]),type='l',lwd=3)
abline(v=mean(gamma.sample[,2]),col='red',
  lty=2,lwd=2)
abline(v=density(gamma.sample[,2])$x
  [which.max(density(gamma.sample[,2])$y)],
  col='blue',lty=3,lwd=2)

hist(sigma.sample,freq=FALSE,
  xlab=expression(sigma^2),
  ylab='Density',main='')
lines(density(sigma.sample),type='l',lwd=3)
abline(v=mean(sigma.sample),col='red',
  lty=2,lwd=2)
abline(v=density(sigma.sample)$x
  [which.max(density(sigma.sample)$y)],
  col='blue',lty=3,lwd=2)

## summary result
summary.vec95=function(vec){
  c(unnname(quantile(vec,0.025)),
  mean(vec),unnname(quantile(vec,0.975)))
}

gamma1.sum=summary.vec95(gamma.sample[,1])
gamma2.sum=summary.vec95(gamma.sample[,2])
sigma.sum=summary.vec95(sigma.sample)
gamma1.mode=density(gamma.sample[,1])$x
[which.max(density(gamma.sample[,1])$y)]
gamma2.mode=density(gamma.sample[,2])$x
[which.max(density(gamma.sample[,2])$y)]
sigma.mode=density(sigma.sample)$x
[which.max(density(sigma.sample)$y)]

```

```

[which.max(density(sigma.sample)$y)]

## sensitive analysis
## try another set of hyperparameters

mu.gamma=c(0,0)
V.gamma=diag(c(100,100))
a.sigma=1
b.sigma=1

## sample and plot
library(mvtnorm)

## calculate parameter
X=model.matrix(lm3)
y=matrix(dat$height,ncol=1)
n=nrow(dat)
V.gamma.inverse=chol2inv(chol(V.gamma))
V.gamma.star=chol2inv(
  chol(V.gamma.inverse+t(X)%*%X))
V.gamma.star.inv=chol2inv(
  hol(V.gamma.star))
mu.gamma.star=V.gamma.star%*%
  (V.gamma.inverse%*%mu.gamma+t(X)%*%y)
a.sigma.star=a.sigma+n/2
b.sigma.star=b.sigma+1/2*(t(mu.gamma)%*%
  V.gamma.inverse%*%mu.gamma+t(y)%*%y
  -t(mu.gamma.star)%*%V.gamma.star.inv
  %*%mu.gamma.star)

## save samples
gamma.sample=matrix(0,nrow=5000,ncol=2)
sigma.sample=rep(0,5000)

## sample
for (i in 1:5000) {
  set.seed(i)
  sigma.sample[i]=1/rgamma(1,
  shape=a.sigma.star,rate=b.sigma.star)
  gamma.sample[i,]=rmvnorm(1,
  mean=mu.gamma.star,sigma=sigma.sample[i]
  *V.gamma.star)
}

## summary
gamma1.sum=summary.vec95(gamma.sample[,1])
gamma2.sum=summary.vec95(gamma.sample[,2])
sigma.sum=summary.vec95(sigma.sample)
gamma1.mode=density(gamma.sample[,1])$x
[which.max(density(gamma.sample[,1])$y)]
gamma2.mode=density(gamma.sample[,2])$x
[which.max(density(gamma.sample[,2])$y)]
sigma.mode=density(sigma.sample)$x
[which.max(density(sigma.sample)$y)]

```

```

## estimate regression function
plot(x=dat$age, y=dat$height,col=colors,
     pch = 16,cex=1.6,
     xlab = "Age",ylab = "Height (cm)",
     main='',ylim=c(105,152))
legend('bottomright',
      legend = levels(dat$group),horiz = TRUE,
      col = colors.group,pt.cex = 1.6,
      pch = 16,title='Group')
abline(gamma1.sum[2], gamma2.sum[2],
      lty=1, lwd=2, col = 'red')
abline(gamma1.sum[3], gamma2.sum[3],
      lty=2, lwd=2, col = 'purple')
abline(gamma1.sum[1], gamma2.sum[1],
      lty=2, lwd=2, col = 'purple')

## interval of posterior predictive
x.grid=seq(6,10,by=1)
pp.sample=matrix(0,nrow=5000,ncol=5)

for (i in 1:5000) {
  set.seed(i)
  for (j in 1:5) {
    mean.cur=gamma.sample[i,1]+
      gamma.sample[i,2]*x.grid[j]
    sd.cur=sqrt(sigma.sample[i])
    pp.sample[i,j]=rnorm(1,mean=mean.cur,
                        sd=sd.cur)
  }
}

mtx.ci=apply(pp.sample, MARGIN = 2,
             summary.vec95)

plot(x=dat$age, y=dat$height,col=colors,
     pch = 16,cex=1.6,
     xlab = "Age",ylab = "Height (cm)",
     main='',ylim=c(105,152))
legend('bottomright',
      legend = levels(dat$group),horiz = TRUE,
      col = colors.group,pt.cex = 1.6,
      pch = 16,title='Group')
for (i in 1:5){lines(c(i+5,i+5,i+5),mtx.ci[,i],
                    type="b",lwd=3,
                    pch=4,cex=3)}

```

B.4 Code for Bayesian Hierarchical Model in Section 4

```

library(MCMCpack)

## prepare hyperparameters
y.mean=aggregate(plotdata$height,

```

```

      list(plotdata$age), mean)[,2]
hyper.df=data.frame(y=y.mean,
                    x=unique(dat$age))
lm.hyper=lm(y~x,data=hyper.df)
mu0=lm.hyper$coefficient
kappa0=nu0=2+2
Lambda0=unname(vcov(lm.hyper))
a=2+(sigma(lm.hyper))^4/10
b=(sigma(lm.hyper))^2*(a-1)

X.k=model.matrix(lm.hyper)

## sample functions
sample.gamma.k=function(yk,
                       sigma2,mu,Sigma){
  Sigma.inv=chol2inv(chol(Sigma))
  Sigma.inv.k=chol2inv(chol(Sigma.inv
    +t(X.k)%*%X.k/sigma2))
  mean.k=Sigma.inv.k%*%
    (Sigma.inv%*%mu+t(X.k)%*%yk/sigma2)
  gamma.k=rmvnorm(1,
                  mean=mean.k,sigma=Sigma.inv.k)
  return(gamma.k)
}

sample.sigma=function(gamma.vec){
  a.star=a+n/2
  cal_error=function(k){
    y.k=matrix(dat[dat$child==k,1],ncol=1)
    coef.k=matrix(gamma.vec
                  [(2*k-1):(2*k)],ncol=1)
    error=t(y.k-X.k%*%coef.k)
    %*%(y.k-X.k%*%coef.k)
    return(error)
  }
  b.star=b+0.5*sum(sapply(
    seq(1,20,by=1), cal_error))
  sigma=1/rgamma(1,shape=a.star,
                 rate=b.star)
  return(sigma)
}

sample.mu=function(gamma.vec,Sigma){
  gamma.mtx=matrix(gamma.vec,
                  nrow=2,byrow=FALSE)
  gamma.mean=rowMeans(gamma.mtx)
  mu.star=kappa0/(kappa0+20)*mu0+20/
    (kappa0+20)*gamma.mean
  kappa.star=20+kappa0
  mu.sample=rmvnorm(1,mean=mu.star,
                    sigma=Sigma/kappa.star)
  return(mu.sample)
}

sample.cov.mtx=function(gamma.vec){

```

```

gamma.mtx=matrix(gamma.vec,
                 nrow=2,byrow=FALSE)
gamma.mean=rowMeans(gamma.mtx)
gamma.error=apply(gamma.mtx-gamma.mean,
                 MARGIN = 2,function(x) x%*%t(x))
S=apply(gamma.error, MARGIN = 1 , sum)
S=matrix(S,nrow=2)
nu.star=20+nu0
Lambda.star=Lambda0+S+20*kappa0*
(gamma.mean-mu0)%*%t(gamma.mean-mu0)/
(kappa0+20)
cov.mtx.sample=riwish(v=nu.star,
                     S=Lambda.star)
return(cov.mtx.sample)
}

```

record samples

```

gamma.hier=matrix(0,ncol=40,nrow=6000)
sigma.hier=rep(0,6000)
mu.hier=matrix(0,ncol=2,nrow=6000)
cov.mtx.hier=matrix(0,ncol=4,nrow=6000)

```

```

gamma.hier[1,]=rep(unnamed(mu0),times=20)
sigma.hier[1]=(sigma(lm.hyper))^2
mu.hier[1,]=mu0
cov.mtx.hier[1,]=as.vector(Lambda0)

```

gibbs sampler

```

for (i in 1:5999) {
  set.seed(i)

  cov.mtx.cur=matrix(cov.mtx.hier[i,],nrow=2)

  ## sample gamma->sigma->mu->Sigma
  for (j in 1:20) {
    y.cur=dat[dat$child==j,1]
    gamma.hier[(i+1),(2*j-1):(2*j)]=
      sample.gamma.k(y.cur,sigma.hier[i],
                    mu.hier[i,],cov.mtx.cur)
  }
  sigma.hier[i+1]=sample.sigma(gamma.hier[(i+1),])
  mu.hier[(i+1),]=sample.mu(gamma.hier[(i+1),],
                           cov.mtx.cur)
  cov.mtx.new=sample.cov.mtx(gamma.hier[(i+1),])
  cov.mtx.hier[(i+1),]=as.vector(cov.mtx.new)
}

```

make summary plot

```

gamma.hier.sample=gamma.hier[1001:6000,]
sigma.hier.sample=sigma.hier[1001:6000]
gamma.hier.sample0=gamma.hier.sample
[,seq(1,by=2,len=20)]
gamma.hier.sample1=gamma.hier.sample

```

```

[,seq(2,40,2)]

```

```

mtx.ci.gamma0=apply(gamma.hier.sample0,
                   MARGIN = 2,summary.vec95)
mtx.ci.gamma1=apply(gamma.hier.sample1,
                   MARGIN = 2,summary.vec95)

```

```

par(mfrow=c(1,2))
plot(c(0,0,0),type="n",lwd=3,xlim=c(-1,21),
     ylim=c(77,88), xlab="Child",
     ylab=expression(gamma[0]),xaxt='n')
for (i in 1:20){lines(c(i,i,i),
                    mtx.ci.gamma0[,i],type="b")}}
xtick<-seq(1,20,by=1)
axis(side=1, at=seq(1,20,by=1), labels = FALSE)
text(x=seq(1,20,by=1),
     par("usr")[3], labels = xtick,
     pos = 1, xpd = TRUE,offset = 1.6)

```

```

plot(c(0,0,0),type="n",lwd=3,xlim=c(-1,21),
     ylim=c(4.5,7.2), xlab="Child",
     ylab=expression(gamma[1]),xaxt='n')
for (i in 1:20){lines(c(i,i,i),
                    mtx.ci.gamma1[,i],type="b")}}
xtick<-seq(1,20,by=1)
axis(side=1, at=seq(1,20,by=1), labels = FALSE)
text(x=seq(1,20,by=1),
     par("usr")[3], labels = xtick,
     pos = 1, xpd = TRUE,offset = 1.6)

```

check fit

```

par(mfrow=c(5,4))
for(k in 1:20){
  y.cur=dat[dat$child==k,1]
  gamma1.sum.cur=summary.vec95(
    gamma.hier.sample[, (2*k-1)])
  gamma2.sum.cur=summary.vec95(
    gamma.hier.sample[, (2*k)])
  y.max=max(gamma1.sum.cur[3]
            +gamma2.sum.cur[3]*10,y.cur)
  y.min=min(gamma1.sum.cur[1]
            +gamma2.sum.cur[1]*6,y.cur)
  plot(x=unique(dat$age), y=y.cur,
       type='p',cex=2,pch=16,
       ylim=c(y.min,y.max),xlab = "Age",
       ylab = paste("Height (cm) of girl",k),
       main='')
  abline(gamma1.sum.cur[2],
        gamma2.sum.cur[2],lty=1,
        lwd=2, col = 'red')
  abline(gamma1.sum.cur[3],
        gamma2.sum.cur[3],lty=2,
        lwd=2, col = 'blue')
  abline(gamma1.sum.cur[1],
        gamma2.sum.cur[1],lty=2,
        lwd=2, col = 'blue')
}

```



```

}

## Bayesian p-value
test.pp=function(vec){
  c(max(vec),min(vec),mean(vec),sd(vec))
}
y.true=test.pp(dat$height)

pp.dist.hier=matrix(0,nrow=5000,ncol=100)
pp.p.value=matrix(0,nrow=5000,ncol=4)
k.idx=as.numeric(dat$child)

for (i in 1:5000) {
  set.seed(i)
  for (j in 1:100) {
    k.cur=k.idx[j]
    mean.cur=gamma.hier.sample
      [i,(2*k.cur-1)]
    +gamma.hier.sample[i,(2*k.cur)]*dat$age[j]
    sd.cur=sqrt(sigma.hier.sample[i])
    pp.dist.hier[i,j]=
      rnorm(1,mean=mean.cur,sd=sd.cur)
  }
  test.pp.cur=test.pp(pp.dist.hier[i,])
  pp.p.value[i,]=test.pp.cur>=y.true
}

pp.hier.test=apply(pp.dist.hier, MARGIN = 1,
  test.pp)
pp.p.value=colMeans(pp.p.value)

test.stat=c('T(y)=max(y) ',
  'T(y)=min(y) ',
  'T(y)=mean(y) ', 'T(y)=sd(y) ')

par(mfrow=c(2,2))
for (i in 1:4) {
  hist(pp.hier.test[i,],freq=FALSE,
    xlab=test.stat[i],main="")
  abline(v=y.true[i],col='red',
    lty=2,lwd=3)
  mtext(paste0('p=',pp.p.value[i]),
    cex=1.6)
}

```

B.5 Code for Model Comparison

```

## calculate CPO

CPO.lm=rep(0,100)
CPO.hier=rep(0,100)
for (i in 1:100) {
  den.lm.cur=rep(0,5000)

```

```

  den.hier.cur=rep(0,5000)
  for (j in 1:5000) {
    mean.lm=gamma.sample[j,1]
    +gamma.sample[j,2]*dat$age[i]
    sd.lm=sqrt(sigma.sample[j])
    k.cur=k.idx[i]
    mean.hier=gamma.hier.sample
      [j,(2*k.cur-1)]
    +gamma.hier.sample[j,(2*k.cur)]*dat$age[i]
    sd.hier=sqrt(sigma.hier.sample[j])
    den.lm.cur[j]=dnorm(dat$height[i],
      mean=mean.lm,sd=sd.lm)
    den.hier.cur[j]=dnorm(dat$height[i],
      mean=mean.hier,sd=sd.hier)
  }
  integral.lm.cur=mean(1/den.lm.cur)
  integral.hier.cur=mean(1/den.hier.cur)
  CPO.lm[i]=1/integral.lm.cur
  CPO.hier[i]=1/integral.hier.cur
}

LS.cpo1=-mean(log(CPO.lm))
LS.cpo2=-mean(log(CPO.hier))

## posterior predictive loss
pp.loss=function(mtx){
  exp.dat=colMeans(mtx)
  exp.dat.sq=colMeans(mtx*mtx)
  g=sum((dat$height-exp.dat)^2)
  p=sum(exp.dat.sq-exp.dat^2)
  d=g+p
  return(c(g,p,d))
}

round(pp.loss(pp.dist.hier),3)

pp.dist.lm=matrix(0,nrow=5000,ncol=100)

for (i in 1:5000) {
  set.seed(i)
  for (j in 1:100) {
    mean.cur=gamma.sample[i,1]
    +gamma.sample[i,2]*dat$age[j]
    sd.cur=sqrt(sigma.sample[i])
    pp.dist.lm[i,j]=rnorm(1,
      mean=mean.cur,sd=sd.cur)
  }
}

round(pp.loss(pp.dist.lm),3)

```