Predicting Children's Weight with Frequentist & Bayesian Mixed Modelling

Author: P. Z. Vivian Hsu • Date: April 17, 2025

Introduction

Data Exploration

We first clean up the data by excluding height, converting sex to a factor, and removing missing values from the data set. The rest of this section will use the clean data and so this will be a complete case analysis.

```
      id sex   wt mage lit died alive age
1 120011   1 12.8   35   0    2     5  41
2 120011   1 12.8   35   0    2     5  45
3 120011   1 13.1   35   0    2     5  49
4 120011   1 13.8   35   0    2     5  53
6 120012   2 14.9   35   0    2     5  57
```

Below we perform data visualization. From the scatter plots, there appears to be:

- An upward linear trend between weight and the child's age.
- A somewhat upward linear trend between weight and the mother's age.
- No strong trend between weight and the number of children the mother has had that died.
- No apparent trend between weight and the number of children the mother had born alive.

These trends do not appear to differ by sex. As for the box plot:

- For male children, the mean weight appears to be larger for those with illiterate mothers vs literate mothers.

1

- For female children, the mean weight appears to be larger for those with literate mothers vs illiterate mothers.
- However, there is a quite a bit of variability (and overlapping variability as well), so it's hard to tell whether the literacy status has an effect on the child's weight or not.

3
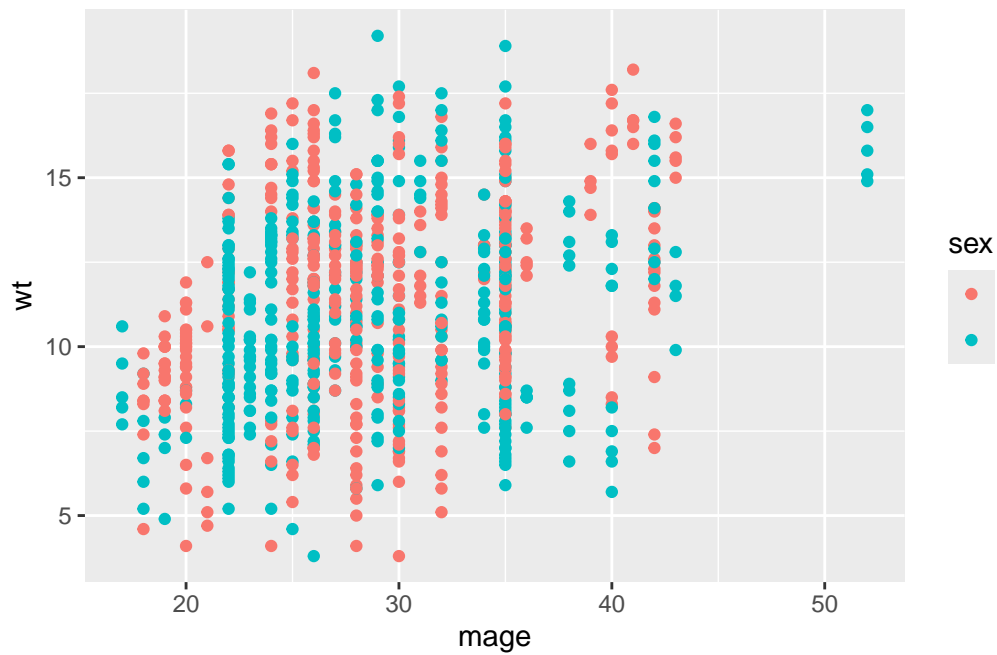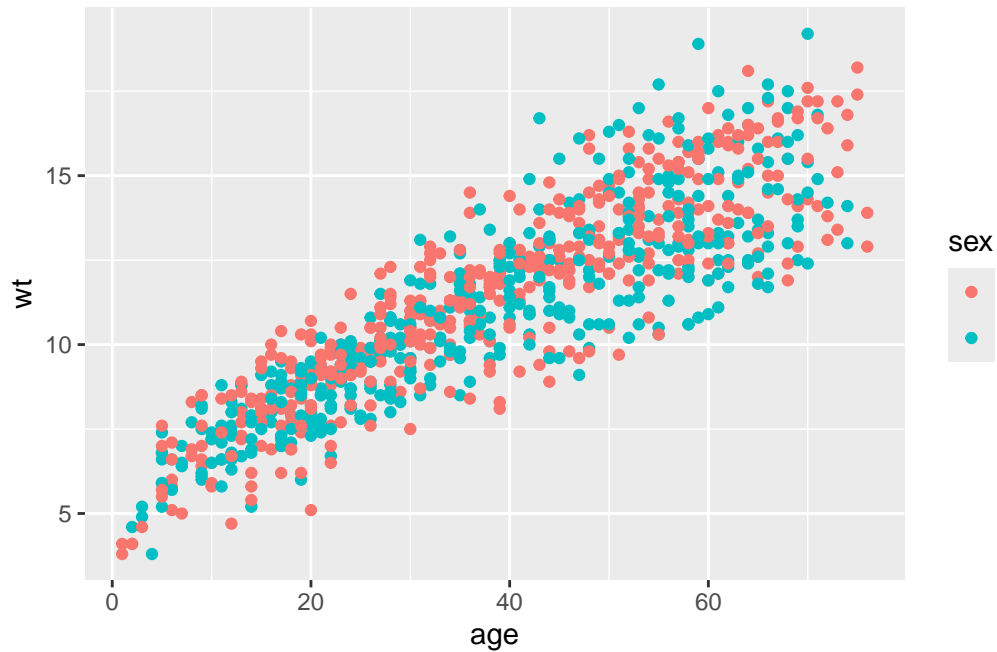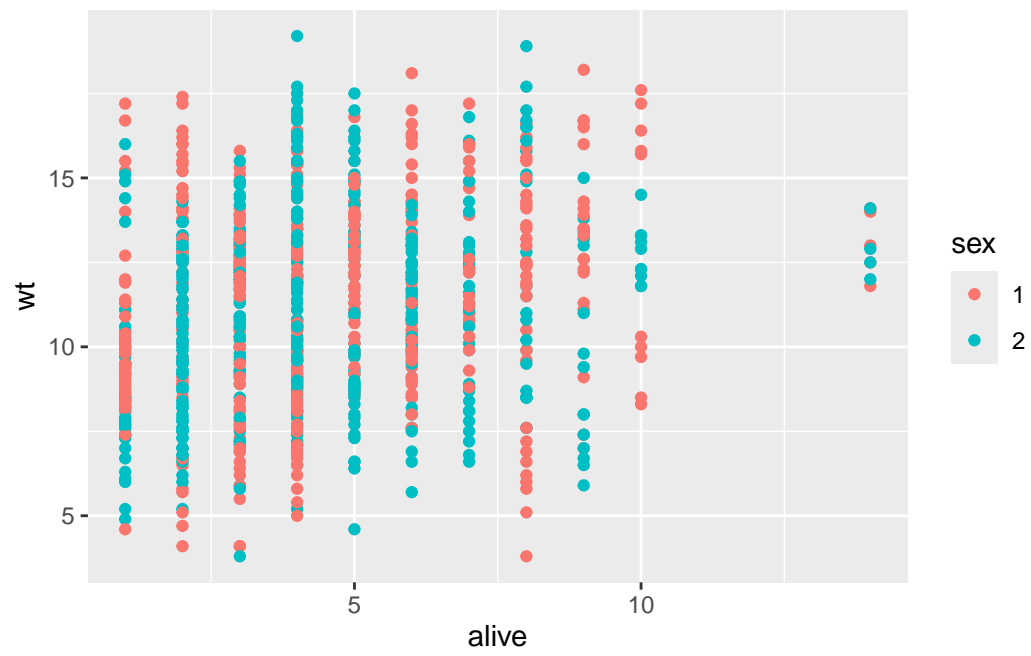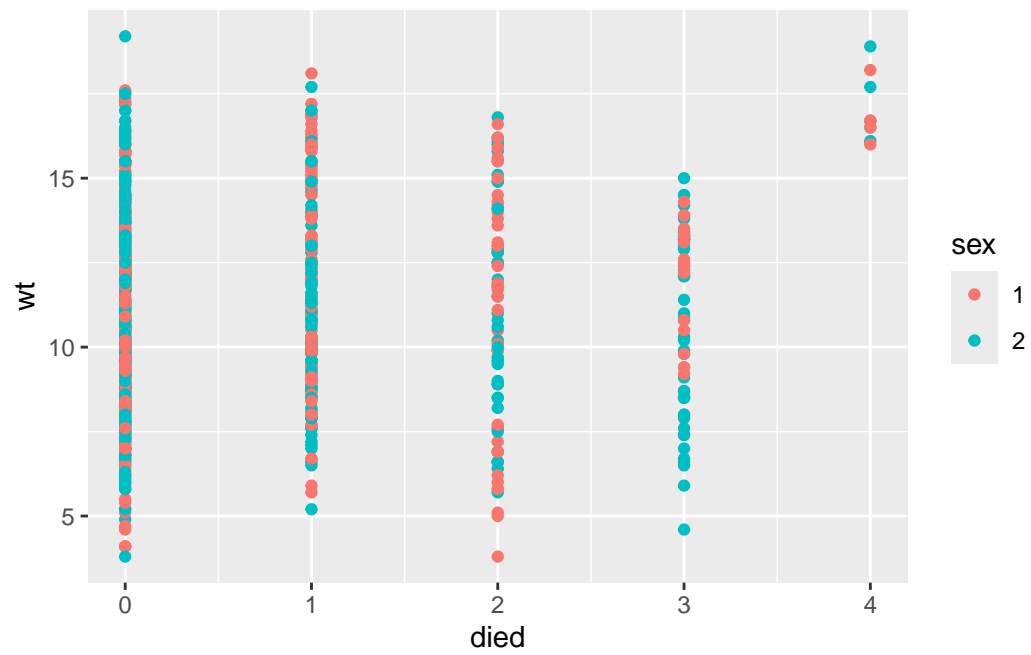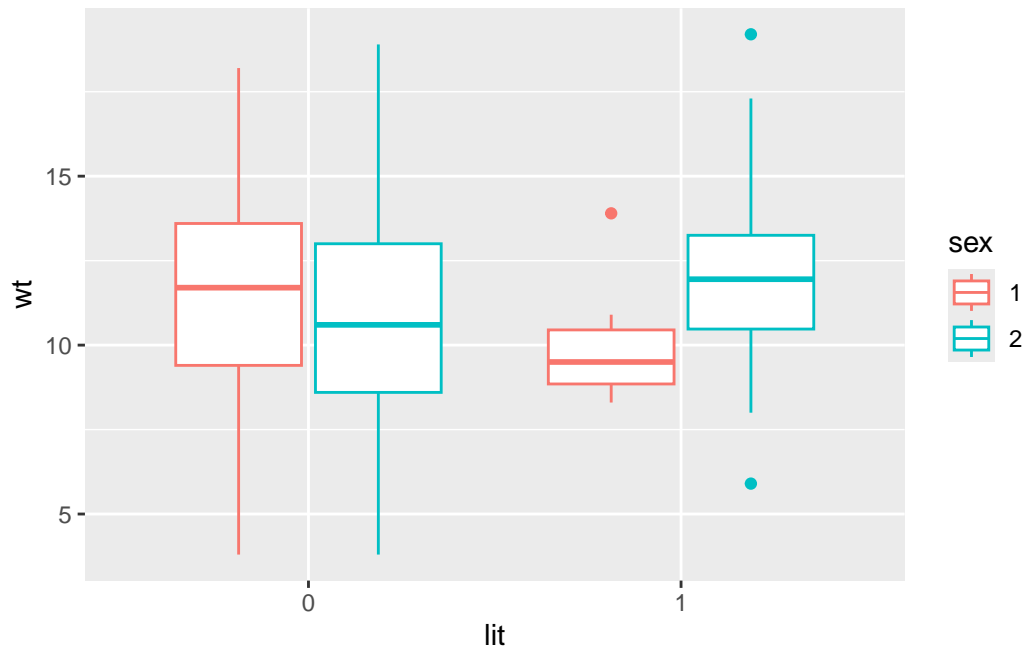
Frequentist Mixed Model

We now fit a mixed model. Based on the visualization, we will use sex, age, mother's age, and mother's literacy status as predictors since they appear to show at least some trend with weight. We exclude the number of children who died and the number of children born alive because they do not show any trends. Age will be a random effect in the model since weight may vary by age.

Below are the model diagnostics:

- In the posterior predictive check, the model-predicted lines mostly resemble the observed data line.
- In the fitted vs. residuals plot, the reference line appears to be mostly flat except at the start of the chart. This suggests the linear assumption may be violated.
- In the scale-location plot, the reference line is curved rather than flat. This suggests there may be heteroscedasticity.
- In the leverage vs. residuals plot, none of the points are outside of Cook's distance. Thus, there are most likely no outliers.
- In the collinearity plot, there appears age and sex appear to have a high collinearity. However, we will keep this in the model because dropping age and sex from the model would

mean we would have to interpret the model in their absence, which is not representative of reality.

- In the QQ-plot, most of the points are on the line. However, the points on the tails of the chart appear to not follow the straight line. This suggests that the residuals may not be normal.
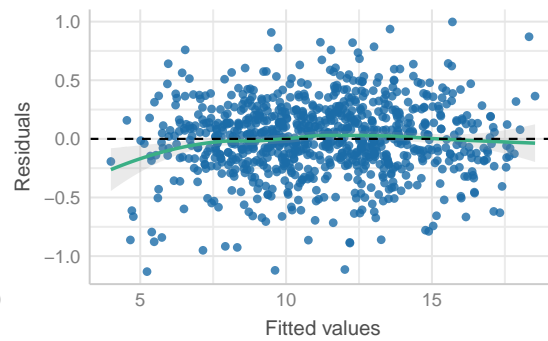
Posterior Predictive Check
Model–predicted lines should resemble observed data line

— Observed data — Model–predicted data
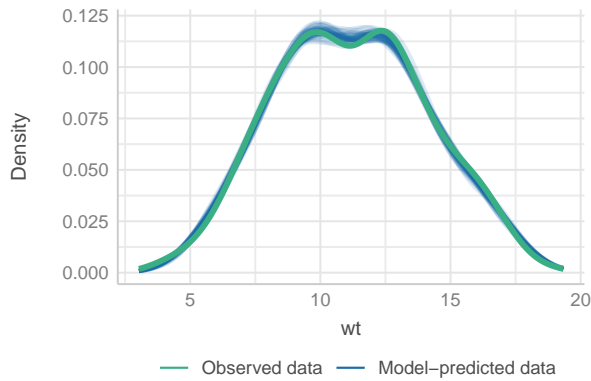
Linearity
Reference line should be flat and horizontal
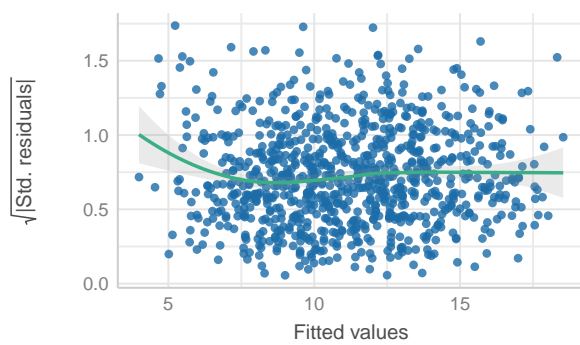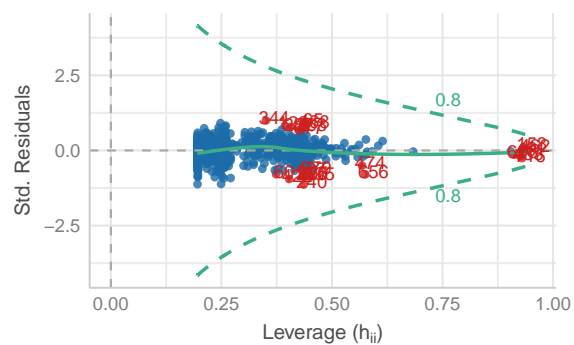

Homogeneity of Variance
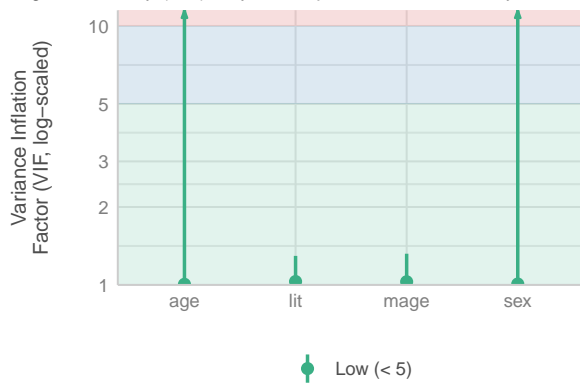Reference line should be flat and horizontal

Influential Observations
Points should be inside the contour lines
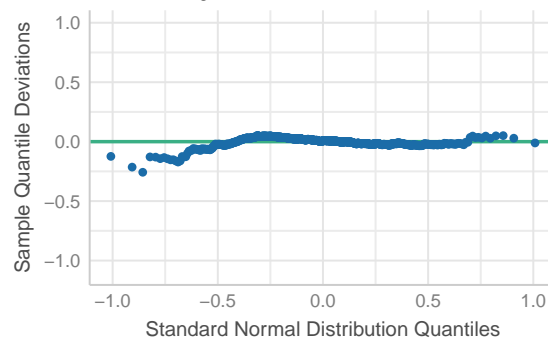

Collinearity
High collinearity (VIF) may inflate parameter uncertainty
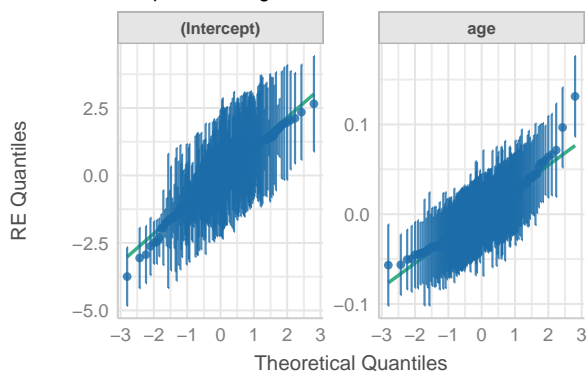

● Low (< 5)

Normality of Residuals
Dots should fall along the line

Normality of Random Effects (id)
Dots should be plotted along the line

In attempt to fix the non-linearity and heteroscedasticity, we tried applying log, inverse, and square root transformations to the response variable. However, these transformations did not improve the diagnostics (in fact, log and inverse transformations appeared to worsen the problems). I have left out the model diagnostics for these transformed models as they are lengthy to read.

Since we are unable to obtain a better model without a substantial amount of effort, we will keep the original model as shown in the summary below. We highlight some of the key takeaways from the summary:

- From the random effects section, we can see that age varies by a factor of about 0.0013 months across individuals.
- From the fixed effects section, we can see that age, mother's age, and sex appear to have a significant impact on weight (in order from greatest significance to least).

  - The mother's literacy status does not have a significant impact on weight.
  - For every month of increase in the child's age, weight increases by a factor of 2.524.
  - For every year of increase in the mother's age, the child's weight increases by a factor of 0.288.

- From the correlation of fixed effects section, there does not appear to be any high correlation between the predictors. Correlation is low.

```
Linear mixed model fit by REML ['lmerMod']
Formula: wt ~ sex + mage + lit + age + (age | id)
   Data: nepali2


REML criterion at convergence: 1702.1


Scaled residuals:
     Min        1Q    Median        3Q       Max
-3.01577  -0.45154   0.01913   0.50514   2.65686


Random effects:
 Groups   Name         Variance Std.Dev. Corr
```

```
 id        (Intercept) 1.933360 1.39045
           age              0.001297 0.03601   -0.58
 Residual                   0.140965 0.37545
Number of obs: 877, groups:  id, 197


Fixed effects:
             Estimate Std. Error t value
(Intercept)  5.037572    0.438910   11.477
sex2        -0.435561    0.179664   -2.424
mage         0.046251    0.014681    3.150
lit1         0.306055    0.446683    0.685
age          0.136673    0.003387   40.349


Correlation of Fixed Effects:
     (Intr) sex2    mage    lit1
sex2 -0.165
mage -0.935 -0.032
lit1 -0.166 -0.087  0.152
age  -0.149  0.023 -0.072 -0.023
```

Now we use the model to predict weight. From the scatterplot below, we can see that the model predicts the data fairly well. This is especially the case when age is low as the distance between predictions and observations is relatively low. As age increases, the distance between the predictions and observations also increase. Further work to reduce the heteroscedasticity may improve future model predictions.

Predicted values of wt
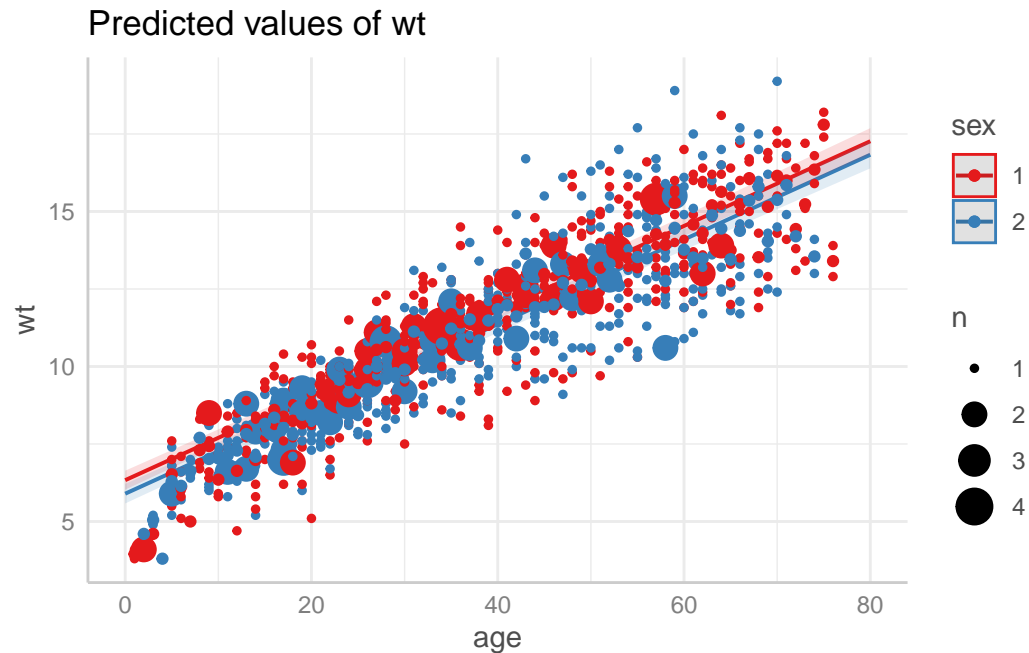
The original paper by West et al. (Jr. et al., 1997) performs this analysis using different variables. Therefore, it is difficult to compare their model with ours.

Bayesian Mixed Model

We will redo Question 1 using the rstanarm and MCMCglmm packages. We'll start with rstanarm.

```
Priors for model 'priorpred'
------
Intercept (after predictors centered)
  Specified prior:
    ~ normal(location = 11, scale = 2.5)
  Adjusted prior:
    ~ normal(location = 11, scale = 7.4)

Coefficients
  Specified prior:
    ~ normal(location = [0,0,0,...], scale = [2.5,2.5,2.5,...])
```

```
   Adjusted prior:

     ~ normal(location = [0,0,0,...], scale = [7.36,7.36,7.36,...])


Auxiliary (sigma)
  Specified prior:
     ~ exponential(rate = 1)
  Adjusted prior:
     ~ exponential(rate = 0.34)


Covariance
 ~ decov(reg. = 1, conc. = 1, shape = 1, scale = 1)
------
See help('prior_summary.stanreg') for more details


SAMPLING FOR MODEL 'continuous' NOW (CHAIN 1).
Chain 1:
Chain 1: Gradient evaluation took 0.000559 seconds
Chain 1: 1000 transitions using 10 leapfrog steps per transition would take 5.59 seconds.
Chain 1: Adjust your expectations accordingly!
Chain 1:
Chain 1:
Chain 1: Iteration:    1 / 2000 [  0%]  (Warmup)
Chain 1: Iteration:  200 / 2000 [ 10%]  (Warmup)
Chain 1: Iteration:  400 / 2000 [ 20%]  (Warmup)
Chain 1: Iteration:  600 / 2000 [ 30%]  (Warmup)
Chain 1: Iteration:  800 / 2000 [ 40%]  (Warmup)
Chain 1: Iteration: 1000 / 2000 [ 50%]  (Warmup)
Chain 1: Iteration: 1001 / 2000 [ 50%]  (Sampling)
Chain 1: Iteration: 1200 / 2000 [ 60%]  (Sampling)
Chain 1: Iteration: 1400 / 2000 [ 70%]  (Sampling)
Chain 1: Iteration: 1600 / 2000 [ 80%]  (Sampling)
Chain 1: Iteration: 1800 / 2000 [ 90%]  (Sampling)
```

```
Chain 1: Iteration: 2000 / 2000 [100%]  (Sampling)
Chain 1:
Chain 1:  Elapsed Time: 218.475 seconds (Warm-up)
Chain 1:                115.449 seconds (Sampling)
Chain 1:                333.924 seconds (Total)
Chain 1:


SAMPLING FOR MODEL 'continuous' NOW (CHAIN 2).
Chain 2:
Chain 2: Gradient evaluation took 0.000716 seconds
Chain 2: 1000 transitions using 10 leapfrog steps per transition would take 7.16 seconds.
Chain 2: Adjust your expectations accordingly!
Chain 2:
Chain 2:
Chain 2: Iteration:    1 / 2000 [  0%]  (Warmup)
Chain 2: Iteration:  200 / 2000 [ 10%]  (Warmup)
Chain 2: Iteration:  400 / 2000 [ 20%]  (Warmup)
Chain 2: Iteration:  600 / 2000 [ 30%]  (Warmup)
Chain 2: Iteration:  800 / 2000 [ 40%]  (Warmup)
Chain 2: Iteration: 1000 / 2000 [ 50%]  (Warmup)
Chain 2: Iteration: 1001 / 2000 [ 50%]  (Sampling)
Chain 2: Iteration: 1200 / 2000 [ 60%]  (Sampling)
Chain 2: Iteration: 1400 / 2000 [ 70%]  (Sampling)
Chain 2: Iteration: 1600 / 2000 [ 80%]  (Sampling)
Chain 2: Iteration: 1800 / 2000 [ 90%]  (Sampling)
Chain 2: Iteration: 2000 / 2000 [100%]  (Sampling)
Chain 2:
Chain 2:  Elapsed Time: 196.865 seconds (Warm-up)
Chain 2:                119.228 seconds (Sampling)
Chain 2:                316.093 seconds (Total)
Chain 2:
```

```
SAMPLING FOR MODEL 'continuous' NOW (CHAIN 3).

Chain 3:

Chain 3: Gradient evaluation took 0.00049 seconds

Chain 3: 1000 transitions using 10 leapfrog steps per transition would take 4.9 seconds.

Chain 3: Adjust your expectations accordingly!

Chain 3:

Chain 3:

Chain 3: Iteration:    1 / 2000 [  0%]  (Warmup)

Chain 3: Iteration:  200 / 2000 [ 10%]  (Warmup)

Chain 3: Iteration:  400 / 2000 [ 20%]  (Warmup)

Chain 3: Iteration:  600 / 2000 [ 30%]  (Warmup)

Chain 3: Iteration:  800 / 2000 [ 40%]  (Warmup)

Chain 3: Iteration: 1000 / 2000 [ 50%]  (Warmup)

Chain 3: Iteration: 1001 / 2000 [ 50%]  (Sampling)

Chain 3: Iteration: 1200 / 2000 [ 60%]  (Sampling)

Chain 3: Iteration: 1400 / 2000 [ 70%]  (Sampling)

Chain 3: Iteration: 1600 / 2000 [ 80%]  (Sampling)

Chain 3: Iteration: 1800 / 2000 [ 90%]  (Sampling)

Chain 3: Iteration: 2000 / 2000 [100%]  (Sampling)

Chain 3:

Chain 3:  Elapsed Time: 186.515 seconds (Warm-up)

Chain 3:                117.239 seconds (Sampling)

Chain 3:                303.754 seconds (Total)

Chain 3:


SAMPLING FOR MODEL 'continuous' NOW (CHAIN 4).

Chain 4:

Chain 4: Gradient evaluation took 0.000492 seconds

Chain 4: 1000 transitions using 10 leapfrog steps per transition would take 4.92 seconds.

Chain 4: Adjust your expectations accordingly!

Chain 4:

Chain 4:
```

```
Chain 4: Iteration:    1 / 2000 [  0%]  (Warmup)
Chain 4: Iteration:  200 / 2000 [ 10%]  (Warmup)
Chain 4: Iteration:  400 / 2000 [ 20%]  (Warmup)
Chain 4: Iteration:  600 / 2000 [ 30%]  (Warmup)
Chain 4: Iteration:  800 / 2000 [ 40%]  (Warmup)
Chain 4: Iteration: 1000 / 2000 [ 50%]  (Warmup)
Chain 4: Iteration: 1001 / 2000 [ 50%]  (Sampling)
Chain 4: Iteration: 1200 / 2000 [ 60%]  (Sampling)
Chain 4: Iteration: 1400 / 2000 [ 70%]  (Sampling)
Chain 4: Iteration: 1600 / 2000 [ 80%]  (Sampling)
Chain 4: Iteration: 1800 / 2000 [ 90%]  (Sampling)
Chain 4: Iteration: 2000 / 2000 [100%]  (Sampling)
Chain 4:
Chain 4:  Elapsed Time: 199.239 seconds (Warm-up)
Chain 4:                120.863 seconds (Sampling)
Chain 4:                320.102 seconds (Total)
Chain 4:
```

The following are diagnostics for the model:

- According to the (recent) rule of thumb, Rhat < 1.01 means the Markov chains have converged (i.e. they have mixed well). For all parameters, Rhat is below this 1.01, so the chains appear to have mixed well.
- The effective sample sizes (ESS) for all of the parameters are greater than 1000, so we can say it is sufficient for stable estimates.
- The Monte Carlo standard error (MCSE) appears to be fair since all of the values are small.

```
      Parameter   Rhat  ESS     MCSE
1   (Intercept) 1.0003 1045 0.002954
2           age 0.9996 3491 0.001074
397         lit 1.0006 2245 0.001715
398        mage 1.0009 1159 0.002746
399         sex 1.0013 1309 0.002495
```

Below is the trace plot. In each of the plots, the chains are all following a similar "caterpillar-like" pattern. In other words, the pattern is not too flat and does not take too many consecutive steps in the same direction. Thus, we can say the model appears to be mixing well.
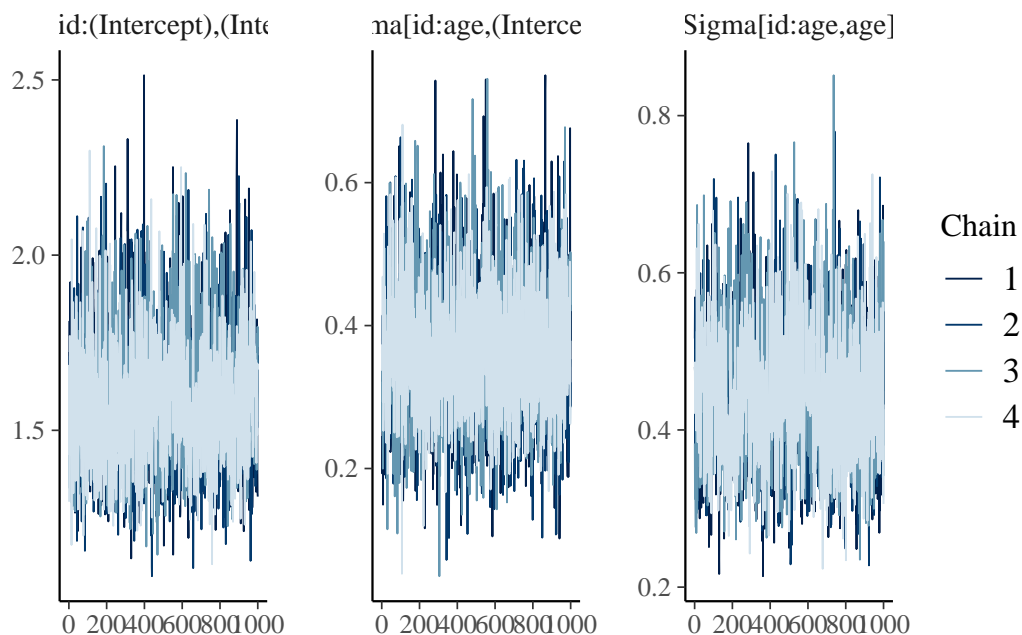
Below is the coefficient plot for the fit. The coefficients for all variables except sex are positive. This means an increase in age, mother's age, and mother's literacy are associated with an increase in weight. Age has a considerably large estimate compared to the other variables. Additionally, when sex increases (i.e. is female), then weight decreases.

It's important to note that the 95% credible interval for mother's literacy contains 0 so there isn't enough statistical evidence to reject the possibility that mother's literacy does not have an influence on weight. All other credible intervals do not contain 0, so there is statistical evidence to reject the possibility that the variables do not have an influence on weight.
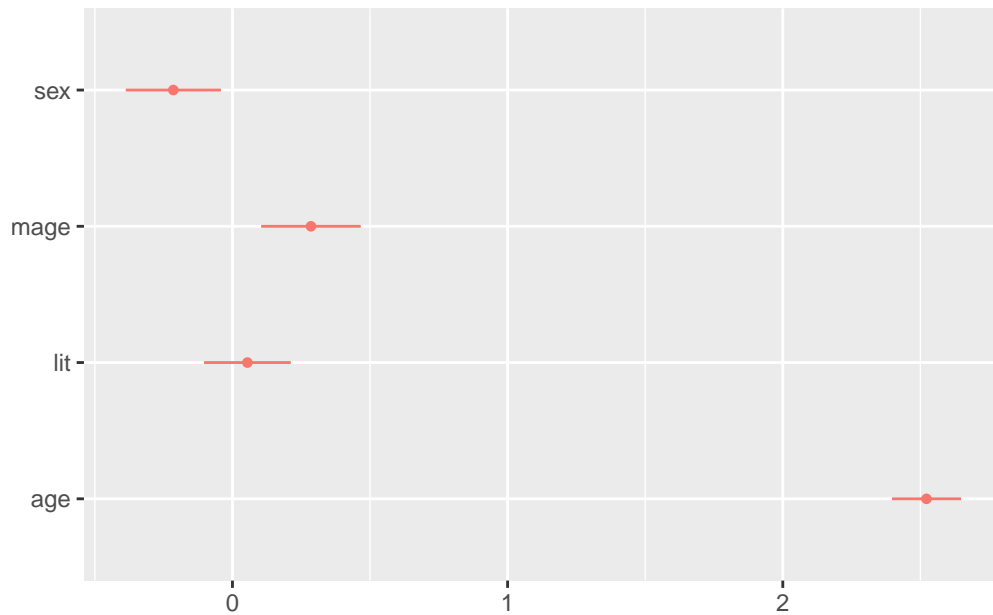
We now fit the model using MCMCglmm.

Below is the trace plot. Similar to the previous model, the trace plots for the MCMC fit has a "caterpillar-like" pattern too. The pattern is not too flat and does not take too many consecutive steps in the same direction. Thus, we can say the model appears to be mixing well.
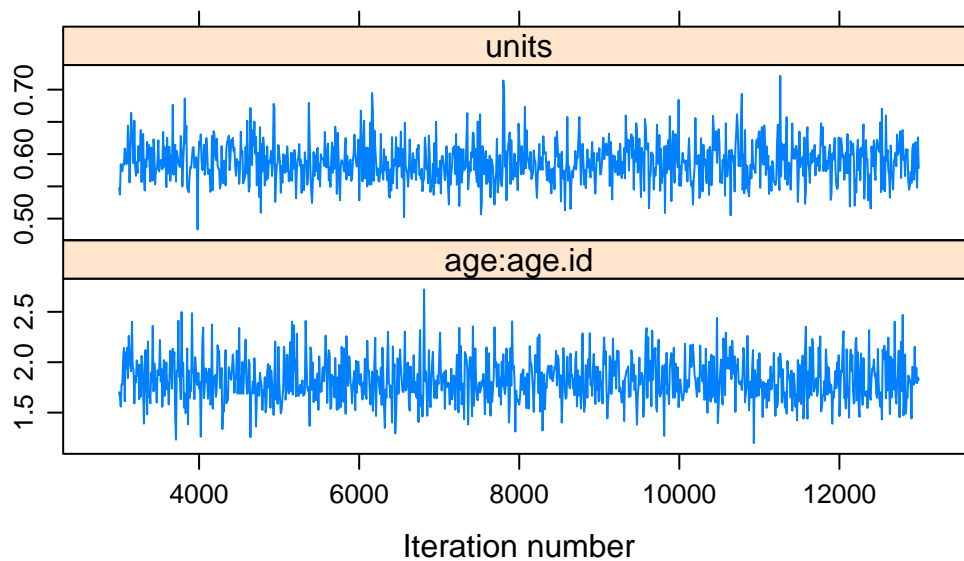

Below is the coefficient plot for the model. The results are very similar to the coefficient plot for the rstanarm model. The coefficients for all variables except sex are positive. This means an increase in age, mother's age, and mother's literacy are associated with an increase in weight. Age has a considerably large estimate compared to the other variables. Additionally, when sex

increases (i.e. is female), then weight decreases.

The 95% credible interval for mother's literacy contains 0 so there isn't enough statistical evidence to reject the possibility that mother's literacy does not have an influence on weight. All other credible intervals do not contain 0, so there is statistical evidence to reject the possibility that the variables do not have an influence on weight. The credible interval for mother's literacy is also slightly wider than the remaining variables.
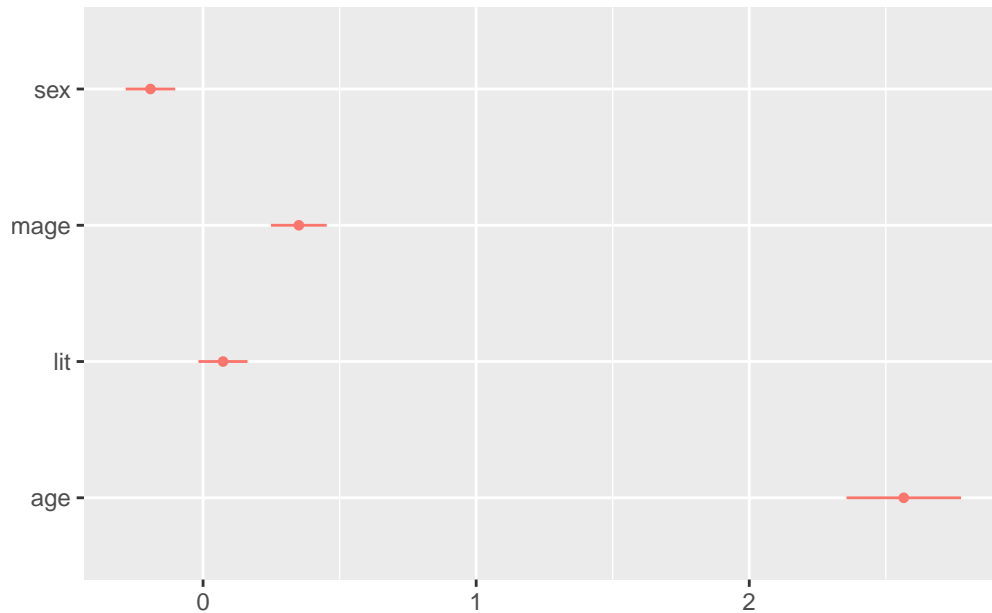
Here is the summary of the model. From the p-values, it appears that age, mage, and sex are the significant while sex is not (for a significance level of 0.05 or less).

```
 Iterations = 3001:12991

 Thinning interval  = 10

 Sample size  = 1000


 DIC: 2192.766


 G-structure:  ~us(age):id


          post.mean l-95% CI u-95% CI eff.samp
age:age.id    1.828    1.39    2.262     1000
```

```
 R-structure:  ~units


      post.mean l-95% CI u-95% CI eff.samp
units    0.5883   0.5258    0.652      1000


 Location effects: wt ~ sex + mage + lit + age


          post.mean l-95% CI u-95% CI eff.samp  pMCMC
(Intercept)  11.43222 11.33603 11.52048     1000 <0.001 ***
sex          -0.19270 -0.28607 -0.10520     1000 <0.001 ***
mage          0.35076  0.24832  0.44663     1099 <0.001 ***
lit           0.07308 -0.01669  0.15771     1170  0.114
age           2.56570  2.36217  2.77875     1000 <0.001 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

References

Jr., K. P. W., LeClerq, S. C., Shrestha, S. R., Wu, L. S.-F., Pradhan, E. K., Khatry, S. K., Katz, J., Adhikari, R., & Sommer, A. (1997). Effects of vitamin a on growth of vitamin a-deficient children: Field studies in nepal. In *The Journal of Nutrition* (10; Vol. 127, pp. 1957–1965). https://doi.org/10.1093/jn/127.10.1957

Appendix

```r
# ----- SETUP ----- #
library(tidyverse)
library(faraway)
library(lme4)
library(performance)
library(ggeffects)
library(rstanarm)
library(MCMCglmm)
library(bayesplot)
library(bayestestR)
library(dotwhisker)
library(lattice)
library(broom.mixed)


# ----- DATA TRANSFORMATION ----- #
nepali2 <- nepali %>%
  filter(!is.na(wt)) %>%
  dplyr::select(-ht) %>%
  mutate(sex = factor(sex),
         lit = factor(lit))
head(nepali2,5)


# ----- DATA VISUALIZATION ----- #
ggplot(data=nepali2, aes(x=age, y=wt, color=sex)) + geom_point()
ggplot(data=nepali2, aes(x=mage, y=wt, color=sex)) + geom_point()
ggplot(data=nepali2, aes(x=died, y=wt, color=sex)) + geom_point()
ggplot(data=nepali2, aes(x=alive, y=wt, color=sex)) + geom_point()
ggplot(data=nepali2, aes(x=lit, y=wt, color=sex)) + geom_boxplot()


# ----- FREQUENTIST MIXED MODEL ----- #
```

18

```r
nep_mod_mm <- lmer(wt ~ sex + mage + lit + age + (age | id), data = nepali2, REML = TRUE

performance::check_model(nep_mod_mm)

summary(nep_mod_mm)

nep_mod_mm_pred <- ggpredict(nep_mod_mm, terms = c("age", "sex"))

pp <- plot(nep_mod_mm_pred)

bb <- ggplot_build(pp)


pp + stat_sum(data = nepali2,

              aes(x = age, y = wt, colour = sex),

              fill = NA) +

    stat_summary(data = nepali2,

                 aes(x = age, y = wt, colour = sex),

                 fill = NA,

                 geom = "point",

                 fun = mean)


# ----- DATA TRANSFORMATION ----- #

nepali3 <- nepali %>%

  filter(!is.na(wt)) %>%

  dplyr::select(-ht) %>%

  mutate(sex = scale(as.numeric(sex)),

         lit = scale(as.numeric(lit)),

         mage = scale(mage),

         died = scale(died),

         alive = scale(alive),

         age = scale(age))

# Prior prediction

priorpred <- rstanarm::stan_lmer(wt ~ sex + mage + lit + age + (age | id),

                      prior_PD = TRUE, data = nepali3, chains = 1,

                      seed = 101,

                      refresh = 0)
```

```r
prior_summary(priorpred)
# Posterior prediction
stanfit <- rstanarm::stan_lmer(wt ~ sex + mage + lit + age + (age | id),
                data = nepali3,
                chains = 4,
                seed = 101)
# Diagnostics
print(bayestestR::diagnostic_posterior(stanfit), digits = 4)
# Trace plot
bayesplot::mcmc_trace(stanfit, regex_pars="Sigma")
# Coefficient plot
dotwhisker::dwplot(tidy(stanfit,effects = "fixed"))
# Using MCMCglmm
mcmcfit <- MCMCglmm(wt ~ sex + mage + lit + age, random = ~us(age):id,
                data = nepali3,
                verbose=FALSE)
# Trace plot
lattice::xyplot(mcmcfit$VCV)
# Coefficient plot
dotwhisker::dwplot(tidy(mcmcfit,effects = "fixed"))
summary(mcmcfit)
```