Homework 3 Solution

```r
knitr::opts_chunk$set(warning = FALSE)
library(nlme)
library(tidyverse)
library(gee)
library(lmtest)
library(splines)
library(ggplot2)
library(gridExtra)
library(dplyr)

options(digits = 3)
```

Part I. Get familiar with the data

```r
load("NepalAnthro.rdata")
d1 = nepal.anthro[,c("id", "alive", "age", "wt","fuvisit")]
d1$parity = factor(ifelse(d1$alive <= 4, 0, 1),
            levels = 0 : 1,
            labels = c("1 to 4 live births",
                       "5 or more live births"))
d2 = d1[complete.cases(d1),]
```

A. Mother's Parity

Make a table of mother's parity (alive variable). Ideally, we would compare children of nulliparous women to categories of women of parity > 0. However, in this dataset, there are only 19 children from nulliparous women. So, we will create two categories of women: parity le 3 (i.e. 1 to 4 live births) vs. parity > 3 (5 or more live births).

```r
## Table for alive
table(d2 %>%
  group_by(id) %>%
  summarise(alive = alive[1]) %>%
  select(alive))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
##
##  1  2  3  4  5  6  7  8  9 10 14
## 21 28 34 36 21 20 12 12  7  4  2
```

```r
## Table for parity
table(d2 %>%
  group_by(id) %>%
```

```
  summarise(parity = parity[1]) %>%
  select(parity))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

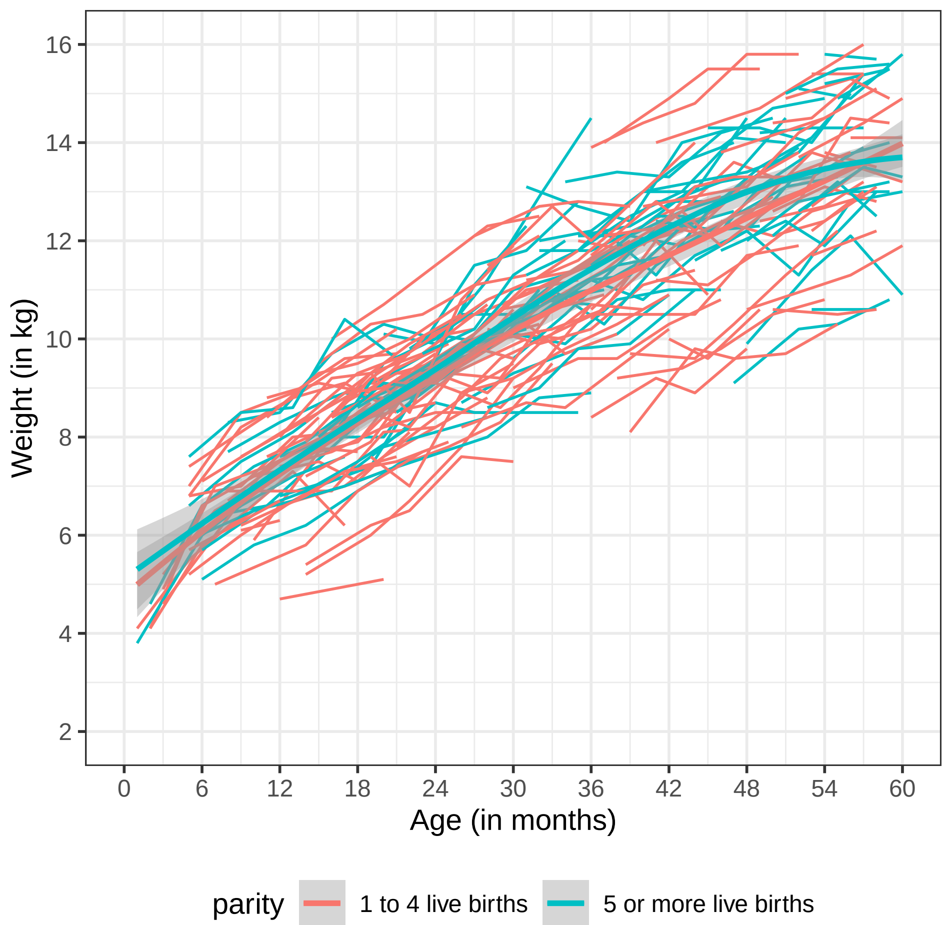
```
##
##    1 to 4 live births 5 or more live births
##                   119                    78
```

B. Spaghetti plot

Make a spaghetti plot of children's weight as a function of age; connecting the measured weights within a child over time. Color code the data by parity group. Add smoothing splines for each parity group. Note any similarities or differences in the growth rates across the two parity groups.

```
# load and wrangle data
d2 %>%
  ggplot(aes(x = age, y = wt, col = parity)) +
  geom_line(aes(group = id)) +
  labs(x="Age (in months)",y="Weight (in kg)") +
  scale_y_continuous(breaks=seq(2,16,2),limits=c(2,16)) +
  scale_x_continuous(breaks=seq(0,60,6),limits=c(0,60))+
  theme_bw()+
  theme(legend.position="bottom",legend.box="horizontal")+ geom_smooth()
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



The average weight as a function of age looks roughly similar across the two parity groups. In addition, the growth rate for weight changes over the course of age (similar to our findings in prior analyses of these data).

Part II. Model checking and recommendations

1. Model checking

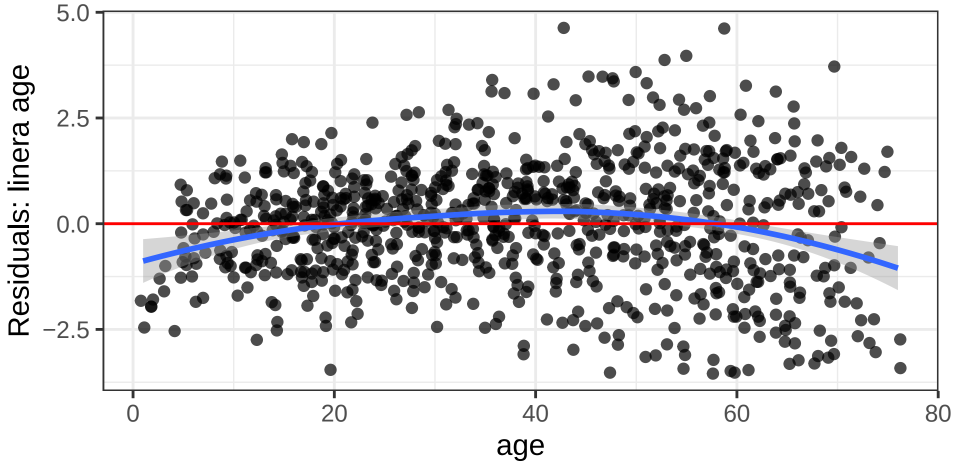
Conduct appropriate checking of this model; i.e. check for appropriateness of the mean model, and the independence and constant variance assumptions for the residuals.

```
lm = lm(wt ~ age * parity, data = d2)
```

We first assess the appropriateness of the mean function

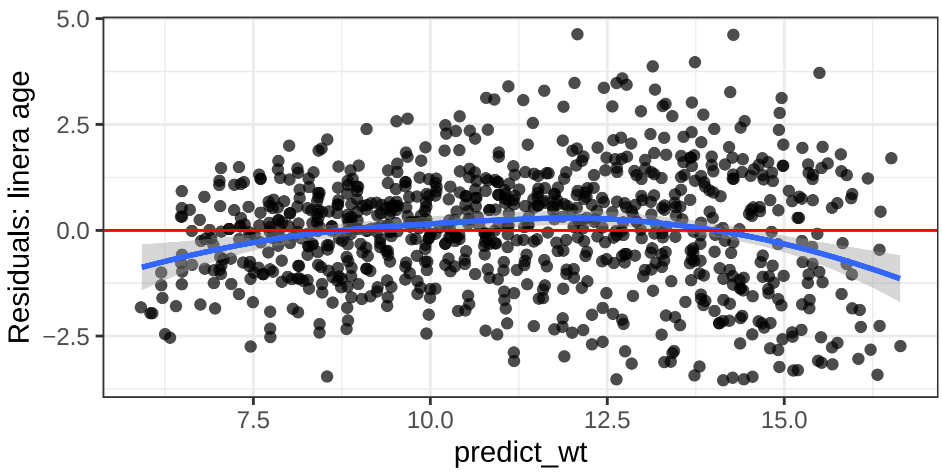
```
## Test age against residuals
tibble(age = d2$age, residual = residuals(lm)) %>%
  ggplot(aes(x = age, y = residual)) +
  geom_jitter(alpha = 0.7) +
  geom_smooth() + theme_bw() +
  geom_hline(yintercept = 0, color = "red") +
  labs(y = "Residuals: linera age")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```




```r
## Test predicted weight against residuals
tibble(predict_wt = predict(lm), residual = residuals(lm)) %>%
  ggplot(aes(x = predict_wt, y = residual)) +
  geom_jitter(alpha = 0.7) +
  geom_smooth() + theme_bw() +
  geom_hline(yintercept = 0, color = "red") +
  labs(y = "Residuals: linera age")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



Based on the plots comparing the estimated residuals vs. age and residuals vs. predicted weight, we observe that the model assuming average weight changes linearly with age underestimates weight for ages < 20 months, overestimates weight for ages between roughly 20 to 60 months and underestimates weight for ages > 60 months. Assuming that the average weight is a linear function of age is not supported by the data; we should add some non-linearity to the average weight / age relationship.

Next, we test the independence assumption

```r
nep.wide <- d2 %>%
  mutate(residual = residuals(lm)) %>%
  select("id", "residual", "fuvisit") %>%
  spread(fuvisit, residual)

## Use pairwise complete observations
```

```r
nep.wide %>%
  select("0","1","2","3","4") %>%
  cor(use = "pairwise.complete.obs")
```

```
##       0     1     2     3     4
## 0 1.000 0.915 0.888 0.879 0.858
## 1 0.915 1.000 0.938 0.907 0.872
## 2 0.888 0.938 1.000 0.950 0.924
## 3 0.879 0.907 0.950 1.000 0.934
## 4 0.858 0.872 0.924 0.934 1.000
```

```r
## Use complete observations
nep.wide %>%
  select("0","1","2","3","4") %>%
  cor(use = "complete.obs")
```

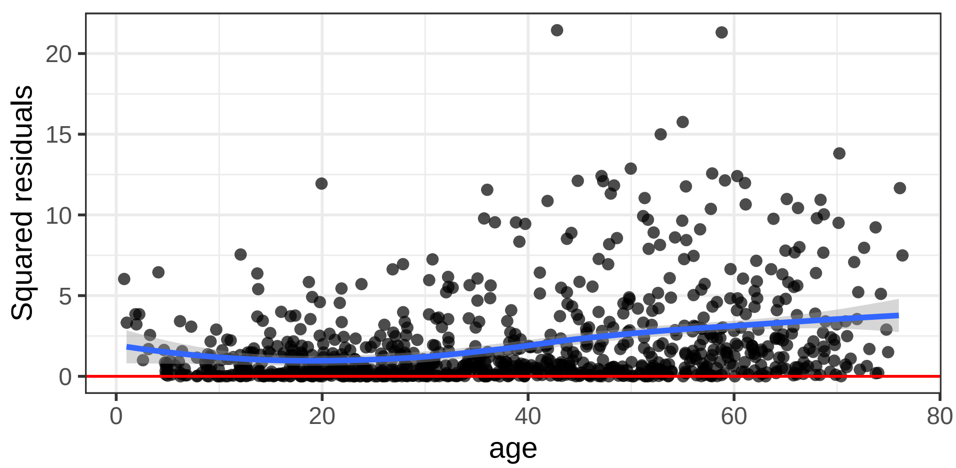
```
##       0     1     2     3     4
## 0 1.000 0.913 0.887 0.879 0.852
## 1 0.913 1.000 0.936 0.909 0.878
## 2 0.887 0.936 1.000 0.946 0.920
## 3 0.879 0.909 0.946 1.000 0.932
## 4 0.852 0.878 0.920 0.932 1.000
```

Based on the correlations between residuals from the same child but at different assessments, the residuals are highly correlated with each other when the assessments are 4 months apart and the correlation decays as the time between assessments increases. Therefore, this is evidence that the independence assumption does not hold. In fact, we anticipated some correlation in the data based on the data collection process.

Finally we check the constant variance assumption:

```r
## Test age against residuals
tibble(age = d2$age, residual = residuals(lm)^2) %>%
  ggplot(aes(x = age, y = residual)) +
  geom_jitter(alpha = 0.7) +
  geom_smooth() + theme_bw() +
  geom_hline(yintercept = 0, color = "red") +
  labs(y = "Squared residuals")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



The plot of the squared residuals vs. age shows that the variance (mean of the squared residuals) increases starting at ages > 30 months.

2. Alternative model

Based on your model checking, propose an alternative model for the data that can address the first goal of the analysis (i.e. determine if the growth rates of children differ by mother's parity (number of previous live births) while satisfying the observed patterns in data with respect to the mean model and distribution of residuals. NOTE: If you modify the mean model, you may want to iterate between model checking for the mean.

Here, we fit a natural cubic spline regression model,

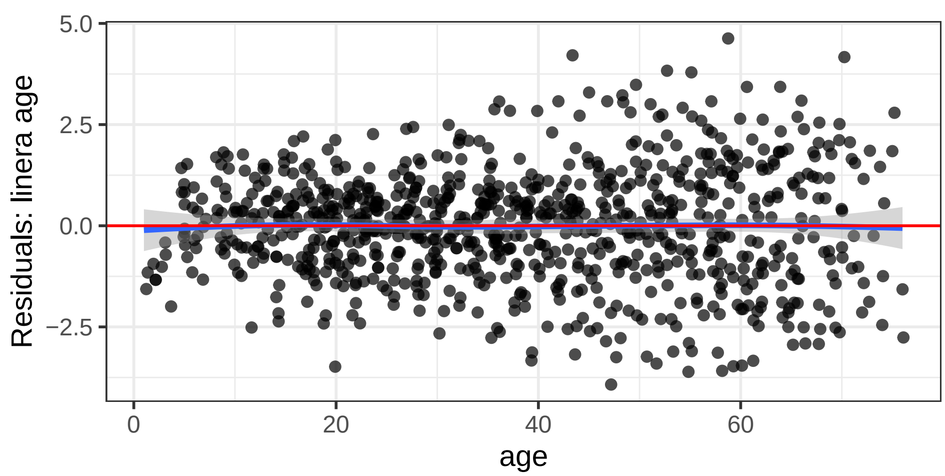
$$E(Y_{ij}|age_{ij}, parity_i) = B_0 + B_{ns,0} \cdot ns(age_{ij}, 3) + B_1 \cdot parity + B_{ns,1} \cdot ns(age_{ij}, 3) \cdot parity_i$$

Then we have $Y_i \sim MVN(\mu_i, V_i)$ where the diagonal elements of V_i are $Var(Y_{ij}) = \sigma^2_{ij}$, where $\sigma^2_{ij} = f(age_i j)$ and $Corr(Y_{ij}, Y_{ik}) = \rho^{|j-k|}$.

Below we confirm that this more flexible mean model has mean zero residuals as a function of age and the predicted weights.

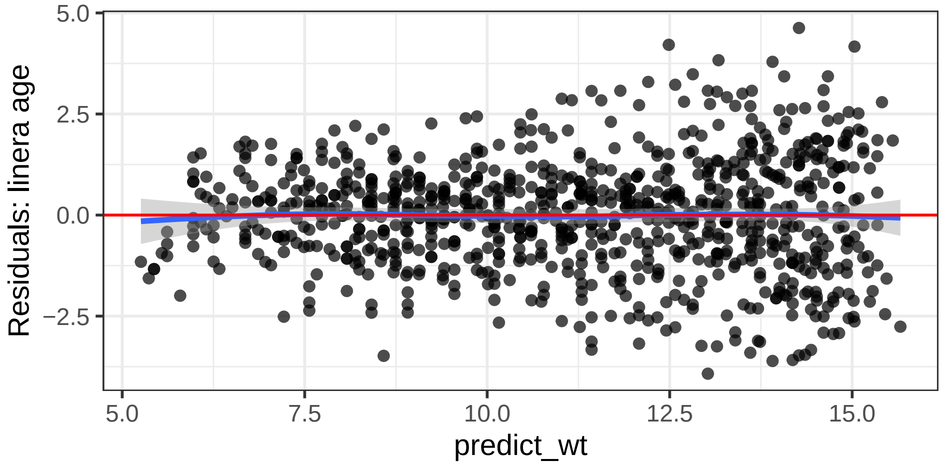
```
lm2 = lm(wt ~ ns(age, 3) * parity, data = d2)
d2$predicted_wt = lm2$fitted.values

## Test age against residuals
tibble(age = d2$age, residual = residuals(lm2)) %>%
  ggplot(aes(x = age, y = residual)) +
  geom_jitter(alpha = 0.7) +
  geom_smooth() + theme_bw() +
  geom_hline(yintercept = 0, color = "red") +
  labs(y = "Residuals: linera age")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```




```
## Test predicted weight against residuals
tibble(predict_wt = predict(lm2), residual = residuals(lm2)) %>%
  ggplot(aes(x = predict_wt, y = residual)) +
  geom_jitter(alpha = 0.7) +
  geom_smooth() + theme_bw() +
  geom_hline(yintercept = 0, color = "red") +
  labs(y = "Residuals: linera age")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



Part III: Marginal model for longitudinal data

1. Fit the model

Use the gls function in R to fit the model you proposed in Part II.

```r
## Fit a full model
gls_full <- gls(wt ~ ns(age, 3) * parity, data = d2, corAR1(, form =  ~fuvisit | id ), weights = varFunc
summary(gls_full)
```

```
## Generalized least squares fit by REML
##   Model: wt ~ ns(age, 3) * parity
##   Data: d2
##    AIC  BIC logLik
##   2055 2102  -1017
##
## Correlation Structure: ARMA(1,0)
##  Formula: ~fuvisit | id
##  Parameter estimate(s):
##  Phi1
## 0.901
## Variance function:
##  Structure: fixed weights
##  Formula: ~age
##
## Coefficients:
##                                        Value Std.Error t-value p-value
## (Intercept)                             4.33     0.084    51.3   0.000
## ns(age, 3)1                             6.06     0.338    17.9   0.000
## ns(age, 3)2                            15.50     0.412    37.6   0.000
## ns(age, 3)3                             8.59     0.522    16.5   0.000
## parity5 or more live births            -0.11     0.139    -0.8   0.434
## ns(age, 3)1:parity5 or more live births  0.59     0.508     1.2   0.244
## ns(age, 3)2:parity5 or more live births  0.82     0.610     1.3   0.179
## ns(age, 3)3:parity5 or more live births  0.11     0.731     0.1   0.882
```

```
##
##  Correlation:
##                                        (Intr) n(,3)1 n(,3)2 n(,3)3 p5omlb
## ns(age, 3)1                            -0.107
## ns(age, 3)2                            -0.359  0.011
## ns(age, 3)3                            -0.080 -0.036  0.748
## parity5 or more live births            -0.605  0.065  0.217  0.049
## ns(age, 3)1:parity5 or more live births  0.071 -0.666 -0.007  0.024 -0.157
## ns(age, 3)2:parity5 or more live births  0.243 -0.007 -0.676 -0.506 -0.376
## ns(age, 3)3:parity5 or more live births  0.057  0.025 -0.535 -0.715 -0.100
##                                        n(,3)1omlb n(,3)2omlb
## ns(age, 3)1
## ns(age, 3)2
## ns(age, 3)3
## parity5 or more live births
## ns(age, 3)1:parity5 or more live births
## ns(age, 3)2:parity5 or more live births  0.024
## ns(age, 3)3:parity5 or more live births  0.005      0.686
##
## Standardized residuals:
##     Min      Q1     Med      Q3     Max
## -3.0928 -0.5239  0.0877  0.6488  4.0520
##
## Residual standard error: 0.258
## Degrees of freedom: 877 total; 869 residual
```

2. Likelihood Ratio Test

Conduct a likelihood ratio test to address the first goal of the analysis; i.e. to determine if the growth rates of children differ by mother's parity (number of previous live births).

To test whether the growth rates of weight differ by parity group, you want to compare the model with a) main terms for age, main term for parity and interaction of age terms with parity to a model with b) main terms for age and a main term for parity. So that the following model needs to be fitted

```r
## Fit a reduced model
gls_reduced <- gls(wt ~ ns(age, 3) + parity, data = d2, corAR1(, form =  ~fuvisit | id ), weights = varF
summary(gls_reduced)
```

```
## Generalized least squares fit by REML
##   Model: wt ~ ns(age, 3) + parity
##   Data: d2
##     AIC  BIC logLik
##   2055 2088  -1020
##
## Correlation Structure: ARMA(1,0)
##  Formula: ~fuvisit | id
##  Parameter estimate(s):
##  Phi1
## 0.901
## Variance function:
##  Structure: fixed weights
##  Formula: ~age
##
```

```
## Coefficients:
##                            Value Std.Error t-value p-value
## (Intercept)                 4.29     0.081    53.1   0.000
## ns(age, 3)1                 6.33     0.252    25.1   0.000
## ns(age, 3)2                15.83     0.301    52.5   0.000
## ns(age, 3)3                 8.65     0.365    23.7   0.000
## parity5 or more live births  0.02     0.124     0.1   0.899
##
##  Correlation:
##                             (Intr) n(,3)1 n(,3)2 n(,3)3
## ns(age, 3)1                 -0.080
## ns(age, 3)2                 -0.300  0.028
## ns(age, 3)3                 -0.052  0.002  0.699
## parity5 or more live births -0.555 -0.066 -0.023 -0.045
##
## Standardized residuals:
##     Min      Q1     Med      Q3     Max
## -3.1470 -0.5618  0.0803  0.6510  3.9839
##
## Residual standard error: 0.258
## Degrees of freedom: 877 total; 872 residual
```

```r
## Likelihood ratio test
lrtest(gls_reduced, gls_full)
```

```
## Likelihood ratio test
##
## Model 1: wt ~ ns(age, 3) + parity
## Model 2: wt ~ ns(age, 3) * parity
##   #Df LogLik Df Chisq Pr(>Chisq)
## 1   7  -1020
## 2  10  -1017  3  6.09       0.11
```

The likelihood ratio test returns a p-value of 0.11, which indicates that the data do not support rejecting the null hypothesis; i.e. there is no evidence in the data suggesting that the growth rates of weight differ by mother's parity.

3. Additional models

In addition, fit the mean model you proposed in Part II using the gee function but where you allow the correlation structure to be: independence, exchangeable and unstructured. Note similarities and differences in the "naive" (model) and "robust" standard error estimates. Compare the estimated coefficients/standard errors from the gls fit to those obtained using gee.

```r
## GEE under independent condition
gee_independence <- suppressMessages(gee(wt ~ ns(age, 3) * parity, data = d2, id = id, corstr = "indepe
```

```
##                             (Intercept)                             ns(age, 3)1
##                                  5.2563                                  6.3615
##                             ns(age, 3)2                             ns(age, 3)3
##                                 13.1691                                  8.0245
##             parity5 or more live births ns(age, 3)1:parity5 or more live births
##                                  0.1078                                  0.7742
## ns(age, 3)2:parity5 or more live births ns(age, 3)3:parity5 or more live births
##                                  0.0471                                 -0.2718
```

```r
## GEE under exchangealbe condition
gee_exchangeable <- suppressMessages(gee(wt ~ ns(age, 3) * parity, data = d2, id = id, corstr = "exchang
```

```
##                               (Intercept)                             ns(age, 3)1
##                                    5.2563                                  6.3615
##                               ns(age, 3)2                             ns(age, 3)3
##                                   13.1691                                  8.0245
##              parity5 or more live births ns(age, 3)1:parity5 or more live births
##                                    0.1078                                  0.7742
## ns(age, 3)2:parity5 or more live births ns(age, 3)3:parity5 or more live births
##                                    0.0471                                 -0.2718
```

```r
## GEE under unstructed condition
gee_unstructured <- suppressMessages(gee(wt ~ ns(age, 3) * parity, data = d2, id = id, corstr = "unstru
```

```
##                               (Intercept)                             ns(age, 3)1
##                                    5.2563                                  6.3615
##                               ns(age, 3)2                             ns(age, 3)3
##                                   13.1691                                  8.0245
##              parity5 or more live births ns(age, 3)1:parity5 or more live births
##                                    0.1078                                  0.7742
## ns(age, 3)2:parity5 or more live births ns(age, 3)3:parity5 or more live births
##                                    0.0471                                 -0.2718
```

Based on the results from the GEE models, we observe that the "robust" standard error estimates are smaller than the "naive" (or model-based) estimates. Also, when comparing the standard error estimates from the GEE approaches to the GLS, the robust standard errors obtained from GEE are smaller than the estimates from the GLS approach.

```r
## Table to compare coefficients for different fitting strategies
coefficient_output = cbind(gee_independence$coefficients, gee_exchangeable$coefficients, gee_unstructure
names(coefficient_output) = c("coeff_gee_independence", "coeff_gee_exchangeable", "coeff_gee_unstructure
rownames(coefficient_output) = c("intercept", "ns1", "ns2", "ns3", "parity", "ns1:parity", "ns2:parity"
print(coefficient_output)
```

```
##              coeff_gee_independence coeff_gee_exchangeable coeff_gee_unstructured
## intercept                    5.2563                  5.488                 5.2712
## ns1                          6.3615                  5.939                 6.1207
## ns2                         13.1691                 12.703                13.2048
## ns3                          8.0245                  7.816                 7.9855
## parity                       0.1078                  0.115                 0.1572
## ns1:parity                   0.7742                  0.295                 0.2899
## ns2:parity                   0.0471                  0.467                 0.3427
## ns3:parity                  -0.2718                  0.201                -0.0846
##              coeff_gls_full
## intercept             4.332
## ns1                   6.057
## ns2                  15.497
## ns3                   8.593
## parity               -0.109
## ns1:parity            0.592
## ns2:parity            0.820
## ns3:parity            0.108
```

```r
## Table to compare naive SE for different fitting strategies
naive_se_output = cbind(sqrt(diag(gee_independence$naive.variance)), sqrt(diag(gee_exchangeable$naive.va
```

```r
names(naive_se_output) = c("naive_se_gee_independence", "naive_se_gee_exchangeable", "naive_se_gee_unst
rownames(naive_se_output) = c("intercept", "ns1", "ns2", "ns3", "parity", "ns1_parity", "ns2_parity", "n
print(naive_se_output)
```

```
##           naive_se_gee_independence naive_se_gee_exchangeable
## intercept                     0.276                     0.192
## ns1                           0.266                     0.202
## ns2                           0.678                     0.364
## ns3                           0.318                     0.235
## parity                        0.477                     0.330
## ns1_parity                    0.412                     0.312
## ns2_parity                    1.130                     0.604
## ns3_parity                    0.451                     0.339
##           naive_se_gee_unstructured naive_se_gls_full
## intercept                     0.246            0.0844
## ns1                           0.294            0.3380
## ns2                           0.528            0.4124
## ns3                           0.332            0.5223
## parity                        0.450            0.1395
## ns1_parity                    0.463            0.5078
## ns2_parity                    0.925            0.6098
## ns3_parity                    0.485            0.7310
```

```r
## Table to compute robust SE for different fitting strategies
robust_se_output = cbind(sqrt(diag(gee_independence$robust.variance)), sqrt(diag(gee_exchangeable$robus
names(robust_se_output) = c("robust_gee_independence", "robust_gee_exchangeable", "robust_gee_unstructu
rownames(robust_se_output) = c("intercept", "ns1", "ns2", "ns3", "parity", "ns1_parity", "ns2_parity",
print(robust_se_output)
```

```
##           robust_gee_independence robust_gee_exchangeable
## intercept                   0.269                   0.207
## ns1                         0.514                   0.196
## ns2                         0.794                   0.458
## ns3                         0.716                   0.296
## parity                      0.521                   0.394
## ns1_parity                  0.715                   0.373
## ns2_parity                  1.302                   0.827
## ns3_parity                  0.931                   0.386
##           robust_gee_unstructured naive_se_gls_full
## intercept                   0.202            0.0844
## ns1                         0.246            0.3380
## ns2                         0.505            0.4124
## ns3                         0.402            0.5223
## parity                      0.347            0.1395
## ns1_parity                  0.358            0.5078
## ns2_parity                  0.808            0.6098
## ns3_parity                  0.494            0.7310
```

Part IV: Linear mixed model

1. Fit two linear mixed models.

Fit the following two models using the lme function in R.

```
d2$age_6 = d2$age - 6
d2$age_24_plus = ifelse(d2$age>24,d2$age-24,0)
d2$age_48_plus = ifelse(d2$age>48,d2$age-48,0)
## Random Intercept Model:
lme_intercept = lme(wt ~ 1 + age_6 + age_24_plus +
                      age_48_plus + parity +
                      parity:age_6 + parity:age_24_plus +
                      parity:age_48_plus, random = ~1|id, data = d2)
summary(lme_intercept)
```

```
## Linear mixed-effects model fit by REML
##  Data: d2
##    AIC  BIC logLik
##   1789 1837   -884
##
## Random effects:
##  Formula: ~1 | id
##         (Intercept) Residual
## StdDev:        1.34    0.423
##
## Fixed effects: wt ~ 1 + age_6 + age_24_plus + age_48_plus + parity + parity:age_6 +      parity:age_2
##                                           Value Std.Error  DF t-value p-value
## (Intercept)                                6.41    0.1625 674    39.5  0.0000
## age_6                                      0.16    0.0066 674    24.5  0.0000
## age_24_plus                               -0.03    0.0097 674    -3.1  0.0022
## age_48_plus                               -0.02    0.0096 674    -2.2  0.0257
## parity5 or more live births                0.17    0.2757 195     0.6  0.5360
## age_6:parity5 or more live births          0.01    0.0112 674     0.6  0.5633
## age_24_plus:parity5 or more live births    0.00    0.0154 674    -0.1  0.9464
## age_48_plus:parity5 or more live births    0.00    0.0136 674    -0.3  0.7487
##  Correlation:
##                                          (Intr) age_6  ag_24_ ag_48_ p5omlb
## age_6                                    -0.516
## age_24_plus                               0.205 -0.810
## age_48_plus                               0.106  0.137 -0.530
## parity5 or more live births              -0.590  0.304 -0.121 -0.062
## age_6:parity5 or more live births         0.305 -0.590  0.478 -0.081 -0.568
## age_24_plus:parity5 or more live births -0.130  0.512 -0.632  0.335  0.262
## age_48_plus:parity5 or more live births -0.074 -0.097  0.373 -0.705  0.126
##                                          a_6omlb a_2omlb
## age_6
## age_24_plus
## age_48_plus
## parity5 or more live births
## age_6:parity5 or more live births
## age_24_plus:parity5 or more live births -0.832
## age_48_plus:parity5 or more live births  0.119  -0.506
```

```
##
## Standardized Within-Group Residuals:
##    Min     Q1    Med     Q3    Max
## -3.237 -0.477  0.018  0.523  4.084
##
## Number of Observations: 877
## Number of Groups: 197
```

```r
## Random Intercept and Random Slope for Age Model
lme_intercept_slope = lme(wt ~ 1 + age_6 + age_24_plus +
                            age_48_plus + parity + parity:age_6 +
                            parity:age_24_plus +
      parity:age_48_plus,
      random = ~1 + age_6|id, data = d2)
summary(lme_intercept_slope)
```

```
## Linear mixed-effects model fit by REML
##  Data: d2
##    AIC  BIC logLik
##   1731 1789   -854
##
## Random effects:
##  Formula: ~1 + age_6 | id
##  Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev Corr
## (Intercept) 1.1153 (Intr)
## age_6       0.0312 -0.269
## Residual    0.3793
##
## Fixed effects: wt ~ 1 + age_6 + age_24_plus + age_48_plus + parity + parity:age_6 +      parity:age_2
##                                      Value Std.Error  DF t-value p-value
## (Intercept)                           6.38    0.1543 674    41.3  0.0000
## age_6                                 0.16    0.0078 674    21.0  0.0000
## age_24_plus                          -0.03    0.0111 674    -2.7  0.0073
## age_48_plus                          -0.02    0.0104 674    -2.2  0.0254
## parity5 or more live births           0.22    0.2732 195     0.8  0.4251
## age_6:parity5 or more live births     0.01    0.0136 674     0.4  0.6728
## age_24_plus:parity5 or more live births  0.00    0.0180 674    -0.1  0.9060
## age_48_plus:parity5 or more live births  0.00    0.0148 674     0.0  0.9756
##  Correlation:
##                                      (Intr) age_6  ag_24_ ag_48_ p5omlb
## age_6                                -0.647
## age_24_plus                           0.381 -0.806
## age_48_plus                           0.122  0.136 -0.460
## parity5 or more live births          -0.565  0.365 -0.215 -0.069
## age_6:parity5 or more live births     0.370 -0.572  0.461 -0.078 -0.704
## age_24_plus:parity5 or more live births -0.235  0.497 -0.617  0.284  0.456
## age_48_plus:parity5 or more live births -0.085 -0.095  0.323 -0.702  0.141
##                                      a_6omlb a_2omlb
## age_6
## age_24_plus
## age_48_plus
## parity5 or more live births
## age_6:parity5 or more live births
## age_24_plus:parity5 or more live births -0.838
```

```
## age_48_plus:parity5 or more live births  0.121  -0.440
##
## Standardized Within-Group Residuals:
##     Min      Q1     Med      Q3     Max
## -3.1007 -0.4652  0.0255  0.5149  2.6854
##
## Number of Observations: 877
## Number of Groups: 197
```

2. Model interpretation: Random intercept model

Interpret the random intercept variance from the Random Intercept Model.

The intercept β_0 in this model is the population average weight of 6-month old children born to mothers with parity 0 to 3; $\beta_0 + \beta_4$ represents the population average weight of 6-month old children born to mothers with parity > 3. The random intercept variance represents an estimate of the variance in 6-month old children's expected weights.

The estimated random intercept standard deviation is 1.34 and the estimated population average weight for 6-month olds born to mothers with parity 0 to 3 and > 3 is 6.41 and 6.57 kg, respectively. Therefore, we expect that 95% of 6-month old children born to mothers with parity 0 to 3 and > 3 will have weights within [3.784, 9.036] and [3.764,9.016] kg, respectively.

3. Model Interpretation: Random intercept and slope for age model

Interpret the random intercept and random slope for age variance from the Random Intercept and Random Slope for Age Model.

The random slope variance is an estimate of variation in growth rates compared to the population average growth rate. The random slope for age standard deviation is 0.03. The population average monthly growth rate in the first 24 months of life is 0.16 kg per month for children born to mothers with parity 0 to 3 and 0.17 kg per month for children born to mothers with parity > 3. So we expect that the growth rate in the first 24 months of life will vary from child to child and estimate that 95% of children born to mothers with parity 0 to 3 or > 3 will have growth rates between [0.101,0.219] or [0.111,0.229]

4. Estimates of variance in weights

Using the fit of the Random Intercept and Random Slope for Age Model, estimate the variance in weights of children at 6-months, 12-months and 36-months, from the parity le 3 group.

To compute the variance in weights of children at different ages, take variance on both sides of the mixed linear models, we have

$$\begin{aligned} Var(Y_{ij}|age_{ij}, parity_i = 0) &= var(u_{0i} + u_{1i}(age_{ij} - 6) + \epsilon_{ij}) \\ &= var(u_{0i}) + (age_{ij} - 6)^2 var(u_{1i}) + 2 \times (age_{ij} - 6) \times cov(u_{0i}, u_{1i}) + var(\epsilon_{ij}) \end{aligned}$$

Plug in the estimated variance for u_{0i} ($\hat{\tau}_0^2 = 1.11^2$) and u_{1i} ($\hat{\tau}_1^2 = 0.031^2$), covariance for u_{0i} and u_{1i} ($\hat{\tau}_{01} = -0.27 \times 1.11 \times 0.031$) and variance for ϵ_{ij} ($\hat{\sigma}^2 = 0.38^2$).

For childeren at 6-months, the variance in weights of children is $1.11^2 + 0.031^2 * (6-6)^2 + 2 * (6-6) * (-0.27) * 1.11 * 0.031 + 0.38^2 = 1.377$.

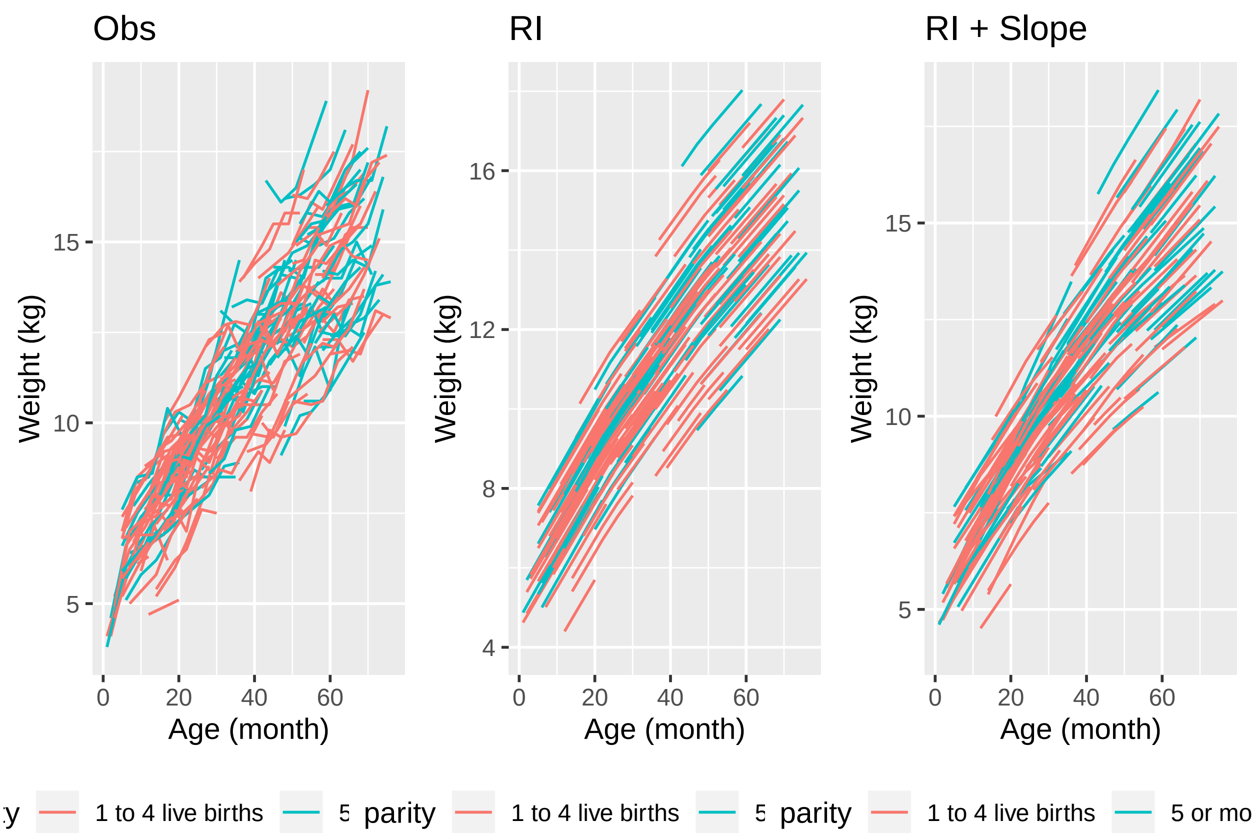
For childeren at 12-months, the variance in weights of children is $1.11^2 + 0.031^2 * (12-6)^2 + 2 * (12-6) * (-0.27) * 1.11 * 0.031 + 0.38^2 = 1.3$.

For childeren at 36-months, the variance in weights of children is $1.11^2 + 0.031^2 * (36-6)^2 + 2 * (36-6) * (-0.27) * 1.11 * 0.031 + 0.38^2 = 1.684$.

5. Observed vs. fitted

Obtain the fitted values from the model and make a plot with 3 panels comparing the observed data (spaghetti plot) and the fitted values from both the Random Intercept and Random Intercept and Random Slope for Age Models (i.e. spaghetti plots of fitted values). Comment on which of the two linear mixed models you think is most consistent with the data.

```r
## Fitted values from the random intercept model
d2$predict_int = lme_intercept$fitted[,2]
d2$predict_intslope = lme_intercept_slope$fitted[,2]

plot.int = ggplot(d2) + geom_line(aes(x = age, y = predict_int,group = id,col=parity)) +
  xlab("Age (month)") +
  ylab("Weight (kg)") +
  theme(legend.position='bottom', legend.box='horizontal') +
  ggtitle("RI")

## Fitted values from the Random Intercept and Random Slope for Age Model
plot.intslope = ggplot(d2) + geom_line(aes(x = age, y = predict_intslope,group = id,col=parity)) +
  xlab("Age (month)") +
  ylab("Weight (kg)") +
  theme(legend.position='bottom', legend.box='horizontal') +
  ggtitle("RI + Slope")

plot.obs = ggplot(d2) + geom_line(aes(x = age, y = wt, col = parity,group = id)) +
  xlab("Age (month)") +
  ylab("Weight (kg)") +
  theme(legend.position='bottom', legend.box='horizontal') +
  ggtitle("Obs")

grid.arrange(plot.obs, plot.int, plot.intslope, nrow = 1)
```



When comparing the fit of the two random effects models, the Random Intercept and Random Slope for Age Model is more consistent with the data. Note that the Random Intercept and Random Slope for Age Model does a better job mimicking the features of the data; that is, variation in growth rates. Whereas, the Random Intercept model assumes that the rate of growth may vary by parity group but across children from the same parity group, the growth rates are the same.

Part V: Summarize your findings

Objectives:

Using data generated from a longitudinal study of Nepalese children aged 1 to 60 months of age, we determined if the growth rates of children differed by mother's parity (number of previous live births) and estimated the population variation in weights of 6-month old Nepalese children and the population variation in annual growth rates of children.

Methods:

We fit a weighted least squares model where the mean weight was modeled as a smooth function of age (natural cubic spline with 3 degrees of freedom) separately for children of mothers with parity 0 to 3 vs. greater than 3, the variance in weights was allowed to vary as a function of age (linear on the log variance scale) and the correlation between weights from the same child followed a first order autoregressive model. To estimate the heterogeneity of children's weights at 6-months of age and heterogeneity in annual growth rates, we fit a random effects model that allowed the mean weight to change as a linear spline model for age (with knots at 24 and 48 months), separately for each parity group with a random intercept and random slope for age (linear term) defined for each child.

Results

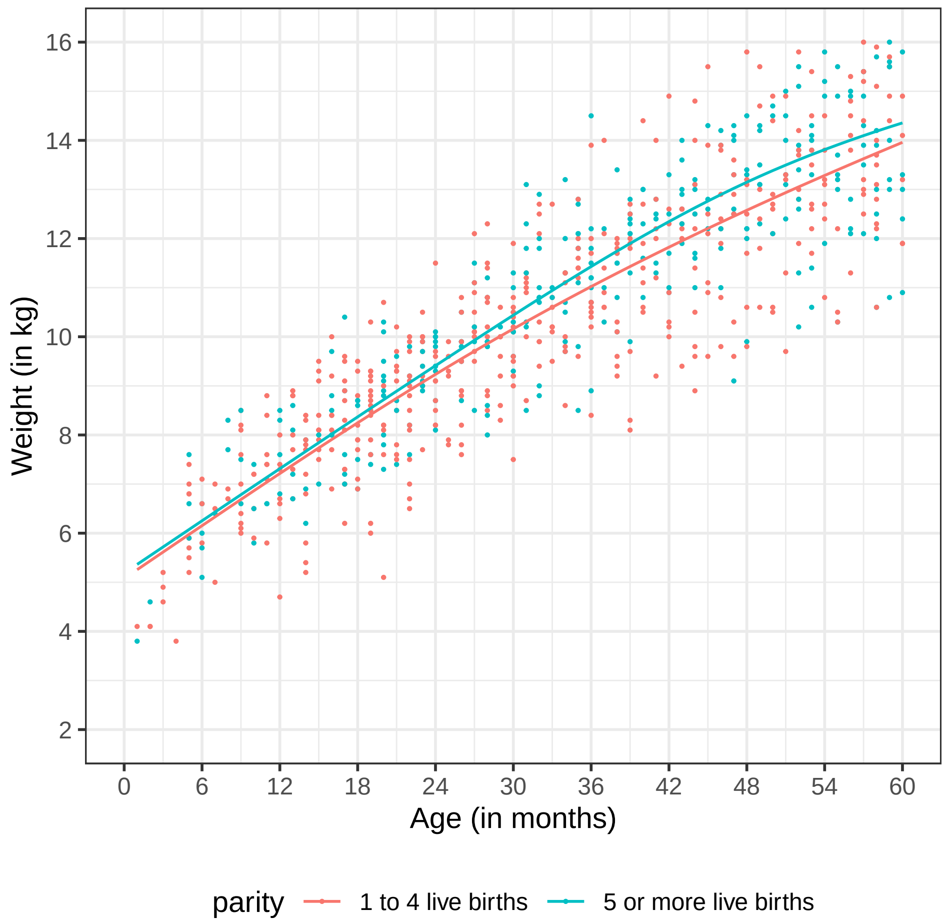


Figure 1 displays the observed and estimated mean weight as a function of age, separately for each parity group based on the weighted least squares model. We found no evidence in the data to suggest that the growth rate of weight differed by parity group (likelihood ratio test p-value 0.11).

From the random effects model, we estimated the population average weight of 6-month old Nepali children is 6.42 kg and 6.48 kg among children born to mothers with parity 0 to 3 and > 3, respectively. The random effects model estimated the standard deviation in expected weights for 6-month olds (SD = 1.34 kg) such that we anticipate that 95% of 6-month old Nepali children will have weights ranging from [3.794, 9.046] and [3.854, 9.106] among children born to mothers with parity 0 to 3 and > 3, respectively.

We estimated that annual growth rates will vary from child to child (estimated variance in annual growth rates is 0.139 and standard deviation 0.373).

For example, among children born to mothers with parity 0 to 3, we estimate that the annual growth rate for ages 0 to 24 months, 24 to 48 months and > 48 months is 1.92 kg, 1.56 kg and 1.2 kg, respectively, and that 95% of children's annual growth rates will fall within [1.189, 2.651],[0.829, 2.291] and [0.469, 1.931].

Among children born to mothers with parity larger than 3, we estimate that the annual growth rate for ages 0 to 24 months, 24 to 48 months and > 48 months is 2.04 kg, 1.56 kg and 1.2 kg, respectively, and that 95% of children's annual growth rates will fall within [1.309, 2.771],[0.829, 2.291] and [0.469, 1.931].

Summary

Based on the data from the longitudinal study, there is no evidence that the growth rates of children differ by mother's parity. There is variation in weights among 6-month old Nepali children as well as annual growth rates.