Initial Modeling Report

Giang (Jessie) Le, Leon Zhou, Calvin Ochi

**I - Introduction**

Life expectancy at birth is defined by the United Nations Human Development Report as “the years a newborn infant would live if prevailing patterns of age-specific mortality rates at the time of birth were to stay the same throughout the infant’s life” (Human Development Report, 2019). Often life expectancy is used as an indicator of national development and quality of life for the average citizen. Since the 1990’s, life expectancy has increased globally with a drastic increase during the Millennium Development Goals (MDGs) era of 2000-2015 (WHO, 2015). Despite this, high-income countries tend to have higher average life expectancies than lower-income countries (WHO, 2015). Learning more about what factors are related to the difference in life expectancy in developing versus developed countries, continents, and regions may help inform policy decision-making and resource allocation to narrow this gap. For this project, we are interested in exploring what are some potential health related factors and socioeconomic factors that are most related to one country’s life expectancy.

Previous studies showed that some important socioeconomic determinants of life expectancy include Gross Domestic Product (GDP) per capita, percent expenditure on health, and education, represented by years of schooling (Lin et al 2012 and Kabir 2008). Moreover, life expectancy can be attributed to health factors including alcohol consumption, percent immunization, BMI, and HIV/AIDS prevalence (Mondal and Shinta 2014 and Khan et al 2016). In our project, we will integrate these socioeconomic and health related factors into our models in order to explore their relationships with life expectancy.

**Baseline models:**

First of all, we created an indicator variable for developing countries (developing=1). Then, based on EDA, we found that Africa has a lower average life expectancy than other continents. Thus, we were especially interested in the coefficient of Africa for all the models. The default used Africa as the reference level, so we releveled Oceania continent as the reference for continent indicator.

We explored the effect of country status and continents as baseline regional models. First, we began with an unconditional model (Table 1- Model 1) with no predictors. The estimated intercept is 69.2, which is the average life expectancy across different countries and across different years.

Then, we added country status as a level two predictor (Table 1- Model 2). The model estimated the coefficient of country status to be -12.1, meaning that on average, a developing country has life expectancy 12.1 years less than a developed country. We also ran anova 𝜒2 test between the unconditional model (Table 1- Model 1) and the country status model (Table 1 - Model 2). The test statistics was 52.9 and corresponding p-value is smaller than 0.05, indicating that country status is a significant predictor for life expectancy.

We added the continent indicator as the only level two predictors (Table 1 - Model 3). Life expectancy of an African country is 12.6 years shorter than that of an Oceanian country, on average. We ran an anova 𝜒2 test between the continent model (Table 1 - Model 3) with the unconditional model (Table 1 - Model 1). The 𝜒2 test statistic is 177 and p-value is smaller than 0.05, indicating continent indicator should be included in our model. Then we included both country status and continent as our level two predictors (Table 1 - Model 4) and compared this composite model with the unconditional model. We obtained a 𝜒2 test statistic 197.08 with p-value smaller than 0.05, indicating both terms are necessary for the model.

We also modeled life expectancy using longitudinal model approach with country status and continent as level two predictors and year as the only level one predictor. After testing different interaction terms with years using ANOVA 𝜒2 test, we found that an interaction term between continent and year is necessary to include in the model (Table 1 - Model 5).

**Health models:**

When building models for life expectancy using only health related variables, we compared each model with the unconditional model. Then, we added level one predictors log(HIV), log(diphtheria), log(polio), BMI, log(Hepatitis B), and Measles cases per 1,000 population one at a time. Each time we added a new predictor, we used ANOVA 𝜒2 test to see if the additional predictor would improve the model, and we only included the predictor with significant 𝜒2 test statistics. We included both continent and country status as level two predictors. In addition, for each of the level one predictors, we tested if there is any interaction with continent or country status or both. For testing the significance of the interaction terms, we used ANOVA 𝜒2 test as well. After all, we found that adding Measles does not improve the model performance, and we obtained an 𝜒2 test statistics of 12.86 with p-value of 0.117 when comparing the model with Measles and the model without. Thus, we excluded Measles from our model. Almost all the level one predictors have significant interactions with both continent and country status, except that BMI does not have significant interaction with continent, and alcohol does not have a significant interaction with country status.

In addition, we also explored using longitudinal approach (Table 3 -Model 2) including year as the only level one predictor and log(HIV), log(diphtheria), log(polio), BMI, log(Hepatitis B) as level two predictors. However, when comparing BIC of this model with other multilevel models, the model using year as the only level one predictor did not perform as well as other models.

**Socioeconomics models:**

Based on EDA, percent expenditure on health and income composition of resources are both highly correlated with GDP (cor>0.9). In addition, based on the variable summary, the percent expenditure on health is calculated as a percentage of GDP. Thus, we decided to use GDP as a predictor and dropped percent expenditure on health and income composition of resources.

We began by first fitting an unconditional model. Then we added log(GDP) and schooling as level one predictors one at a time. Each time we added a new predictor, we used ANOVA 𝜒2 test to compare the model with one additional term to the prior model, and we only included the predictor if we obtained a significant p-value. In addition, we included country status and continent as level two predictors. We also tested different combinations of interaction terms of continent and country status with GDP and schooling. We added those interaction terms one at a time and compared with the model that does not have any interaction term using ANOVA 𝜒2 test. After all, only an interaction between continent and school is necessary to include in our model. We also compared the final model with the unconditional model using ANOVA 𝜒2 test. We obtained a 𝜒2 statistic 1245.2 with p-value smaller than 0.05 indicating an overall improvement by adding those predictors (Table 4 - Model 1).

After that, we explored using longitudinal model (Table 4 - Model 2) including year as the only level one predictor and continent, country status, log(GDP), schooling, and income composition of resource as level two predictors. For those level two predictors, we used their values in 2000 as baseline values. When comparing BIC for the two models in Table 4, the model 2 yielded a lower BIC.

**Longitudinal combined-factor model:**

Next, we examined how combination of healthcare factors and socioeconomic factors could improve model performance. We considered observations in 2000 as baseline variables, which could be applied as level-two covariates. We tested a longitudinal model, with only centered years at level one. Level-two covariates for the intercept include: continents, country status, baseline alcohol consumption, BMI, baseline percent immunization of Hepatitis B, Polio, and Diphtheria, baseline death rate among HIV live births, baseline log(GDP), and average years of schooling. Additionally, the level-two covariate for the slope is Africa indicator because there is an evident difference in life expectancy rate between African countries and the others. By observing t-values, BMI, baseline percent immunization of Hepatitis B, Polio, and Diphtheria might not be significant variables (t-values < 2). After testing the significance of each variable by using 𝜒2 test, we removed baseline percent immunization of Hepatitis B, Polio, Diphtheria, BMI and all interaction terms with BMI. This might be because there is a moderate correlation between BMI and years of schooling (cor=0.560). For confirmation, the adjusted model Table 2 - Model 1) shows similar performance with the larger model, based on ANOVA 𝜒2 test. Also, we found that baseline alcohol consumption is not significant.

Then, we explored multilevel model containing various numerical level-one covariates including centered years, log(GDP), log(HIV deaths), alcohol consumption, log(Hepatitis B), log(Polio), log(Diphtheria), and years of schooling, as well as interaction between baseline variables with the annual rate of life expectancy. Significance of variables was tested with ANOVA 𝜒2 test by removing variables one by one. We found that the following variables did not significantly improve the model performance: alcohol consumption, log(Hepatitis B), log(Polio), and log(Diphtheria). Next, we tested to see whether the interactions between the annual rate of life expectancy and baseline log(GDP) and baseline log(HIV deaths) are significant. Comparing with model that includes both baseline interactions and continuous variables log(GDP) and log(HIV deaths) (Table 2 - Model 2), keeping continuous variables log(GDP) and log(HIV) at level one improves model performance in explaining variability. On the other hand, the interactions between years and baseline variables were found to be insignificant to model performance.

**Tentative final model:**

We finalized our tentative model (Table 2 - Model 4) to include years, log(GDP), log(HIV deaths), and years of schooling at level one. Because those variables change significantly over the years, it is not reasonable to apply one value for each variable at level two. Level-two variables are continent and country status (Equation 1).

Among non-African countries, the annual increase of life expectancy is 0.21 years, after adjusting for continent, status, GDP, death rate among HIV live births, and years of schooling. On the other hand, for African countries, the annual increase of life expectancy is 0.10 years, after adjusting for GDP, death rate of HIV live birth, and years of schooling. On average, a developing country has 3 years less in life expectancy than a developed country, after adjusting for continent, status, GDP, death rate among HIV live births, and years of schooling. After controlling for continent, status, GDP, years, and years of schooling, every increase in deaths per 1,000 HIV live births is associated with a decrease in life expectancy by a factor of 6.88. A one-year increase in years of schooling is associated with a 0.40-year increase in life expectancy, after adjusting for continent, GDP, and death rate among HIV live births.

**III - Appendix**

**Table 1. Table of Comparison Between Baseline Models.**

| Predictors | Coefficients | | | | |
| --- | --- | --- | --- | --- | --- |
| Model 1  (Unconditional model) | Model 2  (Country status) | Model 3  (Continent) | Model 4  (Composite Multilevel) | Model 5  (Longitudinal) |
| Intercept | 69.224 | 79.2 |  | 76.43 | 74.98 |
| Developing |  | -12.08 |  | -6.52 | -6.27 |
| Continent: Oceania (ref.) Africa Asia Europe Americas |  |  | - -12.6 -0.02 6.22 2.27 | - -11.30 0.87 3.17 3.38 | - 3.30 -14.42 0.006 3.21 |
| Year |  |  |  |  | 0.17 |
| Africa x Year |  |  |  |  | 0.41 |
| Asia x Year |  |  |  |  | 0.11 |
| Europe x Year |  |  |  |  | 0.15 |
| Americas x Year |  |  |  |  | 0.02 |
| BIC | 8194.69 | 15019.8 | 14912.99 | 14899.13 | 12746.03 |

**Table 2. Table of Comparison Between Various Combined-Factor Models.**

| Predictors | Longitudinal Multilevel model | | | |
| --- | --- | --- | --- | --- |
| Model 1:  Longitudinal | Model 2: Multilevel with both baseline and continuous | Model 3: Multilevel with baseline interactions | Model 4: Multilevel without baseline interaction |
| Intercept | 46.8 | 61.8 | 64.0 | 61.7 |
| Years (centered) | 0.26 | -0.052 | -0.10 | 0.21 |
| Status Developed (ref.) Developing | - -1.07 | - -2.86 | - -2.29 | - -3.05 |
| Continent Oceania (ref.) Africa Asia Americas Europe | - -4.00 1.56 4.51 2.64 | - 6.13 3.82 6.16 4.61 | - -5.23 3.16 3.73 5.69 | - 6.16 3.54 6.32 4.50 |
| Baseline log(GDP) | 0.48 |  |  |  |
| Baseline log(HIV deaths) | -1.97 |  |  |  |
| Mean years of schooling | 1.11 |  |  |  |
| Years of Schooling |  | 0.40 | 0.44 | 0.40 |
| Log(GDP) |  | -0.066 |  | -0.073 |
| Log(HIV deaths) |  | -1.80 |  | -1.93 |
| African х Years (centered) | 0.36 | -0.117 | -0.20 | -0.104 |
| Baseline log(GDP) х Years (centered) |  | 0.036 | 0.060 |  |
| Baseline log(HIV) х Years (centered) |  | 0.0021 | 0.053 |  |
| BIC | 3740.782 | 3742.006 | 3775.514 | 3721.293 |

**Table 3. Table of Comparison Between Health Models.**

| Predictors | Coefficients | |
| --- | --- | --- |
| Model 1  (Multilevel) | Model 2  (Longitudinal with each predictor at baseline level) |
| Intercept | 62.79 | 46.17 |
| Log HIV | 0.98 | -2.91 |
| Log Diph | 2.09 | 1.25 |
| Log Polio | 1.12 | 2.25 |
| BMI | -0.0008 | 0.08 |
| Log HepB | 0.05 | 0.10 |
| Alcohol | 0.13 |  |
| Log HIV x Africa | -2.85 |  |
| Log HIV x Americas | 0.26 |  |
| Log HIV x Asia | 0.95 |  |
| Log HIV x Europe | -3.28 |  |
| Log HIV x Developing | -2.26 |  |
| Log Diph x Africa | -0.19 |  |
| Log Diph x Americas | 0.19 |  |
| Log Diph x Asia | 0.06 |  |
| Log Diph x Europe | -0.51 |  |
| Log Diph x Developing | -1.72 |  |
| Log Polio x Africa | -0.58 |  |
| Log Polio x Americas | -0.28 |  |
| Log Polio x Asia | -0.38 |  |
| Log Polio x Europe | -0.41 |  |
| Log Polio x Developing | -0.71 |  |
| Developing x BMI | 0.01 |  |
| Log HepB x Africa | 0.14 |  |
| Log HepB x Americas | 0.13 |  |
| Log HepB x Asia | 1.12 |  |
| Log HepB x Europe | 0.70 |  |
| Lof HepB x Developing | -0.15 |  |
| Alcohol x Africa | -0.22 |  |
| Alcohol x Americas | 0.59 |  |
| Alcohol x Asia | -0.16 |  |
| Alcohol x Europe | -0.34 |  |
| BIC | 10213.53 | 3876.286 |

**Table 4. Table of Comparison Between Socioeconomic Models.**

| Predictors | Coefficients | |
| --- | --- | --- |
| Model 1  (Multilevel) | Model 2  (Longitudinal with baseline variables ) |
| Intercept | 44.73 | 58.49 |
| Log GDP | 0.07 | 0.85 |
| Developing |  | -0.29 |
| Schooling (5 years) | 9.77 | 3.44 |
| Schooling (5 years) x Africa | -1.59 |  |
| Schooling (5 years) x Americas | 1.27 |  |
| Schooling (5 years) x Asia | 0.94 |  |
| Schooling (5 years) x Europe | 0.61 |  |
| Africa |  | -5.66 |
| Americas |  | 2.53 |
| Asia |  | 2.22 |
| Europe |  | 5.54 |
| BIC | 7507.777 | 6937.419 |

**Equation 1**

**R Code:**

> library(mosaic)

> library(tidyverse)

> library(ggplot2)

> library(gridExtra)

> library(dplyr)

> library(stats)

> library(lme4)

> library(GGally)

> library(Hmisc)

> library(broom)

> library(MASS)

> library(mnormt)

library(readr)

library(dplyr)

library(mosaic)

library(ggplot2)

Life\_Analysis <- read\_csv("~/Stats 316 F19/Project/Leon-Jessie-Calvin/Leon-folder/Life-Analysis.csv")

life.df <- Life\_Analysis

#Preliminary linear models

life <- life.df %>%

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

mutate(cyear = Year - 2000,

log.GDP = log(GDP),

log.pct.exp = log(percentageexpenditure+0.5),

log.HepB = log(HepatitisB),

log.Polio = log(Polio),

log.Diph = log(Diphtheria),

log.HIV = log(`HIV/AIDS`),

Africa = ifelse(Continent == "Africa",1,0),

Americas = ifelse(Continent == "Americas",1,0),

Asia = ifelse(Continent == "Asia",1,0),

Oceania = ifelse(Continent == "Oceania",1,0),

Europe = ifelse(Continent == "Europe",1,0),

developing = (ifelse(Status == "Developing",1,0)))

life <- mutate\_if(life, is.character, as.factor)

life$Continent = relevel(life$Continent, ref="Oceania")

life0 <- na.omit(life)

summary(life)

#Multi-level models

##initial model

#unconditional model

model.0 <- lmer(Lifeexpectancy~ 1 + (1|Country), REML=T, data=life0)

summary(model.0)

##Baseline(regional) multi-level model

###Multi-level

#status as level 2

model.1 <- lmer(Lifeexpectancy~ developing+ (1|Country), REML=T, data=life)

summary(model.1)

anova(model.1, model.0)

#status and continent as level 2

model.2 <- lmer(Lifeexpectancy~ developing + Continent + (1|Country), REML=T, data=life)

summary(model.2)

anova(model.2, model.1)

anova(model.2, model.0)

model.3 <- lmer(Lifeexpectancy~ Continent + (1|Country), REML=T, data=life)

summary(model.3)

anova(model.3, model.0)

###Longitudinal

model.3a <- lmer(Lifeexpectancy~ developing + Continent + cyear + (1 + cyear|Country), REML=T, data=life)

summary(model.3a)

model.3b <- lmer(Lifeexpectancy~ developing + Continent + Year + developing:Year + (1 + Year|Country), REML=T, data=life)

summary(model.3b)

anova(model.3b, model.3a)

model.3c <- lmer(Lifeexpectancy~ developing + Oceania + cyear + Continent:cyear + (1 + cyear|Country), REML=T, data=life)

summary(model.3c)

anova(model.3c, model.3a)

ggplot(data=life, aes(x=Year, y=Lifeexpectancy, col=Status)) +

geom\_point() +

geom\_smooth()

ggplot(data=life, aes(x=Year, y=Lifeexpectancy, col=Continent)) +

geom\_point() +

geom\_smooth()

model.3d <- lmer(Lifeexpectancy~ developing + Continent + Year + Continent:Year + developing:Year +

(1 + Year + developing|Country), REML=T, data=life)

summary(model.3d)

anova(model.3d, model.3c)

##Healthcare multi-level model

healthcare factors:

###Multi-level

Hmodel.1 <- lmer(Lifeexpectancy~ log.HIV + (1 + log.HIV|Country), REML=T, data=life)

summary(Hmodel.1)

Hmodel.1a <- lmer(Lifeexpectancy~ log.HIV + Continent + (1+log.HIV|Country), REML=T, data=life)

summary(Hmodel.1a)

anova(Hmodel.1a, Hmodel.1)

Hmodel.1b <- lmer(Lifeexpectancy~ log.HIV + Continent + developing + (1+log.HIV|Country), REML=T, data=life)

summary(Hmodel.1b)

anova(Hmodel.1b, Hmodel.1a)

Hmodel.1c <- lmer(Lifeexpectancy~ log.HIV + Continent + developing + Continent:log.HIV + (1 + log.HIV|Country), REML=T, data=life)

summary(Hmodel.1c)

anova(Hmodel.1c, Hmodel.1b)

Hmodel.1d <- lmer(Lifeexpectancy~ log.HIV + Continent + developing + Continent:log.HIV + developing\*log.HIV + (1 + log.HIV|Country), REML=T, data=life)

summary(Hmodel.1d)

anova(Hmodel.1d, Hmodel.1c)

add log.Diph

Hmodel.2 <- lmer(Lifeexpectancy~ log.HIV + log.Diph + (1 + log.HIV + log.Diph|Country), REML=T, data=life)

summary(Hmodel.2)

##with log values

h.comp1 <- lmer(Lifeexpectancy ~ log.HIV + log.Diph + log.Polio + log.HepB + (1 + log.HIV + log.Diph + log.Polio + log.HepB|Country), REML=T, data=life )

summary(h.comp1)

h.comp2 <- lmer(Lifeexpectancy ~ BMI + Alcohol + I(Measles/1000) + (1 + BMI + Alcohol + I(Measles/1000)|Country), REML=T, data=life )

summary(h.comp2)

try composite model:

log.HIV + log.Diph + log.Polio + log.HepB + BMI + Alcohol + I(Measles/1000)

##overall composit

h.comp.all0 <- lmer(Lifeexpectancy ~ log.HIV + (1 + log.Diph|Country), REML=T, data=life )

summary(h.comp.all0)

h.comp.all1 <- lmer(Lifeexpectancy ~ log.HIV + log.Diph+ (1 + log.Diph+log.Diph|Country), REML=T, data=life )

anova(h.comp.all1, h.comp.all0)

h.comp.all2 <- lmer(Lifeexpectancy ~ log.HIV + log.Diph + log.Polio +(1 + log.Diph+log.Diph+log.Polio|Country), REML=T, data=life )

anova(h.comp.all2, h.comp.all1)

h.comp.all3 <- lmer(Lifeexpectancy ~ log.HIV + log.Diph + log.Polio + BMI +(1 + log.Diph+log.Diph+log.Polio+ BMI|Country), REML=T, data=life )

anova(h.comp.all3, h.comp.all2)

h.comp.all4 <- lmer(Lifeexpectancy ~ log.HIV + log.Diph + log.Polio + BMI + log.HepB+(1 + log.Diph+log.Diph+log.Polio+ BMI+log.HepB|Country), REML=T, data=life )

anova(h.comp.all4, h.comp.all3)

h.comp.all5 <- lmer(Lifeexpectancy ~ log.HIV + log.Diph + log.Polio + BMI + log.HepB+Alcohol+(1 + log.Diph+log.Diph+log.Polio+ BMI+log.HepB+Alcohol|Country), REML=T, data=life )

anova(h.comp.all5, h.comp.all4)

h.comp.all6 <- lmer(Lifeexpectancy ~ log.HIV + log.Diph + log.Polio + BMI + log.HepB+Alcohol+I(Measles/1000)+(1 + log.Diph+log.Diph+log.Polio+ BMI+log.HepB+Alcohol+I(Measles/1000)|Country), REML=T, data=life )

anova(h.comp.all6, h.comp.all5)

h.comp.all7 <- lmer(Lifeexpectancy ~ log.Diph + log.Polio + BMI + log.HepB+Alcohol+I(Measles/1000)+(1 + log.Diph+log.Diph+log.Polio+ BMI+log.HepB+Alcohol+I(Measles/1000)|Country), REML=T, data=life )

anova(h.comp.all6, h.comp.all7)

I(Measles/1000)

final model:

health.all <- lmer(Lifeexpectancy ~ log.HIV + log.Diph + log.Polio + BMI + log.HepB+Alcohol+(1 + log.Diph+log.Diph+log.Polio+ BMI+log.HepB+Alcohol|Country), REML=T, data=life )

summary(health.all)

#Test interactions

health.all0 <- lmer(Lifeexpectancy ~ log.HIV + Continent:log.Diph + log.Polio + BMI + log.HepB+Alcohol+(1 + log.Diph+log.Diph+log.Polio+ BMI+log.HepB+Alcohol|Country), REML=T, data=life )

anova(health.all0, health.all)

health.all1 <- lmer(Lifeexpectancy ~ log.HIV + Continent:log.Diph + developing:log.Diph + log.Polio + BMI + log.HepB+Alcohol+(1 + log.Diph+log.Diph+log.Polio+ BMI+log.HepB+Alcohol|Country), REML=T, data=life )

anova(health.all1, health.all0)

health.all2 <- lmer(Lifeexpectancy ~ log.HIV + log.Diph + Continent:log.Polio + BMI + log.HepB+Alcohol+(1 + log.Diph+log.Diph+log.Polio+ BMI+log.HepB+Alcohol|Country), REML=T, data=life )

anova(health.all2, health.all)

health.all2a <- lmer(Lifeexpectancy ~ log.HIV + log.Diph + Continent:log.Polio + developing:log.Polio + BMI + log.HepB+Alcohol+(1 + log.Diph+log.Diph+log.Polio+ BMI+log.HepB+Alcohol|Country), REML=T, data=life )

anova(health.all2a, health.all2)

health.all3 <- lmer(Lifeexpectancy ~ log.HIV + log.Diph + log.Polio + developing:BMI + log.HepB+Alcohol+(1 + log.Diph+log.Diph+log.Polio+ BMI+log.HepB+Alcohol|Country), REML=T, data=life )

anova(health.all3, health.all)

health.all4 <- lmer(Lifeexpectancy ~ log.HIV + log.Diph + log.Polio + BMI + Continent:log.HepB + Alcohol+(1 + log.Diph+log.Diph+log.Polio+ BMI+log.HepB+Alcohol|Country), REML=T, data=life )

anova(health.all4, health.all)

health.all4a <- lmer(Lifeexpectancy ~ log.HIV + log.Diph + log.Polio + BMI + log.HepB+ developing:log.HepB + Alcohol+(1 + log.Diph+log.Diph+log.Polio+ BMI+log.HepB+Alcohol|Country), REML=T, data=life )

anova(health.all4a, health.all)

health.all5 <- lmer(Lifeexpectancy ~ log.HIV + log.Diph + log.Polio + BMI + log.HepB+ log.HepB + developing:Alcohol+(1 + log.Diph+log.Diph+log.Polio+ BMI+log.HepB+Alcohol|Country), REML=T, data=life )

anova(health.all5, health.all)

final.health <- lmer(Lifeexpectancy ~ log.HIV + log.HIV:Continent + log.HIV:developing +

log.Diph + log.Diph:Continent + log.Diph:developing +

log.Polio + log.Polio:Continent + log.Polio:developing +

BMI + BMI:developing +

log.HepB + log.HepB:Continent + log.HepB:developing +

Alcohol + Alcohol:Continent +

(1 + log.Diph+log.Diph+log.Polio+ BMI+log.HepB+Alcohol|Country), REML=T, data=life)

summary(final.health)

###Longitudinal

baseline.health <- life %>% filter(Year == 2000) %>%

select(Country, log.HIV, Alcohol, Continent, developing, log.HepB, BMI, log.Polio, log.Diph) %>%

rename(

base\_alcohol = Alcohol,

base\_HIV = log.HIV,

base\_HepB = log.HepB,

base\_BMI = BMI,

base\_Polio = log.Polio,

base\_Diph = log.Diph)

merging\_data <- life %>%

right\_join(baseline.health, by = "Country") %>%

na.omit()

#lon-spag1.eps

ggplot(life,aes(x=Year,y=Lifeexpectancy)) +

geom\_line(aes(group=Country),color="dark grey") +

geom\_smooth(aes(group=1),color="black",size=1) +

labs(x="Years",y="LE")

#lon-spag1.eps

ggplot(life,aes(x=Year,y=Lifeexpectancy)) +

geom\_line(aes(group=Country),color="dark grey") +

facet\_grid(~Status) +

geom\_smooth(aes(group=1),color="black",size=1) +

labs(x="Years",y="LE")

#lon-spag1.eps

ggplot(life,aes(x=Year,y=Lifeexpectancy)) +

geom\_line(aes(group=Country),color="dark grey") +

facet\_grid(~Continent) +

geom\_smooth(aes(group=1),color="black",size=1) +

labs(x="Years",y="LE")

Vars

log.HIV + log.Diph + log.Polio + log.HepB + BMI + Alcohol + I(Measles/1000)

Test if contant:

HIV

#lon-spag1.eps

ggplot(life,aes(x=Year,y=log.HIV)) +

geom\_line(aes(group=Country),color="dark grey") +

geom\_smooth(aes(group=1),color="black",size=1) +

labs(x="Years")

#lon-spag1.eps

ggplot(life,aes(x=Year,y=log.HIV)) +

geom\_line(aes(group=Country),color="dark grey") +

facet\_grid(~Continent) +

geom\_smooth(aes(group=1),color="black",size=1) +

labs(x="Years")

Diph

#lon-spag1.eps

ggplot(life,aes(x=Year,y=log.Diph)) +

geom\_line(aes(group=Country),color="dark grey") +

geom\_smooth(aes(group=1),color="black",size=1) +

labs(x="Years")

Polio

#lon-spag1.eps

ggplot(life,aes(x=Year,y=log.Polio)) +

geom\_line(aes(group=Country),color="dark grey") +

geom\_smooth(aes(group=1),color="black",size=1) +

labs(x="Years")

HepB

#lon-spag1.eps

ggplot(life,aes(x=Year,y=log.HepB)) +

geom\_line(aes(group=Country),color="dark grey") +

geom\_smooth(aes(group=1),color="black",size=1) +

labs(x="Years")

BMI

#lon-spag1.eps

ggplot(life,aes(x=Year,y=BMI)) +

geom\_line(aes(group=Country),color="dark grey") +

geom\_smooth(aes(group=1),color="black",size=1) +

labs(x="Years")

Alcohol

#lon-spag1.eps

ggplot(life,aes(x=Year,y=Alcohol)) +

geom\_line(aes(group=Country),color="dark grey") +

geom\_smooth(aes(group=1),color="black",size=1) +

labs(x="Years")

Measles

#lon-spag1.eps

ggplot(life,aes(x=Year,y=Measles/1000)) +

geom\_line(aes(group=Country),color="dark grey") +

geom\_smooth(aes(group=1),color="black",size=1) +

labs(x="Years")

health.long.full <- lmer(Lifeexpectancy ~ base\_HIV + base\_Diph + base\_Polio + base\_HepB + base\_BMI + cyear + (1 + cyear |Country), REML=T,data=merging\_data)

summary(health.long.full)

##Socioeconomics multi-level models

###Multi-level

life0 <- na.omit(life)

Smodel.a <- lmer(Lifeexpectancy ~ log.GDP + (1 + log.GDP|Country), REML=T, data=life0)

Smodel.b <- lmer(Lifeexpectancy ~ log.GDP + log.pct.exp + (1 + log.GDP+log.pct.exp|Country), REML=T, data=life0)

anova(Smodel.b,Smodel.a)

Smodel.c <- lmer(Lifeexpectancy ~ log.GDP + log.pct.exp + Schooling+(1 + log.GDP + log.pct.exp + Schooling|Country), REML=T, data=life0)

anova(Smodel.c,Smodel.b)

Smodel.d <- lmer(Lifeexpectancy ~ log.GDP + log.pct.exp + Schooling+Incomecompositionofresources+(1 + log.GDP + log.pct.exp + Schooling+Incomecompositionofresources|Country), REML=T, data=life0)

anova(Smodel.d,Smodel.c)

social.1 <- lmer(Lifeexpectancy ~ log.GDP + log.pct.exp + Schooling+Incomecompositionofresources+(1 + log.GDP + log.pct.exp + Schooling+Incomecompositionofresources|Country), REML=T, data=life0)

summary(social.1)

social.2 <- lmer(Lifeexpectancy ~ log.GDP + log.pct.exp + Schooling +Incomecompositionofresources + Incomecompositionofresources:developing+(1 + log.GDP + log.pct.exp + Schooling+Incomecompositionofresources|Country), REML=T, data=life0)

anova(social.2, social.1)

social.final <- lmer(Lifeexpectancy ~ log.GDP + I(Schooling/5) + I(Schooling/5):Continent + (1 + log.GDP + Schooling|Country), REML=T, data=life0)

summary(social.final)

anova(social.final, model.0)

###Longitudinal

baseline.social <- life %>% filter(Year == 2000) %>%

select(Country,log.GDP, Continent, Schooling) %>%

rename(

base\_GDP = log.GDP,

base\_Schooling = Schooling)

merging\_data\_social <- life %>%

right\_join(baseline.social, by = "Country") %>%

na.omit()

#lon-spag1.eps4

ggplot(life,aes(x=Year,y=log.pct.exp)) +

geom\_line(aes(group=Country),color="dark grey") +

geom\_smooth(aes(group=1),color="black",size=1) +

labs(x="Years")

#lon-spag1.eps

ggplot(life,aes(x=Year,y=log.pct.exp)) +

geom\_line(aes(group=Country),color="dark grey") +

facet\_grid(~Continent) +

geom\_smooth(aes(group=1),color="black",size=1) +

labs(x="Years")

#lon-spag1.eps

ggplot(life,aes(x=Year,y=Incomecompositionofresources)) +

geom\_line(aes(group=Country),color="dark grey") +

facet\_grid(~Status) +

geom\_smooth(aes(group=1),color="black",size=1) +

labs(x="Years")

social.final.long <- lmer(Lifeexpectancy ~ Continent.x + developing + base\_GDP + I(base\_Schooling/5)+ (1 + cyear|Country),REML=T, data= merging\_data\_social)

summary(social.final.long)

ggplot(data=life, aes(x=Year, y=Lifeexpectancy, col=Continent)) +

geom\_point() +

geom\_smooth()

##final composite model

ggplot(life,aes(x=Year,y=Lifeexpectancy)) +

geom\_line(aes(group=Country),color="dark grey") +

geom\_smooth(aes(group=1),color="black",size=1) +

labs(x="Years")

ggplot(life,aes(x=Year,y=Lifeexpectancy)) +

geom\_line(aes(group=Country),color="dark grey") +

facet\_grid(~Continent) +

geom\_smooth(aes(group=1),color="black",size=1) +

labs(x="Years")

> life <- read\_csv("~/Stats 316 F19/Project/Leon-Jessie-Calvin/Leon-folder/Life-Analysis.csv")

Parsed with column specification:

cols(

.default = col\_double(),

Country = col\_character(),

Status = col\_character(),

Continent = col\_character()

)

See spec(...) for full column specifications.

> life <- life %>%

+ filter(!is.na(Lifeexpectancy) & !is.na(Alcohol)) %>%

+ mutate(cyear = Year - 2000,

+ log.GDP = log(GDP),

+ log.pct.exp = log(percentageexpenditure+0.5),

+ log.HepB = log(HepatitisB),

+ log.Polio = log(Polio),

+ log.Diph = log(Diphtheria),

+ log.HIV = log(`HIV/AIDS`),

+ africa = ifelse(Continent == "Africa",1,0),

+ asia = ifelse(Continent == "Asia",1,0),

+ europe = ifelse(Continent == "Europe",1,0),

+ america = ifelse(Continent == "Americas",1,0),

+ oceania = ifelse(Continent == "Oceania",1,0),

+ african2 = ifelse(africa==1, "African","Non-African"),

+ developing = (ifelse(Status == "Developing",1,0)))

**Linear Regression Model (Assuming independent observations)**

> linear\_model <- lm(Lifeexpectancy ~ cyear + african + developing + log.GDP + log.pct.exp + cyear:african + cyear:developing, data= life)

Error in eval(predvars, data, env) : object 'cyear' not found

> summary(linear\_model)

Call:

lm(formula = Lifeexpectancy ~ cyear + african + developing + log.GDP + log.pct.exp +   
 cyear:african + cyear:developing, data = life)

Residuals:

Min 1Q Median 3Q Max

-34.135 -3.268 -0.064 3.092 21.304

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 66.00416 0.85766 76.959 < 2e-16 \*\*\*

cyear 0.27893 0.06064 4.600 4.46e-06 \*\*\*

african -14.09999 0.49715 -28.362 < 2e-16 \*\*\*

developing -3.50098 0.60653 -5.772 8.87e-09 \*\*\*

log.GDP 0.54383 0.16252 3.346 0.000833 \*\*\*

log.pct.exp 0.90541 0.14673 6.171 8.01e-10 \*\*\*

cyear:african 0.34510 0.05998 5.754 9.87e-09 \*\*\*

cyear:developing -0.16538 0.07054 -2.344 0.019139 \*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.412 on 2320 degrees of freedom

(407 observations deleted due to missingness)

Multiple R-squared: 0.6896, Adjusted R-squared: 0.6887

F-statistic: 736.5 on 7 and 2320 DF, p-value: < 2.2e-16

### Large BIC value indicates that linear regression model may not be a good fit.

> BIC(linear\_model)

[1] 14530.66

**# Unconditional means model**

> life.0 <- lmer(Lifeexpectancy ~ 1 + (1|Country), REML=T, data = life)

> summary(life.0)

Linear mixed model fit by REML ['lmerMod']

Formula: Lifeexpectancy ~ 1 + (1 | Country)

Data: life

REML criterion at convergence: 15042.6

Scaled residuals:

Min 1Q Median 3Q Max

-8.8062 -0.5154 -0.0394 0.4191 6.1195

Random effects:

Groups Name Variance Std.Dev.

Country (Intercept) 83.941 9.162

Residual 7.193 2.682

Number of obs: 2928, groups: Country, 183

Fixed effects:

Estimate Std. Error t value

(Intercept) 69.2249 0.6791 101.9

**#Unconditional slope model**

> life.00 <- lmer(Lifeexpectancy ~ cyear + (cyear|Country), REML=T, data=life)

summary(life.00)

Linear mixed model fit by REML ['lmerMod']

Formula: Lifeexpectancy ~ cyear + (cyear | Country)

Data: life

REML criterion at convergence: 11932.7

Scaled residuals:

Min 1Q Median 3Q Max

-14.4798 -0.3608 -0.0463 0.1180 5.0236

Random effects:

Groups Name Variance Std.Dev. Corr

Country (Intercept) 108.47434 10.4151

cyear 0.08158 0.2856 -0.66

Residual 2.67033 1.6341

Number of obs: 2735, groups: Country, 182

Fixed effects:

Estimate Std. Error t value

(Intercept) 66.64898 0.77431 86.08

cyear 0.35713 0.02237 15.97

Correlation of Fixed Effects:

(Intr)

cyear -0.646

convergence code: 0

Model failed to converge with max|grad| = 0.0123045 (tol = 0.002, component 1)

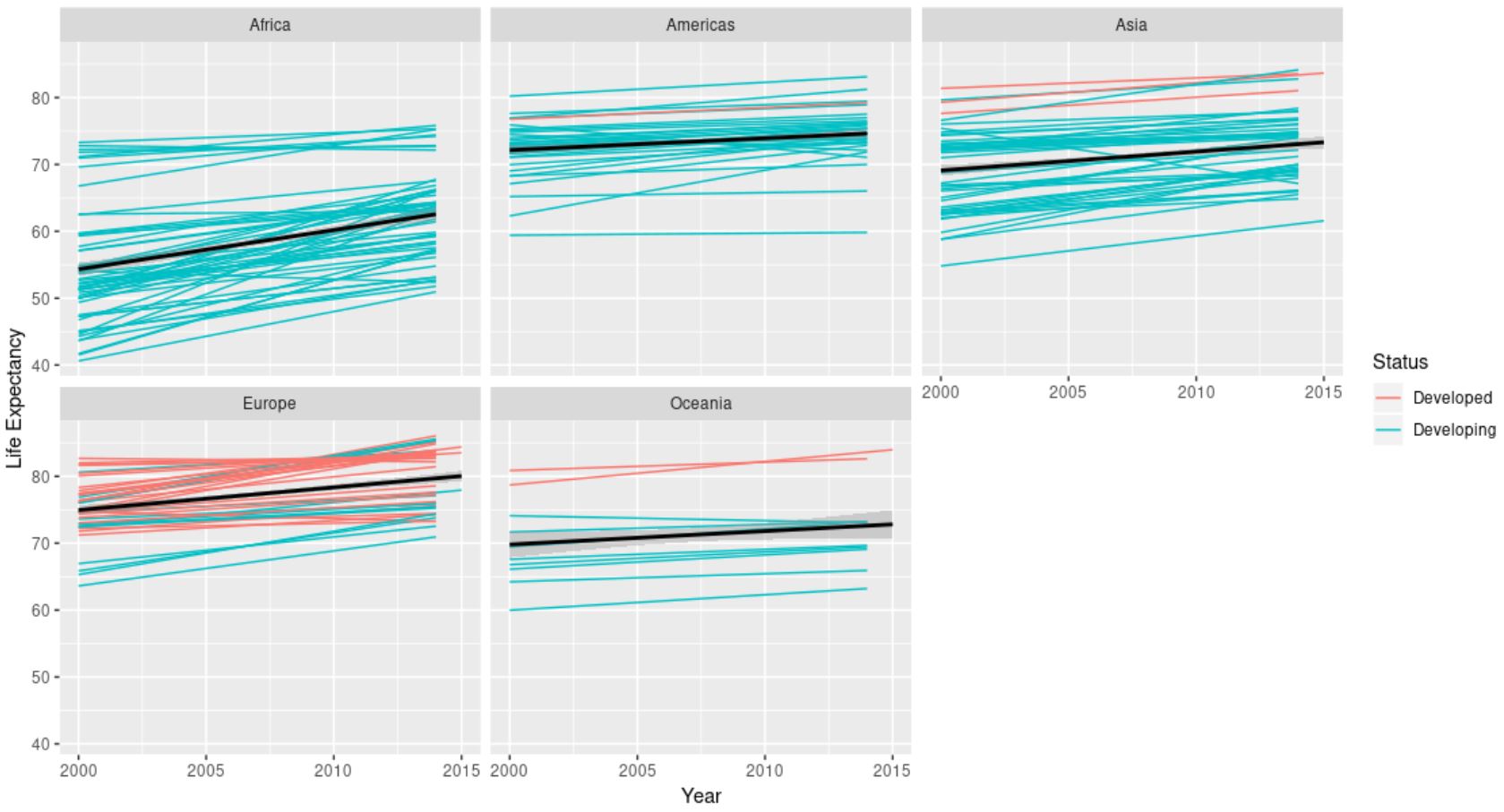
**Different life expectancy trends among different continents**

>ggplot(data=life, aes(x=Year,y=Lifeexpectancy, color = Status)) +

geom\_smooth(aes(group=Country), na.rm = T, method="lm",se=F, size=0.5, alpha=0.05) + facet\_wrap(~Continent,ncol=3) +

geom\_smooth(aes(group=1), method = "lm", color = "black") +

labs(y="Life Expectancy")



>baseline <- life %>% filter(Year == 2000) %>%

select(Country, percentageexpenditure, log.pct.exp, log.HIV, log.GDP, Alcohol, HepatitisB, Polio, Diphtheria) %>%

rename(base\_pctexp = percentageexpenditure,

base\_log.pct.exp = log.pct.exp,

base\_alcohol = Alcohol,

base\_HIV = log.HIV,

base\_GDP = log.GDP,

base\_HepB = HepatitisB,

base\_Polio = Polio,

base\_Diph = Diphtheria)

>mean.schooling <- life %>%

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

group\_by(Country) %>%

summarise(mean.schooling = mean(Schooling))

> merging\_data <- life %>%

right\_join(baseline, by = "Country") %>%

right\_join(mean.schooling, by = "Country") %>%

na.omit()

**#Longitudinal Combined-Factor Model using baseline variables at level two**

>life.2a <- lmer(Lifeexpectancy ~ africa + asia + europe + america + developing + base\_alcohol + BMI + base\_HIV + base\_HepB + base\_Polio + base\_Diph + base\_GDP + base\_log.pct.exp + mean.schooling + cyear + african:cyear + developing:cyear + (cyear|Country),REML=T, data=merging\_data)

>summary(life.2a)

Linear mixed model fit by REML ['lmerMod']

Formula:

Lifeexpectancy ~ africa + asia + europe + america + developing +

base\_alcohol + BMI + base\_HIV + base\_HepB + base\_Polio +

base\_Diph + base\_GDP + base\_log.pct.exp + mean.schooling +

cyear + african:cyear + developing:cyear + (cyear | Country)

Data: merging\_data

REML criterion at convergence: 3660.9

Scaled residuals:

Min 1Q Median 3Q Max

-2.4379 -0.3631 -0.0367 0.1284 5.6680

Random effects:

Groups Name Variance Std.Dev. Corr

Country (Intercept) 19.2494 4.3874

cyear 0.0569 0.2385 -0.53

Residual 2.0626 1.4362

Number of obs: 910, groups: Country, 61

Fixed effects:

Estimate Std. Error t value

(Intercept) 49.1867889 6.0092034 8.185

africa -5.0010739 2.6680721 -1.874

asia 0.5023752 1.7831220 0.282

europe 0.7820582 2.7216857 0.287

america 3.5044858 2.0996877 1.669

developing -0.6976970 2.1843029 -0.319

base\_alcohol 0.0025333 0.2535587 0.010

BMI -0.0005608 0.0035105 -0.160

base\_HIV -1.8688083 0.4341843 -4.304

base\_HepB -0.0241893 0.0205936 -1.175

base\_Polio 0.0252058 0.0337550 0.747

base\_Diph 0.0333454 0.0231911 1.438

base\_GDP -0.6229040 0.7758717 -0.803

base\_log.pct.exp 1.0434837 0.6571094 1.588

mean.schooling 1.1169167 0.3967624 2.815

cyear 0.3177993 0.0703112 4.520

cyear:african 0.3798621 0.0982494 3.866

developing:cyear -0.0752417 0.0810480 -0.928

**> # Remove BMI because BMI is correlated to Schooling**

**> # Remove log.pct.exp because it is correlated to log.GDP**

> life.2b <- lmer(Lifeexpectancy ~ africa + asia + europe + america + developing + base\_alcohol + base\_HIV + base\_HepB + base\_Polio + base\_Diph + base\_GDP + mean.schooling + cyear + african:cyear + (cyear|Country),REML=T, data=merging\_data)

> summary(life.2b)

Linear mixed model fit by REML ['lmerMod']

Formula:

Lifeexpectancy ~ africa + asia + europe + america + developing +

base\_alcohol + base\_HIV + base\_HepB + base\_Polio + base\_Diph +

base\_GDP + mean.schooling + cyear + african:cyear + (cyear |

Country)

Data: merging\_data

REML criterion at convergence: 3652.6

Scaled residuals:

Min 1Q Median 3Q Max

-2.4012 -0.3594 -0.0364 0.1257 5.6877

Random effects:

Groups Name Variance Std.Dev. Corr

Country (Intercept) 19.74340 4.4434

cyear 0.05682 0.2384 -0.52

Residual 2.05965 1.4351

Number of obs: 910, groups: Country, 61

Fixed effects:

Estimate Std. Error t value

(Intercept) 44.95854 5.26227 8.544

africa -5.05699 2.69630 -1.876

asia 0.56679 1.81158 0.313

europe 0.51602 2.76132 0.187

america 3.71599 2.13064 1.744

developing -1.47808 2.08029 -0.711

base\_alcohol 0.05275 0.25583 0.206

base\_HIV -2.03451 0.42815 -4.752

base\_HepB -0.01258 0.01957 -0.643

base\_Polio 0.01540 0.03373 0.456

base\_Diph 0.03528 0.02355 1.498

base\_GDP 0.43251 0.40749 1.061

mean.schooling 1.22008 0.39747 3.070

cyear 0.26094 0.03490 7.476

cyear:african 0.36125 0.09613 3.758

Correlation matrix not shown by default, as p = 15 > 12.

Use print(x, correlation=TRUE) or

vcov(x) if you need it

convergence code: 0

Model failed to converge with max|grad| = 0.0704962 (tol = 0.002, component 1)

> anova(life.2b, life.2a)

refitting model(s) with ML (instead of REML)

Data: merging\_data

Models:

life.2b: Lifeexpectancy ~ africa + asia + europe + america + developing +

life.2b: base\_alcohol + base\_HIV + base\_HepB + base\_Polio + base\_Diph +

life.2b: base\_GDP + mean.schooling + cyear + african:cyear + (cyear |

life.2b: Country)

life.2a: Lifeexpectancy ~ africa + asia + europe + america + developing +

life.2a: base\_alcohol + BMI + base\_HIV + base\_HepB + base\_Polio +

life.2a: base\_Diph + base\_GDP + base\_log.pct.exp + mean.schooling +

life.2a: cyear + african:cyear + developing:cyear + (cyear | Country)

Df AIC BIC logLik deviance Chisq Chi Df Pr(>Chisq)

life.2b 19 3676.2 3767.7 -1819.1 3638.2

life.2a 22 3678.1 3784.0 -1817.0 3634.1 4.1243 3 0.2483

> ## The large model is not significantly better than the smaller one.

> ## Removing terms one by one and testing anova.

## Correlation between immunization variables

### Remove base\_Diph

>life.2c <- lmer(Lifeexpectancy ~ africa + asia + europe + america + developing + base\_alcohol + base\_HIV + base\_HepB + base\_Polio + base\_GDP + mean.schooling + cyear + african:cyear + (cyear|Country),REML=T, data=merging\_data)

>anova(life.2c, life.2b)

Data: merging\_data

Models:

life.2c: Lifeexpectancy ~ africa + asia + europe + america + developing +

life.2c: base\_alcohol + base\_HIV + base\_HepB + base\_Polio + base\_GDP +

life.2c: mean.schooling + cyear + african:cyear + (cyear | Country)

life.2b: Lifeexpectancy ~ africa + asia + europe + america + developing +

life.2b: base\_alcohol + base\_HIV + base\_HepB + base\_Polio + base\_Diph +

life.2b: base\_GDP + mean.schooling + cyear + african:cyear + (cyear |

life.2b: Country)

Df AIC BIC logLik deviance Chisq Chi Df Pr(>Chisq)

life.2c 18 3677.0 3763.6 -1820.5 3641.0

life.2b 19 3676.2 3767.7 -1819.1 3638.2 2.7522 1 0.09712 .

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

### Remove base\_Polio

>life.2d <- lmer(Lifeexpectancy ~ africa + asia + europe + america + developing + base\_alcohol + base\_HIV + base\_HepB + base\_GDP + mean.schooling + cyear + african:cyear + (cyear|Country),REML=T, data=merging\_data)

>anova(life.2d, life.2c)

Data: merging\_data

Models:

life.2d: Lifeexpectancy ~ africa + asia + europe + america + developing +

life.2d: base\_alcohol + base\_HIV + base\_HepB + base\_GDP + mean.schooling +

life.2d: cyear + african:cyear + (cyear | Country)

life.2c: Lifeexpectancy ~ africa + asia + europe + america + developing +

life.2c: base\_alcohol + base\_HIV + base\_HepB + base\_Polio + base\_GDP +

life.2c: mean.schooling + cyear + african:cyear + (cyear | Country)

Df AIC BIC logLik deviance Chisq Chi Df Pr(>Chisq)

life.2d 17 3676.7 3758.6 -1821.4 3642.7

life.2c 18 3677.0 3763.6 -1820.5 3641.0 1.7649 1 0.184

### Remove base\_HepB

>life.2e <- lmer(Lifeexpectancy ~ africa + asia + europe + america + developing + base\_alcohol + base\_HIV + base\_GDP + mean.schooling + cyear + african:cyear + (cyear|Country),REML=T, data=merging\_data)

>anova(life.2e, life.2d)

Data: merging\_data

Models:

life.2e: Lifeexpectancy ~ africa + asia + europe + america + developing +

life.2e: base\_alcohol + base\_HIV + base\_GDP + mean.schooling + cyear +

life.2e: african:cyear + (cyear | Country)

life.2d: Lifeexpectancy ~ africa + asia + europe + america + developing +

life.2d: base\_alcohol + base\_HIV + base\_HepB + base\_GDP + mean.schooling +

life.2d: cyear + african:cyear + (cyear | Country)

Df AIC BIC logLik deviance Chisq Chi Df Pr(>Chisq)

life.2e 16 3674.8 3751.8 -1821.4 3642.8

life.2d 17 3676.7 3758.6 -1821.4 3642.7 0.0736 1 0.7861

### Remove base\_alcohol

life.2f <- lmer(Lifeexpectancy ~ africa + asia + europe + america + developing + base\_HIV + base\_GDP + mean.schooling + cyear + african:cyear + (cyear|Country),REML=T, data=merging\_data)

>anova(life.2f, life.2e)

Data: merging\_data

Models:

life.2f: Lifeexpectancy ~ africa + asia + europe + america + developing +

life.2f: base\_HIV + base\_GDP + mean.schooling + cyear + african:cyear +

life.2f: (cyear | Country)

life.2e: Lifeexpectancy ~ africa + asia + europe + america + developing +

life.2e: base\_alcohol + base\_HIV + base\_GDP + mean.schooling + cyear +

life.2e: african:cyear + (cyear | Country)

Df AIC BIC logLik deviance Chisq Chi Df Pr(>Chisq)

life.2f 15 3672.9 3745.1 -1821.5 3642.9

life.2e 16 3674.8 3751.8 -1821.4 3642.8 0.1332 1 0.7152

>summary(life.2f)

Linear mixed model fit by REML ['lmerMod']

Formula:

Lifeexpectancy ~ africa + asia + europe + america + developing +

base\_HIV + base\_GDP + mean.schooling + cyear + african:cyear +

(cyear | Country)

Data: merging\_data

REML criterion at convergence: 3638.6

Scaled residuals:

Min 1Q Median 3Q Max

-2.3969 -0.3551 -0.0373 0.1270 5.6776

Random effects:

Groups Name Variance Std.Dev. Corr

Country (Intercept) 20.25546 4.5006

cyear 0.05674 0.2382 -0.55

Residual 2.06011 1.4353

Number of obs: 910, groups: Country, 61

Fixed effects:

Estimate Std. Error t value

(Intercept) 46.78616 4.88723 9.573

africa -4.00365 2.56778 -1.559

asia 1.55684 1.72064 0.905

europe 2.63846 2.38021 1.109

america 4.50988 1.94222 2.322

developing -1.06727 1.85039 -0.577

base\_HIV -1.97193 0.41351 -4.769

base\_GDP 0.48134 0.39000 1.234

mean.schooling 1.22241 0.35334 3.460

cyear 0.26089 0.03488 7.479

cyear:african 0.36130 0.09607 3.761

Correlation of Fixed Effects:

(Intr) africa asia europe americ dvlpng bs\_HIV bs\_GDP mn.sch cyear

africa -0.017

asia -0.125 0.499

europe 0.009 0.440 0.601

america 0.006 0.558 0.659 0.569

developing -0.538 -0.051 0.018 0.439 -0.056

base\_HIV -0.104 -0.480 -0.010 -0.136 -0.224 0.010

base\_GDP -0.292 -0.300 -0.091 -0.082 -0.166 0.181 0.210

mean.schlng -0.705 -0.090 -0.104 -0.405 -0.194 0.090 0.201 -0.262

cyear -0.070 0.135 0.003 0.001 0.000 0.000 0.000 0.002 -0.002

cyear:afrcn 0.025 -0.371 -0.001 0.000 0.000 0.000 0.000 -0.001 0.001 -0.363

life.3 <- lmer(Lifeexpectancy ~ africa + asia + europe + america + developing + cyear + african:cyear + base\_GDP:cyear + base\_HIV:cyear + log.GDP + log.HIV + log.HepB + log.Polio + log.Diph + Alcohol + Schooling + (cyear + log.GDP + log.HIV+ log.HepB + log.Polio + log.Diph + Alcohol + Schooling|Country), REML=T, data=merging\_data)

summary(life.3)

#Remove alcohol.

life.3a <- lmer(Lifeexpectancy ~ africa + asia + europe + america + developing + cyear + african:cyear + base\_GDP:cyear + base\_HIV:cyear + log.GDP + log.HIV + log.HepB + log.Polio + log.Diph + Schooling + (cyear + log.GDP + log.HIV+ log.HepB + log.Polio + log.Diph + Schooling|Country), REML=T, data=merging\_data)

summary(life.3a)

anova(life.3a, life.3)

# Remove immunization one by one.

life.3b <- lmer(Lifeexpectancy ~ africa + asia + europe + america + developing + cyear + african:cyear + base\_GDP:cyear + base\_HIV:cyear + log.GDP + log.HIV + log.HepB + log.Diph + Schooling + (cyear + log.GDP + log.HIV + log.HepB + log.Diph + Schooling|Country), REML=T, data=merging\_data)

anova(life.3b, life.3a)

## Polio is correlated to the other two.

life.3c <- lmer(Lifeexpectancy ~ africa + asia + europe + america + developing + cyear + african:cyear + base\_GDP:cyear + base\_HIV:cyear + log.GDP + log.HIV + log.HepB + Schooling + (cyear + log.GDP + log.HIV + log.HepB + Schooling|Country), REML=T, data=merging\_data)

anova(life.3c, life.3b)

life.3d <- lmer(Lifeexpectancy ~ africa + asia + europe + america + developing + cyear + african:cyear + base\_GDP:cyear + base\_HIV:cyear + log.GDP + log.HIV + Schooling + (cyear + log.GDP + log.HIV + Schooling|Country), REML=T, data=merging\_data)

anova(life.3d, life.3c)

# Only use baseline variables for GDP and HIV

>life.3e <- lmer(Lifeexpectancy ~ africa + asia + europe + america + developing + cyear + african:cyear + base\_GDP:cyear + base\_HIV:cyear + Schooling + (cyear + Schooling|Country), REML=T, data=merging\_data)

>anova(life.3e, life.3d)

## Anova indicates that including logGDP is significant

**# Tentative final model**

> life.4 <- lmer(Lifeexpectancy ~ africa + asia + europe + america + developing + cyear + african:cyear + log.GDP + log.HIV + Schooling + (cyear + log.GDP + log.HIV + Schooling|Country), REML=T, data=merging\_data)

> summary(life.4)

Linear mixed model fit by REML ['lmerMod']

Formula: Lifeexpectancy ~ africa + asia + europe + america + developing +

cyear + african:cyear + log.GDP + log.HIV + Schooling + (cyear + log.GDP + log.HIV + Schooling | Country)

Data: merging\_data

REML criterion at convergence: 3537.3

Scaled residuals:

Min 1Q Median 3Q Max

-2.6414 -0.3719 -0.0352 0.1536 5.2931

Random effects:

Groups Name Variance Std.Dev. Corr

Country (Intercept) 67.50387 8.2161

cyear 0.02869 0.1694 -0.16

log.GDP 0.03079 0.1755 0.45 0.07

log.HIV 4.33134 2.0812 0.18 -0.59 -0.06

Schooling 0.39575 0.6291 -0.93 -0.21 -0.49 0.04

Residual 1.76161 1.3273

Number of obs: 910, groups: Country, 61

Fixed effects:

Estimate Std. Error t value

(Intercept) 61.71699 2.83068 21.803

africa 6.16405 1.92341 3.205

asia 3.54094 1.41681 2.499

europe 4.50147 1.83839 2.449

america 6.31803 1.43891 4.391

developing -3.04860 1.71021 -1.783

cyear 0.20600 0.03152 6.536

log.GDP -0.07322 0.04444 -1.648

log.HIV -1.92890 0.37216 -5.183

Schooling 0.39790 0.15261 2.607

cyear:african -0.10385 0.07429 -1.398

Correlation of Fixed Effects:

(Intr) africa asia europe americ dvlpng cyear lg.GDP lg.HIV Schlng

africa -0.309

asia -0.356 0.561

europe -0.472 0.439 0.614

america -0.363 0.563 0.751 0.586

developing -0.654 0.012 0.036 0.491 0.015

cyear 0.202 0.153 0.060 0.158 0.086 -0.032

log.GDP -0.004 -0.004 -0.019 -0.005 -0.026 0.029 -0.056

log.HIV 0.263 -0.117 -0.044 -0.006 -0.109 -0.194 -0.251 -0.019

Schooling -0.601 -0.018 -0.083 -0.201 -0.064 0.030 -0.507 -0.133 0.068

cyear:afrcn -0.024 -0.605 -0.002 -0.029 -0.004 0.042 -0.284 0.013 -0.086 0.025

convergence code: 0

Model failed to converge with max|grad| = 0.106524 (tol = 0.002, component 1)

> BIC(life.4)

[1] 3721.293

**Note:**

**Use baseline levels of predictors (GDP, Income) do not use baseline predictors of response**

**If there is a variable in an interaction keep using (use africa with year)**

**Use BIC to compare models if they aren’t nested**

**Can ignore boundary warning if correlation values are reasonable**

**Write out at least 1 model formula, could be in report or appendix**

**Level 1 Model:**

Modeling using all significant level one variables. Further ANOVA testing needed to narrow down model even more.

**Model comparison criteria: BIC, ANOVA if nested, t for variable selection**