

Coursework 4: Gapminder and Multilevel Modelling

1. Fit simple linear regression, with 'lifeExp' as dependent variable and 'gdpPercapita' as predictor. Visualize the fitted line and summarize the model. Briefly interpret the results.

#R code

#Loaded the packages and the data set

> library(ggplot2)

> library(dplyr)

> library(lattice)

> library(moderndiver)

> library(skimr)

> library(lme4)

> library(foreign)

> library(gapminder)

> str(gapminder) #show the data str type

Fit simple linear regression

> model1 <- lm (gapminder \$ lifeExp ~ gapminder \$ gdpPercap)

> summary(model1)

> gapminder %>%

ggplot (aes (x = gdpPercap, y = lifeExp)) +

geom_point () + geom_smooth (method = "lm", se= FALSE)

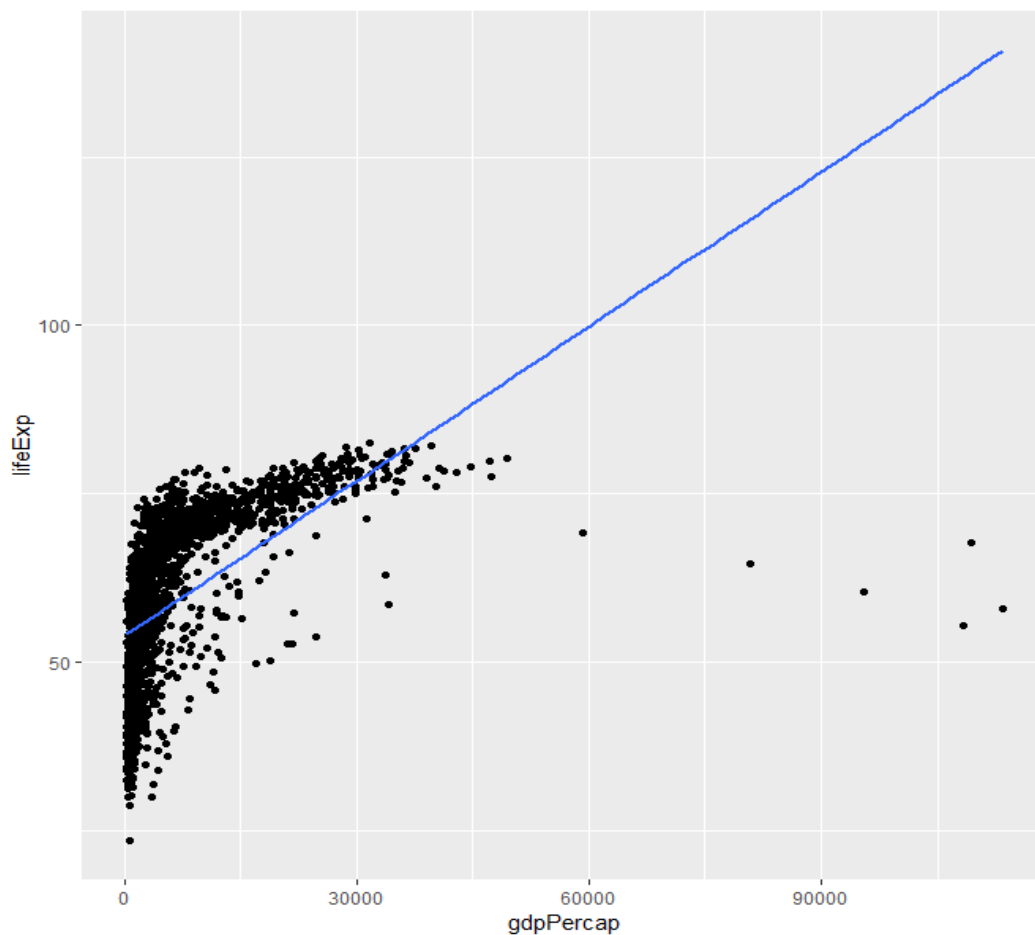
Output:

```
Call:
lm(formula = gapminder$lifeExp ~ gapminder$gdpPercap)

Residuals:
    Min       1Q   Median       3Q      Max
-82.754  -7.758   2.176   8.225  18.426

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    5.396e+01  3.150e-01  171.29  <2e-16 ***
gapminder$gdpPercap 7.649e-04  2.579e-05   29.66  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.49 on 1702 degrees of freedom
Multiple R-squared:  0.3407,    Adjusted R-squared:  0.3403
F-statistic: 879.6 on 1 and 1702 DF,  p-value: < 2.2e-16
```



Interpret: From the output, we can see that all coefficients are significant and there is a positive association between the gdpPercap and lifeExp.

2. Create a plot where you check if the fitted line would vary if continent information were to be included in the model. Should we include the nested structure of the data in to our model specification? Explain your answer.

#R code

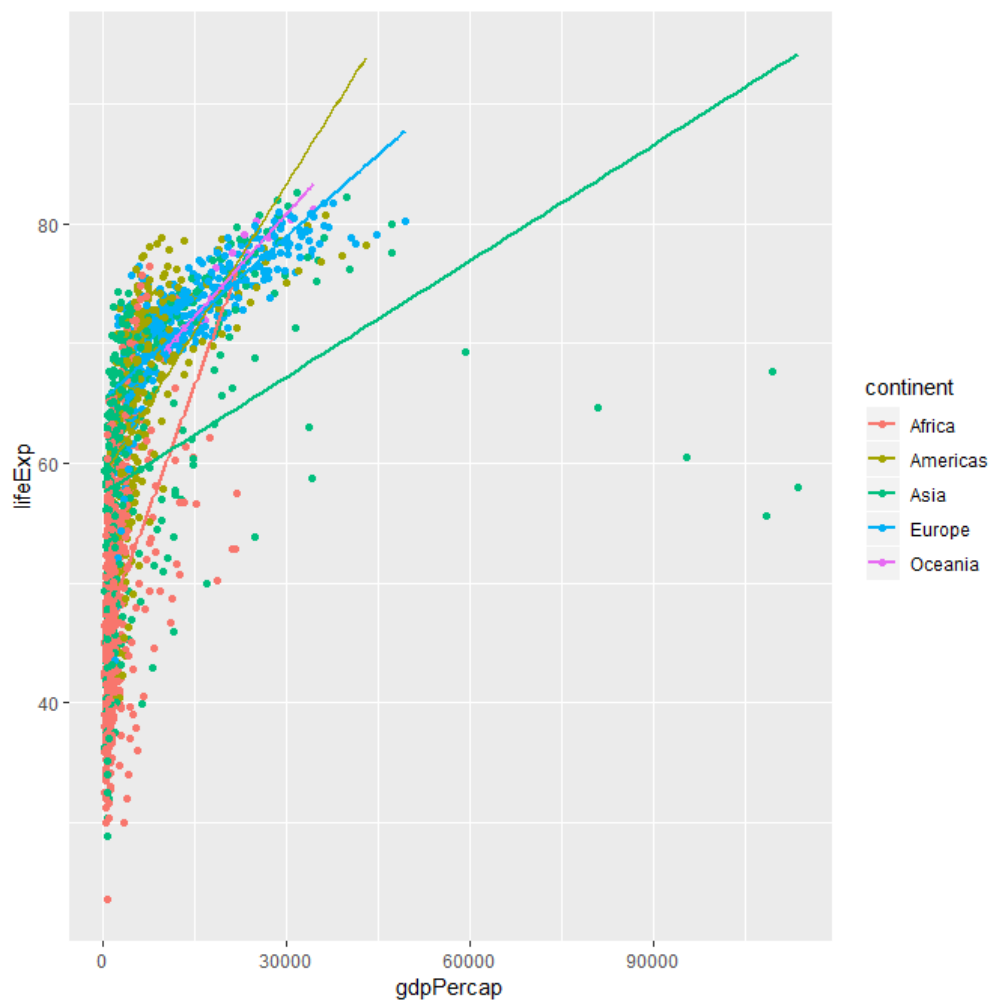
```
> gapminder %>%
```

```
ggplot(aes(x = gdpPercap, y = lifeExp, colour = continent))
```

```
# continent is factor variables (can be seen by running str function)
```

```
geom_point() + geom_smooth(method = "lm", se= FALSE)
```

Output:



Interpret:

Here we see that each continent has a different association as explained by the fitted line. They are positive but have different slopes between the 5 continents. This indicates that continents differ substantially in their relationships between 'gdpPercap' and 'lifeExp'. Therefore, we should include the nested structure of the data (continent information) into our model.

3. Fit a random intercept model with 'lifeExp' as dependent variable and 'gdpPercapita' as predictor:

- write the model down
- show summary of the model fit
- give the estimated regression line
- briefly interpret the results

R code

```
> attach(gapminder)
```

```
# fitting a random intercept model
```

```
#  $\text{lifeExp}_{ij} = \beta_0 + \beta_1(\text{gdpPercap}_{ij}) + \mu_{0j} + \varepsilon_{ij}$ 
```

```
> model2 <- lmer (lifeExp ~ gdpPercap + (1|continent), REML =  
FALSE)
```

```
> summary(model2)
```

```
> prelife <- fitted(model2)
```

```
> xyplot (prelife ~ gdpPercap, gapminder, groups=continent, typ=c  
("p", "l"))
```

Output:

```
Linear mixed model fit by maximum likelihood ['lmerMod']
Formula: lifeExp ~ gdpPercap + (1 | continent)
```

AIC	BIC	logLik	deviance	df.resid
12116.7	12138.5	-6054.3	12108.7	1700

Scaled residuals:

Min	1Q	Median	3Q	Max
-5.8973	-0.5374	0.0420	0.6088	2.9922

Random effects:

Groups	Name	Variance	Std.Dev.
continent	(Intercept)	44.42	6.665
Residual		70.35	8.388

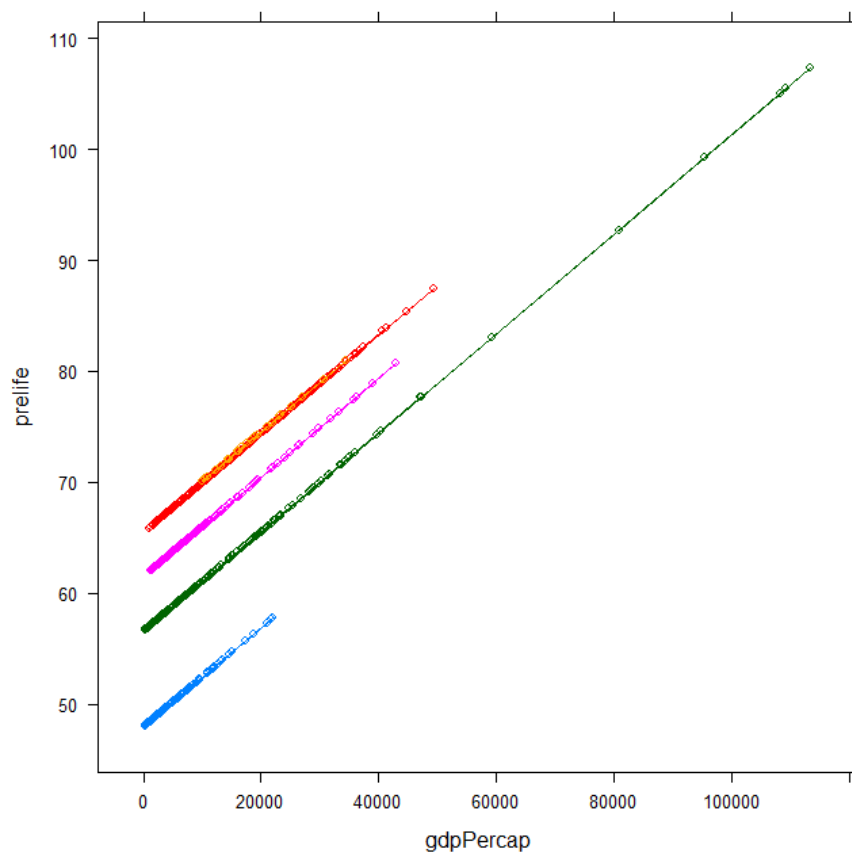
Number of obs: 1704, groups: continent, 5

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	5.938e+01	3.013e+00	19.70
gdpPercap	4.474e-04	2.347e-05	19.06

Correlation of Fixed Effects:

	(Intr)
gdpPercap	-0.078



Model:

The equation of average fitted line across continents is:

$$\text{lifeExp}_{ij} = 59.38 + 0.0004474 (\text{gdpPercap}_{ij})$$

Interpret:

In this model, the changes in 'gdpPercap' is fixed across all continents. The fitted line for a given continent is differ from this average line in its intercept by an amount of μ_{0j} for each continent j (intercepts are random). Thus, the plot of the predicted continent lines shown a set of parallel lines. Each continent has their own fitted line with the same slope but different intercepts.

4. Fit a random slope and intercept model with 'lifeExp' as dependent variable and 'gdpPercapita' as predictor:

- write the model down
- show summary of the model fit
- give the estimated regression line
- briefly interpret the results

R code

```
> model3<-lmer(lifeExp ~ gdpPercap + (1 + gdpPercap|continent),
REML = FALSE)

> summary(model3)

> predlife<-fitted(model3)

> xyplot (predlife ~ gdpPercap, gapminder, groups=continent,
type=c ("p", "l"))
```

Output:

```
Linear mixed model fit by maximum likelihood ['lmerMod']  
Formula: lifeExp ~ gdpPercap + (1 + gdpPercap | continent)
```

AIC	BIC	logLik	deviance	df.resid
12031.0	12063.6	-6009.5	12019.0	1698

Scaled residuals:

Min	1Q	Median	3Q	Max
-4.6057	-0.5333	0.0477	0.6216	2.6193

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
continent	(Intercept)	6.133e+01	7.831323	
	gdpPercap	1.673e-07	0.000409	-0.86
Residual		6.625e+01	8.139518	

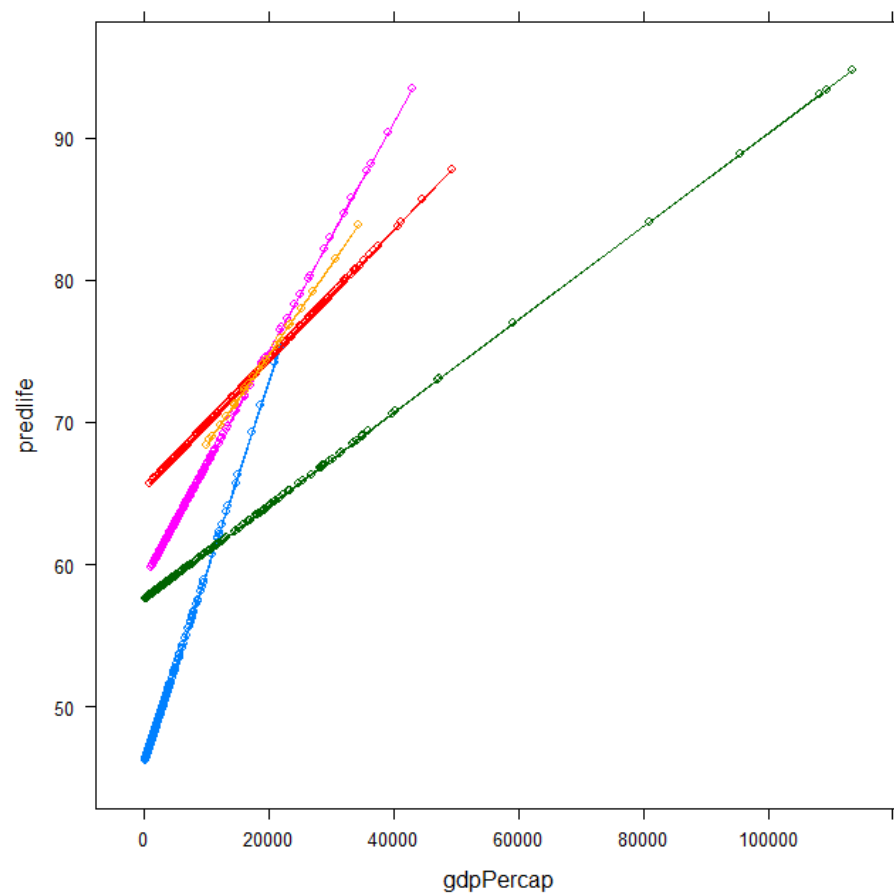
Number of obs: 1704, groups: continent, 5

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	5.794e+01	3.625e+00	15.983
gdpPercap	7.135e-04	1.909e-04	3.738

Correlation of Fixed Effects:

	(Intr)
gdpPercap	-0.855



By defining the random intercepts and slopes together, we indirectly specify that we want the random intercepts and slopes to covary.

Model:

According to the results, we can fit regression line for continent j is

$$\text{lifeExp}_{ij} = 57.94 + \hat{\mu}_{0j} + (0.0007135 + \hat{\mu}_{1j}) \text{gdpPercap}_{ij}$$

Interpret:

From the results we can see that the gdpPercap effect for continent j is estimated as $0.0007135 + \hat{\mu}_{1j}$, and the between continent variance in the slopes is estimated as 1.673×10^{-7} . For an 'average' continent we predict an increase of 0.7135 years in lifeExp achievement for extra 1000 GDP per capita. The intercept variance of 57.94 is interpreted as the between continent variance when gdpPercap=0.

In terms of the plot of the predicted continent lines, each continent has a different slope and intercept. However, they all have a positive relationship between the gdpPercap and lifeExp, this means that more gdpPercap will definitely contribute to the longer lifeExp.