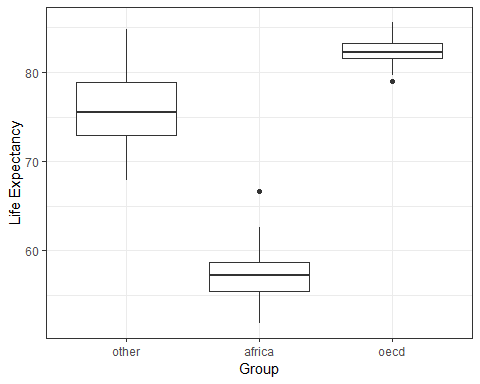
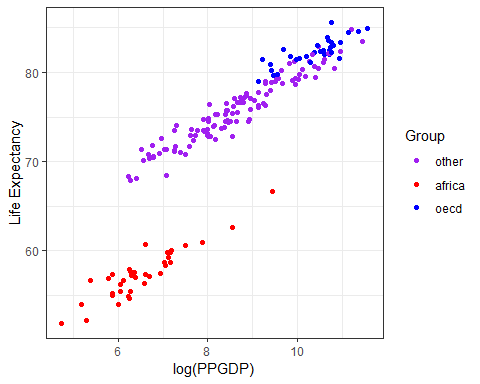
life expectancy

Ethan Kemeny & Chris Trippe

Macroeconomists hypothesize that there is a relationship between life expectancy and the economic well-being of the country. The given data provides an economic group for which each country is a part of (OECD–the organization for Economic Co-operation and Development, a think tank set towards improving global social and economic well-being, Africa, or another group) and the per person GDP (PPDGP). Building a regression model will allow for inference on which group is associated with higher life expectancy and help understand the effect PPDGP has on life expectancy.



When looking at the scatterplot of Life Expectancy and log(PPGDP), group Africa and other have about the same slope while group OECD slope is different. This means that depending on whether the group is OECD or not, the effect that log(PPGDP) has on Life Expectancy differs. This implies an interaction.

: the intercept; for countries in groups other than Africa and OECD with a log(PPGDP) of zero, the life expectancy is , on average.

: Holding all else constant and being in group other, as log(PPGDP) goes up by 1, the life expectancy goes up by , on average.

: For equal log(PPGDP) levels, a country in group Africa has an increase of life expectancy than a country in group “other”, on average.

: For equal log(PPGDP) levels, a country in group OECD has a higher life expectancy than a country in group “other”, on average.

: Life expectancy is expected to increase by more for an African country than a country in the group other as log(PPGDP) goes up by 1.

: Life expectancy is expected to increase by more for an OECD country than a country in the group other as log(PPGDP) goes up by 1.

This model assumes linear data, independence, the normality of the residuals and equal variance.

The purpose of this model is to make predictions of life expectancy. For group other, use to predict life expectancy. For group Africa, use to predict life expectancy. Lastly, for group OECD use to predict life expectancy.

Estimate Std. Error t value Pr(>|t|)  
(Intercept) 50.42403070 0.71286344 70.7344884 1.151147e-130  
log(PPGDP) 2.93881784 0.08187025 35.8960410 1.260139e-82  
Groupafrica -11.89511233 1.47534582 -8.0625926 1.131235e-13  
Groupoecd 11.29201446 3.21336844 3.5140740 5.617474e-04  
log(PPGDP):Groupafrica -0.04128095 0.21248820 -0.1942741 8.461866e-01  
log(PPGDP):Groupoecd -0.95267636 0.31279037 -3.0457343 2.680205e-03

: the intercept; for countries in groups other than Africa and OECD with a log(PPGDP) of zero, the life expectancy is 50.42, on average.

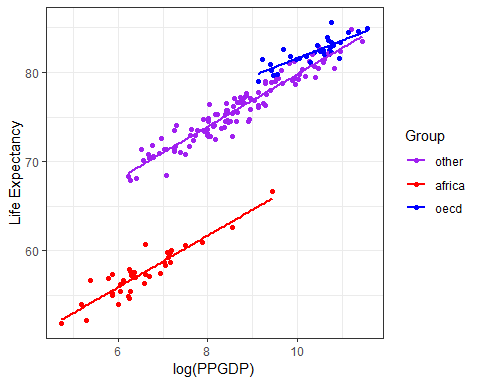
: Holding all else constant and being in group other, as log(PPGDP) goes up by 1, the life expectancy goes up by 2.94, on average.

: For equal log(PPGDP) levels, a country in group Africa has an 11.9 lower life expectancy than a country in group “other”, on average.

: For equal log(PPGDP) levels, a country in group OECD has an 11.29 higher life expectancy than a country in group “other”, on average.

: Life expectancy is expected to decrease by 11.9 + 0.04 or 11.94 if a group changes from other to African and the log(PPGDP) goes up by 1.

: Life expectancy is expected to increase by 11.29 - 0.95 or 10.34 if a group changes from other to African and the log(PPGDP) goes up by 1.



Using a cross-validation study of 500 samples we are able to get a sense of how accurate our model is at predicting life expectancy. The bias is -0.0022. This means our model typically predicts life expectancy -0.0022 below the actual value on average. That value is relatively small so we feel our model isn’t very biased. The root prediction mean squared error value of 0.1722 tells us our predictions miss the mark by an average of 0.1722. This is also very low so we feel that our model is accurate in its predictions.

We also wanted to look at our prediction intervals or the low and high end of our estimated values. The study returned a coverage of 0.9543704 which is the percentage of prediction intervals that contain the actual value. With a width of 4.2653756, we can conclude that our prediction intervals are not very large relative to the data.

The for this model is 0.9865 this means our model of log(PPGDP) + the groups and the interactions can explain 98.65% of the variation in life expectancy. This value is quite high, because of this value and the other statistics from the study so we feel confident in our model’s ability to predict values.

$fstatistic  
 value numdf dendf   
2550.608 5.000 175.000

p-value: 2.2e-16

With a p-value of 2.2e-16 from the F-test, we can conclude that at least one of the variable coefficients is significant.

2.5 % 97.5 %  
(Intercept) 49.0171145 51.8309469  
log(PPGDP) 2.7772377 3.1003980  
Groupafrica -14.8068732 -8.9833514  
Groupoecd 4.9500705 17.6339585  
log(PPGDP):Groupafrica -0.4606503 0.3780884  
log(PPGDP):Groupoecd -1.5700033 -0.3353494

For log(PPGDP), we are 95% confident that the true increase in life expectancy for the group other, as log(PPGDP) goes up by 1, is between 2.77 and 3.1004.

For the interaction between log(PPGDP) and Group: OECD we are 95% confident that life expectancy will decrease between 1.57 and .335 for countries that change from other to the OECD when the log(PPGDP) goes up by 1.

With this data, we can conclude that countries with higher GDP will have a higher life expectancy.

Analysis of Variance Table  
  
Model 1: LifeExp ~ log(PPGDP) + Group + log(PPGDP) \* Group  
Model 2: LifeExp ~ log(PPGDP) + Group  
 Res.Df RSS Df Sum of Sq F Pr(>F)   
1 175 195.12   
2 177 205.48 -2 -10.357 4.6447 0.01083 \*  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

With a p-value of .01083, we reject the null hypothesis that the interaction is not significant. This means we know that the group a country is in has a significant impact on how log(PPGDP) affects life expectancy. This means that when looking at the relationship between log(PPGDP) and life expectancy, the group variable must be taken into account.