Econometrics Project

Does GDP Affect Happiness?

Econometrics Project

Does GDP Affect Happiness?

by

Jack Curtin
16181484

EC6062: Applied Econometrics for Business

Contents

1. Introduction

We've all heard the saying, "Money doesn't buy happiness", but what if it does? We are going to examine if this statement has any validity. To do so, we're going to analyse the affect of Gross Domestic Product (GDP) per capita on happiness, using data from the Gallup World Poll surveys which has been compiled by the World Happiness Report (World Happiness Report 2022; Helliwell et al. 2022). The result of our analysis will have larger implications, namely if GDP per capita is a sufficient factor in determining one's happiness, or if other factors such as life expectancy have a greater influence.

The inspiration for this project came from a genuine curiosity as to how much influence income has on an individual's happiness. Of course, there is no avoiding the external influences on happiness which we cannot measure with our current data set. Some of these influences may include how robust a country's healthcare infrastructure is, the average educational level among citizens, and the affordability of housing. Regardless, we will create a linear model using GDP per capita as a predictor variable to predict happiness. To do so, we will begin by exploring our data set to see if anything appears interesting or unusual, and if any values need to be removed. Next, we will preform a deeper analysis of the data set to spot any trends or correlations which will help us determine if any other variables should be included in our model. Finally, we will create and assess a linear model, and discuss any improvements which could be made for future analysis.

1.1. The Data Set

The data set we are looking at consists of survey data collected from the Gallup World Poll surveys from 2005 to 2020 (Helliwell et al. 2022). This data was collected from phone interviews, where a random-digital-dial method was used to obtain random phone numbers within each country. Then, at least 1,000 people are interviewed over the phone, and to avoid selection bias from the likelihood of interviewing those who live in larger cities, the results are weighted to correct this probability. The original data set contained 1,949 rows ($n = 1,949$) with 11 columns, though we decided to create a new column which contained the log transformation of *GDP per capita* (measured in USD), and we shall explain why in the coming section. The first five rows can be seen in **Figure 1.1** below, and our variables can be explained as follows:

- *Happiness* - Mean value from 0 to 10, where 10 corresponds to survey participants believing they have the best possible standard of life, and 0 otherwise.
- *Social support* - Mean value from 0 to 1, where 1 corresponds to survey participants believing they could count on friends and family for support, and 0 otherwise.
- *Freedom to make life choices* - Mean value from 0 to 1, where 1 corresponds to survey participants believing they have freedom of choice, and 0 otherwise.
- *Generosity* - Regression residual where survey participants where asked if they had donated to a charity within the last month.

- *Perceptions of corruption* - Mean value from 0 to 1, where 1 corresponds to survey participants who believed corruption was present in their respective government, and 0 otherwise.
- *Positive affect* - Mean value from 0 to 1, where 1 corresponds to survey participants who felt feelings of happiness the previous day, and 0 otherwise.
- *Negative affect* - Mean value from 0 to 1, where 1 corresponds to survey participants who felt feelings of negativity the previous day, and 0 otherwise.

Country name	year	Happiness	GDP per capita	Log GDP per capita	Social support	Healthy life expectancy at birth	Freedom to make life choices	Generosity	Perceptions of corruption	Positive affect	Negative affect
Afghanistan	2008	3.724	1587.792	7.370	0.451	50.800	0.718	0.168	0.882	0.518	0.258
Afghanistan	2009	4.402	1881.778	7.540	0.552	51.200	0.679	0.190	0.850	0.584	0.237
Afghanistan	2010	4.758	2093.744	7.647	0.539	51.600	0.600	0.121	0.707	0.618	0.275
Afghanistan	2011	3.832	2037.609	7.620	0.521	51.920	0.496	0.162	0.731	0.611	0.267
Afghanistan	2012	3.783	2220.481	7.705	0.521	52.240	0.531	0.236	0.776	0.710	0.268

Figure 1.1: Top five excel rows.

We can see a better summary of our variables from **Figure 1.2**, which describes a number of descriptive statistical values for each of our variables. Some Interesting insights to note from **Figure 1.2** are:

- Denmark has the highest happiness rating of 8.02 which it obtained in 2005, and this high happiness value seems concurrent with the other Nordic countries. This is due to many factors, primarily their advanced healthcare systems, low unemployment rates, and low crime rates.
- Afghanistan has the lowest happiness rating of 2.375, which it obtained in 2019, and is of course due to it's war-torn status. Low happiness values seem generally more apparent in African countries, which agrees with our theory that low GDP per capita countries aren't as happy as high GDP per capita countries.
- Luxembourg has the highest GDP per capita of $114,482 from 2019, along with the highest GDP per capita for every year from 2009 to 2019. This is due to their thriving finance sector and their status as a major player in private banking.
- Burundi has the lowest GDP per capita, of $761, which it obtained in 2018, and in general we see low GDP per capita values for countries on the continent of Africa.
- Singapore has the highest life expectancy of 77.1 years, which it obtained in 2019, and this high life expectancy trend also appears in Japan. This is due to both countries commitment to healthcare and caring for their elderly.
- Hati has the lowest life expectancy of 32.8 years, which it obtained in 2010, and the lowest value from 2020 was Nigeria with 50.5 years. This trend is common among countries with poor healthcare infrastructures, leading to deaths from curable diseases and pneumonia.
- Hungary has the highest perception of corruption, with a value of 0.983, which it obtained in 2010. This high perceived corruption is common among countries in Eastern Europe, and alike Hungary, it is due to citizens lack of trust in their governments and subsequent services.

- Singapore has the lowest perception of corruption, with a value of 0.035, which it obtained in 2009. This is due to their robust public services and their government's transparency.

In terms of missing values, we find the following counts listed as *NA's* in **Figure 1.2**. However, these missing values occur at random, and as such shouldn't introduce bias but may result in larger standard errors. Thus, when creating our regression model, we will simply omit any of these missing values. We should also note some significant outliers exist, being the 23 cases where the value for GDP per capita is listed as $1. In order to create a more accurate model we will omit these values; thus our new sample size will be $n = 1926$.

```
 Happiness       GDP per capita   Social support   Life expectancy  Freedom
 Min.   :2.375   Min.   :     1   Min.   :0.2902   Min.   :32.30    Min.   :0.2575
 1st Qu.:4.640   1st Qu.:  4628   1st Qu.:0.7494   1st Qu.:58.69    1st Qu.:0.6471
 Median :5.386   Median : 12704   Median :0.8352   Median :65.20    Median :0.7635
 Mean   :5.467   Mean   : 20139   Mean   :0.8126   Mean   :63.36    Mean   :0.7426
 3rd Qu.:6.283   3rd Qu.: 31158   3rd Qu.:0.9053   3rd Qu.:68.59    3rd Qu.:0.8560
 Max.   :8.019   Max.   :114482   Max.   :0.9873   Max.   :77.10    Max.   :0.9852
                                  NA's   :13       NA's   :55       NA's   :32

Generosity         Perceptions of corruption Positive affect  Negative affect
Min.   :-0.33504   Min.   :0.0352            Min.   :0.3217   Min.   :0.08274
1st Qu.:-0.11297   1st Qu.:0.6903            1st Qu.:0.6254   1st Qu.:0.20640
Median :-0.02539   Median :0.8024            Median :0.7224   Median :0.25812
Mean   : 0.00011   Mean   :0.7471            Mean   :0.7100   Mean   :0.26855
3rd Qu.: 0.09097   3rd Qu.:0.8719            3rd Qu.:0.7993   3rd Qu.:0.31972
Max.   : 0.69810   Max.   :0.9833            Max.   :0.9436   Max.   :0.70459
NA's   :89         NA's   :110               NA's   :22       NA's   :16
```

Figure 1.2: Summary of our variables including counts of missing values.

2. Statistical Analysis

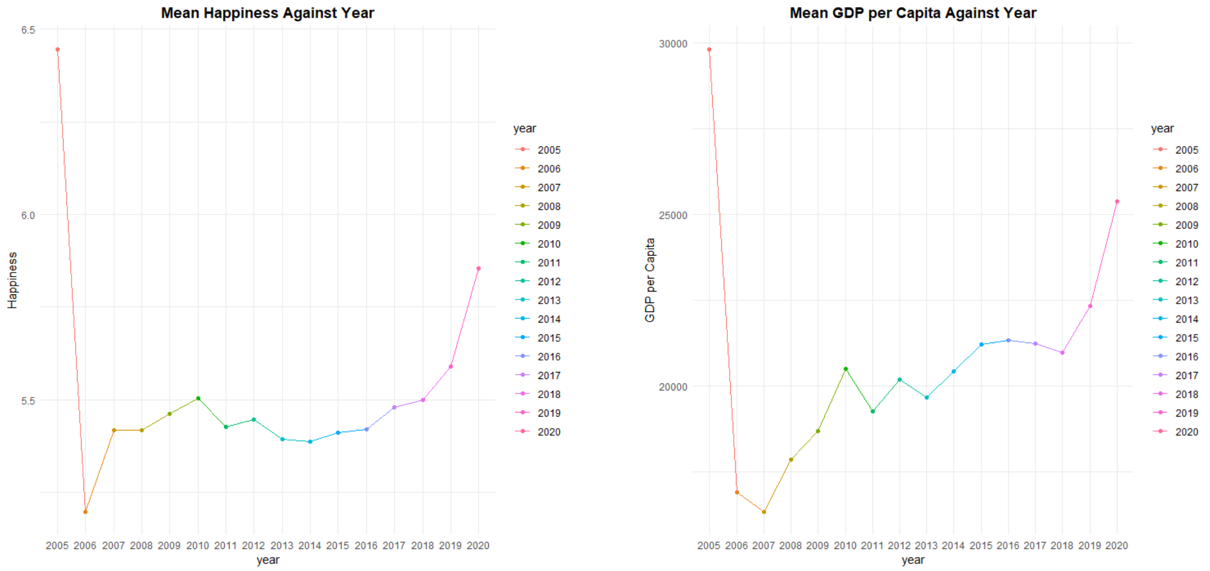

Figure 2.1: Plots of happiness and GDP per capita against year.

When exploring the data, we find a general upward trend for both the mean of happiness and the mean of GDP per capita per year, as shown in **Figure 2.1**. This suggests that a linear relationship may exist between the two. It's also apparent that in both scenarios they had maximum values in the year 2005 and their lowest values in 2006 and 2007 respectively, as seen from the graph and from the table in **Figure 2.2** below. This raises the question as to how perceived happiness can drop from 6.4 to 5.2, and how GDP per capita could drop from $29.8k to $16.9k in the span of 1-2 years.

The reason for this drop in GDP per capita is of course due to the housing market crash, which occurred in 2007, and resulted in unprecedented layoffs around the world (Duca, Muellbauer, and Murphy 2010). Thus, considering our assumption that GDP per capita affects happiness, we could reasonably assume that the loss of jobs which resulted from the crash, and the resulting reduction in income per person, could explain the drop in happiness. As GDP per capita is a variable which grows exponentially over time, we shall preform a log transformation of this variable to make it more linear and thus easier to manage when we create a linear model.

Year	Happiness	GDP per capita	Social support	Life expectancy	Freedom to make life choices	Generosity	Perceptions of corruption	Positive affect	Negative affect
2005	6.446	$ 29,826.40	0.897	67.007	0.830	0.256	0.716	0.749	0.256
2006	5.197	$ 16,916.06	0.836	60.148	0.731	0.007	0.756	0.719	0.255
2007	5.418	$ 16,335.16	0.808	61.305	0.687	0.015	0.792	0.709	0.250
2008	5.419	$ 17,858.08	0.784	61.250	0.688	0.021	0.764	0.706	0.244
2009	5.462	$ 18,693.93	0.819	62.419	0.687	-0.005	0.766	0.705	0.248
2010	5.504	$ 20,502.47	0.832	63.069	0.707	0.003	0.760	0.712	0.242
2011	5.427	$ 19,266.98	0.803	62.212	0.731	-0.014	0.758	0.701	0.250
2012	5.446	$ 20,193.59	0.809	63.217	0.711	-0.002	0.761	0.707	0.261
2013	5.393	$ 19,669.72	0.806	63.326	0.728	0.000	0.764	0.713	0.270
2014	5.386	$ 20,444.96	0.806	63.296	0.734	0.020	0.739	0.709	0.270
2015	5.411	$ 21,229.36	0.801	63.862	0.749	0.020	0.740	0.709	0.278
2016	5.420	$ 21,338.31	0.816	64.061	0.764	0.001	0.749	0.711	0.284
2017	5.479	$ 21,246.81	0.807	64.182	0.783	-0.007	0.728	0.706	0.290
2018	5.499	$ 20,977.55	0.812	64.555	0.784	-0.023	0.735	0.711	0.293
2019	5.590	$ 22,333.64	0.816	65.037	0.799	-0.020	0.721	0.714	0.286
2020	5.855	$ 25,382.00	0.839	67.107	0.823	-0.008	0.706	0.721	0.291

Figure 2.2: Mean values of our numeric variables by year.

From the correlation plot below (**Figure 2.3**), we can see that there exists a really strong correlation between the log of GDP per capita, life expectancy and social support, with happiness (0.79, 0.75, 0.7 respectively). This suggests that when creating a linear model we should consider each of these three variables. Interestingly, perceptions of corruption have quite a moderate negative effect on happiness (-0.43), and surprisingly, generosity has a relatively weak positive effect on happiness (0.2). Another interesting observation is the really strong positive correlation between log of GDP per capita and life expectancy (0.84), which strongly implies that those who are wealthier live longer. Though of course this is partly to the correlation between wealthy countries and having access to the best medical care.

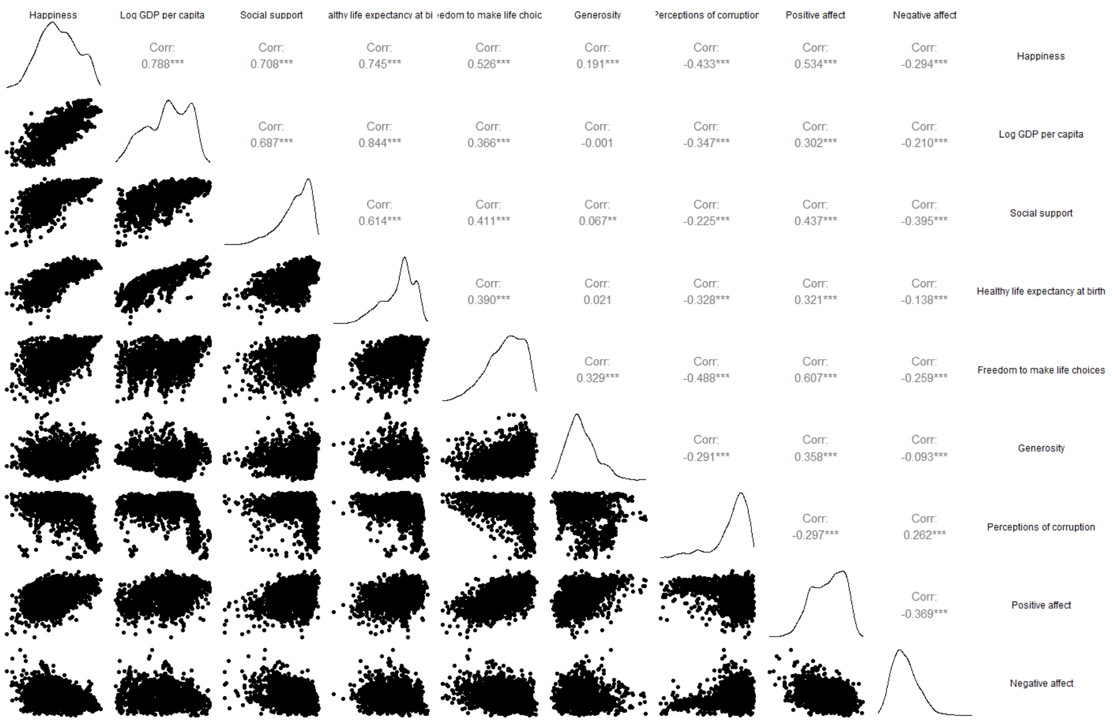

Figure 2.3: Correlation Plot.

3. Linear Model

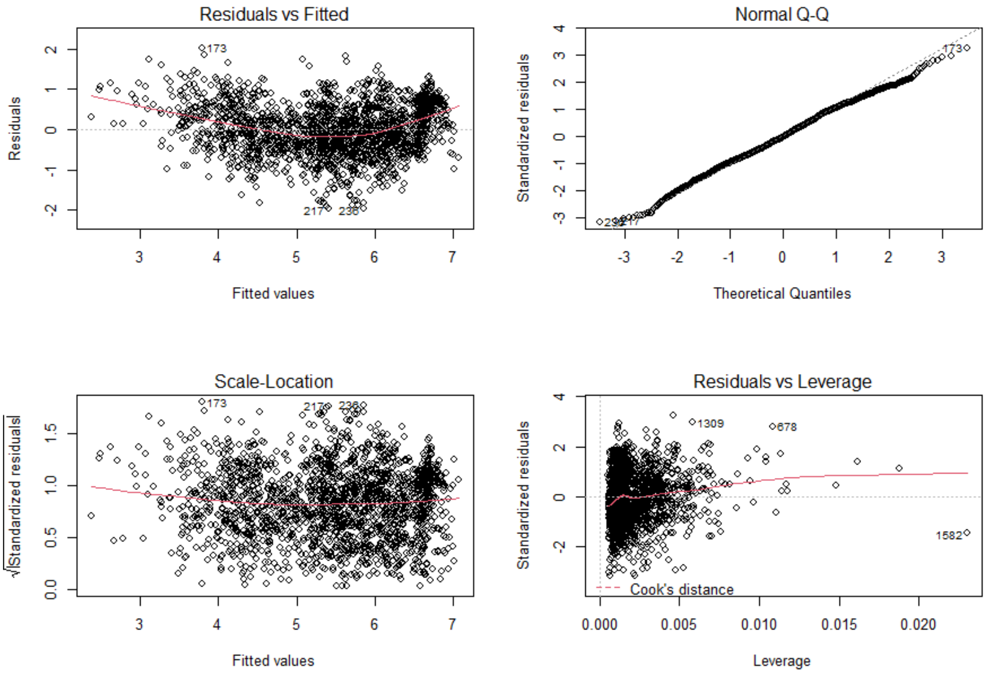

Figure 3.1: Residual and fitted values plots.

When creating the linear model to predict happiness, we decided against incorporating the year as a predictor variable. This decision was made to ensure our model could be used for future years, and to ensure that more emphasis was placed on log GDP per capita as a predictor variable. Thus, when we started by using only log GDP per capita as a predictor variable in our model, which resulted in an R^2 of 62.17%. Although, when we additionally include life expectancy and social support, our R^2 increases to a relatively respectable 69.26%. But we also need to check for heteroscedasticity, which we can visualize from **Figure 3.1**, and from our residual vs fitted plot we can see evidence of heteroscedasticity being present. To correct for this we will use robust standard errors, which results in the output seen in **Figure 3.2**:

```
Original standard errors:
                                        Estimate Std. Error t value Pr(>|t|)
    (Intercept)                        -2.576459   0.127142 -20.264   <2e-16 ***
    `Log GDP per capita`                0.377304   0.025666  14.701   <2e-16 ***
    `Healthy life expectancy at birth`  0.035309   0.003608   9.785   <2e-16 ***
    `Social support`                    2.797939   0.167462  16.708   <2e-16 ***
    ---
    Residual standard error: 0.6243 on 1867 degrees of freedom
      (55 observations deleted due to missingness)
    Multiple R-squared:  0.6926,        Adjusted R-squared:  0.6922
    F-statistic:  1402 on 3 and 1867 DF,  p-value: < 2.2e-16

Heteroskedasticity-robust standard errors:
                                         Estimate Std. Error  t value  Pr(>|t|)
    (Intercept)                        -2.5764593  0.1320160 -19.5163 < 2.2e-16 ***
    `Log GDP per capita`                0.3773044  0.0253218  14.9004 < 2.2e-16 ***
    `Healthy life expectancy at birth`  0.0353095  0.0037323   9.4606 < 2.2e-16 ***
    `Social support`                    2.7979392  0.1755748  15.9359 < 2.2e-16 ***
```

Figure 3.2: Residuals and output from R.

What we notice immediately is that all of our values are statistically significant, which can be seen from the P-values for each coefficient being essentially zero. Due to our large sample size ($n = 1926$), we can see that our new standard errors are very similar to our previous standard errors. Thus, from our estimate values our linear regression equation looks as follows:

$$Y_i = -2.576 + 0.377X_{1i} + 0.035X_{2i} + 2.798X_{3i} + \epsilon_i. \tag{3.1}$$

Where:

$$\begin{aligned} Y_i &= \text{Happiness score}, & X_{3i} &= \text{Social Support}, \\ X_{1i} &= \text{Log GDP per capita}, & \epsilon_i &= \text{Error term}. \\ X_{2i} &= \text{Life expectancy}, \end{aligned}$$

It should be noted that the error term, ϵ_i, could be the effect Covid has on happiness, which is something we can't reasonably measure and thus cannot include in our model.

A plot of our predicted vs the actual happiness against year can be seen in **Figure 3.3** below, just note we used the mean values of Log GDP per capita, life expectancy and perceived social support per year. We decided to use these mean values as it gives a better idea as to how accurate our model is overall. We also plotted happiness against GDP per capita and the log of GDP per capita. The plot containing GDP per capita seems to taper off from $70,000 - $80,000, at a happiness level of 7, which agrees with the conclusion that other articles came to when examining the theory that money buys happiness. This conclusion is that once a person's income reaches roughly $70,000 per year, any further increase in income has no major impact on happiness (Frank 2005; Gatina 2016).

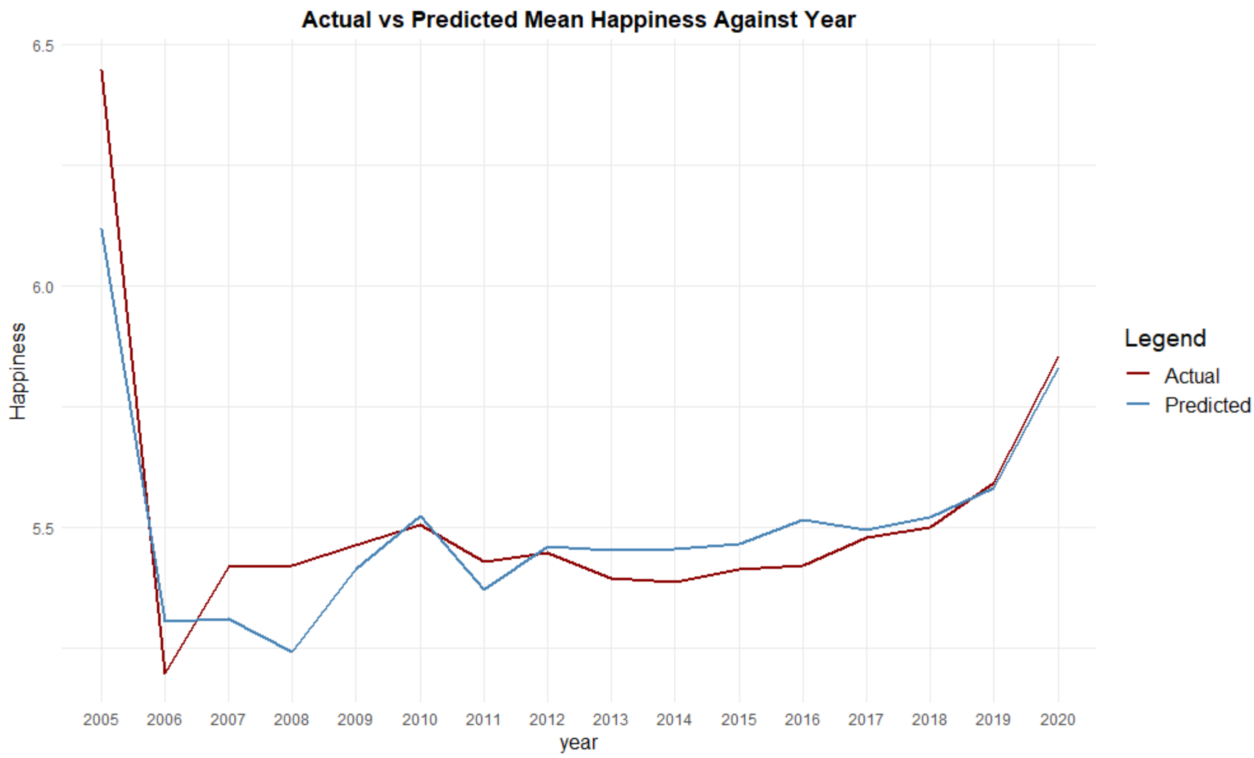

Figure 3.3: Actual vs predicted mean happiness against year.

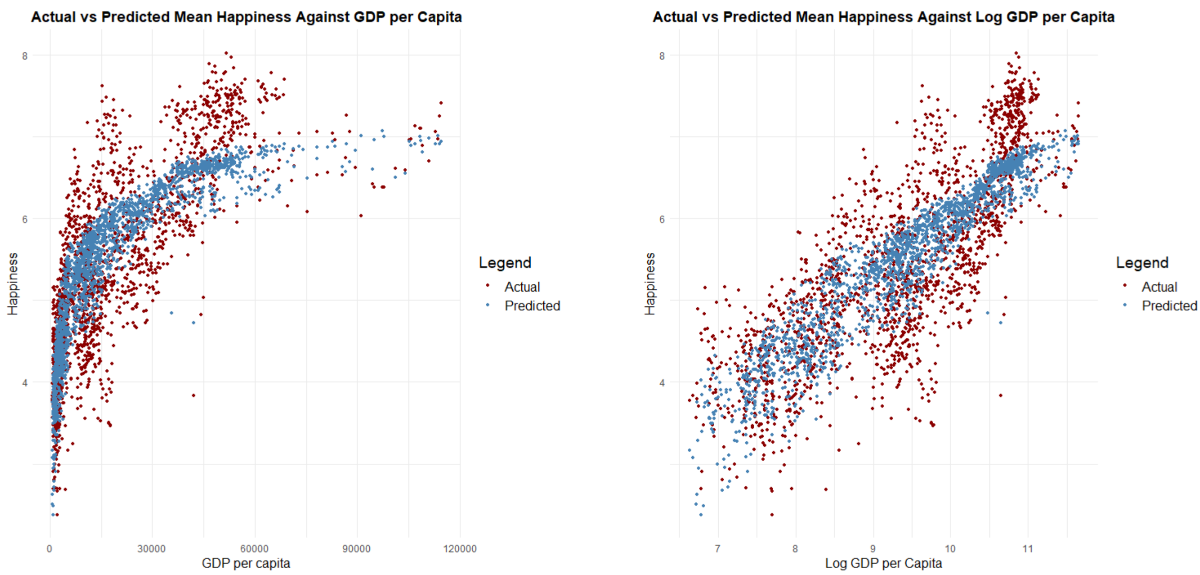

Figure 3.4: Plots of happiness against GDP per capita and log of GDP per capita.

4. Summary

We have thus successfully created a model to predict happiness using on the log of GDP per capita (**Equation 3.1**), life expectancy and perceived social support. We began by defining what each of our variables in the data set represented, and how they were measured. We then explored this data set and highlighted the more interesting observations and apparent trends which we could reasonably determine. This brief analysis led us to discover some significant outliers, namely countries which had a GDP per capita value of $1, and to ensure our model was accurate we removed the 23 occurrences of this error. We also determined it would be best to create a new variable, the log of GDP per capita, as GDP per capita is a variable which grows exponentially over time.

This led us to perform some more complex analysis of our data set, which began with plots of both happiness and GDP per capita against year. We noticed from an initial visual glance that these plots seemed to be correlated, and this theory was later confirmed with a correlation plot. Evident in both plots was a significant drop from 2005 to 2006/2007, which we discovered was due to the infamous housing market crash. This correlation plot also showed that strong correlation also existed between life expectancy and perceived social support with happiness, and that really significant correlation existed between log GDP per capita and life expectancy.

When creating a model to predict happiness, we objectively decided against including the year as a predictor variable, this was to ensure our model could work for future years. As such, our first attempt had considered using only the log of GDP per capita. Though after some trial and error we found that the best model to predict happiness additionally included using life expectancy and perceived social support as predictor variables. This model had an R^2 value of 69.26%, which implies our predictor variables describe happiness quite well. Due to the presence of heteroscedasticity, our initial standard errors were incorrect, so we calculated the robust standard errors to correct this error. From our plots of both the actual and predicted happiness against GDP per capita, we could visually confirm that money buys happiness, although past a GDP per capita of around $75,000 happiness seems to plateau.

5. Discussion

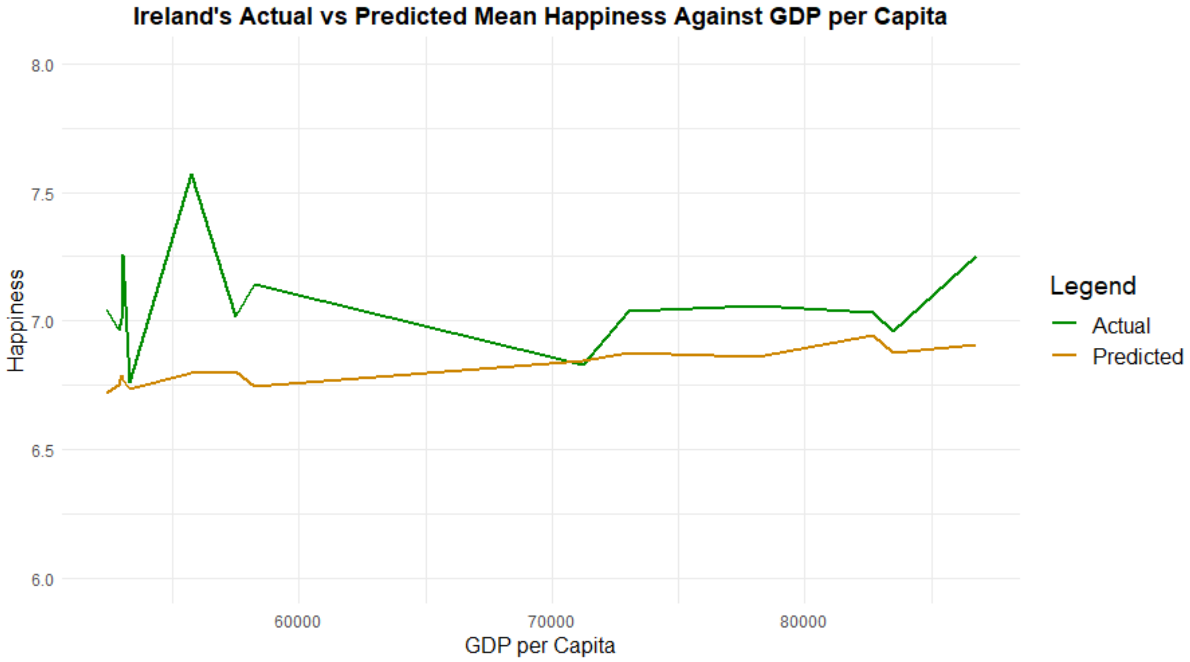

Figure 5.1: Ireland's actual vs predicted mean happiness against GDP per capita.

Figure 5.1 show's Ireland's happiness compared to GDP per capita, though we should note the peak happiness of 7.6 was during the Celtic Tiger, where Ireland experienced unprecedented economic growth, which was followed by a financial crisis (Kirby 2016). From our analysis and subsequent linear model (**Equation 3.1**), we can reasonably conclude that money does in fact buy happiness. This result aligns with the belief many economists share, and which many articles have proven (DeVoe and Pfeffer 2009; Matz, Gladstone, and Stillwell 2016; Tella, MacCulloch, and Oswald 2003). Though it's important to note that we are not insinuating that GDP per capita is the only factor directly correlated with happiness. Our model suggested that life expectancy and perceived social support also contribute to happiness. Life expectancy being significant could however be due to causality, as those from countries with high GDPs tend to have access to better healthcare infrastructures.

So how could our analysis be improved further? One obvious suggestion would be to somehow incorporate the affect Covid-19 has on happiness, as this is a variable which we couldn't incorporate in our predictive model. Of course the affect which Covid-19 may have on happiness would only be relevant from 2020 until early 2022, depending on which country we consider. As such, perhaps a better determinant of happiness could be the average level of education received, although we may again see some correlation between higher educated countries having higher paying jobs, and subsequent higher GDP per capita values.

References

DeVoe, Sanford E and Jeffrey Pfeffer (2009). “When is happiness about how much you earn? The effect of hourly payment on the money—happiness connection”. In: *Personality and Social Psychology Bulletin* 35.12, pp. 1602–1618. URL: https://journals.sagepub.com/doi/abs/10.1177/0146167209346304?casa_token=xZL1ZpVOuYkAAAAA:mF7ywsrQZPg4uiJK89OJnowAN5Sj1APM84ifogV5eH-bd-5WovY4B8GOe4D5o6J-YmAHcwS51C6F.

Duca, John V, John Muellbauer, and Anthony Murphy (2010). “Housing markets and the financial crisis of 2007–2009: lessons for the future”. In: *Journal of financial stability* 6.4, pp. 203–217. URL: https://www.sciencedirect.com/science/article/pii/S1572308910000343?casa_token=FOMgKcjifRIAAAAA:EUPUBh9BkXJqieJZFsxWhosGMyL-ebHRHguHZODDjnTNzHpxthoJosDrqH_56B_YN4OQ1UXK.

Frank, Robert H (2005). “Does money buy happiness”. In: *The science of well-being*, pp. 461–473. URL: https://books.google.ie/books?hl=en&lr=&id=DDFqqeko9boC&oi=fnd&pg=PA461&dq=does+money+buy+happiness&ots=VrhOPxK4d7&sig=OiHRRd18ge3cHNh9zAwAtOo_-eE&redir_esc=y#v=onepage&q=does%20money%20buy%20happiness&f=false.

Gatina, Liliya (2016). “Does money buy happiness? Financial and general well-being of immigrants in Australia”. In: *Journal of Behavioral and Experimental Economics* 63, pp. 91–105. URL: https://www.sciencedirect.com/science/article/pii/S2214804316300325?casa_token=vWdHIjRkdtgAAAAA:qNKlQHvOl44HG5Oi3gEO4YbO2K5nnhmFvBqGzoqHuPzi4l8qIp_eOWp72Vb1D6jXSkUlIwrX.

Helliwell, John F et al. (2022). *Statistical Appendix for “Happiness, benevolence, and trust during COVID-19 and beyond,” Chapter 2 of World Happiness Report 2022*. URL: https://happiness-report.s3.amazonaws.com/2022/Appendix_1_StatiscalAppendix_Ch2.pdf.

Kirby, Peadar (2016). *The Celtic tiger in distress: Growth with inequality in Ireland*. Springer.

Matz, Sandra C, Joe J Gladstone, and David Stillwell (2016). “Money buys happiness when spending fits our personality”. In: *Psychological science* 27.5, pp. 715–725. URL: https://pubmed.ncbi.nlm.nih.gov/27056977/.

Tella, Rafael Di, Robert J MacCulloch, and Andrew J Oswald (2003). “The macroeconomics of happiness”. In: *Review of Economics and Statistics* 85.4, pp. 809–827. URL: https://www.jstor.org/stable/pdf/3211807.pdf?casa_token=lnYfrSNpte4AAAAA:EtNeJwOewka1GTEfkg2j2eN3O9Q0pUGegOzHuwkJn-2_tcQHryaK8tPVcL4DKu-BkKlF7IiNyk4yOEBde4S9GJoqStSOkP_kxlPNXcNIbyqLN32SfpI.

World Happiness Report (2022). *World Happiness Report Dataset*. URL: https://happiness-report.s3.amazonaws.com/2021/DataPanelWHR2021C2.xls.