Comparative Analysis of UK NHS and OECD Countries' Health Systems in 2011

May 2, 2024

Background

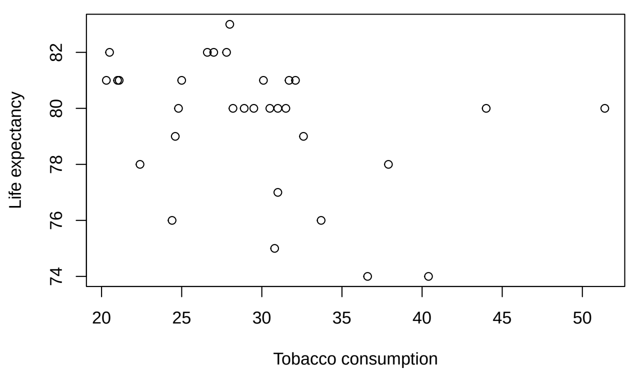
In 2011, the UK Government initiated a significant reorganization of the National Health Service, aimed at improving outcomes and efficiency by decentralising budget control.However,some critiques of prior healthcare provision highlighted poorer survival rates for several major diseases compared to other OECD countries, prompting debates on the effectiveness of health expenditure.In this report, we analyse data from the World Health Organization (WHO) to identify variables that are associated with life expectancy and to determine their influence on our appropriately constructed model.The result shows that UK health care system had a rather good performance comparing to other OECD countries, refuting the Andrew Lansley's claim.

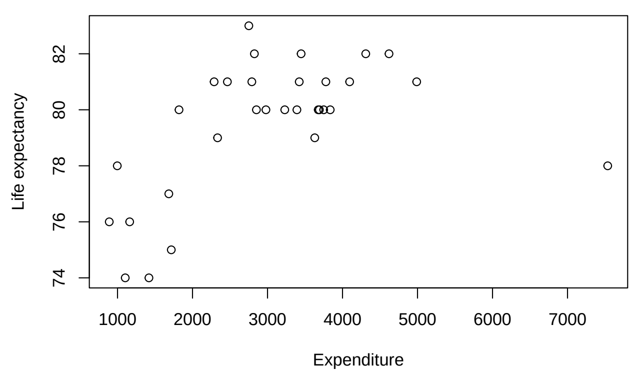
Results

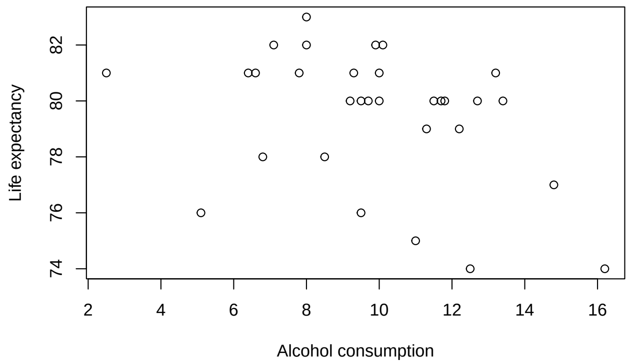
The parameters for our model are defined as follows:

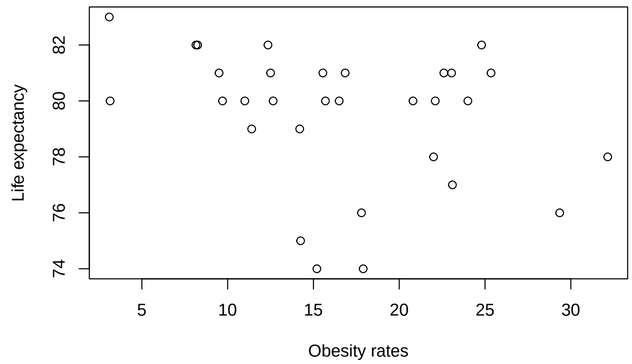
- y_i: Life expectancy of citizens at birth,in years.
- x_i: Total expenditure per person on health.
- z_i: Percentage of tobacco use among adults.
- m_i: The alcohol consumption among adults.
- k_i: The percentage of adults classified as obese.

To learn about the possible relationships,the dependent variable y_i is individually plotted against each independent variable in four scatter plots.

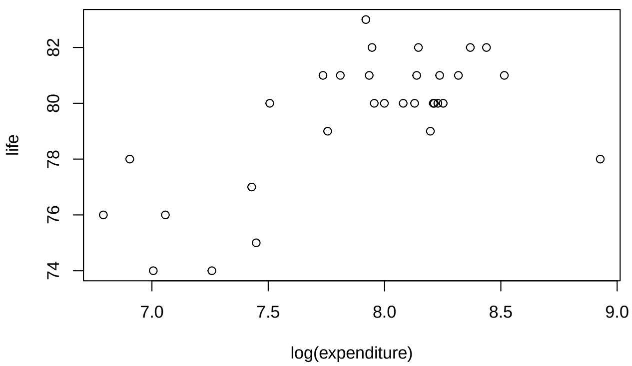




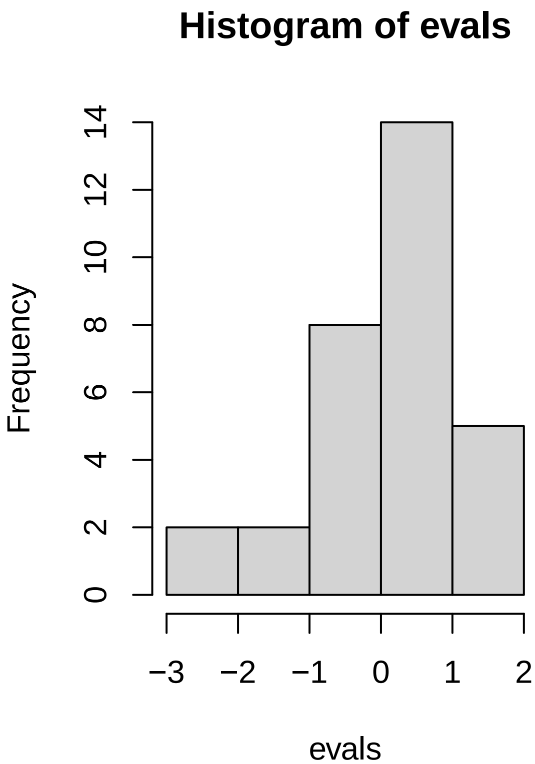


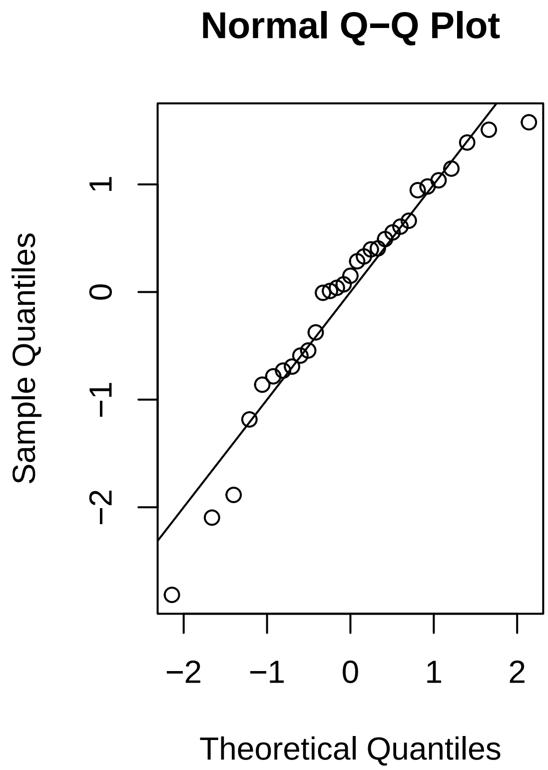


Upon reviewing the plots, the expenditure has a much larger range of values than tobacco consumption, alcohol consumption, and obesity rate. It suggests that expenditure disproportionately influences the model. The logarithmic transformation is needed to apply to moderate the range of expenditure values, ensuring that the relationship is not disproportionately influenced by larger values and aligning it more closely with the scales of the other variables.



In addition, the correlation coefficient between life expectancy and expenditure is 0.51, indicating a moderate positive correlation. After applying a logarithmic transformation to the data, the fit becomes more linear, and the correlation coefficient increases to 0.67, suggesting a strong correlation.Besides this, the correlation coefficients between life expectancy and tobacco consumption, alcohol consumption, and obesity rates are -0.34, -0.36, and -0.3, respectively, all indicating weak negative correlations.





From the result above, the full model is defined as:

$$M_f : y_i = \beta_0 + \beta_x \log(x_i) + \beta_z z_i + \beta_m m_i + \beta_k k_i + \epsilon_i, \text{ where } \epsilon_i \sim N(0, \sigma^2)$$

To validate the assumption, a Kolmogorov-Smirnov (K-S) test is applied to check whether the residuals of the full model are normally distributed. The resulting p-value of 0.511 indicates that there is no evidence to reject the null hypothesis of normality, suggesting that our model is appropriate with i.i.d. variances.

Now all the models nested within M_f as follow

$$\begin{aligned}
M_0 &: y_i = \beta_0 + \epsilon_i \\
M_1 &: y_i = \beta_0 + \beta_z z_i + \beta_m m_i + \beta_k k_i + \epsilon_i \\
M_2 &: y_i = \beta_0 + \beta_x \log(x_i) + \beta_m m_i + \beta_k k_i + \epsilon_i \\
M_3 &: y_i = \beta_0 + \beta_x \log(x_i) + \beta_k k_i + \epsilon_i \\
M_4 &: y_i = \beta_0 + \beta_x \log(x_i) + \beta_m m_i + \epsilon_i
\end{aligned}$$

where $\epsilon_i \sim N(0, \sigma^2)$.

To determine which model best fits the data, one-way ANOVA tests is needed to be applied.

Firstly, we apply the test to model M_0, which assumes no relationship between life expectancy and any of the parameters, and to the full model M_f. The resulting p-value is 2.3×10^{-5}, providing strong evidence to reject the null hypothesis H_0 in favour of the full model.Similarly,we use the same method to compare M_1, which assumes there is no relationship between the life expectancy and the expenditure, with M_f, yielding a p-value of 0.0001061, which strongly suggests that the full model M_f cannot be reduced to M_1.

Comparing M_2, which assumes no relationship between tobacco consumption and life expectancy, with M_f, the p-value is 0.3238. This suggests there is strong evidence to support that tobacco consumption has no significant effect on the model, which can be reduced.Thus, M_2 can be considered as a good reduced model.

Finally, M_2 is compared with M_3 (assuming there is no relationship among life expectancy, alcohol consumption, and tobacco consumption) and M_4 (assuming there is no relationship among life expectancy, obesity rate, and tobacco consumption), both of which are further simplified from M_2. The p-values given by the output are 0.007788 and 0.04133, suggesting there is no evidence to reject the null hypotheses, thereby preventing further reduction of the model simplified from M_2.Thus M_2 is the most suitable model. When the model is summarized, the adjusted R-squared is 0.6146, suggesting that M_2 accounts for about 60 percent of the variance in the data.

Within all above, the estimated life expectancy is 78.8962 years.

Conclusion

The average life expectancy in OECD countries is 79.48 years, while in the UK, it is 78.8962 years, which is slightly above the OECD average.It suggests the UK Health Care System had a rather good performance comparing with other OECD members,which is not enough evidence to support Andrew Lansley's claim.

R Commands

```r
library(MASS)
load("MAS2010.RData")
attach(who)
```

```r
# Scatter plots of life expectancy against each independent variable.

plot(tobacco,life,xlab = "Tobacco consumption",ylab = "Life expectancy")
```

```r
plot(expenditure,life,xlab = "Expenditure",ylab = "Life expectancy")
```

```r
plot(alcohol,life,xlab = "Alcohol consumption",ylab = "Life expectancy")
```

```r
plot(obesity,life,xlab = "Obesity rates",ylab = "Life expectancy")
```

```r
plot(log(expenditure), life,xlab = "Log(Expenditure)",ylab = "Life expectancy")
```

```r
# Correlations between 'life' and other variables
cor(expenditure, life)
```

```
## [1] 0.5084863
```

```r
cor(tobacco, life)
```

```
## [1] -0.3457086
```

```r
cor(alcohol, life)
```

```
## [1] -0.364971
```

```r
cor(obesity, life)
```

```
## [1] -0.3005739
```

```r
cor(log(expenditure), life)
```

```
## [1] 0.6718216
```

```r
# The histogram and Q-Q plot, and K-S test
lmf <- lm(life ~ log(expenditure) + tobacco + alcohol + obesity)
evals <- stdres(lmf)
```

```r
par(mfrow = c(1, 2))
hist(evals)
qqnorm(evals)
abline(0, 1)
```

```r
ks.test(evals, pnorm, 0, 1)
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  evals
## D = 0.14235, p-value = 0.511
## alternative hypothesis: two-sided
```

```r
# Linear models and One-Way ANOVA Tests
lmf <- lm(life ~ log(expenditure) + tobacco + alcohol + obesity)
lm0 <- lm(life ~ 1)
lm1 <- lm(life ~ tobacco + alcohol + obesity)
lm2 <- lm(life ~ log(expenditure) + alcohol + obesity)
lm3 <- lm(life ~ log(expenditure) + obesity)
lm4 <- lm(life ~ log(expenditure) + alcohol)
```

```r
anova(lmf, lm0)
```

```
## Analysis of Variance Table
##
## Model 1: life ~ log(expenditure) + tobacco + alcohol + obesity
## Model 2: life ~ 1
##   Res.Df     RSS Df Sum of Sq      F    Pr(>F)
## 1     26  63.711
## 2     30 171.742 -4   -108.03 11.022 2.313e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(lmf, lm1)
```

```
## Analysis of Variance Table
##
## Model 1: life ~ log(expenditure) + tobacco + alcohol + obesity
## Model 2: life ~ tobacco + alcohol + obesity
##   Res.Df     RSS Df Sum of Sq      F    Pr(>F)
## 1     26  63.711
## 2     27 114.767 -1   -51.056 20.836 0.0001061 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(lmf, lm2)
```

```
## Analysis of Variance Table
##
## Model 1: life ~ log(expenditure) + tobacco + alcohol + obesity
## Model 2: life ~ log(expenditure) + alcohol + obesity
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     26 63.711
## 2     27 66.190 -1   -2.4789 1.0116 0.3238
```

```r
anova(lm2, lm3)
```

```
## Analysis of Variance Table
##
## Model 1: life ~ log(expenditure) + alcohol + obesity
## Model 2: life ~ log(expenditure) + obesity
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1     27 66.190
## 2     28 86.452 -1   -20.263 8.2656 0.007788 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(lm2, lm4)
```

```
## Analysis of Variance Table
##
## Model 1: life ~ log(expenditure) + alcohol + obesity
## Model 2: life ~ log(expenditure) + alcohol
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1     27 66.190
## 2     28 77.442 -1   -11.252 4.5898 0.04133 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(lm2)
```

```
##
## Call:
## lm(formula = life ~ log(expenditure) + alcohol + obesity)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4763 -0.7807  0.3002  0.9860  2.5276
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      61.09510    4.87083  12.543 8.95e-13 ***
## log(expenditure)  2.87028    0.57045   5.032 2.80e-05 ***
## alcohol          -0.28663    0.09970  -2.875  0.00779 **
## obesity          -0.08730    0.04075  -2.142  0.04133 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.566 on 27 degrees of freedom
## Multiple R-squared:  0.6146, Adjusted R-squared:  0.5718
## F-statistic: 14.35 on 3 and 27 DF,  p-value: 8.763e-06
```

```r
#  UK Life Expectancy
uk_values <- who[who$country == "United Kingdom",]
predictions.PI <- data.frame(predict.lm(lm2,uk_values,
interval = "prediction",
level = 0.95))
predictions.PI
```

```
##        fit      lwr      upr
## 30 78.8962 75.53536 82.25705
```

```r
#  All OCED Countries Life Expectancy
lm2 <- lm(life ~ log(expenditure) + alcohol + obesity)
predictions.PI <- predict(lm2, newdata = who, interval = "prediction", level = 0.95)
predictions.PI <- data.frame(predictions.PI)
predictions.PI
```

```
##         fit      lwr      upr
## 1  79.47243 76.11625 82.82860
## 2  80.18073 76.81964 83.54182
## 3  80.54348 77.25516 83.83181
## 4  80.72073 77.35314 84.08832
## 5  77.04397 73.50520 80.58274
## 6  76.15947 72.66482 79.65411
## 7  80.38865 77.07318 83.70413
## 8  74.99764 71.36850 78.62679
## 9  79.81857 76.55185 83.08528
## 10 79.48297 76.11699 82.84894
## 11 79.50197 76.16295 82.84098
## 12 79.36408 76.06585 82.66231
## 13 77.01695 73.63338 80.40052
## 14 82.00238 78.63446 85.37031
## 15 79.37764 76.00830 82.74698
## 16 80.60837 76.99587 84.22087
## 17 80.89860 77.52796 84.26925
## 18 81.26329 77.77290 84.75369
## 19 76.56369 72.82235 80.30504
## 20 81.14558 77.81618 84.47498
## 21 78.63311 75.29097 81.97525
## 22 82.87175 79.41887 86.32464
## 23 77.07633 73.66717 80.48549
## 24 78.61858 75.31452 81.92265
## 25 78.98396 75.47912 82.48880
## 26 78.07698 74.75769 81.39627
## 27 79.64455 76.37922 82.90989
## 28 81.47123 78.11011 84.83235
## 29 81.69981 78.32148 85.07814
## 30 78.89620 75.53536 82.25705
## 31 81.47629 77.71315 85.23943
```

```r
average_predicted_life <- mean(predictions.PI$fit, na.rm = TRUE)


list(Predictions = predictions.PI, Average_Predicted_Life = average_predicted_life)
```

```
## $Predictions
##         fit      lwr      upr
## 1  79.47243 76.11625 82.82860
## 2  80.18073 76.81964 83.54182
## 3  80.54348 77.25516 83.83181
## 4  80.72073 77.35314 84.08832
## 5  77.04397 73.50520 80.58274
## 6  76.15947 72.66482 79.65411
## 7  80.38865 77.07318 83.70413
## 8  74.99764 71.36850 78.62679
```

```
## 9  79.81857 76.55185 83.08528
## 10 79.48297 76.11699 82.84894
## 11 79.50197 76.16295 82.84098
## 12 79.36408 76.06585 82.66231
## 13 77.01695 73.63338 80.40052
## 14 82.00238 78.63446 85.37031
## 15 79.37764 76.00830 82.74698
## 16 80.60837 76.99587 84.22087
## 17 80.89860 77.52796 84.26925
## 18 81.26329 77.77290 84.75369
## 19 76.56369 72.82235 80.30504
## 20 81.14558 77.81618 84.47498
## 21 78.63311 75.29097 81.97525
## 22 82.87175 79.41887 86.32464
## 23 77.07633 73.66717 80.48549
## 24 78.61858 75.31452 81.92265
## 25 78.98396 75.47912 82.48880
## 26 78.07698 74.75769 81.39627
## 27 79.64455 76.37922 82.90989
## 28 81.47123 78.11011 84.83235
## 29 81.69981 78.32148 85.07814
## 30 78.89620 75.53536 82.25705
## 31 81.47629 77.71315 85.23943
##
## $Average_Predicted_Life
## [1] 79.48387
```