Investigating the performance of UK healthcare in comparison to that of other OECD countries

170183149

20th October 2020

1 Introduction

Recent criticisms of the NHS include the belief that it is under-performing compared to other European healthcare services. Politicians are intent on altering the budget for the NHS to improve what it is seen by some as a health provider inferior to those of the rest of Europe. This investigation aims to examine these claims by using data provided by the World Health Organisation (WHO) in 2011 to find out whether the UK really has a comparatively failing health service, indicated by expenditure, or if lifestyle choices are a bigger factor in life expectancy, which is the statistic this report will be using as an indicator of the nation's health. To get an initial picture of the data, Figure 1 [1] is a table showing the UK health data versus the average health data for other countries in the Organisation for Economic Development and Co-ordination (OECD). The data shows that the United Kingdom has a slightly higher life expectancy than other OECD countries, with higher expenditure on healthcare and less smoking, but also more obesity and more drinking. The variables that are most responsible for the UKs longer life expectancy will be investigated in this report.

	country	life	tobacco	expenditure	alcohol	obesity
30	United Kingdom	80.00000	24.80	3230.000	11.500000	24.00000
1	Other OECD Countries	79.46667	30.02	3008.367	9.826667	16.36667

Figure 1: Life expectancy, expenditure (per person on health), tobacco (percentage of adult users), alcohol (average litres drunk over a year) and obesity (percentage of obese adults) in the UK versus other countries.

2 Results

The data was initially analysed to investigate the relationship with each individual factor and life expectancy resulting in the following scatter diagrams [2]. From the graphs alone it is clear that expenditure has the strongest relationship

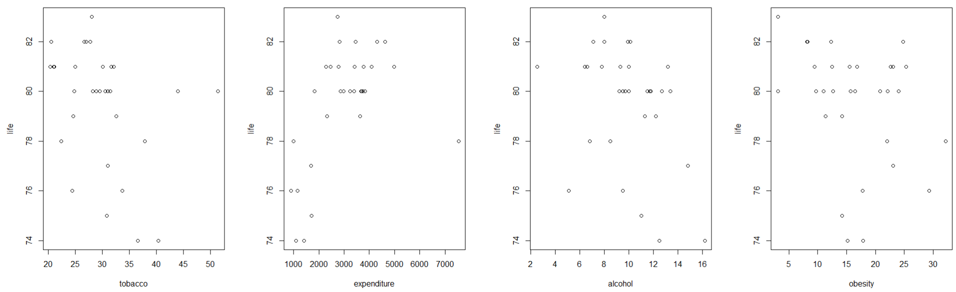

Figure 2: Scatter plots.

with life expectancy and **R** can be used to calculate the exact correlations [3] which, with the variables as listed in Figure 1 in relation to life expectancy, are -0.3457086, 0.5084863, -0.364971 and -0.3005739, respectively.

To decide which variables are important in determining life expectancy, an appropriate model needs to be chosen. The model

$$y_i = \beta_0 + \beta_V V_i + \beta_W W_i + \beta_X X_i + \beta_Z Z_i + \varepsilon_i$$

has been chosen for this report as the most complex model willing to be considered. Let y_i be life expectancy, V be expenditure, W be alcohol, X be tobacco and Z be obesity with $\beta_0, \beta_V, \beta_W, \beta_X$ and β_Z as unknown constants and ε_i as the standard error.

The assumption that $\varepsilon \sim N(0, \sigma^2)$ needs to be checked by fitting [4] the most complex model to the data to obtain the standardised residuals [5] and using them to plot a histogram [6] and QQ plot [7] as in Figure 3. Both plots

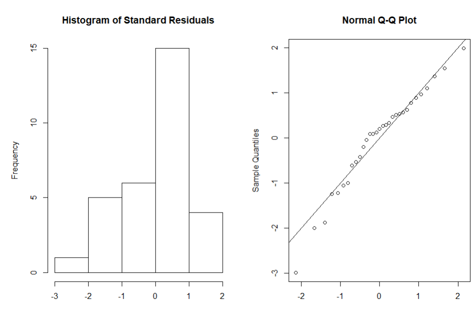

Figure 3: Histogram and QQ plot to check normality.

here suggest normality although the histogram is slightly skewed to the left. To double check, the Kolmogorov-Smirnov goodness of fit test [8] is also performed resulting in the table in Figure 4. As the p-value from this is bigger than 0.05,

```
One-sample Kolmogorov-Smirnov test

data:  evals
D = 0.1488, p-value = 0.4551
alternative hypothesis: two-sided
```

Figure 4: Kolmogorov-Smirnov test

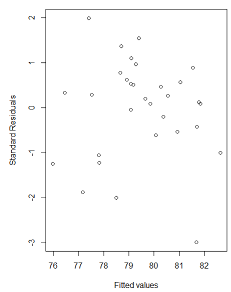

Figure 5: Standardised residuals vs fitted points

which will be the significance level for p-values in this report, the test also implies normality. Looking at the plot [9] in Figure 5 there is no visible pattern, implying that there is constant variance.

The best model is now found by comparing the more complex model to a reduced model, i.e one where one or more coefficient is 0. Starting with the most complex model:

$$y_i = \beta_0 + \beta_V V_i + \beta_W W_i + \beta_X X_i + \beta_Z Z_i + \varepsilon_i,$$

the null hypothesis

$$H_0 : \beta_Z = 0$$

can be tested by comparing the first model to the model

$$y_i = \beta_0 + \beta_V V_i + \beta_W W_i + \beta_X X_i + \varepsilon_i$$

in **R** using an ANOVA test [10] (shown below for every step). β_Z is being tested first as it is the coefficient corresponding to Z, obesity, which has the weakest correlation with life expectancy so is least likely to be included in the model. The F statistic is high and the p-value is low therefore there is significant evidence to reject the null hypothesis. Next we go back to the first model as it has been shown that β_Z does not equal 0 and test the null hypothesis

$$H_0 : \beta_X = 0$$

. The ANOVA test for this doesn't give enough evidence to reject the hypothesis so the model

$$y_i = \beta_0 + \beta_V V_i + \beta_W W_i + \beta_Z Z_i + \varepsilon_i,$$

is a better model. From this model the null hypotheses $H_0 : \beta_W = 0$ and $H_0 : beta_V = 0$ were tested in a similar manner but both were rejected so the best model found is

$$y_i = \beta_0 + \beta_V V_i + \beta_W W_i + \beta_Z Z_i + \varepsilon_i.$$

```
Model 1: life ~ tobacco + expenditure + alcohol
Model 2: life ~ tobacco + expenditure + alcohol + obesity
  Res.Df     RSS Df Sum of Sq      F  Pr(>F)
1     27 108.905
2     26  85.952  1    22.954 6.9434 0.01399

Model 1: life ~ expenditure + alcohol + obesity
Model 2: life ~ tobacco + expenditure + alcohol + obesity
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1     27 91.049
2     26 85.952  1    5.0972 1.5419 0.2254

Model 1: life ~ expenditure + obesity
Model 2: life ~ expenditure + alcohol + obesity
  Res.Df     RSS Df Sum of Sq      F Pr(>F)
1     28 111.206
2     27  91.049  1    20.157 5.9775 0.0213

Model 1: life ~ alcohol + obesity
Model 2: life ~ expenditure + alcohol + obesity
  Res.Df     RSS Df Sum of Sq      F   Pr(>F)
1     28 128.254
2     27  91.049  1    37.205 11.033 0.002576
```

3 Conclusion

The two factors which could've contributed to the UK having a higher life expectancy are tobacco, which is used less than the average and expenditure; more is spent than the average. As tobacco use isn't included in the final model it must be expenditure that is causing UK citizens to live longer lives than their counterparts in OECD countries. In conclusion, the NHS is performing better not worse than other European health providers.

4 Appendix

References

[1]
```
other_countries<-who[-(30),]
uk<-who[30,]
m<-data.frame(country="Other OECD Countries",life=mean(other_countries$life),
tobacco=mean(other_countries$tobacco),expenditure=mean(other_countries$expenditure),
alcohol=mean(other_countries$alcohol),obesity=mean(other_countries$obesity))
a<-rbind(uk,m)
a
```

[2]
```
plot(tobacco, life)
plot(expenditure, life)
plot(alcohol, life)
plot(obesity, life)
```

[3]
```
cor(life,tobacco)
cor(life,expenditure)
cor(life,alcohol)
cor(life,obesity)
```

[4]
```
lm1<-lm(life~tobacco+expenditure+alcohol+obesity, who)
```

[5]
```
evals<-stdres(lm1)
```

[6]
```
hist(evals, xlab="Standard Residuals", main="Histogram of Standard Residuals")
```

[7]
```
qqnorm(evals, xlab ="Standard Residuals")
abline(0,1)
```

[8]
```
ks.test(evals,pnorm,0,1)
```

[9]
```
plot(fitted(lm1),evals, xlab="Fitted values", ylab="Standard Residuals")
```

[10]
```
lm2<-lm(life~tobacco+expenditure+alcohol, who)
lm3<-lm1<-lm(life~expenditure+alcohol+obesity, who)
lm4<-lm(life~expenditure+obesity, who)
anova(lm2, lm1)
anova(lm3, lm1)
anova(lm4, lm3)
anova(lm5,lm3)
```

”No brown M& Ms”