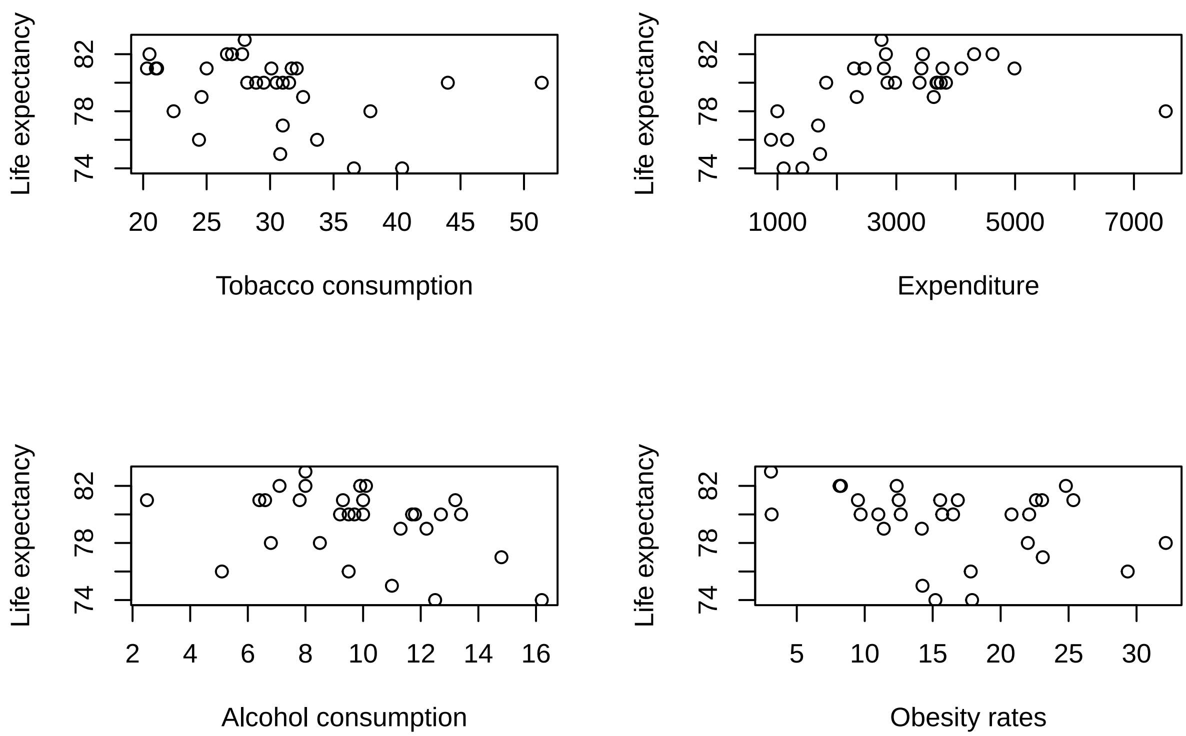
Analysis of UK Healthcare Performance in 2011

In 2011, UK's former Health Secretary Andrew Lansley criticized that UK's previous health-care provision did not get the results compared with other European countries. This report will assess the performance of UK healthcare in 2011 using the data from WHO, which includes life expectancy, expenditure, tobacco consumption, alcohol consumption, and obesity rates in OECD countries. After comparing the 2011 performance of UK healthcare with healthcare in other OECD countries, not enough evidence has been found to support Andrew Lansley's claim.

In this report, the indicator of health experience is set to be life expectancy. There are several variables which may affect health outcomes: expenditure, tobacco consumption, alcohol consumption, and obesity rates. The relationship between those variables and life expectancy in OECD countries other than the UK will be analyzed first. Then UK's data will be assessed to see whether it fits that relation well.

First, to visualize the data points, scatter plots of life expectancy against each independent variable (tobacco consumption, expenditure, alcohol consumption, and obesity rates) are produced below:

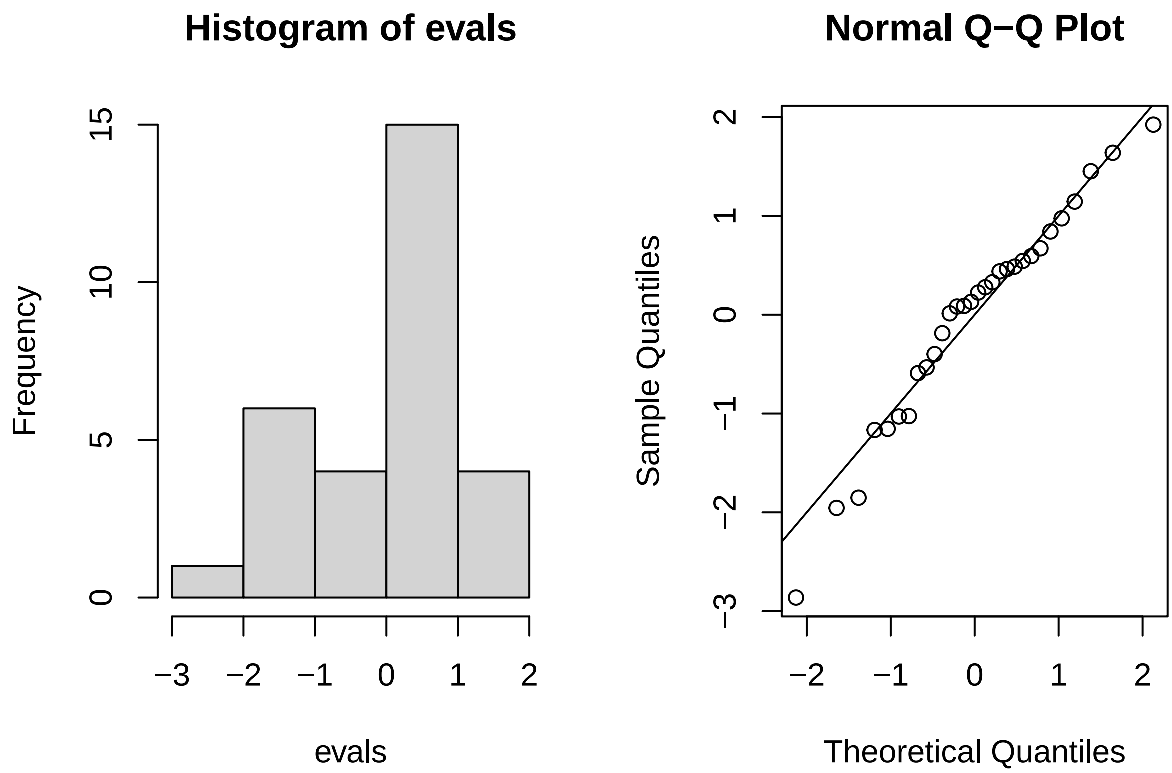


The model being chosen for the dataset is:

$$M_f : y_i = \beta_0 + \beta_1 v_i + \beta_2 w_i + \beta_3 x_i + \beta_4 z_i + \epsilon_i,$$

where y_i is the life expectancy in years, v_i is percentage of tobacco user among adults, w_i is the total expenditure per person on health, x_i is alcohol consumption among adults, and z_i is percentage of adults classified as obese. It is assumed that $\epsilon_i \sim N(0, \sigma^2)$.

To check the assumption $\epsilon_i \sim N(0, \sigma^2)$, a histogram and a Q-Q plot for the standardized residuals of the fitted model are produced below:



In addition, the KS test gives the p-value 0.5636, which is large. These all suggest that the normality assumption is appropriate.

We can then proceed to identify the most suitable model for the data. First, we test the null hypothesis

$$H_0 : \beta_4 = 0.$$

The full model is $M_f : y_i = \beta_0 + \beta_1 v_i + \beta_2 w_i + \beta_3 x_i + \beta_4 z_i + \epsilon_i$, and the reduced model is $M_1 : y_i = \beta_0 + \beta_1 v_i + \beta_2 w_i + \beta_3 x_i + \epsilon_i$. After the F-test is applied, the F-statistic is 7.3535 and the p-value is 0.01193 so there is enough evidence against the null hypothesis. We conclude that $\beta_4 \neq 0$.

Then we test the null hypothesis

$$H_0 : \beta_2 = 0,$$

which claims that the life expectancy is independent of expenditure. The result gives F-statistic 8.5583 and p-value 0.007217 so there is enough evidence against the null hypothesis. We conclude that $\beta_2 \neq 0$.

After that we test the null hypothesis

$$H_0 : \beta_1 = 0.$$

The result gives F-statistic 1.2033 and p-value 0.2831 so there is not enough evidence against the null hypothesis. So there is no evidence to reject the model

$$M_3 : y_i = \beta_0 + \beta_2 w_i + \beta_3 x_i + \beta_4 z_i + \epsilon_i$$

in favor the full model M_f.

However, when we test the null hypothesis

$$H_0 : \beta_3 = 0,$$

the resulting F-statistic is 4.097 and the p-value is 0.05377, which ia quite ambiguous. So we now treat

$$M_3 : y_i = \beta_0 + \beta_2 w_i + \beta_3 x_i + \beta_4 z_i + \epsilon_i$$

as the full model and compare

$$M_4 : y_i = \beta_0 + \beta_2 w_i + \beta_4 z_i + \epsilon_i$$

against M_3. The null hypothesis is still

$$H_0 : \beta_3 = 0.$$

The given F-statistic is 6.4786 and the p-value is 0.01719 so there is enough evidence against the null hypothesis.

Therefore, the most suitable model for the data should be

$$M_3 : y_i = \beta_0 + \beta_2 w_i + \beta_3 x_i + \beta_4 z_i + \epsilon_i.$$

Now we can put the data of the UK into our model to check whether the relationship between life expectancy and other variables in the UK is similar to the relationship in other OECD countries, so that we can know whether the 2011 performance of UK healthcare is close to healthcare in other OECD countries.

Given data about health expenditure, alcohol consumption, and obesity rates in the UK, we can predict the life expectancy in the UK according to our model. The 95% prediction interval for life expectancy in the UK is $[74.22282, 82.13167]$, and the actual life expectancy is 80 which is in the prediction interval. Therefore, there is not enough evidence to show that there exists significant difference between the 2011 performance of UK healthcare and healthcare in other OECD countries.

To conclude, firstly, it has been found that life expectancy depends mostly on health expenditure, alcohol consumption, and obesity rates. In addition, a linear model has been fitted in this report to describe the relationship between the life expectancy and those variables in OECD countries except the UK. Furthermore, given data about health expenditure, alcohol consumption, and obesity rates in the UK, a prediction of life expectancy as well as the prediction interval has been produced, and this prediction fits the real situation well, which suggests that there is not enough evidence to support Andrew Lansley's claim.

Appendix: R commands

```r
library(MASS)
load("MAS223.RData")
# Choose the data except the data in the UK.
who_exUK <- who[who$country != "United Kingdom",]
attach(who_exUK)
```

```r
# Scatter plots of life expectancy against each independent variable.
par(mfrow=c(2,2))
plot(tobacco,life,xlab = "Tobacco consumption",ylab = "Life expectancy")
plot(expenditure,life,xlab = "Expenditure",ylab = "Life expectancy")
plot(alcohol,life,xlab = "Alcohol consumption",ylab = "Life expectancy")
plot(obesity,life,xlab = "Obesity rates",ylab = "Life expectancy")
```

```r
# The full model.
lm_f <- lm(life~tobacco + expenditure + alcohol + obesity)
lm_f
```

```
##
## Call:
## lm(formula = life ~ tobacco + expenditure + alcohol + obesity)
##
## Coefficients:
## (Intercept)      tobacco  expenditure      alcohol      obesity
##  83.7068044   -0.0595233    0.0007318   -0.2545282   -0.1315796
```

```
# The histogram and Q-Q plot for the standardized residuals and KS test.
library(MASS)
par(mfrow=c(1,2))
evals <- stdres(lm_f)
hist(evals)
qqnorm(evals)
abline(0,1)
```

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

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  evals
## D = 0.13872, p-value = 0.5636
## alternative hypothesis: two-sided
```

```
# F-tests.
lm1 <- lm(life~tobacco + expenditure + alcohol)
anova(lm_f,lm1)
```

```
## Analysis of Variance Table
##
## Model 1: life ~ tobacco + expenditure + alcohol + obesity
## Model 2: life ~ tobacco + expenditure + alcohol
##   Res.Df     RSS Df Sum of Sq      F  Pr(>F)
## 1     25  83.969
## 2     26 108.667 -1   -24.699 7.3535 0.01193 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
lm2 <- lm(life~tobacco + alcohol + obesity)
anova(lm_f,lm2)
```

```
## Analysis of Variance Table
##
## Model 1: life ~ tobacco + expenditure + alcohol + obesity
## Model 2: life ~ tobacco + alcohol + obesity
##   Res.Df     RSS Df Sum of Sq      F   Pr(>F)
## 1     25  83.969
## 2     26 112.714 -1   -28.745 8.5583 0.007217 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
lm3 <- lm(life~expenditure + alcohol + obesity)
anova(lm_f,lm3)
```

```
## Analysis of Variance Table
##
```

```
## Model 1: life ~ tobacco + expenditure + alcohol + obesity
## Model 2: life ~ expenditure + alcohol + obesity
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     25 83.969
## 2     26 88.010 -1   -4.0416 1.2033 0.2831
```

```r
lm4 <- lm(life~tobacco + expenditure + obesity)
anova(lm_f,lm4)
```

```
## Analysis of Variance Table
##
## Model 1: life ~ tobacco + expenditure + alcohol + obesity
## Model 2: life ~ tobacco + expenditure + obesity
##   Res.Df    RSS Df Sum of Sq     F  Pr(>F)
## 1     25 83.969
## 2     26 97.729 -1   -13.761 4.097 0.05377 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
lm5 <- lm(life~expenditure + obesity)
anova(lm3,lm5)
```

```
## Analysis of Variance Table
##
## Model 1: life ~ expenditure + alcohol + obesity
## Model 2: life ~ expenditure + obesity
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1     26  88.01
## 2     27 109.94 -1    -21.93 6.4786 0.01719 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Make prediction and compute prediction interval
uk_values <- who[who$country == "United Kingdom",]
predictions.PI <- data.frame(predict.lm(lm3,uk_values,
                                        interval = "prediction",
                                        level = 0.95))
predictions.PI
```

```
##         fit      lwr      upr
## 30 78.17724 74.22282 82.13167
```

```r
uk_values$life
```

```
## [1] 80
```

No brown m&m's