# Does *education*, *sex* and *age* have an impact on *subjective wellbeing*?

This is Exercise 3 from Lab8. Needless to say, you cannot do this exact research question for your assignment, but whichever research question you choose, the process will be similar.

## Introduction

Here I will introduce the issue of “subjective wellbeing”; I can combine it with the ‘literature review’ if I want, and then I would also briefly introduce some findings that we already know from the existing literature that I have reviewed.

## Literature review

Or I can have a separate brief section for the literature review, and then this is where I place my research question in the context of existing literature, summarizing what we already know about the topic based on a few academic sources (around 5 should be enough for this assignment).

## Data and methods

The main research question I have is whether “education” has an impact on someone’s sense of “wellbeing” (satisfaction with their lives). I also want to account for the effect that age and sex may have on wellbeing.

I will be using the UKHLS dataset to answer my research question. In the assignment proper I will say a few more words about this data based on the information available on the UKHLS website, and I’ll include a citation to the original data. I know how to do all this because I have read everything on Canvas and I’ve attended the labs this semester.

The variable that are useful for answering my research question are: - scghq1\_dv (“subjective wellbeing”), the outcome variable I want to explain; - hiqual\_dv (“educational qualification”) - age\_dv (“age”) - sex (“biological sex”)

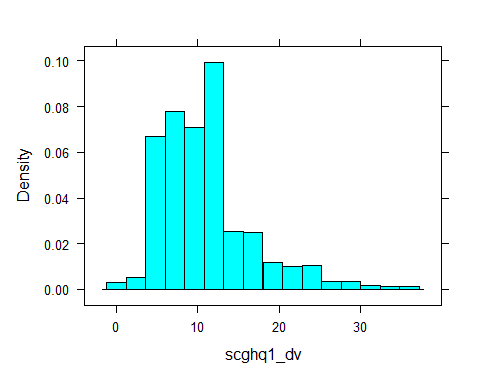
The variables look like this:

# First, I'll load the dataset  
  
ukhls <- readRDS("ukhls\_w8.rds")

# Then I'll do a number of summary statistics for my variables  
# First, summary stats for scghq1\_dv:  
  
summary(ukhls$scghq1\_dv)

Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
 0.00 7.00 10.00 11.12 13.00 36.00 3351

# Maybe also a histogram:  
  
histogram( ~ scghq1\_dv, data = ukhls)



I can briefly explain here what the above summary tells me.

# Then a frequency table for hiqual\_dv  
  
ukhls %>% sjmisc::frq(hiqual\_dv)

Highest qualification, UKHLS & BHPS samples (hiqual\_dv) <categorical>   
# total N=39293 valid N=38435 mean=3.07 sd=1.65  
  
Value | N | Raw % | Valid % | Cum. %  
------------------------------------------------------  
Degree | 10058 | 25.60 | 26.17 | 26.17  
Other higher degree | 4659 | 11.86 | 12.12 | 38.29  
A-level etc | 8326 | 21.19 | 21.66 | 59.95  
GCSE etc | 7726 | 19.66 | 20.10 | 80.05  
Other qualification | 3388 | 8.62 | 8.81 | 88.87  
No qualification | 4278 | 10.89 | 11.13 | 100.00  
<NA> | 858 | 2.18 | <NA> | <NA>

I can briefly explain what we find from the table.

# Then a frequency table for sex  
  
ukhls %>% sjmisc::frq(sex)

sex (sex) <categorical>   
# total N=39293 valid N=39293 mean=1.54 sd=0.50  
  
Value | N | Raw % | Valid % | Cum. %  
-----------------------------------------  
male | 17925 | 45.62 | 45.62 | 45.62  
female | 21368 | 54.38 | 54.38 | 100.00  
<NA> | 0 | 0.00 | <NA> | <NA>

I can briefly explain what we find from the table.

# Then summary stats for age  
  
summary(ukhls$age\_dv)

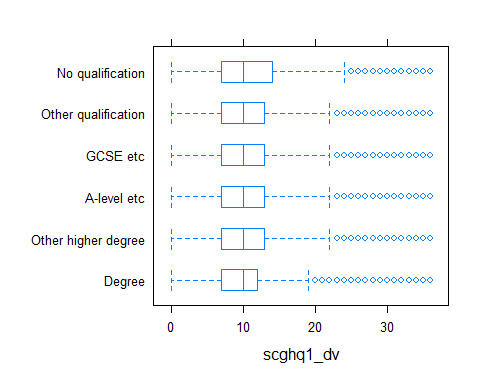
Min. 1st Qu. Median Mean 3rd Qu. Max.   
 16.00 34.00 49.00 48.77 63.00 102.00

I can briefly explain here what the above summary tells me.

To answer my research question, I will first explore visually the relationship between my outcome variable (“subjective wellbeing”) and the main predictor variable whose effect I am interested in (“educational qualification”). Then, because my outcome variable is measured on a “numeric” scale, I will use a linear regression model to estimate the impact of education, age and sex on my outcome of interest.

## Results

# First, let's look at the relationship between wellbeing and education  
# I have a numeric and a factor variable, so the best method to visualise their relationship is using a boxplot  
  
bwplot(hiqual\_dv ~ scghq1\_dv, data = ukhls, pch = "|")



We see in this boxplot that there are some minor differences in the median level of distress felt by respondents with different educational qualifications, but we cannot see any large obvious differences. In the assignment propers I’ll include a bit more elaborate interpretation of this graph.

# I'll fit my regression model now and ask for some nice summary statistics using the summ() function from the jtools package  
  
  
assignment\_reg <- lm(scghq1\_dv ~ age\_dv + sex + hiqual\_dv, data = ukhls)  
  
jtools::summ(assignment\_reg, digits = 3)

MODEL INFO:  
Observations: 35267 (4026 missing obs. deleted)  
Dependent Variable: scghq1\_dv  
Type: OLS linear regression   
  
MODEL FIT:  
F(7,35259) = 80.694, p = 0.000  
R² = 0.016  
Adj. R² = 0.016   
  
Standard errors: OLS  
------------------------------------------------------------------  
 Est. S.E. t val. p  
------------------------------- -------- ------- --------- -------  
(Intercept) 10.913 0.103 106.431 0.000  
age\_dv -0.016 0.002 -9.493 0.000  
sexfemale 1.142 0.059 19.260 0.000  
hiqual\_dvOther higher 0.206 0.101 2.043 0.041  
degree   
hiqual\_dvA-level etc 0.292 0.085 3.437 0.001  
hiqual\_dvGCSE etc 0.567 0.087 6.542 0.000  
hiqual\_dvOther 0.697 0.117 5.963 0.000  
qualification   
hiqual\_dvNo qualification 1.085 0.112 9.703 0.000  
------------------------------------------------------------------

Now I have my results, so I will write down my interpretation here. This is where I need to be very careful and think deeply, go back to the lab worksheets and maybe even the assigned readings, because I need to give accurate interpretations of the results.

## Discussion and conclusions

This is where I take the broader picture of what my analysis and outputs have shown and connect it back to the existing literature, stating how my findings relate to what we already know.

## References

Here I will include a reference list for the literature that I cited above.