# RECSM: Quantitative Methods in Social Research

Day 2 - 03 07 2025

Burak Sonmez

---

**Please skim through all of the instructions before getting started.**

## Problem set

**Step 1**: Open a new script and save it as day2. Clear your workspace.

**Step 2**: Load the Asylum dataset from yesterday.

**Step 3**: Explore the potential correlations between perception bias towards Syrian asylum seekers and other variables in the dataset

**Step 4**: Estimate a linear regression model that explains perception bias towards Syrian asylum seekers using only one independent variable. Please justify your choice. Try to come up with a reasonable argument (e.g.) why more of x should lead to more/less of y).

**Step 5**: Plot a scatterplot of that relationship and add the best fit line to the plot.

**Step 6**: Interpret the regression output and try to imagine that you are communicating these results to your neighbours who are statistically illiterate

**Step 7**: Reiterate the process from step 3. Estimate another model – you can choose a different independent variable on the same dependent variable.

**Step 8**: Interpret theese new regression output.

**Step 9**: Please compare the two models and explain which one you would choose.

**Step 10**: Produce a regression table with both models next to each other in some text document, using 'texreg'.

**Step 11**: Add some additional explanatory variables (no more than 4) to your model.

**Step 12**: Interpret the output of the new model in relation to the previous model.

**Step 13**: Check whether the model assumptions are violated (e.g. collinearity issues; normality assumption; heteroskedasticity). To do so, you can use 'performance' package.

**Step 14**: Add an interaction term to the model (e.g. age x news). First, summarise the model and interpret the output of the new model. Then plot the relationship between perception bias and age moderated by tabloid readership.

## Solutions for the problem set

```r
##Instal the packages first if those packages are not installed before
library(ggplot2)
library(texreg)
```

```
## Version:  1.39.3
## Date:     2023-11-09
## Author:   Philip Leifeld (University of Essex)
##
## Consider submitting praise using the praise or praise_interactive functions.
## Please cite the JSS article in your publications -- see citation("texreg").
```

```r
library(corrplot)
```

```
## Warning: package 'corrplot' was built under R version 4.3.3
```

```
## corrplot 0.95 loaded
```

```r
library(performance)
```

```
## Warning: package 'performance' was built under R version 4.3.3
```

```r
library(RColorBrewer)
library(effects)
```

```
## Loading required package: carData
```

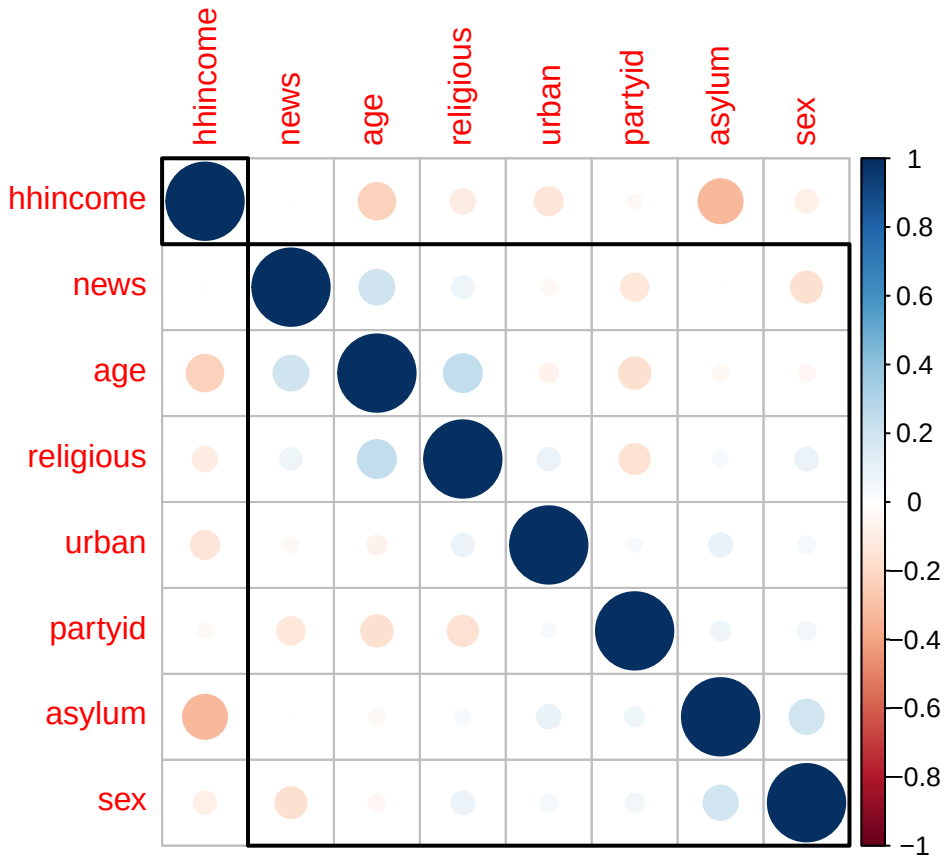
```
## lattice theme set by effectsTheme()
## See ?effectsTheme for details.
```

```r
#Remove objects from the environment
rm(list=ls())
#Set your working directory
setwd("~/Downloads/RECSM workshop")

load("asylum_data.RData")
names(asy)
```

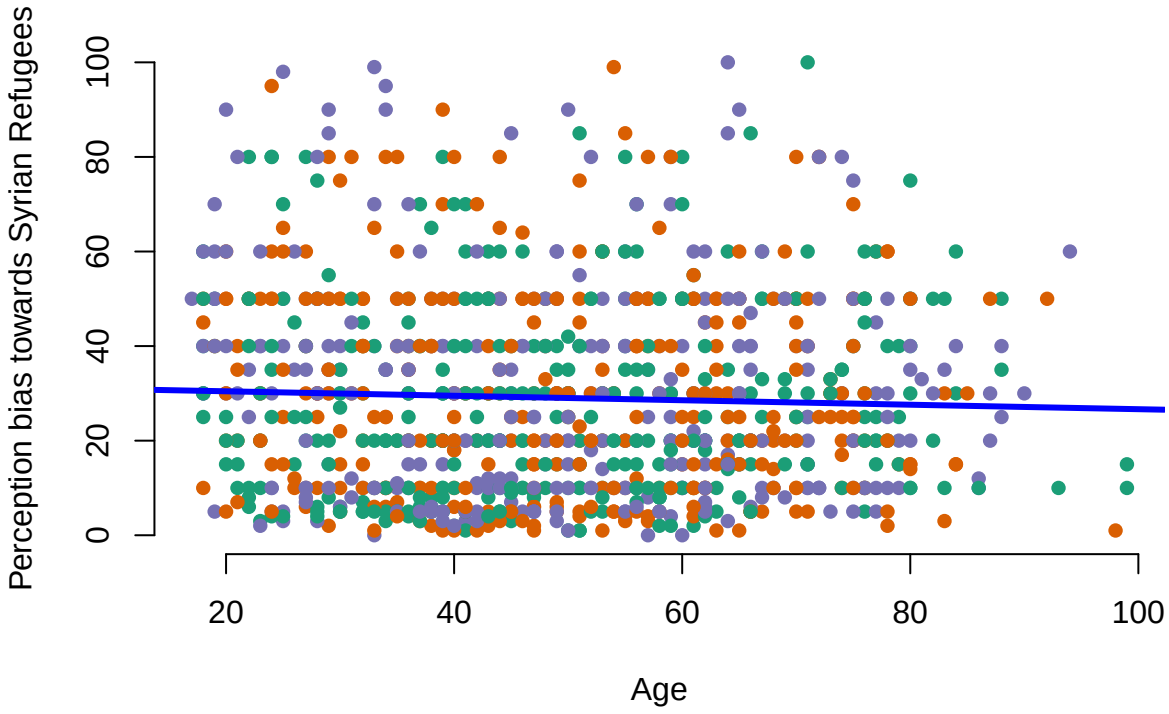
```
## [1] "asylum"    "sex"       "age"       "news"      "religious" "urban"
## [7] "hhincome"  "partyid"
```

```r
#Visualise correlations
cor <- cor(asy)
corrplot(cor, order = "hclust", addrect = 2)
```




```
#Estimating a bivariate linear regression model
model1 <- lm(asylum ~ age, data = asy)

#Plotting the relationship and adding the fit line
plot(
  asylum ~ age, data = asy,
  xlab = "Age",
  ylab = "Perception bias towards Syrian Refugees",
  frame.plot = FALSE,
  pch = 16,
  col = brewer.pal(n = 3, name = "Dark2"))
abline(model1, lwd = 3,
       col = "blue")
```




```
##Interpret the regression output in a table
screenreg(model1)
```

```
##
## ========================
##               Model 1
## ------------------------
## (Intercept)    31.38 ***
##                (1.95)
## age            -0.05
##                (0.04)
## ------------------------
## R^2             0.00
## Adj. R^2        0.00
## Num. obs.    1049
## ========================
## *** p < 0.001; ** p < 0.01; * p < 0.05
```

```
#Estimating another bivariate linear regression model
#Don't forget to convert it into a factor
asy$partyid <- factor(asy$partyid, labels = c("Tories", "Labour", "SNP", "Greens", "Ukip", "BNP", "othe

model2 <- lm(asylum ~ partyid, data = asy)

screenreg( list(model1, model2))
```

```
##
## ======================================
##                Model 1      Model 2
## --------------------------------------
## (Intercept)      31.38 ***    27.73 ***
##                  (1.95)       (1.24)
## age              -0.05
##                  (0.04)
## partyidLabour                  0.17
##                               (1.76)
## partyidSNP                     6.02
##                               (5.38)
## partyidGreens                 -2.82
##                               (4.54)
## partyidUkip                   -3.44
##                               (3.96)
## partyidBNP                    14.46 ***
##                               (3.90)
## partyidother                   2.43
##                               (1.64)
## --------------------------------------
## R^2               0.00         0.02
## Adj. R^2          0.00         0.01
## Num. obs.      1049         1049
## ======================================
## *** p < 0.001; ** p < 0.01; * p < 0.05
```

```r
#Multiple linear regression
asy$news <- factor(asy$news, labels = c("Don't read", "Read"))
asy$sex <- factor(asy$sex, labels = c("Male", "Female"))

model3 <- lm(asylum ~ age + sex + news+ partyid, data = asy)

screenreg(list(model1, model2, model3))
```

```
##
## ==================================================
##                Model 1      Model 2      Model 3
## --------------------------------------------------
## (Intercept)      31.38 ***    27.73 ***    23.16 ***
##                  (1.95)       (1.24)       (2.47)
## age              -0.05                     -0.02
##                  (0.04)                    (0.04)
## partyidLabour                  0.17        -0.07
##                               (1.76)       (1.73)
## partyidSNP                     6.02         6.19
##                               (5.38)       (5.27)
## partyidGreens                 -2.82        -3.74
##                               (4.54)       (4.46)
## partyidUkip                   -3.44        -2.85
##                               (3.96)       (3.88)
## partyidBNP                    14.46 ***    16.75 ***
##                               (3.90)       (3.88)
## partyidother                   2.43         1.79
```

```
##                                (1.64)       (1.63)
## sexFemale                                    9.02 ***
##                                             (1.30)
## newsRead                                     2.29
##                                             (1.33)
## ---------------------------------------------------
## R^2               0.00         0.02         0.06
## Adj. R^2          0.00         0.01         0.05
## Num. obs.      1049         1049         1049
## ===================================================
## *** p < 0.001; ** p < 0.01; * p < 0.05
```

```
#Model assumptions check and interpret the plots
#check_model(model1)
#check_model(model2)
#check_model(model3)
#Interaction model
int.model <- lm(asylum ~ sex + age*news+ partyid, data = asy)
summary(int.model)
```

```
##
## Call:
## lm(formula = asylum ~ sex + age * news + partyid, data = asy)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -43.041 -15.371  -3.585  13.924  74.798
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   22.89093    2.96064   7.732 2.50e-14 ***
## sexFemale      9.02636    1.30025   6.942 6.80e-12 ***
## age           -0.01734    0.05140  -0.337    0.736
## newsRead       2.89760    3.98360   0.727    0.467
## partyidLabour -0.08737    1.73288  -0.050    0.960
## partyidSNP     6.17272    5.27638   1.170    0.242
## partyidGreens -3.74211    4.45978  -0.839    0.402
## partyidUkip   -2.90341    3.89597  -0.745    0.456
## partyidBNP    16.78212    3.88733   4.317 1.73e-05 ***
## partyidother   1.78620    1.63580   1.092    0.275
## age:newsRead  -0.01204    0.07433  -0.162    0.871
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.5 on 1038 degrees of freedom
## Multiple R-squared:  0.0622, Adjusted R-squared:  0.05317
## F-statistic: 6.885 on 10 and 1038 DF,  p-value: 1.739e-10
```

```
#Plotting
#plot(allEffects(int.model)$age)
```