lab6

Jason Seda, Riyesh Nath

2025-10-31

```r
library(haven)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```
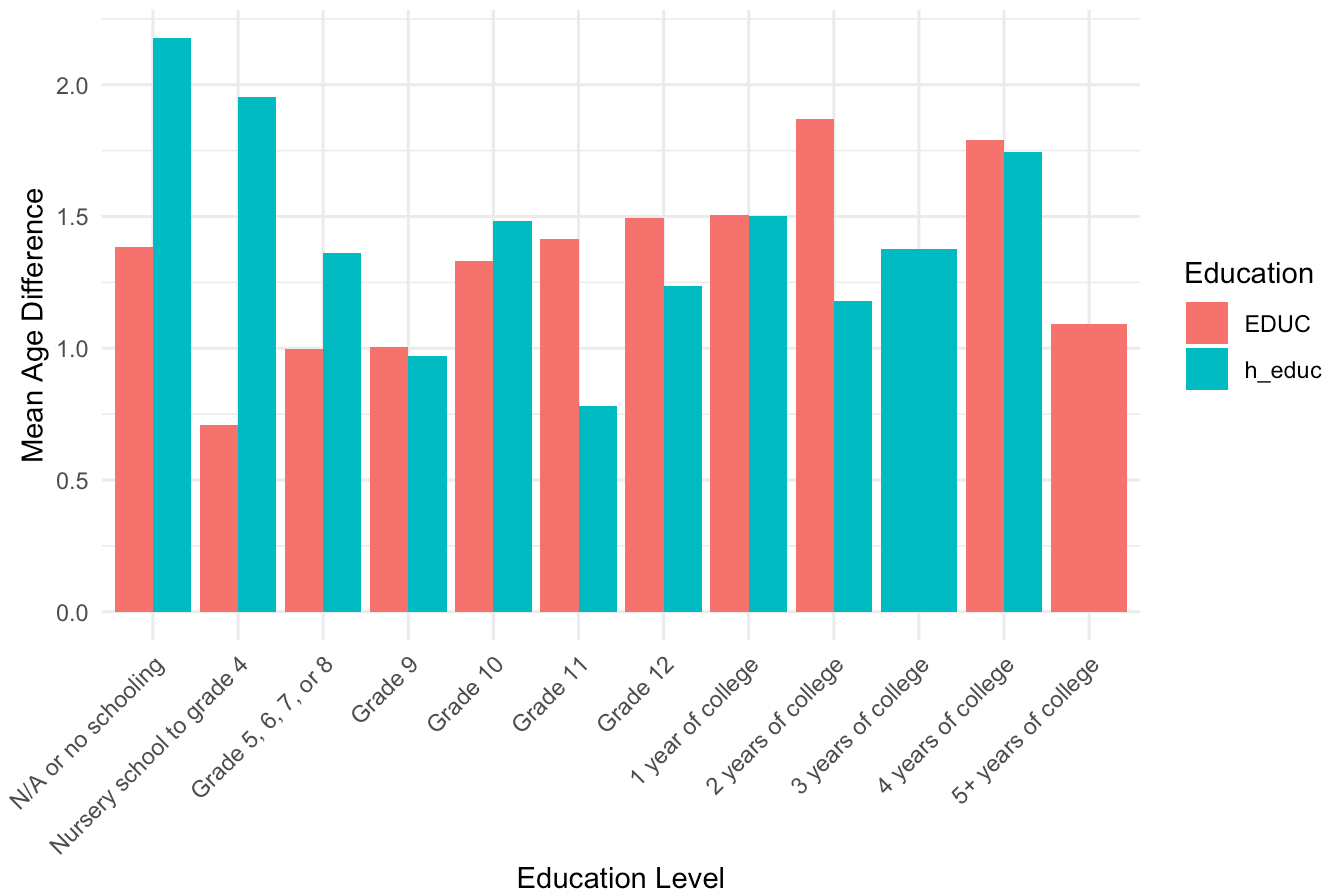
```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(forcats)
library(tibble)
library(ggplot2)
library(class)
library(tidyr)
setwd("~/Documents/Masters/Econometrics/data")# your directory structure will be different
load("ACS_2021_couples.RData")
```

This lab investigates whether educational attainment and immigrant status influence the average spousal age difference in U.S. married couples using data from the 2021 American Community Survey (ACS).
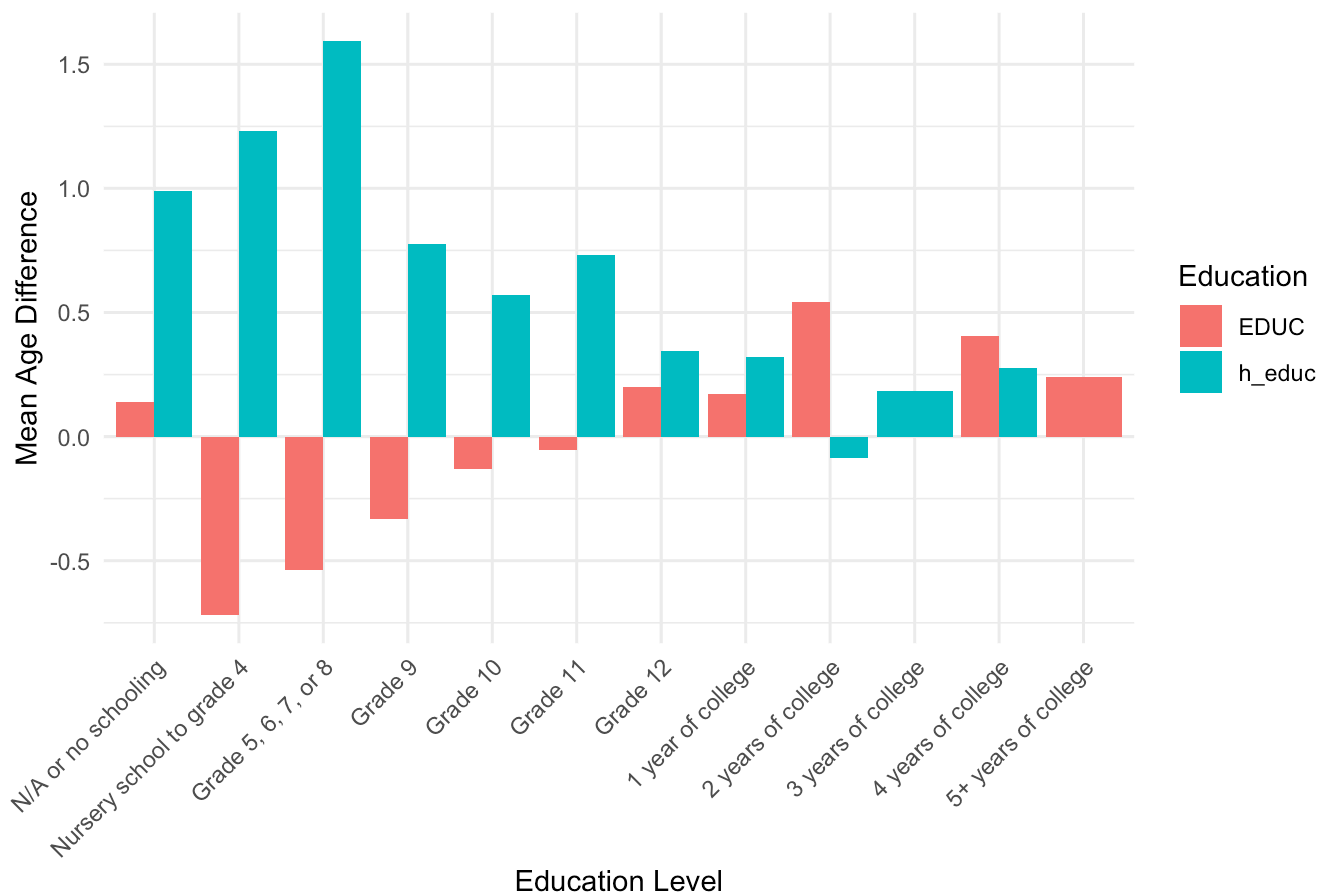
```
## `summarise()` has grouped output by 'Education_Type'. You can override using
## the `.groups` argument.
```

Mean Age Difference by Education (Respondent vs Spouse) Not Born in US

Education
- EDUC
- h_educ

Education Level

Mean Age Difference

```
## `summarise()` has grouped output by 'Education_Type'. You can override using
## the `.groups` argument.
```

Mean Age Difference by Education (Respondent vs Spouse) Not Born in US


```
anova_EDUC = aov(Age_DIFF ~ EDUC, data = acs2021_couples_spouse)
summary(anova_EDUC)
```

```
##                  Df    Sum Sq Mean Sq F value Pr(>F)
## EDUC             10     14903  1490.3   50.98 <2e-16 ***
## Residuals    647443 18925824    29.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova_H_EDUC = aov(Age_DIFF ~ h_educ, data = acs2021_couples_spouse)
summary(anova_H_EDUC)
```

```
##                  Df    Sum Sq Mean Sq F value Pr(>F)
## h_educ           10     37748    3775   129.3 <2e-16 ***
## Residuals    647443 18902979      29
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```
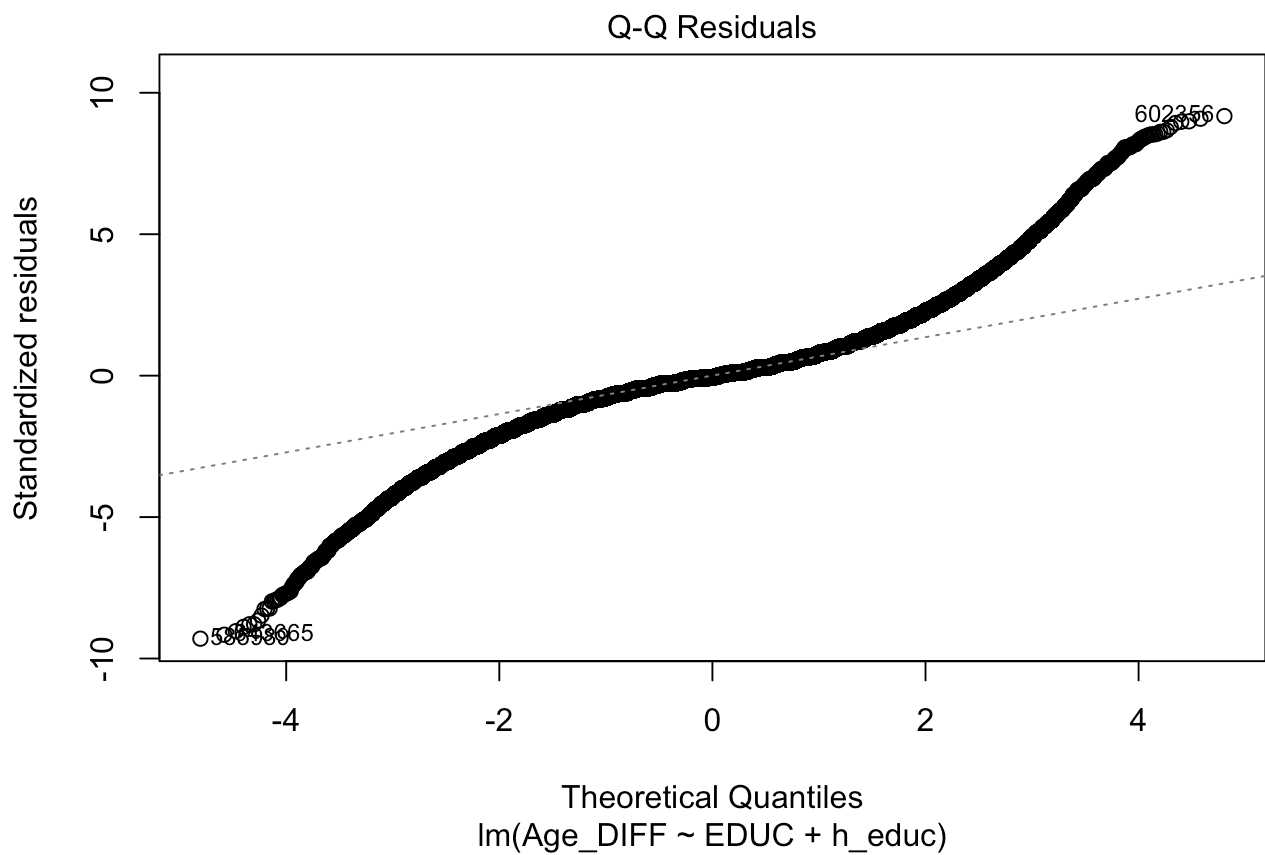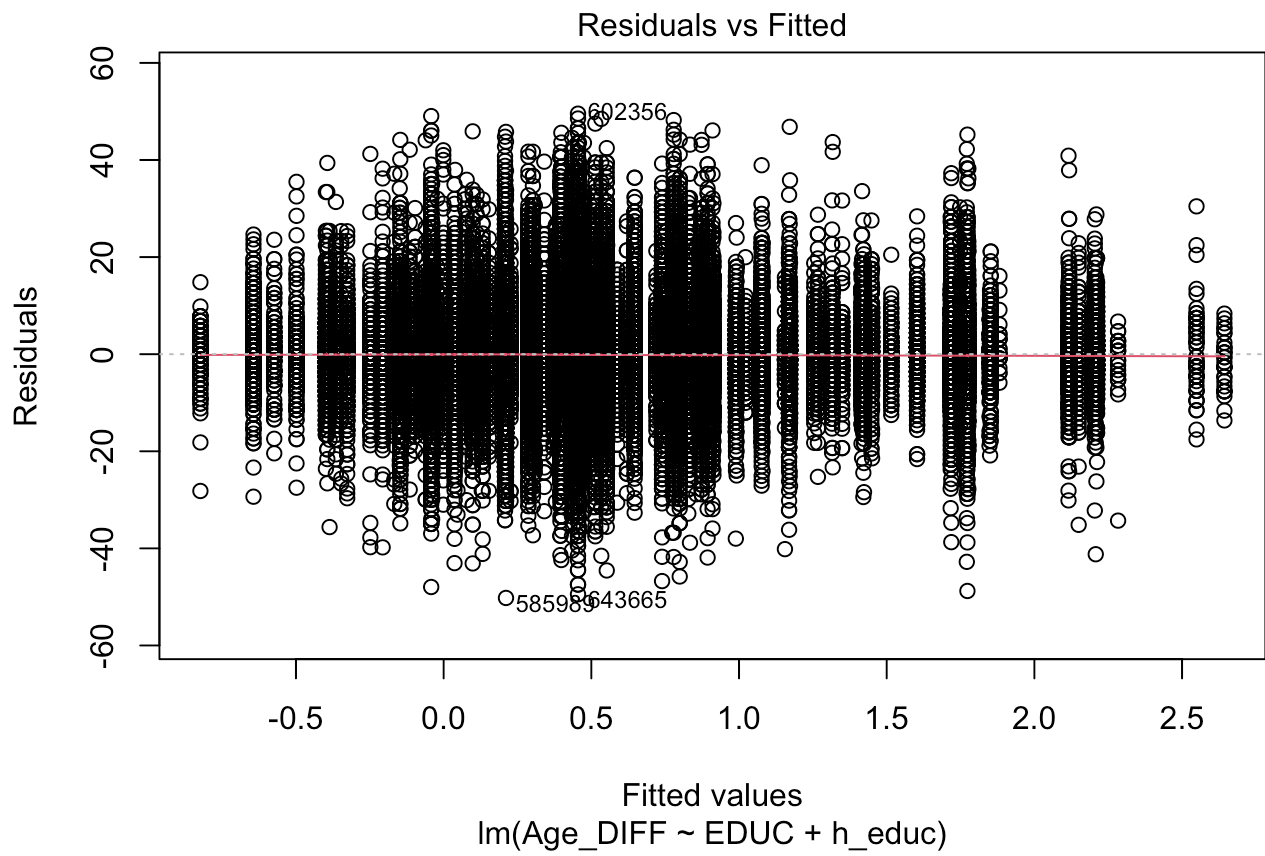
```
lm <- lm(data=acs2021_couples_spouse, formula = Age_DIFF ~ EDUC + h_educ)
summary(lm)
```

```
## 
## Call:
## lm(formula = Age_DIFF ~ EDUC + h_educ, data = acs2021_couples_spouse)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max
## -50.211  -2.458  -0.302   2.487  49.545
## 
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      1.718678   0.068027  25.265  < 2e-16 ***
## EDUCNursery school to grade 4   -0.726798   0.126449  -5.748 9.05e-09 ***
## EDUCGrade 5, 6, 7, or 8         -0.547083   0.078530  -6.967 3.25e-12 ***
## EDUCGrade 9                     -0.401408   0.093975  -4.271 1.94e-05 ***
## EDUCGrade 10                    -0.301892   0.087687  -3.443 0.000576 ***
## EDUCGrade 11                    -0.268148   0.085453  -3.138 0.001701 **
## EDUCGrade 12                     0.054614   0.059916   0.912 0.362022
## EDUC1 year of college            0.057823   0.062427   0.926 0.354320
## EDUC2 years of college           0.493412   0.063312   7.793 6.53e-15 ***
## EDUC4 years of college           0.398959   0.061386   6.499 8.08e-11 ***
## EDUC5+ years of college          0.133497   0.062465   2.137 0.032587 *
## h_educNursery school to grade 4  0.431267   0.144792   2.979 0.002897 **
## h_educGrade 5, 6, 7, or 8       -0.002089   0.089567  -0.023 0.981389
## h_educGrade 9                   -0.696701   0.106580  -6.537 6.29e-11 ***
## h_educGrade 10                  -0.862998   0.103519  -8.337  < 2e-16 ***
## h_educGrade 11                  -0.944959   0.099656  -9.482  < 2e-16 ***
## h_educGrade 12                  -1.318417   0.068225 -19.325  < 2e-16 ***
## h_educ1 year of college         -1.378064   0.070098 -19.659  < 2e-16 ***
## h_educ2 years of college        -1.815429   0.070898 -25.606  < 2e-16 ***
## h_educ3 years of college        -1.565444   0.069167 -22.633  < 2e-16 ***
## h_educ4 years of college        -1.338748   0.069817 -19.175  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 5.4 on 647433 degrees of freedom
## Multiple R-squared:  0.003319,   Adjusted R-squared:  0.003288
## F-statistic: 107.8 on 20 and 647433 DF,  p-value: < 2.2e-16
```

```
plot(lm)
```

Residuals vs Fitted

Residuals

602356

585985643665

Fitted values
lm(Age_DIFF ~ EDUC + h_educ)

Q-Q Residuals

Standardized residuals

602356

585985643665

Theoretical Quantiles
lm(Age_DIFF ~ EDUC + h_educ)

Scale-Location

√|Standardized residuals|

58598.02866

Fitted values
lm(Age_DIFF ~ EDUC + h_educ)

Residuals vs Leverage

Standardized residuals

201790 516672
97985

Cook's distance

Leverage
lm(Age_DIFF ~ EDUC + h_educ)

```
lm <- lm(data=acs2021_couples_spouse, formula = Age_DIFF ~ EDUC + h_educ + born_in_USsta
te)
summary(lm)
```

```
##
## Call:
## lm(formula = Age_DIFF ~ EDUC + h_educ + born_in_USstate, data = acs2021_couples_spous
e)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -50.049  -2.448  -0.236   2.533  49.202
##
## Coefficients:
##                                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)                              1.973147   0.067896  29.061  < 2e-16
## EDUCNursery school to grade 4           -0.851896   0.126024  -6.760 1.38e-11
## EDUCGrade 5, 6, 7, or 8                 -0.503437   0.078260  -6.433 1.25e-10
## EDUCGrade 9                             -0.204854   0.093694  -2.186   0.0288
## EDUCGrade 10                             0.099435   0.087587   1.135   0.2563
## EDUCGrade 11                             0.180536   0.085417   2.114   0.0346
## EDUCGrade 12                             0.517194   0.060103   8.605  < 2e-16
## EDUC1 year of college                    0.550161   0.062640   8.783  < 2e-16
## EDUC2 years of college                   0.968474   0.063487  15.255  < 2e-16
## EDUC4 years of college                   0.844911   0.061532  13.731  < 2e-16
## EDUC5+ years of college                  0.549516   0.062555   8.784  < 2e-16
## h_educNursery school to grade 4          0.277228   0.144308   1.921   0.0547
## h_educGrade 5, 6, 7, or 8                0.004627   0.089256   0.052   0.9587
## h_educGrade 9                           -0.632922   0.106215  -5.959 2.54e-09
## h_educGrade 10                          -0.606270   0.103230  -5.873 4.28e-09
## h_educGrade 11                          -0.659622   0.099401  -6.636 3.23e-11
## h_educGrade 12                          -1.005766   0.068147 -14.759  < 2e-16
## h_educ1 year of college                 -1.055993   0.070019 -15.082  < 2e-16
## h_educ2 years of college                -1.493614   0.070814 -21.092  < 2e-16
## h_educ3 years of college                -1.275980   0.069061 -18.476  < 2e-16
## h_educ4 years of college                -1.096392   0.069669 -15.737  < 2e-16
## born_in_USstateborn in a state in the US -1.198556   0.017833 -67.211  < 2e-16
##
## (Intercept)                             ***
## EDUCNursery school to grade 4           ***
## EDUCGrade 5, 6, 7, or 8                 ***
## EDUCGrade 9                             *
## EDUCGrade 10
## EDUCGrade 11                            *
## EDUCGrade 12                            ***
## EDUC1 year of college                   ***
## EDUC2 years of college                  ***
## EDUC4 years of college                  ***
## EDUC5+ years of college                 ***
## h_educNursery school to grade 4         .
## h_educGrade 5, 6, 7, or 8
## h_educGrade 9                           ***
## h_educGrade 10                          ***
## h_educGrade 11                          ***
## h_educGrade 12                          ***
## h_educ1 year of college                 ***
```

```
## h_educ2 years of college                 ***
## h_educ3 years of college                 ***
## h_educ4 years of college                 ***
## born_in_USstateborn in a state in the US ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.381 on 647432 degrees of freedom
## Multiple R-squared:  0.01022,    Adjusted R-squared:  0.01019
## F-statistic: 318.5 on 21 and 647432 DF,  p-value: < 2.2e-16
```

```
plot(lm)
```

Residuals vs Fitted

20740

5859843665

Residuals

Fitted values
lm(Age_DIFF ~ EDUC + h_educ + born_in_USstate)

Q-Q Residuals

20740

5859843665

Standardized residuals

Theoretical Quantiles
lm(Age_DIFF ~ EDUC + h_educ + born_in_USstate)

Scale-Location

20740 2585968 43665

√|Standardized residuals|

Fitted values
lm(Age_DIFF ~ EDUC + h_educ + born_in_USstate)

Residuals vs Leverage

Standardized residuals

Cook's distance

59869
97985
516672

Leverage
lm(Age_DIFF ~ EDUC + h_educ + born_in_USstate)