Econometrics
Homework 6
10/31/25

Professor: Kevin Foster
Student: Ege Yanik
Group member: Aqib Shahzad

Data Table

	M 1	M 2
(Intercept)	-0.756***	-0.650***
	(0.007)	(0.008)
educ_diff	0.007**	-0.038***
	(0.003)	(0.004)
I(educ_diff^2)		-0.014***
		(0.000)
I(educ_diff^3)		0.000***
		(0.000)
Num.Obs.	776064	776064
R2	0.000	0.003
R2 Adj.	0.000	0.003
AIC	5085935.4	5083663.0
BIC	5085970.1	5083720.8
Log.Lik.	-2542964.709	-2541826.519
F	7.212	762.314
RMSE	6.41	6.40

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

```
library(ggplot2)
library(tidyverse)
library(haven)
#Lab6
   acs2021_couples$age_diff <- acs2021_couples$AGE -
                acs2021_couples$h_age
```

```
summary(acs2021_couples$age_diff[(acs2021_couples$SEX
                                  ==
          "Female")&(acs2021_couples$h_sex == "Male")])
      summary(acs2021_couples$age_diff[(acs2021_couples$SEX ==
           "Male")&(acs2021_couples$h_sex == "Female")])
      summary(acs2021_couples$age_diff[(acs2021_couples$SEX ==
            "Male")&(acs2021_couples$h_sex == "Male")])
      summary(acs2021_couples$age_diff[(acs2021_couples$SEX ==
          "Female")&(acs2021_couples$h_sex == "Female")])
        summary(acs2021_couples$AGE[(acs2021_couples$SEX ==
           "Female")&(acs2021_couples$h_sex == "Male")])
       summary(acs2021_couples$h_age[(acs2021_couples$SEX ==
           "Female")&(acs2021_couples$h_sex == "Male")])
          acs2021_couples$educ_numeric <-
 fct_recode(acs2021_couples$EDUC, "0" = "N/A or no
                    schooling",
                          "2" = "Nursery school to grade 4",
                            "6.5" = "Grade 5, 6, 7, or 8",
                                  "9" = "Grade 9",
                                 "10" = "Grade 10",
                                 "11" = "Grade 11",
                                 "12" = "Grade 12",
                              "13" = "1 year of college",
                             "14" = "2 years of college",
                             "15" = "3 years of college",
                             "16" = "4 years of college",
                            "17" = "5+ years of college")
acs2021_couples$educ_numeric <-
as.numeric(levels(acs2021_couples$educ_numeric))[acs2021_coupl
     es$educ_numeric] acs2021_couples$h_educ_numeric <-
     fct_recode(acs2021_couples$h_educ, "0" = "N/A or no
                          schooling",
                    "2" = "Nursery school to grade 4",
                    "6.5" = "Grade 5, 6, 7, or 8",
                    "9" = "Grade 9",
                    "10" = "Grade 10",
                    "11" = "Grade 11",
                    "12" = "Grade 12",
                    "13" = "1 year of college",
                    "14" = "2 years of college",
                    "15" = "3 years of college",
                    "16" = "4 years of college",
```

```
                    "17" = "5+ years of college")
acs2021_couples$h_educ_numeric <-
as.numeric(levels(acs2021_couples$h_educ_numeric))[acs2021_couples
 $h_educ_numeric] acs_subgroup <- acs2021_couples %>% filter((AGE
             >= 25) & (AGE <= 55) & (LABFORCE == 2)
                                    & (WKSWORK2 > 4)
                                  & (UHRSWORK >= 35) )
                               library(AER)
                     m1 <- lm(age_diff ~ educ_diff,
                          data = acs2021_couples)
  m2 <- lm(age_diff ~ educ_diff + I(educ_diff^2) + I(educ_diff^3),
                          data = acs2021_couples)
                      coeftest(m1, vcov = vcovHC)
                                    ##
                     ## t test of coefficients:
                                    ##
                ## Estimate Std. Error t value Pr(>|t|)
          ## (Intercept) -0.7556547 0.0072612 -104.0678 <2e-16 ***
             ## educ_diff 0.0072887 0.0041757 1.7455 0.0809 .
                                 ## ---
  ## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
                    waldtest(m1, m2, vcov = vcovHC)
                              ## Wald test
                                    ##
                    ## Model 1: age_diff ~ educ_diff
  ## Model 2: age_diff ~ educ_diff + I(educ_diff^2) + I(educ_diff^3)
                          ## Res.Df Df F Pr(>F)
                               ## 1 776062
                    ## 2 776060 2 368.8 < 2.2e-16 ***
                                 ## ---
  ## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
                          library(modelsummary)
                             models <- list(
       "M 1" = lm(age_diff ~ educ_diff, data = acs2021_couples),
         "M 2" = lm(age_diff ~ educ_diff + I(educ_diff^2) +
 I(educ_diff^3), data = acs2021_couples) ) modelsummary(models,
                          stars = TRUE)
```

3.https://pmc.ncbi.nlm.nih.gov/articles/PMC7838214/#:~:text=Abstract,kicks%20were%20not%20 statistically%20significant.

I chose this article because I am planning on comparing the best offensive teams and

the best defensive teams within the top five leagues. This study uses match and goal data from the top five European leagues, the Premier League, La Liga, Bundesliga, Serie A, and Ligue 1, covering the 2009/10 to 2018/19 seasons. The authors classify each goal by factors like build-up type, assist, shot location, and timing. Using the Kruskal Wallis test, they compare scoring patterns across leagues. They find that La Liga teams score more from throw-ins, while the Bundesliga has more counterattack goals. The main goal is to see how goal-scoring styles differ between leagues. The authors mention that the data is publicly available.

https://www.researchgate.net/publication/395734322_Are_Counterattacks_More_Effective_than_Positional_Attacks_in_Soccer_A_Comparative_Analysis_of_Influencing_Factors

I like this article because it shows how different teams score. This paper compares how effective counterattacks and positional attacks are in professional soccer. It uses data from about 5,700 matches from 2017 to 2022 across La Liga, Serie A, and the Premier League, taken from Instat, which is a paid soccer analytics platform. The authors use Anova to compare attack efficiency and multiple regression to find which factors like attack zone, shot type, or possession length predict success. They find that counterattacks are much more efficient than positional attacks. The study aims to understand which styles of play lead to better scoring chances.