Chapter 7 Lab

Allen Church

November 20, 2019

## Preparation

require(knitr)  
require(haven)  
require(car)  
require(AER)  
require(lm.beta)  
  
opts\_chunk$set(echo = TRUE)  
options(digits = 6)  
  
setwd("C:/Users/WB537822/Desktop/stats1")  
load("Chapter7\_lab.RData")

#### (a) Estimate a model for women predicting wages in 1996 as a function of height in 1985 and 1981, siblings and esteem from 1980. Use standardized coefficients and report results. How do the t statistics compare to t statistics in an unstandardized model? (You don’t need to report the unstandardized results.)

#Create two subsetted datasets for women and men, respectively  
dtawomen <- dta[dta$male == 0,]  
dtamen <- dta[dta$male == 1,]  
  
results\_female\_standardized <- lm(scale(dtawomen$wage96) ~ scale(dtawomen$height81) + scale(dtawomen$height85) + scale(dtawomen$siblings) + scale(dtawomen$esteem80))  
  
summary(results\_female\_standardized)

##   
## Call:  
## lm(formula = scale(dtawomen$wage96) ~ scale(dtawomen$height81) +   
## scale(dtawomen$height85) + scale(dtawomen$siblings) + scale(dtawomen$esteem80))  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.55 -0.21 -0.10 0.04 43.39   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -0.00174 0.01835 -0.09 0.9246   
## scale(dtawomen$height81) -0.00947 0.04034 -0.23 0.8144   
## scale(dtawomen$height85) 0.02811 0.03978 0.71 0.4799   
## scale(dtawomen$siblings) -0.05066 0.01901 -2.66 0.0077 \*\*  
## scale(dtawomen$esteem80) 0.05817 0.01914 3.04 0.0024 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.02 on 3097 degrees of freedom  
## (3181 observations deleted due to missingness)  
## Multiple R-squared: 0.00668, Adjusted R-squared: 0.0054   
## F-statistic: 5.21 on 4 and 3097 DF, p-value: 0.000353

Estimating unstandardized model below:

results\_female\_unstandardized <- lm(dtawomen$wage96 ~ dtawomen$height81 + dtawomen$height85 + dtawomen$siblings + dtawomen$esteem80)  
summary(results\_female\_unstandardized)

##   
## Call:  
## lm(formula = dtawomen$wage96 ~ dtawomen$height81 + dtawomen$height85 +   
## dtawomen$siblings + dtawomen$esteem80)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -19.3 -7.3 -3.5 1.3 1516.7   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -15.766 16.374 -0.96 0.3357   
## dtawomen$height81 -0.121 0.516 -0.23 0.8144   
## dtawomen$height85 0.354 0.500 0.71 0.4799   
## dtawomen$siblings -0.662 0.249 -2.66 0.0077 \*\*  
## dtawomen$esteem80 0.711 0.234 3.04 0.0024 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 35.6 on 3097 degrees of freedom  
## (3181 observations deleted due to missingness)  
## Multiple R-squared: 0.00668, Adjusted R-squared: 0.0054   
## F-statistic: 5.21 on 4 and 3097 DF, p-value: 0.000353

As seen in the standardized vs. unstandardized models above, the t statistics are the same (excluding the Intercept), at 0.23, 0.71, 2.66, and 3.04 for height81, height85, siblings, and esteem80, respectively. This reinforces that model fit is the same between standardized and unstandardized variables. The standardized model suggests that the only statistically significant coefficients in the model are siblings and esteem, with t statistics of 2.66 and 3.04, respectively. These results suggest that an increase in 1 standard deviation of siblings is associated with a decrease in -0.05066 standard deviations of wage. Also, the model suggests that an increase in 1 standard deviation of esteem is associated with an increase of 0.711 standard deviations of wage.

#### (b) Add several covariates (your choice) to the above model, including dummy variables for race/ethnicity. Use standardized coefficients as appropriate to each variable and briefly discuss effects, focusing on effect of race relative to the esteem variable (as an example of a continuous covariate).

# Estimate regression models using scale command for continuous variables  
results2\_female <- lm(scale(dtawomen$wage96) ~ scale(dtawomen$height85) + scale(dtawomen$height81) + scale(dtawomen$esteem80) + scale(dtawomen$siblings) + dtawomen$black + dtawomen$hispanic + scale(dtawomen$daded79)+ scale(dtawomen$momed79))  
summary(results2\_female)

##   
## Call:  
## lm(formula = scale(dtawomen$wage96) ~ scale(dtawomen$height85) +   
## scale(dtawomen$height81) + scale(dtawomen$esteem80) + scale(dtawomen$siblings) +   
## dtawomen$black + dtawomen$hispanic + scale(dtawomen$daded79) +   
## scale(dtawomen$momed79))  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.57 -0.22 -0.11 0.03 43.39   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -0.00896 0.03047 -0.29 0.769   
## scale(dtawomen$height85) 0.03439 0.04827 0.71 0.476   
## scale(dtawomen$height81) -0.02070 0.04941 -0.42 0.675   
## scale(dtawomen$esteem80) 0.04414 0.02289 1.93 0.054 .  
## scale(dtawomen$siblings) -0.03685 0.02495 -1.48 0.140   
## dtawomen$black 0.01032 0.05356 0.19 0.847   
## dtawomen$hispanic 0.07275 0.06699 1.09 0.278   
## scale(dtawomen$daded79) 0.03297 0.02940 1.12 0.262   
## scale(dtawomen$momed79) 0.03454 0.03011 1.15 0.251   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.1 on 2613 degrees of freedom  
## (3661 observations deleted due to missingness)  
## Multiple R-squared: 0.00747, Adjusted R-squared: 0.00443   
## F-statistic: 2.46 on 8 and 2613 DF, p-value: 0.0119

The model above suggests that, among coefficients included, esteem is the only one that is statistically signicicant. Compared to the model in (a) where an increase in 1 standard deviation of esteem was associated with an increase of 0.711 standard deviations of wage, the above model suggests that an increase in 1 standard deviation of esteem was associated with an increase of 0.04414 standard deviations of wage. The t statistics associated with the Black and Hispanic variables are 0.19 and 1.09, and associated coefficients are 0.01032 and 0.07275, respectively.

#### (c) Models with wages often have logged variables in order to be able to provide results in percentage terms rather than absolute dollars. Estimate a log-linear model for women. To keep things simple, use only *height85*, *height81*, *esteem80*, *black* and *siblings* as the covariates.

#Convert 0 values to NA  
dta$wage96.nona <- dta$wage96  
dta$wage96.nona[dta$wage96==0] <- NA  
  
results3\_female <- lm(log(wage96.nona) ~ height85 + height81 + esteem80 + black + siblings, dta[dta$male == 0,])  
summary(results3\_female)

##   
## Call:  
## lm(formula = log(wage96.nona) ~ height85 + height81 + esteem80 +   
## black + siblings, data = dta[dta$male == 0, ])  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -5.004 -0.380 0.042 0.443 5.051   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -0.12558 0.36314 -0.35 0.73   
## height85 0.01376 0.01125 1.22 0.22   
## height81 0.00688 0.01160 0.59 0.55   
## esteem80 0.05108 0.00517 9.89 < 2e-16 \*\*\*  
## black -0.16770 0.03169 -5.29 1.3e-07 \*\*\*  
## siblings -0.03121 0.00562 -5.56 3.0e-08 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.777 on 3014 degrees of freedom  
## (3263 observations deleted due to missingness)  
## Multiple R-squared: 0.0664, Adjusted R-squared: 0.0648   
## F-statistic: 42.9 on 5 and 3014 DF, p-value: <2e-16

The log-linear model above suggests that a 1 unit increase in esteem is associated with a 5% increase in wages for women, with a t statistic of 9.89. Also, the model suggests that being an African American woman is associated with a 16% decrease in wages, with a t statistic of 5.29. Lastly, the model suggests that an additional sibling is associated with a 3% decrease in wages for women, with a t statistic of 5.56.

#### (d) Starting with the above model (for women only), create a model in which the effect of siblings is potentially non-linear via a quadratic equation. Discuss the results and note the effect of siblings in general term and for specific cases when *siblings* equals 1 and when *siblings* equals 5. (For fun, estimate the same model for men. You don’t need to report or discuss those results.)

dta$siblings\_squared <- dta$siblings^2  
  
results4\_female <- lm(log(wage96.nona) ~ height85 + height81 + esteem80 + black + siblings + siblings\_squared, data=dta[dta$male==0,])  
  
summary(results4\_female)

##   
## Call:  
## lm(formula = log(wage96.nona) ~ height85 + height81 + esteem80 +   
## black + siblings + siblings\_squared, data = dta[dta$male ==   
## 0, ])  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4.975 -0.384 0.046 0.446 4.987   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -0.06965 0.36354 -0.19 0.848   
## height85 0.01348 0.01124 1.20 0.230   
## height81 0.00728 0.01159 0.63 0.530   
## esteem80 0.05105 0.00516 9.89 < 2e-16 \*\*\*  
## black -0.16698 0.03167 -5.27 1.4e-07 \*\*\*  
## siblings -0.06512 0.01484 -4.39 1.2e-05 \*\*\*  
## siblings\_squared 0.00307 0.00124 2.47 0.014 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.776 on 3013 degrees of freedom  
## (3263 observations deleted due to missingness)  
## Multiple R-squared: 0.0683, Adjusted R-squared: 0.0664   
## F-statistic: 36.8 on 6 and 3013 DF, p-value: <2e-16

The above model suggests that an additional sibling, in general, is associated with a 6% drop in wages for women, with a t statistic of 4.39. In specific cases, when we add one sibling, this is associated with a 5% drop in wages for women (using the equation ), or . Additionally, five siblings is associated with a 3% drop in wages for women, using the previous equation, or .

Estimating above sibling model for men:

dta$siblings\_squared <- dta$siblings^2  
  
results4\_male <- lm(log(wage96.nona) ~ height85 + height81 + esteem80 + black + siblings + siblings\_squared, data=dta[dta$male==1,])

#### (e) Estimate a model for women only in which the dependent variable is log of wages and the independent variables are *momed79*, *daded79*, *height85*, *height81*, *black* and *hispanic*. Test the null hypothesis that the effect of mother’s education is the same as the effect of father’s education. Report unrestricted and restricted models and then show the calculation of the F-statistic and explain the results.

#Unrestricted model  
results\_female5 <- lm(log(wage96.nona) ~ momed79 + daded79 + height85 + height81 + black + hispanic, data=dta[dta$male==0,])  
  
#Calculate F-statistic using car package  
library(car)  
linearHypothesis(results\_female5, c("momed79 = daded79"))

## Linear hypothesis test  
##   
## Hypothesis:  
## momed79 - daded79 = 0  
##   
## Model 1: restricted model  
## Model 2: log(wage96.nona) ~ momed79 + daded79 + height85 + height81 +   
## black + hispanic  
##   
## Res.Df RSS Df Sum of Sq F Pr(>F)  
## 1 2578 1545   
## 2 2577 1545 1 0.1424 0.238 0.626

#Restricting model by imposing the null that momed79 = daded79  
dta$MomDadEd = dta$momed79 + dta$daded79  
  
results\_female5\_restrict1 <- lm(log(wage96.nona) ~ MomDadEd + height85 + height81 + black + hispanic, data = dta[dta$male == 0,])  
  
#calculate F stat  
F.stat.top = ((summary(results\_female5)$r.squared - summary(results\_female5\_restrict1)$r.squared)/1)  
F.stat.bottom = ((1-summary(results\_female5)$r.squared)/(summary(results\_female5)$df[2]))  
F.stat = F.stat.top/F.stat.bottom  
  
#Critical value for F-test  
qf(1-0.05, df1=1, df2= summary(results\_female5)$df[2])

## [1] 3.84507

Conducting an test above produces an statistic of 0.238, which is based on values from estimating an unrestricted model and estimating a restricted model. The unrestricted model is the full model that includes logged wages as the dependent variable and momed79, daded79, height85, height81, black, and hispanic as independent variables. The restricted model imposes the restriction that the null hypothesis is true - in this case, the null posits that the mother’s education carries the same effect as the father’s education. The statistic above of 0.238 is lower than the critical value from the test of 3.84507, meaning that we cannot reject the null. Lastly, the p value above of 0.626 indicates that we cannot reject the null, as it suggests a 0.626 probability of observing these results under the null.