Brandt Assignment 3 SOCI709

Kate Brandt

February 5, 2019

### Setup

# Dummy variables and interaction terms - Lecture D  
  
setwd("C:/Users/kbran/OneDrive/Documents/R/soci709/")  
# load necessary packages  
library(foreign)  
library(dplyr)  
  
  
#import data  
morg = read.dta("./data/morg07\_small\_1.dta", convert.factors = F)

### 1. Recode race variable to create categories

# recode race  
# Create recode column  
morg$re = morg$race  
  
# recoding categories into dummy variables  
morg$re[morg$race > 4] = 5  
morg$re[morg$hisp == 1] = 6  
morg$re\_r = as.factor(morg$re)  
morg$re\_r = recode\_factor(morg$re\_r,  
   
 `1` = 'white',  
   
 `2` = 'black',  
   
 `3` = 'am indian',  
   
 `4` = 'asian',  
   
 `5` = 'other',  
   
 `6` = 'hispanic')  
  
  
#set reference group to white  
morg = within(morg, re\_r <- relevel(re\_r, ref = 'white'))

### *Skip question 2*

### 3. Models

#### (a) Table of average wage by race

# By RACE  
aggregate(wage\_re ~ re\_r, morg, mean)

re\_r wage\_re  
1 white 18.40674  
2 black 14.93366  
3 am indian 13.70536  
4 asian 20.56796  
5 other 15.37197  
6 hispanic 13.04043

#### Model 1

# Create model for RACE gaps, Excl:white  
mod\_1 <- lm(morg$wage\_re~morg$re\_r)

Model 1 Results  
===============================================  
 Dependent variable:   
 ---------------------------  
 wage\_re   
-----------------------------------------------  
re\_rblack -3.473\*\*\*   
 (0.130)   
   
re\_ram indian -4.701\*\*\*   
 (0.393)   
   
re\_rasian 2.161\*\*\*   
 (0.184)   
   
re\_rother -3.035\*\*\*   
 (0.257)   
   
re\_rhispanic -5.366\*\*\*   
 (0.107)   
   
Constant 18.407\*\*\*   
 (0.040)   
   
-----------------------------------------------  
Observations 112,318   
R2 0.029   
Adjusted R2 0.029   
Residual Std. Error 11.540 (df = 112312)   
F Statistic 681.832\*\*\* (df = 5; 112312)  
===============================================  
Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

#### Model 2

# set referece grouo for SEX to MALE  
morg$re2 = morg$sex  
morg$re\_sex = as.factor(morg$re2)  
morg$re\_sex = recode\_factor(morg$re\_sex,   
 `1` = "female",  
 `2` = "male")  
morg = within(morg, re\_sex <- relevel(re\_sex, ref = "female"))  
# By SEX  
aggregate(wage\_re ~ re\_sex, morg, mean)  
  
# Create model for RACE and SEX, Excl: {white, male}  
mod\_2 <- lm(morg$wage\_re ~   
 morg$re\_r +   
 morg$re\_sex)

Model 2 Results  
=================================================  
 Dependent variable:   
 -----------------------------  
 wage\_re   
-------------------------------------------------  
re\_rblack -3.189\*\*\*   
 (0.128)   
   
re\_ram indian -4.580\*\*\*   
 (0.388)   
   
re\_rasian 2.095\*\*\*   
 (0.181)   
   
re\_rother -3.020\*\*\*   
 (0.254)   
   
re\_rhispanic -5.608\*\*\*   
 (0.106)   
   
re\_sexmale -3.677\*\*\*   
 (0.068)   
   
Constant 20.252\*\*\*   
 (0.052)   
   
-------------------------------------------------  
Observations 112,318   
R2 0.054   
Adjusted R2 0.054   
Residual Std. Error 11.393 (df = 112311)   
F Statistic 1,068.443\*\*\* (df = 6; 112311)  
=================================================  
Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

#### Model 2b

# Create model for RACE, SEX, and RACExSEX interaction, Excl: {white, male}  
mod\_2b <- lm(morg$wage\_re ~   
 morg$re\_r +   
 morg$re\_sex +   
 morg$re\_r:morg$re\_sex)

Model 2b Results  
================================================  
 Dependent variable:   
 ----------------------------  
 wage\_re   
------------------------------------------------  
re\_rblack -4.979\*\*\*   
 (0.196)   
   
re\_ram indian -5.842\*\*\*   
 (0.568)   
   
re\_rasian 2.103\*\*\*   
 (0.252)   
   
re\_rother -3.907\*\*\*   
 (0.361)   
   
re\_rhispanic -7.016\*\*\*   
 (0.142)   
   
re\_sexmale -4.338\*\*\*   
 (0.079)   
   
re\_sexmale 3.179\*\*\*   
 (0.259)   
   
re\_sexmale 2.400\*\*\*   
 (0.777)   
   
re\_sexmale -0.042   
 (0.362)   
   
re\_sexmale 1.759\*\*\*   
 (0.507)   
   
re\_sexmale 3.129\*\*\*   
 (0.213)   
   
Constant 20.584\*\*\*   
 (0.056)   
   
------------------------------------------------  
Observations 112,318   
R2 0.057   
Adjusted R2 0.057   
Residual Std. Error 11.376 (df = 112306)   
F Statistic 615.771\*\*\* (df = 11; 112306)  
================================================  
Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

#### Model 3

# Create model for RACE, SEX, RACExSEX interaction, AGE, and AGE^2. Excl: {white, male}   
# Generate AGE^2 variable  
morg$age2 = (morg$age)^2  
mod\_3 <- lm(morg$wage\_re ~   
 morg$re\_r +   
 morg$re\_sex +  
 morg$age +  
 morg$age2)

Model 3 Results  
=================================================  
 Dependent variable:   
 -----------------------------  
 wage\_re   
-------------------------------------------------  
re\_rblack -3.139\*\*\*   
 (0.123)   
   
re\_ram indian -4.659\*\*\*   
 (0.370)   
   
re\_rasian 2.039\*\*\*   
 (0.173)   
   
re\_rother -1.899\*\*\*   
 (0.242)   
   
re\_rhispanic -4.774\*\*\*   
 (0.102)   
   
re\_sexmale -3.667\*\*\*   
 (0.065)   
   
age 1.155\*\*\*   
 (0.014)   
   
age2 -0.012\*\*\*   
 (0.0002)   
   
Constant -5.437\*\*\*   
 (0.275)   
   
-------------------------------------------------  
Observations 112,318   
R2 0.139   
Adjusted R2 0.139   
Residual Std. Error 10.866 (df = 112309)   
F Statistic 2,275.518\*\*\* (df = 8; 112309)  
=================================================  
Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

#### Model 4

# Create model for RACE, SEX, RACExSEX interaction, AGE, AGE^2, EDUCATION. Excl: {white, male}   
mod\_4 <- lm(morg$wage\_re ~   
 morg$re\_r +   
 morg$re\_sex +  
 morg$age +  
 morg$age2 +  
 morg$edyrs)

Model 4 Results  
=================================================  
 Dependent variable:   
 -----------------------------  
 wage\_re   
-------------------------------------------------  
re\_rblack -2.170\*\*\*   
 (0.114)   
   
re\_ram indian -3.177\*\*\*   
 (0.343)   
   
re\_rasian 1.068\*\*\*   
 (0.160)   
   
re\_rother -1.202\*\*\*   
 (0.224)   
   
re\_rhispanic -0.427\*\*\*   
 (0.099)   
   
re\_sexmale -4.028\*\*\*   
 (0.060)   
   
age 0.888\*\*\*   
 (0.013)   
   
age2 -0.009\*\*\*   
 (0.0002)   
   
edyrs 2.072\*\*\*   
 (0.015)   
   
Constant -27.292\*\*\*   
 (0.300)   
   
-------------------------------------------------  
Observations 112,318   
R2 0.263   
Adjusted R2 0.263   
Residual Std. Error 10.058 (df = 112308)   
F Statistic 4,448.806\*\*\* (df = 9; 112308)  
=================================================  
Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

#### Model 5

# Create model which builds on mod\_4 to add interaction term for RACExSEX  
mod\_5 <- lm(morg$wage\_re ~   
 morg$re\_r +   
 morg$re\_sex +  
 morg$age +  
 morg$age2 +   
 morg$edyrs +   
 morg$re\_r:morg$re\_sex)

Model 5 Results  
==================================================  
 Dependent variable:   
 ------------------------------  
 wage\_re   
--------------------------------------------------  
re\_rblack -3.841\*\*\*   
 (0.173)   
   
re\_ram indian -3.993\*\*\*   
 (0.502)   
   
re\_rasian 0.615\*\*\*   
 (0.223)   
   
re\_rother -1.948\*\*\*   
 (0.319)   
   
re\_rhispanic -0.896\*\*\*   
 (0.132)   
   
re\_sexmale -4.440\*\*\*   
 (0.070)   
   
age 0.887\*\*\*   
 (0.013)   
   
age2 -0.009\*\*\*   
 (0.0002)   
   
edyrs 2.067\*\*\*   
 (0.015)   
   
re\_sexmale 2.937\*\*\*   
 (0.229)   
   
re\_sexmale 1.545\*\*   
 (0.686)   
   
re\_sexmale 0.926\*\*\*   
 (0.320)   
   
re\_sexmale 1.475\*\*\*   
 (0.448)   
   
re\_sexmale 0.990\*\*\*   
 (0.189)   
   
Constant -27.009\*\*\*   
 (0.301)   
   
--------------------------------------------------  
Observations 112,318   
R2 0.264   
Adjusted R2 0.264   
Residual Std. Error 10.049 (df = 112303)   
F Statistic 2,878.317\*\*\* (df = 14; 112303)  
==================================================  
Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01