Problem 3. *This problem may be done in either R or Stata. Load the data set asianwhiteMen.dta from Canvas. This data set is from the 2021 American Commu- nity Survey and contains wage data for Asian and white (non-Hispanic) men between the ages of 25 and 55 who have exactly a Bachelor's degree, work at least 30 hours a week and at least 48 weeks per year.*

a. Create a table of summary statistics of the data. Include the mean and median age of the sample (variable: age), the percent of the sample that is Asian (variable: asian), the mean and median wage (variable: lwage), and the percent of the sample that is US born (variable: usborn). Note: The wage variable is log wages.

Summary Statistics of the Sample

Mean_age	Median_age	Percent_Asian	Mean_lnwage	Median_lnwage	Percent_USBorn
39.22758	39	11.757	11.34827	11.35041	87.62068

Figure 3: Summary Statistics of Selected Data

HW5_Q3

QilinZhou

2023-11-17

```
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
```

```
library(knitr)
```

```
setwd("/Users/qilinzhou/Desktop")
data <- read_dta("asianwhiteMen.dta")

summary_stats <- data %>%
  summarise(
    Mean_age = mean(age, na.rm = TRUE),
    Median_age = median(age, na.rm = TRUE),
    Percent_Asian = mean(asian, na.rm = TRUE) * 100,
    Mean_lnwage = mean(lwage, na.rm = TRUE),
    Median_lnwage = median(lwage, na.rm = TRUE),
    Percent_USBorn = mean(usborn, na.rm = TRUE) * 100
  )
kable(summary_stats, caption = "Summary Statistics of the Sample")
```

Table 1: Summary Statistics of the Sample

Mean_age	Median_age	Percent_Asian	Mean_lnwage	Median_lnwage	Percent_USBorn
39.22758	39	11.757	11.34827	11.35041	87.62068

Figure 4: R Code for Q3a

b. Create a table of summary statistics of the data for Asian men and white men separately. Include the mean age of the sample (variable: age), the mean wage (variable: lwage), and the percent of the sample that is US born (variable: usborn). Note: The wage variable is log wages.

Summary Statistics of Asian Men in the Sample		
Mean_age	**Mean_lnwage**	**Percent_USBorn**
39.0138	11.37966	32.3428

Summary Statistics of White Men in the Sample		
Mean_age	**Mean_lnwage**	**Percent_USBorn**
39.25606	11.34408	94.98559

Figure 5: Summary Statistics of Selected Data for Asian and White Men

```r
# Summary statistics for Asian men
summary_asian_men <- data %>%
  filter(asian == 1) %>%
  summarise(
    Mean_age = mean(age, na.rm = TRUE),
    Mean_lnwage = mean(lwage, na.rm = TRUE),
    Percent_USBorn = mean(usborn, na.rm = TRUE) * 100
  )
```

1

```r
# Summary statistics for White men
summary_white_men <- data %>%
  filter(asian == 0) %>%
  summarise(
    Mean_age = mean(age, na.rm = TRUE),
    Mean_lnwage = mean(lwage, na.rm = TRUE),
    Percent_USBorn = mean(usborn, na.rm = TRUE) * 100
  )

# Tables
kable(summary_asian_men, caption = "Summary Statistics of Asian Men in the Sample")
```

Table 2: Summary Statistics of Asian Men in the Sample

Mean_age	Mean_lnwage	Percent_USBorn
39.0138	11.37966	32.3428

```r
kable(summary_white_men, caption = "Summary Statistics of White Men in the Sample")
```

Table 3: Summary Statistics of White Men in the Sample

Mean_age	Mean_lnwage	Percent_USBorn
39.25606	11.34408	94.98559

Figure 6: R Code for Q3b

c. Using OLS, estimate the effect of being Asian versus white on log wages. Run 3 models: 1) with age as a control, 2) with age and age squared (agesq) as controls and 3) a fully saturated regression model.

```r
library(stargazer)
```

```
##
## Please cite as:
##  Hlavac, Marek (2022). stargazer: Well-Formatted Regression and Summary Statistics Tables.
##  R package version 5.2.3. https://CRAN.R-project.org/package=stargazer
```

```r
# Model 1: Control for age
model1 <- lm(lwage ~ asian + age, data = data)

# Model 2: Control for age and age squared
model2 <- lm(lwage ~ asian + age + agesq, data = data)

# Model 3: Fully saturated model
data$age_factor <- factor(data$age)
model3 <- lm(lwage ~ asian + age_factor, data = data)

stargazer(model1, model2, type = "text")
```

```
##
## ============================================================================
##                                        Dependent variable:
##                        -----------------------------------------------------
##                                               lwage
##                                    (1)                        (2)
## ----------------------------------------------------------------------------
```

2

```
## asian                            0.041***                   0.034***
##                                  (0.008)                    (0.008)
##
## age                              0.021***                   0.110***
##                                  (0.0003)                   (0.003)
##
## agesq                                                       -0.001***
##                                                             (0.00004)
##
## Constant                        10.519***                   8.843***
##                                  (0.012)                    (0.057)
##
## ----------------------------------------------------------------------------
## Observations                      64,115                     64,115
## R2                                0.074                      0.087
## Adjusted R2                       0.074                      0.087
## Residual Std. Error      0.658 (df = 64112)          0.653 (df = 64111)
## F Statistic         2,565.575*** (df = 2; 64112) 2,036.804*** (df = 3; 64111)
## ============================================================================
## Note:                                            *p<0.1; **p<0.05; ***p<0.01
```

```r
stargazer(model3, type = "text")
```

```
##
## ==============================================
##                        Dependent variable:
##                        ---------------------------
##                                  lwage
## ----------------------------------------------
## asian                           0.035***
##                                  (0.008)
##
## age_factor26                    0.053***
##                                  (0.021)
##
## age_factor27                    0.139***
##                                  (0.020)
##
## age_factor28                    0.189***
##                                  (0.020)
##
## age_factor29                    0.227***
##                                  (0.020)
##
## age_factor30                    0.284***
##                                  (0.020)
##
## age_factor31                    0.304***
##                                  (0.020)
##
## age_factor32                    0.365***
##                                  (0.020)
##
## age_factor33                    0.401***
##                                  (0.020)
```

3

```
##
## age_factor34                0.435***
##                             (0.020)
##
## age_factor35                0.474***
##                             (0.020)
##
## age_factor36                0.476***
##                             (0.020)
##
## age_factor37                0.519***
##                             (0.020)
##
## age_factor38                0.524***
##                             (0.020)
##
## age_factor39                0.584***
##                             (0.020)
##
## age_factor40                0.580***
##                             (0.021)
##
## age_factor41                0.599***
##                             (0.021)
##
## age_factor42                0.606***
##                             (0.021)
##
## age_factor43                0.651***
##                             (0.021)
##
## age_factor44                0.661***
##                             (0.021)
##
## age_factor45                0.646***
##                             (0.021)
##
## age_factor46                0.653***
##                             (0.021)
##
## age_factor47                0.687***
##                             (0.022)
##
## age_factor48                0.670***
##                             (0.021)
##
## age_factor49                0.664***
##                             (0.021)
##
## age_factor50                0.665***
##                             (0.021)
##
## age_factor51                0.668***
##                             (0.021)
```

4

```
##
## age_factor52                0.686***
##                             (0.021)
##
## age_factor53                0.681***
##                             (0.021)
##
## age_factor54                0.677***
##                             (0.021)
##
## age_factor55                0.664***
##                             (0.022)
##
## Constant                    10.861***
##                             (0.015)
##
## ---------------------------------------------
## Observations                 64,115
## R2                            0.087
## Adjusted R2                   0.087
## Residual Std. Error     0.653 (df = 64083)
## F Statistic        198.163*** (df = 31; 64083)
## =============================================
## Note:              *p<0.1; **p<0.05; ***p<0.01
```

Figure 7: OLS Models for Racial Effect on Log Wages

Effects of being Asian under three models, based on three regression tables above:
Holding age constant, being Asian is associated with an average increase of 0.041 in the logarithm of wages, compared to being White with statistical significance at 0.01 level. Or equivalently, holding age constant, being Asian is associated with an average increase of $4.19\% \approx (e^{0.041} - 1) \times 100$ in wages, compared to being White with statistical significance

at 0.01 level.

Holding age and age squared constant, being Asian is associated with an average increase of 0.034 in the logarithm of wages, compared to being White with statistical significance at 0.01 level. Or equivalently, holding age and age squared constant, being Asian is associated with an average increase of 3.46% $\approx (e^{0.034} - 1) \times 100$ in wages, compared to being White with statistical significance at 0.01 level.

Holding age group constant, being Asian is associated with an average increase of 0.035 in the logarithm of wages, compared to being White with statistical significance at 0.01 level. Or equivalently, holding age group constant, being Asian is associated with an average increase of 3.56% $\approx (e^{0.035} - 1) \times 100$ in wages, compared to being White with statistical significance at 0.01 level.

d. Now using exact matching (matching on age), estimate the average treatment effect of being Asian versus white on log wages. Run 4 models: 1) with no controls, 2) with age as a control, 3) with age and age squared (agesq) as controls and 4) a fully saturated regression model. How do these results differ from the OLS estimates from part (c)? Code this by hand. Do not use any packages or libraries in R or Stata that implement matching for you.

```r
# Exact Matching on age
# Calculate the number of Asians and Whites in each age group
age_group_counts <- data %>%
  group_by(age) %>%
  summarise(
    total = n(),
    asian_count = sum(asian),
    white_count = total - asian_count
  )

# Calculate p_hat (proportion of Asians) for each age group
age_group_counts <- age_group_counts %>%
  mutate(p_hat = asian_count / total)

# Calculate IPW weights for treated (Asian) and untreated (White) units
age_group_counts <- age_group_counts %>%
  mutate(
    weight_asian = ifelse(p_hat > 0, 1 / p_hat, 0),   # Avoid division by zero
    weight_white = ifelse(p_hat < 1, 1 / (1 - p_hat), 1)
  )

# Add IPW weights to each individual
data_with_weights <- data %>%
  left_join(age_group_counts, by = "age") %>%
  mutate(
    wt_ipw = ifelse(asian == 1, weight_asian, weight_white)
  )
```

```r
# Fit Models with calculated IPW weights
# Model 0: with no controls
```



```r
model0_ipw <- lm(lwage ~ asian, data = data_with_weights, weights=wt_ipw)
# Model 1: Control for age
model1_ipw <- lm(lwage ~ asian + age, data = data_with_weights, weights=wt_ipw)

# Model 2: Control for age and age squared
model2_ipw <- lm(lwage ~ asian + age + agesq, data = data_with_weights, weights=wt_ipw)

# Model 3: Fully saturated model
model3_ipw <- lm(lwage ~ asian + age_factor, data = data_with_weights, weights=wt_ipw)

stargazer(model0_ipw, type = "text")
```

```
##
## ===============================================
##                         Dependent variable:
##                     ---------------------------
##                                lwage
## -----------------------------------------------
## asian                         0.030***
##                                (0.006)
##
## Constant                      11.344***
##                                (0.004)
##
## -----------------------------------------------
## Observations                   64,115
## R2                             0.0005
## Adjusted R2                    0.0005
## Residual Std. Error      0.989 (df = 64113)
## F Statistic          30.044*** (df = 1; 64113)
## ===============================================
## Note:               *p<0.1; **p<0.05; ***p<0.01
```

```r
stargazer(model1_ipw, model2_ipw, type = "text")
```

```r
stargazer(model1_ipw, model2_ipw, type = "text")
```

```
##
## ==============================================================================
##                                       Dependent variable:
##                     ----------------------------------------------------------
##                                              lwage
##                                 (1)                           (2)
## ------------------------------------------------------------------------------
## asian                         0.030***                      0.030***
##                               (0.005)                       (0.005)
##
## age                           0.017***                      0.124***
##                               (0.0003)                      (0.003)
##
## agesq                                                      -0.001***
##                                                            (0.00004)
##
## Constant                     10.670***                      8.650***
##                               (0.013)                       (0.059)
##
```



```
## ------------------------------------------------------------------------------
## Observations                   64,115                        64,115
## R2                              0.048                         0.066
## Adjusted R2                     0.048                         0.066
## Residual Std. Error      0.965 (df = 64112)            0.956 (df = 64111)
## F Statistic         1,604.199*** (df = 2; 64112) 1,499.164*** (df = 3; 64111)
## ==============================================================================
## Note:                                              *p<0.1; **p<0.05; ***p<0.01
```

```r
stargazer(model3_ipw, type = "text")
```

```
##
## ===============================================
##                         Dependent variable:
##                     ---------------------------
##                                lwage
## -----------------------------------------------
## asian                         0.030***
##                               (0.005)
##
## age_factor26                   0.036*
##                               (0.021)
##
## age_factor27                  0.109***
##                               (0.021)
##
## age_factor28                  0.193***
##                               (0.021)
##
## age_factor29                  0.191***
##                               (0.021)
##
## age_factor30                  0.266***
##                               (0.020)
##
## age_factor31                  0.299***
##                               (0.020)
##
## age_factor32                  0.367***
##                               (0.021)
##
## age_factor33                  0.372***
##                               (0.021)
##
## age_factor34                  0.413***
##                               (0.021)
##
## age_factor35                  0.441***
##                               (0.021)
##
## age_factor36                  0.433***
##                               (0.021)
##
## age_factor37                  0.542***
##                               (0.021)
```

```
##
## age_factor38                 0.523***
##                              (0.021)
##
## age_factor39                 0.567***
##                              (0.021)
##
## age_factor40                 0.533***
##                              (0.021)
##
## age_factor41                 0.583***
##                              (0.021)
##
## age_factor42                 0.560***
##                              (0.022)
##
## age_factor43                 0.608***
##                              (0.022)
##
## age_factor44                 0.609***
##                              (0.022)
##
## age_factor45                 0.573***
##                              (0.022)
##
## age_factor46                 0.563***
##                              (0.022)
##
## age_factor47                 0.633***
##                              (0.023)
##
## age_factor48                 0.624***
##                              (0.022)
##
## age_factor49                 0.611***
##                              (0.022)
##
## age_factor50                 0.556***
##                              (0.021)
##
## age_factor51                 0.567***
##                              (0.022)
##
## age_factor52                 0.594***
##                              (0.022)
##
## age_factor53                 0.597***
##                              (0.022)
##
## age_factor54                 0.552***
##                              (0.022)
##
## age_factor55                 0.478***
##                              (0.022)
```

8

```
##
## Constant                     10.904***
##                              (0.016)
##
## -----------------------------------------------
## Observations                  64,115
## R2                             0.067
## Adjusted R2                    0.066
## Residual Std. Error      0.955 (df = 64083)
## F Statistic         147.905*** (df = 31; 64083)
## ===============================================
## Note:               *p<0.1; **p<0.05; ***p<0.01
```

Figure 8: Four Models for Racial Effect on Log Wages after IPW Weighting

As we used the balancing function to balance the age covariate across Asian and White groups, four regression models yield exactly the same values of $\hat{\sigma}$, which is a different phenomenon from the OLS estimates across different regression functional forms. The difference in the IPW-weighted mean outcomes by race (Asian vs. White) is around 0.03, which is slightly smaller from the OLS estimates, as we balanced the age covariate.

In terms of interpretation of 0.03, after balancing age covariate across racial groups, being Asian is associated with an average increase of 0.03 in the logarithm of wages, compared to being White with statistical significance at 0.01 level. Or equivalently, holding age group constant, being Asian is associated with an average increase of $3.05\% \approx (e^{0.03} - 1) \times 100$ in wages, compared to being White with statistical significance at 0.01 level.

e. Is the common support assumption satisfied? What would the conditional independence assumption mean in this context?

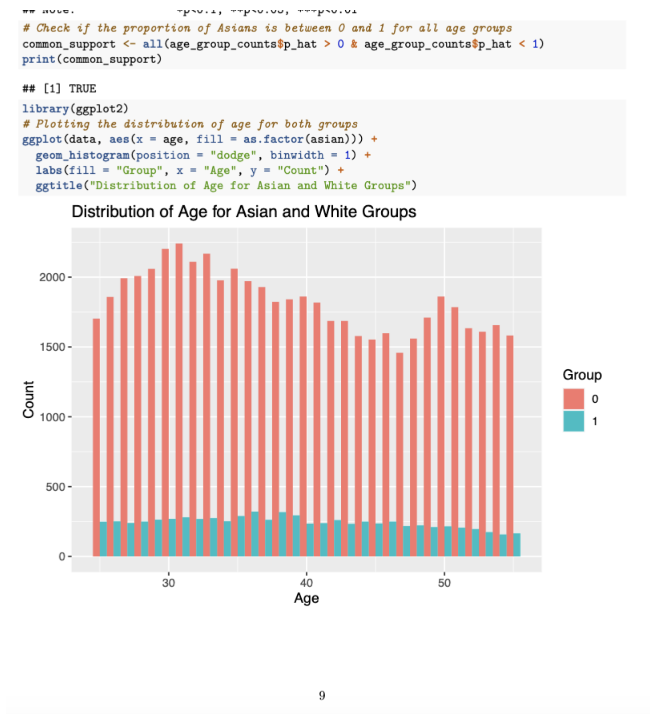

Figure 9: CSA Evaluation for Asian as Treatment Effect

Common Support Assumption states that $0 < Pr(D_i|X = x_0) < 1, \forall x_0$. In this case, at each age group, we have both Asian and White observations, after examining the propensity score at each age group. Hence, the common support assumption is satisfied.

Conditional independence assumption in this context means that being Asian or White is independent of the potential wage outcome, given the observed age. Or more explicitly, after controlling for age, any difference in wages between Asian and White individuals can be causally attributed to their racial background.

f. Redo part (c) using only the US-born sample.

```r
setwd("/Users/qilinzhou/Desktop")
data <- read_dta("asianwhiteMen.dta")
```

9

```r
# Filter data for only US-born samples
usborn_data <- subset(data, usborn == 1)
# Model 1: Control for age
model1_us <- lm(lwage ~ asian + age, data = usborn_data)

# Model 2: Control for age and age squared
model2_us <- lm(lwage ~ asian + age + agesq, data = usborn_data)

# Model 3: Fully saturated model
usborn_data$age_factor <- factor(usborn_data$age)
model3_us <- lm(lwage ~ asian + age_factor, data = usborn_data)

stargazer(model1_us, model2_us, type = "text")
```

```
##
## ============================================================================
##                                       Dependent variable:
##                        -----------------------------------------------------
##                                              lwage
##                                   (1)                        (2)
## ----------------------------------------------------------------------------
## asian                          0.074***                   0.078***
##                                (0.013)                    (0.013)
##
## age                            0.022***                   0.103***
##                               (0.0003)                    (0.003)
##
## agesq                                                    -0.001***
##                                                          (0.00004)
##
## Constant                      10.482***                   8.944***
##                                (0.012)                    (0.059)
##
## ----------------------------------------------------------------------------
## Observations                    56,178                     56,178
## R2                              0.082                      0.093
## Adjusted R2                     0.082                      0.093
## Residual Std. Error      0.645 (df = 56175)         0.641 (df = 56174)
## F Statistic         2,504.833*** (df = 2; 56175) 1,925.466*** (df = 3; 56174)
## ============================================================================
## Note:                                            *p<0.1; **p<0.05; ***p<0.01
```

```r
stargazer(model3_us, type = "text")
```

```
##
## ===============================================
##                         Dependent variable:
##                     ---------------------------
##                                lwage
## -----------------------------------------------
## asian                         0.078***
##                                (0.013)
##
## age_factor26                  0.054***
```

10

```
##                                (0.021)
##
## age_factor27                  0.145***
##                                (0.021)
##
## age_factor28                  0.185***
##                                (0.021)
##
## age_factor29                  0.229***
##                                (0.021)
##
## age_factor30                  0.282***
##                                (0.020)
##
## age_factor31                  0.296***
##                                (0.020)
##
## age_factor32                  0.353***
##                                (0.021)
##
## age_factor33                  0.400***
##                                (0.020)
##
## age_factor34                  0.428***
##                                (0.021)
##
## age_factor35                  0.476***
##                                (0.021)
##
## age_factor36                  0.467***
##                                (0.021)
##
## age_factor37                  0.501***
##                                (0.021)
##
## age_factor38                  0.506***
##                                (0.021)
##
## age_factor39                  0.578***
##                                (0.021)
##
## age_factor40                  0.579***
##                                (0.021)
##
## age_factor41                  0.594***
##                                (0.021)
##
## age_factor42                  0.605***
##                                (0.022)
##
## age_factor43                  0.650***
##                                (0.022)
##
## age_factor44                  0.662***
```

```
##                                (0.022)
##
## age_factor45                   0.647***
##                                (0.022)
##
## age_factor46                   0.662***
##                                (0.022)
##
## age_factor47                   0.686***
##                                (0.023)
##
## age_factor48                   0.665***
##                                (0.023)
##
## age_factor49                   0.668***
##                                (0.022)
##
## age_factor50                   0.679***
##                                (0.022)
##
## age_factor51                   0.678***
##                                (0.022)
##
## age_factor52                   0.701***
##                                (0.022)
##
## age_factor53                   0.692***
##                                (0.022)
##
## age_factor54                   0.695***
##                                (0.022)
##
## age_factor55                   0.695***
##                                (0.022)
##
## Constant                      10.854***
##                                (0.015)
##
## ---------------------------------------------
## Observations                   56,178
## R2                              0.094
## Adjusted R2                     0.093
## Residual Std. Error     0.641 (df = 56146)
## F Statistic         187.536*** (df = 31; 56146)
## =============================================
## Note:               *p<0.1; **p<0.05; ***p<0.01
```

Figure 10: OLS Models for Racial Effect on Log Wages for USborn Samples

Effects of being USborn Asian under three models, based on three regression tables above: Holding age constant, being USborn Asian is associated with an average increase of 0.074 in the logarithm of wages, compared to being USborn White with statistical significance at 0.01 level. Or equivalently, holding age constant, being USborn Asian is associated with an average increase of $7.68\% \approx (e^{0.074} - 1) \times 100$ in wages, compared to being USborn White with statistical significance at 0.01 level.

Holding age and age squared constant, being USborn Asian is associated with an average increase of 0.078 in the logarithm of wages, compared to being USborn White with statistical significance at 0.01 level. Or equivalently, holding age and age squared constant, being USborn Asian is associated with an average increase of $8.11\% \approx (e^{0.078} - 1) \times 100$ in wages, compared to being USborn White with statistical significance at 0.01 level.

Holding age group constant, being USborn Asian is associated with an average increase of 0.078 in the logarithm of wages, compared to being USborn White with statistical significance at 0.01 level. Or equivalently, holding age group constant, being USborn Asian is associated with an average increase of $8.11\% \approx (e^{0.078} - 1) \times 100$ in wages, compared to being USborn White with statistical significance at 0.01 level.

g. Redo part (d) using only the US-born sample.

```r
# Exact Matching on age for USBorn
# Calculate the number of Asians and Whites in each age group
age_group_counts <- usborn_data %>%
  group_by(age) %>%
  summarise(
    total = n(),
    asian_count = sum(asian),
```

12

```r
    white_count = total - asian_count
  )

# Calculate p_hat (proportion of Asians) for each age group
age_group_counts <- age_group_counts %>%
  mutate(p_hat = asian_count / total)

# Calculate IPW weights for treated (Asian) and untreated (White) units
age_group_counts <- age_group_counts %>%
  mutate(
    weight_asian = ifelse(p_hat > 0, 1 / p_hat, 0),    # Avoid division by zero
    weight_white = ifelse(p_hat < 1, 1 / (1 - p_hat), 1)
  )

# Add IPW weights to each individual
data_with_us_weights <- usborn_data %>%
  left_join(age_group_counts, by = "age") %>%
  mutate(
    wt_ipw = ifelse(asian == 1, weight_asian, weight_white)
  )
```

```r
# Model 0: with no controls
model0_us_ipw <- lm(lwage ~ asian, data = data_with_us_weights, weights=wt_ipw)
# Model 1: Control for age
model1_us_ipw <- lm(lwage ~ asian + age, data = data_with_us_weights, weights=wt_ipw)

# Model 2: Control for age and age squared
model2_us_ipw <- lm(lwage ~ asian + age + agesq, data = data_with_us_weights, weights=wt_ipw)

# Model 3: Fully saturated model
model3_us_ipw <- lm(lwage ~ asian + age_factor, data = data_with_us_weights, weights=wt_ipw)

stargazer(model0_us_ipw, type = "text")
```

```
##
## ===============================================
##                         Dependent variable:
##                     ---------------------------
##                                lwage
## -----------------------------------------------
## asian                        0.054***
##                               (0.006)
##
## Constant                     11.331***
##                               (0.004)
##
## -----------------------------------------------
## Observations                   56,178
## R2                              0.002
## Adjusted R2                     0.002
## Residual Std. Error      0.978 (df = 56176)
## F Statistic          86.668*** (df = 1; 56176)
## ===============================================
## Note:               *p<0.1; **p<0.05; ***p<0.01
```

13

```r
stargazer(model1_us_ipw, model2_us_ipw, type = "text")
```

```
##
## ============================================================================
##                                        Dependent variable:
##                        -----------------------------------------------------
##                                               lwage
##                                   (1)                         (2)
## ----------------------------------------------------------------------------
## asian                          0.054***                    0.054***
##                                (0.006)                     (0.006)
##
## age                            0.018***                    0.110***
##                                (0.0003)                    (0.003)
##
## agesq                                                     -0.001***
##                                                           (0.00004)
##
## Constant                      10.612***                    8.883***
##                                (0.013)                     (0.062)
##
## ----------------------------------------------------------------------------
## Observations                    56,178                      56,178
## R2                              0.058                       0.071
## Adjusted R2                     0.058                       0.071
## Residual Std. Error      0.950 (df = 56175)          0.943 (df = 56174)
## F Statistic         1,717.221*** (df = 2; 56175) 1,435.834*** (df = 3; 56174)
## ============================================================================
## Note:                                          *p<0.1; **p<0.05; ***p<0.01
```

```r
stargazer(model3_us_ipw, type = "text")
```

```
##
## ===============================================
##                         Dependent variable:
##                        ---------------------------
##                                  lwage
## -----------------------------------------------
## asian                          0.054***
##                                 (0.006)
##
## age_factor26                    0.046**
##                                 (0.022)
##
## age_factor27                   0.129***
##                                 (0.021)
##
## age_factor28                   0.187***
##                                 (0.021)
##
## age_factor29                   0.205***
##                                 (0.021)
##
## age_factor30                   0.258***
```

14

```
##                               (0.021)
##
## age_factor31                 0.292***
##                               (0.021)
##
## age_factor32                 0.302***
##                               (0.021)
##
## age_factor33                 0.371***
##                               (0.021)
##
## age_factor34                 0.385***
##                               (0.022)
##
## age_factor35                 0.455***
##                               (0.021)
##
## age_factor36                 0.405***
##                               (0.022)
##
## age_factor37                 0.550***
##                               (0.022)
##
## age_factor38                 0.515***
##                               (0.022)
##
## age_factor39                 0.523***
##                               (0.022)
##
## age_factor40                 0.512***
##                               (0.022)
##
## age_factor41                 0.637***
##                               (0.022)
##
## age_factor42                 0.561***
##                               (0.023)
##
## age_factor43                 0.644***
##                               (0.023)
##
## age_factor44                 0.578***
##                               (0.023)
##
## age_factor45                 0.525***
##                               (0.023)
##
## age_factor46                 0.559***
##                               (0.023)
##
## age_factor47                 0.619***
##                               (0.024)
##
## age_factor48                 0.550***
```

```
##                               (0.023)
##
## age_factor49                  0.685***
##                               (0.023)
##
## age_factor50                  0.571***
##                               (0.022)
##
## age_factor51                  0.562***
##                               (0.022)
##
## age_factor52                  0.600***
##                               (0.023)
##
## age_factor53                  0.677***
##                               (0.023)
##
## age_factor54                  0.628***
##                               (0.023)
##
## age_factor55                  0.498***
##                               (0.023)
##
## Constant                     10.896***
##                               (0.016)
##
## ----------------------------------------------------
## Observations                   56,178
## R2                              0.075
## Adjusted R2                     0.074
## Residual Std. Error     0.942 (df = 56146)
## F Statistic         146.464*** (df = 31; 56146)
## ====================================================
## Note:               *p<0.1; **p<0.05; ***p<0.01
```

16

Figure 11: Four Models for Racial Effect on Log Wages after IPW Weighting for USborn Samples

The difference in the IPW-weighted mean outcomes by race (USborn Asian vs. USborn White) is around 0.054.

In terms of interpretation of 0.054, after balancing age covariate across racial groups, being USborn Asian is associated with an average increase of 0.054 in the logarithm of wages, compared to being USborn White with statistical significance at 0.01 level. Or equivalently, holding age group constant, being USborn Asian is associated with an average increase of $5.55\% \approx (e^{0.054} - 1) \times 100$ in wages, compared to being USborn White with statistical significance at 0.01 level.

h. How do your results in parts (f) and (g) differ from those in parts (c) and (d)? What does this tell you?

After changing our samples to include only USborn participants, the positive effect of being Asian on wages in logarithm increases, for both OLS models and regression models with IPW weighting. This suggests that birthplace is another confounder(covariate) we have not ruled out in previous samples that include both USborn and non USborn participants.

Among USborn participants, being Asian is associated with an average larger increase percentage in wages, with statistical significance at 0.01 level. Specifically, non USborn Asian participants might be negatively influenced by their immigration status or language barriers, which in turn impacts their wage level. While a large proportion of Asian participants are non USborn from part(b), the larger effect of being Asian on wages considering only USborn samples, compared to being White, might due to this composition bias.