PS3 Solutions

Jingle Fu

Solution (a).

Gender-Based Wage Differentials

1. On average, the average hourly earnings for men is \$20.114, while women is about \$17.484.

2. Coefficient for female: -2.63, meaning females earn on average \$2.63 less per hour than males, holding other factors constant.

3. Statistical Significance: The t-test for the gender coefficient is highly significant (p-value < 0.01), confirming that males earn significantly more than females in this dataset.

Education-Based Wage Differentials

1. On average, the average hourly earnings for people with a bachelor degree is \$22.908, while for people only graduated from a high-school, the number is \$15.332.

2. Coefficient for bachelor: 7.58, meaning workers with a bachelor's degree earn on average \$7.58 more per hour than those with only a high school diploma.

3. Statistical Significance: The t-test for the education coefficient is also highly significant (p-value < 0.01), confirming that higher education is associated with higher earnings.

```r
rm(list=ls())
library(tidyverse)
library(readr)
library(dplyr)
library(broom)
library(ggplot2)
```

```r
library(xtable)
library(stargazer)
data <- read_csv("dat_CPS08.csv")

avg_ahe_gender = aggregate(ahe ~ female, data = data, FUN = mean)
c(avg_ahe_gender)
avg_ahe_education = aggregate(ahe ~ bachelor, data = data, FUN = mean)
c(avg_ahe_education)

gender_model <- lm(ahe ~ female, data=data)
education_model <- lm(ahe ~ bachelor, data=data)
gender_coeff <- coef(gender_model)["female"]
education_coeff <- coef(education_model)["bachelor"]
c(gender_coeff, education_coeff)

```

Table 1: Regression Results for Gender and Education

	Dependent variable:	
	ahe	
	(1)	(2)
female	-2.630^{***}	
	(0.231)	
bachelor		7.577^{***}
		(0.214)
Constant	20.114^{***}	15.332^{***}
	(0.152)	(0.149)
Observations	7,711	7,711
R^2	0.017	0.139
Adjusted R^2	0.016	0.139
Residual Std. Error (df = 7709)	10.056	9.407
F Statistic (df = 1; 7709)	129.457^{***}	$1{,}248.792^{***}$
Note:	*p<0.1; **p<0.05; ***p<0.01	

Solution (b).

Regression Coefficients

1. **Age Coefficient:** 0.585, implying an increase of approximately \$0.59 in hourly

earnings for each additional year of age, holding other factors constant.

2. **Gender (`female`) Coefficient:** -3.664, indicating that females earn about \$3.66 less per hour than males, controlling for age and education.

3. **Education (`bachelor`) Coefficient:** 8.083, showing that those with a bachelor's degree earn about \$8.083 more per hour compared to those with only a high school diploma, controlling for age and gender.

Changes in Earnings with Age:

1. **From Age 28 to 29:** Earnings are expected to increase by about \$0.59.

2. **From Age 37 to 38:** Earnings are expected to increase by the same amount, about \$0.59.

```r
age_model <- lm(ahe ~ age + female + bachelor, data=data)
age_coeff <- coef(age_model)["age"]
change_28_to_29 <- age_coeff * 1
change_37_to_38 <- age_coeff * 1

```

Table 2: Regression Results for Age, Gender, and Education

	Dependent variable:
	ahe
age	0.585^{***}
	(0.036)
female	-3.664^{***}
	(0.211)
bachelor	8.083^{***}
	(0.209)
Constant	-0.636
	(1.085)

Continued on next page

Table 2 – continued from previous page

	Dependent variable:
	ahe
Observations	7,711
R^2	0.200
Adjusted R^2	0.199
Residual Std. Error	9.072 (df = 7707)
F Statistic	641.492*** (df = 3; 7707)
Note:	*p<0.1; **p<0.05; ***p<0.01

Solution (c).

For the logarithm of earnings regression:

1. **Age Coefficient:** 0.0273, implying that each additional year of age results in about a 2.77% increase in earnings, holding other factors constant.

Expected Changes in Log-Earnings:

1. **From Age 28 to 29:** Earnings are expected to increase by about 2.77%.
2. **From Age 37 to 38:** Earnings are expected to increase by the same percentage, about 2.77%.

```
data$log_ahe <- log(data$ahe)
log_age_model <- lm(log_ahe ~ age + female + bachelor, data=data)
log_age_coeff <- coef(log_age_model)["age"]
percent_change_28_to_29 <- (exp(log_age_coeff)-1) * 100
percent_change_37_to_38 <- (exp(log_age_coeff)-1) * 100

```

Table 3: Logarithm of Earnings Regression

	Dependent variable:
	log_ahe
age	0.027***

Continued on next page

Table 3 – continued from previous page

	Dependent variable:
	log_ahe
	(0.002)
female	-0.186^{***}
	(0.011)
bachelor	0.428^{***}
	(0.011)
Constant	1.876^{***}
	(0.056)
Observations	7,711
R^2	0.201
Adjusted R^2	0.200
Residual Std. Error	0.469 (df = 7707)
F Statistic	644.876^{***} (df = 3; 7707)
Note:	*p<0.1; **p<0.05; ***p<0.01

Solution (d).

For the regression using the logarithm of age:

1. **Log-Age Coefficient:** 0.804, which translates into a change in log-earnings:
   (a) **From Age 28 to 29:** About 2.82% increase in earnings.
   (b) **From Age 37 to 38:** About 2.14% increase in earnings.
2. **Gender and Education Coefficients** remain the same as in the previous model, reflecting similar percentage impacts on earnings.

```r
data$log_age <- log(data$age)
log_age_transform_model <- lm(log_ahe ~ log_age + female + bachelor,
    data=data)
log_age_transform_coeff <- coef(log_age_transform_model)["log_age"]
approx_percent_change_28_to_29 <- log_age_transform_coeff * (log(29) -
    log(28)) * 100
approx_percent_change_37_to_38 <- log_age_transform_coeff * (log(38) -
    log(37)) * 100
```

```
list(
    log_age_transform_coeff = log_age_transform_coeff,
    approx_percent_change_28_to_29 = approx_percent_change_28_to_29,
    approx_percent_change_37_to_38 = approx_percent_change_37_to_38
)

```

Table 4: Log Earnings on Log Age, Gender, and Education

	Dependent variable:
	log_ahe
log_age	0.804^{***}
	(0.055)
female	-0.186^{***}
	(0.011)
bachelor	0.428^{***}
	(0.011)
Constant	-0.035
	(0.186)
Observations	7,711
R^2	0.201
Adjusted R^2	0.200
Residual Std. Error	0.469 (df = 7707)
F Statistic	645.318^{***} (df = 3; 7707)
Note:	*p<0.1; **p<0.05; ***p<0.01

Solution (e).

For the regression including quadratic age effects:

1. **Age Coefficient:** 0.081

2. **Age Squared Coefficient:** -0.001, indicating a slight decrease in logarithm of average hourly earnings as age progresses.

3. **Expected Changes in Log-Earnings:**

(a) **From Age 28 to 29:** Earnings are expected to increase by about 3.01%.

(b) **From Age 37 to 38:** Earnings are expected to increase by about 1.37%.

```
data$age2 <- data$age^2
age_quadratic_model <- lm(log_ahe ~ age + age2 + female + bachelor, data
    =data)
age_quadratic_coeff <- coef(age_quadratic_model)["age"]
age2_quadratic_coeff <- coef(age_quadratic_model)["age2"]
percent_change_quadratic_28_to_29 <- (age_quadratic_coeff + 2 * age2_
    quadratic_coeff * 28)  * 100
percent_change_quadratic_37_to_38 <- (age_quadratic_coeff + 2 * age2_
    quadratic_coeff * 37) * 100
list(
    age_quadratic_coeff = age_quadratic_coeff,
    age2_quadratic_coeff = age2_quadratic_coeff,
    percent_change_quadratic_28_to_29 = percent_change_quadratic_28_to_
    29,
    percent_change_quadratic_37_to_38 = percent_change_quadratic_37_to_
    38
)

```

Table 5: Quadratic Age Effects in Earnings Regression

	Dependent variable:
	log_ahe
age	0.081*
	(0.043)
age^2	−0.001
	(0.001)
female	−0.186***
	(0.011)
bachelor	0.428***
	(0.011)
Constant	1.085*

Continued on next page

Table 5 – continued from previous page

	Dependent variable:
	log_ahe
	(0.638)
Observations	7,711
R^2	0.201
Adjusted R^2	0.200
Residual Std. Error	0.469 (df = 7706)
F Statistic	484.078*** (df = 4; 7706)
Note:	*p<0.1; **p<0.05; ***p<0.01

Solution (f).

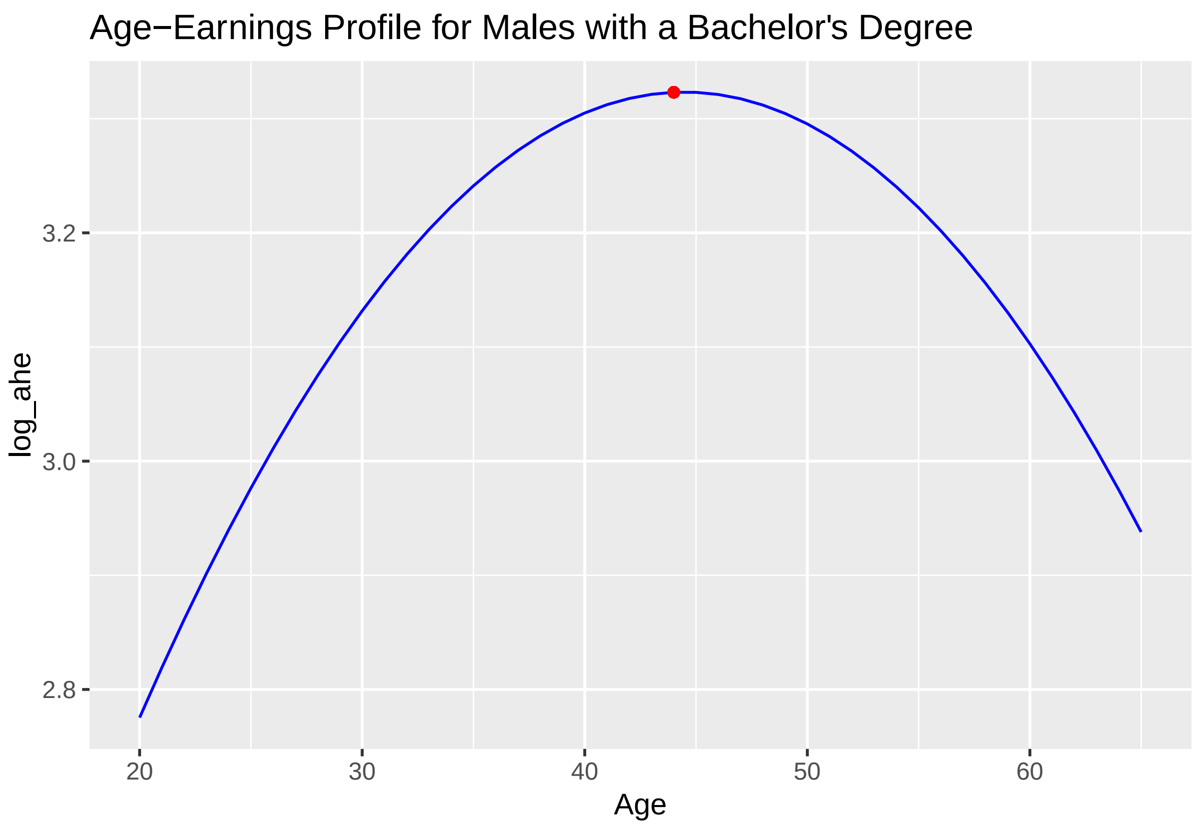
The plot above shows the age-earnings profile for males with a bachelor's degree, covering ages 20 to 65. The earnings peak at age 44, indicating the age at which these individuals earn the highest average hourly rate. (The peak is 44.474 actually, but since our age could only be integers, we take 44.)

```
age_range <- 20:65
predicted_earnings <- predict(age_quadratic_model, newdata = data.frame(
    age = age_range, age2 = age_range^2, female = 0, bachelor = 1))
peak_age <- age_range[which.max(predicted_earnings)]
list(peak_age = peak_age)

ggplot(data.frame(age = age_range, earnings =predicted_earnings), aes(x=
    age, y=earnings)) +
  geom_line(color="blue") +
  geom_point(aes(x=peak_age, y=max(predicted_earnings)), color="red") +
  labs(title="Age-Earnings Profile for Males with a Bachelor's Degree",
    x="Age", y="log_ahe")

cat(sprintf("Peak age for earnings: %d", peak_age), sep="\n")

```



Solution (g).

For the regression including the interaction between age and gender:

1. **Age Coefficient:** 0.033, suggesting an overall increase of about 3.46% in earnings per year of age for males.

2. **Female Coefficient:** 0.31, indicating an overall higher base log-earnings for females when not considering age effects.

3. **Interaction Coefficient (Age*Female):** -0.017, which implies that age on earnings has a negative impact on females. Each additional year of age results in a reduced earnings of about 1.7% for females compared to males.

This interaction term suggests a significant difference in the effect of age on earnings between genders, with age having a negative effect on earnings for females.

```r
data$age_female <- data$age * data$female
interaction_model <- lm(log_ahe ~ age + female + bachelor + age_female,
    data=data)
interaction_coeff <- coef(interaction_model)["age_female"]
percent_change_interaction = (exp(interaction_coeff) - 1) * 100

```

Table 6: Interaction Model of Age and Gender on Earnings

	Dependent variable:
	log_ahe
age	0.035^{***}
	(0.002)
female	0.310^{***}
	(0.112)
bachelor	0.427^{***}
	(0.011)
age_female	-0.017^{***}
	(0.004)
Constant	1.660^{***}
	(0.074)
Observations	7,711
R^2	0.203
Adjusted R^2	0.202
Residual Std. Error	0.469 (df = 7706)
F Statistic	489.765^{***} (df = 4; 7706)
Note:	*p<0.1; **p<0.05; ***p<0.01