Written answers for Introduction to Statistics – HW3

1. Model 1 *(40 points)*

In Model 1, you do not consider different slopes of education on income across race types, but you do consider different intercepts (starting points). Note that your reference group is\White," and you compare the income of \Black" and \Other" to that of \White."

1. State the null hypothesis (H0) and the alternative hypothesis (H1) for the coefficient of the continuous variable, \Years of education" (educ).

   Ans: The null hypothesis and the alternative hypothesis for the coefficient of the continuous variable, "years of education" (*educ)* is as follows:

   H_0: The year of education (*educ)* has no effect on the annual household income *(income),* holding other variables constant.
   H_1: The years of education (*educ)* has an effect on the annual household income *(income),* holding other variables constant.

2. State the null hypothesis (H0) and the alternative hypothesis (H2) for the coefficient of the dummy variable \Black" (Black_dummy).

   Ans: The null hypothesis and the alternative hypothesis for the coefficient of the dummy variable, "Black" (*Black_dummy)* is as follows:

   H_0: Coming from the Black race category, represented by *Black_dummy,* has no effect on the annual household income *(income),* holding other variables constant.
   H_2: Coming from the Black race category, represented by *Black_dummy,* has an effect on the annual household income *(income)* relative to the reference race category, White (*White_dummy*), holding other variables constant.

3. State the null hypothesis (H0) and the alternative hypothesis (H3) for the coefficient of the dummy variable \Other" (Other_dummy).

   Ans: The null hypothesis and the alternative hypothesis for the coefficient of the dummy variable, "Other" (*Other_dummy)* is as follows:

   H_0: Coming from the Other race category, represented by *Other_dummy,* has no effect on the annual household income *(income),* holding other variables constant.
   H_3: Coming from the Other race category, represented by *Other_dummy,* has an effect on the annual household income *(income)* relative to the reference race category, White (*White_dummy),* holding other variables constant.

4. Conduct the appropriate multiple regression analysis to test your hypothesis and present the results in a table format. For dummy variables, label them with 'dummy' in their names within the dataset (e.g., Black dummy). Do not use the \race" variable, as it is a nominal variable.

Ans: Please find the results of my regression analysis between the income (DV) and years of education (IV) and the dummy variables (Z), Black and Other in Table 1 using the "\Stargazer" package. This analysis will also be saved as "model1.txt" when you run the code.

Table 1

Regression Results for Model 1

	Dependent variable:
	Annual Income (USD)
Education (Years)	770.380***
	(41.446)
Black Dummy	-2,764.693***
	(361.325)
Other Dummy	-604.144
	(441.443)
Age	-8.979
	(8.032)
Number of Children	188.458**
	(80.714)
Constant	7,980.009***
	(708.543)
Observations	1,766
R2	0.203
Adjusted R2	0.200
Residual Std. Error	5,132.158 (df = 1760)
F Statistic	89.400*** (df = 5; 1760)
Note:	*p<0.1; **p<0.05; ***p<0.01

5. Visualize the OLS coefficients by creating plots (Note: This refers specifically to the visualization of the coefficients, not the predictions)

Ans: Please find the coefficient plot of the regression analysis conducted in Q4 in Figure 1 below. This plot will also be saved as "coefplot1.png" when you run the code.

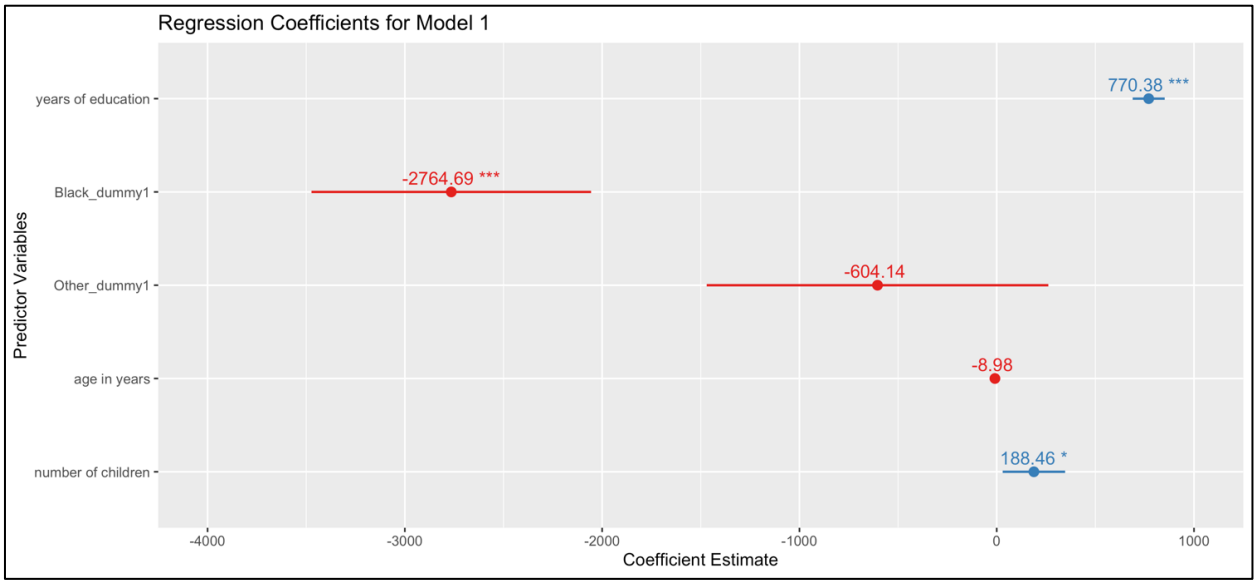

Figure 1

6. Interpret the regression results (from the table in question 4 and the coefficients plot in question 5), focusing on their statistical significance and how they align with the hypotheses (e.g., whether the null hypothesis is rejected).

   Ans: The regression results in *Table 1* and coefficient plot in *Figure 1* show that the effect of *educ* and *black_dummy* are statistically significant at 1% level. In addition to that, the effect of *childs* is also statistically significant at 5% level. However, the effects of variables, *Other_dummy* and *age* are not statistically significant. The F-Statistic of the regression coefficients is significant at 1% level indicating that the *years of education* with *black dummy, other dummy, age* and *number of children* can explain the variance in the dependent variable, annual income, than an intercept-only model.

   With these results, we can reject the H_0 in favour of H_1 in the Q1 meaning that years of education have an effect on the annual income keeping all variables constant. The results show that with a year of increase in education, there is *USD 770.380* increase in the annual income. Similarly, we can reject the H_0 in favour of H_2 in the Q2 stating that coming from a black race has an effect on annual income, relative to coming from a white race while keeping other variables constant. There is a decrease of *USD 2764.693* in the starting annual income of an individual coming from black race vs an individual coming from the white race.

   We fail to reject H_0 in favour of H_3 in the Q3 which means that coming from Other race has no effect on the annual income relative to coming from the white race, while keeping other variables constant.

2. Model 2 *(30 points)*

In Model 2, you consider that the effect of education on income varies across different races, allowing for different slopes of education on income for each racial group.

1. Write the regression equation for our model using symbols, such as α, β_1 etc., rather than using OLS coefficients from the results.

   Ans: The regression equation for Model 2 is as follows:

$$\text{Annual Income } (Y) = \alpha + \beta_1 Years\ of\ Education\ (X) + \beta_2 Black\ Dummy + \beta_3 Other\ Dummy + \beta_4\ Age + \beta_5\ Childs + \beta_6\ (Black\ Dummy\ X\ Years\ of\ Education) + \beta_7\ (Other\ Dummy\ X\ Years\ of\ Education) + \mu$$

……. (I)

where α is the constant and β_1 denotes the effect of X on Y, keeping other variables constant. β_2 *and* β_3 denotes the relative effects of Black race category and Other race category to reference category, White, respectively, while keeping other variables constant. β_4 *and* β_5 represent the effects of age and number of children respectively, while keeping other variables constant. β_6*and* β_7 is the relative effect of race (vs white), black and other category respectively, on education, holding other variables constant. The symbol μ represents the random component of the dependent variable, income.

2. Write the formula for the marginal effects of education according to the regression equation above.

Ans: Please find below the marginal effects of education according to equation (I):

$$Marginal\ Effects\ of\ Education\ on\ Income\ for\ White\ race = \beta_1$$
$$Marginal\ Effects\ of\ Education\ on\ Income\ for\ Black\ race = \beta_1 + \beta_6\ Black\ Dummy$$
$$Marginal\ Effects\ of\ Education\ on\ Income\ for\ Other\ race = \beta_1 + \beta_7\ Other\ Dummy$$

3. Conduct the appropriate multiple regression analysis to test your hypotheses and present the results in a table format.

Ans: The hypothesis of the multiple regression will be as follows:

H_0 : The effect of education on income doesn't differ between individuals of different race taking white as a reference group, and keeping all other variables constant.

H_A: The effect of education on income differs between individuals of different race taking white as a reference group and keeping all other variables constant.

Please find the results of the regression analysis between the income (DV) and years of education (IV) incorporating the effects of race (*Black_dummy and Other_dummy in reference to White_dummy)* on education in *Table 2* using the "\Stargazer" package. This analysis will also be saved as "model2.txt" when you run the code. The results suggest that the variation of the effect of education on income by race is not statistically significant at the 5% level, therefore, we fail to reject the null in this Model.

Table 2

Interaction: educ X Black X Other

	Dependent variable:
	Annual Income (USD)
Education (Years)	719.933***
	(47.171)
Black dummy	-6,273.571***
	(1,840.176)
Other dummy	-2,883.156*
	(1,552.760)
Age	-9.557

	(8.028)
Number of Children	193.730**
	(80.695)
Educ x Black	271.811*
	(140.665)
Educ x Other	174.583
	(115.269)
Constant	8,689.284***
	(776.154)
Observations	1,766
R2	0.205
Adjusted R2	0.202
Residual Std. Error	5,127.288 (df = 1758)
F Statistic	64.742*** (df = 7; 1758)
Note:	*p<0.1; **p<0.05; ***p<0.01

4. Plot the marginal effects of education (Black vs. White).

   Ans: Please find the plot of the marginal effects of education by race in *Figure 2* below. This plot will also be saved as "ame_race_educ.png" when you run the code.

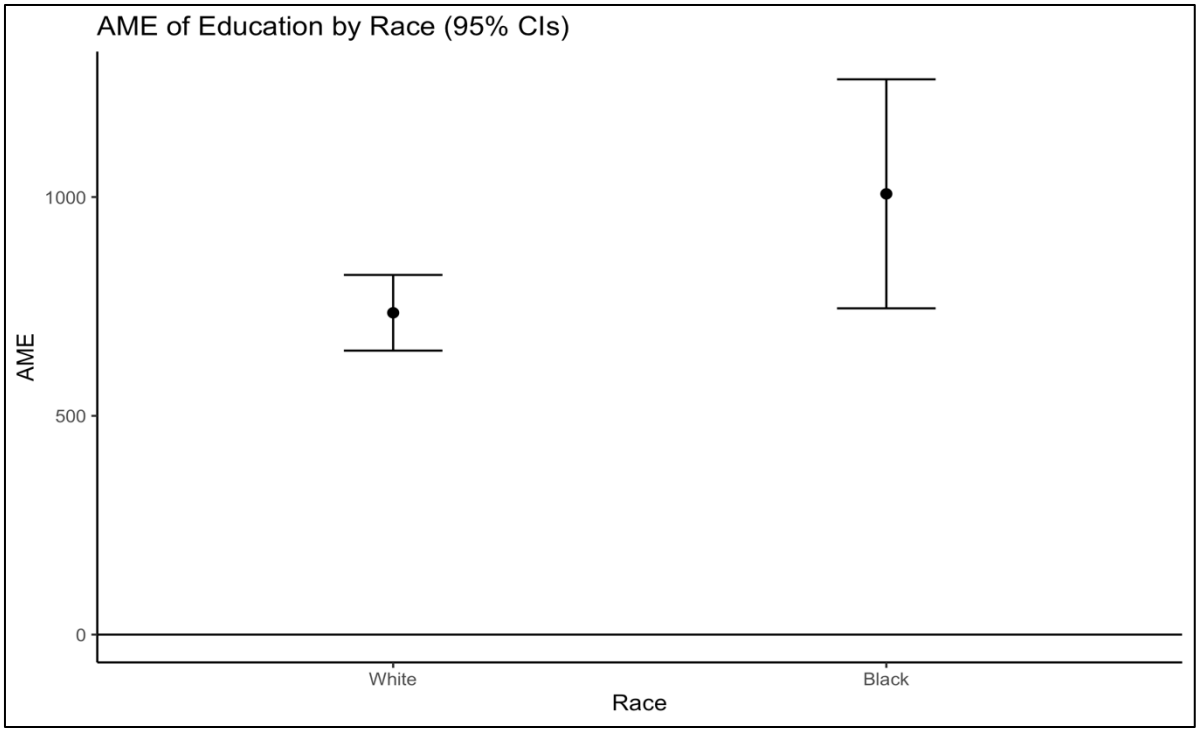

Figure 2

5. Interpret these marginal effects plot of education. Write a hypothesis that this marginal effect plot empirically supports.

   Ans: From the marginal effects plot of education by race in *Figure 2*, we can hypothesise that the effect of education on annual income varies by race to some extent. Additionally, the effect of education on annual income is positive for white race and it is stronger for the black race. This means that for a white individual, the effect of education on income will be *USD 735.5* whereas the effect of education on income for a black individual will be greater by *USD 271.811* which comes to *USD 1007.3.* Please note

that the regression table 2 shows a bit of a difference in the values of these values (Education, Educ x Black), may be as it includes another variable, *Other_dummy* as well.
It is important to take into account that in *Table 2,* the effect of education on income by black race is not significant at 5% level but they are significant at 10%. This helps us understand that if we want to account for the variation of the effect of education on income by race, we will have to take it at 10% level.

3. Model 3 *(30 points)*

In Model 3, you examine whether the effect of education on income varies with the number of children, allowing for different slopes of education on income for individuals with different number of children.

Unlike Model 2, you do not consider variations in the effect of education on income across different races. Instead, you analyse differences in income among White, Black, and Other race groups by allowing for different intercepts (starting points) but not varying slopes of education on income across these groups. Using White as the reference group, you also examine whether the income of Black or Other race individuals differs from that of White individuals in Model 3.

1. Write the regression equation of Model 3 using symbols such as α, β_1 etc., rather than OLS coefficients obtained from the results.

Ans: The regression equation for Model 3 is as follows:

$$\text{Annual Income } (Y) = \alpha' + \beta_1' \, Years\ of\ Education\ (X) + \beta_2' \, Black\ Dummy + \beta_3' Other\ Dummy + \beta_4' \, Age + \beta_5' \, childs + \beta_6' \, (childs\ X\ Years\ of\ Education) + \boldsymbol{\mu} \quad \ldots\ldots\ldots\ldots\ldots \text{(II)}$$

where α' is the constant and β_1' denotes the effect of X on Y, keeping other variables constant. β_2' *and* β_3' denotes the relative effects of Black race category and Other race category to reference category, White, respectively, while keeping other variables constant. β_4' *and* β_5' represent the effects of age and number of children *(childs*) respectively, while keeping other variables constant. β_6' is the effect of number of children on education, holding other variables constant. The symbol $\boldsymbol{\mu}$ represents the random component of the dependent variable, income.

2. Write the formula for the marginal effects of education according to the regression equation above.

Ans: Please find below the marginal effects of education according to equation (II):

$$Marginal\ Effects\ of\ Education\ on\ Income\ by\ Number\ of\ Children = \beta_1' + \beta_6' \, childs$$

3. Conduct the appropriate multiple regression analysis to test your hypotheses and report the results in a table format.

Ans: Please find the results of my regression analysis between the income (DV) and years of education (IV) incorporating the effects of number of children on education along with the intercept effect coming from race (*black_dummy*) on education in Table 3 using the "\Stargazer" package. This analysis will also be saved as "model3.txt" when you run the code.

Table 3

Interaction: educ X childs

	Dependent variable:
	Annual Income (USD)
Education (Years)	651.999***
	(60.504)
Black dummy	-2,774.269***
	(360.709)
Other dummy	-585.428
	(440.724)
Age	-8.817
	(8.018)
Number of Children	-536.352*
	(282.056)
Educ X N Children	56.680***
	(21.138)
Constant	9,556.098***
	(919.645)
Observations	1,766
R2	0.206
Adjusted R2	0.203
Residual Std. Error	5,123.156 (df = 1759)
F Statistic	75.961*** (df = 6; 1759)
Note:	*p<0.1; **p<0.05; ***p<0.01

Variation in the starting points of income by race:

From Equation (II), lets check the effect of race on the intercept of Y:

a) For White race, the Y-intercept will start from α' that is the starting annual income for white race individuals will be *USD 9556.098*

b) For Black race, the Y-intercept will start from $\alpha' + \beta_2'$ that is the starting annual income for black race individuals will be *USD 9556.098 – USD 2,774.269 = USD 6781.829*

c) For Other race, the Y-intercept will start from $\alpha' + \beta_3'$ that is the starting annual income for black race individuals will be *USD 9556.098 – USD 585.428 = USD 8970.67*

These calculations suggest that the starting annual income of individuals from White race is higher than that of Black/Other race. One important thing to note is that the starting annual income of individuals from black category is at least *USD 2188.841* lower than the remaining two categories, making it the lowest amongst the three.

4. Plot the marginal effects of education by the number of children.

Ans: Please find the plot of the marginal effects of education by number of children in *Figure 3* below. This plot will also be saved as “ame_childs_educ.png” when you run the code.

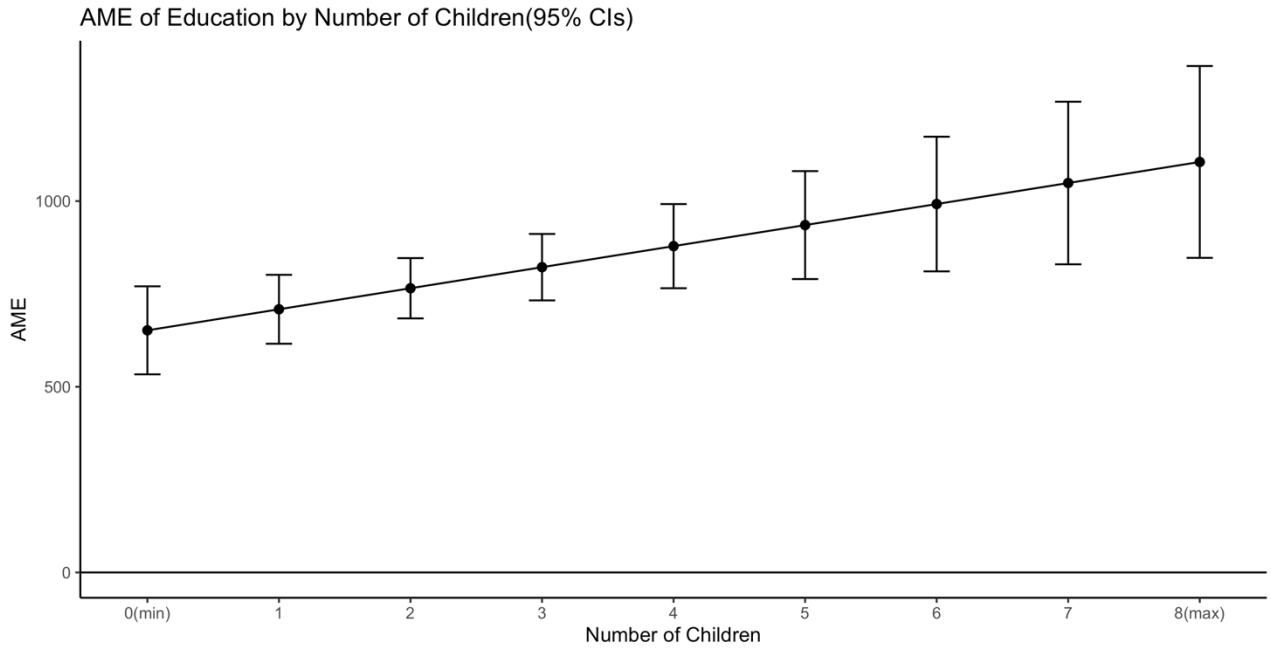

Figure 3

5. Interpret these marginal effects plot of education. Write a hypothesis that this marginal effect plot empirically supports.

Ans: From the marginal effects plot of education in *Figure 3,* we can hypothesise that the effect of education on annual income varies positively by number of children. In addition to that, the impact of *number of children* on education increases as the *number of children* grow.

The results of the regression in *Table 3* also show that the interactive relationship between education and number of children is statistically significant at 1% level. According to the results and the plot, we conclude that with a unit increase in *number of children,* the effect of education on annual income will increase by *USD 56.680*. The marginal effect plot shows that at number of children equals 0 (min), the effect of education on income is *USD 652* and with every unit increase in number of children, the effect of education on income increases by *USD 56.780* with its highest value, *USD 1105.4* at number of children equals 8.