Question 1.

Effects of Physical Attractiveness on Wage

Hamermesh and Biddle (1994) used measures of physical attractiveness in a wage equation. Each person in the sample was ranked by an interviewer for physical attractiveness using five categories (homely, quite plain, average, good looking, and strikingly beautiful or handsome). Because there are so few people at the two extremes, the authors put people into one of three groups for the regression analysis: "average", "below average", and "above average", where **the base or reference group is "average"**. Using data from the 1977 Quality of Employment Survey, after controlling for the usual productivity characteristics, the following two regressions were estimated using data on n = 1,260:

Estimate the model (1.1) reports in the Table 1.1

$$\log(wage_i) = \beta_1 + \beta_2 educ_i + \beta_3 exper_i + \beta_4 expersq_i + \beta_5 union_i + \beta_6 female_i + u_i \qquad \textbf{(1.1)}$$

Table 1.1

Source	SS	df	MS			
				Number of obs	=	1,260
				F(5, 1254)	=	149.25
Model	166.011417	5	33.2022834	Prob > F	=	0.0000
Residual	278.96855	1,254	.222462959	R-squared	=	0.3731
				Adj R-squared	=	0.3706
Total	444.979967	1,259	.353439211	Root MSE	=	.47166

lwage	Coefficient	Std. err.	t	P>|t|	[95% conf. interval]
educ	.0708503	.0052325	Omitted for the purpose of this exam		
exper	.0389808	.0043524			
expersq	-.0005986	.0000975			
union	.1924593	.0301994			
female	-.4421609	.0289766			
_cons	.443611	.078859			

Estimate the model (1.2) reports in the Table 1.2

$$\log(wage_i) = \beta_1 + \beta_2 educ_i + \beta_3 exper_i + \beta_4 expersq_i + \beta_5 union_i + \beta_6 female_i + \beta_7 belavg_i + \beta_8 abvavg_i + u_i \qquad \textbf{(1.2)}$$

where $\log(wage_i)$ or $lwage$ = logarithm of hourly wage (in USD)

- $educ_i$ = years of schooling
- $exper_i$ = years of workforce experience
- $expersq_i$ = years of workforce experience squared
- $union_i$ = 1 if union member
- $female_i$ = 1 if female
- $belavg_i$ = 1 if in below average physical attractiveness
- $abvavg_i$ = 1 if in above average physical attractiveness

Table 1.2

Source	SS	df	MS			
				Number of obs	=	1,260
				F(7, 1252)	=	109.21
Model	168.697151	7	24.099593	Prob > F	=	0.0000
Residual	276.282816	1,252	.220673176	R-squared	=	0.3791
				Adj R-squared	=	0.3756
Total	444.979967	1,259	.353439211	Root MSE	=	.46976

lwage	Coefficient	Std. err.	t	P>|t|	[95% conf. interval]
educ	.0691306	.00525	Omitted for the purpose of this exam		
exper	.0395785	.0043428			
expersq	-.0006081	.0000971			
union	.1884632	.0301843			
female	-.4388235	.028877			
belavg	-.1388291	.0417749			
abvavg	.0070104	.0302809			
_cons	.4737302	.0795614			

Answer the following questions.

1.a) Based on the regression results provided, write out the estimated coefficients in the form of regression equation (1.1). Interpret the estimated coefficients associated with $educ_i$. Based on Model (1.1), test whether education has an impact on logarithm of hourly wage. Show your work. (Use α = 0.05)

1.b) What is the overall significance of the regression from Model (1.2)? What test do you use? (Use α = 0.05)

1.c) If we are interested in testing whether "physical attractiveness" has an impact on logarithm of hourly wage at all, what kind of null/alternative hypothesis would we be testing? Perform the test and discuss your finding. (Use α = 0.05)

1.d) Is there convincing evidence that women with above average looks earn more than women with average looks? Explain.

1.a $\log(wage) = \beta_1 + \beta_2 educ_i + \beta_3 exper_i + \beta_4 expersq_i + \beta_5 union_i + \beta_6 female + \beta_7 belavg + \beta_8 abvavg_i + u_i$

log (wage) change β_2 (0.0708503) %. when education change 1%.

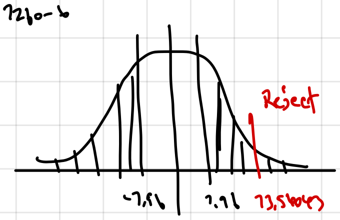
for this case we can say that when education increase by 1%.
log(wage) increase 0.0708503%. in average

$H_0 : \beta_2 = 0$
$H_1 : \beta_2 \neq 0$

$$t_{cal} = \frac{\hat{\beta}_2 - \beta_2}{se\,\hat{\beta}_2} = \frac{0.0708503}{0.0052325} = 13.54043$$

$t_{\frac{\alpha}{2}, n-k} = \pm 1.96$

0.025 1260-6



We can reject H_0 that β_2 is not zero 95 out of 100

it shows that education impact l(wage) with 0.05 significant level

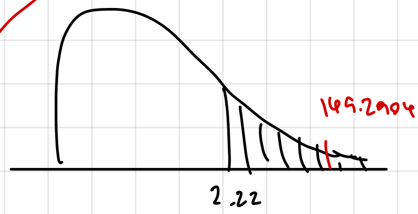
1.b to test the overall significance, we rely on F-test

H_0 : All the β_k are simultaneously equal to zero

H_1 : otherwise

$$F_{cal} = \frac{\frac{ESS}{k-1}}{\frac{RSS}{n-k}} = \frac{\frac{166.011417}{6-1}}{\frac{271.96855}{1210-6}} = \frac{33.2022}{0.2224} = 149.2904$$

$F_{(0.05,\ 6-1,\ 1210-6)} = 2.2212$



∴ since $F_{cal} > F_{critical}$ therefore we can reject H_0

we can say that all coefficient are not simultaneously zero 95 out of 100 times #

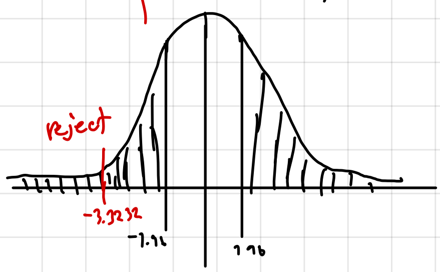
1.C H_0 : $\beta_{belavg} = 0$

H_1 : $\beta_{belavg} \neq 0$

$$t_{cal} = \frac{-0.1388291}{0.0417749} = -3.3232$$

$t_{\frac{\alpha}{2}, n-k} = \pm 1.96$

$\frac{0.05}{2}, 1252$



∴ since t_{cal} falls in rejection area, we can reject H_0

Below average physical attractiveness has an impact on lwage with 0.05 significance #

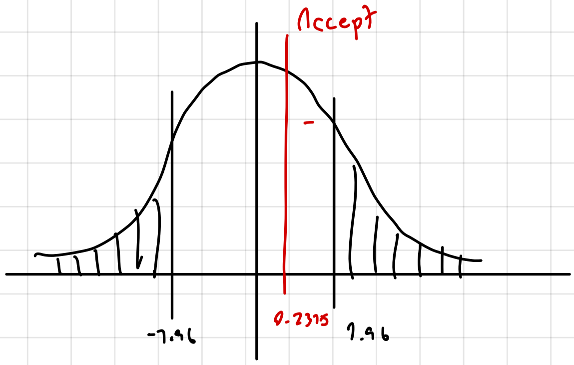
H_0 : $\beta_{abvavg} = 0$

H_1 : $\beta_{abvavg} \neq 0$

$$t_{cal} = \frac{0.0070704}{0.0302809} = 0.2315$$

$t_{\frac{\alpha}{2}, n-k} = \pm 1.96$

$\frac{0.05}{2}, 1252$



∴ since t_{cal} falls in acceptance area, we cannot reject H_0

Above average physical attractiveness has no impact on lwage with 0.05 significance #

∴ we can conclude that a person who is physically attractive doesn't get higer wage

On the other hand, a person who is less physically attractive than average is likely to get lower wage than others. #

1.D According to 1.C, we test the hypothesis wheather physical attractiveness has an impact on lwage.

We found that those with above average physical attractiveness do not earn more wage.

Therefore, there is no convincing evidence that suggest women with above average physical attractiveness earn more wage than those with average attractiveness #

Question 2.

A household expenditure model is given by

$$hhexp_i = \beta_1 + \beta_2 area_i + \beta_3 child_i + u_i$$

0 ไม่อยู่คอนโด

1 อยู่คอนโด

where $hhexp_i$ = household expenditure per month

$area_i$ = a dummy variable for household location: (0 if in a municipal area and 1 if otherwise)

$child_i$ = number of children in household i, aged under 15

Using socio-economic dataset collected in 2018 with 14,908 households, the result is given below with **t value in parentheses**. Answer the following questions.

$$\widehat{hhexp}_i = \underset{(43.83)}{9{,}736} - \underset{(-15.8)}{2{,}835} area_i + \underset{(6.82)}{881} child_i + \hat{u}_i$$

2.a) Do all the signs for each coefficient make economic sense? Explain.

2.b) Test each parameter separately if they are significantly different from zero or not. (Use α = 0.01)

2.c) Find the expected value of a household expenditure not living in a municipal area with 3 children aged under 15.

2.d) When an interaction term is included in this model, the result becomes with **t value in parentheses**.

$$\widehat{hhexp}_i = \underset{(34.38)}{9{,}693} - \underset{(-6.55)}{2{,}742} area_i + \underset{(5.17)}{910} child_i - \underset{(-0.25)}{64} (area_i * child_i) + \hat{u}_i$$

Draw a diagram for this model displaying sampled regression functions (SRF) with expected value of household expenditure on the vertical axis and number of children on the horizontal axis, taking **only significant parameter(s)** into account. Indicate the intercept and slope for each SRF where applicable. Testing of significance can be shortened.

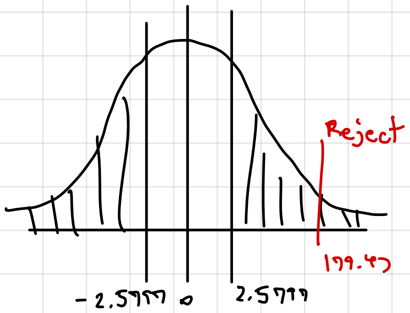
2.a It makes economically sense, because there is no need to pay public service and facility fee in a municipal area, for example; waste management, public maintainance, site drainage system and etc. Therefore, total household expenditure in a municipal area is likely to be lower than in the city. #

2.b $H_0: \beta_2 = 0$

$H_1: \beta_2 \neq 0$

$t_{cal} = \dfrac{-2{,}835}{-15.8} = 179.43$

$t_{\frac{0.01}{2},\, 14908-2} = \pm 2.5797$



∴ Since, tcal falls in Rejection area, we can reject H_0.

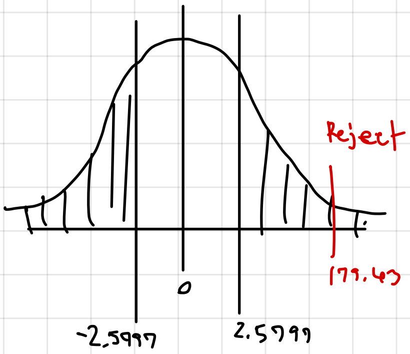
It can be concluded that area has an impact on household expense

$H_0: \beta_3 = 0$

$H_1: \beta_3 \neq 0$

$t_{cal} = \dfrac{881}{6.82} = 129.1788$

$t_{\frac{0.01}{2},\, 14908-2} = \pm 2.5797$



∴ Since, tcal falls in Rejection area, we can reject H_0.

It can be concluded that children have an impact on household expense.

2.c 3 children not living in a municipal area

$\widehat{hhexp_i} = 9{,}936 - 2{,}835(1) + 881(3)$

$\widehat{hhexp_i} = 9{,}936 - 2{,}835 + 2{,}643$

$\widehat{hhexp_i} = 9{,}544$

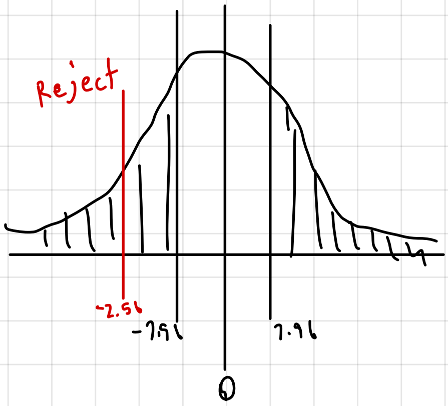
$\therefore$ The expected value of a household expenditure not living in a municipal area with 3 children equals to 9,544 #

2.d H_0 : area * child $= 0$

H_1 : area * child $\neq 0$

$t_{cal} = \frac{-64}{0.25} = -2.56$

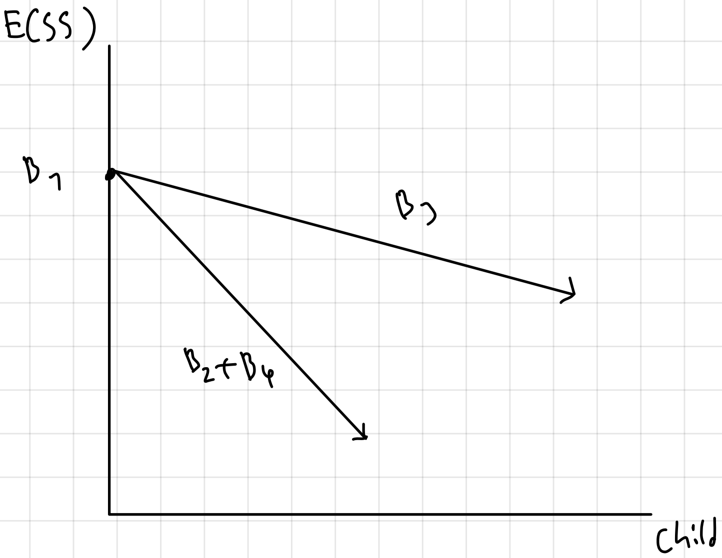
$t_{crit} = \pm 1.96$



We can reject H_0

It shows that area * child coefficient is significantly different from zero in area has different slope

Compared to outarea=0 affect some squar for inarea=1 and outarea=0 differently



Question 3.

Assume a multiple linear regression model as

$$hours_i = \beta_1 + \beta_2 sex_i + \beta_3 age_i + \beta_4 agesq_i + \beta_5 weekot_i + u_i$$

where $hours_i$ is hours worked in a week
sex_i is a dummy variable: 0 = male and 1 = otherwise
age_i is age of observation i
$agesq_i$ is age square observation i
$weekot_i$ is nominal overtime paid per week

Answer the following questions.

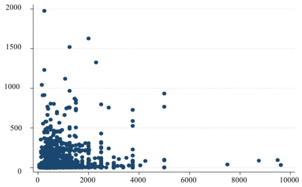
3.a) A VIF and tolerance table (postestimation) is given below

Variable	VIF	1/VIF
2.sex	1.02	0.979129
age	50.61	0.019759
agesq	50.68	0.019731
weekot	1.01	0.985618
Mean VIF	25.83	

Given that you are exploring multicollinearity assumption, which pair of variables that you suspect they might be linearly correlated? Provide clear explanation what criteria (ion) that you rely on making that judgement.

3.b) From **(3.a)**, do you consider removing one of the variables from the model? Why or why not and which one that you choose to remove, if that is the case?

3.c) The graph provided below is a scatter plot between $\hat{u}_i^2$ (vertical axis) and $weekot_i$ (horizontal axis). Using the graphical method, do you conclude that heteroscedasticity is present in this model or not. Explain clearly to support your answer.



3.d) An auxiliary model here is estimated and the result is given in the table below.

$$\hat{u}_i^2 = \beta_1 + \beta_2 sex_i + \beta_3 age_i + \beta_4 agesq_i + \beta_5 weekot_i + v_i$$

Source	SS	df	MS
Model	829063.863	4	207265.966
Residual	44148135	2,027	21780.037
Total	44977198.8	2,031	22145.3465

Number of obs	=	2,032
F(4, 2027)	=	9.52
Prob > F	=	0.0000
R-squared	=	0.0184
Adj R-squared	=	0.0165
Root MSE	=	147.58

uhat2	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
2.sex	-5.648899	6.630832	-0.85	0.394	-18.65286	7.355058
age	-2.490434	2.37094	-1.05	0.294	-7.140168	2.1593
age2	.044175	.0301279	1.47	0.143	-.0149098	.1032599
weekot	.0229916	.0043502	5.29	0.000	.0144603	.0315229
_cons	83.8484	44.4418	1.89	0.059	-3.307973	171.0048

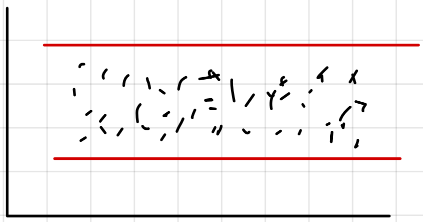
From the table, setup the hypotheses and perform the Breusch-Pagan test to check that heteroscedasticity is present in the original model or not.

****

3.a age and agesq should be less than 10 and TOL should be near 1 from rule of thump. Therefore, age and agesq may be linearly correlated

3.b One of the variables, either age or agesq, should be eliminated if they are highly colinear,

3.c This model we conclude that heterocedasticity is not present in this model because it is not like this picture



linear

3d H_0 : homoscedasticity

H_1 : heterocedasticity

$$F_{cal} = \frac{R^2_{\hat{v}}/k}{(1-R^2_{\hat{u}_i})/(n-k-1)} = \frac{0.0184/4}{(1-0.0184)/(2032-4-1)} = 9.498$$

$F(4, 2027) = 9.52$

$F_{cal} < F_\alpha$

We cannot reject H_0 that heteroscedasticity is not in the original model