Assignment 2
Assigned on Nov 9th, 2021. Due date Nov 25th, 2021 before midnight.

Instructions

(1) Please read the instruction carefully. Also take this habit with you into the exam room.
(2) Please read each question carefully and answer the questions straightforwardly. Always provide economic reasons at lease a paragraph for your analysis, or a graph when necessary, even when the question does not indicate so.
(3) Handing and submitting assignments are only available via BE Moodle.

Answering the questions and preparing answer sheets

(1) Answers are to be handwritten, in either digital or analog form, in a blank canvas or any clean paper. Make sure that your handwriting is clearly visible and readable.
(2) There is no need to rewrite the question. Just indicate the question number clearly for each of the answer, such as 1.a).
(3) Default decimal point is 4.
(4) Choose precise wordings, especially when you want to interpret the meaning of a test, confidence interval, or coefficients.
(5) When done, for the digital case, collage all the pages into a single PDF file. For those who write on sheets of paper, take photo of all pages then convert all of them into a single PDF file as well.
(6) Name your PDF file as StudentID_YourNickname, such as 640123456_Bo.

Submitting your answers

(1) Make sure your file does not exceed 10MB. This is the maximum file size for BE Moodle upload.
(2) Login to BE Moodle, head into the course, then the assignment topic.
(3) Choose your file to submit. Done. There will be timestamp for your upload date and time, so please make sure to not submit later than that.

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
Model	166.011417	5	33.2022834
Residual	278.96855	1,254	.222462959
Total	444.979967	1,259	.353439211

Number of obs	=	1,260
F(5, 1254)	=	149.25
Prob > F	=	0.0000
R-squared	=	0.3731
Adj R-squared	=	0.3706
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 $\alpha = 0.05$)

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

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 $\alpha = 0.05$)

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

Question 2.

A household expenditure model is given by

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

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.

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.

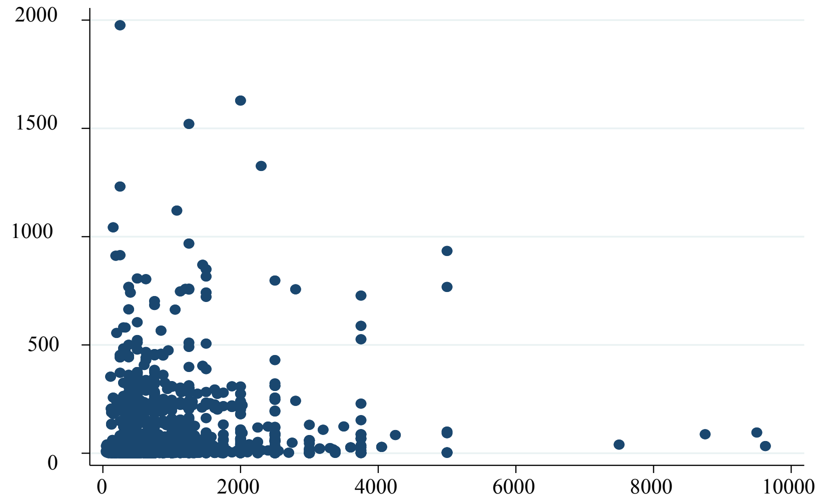
3.a) A VIF and tolerance table (postestimation) is given below

```
    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      Number of obs   =     2,032
-------------+----------------------------------   F(4, 2027)      =      9.52
       Model |  829063.863         4  207265.966   Prob > F        =    0.0000
    Residual |    44148135     2,027   21780.037   R-squared       =    0.0184
-------------+----------------------------------   Adj R-squared   =    0.0165
       Total |  44977198.8     2,031  22145.3465   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.

****

1a.) ① $\log(wage_i) = 0.4436 + 0.0708\, educ_i + 0.03898\, exper_i - 0.000548\, expersq + 0.1925\, union_i - 0.044216\, female_i$

② if years of school increase by 1 year, logarithm of hourly wage will increase by 0.0708 dollar

③ $H_0;\ \beta_2 = 0$; null hypothesis
$H_a;\ \beta_2 \neq 0$

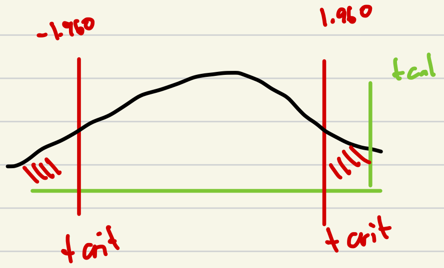
- $t\,cal = \dfrac{\hat{\beta}_2 - \beta_2}{se\,\hat{\beta}_2}$

$= \dfrac{0.07085 - 0}{0.00523}$

$= 13.5468$ //

- $\alpha = 0.05$ d.f. = 1254

$t\,critical = 1.960$ ✓



∴ We can reject null hypothesis and we can make sure 95% that education impact on logarithm of wage.

1.b.) Used: F-test
$H_0;\ \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = 0$
$H_a;$ otherwise

$F\,cal: \dfrac{ESS/k-1}{RSS/n-k} = \dfrac{168.69915/7}{276.282816/1252} = 109.5436$

$\alpha = 0.05$
$F_{cri}\,(7, 1252) = 2.0096$

∴ We can reject null hypothesis and we can make sure 95% that overall variables are significant in 1.2 model.

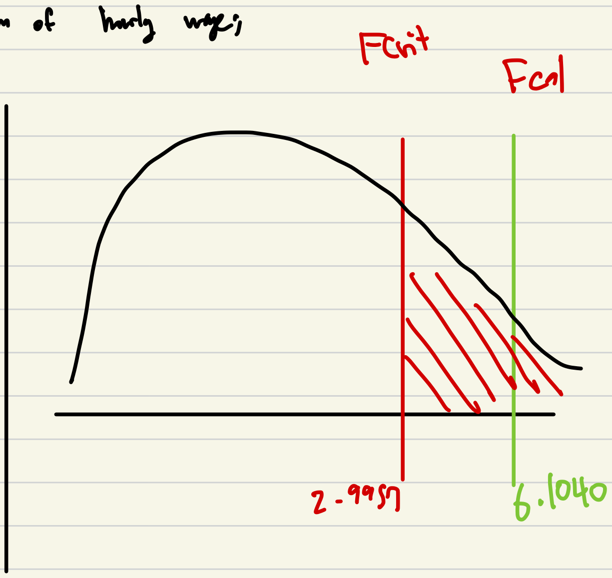
1c.) : F-test ; marginal contribution
H_0: Physical attractiveness has no impact on logarithm of hourly wage;
H_a; otherwise

$F\,cal = \dfrac{ESS_{1.2} - ESS_{1.1} / 2}{RSS_{1.2} / n-k_{1.2}}$

$= \dfrac{(168.699151) - (166.011412) / 2}{(296.282816) / 1260 - 8}$

$= \dfrac{1.3429}{0.22} = 6.1040$ //

$\alpha = 0.05$
$F_{cri}\,(2, 1252) = 2.9954$



1 d.) 0 woman /average locks → belowavg$_i$ = 0 aboveavg$_i$ = 0

log(wage) = 0.4737 + β_2 educ$_i$ + β_3 exper$_i$ + β_4 expersq$_i$ + β_5 union$_i$ - 0.4388235 (1) + 0 + 0
log (wage) = 0.034877 + β_3 exper$_i$ + β_4 expersq$_i$ + β_5 union - woman with average looks

women (1), above average (1)

log (wage) = 0.4737 + β_2 educ$_i$ + β_3 exper$_i$ + β_4 expersq$_i$ + β_5 union$_i$ - 0.4388235 (1)
0 + 0.0070104 (1)
log (wage) = 0.45071 + β_2 educ$_i$ + β_3 exper$_i$ + β_4 expersq$_i$ + β_5 union$_i$ - woman with above wage looks

∴ explain : the intercept of two groups are different, which woman with above rate looks have higher value of intercept.

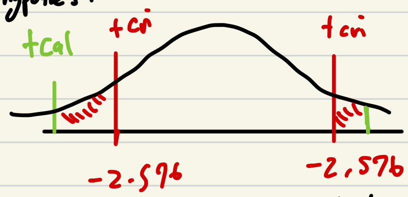
2 a.) β_2 is negative value, this means not living in municiple have lower expense by 2,835 this make economical sense because people who live in municiple area are more likely to have less expense due to lower income and lower cost of living

β_3 is positive, this means the more children's family have, the more expense in that house. This is obvious that more people consume food and money for living.

2b.)

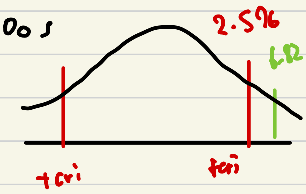
$$\widehat{hhexp_i} = 9{,}736 - 2{,}835 area_i \ + 881 child_i + \hat{u}_i$$
(43.83) (−15.8) (6.82)

Ho: $\beta_2 = 0$ - null hypothesis
H₁: $\beta_2 \neq 0$
α = 0.01 → $\frac{\alpha}{2}$ = 0.005
t cri = 2.576
df = 14908 - 3
= 14905



∴ we can reject Ho

α = 0.01 → $\frac{\alpha}{2}$ = 0.005
tcri = 2.576
df = 14 908 - 3
= 14905



2.c) HH expenditure municipal area = 1, no children = 3

$$\widehat{hhexp_i} = 9{,}736 - 2{,}835 area_i \ + 881 child_i + \hat{u}_i$$

= 9736 - 2835 + 881(3)
= 9544

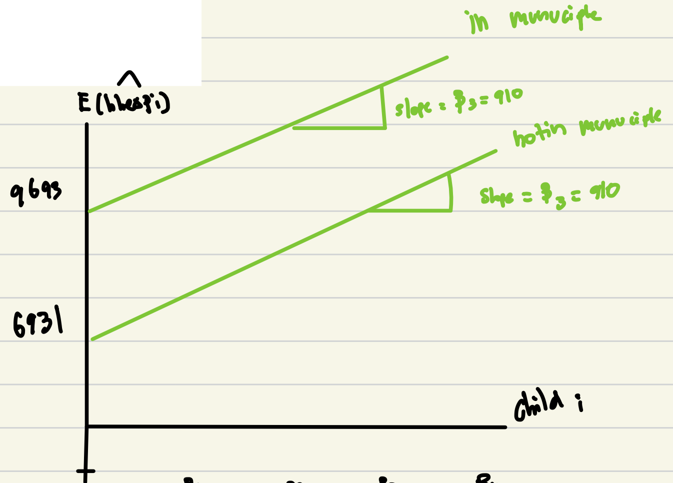
children aged under 15.

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

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

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.

$H_0: \hat{\beta}_2 = 0$	$H_0: \beta_3 = 0$	$H_0: \beta_4 = 0$
$H_a: \hat{\beta}_2 \neq 0$	$H_a: \beta_3 \neq 0$	$H_a: \beta_4 \neq 0$
$t_{cal} = -6.55$	$t_{cal} = 5.17$	$t_{cal} = -0.25$
$t_{cri} = 2.576$	$t_{cri} = 2.576$	$t_{cri} = 2.576$
$\alpha = 0.01$ can reject H_0	can reject H_0	cannot reject H_0
$\frac{\alpha}{2} = 0.05$		



β1 β2 β3 β4

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

$\widehat{hhexp}_i = 9693 - 2742(0) + 910(0) = 9693$; base case: in municipality

$\widehat{hhexp}_i = 9693 - 2742(1) + 910(0) = 6951$; no kid. not in municipality

$\widehat{hhexp}_i = 9693 - 2742(0) + 910(1) = 10603$ or $\beta_1 + \beta_3$ (child)

$\widehat{hhexp}_i = 9693 - 2742(1) + 910(1) = 7797$

$= (\beta_1 - \beta_2) + (\beta_3)(child_i)$

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.

: age and agesq are suspected to be linearly correlation. Because the VIF value of these two variables are exceed 10 and TOL value or $\frac{1}{VIF}$ are nearly 0. This means r^2 between these 2 independences are high.

3.b.) No, we don't have other data to make sure that which variable should be eliminated.

3c.) yes, heteroscedasticity is detected in this model because as week of increase. There are increase in $\hat{u}$ as well.

3.D)

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.

****

Ho: the model is homoscedasticity

Ha: otherwise

$$F_{cal} = \frac{R^2_{\hat{u}_i^2} / (k)}{(1-R^2_{\hat{u}_i^2}) / (n-k-1)} = \frac{(0.0184)/5}{(1-0.0184)/(2032-5-1)} = \frac{0.00368}{0.00048} = 7.6667$$

$F_{cri}(5, 2026) = 2.2141$

$\alpha = 0.05$

$\therefore F_{cal} > F_{cri}$

$\therefore$ we can reject Ho, this means that as time of observation, heteroscedast is present.