1. If age increase from 25 to 26 when all other variables are the same we have an increase of ahe of 0.31. The same for 33 to 34.

2. If age increase from 25 to 26 when all other variables are the same we have an increase of ahe of 2%. The same for 33 to 34.

3. In this case has no sense explain the age increase from 25 to 26 because we used a log-log scale that explain the increase by 1% of age and not by 1 unit. If we want to interpreter the coefficient before age we can say that if age increase by 1%, ahe increase by 0.006369

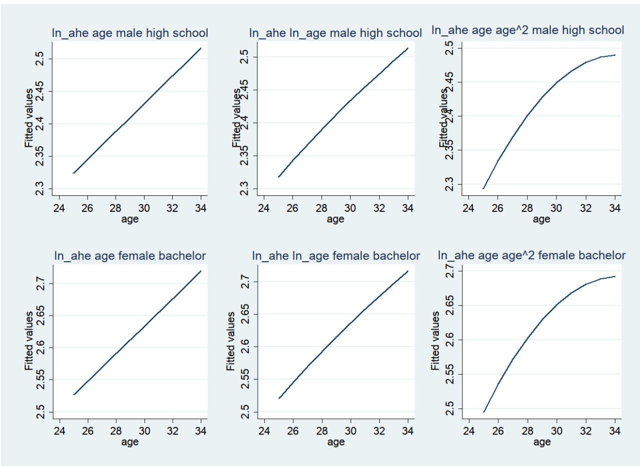
4. Since when age increase by one the variable age^2 also increase we cannot interpreter the coefficient before age

5. Between the 3 and 2 we prefer the 2 because the explain the increase of age by 1 $unit$ instead of 1%

6. We prefer the 2 because the 4 with age^2 doesn't explain anything more than 2

7. We prefer the 3 for the same reason of the point 5 of this exercise

8. If we see the first two plots are very similar, maybe the second line is a little bit curved while the third with age^2 is very curved at the end. The form of the plots for the female with bachelor are basically the same except for the range of values of y that is between $2.50 - 2.70$ in contrast to $2.32 - 2.5$ for the male with high school



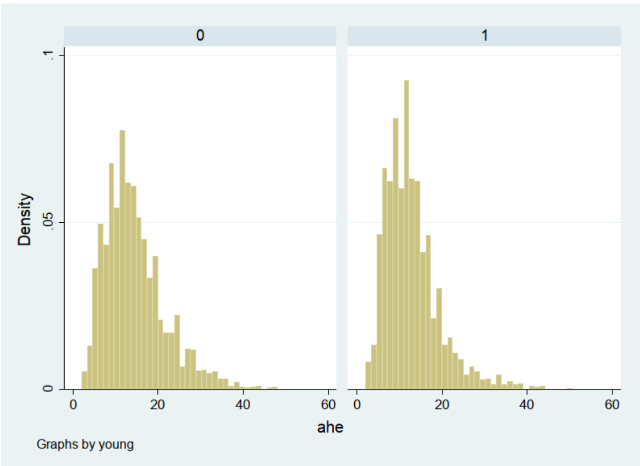
9. The interaction coefficient say that if we have two $female$ with the same age, one with $bachelor$ and one without, the first have an increase of ahe by $0.345 + 0.084 = 0.43 = 43\%$.
If we increase by one the age Jane has a $\ln(AHE) = 2.63$ and $AHE = 13.92$ while Alexis has a $\ln(AHE) = 2.20$ and a $AHE = 9.05$. The difference between the two female is $2.63 - 2.20 = 0.43$ that is exactly the interpretation given before. Bob has a $\ln(AHE) = 2.77$ and $AHE = 15.95$ while Jim has a $\ln(AHE) = 2.421$ and a $AHE = 11.26$. The difference

between the two female is $2.77 - 2.421 = 0.349$ that is exactly the bachelor coefficient but without the interaction because $female = 0$

10. First we centre the age values. Then making a regression with the age_center, $female$ and the interaction between the last two, we obtain a coefficient equal to -0.0070067 and a p-value of 0.124, so there isn't statistical significance for this value then we can't use this value for regression. In conclusion we can say that there isn't a difference on earning between male and female in relation to age

11. We used the age centred in the mean calculated in the latter exercise. In this case the regression use the variables $age\_center, bachelor$ and the interaction between the last two. The coefficient of the interaction is equal to 0.00469 with a $p-value = 0.272$, so also in this case we haven't enough statistical significance to use this value, then we can't assert that there is an effect between age and $bachelor$ variables in relation to earning

12. First, it is created a dummy variable $young$ that has value 1 if the age of the worker is less or equal to 28, else 0. Then with a log-linear regression on ahe with variables $age, young$. The results are:



Source	SS	df	MS
Model	257.861093	3	85.9536976
Residual	1178.27561	5907	.199471069
Total	1436.1367	5910	.243001133

Number of obs =	5911
F(3, 5907) =	430.91
Prob > F =	0.0000
R-squared =	0.1796
Adj R-squared =	0.1791
Root MSE =	.44662

ln_ahe	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
young	-.1099861	.0120547	-9.12	0.000	-.1336178	-.0863544
female	-.1796929	.0118448	-15.17	0.000	-.2029131	-.1564727
bachelor	.3814208	.0117502	32.46	0.000	.3583861	.4044555
_cons	2.465655	.0099196	248.56	0.000	2.446209	2.485101

We can clearly see that the coefficient of $young$ variable is negative, so it mean that if a worker is young, it has a decrease by 10% of AHE