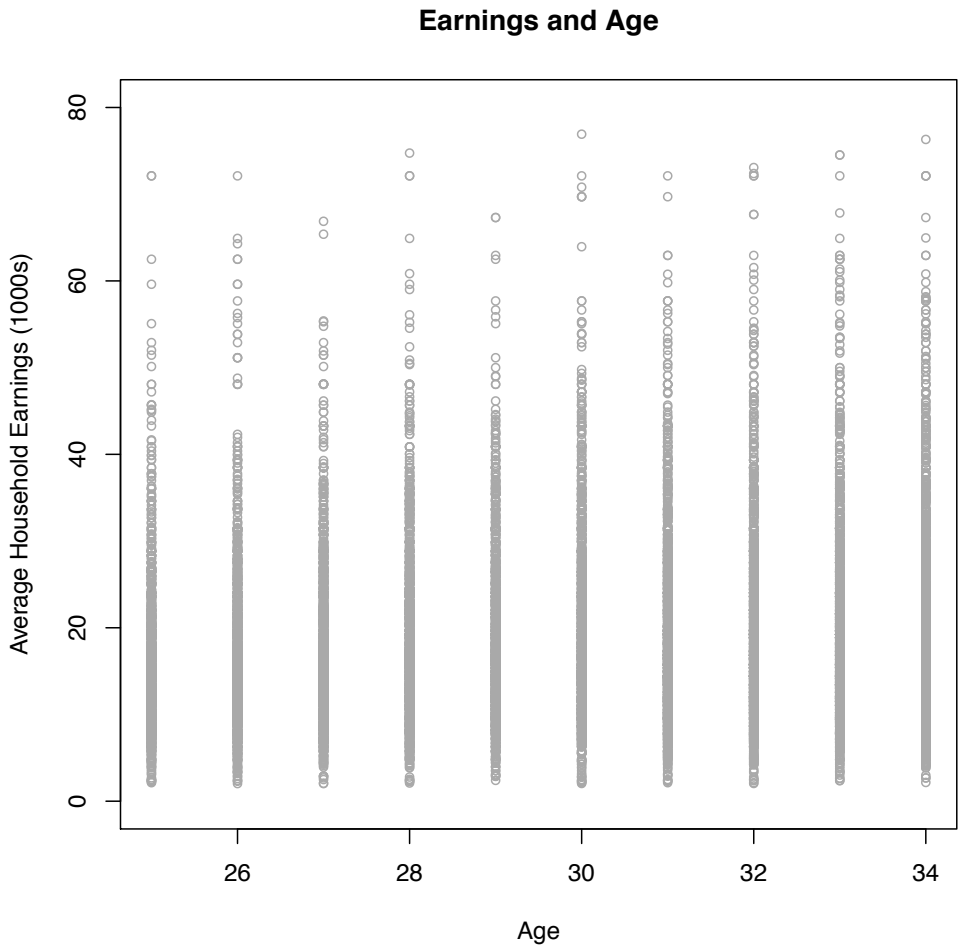
ECOM20001: Econometrics 1

Tutorial 10: Suggested Solutions

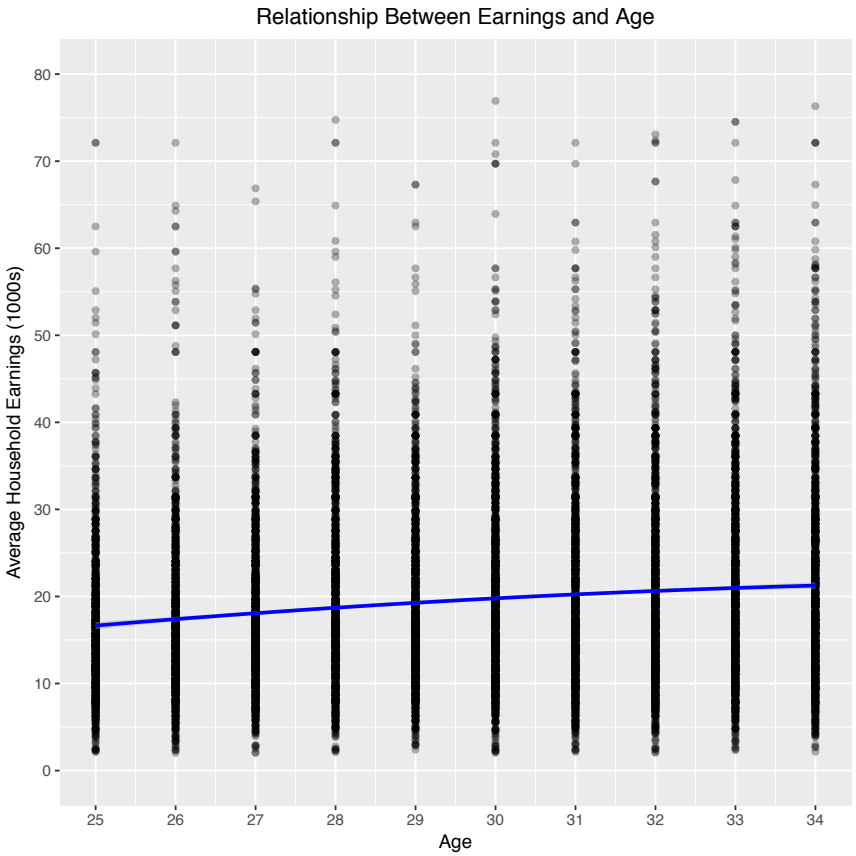
---

1. The graphs are plotted below. There appears to be a quadratic relationship between age and ahe. Intuitively, at younger ages earnings rise faster as people initially progress through their career within their jobs and moving across jobs. As they get older however, people settle into their jobs, and income growth stabilises as their labour markets stable.

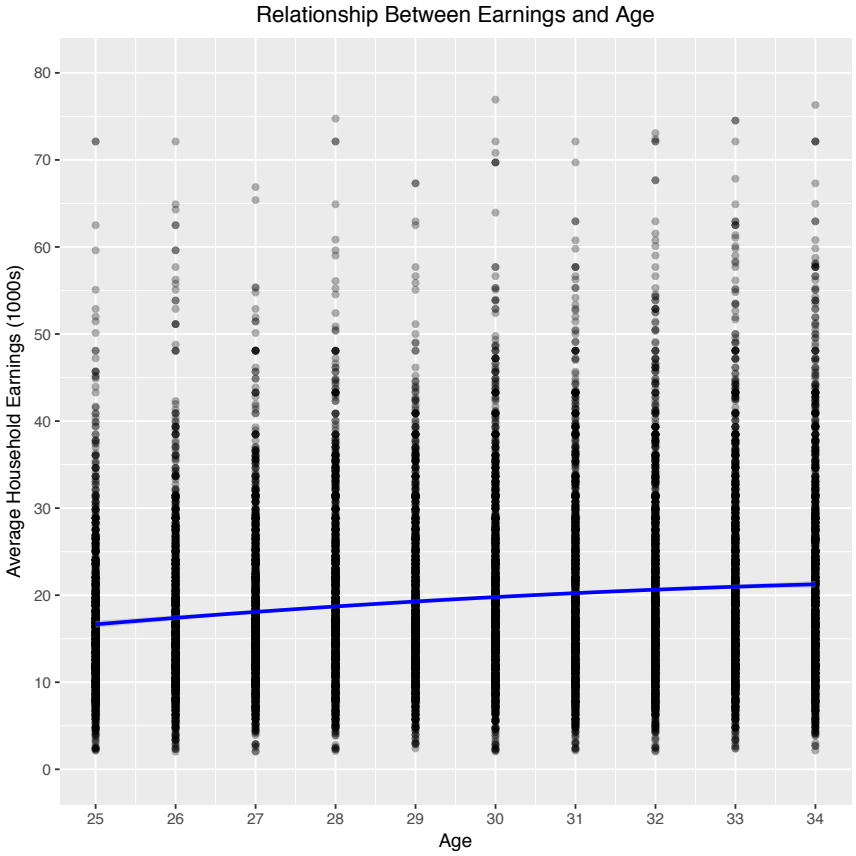
Scatterplot with the plot() command:



Scatterplot with the ggplot() command and a **quadratic** curve fit to the data:



Scatterplot with the ggplot() command and a **cubic** curve fit to the data:



2. Regression results with heteroskedasticity-robust standard errors are contained in the following table:

	Dependent variable:				
	Annual Household Earnings (AHE)				
	(1)	(2)	(3)	(4)	(5)
Age	−212.59	10.04	1.98***	0.54***	2.24***
	(182.20)	(10.81)	(0.60)	(0.03)	(0.66)
Age Squared	11.10	−0.30	−0.02**		−0.03***
	(9.34)	(0.37)	(0.01)		(0.01)
Age Cubed	−0.26	0.003			
	(0.21)	(0.004)			
Age Quartic	0.002				
	(0.002)				
Bachelor Degree	7.75***	7.75***	7.75***	7.75***	
	(0.15)	(0.15)	(0.15)	(0.15)	
Female	−3.53***	−3.53***	−3.53***	−3.53***	
	(0.15)	(0.15)	(0.15)	(0.15)	
Constant	1,525.86	−98.28	−19.79**	1.28*	−21.05**
	(1,327.78)	(105.11)	(8.73)	(0.77)	(9.57)
Observations	15,052	15,052	15,052	15,052	15,052
R2	0.19	0.19	0.19	0.19	0.02
Adjusted R2	0.19	0.19	0.19	0.19	0.02
Residual Std. Error	8.98 (df = 15045)	8.98 (df = 15046)	8.98 (df = 15047)	8.99 (df = 15048)	9.86 (df = 15049)
F Statistic	580.10*** (df = 6; 15045)	695.81*** (df = 5; 15046)	869.65*** (df = 4; 15047)	1,157.28*** (df = 3; 15048)	161.27*** (df = 2; 15049)
Note:					*p<0.1; **p<0.05; ***p<0.01

- Sequential hypothesis testing suggests using a quadratic model as the cubic and quartic regressors in age are statistically insignificant.
- The overall F-statistic for the quadratic regression is F=784.73 with df1=4 and df2=15047, and associated p-value<0.00001. The test result implies that we reject the null at the 1% level of significance that the model has no ability to explain variation in ahe. In other words, the quadratic regression is useful, statistically, for explaining ahe.
- There's clearly an imperfect collinearity problem with the cubic and quartic regressions. This can be seen by how the standard errors blow up on the regression coefficients for age, age^2, age^3, age^4 when either age^3 or age^4 is included in the regression.
  - This is a very common problem to be aware of when formulating and estimating polynomial regressions: higher-order polynomials can lead to imperfect multicollinearity problems when there is not sufficient curvature in the relationship of interest (in our case age and ahe) for the data to allow one to estimate a higher-order polynomial (like cubic or quartic in our case).
  - Also notable is how the quartic regression coefficients themselves appear to explode compared to the cubic, quadratic and linear models. This reflects the fact that the collinearity problem is so severe with the

quartic model is basically bordering on a perfect collinearity problem, and is causing the OLS estimator to completely break down.

- In contrast, there is no problem of imperfect multicollinearity with the quadratic and linear models.

- The t-statistic and p-value on the age^2 coefficient in the quadratic reject are -2.40 and 0.017, respectively. That is, we reject the null that the age^2 coefficient equals 0 at the 5% level of significance. This implies we reject the null of no nonlinearity in the relationship between age on ahe.

3. The regression results with only age, age^2 as regressors are presented in column (5) of the table on the previous page.
   - There are large changes in the coefficient on age (1.979 in column 3 to 2.242 in column 5), and on age^2 (-0.024 in column 3 to -0.029 in column 5).
   - We also see a substantial rise in the R-squared in column 3 to 0.188 from 0.021 in column 5, once bachelor and female are controlled for.
   - These large changes in the coefficients and R-Squared's suggests that the quadratic regression with bachelor and female as controls is the most appropriate model for the age on ahe relationship.
4. If we only focused on the coefficient on age for estimating the partial effect of increasing age from 25 to 28, and from 28 to 31 on ahe we would obtain a partial effect of 1.979 x 3 x $1000 = $5937 increase in earnings from either change. The incorrect partial effects are the same because working with a linear model assumes the same partial effects everywhere, no matter the level of age.
5. Partial effects on ahe are as follows:
   - increasing age from **25 to 28**: 2.072 or $2072
   - increasing age from **28 to 31**: 1.634 or $1634
   - increasing age from **31 to 35**: 1.498 or $1498
   - Consistent with the figures we saw in question 1 (particularly the ggplot() graph with the quadratic model), we see a positive partial effect that is diminishing with larger values of age. This is consistent with our parameter estimates in the quadratic regression model which has a positive coefficient on age (implying a positive partial effect) and a negative coefficient on age^2 (implying a diminishing partial effect).

- The partial effects results differ in questions 4. and 5. Because we are now accounting for the fact that the level of age influences the partial effect of age on ahe in our nonlinear quadratic regression model.

6. Standard errors and 95% CIs for the partial effect on age from:
   - increasing age from **25 to 28**:
     - SE=0.194 or $194
     - 95% CI=[1.692,2.451] or [$1692, $2451]
     - 95% CI width = 0.759 or $759
   - increasing age from **28 to 31**:
     - SE=0.078 or $78
     - 95% CI=[1.482,1.786] or [$1482, $1786]
     - 95% CI width = 0.304 or $304
   - increasing age from **31 to 35**:
     - SE=0.309 or $309
     - 95% CI=[0.892,2.103] or [$892, $2103]
     - 95% CI width = 1.211 or $1211
   - The 95% CI indeed differs across the three partial effects computed above. This again is a function of the nonlinearity in the regression model.
   - The confidence intervals are tighter around the partial effect for the change in age from 28 to 31 relative to the other partial effects because there are more data points in the sample in the 28-31 range that in the lower 25-28 range and the higher 31-35 range (which are the relatively more extreme age ranges in the sample relative to 28-31, which sits in the middle of the sample).