Tripos, Part 2A, Paper 3
Supervision 2

1. In order to understand the determinants of years of schooling among women in the UK, we collected a random sample of 857 individual females aged 28-38 from across the UK and gathered information on a number of background characteristics including their score on an IQ test. The summary statistics are as follows.

Table 1

Variable	Mean	Std Dev	Min	Max
educ	13.58	2.20	9	18
age	32.98	3.09	28	38
black	0.12	0.32	0	1
meduc	10.68	2.85	0	18
IQ	101.80	15.01	50	145

Here, educ represents years of schooling, age is recorded in years, meduc is number of years of schooling for the individual's mother and IQ is the score on an IQ test. Black is a dummy denting whether the woman has African origin. Define "leduc" to be the natural log of years of education.
The OLS regression output with *leduc* as the dependent variable is reported below. For the regression, we have N=857, $R^2 = 0.33$.

Table 2

Regressor	Coeff	Std error	2-sided p-value
IQ	0.0050	0.0003	0
age	0.1202	0.0361	0.001
agesq	-0.0018	0.0005	0.001
black	-0.0314	0.0152	0.039
meduc	0.0126	0.0017	0
_cons	-0.0569	0.5981	0.924

Now, answer the following questions.

a. Interpret the coefficient 0.005 on IQ.
b. How would you test if educational attainment is affected by age, in this model?
c. How would the coefficients and the intercept change if age was instead recorded in months? What would happen to the corresponding t-statistics? Why?
d. How would you test that the true coefficient on mother's education is 1/3 the negative of the true coefficient on black?

2. **Please recap the the law of iterated expectations. We will see an application in the following question.**

3. Consider the regression model $Y_i = \beta_0 + \beta_1 X_i + U_i$ for an I.I.D. sample with $N = 1000$ observations. Suppose $U_i \sim$ I.I.D. $(0, \sigma^2)$ and the X_i are I.I.D. for $i = 1, 2, ..., 1000$, and

that X_i is independent of U_i. Let $\hat{\beta}_1$ denote the OLS estimator of β_1, and consider another estimator $\tilde{\beta}_1$ of β_1, constructed in the following way:

$$\tilde{\beta}_1 = \frac{Y_3 + Y_1 - 2Y_2}{X_3 + X_1 - 2X_2}.$$

You can assume that X_i are continuously distributed and that $X_3 + X_1 - 2X_2$ never takes the value 0.

(a) Is $\tilde{\beta}_1$ an unbiased estimator of β_1? Why?

(b) Can $\tilde{\beta}_1$ be a better estimator than the OLS estimator? Why?

(c) Can you state the general result for estimators of the form $\sum_{i=1}^{n} a_i Y_i / \sum_{i=1}^{n} a_i X_i$?