Answer key for EMET2008 midterm exam.

1. (a) Expectations on both sides: $E[Y_i] = \beta_0$
   (b) $\bar{Y} := \sum Y_i / n$
   (c) Define: $\hat{\beta}_0 := argmin \sum (Y_i - \beta_0)^2$.
   Derive: take derivative and set equal to zero:
   $$2 \sum (Y_i - \hat{\beta}_0) = 0 \qquad \Leftrightarrow \qquad \hat{\beta}_0 = \bar{Y}$$
   It is equal to the sample average.
   (d) $se := s/\sqrt{n} \qquad$ where $s^2 := \sum (Y_i - \bar{Y})^2 / n$
   (e) $E[\hat{\beta}_0] = E[\bar{Y}] = \sum E[Y_i]/n = E[Y_i]$

2. (a) False. Example: $\bar{Y} + 1/n$ for the population mean.
   (b) False. For example, measurement error could bias the results.
   (c) Tricky one. If the researchers conducting the randomized control trial are aware of the non-compliance and, in addition, observe the fraction of subjects in the treatment group who do not comply then, yes, it does not threaten internal validity because the actually treated are a random subset of the treatment group and the researchers are aware of it. If, on the other hand, the researchers are unaware of the non-compliance then the statement is false. (Example: 30% of subjects in treatment group do not take their blood pressure medication but researchers erroneously believe they do. If the medication actually reduces blood pressure in people who comply, then the true treatment effect would be underestimated.) Marking is open-minded towards either argument.

3. (a) No b/c of endogeneity. Endogeneity: reverse causality (body size is a function of wages) and omitted variables (socio-economic and genetic factors and other obvious observables that were included in the paper by Leigh and Kortt). Measurement error in the BMI variables is also a potential problem leading to bias.
   (b) IV: average BMI of biological family members. (Other IV suggestions are also tolerated.) Briefly discuss IV relevance and IV exogeneity.

4. Note: Some of you write stuff when it is obvious that they don't know stuff. This feels a bit insulting. Don't write stuff when you don't know stuff! Simple rule.

   – What is the main research question?
     What is the causal effect on academic participation and achievement of putting a school in an Afghan village?
   – What is the main endogeneity problem?
     Schools are not randomly assigned. Governments put them in districts where there is demand; or alternatively governments may target low demand areas. In any case, bias may arise.
   – What econometric method do they use to address the endogeneity problem?
     Randomized control trial, simple OLS estimation
   – Does their econometric method 'solve' the endogeneity problem?
     Yes, b/c of random allocation of schools. They provide comparison of means b/w control and treatment group.
   – What are their main outcome variables?
     formal enrollment and test scores.
   – What is their main finding?
     Main finding: enrollment jumps up, enrollment gap between boys and girls shrinks dramatically, test scores improve.
   – What problems/shortcomings do you see in their research? Problems: (as discussed in tute) how can test scores improve in such a short time; the fact that enrollments go up is not too surprising as is the fact that the gender gap diminishes.

5. (a) $r_i = \beta_1(X_i - \tilde{X}_i) + u_i$.

   (b) Plugging in,

$$\begin{aligned} E[\tilde{X}_i r_i] &= E[\tilde{X}_i(\beta_1(X_i - \tilde{X}_i) + u_i)] \\ &= \beta_1 E[\tilde{X}_i(X_i - \tilde{X}_i)] \\ &= -\beta_1 E[(X_i + w_i)w_i] \\ &= -\beta_1 \sigma_w^2, \end{aligned}$$

   where the second equality follows because X_i and w_i are independent of u_i, the third equality follows by definition and the last equality follows because X_i is independent of w_i and $E[w_i^2] =: \sigma_w^2$.

   (c) In equation (2), the regressor $\tilde{X}_i$ and the error term r_i are correlated with each other, as shown in part(b). The generic formula for omitted variables bias, in words, is covariance of regressor and error divided by the variance of the regressor. Applied to this case, the bias should therefore be $E[\tilde{X}_i r_i]/\sigma_{\tilde{X}}^2$. Working out the rest:

$$\text{Bias} = E[\tilde{X}_i r_i]/\sigma_{\tilde{X}}^2 = -\beta_1 \sigma_w^2/\sigma_{\tilde{X}}^2 = -\beta_1 \sigma_w^2/(\sigma_X^2 + \sigma_w^2),$$

   where the last equality follows by independence of X_i and w_i.