Midterm 1 (2018) - Answers

#1 a) $sleep = \beta_0 + \beta_1 age + u$

b) holding other factors constant, being one year older leads to 3.54 minute increase in sleep.

c) stress = measure of stress on a one to ten scale. The more stressed you are the harder it is to fall asleep so you sleep less

noise = measure of noise in your neighborhood on a scale of 1 (the least) to 10 (the most). Its hard to fall asleep when it is noisy, so sleep less

#2 A random sample implies that each draws are independently and identically distributed; they all come from the same pop of interest. It is a property we need for the OLS estimator to be unbiased; that way a guess is not systematically too high or too low.

#3 a) u is the error term in the pop. model. It is not observed, and it represents other factors that affect the dep. variable.

$\hat{u}$ is the residual. It is observed and represents the difference between the observed y and the y predicted by the estimated line, ie $y_i - \hat{y}_i$. It can represent other factors that affect y and/or the fact that the true line is estimated with error.

b) it would hold true if by fluke $\hat{\beta}_0 = \beta_0$ and $\hat{\beta}_1 = \beta_1$

#4 The topic was not covered this year.

#5 $\hat{y}_i - \bar{Y} = \hat{\beta}_1 (x_i - \bar{x})$

<u>proof</u>

$\hat{y}_i - \bar{y}$

$= (\hat{\beta}_0 + \hat{\beta}_1 x_i) - \sum_{i=1}^{n} y_i$

$= \hat{\beta}_0 + \hat{\beta}_1 x_i - \sum_{i=1}^{n} (\hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{u}_i)/n$

$= \hat{\beta}_0 + \hat{\beta}_1 x_i - \frac{\sum_{i=1}^{n} \hat{\beta}_0}{n} - \frac{\sum_{i=1}^{n} \hat{\beta}_1 x_i}{n} - \sum_{i=1}^{n} \hat{u}_i / n$

$= \hat{\beta}_0 + \hat{\beta}_1 x_i - \frac{\cancel{n}\hat{\beta}_0}{\cancel{n}} - \hat{\beta}_1 \frac{\sum_{i=1}^{n} x_i}{n} - 0$

$= \hat{\beta}_1 x_i - \hat{\beta}_1 \bar{x}$

$= \hat{\beta}_1 (x_i - \bar{x})$

#6 - no question

#7 the model would be

$$grades = \beta_0 + \beta_1 income + u$$

yes this true that many factors other than income (e.g. innate ability) will determine grades, but ~~as~~ as long as SLR.1, SLR.2, SLR.3 and SLR.4 hold the estimator will be unbiased

The biggest concern is SLR.4. If there is a variable in u that is correlated with income, the the estimator is biased. I would think that not accounting for innate ability would be the biggest concern and that is probably not correlated with family income so I'm not concerned

#8 $\sum_{i=1}^{n} (X_i - \bar{X})^2$

$$= \sum_{i=1}^{n} (X_i - \bar{X})(X_i - \bar{X})$$

$$= \sum_{i=1}^{n} \left[(X_i - \bar{X}) X_i - (X_i - \bar{X}) \bar{X} \right]$$

$$= \sum_{i=1}^{n} (X_i - \bar{X}) X_i - \sum_{i=1}^{n} (X_i - \bar{X}) \bar{X}$$

$$= \sum_{i=1}^{n} (X_i - \bar{X}) X_i - \bar{X} \left(\sum_{i=1}^{n} (X_i - \bar{X}) \right) = 0$$

$$= \sum_{i=1}^{n} (X_i - \bar{X}) X_i$$

#9 a) no causality comes from theory & common sense, not from STATA output

b) we do not know what X and Y are, and not even their units, so cannot tell if econ. significant

c) The underlying model is linear in the parameters. So all we know SLR.1 holds. In addition there is variation in the Xs because one can estimate $\hat{\beta}_1$. So SLR.3 holds. No info on SLR.2 or SLR.4, so cannot make that conclusion.