

**Suggested Answers for Midterm**

1. a) this is the first normal equation; by substituting back in the formulas for  $\hat{u}$  and for  $\beta_0$ , it will all end up being zero  
b) this is the second normal equation; same principle as for (a) (also substitute in the formula for  $\beta_1$ ).  
c) this means that the OLS method constructs the  $\hat{u}$  vector to be orthogonal to all the  $X$ s. This means that they are not correlated with the  $X$ s, and this can be useful for various forms of auxiliary testing.
2. They're all bad. You aren't getting any useful variation in any of these cases (more on this later). But stable is definitely the worse as there you are getting no variation. The other two are the same in terms of usefulness.
3. a) 4 (with the prior that cats first have to pass through the 0-2 age range; otherwise 2)  
b) 2 (or 3 if you answered the first problem as 2)  
c) 3  
d) 1
4. a) passing means that the null hypothesis that all the slope coefficients are equal to zero is rejected at the chosen significance level  
b) the "pass" or fail" must be pretty close to the line (i.e., the chosen significance level). The Wald test gives the largest value of these three methods, so is just over the line, the other two under it.
5.  $y = \beta_0 + \beta_2(x_1 + x_2) + \beta_3(2x_1 + x_3) + \beta_5x_5$

6. a) recall that  $SSR = S_{yy}(1 - R^2)$ . So  $SSR_{ur} = .05S_{yy}$  and  $SSR_r = .20S_{yy}$ . Then the F-statistic is:

$$F = [(.20S_{yy} - .05S_{yy})/1]/[.05S_{yy}/(24-4)] = 60 \text{ with d.f. } 1, 20$$

The F-value for d.f. 1, 20 for the .05 level of confidence is 4.35 so this is way over; even the F-value for the .001 level of confidence is 16.1, again easily exceeded

so we can reject the hypothesis that  $\beta_2 = \beta_3$  with a high level of confidence

- b) that the errors are distributed independent normal  $N(0, \sigma^2)$  under both  $H_0$  and  $H_A$
- c) fall
- d) fall
7. a) information on students' prior academic performance before attending the given school, information on family background, etc.; score = f(charter school dummy, control variables)
- b) perhaps an inherent ability measure, or something about family motivation to succeed
- c) self selection bias; students/families choose whether or not to apply to/attend schools, and also tend to go to schools near their home (or in some districts, have to attend schools near their home, so correlation with neighborhood effects); schools may also select on criteria the researcher does not measure that are correlated with test-taking ability
- d) randomly assign students to charter and noncharter schools
8. a) False. The intercept is the same as before, and the other coefficients are all twice as large.
- b) can answer either as True, It is lower-bounded at zero as it is a squared (multiple) correlation coefficient; or False, this could happen if the model is constrained to fit worse than a horizontal line at the mean of Y.
- c) False. They can be any number and their sign is the same as the sign of the coefficient
- d) True. The F-distribution is lower-bounded at zero.