**MKTG 666 (PhD Seminar: Research Methods 2)**

*Set-up for 1-5*: Suppose the appropriate context for an analysis of *Y* includes predictors . Suppose that model A has been determined to be a valid linear regression model. Model A has two predictors,  and . Suppose that the addition of predictor to model A would also yield a valid regression model. If added to the model, predictor  would have a coefficient that is truly nonzero. In fact, the *t*-statistic for testing the significance of the coefficient of would be 30. But  is uncorrelated with both  and .

1. Does the omission of  lead to an omitted variable bias? Briefly explain why or why not.

2. If  is added to model A , what effect will the addition have on the values of the coefficients in model A?

3. If  is added to model A , what effect will the addition have on the R-square?

4. If  is added to model A , what effect will the addition have on the values of the Root MSE?

5. If  is added to the model, what effect will the addition have on the adjusted R-square?

6. Appraise the desirability of following this advice: “In general, if a model has more than one superfluous variable, it is a good idea to delete all of them at once in order to avoid the possibility that one or more of them may end up in the model if you try to delete them one at a time.”

*Set-up for 7 to 16*: Refer to the data set “**DataCom employees.dta**”, which shows the salaries and other information on a sample of 50 DataCom employees. Use the “salary” variable as *y*. Suppose that the appropriate context for analysis includes all of the other variables as *x*’s.

7. Do you think there are any superfluous variables, in the context of a linear regression model that includes all predictors? If so, which ones and why?

8. Suppose that you have verified that a model with the two predictors Years\_Employed and Years\_Education is a valid regression model. Is Dept an omitted variable?

9. Graphically appraise this model for conformity with the **H** (Homoscedasticity) specification.

10. Quantitatively test this model for conformity with the **H** specification (use a p value of 0.10)

11. Assuming you find the H specification being violated in question 10, implement two methods we discussed in class to address the violation of the H specification. Please explain your method, and share your results. Your first method should involve some transformation of the dependent variable, and your second method should involve something involving the way you report errors. For your first method, please also show, both graphically, and quantitatively, that your transformation helps address the violation of the H specification.

12. Graphically appraise the model in question 8 with the N (normality) specification.

13. Quantitatively test this model for conformity with the N (normality) specification (using a p value of 0.10).

14. Assuming you find the N specification being violated in question, implement the bootstrapping method to address the violation of the N specification. Briefly explain what bootstrapping does, implement bootstrapping, and share your results. How do your bootstrapping results differ from the results in which we had assumed the N specification?

15. Test the hypothesis that the impact of Years\_employed on salary depends on gender. (Hint, start with the model in question 8, and add two more predictors to the model).

16. Do either of the two independent variables in question 8 have a curvilinear rather than a linear effect on the dependent variable? Please carry out appropriate tests to answer this question.

*Set-up for 17-22*: **InsuranceCoSample2.dta** has data on a random sample of 100 life insurance companies. The response variable **Yield** is the ratio of investment income to total invested assets. The predictor variables **pBonds**, **pStocks**, **pMortgages**, **pRealEstate**, **pOtherInvestedAssets**, **pCashSTinvest** are the proportions of total invested assets that are invested in the indicated asset classes (“CashSTinvest” means “cash and short-term investments”). Suppose that the appropriate context for the analysis includes all of the “**p**” variables. Finally, suppose that you have verified that Model A with the predictors pStocks, pRealEstate, and pBonds is a valid linear regression model.

17. Is any predictor variable superfluous for Model A?

18. Is any predictor variable an omitted variable for Model A?

19. Graphically appraise Model A for conformity with the **L** specification.

20. Quantitatively test Model A for conformity with the **L** specification.

21. Quantitatively test Model A for conformity with the **H** specification (choose appropriate p values).

22. Quantitatively test Model A for conformity with the N specification (choose appropriate p values)