Principal Components Regression - Worksheet

Part One - *Give explanation for True/False*

1. *(True/False.)* Since principal components are orthogonal (completely uncorrelated) I can solve the problem of multicollinearity in regression by replacing all my variables with all of the principal components.

   False. Using all of the PCs in a regression will only hide the problem of multicollinearity.

2. *(True/False.)* In order to solve the problem of multicollinearity in regression using principal components, I must omit principal components that have eigenvalues close to zero.

   True. You must omit PCs with small (almost zero) eigenvalues in order to solve the problem.

3. *(True/False.)* The problem with principal components regression is that you cannot interpret the parameter estimates. This problem cannot be solved.

   False. The problem can be solved because PCs can be written in terms of your original variables so its just a matter of substitution and doing the algebra. Or you can use PROC PLS

4. *(True/False.)* Principal components regression is called a biased regression technique because the parameter estimates are biased. True. Once you omit principal components, you've approximated your data. While its a good approximation, its not exact. This error that you impose in the data matrix means your parameter estimates are biased. [statistically speaking, this means that the expected value of each beta-hat is not the population parameter]

5. Can you determine the amount of bias in a principal components regression model?

   no. A model is biased because the statistical theory assuring unbiased parameters breaks down, not because we've "shifted" the parameter estimates in some sense. There is no way to calculate or estimate the amount of bias in the sampling distribution without knowing the population truth.

6. Why would you have to use principal components regression?

   When you have severe multicollinearity and you cannot/don't want to solve it by removing variables. PCR is a good predictive model that might be used when you're solely interested in predictions rather than inference about parameters.

7. What procedure will perform principal components regression and translate the parameter estimates in terms of the original variables?

   PROC PLS

List of Key Words/Phrases.

principal components regression	multicollinearity
variance inflation (VIF)	small eigenvalues
biased regression	PROC PLS