This document shows practice problems for the final exam. The data sets for these problems are in the file "data_practice_problems_final.xlsx" available at the module **Final Exam** on Bruin Learn. The solutions to these problems are not available. To this end, you may compare your answers with those of your classmates.

Problem 1

The data from a patient satisfaction survey in a hospital are available on the file sheet *hospital*. The regressor variables are the patient's age, an illness severity index (higher values indicate greater severity), an indicator (dummy) variable denoting whether the patient is a medical patient (0) or a surgical patient (1), and an anxiety index (higher values indicate greater anxiety).

a. Fit a multiple linear regression model to the satisfaction response using age, illness severity, and the anxiety index as the regressors.

b. Estimate σ^2.

c. Find the standard errors of the regression coefficients.

d. Construct a t-test on each regression coefficient. What conclusions can you draw about the predictors in this model? Use $\alpha = 0.05$.

e. Find 95% confidence intervals on the regression coefficients. Interpret the confidence interval of one of the predictors under study.

f. Does the data contain high leverage points in the data? Are these good or bad?

g. Conduct a full residual analysis.

h. Based on the residual analysis, are your conclusions in (d) and (e) reliable?

Problem 2

An article in IEEE Transactions on Instrumentation and Measurement ["Measurement and Calculation of Powered Mixture Permittivities" (2001, Vol. 50, pp. 1066–1070)] reported on a study that had analyzed powdered mixtures of coal and limestone for permittivity. The errors in the density measurement was the response. The data are in the sheet *density_experiment*.

a. Fit a multiple linear regression model to these data with the density as the response.

b. Estimate σ^2 and the standard errors of the regression coefficients.

c. Use the model to predict the density when the dielectric constant is 2.5 and the loss factor is 0.03.

d. Test for the significance of the full regression model using $\alpha = 0.05$. What is the p-value for this test?

e. What portion of variability is explained by the full model?

f. Construct a t-test on each regression coefficient. What conclusions can you draw about the predictors in this model? Use $\alpha = 0.05$.

g. Conduct regression diagnostics on this model.

h. Based on the regression diagnostics, are your conclusions in (d) and (e) reliable?

Problem 3

An article in the Journal of Pharmaceuticals Sciences ["Statistical Analysis of the Extended Hansen Method Using the Bootstrap Technique" (1991, Vol. 80, pp. 971–977)] presents data on the observed mole fraction solubility of a solute at a constant temperature and the dispersion, dipolar, and hydrogen-bonding Hansen partial solubility parameters. The data are shown in the file sheet *solubility*, where y is the negative logarithm of the mole fraction solubility, x_1 is the dispersion partial solubility, x_2 is the dipolar partial solubility, and x_3 is the hydrogen-bonding partial solubility.

a. Fit the model $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{33} x_3^2 + \epsilon$

b. Test for significance of regression using $\alpha = 0.05$.

c. Plot the residuals and comment on model adequacy.

d. Use an appropriate F-test to test the contribution of the second-order terms using $\alpha = 0.05$.

Remark: If one the residual analysis is not satisfactory in a problem, you can practice by finding appropriate transformations of the predictors and responses and repeat the sub-problems again using the transformed variables.