Statistics 108, Homework 5

Due: November 9th, 2018, In Class (turn in paper form)

*You need to show the steps to get the full credits.

This homework is to practice on multiple linear regression and matrix manipulation. Attach the complete R codes for Problem 2 at the end of the homework. Total: 90 points.

1. (40 points) *Matrix calculation.* Consider the multiple linear regression model in the matrix form $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ with $\mathrm{E}(\boldsymbol{\epsilon}) = \mathbf{0}$ and $\mathrm{Var}(\boldsymbol{\epsilon}) = \sigma^2 I_n$. Let $\hat{\boldsymbol{\beta}}$ be the least squares estimator of $\boldsymbol{\beta}$, $\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ be the fitted values, and $\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}}$ be the residuals.

   (a) (30 pts) Express $\mathrm{Var}(\hat{\boldsymbol{\beta}})$, $\mathrm{Var}(\hat{\mathbf{Y}})$, and $\mathrm{Var}(\mathbf{e})$ in terms of σ^2 and $\mathbf{X}$. Detail the steps in obtaining the expressions.

   (b) (10 pts) Show that $\mathbf{e}^T 1_n = 0$.

2. (50 points) *Data analysis.* A commercial real estate company evaluates age (X_1), operating expenses (X_2, in thousand dollar), vacancy rate (X_3, in percentage), total square footage (X_4, in ten thousand square feet) and rental rates (Y, in thousand dollar) for commercial properties in a large metropolitan area in order to provide clients with quantitative information upon which to make rental decisions. The data are taken from 81 suburban commercial properties. Consider the multiple linear regression model

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \epsilon_i, \; i = 1, \ldots, n, \text{ with } \epsilon_i \overset{iid}{\sim} N(0, \sigma^2).$$

   We use matrix manipulation to fit the model. (For numerical results, keep three digits after the decimal point.)

   (a) In R, create the design matrix $\mathbf{X}$ and the response vector $\mathbf{Y}$. What are their dimensions? Compute $\mathbf{X}^T\mathbf{X}$, $\mathbf{X}^T\mathbf{Y}$, $\left(\mathbf{X}^T\mathbf{X}\right)^{-1}$, and least squares estimator of $\boldsymbol{\beta}$. Copy your results here. Compute the fitted values and residuals. Copy those for the first 6 cases here. Based on the residuals, give an estimate of σ^2.

   (b) Test whether $\beta_2 = 0$ or not at 0.01 significance level.

   (c) Consider a property with the following characteristics: $X_1 = 4, X_2 = 10, X_3 = 10, X_4 = 8$. Give a 95% confidence interval for the expected rental rate of this property.

   (d) Give a 95% prediction interval for the rental rate of the property in (c).

   (e) Continue from part (c), suppose you are also interested in another property with $X_1 = 5, X_2 = 11, X_3 = 8, X_4 = 12$, provide prediction intervals for the rental rate of the two properties with at 0.95 family confidence level.