Exercise: Multiple Regression

SDV2011, Chapter 18, Problems 22, 27 and 28 Chapter 19, Problem 5. Problems marked with a star ("*") are not compulsory.

Problem 1. *Wal-Mart is the second largest retailer in the world. The data file* `Wal-Mart_revenue.csv` *contains monthly data on Wal-Marts's revenue, along with several possibly related economic variables.*

1. *Create a scatterplot matrix with the variables in the data set. Comment!*
2. *Fit a regression with response* `WalMartRevenue` *and explanatory variables* `RetailSalesIndex`, `PersonalConsumption` *and* `CPI`.
3. *Comment the diagnostic plots and identify outliers and leverage points!*
4. *Does it seem that Wal-Marts's revenue is closely related to the general state of the economy?*
5. *Calculate the standardised regression coefficients for the three explanatory variables and discuss the relevance of the variables.*

Problem 2. *We use the indicator variable for the month of December as explanatory variable.*

1. *Calculate the regression model with the four explanatory variables* `RetailSalesIndex`, `PersonalConsumption`, `CPI` *and* `December`. *Discuss the individual coefficients including their significance and explain what the coefficient for* `December` *means.*
2. *Check the diagnostic plots. Comments?*
3. *Does it seem that Wal-Marts's revenue is closely related to the general state of the economy? The general state of the economy here is represented by the variables in the regression equation.*

Problem 3. *We keep* `December` *as an indicator variable.*

1. **Use the full model with all variables to find a best subset model with the BIC criterion (Use Models/subset model selection ...).*
2. *Drop the variables* `RetailSalesIndex` *and* `PersonalConsumption` *from the regression, i.e. recalculate the regression with just* `CPI` *and* `December` *as explanatory variables.*
3. *Check the diagnostic plots.*
4. *Does it seem that Wal-Marts's revenue is closely related to the general state of the economy?*
5. **Compare this last model with two explanatory variables with the full model containing four explanatory variables with an F-test.*
6. **Use the Durbin-Watson test to see whether the residuals exhibit autocorrelation.*