STAT 203 PROBLEM SET 1

Due date: Jan. 21

(1) Let $e_1, \ldots, e_n$ be the residuals from the regression of $y_1, \ldots, y_n$ on $x_1, \ldots, x_n$.
  (a) Show that $\sum_{i=1}^n e_i = 0$.
  (b) One of the assumptions for simple least squares regression is the following: *The errors ϵ_i are independent and identically distributed, with mean 0 and variance σ^2.* Does $\sum_{i=1}^n e_i = 0$ help validate the above assumption? Why or why not?

(2) Consider least squares linear regression of $(y_1, \ldots, n_n)$ on $(x_1, \ldots, x_n)$ by the model:

$$y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2),$$

where the y's are assumed to be independent.
  (a) Express the least squares estimator $\hat{\beta}$ and the residual vector

$$(y_1 - \hat{y}_1, \ldots, y_n - \hat{y}_n)$$

  in matrix notation as a linear transformation of y.
  (b) For $y \sim N(\mu, \Sigma)$, $u = Ay$, $v = By$, the covariance between u and v is

$$A \Sigma B^t.$$

  Use this property to show that r is independent of $\hat{\beta}$.
  (c) Show that

$$\sum_{i=1}^n (y_i - \hat{y})^2 \sim \chi^2_{n-2}.$$

(3) RABE Exercise 3.4 (Data file: Examination.txt)

(4) RABE Exercise 3.14 (Data file: Cigarette.txt)

(5) RABE Exercise 4.7 (Data file: Cigarette.txt)

RABE: Regression Analysis by Example by Chatterjee and Hadi, Ed. 4.