The problems concern the linear regression model

\[ y = X\beta + u \]

where \( y \) is the \( n \)-vector of observations on the dependent variable, \( X \) is the \( n \times k \) matrix of observations on the explanatory variables, \( \beta \) is a \( k \)-vector of unknown parameters, and \( u \) is the \( n \)-vector of unobserved disturbances. Let \( \hat{\beta} \) denote the ordinary least squares (OLS) estimate of \( \beta \), let \( \hat{y} = X\hat{\beta} \) denote the vector of fitted values, and let \( e = y - \hat{y} \) denote the vector of residuals. Let \( y_i, u_i \) and \( e_i \) denote the \( i \)th elements of \( y, u \) and \( e \) respectively, let \( x_i \) denote the \( i \)th row of \( X \), and let \( x_{ij} \) denote the element of \( X \) in the \( i \)th row and \( j \)th column.

1. Let \( F \) be the F-statistic for testing the null hypothesis that \( \beta = 0 \).
   (a) Show that
   \[ F = \frac{\hat{y}'\hat{y}}{e'e} \cdot \frac{n-k}{k} \]
   (b) If all the variables in the model have been “centered” to have mean zero, and there is no intercept (i.e. the matrix of explanatory variables does not have a column of ones), then the coefficient of determination \( R^2 \) is defined by
   \[ R^2 = \frac{\hat{y}'\hat{y}}{y'y}. \]
   (See Subsection 6.6 of the Notes.) Show in this case that
   \[ F = \frac{R^2}{1-R^2} \cdot \frac{n-k}{k} \]

2. An alternative to OLS that was occasionally used in the past is the “least absolute residuals” (LAR) estimator. The estimate \( \hat{\beta} \) is defined as the solution to

\[ \min_{\beta} \sum_{i=1}^{n} |y_i - x_i\beta|. \]
(a) Show that \( \hat{\beta} \) is the ML estimate if the disturbances \( u_i \) are i.i.d. with density

\[
f(t) = \frac{1}{2\theta} \exp\left\{ -\frac{|t|}{\theta} \right\} \quad \text{for } -\infty < t < \infty
\]

conditional on \( X \). (Here the scale parameter \( \theta \) must be positive.) What is the ML estimate \( \hat{\theta} \) for \( \theta \)?

(b) If \( y \) is replaced by \( y + X\alpha \) for some \( k \)-vector \( \alpha \), and \( X \) is replaced by \( XG \) for some non-singular \( k \times k \) matrix \( G \), show that the LAR estimate changes from \( \hat{\beta} \) to

\[
\hat{\beta}_{\text{new}} = G^{-1}(\hat{\beta} + \alpha).
\]

[Note: Equation 83 on page 44 of the Notes shows that OLS also has this property.]

3. This exercise leads you through a simple approach for dealing with heteroskedasticity. Access the data for this problem by clicking on “Data for problem 3 in Problem Set 7” on the course web page. The data set has 45 rows and 3 columns. The first column is \( y \). Columns 2 and 3 of the data set are columns 2 and 3 of \( X \). The first column of \( X \) is a vector of ones. Thus, \( n = 45 \) and \( k = 3 \).

(a) Compute the OLS coefficient vector \( b \) and the vector \( e \) of residuals. Examine scatter-plots of the residuals versus the second and third columns of \( X \). The plot versus the second column of \( X \) is noteworthy: It appears that the variance of the residual is larger for larger values of that explanatory variable. Let \( \sigma_i^2 \) denote the variance of the ith disturbance \( u_i \), and suppose \( \sigma_i^2 \) is related to \( x_{i2} \) by

\[
\sigma_i^2 = \alpha_1 + \alpha_2 x_{i2}^2
\]

where \( \alpha_1 \) and \( \alpha_2 \) are unknown parameters. (This is a simple special case of the general formulation on page 54 of the Notes.) Since \( \mathcal{E}(u_i^2|X) = \sigma_i^2 \) we can write

\[
u_i^2 = \alpha_1 + \alpha_2 x_{i2}^2 + v_i
\]

where \( \mathcal{E}(v_i|X) = 0 \). The disturbances are unobserved, but we can use \( e_i \) as a proxy for \( u_i \).
(b) Compute the least-squares coefficients $\hat{\alpha}_1$ and $\hat{\alpha}_2$ for the regression
\[ e_i^2 = \alpha_1 + \alpha_2 x_i^2 + v_i \quad i = 1, \ldots, n. \]
Set $\sigma_i^2 = \hat{\alpha}_1 + \hat{\alpha}_2 x_i^2$ for each $i$. Find an $n \times n$ matrix $P$ such that the transformed vector of disturbances $\tilde{u} = Pu$ has covariance matrix proportional to the identity matrix. Compute $\tilde{y} = Py$ and $\tilde{X} = PX$, and for the rest of this problem set regard $\tilde{y}$ and $\tilde{X}$ as the data.

(c) Compute the GLS estimate $\hat{\beta} = \left( \tilde{X}^t \tilde{X} \right)^{-1} \tilde{X}^t \tilde{y}$ and the estimated standard error $\hat{\sigma}$ of the components of the vector $\tilde{u}$ of disturbances. Also compute the estimated standard errors of the components of $\hat{\beta}$, and the t-statistics for the components of $\hat{\beta}$. Which coefficients are “significant” at the 1% level?

(d) We wish to test the null hypothesis that $\beta$ satisfies the following two conditions:
\[ 2.5 \beta_2 = 2 \beta_3 + 4.5 \]
\[ \beta_3 = -0.8 \]
Find a matrix $R$ and a vector $r$ such that the null hypothesis can be expressed as $R\beta = r$. Assuming that the disturbances are normally distributed, compute the F-statistic for testing the null hypothesis. Can you reject the null hypothesis at the 5% level of significance? What about at the 1% level of significance?