1. HPRICE1.RAW contains observations collected from the real estate pages of the Boston Globe during 1990. These are homes that sold in the Boston, MA area. The variables of interest are:

1. price  house price, $1000s
2. assess  assessed value, $1000s
3. bdrms  number of bedrooms
4. lotsize  size of lot in square feet
5. sqft  size of house in square feet
6. colonial  =1 if home is colonial style
7. lprice  log(price)
8. lasses  log(assess)
9. llotsize  log(lotsize)
10. lsqft  log(sqft)

(i) Use the data to estimate the model

$$price = \beta_0 + \beta_1 sqft + \beta_2 bdrms + u,$$

(ii) In estimating equation (1), TSP reports:

$$\text{Mean of dep. var.} = 293.546 \quad \text{LM het. test} = 12.2220 \ [0.000]$$
$$\text{Std. dev. of dep. var.} = 102.713 \quad \text{Durbin-Watson} = 1.05807 \ [<.322]$$
$$\text{Sum of squared residuals} = 337845 \quad \text{Jarque-Bera test} = 44.9726 \ [0.000]$$
$$\text{Variance of residuals} = 3974.65 \quad \text{Ramsey's RESET2} = 5.89298 \ [0.017]$$
$$\text{Std. error of regression} = 63.0448 \quad F \ (\text{zero slopes}) = 72.9635 \ [0.000]$$
$$R^2 = .631918 \quad \text{Schwarz B.I.C.} = 494.715$$
$$\text{Adjusted R}^2 = .622288 \quad \text{Log likelihood} = -407.963$$

**Estimated Standard**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Error</th>
<th>t-statistic</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>-19.9150</td>
<td>31.0466</td>
<td>-.622129</td>
<td>[.536]</td>
</tr>
<tr>
<td>SQFT</td>
<td>.128436</td>
<td>.013824</td>
<td>9.29051</td>
<td>[.000]</td>
</tr>
<tr>
<td>BDRMS</td>
<td>15.1982</td>
<td>9.48352</td>
<td>1.60259</td>
<td>[.113]</td>
</tr>
</tbody>
</table>
(a) Write out the results in equation form.

(b) What percentage of the variation in price is explained by square footage and number of bedrooms?

(c) Interpret the slope coefficients. What about their significance? Are they jointly significant?

(d) Why you might keep \textit{bdrms} in the regression even though it is not statistically significant at 10\% confidence level?

(iii) The variable \textit{assess} is the assessed housing value before the house was sold. Suppose now we would like to test whether the assessed housing price is a rational valuation.

(a) For this purpose a simple regression model

\[ \text{price} = \beta_0 + \beta_1 \text{assess} + u \]  

will be estimated. State the null and alternative hypotheses for \(\beta_0\) and \(\beta_1\) carefully. [Hint: if an assessment is rational, then it has incorporated all the relevant information regarding the house.]

(b) The estimation of equation (2) in TSP shows:

\[
\begin{array}{c}
\text{Dependent variable: PRICE} \\
\text{Current sample: 1 to 88} \\
\text{Number of observations: 88} \\
\text{Mean of dep. var.} = 293.546 \\
\text{Std. dev. of dep. var.} = 102.713 \\
\text{Sum of squared residuals} = 165645. \\
\text{Variance of residuals} = 1926.10 \\
\text{Std. error of regression} = 43.8873 \\
\text{R-squared} = .819531 \\
\text{Adjusted R-squared} = .817432 \\
\text{LM het. test} = 2.68713 \ [0.101] \\
\text{Durbin-Watson} = 1.92516 \ [0.393] \\
\text{Jarque-Bera} = 55.5000 \ [0.000] \\
\text{Ramsey's RESET2} = 5.78579 \ [0.018] \\
\text{F (zero slopes)} = 390.535 \ [0.000] \\
\text{Schwarz B.I.C.} = 461.118 \\
\text{Log likelihood} = -456.639 \\
\end{array}
\]

Base on the result, test the hypotheses you made in (a) by conducting \(t\) tests against two-sided alternative. What do you conclude?

(c) \(F\) test can be performed to test the joint hypothesis on \(\beta_0\) and \(\beta_1\). Write down the restricted model in this case.

(d) Suppose in the restricted model, SSR=209,448.99. Carry out the \(F\) test for the joint hypothesis.

(e) Now, test \(H_0: \beta_2 = 0, \beta_3 = 0, \text{ and } \beta_4 = 0\) in the model

\[ \text{price} = \beta_0 + \beta_1 \text{assess} + \beta_2 \text{lotsize} + \beta_3 \text{sqrft} + \beta_4 \text{bdrms} + u. \]  

\[ (3) \]
(iv) *colonial* is a binary variable equal to one if the house is of the colonial style.

Estimate the equation

\[
\log(price) = \beta_0 + \beta_1 \log(lotsize) + \beta_2 \log(sqrft) + \beta_3 bdrms + \beta_4 colonial + u
\]  

and TSP gives:

![Output from TSP](image)

What does the coefficient on *colonial* mean?
(v) A special form of the White test can be applied to test for heteroskedasticity in the regression

$$\log(price) = \beta_0 + \beta_1 \log(lotsize) + \beta_2 \log(sqrft) + \beta_3 bdrms + u.$$ 

The regression of $u^2$ on $\log(price)$ and $\log(price)^2$ have been obtained by using TSP:

**Method of estimation = Ordinary Least Squares**

Dependent variable: SQR_RESIDUAL  
Current sample: 1 to 88  
Number of observations: 88

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimated Coefficient</th>
<th>Standard Error</th>
<th>t-statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>5.04684</td>
<td>3.34500</td>
<td>1.50877</td>
<td>.135</td>
</tr>
<tr>
<td>LPRICE_HAT</td>
<td>-1.70922</td>
<td>1.18333</td>
<td>-1.46924</td>
<td>.145</td>
</tr>
<tr>
<td>SQR_LPRICE_HAT</td>
<td>.145135</td>
<td>.100992</td>
<td>1.43709</td>
<td>.154</td>
</tr>
</tbody>
</table>

(a) Form the *LM* statistic base on the output. Which kind of distribution does the *LM* statistic follow? Be sure to mention the degree of freedom.
(b) The corresponding p-value is computed in TSP and the result shows: 
Upper tail area: .17842.
What conclusion do you get?

2. TRAFFIC1.RAW contains data for 1985 and 1990 for all 50 states and the District of Columbia. The relevant variables are:

1. state  
   state postal code
2. admn90  
   =1 if admin. revocation, '90
3. admn85  
   =1 if admin. revocation, '85
4. open90  
   =1 if open container law, '90
5. open85  
   =1 if open container law, '85
6. dthrte90  
   deaths per 100 mill. miles, '90
7. dthrte85  
   deaths per 100 mill. miles, '85
8. speed90  
   =1 if 65 mph speed limit, 1990
9. speed85  
   =0 always
10. cdthrte  
    dthrte90 - dthrte85
Suppose we are about to estimate the effects of different policies in an attempt to curb drunk driving. Two types of laws that we will study here are open container laws—which make it illegal for passengers to have open containers of alcoholic beverages—and administrative per se laws—which allow courts to suspend licenses after a driver is arrested for drunk driving but before the driver is convicted.

(i) One possible analysis is to use a single cross section of states to regress driving fatalities (or those related to drunk driving) on dummy variable indicators for whether each law is present. What do you think of this approach? Is it going to work well? Why?

(ii) An alternative way is to use first-differenced equation. TSP estimates:

Method of estimation = Ordinary Least Squares

Dependent variable: CDRMRT
Current sample: 1 to 51
Number of observations: 51

Mean of dep. var. = -.545098  LM het. test = .175599 [.675]
Std. dev. of dep. var. = .358505  Durbin-Watson = 2.41223 [<.959]
Sum of squared residuals = 5.66369  Jarque-Bera test = 30.9632 [.000]
Variance of residuals = .117994  Ramsey's RESET2 = 1.31382 [.250]
Std. error of regression = .343502  F (zero slopes) = 3.23144 [.040]
R-squared = .118666  Schwarz B.I.C. = 22.2210
Adjusted R-squared = .081944  Log likelihood = -16.3233

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>t-statistic</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>-.496787</td>
<td>.052426</td>
<td>-9.47604</td>
<td>[.000]</td>
</tr>
<tr>
<td>COPEN</td>
<td>-.419679</td>
<td>.205595</td>
<td>-2.04129</td>
<td>[.047]</td>
</tr>
<tr>
<td>CADMN</td>
<td>-.150602</td>
<td>.116022</td>
<td>-1.25916</td>
<td>[.204]</td>
</tr>
</tbody>
</table>

(a) Write out the result in equation form.
(b) Interpret the intercept and the slope coefficients.