1. Consider the home prices data found in the file “homes.dat”.

(a) Use statistical software to fit a multiple linear regression of home price on the $p = 2$ predictors, $x_1 =$ square feet and $x_2 =$ # of Bedrooms. Report the ANOVA table. Report $r^2 = \frac{SS[Reg]}{SS[TOT]}$ (also called the multiple coefficient of determination).

```
proc reg data=homes;
   model P=S be;
run;
```

(b) Fit a multiple linear regression of home price on the $p = 4$ predictors, $x_1, x_2, x_3 =$ # of bathrooms and $x_4 =$ New. Where $x_4$ is what is called an indicator variable taking values according to

$$x_4 = \begin{cases} 
1 & \text{if house is new} \\
0 & \text{else}
\end{cases}$$

```
proc reg data=homes;
   model P=S be ba new;
run;
```

Report the ANOVA table. Report the multiple coefficient of determination.

(c) Is the $p = 2$ model in part (a) nested in the $p = 4$ model in part (b)?

(d) Use an $F$-test to formally compare the models in parts (a) and (b).

i. Express the comparison as a test of hypotheses about the regression coefficients: $H_0 : ?$. To do this, you should state both models clearly.

ii. Report the appropriate observed $F$-ratio.

iii. Report the appropriate critical value to which the $F$-ratio can be compared for a test with significance level $\alpha = .05$. That is, report $F(.05, ?, ?) =$?.

iv. By comparing the observed $F$-ratio in part (ii) with the critical value in (iii), draw a conclusion regarding whether or not to reject the null hypothesis $H_0$ specified above, using level of significance $\alpha = .05$.

v. Draw a verbose conclusion using the contextual language of the problem (bedrooms, bathrooms and so forth).

vi. In model (b) the estimated partial slope for the $x_2$ variable is $\hat{\beta}_2 = -2.76$ (thousands of dollars per bedroom) with $SE = 3.96$. Consider a comparison of the full model in part (b) versus the reduced model which includes only the $p = 3$ predictors $x_1, x_3$ and $x_4$ (excludes $x_2 =$ # of bedrooms). Report the $F$-ratio for a comparison of these two models using only $\hat{\beta}_2$ and $SE(\hat{\beta}_2)$.

vii. Let $X$ denote the design matrix for model (b). Give the dimension of this matrix: $(rows \times columns) = (?)$.

viii. Give the dimensions of $X'X$.

ix. Report $(X'X)^{-1}X'Y$

x. Report the estimated variance-covariance matrix of the vector of estimated regression coefficients,

$$\hat{VAR}(\hat{\beta}) = (X'X)^{-1}MS(E) = ?$$

xi. Consider the (sub)-population of new ($x_4 = 1$) homes that are $x_1 = 2$ (thousand square feet), have $x_2 = 3$ bedrooms and $x_3 = 2.5$ bathrooms.

A. Report an estimate of the standard deviation of the prices of these homes.

B. Express estimated mean value among these homes, $\hat{\mu}(x_1, x_2, x_3, x_4)$ as a linear combination of regression coefficients. Also, express this linear combination as a product of two vectors ($a' \hat{\beta}$).