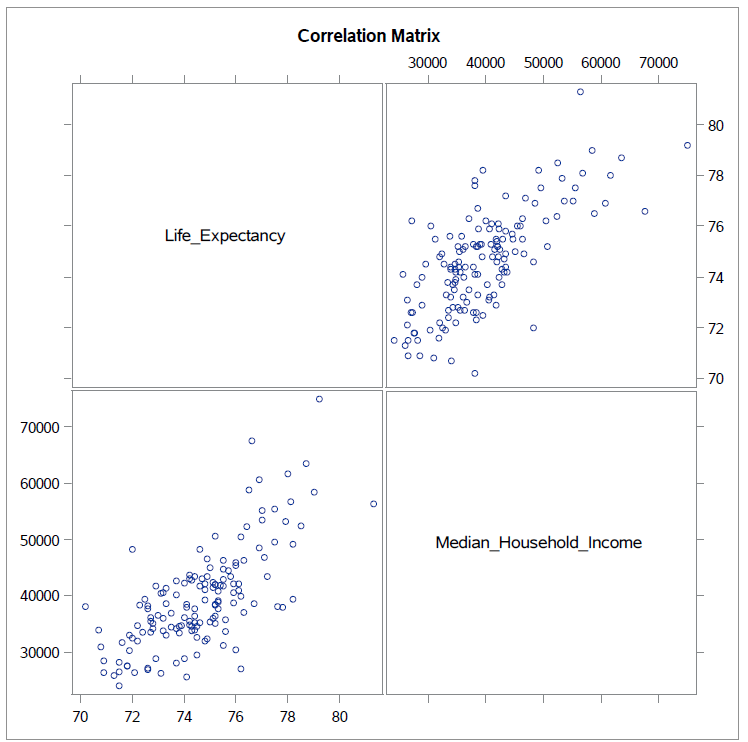
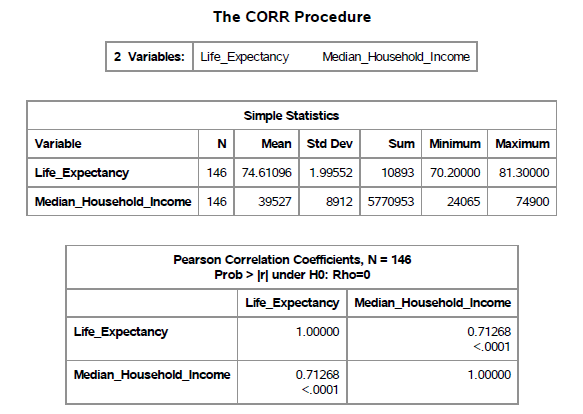
**Multivariate Regression in SAS**

**STA 9705**

From the data gathered, the response variables chosen were Life Expectancy and Median Household Income. Before multivariate regression was decided upon as a possible analysis method, a correlation plot and table between the two response variables were done in order to check for any linearity. The outputs are shown below:

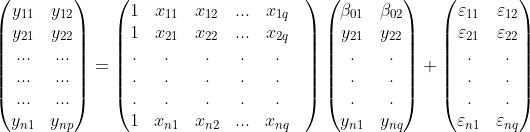


Seeing from the graph on the left as well as the table on the right that the correlation between the two responses were high, it was more efficient to conduct a multivariate approach rather than a univariate approach which would have been done instead if the correlation between the two was small. Another correlation table was produced, this time using both the predictors and the responses. Only the predictors that had an absolute correlation value greater than 0.5 were kept as seen below and then used in conducting our tests.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **x1** | **x2** | **x3** | **x4** | **x5** | **x6** | **x7** | **x8** |
| Population\_Under\_18 | Population\_African\_American | Child\_Poverty\_Rate | Food\_Insecure\_Rate | Uninsured\_Adults\_Prev | Physically\_Inactive\_Rate | Excessive\_Drinking\_Rate, | Teen\_Birth\_Rate |



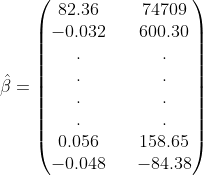
First, our model for the Multivariate Regression has the form:

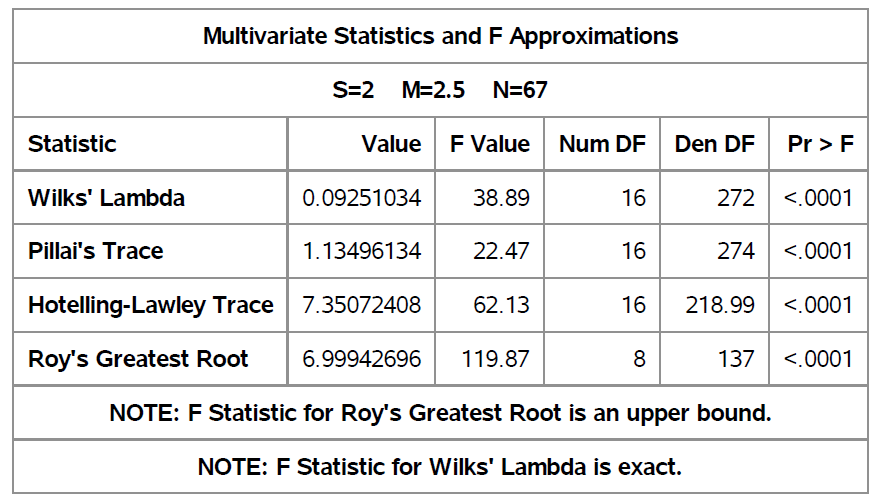
Which in matrix form can be represented as follows:

**Test of Overall Regression**

The test of overall regression seeks to obtain information on whether there is a linear association between the response variables and the set of predictor variables.

For our data set, n = 146 (number of observations), p = 2 (predictor variables) and q = 8 (number of response variables), VH = 8, VE = 137 (n-q-1), s=2, M=2.5, N =67.

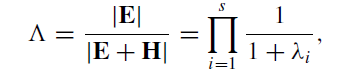
Before we get to testing our hypothesis, we can obtain the estimated coefficients for our model. Usually we have a data set with a collection of observations, with no knowledge of the coefficients β0, β1,…,βk. These can be estimated from our data using the least squares principle which minimizes our solution and gives the least value for the sum of squared errors. From our output, a partial list of the least squares estimate was:



**Hypothesis Test**

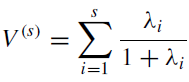
H0: B1 = 0 vs Ha: B1 ≠ 0

Part of the SAS output is as follows

**Wilk’s Test**

Test statistic Λ = 0.0925 (from the table) and the critical value Λ0.5(2,8,137) > Λ0.5(2,8,120) = 0.807.

Since Λ is < Λ.05(2,8,120) we reject the null hypothesis, H0.

**Pillai Test**

Test statistics V(s) = 1.135 (taken from the table) and the critical value V.05(2,3,80)> V.05(2,3,25) = 0.445

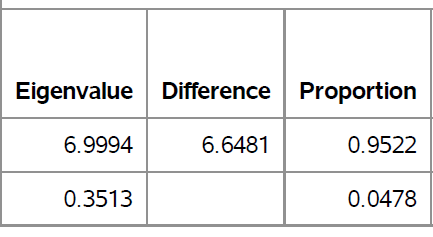
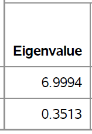
Since V(s) > V.05(2,3,80) we reject the null hypothesis, H0.

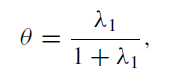
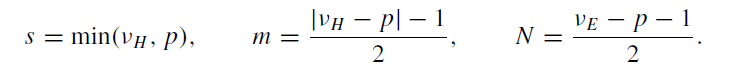
**Lawley-Hotelling Test**

Test statistics is (VE/VH) U(s) , where U(s) = 7.35 (taken from the table) and therefore we have

(VE/VH)U(s) = (137/8)\*7.35 = 125.87 and the critical value is U.05(2,8,137) > U.05(2,8,100) = 3.48

Since U(s) > U.05(2,8,100) we reject the null hypothesis, H0.

****

**Roy’s Test**

Test statistic θ = (6.99/1+6.99) = 0.875 and the critical value θ.05(2,3,80) = .119.

Since θ > θ.05(2,3,80) we reject the null hypothesis, H0.

Based on the four tests conducted at the α = .05 level, in which each test rejected H0 we can conclude that there is a linear regression between the responses, life expectancy and median household income and the predictors.

With the presence of strong linear associations between the responses and the predictors we can now perform some tests on subsets of the data.

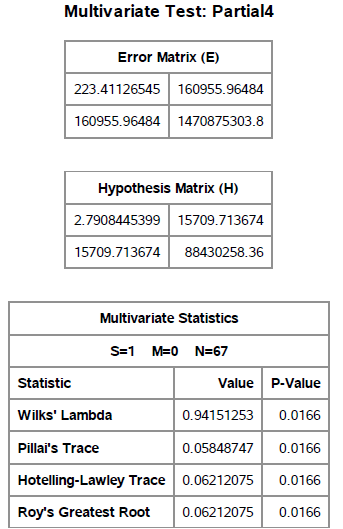
**Tests of Significance**

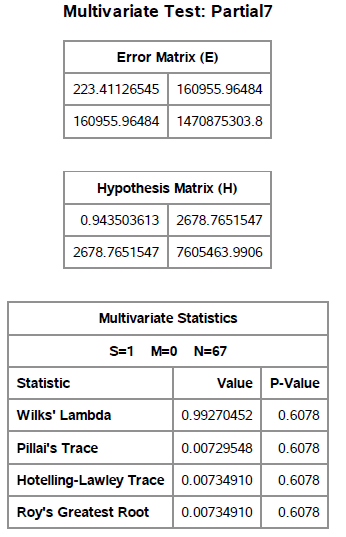
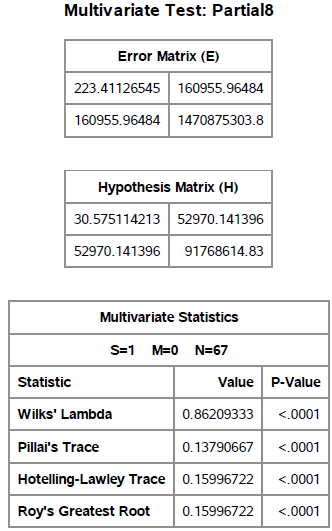
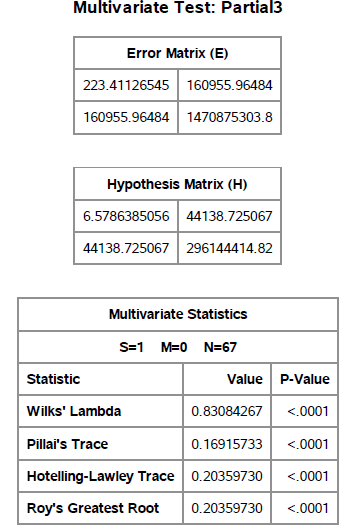
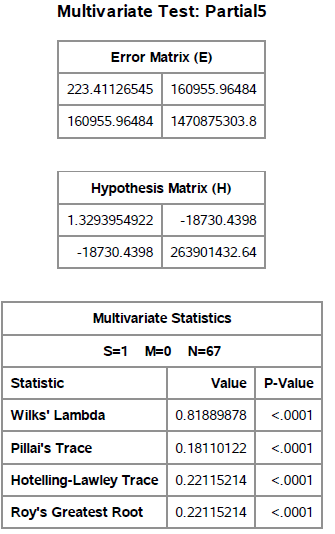
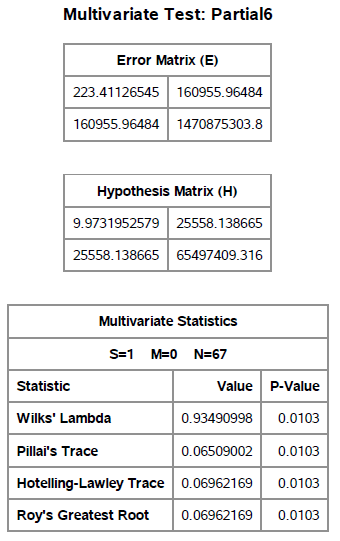
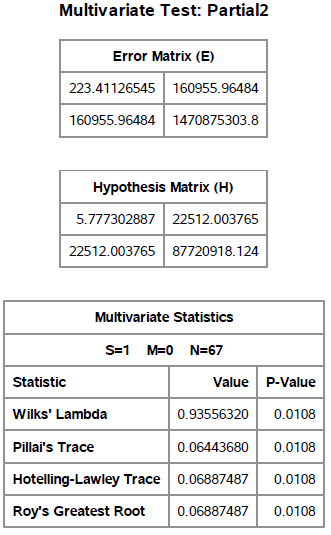
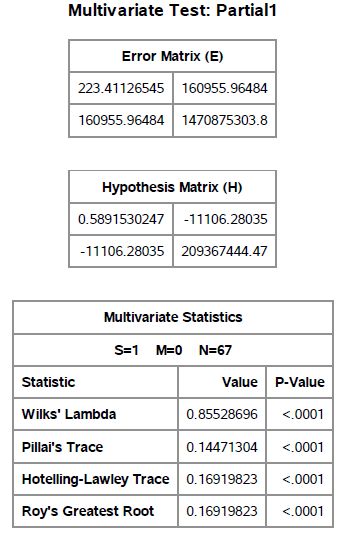
Another area of analysis we can now attempt is seeing if different subsets of the data will prove to be significant on the responses. When testing subsets of the data the B matrix is split in two and now becomes where Bd contains the last h (the number of variables being subset) rows of B.

****This then allows for Xr to contain the last columns in X for Br leading us to compare the reduced model to the full model i.e.

In doing these tests of significance the null and alternate hypothesis take on the form:

H0: Bd = 0 vs Ha: Bd ≠ 0.

First, we test the significance of each predictor on every other predictor. The results from the SAS output are as follows:

********

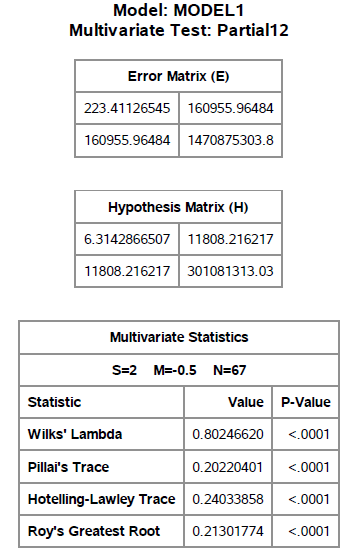
In this case, each of our individual test has VH = h = 1. This means that each of the four test i.e Wilk’s, Pillai’s, Lawley-Hotelling and Roy’s are equivalent to each other and they also have exact F transformations. Therefore, by looking at the p-value in each partial test we can see that all of the predictors were significant except Excessive Drinking Rate(x7) after being adjusted for the other predictors individually.

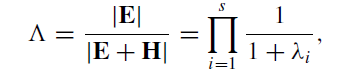
Next, we will test a larger subset such as x1,x2 adjusted for the other predictors. Keep in mind that for our data set, n = 146 p = 2 and VH = 2 (based on subset), VE = 137, s=2, M=0.5, N =67. We test the following:

**Hypothesis Test**

H0: Bd = 0 vs Ha: Bd ≠ 0 where Bd contains the slopes related to x1 and x2.

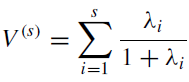
Below is the SAS output from which we will take test stat values.



**Wilk’s Test**

Test statistic Λ = 0.802 (from the table) and the critical value Λ0.5(2,2,137) > Λ0.5(2,2,120) = 0.924.

Since Λ is < Λ.05(2,2,120) we reject the null hypothesis, H0.

**Pillai Test**

Test statistics V(s) = .202 (from the table) and the critical value V.05(2,0.5,67)> V.05(2,1,25) = 0.304.

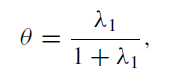
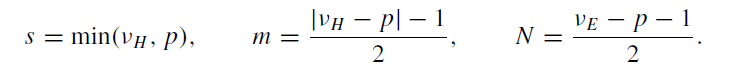
Table A.11 of the Methods of Multivariate Analysis does not have values for Pillai when N is greater than 25. Therefore, the exact p-value is used to conclude the test. Since the Pillai’s p-value as seen in the table above is less than α = .05 we reject, H0.

**Lawley-Hotelling Test**

Test statistics is (VE/VH) U(s) , where U(s) = .2403 (taken from the table) and therefore we have

(VE/VH)U(s) = (137/2)\*.2403 = 16.46 and the critical value is U.05(2,2,137) > U.05(2,2,100) = 4.96

Since U(s) > U.05(2,2,100) we reject the null hypothesis, H0.

**Roy’s Test**

Test statistic θ = (0.213/1+0.213) = 0.176 and the critical value θ.05(2,1,80) = 0.085.

Since θ > θ.05(2,1,80) we reject the null hypothesis, H0.

Based on the four tests conducted at the α = .05 level, in which each test rejected H0 we can conclude that the predictors Population Under 18(x1) and Population African American(x2) are significant predictors when adjusted for the other predictors.