STAT 530 – Homework 10

The computer code and output from SAS for the following problems are provided below. Equivalent code and output from R will be provided early next week.

The data used for the first portion of this homework is from G.C. McDonald and J.A. Ayers (1978) as found in Rasmey and Schafer (1002). It can be found at:
http://www.stat.sc.edu/~habing/courses/data/sascity.txt
It contains the following 16 quantitative variables for each of 60 cities:

PRECIP - average annual precipitation in inches
JANTEMP - average January temperature
JULYTEMP - average July temperatures
OVER65- percent of population over age 65
HOUSE - average population per household
EDUC - median educational attainment in years
SOUND - percentage of housing that was judged to be sound
DENSITY - population density per square mile
NONWHITE - percent non-white
WHITECOL – Percent employed in white-collar occupations
POOR - percent below the poverty line
HC - relative pollution potential of hydrocarbons
NOX - relative pollution potential of nitrogen oxide
SO2 - relative pollution potential of sulphur dioxides
HUMIDITY - average annual humidity
MORTAL - mortality rate per 100,000
CITY - cities abbreviated name

A canonical correlation analysis was conducted with the seven of the “environmental variables” as one group and nine “demographic variables” as the other using the following code.

```
PROC CANCORR DATA=city VPREFIX=envir WPREFIX=people ALL;
VAR PRECIP JANTEMP JULYTEMP  HC  NOX SO2 HUMIDITY;
WITH MORTAL OVER65 HOUSE EDUC SOUND DENSITY NONWHITE WHITECOL POOR;
RUN;
```

The major excerpts from the output are as follows:

```
                                     The CANCORR Procedure

                                 VAR Variables            7
                                 WITH Variables           9
                                 Observations            60

                                Canonical Correlation Analysis

                                      Adjusted    Approximate        Squared
                       Canonical     Canonical       Standard      Canonical
                     Correlation   Correlation          Error    Correlation

                   1    0.922897      0.903014       0.019302       0.851740
                   2    0.810515      0.767463       0.044663       0.656934
                   3    0.610850      0.498803       0.081611       0.373137
                   4    0.520954      0.452388       0.094857       0.271393
                   5    0.368551      0.263663       0.112505       0.135830
                   6    0.252849      0.177047       0.121866       0.063933
                   7    0.139741      0.100686       0.127647       0.019527
```

	Eigenvalues of Inv(E)*H = CanRsq/(1-CanRsq)				Test of H0: The canonical correlations in the current row and all that follow are zero				
	Eigenvalue	Difference	Proportion	Cumulative	Likelihood Ratio	Approximate F Value	Num DF	Den DF	Pr > F
1	5.7449	3.8300	0.6475	0.6475	0.01842512	4.16	63	253.92	<.0001
2	1.9149	1.3196	0.2158	0.8633	0.12427539	2.48	48	225.48	<.0001
3	0.5952	0.2228	0.0671	0.9304	0.36224921	1.53	35	195.93	0.0383
4	0.3725	0.2153	0.0420	0.9723	0.57787662	1.17	24	165.17	0.2751
5	0.1572	0.0889	0.0177	0.9901	0.79312528	0.78	15	132.91	0.7019
6	0.0683	0.0484	0.0077	0.9978	0.91778827	0.54	8	98	0.8261
7	0.0199		0.0022	1.0000	0.98047255	0.33	3	50	0.8023

The CANCORR Procedure

Canonical Structure

Correlations Between the VAR Variables and Their Canonical Variables

	envir1	envir2	envir3	envir4	envir5	envir6	envir7
PRECIP	0.1240	0.8138	-0.3682	0.2158	-0.1449	-0.3175	-0.1353
JANTEMP	-0.9031	0.2668	-0.0509	-0.2819	-0.1387	0.0175	-0.1081
JULYTEMP	-0.4946	0.4600	-0.1942	0.6244	0.2546	-0.1414	0.1774
HC	-0.2812	-0.1230	0.3676	-0.7683	0.2144	0.0484	0.3633
NOX	-0.2261	-0.0142	0.3975	-0.7485	0.2253	0.1727	0.3872
SO2	0.1820	0.3923	0.8048	-0.0264	0.0989	0.3870	0.0706
HUMIDITY	0.0946	-0.1236	-0.0934	-0.3914	0.1827	0.0969	-0.8781

Correlations Between the WITH Variables and Their Canonical Variables

	people1	people2	people3	people4	people5	people6	people7
MORTAL	0.1902	0.8181	0.1956	0.2528	0.1783	0.3695	-0.1503
OVER65	0.6002	0.0331	-0.1386	-0.4950	-0.4013	-0.1964	0.1115
HOUSE	0.1717	0.1913	-0.1738	0.7324	0.1061	-0.0204	-0.0098
EDUC	-0.2318	-0.6104	-0.0853	-0.4229	0.2617	0.2748	-0.1643
SOUND	-0.0492	-0.4646	0.5452	-0.5534	-0.1712	0.0532	-0.0666
DENSITY	0.1492	0.2197	0.6501	-0.0134	-0.0632	0.1308	0.6505
NONWHITE	-0.4327	0.6944	-0.0084	0.2458	0.4372	0.0351	-0.0710
WHITECOL	-0.3481	-0.3098	0.1750	-0.1321	0.0740	0.0100	-0.3090
POOR	-0.5302	0.6667	-0.4413	0.2514	-0.0670	0.0274	0.0751

1) How many of the canonical relationships are statistically significant at an $\alpha=0.05$ level?

2) If an r^2 of 0.10 or higher is used as a cut-off for practical importance then how many of the canonical relationships are practically important?

3) Briefly describe the first "environmental" canonical variate and the first "demographic" canonical variate in terms of the original variables.

The data set analyzed below is from Feynman (1988) and is reported in Ramsey (1997). It concerns the relationship between the temperature at the time the space shuttle was launched and whether or not an O-ring seal on the space shuttle failed (a 1 indicates a failure and a 0 indicates a successful launch).

```
DATA oring;
INPUT temp fail @@;
CARDS;
53      1    56     1      57     1
63      0    66     0      67     0
67      0    67     0      68     0
69      0    70     0      70     1
70      1    70     1      72     0
73      0    75     0      75     1
76      0    76     0      78     0
79      0    80     0      81     0
;

PROC LOGISTIC DATA=oring DESCENDING;
MODEL fail = temp /LACKFIT;
RUN;
```

```
                                                     The LOGISTIC Procedure


                                             Testing Global Null Hypothesis: BETA=0

                                     Test                 Chi-Square       DF     Pr > ChiSq
                                     Likelihood Ratio         5.9441        1         0.0148


                                            Analysis of Maximum Likelihood Estimates
                                                              Standard
                               Parameter    DF    Estimate       Error    Chi-Square    Pr > ChiSq
                               Intercept     1     10.8753      5.7031        3.6363        0.0565
                               temp          1     -0.1713      0.0834        4.2155        0.0401

                                            Hosmer and Lemeshow Goodness-of-Fit Test

                                               Chi-Square       DF     Pr > ChiSq
                                                  13.9837        8         0.0822
```

4) Use the likelihood ratio test to test the null hypothesis that temperature and O-ring failure are unrelated at the α=0.05 level against the alternative that temperature does predict O-ring failure.

5) What is the predicted chance that an O-ring would fail if the outside temperature was 29 degrees? (Note in the output that β_0 is the intercept and β_1 is the coefficient of temp).