Homework Eleven

Jeremy Harper
STA4702
03.24.12

6.23 There was a main effect of iris species $\Lambda = 0.0383$, $p < 0.0001$.

The GLM Procedure

Class Level Information

Class	Levels	Values
fac	3	1 2 3

Number of Observations Read	150
Number of Observations Used	150

MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall fac Effect
H = Type III SSCP Matrix for fac
E = Error SSCP Matrix

S=2   M=-0.5   N=72

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.03831574	299.94	4	292	<.0001
Pillai's Trace	1.14376219	98.18	4	294	<.0001
Hotelling-Lawley Trace	20.34689616	741.00	4	174.17	<.0001
Roy's Greatest Root	20.11060086	1478.13	2	147	<.0001

NOTE: F Statistic for Roy's Greatest Root is an upper bound.
NOTE: F Statistic for Wilks' Lambda is exact.

See attached paper for simultaneous confidence intervals. Using Batlett's test for homogeneity of covariance matrices, we can reject the null hypothesis of covariance homogeneity.

The DISCRIM Procedure
Test of Homogeneity of Within Covariance Matrices

```
Notation: K    = Number of Groups

          P    = Number of Variables

          N    = Total Number of Observations - Number of Groups

          N(i) = Number of Observations in the i'th Group - 1

                   __                       N(i)/2
                   ||   |Within SS Matrix(i)|
          V    = -----------------------------------
                                          N/2
                        |Pooled SS Matrix|

                         _                 _      2
                        |       1        1  |  2P + 3P - 1
          RHO  = 1.0 -  | SUM -----  -  --- | -------------
                        |_     N(i)      N _|  6(P+1)(K-1)

          DF   = .5(K-1)P(P+1)
                                               _                  _
                                              |    PN/2            |
                                              |   N        V       |
Under the null hypothesis:      -2 RHO ln     | ------------------ |
                                              |   __      PN(i)/2  |
```

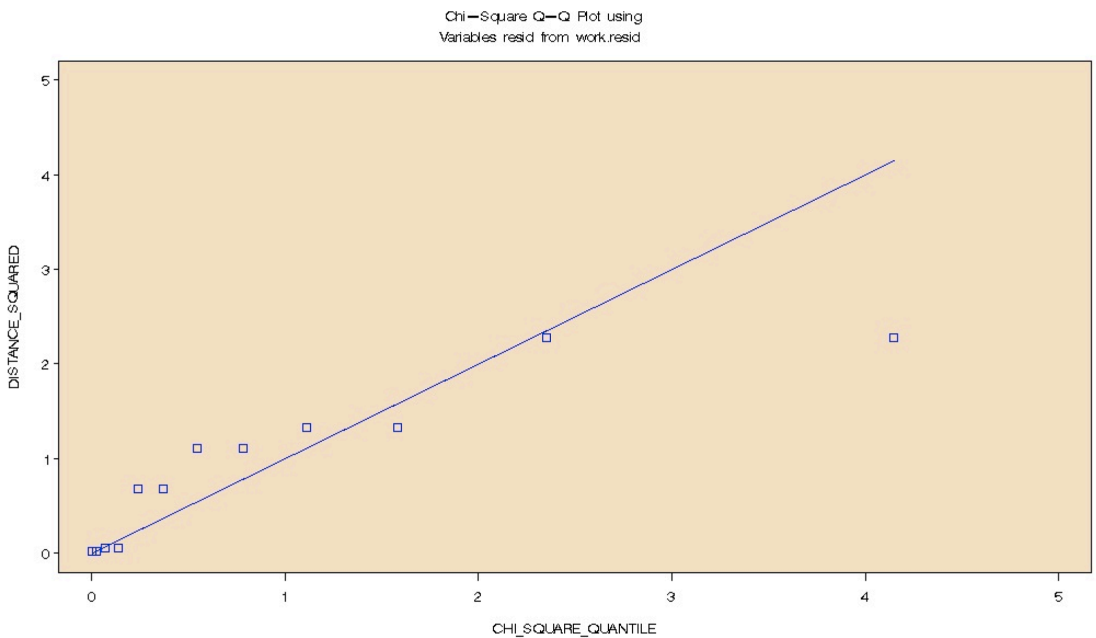
```
                                            |_  ||  N(i)        _|

            is distributed approximately as Chi-Square(DF).


                         Chi-Square        DF    Pr > ChiSq

                          51.794140         6        <.0001

   Since the Chi-Square value is significant at the 0.1 level, the within covariance
            matrices will be used in the discriminant function.
            Reference: Morrison, D.F. (1976) Multivariate Statistical
            Methods p252
```

6.31a There is a significant location effect (Λ = 0.107, *p* = 0.0205), a significant variety effect (Λ = 0.012, *p* = 0.0019), but no significant interaction (Λ = 0.074, *p* = 0.0508)

```
                              The GLM Procedure

                           Class Level Information

                        Class         Levels    Values

                        fac1               2    1 2

                        fac2               3    5 6 8


                    Number of Observations Read          12
                    Number of Observations Used          12

  MANOVA Test Criteria and Exact F Statistics for the Hypothesis of No Overall fac1 Effect
                         H = Type III SSCP Matrix for fac1
                               E = Error SSCP Matrix

                             S=1    M=0.5    N=1

     Statistic                        Value    F Value    Num DF    Den DF    Pr > F

     Wilks' Lambda               0.10651620      11.18         3         4    0.0205
     Pillai's Trace              0.89348380      11.18         3         4    0.0205
     Hotelling-Lawley Trace      8.38824348      11.18         3         4    0.0205
     Roy's Greatest Root         8.38824348      11.18         3         4    0.0205

                              The GLM Procedure
                         Multivariate Analysis of Variance

   MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall fac2 Effect
                         H = Type III SSCP Matrix for fac2
                               E = Error SSCP Matrix

                              S=2    M=0    N=1

     Statistic                        Value    F Value    Num DF    Den DF    Pr > F

     Wilks' Lambda               0.01244417      10.62         6         8    0.0019
     Pillai's Trace              1.70910921       9.79         6        10    0.0011
     Hotelling-Lawley Trace     21.37567504      14.25         6         4    0.0113
     Roy's Greatest Root        18.18761127      30.31         3         5    0.0012

 MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall fac1*fac2 Effect
                       H = Type III SSCP Matrix for fac1*fac2
                               E = Error SSCP Matrix

                              S=2    M=0    N=1

     Statistic                        Value    F Value    Num DF    Den DF    Pr > F

     Wilks' Lambda               0.07429984       3.56         6         8    0.0508
```

```
Pillai's Trace              1.29086073       3.03          6          10     0.0587
Hotelling-Lawley Trace      7.54429038       5.03          6           4     0.0699
Roy's Greatest Root         6.82409388      11.37          3           5     0.0113
```

6.31b Upon investigation of the chi-square plot, the residuals do not appear to be normal.



Using Batlett's test for homogeneity of covariance matrices, we cannot reject the null hypothesis of covariance homogeneity for either factors. Thus, the assumption of covariance homogeneity is satisfied, but the assumption of residual normality is not.

Factor 1:

Chi-Square	DF	Pr > ChiSq
12.477965	6	0.0521

Factor 2:

Chi-Square	DF	Pr > ChiSq
10.563044	12	0.5667

CODE

```
/* 6.23 */
data t115;
     infile '\\psf\Home\Documents\University\Spring_2012\STA4702\Datasets
\T11-5.dat';
     input x1 x2 x3 x4 fac;
run;

/* p = 2, g = 3*/
proc glm data=t115;
     class fac ;
     model x2 x4 = fac /ss3;
```

```
     manova h = fac /printe;
     means fac;
run; quit;

proc discrim  data=t115  pool=test wcov pcov;
     class fac;
     var x2 x4;
run;

/* 6.23a */
data t617;
     infile '\\psf\Home\Documents\University\Spring_2012\STA4702\Datasets
\T6-17.dat';
     input fac1 fac2 x1 x2 x3;
run;
ods graphics on;
/* p = 2, g = 3*/
proc glm data=t617;
     class fac1 fac2;
     model x1 x2 x3 = fac1 fac2 fac1*fac2 /ss3;
     manova h = fac1 fac2 fac1*fac2 /printe;
     means fac1 fac2 fac1*fac2;
     output out = resid r = resid;
run; quit;

/* 6.23b */
proc discrim  data=t617  pool=test wcov pcov;
     class fac1 fac2;
     var x1 x2 x3;
run;

proc discrim  data=t617  pool=test wcov pcov;
     class fac2;
     var x1 x2 x3;
run;

proc sgplot data = resid;
     scatter y = resid x = fac1;
run;

proc sgplot data = resid;
     scatter y = resid x = fac2;
run;

/* sas program for generating data for chi-square q-q plots */
%let inputdata = work.resid;  /* this line must be edited */
%let varlist   = resid;  /* this line must be edited */
proc iml;
   use &inputdata;
   read all var { &varlist } into X;
   n = nrow(X);
   p = ncol(X);
   One = J(n,n,1);            /* just a n x n square matrix full of 1s (nxn)*/
   Xd = X - (One / n)` * X;  /* mean-centered data matrix (nxp)*/
   S = (1 / (n-1)) * Xd`*Xd; /* covariance matrix  (pxp) */
   Sinv = inv(S);
   chisq = j(n,1,0);
     do i = 1 to n;
     chisq[i] = Xd[i,] * Sinv * Xd[i,]`;  /*Distance from obs i to the mean */
```

```
     end;
   probs = (rank(chisq) - j(n,1,.5))/n;   /* contains (r-.5)/n  values */
   quants = quantile('chisquare', probs, p);      /* contains chi-square quantiles */
   plotdata = quants||chisq;
   create chisqqdata(rename=(col1=chi_square_quantile col2=distance_squared)) from 
plotdata;
   append from plotdata;
   quit;
   data chisqqdata;
      merge chisqqdata &inputdata;
   run;
title "Chi-Square Q-Q Plot using";
title2 "Variables &varlist from &inputdata ";
goptions ftext=SWISS ctext=BLACK htext=1 cells;
axis1 width=1 offset=(3 pct) label=(a=90 r=0);
axis2 width=1 offset=(3 pct);
symbol1 c=BLUE ci=BLUE v=SQUARE height=1 cells
        interpol=NONE l=1 w=1;
symbol2 c=BLACK ci=BLUE v=none height=1 cells
        interpol=spline l=1 w=1;
proc gplot data=Work.Chisqqdata ;
   plot distance_squared * chi_square_quantile
          chi_square_quantile *  chi_square_quantile / overlay
      description="Scatter Plot of DISTSQ * CHIQUANT"
      caxis = BLACK
      ctext = BLACK
      cframe = CXF7E1C2
      hminor = 0
      vminor = 0
      vaxis = axis1
      haxis = axis2
      ;
      run;
quit;
goptions ftext= ctext= htext=;
symbol1; symbol2;
axis1; axis2; title;
```