```
%let progname = tlc001.sas;
* input: tlc.dat
* output:
* xref:
* does: Hotelling's T^2
        Profile analysis
        Treatment of Lead Exposed Children Trial (TLC)

*********************************************************************;
title1 "&progname Treatment of Lead Exposed Children (TLC) Trial";

filename INF   "tlc.dat";
ods pdf file = "tlc001.pdf";

*********************************************************************;

data A;
  infile INF  firstobs=2;
  input id group $ lead0 lead1 lead4 lead6;
  dif1 = lead1 - lead0;
  dif4 = lead4 - lead0;
  dif6 = lead6 - lead0;

  * H0: E[lead0] = E[lead1] = E[lead4] = E[lead6]
  is equivalent to
  *     E[(dif1 dif4 dif6)] = (0 0 0);

  run;

***********************************************;

title2 "proc IML";

proc iml ;

*************************************************;
start hot1(y, mu0);
 * Hotelling's T^2 one-sample multivariate test;
 * y = data matrix, mu0 = column vector;

  n = nrow(y);
  d = ncol(y);
  ybar = t(mean(y));
  r = ybar - mu0;
  s = cov(y);
  a = solve(s, r);
  t2 = n * (t(r) * a);               * ~ T^2(d, n-1) under H0;
  t2_df1 = d;
  t2_df2 = n - 1;
  f   = t2 * (n - d) / (d * (n - 1)); * ~ F(d, n-d) under H0;
  f_df1 = d;
  f_df2 = n - d;
  pval = 1 - probf(f, f_df1, f_df2);
  maxroot = t2 / (n - 1);

  print "Hotelling's T^2 one-sample test";
  print n d;
  print ybar s;
```

```
  print "T^2:" ,  t2  t2_df1 t2_df2 ;
  print "F:"   ,  f   f_df1  f_df2;
  print pval;
  print "The max eigenvalue of E^{-1} H: "  maxroot;

  print "Coefficients:", a  [format=16.10];
  a2 = a / sqrt(a[##]);      print "Norm=1:", a2 [format=16.10];
  a3 = a / sqrt(t2*(n-1)/n); print "sas:",    a3 [format=16.10];

finish;
*******************************************************;

* IML execution starts here;

use A;

print "Group: A";
read all var  {dif1 dif4 dif6} into y where (group = "A");
run hot1(y, j(3, 1, 0));  * H0: E[(dif1 dif4 dif6)] = (0 0 0);

print "Group: P";
read all var  {dif1 dif4 dif6} into y where (group = "P");
run hot1(y, j(3, 1, 0));  * H0: E[(dif1 dif4 dif6)] = (0 0 0);

*******************************************************;

title2 "MANOVA in proc GLM";

proc sort data = A; by group; run;
* MANOVA in glm;
proc glm data = A;
  model dif1 dif4 dif6 = ;
  manova h = intercept ;
  by group;
  title2 "proc glm";
  run;

*=================================================================*;
*
Hotelling's T^2 test above can be viewed as a one-sample t-test adjusted
for a specific data-dependent choice of weights.
Demonstration ->
;

data A;
  infile INF  firstobs=2;
  input id group $ lead0 lead1 lead4 lead6;
  dif1 = lead1 - lead0;
  dif4 = lead4 - lead0;
  dif6 = lead6 - lead0;


  * The weights = a = solve(s, r), see the IML code;
  if (group = "P") then
  ay =
   -0.0108130163 * dif1
   -0.0195638415 * dif4
   -0.0195823747 * dif6;
```

```
  else  /* group = "A" */
  ay =
   -0.0161850617 * dif1
   -0.0043244324 * dif4
   -0.0000215889 * dif6;

   label ay = "A linear combination that maximizes the univariate t^2";

   * Comment on scaling:
    The SAS scaling of the coefficients makes the univariate
    sample SD equal to 1/sqrt(n-1).
    Then the univariate t will be t = xbar sqrt{n (n-1)} and
    the its square t^2 = xbar^2 n (n-1).

    Of course, the univariate t^2 = Hotelling's T^2, regardless of the
    scaling. The word "coefficients" refers to
                a := S^{-1} (xbar - mu0).
  ;

  run;

*********************************************;

proc sort data = A; by group; run;

title2 "Hotelling T^2 = t^2 based on the optimal linear combination";
proc univariate data = A;
  var ay;
  by group;
  run;
  * Verify that the value of "Student's t" in the output
  (ignoring the sign) is the square root of Hotelling's T^2 statsitic.
  Note that the p-value printed next to "Student's t" is not correct
  as it does not take into account that the weights were data-dependent
  and specifically chosen to maximize |t| for this particular data matrix.
```