STAT 8320 Spring 2015 Assignment 2

Peng Shao 14221765

May 4, 2015

```
libname da2 'C:\Users\psy6b\Desktop\8320 datasets';
ods graphics on;
options ls=65 ps=35;
title;

data da2.h7q1;
infile 'C:\Users\psy6b\Desktop\8320 datasets\decathlon.dat';
input run100 Ljump shot Hjump run400 hurdle discus polevlt javelin run1500 scor
race_100=-race_100;
race_400=-race_400;
hurdles=-hurdles;
run_1500=-run_1500;
run;
proc univariate data=da2.h7q1 plots;
var score;
run;
data da2.h7q1;
set da2.h7q1;
if score>6000;
run;
proc princomp data=da2.h7q1 out=q1out;
var run100--run1500;
run;
proc rank data=q1out out=q1out descending;
var score;
```

```
ranks posn;
run;
data labels;
set q1out;
retain xsys ysys '2';
y=prin1;
x=prin2;
text=put(posn,2.);
keep xsys ysys x y text;
run;
proc gplot data=q1out;
plot prin1*prin2 / annotate=labels; symbol v=none;
run;
goptions reset=symbol;
proc gplot data=q1out;
plot score*(prin1 prin2);
run;
proc corr data=q1out;
var score prin1 prin2;
run;
```

```
data da2.h7q2(type=corr);
infile 'C:\Users\psy6b\Desktop\8320 datasets\pain.dat' missover;
input _type_ $ _name_ $ p1-p9;
run;
data pain(type=corr);
set da2.h7q2;
run;
proc factor data=pain method=ml n=2 scree;
var p1-p9;
title 'Maximum Likelihood with 2 Factors';
run;
proc factor data=pain method=ml n=3 rotate=varimax;
```

```
var p1-p9;
title 'Maximum Likelihood with 3 factors';
run;
PROC FACTOR DATA=pain PRIORS=SMC PREPLOT PLOT
ROTATE=varimax REORDER OUTSTAT=fact_all n=3;
TITLE 'Principal factor with SMC and varimax';
RUN;
PROC FACTOR DATA=pain PRIORS=SMC PREPLOT PLOT
ROTATE=promax REORDER OUTSTAT=fact_all n=3;
TITLE 'Principal factor with SMC and PROMAX';
run;
title;

data da2.h7q4;
infile 'C:\Users\psy6b\Desktop\8320 datasets\spamdetect_train.dat' dsd;
input attr1-attr57 spam;
run;
data da2.h7q4test;
infile 'C:\Users\psy6b\Desktop\8320 datasets\spamdetect_test.dat' dsd;
input attr1-attr57 spam;
run;
PROC DISCRIM DATA=da2.h7q4 METHOD=NORMAL POOL=YES CROSSVALIDATE outstat=spam;
CLASS spam;
PRIORS PROP;
VAR attr1-attr57;
RUN;
PROC DISCRIM DATA=spam TESTDATA=da2.h7q4test TESTOUT=tout;
CLASS spam;
VAR attr1-attr57;
RUN;
PROC DISCRIM DATA=da2.h7q4 METHOD=NORMAL POOL=NO CROSSVALIDATE outstat=spam2;
CLASS spam;
PRIORS PROP;
VAR attr1-attr57;
RUN;
PROC DISCRIM DATA=spam2 TESTDATA=da2.h7q4test TESTOUT=tout2;
CLASS spam;
VAR attr1-attr57;
```

```
RUN;
proc stepdisc data=da2.h7q4 short METHOD=STEPWISE SLENTRY=.15 SLSTAY=.15;
class spam;
var attr1-attr57;
%put &_stdvar;
run;
proc discrim DATA=da2.h7q4 outstat=spam3 noprint;
CLASS spam;
var &_stdvar;
run;
PROC DISCRIM DATA=spam3 TESTDATA=da2.h7q4test TESTOUT=tout3;
CLASS spam;
var &_stdvar;
RUN;
title 'Model 1';
PROC DISCRIM DATA=da2.h7q4 METHOD=npar K=2 CROSSVALIDATE;
CLASS spam;
PRIORS PROP;
VAR attr1-attr57;
RUN;
title 'Model 2';
PROC DISCRIM DATA=da2.h7q4 METHOD=npar K=3 CROSSVALIDATE;
CLASS spam;
PRIORS PROP;
VAR attr1-attr57;
RUN;
title 'Model 3';
PROC DISCRIM DATA=da2.h7q4 METHOD=npar K=5 CROSSVALIDATE;
CLASS spam;
PRIORS PROP;
VAR attr1-attr57;
RUN;
```

Figure 1: Regression Analysis

The GLM Procedure

Class Level Information

Class	Levels	Values
Gender	2	F M

Number of Observations Read 27
Number of Observations Used 27

The GLM Procedure
Repeated Measures Analysis of Variance

Repeated Measures Level Information

Dependent Variable	Age8	Age10	Age12	Age14
Level of AGE	1	2	3	4

Partial Correlation Coefficients from the Error SSCP Matrix / Prob > |r|

DF = 25	Age8	Age10	Age12	Age14
Age8	1.000000	0.570699 0.0023	0.661320 0.0002	0.521583 0.0063
Age10	0.570699 0.0023	1.000000	0.563167 0.0027	0.726216 <.0001
Age12	0.661320 0.0002	0.563167 0.0027	1.000000	0.728098 <.0001
Age14	0.521583 0.0063	0.726216 <.0001	0.728098 <.0001	1.000000

Figure 1: *continued*

```
     AGE_N represents the nth degree polynomial contrast for AGE

              M Matrix Describing Transformed Variables

                  Age8           Age10           Age12           Age14

AGE_1     -.6708203932    -.2236067977    0.2236067977    0.6708203932
AGE_2     0.5000000000    -.5000000000    -.5000000000    0.5000000000
AGE_3     -.2236067977    0.6708203932    -.6708203932    0.2236067977
                         E = Error SSCP Matrix

     AGE_N represents the nth degree polynomial contrast for AGE

                           AGE_1         AGE_2         AGE_3

          AGE_1         59.1673      -11.2242        4.5278
          AGE_2        -11.2242       26.0412       -1.2819
          AGE_3          4.5278       -1.2819       62.9193
  Partial Correlation Coefficients from the Error SSCP Matrix of the
    Variables Defined by the Specified Transformation / Prob > |r|

         DF = 25           AGE_1          AGE_2          AGE_3

         AGE_1          1.000000      -0.285945       0.074209
                                         0.1567         0.7186

         AGE_2         -0.285945       1.000000      -0.031669
                          0.1567                        0.8779

         AGE_3          0.074209      -0.031669       1.000000
                          0.7186         0.8779
```

Figure 1: *continued*

Sphericity Tests

Variables	DF	Mauchly's Criterion	Chi-Square	Pr > ChiSq
Transformed Variates	5	0.7353334	7.2929515	0.1997
Orthogonal Components	5	0.7353334	7.2929515	0.1997

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

S=1 M=0.5 N=10.5

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.19479424	31.69	3	23	<.0001
Pillai's Trace	0.80520576	31.69	3	23	<.0001
Hotelling-Lawley Trace	4.13362211	31.69	3	23	<.0001
Roy's Greatest Root	4.13362211	31.69	3	23	<.0001

MANOVA Test Criteria and Exact F Statistics
for the Hypothesis of no AGE*Gender Effect
H = Type III SSCP Matrix for AGE*Gender
E = Error SSCP Matrix

S=1 M=0.5 N=10.5

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.73988739	2.70	3	23	0.0696
Pillai's Trace	0.26011261	2.70	3	23	0.0696
Hotelling-Lawley Trace	0.35155702	2.70	3	23	0.0696
Roy's Greatest Root	0.35155702	2.70	3	23	0.0696

Figure 1: *continued*

```
                        The GLM Procedure
              Repeated Measures Analysis of Variance
          Tests of Hypotheses for Between Subjects Effects

Source                     DF    Type III SS    Mean Square   F Value

Gender                      1    140.4648569    140.4648569      9.29
Error                      25    377.9147727     15.1165909

                    Source                Pr > F

                    Gender                0.0054
                    Error
                        The GLM Procedure
              Repeated Measures Analysis of Variance
     Univariate Tests of Hypotheses for Within Subject Effects

Source                     DF    Type III SS    Mean Square   F Value

AGE                         3    209.4369739     69.8123246     35.35
AGE*Gender                  3     13.9925295      4.6641765      2.36
Error(AGE)                 75    148.1278409      1.9750379

                                            Adj Pr > F
          Source                Pr > F     G - G     H-F-L

          AGE                   <.0001    <.0001    <.0001
          AGE*Gender            0.0781    0.0878    0.0797
          Error(AGE)
                Greenhouse-Geisser Epsilon      0.8672
                Huynh-Feldt-Lecoutre Epsilon    0.9769
```

Figure 1: *continued*

```
                        The GLM Procedure
              Repeated Measures Analysis of Variance
            Analysis of Variance of Contrast Variables

AGE_N represents the nth degree polynomial contrast for AGE

Contrast Variable: AGE_1


Source                      DF    Type III SS    Mean Square   F Value

Mean                         1    208.2660038    208.2660038     88.00
Gender                       1     12.1141519     12.1141519      5.12
Error                       25     59.1673295      2.3666932

                    Source                Pr > F

                    Mean                  <.0001
                    Gender                0.0326
                    Error
Contrast Variable: AGE_2


Source                      DF    Type III SS    Mean Square   F Value

Mean                         1     0.95880682     0.95880682      0.92
Gender                       1     1.19954756     1.19954756      1.15
Error                       25    26.04119318     1.04164773

                    Source                Pr > F

                    Mean                  0.3465
                    Gender                0.2935
                    Error
```

Figure 1: *continued*

```
Contrast Variable: AGE_3


Source                       DF    Type III SS    Mean Square   F Value

Mean                          1     0.21216330     0.21216330      0.08
Gender                        1     0.67882997     0.67882997      0.27
Error                        25    62.91931818     2.51677273

                    Source                 Pr > F

                    Mean                   0.7739
                    Gender                 0.6081
                    Error
                         The Mixed Procedure

                          Model Information

        Data Set                     WORK.UNIDATA
        Dependent Variable           measure
        Covariance Structure         Compound Symmetry
        Subject Effect               ID
        Estimation Method            REML
        Residual Variance Method     Profile
        Fixed Effects SE Method      Model-Based
        Degrees of Freedom Method    Between-Within
                      Class Level Information

          Class     Levels    Values

          Gender         2    F M
          age            4    8 10 12 14
```

Figure 1: *continued*

Dimensions

Covariance Parameters	2
Columns in X	15
Columns in Z	0
Subjects	27
Max Obs per Subject	4

Number of Observations

Number of Observations Read	108
Number of Observations Used	108
Number of Observations Not Used	0

Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	470.49084642	
1	1	423.40853283	0.00000000

Convergence criteria met.

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
CS	ID	3.2854
Residual		1.9750

Fit Statistics

-2 Res Log Likelihood	423.4
AIC (Smaller is Better)	427.4
AICC (Smaller is Better)	427.5
BIC (Smaller is Better)	430.0

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
1	47.08	<.0001

Figure 1: *continued*

Solution for Fixed Effects

Effect	Gender	age	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept			27.4687	0.5734	25	47.91	<.0001
Gender	F		-3.3778	0.8983	25	-3.76	0.0009
Gender	M		0	.	.	.	.
age		8	-4.5937	0.4969	75	-9.25	<.0001
age		10	-3.6562	0.4969	75	-7.36	<.0001
age		12	-1.7500	0.4969	75	-3.52	0.0007
age		14	0	.	.	.	.
Gender*age	F	8	1.6847	0.7784	75	2.16	0.0336
Gender*age	F	10	1.7926	0.7784	75	2.30	0.0241
Gender*age	F	12	0.7500	0.7784	75	0.96	0.3384
Gender*age	F	14	0	.	.	.	.
Gender*age	M	8	0	.	.	.	.
Gender*age	M	10	0	.	.	.	.
Gender*age	M	12	0	.	.	.	.
Gender*age	M	14	0	.	.	.	.

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
Gender	1	25	9.29	0.0054
age	3	75	35.35	<.0001
Gender*age	3	75	2.36	0.0781

Figure 1: *continued*

The Mixed Procedure

Model Information

Data Set	WORK.UNIDATA
Dependent Variable	measure
Covariance Structure	Compound Symmetry
Subject Effect	ID
Estimation Method	REML
Residual Variance Method	Profile
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Between-Within

Class Level Information

Class	Levels	Values
Gender	2	F M

Dimensions

Covariance Parameters	2
Columns in X	6
Columns in Z	0
Subjects	27
Max Obs per Subject	4

Number of Observations

Number of Observations Read	108
Number of Observations Used	108
Number of Observations Not Used	0

Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	483.55911746	
1	1	433.75724920	0.00000000

Convergence criteria met.

Figure 1: *continued*

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
CS	ID	3.2986
Residual		1.9221

Fit Statistics

-2 Res Log Likelihood	433.8
AIC (Smaller is Better)	437.8
AICC (Smaller is Better)	437.9
BIC (Smaller is Better)	440.3

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
1	49.80	<.0001

Solution for Fixed Effects

Effect	Gender	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		16.3406	0.9813	25	16.65	<.0001
Gender	F	1.0321	1.5374	25	0.67	0.5082
Gender	M	0	.	.	.	.
age		0.7844	0.07750	79	10.12	<.0001
age*Gender	F	-0.3048	0.1214	79	-2.51	0.0141
age*Gender	M	0	.	.	.	.

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
Gender	1	25	0.45	0.5082
age	1	79	108.36	<.0001
age*Gender	1	79	6.30	0.0141

Figure 1: *continued*

Overall Reinforcement Schedule Effect

The GLM Procedure

Class Level Information

Class	Levels	Values
Rein	3	1 2 3

Number of Observations Read 15
Number of Observations Used 15

Overall Reinforcement Schedule Effect

The GLM Procedure
Multivariate Analysis of Variance

E = Error SSCP Matrix

	Cond1	Cond2	Cond3	Cond4
Cond1	386.8	-99.8	-5.4	535.8
Cond2	-99.8	194.8	87.4	10.2
Cond3	-5.4	87.4	328.4	452.4
Cond4	535.8	10.2	452.4	2553.2

Figure 1: *continued*

```
                Partial Correlation Coefficients from
                  the Error SSCP Matrix / Prob > |r|

DF = 12          Cond1          Cond2          Cond3          Cond4

Cond1         1.000000      -0.363574      -0.015151       0.539159
                                0.2220         0.9608         0.0572

Cond2        -0.363574       1.000000       0.345554       0.014463
                0.2220                         0.2475         0.9626

Cond3        -0.015151       0.345554       1.000000       0.494059
                0.9608         0.2475                         0.0862

Cond4         0.539159       0.014463       0.494059       1.000000
                0.0572         0.9626         0.0862
               Overall Reinforcement Schedule Effect

                          The GLM Procedure
                  Multivariate Analysis of Variance

                  H = Type III SSCP Matrix for Rein

               Cond1           Cond2           Cond3           Cond4

Cond1    1468.1333333    -82.13333333    -185.5333333    1402.1333333
Cond2    -82.13333333    26.133333333    34.533333333    -82.13333333
Cond3    -185.5333333    34.533333333    50.533333333    -181.3333333
Cond4    1402.1333333    -82.13333333    -181.3333333    1339.7333333
```

Figure 1: *continued*

Characteristic Roots and Vectors of: E Inverse * H, where
H = Type III SSCP Matrix for Rein
E = Error SSCP Matrix

Characteristic Root	Percent	Characteristic Vector V'EV=1 Cond1	Cond2	Cond3	Cond4
4.52748560	95.57	0.06117242	0.02992718	-0.00648850	-0.00508768
0.20978723	4.43	0.02215228	0.04489182	0.04553803	-0.01518185
0.00000000	0.00	-0.00272754	-0.04849369	0.04413792	0.00585571
0.00000000	0.00	-0.02461117	0.03992289	-0.03194950	0.02388060

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

S=2 M=0.5 N=3.5

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.14954207	3.57	8	18	0.0119
Pillai's Trace	0.99249428	2.46	8	20	0.0488
Hotelling-Lawley Trace	4.73727283	5.09	8	10.78	0.0081
Roy's Greatest Root	4.52748560	11.32	4	10	0.0010

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

Figure 1: *continued*

Overall Reinforcement Schedule Effect

The GLM Procedure

Class Level Information

Class	Levels	Values
Rein	3	1 2 3

Number of Observations Read 15
Number of Observations Used 15

Overall Reinforcement Schedule Effect

The GLM Procedure
Multivariate Analysis of Variance

M Matrix Describing Transformed Variables

	Cond1	Cond2	Cond3	Cond4
MVAR1	1	-1	0	0
MVAR2	0	1	-1	0
MVAR3	0	0	1	-1

Figure 1: *continued*

```
                Overall Reinforcement Schedule Effect

                          The GLM Procedure
                   Multivariate Analysis of Variance

      Characteristic Roots and Vectors of: E Inverse * H, where
   H = Contrast SSCP Matrix for Compare Schedule 1 v.s. Schedule 2
                        E = Error SSCP Matrix

           Variables have been transformed by the M Matrix

Characteristic              Characteristic Vector  V'EV=1
          Root   Percent           MVAR1           MVAR2           MVAR3

    0.00381869    100.00      0.01485376      0.03301340    -0.01683428
    0.00000000      0.00      0.00840053      0.04788857     0.01044145
    0.00000000      0.00      0.03999183    -0.00230315      0.01681033
    MANOVA Test Criteria and Exact F Statistics for the Hypothesis
       of No Overall Compare Schedule 1 v.s. Schedule 2 Effect
       on the Variables Defined by the M Matrix Transformation
   H = Contrast SSCP Matrix for Compare Schedule 1 v.s. Schedule 2
                        E = Error SSCP Matrix

                         S=1    M=0.5    N=4

Statistic                        Value  F Value  Num DF  Den DF  Pr > F

Wilks' Lambda                0.99619584     0.01       3      10  0.9979
Pillai's Trace               0.00380416     0.01       3      10  0.9979
Hotelling-Lawley Trace       0.00381869     0.01       3      10  0.9979
Roy's Greatest Root          0.00381869     0.01       3      10  0.9979
```

Figure 1: *continued*

```
     Characteristic Roots and Vectors of: E Inverse * H, where
 H = Contrast SSCP Matrix for Compare Schedule 1 & 2 v.s. Schedule 3
                       E = Error SSCP Matrix

          Variables have been transformed by the M Matrix

Characteristic             Characteristic Vector  V'EV=1
          Root   Percent          MVAR1          MVAR2          MVAR3

   2.80326909    100.00     0.03454831     0.02736818    -0.00461007
   0.00000000      0.00     0.02617305     0.00124065     0.02556856
   0.00000000      0.00    -0.00345477     0.05136096     0.00000000
   MANOVA Test Criteria and Exact F Statistics for the Hypothesis
    of No Overall Compare Schedule 1 & 2 v.s. Schedule 3 Effect
      on the Variables Defined by the M Matrix Transformation
 H = Contrast SSCP Matrix for Compare Schedule 1 & 2 v.s. Schedule 3
                       E = Error SSCP Matrix

                        S=1    M=0.5    N=4

Statistic                      Value  F Value  Num DF  Den DF  Pr > F

Wilks' Lambda             0.26293170     9.34       3      10  0.0030
Pillai's Trace            0.73706830     9.34       3      10  0.0030
Hotelling-Lawley Trace    2.80326909     9.34       3      10  0.0030
Roy's Greatest Root       2.80326909     9.34       3      10  0.0030
```

Figure 2: Graphs for Regression Analysis