Question 1

Problems graded on completion:
1e, 1f, 2a, 3b

Homework 4
764 Multivariate
Qianqian Yao
1253177

35/35

previous in HW3.

$\underline{y}_{ij} = \underline{\mu} + \alpha_i + \epsilon_{ij}$

$i = 1, 2, 3, 4, 5, 6$

$j = 1, 2, \cdots 20$

$k = 6$
$p = 4$
$n = 20$

$\underline{y}_{ij}$ is the 4x1 observation vector.

$\underline{\mu}$ is the 4x1 overall population mean vector.

α_i is the 4x1 effect vector of ith charring process group.

$\epsilon_{ij} \sim N(\underline{\mu}, \Sigma)$ is the 4x1 random error vector associated with jth grain of rice in ith charring process group.

Assumptions of MANOVA:

- The six samples of rice grains are randomly selected and mutually independent.
- For the grains of rice in each charring process group, the observation vectors of four changes in morphological measurements have a multivariate Normal Distribution.
- The population covariance matrix of observation vectors, Σ, is the same for each group of rice grains.

a. PROC DISCRIM

$H_0: \Sigma_1 = \Sigma_2 = \Sigma_3 = \Sigma_4 = \Sigma_5 = \Sigma_6$

H_a: At Least One Inequity. (ALOI)

output:

Chi-square	DF	P-value
178.314	30	<0.0001

P-value $< \alpha = 0.10$,

Thus, Reject Ho.

At significant level of $\alpha = 0.10$, there's sufficient evidence to claim Ha, which is at least one Σ is inequality.

b. output:

group	$\ln\lvert S_i\rvert$
1	-9.921
2	-10.927
3	-8.290
4	-8.281
5	-9.200
6	-7.328
pooled	-7.246

$n=20$

$p=4$

$k=6$

$v_i=19$

$$C_1=\left(\sum_{i=1}^{k}\frac{1}{v_i}+\frac{1}{\sum_i v_i}\right)\left(\frac{2p^2+3p+1}{6(p+1)(k-1)}\right)$$

$$=\frac{(k+1)(2p^2+3p+1)}{6kv(p+1)} \quad (\text{if } v_1=v_2=\cdots=v_k=v)$$

$$=\frac{(6+1)(2\times4^2+3\times4+1)}{6\times6\times19\times5}$$

$$\doteq 0.092$$

$$\ln M=\frac{1}{2}\sum_{i=1}^{k}v_i\ln\lvert S_i\rvert-\frac{1}{2}\left(\sum_{i=1}^{k}v_i\right)\ln\lvert S_{pl}\rvert$$

$$=\frac{1}{2}\times19\times(-9.921-10.927-8.290-8.281-9.200-7.328)-\frac{1}{2}\times6\times19\times(-7.246)$$

$$=-97.7645$$

$$u=-2(1-C_1)\ln M$$

$$=-2\times(1-0.092)\times(-97.7645)$$

$$=177.540 \checkmark$$

SAS output

chi-square = 178.314

Critical Value: $\chi^2_{0.10,\ \frac{1}{2}(k-1)p(p+1)}=\chi^2_{0.10,50}=63.167$

Reject Ho if $U>\chi^2_{0.10,50}$

$U=177.540>63.167$. Thus, Reject Ho.

At significant level of $\alpha=0.10$, there's sufficient evidence to claim Ha, which is at least one Σ is inequality

c. T^2 and MANOVA tests are fairly robust to heterogeneity of covariance matrices as long as the sample sizes are large and equal. But it is necessary to test for some cases. $H_0\ \Sigma_1 = \Sigma_2 = \cdots = \Sigma_k$

For testing $H_0: \Sigma_1 = \Sigma_2 = \cdots \Sigma_6$,

* we assume independent samples of size 20 are from multivariate Normal Distributions.

d. $E^{-1}H$:

	$\propto$ sample	$\propto$ population
$\lambda_1 = 1.0576$	a_1	α_1
$\lambda_2 = 0.3301$	a_2	α_2
$\lambda_3 = 0.1116$	a_3	α_3
$\lambda_4 = 0.0037$	a_4	α_4

discriminant functions: $\lambda = \dfrac{a'Ha}{a'Ea}$ is maximized by λ_1, the remaining eigenvalues $\lambda_2 \geq \lambda_3 \geq \lambda_4$ correspond to the remaining 3 discriminant functions.

$H_{01}: \alpha_i = 0$ for $i = 1, 2, 3, 4$

$k=6$
$P=4$
$N = 20\times 6 = 120$

$$V_1 = \left(N-1-\tfrac{1}{2}(P+k)\right)\sum_{i=1}^{4} \ln(1+\lambda_i)$$

$= (120-1-\frac{1}{2}\times 10)\times(\ln 2.0576 + \ln 1.3301 + \ln 1.1116 + \ln 1.0037)$

$= 127.257$

Critical value: $\chi^2_{0.05,\, P(k-1)} = \chi^2_{0.05,\, 4\times 5=20} = 31.410$

Reject H_{01} if $V_1 > \chi^2_{0.05,20}$

$V_1 = 127.257 > 31.410$, thus, Reject H_0.

At significant level of $\alpha = 0.05$, there's sufficient evidence to claim that at least the first discriminant function is significantly different from 0.

$H_{02}: \alpha_i = 0$ for $i = 2, 3, 4$

$$V_2 = \left(N-1-\tfrac{1}{2}(P+k)\right)\sum_{i=2}^{4} \ln(1+\lambda_i)$$

$= (120-1-\frac{1}{2}\times 10)\times(\ln 1.3301 + \ln 1.1116 + \ln 1.0037)$

$= 45.001$

Critical value: $\chi^2_{0.05,\,(P-1)(k-2)} = \chi^2_{0.05,\, 3\times 4=12} = 21.026$

Reject H_{02} if $V_2 > \chi^2_{0.05,12}$

$V_2 = 45.001 > 21.026$, Thus, Reject H_0.

At significance level of $\alpha = 0.05$, there's sufficient evidence to claim that at least α_2 are significantly different from 0.

H_{03}: $\underline{\alpha_i} = \underline{0}$ for $i = 3, 4$

$$V_3 = (N - 1 - \tfrac{1}{2}(P+k)) \sum_{i=3}^{4} \ln(1+\lambda_i)$$

$$= (120 - 1 - \tfrac{1}{2} \times 10) \times (\ln 1.1116 + \ln 1.0037)$$

$$= 12.482$$

Critical Value: $\chi^2_{0.05, (P-2)(k-3)} = \chi^2_{0.05, 2\times3=6} = 12.592$

Reject H_0 if $V_3 > \chi^2_{0.05, 6}$

$V_3 = 12.482 < 12.592$, Thus, Do Not Reject H_0.

At significance level of $\alpha = 0.05$, there's no sufficient evidence to claim that $\underline{\alpha}_3'\underline{y}$ are significantly different from 0.

Thus, $\underline{a}_1'\underline{y}$ and $\underline{a}_2'\underline{y}$ are the discriminant functions significantly different from 0.

e. $E^{-1}H$: $\lambda_1 = 1.0576$, $\underline{a}_1 = (-0.1299, -0.1268, 0.0960, 0.0163)'$
$\lambda_2 = 0.3301$, $\underline{a}_2 = (-1.6060, -0.9601, -1.2530, 0.0590)'$

$$S_{pl} = \frac{E}{k(n-1)} = \frac{1}{6 \times 19} \begin{bmatrix} \underline{8.2903} & 1.1700 & -4.0292 & 154.4974 \\ 1.1700 & \underline{16.8731} & -0.4463 & 284.4113 \\ -4.0292 & -0.4463 & \underline{5.1240} & -11.4755 \\ 154.4974 & 284.4113 & -11.4755 & \underline{8475.7355} \end{bmatrix}$$

$[\mathrm{diag}(S_{pl})]^{\frac{1}{2}}$:
Diagonal elements

$$\begin{pmatrix} \sqrt{\frac{8.2903}{114}} \\ \sqrt{\frac{16.8731}{114}} \\ \sqrt{\frac{5.1240}{114}} \\ \sqrt{\frac{8475.7355}{114}} \end{pmatrix} = \begin{pmatrix} 0.2697 \\ 0.3847 \\ 0.2120 \\ 8.6226 \end{pmatrix}$$

Standardized coefficient vector:

$$a^* = [\mathrm{diag}(S_{pl})]^{1/2} a$$

$$a_1^* = \begin{pmatrix} 0.2697 \times (-0.1299) \\ 0.3847 \times (-0.1268) \\ 0.2120 \times 0.0960 \\ 8.6226 \times 0.0163 \end{pmatrix} = \begin{pmatrix} -0.035 \\ -0.049 \\ 0.020 \\ 0.141 \end{pmatrix}$$

$$a_2^* = \begin{pmatrix} 0.2697 \times (-1.6060) \\ 0.3847 \times (-0.9601) \\ 0.2120 \times (-1.2530) \\ 8.6226 \times 0.0590 \end{pmatrix} = \begin{pmatrix} -0.433 \\ -0.369 \\ -0.266 \\ 0.509 \end{pmatrix}$$

✓

f. $a_1^* = \begin{pmatrix} -0.035 \\ -0.049 \\ 0.020 \\ 0.141 \end{pmatrix} \begin{matrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{matrix}$ Rank: y_4, y_2, y_1, y_3

$a_1 = \begin{pmatrix} -0.1299 \\ -0.1268 \\ 0.0960 \\ 0.0163 \end{pmatrix}$ Rank: y_1, y_2, y_3, y_4

$a_2^* = \begin{pmatrix} -0.433 \\ -0.369 \\ -0.266 \\ 0.509 \end{pmatrix} \begin{matrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{matrix}$ Rank: y_4, y_1, y_2, y_3

$a_2 = \begin{pmatrix} -1.6060 \\ -0.9601 \\ -1.2530 \\ -0.0590 \end{pmatrix}$ Rank y_1, y_3, y_2, y_4

✓

variable Rankings based on standardized coefficient vectors and unstandardized coefficient vectors are different.

g. See AS the Attachment.

y_4 first selected,
y_3 second selected.

y_4 selected can be supported by a_1^* and a_2^*. But By a_1^* and a_2^*, y_3 selected cannot be supported.

about y_3, can be supported by a_3^*. But a_3^* is tested Not significant.

if we calculate $a_3^* = \begin{pmatrix} 0.2691 \times 0.1244 \\ 0.3841 \times 0.0234 \\ 0.2120 \times (-0.3009) \\ 8.6226 \times 0.0012 \end{pmatrix} = \begin{pmatrix} 0.034 \\ 0.009 \\ -0.064 \\ 0.010 \end{pmatrix}$ Rank: y_3, y_1, y_4, y_2

$\lambda_3 = 0.1116$ $a_3 = (0.1244, 0.0234, -0.3009, 0.0012)'$

$\lambda_4 = 0.0037$ $a_4 = (0.3221, -0.1771, 0.2672, -0.0015)'$

Non significant discriminant functions in d.

Question 2.

a. $H_0: \Sigma = \begin{bmatrix} \sigma_{11} & 0 & & \\ & \sigma_{22} & & 0 \\ 0 & & \sigma_{33} & \\ & & & \sigma_{44} \end{bmatrix}$

or

$H_0: \sigma_{jk} = 0 \ \forall j \neq k$

$\ln|S_1| = -9.922$

$|S_1| = 4.908 \times 10^{-5}$

$u = \dfrac{|S_1|}{S_{11} S_{22} S_{33} S_{44}}$

$= \dfrac{4.908 \times 10^{-5}}{0.0443 \times 0.4533 \times 0.0198 \times 110.2107}$

$\nu = n-1 = 19$

$p = 4$

$= 1.120 \times 10^{-3}$

$u' = -(\nu - \frac{1}{6}(2p+5)) \ln u$

$= -(19 - \frac{1}{6} \times (2 \times 4 + 5)) \ln 0.001120$

16.833×6.794

$= 114.371$

Critical value: $\chi^2_{0.05,\ \frac{1}{2}p(p-1)} = \chi^2_{0.05,\ \frac{1}{2} \times 4 \times 3 = 6} = 12.592$

Reject H_0 if $u' > \chi^2_{0.05,6}$

$u' = 114.371 > 12.592$, Thus, Reject H_0.

At significance level of $\alpha = 0.05$, there's sufficient evidence to claim H_a that, the four variables are not mutually independent.

b. Let $\underline{y} = \begin{pmatrix} y_1 \\ y_3 \end{pmatrix}$ $\underline{x} = \begin{pmatrix} y_2 \\ y_4 \end{pmatrix}$

$\begin{pmatrix} \underline{y} \\ \underline{x} \end{pmatrix} = \begin{pmatrix} y_1 \\ y_3 \\ y_2 \\ y_4 \end{pmatrix}$

Covariance Matrix =

$$\begin{array}{c|cc|cc} & y_1 & y_3 & y_2 & y_4 \\ \hline y_1 & 0.0443 & -0.0185 & 0.0072 & 0.6815 \\ y_3 & -0.0185 & 0.0198 & 0.0028 & -0.0070 \\ \hline y_2 & 0.0072 & 0.0028 & 0.4533 & 6.7150 \\ y_4 & 0.6815 & -0.0070 & 6.7150 & 110.2107 \end{array}$$

S_{yy} (upper left), S_{yx} (upper right), S_{xy} (lower left), S_{xx} (lower right)

$H_0: \Sigma = \begin{pmatrix} \Sigma_{yy} & 0 \\ 0 & \Sigma_{xx} \end{pmatrix}$

$|S_{yy}| = 0.0443 \times 0.0198 - 0.0185^2 = 5.3489 \times 10^{-4}$

$|S_{xx}| = 0.4533 \times 110.2107 - 6.7150^2 = 4.867$

$\Lambda = \dfrac{|S|}{|S_{yy}|\,|S_{xx}|} = \dfrac{4.908 \times 10^{-5}}{5.3489 \times 10^{-4} \times 4.867}$

$= 0.019$

$p = 2$

$q = 2$

$n = 20$

$n-1-q = 20-1-2 = 17$

Critical Value: $\Lambda_{p,q,n-1-q} = \Lambda_{2,2,17} = 0.562$

Reject H_0 if $\Lambda < \Lambda_{2,2,17}$

$\Lambda = 0.019 < 0.562$, Thus, Reject H_0.

At significance level of $\alpha = 0.05$, There's sufficient evidence to claim H_a, that the variables in each of the two subvectors are not mutually independent.

c. Assumption of $H_0: \Sigma = \begin{bmatrix} \sigma_{11} & & & 0 \\ & \sigma_{22} & & \\ & & \sigma_{33} & \\ 0 & & & \sigma_{44} \end{bmatrix}$ or $H_0: P_\rho = I$

All variables are mutually independent.

$\sigma_{jk} = 0 \quad j \neq k$, No Restrictions on σ_{jk}.

Assumption of $H_0: \Sigma = \begin{bmatrix} \Sigma_{yy} & 0 \\ 0 & \Sigma_{xx} \end{bmatrix}$

every variable in $\underline{y}$ is independent of every variable in $\underline{x}$

No Restrictions on Σ_{yy}, Σ_{xx}.

d. * sphericity : $H_0: \Sigma = \sigma^2 I$

All Variables are independent,

And have same variances. → But given the sample covariance S,

$\sigma_{44} = 110.2107$ is much larger than other variances of 0.0443, 0.4533, 0.0198,

Thus, Assumptions are unmet.

Not appropriate to perform sphericity test.

*Compound symmetry : $H_0: \Sigma = \sigma^2[(1-\rho)I + \rho J]$

Variables to be correlated But Restricts every variable to have the same variance and every pair of variables to have same covariance.

→ But given the sample covariance S,

$\sigma_{11} = 0.0443$, $\sigma_{22} = 0.4533$ $\sigma_{33} = 0.0198$, $\sigma_{44} = 110.2107$,

every variable Don't have same variance.

Covariances are also not same

Thus, Assumptions are unmet.

Not appropriate to perform compound symmetry test.

Question 3

$$S=\begin{pmatrix} 101.84 & 36.99 & 32.59 \\ 36.99 & 167.54 & 66.53 \\ 32.59 & 66.53 & 177.88 \end{pmatrix}$$

a. $H_0: \Sigma = \sigma^2 I \quad H_a: \Sigma \neq \sigma^2 I$

$$LR = \left(\frac{|S|}{[tr(S)/p]^p}\right)^{n/2}$$

$$u = (LR)^{2/n}$$

$$= \left(\frac{|S|}{[tr(S)/p]^p}\right)^{\frac{n}{2}\cdot\frac{2}{n}}$$

$$= \frac{|S|}{[tr(S)/p]^p}$$

$$= \frac{p^p\,|S|}{(tr(S))^p} = \frac{3^3 \times 2323342}{(447.26)^3}$$

$$= 0.701$$

$|S| = 2323342$

$tr(S) = 101.84 + 167.54 + 177.88 = 447.26$

$p = 3$

$n = 15$

$\nu = 15 - 1 = 14$

12.722×0.355

$$u' = -\left(\nu - \frac{2p^2+p+2}{6p}\right)\ln u$$

$$= -\left(14 - \frac{2\times 9+3+2}{6\times 3}\right)\ln 0.701$$

$$= 4.519$$

Critical value: $\chi^2_{0.05,\,\frac{1}{2}p(p+1)-1} = \chi^2_{0.05,\,\frac{1}{2}\times 3\times 4-1=5} = 11.070$

Reject H_0 if $u' > \chi^2_{0.05,5}$

$u' = 4.519 < 11.070$, Thus, Do Not Reject H_0.

At significance level of $\alpha = 0.05$, there's No sufficient evidence to claim H_a, that the 3 variables have sphericity.

b. $H_0: \Sigma = \sigma^2 \cdot [(1-\rho)I + \rho J]$ $H_a: \Sigma \neq \sigma^2[(1-\rho)I + \rho J]$

$$s^2 = \frac{1}{p}(S_{11} + S_{22} + S_{33} + S_{44})$$

$$= \frac{1}{3}(101.84 + 167.54 + 177.88)$$

$$= 149.087$$

$$s^2 r = \frac{2}{p(p-1)} \sum_{j<k} S_{jk}$$

$$= \frac{2}{3\times 2} \times (36.99 + 32.59 + 66.53)$$

$$= 45.37$$

$$r = \frac{s^2 r}{s^2} = \frac{45.37}{149.087} = 0.304$$

$$u = \frac{|S|}{(s^2)^p (1-r)^{p-1} [1+(p-1)r]}$$

$$= \frac{2323342}{(149.087)^3 \underbrace{(1-0.304)^2}_{0.484416} \underbrace{[1+2\times 0.304]}_{1.608}} = \frac{2323342}{2581213.043}$$

$$= 0.900$$

$$u' = -\left[v - \frac{p(p+1)^2(2p-3)}{6(p-1)(p^2+p-4)}\right] \ln u$$

$$= -\left[14 - \frac{3\times 4^2 \times (6-3)}{6\times 2\times (9+3-4)}\right] \ln 0.900$$

$$= 1.317$$

Critical Value: $\chi^2_{0.05,\ \frac{1}{2}p(p+1)-2} = \chi^2_{0.05,\ \frac{1}{2}\times 3\times 4-2=4} = 9.488$

Reject H_0 if $u' > \chi^2_{0.05,4}$

$u' = 1.317 < 9.488$, Thus, Do Not Reject H_0.

At significance level of $\alpha = 0.05$, there's No sufficient evidence to claim H_a, that 3 variables do not exibit compound symmetry.

c. similar to answer for d. in question 2.

Assumptions for sphericity: $H_0: \Sigma = \sigma^2 I$

All variables are independent.

And all variables have same variance.

Assumptions for compound symmetry: $H_0: \Sigma = \sigma^2[(1-\rho)I + \rho J]$

All Variables are related.

All variables have same variance.

Every pair of variables have same covariance.

The SAS System

The DISCRIM Procedure

Total Sample Size	120	DF Total	119
Variables	4	DF Within Classes	114
Classes	6	DF Between Classes	5

Number of Observations Read	120
Number of Observations Used	120

Class Level Information					
group	Variable Name	Frequency	Weight	Proportion	Prior Probability
1	_1	20	20.0000	0.166667	0.166667
2	_2	20	20.0000	0.166667	0.166667
3	_3	20	20.0000	0.166667	0.166667
4	_4	20	20.0000	0.166667	0.166667
5	_5	20	20.0000	0.166667	0.166667
6	_6	20	20.0000	0.166667	0.166667

Within Covariance Matrix Information		
group	Covariance Matrix Rank	Natural Log of the Determinant of the Covariance Matrix
1	4	-9.92175
2	4	-10.92651
3	4	-8.29031
4	4	-8.28070
5	4	-9.19996
6	4	-7.32765
Pooled	4	-7.27604

$\ln|S_i|$

The DISCRIM Procedure
Test of Homogeneity of Within Covariance Matrices

Chi-Square	DF	Pr > ChiSq
178.313647	50	<.0001

The SAS System

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.

The DISCRIM Procedure

Generalized Squared Distance to group						
From group	**1**	**2**	**3**	**4**	**5**	**6**
1	-9.92175	-6.02977	4.06297	-3.81662	12.94780	4.28403
2	-0.71456	-10.92651	-7.24773	-7.85618	-4.62142	-5.81940
3	45.44667	-6.52316	-8.29031	-5.85606	-7.86816	-6.86880
4	3.82052	-9.75623	-7.67764	-8.28070	-7.07232	-6.65302
5	132.07608	8.31366	-6.85477	4.02345	-9.19996	-5.14409
6	22.47290	-4.84502	-7.41469	-7.75803	-6.87492	-7.32765

The DISCRIM Procedure
Classification Summary for Calibration Data: WORK.RICE
Resubstitution Summary using Quadratic Discriminant Function

Number of Observations and Percent Classified into group							
From group	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
1	14 70.00	5 25.00	1 5.00	0 0.00	0 0.00	0 0.00	20 100.00
2	2 10.00	15 75.00	2 10.00	1 5.00	0 0.00	0 0.00	20 100.00
3	0 0.00	6 30.00	4 20.00	1 5.00	6 30.00	3 15.00	20 100.00
4	0 0.00	7 35.00	1 5.00	11 55.00	1 5.00	0 0.00	20 100.00
5	0 0.00	1 5.00	2 10.00	4 20.00	12 60.00	1 5.00	20 100.00
6	3 15.00	1 5.00	3 15.00	4 20.00	3 15.00	6 30.00	20 100.00
Total	19 15.83	35 29.17	13 10.83	21 17.50	22 18.33	10 8.33	120 100.00
Priors	0.16667	0.16667	0.16667	0.16667	0.16667	0.16667	

The SAS System

Error Count Estimates for group							
	1	2	3	4	5	6	Total
Rate	0.3000	0.2500	0.8000	0.4500	0.4000	0.7000	0.4833
Priors	0.1667	0.1667	0.1667	0.1667	0.1667	0.1667	

The STEPDISC Procedure

The Method for Selecting Variables is STEPWISE			
Total Sample Size	120	**Variable(s) in the Analysis**	4
Class Levels	6	**Variable(s) Will Be Included**	0
		Significance Level to Enter	0.15
		Significance Level to Stay	0.15

Number of Observations Read	120
Number of Observations Used	120

Class Level Information				
group	Variable Name	Frequency	Weight	Proportion
1	_1	20	20.0000	0.166667
2	_2	20	20.0000	0.166667
3	_3	20	20.0000	0.166667
4	_4	20	20.0000	0.166667
5	_5	20	20.0000	0.166667
6	_6	20	20.0000	0.166667

The STEPDISC Procedure
Within-Class Correlation Coefficients / Pr > |r|

STEPWISE SELECTION

group = 1				
Variable	**width**	**thickness**	**lwratio**	**volume**
width	1.0000	0.0511 0.8305	-0.6235 0.0033	0.3084 0.1859
thickness	0.0511 0.8305	1.0000	0.0291 0.9032	0.9500 <.0001
lwratio	-0.6235 0.0033	0.0291 0.9032	1.0000	-0.0047 0.9841
volume	0.3084 0.1859	0.9500 <.0001	-0.0047 0.9841	1.0000

group = 2				
Variable	**width**	**thickness**	**lwratio**	**volume**
width	1.0000	0.4808 0.0319	-0.4241 0.0624	0.8377 <.0001
thickness	0.4808 0.0319	1.0000	-0.3852 0.0935	0.7203 0.0003
lwratio	-0.4241 0.0624	-0.3852 0.0935	1.0000	-0.0804 0.7361
volume	0.8377 <.0001	0.7203 0.0003	-0.0804 0.7361	1.0000

group = 3				
Variable	**width**	**thickness**	**lwratio**	**volume**
width	1.0000	0.1417 0.5513	-0.8339 <.0001	0.6203 0.0035
thickness	0.1417 0.5513	1.0000	0.0010 0.9967	0.7758 <.0001
lwratio	-0.8339 <.0001	0.0010 0.9967	1.0000	-0.2568 0.2744
volume	0.6203 0.0035	0.7758 <.0001	-0.2568 0.2744	1.0000

STEPWISE SELECTION

group = 4				
Variable	**width**	**thickness**	**lwratio**	**volume**
width	1.0000	-0.1580 0.5059	-0.3360 0.1475	0.5343 0.0152
thickness	-0.1580 0.5059	1.0000	0.0650 0.7856	0.5503 0.0119
lwratio	-0.3360 0.1475	0.0650 0.7856	1.0000	0.3375 0.1456
volume	0.5343 0.0152	0.5503 0.0119	0.3375 0.1456	1.0000

group = 5				
Variable	**width**	**thickness**	**lwratio**	**volume**
width	1.0000	0.5042 0.0234	-0.6156 0.0039	0.7837 <.0001
thickness	0.5042 0.0234	1.0000	-0.3855 0.0932	0.8311 <.0001
lwratio	-0.6156 0.0039	-0.3855 0.0932	1.0000	-0.2726 0.2449
volume	0.7837 <.0001	0.8311 <.0001	-0.2726 0.2449	1.0000

group = 6				
Variable	**width**	**thickness**	**lwratio**	**volume**
width	1.0000	-0.0827 0.7289	-0.7380 0.0002	0.6075 0.0045
thickness	-0.0827 0.7289	1.0000	-0.0148 0.9507	0.5082 0.0221
lwratio	-0.7380 0.0002	-0.0148 0.9507	1.0000	-0.1311 0.5817
volume	0.6075 0.0045	0.5082 0.0221	-0.1311 0.5817	1.0000

The STEPDISC Procedure

STEPWISE SELECTION

Total-Sample Correlation Coefficients / Pr > |r|				
Variable	**width**	**thickness**	**lwratio**	**volume**
width	1.0000	0.2484 0.0062	-0.4671 <.0001	0.6214 <.0001
thickness	0.2484 0.0062	1.0000	0.1258 0.1710	0.8183 <.0001
lwratio	-0.4671 <.0001	0.1258 0.1710	1.0000	0.1757 0.0549
volume	0.6214 <.0001	0.8183 <.0001	0.1757 0.0549	1.0000

The STEPDISC Procedure
Simple Statistics

Total-Sample					
Variable	**N**	**Sum**	**Mean**	**Variance**	**Standard Deviation**
width	120	10.18000	0.08483	0.08298	0.2881
thickness	120	43.27000	0.36058	0.19207	0.4383
lwratio	120	20.12618	0.16772	0.05413	0.2327
volume	120	1398	11.65018	131.45283	11.4653

---

group = 1					
Variable	**N**	**Sum**	**Mean**	**Variance**	**Standard Deviation**
width	20	-2.41000	-0.12050	0.04431	0.2105
thickness	20	-2.05000	-0.10250	0.45335	0.6733
lwratio	20	0.66228	0.03311	0.01980	0.1407
volume	20	-66.69580	-3.33479	110.21067	10.4981

STEPWISE SELECTION

group = 2					
Variable	**N**	**Sum**	**Mean**	**Variance**	**Standard Deviation**
width	20	3.12000	0.15600	0.04863	0.2205
thickness	20	6.80000	0.34000	0.06796	0.2607
lwratio	20	1.28714	0.06436	0.01872	0.1368
volume	20	200.80410	10.04021	49.78572	7.0559

group = 3					
Variable	**N**	**Sum**	**Mean**	**Variance**	**Standard Deviation**
width	20	4.25000	0.21250	0.09243	0.3040
thickness	20	9.66000	0.48300	0.09382	0.3063
lwratio	20	2.76271	0.13814	0.05682	0.2384
volume	20	329.02987	16.45149	73.12798	8.5515

group = 4					
Variable	**N**	**Sum**	**Mean**	**Variance**	**Standard Deviation**
width	20	0.82000	0.04100	0.08349	0.2890
thickness	20	7.53000	0.37650	0.10034	0.3168
lwratio	20	3.64553	0.18228	0.05961	0.2442
volume	20	209.76339	10.48817	65.99479	8.1237

group = 5					
Variable	**N**	**Sum**	**Mean**	**Variance**	**Standard Deviation**
width	20	3.78000	0.18900	0.05394	0.2322
thickness	20	12.04000	0.60200	0.09348	0.3057
lwratio	20	4.99402	0.24970	0.01916	0.1384
volume	20	430.21934	21.51097	87.81837	9.3711

STEPWISE SELECTION

group = 6					
Variable	**N**	**Sum**	**Mean**	**Variance**	**Standard Deviation**
width	20	0.62000	0.03100	0.11354	0.3370
thickness	20	9.29000	0.46450	0.07911	0.2813
lwratio	20	6.77449	0.33872	0.09557	0.3091
volume	20	294.90092	14.74505	59.15381	7.6912

The STEPDISC Procedure

Total-Sample Standardized Class Means						
Variable	**1**	**2**	**3**	**4**	**5**	**6**
width	-0.712823570	0.247058169	0.443200369	-0.152169317	0.361619100	-0.186884751
thickness	-1.056633450	-0.046965712	0.279322393	0.036317615	0.550848858	0.237110296
lwratio	-0.578565795	-0.444276361	-0.127153693	0.062575669	0.352386065	0.735034115
volume	-1.306986103	-0.140421842	0.418769388	-0.101350468	0.860055612	0.269933414

Pooled Within-Class Standardized Class Means						
Variable	**1**	**2**	**3**	**4**	**5**	**6**
width	-0.761424807	0.263902916	0.473418346	-0.162544419	0.386274760	-0.199626796
thickness	-1.203689817	-0.053502139	0.318196930	0.041372099	0.627512939	0.270109988
lwratio	-0.634901115	-0.487535833	-0.139534729	0.068668701	0.386698120	0.806604855
volume	-1.737879511	-0.186716785	0.556831276	-0.134764174	1.143602846	0.358926348

The STEPDISC Procedure
Stepwise Selection: Step 1

Statistics for Entry, DF = 5, 114				
Variable	**R-Square**	**F Value**	**Pr > F**	**Tolerance**
width	0.1604	4.36	0.0012	1.0000
thickness	0.2618	8.09	<.0001	1.0000
lwratio	0.2045	5.86	<.0001	1.0000
volume	0.4582	19.28	<.0001	1.0000

Variable volume will be entered.

STEPWISE SELECTION

Variable(s) That Have Been Entered
volume

Multivariate Statistics					
Statistic	**Value**	**F Value**	**Num DF**	**Den DF**	**Pr > F**
Wilks' Lambda	0.541827	19.28	5	114	<.0001
Pillai's Trace	0.458173	19.28	5	114	<.0001
Average Squared Canonical Correlation	0.091635				

The STEPDISC Procedure
Stepwise Selection: Step 2

Statistics for Removal, DF = 5, 114			
Variable	**R-Square**	**F Value**	**Pr > F**
volume	0.4582	19.28	<.0001

No variables can be removed.

Statistics for Entry, DF = 5, 113				
Variable	**Partial R-Square**	**F Value**	**Pr > F**	**Tolerance**
width	0.0968	2.42	0.0398	0.6138
thickness	0.0296	0.69	0.6333	0.3304
lwratio	0.1816	5.02	0.0003	0.9691

Variable lwratio will be entered.

Variable(s) That Have Been Entered	
lwratio	volume

STEPWISE SELECTION

Multivariate Statistics					
Statistic	**Value**	**F Value**	**Num DF**	**Den DF**	**Pr > F**
Wilks' Lambda	0.443414	11.34	10	226	<.0001
Pillai's Trace	0.606951	9.93	10	228	<.0001
Average Squared Canonical Correlation	0.121390				

The STEPDISC Procedure
Stepwise Selection: Step 3

Statistics for Removal, DF = 5, 113			
Variable	**Partial R-Square**	**F Value**	**Pr > F**
lwratio	0.1816	5.02	0.0003
volume	0.4426	17.95	<.0001

No variables can be removed.

Statistics for Entry, DF = 5, 112				
Variable	**Partial R-Square**	**F Value**	**Pr > F**	**Tolerance**
width	0.0223	0.51	0.7674	0.2711
thickness	0.0287	0.66	0.6536	0.3251

No variables can be entered.

No further steps are possible.

The STEPDISC Procedure

Stepwise Selection Summary										
Step	**Number In**	**Entered**	**Removed**	**Partial R-Square**	**F Value**	**Pr > F**	**Wilks' Lambda**	**Pr < Lambda**	**Average Squared Canonical Correlation**	**Pr > ASCC**
1	1	volume		0.4582	19.28	<.0001	0.54182683	<.0001	0.09163463	<.0001
2	2	lwratio		0.1816	5.02	0.0003	0.44341391	<.0001	0.12139023	<.0001

```
ods rtf file='hw4.rtf';

DATA rice;
  INFILE 'grain.csv' firstobs=2 dsd;
  INPUT group width thickness lwratio volume;
  RUN;

  ODS STARTPAGE = NO;


/*Add SAS code for Homework 4 below this line and above green line of stars below*/
PROC DISCRIM DATA=rice POOL=TEST;
CLASS group;
run;


PROC STEPDISC STEPWISE SIMPLE STDMEAN TCORR WCORR;
CLASS GROUP;
TITLE 'STEPWISE SELECTION';
RUN;

/*******************************************/

ods rtf close;
```

HW3 output

The SAS System

The GLM Procedure

Class Level Information		
Class	**Levels**	**Values**
group	6	1 2 3 4 5 6

Number of Observations Read	120
Number of Observations Used	120

The GLM Procedure

Dependent Variable: width

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	1.58390667	0.31678133	4.36	0.0012
Error	114	8.29029000	0.07272184		
Corrected Total	119	9.87419667			

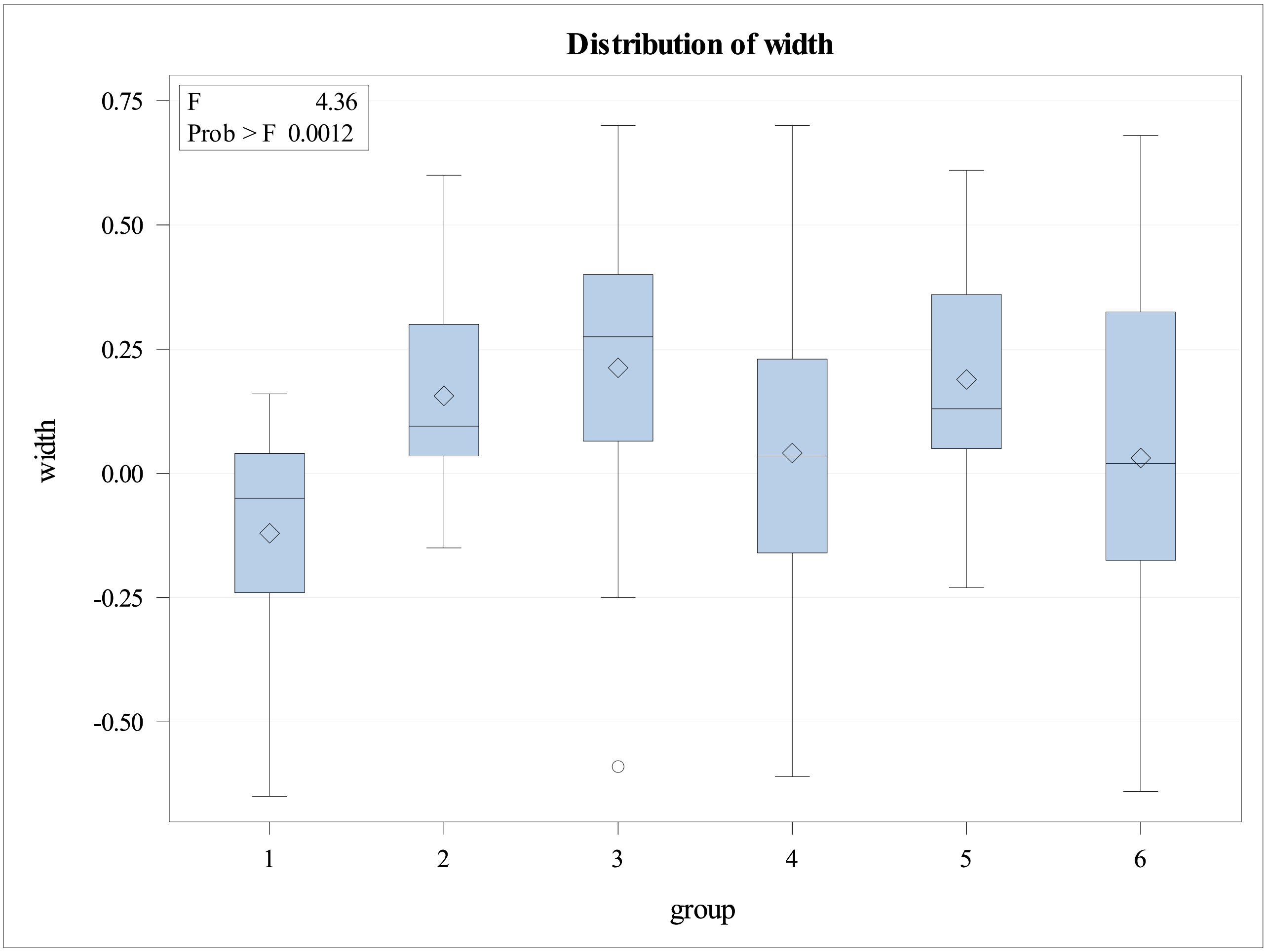
R-Square	Coeff Var	Root MSE	width Mean
0.160409	317.8820	0.269670	0.084833

Source	DF	Type I SS	Mean Square	F Value	Pr > F
group	5	1.58390667	0.31678133	4.36	0.0012

Source	DF	Type III SS	Mean Square	F Value	Pr > F
group	5	1.58390667	0.31678133	4.36	0.0012

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
contrast	1	0.00936333	0.00936333	0.13	0.7204

The SAS System



The GLM Procedure

Dependent Variable: thickness

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	5.98379417	1.19675883	8.09	<.0001
Error	114	16.87306500	0.14800934		
Corrected Total	119	22.85685917			

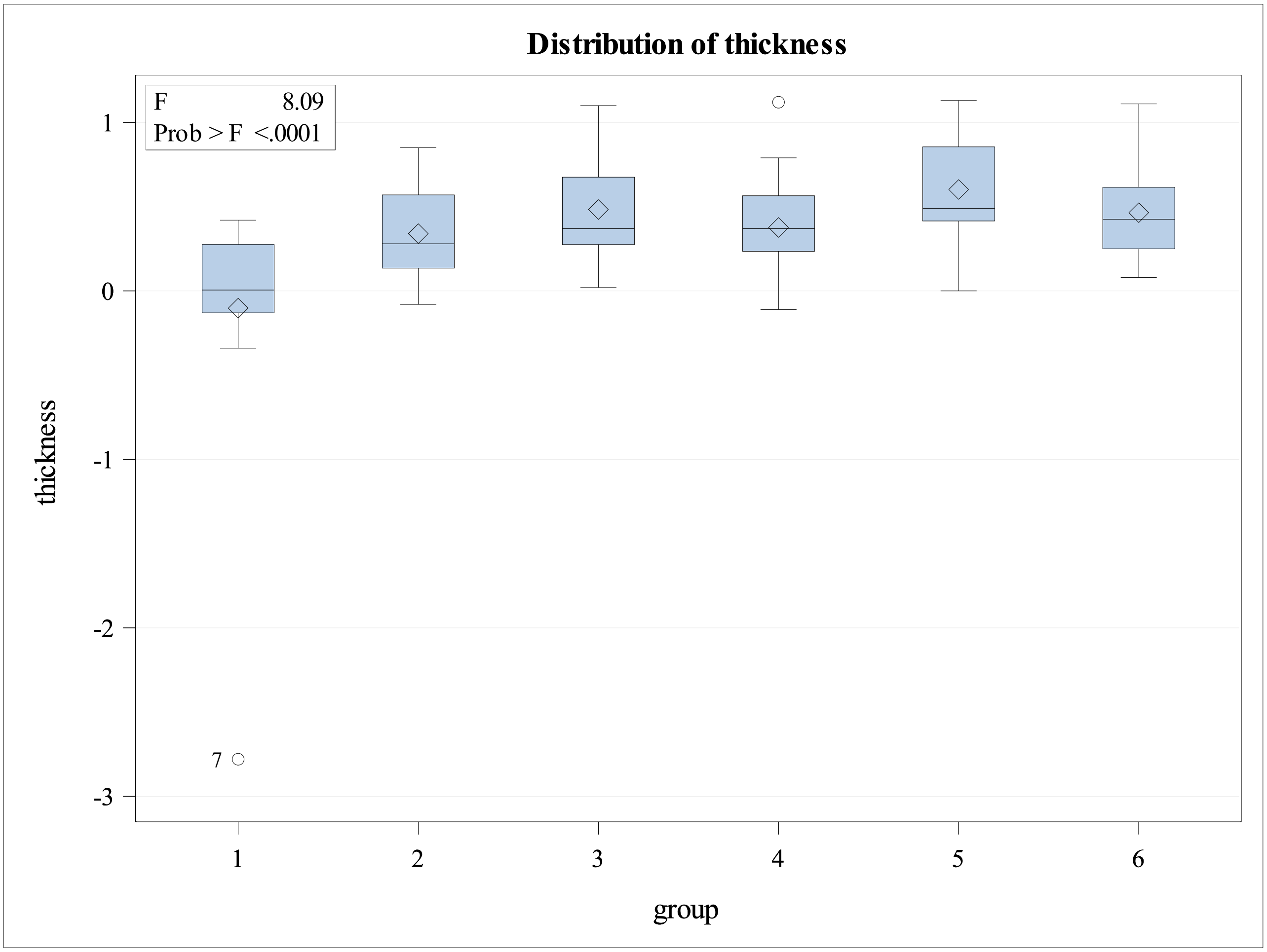
R-Square	Coeff Var	Root MSE	thickness Mean
0.261794	106.6937	0.384720	0.360583

Source	DF	Type I SS	Mean Square	F Value	Pr > F
group	5	5.98379417	1.19675883	8.09	<.0001

The SAS System

Source	DF	Type III SS	Mean Square	F Value	Pr > F
group	5	5.98379417	1.19675883	8.09	<.0001

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
contrast	1	0.13134083	0.13134083	0.89	0.3482



The GLM Procedure

Dependent Variable: lwratio

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	1.31706575	0.26341315	5.86	<.0001
Error	114	5.12399023	0.04494728		
Corrected Total	119	6.44105598			

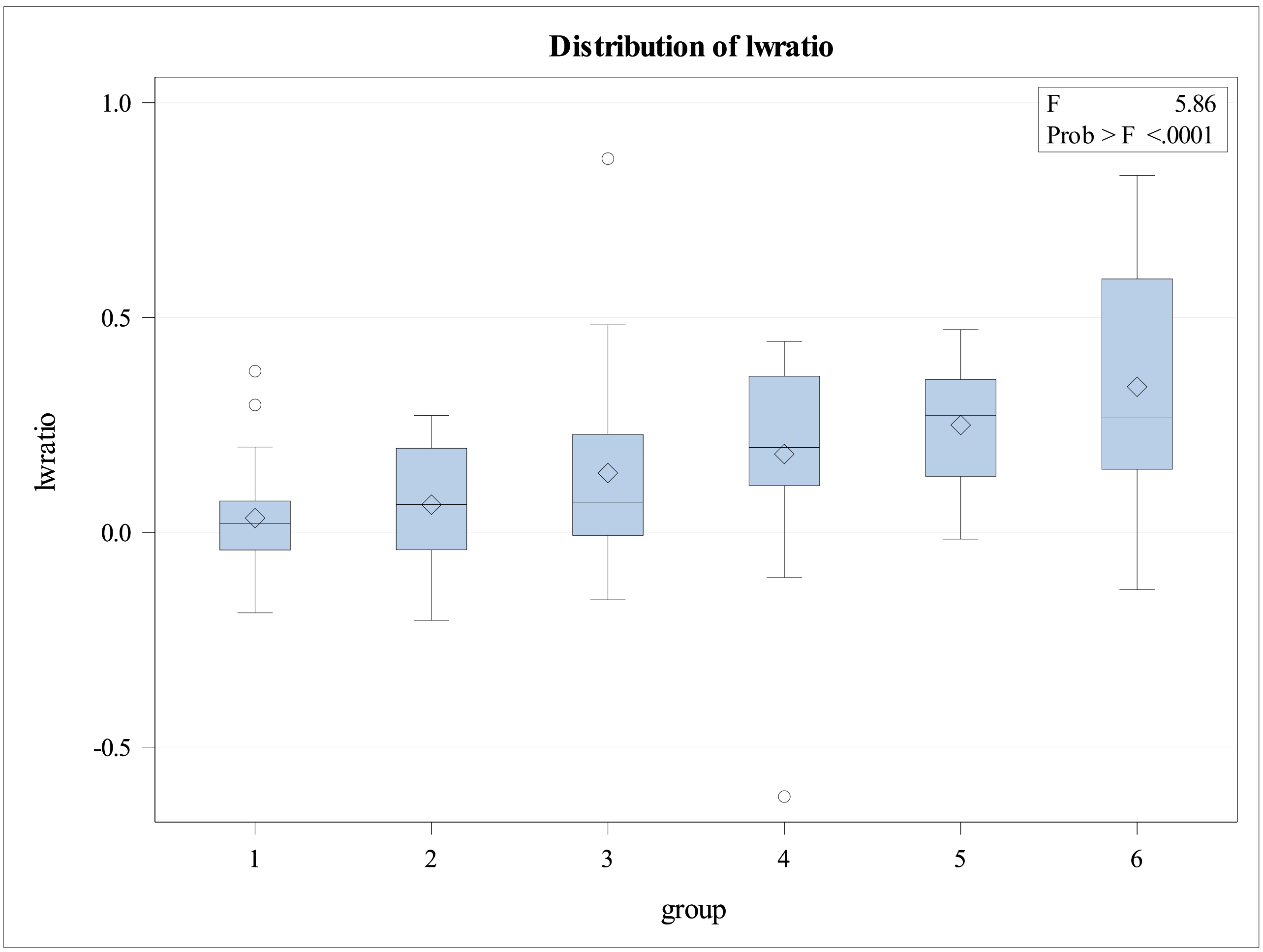
The SAS System

R-Square	Coeff Var	Root MSE	lwratio Mean
0.204480	126.4071	0.212008	0.167718

Source	DF	Type I SS	Mean Square	F Value	Pr > F
group	5	1.31706575	0.26341315	5.86	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
group	5	1.31706575	0.26341315	5.86	<.0001

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
contrast	1	0.09009865	0.09009865	2.00	0.1596



The GLM Procedure

Dependent Variable: volume

The SAS System

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	7167.15091	1433.43018	19.28	<.0001
Error	114	8475.73547	74.34856		
Corrected Total	119	15642.88637			

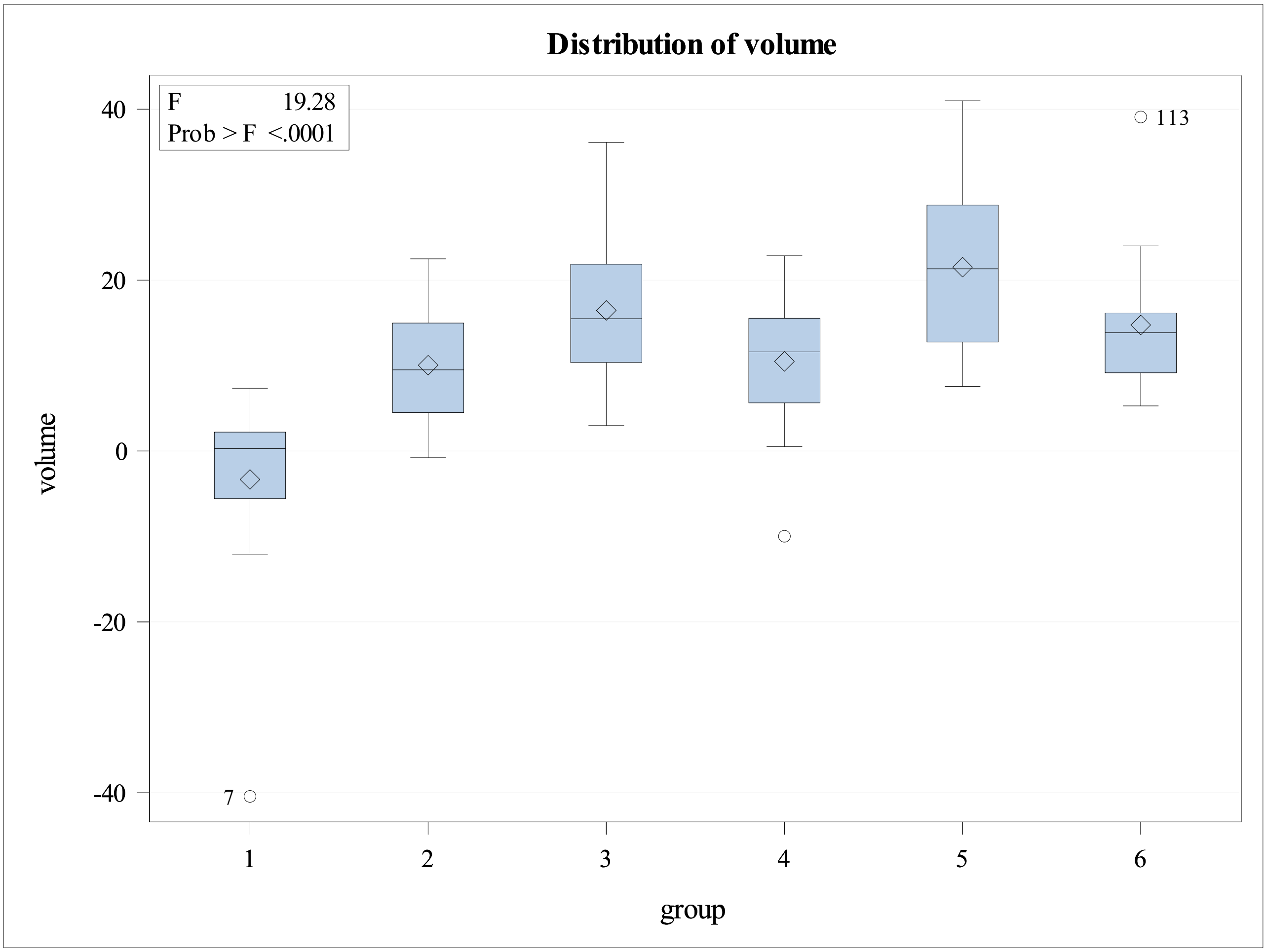
R-Square	Coeff Var	Root MSE	volume Mean
0.458173	74.01224	8.622561	11.65018

Source	DF	Type I SS	Mean Square	F Value	Pr > F
group	5	7167.150905	1433.430181	19.28	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
group	5	7167.150905	1433.430181	19.28	<.0001

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
contrast	1	1.38997817	1.38997817	0.02	0.8915

The SAS System



The GLM Procedure
Multivariate Analysis of Variance

E

E = Error SSCP Matrix				
	width	**thickness**	**lwratio**	**volume**
width	8.29029	1.170045	-4.029202104	154.49735915
thickness	1.170045	16.873065	-0.44628408	284.4113197
lwratio	-4.029202104	-0.44628408	5.1239902254	-11.47549504
volume	154.49735915	284.4113197	-11.47549504	8475.7354672

The SAS System

Partial Correlation Coefficients from the Error SSCP Matrix / Prob > |r|				
DF = 114	**width**	**thickness**	**lwratio**	**volume**
width	1.000000	0.098928 0.2929	-0.618201 <.0001	0.582838 <.0001
thickness	0.098928 0.2929	1.000000	-0.047997 0.6105	0.752075 <.0001
lwratio	-0.618201 <.0001	-0.047997 0.6105	1.000000	-0.055065 0.5589
volume	0.582838 <.0001	0.752075 <.0001	-0.055065 0.5589	1.000000

The GLM Procedure
Multivariate Analysis of Variance

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

	Characteristic Root	Percent	Characteristic Vector V'EV=1			
			width	**thickness**	**lwratio**	**volume**
λ_1	**1.05758634**	70.37	-0.12987286	-0.12681120	0.09598563	0.01633480
λ_2	**0.33007068**	21.96	-1.60599251	-0.96010780	-1.25300712	0.05899748
λ_3	**0.11160307**	7.43	0.12443219	0.02337911	-0.30090176	0.00121369
λ_4	**0.00373024**	0.25	0.32207304	-0.17711522	0.26723234	-0.00148523

MANOVA Tests for the Hypothesis of No Overall group Effect
H = Type III SSCP Matrix for group
E = Error SSCP Matrix

S=4 M=0 N=54.5

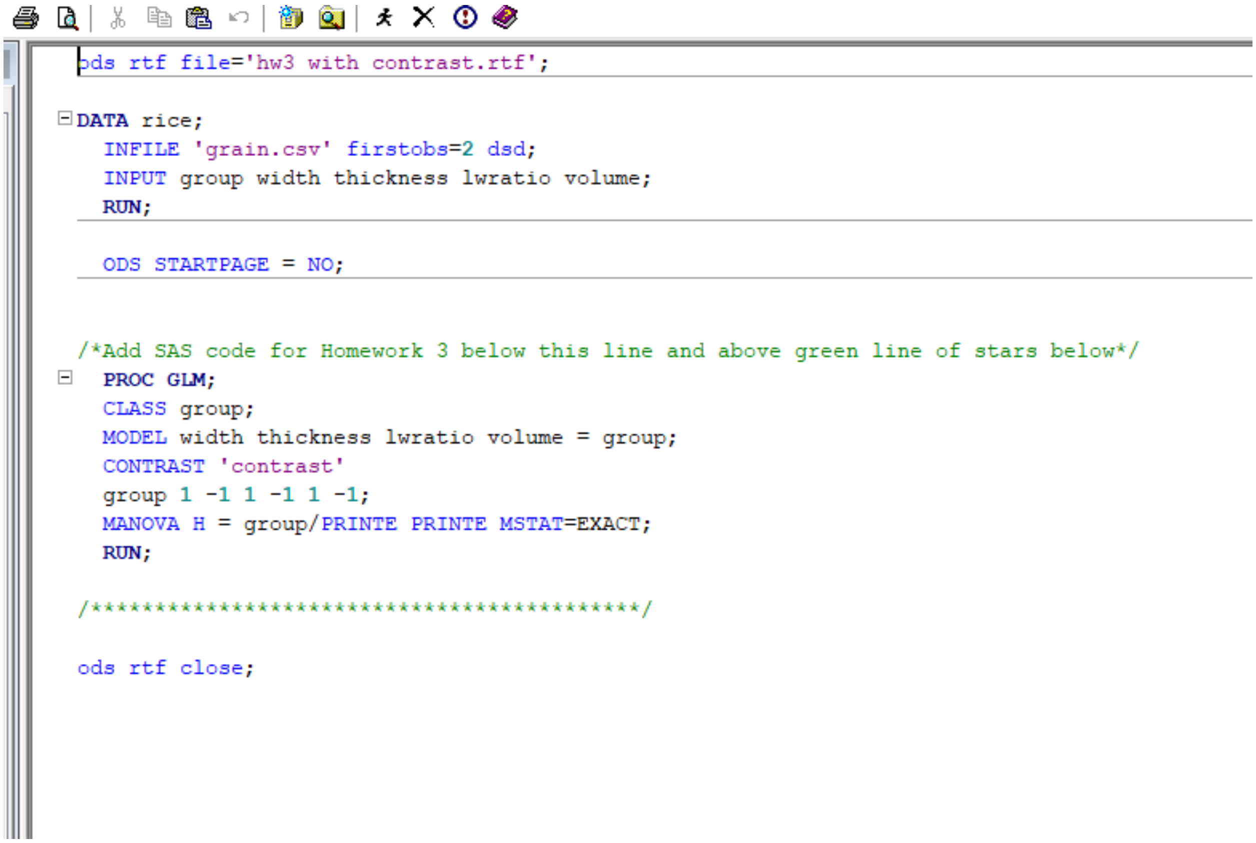
Statistic	Value	P-Value
Wilks' Lambda	0.32749183	<.0001
Pillai's Trace	0.86626861	<.0001
Hotelling-Lawley Trace	1.50299033	<.0001
Roy's Greatest Root	1.05758634	<.0001

The SAS System

Characteristic Roots and Vectors of: E Inverse * H, where H = Contrast SSCP Matrix for contrast E = Error SSCP Matrix					
Characteristic Root	Percent	Characteristic Vector V'EV=1			
		width	thickness	lwratio	volume
0.25815117	100.00	-1.54049148	-0.94110022	-1.27949481	0.05765378
0.00000000	0.00	0.45018911	0.06238517	0.08481721	-0.00382185
0.00000000	0.00	-0.32426570	0.07472546	-0.23663604	0.01066277
0.00000000	0.00	-0.18536681	-0.27323805	0.20223140	0.01729011

M

MANOVA Tests for the Hypothesis of No Overall contrast Effect H = Contrast SSCP Matrix for contrast E = Error SSCP Matrix S=1 M=1 N=54.5		
Statistic	Value	P-Value
Wilks' Lambda	0.79481705	<.0001
Pillai's Trace	0.20518295	<.0001
Hotelling-Lawley Trace	0.25815117	<.0001
Roy's Greatest Root	0.25815117	<.0001



```
ods rtf file='hw3 with contrast.rtf';

DATA rice;
  INFILE 'grain.csv' firstobs=2 dsd;
  INPUT group width thickness lwratio volume;
  RUN;

  ODS STARTPAGE = NO;


/*Add SAS code for Homework 3 below this line and above green line of stars below*/
  PROC GLM;
  CLASS group;
  MODEL width thickness lwratio volume = group;
  CONTRAST 'contrast'
  group 1 -1 1 -1 1 -1;
  MANOVA H = group/PRINTE PRINTE MSTAT=EXACT;
  RUN;

/*******************************************/

ods rtf close;
```


```
ods rtf file='hw3 with contrast.rtf';

DATA rice;
  INFILE 'grain.csv' firstobs=2 dsd;
  INPUT group width thickness lwratio volume;
  RUN;

  ODS STARTPAGE = NO;


/*Add SAS code for Homework 3 below this line and above green line of stars below*/
  PROC GLM;
  CLASS group;
  MODEL width thickness lwratio volume = group;
  CONTRAST 'contrast'
  group 1 -1 1 -1 1 -1;
  MANOVA H = group/PRINTE PRINTE MSTAT=EXACT;
  RUN;

/*******************************************/

ods rtf close;
```