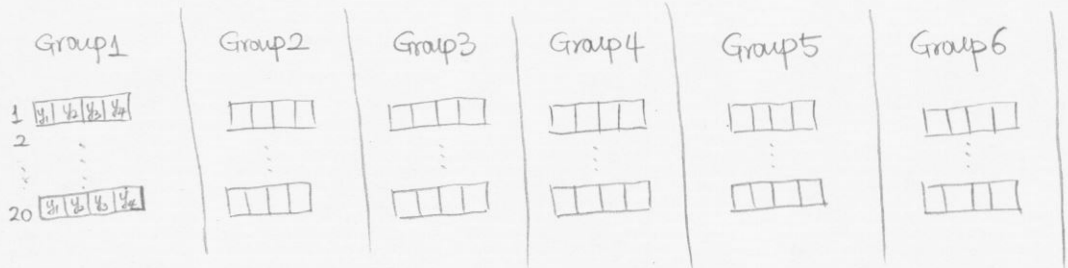
Question 1.

Homework 3
Stat 764
Qianqian Tao
1253177

37/40



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

a. Assumptions:

- The observation vectors from each of the 6 populations come from a multivariate Normal distributions.
- Each of 6 populations has the same population covariance matrix (Σ).
- 6 independent samples from 6 different populations.
- ~~No multicollinearity~~, 4 dependent variables moderately correlated with each other.

Assumptions not stated in context -0.5

This is not an assumption -0.5

b. Factor Effects Model

$$\underline{y}_{ij} = \underline{\mu} + \underline{\alpha}_i + \underline{\epsilon}_{ij}$$

$$= \cancel{\underline{\mu}_i + } \underline{\epsilon}_{ij}$$

$i = 1, 2, 3, 4, 5, 6$ $(k=6)$

$j = 1, 2, \cdots, 20$ $(n=20)$

Define y_{ij} -0.5

Definitions not in context -0.5

$\underline{\mu}$: overall population mean vector

$\underline{\alpha}_i$: effect vector of ith population

$\underline{\mu}_i = \underline{\mu} + \underline{\alpha}_i$: population mean vector of ith population.

$\underline{\epsilon}_{ij} \sim N_p(\underline{0}, \Sigma)$: error vector for jth observation in ith sample.

c. Assumptions of Multivariate Normal Distributions.

By Q-Q plot for y_1 y_2 y_3 y_4 in each group of 6 groups,

All points are close to a straight line, → There are clear outliers in some of the Q-Q plots

There's no indication of departure from normality,

Thus, all variables of y_1 y_2 y_3 y_4 in each group follow Normality seperately.

By Scatterplots of y_1 vs y_2, y_1 vs. y_3 y_1 vs. y_4

y_2 vs. y_3, y_2 vs. y_4

y_3 vs. y_4

There's No indication of a curved trend or other obvious nonnormality.
Although there are few outliers in group 1, group 4, But generally it is still can be inferred that each pair of variables has a bivariate normal distribution.

There's No indication of a clear deviation from MVN distribution. It's appropriate to assume that these observation vectors come from a multivariate Normal distribution.

d. $H_0: \underline{\mu}_1 = \underline{\mu}_2 = \underline{\mu}_3 = \underline{\mu}_4 = \underline{\mu}_5 = \underline{\mu}_6$

H_a: at least two population mean vectors differ. ✓

e. see as attachments ✓

f. Non-zero eigenvalues of $E^{-1}H$:

$\lambda_1 = 1.0576$

$\lambda_2 = 0.3301$ ✓

$\lambda_3 = 0.1116$

$\lambda_4 = 0.0037$

g.

Wilks' Test: $\Lambda = \frac{|E|}{|E+H|} = \prod_{i=1}^{s} \frac{1}{1+\lambda_i}$

$= \frac{1}{1+1.0576} \cdot \frac{1}{1+0.3301} \cdot \frac{1}{1+0.1116} \cdot \frac{1}{1+0.0037}$

$= \frac{1}{2.0576} \cdot \frac{1}{1.3301} \cdot \frac{1}{1.1116} \cdot \frac{1}{1.0037}$

$= \frac{1}{3.0535} \doteq 0.3275$ ✓

$p=4$
$k=6$
$n=20$
$\nu_H = k-1 = 5$
$\nu_E = k(n-1) = 6 \times 19 = 114$

$S = \text{rank}(H)$
$= \min(p, \nu_H)$
$= \min(4, 5)$
$= 4$

$\Lambda_{\alpha, p, \nu_H, \nu_E} = \Lambda_{0.05, 4, 5, 114} \doteq 0.730$ ✓

Reject Ho if $\Lambda < \Lambda_{0.05, 4, 5, 114}$

$\Lambda \doteq 0.3275 < 0.730$

Thus, Reject Ho.

At significance level of $\alpha = 0.05$, there's sufficient evidence to claim Ha, which is that at least two population mean vectors differ. state in context of problem

h. Roy's Largest Root:

$\theta = \frac{\lambda_1}{1+\lambda_1} = \frac{1.0576}{1+1.0576} \doteq 0.5140$ ✓

$\theta_{\alpha, s, m, N} = \theta_{0.05, 4, 0, 54.5} \doteq 0.171$ ✓

$S = \min(p, \nu_H) = \min(4, 5) = 4$
$m = \frac{1}{2}(|\nu_H - p| - 1) = \frac{1}{2}(|5-4|-1) = 0$
$N = \frac{1}{2}(\nu_E - p - 1) = \frac{1}{2}(114-4-1)$
$= 54.5$

Reject Ho if $\theta \geq \theta_{0.05, 4, 0, 54.5}$

$\theta \doteq 0.5140 > 0.171$ ✓

Thus, Reject Ho. ✓

At significance level of $\alpha = 0.05$, there's sufficient evidence to claim Ha, which is that at least two population mean vectors differ.

i. Pillai's test

$$V^{(s)} = \sum_{i=1}^{s} \frac{\lambda_i}{1+\lambda_i} = \frac{1.0546}{1+1.0546} + \frac{0.3301}{1+0.3301} + \frac{0.1116}{1+0.1116} + \frac{0.0037}{1+0.0037} \qquad (s=4)$$

$$= 0.8662$$

$$V^{(s)}_{\alpha, m, N} = V^{(4)}_{0.05,\, 0,\, 54.5} \doteq 0.602$$

Reject H_0 if $V^{(s)} > V^{(s)}_{\alpha, m, N}$

$$V^{(4)} = 0.8662 > 0.602$$

Thus, Reject H_0.

At significance level of $\alpha = 0.05$, there's sufficient evidence to claim H_a, which is that at least two population mean vectors differ.

j. Lawley - Hotelling test

$$U^{(s)} = \sum_{i=1}^{s} \lambda_i$$

$$= 1.0546 + 0.3301 + 0.1116 + 0.0037$$

$$= 1.503$$

$$\frac{V_E}{V_H} U^{(s)} = \frac{114}{5} \times 1.503 = 34.268$$

$$\frac{V_E}{V_H} U^{(s)}_{\alpha, p} = \frac{114}{5} \times U^{(4)}_{0.05,\, 4} \doteq 3.8557$$ -0.5

Reject H_0 if $\frac{V_E}{V_H} U^{(s)} \geq \frac{V_E}{V_H} U^{(s)}_{\alpha, p}$

$$\frac{V_E}{V_H} U^{(s)} = 34.268 > 3.8557$$

Thus, Reject H_0.

At significance level of $\alpha = 0.05$, there's sufficient evidence to claim H_a, which is that at least two population mean vectors differ.

k.

Based on g. h. i. j. , Ho is not true.

Thus, the power of these tests differ in different situations.

The mean vectors are collinear, Because $E^{-1}H$ has one Large eigenvalue and the rest of the eigenvalues are small. I recommend λ_1 being at least 80% of sum of eigenvalues to indicate multicollinearity.

The power of the tests are ordered as:

$$\Theta \geq U^{(s)} \geq \Lambda \geq V^{(s)}$$

Roy's Largest Root ≥ Lawley-Hotelling test ≥ Wilks' test ≥ Pillai's test.

There are few outliers in group 1 and group 4, any of the other three tests other than Roy's test will still perform Reasonably well.

Λ remains the most popular test statistic for MANOVA because of its flexibility and historical dominance.

Based on power of the tests in this case,

I Recommend Lawley-Hotelling test first, then Wilks' test second.

l.

Variables	univariate F-Statistics	P-value	coefficient in a_1
width	4.36	0.0012	-0.1299
thickness	8.09	<0.0001	-0.1268
Lwratio	5.86	<0.0001	0.0960
Volumn	19.28	<0.0001	0.0163

significant ↓ less significant	Volumn Thickness ✓ Lwration width

by results of the univariate one-way ANOVAs.

Significant ↓ Less significant	width thickness ✓ Lwratio Volumn

by coefficients of the discriminant function.

Using the (unstardized) discriminant function is only appropriate if the variances are similar. -0.5

m. Ranking variables based on the coefficients of the discriminant function takes the correlations of the variables into account, it generally provides more pertinent information than Ranking Variables by a univariate ANOVA if the sample variances are similar among the different variables. I Recommed coefficients of the discriminant function.

Question 2. → Graded on completion

	Group1	Group2	Group3	Group4	Group5	Group6
	3H	5H	3H	5H	3H	5H
y_1	-0.120	0.156	0.212	0.041	0.189	0.031
y_2	-0.102	0.340	0.483	0.376	0.602	0.464
y_3	0.033	0.064	0.138	0.182	0.250	0.339
y_4	-3.335	10.040	16.451	10.488	21.511	14.745

a. $H_0: \mu_1 + \mu_3 + \mu_5 - \mu_2 - \mu_4 - \mu_6 = 0$ $\quad \sum c_i = 1+1+1-1-1-1=0$

$H_a: \mu_1 + \mu_3 + \mu_5 - \mu_2 - \mu_4 - \mu_6 \neq 0$

b. $\hat{\delta} = \sum_{i=1}^{6} c_i \bar{y}_{i.}$

$$= \begin{pmatrix} -0.120 \\ -0.102 \\ 0.033 \\ -3.335 \end{pmatrix} + \begin{pmatrix} 0.212 \\ 0.483 \\ 0.138 \\ 16.451 \end{pmatrix} + \begin{pmatrix} 0.189 \\ 0.602 \\ 0.250 \\ 21.511 \end{pmatrix} - \begin{pmatrix} 0.156 \\ 0.340 \\ 0.064 \\ 10.040 \end{pmatrix} - \begin{pmatrix} 0.041 \\ 0.376 \\ 0.182 \\ 10.488 \end{pmatrix} - \begin{pmatrix} 0.031 \\ 0.464 \\ 0.339 \\ 14.745 \end{pmatrix}$$

$$= \begin{pmatrix} -0.120 + 0.212 + 0.189 - 0.156 - 0.041 - 0.031 \\ -0.102 + 0.483 + 0.602 - 0.340 - 0.376 - 0.464 \\ 0.033 + 0.138 + 0.250 - 0.064 - 0.182 - 0.339 \\ -3.335 + 16.451 + 21.511 - 10.040 - 10.488 - 14.745 \end{pmatrix} = \begin{pmatrix} 0.053 \\ -0.197 \\ -0.164 \\ -0.646 \end{pmatrix}$$

c. $\text{Cov}(\hat{\delta}) = \Sigma \cdot \frac{1}{n} \cdot \sum_{i=1}^{k} c_i^2$

$$= \frac{E}{V_E} \cdot \frac{1}{n} \cdot \sum_{i=1}^{k} c_i^2$$

$$= E \cdot \frac{1}{V_E} \cdot \frac{1}{n} \cdot \sum_{i=1}^{k} c_i^2$$

$$= \begin{bmatrix} 8.290 & 1.170 & -4.029 & 154.497 \\ 1.170 & 16.873 & -0.446 & 284.411 \\ -4.029 & -0.446 & 5.124 & -11.475 \\ 154.497 & 284.411 & -11.475 & 8475.735 \end{bmatrix} \cdot \frac{1}{114} \cdot \frac{1}{20} \cdot 6$$

$$= \begin{bmatrix} 8.290 & 1.170 & -4.029 & 154.497 \\ 1.170 & 16.873 & -0.446 & 284.411 \\ -4.029 & -0.446 & 5.124 & -11.475 \\ 154.497 & 284.411 & -11.475 & 8475.735 \end{bmatrix} \times 0.00263$$

d. See Attach Files

contrast "contrast"

group 1 -1 1 -1 1 -1;

e. Non-zero eigenvalue of $E^{-1}H_1$

$\lambda = 0.2582$

f. Wilks' test:

$$\Lambda = \frac{|E|}{|E+H_1|} = \prod_{i=1}^{s} \frac{1}{1+\lambda_i} = \frac{1}{1+0.2582} \doteq 0.795$$

$\Lambda_{\alpha, p, 1, \nu_E} = \Lambda_{0.05, 4, 1, 114} \doteq 0.962$

Reject H_0 if $\Lambda < \Lambda_{\alpha, p, 1, \nu_E}$

$\Lambda = 0.795 < 0.962$,

Thus, Reject H_0.

At significance level of $\alpha = 0.05$, there's sufficient evidence to claim H_a, which is that the average of the population mean vectors for groups with a charring time of 3 hours differs from the average of the population mean vectors for groups with a charring time of 5 hours.

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			

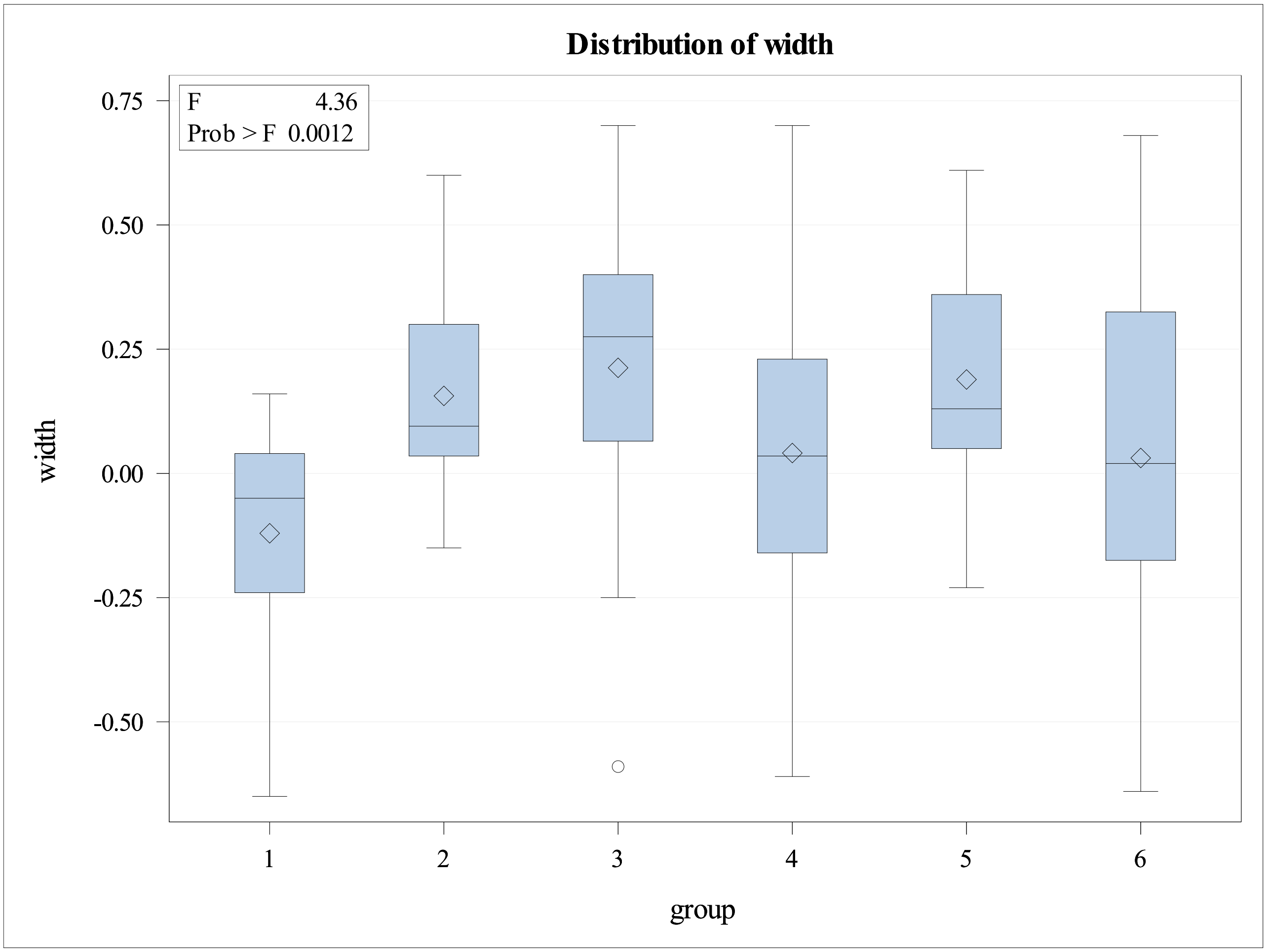
R-Square	Coeff Var	Root MSE	width Mean
0.160409	317.8820	0.269670	0.084833

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			

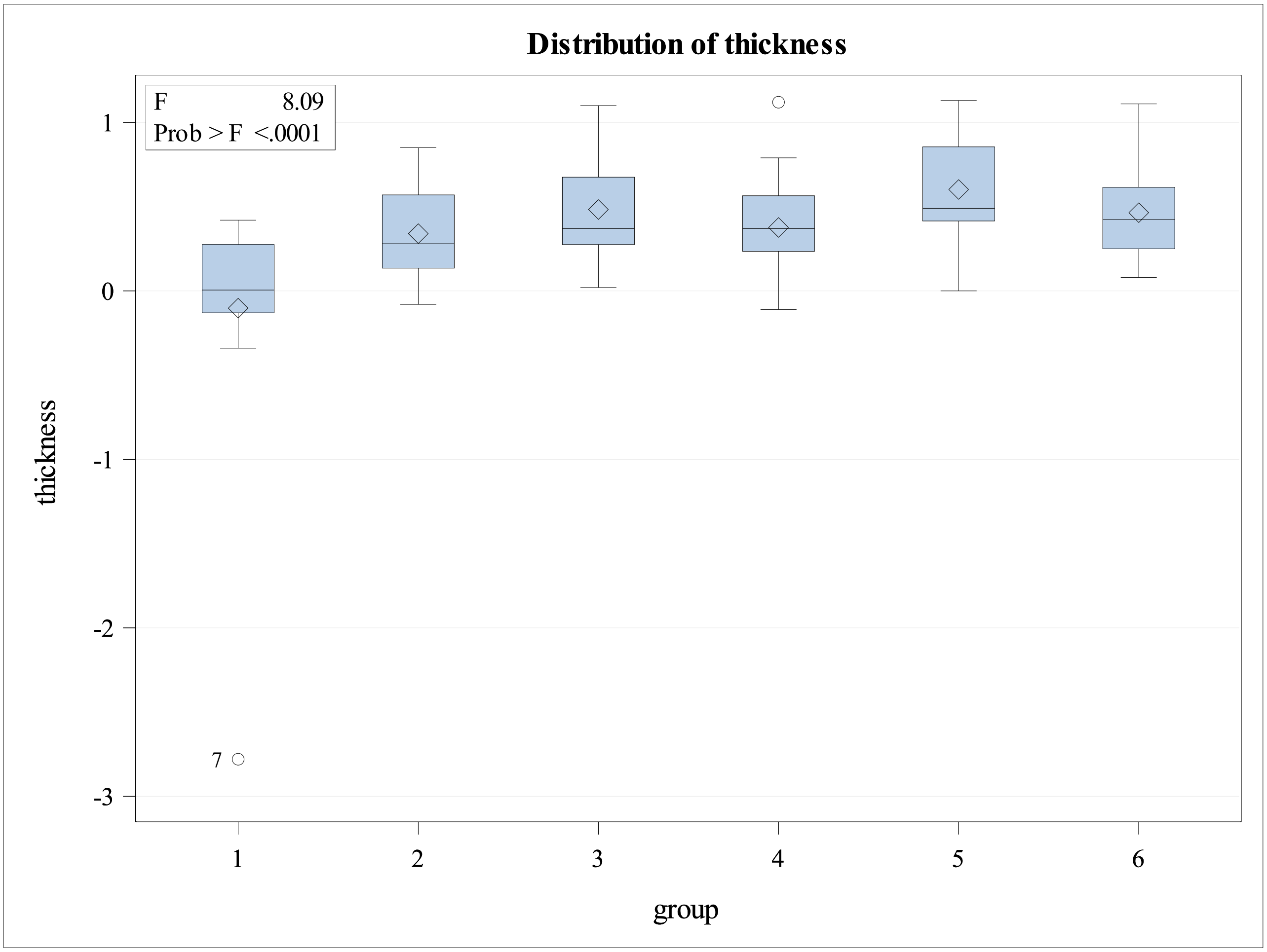
R-Square	Coeff Var	Root MSE	thickness Mean
0.261794	106.6937	0.384720	0.360583

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			

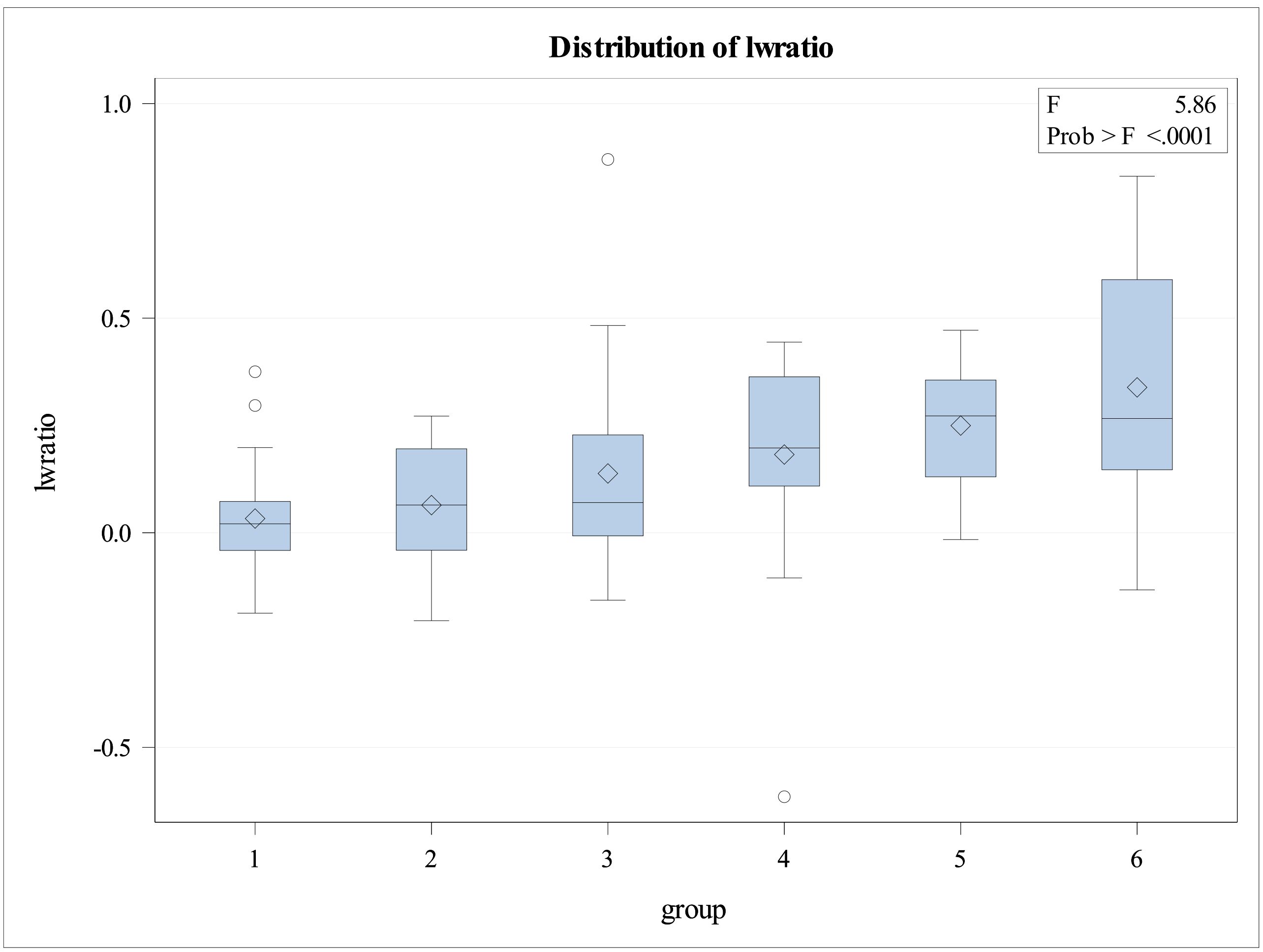
The SAS System

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			

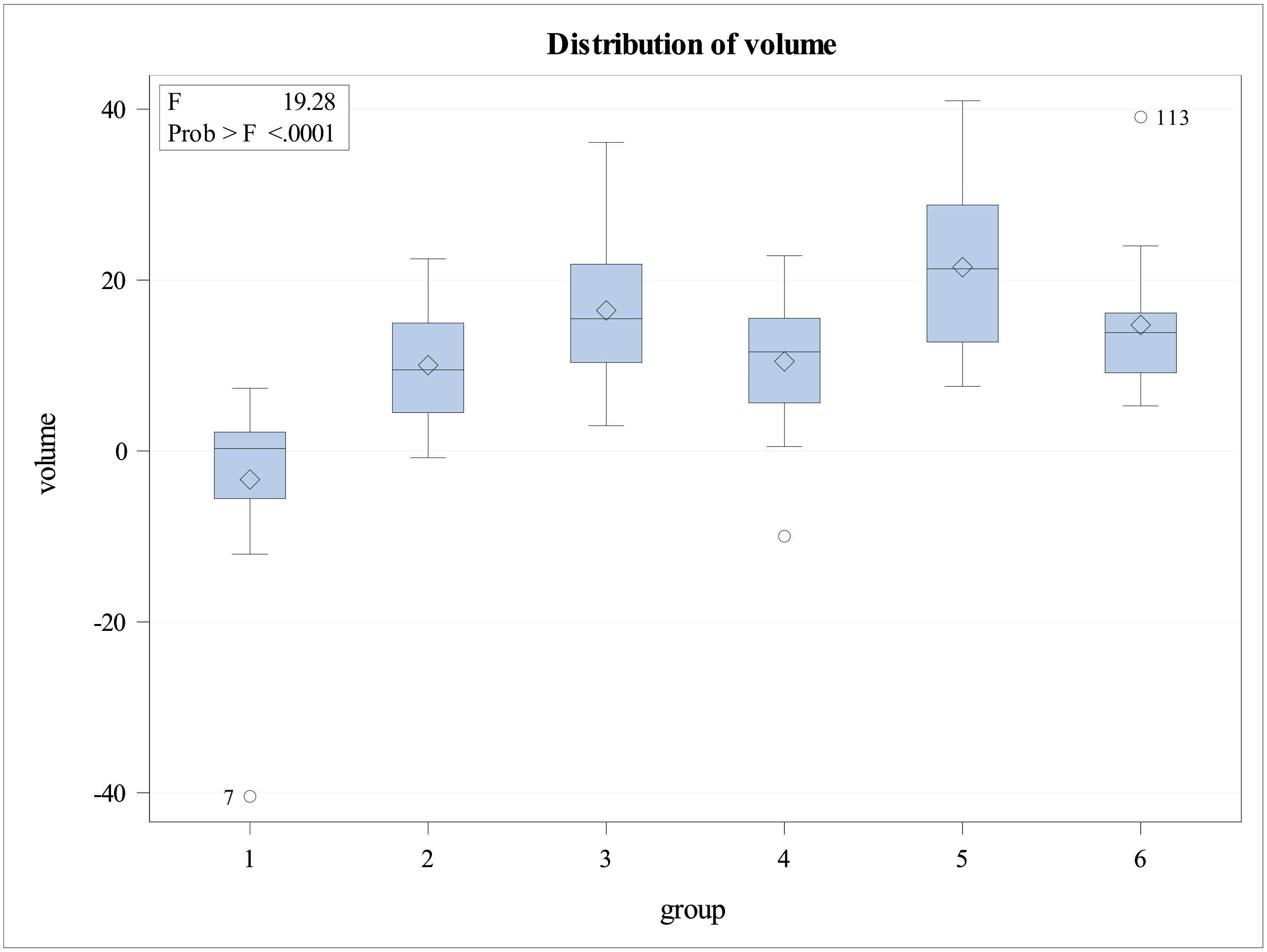
R-Square	Coeff Var	Root MSE	volume Mean
0.458173	74.01224	8.622561	11.65018

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;
```