Chapter 6. Comparison of Several Multivariate Means.

This chapter addresses comparison of several multivariate means. We begin with paired comparison followed by repeated measurement. In these two settings, comparison of two multivariate means or several univaraite means is transformed, by taking differences or using contrasting matrices, to the setting of one population, which is deemed as one-sample problem with the statistical methodology already introduced in the last chapter.

The next is the so called two sample problem with two independent samples each coming from one population. The objective is to compare the means of the two populations. The two-sample problem is then extended to several sample problem, which is treated by using the multivaraite analysis of variance (MANOVA).

6.1 Paired comparison.

(i). The univariate case — a review.

For a univariate paired comparison of the means of X and Y based on iid samples

$$(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n),$$

set $d_i = X_i - Y_i$ be the difference of X_i and Y_i. Let

$$\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i \qquad \text{and} \qquad s_d^2 = \frac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d})^2$$

be the sample mean and sample variance of d_i. Since d_i are iid with mean $\mu_X - \mu_Y$. Then $\mu_X - \mu_Y$ is estimated by $\bar{d}$ and inference with $\mu_X - \mu_Y$ is based on the fact that

$$\frac{\bar{d} - (\mu_X - \mu_Y)}{s_d/\sqrt{n}} \sim t_{n-1}.$$

In summary, paired comparison uses the same statistical procedure as that of one-sample problem. Here the "one-sample" refers to $d_1, ..., d_n$.

(ii). Multivariate paired comparison.

We wish to compare the mean vector of two p-dimensional random vectors X and Y, may be dependent of each other. With iid samples

$$(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n),$$

set $d_i = X_i - Y_i$. We shall only need the assumption

$$d_i \equiv \begin{pmatrix} d_{i1} \\ \vdots \\ d_{ip} \end{pmatrix} \quad iid \ \sim \ MN(\mu, \Sigma).$$

$\bar{d}$ is an obvious candidate of estimator of μ. As with the univariate paired comparison, the inference about μ are based on the "one sample": $d_1, ..., d_n$. Specifically, T^2-confidence region, T^2 simultaneous confidence intervals and Bonferroni's confidence intervals all apply.

Example 6.1. Analysis of the Effluent Data *Example 6.1 in the textbook* Municipal wastewater treatment plants are required by law to monitor their discharges into rivers and streams on a regular basis. Concern about the reliability of data from on e of these self-monitoring programs led to a study in which samples of effluent were divided and sent to two laboratories for testing. *One-half of each sample was sent to the Wisconsin State laboratory of Hygiene , and one-half was sent to a private commercial laboratory routinely used in the monitoring program.* Measurements of biochemical oxygen demand (BOD) and suspended solids (SS) were obtained, for $n = 11$ *sample splits* from the two laboratories. The data are displayed here:

Effluent Data:				
	Commercial lab		State lab	
Sample j	(BOD)	(SS)	(BOD)	SS
1	6	27	25	15
2	6	23	28	13
3	18	64	36	22
4	8	44	35	29
5	11	30	15	31
6	34	75	44	64
7	28	26	42	30
8	71	124	54	64
9	43	54	34	56
10	33	30	29	20
11	20	14	39	21

Our question in concern is whether there is enough statistical evidence to indicate the two lab analysis procedures are different in the sense that they produce systematically different results. In this example, sample size $n = 11$ and $p = 2$.

$$\bar{d} = \begin{pmatrix} -9.36 \\ 13.27 \end{pmatrix} \qquad \mathbf{S}_d = \begin{pmatrix} 199.26 & 88.38 \\ 88.38 & 418.61 \end{pmatrix}, \qquad \mathbf{S}_d^{-1} = \begin{pmatrix} .0055 & -.0012 \\ -.0012 & .0026 \end{pmatrix}.$$

(1). T^2 confidence region for $\mu_1 - \mu_2$ at confidence level 95%:

$$\left\{\mu_1 - \mu_2 \in R^2 : (\bar{d} - \mu_1 + \mu_2)' \mathbf{S}_d^{-1} (\bar{d} - \mu_1 + \mu_2) \leq \frac{(n-1)p}{(n-p)n} F_{p,n-p}(0.05) = \frac{20}{99} F_{2,9}(0.05) = 0.86\right\}$$

which is an ellipse centered at $\bar{d}$. Note that the origin ($\mu_1 - \mu_2 = 0$) of R^2 is not in this ellipse, since

$$\bar{d}' \mathbf{S}_d^{-1} \bar{d} = 1.23 > 0.86.$$

From this fact, we conclude, at significance level 5% (with 5% chance of being mistaken), that the two lab analysis procedures are different. This is formally addressed in the framework of hypothesis testing in the following.

(2). Consider test of hypothesis

$$\begin{cases} H_0 : & \mu_1 = \mu_2 \\ H_a : & \mu_1 \neq \mu 2 \end{cases}$$

The p-value is

$$P(F_{2,9} > \frac{11 \times 9}{10 \times 2} \bar{d}' \mathbf{S}_d^{-1} \bar{d} = 6.12) = 0.0209 < 0.05$$

At significant level 5%, we reject the null hypothesis and conclude the two lab analysis procedures are different.

(3). T^2 simultaneous confidence intervals at 95% confidence level:

$$\text{for } \mu_{11} - \mu_{21}: \qquad \bar{d}_1 \pm c\sqrt{s_{d,11}/n} = 9.36 \pm \sqrt{9.47}\sqrt{199.26/11} = (-22.46,\ 3.74)$$

$$\text{for } \mu_{12} - \mu_{22}: \qquad \bar{d}_2 \pm c\sqrt{s_{d,22}/n} = 13.27 \pm \sqrt{9.47}\sqrt{418.61/11} = (-5.71,\ 32.25)$$

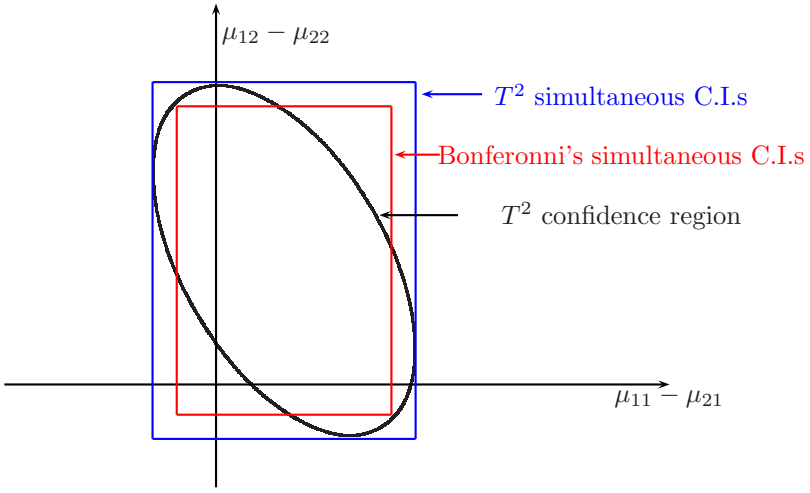
where $\mu_{11} - \mu_{21}$ is the first component of $\mu_1 - \mu_2$, standing for BOD, and $\mu_{11} - \mu_{21}$ is the second component of $\mu_1 - \mu_2$, standing for SS; and $c = \sqrt{(n-1)p/(n-p)F_{p,n-p}(.05)} = \sqrt{9.47} = 3.077$.

(4). Bonferroni's simultaneous confidence intervals at 95% confidence level:

$$\text{for } \mu_{11} - \mu_{21}: \qquad \bar{d}_1 \pm t_{n-1}(\frac{0.05}{2 \times 2})\sqrt{s_{d,11}/n} = -9.36 \pm 2.634\sqrt{199.26/11} = (-20.57\ 1.85)$$

$$\text{for } \mu_{12} - \mu_{22}: \qquad \bar{d}_2 \pm t_{n-1}(\frac{0.05}{2 \times 2})\sqrt{s_{d,22}/n} = 13.27 \pm 2.634\sqrt{418.61/11} = (-3.51,\ 29.51)$$

It seems to be contradicting that both T^2 and Bonferroni's simultaneous confidence intervals contain the origin of R^2, while the T^2 confidence region does not. This can be explained by the fact that the actual levels of T^2 and Bonferroni's simultaneous confidence intervals are both larger than the nominal level 95%, and the T^2 confidence region's actual confidence level is the same as the nominal level 95%. The following picture is an *ad hoc* illustration.



In this graph, the origin is in the T^2 and Bonferroni's simultaneous confidence intervals but not in the T^2 confidence region.

6.2 Repeated measurements of one variable.

Example 6.2 Analysis of Sleeping Dog Data In a clinical trial to test the anesthetizing effects of CO_2 pressure and halothane with different combinations. There are totally 19 dogs, each taking all four treatments with sleeping time recorded. Treatments 1, 2, 3 and 4 represent, respectively, treatments of high CO_2 without halothane, low CO_2 without halothane, high CO_2 with halothane and low CO_2 with halothane. The data are presented in the following table:

Sleeping Dog Data:				
	Treatment			
Dog	1	2	3	4
1	426	609	556	600
2	253	236	392	395
3	359	433	349	357
⋮	⋮	⋮	⋮	⋮
18	420	395	508	521
19	397	556	645	625

This is a typical example of a class of statistical problems called "repeated measurements".

Let

$$X_1, ..., X_n \; iid \sim MN(\mu, \Sigma) \qquad \text{with } \mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_p \end{pmatrix} \text{ and } \Sigma = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1p} \\ \vdots & \vdots & \vdots \\ \sigma_{p1} & \cdots & \sigma_{pp} \end{pmatrix}.$$

We are interested in comparison of $\mu_1, ..., \mu_p$. The data $X_1, ..., X_p$ are obviously *one sample*. The difference between "repeated measurements" and the standard one sample problem is that, in the former we are interested in comparison $\mu_1, ..., \mu_p$, the components of the mean vector μ, while in the

later we are interested in the inference with μ. Most commonly, in the "repeated measurements" problem, we care whether $\mu_1, ..., \mu_p$ are the same or not. This is the statistical hypothesis:

$$\begin{cases} H_0: & \mu_1 = \cdots = \mu_p \\ H_a: & \text{otherwise} \end{cases}$$

Similar to treating the paired comparison problem, we construct the differences. Suppose the data set is

$$\mathbf{X}_{n\times p} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \vdots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix}.$$

Define

$$\mathbf{C}_{(p-1)\times p} = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 \\ -1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -1 & 0 & 0 & \cdots & 1 \end{pmatrix}.$$

$\mathbf{C}$ is one of typical *contrast matrices*, those matrices with sum of rows being 0. Clearly

$$\mathbf{C}X_1, ..., \mathbf{C}X_n \; iid \sim MN(\mathbf{C}\mu, \mathbf{C}\Sigma\mathbf{C}') \qquad \text{with } \mathbf{C}\mu = \begin{pmatrix} \mu_2 - \mu_1 \\ \vdots \\ \mu_p - \mu_1 \end{pmatrix}_{(p-1)\times p}$$

The sample mean and sample variance of $\mathbf{C}X_1, ..., \mathbf{C}X_n$, of $p-1$ dimension, are $\mathbf{C}\bar{X}$ and $\mathbf{CSC}'$. It follows that

$$n(\mathbf{C}\bar{X} - \mathbf{C}\mu)'(\mathbf{CSC}')^{-1}(\mathbf{C}\bar{X} - \mathbf{C}\mu) \sim \frac{(n-1)(p-1)}{n-(p-1)} F_{p-1,n-(p-1)}.$$

Translating the hypothesis

$$\begin{cases} H_0: & \mu_1 = \cdots = \mu_p \\ H_a: & \text{otherwise} \end{cases}$$

into

$$\begin{cases} H_0: & \mathbf{C}\mu = 0 \\ H_a: & \text{otherwise} \end{cases}$$

Then, T^2 test of this hypothesis can be carried out with "one sample": $\mathbf{C}X_1, ..., \mathbf{C}X_n$. The p-value is

$$P\Big(F_{p-1,n-(p-1)} > \frac{n(n-p+1)}{(n-1)(p-1)} \times \text{"the observed value of } (\mathbf{C}\bar{X})'(\mathbf{CSC}')^{-1}\mathbf{C}\bar{X}\text{"} \Big).$$

Simultaneous inferences, either by the T^2 method or by the Bonferroni's method can also be carried out analogously, with "one sample": $\mathbf{C}X_1, ..., \mathbf{C}X_n$.

6.3 Mean comparison for two populations (Two sample problem).

(i). Univariate two sample problem — a review.

Suppose $X_{1,1}, ..., X_{1,n_1}$ are iid from population 1: $\sim N(\mu_1, \sigma_1^2)$, and $X_{2,1}, ..., X_{2,n_2}$ are iid from population 2: $\sim N(\mu_2, \sigma_2^2)$. The "one sample" $\{X_{1,1}, ..., X_{1,n_1}\}$ is independent of the other "one sample" $\{X_{2,1}, ..., X_{2,n_2}\}$. All random variables or observations here are of 1 dimension. We are interested in comparing the two population means μ_1 and μ_2. Let $\bar{X}_1$ and s_1^2 be the sample mean and sample variance for the first sample $X_{1,1}, ..., X_{1,n_1}$, and likewise $\bar{X}_2$ and s_2^2 for the second

sample $X_{2,1}, ..., X_{2,n_2}$. The primary question is whether they are the same or not. The key facts here are

$$\frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \sim N(0,1);$$
$$\frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{s_{\text{pooled}}\sqrt{1/n_1 + 1/n_2}} \sim t_{n_1+n_2-2} \qquad \text{if } \sigma_1^2 = \sigma_2^2$$

where

$$s^2_{\text{pooled}} = \frac{1}{n_1 + n_2 - 2}[\sum_{i=1}^{n_1}(X_{1,i} - \bar{X}_1)^2 + \sum_{i=1}^{n_2}(X_{2,i} - \bar{X}_2)^2]$$

is the so-called pooled estimator of $\sigma_1^2 = \sigma_2^2$.

Case 1: $\sigma_1^2 = \sigma_2^2$. Confidence intervals for $\mu_1 - \mu_2$ is

$$\bar{X}_1 - \bar{X}_2 \pm t_{n_1+n_2-2}(\alpha/2)s_p\sqrt{1/n_1 + 1/n_2}.$$

Tests for hypothesis $H_0 : \mu_1 = \mu_2$ can be likewise carried out using t-method.

Case 2: $\sigma_1^2 \neq \sigma_2^2$. A (conservative) confidence interval for $\mu_1 - \mu_2$ at norminal confidence level:

$$\bar{X}_1 - \bar{X}_2 \pm t_{k-2}(\alpha/2)\sqrt{s_1^2/n_1 + s_2^2/n_2}$$

where $k = \min(n_1, n_2)$

(ii). Multivariate two sample problem.

The following two examples are illustration of the setup of multivariate two sample problem.

Example 6.3 CARAPACE MEASUREMENTS FOR PAINTED TURTLES *Exercise 6.18 of the textbook* Painted turtles are a kind of water turtles living in North America. Some researchers wish to compare the sizes of male and female painted turtles by comparing the sizes of their carapaces (shells). The carapace is measured in length, width and height. Totally $n_1 = 24$ of females and $n_2 = 24$ of males are measured and the data is illustrated in the following (See Table 6.9 of the textbook for complete data):

Carapace measurements for painted turtles:					
Female			Male		
Length	Width	Height	Length	Width	Height
98	81	38	93	74	37
103	84	38	94	78	35
103	86	42	96	80	35
⋮	⋮	⋮	⋮	⋮	⋮
162	124	61	131	95	46
177	132	67	135	106	47

Note that the females are males are not paired with each other. In fact, they are all unrelated with each other. This is an example of two sample problem.

Example 6.4 ANACONDA DATA *(Exercise 6.39 of the textbook)* Anacondas are some of the largest snakes in the world. Some researchers capture the snakes and measure their snout vent length (cm) and weight (kg). The data contain $n_1 = 28$ female and $n_2 = 28$ male snakes are illustrated in the following (See Table 6.19 in the textbook for complete data):

Anaconda Data:			
Female		Male	
Length	Width	Length	Width
271.0	18.50	176.7	3.00
477.0	82.50	259.5	9.75
306.3	23.40	258.0	10.07
⋮	⋮	⋮	⋮
438.6	57.00	236.7	6.49
377.1	61.50	235.3	6.00

Note that "Length" refers to snout vent length, not body length. The females and the males, as in the last example, are not related with each other. This is an example of two-sample problem.

The setup of multivariate two sample problem is the same as that of univariate two sample problem, except for the dimensionality. Suppose $X_{1,1}, ..., X_{1,n_1}$ are iid from population 1: $\sim MN(\mu_1, \Sigma_1)$, and $X_{2,1}, ..., X_{2,n_2}$ are iid from population 2: $\sim MN(\mu_2, \Sigma_2)$. The "one sample" $\{X_{1,1}, ..., X_{1,n_1}\}$ is independent of the other "one sample" $\{X_{2,1}, ..., X_{2,n_2}\}$. All of them are of p-dimension. We are interested in the difference of the two population means μ_1 and μ_2, especially whether they are equal or not. Let $\bar{X}_1$ and $\mathbf{S}_1$ be the sample mean and sample variance for the first sample $X_{1,1}, ..., X_{1,n_1}$, and likewise $\bar{X}_2$ and $\mathbf{S}_2$ for the second sample $X_{2,1}, ..., X_{2,n_2}$.

Case 1. $\Sigma_1 = \Sigma_2$.

$$\frac{n_1 n_2}{n_1+n_2}(\bar{X}_1 - \bar{X}_2 - (\mu_1-\mu_2))'\mathbf{S}_{\text{pooled}}^{-1}(\bar{X}_1 - \bar{X}_2 - (\mu_1-\mu_2)) \sim \frac{(n_1+n_2-2)p}{n_1+n_2-p-1}F_{p,n_1+n_2-p-1}$$

where

$$\begin{aligned}\mathbf{S}_{\text{pooled}} &= \frac{1}{n_1+n_2-2}[\sum_{i=1}^{n_1}(X_{1,i}-\bar{X}_1)(X_{1,i}-\bar{X}_1)' + \sum_{i=1}^{n_2}(X_{2,i}-\bar{X}_2)(X_{2,i}-\bar{X}_2)'] \\ &= \frac{1}{n_1+n_2-2}[(n_1-1)\mathbf{S}_1 + (n_2-1)\mathbf{S}_2]\end{aligned}$$

is the so-called pooled estimator of $\Sigma_1 = \Sigma_2$. This ensures, for example, the T^2 confidence region for $\mu_1 - \mu_2$ at confidence level $1-\alpha$ as:

$$\Big\{\mu_1-\mu_2 \in R^p : \frac{n_1 n_2}{n_1+n_2}(\bar{X}_1 - \bar{X}_2 - (\mu_1-\mu_2))'\mathbf{S}_{\text{pooled}}^{-1}(\bar{X}_1 - \bar{X}_2 - (\mu_1-\mu_2)) \\ \le \frac{(n_1+n_2-2)p}{n_1+n_2-p-1}F_{p,n_1+n_2-p-1}(\alpha)\Big\},$$

which is an ellipse in R^p centered at $\bar{X}_1 - \bar{X}_2$. Simultaneous confidence intervals can also be constructed. We omit the details.

Case 2: $\Sigma_1 \neq \Sigma_2$. For large n_1 and n_2,

$$(\bar{X}_1 - \bar{X}_2 - (\mu_1-\mu_2))'(\frac{1}{n_1}\mathbf{S}_1 + \frac{1}{n_2}\mathbf{S}_2)^{-1}(\bar{X}_1 - \bar{X}_2 - (\mu_1-\mu_2)) \sim \chi_p^2 \qquad \text{approximately.}$$

This approximation provides theoretical ground for inference approaches to constructing confidence regions or simultaneous confidence intervals. Details are omitted.

Remark. Note that Bonferroni's method of simultaneous inferences does not rely on F-distribution or the approximate χ^2 distribution of the relevant statistics, but rather on the t distribution of relevant *univariate* statistics.

6.4 Comparing several multivariate population means. (MANOVA)

(i). ANOVA for univariate mean comparison — a review.

The random variables or their observations can be presented as follows:

$$\text{all independent} \begin{cases} \text{Group } 1: & X_{11}, ..., X_{1n_1} \sim N(\mu_1, \sigma^2) \\ \text{Group } 2: & X_{21}, ..., X_{2n_2} \sim N(\mu_2, \sigma^2) \\ \vdots & \vdots \\ \text{Group } g: & X_{g1}, ..., X_{gn_g} \sim N(\mu_g, \sigma^2) \end{cases}$$

And we are interested in whether the population means of these g groups, $\mu_1, ..., \mu_g$, are same or not. Note that there are n_i observations from the i-th group And the population variances σ^2 are assumed all same.

Let $n = n_1 + \cdots + n_k$ be the total sample size. Set $\bar{X}_k = (1/n_k)\sum_{j=1}^{n_k} X_{kj}$ as the sample mean for the k-th group, and $\bar{X} = (1/n)\sum_{k=1}^{g}\sum_{j=1}^{n_k} X_{kj} = \sum_{k=1}^{g}(n_k/n)\bar{X}_k$ as the total sample mean.

Consider hypothesis:

$$\begin{cases} H_0: & \mu_1 = \mu_2 = \cdots = \mu_g \\ H_a & \text{otherwise.} \end{cases}$$

Variance decomposition:

SS_{Total}	$=$	$SS_{Between}$	$+$	SS_{Within}
$\sum_{i=1}^{g}\sum_{j=1}^{n_i}(X_{ij} - \bar{X})^2$	$=$	$\sum_{i=1}^{g} n_i(\bar{X}_i - \bar{X})^2$	$+$	$\sum_{i=1}^{g}\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_i)^2$
Total variation	$=$	Variation between groups	$+$	Variation within groups

The following ANOVA Table is formed based on the above variance decomposition.

Source of variation	Sum of squares	Degree of freedom	Mean Squares	F-statistic
"Between"	$SS_{Between}$	$g-1$	$SS_{Between}/(g-1)$	$F = \frac{SS_{Between}/(g-1)}{SS_{Within}/(n-g)}$
"Within"	SS_{Within}	$n-g$	$SS_{Within}/(n-g)$	
"Total"	SS_{Total}	$n-1$		

And the p-value is

$$P(F_{g-1,n-g} > \text{"the observed value of the } F\text{-statistic"})$$

This is because, under H_0, the F-statistic follows $F_{g-1,n-g}$ distribution.

(ii). MANOVA for multivariate mean comparison.

Example 6.5. Analysis of Wisconsin nursing home data *(Example 6.10 in the text book)* We wish to investigate whether private, nonprofit or government sponsored nursing homes are different in terms of their costs. For variables are chosen to measure their costs per-patient-day: 1. cost of nursing labor; 2. cost of dietary labor; 3. cost of plant operation and maintenance labor; and 4. cost of housekeeping and laundry labor. The four variables are observed for each of 271 private, 138 nonprofit and 107 government sponsored nursing homes. The sample mean and sample variances are

$$\text{private nursing homes:} \quad \bar{X}_1 = \begin{pmatrix} 2.066 \\ .480 \\ .082 \\ 0.360 \end{pmatrix} \quad \mathbf{S}_1 = \begin{pmatrix} .291 & & & \\ -.001 & .001 & & \\ .002 & .000 & .001 & \\ .010 & .003 & .000 & .010 \end{pmatrix}$$

$$\text{nonprofit nursing homes:} \quad \bar{X}_2 = \begin{pmatrix} 2.167 \\ 0.596 \\ .124 \\ .418 \end{pmatrix} \quad \mathbf{S}_2 = \begin{pmatrix} .561 & & & \\ .011 & .025 & & \\ .001 & .004 & .005 & \\ .037 & .007 & .002 & .019 \end{pmatrix}$$

$$\text{government nursing homes:} \quad \bar{X}_3 = \begin{pmatrix} 2.273 \\ .521 \\ .125 \\ .383 \end{pmatrix} \quad \mathbf{S}_3 = \begin{pmatrix} .261 & & & \\ .030 & .017 & & \\ .003 & -.000 & .004 & \\ .018 & .006 & .001 & .013 \end{pmatrix}$$

We shall answer the question after introducing the general setup and methodology of multivariate analysis of variance, so-called MANOVA.

The setup for comparison of several multivariate means is analogous to that of univariate case:

$$\text{all independent} \begin{cases} \text{Group } 1: & X_{11}, ..., X_{1n_1} \sim N(\mu_1, \Sigma) \\ \text{Group } 2: & X_{21}, ..., X_{2n_2} \sim N(\mu_2, \Sigma) \\ \vdots & \vdots \\ \text{Group } g: & X_{g1}, ..., X_{gn_g} \sim N(\mu_g, \Sigma) \end{cases}$$

Now X_{ij} and μ_k are p-vectors and Σ is a $p \times p$ matrix. And we are interested in whether the population means of these g groups, $\mu_1, ..., \mu_g$, are same or not. Again the population variances σ^2 are assumed all same.

There are n_i observations from the i-th group. Let $n = n_1 + \cdots + n_k$ be the total sample size. Set $\bar{X}_k = (1/n_k) \sum_{j=1}^{n_k} X_{kj}$ as the sample mean for the k-th group, and $\bar{X} = (1/n) \sum_{k=1}^{g} \sum_{j=1}^{n_k} X_{kj} = \sum_{k=1}^{g} (n_k/n) \bar{X}_k$ as the total sample mean.

Consider hypothesis:

$$\begin{cases} H_0: & \mu_1 = \mu_2 = \cdots = \mu_g \\ H_a & \text{otherwise.} \end{cases}$$

Variance decomposition:

$$\begin{array}{ccccccc} SS_{Total} & = & SS_{Between} & + & SS_{Within} \\ \sum_{i=1}^{g} \sum_{j=1}^{n_i} (X_{ij} - \bar{X})^{\otimes 2} & = & \sum_{i=1}^{g} n_i (\bar{X}_i - \bar{X})^{\otimes 2} & + & \sum_{i=1}^{g} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^{\otimes 2} \end{array}$$

where $\otimes 2$ is the outer product of a p-vector, i.e., for any p-vector a, $a^{\otimes 2} = aa'$ which is a $p \times p$ matrix. Note that $SS_{Total}, SS_{Between}$ and SS_{Within} are all $p \times p$ matrices.

With the above variance decomposition, one can construct a MANOVA Table:

Source of variation	Sum of squares	Degree of freedom	Λ^*
"Between" (Treatment)	$\mathbf{B} \equiv SS_{Between}$	$g - 1$	$\frlap{\frac{\lvert\mathbf{W}\rvert}{\lvert\mathbf{W}+\mathbf{B}\rvert}}$
"Within" (Error)	$\mathbf{W} \equiv SS_{Within}$	$n - g$	
"Total"	$\mathbf{B} + \mathbf{W} = SS_{Total}$	$n - 1$	

For the hypothesis testing, we first notice that large values of Λ^* indicates evidence of H_0 being true. If g or p is small the distribution of Λ^* under H_0 is known; see Table 6.3 in the textbook. If n is large, an approximation is, under H_0,

$$-(n - 1 - \frac{p+g}{2}) \log \Lambda^* \sim \chi^2_{p(g-1)} \qquad \text{approximately.}$$

With this approximation, the p-value is

$$P\Big(\chi^2_{p(g-1)} > -(n - 1 - \frac{p+g}{2}) \log(\text{"the observed value of } \Lambda^*\text{"})\Big).$$

At significance level α, we reject H_0 when the p-value is smaller than α, or, equivalently, when

$$-(n - 1 - \frac{p+g}{2}) \log(\text{"the observed value of } \Lambda^*\text{"}) > \chi^2_{p(g-1)}(\alpha).$$

For Example 6.5, $g = 3, p = 4, n_1 = 271, n_2 = 138, n_3 = 107$ and $n = 516$. And

$$\bar{X} = \frac{n_1 \bar{X}_1 + n_2 \bar{X}_2 + n_3 \bar{X}_3}{n} = \begin{pmatrix} 2.136 \\ .519 \\ .102 \\ .380 \end{pmatrix}$$

$$\mathbf{W} = (n_1-1)\mathbf{S}_1+(n_2-1)\mathbf{S}_2+(n_3-1)\mathbf{S}_3 = \begin{pmatrix} 182.962 & & & \\ 4.408 & 8.200 & & \\ 1.695 & .633 & 1.484 & \\ 9.591 & 2/418 & .394 & 6.538 \end{pmatrix}$$

$$\text{and } \mathbf{B} = \sum_{k=1}^{3} n_k(\bar{X}_k-\bar{X})(\bar{X}_k-\bar{X})' = \begin{pmatrix} 3.475 & & & \\ 1.111 & 1.225 & & \\ .821 & .453 & .235 & \\ .584 & .610 & .230 & .304 \end{pmatrix}$$

Then, the MANOVA Table is

Source of variation	Sum of squares	Degree of freedom	Λ^*
"Between"(Treatment)	$\mathbf{B} \equiv SS_{Between}$	$g-1=2$	$\frac{\lvert\mathbf{W}\rvert}{\lvert\mathbf{W}+\mathbf{B}\rvert} = 0.7714$
"Within"(Error)	$\mathbf{W} \equiv SS_{Within}$	$n-g=513$	
"Total"	$\mathbf{B}+\mathbf{W} = SS_{Total}$	$n-1=515$	

where $\mathbf{W}$ and $\mathbf{B}$ are as given above. We try two approach for test of the hypothesis

$$\begin{cases} H_0: & \mu_1=\mu_2=\mu_3 \\ H_a: & \text{otherwise} \end{cases}$$

Approach 1: Since $g=3$ is small, from Table 6.2 in the text book, we know, under H_0,

$$\frac{n-p-2}{p}\cdot\frac{1-\sqrt{\Lambda^*}}{\sqrt{\Lambda^*}} \sim F_{2p,2(n-p-2)}.$$

At significance level α, H_0 is rejected when

$$\frac{n-p-2}{p}\cdot\frac{1-\sqrt{\Lambda^*}}{\sqrt{\Lambda^*}} > F_{2p,2(n-p-2)}(\alpha).$$

In this example, the left hand side is

$$\frac{516-4-2}{4}\cdot\frac{1-\sqrt{0.7714}}{\sqrt{0.7714}} = 17.67$$

Notice that $2p=8$, $2(n-p-2)=1020$. Using any statistical software, one can check that the p-value is

$$P(F_{8,1020} > 17.67) \approx 0.0000$$

The p-value is so small, that one is certain to reject H_0 at any sensible significance level. (If using the R, one can check that $F_{8,1020}(0.00001) = 4.738 < 17.4$, so we should reject H_0 even at significance level 0.00001.)

Approach 2. Since n is large, we use the fact

$$-(n-1-\frac{p+g}{2})\log\Lambda^* \sim \chi^2_{p(g-1)} \qquad \text{approximately.}$$

With some calculation, the left hand is

$$-(516-1-7/2)\log(.7714) = -511.5\log(.7714) = 132.76$$

The approximate p-value is

$$P(\chi^2_8 > 132.76) \approx 0.00000$$

which is also extremely small. Thus, this approach also leads to rejection of H_0 at any sensible significance level. (If using the R, one can check that $\chi^2_8(0.00001) = 37.33 < 132.76$, so we should reject H_0 even at significance level 0.00001. You may notice that $\chi^2_8(0.00001) = 37.33 \approx 8\times 4.738 = 8\times F_{8,1020}(0.00001)$, why?)