Lecture 5 Comparisons of Two Multivariate Means

Shiwei Lan[1]

[1]School of Mathematical and Statistical Sciences
Arizona State University

STP533 Multivariate Analysis
Spring 2025

Table of Contents

Paired Comparisons

- Previously, we discussed the inference about a population mean vector. The ideas can be extended to handle problems involving the comparison of several mean vectors.
- We will review the comparison of multiple (univariate) means and then generalize it to the multivariate cases.
- Measurements are often recorded under different experimental conditions. We are interested in comparing their results (effects).
- Consider a single response (univariate) case. Denote X_{ij} as the response to treatment $j (= 1, 2)$ for the i-th trial.
- We are interested in comparing their difference $D_i = X_{i1} - X_{i2}$ for $i = 1, \cdots, n$. Denote the difference $\delta = \mu_1 - \mu_2$. We test

$$H_0 : \delta = 0, \quad H_1 : \delta \neq 0$$

Paired Comparisons

- Assume that $D_i \overset{iid}{\sim} N(\delta, \sigma_d^2)$. Then we construct the t-test statistic

$$t = \frac{\bar{D} - \delta}{s_d/\sqrt{n}}, \quad \bar{D} = \frac{1}{n}\sum_{i=1}^n D_i, \quad s_d^2 = \frac{1}{n-1}\sum_{i=1}^n (D_i - \bar{D})^2$$

- The hypothesis can be conducted by comparing $|t|$ with $t_{1-\alpha/2}(n-1)$ at the the significance level α.
- The $100(1-\alpha)\%$ confidence interval for the mean difference δ is

$$\bar{d} - t_{1-\alpha/2}(n-1)s_d/\sqrt{n} \leq \delta \leq \bar{d} + t_{1-\alpha/2}(n-1)s_d/\sqrt{n}$$

Multivariate Paired Comparisons

- Now we generalize such paired comparisons to the multivariate cases.
- Let X_{ijk} denote the response for the i-th trial, to treatment $j(=1,2)$, and for the k-th variable.
- Denote the p paired difference random variables

$$D_{ik} = X_{i1k} - X_{i2k}, \quad i = 1, \cdots, n, \quad k = 1, \cdots, p$$

- Let $\mathbf{D}_i = [D_{ik}]_{p\times 1}$. Assume $\mathrm{E}[\mathbf{D}_i] = \boldsymbol{\delta}$, $\mathrm{Cov}[\mathbf{D}_i] = \boldsymbol{\Sigma}_d$, and $\mathbf{D}_i \overset{iid}{\sim} N_p(\boldsymbol{\delta}, \boldsymbol{\Sigma}_d)$.
- We can construct T^2-statistic

$$T^2 = n(\bar{\mathbf{D}}-\boldsymbol{\delta})^T \mathbf{S}_d^{-1}(\bar{\mathbf{D}}-\boldsymbol{\delta}), \quad \bar{\mathbf{D}} = \frac{1}{n}\sum_{i=1}^n \mathbf{D}_i, \; \mathbf{S}_d = \frac{1}{n-1}\sum_{i=1}^n (\mathbf{D}_i-\bar{\mathbf{D}})(\mathbf{D}_i-\bar{\mathbf{D}})^T$$

- Then we have $T^2 \sim \frac{(n-1)p}{(n-p)} F(p, n-p)$. And $T^2 \dot{\sim} \chi^2(p)$ when both p and $n-p$ are large.

Testing Paired Comparisons

- Given the observed differences $\mathbf{d}_i \in \mathbb{R}^p$, $i = 1, \cdots, n$, α-level test $H_0 : \boldsymbol{\delta} = 0$ vs $H_1 : \boldsymbol{\delta} \neq 0$ for an $N_p(\boldsymbol{\delta}, \boldsymbol{\Sigma})$ population rejects H_0 if

$$T^2 = n\bar{\mathbf{d}}^T \mathbf{S}_d^{-1} \bar{\mathbf{d}} > \frac{(n-1)p}{(n-p)} F_{1-\alpha}(p, n-p)$$

- The $100(1-\alpha)\%$ confidence region (CR) for $\boldsymbol{\delta}$ can be determined by

$$(\bar{\mathbf{d}} - \boldsymbol{\delta})^T \mathbf{S}_d^{-1} (\bar{\mathbf{d}} - \boldsymbol{\delta}) \leq \frac{(n-1)p}{n(n-p)} F_{1-\alpha}(p, n-p)$$

- Also, $100(1-\alpha)\%$ simultaneous confidence intervals (SCI) for individual differences δ_i:

$$\bar{d}_i \pm \sqrt{\frac{(n-1)p}{(n-p)} F_{1-\alpha}(p, n-p) \frac{s_{d_i}}{n}}$$

- For $n-p$ large, $\frac{(n-1)p}{(n-p)} F(p, n-p) \doteq \chi^2(p)$ and normality need not be assumed.
- The Bonferroni $100(1-\alpha)\%$ SCI for δ_i:

$$\bar{d}_i \pm t_{1-\alpha/(2p)}(n-1) \frac{s_{d_i}}{n}$$

Multivariate Paired Comparisons

Example 6.1 (Checking for a mean difference with paired observations) 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 one 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 in Table 6.1.

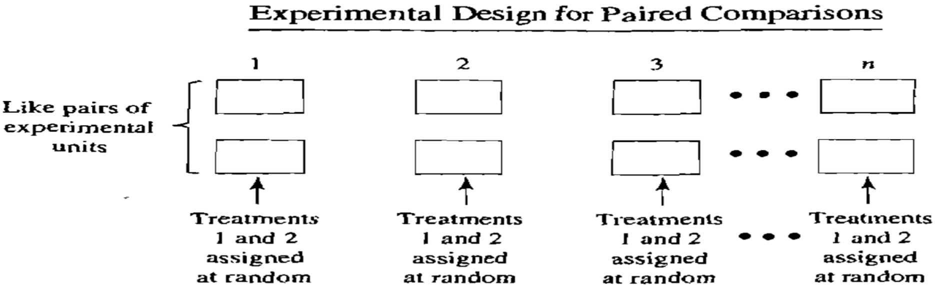
Table 6.1 Effluent Data

	Commercial lab		State lab of hygiene	
Sample j	x_{1j1} (BOD)	x_{1j2} (SS)	x_{2j1} (BOD)	x_{2j2} (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

Source: Data courtesy of S. Weber.

Experimental Design for Paired Comparisons

- Whenever an investigator can control the assignment of treatments to experimental units, an appropriate randomized assignment of treatments can enhance the statistical analysis.
- For paired differences, random assignment to treatments will help eliminate the systematic effects of uncontrolled sources of variation.
- Such randomization can be implemented by flipping a coin to determine between two treatments.



Experimental Design for Paired Comparisons

- Denote the full-sample quantities of mean and covariance as

$$\bar{\mathbf{x}} = [\bar{x}_{11}, \bar{x}_{12}, \cdots, \bar{x}_{1p}, \bar{x}_{21}, \bar{x}_{22}, \cdots, \bar{x}_{2p}]^T, \quad \mathbf{S}_{2p\times 2p} = \begin{bmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{bmatrix}$$

- Now define the *contrast matrix*

$$\mathbf{C}_{p\times 2p} = \begin{bmatrix} \mathbf{I}_p & -\mathbf{I}_p \end{bmatrix}$$

- Then we can verify that

$$\mathbf{d}_i = \mathbf{C}\mathbf{x}_i, \; i = 1, \cdots, n, \quad \bar{\mathbf{d}} = \mathbf{C}\bar{\mathbf{x}}, \; \mathbf{S}_d = \mathbf{C}\mathbf{S}\mathbf{C}^T$$

- Hence it is not necessary to calculate the differences:

$$T^2 = n\bar{\mathbf{x}}^T\mathbf{C}^T(\mathbf{C}\mathbf{S}\mathbf{C}^T)^{-1}\mathbf{C}\bar{\mathbf{x}}$$

A Repeated Measures Design for Comparing Treatments

- Another generalization of univariate paired t-statistic arises when we have q-treatments.
- Let X_{ijk} denote the response for the i-th trial, to treatment $j(=1,2,\cdots,q)$, and for the k-th variable.
- Consider a single response, i.e. $p=1$. Each subject (a.k.a. experiment unit) i receives *repeated measures* for all q treatments. Denote i-th observations for the repeated measurements as $\mathbf{X}_i = [X_{ij}]_{q\times 1}$ and $\mathrm{E}[\mathbf{X}_i] = \boldsymbol{\mu}$.
- We consider contrasts $\mathbf{C}_1\boldsymbol{\mu}$ or $\mathbf{C}_2\boldsymbol{\mu}$ with

$$\mathbf{C}_1 = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 \\ 1 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \vdots & -1 \end{bmatrix}, \quad \textit{or} \quad \mathbf{C}_2 = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \vdots & -1 & 1 \end{bmatrix}$$

Test for Equality of Treatments

- Consider more general *contrast matrix* (each row sums to 0) matrix $\mathbf{C}$. We are interested in testing $H_0 : \mathbf{C}\boldsymbol{\mu} = \mathbf{0}$ vs $H_1 : \mathbf{C}\boldsymbol{\mu} \neq \mathbf{0}$.
- Assume $\mathbf{X}_i \overset{iid}{\sim} N_q(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then we reject H_0 at the significance level α if

$$T^t = n(\mathbf{C}\bar{\mathbf{x}})^T(\mathbf{CSC}^T)^{-1}\mathbf{C}\bar{\mathbf{x}} > \frac{(n-1)(q-1)}{(n-q+1)}F_{1-\alpha}(q-1, n-q+1)$$

where $\bar{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i$ and $\mathbf{S} = \frac{1}{n-1}\sum_{i=1}^n(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T$.
- Does T^2 depend on the choice of $\mathbf{C}$?
- The $100(1-\alpha)\%$ CR for contrasts $\mathbf{C}\boldsymbol{\mu}$ with normal population mean $\boldsymbol{\mu}$ is

$$n(\mathbf{C}\bar{\mathbf{x}} - \mathbf{C}\boldsymbol{\mu})^T(\mathbf{CSC}^T)^{-1}(\mathbf{C}\bar{\mathbf{x}} - \mathbf{C}\boldsymbol{\mu}) \leq \frac{(n-1)(q-1)}{(n-q+1)}F_{1-\alpha}(q-1, n-q+1)$$

- The $100(1-\alpha)\%$ SCI for single contrasts $\mathbf{c}^T\boldsymbol{\mu}$ for any contrast vector

$$\mathbf{c}^t\boldsymbol{\mu} : \mathbf{c}^T\bar{\mathbf{x}} \pm \sqrt{\frac{(n-1)(q-1)}{(n-q+1)}F_{1-\alpha}(q-1, n-q+1)}\sqrt{\mathbf{c}^T\mathbf{S}\mathbf{c}/n}$$

Test for Equality of Treatments

Example 6.2 (Testing for equal treatments in a repeated measures design) Improved anesthetics are often developed by first studying their effects on animals. In one study, 19 dogs were initially given the drug pentobarbitol. Each dog was then administered carbon dioxide CO_2 at each of two pressure levels. Next, halothane (H) was added, and the administration of CO_2 was repeated. The response, milliseconds between heartbeats, was measured for the four treatment combinations:

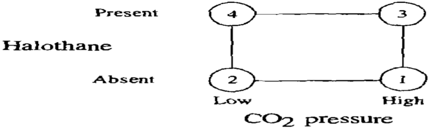


Table 6.2 contains the four measurements for each of the 19 dogs, where

Treatment 1 = high CO_2 pressure without H
Treatment 2 = low CO_2 pressure without H
Treatment 3 = high CO_2 pressure with H
Treatment 4 = low CO_2 pressure with H

We shall analyze the anesthetizing effects of CO_2 pressure and halothane from this repeated-measures design.

There are three treatment contrasts that might be of interest in the experiment. Let μ_1, μ_2, μ_3, and μ_4 correspond to the mean responses for treatments 1, 2, 3, and 4, respectively. Then

$$(\mu_3 + \mu_4) - (\mu_1 + \mu_2) = \begin{pmatrix} \text{Halothane contrast representing the} \\ \text{difference between the presence and} \\ \text{absence of halothane} \end{pmatrix}$$

$$(\mu_1 + \mu_3) - (\mu_2 + \mu_4) = \begin{pmatrix} CO_2 \text{ contrast representing the difference} \\ \text{between high and low } CO_2 \text{ pressure} \end{pmatrix}$$

$$(\mu_1 + \mu_4) - (\mu_2 + \mu_3) = \begin{pmatrix} \text{Contrast representing the influence} \\ \text{of halothane on } CO_2 \text{ pressure differences} \\ (\text{H}-CO_2 \text{ pressure “interaction”}) \end{pmatrix}$$

Table of Contents

Comparing Mean Vectors from Two Populations

- A T^2-statistic for testing the equality of vector means from two multivariate populations can be developed by analogy with the univariate procedure.
- The two populations may correspond to different sets of experimental settings independent from each other, as in paired comparison.
- The inferences in the two-population case are applicable to a more general collection of experimental units where homogeneity is not required.
- Consider a random sample of size n_1 from population 1 and a sample of size n_2 from population 2. We have the following notations.

Sample	Summary statistics	
(Population 1) $\mathbf{x}_{11}, \mathbf{x}_{12}, \ldots, \mathbf{x}_{1n_1}$	$\bar{\mathbf{x}}_1 = \frac{1}{n_1}\sum_{j=1}^{n_1} \mathbf{x}_{1j}$	$\mathbf{S}_1 = \frac{1}{n_1 - 1}\sum_{j=1}^{n_1} (\mathbf{x}_{1j} - \bar{\mathbf{x}}_1)(\mathbf{x}_{1j} - \bar{\mathbf{x}}_1)'$
(Population 2) $\mathbf{x}_{21}, \mathbf{x}_{22}, \ldots, \mathbf{x}_{2n_2}$	$\bar{\mathbf{x}}_2 = \frac{1}{n_2}\sum_{j=1}^{n_2} \mathbf{x}_{2j}$	$\mathbf{S}_2 = \frac{1}{n_2 - 1}\sum_{j=1}^{n_2} (\mathbf{x}_{2j} - \bar{\mathbf{x}}_2)(\mathbf{x}_{2j} - \bar{\mathbf{x}}_2)'$

Common Covariance When n_1 and n_2 Are Small

- We the following assumptions for the two samples
  1. $\mathbf{X}_{1j} \overset{iid}{\sim} (\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$,
  2. $\mathbf{X}_{2j} \overset{iid}{\sim} (\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$.
  3. $\{\mathbf{X}_{1j}\} \perp \{\mathbf{X}_{2j'}\}$.
- When both n_1 and n_2, we further assume both population are normal with a common variance, i.e. $\mathbf{X}_{1j} \overset{iid}{\sim} N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$ and $\mathbf{X}_{2j} \overset{iid}{\sim} N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$.
- Then we consider pooling the data to estimate the common covariance $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$ with

$$\mathbf{S}_p = \frac{(n_1 - 1)\mathbf{S}_1 + (n_2 - 1)\mathbf{S}_2}{n_1 + n_2 - 2}$$

where $\mathbf{S}_i = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)^T$ for $i = 1, 2$.

Test Difference of Mean Vectors

- We want to test $H_0 : \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \boldsymbol{\delta}_0$ vs $H_1 : \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 \neq \boldsymbol{\delta}_0$.
- The likelihood ratio test is based on the square of the statistical distance T^2:

$$T^2 = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2 - \delta_0)^T \left[\left(\frac{1}{n_1} + \frac{1}{n_2} \right) \mathbf{S}_p \right]^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2 - \delta_0).$$

- Under normality assumption, one can show that

$$T^2 \sim \frac{(n_1 + n_2 - 2)p}{n_1 + n_2 - p - 1} F(p, n_1 + n_2 - p - 1).$$

- It suffices to show that $(n_1 + n_2 - 2)\mathbf{S}_p \sim W_{n_1+n_2-2}(\boldsymbol{\Sigma})$.

Example 6.3 (Constructing a confidence region for the difference of two mean vectors) Fifty bars of soap are manufactured in each of two ways. Two characteristics, X_1 = lather and X_2 = mildness, are measured. The summary statistics for bars produced by methods 1 and 2 are

$$\bar{\mathbf{x}}_1 = \begin{bmatrix} 8.3 \\ 4.1 \end{bmatrix}, \qquad \mathbf{S}_1 = \begin{bmatrix} 2 & 1 \\ 1 & 6 \end{bmatrix}$$

$$\bar{\mathbf{x}}_2 = \begin{bmatrix} 10.2 \\ 3.9 \end{bmatrix}, \qquad \mathbf{S}_2 = \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix}$$

Obtain a 95% confidence region for $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$.

Simultaneous Confidence Intervals

- The SCI for $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ can be derived from a consideration of all possible linear combinations of the mean difference vector.
- Let $c^2 = \frac{(n_1+n_2-2)p}{n_1+n_2-p-1} F_{1-\alpha}(p, n_1+n_2-p-1)$. With probability $1-\alpha$

$$\mathbf{a}^T(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2) \pm c\sqrt{\mathbf{a}^T\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\mathbf{S}_p\mathbf{a}}$$

  will cover $\mathbf{a}^T(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$ for all $\mathbf{a}$.
- In particular, SCI for $\mu_{1k} - \mu_{2k}$ is

$$(\bar{X}_{1k} - \bar{X}_{2k}) \pm c\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)s_{p,kk}}, \quad k = 1, \cdots, p.$$

Example 6.4 (Calculating simultaneous confidence intervals for the differences in mean components) Samples of sizes $n_1 = 45$ and $n_2 = 55$ were taken of Wisconsin homeowners with and without air conditioning, respectively. (Data courtesy of Statistical Laboratory, University of Wisconsin.) Two measurements of electrical usage (in kilowatt hours) were considered. The first is a measure of total *on*-peak consumption (X_1) during July, and the second is a measure of total *off*-peak consumption (X_2) during July. The resulting summary statistics are

$$\bar{\mathbf{x}}_1 = \begin{bmatrix} 204.4 \\ 556.6 \end{bmatrix}, \quad \mathbf{S}_1 = \begin{bmatrix} 13825.3 & 23823.4 \\ 23823.4 & 73107.4 \end{bmatrix}, \quad n_1 = 45$$

$$\bar{\mathbf{x}}_2 = \begin{bmatrix} 130.0 \\ 355.0 \end{bmatrix}, \quad \mathbf{S}_2 = \begin{bmatrix} 8632.0 & 19616.7 \\ 19616.7 & 55964.5 \end{bmatrix}, \quad n_2 = 55$$

Two-Sample with Unequal Covariances

- When $\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$, the previous argument based on T^2 fails.
- Bartlett's test can be used to test $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$. However the result can be misleading for non-normal populations.
- In practice, when both n_1 and n_2 are large, we can consider the approximate $100(1-\alpha)\%$ CR for $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ with unequal covariances.

$$[\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)]^T \left[\frac{1}{n_1}\mathbf{S}_1 + \frac{1}{n_2}\mathbf{S}_2\right]^{-1} [\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)] \le \chi^2_{1-\alpha}(p)$$

- The approximate $100(1-\alpha)\%$ SCI for all linear $\mathbf{a}^T(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$ is

$$\mathbf{a}^T(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) : \mathbf{a}^T(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) \pm \sqrt{\chi^2_{1-\alpha}(p)} \sqrt{\mathbf{a}^T \left(\frac{1}{n_1}\mathbf{S}_1 + \frac{1}{n_2}\mathbf{S}_2\right) \mathbf{a}}.$$

Approximate SCIs of Mean Difference

Lecture 5

S.Lan

Example 6.5 (Large sample procedures for inferences about the difference in means) We shall analyze the electrical-consumption data discussed in Example 6.4 using the large sample approach. We first calculate

$$\frac{1}{n_1}\mathbf{S}_1 + \frac{1}{n_2}\mathbf{S}_2 = \frac{1}{45}\begin{bmatrix} 13825.3 & 23823.4 \\ 23823.4 & 73107.4 \end{bmatrix} + \frac{1}{55}\begin{bmatrix} 8632.0 & 19616.7 \\ 19616.7 & 55964.5 \end{bmatrix}$$

$$= \begin{bmatrix} 464.17 & 886.08 \\ 886.08 & 2642.15 \end{bmatrix}$$

The 95% simultaneous confidence intervals for the linear combinations

$$\mathbf{a}'(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = [1, 0]\begin{bmatrix} \mu_{11} - \mu_{21} \\ \mu_{12} - \mu_{22} \end{bmatrix} = \mu_{11} - \mu_{21}$$

and

$$\mathbf{a}'(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = [0, 1]\begin{bmatrix} \mu_{11} - \mu_{21} \\ \mu_{12} - \mu_{22} \end{bmatrix} = \mu_{12} - \mu_{22}$$

are (see Result 6.4)

$$\mu_{11} - \mu_{21}: \quad 74.4 \pm \sqrt{5.99}\,\sqrt{464.17} \quad \text{or} \quad (21.7, 127.1)$$

$$\mu_{12} - \mu_{22}: \quad 201.6 \pm \sqrt{5.99}\,\sqrt{2642.15} \quad \text{or} \quad (75.8, 327.4)$$