MATH 189
Multivariate Analysis of Variance

Wenxin Zhou
UC San Diego

Time: 2:00—3:20 & 3:30—4:50pm TueThur
Location: CENTR 115

Outline

- In the last lecture, we considered hypothesis testing problems on comparing population mean vectors from two samples.
  - Hotelling's T^2
  - Paired samples
  - Two-sample tests

- Today we will discuss how to compare the means from more than two samples.
  - Multiple-sample testing
  - Analysis of variance
  - Multivariate analysis of variance

Comparing More Than Two Populations

- We have learned how to test if one population mean vector equals a specific vector
$$H_0\colon \boldsymbol{\mu}_1 = \boldsymbol{\mu}_0 \quad \text{vs} \quad H_1\colon \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_0.$$
- Also, we have learned how to test if two populations have equal mean vectors
$$H_0\colon \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \quad \text{vs} \quad H_1\colon \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2.$$

- What if we want to test the equivalence of mean vectors among three or more populations?

- Suppose $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_g$ are mean vectors of g populations. We are interested in the following hypothesis testing problem
$$H_0\colon \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \cdots = \boldsymbol{\mu}_g \quad \text{vs} \quad H_1\colon \boldsymbol{\mu}_i \neq \boldsymbol{\mu}_j \text{ for at least one pair } i \text{ and } j.$$
- The null hypothesis assumes all mean vectors to be the same. The null hypothesis is false (alternative is true) if there exist two different mean vectors.

Example: Romano-British Pottery Data

- Romano-British Pottery shards are collected from four sites in the British Isles:
  - *L*: Llanedeyrn
  - *C*: Caldicot
  - *I*: Isle Thorns
  - *A*: Ashley Rails

- Each pottery sample was returned to the laboratory for chemical test. In these tests the concentrations of five different chemicals were measured:
  - *Al*: Aluminum
  - *Fe*: Iron
  - *Mg*: Magnesium
  - *Ca*: Calcium
  - *Na*: Sodium

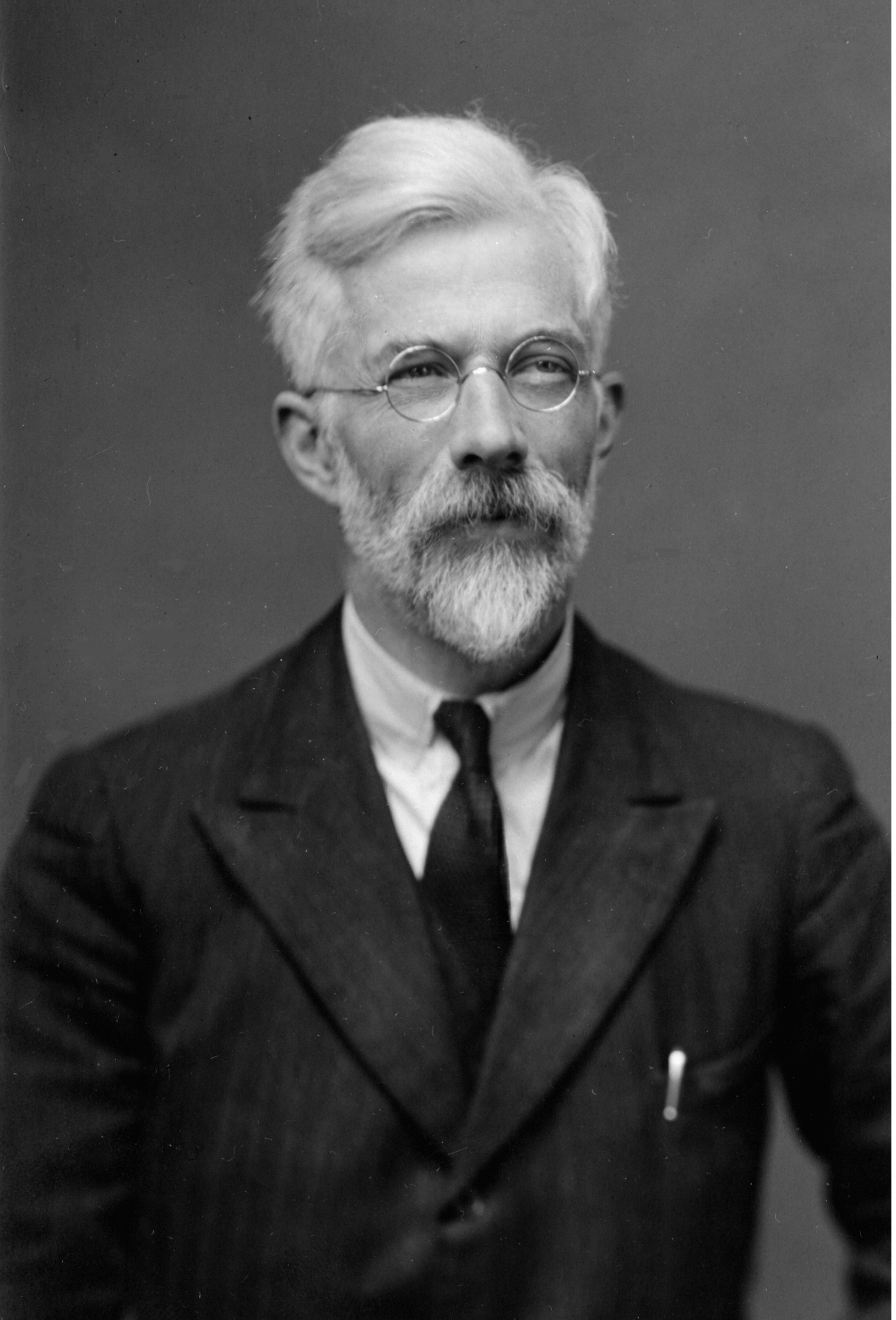
Whether the chemical content of the pottery depends on the site where the pottery was discovered?

If yes, we can use the chemical content of a pottery sample of unknown origin to hopefully determine which site the sample came from.

A Peek at the Data

- Dataset contains 26 ancient pottery shards found at four sites in British Isles.
- For each of 26 samples of pottery, the percentages of oxides of five metals are measured.

Location	Al	Fe	Mg	Ca	Na
Llanederyn	14.4	7.00	4.30	0.15	0.51
Caldicot	11.8	5.44	3.94	0.30	0.04
Island Thorns	18.3	1.28	0.67	0.03	0.03
Ashley Rails	17.7	1.12	0.56	0.06	0.06



Analysis of Variance (ANOVA)

- ANalysis Of VAriance (ANOVA) is a set of statistical tools used to analyze the differences among population means given their samples. ANOVA is useful for comparing (testing) three or more group means for statistical significance.
- ANOVA was developed by British statistician Ronald Fisher.
- ANOVA is conceptually similar to multiple two-sample tests, but is more conservative (fewer type I errors). It is suited to a wide range of practical problems.

Univariate ANOVA

- Let's start with the univariate case. Suppose we have g treatments (samples). For the i-th treatment, we observe its effects on a group of patients of size n_i. The dataset can be summarized in the following table.

Treatments (samples)

Observations

	1	**2**	$\cdots$	**g**
1	x_{11}	x_{21}	$\cdots$	x_{g1}
2	x_{12}	x_{22}	$\cdots$	x_{g2}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
n_i	x_{1n_1}	x_{2n_2}	$\cdots$	x_{gn_g}

Notations:

- $x_{ij} = j$-th observation in sample i.
- n_i = number of observations in sample i.
- $N = n_1 + \cdots + n_g$ = total sample size.
- g = number of treatments/populations/samples

Univariate ANOVA (cont.)

- Assumptions for the ANOVA are listed as follows:
  1. The data from group i has common mean μ_i, i.e. $\mathbb{E}(x_{ij}) = \mu_i$.
  2. *Homoskedasticity*: The data from all groups have common variance σ^2.
  3. *Independence*: The observations are independently sampled.
  4. *Normality*: The data are normally distributed.

- The hypothesis of interest is that all of the means are equal. Mathematically this is formulated as:

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_g \quad \text{vs} \quad H_1: \mu_i \neq \mu_j \text{ for at least one } i \neq j.$$

- The alternative hypothesis indicates there exists at least one pair of different group population means .

Total Sum of Squares

- Consider the following notation:
  - $\bar{x}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}$ = Sample mean of i-th sample
  - $\bar{x} = \frac{1}{N}\sum_{i=1}^{g}\sum_{j=1}^{n_i} x_{ij}$ = Grand mean (overall sample mean)

- The Analysis of Variance involves a partitioning of the total sum of squares, defined as:

$$SS_{total} = \sum_{i=1}^{g}\sum_{j=1}^{n_i}\left(x_{ij} - \bar{x}\right)^2 \quad .$$

- SS_{total} is the sum of squared differences between each observation and the grand mean, which measures variation of the data around the grand mean.
  - When observations are far away from the grand mean, it tends to take a large value.
  - When observations are close to the grand mean, it tends to take a small value.

Partitioning Total Sum of Squares

- An Analysis of Variance (ANOVA) is a partitioning of the total sum of squares.

$$
\begin{aligned}
SS_{total} &= \sum_{i=1}^{g}\sum_{j=1}^{n_i}\left(x_{ij}-\bar{x}\right)^2 \\
&= \sum_{i=1}^{g}\sum_{j=1}^{n_j}\left[\left(x_{ij}-\bar{x}_i\right)+\left(\bar{x}_i-\bar{x}\right)\right]^2 \\
&= \underbrace{\sum_{i=1}^{g}\sum_{j=1}^{n_i}\left(x_{ij}-\bar{x}_i\right)^2}_{\text{Error SS}} + \underbrace{\sum_{i=1}^{g} n_i(\bar{x}_i-\bar{x})^2}_{\text{Treatment SS}}
\end{aligned}
$$

- The first term is called the error sum of squares and measures the variation in the data towards their group means. Denoted as SS_{error}.
- The second term is called the treatment sum of squares and involves the differences between the group means and the Grand mean. Denoted as SS_{treat}.

Partitioning Total Sum of Squares (cont.)

$$SS_{total} = \underbrace{\sum_{i=1}^{g}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2}_{\text{Error SS}} + \underbrace{\sum_{i=1}^{g} n_i(\bar{x}_i - \bar{x})^2}_{\text{Treatment SS}}$$

- Error sum of squares (SS_{error}) measures the within-group variability.
- Treatment sum of squares (SS_{treat}) measures the between-group variability.

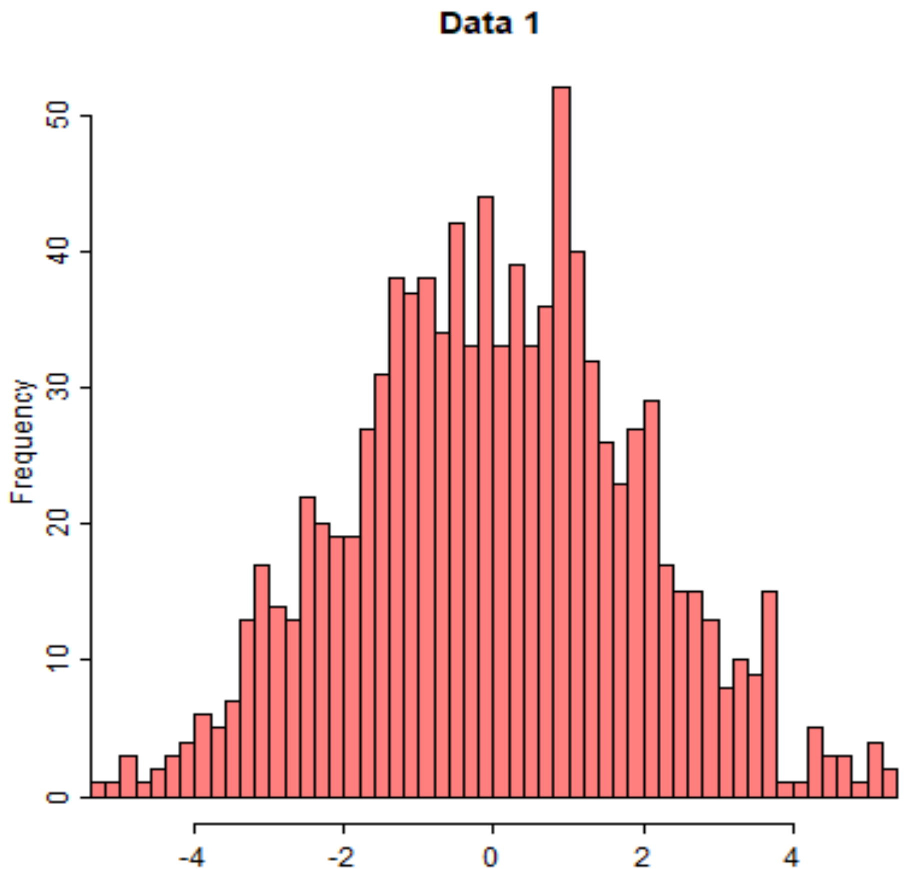
Some intuitions: Why?

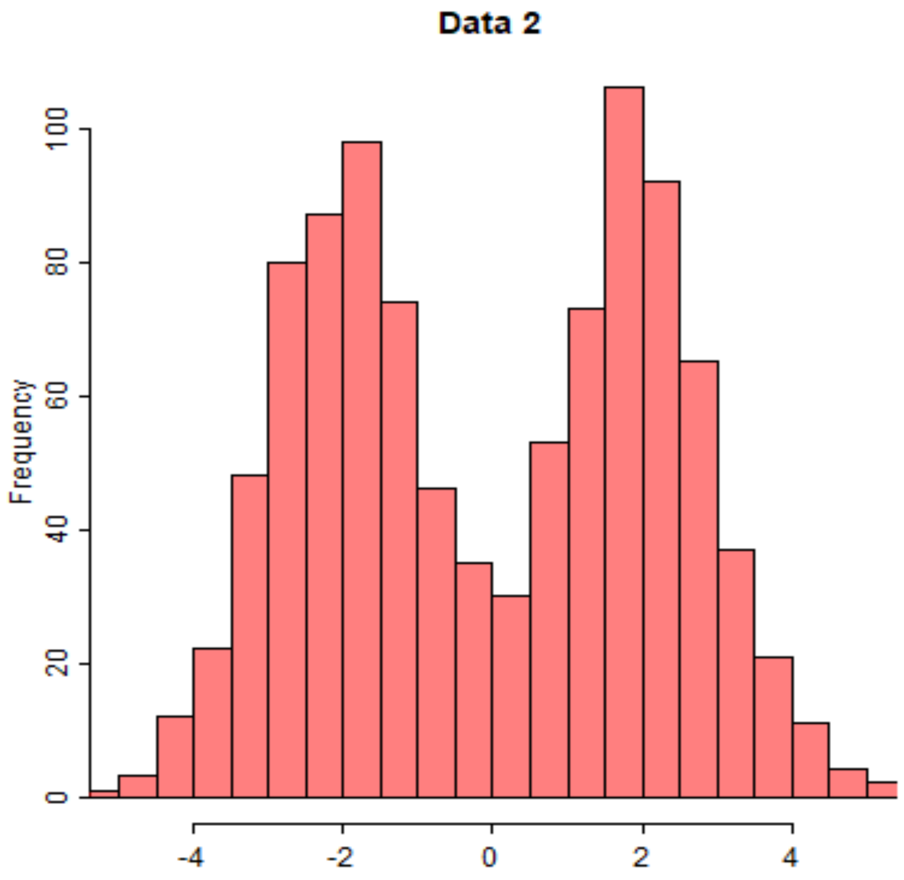
When the null hypothesis is true, i.e. $\mu_1 = \mu_2 = \cdots = \mu_g$, we would expect the between-group variability to be **small** and within-group variability to be **large**.

On the other hand, when the null hypothesis is false, we would expect the between-group variability to be **large** and within-group variability to be **small**.

A Simulated Example

- We simulate two datasets, each contains two samples (to compare).
- Here we draw the histogram for each dataset.
- Can you tell if the population means of the two samples equal or not?





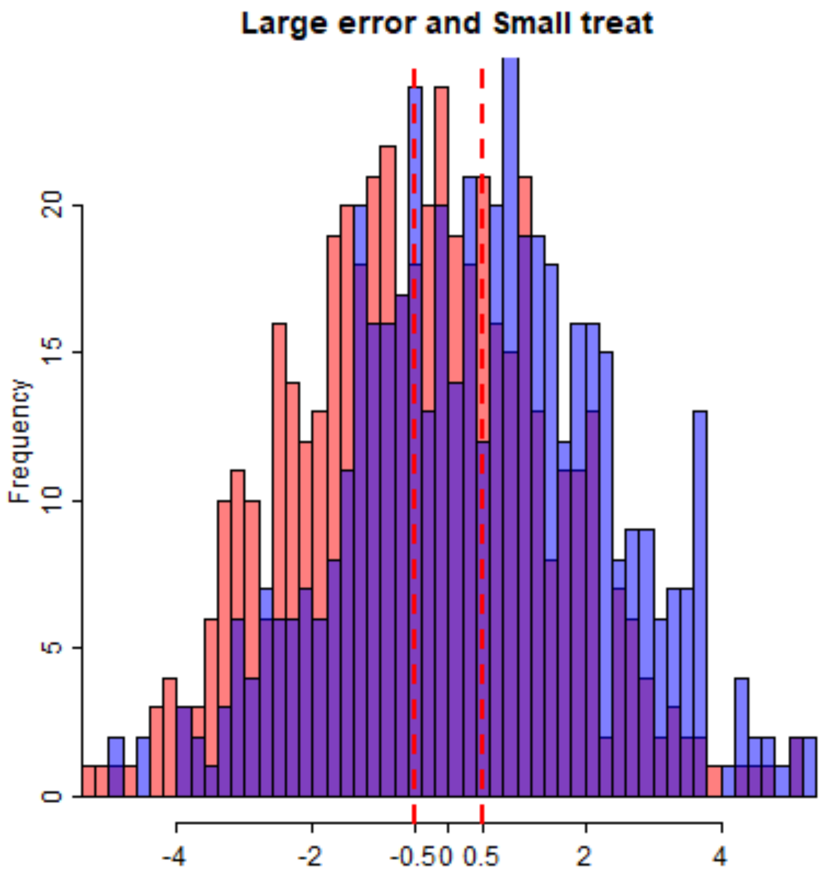
True Data Generating Process

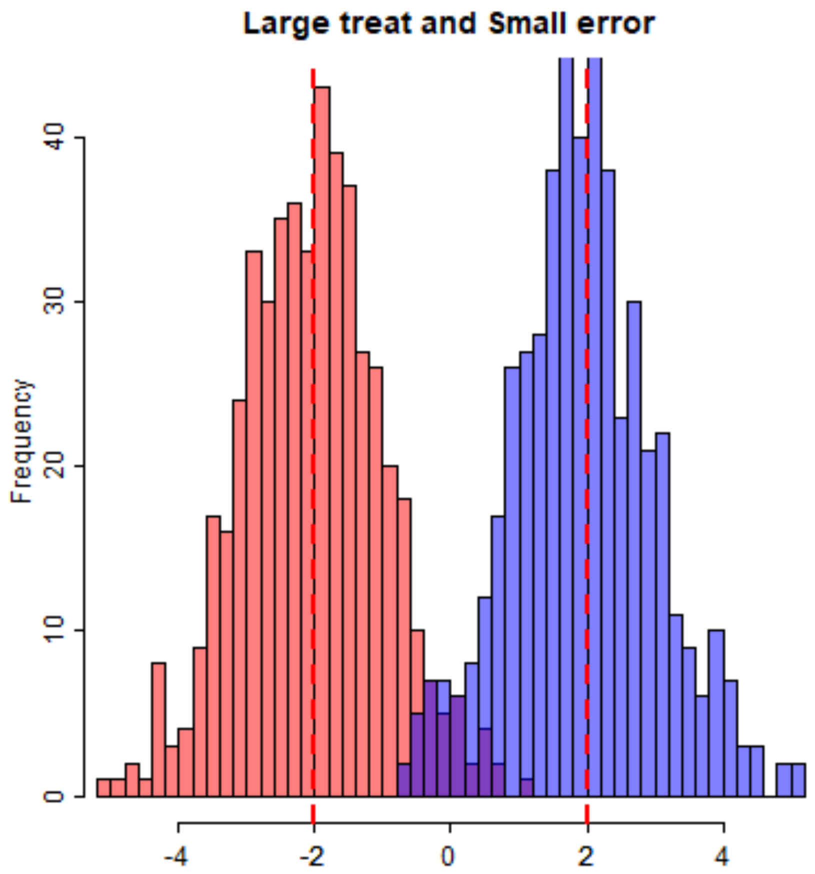
- Dataset 1:
  1. Sample 1 is generated from $N(-0.5, 2^2)$. The sample size is 500.
  2. Sample 2 is generated from $N(0.5, 2^2)$. The sample size is 500.
  3. $SS_{error} = 4027.62$. Large within-group variability.
  4. $SS_{treat} = 0.5278$. Small between-group variability.

- Dataset 2:
  1. Sample 1 is generated from $N(-2, 1^2)$. The sample size is 500.
  2. Sample 2 is generated from $N(2, 1^2)$. The sample size is 500.
  3. $SS_{error} = 923.36$. Small within-group variability.
  4. $SS_{treat} = 7.8262$. Large between-group variability.

Both datasets have un-equal population means. Why the second one is easier to tell?

- Now we use red and blue to discriminate the two samples.
- The red dashed line denotes the true population mean of each sample.
- Error sum of squares (SS_{error}) measures within-group variability.
- Treatment sum of squares (SS_{treat}) measures between-group variability.





Design of Test Statistic

$$SS_{total} = \underbrace{\sum_{i=1}^{g} \sum_{j=1}^{n_i} \left(x_{ij} - \bar{x}_i\right)^2}_{\text{Error SS}} + \underbrace{\sum_{i=1}^{g} n_i(\bar{x}_i - \bar{x})^2}_{\text{Treatment SS}}$$

- Error sum of squares (SS_{error}) measures within-group variability.
- Treatment sum of squares (SS_{treat}) measures between-group variability.

Motivated by this toy example, we should reject the null hypothesis when the Treatment Sum of Squares (SS_{treat}) is large. Given the same level of total sum of squares, a large SS_{treat} will result in a small SS_{error}.

We want to find a test statistic which is proportional to SS_{treat} and inverse proportional to SS_{error}.

A naive approach is to use the ratio

$$SS_{treat}/SS_{error}$$

Design of Test Statistic (cont.)

Let us calculate this ratio, SS_{treat}/SS_{error}, for our simulated datasets:

- Dataset 1: SS_{treat}/SS_{error} = 0.5278/4027.62 = **0.00013**
- Dataset 2: SS_{treat}/SS_{error} = 7.8262/923.36 = **0.00847**
- Some problems:
  1. Both ratios are very small! This can make the test insensitive!
  2. SS_{error} is sensitive to the size of each sample n_i, while SS_{treat} is less sensitive to n_i. Therefore, this ratio will be sensitive to the sample sizes n_i which is not ideal!
- Solution:
  1. We rescale the sum of squares by dividing their degree of freedoms.
  2. Change from ratio between sum of squares to the ratio between mean of squares!

Degrees of Freedom (DOF)

- Let us check the degrees of freedom of these sum of squares
- Error sum of squares

$$SS_{error} = \sum_{i=1}^{g} \sum_{j=1}^{n_i} \left(x_{ij} - \bar{x}_i\right)^2 \sim \chi^2_{N-g}.$$

- Treatment sum of squares

$$SS_{treat} = \sum_{i=1}^{g} n_i (\bar{x}_i - \bar{x})^2 \sim \chi^2_{g-1}.$$

- Total sum of squares

$$SS_{total} = \sum_{i=1}^{g} \sum_{j=1}^{n_i} \left(x_{ij} - \bar{x}\right)^2 \sim \chi^2_{N-1}.$$

- Therefore, the three degrees of freedom satisfy

$$DF_{total} = DF_{error} + DF_{treat}.$$

An F-Statistic as the Ratio of Means of Squares

- Error **mean** of squares

$$MS_{error} = \frac{1}{N-g} SS_{error} = \frac{1}{N-g} \sum_{i=1}^{g} \sum_{j=1}^{n_i} \left(x_{ij} - \bar{x}_i\right)^2.$$

- Treatment **mean** of squares

$$MS_{treat} = \frac{1}{g-1} SS_{treat} = \frac{1}{g-1} \sum_{i=1}^{g} n_i (\bar{x}_i - \bar{x})^2.$$

- Then we can define an F-statistic as the ratio between MS_{treat} and MS_{error},

$$F = \frac{MS_{treat}}{MS_{error}} \sim F_{g-1,N-g}.$$

At significance level α, we reject the null hypothesis $H_0: \mu_1 = \mu_2 = \cdots = \mu_g$ if

$$F > F_{g-1,N-g,\alpha}.$$

Example: Romano-British Pottery Data

Question: Which chemical element varies significantly across sites?

- As a naive approach to assess the significance of individual variables (chemical elements), we consider the following univariate multi-sample testing:

$$H_0^{(k)}: \mu_{1k} = \mu_{2k} = \cdots = \mu_{gk} \text{ for } k = 1, \dots, m.$$

- For the k-th chemical, we apply a univariate ANOVA and calculate an F-statistic

$$F^{(k)} \text{ for } k = 1, \dots, m.$$

- For a significance level α, we reject the null hypothesis $H_0^{(k)}$ if

$$F^{(k)} > F_{g-1,N-g,\alpha}.$$

- *Sounds a good plan? How about FWER?*

Example: Romano-British Pottery Data

- To control the type I errors among multiple testing, we can refer to an old friend – Bonferroni correction.

- To control the FWER at level α, we adjust the significance level for each individual test to $\frac{1}{m}\alpha$ instead of α.

- At a significance level α, we reject the null hypothesis $H_0^{(k)}$ if
$$F^{(k)} > F_{g-1,N-g,\alpha/m}.$$

- Other type I error control methods can also be applied depending on your goals.

Example: Romano-British Pottery Data

- Let's get back to the Pottery data, which contains $m = 5$ variables, $g = 4$ groups, and a total number of $N = 26$ observations. For an $\alpha = 0.05$ level test, we reject the k-th null hypothesis $H_0^{(k)}$ if

$$F^{(k)} > F_{3,22,0.01} = 4.82.$$

- The testing results are listed below. As all F statistics exceed the critical value 4.82, we reject all the null hypotheses.

Element	F statistic	P-value	Decision
Al: Aluminum	26.67	<0.0001	Reject
Fe: Iron	89.88	<0.0001	Reject
Mg: Magnesium	49.12	<0.0001	Reject
Ca: Calcium	29.16	<0.0001	Reject
Na: Sodium	9.50	0.0003	Reject

Conclusion: All chemical elements differ significantly among the sites. Each element is significantly different between at least one pair of sites.

Multivariate ANOVA

- Suppose we measure m variables over g populations. For the i-th population, we observe a sample of size n_i.

- The dataset can be summarized in the table on the right.

Treatments

Observations

	1	$\cdots$	g
1	$\boldsymbol{x}_{11} = \begin{pmatrix} x_{11,1} \\ \vdots \\ x_{11,m} \end{pmatrix}$	$\cdots$	$\boldsymbol{x}_{g1} = \begin{pmatrix} x_{g1,1} \\ \vdots \\ x_{g1,m} \end{pmatrix}$
2	$\boldsymbol{x}_{12} = \begin{pmatrix} x_{12,1} \\ \vdots \\ x_{12,m} \end{pmatrix}$	$\cdots$	$\boldsymbol{x}_{g2} = \begin{pmatrix} x_{g2,1} \\ \vdots \\ x_{g2,m} \end{pmatrix}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$
n_i	$\boldsymbol{x}_{1n_1} = \begin{pmatrix} x_{11,1} \\ \vdots \\ x_{1n_1,m} \end{pmatrix}$	$\cdots$	$\boldsymbol{x}_{gn_g} = \begin{pmatrix} x_{gn_g,1} \\ \vdots \\ x_{gn_g,m} \end{pmatrix}$

Notations:

- $\boldsymbol{x}_{ij} = j$-th observation in group i.
- $n_i =$ number of observations in group i.
- $N = n_1 + \cdots + n_g =$ total sample size.
- $g =$ number of treatments/populations/groups

Multivariate ANOVA (cont.)

- The assumptions here are essentially the same as the univariate ANOVA, except we have m variables:
    1. The data from group i has common **mean vector** $\boldsymbol{\mu}_i$, i.e. $\mathbb{E}(\boldsymbol{x}_{ij}) = \boldsymbol{\mu}_i$.
    2. *Homoskedasticity*: The data from all groups have common **covariance matrix $\boldsymbol{\Sigma}$**.
    3. *Independence*: The observations are independently sampled.
    4. *Normality*: The data are **multivariate** normally distributed.

- We are interested in testing the null hypothesis that all group mean vectors are equal:

$$H_0: \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \cdots = \boldsymbol{\mu}_g \;\text{ vs }\; H_1: \mu_{ik} \neq \mu_{jk} \text{ for at least one } i \neq j \text{ and at least one } k.$$

- This says that the null hypothesis is false if at least one pair of treatments is different on at least one variable.

Total Sum of Squares and Cross Products

- Consider the following notation:
  - $\bar{\boldsymbol{x}}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} \boldsymbol{x}_{ij}$ = Sample mean **vector** of i-th sample
  - $\bar{\boldsymbol{x}} = \frac{1}{N}\sum_{i=1}^{g}\sum_{j=1}^{n_i} \boldsymbol{x}_{ij}$ = Grand mean **vector** (overall sample mean)

- The multivariate analog is the Total Sum of Squares and is a cross products matrix of size $m \times m$:

$$\mathbf{T} = \sum_{i=1}^{g}\sum_{j=1}^{n_i}(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}})(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}})' \ .$$

  - The diagonal entries of $\mathbf{T}$ are the total sum of squares of each variable.
  - The off-diagonal entries of $\mathbf{T}$ measures the dependence of two variables across all observations.

Partitioning Total Sum of Squares $\mathbf{T}$

- We may partition the total sum of squares and cross products as follows:

$$
\begin{aligned}
\mathbf{T} &= \sum_{i=1}^{g}\sum_{j=1}^{n_i}(\boldsymbol{x}_{ij}-\overline{\boldsymbol{x}})(\boldsymbol{x}_{ij}-\overline{\boldsymbol{x}})' \\
&= \sum_{i=1}^{g}\sum_{j=1}^{n_j}\{(\boldsymbol{x}_{ij}-\overline{\boldsymbol{x}}_i)+(\overline{\boldsymbol{x}}_i-\overline{\boldsymbol{x}})\}\{(\boldsymbol{x}_{ij}-\overline{\boldsymbol{x}}_i)+(\overline{\boldsymbol{x}}_i-\overline{\boldsymbol{x}})\}' \\
&= \sum_{i=1}^{g}\sum_{j=1}^{n_i}\underbrace{(\boldsymbol{x}_{ij}-\overline{\boldsymbol{x}}_i)(\boldsymbol{x}_{ij}-\overline{\boldsymbol{x}}_i)'}_{\text{Error SS: }\mathbf{E}}+\sum_{i=1}^{g}\underbrace{(\overline{\boldsymbol{x}}_i-\overline{\boldsymbol{x}})(\overline{\boldsymbol{x}}_i-\overline{\boldsymbol{x}})'}_{\text{Hypothesis SS: }\mathbf{H}}
\end{aligned}
$$

- $\mathbf{E}$ is the *Error Sum of Squares and Cross Products*, and $\mathbf{H}$ is the *Hypothesis Sum of Squares and Cross Products*.
- We tend to reject the null hypothesis

$$H_0\colon \boldsymbol{\mu}_1=\boldsymbol{\mu}_2=\cdots=\boldsymbol{\mu}_g,$$

if *Hypothesis Sum of Squares and Cross Products* matrix $\mathbf{H}$ is large relative to the *Error Sum of Squares and Cross Products* matrix $\mathbf{E}$.

Wilks's Lambda (Ratio of Determinants)

- The first test statistic for Multivariate ANOVA is named Wilks's Lambda which is named after American statistician Samuel S. Wilks,

$$\Lambda = \frac{|\mathbf{E}|}{|\mathbf{T}|} = \frac{|\mathbf{E}|}{|\mathbf{E} + \mathbf{H}|}$$

- where $\mathbf{E}$ is the error sum of squares, $\mathbf{H}$ is the hypothesis sum of squares and $\mathbf{T}$ is the total sum of squares. $|\cdot|$ denotes the determinant of a matrix.

- The Wilks's Lambda statistic follows a Lambda distribution. In general, we will reject the null hypothesis if Wilks's lambda is small (close to zero).

- Wilks's Lambda statistic can be transformed to an F-statistic in a complicated way.

Pillai’s Trace (Trace of Ratio)

- The second test statistic for Multivariate ANOVA is named Pillai’s Trace:

$$\mathrm{V} = \mathrm{trace}(\mathbf{H}[\mathbf{H}+\mathbf{E}]^{-1}),$$

where $\mathbf{E}$ is the error sum of squares and $\mathbf{H}$ is the hypothesis sum of squares.
Here $\mathrm{trace}(\cdot)$ denotes the trace of a matrix.

- If $\mathbf{H}$ is large relative to $\mathbf{E}$, then the Pillai’s Trace will take a large value. Thus, we reject the null hypothesis if this test statistic is large.

- Pillai’s Trace can be transformed to an F-statistic up to a scaling factor depending on the data.

$$F = \frac{s_1 \mathrm{V}}{s_2 - \mathrm{V}},$$ for some data-dependent parameters s_1 and s_2.

Hotelling-Lawley Trace (Trace of Ratio)

- The third test statistic for Multivariate ANOVA is named Hotelling-Lawley Trace, which is defined as,

$$\mathrm{U} = \mathrm{trace}(\mathbf{H}\mathbf{E}^{-1}),$$

where $\mathbf{E}$ is the error sum of squares and $\mathbf{H}$ is the hypothesis sum of squares.

- If $\mathbf{H}$ is large relative to $\mathbf{E}$, then the Hotelling-Lawley Trace will take a large value. Thus, we reject the null hypothesis if this test statistic is large.
- Hotelling-Lawley Trace can be transformed to an F-statistic up to a scaling factor depending on the data:

$$F = s \cdot \mathrm{U}, \text{ for some data-dependent parameter } s.$$

Simpler than Pillai's Trace!

Roy's Maximum Root (Largest Eigenvalue of Ratio)

- The fourth test statistic for Multivariate ANOVA is named Roy's Maximum Root which is defined as,
$$\mathrm{R} = \lambda(\mathbf{H}\mathbf{E}^{-1}),$$

  where $\mathbf{E}$ is the error sum of squares and $\mathbf{H}$ is the hypothesis sum of squares.
  Here $\lambda(\cdot)$ denotes the largest eigenvalue of a matrix.

- If $\mathbf{H}$ is large relative to $\mathbf{E}$, then the Roy's Maximum Root will take a large value. Thus, we reject the null hypothesis if this test statistic is large.

- Roy's Maximum Root can also be transformed to an F-statistic up to a scaling factor depending on the data.
$$F = t \cdot \mathrm{R}, \text{ for some data-dependent parameter } t.$$

Simple in calculation at a price of accuracy as it only uses partial information!

Example: Romano-British Pottery data

Recall the null hypothesis

$$H_0: \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \cdots = \boldsymbol{\mu}_g,$$

and alternative hypothesis

$$H_1: \mu_{ik} \neq \mu_{jk} \text{ for at least one } i \neq j \text{ and at least one } k.$$

Associated Question: Is there at least one chemical varies significantly over at least one pair of sites?

What is the difference compared with univariate ANOVA?

1. The alternative hypothesis is different. (Answering different questions!)
2. The correlation among variables is ignored in univariate case.
3. Interpretation will be different.

Example: Romano-British Pottery data

- Let's compare the MANOVA test results for the four test statistics
    1. Wilks' Lambda: $\Lambda = \frac{|\mathbf{E}|}{|\mathbf{T}|} = \frac{|\mathbf{E}|}{|\mathbf{E}+\mathbf{H}|}$.
    2. Pillai's Trace: $\mathrm{V} = \mathrm{trace}(\mathbf{H}[\mathbf{H}+\mathbf{E}]^{-1})$.
    3. Hotelling-Lawley Trace: $\mathrm{U} = \mathrm{trace}(\mathbf{H}\mathbf{E}^{-1})$.
    4. Roy's Maximum Root: $\mathrm{R} = \lambda(\mathbf{H}\mathbf{E}^{-1})$.

- A common procedure can be summarized as follows:
    1. Calculate $\mathbf{E}$ and $\mathbf{H}$.
    2. Calculate test statistic.
    3. Transform the test statistic to an F statistic.
    4. Calculate critical value/p-value, and make conclusion.

Example: Romano-British Pottery data

- Error Sum of Squares and Cross Products matrix $\mathbf{E}$ can be calculated as follows:

$$\mathbf{E} = \sum_{i=1}^{g} \sum_{j=1}^{n_i} (\boldsymbol{x}_{ij} - \overline{\boldsymbol{x}}_i)(\boldsymbol{x}_{ij} - \overline{\boldsymbol{x}}_i)' \in \mathbb{R}^{m \times m}.$$

The results are listed in the table below:

Error SS	Al	Fe	Mg	Ca	Na
Al	48.288	7.080	0.608	0.106	0.588
Fe	7.080	10.950	0.527	-0.155	0.066
Mg	0.608	0.527	15.429	0.435	0.027
Ca	0.106	-0.155	0.435	0.051	0.010
Na	0.588	0.066	0.027	0.010	0.199

Example: Romano-British Pottery data

- Hypothesis Sum of Squares and Cross Products matrix $\mathbf{H}$ can be calculated as follows:

$$\mathbf{H} = \sum_{i=1}^{g} (\overline{\boldsymbol{x}}_i - \overline{\boldsymbol{x}})\,(\overline{\boldsymbol{x}}_i - \overline{\boldsymbol{x}})' \in \mathbb{R}^{m\times m}.$$

The results are listed in the table below:

Treatment SS	Al	Fe	Mg	Ca	Na
Al	175.610	-149.295	-138.809	-5.889	-5.372
Fe	-149.295	134.221	117.745	4.822	5.326
Mg	-138.809	117.745	103.350	4.209	4.711
Ca	-5.889	4.822	4.209	0.205	0.155
Na	-5.372	5.326	4.711	0.155	0.258

Example: Romano-British Pottery data

- After calculated $\mathbf{E}$ and $\mathbf{H}$, we can calculate the four test statistics.
- In the table below, we list the calculated test statistics, F statistics and their p-values.

Test Statistic	Statistic Value	F statistic	P-value	Decision
Wilks' Lambda	0.0123	13.09	<0.0001	Reject
Pillai's Trace	1.554	4.30	<0.0001	Reject
Hotelling-Lawley Trace	35.438	40.59	<0.0001	Reject
Roy's Greatest Root	34.131	136.64	<0.0001	Reject

Conclusion: At least one chemical varies significantly over at least one pair of sites.

ANOVA vs MANOVA

Univariate Analysis of Variance

- Compare univariate means among multiple populations
- Decompose total sum of errors into two errors
- F-statistic is defined as the ratio of two errors (up to some factor)
- Can be applied in multiple testing way for multivariate data

Multivariate Analysis of Variance

- Compare mean vectors among multiple populations
- Decompose total sum of errors matrix into two error matrices
- Four statistics are defined on the function of the ratio between two error matrices