



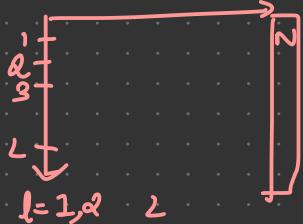
MANOVA [Multivariate Analysis of Variance]

Conceptual model

# populations (l)	# variables (p)	Hypothesis	Technique used
$l = 2$	$p = 1$	$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 \neq \mu_2$	t-test
\approx	$p \geq 2$	$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 \neq \mu_2$	Hotelling's T-square
$l \geq 2$	$p = 1$	$H_0: \mu_1 = \mu_2 = \dots = \mu_L$ $H_1: \text{at least one pair } (\mu_i = \mu_m) \text{ is not equal}$	ANOVA
	$p \geq 2$	$H_0: \mu_1 = \mu_2 = \dots = \mu_L$ $H_1: \text{at least one pair } (\mu_i = \mu_m) \text{ is not equal}$	MANOVA

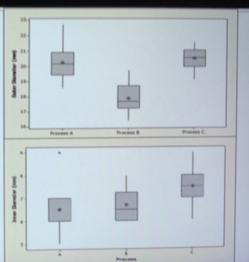
$l > 3 \rightarrow \text{MANOVA}$

ANOVA $i = 1, 2, \dots, l$ n observations



Conceptual model: An example

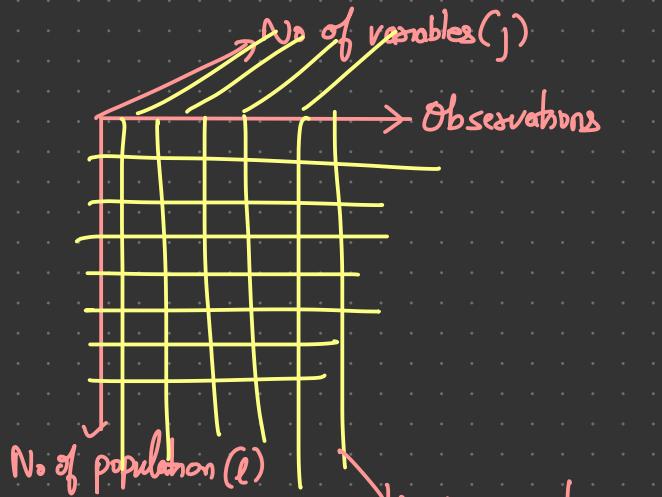
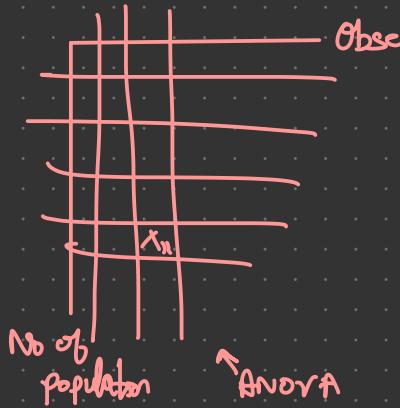
#No	Process A		Process B		Process C	
	OD	ID	OD	ID	OD	ID
1	20	6	17	6	20	8
2	21	6	17	6	20	7
3	20	9	19	7	21	8
4	21	6	17	8	20	7
5	23	7	16	6	21	8
6	19	7	19	7	21	9
7	20	6	18	7	22	7
8	19	7	18	6	19	7
9	19	5	18	6	22	6
10	20	6	20	8	20	8



} will tab collectively

$b = p = 2, l = 3 \rightarrow \text{MANOVA is used}$

$p \geq l$ for MANOVA



General Observations

$$X_{1l} = \begin{bmatrix} X_{1l1} \\ X_{1l2} \\ \vdots \\ X_{1lp} \end{bmatrix}_{p \times 1}$$

$$\mu_1 = \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \vdots \\ \mu_{1p} \end{bmatrix}_{p \times 1}$$

$$\sum_l = \begin{bmatrix} & & \\ & & \\ & & \end{bmatrix}_{p \times p}$$

$$l = 1, 2, \dots, L$$

$$\sum_1, \sum_2, \dots, \sum_L$$

$$\mu_1, \mu_2, \dots, \mu_L$$

Conceptual model: Hypothesis

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_L$$

$$H_1 : \mu_\ell \neq \mu_m \text{ for at least one pair of } \ell \text{ and } m, \ell \neq m, \ell=1,2,\dots,L \text{ and } m=1,2,\dots,L$$

$$X_{1l} = \mu + (\mu_\ell - \mu) + (X_{1l} - \mu_\ell)$$

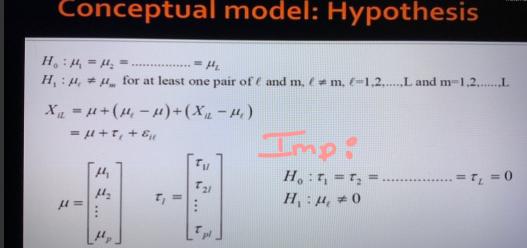
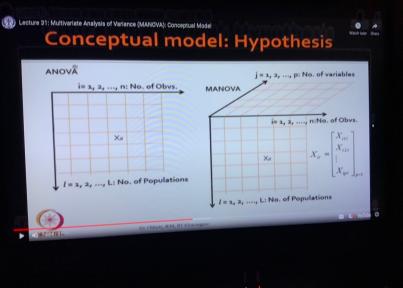
$$= \mu + \tau_\ell + \varepsilon_{1l}$$

Imp:

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} \quad \tau_\ell = \begin{bmatrix} \tau_{1\ell} \\ \tau_{2\ell} \\ \vdots \\ \tau_{p\ell} \end{bmatrix}$$

$$H_0 : \tau_1 = \tau_2 = \dots = \tau_L = 0$$

$$H_1 : \tau_\ell \neq 0$$



ANOVA

$$X_{1L} = \mu + (\mu_L - \mu) + (X_{1L} - \mu_L)$$

$$= \mu + \gamma_L + \epsilon_{1L}$$

↓
 Grand
Mean

 ↑
 Population
Effect

 Random
Error

M ANOVA

$$X_{1L} = \mu + cS_L + \epsilon_{1L}$$

$$X_{1L} = \mu + (\mu_L - \mu) + (X_{1L} - \mu_L)$$

$$\begin{bmatrix} X_{1L_1} \\ X_{1L_2} \\ \vdots \\ X_{1L_p} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} + \begin{bmatrix} \mu_{L_1} - \mu_1 \\ \mu_{L_2} - \mu_2 \\ \vdots \\ \mu_{L_p} - \mu_p \end{bmatrix} + \begin{bmatrix} X_{1L_1} - \mu_{L_1} \\ X_{1L_2} - \mu_{L_2} \\ \vdots \\ X_{1L_p} - \mu_{L_p} \end{bmatrix}$$

General Obs
Vectors

Grand Mean
Vector

Population
Effect
Vector

Random
Error
Vector

21. Multivariate Analysis of Variance (MANOVA): Conceptual Model

$$\begin{aligned}
 H_0: \mu_1 = \mu_2 = \dots = \mu_L & \quad \text{H}_0 \\
 H_1: \mu_L \neq \mu_m & \quad \text{H}_1
 \end{aligned}$$

$$\begin{aligned}
 \gamma_L = \mu_L - \mu & \quad \mu_L = \frac{n_1\mu_1 + n_2\mu_2 + \dots + n_L\mu_L}{n_1 + n_2 + \dots + n_L} \\
 \mu_L = \mu & \quad \text{if } H_0 \text{ is true} \\
 \gamma_L = \mu_L - \mu & \quad = \frac{(n_1\mu_1 + \dots + n_L\mu_L)}{n_1 + n_2 + \dots + n_L}, \mu = \\
 H_0: \gamma_L = 0, \quad L & \quad \vdots \\
 H_1: \gamma_L \neq 0, \quad \text{for at least one } L.
 \end{aligned}$$

$$X_{1L} = \mu + \gamma_L + \epsilon_{1L}$$

$$\sum_{l=0}^L \gamma_L = 0 \rightarrow \text{Equal sample size}$$

$$\sum_{l=1}^L n_l \gamma_L = 0 \rightarrow \text{Unequal Sample size}$$

$$\sum_{l=1}^L n_l \gamma_L = 0$$

$$\text{LHS} = \sum n_l (\mu_l - \mu)$$

$$\text{LHS} = \sum n_l \mu_l - \sum n_l \mu$$

$$\text{LHS} = (n_1\mu_1 + n_2\mu_2 + \dots + n_L\mu_L) - (n_1 + n_2 + \dots + n_L)\mu$$

$$\text{LHS} = 0$$

$$\mu = \text{Grand Mean} = \frac{n_1\mu_1 + n_2\mu_2 + \dots + n_L\mu_L}{n_1 + n_2 + \dots + n_L}$$

Lecture-2

Assumptions:

- ① Population covariances are equal
- ② Errors are normally distributed
- ③ Errors are IID

Lecture 32: Multivariate Analysis of Variance (MANOVA) (SS20)

Test of equality of population covariances: Box M test

Hypothesis $H_0: \Sigma_1 = \Sigma_2 = \dots = \Sigma_L$
 $H_1: \Sigma_\ell \neq \Sigma_m$, for at least pair of (ℓ, m) .

Statistic $D = (1-u)M$

$$M = -2 \ln \left[\prod_{\ell=1}^L \frac{|S_\ell|}{|S_{\text{pooled}}|} \right]^{(n_\ell-1)/2}$$

$$u = \left[\sum_{\ell} \frac{1}{(n_\ell-1)} - \frac{1}{\sum_{\ell} (n_\ell-1)} \right] \left[\frac{2p^2 + 3p - 1}{6(p+1)(L-1)} \right]$$

Decision Reject H_0 when $D > \chi_{\alpha, \nu}^2$, $\nu = \frac{1}{2} p(p+1)(L-1)$

On IIT-BHU, YouTube

$$S_{\text{pooled}} = \frac{(n_1-1)S_1 + (n_2-1)S_2 + \dots + (n_L-1)S_L}{[n_1+n_2+\dots+n_L]-L}$$

If $\Sigma_1 = \Sigma_2 = \dots = \Sigma_L$, then $|S_\ell| = |S_{\text{pooled}}|$
 ↳ ratio will become one

In us tation for linearisation

$V = \frac{1}{2} p(p+1)(L-1)$ If $D > \chi_{\alpha, \nu}^2$ we reject H_0
 ↳ variances aren't equal
 DOF for D

Example^o

Lecture 32: Multivariate Analysis of Variance (MANOVA): (SSCP)

Box M test

#N	Process A	Process B	Process C
OD	ID	OD	OD
1	6	17	6
2	21	6	17
3	20	9	19
4	23	6	17
5	23	7	16
6	19	7	19
7	20	6	18
8	19	7	18
9	19	5	18
10	20	6	20

	S ₁	S ₂	S ₃	
1.51	0.11	1.43	0.52	
0.11	1.17	0.52	0.68	
			-0.11 0.72	
	Spooled	M	1.04	
	1.29	0.17	U	0.11
	0.17	0.86	D	0.93
dof	6			
chi-sq(6, 0.05)	12.59			
		Decision		
		Accept Ho		
		So, $\Sigma_1 = \Sigma_2 = \Sigma_3 = \Sigma$		

$\chi_{\text{exp}}^2 = \bar{x}_{\text{p}x1}^T (X - \bar{X})^T S^{-1} (X - \bar{X}) \bar{x}_{\text{p}x1}$

$\text{Spooled} = \frac{(n_1-1)S_1 + (n_2-1)S_2 + (n_3-1)S_3}{10+10+10-3} = \frac{9}{27} (S_1 + S_2 + S_3) = \frac{1}{3} (S_1 + S_2 + S_3)$

$\gamma = \frac{1}{2} p(\text{p}x1) (L-1) = \frac{1}{2} \times 2 \times 3 \times (3-1) = 6$

$$\bar{x}_{\text{p}x1} \quad \bar{x}_{\text{p}x2} \quad \rightarrow \text{Same formula is used here}$$

$$(\bar{X} - 1 \cdot \bar{X}^T)^T (\bar{X} - 1 \cdot \bar{X}^T) = (n-1) \sum$$

Lecture 32: Multivariate Analysis of Variance (MANOVA): (SSCP)

Decomposition of total sum of squares

$$x_{i\ell} - \bar{x} = \bar{x}_\ell - \bar{x} + x_{i\ell} - \bar{x}_\ell$$

$$\sum_{\ell=1}^L \sum_{i=1}^{n_\ell} (x_{i\ell} - \bar{x})(x_{i\ell} - \bar{x})^T = \sum_{\ell=1}^L n_\ell (\bar{x}_\ell - \bar{x})(\bar{x}_\ell - \bar{x})^T + \sum_{\ell=1}^L \sum_{i=1}^{n_\ell} (x_{i\ell} - \bar{x})(x_{i\ell} - \bar{x})^T$$

$$\text{SSCP}_B = \sum_{\ell=1}^L n_\ell (\bar{x}_\ell - \bar{x})(\bar{x}_\ell - \bar{x})^T$$

$$\text{SSCP}_E = (n_1-1)S_1 + (n_2-1)S_2 + \dots + (n_L-1)S_L$$

$$\begin{aligned} \text{SSCP}_T &= \text{SSCP}_B + \text{SSCP}_E \\ N-1 &= L-1 + N-L \end{aligned}$$

$$N = \sum_{\ell=1}^L n_\ell$$

For ANOVA,

$$\begin{aligned} x_{1L} - \mu &+ (\mu_L - \mu) + (x_{1L} - \mu_L) \\ \hat{\mu} &= \bar{x}, \hat{\mu}_L = \bar{x}_L \\ x_{1L} &= \bar{x} + (\bar{x}_L - \bar{x}) + (x_{1L} - \bar{x}_L) \end{aligned}$$

$\left. \right\} \text{Similarly done for MANOVA}$

$$(p \times 1) \quad (p \times 1) \quad (p \times 1)$$

lecture 32: Multivariate Analysis of Variance (MANOVA): (SSCP)

$$\begin{aligned}
 & \sum_{l=1}^L \sum_{i=1}^{n_l} (\bar{x}_{il} - \bar{\bar{x}}) = (\bar{\bar{x}}_l - \bar{\bar{x}}) + (\bar{x}_{il} - \bar{\bar{x}}_l) \\
 & (\bar{x}_{il} - \bar{\bar{x}})(\bar{x}_{il} - \bar{\bar{x}})^T = \underbrace{[(\bar{\bar{x}}_l - \bar{\bar{x}}) + (\bar{x}_{il} - \bar{\bar{x}}_l)]}_\text{pxp} \times \underbrace{[(\bar{\bar{x}}_l - \bar{\bar{x}}) + (\bar{x}_{il} - \bar{\bar{x}}_l)]^T}_\text{pxp} \\
 & \sum_{l=1}^L (\bar{x}_{il} - \bar{\bar{x}})(\bar{x}_{il} - \bar{\bar{x}})^T = \sum_{l=1}^L (\bar{\bar{x}}_l - \bar{\bar{x}})(\bar{\bar{x}}_l - \bar{\bar{x}})^T + \sum_{l=1}^L (\bar{\bar{x}}_l - \bar{\bar{x}})(\bar{x}_{il} - \bar{\bar{x}}_l)^T \\
 & \quad + \sum_{l=1}^L (\bar{x}_{il} - \bar{\bar{x}}_l)(\bar{\bar{x}}_l - \bar{\bar{x}})^T + \sum_{l=1}^L (\bar{x}_{il} - \bar{\bar{x}}_l)(\bar{x}_{il} - \bar{\bar{x}})^T \\
 & \sum_{l=1}^L n_l = n_L \quad \sum_{l=1}^L \bar{x}_l = n_L \bar{\bar{x}}
 \end{aligned}$$

lecture 32: Multivariate Analysis of Variance (MANOVA): (SSCP)

$$\begin{aligned}
 & \sum_{l=1}^L \sum_{i=1}^{n_l} (\bar{x}_{il} - \bar{\bar{x}})(\bar{x}_{il} - \bar{\bar{x}})^T = \sum_{l=1}^L \sum_{i=1}^{n_l} (\bar{\bar{x}}_l - \bar{\bar{x}})(\bar{\bar{x}}_l - \bar{\bar{x}})^T + \sum_{l=1}^L \sum_{i=1}^{n_l} (\bar{x}_{il} - \bar{\bar{x}}_l)(\bar{x}_{il} - \bar{\bar{x}}_l)^T \\
 & = \sum_{l=1}^L n_l (\bar{\bar{x}}_l - \bar{\bar{x}})(\bar{\bar{x}}_l - \bar{\bar{x}})^T + \sum_{l=1}^L \sum_{i=1}^{n_l} (\bar{x}_{il} - \bar{\bar{x}}_l)(\bar{x}_{il} - \bar{\bar{x}}_l)^T \\
 & \left[\begin{array}{c} SSCP_T \\ \vdots \\ SSCP_E \end{array} \right]_{pxp} = \left[\begin{array}{c} SSCP_B \\ \vdots \\ SSCP_E \end{array} \right]_{pxp} + \left[\begin{array}{c} SSCP_E \\ \vdots \\ SSCP_E \end{array} \right]_{pxp} \\
 & N - L = L - 1 + \dots + L - L
 \end{aligned}$$

$$SSCP_B = \sum_{l=1}^L n_l (\bar{\bar{x}}_l - \bar{\bar{x}})(\bar{\bar{x}}_l - \bar{\bar{x}})^T$$

$$SSCP_T = (n_1 - 1)S_1 + (n_2 - 1)S_2 + \dots + (n_L - 1)S_L$$

$$SSCP_T = SSCP_B + SSCP_E$$

↳ Decomposition of SSCP Matrix

Lecture 3

→ Example solving at sheet

One way MANOVA [Population of one find]

Source of Variation	SSCP	Dof	Wilks' Λ
population	$SSCP_B$	$L-1$	$\Lambda = \frac{ SSCP_E }{ SSCP_T }$
error	$SSCP_E$	$N-L$	$\Lambda = \frac{ SSCP_E }{ SSCP_T \cdot SSCP_B }$
Total	$SSCP_T$	$N-1$	↓ $N = \sum_{l=1}^L n_l$

Example in end

Lecture 4

Using example of last class

Estimation of parameters

$$\hat{\mu} = \bar{x} \quad \text{and} \quad \hat{\mu}_l = \bar{x}_l$$

$$\hat{\tau}_\ell = \hat{\mu}_\ell - \mu = \hat{x}_\ell - \bar{x} \quad \hat{\epsilon}_{\ell\ell} = x_{\ell\ell} - \bar{x}_\ell$$

$$\hat{\tau}_1 = \begin{pmatrix} 0.63 \\ -0.4 \end{pmatrix}; \quad \hat{\tau}_2 = \begin{pmatrix} -1.67 \\ -0.2 \end{pmatrix}; \quad \hat{\tau}_3 = \begin{pmatrix} 1.03 \\ 0.6 \end{pmatrix}$$

$$\hat{x}_\ell - \hat{x}_m = (\hat{x}_\ell - \bar{x}) - (\hat{x}_m - \bar{x}) = \bar{x}_\ell - \bar{x}_m$$

$$\bar{x}_1 - \bar{x}_2 = \begin{pmatrix} 2.3 \\ -0.2 \end{pmatrix}; \quad \bar{x}_1 - \bar{x}_2 = \begin{pmatrix} -0.4 \\ -1.00 \end{pmatrix}; \quad \bar{x}_1 - \bar{x}_2 = \begin{pmatrix} -2.7 \\ -0.8 \end{pmatrix}$$



Dr J Maiti, IEM, IIT Kharagpur

$$E(\hat{\gamma}_\ell - \hat{\gamma}_m) = E[\bar{x}_\ell - \bar{x}_m]$$

$$= E[\bar{x}_\ell] - E[\bar{x}_m] \quad [\text{As independent}]$$

$$V(\hat{\gamma}_\ell - \hat{\gamma}_m)$$

$$V(\bar{x}_\ell - \bar{x}_m) = V(\dot{\bar{x}}_\ell) + V(\bar{x}_m)$$

$$\downarrow \text{Covably} \quad = \frac{\sum_\ell}{n_\ell} + \frac{\sum_m}{n_m} = \boxed{\left(\frac{1}{n_\ell} + \frac{1}{n_m} \right) \sum}$$

$$\text{Assume } \Sigma_1 = \Sigma_2 = \Sigma_L$$

Figure 34: Multivariate Analysis of Variance (MANOVA). Estimation and Hypothesis testing

$$\gamma_L = \begin{bmatrix} \gamma_{11} \\ \gamma_{12} \\ \vdots \\ \gamma_{1L} \end{bmatrix} \quad E(\hat{\gamma}_\ell - \hat{\gamma}_m) = \begin{bmatrix} \hat{\gamma}_{11} - \hat{\gamma}_{m1} \\ \hat{\gamma}_{12} - \hat{\gamma}_{m2} \\ \vdots \\ \hat{\gamma}_{1L} - \hat{\gamma}_{mL} \end{bmatrix} = \begin{bmatrix} \mu_{11} - \mu_{m1} \\ \vdots \\ \mu_{1j} - \mu_{mj} \\ \vdots \\ \mu_{1p} - \mu_{mp} \end{bmatrix}$$

$$E(\hat{\gamma}_\ell - \hat{\gamma}_m) = E(\bar{x}_\ell - \bar{x}_m) = \mu_L - \mu_m$$

$$\hat{\gamma}_\ell - \hat{\gamma}_m \Rightarrow \hat{\gamma}_{1j} - \hat{\gamma}_{mj}$$

$$E(\hat{\gamma}_{1j} - \hat{\gamma}_{mj}) = \mu_{1j} - \mu_{mj}$$

$$V(\hat{\gamma}_\ell - \hat{\gamma}_m) = \left(\frac{1}{n_\ell} + \frac{1}{n_m} \right) \Sigma = C \sum$$

$$\sum = \begin{bmatrix} \sum_{11} & \dots & \sum_{1j} \\ \vdots & \ddots & \vdots \\ \sum_{ij} & \dots & \sum_{jj} \end{bmatrix}$$

$$\nu(\hat{\mu}_{ij} - \hat{\mu}_{mj}) = \nu \sum_{jj} = \left(\frac{1}{n_1} + \frac{1}{n_m} \right) \sum_{jj}$$

$$\sum = \frac{\sum_{jj}}{N-L} = \frac{SSCP_E}{N-L} = \frac{MSE}{L-L} = W = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1P} \\ w_{21} & w_{22} & \dots & w_{2P} \\ \vdots & \vdots & \ddots & \vdots \\ w_{P1} & w_{P2} & \dots & w_{PP} \end{bmatrix}$$

$$\hat{\sigma}_{ij}^2 = w_{ij}$$

20:23 / 30:00

Find interval estimate
Random Variable: $\hat{\mu}_{ej} - \hat{\mu}_{mj}$
Expectation: $\mu_{ej} - \mu_{mj}$

$$\text{Var}(\hat{\mu}_{ej} - \hat{\mu}_{mj}) = \left(\frac{1}{n_e} + \frac{1}{n_m} \right) w_{jj}$$

Estimation of parameters

No of comparisions = $m = pL(L-1)/2$

Bonferroni SCI

$$(\bar{x}_{ij} - \bar{x}_{mj}) - t_{N-L}(\alpha/2m) \sqrt{w_{jj} \left(\frac{1}{n_e} + \frac{1}{n_m} \right)} \leq \mu_{ij} - \mu_{mj}$$

$$\leq (\bar{x}_{ij} - \bar{x}_{mj}) + t_{N-L}(\alpha/2m) \sqrt{w_{jj} \left(\frac{1}{n_e} + \frac{1}{n_m} \right)}$$

22:24 / 30:00

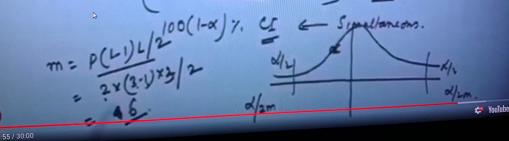
Dr J Marti, IITM, BT Kharagpur

Random variable: $\hat{\mu}_{ej} - \hat{\mu}_{mj}$ Expected values: $\frac{\mu_{ej} - \mu_{mj}}{2(n_e - n_m)}$

$$\text{Var}(\hat{\mu}_{ej} - \hat{\mu}_{mj}) = \left(\frac{1}{n_e} + \frac{1}{n_m} \right) w_{jj}$$

$$P \left\{ \hat{\mu}_{ej} - \hat{\mu}_{mj} \leq \frac{\mu_{ej} - \mu_{mj}}{2(n_e - n_m)} \leq u \right\} = 1 - \alpha.$$

$$m = \frac{p(L-1)L/2}{2 \times (2-1) \times 3/2} = 4.6$$



25:55 / 30:00

If any interval contains zero, then that variable is not differentiating.

$$\text{Ex: } -0.25 \leq \underline{\quad} \leq 1.80$$

Contains 0 [Not creating differentiating variable]