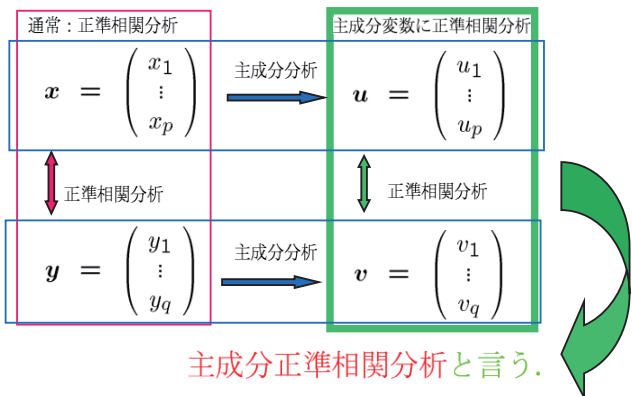


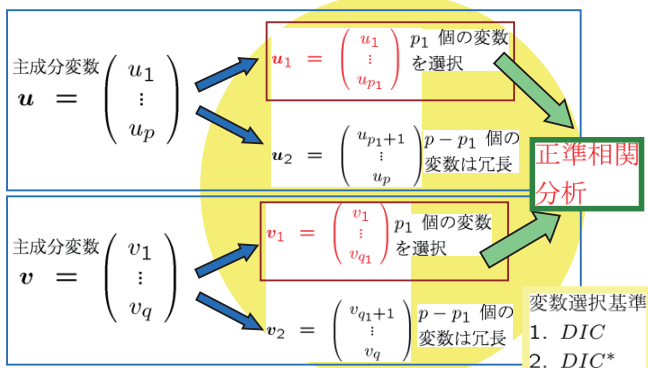
# 多変量モデルの選択における新展開

研究代表者 杉山高一 研究員

## 主成分正準相関分析におけるモデル選択



モデル選択 (冗長性の検討)



数値例  $x$  : 高校での学業成績、 $y$  : センター試験の得点

表1:主成分正準相関分析によるDIC, DIC\* (昇順)

$p_1$	$q_1$	DIC	DIC*	
2( $u_1, u_2$ )	3( $v_1, v_2, v_3$ )	100.9556	105.8770	最適
2( $u_1, u_2$ )	4( $v_1, v_2, v_3, v_4$ )	101.3951	107.0795	
3( $u_1, u_2, u_3$ )	3( $v_1, v_2, v_3$ )	102.7035	108.3584	
2( $u_1, u_2$ )	5( $v_1, v_2, v_3, v_4, v_5$ )	104.0097	111.1101	
3( $u_1, u_2, u_3$ )	4( $v_1, v_2, v_3, v_4$ )	105.0182	111.6681	
5( $u_1, u_2, u_3, u_4, u_5$ )	5( $v_1, v_2, v_3, v_4, v_5$ )	110.0000	131.7739	

表2:主成分正準相関係数

$p_1$	$q_1$	$r_1^2$	$r_2^2$	$r_3^2$	$r_4^2$	$r_5^2$
2( $u_1, u_2$ )	3( $v_1, v_2, v_3$ )	0.5270	0.1195			
2( $u_1, u_2$ )	4( $v_1, v_2, v_3, v_4$ )	0.5520	0.1207			
3( $u_1, u_2, u_3$ )	3( $v_1, v_2, v_3$ )	0.5308	0.1200	0.0562		
2( $u_1, u_2$ )	5( $v_1, v_2, v_3, v_4, v_5$ )	0.5568	0.1458	0.1049	0.0119	
3( $u_1, u_2, u_3$ )	4( $v_1, v_2, v_3, v_4$ )	0.5564	0.1211	0.0565		
5( $u_1, u_2, u_3, u_4, u_5$ )	5( $v_1, v_2, v_3, v_4, v_5$ )	0.5934	0.2162	0.1172	0.0518	0.0054

## 数量化2類における変数の冗長性検定法の開発

質的データの数値化  $\{1, 0\}$  変数  $X_1, \dots, X_p$   
多群 ( $q$  群) の判別分析

$X_{k+1}, \dots, X_p$  は冗長か?  
( $X_{k+1}, \dots, X_p$  は十分か?)

検定統計量と分布

$$L = |S_\Omega| / |S_\omega| \quad \text{多変量回帰法}$$

$$= |W_\Omega| / |W_\omega| \quad \text{サンプルスコア法}$$

$$\sim \Lambda_{p-k}(q-1, n-q-k)$$

検定統計量の理論分布 (近似)  
とシミュレーション

$p = 7, q = 4, k = 5, n = 300$

