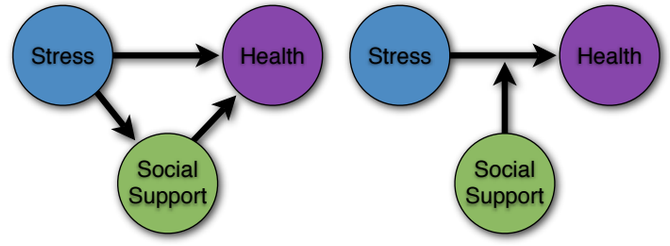
Review

- Interpreting outliers
  - Whether you regard an outlier as signal (i.e., as evidence for whatever you're measuring) or disregarding it as noise (i.e, as evidence for something other than what you're measuring – will depend on the context of the outlier.
- Regression diagnostics
  - Homoscedasticity, multicollinearity, and singularity.
- A *substantive* change...
  - ...is one that changes the interpretation of the relationship.
- Sequential (hierarchical) Regression
  - Rather than being entered all at once, predictors enter the equation in groups specified by the researcher.
  - R^2_{change} represents the improvement in R^2 when the second predictor is added. R^2_{change} is tested with an F test, which is referred to as the F Change. A significant F Change means that the variables added in that step significantly improved the prediction.

Review

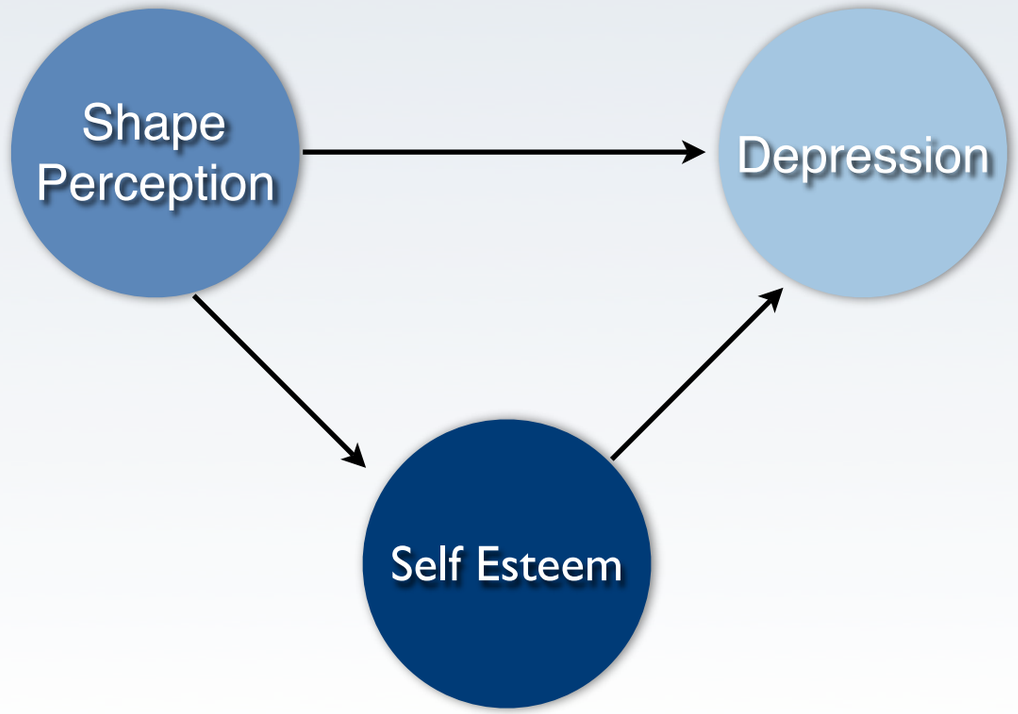
- ANOVA via multiple regression
  - Convert a categorical variable into multiple dichotomous variables, then do an ANOVA using multiple regression.
  - Provides the ‘missing link’ between the correlational and analysis of variance methods.
  - Dummy coding of categorical variables
  - Tests of significance
  - Use of categorical variables in multiple regression

- Moderated Multiple Regression
  - The case of the third variable
  - Mediating or Moderating
  - We can perform a 2x2 ANOVA by multiple regression to test whether the effects of A on Y is moderated by B.
  - That is, if the interaction contributes to the prediction (tested by R^2_{change} in the same way as sequential multiple regression).
  - A ‘simple slopes’ analysis will verify that the pattern of relationships is what you expected them to be.

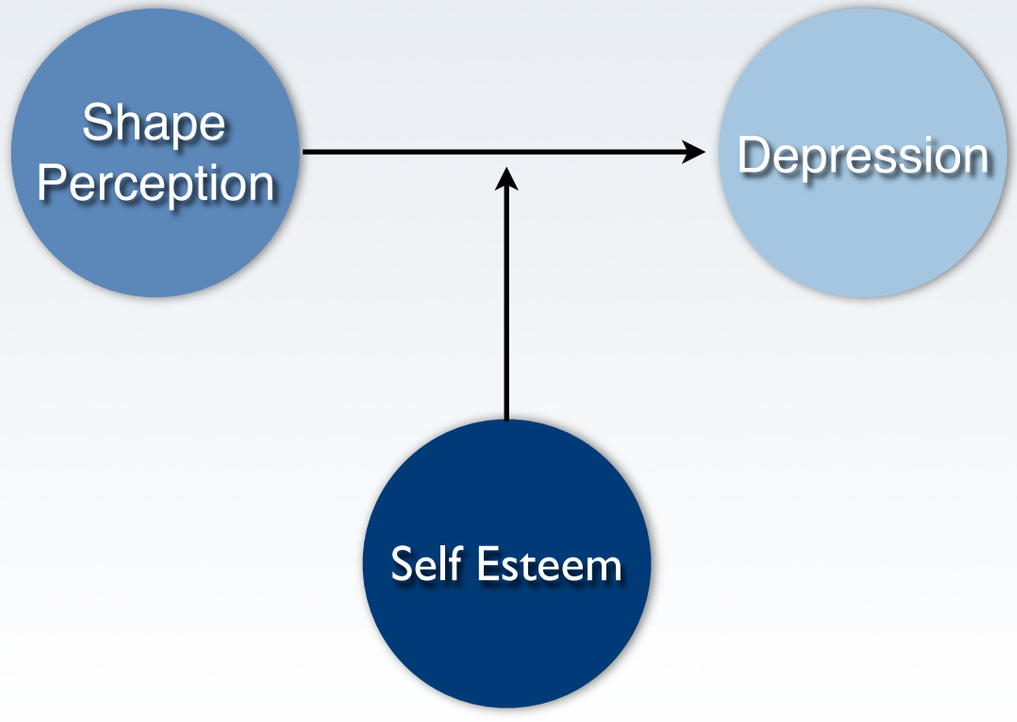


Mediating Moderating

Mediated Effects Model



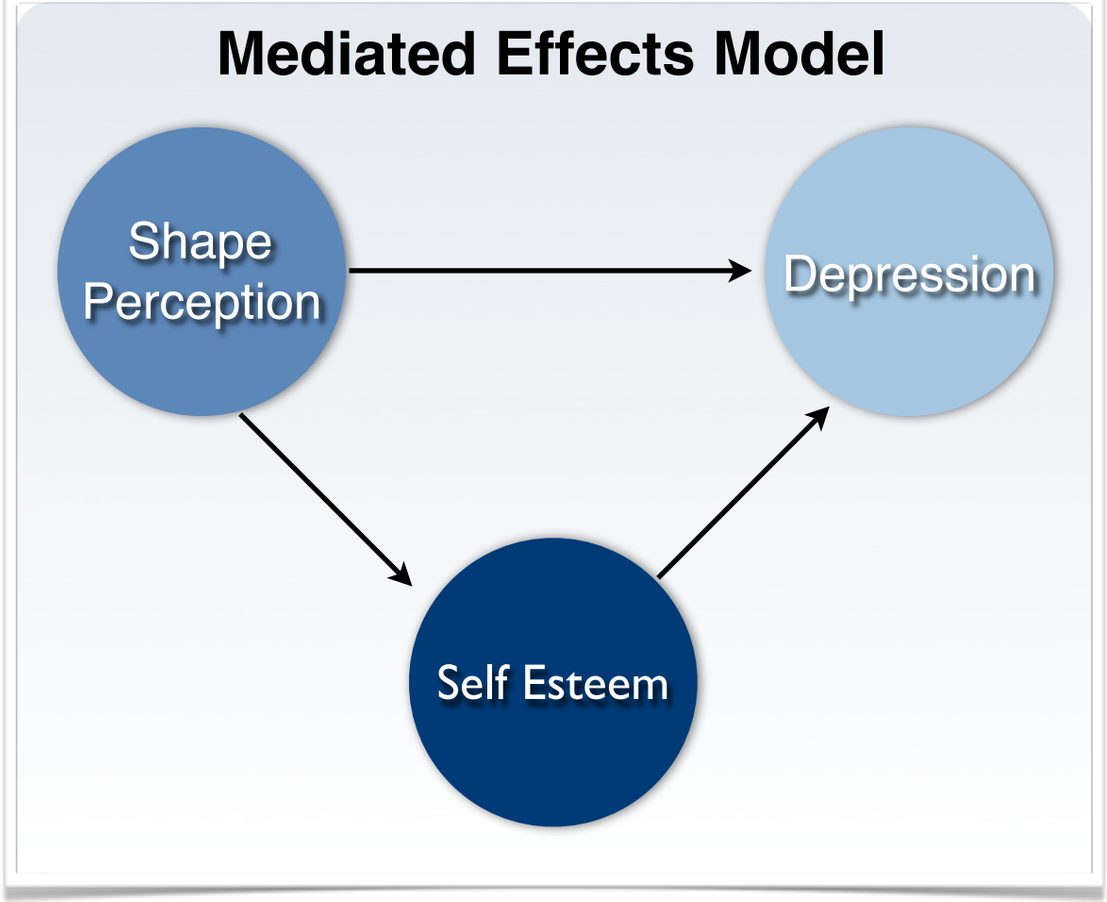
Moderated Effects Model



"A variable may be considered a mediator to the extent to which it carries the influence of a given independent variable (IV) to a given dependent variable (DV). Generally speaking, mediation can be said to occur when:

1. the IV significantly affects the mediator,
2. the IV significantly affects the DV in the absence of the mediator,
3. the mediator has a significant unique effect on the DV, and
4. the effect of the IV on the DV shrinks upon the addition of the mediator to the model."

- Preacher & Leonardelli, 2001



Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	24.023	2.994		8.023	.000
	shpercep	2.852	.686	.417	4.158	.000

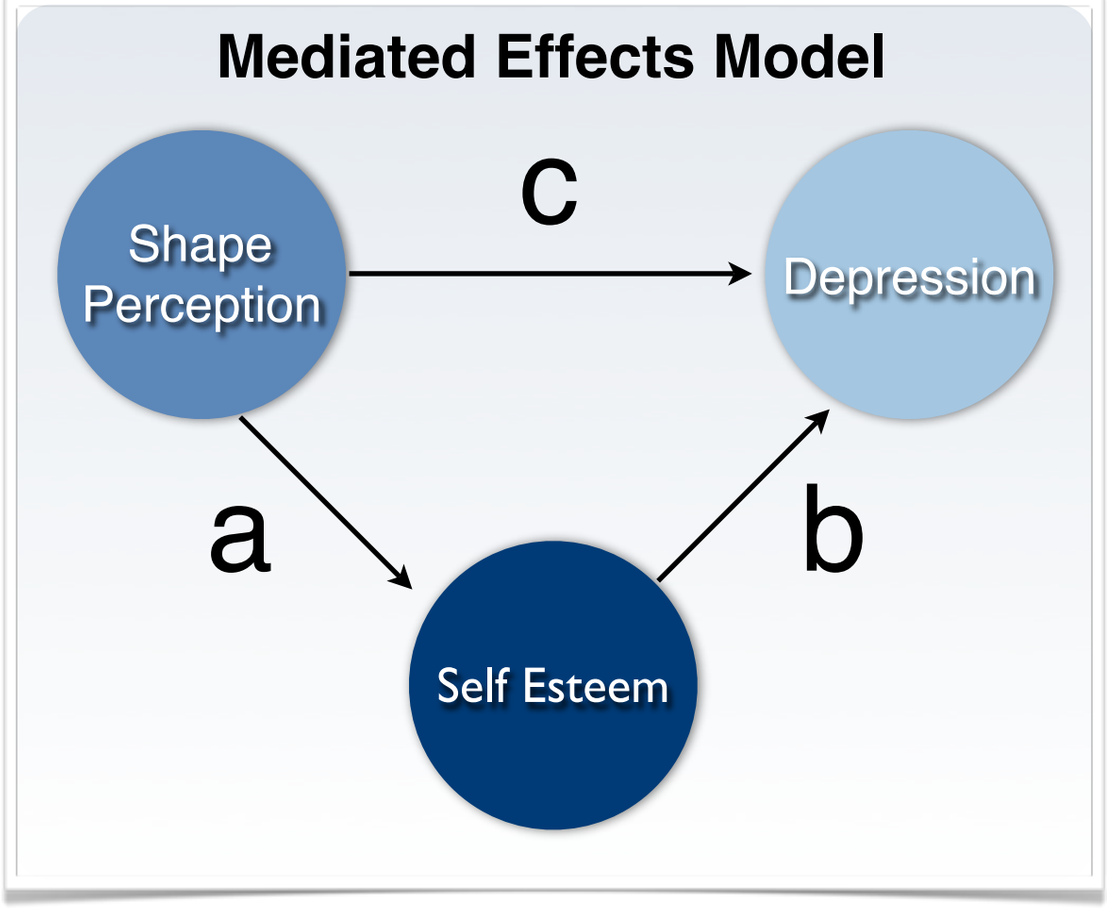
a. Dependent Variable: rses

Coefficients[a]

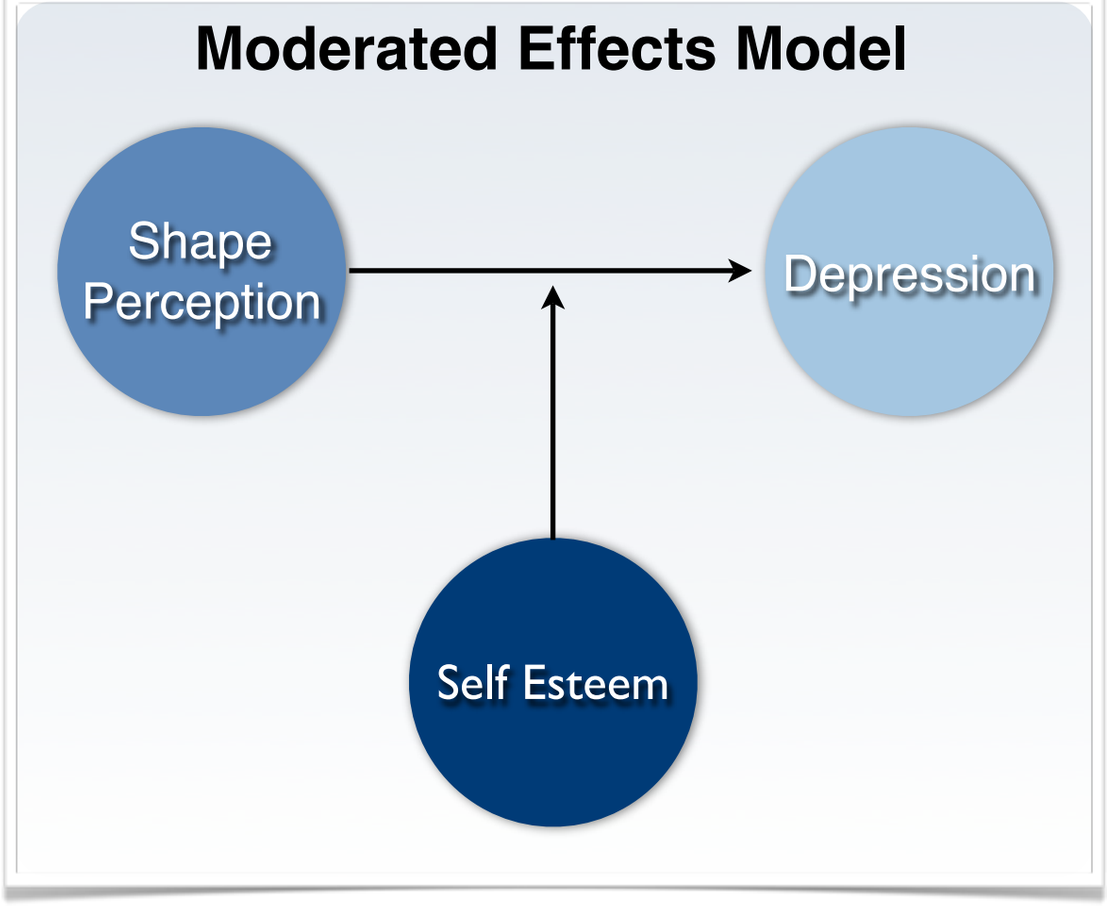
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	37.674	2.282		16.507	.000
	shpercep	-1.029	.431	-.172	-2.390	.019
	rses	-.635	.063	-.723	-10.075	.000

a. Dependent Variable: bdi

	Input:		Test statistic:	*p*-value:
a	2.852	Sobel test:	-3.84333366	0.00012137
b	-.635	Aroian test:	-3.82726989	0.00012957
s_a	.686	Goodman test:	-3.85960142	0.00011357
s_b	.063	Reset all	Calculate	

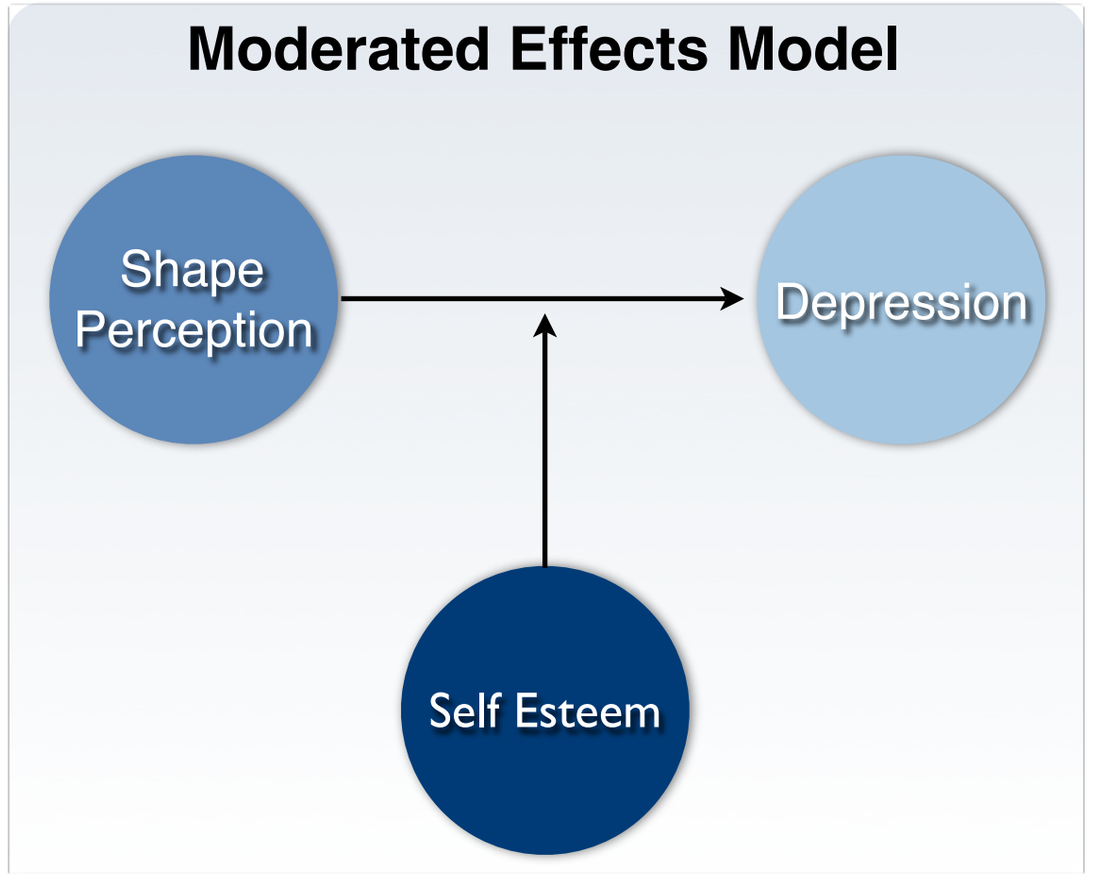


Another way of thinking about the influence of Self Esteem on the relationship between Shape Perception and Depression is to suggest that for those women with good Self Esteem, there will be little relationship between Shape Perception and Depression. However, for women with low Self Esteem, there will be a strong relationship between Shape Perception and Depression. This describes Self Esteem as *moderating* the relationship.



Depression ← Shape Perception, Self Esteem
explains 65.6% of the variability in Depression.

Depression ← Shape Perception, Self Esteem,
Self Esteem x Shape Perception
explains 66.3% of the variability in Depression.



Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.810[a]	.656	.647	4.27748
2	.814[b]	.663	.651	4.25599

a. Predictors: (Constant), mod, idv

b. Predictors: (Constant), mod, idv, x

ANOVA[d]

Model		Sum of Squares	df	Mean Square	F	Sig.	R Square Change
1	Regression	2820.673	2	1410.337	77.081	.000[a]	
	Residual	1482.047	81	18.297			
	Total	4302.720	83				
2	Subset Tests x	32.968	1	32.968	1.820	.181[b]	.008
	Regression	2853.642	3	951.214	52.514	.000[c]	
	Residual	1449.078	80	18.113			
	Total	4302.720	83				

a. Predictors: (Constant), mod, idv

b. Tested against the full model.

c. Predictors in the Full Model: (Constant), mod, idv, x.

d. Dependent Variable: dv

We're testing the difference between these two models (i.e., the change in R^2).

Discriminant analysis

- Motivational example
- Schematic representation and purposes
- The simplest extension of the t-test - the two group example and linear composites
- Partitioning sums of squares
- An extension: Three groups and linear composites
- Discriminant ratios and eigenvalues
- Canonical correlations
- Significance tests in discriminant analysis

A Motivational Example

Imagine you are in prison and enrolled in fourth year. During your stay, you observe that there seem to be consistent differences between your fellow prisoners depending on the type of crime, e.g. people convicted of fraud tend to be liars etc. You need an Honours topic. So, you get permission to do a study. You convince 200 prisoners to fill out questionnaires. These include; The EPQ, the Hostility questionnaire, the Attitudes to criminals scale, the Rosenberg Self Esteem scale. Counting subscales, you end up with 11 variables. You get five groups of 20 prisoners convicted of Rape, Fraud, Murder, Victimless Crime, and Armed Robbery respectively. You also get 100 other prisoners convicted of crimes other than these. You want to describe differences between the groups. You call your friendly tutor at the University for advice on the appropriate data analysis.

A Motivational Example

Groups	N	Variables (11 = p)
Rape	20	[] [] [] [] [] [] [] [] [] [] []
Fraud	20	[] [] [] [] [] [] [] [] [] [] []
Murder	20	[] [] [] [] [] [] [] [] [] [] []
Victimless crime	20	[] [] [] [] [] [] [] [] [] [] []
Armed Robbery	20	[] [] [] [] [] [] [] [] [] [] []
Other	100	[] [] [] [] [] [] [] [] [] [] []

Each [] is a column of 20 scores.

Questions of interest

- Given information about their personality, can we *predict* what sort of crime a person will commit?
- Can we *describe* the differences between existing criminals in terms of personality measures?
- More generally: Is there a *relationship* between personality factors and type of crime?

Overview

- Multiple continuous (or dichotomous) variables, are known as measured variables or discriminant variables.
  - Number of discriminant variables = *p*
- A single categorical variable with multiple groups of cases, known as the grouping variable.
  - Number of groups = *k*

Y ← $X_1, X_2, \ldots X_p$

categorical, *k* levels ← *p* continuous variables

⇐ Predictive Discriminant Analysis

$Y_1, Y_2, \ldots Y_p$ ← X

p continuous variables ← categorical, *k* levels

⇐ Descriptive Discriminant Analysis

Linear Composites in Discriminant Analysis

Discriminant Analysis looks for a relationship between a **categorical** variable and a set of variables:

$$X_{cat} \leftarrow Y_1, Y_2, Y_3$$

Pick some weights: w_1, w_2, w_3

Create a linear composite:

$$C_1 = w_1 Y_1 + w_2 Y_2 + w_3 Y_3$$

Resulting in a **t-test** or **F-test**:

$$X_{cat} \leftarrow C_1$$

Recall from Lecture 2...

Bald

Hairy

Hair Density (2 Levels)	←	Entry GPA	Gene Quality	Hand Span
X_{cat}	←	Y_1	Y_2	Y_3

Recall from Lecture 2...

	GPA	Gene Quality	Hand Span
	6	2	10
	4	3	9
	4	4	11
	7	2	11
	5	2	10
Mean	5.2	2.6	10.2
	7	3	8
	5	3	7
	4	4	9
	8	2	8
	5	2	5
Mean	5.8	2.8	7.4
t-value	-.64	-.37	3.61

*These weights are arbitrary in this example. Later, we'll cover how to find optimal weights.

By doing three t-tests (one for each variable), the three variables may be correlated.

So the interpretations of the t-tests are not independent (i.e., we aren't properly assessing the effect of GPA and gene quality *independently* because GPA and gene quality may, in fact, be correlated.)

Another approach is to combine the three variables into a composite variable and perform a t-test on this composite variable.

But how do we combine the scores?

	w_1	w_2	w_3
C_1	1	1	1
C_2	1	2	-1
C_3	1	1	-2

Recall from Lecture 2...

	GPA	Gene Quality	Hand Span	C_1 $(1,1,1)$	C_2 $(1,2,-1)$	C_3 $(1,1,-2)$
	6	2	10	18	0	-12
	4	3	9	16	1	-11
	4	4	11	19	1	-14
	7	2	11	20	0	-13
	5	2	10	17	-1	-13
Mean	5.2		10.2	18	0.2	-12.6
				18	5	-6
				15	4	-6
				17	3	-10
				18	4	-6
				12	4	-3
Mean	5			16	4	-6.2
t-value	-.64	-.37	3.61	1.49	-7.76	5.23

The goal here is to find the linear composite such that the t-value for the differences between the groups is as large as possible. The weights give the 'relative importance' of the variables. The optimum weights depend essentially on the pattern of correlations among the variables.

Recall from Lecture 2...

Bald

With more than two groups, a t-value is no longer appropriate. Instead, an F-value is the appropriate index of between-group differences. The goal now would be to find the linear composite such that the F-value for the difference between groups is as large as possible.

airy

Not Quite Bald

Recall from Lecture 2...

Research Questions

- Is the overall relationship statistically significant and how strong is the relationship?
- What variables are individually important in separating (discriminating) between the groups?

A simple example

2 group Discriminant Analysis

Two groups of inmates:

- Group 1 = convicted for murder
- Group 2 = convicted for fraud

Two measured variables:

- a measure of intelligence (Y_1)
- a measure of aggression (Y_2)

$Y_1 Y_2 \quad \leftarrow \quad X$

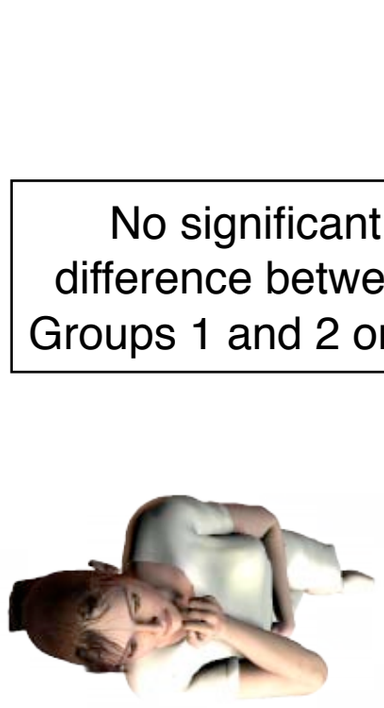
2 continuous variables

categorical 2 levels

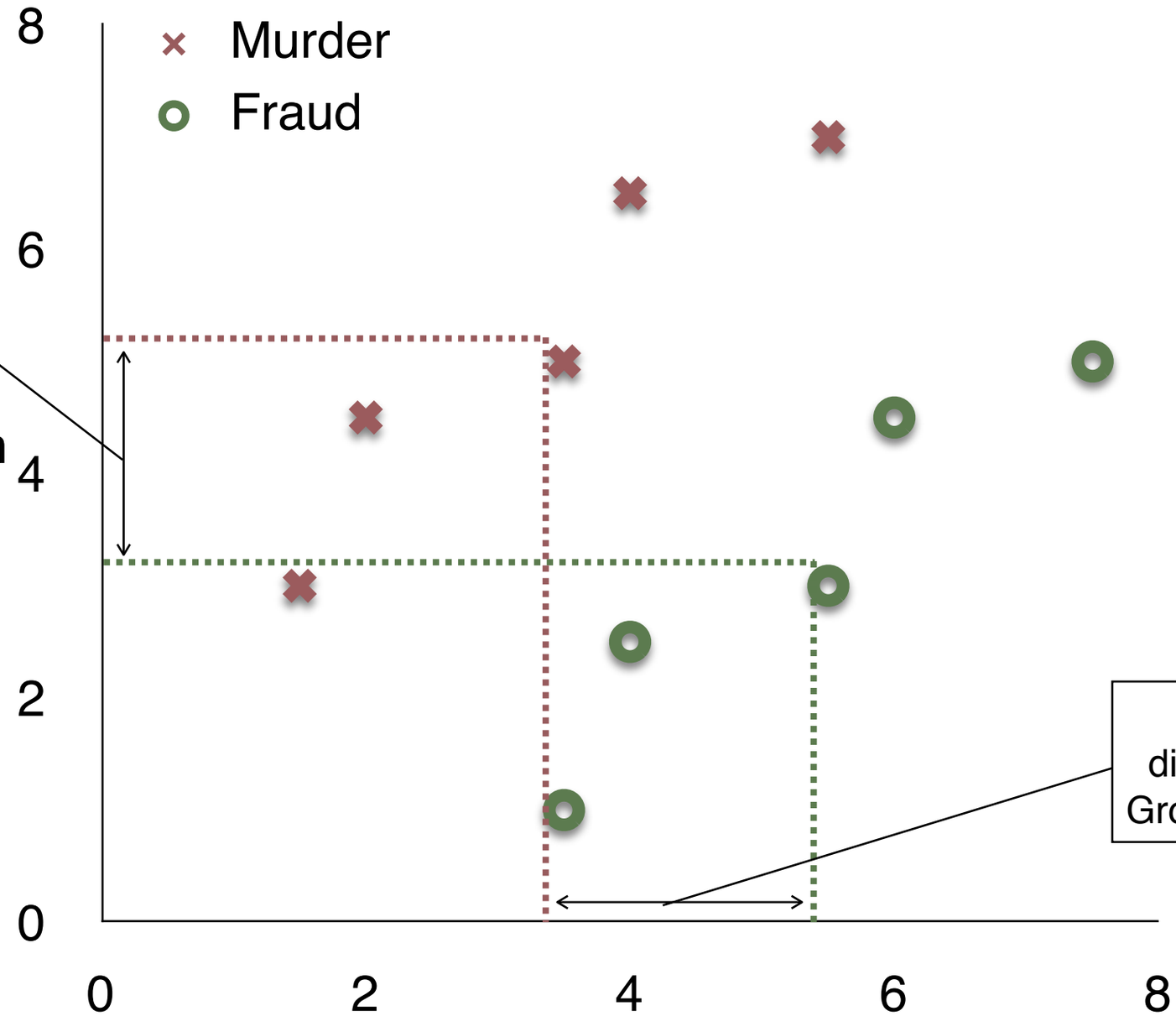
	Group (X)	Intelligence (Y_1)	Aggression (Y_2)
Murder	1	1.5	3.0
	1	2.0	4.5
	1	3.5	5.0
	1	4.0	6.5
	1	5.5	7.0
	Mean	3.3	5.2
Fraud	2	3.5	1.0
	2	4.0	2.5
	2	5.5	3.0
	2	6.0	4.5
	2	7.5	5.0
	Mean	5.3	3.2
	t value	-1.97	1.97

2 group Discriminant Analysis

Small differences between groups on Y_1 and Y_2



Aggression (Y_2)

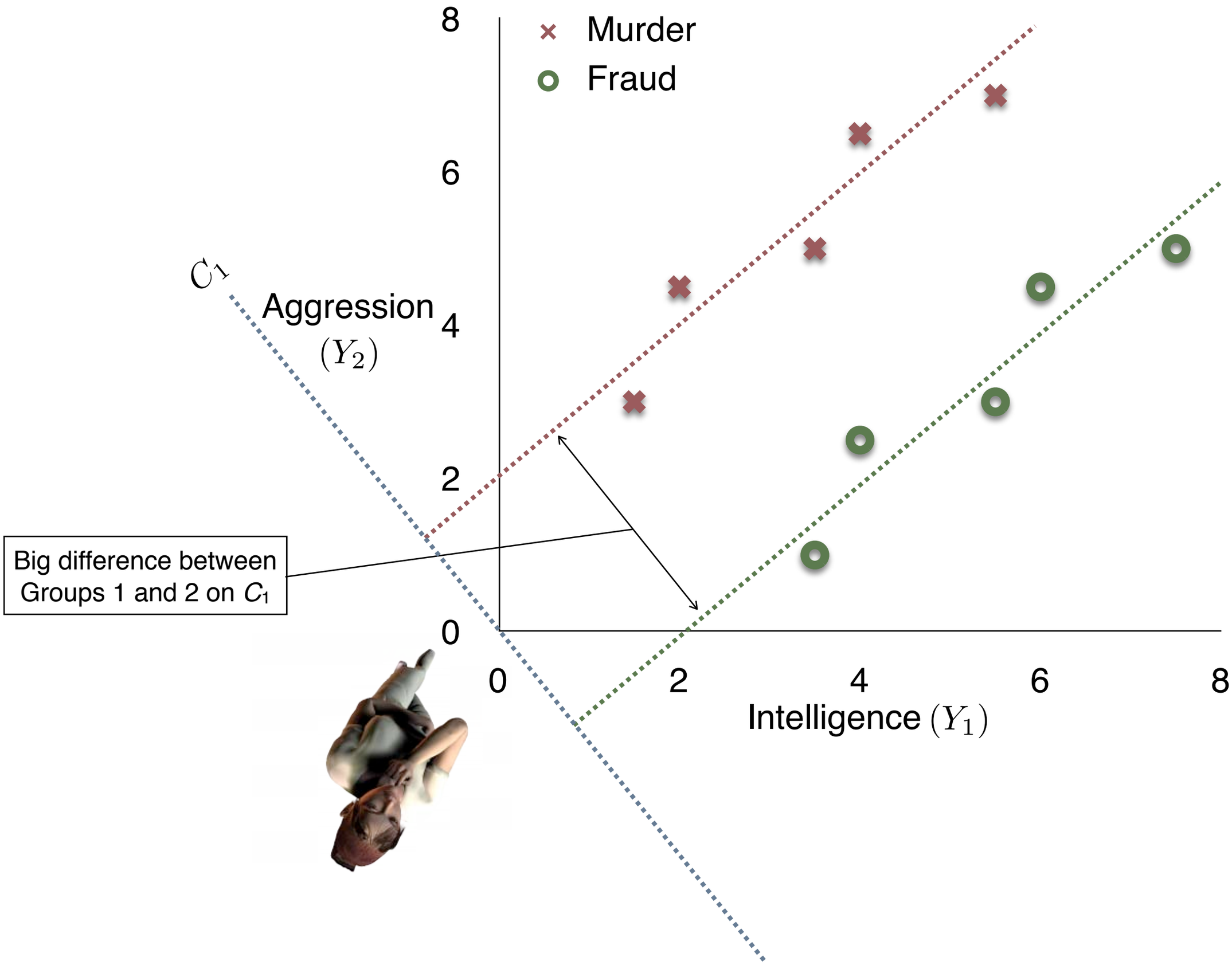


No significant difference between Groups 1 and 2 on Y_2

No significant difference between Groups 1 and 2 on Y_1

2 group Discriminant Analysis

The combination, C_1, maximises differences between groups



2 group Discriminant Analysis

The combination, C_1, maximises differences between groups

- Eyeball approach suggests big group differences if we view the data from a new direction, between Y_1 and Y_2.
- Viewing the data from a new direction is the spatial equivalent of calculating a linear composite of Y_1 and Y_2.
- In this example, the new direction implies a linear composite with weights of -1 for Y_1 and +1 for Y_2.

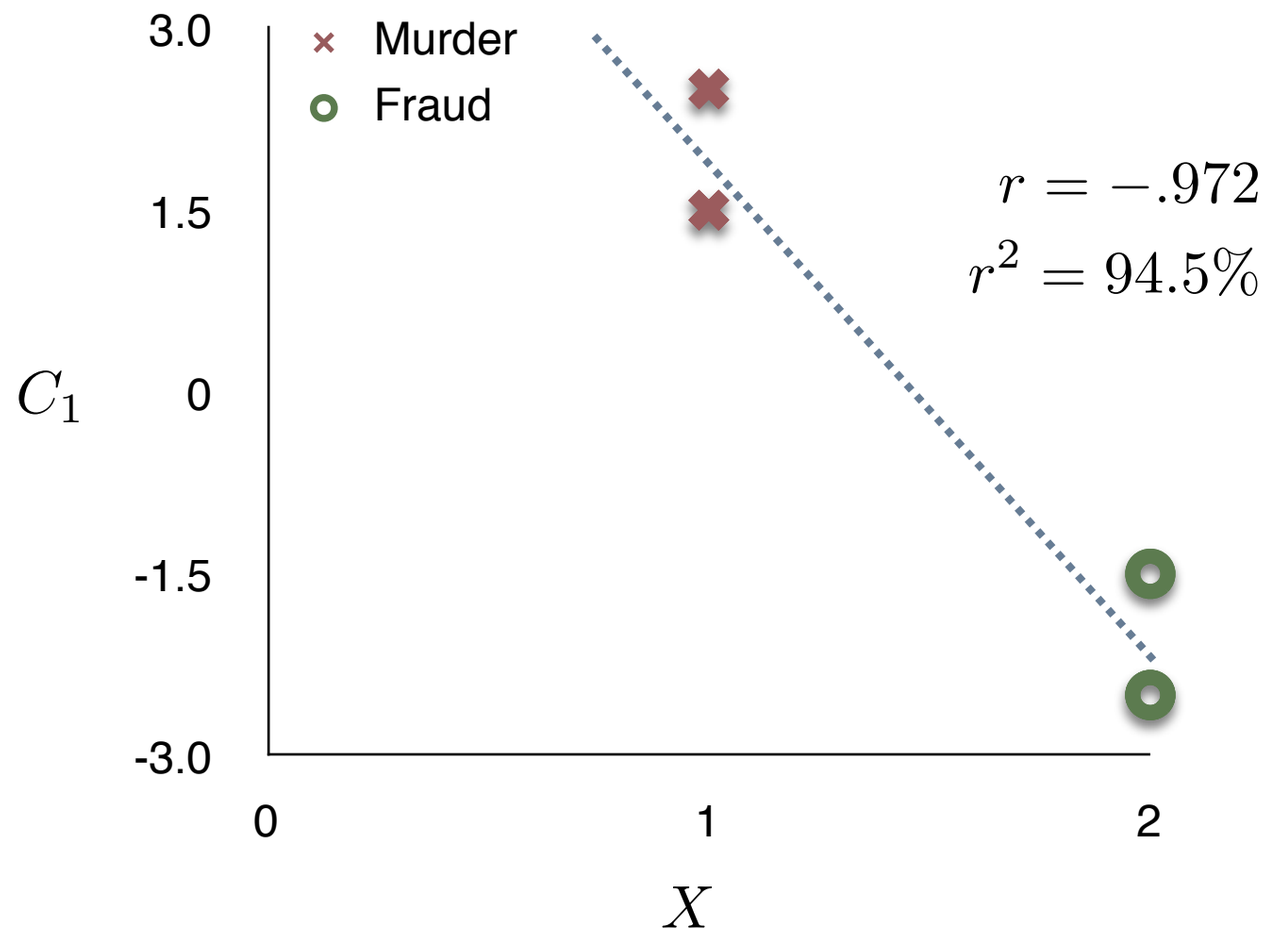
This is an extension of the t-test to two measured variables. Earlier, the goal of discriminant analysis was stated as finding the linear composite such that the t-value for the difference between the groups is as large as possible. Another way to state the discriminant analysis problem is that we wish to find the direction or dimension in the space of the discriminant variables that maximally separates the groups.

	Group (X)	Composite $(C_1 : -1, +1)$
Murder	1	1.5
	1	2.5
	1	1.5
	1	2.5
	1	1.5
Fraud	2	-2.5
	2	-1.5
	2	-2.5
	2	-1.5
	2	-2.5

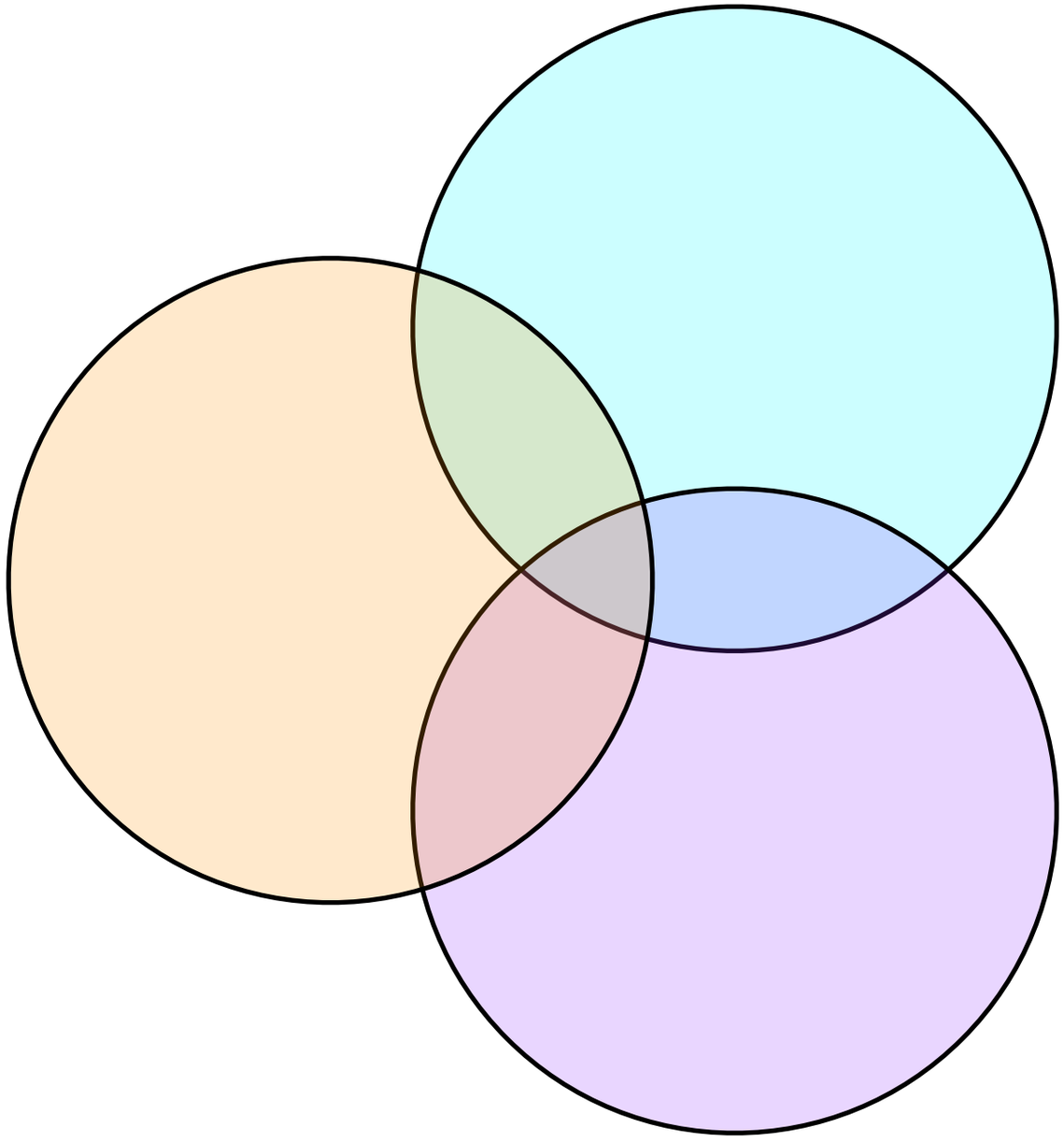
2 group Discriminant Analysis

Scatterplot of X and C_1

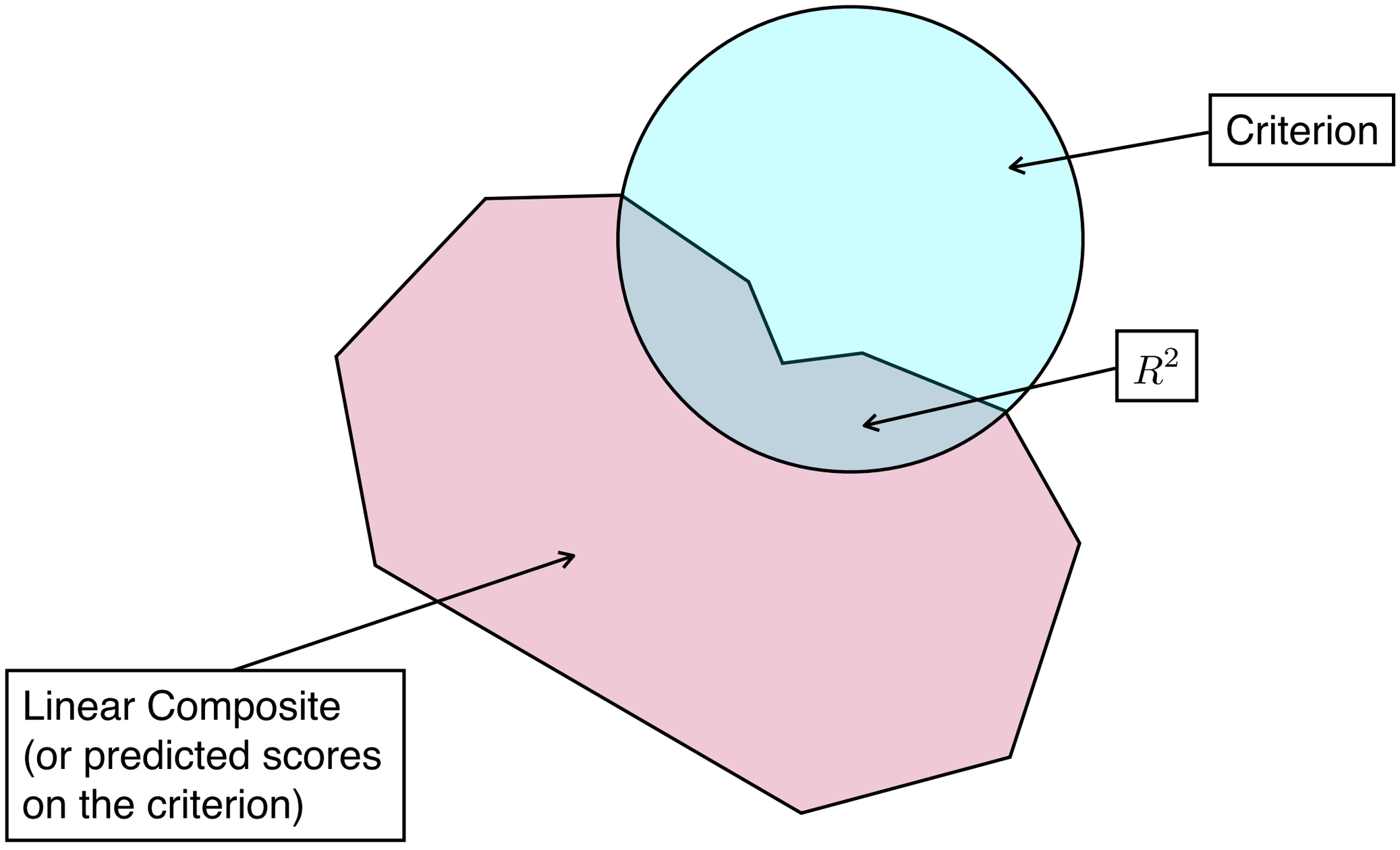
	Group (X)	Composite $(C_1 : -1, +1)$
Murder	1	1.5
	1	2.5
	1	1.5
	1	2.5
	1	1.5
Fraud	2	-2.5
	2	-1.5
	2	-2.5
	2	-1.5
	2	-2.5



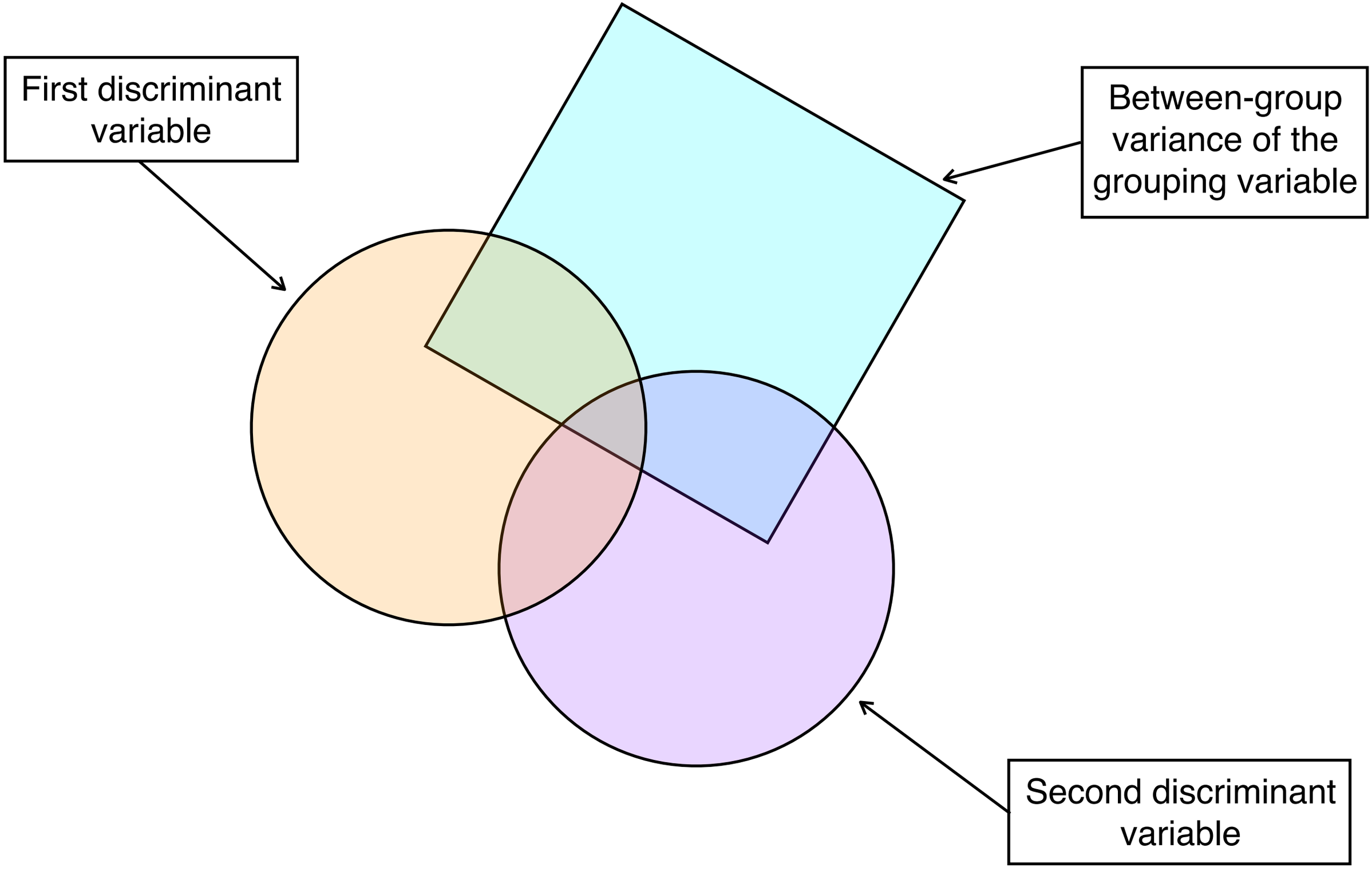
Note that the correlation between *X*, the grouping variable, and C_1, the linear composite, is -0.972. This is a point-biserial correlation. Later it will be referred to as a *canonical correlation*. The square of this correlation gives the proportion of variance in the linear composite accounted for by the grouping variable, (94.5%). C_1 summarises the information in Y_1 and Y_2 in such a way that maximises the canonical correlation.



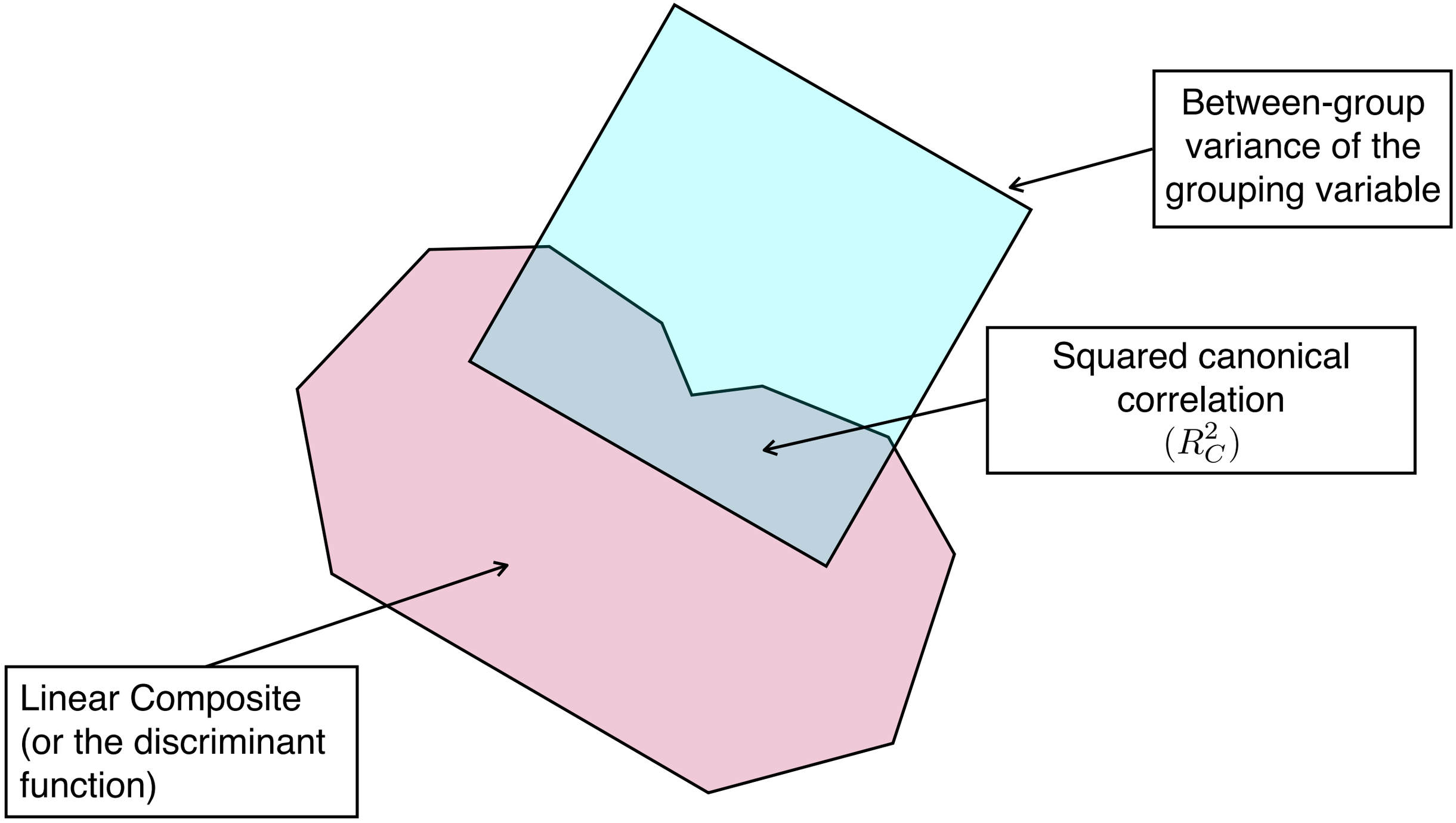
First, reconsider multiple regression with 2 predictors



The predictors are used to create a linear composite



2-Group Discriminant Analysis

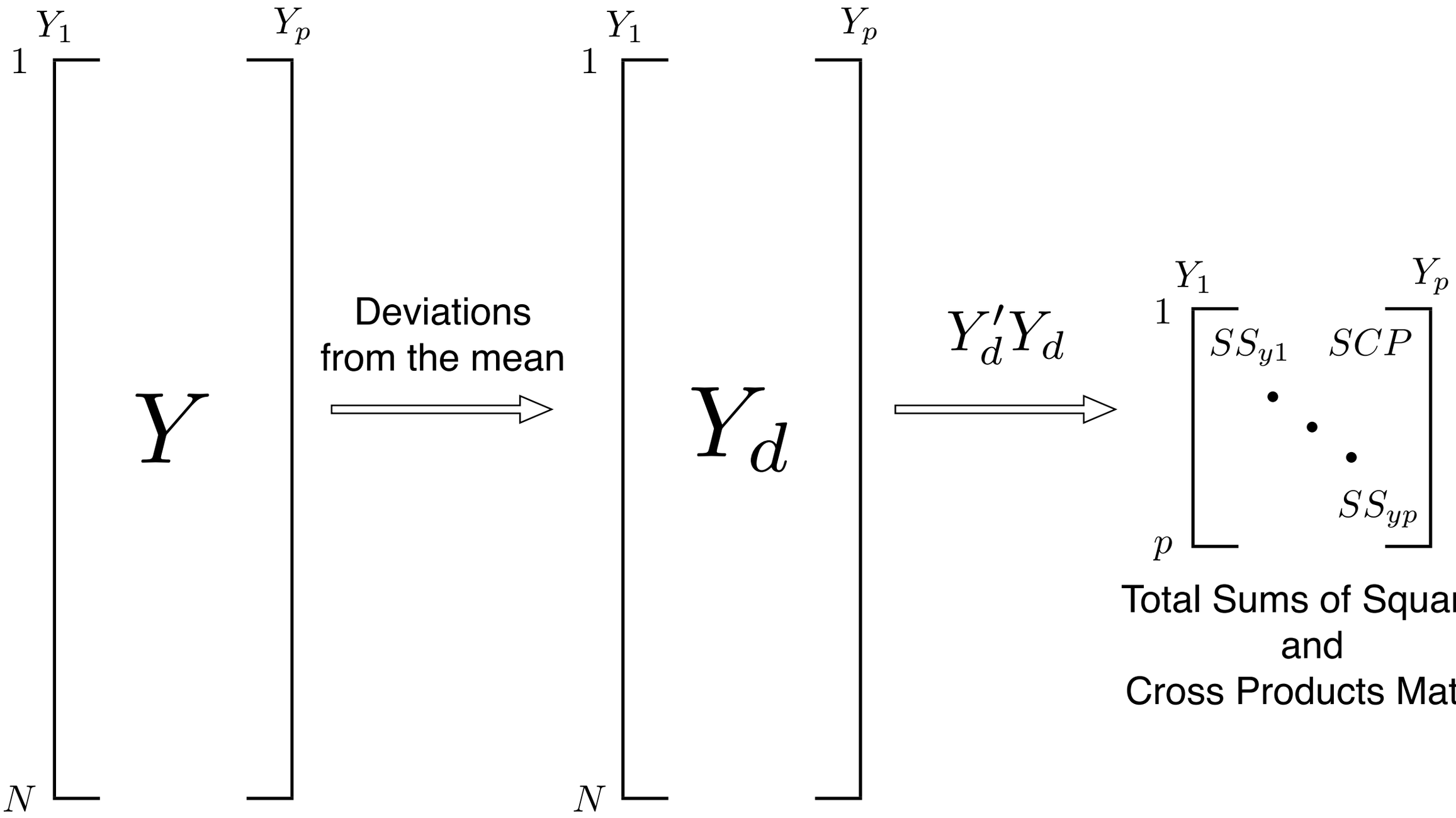


The predictors are used to create a linear composite

To summarise so far…

- 2 group discriminant analysis involves producing a linear composite that distinguishes between the groups.
- Finding the *best* linear composite is a complex process.
- The best linear composite
  - explains the most between-groups variance.
  - maximises the differences between the groups.
  - maximises the t-value or F-value on the linear composite.
  - maximises the ratio of $\frac{SS_{between\ on\ C_1}}{SS_{within\ on\ C_1}}$
- So we want to find a set of weights which maximises the value of $\frac{SS_{between\ on\ C_1}}{SS_{within\ on\ C_1}}$
  - This involves partitioning the variance of the linear composite, C_1 which can be represented in matrix notation.

Calculating the Total SSCP from raw data



Partitioning the Total SSCP matrix

$$\mathbf{T} = \begin{array}{c} \\ 1 \\ \\ \\ p \end{array} \begin{array}{c} \begin{array}{cc} Y_1 & \qquad\qquad Y_p \end{array} \\ \begin{bmatrix} SS_{Ty1} & SCP_{Ty1yp} \\ & \ddots \\ & SS_{Typ} \end{bmatrix} \end{array}$$

Total Sums of Squares
and
Cross Products Matrix

$$\mathbf{B} = \begin{array}{c} \\ 1 \\ \\ \\ p \end{array} \begin{array}{c} \begin{array}{cc} Y_1 & \qquad\qquad Y_p \end{array} \\ \begin{bmatrix} SS_{By1} & SCP_{By1yp} \\ & \ddots \\ & SS_{Byp} \end{bmatrix} \end{array}$$

Between Groups Sums of Squares
and
Cross Products Matrix

$$=$$

$$+$$

$$\mathbf{W} = \begin{array}{c} \\ 1 \\ \\ \\ p \end{array} \begin{array}{c} \begin{array}{cc} Y_1 & \qquad\qquad Y_p \end{array} \\ \begin{bmatrix} SS_{Wy1} & SCP_{Wy1yp} \\ & \ddots \\ & SS_{Wyp} \end{bmatrix} \end{array}$$

Within Groups Sums of Squares
and
Cross Products Matrix

In matrix notation...

$\mathbf{v} =$ the weights used to form the linear composite (a vector of weights).

- also known as the *eigenvector*.

Applying weights to both the between and within groups variability matrices creates a linear composite:

$$SS_{between\,on\,C_1} = \mathbf{v'Bv}$$

$$SS_{within\,on\,C_1} = \mathbf{v'Wv}$$

so now we are looking for a vector of weights that maximises the value of:

$$\frac{SS_{between\,on\,C_1}}{SS_{within\,on\,C_1}} = \frac{\mathbf{v'Bv}}{\mathbf{v'Wv}} = \lambda$$

$\lambda =$ the *discriminant ratio* or an *eigenvalue*.

Unknown Eigenvalue

$$\lambda =$$

Unknown Eigenvectors

$$\mathbf{v}'$$

$$\begin{array}{c} \\ 1 \\ \\ \\ \\ p \end{array} \begin{array}{c} \begin{array}{cc} Y_1 & \qquad\qquad Y_p \end{array} \\ \begin{bmatrix} SS_{By1} & SCP_{By1yp} \\ & \ddots \\ & & SS_{Byp} \end{bmatrix} \end{array}$$

Between Groups Sums of Squares
and
Cross Products Matrix

Unknown Eigenvectors

$$\mathbf{v}$$

Unknown Eigenvectors

$$\mathbf{v}'$$

$$\begin{array}{c} \\ 1 \\ \\ \\ \\ p \end{array} \begin{array}{c} \begin{array}{cc} Y_1 & \qquad\qquad Y_p \end{array} \\ \begin{bmatrix} SS_{Wy1} & SCP_{Wy1yp} \\ & \ddots \\ & & SS_{Wyp} \end{bmatrix} \end{array}$$

Within Groups Sums of Squares
and
Cross Products Matrix

Unknown Eigenvectors

$$\mathbf{v}$$

Unknown Eigenvalue

$$\lambda = \frac{\text{Unknown Eigenvectors } \mathbf{v}' \quad \begin{array}{c} \\ 1 \\ \\ p \end{array} \overset{\begin{array}{cc} Y_1 & \quad Y_p \end{array}}{\begin{bmatrix} SS_{By1} & SCP_{By1yp} \\ \cdot & \\ & \cdot \\ & SS_{Byp} \end{bmatrix}} \quad \text{Unknown Eigenvectors } \mathbf{v}}{\text{Unknown Eigenvectors } \mathbf{v}' \quad \begin{array}{c} \\ 1 \\ \\ p \end{array} \overset{\begin{array}{cc} Y_1 & \quad Y_p \end{array}}{\begin{bmatrix} SS_{Wy1} & SCP_{Wy1yp} \\ \cdot & \\ & \cdot \\ & SS_{Wyp} \end{bmatrix}} \quad \text{Unknown Eigenvectors } \mathbf{v}}$$

Between Groups Sums of Squares and Cross Products Matrix

Within Groups Sums of Squares and Cross Products Matrix

The eigenvalue, λ, is a measure of the ratio of between group variability to within group variability.

The weights, $\mathbf{v}$, the *eigenvector*, are called the discriminant function coefficients and are one measure of the relative importance of the discriminant variables to the separation of the groups.

Another example

3 group Discriminant Analysis

Three groups of inmates:

- Group 1 = convicted for murder
- Group 2 = convicted for fraud
- Group 3 = convicted for armed robbery

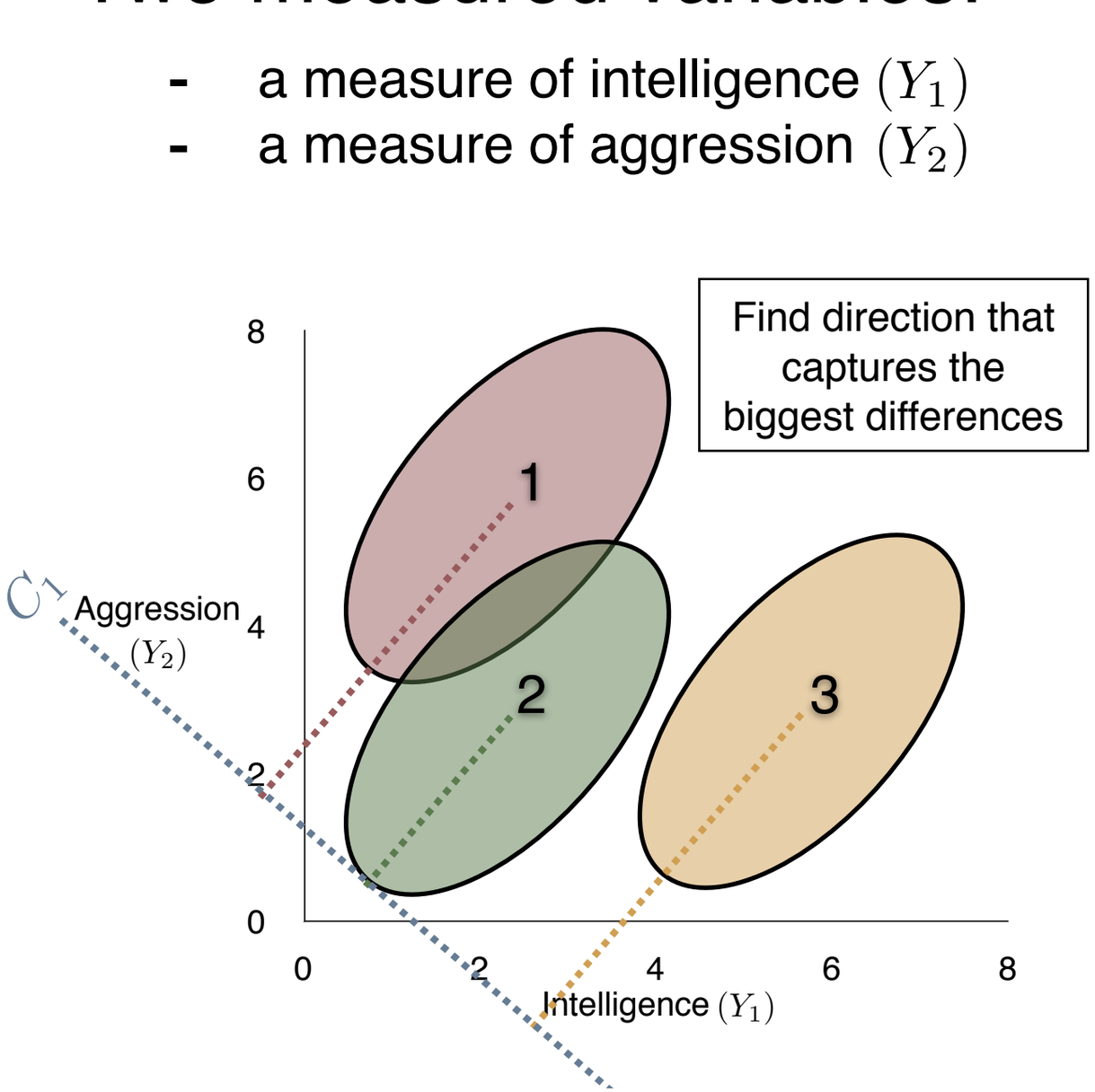
Two measured variables:

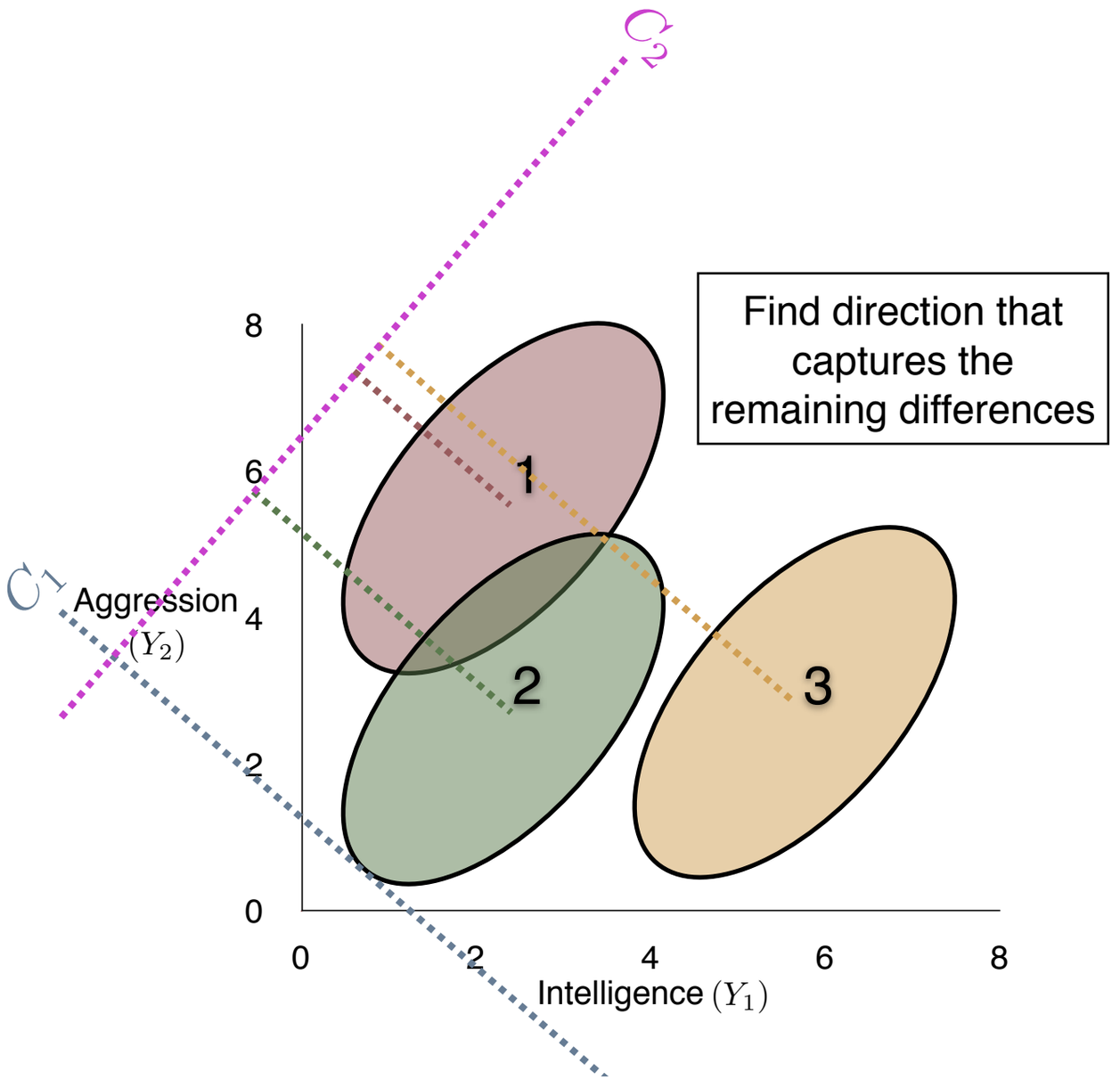
- a measure of intelligence (Y_1)
- a measure of aggression (Y_2)

$$Y_1 Y_2 \quad \leftarrow \quad X$$

2 continuous variables

categorical 3 levels





Another example

3 group Discriminant Analysis

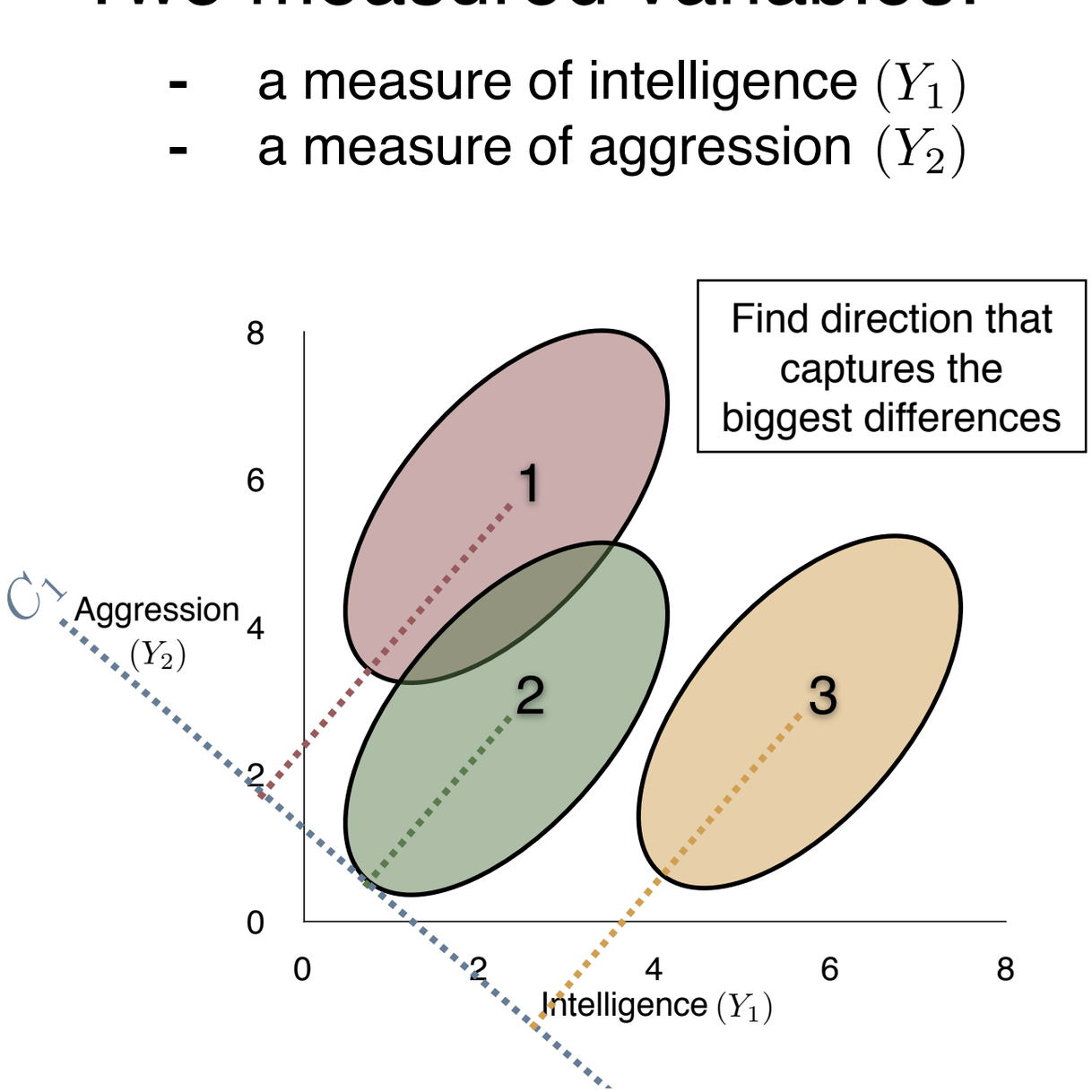
Three groups of inmates:

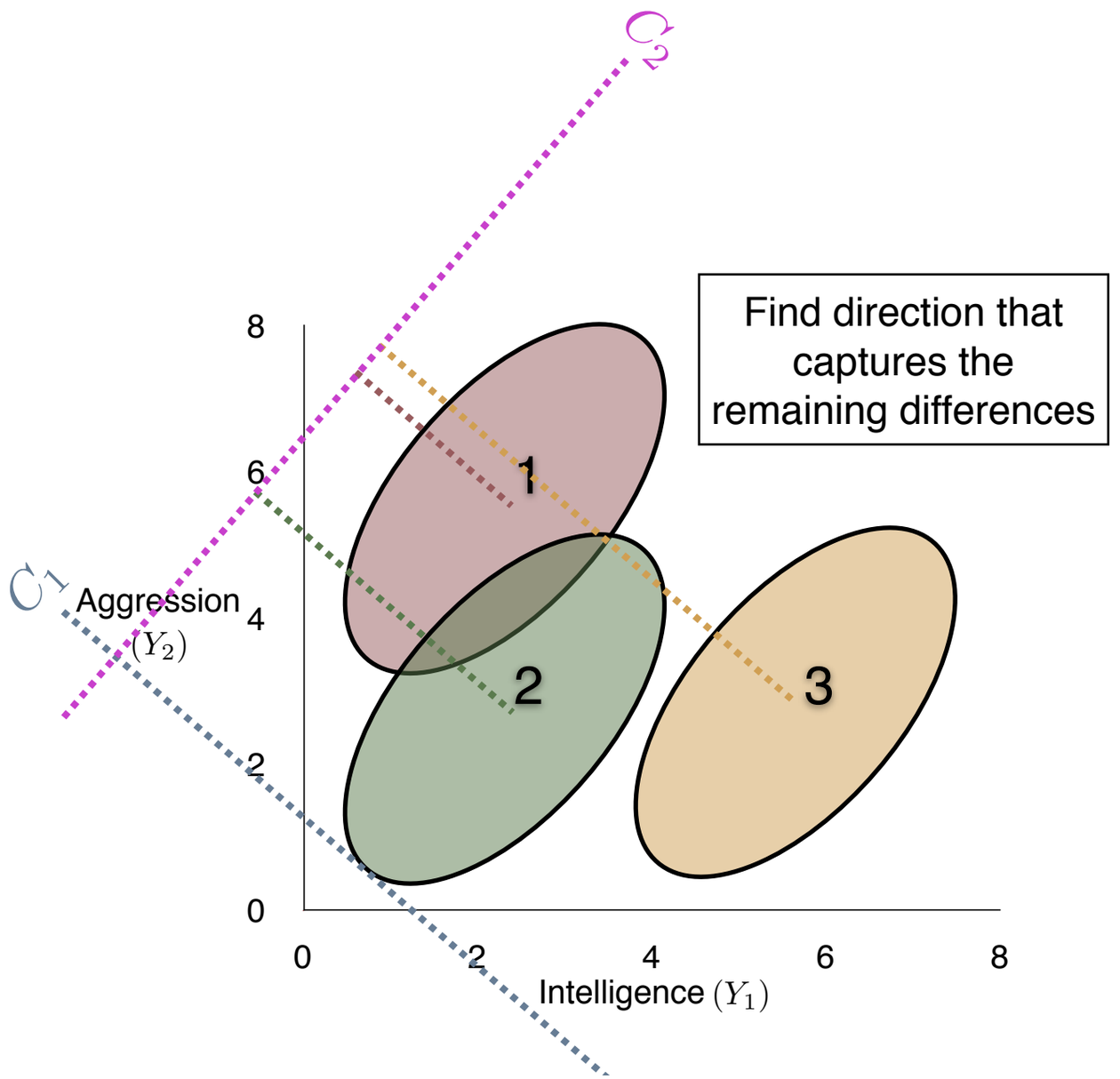
- Group 1 = convicted for murder
- Group 2 = convicted for fraud
- Group 3 = convicted for armed robbery

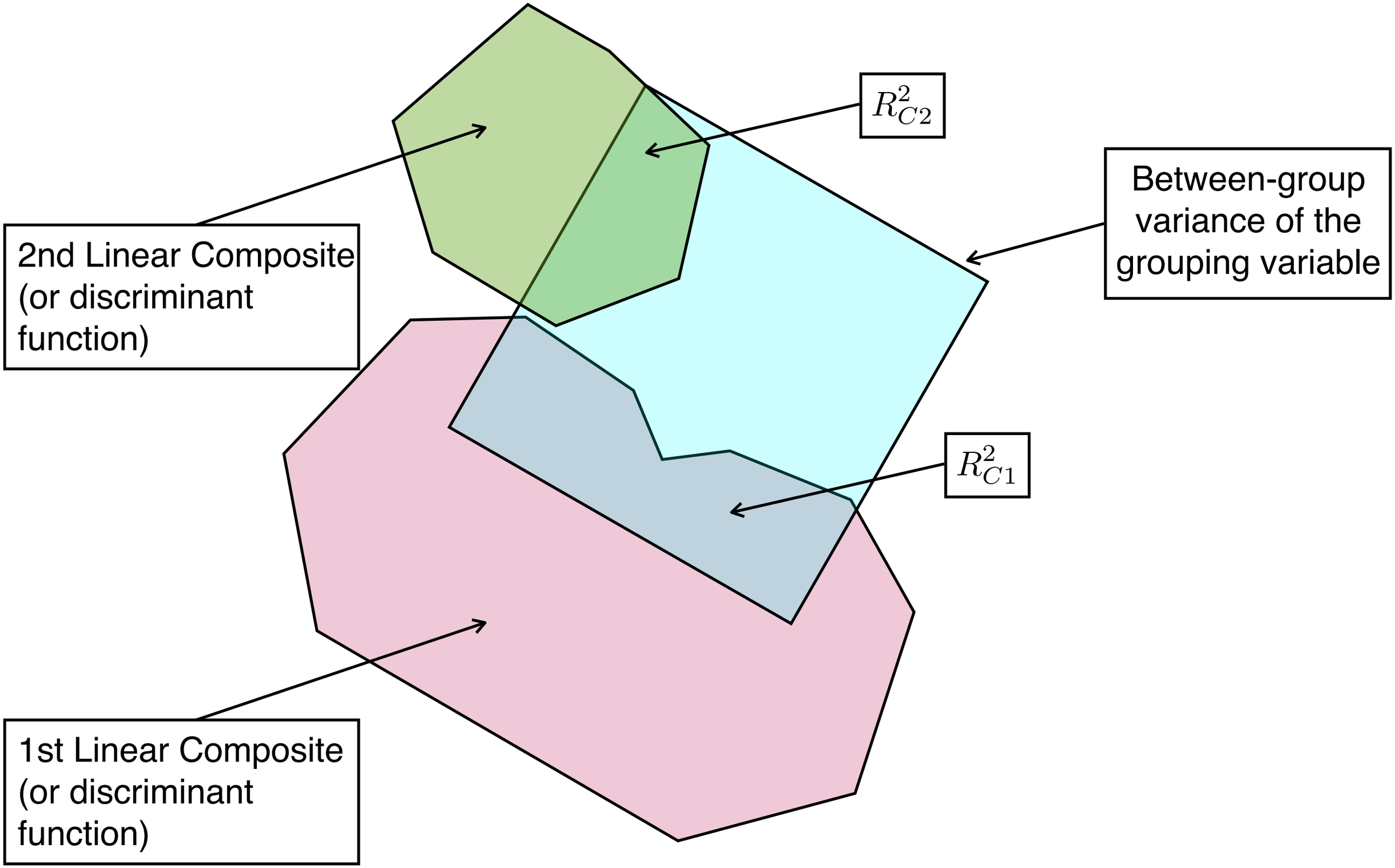
Two measured variables:

- a measure of intelligence (Y_1)
- a measure of aggression (Y_2)

Given *k* groups and *p* discriminating variables, this 'extraction' of discriminant functions continues until the number of discriminant functions is (*k*-1) or *p*, whichever is the smaller.







3 group discriminant analysis

Earlier, we expressed the ratio of the between to within sums of squares in terms of the weights for the linear composite:

$$\lambda = \frac{\overbrace{\begin{bmatrix}\mathbf{v}'\end{bmatrix}}^{\text{Unknown Eigenvectors}} \overbrace{\begin{array}{c} \\ 1 \\ \\ \\ p \end{array}\!\!\begin{array}{c} \begin{array}{cc} Y_1 & \quad Y_p \end{array} \\ \begin{bmatrix} SS_{By1} & SCP_{By1yp} \\ & \ddots \\ & SS_{Byp} \end{bmatrix} \end{array}}^{\text{Between Groups Sums of Squares and Cross Products Matrix}} \overbrace{\begin{bmatrix}\mathbf{v}\end{bmatrix}}^{\text{Unknown Eigenvectors}}}{\underbrace{\begin{bmatrix}\mathbf{v}'\end{bmatrix}}_{\text{Unknown Eigenvectors}} \underbrace{\begin{array}{c} \\ 1 \\ \\ \\ p \end{array}\!\!\begin{array}{c} \begin{array}{cc} Y_1 & \quad Y_p \end{array} \\ \begin{bmatrix} SS_{Wy1} & SCP_{Wy1yp} \\ & \ddots \\ & SS_{Wyp} \end{bmatrix} \end{array}}_{\text{Within Groups Sums of Squares and Cross Products Matrix}} \underbrace{\begin{bmatrix}\mathbf{v}\end{bmatrix}}_{\text{Unknown Eigenvectors}}} = \frac{\mathbf{v}'\mathbf{B}\mathbf{v}}{\mathbf{v}'\mathbf{W}\mathbf{v}}$$

Unknown Eigenvalue

Just knowing this formula does not allow us to find the weights that maximise the ratio.

The goal here is to maximise this discriminant ratio, this eigenvalue λ. How?

We could use trial and error, but the use of calculus will give us an analytic solution...

The equation for the first discriminant function is:

$$(\mathbf{W^{-1}B})\mathbf{v_1} = \lambda_1 \mathbf{v_1}$$

which can be rewritten as

$$(\mathbf{W^{-1}B} - \lambda_1 \mathbf{I})\mathbf{v_1} = \mathbf{0}$$

For the second discriminant function:

$$(\mathbf{W^{-1}B} - \lambda_2 \mathbf{I})\mathbf{v_2} = \mathbf{0}$$

The matrices $\mathbf{W}$ and $\mathbf{B}$ are known. The scalars λ_1 and λ_2 (the first and second eigenvalues) as well as $\mathbf{v_1}$ and $\mathbf{v_2}$ (the first and second eigenvectors) are unknown and need to be calculated. We also fix $\mathbf{v_1}$ and $\mathbf{v_2}$ to be independent of each other by the constraint: $\mathbf{v_1'v_2} = \mathbf{0}$. There is an eigen equation for each discriminant function.

Eigenvectors and eigenvalues are also referred to as *characteristic vectors* (in German, “eigen” means “specific of ” or “characteristic of ”). The set of eigenvalues of a matrix is also called its *spectrum*. (Abdi, 2007)

The eigenvalue (λ)

Why the name ‘eigenvalue’? Consider a simplified form of the eigenvalue equation:

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$$

where $\mathbf{A}$ is a matrix, and λ is a scalar (a number). What if the equation was a scalar equation $a\mathbf{v} = \lambda\mathbf{v}$? This would imply that $a = \lambda$. However, the eigenvalue equation is a matrix equation, so $\mathbf{A} \neq \lambda$.

In this sense, λ has a special relationship with $\mathbf{A}$, as it is often referred to as the ‘singular value’ of the matrix $\mathbf{A}$. The process of finding the eigenvalues of a matrix is called the ‘singular value decomposition’.

The eigenvalue (λ)

In discriminant analysis, we're decomposing the relationship between the grouping variable (*X)* and the discriminant variables (*Y*s).

$$Y_1, Y_2, \ldots Y_p \leftarrow X$$

- An eigenvalue is a measure of concentration of shared variance between the grouping variable and a linear combination of the multiple discriminant variables.
- Multivariate test statistics are all computed as functions of these eigenvalues.
- The eigenvalue can be used to calculate the squared canonical correlation.

Canonical Correlations

The canonical correlation for a discriminant function is useful because it has a meaning just like a multiple correlation in multiple regression. There is one for each discriminant function and can be calculated in a number of ways.

The canonical correlation:

$$R_{Cj} = \sqrt{\frac{\lambda_j}{(1 + \lambda_j)}}$$

is a function of the eigenvalues. That is, it measures the 'concentration of shared variance' in a more interpretable form.

Canonical Correlations

Interpretation

Each discriminant function is related to the grouping variable:

$$Y_1 Y_2 \quad \leftarrow \quad X$$

2 continuous variables — categorical 3 levels

$$C_1 \quad \leftarrow \quad X$$

Composite variable — categorical 3 levels

$$C_1 \quad \leftarrow \quad X_1 X_2 X_3$$

Composite variable — 3 dichotomous variables

This allows a normal regression to be performed.

Canonical Correlations

Interpretation

$$C_1 \quad \leftarrow \quad X_1 X_2 X_3$$

Composite variable

3 dichotomous variables

This allows a normal regression to be performed.

- the R^2 from this regression is the squared canonical correlation.
- R^2_C is the proportion of variance of the discriminant function that's predictable from group membership.
- R^2_C is the proportion of between-group variability accounted for by the discriminant function.

Canonical Correlations

Interpretation

Be sure not to confuse R^2_{Cj} with the '% variance' reported in SPSS.

$$R^2_{Cj}$$

$$\sqrt{\frac{\lambda_j}{(1+\lambda_j)}}$$

How much of the between groups variability is accounted for by that function.

% variance

$$\frac{\lambda_j}{(\sum \lambda_j)}$$

How well one discriminant function discriminates between groups in comparison to the all other discriminant functions in the analysis

Canonical Correlations

Interpretation

- Canonical Correlations explain how each discriminant function performs in the analysis.
- Canonical Correlations do not provide an overall measure of statistical significance like an R in multiple regression.
- Wilk's Lambda (and Bartlett's V) provides measures of overall significance.

Significance testing in Discriminant Analysis

- testing the strength of the overall relationship.
- testing for the number of significant discriminant functions.

Testing the strength of the overall relationship

- Wilk's Lambda (Λ) measures the overall relationship between the grouping variable and the predictors.
- Wilk's Lambda summarises the information from all discriminant functions.
- You can think about the overall expression:

$$\Lambda = \frac{|\mathbf{W}|}{|\mathbf{T}|} = \frac{|\mathbf{W}|}{|\mathbf{W}+\mathbf{B}|} = \prod_{j=1}^{r} \frac{1}{1+\lambda_j} = \prod_{j=1}^{r} \left(1 - R^2_{Cj}\right)$$

- in three different ways:

$|\mathbf{W}|$ and $|\mathbf{T}|$ are the determinants of $\mathbf{W}$ and $\mathbf{T}$.

Testing the strength of the overall relationship

$$\Lambda = \frac{|\mathbf{W}|}{|\mathbf{T}|} = \frac{|\mathbf{W}|}{|\mathbf{W}+\mathbf{B}|} = \prod_{j=1}^{r} \frac{1}{1+\lambda_j} = \prod_{j=1}^{r} \left(1 - R_{Cj}^2\right)$$

1. In terms of within and between group variance.
   - This is similar to the reciprocal of an F value: $\left(\frac{1}{F}\right)$
   - The bigger the effects of differences between groups the smaller the value of Λ.

$|\mathbf{W}|$ and $|\mathbf{T}|$ are the determinants of $\mathbf{W}$ and $\mathbf{T}$.

Testing the strength of the overall relationship

$$\Lambda = \frac{|\mathbf{W}|}{|\mathbf{T}|} = \frac{|\mathbf{W}|}{|\mathbf{W}+\mathbf{B}|} = \prod_{j=1}^{r} \frac{1}{1+\lambda_j} = \prod_{j=1}^{r} \left(1 - R_{Cj}^2\right)$$

2. In terms of eigenvalues (λ)

- λ_j is the 'discriminant ratio' for a discriminant function.
- Λ can be considered as summarising the 'discriminant ratios' for all discriminant functions.

Testing the strength of the overall relationship

$$\Lambda = \frac{|\mathbf{W}|}{|\mathbf{T}|} = \frac{|\mathbf{W}|}{|\mathbf{W}+\mathbf{B}|} = \prod_{j=1}^{r} \frac{1}{1+\lambda_j} = \prod_{j=1}^{r} \left(1 - R_{Cj}^2\right)$$

3. In terms of 1 minus the squared canonical correlations for each function.

- it measures the lack of fit of the linear model.
- Compare with $(1 - R^2)$ from multiple regression.

However, the distribution of Wilk's Lambda is not as friendly or useable as a *t* or *F* distribution. This is because Λ doesn't have a convenient or easily accessible distribution that can be looked up in tables. So Bartlett worked out the transformation that enables the chi-square distribution to be used.

Bartlett's *V*

Testing the strength of the overall relationship

Bartlett's V

$$V = -\left(\frac{N - 1 - (p + k)}{2}\right) \ln(\Lambda)$$

where

$N =$ the total number of cases

$\ln =$ the natural log

V is distributed approximately χ^2

with $p(k-1)$ degrees of freedom.

In the MANOVA and GLM procedures, SPSS give the results of other multivariate tests of significance. These are Roy's θ, Pillai's V, and Hotelling's T. Haase and Ellis (1987) state that Roy's is the most powerful when there is a very strong first discriminant function. If the 'variance' is diffused across the discriminant functions, then the other three, including Wilk's Λ, are about equally powerful.

Testing for the number of significant discriminant functions

- The purpose is to find the smallest number of functions that adequately describe differences between the groups.
- A Bartlett's V is associated with each discriminant function.
- The components corresponding to the first, second, etc. functions are subtracted from V and the remainder is tested for significance.
- As soon as the remainder, after removing the first *s* functions, becomes 'nonsignificant' at some alpha level, then we conclude that only the first *s* functions are significant.

Relative importance of the individual predictor variables

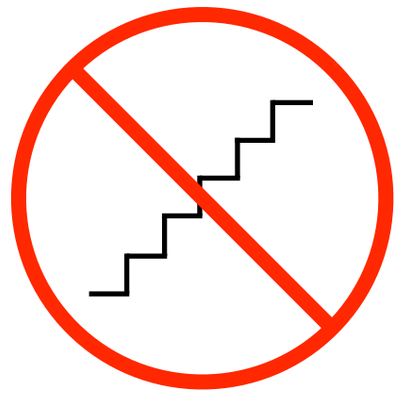
- What are the available measures of the importance of individual predictors?
- Like multiple regression this is not an easy question to answer because there are many different statistics suggested.
- In this course we will consider five of them:
  - Univariate F-Ratios
  - F-TO-REMOVE statistics
  - Structure Coefficients
  - Standardised Discriminant Function Coefficients
  - Relative Weights

These will be defined later and will be discussed in the context of an interpretation of a discriminant analysis.

Stepwise Methods

- Like multiple regression, there are several types of discriminant analysis:
  - Direct
  - Stepwise
  - Hierarchical

  Carl Huberty, in a talk to the Psychology Department, stated emphatically that stepwise methods were not generally appropriate (he offered the following diagram to be displayed). We cover exclusively the direct or standard method. The hierarchical method is analogous to hierarchical or sequential multiple regression. Thompson (1995) reiterates this theme offering very good reasons for NOT using stepwise methods.



NO Stepwise Methods

Summary

$Y_1, Y_2, \ldots Y_p$ ← X

p continuous variables

categorical
k levels

⇐ Descriptive Discriminant Analysis (MANOVA)

Discriminant Analysis describes the effects of some grouping variable (with k levels) on a set of *p* continuous discriminant (response) variables.

In our motivational example, we were looking at whether we could describe the difference between convicted murderers and fraudsters on the basis of personality measures aimed at intelligence and aggression?

We're interested in two research questions:

1. Is the overall relationship statistically significant and how strong is the relationship?
2. What variables are individually important in separating (discriminating) between the groups?

Summary

Discriminant Analysis involves producing a linear composite that distinguishes between the groups.

These composites are created so that the between group (*k* groups) means on these composites are as different as possible.

Eigenvalues (λ) are a measure of how well each discriminant function is able to maximally separate the groups.

Eigenvectors $(\mathbf{v})$ are the weights applied to the variables to create these linear composites.

The Squared Canonical Correlation (R^2_{Cj}) estimates the proportion of between group variability accounted for by the jth discriminant function.

Wilk's Lambda (Λ) provides us with an overall test and can be considered as summarising the 'discriminant ratios' for all discriminant functions.

Bartlett's *V* provides us with a transformation of Wilk's Lambda that uses a χ^2 distribution.