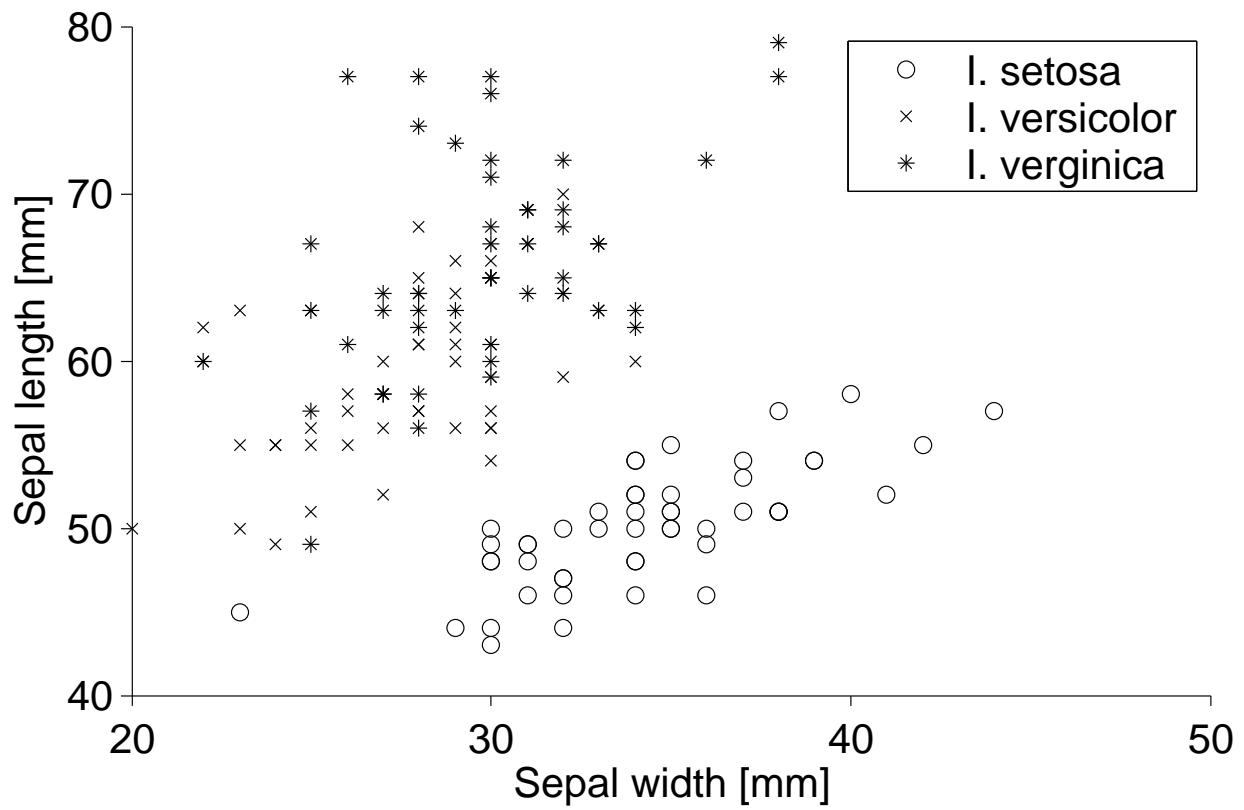
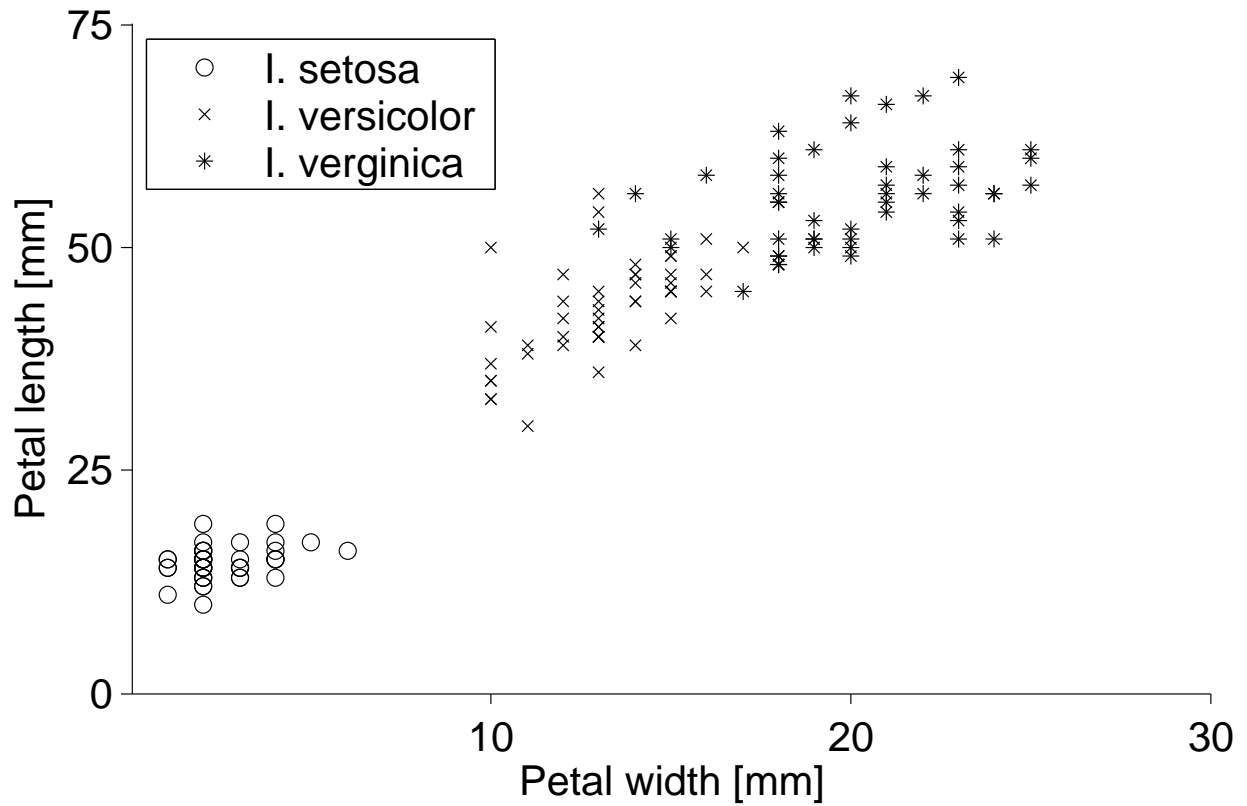


## Fisher's Iris Data (cont.)



Three species of Iris ( $g = 3$ )



Four variables ( $p = 4$ ):

- sepal length (SL)
- sepal width (SW)
- petal length (PL)
- petal width (PW)

## Discriminant Analysis

Discriminant analysis of log-transformed variables yields canonical variate

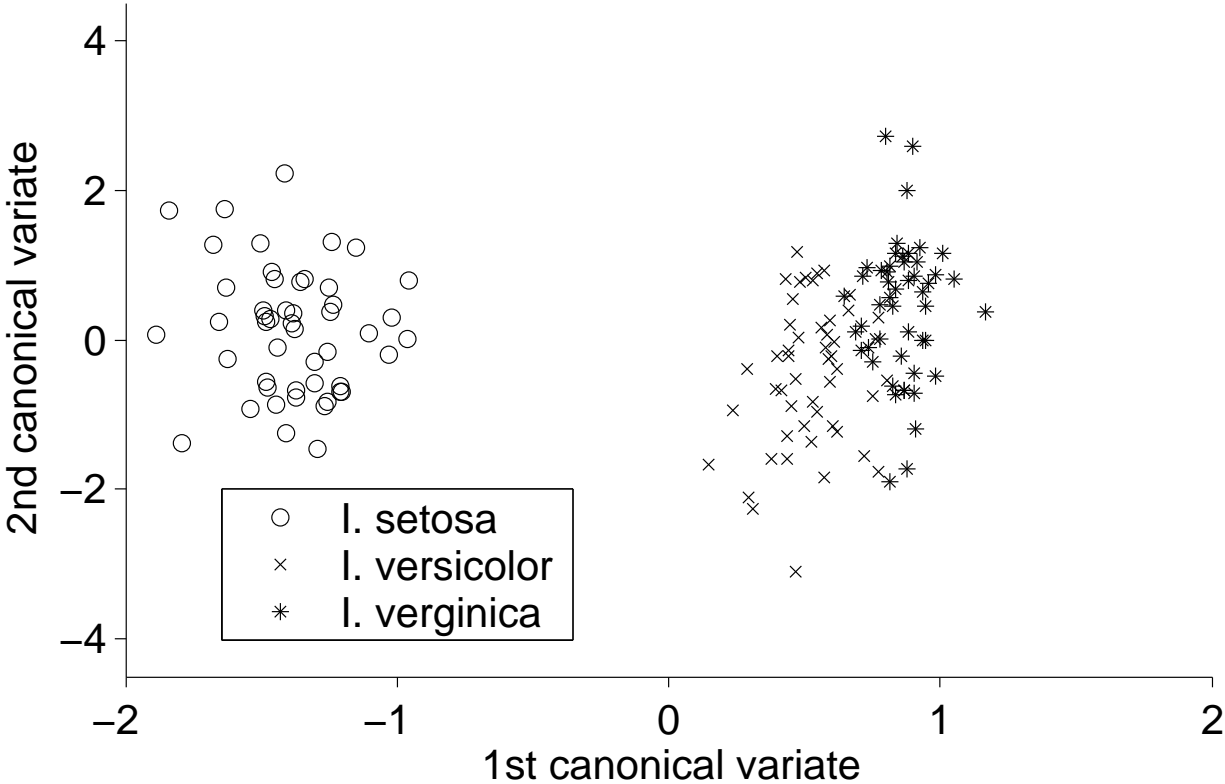
$$z_1 = 0.25 \log(\text{PW}) + 1.32 \log(\text{PL}) \\ - 0.58 \log(\text{SW}) - 0.58 \log(\text{SL}).$$

Ratio  $\gamma_i = (u_i^T \Sigma_a u_i) / (u_i^T \Sigma_w u_i)$  of among-group to within-group variance:

$$\gamma_1 = 3012.6$$

$$\gamma_2 = 11.7$$

# Scatterplot canonical variates



## Classification

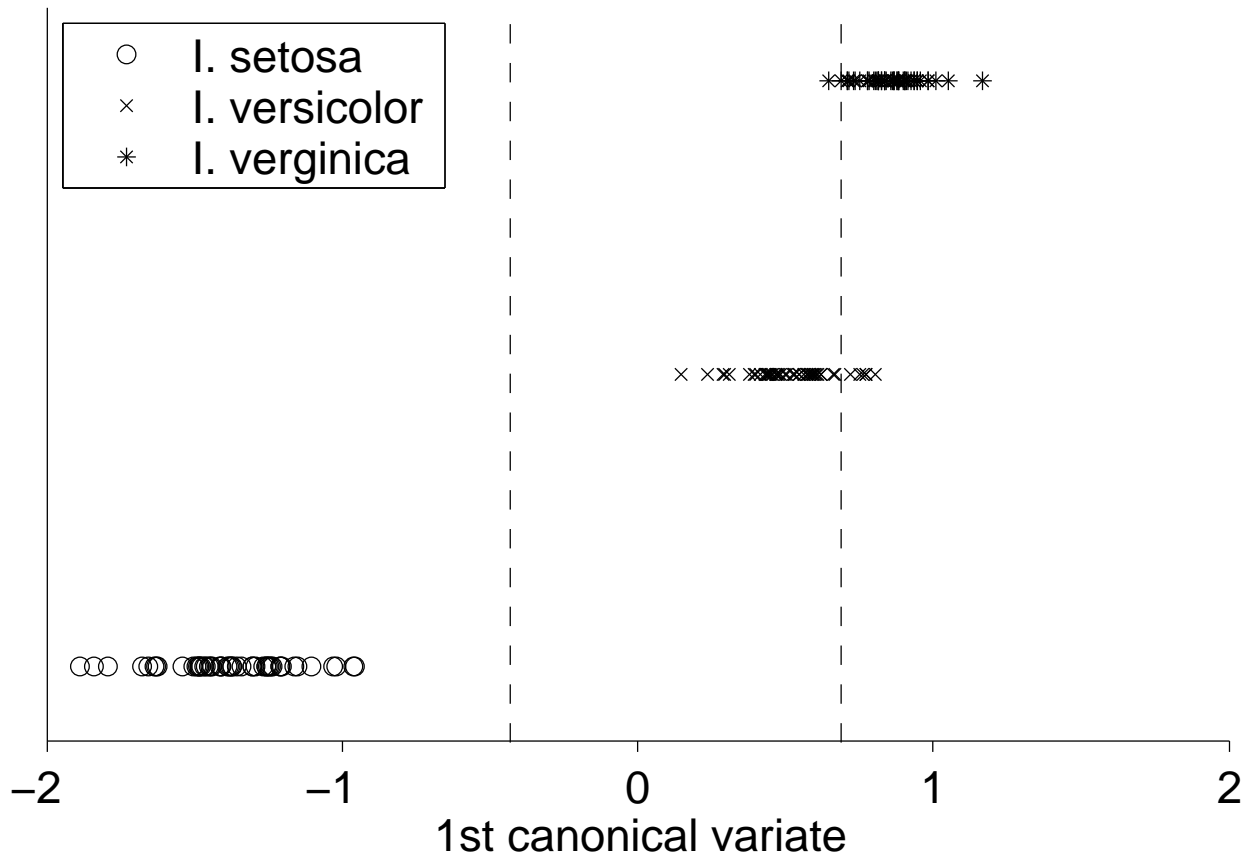
Data suggest to use only first canonical variate.

Classification rule: allocate an observation  $x$  ( $1 \times p$  vector) to the group  $i$  for which

$$\|(x - \mu_i)u_1\|^2$$

is smallest.

# Decision Boundaries



## Confusion Matrix

Rows: actual species; columns: decision

	I. setosa	I. versicolor	I. virginica
I. setosa	50	0	0
I. versicolor	0	44	6
I. virginica	0	1	49