Discriminant analysis (a type of regression)

1. **Step:** Preprocessing: Transform the nonlinear world of benefits for illness to a linear framework

2. **Step:** Discriminant analysis
   - Build clusters of values (groups) of the random variable $Y$ (cost of "linearized" illness), that is to be predicted
   - The random variables $X_i$ (representing various diseases, $i=1,...,15,000$) will predict $Y$
   - Modelling the total (linearized) cost of illness using "discriminant parameters $b_i$":
     \[ Y_D = b_0 + b_1X_1 + b_2X_2 + \ldots + b_mX_m \] (Discriminant function)
   - Determine $b_i$ so, such that:
     - the distance of the mean values of the groups is maximal
     - the variance within the groups is minimal
   - Select adequate predictors $X_i$ such that the loss of information is minimal.

3. **Step:** Postprocessing Return to the nonlinear world of cost of illness by adjusting functions

**Goals of the selection of coefficients**
- large distance between the "centroids" (mean value of the considered group)
- small variance within groups
- large variance between groups
- metric is the Mahalanobis distance

![Graph showing distance between centroids for discriminant analysis](image-url)
DECISIONS by DISCRIMINANT FUNCTION

- **$Y_D = b_0 + b_1X_1 + b_2X_2 + \ldots + b_mX_m$** (Discriminant function)
- evaluate discriminant function for known customers with special values of the predictors $X_i$
- the discriminant functions’ values are assigned to points on the discriminant axis.
- compare the values with the group-centroids.

<table>
<thead>
<tr>
<th>Group A</th>
<th>Group B</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_A$</td>
<td>$Y_B$</td>
</tr>
<tr>
<td>$\bar{Y}_A, \bar{Y}_B$ = Group-centroids = mean value of discriminant function $Y_D$ for the group A or B respectively.</td>
<td></td>
</tr>
<tr>
<td>$Y^\ast$ = criterion of separation</td>
<td></td>
</tr>
</tbody>
</table>

- evaluate $Y_D$ for new customers with special values of the predictors $X_i$ according to the value of separation $Y^\ast$ of the group A or B

SELECTION of SIGNIFICANT PREDICTORS for DISCRIMINANT ANALYSIS

- selection of significant predictors is realized step by step.
  - select an additional variable in every step, such that discrimination is optimized between groups with the smallest (Mahalanobis-) distance ($F$ to enter)
  - if variables contribute only insignificantly to discrimination in the current multivariate system, they can be removed in a later step
- measure the quality of discrimination by Wilks Lambda $\lambda$:
  $$\lambda = \frac{\text{non explained scattering}}{\text{total scattering}}$$
- the level of significance of the discriminant analysis results from $\lambda$ as a $\chi^2$-distributed variable.
  $$\chi^2 = \left( N - J \frac{ \hat{G} }{ 2 } - 1 \right) \ln(\lambda)$$
  $N$ = number of customers in the sample, $J$ = number of predictors, $G$ = number of groups
- the level of significance is better (lower) than $\alpha = 0.001$ for the present study
DISCRIMINANCE BETWEEN 2 GROUPS

Group 1 = no benefit paid  Group 2 = benefit paid  i. e. G = 2  (2 groups)

\[ X_{kl}^{(i)} \] = random variable, describing the predictor number \( k \) of the person number \( l \),
who belongs to the with benefits \( (i) \), \( l=1...n_i \); \( n = n_1 + n_2 \)

\[ Y_{l}^{(i)} \] = random variable, describing the "probability to benefits" of person \( l \) belonging to group \( i \)

\[ Y_{1}^{(1)} := b_1X_{1l}^{(1)} + b_2X_{2l}^{(1)} + ... + b_mX_{ml}^{(1)} \]  \( \text{group (1)} \)

\[ Y_{1}^{(2)} := b_1X_{1l}^{(2)} + b_2X_{2l}^{(2)} + ... + b_mX_{ml}^{(2)} \]  \( \text{group (2)} \)

Group centroids:

\[ Y_1 := \frac{1}{n_1} \sum_{l=1}^{n_1} Y_{l}^{(1)} \]

\[ Y_2 := \frac{1}{n_2} \sum_{l=1}^{n_2} Y_{l}^{(2)} \]

\[ X_k^{(i)} = \text{mean value of the predictor } k \text{ in group } i \]

CRITERION of DISCRIMINANCE

\[ S = S(b_1, ..., b_m) = Y_1 - Y_2 = \text{distance of the centre of the group} \] = \text{max!}

\[ T = T(b_1, ..., b_m) = \sum_{l=1}^{n_1} (Y_{l}^{(1)} - Y_1)^2 + \sum_{l=1}^{n_2} (Y_{l}^{(2)} - Y_2)^2 = \text{variance between groups} \] = \text{min!}

\[ Q = \frac{S^2}{T} \] = \text{max!} \quad \Rightarrow \quad \frac{\partial Q}{\partial b_k} = 0 \quad \text{under minor condition} \quad S^2 = 1

Solution \( \hat{b}_1, ..., \hat{b}_m \) is inserted in \( S \)

\[ S(\hat{b}_1, ..., \hat{b}_m) = (n - g) \sum_{i,j=1}^{m} w_{ij} (X_{i1}^{(1)} - \bar{X}_i^{(1)})(X_{j2}^{(2)} - \bar{X}_j^{(2)}) \]

\( w_{ij} \) = elements of inversion of the co variance-matrix (within-groups)

\[ D^2 = S \] is called Mahalanobis - distance

© 2005 RISK-CONSULTING Prof. Dr. Weyer GmbH
Criterion of selection of significant predictors

• not all predictors $X_i$ are considered immediately.
  search for predictors, that separate the groups optimally

• for that purpose: calculate $S$ depending only on some of the $b_i$

• first select such $X_i$ (or $b_i$ respectively), maximizing the distance $S$ between groups, that are nearest

• $F = \frac{(n - 1 - m) m n_1 n_2}{m (n - 2) (n_1 + n_2)} S$ is F-distributed.