Stockholms universitet

STATISTISKA INSTITUTIONEN HT-2014

Written exam in Multivariate Methods, 7.5 ECTS credits
Thursday, 26th November 2014, 9:00 – 14:00
Time allowed: FIVE hours
Examination Hall: Brunnsvikssalen

You are required to answer all **7 (seven)** questions as well as motivate your solutions. The total amount of points is 80. In order to pass this part, you need to get at least 40 points. Points from this exam will be added to your results from the home assignment. The final grades are assigned as follows: **A** (91+), **B** (75-90), **C** (66-74), **D** (58-65), **E** (50-57), **Fx** (30-49), and **F** (0-29).

You are **allowed** to use a pocket calculator, a language dictionary, and a list of formulas (attached). In addition, you are **allowed** to use a one-sided A4 containing your own formulae, but excluding proofs and solutions. The A4 must be approved (signed) by the teacher; and, it must be submitted along with your solutions. If the A4 is not signed by the teacher and discovered, student might be accused in cheating on exam.

The teacher reserves the right to examine the students **orally** on the questions in this examination.

1. (10 points) The coordinates of three points with respect to the orthogonal basis vectors $\mathbf{e}_1$ and $\mathbf{e}_2$ are as follows:

   A = (2,-1) B = (1.5, -0.15) C = (-3, 7.5)

   Show that A, B and C lie on a straight line.
   The set of oblique basis vectors f_1 and f_2 are related to vectors $\mathbf{e}_1$ and $\mathbf{e}_2$ as follows:

   $$\mathbf{e}_1 = .804\mathbf{f}_1 + .629\mathbf{f}_2$$
   $$\mathbf{e}_2 = .273\mathbf{f}_1 - .588\mathbf{f}_2$$

   Compute the coordinates of A, B and C with respect to $\mathbf{f}_1$ and $\mathbf{f}_2$. Do they still lie on a straight line? Justify your answer.

2. (12 points) Use the data in the contingency Table below to answer the following questions:

Blood Cholesterol (mg/100 cc)	Heart Disease Present	Heart Disease Absent
<200	6	5
200-219	10	6
220-259	30	5
>259	45	7

a. What is the probability that
   (i) Heart disease is present?
   (ii) Blood cholesterol is less than 219 mg/100 cc?

b. What are the odds that

   (i) Heart disease is present given that blood cholesterol is less than 200 mg/100 cc?

   (ii) Blood cholesterol is greater than 219 mg/100 cc given that heart disease is present?

c. Compute the log of odds that

   (i) Heart disease is present given that blood cholesterol is greater than 259 mg/100 cc.

   (ii) Heart disease is present given that blood cholesterol is between 200 and 219 mg/100 cc.

3. (12 points) Consider the two-indicator two-factor model represented by the following equations:

$$X_1 = 0.104F_1 + 0.824F_2 + U_1$$
$$X_2 = 0.065F_1 + 0.959F_2 + U_2$$
$$X_3 = 0.065F_1 + 0.725F_2 + U_3$$
$$X_4 = 0.906F_1 + 0.134F_2 + U_4$$
$$X_5 = 0.977F_1 + 0.116F_2 + U_5$$
$$X_6 = 0.827F_1 + 0.016F_2 + U_6$$

The usual assumptions hold for the above model. Answer the following questions assuming that the correlation between the common factors F_1 and F_2 is given by
Corr(F_1, F_2)= Φ_{12} = -0.3

(a) What are the pattern loadings of indicators X_1, X_4 and X_6 on the factors F_1 and F_2?

(b) Compute the correlation between the indicators X_1 and X_2.

(c) What percentage of the variance of indicators X_1 and X_2 is not accounted for by the common factors F_1 and F_2?

4. (10 points) Perform principal component analysis on the following data by hand. In other words, determine the angle (with best precision you can: 1%, 5%, 10%: your calculators should be of help) between the new axis and the old axis that would give a new variable, which accounts for the maximum variance in the data. What conclusions can you draw?

X_1	X_2
1	4
1	1
2	2
2	3
3	2
3	2
4	1
4	4

5. (15 points) Consider the following single-factor model

$$x_1 = \lambda_1 \xi + \delta_1$$
$$x_2 = \lambda_2 \xi + \delta_2$$
$$x_3 = \lambda_3 \xi + \delta_3$$

If the sample covariance matrix of the indicators is given by:

$$S = \begin{pmatrix} 1.20 & 0.93 & -0.45 \\ 0.93 & 1.56 & 0.27 \\ -0.45 & 0.27 & 2.15 \end{pmatrix}$$

Compute the estimates of the model parameters ($\lambda_1, \lambda_2, \lambda_3, Var(\delta_1), Var(\delta_2), Var(\delta_3)$) using hand calculations. Are the parameter estimates unique?

6. (15 points) The correlation matrix for a hypothetical data set is given in the following table:

	X_1	X_2	X_3	X_4
X_1	1.000			
X_2	0.690	1.000		
X_3	0.280	0.255	1.000	
X_4	0.350	0.195	0.610	1.000

The following estimated factor loadings were extracted by the principal axis factoring procedure:

Variable	F_1	F_2
X_1	0.80	0.20
X_2	0.70	0.15
X_3	0.10	0.90
X_4	0.20	0.70

Compute and discuss the following: (a) specific variances; (b) communalities; (c) proportion of variance explained by each factor; (d) estimated or reproduced correlation matrix; and (e) residual matrix.

7. (6 points) Describe in detail and compare hierarchical and non-hierarchical clustering methods. Do they require a priori knowledge of the number of clusters? Discuss weaknesses and strengths of each of the methods.

MULTIVARIATE METHODS

PCA

$\xi_1 = W_{11}X_1 + W_{12}X_2 + \ldots + W_{1p}X_p$

$\xi_2 = W_{21}X_1 + W_{22}X_2 + \ldots + W_{2p}X_p$

$\vdots$

$\xi_p = W_{p1}X_1 + W_{p2}X_2 + \ldots + W_{pp}X_p$

- The weights: $W_{i1}^2 + W_{i2}^2 + \ldots + W_{ip}^2 = 1 \quad , i=1,\ldots,p$

  $W_{i1}W_{j1} + W_{i2}W_{j2} + \ldots + W_{ip}W_{jp} = 0 \quad \forall i \neq j$

- Characteristic equation: $\det(\lambda I - A) = 0$

- Loadings = correlation between the original and the new variables

  $l_{ij} = \frac{W_{ij}}{s_j}\sqrt{\lambda_i}$

FA

$X_1 = \lambda_{11}\xi_1 + \lambda_{12}\xi_2 + \ldots + \lambda_{1m}\xi_m + \varepsilon_1$

$X_2 = \lambda_{12}\xi_1 + \lambda_{22}\xi_2 + \ldots + \lambda_{2m}\xi_m + \varepsilon_2$

$\vdots$

$X_p = \lambda_{p1}\xi_1 + \lambda_{p2}\xi_2 + \ldots + \lambda_{pm}\xi_m + \varepsilon_p$

- Assumptions:
  1) Means of indicators, common factors and unique factors are zero.
  2) Variances of indicators and common factors are one.
  3) The unique factors are not correlated among themselves or with the common factors. $\Rightarrow E(\xi_i\varepsilon_j) = E(\varepsilon_i\varepsilon_j) = 0 \quad \forall i \neq j$

Model	Equations	The variance of any indicator x_j	Structure loading	Shared variance = (structure loading)^2	Correlation between indicators
One-factor model	$X_1 = \lambda_1\xi + e_1$ ⋮ $X_p = \lambda_p\xi + \varepsilon_p$	$V(x_j) = \lambda_j^2 + V(\varepsilon_j)$	$Cor(x_j, \xi) = \lambda_j$	λ_j^2	$Corr(x_j, x_k) = \lambda_j\lambda_k$
Two-factor model	$X_1 = \lambda_{11}\xi_1 + \lambda_{12}\xi_2 + \varepsilon_1$ ⋮ $X_p = \lambda_{p1}\xi_1 + \lambda_{p2}\xi_2 + \varepsilon_p$	$V(x_j) = \lambda_{j1}^2 + \lambda_{j2}^2 + 2\lambda_{j1}\lambda_{j2} + V(\varepsilon_j)$	$Cor(x_j, \xi_2) = \lambda_{j2} + \lambda_{j1}\phi$ $Cor(x_j, \xi_1) = \lambda_{j1} + \lambda_{j2}\phi$	$(\lambda_{j2} + \lambda_{j1}\phi)^2$ $(\lambda_{j1} + \lambda_{j2}\phi)^2$	$Cor(x_j, x_k) = \lambda_{j1}\lambda_{k1} + \lambda_{j2}\lambda_{k2} + [\lambda_{j1}\lambda_{k2} + \lambda_{j2}\lambda_{k1}]\phi$

$Cor(\xi_1, \xi_2) = \phi$

CFA

Underidentified model: # equations < # variables

Just-identified model: # equations = # variables

Overidentified model: # equations > # variables

$P(x_1, x_2, x_3) = P(x_1) \cdot P(x_2|x_1) \cdot P(x_3|x_1, x_2)$

CA

$D_{AB}^2 = \sum_{j=1}^{p}(a_j - b_j)^2 \qquad SD_{ik}^2 = \sum_{j=1}^{p}\left(\frac{x_{ij} - x_{kj}}{s_j}\right)^2 \qquad MD_{ik}^2 = \frac{1}{1-r^2}\left[\frac{(x_{i1}-x_{k1})^2}{s_1^2} + \frac{(x_{i2}-x_{k2})^2}{s_2^2} - \frac{2r(x_{i1}-x_{k1})(x_{i2}-x_{k2})}{s_1 s_2}\right]$

DA (two groups)

$\lambda = \frac{SS_B}{SS_W} \rightarrow max$

If $\xi = \bar{X}^T\bar{\gamma} \Rightarrow$ The estimation of $\bar{\gamma}$: $\bar{\gamma}^T = (\bar{\mu}_1 - \bar{\mu}_2)^T \Sigma^{-1}$

$SSCP_W = SSCP_1 + SSCP_2$

$SSCP_T = SSCP_W + SSCP_B$

LOG-REG

$odds = \frac{p}{1-p} \qquad p = \frac{odds}{1+odds}$

$\ln\frac{p}{1-p} = \beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k$

$P(Y=1) = p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k)}}$

If $Y = \beta_0 + \beta_1 X$:

$\hat{\beta}_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} \quad , \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X}$

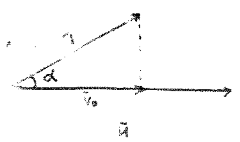
- Rotation $\Rightarrow$ $a_1^* = \cos\theta a_1 + \sin\theta a_2$

  $a_2^* = -\sin\theta a_1 + \cos\theta a_2$

$S^2 = \frac{SS}{df} = \frac{\sum(x_i - \bar{x})^2}{n-1} = \frac{\sum x_i^2 - n\bar{x}^2}{n-1}$

$S_{xy} = \frac{SCP}{df} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n-1} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{n-1}$

- Projection: of $\bar{v}$ onto $\bar{u}$



- The projection vector: $\bar{v}_p = \frac{\|\bar{v}_p\|}{\|\bar{u}\|}\bar{u}$
- $\|\bar{v}_p\| = \|\bar{v}\|\cos\alpha = \frac{\bar{v} \cdot \bar{u}}{\|\bar{u}\|}$

- Direction cosines = The cosines of the angle between a vector and the axes

Formula Sheet, Multivariate Methods

Matrices

Transpose – exchange rows and columns
Identity (I) – diag (1,1...) of order n*n
Inverse of A (A^{-1}): $AA^{-1} = A^{-1}A = I$
$A + B = B + A; x(A + B) = xA + xB; AB \neq BA$ $(in\ general)$;
If order (A)=m*n, order (B)=n*p, then C=AB is of order m*p

$$D=\det A=\begin{vmatrix} a_{11} & \dots & a_{1n} \\ \dots & \dots & \dots \\ a_{n1} & \dots & a_{nn} \end{vmatrix}$$

detA=$a_{i1}A_{i1} + a_{i2}A_{i2} + \dots + a_{in}A_{in}$ where cofactor $A_{ij} = (-1)^{i+j}D_{ij}$ (i-row, j-column of D)
Cramer's rule: $x_j = D_j/D$ where D=detA and D_j is the determinant that arises when the j column of D is replaced by the column elements $b_1, \dots, b_n$. (**Ax=b**)

Vectors

$\boldsymbol{a} = (a_1 a_2 \dots a_p)$
A right-angle triangle: α - angle between a and c; c – hypotenuse; $\cos\alpha = \frac{a}{c}$, $\sin\alpha = \frac{b}{c}$
Length of vector **a**=$\|\boldsymbol{a}\| = \sqrt{a_1^2 + a_2^2}$
Basis vectors $\boldsymbol{e_1} = (1\ 0), \boldsymbol{e_2} = (0\ 1)$
$\boldsymbol{a} = a_1\boldsymbol{e_1} + a_2\boldsymbol{e_2}$
Scalar product $\boldsymbol{ab} = a_1b_1 + a_2b_2 + \dots + a_pb_p$; $\boldsymbol{ab} = \|\boldsymbol{a}\|\|\boldsymbol{b}\|\cos\alpha$
Length of the projection: $\|\boldsymbol{a_p}\| = \|\boldsymbol{a}\|\cos\alpha$
Variance of x_i: $s_1^2 = \frac{\|x_i\|^2}{n-1}$; Generalized variance: $GV = \left(\frac{\|x_1\|\cdot\|x_2\|}{n-1}\cdot\sin\alpha\right)^2$

Distances

Euclidean: $D_{AB} = \sqrt{\sum_{j=1}^{p}(a_j - b_j)^2}$

Statistical: $SD_{ij}^2 = \left(\frac{x_i - x_j}{s}\right)^2$, s-standard deviation

Mahalanobis: $MD_{ik}^2 = \frac{1}{1-r^2}\left[\frac{(x_{i1}-x_{k1})^2}{s_1^2} + \frac{(x_{i2}-x_{k2})^2}{s_2^2} - \frac{2r(x_{i1}-x_{k1})(x_{i2}-x_{k2})}{s_1 s_2}\right]$

Variance, Sum of Squares, and Cross Products

Variance: $s_j^2 = \frac{\sum_{i=1}^{n} x_{ij}^2}{n-1} = \frac{SS}{df}$ (sum of squares/degrees of freedom)
Covariance: $s_{jk} = \frac{\sum_{i=1}^{n} x_{ij}x_{ik}}{n-1} = \frac{SCP}{df}$ (sum of the cross products/degrees of freedom)
SSCP – sum of squares and cross products matrix $\begin{pmatrix} SSX_1 & SCP \\ SCP & SSX_2 \end{pmatrix}$
S – covariance matrix $S_t = \frac{SSCP_t}{df}$
Within-Group Analysis: $SSCP_w = SSCP_1 + SSCP_2$ (pooled SSCP matrix) $S_w = \frac{SSCP_w}{n_1+n_2-2}$ (pooled cov m)
Between-Group Analysis: $SS_j = \sum_{g=1}^{G} n_g(\bar{x}_{jg} - \bar{x}_j)^2$; $SCP_{jk} = \sum_{g=1}^{G} n_g(\bar{x}_{jg} - \bar{x}_j)(\bar{x}_{kg} - \bar{x}_k)$
$SSCP_t = SSCP_w + SSCP_b$

Principal Components Analysis

$x_1^* = \cos\theta * x_1 + \sin\theta * x_2$; $x_2^* = -\sin\theta * x_1 + \cos\theta * x_2$
Σ covariance matrix; λ-eigenvalues; $|\boldsymbol{\Sigma} - \boldsymbol{\lambda I}| = \boldsymbol{0}$; $\boldsymbol{\gamma}$-eigenvector; $(\Sigma - \lambda I)\gamma = 0$; $\gamma'\boldsymbol{\gamma} = \boldsymbol{1}$;

Factor Analysis

Assumptions: 1. Means of indicators, common factor, unique factors are zero.
2. Variances of indicators and common factors are one. 3.$E(\xi_i\varepsilon_i) = 0$ and $E(\varepsilon_i\varepsilon_j) = 0$

Two-Factor Model: $x_1 = \lambda_{11}\xi_1 + \lambda_{12}\xi_2 + \varepsilon_1$

$x_2 = \lambda_{21}\xi_1 + \lambda_{22}\xi_2 + \varepsilon_2$

$\vdots$

$x_p = \lambda_{p1}\xi_1 + \lambda_{p2}\xi_2 + \varepsilon_p$

The variance of x: $E(x^2) = E(\lambda_1\xi_1 + \lambda_1\xi_2 + \varepsilon_1)^2$; $Var(x) = \lambda_1^2 + \lambda_2^2 + Var(\varepsilon) + 2\lambda_1\lambda_2\phi$

The correlation between any indicator and any factor (the structure loading):

$E(x\xi_1) = E[(\lambda_1\xi_1 + \lambda_1\xi_2 + \varepsilon_1)\xi_1]$; $Corr(x\xi_1) = \lambda_1 + \lambda_2\phi$

The shared variance between the factor and an indicator: $Shared\ variance = (\lambda_1 + \lambda_2\phi)^2$

The correlation between two indicators:

$$E(x_j x_k) = E[(\lambda_{j1}\xi_1 + \lambda_{j2}\xi_2 + \varepsilon_j)(\lambda_{k1}\xi_1 + \lambda_{k2}\xi_2 + \varepsilon_k)]$$
$$Corr(x_j x_k) = \lambda_{j1}\lambda_{k1} + \lambda_{j2}\lambda_{k2} + (\lambda_{j1}\lambda_{k2} + \lambda_{j2}\lambda_{k1})\phi$$

Confirmatory Factor Analysis

The covariance matrix (one-factor model, two indicators): $\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix}$

Evaluating model fit: χ^2-test H_0: $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$ H_a: $\boldsymbol{\Sigma} \neq \boldsymbol{\Sigma}(\boldsymbol{\theta})$ (test whether the difference between the sample and the estimated covariance matrix is a zero matrix)

$\chi^2 = \sum_{i=1}^{k} \frac{[n_i - E(n_i)]^2}{E(n_i)}$

Cluster Analysis

Measure of similarity – squared Euclidean distance between two points

Hierarchical clustering:

Centroid method – each group is replaced by centroid

Nearest-neighbor or single-linkage method – the distance between two clusters is represented by the minimum of the distance between all possible pair of subjects in the two clusters

Farthest-neighbor or complete-linkage method - … the maximum of the distances…

Average-linkage method - … the average distance…

Ward's method – does not compute distances between clusters. Method tries to minimize the total within-group sums of squares.

Discriminant Analysis

Assumptions: multivariate normality, equality of covariance matrices

Discriminant function: $Z = w_1x_1 + w_2x_2$

$\lambda = \frac{between\ -group\ sum\ of\ squares}{within-group\ sum\ of\ squares}$

Σ-variance-covariance matrix, **T**-total SSCP matrix. $\boldsymbol{\gamma}$-vector of weights.

Discriminant function $\boldsymbol{\xi} = \boldsymbol{X}'\boldsymbol{\gamma}$. **B** and **W** are between-groups and within-group SSCP matrices.

Maximize $\lambda = \frac{\gamma' B\gamma}{\gamma' W\gamma}$

$|\boldsymbol{W}^{-1}\boldsymbol{B} - \boldsymbol{\lambda I}| = \boldsymbol{0}$; $\boldsymbol{\gamma} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$ - Fisher's discriminant function

Logistic regression

$odds = \frac{p}{1-p}$

$\ln odds = \beta_0 + \beta_1X_1 + \cdots + \beta_kX_k$

$p = \frac{1}{1+e^{-(\beta_0+\beta_1X_1+\cdots+\beta_kX_k)}}$

Maximum likelihood estimation: $P(Y = 1) = p = \frac{e^{\beta X}}{1+e^{\beta X}}$

$L = \prod_{i=1}^{n} p_i^{y_i}(1 - p_i)^{1-y_i}$

Quadratic equations: $ax^2 + bx + c = 0; x = \frac{-b \pm \sqrt{b^2-4ac}}{2a}$

Cubic equations:

$y^3 + ay^2 + by + c = 0; y = x - \frac{a}{3}; x^3 + px + q = 0; x_1 = \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}$

Madeleine Hejberg

MULTIVARIATE METHODS

PCA

$\xi_1 = W_{11}X_1 + W_{12}X_2 + \ldots + W_{1p}X_p$
$\xi_2 = W_{21}X_1 + W_{22}X_2 + \ldots + W_{2p}X_p$
$\vdots$
$\xi_p = W_{p1}X_1 + W_{p2}X_2 + \ldots + W_{pp}X_p$

- The weights: $W_{i1}^2 + W_{i2}^2 + \ldots + W_{ip}^2 = 1 \quad , i = 1, \ldots, p$

  $W_{i1}W_{j1} + W_{i2}W_{j2} + \ldots + W_{ip}W_{jp} = 0 \quad \forall i \neq j$

- Characteristic equation: $\det(\lambda I - A) = 0$

- Loadings = correlation between the original and the new variables

  $\lambda_{ij} = \frac{W_{ij}}{\hat{s}_j}\sqrt{\lambda_i}$

FA

$x_1 = \lambda_{11}\xi_1 + \lambda_{12}\xi_2 + \ldots + \lambda_{1m}\xi_m + \varepsilon_1$
$x_2 = \lambda_{12}\xi_1 + \lambda_{22}\xi_2 + \ldots + \lambda_{2m}\xi_m + \varepsilon_2$
$\vdots$
$x_p = \lambda_{p1}\xi_1 + \lambda_{p2}\xi_2 + \ldots + \lambda_{pm}\xi_m + \varepsilon_p$

- Assumptions:
  1) Means of indicators, common factors and unique factors are zero.
  2) Variances of indicators and common factors are one.
  3) The unique factors are not correlated among themselves or with the common factors. $\Rightarrow E(\xi_i\varepsilon_j) = E(\varepsilon_i\varepsilon_j) = 0 \quad \forall i \neq j$

Model	Equations	The variance of any indicator x_j	Structure loading	Shared variance =(structure loading)^2	Correlation between indicators
One-factor model	$x_1 = \lambda_1\xi + e_1$ ⋮ $x_p = \lambda_p\xi + \varepsilon_p$	$V(x_j) = \lambda_j^2 + V(\varepsilon_j)$	$Cor(x_j, \xi) = \lambda_j$	λ_j^2	$Corr(x_j, x_k) = \lambda_j\lambda_k$
Two-factor model	$x_1 = \lambda_{11}\xi_1 + \lambda_{12}\xi_2 + \varepsilon_1$ ⋮ $x_p = \lambda_{p1}\xi_1 + \lambda_{p2}\xi_2 + \varepsilon_p$	$V(x_j) = \lambda_{j1}^2 + \lambda_{j2}^2 + 2\lambda_{j1}\lambda_{j2} + V(\varepsilon_j)$	$Cor(x_j, \xi_2) = \lambda_{j2} + \lambda_{j1}\phi$ $Cor(x_j, \xi_1) = \lambda_{j1} + \lambda_{j2}\phi$	$(\lambda_{j2} + \lambda_{j1}\phi)^2$ $(\lambda_{j1} + \lambda_{j2}\phi)^2$	$Cor(x_j, x_k) = \lambda_{j1}\lambda_{k1} + \lambda_{j2}\lambda_{k2} + [\lambda_{j1}\lambda_{k2} + \lambda_{j2}\lambda_{k1}]\phi$

$Cor(\xi_1, \xi_2) = \phi$

CFA

Underidentified model: #equations < #variables
Just-identified model: #equations = #variables
Overidentified model: #equations > #variables

$P(x_1, x_2, x_3) = P(x_1) \cdot P(x_2|x_1) \cdot P(x_3|x_1, x_2)$

CA

$D_{AB}^2 = \sum_{j=1}^{p}(a_j - b_j)^2 \qquad SD_{ik}^2 = \sum_{j=1}^{p}\left(\frac{x_{ij} - x_{kj}}{s_j}\right)^2$

$MD_{ik}^2 = \frac{1}{1-r^2}\left[\frac{(x_{i1} - x_{k1})^2}{s_1^2} + \frac{(x_{i2} - x_{k2})^2}{s_2^2} - \frac{2r(x_{i1} - x_{k1})(x_{i2} - x_{k2})}{s_1 s_2}\right]$

DA
(two groups)

$\lambda = \frac{SS_B}{SS_W} \rightarrow max$

If $\xi = \bar{X}^T\hat{\gamma} \Rightarrow$ The estimation of $\hat{\gamma}$: $\hat{\gamma}^T = (\bar{\mu}_1 - \bar{\mu}_2)^T \Sigma^{-1}$

$SSCP_W = SSCP_1 + SSCP_2$
$SSCP_T = SSCP_W + SSCP_B$

LOG-REG

$odds = \frac{p}{1-p}$

$p = \frac{odds}{1 + odds}$

$\ln\frac{p}{1-p} = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k$

$P(Y=1) = p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k)}}$

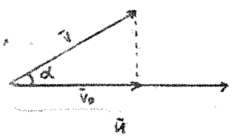
If $Y = \beta_0 + \beta_1 X$:
$\hat{\beta}_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}, \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$

- Rotation $\Rightarrow$ $a_1^* = \cos\theta\, a_1 + \sin\theta\, a_2$
  $a_2^* = -\sin\theta\, a_1 + \cos\theta\, a_2$

$S^2 = \frac{SS}{df} = \frac{\sum(x_i - \bar{x})^2}{n-1} = \frac{\sum x_i^2 - n\bar{x}^2}{n-1}$

$S_{xy} = \frac{SCP}{df} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n-1} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{n-1}$

- Projection: of $\bar{v}$ onto $\bar{u}$



- The projection vector: $\bar{v}_p = \frac{\|\bar{v}_p\|}{\|\bar{u}\|}\bar{u}$
- $\|\bar{v}_p\| = \|\bar{v}\|\cos\alpha = \frac{\bar{v} \cdot \bar{u}}{\|\bar{u}\|}$

- Direction cosines = The cosines of the angle between a vector and the axes