ACTA UNIVERSITATIS LODZIENSIS
FOLIA OECONOMICA 156, 2002

*Teresa H. Jelenkowska**

BAYESIAN AND AKAIKE'S INFORMATION CRITERIONS FOR SOME MULTIVARIATE TESTS OF HOMOGENEITY WITH APPLICATIONS IN MULTISAMPLE CLUSTERING

Abstract. This paper studies the AIC and BIC (Akaike's and Bayesian Information Criterion) replacement for:

- Box's (1949) M test of the homogeneity of covariances,
- Wilks' (1932) Λ criterion for testing the equality of mean vectors and
- likelihood ratio test of the complete homogeneity as two of model – selection criterions.

AIC and BIC are new procedures for comparing means and samples, and selecting the homogeneous groups from heterogenous ones in multi-sample data analysis problems.

From the Bayesian view-point, the approach to the model-selection problem is to specify the prior probability of each model, prior distributions for all parameters in each model and compute the posterior probability of each model given the data. That model for which the estimated posterior probability is the largest is chosen to be the best one.

A clustering technique is presented to generate all possible choices of clustering alternatives of groups and indentify the best clustering among the alternative clusterings.

Key words: Model selection, Akaike's and Bayesian information criterions, multisample clustering.

I. INTRODUCTION

Many practical situations require the presentation of multivariate data from several structured samples for comparative inference and the grouping of the heterogeneous samples into homogeneous sets of samples.

In statistical literature, the Multivariate Analysis of Variance (MANOVA) is a widely used model for comparing two or more multivariate samples. The formal analyses involved in MANOVA are not informative. The model

---

* Dr., Institute of Applied Mathematics, University of Agriculture in Lublin.

considers the variabilities only in pairs of samples, and it ignores the variabilities in other groups. For this reason, in any problem where a set of parameters is to be partitioned into groups, it is reasonable to provide practically useful statisitcal procedures that would tell us which samples should be clustered together and which samples should not be cluster together.

A common problem in clustering techniques is the difficulty of deciding on the number of clusters present in a given data set. This problem has been noted by many authors such as Beale (1969), Marriot (1971), Caliński and Harabasz (1974), Maronna and Jacovkis (1974), Duran and Odell (1974), Hartigan (1977), Matusita and Ohsumi (1980), Bozdagan (1988, 1990, 1994) proposed and estabilished AIC (1973) and he developed a new informational measure of complexity (ICOP) criterion as performance measure for choosing the number of clusters. Maddala (1992) discusses the implied critical F-values for the different classical model – selection criteria (maximum R^2, Amemiya's PC, Mallows'Cp, Hocking's Sp and Akaike's AIC), as well as the F-ratios presented by Leamar (1978) based on his posterior adds criterion. Andrews (1994) discusses the large-sample correspondence between classical hypothesis tests and Bayesian posterior adds tests.

From the Bayesian viewpoint, the approach to the model-selection problem is to specify the prior probability of each model, prior distribution for all parameters in each model and compute the posterior probability of each model given the data. We choose this model for which the estimated posterior probability is the largest. Next, individual sample assigned to this cluster for which also, the estimated posterior probability is the largest.

To ilustrate two of the important virtues of model selection criterions, in this paper, we compute AIC and BIC replacement for three tests of multivariate homogeneity by varying p, the number of variables; K – the number of samples; and n – the sample size.

We present the applications of the result obtained in this paper to multi-sample cluster analysis, the problem of clustering data matrices combinatorially without forcing an arbitrary choice among the clustering alternatives to achieve a parsimonious grouping of samples. The remainder of this paper is organised as follows.

In Section 2, we discuss three linked multisample models and as an alternative to conventional tests procedure we derive AIC and BIC for the test of homogeneity of covariance model, MANOVA model and complete homogeneity model. In Section 3, we give the AIC and BIC – replacements for multivariate classical tests of homogeneity. In Section 4 we shall propose Multi-Sample Cluster Analysis (MSCA) as an alternative to conventional Multiple Comparison Procedure (MCP).

II. AIC'S AND BIC PROCEDURES FOR MULTIVARIATE MODELS

II.1. AKAIKE AND BAYESIAN CRITERIONS

Throughout this section we shall suppose that we have independent data matrices $Y_1, Y_2, \dots, Y_K$, where the rows of $Y_i\,(n_i \times p)$ are independent and identically distributed $N_p(\mu_i, \Sigma_i)$, $i = 1, 2, \dots, K$. In terms of the parameters $\theta = (\mu_1, \dots, \mu_K, \Sigma_1, \dots, \Sigma_K)$, the models we are going to consider are as follows:

$$1)\ \mathcal{M}_1:\quad \theta. = (\underbrace{\mu_1, \mu_2, \dots, \mu_K}, \underbrace{\Sigma_1, \dots, \Sigma_K})$$

(varying mean vectors and varying covariances)

with $m = Kp + Kp(p+1)/2$ parameters, where K is the number of samples and, p is the number of variables.

$$2)\ \mathcal{M}_2:\quad \theta.. = (\underbrace{\mu_1, \mu_2, \dots, \mu_K}, \underbrace{\Sigma, \dots, \Sigma, \Sigma})$$

(varying mean vectors and common covariances)

with $m = Kp + p(p+1)/2$ parameters.

$$3)\ \mathcal{M}_3:\quad \theta... = (\underbrace{\mu, \mu, \dots, \mu}, \underbrace{\Sigma, \dots, \Sigma, \Sigma})$$

(common mean vectors and common covariances)

with $m = p + p(p+1)/2$ parameters. The three models which we considered above, are the most common models which are linked in their parameters, and thus, are related to one another. If we denote $\Omega.$, to be the unrestricted parameter space of $\theta.$, $\Omega..$ to be the conditional parameter space of $\theta..$ and $\Omega...$ to be the restricted parameter space of $\theta...$, then the relationship between these three models in terms of their parameter spaces can be writen as $\Omega... \subset \Omega.. \subset \Omega.$.

In this section we shall derive the forms of AIC and BIC for testing the equality of covariance materices and means for the above models.

First, we recall the definition of AIC:

Definition 2.1

Let $(M_j\colon j \in J)$ be a set of competing models indexed by $j = 1, 2, \dots, J$. Then, the criterion

$$AIC(j) = -2\ln L[(\hat{\theta}(j)] + 2m(j) \tag{1}$$

which is minimized to choose a model M_j over the set of models is called Akaike's. Information Criterion (AIC) (see Akaike (1973) and Sakamoto (1986)).

In (1) $L[\hat{\theta}(j)]$ is the likelihood function of observations, $\hat{\theta}(j)$ is the maximum likelihood estimate of the parameter vector θ under the model M_j and $m(j)$ is the number of independent parameters estimated when M_j is the model.

The AIC statistic is an estimator of the risk of the model under the maximum likelihood estimations. That model, which optimizes the criterion, is chosen to be the best model.

Now, we define Bayesian Model – Selection Criterion (BIC): Assume the each possible alternative models M_j, $j = 1, 2, ..., J$ is generated by the densities

$$p_1(Y/\theta_1, \quad p_2(Y/\theta_2), \; ..., \; p_J(Y/\theta_J)$$

with different numbers of parameters θ_j. Suppose that a prior has been specified for ech model M_j and parameters θ_j, $j = 1, 2, ..., J$. Let these priors be denoted respectively by $\Pi(M_j)$ and $\Pi(\theta_j)$ and the likelihood function by $p_j(Y/\theta_j)$. Then the posterior probability associated with model M_j is given by

$$p(M_j/Y) = \Pi(M_j)\int\Pi(\theta_j)P_j(Y/\theta_j)d\theta_j \tag{2}$$

From (2) we see that the posterior is equal to the priot probability $\Pi(M_j)$ of the j-th model being the true one. $\Pi(M_j)$ times the averaged likelihood with the prior $\Pi(\theta_j)$.

For example the prior probability associated with the model M_j can be $\Pi(M_j) = p(w_j)$, where

$$w_j = \begin{cases} 0 \text{ if } M_j & \text{is false} \\ 1 \text{ if } M_j & \text{is true} \end{cases}$$

denotes a dichotomous random variable and $\sum_{j=1}^{J} \Pi(M_j) = 1$.

The amount and kind of prior information to be employed in an analysis will depend on what we know and what we judge appropriate to incorporate into the analysis. We recognize that there are situations in which we know very little and thus want procedures for comparing models with the use of little prior information. When we have prior information, say from analyses of part samples of data, and wish to incorporate it in our comparison of models.

We argue, however, that this framework may not be appropriate in the important case where the models are nested. Thus, if M_i is a particular case of M_j, and M_i is of smaller dimesion that M_j, that this strictly positive probability $\Pi(M_i)$ has been assigned to a set of zero Lebesque measure under the larger model M_j, which is assumed to be appropriate.

Thus a comparison of the posterior probabilities will provide a basis for choosing the best model among all models M_j, $j = 1, 2, ..., J$. The Bayes solution consists of selecting the model that is a posterior most probable. Via Bayes formula that is equivalent to choosing the j, that maximises

$$BIC(j) = \ln P(M_j / Y) \tag{3}$$

where ln denotes the natural logarithm. The above formula is a Bayesian model – selection criterion called BIC.

MODEL $\mathcal{M}_1$

II.2. AIC FOR THE TEST OF HOMOGENEITY OF COVARIANCES MODEL

Let Y_i, $i = 1, 2, ..., k$ be a random sample of observations from the i-th population $N_p(\mu_i, \Sigma_i)$. The basic null and the alternative hypotheses are given by

$$H_0^{\cdot}: \quad \Sigma_1 = \Sigma_2 = ... = \Sigma_k$$

versus $H_0^{\cdot}$: the K population covariances are not all equal.

Now, we derive the form of Akaike's Information Criterion (AIC) to test the hypothesis that the covariance matrices of these populations are equal. The likelihood function of all the sample observations is given by

$$L(\theta.) = p(Y_1, \; ..., \; Y_K / \theta.) = \prod_{i=1}^{K} p(Y_i / \theta.) \tag{4}$$

The log likelihood function, $l(\theta.)$, is the natural logarithm of $L(\theta.)$ and is defined by

$$l(\theta.) = \ln L(\theta.) = \sum_{i=1}^{K} \ln p(Y_i / \theta.) \tag{5}$$

where

$$p(Y_i / \theta.) = (2\Pi)^{-np/2} |\Sigma_i|^{-n_i/2} \tag{6}$$

$$\exp\left\{-\frac{1}{2}tr\sum_{i=1}^{K}\Sigma_i^{-1}(S_i+n_i(\overline{Y}_i-\mu_i)'(\overline{Y}_i-\mu_i))\right\}$$

The log likelihood function is

$$l(\theta.)=-\left(\frac{np}{2}\right)\ln 2\Pi-\frac{1}{2}\sum_{i=1}^{K}n_i\ln|\Sigma_i|-\frac{1}{2}tr\left(\sum_{i=1}^{K}\Sigma_i^{-1}S_i\right)$$

$$-\frac{1}{2}tr\left[\sum_{i=1}^{K}n_i\Sigma_i^{-1}(\overline{Y}_i-\mu_i)'(\overline{Y}_i-\mu_i)\right],\quad \text{where}\quad S_i=\sum_{l=1}^{n_i}(Y_{il}-\overline{Y}_i)'(Y_{il}-\overline{Y}_i) \tag{7}$$

The maximum likelihood estimators (MLE's) of μ_i and Σ_i are

$$\hat{\mu}_i=\overline{\mathrm{Y}}_i \quad\text{and}\quad \hat{\Sigma}_i=\frac{S_i}{n_i},\quad i=1,\ ...,\ K \tag{8}$$

Substituting the MLE's into (2.8) simplifying, the maximised log likelihood becomes

$$l(\hat{\theta}.)=-\left(\frac{np}{2}\right)\ln 2\Pi-\frac{1}{2}\sum_{i=1}^{K}n_i\ln|n_i^{-1}S_i|-\frac{np}{2} \tag{9}$$

Since

$$AIC=-2l(\hat{\theta}.)+2m \tag{10}$$

where $m=Kp+Kp(p+1)/2$ is the number of parameters, the AIC becomes

$$AIC(\mu_i,\Sigma_i)=\sum_{i=1}^{K}n_i\ln|n_i^{-1}S_i|+np(\ln 2\Pi+1)$$
$$+2[Kp+Kp(p+1)/2 \tag{11}$$

II.3. BIC FOR THE TEST OF HOMOGENEITY OF COVARIANCES MODEL

Denote BIC $(\mu_i, \Sigma_i)\equiv$ BIC (varying μ and varying Σ).

Now, we derive the form of Bayes' Information Criterion (BIC) to test the hypothesis that the covariance matrices $\Sigma_1, ..., \Sigma_k$ are equal.

So that, consider K normal populations with different mean vectors μ_i, and different covariance matrices Σ_i, $i=1, ..., K$.

As regards prior for the parameters $\underset{1\times p}{\mu_i}$ and $\underset{p\times p}{\Sigma_i}$ we employ the following natural conjugate forms for $i = 1, \ldots, K$.

The joint prior becomes

$$p(\mu_i, \Sigma_i) = p(\mu_i / \Sigma_i) p(\Sigma_i) \tag{12}$$

where

$$p(\underset{1\times p}{\mu_i} \;/\; \underset{p\times p}{\Sigma_i}) \sim N_p(0, \Sigma_i b_i) \tag{13}$$

In (13) we have assumed a proper normal prior for the elements of μ_i, given Σ_i, with prior mean vector zero and covariance matrix $\Sigma_i b_i$, where b_i is an arbitrary positive scalar obtained by enrichment from n_i, and

$$p(\Sigma_i) \sim W^{-1}(V_i, p, v_i), \quad \Sigma_i > 0, \; V_i > 0 \quad \text{and} \quad v_i > p+1 \tag{14}$$

Σ_i will be said to have an inverted Wishart distribution with scale matrix V_i, dimension p, and v_i degrees of freedom. Substituting from (13) and (14) into (12) shows that the joint natural conjugate prior distribution for μ_i and Σ_i^1 is given by the Normal – inverted Wishart distribution as

$$p(\mu_i, \Sigma_i) = \prod_{i=1}^{K} \left|\Sigma_i^{-1}\right|^{\frac{1}{2}(v_i + np - p - 1)} \exp\left\{\left(-\frac{1}{2}\right) tr\Sigma_i^{-1} \left[\sum_{i=1}^{K} (n_i b_i \mu_i' \mu_i + V_i)\right]\right\} \tag{15}$$

The likelihood function of all the sample observations is given by

$$p(Y/\mu_i \Sigma_i) \propto \prod_{i=1}^{K} \left|\Sigma_i^{-1}\right|^{\frac{n}{2}} \exp\left(-\frac{1}{2}\right) tr\Sigma_i^{-1} \left\{\sum_{i=1}^{K} \left[S_i + n_i(\overline{Y}_i - \mu_i)'(\overline{Y}_i - \mu_i\right]\right\} \tag{16}$$

The joint posterior density for μ_i, Σ_i is found by multiplying (15) and (16):

$$p(\mu_i \Sigma_i / Y) \propto \prod_{i=1}^{K} \left|\Sigma_i^{-1}\right|^{\frac{1}{2}(v_i + n_i + (n_i - 1)p - 1)} \exp\left\{-\frac{1}{2} tr\Sigma_i^{-1} A_i\right\} \tag{17}$$

where

$$A_i = \sum_{i=1}^{K} \left[n_i b_i \mu_i' \mu_i + V_i + S_i + n_i(\overline{Y}_i - \mu_i)'(\overline{Y}_i - \mu_i)\right]$$

Integrating (17) with respect to μ_i and Σ_i, and assuming that there is no reason to believe more in one model than the other, we would take $\Pi(M_1) = \Pi(M_2) = \Pi(M_3) = \frac{1}{3}$, then BIC becomes:

$$BIC(1) = BIC(\mu_i, \Sigma_i) = \ln P(M_1 / Y) \tag{18}$$

where

$$P(M_1 / Y) = \frac{1}{3}\prod_{i=1}^{K}\left|\sum_{i=1}^{K}\left[V_i + S_i + (n_i - d_i^{-1})\,\overline{Y}_i'\overline{Y}_i\right]d_i\right|^{\frac{1}{2}(v_i + n_i + n_i p)} \tag{19}$$

and

$$d_i = n_i(1 + b)$$

MODEL $\mathcal{M}_2$

II.4. AIC FOR THE TEST OF HOMOGENEITY OF MEANS

Consider in this case again K normal populations with different mean vectors μ_i, $i = 1, \ldots, K$, but each population is assumed to have the same covariance matrix Σ.

In terms of testing the significant of the equality of the mean vectors of K samples, the MANOVA hypotheses are:

$$H_0^{oo}\colon\ \mu_1 = \mu_2 = \ldots = \mu_k \quad \text{given that} \quad \Sigma_1 = \Sigma_2 = \ldots = \Sigma_k = \Sigma$$

versus: H_1^{oo}: Nost all μ_s''s are equal.

To derive Akaike's Information Criterion (AIC) in this case, we use the log likelihood function given in (7). Since each population is assumed to have the same covariance matrix Σ, the log likelihood function becomes

$$\begin{aligned} l(\theta..) = {} & -\left(\frac{np}{2}\right)\ln 2\Pi - \frac{n}{2}\ln|\Sigma| - \frac{1}{2}tr\left(\Sigma^{-1}\sum_{i=1}^{K}S_i\right) \\ & -\frac{1}{2}tr\left[\Sigma^{-1}\sum_{i=1}^{K}n_i(\overline{Y}_i - \mu_i)(\overline{Y}_i - \mu_i)\right] \end{aligned} \tag{20}$$

The maximum likelihood estimators of μ_i and Σ are

$$\hat{\mu}_i = \overline{Y}_i \quad \text{and} \quad \hat{\Sigma} = n^{-1}\sum_{i=1}^{K}S_i, \quad i = 1, \ldots, K$$

Substituting the MLE's into (19) and simplifying, the maximised log likelihood becomes

$$l(\hat{\theta}..) = -\frac{np}{2}\ln(2\Pi) - \frac{n}{2}\ln\left|n^{-1}\sum_{i=1}^{K}S_i\right| - \frac{np}{2} \tag{21}$$

Since

$$AIC = -2l(\hat{\theta}) + 2m$$

where $m = Kp + p(p+1)/2$, then AIC becomes

$$AIC(\mu_i, \Sigma) = np(\ln 2\Pi) + n\ln\left| n_i \sum_{i=1}^{K} S_i \right| + np + 2[Kp + p(p+1)/2] \tag{22}$$

II.5. BIC FOR THE TEST OF HOMOGENEITY OF MEANS

Denote $BIC(\mu_i, \Sigma) \equiv BIC$ (varying μ and common Σ).

In this case, we derive the information criteria in a similar fashion as in Section II.3 by assuming the same covariance matrix Σ in the likelihood function and the same type of prior distributions for μ_i and Σ. Thus, the joint prior density function of the parameters μ_i and Σ is

$$p(\mu_i, \Sigma) = \left|\Sigma^{-1}\right|^{\frac{1}{2}(v+np-p-1)} \exp\left(-\frac{1}{2}\right) tr\Sigma^{-1}\left\{\sum_{i=1}^{K}(n_i b_i \mu_i' \mu_i + V)\right. \tag{23}$$

where $\Sigma > 0$, $V > 0$ and $v > p + 1$.

Then the joint posterior density for μ_i and Σ is

$$p(\mu_i, \Sigma / Y) \propto \left|\Sigma^{-1}\right|^{\frac{1}{2}(v+n+np-p-1)}$$

$$\cdot \exp\left(-\frac{1}{2}\right) tr\,\Sigma^{-1}\left\{\sum_{i=1}^{K}\left[n_i b_i \mu_i' \mu_i + V_i + S_i + n_i(\overline{Y}_i - \mu_i)'(\overline{Y}_i - \mu_i)\right]\right\} \tag{24}$$

After integrating with respect to μ_i and Σ we obtain

$$BIC(2) = BIC(\mu_i, \Sigma) = \ln P(M_2 / Y) \tag{25}$$

where

$$P(M_2 / Y) = \frac{1}{3}\left|\sum_{i=1}^{K}\left[V + S_i + (n_i - d_i^{-1})\overline{Y}_i'\overline{Y}_i\right]d_i\right|^{\frac{1}{2}(v+n+np)} \tag{26}$$

MODEL $\mathcal{M}_3$

II.6. AIC FOR THE TEST OF COMPLETE HOMOGENEITY MODEL

The null hypothesis to test of complete homogeneity is

$$H_0^{\cdots}:\quad \mu_1 = \mu_2 = \ldots = \mu_k \text{ and } \Sigma_1 = \Sigma_2 = \ldots = \Sigma_k$$

versus: $H_1^{\cdots}$: Not all K mean vectors and covariance matrices are equal.

To derive Akaike's Information Criterion for model M_3, we set all μ_i's equal to μ and all the Σ_i's equal to Σ and obtain the log likelihood function which is given by

$$l(\theta\ldots) = -\frac{np}{2}(2\Pi) - \frac{n}{2}\ln|\Sigma| - \frac{1}{2}tr(\Sigma^{-1}T) - \frac{n}{2}tr\left[\Sigma^{-1}(\overline{Y}-\mu)'(\overline{Y}-\mu\right] \quad (27)$$

where T is a total sum square matris.

The MLE's of μ and Σ are

$$\hat{\mu} = \overline{Y} \quad \text{and} \quad \hat{\Sigma} = n^{-1}T$$

Substituting these back into (25), we have the maximum log likelihood

$$l(\hat{\theta}\ldots) = -\frac{np}{2}\ln(2\Pi) - \frac{n}{2}\ln|n^{-1}T| - \frac{np}{2} \quad (28)$$

Thus, using the equation of AIC in (10) again, where $m = p + p(p+1)/2$ the AIC becomes

$$AIC(m, \Sigma) = np\ln(2\Pi) + \mathrm{n}\ln|n^{-1}T| + np + 2[p + p(p+1)/2] \quad (29)$$

II.7. BIC FOR THE TEST OF COMPLETE HOMOGENEITY MODEL

Denote $BIC(\mu, \Sigma) \equiv BIC$ (common μ and common Σ).

To derive the information criteria under the assumption that all μ_i, $i = 1, \ldots, k$ are equal to μ and the Σ_i are equal to Σ, we assume the same type of prior distributions for μ and Σ as before. So that, the joint posterior density for μ and Σ is the following:

$$p(\mu, \Sigma / Y) \propto \left|\Sigma^{-1}\right|^{\frac{1}{2}(v+n+np-p-1)}$$

$$\exp\left(-\frac{1}{2}\right) tr\Sigma^{-1}\left\{nb\mu'\mu + KV + \sum_{i=1}^{K}[S_i + n_i(\overline{Y}_i - \mu)'(\overline{Y}_i - \mu)]\right\} \tag{30}$$

where b and V are hyperparameters positive defined for the Normal – inverted Wishart joint prior distribution.

After some work we obtain

$$p(\mathscr{M}_3 / Y) = BIC(\mu, \Sigma) = \frac{1}{3}\left|KV + \sum_{i=1}^{K}[S_i + (n_i - d_i^{-1}\overline{Y}_i\,\overline{Y}_i]\,d_i\right|^{\frac{1}{2}(v+n+np)} \tag{31}$$

where $d_i = n_i(1 + b)$.

III. AIC AND BIC – REPLACEMENTS FOR CLASSICAL TESTS OF HOMOGENEITY

In Section II we derived the exact formulas for AIC and BIC for each of the multivariate models. In this section, we use the formulas of AIC and BIC replacements for multivariate classical test of homogeneity.

Proposition 3.1

We reject $H_0^{\cdot}$ if

$$\begin{gathered}\Delta AIC(H_0^{o}, H_1^{o}) = AIC(\mu_i, \Sigma) - AIC(\mu_i, \Sigma) > 0 \\ \Leftrightarrow n\ln\left|n^{-1}\sum_{i=1}^{K} S_i\right| - \sum_{i=1}^{K} n_i \ln|n_i^{-1}S_i| > p(p+1)(K-1)\end{gathered} \tag{32}$$

or

$$\Delta BIC(H_0^{\cdot}, H_1^{\cdot}) = BIC(\mu_i, \Sigma) - BIC(\mu_i, \Sigma_i) < 0 \tag{33}$$

where $BIC(\mu_i, \Sigma_i)$ is given in (18), and $BIC(\mu_i, \Sigma)$ is given in (25).

Proposition 3.2

We reject $H_0^{\cdot\cdot}$ if

$$\Delta AIC(H_0^{\cdot\cdot}, H_1^{\cdot\cdot}) = AIC(\mu, \Sigma) - AIC(\mu_i, \Sigma) > 0$$

$$\Leftrightarrow n\ln|n^{-1}T| - n\ln\left|n^{-1}\sum_{i=1}^{K} S_i\right| > 2p(K-1) \tag{34}$$

or

$$\Delta BIC(H_0'', H_1'') = BIC(\mu, \Sigma) - BIC(\mu_i, \Sigma) < 0 \tag{35}$$

where $BIC(\mu, \Sigma)$ is given in (31).

Proposition 3.3
We reject H_0''' if

$$\Delta AIC(H_0''', H_1''') = AIC(\mu, \Sigma) - AIC(\mu_i, \Sigma_i) > 0$$

$$\Leftrightarrow n \ln|n^{-1}T| - \sum_{i=1}^{K} n \ln|n_i^{-1}S_i| > p(p+3)(K-1) \tag{36}$$

or

$$\Delta BIC(H_0''', H_1''') = BIC(\mu, \Sigma) - BIC(\mu_i, \Sigma_i) < 0 \tag{37}$$

Using AIC and BIC procedures we avoid any restrictions on p and K and we do not need to assume any level of significance α. In large sample (as the sample size goes to infinity) BIC formula for a given prior corresponds to classical hypothesis tests for some significance level α. Thus the choice of significance level for classical tests is seen to be analogous to the choice of prior for Bayesian criterion and vice versa.

IV. MULTI-SAMPLE CLUSTER ANALYSIS

The aim of cluster analysis is to put the K samples into k homogeneous groups, where k is unknown and varying, and $k \leqslant K$.
Let

$$\underset{n \times p}{Y} = \begin{bmatrix} \underset{n_1 \times p}{Y_1} \\ \vdots \\ \underset{n_k \times p}{Y_k} \end{bmatrix} \tag{38}$$

be a data matrix of K samples, where $\underset{n_i \times p}{Y_i}$ represents the observations on p dependent variables from i-th sample, $i = 1, \ldots, K$, and $n = \sum_{i=1}^{K} n_i$.

First, we should know the total number of clusters of samples.

Bozdogan (1986) generated all possible clustering alternatives of groups on the computer using efficient combinatorial algorithms and complete enumerations technique. The total number of ways of clustering alternatives K samples into k-sample clusters where $k \leqslant K$ (k is known in advance) such that none of the k-sample clusters is empty, is given by

$$S(K, k) = \frac{1}{k!} \sum_{i=1}^{K} (-1)^{i} \binom{k}{j} (k-i)^{K} \tag{39}$$

If k is not specified a prior and it is unknown, then the total number of clustering alternatives is given by

$$\sum_{i=1}^{K} S(K, k) \tag{40}$$

$S(K, k)$ is writen in terms of the recursive formula

$$S(K, k) = kS(K-1, k) + S(K-1, k-1) \tag{41}$$

with $S(1, 1) = 1$ and $S(1, k) = 0$ for $k \neq 1$ and $S(K, 2) = 2^{K-1} - 1$.

If we identify the best fitting model using BIC or AIC criterion we should determine the number of clustering alternatives and next we should choose the best alternative clustering. If we want to answer the question: „which clustering alternative do we choose?" we should compute the AIC for all clustering alternatives and we should choose this clustering alternative for that the minimum AIC occurred.

From Bayesian point of view, we should calculate the posterior probability of each alternative clustering. The best one is assigned to this cluster for which the estimated posterior probability is the largest.

V. CONCLUSIONS

To apply a new approach using AIC or BIC procedures we should:

(i) Identify the best fitting parametric model.

(ii) Curry out multisample cluster analysis (MSCA) under the best fitting model.

(iii) Determine the relevant variables between the samples by using the subset selection procedure.

From the results presented in this paper:

– We can determine whether we should use equal or varying covariance matrices in the analysis of a data set.

– By subset selection of variables across the samples, we can reduce the dimensionality of data set.

– We do not need to assume any arbitrary level of significance α.

– Qualitatively both Bayesian and Akaike's procedures give a mathematical formulation of the principle of parsimony in model building.

– Quantitatively, Bayesian procedure leans more than Akaike's towards lower-dimensional models.

– As the sample size goes to infinity, BIC formula for a given prior corresponds to classical hypothesis tests for some significance level α.

– For large numbers of observations the procedures differ markedly from each other.

– There is no single criterion, which will play the role of a panacea in model selection problems.

REFERENCES

Akaike H. (1973), *Information Theory and Extension of the Maximum Likelihood Principle*, [in:] *Second International Symposium on Information Theory*, Eds. B. N. Petrov, Csaki, Academiai Kiado Budapest, 267–281.

Andrews D. W. K. (1994), *The Large Sample Correspondence Between Classical Hypothesis Tests and Bayesian Posterior Odds Tests*, „Econometrica" **62**, 1207–1232.

Beale E. M. (1969), *Cluster Analysis*, Scientific Control System, London.

Bozdogan H. (1986), *Multi-Sample Cluster Analysis as an Alternative to Multiple Comparison Procedure*, „Bulletin of Informations and Cybernetics" **22**, 1–2, 95–130.

Bozdogan H. (1988), *ICOMP: A New Model Selection Criterion, in Classification and Related Methods of Data Analysis*, Ed. H. Bock, North-Holland, Amsterdam, 599–608.

Bozdogan H. (1990a), *On the Information – Based Measure of Covariance Complexity and its Application to the Evaluation of Multivariate Linear Models*, „Communications in Statistics (Theory and Methods)" **19**(1), 221–278.

Caliński T., Harabasz J. (1974), *A Deudrite Method for Cluster Analysis*, „Communications in Statistics" **3**, 1–27.

Duran B. S., Odell P. L. (1974), *Cluster Analysis. A Survey*, Springer-Verlag, New York.

Hartigan J. A. (1977), *Distribution Problems in Clustering*, [in:] *Classification and Clustering*, Ed. J. Van Ryzin, Academic Press, New York, 45–71.

Leamer E. E. (1978), *Specification Searches*, Wiley, New York.

Maddala G. S. (1992), *Introduction to Econometrics*, Macmillan, New York.

Maronna R., Jacovkis P. M. (1974), *Multivariate Clustering Procedures with Variable Metrics*, „Biometrics", **30**, 499–505.

Marriott F. H. C. (1971), *Practical Problems in a Method of Cluster Analysis*, „Biometrics" **27**, 501–514.

Matusita K., Ohsumi N. (1981), *Evaluation Procedure of Some Clustering Techniques*, Unpublished paper, The Institute of Statistical Mathematics, Tokyo, Japan.

Sakamoto Y., Ishiguro M., Kitagawa G. (1986), *Akaike Information Criterion Statistics*, D. Reidel, Tokyo.

Teresa H. Jelenkowska

KRYTERIUM BAYESOWSKIE I KRYTERIUM AKAIKE DLA TESTÓW HOMOGENICZNOŚCI W MODELU LINIOWYM

(Streszczenie)

W pracy zostały przedstawione dwa kryteria dotyczące selekcji modeli, mianowicie kryterium Akaike: AIC (Akaike's Information Criterion) i kryterium bayesowskie: BIC (Bayesian Information Criterion). Obydwa te kryteria zostały zilustrowane na przykładzie trzech wielowymiarowych modeli liniowych: modelu $\mathcal{M}_1$, w którym zakłada się zróżnicowanie pomiędzy średnimi i macierzami kowariancji badanych cech, modelu $\mathcal{M}_2$, który jest heterogeniczny ze względu na wektory średnich i homogeniczny ze względu na macierze kowariancji oraz modelu $\mathcal{M}_3$, który jest w pełni homogeniczny, tzn. nie występuje zróżnicowanie ani ze względu na średnie ani macierze kowariancji.

Dla każdego z tych modeli zostały wyprowadzone *explicite* funkcje określające obydwa kryteria selekcji modeli: AIC i BIC. Powyższe kryteria służą do wybrania modelu, który najlepiej opisuje nasze dane eksperymentalne w miejsce tradycyjnie stosowanych testów badających homogeniczność średnich, jak i macierzy kowariancji w wielowymiarowych modelach liniowych. Zarówno kryterium BIC, jak i AIC są jednymi z najbardziej nowoczesnych procedur stosowanych w problemie selekcji modeli.