Journal of Statistical Theory and Applications (2021) 20:450–461
https://doi.org/10.1007/s44199-021-00002-4

RESEARCH ARTICLE

# Using Model Selection Criteria to Choose the Number of Principal Components

**Stanley L. Sclove**[1]

Received: 19 June 2021 / Accepted: 31 August 2021 / Published online: 20 September 2021

**Abstract**
The use of information criteria, especially AIC (Akaike's information criterion) and BIC (Bayesian information criterion), for choosing an adequate number of principal components is illustrated.

**Keywords** Information criteria · AIC · BIC · Principal components

### Abbreviations

| | |
|---|---|
| AIC | Akaike's information criterion |
| BIC | Bayesian information criterion |
| DIAS | Diastolic blood pressure |
| HT | Height |
| LC | Linear combination |
| LL | Maximum log likelihood |
| MLE | Maximum likelihood estimate |
| MSE | Mean squared error |
| PC | Principal component |
| SYS | Systolic blood pressure |
| WT | Weight |

## 1 Introduction

This paper applies model selection criteria, especially AIC and BIC, to the problem of choosing a sufficient number of principal components to retain. It applies the concepts of Sclove [13] to this particular problem.

✉ Stanley L. Sclove
slsclove@uic.edu

[1] University of Illinois at Chicago, Chicago, USA

## 2 Background

Other researchers have considered to problem of the choice of number of principal components. For example, Bai et al. [6] examined the asymptotic consistency of the criteria AIC and BIC for determining the number of significant principal components in high-dimensional problems. The focus here is not necessarily on high-dimensional problems.

To begin the discussion here, we first give a short review of some general background on the relevant portions of multivariate statistical analysis, such as may be obtained from textbooks such as Anderson [5] or Johnson and Wichern [9].

### 2.1 Sample Quantities

Suppose we have a multivariate sample $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ of $n$ $p$-dimensional random vectors,

$$\mathbf{x}_i = (x_{1i}, x_{2i}, \ldots x_{pi})', \quad i = 1, 2, \ldots, n.$$

The transpose ($'$) means that we are thinking of the vectors as column vectors. The sample *mean vector* is

$$\bar{\mathbf{x}} = \sum_{i=1}^{n} \mathbf{x}_i / n.$$

The $p \times p$ sample covariance matrix is

$$S = \sum_{i=1}^{n} (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})' / (n-1).$$

### 2.2 Population Quantities and Principal Components

The sample covariance matrix $\mathbf{S}$ estimates the true covariance matrix $\boldsymbol{\Sigma}$ of the random variables

$$X_1, X_2, \ldots, X_p.$$

That is,

$$\boldsymbol{\Sigma} = [\sigma_{u,v}]_{u,v=1,2,\ldots,p},$$

where

$$\sigma_{uv} = \mathcal{C}[X_u, X_v],$$

the covariance of $X_u$ and $X_v$. In particular, $\mathcal{C}[X_v, X_v] = \mathcal{V}[X_v]$, the variance of $X_v$.

The *principal components* of $\mathbf{\Sigma}$ are defined as uncorrelated linear combinations of maximal variance. A linear combination, say LC, of the $p$ variables is $\mathbf{a'X}$, that is

$$LC = \mathbf{a'X} = a_1\mathbf{X}_1 + a_2\mathbf{X}_2 + \ldots + a_p\mathbf{X}_p.$$

Here the vector $\mathbf{a}$ is a vector of scalars $a_1, a_2, \ldots, a_p$:

$$\mathbf{a}' = (a_1\ a_2\ \ldots\ a_p).$$

These are the coefficients in the linear combination. Such linear combinations are called *variates.*

We have

$$\mathcal{V}[LC] = \mathcal{V}[\mathbf{a'X}] = \mathbf{a'\Sigma a}.$$

This is estimated as $\mathbf{a'Sa}$. This is to be maximized over $\mathbf{a}$. The derivative is

$$\partial\mathbf{a'Sa}/\partial\mathbf{a} = \mathbf{Sa}.$$

is A constraint is required for meaningful maximization. A reasonable such constraint $\mathbf{a'a} = 1$, which is equivalent to the length of $\mathbf{a}$, the quantity $\sqrt{\mathbf{a'a}}$, being equal to 1.

The Lagrangian function incorporating the constraint is

$$L(\mathbf{S}, \mathbf{a}; \lambda) = \mathbf{a'Sa} + \lambda(1 - \mathbf{a'a}).$$

The partial derivatives are

$$\partial L/\partial\mathbf{a} = 2\mathbf{Sa} - 2\lambda\mathbf{a}$$

and

$$\partial L/\partial\lambda = \partial\lambda(1 - \mathbf{a}'a)/\partial\lambda = 1 - \mathbf{a'a}.$$

Setting these equal to zero gives the simultaneous linear equations

$$\mathbf{Sa} = \lambda\mathbf{a}, \mathbf{a'a} = 1.$$

The first is the equation

$$\mathbf{Sa} - \lambda\mathbf{a} = \mathbf{0},$$

the zero vector. This is the homogeneous equation

$$(\mathbf{S} - \lambda\mathbf{I})\mathbf{a} = \mathbf{0}.$$

For nontrivial solutions, we must have $\det(\mathbf{S} - \lambda\mathbf{a}) = 0$. This is a polynomial equation of degree $p$ in $\lambda$; denote the roots by $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_p$. These are the *eigenvalues.* Their sum is the trace of $\mathbf{S}$; their product is the determinant of $\mathbf{S}$.

The corresponding eigenequations are

$$\mathbf{S}\mathbf{a}_j = \lambda_j \mathbf{a}_j, \quad j = 1, 2, \ldots, p.$$

The $j$-th PC (principal component), $C_j$, is the linear combination of the form

$$C_j = \mathbf{a}_j'\mathbf{x} = a_{1j}x_1 + a_{2j}x_2 + \cdots + a_{pj}x_p,$$

where $\mathbf{a}_j' = (a_{1j}, a_{2j}, \ldots, a_{pj})$. That is to say, for $j = 1, 2, \ldots, p$, the value of the $j$-th PC for Individual $i$ is $\mathbf{a}_j'\mathbf{x}_i$, $i = 1, 2, \ldots, n$.

The equations for the PCs in terms of the $X$s are $\mathrm{PC}_j = \mathbf{a}_j'\mathbf{X}$, $j = 1, 2, \ldots, p$. Let $\mathbf{C}$ be the $p$-vector of PCs. Then $\mathbf{C} = \mathbf{A}'\mathbf{X}$, where $\mathbf{A} = [\mathbf{a_1}\,\mathbf{a_2}\,\ldots\,\mathbf{a_p}]$ is the matrix whose columns are the eigenvectors. The inverse relation is

$$\mathbf{X} = \mathbf{A}'^{-1}\,\mathbf{C} = \mathbf{L}\mathbf{C},$$

where

$$\mathbf{L} = \mathbf{A}'^{-1},$$

where $\mathbf{L}$ is the matrix of *loadings* of the $X_v$ on the PCs $C_j$. Actually, $\mathbf{A}$ is an orthonormal matrix (its columns are of length one and are pairwise orthogonal), so $\mathbf{A}^{-1} = \mathbf{A}'$. Thus $\mathbf{L} = \mathbf{A}$. So

$$\mathbf{X} = \mathbf{A}'^{-1}\,\mathbf{C} = \mathbf{A}\mathbf{C}.$$

Letting $\mathbf{a}^{(v)'}$ be the $v$-th row of the matrix $\mathbf{A}$, that is

$$\mathbf{a}^{(v)'} = (a_{v1}, a_{v2}, \ldots, a_{vp}),$$

we have

$$X_v = a_{v1}C_1 + a_{v2}C_2 + \cdots + a_{vp}C_p.$$

In terms of the first $k$ PCs, this is

$$X_v = a_{v1}C_1 + a_{v2}C_2 + \cdots + a_{vk}C_k + \varepsilon_v, \quad (*)$$

where the error $\varepsilon_v$ is

$$\varepsilon_v = a_{v\,k+1}C_{k+1} + a_{v\,k+2}C_{k+2} + \cdots + a_{vp}C_p.$$

The covariance matrix can be represented as

$$\mathbf{S} = \sum_{j=1}^{p} \lambda_j \mathbf{a}_j \mathbf{a}_j'.$$

Correspondingly, the best rank $k$ approximation to $\mathbf{S}$ is

$$\mathbf{S}^{(k)} = \sum_{j=1}^{k} \lambda_j \mathbf{a}_j \mathbf{a}_j'.$$

Recall that for a symmetric matrix such as a covariance matrix, the eigenvalues are non-negative.

## 2.3 Ad Hoc Procedures for Determining an Appropriate Number of PCs

### 2.3.1 Procedure Based on the Average Eigenvalue

The average eigenvalue is

$$\bar{\lambda} = \sum_{j=1}^{p} \lambda_j / p.$$

One rule for the number of PCs to retain is the retain those for which the eigenvalues are greater than $\bar{\lambda}$. When $\mathbf{S}$ is taken to be the sample *correlation* matrix, the trace is $p$ and the average eigenvalue $\bar{\lambda}$ is 1.

### 2.3.2 Procedure Based on Retaining a Prescribed Portion of the Total Variance

Another procedure is to retain a number of PCs sufficient to account for, say, 90% of the total variance, trace $\mathbf{S} = \sum_{j=1}^{p} \lambda_j$. Of course the figure ninety percent is somewhat arbitrary and it might be nice to have some somewhat more objective criteria.

### 2.3.3 Procedure Based on the Dropoff of the Eigenvalues

Another procedure is to plot $\lambda_1, \lambda_2, \dots, \lambda_p$ against $1, 2, \dots, p$. One then looks for an elbow in the curve and retains a number of PCs corresponding to the point before the leveling off of the curve, if it does indeed take an elbow shape. Such a plot is called a *scree* plot, “scree” being the debris at the foot of a glacier.

# 3 AIC and BIC for the Number of PCs

Let us see what a Gaussian model would imply. The maximum log likelihood for the model (*) approximating the $p$ variables in terms of $k$ PCs is $(2\pi|\hat{\mathbf{\Sigma}}_k|)^{-n/2} C(n, p, k)$, where $C(n, p, k)$ is a constant depending upon $n, p,$ and $k$ and $|\mathbf{\Sigma}_k|$ denotes the determinant of the residual covariance matrix $\mathbf{\Sigma}_k$.

The determinant of the covariance matrix is the product of the eigenvalues,

$$|\mathbf{\Sigma}| = \Pi_{j=1}^{p} \lambda_j.$$

For a model based on the first $k$ PCs, this is

$$\Pi_{j=1}^{k}\,\lambda_j.$$

The determinant of the residual covariance is $\Pi_{j=k+1}^{p}\lambda_j$. The model-selection criterion AIC—Akaike's information criterion [2–4]—is based on an estimate of the log cross-entropy of $K$ proposed models with a null model.

The Bayesian information criterion BIC [12] is based on a large-sample estimate of the posterior probability $pp_k$ of Model $k$, $k = 1, 2, \dots, K$.

More precisely, $\mathrm{BIC}_k$ is an approximation to $-2\ln pp_k$. These model-selection criteria (MSCs) are thus smaller-is-better criteria and take the form

$$MSC_k = -2\ln\max L_k + a(n)m_k, \quad k = 1, 2, \dots, K,$$

where $L_k$ is the likelihood for Model $k$, $a(n) = \ln n$ for $\mathrm{BIC}_k$, $a(n) = 2$ (not depending upon $n$) for $\mathrm{AIC}_k$ and $m_k$ is the number of independent parameters in Model $k$. Relative to BIC, AIC tends to favor models with a smaller number of parameters. Note that

$$pp_k \;\approx\; C\exp(-\,\mathrm{BIC}_k/2),$$

where $C$ is a constant. Thus BIC values can be converted to a scale of 0 to 1. This is done by exponentiating -$\mathrm{BIC}_k/2$, summing the values, and dividing by the sum.

For the PC model,

$$-2\;\ln\max\;L_k = n\;\ln\;\Pi_{j=k+1}^{p}\,\lambda_k = n\sum_{j=k+1}^{p}\ln\lambda_k.$$

The criteria can be written as

$$\mathrm{MSC}_k = \mathrm{Deviance}_k + \;\mathrm{Penalty}_k,$$

where $\mathrm{Deviance}_k = n\;\ln\max L_k$ is a measure of lack of fit and $\mathrm{Penalty}_k = \;a(N)m_k$. Inclusion of an additional PC is justified if the criterion value decreases, that is if $\mathrm{MSC}_{k+1} < MSC_k$. For PCs, this is

$$n\sum_{j=k+2}^{p}\ln\lambda_j + (k+1)a(n) < n\sum_{j=k+1}^{p}\ln\lambda_j + k\,a(n).$$

This is

$$a(n) < n\ln\lambda_{k+1} = \ln(\lambda_{k+1}^{n}),$$

or

$$\exp[a(n)] < \lambda_{k+1}^{n},$$

or

$$\lambda_{k+1} > \exp\,[a(n)/n]$$

or

**Table 1** Correlation matrix of 5 variables–LA heart data

```
Correlations: AGE, SYS, DIAS, WT, HT
           AGE      SYS     DIAS      WT
SYS      0.342
DIAS     0.354    0.835                     <= NOTE highest r of .835
WT      -0.009    0.261    0.308                  is btw SYS and DIAS
HT      -0.332   -0.088   -0.099    0.426   <= NOTE next highest r of .426
Cell Contents: Pearson correlation                is btw HT and WT
```

**Table 2** PCs of heart data

```
Principal Component Analysis: AGE, SYS, DIAS, WT, HT

Eigenanalysis of the Correlation Matrix
Eigenvalue    2.1894    1.5382    0.6617    0.4485    0.1621
Proportion     0.438     0.308     0.132     0.090     0.032
Cumulative     0.438     0.746     0.878     0.968     1.000

Variable         PC1       PC2       PC3       PC4       PC5
AGE           -0.394    -0.365     0.800    -0.269     0.005
SYS           -0.615     0.050    -0.342    -0.174     0.687
DIAS          -0.624     0.063    -0.291    -0.049    -0.721
WT            -0.252     0.616     0.373     0.642     0.078
HT             0.117     0.694     0.141    -0.695    -0.051
```

$$\lambda_{k+1} > \exp[-a(n)/n].$$

Thus for AIC, inclusion of the additional PC$_{k+1}$ is justified if $\lambda_{k+1}$ is greater than $\exp(-2/n)$.

For BIC, inclusion of an additional PC$_{k+1}$ is justified if $\lambda_{k+1} > \exp(\ln N/N)$ $= [\exp(\ln n)]^{1/n} = n^{1/n}$, which tends to 1 for large $n$. So this is in approximate agreement with the average eigenvalue rule for correlation matrices, stating that one should retain dimensions with eigenvalues larger than 1.

## 4 Example

Here we consider a sample from the LA Heart Study. See, e.g., [8]. The sample is $n = 100$ men. The variables include Age, Systolic blood pressure, Diastolic blood pressure, weight, height and Coronary Incident, a binary variable indicating whether or not the individual had a coronary incident during the course of the study. (Data on the same variables for another 100 men are also given in Dixon and Massey's book. Results can be compared and contrasted between the two samples.) Here we focus on the first five variables. Minitab statistical software was used for the analysis.

Table 1 is the lower-triangular portion of the correlation matrix for the five variables (Table 2).

### 4.1 Principal Component Analysis in the Example

Note that an eigenvector can be multiplied by −1, changing the signs of all its elements. Below, this is done with PC1 so that SYS and DIAS have positive loadings. Interpretations, BPtotal, SIZE, AGE, OVERWT, BPdiff, are given below the eigenvectors. The interpretations are based on which loadings are large and which are small. Taking .6 as a cut-off point, in PC1, SYS and DIAS have loadings above this, while the other variables have loadings less than this (in fact, less than .4), so PC1 can be interpreted asan index of total BP. In PC2, WT and HT have large loadings with the same sign, so PC2 can be interpreted as SIZE (Table 3).

As above, denote the eigensystem by

$$(\lambda_v, \boldsymbol{a}_v),\ v = 1, 2, \ldots, p.$$

Then the eigensystem equations are

$$\mathbf{S}\boldsymbol{a}_v \ = \ \lambda_v \boldsymbol{a}_v,\ v = 1, 2, \ldots, p.$$

Here $\mathbf{S}$ is taken to be the correlation matrix. Let $\mathbf{1}'_v \ = \ (0\,0 \cdots\ 1 \cdots\ 0 \cdots\ )$, the vector with 1 in the $v$-th position and zeroes elsewhere. The covariance between a variable $X_v$ and a PC $C_u$ is $\mathcal{C}[X_v,\, C_u\,] = \mathcal{C}[\mathbf{1}'_v\boldsymbol{X}, \boldsymbol{a}'_u\boldsymbol{X}] = \mathbf{1}'\Sigma\,\boldsymbol{a}_u \ = \ \mathbf{1}'_v\,\lambda_u\,\boldsymbol{a}_u \ = \ \lambda_u a_{uv}$, where $a_{uv}$ is the $v$-th element of the vector $\boldsymbol{a}_u$. The correlation is $\mathrm{Corr}\,[X_v,\, C_u\,] = \mathcal{C}[X_v,\, C_u\,]/SD[X_v]SD[C_u\,] \ = \ \lambda_u\, a_{uv}\,/\,\sigma_v\,\sqrt{\lambda_u} \ = \ \sqrt{\lambda_u}\,a_{uv}\,/\,\sigma_v$. When the correlation matrix is used, $\sigma_v = 1$, and this correlation is $\sqrt{\lambda_u}\ a_{uv}$. A correlation of size greater than .6 corresponds to 36% of variance explained. The variable $X_v$ has a correlation higher than .6 with the component $C_u$ if its loading in $C_u$, the value $a_{uv}$, is greater than $.6\,/\,\sqrt{\lambda_u}$. These values are appended to the table below. Loadings larger than this cut point are in boldface. (The cut-off of .6 is somewhat arbitrary; one might use, for example, a cut-off of .5.)

One can also focus on the pattern of loadings within the different PCs for interpretation of the PCs. To reiterate:

**Table 3** PC1 is multiplied by −1

| Variable | PC1 | PC2 | PC3 | PC4 | PC5 |
|---|---|---|---|---|---|
| AGE | 0.394 | -0.365 | 0.800 | -0.269 | 0.005 |
| SYS | 0.615 | 0.050 | -0.342 | -0.174 | 0.687 |
| DIAS | 0.624 | 0.063 | -0.291 | -0.049 | -0.721 |
| WT | 0.252 | 0.616 | 0.373 | 0.642 | 0.078 |
| HT | - 0.117 | 0.694 | 0.141 | -0.695 | -0.051 |
| Interpretations (edited in by SLS): | | | | | |
| | BPtotal | SIZE | AGEindex | OVERWT | BPdiff |

**Table 4** Loadings corresponding to correlations > .6 are boldface

| Variable | PC1 | PC2 | PC3 | PC4 | PC5 |
|---|---|---|---|---|---|
| AGE | 0.394 | − 0.365 | **0.800** | − 0.269 | 0.005 |
| SYS | **0.615** | 0.050 | − 0.342 | − 0.174 | 0.687 |
| DIAS | **0.624** | 0.063 | − 0.291 | − 0.049 | − 0.721 |
| WT | 0.252 | **0.616** | 0.373 | 0.642 | 0.078 |
| HT | − 0.117 | **0.694** | 0.141 | − 0.695 | − 0.051 |
| Eigenvalue, $\lambda$ | 2.1894 | 1.5382 | 0.6617 | 0.4485 | 0.1621 |
| Square root, $\sqrt{\lambda}$ | 1.48 | 1.24 | 0.81 | 0.67 | 0.40 |
| .6 / $\sqrt{\lambda}$ | 0.40 | 0 .48 | 0.74 | 0.90 | 1.50 |
| Interpretations | BPtotal | SIZE | AGE | OVERWT | BPdiff |

**Table 5** Estimating the number of PCs by various methods

| No. of PCs, $k$ | $\lambda_k$ | $\lambda_k > 1$? | $\ln \lambda_k$ | $N \ln \lambda_k$ | For BIC:<br>$N \ln \lambda_k > -4.61$? | For AIC:<br>$N \ln \lambda_k > -2$? |
|---|---|---|---|---|---|---|
| 1 | 2.19 | Yes | 0.78 | 78.36 | Yes | Yes |
| 2 | 1.54 | Yes | 0.43 | 43.06 | Yes | Yes |
| 3 | 0.66 | No | − 0.41 | − 41.29 | No | No |
| 4 | 0.45 | No | − 0.80 | − 80.18 | No | No |
| 5 | 0.16 | No | − 1.82 | − 181.95 | No | No |

PC1: SYS and DIAS have large loadings with the same sign; we interpret PC1 as BPinex or BPtotal.

PC2: WT and HT have large loadings of the same sign; we interpret PC2 as the man's SIZE.

PC3: Only AGE has a large loading; we interpret PC3 as AGE.

PC4: WT and HT have large loadings with opposite signs; we interpret PC4 as OVERWEIGHT.

PC5: SYS and DIAS have large loadings with opposite signs; we interpret PC5 as BPdrop.

I continue to marvel at how readily interpretable the PCs are. And, this is even without using a factor analysis model and using rotation (Table 4).

### 4.2 Employing the Criteria in the Example

Table 5 shows the eigenvalues and the results according to the various criteria. According to the rule based on the average eigenvalue, the dimension is retained it its eigenvalue is greater than 1 (for a correlation matrix). For BIC, the $k$-th PC is retained if $n \ln \lambda_k > -a(n)$, where $a(n) = \ln n$. Here, $n = 100$ and $\ln n = \ln 100$, approx. 4.61. For AIC, the $k$-th PC is retained if $n \ln \lambda_k > -2$. In this example, the methods agree on retaining $k = 2$ PCs.

I feel that I should remark that, though this is the case, the fourth and fifth PCs do have simple and interesting interpretations. It is just that they do not improve the fit very much.

## 5 Discussion

The focus here has been on determining the number of dimensions needed to represent a complex of variables adequately.

### 5.1 Regression on Principal Components

Given a response variable $Y$ and explanatory variables $X_1, X_2, \ldots, X_p$, one may transform the $X$s to their principal components, as this may aid in the interpretation of the results of the regression. In such regression on principal components (see, e.g., [10]), however, one should not necessarily eliminate the principal components with small eigenvalues, as they may still be strongly related to the response variable. The Bayesian information criterion is

$$BIC_k = -2LL_k + m_k \ln n,$$

for alternative models indexed by $k = 1, 2, \ldots, K$, where $LL_k$ is the maximum log likelihood for Model $k$ and $m_k$ is the number of independent parameters in Model $k$. For linear regression models with Gaussian-distributed errors BIC takes the form

$$BIC_k = n \ln MSE_k + m_k \ln n$$

where $MSE_k$ is the MLE (maximum likelihood estimate) of the MSE (mean squared error) of Model $k$, with divisor $n$, of the error variance. With $p$ explanatory variables, there are $2^p$ alternative models (including the model where no explanatory variables are used and the fitted value of $Y$ is simply $\bar{y}$). It would usually seem to be wise to evaluate all $2^p$ models using $BIC_k$ rather than reducing the number of principal components by just looking at the explanatory variables.

### 5.2 Some Related Recent Literature

Some various applications involving choosing the number of principal components from recent literature include the following. The method presented here could possibly be applied in these applications. For example, a good book on the topic of model selection and testing covering all aspects is Bhatti et al. [7]. In recent years econometricians have examined the problems of diagnostic testing, specification testing, semiparametric estimation and model selection. In addition, researchers have considered whether to use model testing and model selection procedures to decide upon the models that best fit a particular dataset. This book explores both issues with application to various regression models, including arbitrage pricing theory models. Along the lines of model-selection criteria, the book references, e.g., Schwarz [12], the foundational paper for BIC.

Next we mention some recent papers which show applications of model selection in various research areas.

One such paper is Xu et al. [14] an application of principal components analysis and other methods to water quality assessment in a lake basin in China,

Another is Omuya et al. [11], on feature selection for *classification* using principal component analysis.

As mentioned, a particularly interesting application of principal components analysis is in regression and logistic regression. We have mentioned the paper by Massy [10] on using principal components analysis in regression. Another is Aguilera et al. [1] on using principal components in *logistic* regression.

## 6 Conclusions

The information criteria AIC and BIC have been applied here to the choice of the number of principal components to represent a dataset. The results have been compared and contrasted with criteria such as retaining those principal components which explain more than an average amount of the total variance.

**Acknowledgements** There are no further acknowledgements.

**Author Contributions** SLS is the sole author.

**Funding** There was no funding other than the author’s usual salary at the university.

**Availability of Data and Material** The source of data used is a book that is referenced and available.

## Declarations

**Conflict of interest** There are no competing interests.

## References

1. Aguilera, A.M., Escabias, M., Valderrama, M.J.: Using principal components for estimating logistic regression with high-dimensional multicollinear data. Comput. Stat. Data Anal. **50**(8), 1905–1924 (2006)
2. Akaike, H.: Information theory and an extension of the maximum likelihood principle. In: Petrov, B.N., Csáki, F. (eds.) 2nd International Symposium on Information Theory, Tsahkadsor, Armenia, USSR, September 2–8, 1971, pp. 267–281. Akadémiai Kiadó, Budapest (1973). [Republished in Kotz, S., Johnson, N.L. (eds.) (1992) Breakthroughs in Statistics, I. Springer, pp. 610–624 (1973)]
3. Akaike, H.: A new look at the statistical model identification. IEEE Trans. Autom. Control 19(6):716–723.(1974). https://doi.org/10.1109/TAC.1974.1100705, MR 0423716
4. Akaike, H.: Prediction and entropy. In: Atkinson, A.C., Fienberg, S.E. (eds.) A Celebration of Statistics, pp. 1–24. Springer, New York (1985)
5. Anderson, T.W.: An Introduction to Multivariate Statistical Analysis, 3rd edn. Wiley, New York (1958) [Wiley, Hoboken, NJ, 2002]
6. Bai, Z., Choi, K.P., Fujikoshi, Y.: Consistency of AIC and BIC in estimating the number of significant components in high-dimensional principal component analysis. Ann. Stat. **46**(3), 1050–1076 (2018). https://doi.org/10.1214/17-AOS1577
7. Bhatti, M.I., Al-Shanfari, H., Zakir Hossain, M.: Econometric Analysis of Model Selection and Model Testing. Routledge, London (2017)
8. Dixon, W.J., Massey, F.J., Jr.: Introduction to Statistical Analysis, 3rd edn. McGraw-Hill, New York (1969)
9. Johnson, R.J., Wichern, D.W.: Applied Multivariate Statistical Analysis, 6th edn. Pearson, Upper Saddle River (2008)
10. Massy, W.F.: Principal components regression in exploratory statistical research. J. Am. Stat. Assoc. **60**(309), 234–256 (1965). https://doi.org/10.1080/01621459.1965.10480787
11. Omuya, E.O., Okeyo, G.O., Kimwele, M.W.: Feature selection for classification using principal component analysis and information gain. Expert Syst. Appl. **174**, 114765 (2021)
12. Schwarz, G.: Estimating the dimension of a model. Ann. Stat. **6**, 461–464 (1978). Stable URL: http://www.jstor.org/stable/2958889
13. Sclove, S.L.: Application of model-selection criteria to some problems in multivariate analysis. Psychometrika **52**(1987), 333–343 (1987). https://doi.org/10.1007/BF02294360
14. Xu, S., Cui, Y., Yang, C., Wei, S., Dong, W., Huang, L., Liu, C., Ren, Z., Wang, W.: The fuzzy comprehensive evaluation (FCE) and the principal component analysis (PCA) model simulation and its applications in water quality assessment of Nansi Lake Basin, China. Environ. Eng. Res. **26**(2), 222–232 (2021)