Chapter 4

Some Methods for Evaluating Performance of Management Information System

Khu Phi Nguyen and Hong Tuyet Tu

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.74093

Abstract

Recently, several kinds of information systems are developed for purposes and needs of business and play an important role in business organizations and management operations. Management information system, or MIS for short, is a kind of information system. It is a key factor to facilitate and attain efficient decision-making in an organization. Its performance relates to many other information systems, for instance, DSS or decision support system, SIS or strategic information system, etc. Methods of testing statistical hypotheses concerning the performance of MIS are absolutely essential to support management activities and decision-making.

Keywords: management information systems, information theory, rough set theory, decision-making process, ANOVA

1. Introduction

A *system* is a set of interrelated components assembled to accomplish certain objectives or goal. Basic characteristics of a system are highlighted as boundaries, interfaces, input-outputs, and methods of making outputs from inputs. The environment of a system includes people, organizations, and other systems that supply data to or receive data from the system.

Solving problems comes from a system that usually uses the method of *systems approach* taking into account the goals, environment, and internal workings of the system. This method involves the following steps:

i. Define the problem and collect data for the problem.

ii. Identify and evaluate feasible solutions.

iii. Select the best solution and determine whether the solution is working.

An *information system* (IS) consists of components such as hardware, software, databases, personnel, and procedures that managers can use to make better decisions in control business operations. ISs are also used to document and monitor the operations of some other systems, called target systems that are prerequisite for the existence of ISs. On side of infrastructure, information system is an integration of diverse computers, displays and visualizations, database, storage systems, instruments, sensors, etc. via software and networks to share data and to provide aggregate capabilities.

In business operation, the activities of an organization equipped with IS are usually of three kinds: operational, tactical, and strategic planning. In this context, a *strategy* is meant as determination of the basic long-term goals and objectives of an enterprise and the adoption of courses of action and the allocation of resources necessary for achieving these goals. *Operational tasks* are the daily activities of the firm in consuming and acquiring resources. These daily transactions produce basis data for the operational systems.

ISs that provide information for allocation of efficient resources to achieve business objectives are known as *tactical systems*. Tactical systems provide middle-level managers with the information they need to monitor and control operational tasks and to allocate their resources effectively. The time frame for tactical activities may be monthly, quarterly, or yearly. Alternatively, ISs that support the strategic plans of the business are known as *strategic planning systems*. These systems are designed to provide top managers with information that assists them in making long-term planning decisions.

Both of the strategic planning information systems and tactical information systems may use the same data source, so the distinction between them is not always clear. For example, middle-level and top managers use budgeting information to allocate reasonable resources or to plan the long-term or short-term activities, budgeting becomes a tactical decision activity or a strategic planning activity, respectively. Hence, the differences between systems are attributed to whom and what the budgeting data are used.

The top management of the organization carries out strategic planning based on results of operational tasks, tactical systems, and related external information to decide whether to build new plants, new products, facilities, or invest in technology. For making these decisions, strategic planners have to address problems that involve long-range analysis and prediction. The time frame for strategic activities may be months or years.

Some basic business systems that serve the operational level of the organization are called *transaction processing systems* or TPS for short. A TPS that records the daily routine transactions necessary to the conduct of the business monitor and control system physical processes is called *process control system* or PCS. For example, a wastewater treatment plan uses electronic sensors linked to computers to monitor wastewater processes continually and control the water quality process [1]. Similarly, a petroleum refinery uses sensors and computers to monitor chemical processes and make real-time controls to the refinery process. A process control system comprises the whole range of equipment, computer programs, and operating procedures [2].

Knowledge-based IS that supports the creation, organization, and dissemination of business knowledge to employees and managers throughout a company is named as *knowledge management system*. In such a case, knowledge management is the deployment of a comprehensive system that enhances the growth of knowledge. Expert systems are the category of artificial intelligence which has been used most successfully in building commercial applications. An expert system is also considered as a knowledge-based system that provides expert advice and act as expert consultants to users.

A decision support system (DSS) is a computer-based system intended for use by a particular manager or a team of managers at any organizational level in making a decision in the process of solving a semi-structured decision. Database-based management system and a user interface are major components of a DSS. The database consists of information related to production information, market and marketing information, research data, financial transactions, and so forth.

The decision-maker must have suitable knowledge and skills on mining these systems of DSS to address the problem arising and make effective decisions. In traditional approaches to decision-making, usually scientific expertise together with statistical descriptions is needed to support decision-making. Recently, many innovative facilities have been proposed for decision-making process in enterprises with huge databases, together with several heuristic models.

Management information systems (MIS) are a kind of computer ISs that could collect and process information from different sources to make decisions in level of management [3]. This level contains computer systems that are intended to assist operational management in monitoring and controlling the transaction processing activities that occur at clerical level. MIS provides information in the form of prespecified formats to support business decision-making. The next level in the organizational hierarchy is occupied by low-level managers and supervisors. Therefore, MIS takes internal data from the system and summarized it to meaningful and useful forms as management reports to use it to support management activities and decision-making.

MISs encompass a complex and broad topic, that is why, MIS boundaries need to be defined to reduce difficulties in system managing. Firstly, MIS contains a vast number of related activities, so it is hard to review all of them. It may discuss on a selected sample of activities, depending on objectives and viewpoint of researcher. Alternatively, it only focuses on farm levels or on some lesser extent systems enough for researchers addressing problems. Secondly, MISs can be defined and described in several frameworks. Only a few of these frameworks are used to discuss important subject matters. Lastly, MISs are developed as a sense of how these systems have evolved, adapted, and been refined as new technologies have emerged, changing economic conditions, etc.

To evaluate performance of MIS, its output data must be characterized in a set of basic *features* appropriate to functions, objectives, and goals of the system. These output data need to be observed repetitively to evaluate the extent to which MIS is implemented to make successful decisions in organization. Using these observations, methods of data mining in rough set point of view, statistical analysis, etc. can be applied to evaluate the extent to which MISs are used to make effective decisions in planning purposes [4–7].

2. Evaluation of features and making decision rules

In mathematical modeling, an IS can be modeled by a sample $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ of n objects ω_i where i = 1,2,…, n. The ith object ω_i is observed by instances of m *conditional features* $f_1, f_2, \ldots, f_m$, valued as $f_j(\omega_i)$ j = 1,2,…, m. Additionally, a feature d characterizes a specific effect of ω_i denoted by $d(\omega_i)$, the so-called *decision feature*. In case of having s effects for a decision, d is represented by values $d(\omega_i) = d_k$ with $k \in \{1,2,\ldots, s\}$.

Let $F = \{f_1, f_2, \ldots, f_m\}$, then $(\Omega, F \cup \{d\})$ is a *decision information table* or DIT with $n = |\Omega|$ objects, $m = |A|$ conditional features, and a decision d. Objects ω and ω' are *indiscernible* if and only if the following binary relation R_F on Ω with respect to (w.r.t.) F is satisfied:

$$R_A: f_j(\omega) = f_j(\omega') \; j = 1, 2, \ldots, m \tag{1}$$

This is an *equivalence relation*. Equivalent class of $\omega \in \Omega$ with respect to (w.r.t.) F is:

$$[\omega]_F = \{\omega' \in \Omega \mid f_j(\omega') = f_j(\omega) \;\; j = 1, 2, \ldots, m\} \tag{2}$$

Assume that there are r such equivalence classes and named by $C_1, C_2, \ldots, C_r$. They are disjoint subsets and form a partition of Ω by R_F. Similarly, for the decision feature d, another partition of Ω is $D_1, D_2, \ldots, D_s$ defined by the following equivalence relation:

$$R_d : d(\omega) = d_k, k = 1, 2, \ldots, s \tag{3}$$

Here, $D_k = \{\omega' \in \Omega \mid d(\omega') = d_k\}$ is an equivalence classes called the kth *decision class* of the DIT. If $f(D_k) = |D_k|/n$ be frequency of D_k w.r.t Ω, *information entropy* H(d) of decision feature d is

$$H(d) = -\sum_{k=1}^{s} f(D_i) \log_2 f(D_k) \tag{4}$$

On the other hand, let $f(C_i) = |C_i|/n$ be frequency of C_i and $f(D_k \mid C_i) = |D_k \cap C_i|/|C_i|$ conditional frequency of D_k conditioned C_i. The *conditional entropy* H(d|F) of the decision feature d w.r.t condition F is determined by

$$H(d|F) = -\sum_{i=1}^{r} f(C_i) \sum_{k=1}^{s} f(D_k|C_i) \log_2 f(D_k|C_i) \tag{5}$$

From Eqs. (4) and (5), the *mutual information* I(F, d) between F and d is given by

$$I(F, d) = H(d) - H(d|F) \tag{6}$$

The mutual information is nonnegative and symmetric, i.e. I(F, d) = I(d, F). In this case, the *significance* of feature $f \in F$ w.r.t d is defined as

$$Sgnf(f, d) = I(F, d) - I(F - \{f\}, d) \tag{7}$$

The significance of feature a represents the dependency of decision attribute d relative to condition attribute f. This measure reflects the discrimination ability of condition attributes.

The larger Sgnf(f, d), the more stronger of dependency relationships between a and decision attribute d. if Sgnf(f, d) > 0, then f is a *core feature* of DIT or f satisfies

$$I(F - \{f\}, d) < I(F, d) \tag{8}$$

Any core feature is significant and may not be eliminated in mining DIT. Let CFs be a set of all core features, CFs ⊆ F. To find CFs, each feature in F must be verified using Eq. (8) to whether or not include it to CFs.

Example 1: To analyze some features of a service, **Table 1** illustrated a DIT consists of evaluations of nine clients on four features of the service. In which, d is the decision feature, f_1: capacity for innovation; f_2: service capability; f_3: product technologies; and f_4: solution, are conditional features. Values in **Table 1** mean, 0: unpleased, 1: acceptable, and 2: very pleased.

Here, F = {f_1, f_2, f_3, f_4}. Using Eq. (1), four equivalence classes w.r.t F are C_1 = {ω_1, ω_8}, C_2 = {ω_2, ω_7}, C_3 = {ω_3, ω_5, ω_9}, C_4 = {ω_4, ω_6} and from Eq. (3) two decision classes D_0 = {ω_2, ω_5, ω_7, ω_9}, D_1 = {ω_1, ω_3, ω_4, ω_6, ω_8}. From Eq. (4), the information entropy of decision feature d is H(d) = 0.9911 and H(A) = 0.4976. From Eq. (5), the conditional entropy of d is H(d|F) = 0.3061, so the mutual information between F and d is I(F, d) = 0.6850.

	f_1	f_2	f_3	f_4	d
(a) Original data table					
ω_1	1	1	1	0	1
ω_2	0	1	1	0	0
ω_3	1	0	1	2	1
ω_4	1	2	0	1	1
ω_5	1	0	1	2	0
ω_6	1	2	0	1	1
ω_7	0	1	1	0	0
ω_8	1	1	1	0	1
ω_9	1	0	1	2	0
(b) Sorted data table					
ω_2	0	1	1	0	0
ω_5	1	0	1	2	0
ω_7	0	1	1	0	0
ω_9	1	0	1	2	0
ω_1	1	1	1	0	1
ω_3	1	0	1	2	1
ω_4	1	2	0	1	1
ω_6	1	2	0	1	1
ω_8	1	1	1	0	1

Table 1. A decision information system for evaluation service quality.

If the first feature a_1 is eliminated, it is obtained the same H(d), but $H(F - \{f_1\}) = 0.5144$ and $H(d|F - \{f_1\}) = 0.7505$. These imply $I(F - \{f_1\}, d) = 0.2405 < I(F, d)$ and the a_1, capacity for innovation is a core feature. But, $Sgnf(f_4, d) = Sgnf(F, d) - Sgnf(F - \{f_4\}, d) = 0$, so f_4 may be eliminated since it is not significant.

The features F, d can be considered as random quantities with values are represented in rows of a DIT. In theory of information, the mutual information is a measure of average information this random quantity receives from that one in all one's conditions and vice versa. Therefore, I(F, d) measures quantity of average information that the decision feature d receives from conditional features w.r.t. decisional value of d. That is why, it is concerned to the problem of removing redundant conditional features so that the reduced set provides the same effect, e.g., the same quality of classification or decision as the original.

A *coeffect reduced set* R of conditional features set is a subset of A so that I(R, d) = I(F, d), i.e., R contains some conditional features having the same effect as F. Any coeffect reduced set or *reduced set* of F for short can be used as the whole F. An algorithm to find a reduced set R based on mutual information is as follows:

ALGORITHM MIBR // *Mutual Information Based Reduced set.*

// Input: DIT = $(\Omega, F \cup \{d\})$.

// Output: R // *a reduced set of* F.

$S := \varnothing$; $R := CFs$; // *set of core features.*

Repeat.

$S := R$; for any $f \in F - R$, if $I(R \cup \{f\}, d) > I(S, d)$ then $S := R \cup \{f\}$;

$R := S$; // *reassign before doing the next iteration.*

Until I(R, d) = I(F, d);

Example 2: Using data in **Table 1**, the above algorithm is done as follows.

Firstly, $R = CFs = \{f_1\}$, $S = R$ then

i. $f_2 \in F - R$, then $I(R \cup \{f_2\}, d) = 0.6850 > I(S, d) = 0.3198$, so $S = R \cup \{f_2\} = \{f_1, f_2\}$;

ii. $f_3 \in F - R$, $I(R \cup \{f_3\}, d) = 0.6850 = I(S, d)$, S does not change;

iii. $f_4 \in F - R$, $I(R \cup \{f_4\}, d) = 0.6850 = I(S, d)$, S does not change;

$R = S = \{f_1, f_2\}$. By checking, $I(R, d) = 0.6850 = I(F, d)$, the iteration is terminated. It is obtained $R = \{f_1, f_2\}$ is a reduced set of F.

It is noticed that, if the two steps i and ii of the previous treatment are permuted, then the set $R = \{f_1, f_3\}$ is another reduced set of F.

Remark: As shown above, reduced set R of DIT is not unique. Finding minimum reduced set of DIT is an optimization problem. Several algorithms have been proposed to solve this problem, e.g., algorithm of rough set-based feature selection based on ant colony optimization (RSFSACO) in [8], cf. [9], for more detail.

Given X, a subset of Ω in a DIT, *low-approximation* or *upper-approximation* of X w.r.t. F respectively named as L_FX or U_FX, is defined by:

$$L_FX = \{\omega \in \Omega \mid [\omega]_F \subseteq X\}, U_FX = \{\omega \in \Omega \mid [\omega]_F \cap X \neq \varnothing\} \tag{9}$$

It can be shown that $L_FX \subseteq X \subseteq U_FX$. Some other relations between these approximations have been illustrated, e.g., in [5]. The difference set $B_FX = U_FX - L_FX$ is called a *boundary* of X and $\Omega - U_FX$ is the outside region of X. X is a *rough set* if $B_FX \neq \varnothing$, otherwise a *crisp set*.

Example 3: In Example 1, let $X = \{\omega_1, \omega_3, \omega_5, \omega_7, \omega_9\}$. Then, the approximations of X are $L_FX = \{\omega_3, \omega_5, \omega_9\} = C_3$ and $U_FX = \{\omega_1, \omega_2, \omega_3, \omega_5, \omega_7, \omega_8, \omega_9\} = C_1 \cup C_2 \cup C_3$. The boundary $B_FX = \{\omega_2, \omega_8, \omega_9\}$ differs from empty set, so X is a rough set and C_4 is the outside region of X. **Figure 1** shows all these sets w.r.t in Ω.

Any decision class Ω_k in Ω/R_d is subset of Ω, so it has a low approximation $L_F\Omega_k$. Hence, *positive region* in Ω w.r.t d, f is the following subset:

$$P_i(F) = \cup_{k=1}^{s} L_F\Omega_k \tag{10}$$

In data analysis, the dependence between attributes is important. The *dependency* of the decision feature d on the conditional features F is defined by the following ratio:

$$Dep(d, F) = |P_d(F)|/|\Omega| \tag{11}$$

By definition, $0 \leq Dep(d, F) \leq 1$ and if $Dep(d, F) = 1$, d depends *totally* on F. If $Dep(d, F) = 0$, i.e., $P_d(F) = \varnothing$, then d does not depend on F. In case of $0 < Dep(d, F) < 1$, d depends *partially* on F.

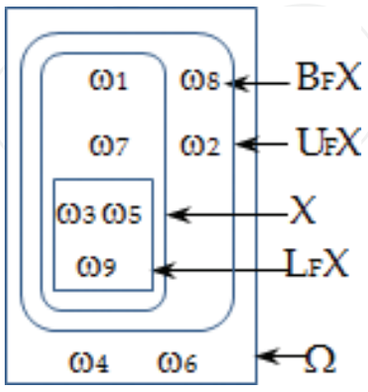

Figure 1. *Approximations of* X.

Using the degree of dependency, a coeffect reduced set R of conditional features in a DIT can also be found by meaning of Dep(d, R) = Dep(d, F).

Example 4: Example 1 gives two decision classes $D_0 = \{\omega_2, \omega_5, \omega_7, \omega_9\}$, $D_1 = \{\omega_1, \omega_3, \omega_4, \omega_6, \omega_7, \omega_8\}$; low approximations of these classes are $L_FD_0 = \{\omega_2, \omega_7\}$, $L_FD_1 = \{\omega_1, \omega_4, \omega_6, \omega_8\}$ thus $P_d(F) = \{\omega_1, \omega_2, \omega_4, \omega_6, \omega_7, \omega_8\}$ and the degree of dependency or quality of approximation is Dep(d, F) = 1/3. Using the coeffect reduced set $R = \{f_1, f_2\}$, it can be shown that all equivalence classes w.r.t R are the same ones in Example 1. Therefore, the above low approximations and positive region are also the same, i.e., $L_RD_0 = L_FD_0$, $L_RD_1 = L_FD_1$ and $P_d(R) = P_d(F)$.

So far, problems of inducing rules from DITs have been studied and developed. The rough set method can be applied to the problems with several advantages [5]. For instance, the lower and upper approximations are applied to describe the inconsistency of a DIT and to induce corresponding rules dynamically from decision systems [6]. These methods of approximation can be used to address incomplete input data for inducing decision rules [7]. Such rules can be applied to partition a set of objects into classifications [10].

Given a DIT, let V_f be the range of $f \in F$, for a $v \in V_f$, $\omega \in \Omega$ a proposition like $f(\omega) = v$ or $f = v$ for short, takes a logic value true or false depending on ω. Assignment, $\phi := f = v$ is to define a logic variable ϕ w.r.t the proposition $f = v$. Then, ϕ is true if there exists $\omega \in \Omega$ so that $f(\omega) = v$ or false in vice versa. Set of logic variables on F and logical operations, like ~: not; ∧: and; ∨: or; set up a set of logic expressions called *decision language* from F, denoted by $\mathcal{L}(F)$. The meaning of ϕ in $\mathcal{L}(F)$, denoted by $\langle\phi\rangle$, is a set of ω in Ω so that the proposition ϕ is true. Additionally, if $\phi := f = v$ then $\langle\phi\rangle = \{\omega \in \Omega / f(\omega) = v\}$, so ϕ takes the set $\langle\phi\rangle$ as its *description*.

A *decision rule* allows individual, team workers, and organization choose effectively specific course of action in response to opportunities and threads and help. Formally, a decision rule is a logic expression defined by proposition $\phi \rightarrow \psi$, read "*if* ϕ *then* ψ", where $\phi \in \mathcal{L}(F)$ and $\psi \in \mathcal{L}(d)$ referred to as *condition* and *decision* of the rule, respectively. A decision rule $\phi \rightarrow \psi$ is true if $\langle\phi\rangle \subseteq \langle\psi\rangle$. Both ϕ andψ are equivalent written as $\phi \leftrightarrow \psi$, if and only if $(\phi \rightarrow \psi) \wedge (\psi \rightarrow \phi)$.

Assume that $\langle\phi\rangle$ and $\langle\psi\rangle$ are nonempty. The *support* of the rule $\phi \rightarrow \psi$ is defined as

$$\text{Supp}(\phi \rightarrow \psi) = |\langle\phi\rangle \cap \langle\psi\rangle| \tag{12}$$

The larger Supp($\phi \rightarrow \psi$), the more power of the rule in DIT. When $|\langle\phi\rangle| \neq \varnothing$, the *certainty* or *accuracy* of $\phi \rightarrow \psi$ denoted by Cert(ϕ,ψ) is

$$\text{Cert}(\phi \rightarrow \psi) = |\langle\phi\rangle \cap \langle\psi\rangle| / |\langle\phi\rangle| \tag{13}$$

This is a percentage objects of $\langle\psi\rangle$ presented in $\langle\phi\rangle$ or percent of objects having property ψ in the set of objects having property ϕ or Cert($\phi \rightarrow \psi$) shows the *confidence* of the rule. In consequences, Cert($\phi \rightarrow \psi$) = 1 is equivalent to $\phi \rightarrow \psi$ is true, the rule is certain or accurate. Alternatively, if $|\langle\psi\rangle| \neq \varnothing$ the *coverage* of $\phi \rightarrow \psi$ is also defined:

$$\mathrm{Covg}(\phi \to \psi) = \left|\langle\phi\rangle \cap \langle\psi\rangle\right| / \left|\langle\psi\rangle\right| \tag{14}$$

The smaller of Covg($\phi \to \psi$), the less power of the rule. Finally, the popularity of $\phi \to \psi$ is measured by the *strength* of the rule:

$$\mathrm{Strg}(\phi \to \psi) = \left|\langle\phi\rangle \cap \langle\psi\rangle\right| / |\Omega| \tag{15}$$

In a given DIT, a coeffect reduced set R of conditional features and corresponding positive region $P_d(R)$ are set up. Then, the DIT is restricted to a new table with features R, d and $P_d(R)$. Such a table is called *decision support table* or DST. Based on the above measures, decision rules extracted from DST are verified before using them in prediction decisions.

It is noted that, there may be pairs of *inconsistent* or *conflicting* decision rules which have the same conditions but different decisions. Such conflicting rules must be excluded. In general, set $\mathfrak{R}$ of τ decision rules $\phi_\alpha \to \psi_\alpha$ selected need to meet the properties:

1. Each $\phi_\alpha \to \psi_\alpha$ in $\mathfrak{R}$ is *admissible*, $\mathrm{Supp}(\phi_\alpha \to \psi_\alpha) \neq 0$,
2. $\mathfrak{R}$ *covers* Ω or $\left|\cup_{\alpha=1}^{\tau}\langle\phi_\alpha\rangle\right| = \left|\cup_{\alpha=1}^{\tau}\langle\psi_\alpha\rangle\right| = |\Omega|$,
3. $\mathfrak{R}$ consists of pairs *mutually independent*, i.e., for $\phi_\alpha \to \psi_\alpha$, $\phi_\beta \to \psi_\beta \in \mathfrak{R}$, it is obtained that $\langle\phi_\alpha\rangle \cap \langle\phi_\beta\rangle = \varnothing$ and $\langle\psi_\alpha\rangle \cap \langle\psi_\beta\rangle = \varnothing$,
4. $\mathfrak{R}$ preserves the *consistency*: $\cup_{i=1}^{\tau} L_F D_i = \left\langle \vee_{\alpha=1}^{\tau} \phi_\alpha \right\rangle$.

Example 5: A coeffect reduced set, e.g., R = {f_1, f_2}, and positive region determined by $P_d(R) = \{\omega_1, \omega_2, \omega_4, \omega_6, \omega_7, \omega_8\}$ as in Example 4. Some decision rules are extracted from **Table 1** and measures of obtained rules are presented in **Table 2**. The supports of the 2nd and 3rd rules are 2, their certainties and strengths are equal to 1 and 22.2%. So, they can be combined together:

$$(f_1 = 1) \wedge [(f_2 = 1) \vee (f_2 = 2)] \to d = 1 \tag{16}$$

The support of this rule is raised to 4, coverage of 100% and strength 44.4%. This rule is supported by the classes C_1, C_4, and can be deduced as follows: "if capacity for innovation is acceptable and service capability is unpleased then the system activity is still acceptable".

The class $C_3 = \{\omega_3, \omega_5, \omega_9\}$ is not in $P_d(R)$, and a rule like ($f_1 = 1$) and ($f_2 = 0$) → (d = 0 or 1) may not be considered. Because, when it was used, this rule would be useless, since it receives nothing in decision.

Decision rules	Coverage (%)	Supported by
1. $(f_1 = 0) \wedge (f_2 = 1) \to (d = 0)$	50.0	C_2: ω_2, ω_7
2. $(f_1 = 1) \wedge (f_2 = 1) \to (d = 1)$	40.0	C_1: ω_1, ω_8
3. $(f_1 = 1) \wedge (f_2 = 2) \to (d = 1)$	40.0	C_4: ω_4, ω_6

Table 2. List of extracted decision rules.

The method of decision-making is also applied to build up decisions for risk warning based on processing historical data. Risk management model includes three sequential basic steps, that are risk identification, risk measurement, and risk warning. Risk identification should be objective itself, all risk levels are assessed by experts based on their work experience, this method ignores the role of historical data. That model does not have enough consideration on the uncertain and imprecision of risk. Alternatively, that method will unavoidably lead to some faulty judgments.

Data to identify risk factors often come from the operation, policy, environment, and management of a system. Collected data including a feature to assess risks are described by the feature d in a DIT. This decision feature d is often of six levels, 0: no risk, 1: little, 2: low-grade, 3: middle-grade, 4: distinct, and 5: dangerous. The historical data are collected factually, so there will be some data fields or features which have less impact on the final risk level. If these redundant features are removed, then there will be produced a simplified feature set which will have a positive impact on risk judgment. Where is the place of finding reduced feature set to ignore unnecessary information while the nature of collected data is still unchanged.

Based on fact-finding of conditional features and observed risk levels on DIT, decision rules to predict risk levels are extracted. This process is only a step of the training stage in machine learning. To improve quality of risk prediction, more observations on DIT and verifications of rules must be done repeatedly.

Example 6: To evaluate security risks of a system, three conditional feature types of the system come from environmental impact, management structure, and control equipment are taken into account. These conditional features are notated as E, M, and C, respectively, and the decision feature d is simplified at two levels, either 1: risk-warning or 0: no-warning. Data are shown in **Table 3**.

From **Table 3**, there are five equivalence classes $C_1 = \{\omega_1\}$, $C_2 = \{\omega_2, \omega_5\}$, $C_3 = \{\omega_3\}$, $C_4 = \{\omega_4\}$, $C_5 = \{\omega_6\}$ and two decision classes $D_1 = \{\omega_4, \omega_5\}$, $D_2 = \{\omega_1, \omega_2, \omega_3, \omega_6\}$.

Using Eqs. (4)–(6), the information entropy of F = {E, M, C} is H(F) = 2.2516, H(d) = 0.9183 and mutual information between F and d I(F, d) = 0.5850. From Eq. (6), $I(F - \{C\}, d) = 0.1258$ less than I(F, d), then a_3 is a core feature with a significance of Sgnf(C, d) = 0.4591.

	E	**M**	**C**	**d**
ω_1	0	1	1	1
ω_2	1	0	1	1
ω_3	1	1	2	1
ω_4	0	1	0	0
ω_5	1	0	1	0
ω_6	0	1	2	1

Table 3. Risk warning data.

Consider F-{M} = {E, C}, from Eq. (5), $H(d|F-\{M\}) = 0.3333$ implies to $I(F-\{M\}, d) = 0.5850 = I(F, d)$. Therefore, {E, C} is a coeffect reduced set of F. Hence, there are formally two decision rules:

$$[(E=0)\wedge(C=0)]\vee[(E=1)\wedge(C=1)]\rightarrow(d=0) \tag{17}$$

$$[(E=1)\wedge(C\#0)]\vee[(E=0)\wedge(C\#0)]\rightarrow(d=1) \tag{18}$$

It is noticed that the first expression of the second disjunction is an implication of the second one in the first rule. Therefore, maybe [(E = 1) ∧ (C = 1)] → [(d = 0) or (d = 1)] happens. Alternatively, the second rule can be written as (C # 0) → (d = 1). However, if E = 1 and C = 1, the first rule gives d = 0 contrary to the just deduced rule. For these reason, the above rules are chosen reasonably as [(E = 1) ∧ (C = 2)] ∨ [(E = 0) ∧ (C≠ 0)] → (d = 1).

Similarly, F-{E} = {M, C} gives $I(F-\{E\}, d) = I(F, d)$, thus {M, C} is also a reduced set of F. Then,

$$[(M=1)\wedge(C=0)]\vee[(M=0)\wedge(C=1)]\rightarrow(d=0) \tag{19}$$

$$[(M=1)\wedge(C\neq0)]\vee[(M=0)\wedge(C=1)]\rightarrow(d=1) \tag{20}$$

It is also noticed that the second expressions of the above disjunctions are identical and it is necessary to ignore them. Because, if (M = 0) ∧ (C = 1) is true, these rules simultaneously imply d = 0, 1 hard to decide.

Consequently, the second and fourth rules in **Table 4** may be used for risk warning w.r.t the collected data in **Table 3**.

The difficulties in choosing decision rules will be increasing with large-scale datasets. To reduce in part this shortcoming and make decision rules more efficiently, techniques of machine learning should be used. For instance, in [11], a back propagation neural network was used for training data in DIT, verifying decision rules in a number of steps to minimize errors in prediction based on decision rules.

Decision rules for risk warning	Coverage (%)	Strength (%)
1. [(E = 0) ∧ (C = 0)] → (d = 0)	50.0	20.0
2. [(E = 1) ∧ (C = 2)] ∨ [(E = 0) ∧ (C ≠ 0)] → (d = 1)	75.0	60.0
3. [(M = 1) ∧ (C = 0)] → (d = 0)	50.0	20.0
4. [(M = 1) ∧ (C≠ 0)] → (d = 1)	75.0	60.0

Table 4. List of extracted decision rules for risk warning.

3. Evaluation of the extent of MIS using ANOVA

For the outcome extent of an MIS, it is assumed that a reduced set of m features, namely f_1, f_2, …, f_m, is considered and evaluated with real numbers. The probability distribution of f_i is assumed that normal $N(\xi_i, \sigma_i^2)$ with expected mean ξ_i and variance σ_i^2.

ANOVA or analysis of variance was derived based on the approach in which the statistical method uses the variance to determine the expected means whether they are different or equal. It assesses the significance of factors, the so-called features here, by comparing the response means of observation samples at different features. In this chapter, ANOVA with single stage and multiple stages are introduced to evaluate features from the extent of an MIS.

In doing ANOVA, it is also assumed that all m features f_i are of the same variances. In a course of consideration, m observation samples at different features are randomly drawn. The ith sample is denoted by $\{\omega_{ij}\}$, j = 1, 2,…, n_i, a manifestation of a random variable f_i from the population of f_i values. Basic characteristics of the ith sample are:

$\overline{\omega}_i = \left(\sum_{j=1}^{n_i} \omega_{ij}\right)/n_i$—sample average, is an estimate for μ_i,

$s_i^2 = \left(\sum_{j=1}^{n_i} \left[\omega_{ij} - \overline{\omega}_i\right]\right)^2/df_i$—sample variance, estimate for σ^2 with degree of freedom $df_i = n_i - 1$.

These calculations are done by using the following three basic sums:

Sum:

$$S_i = \sum_{j=1}^{n_i} \omega_i \tag{21}$$

Sum of squares:

$$SS_i = \sum_{j=1}^{n_i} \omega_i^2 \tag{22}$$

Sum of squares of derivations:

$$SS_i = \sum_{j=1}^{n_i} \left[\omega_{ij} - \overline{\omega}_i\right]^2 \tag{23}$$

Then, it is implied that $\varpi_i = S_i/n_i$ and $SSD_i = SS_i - S_i^2/n_i$, so $s^{*}_i{}^2 = SSD_i/df_i$.

To verify condition that all variance σ_i^2 are equal to the same value σ^2, the Bartlett test based on the χ^2 probability distribution is used at a level of significance α valued from 1 to 5%. If the hypothesis on the equality of all variances is correct, m > 1 and $n_i > 1$ for all i, Bartlett has shown that the statistic χ^2_{cal} has approximately a χ^2-distribution with df = m−1:

$$\chi^2{}_{cal} = 2.3026\left(df \times \log s^2 - \sum_{i=1}^{m} df_i \log s_{*i}^2\right)/c \tag{24}$$

Here, $df = \sum_{i:1..m} df_i$, $c = 1+(\sum_{i:1..m} 1/df_i - 1/df)/[3(m-1)]$, $s^2 = (\sum_{i:1..m} df_i \times s_{*i}{}^2)/df = (\sum_{i:1..m} SSD_i)/df$ is the pool variance, an estimate for σ^2. If a calculated $\chi^2{}_{cal}$ is less than $\chi^2{}_{1-\alpha}$-percentile, it is unreasonable to deny that all variances are the same. It is noticed that the approximation χ^2-distribution is a poor one for $df_i \leq 2$.

In case of $n_1 = n_2 = \ldots = n$, then df = n−1 and Eq. (21) can be quite simple. Indeed, because of $\log s^2 = \log\sum_{i:1..m} SSD_i - \log(df)$ and $\log \underline{s}_i^2 = \log SSD_i - \log(df_i)$, a shortened form of Eq. (24) is

$$\chi^2{}_{\text{cal}} = 2.3026\left(\text{m} \times \text{logs}^2 - \sum_{\text{i}=1}^{\text{m}} \log \text{s}_{*\text{i}}^2\right)\text{df}/\text{c} \tag{25}$$

where, c = 1 + (m + 1)/(3 m[n−1]). The value $\chi^2{}_{\text{cal}}$ in Eq. (25) is calculated by using only all SSDs.

Setting $\text{n} = \sum_{\text{i:1..m}} \text{n}_\text{i}$, $\omega_\text{o} = (\sum_{\text{i:1..ni}} \text{n}_\text{i}\varpi_\text{i})/\text{n}$, $\xi_\text{o} = (\sum_{\text{i:1..ni}} \text{n}_\text{i}\xi_\text{i})/\text{n}$ and $\eta_\text{i} = \xi_\text{i} - \xi_\text{o}$. It is shown the following partitions

$$\begin{aligned}\sum_{\text{i}=1}^{\text{m}} \sum_{\text{j}=1}^{\text{n}_\text{i}} \left[\omega_{ij} - \xi_\text{i}\right]^2 &= \sum_{\text{i}=1}^{\text{m}} \sum_{\text{j}=1}^{\text{n}_\text{i}} \left[\omega_{ij} - \overline{\omega}_\text{i}\right]^2 + \sum_{\text{i}=1}^{\text{m}} \text{n}_\text{i}[\overline{\omega}_\text{i} - \xi_\text{i}]^2 \\ &= \sum_{\text{i}=1}^{\text{m}} \sum_{\text{j}=1}^{\text{n}_\text{i}} \left[\omega_{ij} - \overline{\omega}_\text{i}\right]^2 + \sum_{\text{i}=1}^{\text{m}} \text{n}_\text{i}\left[\overline{\omega}_\text{i} - \omega_\text{o} - \eta_\text{i}\right]^2 + \text{n}[\omega_\text{o} - \xi_\text{o}]^2\end{aligned} \tag{26}$$

According to the χ^2-partition theorem, the sums in the rightmost side of Eq. (26) are of χ^2-distribution with degrees of freedom n−m, m−1, 1, respectively.

If the expected means of m populations are the same, $\xi_\text{i} = \xi_\text{o}$ and $\eta_\text{i} = 0$ for all i. The two first terms of Eq. (26) are variations *within* or *between* samples and determined in turn as follows:

$$\text{s}_1^2 = \left(\sum_{\text{i}=1}^{\text{m}} \sum_{\text{j}=1}^{\text{n}_\text{i}} \left[\omega_{ij} - \overline{\omega}_\text{i}\right]^2\right)/(\text{n} - \text{m}) = \left(\sum_{\text{i}=1}^{\text{m}} \text{SSD}_\text{i}\right)/(\text{n} - \text{m}) \tag{27}$$

$$\text{s}_2^2 = \left(\sum_{\text{i}=1}^{\text{m}} \text{n}_\text{i}[\overline{\omega}_\text{i} - \omega_\text{o}]^2\right)/(\text{m} - 1) = \left(\sum_{\text{i}=1}^{\text{m}} \text{S}_\text{i}^2/\text{n}_\text{i} - \left[\sum_{\text{i}=1}^{\text{m}} \text{S}_\text{i}\right]^2/\text{n}\right)/(\text{m} - 1) \tag{28}$$

The statistics $\text{s}_1{}^2$, $\text{s}_2{}^2$ and $\text{s}_3{}^2 = \text{n}[\omega_\text{o} - \xi_\text{o}]^2$ are unbiased estimates of σ^2. In this case, the total variance between observations and population is determined as follows:

$$\text{s}^2 = \left(\sum_{\text{i}=1}^{\text{m}} \text{n}_\text{i}\left[\omega_\text{ij} - \omega_\text{o}\right]^2\right)/(\text{n} - 1) = \left(\sum_{\text{i}=1}^{\text{m}} \text{SS}_\text{i} - \left[\sum_{\text{i}=1}^{\text{m}} \text{S}_\text{i}\right]^2/\text{n}\right)/(\text{n} - 1) \tag{29}$$

In such a case, the variance ratio $\boldsymbol{v}^2{}_{\text{cal}} = \text{s}_1{}^2/\text{s}_2{}^2$ is of the Fisher probability distribution with $\text{df}_{\text{s1}} = \text{n} - \text{m}$, $\text{df}_{\text{s2}} = \text{m} - 1$. Therefore, the hypothesis about equality of m expected means is tested using the Fisher distribution with a given level of significance α valued from 1 to 5%. If $\boldsymbol{v}^2{}_{\text{cal}} > \text{F}_{1-\alpha}(\text{df}_{\text{s1}}, \text{df}_{\text{s2}})$, the hypothesis of equal means would be rejected, in which $\text{F}_{1-\alpha}(\text{df}_{\text{s1}}, \text{df}_{\text{s2}})$ is the $100(1 - \alpha)$% percentile of the Fisher distribution.

It is noticed that the condition m > 1 and, for all i, $\text{n}_\text{i} > 1$ are essential not only for Bartlett test, but also for doing ANOVA [12]. Conversely, the analysis is trivial when $\text{n}_\text{i} = 1$ for some i. Also, if m = 1, the analysis is pure inference from single population [13].

Example 7: Assume that there are four features need to be tested at the 5% level of significance with data in **Table 5**. Calculations are given in **Table 5**.

Using Eq. (24), $\chi^2{}_{\text{cal}} = 1.328$ is far less than $\chi^2{}_{0.95}(3) = 7.815$, the 95% percentile in the table of χ^2 probabilities with df = 3. Therefore, the hypothesis on equality of variances is accepted. The variation between dataset is estimated by the pool variance, Eq. (29), $\text{s}^2 = 36.3/9 = 4.037$. Using the underlined numbers in **Table 5**, the ANOVA table is presented in **Table 6**.

Features f_i	f_1	f_2	f_3	f_4	
ω_{i1}	7	5	8	7	
ω_{i2}	3	4	3	4	
ω_{i3}	4	6	5	2	
ω_{i4}				5	
{1}. ni.	3	3	3	4	13
Si	14	15	16	18	*63*
SSi	74	77	98	94	343
Si^2/fi	65.33	75	85.33	81	306.67
SSDi	8.667	2	12.67	13	*36.333*
{2}. dfi.	2	2	2	3	9
1/dfi	0.5	0.5	0.5	0.333	1.833
$\underline{si}^2$	4.333	1	6.333	4.333	
$\log(\underline{si}^2)$	0.637	0	0.802	0.637	
$df_i.\log(\underline{si}^2)$	1.274	0	1.603	1.91	4.787
$s^{2:}$	4.037	c.:	1.157	χ^2_{cal}:	*1.328*
$df.\log(s^2)$	5.455	$\Sigma(Si^2/ni)-(\Sigma Si)^2/n$:			*1.359*

Table 5. Calculations for single-stage ANOVA.

Variation sources	**SSD**	**df**	$\mathbf{s^2}$	v^2
Between features	1.359	3	0.453	0.11
Within features	36.333	9	4.037	
Total	37.692	12	$F_{0.95}(3,9) = 3.86$	

Table 6. Single-stage ANOVA table of Example 7.

The calculated basic sums in the first part of **Table 5** are used to set up an ANOVA in **Table 6**. It is shown that $v^2_{cal} = 0.453/4.037 = 0.112 < 3.86$, the 95% percentile in the table of Fisher probabilities w.r.t $\alpha = 5\%$. The hypothesis on equality of the expected means would be accepted at the 5% significance level.

If the hypothesis $\xi_i = \xi_2 = \ldots = \xi_m$ is rejected, all possible differences of these means in form of linear combinations are estimated by using confidence intervals. In such a case, there is a probability of $1 - \alpha$ that all comparisons simultaneously among the expected means satisfy:

$$-\lambda < \sum_{i=1}^{m} \delta_i \overline{\omega}_i - \sum_{i=1}^{m} \delta_i \xi_i < \lambda \tag{30}$$

Here, $\sum_{i=1\ldots m} \delta_i = 0$ and $\lambda^2 = s^2 \times F_{1-\alpha}(m-1, n-k) \times (m-1) \times \sum_{i=1\ldots m} (\delta_i^2/n_i)$, $F_{1-\alpha}(m-1, n-k)$ is the $100(1-\alpha)\%$ percentile of the Fisher probability distribution.

For instance, if m = 3, n = 4, $\varpi_1 = 2.25$, $\varpi_2 = 4.0$, $\varpi_3 = 4.5$ and $s^2 = 4.41$, then $F_{0.95}(2,3 \times 4 - 3) = 4.26$. Using Eq. (30), some 95% confidence intervals are calculated as follows:

$-\delta_1 = 1 = -\delta_2$, $\delta_3 = 0$, $\lambda = 4.33$; the confidence interval of $\xi_1 - \xi_2$ is -1.75 ± 4.297 or $(-2.55, 6.47)$.

$-\delta_1 = 0$, $\delta_2 = 1 = -\delta_3$; similarly, the confidence intervals of $\xi_2 - \xi_3$ is -0.5 ± 4.297 or $(-3.797, 4.797)$.

$-\delta_1 = ½ = \delta_2$, $\delta_3 = -1$, $\lambda = 3.721$. The 95% confidence interval of $½\xi_1 - ½\xi_2 - \xi_3$ is $(-2.436, 5.096)$.

When having several stages need to be tested on equality with expected means of features, multiple-stage ANOVA is applied. This is the case of evaluating the same given m features in k different stages, denoted by $\Gamma_\nu \nu = 1, 2, \ldots, k$. To simplify in presentation, without loss generality, it is assumed that all observed samples in stages have the same size, i.e., $n_i = n$ for all i, and Eq. (25) is used for Bartlett test.

The notations are similar, but an index ν added to the observations in each νth stage. The sums in Eqs. (21)–(23) are renotated as $S_{\nu i}$, $SS_{\nu i}$, $SSD_{\nu i}$. Since, $\varpi_{\nu i} = S_{\nu i}/n$, $s_{\nu i}^2 = SSD_{\nu i}/(n-1)$ are the average and variance of sample of the νth stage. All computations with multistage are similar to the single-stage ANOVA. Then, the results from stage computations are combined as shown at the end part of **Table 7**, to form multistage ANOVA table.

Example 8: Given a two-stage dataset of three features in five first rows of **Table 7**, calculations are illustrated in the parts, notated as {1} and {2}, of the table which aim at presenting schemes for finding basic sums and terms of Bartlett test and ANOVA.

fi	**Stage 1**			**Stage 2**			**Sizes**
	$\mathbf{f_1}$	$\mathbf{f_2}$	$\mathbf{f_3}$	$\mathbf{f_1}$	$\mathbf{f_2}$	$\mathbf{f_3}$	
$\omega_{\nu\ ij}$ 1	5	7	6	9	8	10	k = 2
2	8	4	7	8	7	8	m = 3
3	6	6	5	8	5	7	n = 3
{1} S_{ij}	19	17	18	25	20	25	124
SS_{ij}	125	101	110	209	138	213	896
S_{ij}^2/n	120.33	96.33	108.00	208.33	133.33	208.33	874.67
SSD_{ij}	4.67	4.67	2.00	0.67	4.67	4.67	21.33
$logSSD_{ij}$	0.67	0.67	0.30	−0.18	0.67	0.67	2.80
{2}	Bartlett test:			ANOVA:			
(ΣlogSSDi)/(km)−log(n−1):			0.166	$S^2/(mn)$:		868.44	
S^2/df_i	874.67	$logs^2$	0.250	$S^2/(kmn)$:		854.22	*14.222*
SSD	21.333	c.	1.194	$(\Sigma(S_1i + S_2i)^2)/(km) - S^2/(kmn)$:			*4.778*
logSSDi	2.80	χ^2_{cal}	*1.945*	$SS_1 + SS_2 - S^2/(kmn)$:			*41.778*

Table 7. Calculations for Two-stage ANOVA.

Calculations for the Bartlett test in {2} of **Table 7** show that $\chi^2_{cal} = 1.194 < \chi^2_{0.95}(5) = 11.07$, the hypothesis that population variance is the same for all features is accepted at $\alpha = 5\%$. An estimate of the population variance is $s_1^2 = 21.33/(2 \times 3 \times [3–1]) = 1.778$, cf. **Table 8**. The part {3} of **Table 7** is the calculation scheme for the terms in **Table 8**, where Subtotal equals Total minus Within stages or the sum of Between features within stages, Between stages, and Interaction.

The ratio of the variation between stages to within features is $v^2 = s_3^2/s_1^2 = 14.222/1.778 = 8.0$ which by far exceeds the 95% percentile of Fisher distribution $F_{0.95}(1,12) = 4.75$. That means the difference of the expected means between stages is different significantly. In other words, the effects between stages are significantly discriminated.

Similarly, in comparison of the variation within features and between features within stages, **Table 8** shows that $v^2 = s_2^2/s_1^2 = 3.105/1.778 = 1.747 < F_{0.95}(2,12) = 3.89$. This shows that the difference between the expected means of features within stages is not significant or the effects between features within stages are almost the same.

Beside the above effects, the interaction between stages and features is also a factor need to be considered. The ratio $v^2 = 0.006/0.012 = 0.50$ gives that such an interaction is not present in given dataset. Thus, both the lines labeled "*Interaction*" and "*Within stages*" give the same unbiased estimates of σ^2, since a combination of these lines can improve the estimate of σ^2. The residual mean square is a sum of variations between the Interaction and Within stages. This leads to an updated population variance is 1.525 less than $s_1^2 = 1.778$ in **Table 8**, but obviously increases v^2 ratios. **Table 9** analyzes the interaction without stage of Example 8.

Variation sources	**SSD**	**df**	$\mathbf{s^2}$	$\boldsymbol{v^2}$
Between stages	14.222	1	14.222	8.0
Between features within stages	6.210	2	3.105	1.747
Interaction	0.012	2	0.006	0.50
Subtotal	20.444	5		
Within stages	21.333	12	1.778	
Total	41.778	17		

Table 8. Two-stage ANOVA table of Example 8.

Variation sources	**SSD**	**df**	$\mathbf{s^2}$	$\boldsymbol{v^2}$
Between stages	14.222	1	14.222	9.328
Between features within stages	6.210	2	3.105	2.036
Residual mean square	21.345	14	1.525	
Total	41.778	17		

Table 9. ANOVA table—two-stage without interaction.

http://dx.doi.org/10.5772/intechopen.74093

The ratio $v^2 = s_2^2/s_1^2 = 3.105/1.525 = 2.036 < F_{0.95}(2,14) = 3.74$ or the effects between features within stages are the same. While, $v^2 = s_3^2/s_1^2 = 14.222/1.525 = 9.328$ which also by far exceeds $F_{0.95}(1,14) = 4.60$, the effects between stages are also significantly discriminated, cf. **Table 8**.

4. Case studies

To evaluate the extent to which MIS is being used to attain achievements of long-term planning, short-term planning in the South-West Nigerian universities [14], all selected features are f_1: Construction of building in the university, f_2: Student enrolment projection, f_3: Manpower projection, f_4: Staff recruitment exercises, f_5: Establishment of new faculties and department, f_6: Designing university academic program, f_7: Stock library with books and journals are considered in long-term evaluation. For short-term, f_1: Promotion of Staff, f_2: Staff Training and Development, f_3: Appointment of Deans or Heads of Departments or Divisions, f_4: Appointment of Committee Members, f_5: Allocation of offices to staff, f_6: Allocation of Residential Quarters, f_7: Allocation of Lecture room/theaters, f_8: Full-time equivalent or Teacher-Students Ratio, and f_9: Maximum Teaching Load are considered.

In evaluation of the extent of our university for a 5-year strategy planning, the following features are used f_1: Effectuation rights and obligations of students, f_2: Promotion of international cooperations, f_3: Library, equipment and material facilities, f_4: Potential of Scientific R&D and transfer of technology, f_5: Capacity of organization and management, f_6: Design of university academic programs, f_7: Promotion of academic operations, f_8: Capacity of manpower projection, f_9: Management of finance and resources. These basic features are factors to evaluate whether the university attains its goal and objectives. Each basic feature is evaluated in the scale of 100 but here it is illustrated in the one of 20 points.

Example 9: Let f_i, i = 1, 2,…, 9 be features characterized as the extent of an MIS as above. ω_{ij}, j = 1, 2,…, 12 is a value that is evaluated as the ith feature by the jth evaluator in a shorten marking scheme of 20. Calculations for the single-stage ANOVA table are shown in **Table 10**.

The calculated value $\chi^2_{cal} = 9.432$ in **Table 10** does not exceed $\chi^2_{0.95}(8) = 15.51$, the hypothesis on equality of variances is accepted. The population variance is estimated as $s^2 = 1185.58/99 = 11.976$. The corresponding ANOVA table for this dataset is given in **Table 11**.

Here, as variance ratio $v^2 = 8.907$ far exceeds $F_{0.95}(8,99) = 2.06$, it is unreasonable to assume that all the expected means of features are the same. This can also be seen from **Table 10**, where all sum of features from f_1 to f_4 are less than the ones of features from f_5 to f_9.

A more detailed examination revealed that the nine features can be partitioned into two groups, namely A = {f_1, f_2, f_3, f_4} with the first four features and B = {f_5, f_6, f_7, f_8, f_9} with the remainders. Each group of features can be seen as a treatment and its observation sample includes all observations in the same group. Since, it would be reasonable to consider the variation between features into three portions between: the features from A, the features from B, and between group A and B. Calculations in this consideration are extracted from **Table 10** and illustrated in **Table 12**.

	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9
ω_{i1}	13	2	10	2	14	10	9	9	11
ω_{i2}	1	1	3	13	11	10	10	10	14
ω_{i3}	7	5	0	11	13	7	11	15	15
ω_{i4}	9	15	2	17	10	7	11	13	15
ω_{i5}	3	8	9	7	17	9	7	17	20
ω_{i6}	6	2	5	5	14	10	11	15	17
ω_{i7}	7	2	7	2	13	11	15	11	5
ω_{i8}	10	3	10	10	11	13	13	11	15
ω_{i9}	8	13	7	7	10	11	11	8	17
ω_{i10}	11	6	9	8	9	9	10	14	15
ω_{i11}	5	2	7	8	3	14	7	10	15
ω_{i12}	7	3	3	10	9	15	11	13	11
{1} Si	87	62	72	100	134	126	126	146	170
SSi	753	554	556	1038	1632	1392	1378	1860	2566
Si^2/ni	630.75	320.33	432	833.3	1496.3	1323	1323	1776.3	2408.3
SSDi	122.25	233.67	124	204.7	135.67	69	55	83.667	157.67
logSSDi	2.087	2.369	2.09	2.311	2.133	1.839	1.740	1.923	2.20
{2} Bartlett		df:8				Anova			
$\Sigma logs_i^2/k$:		1.036	c:	1.034	ΣSi:	1023	$\Sigma SSDi$:		*1185.58*
$log.s^2$:		1.078	χ^2_{cal}	*9.432*	$\Sigma Si^2/fi$:	10,543	$\Sigma Si^2/n - S^2/nm$:		*853.33*

Table 10. Calculations for single-stage ANOVA dataset.

Variation sources	**SSD**	**df**	$\mathbf{s^2}$	$\boldsymbol{v^2}$
Between feature	853.333	8	106.667	8.907
Within features	1185.583	99	11.976	
Total	2038.917	107		

Table 11. Single-stage ANOVA table of Example 9.

In comparison with the variance within features s^2, the variance ratios v^2 = 23.243/11.976 = 1.941 < $F_{0.95}(3,99)$ = 2.66 and v^2 = 113.60/11.976 = 2.371 < $F_{0.95}(4,99)$ = 2.43 in **Table 13** show that there is no essential difference between features in the same group. Since the third ratio v^2 = 183.33/11.976 = 15.309 is far greater than $F_{0.95}(1,99)$ = 3.9, the features in group A and B do have different expected mean.

Example 10: Assume that in an MIS, there are two stages that need ANOVA with the same set of features. In each stage, samples of evaluations in marking scheme of 20. Let ω_{vij} be an integral value in marking scheme of 20 that evaluates the ith feature given by the jth evaluator

	Group A		**Group B**		**n = 12**	
n_*	4 × 12		5 × 12		108	
ΣS_i	321	6.6875	702	11.7	1023	
$\Sigma S_i^2/n_i$		2216.4		8327		10,360
$S^2/\underline{S}(ni)$		2146.7		8213.4		9690.1
SSD		*69.729*		*113.6*		*183.33*
Ave.:	6.688		11.7		9.472	

Table 12. Calculations for ANOVA between group A and B.

Variation sources	**SSD**	**df**	$\mathbf{s^2}$	$\boldsymbol{v^2}$
Between features from A	69.729	3	23.243	1.941
Between features from B	113.60	4	28.40	2.371
Between features in A and B	183.33	1	183.33	15.309
Total	853.33	8		

Table 13. ANOVA table of two groups A and B.

from the νth stage, ν = 1, 2, i = 1, 2,…, 7, j = 1, 2,…, 8. This dataset is in **Table 14** including calculation for ANOVA.

Using Bartlett test in {3}, $\chi^2_{cal} = 9.507$ not exceed $\chi^2_{0.95}(15) = 22.36$, so population variances are the same with the pool variance of $s^2 = 1185.58/96 = 2.105$. **Table 15** shows this ANOVA.

The ratio $v^2 = s_2^2/s_1^2 = 3.932/2.063 = 1.906$ is less than $F_{0.95}(6,98) = 2.15$, the difference between the expected means within stages is not significant. Similarly, $v^2 = s_3^2/s_1^2 = 37.723/2.063 = 18.29 > F_{0.95}(1,98) = 3.96$, the expected means between stages are discriminated.

Since $v^2 = 0.057/0.339 = 0.167$ less than the 95% percentile of Fisher distribution, any interaction does not exist. Thus, "Interaction" and "Within stages" variation sources are combined to $s_1^2 = (202.125 + 0.339)/104 = 1.947$ a better estimation for σ^2 than 2.063 in **Table 15**.

The case of m = 1 and k = 2 has been presented in the previous subsection with group A, B. In [15], ANOVA has been used to specify whether a statistical relationship exists between human development index and security index. The authors in [16] have used the ANOVA combined with regression analysis to assess and evaluate student MIS of a university.

In this subsection, the student test is presented in comparison with the effects of f from the two stages or treatments. Let $\{\omega_{ij}\}$ i = 1, 2 and j = 1, 2,…, n_i be two observation samples of sizes n_i drawn from the two treatments of the feature f. Using Eqs. (21)–(23), the means ϖ_1, ϖ_2 and variances s_1^2, s_2^2 are calculated with $df_1 = n_1–1$, $df_2 = n_2–1$.

The equality of population variances is tested using Fisher distribution with $v^2 = s_1^2/s_2^2$. If $v^2 < F_{\alpha/2}(df_1, df_2)$ or $v^2 > F_{1-\alpha/2}(df_1, df_2)$, it is unreasonable to assert that the population

	f_1	f_2	f_3	f_4	f_5	f_6	f_7	
ω_{1i1}	16	15	17	14	14	13	14	k = 2
ω_{1i2}	13	14	11	12	10	13	12	m = 7
ω_{1i3}	14	16	15	14	15	14	16	n = 8
ω_{1i4}	12	14	13	12	14	12	15	
ω_{1i5}	13	15	14	10	13	14	13	
ω_{1i6}	10	12	11	13	12	10	12	
ω_{1i7}	12	14	13	14	13	13	12	
ω_{1i8}	13	14	13	15	13	14	13	
{1} S_1i	103	114	107	104	104	103	107	742
SS_1i	1347	1634	1459	1370	1368	1339	1447	9964
S_1i^2/ni	1326	1625	1431	1352	1352	1326	1431.1	9843
SSD_1i	20.88	9.5	27.88	18	16	12.88	15.88	121
$logSSD_1i$	1.32	0.978	1.445	1.255	1.204	1.11	1.201	8.512
	f_1	f_2	f_3	f_4	f_5	f_6	f_7	
ω_{2i1}	16	17	18	14	15	14	15	
ω_{2i2}	14	15	13	12	11	13	14	
ω_{2i3}	15	18	15	14	16	15	16	
ω_{2i4}	13	14	14	15	14	13	14	
ω_{2i5}	14	16	15	13	15	14	13	
ω_{2i6}	13	14	13	15	14	14	15	
ω_{2i7}	14	15	16	14	13	15	14	
ω_{2i8}	13	14	14	16	14	15	15	
{2} S_2i	112	123	118	113	112	113	116	807
SS_2i	1576	1907	1760	1607	1584	1601	1688	11,723
S_2i^2/ni	1568	1891	1741	1596	1568	1596	1682	11641.9
SSD_2i	8	15.88	19.5	10.88	16	4.875	6	81.125
$logSSD_2i$	0.903	1.201	1.29	1.036	1.204	0.688	0.778	7.101
$(S_1i + S_2i)^2$	46,225	56,169	50,625	47,089	46,656	46,656	4973	343,149
{3} Bartlett					Anova		$S=S_1 + S_2$	1549
$\Sigma logSSDi./(km)-log(n-1)$:				0.270	$S^2/(mn)$:		21460.9	
$\Sigma SSDi$:		*202.1*	$logs^2$:	0.314	$S^2/(kmn)$:		21423.2	*37.723*
$\Sigma logSSDi$:		15.61	c:	1.051	$\Sigma(S_1i + S_2i)^2/(km) - S^2/(kmn)$:			*23.589*
			χ^2_{cal}:	*9.507*	$SS_1 + SS_2 - S^2/(kmn)$:			*263.77*

Table 14. Calculations for two-stage ANOVA dataset.

variances are equal. Otherwise, the pool variance of these treatments is $s^2 = (SSD_1 + SSD_2)/(df_1 + df_2)$.

Variation sources	SSD	df	s^2	v^2
Between stages	37.723	1	37.723	18.290
Between features within stages	23.589	6	3.932	1.906
Interaction	0.339	6	0.057	0.167
Subtotal	61.652	13		
Within stages	202.125	98		
Total	263.777	111		

Table 15. Two-stage ANOVA table of Example 10.

The equality of the expected means from treatments is tested by the student distribution based on the difference $\varpi_0 = \varpi_1 - \varpi_2$. If this hypothesis is correct, there are two cases:

- If the variances in each treatment are equal, the statistics $t_{cal} = \varpi_0/s_o$ with ${s_o}^2 = s^2[1/n_1 + 1/n_2]$ has the student distribution $df = df_1 + df_2$ degrees of freedom,
- If the variances of treatments not equal, $t_{cal} = \varpi_0/s_o$ with ${s_o}^2 = {s_1}^2/n_1 + {s_2}^2/n_2$ approximate the student distribution with $df = c^2/df_1 + (1-c^2)/df_2$, where $c = ({s_1}^2/n_1)/({s_1}^2/n_1 + {s_2}^2/n_2)$.

The hypothesis that the two expected means of the feature f from the treatments are equal is rejected at a level of significance α when $|t_{cal}| > t_{1\ -\ \alpha/2}(df)$. Otherwise, the confidence interval of the difference η between the two means is

$$\varpi_0 + t_{\alpha/2}(df)s_o < \eta < \varpi_0 + t_{1-\alpha/2}(df)s_o \tag{31}$$

where $t_{1\ -\ \alpha/2}(df)$ is the $100(1-\alpha/2)\%$ percentile of the student distribution, $t_{1-\alpha/2}(df) = -t_{\alpha/2}(df)$.

For instance, from **Table 12**, the variances in groups ${s_A}^2 = 69.729/47 = 1.484$ and ${s_B}^2 = 113.60/59 = 1.925$ give $v^2 = {s_A}^2/{s_B}^2 = 1.30$ less than $t_{0.995}(106) = 2.35$. It is accepted the variances in group A and B are equal. The pool variance is estimated by $s^2 = (SSD_1 + SSD_2)/(df_1 + df_2) = 113.60/106 = 1.729$. Also, **Table 12** gives ${s_o}^2 = s^2.[1/47 + 1/59] = 0.06611$ and $t_{cal} = (11.7–6.6875)/s_o = 19.68$, this so far exceeds $t_{0.995}(106) = 2.606$. The student test for these two treatment shows the expected mean of group B so far exceeds the one of A. The 99.5% confidence interval of the difference between these expected means is $11.7 - 6.6875 \pm 2.606 \times \sqrt{0.06611}$ or (4.342, 5.683).

Similarly, **Table 15** shows that there is no difference in evaluating features by evaluators within stages in Example 10. It is reasonable to group features in each stage to each other and using the method of comparison between two treatments of a feature as above.

5. Conclusion

It is dealt with this chapter the useful methods for choosing important features and supporting decisions of a given decision information system, presented in Section 2. The methods of ANOVA are introduced in Section 3 to evaluate features from the extent of an MIS.

The demonstrations of using such methods, through examples and case studies in Section 4 at our Faculty of Information System—University of Information Technology, showed that the efficiency of the proposed methods. The illustrated calculating schemes allow designing and coding computer programs for solving the above problems automatically.

Author details

Khu Phi Nguyen[1]* and Hong Tuyet Tu[2]

*Address all correspondence to: khunp@uit.edu.vn

1 Faculty of Information System, University of Information Technology, Vietnam National University, Ho Chi Minh City, Vietnam

2 Faculty of Information Technology, HCMC University of Technology and Education, Ho Chi Minh City, Vietnam

References

[1] Nguyen KP, Tu HT. Assessment waste water treatment process with in-completed dataset. In: Proceedings of the 20th National Conference on Fluid Mechanics. Vietnam; 2017

[2] Ciortea M. Aspects regarding the types of process control systems. International Conference on Theory and Applications of Mathematics and Informatics. 2004:90-95

[3] Heidarkhani A, Khomami AA, Jahanbazi Q, Alipoor H. The role of management information systems (MIS) in decision-making and problems of its implementation. Universal Journal of Management and Social Sciences. 2013;**3**(3):78-89

[4] Pawlak Z, Skowron A. Rough sets: Some extensions. Information Sciences. 2007;**177**:28-40

[5] Nguyen KP, Tu HT. Data mining based on rough set theory. In: Knowledge Discovery in Databases. USA: Academy Publisher; 2013

[6] Nguyen KP, Bui ST, Tu HT. Revenue evaluation based on rough set reasoning. In: Sobecki et al., editors. Advances Approaches to Intelligent Information and Data-base Systems. Thailand: Springer SCI 551; 2014

[7] Dai J, Xu Q, Wang W. A comparative study on strategies of rule induction for incomplete data based on rough set approach. International Journal of Advanced in Computing Technology. 2011;**3**(3):176-182

[8] Chen Y, Miao D, Wang R. A rough set approach to feature selection based on ant colony optimization. Elsevier Pattern Recognition Letters. 2010;**31**:226-233

[9] Liu Y, Esseghir M, Boulahia LM. Evaluation of parameters importance in cloud service selection using rough sets. Applied Mathematics, Scientific Research Publication. Mar. 2016;**7**:527-541

[10] Nguyen KP, Nguyen VL. Analysis of weather information system in statistical and rough set point of view. In: New Trends in Computational Collective Intelligence, ICCCI, 2014, Korea, 2014

[11] Shi L, Luo F. Research on risk early-warning model in airport flight area based on information entropy attribute reduction and BP neural network. International Journal of Security and Its Applications. 2015;**9**(10):313-322

[12] Using ANOVA. Available from: https://www.educba.com/interpreting-results-using-anova/. May 2016

[13] Ramachandran KM, Tsokos CP. Mathematical Statistics with Applications. San Diego, California, USA: Elsevier Academic Press; 2009

[14] Ajayi IA, Omirin FF. the use of management information systems (MIS) in decision making in the South-West Nigerian. Academy Journal: Educational Research and Review. May 2007;**2**(5):109-116

[15] Sow MT. Using ANOVA to examine the relationship between safety & security and human development. Journal of International Business and Economics;**2**(4):101-106, Published by American Research Institute for Policy Development, Dec 2014

[16] Fetaji B et al. Assessing and evaluating UBT model of student management information system using ANOVA. TEM Journal. Aug. 2016;**5**(3):313-318, ISSN 2217-8309. DOI: 10.18421/TEM53-10

IntechOpen